Supplementary Material

Geometry–induced bursting dynamics in gene expression

B. Meyer,[1] O. Bénichou,[1] Y. Kafri,[2] and R. Voituriez[1]

[1]*UMR 7600, Université Pierre et Marie Curie/CNRS,*

4 Place Jussieu, 75255 Paris Cedex 05 France.

[2]*Department of Physics, Technion, Haifa 32000, Israel.*

I. DERIVATION OF THE CORRELATION FUNCTION OF THE mRNA NUMBER

A. Definitions

The discrete time variable is denoted by t, where the unit time step is τ_0. The discrete Fourier transform of a function g is defined by: $\hat{g}(\omega) = \sum_{t=-\infty}^{+\infty} e^{-i\omega t} g(t)$ with $\omega \in [-\pi, \pi]$. In the limit of $\tau_0 \to 0$ we will make use of the continuous time notation $\hat{g}(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t} g(t) dt$. The mean of a given stationary process $\xi = (\xi(t))_t$ is denoted by μ_ξ. The autocorrelation function (or auto-covariance) of ξ is defined in the stationary state by: $R_\xi(t) = \langle \xi(0)\xi(t) \rangle - \mu_\xi^2$.

For the sake of simplicity, we first consider a single TF in a nucleus (or cell) of volume V. The case of n TFs will be treated later. Let us then denote by $f(T) = f(r_0, T | r_0, 0)$ the distribution of the first-return time T to the target site r_0, with the convention $f(0) = 0$. In turn, $p(t) = p(r_0, t | r_0, 0)$ denotes the propagator, i.e. the probability that the TF, starting from the target site r_0 at $t = 0$, is located again at the target at time t. Both quantities are related by the renewal equation:

$$p(t) = \sum_{t'=0}^{t} f(t') \, p(t - t') + \delta_{t,0}, \tag{1}$$

which reads after Fourier transform:

$$\hat{p}(\omega) = \frac{1}{1 - \hat{f}(\omega)}. \tag{2}$$

The mRNA input process is governed by two stochastic processes: (i) first, the successive visits of the TF to the gene locus, which can be characterized by the waiting time distribution $f(T)$ between each visit to the target; (ii) second, the production of mRNA, which follows a Poisson law: at each visit of the TF at the gene locus (and only in that case), the gene produces m new mRNA molecules with probability $e^{-\langle m \rangle} \langle m \rangle^m / m$, where $\langle m \rangle = k\tau$, with k the synthesis rate and τ the binding time of the TF to its target. For the sake of simplicity, τ is taken here as the elementary time step τ_0. We denote by $E = (E(t))_t$ the resulting production process of mRNA molecules, namely the number of mRNA molecules produced at time step t. In the framework of queuing theory, that we use below, this is called the input or emission process.

2

Last, we denote by θ the random life time of mRNA molecules and $h(\theta)$ its distribution, which characterizes the degradation process of the mRNA molecules. In practice we will consider a geometric decay $h(\theta) = \lambda_d (1 - \lambda_d)^\theta$, and assume $\lambda_d \ll 1$, so that the usual exponential form $h(\theta) \simeq \lambda_d e^{-\lambda_d \theta}$ holds. In Fourier space this yields

$$\hat{h}(\omega) = \frac{\lambda_d}{1 - (1 - \lambda_d)e^{i\omega}}. \tag{3}$$

B. Input process

From Kac's formula, the stationary probability for the TF to be located at the gene locus is given by $p^\infty = 1/\langle T \rangle$, where $\langle T \rangle$ denotes the mean first return time. Hence, the mean of the input process is readily obtained as

$$\mu_E = \langle m \rangle / \langle T \rangle. \tag{4}$$

When $t > 0$, the correlation function of the emission process reads:

$$R_E(t) = \frac{\langle m \rangle^2}{\langle T \rangle} \left(p(t) - p^\infty \right). \tag{5}$$

When $t = 0$, $R_E(0)$ is the variance of the emission process, which reads:

$$\sigma_E^2 = \frac{\langle m \rangle}{\langle T \rangle} \left(1 + \langle m \rangle - \frac{\langle m \rangle}{\langle T \rangle} \right). \tag{6}$$

Finally, one gets for $t \geq 0$:

$$R_E(t) = \frac{\langle m \rangle^2}{\langle T \rangle} \left(p(t) - 1/\langle T \rangle \right) + \frac{\langle m \rangle}{\langle T \rangle} \delta_{t,0}, \tag{7}$$

and the range $t < 0$ is obtained by parity. This yields in Fourier space

$$\hat{R}_E(\omega) = \frac{\langle m \rangle^2}{\langle T \rangle} \left(2\text{Re}\left[\hat{p}(\omega)\right] - 1 - \frac{2\pi}{\langle T \rangle}\delta(\omega) \right) + \langle m \rangle / \langle T \rangle, \tag{8}$$

where Re denotes the real part.

C. Correlation function of the mRNA copy number

The results of this Section are derived from the expressions given by I. Eliazar in [1] in the framework of queuing theory. Queuing theory relates the number of customers in a

3

queue, called queue process (analogous to the number n_M of mRNA molecules in our case) as a function of the input process, i–e the arrival statistics of customers (in our case the synthesis of mRNA molecules E) and the output process or service time statistics (in our case the life time θ of an mRNA molecule). First, it is needed to derive the mean number of mRNA molecules, which is readily obtained as:

$$\mu_M = \frac{\mu_E}{\lambda_d}. \tag{9}$$

On general grounds, the correlation function of the queue process (mRNA number in our case) is then given in terms of the input and output processes by [1] :

$$\hat{R}_M(\omega) = \frac{1 - \mathrm{Re}[\hat{h}(\omega)]}{1 - \cos(\omega)} \mu_E + \left| \frac{1 - \hat{h}(\omega)}{1 - \exp(i\omega)} \right|^2 \left(\hat{R}_E(\omega) - \mu_E \right). \tag{10}$$

This equation, together with Eqs.(4,8) explicitly determines the Fourier transformed correlation function of the mRNA copy number, for any distribution $h(\theta)$ of the mRNA life time. In what follows we will consider the continuous time limit where $\tau_0 \to 0$, and assume an exponential decay of the mRNA life time as given in Eq.(3). In this regime the correlation function takes the following simple form

$$\hat{R}_M(\omega) = \frac{\langle m \rangle}{\langle T \rangle} \left(1 + \frac{\langle m \rangle}{2} \right) \frac{2}{\lambda_d^2 + \omega^2} + \frac{\langle m \rangle^2}{\langle T \rangle} \left(\mathrm{Re} \left(\frac{\hat{f}(\omega)}{1 - \hat{f}(\omega)} \right) - \frac{\pi}{\langle T \rangle} \delta(\omega) \right) \frac{2}{\lambda_d^2 + \omega^2} \tag{11}$$

which is exactly Eq.(1) of the main text.

II. 2-EXPONENTIAL WAITING TIME DISTRIBUTION

In this section we consider the example where f is a sum of 2 exponentials of the form :

$$f(T) = (1 - q)\gamma_1 \, e^{-\gamma_1 T} + q\gamma_2 \, e^{-\gamma_2 T}, \tag{12}$$

with $\gamma_1 > \gamma_2$ so that the mean return time reads $\langle T \rangle = (1 - q)/\gamma_1 + q/\gamma_2$. This functional form will be useful later on. The Fourier transformed distribution reads

$$\hat{f}(\omega) = \frac{1 - q}{1 + i\omega/\gamma_1} + \frac{q}{1 + i\omega/\gamma_2}. \tag{13}$$

It is useful to introduce an extra time scale $1/\tilde{\gamma}$ defined by:

$$\tilde{\gamma} = q\gamma_1 + (1 - q)\gamma_2. \tag{14}$$

4

Then one has:

$$\frac{\hat{f}(\omega)}{1-\hat{f}(\omega)} = \frac{(1+\frac{i\omega}{\gamma_1})(1+\frac{i\omega}{\gamma_2})}{(1+\frac{i\omega}{\gamma_1})(1+\frac{i\omega}{\gamma_2})-(1-q)(1+\frac{i\omega}{\gamma_2})-q(1+\frac{i\omega}{\gamma_1})} - 1$$

$$= \frac{1-\frac{\omega^2}{\gamma_1\gamma_2}+i\omega\left(\frac{1}{\gamma_1}+\frac{1}{\gamma_2}\right)}{i\omega\langle T\rangle - \frac{\omega^2}{\gamma_1\gamma_2}} - 1$$

$$= -\frac{i}{\omega\langle T\rangle}\frac{1}{1+\omega^2/\tilde{\gamma}^2}\left[1+\frac{\omega^2}{\tilde{\gamma}}\left(\frac{1}{\gamma_1}+\frac{1}{\gamma_2}-\langle T\rangle\right)\right.$$

$$\left.+ i\omega\left(\frac{1}{\gamma_1}+\frac{1}{\gamma_2}-\frac{1}{\tilde{\gamma}}+\frac{\omega^2\langle T\rangle}{\tilde{\gamma}^2}\right)\right] - 1. \tag{15}$$

As before, this function is singular for $\omega \to 0$ and needs regularization. The only term that is not well defined for $\omega \to 0$ is the very first one within the square bracket. This term can be regularized as follows:

$$\mathrm{Re}\left[\lim_{\epsilon\to0^+}\frac{-i}{\langle T\rangle}\frac{1}{\omega-i\epsilon}\frac{1}{1+\frac{(\omega-i\epsilon)^2}{\tilde{\gamma}^2}}\right] = \pi\frac{\delta(\omega)}{\langle T\rangle}. \tag{16}$$

Taking the real part of Eq. (15) then leads to the following expression for the Fourier transformed correlation function:

$$\hat{R}_M(\omega) = \frac{2\langle m\rangle/\langle T\rangle}{\lambda_d^2+\omega^2}\left[1+\frac{\langle m\rangle}{2}-\frac{\langle m\rangle\tilde{\gamma}^2}{\omega^2+\tilde{\gamma}^2}\left(1-\frac{1}{\gamma_1\langle T\rangle}\right)\left(1-\frac{1}{\gamma_2\langle T\rangle}\right)\right]. \tag{17}$$

After Fourier inversion, one finally obtains :

$$R_M(t) = \frac{\langle m\rangle}{\lambda_d\langle T\rangle}\left(1+\frac{\langle m\rangle}{2}\right)e^{-\lambda_d t} + \frac{\langle m\rangle^2}{\lambda_d\langle T\rangle}\frac{q(1-q)(\gamma_1-\gamma_2)^2}{\tilde{\gamma}(\lambda_d^2-\tilde{\gamma}^2)}\left(\lambda_d e^{-\tilde{\gamma}t}-\tilde{\gamma}e^{-\lambda_d t}\right). \tag{18}$$

III. EFFECT OF THE NUMBER n OF TFs

Now assume that n copies of TFs can activate the gene, and that n is not too large so that the probability that more than one TF occupies the gene locus is negligible. We denote by $f_n(T)$ the waiting time distribution between successive activation events, and $p_n(t)$ the corresponding propagator, i-e the probability that the gene is activated at time t, knowing that it was activated at $t=0$.

The derivation of the autocorrelation function closely follows the case of a single TF presented above with the substitutions $p \to p_n$, $p_n^\infty = np_1^\infty$, and $\langle T\rangle_n = \langle T\rangle/n$. One finds:

$$\hat{R}_M^{(n)}(\omega) = \frac{n\langle m\rangle}{\langle T\rangle}\left(1+\frac{\langle m\rangle}{2}\right)\frac{2}{\lambda_d^2+\omega^2} + \frac{n\langle m\rangle^2}{\langle T\rangle}\left(\mathrm{Re}\left(\frac{\hat{f}_n(\omega)}{1-\hat{f}_n(\omega)}\right)-\frac{n\pi}{\langle T\rangle}\delta(\omega)\right)\frac{2}{\lambda_d^2+\omega^2}. \tag{19}$$

To determine f_n, we introduce the survival probability defined by $S(t) = \int_t^\infty f(t)\,dt = 1 - \int_0^t f(t)\,dt$. Then one has :

$$S_n = \overline{S}^{n-1}S_1, \tag{20}$$

where $\overline{S}(t)$ is the survival probability averaged over all possible starting positions, derived in [2]. The waiting time distribution for n TFs then reads $f_n(t) = -\frac{dS_n(t)}{dt}$.

We consider below the case where $f(T)$ is characterized by well separated regimes for short and long time scales. This is the case of both non compact and compact transport detailed in the next section. In that case $\hat{R}_M^{(n)}(t)$ can be straightforwardly deduced from $\hat{R}_M^{(1)}(\omega)$. This can be shown explicitly on the simple example where we take:

$$f(T) = (1-q)\gamma_1\,e^{-\gamma_1 T} + q\gamma_2\,e^{-\gamma_2 T}, \tag{21}$$

with $\gamma_1 \gg \gamma_2$. As shown in [2], the first-passage distribution to a target averaged over the initial position $\overline{f}(T)$ can be deduced from $f(T)$, and is characterized only by long time scales. We therefore assume

$$\overline{f}(t) = \gamma_2 e^{-\gamma_2 t}. \tag{22}$$

Eq. (20) then yields :

$$f_n(t) = (1-q)\left(\gamma_1 + (n-1)\gamma_2\right)e^{-(\gamma_1+(n-1)\gamma_2)t} + nq\gamma_2 e^{-n\gamma_2 t}. \tag{23}$$

One then makes use of Eq. (18) and finds :

$$R_M^{(n)}(t) = \frac{\langle m \rangle}{\lambda_d \langle T \rangle_n}\left(1 + \frac{\langle m \rangle}{2}\right)e^{-\lambda_d t} + \frac{\langle m \rangle^2}{\lambda_d \langle T \rangle_n}\frac{q(1-q)(\gamma_2-\gamma_1)^2}{\widetilde{\gamma}_n\left(\lambda_d^2 - \widetilde{\gamma}_n^2\right)}\left(\lambda_d e^{-\widetilde{\gamma}_n t} - \widetilde{\gamma}_n e^{-\lambda_d t}\right), \tag{24}$$

with

$$\overline{\langle T \rangle}_n = \frac{1-q}{\gamma_1 + (n-1)\gamma_2} + \frac{q}{n\gamma_2}, \tag{25}$$

$$\widetilde{\gamma}_n = q\gamma_1 + (n-q)\gamma_2. \tag{26}$$

We then consider the biologically relevant regime (see figures 2 and 3 of the main text) :

$$(1-q)/\gamma_1 \ll q/\gamma_2, \tag{27}$$

that gives $\overline{\langle T \rangle} \simeq q/\gamma_2$. We also assume that n is not too large, so that $\gamma_1 \gg n\gamma_2$ and $(1-q)/\gamma_1 \ll q/n\gamma_2$. Then one has :

$$\overline{\langle T \rangle}_n \simeq \overline{\langle T \rangle}/n, \tag{28}$$

$$\widetilde{\gamma}_n \simeq \widetilde{\gamma}, \tag{29}$$

so that

$$R_M^{(n)}(t) \simeq n R_M^{(1)}(t). \tag{30}$$

These relations are checked numerically in the main text (see figures 2 and 3).

IV. WAITING TIME DISTRIBUTIONS $f_n(T)$ OF GENERIC SCALE INVARIANT TRANSPORT PROCESSES

The first-passage time distribution of a random walker to a target site has been obtained in [3] for generic scale invariant transport processes. It was shown that asymptotic distributions in the large volume limit have universal features controlled essentially by the walk dimension d_w of the process, defined for example through the scaling of the mean square displacement $\langle \Delta r^2 \rangle \propto t^{2/d_w}$, and the fractal dimension d_f of the environment, defined by the scaling of the accessible volume $V \propto r_f^d$ within a ball of radius r.

A. Non compact exploration $(d_w < d_f)$

In the case of non compact exploration, it was found that the distribution of return time to a target for a single random walker reads in the large volume limit:

$$f(T) = (1 - \Pi)\,\delta(T) + \frac{\Pi}{\overline{\langle T \rangle}} \exp\left(-T/\overline{\langle T \rangle}\right), \tag{31}$$

with

$$\Pi = \frac{\langle T \rangle}{\overline{\langle T \rangle}} \tag{32}$$

and $\overline{\langle T \rangle} \propto V$ denotes the mean first-passage time to the target averaged over all possible starting positions. Note that $\langle T \rangle \propto V$ so that Π is of order 1. In the non compact case the survival probability averaged over the starting positions reads [2] :

$$\bar{S}(T) = \exp\left(-T/\overline{\langle T \rangle}\right). \tag{33}$$

Making use of Eq.(20) then yields for n TFs:

$$f_n(T) = (1 - \Pi)\delta(T) + \Pi \frac{n}{\overline{\langle T \rangle}} e^{-nT/\overline{\langle T \rangle}}, \tag{34}$$

7

where $\delta(T)$ accounts for trajectories returning to the gene locus within time scales much shorter than $\overline{\langle T \rangle}$, as stated in the main text. Rewriting $\delta(T) = \lim_{\gamma \to \infty} \gamma e^{-\gamma T}$, the formalism of the previous paragraph for 2-exponential waiting time distributions applies and one gets:

$$R_M(t) \simeq \frac{n \langle m \rangle}{\langle T \rangle \lambda_d} \left(1 + \frac{\langle m \rangle}{2} + \frac{\langle m \rangle (1 - \Pi)}{\Pi} \right) e^{-\lambda_d t} \tag{35}$$

as given in the main text.

B.   Compact exploration $(d_w \geq d_f)$

In the case of compact exploration, the waiting time distribution for one TF reads in the large volume limit: :

$$f(T) = (1 - \Pi) \delta(T) + \Pi \sum_{k=0}^{\infty} b_k a_k e^{-a_k T}, \tag{36}$$

with

$$\Pi = \frac{2 \nu^2}{1 + \nu} \frac{\langle T \rangle}{\overline{\langle T \rangle}}, \tag{37}$$

$$a_k = \frac{z_k^2 \nu}{2(1 - \nu^2) \overline{\langle T \rangle}}, \tag{38}$$

$$b_k = \frac{\Gamma(\nu)}{\Gamma(2 - \nu)} \left(\frac{z_k}{2} \right)^{1 - 2\nu} \frac{J_\nu(z_k)}{J_{1-\nu}(z_k)}, \tag{39}$$

where $z_0 < z_1 < ...$ are the real zeros of the Bessel function $J_{-\nu}$ and $\nu = d_f / d_w$. Note that here $\langle T \rangle \sim V$ and $\overline{\langle T \rangle} \sim V^{d_w/d_f}$, so that $\Pi \sim V^{1-d_w/d_f} \simeq 0$. This is a crucial difference with the non compact case. This large weight of fast trajectories makes the description of short time scales in $f(T)$ important. In the compact case the survival probability averaged over the starting position reads [2] :

$$\bar{S}(T) = \sum_{k=0}^{\infty} c_k e^{-a_k T} \tag{40}$$

with

$$c_k = \frac{2 \nu^2}{1 + \nu} b_k. \tag{41}$$

Making use of Eq.(20) then yields for n TFs:

$$f_n(T) = (1 - \Pi) \delta(T) + \Pi \sum_{k=0}^{\infty} q_k \gamma_k e^{-\gamma_k T} \tag{42}$$

8

which is the expression of the main text (with $\alpha_k = q_k \gamma_k$ and $\beta_k = \gamma_k \overline{\langle T \rangle}/n$). In practice exact expressions for q_k, γ_k are cumbersome, and for practical purposes one uses truncated sums in Eqs. (36, 40) .

Below we go beyond the result of Eq.(42), which only gives the long time regime, and derive an expression that describes in more details also the short time regime. Using the identity $\hat{f}_n/(1 - \hat{f}_n) = \hat{p}_n - 1$ that holds in Fourier (or Laplace) space, we will focus below on the propagator p_n.

a. *Long time regime of the propagator* The long time regime of the propagator is controlled only by the long time regime of the waiting time distribution $f(T)$ (we omit below the index n referring to the number of TFs for the sake of readability). Instead of the exact asymptotics of Eq.(42), we introduce for practical purposes a finite truncation:

$$f_{\text{long}}(T) = \Pi \sum_{k=0}^{n_e} q_k \, \gamma_k e^{-\gamma_k T}. \tag{43}$$

In order to preserve the normalisation, we define :

$$f_{\text{long}}^{(n_e)}(T) = \left(1 - \Pi \sum_{k=0}^{n_e} q_k \right) \delta(T) + \Pi \sum_{k=0}^{n_e} q_k \, \gamma_k e^{-\gamma_k T}. \tag{44}$$

We then write $\widetilde{q}_k = q_k / \sum_{k=0}^{n_e-1} q_k$ and deduce in Laplace space :

$$\hat{p}_{\text{long}}^{(n_e)}(s) = \frac{1}{1 - \hat{f}_{\text{long}}(s)} = \frac{1}{\Pi \left(\sum_{k=0}^{N-1} q_k \right)} \frac{1}{1 - \sum_{k=0}^{n_e-1} \frac{\widetilde{q}_k}{1+s/\gamma_k}}. \tag{45}$$

The Laplace inverse can then be obtained by analysing the pole structure of Eq. (45). Eventually the propagator reads in the long time regime:

$$\Pi \left(\sum_{k=0}^{n_e-1} q_k \right) p_{\text{long}}(t) = \frac{\sum_{k=0}^{n_e-1} q_k}{\langle T \rangle} + \sum_{p=1}^{n_e-1} A_p \, \Lambda_p \, e^{-\Lambda_p t}, \tag{46}$$

where Λ_p and A_p depend on γ_k and $\widetilde{q}_k$. More explicitly, one obtains for $n_e = 2$ (the derivation can be carried out similarly for larger n_e) :

$$p_{\text{long}}^{(2)}(t) = \frac{1}{\Pi} \frac{\gamma_0 \gamma_1}{\widetilde{\gamma}} + \frac{1}{\Pi} \frac{q_0 q_1 \, (\gamma_0 - \gamma_1)^2}{\widetilde{\gamma}(q_0 + q_1)} \times \exp \left(-\frac{\widetilde{\gamma}}{q_0 + q_1} t \right). \tag{47}$$

b. *Short time regime of the propagator.* In the regime of short times, the propagator can be approximated by the infinite space propagator of a single diffusing particle. This quantity has been analyzed in [4] and reads:

$$p_{\text{short}}(t) = \frac{p^{\infty}}{\Gamma(\nu + 1)} \left(\frac{(1 - \nu^2) \overline{\langle T \rangle}}{2\,\nu\,t} \right)^{\nu} \exp \left(-\frac{(1 - \nu^2) \overline{\langle T \rangle}}{2\,\nu\,t} \Pi^{\frac{1}{1-\nu}} \right). \tag{48}$$

c.   Interpolation at all times of the propagator.   We use the following ansatz to obtain an interpolation of the propagator that holds at all times :

$$p(t) = e^{-t/\tau^*} \, p_{\text{short}}(t) + \left(1 - e^{-t/\tau^*}\right) p_{\text{long}}(t). \tag{49}$$

The crossover time τ^* is defined as the smallest root of the equation

$$p_{\text{short}}(t) = p_{\text{long}}(t). \tag{50}$$

In practice this solution exists and provides a very good approximation of the propagator which is valid on all time scales.

[1] I. Eliazar, *EPL (Europhysics Letters)* **82**, 30005 (2008).

[2] O. Bénichou, C. Chevalier, B. Meyer, R. Voituriez, *Physical Review Letters* **106** (2011).

[3] O. Bénichou, C. Chevalier, J. Klafter, B. Meyer, R. Voituriez, *Nat Chem* **2**, 472 (2010).

[4] B. O'Shaughnessy, I. Procaccia, *Physical Review Letters* **54** (1985).