Dynamics of a stochastic spatially extended system predicted by comparing deterministic and stochastic attractors of the corresponding birth-death process

Pawel J. Zuk[1,2], Marek Kochańczyk[1], Joanna Jaruszewicz[1], Witold Bednorz[3], and Tomasz Lipniacki[1,4]

[1] Institute of Fundamental Technological Research, Warsaw 02-106, Poland
[2] Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, Hoza 69, 00-681 Warsaw, Poland
[3] University of Warsaw, Faculty of Mathematics, Informatics and Mechanics, Warsaw 02-097, Poland
[4] Rice University, Department of Statistics, Houston, TX 77-005, USA

E-mail: tlipnia@ippt.gov.pl

Abstract. Living cells may be considered as biochemical reactors of multiple steady states. Transitions between these states are enabled by noise, or, in spatially extended systems, may occur due to the traveling wave propagation. We analyze a one-dimensional bistable stochastic birth-death process by means of potential and temperature fields. The potential is defined by the deterministic limit of the process, while the temperature field is governed by noise. The stable steady state in which the potential has its global minimum defines the *global deterministic attractor*. For the stochastic system, in the low noise limit, the stationary probability distribution becomes unimodal, concentrated in one of two stable steady states, defined in this study as the *global stochastic attractor*. Interestingly, these two attractors may be located in different steady states. This observation suggests that the asymptotic behavior of spatially-extended stochastic systems depends on the substrate diffusivity and size of the reactor. We confirmed this hypothesis within kinetic Monte Carlo simulations of a bistable reaction-diffusion model on the hexagonal lattice. In particular, we found that although the kinase-phosphatase system remains inactive in a small domain, the activatory traveling wave may propagate when a larger domain is considered.

Keywords: cell fate, bistability, stochastic processes, global attractor, thermodynamic interpretation, kinetic Monte Carlo

PACS numbers: 87.10.Mn, 87.10.Rt, 87.15.A-, 87.15.La, 87.16.dj, 87.18.Cf, 87.18.Mp, 87.18.Tt

"This is an author-created, un-copyedited version of an article accepted for publication in *Physical Biology*. IOP Publishing Ltd is not responsible for any errors or omissions in this version of the manuscript or any version derived from it. The definitive publisher-authenticated version is available online at http://dx.doi.org/10.1088/1478-3975/9/5/055002."

1. Introduction

Bistability and stochasticity are the key concepts in molecular biology. Bistable regulatory elements are capable of introducing heterogeneity in cell population and may allow cells in a multicellular organism to specialize and specify their fates [1, 2, 3, 4, 5]. Decisions between cell death, survival, proliferation or senescence are associated with bistability [6, 7, 8]. Ogasawara and Kawato showed that a bistable system of brain-specific protein kinase Mζ can play role in the long-term storage of memory [9]. Transitions between stable steady states occur due to the stochastic switching [10, 11], or, in spatially extended systems, may follow the traveling wave propagation [12].

When the magnitude of noise is relatively large, the stochastic transitions between attracting states are relatively frequent and the stationary probability distribution (SPD) associated with the stochastic process has a characteristic bimodal shape. The maxima of the SPD are approximately determined by the macroscopic steady states (although they do not necessarily exactly overlap). Song *et al* proposed a stochastic bifurcation concept describing the appearance of a new mode in the SPD, accompanying the appearance of a new stable steady state in the bifurcation diagram of the deterministic approximation of the process [13]. As the magnitude of noise decreases, communication between the two attractors ceases. Consequently, the characteristic time in which the probability distribution (PD) converges to the SPD elongates. Kinetics near each attractor can be modeled as Gaussian process [14], while jumping between the attractors can be modeled using the two-lumped-state Markov chain with transition propensities calculated from the full Markov process. The relative stability of two or more steady states depends on the system volume [15]. However, since the characteristic time spent in each attractor basin grows exponentially with the volume of the system, when the volume diverges to infinity, the SPD becomes (generically) unimodal, concentrated in the vicinity of the "most stable steady state" or the "global stochastic attractor" (GSA) [16, 17]. In this limit, bistability is manifested by a rapid transition from one unimodal SPD to the other unimodal SPD in response to the change of the bifurcation parameter.

For spatially extended bistable systems, considered in the deterministic approximation, the most stable steady state can be determined by the direction of the traveling wave propagation. Intuitively, the traveling wave propagates in such direction that the whole domain converges to the most stable steady state – which will be referred to as a global deterministic attractor (GDA). Interestingly, for a given system of reactions and parameters, the GDA and GSA may be different. This implies that SPD of the stochastic but perfectly mixed system will concentrate in one steady state, while the corresponding spatially extended deterministic system may converge to the other steady state due to traveling wave propagation. In the case when GDA and GSA do not collocalize, one could expect that for the corresponding bistable, stochastic but spatially distributed system, the size of the compartment and substrate diffusivity control the relative stability of steady states. We verify this hypothesis simulating a

stochastic kinase–phosphatase reaction–diffusion model on a hexagonal lattice.

The paper is organized as follows. In Preliminaries, we consider a general one-dimensional birth–death (B–D) Markov process, introduce the GDA (subsection 2.1) and GSA (subsection 2.2), and show that these two attractors may not overlap.

In Results, we propose the thermodynamic interpretation using the concepts of the potential and temperature fields associated with the B–D process (subsection 3.1). The concept of the temperature for B–D processes was introduced by Ross and colleagues [16, 18, 19]. Later, Bialek [20] and Lu et al. [7] proposed that the temperature field is not uniform, but proportional to the sum of birth and death rates. We will introduce another temperature definition, which converges to that proposed by Bialek [20] in steady states of the system.

Next, we consider two biological examples. First (subsection 3.2), we investigate analytically a simplified one-dimensional B–D process of kinase auto-activation in an open compartment. We will demonstrate that the temperature field and thus the stationary probability distribution is controlled by fluxes of the active kinase to and out of the compartment, even in the case when these fluxes are equal and do not influence the deterministic mass rate equation. In a relatively broad range of parameters the value of in- and out-flux (which can be associated with substrate diffusivity) determines the global attractor of the system. This observation suggests that for stochastic spatially distributed systems the size of the compartment and substrate diffusivity control the relative stability of steady states.

We verify this hypothesis in subsection 3.3 introducing a more realistic kinase–phosphatase model with diffusion on a hexagonal lattice. In this latter model, by performing kinetic Monte Carlo (KMC) simulations, we demonstrate that, despite the fact that the kinase molecules in a small isolated compartment remain mostly inactive, in a larger compartment the activatory traveling wave may propagate leading to persistent activation of the system.

In Conclusions we review the main results in the biological context. The manuscript is supplemented by three Appendices. In Appendix A.1 we review and discuss the Bialek's derivation of temperature. In Appendix A.2 we demonstrate that when the temperature profile is nonuniform (i.e. in the generic case), an arbitrary steady state may become the GSA, provided that the temperature in the vicinity of this state is sufficiently low. In Appendix A.3 we show that the parameter range for which the bimodal SPD distribution is observed decreases to zero as the system volume diverges to the infinity.

2. Preliminaries

2.1. Global deterministic attractor

We will consider a one-dimensional B–D process:

$$K \to K+1 : V\lambda\left(\tfrac{K}{V}\right); \qquad K \to K-1 : V\mu\left(\tfrac{K}{V}\right), \tag{1}$$

where K is the number of substrate molecules, V is the volume of the reactor and $\lambda\left(\frac{K}{V}\right)$ and $\mu\left(\frac{K}{V}\right)$ denote birth and death intensities, nonnegative for $K > 0$, with $\mu(0) = 0$. The assumption that birth and death intensities depend on the concentration $x = \frac{K}{V}$, rather than on the number of molecules K, allows for comparing reactors of different volumes. In the deterministic limit, when the volume of the reactor tends to infinity, the B–D process follows the law of mass action,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \lambda(x) - \mu(x) =: W(x) =: -\frac{\mathrm{d}U(x)}{\mathrm{d}x}. \tag{2}$$

Function $U(x)$ can be interpreted as a potential. The stable stationary states of (2) are in the minima of $U(x)$. We assume that the potential has two minima in x_1 and x_3, separated by a maximum in x_2. The trajectories of (2) converge to x_1 or x_3 depending on the initial state of the system. Equations (1–2) may satisfactorily describe the state of the reactor in the infinite diffusion limit, in which the concentration x is constant over the reactor. In the case of finite diffusion, (2) should be replaced by the reaction–diffusion equation for the substrate density $x(z,t)$ (here for illustration purposes we assume that x depends on one spatial coordinate, z),

$$\frac{\partial x}{\partial t} = D\frac{\partial^2 x}{\partial z^2} - \frac{\mathrm{d}U(x)}{\mathrm{d}x}, \tag{3}$$

where D is the diffusion coefficient. The above equation yields the traveling wave solutions, $x = x(z - vt) = x(\zeta)$ connecting steady states x_1 and x_3. For this solution $x(z) \to x_1$ for $z \to -\infty$ and $x(z) \to x_3$ for $z \to \infty$. The propagation velocity v may be given in the implicit form (see [21]) as

$$v = \frac{U(x_1) - U(x_3)}{\int_{-\infty}^{\infty}(\frac{\mathrm{d}x}{\mathrm{d}\zeta})^2\mathrm{d}\zeta}. \tag{4}$$

The sign of velocity v, given by $\mathrm{Sgn}(U(x_1)-U(x_3))$, assures that the region of "lower energy" expands, i.e. $x(z,t) \to x_U$ as $t \to \infty$, where $U(x_U)$ is the global minimum. Only in the non-generic case, when $U(x_1) - U(x_3) = 0$, the system has standing wave solutions in which states "x_1" and "x_3" coexist. Intuitively, for a "random" initial condition in the sufficiently large reactor, in some areas of reactor $x(z)$ will be in the basin of attraction of x_U. In such a case the traveling wave or waves will be formed and in the whole reactor the substrate density $x(z,t)$ will converge to x_U. Therefore, for the deterministic, spatially extended system (3), the state x_U, where the potential achieves the global minimum, may be considered a global attractor. We will refer to it as a GDA (global deterministic attractor).

2.2. Global stochastic attractor

Let $p(K,t)$ denote the probability that the number of substrate molecules equals K at time t in the process (1). Probability $p_K(t)$ obeys the evolution (or Master) equation

$$\frac{\partial p(K,t)}{\partial t} = p(K-1,t)\,V\lambda\left(\tfrac{K-1}{V}\right) + p(K+1,t)\,V\mu\left(\tfrac{K+1}{V}\right) - V\left(\lambda\left(\tfrac{K}{V}\right) + \mu\left(\tfrac{K}{V}\right)\right)p(K,t). \tag{5}$$

In the steady state $p(K,t) = p_K = \text{const}$, the net probability flux between neighboring states K and $K+1$ equals zero, i.e.

$$p_K\,\lambda\left(\tfrac{K}{V}\right) - p_{K+1}\,\mu\left(\tfrac{K+1}{V}\right) = 0. \tag{6}$$

Therefore, under the assumption that $\prod_{i=0}^{\infty}\left(\lambda\left(\tfrac{i}{V}\right)/\mu\left(\tfrac{i+1}{V}\right)\right) < \infty$, p_K satisfies

$$p_K = p_0 \prod_{i=0}^{K-1} \frac{\lambda\left(\tfrac{i}{V}\right)}{\mu\left(\tfrac{i+1}{V}\right)}, \tag{7}$$

where p_0 is such that $\sum p_K = 1$. From (7) we have

$$\log p_K = \log p_0 + \sum_{i=0}^{K-1} \log \frac{\lambda\left(\tfrac{i}{V}\right)}{\mu\left(\tfrac{i+1}{V}\right)}. \tag{8}$$

For large V the sum can be replaced by the integral, that leads to [16, 17, 22, 23]

$$p_K \simeq p(x) = p_0 \exp(-V\Phi(x)), \tag{9}$$

where

$$\Phi(x) := -\int_0^x \log \frac{\lambda(y)}{\mu(y)}\mathrm{d}y. \tag{10}$$

The local minima and maxima of $\Phi(x)$ correspond to stable and unstable macroscopic steady states of (2), respectively. The Laplace method implies that for $V \to \infty$, $p(x)$ converges to the Dirac delta, $\delta(x_\Phi)$, where $\Phi(x_\Phi)$ is the proper global minimum of $\Phi(x)$, provided that such minimum exists [24]. Only in the non-generic case, when $\Phi(x)$ has two or more equal minima, the stationary probability distribution (SPD) in the zero noise limit ($V \to \infty$) is distributed between these minima. In this way we showed that, generically, when $V \to \infty$, the SPD concentrates in the macroscopic steady state of the mass rate equation, in which $\Phi(x)$ achieves the global minimum. The steady state x_Φ, in which $\Phi(x)$ achieves the global minimum, may be thus considered the GSA (global stochastic attractor). The local maxima and minima of $\Phi(x)$ correspond to the maxima and minima of $U(x)$. However, the global minimum of $\Phi(x)$ may not correspond to the global minimum of $U(x)$, therefore the GSA may not be collocalized with the GDA. According to our knowledge this observation was first explicitly made in the context of the Schlögl model [25] by Nicolis [23].

Later, in the context of stochastic gene expression, it was observed that the character of noise, defined by the adiabaticity parameter [26, 27] or coarse graining [28] influences the SPD. Recently, for a gene expression model with the additive or multiplicative noise Frigola *et al* showed that the SPD is concentrated in the global minimum of the stochastic rather than the deterministic potential [29]. Frigola *et al* defined the stochastic potential based on the Fokker–Planck approximation and demonstrated that type of noise dictates in which of steady states such potential has the global minimum.

3. Results

3.1. Thermodynamic interpretation

In order to provide the thermodynamic interpretation of the above observation, we consider the problem of particles diffusing in the potential and temperature fields $U(x)$ and $T(x)$. Such particles drift with flow $J_F = -M\rho(x,t)(\mathrm{d}U/\mathrm{d}x)$ where M is the mobility and $\rho(x,t)$ is the local concentration of particles. In thermal equilibrium, the drift flow is balanced by the diffusion flow $J_D = -\frac{\partial}{\partial x}[\Gamma(x)\rho(x,t)]$, where $\Gamma(x) = Mk_\mathrm{B}T(x)$ is the diffusion and k_B is the Boltzman constant. In the steady state $J_D + J_F = 0$, and one obtains

$$\rho\frac{\mathrm{d}U}{\mathrm{d}x} = -k_\mathrm{B}\frac{\mathrm{d}}{\mathrm{d}x}[\rho(x)T(x)], \quad \rho(x) = \rho_0\frac{T_0}{T(x)}\exp\left[-\int\frac{\mathrm{d}U/\mathrm{d}x}{k_\mathrm{B}T(x)}\mathrm{d}x\right]. \tag{11}$$

In the uniform temperature field $T(x) = T_0$, the last expression simplifies to $\rho(x) = \rho_0\exp\left[-U(x)/k_\mathrm{B}T_0\right]$. In such a case, $\rho(x)$ converges to $\delta(x_U)$ as $T_0 \to 0$. However, the above is not true when the temperature field $T = \tau_0 T(x)$ is not uniform. In case when $\tau_0 \to 0$, $\rho(x)$ converges to $\delta(x_\Phi)$, where x_Φ is point of the global minimum of $\int\frac{\mathrm{d}U(x)/\mathrm{d}x}{k_\mathrm{B}T(x)}\mathrm{d}x$ which may be different than x_U. When temperature gradients are not very large, the prefactor $T_0/T(x)$ can be replaced by a constant which leads to

$$\rho(x) = \rho_0\exp\left[-\int\frac{\mathrm{d}U/\mathrm{d}x}{k_\mathrm{B}T(x)}\mathrm{d}x\right]. \tag{12}$$

This approximation is equivalent to neglecting the component of diffusion flow J_D called spurious flow $J_\mathrm{spurious} = -\rho(x,t)\mathrm{d}\Gamma(x)/\mathrm{d}x = -k_\mathrm{B}\rho(x,t)M\mathrm{d}T(x)/\mathrm{d}x$; i.e. flow of particles induced by the diffusion or temperature gradient, known also as Soret effect, see [30] and [31]. By comparing $p(x)$ given by (9) and (10) with the particle density $\rho(x)$ given by (12), we obtain $T(x)$ as

$$T(x) = \frac{1}{k_\mathrm{B}V}\frac{\lambda(x)-\mu(x)}{\log\left(\lambda(x)/\mu(x)\right)}. \tag{13}$$

That is, in the low-noise limit (large V) the SPD of the B–D process is proportional to the density of particles diffusing in the potential $U(x) = \int(\mu(x)-\lambda(x))\mathrm{d}x$ and the temperature field $T(x)$ given by (13). Let us notice that $T(x)$ defined by (13) vanishes when either $\lambda(x)$ or $\mu(x)$ are zero. Points in which $T(x) = 0$ are singular, i.e. they can be passed only in one direction. If such points exist, they bound the absorbing regions.

Another, methodologically rigorous, way of deriving temperature, based on the Fokker–Planck approximation, was proposed by Bialek [20] and then followed by Lu *et al* [7]. It led to $\widetilde{T}(x) = (\lambda(x)+\mu(x))/(2k_\mathrm{B}V)$. We discuss Bialek's approach in Appendix A.1; see also [32] for a recent detailed review. Finally, we mention the recent study by Feng and Wang [33] who introduced the other concept of the effective temperature for gene networks basing on the correlation and the response functions. This is an entirely different approach, leading for example to negative temperatures for a self-repressing gene.

As shown in the Appendix A.2, any of the $U(x)$ minima may become a global attractor depending on the temperature profile $T(x)$. This implies that two stochastic processes converging to the same mass rate equation in the $V \to \infty$ limit may have two different GSAs. In the following section, we illustrate this unintuitive observation and its consequences using a simple bistable kinase auto-activation model, a version of the Schlögl model [25], classified as one of the simplest bistable systems [34]. The auto-activation is characteristic for Src, Syk, and Tec family kinases important in the immune cell signaling [35, 36].

3.2. Simplified kinase auto-activation model

In the model we will assume that the total concentration of kinase molecules remains constant and equals 1, thus their number is equal to the volume of the compartment V. Kinases can be in either active or inactive state, and the number of active kinases will be denoted by K (see figure 1). Inactive kinases, number of which is $V - K$, can be activated by active kinases with rate proportional to the square of the active kinase concentration $x = K/V$, or by other kinase species with some small constant rate c_1. The second order nonlinearity arises either when the active unit of the kinase is a dimer or when double phosphorylation is required to activate the kinase [37, 38]. In turn, active kinases are inactivated with constant rate d_1. Under above assumptions the number of active kinases K follows the B–D process with rates:

$$K \to K+1 : \left(c_1 + c_2 \left(\tfrac{K}{V}\right)^2\right)\left(1 - \left(\tfrac{K}{V}\right)\right) V; \qquad K \to K-1 : d_1 \left(\tfrac{K}{V}\right) V. \tag{14}$$

Let us notice that since the total number of kinases remains constant, the number of active kinases is limited by V, which is reflected by the fact that the birth intensity is zero for $K = V$.

Next, we will assume that active kinases can translocate to and out of the compartment, with fluxes f_1 and f_2, respectively. Such situation arises when the considered compartment is a subvolume of a larger reactor. As a consequence, we obtain the following birth and death intensities, where concentration $x = K/V$ is used instead of the number of molecules K,

$$\lambda(x) = (c_1 + c_2 x^2)(1 - x) + f_1; \qquad \mu(x) = d_1 x + f_2. \tag{15}$$

For the sake of simplicity and illustration purposes we will assume that the in-flux and out-flux are equal, $f_1 = f_2 = f$. Such simplifying assumption assures that the (deterministic) mass rate equation for x,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \lambda(x) - \mu(x) = (c_1 + c_2 x^2)(1 - x) - d_1 x =: W(x) \tag{16}$$

does not depend on f. We will see, however, that f controls the SPD and determines the most stable stochastic attractor. We focus on the bistable case when $W(x)$ has 3 real roots $0 < x_1 < x_2 < x_3 < 1$. In the further analysis we will use the roots x_1, x_2, x_3

as parameters describing polynomial $W(x)$. The original coefficients c_1, c_2, d_1 may be recovered from the roots using Vieta's formulas by the following relations:

$$c_1 = \frac{d_1 x_1 x_2 x_3}{(x_1+x_2)(x_1-1)(x_2-1)}, \qquad c_2 = \frac{d_1}{(x_1+x_2)(x_1-1)(x_2-1)}. \quad (17)$$

Let us notice that since c_1 and c_2 are proportional to d_1, the last coefficient determines only the time scale of the process, $\tau = 1/d_1$. Since the quotient of coefficients at the third and at the second order term in $W(x)$ equals -1, roots of $W(x)$ satisfy $x_1+x_2+x_3 = 1$, and thus the parameter space may be reduced to the two-dimensional domain Δ, defined as follows:

$$\Delta = \left\{ (x_1, x_2) \in \mathbb{R}^2 \;\middle|\; x_1 > 0, \;\; x_2 > 0, \;\; x_1 < x_2, \;\; x_2 < \frac{1}{2} - \frac{x_1}{2} \right\} \quad (18)$$

(see figure 2). Domain Δ splits into two subdomains, Δ_1 and Δ_3, such that for $(x_1, x_2) \in \Delta_1$, the SPD of the process (14-15) in the $V \to \infty$ limit converges to $\delta(x_1)$ and for $(x_1, x_2) \in \Delta_3$, the SPD converges to $\delta(x_3)$. The curves separating Δ_1 and Δ_3 depend on f, $x_1(x_2; f)$. They are showed for $f = 0$, $f = 0.1$, $f = 1$ and $f = \infty$ in figure 2, and are given analytically based on (10) in the implicit form by

$$\Phi(x_3) - \Phi(x_1) = -\int_{x_1}^{x_3} \log \frac{\lambda(y)}{\mu(y)} \mathrm{d}y = 0. \quad (19)$$

In the case with zero flux, $f = 0$, the temperature profile is not uniform and for the symmetric potential $U(x)$, the SPD is not symmetric, but is concentrated in the colder attraction basin of point x_1 (figure 3, left column). The temperature effect may be balanced by the asymmetry of the potential (figure 3, right column). For the point $B = (x_1, x_2(x_1))$ on the separatrix $f = 0$, in the $V \to \infty$ limit SPD $p(x)$ converges to $\alpha\delta(x_1) + \beta\delta(x_3)$. However, this cannot be explicitly numerically demonstrated, because even a tiny deviation from the separatrix causes the SPD to be redistributed either to x_1 or x_3 as $V \to \infty$ (figure 3, right column, two bottom panels). As shown in the Appendix A.3, in figure 9, the bimodal SPD expected for bistable systems may be observed only if the magnitude of noise is sufficiently large. Interestingly, as shown in the lowest panel of figure 3, SPD for large temperatures concentrates mostly in the x_1 basin, and then in the $T \to 0$ limit it converges to $\delta(x_3)$. This confirms the observation made by Vellela and Qian that the relative stability of steady states depends on the system volume [15].

When the system communicates with environment, i.e. when $f > 0$, the temperature profile $T(x)$ is modified; figure 4. For $f = 0$ (figure 4, left column) the temperature is much lower in the left attraction basin and thus the SPD concentrates in this basin. For $f = 1$ (figure 4, right column), the temperature profile is flatter and the SPD concentrates in the global potential minimum x_3 as $V \to \infty$. For $f \to \infty$, the temperature profile $T(x)$ becomes uniform, and SPD converges to $\delta(x_U)$ as $V \to \infty$, where x_U is the steady state in which potential $U(x)$ has the global minimum. It should be noted that for larger f, larger V is required to reach the unimodal SPD, figure 4. In the $f \to \infty$ limit, $\Phi(x)$ becomes proportional to $U(x)$ and SPD $f(x)$ becomes symmetric for the symmetric potential $U(x)$, i.e. when $U(x_1) = U(x_3)$, which implies $x_2 = (x_1 + x_3)/2 = 1/3$.

At this point, let us recall the reaction–diffusion system (3). As follows from (4), for $U(x_1) = U(x_3)$ the traveling wave velocity is $v = 0$. In our specific example, (3) takes the form of the Nagumo equation [21],

$$\frac{\partial x}{\partial t} = D\frac{\partial^2 x}{\partial z^2} + (x_1 - x)(x_2 - x)(x_3 - x), \tag{20}$$

which yields the traveling wave solutions

$$x(z - vt) = x(\zeta) = \frac{x_3 + x_1 \exp\left[(2D)^{-1/2}(x_3 - x_1)\zeta\right]}{1 + \exp\left[(2D)^{-1/2}(x_3 - x_1)\zeta\right]} \tag{21}$$

with

$$v = (x_1 - 2x_2 + x_3)\sqrt{\frac{D}{2}}. \tag{22}$$

The last result implies that in the $f \to \infty$ limit the stochastic process is diffusion–driven and the associated temperature field becomes uniform. As f increases the parameter range for which GSA and GDA are localized in different steady states decreases and only for the $f \to \infty$ limit the two attractors must collocalize.

In summary, we found that there is a broad range of parameters (figure 2, between separatrices for $f = 0$ and $f = \infty$) in which GDA corresponds to the active state x_3, while GSA corresponds to the inactive state x_1. This is, there exists a broad parameter range in which:

- the discrete, perfectly mixed stochastic system, considered in the isolated reactor, will converge to the inactive state,
- while in its spatially extended deterministic counterpart the activatory traveling waves will propagate, leading to the activation of the system.

This suggests that in stochastic spatially distributed systems the global attractor is determined by the diffusion and size of the reactor. In the next section we will confirm this hypothesis considering a more realistic reaction–diffusion model on a hexagonal lattice.

Remark: The SPD may inadequately represent the behavior of bistable systems in the low noise limit. As noise decreases, the communication between steady states ceases, and the characteristic state-to-state transition time lengthens, and thus time in which the PD approaches the SPD may become longer than time in which external conditions defining potential may be considered constant. In the low noise limit the bistable system has two distinct time scales: the intermediate (in which the initially uniform PD becomes bimodal) and the asymptotic time scale in which the PD converges to the unimodal SPD (figure 5). At intermediate time scale the system behavior is close to deterministic. Depending on the nature of the biochemical process, each of the two time scales and associated "limiting" PDs can be important.

3.3. Kinase auto-activation reaction–diffusion model on a 2D hexagonal lattice

In the model we consider two molecular species, kinase $\mathcal{K}$ and phosphatase $\mathcal{P}$. We assume that the kinase molecules can be in one of three states, unphosphorylated $\mathcal{K}$, singly phosphorylated $\mathcal{K}_\mathrm{p}$ and doubly phosphorylated $\mathcal{K}_\mathrm{pp}$. Kinases may activate one another, and in turn are dephosphorylated by phosphatases with intensity d. The phosphorylation intensities c_1, c_2 and c_3, of, respectively, $\mathcal{K}$, $\mathcal{K}_\mathrm{p}$, and $\mathcal{K}_\mathrm{pp}$, increase with the kinase phosphorylation level. The following reactions are considered:
Phosphorylation by an unphosphorylated kinase:

$$\mathcal{K} + \mathcal{K} \ \rightarrow \mathcal{K} + \mathcal{K}_\mathrm{p} : 2c_1, \tag{23a}$$

$$\mathcal{K} + \mathcal{K}_\mathrm{p} \rightarrow \mathcal{K} + \mathcal{K}_\mathrm{pp} : c_1. \tag{23b}$$

Phosphorylation by a singly phosphorylated kinase:

$$\mathcal{K}_\mathrm{p} + \mathcal{K} \ \rightarrow \mathcal{K}_\mathrm{p} + \mathcal{K}_\mathrm{p} : 2c_2, \tag{24a}$$

$$\mathcal{K}_\mathrm{p} + \mathcal{K}_\mathrm{p} \rightarrow \mathcal{K}_\mathrm{p} + \mathcal{K}_\mathrm{pp} : c_2. \tag{24b}$$

Phosphorylation by a doubly phosphorylated kinase:

$$\mathcal{K}_\mathrm{pp} + \mathcal{K} \ \rightarrow \mathcal{K}_\mathrm{pp} + \mathcal{K}_\mathrm{p} : 2c_3, \tag{25a}$$

$$\mathcal{K}_\mathrm{pp} + \mathcal{K}_\mathrm{p} \rightarrow \mathcal{K}_\mathrm{pp} + \mathcal{K}_\mathrm{pp} : c_3. \tag{25b}$$

Dephosphorylations:

$$\mathcal{P} + \mathcal{K}_\mathrm{p} \ \rightarrow \mathcal{P} + \mathcal{K} : d_1, \tag{26a}$$

$$\mathcal{P} + \mathcal{K}_\mathrm{pp} \rightarrow \mathcal{P} + \mathcal{K}_\mathrm{p} : 2d_1. \tag{26b}$$

By convention, the first molecule on both reaction sides is considered the enzyme, while the second represents the substrate. The factor 2 multiplying rates of reactions involving $\mathcal{K}$ and $\mathcal{K}_\mathrm{pp}$ as a substrate reflects the fact that an unphosphorylated kinase can be phosphorylated at any of its two residues, and similarly the doubly phosphorylated kinase can be dephosphorylated at any of its two residues.

The deterministic approximation of the system leads to three partial differential equations for concentrations of $\mathcal{K}$, $\mathcal{K}_\mathrm{p}$ and $\mathcal{K}_\mathrm{pp}$. We confine to the case in which the diffusion coefficient D is equal for all kinase forms regardless of their phosphorylation level. In such a case we may assume that the total surface concentration of kinase $C_\mathcal{K}$ remains constant and uniform over the surface. The phosphatase surface concentration will be denoted by $C_\mathcal{P}$. The fractional concentrations of $\mathcal{K}$, $\mathcal{K}_\mathrm{p}$ and $\mathcal{K}_\mathrm{pp}$ will be denoted by k, k_p and k_pp, thus by definition $k + k_\mathrm{p} + k_\mathrm{pp} = 1$.

$$\frac{\partial k}{\partial t} = D\frac{\partial^2 k}{\partial z^2} + d_1 C_\mathcal{P} k_\mathrm{p} - 2(c_1 k + c_2 k_\mathrm{p} + c_3 k_\mathrm{pp}) k C_\mathcal{K}, \tag{27a}$$

$$\begin{aligned}\frac{\partial k_\mathrm{p}}{\partial t} &= D\frac{\partial^2 k_\mathrm{p}}{\partial z^2} + 2(c_1 k + c_2 k_\mathrm{p} + c_3 k_\mathrm{pp}) k C_\mathcal{K} + 2d_1 C_\mathcal{P} k_\mathrm{pp} \\ &\quad - (c_1 k + c_2 k_\mathrm{p} + c_3 k_\mathrm{pp}) k_\mathrm{p} C_\mathcal{K} - d_1 C_\mathcal{P} k_\mathrm{p},\end{aligned} \tag{27b}$$

$$\frac{\partial k_\mathrm{pp}}{\partial t} = D\frac{\partial^2 k_\mathrm{pp}}{\partial z^2} + (c_1 k + c_2 k_\mathrm{p} + c_3 k_\mathrm{pp}) k_\mathrm{p} C_\mathcal{K} - 2d_1 C_\mathcal{P} k_\mathrm{pp}. \tag{27c}$$

Here for the sake of simplicity we assume the dependence on only one spatial coordinate z. The above system exhibits bistability in a broad range of parameters. Let us notice, that in the deterministic approximation the system dynamics depends on only four parameters, values of which will be set constant for the rest of this considerations,

$$C_1 = c_1 C_{\mathcal{K}} = 0.02; \quad C_2 = c_2 C_{\mathcal{K}} = 0.15; \quad C_3 = c_3 C_{\mathcal{K}} = 4; \quad D_1 = d_1 C_{\mathcal{P}} = 1. \tag{28}$$

The values of parameters are chosen so that the system (27*a*–27*c*) has three steady state solutions, two stable (active with low level of unphosphorylated kinase, $k = 0.15$, and inactive with high level of unphosphorylated kinase, $k = 0.94$) and one unstable:

- inactive: $(k = 0.94,\ k_{\mathrm{p}} = 0.06,\ k_{\mathrm{pp}} = 0.00)$,
- unstable: $(k = 0.50,\ k_{\mathrm{p}} = 0.41,\ k_{\mathrm{pp}} = 0.08)$,
- active: $(k = 0.15,\ k_{\mathrm{p}} = 0.48,\ k_{\mathrm{pp}} = 0.37)$.

For these parameters we found that:

(i) The SPD obtained in Gillespie algorithm simulations for the stochastic perfectly mixed system defined by reactions (23*a*-26*b*) concentrates in the decreasing vicinity of the inactive state as the number of kinases grows, figure 6(*a*–*c*). Values of parameters employed for Gillespie algorithm simulations are $c_1 = C_1/N_{\mathcal{K}}$, $c_2 = C_2/N_{\mathcal{K}}$, $c_3 = C_3/N_{\mathcal{K}}$, $d_1 = D_1/N_{\mathcal{P}}$, where $N_{\mathcal{K}}$ and $N_{\mathcal{P}}$ are the numbers of kinase and phosphatase molecules. That is, the reaction rates are scaled by the number of molecules, which is equivalent to the assumption that the concentrations are independent of the number of substrate molecules. In spatially homogeneous systems the kinases dephosphorylation rate D_1 is a product of phosphatase activity and the number of phosphatases $D_1 = C_{\mathcal{P}} d_1 = 1$. In reactions (23*a*-26*b*) the number of phosphatases remains unchanged, and is unimportant for the system dynamics.

(ii) The system (27*a*-27*c*) with parameters C_1, C_2, C_3, D_1 (28) describes an activatory traveling wave solution i.e. waves that propagate from the active to inactive state, figure 6(*d*).

The above finding implies that for the system defined by reactions (23*a*-26*b*) and parameters (28) the stochastic and deterministic global attractors diverge, and are defined, respectively, by the inactive and active states. Therefore, although a potential may not be defined for the system (27*a*-27*c*), the system exhibits an analogous behavior as the simpler system (16) analyzed in the previous section.

Finally, to confirm the hypothesis stated in the previous section, we performed KMC simulations on hexagonal lattices, 20×20 (with periodic boundary conditions, i.e. toroidal topology) and 20×1000 (with periodic-reflecting boundary conditions, i.e. cylindrical topology). In these two simulations we assumed that fraction of lattice sites $\chi_{\mathcal{K}} = 0.4$ is occupied by kinases, while fraction of lattice sites $\chi_{\mathcal{P}} = 0.1$ is occupied by phosphatases. In KMC simulations molecules can interact only when in contact. All molecules move with the same motility $M = 5000$, that is, the propensity that a given molecule jumps to a neighboring empty site is $M/6$. The relation between diffusion

and motility on 2D lattice is $D = (1 - \chi_{\mathcal{K}} - \chi_{\mathcal{P}})s^2 M/4$, where $(1 - \chi_{\mathcal{K}} - \chi_{\mathcal{P}})$ is the fraction of empty sites on the lattice, and s is the distance between the centers of adjacent hexagonal cells. Thus, in units in which $s = 1$, we obtain for $\chi_{\mathcal{K}} = 0.4$ and $\chi_{\mathcal{P}} = 0.1$, $D = M/8$. For KMC simulations we set $\hat{c}_1 = C_1/n_{\mathcal{K}}$, $\hat{c}_2 = C_2/n_{\mathcal{K}}$, $\hat{c}_3 = C_3/n_{\mathcal{K}}$, $\hat{d}_1 = D_1/n_{\mathcal{P}}$, where $n_{\mathcal{K}} = 6\chi_{\mathcal{K}}$ (and $n_{\mathcal{P}} = 6\chi_{\mathcal{P}}$) are expected values of the number of kinase (and phosphatase) molecules that are in the immediate vicinity of any molecule. The use of coefficients $c_1 = C_1/N_{\mathcal{K}}$, $c_2 = C_2/N_{\mathcal{K}}$, $c_3 = C_3/N_{\mathcal{K}}$, $d_1 = D_1/N_{\mathcal{P}}$ in Gillespie simulations, and coefficients $\hat{c}_1 = C_1/n_{\mathcal{K}}$, $\hat{c}_2 = C_2/n_{\mathcal{K}}$, $\hat{c}_3 = C_3/n_{\mathcal{K}}$, $\hat{d}_1 = D_1/n_{\mathcal{P}}$ in KMC simulations, provides that these two approaches converge in the infinite diffusion limit [39].

In simulations performed on a small toroidal lattice, 20×20, we found, as already expected from Gillespie algorithm simulations shown in figure 6(b), that the system remains in the inactive state for most of the time (figure 7). In these two simulations (shown in figure 6(b) and figure 7) the number of kinase molecules was the same (equal 160). The assumed large motility $M = 5000$ implies that the relatively small reactor 20×20 can be considered to be almost perfectly mixed, and thus the obtained SPD is nearly identical to the one obtained in Gillespie algorithm simulations for 160 kinases, figure 7(b). In both approaches more than 0.99 of mass is concentrated in the vicinity of the inactive state.

In contrast to simulations on the small toroidal lattice, 20×20, in simulations performed on the 20×1000 lattice we observed propagation of an activatory traveling wave, followed by persistent activity of the system (figure 8). The system remains persistently active because, even when any region is inactivated due to stochastic fluctuations, it is quickly sealed by the activatory traveling wave. On the other hand, the motility $M = 5000$ is about 3 orders of magnitude too small to render the 20×1000 reactor mixed.

The behavior observed in KMC simulations remains in agreement with the expectations coming from the deterministic approximation of the system (figure 6). Activation of the systems on the 20×1000 lattice can be interpreted on the basis of the analysis we made in the previous section. The 20×1000 lattice can be viewed as an array of small 20×20 compartments. When these compartments are isolated, as in simulations shown in figure 7, the in-flux and out-flux of active kinase f equals zero, and the system remains in the vicinity of GSA – in this case the inactive state (figure 4). However, when compartments are connected, the flux f makes the temperature profile flatter and the systems converges to GDA, which corresponds to the active state (figure 4 and figure 8). In summary, we demonstrated that the spatially-extended system converges to GSA when diffusion is fast enough to make the reactor perfectly mixed, but when the same system is considered in a larger reactor the traveling waves may form and can drive the system to GDA.

4. Conclusions

Bi- and multistable systems play a prominent role in signal processing. The attractors of molecular dynamical systems control the cell evolution and fate. Transitions between attractors can be due to stochastic switching, or may result from the traveling wave propagation. The first mode is characteristic for perfectly mixed systems for which noise provides the unique possibility of selection of the most stable steady state — termed here the global stochastic attractor. The second transition mode dominates in spatially–extended systems characterized by relatively slow diffusion. These systems may achieve the global deterministic attractor due to the traveling wave propagation in which the more stable steady state expands. Interestingly, as we discussed in this study, in bistable systems these two attractors (i.e. GSA and GDA) can be different.

We studied analytically the one-dimensional birth–death process for which potential and temperature fields may be constructed. In such a case, the GDA is defined by the global minimum of the potential, while the GSA can be in any of the potential minima for a particular temperature profile. As an example, we consider the bistable kinase auto-activation model in the open compartment, such that the active kinase can flow in and out of the compartment. Even in the case when the in-flux and out-flux are equal and do not influence the deterministic mass rate equation and the related potential, they control the temperature profile, and as a consequence – the GSA. When the in-flux and out-flux increase (which can be interpreted as the increase of kinase diffusivity), the temperature grows and its profile becomes uniform, and in the infinite flux limit the GSA is determined by the global minimum of the potential, i.e. it collocates with the GDA. This finding allows us for putting forward the hypothesis that in stochastic spatially-extended biochemical reactors the relative stability of attractors is governed by the substrate diffusivity and size of the compartment. When for a given diffusion, the reactor is small enough to be considered as perfectly mixed, the system of interacting molecules converges to the GSA. In a much larger reactor, in which traveling waves can be formed, the same system converges to GDA.

We confirmed this hypothesis by performing KMC simulations for the kinase-phosphatase system, in the parameter range in which the GSA is located in the inactive steady state, while the GDA is located in the active steady state (figure 6). Accordingly, we found that the kinase–phosphatase system simulated in a small toroidal compartment remains mostly inactive state (figure 7), while the same system simulated in a long cylindrical compartment is activated due to the traveling wave propagation and remains active (figure 8).

In summary, we found that relative stability of attractors in bi- (or multi-) stable systems is controlled by the diffusivity of substrates and size of the compartment. These two parameters can be controlled and modified in cell evolution. Buffers, extracellular ligands can control diffusivity of cytoplasmic or membrane proteins. The effective size of the reactor can be modified by plasma membrane deformation, formation of lipids rafts and other barriers. It is thus tempting to speculate that cells may employ these

mechanism for their activation or inactivation. Such mode of activation may be relevant in the immune cell signaling, which requires the aggregation of membrane receptors.

Appendix

Appendix A.1. Temperature derivation based on the Fokker–Planck approximation.

To derive temperature Bialek [20] used the Fokker–Planck (or diffusion) approximation of the Master equation (5), see e.g. [30],

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(\frac{\mathrm{d}U(x)}{\mathrm{d}x}p(x,t)\right) + \frac{\partial^2}{\partial x^2}\left(\frac{\lambda(x)+\mu(x)}{2V}p(x,t)\right). \tag{A.1}$$

In the stationary case the above equation can be solved explicitly:

$$p(x) = p_0 \frac{2V}{\lambda(x)+\mu(x)} \exp\left[-\int_0^x \frac{\mathrm{d}U(x')}{\mathrm{d}x'}\frac{2V}{\lambda(x')+\mu(x')}\mathrm{d}x'\right]. \tag{A.2}$$

It is known, however, that the Fokker–Planck approximation obtained by truncating the Kramers-Moyal expansion after the second term is not a satisfactory approximation for bistable systems [40]. As a result, the SPD $p(x)$ given by (A.2) differs from the "exact" $p(x)$ obtained directly from the stationary Master equation in the $V \to \infty$ limit, (9) and (10). In particular, equation (A.2) may lead to the incorrect determination of the GSA. In equation (A.1) $\Gamma(x) = k_\mathrm{B}T(x) = \frac{\lambda(x)+\mu(x)}{2V}$ is the diffusion coefficient. Thus the effective temperature is (Bialek writes it simply as $T_\mathrm{eff} = (\lambda(x)+\mu(x))/2$)

$$\widetilde{T}(x) = \frac{\lambda(x)+\mu(x)}{2k_\mathrm{B}V}. \tag{A.3}$$

Our and Bialek's expressions for temperature converge in the limit $|\lambda(x)-\mu(x)|/(\lambda(x)+\mu(x)) \to 0$, i.e. in the steady states of (2). The main difference between Bialek's and our expression is that $T(x)$ defined by (13) vanishes when either $\lambda(x)$ or $\mu(x)$ are zero, that is when the direction of motion is deterministic, while $\widetilde{T}(x)$ vanishes when both birth and death rates are zero, i.e. there is no motion at all. Points in which $T(x) = 0$ bound absorbing regions for the original Markov process, however these points state no barrier when the process is considered in the diffusion approximation. This intuitively explains the observation by Hänggi *et al* [40] that the Fokker–Planck approximation overestimates transition rates between steady states. For that reason we prefer our definition. Finally, let us recall that Ross *et al* and Chu *et al* [16, 19] considered $\Phi(x)$ given in (10) as a "stochastic potential", which leads to a constant temperature equal $1/(k_\mathrm{B}V)$. In a sense our approach is equivalent to that of Ross *et al*, i.e. our temperature (13) combined with deterministic potential $U(x)$ give the same SPD as Ross' temperature and stochastic potential $\Phi(x)$.

Appendix A.2. Each of the minima of the potential $U(x)$ may become a global attractor for a particular noise characteristics.

Two stochastic processes, characterized by B–D rates $\lambda(x)$, $\mu(x)$ and $\lambda^*(x) = \lambda(x) + f(x)$, $\mu^*(x) = \mu(x) + f(x)$ have the same deterministic mass rate equation, but are

associated with different temperature fields and thus have different SPDs. In particular, let us consider the case in which (2) has three steady states $x_1 < x_2 < x_3$, such that x_1 and x_3 are stable and x_2 is unstable. This implies that $W(x) = \lambda(x) - \mu(x) < 0$ on (x_1, x_2) and $W(x) > 0$ on (x_2, x_3). Without loss of generality we may assume that for the process characterized by transitions $\lambda(x)$ and $\mu(x)$ the function $\Phi(x)$ has the global minimum in x_1, and thus the SPD of this process converges to $\delta(x_1)$. We will demonstrate that there exists a function $f(x)$ such that the SPD of the B–D process with transitions $\lambda(x) + f(x)$ and $\mu(x) + f(x)$ converges to $\delta(x_3)$. Function $f(x)$ must thus satisfy

$$\Phi^*(x_3) - \Phi^*(x_1) = -\int_{x_1}^{x_3} \log \frac{\lambda(x) + f(x)}{\mu(x) + f(x)} \mathrm{d}x < 0, \tag{A.4}$$

which holds when

$$\left| \int_{x_1}^{x_2} \log \frac{\lambda(x) + f(x)}{\mu(x) + f(x)} \mathrm{d}x \right| < \int_{x_2}^{x_3} \log \frac{\lambda(x) + f(x)}{\mu(x) + f(x)} \mathrm{d}x. \tag{A.5}$$

The last inequality holds when

(i) $f(x)$ is sufficiently large on (x_1, x_2) and equals zero elsewhere, that is equivalent to the increase of $T(x)$ in the attraction basin of x_1, or

(ii) $f(x)$ is negative on (x_2, x_3) and equals zero elsewhere, and the new death rate $\mu^*(x) < \lambda^*(x)$ is sufficiently small on (x_2, x_3). Such choice of $f(x)$ is equivalent to the decrease of $T(x)$ in the basin of x_3. In particular we can take $f(x)$ such that $\min \mu^*(x) = 0$ on (x_2, x_3). In such a case the system is no longer ergodic, and the domain $x > x_{\min}$ (where $\mu^*(x_{\min}) = 0$) is absorbing.

We thus demonstrated that in the bistable B–D process, the SPD in the zero noise limit (generically) converges to the Dirac delta in one of the two minima of the corresponding potential $U(x)$, but the choice of a particular minimum depends on the temperature field $T(x)$ of the process. The proof for a multistable system is analogous. In the field of evolutionary games it has been shown similarly that the long-run behavior of a population depends on its size and the mutation intensity [41].

Appendix A.3. The parameter range in which the bimodal SPD is observed decreases to zero as the system volume diverges to infinity.

The bimodal probability distributions are typically associated with bistability. This is true when the magnitude of noise is sufficiently large, i.e. when the volume of the system is sufficiently small. However, when the system volume grows, the bimodal distribution is replaced by the unimodal distribution concentrated around the most stable steady state or the GSA. As shown in figure 9 the parameter range in which bimodal SPD $p(x)$ associated with the process (14) is observed decreases to zero as $1/V$. By bimodal SPD in this context we understand $p(x)$ that satisfies

$$0.1 < \int_0^{x_2} p(x) < 0.9. \tag{A.6}$$

Acknowledgments

We would like to thank an anonymous referee for constructive comments which helped us to improve the paper. This research was funded by the following grants: Foundation for Polish Science grant TEAM/2009-3/6, Polish Ministry of Science and Higher Education grant N N501 132936, NSF/NIH grant No. R01-GM086885.

References

[1] Becskei A, Séraphin B and Serrano L 2001 Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion *EMBO J.* **20** 2528–35 (page 2).

[2] Ferrell Jr J E 2002 Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability *Curr. Opin. Cell. Biol.* **14** 140–8 (page 2).

[3] Ptashne M 2004 *A genetic switch: phage lambda revisited* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press) (page 2).

[4] Chatterjee A, Kaznessis Y N and Hu W S 2008 Tweaking biological switches through a better understanding of bistability behavior *Curr. Opin. Biotechnol.* **19** 475–81 (page 2).

[5] Sprinzak D, Lakhanpal A, LeBon L, Santat L A, Fontes M E, Anderson G A, Garcia-Ojalvo J and Elowitz M B 2010 Cis-interactions between Notch and Delta generate mutually exclusive signalling states *Nature* **465** 86–90 (page 2).

[6] Acar M, Becskei A and van Oudenaarden A 2005 Enhancement of cellular memory by reducing stochastic transitions *Nature* **435** 228–32 (page 2).

[7] Lu T, Hasty J and Wolynes P G 2006 Effective temperature in stochastic kinetics and gene networks *Biophys. J.* **91** 84–94 (pages 2, 3, 6).

[8] Zhang Q, Bhattacharya S, Kline D E, Crawford R B, Conolly R B, Thomas R S, Kaminski N E and Andersen M E 2010 Stochastic modeling of B lymphocyte terminal differentiation and its suppression by dioxin *BMC Syst. Biol.* **4** 40 (page 2).

[9] Ogasawara H and Kawato M 2010 The protein kinase Mζ network as a bistable switch to store neuronal memory *BMC Syst. Biol.* **4** 181 (page 2).

[10] Lipniacki T, Hat B, Faeder J R and Hlavacek W S 2008 Stochastic effects and bistability in T cell receptor signaling *J. Theor. Biol.* **254** 110–122 (page 2).

[11] Puszyński K, Hat B and Lipniacki T 2008 Oscillations and bistability in the stochastic model of p53 regulation *J. Theor. Biol.* **254** 452–65 (page 2).

[12] Hat B, Kazmierczak B and Lipniacki T 2011 B cell activation triggered by the formation of the small receptor cluster: a computational study *PLoS Comput. Biol.* **7** e1002197 (page 2).

[13] Song C, Phenix H, Abedi V, Scott M, Ingalls B P, Kærn M and Perkins T J 2010 Estimating the stochastic bifurcation structure of cellular networks *PLoS Comput. Biol.* **6** e1000699 (page 2).

[14] Keizer J 1978 Maxwell-type constructions for multiple nonequilibrium steady states *Proc. Natl Acad. Sci. USA* **75** 3023–6 (page 2).

[15] Vellela M and Qian H 2009 Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited *J. R. Soc. Interface* **6** 925–40 (pages 2, 8).

[16] Ross J, Hunt K L C and Hunt P M 1988 Thermodynamics far from equilibrium: Reactions with multiple stationary states *J. Chem. Phys.* **88** 2719–2729 (pages 2, 3, 5, 14).

[17] Ge H and Qian H 2009 Thermodynamic limit of a nonequilibrium steady state: Maxwell-type construction for a bistable biochemical system *Phys. Rev. Lett.* **103** 148103 (pages 2, 5).

[18] Ross J, Hunt K L C and Hunt P M 1992 Thermodynamic and stochastic theory for nonequilibrium systems with multiple reactive intermediates: The concept and role of excess work *J. Chem. Phys.* **96** 618–629 (page 3).

[19] Chu X, Ross J, Hunt P M and Hunt K L C 1993 Thermodynamic and stochastic theory of reaction-diffusion systems with multiple stationary states *J. Chem. Phys.* **99** 3444–3454 (pages 3, 14).
[20] Bialek W 2000 Stability and noise in biochemical switches *arXiv:cond-mat/0005235* (pages 3, 6, 14).
[21] Murray J D 2003 *Mathematical biology* (New York: Springer) (pages 4, 9).
[22] Matheson I, Walls DF and Gardiner CW 1975 Stochastic Models of First-Order Nonequilibrium Phase Transitions in Chemical Reactions *J. Stat. Phys.* **12** 21–34 (page 5).
[23] Nicolis G and Lefever R 1977 Comment on the kinetic potential and the Maxwell construction in non-equilibrium chemical phase transitions *Phys. Lett. A* **62** 469–471 (page 5).
[24] Barndorff-Nielsen O E and Cox D R 1989 *Asymptotic techniques for use in statistics* (London: Chapman and Hall) (page 5).
[25] Schlögl F 1971 On thermodynamics near a steady state *Z. Phys. A* **248** 446–458 (pages 5, 7).
[26] Hornos E M, Schultz D, Innocentini G C P, Wang J, Walczak A M, Onuchic J N and Wolynes PG 2005 Self-regulating gene: An exact solution *Phys. Rev. E* **72** 051907 (page 5).
[27] Feng H, Han B and Wang J 2011 Adiabatic and Non-Adiabatic Non-Equilibrium Stochastic Dynamics of Single Regulating Genes *J. Phys. Chem. B* **115** 1254–1261 (page 5).
[28] Morelli M J, Allen R J, Tanase-Nicola S and ten Wolde P R 2008 Eliminating fast reactions in stochastic simulations of biochemical networks: A bistable genetic switch *J. Chem. Phys.* **128** 045105 (page 5).
[29] Frigola D, Casanellas L, Sancho J M and Ibañes M 2012 Asymmetric Stochastic Switching Driven by Intrinsic Molecular Noise *PLoS ONE* **7** e31407 (page 5).
[30] Van Kampen N G 2007 *Stochastic Processes in Physics and Chemistry* (Elsevier Science & Technology Books) (pages 6, 14).
[31] Groot S R and Mazur P 1962 Non-Equilibrium Thermodynamics (Amsterdam: North-Holland Publishing Company) (page 6).
[32] Walczak A M, Mugler A and Wiggins C H 2011 *Analytic methods for modeling stochastic regulatory networks* [in] Methods in Molecular Biology Eds. M. Betterton and X. Liu (Springer Verlag). (page 6).
[33] Feng H D and Wang J 2011 Correlation function, response function and effective temperature of gene networks, *Chem. Phys. Lett.* **510** 267-272 (page 6).
[34] Wilhelm T 2009 The smallest chemical reaction system with bistability *BMC Syst. Biol.* **3** 90 (page 7).
[35] Roskoski Jr R 2005 Src kinase regulation by phosphorylation and dephosphorylation *Biochem. Biophys. Res. Commun.* **331** 1–14 (page 7).
[36] Bradshaw J M 2010 The Src, Syk, and Tec family kinases: distinct types of molecular switches *Cell. Signal.* **22** 1175–84 (page 7).
[37] Markevich N I, Hoek J B and Kholodenko B N 2004 Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades *J. Cell. Biol.* **164** 353–9 (page 7).
[38] Gilfillan A M and Rivera J 2009 The tyrosine kinase network regulating mast cell activation *Immunol. Rev.* **228** 149–69 (page 7).
[39] Stamatakis M and Vlachos D G 2011 Equivalence of on-lattice stochastic chemical kinetics with the well-mixed chemical master equation in the limit of fast diffusion *Comput. Chem. Eng.* **35** 2602–2610 (page 12).
[40] Hänggi P, Grabert H, Talker P and Thomas H 1984 Bistable systems: Master equation versus Fokker-Planck modeling *Phys. Rev. A* **29** 371–378 (page 14).
[41] Miekisz J 2005 Equilibrium selection in evolutionary games with random matching of players *J. Theor. Biol.* **232** 47–53 (page 15).

Figure captions

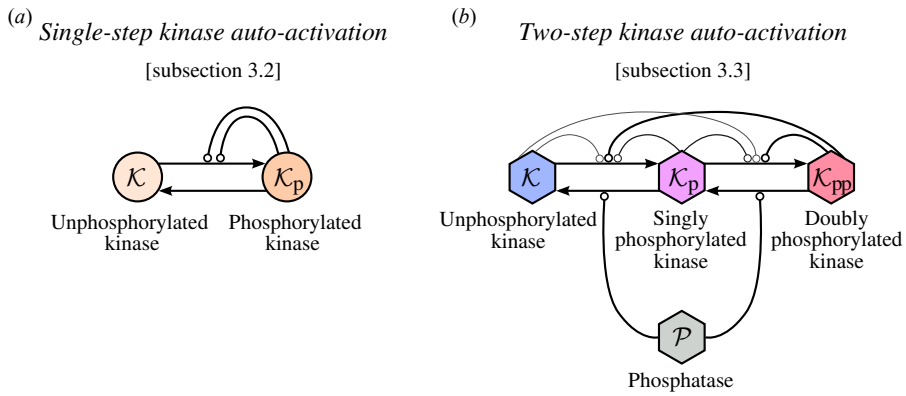

Figure 1. Two bistable reaction systems of kinase auto-phosphorylation. (*a*) A simple model with nonlinear auto-activation (circle-headed double arrow) defined by equations (14). (*b*) A multistate kinase model with explicit phosphatase activity described by equations (27*a*–27*c*). Phosphorylated kinases have higher catalytic activity (as reflected by the width of circle-headed arrows).

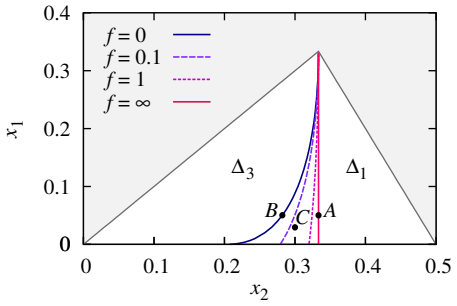

Figure 2. Subdomains Δ_1 and Δ_3 with their separation curves for four values of flux f.

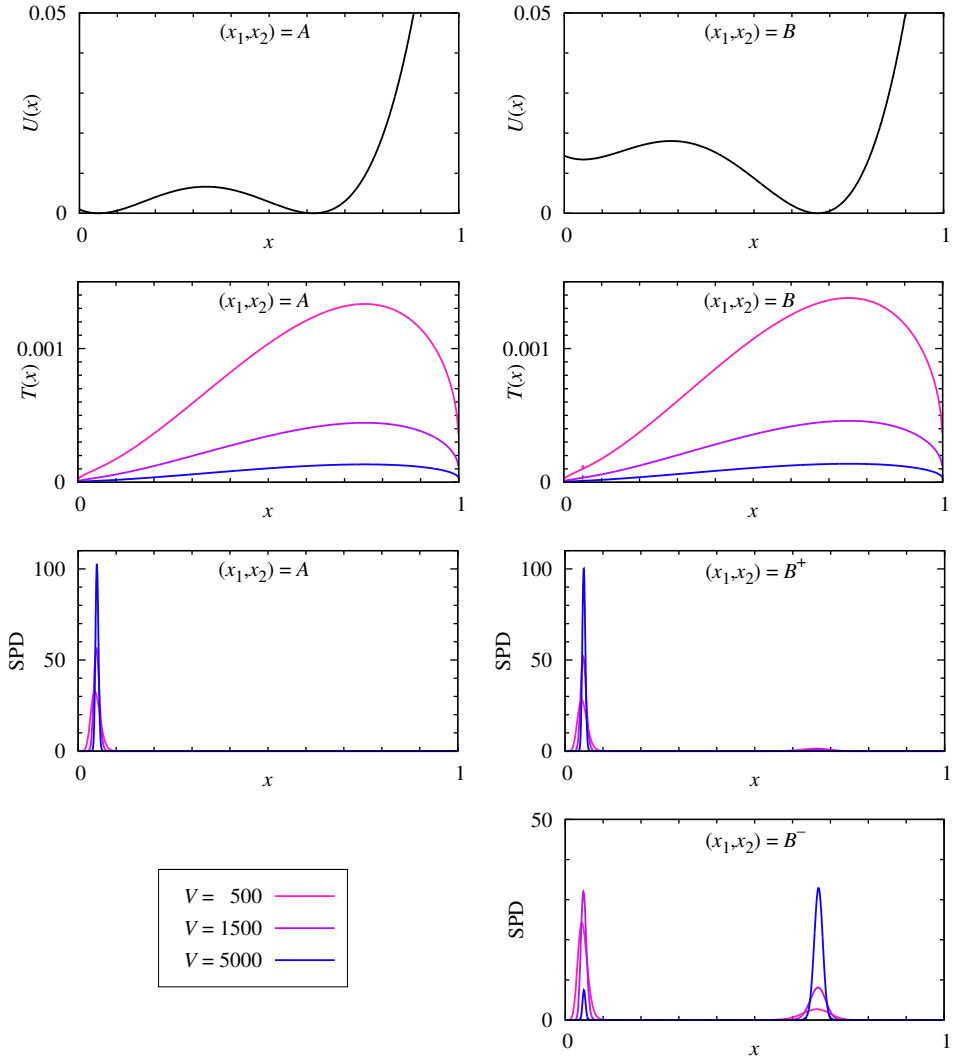

Figure 3. The case of zero flux, $f = 0$. Left column: $(x_1, x_2) = A = (0.05, \frac{1}{3})$ (see figure 2). The potential field $U(x)$ is symmetric with respect to the unstable steady state x_2. In this case, as the system volume V grows, the SPD concentrates in the colder attraction basin of steady state $x_1 = 0.05$. Right column: potential $U(x)$, temperature $T(x)$ and function $\Phi(x)$ for $(x_1, x_2) = B \approx (0.05, 0.2818)$ on the separatrix $f = 0$ shown in figure 2. The SPDs $p(x)$ for $B^- = (0.05, 0.2818 - 0.001)$ below the separatrix and $B^+ = (0.05, 0.2818 + 0.001)$ above the separatrix are shown in two bottom panels. As $V \to \infty$, $p(x; B^-)$ converges to $\delta(x_3)$ and $p(x; B^+)$ converges to $\delta(x_1)$.

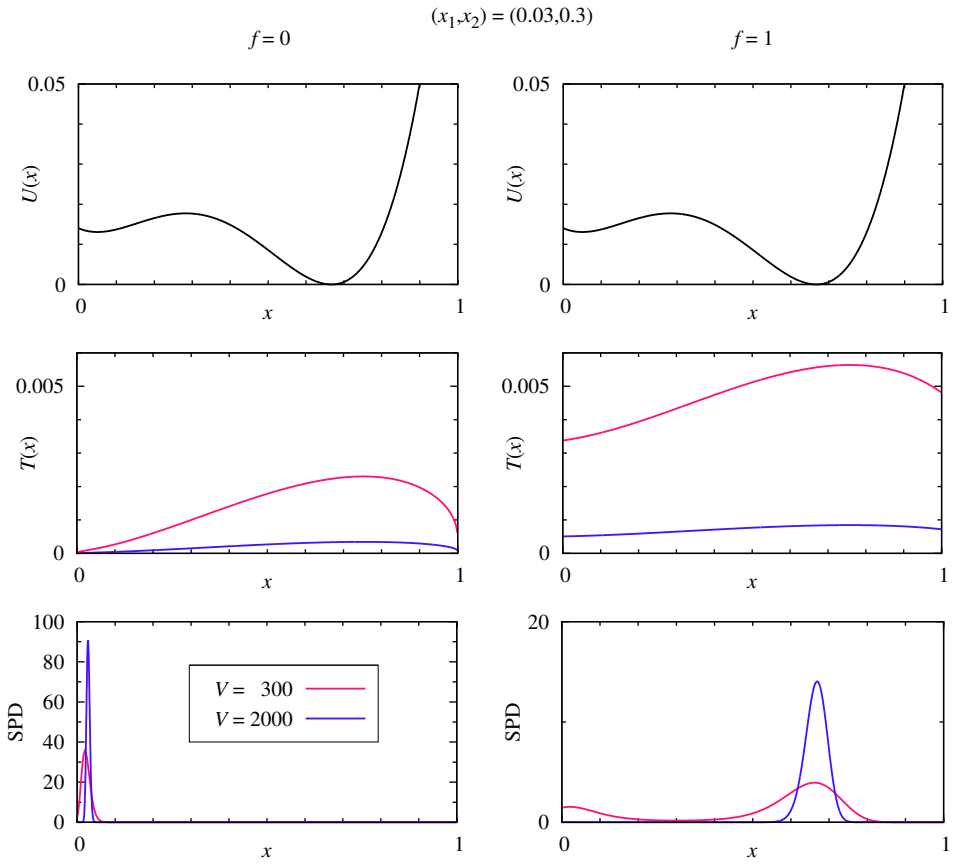

Figure 4. The flux controls temperature profile $T(x)$ and the SPD. Left column: $f = 0$, $(x_1, x_2) = C = (0.03, 0.3)$. The temperature is much lower in the left basin of attraction, and thus the SPD $p(x)$ concentrates in x_1 as $V \to \infty$. Right column: $f = 1$, $(x_1, x_2) = C$; the temperature profile is flatter than for $f = 0$ and the SPD $p(x)$ concentrates in the global potential minima at $x_3 = 0.67$. The potential $U(x)$ is the same for both columns.

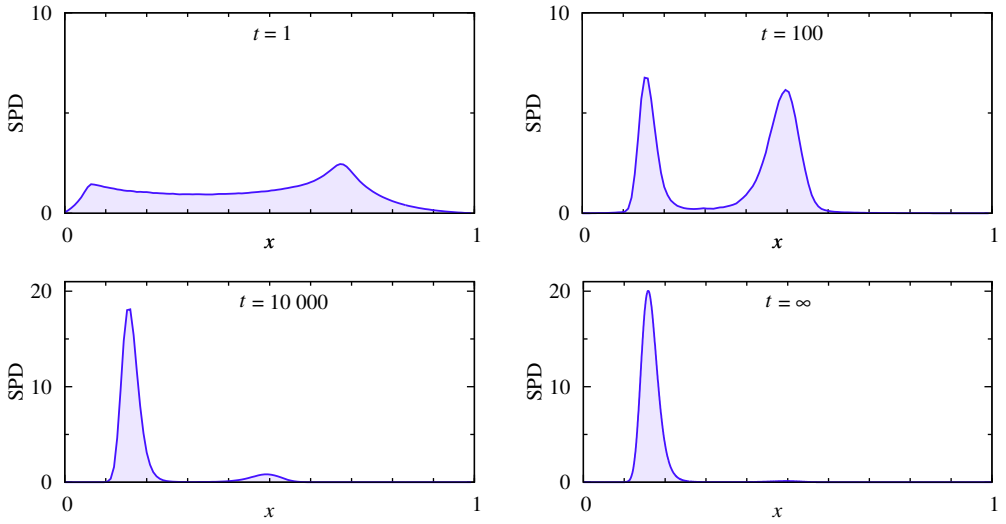

Figure 5. Evolution of PD based on Monte Carlo simulations for $(x_1, x_2) = (\frac{1}{6} - 0.005, \frac{1}{3} + 0.005)$ with $V = 2000$. For $(x_1, x_2) = (\frac{1}{6}, \frac{1}{3})$ and $x_3 = 1 - x_1 - x_2 = \frac{1}{2}$ the potential $U(x)$ is symmetric with respect to x_2. At $t = 0$, PD is uniform; at intermediate times it becomes bimodal, determined by the size of attraction basins; then at $t \to \infty$ it becomes unimodal, determined by the global attractor.

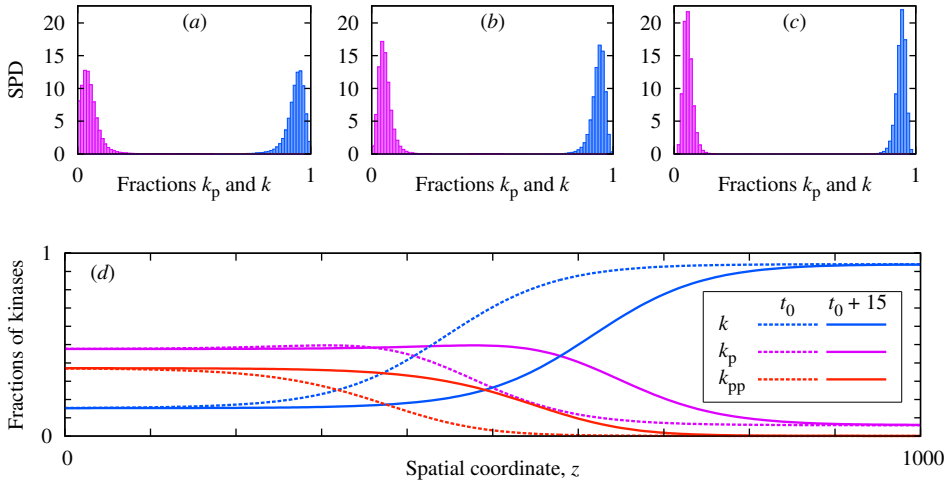

Figure 6. Comparison of behavior of stochastic perfectly mixed systems with their deterministic spatially-extended counterpart. (a), (b) and (c): SPDs estimated in Gillespie algorithm (assuming spatial homogeneity of the system) simulations of the stochastic system (23*a*-26*b*), containing 80, 160 or 320 interacting kinases, respectively. SPDs concentrate in the decreasing vicinity of the inactive state ($k = 0.94, k_p = 0.06, k_{pp} = 0.00$) as the number of kinases grows. (d) Profile of the activatory traveling wave, that propagates from the active to inactive state (the assumed diffusion coefficient $D = 625$ corresponds to the motility $M = 5000$). The SPD in Gillespie algorithm simulations was estimated from long trajectories containing more than 10 switches between the active and inactive state in case of 320 molecules, and much more switches for 160 and 80 molecules.

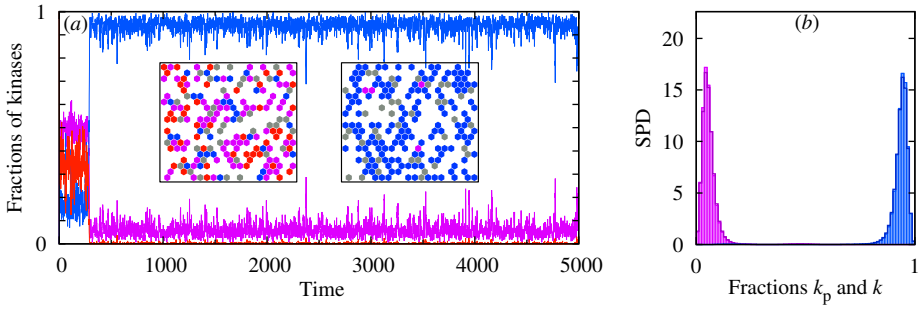

Figure 7. Kinetic Monte Carlo simulations of the system (23*a*-26*b*) on the hexagonal lattice of size 20×20, for motility equal $M = 5000$. A fraction of lattice sites $\chi_{\mathcal{K}} = 0.4$ is occupied by kinase molecules, while phosphatase molecules occupy $\chi_{\mathcal{P}} = 0.1$ of the lattice. (*a*) Fraction of unphosphorylated, singly phosphorylated and doubly phosphorylated kinases in time. Insets: Snapshots of the system in its active (left) and inactive (right) state. (*b*) SPDs obtained in Gillespie simulations (boxes) and KMC (thick lines). In case of KMC simulations, the distribution was estimated from multiple runs starting from the active and inactive state, and the mean first-passage time analysis.

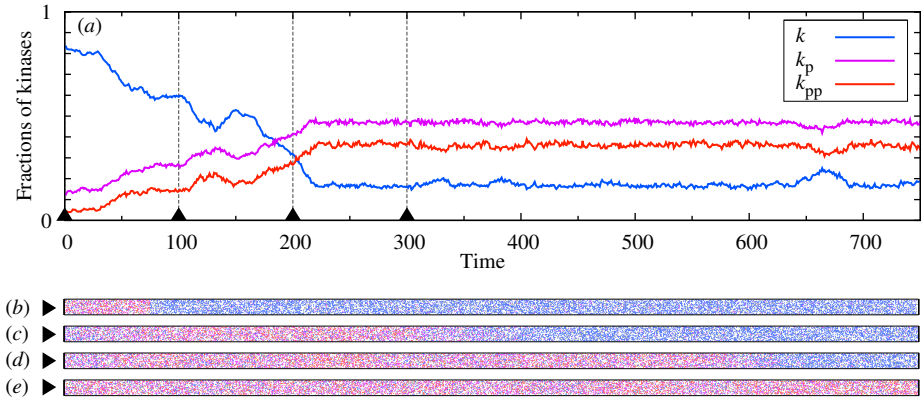

Figure 8. Kinetic Monte Carlo simulations of the system (23*a*-26*b*) on the hexagonal lattice of size 20×1000, showing the activatory traveling wave propagation. The same kinetic parameters, substrate fractions and motility M are assumed as in figure 7. (*a*) Fraction of unphosphorylated, singly phosphorylated and doubly phosphorylated kinases in time. (*b*), (*c*), (*d*), (*e*): four snapshots of the system in time points as marked in (*a*).

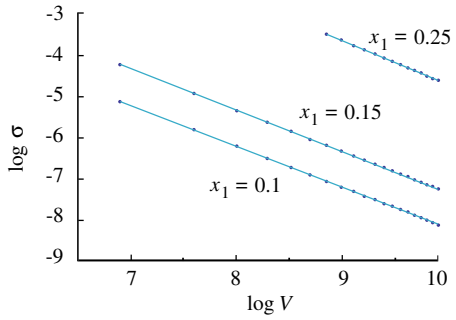

Figure 9. Width of the bistability zone σ in (x_1, x_2) plane as a function of V for three values of x_1. The width σ decreases as $1/V$. For a given system volume V, σ decreases as the distance between the two stable steady states x_1 and x_3 grows (i.e. when the value of x_1 decreases).