



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Frequency domain analysis of fluctuations of mRNA and protein copy numbers within a cell lineage

**Citation for published version:**

Jia, C & Grima, R 2021, 'Frequency domain analysis of fluctuations of mRNA and protein copy numbers within a cell lineage: Theory and experimental validation', *Physical Review X*.  
<https://doi.org/10.1103/PhysRevX.11.021032>

**Digital Object Identifier (DOI):**

[10.1103/PhysRevX.11.021032](https://doi.org/10.1103/PhysRevX.11.021032)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Physical Review X

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Frequency Domain Analysis of Fluctuations of mRNA and Protein Copy Numbers within a Cell Lineage: Theory and Experimental Validation

Chen Jia<sup>1</sup> and Ramon Grima<sup>2,\*</sup>

<sup>1</sup>*Applied and Computational Mathematics Division, Beijing Computational Science Research Center, Beijing 100193, China*

<sup>2</sup>*School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, United Kingdom*

 (Received 13 October 2020; revised 8 February 2021; accepted 8 March 2021; published 11 May 2021)

The stochasticity of gene expression is manifested in the fluctuations of messenger ribonucleic acid and protein copy numbers within a cell lineage over time. While data of this type can be obtained for many generations, most mathematical models are unsuitable to interpret such data since they assume nongrowing cells. Here we develop a theoretical approach that quantitatively links the frequency content of lineage data to subcellular dynamics. We elucidate how the position, height, and width of the peaks in the power spectrum provide a distinctive fingerprint that encodes a wealth of information about mechanisms controlling transcription, translation, replication, degradation, bursting, promoter switching, cell cycle duration, cell division, gene dosage compensation, and cell-size homeostasis. Predictions are confirmed by analysis of single-cell *Escherichia coli* data obtained using fluorescence microscopy. Furthermore, by matching the experimental and theoretical power spectra, we infer the temperature-dependent gene expression parameters, without the need of measurements relating fluorescence intensities to molecule numbers.

DOI: [10.1103/PhysRevX.11.021032](https://doi.org/10.1103/PhysRevX.11.021032)

Subject Areas: Biological Physics,  
Interdisciplinary Physics,  
Statistical Physics

## I. INTRODUCTION

In recent years, measurements of the size, division events, and the content of single cells over extended time (many generations) have been made possible due to advances in microfluidic devices and live-cell imaging [1–3]. While the existing data are typically for proteins, such measurements are also in principle possible for messenger ribonucleic acids (mRNAs), particularly with the advent of new methods to visualize RNA dynamics in live cells using bright and stable fluorescent RNAs [4]. The data in these experiments are sampled at a rate that is much higher than the frequency of cell division, thus providing us with a means to understand the temporal variation of gene expression as a cell progresses through its cell cycle.

A common feature of these time traces is a noisy oscillatory variation of the fluorescence intensity (from fluorescently labeled proteins) with time with a period that

is roughly coincident with the interval between two successive cell division events [2]. A sawtooth type of temporal pattern is expected due to a sharp dip in the protein numbers at cell division stemming from the partitioning of the contents of the mother cell among two daughter cells. While these oscillations are regular in some cases, very often they display a significant degree of noisiness. This noisiness reflects the inherent stochasticity of transcription, translation, and replication [5–7], noise introduced or modified by homeostatic mechanisms such as those that compensate for the doubling of gene copies at replication [8,9], and nongenetic sources of noise such as variability in the cell cycle duration from one generation to the next [1,10,11] and variability in the number of molecules allocated to a newborn cell at cell division [12,13]. Hence, it follows that a measure of the regularity of an oscillation, such as the power spectrum of fluorescence fluctuations calculated over a lineage, encapsulates within it a large amount of information about the inherent chemical and physical processes, both deterministic and stochastic, that shape cellular dynamics.

An essential first step to link the properties of the power spectrum with the underlying dynamic intracellular processes is the derivation of a mathematical formula for the spectrum as a function of gene expression rate parameters. For this purpose, the standard stochastic models in the

\*Corresponding author.

Ramon.Grima@ed.ac.uk

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

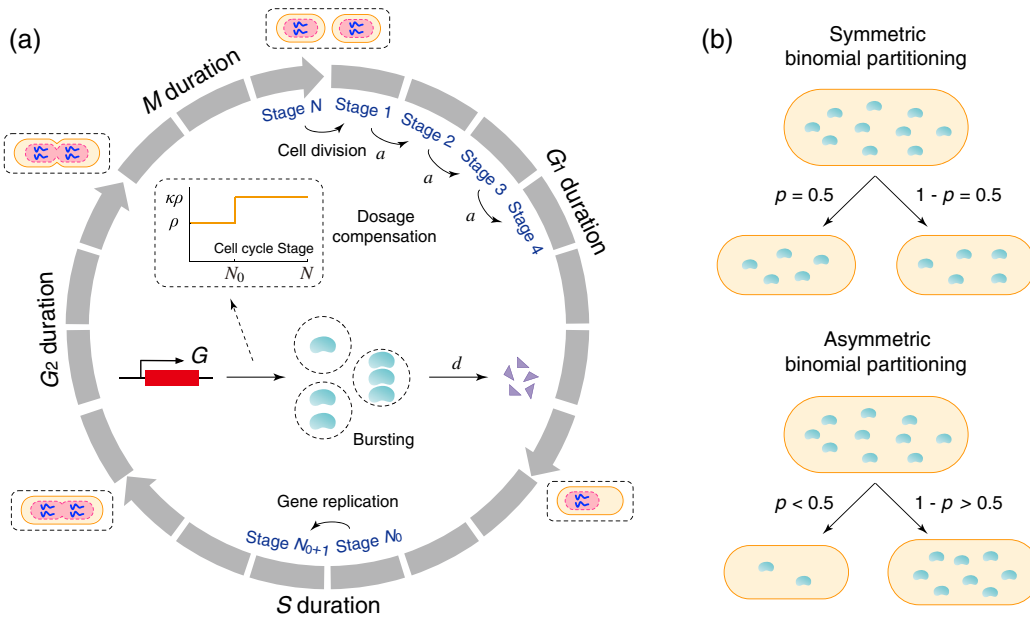


FIG. 1. A detailed gene expression model across the cell cycle. (a) Schematic illustrating the model describing  $N$  effective cell cycle stages, gene replication at stage  $N_0$ , bursty production of the gene product, degradation of the gene product, and gene dosage compensation induced by a change in the burst production rate upon replication from  $\rho$  to  $\kappa\rho$  with  $\kappa < 2$  (see inset graph). (b) At stage  $N$ , a mother cell divides into two daughters that are typically different in size (asymmetric division) with the larger daughter inheriting more molecules. Symmetric division is the special case where the daughters are equisized.

literature that are based on the two-stage or three-stage representation of gene expression [14] are not useful because they do not provide a mathematical description of processes along a cell lineage. These models describe transcription and translation, and implicitly characterize the dilution due to cell division via an effective decay reaction; however, simulated time traces of the protein numbers based on these models using the stochastic simulation algorithm will not display any noisy oscillatory behavior since the partitioning of molecules at cell division is not taken into account. Standard models also lack a description of important cell cycle features such as the high variability in the interdivision time that is a characteristic of experimental time traces. Recently, models that surmount the aforementioned limitations of the standard models have been studied leading to mathematical formulas for the mean and variance of mRNA or protein numbers [15–18] and also for the steady-state distributions of these numbers [19,20] calculated across a cell lineage. These statistical measures provide different information from the power spectrum; notably, the former statistical measures, unlike the latter, neither provide an understanding of the correlations between molecule numbers at two time points nor of the frequency composition of fluctuations in molecule numbers [21].

In this article, for the first time, we calculate in closed form the power spectrum of fluctuations across a lineage in a stochastic gene expression model with a high level of biological realism, including a description of transcription, translation, degradation, bursting, promoter switching,

DNA replication, gene dosage compensation, and symmetric or asymmetric partitioning at cell division. The analytical expressions give insights into how the regularity and noisiness of the oscillations in the mRNA or protein abundance across generations are related to the rate parameters associated with the various subcellular processes at play. The theory also makes various predictions that are then verified by analysis of a publicly available single-cell dataset of *Escherichia coli* followed over 70 generations for three different growth conditions. Finally, we show how matching the experimental and theoretical power spectra enables a reliable inference of gene expression rate parameters.

## II. RESULTS

### A. Model specification

We consider a detailed model of stochastic gene expression dynamics in a single cell which includes transcriptional or translational bursting, degradation, cell cycle duration variability, gene replication, gene dosage compensation, and symmetric or asymmetric cell division [see Fig. 1(a) for an illustration]. The specific meaning of all model parameters is listed in Table I, and the biological values of some key parameters in different cell types are listed in Table II. The model is based on a number of assumptions that are closely tied to experimental data. The assumptions are as follows:

- (i) Before replication, the synthesis of the gene product of interest, mRNA or protein, occurs at a rate  $\rho$  in bursts of a random size sampled from an arbitrary

TABLE I. Model parameters and their meaning.

Meaning of model parameters	
$\rho$	Burst production rate before replication
$\kappa\rho$	Burst production rate after replication
$B$	Mean burst size of the gene product
$d$	Degradation rate of the gene product
$N$	Number of cell cycle stages
$N_0$	Number of cell cycle stages before replication
$w$	Proportion of the cell cycle before replication
$a$	Transition rate from one cell cycle stage to the next
$f$	Cell cycle frequency
$\eta$	Ratio of the degradation rate to the cell cycle frequency
$p$	Allocation probability in asymmetric binomial partitioning
$d_{\text{eff}}$	Effective decay rate of the gene product

distribution  $\mu = (\mu_n)$ . This means that in each burst, there is a probability  $\mu_n$  of producing  $n$  copies of the gene product. In some previous studies [22], the synthesis of the gene product is assumed to be non-bursty; i.e., the molecules are produced one at a time. In this case,  $\mu_n = \delta_{1,n}$  is the Kronecker delta which takes the value of 1 when  $n = 1$  and the value of 0 otherwise. In many other papers [23], the synthesis of the gene product is assumed to be bursty with the burst size sampled from a geometric distribution  $\mu_n = p_B^n(1 - p_B)$ , where  $p_B = B/(1 + B)$  with  $B$  being the mean burst size. Both nonbursty and bursty gene expression are commonly observed in naturally occurring systems [24]. The bursting of mRNA and protein has different biophysical origins [25,26]: Transcriptional bursting is due to a gene that is mostly OFF but transcribes a large number of mRNA molecules when it is ON, whereas translational bursting is due to rapid synthesis of protein molecules from a single short-lived mRNA molecule. Therefore, the arbitrariness of the burst size distribution allows us to analyze both bursty and nonbursty cases within a unified model.

In Appendix A, we also consider a more complicated model where the promoter of the gene of interest can switch between an inactive state  $G_0$  and an active state  $G_1$  [14], each associated with a different burst production rate. Previous studies [25,26] have shown that there is a degree of degeneracy in the model if promoter switching is taken into account since a switching gene coupled with nonbursty production of the gene product can simulate bursty production of the gene product. For this reason, we mainly focus on the model without promoter switching in the main text and defer the analysis of the model with promoter switching to Appendixes A and D.

- (ii) The gene product is degraded via first-order kinetics with rate constant  $d$ , which is a common assumption supported by experiments [27].
- (iii) Each cell can exist in  $N$  effective cell cycle stages denoted by  $1, 2, \dots, N$ , with  $a$  being the transition rate

from one stage to the next, which is assumed to be the same for all stages. Since the transition time between stages is exponentially distributed, the duration of the cell cycle is Erlang distributed with mean  $N/a$ , and thus, the cell cycle frequency is  $f = a/N$ . In our model, the noise in the doubling time characterized by the coefficient of variation squared is equal to  $1/N$ . As  $N \rightarrow \infty$ , the noise vanishes, and thus, the doubling time becomes fixed. Hence, our model allows the investigation of the influence of cell cycle duration variability on stochastic gene expression.

We emphasize that the effective cell cycle stages introduced here do not directly correspond to the four biological cell cycle phases of eukaryotic cells ( $G_1$ ,  $S$ ,  $G_2$ , and  $M$ ) since the durations of the latter are typically not exponentially distributed. In our model, a cell cycle phase corresponds to multiple effective cell cycle stages [Fig. 1(a)]. By introducing a number of effective cell cycle stages, our model has the property that the total duration of the cell cycle and the durations of individual cell cycle phases are all Erlang distributed. This is in agreement with experiments in various cell types [17,28–31].

- (iv) Cell division occurs when the cell transitions from effective stage  $N$  to the next stage 1. At division, most previous papers assume that the mother cell divides into two via symmetric binomial partitioning: Each molecule has an equal chance to be allocated to one of the two daughter cells [20,32]. However, asymmetric cell division is also common in biology [13,33]. For instance, *Saccharomyces cerevisiae* divides asymmetrically into two daughters with different sizes. *Escherichia coli* may also undergo asymmetric division with old daughters receiving fewer gene products than new daughters [34]. Here we extend previous models by considering asymmetric binomial partitioning at cell division: The probability of a molecule being allocated to one daughter is  $p \leq 1/2$ , and the probability of being allocated to the other is  $q = 1 - p$  [Fig. 1(b)]. After division, we randomly track one of the two daughters with probability  $1/2$ ; hence, our model corresponds to cell lineage measurements performed using a mother machine such as in Ref. [35].
- (v) The replication of the locus containing the gene of interest occurs over a period that is much shorter than the rest of the cell cycle. Note that the replication of the whole genome within a cell cannot be assumed to be instantaneous. However, since the replication time of a particular gene is much shorter than the total duration of the  $S$  phase, it is reasonable to consider it to be instantaneous. Specifically, we assume that gene replication occurs instantaneously when the cell transitions from a fixed effective stage  $N_0 \in [1, N - 1]$  to the next stage  $N_0 + 1$ . We refer to the gene copy that is replicated as the mother copy and to the duplicated gene copies as the daughter copies. Under this

TABLE II. Biological values of some of the model parameters for a variety of cell types and experimental conditions. The number  $N$  of effective cell cycle stages is determined by fitting the experimental distribution of doubling time to an Erlang distribution. The lower and upper bounds of  $w$  are determined as the relative durations of the  $G_1$  phase and the  $G_1 + S$  phase, respectively. The bracket in the first column gives the strain, medium, and temperature for prokaryotes or the cell type for eukaryotes. Some more data about mRNA half-lives for other *E. coli* strains can be found in Refs. [38,39]. The brackets in the last two columns give the ranges of the mRNA and protein half-lives for all studied genes in an experiment. The value of  $\eta$  is calculated using the mean doubling time  $T$  and the half-life  $T_H$  via the formula  $\eta = (\ln 2)T/T_H$ .

Cell types	$N$	$w$	Mean doubling time	Median mRNA half-life	$\eta$ for mRNA	Median protein half-life	$\eta$ for protein
<i>E. coli</i> (K10+M9+30°C)	...	...	73 min [38]	2.1 min (0.6–24.6 min) [38]	24.1 (2.1–84.3)	About 20 h [40]	0.042
<i>E. coli</i> (DF261+M9+30°C)	...	...	78 min [38]	2.8 min (0.8–43.5 min) [38]	19.3 (1.2–67.6)	About 20 h [40]	0.045
<i>E. coli</i> (N3433+M9+30°C)	...	...	75 min [38]	3.7 min (1.0–44.2 min) [38]	14.1 (1.2–52.0)	About 20 h [40]	0.043
<i>E. coli</i> (SU02+M9+30°C)	...	...	75 min [38]	4.0 min (1.6–55.5 min) [38]	13.0 (0.9–32.5)	About 20 h [40]	0.043
<i>E. coli</i> (YHC012+M9+30°C)	...	...	84 min [38]	4.0 min (1.2–50.2 min) [38]	14.6 (1.2–48.5)	About 20 h [40]	0.049
<i>E. coli</i> (SH3208+M9+30°C)	...	...	81 min [38]	3.2 min (1.1–24.6 min) [38]	17.5 (2.3–51.0)	About 20 h [40]	0.047
<i>E. coli</i> (BZ453+M9+30°C)	...	...	84 min [38]	6.0 min (1.7–38.6 min) [38]	9.7 (1.5–34.2)	About 20 h [40]	0.049
<i>E. coli</i> (MG1655+M9+30°C)	7 [1]	...	30 min [41]	6.8 min (0.7–34.5 min) [41]	3.1 (0.6–29.7)	About 20 h [40]	0.017
<i>E. coli</i> (JM109+M9+37°C)	5 [42]	...	87 min [41]	...	...	About 20 h [40]	0.050
<i>E. coli</i> (EJ2848+M9+37°C)	9 [43]	...	52 min [41]	...	...	About 20 h [40]	0.030
<i>E. coli</i> (MC4100+LB+25°C)	14 [2]	...	68 min [2]	...	...	...	...
<i>E. coli</i> (MC4100+LB+27°C)	31 [2]	...	53 min [2]	...	...	...	...
<i>E. coli</i> (MC4100+LB+37°C)	31 [2]	...	33 min [2]	...	...	...	...
<i>E. coli</i> (K12+AB+30°C)	...	0.16–0.52 [44]	125 min [44]	...	...	About 20 h [40]	0.072
<i>E. coli</i> (NCM3416+M9+30°C)	...	...	90 min [39]	5.4 min (1.3–31.4 min) [39]	11.6 (2.0–48.0)	About 20 h [40]	0.052
<i>E. coli</i> (NCM3416+LB+30°C)	...	...	30 min [39]	4.7 min (1.1–24.8 min) [39]	4.4 (0.8–18.9)	About 20 h [40]	0.017
<i>M. acetivorans</i> (WWM82+MeOH+37°C)	...	...	7.5 h [45]	1.0 h (4 min–5.8 h) [45]	5.2 (0.9–78.0)	...	...
<i>M. acetivorans</i> (WWM82+TMA+37°C)	...	...	8.9 h [45]	1.1 h (14 min–8.3 h) [45]	5.6 (0.7–26.4)	...	...
<i>M. acetivorans</i> (WWM82+acetate+37°C)	...	...	24.6 h [45]	2.8 h (11 min–9.9 h) [45]	6.1 (1.7–93.0)	...	...
<i>S. cerevisiae</i> (WWM82+acetate+37°C)	31 [46]	0.16–0.43 [47]	1.5–2.5 h [48,49]	20 min (3–>90 min) [27]	5.2(<1.2–34.7)	8.8 h (8 min –>150h) [49]	0.2(<0.01 –13.0)
<i>S. pombe</i>	169 [50]	0.25–0.37 [51]	1.8–4.0 h [49,52]	33 min (10–96 min) [53]	5.0 (1.7–16.6)	12 h (36 min –>150h) [49]	0.2(<0.02–4.6)
Mouse fibroblasts (NIH3T3)	12 [28]	0.40–0.80 [54]	27.5 h [55]	9 h (1.4–40 h) [55]	2.1 (0.5–13.6)	46 h (30 min –>150h) [55]	0.4(<0.13–38.1)
Mouse embryonic stem cells	...	...	About 12 h [56]	7.1 h (0.3–>24h) [57]	1.2(<0.3–27.7)	...	...
Mouse epidermal stem cells	...	...	About 200 h [58]	...	...	...	...
Human (HeLa)	17 [17]	0.65–0.83 [59]	24.7 h [60]	9.5 h (0.8–130 h) [59]	1.9 (0.13–21.4)	...	...

(Table continued)



TABLE II. (Continued)

Cell types	$N$	$w$	Mean doubling time	Median mRNA half-life	$\eta$ for mRNA	Median protein half-life	$\eta$ for protein
Human cancer cells (H1299)	75 [61]	...	22.5 h [61]	...	...	9 h (45 min -22.5 h) [61]	1.7 (0.7-20.8)
Human cancer cells (HepG2)	...	...	24-48 h [62]	10 h (0.1-27.2 h) [62]	3.3 (1.2-332.7)	8.7 h (19 min -36.7 h) [63]	3.8 (0.9-105.1)
Human embryonic stem cells	...	0.19-0.85 [64]	15.8 h [64]	...	...	...	...

assumption, for haploid cells, there is only one mother copy during the first  $N_0$  stages and two daughter copies during the last  $N - N_0$  stages; for diploid cells, the number of gene copies varies from two to four upon replication. For diploid cells, we assume that the two alleles act independently of each other [36,37].

- (vi) If the two daughter copies have the same burst production rate as the mother copy, then the total burst production rate after replication should be equal to  $2\rho$ . Dosage compensation is then modeled as a change in the burst production rate upon replication from  $\rho$  to  $\kappa\rho$  with  $\kappa < 2$ . In the absence of dosage compensation, we have  $\kappa = 2$ .

We next describe our stochastic model for haploid cells. The results for diploid cells can be easily deduced from the haploid case using allelic expression independence (see Appendix B for details). The microstate of the gene of interest can be represented by an ordered pair  $(r, n)$ , where  $r$  is the effective cell cycle stage, and  $n$  is the copy number of the gene product. Let  $p_{r,n}$  denote the probability of observing microstate  $(r, n)$ . Then the evolution of stochastic gene expression dynamics is governed by the master equation

$$\begin{aligned}
 \dot{p}_{1,n} &= \sum_{k=0}^{n-1} \rho \mu_{n-k} p_{1,k} - \sum_{k=1}^{\infty} \rho \mu_k p_{1,n} \\
 &\quad + (n+1) dp_{1,n+1} - n dp_{1,n} \\
 &\quad + \frac{a}{2} \sum_{m=n}^{\infty} C_{m,n} (p^n q^{m-n} + q^n p^{m-n}) p_{N,m} - a p_{1,n}, \\
 \dot{p}_{r,n} &= \sum_{k=0}^{n-1} \rho \mu_{n-k} p_{r,k} - \sum_{k=1}^{\infty} \rho \mu_k p_{r,n} + (n+1) dp_{r,n+1} \\
 &\quad - n dp_{r,n} + a p_{r-1,n} - a p_{r,n}, \quad 2 \leq r \leq N_0, \\
 \dot{p}_{r,n} &= \sum_{k=0}^{n-1} \kappa \rho \mu_{n-k} p_{r,k} - \sum_{k=1}^{\infty} \kappa \rho \mu_k p_{r,n} + (n+1) dp_{r,n+1} \\
 &\quad - n dp_{r,n} + a p_{r-1,n} - a p_{r,n}, \quad N_0 + 1 \leq r \leq N. \quad (1)
 \end{aligned}$$

Here,  $C_{m,n} = m!/n!(m-n)!$  is the combinatorial number, the first two terms on the right-hand side describe bursty production of the gene product, the next two terms describe degradation, and the last two terms describe cell cycle progression. In particular, in the first equation, the last two terms also describe asymmetric partitioning of molecules at cell division.

### B. General expressions for the power spectrum of single-cell measurements across generations

Experiments suggest that the periodicity of the cell cycle can induce oscillatory behavior in gene expression [1]. In fact, if we use a deterministic model to describe the synthesis and degradation of the gene product and assume

that the molecule numbers halve after a deterministic doubling time, then the solution of the deterministic rate equation (the time series of gene product abundances) will be periodic under all choices of rate parameters. However, numerous time-lapse experiments [6,65] have shown that the time course data of expression levels in single cells do not always appear oscillatory due to various sources of noise such as cell cycle duration variability, gene copy number variability, asymmetry of cell division, gene expression bursting, and promoter switching. Here we examine how these sources of noise influence the robustness of sustained oscillations.

Let  $n(t)$  denote the copy number of the gene product in a cell at time  $t$ . Stochastic gene expression oscillations are often characterized by two functions: the autocorrelation function and the power spectrum. The former  $R(t) = \text{Cov}_{\text{SS}}[n(0), n(t)]$  is defined as the steady-state covariance of  $n(0)$  and  $n(t)$ , while the latter  $G(\xi) = \int_{-\infty}^{\infty} R(|t|)e^{-2\pi i \xi t} dt$  is defined as the Fourier transform of the former, where  $\xi \geq 0$  denotes the frequency. In general, sustained oscillations cannot be observed if the power spectrum  $G(\xi)$  is a monotonically decreasing function of  $\xi$ . In contrast, a nonmonotonic power spectrum with one or more off-zero peaks implies the presence of sustained oscillations; the dominant frequencies of these oscillations are the values of  $\xi$  at which the modes of the power spectrum occur.

To obtain the analytical expression of the autocorrelation function, we first calculate the moments of the gene product number. For each cell cycle stage  $r$ , we introduce the generating function  $F_r(z) = \sum_{n=0}^{\infty} p_{r,n} z^n$ . Then, Eq. (1) can be converted into the following partial differential equations:

$$\begin{aligned} \partial_t F_1(z) &= \rho[H(z) - H(1)]F_1(z) + d(1-z)F_1(z) \\ &\quad + \frac{a}{2}[F_N(pz+q) + F_N(qz+p)] - aF_1(z), \\ \partial_t F_r(z) &= \rho[H(z) - H(1)]F_r(z) + d(1-z)F_r(z) \\ &\quad + a[F_{r-1}(z) - F_r(z)], \quad 2 \leq r \leq N_0, \\ \partial_t F_r(z) &= \kappa\rho[H(z) - H(1)]F_r(z) + d(1-z)F_r(z) \\ &\quad + a[F_{r-1}(z) - F_r(z)], \quad N_0 + 1 \leq r \leq N, \end{aligned} \quad (2)$$

where  $H(z) = \sum_{n=1}^{\infty} \mu_n z^n$  is the generating function of the burst size distribution. Recall that the  $k$ th factorial moment of the gene product number in cell cycle stage  $r$  is defined as  $m_{kr} := \sum_{n=0}^{\infty} n(n-1)\cdots(n-k+1)p_{r,n} = F_r^{(k)}(1)$ . Let  $m_k = (m_{kr})$  denote the row vector whose components are the  $k$ th factorial moments in all cell cycle stages. Taking the  $k$ th derivative on both sides of Eq. (2), we obtain

$$\dot{m}_k = m_k W_{kk} + \sum_{l=0}^{k-1} m_l W_{lk}, \quad (3)$$

where  $W_{kk}$  and  $W_{lk}$  are matrices defined as

$$W_{kk} = \begin{pmatrix} -a & a & & & & \\ & -a & a & & & \\ & & \ddots & \ddots & & \\ & & & -a & a & \\ (p^k + q^k)a/2 & & & & & -a \end{pmatrix} - kdI, \quad (4)$$

and  $W_{lk} = C_{k,l} H^{(k-l)}(1)S$ . Here,  $I = \text{diag}(1, \dots, 1)$  is the identity matrix,  $C_{k,l} = k!/l!(k-l)!$  is the combinatorial number, and  $S = \text{diag}(\rho, \dots, \rho, \kappa\rho, \dots, \kappa\rho)$  is the diagonal matrix whose diagonal entries are the burst production rates in all cell cycle stages. In the nonbursty case, the derivative terms in  $W_{lk}$  are given by  $H'(1) = 1$  and  $H^{(k)}(1) = 0$  for any  $k \geq 2$ ; in the bursty case, they are given by  $H^{(k)}(1) = k!B^k$ , where  $B$  is the mean burst size.

Note that Eq. (3) is recursive with respect to  $k$ ; hence, all factorial moments of the gene product number can be solved analytically. Since the autocorrelation function contains only information of moments up to the second order, it can be computed exactly as (see Appendix C for the proof)

$$\begin{aligned} R(t) &= (m_1 + m_2)e^{W_{11}t}\mathbb{1} - (m_1\mathbb{1})^2 \\ &\quad + B \int_0^t m_1 e^{W_{00}s} S e^{W_{11}(t-s)} \mathbb{1} ds, \end{aligned} \quad (5)$$

where  $W_{00}$  and  $W_{11}$  are matrices defined in Eq. (4),

$$\begin{aligned} m_0 &= (1/N, \dots, 1/N), \\ m_1 &= -Bm_0 S W_{11}^{-1}, \\ m_2 &= -[2Bm_1 + H''(1)m_0] S W_{22}^{-1} \end{aligned} \quad (6)$$

are the steady-state zeroth, first, and second factorial moments in all cell cycle stages, and  $\mathbb{1} = (1, \dots, 1)^T$  is the column vector whose components are all 1. Note that  $W_{00}$  is nothing but the generator matrix for the transitions between cell cycle stages. The analytical expressions of the autocorrelation function and power spectrum for the model with promoter switching can be found in Appendix D.

Note that the autocorrelation function  $R(t)$  in Eq. (5) is expressed in matrix form. A more explicit expression can be obtained by expanding the matrix exponentials  $e^{W_{11}t}$  and  $e^{W_{00}s}$  in terms of their eigenvalues and eigenvectors. Since both  $W_{00}$  and  $W_{11}$  are  $N \times N$  circular matrices, their eigenvalues can be computed explicitly. The eigenvalues of  $W_{00}$  are given by

$$\lambda_k = -a + a\omega_k, \quad 0 \leq k \leq N-1, \quad (7)$$

and the eigenvalues of  $W_{11}$  are given by

$$\lambda_{N+k} = -d - a + 2^{-1/N} a \omega_k, \quad 0 \leq k \leq N-1, \quad (8)$$

where  $\omega_k = e^{2\pi k i / N}$  are all the  $N$ th roots of unity. We find that the autocorrelation function (power spectrum) can be rewritten as the linear combination of  $2N-1$  exponential (Lorentzian) functions

$$R(t) = \sum_{k=1}^{2N-1} u_k e^{\lambda_k t}, \quad G(\xi) = \sum_{k=1}^{2N-1} \frac{-2u_k \lambda_k}{4\pi^2 \xi^2 + \lambda_k^2}, \quad (9)$$

where  $\lambda_0, \dots, \lambda_{2N-1}$  are all the eigenvalues of  $W_{00}$  and  $W_{11}$ , and  $u_k$  are suitable constants (see Appendix C for the proof and the specific expressions of  $u_k$ ).

In addition, the steady-state mean of the gene product number can be also calculated explicitly as (see Appendix C for the proof)

$$\langle n \rangle = \frac{\rho B}{\eta f} \left[ w + \kappa(1-w) + \frac{(\kappa-1)a_{\eta,N}^w + 1 - \kappa a_{\eta,N}}{\eta(2a_{\eta,N} - 1)} \right], \quad (10)$$

where  $w = N_0/N$  is the proportion of the cell cycle before replication,  $\eta = d/f$  is the ratio of the degradation rate to the cell cycle frequency which serves as a measure for the stability of the gene product, and  $a_{\eta,N} = (1 + \eta/N)^N \approx e^\eta$  when  $N \gg 1$ .

To validate the analytical expression of the power spectrum, we compare it with the numerical spectrum calculated by means of the Wiener-Khinchin theorem, which states that  $G(\xi) = \lim_{T \rightarrow \infty} \langle |\hat{n}_T(\xi)|^2 \rangle / T$ , where  $\hat{n}_T(\xi) = \int_0^T n(t) e^{-2\pi i \xi t} dt$  is the truncated Fourier transform of a single stochastic trajectory over the interval  $[0, T]$ , and the angled brackets denote the ensemble average over trajectories [Fig. 2(a)]. To guarantee the accuracy of the numerical spectrum, we calculate the ensemble average over 5000 trajectories simulated using Gillespie's stochastic simulation algorithm with the maximum simulation time being chosen as  $30N/a$  (about 30 cell cycles). Here we normalize the power spectrum such that  $G(0) = 1$ . Clearly, the analytical (blue curve) and numerical (red circles) spectra coincide perfectly with each other. In general, the numerical simulations of the power spectrum turn out to be very slow. The analytical solution is hence crucial because it allows a fast exploration of large swathes of parameter space.

### C. Single-cell time traces can be classified into four different types according to their power spectrum shapes

Our gene expression model displays four different types of power spectra [Fig. 2(a)]: (i) the spectrum is unimodal and monotonically decreasing with a peak at zero, (ii) the spectrum is bimodal with the height of the off-zero peak

less than 1, (iii) the spectrum is bimodal with the height of the off-zero peak greater than 1, and (iv) the spectrum is unimodal and bell shaped with the height of the off-zero peak greater than 1. For convenience, we refer to (i)–(iv) as type I, II, III, and IV spectra, respectively. The robustness of oscillations increases as the spectrum changes from type I to type IV [Fig. 2(b)]; this is since the increasing height of the off-zero peak relative to the zero peak implies increasing power in a narrow range of frequencies centered about the cell cycle frequency. To better understand the analytical solution of the power spectrum, we decompose it into two parts:

$$G(\xi) = \sum_{k=1}^{N-1} \frac{-2u_k \lambda_k}{4\pi^2 \xi^2 + \lambda_k^2} + \sum_{k=N}^{2N-1} \frac{-2u_k \lambda_k}{4\pi^2 \xi^2 + \lambda_k^2}, \quad (11)$$

where the first part is the contribution of the eigenvalues of  $W_{00}$ , and the second part is the contribution of those of  $W_{11}$  [see insets of Fig. 2(a) for the decomposition]. Clearly, the first part (green curve) mainly controls the off-zero peak, and the second part (red curve) mainly controls the zero peak as well as the decay of the spectrum. When cell cycle duration variability is small ( $N \gg 1$ ), the first eigenvalue of  $W_{11}$  is given by

$$\begin{aligned} \lambda_N &= -d - a(1 - 2^{-1/N}) = -d - a(1 - e^{-(\ln 2)/N}) \\ &\approx -d - (\ln 2)a/N = -d_{\text{eff}}, \end{aligned} \quad (12)$$

where  $d_{\text{eff}} = d + (\ln 2)f$  is the effective decay rate of the gene product, which is the sum of the decay rates due to active degradation and due to dilution at cell division [67]. This explains why the second part characterizes the decay of the spectrum.

### D. General scaling properties of the height and width of the off-zero spectral peak

To see the influence of cell cycle duration variability on sustained oscillations, we illustrate how the power spectrum given by Eq. (9) varies with  $N$  [Fig. 2(c)]. When  $N$  is very small, the spectrum has only a peak at zero, implying that no regular oscillations can be observed. However, as  $N$  increases, the spectrum becomes nonmonotonic with the off-zero peak becoming higher and closer to (but still less than) the cell cycle frequency  $f$  (shown as a vertical dashed line). This indicates that there is a threshold cell cycle duration variability below which the periodicity of the cell cycle leads to sustained oscillations in gene expression.

Moreover, as  $f$  increases while keeping  $N$  and  $\langle n \rangle$  fixed, the power spectrum becomes wider, but the height of the off-zero peak remains exactly the same [Fig. S1(a) in the Supplemental Material [66]]. In fact, if the cell cycle frequency is increased from  $f$  to  $\alpha f$  with some  $\alpha > 1$ , then the coefficients  $u_k$  in Eq. (9) will remain the same, but the eigenvalues  $\lambda_k$  will be replaced by  $\alpha \lambda_k$  (see Appendix C



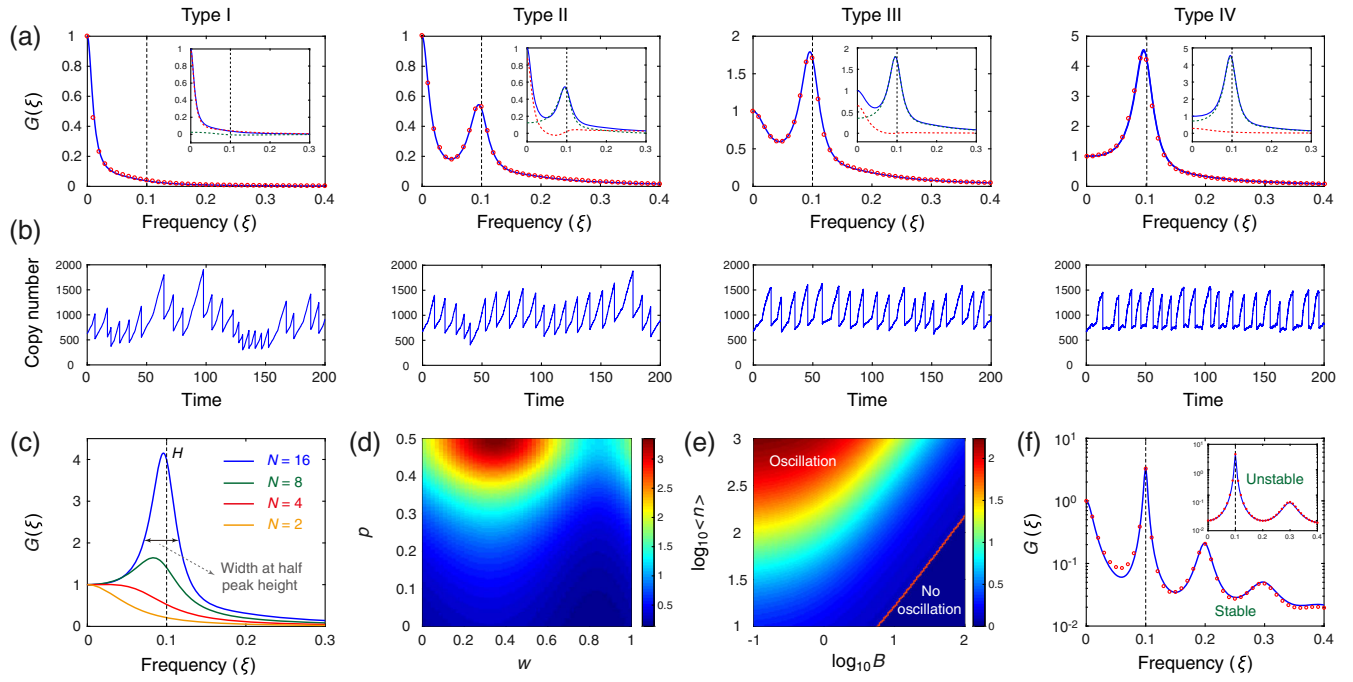


FIG. 2. Properties of the power spectrum. (a) For significant cell cycle duration variability, the spectra can be of four types. Note that the spectra are normalized such that  $G(0) = 1$ . Theory (blue) matches stochastic simulations using Gillespie’s algorithm (red circles). Vertical dashed lines show the cell cycle frequency. Insets show the decomposition of the spectrum into two parts controlling the zero and off-zero peaks (red and green dashed curves, respectively). (b) Typical trajectories for the four types of spectra. (c) The off-zero peak becomes higher as cell cycle duration variability decreases ( $N$  increases). (d),(e) Heat maps showing the dependence of the height of the off-zero peak (relative to the zero peak) on the parameters controlling the asymmetry of partitioning ( $p$ ), the replication point ( $w$ ), the burstiness of gene expression ( $B$ ), and the mean number of gene product molecules ( $\langle n \rangle$ ). Type I spectra are observed in the region marked “no oscillation.” (f) For low cell cycle duration variability, higher-order harmonics of the cell cycle frequency appear; for midway replication, an unstable gene product, e.g., most mRNAs, has a spectrum with no peaks at even harmonics, while a stable gene product, e.g., most proteins, has peaks at all harmonics. Note that the blue curves are computed using the analytical expression given in Eq. (9), and the red circles are computed using the approximate expressions given in Eq. (13) for stable gene products and Eq. (F1) for unstable gene products. To make the higher-order peaks more apparent, the vertical axis is shown on the logarithmic scale. See Sec. S5 in the Supplemental Material [66] for the technical details of this figure.

for the proof). Therefore, the power spectrum (autocorrelation function) will be stretched (compressed) along the horizontal axis by a factor of  $\alpha$ . This explains why the height of the off-zero peak is independent of the cell cycle frequency.

### E. Symmetric division, midway replication, and nonbursty expression enhance the regularity of oscillations

Oscillations are also affected by the gene replication time, asymmetric cell division, random bursting, and gene expression mean. Figures 2(d) and 2(e) illustrate the height of the off-zero peak as a function of  $w$ ,  $p$ ,  $B$ , and  $\langle n \rangle$ . It is clear that the off-zero peak becomes lower as  $B$  increases and as  $p$  and  $\langle n \rangle$  decrease. The decline of the height with increasing  $B$  and decreasing  $p$  and  $\langle n \rangle$  is expected since all of them correspond to an increase in the fluctuations of gene product abundances which counteracts the regularity of oscillations; indeed, noise above a certain threshold can even completely destroy oscillations [Fig. 2(e)]. This

observation is in sharp contrast to a negative feedback genetic loop, where random bursting can promote the regularity of oscillations [68,69]. In addition, we also find that sustained oscillations are the most regular when  $w$  is neither too small nor too large [Fig. 2(d)].

### F. Contrasting the properties of the off-zero peak for stable and unstable gene products

To further understand these observations, we consider two important special cases. In bacteria and yeast, most proteins have very long half-lives, i.e.,  $\eta \ll 1$  (Table II). For such stable gene products, when cell cycle duration variability is not too large, the power spectrum can be simplified as

$$G(\xi) \approx \frac{2A\langle n \rangle f}{3(4\pi^2 \xi^2 + \lambda_N^2)} + \frac{\rho^2 B^2 N}{2f} \sum_{k=1}^{[N/2]} H_k G_k(\xi), \quad (13)$$

where  $A$  and  $H_k$ ,  $k \geq 1$  are suitable constants and

$$G_k(\xi) = \frac{2k^2\pi^2 f^2 - (\pi^2 \xi^2 - a^2 \cos \theta_k \sin^2 \theta_k / 2)}{4k^6 \pi^6 f^4 + N^2 (\pi^2 \xi^2 - a^2 \cos \theta_k \sin^2 \theta_k / 2)^2}, \quad (14)$$

with  $\theta_k = 2k\pi/N$  (see Appendix E for the expressions of  $A$  and  $H_k$  and Sec. S1 in the Supplemental Material [66] for the proof). In Eq. (13), the decay of the power spectrum is mainly controlled by the first term, while the off-zero peak is controlled by the function  $G_1(\xi)$  in the second term. The influence of the remaining functions  $G_k(\xi)$ ,  $k \geq 2$  in the second term will be discussed later.

From Eq. (14) with  $k = 1$ , the position of the off-zero peak is given by  $\xi = (a/\pi) \cos \theta_1 \sin(\theta_1/2) < a\theta_1/2\pi = a/N = f$ , which is smaller than the cell cycle frequency  $f$ . When  $N \gg 1$ , the peak position is approximately equal to  $f$  since  $\sin \theta \approx \theta$  and  $\cos \theta \approx 1$  when  $\theta$  is small [Fig. 2(c)]. Moreover, the width of the off-zero peak characterized by the difference of the frequencies at which the spectrum attains half of its peak value [see Fig. 2(c) for an illustration] is given by  $D = 2\pi f/N$ . In other words, the width is proportional to both the cell cycle frequency and cell cycle duration variability [in agreement with Fig. S1(a) in the Supplemental Material [66]]. In the case of bursty gene expression, the height of the off-zero peak is given by [see Sec. S1 in the Supplemental Material [66] for the proof]

$$H = \frac{J_1 \langle n \rangle N / 6\pi^4}{\langle n \rangle [2J(\gamma + \frac{2}{3N}) + \frac{C}{\pi^2 N} + \frac{3}{\ln 2} + 2B(\gamma - 2 + \frac{3}{\ln 2})]}, \quad (15)$$

where  $\gamma = 2(1 - 4pq)/(1 + 2pq)$  is a function of  $p$ , and

$$J_k = \frac{[w + \kappa(1 - w) + \frac{(\kappa-1) \sin(2k\pi)}{2k\pi}]^2 + \frac{(\kappa-1)^2 \sin^4(k\pi)}{k^2 \pi^2}}{[\kappa - \frac{1}{3}(\kappa-1)w(4-w)]^2},$$

$$J = \frac{[w + \kappa(1 - w)]^2}{[\kappa - \frac{1}{3}(\kappa-1)w(4-w)]^2}, \quad (16)$$

and  $C = \sum_{k=1}^{\infty} J_k/k^2$  are all functions of  $\kappa$  and  $w$ . In the nonbursty case, the term involving  $B$  in Eq. (15) vanishes. Note that since  $p \leq 1/2$  is the probability that a molecule is allocated to one daughter and  $q = 1 - p$  is the probability of being allocated to the other daughter, it follows that  $0 \leq \gamma \leq 2$ ; hence, the parameter  $\gamma$  is a dimensionless measure of the asymmetry of cell division. From Eq. (15), it is easy to see that the height decreases with increasing  $B$  and decreasing  $p$  and  $\langle n \rangle$ . This is in full agreement with the numerical results shown in Figs. 2(d) and 2(e), which are computed using Eq. (9).

On the other hand, in bacteria and yeast, most mRNAs have very short half-lives, i.e.,  $\eta \gg 1$  (Table II). For such unstable gene products, we also derive a simplified expression of the power spectrum which is given in Appendix F. In this case, the width of the off-zero peak is still given by  $D = 2\pi f/N$ , but the height is given by (see Sec. S2 in the Supplemental Material [66] for the proof)

$$H = \frac{\sin^2(w\pi)N^2}{2C(w)\pi^2}, \quad (17)$$

where  $C(w) = \sum_{k=1}^{\infty} \sin^2(kw\pi)/k^2$  is a function of  $w$ . We stress that this formula is derived in the limit of  $\eta \rightarrow \infty$ . In naturally occurring systems, the value of  $\eta$  for an unstable gene product in general does not exceed 100 (Table II), and hence, applying this formula may lead to some errors, especially when  $\kappa$  is very close to 1. Interestingly, for unstable products, the height of the off-zero peak depends less on asymmetric division, gene expression mean, random bursting, and dosage compensation, whereas from Eq. (15), it is clear that the opposite is true for stable products. In both stable and unstable cases, as cell cycle duration variability becomes smaller ( $N$  increases), the off-zero peak becomes narrower and higher.

According to our simulations in Fig. 2(d), oscillations are the most regular when  $w$  is neither too small nor too large. For stable gene products with  $\eta \ll 1$ , it follows from Eq. (15) that the maximal regularity is obtained when  $w \approx 0.29$  for the case of symmetric division, large gene expression mean, and no dosage compensation ( $\kappa = 2$ ). For unstable gene products with  $\eta \gg 1$ , it follows from Eq. (17) that the maximal regularity is obtained when  $w = 0.5$ .

Note that our conclusions about the differences between the power spectra for (unstable) mRNA and (stable) protein are also confirmed by numerical simulations of a more complex model with both mRNA and protein descriptions (rather than an effective protein description with bursting dynamics as described in point 1 in Sec. II A).

### G. Increasing the asymmetry of cell division induces a sharp change in the power spectrum for stable gene products

Our theory further reveals an important difference between symmetric ( $p = 0.5$ ) and asymmetric ( $p < 0.5$ ) cell division for stable gene products. When the gene expression mean is large, it follows from Eq. (15) that the height of the off-zero peak can be simplified as

$$H \approx \frac{J_1 N}{6\pi^4 [2J(\gamma + \frac{2}{3N}) + \frac{C}{\pi^2 N}]}. \quad (18)$$

For the case of symmetric division, we have  $\gamma = 0$ , and thus, the height depends on  $N$  quadratically as  $H \approx J_1 N^2 / 2\pi^2 (4\pi^2 J + 3C)$  [Fig. S2(a) in the Supplemental Material [66]]. For the case of asymmetric division, we have  $0 < \gamma \leq 2$ , and thus, when cycle duration variability is very small ( $N \gg 1$ ), the height depends on  $N$  linearly as  $H \approx J_1 N / 12\pi^4 J \gamma$  [Fig. S2(b) in the Supplemental Material [66]]. As cell division asymmetry becomes stronger, the height transitions sharply from the  $N^2$  law

to the  $N$  law [Fig. S2(c) in the Supplemental Material [66]]. This implies that for a certain cell cycle duration variability (fixed  $N$ ), symmetric division leads to a much higher degree of regularity in oscillations for stable gene products than asymmetric division. In contrast, we find that for unstable gene products, there is no analogous transition because the height of the off-zero peak is always proportional to  $N^2$  and is independent of  $p$  which controls the asymmetry of cell division [see Eq. (17)].

### H. Single-cell time traces can display higher-order harmonics of the cell cycle frequency

Interestingly, when cell cycle duration variability is very small, besides the peak at the cell cycle frequency  $f$ , the power spectrum also has peaks at integer multiples of  $f$  [see Fig. 2(f) and Figs. S2(a) and 2(b) in the Supplemental Material [66]]. This observation indicates that in addition to the fundamental period of the mean doubling time  $T = 1/f$ , the system also has the hidden periods of  $T/2$  and even  $T/3$ . Note that in Figs. 2(a) and 2(c), higher-order harmonics are not observed because the variability in the cell cycle duration is not sufficiently small ( $N$  is not sufficiently large). Similar peaks at higher-order harmonic frequencies have been previously reported for biochemical systems with feedback loops due to the combination of intrinsic noise and nonlinearity in the law of mass action [70]. In the present model, the propensities of the reactions are all linear in molecule numbers, and hence, the hidden periods cannot be attributed to the same mechanism as in Ref. [70].

These hidden frequencies can be better understood using our analytical results. Recall that the spectral peak at  $f$  is controlled by the function  $G_1(\xi)$  in the second term of Eq. (13). In fact, the remaining functions  $G_k(\xi)$ ,  $k \geq 2$  in the second term control the peaks at higher-order harmonics  $kf$ . From Eq. (14), for stable gene products, the height and width of the spectral peak at the  $k$ th harmonic frequency are given by  $J_k H / J_1 k^4$  and  $k^2 D$ , respectively, where  $J_k$  is given in Eq. (16). In particular, when  $w$  is very small or very large, we have  $J_k \approx 1$  for all  $k$ , and hence, the height of the spectral peak at  $f$  is  $2^4 = 16$  times greater than that at  $2f$  and  $3^4 = 81$  times greater than that at  $3f$ . Moreover, the width of the spectral peak at  $f$  is  $2^2 = 4$  times lesser than that at  $2f$  and  $3^2 = 9$  times lesser than that at  $3f$ . This is in full agreement with our simulation results which show that the peaks at higher-order harmonic frequencies are much lower and wider than the peak at the fundamental frequency [Fig. 2(f)].

For unstable gene products, similar phenomena are also observed, but the characteristics of higher-order peaks are slightly different. In fact, the height and width of the spectral peak at the  $k$ th harmonic frequency are given by  $\sin^2(kw\pi)H / \sin^2(w\pi)k^4$  and  $k^2 D$ , respectively (see Appendix F for the proof). When  $w$  is very small or very

large, we have  $\sin^2(kw\pi) / \sin^2(w\pi) \approx k^2$ , and hence, the height of the spectral peak at  $f$  is  $2^2 = 4$  times greater than that at  $2f$  and  $3^2 = 9$  times greater than that at  $3f$ . Another interesting prediction is that for midway replication ( $w = 0.5$ ), there are no peaks at even harmonics since the height at  $2kf$  is zero [inset of Fig. 2(f)]. Therefore, we find that for midway replication, stable gene products yield higher peaks at higher-order harmonics, while for early or late replication, unstable gene products yield higher peaks at higher-order harmonics.

### I. Bifurcations between the four types of power spectra are observed as the gene product stability is varied

The dependence of oscillations on gene product stability  $\eta$  is much more complicated. Figures 3(a)–3(c) illustrate the height of the off-zero peak as a function of  $\eta$ ,  $\kappa$ , and  $N$  when the gene expression mean is large. Since type I spectra are monotonically decreasing, the height of the off-zero peak is set to be zero for convenience. For stable gene products with  $\eta \ll 1$ , it is only possible to observe type I, II, and III spectra. It can be analytically shown that for the case of symmetric division, large gene expression mean, midway replication ( $w = 0.5$ ), and no dosage compensation ( $\kappa = 2$ ), stable gene products yield type I spectra when  $N \leq 6$ , type II when  $7 \leq N \leq 28$ , and type III when  $N \geq 29$  (see Sec. S1 in the Supplemental Material [66] for the proof).

When  $N$  is small and  $\eta$  is not restricted to small values, all four types of power spectra can be observed; as  $\eta$  increases, the system undergoes three stochastic bifurcations from type I to type II, then to type III, and finally to type IV [Fig. 3(a)]; when  $N$  is moderate, type I spectra cannot occur, and the system undergoes two stochastic bifurcations from type II to type III, and then to type IV [Fig. 3(b) and Fig. S1(b) in the Supplemental Material [66]]; when  $N$  is large, both type I and II spectra fail to be observed, and the system undergoes only one stochastic bifurcation from type III to type IV [Fig. 3(c)]. These results show that gene expression oscillations tend to occur when (i)  $N$  is small but  $\eta$  is large (unstable gene products with large cell cycle duration variability) or (ii)  $N$  is moderate or large. In the latter case of moderate or low cell cycle duration variability, oscillations can be observed for all values of  $\eta$  and hence are expected for both stable and unstable gene products.

The reason why unstable gene products yield more robust sustained oscillations can be understood as follows. For stable gene products, since active degradation is weak, the number of molecules increases approximately linearly with time, and hence, the time traces of gene product abundances appear like a sawtooth wave [upper panel of Fig. 3(d)]. Because of cell cycle duration variability and noise due to partitioning at division, the expression levels at the end (or beginning) of each generation have large

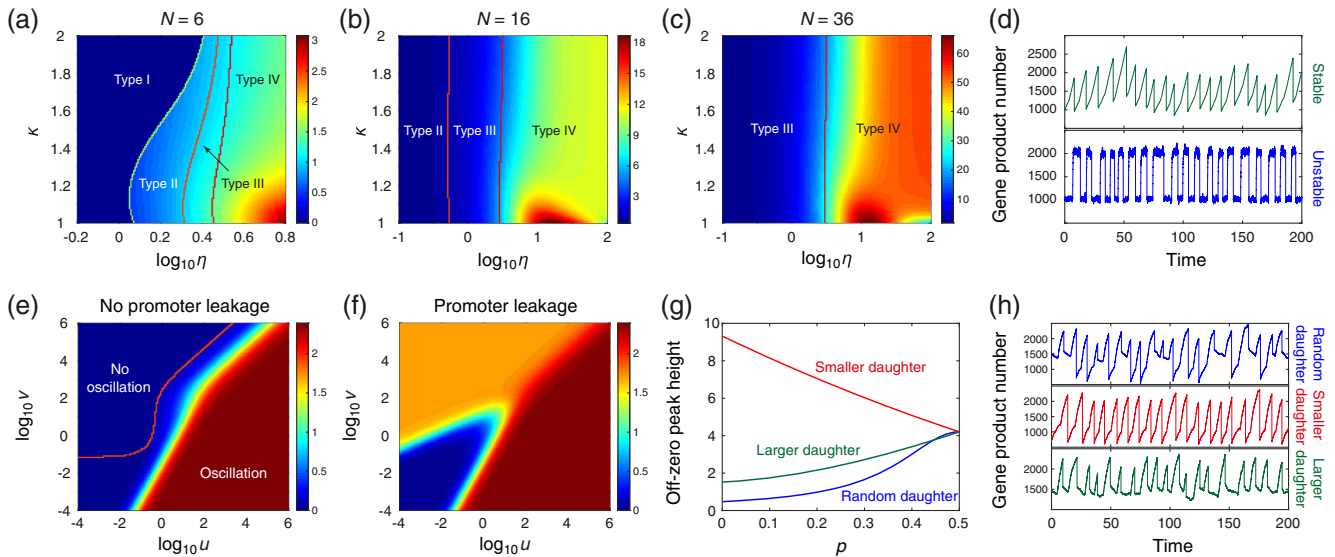


FIG. 3. Further properties of the power spectrum. (a)–(c) Heat maps showing the dependence of the height of the off-zero peak (relative to the zero peak) on the parameters controlling gene product stability ( $\eta$ ) and dosage compensation ( $\kappa$ ). For stable gene products with  $\eta \ll 1$ , it is only possible to observe type I–III spectra; type IV can be produced by unstable gene products with  $\eta \gg 1$ . Strong dosage compensation ( $\kappa = 1$ ) in (b),(c) leads to maximal oscillation regularity at an intermediate gene product stability. (d) Typical trajectories for stable and unstable gene products. With the same gene expression mean, unstable products lead to more robust sustained oscillations than stable ones. (e) Heat map showing the dependence of the height of the off-zero peak (relative to the zero peak) on promoter switching rates  $u$  and  $v$  in the absence of promoter leakage. (f) Same as (e) but in the presence of promoter leakage. Fast promoter switching, leakage, and active gene-state dominance enhance oscillation regularity. (g) Height of the off-zero peak (relative to the zero peak) versus the asymmetric allocation probability  $p$  for three types of tracking protocols at cell division: tracking one of the two daughters randomly (blue), tracking the smaller daughter (red), and tracking the larger daughter (green). (h) Typical trajectories for the three protocols generated using Gillespie’s algorithm. Note that (a)–(c) and (e)–(g) are obtained from numeric evaluation of Eq. (9). See Sec. S6 in the Supplemental Material [66] for the technical details of this figure.

fluctuations. Such noise results in the zero peak of the power spectrum, which explains why for small  $\eta$ , the spectra are of types I–III. In contrast, for unstable gene products, since the degradation rate is large, molecule numbers quickly reach a steady state, and hence, the time traces of gene product abundances appear like a square wave [lower panel of Fig. 3(d)]. The two levels of the square wave correspond to the steady-state levels before and after replication. Once the expression level deviates from the steady-state value, the large degradation rate will help it relax to the steady state rapidly, and hence, the expression levels at the end (or beginning) of each generation will have relatively small fluctuations. This explains why for large  $\eta$ , the spectra are of type IV; i.e., they do not have a peak at zero frequency, but rather the power is concentrated in a narrow bandwidth of frequencies close to the cell cycle frequency.

### J. Strong dosage compensation causes resonancelike behavior at intermediate gene product stability

The pattern of sustained oscillations is also influenced by dosage compensation. When dosage compensation is

weak ( $\kappa$  is close to 2), the height of the off-zero peak (relative to the zero peak) is an increasing function of  $\eta$  [Figs. 3(a)–3(c)]. In this case, the more unstable the gene products are, the more capable they are of exhibiting regular gene expression oscillations. However, when dosage compensation is strong ( $\kappa$  is close to 1) and cell cycle duration variability is not too large, there is an optimal  $\eta$  such that the height is maximized [Figs. 3(b) and 3(c)]; oscillations are the most regular when  $\eta$  is around 15, which falls within the biological range of a typical mRNA in bacteria (Table II). The reason for this phenomenon can be understood as follows. On one hand, oscillations cannot be very regular for small  $\eta$  due to noise in cell cycle duration and partitioning at division, as discussed earlier. On the other hand, for large  $\eta$ , oscillations also cannot be very regular because when  $\kappa \approx 1$ , there is little change in the burst production rate across the cell cycle, and hence, the steady-state expression levels before and after replication [shown as the two levels of the square wave in Fig. 3(d)] will merge into one. As a result, when dosage compensation is strong, sustained oscillations are the most regular at an intermediate  $\eta$  value.



### K. Fast promoter switching, leakage, and active gene-state dominance enhance oscillation regularity

In Appendix A, we consider a more general model with promoter switching, with which we investigate the effect of promoter switching on gene expression oscillations. Figures 3(e) and 3(f) illustrate the height of the off-zero peak as a function of the promoter switching rates,  $u$  and  $v$ , in the absence and presence of promoter leakage, where promoter leakage means that there is a nonzero burst production rate when the gene is in the inactive state. For the case of no leakage, oscillations are manifest when the gene is mostly active, i.e.,  $u \gg v$ , and fail to occur when the gene is mostly inactive, i.e.,  $u \ll v$ , likely due to an exceptionally small gene expression mean. In the presence of leakage, however, we observe a strong oscillation when  $u \gg v$  and a weaker oscillation when  $u \ll v$ . Because of promoter leakage, even when the gene is mostly inactive, there is still a smaller but nonvanishing gene expression mean, which leads to the weaker oscillation observed. In addition, we find that the height of the off-zero peak is exceptionally small when promoter switching is very slow [71], i.e.,  $u, v \ll 1$ , regardless of whether there is promoter leakage or not. We emphasize that while Figs. 3(e) and 3(f) display the results for the inheritance mechanism (the daughter copies inherit the gene state of the mother copy after replication), similar results also hold for the reset mechanism (the daughter copies are reset to the inactive state after replication).

### L. The power spectrum in asymmetrically dividing cells is strongly influenced by the single-cell tracking protocol

In some previous papers [1,72], to track a cell lineage, one of the two daughters was randomly selected at division with probability  $1/2$ . However, for asymmetric division, the two daughters are of different sizes, and another possible protocol is to track the smaller or larger daughter (such as the bud or mother cell in budding yeast) at division [3,73]. Assuming well mixing, the probability for a newborn cell receiving a gene product molecule is equal to the ratio of the volume of the newborn to the volume of the mother cell, and hence, on average the smaller daughter receives fewer gene products than the larger one. We remind the reader that in our model the probability of a molecule being allocated to one daughter is  $p \leq 1/2$ , and the probability of being allocated to the other is  $q = 1 - p$ . Hence, it follows that the daughter with allocation probability  $p < 1/2$  is the smaller daughter [Fig. 1(b)]. Thus far, we assume that one of the two daughters is randomly tracked after division. Now we study two other tracking protocols, namely, where we always follow the smaller or the larger daughter after division.

For smaller daughter tracking, the master equation for the stochastic gene expression dynamics should be modified as

$$\begin{aligned} \dot{p}_{1,n} &= \sum_{k=0}^{n-1} \rho \mu_{n-k} p_{1,k} - \sum_{k=1}^{\infty} \rho \mu_k p_{1,n} \\ &\quad + (n+1)dp_{1,n+1} - ndp_{1,n} \\ &\quad + a \sum_{m=n}^{\infty} C_{m,n} p^n q^{m-n} p_{N,m} - ap_{1,n}, \\ \dot{p}_{r,n} &= \sum_{k=0}^{n-1} \rho \mu_{n-k} p_{r,k} - \sum_{k=1}^{\infty} \rho \mu_k p_{r,n} \\ &\quad + (n+1)dp_{r,n+1} - ndp_{r,n} \\ &\quad + ap_{r-1,n} - ap_{r,n}, \quad 2 \leq r \leq N_0, \\ \dot{p}_{r,n} &= \sum_{k=0}^{n-1} \kappa \rho \mu_{n-k} p_{r,k} - \sum_{k=1}^{\infty} \kappa \rho \mu_k p_{r,n} \\ &\quad + (n+1)dp_{r,n+1} - ndp_{r,n} \\ &\quad + ap_{r-1,n} - ap_{r,n}, \quad N_0 + 1 \leq r \leq N, \end{aligned} \quad (19)$$

where only the first equation differs from that in Eq. (1) due to a different tracking strategy. In this case, the autocorrelation function  $R(t)$  also has the same form as in Eq. (5), where the matrices  $W_{00}$  and  $W_{11}$ , as well as the matrix  $W_{22}$  in Eq. (6), should be replaced by

$$W_{kk} = \begin{pmatrix} -a & a & & & \\ & -a & a & & \\ & & \ddots & \ddots & \\ & & & -a & a \\ p^k a & & & & -a \end{pmatrix} - kdI, \quad k=0,1,2.$$

For larger daughter tracking, we only need to interchange  $p$  and  $q$  in the above two equations.

To compare the three types of tracking protocols at cell division (tracking a random daughter, the smaller daughter, or the larger daughter), we illustrate the height of the off-zero peak as a function of  $p$  for moderately unstable gene products with  $\eta = 2$  [Fig. 3(g)]. Clearly, the three types of tracking protocols lead to the same oscillatory behavior for symmetric division. However, for asymmetric division, the smaller daughter tracking protocol yields a much higher off-zero peak than the other two protocols, implying the most regular oscillations. The reason for this observation is as follows. For unstable gene products, the number of molecules relaxes to the steady-state value rapidly, and hence, the expression levels just before division are roughly the same for the three protocols. However, just after division, the smaller daughter receives fewer molecules than the larger one. Hence, it follows that smaller daughter tracking yields larger amplitudes of oscillations in the time traces than larger daughter tracking [Fig. 3(h)]. Compared to random tracking, smaller or larger daughter tracking leads to less noisy oscillations, presumably since the



TABLE III. The medians of the estimated parameters for all cell lineages at three different temperatures. The distributions of the estimated parameters can be found in Figs. S4 and S5 in the Supplemental Material [66]. The medians rather than the means are reported here since the estimation of the former is more robust than that of the latter, with respect to outliers.

	25°C	27°C	37°C
$f$ (min <sup>-1</sup> )	0.0148	0.0187	0.0307
$T = 1/f$ (min)	67.5676	53.4759	32.5733
$N$	14	31	31
$w$	0.2000	0.1450	0.1224
$\kappa$	2.0621	2.0000	2.9412
$\beta$	42.0719	15.4767	13.0344
$\langle n \rangle$	99.2042	199.5930	287.8168
$\rho$ (min <sup>-1</sup> )	0.1299	0.2303	0.6259
$\kappa\rho$ (min <sup>-1</sup> )	0.2232	0.5319	1.3809
$B$	4.7718	5.2099	4.8606

latter protocol introduces a deterministic element in the tracking process. This explains why random tracking leads generally to less robust oscillations than the other two tracking protocols [Fig. 3(g)]. Note that while increasing cell division asymmetry (decreasing  $p$ ) reduces the robustness of oscillations for random and larger daughter tracking, it leads to the opposite effect for small daughter tracking.

In Fig. S3 in the Supplemental Material [66], we also investigate the differences between the power spectra obtained from the three tracking protocols as functions of  $\eta$  which is a measure of gene product stability. We find that the differences increase with gene product stability. This can be explained as follows. The abrupt change in the number of molecules at division as we switch from a mother cell to a daughter cell is sensitive to the choice of protocol; these low-frequency fluctuations contribute to the height of the zero peak which is sizeable for very stable products and fairly small for very unstable products [Figs. 3(a)–3(c)]. Hence, the choice of protocol has the most effect on the spectra of stable gene products.

### M. Experimental validation of the theory and its application to parameter inference

To test our theory, we apply it to study oscillations in single-cell gene expression data collected for *E. coli* in Ref. [2]. In this dataset, the time course data of fluorescence intensities of a yellow fluorescent protein is recorded every minute for 279 cell lineages over 70 generations using a mother machine under three different growth conditions (25°C, 27°C, and 37°C). At the three temperatures, there is a total of 65, 54, and 160 cell lineages measured, respectively. Based on such data, it is possible to estimate all the parameters involved in our model for each cell lineage. The medians of the estimated parameters for all cell lineage are listed in Table III, and the distributions of

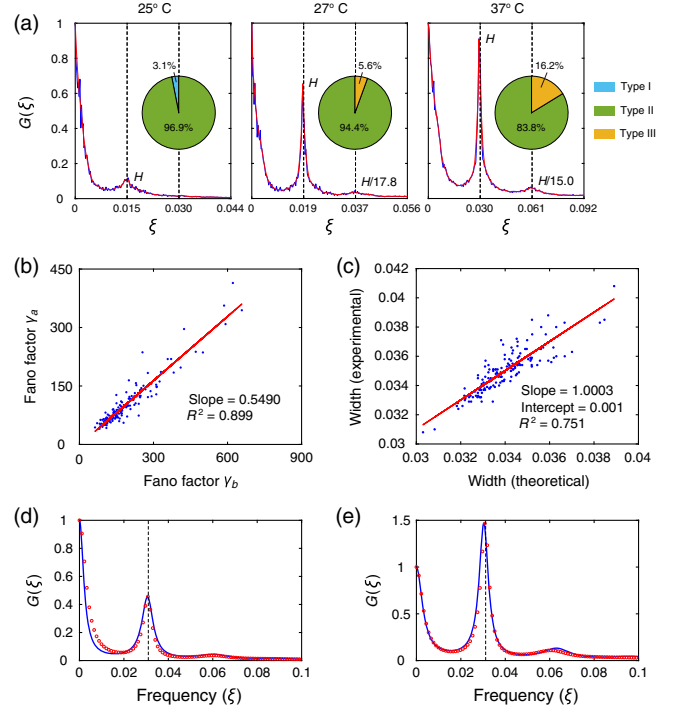


FIG. 4. Analysis of single-cell time course data of fluorescence intensities in *E. coli* published in Ref. [2]. (a) Average power spectra over all cell lineages at three different temperatures (blue curves) and their smoothed approximations (red curves) using the Gaussian filter. The average spectra are estimated by means of the Wiener-Khinchin theorem, where 65, 54, and 160 cell lineages are averaged over, respectively, for the three growth conditions. The power spectrum for each cell lineage is also estimated by fitting the time course data to an AR model. The pie charts display the percentages of various types of power spectra for all cell lineages. (b) Comparison between the Fano factor  $\gamma_b$  of the fluorescence intensities just before division and the Fano factor  $\gamma_a$  of those just after division for all lineages at 37°C. (c) Comparison between the widths of the experimental power spectra obtained using the AR model technique and the theoretical power spectra determined using the estimated parameters for all lineages at 37°C. (d),(e) Comparison between the experimental (blue curve) and theoretical (red circles) power spectra for two typical lineages at 37°C.

the estimated parameters are given in Figs. S4 and S5 in the Supplemental Material [66]. In the following, we briefly describe the estimation procedures.

Since the protein used in the experiment is very stable, it is reasonable to assume that it has negligible degradation, i.e.,  $d = \eta = 0$  [2,72]. Based on the time course data, we estimate the average power spectrum over all cell lineages at each temperature by means of the Wiener-Khinchin theorem [Fig. 4(a)], where we normalize the spectrum such that  $G(0) = 1$ . Clearly, the average power spectra are of type II for all the three growth conditions. As the temperature increases, the position and height of the off-zero peak both increase, implying more robust sustained oscillations. The position of the off-zero peak is very close to the cell

cycle frequency  $f$ , which can be easily estimated from the data of doubling times (Table III).

We find that the doubling time data for all cell lineages are well fitted by an Erlang distribution; the parameter  $N$  can then be estimated as the inverse of the coefficient of the variation squared of this distribution. The medians of the estimated  $N$  for the three growth conditions are 14, 31, and 31, respectively. It can be seen that cells cultured at 27 °C and 37 °C have much smaller cell cycle duration variability than those cultured at 25 °C. Recent studies have suggested the accumulation of certain proteins up to a critical threshold as a putative mechanism in bacterial and yeast cell division [74–77]. Under this mechanism, the multiple cell cycle stages can be interpreted as different levels of the “division protein,” and thus, the number of stages  $N$  can be interpreted as the critical threshold of the protein. The phenomenon that 27 °C and 37 °C correspond to a larger  $N$  is probably because at higher temperatures, a higher level of the division protein is needed for cell division to occur.

Our theory shows that smaller cell cycle duration variability gives rise to a higher off-zero spectral peak [Fig. 2(c)]. This explains why the off-zero peak is significantly higher at higher temperatures [Fig. 4(a)]. Our theoretical result also predicts that when  $N$  is large, the power spectrum has peaks at higher-order harmonics of the cell cycle frequency [Fig. 2(f)]. This is in excellent agreement with our data analysis, which shows that the average power spectra at 27 °C and 37 °C have an apparent peak at the second harmonic frequency. Interestingly, the ratio of the heights of the spectral peaks at  $f$  and  $2f$  is estimated to be 17.8 and 15.0 for the two temperatures, both of which are very close to the theoretical value of  $2^4 = 16$  predicted by our theory for stable gene products when  $w$  is small.

To perform a more detailed analysis, we also estimate the power spectrum for each cell lineage by fitting the time course data to an autoregressive (AR) model, which is a standard model in time series analysis [78], with the order of the AR model being determined by minimizing the Akaike information criterion. Our theory predicts that for stable gene products, only type I–III spectra can be observed [Figs. 3(a)–3(c)]. This is in full agreement with our data analysis with the percentages of the three types of spectra being illustrated by the pie charts in Fig. 4(a). Clearly, type II spectra are dominant for all the three temperatures. For cells at 25 °C, only types I and II are observed; for cells at higher temperatures, only types II and III are observed. The percentage of type III spectra is significantly higher for cells at 37 °C than cells at 27 °C.

Since the burst production rate increases from  $\rho$  to  $\kappa\rho$  upon replication, we can fit the data (recorded per minute) between two cell division times by the following mean-field approximation:

$$\hat{n}(t+1) - \hat{n}(t) = \begin{cases} \rho B, & \text{if } t - T_k \leq w(T_{k+1} - T_k), \\ \kappa\rho B, & \text{if } t - T_k > w(T_{k+1} - T_k), \end{cases} \quad (20)$$

where  $T_k$  is the  $k$ th cell division time, and  $t$  is an arbitrary time point (in minutes) between two consecutive cell division times  $T_k$  and  $T_{k+1}$ . Here we use a piecewise linear function to approximate the time series of protein levels between two division events. The first (second) part of the piecewise linear function approximates the time series before (after) replication and hence should have an average slope of  $\rho B$  ( $\kappa\rho B$ ). By fitting the time course data  $n(t)$  to the approximation  $\hat{n}(t)$ , we obtain the least-squares estimates of  $w$  and  $\kappa$  for each cell lineage by minimizing the (squared) distance  $\sum_{t=0}^{M-1} [n(t) - \hat{n}(t)]^2$  between the two, where  $M$  is the number of time points measured (see Appendix G for details). The medians of the estimated  $w$  for the three growth conditions are 0.20, 0.15, and 0.12, respectively, which decrease with the temperature. This is possibly because at higher temperatures, the growth rate of *E. coli* cells is faster, and thus, gene replication would occur more continuously, giving rise to a shift in  $w$  toward zero [79]. In addition, the medians of the estimated  $\kappa$  are close to 2 for all three temperatures (Table III), implying weak dosage compensation.

The remaining parameters to be estimated are  $\rho$  and  $B$ , where  $B$  is the mean translational burst size, i.e., the average number of protein molecules produced per mRNA lifetime. To estimate them, recall that  $\rho B$  represents the mean number of protein molecules produced per unit time. However, what is measured in the dataset is the fluorescence intensity of protein molecules instead of the real copy number. Hence, it is crucial to determine the proportionality constant between fluorescence intensities and copy numbers. To do this, we note that under the assumption of symmetric binomial partitioning at division, the expression levels just before and just after a particular cell division time are coupled by (see Appendix H for the proof)

$$\gamma_a = \frac{\gamma_b}{2} + \frac{\beta}{2}, \quad (21)$$

where  $\gamma_b$  ( $\gamma_a$ ) is the Fano factor (the variance divided by the mean) of the fluorescence intensities just before (after) division, and  $\beta$  is the fluorescence intensity per protein copy. Both  $\gamma_a$  and  $\gamma_b$  can be easily estimated for each cell lineage by using the data at birth and division times. To test this relationship, we show  $\gamma_a$  as a function of  $\gamma_b$  for all cell lineages at the three temperatures [see Fig. 4(b) for cells at 37 °C and Fig. S6(a) in the Supplemental Material [66] for cells at lower temperatures], from which we observe a strong linear relationship with a high  $R^2$  and a slope close to 0.5. Then the proportionality constant  $\beta$  for a given temperature can be estimated as  $\beta = 2\langle\gamma_a\rangle - \langle\gamma_b\rangle$ , where the angled brackets denote the sample means over all cell lineages at that temperature. The estimated  $\beta$  for the three growth conditions are 42.1, 15.5, and 13.0, respectively. In particular, our analysis shows that the fluorescence

intensity per protein molecule decreases with the temperature, which is likely because lower temperatures are more conducive to the correct folding of the fluorescent protein [80].

From the fluorescence intensity data and the inferred  $\beta$ , it is easy to estimate the gene expression mean  $\langle n \rangle$  for each cell lineage (Table III). Note that cells at 27°C and 37°C have similar cell cycle duration variability with  $N = 31$ , but the latter temperature yields a higher off-zero spectral peak. This is because cells at 37°C have a larger gene expression mean, which augments the height of the off-zero peak. Moreover, since  $\rho$ ,  $B$ ,  $w$ ,  $\kappa$ , and  $\langle n \rangle$  are related by Eq. (10) and we have estimated  $w$  and  $\kappa$ , we can obtain the estimate of  $\rho B$ . Finally, we estimate  $\rho$  and  $B$  separately by equating the heights of the off-zero peaks of the experimental power spectrum obtained using the AR model technique and the theoretical power spectrum determined by Eq. (9) using all estimated parameters. From Table III, we can see that the mean burst frequency (average number of bursts produced per unit time)  $\rho$  increases with temperature approximately linearly, but the mean burst size  $B$  is not temperature dependent.

Thus far, we have shown how, by means of the theoretical expressions of the power spectrum and the gene expression mean, we can estimate all the model parameters for each cell lineage from the time course data (except that the degradation rate  $d$  is assumed to be zero, following Ref. [72]). The accuracy of these estimates is verified in two different ways. First, in Figs. 4(d) and 4(e), we show a good agreement between the experimental and theoretical power spectra for two typical cell lineages. As a second test of the accuracy of parameter inference, we compare the widths of the experimental and theoretical power spectra for all cell lineages [see Fig. 4(c) for cells at 37°C and Fig. S6(b) in the Supplemental Material [66] for cells at lower temperatures]. It can be seen from Fig. 4(c) that the widths for the two spectra show a strong linear relationship with a slope close to 1, a negligible intercept, and an  $R^2$  of 0.75. We emphasize that while the model parameters are not estimated from the full spectrum curve but only from the height of the off-zero peak, the full theoretical spectrum matches the experimental one reasonably well.

### N. Generalization of the model to include cell-size dynamics

Thus far, we have incorporated various cell cycle features, including cell cycle duration variability, gene replication, dosage compensation, and cell division, into the classical stochastic gene expression model for non-growing cells [14] and carried out a frequency domain analysis of lineage measurements. While the model captures a large variety of biological processes, it still does not include the important variable of cell size. Recently, there has been ample evidence [72,81,82] that the amount of growth produced during the cell cycle must be controlled

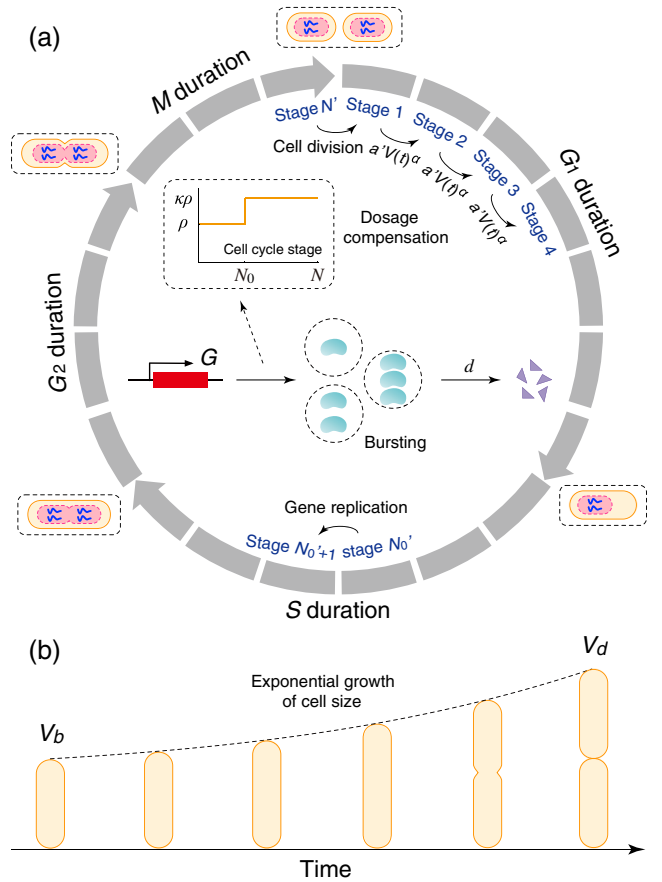


FIG. 5. The model including a cell-size description. (a) Schematic illustrating a detailed model of both gene expression and cell-size dynamics. Each cell can exist in  $N'$  effective cell cycle stages. The transition rate from one stage to the next at a particular time  $t$  is proportional to the  $\alpha$ th power of the cell size  $V(t)$  with  $\alpha > 0$  being the strength of cell-size control and  $a' > 0$  being the proportionality constant. This guarantees that larger cells at birth divide faster than smaller ones to achieve size homeostasis. At stage  $N'$ , a mother cell divides into two daughters that are typically different in size via asymmetric cell division. (b) The cell grows exponentially between birth and division events with the growth rate being a generation-independent constant.

such that, on average, large cells at birth have shorter cell cycle durations than small ones. This mechanism maintains size homeostasis. Many mathematical models have also been proposed to describe cell-size dynamics and control across the cell cycle [83–91].

Next we discuss how to incorporate the important factor of cell size into our model. To this end, we consider a new model which combines the gene expression model illustrated in Fig. 1(a) and the cell-size model investigated in Refs. [88,91] (see Fig. 5 for an illustration). This new model is based on some more assumptions which are listed as follows:

- (i) The size of each cell undergoes deterministic exponential growth in each generation with constant growth rate  $g$ .



- (ii) Each cell can exist in  $N'$  effective cell cycle stages denoted by  $1, 2, \dots, N'$ . The transition rate from one stage to the next at a particular time is proportional to the  $\alpha$ th power of the cell size at that time, with  $a'$  being the proportionality constant [88]. In other words, the transition rate between stages at time  $t$  is equal to  $a'V(t)^\alpha$ , where  $\alpha > 0$  is the strength of cell-size control and  $V(t)$  is the cell size at that time. Under this assumption, larger cells at birth have larger transition rates between stages and thus, on average, have shorter cell cycle durations and lesser volume changes than smaller ones; in this way, size homeostasis is achieved.

The major difference between our previous model illustrated in Fig. 1(a) (model I) and the current model (model II) is the transition rule between stages. The transition rate is assumed to be a constant for the former model, while it is a power law of cell size for the latter model. We stress that while this power law is compatible with a biophysical mechanism where division occurs when a certain type of protein accumulates up to a critical threshold (in this case, the multiple cell cycle stages could be interpreted as different levels of the “division protein”) [88,91], it could also simply be understood as a phenomenological means to model size homeostasis. Note that the parameters  $N$ ,  $N_0$ , and  $a$  in model I are different from the parameters  $N'$ ,  $N'_0$ , and  $a'$  in model II.

Let  $V_b$  and  $V_d$  denote the cell sizes at birth and at division in a generation, respectively. Under the assumption of power-law form of the transition rate, the increment in the  $\alpha$ th power of the cell size across the cell cycle,  $V_d^\alpha - V_b^\alpha$ , has an Erlang distribution with shape parameter  $N'$  and mean  $N'\alpha g/a'$  [88]. Interestingly, the model considered here unifies three popular strategies of cell-size control leading to size homeostasis:  $\alpha = 1$  corresponds to the adder (since the added size  $V_d - V_b$  has an Erlang distribution that is independent of the birth size),  $\alpha \rightarrow 0$  corresponds to the timer (since the doubling time  $T$  has an Erlang distribution that is independent of the birth size), and  $\alpha \rightarrow \infty$  corresponds to the sizer (since the  $\alpha$ th power of the division size  $V_d^\alpha$  has an Erlang distribution that is independent of the birth size). Intermediate strategies are naturally obtained for intermediate values of  $\alpha$ ; timerlike control is obtained when  $0 < \alpha < 1$  and sizerlike control is obtained when  $1 < \alpha < \infty$ .

- (iii) The burst production rate  $\rho$  of the gene product is assumed to be independent of cell size. This assumption is supported by experiments in bacteria [92]. In fission yeast and mammalian cells, however, there has been some evidence [9,93] showing that the transcription rate for many genes

may be proportional to cell size; this effect is not considered here.

- (iv) Cell division occurs when the cell transitions from effective stage  $N'$  to the next stage 1. In naturally occurring systems, the partitioning of cell size at division is often asymmetric and stochastic [72]. Let  $V_d$  and  $V'_b$  denote the cell sizes at division and just after division, respectively. At cell division, we assume that the partition ratio  $z = V'_b/V_d$  has a beta distribution with mean  $p$ , whose probability density function is given by

$$f(z) = \frac{1}{B(p\nu, q\nu)} z^{p\nu-1} (1-z)^{q\nu-1}, \quad 0 < z < 1, \quad (22)$$

where  $B$  is the beta function,  $q = 1 - p$ , and  $\nu > 0$  is referred to as the sample size parameter. The reason behind this assumption is that the partition ratio  $V'_b/V_d$  should be a random number between 0 and 1, which is an important property of the beta distribution [94]. For simplicity, we assume that the mean partition ratio for the cell size is the same as the mean partition ratio for the gene product number.

Based on the four assumptions of model II and the six assumptions of model I (except the third one, which assumes constant transition rates between stages), we are able to simulate both the gene expression and cell-size dynamics in a unified model. In our recent work [91], we have computed the exact distribution of the cell cycle duration for model II and found that it can always be well approximated by an Erlang or gamma distribution (see Fig. 4 in Ref. [91]). This implies that model II can be approximated by model I (since model I has an Erlang-distributed cell cycle duration) with the parameters  $N$  and  $a$  for the approximate model being chosen so that the first and second moments of the cell cycle durations for the two models are matching. Moreover, the parameter  $N_0$  for the approximate model can be determined so that the mean replication times for the two models are matching.

To test the accuracy of this approximation, we illustrate the power spectra for model II and its approximation using model I as  $\alpha$  and  $\eta$  vary (Fig. 6). Clearly, when  $\alpha$  is small (timerlike control), the power spectra of both models are in good agreement with each other [Fig. 6(a)]. This is because a small  $\alpha$  leads to an approximately constant transition rate between stages. When  $\alpha$  is moderate or large (adder or sizerlike control), there are some differences between them for small  $\eta$  [Figs. 6(b) and 6(c)]; compared with model I, model II has a power spectrum with a higher off-zero peak, possibly due to strong cell-size control, which is equivalent to a negative feedback loop [72], enhancing the regularity of oscillations. However, when  $\eta$  is large, they are very close to each other, implying that both models lead to similar oscillatory behavior for unstable gene products.

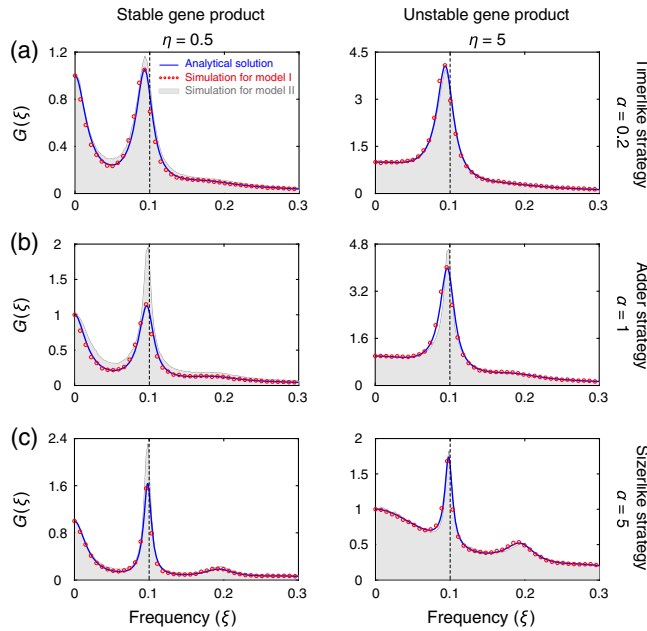


FIG. 6. Power spectra for the model with a cell-size description and its approximation by the model without a cell-size description under different size homeostasis strategies. (a) Timerlike control for stable (left panel) and unstable (right panel) gene products. (b) Adder control. (c) Sizer-like control. In (a)–(c), the gray region represents the simulated spectrum for the model with a cell-size description, the red circles represent the simulated spectrum for the model without a cell-size description, and the blue curve represents the analytical spectrum given in Eq. (9). See Sec. S7 in the Supplemental Material [66] for the technical details of this figure.

In all cases, we can see that the model without cell-size dynamics investigated earlier can indeed capture (with acceptable errors) the oscillatory behavior of the model with cell-size dynamics.

### III. DISCUSSION

In this work, we investigate the frequency decomposition of the copy number fluctuations of a gene product (mRNA or protein) within a cell lineage by deriving expressions for the power spectrum of fluctuations in a detailed stochastic model of gene expression. This model takes into account the salient experimental observations of intracellular dynamics including promoter switching, transcriptional and translational bursting, variability in the duration of the cell cycle, variability in the gene copy number due to replication, the copying or resetting of the gene state during replication, gene dosage compensation, and partitioning of molecules due to symmetric or asymmetric cell division.

Our investigation differs from previous studies in four main respects: (i) Our model is more grounded in biological reality than other models in the literature due to the large number of subcellular and cellular processes that it accounts for, as we mention above. (ii) A number of

studies have derived the distribution of molecule numbers or more commonly the moments for models that have some similarity to ours [15–20]; however, in contrast, here we derive expressions for the power spectrum which provide information about the frequency content of lineage data. This type of analysis has previously been reported only for models of nongrowing cells [21,95,96]. (iii) Rather than a parameter inference based on the matching of the moments or the distribution of molecule numbers calculated from the data to those of a stochastic model [97], we showcase a power-spectrum-based parameter inference method. (iv) Our model takes into account the details of the experimental protocol used for tracking cells across a lineage.

Our novel theory provides expressions for the height and width of the off-zero spectral peak (with its position close to the cell cycle frequency) as a function of all rate parameters in the model. The ratio of the height of this peak to the power at zero frequency provides a means to classify the power spectra into two main types: (i) the spectra with the ratio less than 1 (types I and II), and (ii) the spectra with the ratio greater than 1 (types III and IV). The periodicity in molecule number variation induced by cell division dominates over subcellular noise for (ii) while the reverse is the case for (i). Type I is further differentiated from type II by specifying that in the former type, there is only one peak at zero frequency, whereas in the latter type, there is a dominant peak at zero frequency and a lesser one at approximately the cell cycle frequency. Similarly, type III is further differentiated from type IV by specifying that in the former type, there is a dominant peak at approximately the cell cycle frequency and a lesser one at zero frequency, while in the latter type, there is only one peak at approximately the cell cycle frequency. The theory also predicts that while the spectra for fast decaying (unstable) gene products can be of all four types, the spectra for slowly decaying (stable) gene products can be only of types I–III.

Our analysis of experimental data for *E. coli* shows that the type of spectra of single-cell trajectories depends on the temperature: Lower temperatures favor type I and II spectra, while higher temperatures favor type II and III spectra. Overall, the most common spectrum is type II, implying that for many cells, the “forces” inducing periodicity of molecule numbers are typically slightly less strong than the forces inducing subcellular noise; the strength of the latter increases with decreasing temperature. None of the 279 cell lineages have a type IV spectrum for protein fluctuations, in accordance with the theoretical result that stable gene products cannot display such a spectrum.

Our theory makes a number of other testable predictions: (i) The height of the off-zero spectral peak (relative to the zero peak) increases with decreasing cell cycle duration variability and increasing mean expression levels, while the width of the off-zero peak is proportional to the cell cycle



frequency and to the cell cycle duration variability. (ii) If the cell cycle duration variability is small enough, then the spectra display peaks at higher-order harmonics of the cell cycle frequency. For stable proteins, the height of the spectral peak at the second harmonic (twice the cell cycle frequency) is 16 times larger than that at the cell cycle frequency for early or late replication. Both predictions are confirmed by analysis of lineage data for *E. coli*.

Furthermore, our theory makes numerous other predictions which require the design of new experiments and which cannot be tested using current data. A brief summary of these predictions is as follows: (i) The dependence of the height of the off-zero spectral peak on parameters is very different for gene products that are stable, e.g., most proteins, compared to gene products that are unstable, e.g., most mRNAs. For stable gene products, the height of the off-zero peak decreases with increasing mean burst size, decreasing mean molecule number, and increasing asymmetry of cell division. Furthermore, the height is maximal for replication occurring (almost) a third of the way through the cycle. In contrast, for unstable gene products, the height of the off-zero peak is maximal for replication occurring midway through the cycle but is (almost) independent of asymmetric division, gene expression mean, random bursting, and dosage compensation. Independent of stability, fast promoter switching enhances the height of the off-zero peak (relative to the zero peak), whereas slow switching has the opposite effect; (ii) the strength of gene dosage compensation is reflected in how the spectrum varies with gene product stability. For weak dosage compensation, molecules which have a larger decay rate also exhibit a spectrum with a higher off-zero peak. For strong dosage compensation, molecules which have a decay rate that is neither too large nor too small are the ones which exhibit the most pronounced off-zero peaks. (iii) The power spectrum in asymmetrically dividing cells is strongly influenced by the choice of single-cell tracking protocol. In particular, the spectra obtained from following the smaller daughter after division have a significantly higher off-zero peak than those obtained from following the larger daughter or from random tracking. This last prediction is likely the easiest of the three to check by redoing the cell tracking analysis in budding yeast which displays asymmetric cell division. Predictions (i) and (ii) ideally require the ability to obtain live-cell fluorescence data for mRNA over tens of generations and for different types of mRNAs with widely varying decay rates, which are presently not easy to obtain. The development of such methods, particularly those with minimal perturbation of transcription and translation, is still in an active area of research [4].

To make our model closer to biological reality, we also incorporate cell-size dynamics into our gene expression model by assuming that (i) the cell grows exponentially between birth and division events with the growth rate being a generation-independent constant, (ii) stochastic partitioning of cell size at division, (iii) the rate of moving

from one cell cycle stage to the next is a power law of cell size, and (iv) the burst production rate of the gene product is independent of cell size. By matching the moments of the doubling time and replication time, we approximate the complex model by a model without a cell-size description. We find that both models display similar oscillatory behavior for unstable products. For stable products, explicit modeling of size homeostasis results in a slightly higher regularity of oscillations for adder and sizerlike strategies.

We also showcase the use of the power spectrum to determine the values of all the rate parameters in our model from cell lineage data. The two strengths of our inference procedure are (i) we do not need an experimental determination of the relationship between total fluorescence intensity and molecule number; this is determined automatically from the relationship between fluctuations in the fluorescence intensity just before and just after division; (ii) Our analysis takes into account noise due to partitioning and variability in the cell cycle duration, which has recently been shown to be crucial to obtain an accurate estimation of the burst frequency and the burst size [20]. Our analysis confirms that there is no or very weak dosage compensation in *E. coli*, in agreement with previous studies [79]; i.e., at replication, the expression increases roughly twofold, in agreement with an expected doubling of the number of gene copies. Furthermore, we find that for constitutive expression, while the translational burst frequency increases approximately linearly with the temperature, the translational burst size is temperature invariant. Previous studies were conducted at one temperature [5] and hence could not quantify the thermal dependence of gene expression parameters.

Concluding, we perform an exact frequency analysis of mRNA and protein fluctuations in a detailed model of gene circuit elements central to gene expression control. As we show, there is a wealth of information about subcellular processes hidden in the frequency content of mRNA and protein fluctuations within a cell lineage, and we hope our results will further stimulate work in this area.

## ACKNOWLEDGMENTS

We thank Ruben Perez-Carrasco, Irina Kalita, Abhyudai Singh, and Jiajun Zhang for stimulating discussions. C. J. acknowledges support from the NSAF grant of the National Natural Science Foundation of China (Grant No. U1930402). R. G. acknowledges support from the Leverhulme Trust (Grant No. RPG-2018-423).

## APPENDIX A: MODEL WITH PROMOTER SWITCHING

Here we consider a more detailed model of stochastic gene expression dynamics across the cell cycle that includes promoter switching [see Fig. 7(a) for an illustration]. When promoter switching is taken into account, we need some more assumptions which are as follows.

The specific meaning of the new parameters involved is listed in Table IV.

- (i) The promoter of the gene of interest can switch between an inactive state  $G_0$  and an active state  $G_1$  with switching rates  $u$  and  $v$  before replication [14]. Dosage compensation is modeled as a change in the switching rate of the promoter from the inactive to the active state upon replication with its value being

$u$  before replication and  $gu$  after replication. This assumption is supported by experiments [8].

- (ii) In each gene state  $G_i$  ( $i = 0, 1$ ), the synthesis of the gene product, mRNA or protein, occurs at a rate  $\rho_i$  in bursts of a random size sampled from an arbitrary distribution  $\mu = (\mu_n)$ . In previous papers [22], the synthesis of mRNA in each gene state is assumed to be nonbursty with  $\mu_n = \delta_{1,n}$ . On the other hand, in

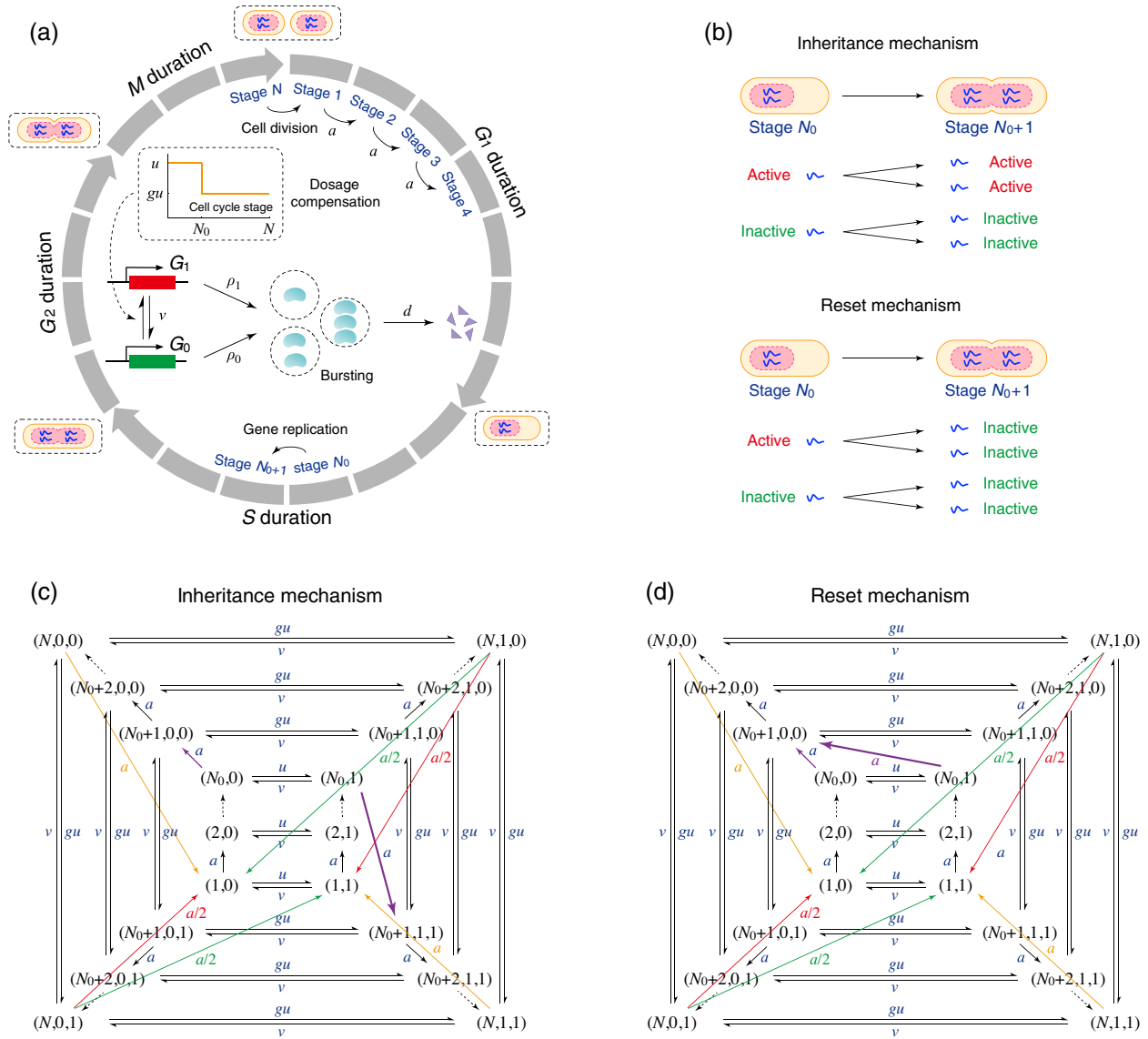


FIG. 7. The model with promoter switching. (a) Schematic illustrating the model describing  $N$  effective cell cycle stages, gene replication at stage  $N_0$ , promoter switching between active (red) and inactive (green) states, bursty production of the gene product in the two gene states, degradation, and gene dosage compensation induced by a decrease in the activation rate of the gene after replication (see inset graph). (b) At replication, the gene states of the two daughter copies can be the same as that of the mother copy (inheritance mechanism) or else they are both reset to the inactive state (reset mechanism). (c),(d) Transition diagram of cellular states under the two mechanisms. Before replication, a cellular state can be represented by an ordered pair  $(r, i)$ , where  $r$  is the cell cycle stage, and  $i$  is the gene state of the mother copy; after replication, a cellular state can be represented by an ordered triple  $(r, i, j)$ , where  $i$  and  $j$  are the gene states of the two daughter copies. See the main text for explanation of the colored arrows. The difference between (c) and (d) is that the cellular state transitions from  $(N_0, 1)$  to  $(N_0 + 1, 1, 1)$  in (c), while it transitions from  $(N_0, 1)$  to  $(N_0 + 1, 0, 0)$  in (d) [see the arrows in (c) and (d) that are shown in bold].

TABLE IV. Model parameters that are not included in Table I and their meaning.

	Meaning of model parameters
$u$	Gene switching rate from OFF to ON before replication
$gu$	Gene switching rate from OFF to ON after replication
$v$	Gene switching rate from ON to OFF
$\rho_0$	Burst production rate when the gene is OFF
$\rho_1$	Burst production rate when the gene is ON

models where mRNA is not explicitly described, the effective synthesis of protein is usually assumed to be bursty with the burst size sampled from a geometric distribution  $\mu_n = p_B^n(1 - p_B)$ ; this is due to rapid synthesis of protein molecules from a single short-lived mRNA molecule [25,26]. Therefore, the arbitrariness of the burst size distribution allows us to analyze the dynamics of both mRNA and protein in a unified model.

- (iii) After replication, the two daughter copies can either inherit the gene state from the mother copy or both be reset to the inactive state [19]. To distinguish between them, we refer to the former as the inheritance mechanism and the latter as the reset mechanism [Fig. 7(b)]. The consideration behind the former mechanism is the copying of the landscape of histone modifications (implicated in gene activation) during DNA replication [98]. One plausible explanation for the latter mechanism is that to avoid the potential risk of conflict between replication and transcription [99], it is likely that in the region where replication is ongoing or just completed, there is no transcription, indicating an inactive state.

We next describe our stochastic model for haploid cells. The results for diploid cells can be easily deduced from the haploid case using allelic expression independence (see Appendix B for details). Based on the stage of the cell cycle and the states of the gene copies, there are many possible cellular states. Before replication, the cell can exist in  $2N_0$  cellular states that can be represented by an ordered pair  $\alpha = (r, i)$ , where  $r \in [1, N_0]$  is the cell cycle stage and  $i = 0, 1$  is the gene state of the mother copy; after replication, the cell can exist in  $4(N - N_0)$  cellular states that can be represented by an ordered triple  $\alpha = (r, i, j)$ , where  $r \in [N_0 + 1, N]$  is the cell cycle stage and  $i, j = 0, 1$  are the gene states of the two daughter copies. In sum, there is a total of  $K = 4N - 2N_0$  possible cellular states.

The transitions among the  $K$  cellular states can be described by a Markovian model whose transition diagram is illustrated in Fig. 7(c) for the inheritance mechanism and Fig. 7(d) for the reset mechanism. The purple arrows show that upon replication, the cellular state will transition from  $(N_0, i)$  to  $(N_0 + 1, i, i)$  for the inheritance mechanism and from  $(N_0, i)$  to  $(N_0 + 1, 0, 0)$  for the reset mechanism. After division, we randomly track one of the two daughters,

and thus, the cell will transition from  $(N, i, j)$  to  $(1, i)$  or  $(1, j)$ . The orange arrows illustrate those transitions from  $(N, i, i)$  to  $(1, i)$  with rate  $a$ , the red arrows illustrate those from  $(N, i, j), i \neq j$  to  $(1, i)$  with rate  $a/2$ , and the green arrows illustrate those from  $(N, i, j), i \neq j$  to  $(1, j)$  with rate  $a/2$ .

The microstate of the gene of interest can be represented by an ordered pair  $(\alpha, n)$ , where  $\alpha$  is the cellular state and  $n$  is the copy number of the gene product. The evolution of the complete stochastic gene expression dynamics is governed by a master equation which is given in Sec. S3 in the Supplemental Material [66]. The analytical solutions of the autocorrelation function and power spectrum for the current model can be found in Appendix D.

A special case occurs when promoter switching is much faster than cell cycle progression and gene product degradation, i.e.,  $f, d \ll u, v$  [37]. In this case, gene switching dynamics will reach rapid equilibrium. As a result, the effective burst production rate is given by  $\rho_{\text{eff}} = (\rho_1 u + \rho_0 v)/(u + v)$  before replication and it is given by  $\kappa \rho_{\text{eff}} = 2(\rho_1 gu + \rho_0 v)/(gu + v)$  after replication, where  $\kappa \geq 1$  is a factor characterizing the change in the burst production rate due to gene replication and dosage compensation. In the absence of dosage compensation, we have  $gu = u$ , and thus,  $\kappa = 2$ . In the fast switching regime, the two gene states can be combined into a single one, and hence, the model with promoter switching considered here reduces to our previous model without promoter switching. In this case, we do not need to distinguish between the inheritance and reset mechanisms because they lead to the same oscillatory behavior.

## APPENDIX B: OSCILLATIONS FOR DIPLOID CELLS

Suppose that we have computed the analytical solutions of the autocorrelation function  $R(t)$  and the power spectrum  $G(\xi)$  for haploid cells. For diploid cells, we assume that the two alleles act independently of each other. Under the assumption of allelic expression independence, let  $n(t)$  and  $n'(t)$  denote the expression levels of the two alleles at time  $t$ , which are independent and identically distributed. Then for diploid cells, the autocorrelation function is given by

$$\begin{aligned} R_{\text{diploid}}(t) &= \text{Cov}_{\text{SS}}[n(0) + n'(0), n(t) + n'(t)] \\ &= 2\text{Cov}_{\text{SS}}[n(0), n(t)] = 2R(t), \end{aligned}$$

and thus, the power spectrum is given by

$$G_{\text{diploid}}(\xi) = 2G(\xi).$$

Since the two functions differ only by a constant for haploid and diploid cells, they lead to the same oscillatory behavior. Thus, we need to focus only on haploid cells in what follows.

**APPENDIX C: DERIVATION OF THE GENE  
EXPRESSION MEAN, AUTOCORRELATION  
FUNCTION, AND POWER SPECTRUM  
FOR THE MODEL WITHOUT  
PROMOTER SWITCHING**

Let  $r(t)$  denote the cell cycle stage and let  $n(t)$  denote the copy number of the gene product in an individual cell at time  $t$ , respectively. To proceed, let

$$\begin{aligned} m_{0r}(t) &= \sum_{n=0}^{\infty} p_{r,n} = \mathbb{P}(r(t) = r), \\ m_{1r}(t) &= \sum_{n=0}^{\infty} n p_{r,n} = \mathbb{E}n(t) I_{\{r(t)=r\}}, \\ m_{2r}(t) &= \sum_{n=0}^{\infty} n(n-1) p_{r,n} = \mathbb{E}n(t)(n(t)-1) I_{\{r(t)=r\}} \end{aligned}$$

be the zeroth, first, and second factorial moments of the gene product number in cell cycle stage  $r$ , where  $I_A$  denotes the indicator function of the set  $A$ . Let  $m_k = (m_{k\alpha})$  denote the row vector whose components are the  $k$ th factorial moments in all cell cycle stages. It then follows from Eq. (3) that  $m_0$ ,  $m_1$ , and  $m_2$  satisfy the following differential equations:

$$\begin{aligned} \dot{m}_0(t) &= m_0(t)W_{00}, \\ \dot{m}_1(t) &= m_1(t)W_{11} + m_0(t)W_{01}, \\ \dot{m}_2(t) &= m_2(t)W_{22} + m_1(t)W_{12} + m_0(t)W_{02}. \end{aligned} \quad (\text{C1})$$

Moreover, it is easy to check that

$$W_{01} = BS, \quad W_{12} = 2BS, \quad W_{02} = H''(1)S,$$

where  $B = \sum_{n=1}^{\infty} n\mu_n$  is the mean burst size of the gene product, and  $S = \text{diag}(\rho, \dots, \rho, \kappa\rho, \dots, \kappa\rho)$  is the diagonal matrix whose diagonal entries are the burst production rates in all cell cycle stages. From Eq. (C1), at the steady state, the zeroth, first, and second factorial moments are given by

$$\begin{aligned} m_0 &= (1/N, \dots, 1/N), \\ m_1 &= -Bm_0SW_{11}^{-1}, \\ m_2 &= -[2Bm_1 + H''(1)m_0]SW_{22}^{-1}. \end{aligned} \quad (\text{C2})$$

In the nonbursty case, it is easy to check that  $H''(1) = 0$ ; in the bursty case, it is easy to check that  $H''(1) = 2B^2$ . Therefore, the steady-state gene expression mean is given by

$$\langle n \rangle = m_1 \mathbb{1} = -Bm_0SW_{11}^{-1} \mathbb{1}.$$

Since we know the expressions of  $m_0$ ,  $S$ , and  $W_{11}$ , straightforward calculations show that

$$\langle n \rangle = \frac{\rho B}{\eta f} \left[ w + \kappa(1-w) + \frac{(\kappa-1)a_{\eta,N}^w + 1 - \kappa a_{\eta,N}}{\eta(2a_{\eta,N} - 1)} \right].$$

This gives the analytical expression of the gene expression mean.

Since Eq. (C1) is a set of linear differential equations, its time-dependent solution is given by

$$m_1(t) = m_1(0)e^{W_{11}t} + B \int_0^t m_0(0)e^{W_{00}s} S e^{W_{11}(t-s)} ds.$$

From now on, we assume that the system has reached the steady state. Given the initial cell cycle stage  $r(0) = r$  and initial copy number  $n(0) = n$  of the gene product, it follows that

$$\begin{aligned} \mathbb{E}n(t) &= m_1(t) \mathbb{1} \\ &= n e_r e^{W_{11}t} \mathbb{1} + B \int_0^t e_r e^{W_{00}s} S e^{W_{11}(t-s)} \mathbb{1} ds, \end{aligned}$$

where  $\mathbb{1}$  denotes the column vector whose components are all 1, and  $e_r$  denotes the row vector whose  $r$ th component is 1 and all other components are all 0. This clearly shows that

$$\begin{aligned} \mathbb{E}[n(t)|r(0), n(0)] \\ &= n(0)e_{r(0)}e^{W_{11}t} \mathbb{1} + B \int_0^t e_{r(0)}e^{W_{00}s} S e^{W_{11}(t-s)} \mathbb{1} ds. \end{aligned}$$

Therefore, at the steady state, we have

$$\begin{aligned} \mathbb{E}n(0)n(t) &= \sum_r \mathbb{E}n(0) I_{\{r(0)=r\}} \mathbb{E}[n(t)|\alpha(0), n(0)] \\ &= \sum_r \mathbb{E}n(0) I_{\{r(0)=r\}} [n(0)e_{r(0)}e^{W_{11}t} \mathbb{1} \\ &\quad + B \int_0^t e_{r(0)}e^{W_{00}s} S e^{W_{11}(t-s)} \mathbb{1} ds] \\ &= \sum_r \mathbb{E}n(0)^2 I_{\{r(0)=r\}} e_r e^{W_{11}t} \mathbb{1} \\ &\quad + B \int_0^t \sum_r \mathbb{E}n(0)^2 I_{\{r(0)=r\}} e_r e^{W_{00}s} S e^{W_{11}(t-s)} \mathbb{1} ds \\ &= \sum_r (m_{1r} + m_{2r}) e_r e^{W_{11}t} \mathbb{1} \\ &\quad + B \int_0^t m_{1r} e_r e^{W_{00}s} S e^{W_{11}(t-s)} \mathbb{1} ds \\ &= (m_1 + m_2) e^{W_{11}t} \mathbb{1} + B \int_0^t m_1 e^{W_{00}s} S e^{W_{11}(t-s)} \mathbb{1} ds, \end{aligned}$$

where  $m_1$  and  $m_2$  are the steady-state first and second factorial moments given in Eq. (C2). Since the autocorrelation function is defined as  $R(t) = \mathbb{E}n(0)n(t) - \mathbb{E}n(0)\mathbb{E}n(t)$ , we finally obtain the explicit expression of the autocorrelation function, which is given in Eq. (5).



We next compute the analytical expression of the power spectrum. Recall that all the eigenvalues  $\lambda_0, \dots, \lambda_{2N-1}$  of  $W_{00}$  and  $W_{11}$  have been computed in Eqs. (7) and (8). Since  $W_{00}$  is a circulant matrix, it is easy to check that there exists a complex orthogonal matrix

$$R = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \omega_0 & \omega_1 & \cdots & \omega_{N-1} \\ \cdots & \cdots & \cdots & \cdots \\ \omega_0^{N-1} & \omega_1^{N-1} & \cdots & \omega_{N-1}^{N-1} \end{pmatrix}$$

such that  $W_{00}$  is diagonalized, i.e.,

$$\bar{R}' W_{00} R = \begin{pmatrix} \lambda_0 & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_{N-1} \end{pmatrix},$$

where  $\bar{R}'$  denotes the conjugate transpose of  $R$ . Similarly,  $W_{11}$  can be diagonalized as

$$\bar{R}' M^{-1} W_{11} M R = \begin{pmatrix} \lambda_N & & & \\ & \lambda_{N+1} & & \\ & & \ddots & \\ & & & \lambda_{2N-1} \end{pmatrix},$$

where  $M$  is a diagonal matrix, which is given by

$$M = \begin{pmatrix} 1 & & & \\ & 2^{-1/N} & & \\ & & \ddots & \\ & & & 2^{-(N-1)/N} \end{pmatrix}.$$

With these notations, it then follows from Eq. (5) that all the coefficients  $u_k$ ,  $0 \leq k \leq N-1$  associated with the eigenvalues of  $W_{00}$  are given by

$$u_k = [(m_1 + m_2)MR]_k [\bar{R}' M^{-1} \mathbb{1}]_k - B \sum_{i=1}^N \frac{[m_1 R]_i [\bar{R}' S M R]_{ik} [\bar{R}' M^{-1} \mathbb{1}]_k}{\lambda_{i-1} - \lambda_{N+k-1}}, \quad (\text{C3})$$

and all the coefficients  $u_k$ ,  $N \leq k \leq 2N-1$  associated with the eigenvalues of  $W_{11}$  are given by

$$u_k = B \sum_{j=1}^N \frac{[m_1 R]_k [\bar{R}' S M R]_{kj} [\bar{R}' M^{-1} \mathbb{1}]_j}{\lambda_{k-1} - \lambda_{N+j-1}}. \quad (\text{C4})$$

Note that if the cell cycle frequency is increased from  $f$  to  $\alpha f$  with some  $\alpha > 1$ , while keeping  $\eta = d/f$ ,  $N$ , and  $\langle n \rangle$  as fixed, then the parameters  $d$ ,  $a$ , and  $\rho$  will become  $\alpha d$ ,  $\alpha a$ , and  $\alpha \rho$ , respectively. It then follows from Eqs. (7) and (8) that all the eigenvalues  $\lambda_k$ ,  $0 \leq k \leq 2N-1$  will become  $\alpha \lambda_k$ . In addition, it follows from Eq. (C2) that both  $m_1$  and  $m_2$  are invariant under the above transformation. Since  $S$  becomes  $\alpha S$  under the above transformation, it follows from Eqs. (C3) and (C4) that all the coefficients  $u_k$ ,  $0 \leq k \leq 2N-1$  will remain the same. These results show that the power spectrum (autocorrelation function) will be stretched (compressed) along the horizontal axis by a factor of  $\alpha$ .

#### APPENDIX D: ANALYTICAL EXPRESSIONS OF THE AUTOCORRELATION FUNCTION AND POWER SPECTRUM FOR THE MODEL WITH PROMOTER SWITCHING

While the model with promoter switching is complex due to the large variety of biological processes that it captures, its autocorrelation function can still be computed exactly as

$$R(t) = (m_1 + m_2) e^{W_{11} t} \mathbb{1} - (m_1 \mathbb{1})^2 + B \int_0^t m_1 e^{Qs} S e^{W(t-s)} \mathbb{1} ds, \quad (\text{D1})$$

where  $m_1 = (m_{1\alpha})$  and  $m_2 = (m_{2\alpha})$  are two row vectors whose components are the first and second factorial moments of the gene product number in all cellular states,  $\mathbb{1}$  is the column vector whose components are all 1,  $S = \text{diag}(\rho_\alpha)$  is the diagonal matrix whose diagonal elements are the burst production rates in all cellular states,  $Q = (q_{\alpha\beta})$  is the generator matrix of cellular state transitions [see Figs. 7(c) and 7(d) for the transition diagram which has some orange, red, and green arrows], and  $W_{11} = (w_{\alpha\beta})$  is another matrix obtained from  $Q$  by replacing the rates of the orange arrows by  $a/2$ , replacing the rates of the red arrows by  $pa/2$ , replacing the rates of the green arrows by  $qa/2$ , and subtracting  $d$  from the diagonal entries (see Sec. S4 in the Supplemental Material [66] for the proof and the detailed expressions of each term).

Note that the autocorrelation function  $R(t)$  in Eq. (D1) is expressed in matrix form. A more explicit expression can be obtained by expanding the matrix exponentials  $e^{W_{11} t}$  and  $e^{Qs}$  in terms of their eigenvalues and eigenvectors. We find that the autocorrelation function (power spectrum) can be rewritten as the linear combination of  $2K-1$  exponential (Lorentzian) functions:

$$R(t) = \sum_{k=1}^{2K-1} u_k e^{\lambda_k t}, \quad G(\xi) = \sum_{k=1}^{2K-1} \frac{-2u_k \lambda_k}{4\pi^2 \xi^2 + \lambda_k^2},$$

where  $K$  is the number of cellular states,  $\lambda_0, \dots, \lambda_{2K-1}$  are all the eigenvalues of the two matrices  $Q$  and  $W_{11}$ , and  $u_k$  are



suitable constants (see Sec. S4 in the Supplemental Material [66] for the proof and the specific expressions of  $u_k$ ). Since  $Q$  and  $W_{11}$  are both  $K \times K$  matrices, they have a total of  $2K$  eigenvalues.

### APPENDIX E: POWER SPECTRUM FOR STABLE GENE PRODUCTS

For stable gene products with  $\eta \ll 1$ , the power spectrum can be simplified as in Eq. (13), where

$$A = 2(\ln 2)^2 J \langle n \rangle \left( \gamma + \frac{2}{3N} \right) + \frac{1}{(\ln 2)^2} + \frac{2B}{3 \ln 2} \left( \gamma - 2 + \frac{3}{\ln 2} \right)$$

if the gene expression is bursty,

$$A = 2(\ln 2)^2 J \langle n \rangle \left( \gamma + \frac{2}{3N} \right) + \frac{1}{(\ln 2)^2}$$

if the gene expression is nonbursty, and

$$H_k = \left[ w + \kappa(1-w) + \frac{(\kappa-1) \sin(2k w \pi)}{2k\pi} \right]^2 + \frac{(\kappa-1)^2 \sin^4(k w \pi)}{k^2 \pi^2}$$

is a function of  $\kappa$  and  $w$ , where  $J$  is given in Eq. (16).

### APPENDIX F: POWER SPECTRUM FOR UNSTABLE GENE PRODUCTS

For unstable gene products with  $\eta \ll 1$ , when cell cycle duration variability is not too large, the power spectrum can be written more explicitly as (see Sec. S2 in the Supplemental Material [66] for the proof)

$$G(\xi) \approx \frac{2(\kappa-1)^2 \rho^2 B^2 N \sin^2(k w \pi)}{\eta^2 f} \sum_{k=1}^{[N/2]} G_k(\xi), \quad (\text{F1})$$

where

$$G_k(\xi) = \frac{2k^2 \pi^2 f^2 + (\pi^2 \xi^2 - a^2 \cos \theta_k \sin^2 \theta_k / 2)}{4k^6 \pi^6 f^4 + N^2 (\pi^2 \xi^2 - a^2 \cos \theta_k \sin^2 \theta_k / 2)^2},$$

with  $\theta_k = 2k\pi/N$ . Here the function  $G_1(\xi)$  controls the first off-zero peak, and the functions  $G_k(\xi)$ ,  $k \geq 2$  control the peaks at higher-order harmonic frequencies. In analogy to the case of stable gene products, the position of the off-zero peak is given by  $\xi = (a/\pi) \cos \theta_1 \sin(\theta_1/2) < a\theta_1/2\pi = a/N = f$ , which is smaller than the cell cycle frequency  $f$ . When  $N \gg 1$ , the peak position is approximately equal to  $f$  since  $\sin \theta \approx \theta$  and  $\cos \theta \approx 1$  when  $\theta$  is small. Moreover, the width of the off-zero peak is given by  $D = 2\pi f/N$ , and the

height of the off-zero peak is given in Eq. (17). From Eq. (F1), the height and width of the spectral peak at the  $k$ th harmonic frequency are given by  $\sin^2(k w \pi) / \sin^2(w \pi) k^4 H$  and  $k^2 D$ , respectively. In particular, the height of the spectral peak at  $f$  is  $2^4 \sin^2(w \pi) / \sin^2(2w \pi)$  times greater than that at  $2f$  and  $3^4 \sin^2(w \pi) / \sin^2(3w \pi)$  times greater than that at  $3f$ . Moreover, the width of the spectral peak at  $f$  is  $2^2 = 4$  times lesser than that at  $2f$  and  $3^2 = 9$  times lesser than that at  $3f$ .

### APPENDIX G: ESTIMATION OF $w$ AND $\kappa$ FOR EACH CELL LINEAGE

For a single-cell lineage, let  $n(t)$  be the time course data at time  $t$  and  $T_k$  be the  $k$ th division time. Since the burst production rate increases from  $\rho$  to  $\kappa\rho$  upon replication, we can fit the data between two division times by the mean-field approximation given in Eq. (20). In other words, we use a piecewise linear function to approximate the gene expression levels between two division events. Since  $w$  represents the proportion of cell cycle before replication, the replication time between  $T_k$  and  $T_{k+1}$  is on average given by  $T_k + w(T_{k+1} - T_k)$ . Thus, the first part of the piecewise linear approximation is given by

$$\hat{n}(t) = n(T_k) + \rho B(t - T_k)$$

for any  $T_k \leq t \leq T_k + w(T_{k+1} - T_k)$ , and the second part of the piecewise linear approximation is given by

$$\hat{n}(t) = n(T_k) + \rho B w (T_{k+1} - T_k) + \kappa \rho B [t - T_k - w(T_{k+1} - T_k)]$$

for any  $T_k + w(T_{k+1} - T_k) < t \leq T_{k+1}$ , where  $n(T_k)$  is the gene expression level at time  $T_k$ . Then the distance between the time course data  $n(t)$  and the piecewise linear approximation  $\hat{n}(t)$  is given by

$$D = \sqrt{\sum_{t=0}^{M-1} [n(t) - \hat{n}(t)]^2}, \quad (\text{G1})$$

where  $M$  is the number of time points measured. When  $\eta = 0$ , it follows from Eq. (10) and l'Hôpital's rule that

$$\langle n \rangle = \frac{\rho B}{2f} \left[ 3\kappa - (\kappa-1)w(4-w) + \frac{1}{N}(\kappa w + 1 - w) \right].$$

Moreover, the sample mean of the time course data can be estimated as

$$\langle n \rangle = \frac{1}{M} \sum_{t=0}^{M-1} n(t).$$

Therefore,  $\rho B$  can be represented using  $w$  and  $\kappa$  as

$$\rho B = \frac{2f\langle n \rangle}{3\kappa - (\kappa - 1)w(4 - w) + \frac{1}{N}(\kappa w + 1 - w)}.$$

Finally, the parameters  $w$  and  $\kappa$  can be estimated by minimizing the distance  $D$  in Eq. (G1) using the least-squares criterion.

### APPENDIX H: ESTIMATION OF THE FLUORESCENCE INTENSITY PER PROTEIN MOLECULE

Suppose that cell division occurs at a particular time. Let  $n_b$  be the protein copy number just before division and let  $n_a$  be the protein copy number just after division. Since we assume symmetric binomial partitioning at cell division,  $n_a$  and  $n_b$  are related by the following equality:

$$\mathbb{P}(n_a = m) = \sum_{n=m}^{\infty} C_{n,m} \left(\frac{1}{2}\right)^n \mathbb{P}(n_b = n). \quad (\text{H1})$$

Therefore, the means of  $n_a$  and  $n_b$  are related by

$$\begin{aligned} \langle n_a \rangle &= \sum_{m=0}^{\infty} m \mathbb{P}(n_a = m) = \sum_{m=0}^{\infty} m \sum_{n=m}^{\infty} C_{n,m} \left(\frac{1}{2}\right)^n \mathbb{P}(n_b = n) \\ &= \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n \mathbb{P}(n_b = n) \sum_{m=0}^n m C_{n,m} \\ &= \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n \mathbb{P}(n_b = n) 2^{n-1} n \\ &= \frac{1}{2} \sum_{n=0}^{\infty} n \mathbb{P}(n_b = n) = \frac{1}{2} \langle n_b \rangle, \end{aligned} \quad (\text{H2})$$

and the second factorial moments of  $n_a$  and  $n_b$  are related by

$$\begin{aligned} \langle n_a(n_a - 1) \rangle &= \sum_{m=0}^{\infty} m(m-1) \mathbb{P}(n_a = m) \\ &= \sum_{m=0}^{\infty} m(m-1) \sum_{n=m}^{\infty} C_{n,m} \left(\frac{1}{2}\right)^n \mathbb{P}(n_b = n) \\ &= \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n \mathbb{P}(n_b = n) \sum_{m=0}^n m(m-1) C_{n,m} \\ &= \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n \mathbb{P}(n_b = n) 2^{n-2} n(n-1) \\ &= \frac{1}{4} \sum_{n=0}^{\infty} n(n-1) \mathbb{P}(n_b = n) = \frac{1}{4} \langle n_b(n_b - 1) \rangle. \end{aligned} \quad (\text{H3})$$

Let  $\mu_b$  be the mean of the fluorescence intensity just before division, let  $\mu_a$  be the mean of the fluorescence intensity just after division, and let  $\beta$  be the fluorescence intensity per protein molecule. It then follows from Eq. (H2) that

$$\mu_a = \beta \langle n_a \rangle = \frac{1}{2} \beta \langle n_b \rangle = \frac{1}{2} \mu_b. \quad (\text{H4})$$

Let  $\sigma_b$  be the standard deviation of the fluorescence intensity just before division and let  $\sigma_a$  be the standard deviation of the fluorescence intensity just after division. Combining Eqs. (H2) and (H3) shows that

$$\begin{aligned} \sigma_a^2 &= \beta^2 [\langle n_a(n_a - 1) \rangle + \langle n_a \rangle - \langle n_a \rangle^2] \\ &= \beta^2 \left[ \frac{1}{4} \langle n_b(n_b - 1) \rangle + \frac{1}{2} \langle n_b \rangle - \frac{1}{4} \langle n_b \rangle^2 \right] \\ &= \frac{1}{4} \sigma_b^2 + \frac{1}{4} \beta \mu_b. \end{aligned} \quad (\text{H5})$$

Let  $\gamma_b = \sigma_b^2 / \mu_b$  be the Fano factor of the fluorescence intensity just before division and let  $\gamma_a = \sigma_a^2 / \mu_a$  be the Fano factor of the fluorescence intensity just after division. Combining Eqs. (H4) and (H5) finally shows that

$$\gamma_a = \frac{1}{2} \gamma_b + \frac{1}{2} \beta.$$

- [1] N. Brenner, E. Braun, A. Yoney, L. Susman, J. Rotella, and H. Salman, *Single-Cell Protein Dynamics Reproduce Universal Fluctuations in Cell Populations*, *Eur. Phys. J. E* **38**, 102 (2015).
- [2] Y. Tanouchi *et al.*, *Long-Term Growth Data of Escherichia coli at a Single-Cell Level*, *Sci. Data* **4**, 170036 (2017).
- [3] M. M. Crane, I. B. Clark, E. Bakker, S. Smith, and P. S. Swain, *A Microfluidic System for Studying Ageing and Dynamic Single-Cell Responses in Budding Yeast*, *PLoS One* **9**, e100042 (2014).
- [4] X. Chen *et al.*, *Visualizing RNA Dynamics in Live Cells with Bright and Stable Fluorescent RNAs*, *Nat. Biotechnol.* **37**, 1287 (2019).
- [5] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, *Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells*, *Science* **329**, 533 (2010).
- [6] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, *Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics*, *Science* **332**, 472 (2011).
- [7] J. R. Peterson, J. A. Cole, J. Fei, T. Ha, and Z. A. Luthey-Schulten, *Effects of DNA Replication on mRNA Noise*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15886 (2015).
- [8] S. O. Skinner, H. Xu, S. Nagarkar-Jaiswal, P. R. Freire, T. P. Zwaka, and I. Golding, *Single-Cell Analysis of Transcription Kinetics across the Cell Cycle*, *eLife* **5**, e12175 (2016).
- [9] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh, and A. Raj, *Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number*

- through Independent Global Transcriptional Mechanisms, *Mol. Cell* **58**, 339 (2015).
- [10] M. Hölzel, F. Kohlhuber, I. Schlosser, D. Hölzel, B. Lüscher, and D. Eick, *Myc/Max/Mad Regulate the Frequency but not the Duration of Productive Cell Cycles*, *EMBO Rep.* **2**, 1125 (2001).
- [11] D. R. Tyson, S. P. Garbett, P. L. Frick, and V. Quaranta, *Fractional Proliferation: A Method to Deconvolve Cell Population Dynamics from Single-Cell Data*, *Nat. Methods* **9**, 923 (2012).
- [12] D. Huh and J. Paulsson, *Random Partitioning of Molecules at Cell Division*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15004 (2011).
- [13] D. Huh and J. Paulsson, *Non-Genetic Heterogeneity from Stochastic Partitioning at Cell Division*, *Nat. Genet.* **43**, 95 (2011).
- [14] V. Shahrezaei and P. S. Swain, *Analytical Distributions for Stochastic Gene Expression*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17256 (2008).
- [15] 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).
- [16] M. Soltani and A. Singh, *Effects of Cell-Cycle-Dependent Expression on Random Fluctuations in Protein Levels*, *R. Soc. Open Sci.* **3**, 160578 (2016).
- [17] R. Perez-Carrasco, C. Beentjes, and R. Grima, *Effects of Cell Cycle Variability on Lineage and Population Measurements of Messenger RNA Abundance*, *J. R. Soc. Interface* **17**, 20200360 (2020).
- [18] R. Dessalles, V. Fromion, and P. Robert, *Models of Protein Production along the Cell Cycle: An Investigation of Possible Sources of Noise*, *PLoS One* **15**, e0226016 (2020).
- [19] Z. Cao and R. Grima, *Analytical Distributions for Detailed Models of Stochastic Gene Expression in Eukaryotic Cells*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4682 (2020).
- [20] C. H. Beentjes, R. Perez-Carrasco, and R. Grima, *Exact Solution of Stochastic Gene Expression Models with Bursting, Cell Cycle and Replication Dynamics*, *Phys. Rev. E* **101**, 032403 (2020).
- [21] M. L. Simpson, C. D. Cox, and G. S. Sayler, *Frequency Domain Analysis of Noise in Autoregulated Gene Circuits*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4551 (2003).
- [22] J. Peccoud and B. Ycart, *Markovian Modeling of Gene-Product Synthesis*, *Theor. Popul. Biol.* **48**, 222 (1995).
- [23] N. Kumar, T. Platini, and R. V. Kulkarni, *Exact Distributions for Stochastic Gene Expression Models with Bursting and Feedback*, *Phys. Rev. Lett.* **113**, 268105 (2014).
- [24] D. Zenklusen, D. R. Larson, and R. H. Singer, *Single-RNA Counting Reveals Alternative Modes of Gene Expression in Yeast*, *Nat. Struct. Mol. Biol.* **15**, 1263 (2008).
- [25] P. Bokes, J. R. King, A. T. Wood, and M. Loose, *Multiscale Stochastic Modelling of Gene Expression*, *J. Math. Biol.* **65**, 493 (2012).
- [26] C. Jia, *Simplification of Markov Chains with Infinite State Space and the Mathematical Theory of Random Gene Expression Bursts*, *Phys. Rev. E* **96**, 032402 (2017).
- [27] Y. Wang, C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown, *Precision and Functional Specificity in mRNA Decay*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5860 (2002).
- [28] C. A. Yates, M. J. Ford, and R. L. Mort, *A Multi-Stage Representation of Cell Proliferation as a Markov Process*, *Bull. Math. Biol.* **79**, 2905 (2017).
- [29] A. Golubev, *Applications and Implications of the Exponentially Modified Gamma Distribution as a Model for Time Variabilities Related to Cell Proliferation and Gene Expression*, *J. Theor. Biol.* **393**, 203 (2016).
- [30] H. X. Chao *et al.*, *Evidence that the Human Cell Cycle Is a Series of Uncoupled, Memoryless Phases*, *Mol. Syst. Biol.* **15** (2019).
- [31] E. Gavagnin *et al.*, *Synchronised Oscillations in Growing Cell Populations are Explained by Demographic Noise*, bioRxiv, <https://doi.org/10.1101/2020.03.13.987032>.
- [32] 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. R. Soc. A* **471**, 20150050 (2015).
- [33] H. D. Summers, P. Rees, M. D. Holton, M. R. Brown, S. C. Chappell, P. J. Smith, and R. J. Errington, *Statistical Analysis of Nanoparticle Dosing in a Dynamic Cellular System*, *Nat. Nanotechnol.* **6**, 170 (2011).
- [34] C. Shi, L. Chao, A. M. Proenca, A. Qiu, J. Chao, and C. U. Rang, *Allocation of Gene Products to Daughter Cells is Determined by the Age of the Mother in Single Escherichia coli Cells*, *Proc. R. Soc. B* **287**, 20200569 (2020).
- [35] L. Robert, J. Ollion, J. Robert, X. Song, I. Matic, and M. Elez, *Mutation Dynamics and Fitness Effects Followed in Single Cells*, *Science* **359**, 1283 (2018).
- [36] Q. Deng, D. Ramsköld, B. Reinius, and R. Sandberg, *Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells*, *Science* **343**, 193 (2014).
- [37] L. A. Sepúlveda, H. Xu, J. Zhang, M. Wang, and I. Golding, *Measurement of Gene Regulation in Individual Cells Reveals Rapid Switching between Promoter States*, *Science* **351**, 1218 (2016).
- [38] J. A. Bernstein, P.-H. Lin, S. N. Cohen, and S. Lin-Chao, *Global Analysis of Escherichia coli RNA Degradosome Function Using DNA Microarrays*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2758 (2004).
- [39] J. A. Bernstein, A. B. Khodursky, P.-H. Lin, S. Lin-Chao, and S. N. Cohen, *Global Analysis of mRNA Decay and Abundance in Escherichia coli at Single-Gene Resolution Using Two-Color Fluorescent DNA Microarrays*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9697 (2002).
- [40] M. A. Moran *et al.*, *Sizing up Metatranscriptomics*, *ISME J.* **7**, 237 (2013).
- [41] D. W. Selinger, R. M. Saxena, K. J. Cheung, G. M. Church, and C. Rosenow, *Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation*, *Genome Res.* **13**, 216 (2003).
- [42] I. Inoue, Y. Wakamoto, and K. Yasuda, *Non-Genetic Variability of Division Cycle and Growth of Isolated Individual Cells in On-Chip Culture System*, *Proc. Jpn. Acad. Ser. B* **77**, 145 (2001).



- [43] Y. Wakamoto, J. Ramsden, and K. Yasuda, *Single-Cell Growth and Division Dynamics Showing Epigenetic Correlations*, *Analyst* **130**, 311 (2005).
- [44] D. Bates and N. Kleckner, *Chromosome and Replisome Dynamics in E. coli: Loss of Sister Cohesion Triggers Global Chromosome Movement and Mediates Chromosome Segregation*, *Cell* **121**, 899 (2005).
- [45] J. R. Peterson, S. Thor, L. Kohler, P. R. A. Kohler, W. W. Metcalf, and Z. Luthey-Schulten, *Genome-Wide Gene Expression and RNA Half-Life Measurements Allow Predictions of Regulation and Metabolic Behavior in Methanosarcina acetivorans*, *BMC Genomics* **17**, 924 (2016).
- [46] S. Di Talia, J. M. Skotheim, J. M. Bean, E. D. Siggia, and F. R. Cross, *The Effects of Molecular Noise and Size Control on Variability in the Budding Yeast Cell Cycle*, *Nature (London)* **448**, 947 (2007).
- [47] B. J. Brewer, E. Chlebowicz-Sledziewska, and W. L. Fangman, *Cell Cycle Phases in the Unequal Mother/Daughter Cell Cycles of Saccharomyces cerevisiae*, *Mol. Cell. Biol.* **4**, 2529 (1984).
- [48] A. Khmelinskii *et al.*, *Tandem Fluorescent Protein Timers for In Vivo Analysis of Protein Dynamics*, *Nat. Biotechnol.* **30**, 708 (2012).
- [49] R. Christiano, N. Nagaraj, F. Fröhlich, and T. C. Walther, *Global Proteome Turnover Analyses of the Yeasts S. cerevisiae and S. pombe*, *Cell Rep.* **9**, 1959 (2014).
- [50] M. Makarova, Y. Gu, J.-S. Chen, J. R. Beckley, K. L. Gould, and S. Olfierenko, *Temporal Regulation of Lipin Activity Diverged to Account for Differences in Mitotic Programs*, *Curr. Biol.* **26**, 237 (2016).
- [51] A. Sveiczer, A. Csikasz-Nagy, B. Györfy, J. J. Tyson, and B. Novak, *Modeling the Fission Yeast Cell Cycle: Quantized Cycle Times in wee1—cdc25Δ Mutant Cells*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7865 (2000).
- [52] D. Siegal-Gaskins and S. Crosson, *Tightly Regulated and Heritable Division Control in Single Bacterial Cells*, *Biophys. J.* **95**, 2063 (2008).
- [53] D. H. Lackner, T. H. Beilharz, S. Marguerat, J. Mata, S. Watt, F. Schubert, T. Preiss, and J. Bühler, *A Network of Multiple Regulatory Layers Shapes Gene Expression in Fission Yeast*, *Mol. Cell* **26**, 145 (2007).
- [54] A. T. Hahn, J. T. Jones, and T. Meyer, *Quantitative Analysis of Cell Cycle Phase Durations and PC12 Differentiation Using Fluorescent Biosensors*, *Cell Cycle* **8**, 1044 (2009).
- [55] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, *Global Quantification of Mammalian Gene Expression Control*, *Nature (London)* **473**, 337 (2011).
- [56] N. Festuccia, I. Gonzalez, and P. Navarro, *The Epigenetic Paradox of Pluripotent ES Cells*, *J. Mol. Biol.* **429**, 1476 (2017).
- [57] L. V. Sharova *et al.*, *Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells*, *DNA Res.* **16**, 45 (2009).
- [58] C. S. Potten and C. Booth, *Keratinocyte Stem Cells: A Commentary*, *J. Invest. Dermatol.* **119**, 888 (2002).
- [59] J. Lim, M. Ha, H. Chang, S. C. Kwon, D. K. Simanshu, D. J. Patel, and V. N. Kim, *Uridylation by TUT4 and TUT7 Marks mRNA for Degradation*, *Cell* **159**, 1365 (2014).
- [60] F. M. Boisvert *et al.*, *A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells*, *Mol. Cell. Proteom.* **11**, M111.011429 (2012).
- [61] E. Eden, N. Geva-Zatorsky, I. Issaeva, A. Cohen, E. Dekel, T. Danon, L. Cohen, A. Mayo, and U. Alon, *Proteome Half-Life Dynamics in Living Human Cells*, *Science* **331**, 764 (2011).
- [62] E. Yang *et al.*, *Decay Rates of Human mRNAs: Correlation with Functional Characteristics and Sequence Attributes*, *Genome Res.* **13**, 1754 (2003).
- [63] W. Chen, J. M. Smeekens, and R. Wu, *Systematic Study of the Dynamics and Half-Lives of Newly Synthesized Proteins in Human Cells*, *Chem. Sci.* **7**, 1393 (2016).
- [64] K. A. Becker, P. N. Ghule, J. A. Therrien, J. B. Lian, J. L. Stein, A. J. van Wijnen, and G. S. Stein, *Self-Renewal of Human Embryonic Stem Cells is Supported by a Shortened G1 Cell Cycle Phase*, *J. Cell. Physiol.* **209**, 883 (2006).
- [65] L. Cai, C. K. Dalal, and M. B. Elowitz, *Frequency-Modulated Nuclear Localization Bursts Coordinate Gene Regulation*, *Nature (London)* **455**, 485 (2008).
- [66] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.11.021032> for the detailed derivation of some important equations and the technical details of some figures.
- [67] N. Friedman, L. Cai, and X. S. Xie, *Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression*, *Phys. Rev. Lett.* **97**, 168302 (2006).
- [68] D. Toner and R. Grima, *Effects of Bursty Protein Production on the Noisy Oscillatory Properties of Downstream Pathways*, *Sci. Rep.* **3**, 2438 (2013).
- [69] C. Jia, M. Q. Zhang, and H. Qian, *Analytic Theory of Stochastic Oscillations in Single-Cell Gene Expression*, *arXiv:1909.09769*.
- [70] P. Thomas, A. V. Straube, J. Timmer, C. Fleck, and R. Grima, *Signatures of Nonlinearity in Single Cell Noise-Induced Oscillations*, *J. Theor. Biol.* **335**, 222 (2013).
- [71] P. Thomas, N. Popović, and R. Grima, *Phenotypic Switching in Gene Regulatory Networks*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6994 (2014).
- [72] Y. Tanouchi, A. Pai, H. Park, S. Huang, R. Stamatov, N. E. Buchler, and L. You, *A Noisy Linear Map Underlies Oscillations in Cell Size and Gene Expression in Bacteria*, *Nature (London)* **523**, 357 (2015).
- [73] C. Zopf, K. Quinn, J. Zeidman, and N. Maheshri, *Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression*, *PLoS Comput. Biol.* **9**, e1003161 (2013).
- [74] K. R. Ghusinga, C. A. Vargas-Garcia, and A. Singh, *A Mechanistic Stochastic Framework for Regulating Bacterial Cell Division*, *Sci. Rep.* **6**, 30229 (2016).
- [75] K. Sekar *et al.*, *Synthesis and Degradation of FtsZ Quantitatively Predict the First Cell Division in Starved Bacteria*, *Mol. Syst. Biol.* **14**, e8623 (2018).
- [76] F. Si, G. Le Treut, J. T. Sauls, S. Vadia, P. A. Levin, and S. Jun, *Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria*, *Curr. Biol.* **29**, 1760 (2019).
- [77] J. O. Patterson, P. Rees, and P. Nurse, *Noisy Cell-Size-Correlated Expression of Cyclin B Drives Probabilistic*



- Cell-Size Homeostasis in Fission Yeast*, *Curr. Biol.* **29**, 1379 (2019).
- [78] Z. Xie, *Case Studies in Time Series Analysis* (World Scientific, Singapore, 1993), <https://doi.org/10.1142/1698>.
- [79] N. Walker, P. Nghe, and S. J. Tans, *Generation and Filtering of Gene Expression Noise by the Bacterial Cell Cycle*, *BMC Biol.* **14**, 11 (2016).
- [80] C. Zhang, M.-S. Liu, and X.-H. Xing, *Temperature Influence on Fluorescence Intensity and Enzyme Activity of the Fusion Protein of GFP and Hyperthermophilic Aylanase*, *Applied Microbiology and Biotechnology* **84**, 511 (2009).
- [81] C. Cadart, S. Monnier, J. Grilli, P. J. Sáez, N. Srivastava, R. Attia, E. Terriac, B. Baum, M. Cosentino-Lagomarsino, and M. Piel, *Size Control in Mammalian Cells Involves Modulation of Both Growth Rate and Cell Cycle Duration*, *Nat. Commun.* **9**, 3275 (2018).
- [82] S. Xie and J. M. Skotheim, *A G1 Sizer Coordinates Growth and Division in the Mouse Epidermis*, *Curr. Biol.* **30**, 916 (2020).
- [83] A. Amir, *Cell Size Regulation in Bacteria*, *Phys. Rev. Lett.* **112**, 208102 (2014).
- [84] M. Osella, E. Nugent, and M. C. Lagomarsino, *Concerted Control of Escherichia coli Cell Division*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3431 (2014).
- [85] C. A. Vargas-García, M. Soltani, and A. Singh, *Conditions for Cell Size Homeostasis: A Stochastic Hybrid System Approach*, *IEEE Life Sci. Lett.* **2**, 47 (2016).
- [86] P. Thomas, *Analysis of Cell Size Homeostasis at the Single-Cell and Population Level*, *Front. Phys.* **6**, 64 (2018).
- [87] J. Lin and A. Amir, *Homeostasis of Protein and mRNA Concentrations in Growing Cells*, *Nat. Commun.* **9**, 1 (2018).
- [88] C. Nieto, J. Arias-Castro, C. Sánchez, C. Vargas-García, and J. M. Pedraza, *Unification of Cell Division Control Strategies through Continuous Rate Models*, *Phys. Rev. E* **101**, 022401 (2020).
- [89] F. Bertaux, J. Von Kügelgen, S. Marguerat, and V. Shahrezaei, *A Bacterial Size Law Revealed by a Coarse-Grained Model of Cell Physiology*, *PLoS Comput. Biol.* **16**, e1008245 (2020).
- [90] P. Thomas and V. Shahrezaei, *Coordination of Gene Expression Noise with Cell Size: Extrinsic Noise versus Agent-Based Models of Growing Cell Populations*, bioRxiv, <https://doi.org/10.1101/2020.10.23.352856>.
- [91] C. Jia, A. Singh, and R. Grima, *Cell Size Distribution of Lineage Data: Analytic Results and Parameter Inference*, *iScience* **24**, 102220 (2021).
- [92] M. Wang, J. Zhang, H. Xu, and I. Golding, *Measuring Transcription at a Single Gene Copy Reveals Hidden Drivers of Bacterial Individuality*, *Nat. Microbiol.* **4**, 2118 (2019).
- [93] X.-M. Sun, A. Bowman, M. Priestman, F. Bertaux, A. Martinez-Segura, W. Tang, C. Whilding, D. Dormann, V. Shahrezaei, and S. Marguerat, *Size-Dependent Increase in RNA Polymerase II Initiation Rates Mediates Gene Expression Scaling with Cell Size*, *Curr. Biol.* **30**, 1217 (2020).
- [94] C. Nieto-Acuña, J. C. Arias-Castro, C. Vargas-García, C. Sánchez, and J. M. Pedraza, *Correlation between Protein Concentration and Bacterial Cell Size Can Reveal Mechanisms of Gene Expression*, *Phys. Biol.* **17**, 045002 (2020).
- [95] C. D. Cox, J. M. McCollum, D. W. Austin, M. S. Allen, R. D. Dar, and M. L. Simpson, *Frequency Domain Analysis of Noise in Simple Gene Circuits*, *Chaos* **16**, 026102 (2006).
- [96] A. Gupta and M. Khammash, *Frequency Spectra and the Color of Cellular Noise*, bioRxiv, <https://doi.org/10.1101/2020.09.15.292664>.
- [97] B. Munsky, B. Trinh, and M. Khammash, *Listening to the Noise: Random Fluctuations Reveal Gene Network Parameters*, *Mol. Syst. Biol.* **5**, 318 (2009).
- [98] N. Reverón-Gómez, C. González-Aguilera, K. R. Stewart-Morgan, N. Petryk, V. Flury, S. Graziano, J. V. Johansen, J. S. Jakobsen, C. Alabert, and A. Groth, *Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication*, *Mol. Cell* **72**, 239 (2018).
- [99] S. Hamperl and K. A. Cimprich, *Conflict Resolution in the Genome: How Transcription and Replication Make It Work*, *Cell* **167**, 1455 (2016).