Biophysical Journal, Volume *117*

Supplemental Information

Transcriptional Bursts in a Nonequilibrium Model for Gene Regulation by Supercoiling

Marco Ancona, Alessandro Bentivoglio, Chris A. Brackley, Giuseppe Gonnella, and Davide Marenduzzo

I. MEAN FIELD THEORY

Here, we develop a mean field theory with some improvements with respect to our previous work [1]. In particular we solve the mean field ordinary differential equation (ODE) with periodic boundary conditions (instead of open boundary conditions, as previously done) and in the presence of topoisomerases.

We consider the case $N = n = 1$, where N is the number of RNAP and n is the number of genes. Besides, we consider a *static* polymerase (i.e. $v = 0$) at the lattice position $x = 0$. If L is the length of the lattice, we assume boundary conditions $\sigma(0) = 0$ and $\sigma(L/2) = \sigma(-L/2)$. In steady state ($\partial\sigma/\partial t = 0$), Eq. (1) reads:

$$\frac{\partial^2\sigma(x)}{\partial x^2} - \frac{J_0}{D}\frac{k_{\rm in}\tau}{k_{\rm in}\tau+1}\frac{\partial\delta(x)}{\partial x} - \frac{k_{\rm topo}}{D}\sigma(x) = 0 \tag{S1}$$

where we have made the mean field approximation

$$\frac{J_{tr}(x,t)}{D} \to \frac{J_0}{D}\frac{k_{\rm in}\tau}{k_{\rm in}\tau+1}\delta(x) \equiv M\delta(x) \tag{S2}$$

with $k_{\rm in}\tau/(k_{\rm in}\tau+1)$ the fraction of time the system spends in the transcribing state.

As the flux term acts only at $x = 0$, solving the model in the mean field approximation is equivalent to solving the following ODE:

$$\begin{cases} \dfrac{\partial^2\sigma(x)}{\partial x^2} - \dfrac{k_{\rm topo}}{D}\sigma(x) = 0 & x \neq 0 \\[2ex] \left.\dfrac{\partial\sigma(x)}{\partial x}\right|_{x=0} = M\delta(0) \\[2ex] \sigma\left(L/2\right) = \sigma\left(-L/2\right). \end{cases} \tag{S3}$$

Since both $\sigma(x)$ and $\sigma(-x)$ are solution of the ODE for $x \neq 0$ the unique solution of Eq. (S3) is a linear combination of $\sigma(x)$ and $\sigma(-x)$. It can be shown that only the antisymmetric combination fulfils the periodic boundary conditions, with $\sigma(L/2) = \sigma(-L/2) = 0$.

The solution of Eq. (S3) with the appropriate parity and boundary conditions is given

by:

$$\sigma(x) = \frac{M}{2} \frac{\sinh\left[\sqrt{\frac{k_{\text{topo}}}{D}}\left(\frac{L}{2} - |x|\right)\right]}{\sinh\left[\sqrt{\frac{k_{\text{topo}}}{D}}\frac{L}{2}\right]} \text{sgn}(x), \tag{S4}$$

where sgn(x) is the *sign function*. From Eq. (S4) it can be easily shown that in the limit $k_{\text{topo}} \to 0$ we obtain:

$$\sigma(x) = \frac{M}{2}\left(1 - \frac{2|x|}{L}\right)\text{sgn}(x). \tag{S5}$$

The term proportional to $1/L$ is the correction due to the periodic boundary conditions, that disappears for $L \to \infty$, recovering the solution in Ref. [1]. In the limit $L \to \infty$, with finite k_{topo}, we have

$$\sigma(x) = \frac{M}{2}\exp\left(-\sqrt{\frac{k_{\text{topo}}}{D}}|x|\right)\text{sgn}(x). \tag{S6}$$

The validity of this mean field theory can be determined by comparing it to the time-average supercoiling profile in our single gene simulations.

Interestingly, from our simulations we found that the point along the gene at which the time-averaged supercoiling profile crosses zero is $\sim 2\lambda/3$, independently of the parameter used. The correct mean field profile of supercoiling for a moving polymerase is then computed by substituting $|x| \to |x - 2\lambda/3|$.

II. THE MODIFIED *IPP* PROCESS

The mechanism which leads to bursty dynamics for transcription in living cells is still not well understood, though several hypothesis have been made. In our model, we have seen that both the action of topoisomerases (1-gene model) and the interaction among genes (10-gene model) can yield bursts, in absence of external factors. The nontrivial nonlinear behaviour predicted by our model can be captured by a simpler kinetic scheme: the *Interrupted Poisson Process* (*IPP*). By solving the *IPP* equations, one can explicitly obtain the double-exponential distribution of waiting times between events, that, when appropriate conditions on the kinetics rates are met, leads to bursting. In this section we modify the *IPP* equations. Nevertheless, the resulting distribution is still a double-exponential (see below), with the exception of the two timescales, which are different from those found in [2].

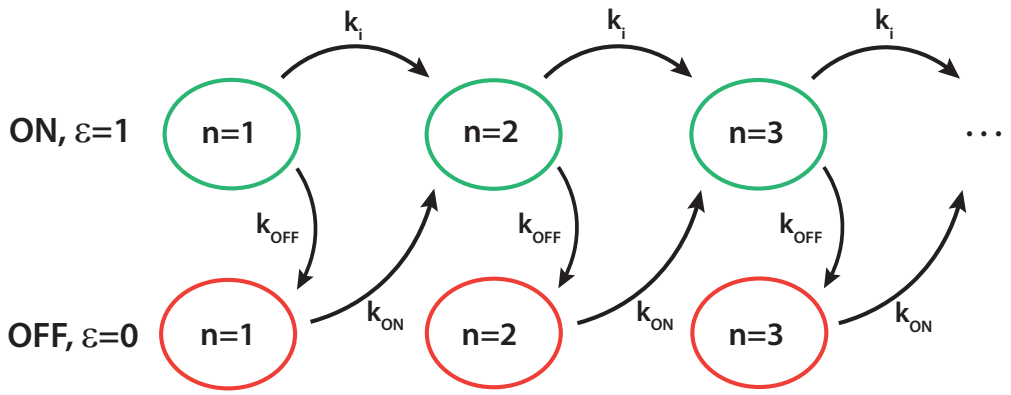


FIG. S1: **Scheme of the discrete states in a *modified IPP*.** Each *box* represent a state, with $\epsilon \in \{0, 1\}$. The index n labels the number of initiation events.

We define (as in Fig. 1a of the main text) an *active* (ON, or $\epsilon = 1$) and an *inactive* (OFF, or $\epsilon = 0$) state of the gene promoter according to the local supercoiling density (i.e., if the supercoiling density is below $(1 - (k_0\tau_2)^{-1})/\alpha$, then the promoter is ON). We then associate the rates k_{OFF} and k_{ON} with the $ON \to OFF$ and $OFF \to ON$ transitions respectively. The gene oscillates between the two states, tracking the typical trajectories of a *Random Telegraph Process* (see Fig. 1a, BOTTOM). Whilst in the ON state, the gene is able to transcribe with rate k_i. Given a time series of events, the *waiting time* t_n is the elapsed time between two consecutive transcriptions, say the $(n-1)$-th and the n-th. In the *IPP*, the waiting time t_n is drawn by the same probability distribution function (*pdf*) for each n (this may not be true in general in our stochastic model for supercoiling-dependent transcription). Moreover, in the standard *IPP* prescription the $OFF \to ON$ transition occurs between states labelled by the same n, see Ref. [2]. Differently, in our modified *IPP* description, the transition $OFF \to ON$ is triggered by transcription (see Fig. S1), meaning that the transition and the first event in the burst occur at the same time. By labelling possible states with the instantaneous value of ϵ (gene ON/OFF) and by the index n, which keeps count of the number of transcriptions, we denote the probability of being in the state $\{n, \epsilon\}$ at time t by $p_{n,\epsilon}(t)$. Then, the set of master equations for our modified *IPP* is:

$$\begin{cases} \dfrac{dp_{1,1}(t)}{dt} = -(k_i + k_{OFF})\ p_{1,1}(t) \\ \dfrac{dp_{n,1}(t)}{dt} = k_i p_{n-1,1}(t) + k_{ON} p_{n-1,0}(t) - (k_i + k_{OFF}) p_{n,1}(t), \qquad n = 2, 3, \dots \\ \dfrac{dp_{n,0}(t)}{dt} = k_{OFF}\ p_{n,1}(t) - k_{ON}\ p_{n,0}(t) \qquad n = 1, 2, \dots \end{cases} \tag{S7}$$

with the initial condition $p_{1,1}(t = 0) = 1$. Clearly, the *pdf* associated with the first transcription event after initialisation ($n = 2$) corresponds to the distribution of waiting times, that is

$$f(t) = k_i\ p_{1,1}(t) + k_{ON}\ p_{1,0}(t). \tag{S8}$$

In order to find $p_{1,1}(t)$ and $p_{1,0}(t)$, and therefore $f(t)$, we need to solve just the following two first order coupled ODEs:

$$\begin{cases} \dfrac{dp_{1,1}(t)}{dt} = -(k_i + k_{OFF})\ p_{1,1}(t), \\ \dfrac{dp_{1,0}(t)}{dt} = k_{OFF}\ p_{1,1}(t) - k_{ON}\ p_{1,0}(t). \end{cases} \tag{S9}$$

By solving Eq. (S9) and using Eq. (S8), we have

$$f(t) = w_1 r_1 e^{-r_1 t} + w_2 r_2 e^{-r_2 t}, \tag{S10}$$

with rates $r_{1,2}$

$$r_1 = k_i + k_{OFF}, \qquad r_2 = k_{ON}, \tag{S11}$$

and weights $w_{1,2}$

$$w_1 = \frac{k_i - r_2}{r_1 - r_2}, \qquad w_1 \in [0, 1]\,, \tag{S12}$$

$$w_2 = 1 - w_1. \tag{S13}$$

In our stochastic model used in the main text, we can identify $k_{ON} \sim k_0$. Conversely, it is not easy to find a value for the rate k_{OFF} without fitting the data, since the transition $ON \rightarrow OFF$ is mainly due to fluctuations, that, in the bursty phase, relax the system towards the initial value of the supercoiling σ_0.

III. 1-GENE ARRAY: BURST PARAMETERS AND ADDITIONAL FIGURES

In our work we use the *sequence-size function*, $\Phi(\tau)$, to analyse burst significance. Although this method has been presented in previous works [2, 3], it has not previously been

applied to simulation of the dynamics of transcription which does not use predetermined kinetic rates for the process.

We use the parameter ξ as defined in Eq. (7) in the main tex, which is different from the parameter proposed in [2], $(\tau_2 - \tau_1)/\tau_2$. Indeed, the former parameter does show variation within the bursty phase, whereas the latter does not. Our parameter ξ yields a reasonable estimate of burst significance, as (i) it is still proportional to $\tau_2 - \tau_1$ and (ii) it fulfils the intuitive expectation that the burst significance should decrease if the system spends more time in intermediate states, at fixed $\tau_2 - \tau_1$.

The analysis of the *ssf* allows us to readily compute other relevant burst parameters in a relatively simple way. The time separation between the two timescales is just $\tau_x = (\tau_1 + \tau_2)/2$ and, from the definition of the *ssf*, we can estimate the *mean burst size* (the average number of transcriptions in a single burst) as $\beta = \Phi(\tau_x)$. This is a useful parameter, since it provides a simple basis to compare with experimental data. As we can see from Fig. S2 higher values of β ($\beta > 4 - 5$) correspond to less significant bursting (see main text, Fig. 2). Conversely, in the region of higher ξ ($\bar{J}/D \sim 1.5 - 2$), the burst size $\beta \simeq 2.2$ is short and close to that experimentally observed in E. Coli [4]. The burst *duration* – i.e. the time duration of a

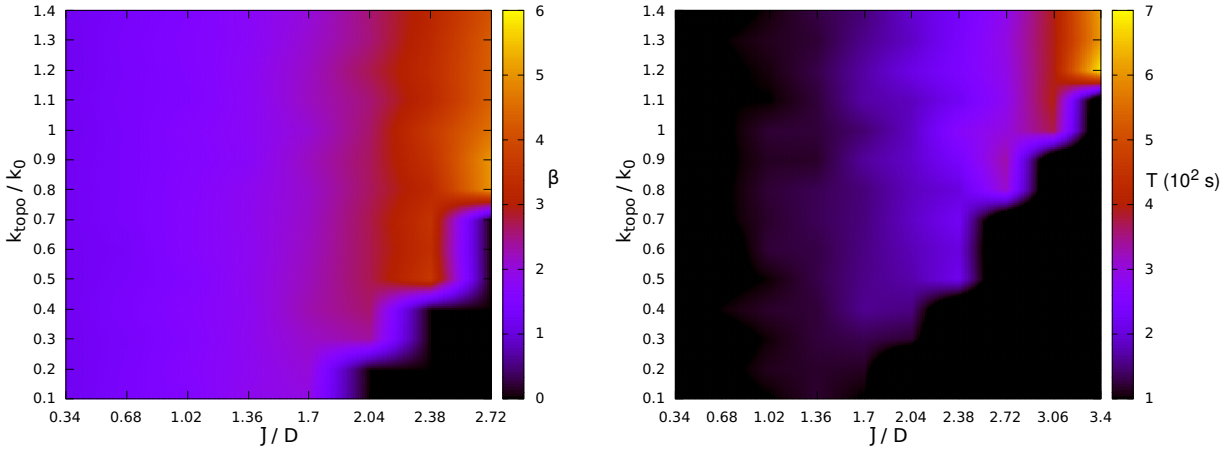
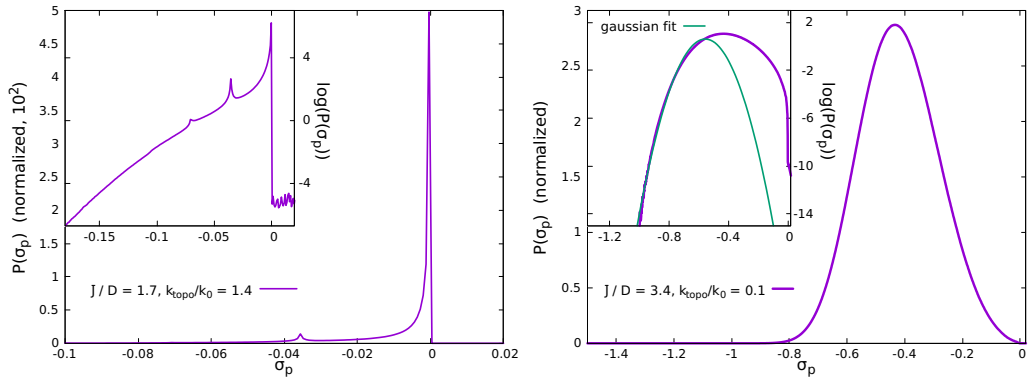


FIG. S2: **Burst size and burst duration for a single gene**. **(a)** Here, we show the burst size β up to $\bar{J}/D = 2.72$, in order to highlight the region of higher burst significance ($\bar{J}/D \sim 1.5 - 2$), in which β is in agreement with experimental measurement of the same parameter [4]. **(b)** Duration of bursts T. We find that in the same region the duration of bursts is also consistent with [4], as $T \sim 3 - 4$ min.

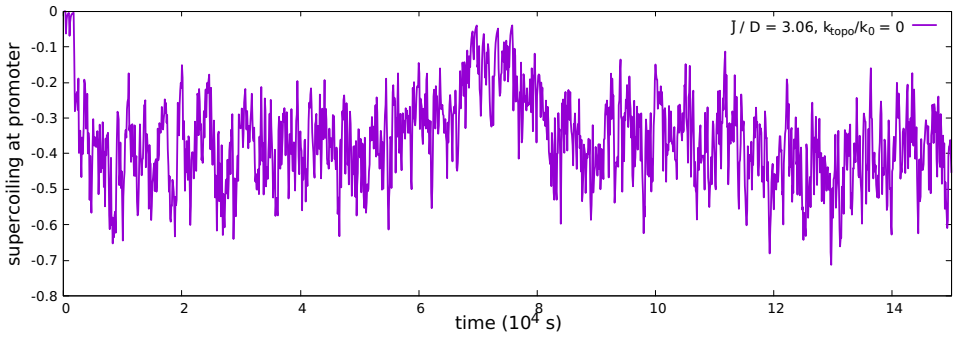
single burst – is estimated by $\beta\tau_x$: in the same parameter region it is also consistent with experimental results [4, 5], $T \sim 3 - 4$ min.

In Fig. S3a,b we present the typical probability distribution of supercoiling at the promoter, respectively in the bursty and non-bursty phases. Within our stochastic model, the supercoiling at the promoter is directly linked to the probability of initiation, and therefore its distribution encodes all of the information about the process. As expected, for bursty dynamics we observe a bimodal distribution of σ_p, while for non-bursty dynamics we have a unimodal distribution, with fluctuations approximately Gaussian.

For completeness, in Fig. S4 we consider a situation where there is a single gene but three polymerases. The rationale is that *in vivo*, at any given time, there can be more than one polymerase available for a given gene, even if the ratio of the total number of RNAP and genes is small. In this multiple polymerase case we find qualitatively similar results to the single polymerase case treated in the text, but only if we increase k_{topo} by a factor of 10. However, the burst significance is remarkably smaller than the case studied in the main text. Nevertheless, the physical features of the bursts are consistent with those of the single polymerase case: e.g., for $\bar{J}/D \sim 1$ we find $\beta \sim 3$ and $T \sim 2$ min.

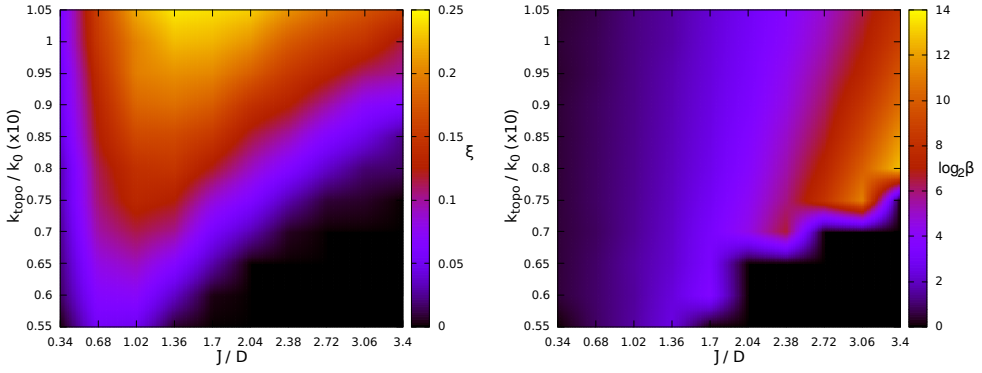


(a) **(b)**

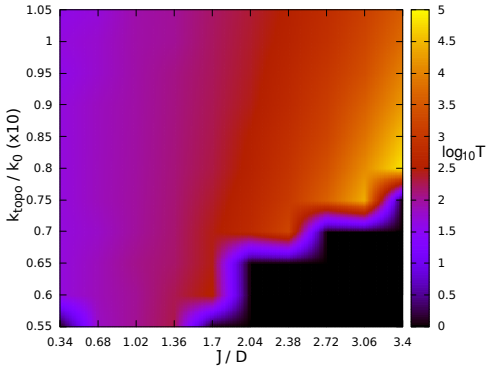


(c)

FIG. S3: **Probability distribution for supercoiling at the promoter**. **(a)** In the bursty phase, supercoiling at the promoter is strongly peaked at $\sigma_p \sim 0$. Another peak appears for more negative value of supercoiling, due to occupation of the ON state. Inset: log-linear plot of the *pdf* in the main panel. **(b)** In the non-bursty phase the distribution is unimodal, with one gaussian tail. The gene tends to more often be in a state with less negative supercoiling for longer time; this results in a non-gaussian positive tail, with a nonzero kurtosis (see main text). **(c)** Time profile of the supercoiling at the promoter in the non-bursty regime. Clearly, the gene is always ON, as the supercoiling does not relax to the initial value $\sigma_0 = 0$.



(a) **(b)**



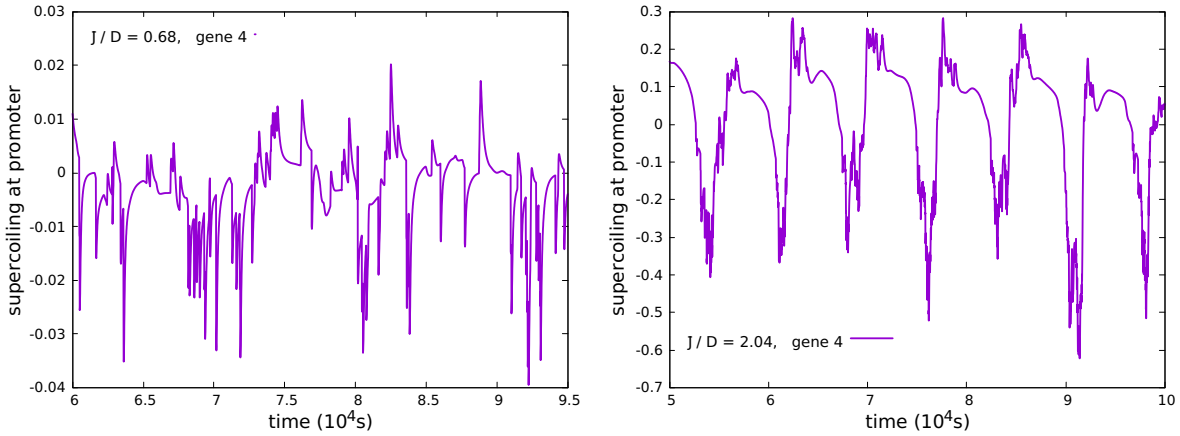
(c)

FIG. S4: **Burst parameters for a single gene and three polymerases**. **(a)** Burst significance ξ. There are still two phases separated by a crossover. **(b,c)** Burst size β and bursts duration T. With respect to the single gene case, here we have a large region of very high values (not biologically relevant) for both the burst size and the duration. However, these correspond to a region of low burst significance; in the region where burst significance is maximal we again find values for β and T consistent with experiments ($\beta \sim 3$, $T \sim 2$ min).

IV. MULTIPLE GENE SIMULATIONS: ADDITIONAL FIGURES

We present some additional results from the 10-gene array simulations for the case $k_{topo} = 0$, for which the main results are presented in the main text.

In the case of tandem genes, for the configuration shown in the main text (see Fig. 5a) the genes 1, 6 and 10 are upregulated by supercoiling. These genes have a larger space upstream of them, so are less affected by the repressive action of positive supercoils generated at their upstream neighbour. This occurs, albeit to a much lesser extent, even in the relaxed regime. This relatively small upregulation is sufficient to yield a sizeable change in the burst significance (see main text, Fig. 5b).



(a) **(b)**

FIG. S5: **Supercoiling dynamics at the promoter gene 4 in a 10-gene array**. **(a)** Promoter supercoiling versus time in the bursty regime. For $\bar{J}/D$ sufficiently small correlation between neighbours genes are established. Positive supercoiling produced by gene 3 transcription often freezes gene 4, yielding the supercoiling value transiently to the absorbing state ($\sigma_0 = 0.01$, $k_{in} = 0$). **(b)** Promoter supercoiling versus time in the supercoiling-regulated regime. In this regime the correlation spreads through the whole lattice, creating a transcription wave. Supercoiling at the promoter now oscillates in time.

In Fig. S5a we show a typical time series for the supercoiling at the promoter of the gene 4 (which is not upregulated), when the burst significance is high ($\bar{J}/D = 0.68$, that is in the bursty transcriptional regime). In Fig. S5b we show the supercoiling time series in the *wavy* regime: a periodic pattern appears so that the dynamics is no longer bursty. In Figure

Fig. S6 we show the probability of transcription for each gene, in the bursty regime, for two different values of the flux $\bar{J}/D$.

In Fig. S7 we show the mean size of bursts β and the duration of bursts T for simulations of tandem genes. Even in this case we have a good agreement with the results for a single gene. Indeed we find that in the region of higher burst significance we have $\beta \gtrsim 2$ and $T \sim 4 - 5$ min.

For completeness, in Fig. S8 we show the non-Gaussian parameters for the distribution of the supercoiling at the promoter σ_p, already computed for a single gene in the main text. The skewness and the kurtosis are not well-correlated to the burst significance, and the values depend strongly on the particular gene considered in a given configuration. However, for each gene individually, the skewness/kurtosis displays a decreasing/increasing trend as a function of ξ.

In arrays with a pair of divergent genes, the transcription probability for different genes starts to differ as soon as the value of the flux is large enough to give rise to supercoiling mediated interaction (positive feedback loop) between the two divergent genes, Fig. S9a. As a consequence, the burst significance ξ also differs among the genes. Since for high value of the flux ($\bar{J}/D \sim 1$) the transcription across all genes is almost totally dominated by the pair of divergent genes, we find that the latter behave like a single upregulated gene. This can

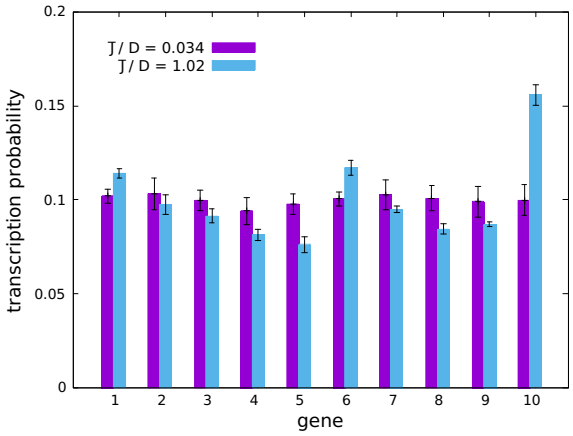
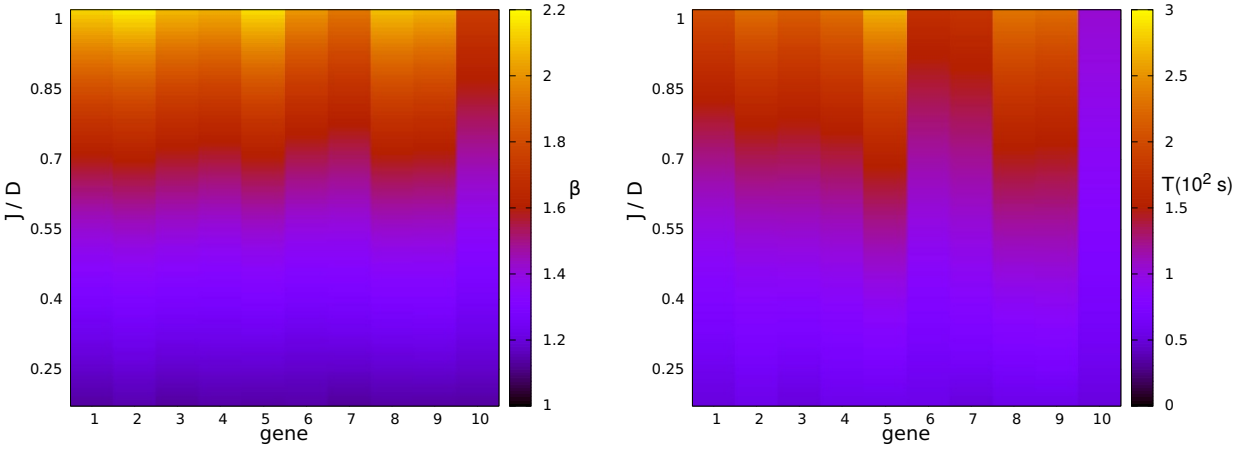


FIG. S6: **Transcriptional probability in the tandem 10-gene array.** The histograms show the transcription probability for each gene, for two different values of the flux, $\bar{J}/D = 0.034$ and $\bar{J}/D = 1.02$, that are the bottom and the top part of the diagram in Fig. 5b in the main text, respectively.

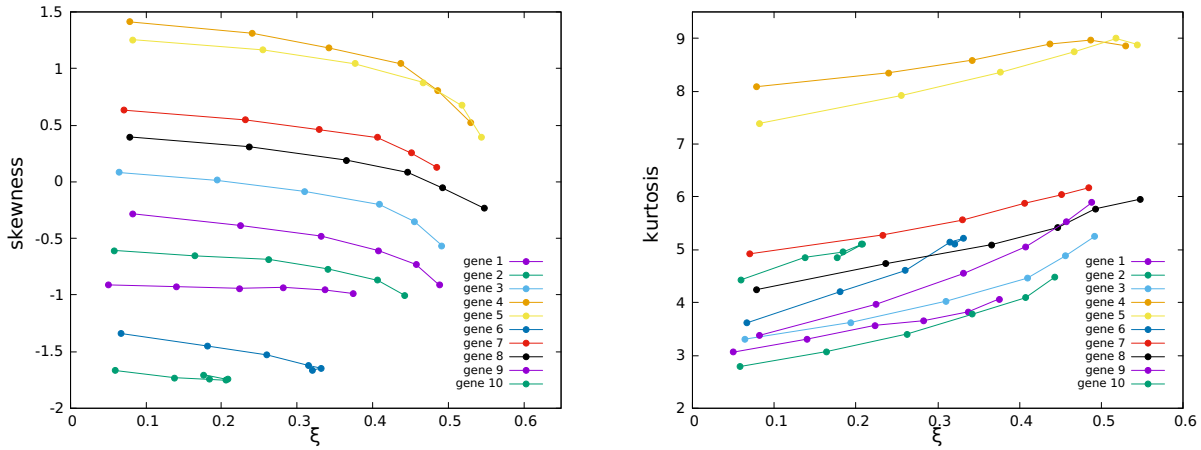


(a) **(b)**

FIG. S7: **Burst size and burst duration in a 10 genes array**. **(a)** Burst size β. **(b)** Burst duration T.

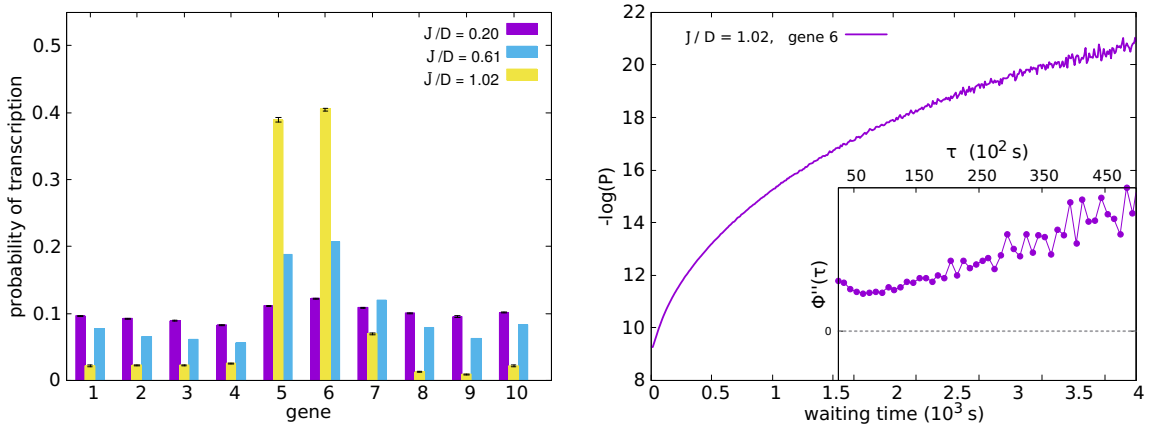
be seen by looking at the distribution of waiting times of one of the two genes (Fig. S9b), which clearly does not display two separate timescales.

As for the tandem setup, we show the burst parameters β and T in the presence of a pair



(a) **(b)**

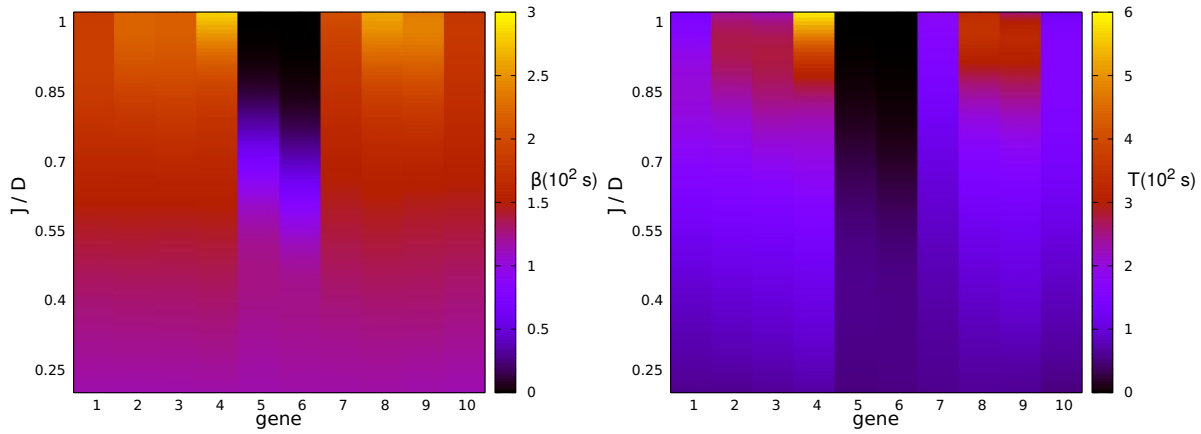
FIG. S8: **Non-Gaussian parameters for simulations with tandem genes**. **(a)** Skewness and **(b)** kurtosis as a function of ξ. For each gene the slowness and the burst significance are computed for different value of the flux $\bar{J}/D$. For each gene the skewness decreases as the bursts significance decreases, whereas the kurtosis increases.



(a) **(b)**

FIG. S9: **Transcription probability and waiting times distribution for gene 6 in a 10 gene array, with a pair of divergent genes**. **(a)** Transcription probability for each gene. For small values of the flux ($\bar{J}/D = 0.2$, purple boxes) genes are almost equally transcribed. As the flux increases, divergent genes start to dominate the dynamics, and their transcription probability increases, while the others are virtually silenced ($\bar{J}/D = 1.02$, yellow boxes). **(b)** Log-linear plot of the waiting time distribution. The system does not display any bistability. Instead, the system visits several states, each of them described by a particular value of supercoiling at the promoter and a corresponding typical waiting time. Inset: the second derivative of $\Phi(\tau)$ does not display zeros, corresponding with the absence of two separate timescales.

of divergent genes (genes 5 and 6, see Fig. S10). We note that the size of bursts is barely greater than 2 for high values of $\bar{J}/D$, since bursty genes are strongly down-regulated.



(a) **(b)**

FIG. S10: **Burst size and burst duration in the presence of divergent genes**. **(a)** Burst size β and **(b)** Burst duration T.

[1] C. A. Brackley, J. Johnson, A. Bentivoglio, S. Corless, N. Gilbert, G. Gonnella, and D. Marenduzzo, *Phys. Rev. Lett.*, **117**, 018101 (2016).

[2] M. Dobrzyński, and F. J. Bruggerman, *Proc. Natl. Acad. Sci.*, **106**, 2583 (2009).

[3] N. Kumar, A. Singh, and R. V. Kulkami, *PLoS Comput. Biol.*, **11**, 1004292 (2015).

[4] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, *Cell*, **123**, 1025 (2005).

[5] S. Chong, C. Chen, H. Ge, and X. S. Xie, *Cell*, **158**, 314 (2014).