

SCIENTIFIC REPORTS



OPEN

Intrinsic and extrinsic noise of gene expression in lineage trees

Philipp Thomas 

Cell-to-cell heterogeneity is driven by stochasticity in intracellular reactions and the population dynamics. While these sources are usually studied separately, we develop an agent-based framework that accounts for both factors while tracking every single cell of a growing population. Apart from the common intrinsic variability, the framework also predicts extrinsic noise without the need to introduce fluctuating rate constants. Instead, extrinsic fluctuations are explained by cell cycle fluctuations and differences in cell age. We provide explicit formulas to quantify mean molecule numbers, intrinsic and extrinsic noise statistics in two-colour experiments. We find that these statistics 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) population snapshots with known cell ages as recorded in time-lapse microscopy, and (iii) snapshots with unknown cell ages as measured from static images or flow cytometry. Applying the method to models of stochastic gene expression and feedback regulation elucidates 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.

Cellular behaviour varies substantially from cell to cell and over time^{1–3}. Identifying the sources of these fluctuations can help reveal the function of gene circuits and signalling networks and understand how clonal cells diversify their responses. When quantifying these fluctuations in gene regulatory networks, 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 realised in bacteria⁴, yeast⁴, mammalian⁵ and plant cells⁶. The difference between the two circuit copies reflect 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^{7–16}.

A less commonly studied but equally important source of extrinsic noise is the population dynamics¹⁷. 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*¹⁸, budding yeast¹⁹ and mammalian cells²⁰ vary about 30–50%. These sources should therefore prevail in growing cells, populations and tissues.

Modelling approaches for understanding the cell cycle effects on gene expression noise are only recently being developed^{21–27}. 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 machine²⁸. Many experiments, however, report cell-to-cell variability across snapshots of an exponentially growing cell population. These approaches either use time-lapse microscopy^{29,30} or analyse snapshots with distributed cell cycle stages as in flow cytometry, smFISH or similar techniques^{31–33}.

Recent studies elucidated that population snapshots and lineages can significantly deviate in their statistics^{27,34}. 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

Department of Mathematics, Imperial College London, London, SW7 2AZ, UK. Correspondence and requests for materials should be addressed to P.T. (email: p.thomas@imperial.ac.uk)

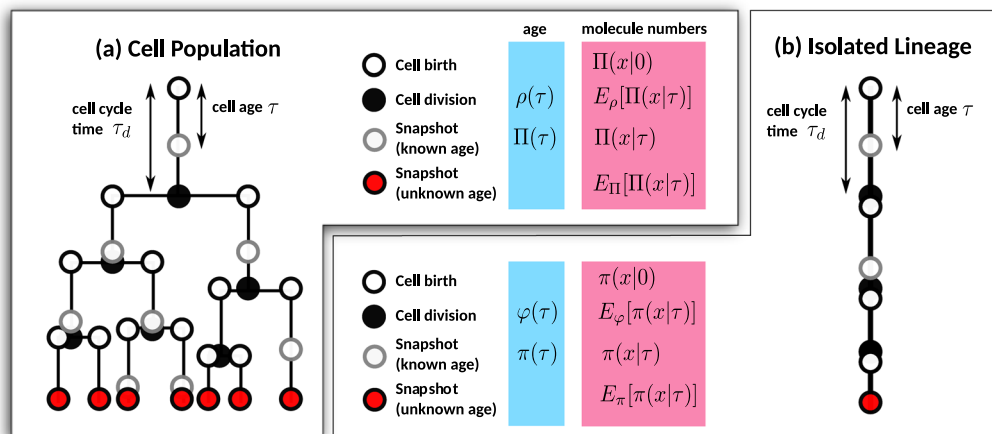


Figure 1. Single-cell statistics of cell populations and isolated lineages. **(a)** Distributions of the molecule numbers x measured in a cell of age τ across a lineage tree from time-lapse movies. The population dynamics is described by the age distribution $\Pi(\tau)$ and the distribution of interdivision times $\rho(\tau)$. Gene expression dynamics is described by the distribution $\Pi(x|\tau)$ for cells of age τ (open grey circles). Accordingly, the distributions are $\Pi(x|0)$ at cell birth (open black circles) and $E_\rho[\Pi(x|\tau_d)]$ at cell division (filled black circles). The final state of the population (red dots) describes a snapshot with distributed cell ages $E_\Pi[\Pi(x|\tau)]$ as obtained from static images or flow cytometry. **(b)** The corresponding statistics for a lineage following an isolated cell over time as observed in the mother machine.

measurements 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 analyse stochastic biochemical reaction networks in a growing and dividing cell population. In this framework, an agent is represented by a cell whose biochemical composition changes due to stochastic reaction kinetics and cell divisions. A distinguishing feature of this theory is that it takes into account the cell age – the time since the last division – which is accessible from time-lapse microscopy or cell cycle markers. We show how to analytically characterise the joint distribution of cell age and molecule content per cell in a lineage tree. We then derive the exact moment equations of this model.

Since stochastic models are rarely solvable, we propose an analytically tractable approximation to mean and covariances. 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, illustrate the results using a simple two-reporter system, and study how circuit dynamics can be tuned to suppress either intrinsic or extrinsic fluctuations.

Methods

There exist a variety of ways of how to pool data from different cells in lineage trees. In Fig. 1a, we summarise the different distributions that characterise tree-structured data. Time-lapse data resolves the age of each cell and hence enables measuring the statistics of interdivision times and molecule numbers at cell birth (open black circles), division (filled black circles) or at a specified position in the cell cycle (open grey circles). If the cell age is unknown, as for static images or flow cytometry data, only a snapshot of the present population is available (red dots). The corresponding measures for an isolated lineage of a single cell tracked over time, as in the mother machine, are summarised in Fig. 1b.

We model such data by investigating 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 are tracked from birth to division. After cell division, the two daughters have zero age and inherit different proportions of the mother's molecules through stochastic partitioning of molecules. Cells divide at random times, so divisions occur asynchronously, which results in distributed cell ages and molecule numbers across the population. Figure 2a illustrates the resulting branching process whose final state represents a snapshot of the population.

Elowitz *et al.* identified the sources of cell-to-cell variation using snapshots of cells expressing green and red fluorescent reporters¹. Reporters expressed at different levels in the cell appear either red or green, a signature of intrinsic noise. Cells with similar reporter levels light up in various shades of yellow, a signature of extrinsic noise. In our model (Fig. 2a), intrinsic variation results from stochasticity in biochemical reactions and partitioning of molecules at division whereas cell age and variability in division timing provide a source of extrinsic noise (Fig. 2b). An isolated cell tracked over time (Fig. 2c) displays different levels of variability than a typical cell in the population (Fig. 2d). We begin with deriving an analytical framework to compare these different distributions.

Agent-based framework for stochastic biochemical kinetics in growing cell populations. To each cell we associate an age τ that counts the time since its last division and a set of biochemical species $X_1, X_2,$

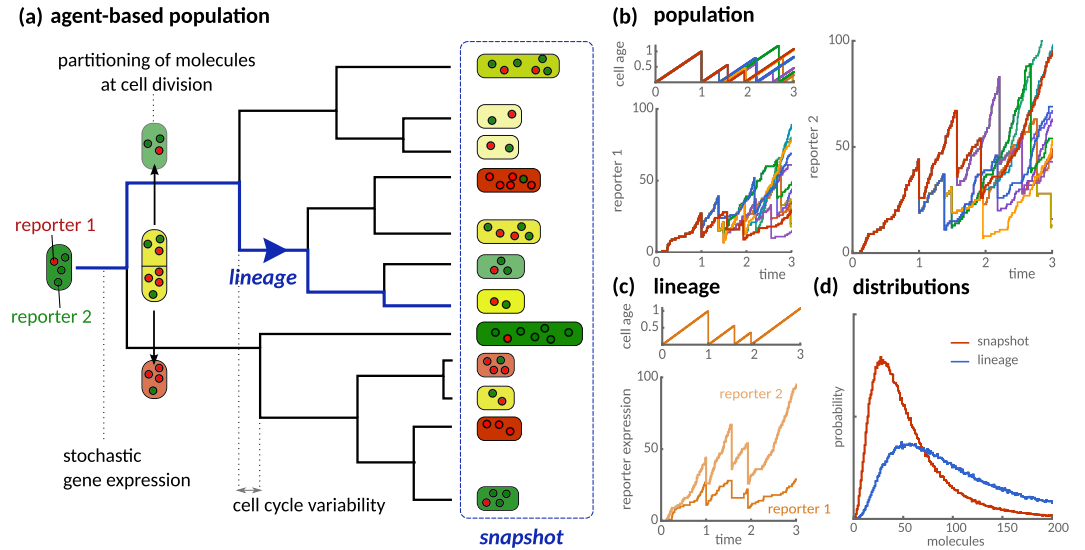
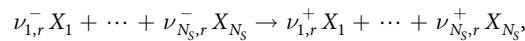


Figure 2. Agent-based model of clonal population dynamics with stochastic gene expression and cell cycle variability. **(a)** Illustration of a growing population as a branching process with stochastic interdivision times. Each cell expresses two identical but non-interacting reporters (green and red) 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 their levels vary over the cell cycle resulting in extrinsic variability. A snapshot of the population (blue dashed box) quantifies the cell-to-cell variability across the population. A lineage (blue 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 simulated distributions of lineages and population snapshots. Exemplary simulations of the reactions (29) assume $k_0 = 10$, $k_m = 1$, $k_s = 10$ for each reporter and lognormal-distributed interdivision times with unit mean and standard deviation.

..., X_{N_S} with molecule numbers $x = (x_1, x_2, \dots, x_{N_S})^T$. These species interact via a network of R intracellular biochemical reactions of the type



where ν_{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.

Master equation for the agent-based population. The state of a cell population is described by the snapshot density $n(\tau, x, t)$ counting the number of cells at time t with age between τ and $\tau + d\tau$ and molecule counts x . Accordingly, the total number of cells in the population is given by

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

where the summation is over all possible molecule number configurations.

We assume that cells divide with an age-dependent rate $\gamma(\tau)$, which can be written in terms of the interdivision time distribution $\varphi(\tau_d)$ via

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

The snapshot density evolves due to age-progression of cells, cell divisions and the change in their molecular numbers in biochemical reactions (see SI Sec. S1 for a detailed derivation)

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

Here, the change in the molecule numbers per cell is expressed by the transition matrix \mathbb{Q} acting as

$$\mathbb{Q}n(\tau, x, t) = \sum_{r=1}^R [w_r(x - \nu_r) n(\tau, x - \nu_r, t) - w_r(x) n(\tau, x, t)], \tag{3b}$$

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

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

whereby the mother cell is replaced with two daughter cells of zero age and 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 x' molecules in the mother cell (see SI Sec. 1). Importantly, since the total amount is conserved in the division, we have $B_2(x|x') = B_1(x' - x|x')$, which implies that, on average, cells inherit equal amounts of molecules $E_B[x|x'] = x'/2$.

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 $N(t) \sim e^{\lambda t}$ with rate λ 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 describe this distribution analytically.

Age-distribution and population growth rate. The fraction of cells with the same age in a snapshot is given by the age-distribution, which follows³⁴

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

The distribution φ 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 λ 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 dividing cell populations, which has been verified in experiments^{18,35,36}.

Distribution of molecules for cells of the same age. We consider the total number of cells with age τ 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 τ 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 verified³⁴ 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 a boundary condition that accounts for the cell divisions

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

The condition states that the distribution at cell birth is the distribution at cell division after partitioning (cf. Fig. 1a). The distribution under the integral sign

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

is the interdivision time distribution in the population^{36,37}. The distribution describes the interdivision times of cells with completed cell divisions and depends explicitly on the population growth rate λ .

Comparison with the lineage framework. A lineage tracks a single cell of the population over successive cell divisions (cf. Fig. 1a). In the long-term, it approaches a stable distribution denoted by $\pi(\tau, x) = \pi(\tau)\pi(x|\tau)$. The molecule number distribution for cells of the same age is then given by $\pi(x|\tau)$ and satisfies

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

The age-distribution $\pi(\tau)$ is given by

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

Analogous expressions to Eq. (12) have been given in^{21,25,38,39}. Our approach accounts for cell cycle variability through the interdivision time distribution $\varphi(\tau_d)$ explicitly. Comparing the above equations with Eqs (6) and (10), we see that the lineage distributions are obtained from the population distributions by substituting the interdivision time distribution $\rho(\tau_d)$ for $\varphi(\tau_d)$ and the age-distribution $\Pi(\tau)$ for $\pi(\tau)$. The intuition behind this result is that cells in the population and lineage follow the same reaction dynamics but experience different levels of cell cycle fluctuations. In particular, comparing Eq. (6) with (13) and Eq. (7) with (11), we observe that slow dividing cells are exponentially suppressed in the population. Thus the stochastic dynamics of biochemical reactions can be analysed within a common framework whether in populations or lineages. Analytical expressions for special cases of the different age and interdivision distributions are derived in SI Sec. S2.

Moment equations for the agent-based model. Solving for the full distribution is infeasible in many practical situations. Summary statistics such as mean molecule numbers and their variances present convenient alternatives as they are more amenable to analysis and we will focus on these quantities in the following.

Exact moment equations for cells of the same age. 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

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

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

Interestingly, these equations 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

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

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

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).

Exact moment equations for cells of unknown age. When the cell age is not explicitly known, age-averaged moments have to be considered. For this purpose, we consider the derivative of a function $f(\tau)$ and average it with respect to the age distribution $\Pi(\tau)$. Using Eq. (6), we compute the identity

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

where we integrated by parts assuming $\lim_{\tau \rightarrow \infty} f(\tau)\Pi(\tau) = 0$ and performed the derivative. The first term describes the effect of dilution due to population growth, while the second term is due to cell division.

The equation for the first moment, $E_\Pi[x] = E_\Pi[E_\Pi[x|\tau]]$, is obtained by setting $f(\tau) = E_\Pi[x|\tau]$ in (15), combining the result with Eq. (14a) and using the boundary condition (14c). The mean number of molecules in the population satisfies,

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

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

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

where the left hand side depends explicitly on the division-time distribution ρ . These equations cannot generally be solved because the hierarchy of moments is not closed when the propensities depend non-linearly on the molecule numbers. Another complication is that they depend on the moments for cells of known age, which have to be solved simultaneously. The exception is the case of a constant age-independent division rate for which cell ages and interdivision times are exponentially distributed and divisions are memoryless (see SI Sec. S3 for further details). In the Results, we present a tractable approximation to these equations.

Results

To circumvent the moment-closure problem, we employ the linear noise approximation to decompose the noise into intrinsic and extrinsic components. In brief, the approximation assumes Gaussian fluctuations and provides closed-form expressions for the mean molecule numbers $E_{\Pi}[x|\tau]$ and the covariance matrix $\text{Cov}_{\Pi}[x|\tau] = E_{\Pi}[xx^T|\tau] - E_{\Pi}[x|\tau]E_{\Pi}[x^T|\tau]$ (see^{40,41} for details), which leads to the moment equations

$$\frac{\partial}{\partial \tau} E_{\Pi}[x|\tau] = \sum_{r=1}^R \nu_r w_r(E_{\Pi}[x|\tau]), \quad (18a)$$

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

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

$$\mathcal{J}_{ij}(\tau) = \sum_{r=1}^R \nu_r \frac{\partial w_r(E_{\Pi}[x|\tau])}{\partial E_{\Pi}[x_j|\tau]}, \quad \mathcal{D}_{ij}(\tau) = \sum_{r=1}^R \nu_r \nu_{jr} w_r(E_{\Pi}[x|\tau]). \quad (19)$$

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

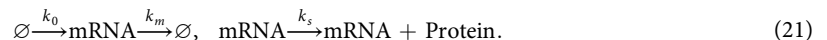
Next, we cast the boundary condition (14d) in terms of the covariance matrix $\text{Cov}_{\Pi}[x|\tau]$, which leads to

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

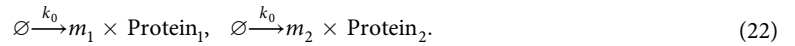
$$\begin{aligned} \underbrace{\text{Cov}_{\Pi}[x_i, x_j|0]}_{\text{variability between daughters}} &= \underbrace{\frac{1}{4} E_{\rho}[\text{Cov}_{\Pi}(x_i, x_j|\tau)]}_{\text{variability between mother cells}} \\ &+ \underbrace{E_{\rho}[E_{\Pi}(\text{Cov}_B[x_i, x_j|x']|\tau)]}_{\text{partitioning of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

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 term stems from differences in the molecule numbers due to different cell cycle lengths. Eq. (20b) does not constitute a noise decomposition itself because the variability between mothers depends implicitly on the variability between daughters. Instead, the second and third terms constitute sources of intrinsic and extrinsic cell-to-cell variability as we will show in the following.

Decomposing noise into intrinsic and extrinsic contributions. 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 network 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.⁴² and SI Sec. S4 for details of the burst approximation).

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 τ becomes

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

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

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

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

Noise decomposition for two-reporter systems. For identical two-reporter systems, the total variance can be decomposed into intrinsic and extrinsic components

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

The individual contributions can be quantified using the two reporters¹

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

Since these components are measured in the same cell, these expressions also account for the correct history dependence^{8,10,43}.

The variance of intrinsic and extrinsic fluctuations follows from using Eq. (25) in (18b) with $\mathcal{J} = 0$ and $\mathcal{D} = k_0 \sum_{m=1}^{\infty} m^2 p(m)$, where $p(m)$ is the geometric distribution (see SI Sec. S4 for details). Rearranging then leads to

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

Integrating these equations we find $\Sigma_{\text{int}}(\tau) = \Sigma_{\text{int}}(0) + b(2b + 1)k_0\tau$ and $\Sigma_{\text{ext}}(\tau) = \Sigma_{\text{ext}}(0)$. To match the boundary condition (20b), we assume that each molecule of the mother cell is 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

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

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

The coefficients of variations quantify the size of fluctuations relative to the mean. The result confirms the intuition that intrinsic noise, Eq. (26a), decreases with the mean number of molecules. The extrinsic noise component, however, reflects the variations in cell cycle duration CV_{ρ}^2 that are transmitted to the protein levels, which are independent of the mean.

Snapshots display higher intrinsic but lower extrinsic noise levels than lineages. Next, we compare the statistics of snapshots with the lineage of an isolated cell over time. As explained in the Methods, we obtain the lineage statistic by substituting the interdivision time distribution ρ for φ 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 interdivision time is longer when averaged over single cells than for cells in the population¹⁸ $E_{\rho}[\tau] \leq E_{\varphi}[\tau]$. 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 interdivision time distribution is.

In Fig. 3a, 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

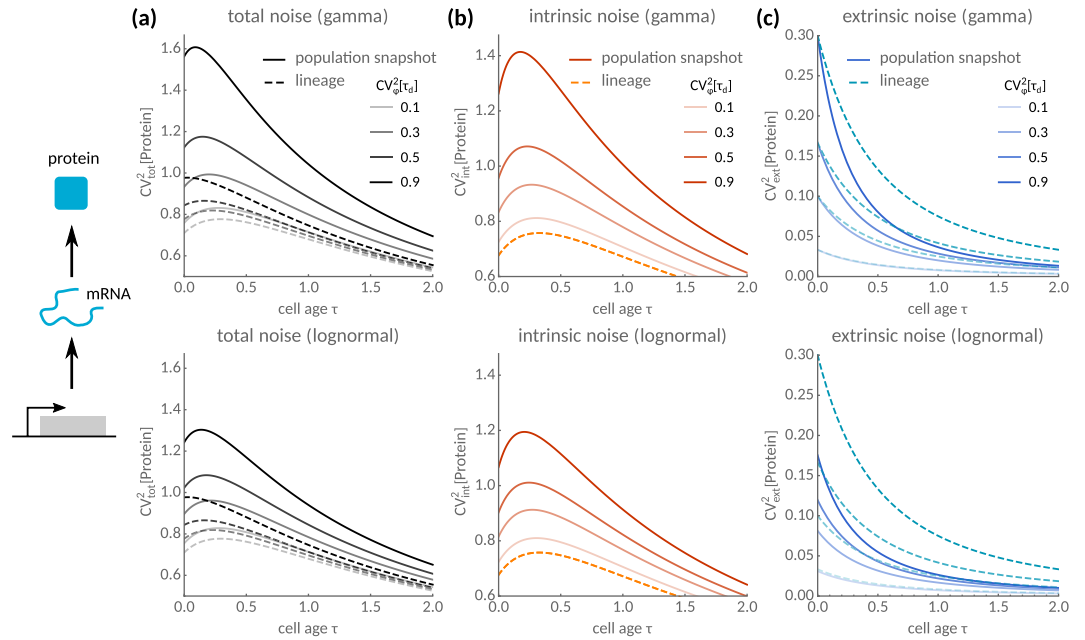


Figure 3. Intrinsic and extrinsic noise propagation over the cell cycle. **(a)** Total noise as a function of cell age τ with gamma (top) and log-normal-distributed (bottom) interdivision times. Population snapshot statistics (solid) are compared to lineages (dashed lines). **(b)** Intrinsic noise peaks 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 interdivision time distributions assume unit mean.

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. 3b) while it is significantly lower in lineages and independent of these fluctuations, which is consistent with lower expression levels in snapshots.

Figure 3b 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 (round brackets in Eq. (26a)). 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 interdivision 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. 3c). 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.

General decomposition for cells of the same age. Our aim is to 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 the subnetworks have the same mean expression levels and covariances. The covariance matrix of the combined 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 sub-matrices satisfy

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

according to Eq. (18b).

The intrinsic and extrinsic noise components can be calculated as

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

Since the overall covariance matrix obeys the linear equations (27), the two noise contributions evolve independently. In particular, the intrinsic and extrinsic covariance matrices satisfy

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

$$\frac{d}{d\tau}\Sigma_{\text{ext}}(\tau) = \mathcal{J}(\tau)\Sigma_{\text{ext}} + \Sigma_{\text{ext}}\mathcal{J}^T(\tau). \quad (28b)$$

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

For binomial partitioning, the covariance matrix is given by $\text{Cov}_B[x_1, x_2|x'] = \text{diag}(x')/4$, where $\text{diag}(x')$ is the diagonal matrix with entries x' on the diagonal. This dependence allows us to split the boundary condition (20b) according to Eq. (24), which results in two independent conditions

$$4\Sigma_{\text{int}}(0) = \underbrace{E_\rho[\Sigma_{\text{int}}(\tau_d)]}_{\text{intrinsic noise in mother}} + \underbrace{E_\rho[E_\Pi[\text{diag}(x)|\tau_d]]}_{\text{partitioning noise}}, \quad (28c)$$

$$4\Sigma_{\text{ext}}(0) = \underbrace{E_\rho[\Sigma_{\text{ext}}(\tau_d)]}_{\text{extrinsic noise in mother}} + \underbrace{\text{Cov}_\rho[E_\Pi[x|\tau_d], E_\Pi[x|\tau_d]]}_{\text{cell cycle noise}}. \quad (28d)$$

The partitioning noise features in Eq. (28c) as a source of intrinsic noise since molecules are partitioned independently between the daughter cells. The cell cycle noise represents a source of extrinsic noise in Eq. (28d) since the division timing of the mother cells affects both daughter cells in the same way. The noise decomposition is fully specified by the mean number of molecules for cells of the same age $E_\Pi[x|\tau]$, the Jacobian \mathcal{J} of the corresponding rate equations, the diffusion matrix \mathcal{D} (see Eq. (19)) and the distribution of interdivision times in the population ρ (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 incorporated into the framework by using a different form of $\text{Cov}_B[x_i, x_j|x']$ ⁴⁴.

General decomposition for cells of unknown age. An obstacle for applying this decomposition in practice is that the cell age is often unknown, 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

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

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.

Practical computation of the noise decomposition and applications. Finally, we apply the noise decomposition to analyse snapshots with unknown cell ages. While the decomposition can be carried out exactly for linear reaction networks, we also outline an efficient numerical method for complex nonlinear networks and demonstrate its efficacy for a protein that regulates its own expression.

Decomposition for linear reaction networks in cells of unknown age. Linear reaction networks have propensities functions that are linear functions of the molecule numbers. This dependence implies $E_\Pi[w_r(x)] = w_r(E_\Pi[x])$ and allows averaging the statistics exactly over all cell ages. Thus from Eq. (16), we obtain the rate equations

$$\lambda E_\Pi[x] = \sum_{r=1}^R \nu_r w_r(E_\Pi[x]), \quad (30)$$

which coincide with the steady state of the traditional deterministic rate equations including an effective dilution term proportional to the population growth rate λ .

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

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

Here, the Jacobian \mathcal{J} is independent of cell age because the propensities are linear. Similarly, averaging Eq. (28b) the extrinsic variance transmitted from cell cycle fluctuations is obtained as

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

An equation for $\bar{\Sigma}_{\text{age}}$ can also be derived (SI Sec. S5), which reads

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

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

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

$$E[p] = \begin{cases} k_0 b / \lambda & \text{snapshot,} \\ bk_0(E_{\varphi}[\tau_d] + E_{\pi}[\tau]) & \text{lineage.} \end{cases} \quad (32)$$

Note that the lineage mean follows from integrating Eq. (23) with φ instead of ρ against the lineage age-distribution (13). Both averages depend implicitly on the cell cycle variability through the average age $E_{\pi}[\tau]$ or the population growth rate λ , respectively (see SI Sec. 2). Interestingly, the population mean is limited by the growth rate but the lineage mean depends both on age and interdivision time. In Fig. 4a, we show that, due to this dependence, molecule numbers increase with cell cycle variability in lineages while they decrease when using snapshot statistics 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

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

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

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

and the one due to the unknown cell age is

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

The noise decomposition crucially depends on the population growth rate λ , while in lineages it depends on the corresponding average cell age $E_{\pi}[\tau]$. More specifically, CV_{cc}^2 and CV_{age}^2 depend on the variations in the age and interdivision time distributions, which are generally different in lineages and population (see SI Sec. S2 for details). 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. 4b) 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. 4a) 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. 4b) 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. 4c), while this not true for the log-normal distribution and large cell cycle variability (black lines, Fig. 4d).

In developed network models, the extrinsic components will also depend on the biochemical properties of the network. We demonstrate this analytically in SI Sec. S5 when the protein is also subject to degradation, which reveals more intricate noise patterns. A straight-forward approach for the noise decomposition in nonlinear biochemical networks is given in the following section.

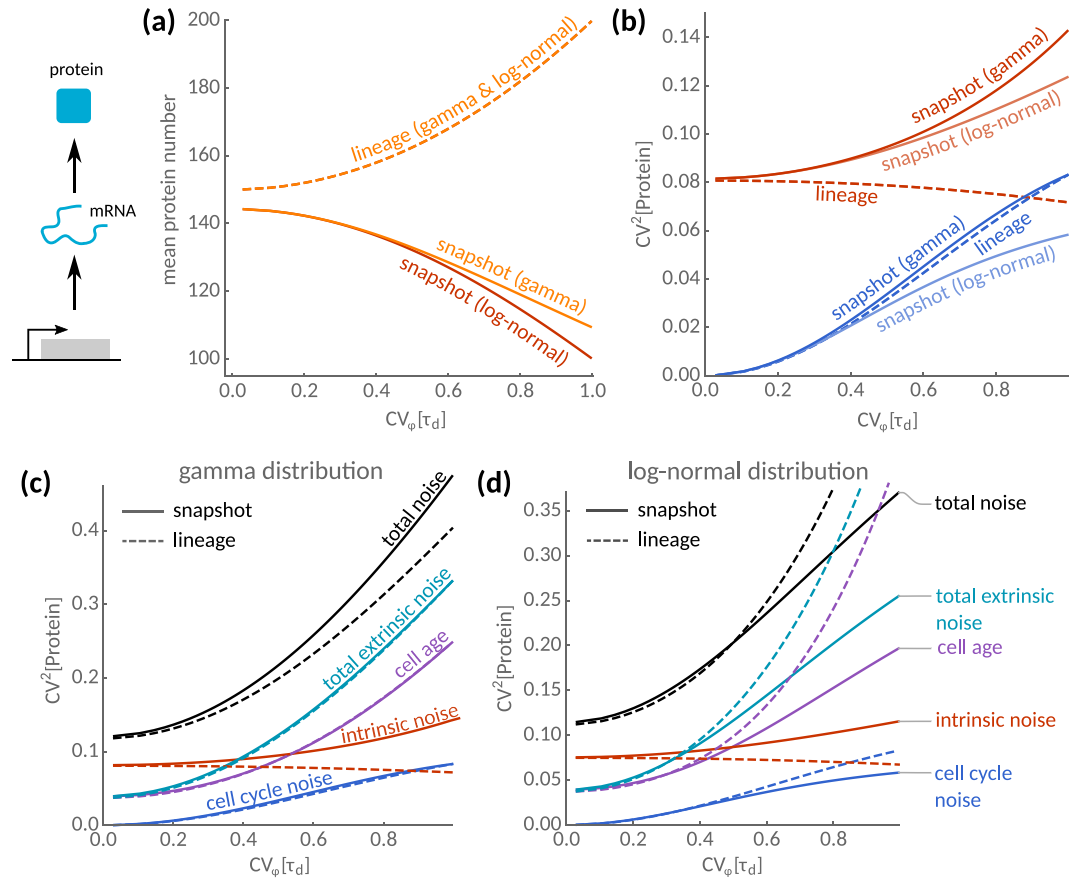


Figure 4. Statistics of population snapshots and isolated lineages for cells of unknown age. **(a)** Mean protein number as a function of the cell cycle variations $CV_\phi[\tau_d]$ in lineages (dashed) and snapshots (solid lines). For lineages, the mean protein number increases with cell cycle variability and is independent of the interdivision 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 are 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 decomposition for the log-normal distribution. Parameters are $k_0 = 10$ and $b = 10$ and $E_\phi[\tau_d] = 1$.

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 algorithm to compute the noise decomposition numerically is the following:

1. Calculate the population growth rate using Eq. (8).
2. Solve for the statistics of cells of the same age, Eqs (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 averages in Eq. (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 using an ODE solver with an initial guess on their values at cell birth, $E_\Pi[x|0]$, $\Sigma_{\text{int}}(0)$ and $\Sigma_{\text{ext}}(0)$, and refining the result 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 large parameter sweeps required for statistical inference.

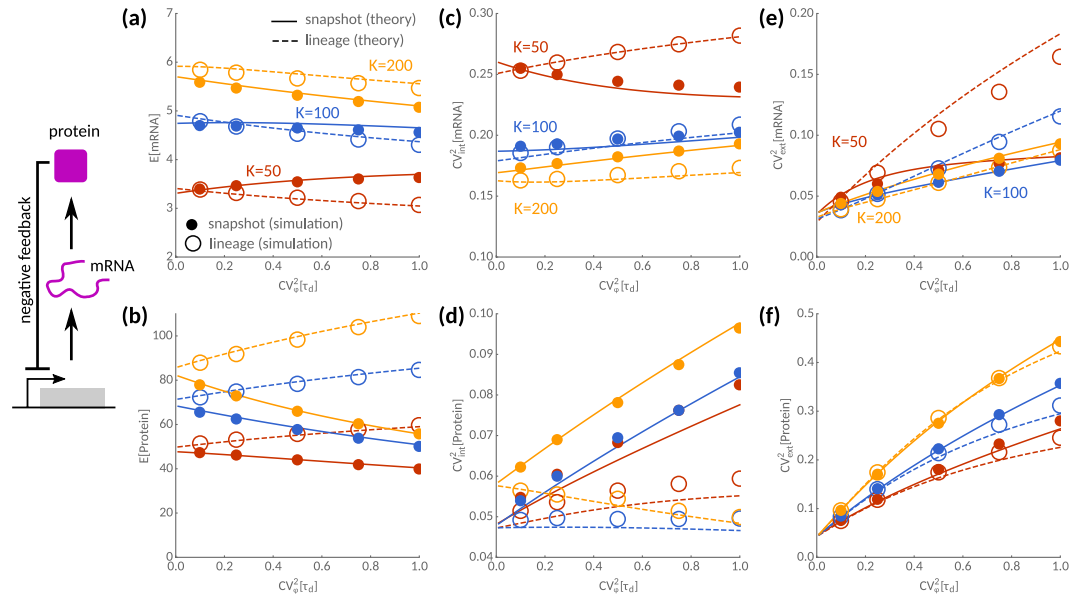
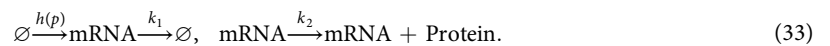


Figure 5. Noise decomposition of a negative feedback circuit. Sensitivities to cell cycle noise $CV_{\varphi}^2[\tau_d]$ of mean, intrinsic and extrinsic noise contributions are shown for weak (yellow, $K=200$), moderate (blue, $K=100$) and strong feedback (red, $K=50$). Predictions by the linear noise 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 log-normal distributed interdivision times with unit-mean.

Suppressing intrinsic or extrinsic noise through feedback mechanisms. Over 40% of known transcription factors in *E. coli* regulate their own expression⁴⁵. 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 $h(p) = \frac{k_0}{1 + (p/K)^4}$, an approximation that implicitly assumes fast promoter-binding⁴⁶. The transcription rate decreases with the protein number p depending on the effective dissociation constant K .

In Fig. 5a we show that, in lineages, mRNA levels 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 when using snapshot statistics for various feedback strengths (Fig. 5b). In agreement with this trend, we find that intrinsic noise in protein expression increases with cell cycle variability in snapshots while the opposite behaviour is observed in lineages for weak to moderate feedback (Fig. 5d). Strikingly, due to the negative feedback regulation, the sensitivity of intrinsic noise of mRNAs and proteins follows an inverse relationship (Fig. 5c), but the differences between lineage and snapshot statistics are most pronounced for weak feedback. 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. 5d,e). In all cases, our approximations are in good agreement with exact stochastic simulations of two non-interacting reporter networks (Fig. 5 dots) carried out using the First-Division Algorithm³⁴.

Finally, we use the noise decomposition to understand how heterogeneity can be controlled in natural and synthetic circuits by sweeping the feedback strength parameter ($1/K$). Negative feedback is widely known to reduce noise but often requires fine-tuned parameters^{47,48}. How this translates to individual functional noise components, such as intrinsic and extrinsic noise, has only been explored in response to fluctuations in parameters and shared components^{7,16,49} but not in the context of cell division dynamics. Here, we are specifically interested in the sensitivity of lineage and population snapshot statistics to cell cycle noise.

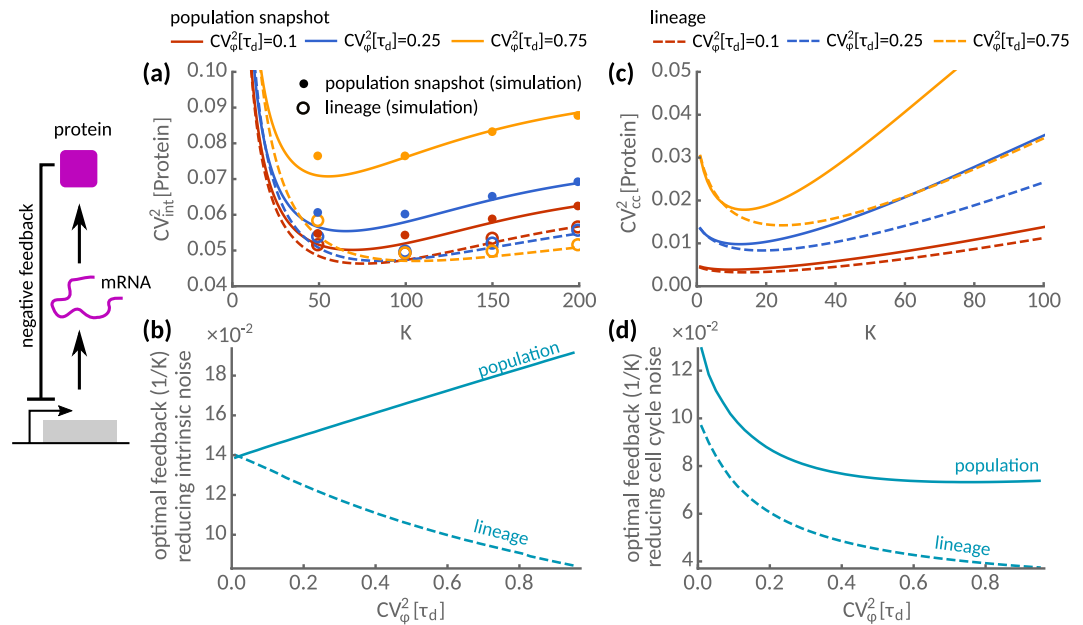


Figure 6. Feedback strategies for noise suppression in lineages and populations. Intrinsic and extrinsic noise statistics of negative autoregulatory feedback circuit are shown as a function of K , the inverse feedback strength, for three different levels of cell cycle noise $CV^2_{\phi}[\tau_d] = 0.1$ (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 Algorithm³⁴ (dots for population, open circles for lineages). **(b)** Optimal feedback strength ($1/K$) to minimise intrinsic noise is shown. The feedback strength increases with cell cycle 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 interdivision time noise both in lineages and to a lesser extent in the population. Parameters as in Fig. 5.

In Fig. 6a, 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. 6b. 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. Tuning the dissociation constant (K) can reduce the transmitted cell cycle noise CV^2_{cc} (Fig. 6c). Comparing the optimal feedback strength ($1/K$) as a function of the cell cycle noise $CV^2_{\phi}[\tau_d]$ (Fig. 6d), we observe that it increases with cell cycle variability in both lineages and snapshots. Vastly different feedback strengths, however, achieve either intrinsic or extrinsic noise suppression (cf. Fig. 6b,d). In population snapshots, we need to increase the feedback strength to decrease noise, while in lineages we need to decrease it. These findings highlight that a single feedback loop is not sufficient to suppress several functional noise components in lineages and population snapshots simultaneously.

Discussion

We present an analytical framework to quantify stochastic biochemical reactions in an exponentially growing cell population. This theory allows us to characterise and systematically decompose noise in general biochemical reaction networks into intrinsic and extrinsic components. We found that a typical cell in the population expresses lower protein levels per cell than an isolated cell tracked over successive cell divisions. The intuition behind this effect is that fast-growing cells are overrepresented in populations, where cells compete for growth. Since these cells accumulate less proteins over the cell cycle, their overall expression decreases. 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 interdivision time variability as it is expected in natural populations. Importantly, this highlights that one needs to account for cell cycle fluctuations when modelling intrinsic or extrinsic noise.

Previous studies^{38,39} focussed on the effect of age-structure but mostly neglected cell cycle variations. If these variations are taken into account the statistics of lineages and population snapshots are not equivalent even when the cell cycle position is known. This implies that measuring cells within narrow stages of the cell cycle as achieved

through gating^{9,31}, for instance, 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^{7,8,14}. 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^{50,51}. Incorporating additional physiological details such as cell size into our framework could thus provide insights to the statistics of intracellular concentrations^{52–54} and extrinsic noise transduced from cell size control and growth rate fluctuations⁵⁵. 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^{46,56}, or close the hierarchy of moments using moment closure approximations⁵⁷.

Statistics obtained from snapshots are often used to say something about a cell's history. By grouping cells of similar ages, as in ergodic rate analysis⁵⁸, one can in principle reconstruct time-course information. We demonstrated that such a procedure produces different results to the lineage statistic (see Fig. 3). Instead, the variability across the population is equivalent to choosing an arbitrary cell from the final population and tracing it backwards in time³⁴. This equivalence provides a sample-path interpretation of snapshot data, but 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 here.

We showed that gene expression noise is coupled to the population growth rate and cell cycle variability. With increasing cell cycle variability, fast-growing cells become increasingly overrepresented in population snapshots because these cells divide faster. In contrast, slow-growing cells become increasingly frequent in isolated lineages because cells are tracked over time. Since faster growing cells produce less molecules, mean molecule levels decrease with cell cycle variability in populations but increase in lineages. Taking into account this dependency implies that cellular heterogeneity displays opposite sensitivities to cell cycle variability across populations and lineages. Thus, reducing intrinsic noise in lineages over time thus comes at the cost of increasing population heterogeneity.

Cells may exploit negative feedback circuits to control this variability. Our findings imply, however, that cells cannot minimise noise of lineages and populations simultaneously (Fig. 6). Since minimising noise over time increases population heterogeneity, cells may use this strategy to diversify in response to stress⁵⁹. Conversely, tuning population 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^{3,60}. Identifying the relevant noise components and cellular objectives will likely depend on the environmental or experimental conditions^{59,61}. These dependencies reveal a fundamental trade-off for the evolution of natural circuits and the design of synthetic circuits in living cells.

While we assumed that exponential population growth, our results hold identically when the population size is kept constant as long as the population size is sufficiently large (100–1000 cells are sufficient in our experience). This includes chemostats where cells are randomly removed. In summary, we presented an agent-based framework that reveals typical features of snapshot data using noise decompositions. The theory shows that noise statistics vary with the experimental setup to grow the cells, whether it is a mother machine or a chemostat. The 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 cells⁶².

Data Availability

All data is available from the corresponding author upon request.

References

1. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186, <https://doi.org/10.1126/science.1070919> (2002).
2. Raser, J. M. & O'shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013, <https://doi.org/10.1126/science.1105891> (2005).
3. Thomas, P., Terradot, G., Danos, V. & Weiß, A. Y. Sources, propagation and consequences of stochasticity in cellular growth. *Nat Commun* **9**, 4528, <https://doi.org/10.1038/s41467-018-06912-9> (2018).
4. Raser, J. M. & O'shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814, <https://doi.org/10.1126/science.1105891> (2004).
5. Harper, C. V. *et al.* Dynamic analysis of stochastic transcription cycles. *PLoS Biol* **9**, e1000607, <https://doi.org/10.1371/journal.pbio.1000607> (2011).
6. Araújo, I. S. *et al.* Stochastic gene expression in *Arabidopsis thaliana*. *Nat Commun* **8**, 2132, <https://doi.org/10.1038/s41467-017-02285-7> (2017).
7. Shahrezaei, V., Ollivier, J. F. & Swain, P. S. Colored extrinsic fluctuations and stochastic gene expression. *Mol Syst Biol* **4**, 196, <https://doi.org/10.1038/msb.2008.31> (2008).
8. Hilfinger, A. & Paulsson, J. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci* **108**, 12167–12172, <https://doi.org/10.1073/pnas.1018832108> (2011).
9. Zechner, C. *et al.* Moment-based inference predicts bimodality in transient gene expression. *Proc Natl Acad Sci* **109**, 8340–8345, <https://doi.org/10.1073/pnas.1200161109> (2012).
10. Bowsler, C. G., Voliotis, M. & Swain, P. S. The fidelity of dynamic signaling by noisy biomolecular networks. *PLoS Comput Biol* **9**, e1002965, <https://doi.org/10.1371/journal.pcbi.1002965> (2013).
11. Singh, A. & Soltani, M. Quantifying intrinsic and extrinsic variability in stochastic gene expression models. *PLoS One* **8**, e84301, <https://doi.org/10.1371/journal.pone.0084301> (2013).
12. Hasenauer, J., Hasenauer, C., Hucho, T. & Theis, F. J. Ode constrained mixture modelling: a method for unraveling subpopulation structures and dynamics. *PLoS Comput Biol* **10**, e1003686, <https://doi.org/10.1371/journal.pcbi.1003686> (2014).

13. Voliotis, M., Thomas, P., Grima, R. & Bowsher, C. G. Stochastic simulation of biomolecular networks in dynamic environments. *PLoS Comput Biol* **12**, e1004923 (2016).
14. Dattani, J. & Barahona, M. Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization. *J Royal Soc Interface* **14**, 20160833, <https://doi.org/10.1098/rsif.2016.0833> (2017).
15. Del Giudice, M., Bo, S., Grigolon, S. & Bosia, C. On the role of extrinsic noise in microRNA-mediated bimodal gene expression. *PLoS Comput Biol* **14**, e1006063 (2018).
16. Kobayashi, T. J., Yokota, R. & Aihara, K. Feedback Regulation and Its Efficiency in Biochemical Networks. *J Stat Phys* **162**, 1425–1449, <https://doi.org/10.1007/s10955-015-1443-2> (2016).
17. Volfson, D. *et al.* Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**, 861–864, <https://doi.org/10.1038/nature04281> (2006).
18. Hashimoto, M. *et al.* Noise-driven growth rate gain in clonal cellular populations. *Proc Natl Acad Sci* 201519412, <https://doi.org/10.1073/pnas.1519412113> (2016).
19. Cerulus, B., New, A. M., Pougach, K. & Verstrepen, K. J. Noise and epigenetic inheritance of single-cell division times influence population fitness. *Curr Biol* **26**, 1138–1147, <https://doi.org/10.1016/j.cub.2016.03.010> (2016).
20. Shen, F. *et al.* Functional proteometrics for cell migration. *Cytometry Part A* **69**, 563–572, <https://doi.org/10.1002/cyto.a.20283> (2006).
21. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci* **99**, 12795–12800, <https://doi.org/10.1073/pnas.162041399> (2002).
22. Chen, L., Wang, R., Kobayashi, T. J. & Aihara, K. Dynamics of gene regulatory networks with cell division cycle. *Phys Rev E* **70**, 011909, <https://doi.org/10.1103/PhysRevE.70.011909> (2004).
23. Lloyd-Price, J., Tran, H. & Ribeiro, A. S. Dynamics of small genetic circuits subject to stochastic partitioning in cell division. *J Theor Biol* **356**, 11–19, <https://doi.org/10.1016/j.jtbi.2014.04.018> (2014).
24. Bierbaum, V. & Klumpp, S. Impact of the cell division cycle on gene circuits. *Phys Biol* **12**, 066003, <https://doi.org/10.1088/1478-3975/12/6/066003> (2015).
25. Johnston, I. G. & Jones, N. S. Closed-form stochastic solutions for non-equilibrium dynamics and inheritance of cellular components over many cell divisions. *Proc Royal Soc A* **471**, 20150050, <https://doi.org/10.1098/rspa.2015.0050> (2015).
26. Soltani, M., Vargas-Garcia, C. A., Antunes, D. & Singh, A. Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes. *PLoS Comput Biol* **12**, e1004972, <https://doi.org/10.1371/journal.pcbi.1004972> (2016).
27. Roy, A. & Klumpp, S. Simulating genetic circuits in bacterial populations with growth heterogeneity. *Biophys J* **114**, 484–492, <https://doi.org/10.1016/j.bpj.2017.11.3745> (2018).
28. Wang, P. *et al.* Robust growth of Escherichia coli. *Curr Biol* **20**, 1099–1103, <https://doi.org/10.1016/j.cub.2010.04.045> (2010).
29. Locke, J. C. & Elowitz, M. B. Using movies to analyse gene circuit dynamics in single cells. *Nat Rev Microbiol* **7**, 383, <https://doi.org/10.1038/nrmicro2056> (2009).
30. Muzzey, D. & van Oudenaarden, A. Quantitative time-lapse fluorescence microscopy in single cells. *Annu Rev Cell Dev Biol* **25**, 301–327, <https://doi.org/10.1146/annurev.cellbio.042308.113408> (2009).
31. Newman, J. R. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840, <https://doi.org/10.1038/nature04785> (2006).
32. So, L.-H. *et al.* General properties of transcriptional time series in Escherichia coli. *Nat Genet* **43**, 554, <https://doi.org/10.1038/ng.821> (2011).
33. Okumus, B. *et al.* Mechanical slowing-down of cytoplasmic diffusion allows *in vivo* counting of proteins in individual cells. *Nat Commun* **7**, 11641, <https://doi.org/10.1038/ncomms11641> (2016).
34. Thomas, P. Making sense of snapshot data: ergodic principle for clonal cell populations. *J Royal Soc Interface* **14**, 20170467, <https://doi.org/10.1098/rsif.2017.0467> (2017).
35. Powell, E. Growth rate and generation time of bacteria, with special reference to continuous culture. *Microbiology* **15**, 492–511, <https://doi.org/10.1099/00221287-15-3-492> (1956).
36. Painter, P. & Marr, A. Mathematics of microbial populations. *Annu Rev Microbiol* **22**, 519–548, <https://doi.org/10.1146/annurev.mi.22.100168.002511> (1968).
37. Wakamoto, Y., Grosberg, A. Y. & Kussell, E. Optimal lineage principle for age-structured populations. *Evolution* **66**, 115–134, <https://doi.org/10.1111/j.1558-5646.2011.01418.x> (2012).
38. Berg, O. G. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol* **71**, 587–603, [https://doi.org/10.1016/0022-5193\(78\)90326-0](https://doi.org/10.1016/0022-5193(78)90326-0) (1978).
39. Schwabe, A. & Bruggeman, F. J. Contributions of cell growth and biochemical reactions to nongenetic variability of cells. *Biophys J* **107**, 301–313, <https://doi.org/10.1016/j.bpj.2014.05.004> (2014).
40. Van Kampen, N. G. *Stochastic processes in physics and chemistry* (Elsevier, 1992).
41. Elf, J. & Ehrenberg, M. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res* **13**, 2475–2484, <https://doi.org/10.1101/gr.1196503> (2003).
42. Shahrezaei, V. & Swain, P. S. Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci* **105**, 17256–17261, <https://doi.org/10.1073/pnas.0803850105> (2008).
43. Bowsher, C. G. & Swain, P. S. Identifying sources of variation and the flow of information in biochemical networks. *Proc Natl Acad Sci* **109**, E1320–E1328, <https://doi.org/10.1073/pnas.1119407109> (2012).
44. Huh, D. & Paulsson, J. Random partitioning of molecules at cell division. *Proc Natl Acad Sci* **108**, 15004–15009, <https://doi.org/10.1073/pnas.1013171108> (2011).
45. Thieffry, D., Huerta, A. M., Pérez-Rueda, E. & Collado-Vides, J. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. *Bioessays* **20**, 433–440, [https://doi.org/10.1002/\(SICI\)1521-1878\(199805\)20:5<433::AID-BIES10>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1521-1878(199805)20:5<433::AID-BIES10>3.0.CO;2-2) (1998).
46. Thomas, P. & Grima, R. Approximate probability distributions of the master equation. *Phys Rev E* **92**, 012120, <https://doi.org/10.1103/PhysRevE.92.012120> (2015).
47. Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* **323**, 785–793, [https://doi.org/10.1016/S0022-2836\(02\)00994-4](https://doi.org/10.1016/S0022-2836(02)00994-4) (2002).
48. Dublanche, Y., Michalodimitrakis, K., Kümmerer, N., Foglierini, M. & Serrano, L. Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol Syst Biol* **2**, 41, <https://doi.org/10.1038/msb4100081> (2006).
49. Singh, A. & Hespanha, J. P. Optimal feedback strength for noise suppression in autoregulatory gene networks. *Biophys J* **96**, 4013–4023, [j.bpj.2009.02.064](https://doi.org/10.1016/j.bpj.2009.02.064) (2009).
50. Bertaux, F., Marguerat, S. & Shahrezaei, V. 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, <https://doi.org/10.1098/rsos.172234> (2018).
51. Vargas-Garcia, C., Ghusinga, K. & Singh, A. Cell size control and gene expression homeostasis in single-cells. *Curr Opin Syst Biol* **8**, 109–116, <https://doi.org/10.1016/j.coisb.2018.01.002> (2018).
52. Blasi, T., Buettner, F., Strasser, M. K., Marr, C. & Theis, F. J. cgcorrect: a method to correct for confounding cell–cell variation due to cell growth in single-cell transcriptomics. *Phys Biol* **14**, 036001, <https://doi.org/10.1088/1478-3975/aa609> (2017).
53. Heerden, J. H. *et al.* Statistics and simulation of growth of single bacterial cells: illustrations with *B. subtilis* and *E. coli*. *Sci Rep* **7**, 16094, <https://doi.org/10.1038/s41598-017-15895-4> (2017).

54. Thomas, P. Analysis of cell size homeostasis at the single-cell and population level. *Frontiers in Physics* **6**, 64, <https://doi.org/10.3389/fphy.2018.00064> (2018).
55. Tsuru, S. *et al.* Noisy cell growth rate leads to fluctuating protein concentration in bacteria. *Phys Biol* **6**, 036015, <https://doi.org/10.1088/1478-3975/6/3/036015> (2009).
56. Grima, R., Thomas, P. & Straube, A. V. How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations? *J Chem Phys* **135**, 084103, <https://doi.org/10.1063/1.3625958> (2011).
57. Schnoerr, D., Sanguinetti, G. & Grima, R. Comparison of different moment-closure approximations for stochastic chemical kinetics. *J Chem Phys* **143**, 11B610_1, <https://doi.org/10.1063/1.4934990> (2015).
58. Kafri, R. *et al.* Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. *Nature* **494**, 480–483, <https://doi.org/10.1038/nature11897> (2013).
59. van Boxtel, C., van Heerden, J. H., Nordholt, N., Schmidt, P. & Bruggeman, F. J. Taking chances and making mistakes: non-genetic phenotypic heterogeneity and its consequences for surviving in dynamic environments. *J Royal Soc Interface* **14**, 20170141, <https://doi.org/10.1098/rsif.2017.0141> (2017).
60. Martins, B. M., Tooke, A. K., Thomas, P. & Locke, J. C. Cell size control driven by the circadian clock and environment in cyanobacteria. *Proc Natl Acad Sci* **115**, E11415–E11424, <https://doi.org/10.1073/pnas.1811309115> (2018).
61. Martins, B. M. & Locke, J. C. Microbial individuality: how single-cell heterogeneity enables population level strategies. *Curr Opin Microbiol* **24**, 104–112, <https://doi.org/10.1016/j.mib.2015.01.003> (2015).
62. Fröhlich, F. *et al.* Inference for stochastic chemical kinetics using moment equations and system size expansion. *PLoS Comput Biol* **12**, e1005030, <https://doi.org/10.1371/journal.pcbi.1005030> (2016).

Acknowledgements

It is a pleasure to thank Vahid Shahrezaei for valuable feedback. P.T. acknowledges a fellowship by The Royal Commission for the Exhibition of 1851 and support by the EPSRC Centre for Mathematics of Precision Healthcare (EP/N014529/1).

Author Contributions

P.T. designed the study, carried out the research and analysed the results.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-35927-x>.

Competing Interests: The author declares no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019