Ab initio scaling laws between noise and mean of gene expression

Guillermo Rodrigo*

Institute for Integrative Systems Biology (I2SysBio), CSIC–University Valencia, 46980 Paterna, Spain

(Received 9 April 2019; published 30 September 2019)

Gene expression is inherently noisy due to fluctuations occurring at the molecular level. From a top-down perspective, noise has been traditionally decomposed into an intrinsic component that scales inversely with the mean expression level and an extrinsic component that is constant in absence of regulatory changes. Here, we adopt a bottom-up approach to reveal that extrinsic noise, by itself, can follow the aforementioned decomposition, which entails that one component of it can be confounded with intrinsic noise. Analytical expressions of the noise-mean relationship were derived for different scenarios, which were in part supported by numerical simulations.

 DOI: [10.1103/PhysRevE.100.032415](https://doi.org/10.1103/PhysRevE.100.032415)

I. INTRODUCTION

Genes are the fundamental, information-encoding pieces into which the DNA of every living organism is divided, and their expression (e.g., the biochemical process by which RNAs and subsequently proteins are produced) is inherently stochastic [1]. Over the last years, much experimental and theoretical work has been conducted with the aim of unveiling and characterizing the mechanisms that underlie cell-to-cell variability in gene expression, what is called molecular noise [2–8]. By defining noise (denoted by η^2) as the square of the coefficient of variation of gene expression, the conventional model states that $\eta^2 = \frac{b}{\langle \cdot \rangle} + B$ [9–11], where the first term accounts for the intrinsic noise (i.e., the variability generated specifically in the gene of study) and the second for the extrinsic noise (i.e., the variability generated at a global scale, shared by different genes) [2]. In this simple but effective model, while the extrinsic noise is constant (B , this holds for nonregulated genes or invariant regulatory activity), the intrinsic noise scales inversely with the mean expression level (denoted by $\langle \cdot \rangle$). In brief, b is the Fano factor and represents the number of molecules produced on average per stochastic fluctuation (when $b = 1$, the noise is Poissonian).

In this work, we challenge this conventional model. For that, we follow a bottom-up approach to derive a general model based on stochastic differential equations (SDEs) in which the extrinsic noise arises as a consequence of fluctuations in the different kinetic parameters that shape gene expression. For instance, the transcription rate will vary if the pool of available RNA polymerases does [12]. Similarly, the mRNA degradation rate will be susceptible to perturbations in ribonucleases [13]. The relationship between extrinsic noise and cellular factors has been previously tackled mathematically [14–16]. Yet, it has been difficult to end with a tractable model, integrating in a simple way the variations in all kinetic parameters as a result of fluctuations in cellular factors, from which to perform analytical calculations. Moreover, the

growing generation of single-cell (and even single-molecule) data urges improved quantitative descriptions [17].

II. RESULTS

A. General framework

Let us denote by x_i the concentration of the i th species at play, whose level depends on a series of biochemical reactions (of synthesis and degradation) occurring in the cell [1]. These reactions can be modeled if we know the different kinetic parameter values (α_k denotes the nominal value of the k th parameter). Then, if f_{ij} denotes the j th reaction affecting the expression of the i th species, by following a perturbative approach the general system of SDEs that model the dynamics of gene expression reads

$$\begin{aligned} \frac{dx_i}{dt} = & \sum_j f_{ij}(t, \{x_i\}, \{\alpha_k\}) \\ & + \sum_j \sqrt{|f_{ij}(t, \{x_i\}, \{\alpha_k\})|} \xi_{ij}(t) \\ & + \sum_k \sum_j \frac{\partial}{\partial \alpha_k} f_{ij}(t, \{x_i\}, \{\alpha_k\}) \Delta \alpha_k(t), \end{aligned} \quad (1)$$

where ξ_{ij} 's are independent Wiener processes that account for intrinsic noise in each biochemical reaction (white noise), and their statistics (mean and autocorrelation) are $\langle \xi_{ij}(t) \rangle = 0$ and $\langle \xi_{ij}(0) \xi_{ij}(t) \rangle = \delta(t)$ [18]. Moreover, $\Delta \alpha_k$ is the uncertainty in the value of the k th parameter, which is associated with extrinsic noise. By assuming that α_k is proportional to the concentration of a given cellular factor that fluctuates with time (z), typically a protein, we can write $\langle \Delta \alpha_k^2 \rangle \propto \langle \Delta z^2 \rangle$. If we further assume that the noise in the cellular factor scales with its mean, as $\langle \Delta z^2 \rangle = b \langle z \rangle + B \langle z \rangle^2$ [9,10], we obtain $\langle \Delta \alpha_k^2 \rangle \propto \alpha_k + \varepsilon_k \alpha_k^2$, with ε_k a constant balancing the scaling law of the extrinsic fluctuation. Therefore, given θ_k , a constant of proportionality, it turns out $\Delta \alpha_k(t) = \theta_k \sqrt{\alpha_k + \varepsilon_k \alpha_k^2} \zeta_k(t)$. Here, ζ_k 's are different Ornstein-Uhlenbeck processes that account for noise in each biochemical parameter (colored

*guillermo.rodrigo@csic.es

noise), and their statistics are $\langle \zeta_k(t) \rangle = 0$ and $\langle \zeta_k(0)\zeta_k(t) \rangle = \frac{1}{2\tau} e^{-|t|/\tau}$, where τ is the correlation time. It has been revealed that τ is given by the effective protein degradation rate (e.g., the growth rate in the case of bacteria) [4].

A sum of independent stochastic processes (of Wiener or Ornstein-Uhlenbeck) is equivalent to one process of the same nature, as $\sum_j q_{ij} \xi_{ij}(t) = \sqrt{\sum_j q_{ij}^2} \xi_i(t)$. Hence, if the fluctuations in the different cellular factors are independent, and provided that each parameter introducing noise only affects a particular biochemical reaction in the system, Eq. (1) can be rewritten as

$$\begin{aligned} \frac{dx_i}{dt} = & \sum_j f_{ij}(t, \{x_i\}, \{\alpha_k\}) + \sqrt{\sum_j |f_{ij}(t, \{x_i\}, \{\alpha_k\})|^2} \xi_i(t) \\ & + \sqrt{\sum_k \theta_k^2 (\alpha_k + \varepsilon_k \alpha_k^2) \left| \sum_j \frac{\partial}{\partial \alpha_k} f_{ij}(t, \{x_i\}, \{\alpha_k\}) \right|^2} \zeta_i(t), \end{aligned} \quad (2)$$

where the same correlation time was considered for all Ornstein-Uhlenbeck processes. This seems reasonable if all cellular factors are proteins with the same half-life. However, if some factors were RNAs (typically more short-lived) or proteins with widely different stability, that consideration would introduce inaccuracies.

B. Expression model

Following Eq. (2), the system of SDEs that model the dynamics of the mRNA and protein concentrations of a given gene (denoted by m and p , respectively) reads

$$\begin{aligned} \frac{dm}{dt} = & k_m - \gamma_m m + \sqrt{k_m + \gamma_m m} \xi_m(t) \\ & + \sqrt{\theta_1^2 (k_m + \varepsilon_1 k_m^2) + \theta_2^2 (\gamma_m + \varepsilon_2 \gamma_m^2) m^2} \zeta_m(t), \\ \frac{dp}{dt} = & k_p m - \gamma_p p + \sqrt{k_p m + \gamma_p p} \xi_p(t) \\ & + \sqrt{\theta_3^2 (k_p + \varepsilon_3 k_p^2) m^2 + \theta_4^2 (\gamma_p + \varepsilon_4 \gamma_p^2) p^2} \zeta_p(t), \end{aligned} \quad (3)$$

where k_m represents the transcription rate, k_p is the translation rate, γ_m is the mRNA degradation rate, and γ_p is the effective protein degradation rate.

Easily, the deterministic solution is $\langle m \rangle = \frac{k_m}{\gamma_m}$ and $\langle p \rangle = \frac{k_p \langle m \rangle}{\gamma_p}$. Note that here first-order kinetics are assumed, but nothing prevents the use of more complex kinetics (e.g., *sensu* Michaelis-Menten [19]) to model the dynamic behavior of the system. Only, the analytical expressions of noise will be more complex as a consequence of the derivatives with respect to the parameters and less accurate if we follow a perturbative approach [20].

C. Analytical noise calculations

1. Comparable Poissonian fluctuations in the kinetic parameters

The value of θ_k might be different for each kinetic parameter, as each cellular factor would enter into the system with a characteristic susceptibility [14,15]. However, we can

consider a scenario of comparable fluctuations in the parameters and then simplify the model with $\theta_k = \theta$. Moreover, we can neglect the second-order contribution in the variance of the cellular factor and then further simplify the model with $\varepsilon_k = 0$ (Poissonian regime). This assumption would hold if the cellular factor were in low or moderate abundance. Hence, under the mean-field approximation [21], the SDE for the fluctuation in the mRNA expression (Δm) reads

$$\begin{aligned} \frac{d\Delta m}{dt} = & -\gamma_m \Delta m + \sqrt{k_m + \gamma_m \langle m \rangle} \xi_m(t) \\ & + \theta \sqrt{k_m + \gamma_m \langle m \rangle^2} \zeta_m(t). \end{aligned} \quad (4)$$

For convenience, the constant modulating of the extrinsic noise level can be redefined as $\theta = \sqrt{\frac{2\sigma^2}{\gamma_p}}$ (also with $\tau = \frac{1}{\gamma_p}$). By applying the Fourier transform [22], it turns out that

$$S_{\Delta m} = \frac{2\gamma_m \langle m \rangle}{\omega^2 + \gamma_m^2} + \frac{2\sigma^2 \gamma_m \gamma_p (\langle m \rangle + \langle m \rangle^2)}{(\omega^2 + \gamma_m^2)(\omega^2 + \gamma_p^2)}, \quad (5)$$

where $S_{\Delta m}$ is the power spectrum of Δm (ω represents the frequency variable). By applying the Wiener-Khinchin theorem, and in the limit $\gamma_m \gg \gamma_p$ (i.e., faster degradation of mRNA than protein), it turns out an expression for $\langle \Delta m^2 \rangle$, the variance of the fluctuation. Finally, the noise in the mRNA expression ($\eta_m^2 = \frac{\langle \Delta m^2 \rangle}{\langle m \rangle^2}$) is

$$\eta_m^2 = \left(1 + \frac{\sigma^2}{\gamma_m}\right) \frac{1}{\langle m \rangle} + \frac{\sigma^2}{\gamma_m}, \quad (6)$$

which can be decomposed by defining $b_{in} = 1$, $b_{ex} = \frac{\sigma^2}{\gamma_m}$, and $B_{ex} = \frac{\sigma^2}{\gamma_m}$, in intrinsic and extrinsic contributions ($\eta_m^2 = \eta_{m,in}^2 + \eta_{m,ex}^2$) as

$$\begin{aligned} \eta_{m,in}^2 &= \frac{b_{in}}{\langle m \rangle}, \\ \eta_{m,ex}^2 &= \frac{b_{ex}}{\langle m \rangle} + B_{ex}. \end{aligned} \quad (7)$$

This is illustrated in Fig. 1(a).

In addition, the SDE for the fluctuation in protein expression (Δp) reads

$$\begin{aligned} \frac{d\Delta p}{dt} = & k_p \Delta m - \gamma_p \Delta p + \sqrt{k_p \langle m \rangle + \gamma_p \langle p \rangle} \xi_p(t) \\ & + \theta \sqrt{k_p \langle m \rangle^2 + \gamma_p \langle p \rangle^2} \zeta_p(t). \end{aligned} \quad (8)$$

By applying the Fourier transform, and noting that Δm , ξ_p , and ζ_p are independent fluctuations, it turns out that

$$S_{\Delta p} = \frac{k_p^2 S_{\Delta m}}{\omega^2 + \gamma_p^2} + \frac{2\gamma_p \langle p \rangle}{\omega^2 + \gamma_p^2} + \frac{2\sigma^2 \gamma_p^2 \left(\frac{\gamma_p}{k_p} + 1\right) \langle p \rangle^2}{(\omega^2 + \gamma_p^2)^2}, \quad (9)$$

where $S_{\Delta p}$ is the power spectrum of Δp . Again, by applying the Wiener-Khinchin theorem, in the limit $\gamma_m \gg \gamma_p$ as before and further assuming $k_p \gg \gamma_p$, it turns out an expression for $\langle \Delta p^2 \rangle$. Finally, the noise in the protein expression ($\eta_p^2 = \frac{\langle \Delta p^2 \rangle}{\langle p \rangle^2}$) is

$$\eta_p^2 = \left(1 + \frac{k_p}{\gamma_m} + \frac{\sigma^2 k_p}{2\gamma_m \gamma_p}\right) \frac{1}{\langle p \rangle} + \frac{\sigma^2}{2\gamma_p}, \quad (10)$$

FIG. 1. Noise dependence (analytical) on average expression level (in number of molecules per cell) according to the novel model presented in this work, when comparable Poissonian fluctuations in the kinetic parameters are considered. a) For mRNA. b) For protein. Parameters: $\sigma^2/\gamma_m = 0.1$, $k_p/\gamma_m = 1$ mol., and $\gamma_m/\gamma_p = 10$.

which again can be decomposed, by defining $b_{in} = 1 + \frac{k_p}{\gamma_m}$, $b_{ex} = \frac{\sigma^2 k_p}{2\gamma_m \gamma_p}$, and $B_{ex} = \frac{\sigma^2}{2\gamma_p}$, in intrinsic and extrinsic contributions ($\eta_p^2 = \eta_{p,in}^2 + \eta_{p,ex}^2$) as

$$\begin{aligned} \eta_{p,in}^2 &= \frac{b_{in}}{\langle p \rangle}, \\ \eta_{p,ex}^2 &= \frac{b_{ex}}{\langle p \rangle} + B_{ex}. \end{aligned} \quad (11)$$

This is illustrated in Fig. 1(b).

It is worthwhile to note at this point that if the protein of interest were short-lived ($\tau \gg \frac{1}{\gamma_p}$), we would have $b_{ex} = \frac{\sigma^2 k_p}{\gamma_m \gamma_p}$ and $B_{ex} = \frac{\sigma^2}{\gamma_p}$ instead.

2. Comparable non-Poissonian fluctuations in the kinetic parameters

If the cellular factors were in high abundance, their expression noise would be mostly constant. We can now consider this scenario and simplify the model accordingly with $\varepsilon_k \gg \frac{1}{\alpha_k}$, together with $\theta_k = \theta$ and $\varepsilon_k = \varepsilon$ for comparable fluctuations in the parameters. Hence, the SDE for Δm reads

$$\begin{aligned} \frac{d\Delta m}{dt} &= -\gamma_m \Delta m + \sqrt{k_m + \gamma_m \langle m \rangle} \xi_m(t) \\ &+ \theta \sqrt{\varepsilon(k_m^2 + \gamma_m^2 \langle m \rangle^2)} \zeta_m(t). \end{aligned} \quad (12)$$

With $\theta = \sqrt{\frac{2\sigma^2}{\gamma_p}}$ and $\gamma_m \gg \gamma_p$, it turns out that

$$\eta_m^2 = \frac{1}{\langle m \rangle} + 2\sigma^2 \varepsilon. \quad (13)$$

Note that now $b_{ex} = 0$ (i.e., the extrinsic noise is constant). Equation (7) stands with $b_{in} = 1$ and $B_{ex} = 2\sigma^2 \varepsilon$.

In addition, the SDE for Δp reads

$$\begin{aligned} \frac{d\Delta p}{dt} &= k_p \Delta m - \gamma_p \Delta p + \sqrt{k_p \langle m \rangle + \gamma_p \langle p \rangle} \xi_p(t) \\ &+ \theta \sqrt{\varepsilon(k_p^2 \langle m \rangle^2 + \gamma_p^2 \langle p \rangle^2)} \zeta_p(t). \end{aligned} \quad (14)$$

Thus, we end with

$$\eta_p^2 = \left(1 + \frac{k_p}{\gamma_m}\right) \frac{1}{\langle p \rangle} + 2\sigma^2 \varepsilon, \quad (15)$$

in which again $b_{ex} = 0$. Equation (11) stands with $b_{in} = 1 + \frac{k_p}{\gamma_m}$ and $B_{ex} = 2\sigma^2 \varepsilon$.

3. Extrinsic noise dominated by fluctuations in the transcription rate

It has been suggested that transcription is the biochemical process that mostly contributes to noise in gene expression [23,24], especially in eukaryotes. We can take into account this scenario by setting $\theta_1 \gg \theta_k$, then leading to

$$\begin{aligned} \frac{d\Delta m}{dt} &= -\gamma_m \Delta m + \sqrt{k_m + \gamma_m \langle m \rangle} \xi_m(t) \\ &+ \theta_1 \sqrt{k_m + \varepsilon_1 k_m^2} \zeta_m(t). \end{aligned} \quad (16)$$

Again with $\theta_1 = \sqrt{\frac{2\sigma^2}{\gamma_p}}$ and $\gamma_m \gg \gamma_p$, it turns out that

$$\eta_m^2 = \left(1 + \frac{\sigma^2}{\gamma_m}\right) \frac{1}{\langle m \rangle} + \sigma^2 \varepsilon_1. \quad (17)$$

Thus, with $b_{in} = 1$, $b_{ex} = \frac{\sigma^2}{\gamma_m}$, and $B_{ex} = \sigma^2 \varepsilon_1$, we recover Eq. (7).

In addition, the SDE for Δp is given by Eq. (8) with $\theta = 0$, so

$$\eta_p^2 = \left(1 + \frac{k_p}{\gamma_m} + \frac{\sigma^2 k_p}{2\gamma_m \gamma_p}\right) \frac{1}{\langle p \rangle} + \frac{\sigma^2 \varepsilon_1}{2}, \quad (18)$$

leading to Eq. (11) with $b_{in} = 1 + \frac{k_p}{\gamma_m}$, $b_{ex} = \frac{\sigma^2 k_p}{2\gamma_m \gamma_p}$, and $B_{ex} = \frac{\sigma^2 \varepsilon_1}{2}$.

It is important to note that the main difference with respect to the results obtained in Sec. II C 1 is that now, in terms of protein, B_{ex} presents a lower value, as $\varepsilon_1 \ll \frac{1}{\gamma_p}$ can be considered. This is illustrated in Fig. 2.

D. Numerical simulations

To validate, at least in part, our analytical calculations, we performed numerical simulations of the system of SDEs presented in Eq. (3). For that, we allowed propensities to vary with time, beyond the mean-field approximation (i.e., using m and p rather than $\langle m \rangle$ and $\langle p \rangle$). We considered different sets of kinetic parameters to reach different expression levels. The parameters were directly Ornstein-Uhlenbeck processes,

FIG. 2. Noise dependence (analytical) on average protein expression level (in number of molecules per cell) according to the novel model presented in this work, when extrinsic noise is dominated by fluctuations in the transcription rate. Parameters: $\sigma^2/\gamma_m = \sigma^2\varepsilon_1 = 0.1$, $k_p/\gamma_m = 1$ mol, and $\gamma_m/\gamma_p = 10$.

avoiding a first-order expansion on them. Simulations were done for 10^4 h with an increment of 0.1 h by employing a practical numerical approach [25,26] implemented in MATLAB (MathWorks).

Specifically, we considered different nominal transcription rates ($k_m = 0.1$ – 100 mol/h), keeping fixed the nominal mRNA degradation rate ($\gamma_m = 1$ h $^{-1}$), to achieve different values of $\langle m \rangle$, covering several orders of magnitude. Moreover, we considered a single value for the nominal translation rate ($k_p = 1$ h $^{-1}$) and two values for the nominal protein degradation rate ($\gamma_p = 0.04$ or 0.4 h $^{-1}$), the greater to achieve low protein levels. Despite this, the frequency of the extrinsic fluctuations was always $\frac{1}{\tau} = 0.04$ h $^{-1}$. Accordingly, $\langle p \rangle$ also spanned over several orders of magnitude. Of note, these nominal values are compatible with experimental determinations in eukaryotes [27]. To model the fluctuations in the kinetic parameters, we set $\theta_1 = \theta_2 = \theta_3 = 1.41$, $\theta_4 = 0.14$, $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 1$, and $\varepsilon_4 = 25$ (each in its appropriate units). Numerical simulations were then carried out by considering only intrinsic noise, only extrinsic noise, or both types of noise (Fig. 3).

Analytically, when $k_m \approx 1$ mol/h, the Poissonian regime can be invoked, then with $\sigma^2 = \theta_1^2 \gamma_p$ (as $\varepsilon_1 k_m \approx 1$ and $\varepsilon_2 \gamma_m = 1$). This leads to $b_{ex} = \frac{\sigma^2}{\gamma_m} = 0.08$, in terms of mRNA. However, this regime cannot be assumed to calculate B_{ex} , because at high mRNA levels the transcriptional fluctuations are non-Poissonian. Instead, we shall write $B_{ex} = 2\sigma^2\varepsilon_1 = 0.16$. Of relevance, we found that Eqs. (7) with those values of b_{ex} and B_{ex} , together with $b_{in} = 1$, explain well the data [Fig. 3(a)].

In terms of protein, in addition, we can write $b_{ex} = \frac{\sigma^2 k_p}{\gamma_m \gamma_p} = 0.20$, as for low $\langle p \rangle$ values the protein is short-lived ($\gamma_p = 0.4$ h $^{-1}$) and the Poissonian regime holds. However, as before, we need to write $B_{ex} = 2\sigma^2\varepsilon_1 = 0.16$. Of relevance, we also found that Eqs. (11) with those values of b_{ex} and B_{ex} , together with $b_{in} = 1 + \frac{k_p}{\gamma_m} = 2$, correctly explain the numerical data [Fig. 3(b)].

III. DISCUSSION

In sum, we have relied on *ab initio* calculations to develop a simple theory that provides original noise-mean scaling

FIG. 3. Noise dependence (analytical and numerical) on average expression level (in number of molecules per cell). Points correspond to numerical simulations and lines to analytical models. (a) For mRNA. (b) For protein. Different parameter values were used to obtain a wide range of expression levels. The insets show particular stochastic dynamics with total noise.

laws in gene expression. Importantly, this noise description can be directly tested against experimental data (measuring expression levels for a large cell population, to then calculate means and variances). This can be done from expression data of genomic approaches [11,17] or externally regulable genes [28]. The slope (at low expression levels) and the floor (at high levels) of the noise-mean scaling law shall reveal the values of $b_{in} + b_{ex}$ and B_{ex} , respectively. To extract the value of b_{ex} when b_{in} is not known (>1), additional data will be required, such as from experiments with a second gene copy [15]. Moreover, we provide *ab initio* calculations for different scenarios (e.g., when all reactions equally contribute to the extrinsic noise or when transcription is the main determinant). Of note, $b_{ex} = 0$ if the cellular resources used for gene expression are in great excess and a Poissonian regime cannot be invoked ($\langle \Delta z^2 \rangle \propto \langle z \rangle^2$). In principle, by analyzing the $\frac{b_{ex}}{B_{ex}}$ ratio, both in terms of mRNA and protein, we might get certain mechanistic insight about the underlying noise scheme.

Our calculations indeed highlight the importance of distilling with precision the different contributions to the Fano factor in order to appreciate the effect of the stochastic fluctuations of extrinsic nature, in contrast to the conventional wisdom [9–11]. In terms of mRNA, our model is more

conceptual than practical, as $b_{ex} \ll b_{in} = 1$ (i.e., $\sigma^2 \ll \gamma_m$, if we follow the Poissonian regime of extrinsic fluctuations). This is appreciated at low mRNA abundance, where $\eta_{m,in}^2 \approx \eta_m^2$ [Fig. 1(a)]. Indeed, by looking at the variability of highly expressed genes in mammalian cells, it results that $B_{ex} \approx 0.06$ [11]. Thus, assuming a half-life of about 1 h for transcripts, we have $\sigma \approx 0.24 \text{ h}^{-1/2}$ (with $\frac{b_{ex}}{b_{in}} \approx 0.06$). In terms of protein, by contrast, the model predicts a higher $\frac{b_{ex}}{b_{in}}$ ratio (approximated by $\frac{\sigma^2}{2\gamma_p}$), due to time averaging of stochastic fluctuations from transcription to translation ($\gamma_p \ll \gamma_m$). Figure 1(b) shows how the noise-mean scaling is maintained for $\eta_{p,in}^2$ and η_p^2 , but with different Fano factors. By using that inferred value of σ and considering a half-life of about 1 d for mammalian proteins, we have $\frac{b_{ex}}{b_{in}} \approx 0.72$. This means that stochastic protein bursts are substantially contributed by changes from time to time in the transcription rate [24], and not only by intrinsic fluctuations in mRNA levels as previously thought [9]. This entails that such bursts might appear even in scenarios of high expression or might be correlated between different proteins [15].

So far, we considered a gene transcribed homogeneously at a given rate. However, increasing evidence suggests that many genes, mainly in eukaryotic cells, are transcribed by bursts [6,29]. In Appendix A, we provide a new noise calculation by considering this scenario. Now, as the bursting process typically dominates the amplitude of the stochastic fluctuations of intrinsic nature ($k_m \gg k_{off}$ in mammalian cells, then $b_{in} \approx \frac{k_m}{k_{off}} \gg 1$ for mRNA) [28], the impact of extrinsic noise on the Fano factor is reduced ($b_{ex} \ll b_{in}$). Yet, our results revealed that if transcription remains off for large periods, the bursting process also determines the fluctuations of extrinsic nature ($\gamma_m \gg k_{on}$, then $B_{ex} \approx \frac{\sigma^2}{k_{on}}$ for mRNA; although this should be taken more in qualitative terms due to the use of an approximative model) [30]. Perhaps perturbations in chromatin remodelers that are propagated to the off-on switching kinetics can be explanatory [31].

Moreover, if the gene is regulated (e.g., by an upstream transcription factor), the major contribution to the extrinsic noise will be given by the uncertainty associated with the corresponding molecular binding (e.g., a protein-DNA binding). As observed experimentally [2,5], this leads to a peak-shape function by following a description based on Hill functions [32], which will only affect B_{ex} .

We might also consider an alternative scenario by assuming that the different stochastic fluctuations of extrinsic nature are correlated (nonindependent), provided they are the result of global changes in the cell. For instance, a variation in the cell growth rate (in bacteria) could lead to systematic changes in the different resources used for gene expression [33]. In Appendix B, we provide a new noise calculation by considering the ideal case of common fluctuations, revealing that more complex noise-mean scaling laws can appear.

Finally, a limitation of a noise-mean scaling law is that its parametrization (the values of b_{in} , b_{ex} , and B_{ex}) can vary from gene to gene, especially when the translation rate changes. We would then expect a tighter genome-wide distribution in the case of mRNA than in the case of protein expression [17]. The parametrization can even change for a given gene if we are able to tune its expression level. Another drawback here

is the fact that at low expression levels (<10 mol) a continuous model given by SDEs might be not accurate enough. Certainly, the transition from a master equation to a Langevin equation is typically justified when the number of molecules is high; yet, the first and second moments analytically calculated from those approaches always coincide when the system is linear (i.e., based on first-order kinetics) [34]. In addition, more complex analytical treatments might be adopted for noise calculation, beyond mean-field theory; treatments that might lead to more accurate predictions [35]. Overall, the results presented in this work contribute to enlarge our knowledge on gene expression noise, providing a basic framework from which to assess from scratch the impact of extrinsic noise.

ACKNOWLEDGMENT

This work was supported by the Spanish Ministry of Science, Innovation, and Universities grant PGC2018-101410-B-I00 (cofinanced by the ERDF).

APPENDIX A: NONHOMOGENEOUS TRANSCRIPTION

If the gene is expressed through transcriptional bursts (i.e., pulses in the transcriptional activity of the promoter), the system of SDEs that model the dynamics reads (for comparable Poissonian fluctuations in the kinetic parameters)

$$\begin{aligned} \frac{dm}{dt} &= k_m \psi(t) - \gamma_m m + \sqrt{k_m \psi(t) + \gamma_m m} \xi_m(t) \\ &\quad + \theta \sqrt{k_m \psi(t)^2 + k_{on} \left(\frac{\partial k_m \psi(t)}{\partial k_{on}} \right)^2} \\ &\quad + k_{off} \left(\frac{\partial k_m \psi(t)}{\partial k_{off}} \right)^2 + \gamma_m m^2 \zeta_m(t), \\ \frac{dp}{dt} &= k_p m - \gamma_p p, \end{aligned} \quad (A1)$$

where ψ is a telegraph process [6], which is described by the kinetics k_{on} (transcription activation) and k_{off} (transcription deactivation). In the limit $k_{on} \ll k_{off}$, its statistics are $\langle \psi(t) \rangle = \rho$ and $\langle \psi(0)\psi(t) \rangle = \rho(1 - \rho)e^{-k_{off}|t|}$, with $\rho = \frac{k_{on}}{k_{off}}$. The SDE for Δm then reads

$$\begin{aligned} \frac{d\Delta m}{dt} &= k_m \Delta \psi(t) - \gamma_m \Delta m + \sqrt{k_m \rho + \gamma_m \langle m \rangle} \xi_m(t) \\ &\quad + \theta \sqrt{k_m \rho^2 + \frac{k_m^2 \rho^2}{k_{on}} + \frac{k_m^2 \rho^2}{k_{off}} + \gamma_m \langle m \rangle^2} \zeta_m(t). \end{aligned} \quad (A2)$$

By applying the Fourier transform, it turns out that

$$\begin{aligned} S_{\Delta m} &= \frac{2k_m k_{off} \gamma_m \langle m \rangle}{(\omega^2 + \gamma_m^2)(\omega^2 + k_{off}^2)} + \frac{2\gamma_m \langle m \rangle}{\omega^2 + \gamma_m^2} \\ &\quad + \frac{2\sigma^2 \gamma_m \gamma_p}{(\omega^2 + \gamma_m^2)(\omega^2 + \gamma_p^2)} \left(\rho \langle m \rangle + \left(\frac{\gamma_m}{k_{on}} + 1 \right) \langle m \rangle^2 \right). \end{aligned} \quad (A3)$$

By applying the Wiener-Khinchin theorem, and in the limit $k_{off} \gg \gamma_m \gg \gamma_p$ (i.e., shorter period of active transcription than mRNA half-life and faster degradation of mRNA than

protein), it turns out that the noise in the mRNA expression is

$$\eta_m^2 = \left(\frac{k_m}{k_{\text{off}}} + 1 + \frac{\sigma^2 \rho}{\gamma_m} \right) \frac{1}{\langle m \rangle} + \sigma^2 \left(\frac{1}{k_{\text{on}}} + \frac{1}{\gamma_m} \right), \quad (\text{A4})$$

which can be decomposed in intrinsic and extrinsic contributions, with $b_{\text{in}} = \frac{k_m}{k_{\text{off}}} + 1$, $b_{\text{ex}} = \frac{\sigma^2 \rho}{\gamma_m}$, and $B_{\text{ex}} = \sigma^2 \left(\frac{1}{k_{\text{on}}} + \frac{1}{\gamma_m} \right)$, to end with Eq. (7).

In addition, the SDE for Δp is given by Eq. (8). If the noise generated during the translation process is neglected (with respect to the amplified noise in the mRNA expression, provided $k_m \gg k_{\text{off}}$ and/or $k_p \gg \gamma_m$) [36], the noise in the protein expression is

$$\eta_p^2 = \frac{k_p}{\gamma_m} \left(\frac{k_m}{k_{\text{off}}} + 1 + \frac{\sigma^2 \rho}{2\gamma_p} \right) \frac{1}{\langle p \rangle} + \frac{\sigma^2}{2} \left(\frac{1}{k_{\text{on}}} + \frac{1}{\gamma_m} \right), \quad (\text{A5})$$

which again can be decomposed in intrinsic and extrinsic contributions, with $b_{\text{in}} = \frac{k_p}{\gamma_m} \left(\frac{k_m}{k_{\text{off}}} + 1 \right)$, $b_{\text{ex}} = \frac{\sigma^2 k_p \rho}{2\gamma_m \gamma_p}$, and $B_{\text{ex}} = \frac{\sigma^2}{2} \left(\frac{1}{k_{\text{on}}} + \frac{1}{\gamma_m} \right)$, to end with Eq. (11).

APPENDIX B: CORRELATION OF EXTRINSIC FLUCTUATIONS

In the ideal case of common fluctuations of extrinsic nature ($\zeta_k = \zeta$; e.g., due to a global dependence of the different resources used for gene expression on a varying determinant),

we would have

$$\sum_k \theta_k \sqrt{\alpha_k + \varepsilon_k \alpha_k^2} \sum_j \frac{\partial}{\partial \alpha_k} f_{ij}(t, \{x_i\}, \{\alpha_k\}) \zeta(t) \quad (\text{B1})$$

as the term accounting for extrinsic fluctuations in Eq. (2). Thus, in the case of comparable Poissonian fluctuations in the kinetic parameters, the SDE for Δm would read

$$\begin{aligned} \frac{d\Delta m}{dt} = & -\gamma_m \Delta m + \sqrt{k_m + \gamma_m \langle m \rangle} \xi_m(t) \\ & + \theta(\sqrt{k_m} + \sqrt{\gamma_m \langle m \rangle}) \zeta(t), \end{aligned} \quad (\text{B2})$$

which would lead to

$$S_{\Delta m} = \frac{2\gamma_m \langle m \rangle}{\omega^2 + \gamma_m^2} + \frac{2\sigma^2 \gamma_m \gamma_p (\langle m \rangle + 2\langle m \rangle^{3/2} + \langle m \rangle^2)}{(\omega^2 + \gamma_m^2)(\omega^2 + \gamma_p^2)}, \quad (\text{B3})$$

and finally

$$\eta_m^2 = \left(1 + \frac{\sigma^2}{\gamma_m} \right) \frac{1}{\langle m \rangle} + \frac{2\sigma^2}{\gamma_m \langle m \rangle^{1/2}} + \frac{\sigma^2}{\gamma_m}, \quad (\text{B4})$$

which would not be compatible with Eq. (7). To further obtain an expression for the noise in the protein expression, one would have to note that Δm and ζ are now nonindependent fluctuations.

Alternatively, when extrinsic noise is dominated by fluctuations in the transcription rate, the noise in the mRNA expression will follow the very same law, previously derived in Eq. (17).

-
- [1] J. M. Raser and E. K. O'Shea, Noise in gene expression: origins, consequences, and control, *Science* **309**, 2010 (2005).
 - [2] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, *Science* **297**, 1183 (2002).
 - [3] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Regulation of noise in the expression of a single gene, *Nat. Genet.* **31**, 69 (2002).
 - [4] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, Gene regulation at the single-cell level, *Science* **307**, 1962 (2005).
 - [5] J. M. Pedraza and A. van Oudenaarden, Noise propagation in gene networks, *Science* **307**, 1965 (2005).
 - [6] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, Stochastic mRNA synthesis in mammalian cells, *PLoS Biol.* **4**, e309 (2006).
 - [7] M. J. Dunlop, R. S. Cox III, J. H. Levine, R. M. Murray, and M. B. Elowitz, Regulatory activity revealed by dynamic correlations in gene expression noise, *Nat. Genet.* **40**, 1493 (2008).
 - [8] N. Ji, T. C. Middelkoop, R. A. Mentink, M. C. Betist, S. Tonegawa, D. Mooijman, H. C. Korswagen, and A. van Oudenaarden, Feedback control of gene expression variability in the *Caenorhabditis elegans* Wnt pathway, *Cell* **155**, 869 (2013).
 - [9] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, Noise in protein expression scales with natural protein abundance, *Nat. Genet.* **38**, 636 (2006).
 - [10] G. Hornung, R. Bar-Ziv, D. Rosin, N. Tokuriki, D. S. Tawfik, M. Oren, and N. Barkai, Noise-mean relationship in mutated promoters, *Genome Res.* **22**, 2409 (2012).
 - [11] D. Grün, L. Kester, and A. van Oudenaarden, Validation of noise models for single-cell transcriptomics, *Nat. Methods* **11**, 637 (2014).
 - [12] S. Yang, S. Kim, Y. R. Lim, C. Kim, H. J. An, J. H. Kim, J. Sung, and N. K. Lee, Contribution of RNA polymerase concentration variation to protein expression noise, *Nat. Commun.* **5**, 4761 (2014).
 - [13] A. Baudrimont, V. Jaquet, S. Wallerich, S. Voegeli, and A. Becskei, Contribution of RNA degradation to intrinsic and extrinsic noise in gene expression, *Cell Rep.* **26**, 3752 (2019).
 - [14] V. Shahrezaei, J. F. Ollivier, and P. S. Swain, Colored extrinsic fluctuations and stochastic gene expression, *Mol. Syst. Biol.* **4**, 196 (2008).
 - [15] A. Hilfinger and J. Paulsson, Separating intrinsic from extrinsic fluctuations in dynamic biological systems, *Proc. Natl. Acad. Sci. USA* **108**, 12167 (2011).
 - [16] J. Lei, Stochasticity in single gene expression with both intrinsic noise and fluctuation in kinetic parameters, *J. Theor. Biol.* **256**, 485 (2009).
 - [17] 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).
 - [18] D. T. Gillespie, The chemical Langevin equation, *J. Chem. Phys.* **113**, 297 (2000).

- [19] L. A. Segel, On the validity of the steady state assumption of enzyme kinetics, *Bull. Math. Biol.* **50**, 579 (1988).
- [20] P. Thomas, H. Matuschek, and R. Grima, How reliable is the linear noise approximation of gene regulatory networks?, *BMC Genom.* **14**, S5 (2013).
- [21] J. M. Horowitz, Diffusion approximations to the chemical master equation only have a consistent stochastic thermodynamics at chemical equilibrium, *J. Chem. Phys.* **143**, 044111 (2015).
- [22] M. L. Simpson, C. D. Cox, and G. S. Sayler, Frequency domain analysis of noise in autoregulated gene circuits, *Proc. Natl. Acad. Sci. USA* **100**, 4551 (2003).
- [23] W. J. Blake, M. Kaern, C. R. Cantor, and J. J. Collins, Noise in eukaryotic gene expression, *Nature (London)* **422**, 633 (2003).
- [24] M. S. Sherman, K. Lorenz, M. H. Lanier, and B. A. Cohen, Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression, *Cell Syst.* **1**, 315 (2015).
- [25] R. F. Fox, I. R. Gatland, R. Roy, and G. Vemuri, Fast, accurate algorithm for numerical simulation of exponentially correlated colored noise, *Phys. Rev. A* **38**, 5938 (1988).
- [26] R. Mannella and V. Palleschi, Fast and precise algorithm for computer simulation of stochastic differential equations, *Phys. Rev. A* **40**, 3381 (1989).
- [27] B. Schwanhaussner, 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).
- [28] R. D. Dar, S. M. Shaffer, A. Singh, B. S. Razooky, M. L. Simpson, A. Raj, and L. S. Weinberger, Transcriptional bursting explains the noise-versus-mean relationship in mRNA and protein levels, *PLoS ONE* **11**, e0158298 (2016).
- [29] 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).
- [30] R. D. Dar, B. S. Razooky, L. S. Weinberger, C. D. Cox, and M. L. Simpson, The low noise limit in gene expression, *PLoS ONE* **10**, e0140969 (2015).
- [31] C. R. Brown, C. Mao, E. Falkovskaia, M. S. Jurica, and H. Boeger, Linking stochastic fluctuations in chromatin structure and gene expression, *PLoS Biol.* **11**, e1001621 (2013).
- [32] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Transcriptional regulation by the numbers: models, *Curr. Opin. Genet. Dev.* **15**, 116 (2005).
- [33] S. Klumpp, Z. Zhang, and T. Hwa, Growth rate-dependent global effects on gene expression in bacteria, *Cell* **139**, 1366 (2009).
- [34] R. Khanin and D. J. Higham, Chemical Master Equation and Langevin regimes for a gene transcription model, *Theor. Comput. Sci.* **408**, 31 (2008).
- [35] 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).
- [36] M. M. Hansen, R. V. Desai, M. L. Simpson, and L. S. Weinberger, Cytoplasmic amplification of transcriptional noise generates substantial cell-to-cell variability, *Cell Syst.* **7**, 384 (2018).