Stochastic models of gene transcription with upstream drives:
exact solution and sample path characterization

Justine Dattani1,† and Mauricio Barahona1

1Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom

Gene transcription is a highly stochastic and dynamic process. As a result, the mRNA copy number of a given gene is heterogeneous both between cells and across time. We present a framework to model gene transcription in populations of cells with time-varying (stochastic or deterministic) transcription and degradation rates. Such rates can be understood as upstream cellular drives representing the effect of different aspects of the cellular environment. We show that the full solution of the master equation contains two components: a model-specific, upstream effective drive, which encapsulates the effect of the cellular drives (e.g., entrainment, periodicity or promoter randomness), and a downstream transcriptional Poissonian part, which is common to all models. Our analytical framework allows us to treat cell-to-cell and dynamic variability consistently, unifying several approaches in the literature. We apply the obtained solution to characterize several gene transcription models of experimental relevance, and to explain the influence on gene transcription of synchrony, stationarity, ergodicity, as well as the effect of time-scales and other dynamic characteristics of drives. We also show how the solution can be applied to the analysis of single-cell data, and to reduce the computational cost of sampling solutions via stochastic simulation.

I. INTRODUCTION

Gene transcription, the cellular mechanism through which DNA is copied into mRNA transcripts, is a complex, stochastic process [1]. As a result, the number of mRNA copies for most genes is highly heterogeneous both within each cell over time, and across cells in a population [2–4]. Such fundamental randomness is biologically relevant: it underpins the cell-to-cell variability linked with phenotypic outcomes and cell decisions [5–9].

The full mathematical analysis of gene expression variability requires the solution of master equations. Given a gene transcription model, its master equation (ME) is a differential-difference equation that describes the evolution of $P(n,t)$, the probability of having $n$ mRNA molecules in a single cell at time $t$. However, MEs are problematic to solve, both analytically and numerically, due to the difficulties associated with discrete stochastic variables—the molecule number $n$ is an integer [10]. Indeed, most existing analytical solutions of the ME are specific to particular models, typically obtained via the probability generating function and under stationarity assumptions [11–17]. When analytical solutions are intractable, the first few moments of the distribution are approximated, usually at stationarity, but error bounds are difficult to obtain [18–19]. Alternatively, full stochastic simulations are used, although the computational cost to sample $P(n,t)$ at each $t$ is often impractical, and many methods lead to estimation bias in practice [20].

With the emergence of accurate measurements of time-courses of mRNA counts in single cells [3,4,21,22], the application of MEs to data analysis faces new challenges. Mathematically, ME models must be able to describe time-dependent gene transcription in single cells within a population, allowing for a degree of synchrony or cell-to-cell correlation. However, current stationary solutions tacitly assume that gene expression is uncorrelated between cells; hence most current analytical models cannot account for the dynamic variability due to upstream biological drives, such as chromatin remodeling [24], transcription factor binding [25], circadian rhythms and cell cycle [26,27], external signaling [28], or stimulus-induced modulation or entrainment [29,30]. Full solutions of the ME that capture temporal heterogeneity from the single-cell to the population level could help unravel how the dynamics of biological phenomena affect gene transcription and regulation [31]. These examples highlight the methodological need for solvable models flexible enough for the analysis of a range of datasets, thus enabling the formulation of hypotheses in conjunction with experiments.

Here, we consider a simple, yet generic, framework for the solution of the ME of gene transcription and degradation for single cells under upstream drives, i.e., when the transcription and degradation parameters can be time-dependent functions or stochastic variables. We show that the exact solution $P(n,t)$ for such a model naturally decouples into two parts: a discrete transcriptional Poissonian component, which is common to all transcription models of this kind, and a model-specific continuous component, which describes the dynamics of the parameters reflecting the upstream variation. To obtain the full $P(n,t)$ one only needs to calculate the probability density for the model-specific upstream drive, which we show to correspond to a continuous variable satisfying a linear random differential equation directly related to traditional differential rate equations of chemical kinetics. Below we present the properties of the general solution, including its moments and noise characteristics in terms of the moments of the upstream component. We also clarify the different effects of stochastic and deter-
ministic drives by considering the Fano factor across the population and across time. To illustrate the utility of our approach, we present analytical and numerical analyses of several models in the literature, which are shown to correspond to different upstream drives. Finally, we provide examples of its use for data analysis.

II. THE MASTER EQUATION FOR GENE TRANSCRIPTION IN POPULATIONS OF CELLS WITH UPSTREAM DRIVES

Notation and formulation of the problem

To study gene expression in a single cell with time-dependent upstream drives, we consider the stochastic process in continuous time $t$, $N = \{N_t \in \mathbb{N} : t \geq 0\}$, where $N_t$ is a discrete random variable describing the number of mRNA molecules in the cell. We look to obtain the probability mass function, $P(n,t) := \Pr(N_t = n)$.

The mRNA copy number increases via transcription events and decreases via degradation events but, importantly, the transcription and degradation rates can depend on time and can be different for each cell (Fig. 1). To account for such variability, we describe transcription and degradation rates as stochastic processes

$$\mu(t) \in \mathbb{R}_+ \quad \text{and} \quad \lambda(t) \in \mathbb{R}_+$$

without specifying any functional form except requiring that $M$ and $L$ do not depend on the number of mRNA molecules already present. Deterministic time-varying transcription/degradation rates, with or without cell-to-cell correlations, are a particular case of this definition.

Following standard notation in the stochastic processes literature, $M_t$ and $L_t$ denote the random variables at time $t$. To simplify notation, however, we depart from standard notation and denote the sample paths (i.e., realizations) of $M$ and $L$ by $\{\mu(t)\}_{t \geq 0}$ and $\{\lambda(t)\}_{t \geq 0}$, respectively, thinking of them as particular functions of time describing the transcription and degradation rates under the changing cellular state and environmental conditions in an ‘example’ cell (Fig. 1). The sample paths of other random variables are denoted similarly, e.g., the sample paths of $N_t$ are $\{N_t\}_{t \geq 0}$.

The sample paths $\{\mu(t)\}_{t \geq 0}$ and $\{\lambda(t)\}_{t \geq 0}$ represent cellular drives encapsulating the variability across time and across the population consistently. This formulation unifies several models in the literature, which implicitly or explicitly assume time-varying transcription and/or degradation processes [4 13 62 39], and can be shown to correspond to particular types of dynamic upstream variability. In addition, the framework allows us to specify cell-to-cell correlations across the population, which we refer to as the ‘degree of synchrony’. A population will be perfectly synchronous when the sample paths of the drives for every cell in the population are identical, i.e., if $M_t$ and $L_t$ have zero variance. If, however, transcription and/or degradation rates differ between cells, $M_t$ and $L_t$ themselves emerge from a probability density: the wider the density, the more asynchronous the cellular drives are (Fig. 1).

Our aim is to obtain the probability distribution of the copy number $N_t$ under upstream time-varying cellular drives $M_t$ and $L_t$, themselves containing stochastic parameters reflecting the cell-to-cell variability. To do this, we proceed in two steps: first, we solve the synchronous system without cell-to-cell variability; then we consider the general asynchronous case.

A. Perfectly synchronous population

As a first step to the solution of the general case, consider a population of cells with perfectly synchronous transcription and degradation rate functions, $M = \{\mu(t)\}_{t \geq 0}$ and $L = \{\lambda(t)\}_{t \geq 0}$; i.e., the transcription and degradation processes are defined by the same sample path for the whole population and the stochastic pro-
cesses $M$ and $L$ have zero variance at all times (Fig. 1).

In the perfectly synchronous case, we have an immigration-death process with reaction diagram:

$$
\emptyset \xrightarrow{\mu(t)} \text{mRNA} \xrightarrow{\lambda(t)} \emptyset,
$$

and its ME is standard:

$$
\frac{d}{dt} P(n, t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}) = \mu(t) P(n-1, t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]})
+ (n+1)\lambda(t) P(n+1, t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]})
- (\mu(t) + n\lambda(t)) P(n, t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}),
$$

where $P(n, t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]})$ denotes the probability of having $n$ mRNAs at time $t$ for the given history of the cellular drives $\{\mu(\tau)\}_{\tau \in [0,t]}$ and $\{\lambda(\tau)\}_{\tau \in [0,t]}$.

Using the probability generating function

$$
G(z, t) = \sum_{n=0}^{\infty} z^n P(n, t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}),
$$

we transform the ME (2) into

$$
\frac{\partial G}{\partial t} = (z - 1) \mu(t) G - (z - 1) \lambda(t) \frac{\partial G}{\partial z}.
$$

Without loss of generality, let us first consider an initial condition with $n_0$ mRNA molecules. Using the method of characteristics, we obtain the solution:

$$
G(z, t|n_0) = \left((z - 1)^t\int_0^t e^{-\int_\tau^t \lambda(\tau) d\tau} d\tau + 1\right)^{n_0} e^{\chi(t)(z-1)},
$$

which is given in terms of

$$
\chi(t) := \int_0^t \mu(\tau) e^{-\int_\tau^t \lambda(\tau) d\tau} d\tau.
$$

We will refer to the time-varying continuous function $\chi(t)$ as the effective drive, as it integrates the effect of both cellular drives.

Notice that the solution (3) can be rewritten as the product of two probability generating functions:

$$
G(z, t|n_0) = G_{\text{Bin}}(z, t|n_0) G_{\text{Poi}}(z, t),
$$

corresponding to a binomial and a Poisson distribution, respectively. Hence, for the perfectly synchronous case, the solution is given by:

$$
N_t^i = N_{t|n_0}^i | n_0 + N_t^s, \quad \text{with}
$$

$$
N_{t|n_0}^i \sim \text{Bin} \left(n_0, e^{-\int_0^t \lambda(\tau) d\tau}\right),
$$

$$
N_t^s \sim \text{Poi} (\chi(t)),
$$

where $N_{t|n_0}^i$ is a binomial random variable with $n_0$ trials and success probability $e^{-\int_0^t \lambda(\tau) d\tau}$, and $N_t^s$ is a Poisson random variable with parameter $\chi(t)$. The physical interpretation of this breakdown is that $N_t^i$ describes the mRNA transcripts that were initially present in the cell and still remain at time $t$, whereas $N_t^s$ describes the number of mRNAs transcribed since $t = 0$.

Since $N_{t|n_0}^i$ and $N_t^s$ are independent, it is easy to read off the first two moments directly:

$$
E [N_t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}, n_0] = E [N_{t|n_0}^i] + E [N_t^s]
$$

$$
= n_0 e^{-\int_0^t \lambda(\tau) d\tau} + \chi(t)
$$

$$
\text{Var}(N_t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}, n_0) = \text{Var}(N_{t|n_0}^i) + \text{Var}(N_t^s)
$$

$$
= n_0 e^{-\int_0^t \lambda(\tau) d\tau} \left(1 - e^{-\int_0^t \lambda(\tau) d\tau}\right) + \chi(t).
$$

From (5)-(7), the distribution is:

$$
P(n, t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}, n_0) = \sum_{k=0}^{n} \text{Pr}(N_{t|n_0}^i = k|n_0) \text{Pr}(N_t^s = n-k)
$$

$$
= \sum_{k=0}^{n} \left(\frac{n_0}{k}\right) \left(e^{-\int_0^t \lambda(\tau) d\tau}\right)^k \left(1 - e^{-\int_0^t \lambda(\tau) d\tau}\right)^{n-k} \frac{\chi(t)^{n-k}}{(n-k)!} e^{-\chi(t)}.
$$

If the initial state is itself described by a random variable $N_0$ with its own probability distribution, we apply the law of total probability to obtain the solution in full generality as (see Appendix A):

$$
P(n, t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}, n_0) = \sum_{n_0} P(n, t|\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}, n_0) \text{Pr}(N_0 = n_0)
$$

$$
= \sum_{k=0}^{n} \text{Pr}(N_t^s = n-k) \text{Pr}(N_{t|n_0}^i = k),
$$

where $N_t^s$ is distributed according to (5)-(7), and $\text{Pr}(N_{t|n_0}^i = k)$ is the mixture of the time-dependent bi-
nominal distribution \([5]\) and the distribution of the initial condition \(N_0\).

**The initial transient ‘burn in’ period.** For biologically realistic degradation rates \(\{X(t)\}_{t \geq 0}\), the contribution from the initial condition \((N_t^\text{ini})_t\) decreases exponentially. Hence as time grows the transcripts present at \(t = 0\) degrade, and the population is expected to be composed of mRNAs transcribed after \(t = 0\). If the initial distribution of \(N_0\) is not the stationary distribution of the ME (or, more generally, not equal to the attracting distribution of the ME, as defined in Appendix \([A]\), there is an initial time-dependence of \(P(n,t)\) lasting over a time scale \(T^\text{inc}\) (given by \(\int_0^{T^\text{inc}} \lambda(t) \, dt \approx 1\)), which corresponds to a ‘burn-in’ transient associated with the decay of the initial condition. For instance, the time-dependence described in Ref. \([13]\) for the random telegraph model (Fig. 6) corresponds to this ‘burn-in’ transient.

On the other hand, when the initial distribution of \(N_0\) is the stationary distribution (or the attracting distribution) of the ME, the component containing the initial condition \((N_t^\text{ini})_t\) and the long-term component \((N_t^\text{L})_t\) balance each other at every point in time, maintaining stationarity (or the attracting distribution), as shown analytically in Appendix \([A]\).

**The long-term behaviour of the solution.** In this work, we focus on the time dependence of \(P(n,t)\) induced through non-stationarity of the parameters and/or correlated behaviour of the cells within the population. Hence for the remainder of the paper, we neglect the transient terms. Consequently, for perfectly synchronous cellular drives, the solution of the ME \([3]\) is a Poisson random variable with time-dependent rate equal to the effective upstream drive, \(\chi(t)\):

\[
[N_t|\chi(t)] \approx [N_t^\text{L}|\chi(t)] \sim \text{Poi}(\chi(t)),
\]

with distribution

\[
P(n,t)[\mu(\tau)\tau \in [0,t], \lambda(\tau)\tau \in [0,t]) = P(n,t|\chi(t)) = \frac{\chi(t)^n}{n!} e^{-\chi(t)},
\]

which makes explicit the dependence on the *history* of the sample paths \(\{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}\), which is encapsulated in the value of the effective drive \(\chi(t)\) at time \(t\).

Indeed, from \([4]\) it follows that the sample path \(\{\chi(t)\}_{t \geq 0}\) satisfies a first order linear ordinary differential equation with time-varying coefficients:

\[
\frac{d\chi}{dt} + \lambda(t) \chi = \mu(t),
\]

which is the rate law for a chemical reaction with zeroth-order production with rate \(\mu(t)\), and first-order degradation with rate \(\lambda(t)\) per mRNA molecule. For biologically realistic (i.e., positive and finite) cellular drives, \(\chi(t)\) is a continuous function.

**B. The general asynchronous case: cell-to-cell variability in the cellular drives**

Consider now the general case where different sample paths for the cellular drives are possible, i.e., the cell population has some degree of asynchrony and \(M_t\) and \(L_t\) have non-zero variance for at least some \(t \geq 0\). The transcription and degradation rates are then described by stochastic processes \(M\) and \(L\):

\[
\emptyset \overset{M_t}{\rightarrow} \text{mRNA} \overset{L_t}{\rightarrow} \emptyset,
\]

and the collection of all differential equations of the form \([11]\) is represented formally by the random differential equation

\[
dX/dt + LX_t = M_t,
\]

Equations of this form appear in many sciences, and a large body of classical results allows us to determine the probability density function of the upstream process, \(X_t\) \([37–39]\). Below, we use such results to obtain \(f_{X_t}(x,t)\) for biologically relevant models.

Note that from Eq. (10) and the law of total probability, it follows that the probability mass function for the random variable \(N_t\) under cellular drives described by the random processes \(M\) and \(L\) is given by the Poisson mixture (or compound) distribution:

\[
P(n,t) = P_{X_t}(n,t) = \int x^n e^{-\chi} f_{X_t}(x,t) \, dx,
\]

where the density \(f_{X_t}(x,t)\) of the effective drive \(X_t\) (to be determined) can be understood as a *mixing density*. The notation \(P_{X_t}(n,t)\) recalls explicitly the dependence of the solution on the density of \(X_t\), but we drop this reference and use \(P(n,t)\) below to simplify notation. The problem of solving the full ME is thus reduced to finding the mixing density \(f_{X_t}(x,t)\). Note that for synchronous drives, we have \(f_{X_t}(x,t) = \delta(\chi(t) - x)\), where \(\delta\) is the Dirac delta function, and (14) reduces to (10).

Equation (14) also shows that there are two separate sources of variability in gene expression, contributing to the distribution of \(N_t\). One source of variability is the Poisson nature of transcription and degradation, common to every model of the form considered here; the second source is the time-variation or uncertainty in the cellular drives, encapsulated in the upstream process \(X_t\) describing the ‘degree of synchrony’ between cells and/or their variability over time. In this sense, Eq. (14) connects with the concept of separable ‘intrinsic’ and ‘extrinsic’ components of gene expression noise pioneered by Swain *et al.* \([40–43]\). Yet rather than considering moments, the full distribution \(P(n,t)\) is separable into a model-dependent ‘upstream component’ given

---

1. We do not use the term *stochastic* differential equation (SDE), because SDEs are usually associated with random white noise.
by \(f_{X_t}(x,t)\), and a downstream transcriptional ‘Poisson component’ common to all models of this type.

### III. THE EFFECTIVE UPSTREAM DRIVE IN GENE TRANSCRIPTION MODELS

The generic model of gene transcription and degradation with time-dependent drives introduced above provides a unifying framework for several models previously considered in isolation. In this section, we exemplify the tools to obtain the density of the effective drive \(f_{X_t}(x,t)\) analytically or numerically through relevant examples.

#### A. Gene transcription under upstream drives with static randomness

In this first section, we consider models of gene transcription where the upstream drives are deterministic, yet with random parameters representing cell variability.

1. Random entrainment to upstream sinusoidal drives: random phase offset in transcription or degradation rates

Equation (13) can sometimes be solved directly to obtain \(f_{X_t}(x,t)\) from a transformation of the random variables \(M_t\) and \(L_t\). We now show two such examples, where we explore the effect of entrainment of gene transcription and degradation to an upstream periodic drive [44].

First, consider a model of gene transcription of the form (12) with transcription rate given by a sinusoidal function and where each cell has a random phase. This random entrainment (RE) model is a simple representation of a cell population with transcription entrained to an upstream rhythmic signal, yet with a random phase offset for each cell:

\[
M_t := 
\frac{m}{2} \left( 1 + \cos(\omega t + \phi) \right),
\]

\[
L_t := b + \alpha \cos(\omega t + \phi),
\]

where \(m, \alpha, b, \omega\) are given constants, and \(\phi\) is a (static) random variable. Equation (13) can be solved approximately [44] to give

\[
X_t = B + A \sin(\omega t + \phi^*),
\]

where \(A = 2ma\sqrt{1 + (\phi^*/b)^2}/[2(b^2 + \omega^2) - a^2]\), \(B = Ab\sqrt{1 + (\phi^*/b)^2}/a\), and \(\phi^* = \pi/2 - \arctan(\phi^*/\phi)\). As before, if \(\phi^*\) is uniform on \([-r, r]\), the density of the effective drive takes the same form (16) as above.

2. Upstream Kuramoto promoters with varying degree of synchronization

As an illustrative computational example, we study a population of cells whose promoter strengths display a degree of synchronization across the population. To model this upstream synchronization, consider the Kuramoto promoter model, where the promoter strength of each cell depends on an oscillatory phase \(\theta, t\), and cells are coupled via a Kuramoto model [45, 47]. We then have a model of the form (12) with:

\[
M_t := \frac{m}{2} \left( 1 + \cos(\Theta(t; \Omega)) \right),
\]

\[
L_t := 1.
\]

Here \(m, b\) are constants and \(\{\theta, t\}\) are the phase variables for the \(C\) cells governed by the globally coupled Kuramoto model:

\[
\frac{d\theta_j}{dt} = \omega_i + \frac{K}{C} \sum_{j=1}^{C} \sin(\theta_j - \theta_i),
\]

where \(K\) is the coupling parameter and the intrinsic frequency of each cell, \(\omega_i\), is drawn from the random distribution \(\Omega \sim N(0, 0.05^2)\). The Kuramoto model allows us to tune the degree of synchrony through the coupling
K: for small K, the cells do not display synchrony since they all have a slightly different intrinsic frequency; as K is increased, the population becomes more synchronized.

This model is a simple representation of nonlinear synchronization processes in cell populations with intrinsic heterogeneity [48–51]. In Figure 5(a), we show how the synchronization processes in cell populations with intrinsic heterogeneity influence the degree of synchrony in the population. The desired density \( f_{X_t}(x, t) \):}

\[
\frac{\partial}{\partial t} f_{X_t}(x, y, t) = -\frac{\partial}{\partial x} \left[ (\mu(y, t) - \lambda(y, t)) f_{X_t}(x, y, t) \right] + \mathcal{L}_{Y_t} [f_{X_t}(x, y, t)],
\]

which follows from conservation of probability. In Equation (20), the differential operator for \( X \), which follows from (13), is the first jump moment [52] conditional upon \( Y_t = y \) (and hence upon \( M_t = \mu(y, t) \) and \( L_t = \lambda(y, t) \)); the second term \( \mathcal{L}_{Y_t} \) is the infinitesimal generator of the upstream processes. In particular, for continuous stochastic processes \( \mathcal{L}_{Y_t} \) is of Fokker-Planck type, and for Markov chains \( \mathcal{L}_{Y_t} \) is a transition rate matrix. The desired density \( f_{X_t}(x, t) \) can then be obtained via marginalization.

Equation (20) can be employed to analyze the widely used class of transcription models with asynchronous, random promoter switching between discrete states, where each state has different transcription and degradation rates representing different levels of promoter activity due to, e.g., transcription factor binding or chromatin remodeling [54]. A classic example is the random telegraph (RT) model, first used by Ko in 1991 [55] to explain cell-to-cell heterogeneity and bursty transcription (Fig. 3a).

In our framework, such random promoter switching can be understood as an upstream stochastic process driving transcription as follows. Let us assume that

B. Asynchronous transcription under stochastic multistate promoters

In the previous section, we obtained \( f_{X_t}(x, t) \) by capitalizing on the precise knowledge of the sample paths of \( M \) and \( L \) to solve (15) explicitly. In other cases, we can obtain \( f_{X_t}(x, t) \) by following the usual procedure of writing down an evolution equation for the probability density of an expanded state that is Markovian, and then marginalizing. More specifically, let the vector process \( Y = (X_t, Y_t) \) prescribes the upstream drives, so that \( M = M(Y, t) \) and \( L = L(Y, t) \), and consider the expanded state \( X_t := (X_t, Y_t) \). Note that since \( Y \) is upstream, it prescribes \( X \) (and not vice versa). We can then write the evolution equation for the joint probability density \( f_{X_t}(x, y, t) \):

\[
\frac{\partial}{\partial t} f_{X_t}(x, y, t) = -\frac{\partial}{\partial x} \left[ (\mu(y, t) - \lambda(y, t)) f_{X_t}(x, y, t) \right] + \mathcal{L}_{Y_t} [f_{X_t}(x, y, t)],
\]

where each state has different transcription and degradation rates representing different levels of promoter activity due to, e.g., transcription factor binding or chromatin remodeling [54]. A classic example is the random telegraph (RT) model, first used by Ko in 1991 [55] to explain cell-to-cell heterogeneity and bursty transcription (Fig. 3a).
the promoter can attain D states s, and each state has constant transcription rate \( m_s \) and constant degradation rate \( \ell_s \). The state of the promoter is described by a random process \( S = \{ S_t \in \{ 1, 2, \ldots, D \} : t \geq 0 \} \), with sample paths denoted by \( \{ \{ s(t) \} \}_{t \geq 0} \), and its evolution is governed by the D-state Markov chain with transition rate \( k_{sr} \) from state \( r \) to state \( s \). The state of the promoter \( S_t = s \) prescribes that \( M_t = m_s \) and \( L_t = \ell_s \). Hence, the sample paths of \( M \) and \( L \) are a succession of step functions with heights \( m_s \) and \( \ell_s \), respectively, occurring at exponentially distributed random times.

As described above, we expand the state space of the cellular drives to include the promoter state \( X_t = \{ X_t, S_t \} \). The evolution equation (20) is then given by \( D \) coupled equations:

\[
\frac{\partial}{\partial t} f_{X,S}(x,s,t) = -\frac{\partial}{\partial x} \left[ (\mu_s - \lambda_s) f_{X,S}(x,s,t) \right] + \sum_{j=1}^{D} k_{sj} f_{X,S}(x,j,t) - \sum_{j=1}^{D} k_{js} f_{X,S}(x,s,t),
\]

\[
\quad s = 1, 2, \ldots, D,
\]

which can be thought of as a set of multistate Fokker-Planck-Kolmogorov equations [52]. Marginalization then leads to the density of the effective drive:

\[
f_{X}(x,t) = \sum_{s=1}^{D} f_{X,S}(x,s,t),
\]

and the full ME solution is obtained from (22) and [14].

We illustrate this approach more explicitly with two examples (Fig. 3): a re-derivation of the known solution of the standard RT model; and the solution of the 3-state cyclic model with a refractory state. Results for other promoter architectures are discussed in [54].

1. The Random Telegraph model (2 states)

Although the RT model has been solved by several methods [13, 32, 33], we briefly rederive its solution within the above framework to clarify its generalisation to other promoter architectures.

Consider the standard RT model (Fig. 3a), with promoter switching stochastically between the active state \( s_{on} = 1 \), with constant transcription rate \( m_1 = m \), and the inactive state \( s_{off} = 0 \), where no transcription takes place, \( m_0 = 0 \). The transition rates between the two states are \( k_{10} = k_{on} \) and \( k_{01} = k_{off} \). Without loss of generality, we assume \( \ell_1 = \ell_0 = \lambda(t) \equiv 1 \). The transcription model is of the form (12):

\[
M_t = m S_t, \quad L_t = 1,
\]

where \( S = \{ S_t \in \{ 0, 1 \} : t \geq 0 \} \) with waiting times drawn from exponential distributions: \( \tau_{on} \sim \exp(1/k_{on}) \) and \( \tau_{off} \sim \exp(1/k_{off}) \).

Let \( Z_t = X_t/m, \) and denote \( f_{on}(z,t) = f_{Z,S}(z,s_{on},t) \) and \( f_{off}(z,t) = f_{Z,S}(z,s_{off},t) \), with \( z \in (0, 1) \). Then the multistate Fokker-Planck-Kolmogorov equations (21) are:

\[
\frac{\partial f_{on}}{\partial t} = -\frac{\partial}{\partial z} [(1-z)f_{on}] - k_{off} f_{on} + k_{on} f_{off},
\]

\[
\frac{\partial f_{off}}{\partial t} = -\frac{\partial}{\partial z} [zf_{off}] + k_{off} f_{off} - k_{on} f_{on},
\]

\[
f_{Z,t} = f_{on} + f_{off}
\]

with integral conditions \( \int_{0}^{1} f_{on}(z,t) \, dz = P(S_t = s_{on}) \) and \( \int_{0}^{1} f_{off}(z,t) \, dz = P(S_t = s_{off}) \).

At stationarity, it then follows [55] that

\[
f_{Z,t}(z) = \frac{z^{k_{off}-1}(1-z)^{-1}}{B(k_{on},k_{off})},
\]

where \( B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b) \) is the Beta function. In other words, at stationarity, the normalised effective drive is described by a Beta distribution:

\[
Z_t \sim \text{Beta}(k_{on},k_{off}), \quad \forall t.
\]

Using (24) and (14), we obtain that the full stationary solution is the Poisson-Beta mixture:

\[
P(n) = \int_{0}^{1} \frac{(mz)^n}{n!} e^{-mz} \frac{z^{k_{off}-1}(1-z)^{-1}}{B(k_{on},k_{off})} \, dz
\]

\[
= \frac{\Gamma(k_{on} + n)}{\Gamma(k_{on})} \frac{\Gamma(k_{off} + k_{on} + n)}{\Gamma(k_{on} + k_{off} + n)} \frac{m^n}{\Gamma(n+1)} \times 1 F_1 (k_{on} + n, k_{on} + k_{off} + n; -m),
\]

\[\text{(25)}\]
where $F_1(a, b; z)$ is the confluent hypergeometric function \[36,37\].

### 2. The Refractory Promoter model (3 states)

In the standard RT model, the waiting times in each state are exponentially distributed. In recent years, time course data have shown that the $\tau_{on}$ do not conform to an exponential distribution, leading some authors to incorporate a second inactive (refractory) state, which needs to be cycled through before returning to the active state \[36,37\]. The net ‘OFF’ time is then the sum of two exponentially distributed waiting times.

In this **refractory promoter model** (Fig. 3b), the promoter switches through the states $s_a$, $s_1$, and $s_2$ with rates $k_s$, $k_1$ and $k_2$. Transcription takes place at constant rate $m$ only when the promoter is in the active state $s_a$ and, without loss of generality, we assume a constant degradation rate $\lambda(t) \equiv 1$ for all states. This model is of the same form as \[23\], and is solved similarly.

Making the change of variables $Z_t = \lambda X_t / m = X_t / m$ and, using the notation $f_i(z, t) := f_i, z \{z, t, s_i\}$, the multistate Fokker-Planck-Kolmogorov equations are

$$\frac{\partial f_s}{\partial t} = -\frac{\partial}{\partial z} \left[(1-z)f_s - k_s f_s + k_2 f_2\right],$$
$$\frac{\partial f_1}{\partial t} = -\frac{\partial}{\partial z} \left[-zf_1 + k_s f_s - k_1 f_1\right],$$
$$\frac{\partial f_2}{\partial t} = -\frac{\partial}{\partial z} \left[-zf_2 + k_1 f_1 - k_2 f_2\right],$$
$$f_{Zt} = f_s + f_1 + f_2,$$

with three integral conditions $\int_0^1 f_i(z) \, dz = P(S = s_i)$. At stationarity, we find

$$f_{Zt}(z) = C_1 \ z^{k_2-1} F_1 \left[\binom{a}{2}; a; 1 + k_1 - k_2; z \right] + C_2 \ z^{k_2-2} F_1 \left[\binom{a}{2}, a - 1; 1 + k_1 - k_2; z \right],$$

with $a^{(1)} = \frac{1}{2} \ (2 + k_1 - k_2 - k_s \pm d)$

$$a^{(2)} = \frac{1}{2} \ (2 - k_1 + k_2 - k_s \pm d),$$

where $k_s, (k_1 - k_2)$, $d = \sqrt{(k_1 - k_2)^2 - 4 k_1 k_2} \not\in \mathbb{Z}$ and $\binom{a}{2} = \binom{a}{2} \left(a; b; c; z \right)$ is the Gauss hypergeometric function \[36\]. The full stationary solution $P(n)$ is then obtained from \[14\].

For a detailed derivation (including expressions for the integration constants $C_1$ and $C_2$), see Appendix B.

### C. Asynchronous multistate models with upstream promoter modulation

Finally, we consider a model of gene transcription that incorporates features of models described in Sections IIIA and IIIB. Such a situation is of biological interest and appears when individual cells exhibit correlated dynamics in response to upstream factors (e.g., changing environmental conditions, drives or stimulations), but still maintain asynchrony in internal processes, such as transcription factor binding \[31,35\].

To illustrate this concept, we consider the **modulated random telegraph** (MRT) model, a combination of the RE model \[15\] and the RT model \[23\], i.e., the promoter strength is modulated by an upstream sinusoidal drive with random phase $\phi$, as in the RE model, yet the promoter switches stochastically between active/inactive states with rates $k_{on}$ and $k_{off}$, as in the RT model. In this model, the transcription rate is correlated between cells through the entrainment to the upstream sinusoidal drive as a simple model for, e.g., circadian gene expression:

$$M_t := R_t(\Phi) S_t = m \frac{1 + \cos(\omega t + \Phi)}{2} S_t,$$
$$L_t := 1,$$

where $m, \omega > 0$ are constants; $\Phi$ is the random phase across the cell population; and $S = \{S_t \in \{0, 1\} : t \geq 0\}$, with exponential waiting times, describes the stochastic integration constants. \[58\].
promoter switching (Fig. 4h).

The solution of this model follows from the RT probability density \[ f_{\rho}(t) \] conditioned on the random phase \( \Phi \), which prescribes the sample path \( \{ \rho(t; \phi) \}_{t \geq 0} \) of the promoter strength \( R \). The resulting scaled Beta distribution

\[
f_{X_1}(x, t|\phi) = \frac{x^{k_{off}-1}(\rho(t; \phi) - x)^{k_{off}-1}}{B(k_{on}, k_{off}) \rho(t; \phi)^{k_{on}+k_{off}-1}}
\]

is then marginalized over the phase \( \Phi \) to obtain the density \( f_{X_1}(x, t) \) of the effective drive. For instance, if the phases are normally distributed \( \Phi \sim \mathcal{N}(0, \sigma^2) \), we have (see Fig. 4i):

\[
f_{X_1}(x, t) = \int f_{X_1}(x, t|\phi) f_\phi(\phi) \, d\phi
\]

As \( \sigma \to \infty \), the population becomes asynchronous in the promoter strength, as well as in the state transitions, and time dependence wanes (Fig. 4i).

**IV. ENSEMBLE NOISE CHARACTERISTICS IN TIME-VARYING POPULATIONS**

In the previous sections, we were concerned with the full time-dependent probability distribution \( P(n, t) \) for the mRNA copy number \( N \). However, in many circumstances such detailed information is not required, and simpler characterizations based on ensemble averages (e.g., Fano factor, coefficient of variation) are of interest. Simple corollaries from the Poisson mixture expression [16] allow us to derive expressions for the ensemble moments and other noise characteristics, as shown below. We remark that, in this section, all the expectations are taken over the distribution describing the cell population.

**A. Time-dependent ensemble moments over the distribution of cells**

To quantify noise characteristics of gene expression in a population, the ensemble moments \( \mathbb{E}[N^k] \), \( k \in \mathbb{N}, t \geq 0 \) are often determined via the probability generating function [14] [32] [59] or by integrating the master equation [19] [34] [60]. However, stationarity is usually assumed and the moments derived are not suitable for time-varying systems. Here we use corollaries of the Poisson mixture expression [14] to derive expressions for the ensemble moments for time-varying systems under upstream drives.

From [10] we have \( [N_t|X_t = x] \sim \text{Poi}(x) \); hence

\[
\mathbb{E}[N^k_t|X_t = x] = \sum_{r=1}^{k} x^r S(k, r),
\]

where

\[
S(k, r) = \sum_{j=0}^{r} \frac{(r)^{-j} j^k}{r!}
\]

are the Stirling numbers of the second kind [61]. The law of total probability then gives

\[
\mathbb{E}[N^k_t] = \sum_{r=1}^{k} S(k, r) \mathbb{E}[X^r_t], \quad (26)
\]

or, equivalently,

\[
\begin{pmatrix}
\mathbb{E}[N^1_t] \\
\mathbb{E}[N^2_t] \\
\vdots \\
\mathbb{E}[N^k_t]
\end{pmatrix} = \begin{pmatrix}
S(1, 1) & 0 & \ldots & 0 \\
S(2, 1) & S(2, 2) & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
S(k, 1) & S(k, 2) & \ldots & S(k, k)
\end{pmatrix} \begin{pmatrix}
\mathbb{E}[X^1_t] \\
\mathbb{E}[X^2_t] \\
\vdots \\
\mathbb{E}[X^k_t]
\end{pmatrix}
\]

Therefore the ensemble moments of the mRNA copy number \( \mathbb{E}[N^k_t] \) can be obtained in terms of the moments of the effective drive \( \mathbb{E}[X^k_t] \), and vice versa. For instance, \( \mathbb{E}[X_t] = \mathbb{E}[N^1_t] \); \( \text{Var}(X_t) = \mathbb{E}[N_t] - \mathbb{E}[N^2_t] \), and the skewness \( \gamma_1(X_t) = (\mathbb{E}[N^3_t] - 3 \mathbb{E}[N_t] \text{Var}(N_t) - 3 \text{Var}(N_t) \mathbb{E}[N_t])^{1/2} / (\mathbb{E}[N_t])^{3/2} \).

The expressions (26) complement those used by Swain et al. in their seminal work on intrinsic/extrinsic noise [11]. and those by Hillinger and Paulsson to separate intrinsic/extrinsic fluctuations [12].

**B. Time-dependent ensemble Fano factor: a measure of synchrony in the population**

A commonly used measure of variability in the population is the ensemble Fano factor:

\[
\text{Fano}(N_t) := \frac{\text{Var}(N_t)}{\mathbb{E}[N^2_t]},
\]

(27)

which is unity for the Poisson distribution. Its use has been popularized as a measure of the deviation from the stationary solution of the transcription of an unregulated gene with constant rates [62] [63], which is Poisson; hence with \( \text{Fano}(N_t) \equiv 1, \forall t \).

For time-varying systems, however, the ensemble Fano factor conveys how the dynamic variability in single cells combines at the population level. Indeed, \( \text{Fano}(N_t) \) can be thought of as a measure of synchrony in the population at time \( t \). For instance, it follows from Eq. 10 that the ensemble Fano factor for a population with perfectly synchronous drives is always equal to one, \( \text{Fano}(N_t) \equiv 1, \forall t \). Even if the upstream drive \( \chi(t) \) changes in time, the population remains synchronous and has a Poisson distribution at all times. On the other hand, under the assumptions of our model, when \( \text{Fano}(N_t) \) varies in time
from (26) it follows that: $X$ cells, as captured by the upstream drive it reflects a change in the degree of synchrony between mRNA counts $N$ are shown. The mean, variance, and ensemble Fano factor of $N$ were calculated from the sample paths of $N$ (blue lines) and, more efficiently, from the sample paths of $X$ (black lines). The last row shows the Kuramoto order parameter $r(t)$ measuring the cell synchrony, signaled by $r(t) \rightarrow 1$. (b) Ensemble Fano factor (averaged over the simulated time courses) against coupling parameter $K \in (0, 0.4)$. As $K$ is increased, the oscillators become synchronized and the ensemble Fano factor decreases towards the Poisson value of unity. (c) Scatter plot of the ensemble Fano factor against the order parameter $r(t)$ (both averaged over the simulated time courses). As the oscillators become synchronized ((r) $\rightarrow 1$), the ensemble Fano factor also approaches 1, signifying that the distribution is Poissonian at all times.

![Diagram of noise characteristics of the Kuramoto promoter model](image)

FIG. 5. Noise characteristics of the Kuramoto promoter model (18). (a) Numerical simulations for $C = 100$ oscillatory cells and different coupling parameters: $K = 0.002, 0.1, 0.4$ (left, middle, right columns). For each coupling, the sample paths of the upstream effective drive $X$ and mRNA counts $N$ are shown. The mean, variance, and ensemble Fano factor of $N$ were calculated from the sample paths of $N$ (blue lines) and, more efficiently, from the sample paths of $X$ (black lines). The last row shows the Kuramoto order parameter $r(t)$ measuring the cell synchrony, signaled by $r(t) \rightarrow 1$. (b) Ensemble Fano factor (averaged over the simulated time courses) against coupling parameter $K \in (0, 0.4)$. As $K$ is increased, the oscillators become synchronized and the ensemble Fano factor decreases towards the Poisson value of unity. (c) Scatter plot of the ensemble Fano factor against the order parameter $r(t)$ (both averaged over the simulated time courses). As the oscillators become synchronized ((r) $\rightarrow 1$), the ensemble Fano factor also approaches 1, signifying that the distribution is Poissonian at all times.

The greater the synchrony at time $t$, the closer $\text{Fano}(N_t)$ is to unity, since the deviation from the Poisson distribution emanates from the ensemble Fano factor of the upstream drive $X_t$.

As an example, consider the Kuramoto promoter model (18) introduced in Section III A 2 where the cells in the population become more synchronized as the value of the coupling $K$ is increased. Figure 5 shows simulation results for 100 cells with a range of couplings. The order parameter $r(t) \in [0, 1]$ measures the phase coherence of the oscillators at time $t$; as it grows closer to 1, so grows the degree of synchrony. Using the Kuramoto numerics, we calculate the ensemble Fano factor $\text{Fano}(N_t)$ for the transcription model. As seen in Fig. 5(b)-(c), the more synchronous, the closer the Fano factor is to the Poisson value: $r(t) \rightarrow 1 \implies \text{Fano}(N_t) \rightarrow 1$.

Figure 5 also illustrates the computational advantages of our method. The cost to approximate the time-varying ensemble moments is drastically reduced by using (26), because transcription and degradation events do not have to be simulated. The sample paths of the effective drive $X_t(t)$ were used to estimate the time-varying moments: $\mathbb{E}[N_t] = \mathbb{E}[X_t]$ and $\text{Var}(N_t) = \text{Var}(X_t)$ (shown in black). These correspond to the numerical simulation of ODEs, and are far more efficient than sampling from realisations $\nu_i(t)$ of the mRNA copy number.

V. VARIABILITY OVER TIME: STATIONARITY AND ERGODICITY

Our results up to now have not assumed stationarity—in general the distribution (14) and ensemble moments (26) depend on time. If the cells in the population are uncorrelated and $M$ and $L$ are stationary (i.e., their statistics do not change over time), then $f_{X_t}(x, t)$ tends to a stationary density $f_{X_t}(x)$ (25), and the full solution $P(n, t)$ also tends to a stationary distribution $P(n)$. Under such assumptions leading to stationarity, any time dependence in the solution $P(n, t)$ describes the ‘burn-in’ transient from an initial condition towards the attracting stationary distribution, as discussed in Section (11A.1). An example of such transience in the random telegraph model (25) appeared in (13) describing how the distribution $P(n, t)$ settles from an initial Kronecker delta distribution $P(n, 0) = \delta_{n0}$ to the stationary distribution (25) (see Fig. 6 and Appendix A.2).

If, in addition to stationarity, we assume the cells to be
Sinusoidal telegraph Random Square wave

The delta distribution at decay as the system approaches its stationary distribution. For this initial probability distribution exhibits an exponential the inactive state [13]. (b) The full solution of the RT model for an initial condition $X_{\text{effective upstream drive}}$ and the number of mRNAs $N$ from a time snapshot at stationarity of a population of $T$ cells, as $C \rightarrow \infty$. Hence the probability distribution obtained from a single cell over a time distribution (14) under ergodicity using (28)-(29). Note that when $\chi(t)$ is periodic with period $T$, the limit in Eq. (29) is not required. In Figure 7 we show the time-averaged distribution $\langle P(n) \rangle$ for a cell under three different upstream drives $\mu(t)$: (i) a continuous sinusoidal form; (ii) a discontinuous square wave form; (iii) a random telegraph (RT) form, which can be thought of as the stochastic analogue of the square wave. In all cases, the drive $\{\mu(t)\}_{t \geq 0} \in [0, 20]$ with the same period, or expected period, $T$. For simplicity, we set $\lambda(t) \equiv 1$.

indistinguishable, the population is ergodic. In this case, the distribution obtained from a single cell over a time $T$, as $T \rightarrow \infty$, is equivalent to the distribution obtained from a time snapshot at stationarity of a population of $C$ cells, as $C \rightarrow \infty$:  

$$P(n) = \langle P(n) \rangle$$  

(28)

where $P(n) = \lim_{T \rightarrow \infty} P(n, t) = \int \frac{x^n}{n!} e^{-x} f_X(x) \, dx$, and $\langle P(n) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T P(n, t|\chi(t)) \, dt$  

$$= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \frac{\chi(t)^n}{n!} e^{-\chi(t)} \, dt. $$  

(29)

where $\langle \cdot \rangle$ denotes time-averaging. Under the assumption of ergodicity, the averages computed over single-cell sample paths can be used to estimate the stationary distribution of the population.

A. Ergodic systems: stochastic vs deterministic drives

It has been suggested that stochastic and periodic drives lead to distinct properties in the noise characteristics within a cell population [12]. We investigate the effect of different temporal drives on the full distribution [14] under ergodicity using [28]-[29]. Note that as the period $T$ is varied, the similarity between the distributions under the three upstream drives varies considerably (Fig. 7). At small $T$, the distributions under sinusoidal and square wave forms are most similar; whereas at large $T$, the distributions under square wave and RT forms become most similar. In general, the distribution of the RT model has longer tails (i.e., $n$ low and high) as a consequence of long (random) waiting times that allow the system to reach equilibrium in the active and inactive states, although this effect is less pronounced when the promoter switching is fast relative to the time scales of transcription and degradation (e.g., $T = 1/5$). On the other hand, as $T$ grows, the square wave and RT drives are slow and the system is able to reach the equilibrium in both active and inactive states. Hence the probability distributions of the square wave and RT drives become similar, with a more prominent bimodality.
B. The temporal Fano factor: windows of stationarity in single-cell time course data

The **temporal Fano factor** (TFF) is defined similarly to the ensemble version (27), but is calculated from the variance and mean of a single time series \( \{\nu(t)\}_{t \geq 0} \) over a time window \( W := (t_1, t_2) \):

\[
\text{TFF}_{\{\nu(t)\}}(W) = \frac{\langle \nu(t)^2 \rangle_{t \in W} - \langle \nu(t) \rangle^2_{t \in W}}{\langle \nu(t) \rangle^2_{t \in W}}. \tag{30}
\]

In fact, this is the original definition of the Fano factor [64], which is used in signal processing to estimate statistical fluctuations of a count variable over a time window. Although \( N_t \) is not a count variable (it decreases with degradation events), the TFF can be used to detect windows of stationarity in single-cell time courses.

Figure 8a shows a single-cell sample path \( \{\nu(t)\}_{t \geq 0} \) generated by the (leaky) random telegraph model with constant degradation rate \( \lambda \), and transcription rates \( \mu_1 > \mu_0 > 0 \) for the active and inactive promoter states. The leaky RT model is equivalent to the standard RT model, and switches between two states with expectations \( \mu_1/\lambda \) and \( \mu_0/\lambda \). In the time windows \( W \) between promoter switching, \( \{\nu(t)\}_{t \in W} \) can be considered almost at stationarity and described by a Poisson distribution with parameter \( \mu_0/\lambda \) (resp. \( \mu_1/\lambda \)) in the inactive (resp. active) state. Hence \( \text{TFF}_{\{\nu(t)\}}(W) \approx 1 \) across most of the sample path, except over the short transients \( W_{\text{trans}} \) when the system is switching between states, where \( \text{TFF}_{\{\nu(t)\}}(W_{\text{trans}}) > 1 \) (Fig. 8p).

Alternatively, this information can be extracted robustly from the **cumulative Fano factor** (cTFF):

\[
\text{cTFF}_{\{\nu(t)\}}(t_1, t) = \text{TFF}_{\{\nu(t)\}}((t_1, t)), \quad t \geq t_1 \tag{31}
\]

which is computed over a growing window from a fixed starting time \( t_1 \). The cTFF is a cumulative moving average giving an integrated description of how the stationary regimes are attained between switching events indicated by the step-like structure of the heatmap in Fig. 8.

VI. DISCUSSION

In this paper, we have presented the solution of the master equation for gene transcription with upstream dynamical variability, in a setting that allows a unified treatment of a broad class of models. The framework allows quantitative biologists to go beyond stationary solutions when using inference to analyze noise in single-cell experiments. As an alternative to computational approaches where many cells are explicitly simulated to account for observed variability, our work takes a parsimonious approach and uses a simple gene transcription model of Poissonian type, but considers explicitly the effect of dynamical and cell-to-cell upstream variability in the solution of the master equation. The solution obtained can be viewed as combining an upstream component (dynamic or static, deterministic or stochastic) with a downstream Poissonian immigration-death process. This structure can describe both time-dependent snapshots of the population as well as the variability over single-cell time courses in a coherent fashion.

Our procedure solves the general transcription-degradation model first, and then arrives at its Poisson mixture form (11). Since only the upstream process is model-specific, different models are solved by obtaining the mixing density \( f_X \) of the upstream process. It is interesting to note that our solution conforms with the "Poisson representation", which was proposed by Gardiner and Chaturvedi [63, 60] as an ansatz to solutions of the ME. In their representation, \( f_X \) had support on the complex plane as a means to obtain asymptotic expansions for certain stationary systems [66], and assumptions
regarding vanishing boundary conditions were required. In our case, the Poisson mixture [14] is obtained directly as a result of transcription-degradation models, and the mixing density \( f_{X_t} \) with support on the positive real line, has a clear physical interpretation in terms of single-cell sample paths for a range of time-varying systems.

Our solution confers to us two broad advantages. The first is pragmatic: Since \( X_t \) is a continuous random variable satisfying a linear random differential equation, we can draw upon the rich theory and analytical results for \( f_{X_t} \), even for non-stationary models, or we can use ODE and PDE solvers as further options to solve the differential equation for \( f_{X_t} \). If simulations are still necessary, sampling \( P(n,t|M,L) \) directly using stochastic simulation algorithms becomes computationally expensive, particularly if the upstream processes \( M \) and \( L \) are time-varying [67]. Instead, we can sample \( f_{X_t}(x,t) \) directly and then obtain the full distribution via numerical integration using [14]. This approach leads to a significant reduction in computational cost, as shown in Fig. 9.

Our approach can be used to analyze noise characteristics in conjunction with biological hypotheses. For instance, if measurements of additional cellular variables (e.g., cell cycle) are available, they can be incorporated as a source of variability for gene regulation, with the possibility to test biological hypotheses computationally against experimental data. Conversely, it is possible to discount the Poissonian component from data, so as to fit different promoter models to experimental data and perform model comparison [54].

The second advantageous feature of our approach is conceptual. Studying the natural decoupling of the solution into a discrete, Poisson component and a continuous, mixing component, allows us to derive expressions and properties for both ensemble and temporal moments, extending the concept of ‘intrinsic’/‘extrinsic’ noise to dynamic upstream cellular drives. Importantly, all upstream variability gets effectively imbricated through the upstream effective drive \( X \), which can be interpreted in terms of biochemical differential rate equations. This analysis clarifies how upstream fluctuations are combined to affect the probability distribution of the mRNA copy number, providing further intuition of the sources of noise and their characteristics. Indeed, stripping the model down to its extrinsic component \( f_{X_t} \) can provide us with additional understanding of its structure and time scales [54]. Further extensions of our solution could lead to physical interpretations of Gardiner’s complex-valued “Poisson representation” [63, 64], and deeper understanding of models with and without feedback [68]. These extensions will be the subject of future work.

VII. ACKNOWLEDGMENTS

We are grateful for insightful comments and extended discussions with Juan Kuntz, Philipp Thomas, Martin Hemberg, and Andrew Parry. JD was funded by an EPSRC PhD studentship. MB acknowledges funding from the EPSRC through grants EP/I017267/1 and EP/N014529/1.

Appendix A: The ‘burn-in’ transient towards stationarity

In Section II A, it was stated that the contribution from the initial condition \( N_0^\infty \) decreases exponentially for biophysically realistic degradation rates \( \{\lambda(t)\}_{t\geq0} \). As a result, the transcripts that were present at \( t = 0 \) are expected to degrade in finite time, and the long-term population is expected to be composed only of mRNA molecules that were transcribed since \( t = 0 \).

Let the initial condition be described by a random variable \( N_0 \) with a given probability distribution. It follows from Eqs. (8)-(9) that
where \( N_{ic}^t | n_0 \) and \( N_i^s \) are distributed according to Eqs (5, 7), and \( \Pr(N_{ic}^t = k) \) is the mixture of the time-dependent binomial distribution (5) and the distribution of the initial condition \( N_0 \).

1. ‘Burn-in’ transience in the model with constant transcription and degradation

The decay towards stationarity

To understand the ‘burn-in’ period more explicitly, consider the simplest example of the gene transcription model \([1]\) with constant transcription and degradation rates \( \mu \) and \( \lambda \), and assume that there are initially \( n_0 \) mRNA transcripts. Given that \( N_0 \sim \delta_{0,n_0} \), the solution is given by:

\[
N_t = N_{ic}^t | n_0 + N_i^s
\]

\[
N_{ic}^t | n_0 \sim \text{Bin}\left(n_0, e^{-\lambda t}\right),
\]

\[
N_i^s \sim \text{Poi}\left(\frac{\mu}{\lambda} \left(1 - e^{-\lambda t}\right)\right).
\]

Hence as \( t \to \infty \), the distribution will tend towards \( \text{Poi}(\mu/\lambda) \), the stationary distribution of the population. This is a well-known result in the literature \([66, 69]\).

\[
\begin{align*}
P(n, t | \{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}, n_0) &= \sum_{n_0} P(n, t | \{\mu(\tau)\}_{\tau \in [0,t]}, \{\lambda(\tau)\}_{\tau \in [0,t]}, n_0) \Pr(N_0 = n_0) \\
&= \sum_{n_0} \Pr(N_{ic}^t = k | N_0 = n_0) \Pr(N_i^s = n - k) \Pr(N_0 = n_0) \\
&= \sum_{k=0}^n \Pr(N_i^s = n - k) \left[ \sum_{n_0} \Pr(N_{ic}^t = k | N_0 = n_0) \Pr(N_0 = n_0) \right] \\
&= \sum_{k=0}^n \Pr(N_i^s = n - k) \Pr(N_{ic}^t = k).
\end{align*}
\]

The distribution of \( N_i^s \) is still given by (A1) and the contribution of \( N_{ic}^t \) is given by

\[
\begin{align*}
\Pr(N_{ic}^t = k) &= \sum_{n_0=k}^{\infty} \Pr(N_{ic}^t = k | N_0 = n_0) \Pr(N_0 = n_0) \\
&= \sum_{n_0=k}^{\infty} \binom{n_0}{k} (e^{-\lambda t})^k \left(1 - e^{-\lambda t}\right)^{n_0-k} \left(\frac{\mu}{\lambda}\right)^{n_0} e^{-\mu/\lambda} \frac{1}{n_0!} \\
&= \left(\frac{\mu}{\lambda}\right)^k (e^{-\lambda t})^k e^{-\mu/\lambda} \frac{1}{k!} \sum_{r=0}^{\infty} e^{\mu/\lambda} \left(1 - e^{-\lambda t}\right)^r \\
&= \left(\frac{\mu}{\lambda}\right)^k (e^{-\lambda t})^k e^{-\mu/\lambda} e^{\frac{\mu}{\lambda} (1 - e^{-\lambda t})} \\
&= \left(\frac{\mu}{\lambda}\right)^k e^{-\frac{\mu}{\lambda} (1 - e^{-\lambda t})}.
\end{align*}
\]

In other words, \( N_{ic}^t \sim \text{Poi}\left(\frac{\mu}{\lambda} e^{-\lambda t}\right) \), which cancels the contribution from \( N_i^s \sim \text{Poi}\left(\frac{\mu}{\lambda} \left(1 - e^{-\lambda t}\right)\right) \). Therefore

\[
N_t \sim \text{Poi}\left(\frac{\mu}{\lambda}\right), \quad \forall t.
\]

This example shows that \( N_{ic}^t \) and \( N_i^s \) will combine to reproduce a stationary distribution at all times \( t > 0 \), when the system starts at stationarity so that there is no ‘burn-in’ transient.

Starting at stationarity: the time-dependent \( N_{ic}^t \) and \( N_i^s \) balance each other at all times

It is illustrative to consider the dynamics of this system when the initial condition is chosen to be the stationary distribution. In this case, the breakdown of \( N_t \) into the time-dependent components \( N_{ic}^t \) and \( N_i^s \) will need to reproduce the stationary distribution at all times \( t > 0 \), with no ‘burn-in’ period.

To see this, let the initial distribution start at stationarity, i.e., \( N_0 \sim \text{Poi}(\mu/\lambda) \) and

\[
\Pr(N_0 = n_0) = \left(\frac{\mu}{\lambda}\right)^n e^{-\mu/\lambda} \frac{1}{n_0!}.
\]
as an inherent property of the system, analogous to the stable point of a dynamical system that moves in time (sometimes called a **chronotaxic system** [70]).

If \( \Pr(N_0 = n_0) = P_*(n_0, 0) \) for all \( n_0 \) in Eq. [9], the contributions from \( N_t^c \) and \( N_t^s \) balance each other as they did in the case of stationarity with \( N_0 \sim \text{Poi}(\mu/\lambda) \), and we would have \( P(n, t) = P_*(n, t) \), \( \forall n, t > 0 \) (recall that the breakdown \( N_t = N_t^c + N_t^s \) simply resolves the existing mRNA molecules into those that were present at \( t = 0 \), and those that were transcribed since \( t = 0 \)). Thus we only observe an initial transient period if the initial distribution starts away from its attracting distribution at \( t = 0 \). In all other cases, the following mathematical formulations are equivalent: \( i) \) assume that the system was initialised at \( t = -\infty \) and consider only \( N_t^s \), or \( ii) \) use the initial distribution \( P_*(n, 0) \) for all \( n \) at \( t = 0 \), and consider \( N_t^c + N_t^s \).

In this work, we focus on the time dependence of \( P(n, t) \) induced through non-stationarity of the parameters, and/or synchronous behaviour of the cells within the population. Hence, unless otherwise stated, in this work we assume that the system was initialised at \( t = -\infty \) and that the distribution of \( N_t^s \) is the attracting distribution \( P_*(n, t) \) for all \( t > 0 \), i.e., we neglect the contribution from \( N_t^c \).

**2. ‘Burn-in’ transience in the random telegraph model**

A time-dependent solution of the probability generating function for the random telegraph model appeared in Ref. [13], although the explicit expression for \( P(n, t) \) was omitted. As discussed above, the RT model represents asynchronous and stationary behaviour, hence the time-dependence appears only through convergence to the stationary distribution from the initial condition. We include the derivation here for completeness and to complement Fig. [6].

Consider the RT model depicted in Fig. [3]. Assuming that every cell is initialised in the inactive state with \( n_0 = 0 \) mRNA molecules, the probability generating function for the cell population is [13]:

\[
G(z, t) = F_s(z, t)\Phi_s(z, t) - \frac{\mu k_{on} e^{-(k_{on} + k_{off})t}}{(k_{on} + k_{off})(1 - k_{on} - k_{off})} (1 - z) F_{ns}(z, t)\Phi_{ns}(z, t)
\]

where \( F_s(z, t) := \sum_{r = 0}^{\infty} \frac{\mu e^{-t}}{r!} \frac{m_{off} r^m}{(k_{on} + k_{off})^m} \left( \frac{1 - m_{on} - m_{off}}{1 - m_{off}} \right) \left( \frac{m_{on}}{1 - m_{off}} \right)^{r - m_{off}} F_1 \left[ k_{off} + r, 1 + k_{off} + m_{off} + n - r; -\mu \right] \)

\[
\times F_1 \left[ k_{off} + r, 1 + k_{off} + m_{off} + n - r; \mu e^{-t} \right] - \frac{k_{on} m_{off} e^{-(k_{on} + k_{off})t}}{(k_{on} + k_{off})(1 - k_{on} - k_{off})} \sum_{r = 0}^{n - 1} \frac{m_{off} r^{m_{off}}}{(1 + k_{off} + k_{off}) r (2 - k_{on} - k_{off})^{n - r}} \]

\[
\times F_1 \left[ k_{off} + r, 1 + k_{off} + k_{off} + n - r; -\mu \right] - \frac{k_{on} \mu m_{off} e^{-(k_{on} + k_{off})t}}{(k_{on} + k_{off})(1 - k_{on} - k_{off})} \sum_{r = 0}^{n - 1} \frac{m_{off} r^{m_{off}}}{(1 + k_{off} + k_{off}) r (2 - k_{on} - k_{off})^{n - r}} \]

\[
\times F_1 \left[ k_{off} + r, 1 + k_{off} + k_{off} + n - r; \mu e^{-t} \right] - \frac{k_{on} \mu m_{off} e^{-(k_{on} + k_{off})t}}{(k_{on} + k_{off})(1 - k_{on} - k_{off})} \sum_{r = 0}^{n - 1} \frac{m_{off} r^{m_{off}}}{(1 + k_{off} + k_{off}) r (2 - k_{on} - k_{off})^{n - r}} \]

\[
\times F_1 \left[ k_{off} + r, 1 + k_{off} + k_{off} + n - r; -\mu \right],
\]

where \( (a)_m := a(a + 1) \ldots (a + m - 1) \) is Pochhammer’s function [56].

As \( \zeta \to 0, F_1 [a, b; \zeta] \to 1 \). Hence as \( t \to \infty, P(n, t) \to P(n) \), and we recover the known stationary solution:

\[
P(n) = \frac{\mu^n (k_{on}) n}{n!} F_1 \left[ k_{on} + n, k_{on} + k_{off} + n; -\mu \right].
\]
Appendix B: The stationary solution for the refractory promoter model (3 cyclic promoter states)

As explained in Section III B2, the stationary solution of the 3-state cyclic model describing a refractory promoter is obtained by solving the set of equations:

$$\frac{d}{dz} [(1 - z)f_z(z)] = -k_1 f_z(z) + k_2 f_1(z)$$ \hspace{1cm} (B1)

$$\frac{d}{dz} [-zf_1(z)] = -k_1 f_1(z) + k_2 f_z(z)$$ \hspace{1cm} (B2)

$$\frac{d}{dz} [-zf_2(z)] = -k_2 f_2(z) + k_2 f_1(z)$$ \hspace{1cm} (B3)

to obtain an expression for $f_z = f_z + f_1 + f_2$.

Note that the transition matrix $[k_{zs}]$ of the 3-state cyclic model is singular and hence $\lambda = 0$ is an eigenvalue. Furthermore, by Gershgorin’s circle theorem the non-zero eigenvalues of $[k_{zs}]$ have negative real parts, so a stationary solution always exists, i.e., the probabilities $\pi_i = \int_0^1 f_i(z) dz$ of being in state $S_i$, evolve to an equilibrium state given by the eigenvector $\pi$ associated with the eigenvalue $\lambda = 0$. Note that $\pi$ must be normalised so that the elements sum to 1. It can easily be shown that

$$\pi_i = \int_0^1 f_i(z) dz = \frac{k_1 k_2 k_3}{k_1 k_2 + k_1 k_3 + k_2 k_3} \hspace{1cm} (B4)$$

Now, integrating Eqs (B1) - (B3) and using Eq. (B4) we obtain the boundary values $f_1(1) = f_2(1) = 0$ and $f_1(0) = 0$. Also, summing Eqs (B1) - (B3) and integrating gives

$$-z f_1(z) - z f_2(z) + (1 - z) f_3(z) = C, \hspace{1cm} (B5)$$

where $C$ is a constant. Eq. (B5) is true for all $z \in [0, 1]$, so we can substitute in $z = 0$ or $z = 1$ and use the fact that $f_1(1) = 0$ and $f_1(1) = 0$, or $f_3(0) = 0$, to show that $C = 0$. Hence

$$f_z(z) = f_1(z) + f_2(z) + f_3(z) = \frac{1}{z} f_3(z), \hspace{1cm} (B6)$$

so we need only solve Eqs (B1) - (B3) for $f_z(z)$, the marginal probability density corresponding to the active state. Using (B6) and substituting into (B1) - (B3), we then obtain the following equation for $f_z(z)$:

$$0 = z^2 (1 - z) f_z''(z) + z \left[ -(3 + k_1 + k_2 + k_3) z + 1 - k_1 - k_2 \right] f_z'(z) + \left[ -k_1 k_2 (k_1 + k_2) c + c^2 \right] u(z)$$ \hspace{1cm} (B7)

Set $f(z) = z^c u(z)$, where $c$ is a constant that we can choose, to transform (B7) into

$$0 = z^{c+2} (1 - z) u''(z) + z^{c+1} \left[ -(3 + K - 2c) z + 1 - k_1 - k_2 + 2c \right] u'(z) + \left[ -k_1 k_2 (1 - k_1) k_2 c + c^2 \right] u(z)$$

where $K := k_1 + k_2 + k_3$ and $\tilde{K} := k_1 k_2 + k_1 k_3 + k_2 k_3$. From here, we set the last term on the right hand side to zero by choosing $c = k_1$ or $c = k_2$. We can then divide through by $z^{c+1}$ to obtain an equation in the form of the hypergeometric equation [56]. For example, for $c = k_1$ we obtain

$$0 = z (1 - z) u''(z) + \left[ -(3 - k_1 + k_2 + k_3) z + 1 + k_1 - k_2 \right] u'(z) + \left[ -1 - k_1 + k_2 - k_3 - k_2 k_3 \right] u(z).$$

When none of $\sqrt{(k_1 - k_2 - k_3)^2 - 4 k_2 k_3, k_1 - k_2, k_3}$ are integers, we can write down the solution [59]:

$$u(z) = C_1 z F \left[ a^{(1)}, a^{(1)}; 1 + k_1 - k_2; z \right] + C_2 z^{-k_1 + k_2} F \left[ a^{(2)}, a^{(2)}; 1 - k_1 + k_2; z \right]$$

where $C_1$ and $C_2$ are constants of integration.

The solutions for the other cases are similar and are also given in [59]. Hence $f_z(z) = z^k u(z)$ is given by

$$f_z(z) = C_1 z^{k_1} z F \left[ a^{(1)}, a^{(1)}; 1 + k_1 - k_2; z \right] + C_2 z^{k_2} z F \left[ a^{(2)}, a^{(2)}; 1 - k_1 + k_2; z \right].$$

The same expression for $f_z(z)$ is obtained if we choose $c = k_2$ instead, so finally we can write down the general solution for $f_z(z) = f_z(z)/z$:

$$f(z) = C_1 z^{k_1} z F \left[ a^{(1)}, a^{(1)}; 1 + k_1 - k_2; z \right] + C_2 z^{k_2} z F \left[ a^{(2)}, a^{(2)}; 1 - k_1 + k_2; z \right]$$

where

$$C_1 := \frac{k_1}{k_1 + k_2 + k_3} \quad \text{and} \quad C_2 := \frac{k_2}{k_1 + k_2 + k_3}.$$
are satisfied. These constants are conveniently obtained from a Mellin transform identity (see Ref. [71], p. 152):

\[
C_1 = \frac{\Gamma(k_2 - k_1)}{\Gamma(k_1) \Gamma(k_2)} \frac{\Gamma(1 + k_1 - a_+^{(1)})}{\Gamma(1 - a_+^{(1)})} \Gamma(1 + k_1 - a_-^{(1)}) \Gamma(1 - a_-^{(1)}),
\]

\[
C_2 = \frac{\Gamma(k_1 - k_2)}{\Gamma(k_1) \Gamma(k_2)} \frac{\Gamma(1 + k_2 - a_+^{(2)})}{\Gamma(1 - a_+^{(2)})} \Gamma(1 + k_2 - a_-^{(2)}) \Gamma(1 - a_-^{(2)}),
\]

where \(\Gamma(.)\) is the Gamma function [56]. Eq. (B10) is useful for comparisons with the 2-state random telegraph model.


[60] Alvaro Sánchez and Jané Kondev, “Transcriptional control of noise in gene expression,” Proceedings of the Na-


