Bursting noise in gene expression dynamics: Linking microscopic and mesoscopic models
Yen Ting Lin ∗ and Tobias Galla†
Theoretical Physics, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK
(Dated: August 5, 2015)

The dynamics of short-lived mRNA results in bursts of protein production in gene regulatory networks. We investigate the propagation of bursting noise between different levels of mathematical modelling, and demonstrate that conventional approaches based on diffusion approximations can fail to capture bursting noise. An alternative coarse-grained model, the so-called piecewise deterministic Markov process (PDMP), is seen to outperform the diffusion approximation in biologically relevant parameter regimes. We provide a systematic embedding of the PDMP model into the landscape of existing approaches, and we present analytical methods to calculate its stationary distribution and switching frequencies.

Transcription and translation in the process of gene expression occur at the molecular level and in environments of relatively small copy numbers. The discreteness of the molecular dynamics and the inherent randomness with which reactions occur are known as ‘intrinsic noise’. It is now widely accepted that intrinsic noise plays an important role in gene regulatory networks1–3. It promotes epigenetic diversity and enhances the adaptability of a single phenotype in changing environments4,5. To investigate the effects of intrinsic noise, mathematical models at different levels have been constructed, ranging from microscopic models6,6–10 describing the finer origins of intrinsic noise to mesoscopic models11–15. While the former capture the biological processes in more detail, the latter are computationally scalable and constructed to model more complex networks. These models all capture some signatures of intrinsic noise, but the detailed implementation of stochasticity varies from model to model. It is then important to consider how noise propagates between different levels of mathematical modelling. At present coarse-grained models are often proposed ad hoc and not derived from the more detailed lower-scale models. Is this always mathematically appropriate? What statistics of noise should modellers use at the different levels of coarse graining? What are the consequences of the choice of noise statistics, and what are the pitfalls in deriving models on the meso-level from finer models on smaller scales? These are some of the questions we aim to address in this work.

The above difficulties in transitioning between different levels of modelling can nicely be illustrated in the context of biological switches. These are systems with different metastable states and the possibility to ‘switch’ between those states. Biological organisms with such behaviour include the Lac switch6 in Escherichia coli and the Enterobacteria phage λ switch7. Computational and mathematical models of these range from very detailed descriptions6,7 over individual-molecule approaches8–10,16 to mesoscopic models11,12,14,15.

The difficulties in connecting these different levels of modelling biological switches are amplified by the recent recognition that the mRNA populations are essential to describing the statistics of regulatory processes16,17. Biologically, mRNA molecules are a relatively short-lived source compared to the proteins into which they ultimately translate. Protein production from a given mRNA molecule proceeds while it exists, but ceases after the mRNA decays. This leads to a production of protein in bursts—that is, the production is active for a relatively short and random period of mRNA lifetime, and during that time a random number of proteins is generated. This phenomenon is termed translational bursting1 and it can be observed in single-molecule experiments18. While some mesoscopic models account for such bursting14,15, the theoretical investigation of these processes is often limited to their stationary distribution and frequently does not include dynamic features such as switching times.

The aim of our work is to investigate the effects bursting noise in gene regulatory networks9–11,13,16,19–21, and to construct connections between individual-based models and mesoscopic approaches. Specifically, we start from microscopic and individual-molecule-based models of a toggle switch and set out to construct coarse grained, mathematically tractable models without systematically biasing the outcomes.

RESULTS

Different scales of individual-based models of a toggle switch network. We compare four individual-based models and investigate the effect of bursting noise in a toggle switch network. The first model we consider describes both the mRNA and the protein population dynamics16. Fig. 1a illustrates the Markovian model of the regulatory network. Genes X and Y are transcribed into mRNA X and mRNA Y, respectively, which in turn are translated to produce proteins X and Y. The transcription of each of the two genes is suppressed by pro-
teins of the respective other type via a Hill function $H(N) = K\left[\gamma_0 + r/[1 + (N/K)^n]\right]$, where $N$ stands for the number of suppressing proteins. The model parameter $K$ represents a typical population scale of the proteins, and the parameters $r$ and $r_0$ set the minimal $(r_0 K)$ and maximal transcription rates $((r_0 + r)K)$. The parameter $n > 0$ is the so-called Hill coefficient which models the cooperative binding of the repressors. More details of the reaction scheme can be found in the Supplementary Information. Proteins of either type, and the mRNA molecules degrade with constant rates $\gamma_0$ and $\gamma$ respectively. Biologically, mRNA molecules degrade much faster than the proteins do $(\gamma \gg \gamma_0)^{2,14,18}$. The translation rate of the mRNA is parametrised by $\gamma B$ where the parameter $B$ is the relative frequency of protein production to mRNA degradation. In this parametrisation, the number of proteins one single mRNA molecule produces during its lifetime is a geometrically distributed random variable with mean $B$ (see Supplementary Information). Biologically the parameter $B$ varies depending on the type of product protein. We assume $B \gtrsim 10$ in this work$^2$ to investigate the effect of translational bursting. Together with the relatively short lifetime of mRNA molecules, this constitutes the origin of ‘translational bursting’ in the model$^{14,25}$: a relatively large number of protein molecules is synthesized in a relatively short period of time.

For simplicity, the process in Fig. 1a is assumed to be symmetric with respect to X and Y, but the analysis is easily generalised to asymmetric circuits. In Table I we list a set of estimated values of the parameters for the model organism E. coli, along with relevant references.

In the context of this work the model just described constitutes the most detailed model we will investigate and compare against. It serves as a starting point for the derivation of more coarse grained models, and for these purposes we will refer to it as the ‘full model’ (FM) in the following.

The FM describes both the mRNA and the protein populations, hence it constitutes a relatively high-dimensional system which complicates the mathematical analysis. Notably, the only role of mRNA in the FM is to generate proteins, and so mRNA can be left out, so long as the correct statistics of protein production is retained. The timescale separation between the mRNA and protein lifetimes leads to the following reduced model describing only the protein dynamics. In the limit of infinitely-fast mRNA degradation $(\gamma \gg \gamma_0)$, proteins are generated instantaneously in bursts of geometrically distributed sizes with a mean $B$, and in between bursting events protein populations decay with rate $\gamma_0$. We will refer to the reduced model as the GB model (geometrically distributed bursts), see Fig. 1b$^{17,24}$. In the GB model, the transcription rates are regulated via the Hill function exactly as before in the FM.

A further reduction of the GB model involves replac-
Table I. Parameter set.

| Parameter | Description                                    | Value | Unit          | Reference |
|-----------|------------------------------------------------|-------|---------------|-----------|
| B         | Average number of protein each mRNA produces  | 30    | molecule      | 2         |
| γ         | mRNA degradation rate                          | 30    | 1/(cell cycle)| 2         |
| γ₀        | Protein degradation rate                       | 1.0   | 1/(cell cycle)| 13,22     |
| r         | Maximum suppressed transcription rate           | 6/100 | 1/(cell cycle)| 13,23a    |
| r₀        | Basal transcription rate                       | 1/150 | 1/(cell cycle)| 13,23b    |
| K         | A typical population scale of the proteins     | 200   | molecule      | 13,23c    |
| n         | Hill coefficient                               | 3.0   | Dimensionless | 13,19,23  |

a In $r = 1.8$ and the time unit is defined as the inverse of the protein degradation rate. In our full model we use this value, normalized by to the mean burst size $B = 30$ molecules ($r = 1.8/30 = 0.06$).

b In $r₀ = 0.2$. After normalising with respect to the burst size 30, we obtain $1/150$. In $r₀ = 0.5r$, which is of the same order as $K$.

c In $K$ is set to be 200 molecules. In only the deterministic dynamics are provided and $r + r₀ = 4.0$. To match the protein population scale $≈ 400$ in $r₀ = 400$, resulting in a typical population scale of the proteins $K ~ 100$ molecules, which is of the same order as that of $K$. 

Only the GB model approximates the stationary distribution of the FM. Numerical simulations of each of the models are carried out using standard methods. In the following we present statistical properties of the models, leaving typical sample paths to the Supplementary Information. Fig. 2 displays the numerically computed stationary distributions for the FM, GB, CB and NB models. They illustrate that the profiles of protein expressions in different model settings are quite distinct. This is due to the different representation of the underlying intrinsic noise.

While the stationary distributions of the FM and the GB model are in good agreement with each other, substantial discrepancies from the full model are found in the CB and NB models. In the CB model the stationary distribution of protein numbers is very localised compared to the FM and the GB model. In the NB model the probability distribution is even more sharply concentrated. This is because the NB model misses out two pertinent sources of noise. Bursting production in the CB model amplifies the stochasticity of transcription events and leads to a broadening of the protein distribution. Adding randomly distributed burst sizes (GB model) introduces further stochasticity, and diversifies of protein numbers even further.

Based on these results, we conclude that the bursting noise introduced by the mRNA populations significantly broadens the stationary distribution. In addition, the GB model approximates the mean first switching time of the FM. The toggle switch has two dynamic attractors, one in which protein X is highly expressed and where protein Y has a low concentration, and the other with inverted roles by symmetry. Starting from one attractor the switch can be driven to the other attractor by fluctuations. The timescale of such a transition quantifies the dynamical stability: the longer the timescale, the more stable the system is at the initial position. As we will study next, the way in which the bursting production of protein is implemented significantly affects the timescale of these switching processes.

Starting from initial condition $N_X(0) = n_{x,0}$ and $N_Y(0) = n_{y,0}$, we define the first switching time as the time it takes a sample path to reach the symmetric boundary $N_X = N_Y$. Mathematically, the first switching time is a random variable. The mean first switching time (MFST) is then the average value of the random first switching time. The MFST depends on the initial condition $(n_{x,0}, n_{y,0})$.

Sweeping across the space of possible initial configuration, the MFST of the FM and of the GB model are measured in simulations and presented in Fig. 3. We show the MFST of the CB in the Supplementary Information. As with the stationary distributions, the data in Fig. 3 indicates that the GB model approximates the switching times of the full model to a good accuracy. We remark that the MFST of the CB model is almost
Diffusion approximation of the GB model. The evolution of the protein population in the GB model is described by a master equation (Supplementary Information). Solving master equations mathematically is however difficult and mostly limited to linear dynamics. The only realistic way forward for a theoretical analysis is often the so-called diffusion approximation.

In the diffusion approximation, the discrete-molecule process is approximated by a Gaussian process for continuous concentrations—numbers of the different types of molecules normalized by a typical population scale. The Gaussian process satisfies a diffusion equation (the Fokker–Planck equation)\(^{29,30}\). Based on these methods, it is often possible to calculate or approximate the stationary behaviour and switching times of model gene networks. For existing studies in the context of toggle switches see\(^{11-13}\).

Deriving the diffusion approximation of the GB model requires modest modifications to the standard Kramers–Moyal expansion\(^{29,30}\). These modifications are necessary to account for the randomness induced by the geometrically distributed burst size. Details of the derivation can be found in the Supplementary Information, we here only report the final outcome. The expansion results in\(^\text{two coupled Itō stochastic differential equations for the concentrations } x_t = N_x(t)/K \text{ and } y_t = N_y(t)/K. \text{ These are valid in the limit of large but finite populations}^{31} \text{ and of the form}\)

\[
\begin{align*}
    dx_t &= v(x_t, y_t)dt + \sqrt{D(x_t, y_t)}dW_t^{(x)}, \\
    dy_t &= v(y_t, x_t)dt + \sqrt{D(y_t, x_t)}dW_t^{(y)},
\end{align*}
\]

with drift \(v\) and diffusion \(D\) given by

\[
\begin{align*}
    v(w, z) := B \left( r_0 + \frac{r}{1 + z^n} \right) - \gamma_0 w, \\
    D(w, z) := \frac{B}{K} \left( 2B + 1 \right) \left( r_0 + \frac{r}{1 + z^n} \right) + \frac{\gamma_0}{B} w.
\end{align*}
\]

The quantities \(dW_t^{(x)}\) and \(dW_t^{(y)}\) represent independent Wiener processes.

The diffusion approximation can only be expected to be accurate when molecule numbers are large so that the concentrations \(x_t\) and \(y_t\) are effectively continuous. In principle, a similar analysis can also be applied to the master equations of the full model. In the FM mRNA numbers are rather small though (typically < 5, see Supplementary Information), so the Gaussian approximation does not capture the statistics of the intrinsic noise faithfully. Similarly further analysis of the CB and NB models can be carried out based on the diffusion approximation. Given that CB and NB models fail to reproduce the behaviour of the FM, these results are relegated to the Supplementary Information.

Results from simulating the Gaussian process of equations (1) are shown in Fig. 4. While the data for the stationary distribution (Fig. 4a) looks similar to that of the full model (Fig. 2a), noticeable discrepancies are manifest in the mean first switching times (compare Fig. 4b and Fig. 3a). In Fig. 4c and d, we show the differences between simulation outcomes of the full model and those of the diffusion approximation of the GB model. Although the GB model itself approximates...
Constructing a mesoscopic piecewise deterministic Markov process. We have seen that the diffusion approximation of the GB model fails to reproduce the statistics of the full model. This underlines the need to construct coarse-grained models directly from the full model and without the intermediate step of a protein-only dynamics. We now proceed to introduce such a model. As before we describe protein concentrations by continuous variables, $x$ and $y$. The mRNA dynamics are captured by introducing three ‘states’: The 0-state describes phases in which no mRNA is present. In the X-state there is one mRNA of type X and protein X is generated with rate $\gamma b$. The quantity $b = B/K$ is the mean burst size in the unit of protein concentration. No proteins of type Y are produced in the X-state. Similarly, in the Y-state protein Y is generated with rate $\gamma b$. Both types of protein are subject to natural degradation with rate $\gamma_0$ in any of the three states.

This is described by the following deterministic differential equations:

$$
\begin{align*}
0\text{-state: } & \dot{x} = -\gamma_0 x \quad \text{and} \quad \dot{y} = -\gamma_0 y, \\
X\text{-state: } & \dot{x} = \gamma b - \gamma_0 x \quad \text{and} \quad \dot{y} = -\gamma_0 y, \\
Y\text{-state: } & \dot{x} = -\gamma_0 x \quad \text{and} \quad \dot{y} = \gamma b - \gamma_0 y.
\end{align*}
$$

The rates with which the system transits between the states are based on the dynamics of the FM:

$$
\begin{align*}
0\text{-state} \xrightarrow{H(Ky)} X\text{-state}, & \quad X\text{-state} \xrightarrow{\gamma} 0\text{-state}, \\
0\text{-state} \xrightarrow{H(Kx)} Y\text{-state}, & \quad Y\text{-state} \xrightarrow{\gamma} 0\text{-state}.
\end{align*}
$$

No transitions occur directly between the X and Y states. The kinetic scheme is illustrated in Fig. 1d.

The stochasticity and discreteness of the mRNA populations is reflected in the random transitioning between the 0-, X- and Y-states. Between these Markovian events the protein concentrations evolve deterministically. We will refer to this model as the piecewise deterministic Markov process (PDMP).

Notably, at most one mRNA molecule of either type can be present in PDMP at any time. Although the model can be generalised to allow more than one mRNA molecule, the analysis below shows that the lowest-order approximation is sufficient to capture the relevant fluctuations of the mRNA dynamics.
The PDMP approximation outperforms the diffusion approximation of the GB model. As in the GB model, we work in the limit of infinitely fast degrading mRNA ($\gamma \to \infty$). Simulations of the PDMP model in this limit can be carried out using a minor modification of a previously proposed algorithm \cite{15}. We measure the stationary distribution of the PDMP model and the mean first switching times for different initial protein numbers. Results are shown in Fig. 5a and b, and we compare the outcome against that of the full model in Fig. 5c and d.

The simulation data indicate that the PDMP approximation outperforms the diffusion approximation of the GB model, and it provides a more faithful approximation to the FM. This is because the diffusion approximation introduces Gaussian noise. It retains some information about the variance of protein production and degradation, but it does not capture the geometrically distributed burst sizes in the GB model well enough. The PDMP approximation, on the other hand, models exponentially distributed bursts in protein concentration. The exponential distribution in the PDMP model is the analogue of the geometric distribution in the discrete-molecule GB model. While the PDMP model is an approximation as well, it retains the typical characteristics of the stationary distribution and switching times of the original model. At the same time the PDMP model is suitable for further mathematical analysis (see below).

When does the PDMP outperform the diffusion approximation? We now investigate the robustness of these findings. In Fig. 6 we vary two essential parameters, the mean burst size $B$, and the population scale $K$, while keeping the other parameters fixed. We measure the Jensen–Shannon distance \cite{32,33} between the resulting stationary distributions of the PDMP and that of the full model. Data is shown in Fig. 6a and c. We also compare the mean first switching times starting from one of the stable modes, see Fig. 6b and d. The figure also shows results from the diffusion approximation of the GB model.

Results indicate that PDMP model outperforms the diffusion approximation of the GB model for mean burst sizes of $B \gtrsim 5$. We conclude that the bursting noise has to be considered in this biologically relevant regime \cite{24}. The PDMP model incorporates only the bursting noise and neglects the demographic noise from random degradation of the proteins. The strength of this demographic noise is proportional to $1/\sqrt{K}$. The results in Fig. 6c and d indicate that the difference in describing intrinsic noise propagates to physical observables even when the noise is weak ($K \approx 1000$ for fixed $B = 30$).

Analytic investigation of the PDMP process. The simplicity of the PDMP approach allows us to proceed with a mathematical analysis. We here only outline the main steps, further details are reported in the Supplementary Information. We denote the probability density that the system is in the 0-state and with protein densities $x, y$ at time $t$ by $p_0(x, y, t)$. Similarly we write $p_X(x, y, t)$ and $p_Y(x, y, t)$ when the system is in the $X$- or $Y$-states. The evolution of these distributions then follows the forward equation

$$
\frac{\partial}{\partial t} \begin{bmatrix} p_0 \\ p_X \\ p_Y \end{bmatrix} = \begin{bmatrix} L^+_d + L^+_s \end{bmatrix} \begin{bmatrix} p_0 \\ p_X \\ p_Y \end{bmatrix},
$$

where $L^+_d$ and $L^+_s$ drive the deterministic flow and the random switching between states respectively. These operators are of the form

$$
L^+_d := \begin{bmatrix} (L^+_d)_{11} & 0 & 0 \\ 0 & (L^+_d)_{22} & 0 \\ 0 & 0 & (L^+_d)_{33} \end{bmatrix},
$$

$$
L^+_s := \begin{bmatrix} -H(Kx) & -H(Ky) & \gamma & \gamma \\ H(Ky) & -\gamma & 0 & 0 \\ H(Kx) & 0 & -\gamma \end{bmatrix},
$$

$H$ is a Heaviside function.
with

\[
\begin{align*}
\left( L^\dagger_d \right)_{11} & := \gamma_0 \partial_x (x) + \gamma_0 \partial_y (y), \\
\left( L^\dagger_d \right)_{22} & := \partial_x (-\gamma b + \gamma_0 x) + \gamma_0 \partial_y (y), \\
\left( L^\dagger_d \right)_{33} & := \gamma_0 \partial_x (x) + \partial_y (-\gamma b + \gamma_0 y).
\end{align*}
\]

(7a), (7b), (7c)

The differential operators \( \partial_x \) and \( \partial_y \) act on all that follows to their right, including the probability densities \( p_0, p_X \) and \( p_Y \) outside the matrix notation in equations (5) and (6).

The PDMP approximation applies in the limit \( \gamma \to \infty \), i.e., fast return into the 0-state. The resident time in the X- and Y-states is exponentially distributed and scales as \( \gamma^{-1} \). It formally tends to zero as \( \gamma \to \infty \). On the other hand the translation rate \( \gamma B \) tends to infinity in this limit. Combining the limiting behaviours of resident time and translation rate results in an exponentially distributed increment of protein concentration in each cycle of switching from the 0-state to the X- or Y-state, and then returning to the 0-state. As a consequence the PDMP converges to previously proposed continuous-state bursting models\textsuperscript{14,15,18} in the limit \( \gamma \to \infty \), and \( p_0(x, y, t) \) satisfies

\[
\begin{align*}
\partial_t p_0 & = \partial_x (\gamma_0 x p_0) + \partial_y (\gamma_0 y p_0) - [H(Kx) + H(Ky)] p_0 \\
& \quad + H(Kx) \int_0^x \frac{1}{b} e^{-\frac{x-x'}{b}} p_0 (x', y', t) \, dx' \\
& \quad + H(Ky) \int_0^y \frac{1}{b} e^{-\frac{y-y'}{b}} p_0 (x, y', t) \, dy',
\end{align*}
\]

(8)
as detailed in the Supplementary Information.

**Analytic investigation of the mean first switching time.** One of the strengths of the PDMP formulation (equations (5) and (6)) is the relative ease with which mean first switching times can be obtained. We first proceed by computing mean escape time from an arbitrary open domain \( \Omega \). The mean first switching time can be calculated by setting \( \Omega = \{ (x, y) : x < y \} \), recognising that the process can only exit this domain by crossing the boundary \( x = y \).

Suppose, the system is initially at \( (x, y) \in \Omega \), and in state \( Z \in \{0, X, Y\} \). We write \( T_Z(x, y) \) for the mean first time at which the process exits the domain \( \Omega \). The quantities \( T_Z \) then satisfy the following backward equation\textsuperscript{29,30}

\[
- \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = (L_d + L_s) \begin{bmatrix} T_0(x, y) \\ T_X(x, y) \\ T_Y(x, y) \end{bmatrix},
\]

(9)

where \( L_d \) and \( L_s \) are adjoint to the operators in equations (5) and (6). They are given by

\[
L_d := \begin{bmatrix}
(L_d)_{11} & 0 & 0 \\
0 & (L_d)_{22} & 0 \\
0 & 0 & (L_d)_{33}
\end{bmatrix},
\]

(10a)

\[
L_s := \begin{bmatrix}
-H(Kx) - H(Ky) & H(Kx) & H(Ky) \\
\gamma & -\gamma & 0 \\
\gamma & 0 & -\gamma
\end{bmatrix}
\]

(10b)

In the infinitely-fast degrading mRNA limit \( (\gamma \to \infty) \), and using appropriate boundary conditions (Supplementary Information) we arrive at

\[
-1 = \left[ -\gamma_0 \partial_x - \gamma_0 y \partial_y - H(Kx) - H(Ky) \right] T_0(x, y)
\]

\[
+ H(Ky) \int_x^y \frac{e^{-\frac{x-x'}{b}}}{b} T_0(x', y) \, dx'
\]

\[
+ H(Kx) \int_y^\infty \frac{e^{-\frac{y-y'}{b}}}{b} T_0(x, y') \, dx'.
\]

(12)

This is the adjoint equation\textsuperscript{29} of the expression in equation (8) on the open domain \( \Omega \). Equation (12) is solved by a finite difference method, noting that it is self-consistent and no boundary condition needs to be specified. The solution is shown in Fig. 7, and reproduces the simulation outcome of the FM well.

We remark that equation (12) is only valid for the half-plane \( \Omega \). A detailed discussion can be found in the Supplementary Information.

**Analytic investigation of the weak-noise limit.** The analytical calculation of the stationary distributions
of the PDMP model can be pursued further using the so-called Wentzel-Kramers-Brillouin (WKB) method. This technique is based on the ansatz

\[ p_{\text{stat}}(x,y) = \exp \left[ -\frac{1}{b} \sum_{\ell=0}^{\infty} b^\ell S_\ell(x,y) \right], \tag{13} \]

where \( b = B/K \ll 1 \). One proceeds by considering \( (L_d^+ + L_s^-) p_{\text{stat}}(x,y) = 0 \) order-by-order in \( b \). To leading order we find the Hamilton-Jacobi equation

\[
0 = \left[ \gamma_0 x - B h(y) \right] \partial_x S_0 + \left[ \gamma_0 y - B h(x) \right] \partial_y S_0 + \gamma_0 x (\partial_x S_0)^2 + \gamma_0 y (\partial_y S_0)^2 + \gamma_0 x (\partial_x S_0) (\partial_y S_0) + \gamma_0 y (\partial_x S_0) (\partial_y S_0), \tag{14}
\]

where \( h(z) := H(Kz)/K \). This equation is then numerically solved using the algorithm of Heymann and Vanden–Eijnden\(^{34} \). Results are shown in Fig. 8. Even though this only provides a first-order approximation and despite the fact that we have used \( b = 0.15 \) (which is not very small) we obtain a reasonable agreement with the stationary distribution in Fig. 5a.

For completeness we have also carried out a WKB analysis of the diffusion approximation of the GB, CB and NB models. These are presented in the Supplementary Information.

The leading order function \( S_0(x,y) \) is the so-called ‘rate function’ which quantifies the rare-event statistics of the process in the weak-noise limit \( b \ll 1 \).\(^{35,36} \) Several studies have suggested that \( S_0(x,y) \) is a suitable candidate for a ‘landscape’ of the non-equilibrium random processes in models of gene regulatory networks\(^{13,21,37} \).

The Hamilton–Jacobi equation (14) contains cubic terms such as \( (\partial_x S_0)^2 (\partial_y S_0) \), while diffusion equations are quadratic in derivatives of \( S_0 \). This illustrates the fundamental difference between the statistics of intrinsic noise in the diffusion approximation and the bursting noise in PDMP. Further more rigorous mathematical investigations into these differences would be very welcome in our view.

We compare the functions \( S_0 \) of the PDMP and the diffusion approximation of the GB model in Fig. 8. One observes a much ‘shallower’ rate function in the PMDP model, especially at larger protein numbers \( (N_X, N_Y \approx 700) \). This is due to the long tails in the exponential bursting kernel of the PMDP model, which are not present in the diffusion approximation of the GB model. Such a fat-tail bursting kernel enhances the probability for the system to evolve to high protein concentrations. We identify this as the origin of the qualitatively distinct rare-event statistics in the two models.

FIG. 8. Rate functions \( S_0 \) as functions of the protein numbers \( 0 \leq N_X, N_Y \leq 700 \) on a linear scale. (a) PDMP model; (b) Diffusion approximation of the GB model.

Effects of bursting noise in a multi-switch network. Recently, multi-switch systems have gained interest\(^{13,21,37} \). A schematic diagram of the three-way switch network proposed by\(^{13} \) is shown in Fig. 9a. It is obtained from the classical toggle switch network by including a self-enhancing autoregulation. Our computational and mathematical setup requires only minor modifications to include generalisation to this case. Specifically, we replace the earlier Hill functions by

\[
G(N_X, N_Y) = q_0 \left( 1 + \frac{r_1}{(N_X/K_1)^{n_1} + 1} \right) \times \left( 1 + \frac{r_2}{(N_Y/K_2)^{n_2} + 1} \right), \tag{15}
\]

with parameters\(^{13} \) \( q_0 = 4 \), \( r_1 = -4/5 \), \( r_2 = 7/3 \), \( n_1 = 3 \), \( n_2 = 1 \), \( K_1 = 160 \), and \( K_2 = 320 \). The rest of the parameters follows Table I. The negative value of \( r_1 \) reflects the positive autoregulation. To evaluate the effects of bursting noise on this multi-switch model, we consider again the full model, the diffusion approximation of the GB model, as well as the CB and NB models of the extended network.

Fig. 9 displays the stationary distribution to illustrate the effects of the bursting noise in the multi-switch network. The model without bursts (NB, panel f) has a stationary distribution consisting of three modes, as reported earlier\(^{13} \). Inclusion of constant bursts (CB, panel e) diversifies the protein expression and reduces the stability of the mode located at \( N_X = N_Y \approx 230 \). In the full model (panel b) there is no discernible concentration of probability in the symmetric mode, hence the three-way switching capability appears to be absent. We also notice that the saddle of the distribution in the FM is located at a state with much lower number of proteins compared to the NB and CB models. The most likely switching path\(^{6} \) from one of the asymmetric modes to the other will differ significantly between the different variants of the model. The diffusion approximation of the GB model (panel c) does not capture the outcome of the FM either.
Overall these findings confirm again that the inclusion of bursty noise statistics has significant effects on the model outcome. Finally, we observe in Fig. 9d that the PDMP model approximates the full model of the three-way switch well. We conclude that randomly distributed burst sizes are again the predominant form of intrinsic noise in the multi-switch network.

DISCUSSION AND CONCLUSION

Explicitly including mRNA dynamics in gene regulatory models inevitably introduces more complexity. We have quantitatively studied the effects of bursting noise in a biologically relevant regime or the model organism *E. coli*. To our knowledge, this is one of the first which attempts to build a rigorous connection between existing individual-based models and more coarse-grained models. Results of our simulations indicate that the bursting statistics of transcription and translation are essential ingredients of models of gene regulation. Coarse grained models need to account for bursting to retain correct statistics of noise-driven phenomena such as the switching between different dynamic attractors.

The implications of our observations are relevant to the abstract modelling of regulatory networks in different ways. We are now in a better position to address our opening question, and to say how noise propagates between different levels of modelling. Perhaps more importantly, our study may ultimately help to decide what level of modelling is most appropriate to study gene regulatory circuits computationally. The answer will of course depend on the question in the focus of the investigation.

We have examined different levels of coarse graining, and we have identified the steps in these reduction procedures at which significant alternations to different model outcomes are introduced.

Systematically choosing a suitable level of coarse-graining also facilitates the mathematical analysis of regulatory networks. The high dimensionality of full regulatory network effectively makes them intractable. Model reduction is needed to make progress, and our analysis demonstrates that the PDMP formulation is a powerful way forward, and that it can be more suitable than the conventional diffusion approximation. The PDMP model explicitly retains the bursting noise originating from the mRNA dynamics. Even though it effectively disregards the demographic noise from random degradation of the proteins, it delivers accurate predictions for stationary distributions and switching times.

As another strength, the PDMP formulation can relatively easily be generalised to accommodate more complex reactions. For example, in the *Enterobacteria phage λ* switch it is not the monomer of the synthesized proteins which acts as the repressor to regulate transcription, but instead their dimer. Modelling these processes requires the inclusion of dimerization further downstream after transcription and translation. Preliminary results not shown here reveal that the PDMP approximates such dynamics well.

The fact that the piecewise deterministic Markov process is successful in approximating the full model opens a relatively new type of modelling paradigm. We acknowledge that we are not the first to propose this. Our contribution consists in a first analytical treatment of PDMP models and in a systematic embedding into a wider landscape of modelling approaches. The bursting phenomenon is ubiquitous whenever there is a separation of time scales between the source and the product of a biological process. These are mRNA and protein in models of gene regulation, but we expect that these ideas can be applied to other biological problems with similar time-scale separation.
METHODS

Sample paths of the individual-based processes (FM, CB, NB, and GB) are generated using the standard Gillespie algorithm. The PDMP process is simulated using the algorithm discussed by Bokes et al. Simulations of the stochastic differential equations resulting from the diffusion approximation are performed with a standard Euler–Maruyama scheme. The geometric minimum action method is implemented using MATLAB R2010a, as is the finite-difference scheme to solve the backward equation (12). Further details can be found in the Supplementary Information.

* yenting.lin@manchester.ac.uk
† tobias.galla@manchester.ac.uk
1 Kaern, M., Elston, T. C., Blake, W. J., & Collins, J. J. Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
2 Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **98**, 8614–8619 (2001).
3 Walczak, A. M., Mugler, A. & Wiggins, C. H. Analytic methods for modeling stochastic regulatory networks. *Methods Mol. Biol.* **880**, 273–322 (2012).
4 Thattai, M. & van Oudenaarden, A. Stochastic gene expression in fluctuating environments. *Genetics* **167**, 523–530 (2004).
5 Acar, M., Mettetal, J. T., & van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* **40**, 471–475 (2008).
6 Roberts, E., Magis, A., Ortiz, J. O., Baumeister, W. & Luthey-Schulten, Z. Noise contributions in an inducible genetic switch: A whole-cell simulation study. *PLoS Comput. Biol.* **7**, e1002010 (2011).
7 Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells. *Genetics* **149**, 1633–1648 (1998).
8 Walczak, A. M., Sasai, M. & Wolynes, P. G. Self-consistent proteomic field theory of stochastic gene switches. *Biophys. J.* **88**, 828–850 (2005).
9 Roma, D. M., O’Flanagan, R. A., Ruckenstein, A. E., Sen-gupta, A. M. & Mukhopadhyay, R. Optimal path to epigenetic switching. *Phys. Rev. E* **71**, 011902 (2005).
10 Warren, P. B. & ten Wolde, P. R. Chemical models of genetic toggle switches. *J. Phys. Chem. B* **109**, 6812–6823 (2005).
11 Wang, J., Zhang, K., Xu, L. & Wang, E. Quantifying the waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. USA* **108**, 8257–8262 (2011).
12 Wang, J., Xu, L., Wang, E. & Huang, S. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys. J.* **99**, 29–39 (2010).
13 Lu, M. Y., Omuchic, J. & Ben-Jacob, E. Construction of an effective landscape for multistate genetic switches. *Phys. Rev. Lett.* **113**, 5 (2014).
14 Friedman, N., Cai, L., Xie & X. S. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006).
15 Bokes, P., King, J., Wood, A. A. & Loose, M. Transcriptional bursting diversifies the behaviour of a toggle switch: Hybrid simulation of stochastic gene expression. *Bull. Math. Biol.* **75**, 351–371 (2013).
16 Strasser, M., Theis, F. J. & Marr, C. Stability and multi-attractor dynamics of a toggle switch based on a two-stage model of stochastic gene expression. *Biophys. J.* **102**, 19–29 (2012).
17 Assaf, M., Roberts, E. & Luthey-Schulten, Z. Determining the stability of genetic switches: Explicitly accounting for mrna noise. *Phys. Rev. Lett.* **106**, 248102 (2011).
18 Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
19 Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in escherichia coli. *Nature* **403**, 339–342 (2000).
20 Wang, J. W., Zhang, J. J., Yuan, Z. J. & Zhou, T. S. Noise-induced switches in network systems of the genetic toggle switch. *BMC Syst. Biol.* **1**, 14 (2007).
21 Lu, M. et al. Tristability in cancer-associated microRNA-TF chimera toggle switch. *J. Phys. Chem. B* **117**, 13164–13174 (2013).
22 Taniguchi, Y. et al. Quantifying e coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
23 Kobayashi, H. et al. Programmable cells: Interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. USA* **101**, 8414–8419 (2004).
24 Shahrezaei, V. & Swain, P. S. Analytical distributions for stochastic gene expression models with bursting and feedback. *Phys. Rev. Lett.* **105**, 17256–17261 (2008).
25 Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
26 Schwartz, R. *Biological modeling and simulation : a survey of practical models, algorithms, and numerical methods*. Cambridge, MA: MIT Press (2008).
27 Gillespie, D. T. Exact stochastic simulation of coupled chemical-reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
28 Kumar, N., Platini, T. & Kulkarni, R. V. Exact distributions for stochastic gene expression models with bursting and feedback. *Phys. Rev. Lett.* **113**, 5 (2014).
29 van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*. Amsterdam: Elsevier Science B.V. (2007).
30 Gardiner, C. W. *Handbook of Stochastic Methods*. Springer, New York (2009).
31 Kurtz, T. G. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probab.* **8**, 344–356 (1971).
32 Lin, J. H. Divergence measures based on the shannon entropy. *IEEE Trans. Inf. Theory* **37**, 145–151 (1991).
33 Sekine, D. M. & Schindelin, J. E. A new metric for probability distributions. *IEEE Trans. Inf. Theory* **49**, 1858–1860 (2003).
34 Heymann, M. & Vanden-Eijnden, E. The geometric minimum action method: A least action principle on the space of curves. *Comm. Pure Appl. Math.* **61**, 1052–1117 (2008).
35 Freidlin, M. I. & Wentzell, A. D. *Random Perturbations of Dynamical Systems*. New York: Springer-Verlag (2012).
36 Zhou, J. X., Aliyu, M. D. S., Aurell, E. & Huang, S. Quasi-potential landscape in complex multi-stable systems. J. R. Soc. Interface 9, 3539–3553 (2012).

37 Guantes, R. & Poyatos, J. F. Multistable decision switches for flexible control of epigenetic differentiation. PLoS Comput. Biol. 4 e1000235. (2008).

38 Hasty, J., Pradines, J., Dolnik, M. & Collins, J. J. Noise-based switches and amplifiers for gene expression. Proc. Natl. Acad. Sci. USA 97, 2075–2080 (2000).

39 Zeiser, S., Franz, U., Wittich, O. & Liebscher, V. Simulation of genetic networks modelled by piecewise deterministic markov processes. IET Syst. Biol. 2, 113–135 (2008).

40 Zeiser, S., Franz, U. & Liebscher, V. Autocatalytic genetic networks modeled by piecewise-deterministic markov processes. J. Math. Biol. 60, 207–246 (2010).

41 Huang, G. R., Saakian, D. B., Rozanova, O., Yu, J. L. & Hu, C. K. Exact solution of master equation with gaussian and compound poisson noises. J. Stat. Mech. P11033 (2014).

ACKNOWLEDGEMENT

The authors acknowledge supported by the Engineering and Physical Sciences Research Council EPSRC (UK) (grant reference EP/K037145/1). We thank Peter Ashcroft, Luis Fernandez-Laifierza, Louise Dyson, and Charlie Doering for fruitful discussions.

AUTHOR CONTRIBUTIONS

Y.T.L. conceived the study, performed analytical calculations and numerical simulations, prepared the figures. T.G. provided general direction of the study, confirming analytical calculations. Both authors wrote the manuscript.

ADDITIONAL INFORMATION

Supplementary Information accompanies this manuscript.

Competing financial interests: The authors declare no competing financial interests.
Supplementary Information

Bursting noise in gene expression dynamics: Linking microscopic and mesoscopic models

Yen Ting Lin* and Tobias Galla†

Theoretical Physics, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK
(Dated: August 5, 2015)

*yentingl@umich.edu
†tobias.galla@manchester.ac.uk

arXiv:1508.00608v1 [q-bio.MN] 3 Aug 2015
CONTENTS

I. Notation 4

II. Master equations of the individual-based models 4
   A. Master equation of full model (FM) 4
   B. Infinitely fast-degrading mRNA limit: the GB model 4
   C. Master equation of the GB model 5
   D. Master equation of CB model 6
   E. Master equation of the model without bursts (NB) 6

III. Deriving the diffusion approximations 6
   A. Diffusion approximation of the GB model 6
   B. Diffusion approximation of the CB and NB models 8

IV. The piecewise deterministic Markov process (PDMP) 8
   A. Construction of the PDMP model 8
   B. Forward equation in the limit $\gamma \to \infty$ 11
   C. Equations for the mean first switching time 12

V. WKB analysis 13
   A. WKB ansatz 13
   B. DA of the GB model 13
   C. DA of the CB model 14
   D. DA of the NB model 14
   E. PDMP model 14

VI. Numerical methods 15

VII. Sample paths of the different models 16
   A. Full model 16
   B. Geometrically distributed burst model (GB) 17
   C. Constant burst model (CB) 17
   D. No-burst model (NB) 18
   E. PDMP model 18
   F. Diffusion approximation of the GB model 19
G. Diffusion approximation of the CB model 19
H. Diffusion approximation of the NB model 20

VIII. Comparison of stationary distributions 21

IX. Comparison of the mean first switching times 22

X. Comparing WKB results 23

References 24
I. NOTATION

We briefly summarise the notation used in the main manuscript and in this supplement:

- Discrete numbers of the two types of protein are denoted by \( N_X \) and \( N_Y \). We write \( M_X \) and \( M_Y \) for the number of mRNA molecules of the two types.

- Variables such as \( x(t) \) denote continuous particle densities (or concentrations). Specifically \( x(t) \) and \( y(t) \) are protein densities, i.e., \( x = N_X/K \) and \( y = N_Y/K \) in the limit \( K \gg 1 \).

- We denote the probabilities in the master equations by capital \( P \) (discrete particle numbers).

- The lower-case notation \( p \) is used for probability density functions in the diffusion approximation (continuous particle densities/concentrations).

II. MASTER EQUATIONS OF THE INDIVIDUAL-BASED MODELS

The different individual-based models in the main manuscript are uniquely defined by their master equations. Here we briefly summarise the master equations of the FM and of the GB, CB and NB models.

A. Master equation of full model (FM)

We write \( P_{a,b,c,d} \) for the probability that the system is in state \( M_X = a, M_Y = b, N_X = c, N_Y = d \) at time \( t \). The master equation of the FM is then

\[
\frac{d}{dt} P_{a,b,c,d} = - \{ H(c) + H(d) + a \gamma [1 + B] + b \gamma [1 + B] - \gamma_0 c - \gamma_0 d \} P_{a,b,c,d} \\
+ H(c) P_{a,b-1,c,d} + H(d) P_{a-1,b,c,d} + \gamma (a + 1) P_{a+1,b,c,d} + \gamma (b + 1) P_{a,b+1,c,d} \\
+ B \gamma a P_{a-1,b,c,d-1} + B \gamma b P_{a,b,c,d-1} + \gamma_0 (c + 1) P_{a,b,c+1,d} + \gamma_0 (d + 1) P_{a,b,c,d+1}. 
\] (1)

The probability of a state is zero if any of the variables \( a, b, c \) or \( d \) are negative.

B. Infinitely fast-degrading mRNA limit: the GB model

In the kinetic scheme of the full model the mRNA decays with a rate \( \gamma \), but synthesizes a protein with a rate \( \gamma B \). Both of the rates are constants. One an mRNA is created the next event involving this mRNA particle is either the production of protein or the decay of the mRNA molecule. The
The probability that the next event is the synthesis of a protein is \( \frac{B}{B+1} \), and the probability that a decay occurs next (before production of a protein) is \( \frac{1}{B+1} \). The random number, \( \ell \), of protein molecules generated by one particular mRNA molecule during its lifetime then follows a geometric distribution

\[
g(\ell) = \left( \frac{B}{1+B} \right)^\ell \left( \frac{1}{1+B} \right).
\]  

(2)

The lifetime of an mRNA molecule is of order \( \mathcal{O}(1/\gamma) \). As a consequence, we can think of the protein-generating process as follows in the infinitely-fast decaying mRNA limit (\( \gamma \to \infty \)): As soon as an mRNA is transcribed, it immediately releases a random number of proteins \( \ell \) drawn from the distribution (2) and then decays.

### C. Master equation of the GB model

In the limit \( \gamma \to \infty \), the dynamics of the full model can effectively be coarse-grained into a single-species model

\[
\begin{align*}
\text{Gene } X & \xrightarrow{H(X)} \ell \times \text{Protein } X & \text{(transcription and translation of protein } X), & \text{(3a)} \\
\text{Gene } Y & \xrightarrow{H(Y)} \ell \times \text{Protein } Y & \text{(transcription and translation of protein } Y), & \text{(3b)} \\
\text{Protein } X & \xrightarrow{\gamma_0} \emptyset & \text{(degradation of protein } X), & \text{(3c)} \\
\text{Protein } Y & \xrightarrow{\gamma_0} \emptyset & \text{(degradation of protein } Y), & \text{(3d)}
\end{align*}
\]

where \( \ell \) is drawn from the above geometric distribution, every time one of the first two reactions fires. The master equation of this process is

\[
\begin{align*}
\frac{d}{dt} P_{c,d} &= -[H(c) + H(d) + \gamma_0 c + \gamma_0 d] P_{c,d} + \gamma_0 (c+1) P_{c+1,d} + \gamma_0 (d+1) P_{c,d+1} \\
&\quad + \sum_{\ell=0}^{c} H(d) \left( \frac{B}{1+B} \right)^\ell \left( \frac{1}{1+B} \right) P_{c-\ell,d} + \sum_{\ell=0}^{d} H(c) \left( \frac{B}{1+B} \right)^\ell \left( \frac{1}{1+B} \right) P_{c,d-\ell}. & \text{(4)}
\end{align*}
\]

We have written \( P_{c,d}(t) \) for the probability that the system is in state \( N_X = c, N_Y = d \) at time \( t \). Again, the probability of a state is zero if \( c \) or \( d \) are negative.
D. Master equation of CB model

The master equation for the CB model is obtained by replacing $g(\ell) \rightarrow \delta_{\ell,B}$, i.e., $\ell$ takes value $\ell = B$ with probability one. We find

$$\frac{d}{dt}P_{c,d} = - [H(c) + H(d) + \gamma_0 c + \gamma_0 d] P_{c,d}$$
$$+ H(d) P_{c-B,d} + H(c) P_{c-d-B} + \gamma_0 (c+1) P_{c+1,d} + \gamma_0 (d+1) P_{c,d+1}. \quad (5)$$

E. Master equation of the model without bursts (NB)

In this case we have

$$\frac{d}{dt}P_{c,d} = - [BH(c) + BH(d) + \gamma_0 c + \gamma_0 d] P_{c,d}$$
$$+ BH(d) P_{c-1,d} + BH(c) P_{c,d-1} + \gamma_0 (c+1) P_{c+1,d} + \gamma_0 (d+1) P_{c,d+1}. \quad (6)$$

III. DERIVING THE DIFFUSION APPROXIMATIONS

A. Diffusion approximation of the GB model

Simulations of the full model (FM) show that the number of mRNA molecules present at any one time is typically very small ($M_X, M_Y < 10$) when biologically relevant parameters are used. The conventional diffusion approximation relies on large particle numbers, and so it is not adequate for the full model.

Instead, we perform the diffusion approximation to the master equation of the GB model, equation (4). This is a standard method, and we proceed along the lines of [1]. The only complication is the presence of the geometrically distributed random numbers (denoted by $\ell$) in the protein-generation reactions. As we will discuss below this requires only modest modifications to the standard Kramers-Moyal expansion.

We assume that the scale of the population size, $K$, is large but finite, i.e., $K \gg 1$, and we write $x = N_X/K$ and $y = N_Y/K$, and replace $P_{c,d}(t)$ in favour of $p(x,y,t)$. The master equation (4) then
becomes
\[
\partial_t p(x, y, t) = -[H(Kx) + H(Ky) + \gamma_0 Kx + \gamma_0 Ky] p(x, y, t)
+ \gamma_0 K \left( x + \frac{1}{K} \right) p \left( x + \frac{1}{K}, y, t \right) + \gamma_0 K \left( y + \frac{1}{K} \right) p \left( x, y + \frac{1}{K}, t \right)
+ \sum_{\ell=0}^{\infty} H(Ky) \left( \frac{B}{1 + B} \right)^{\ell} \left( \frac{1}{1 + B} \right) p \left( x - \frac{\ell}{K}, y, t \right)
+ \sum_{\ell=0}^{\infty} H(Kx) \left( \frac{B}{1 + B} \right)^{\ell} \left( \frac{1}{1 + B} \right) p \left( x, y - \frac{\ell}{K}, t \right).
\]

In the last two terms we have extended the summation over \( \ell \) to infinity, terms in which \( x - \ell/K \) or \( y - \ell/K \) become negative are automatically suppressed as the corresponding probabilities \( p(x - \ell/K, y, t) \) and \( p(x, y - \ell/K, t) \) vanish.

The above expression can then be written as
\[
\partial_t p(x, y, t) = -[H(Kx) + H(Ky) + \gamma_0 Kx + \gamma_0 Ky] p(x, y, t)
+ \gamma_0 K \left( x + \frac{1}{K} \right) p \left( x + \frac{1}{K}, y, t \right) + \gamma_0 K \left( y + \frac{1}{K} \right) p \left( x, y + \frac{1}{K}, t \right)
+ H(Ky) \left\langle p \left( x - \frac{\ell}{K}, y, t \right) \right\rangle_{\ell} + H(Kx) \left\langle p \left( x, y - \frac{\ell}{K}, t \right) \right\rangle_{\ell},
\]

where \( \langle \cdot \cdot \cdot \rangle_{\ell} \) denotes an average with respect to a geometrically distributed random number \( \ell \), i.e.,
\[
\langle f_{\ell} \rangle_{\ell} = (1 + B)^{-1} \sum_{\ell} \left( \frac{B}{1 + B} \right)^{\ell} f_{\ell}.
\]

We next expand the above equation in powers of \( 1/K \), keeping only the leading and sub-leading order terms [1]. We also use the explicit expressions \( \langle \ell \rangle_{\ell} = B \) and \( \langle \ell^2 \rangle_{\ell} = B(2B+1) \) for the first two moments of the geometric distribution.

We arrive at the Fokker–Planck equation
\[
\partial_t p(x, y, t) = -\partial_x \left\{ \left[ \frac{B}{K} H(Ky) - \gamma_0 x \right] p(x, y, t) \right\} - \partial_y \left\{ \left[ \frac{B}{K} H(Kx) - \gamma_0 y \right] p(x, y, t) \right\}
+ \frac{1}{2K} \partial_x^2 \left\{ \left[ \frac{B(2B+1)}{K} H(Ky) + \gamma_0 x \right] p(x, y, t) \right\}
+ \frac{1}{2K} \partial_y^2 \left\{ \left[ \frac{B(2B+1)}{K} H(Kx) + \gamma_0 y \right] p(x, y, t) \right\},
\]

where we have written \( \partial_x = \frac{\partial}{\partial x} \) and similarly for \( \partial_y \). Realisations of the random process described by the Fokker-Planck equation (9) can be obtained as the solutions of the coupled Itô stochastic differential equations

\[
dx_t = v(x_t, y_t) dt + \sqrt{D(x_t, y_t)} dW_t^{(x)}, \quad (10a)
dy_t = v(x_t, y_t) dt + \sqrt{D(y_t, x_t)} dW_t^{(y)}, \quad (10b)
\]
with the defined drift $v$ and diffusion $D$ given by

$$v(w,z) := B \left( r_0 + \frac{r}{1 + z^n} \right) - \gamma_0 w,$$  \hspace{1cm} (11a)

$$D(w,z) := \frac{B}{K} \left[ (2B + 1) \left( r_0 + \frac{r}{1 + z^n} \right) + \frac{1}{B} \gamma_0 w \right].$$  \hspace{1cm} (11b)

The quantities $W_i^{(x)}$ and $W_i^{(y)}$ are independent Wiener processes.

**B. Diffusion approximation of the CB and NB models**

The same procedure can be applied to the master equation of the CB and NB models, and for completeness we report the resulting Fokker-Planck equations.

For the CB model one finds

$$\partial_t p(x,y,t) = - \partial_x \left\{ \left[ \frac{B}{K} H(Ky) - \gamma_0 x \right] p(x,y,t) \right\} - \partial_x \left\{ \left[ \frac{B}{K} H(Kx) - \gamma_0 y \right] p(x,y,t) \right\}$$

$$+ \frac{B}{2K} \partial_x^2 \left\{ \left[ \frac{B}{K} H(Ky) + \frac{1}{B} \gamma_0 x \right] p(x,y,t) \right\}$$

$$+ \frac{B}{2K} \partial_y^2 \left\{ \left[ \frac{B}{K} H(Kx) + \frac{1}{B} \gamma_0 y \right] p(x,y,t) \right\},$$  \hspace{1cm} (12)

and for the model without bursts (NB) one has

$$\partial_t p(x,y,t) = - \partial_x \left\{ \left[ \frac{1}{K} B H(Ky) - \gamma_0 x \right] p(x,y,t) \right\} - \partial_x \left\{ \left[ \frac{1}{K} B H(Kx) - \gamma_0 y \right] p(x,y,t) \right\}$$

$$+ \frac{1}{2K} \partial_x^2 \left\{ \left[ \frac{1}{K} B H(Ky) + \gamma_0 x \right] p(x,y,t) \right\}$$

$$+ \frac{1}{2K} \partial_y^2 \left\{ \left[ \frac{1}{K} B H(Kx) + \gamma_0 y \right] p(x,y,t) \right\}. $$  \hspace{1cm} (13)

**IV. THE PIECEWISE DETERMINISTIC MARKOV PROCESS (PDMP)**

**A. Construction of the PDMP model**

In this section we outline the construction of the PDMP approximation, starting from the full model. For this purpose it is useful to introduce the notation

$$P^{(a,b)}(c,d,t) = P(M_X = a, M_Y = b, N_X = c, N_Y = d, t).$$  \hspace{1cm} (14)

Thus the upper indices $(a, b)$ denote the number of mRNA molecules of either type in the system, and the arguments $a, b$ stand for protein numbers.
The master equation (1) can then be written in matrix form

\[
\frac{d}{dt} \begin{bmatrix}
P^{(0,0)}(c,d,t) \\
P^{(1,0)}(c,d,t) \\
P^{(0,1)}(c,d,t)
\end{bmatrix} = \mathcal{L}^{\dagger} \begin{bmatrix}
P^{(0,0)}(c,d,t) \\
P^{(1,0)}(c,d,t) \\
P^{(0,1)}(c,d,t)
\end{bmatrix}.
\] (15)

We have introduced

\[
\mathcal{L}^{\dagger} := \begin{bmatrix}
\mathcal{L}^{(0,0)} - H(c) - H(d) & \gamma & \gamma & \cdots \\
H(d) & \mathcal{L}^{(1,0)} - H(c) - H(d) - \gamma & 0 & \cdots \\
H(c) & 0 & \mathcal{L}^{(1,1)} - H(c) - H(d) - \gamma & \cdots \\
\cdots & \cdots & \cdots & \cdots
\end{bmatrix},
\] (16)

where the operator \( \mathcal{L}^{(m,n)} \) describes the forward evolution of protein numbers when there are \( M_X = m \) and \( M_Y = n \) mRNA molecules in the system. Specifically,

\[
\mathcal{L}^{(m,n)} = m \gamma B \left( \mathcal{E}^{-1,0} - 1 \right) + n \gamma B \left( \mathcal{E}^{0,-1} - 1 \right) + \gamma_0 \left[ \left( \mathcal{E}^{0,1} - 1 \right) + \left( \mathcal{E}^{0,1} - 1 \right) \right],
\] (17)

where \( \mathcal{E}^{i,j} \) are the shift operators [1] acting on functions of protein numbers. They are defined through

\[
\mathcal{E}^{i,j} f(m,n) \equiv f(m + i, n + j).
\] (18)

Next we consider the limit of fast mRNA decay, that is, large values of \( \gamma \). More specifically mRNA molecules of either type are generated with rates \( H(N_Y) \) and \( H(N_X) \) respectively, and we assume that \( \gamma \) is much larger than either of these two rates (\( \gamma \gg H(N_Y), \gamma \gg H(N_X) \), for any values of \( N_X \) and \( N_Y \)). In this limit, the system is almost always in the state without mRNA molecules (i.e., \( M_X = 0, M_Y = 0 \)), except for short spells during which there is either one molecule of mRNA of type \( X \), or one of type \( Y \). The duration of the episodes spent in these \((1,0)\) and \((0,1)\) states is of order \( \gamma^{-1} \), then a switch back to the \((0,0)\) state occurs. The probability to find the system in states with \( M_X > 1 \) or \( M_Y > 1 \) is even smaller, specifically of order \( (H/\gamma)^2 \), and we neglect contributions from these states. Equation (15) can then be simplified into a forward equation of a three-state model:

\[
\frac{d}{dt} \begin{bmatrix}
P^{(0,0)}(c,d,t) \\
P^{(1,0)}(c,d,t) \\
P^{(0,1)}(c,d,t)
\end{bmatrix} = \mathcal{L}^{\text{approx}} \begin{bmatrix}
P^{(0,0)}(c,d,t) \\
P^{(1,0)}(c,d,t) \\
P^{(0,1)}(c,d,t)
\end{bmatrix},
\] (19)

with

\[
\mathcal{L}^{\text{approx}} := \begin{bmatrix}
\mathcal{L}^{(0,0)} - H(c) - H(d) & \gamma & \gamma \\
H(d) & \mathcal{L}^{(1,0)} - \gamma & 0 \\
H(c) & 0 & \mathcal{L}^{(1,1)} - \gamma
\end{bmatrix}.
\] (20)
In the next step we consider the limit of large values of $K$. Formally we take the limit $K \to \infty$. The system can then be described by the protein concentration $x = N_X/K$ and $y = N_Y/K$. The corresponding probability distributions in the continuum limit are

\begin{align}
  p_0(x, y) &:= P^{(0,0)}(Kx, Ky) K^2, \\
  p_X(x, y) &:= P^{(1,0)}(Kx, Ky) K^2, \\
  p_Y(x, y) &:= P^{(0,1)}(Kx, Ky) K^2.
\end{align}

On the left-hand side we have introduced the notation $0, X$ and $Y$ to describe the states in which there are no mRNA molecules ($M_X = M_Y = 0$), one mRNA molecule of type $X$ ($M_X = 1, M_Y = 0$) and one mRNA molecule of type $Y$ respectively ($M_X = 0, M_Y = 1$). This is in-line with the notation in the main manuscript.

The time evolution of the protein concentrations between the random switching events between these three states is then taken to be deterministic. Mathematically this corresponds to expanding the discrete operators $L^{(a,b)}$ in powers of $K^{-1}$, and keeping only the lowest-order advection terms. This generates so-called Liouville operators, and leads to

\[
\frac{\partial}{\partial t} \begin{bmatrix} p_0 \\ p_X \\ p_Y \end{bmatrix} = \left( L^d_1 + L^s_1 \right) \begin{bmatrix} p_0 \\ p_X \\ p_Y \end{bmatrix},
\]

where $L^d_1$ and $L^s_1$ are the forward operators driving the deterministic flow and the random switching between states, respectively. They are given by

\[
L^d_1 := \begin{bmatrix}
(L^d_1)_{11} & 0 & 0 \\
0 & (L^d_1)_{22} & 0 \\
0 & 0 & (L^d_1)_{33}
\end{bmatrix},
\]

\[
\begin{bmatrix}
-L(Kx) - H(Ky) & \gamma & \gamma \\
H(Ky) & -\gamma & 0 \\
H(Kx) & 0 & -\gamma
\end{bmatrix},
\]

and

\[
(L^d_1)_{11} := \gamma_0 \partial_x (x) + \gamma_0 \partial_y (y),
\]

\[
(L^d_1)_{22} := \partial_x (-\gamma b + \gamma_0 x) + \gamma_0 \partial_y (y),
\]

\[
(L^d_1)_{33} := \gamma_0 \partial_x (x) + \partial_y (-\gamma b + \gamma_0 y).
\]
Nature of the approximation
In deriving Eq. (22) we have made several assumptions and approximations:

(i) First, we have assumed that $\gamma/H \gg 1$, where $H$ stands for the maximum value $H(N_X)$ and $H(N_Y)$ can attain. We recall that $H(N) = K \left[ r_0 + \frac{r}{1 + (r/N)(K/N)} \right]$. The function $h(x) = r_0 + r/(1 + x^n)$ does not involve $K$ or $\gamma$, and its maximum value is $r_0 + r$. In dimensionless units, the assumption $\gamma/H \gg 1$ is thus fulfilled if $\gamma \gg (r_0 + r)K$.

(ii) We have replaced the discrete operators $L^{(a,b)}$ by deterministic Liouville operators, i.e., we neglected demographic stochasticity of the protein degradation. The purpose of this is to isolate the contribution of the bursting noise, originating from the random switching of the mRNA state ($0, X$ and $Y$). Making the deterministic approximation for the protein concentrations is formally valid only in the limit of very large protein populations, $K \gg 1$ ($K$ sets the scale of the numbers of protein molecules).

In summary we assume $\gamma \gg (r_0 + r)K$ and $K \gg 1$. We expect our approximations to be accurate when both of these are fulfilled, in particular the typical value of $\gamma$ above which our theory can be expected to be accurate will depend on the choice of $K$, which in turn must be chosen large enough to justify the deterministic approximation of the protein dynamics.

The data in the main manuscript reveals that the mathematical approximation agrees well with simulations for $\gamma = 30$ and $K = 200$. In our simulations we use $r_0 \approx 0.007$ and $r = 0.06$.

B. Forward equation in the limit $\gamma \to \infty$

We start from

\begin{align}
\partial_t p_0 = & [\gamma_0 \partial_x x + \gamma_0 \partial_y y - H(Kx) - H(Ky)] p_0 + \gamma p_X + \gamma p_Y, \\
\partial_t p_X = & [\partial_x (\gamma_0 x - \gamma b) + \gamma_0 \partial_y - \gamma] p_X + H(Ky) p_0, \\
\partial_t p_Y = & [\partial_y (\gamma_0 y - \gamma b) + \gamma_0 \partial_x - \gamma] p_Y + H(Kx) p_0.
\end{align}

(25a)

(25b)

(25c)
Applying the operator \[ \partial_y \left( \gamma_0 y - b \right) + \gamma_0 \partial_x - 1 \] \[ \partial_x \left( \gamma_0 x - b \right) + \gamma_0 \partial_y - 1 \] to both sides of equation (25a) results in
\[
\partial_t \left[ \partial_y \left( \frac{\gamma_0}{\gamma} y - b \right) + \gamma_0 \partial_x - 1 \right] \left[ \partial_x \left( \frac{\gamma_0}{\gamma} x - b \right) + \gamma_0 \partial_y - 1 \right] p_0
\]
\[ = \left[ \partial_y \left( \frac{\gamma_0}{\gamma} y - b \right) + \gamma_0 \partial_x - 1 \right] \left[ \partial_x \left( \gamma_0 y - b \right) + \gamma_0 \partial_y - 1 \right] \left[ \partial_t p_X - H(K_y) p_0 \right] + \left[ \partial_x \left( \frac{\gamma_0}{\gamma} x - b \right) + \gamma_0 \partial_y - 1 \right] \left[ \partial_t p_Y - H(K_x) p_0 \right]
\[ + \left[ \partial_y \left( \frac{\gamma_0}{\gamma} y - b \right) + \gamma_0 \partial_x - 1 \right] \left[ \partial_x \left( \frac{\gamma_0}{\gamma} x - b \right) + \gamma_0 \partial_y - 1 \right] \left[ \gamma_0 \partial_x x + \gamma_0 \partial_y y - H(K_x) - H(K_y) \right] p_0.
\] (26)

We note that this equation is not closed in \( p_0 \).

Next, we take the \( \gamma \to \infty \) limit, keeping in mind that \( H \) and \( \gamma_0 \) are finite. The system then almost-surely stays in the 0-state, and consequently \( p_X, p_Y \to 0 \). Equation (26) then reduces to
\[
\partial_t (-b \partial_y - 1) (-b \partial_x - 1) p_0 = (b \partial_y + 1) [H(K_y) p_0] + (b \partial_x + 1) [H(K_x) p_0].
\] (27)

The inverse operator of \( 1 + b \partial_z \) is
\[
(1 + b \partial_z)^{-1} f(z) = \int \frac{e^{-\frac{z+z'}{b}}}{b} f(z') dz,
\] (28)
and so equation (27) turns into the ‘forward equation’ presented in the main text:
\[
\partial_t p_0 = \partial_x (\gamma_0 x p_0) + \partial_y (\gamma_0 y p_0) - [H(K_x) + H(K_y)] p_0
\[ + H(K_y) \int_0^x \frac{1}{b} e^{-\frac{z+z'}{b}} p_0 (x', y, t) dx' + H(K_x) \int_0^y \frac{1}{b} e^{-\frac{z+z'}{b}} p_0 (x, y', t) dy'.
\] (29)

C. Equations for the mean first switching time

Here we illustrate the detail derivation to the adjoint equation in the main text. We focus on initial conditions \( y > x \) and our goal is to calculate the mean time it takes the dynamics to reach states with \( x = y \). We write \( T_Z(x, y) \) for the time it takes the dynamics to reach a state in which \( x = y \) if started from initial condition \( x, y \), and in mRNA state \( Z \in \{0, X, Y\} \). The \( T_Z(x, y) \) then satisfy the following adjoint equation [1]
\[
\begin{bmatrix}
\Gamma \\
\end{bmatrix}
= \begin{bmatrix}
\Theta_0(x, y) \\
\Theta_X(x, y) \\
\Theta_Y(x, y)
\end{bmatrix},
\] (30)

where \( L_d \) and \( L_s \) are the adjoint operators of \( L_d^† \) and \( L_s^† \). They are given by
\[
L_d := \begin{bmatrix}
(L_d)_{11} & 0 & 0 \\
0 & (L_d)_{22} & 0 \\
0 & 0 & (L_d)_{33}
\end{bmatrix}
\] and \( L_s = \begin{bmatrix}
-H(K_x) - H(K_y) & H(K_y) & H(K_x) \\
\gamma & -\gamma & 0 \\
\gamma & 0 & -\gamma
\end{bmatrix},
\] (31)
with

\[(L_d)_{11} = -\gamma_0 x \partial_x - \gamma_0 y \partial_y, \quad (32a)\]

\[(L_d)_{22} = (\gamma_b - \gamma_0 x) \partial_x - \gamma_0 y \partial_y, \quad (32b)\]

\[(L_d)_{33} = -\gamma_0 x \partial_x + (\gamma_b - \gamma_0 y) \partial_y. \quad (32c)\]

In the infinitely fast degrading mRNA limit, \(\gamma \to \infty\), equations (30) can be seen to converge to

\[\begin{pmatrix}
1 \\
0 
\end{pmatrix} = \begin{bmatrix}
-\gamma_0 x \partial_x - \gamma_0 y \partial_y - H(Kx) - H(Ky) & H(Ky) & H(Kx) \\
1 & b\partial_x - 1 & 0 \\
1 & 0 & b\partial_y - 1
\end{bmatrix}\cdot \begin{bmatrix}
T_0(x, y) \\
T_X(x, y) \\
T_Y(x, y)
\end{bmatrix}, \quad (33)\]

The boundary conditions for the mean first exit times are determined by \(T_Z(x_b, y_b) = 0\), for all locations \((x_b, y_b) \in \partial\Omega\) at which the deterministic flow driven by \(L_d^\dagger\) flows out of the domain \(\Omega\) in state \(Z\). Next, we specify a bounded domain \(\Omega_C := \{(x, y) : 0 < x < y, y < C\}\). The boundary conditions of equations (33) are then \(T_X(z, z) = 0\) and \(T_Y(z, C) = 0\) \(\forall z < C\). We now use these boundary conditions, and integrate the second and the third components of the expression in equation (33). Subsequently we send \(C \to \infty\) and arrive at

\[-1 = \begin{bmatrix}
-\gamma_0 x \partial_x - \gamma_0 y \partial_y - H(Kx) - H(Ky)
\end{bmatrix} T_0(x, y)
+ H(Ky) \int_x^y \frac{e^{-\frac{x'-x}{b}}}{b} T_0(x', y) \, dx' + H(Kx) \int_y^\infty \frac{e^{-\frac{y'-y}{b}}}{b} T_0(x, y') \, dx'. \quad (34)\]

\[\text{V. WKB ANALYSIS}\]

A. WKB ansatz

In order to find the quasi-stationary distribution of the PDMP model and of the diffusion approximation of the GB and CB models, one uses the ansatz

\[p_{\text{stat}} = \exp \left\{ -\frac{1}{\epsilon} \left[ S_0(x, y) + \mathcal{O}\left(\frac{B}{K}\right) \right] \right\} \]

where \(\epsilon \propto K^{-1}\) is the magnitude of the intrinsic noise in the protein dynamics. For the purposes of the WKB analysis the noise is assumed to be weak, i.e., \(\epsilon \ll 1\).

B. DA of the GB model

In the context of the diffusion approximation of the GB model we use \(\epsilon = B/K\). To leading order \((\mathcal{O}(K^0))\) one finds a Hamilton–Jacobi equation of the form

\[0 = \frac{1}{2} (\nabla S_0)^T \mathbf{D} (\nabla S_0) + \mathbf{v}^T \cdot \nabla S_0. \quad (36)\]
The vector \( \mathbf{v} \) denotes the deterministic flow
\[
\mathbf{v}(x, y) := \begin{bmatrix} \frac{B}{K} H(y) - \gamma_0 x \\ \frac{B}{K} H(x) - \gamma_0 y \end{bmatrix},
\]
and the (scaled) diffusion matrix \( \mathbf{D} \) is given by
\[
\mathbf{D}(x, y) := \begin{bmatrix} D_{11}(x, y) & 0 \\ 0 & D_{22}(x, y) \end{bmatrix},
\]
with entries
\[
D_{11}(x, y) = \frac{2B + 1}{K} H(Ky) + \frac{1}{B} \gamma_0 x, \quad D_{22}(x, y) = \frac{2B + 1}{K} H(Kx) + \frac{1}{B} \gamma_0 y.
\]

C. DA of the CB model

As before we use \( \epsilon = B/K \). A similar leading-order calculation delivers the Hamilton--Jacobi equation, which is again of the form described in equation (36). The only differences are minor modifications in the diffusion matrix, which now has entries
\[
D_{11}(x, y) = \frac{B}{K} H(Ky) + \frac{1}{B} \gamma_0 x, \quad D_{22}(x, y) = \frac{B}{K} H(Kx) + \frac{1}{B} \gamma_0 y.
\]

D. DA of the NB model

It is now convenient to use \( \epsilon = 1/K \). Again one finds a Hamilton-Jacobi equation of the form as above. The diffusion matrix now has entries
\[
D_{11}(x, y) = \frac{B}{K} H(Ky) + \gamma_0 x, \quad D_{22}(x, y) = \frac{B}{K} H(Kx) + \gamma_0 y.
\]

E. PDMP model

For the PDMP model, similar calculations deliver the Hamilton--Jacobi equation
\[
0 = \left[ \gamma_0 x - \frac{B}{K} H(Ky) \right] \partial_x S_0 + \left[ \gamma_0 y - \frac{B}{K} H(Kx) \right] \partial_y S_0 \\
+ \left[ \gamma_0 x + \gamma_0 y - \frac{B}{K} H(Kx) - \frac{B}{K} H(Ky) \right] (\partial_x S_0) (\partial_y S_0) \\
+ \gamma_0 x (\partial_x S_0)^2 + \gamma_0 y (\partial_y S_0)^2 + \gamma_0 x (\partial_x S_0)^2 (\partial_y S_0) + \gamma_0 y (\partial_x S_0) (\partial_y S_0)^2.
\]
VI. NUMERICAL METHODS

Sample paths of the individual-based processes (FM, CB, NB, and GB) are generated by the standard kinetic Monte Carlo algorithm [2, 3] implemented in C++. The PDMP process is simulated using the algorithm proposed by Bokes et al [4]. Simulations of the diffusion approximations are performed using the standard Euler–Maruyama algorithm with a constant time step $\delta t = 10^{-4}$. In all cases $10^6$ sample paths are simulated for a sufficiently long time to measure stationary distributions. For the mean first switching times, we sample 105 initial states on a lattice on the domain $0 \leq N_X(0) < N_Y(0) \leq 700$. For each initial state, we simulate $10^4$ sample paths, each until they cross the boundary $N_X = N_Y$ to measure the mean first switching times.

The geometric minimum action method proposed by Heymann and Vanden–Eijnden [5] is implemented using MATLAB R2010a, and used to find the rate function $S_0$ of the WKB method. For each model, we sample at least 150 end points and solve for the least-action paths, discretized into 257 equidistant points, connecting one of the fixed points and the end point. The final landscapes are generated by linear interpolation of the rate functions so obtained.

The finite-difference scheme to solve the adjoint equation was implemented in MATLAB R2010a, discretizing the domain $0 \leq x, y \leq C = 2000$ into $150 \times 150$ grid points. The adjoint equation is then transformed to a set of $22500$ linear equations, which is solved using a built-in numerical solver in MATLAB R2010a.
VII. SAMPLE PATHS OF THE DIFFERENT MODELS

A. Full model

FIG. 1. One sample path of the full model (FM). Left panel: short time scale. Right panel: the protein expressions switches at a longer time scale driven by intrinsic noise.
B. Geometrically distributed burst model (GB)

FIG. 2. One sample path of the model with geometrically distributed burst size (GB). Left panel: short time scale. Right panel: long time scale.

C. Constant burst model (CB)

FIG. 3. One sample path of the model with constant bursts (CB). Left panel: short time scale. Right panel: long time scale.
D. No-burst model (NB)

FIG. 4. One sample path of the model without bursts. Left panel: short time scale. Right panel: long time scale. In 1000 cell cycles, we observe no switching event in this sample path.

E. PDMP model

FIG. 5. One sample path of the PDMP. Left panel: short time scale. Right panel: long time scale.
F. Diffusion approximation of the GB model

FIG. 6. One sample path of the diffusion approximation of the GB model. Left panel: short time scale. Right panel: long time scale.

G. Diffusion approximation of the CB model

FIG. 7. One sample path of the diffusion approximation of the CB model. Left panel: short time scale. Right panel: long time scale.
H. Diffusion approximation of the NB model

FIG. 8. One sample path of the diffusion approximation of the single-stage model without bursts. Left panel: short time scale. Right panel: long time scale. Similar to the NB model, no switching event occurs in 1000 cell cycles in this sample path.
FIG. 9. Stationary distribution measured in simulations. All axes show $0 \leq N_X, N_Y \leq 700$ on a linear scale. Insets show the distribution as viewed from the point $N_X = N_Y = 700$ facing towards the origin. (a) Full model; (b) PDMP; (c) GB model; (d) Diffusion approximation (DA) of GB; (e) CB model; (f) DA of CB; (g) NB model; (a) DA of NB. The same colour scale is used in all panels, except for panels e and f.
IX. COMPARISON OF THE MEAN FIRST SWITCHING TIMES

FIG. 10. Mean first switching times. All graphs show $0 \leq N_X, N_Y \leq 700$ on linear scales. (a) Full model; (b) PDMP; (c) GB model; (d) Diffusion approximation (DA) of GB; (e) CB model; (f) Numerical solution of the adjoint equation of the PDMP. Data are plotted on the same colour scale in all panels to allow comparison.
FIG. 11. Results from the WKB analysis. Panels a, c, e, and g show the WKB rate functions $S_0(N_X, N_Y)$, and panels b, d, f, and h the corresponding approximation for the stationary probability distribution $N \exp[-S_0(N_X, N_Y) / \epsilon]$ where $N$ is the normalisation factor. All panels show $0 \leq N_X, N_Y \leq 700$ on a linear scale. The insets show the stationary distributions viewed from $(N_X, N_Y) = (700, 700)$. (a, b) PDMP; (c, d) DA of GB; (e, f) DA of CB; (g, h) DA of NB.
[1] van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*. Amsterdam: Elsevier Science B.V. (2007).

[2] Schwartz, R. *Biological modeling and simulation: a survey of practical models, algorithms, and numerical methods*. Cambridge, MA: MIT Press (2008).

[3] Gillespie, D. T. Exact stochastic simulation of coupled chemical-reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

[4] Bokes, P., King, J., Wood, A. A. & Loose, M. Transcriptional bursting diversifies the behaviour of a toggle switch: Hybrid simulation of stochastic gene expression. *Bull. Math. Biol.* **75**, 351–371 (2013).

[5] Heymann, M. & Vanden-Eijnden, E. The geometric minimum action method: A least action principle on the space of curves. *Comm. Pure Appl. Math.* **61**, 1052–1117 (2008).