Early signatures of regime shifts in gene expression dynamics

Mainak Pal,* Amit Kumar Pal,† Sayantari Ghosh,‡ and Indrani Bose§
Department of Physics, Bose Institute, 93/1, Acharya Prafulla Chandra Road, Kolkata - 700009, India

Recently, a large number of studies have been carried out on the early signatures of sudden regime shifts in systems as diverse as ecosystems, financial markets, population biology and complex diseases. Signatures of regime shifts in gene expression dynamics are less systematically investigated. In this paper, we consider sudden regime shifts in the gene expression dynamics described by a fold-bifurcation model involving bistability and hysteresis. We consider two alternative models, Models 1 and 2, of competence development in the bacterial population *B.subtilis* and determine some early signatures of the regime shifts between competence and vegetative state. We use both deterministic and stochastic formalisms for the purpose of our study. The early signatures studied include the critical slowing down as a transition point is approached, rising variance and the lag-1 autocorrelation function, skewness and a ratio of two mean first passage times. Some of the signatures could provide the experimental basis for distinguishing between bistability and excitability as the correct mechanism for the development of competence.

PACS numbers: 87.10.Mn, 87.16.Yc, 87.17.Aa, 87.18.Cf, 87.18.Tt

I. INTRODUCTION

Complex dynamical systems, ranging from ecosystems and the climate to gene regulatory networks and financial markets, are known to exhibit abrupt shifts from one dynamical regime to a contrasting one at critical parameter values [1–4]. Examples of sudden regime shifts include the collapse of vegetation when semi-arid conditions prevail, the transition from a clear lake to a turbid one, sudden changes in fish/wildlife populations [4, 5], distinct changes in the climate and patterns of oceanic circulation [1, 6], financial markets undergoing global crashes [1], spontaneous systemic failures such as asthma attacks [7], or epileptic seizures [8] etc. In a gene regulatory network, a sudden transition may occur from one stable gene expression state to another at a critical parameter value [9–12]. The induction of the *lac* operon in *E. coli* results in the synthesis of the protein *β-galactosidase* required for breaking up sugar molecules and releasing energy to the cell. There is now experimental evidence [9] that an abrupt transition from the uninduced (*β-galactosidase* level low) state to the induced (*β-galactosidase* level high) state of the *lac* operon occurs at a critical value of an inducer concentration.

The regime/state shifts in the examples mentioned above are mostly a consequence of the fold-bifurcation (or the fold-catastrophe), well-characterized in dynamical systems theory [1, 2, 6, 13]. Figure 1 illustrates a specific type of the fold-bifurcation based on bistability and hysteresis, which provides a physical understanding of the features associated with sudden regime shifts. The plot represents the steady states of a dynamical system versus a specific parameter. The state of the dynamical system at time t is defined by the magnitudes of the relevant variables at t. In the steady state, the net rate of change in the magnitude of a variable is zero so that there is no further time evolution. The solid lines in Figure 1 denote stable steady states separated by a branch (dotted line) of unstable steady states. The steady state curve is folded backwards giving rise to bistability, i.e., the existence of two stable steady states in the shaded region. The parameter values u_1 and u_2 represent the bifurcation points at which the abrupt regime shifts from one stable steady state to another occur. On crossing the bifurcation point, the system loses bistability and becomes monostable. The transition from one branch of stable steady states to the other is not reversible but exhibits hysteresis. This implies that if a transition takes place at a bifurcation point, say the upper one, the reverse transition from the upper to the lower branch occurs at the lower bifurcation point and not at the upper bifurcation point itself. Bistability owes its origin to the presence of one or more positive feedback loops governing the system dynamics with the added condition that the dynamics be sufficiently nonlinear [10–12]. Each stable steady state has its own basin of attraction in the state space, i.e., the space of all states. In the case of more than one stable steady state, the location of the initial state of the system in a basin of attraction determines the steady state attained by the system in the course of time.

The stability of a steady state indicates that the system returns to the state on being weakly perturbed from it. This is shown by arrow directions in Figure 1. If the perturbation is sufficiently strong so that a transition takes place from one basin of attraction to the other, a switch occurs between the stable steady states even before the bifurcation point is reached. Closer the system is to the bifurcation point, the lesser is the magnitude of the perturbation needed to bring about the switch. In the example of Figure 1, the branch of unstable steady states constitutes the border between the two basins of attraction. The distance of a stable steady state from the corresponding unstable steady state is a measure of the resilience (robustness against perturbation) of the system. The resilience is gradually weakened as the system approaches a bifurcation point. The dynamics of natural systems, in general, have a stochastic component due to the presence of random external influences and the inherently probabilistic nature of the processes involved in the dynamics. Consider the dynamics of gene expression in a gene regulatory network. Gene expression consists of two major steps: transcription and translation during

* mainakpl@bosemain.boseinst.ac.in
† ak.pal@bosemain.boseinst.ac.in
‡ sayantari@bosemain.boseinst.ac.in
§ indrani@bosemain.boseinst.ac.in

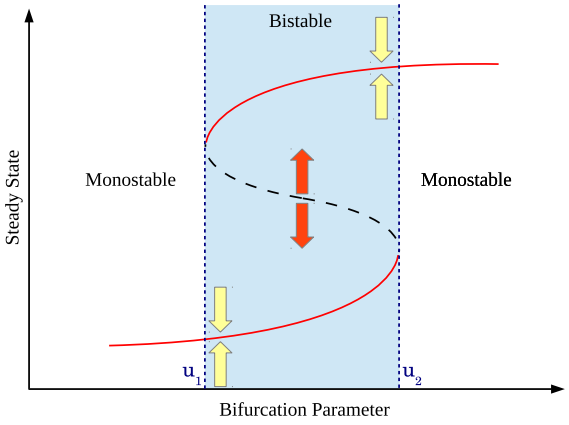

Figure 1. A generic steady state versus bifurcation parameter diagram. The shaded region represents the region of bistability separating two regions of monostability. The solid lines correspond to stable steady states and the dashed line represents unstable steady states. The points u_1 and u_2 are the lower and upper bifurcation points respectively. The arrows indicate the time evolution of a weakly perturbed system with the system moving towards stable steady states and moving away from unstable steady states

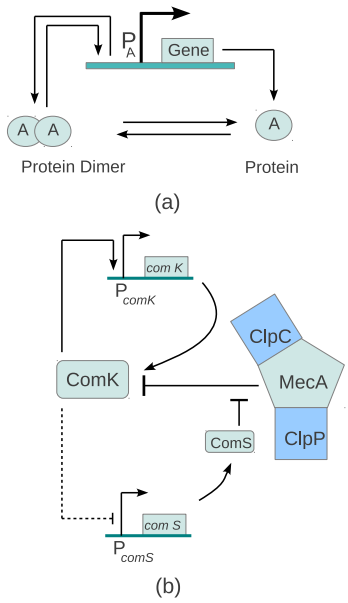

Figure 2. Two models, Model 1 (a) and Model 2 (b) of competence development. (a) The *comK* gene expresses ComK proteins (represented by A) which form dimers. A dimer binds the promoter region of the gene and activates the initiation of transcription. The dimer can also unbind from the promoter region and dissociate into free monomers. (b) The autoregulatory positive feedback loop mediated by ComK proteins is present. In addition there is a negative feedback loop in which ComK inhibits the expression of the *comS* gene and the ComS proteins repress the activity of the MecA-ClpP-ClpC Complex which targets ComK proteins for degradation.

which messenger RNA (mRNA) and protein molecules respectively are synthesized. The biochemical events (e.g., the binding of a RNA polymerase at the promoter region of the DNA to initiate transcription) constituting gene expression are inherently stochastic in character resulting in fluctuations (noise) around mean mRNA and protein levels [14, 15]. Extrinsic influences like variability in the number of regulatory molecules also contribute to the noise. Instead of a single steady state protein level, as in the case of deterministic time evolution, one now has a steady state probability distribution in the protein levels reflecting the stochastic nature of the time evolution. In the presence of low/moderate amounts of noise, the physical picture underlying sudden regime shifts still remains valid. As in the case of applied perturbations, fluctuations of sufficiently strong magnitude can bring about regime shifts before the bifurcation point is crossed. A number of early signatures of regime shifts have been proposed so far [1, 6, 16, 17] in the scenario of the fold-bifurcation. These are the critical slowing down (CSD) and its associated effects, namely, rising variances and the lag-1 autocorrelation function as the critical transition point is approached, increased skewness in the steady state probability distribution and the presence of flickering transitions between the stable steady states. The utility of such signals in the cases of ecological and financial systems [1] and complex deteriorating diseases [18] is significant, specially, in developing appropriate risk-aversion/management strategies. Also, in the absence of physical models capturing the essential dynamics, one can obtain quantitative measures of the early signatures by analyzing time-series data.

In this paper, we compute quantities providing early signatures of abrupt regime shifts in gene expression dynamics. We specifically focus on the phenomenon of competence development in the soil bacteria *B.subtilis* involving binary gene expression , i.e , the existence of two expression states. We consider two alternative models of competence development based on bistable and excitable dynamics. In Section 2, the two models are described and their different dynamics highlighted. In Section 3, we define the quantitative measures of the early signatures of sudden regime shifts. Section 4 contains the results of our computations on these early signatures. Section 5 contains concluding remarks.

II. MODELS OF COMPETENCE DEVELOPMENT

Microorganisms like bacteria have to cope with a number of stresses during their lifetime. The bacteria adopt a number of strategies to enhance their chances of survival under changed circumstances [12, 19]. One of these is the generation of phenotypic heterogeneity (no genetic change) in the bacterial population so that a fraction of the population, even if small, may survive under stressful conditions. In the *B.subtilis* population, only a part of the population, in which the level of a key transcription factor ComK is high, develops competence [12, 20]. The rest of the population is in the so-called vegetative state. The ComK protein acts as a master regulator activating the transcription of several genes including those essential for the uptake of foreign DNA from the environment. The incorporation of the new DNA into the bacterial genome confers favorable traits on the bacterial subpopulation with high ComK level, enabling the subpopulation to adapt to stress. The ComK activity resulting in the development of competence is in turn controlled

by a host of other proteins. The core module of the complex regulatory network consists of an autoregulatory positive feedback loop in which the ComK proteins promote their own production. The positive feedback gives rise to binary gene expression in the cell population, i.e., two distinct subpopulations with low and high ComK levels respectively. Two independent experiments [21, 22] provide confirmation that an autoregulatory positive feedback loop of ComK production is by itself sufficient to establish binary gene expression in a bacterial population. Considering deterministic time evolution, the steady state ComK level versus an appropriate gene expression parameter exhibits a hysteresis curve, resulting from the fold-bifurcation, similar to the one shown in Figure 1 [23]. In this scenario, if the cells in a population are prepared to be in the same initial state, all the cells evolve to the same final state giving rise to a homogeneous cell population. The generation of heterogeneity in the form of two distinct subpopulations is brought about by fluctuation-driven transitions between the low and high ComK expression states, the fluctuations being associated with the ComK levels. This gives rise to the experimentally observed [12, 20] bimodal distribution in the ComK levels, i.e., a distribution with two prominent peaks. There is now experimental evidence [24] that a lower fraction of the *B.subtilis* population develops competence with reduced noise in the low ComK level. An alternative physical mechanism underlying competence development has been proposed by Süel et. al. [25] in terms of excitability in the dynamics of the genetic circuit. The excitable core module consists of both positive and negative feedback loops which bring about a transient activation to the high ComK state. In this case, there is only one stable steady state (low ComK level) and two unstable steady states the lower of which, in terms of the ComK level, sets a threshold for switching [25]. Fluctuations in the low ComK level activate the switch to expression states in the neighborhood of the state with high ComK level which constitutes an unstable steady state. The transient activation is followed by an ultimate return to the stable low ComK state. At any instant of time, the population divides into two subpopulations with low and high ComK levels respectively, signifying a different origin for the bimodal distribution. Quantitative time lapse fluorescence microscopy provides experimental evidence [25] of the probabilistic and transient differentiation into competence.

We first consider Model 1 in which the protein product of a single gene autoactivates its own production via a positive feedback loop. The genetic circuit is shown in Figure 2(a) and, as mentioned earlier, constitutes a core module of the complex network resulting in competence development. The proteins synthesized by the *comK* gene form dimers. The dimer molecules bind at the promoter region of the DNA and activate gene expression. The gene also synthesizes proteins at a basal level. A kinetic scheme of the model is as follows [23]:

$$\begin{aligned} P_2 + G &\underset{k_2}{\overset{k_1}{\rightleftharpoons}} GP_2 \underset{k_d}{\overset{k_a}{\rightleftharpoons}} G^* \\ G &\xrightarrow{J_0} P \\ G^* &\xrightarrow{J_1} P \\ P + P &\underset{K_e}{\rightleftharpoons} P_2 \\ P &\xrightarrow{k_p} \phi \end{aligned} \tag{1}$$

The gene can be in two possible states: inactive (G) and active (G^*). Proteins are synthesized with rate constant $J_1(J_0)$ in the state $G^*(G)$ with $J_0 \ll J_1$. The synthesized proteins dimerize with K_e as the equilibrium dissociation constant. The protein dimer P_2 binds to the gene in state G and forms the complex GP_2 with k_1 and k_2 being the rate constants for binding and unbinding. The complex GP_2 in turn is activated to the state G^*, the rate constants k_a and k_d being the activation and deactivation rate constants. The synthesized proteins are degraded with a rate constant, k_p, ϕ denoting the degradation product. As shown by Karmakar and Bose [23], the kinetic scheme displayed in Equation (1) can be mapped onto a simpler scheme

$$\begin{aligned} G &\underset{k'_d}{\overset{k'_a}{\rightleftharpoons}} G^* \\ G &\xrightarrow{J_0} P \\ G^* &\xrightarrow{J_1} P \\ P + P &\underset{K_e}{\rightleftharpoons} P_2 \\ P &\xrightarrow{k_p} \phi \end{aligned} \tag{2}$$

The effective activation and deactivation rate constants $k'_a(x)$ and k'_d are:

$$k'_a(x) = k_a \frac{(x/k_s)^2}{1 + (x/k_s)^2},\ k'_d = k_d \tag{3}$$

In Equation (3), x denotes the protein concentration and $k_s = \sqrt{\frac{k_2}{k_1} K_e}$. In the simplified scheme, the rate of change of the protein concentrations:

$$\frac{dx}{dt} = \frac{J_1 k'_a}{k'_a + k_d} + \frac{J_0 k_d}{k'_a + k_d} - k_p x \tag{4}$$

The steady state condition $\frac{dx}{dt} = 0$ yields three solutions in a specific parameter regime corresponding to two stable steady states separated by an unstable steady state. The model dynamics exhibit the fold-bifurcation of the type shown in Figure 1. The rate constants J_0, J_1, and k_a serve as the bifurcation parameters. Figure 2(b) shows the genetic circuit (Model 2), proposed in Ref. [25], as governing the dynamics of competence development via excitability. The circuit contains the autoregulatory positive feedback loop of Figure 2(a). In addition there is a negative feedback loop in which the ComK protein represses the expression of the *comS* gene whereas the ComS protein inhibits the degradation of ComK by the MecA-ClpP-ClpC complex. The repression of *comS* by ComK is however,

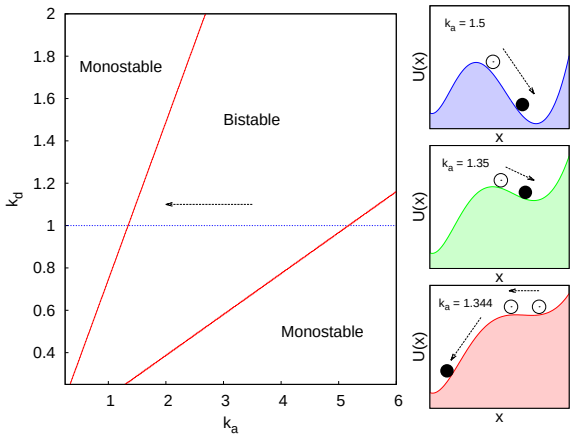

Figure 3. The $k_a - k_d$ phase diagram of Model 1 showing a region of bistability separating two regions of monostability. The stability landscape $U(x)$ versus x for three values of the bifurcation parameter is also shown. As one moves toward the lower bifurcation point $k_a = 1.344$, indicated by an arrow in the region of bistability, the basin of attraction corresponding to the high stable expression state becomes flatter. At the bifurcation point itself, the return time T_R, associated with the CSD, becomes zero.

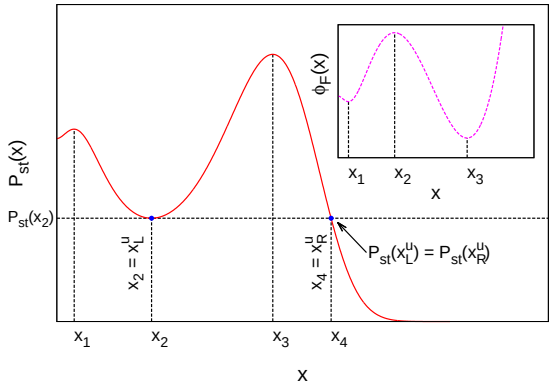

Figure 4. The steady SSPD $P_{st}(x)$ of protein levels in the region of bistability. The maxima of $P_{st}(x)$ at x_1 and x_3 represent low and high expression states respectively. The points x_2 and x_4 denote the lower and upper cut-off points of the probability distribution $P_u^n(x)$ corresponding to the high expression state. The inset shows the stochastic potential $\phi_F(x)$ versus x.

not well established under wild-type expression conditions [20]. The dynamics of the model are described in terms of the differential equations [25]:

$$\frac{dK}{dt} = a_k + \frac{b_k\, K^n}{k_0^n + K^n} - \frac{K}{1+K+S} \tag{5}$$

$$\frac{dS}{dt} = \frac{b_s}{1+(K/k_1)^p} - \frac{S}{1+K+S} \tag{6}$$

The variables K and S denote the concentrations of the ComK and ComS proteins respectively, a_k and b_k represent the basal and fully activated rates of ComK synthesis and k_0 is the ComK concentration needed for 50% activation. The Hill coefficients n and p are indicative of the cooperativities of ComK autoactivation and ComS repression respectively. The expression rate of ComS has maximal value b_s and is half-maximal at $K = k_1$. The non-linear degradation terms are a consequence of the MecA complex-mediated degradation with a competitive inhibition of the degradation by ComS.

III. EARLY SIGNATURES OF REGIME SHIFTS

The deterministic dynamics of Model 1 are governed by the rate equation shown in Equation (4). A simple stochastic version of the model has been studied in Ref. [23]. In this model, the only stochasticity considered is associated with the random transitions of the gene between the inactive (G) and active (G^*) states. The rest of the processes undergo deterministic time evolution. The simple model yields bimodal protein distributions in the steady state in certain parameter regions. In this paper, the stochastic dynamics of the model are investigated using the formulations based on the Langevin and Fokker-Planck (FP) equations. The steady state analysis of a bistable system in these formalisms is described in Refs. [26, 28–31]. The one-variable Langevin equation (LE) including both multiplicative and additive noise terms is given by

$$\frac{dx}{dt} = f(x) + g_1(x)\varepsilon(t) + \Gamma(t) \tag{7}$$

where $\varepsilon(t)$ (multiplicative noise) and $\Gamma(t)$ (additive noise) are Gaussian white noises with mean zero and correlations:

$$\begin{aligned} \langle \varepsilon(t)\varepsilon(t')\rangle &= 2d_1\delta(t-t') \\ \langle \Gamma(t)\Gamma(t')\rangle &= 2d_2\delta(t-t') \\ \langle \varepsilon(t)\Gamma(t')\rangle &= \langle \Gamma(t)\varepsilon(t')\rangle = 2\lambda_d\sqrt{d_1 d_2}\,\delta(t-t') \end{aligned} \tag{8}$$

In Equation (8), d_1 and d_2 denote the strengths of the noises $\varepsilon(t)$ and $\Gamma(t)$ respectively and λ_d is the degree of correlation between them. The first term in Equation (7) describes the deterministic dynamics. In the case of Model 1, $f(x)$ is given by the right hand side expression in Equation (4). The FP equation is a rate equation for the probability distribution $P(x,t)$, obtained from the LE as [28, 36]:

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}[A(x)P(x,t)] + \frac{\partial^2}{\partial x^2}[B(x)P(x,t)] \tag{9}$$

where

$$A(x) = f(x) + d_1 g_1(x) g_1'(x) + \lambda_d\sqrt{d_1 d_2}\, g_1'(x) \tag{10}$$

and

$$B(x) = d_1[g_1(x)]^2 + 2\lambda_d\sqrt{d_1 d_2}\, g_1(x) + d_2 \tag{11}$$

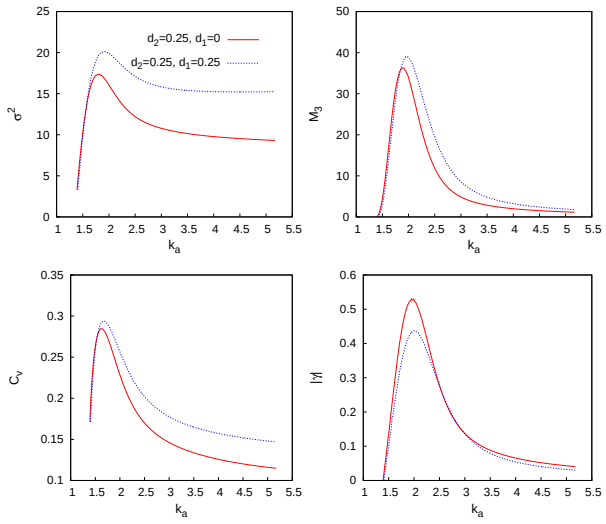

Figure 5. For Model 1, the plots of the variance σ^2, the third moment M_3, the coefficient of variation C_v and the modulus $|\gamma|$ where γ is the skewness of the normalized SSPD $P_u^n(x)$ (Equation (24)) as the bifurcation parameter k_a is decreased towards the lower bifurcation point. The solid lines correspond to the case when only additive noise is present. The dotted lines are obtained when both additive and multiplicative noise terms are present in the LE (Equation (7)). The strengths of the multiplicative and additive noise are $d_1 = 0.25$ and $d_2 = 0.25$ respectively. The other parameter values are $k_d = 1$, $k_s = 15$, $k_p = 0.03$, $J_0 = 0.01$ and $J_1 = 1$.

The steady state probability distribution (SSPD) is given by [28–30]

$$\begin{aligned} P_{st}(x) &= \frac{N}{B(x)} \exp\left[\int^x \frac{A(x)}{B(x)} dx\right] \\ &= \frac{N}{\{d_1[g_1(x)]^2 + 2\lambda_d\sqrt{d_1 d_2} g_1(x) + d_2\}^{\frac{1}{2}}} \\ &\times \exp\left[\int^x \frac{f(x')dx'}{d_1[g_1(x')]^2 + 2\lambda_d\sqrt{d_1 d_2} g_1(x') + d_2}\right] \end{aligned} \tag{12}$$

where N is the normalization constant. Equation (12) can be rewritten in the form

$$P_{st}(x) = N e^{-\phi_F(x)} \tag{13}$$

with

$$\begin{aligned} \phi_F(x) &= \frac{1}{2} \ln\left[d_1[g_1(x)]^2 + 2\lambda_d\sqrt{d_1 d_2} g_1(x) + d_2\right] \\ &- \int^x \frac{f(y)dy}{d_1\left[g_1(y)\right]^2 + 2\lambda_d\sqrt{d_1 d_2} g_1(y) + d_2} \end{aligned} \tag{14}$$

$\phi_F(x)$ defines the "stochastic potential" of the dynamics. Once the SSPD is determined, quantities like the variance and skewness, to be defined below, which provide early signatures of regime shifts can be determined.

We next define the quantitative measures of the early signatures of regime shifts. The first signature is that of the CSD which is a distinguishing feature of the dynamics close to a bifurcation point [16, 17, 32]. This involves a progressively larger relaxation time, as the bifurcation point is approached, to an attractor of the dynamics (say, a stable steady state) after being weakly perturbed from it. The physical origin of this phenomenon can be understood in terms of the stability landscape $U(x)$. The rate equation governing the dynamics of Model 1 (Equation (4)) can be written as $\frac{dx}{dt} = f(x) = -\frac{\partial U(x)}{\partial x}$. Figure 3(a) shows the phase diagram of the model in the $k_a - k_d$ plane with a region of bistability separating two regions of monostability. The other parameter values are $J_1 = 1$, $J_0 = 0.01$, $k_p = 0.03$ in appropriate units. Figure 3(b) shows the three stability landscapes at three successively decreasing values of k_a as one approaches the bifurcation point in the direction of the arrow from within the region of bistability. At point 1 ($k_a = 1.5$), the states corresponding to the minima of $U(x)$ represent the stable steady states. The location of the "ball" represents the state of the dynamical system. As one progresses from point 1 to point 2 ($k_a = 1.35$), the steady state with high value of x becomes less stable. The associated basin of attraction becomes flatter with reduced size so that it takes a longer time for a perturbed state (ball shifted from the minimum position) to relax back to the stable steady state. The relaxation time diverges at the bifurcation point where the stable steady state is on the verge of losing stability (point 3, $k_a = 1.344$). A weak perturbation pushes the ball to the minimum with low value of x, i.e., the system does not relax back to the high x state. One can define a return time T_R which provides a measure of the time taken by the dynamical system to regain a stable steady state after being weakly perturbed from it. Let x_{st} be the stable steady state value of x and $\eta(t) = x(t) - x_{st}$ be the small deviation from the steady state value under weak perturbation. The deterministic rate equation is given by $\frac{dx}{dt} = f(x)$. On Taylor expanding $f(x)$ around $x = x_{st}$ and keeping terms upto the order of $\eta(t)$, one obtains

$$\frac{d\eta}{dt} = \lambda\eta(t),\ \lambda = \left.\frac{df(x)}{dx}\right|_{x=x_{st}} \tag{15}$$

The solution of the Equation (15) is

$$\eta(t) = \eta(0)e^{\lambda t} \tag{16}$$

where $\eta(0)$ is the initial value of $\eta(t)$ at time $t = 0$. The sign of the parameter λ indicates the stability of the steady state. If λ is < 0 (> 0), the steady state is stable (unstable). We designate λ as the stability parameter. It is a well-known result from dynamical systems theory that at a bifurcation point, the stability parameter λ, associated with the steady state losing stability, becomes zero [13, 16, 17, 32]. In the case of Model 1, one can check that $\lambda = 0$ at the two bifurcation points. The return time $T_R = \frac{1}{|\lambda|}$ thus diverges at a bifurcation point. If the dynamical system is described by more than one variable, the stability of a steady state is determined by the eigenvalues of the Jacobian matrix governing the dynamics of the perturbed system [13]. Let λ_{max} be the real part of the dominant eigenvalue of the Jacobian matrix (for stability, the real parts of all the λ's should be negative). The dominant eigenvalue is the one with the largest real part. The return time T_R is given by $T_R = \frac{1}{|\lambda_{max}|}$. Examples of the experimental observations of the CSD include the transition from the $G2$ growth phase to the mitotic phase

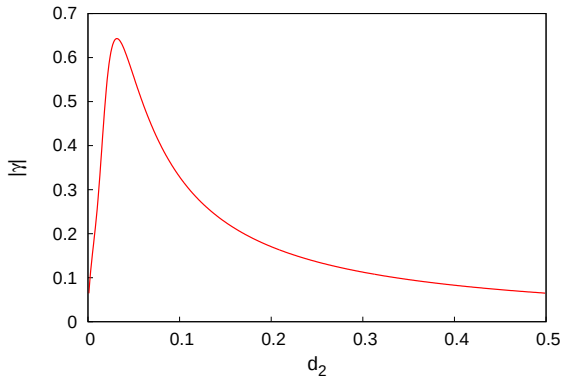

Figure 6. For Model 1, the plot of $|\gamma|$, where γ is the skewness of the normalized steady SSPD, versus the strength of additive noise d_2 for the parameter values $k_a = 1.5$, $k_d = 1.5$, $k_p = 0.03$, $J_0 = 0.01$, $J_1 = 1$, and $d_1 = 0$ (no multiplicative noise). The value $k_a = 1.5$ falls in the region of bistability.

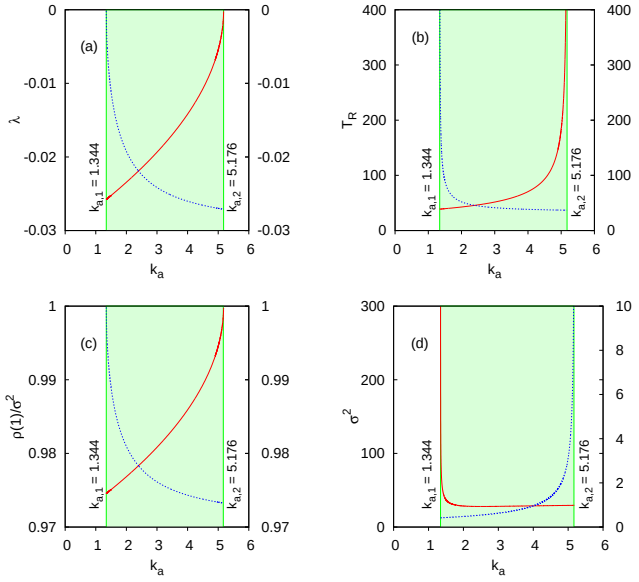

Figure 7. For Model 1, the variation of the stability parameter λ (Equation (15)), the return time T_R, the lag-1 autocorrelation $\rho_\eta(1)$ and the variance σ^2 (Equations (17) and (18)) as a function of the bifurcation parameter. The plots exhibit characteristic features at the bifurcation points $k_{a,1} = 1.344$ and $k_{a,2} = 5.176$. The parameter values are $k_d = 1$, $k_s = 15$, $k_p = 0.03$, $J_0 = 0.01$, $J_1 = 1$ and $\sigma_d = 0.25$.

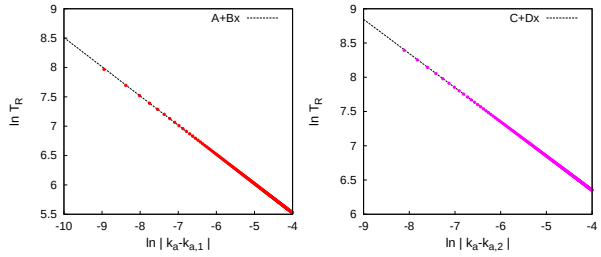

Figure 8. The computed data points in the $\ln T_R$ versus (a) $\ln|k_a - k_{a,1}|$ and (b) $\ln|k_a - k_{a,2}|$ plots are fitted by straight lines (a) $y = A + Bx$ and (b) $y = C + Dx$ respectively. The parameter values are the same as in the case of Figure 7. The values of A and C are $A = 3.53401 \pm 0.00068$ and $C = 4.35482 \pm 0.00049$. The exponents have values very close to $-\frac{1}{2}$ ($B = -0.49752 \pm 0.00013$ and $D = -0.49901 \pm 0.00099$).

of the eukaryotic cell division cycle [31], a direct observation of the CSD in a laboratory population of budding yeast before population collapse occurs at a critical experimental condition [33] and the demonstration of the CSD in a population of cyanobacteria [34].

In the presence of noise, the variance σ^2 is determined from the fluctuation-dissipation (FD) relation (Equation (A.5)) in Appendix A. In the case of a one-variable model (Model 1), the matrix $\mathbf{A}$ consists of a single element λ, the stability parameter defined in Equation (15). The steady state covariance matrix $\mathbf{C}$ reduces to a single element, the variance σ^2, which is determined as

$$\sigma^2 = \frac{D}{2 \mid \lambda \mid} \tag{17}$$

with λ negative. Also, the lag-1 autocorrelation in the stationary state (Equations (A.8) and (A.9) with $\tau = 1$) is given by

$$\rho(1) = e^{\lambda} \tag{18}$$

The variance σ^2 diverges and the lag-1 autocorrelation $\rho(1)$ reaches its maximum value at the bifurcation point since the stability parameter λ is zero at this point. Rising variance and autocorrelation thus provide early signatures of impending regime shifts. The CSD close to the bifurcation point implies that the system's intrinsic rates of change are decreased so that the state of the system at time t resembles closely the state at time $t - 1$. This increased memory is measured by the lag-1 autocorrelation function. Also, because of a flatter basin of attraction close the transition point, a given perturbation brings about a greater shift in the system's state, i.e., an increasing variance.

Two other early signatures of a regime shift which are not related to the CSD are skewness and flickering [1, 5]. The skewness γ of a probability distribution $P(x)$ is a dimensionless measure of its degree of asymmetry and is defined as

$$\gamma = \frac{\int (x-\mu)^3 P(x) dx}{\sigma^3} \tag{19}$$

where μ and σ are respectively the mean and standard deviation of $P(x)$. The variance σ^2, the coefficient of variation C_v and the third moment M_3 of $P(x)$ are expressed as

$$\sigma^2 = \int (x-\mu)^2 P(x) dx$$
$$M_3 = \int (x-\mu)^3 P(x) dx$$
$$C_v = \frac{\sigma}{\mu} \tag{20}$$

The skewness of a probability distribution increases as the bifurcation point is approached. This is because of the asymmetry in fluctuations with a system exhibiting greater amplitude deviations in the direction of the state it is fated to switch to than in the other direction. The

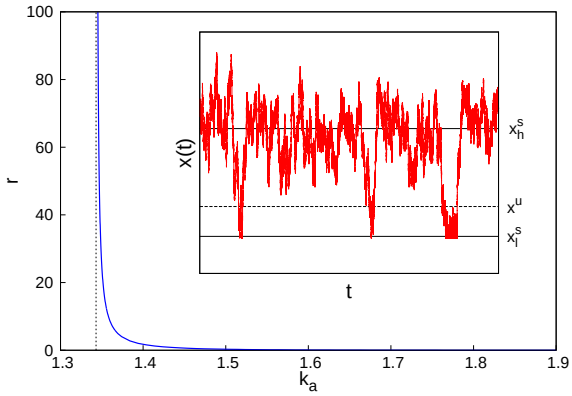

Figure 9. For Model 1, the plot of $r = \frac{T_1}{T_2}$ versus the bifurcation parameter k_a where T_1 and T_2 are the MFPTs. The ratio r is seen to diverge at the lower bifurcation point. Inset shows the time-series data of $x(t)$ versus t obtained by solving the LE. The straight lines are drawn at the steady state points x_h^s, x^u, x_l^s obtained by solving the deterministic rate equation, Equation (4). The parameter values are $k_d = 1$, $k_s = 15$, $k_p = 0.03$, $J_0 = 0.01$, $J_1 = 1$ and $d_2 = 0.25$.

phenomenon of flickering is observed in the region of bistability with the system switching back and forth between the two attractor states before the bifurcation point is reached. We propose a quantitative measure of flickering which serves as an early signature of regime shift. In the region of bistability, the stability landscape has two minima corresponding to the two stable steady states. Noise-induced transitions take the system from one valley to the other. The mean first passage time (MFPT) refers to the average first exit time from a valley [26, 28]. Let T_1 and T_2, be the MFPTs for the exits respectively from valley 1 and valley 2. The times are indicative of the amount of flickering present in the system. The MFPT T_2 becomes zero at the bifurcation point where the steady state 2 loses stability. The ratio $r = \frac{T_1}{T_2}$ measures the asymmetry in the exit times and diverges as the bifurcation point, at which $T_2 \to 0$ is approached. The quantity r thus provides an early signature of regime shift.

IV. RESULTS ON EARLY SIGNATURES

A. Model 1 (one variable)

We first present the results on early signatures in the case of Model 1. In the region of bistability, the SSPD $P_{st}(x)$ is bimodal with two distinct peaks. $P_{st}(x)$ may be obtained via a solution of the FP equation (Equation (12)) or obtained by a numerical solution of the LE (Equation(7)). Figure 4 shows an example of a bimodal distribution with two distinct peaks at $x = x_1$ and x_3. The inset of the figure shows the corresponding stochastic potential profile (Equation (14)) with x_2 denoting the location of the hill and x_1, x_3 denoting the minima of the left and right valley respectively. We find the normalized probability distributions, $P_l^n(x)$ and $P_u^n(x)$, corresponding to the low and high expression states respectively, using the following procedure:

1. From the stochastic potential profile $\phi_F(x)$, determine the position of the hill at x_2 and compute $P_{st}(x_2)$. The point x_2 corresponds to the minimum of the probability distribution between its two peaks at $x_1$and x_3. The point x_2 is chosen to be the left cut-off point, x_L^u, for $P_u^n(x)$. The right cut-off point, x_R^u, of $P_u^n(x)$ is obtained as a solution of the equation

$$P_{st}(x) = P_{st}(x_2^u),\ x > x_3 \tag{21}$$

2. If $P_{st}(0) < P_{st}(x_2)$, then the lower-cutoff point, x_L^l, of $P_l^n(x)$ is determined from the solution of the equation

$$P_{st}(x) = P_{st}(x_R^l),\ x < x_1 \tag{22}$$

where x_R^l, the right cut-off point of $P_l^n(x)$, is given by $x_R^l = x_L^u$. If $P_{st}(0) > P_{st}(x_2)$, $x_L^l = 0$.

Other cut-off procedures, e.g., that in Ref. [5], have been proposed to isolate the low and high expression probability distributions but the general results on the early signatures are qualitatively similar.

In the case of Model 1, we first consider only an additive noise in the LE (Equation (7), $g_1(x) = 0$). The additive noise $\Gamma(t)$ represents noise arising from an external perturbative influence or originating from some missing information because of rate equation approximations [31]. The SSPD, $P_{st}(x)$, in the region of bistability is obtained from Equation (12) putting $g_1(x)$, $g_1(x') = 0$. Following the prescription already given, one determines the cut-off points of the distribution $P_l^n(x)$. The normalized distributions are obtained from

$$P_l^n(x) = \frac{P'_{st}(x)}{\int_{x_L^l}^{x_R^l} P_{st}(x)dx} \quad \text{for} \quad x_L^l < x < x_R^l \tag{23}$$

$$P_u^n(x) = \frac{P'_{st}(x)}{\int_{x_L^u}^{x_R^u} P_{st}(x)dx} \quad \text{for} \quad x_L^u < x < x_R^u \tag{24}$$

In Equations (23) and (24), $P_{st}^{'}(x)$ is not normalized. With a knowledge of the normalized probability distributions $P_l^n(x)$ and $P_u^n(x)$, one can compute the skewness γ, variance σ^2, third moment M_3 and the coefficient of variation C_v using Equations (19) and (20). Figure 5 shows plots of these quantities (solid lines) versus the bifurcation parameter k_a for the probability distribution $P_u^n(x)$, i.e., considering the system to be in the high expression state. The sudden regime shift in the deterministic case occurs at the lower bifurcation point $k_{a,1} = 1.334$. The parameter values used for the computation are: $k_d = 1$, $k_s = 15$, $k_p = 0.03$, $J_0 = 0.01$ and $J_1 = 1$. The strength d_2 of the additive noise is $d_2 = 0.25$. One finds that all the four quantities $|\gamma|$, σ^2, M_3 and C_v increase as the lower bifurcation point is approached thus providing early signatures of a regime shift. The quantities, however, reach their maxima before the deterministic bifurcation point is reached and then start decreasing. The quantities, though providing early signatures, cannot provide knowledge of the bifurcation point. We next include an additional mul-

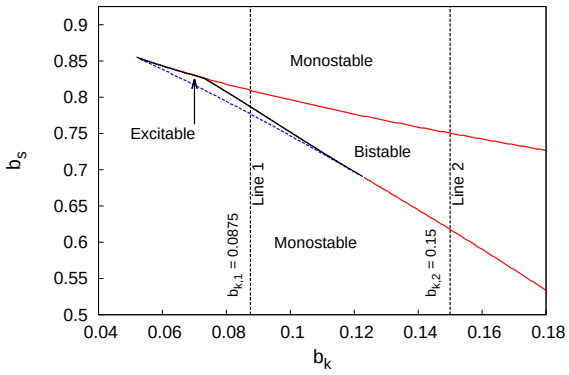

Figure 10. The phase diagram of Model 2 in the $b_k - b_s$ plane. It shows two regions of monostability, one region of bistability and one region of excitability. The boundaries shown as solid lines correspond to saddle-node (SN) bifurcations. Lines 1 and 2 mark the values of the parameter b_s for which computations are carried out. The other parameter values are the same as in Ref. [25], namely, $a_k = 0.004$, $k_0 = 0.2$, $k_1 = 0.222$, $n = 2$ and $p = 5$.

tiplicative noise term in the LE (Equation (7)) with

$$g_1(x) = \frac{k'_a}{k'_a + k_d} \tag{25}$$

The multiplicative noise is associated with the protein synthesis rate constant J_1 in the active state, i.e., $J_1 \to J_1 + \varepsilon(t)$. The origin of multiplicative noise lies in the fact that the rate constants are expected to fluctuate in time due to the inherently stochastic nature of gene expression as well as due to stochastic influences like fluctuations in the number of regulatory molecules and RNA polymerases. With both the additive and multiplicative noise terms present, the SSPDs $P_l^n(x)$ and $P_u^n(x)$ are computed following the procedure already described. The dotted curves in Figure 5 show the variations of $|\gamma|$, σ^2, M_3 and C_v , associated with the probability distribution $P_u^n(x)$, as a function of the bifurcation parameter k_a. The parameter values are the same as before with d_1, the strength of the multiplicative noise term, having the value $d_1 = 0.25$. The cross-correlation coefficient λ_d in Equation (8) is taken to be zero. Figure 6 shows how the skewness of $P_u^n(x)$ changes as a function of the additive noise strength d_2 for the parameter values $k_a = 1.5$, $k_d = 1$, $k_s = 15$, $k_p = 0.03$, $J_0 = 0.01$, $J_1 = 1$ and $d_1 = 0$ (no multiplicative noise). The value $k_a = 1.5$ is greater than the value of $k_{a,1} = 1.344$. The rising skewness is thus a signature of noise-induced regime shift. In the case of the SSPD $P_l^n(x)$, one obtains early signatures of the upper bifurcation point $k_{a,2} = 5.176$ similar to the ones shown in Figures 5 and 6 in the case of $P_u^n(x)$, though the quantitative measures exhibit less prominent variation. We find that the early signatures of regime shifts are obtained in both the cases, (i) only additive noise is present and (ii) additive as well as multiplicative types of noise are present.

Considering Model 1, we next calculate the stability parameter λ (Equation (15)), with $f(x)$ given by the expression on the r.h.s. of Equation (4), in the region of bistability. Knowing λ, the return time $T_R (= \frac{1}{|\lambda|})$ as well as the lag-1 autocorrelation $\rho(1)$ and the variance σ^2 (Equations (17) and (18)) are also determined. Figure 7 shows the variation of λ, T_R, $\rho(1)$ and σ^2 as a function of the bifurcation parameter k_a and for the parameter values $k_d = 1$, $k_s = 15$, $k_p = 0.03$, $J_0 = 0.01$, $J_1 = 1$ and $\sigma_d = 0.25$. The stability parameter λ becomes zero at the lower ($k_{a,1} = 1.344$) and the upper ($k_{a,2} = 5.176$) bifurcation points as expected. The associated return time T_R diverges at the bifurcation points, a characteristic feature of the CSD. The lag-1 autocorrelation $\rho(1)$ reaches the maximum value and the variance σ^2 diverges at the bifurcation points. These quantities are good indicators of regime shifts and carry distinct signatures of the bifurcation points. The return time T_R is known to satisfy a general scaling law $T_R \sim |B - B_i|^{-\frac{1}{2}}$ where B and B_i stand for the bifurcation parameter and point respectively [32]. In Figure 8, the scaling is demonstrated for the bifurcation parameter k_a and the bifurcation points $k_{a,1} = 1.344$ and $k_{a,2} = 5.176$. The exponent is very close to $-\frac{1}{2}$ in each case.

Figure 9 shows the variation of $r = T_1/T_2$ versus the bifurcation parameter k_a. The ratio is seen to diverge at the lower bifurcation point $k_{a,1} = 1.344$. The parameter values are $k_d = 1$, $k_s = 15$, $k_p = 0.03$, $J_0 = 0.01$, $J_1 = 1$ and $d_2 = 0.25$ (only additive noise is considered). The time-series data shown in the inset of Figure 9 is obtained via numerical solution of the LE (Equation (7)) using the algorithm described in Ref. [36]. In the limit of large times, the steady state is assumed to be reached. The solid lines in the inset mark the stable expression states x_l^s and x_h^s and the dotted line corresponds to the unstable steady state x^u. The values are obtained from a solution of the deterministic rate equation. The MFPTs T_1 and T_2 are computed using the method outlined in Ref. [37]. Let us consider a bistable potential with the stable steady states at x_1 and x_3 ($x_1 < x_3$) which are separated by an unstable steady state at x_2, termed the barrier state (Figure 4). The MFPT, $T(x; a, b)$ is the average time of the first exit from the interval (a, b) and satisfies the equation [28, 35]

$$-1 = A(x)\frac{dT(x)}{dx} + \frac{1}{2}B(x)\frac{d^2T(x)}{dx^2} \tag{26}$$

where $A(x)$ and $B(x)$ appear in the associated FP equation (Equation (9)). The MFPT $T(x_1)(= T_1)$ for exit from the basin of attraction of the stable steady state at x_1 is obtained as a solution of Equation (26) with the interval $(a, b) = (0, x_2)$ and boundary conditions given by

$$T'(a; a, b) = 0 \ \text{ and } \ T(b; a, b) = 0 \tag{27}$$

The prime denotes differentiation with respect to x, with reflecting and absorbing boundary conditions prevailing at a and b respectively [28, 35]. Following the same procedure, the MFPT $T(x_3)(= T_2)$ for exit from the basin with the stable steady state at x_3 can be calculated from Equation (26). The interval now is $(a, b) = (x_2, \infty)$ with x_2 and ∞ serving as absorbing and reflecting boundary points respectively, i.e.,

$$T(a; a, b) = 0 \ \text{ and } \ T'(b; a, b) = 0 \tag{28}$$

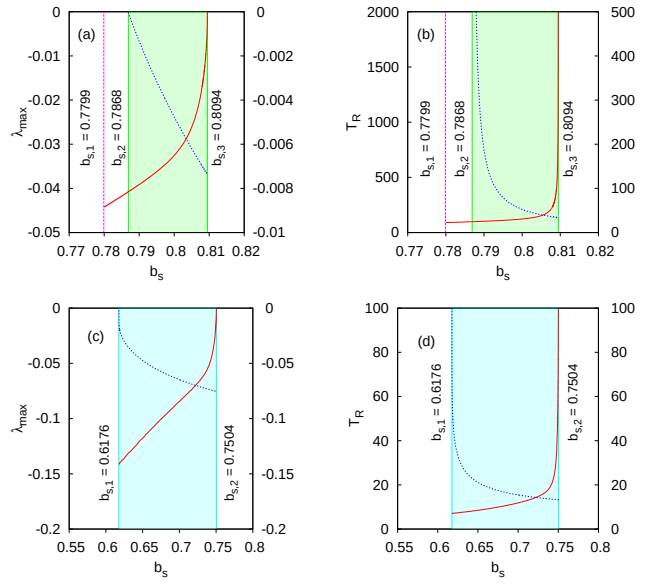

Figure 11. For Model 2, the variation of λ_{max} and T_R as a function of b_s. (a), (b): the computations are carried out along the line $b_{k,1} = 0.0875$. (c), (d): the computations are carried out along the line $b_{k,2} = 0.15$. The parameter values are the same as in the case of Figure 10.

B. Model 2 (two variables)

We next consider Model 2 the dynamics of which are governed by the set of two equations (5) and (6). In Ref. [25], the phase diagram of the model in the $b_k - b_s$ plane has been determined which has four different regions: (i) monostability with only one fixed point, (ii) bistability with three fixed points, two stable and one unstable, (iii) excitability involving three fixed points only one of which, corresponding to low ComK value, is stable. The competent fixed point (high ComK level) has the characteristic of an unstable spiral and the mid-ComK fixed point is a saddle point, (iv) only one fixed point exists which is unstable. The system exhibits limit cycle oscillations between the mid-ComK and high-ComK levels. For the purpose of our study, a part of the $b_k - b_s$ phase diagram has been recomputed and shown in Figure 10. The diagram shows three different regions: monostable, bistable and excitable. The monostable region is again of two types: monostable low (low ComK level) and monostable high (high ComK level). The boundaries between the regions depicted by solid lines correspond to the saddle-node (SN) bifurcation [13]. Two vertical lines 1 and 2 are drawn in the phase diagram at the points $b_{k,1} = 0.0875$ and $b_{k,2} = 0.15$ respectively. Line 1 intersects the phase boundaries at the three points $b_{s,1} = 0.7799$, $b_{s,2} = 0.7868$ and $b_{s,3} = 0.8094$. Line 2 has two points of intersection: $b_{s,1} = 0.6175$ and $b_{s,2} = 0.7504$. Figure 11 shows the plots of λ_{max}, the real part of the dominant eigenvalue of the Jacobian matrix computed from Equations (5) and (6), and the return time $T_R = \frac{1}{|\lambda_{max}|}$ versus b_s along line 1 ((a) and (b)) and along line 2 ((c) and (d)). For a specific steady state, the Jacobian matrix $\mathbf{J}$ has the following structure [13]:

$$\mathbf{J} = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}_{SS} \tag{29}$$

The suffix SS stands for steady state, i.e., the matrix elements are to be computed at the fixed point (x^*, y^*). A matrix element $f_{ij} = \frac{\partial f_i}{\partial x_j}$ $(i = 1, 2, j = 1, 2)$ where

$$f_1(x_1, x_2) = a_k + \frac{b_k\, x_1^n}{k_0^n + x_1^n} - \frac{x_1}{1 + x_1 + x_2} \tag{30}$$

$$f_2(x_1, x_2) = \frac{b_s}{1 + (x_1/k_1)^p} - \frac{x_2}{1 + x_1 + x_2} \tag{31}$$

with $x_1 = K$, $x_2 = S$, $n = 2$ and $p = 5$. The parameter values are the same as in Ref. [25]: $a_k = 0.005$, $k_0 = 0.2$, and $k_1 = 0.222$. The eigenvalue of $\mathbf{J}$ are λ_1 and λ_2 and λ_{max} is the real part of the dominant eigenvalue. In Figure 11, one notes that λ_{max} becomes zero at the SN bifurcation points as expected and the corresponding return time T_R diverges at a bifurcation point. The shaded regions in the Figure denote the regions of bistability in which the two stable steady states correspond to low and high ComK levels respectively. The level of ComS is anticorrelated with that of ComK. The solid (dotted) lines in Figure 11 are associated with the stable steady states representing low (high) ComK levels. Along line 2 and at the bifurcation point $b_{s,1} = 0.6176$ (Figures 11(c) and (d)) the stable steady state corresponding to the high ComK level loses stability ($\lambda_{max} = 0$, T_R diverges). At the boundary point $b_{s,2} = 0.7504$, the state representing the low ComK level loses stability. A steady state is stable if the real parts of both λ_1 and λ_2 are negative. At the point $b_{s,1} = 0.7799$ along line 1 (Figures 11(a) and (b)), there is no loss of stability of a steady state and hence no CSD with diverging T_R is observed at this point. The low ComK level of the monostable region continues to remain stable as one enters the region of excitability at the point $b_{s,1}$ along line 1. At the point $b_{s,2} = 0.7868$, the system enters the region of bistability from a region of excitability. The high ComK stable steady state loses its stability at this point with $\lambda_{max} = 0$ and a divergent T_R (dotted branch in Figures 11(a) and (b)). The low ComK state loses stability at the point $b_{s,3} = 0.8094$ when the system passes from a region of bistability to a region of monostable high ComK level. The eigenvalue $\lambda_{max} = 0$ and T_R diverges at this point. The main point to note from the results is that there is no CSD and diverging return time T_R at the transition point $b_{s,1}$ along line 1 between the regions of monostability and excitability. On the other hand, at the point $b_{s,1}$ along line 2, separating the regions of monostability and bistability, one of the expression states, namely, the high ComK state, loses stability with a divergent T_R.

We next consider the time evolution of the ComK-ComS system to be stochastic in nature. As in the case of Model 1, the FD relation (Equation (A.5)) can be used to calculate the stationary state variances of the fluctuations around the steady state and also the lag-1 autocorrelation function $\rho_{11}(\tau = 1)$ (Equations (A.8) and (A.9)). The Jacobian matrix A has the form shown in Equation (29) with

the elements calculated from Equations (30) and (31) using the relationship $f_{ij} = \frac{\partial f_i}{\partial x_j}$ $(i = 1, 2,\ j = 1, 2)$. The stoichiometric matrix is given by

$$\mathbf{S} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \tag{32}$$

The first and second rows of $\mathbf{S}$ correspond to ComK and ComS molecular numbers respectively. The "elementary complex reactions"[26, 27] considered for the computations are the composite reactions (Equations (5) and (6) with $x_1 = K$ and $x_2 = S$),

$$\begin{aligned} X_1 &\xrightarrow{a_k + \frac{b_k x_1^n}{k_0^n + x_1^n}} X_1 + 1 \\ X_1 &\xrightarrow{\frac{x_1}{1+x_1+x_2}} X_1 - 1 \\ X_2 &\xrightarrow{\frac{b_s}{1+(\frac{x_1}{k_1})^p}} X_2 + 1 \\ X_2 &\xrightarrow{\frac{x_2}{1+x_1+x_2}} X_2 - 1 \end{aligned} \tag{33}$$

The reaction propensity vector (Equation (A.2)) is given by

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} a_k + \frac{b_k x_1^n}{k_0^n + x_1^n} & 0 & 0 & 0 \\ 0 & \frac{x_1}{1+x_1+x_2} & 0 & 0 \\ 0 & 0 & \frac{b_s}{1+(\frac{x_1}{k_1})^p} & 0 \\ 0 & 0 & 0 & \frac{x_2}{1+x_1+x_2} \end{pmatrix} \tag{34}$$

With knowledge of the stoichiometric matrix $\mathbf{S}$ and the reaction propensity vector $\mathbf{f}(\mathbf{x})$, the diffusion matrix D in the FD relation can be calculated using Equation (A.7). Substituting the computed $\mathbf{A}$ and $\mathbf{D}$ matrices in the FD relation, the variances and the covariances are determined. Similarly, with the help of Equations (A.8) and (A.9), the lag-1 autocorrelation function $\rho_{11}(\tau) = \langle \delta x_1(t+\tau)\delta x_1(t)\rangle$ with $\tau = 1$ is calculated. Figures 12 (a) and (b) exhibit the plots of the variance $\sigma^2 = \langle \delta x_1^2 \rangle$, i.e. , the variance of the ComK fluctuations, as a function of the bifurcation parameters b_s. Figures 12 (c) and (d) exhibit the lag-1 autocorrelation function $\rho_{11}(1)$, associated with the ComK fluctuations, as a function of the bifurcation parameter b_s. The parameter values used in the computations are the same as in the case of Figure 11. In the cases of Figures 12(a) and (b), the computations are carried out along the line $b_{k,1} = 0.0875$ which traverses successively through the regimes of monostability, excitability, bistability and monostability. In the cases of Figures 12(c) and (d), the calculations are carried out along the line $b_{k,2} = 0.15$. The shaded regions in Figure 12 represent the regions of bistability with the solid (dotted) lines associated with low (high) ComK levels. From the plots one finds that along the line 2 ($b_{k,2} = 0.15$), the high ComK level loses stability at the bifurcation point $b_{s,1} = 0.6176$ (Figures 12(c) and (d)). The sudden regime shift from the high to the low ComK state is signaled by a diverging variance as the bifurcation point is approached and the lag-1 autocorrelation function attaining its maximum value at the point. Similar signatures are obtained at the other bifurcation point $b_{s,2} = 0.7504$ where the low

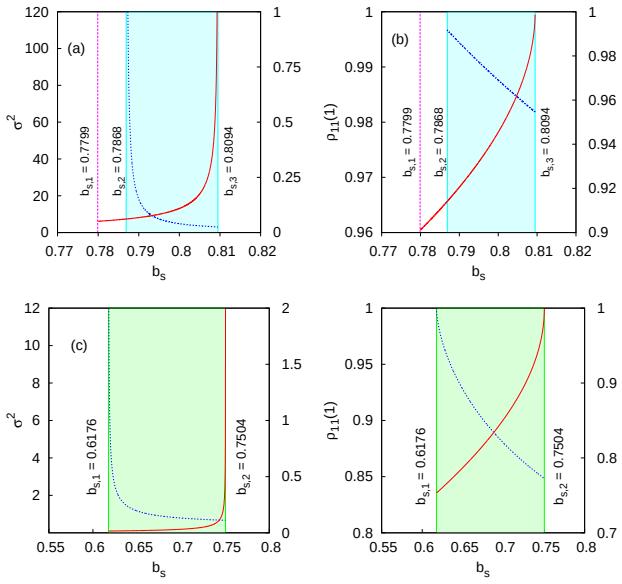

Figure 12. For Model 2, the variation of σ^2 and $\rho_{11}(1)$ as a function of b_s. (a), (b): the computations are carried out along the line $b_{k,1} = 0.0875$. (c), (d): the computations are carried out along the line $b_{k,2} = 0.15$. The parameter values are the same as in the case of Figure 10.

ComK state loses stability. In Figures 12(a) and (b), at the point $b_{s,1} = 0.7799$ along line 1 ($b_{k,1} = 0.0875$), as one enters a region of excitability from a region of monostability, the low ComK level of the monostable region continues to be stable. Since no state loses stability at the point, the variance σ^2 and the lag-1 autocorrelation function $\rho_{11}(1)$ do not provide any signatures of regime change. At the point $b_{s,2} = 0.7868$, the system enters the region of bistability from a region of excitability. The high ComK state loses its stability at this point signaled by a diverging variance and with $\rho_{11}(1)$ attaining a maximum at this point. The low ComK level loses stability at the point $b_{s,3} = 0.8094$ when the system transactions from a region of bistability to a region of monostability with high ComK level. Experimentally, the entry into a region of excitability/bistability from a region of monostability (low ComK level) is identified by the appearance of a bimodal distribution in the ComK levels as observed in single cell flow-cytometry measurements. In the case of excitability, however, there are no accompanying signatures in the measurable quantities like variance and lag-1 autocorrelation function. In the case of bistability, sharp rises in σ^2 and $\rho_{11}(1)$ indicate an impending regime shift.

V. CONCLUDING REMARKS

In this paper, we investigate the early signatures of sudden regime shifts occurring at the bifurcation points associated with a fold-bifurcation model. The basic concepts and methodology are applicable to a general class of bifurcation phenomena occurring in diverse systems. Our focus, however, is on obtaining quantitative estimates of the early signatures of regime shifts in the gene expression dynamics of competence development in *B.subtilis*. While quantities like λ_{max} and T_R are computed with a knowl-

edge of the deterministic rate equations, stochastic formalisms based on the LE, the FP equation and the LNA have been employed to calculate the other quantities. The early indicators of regime shifts, which include the CSD, variance, lag-1 autocorrelation, skewness and the ratio r of MFPTs, are shown to exhibit distinctive variations as a bifurcation point is approached. Though the literature on the early signatures of regime shifts in ecosystems, financial markets and complex diseases is extensive [1–8], the issue has not been systematically addressed in the case of gene expression dynamics, a fundamental activity in the living cell. Our study is the first in this direction and illustrates how the early signatures provide knowledge in advance of an impending regime shift. Some of these signatures may also provide clues on the physical principles underlying gene expression dynamics. As already pointed out, λ_{max} and T_R do not exhibit any distinctive features (Figures 11 (a) and (b)) when the system passes from a region of excitability (bimodal distributions in protein levels) to a region of monostability (unimodal distribution of low ComK levels). This is in contrast to the case when the system passes from a region of bistability to one of monostability. The quantity λ_{max}, associated with the steady state which loses stability, becomes zero at the bifurcation point and the return time T_R diverges (Figures 11 (c) and (d)). Experiments detecting the CSD, as the bifurcation point is approached, are difficult to carry out. Observation of the CSD in recent experiments [33, 34] provides pointers for carrying out further such experiments. The CSD experiment, if designed in the appropriate manner, would be able to distinguish between the bistability versus excitability paradigm in the case of competence development in *B.subtilis*. Experiments based on flow-cytometry and time-lapse fluorescence microscopy are easier to carry out and provide estimates of the variance σ^2 and the lag-1 autocorrelation function $\rho_{11}(1)$ [38, 39]. Similar to Figure 11, no distinctive signatures are obtained in terms of σ^2 and $\rho_{11}(1)$ as one crosses from a region of excitability to that of monostability (Figures 12 (a) and (b)) whereas prominent signatures indicate the passage from bistability to monostability (Figures 12 (c) and (d)). Thus flow-cytometry and time-lapse fluorescence microscopy measurements for a range of values of the bifurcation parameter would be able to distinguish between bistability and excitability as the physical mechanism underlying competence developement.

One utility of obtaining knowledge of the early signatures is that specific risk aversion strategies may be developed if the sudden transition is to a regime which is undesirable due to various considerations. Examples of such regime shifts include asthma attacks [7], epileptic seizures [8] and the sudden deterioration of complex diseases [18]. One may add the development of persistence of pathogens like *M. tuberculosis* in the human lung granulomas to the list. Recent experiments [40, 41] on a sister species *M. smegmatis* provide evidence that a fraction of the mycobacterial population (the population of persisters) is able to survive under nutrient depletion. To achieve this, the mycobacteria adopt the strategy of generating phenotypic heterogeneity in the form of two distinct subpopulations. The concentration of a key regulatory protein *Rel* is high in one subpopulation (which develops into the persister subpopulation) and low in the other. In the subpopulation with high *Rel* level, the stringent response pathway is initiated which help the mycobacteria to avoid death and adapt to nutrient depletion. The experiments [40, 41] show that the principle underlying the development of phenotypic heterogeneity is based on bistability and noise-induced transitions between the expression states corresponding to low and high *Rel* levels. The problem of persisters is that they are not killed by antibiotic drugs and wait for the opportune moment to restart an infection [42]. Early signatures of a regime shift from the normal to the persistent state would help in developing measures preventing the switch to persistence. Similar studies could be carried out on other systems in which bifurcation phenomena are responsible for the development of heterogeneity, choice of cell fate [19] or regime shifts leading to new types of dynamical behaviour.

Appendix: A. Linear Noise Approximation

In the Appendix A, we describe briefly the linear noise approximation (LNA) to the Chemical Master Equation (CME) [26, 27] and introduce the notations for the relevant quantities. We consider an intracellular biochemical system with volume Ω and N different chemical components. The concentrations of the components are represented in the form of the vector $\mathbf{x} = (x_1, x_2, ..., x_N)^T$ where T denotes the transpose. The chemical constituents take part in R elementary reactions. The state of the system is given by $\mathbf{x}$ which changes due to the occurrence of any one of the R reactions. We define the integers S_{ij}, $i = 1, 2, ..., N$, $j = 1, 2, ..., R$ to be the elements of a stoichiometric matrix $\mathbf{S}$. The number of molecules of the chemical components i changes from X_i to $X_i + S_{ij}$ when the jth reaction takes place.

The deterministic dynamics of the system are described by the rate equations

$$\frac{dx_i}{dt} = \sum_{j=1}^{R} S_{ij}\, f_j(x) \;\; (i = 1, 2, ..., N) \tag{A.1}$$

In a compact notation

$$\dot{\mathbf{x}} = \mathbf{S}\mathbf{f}(\mathbf{x}), \;\; \mathbf{f}\,(\mathbf{x}) = (f_1(\mathbf{x}), \cdots, f_R(\mathbf{x}))^T \tag{A.2}$$

where $\mathbf{f}(\mathbf{x})$ defines the reaction propensity vector. In the steady state, $\dot{\mathbf{x}} = 0$ with the state vector $\mathbf{x}_s$ determined from the condition $\mathbf{f}\,(\mathbf{x}_s) = 0$. Let $\delta\mathbf{x}$ denote a weak perturbation applied to the steady state, i.e., the new state vector $\mathbf{x} = \mathbf{x}_s + \delta\mathbf{x}$. On Taylor expansion of the rate vector $\frac{d\mathbf{x}}{dt}$ about the steady state and retaining only terms linear in $\delta\mathbf{x}$, one gets

$$\frac{d}{dt}\delta\mathbf{x} = \mathbf{A}\delta\mathbf{x} \tag{A.3}$$

where $\mathbf{A}$ is the Jacobian matrix the elements of which are given by

$$A_{ij} = \sum_{k=1}^{R} S_{ik} \ \frac{\partial f_k}{\partial x_j} \tag{A.4}$$

The state $\mathbf{x}_s$ is stable if all the eigenvalues of $\mathbf{A}$ have negative real parts. A deterministic dynamical model, as described above, is appropriate for describing the dynamics of a system when the number of molecules, $\mathbf{X}_i$ $(i = 1, 2, ..., N)$ is large. In reality, the biomolecules participating in cellular reactions are mostly small in number so that a stochastic description of the dynamics is more valid.

The CME describes the rate of change of the probability distribution $P(X_1, X_2, ..., X_N, t)$ of the numbers of the different chemical components. The CME is not exactly solvable in most cases and one has to take recourse to various approximate methods in order to solve the equation. The LNA provides an approximation to the CME via a large volume (Ω) expansion around the macroscopic steady state. Noise in the form of fluctuations around the steady state is expected to be small in the large Ω limit as the number of the molecules scales with the volume. To the first order in the expansion, one obtains the set of deterministic rate equations whereas in the second order, one gets the linear FPE describing fluctuations about the steady state. The stationary solution of the linear FPE is given by a multivariate Gaussian distribution. One can further show that the covariance of the fluctuations about the deterministic steady state is given by the fluctuation-dissipation (FD) relation

$$\mathbf{AC} + (\mathbf{AC})^T + \mathbf{D} = 0 \tag{A.5}$$

where $\mathbf{A}$ is the Jacobian matrix (Equation (A.4)), $\mathbf{C} = \langle \delta\mathbf{x} \, \delta\mathbf{x}^T \rangle$ is the covariance matrix, the diagonal elements of which are the variances, and $\mathbf{D}$ is the diffusion matrix. The elements of the covariance matrix are

$$C_{ij} = \langle \delta x_i \, \delta xj \rangle \quad (i, j = 1, 2, ..., N) \tag{A.6}$$

The matrix $\mathbf{D}$ has the form

$$\mathbf{D} = \mathbf{S} \ \mathrm{diag}(\mathbf{f}(\mathbf{x})) \ \mathbf{S}^T \tag{A.7}$$

where $\mathrm{diag}(\mathbf{f}(\mathbf{x}))$ is a diagonal matrix with the elements $\mathbf{f}_j(\mathbf{x})$, $j = 1, 2, ..., R$. The matrices $\mathbf{A}$ and $\mathbf{D}$ are computed at the stationary state $\mathbf{x} = \mathbf{x}_s$. Also, $\langle \delta x_i \rangle = 0$, $i = 1, 2, ..., N$. Once $\mathbf{A}$ and $\mathbf{D}$ are determined, one can determine the elements of the covariance matrix (specially, the variances) from the FD relation (A.5). The time correlation matrix for $\delta\mathbf{x}$ is given by

$$\langle \delta\mathbf{x}(t + \tau)(\delta\mathbf{x}(t))^T \rangle = \exp(\mathbf{A}\tau) \ \mathbf{C}$$

with

$$\langle \delta x_i(t + \tau) \delta x_j(t) \rangle = \sum_n \sum_m e^{\lambda_m \tau} (T^{-1})_{mn} T_{im} C_{nj} \tag{A.8}$$

In Equation (A.8), λ_m's are the eigenvalues of the Jacobian matrix $\mathbf{A}$ and $\mathbf{T}$ is a matrix the columns of which represent the right eigenvectors of $\mathbf{A}$.
The diagonal elements of the matrix are the autocovariances and τ defines the lag time. The lag-τ autocorrelation for the ith chemical component is

$$\rho(\tau) = \frac{\langle \delta x_i(t + \tau) \delta x_j(t) \rangle}{\sqrt{\mathrm{var}\,(x_i(t + \tau))} \sqrt{\mathrm{var}\,(x_i(t))}} \tag{A.9}$$

Acknowledgments

MP acknowledges the support by UGC, India, vide sanction Lett. No. F.2-8/2002(SA-I) dated 23.11.2011. SG acknowledges the support by CSIR, India, under Grant No. 09/015(0361)/2009-EMR-I.

[1] Scheffer M et al. 2009 Nature 461, 53
[2] Scheffer M 2009 *Critical Transitions in Nature and Society*(Princeton University Press)
[3] Strange C J 2007 Bioscience 57(11), 920
[4] Scheffer M and Carpenter S R 2003 Trends in Ecology and Evolution 18, 648
[5] Guttal V and Jayaprakash C 2008 Ecology Letters 11(5), 450
[6] Lenton T M 2011 Nature Climate Change 1, 201
[7] Venegas J G et al. 2005 Nature 434, 777
[8] McSharry P E, Smith L A and Tarassenko L 2003 Nature Medicine 9(3), 241
[9] Ozbudak E M, Thattai M, Lim H N, Shraiman B I and Van Oudenaarden A 2004 Nature 427, 737
[10] Pomerening J R 2008 Curr. Opin. Biotechnol. 19(4), 381
[11] Ferrell J E and Xiong W 2001 Chaos 11, 227
[12] Veening J-W, Smits W K and Kuipers O P 2008 Annu. Rev. Microbiol. 62, 193
[13] Strogatz S H 1994 *Nonlinear dynamics and chaos - with applications to physics, biology, chemistry and engineering* (Addison-Wesley)
[14] Kærn M, Elston T C, Blake W J and Collins J J 2005 Nature Rev. Genet. 6, 451
[15] Raj A and Van Oudenaarden A, 2008 Cell 135, 216
[16] Van Nes E H and Scheffer M 2007 American Naturalist 169, 738
[17] Dakos V, Van Nes E H, D'Odorico P and Scheffer M 2012 Ecology 93(2), 264
[18] Chen L, Liu R, Liu Z-P, Li M and Aihara K 2012 Scientific Reports 2: 342, 1
[19] Balázsi G, Van Oudenaarden A and Collins J J 2011 Cell 144, 910
[20] Samoilov M S, Price G and Arkin A P 2006 Sci. STKE, re17
[21] Smits W K, Eschevins C C, Susanna K A, Bron S, Kuipers O P and Hamoen L W 2005 Mol. Microbiol. 56, 604
[22] Maamar H and Dubnau D 2005 Mol. Microbiol. 56, 615

[23] Karmakar R and Bose I 2007 Phys. Biol. 4, 29
[24] Maamar H, Raj A and Dubnau D 2007 Science 317, 526
[25] Süel G M, Garcia-Ojalvo J, Liberman L M and Elowitz M B 2006 Nature 440, 545
[26] Van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam)
[27] Elf J and Ehrenberg M 2003 Genome Res. 13(11) 2475
[28] Gardiner C W 1983 *Handbook of Stochastic Methods* (Springer-Verlag, Berlin)
[29] Risken H 1984 *The Fokker-Planck Equation* (Springer-Verlag, Berlin)
[30] Bialek W 2001 Adv. in Neural Information Processing 13, *Stability and noise in biochemical switches*, Leen T K, Dietterich T G and Tresp V eds, pp 103 (MIT Press, Cambridge)
[31] Hasty J, Pradines J, Dolink M and Collins J J 2000 Proc. Natl. Acad. Sci. USA 97, 2075
[32] Wissel C 1984 Oecologia 65, 101
[33] Dai L, Vorselen D, Korolev K S and Gore J 2012 Science 336, 1175
[34] Veraart A J, Faassen E J, Dakos V, Van Nes E H, Lürling M and Scheffer M 2012 Nature 481, 357
[35] Gillespie D 1992 *Markov Processes: An Introduction for Physical Scientists* (Academic. Press, San Diego, CA)
[36] Fox R F, Gatland I R, Roy R and Vemuri G 1988 Phys. Rev. A 38, 5938
[37] Ghosh S, Banerjee S and Bose I 2012 Eur. Phys. J. E, 35(2), 11
[38] Weinberger L S, Dar R D and Simpson M L 2008 Nature Genet. 40, 466
[39] Kaufmann B B and van Oudenarden A 2007 Curr. Opin. Genet. Dev. 17, 107
[40] Sureka K, Ghosh B, Dasgupta A, Basu J, Kundu M and Bose I 2008 PLoS ONE 3, e1771
[41] Ghosh S, Sureka K, Ghosh B, Bose I, Basu J and Kundu M 2011 BMC Syst. Biol. 5, 18
[42] Balaban N Q 2011 Curr. Opin. Genet. Dev. 21, 768