Supporting Online Material for

Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression

Juan M. Pedraza and Johan Paulsson*

*To whom correspondence should be addressed.
E-mail: johan_paulsson@hms.harvard.edu

Published 18 January 2008, *Science* **319**, 339 (2008)

DOI: 10.1126/science.1144331

This PDF file includes:

Supporting Online Material

In this document we derive the equations presented in the main text and comment on some of the findings. To increase accessibility, we do not derive the full Eq. (1) immediately, but first present the standard result and then gradually generalize it using straightforward methods. We apologize for the somewhat algebraic derivations that result from this approach. Using results from renewal and queuing theory is likely to produce shorter proofs.

The document is organized as follows:

1. Some technical comments on the analysis

This section briefly discusses protein bursts, time-correlated noise, non-exponential times, and the role of simulations, to verbally clarify some common technical issues before the mathematical analysis.

2. Introducing master and moment equations.

This section introduces the chemical master equation and general moment equations for variances.

3. Re-deriving the standard model for gene expression.

This section derives the standard mRNA-protein model and shows how it can sometimes be condensed into geometric bursts. The derivation is based on moment equations, which are exact for this process.

4. External noise from linear response theory

This section presents an alternative derivation of the mRNA contribution to protein noise in the model above, using linear response theory. This method is also exact for this process.

5. Bursts from moment equation

This section shows a general and simple result for any bursting process, obtained simply by reinterpreting the terms in the moment equation.

6. Using master equations for non-exponential waiting times

This section discusses ways to use master equations to approximate non-exponential transitions, by first expanding the process into more variables, and then contracting the result before interpretation.

7. General timing and bursting statistics

This section derives the noise in the mRNA levels in the main example, using a combination of transform methods and other tricks. The protein noise can then be straightforwardly derived using linear response theory.

8. Senescence of mRNA

This section derives protein noise for the case when mRNAs are degraded at the end of an N step process and the autocorrelation function for fluctuations in mRNA levels when these senesce. This result is then used to calculate the corresponding protein noise. It also sketches out an alternative proof using similar methods as in section 5 above, expanding and contracting the master equation.

9. Parameters used in the figures

This section explains in detail the parameters used in the figures.

1. Some technical comments on the analysis

Protein bursts Our analysis focuses on nontrivial timing and bursting in the mRNA production and degradation. The analysis could easily be extended to the protein by similar principles. However, the geometric bursts of protein synthesis that have been emphasized in the literature simply reflect the underlying mRNA dynamics, so protein bursting is already included in some sense. Furthermore, by making the transcription bursty, protein synthesis automatically becomes bursty in more complex ways.

Time-correlated noise in the waiting times. In principle, there could also be long-term memory in T or b. However, the analysis cannot allow for any type of production statistics and still obtain explicit results. In particular, we see no other way to account for that option than to include more variables into the description, as it otherwise would be non-Markovian on long time-scales. One example of time-correlations is in fact already included in the analysis, by the fact that mRNAs are not assumed to be much more rapidly degraded than their encoded proteins. Our analyses thus attempt to generalize the possible statistical mechanisms that can be captured in each concentration variable, illustrating a general principle rather than deriving a theory that accounts for all sources of noise in the cell.

Non-exponential waiting times. As mentioned in footnote 11 of the main text, non-exponential waiting times could in principle be modeled simply by including many more variables into the standard context of chemical master equations. However, that would be similar to reconstructing nonlinear functions from large sets of piecewise linear functions: it may be possible but is physically misleading and makes analyses so complicated that general conclusions are precluded. The chemical master equation is simply poorly suited to describe many types of phenomena in biology. Exponentials certainly deserve a very special position in kinetics, but not all processes should be tweaked to fit a multi-exponential mold. The problems of the master equation approach are outlined in more technical detail in section 6.

Simulations vs analytical results All analytical results were double-checked by running Gillespie simulations (*1*) for a wide range of different systems, using a large sample size for accuracy. Because the Gillespie algorithm is simply a reformulation of the master equation, and the analytical results are either exact or virtually exact, the results must agree unless our implementations are flawed. In all cases, the simulations perfectly matched the analytical equations.

2. Introducing master and moment equations

Most models of stochastic kinetics formulate chemical master equations (CMEs): continuous-time Markov processes for the probability $P(\mathbf{x},t)$ where $\mathbf{x}$ is the state vector of x_i molecules of species X_i. With $\mathbf{s}^k$ as the vector of jump sizes associated with reaction number k, occurring with rate r_k and adding s_i^k molecules of X_i, the CME follows

$$\partial_t P(\mathbf{x},t) = \sum_k r_k(\mathbf{x}-\mathbf{s}^k)P(\mathbf{x}-\mathbf{s}^k,t) - r_k(\mathbf{x})P(\mathbf{x},t)\,. \qquad \text{(A.1)}$$

The CME can be formulated as a Monte-Carlo algorithm for generating sample paths (*2*), known in kinetics (*1, 3*) as the Gillespie-algorithm. It can also be used to formulate moment equations. Defining

$$J_i^{tot} = J_i^+ - J_i^- = \sum_k s_i^k r_k(\mathbf{x}) \quad \text{and} \quad B_{ij} = \sum_k s_i^k s_j^k r_k(\mathbf{x})\,, \qquad \text{(A.2)}$$

where J_i^+ and J_i^- are the total production and elimination fluxes of X_i respectively, changes in the average number of molecules $\langle \mathbf{x} \rangle$ and the covariance matrix **C**, where $C_{ij} = Cov(x_i, x_j)$, follow (*4*, *5*)

$$\partial_t \langle \mathbf{x} \rangle = \left\langle \mathbf{J}^{tot} \right\rangle \quad \text{and} \quad \partial_t \mathbf{C} = \left\langle \mathbf{J}^{tot} (\mathbf{x} - \langle \mathbf{x} \rangle)^T \right\rangle + \left\langle (\mathbf{x} - \langle \mathbf{x} \rangle) \mathbf{J}^{totT} \right\rangle + \langle \mathbf{B} \rangle . \quad \text{(A.3)}$$

Formulating CMEs requires enough state variables for changes to depend only on the current state, yet few enough to allow analysis. This necessitates condensations of fast transitions between short-lived states into single reaction steps, as bimolecular reactions in well-stirred systems are approximated as 'elementary' without accounting for spatial positions. The requirement of separated time-scales can also apply to complicated processes, which are said to be elementary complex (*6*).

When the reaction rates are linear in the state variables, as they are for all analytical results in the main body of the paper, then $\mathbf{J} = \mathbf{K} \times (\mathbf{x} - \langle \mathbf{x} \rangle)^T$ where **K** is the Jacobian coefficient matrix for changes in the averages. With $\eta_{ij} = C_{ij} / (\langle x_i \rangle \langle x_j \rangle)$ and $\mathbf{A}(t) = Cov(x_i(t'), x_j(t'+t)) / (\langle x_i(t') \rangle \langle x_j(t'+t) \rangle)$ as the normalized stationary covariances and autocovariances respectively, this leads to the following *exact* matrix equations

$$\mathbf{M}\boldsymbol{\eta} + \boldsymbol{\eta}\mathbf{M}^T = \mathbf{D} \quad \text{and} \quad \mathbf{A}(t) = \boldsymbol{\eta} Exp(-\mathbf{M}t), \quad \text{(A.4)}$$

where the diffusion matrix **D** is a normalized version of **B** in Eq. (A.2), and the drift matrix **M** is a normalized version of the Jacobian matrix **K**. The problem is that the conventional formulations of these matrices depend on the size and rate of each reaction event. When each chemical event adds or removes a single molecule, then we have previously shown that these matrices can be systematically reevaluated in terms of intuitive biological observables(*7, 8*) that can be directly measured:

$$D_{ii} = \frac{2}{\tau_i} \frac{1}{\langle x_i \rangle} \quad \text{and} \quad M_{ij} = \frac{H_{ij}}{\tau_i} \quad \text{where} \quad H_{ij} = \frac{\partial \ln (J_i^- / J_i^+)}{\partial \ln x_j} . \quad \text{(A.5)}$$

without any approximations. The drift matrix **M** describes how the system adjusts to steady state, and can be written in terms of average life-times τ_i and elasticities H_{ij}. The elasticities measure how the balance between degradation and synthesis of x_i responds to changes in x_j, as in metabolic control analysis (*9*). The diffusion matrix **D** reflects the randomness of the individual events, set by the average numbers of molecules $\langle x_i \rangle$ and the average step sizes of the individual chemical events, at this point simply 1.

3. Re-deriving the standard model for gene expression

This simple model summarizes the main principles of most previous work, and has been used to interpret many experimental analyses. To our knowledge it was first published by Paulsson & Ehrenberg (*10*), but related models were published by Berg and Rigney in the late 1970s (*11, 12*). Because the analysis uses matrix formulations, we use subscripts to separate the two chemical species, with x_1 as the number of mRNA molecules per cell and x_2 as the number of protein molecules. If transcription occurs with a constant rate, translation occurs with a constant rate per transcript, and both molecules decay exponentially, we have

$$\begin{array}{l} x_1 \xrightarrow{\lambda_1} x_1 + 1 \\ x_1 \xrightarrow{x_1/\tau_1} x_1 - 1 \end{array} \quad \text{and} \quad \begin{array}{l} x_2 \xrightarrow{\lambda_2 x_1} x_2 + 1 \\ x_2 \xrightarrow{x_2/\tau_2} x_2 - 1 \end{array} \quad \text{(A.6)}$$

which exactly gives

$$D = \begin{bmatrix} \frac{2}{\tau_1}\frac{1}{\langle x_1 \rangle} & 0 \\ 0 & \frac{2}{\tau_2}\frac{1}{\langle x_2 \rangle} \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} \frac{1}{\tau_1} & 0 \\ -\frac{1}{\tau_2} & \frac{1}{\tau_2} \end{bmatrix} \tag{A.7}$$

Multiplying the matrices, and solving for the covariances then simply results in

$$\overbrace{\frac{\sigma_2^2}{\langle x_2 \rangle^2}}^{\text{Total protein noise}} = \overbrace{\underbrace{\frac{1}{\langle x_2 \rangle}}_{\text{Poisson}}}^{\text{From individual births and deaths}} + \overbrace{\underbrace{\frac{1}{\langle x_1 \rangle}}_{\text{Poisson}} \times \underbrace{\frac{\tau_1}{\tau_2 + \tau_1}}_{\text{Time-averaging}}}^{\text{From spontaneous mRNA noise}} \tag{A.8}$$

Protein noise can thus come from low numbers of either the protein or the corresponding mRNA, although proteins partly average out mRNA fluctuations (low-pass filter) because protein levels cannot adjust immediately. Many studies have instead interpreted Eq. (A.8) in terms of bursts – brief periods of high expression intensity followed by long periods of low intensity – motivated by the fact that proteins often decay slowly relative to their transcripts, $\tau_2 >> \tau_1$, so that the Fano factor (variance over average) simplifies to:

$$\frac{\sigma_2^2}{\langle x_2 \rangle} = 1 + \frac{\langle x_2 \rangle}{\langle x_1 \rangle}\frac{\tau_1}{\tau_1 + \tau_2} \approx 1 + \frac{\langle x_2 \rangle}{\langle x_1 \rangle}\frac{\tau_1}{\tau_2} = 1 + \frac{\lambda_1\tau_1\lambda_2\tau_2}{\lambda_1\tau_1}\frac{\tau_1}{\tau_2} = 1 + \lambda_2\tau_1 = 1 + \langle b \rangle, \tag{A.9}$$

where b is the average number of translation events per transcript, on the order of 100 for an average bacterial gene. However, having $\tau_2 >> \tau_1$ and $\langle b \rangle >> 1$ is not a sufficient condition to get burst-like time series. For example, some mRNAs are present in hundreds of copies per cell at any given time, but are still so unstable that the approximation in Eq. (A.9) is close to exact even if protein synthesis is not burst-like at all: the conditions for a mathematically correct expression in terms of $\langle b \rangle$ are more relaxed than the conditions for a physically sound burst interpretation. However, several experiments have demonstrated gene expression close to the bursting limit.

4. External noise from linear response theory

Another strategy for calculating the mRNA contribution to the protein noise above is to use linear response theory. This theory states a simple rule for calculating the variance (or autocorrelation function) of the output of a linear system, from the autocorrelation function of the input. To use it for the example above, we thus first have to calculate the stationary autocorrelations of the mRNA fluctuations. This can be simply and exactly done using Eq. (A.4). Because mRNAs are not affected by proteins in the model, this is a one-variable process, so that Eq. (A.4) immediately becomes

$$\mathrm{A}_{11}(t) = \eta_{11} Exp(-t/\tau_1) = \frac{1}{\langle x_1 \rangle} e^{-t/\tau_1} \tag{A.10}$$

Next we calculate the transfer function for how mRNAs affect the protein dynamics. Because this again is a one-variable linear system (subjected to fluctuations in the mRNA), adjustment to steady

state is simply set by a single exponential, with a time-constant set by the average lifetime of the molecules, $\tau_2^{-1} e^{-t/\tau_2}$. The mRNA contribution to the protein noise now simply follows (see for example (*13*) for more details)

$$\int_0^\infty \underbrace{\frac{1}{\tau_2} e^{-\frac{t}{\tau_2}}}_{\substack{\text{Transfer function} \\ \text{of mRNA-proteins}}} \times \underbrace{\frac{1}{\langle x_1 \rangle} e^{-t/\tau_1}}_{\substack{\text{Autocorrelation function} \\ \text{of input mRNA fluctuations}}} dt = \underbrace{\frac{1}{\langle x_1 \rangle} \frac{\tau_1}{\tau_1 + \tau_2}}_{\substack{\text{mRNA contribution} \\ \text{to protein variance}}} \tag{A.11}$$

This calculation gives exactly the same result as the moment equations above, as they must: when the dynamics is linear, both methods are exact. For linear Markov processes, linear response theory is a convenient alternative method of calculating how much of the underlying fluctuations that are averaged out by the system.

5. General mRNA bursts

Here we return to the moment equations in (A.4), and show how the diffusion matrix *D* can be rewritten in terms of the burstiness of the underlying process, without introducing any approximations. For one-variable processes, the definition of *D* follows immediately from *B* in Eq. (A.2):

$$D = \frac{\sum_k r_k \left(s^k\right)^2}{\langle x \rangle^2} \tag{A.12}$$

Note that the superscript *k* is simply an index that keeps track of the reactions, and should not be read as an exponent. Here we rewrite this equation for the special case that the deaths occur one molecule at a time in exponential decay, while births can be made in arbitrary steps, as in Eq. (1) of the main text. We can then write

$$\sum_k r_k \times \left(s^k\right)^2 = \frac{\langle x \rangle}{\tau} + \sum_{\substack{\text{synthesis} \\ \text{reactions}}} r_k \times \left(s^k\right)^2 \tag{A.13}$$

The second term is thus summed over all possible synthesis reactions, regardless of the number of molecules made. We then write

$$\sum r_k \times \left(s^k\right)^2 = \sum r_k \times \sum \left(\frac{r_k}{\sum r_k} \right) \times \left(s^k\right)^2 \tag{A.14}$$

The ratio between the rate (probability per second) of synthesis reaction *k* and the sum of all synthesis reactions is simply the probability ρ_k that the next synthesis reaction is reaction number *k*

$$\rho_k = \frac{r_k}{\sum r_k} \tag{A.15}$$

We can thus identify the average square burst size as

$$\sum \rho_k \times \left(s^k\right)^2 = \langle b^2 \rangle = \langle b \rangle^2 + \sigma_b^2 \tag{A.16}$$

where the second step follows from the definition of variances. The price we paid in Eq. (A.14) is that we also multiplied by the sum of all synthesis reaction rates, $\sum r_k$. However, following a similar trick we can now write

$$\sum r_k = \sum r_k s^k \times \frac{\sum r_k}{\sum r_k s^k} \tag{A.17}$$

where

$$\frac{\sum r_k s^k}{\sum r_k} = \langle b \rangle \tag{A.18}$$

similarly to (A.16). Finally, we identify that $\sum r_k s^k$ is the sum of all synthesis fluxes, which at the average steady state perfectly balances the average degradation reaction. Hence

$$\sum r_k s^k = \frac{\langle x \rangle}{\tau} \tag{A.19}$$

This means that

$$\sum r_k = \frac{\langle x \rangle}{\tau} \frac{1}{\langle b \rangle} \tag{A.20}$$

Putting everything together in Eq. (A.12) gives

$$D = \frac{\frac{\langle x \rangle}{\tau} + \left(\langle b \rangle^2 + \sigma_b^2 \right) \frac{\langle x \rangle}{\tau} \frac{1}{\langle b \rangle}}{\langle x \rangle^2} = \frac{1}{\tau} \frac{1 + \langle b \rangle \left(1 + \frac{\sigma_b^2}{\langle b \rangle^2} \right)}{\langle x \rangle} \tag{A.21}$$

Solving the one-variable moment equation now gives

$$\eta_{11} = \frac{\sigma_1^2}{\langle x_1 \rangle^2} = \frac{D}{2M} = \frac{D}{2/\tau} = \frac{1}{\langle x \rangle} \times \underbrace{\left(1 + \langle b \rangle \left(1 + \frac{\sigma_b^2}{\langle b \rangle^2} \right) \right)}_{\substack{\text{Coarse-graining factor for} \\ \text{exponential waiting times}}} \tag{A.22}$$

Solving the two-variable process for mRNAs and proteins immediately gives the coarse-graining factor for the protein in Eq. (1) of the main text, in the case for exponential waiting times where $\sigma_T = \langle T \rangle$, as can be seen by simply replacing D_{11} in Eq. (A.7) by D in Eq. (A.21).

6. Using master equations for non-exponential waiting times

Before deriving the general Eq. (1) we discuss why non-exponential events cannot be conveniently summarized by CMEs. Each step of a CME is memoryless – the probability of jumping only depends

on the state itself, not on when the system (by chance) entered that state[1]. It is still possible to model non-exponential waiting times between using CMEs, but these require an extended state space. For example assume that the waiting time distribution corresponds to a sum of two exponentials. It is then possible to introduce a pseudo-variable for the intermediate, making each sub-step exponential. However, the results from the master equation are then no longer functions of the original variables, but rather come out in terms of the details of the extended state space. This problem can be fixed by introducing a second pseudo-variable that 'contracts' other variables. For example, if we are interested in the variance in the sum of two (dependent) variables X_1 and X_2 in the CME, we can define a new variable X_3 that satisfies the following criteria:

(1) Its synthesis rate is proportional to the sum of x_1 and x_2, i.e., $k(x_1+x_2)$.

(2) The degradation rate is set to kx_3.

(3) The rate constant k goes to infinity, and the corresponding elements in the diffusion matrix D (all elements in row or column 3 in this case) are set to zero.

These assumptions ensure that X_3 immediately and deterministically tracks changes in the sum of X_1 and X_2. Using this trick, we can use the straightforward moment equations above to derive Eq. (1) for a range of non-exponential switching times, for example when the time between transcription events is the sum of many (non-identical) exponential steps. However, in many cases the switching time distribution is more complicated than simply a sum of exponentials, as in Fig. 2 in the main text. The approach above then requires complicated extensions of the state space, requiring infinite complications to account for arbitrary switching times. The extreme clumsiness and uselessness of this approach reflects the fact that the CME is formulated in terms of exponential waiting times between events, and is not an appropriate framework for non-exponential waiting times.

7. General timing and bursting statistics

To obtain the stationary mRNA noise for a process with arbitrary bursting and timing, we first examine a case with an arbitrary distribution of creation times but no bursts nor decay, then introduce the bursting complication, and finally add decay. Let *f(t)* be the probability distribution of a creation event happening since the last even happened (i.e., an event is assumed to have happened at $t = 0$), and let n be the total number of creation events since $t = 0$.

$$
\begin{aligned}
&P(t=T)dt \equiv f(t)dt \\
&P(t<T) = \int_0^T f(t)dt \equiv F(t) \\
&P(n=0 \mid t=T) = P(t>T) = 1-\int_0^T f(t)dt = 1-F(t) \\
&P(n=1 \mid t=T) = \int_{t_1} P(t=t_1)P(t>T-t_1)dt_1 = \int_0^T f(t_1)\big(1-F(T-t_1)\big)dt_1 = f*(1-F)\,|_T
\end{aligned}
\tag{A.23}
$$

Analogously, it can be shown by induction that

[1] In some cases, there may be an explicit time-dependence in the equations, as when there is deterministic seasonality to the dynamics, where the intensities change predictably in real time. Such master equations are said to be non-homogenous and corresponds to very different principles than the statistical memory that we describe here.

$$P(n = N \mid t = T) = f * P(n = N - 1 \mid t)\,|_T . \tag{A.24}$$

Using the Laplace transform, $L\big(f(t)\big) \equiv \hat{f}(s)$,

$$\begin{aligned} \hat{F}(s) &= \frac{1}{s}\hat{f}(s) \\ \hat{P}(0,s) &= \frac{1}{s}(1 - \hat{f}(s)) \\ \hat{P}(n,s) &= \hat{f}(s)\hat{P}(n-1,s) = \frac{1}{s}\hat{f}^{\,n}(s)(1 - \hat{f}(s)) \end{aligned} \tag{A.25}$$

Defining the moment generating function as

$$G(z,t) = \sum_{n=0}^{\infty} z^n P(n,t)\,, \tag{A.26}$$

with the usual properties

$$G(1,t) = 1, \qquad \frac{\partial}{\partial z} G(z,t)\bigg|_1 = \langle n \rangle_{(t)}, \qquad \frac{\partial^2}{\partial z^2} G(z,t)\bigg|_1 = \left\langle n^2 - n \right\rangle_{(t)} \tag{A.27}$$

we can obtain its Laplace transform as

$$\hat{G}(z,s) = \sum_{n=0}^{\infty} z^n \hat{P}(n,s) = \frac{1}{s}(1 - \hat{f}(s)) \sum_{n=0}^{\infty} z^n \hat{f}^{\,n}(s) = \frac{1}{s}\frac{1 - \hat{f}(s)}{1 - z\hat{f}(s)}\,. \tag{A.28}$$

Note that the expansion is valid since $\hat{f}(s) \le 1$ and z will be evaluated at 1.
We can thus obtain the moments in Laplace space as

$$\begin{aligned} \langle \hat{n} \rangle_{(s)} &= \frac{\partial}{\partial z}\hat{G}(z,s)\bigg|_1 = \frac{1}{s}\frac{\hat{f}(s)}{1 - \hat{f}(s)} \\ \left\langle \hat{n}^2 \right\rangle_{(s)} - \langle \hat{n} \rangle_{(s)} &= \frac{\partial^2}{\partial z^2}\hat{G}(z,s)\bigg|_1 = \frac{2}{s}\left(\frac{\hat{f}(s)}{1 - \hat{f}(s)}\right)^2 \end{aligned} \tag{A.29}$$

From the definition of the Laplace transform, it is easy to see that

$$\hat{f}(0) = 1, \qquad \frac{d}{ds}\hat{f}(s)\bigg|_0 = -\langle t \rangle, \qquad \frac{d^2}{ds^2}\hat{f}(s)\bigg|_0 = \left\langle t^2 \right\rangle \tag{A.30}$$

This implies that $1-\hat{f}(s)\to 0$ as $s\to 0$, so to inverse-transform the mean we need the second order residue at $s=0$:

$$\begin{aligned}\langle n\rangle_{(t)} &= L^{-1}\left(\langle \hat{n}\rangle_{(s)}\right)=\frac{1}{2\pi i}\oint e^{st}\frac{1}{s}\frac{\hat{f}(s)}{1-\hat{f}(s)}ds=\mathrm{Res}_2\left(\frac{e^{st}}{s}\frac{\hat{f}(s)}{1-\hat{f}(s)}\right)_{s=0}=\frac{d}{ds}\left(\frac{se^{st}\hat{f}(s)}{1-\hat{f}(s)}\right)_{s=0}\\ &=\frac{t}{-\hat{f}'(0)}+\frac{\hat{f}''(0)}{2(\hat{f}'(0))^2}-1=\frac{t}{\langle t\rangle}+\left(\frac{\langle t^2\rangle}{2\langle t\rangle^2}-1\right)\equiv\frac{t}{\langle t\rangle}+C_1\end{aligned} \tag{A.31}$$

For the second moment, we require the third order residue:

$$\begin{aligned}\langle n^2-n\rangle_{(t)} &= L^{-1}\left(\langle \hat{n}^2-\hat{n}\rangle_{(s)}\right)=\frac{1}{2\pi i}\oint e^{st}\frac{2}{s}\left(\frac{\hat{f}(s)}{1-\hat{f}(s)}\right)^2 ds=\mathrm{Res}_3\left(\frac{2e^{st}}{s}\left(\frac{\hat{f}(s)}{1-\hat{f}(s)}\right)^2\right)_{s=0}\\ &=\frac{1}{2}\frac{d^2}{ds^2}\left(2e^{st}\left(\frac{s\hat{f}(s)}{1-\hat{f}(s)}\right)^2\right)_{s=0}=\frac{t^2}{\langle t\rangle^2}+\frac{4t}{\langle t\rangle}\left(\frac{\langle t^2\rangle}{2\langle t\rangle^2}-1\right)+2\left(1-\frac{\langle t^2\rangle}{\langle t\rangle^2}+\frac{3\langle t^2\rangle^2}{4\langle t\rangle^4}+\frac{\langle t^3\rangle}{3\langle t\rangle^3}\right)\end{aligned} \tag{A.32}$$

which in turn implies that

$$\sigma_n^2=\langle n^2\rangle_{(t)}-\langle n\rangle^2{}_{(t)}=\frac{t}{\langle t\rangle}\left(\frac{\langle t^2\rangle}{\langle t\rangle^2}-1\right)+\left(-\frac{\langle t^2\rangle}{2\langle t\rangle^2}+\frac{5\langle t^2\rangle^2}{4\langle t\rangle^4}-\frac{2\langle t^3\rangle}{3\langle t\rangle^3}\right)\equiv\frac{t}{\langle t\rangle}\left(\frac{\langle t^2\rangle}{\langle t\rangle^2}-1\right)+C_2 \tag{A.33}$$

Note that for exponential times, $\eta_t=1$ and $C_1=C_2=0$. The constant terms arise from the assumption that an event happened at t = 0. This is exact for the exponential (in the sense that it doesn't affect the likelihood of following events) when describing the system from an arbitrary starting point. For other time distributions this introduces an approximation which diminishes in importance with time, and for a system with decay (as will be considered here) the original time can be assumed far into the past.

To include the effect of bursts of creation, let b_i be the random variables chosen from the distribution of burst sizes and consider the random variable

$$x=\sum_{i=0}^{n}b_i \tag{A.34}$$

where the number of terms in the sum is given by the random number of creation events for a given time interval. The probability of having a total number of events a is given by

$$P_{Tot}(x=a)=\sum_{n=0}^{\infty}P_{Tot}(x=a\,|\,n)P(n) \tag{A.35}$$

Its characteristic function $\wp(r)$ is given by

$$\begin{aligned}\wp(r) &= \left\langle e^{xr} \right\rangle_{Tot} = \sum_{a=0}^{\infty} e^{xr} P_{Tot}(x=a) = \sum_{a=0}^{\infty} e^{xr} \sum_{n=0}^{\infty} P_{Tot}(x=a \mid n) P(n) \\ &= \sum_{n=0}^{\infty} \sum_{a=0}^{\infty} e^{xr} P_{Tot}(x=a \mid n) P(n) = \sum_{n=0}^{\infty} \left\langle e^{xr} \right\rangle_{Tot|n} P(n) = \sum_{n=0}^{\infty} \left\langle e^{br} \right\rangle_b^n P(n) = \left\langle \left\langle e^{br} \right\rangle_b^n \right\rangle_e\end{aligned} \tag{A.36}$$

where $\langle\ \rangle_b$ and $\langle\ \rangle_e$ indicate the expectation with respect to the distribution of bursts and events, respectively. From the characteristic function we can obtain the moments as follows:

$$\begin{aligned}\left\langle x \right\rangle &= \frac{d}{dr}\wp(r)\bigg|_0 = \frac{d}{dr}\left\langle \left\langle e^{br} \right\rangle_b^n \right\rangle_e\bigg|_0 = \left\langle n \left\langle e^{br} \right\rangle_b^{n-1} \left\langle x e^{br} \right\rangle_b \right\rangle_e\bigg|_0 = \left\langle n \left\langle b \right\rangle_b \right\rangle_e = \left\langle n \right\rangle_e \left\langle b \right\rangle_b \\ \left\langle x^2 \right\rangle &= \frac{d^2}{dr^2}\wp(r)\bigg|_0 = \left\langle n(n-1) \left\langle e^{br} \right\rangle_b^{n-2} \left\langle b e^{br} \right\rangle_b^2 + n \left\langle e^{br} \right\rangle_b^{n-1} \left\langle x^2 e^{br} \right\rangle_b \right\rangle_e\bigg|_0 = \left\langle (n^2 - n) \right\rangle_e \left\langle b \right\rangle_b^2 + \left\langle n \right\rangle_e \left\langle b^2 \right\rangle_b \\ \sigma_x^2 &= \sigma_n^2 \left\langle b \right\rangle_b^2 + \left\langle n \right\rangle_e \sigma_b^2\end{aligned} \tag{A.37}$$

Defining the noise as standard deviation over average (coefficient of variation), its square is

$$\eta_x^2 \equiv \frac{\sigma_x^2}{\left\langle x \right\rangle^2} = \frac{\sigma_n^2 \left\langle b \right\rangle_b^2 + \left\langle n \right\rangle_e \sigma_b^2}{\left(\left\langle b \right\rangle_b \left\langle n \right\rangle_e \right)^2} = \eta_n^2 + \frac{1}{\left\langle n \right\rangle} \eta_b^2 \tag{A.38}$$

Using expressions (A.31) and (A.33),

$$\eta_n^2 = \frac{\sigma_n^2}{\left\langle n \right\rangle^2} = \frac{\eta_t^2}{\left\langle n \right\rangle} \tag{A.39}$$

which implies

$$\eta_x^2 = \frac{\eta_t^2}{\left\langle n \right\rangle} + \frac{\eta_b^2}{\left\langle n \right\rangle} = \frac{\left\langle b \right\rangle (\eta_t^2 + \eta_b^2)}{\left\langle x \right\rangle} \tag{A.40}$$

with

$$\left\langle x \right\rangle = \left\langle b \right\rangle \frac{t}{\left\langle t \right\rangle}. \tag{A.41}$$

This result holds for a pure birth process, and next we account for decay of molecules. We next calculate the effect of decay by introducing an uneven instantaneous binomial partitioning. With $P_{Dr}(m \mid n)$ as the probability of finding *m* molecules in volume fraction *r* given *n* molecules before division, then

$$P_{Dr}(m \mid n) = \binom{n}{m} r^n (1-r)^{n-m} . \tag{A.42}$$

Let $P_{Br}(n)$ be the probability of having *n* molecules before division and $P_{Ar}(m)$ be the probability of having *m* molecules after division. For a fixed volume fraction at division, we have

$$P_{Ar}(m) = \sum_{n=0}^{\infty} P_{Dr}(m \mid n) P_{Br}(n) = \sum_{n=0}^{\infty} \binom{n}{m} r^n (1-r)^{n-m} P_{Br}(n) , \tag{A.43}$$

with moment generating function

$$G_{Ar}(z) = \sum_{m=0}^{\infty} z^m \sum_{n=0}^{\infty} P_{Dr}(m \mid n) P_{Br}(n) = \sum_{n=0}^{\infty} (zr + (1-r))^n P_{Br}(n) . \tag{A.44}$$

Since the creation probability is independent of the current number of molecules, the distribution at the end of a growth stage $P_{Br}(n)$ will be the convolution of the distribution at the beginning $P_{Ar}(n)$ and the distribution from the creation process $P_{C\tau}(n) \equiv P_{Tot}(n)|_{t=\tau}$. Thus the generating function for the distribution before division will be the product of the generating function after division $G_{Ar}(z)$ and the generating function for the creation process $G_{C\tau}(z)$:

$$G_{Br}(z) = G_{Ar}(z) G_{C\tau}(z) . \tag{A.45}$$

We can now obtain the moments by differentiation as before. The averages before and after division are

$$\langle n \rangle_{Br} = \frac{\partial}{\partial z} G_{Br}(z)\Big|_{z=1} = G_{Ar}(1) \frac{\partial}{\partial z} G_{C\tau}(z)\Big|_{z=1} + \frac{\partial}{\partial z} G_{Ar}(z)\Big|_{z=1} G_{C\tau}(1) = \langle n \rangle_{C\tau} + \langle n \rangle_{Ar}$$

$$\langle n \rangle_{Ar} = \frac{\partial}{\partial z} G_{Ar}(z)\Big|_{z=1} = \frac{\partial}{\partial z} \sum_{n=0}^{\infty} (zr + (1-r))^n P_{Br}(n)\Big|_{z=1} = \sum_{n=0}^{\infty} nr P_{Br}(n) = r \langle n \rangle_{Br} \tag{A.46}$$

which, together, imply

$$\langle n \rangle_{Br} = \frac{1}{1-r} \langle n \rangle_{C\tau} \text{ and } \langle n \rangle_{Ar} = r \langle n \rangle_{Br} = \frac{r}{1-r} \langle n \rangle_{C\tau} . \tag{A.47}$$

The variances are obtained from

$$\begin{aligned} \langle n^2 \rangle_{Br} - \langle n \rangle_{Br} &= \frac{\partial^2}{\partial z^2} G_{Br}(z)\Big|_{z=1} = G_{Ar}(1) \frac{\partial^2}{\partial z^2} G_{C\tau}(z)\Big|_{z=1} + 2 \frac{\partial}{\partial z} G_{Ar}(z)\Big|_{z=1} \frac{\partial}{\partial z} G_{C\tau}(z)\Big|_{z=1} + \frac{\partial^2}{\partial z^2} G_{Ar}(z)\Big|_{z=1} G_{C\tau}(1) \\ &= \left(\langle n^2 \rangle_{C\tau} - \langle n \rangle_{C\tau} \right) + 2 \langle n \rangle_{Ar} \langle n \rangle_{C\tau} + \left(\langle n^2 \rangle_{Ar} - \langle n \rangle_{Ar} \right) \end{aligned} \tag{A.48}$$

$$\left\langle n^2\right\rangle_{Ar} - \left\langle n\right\rangle_{Ar} = \frac{\partial^2}{\partial z^2} G_{Ar}(z)\big|_{z=1} = \frac{\partial^2}{\partial z^2}\sum_{n=0}^{\infty}(zr+(1-r))^n P_{Br}(n)\bigg|_{z=1} = \sum_{n=0}^{\infty} n(n-1)r^2 P_{Br}(n) = r^2\left(\left\langle n^2\right\rangle_{Br} - \left\langle n\right\rangle_{Br}\right)$$

which imply

$$\begin{aligned}\sigma^2{}_{Br} &= \sigma^2{}_{C\tau} + \left\langle n\right\rangle^2_{C\tau} - \left\langle n\right\rangle_{C\tau} + 2r(1-r)\left\langle n\right\rangle^2_{Br} + r^2\left(\left\langle n^2\right\rangle_{Br} - \left\langle n\right\rangle_{Br}\right) + \left\langle n\right\rangle_{Br} - \left\langle n\right\rangle^2_{Br} \\ &= \sigma^2{}_{C\tau} + (1-r)^2\left\langle n\right\rangle^2_{Br} - (1-r)\left\langle n\right\rangle_{Br} + 2r(1-r)\left\langle n\right\rangle^2_{Br} + r^2\sigma^2{}_{Br} + (r^2-1)\left\langle n\right\rangle^2_{Br} + (1-r^2)\left\langle n\right\rangle_{Br} \\ &= \frac{1}{1-r^2}\left[\sigma^2{}_{C\tau} + r(1-r)\left\langle n\right\rangle_{Br}\right] = \frac{1}{1-r^2}\sigma^2{}_{C\tau} + \frac{r}{1+r}\left\langle n\right\rangle_{Br}\end{aligned}$$

$$\eta^2{}_{Br} = \frac{1-r}{1+r}\eta^2{}_{C\tau} + \frac{r}{1+r}\frac{1}{\left\langle n\right\rangle_{Br}} \tag{A.49}$$

$$\begin{aligned}\sigma^2{}_{Ar} &= r^2\left(\left\langle n^2\right\rangle_{Br} - \left\langle n\right\rangle_{Br}\right) + \left\langle n\right\rangle_{Ar} - \left\langle n\right\rangle^2_{Ar} = \frac{r^2}{1-r^2}\sigma^2{}_{C\tau} + \frac{r^3}{1+r}\left\langle n\right\rangle_{Br} + r^2\left\langle n\right\rangle^2_{Br} - r^2\left\langle n\right\rangle_{Br} + \left\langle n\right\rangle_{Ar} - \left\langle n\right\rangle^2_{Ar} \\ &= \frac{r^2}{1-r^2}\sigma^2{}_{C\tau} + \frac{r^2}{1+r}\left\langle n\right\rangle_{Ar} + \left\langle n\right\rangle^2_{Ar} - r\left\langle n\right\rangle_{Ar} + \left\langle n\right\rangle_{Ar} - \left\langle n\right\rangle^2_{Ar} = \frac{r^2}{1-r^2}\sigma^2{}_{C\tau} + \frac{1}{1+r}\left\langle n\right\rangle_{Ar}\end{aligned}$$

$$\eta^2{}_{Ar} = \frac{1-r}{1+r}\eta^2{}_{C\tau} + \frac{1}{1+r}\frac{1}{\left\langle n\right\rangle_{Ar}} \tag{A.50}$$

Incidentally, this includes the case of normal dilution by division ($r = \frac{1}{2}$, $\tau =$ generation time, $\left\langle n\right\rangle_B \equiv \left\langle n\right\rangle_{Max} = 2\left\langle n\right\rangle_A \equiv 2\left\langle n\right\rangle_{Min}$), where

$$\begin{aligned}\eta^2{}_{\text{Before division}} &= \frac{1}{3}\eta^2{}_{C\tau} + \frac{1}{3}\frac{1}{\left\langle n\right\rangle_{Max}} \\ \eta^2{}_{\text{After division}} &= \frac{1}{3}\eta^2{}_{C\tau} + \frac{2}{3}\frac{1}{\left\langle n\right\rangle_{Min}} \\ &\Rightarrow \\ \eta^2{}_{\text{After division}} &= \eta^2{}_{\text{Before division}} + \frac{1}{\left\langle n\right\rangle_{Max}}\end{aligned} \tag{A.51}$$

We can now approximate the process of continuous decay as division in the limit of small lost volume, $r \to 1$. Since the number of created molecules corresponds to the number to be replaced,

$$\left\langle n\right\rangle_{C\tau} = \left\langle n\right\rangle_B (1-r)\,. \tag{A.52}$$

From equation (A.40),

$$\eta^2_{C\tau} = \frac{\left\langle b\right\rangle(\eta_t^2 + \eta_b^2)}{\left\langle n\right\rangle_{C\tau}} = \frac{1}{1-r}\frac{\left\langle b\right\rangle(\eta_t^2 + \eta_b^2)}{\left\langle n\right\rangle_{Br}}, \tag{A.53}$$

so in the limit $r \to 1$ we have

$$\lim_{r\to 1} \langle n \rangle_{Ar} = \lim_{r\to 1} r\langle n \rangle_{Br} = \langle n \rangle \tag{A.54}$$

and

$$\eta^2 = \lim_{r\to 1} \eta^2{}_{Br} = \lim_{r\to 1} \frac{1}{1+r}\frac{\langle b \rangle(\eta_t^2+\eta_b^2)}{\langle n \rangle_{Br}} + \frac{r}{1+r}\frac{1}{\langle n \rangle_{Br}} = \frac{\langle b \rangle(\eta_t^2+\eta_b^2)+1}{2\langle n \rangle} \tag{A.55}$$

This corresponds to the coarse-graining factor in Eq. (1) of the main text. The autocorrelation function is effectively unchanged and still follows a single exponential: the process can still be summarized by a single variable without time-correlated input noise, because subsequent events are independent. This can only provide a slight deviation, and only at such low average numbers that the time-interval between synthesis events is comparable to the lifetime of the molecules. Even then the effect is small because the input noise comes from both the synthesis and degradation events, and the latter are exponential even when the former are perfectly precise. In other words, because the average dynamics is unchanged, with a single exponential for adjustment to steady state, and the synthesis events are independent, the autocorrelations are still exponential with a time-constant set by the average lifetime.

8. Senescence of mRNA

The result in Eq. (2) can be derived in either of two ways. The simplest way is perhaps to calculate the autocorrelation function for the total mRNA level, using a combination of the tricks in sections 5 and 6 above.

Assume that mRNAs are made at a constant rate and that they senesce through a series of states before finally degrading. If the process goes through a series of (identical) exponential steps, the lifetimes will be Gamma distributed. This can be recreated in the master equation formalism by letting x_1 be the number of 'newborn' transcripts per cell, which then go through a series of conversions, as

$$\textit{Transcription} \to \mathrm{X}_1 \to \mathrm{X}_2 \to ... \to \mathrm{X}_\mathrm{N} \to \textit{Degradation}$$

The stochastic jumps for the mRNAs are thus

$$\begin{aligned} &x_1 \xrightarrow{\lambda_1} x_1 + 1 \\ &\{x_i, x_{i+1}\} \xrightarrow{\beta_1 x_i} \{x_i - 1, x_{i+1} + 1\} \quad \text{for } 0 < i < n \\ &x_n \xrightarrow{\beta_1 x_n} x_n - 1 \end{aligned} \tag{A.56}$$

and the number of proteins per cell, now called x_{n+1} for convenience, in turn changes in the reactions

$$\begin{aligned} &x_{n+1} \xrightarrow{\lambda_2 \sum_{i=1}^{n} x_n} x_{n+1} + 1 \\ &x_{n+1} \xrightarrow{\beta_2 x_{n+1}} x_{n+1} - 1 \end{aligned} \tag{A.57}$$

The averages thus follow

$$\begin{aligned}
\frac{d}{dt}\langle x_1\rangle &= \lambda_1 - \beta_1\langle x_1\rangle \\
\frac{d}{dt}\langle x_{i+1}\rangle &= \beta_1(\langle x_i\rangle - \langle x_{i+1}\rangle) \quad \text{for } 0 < i < n \\
\frac{d}{dt}\langle x_{n+1}\rangle &= \lambda_2\sum_{i=1}^{n}\langle x_i\rangle - \beta_2\langle x_{n+1}\rangle
\end{aligned} \tag{A.58}$$

From symmetry we get that the lifetimes and steady state averages of the different conformations of the mRNA follow

$$\begin{aligned}
\langle x_i\rangle &= \frac{\lambda_1}{\beta_1} \quad \text{for } 0 < i \le n \\
\tau_i &= \frac{1}{\beta_1} \qquad \text{for } 0 < i \le n
\end{aligned} \tag{A.59}$$

If we define the total mRNA level as $m = \sum_{i=1}^{n} x_i$, then

$$\begin{aligned}
\langle m\rangle &= \lambda_1\tau_0 \\
\tau_m &= \frac{n}{\beta_1}
\end{aligned} \tag{A.60}$$

The symmetry of the dynamics and the compounded variable x_0 give rise to simple drift and diffusion matrices *M* and *D*. To simplify notation in the final answer, we return to $x_{n+1} = p$ for the number of protein molecules per cell. If the mRNA goes through *n* conformations before finally degrading, then the $(n+1)\times(n+1)$ matrix is defined by

$$\begin{aligned}
M_{ii} &= \frac{n}{\tau_m} \qquad \text{for } 1 \le i \le n \\
M_{i+1,i} &= -\frac{n}{\tau_m} \quad \text{for } 1 \le i \le n \\
M_{n+1,n+1} &= \frac{1}{\tau_p} \\
M_{n+1,i} &= \frac{1}{n\tau_p} \qquad \text{for } 1 \le i \le n \\
M_{ij} &= 0 \qquad \text{for all other } i \text{ and } j
\end{aligned} \tag{A.61}$$

The matrix thus has a non-zero main diagonal, a non-zero diagonal below the main, and a non-zero last row, while all other entries are zero. The (symmetric) diffusion matrix *D* is equally sparse, and defined by

$$
\begin{aligned}
D_{ii} &= \frac{2}{\tau_i}\frac{1}{\langle x_i \rangle} = \frac{2}{\tau_m}\frac{n^2}{\langle m \rangle} && \text{for } 1 \le i \le n \\
D_{i+1,i} &= D_{i,i+1} = -\frac{1}{\tau_i}\frac{1}{\langle x_i \rangle} = -\frac{1}{\tau_i}\frac{n^2}{\langle m \rangle} && \text{for } 1 \le i \le n \\
D_{n+1,n+1} &= \frac{(+1)(-1)\beta_i \langle x_i \rangle}{\langle x_i \rangle \langle x_{i+1} \rangle} = -\frac{1}{\tau_m}\frac{n^2}{\langle m \rangle} && \\
D_{ij} &= 0 && \text{for all other } i \text{ and } j
\end{aligned}
\tag{A.62}
$$

The upper left $n \times n$ block is thus a tridiagonal band matrix, while the last row and last column are all zeroes, except for the last element on the main diagonal. Because the matrices are so simple, we can solve the matrix equations in Eq. (A.4) for the general case. This step is algebraic, but simply involves matrix-matrix multiplication and addition, and solving a linear equation system. This generates Eq. (2) in the main body of the paper. The corresponding result for the total mRNA level is simply found by assuming an infinite number of proteins (1st term in Eq. (2) of main paper goes to zero), and taking the limit where proteins are turned over infinitely fast compared to the mRNA. The protein then perfectly tracks the sum of all mRNA molecules, and thus has the same normalized variance. A similar approach can be used to shed some light on the physical processes at play. By again letting proteins be deterministic and infinitely fast (corresponding to the pseudo-variable trick in section 6 above), we can take the matrix exponent to calculate the autocorrelation function of the total mRNA level. Because of the bi-diagonal structure of the upper left $n \times n$ block, this is a straightforward operation, and the normalized autocorrelation function of the mRNA can be shown to follow:

$$
\frac{A_m}{\sigma_m^2 / \langle m \rangle^2} = \frac{e^{-n\frac{t}{\tau_m}}}{n} \sum_{j=0}^{n-1} (n-j) \frac{\left(n\frac{t}{\tau_m}\right)^j}{j!}
\tag{A.63}
$$

It is thus an exponential multiplied by a polynomial of order $n - 1$, as expected from the serial degradation mechanism. We can also define an autocorrelation time as the integral of the normalized autocorrelation function:

$$
\begin{aligned}
\tau_{auto} &= \int_0^\infty \frac{e^{-n\frac{t}{\tau_m}}}{n} \sum_{j=0}^{n-1} (n-j) \frac{\left(n\frac{t}{\tau_m}\right)^j}{j!} dt = \frac{1}{n} \sum_{j=0}^{n-1} \frac{(n-j)}{j!} n^j \int_0^\infty e^{-n\frac{t}{\tau_m}} \left(\frac{t}{\tau_m}\right)^j dt = \\
&\frac{1}{n} \sum_{j=0}^{n-1} \frac{(n-j)}{j!} n^j \frac{j!}{n^{j+1}} \tau_m = \frac{1}{n^2} \sum_{j=0}^{n-1} n - j = \frac{n(n+1)/2}{n^2} \tau_m = \frac{n+1}{2n} \tau_m
\end{aligned}
\tag{A.64}
$$

For $n = 1$, the autocorrelations are simply exponential e^{-t/τ_m} with $\tau_{auto} = \tau_m$, but when the lifetime distribution narrows, we get:

$$\lim_{n\to\infty}\frac{e^{-n\frac{t}{\tau_m}}}{n}\sum_{j=0}^{n-1}(n-j)\frac{\left(n\frac{t}{\tau_m}\right)^j}{j!}=1-\frac{t}{\tau_m} \qquad \text{(A.65)}$$

$$\lim_{n\to\infty}\tau_{auto}=\frac{1}{2}\tau_m$$

This shows that, in the limit where each individual molecule lives for the same time, without significant statistical fluctuations, the autocorrelations drop linearly to zero. As can be seen in the figure below, the autocorrelations for a multistep decay function thus drop to zero more quickly than for single step decay, when compared for the *same average lifetime* (see Fig A1). This means that the mRNA dynamics effectively speed up, without changing the overall turnover rate of the molecules, thus providing speed without need for increased turnover or nonlinear control loops.

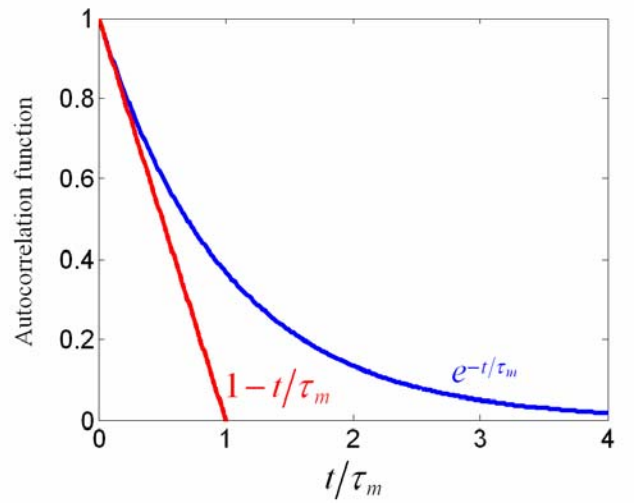

Figure A1 Normalized autocorrelation function for fluctuations in mRNA levels vs normalized time. The blue curve corresponds to exponentially distributed lifetimes of the individual molecules, and the red curve corresponds to the limit when each mRNA lives for exactly the same amount of time (delta function). The area under the curves (τ_{auto}) is twice as large for the blue curve.

We can also calculate the reverse scenario, by assuming that proteins instead go through multiple steps. This follows exactly the same strategy as above, and the exact results are used in Fig. 3 in the main text. We have not yet found a way to simplify the general expression, but with the simplified notation that $\tau \equiv \tau_m/\tau_p$, the result for *N* steps in the degradation pathway is

$$
\begin{aligned}
N=1 &\Rightarrow \frac{\sigma_p^2}{\langle p\rangle^2}=\frac{1}{\langle p\rangle}+\frac{1}{\langle m\rangle}\times\frac{\tau}{1+\tau} \\
N=2 &\Rightarrow \frac{\sigma_p^2}{\langle p\rangle^2}=\frac{1}{\langle p\rangle}+\frac{1}{\langle m\rangle}\times\frac{\tau(5+16\tau)}{4(1+2\tau)^2} \\
N=3 &\Rightarrow \frac{\sigma_p^2}{\langle p\rangle^2}=\frac{1}{\langle p\rangle}+\frac{1}{\langle m\rangle}\times\frac{\tau(11+91\tau+216\tau^2)}{8(1+3\tau)^3} \\
N=4 &\Rightarrow \frac{\sigma_p^2}{\langle p\rangle^2}=\frac{1}{\langle p\rangle}+\frac{1}{\langle m\rangle}\times\frac{\tau(93+1424\tau+7824\tau^2+16384\tau^3)}{64(1+4\tau)^4} \\
&\ldots
\end{aligned}
\tag{A.66}
$$

9. Parameters used in the figures

The parameters used in Figure 2 – both for the Monte Carlo simulations and for the analytical expression (from Eq. 1) – are as follows:

γ_m	log(2)/5	Decay rate of mRNA (min^{-1})
γ_p	log(2)/30	Decay rate of protein (min^{-1})
$k_m=\frac{1}{\langle t_m\rangle}$	0.4 log(2)→ 2.4 log(2)	Average creation rate of mRNA (min^{-1})
k_p	10 log(2)	Average creation rate of protein per mRNA (min^{-1})
$\eta_t=\frac{\sigma_t}{\langle t\rangle}$	0.8, 1, 1.2	Noise in the timing of mRNA creation events
b	1	Number of mRNAs created per creation event (“bursts”)

Table S.1 Biochemical parameters used in the simulations

These parameters result in average mRNA copy numbers $\langle m\rangle$ between 2 and 12, average protein numbers $\langle p\rangle$ between 600 and 3600, an average of 50 proteins produced by each mRNA, and a ratio between the mRNA and protein lifetimes of 1/6. They were chosen as representative for *Escherichia coli*. Similar fits were obtained for a broad range of distributions and parameter values.

The distributions of creation times are given by

$$
P_{\text{Gamma}}(x)=\frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\theta^{\alpha}} \tag{A.67}
$$

with $\alpha = \eta_t^{-2}$ and $\theta = \dfrac{\langle t \rangle}{\alpha}$, and

$$P_{\text{Bimodal}}(x) = Ae^{\frac{-(x-(1-0.7f)\bar{t})^2}{2\sigma_a^2}} + Be^{\frac{-(x-(1+1.3f)\bar{t})^2}{2\sigma_b^2}} \qquad \text{(A.68)}$$

with $A/B = 1.3/0.7$, and their magnitude determined by normalization. The parameters f, σ_a^2 and σ_b^2 are determined numerically to provide good visual separation between peaks and to match the desired $\langle t \rangle$ and η_t. This distribution is chosen as an example of an "unnatural" distribution, far from the usually assumed exponential. However, a bimodal distribution could in principle be observed in cells, for example when a creation event arises from two parallel pathways.

Simulations are done following 10^4 individual cells over 100 minutes for each set of parameters. We used a direct generalization of the Gillespie algorithm (*3*), where the random times t that follow any distribution $f(t)$ are obtained from a uniformly generated random number x using

$$t = F^{-1}(1-x), \qquad \text{(A.68)}$$

where $F(y) = \int_0^y f(t)dt$ is the cumulative distribution and the integration and inversion were performed numerically.

In addition to the parameters and distributions used in the figure, the accuracy of Eq. (1) was confirmed over multiple distributions and parameter values. In fact, the small deviations observed in Figure 2 disappear if the values for $\langle t \rangle, \eta_t, \langle m \rangle$ and $\langle p \rangle$ resulting from a particular run are used instead of the preset parameters.

For Figure 3C, we used the same parameters as for the case of the exponential distribution with $\langle t \rangle = 1/(0.4\,Log(2))$, $\eta_t = 1$, $\langle m \rangle = 2$ and $\langle p \rangle = 600$. The analytical results are plotted, but they correspond exactly to simulations performed by directly including the intermediate steps in a standard Gillespie algorithm.

For Figure 4, the curves for transcriptionally varied averages were generated with most of the same parameters used in Figure 2: the red solid line corresponds to the case of the exponential distribution with $\eta_t = 1$, $\eta_b = 0$, the blue solid line corresponds to $\eta_t = 0.6$, $\eta_b = 0$ and the blue circles to $\eta_t = 0.6$, $\eta_b = 0.8$. In all cases, $\langle b \rangle = 1$. For the translationally varied averages, the average mRNA creation rate was fixed at $\langle t \rangle = 1/(1.6\,Log(2))$ and the average burst at $\langle b \rangle = 1$, which results in $\langle m \rangle = 8$. The protein creation rate k_p was varied to have the average in the range $\langle p \rangle \in (500, 3500)$. The mRNA creation timing and burst noise were varied as before, and now the red dotted line corresponds to $\eta_t = 1$, $\eta_b = 0$, the blue dotted line corresponds to $\eta_t = 0.6$, $\eta_b = 0$ and the blue triangles to $\eta_t = 0.6$, $\eta_b = 0.8$.

In Figure 5A, the reference gene (green curves) uses the same parameters as the case of the exponential distribution from Fig. 2, with $k_m = 0.4\,Log(2)$, $\eta_t = 1$, $\langle m \rangle = 2$ and $\langle p \rangle = 600$, until $t = 10h$, where the transcription rate is doubled. The blue curves correspond to a similar case but with the transcription rate modulated by $k_m \to k_m \frac{2}{1+(p/600)^2}$ where *p* represents the instantaneous protein number. After $t = 10h$, k_m is doubled. The red curves correspond to a gene with gestation in the transcription times, specifically a Gamma distribution of creation times with $\alpha = 8$ and $\theta = \frac{\langle t \rangle}{\alpha} = \frac{1}{\alpha k_m}$. Other parameters are as before. After $t = 10h$, a Gamma distribution with $\theta = \frac{1}{2\alpha k_m}$ is used. In Figure 5.B, the reference gene is as explained above until $t = 10h$, where the transcription rate is increased by 50%. The blue curves correspond to a case where the transcription rate is modulated by $k_m \to k_m \left(0.01 + 0.99 \frac{2p}{600 + p} \right)$. The red curves correspond to a gene with the same feedback but where the mRNA undergoes a five step decay, with the intermediate rates adjusted to maintain the overall decay rate. Note that the feedbacks are idealized, so they react instantaneously and depend directly on the protein concentration. In a real system further complications arise due to time delays and possible additional fluctuations coming from feedback intermediaries, increasing the relative advantage of small step-size mechanisms.

References

1. D. T. Gillespie, *Journal of Computational Physics* **22**, 403 (1976).
2. J. L. Doob, *Transactions of the American Mathematical Society* **58**, 455 (1945).
3. D. T. Gillespie, *Journal of Physical Chemistry* **81**, 2340 (1977).
4. M. Lax, *Rev Modern Physics*, 25 (1960).
5. J. Paulsson, *Nature Genetics* **37**, 925 (2005).
6. J. Keizer, *Statistical Thermodynamics of Nonequilibrium Processes* (Springer, Berlin, 1987), pp. 506.
7. J. Paulsson, *Nature (London, United Kingdom)* **427**, 415 (2004).
8. J. Paulsson, M. Ehrenberg, *Physical Review Letters* **84**, 5447 (2000).
9. H. Kacser, J. A. Burns, *Symposia of the Society for Experimental Biology* **27**, 65 (1973).
10. J. Paulsson, M. Ehrenberg, *Quarterly Reviews of Biophysics* **34**, 1 (2001).
11. O. G. Berg, *Journal of Theoretical Biology* **71**, 587 (1978).
12. D. R. Rigney, W. C. Schieve, *Journal of Theoretical Biology* **69**, 761 (1977).
13. S. U. P. A. Papoulis, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, New York, 2002), pp.