

# Population growth affects intrinsic and extrinsic noise in gene expression

Philipp Thomas<sup>1</sup>

*Department of Mathematics, Imperial College London, London SW7 2AZ,  
UK<sup>a)</sup>*

Clonal cells of exponentially growing populations vary substantially from cell to cell. The main drivers of this heterogeneity are the population dynamics and stochasticity in the intracellular reactions, which are commonly studied separately. Here we develop an agent-based framework that allows tracking of the biochemical dynamics in every single cell of a growing population that accounts for both of these factors. Apart from the common intrinsic variability of the biochemical reactions, the framework also predicts extrinsic noise arising from fluctuations in the histories of cells without the need to introduce fluctuating rate constants. Instead, these extrinsic fluctuations are explained by cell cycle fluctuations and differences in cell age, which are ubiquitously observed in growing populations. We give explicit formulas to quantify mean molecule numbers, intrinsic and extrinsic noise statistics as measured in two-colour experiments. We find that these statistics may differ significantly depending on the experimental setup used to observe the cells. We illustrate this fact using (i) averages over an isolated cell lineage tracked over many generations as observed in the mother machine, (ii) snapshots of a growing population with known cell ages as recorded in time-lapse microscopy, and (iii) snapshots of unknown cell ages as measured from static images. Our integrated approach applies to arbitrary biochemical networks and generation time distributions. By employing models of stochastic gene expression and feedback regulation, we elucidate that isolated lineages, as compared to snapshot data, can significantly overestimate the mean number of molecules, overestimate extrinsic noise but underestimate intrinsic noise and have qualitatively different sensitivities to cell cycle fluctuations.

---

<sup>a)</sup>Electronic mail: [p.thomas@imperial.ac.uk](mailto:p.thomas@imperial.ac.uk)

# I Introduction

The behaviour of clonal cells varies substantially from cell to cell and over time<sup>1,2</sup>. Identifying the sources of these fluctuations can help us to understand how clonal cells diversify their responses and to reveal the function of gene circuits and signalling networks. For their quantification, it is often convenient to break down the experimentally observed variability into functional components. Commonly one wishes to separate fluctuations inherent in the circuit dynamics itself, called *intrinsic noise*, from fluctuations that arise from embedding the circuit in the environment of the cell, called *extrinsic noise*.

A possible resolution to this problem is to place and simultaneously measure a second independent copy of the circuit in the cell, as has been done in *E. coli*<sup>1</sup>, yeast<sup>3</sup>, mammalian cells<sup>4</sup> and plants<sup>5</sup>. The difference between the two circuit copies measures the intrinsic noise whereas their correlations measure the extrinsic noise component. Intrinsic noise arises from the random nature of the involved biochemical reactions. Extrinsic noise originates from factors affecting both circuits in the same way. These can, for instance, be modelled by reaction rates that fluctuate between cells or over time due to shared resources, promoter architecture or upstream pathways. Such sources of extrinsic noise have been studied extensively in the literature<sup>6–14</sup>.

A less commonly studied but equally important source of extrinsic noise is the population dynamics<sup>15</sup>. Since intracellular molecule numbers must double over the cell division cycle, a two-fold variation of expression levels is expected from cell proliferation alone. Moreover, the cell cycle itself is subject to tremendous variation providing an additional source of extrinsic variability. For example, generation times in *Escherichia coli*<sup>16</sup>, budding yeast<sup>17</sup> and mammalian cells<sup>18</sup> vary about 30 – 50%. These sources should therefore prevail in growing cells, populations and tissues.

Modelling approaches for understanding the effects of the cell cycle on gene expression noise are only recently being developed<sup>19–24</sup>. These studies are often restricted to a single isolated cell observed over successive cell divisions and measuring variability over time, similar to a lineage in the mother machine<sup>25</sup>. Many experiments, however, report cell-to-cell variability across snapshots of an exponentially growing cell population. These approaches either use time-lapse microscopy<sup>26,27</sup> or analyse snapshots with distributed cell ages as observed in flow cytometry, smFISH or similar techniques<sup>28–30</sup>.

Recent studies elucidated that population snapshots and lineages can significantly deviate from each other in their statistics<sup>24,31</sup>. To-date, however, there exists no general analytical framework with which to quantify the gene expression fluctuations in populations. We are thus lacking the means with which to understand, compare and integrate the knowledge from different experiments such as mother machines, time-lapse microscopy or fixed-cell images. Agent-based approaches allowing to track the expression levels of each individual cell in a growing population are ideally suited to address this issue.

In this manuscript, we develop such an approach to characterise the statistics of biochemical reaction networks in a growing and dividing cell population. In this framework, an agent is represented by a cell whose biochemical decomposition changes due to stochastic reaction kinetics and cell divisions. In Sec. II A we show how to analytically characterise the joint distribution of cell age and molecule content per cell in a snapshot of the population. We then, in Sec. II B, derive the exact moment equations of this model.

Since stochastic models are rarely solvable, we propose an analytically tractable approximation to mean and covariances in Sec. III A. Intrinsic and extrinsic noise sources as they are measured using two-reporter systems are in-built in the agent-based approach, and we explain how to decompose the apparent noise into the respective components. We further elaborate on the decomposition in cases where the cell age is unknown, a situation commonly encountered when analysing data from population snapshots or flow cytometry. We demonstrate how to practically compute the noise decomposition in Sec. III B, illustrate the results using a simple two-reporter system, and study how circuit dynamics can be tuned to suppress either intrinsic or extrinsic fluctuations.

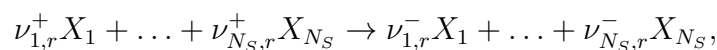
## II Methods

We model the dynamics of a dividing population of cell agents. The state of each cell is given by its age and the number of intracellular molecules, which evolve from birth to division. After cell division, the mother's molecules are inherited by the two daughters through stochastic partitioning of molecules. Cell divisions occur asynchronously in the population because cells divide at random times. In consequence, cell ages and molecule numbers are heterogeneous in the population. Fig. 1a illustrates the resulting branching process whose final state is a snapshot of the cell population.

The seminal experiment by *Elowitz et al.* identified the sources of cell-to-cell variation using snapshots of cells expressing green and red fluorescent reporters<sup>1</sup>. Reporters expressed at different levels appear either red or green, a signature of intrinsic noise. Cells with similar reporter levels appear yellow but with variable intensities, a signature of extrinsic noise. In our model (Fig. 1a), similar effects are observed since stochasticity in biochemical reactions and partitioning of molecules at division account for intrinsic variation across the population. Cell age and variability in division timing provide a source of extrinsic noise (Fig. 1b). In contrast to the population-view, the dynamics of isolated cells can also be tracked over time, which we will refer to as lineage (Fig. 1c), which corresponds to a random path in the tree. These statistics can differ significantly from population snapshots (Fig. 1d). To develop a quantitative understanding of these effects, we begin with deriving analytical framework to quantify these populations.

## A Agent-based framework for stochastic biochemical kinetics in growing cell populations

To each cell we associate an age  $\tau$  that counts the time since its last division and a number of set of biochemical species  $X_1, X_2, \dots, X_{N_S}$  present in amounts  $x = (x_1, x_2, \dots, x_{N_S})$ . These species interact via a network of  $R$  intracellular biochemical reactions of the type



where  $\nu_{ir}^\pm$  are the stoichiometric coefficients and  $r = 1, \dots, R$ . In the following, we outline the master equation that allows to analytically study these networks in an agent-based context.

### 1 Master equation for the agent-based population

The state of the overall cell population can be characterised by the snapshot density  $n(\tau, x, t)$  that counts the number of cells at time  $t$  with age between  $\tau$  and  $\tau + d\tau$  and molecule counts  $x$ . Accordingly, the total number of cells in the population is given by

$$N(t) = \int_0^\infty d\tau \sum_x n(\tau, x, t), \quad (1)$$

where the summation is over all possible molecule number configurations  $x$ .

We assume that cells divide with an age-dependent rate  $\gamma(\tau)$ , which is related to the interdivision time distribution  $\varphi(\tau_d)$  via

$$\gamma(\tau) = \frac{\varphi(\tau)}{\int_{\tau}^{\infty} d\tau_d \varphi(\tau_d)}. \quad (2)$$

The snapshot density then evolves due to age-progression of cells, cell divisions and the change in their molecular decomposition due to biochemical reactions

$$\left( \frac{\partial}{\partial t} + \frac{\partial}{\partial \tau} + \gamma(\tau) \right) n(\tau, x, t) = \mathbb{Q} n(\tau, x, t). \quad (3a)$$

Here, the change in the molecule numbers per cell is expressed by the transition matrix  $\mathbb{Q}$  acting as  $\mathbb{Q}n(\tau, x, t) = \sum_{r=1}^R [w_r(x - \nu_r) n(\tau, x - \nu_r, t) - w_r(x) n(\tau, x, t)]$ , where  $(\nu_r)_i = \nu_{ir}^+ - \nu_{ir}^-$  is the stoichiometric vector of the  $r^{th}$  reaction. Cell birth is described by the boundary condition

$$n(0, x, t) = 2 \int_0^{\infty} d\tau \sum_{x'} B(x|x') \gamma(\tau) n(\tau, x', t), \quad (3b)$$

whereby the mother cell is replaced with two daughter cells of zero age with its molecules being partitioned between them. The division kernel  $B(x|x')$  is the probability of partitioning the molecule numbers  $x'$  to  $x$  of any daughter cell and is given by

$$B(x|x') = \frac{1}{2} B_1(x|x') + \frac{1}{2} B_2(x|x'), \quad (4)$$

where  $B_1$  and  $B_2$  are the marginal probabilities for the two daughter cells to inherit  $x$  of the mother cell's molecules  $x'$ . Importantly, if the total amount is conserved in the division, we have  $B_2(x|x') = B_1(x' - x|x')$ . Thus each cell inherits an equal amount of molecules  $E_B[x|x'] = x'/2$ , because we do not distinguish the daughters.

Since resolving the time-evolution of the snapshot density is a formidable task, we concentrate on the long-term evolution of Eq. (3), which describes the exponential growth phase or balanced growth condition. In this limit, the total number of cells grows exponentially  $N(t) \sim N_0 e^{\lambda t}$  with rate  $\lambda$  and the fraction of cells with a certain cell age and molecule content is constant

$$\Pi(\tau, x) = \lim_{t \rightarrow \infty} \frac{n(\tau, x, t)}{N(t)}, \quad (5)$$

due to the balance between cell births, divisions and the biochemical reactions. In the following, we summarise how to characterise this distribution analytically.

## 2 Age-distribution and population growth rate

The fraction of cells with the same age in a snapshot is given by the age-distribution, which follows<sup>31</sup>

$$\Pi(\tau) = \lim_{t \rightarrow \infty} \sum_x \frac{n(\tau, x, t)}{N(t)} = 2\lambda e^{-\lambda\tau} \int_{\tau}^{\infty} d\tau' \varphi(\tau'). \quad (6)$$

The distribution  $\varphi$  characterises the interdivision times

$$\varphi(\tau_d) = \gamma(\tau_d) e^{-\int_0^{\tau_d} d\tau' \gamma(\tau')}, \quad (7)$$

as also seen from Eq. (2). The age distribution, Eq. (6), depends on the population growth rate  $\lambda$  that is uniquely determined by the Euler-Lotka equation

$$1 = 2 \int_0^{\infty} d\tau_d e^{-\lambda\tau_d} \varphi(\tau_d). \quad (8)$$

The above equations constitute the fundamental age-structure of microbial populations, which has been verified in experiments<sup>16,32,33</sup>.

## 3 Distribution of molecules for cells of the same age

We consider the total number of cells with age  $\tau$  and molecule count  $x$  divided by the number of cells at that age. This conditional probability quantifies the likelihood of observing  $x$  molecules in a cell of age  $\tau$  and is given by

$$\Pi(x|\tau) = \lim_{t \rightarrow \infty} \frac{n(\tau, x, t)}{\sum_x n(\tau, x, t)} = \frac{\Pi(\tau, x)}{\Pi(\tau)}. \quad (9)$$

It can be verified<sup>31</sup> that  $\Pi(x|\tau)$  obeys

$$\frac{\partial}{\partial \tau} \Pi(x|\tau) = \mathbb{Q} \Pi(x|\tau), \quad (10a)$$

which is similar to the chemical master equation (with time replaced by cell age). An important difference though, is that it has to be solved subject to the boundary condition

$$\Pi(x|0) = \int_0^{\infty} d\tau_d \sum_{x'} B(x|x') \Pi(x'|\tau_d) \rho(\tau_d), \quad (10b)$$

which accounts for the cell divisions. The distribution under the integral

$$\rho(\tau_d) = 2\varphi(\tau_d) e^{-\lambda\tau_d}, \quad (11)$$

is the interdivision time distribution in the population<sup>32,34</sup>. The distribution describes the interdivision times of cells with completed cell divisions and depends explicitly on the population growth rate  $\lambda$ .

#### 4 Comparison with the lineage framework

A lineage tracks a single of the population over successive cell divisions. In the long term its evolution also approaches a stable distribution, which we denote by  $\pi(\tau, x)$ . The molecule number distribution for cells of the same age in a lineage is given by  $\pi(x|\tau) = \pi(x, \tau)/\pi(\tau)$  and satisfies

$$\begin{aligned} \frac{\partial}{\partial \tau} \pi(x|\tau) &= \mathbb{Q} \pi(x|\tau), \\ \pi(x|0) &= \int_0^\infty d\tau_d \sum_{x'} B(x|x') \Pi(x'|\tau_d) \varphi(\tau_d). \end{aligned} \quad (12)$$

By comparing the above equations with Eqs. (10), we notice that this distribution is obtained by substituting the division time distribution  $\rho$  by  $\varphi$ . Thus cells of the same age can be analysed using a unified framework whether in populations or lineages. The age-distribution in a lineage, however, which is

$$\pi(\tau) = \frac{1}{E_\varphi[\tau]} \int_\tau^\infty d\tau' \varphi(\tau'), \quad (13)$$

differs significantly from the population, Eq. (6).

## B Moment equations for the agent-based model

In many practical situations, solving for the full distribution is infeasible. Summary statistics such as means and variances, which we will focus on in the following, present convenient alternatives as they are more amenable to analysis.

## 1 Exact moment equations for cells of the same age

141

In brief, the moment equations are obtained by multiplying Eq. (10a) by  $x$  or  $xx^T$  and summing over all possible states. The results for the first and second moments are

142

143

$$\frac{\partial E_{\Pi}[x|\tau]}{\partial \tau} = \sum_{r=1}^R \nu_r E_{\Pi}[w_r(x)|\tau], \quad (14a)$$

$$\begin{aligned} \frac{\partial E_{\Pi}[xx^T|\tau]}{\partial \tau} = & \sum_{r=1}^R \left( \nu_r E_{\Pi}[x^T w_r(x)|\tau] + E_{\Pi}[x w_r(x)|\tau] \nu_r^T \right) \\ & + \sum_{r=1}^R \left( \nu_r E_{\Pi}[w_r(x)|\tau] \nu_r^T \right). \end{aligned} \quad (14b)$$

Interestingly, these are the same moment equations that appear in the study of systems without age-dependence (with age being replaced by the observational time). The key difference is the boundary condition subject to which the moment equations have to be solved. These conditions follow from Eq. (10b) and the conservation of molecules in Eq. (4), which implies  $E_B[x|x'] = x'/2$ . They read

144

145

146

147

148

$$E_{\Pi}[x|0] = \frac{1}{2} E_{\rho}[E_{\Pi}[x|\tau_d]], \quad (14c)$$

$$E_{\Pi}[xx^T|0] = E_{\rho}[E_{\Pi}(E_B[x'x'^T|x]|\tau_d)]. \quad (14d)$$

The first condition states that, on average, molecule numbers need to double over one cell cycle. The second condition relates the second moments to the partitioning of molecules described by the division kernel, Eq. 4.

149

150

151

## 2 Exact moment equations for cells of unknown age

152

We now consider the snapshot moments of molecule numbers irrespective of age.  $E_{\Pi}[x] = E_{\Pi}[E_{\Pi}[x|\tau]]$  and  $E_{\Pi}[xx^T] = E_{\Pi}[E_{\Pi}[xx^T|\tau]]$  are obtained by multiplying Eq. (14) with  $\Pi(\tau)$  and performing the integration. For this purpose, consider the expected value of a function  $f(\tau)$  with respect to the age distribution, which satisfies

153

154

155

156

$$\begin{aligned} & \int_0^{\infty} d\tau \Pi(\tau) \frac{\partial f(\tau)}{\partial \tau} \\ & = - \int_0^{\infty} d\tau f(\tau) \frac{\partial}{\partial \tau} \Pi(\tau) - f(0) \Pi(0) \\ & = \lambda E_{\Pi}[f(\tau)] + \lambda E_{\rho}(f(\tau) - 2f(0)). \end{aligned} \quad (15)$$



In the first line, we integrate by parts assuming  $\lim_{\tau \rightarrow \infty} f(\tau)\Pi(\tau) = 0$ , and in the second line we substituted Eq. (6) for  $\Pi(\tau)$  and performed the derivative. The first term captures the effect of dilution, while the second term describes discrete changes during cell division.

Setting now  $f(\tau) = E_{\Pi}[x|\tau]$  in (15) and combining the result with Eq. (14a) and the boundary conditions (14c), we find an equation for the mean number of molecules in the population,

$$\lambda E_{\Pi}[x] = \sum_{r=1}^R \nu_r E_{\Pi}[w_r(x)]. \quad (16)$$

Similarly, using Eq. (15) with  $f(\tau) = E_{\Pi}[xx^T|\tau]$ , Eq. (14b) and (14d), the equation for the second moment becomes

$$\begin{aligned} & \lambda E_{\Pi}[xx^T] + \frac{\lambda}{2} E_{\rho}[E_{\Pi}(xx^T|\tau_d)] - 4E_{\Pi}(\text{Cov}_B[x'|x]|\tau) \\ &= \sum_{r=1}^R \left( \nu_r E_{\Pi}[x^T w_r(x)] + E_{\Pi}[x w_r(x)] \nu_r^T \right) \\ & \quad + \sum_{r=1}^R \nu_r E_{\Pi}[w_r(x)] \nu_r^T, \end{aligned} \quad (17)$$

where the left hand side depends explicitly on the division-time distribution  $\rho$ . Obviously, these equations cannot be solved in general, not only because the hierarchy of moments is not closed but also because they depend on moments for cells of known age. The conditions for which these equations are closed and can be solved exactly are discussed in SI V C.

### III Results

#### A Decomposing noise into intrinsic and extrinsic contributions

To circumvent the moment-closure problem, we employ the linear noise approximation to decompose the noise into intrinsic and extrinsic components (see<sup>35,36</sup> for details of the approximation). In brief, the approximation assumes Gaussian fluctuations and provides closed-form expressions for the mean molecule numbers and their covariances. Writing short  $\text{Cov}_{\Pi}[x|\tau] = E_{\Pi}[xx^T|\tau] - E_{\Pi}[x|\tau]E_{\Pi}[x^T|\tau]$  for the covariance matrix, the result is

$$\frac{\partial}{\partial \tau} E_{\Pi}[x|\tau] = \nu w(E_{\Pi}[x|\tau]), \quad (18a)$$

$$\frac{\partial}{\partial \tau} \text{Cov}_{\Pi}[x|\tau] = \mathcal{J} \text{Cov}_{\Pi}[x|\tau] + \text{Cov}_{\Pi}[x|\tau] \mathcal{J}^T + \mathcal{D}. \quad (18b)$$

where the Jacobian  $\mathcal{J}$  and the diffusion matrix  $\mathcal{D}$  are defined as

$$\begin{aligned}\mathcal{J}_{ij}(\tau) &= \sum_{r=1}^R \nu_{ir} \frac{\partial w_r(E_{\Pi}[x|\tau])}{\partial E_{\Pi}[x_j|\tau]}, \\ \mathcal{D}_{ij}(\tau) &= \sum_{r=1}^R \nu_{ir} \nu_{jr} w_r(E_{\Pi}[x|\tau]),\end{aligned}\tag{19}$$

which depend on cell age through the mean molecule numbers  $E_{\Pi}[x|\tau]$ . Comparison of Eqs. (18) with (14) shows that these equations are exact whenever the propensities are linear in the molecule numbers. In all other cases, we consider these as an approximation valid in the limit of large molecule numbers.

Next, we cast the boundary condition (14d) in terms of the covariance matrix  $\text{Cov}_{\Pi}[x|\tau]$ , which leads to

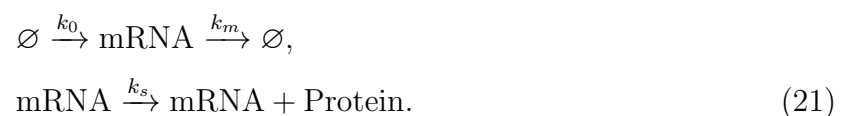
$$E_{\Pi}[x|0] = \frac{1}{2} E_{\rho}[E_{\Pi}(x|\tau_d)],\tag{20a}$$

$$\begin{aligned}\text{Cov}_{\Pi}[x_i, x_j|0] &= \underbrace{\frac{1}{4} E_{\rho}[\text{Cov}_{\Pi}(x_i, x_j|\tau)]}_{\text{variability between mother cells}} \\ &\quad + \underbrace{E_{\rho}[E_{\Pi}(\text{Cov}_B[x_i, x_j|x']|\tau)]}_{\text{partitioning of molecules}} \\ &\quad + \underbrace{\frac{1}{4} \text{Cov}_{\rho}[E_{\Pi}(x_j|\tau), E_{\Pi}(x_j|\tau)]}_{\text{cell cycle variability}}.\end{aligned}\tag{20b}$$

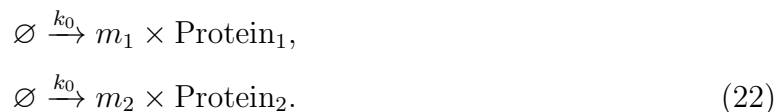
The first term is the contribution due to fluctuations in the number of molecules before division. The second term denotes the variation due to random partitioning of molecules at cell division, while the third contribution stems from differences in the molecule numbers due to different cell cycle lengths. We note that Eqs. (20b) themselves do not constitute a noise decomposition since these contribute do not propagate independently. Instead, they represent the sources of cell-to-cell variability for the two daughter cells.

## 1 Noise decomposition using dual reporter systems

To investigate how the different sources of variations affect biochemical reaction dynamics, we consider the synthesis and degradation of mRNA molecules and translation into proteins



We do not account for protein degradation explicitly in this model since stable proteins are diluted in the population by cell division, the effect of which we will study in the following. For simplicity, we assume that mRNA degradation is faster than the population growth such that the reactions can be approximated by a single reaction synthesising proteins in stochastic bursts. At the same time, for the purpose of the noise decomposition, we consider an additional, identical copy of the same circuit in the cell



The stochastic burst size of the first and second copy are denoted by  $m_1$  and  $m_2$ , respectively, and both follow a geometric distribution with mean  $b = k_s/k_m$  (see Ref.<sup>37</sup> and SI V D for details of the burst approximation).

*a. Mean number of proteins.*

Since the two reporter proteins are expressed identically in the cell, their mean expression levels must be the same. Denoting the protein numbers of the two reporters by  $x_1$  and  $x_2$ , we have  $E_{\Pi}[x_1|\tau] = E_{\Pi}[x_2|\tau]$ . The rate equation (18a) for the average number of proteins for a cell of given age becomes

$$\frac{\partial E_{\Pi}[x_1|\tau]}{\partial \tau} = k_0 b.$$

The solution that respects the boundary condition (20a) is

$$E_{\Pi}[x_1|\tau] = k_0 b(\tau + E_{\rho}[\tau]).\tag{23}$$

The number of proteins inherited after cell division ( $\tau = 0$ ) is thus  $k_0 b E_{\rho}[\tau]$ , which depends on the mean division time  $E_{\rho}[\tau]$  in the population.

*b. Separating noise into intrinsic and extrinsic components.*

For identical two-reporter systems, the overall variance can be decomposed into intrinsic and extrinsic components as follows

$$\Sigma(\tau) = \Sigma_{\text{int}}(\tau) + \Sigma_{\text{ext}}(\tau).\tag{24}$$

The individual contributions can be quantified using<sup>1</sup>

$$\begin{aligned}\Sigma_{\text{int}}(\tau) &= \frac{1}{2} E_{\Pi}[(x_1 - x_2)^2|\tau], \\ \Sigma_{\text{ext}}(\tau) &= \text{Cov}_{\Pi}[x_1, x_2|\tau].\end{aligned}\tag{25}$$

Since these components are measured in the same cell, they also account for the correct history dependence<sup>7,9,38</sup>.

The variance of intrinsic and extrinsic fluctuations follows from using Eqs. (25) in (18b) and rearranging, which leads to

$$\frac{\partial}{\partial \tau} \Sigma_{\text{int}}(\tau) = b(2b + 1)k_0, \quad \frac{\partial}{\partial \tau} \Sigma_{\text{ext}}(\tau) = 0.$$

Its solution is obtained by straight-forward integration and is given by  $\Sigma_{\text{int}}(\tau) = \Sigma_{\text{int}}(0) + b(2b + 1)k_0\tau$  and  $\Sigma_{\text{ext}}(\tau) = \Sigma_{\text{ext}}(0)$ . To fix the boundary condition (20b), we assume that each molecule of the mother cell being partitioned with equal probability between the two daughter cells. In this case, the division kernel in Eq. (4) is binomial with covariance  $\text{Cov}_B[x_i, x_j|x'] = \delta_{ij}x'_i/4$ . We then find that  $\Sigma_{\text{int}}(0) = \frac{1}{3}b(2b + 3)k_0E_\rho[\tau]$  and  $\Sigma_{\text{ext}}(0) = \frac{1}{3}b^2k_0^2E_\rho[\tau]^2\text{CV}_\rho^2$ , and finally

$$\text{CV}_{\text{int}}^2(\tau) = \frac{\Sigma_{\text{int}}(\tau)}{E_\Pi[x_1|\tau]^2} = \frac{1}{E_\Pi[p_1|\tau]} \left( 1 + \frac{2}{3}b \left( 1 + \frac{2\tau}{E_\rho[\tau] + \tau} \right) \right), \quad (26a)$$

$$\text{CV}_{\text{ext}}^2(\tau) = \frac{\Sigma_{\text{ext}}(\tau)}{E_\Pi[x_1|\tau]^2} = \frac{E_\rho[\tau]^2}{(E_\rho[\tau] + \tau)^2} \frac{\text{CV}_\rho^2}{3}. \quad (26b)$$

The coefficients of variations quantify the size of fluctuations relative to the mean. The result confirms the intuition that intrinsic noise decreases with the mean number of molecules. The extrinsic noise component, however, reflects the variations in cell cycle duration  $\text{CV}_\rho^2$  that are transmitted to the protein levels.

*c. Snapshots display higher intrinsic but lower extrinsic noise levels than lineages.*

Next, we compare the statistics of snapshots of a growing population with the one of a lineage of an isolated cell over time. As explained in Sec. II A 4, we obtain the lineage statistic by substituting the division time distribution  $\rho$  for  $\varphi$  in Eqs. (23), (26a) and (26b).

Interestingly, the deviations between these two statistics is apparent even on the mean level. To see this, we notice that the mean number of molecules, Eq. (23), increases with the duration of the cell cycle. It is well known that the cell cycle time is longer when averaged over single cells than for cells in the population<sup>16</sup>  $E_\rho[\tau] \leq E_\varphi[\tau]$ . An intuitive explanation of this fact is that fast dividing cells are over-represented in the population. It hence follows from Eq. (23) that the expected number of molecules is lower in populations compared to lineages, no matter what the division time distribution is.

In Fig. 2a, we compare the total noise  $\text{CV}_{\text{tot}}^2(\tau) = \text{CV}_{\text{int}}^2(\tau) + \text{CV}_{\text{ext}}^2(\tau)$  for gamma and log-normal distributed interdivision times. In both cases, we observe that the noise exhibits

a maximum for low cell cycle variability. With increasing cell cycle variability, we find that the maximum flattens in the lineage but not in snapshot statistics. Albeit the two statistics are collected from different samples of the same population, snapshots are more noisy than lineages in both cases. To understand this noise propagation, we decompose the total noise into intrinsic and extrinsic components via Eqs. (26a) and (26b). We observe that intrinsic noise in snapshots increases with cell cycle variability (Fig. 2b) while it is significantly lower in lineages and independent of these fluctuations, which is consistent with lower expression levels in snapshots.

Fig. 2a also reveals a non-monotonic dependence of the intrinsic noise component on cell age. To explain this phenomenon, we notice that intrinsic noise, Eq. (26a), increases with cell age due to an increase in the Fano factor. For older cells, however, intrinsic noise decreases with age as these cells express higher protein levels. Combining these findings explains the noise maximum at a well-defined cell age. By maximising Eq. (26a) over all possible cell ages, the age at which noise peaks is  $\tau/E_\rho[\tau_d] = \frac{2b-3}{6b+3}$  whenever  $b > 3/2$ . This ratio only depends on the burst size  $b$  and approaches 1/3 of the mean cell cycle time for large  $b$ . By contrast, we find that extrinsic noise is lower and decays slower over the cell cycle in snapshots than in lineages (Fig. 2c). We conclude that lineage statistics may significantly underestimate intrinsic heterogeneity but overestimate extrinsic noise in the population. In the next subsection, we extend this method to general stochastic reaction networks.

## 2 General decomposition for cells of the same age

We now generalise the decomposition to two-reporter systems involving an arbitrary network of biochemical reactions. As before, we assume that the two copies of our network with molecule numbers  $x_1$  and  $x_2$  do not interact and involve the same type reactions, and therefore they have the same mean expression level. The covariance of the two-reporter system is

$$\text{Cov}_\Pi[x_1, x_2|\tau] = \begin{pmatrix} \Sigma_{11}(\tau) & \Sigma_{12}(\tau) \\ \Sigma_{12}(\tau) & \Sigma_{11}(\tau) \end{pmatrix},$$

whose individual components satisfy

$$\begin{aligned}\frac{\partial}{\partial \tau} \Sigma_{11}(\tau) &= \mathcal{J} \Sigma_{11} + \Sigma_{11} \mathcal{J}^T + \mathcal{D}, \\ \frac{\partial}{\partial \tau} \Sigma_{12}(\tau) &= \mathcal{J} \Sigma_{12} + \Sigma_{12} \mathcal{J}^T,\end{aligned}\tag{27}$$

according to Eqs. (18b).

The intrinsic and extrinsic noise components can be expressed via

$$\begin{aligned}\Sigma_{\text{int}}(\tau) &= \frac{1}{2} E[(x_1 - x_2)^2 | \tau] = \Sigma_{11} - \Sigma_{12}, \\ \Sigma_{\text{ext}}(\tau) &= \text{Cov}[x_1, x_2 | \tau] = \Sigma_{12}.\end{aligned}$$

Since the covariances obey the linear equations (27), the two noise contributions evolve independently. In particular, the intrinsic and extrinsic covariances satisfy

$$\frac{d}{d\tau} \Sigma_{\text{int}} = \mathcal{J} \Sigma_{\text{int}} + \Sigma_{\text{int}} \mathcal{J}^T + \mathcal{D},\tag{28a}$$

$$\frac{d}{d\tau} \Sigma_{\text{ext}} = \mathcal{J} \Sigma_{\text{ext}} + \Sigma_{\text{ext}} \mathcal{J}^T,\tag{28b}$$

where only the intrinsic component involves the biochemical noise from the intracellular reactions through the diffusion matrix  $\mathcal{D}$ .

Assuming again binomial partitioning with covariance  $\text{Cov}_B[x_i, x_j | x'] = \delta_{ij} x'_i / 4$ , allows us to split the boundary condition (20b) according to Eq. (24). The results are two independent conditions

$$4\Sigma_{\text{int}}(0) = E_\rho[\Sigma_{\text{int}}(\tau_d)] + E_\rho[E_\Pi[\text{diag}(x) | \tau_d]],\tag{28c}$$

$$\begin{aligned}4\Sigma_{\text{ext}}(0) &= E_\rho[\Sigma_{\text{ext}}(\tau_d)] \\ &+ \text{Cov}_\rho[E_\Pi[x | \tau_d], E_\Pi[x | \tau_d]].\end{aligned}\tag{28d}$$

The noise decomposition is fully specified by the mean number of molecules for cells of the same age, the Jacobian  $\mathcal{J}$  of the corresponding rate equations, the diffusion matrix  $\mathcal{D}$  (see Eqs. (19)) and the distribution of interdivision times in the population  $\rho$  (see Eq. (11)). Importantly, Eq. (28c) shows that partitioning is a noise source to intrinsic fluctuations, while Eq. (28d) shows that cell cycle variations contribute to extrinsic fluctuations. We conclude that conditioning on the cell cycle position is not enough to eliminate all extrinsic noise. Before we continue, we note that other types of partitioning, such as asymmetric cell division, can be easily incorporated into the framework but using a different form of  $\text{Cov}_B[x_i, x_j | x']$ <sup>39</sup>.

### 3 General decomposition for cells of unknown age

An obstacle for applying this decomposition in practice is that in many situation the cell age is not known, and this is especially true for population snapshots. For this reason, the mean of the molecule number has to be averaged over all possible cell ages

$$E_{\Pi}[x] = E_{\Pi}(E_{\Pi}[x|\tau]) = \int_0^{\infty} d\tau E_{\Pi}[x|\tau]\Pi(\tau).$$

Similarly, we use the law of total variance to decompose the snapshot-variance as

$$\bar{\Sigma} = \underbrace{\bar{\Sigma}_{\text{int}}}_{\text{intrinsic}} + \underbrace{\bar{\Sigma}_{\text{cc}} + \bar{\Sigma}_{\text{age}}}_{\text{extrinsic}}, \quad (29a)$$

with

$$\begin{aligned} \bar{\Sigma}_{\text{int}} &= E_{\Pi}(\Sigma_{\text{int}}(\tau)), \\ \bar{\Sigma}_{\text{cc}} &= E_{\Pi}(\Sigma_{\text{ext}}(\tau)), \\ \bar{\Sigma}_{\text{age}} &= \text{Cov}_{\Pi}(E_{\Pi}[x|\tau], E_{\Pi}[x|\tau]). \end{aligned} \quad (29b)$$

The first term in Eq. (29a) is the intrinsic variance measured across a population, the second term is the extrinsic variance transmitted from cell cycle variations, and the third term is another extrinsic component that comes from averaging over cells of different ages. The total extrinsic noise, which is measured in a two-reporter system, is the sum of the second and third term. The practical use of this noise decomposition is demonstrated in the following section.

## B Practical computation of the noise decomposition and applications

Finally, we apply the noise decomposition to analyse snapshots in which the age of individual cells is not known. While the decomposition can be carried out exactly for linear reaction networks, we also outline a numerical method with which the decomposition can be carried out efficiently for complex nonlinear networks as we demonstrate for a protein that regulates its own expression.

# 1 Decomposition for linear reactions in cells of unknown age

For linear reaction networks in which the propensities are linear functions of the molecule numbers. This dependence allows to average the statistics exactly over all cell ages. Thus from Eq. (16), we obtain the rate equations

$$\lambda E_{\Pi}[x] = \nu w(E_{\Pi}[x]), \quad (30)$$

since  $E_{\Pi}[w(x)] = w(E_{\Pi}[x])$ . These equations coincide precisely with the steady state of the traditional deterministic rate equations including an effective dilution term proportional to the population growth rate  $\lambda$ .

Averaging Eq. (28a) over all ages and accounting for the boundary terms using Eq. (15), the intrinsic variance becomes

$$\begin{aligned} \lambda \bar{\Sigma}_{\text{int}} &= \mathcal{J} \bar{\Sigma}_{\text{int}} + \bar{\Sigma}_{\text{int}} \mathcal{J}^T + E_{\Pi}[\mathcal{D}(\tau)] \\ &+ \lambda (2\Sigma_{\text{int}}(0) - E_{\rho}[\Sigma_{\text{int}}(\tau_d)]), \end{aligned} \quad (31a)$$

where the Jacobian  $\mathcal{J}$  is assumed to be independent of cell age. Similarly, averaging Eq. (28b) the extrinsic variance transmitted from cell cycle fluctuations is obtained as

$$\begin{aligned} \lambda \bar{\Sigma}_{\text{cc}} &= \mathcal{J} \bar{\Sigma}_{\text{cc}} + \bar{\Sigma}_{\text{cc}} \mathcal{J}^T \\ &+ \lambda (2\Sigma_{\text{ext}}(0) - E_{\rho}[\Sigma_{\text{ext}}(\tau_d)]). \end{aligned} \quad (31b)$$

Similarly, an equation for  $\bar{\Sigma}_{\text{age}}$  can be derived (see SI V B for details), which reads

$$\begin{aligned} \lambda \bar{\Sigma}_{\text{age}} &= \mathcal{J} \bar{\Sigma}_{\text{age}} + \bar{\Sigma}_{\text{age}} \mathcal{J}^T \\ &+ \lambda \left( E_{\Pi}[x] E_{\Pi}[x^T] - 2E_{\Pi}[x|0] E_{\Pi}[x^T|0] \right. \\ &\quad \left. - \text{Cov}_{\rho}(E_{\Pi}[x|\tau_d], E_{\Pi}[x|\tau_d]) \right). \end{aligned} \quad (31c)$$

This decomposition exactly characterises the variability of linear intracellular reaction networks across snapshots.

## a. Application to stochastic reporter expression.

We return to the two-reporter system (22) and apply the decomposition developed in the previous section. From Eq. (30), we find that the mean molecule number is given by

$$E[p] = \begin{cases} k_0 b / \lambda & \text{snapshot,} \\ b k_0 (E_{\varphi}[\tau_d] + E_{\pi}[\tau]) & \text{lineage.} \end{cases} \quad (32)$$



Note that the lineage mean follows from integrating Eq. (23) with  $\varphi$  instead of  $\rho$  against the lineage age-distribution (13). Interestingly, only the population mean agrees with the intuition in which the ratio of synthesis and dilution rates yields the steady state levels. However, both averages depend implicitly on the cell cycle variability through the average age  $E_\pi[\tau]$  or the population growth rate  $\lambda$ , respectively. In Fig. 3a, we show that molecule numbers in the lineage increase with cell cycle variability while they decrease in the snapshot statistic under the same conditions. These quantities thus exhibit opposite sensitivities to cell cycle variability.

Next, we explore the noise properties of the reporter system using the decomposition (31). We find that the contributions of intrinsic noise are

$$CV_{\text{int}}^2 = \begin{cases} \frac{1}{E_\Pi[p]} (1 + \frac{2}{3}b(1 + 2\lambda E_\Pi[\tau])) & \text{snapshot,} \\ \frac{1}{E_\pi[p]} \left(1 + 2b - \frac{4bE_\varphi[\tau_d]}{3(E_\varphi[\tau_d] + E_\pi[\tau])}\right) & \text{lineage,} \end{cases}$$

which is inversely proportional to the mean number of proteins. The contribution of extrinsic noise due to stochasticity in cell cycle duration is

$$CV_{\text{cc}}^2 = \begin{cases} \frac{1}{3}(E_\rho[\tau_d]\lambda)^2 CV_\rho^2[\tau_d] & \text{snapshot,} \\ \frac{1}{3} \frac{E_\varphi[\tau_d]^2}{(E_\varphi[\tau_d] + E_\pi[\tau])^2} CV_\varphi^2[\tau_d] & \text{lineage,} \end{cases}$$

and the one due to the unknown cell age is

$$CV_{\text{age}}^2 = \begin{cases} (\lambda E_\Pi[\tau])^2 CV_\Pi^2[\tau] & \text{snapshot,} \\ \frac{E_\pi[\tau]^2}{(E_\varphi[\tau_d] + E_\pi[\tau])^2} CV_\pi^2[\tau] & \text{lineage.} \end{cases}$$

The noise decomposition crucially depends on the population growth rate  $\lambda$ , while in lineages it depends on the corresponding average cell age  $E_\pi[\tau]$ . More specifically,  $CV_{\text{cc}}^2$  and  $CV_{\text{age}}^2$  depend the variations in the age and interdivision time distributions, which are generally different in lineages and population. We illustrate this dependence using the analytical decompositions for two different interdivision-time distributions with the same mean and variance.

For both the gamma and the log-normal distribution, intrinsic noise (red lines, Fig. 3b) exhibits opposite sensitivities on cell cycle variability comparing lineage (dashed) and snapshot statistics (solid). This observation is explained by smaller mean expression levels in snapshots (cf. Fig. 3a) because intrinsic noise is expected to scale inversely with the mean

molecule number. For the gamma distribution, the extrinsic noise transmitted from cell cycle variations (blue lines, Fig. 3b) is (almost) identical for these measures. For the log-normal distribution, however, extrinsic noise in the lineage is smaller than in the snapshot. Interestingly, we find that the total noise is higher in snapshots than in lineages for the gamma distribution (black lines, Fig. 3c), while this not true for the log-normal distribution and large cell cycle variability (black lines, Fig. 3d).

In developed network models, the extrinsic components will also depend on the biochemical properties of the network. We demonstrate this analytically in SI V D when the protein is also subject to degradation, which reveals intricate noise patterns. A straight-forward approach for the noise decomposition in complex biochemical network is given in the following section.

## 2 Decomposition for nonlinear reaction networks

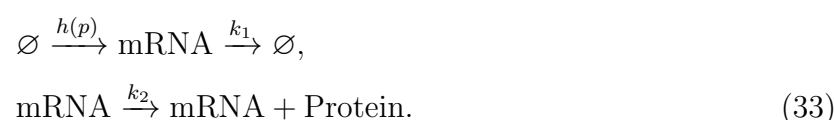
For nonlinear reaction networks, it is generally difficult to carry out the noise decomposition analytically. This is because the statistics of known and unknown cell age are intricately coupled and can be solved simultaneously only in simple cases. An efficient and generally applicable procedure to compute the numerical noise decomposition is the following:

1. Calculate the population growth rate using Eq. (8).
2. Solve for the statistics of cells of the same age, Eq. (28a) and (28b) and use the shooting method to match the boundary conditions (28c) and (28d).
3. Obtain the noise decomposition (29a) irrespectively of cell age by performing the average in Eqs. (29b).

Step 1 can be efficiently computed using numerical root-finding methods. The shooting method in Step 2 consists of an iterative procedure by which the mean molecule number, intrinsic variance and extrinsic variance are obtained through an initial guess on their values after cell division,  $E_\rho[x|0]$ ,  $\Sigma_{\text{int}}(0)$  and  $\Sigma_{\text{cc}}(0)$ , and the result is then refined using standard root-finding methods until the boundary conditions (28c) and (28d) are matched. Step 3 is easily carried out alongside the numerical integration of Step 2. The procedure typically evaluates the noise decomposition in seconds on a desktop computer and may therefore be adequate for statistical inference.

# *a. Suppressing intrinsic or extrinsic noise through feedback mechanisms.*

Over 40% of known transcription factors in *E. coli* regulate their own expression<sup>40</sup>. We here investigate the sensitivity of negative autoregulatory feedback to cell cycle fluctuations. We consider transcription and degradation of mRNA molecules from which proteins are synthesised



The effect of negative feedback is modelled via a Hill-function for the transcription rate  $h(p) = \frac{k_0}{1+(p/K)^4}$ , which decreases with the protein number  $p$ . This approximation is appropriate when the promoter-binding is extremely fast<sup>41</sup>.

In Fig. 4a we show mRNA levels in lineages decrease with cell cycle variability for various feedback strength (the inverse of the dissociation constant,  $1/K$ ). Mean mRNA numbers in the snapshot statistic either decrease (weak, moderate feedback) or increase with cell cycle variability (strong feedback) depending on the feedback strength. In contrast, protein levels increase with cell cycle variability in lineages but decrease in the snapshot for various feedback strengths (Fig. 4b). In agreement with this trend, we find that intrinsic noise always increases with cell cycle variability while the opposite holds for weak to moderate feedback but not for strong feedback (Fig. 4d). Strikingly, due to the negative feedback regulation, the sensitivity of intrinsic noise of mRNAs is the opposite (Fig. 4c). In contrast to the intrinsic noise properties of the circuit, the total extrinsic noise of the circuit always increases with cell cycle variability, both in lineages and snapshots (Fig. 4d,e). In all cases, our approximations are in good agreement with exact stochastic simulations (Fig. 4 dots) carried out using the First-Division Algorithm given in Ref.<sup>31</sup> including two non-interacting reporter networks.

Finally, we use the noise decomposition to understand how heterogeneity can be controlled by natural and synthetic circuits. Negative feedback is widely known to reduce noise but often requires fine-tuned parameters<sup>42–44</sup>. How this translates to individual functional noise components, such as intrinsic and extrinsic noise, has only been explored in response to parameter fluctuations<sup>6,45</sup> but not in the context of the ubiquitous population dynamics. Here, we are specifically interested in the sensitivity of lineage and population snapshot statistics to cell cycle noise.

In Fig. 5a we show that negative feedback can efficiently suppress intrinsic noise as the feedback strength is varied. Intriguingly, comparing the minimum noise levels in lineages and snapshots, vastly different values of the dissociation constants achieve noise suppression in these measures. To study this dependency in more detail, we compute the optimal feedback strength that minimises intrinsic noise as shown in Fig. 5b. Intriguingly, the optimal values exhibit opposite sensitivities to the cell cycle variability in lineages than in the population snapshots. To efficiently suppress intrinsic noise in a lineage, we must decrease the feedback strength in response to an increase in cell cycle variability. To compensate for intrinsic variability across the population, however, the feedback strength must increase by almost a two-fold of what would be required in the lineage.

In other situations, it may be advantageous to reduce the extrinsic instead of the intrinsic noise component. In Fig. 5c, we show that tuning the dissociation constant ( $K$ ) can similarly reduce the transmitted cell cycle noise  $CV_{cc}^2$ . Comparing the optimal feedback strength ( $1/K$ ) as a function of the cell cycle noise  $CV[\tau_d]$  (Fig. 5d), we observe that they increase with cell cycle variability in both lineage and snapshot. It is worth noting that vastly different feedback strengths achieve either intrinsic or extrinsic noise suppression (cf. Fig. 5c,d). These findings highlight that a single feedback loop may not be sufficient to simultaneously suppress both noise components whether in lineages or population snapshots.

## IV Discussion

We presented an analytical framework to analyse stochastic biochemical reactions in an exponentially growing cell population. This theory allows us to characterise and systematically decomposes cellular noise into intrinsic and extrinsic components, which applies to general stochastic biochemical networks. We found that a typical cell in the population expresses lower levels of proteins per cell than an isolated cell tracked over successive cell divisions. As a consequence, we observed higher levels of intrinsic noise but, for the examples studied, the extrinsic noise component was significantly reduced. These effects are most pronounced in the presence of division time variability as it is the case in natural populations. Importantly, this highlights that one needs to account for cell cycle fluctuations when modelling either intrinsic or extrinsic noise components.

Previous studies<sup>46,47</sup> focussed solely on the effect of age-structure but mostly neglected cell

cycle variations. We demonstrated that the statistics of lineages and population snapshots are not equivalent even when the cell cycle position is known. Although these differences appear to be small when divisions occur deterministically, they will be pronounced in the presence of division time variability (Figs. 3 and 4). In particular, we showed that measuring cells within a narrow range of cell cycle stages, as for instance achieved through gating<sup>8,28</sup>, does not eliminate all extrinsic noise due to cell cycle fluctuations. In reality, cells are affected by more than one type of extrinsic noise source as reaction rates may fluctuate over time and between cells<sup>6,7,13</sup>. These effects should be added to the transmitted extrinsic noise. We anticipate, however, that it will be difficult in practice to distinguish these fluctuations from the variations induced by cell cycle variability.

While this study focused on the effects of age-structure on biochemical dynamics, several simulation studies suggest that cell size also coordinates gene expression<sup>48,49</sup>. Incorporating additional physiological details such as cell size into our framework could thus provide insights to the statistics of intracellular concentrations<sup>50–52</sup> and extrinsic noise transduced from cell size control and growth rate fluctuations<sup>53</sup>. A different limitation of this study is that it is based on the linear noise approximation, which albeit being exact for linear reaction networks, represents an approximation assuming large molecule numbers. Its estimates can be inaccurate for nonlinear reaction networks involving low numbers of molecules. An improvement to this approximation could employ higher order terms in the system size expansion<sup>41,54</sup>, or close the hierarchy of moments using moment closure approximations<sup>55</sup>.

Heterogeneity inferred from snapshots is often used to say something about a cell's history. By grouping cells of similar ages, as in ergodic rate analysis<sup>56</sup>, one can in principle reconstruct time-course information. We demonstrated that such a procedure produces different results to the lineage statistic (see Fig. 2). Instead, the variability across the population is equivalent to choosing an arbitrary cell from the final population and tracing it backwards in time<sup>31</sup>. Although this equivalence provides a sample-path interpretation of snapshot data, it is worth pointing out that it does not apply when cell ages are unidentified. In this case, understanding the relationship between single cell fluctuations and population heterogeneity requires an agent-based framework as the one presented.

We showed that gene expression noise in populations is coupled to the population growth rate, as observed in population studies<sup>57</sup>. This dependency is crucial when quantifying summary statistics such as mean and variances. We found that the cellular heterogeneity

displays opposite sensitivities to cell cycle variability across populations and lineages. For negative feedback circuits, this implies that no parameter tuning enables cells to minimise noise of both measures efficiently (Fig. 5). Reducing noise in lineages over time comes at the cost of increased population heterogeneity, a strategy cells could exploit to diversify in response to stress<sup>58</sup>. Conversely, tuning snapshot homogeneity sacrifices lineage-optimality, which could confer advantages when gene expression couples to global physiological factors such as cell size, growth rate or cell division<sup>59</sup>.

Cells may thus perform suboptimally depending on which experimental setup is used to study them, whether it is a mother machine or a chemostat. Identifying the relevant noise components and cellular objectives will likely depend on the environmental and experimental conditions, or even on the particular application<sup>58,59</sup>. These dependencies thus reveal a fundamental trade-off for the evolution of natural circuits and the design of synthetic circuits in living cells.

In summary, we presented an agent-based framework for the statistical analysis of population snapshots. Inherent in this approach are several noise sources that reveal typical features of snapshot data using noise decompositions. The present framework is widely applicable and as such it also applies to large gene regulatory, signalling or metabolic networks. We, therefore, envision that the proposed moment-based approach could prove especially useful for parameter inference from snapshots of living cells<sup>60</sup>.

## V Supporting information

### A Statistics of interdivision times and age distributions

We here characterise the statistics of division times and age-distributions in lineages and populations. To this end, it is useful to recall the definition of the Laplace transform of the interdivision time distribution  $\varphi(\tau_d)$  in a lineage

$$\hat{\varphi}(s) = \int_0^\infty d\tau_d e^{-s\tau_d} \varphi(\tau_d). \quad (34)$$

We assume that either  $\varphi(\tau_d)$  is known or can be calculated from the division rate via Eq. (7).

## 1 Moments of interdivision times in a population

To calculate the statistics of the interdivision times in the population, we employ the Laplace transform of Eq. (11), which is

$$\hat{\rho}(s) = 2\hat{\varphi}(\lambda + s). \quad (35)$$

and note that  $\hat{\varphi}(\lambda) = 1/2$  due to characteristic equation (8). The moments can thus be expressed in terms of the Laplace transform

$$E_{\rho}[\tau_d^n] = 2 \left( -\frac{\partial}{\partial \lambda} \right)^n \hat{\varphi}(\lambda). \quad (36)$$

From these, we can compute

$$E_{\rho}[\tau_d] = -2\hat{\varphi}'(\lambda), \quad \text{Var}_{\rho}[\tau_d] = 2\hat{\varphi}''(\lambda) - 4\hat{\varphi}'(\lambda)^2. \quad (37)$$

and hence

$$\text{CV}_{\rho}^2[\tau_d] = \frac{\hat{\varphi}''(\lambda)}{2\hat{\varphi}'(\lambda)^2} - 1. \quad (38)$$

## 2 Age-distribution in lineages

The age-distribution yields the frequency of cell ages observed for different single cell measures. To compute the moments of the age-distribution in a lineage, we compute the Laplace transform of Eq. (13), which gives

$$\hat{\pi}(s) = \frac{1 - \hat{\varphi}(s)}{sE_{\varphi}[\tau_d]}. \quad (39)$$

By differentiating the above expression repeatedly at  $s = 0$ , we find

$$E_{\pi}[\tau] = \frac{E_{\varphi}[\tau^2]}{2E_{\varphi}[\tau]}, \quad \text{Var}_{\pi}[\tau] = \frac{E_{\varphi}[\tau^3]}{3E_{\varphi}[\tau]} \quad (40)$$

and

$$\text{CV}_{\pi}^2[\tau] = \frac{4E_{\varphi}[\tau]E_{\varphi}[\tau^3]}{3E_{\varphi}[\tau^2]}, \quad (41)$$

which concludes the the age-statistics in the population snapshot.

### 3 Age-distribution in populations

Similarly, we consider the Laplace transform of the age-distribution in a population snapshot, Eq. (6), which evaluates to

$$\begin{aligned}\hat{\Pi}(s) &= E_{\Pi}[e^{-s\tau}] = 2\lambda \int_0^{\infty} d\tau \Phi(\tau) e^{-(s+\lambda)\tau} \\ &= \frac{2\lambda}{s+\lambda} (1 - \hat{\varphi}(s+\lambda)),\end{aligned}\quad (42)$$

Repeated differentiation at  $s = 0$ , gives

$$E_{\Pi}[\tau] = \frac{1}{\lambda} + 2\hat{\varphi}'(\lambda), \quad \text{Var}_{\Pi}[\tau] = \frac{1}{\lambda^2} - 2\hat{\varphi}''(\lambda) - 4\hat{\varphi}'(\lambda)^2. \quad (43)$$

It is straightforward to evaluate these statistics numerically as we do for the log-normal distribution in Fig. 2 and 3. For the gamma-distribution, the population growth rate, the age- and interdivision-time distributions can be obtained in close form as we show in the following.

### 4 Gamma distribution: Explicit solutions to population growth rate and the age/interdivision time distributions

We fix the division time to be gamma distributed with density function

$$\varphi(\tau_d) = \frac{e^{-\frac{\tau_d}{c\mu}} \left(\frac{\tau_d}{c\mu}\right)^{1/c}}{\Gamma\left(\frac{1}{c}\right) \tau_d}, \quad (44)$$

where  $\Gamma$  is the gamma function, such that  $E_{\varphi}[\tau_d] = \mu$  and  $c = \text{CV}_{\varphi}^2[\tau_d]$ . The Laplace transform of the distribution is

$$\hat{\varphi}(s) = (c\mu s + 1)^{-1/c}. \quad (45)$$

Recasting now the Euler-Lotka equation in the form  $\hat{\varphi}(\lambda) = 1/2$ , we can solve for  $\lambda$  to obtain

$$\lambda = \frac{2^c - 1}{c\mu}. \quad (46)$$

The division time distribution in the population is then given by

$$\rho(\tau_d) = 2e^{-\lambda\tau_d} \varphi(\tau_d) = \frac{2e^{-\frac{2^c\tau_d}{c\mu}} \left(\frac{\tau_d}{c\mu}\right)^{1/c}}{\Gamma\left(\frac{1}{c}\right) \tau_d}, \quad (47)$$



which is also a gamma distribution but with a shorter mean division time but the same  
coefficient of variation

$$E_{\rho}[\tau_d] = 2^{-c}\mu, \quad \text{CV}_{\rho}^2[\tau_d] = c. \quad (48)$$

Further, the age-distribution in a lineage is

$$\pi(\tau) = \frac{1}{\mu} \frac{\Gamma\left(\frac{1}{c}, \frac{\tau}{c\mu}\right)}{\Gamma\left(\frac{1}{c}\right)}, \quad (49)$$

where  $\Gamma(\cdot, \cdot)$  is the upper incomplete gamma function. Its statistics are with statistics

$$E_{\pi}[\tau] = \frac{\mu}{2}(1+c), \quad \text{CV}_{\pi}^2[\tau] = \frac{5c+1}{3c+3}. \quad (50)$$

Similarly, the age-distribution in the population becomes

$$\Pi(\tau) = 2\lambda e^{-\lambda\tau} \frac{\Gamma\left(\frac{1}{c}, \frac{\tau}{c\mu}\right)}{\Gamma\left(\frac{1}{c}\right)} \quad (51)$$

with statistics

$$E_{\Pi}[\tau] = \mu \left( \frac{c}{2^c - 1} - \frac{1}{2^c} \right),$$

$$\text{CV}_{\Pi}^2[\tau] = \frac{2c}{c + 2^{-c} - 1} - \frac{(2^c - 1)^2 (c + 1)}{(2^c(c - 1) + 1)^2} - 1. \quad (52)$$

Interestingly, it follows that  $E_{\Pi}[\tau] < E_{\pi}[\tau]$ , but  $\text{CV}_{\Pi}^2[\tau] > \text{CV}_{\pi}^2[\tau]$  for  $c < 1$  and  $\text{CV}_{\Pi}^2[\tau] < \text{CV}_{\pi}^2[\tau]$  for  $c > 1$ .

## B An explicit formula for the uncertainty due to unknown cell age

We here verify Eq. (31c) of the main text, which holds for linear reaction networks. To this end we define  $\epsilon(\tau) = E_{\Pi}[x|\tau] - E_{\Pi}[x]$  such that  $\bar{\Sigma}_{\text{age}} = E_{\Pi}[\epsilon\epsilon^T]$  and use Eq. (18a) to write

$$\frac{\partial}{\partial\tau}\epsilon(\tau) = \nu w(E[x|\tau]) = c + \mathcal{J}(\epsilon(\tau) + E_{\Pi}[x]).$$

We used the fact that  $\nu w(E[x|\tau]) = c + \mathcal{J}E[x|\tau]$ , where  $c$  is a constant vector, since for linear reaction networks the propensities are linear in the number of molecules. Making use of Eq. (53) we then compute

$$\begin{aligned} E_{\Pi} \left[ \frac{\partial}{\partial\tau}\epsilon(\tau)\epsilon^T(\tau) \right] &= E_{\Pi} \left[ \epsilon(\tau) \frac{\partial}{\partial\tau}\epsilon^T(\tau) \right] + E_{\Pi} \left[ \epsilon^T(\tau) \frac{\partial}{\partial\tau}\epsilon(\tau) \right] \\ &= \mathcal{J}\bar{\Sigma}_{\text{age}} + \bar{\Sigma}_{\text{age}}\mathcal{J}^T. \end{aligned}$$

On the other hand, using Eq. (15) of the main text, it follows that

$$\begin{aligned} E_{\Pi} \left[ \frac{\partial}{\partial \tau} \epsilon(\tau) \epsilon^T(\tau) \right] &= \lambda \left( \bar{\Sigma}_{\text{age}} + E_{\rho}[\epsilon(\tau) \epsilon^T(\tau)] - 2\epsilon(0) \epsilon^T(0) \right) \\ &= \lambda \left( \bar{\Sigma}_{\text{age}} + \text{Cov}_{\rho}[E[x|\tau]] - E_{\Pi}[x] E_{\Pi}[x^T] + 2E_{\rho}[x|0] E_{\rho}[x^T|0] \right). \end{aligned}$$

Combining the last two equation gives the result (31c) of the main text.

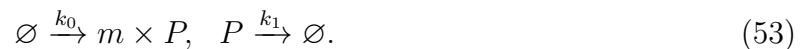
## C Detailed discussion of the moment-closure conditions

While the moment equations derived in Sec. II B are exact, the equations for cells of the same age are only closed when  $w_r(x)$  depends at most linearly on the molecule numbers  $x$  and the covariance of the partitioning kernel  $\text{Cov}_B[x|x']$  depends at most quadratically on the number molecules in the mother cell  $x'$ . This holds, for instance, for biochemical composed solely from unimolecular reactions and independent binomial partitioning. Similarly, it holds true for the mean of cells with unknown age, but not generally for their corresponding variances. Specifically, the covariance for cells of unknown age also depends on the moments for cells of known age and thus they must explicitly depend on the division time distribution.

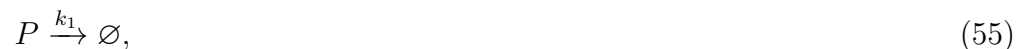
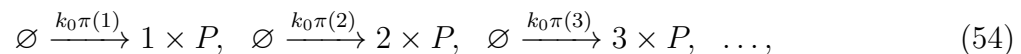
There are now two scenarios in which the variances are independent of the division time distribution. The first case is when the age-distribution coincides with the division time distribution  $\Pi(\tau) = \rho(\tau)$ , which follows only when the division rate  $\gamma$  is constant and independent of age, i.e. the division times are exponentially distributed. The second case assumes a particular division kernel  $B(x|x')$  that satisfies  $\text{Cov}_B[x|x'] = \frac{1}{4}x'x'^T$ , which follows when all molecules are inherited by only one of the daughter cells. In all other cases, which seem most relevant in practice, the moment equations for unknown cell age involve the moments for cells of known age. Thus, for general nonlinear reaction networks, they involve two hierarchies of moments that cannot be easily closed. A simple and generally applicable approximation that circumvents this problem is given in Sec. III A using the linear noise approximation.

## D Analytical noise decomposition for gene expression with degradation

We consider a simple system in which a protein  $P$  is translated in stochastic bursts of size  $m$  and is subsequently degraded



Because the burst size is a random variable we can recast the synthesis reaction into a series of reactions with reaction rates  $k_0\pi(m)$ ,



where  $\pi(m)$  is the distribution of burst sizes, which is geometric for the two-stage model of gene expression.

### 1 Mean protein number of the same age

The equation for the mean number of molecules is then

$$\frac{\partial E_{\Pi}[x|\tau]}{\partial \tau} = k_0 \left( \sum_{m=1}^{\infty} m\pi(m) \right) - k_1 E_{\Pi}[x|\tau] = bk_0 - k_1 E_{\Pi}[x|\tau] \quad (56)$$

with solution

$$E_{\Pi}[x|\tau] = E_{\Pi}[x|0]e^{-k_1\tau} + \frac{bk_0}{k_1} (1 - e^{-k_1\tau}) \quad (57)$$

Substituting the solution into the boundary condition  $E_{\rho}(E_{\Pi}[x|\tau]) = 2E_{\Pi}[x|0]$  and solving for  $E_{\Pi}[x|0]$  yields the final result

$$E_{\Pi}[x|\tau] = \frac{bk_0}{k_1} \left( 1 - \frac{e^{-k_1\tau}}{2 - \hat{\rho}(k_1)} \right), \quad (58)$$

where  $\hat{\rho}$  is the Laplace transform of the division time distribution

$$\hat{\rho}(s) = \int_0^{\infty} d\tau_d e^{-s\tau_d} \rho(\tau_d). \quad (59)$$

As we will show in the following the statistics of gene expression in growing populations depends crucially on this function.

## 2 Protein fluctuations for cells of the same age

568

To compute the protein fluctuations, we see from (58), that the Jacobian is

569

$$\mathcal{J} = -k_1. \quad (60)$$

The diffusion matrix then follows from Eq. (19) then follows

570

$$\mathcal{D} = k_0 \sum_{m=1}^{\infty} m^2 \pi(m) + k_1 E_{\Pi}[x|\tau] = k_0 b(2b+1) + k_1 E_{\Pi}[x|\tau]. \quad (61)$$

The variance of intrinsic and extrinsic fluctuations obeys Eqs. (18b, which read explicitly

571

$$\frac{\partial}{\partial \tau} \Sigma_{\text{int}}(\tau) = -2k_1 \Sigma_{\text{int}}(\tau) + k_0 b(2b+1) + k_1 E_{\Pi}[x|\tau] \quad (62)$$

$$\frac{\partial}{\partial \tau} \Sigma_{\text{ext}}(\tau) = -2k_1 \Sigma_{\text{ext}}(\tau) \quad (63)$$

$$\begin{aligned} \Sigma_{\text{int}}(\tau) = & \frac{bk_0 e^{-k_1 t}}{k_1 (g(k_1) - 2)} + \frac{b(b+1)k_0}{k_1} \\ & + e^{-2k_1 t} \left( \frac{bk_0 (-(b+1)g(k_1) + 2b+1)}{k_1 (g(k_1) - 2)} + \Sigma_{\text{int}}(0) \right) \end{aligned} \quad (64)$$

$$\Sigma_{\text{ext}}(\tau) = \Sigma_{\text{ext}}(0) e^{-2k_1 t}, \quad (65)$$

where  $\Sigma_{\text{int}}(0)$  and  $\Sigma_{\text{ext}}(0)$  are the intrinsic and extrinsic variation at cell division, which

572

have to be determined from the boundary conditions. According to Eq. (28c) and (28d),

573

the boundary conditions are

574

$$\begin{aligned} 4\Sigma_{\text{int}}(0) &= E_{\rho}[\Sigma_{\text{int}}(\tau_d)] + E_{\rho}[E_{\Pi}[x|\tau_d]], \\ 4\Sigma_{\text{ext}}(0) &= E_{\rho}[\Sigma_{\text{ext}}(\tau_d)] + \text{Cov}_{\rho}[E_{\Pi}[x|\tau_d], E_{\Pi}[x|\tau_d]]. \end{aligned} \quad (66)$$

To compute these values we notice that the variances at cell division follow from averaging

575

Eqs. (64) over the division time distribution  $\rho$ , which results in

576

$$\begin{aligned} E_{\rho}[\Sigma_{\text{int}}(\tau_d)] &= \frac{bk_0 \hat{\rho}(k_1)}{k_1 (g(k_1) - 2)} + \frac{b(b+1)k_0}{k_1} \\ &+ \hat{\rho}(2k_1) \left( \frac{bk_0 (-(b+1)g(k_1) + 2b+1)}{k_1 (g(k_1) - 2)} + \Sigma_{\text{int}}(0) \right) \\ E_{\rho}[\Sigma_{\text{ext}}(\tau_d)] &= \Sigma_{\text{ext}}(0) \hat{\rho}(2k_1). \end{aligned} \quad (67)$$

Further, we evaluate

577

$$E_{\rho}(E_{\Pi}[x|\tau_d]) = 2E_{\Pi}[x|0] = 2\frac{bk_0}{k_1} \left( \frac{1 - \hat{\rho}(k_1)}{2 - \hat{\rho}(k_1)} \right), \quad (68)$$

and

578

$$\begin{aligned} \text{Cov}_{\rho}[E_{\Pi}[x|\tau_d], E_{\Pi}[x|\tau_d]] &= E_{\rho}(E_{\Pi}[x|\tau] - E_{\rho}(E_{\Pi}[x|\tau]))^2 \\ &= \left( \frac{bk_0}{k_1(2 - \hat{\rho}(k_1))} \right)^2 E_{\rho}(\hat{\rho}(k_1) - e^{-k_1\tau})^2 \\ &= \left( \frac{bk_0}{k_1(2 - \hat{\rho}(k_1))} \right)^2 (\hat{\rho}(2k_1) - \hat{\rho}^2(k_1)). \end{aligned} \quad (69)$$

Plugging Eqs. (67), (68) and (69) into (66), solving for  $\Sigma_{\text{int}}(0)$  and  $\Sigma_{\text{ext}}(0)$  and using the result in Eqs. (64), we finally arrive at

579

580

$$\begin{aligned} \Sigma_{\text{int}}(\tau) &= \frac{3b^2k_0e^{-2k_1t}}{k_1(\hat{\rho}(2k_1) - 4)} + \frac{bk_0e^{-k_1t}}{k_1(\hat{\rho}(k_1) - 2)} + \frac{(b+1)bk_0}{k_1}, \\ \Sigma_{\text{ext}}(\tau) &= \frac{b^2k_0^2(\hat{\rho}(2k_1) - \hat{\rho}^2(k_1))e^{-2k_1t}}{k_1^2(\hat{\rho}(k_1) - 2)^2(4 - \hat{\rho}(2k_1))}, \end{aligned} \quad (70)$$

which determines the progression of intrinsic and extrinsic fluctuations over the cell cycle.

581

### 3 Protein statistics for cells of unknown age

582

The mean protein number is given by

583

$$E_{\Pi}[x] = \frac{bk_0}{k_1 + \lambda}. \quad (71)$$

Thus the mean number is determined from the balance between the rates of translation, degradation and dilution due to cell divisions. From Eqs. (31) we compute

584

585

$$\bar{\Sigma}_{\text{int}} = \frac{b(b+1)k_0}{k_1} - \frac{b\lambda k_0}{\lambda k_1 + k_1^2} + \frac{3b^2\lambda k_0(\hat{\rho}(2k_1) - 2)}{k_1(4 - \hat{\rho}(2k_1))(\lambda + 2k_1)}, \quad (72)$$

$$\bar{\Sigma}_{\text{ext}} = \frac{b^2\lambda k_0^2(\hat{\rho}(2k_1) - \hat{\rho}^2(k_1))(2 - \hat{\rho}(2k_1))}{k_1^2(2 - \hat{\rho}(k_1))^2(4 - \hat{\rho}(2k_1))(\lambda + 2k_1)}, \quad (73)$$

$$\bar{\Sigma}_{\text{age}} = \frac{b^2\lambda k_0^2(2 - \hat{\rho}(2k_1))}{k_1^2(\hat{\rho}(k_1) - 2)^2(\lambda + 2k_1)} - \frac{b^2\lambda^2 k_0^2}{k_1^2(\lambda + k_1)^2}. \quad (74)$$

Finally, we compute  $\text{CV}_y^2 = \bar{\Sigma}_y/E_{\Pi}[x]^2$  to arrive at the expressions for the coefficient of

586

variations

587

$$CV_{\text{int}}^2 = \frac{1}{E_{\Pi}[x]} \left( 1 + b \left( 1 + \frac{b\lambda}{k_1} \right) - \frac{3b\lambda (2 - \hat{\rho}(2k_1)) (\lambda + k_1)}{k_1 (4 - \hat{\rho}(2k_1)) (\lambda + 2k_1)} \right), \quad (75)$$

$$CV_{\text{cc}}^2 = \frac{\lambda (\hat{\rho}(2k_1) - \hat{\rho}(k_1))^2 (\hat{\rho}(2k_1) - 2) (\lambda + k_1)^2}{k_1^2 (\hat{\rho}(k_1) - 2)^2 (\hat{\rho}(2k_1) - 4) (\lambda + 2k_1)}, \quad (76)$$

$$CV_{\text{age}}^2 = \frac{\lambda (2 - \hat{\rho}(2k_1)) (\lambda + k_1)^2}{k_1^2 (\hat{\rho}(k_1) - 2)^2 (\lambda + 2k_1)} - \frac{\lambda^2}{k_1^2}, \quad (77)$$

which denote the intrinsic noise, the transmitted noise from cell cycle fluctuations and the 588  
uncertainty due to the unknown cell age. It is obvious that these expressions are much 589  
more involved than for the case without degradation because they depend on the Laplace 590  
transform  $\hat{\rho}$  of the division time distribution. Interestingly, the last two components, which 591  
together represent the extrinsic noise, also depend on the degradation rate  $k_1$  meaning that 592  
the total extrinsic noise is constant only when measured the mean concentration is varied 593  
through the either transcription rate or burst size. 594

## Acknowledgments

595

It is a pleasure to thank Vahid Shahrezaei for valuable feedback.

596

## Funding

597

PT acknowledges a fellowship by The Royal Commission for the Exhibition of 1851 and 598  
support by the EPSRC Centre for Mathematics of Precision Healthcare (EP/N014529/1). 599

## References

- <sup>1</sup>M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, “Stochastic gene expression in a single cell,” *Science* **297**, 1183–1186 (2002).
- <sup>2</sup>J. M. Raser and E. K. O’shea, “Noise in gene expression: origins, consequences, and control,” *Science* **309**, 2010–2013 (2005).
- <sup>3</sup>J. M. Raser and E. K. O’shea, “Control of stochasticity in eukaryotic gene expression,” *Science* **304**, 1811–1814 (2004).

- <sup>4</sup>C. V. Harper, B. Finkenstädt, D. J. Woodcock, S. Friedrichsen, S. Semprini, L. Ashahrezaei, D. G. Spiller, J. J. Mullins, D. A. Rand, J. R. Davis, *et al.*, “Dynamic analysis of stochastic transcription cycles,” *PLoS Biol* **9**, e1000607 (2011).
- <sup>5</sup>I. S. Araújo, J. M. Pietsch, E. M. Keizer, B. Greese, R. Balkunde, C. Fleck, and M. Hülkamp, “Stochastic gene expression in *Arabidopsis thaliana*,” *Nat Commun* **8**, 2132 (2017).
- <sup>6</sup>V. Shahrezaei, J. F. Ollivier, and P. S. Swain, “Colored extrinsic fluctuations and stochastic gene expression,” *Mol Syst Biol* **4**, 196 (2008).
- <sup>7</sup>A. Hilfinger and J. Paulsson, “Separating intrinsic from extrinsic fluctuations in dynamic biological systems,” *Proc Natl Acad Sci* **108**, 12167–12172 (2011).
- <sup>8</sup>C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl, “Moment-based inference predicts bimodality in transient gene expression,” *Proc Natl Acad Sci* **109**, 8340–8345 (2012).
- <sup>9</sup>C. G. Bowsher, M. Voliotis, and P. S. Swain, “The fidelity of dynamic signaling by noisy biomolecular networks,” *PLoS Comput Biol* **9**, e1002965 (2013).
- <sup>10</sup>A. Singh and M. Soltani, “Quantifying intrinsic and extrinsic variability in stochastic gene expression models,” *PLoS One* **8**, e84301 (2013).
- <sup>11</sup>J. Hasenauer, C. Hasenauer, T. Hucho, and F. J. Theis, “Ode constrained mixture modelling: a method for unraveling subpopulation structures and dynamics,” *PLoS Comput Biol* **10**, e1003686 (2014).
- <sup>12</sup>M. Voliotis, P. Thomas, R. Grima, and C. G. Bowsher, “Stochastic simulation of biomolecular networks in dynamic environments,” *PLoS Comput Biol* **12**, e1004923 (2016).
- <sup>13</sup>J. Dattani and M. Barahona, “Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization,” *J Royal Soc Interface* **14**, 20160833 (2017).
- <sup>14</sup>M. Del Giudice, S. Bo, S. Grigolon, and C. Bosia, “On the role of extrinsic noise in microRNA-mediated bimodal gene expression,” *PLoS Comput Biol* **14**, e1006063 (2018).
- <sup>15</sup>D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty, “Origins of extrinsic variability in eukaryotic gene expression,” *Nature* **439**, 861–864 (2006).
- <sup>16</sup>M. Hashimoto, T. Nozoe, H. Nakaoka, R. Okura, S. Akiyoshi, K. Kaneko, E. Kussell, and Y. Wakamoto, “Noise-driven growth rate gain in clonal cellular populations,” *Proc Natl Acad Sci* , 201519412 (2016).

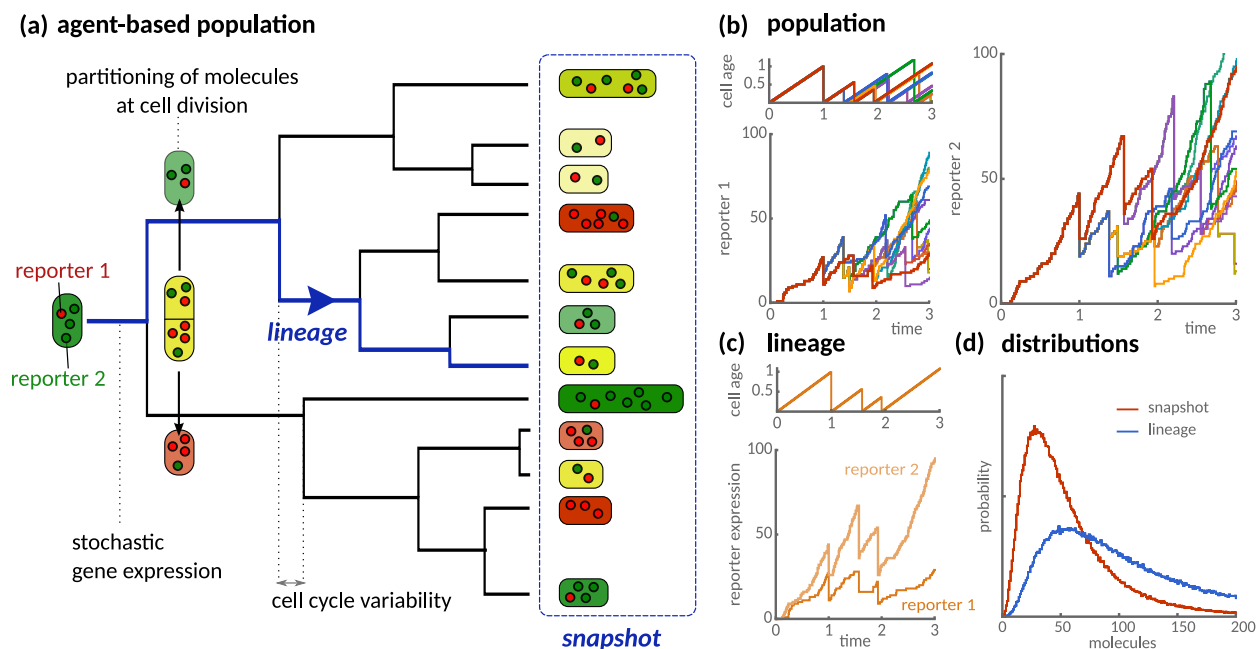
- <sup>17</sup>B. Cerulus, A. M. New, K. Pougach, and K. J. Verstrepen, “Noise and epigenetic inheritance of single-cell division times influence population fitness,” *Curr Biol* **26**, 1138–1147 (2016).
- <sup>18</sup>F. Shen, L. Hodgson, A. Rabinovich, O. Pertz, K. Hahn, and J. H. Price, “Functional proteometrics for cell migration,” *Cytometry Part A* **69**, 563–572 (2006).
- <sup>19</sup>P. S. Swain, M. B. Elowitz, and E. D. Siggia, “Intrinsic and extrinsic contributions to stochasticity in gene expression,” *Proc Natl Acad Sci* **99**, 12795–12800 (2002).
- <sup>20</sup>J. Lloyd-Price, H. Tran, and A. S. Ribeiro, “Dynamics of small genetic circuits subject to stochastic partitioning in cell division,” *J Theor Biol* **356**, 11–19 (2014).
- <sup>21</sup>V. Bierbaum and S. Klumpp, “Impact of the cell division cycle on gene circuits,” *Phys Biol* **12**, 066003 (2015).
- <sup>22</sup>I. G. Johnston and N. S. Jones, “Closed-form stochastic solutions for non-equilibrium dynamics and inheritance of cellular components over many cell divisions,” *Proc Royal Soc A* **471**, 20150050 (2015).
- <sup>23</sup>M. Soltani, C. A. Vargas-Garcia, D. Antunes, and A. Singh, “Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes,” *PLoS Comput Biol* **12**, e1004972 (2016).
- <sup>24</sup>A. Roy and S. Klumpp, “Simulating genetic circuits in bacterial populations with growth heterogeneity,” *Biophys J* **114**, 484–492 (2018).
- <sup>25</sup>P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, “Robust growth of *Escherichia coli*,” *Curr Biol* **20**, 1099–1103 (2010).
- <sup>26</sup>J. C. Locke and M. B. Elowitz, “Using movies to analyse gene circuit dynamics in single cells,” *Nat Rev Microbiol* **7**, 383 (2009).
- <sup>27</sup>D. Muzzey and A. van Oudenaarden, “Quantitative time-lapse fluorescence microscopy in single cells,” *Annu Rev Cell Dev Biol* **25**, 301–327 (2009).
- <sup>28</sup>J. R. Newman, S. Ghaemmamghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman, “Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise,” *Nature* **441**, 840 (2006).
- <sup>29</sup>L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, “General properties of transcriptional time series in *Escherichia coli*,” *Nat Genet* **43**, 554 (2011).
- <sup>30</sup>B. Okumus, D. Landgraf, G. C. Lai, S. Bakshi, J. C. Arias-Castro, S. Yildiz, D. Huh, R. Fernandez-Lopez, C. N. Peterson, E. Toprak, *et al.*, “Mechanical slowing-down of cy-



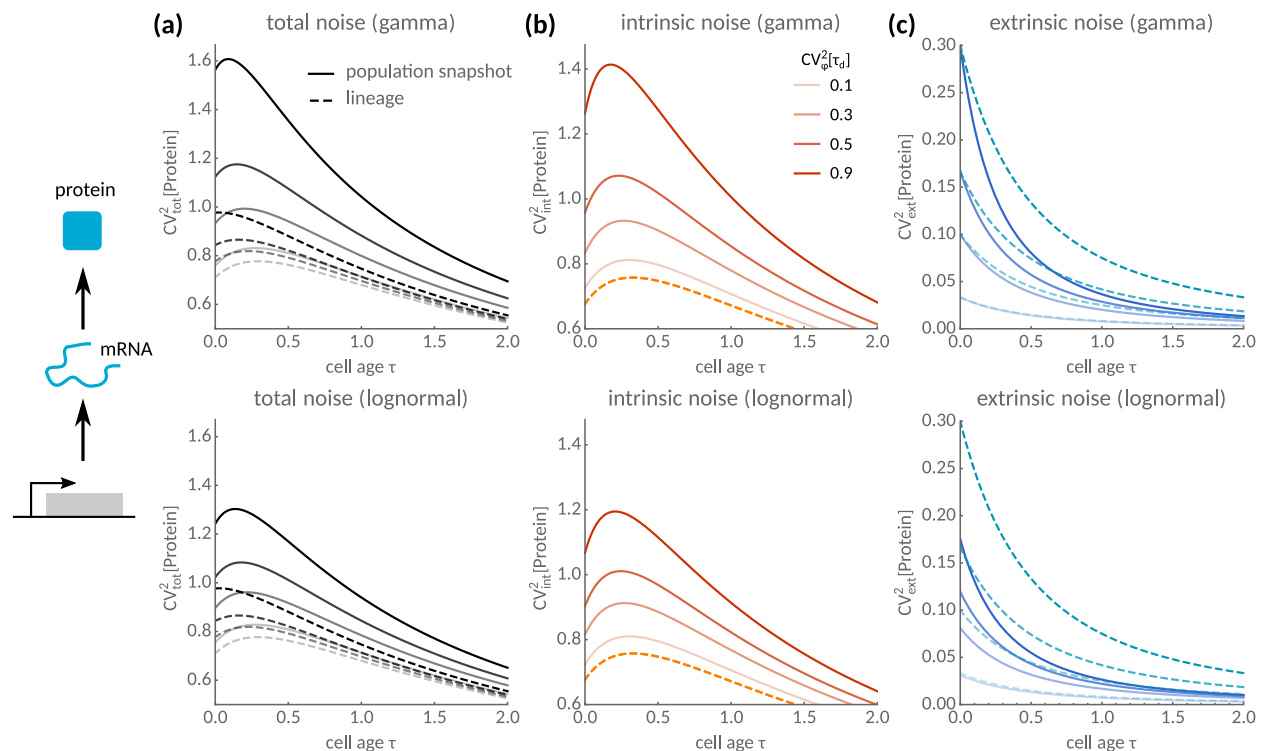
- toplasmic diffusion allows in vivo counting of proteins in individual cells,” *Nat Commun* **7**, 11641 (2016).
- <sup>31</sup>P. Thomas, “Making sense of snapshot data: ergodic principle for clonal cell populations,” *J Royal Soc Interface* **14**, 20170467 (2017).
- <sup>32</sup>E. Powell, “Growth rate and generation time of bacteria, with special reference to continuous culture,” *Microbiology* **15**, 492–511 (1956).
- <sup>33</sup>P. Painter and A. Marr, “Mathematics of microbial populations,” *Annu Rev Microbiol* **22**, 519–548 (1968).
- <sup>34</sup>Y. Wakamoto, A. Y. Grosberg, and E. Kussell, “Optimal lineage principle for age-structured populations,” *Evolution* **66**, 115–134 (2012).
- <sup>35</sup>N. G. Van Kampen, *Stochastic processes in physics and chemistry* (Elsevier, 1992).
- <sup>36</sup>J. Elf and M. Ehrenberg, “Fast evaluation of fluctuations in biochemical networks with the linear noise approximation,” *Genome Res* **13**, 2475–2484 (2003).
- <sup>37</sup>V. Shahrezaei and P. S. Swain, “Analytical distributions for stochastic gene expression,” *Proc Natl Acad Sci* **105**, 17256–17261 (2008).
- <sup>38</sup>C. G. Bowsher and P. S. Swain, “Identifying sources of variation and the flow of information in biochemical networks,” *Proc Natl Acad Sci* **109**, E1320–E1328 (2012).
- <sup>39</sup>D. Huh and J. Paulsson, “Random partitioning of molecules at cell division,” *Proc Natl Acad Sci* **108**, 15004–15009 (2011).
- <sup>40</sup>D. Thieffry, A. M. Huerta, E. Pérez-Rueda, and J. Collado-Vides, “From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*,” *Bioessays* **20**, 433–440 (1998).
- <sup>41</sup>P. Thomas and R. Grima, “Approximate probability distributions of the master equation,” *Phys Rev E* **92**, 012120 (2015).
- <sup>42</sup>N. Rosenfeld, M. B. Elowitz, and U. Alon, “Negative autoregulation speeds the response times of transcription networks,” *J Mol Biol* **323**, 785–793 (2002).
- <sup>43</sup>Y. Dublanche, K. Michalodimitrakakis, N. Kümmerer, M. Foglierini, and L. Serrano, “Noise in transcription negative feedback loops: simulation and experimental analysis,” *Mol Syst Biol* **2**, 41 (2006).
- <sup>44</sup>D. A. Oyarzun, J.-B. Lugagne, and G.-B. V. Stan, “Noise propagation in synthetic gene circuits for metabolic control,” *ACS Synth Biol* **4**, 116–125 (2014).

- <sup>45</sup>A. Singh and J. P. Hespanha, “Optimal feedback strength for noise suppression in autoregulatory gene networks,” *Biophys J* **96**, 4013–4023 (2009).
- <sup>46</sup>O. G. Berg, “A model for the statistical fluctuations of protein numbers in a microbial population,” *J Theor Biol* **71**, 587–603 (1978).
- <sup>47</sup>A. Schwabe and F. J. Bruggeman, “Contributions of cell growth and biochemical reactions to nongenetic variability of cells,” *Biophys J* **107**, 301–313 (2014).
- <sup>48</sup>F. Bertaux, S. Marguerat, and V. Shahrezaei, “Division rate, cell size and proteome allocation: impact on gene expression noise and implications for the dynamics of genetic circuits,” *Royal Soc Open Sci* **5**, 172234 (2018).
- <sup>49</sup>C. Vargas-Garcia, K. Ghusinga, and A. Singh, “Cell size control and gene expression homeostasis in single-cells,” *Curr Opin Syst Biol* **8**, 109–116 (2018).
- <sup>50</sup>T. Blasi, F. Buettner, M. K. Strasser, C. Marr, and F. J. Theis, “cgcorrect: a method to correct for confounding cell–cell variation due to cell growth in single-cell transcriptomics,” *Phys Biol* **14**, 036001 (2017).
- <sup>51</sup>J. H. Heerden, H. Kempe, A. Doerr, T. Maarleveld, N. Nordholt, and F. J. Bruggeman, “Statistics and simulation of growth of single bacterial cells: illustrations with *B. subtilis* and *E. coli*,” *Sci Rep* **7**, 16094 (2017).
- <sup>52</sup>P. Thomas, “Analysis of cell size homeostasis at the single-cell and population level,” *Frontiers in Physics* **6**, 64 (2018).
- <sup>53</sup>S. Tsuru, J. Ichinose, A. Kashiwagi, B.-W. Ying, K. Kaneko, and T. Yomo, “Noisy cell growth rate leads to fluctuating protein concentration in bacteria,” *Phys Biol* **6**, 036015 (2009).
- <sup>54</sup>R. Grima, P. Thomas, and A. V. Straube, “How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations?” *J Chem Phys* **135**, 084103 (2011).
- <sup>55</sup>D. Schnoerr, G. Sanguinetti, and R. Grima, “Comparison of different moment-closure approximations for stochastic chemical kinetics,” *J Chem Phys* **143**, 11B610.1 (2015).
- <sup>56</sup>R. Kafri, J. Levy, M. B. Ginzberg, S. Oh, G. Lahav, and M. W. Kirschner, “Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle,” *Nature* **494**, 480–483 (2013).
- <sup>57</sup>L. Keren, D. Van Dijk, S. Weingarten-Gabbay, D. Davidi, G. Jona, A. Weinberger, R. Milo, and E. Segal, “Noise in gene expression is coupled to growth rate,” *Genome Res* **25**, 1893–1902 (2015).

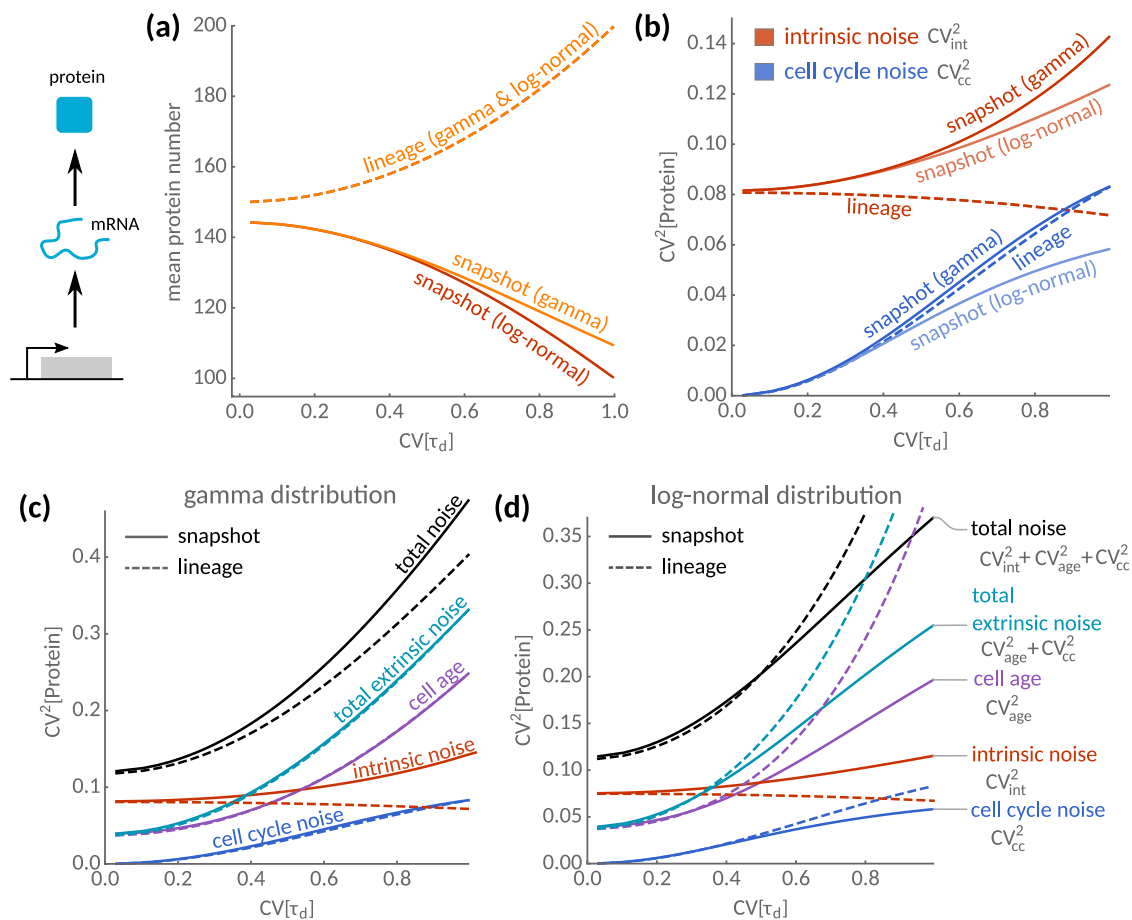
- <sup>58</sup>C. van Boxtel, J. H. van Heerden, N. Nordholt, P. Schmidt, and F. J. Bruggeman, “Taking chances and making mistakes: non-genetic phenotypic heterogeneity and its consequences for surviving in dynamic environments,” *J Royal Soc Interface* **14**, 20170141 (2017).
- <sup>59</sup>B. M. Martins and J. C. Locke, “Microbial individuality: how single-cell heterogeneity enables population level strategies,” *Curr Opin Microbiol* **24**, 104–112 (2015).
- <sup>60</sup>F. Fröhlich, P. Thomas, A. Kazeroonian, F. J. Theis, R. Grima, and J. Hasenauer, “Inference for stochastic chemical kinetics using moment equations and system size expansion,” *PLoS Comput Biol* **12**, e1005030 (2016).



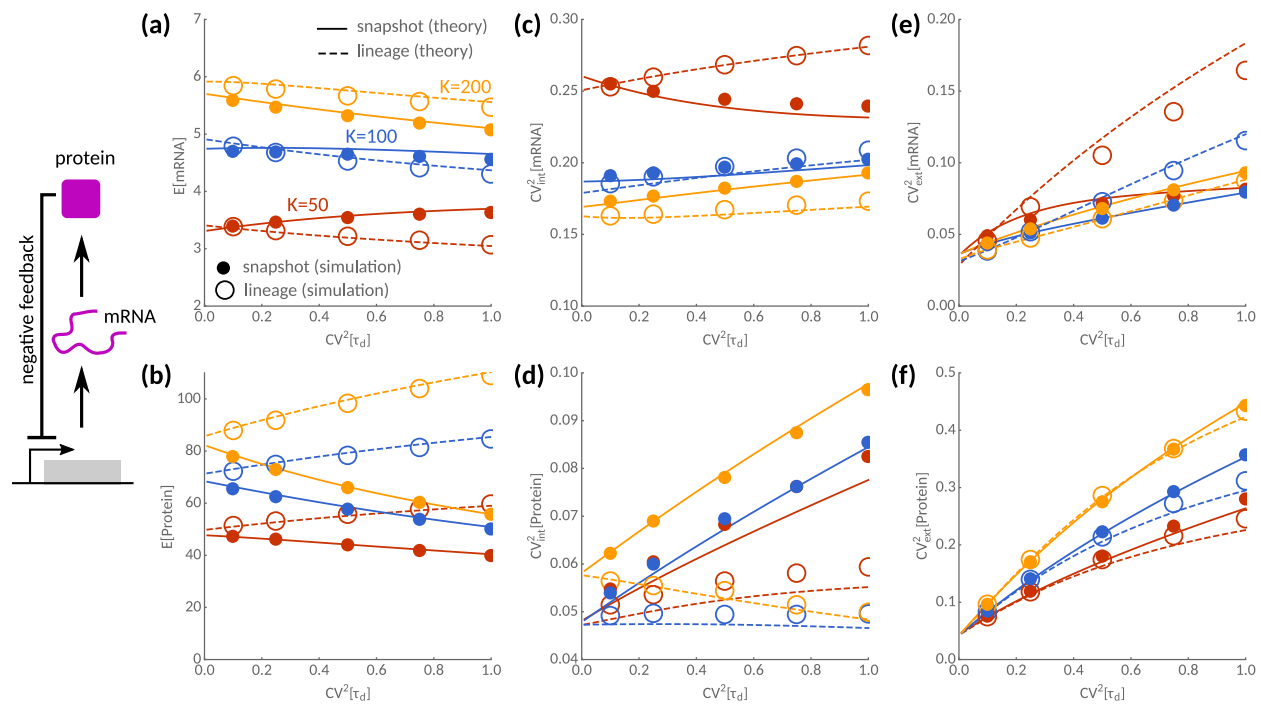
**FIG. 1. Agent-based model of clonal population dynamics with stochastic gene expression and cell cycle variability.** (a) Illustration of a growing population as a stochastic branching process with stochastic interdivision times. Each cell expresses two identical but non-interacting reporters (green and red) that are partitioned randomly at cell division. Red and green cells express more molecules of either reporter, which indicates intrinsic variability between cells. Yellow cells express similar levels of reporter molecules, but vary in their absolute amounts at different cell cycle stages, which constitutes extrinsic variability. A snapshot of the population (blue dashed box) quantifies the cell-to-cell variability across the population. A lineage (red path) quantifies variability over time and tracks an isolated cell over successive cell divisions by randomly selecting one of the daughter cells. (b) Simulated trajectories of cell age and stochastic protein expression of two identical reporters on a branched tree. Line colour indicates reporter expression in the same cell. (c) Cell age and reporter expression of an isolated cell lineage. (d) Comparison of distributions obtained from lineages and population snapshots. Simulations of the reactions (21) assume  $k_0 = 10$ ,  $k_m = 1$ ,  $k_s = 10$  for each reporter and lognormal-distributed division times with unit mean and standard deviation.



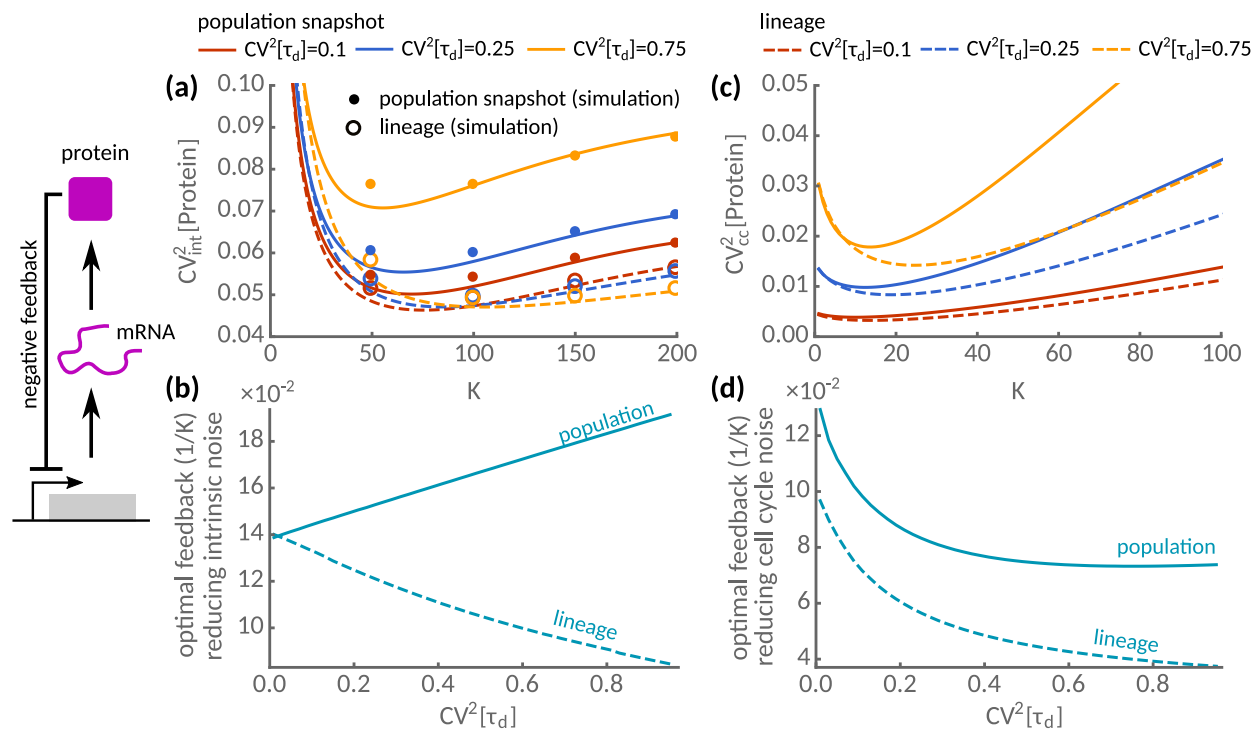
**FIG. 2. Intrinsic and extrinsic noise propagation over the cell cycle.** (a) Total noise as a function of cell age  $\tau$  with gamma (top) and log-normal-distributed (bottom) interdivision times. Population snapshot statistics (solid) are compared to lineages (dashed lines). Noise is monotonic for large cell cycle fluctuations  $CV^2[\tau_d]$  in lineages but not in snapshots. (b) Intrinsic noise peaks as a function of cell age and increases with cell cycle fluctuations in populations but not in lineages. (c) Extrinsic noise is lower in the population than in lineages. Parameters are  $k_0 = 1$ ,  $b = 100$  and division time distributions assume unit mean.



**FIG. 3. Statistics of population snapshots and isolated lineages for cells of unknown age.** (a) Mean protein number as a function of the cell cycle variations  $CV[\tau_d]$  in lineages (dashed) and snapshots (solid lines). For lineages, the mean protein number increases with cell cycle variability and is independent of the division time distribution. In snapshots, the mean decreases with cell cycle variability with a rate that depends on higher moments of the distribution. The predictions for gamma- and log-normal distributed interdivision times is shown. (b) Sensitivity of intrinsic and extrinsic noise sources to cell cycle fluctuations. Intrinsic noise (red lines) increases in lineages but decreases in snapshots consistent with the dependence of the respective means shown in (a). The transmitted cell cycle noise (blue lines) shows a similar dependence on cell-cycle variability in lineages and snapshots for the gamma-distribution, but is lower in snapshots for the log-normal distribution. (c) Total noise (black lines) broken down into individual noise components for the gamma-distribution. Transmitted cell cycle noise and the uncertainty due to distributed cell ages (purple lines) contribute to the total extrinsic noise (teal). (d) The corresponding break-down for the log-normal distribution. Parameters are  $k_0 = 10$  and  $b = 10$  and  $E_\varphi[\tau_d] = 1$ .



**FIG. 4. Noise decomposition of a negative feedback circuit.** Sensitivities to cell cycle noise  $CV^2[\tau_d]$  of mean, intrinsic and extrinsic noise contributions are shown for weak (yellow,  $K = 200$ ), moderate (blue,  $K = 100$ ) and strong feedback (red,  $K = 50$ ). Predictions by the linear noise approximation (solid lines) are in good qualitative agreement with stochastic simulations (dots). **(a)** In lineages, the mean mRNA number always decreases with cell cycle variability while this is not true in snapshots for moderate to high feedback. **(b)** In contrast, protein levels always increase in lineages but decrease in snapshots. **(c)** The corresponding intrinsic noise profiles of mRNAs typically increase with cell cycle noise except in snapshots with strong feedback. **(d)** Intrinsic noise of proteins always increases with cell cycle noise in snapshots but not in lineages. **(e,f)** Total extrinsic noise increases with cell cycle variability for mRNAs and proteins. However, strong feedback may significantly reduce extrinsic noise in snapshots compared to lineages. Deviations between the approximation (lines) and the simulations (dots) are most pronounced for strong feedback. Parameters are  $k_0 = 10$ ,  $k_m = 1$ ,  $k_s = 10$  and unit-mean interdivision times.



**FIG. 5. Feedback strategies for noise suppression in lineages and populations.** Intrinsic and extrinsic noise statistics of negative autoregulatory feedback circuit are shown as a function of  $K$ , the inverse feedback strength, for three different levels of cell cycle noise  $CV^2[\tau_d] = 0.1$  (red), 0.25 (blue) and 0.75 (yellow). **(a)** Intrinsic noise exhibits a minimum as a function of the repression strength both in lineage (dashed) and in snapshot statistics (solid lines). The predictions obtained using the linear noise approximation (lines) are in good agreement with exact stochastic simulations using the First-Division Algorithm<sup>31</sup> (dots for population, open circles for lineages). **(b)** Optimal feedback strength ( $1/K$ ) to minimise intrinsic noise is shown. The feedback strength increases with interdivision time noise in lineages but decreases in population snapshots. **(c)** The transmitted cell cycle noise shows a minimum in dependence of the repression strength both in lineage (dashed) and in snapshot statistics (solid lines). **(d)** The optimal feedback strength to minimise transmitted cell cycle noise decreases with division time noise both in lineages and to a lesser extent in the population. Parameters as in Fig. 4.