Amplified Biochemical Oscillations in Cellular Systems

A. J. McKane,[1] J. D. Nagy,[2,3] T. J. Newman,[3,4] and M. O. Stefanini[4]

We describe a mechanism for pronounced biochemical oscillations, relevant to microscopic systems, such as the intracellular environment. This mechanism operates for reaction schemes which, when modeled using deterministic rate equations, fail to exhibit oscillations for any values of rate constants. The mechanism relies on amplification of the underlying stochasticity of reaction kinetics within a narrow window of frequencies. This amplification means that fluctuations have a dominant effect, even though the number of molecules in the system is relatively large. The mechanism is quantitatively studied within simple models of self-regulatory gene expression, and glycolytic oscillations.

1. INTRODUCTION

Biochemical oscillations have been studied from complementary experimental and theoretical perspectives for many years. They appear to be generic processes in biological systems, with numerous examples known in both epigenetic and metabolic contexts.[1−4] Well known instances are circadian rhythms (e.g. in microorganisms[5,6]), and the oscillation of ATP and ADP concentrations during phosphorylation of fructose-6-phosphate (F6P), a key step in glycolysis.[7,8] The study of biochemical oscillations has benefited from an intimate collaboration between experimentalists and theorists: reaction networks are pieced together in

---

[1] Theoretical Physics, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK.

[2] Life Science Department, Scottsdale Community College, Scottsdale, AZ 85256.

[3] School of Life Sciences, Arizona State University, Tempe, AZ 85287.

[4] Department of Physics, Arizona State University, PO Box 871504, Tempe, AZ 85287; e-mail: timothy.newman@asu.edu.

the laboratory, with hypotheses severely constrained by results from mathematical models. [4,8] These mathematical models are generally constructed using deterministic chemical rate equations. Within this modeling framework, the signature of oscillatory behavior is the existence of a limit cycle. The absence of a limit cycle is assumed to imply that the underlying reaction scheme, upon which the model is based, does not have cyclic behavior. Recently, a number of groups have challenged this assumption,[9–14] mainly in the context of calcium oscillations. They have studied deterministic rate equations which support a limit cycle over a certain range of parameter values, and have found, from computer simulations, that the addition of noise can expand the region of parameter space in which cycles occur. These authors demonstrate that this effect can be maximized for particular values of the system size or noise strength. On this basis, the effect has been termed "internal noise stochastic resonance." We show here that internal noise can have a far more profound effect on biochemical reaction kinetics; namely, inducing amplified oscillations in systems which, when modeled with rate equations, lack a limit cycle throughout their entire parameter space. Thus, from the view of conventional deterministic modeling, these reaction schemes would be immediately ruled out as candidates to describe biochemical oscillations.

In addition to the work already mentioned, the possibility of noise due to the discreteness of the components of a system giving rise to cycles where none previously existed, has long been discussed. [15,16] However, much of the discussion was qualitative and lacked a theoretical framework. In this paper, we develop such a framework, using the Van Kampen system-size expansion,[17–19] which allows an exact analytic description of these stochastically induced cycles—indeed, our theory reduces to a problem in linear algebra regardless of the reaction scheme under study, thereby allowing straightforward analysis. A crucial point is that this effect operates only in systems composed of a relatively modest number of molecules—typically in the range 10^2–10^6. Many biochemical reactions within cells operate with numbers of molecules in this range, which leads us to believe that intracellular processes may well exploit this amplification mechanism. Our result is at odds with the intuitive notion that fluctuations can be safely ignored for systems composed of many thousands of molecules—the reason being that, here, the amplitude of fluctuations is composed of two factors: the usual statistical factor $1/\sqrt{N}$ (where N is the typical number of molecules in the system), and a large factor $R \gg 1$ arising from the amplification of the underlying noise. If $R/\sqrt{N} \sim O(1)$ or larger, the intuitive conclusion that fluctuations can be ignored for $N \gg 1$, is incorrect.

For ease of presentation we shall introduce this mechanism through two relatively simple examples—one epigenetic (gene regulation), the other metabolic (glycolysis). Generalizations to more complex reaction schemes is straightforward, since, as already mentioned, the analysis of fluctuations reduces to an exercise in linear algebra. We will develop the theoretical ideas using the language of the gene
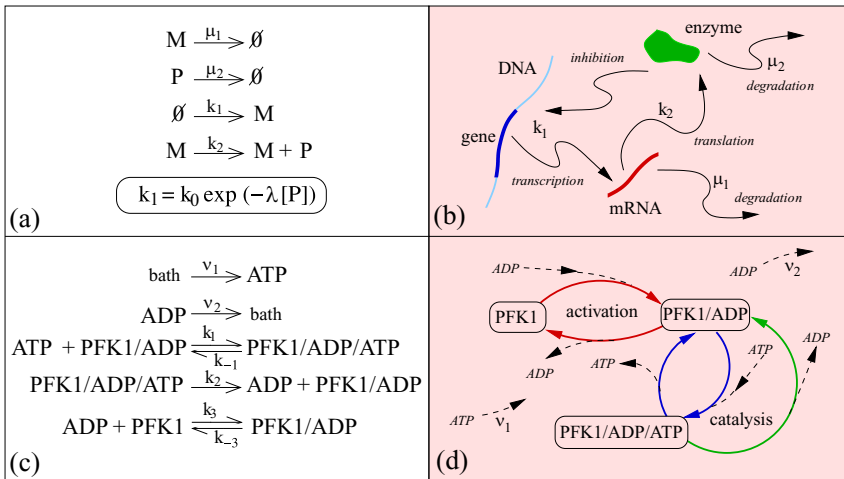
Fig. 1. Reaction networks for the two models considered in this paper. In the gene regulation model (panels a,b), mRNA (M) and the corresponding enzyme (P) have rates of degradation μ_1 and μ_2 respectively. Transcription occurs with rate k_1 (with inhibition), and translation occurs with rate k_2. The functional form of inhibition is chosen to be an exponential function of relative enzyme concentration, with "regulation parameter" λ. In Sel'kov's model for phosphorylation of fructose-6-phosphate (panels c,d), ATP enters the system with rate ν_1 and ADP leaves the system with rate ν_2. The enzyme PFK1 is activated by ADP, and this complex catalyzes the reaction with F6P (not shown) and ATP. Crucially, we assume that only one molecule of ADP is required to activate PFK1.

regulation model (defined in Fig. 1a) for concreteness, because it involves only the populations of two constituents, unlike the model of glycolytic oscillations which involves four.

The outline of the remainder of this paper is as follows. In Sec. 2 we introduce a bath model representation of the gene regulation model, and proceed to instantiate this as a master equation. The deterministic limit of the master equation is derived in Sec. 3, and is shown to correspond to the usual chemical rate equation description. We calculate the effect of weak fluctuations in Sec. 4, using the Van Kampen system-size expansion. We obtain a description of the fluctuations as a set of *linear* stochastic differential equations, which can therefore be analyzed exactly. The power-spectra obtained from these equations have resonance peaks which considerably enhance the $1/\sqrt{N}$ effects we would naively expect from fluctuations. The analysis (master equation, deterministic limit, first-order fluctuations) is repeated in a condensed form for Sel'kov's model of glycolysis in Sec. 5. Again, we find a strong peak in power spectrum of the fluctuations indicating amplified oscillations. We end with a discussion of our results in the broader context of biochemical reaction networks.

2. INDIVIDUAL-BASED STOCHASTIC MODEL FOR THE GENE REGULATION MODEL

Consider, first, a simple reaction scheme in which an enzyme inhibits the transcription of its parent gene (Fig. 1a and b). More elaborate reaction schemes based on this simple model have long been proposed to explain circadian rhythms. [5,20] It is well-known that a deterministic model of this simplest reaction scheme involving two chemical agents (mRNA and the enzyme) will fail to produce cycles—at least three chemical agents (the mRNA, the enzyme, and an intermediate form, e.g. the primary RNA transcript) are required. It is important to recall that the deterministic model is an approximate representation of the actual chemical kinetics—it is strictly accurate for an infinite bath of molecules which are well-mixed. In order to describe a well-mixed system with a finite number of molecules it is necessary to use a discrete stochastic formulation in which one tracks the probability distribution of chemical concentrations over time. We have developed a standard stochastic treatment of the two-agent reaction scheme in Fig. 1a and b. The calculational steps used in the approach are illustrated in a flow-chart (Fig. 2). By taking the limit of the number of molecules in our system, N, to be infinite, we indeed recover the deterministic theory—this is an important benchmark. To handle the case of a finite number of molecules, we perform a system size expansion, [17] which is a standard technique from the theory of stochastic processes, in which fluctuations are accounted for within a perturbative treatment. The second order terms in this expansion describe the fluctuations about the deterministic theory. These fluctuations satisfy linear equations and their statistics can be solved exactly. In particular, we calculate the power spectrum of these fluctuations. We find that for a wide range of parameters, the power spectrum has a pronounced maximum within a narrow window of frequencies.

In order to perform systematic numerical comparisons between integrating the rate equations and simulating the stochastic model it is convenient to envisage the reactions (Fig. 1a and b) occurring within two baths, as shown in Fig. 3a. The empty state Ø has been replaced by null constituents E_1 and E_2 to be discussed further below. The reason for introducing two baths is simply because it is easier to set up the dynamics of the system and is closer in spirit to other systems (grounded in population dynamics) which have been analyzed in a similar way, and are well-understood. [18,19] In addition to the pedagogical simplicity of the bath formulation, it is also biochemically meaningful to describe the statistics of molecular collisions in terms of baths. The solution of reagents consists of a system, which at any given time, can be spatially decomposed into various objects, viz. the particular molecules of interest as included in the reaction system and microscopic regions of aqueous solution. In a well-mixed system the statistics of inter-molecular collisions follows the law of mass-action. The same statistics clearly can be represented by drawing various objects from baths, as will be
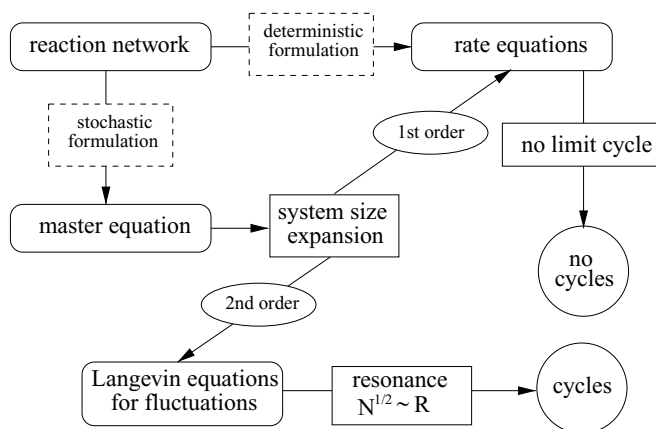
Fig. 2. A flow diagram indicating the connections between deterministic and stochastic models of the same reaction scheme. Within the deterministic model, if limit cycles are absent, then the model predicts that cycling is absent. The stochastic model is analyzed using the system size expansion. At leading order the deterministic theory is retrieved. The corrections account for stochastic fluctuations. If these are amplified (with a factor R), then the model predicts large cycles, so long as the typical number of molecules N satisfies $N^{1/2} < R$.

described in more detail shortly. This algorithm is not meant to be a realistic model of the actual dynamics, but then this is the case for any non-spatial stochastic model of chemical kinetics. Dynamically realistic models are far more complicated, and need to be formulated in terms of spatio-temporal distribution functions, such as with Boltzmann's transport equation. So long as we assume that the system is well mixed, a non-spatial approach is satisfactory.

Our aim is to build an individual-based stochastic model, and so the basic ingredients will be the number of M (mRNA) and E_1 (null) constituents in bath 1 and of P (enzyme) and E_2 (null) constituents in bath 2—in a given stochastic realization at a given time t. The dynamics will consist of picking constituents from the baths at each time step and attempting the specified reactions. Performing many runs of such reactions will enable us to collect a large number of realizations and extract average behavior. The reactions proceed according to the rates shown in Fig. 3a, and if the selection of constituents does not correspond to one of the four reactions, then the molecules are returned to their respective baths without any action being taken.

On the basis of the specification so far, and summarized in Fig. 3a, numerical simulations of the model can be carried out. We use the elegant Gillespie algorithm[21] which uses the information encoded in the reaction scheme to generate random time increments, in each of which a randomly selected reaction is forced to occur. The time increments and reactions are selected according to weighted
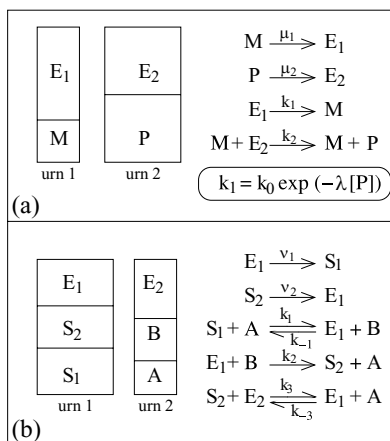
Fig. 3. In panel (a), the gene regulation reaction network (Fig. 1a and 1b) is recast in terms of two baths. Available units of space (or resources) for mRNA in bath 1 are represented by constituents E_1. Likewise, available units for the enzyme in bath 2 are represented by constituents E_2. Sel'kov's model for phosphorylation of F6P (Fig. 1c and 1d) is similarly recast in terms of two baths in panel (b). ATP (S_1) and ADP (S_2) molecules, along with available units E_1 are contained in bath 1, while the PFK1 enzyme (E_2), and its complexes (A and B) are contained in bath 2.
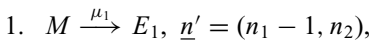
distributions in such a way that the probability distribution of the stochastic time series generated is exact. The results of some of these simulations are displayed in Figs. 4 and 7.

It is also possible to derive a set of equations which describe the stochastic process which is defined by the model we have specified. To do this, we first note that there has been an implication that the choice of constituents at a given time step is a random process, only dependent on the numbers of the various molecules in the baths at that time, and not on choices or availability at previous time steps. Assumptions of this kind imply that the process is Markov, and so can be modeled using a master equation.[17,22] This is essentially a continuous time version of a Markov chain. Before we can write down this equation, we need to define some more quantities.
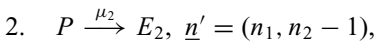
Let us denote the number of molecules of the various kinds as follows: in bath 1 the number of molecules of M is n_1 and in bath 2 the number of molecules of P is n_2. The state of the system is then denoted by the two numbers (n_1, n_2). We will frequently write this as $\underline{n}$ when we simply want to refer to the general state of the system. Note that we do not confer the status of independent variables on the numbers of E_1 or E_2 constituents. If we denote the total number of constituents in bath 1 by N_1 and that in bath 2 by N_2, then the number of E_1 and E_2 constituents are simply what is required to make up these numbers: $(N_1 - n_1)\, E_1$ and $(N_2 - n_2)$ E_2 constituents. This is why it was necessary to introduce the null constituents:

if they had not been included, the number of M and P molecules would not have the freedom to vary—they provide room for independent changes in the numbers of both M and P molecules.
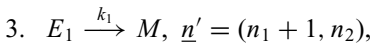
We may now define transition rates from one state, $\underline{n}$, to a different state $\underline{n}'$. For instance, in the fourth reaction, the number of P molecules increases by 1 (recall that the number of E_2 molecules is not an independent variable). So in this case, $\underline{n} = (n_1, n_2)$ and $\underline{n}' = (n_1, n_2 + 1)$. We denote the transition rate by $T(\underline{n}'|\underline{n})$. In our convention, initial states are on the right and final states on the left. When picking the constituents, there is a probability of n_1/N_1 of M being chosen, $(N_2 - n_2)/N_2$ that an E_2 is chosen, and the reaction happens at a rate k_2. This gives the result $k_2(n_1/N_1)(N_2 - n_2)/N_2$ for this particular transition rate. Others may be found in the same way. A complete listing with $\underline{n} = (n_1, n_2)$ is
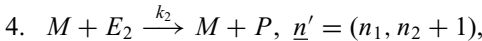
1. $M \xrightarrow{\mu_1} E_1$, $\underline{n}' = (n_1 - 1, n_2)$,

$$T(\underline{n}'|\underline{n}) = \mu_1 \frac{n_1}{N_1}.$$

2. $P \xrightarrow{\mu_2} E_2$, $\underline{n}' = (n_1, n_2 - 1)$,

$$T(\underline{n}'|\underline{n}) = \mu_2 \frac{n_2}{N_2}.$$

3. $E_1 \xrightarrow{k_1} M$, $\underline{n}' = (n_1 + 1, n_2)$,

$$T(\underline{n}'|\underline{n}) = k_1 \frac{(N_1 - n_1)}{N_1} = k_0 \exp(-\lambda n_2/N_2) \frac{(N_1 - n_1)}{N_1}.$$

4. $M + E_2 \xrightarrow{k_2} M + P$, $\underline{n}' = (n_1, n_2 + 1)$,

$$T(\underline{n}'|\underline{n}) = k_2 \frac{n_1}{N_1} \frac{(N_2 - n_2)}{N_2}.$$

Note, we use an exponential function $\exp(-\lambda[P])$ to model the down-regulation of transcription. The primary reason is this form introduces one extra parameter (λ) into the model, unlike the typical Hill form[20] $(1 + \alpha[P]^q)^{-1}$ which introduces two new parameters (α and q). Presumably the non-linear biological down-regulation mechanism is extremely complicated, so the choice of the function for a phenomenological study such as ours is somewhat arbitrary. The amplification mechanism is not sensitive to the detailed form of this function.

Having defined the transition rates, we are now in a position to write down the master equation. It has the general form[17,22]

$$\frac{d}{dt} P(\underline{n}, t) = \sum_{\underline{n}' \neq \underline{n}} T(\underline{n}|\underline{n}') P(\underline{n}', t) - \sum_{\underline{n}' \neq \underline{n}} T(\underline{n}'|\underline{n}) P(\underline{n}, t), \tag{1}$$

where $P(\underline{n}, t)$ is the probability that the system is in the state $\underline{n}$ at time t. This equation has a simple interpretation: the first term on the right hand side is the sum of the transition rates into the state $\underline{n}$ from all other states $\underline{n}'$ and the second term is the sum of the transition rates out of the state $\underline{n}$ into all other states $\underline{n}'$. When the second term is subtracted from the first, it gives the rate of change of the probability P.

So far we have formulated a discrete stochastic model and given specified analytic forms for transition probabilities between states, which can be used in conjunction with the master Eq. (1) to, in principle, solve for the probability, $P(\underline{n}, t)$, that the system is in state $\underline{n}$ at time t. We will now start from the master equation and (i) determine the form that the model takes in the mean-field limit, and (ii) implement the system-size expansion to study the origin of the cycling behavior, which is absent in the deterministic (mean-field) limit.

3. THE DETERMINISTIC LIMIT

A straightforward way of obtaining the deterministic version of the model, valid in the limit of very large system size, from the master equation is to multiply (1) by n_1 and n_2 in turn and to sum over all the states. This generates rate equations which are the deterministic equations if correlations between the variables are ignored. Let us illustrate the method in the case of n_2. We wish to calculate $\langle n_2 \rangle = \sum_{\underline{n}} n_2 P(\underline{n}, t)$ by multiplying the master equation by n_2 and summing over $\underline{n}$. On the left-hand side we find $d\langle n_2 \rangle/dt$. On the right-hand side are two terms which are nearly equal and opposite—if it was not for the n_2 factor, they would be. The only difference between the two terms is that $\underline{n}$ and $\underline{n}'$ are interchanged. So if n_2 does not change in a reaction, as for reactions 1 and 3, then the two contributions do in fact cancel out. If it decreases by 1, as in reaction 2, then a shift in the sum over n_2 gives an overall contribution of -1, and if it increases by 1, as in the reaction 4, then a shift in the sum over n_2 the other way, gives an overall contribution of $+1$. This gives the result

$$\frac{d}{dt}\langle n_2 \rangle = -\mu_2 \left\langle \frac{n_2}{N_2} \right\rangle + k_2 \left\langle \frac{n_1}{N_1} \frac{(N_2 - n_2)}{N_2} \right\rangle.$$

So far this is exact. The mean-field approximation enters through ignoring correlations, which vanish as $N_1, N_2 \to \infty$. So, for example, this means that $\langle n_i n_j \rangle = \langle n_i \rangle \langle n_j \rangle$ for $i, j = 1, 2$. If we make this approximation and introduce the fractions of M and P to be ϕ_1 and ϕ_2 respectively, in the limit $N_1, N_2 \to \infty$, then we find

$$\frac{d\phi_2}{dt} = -\frac{\mu_2}{N_2}\phi_2 + \frac{k_2}{N_2}\phi_1(1 - \phi_2).$$

This is the required deterministic equation. So in summary, if we define $\phi_i = n_i/N_i$ and scale the time by introducing $\tau = t/N_1$, then the deterministic equations corresponding to the individual based stochastic model we have defined are

$$\frac{d\phi_1}{d\tau} = -\mu_1\phi_1 + k_0 \exp(-\lambda\phi_2)(1 - \phi_1), \tag{2}$$

$$\sigma^{-1}\frac{d\phi_2}{d\tau} = -\mu_2\phi_2 + k_2\phi_1(1 - \phi_2), \tag{3}$$

where $\sigma = N_1/N_2$. Note that the mean-field approximation as applied to the first equation implies that $\langle\exp(-\lambda n_2/N_2)\rangle = \exp(-\lambda\langle n_2\rangle/N_2)$.

Most of the theoretical investigations of biochemical reactions start from a set of differential equations of the type (2) and (3). One of the first quantities of interest in such studies are the fixed points of the system. If we denote the fixed points with an asterisk and define $X = \lambda\phi_2^*$, then X satisfies the transcendental equation

$$k_0\,e^{-X} = \frac{\mu_1\mu_2 X}{k_2\lambda - (k_2 + \mu_2)X}, \tag{4}$$

with the fixed points for ϕ_1 given by

$$\phi_1^* = \frac{\mu_2 X}{k_2(\lambda - X)}. \tag{5}$$

Note that $1 - \phi_1^* > 0$ and so the denominator of the right-hand side of Eq. (4) is never zero. Also, since the left-hand side of this equation is monotonically decreasing from its value at $X = 0$ and the right-hand side is monotonically increasing from its value at $X = 0$, it follows that there is always a unique solution for X and therefore always just one fixed point. The condition $0 \leq \phi_2^* \leq 1$ implies that $0 \leq X \leq \lambda$, but the condition $0 \leq \phi_1^* \leq 1$ gives the stronger constraint

$$X \leq \frac{k_2\lambda}{k_2 + \mu_2}. \tag{6}$$

The stability of these fixed points would always be of interest, but the question takes on an added significance in our case, since the stability matrix associated with the non-trivial fixed point Eqs. (4) and (5) plays a central role in the analysis of the cycling phenomenon. To determine it, let us write the mean-field Eqs. (2) and (3) in the form $d\phi_i/d\tau = f_i(\phi)$ where $i = 1, 2$. The fixed points are found from solving $f_i(\phi) = 0$ and a linear stability analysis consists of writing $\phi_i = \phi_i^* + (\hat{\phi}_i/\sqrt{N_i})$, where the $\hat{\phi}_i$ are small deviations from the fixed point.[23] Note that we have included extra factors of $1/\sqrt{N_i}$ in the small deviations from the fixed point—which simply amounts to a re-scaling of $\hat{\phi}_i$—compared to the usual linear stability analysis, so as to make contact with the system-size expansion in the next section. Linearizing about the fixed point gives $d\hat{\phi}_i/d\tau = \sum_j M_{ij}\hat{\phi}_j$,

where M_{ij} is defined by $M_{ij} = \partial f_i / \partial \phi_j |_{FP}$. Here FP means "evaluated at the fixed point". The explicit forms for the entries of the matrix M are

$$M_{11} = -\mu_1 - k_0 \exp(-\lambda \phi_2^*),$$
$$M_{12} = -k_0 \lambda \sigma^{1/2} \exp(-\lambda \phi_2^*)(1 - \phi_1^*),$$
$$M_{21} = k_2 \sigma^{1/2}(1 - \phi_2^*),$$
$$M_{22} = -\sigma \mu_2 - \sigma k_2 \phi_1^*. \tag{7}$$

These entries may be rewritten using the fixed-point equations to give

$$M_{11} = -\frac{\mu_1}{1 - \phi_1^*}, \quad M_{12} = -\lambda \mu_1 \sigma^{1/2} \phi_1^*,$$
$$M_{21} = \mu_2 \sigma^{1/2} \frac{\phi_2^*}{\phi_1^*}, \quad M_{22} = -\sigma k_2 \frac{\phi_1^*}{\phi_2^*}. \tag{8}$$

From these expressions it is clear that all entries of M have a definite sign: M_{11}, M_{12} and M_{22} are negative, and M_{21} is positive. Therefore, the determinant and the trace of M are positive and negative respectively for all parameter choices. This already tells us that the fixed point is stable, but further analysis is required if we wish to know whether or not the fixed point is approached in an oscillatory fashion. This will be discussed again in the next section.

4. ANALYSIS OF THE FLUCTUATIONS

In contrast with the innate discreteness of the stochastic model, the mean-field equations involve functions of continuous variables. This is one of the reasons why they are more amenable to analytic treatment. It is the limit $N_1, N_2 \to \infty$ which leads to the continuity of the mean-field equations, as well as to the elimination of the fluctuations. In this section we will discuss how we can keep continuity, but still not lose the stochastic nature of the system. This is achieved by using new continuous variables x_1, x_2 in place of the previously used discrete variables n_1, n_2 to describe the probability distribution. The explicit form of the replacements are

$$\frac{n_1}{N_1} = \phi_1 + \frac{x_1}{\sqrt{N_1}}, \quad \frac{n_2}{N_2} = \phi_2 + \frac{x_2}{\sqrt{N_2}}. \tag{9}$$

The $1/\sqrt{N}$ terms are present since, by the central-limit theorem, we expect fluctuations to be of the order of $1/\sqrt{N}$ when the variables n_1, n_2 are expressed in terms of the fractions n_1/N_1 and n_2/N_2. As $N_1, N_2 \to \infty$, these fluctuations vanish, and the system is entirely described by the mean-field variables $\phi_i(\tau)$, which can be found, in principle, by solving Eqs. (2) and (3) with given initial conditions. If we imagine a plot of the probability distribution P along the vertical axis, and the variables n_i along the horizontal axes, for various values of τ, then at the

initial time $\tau = 0$, P is a delta-function spike at the starting values of the n_i. As τ increases, not only does the position of the peak move, but the probability distribution also broadens due to fluctuations. The $\phi_i(\tau)$, which are the solutions of Eqs. (2) and (3), tell us the position of the peak of the distribution and the variables $x_i(\tau)$ tell us something about the distribution itself. Actually it turns out that to order $1/\sqrt{N}$ (here we use N to mean N_1 or N_2, since they are assumed to be of the same order), the probability distribution is Gaussian, and since the position of the peak is specified, all that is left to determine is the variance. Higher terms in the $1/\sqrt{N}$ expansion give deviations from the Gaussian form, which our numerical simulations show are very small for reasonable values of N. Once the leading (deterministic) contributions have been subtracted out from (9) (by, in effect, continuously moving the origin to the peak of the probability distribution) only fluctuations remain. If the $1/\sqrt{N_1}$ and $1/\sqrt{N_2}$ terms are then factored out, we may once again take $N \rightarrow \infty$. In this way the x_i become continuous variables, and terms of different orders can be identified in the master equation.

The actual implementation of the system-size expansion is straightforward, if tedious. It is discussed clearly in Ref. 17 and in some detail in Ref. 18. We shall illustrate it by applying it to one term in the master equation only. The reader should be able to understand the essential features of the method from this, and can then consult the above references to get a broader picture. Let us take reaction 2, in which the number of P molecules decreases by 1. It gives a contribution to the term $T(\underline{n}'|\underline{n})P(\underline{n}, t)$ in the master Eq. (1) which is equal to $\mu_2(n_2/N_2)P(n_1, n_2, t)$. It also gives a contribution to $T(\underline{n}|\underline{n}')P(\underline{n}', t)$ in the same equation which is equal to $\mu_2([n_2 + 1]/N_2)P(n_1, n_2 + 1, t)$. We can combine these two terms as follows:

$$\text{Reaction 2} : \quad (\mathcal{E}_2 - 1)\left[\mu_2 \frac{n_2}{N_2} P(n_1, n_2, t)\right], \tag{10}$$

where $\mathcal{E}_2$ is a step operator which is defined in terms of its action on functions of the n_i by $\mathcal{E}_2^{\pm 1}\psi(n_1, n_2) = \psi(n_1, n_2 \pm 1)$. Similar operators can be defined for the other variables. The advantage of using these operators is that, within the replacement scheme (9), they have a simple form for large N. For example,

$$\mathcal{E}_2^{\pm 1} = 1 \pm \frac{1}{\sqrt{N_2}}\frac{\partial}{\partial x_2} + \frac{1}{2N_2}\frac{\partial^2}{\partial x_2^2} + \ldots. \tag{11}$$

Substituting (9) and (11) in (10), and expanding in inverse powers of $\sqrt{N_2}$ one finds a contribution proportional to $1/\sqrt{N_2}$:

$$\frac{1}{\sqrt{N_2}}\mu_2\phi_2 \frac{\partial \Pi}{\partial x_2}, \tag{12}$$

and a term of order $1/N_2$:

$$\frac{1}{N_2} \mu_2 \left[\frac{\partial}{\partial x_2} (x_2 \Pi) + \frac{1}{2} \phi_2 \frac{\partial^2 \Pi}{\partial x_2^2} \right]. \tag{13}$$

There are higher order terms, but this is as far as we need to go in the expansion. The quantity Π which appears in Eqs. (12) and (13) is numerically equal to P, but is instead a function of the x_i and of t. This means that the left-hand side of (1) now reads

$$\frac{\partial P}{\partial t} = \frac{\partial \Pi}{\partial t} - \sqrt{N_1} \frac{d\phi_1}{dt} \frac{\partial \Pi}{\partial x_1} - \sqrt{N_2} \frac{d\phi_2}{dt} \frac{\partial \Pi}{\partial x_2}$$

$$= \frac{1}{N_1} \frac{\partial \Pi}{\partial \tau} - \frac{1}{\sqrt{N_1}} \frac{d\phi_1}{d\tau} \frac{\partial \Pi}{\partial x_1} - \sigma^{-1} \frac{1}{\sqrt{N_2}} \frac{d\phi_2}{d\tau} \frac{\partial \Pi}{\partial x_2}. \tag{14}$$

Equating the right-hand side of the master equation (Eqs. (12) and (13)) to the left-hand side Eq. (14) order by order, we find that $\sigma^{-1} d\phi_2/d\tau$ has a contribution $-\mu_2\phi_2$ (the $\partial\Pi/\partial x_2$ cancel out) and

$$\frac{\partial \Pi}{\partial \tau} = \mu_2 \sigma \left[\frac{\partial}{\partial x_2} (x_2 \Pi) + \frac{1}{2} \phi_2 \frac{\partial^2 \Pi}{\partial x_1^2} \right] + \dots, \tag{15}$$

where the dots mean that the reactions other than 2 will also give a contribution.

This partial, but explicit, calculation allows us to see more clearly how the expansion works. At leading order we have a contribution to the mean-field equation for $\sigma^{-1} d\phi_2/d\tau$ which is equal to $-\mu_2\phi_2$, which indeed appears in Eq. (3), and in none of the other mean-field equations. Therefore, an alternative, and in some sense more systematic, way of obtaining the mean-field equations is as the leading order terms in the large-N expansion. The next-to-leading terms give a partial differential equation for the probability distribution $\Pi(\underline{x}, t)$ which, when we include all the other reactions, has the form

$$\frac{\partial \Pi}{\partial \tau} = -\sum_i \frac{\partial}{\partial x_i} (A_i(\underline{x}) \, \Pi) + \frac{1}{2} \sum_{i,j} B_{ij} \frac{\partial^2 \Pi}{\partial x_i \partial x_j}. \tag{16}$$

From (15) we see that $A_2(\underline{x}) = -\mu_2 \sigma x_2 + \dots$ and $B_{11} = \mu_2 \sigma \phi_2 + \dots$. In fact when all the reactions are included, the $A_i(\underline{x})$ remain linear functions of the x_j and the B_{ij} remain independent of them. We may therefore write

$$A_i(\underline{x}) = \sum_{j=1}^{2} M_{ij} x_j. \tag{17}$$

This means that the probability distribution at next-to-leading order, $\Pi(\underline{x}, \tau)$, is completely determined by two 2×2 matrices: M and B, whose elements are independent of the x_j, and only functions of the ϕ_j. In fact it is a characteristic

of the large-N expansion that M is nothing else but the Jacobian matrix with elements $\partial f_i / \partial \phi_j$, which may be calculated directly from the mean-field Eqs. (2) and (3). If the initial transients have died away and the solution to the deterministic equations has approached a fixed-point ϕ^*, then M will simply be the stability matrix at that fixed point. The stability matrix for the fixed point (4) and (5) is given in the previous section. The matrix B has entries

$$B_{11} = \mu_1 \phi_1 + k_0 \exp(-\lambda \phi_2)(1 - \phi_1)$$

$$B_{22} = \mu_2 \sigma \phi_2 + k_2 \sigma \phi_1 (1 - \phi_2)$$

$$B_{12} = B_{21} = 0. \tag{18}$$

So, to summarize the position so far, we know the M_{ij} and B_{ij} in terms of the parameters which define the individually-based stochastic model, and therefore by solving Eq. (16) we can find all we need to know about the fluctuations for large N_1 and N_2. The partial differential Eq. (16) is a Fokker-Planck equation—a continuous version of the master equation—and can be solved, in principle, given the initial condition that Π is a delta-function spike at $\tau = 0$. In fact, for the simple case where the B_{ij} are independent of x_j and the A_i are linear functions of the x_j, it can be solved exactly.[17] The result is a multi-variate Gaussian with $\langle x_i \rangle = 0$, as already mentioned.

Our main aim in this paper is to understand oscillations and for this one of the main tools is Fourier analysis. The form of Eq. (16) is not so useful for this purpose, but fortunately there is a completely equivalent formulation of the stochastic process which is ideally suited to investigation using Fourier transforms. Rather than write an equation for the probability distribution function Π, an equation for the actual stochastic variables $x_i(\tau)$ can be given; in other words, the problem may be formulated as a set of stochastic differential equations of the Langevin type. The Langevin equations which are equivalent to (16) are[17]

$$\frac{dx_i}{d\tau} = A_i(\underline{x}) + \eta_i(\tau), \tag{19}$$

where $\eta_i(\tau)$ is a Gaussian noise with zero mean and with a correlation function given by

$$\langle \eta_i(\tau) \eta_j(\tau') \rangle = B_{ij} \delta(\tau - \tau'). \tag{20}$$

The system defined by Eqs. (19) and (20) is ideally suited to Fourier analysis, since the Eqs. (19) are linear (since the A_i are) and Eq. (20) implies that the noise is white, that is, the Fourier transform of its correlation function is frequency-independent.

Taking the Fourier transform of (19) gives

$$-i\omega \tilde{x}_i(\omega) = \sum_{j=1}^{2} M_{ij} \tilde{x}_j(\omega) + \tilde{\eta}_i(\omega), \tag{21}$$

where the tilde denotes the Fourier transform. We may write this as

$$\sum_{j=1}^{2} \Phi_{ij}(\omega)\tilde{x}_j(\omega) = \tilde{\eta}_i(\omega), \quad \Phi_{ij}(\omega) \equiv -i\omega\delta_{ij} - M_{ij}. \tag{22}$$

The Fourier transform of $\eta_i(t)$ has the correlation function

$$\langle \tilde{\eta}_i(\omega)\tilde{\eta}_j(\omega') \rangle = B_{ij}(2\pi)\delta(\omega + \omega'). \tag{23}$$

From Eq. (22) we obtain $\tilde{x}_i(\omega) = \sum_j \Phi_{ij}^{-1}(\omega)\tilde{\eta}_j(\omega)$, and averaging the squared modulus of $\tilde{x}_i$ gives the power-spectra

$$P_i(\omega) = \langle |\tilde{x}_i(\omega)|^2 \rangle = \sum_{j=1}^{2} \sum_{k=1}^{2} \Phi_{ij}^{-1}(\omega)B_{jk}\left(\Phi^\dagger\right)_{ki}^{-1}(\omega), \tag{24}$$

where we have used $\Phi_{ij}(-\omega) = \Phi_{ji}^\dagger(\omega)$. We have omitted the proportionality factor $2\pi\delta(0)$. In practice, when comparing the analytically calculated power spectra to those generated from a numerical time series, one uses a discrete Fourier transform, and the proportionality factor is simply equal to the time increment of the recorded values in the time series.

To isolate the resonance, we note that $P_i(\omega)$ has the form of the ratio of two power series in ω. The denominator, which will largely control the position of the resonance, can be simply expressed as the determinant of the matrix Φ. If we define $D(\omega) = \det \Phi(\omega)$, then the denominator is just $|D(\omega)|^2$.

In Eqs. (21)–(24) we have written everything in a rather general form, since the formalism can be seen to generalize to the case of an arbitrary number of constituents rather easily, such as in the model of glycolysis in the next section which has 4 constituents. However, for the model we are presently considering, there are only two constituents and the expressions for $P_i(\omega)$, $i = 1, 2$ have simple explicit forms:

$$P_i(\omega) = \langle |\tilde{x}_i(\omega)|^2 \rangle = \frac{\alpha_i + \beta_i\omega^2}{|D(\omega)|^2}, \tag{25}$$

where $D(\omega) = -\omega^2 + i\omega \operatorname{tr} M + \det M$ and where

$$\alpha_1 = B_{11}M_{22}^2 + B_{22}M_{12}^2, \quad \beta_1 = B_{11},$$
$$\alpha_2 = B_{11}M_{21}^2 + B_{22}M_{11}^2, \quad \beta_2 = B_{22}. \tag{26}$$

Since the elements of the M and B matrices are known in terms of the parameters of the model and the fixed point values of the ϕ_i, so are the α_i and β_i.

The denominator of the power spectrum in Eq. (25) is given by

$$|D(\omega)|^2 = (\omega^2 - \det M)^2 + (\operatorname{tr} M)^2 \omega^2$$
$$= \left(\omega^2 - \Omega_0^2\right)^2 + \Gamma^2\omega^2, \tag{27}$$

where, since in Sec. 3 we found $\det M > 0$ and $\text{tr} M < 0$, we have introduced $\Omega_0^2 = \det M$ and $\Gamma = -\text{tr} M$. So, we may write the power spectrum as

$$P_i(\omega) = \langle |\tilde{x}_i(\omega)|^2 \rangle = \frac{\alpha_i + \beta_i \omega^2}{\left[\left(\omega^2 - \Omega_0^2 \right)^2 + \Gamma^2 \omega^2 \right]}. \qquad (28)$$

This form for the power-spectrum shows clearly the existence of a resonance: for a value of ω^2 the denominator becomes small, and the power spectrum has a large peak centered on this frequency.

To analyze the nature of the resonance in more detail, let us set $z = \omega^2$ and ask for what values of z the power spectrum (28) has a maximum. The condition $dP/dz = 0$ gives

$$\beta z^2 + 2\alpha z + \left[\alpha \left(\Gamma^2 - 2\Omega_0^2 \right) - \beta \Omega_0^4 \right] = 0, \qquad (29)$$

where we have dropped the index i on α_i and β_i. Let us begin by neglecting the term $\beta_i \omega^2$ in the numerator of Eq. (28) which may be justified for some parameter choices. Then the condition (29) simply becomes $z = (2\Omega_0^2 - \Gamma^2)/2$. Since we require $z = \omega^2 > 0$, this implies $2\Omega_0^2 > \Gamma^2$. In terms of the stability matrix M, this condition reads $2 \det M > (\text{tr} M)^2$, which implies that the eigenvalues of M are complex. In the last section we showed that the eigenvalues of the stability matrix were either real and negative or complex with a negative real part. The condition that the power spectrum has an extremum imposes the latter.

In reality, the presence of the factor $\beta_i \omega^2$ in the numerator of the power spectrum may have a significant effect, and the full condition (29) has to be used. From Eqs. (18) and (26) we see that the α_i and β_i are always positive, and so from Eq. (29) we see that the sum of the roots of this equation is negative and so at least one of the roots is negative. For the other one to be positive we require

$$\alpha \left(\Gamma^2 - 2\Omega_0^2 \right) - \beta \Omega_0^4 < 0. \qquad (30)$$

This is the general condition for an extremum of the power spectrum to occur. It goes beyond the previous condition, which only involved the eigenvalues of the matrix M, since it contains the β_i which come from consideration of the fluctuations about the fixed point. We may now ask if the extremum is a maximum. It is straightforward to show by calculating d^2P/dz^2 that, if a positive solution to Eq. (29) exists, then it is automatically a maximum.

A peak in the power spectrum indicates a resonant frequency, and a corresponding oscillation at this frequency. We must stress that our use of the word "resonance" here follows the more general usage; namely, a peak in the power spectrum as a function of frequency. In "stochastic resonance" and its derivatives [10,13] (and references therein), the "resonance" refers to a maximized response of the system as a function of some parameter, such as the strength of the internal noise. Because of the potential for confusion, we often use the term "amplification"

for the effect under consideration in this work. The height of the peak indicates the strength of the amplification. The width of the peak indicates the amount of frequency dispersion one would observe about the resonant frequency. We have calculated the power spectra exactly for this epigenetic model. For a specific set of parameters, we show the power spectra for fluctuations in mRNA and enzyme concentrations (Fig. 4c and d respectively). We have also computed stochastic realizations of time series for this reaction, using the exact Gillespie algorithm[21] (Fig. 4a and b). Note, the resonant amplification R for mRNA (defined as the height of the peak of the power spectrum relative to $P(0)$) is quite pronounced ($R \sim 30$), and, indeed, regular oscillatory peaks are discernible in the corresponding time series (Fig. 4a). The constant dashed lines in Fig. 4a and b are the results from the deterministic theory—which predicts a complete absence of cycling. We emphasize that there are two competing quantities which combine to determine the importance of these amplified oscillations—these are i) the relative height of the peak in the power spectrum R, and ii) the mean number of molecules in the system N. The amplitude of the oscillations in a time series measurement of the concentration will be of the order $R/\sqrt{N}$. This is why the effect dies away for macroscopic systems in which N is extremely large. In the example given here $R \sim 30$ (for mRNA fluctuations), and so we can estimate that the oscillations will be negligible in intracellular systems for which $N_{mRNA} \gg 1000$. Given the simplicity of this two component model, and the modest estimate above, this mechanism may be relevant to circadian rhythms in procaryotic cells, such as cyanobacteria,[5] since such cells are small, and have no nuclear membrane separating transcription and translation processes (justifying, to some degree, a two-component model).

In Fig. 5 we contrast the time series for the protein concentration (as shown in Fig. 4b) with a similar time series for which the amplification mechanism is not in effect (same parameters, but with $\lambda = 10$ instead of $\lambda = 100$). This second time series lacks regularity and has fluctuations which are an order of magnitude smaller, thus emphasizing the dramatic effect of the amplification mechanism on the underlying fluctuations in the system.

In Fig. 6 we compare the exact form for the mRNA power spectrum, given above in (25), to a numerically generated power spectrum. The latter is computed from discrete Fourier transforms of a large number (10,000) of long (60,000 iterations) time series generated using the exact Gillespie algorithm described earlier. There is very good agreement between the exact and numerical spectra, as expected.

5. SEL'KOV'S MODEL OF GLYCOLYSIS

We now turn to an example from intracellular metabolic dynamics. Consider the key step in glycolysis in which the enzyme PFK1 catalyzes the phosphorylation of F6P. This reaction is actually rather complex, and rather sophisticated models of
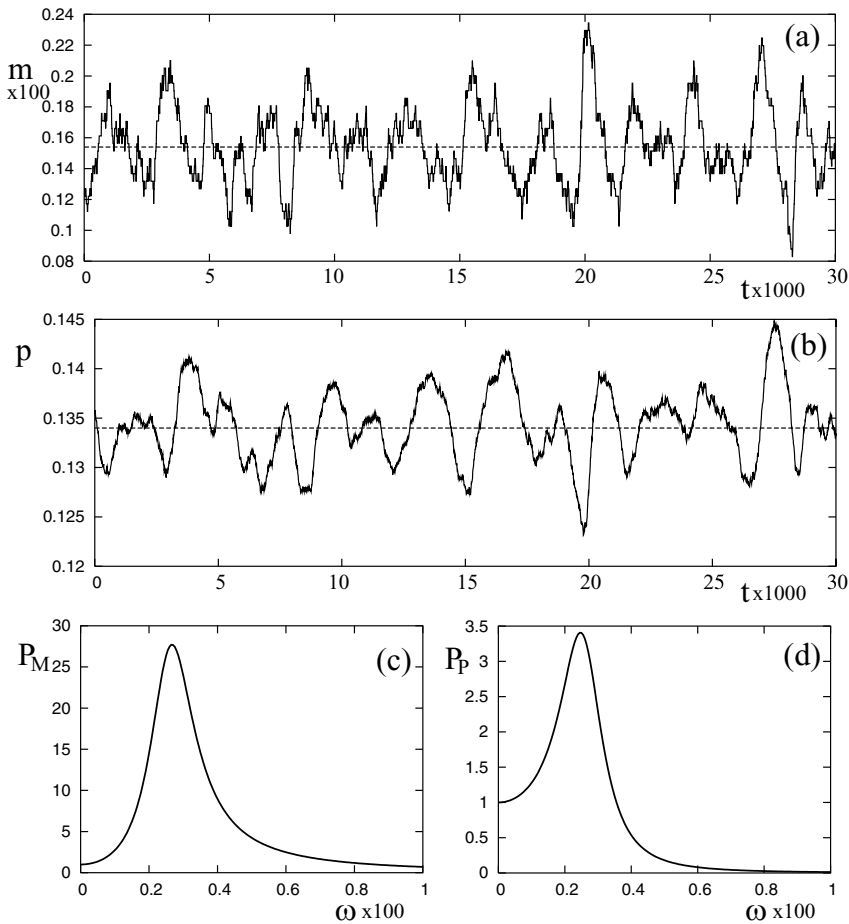
Fig. 4. Dynamics of mRNA and enzyme concentrations in the simple gene regulation model defined in Figs. 1a and 1b. The rate constants are $k_0 = 1.0$, $k_2 = 0.1$, $\mu_1 = \mu_2 = 0.001$, the regulation parameter is $\lambda = 100.0$, and the bath sizes are $N_1 = N_2/2 = 20480$. Panels (a) and (b) show time series of the relative concentrations of mRNA (m) and enzyme (p) respectively. The constant dashed lines are the predictions from the deterministic theory. Panels (c) and (d) show the power spectra (normalized so that $P(0) = 1$) associated with the fluctuations in the mRNA and enzyme concentrations respectively. The resonance peaks are the signature of this cycling mechanism. The relative amplification of fluctuations is $R = 27.9$ for mRNA, and $R = 3.4$ for the enzyme.

this process have been developed over the years, guided by ever more quantitatively precise experiments. [3,7,8] For the purposes of illustration, we shall concentrate on one of the first, and simplest, models of this reaction due to Sel'kov. [24] Although this model is no longer regarded as being an accurate representation of this reaction,
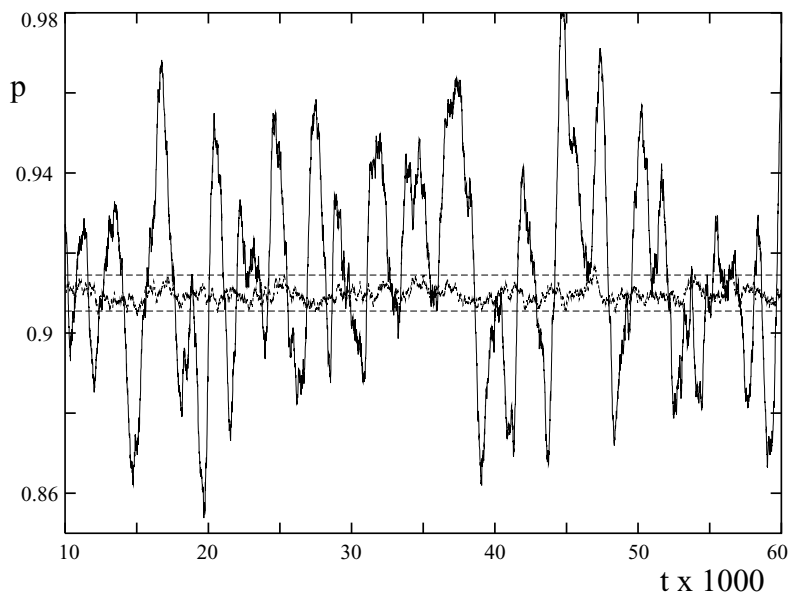
Fig. 5. Dynamics of the enzyme concentration for the gene regulation model for two different values of λ, with all other parameters as in Fig. 4. The solid curve corresponds to $\lambda = 100$ as in Fig. 4b (with a scaled ordinate), while the dashed curve corresponds to $\lambda = 10$. In this latter case the condition for oscillations is not satisfied and the amplified oscillations are replaced by typical noise. The horizontal lines correspond to the mean concentration shifted up and down by a factor of $1/\sqrt{N_2}$ and thus give an estimate of expected fluctuations based on standard central limit theorem arguments.

it captures the essence of the process. Sel'kov's reaction scheme is illustrated in Fig. 1c and d. A key parameter in this scheme is the number of ADP molecules (conventionally denoted by γ) required to activate the PFK1 enzyme. Within this deterministic modeling framework it has been shown that $\gamma > 1$ is required for cyclic behavior to emerge from the model (even though the biochemistry of PFK1 demands $\gamma = 1$), and that even then, the cycling exists in a very narrow range of parameter values. We have reformulated Sel'kov's reaction scheme using the stochastic framework, but have restricted our attention to the more biologically plausible case of $\gamma = 1$.

As before, the deterministic model is retrieved when the number of molecules in the system is taken to be infinitely large, and predicts constant, non-cycling, concentrations. When this number is finite, we account for the fluctuations in the system using the system size expansion, and solve the simple linear theory in order to calculate the power spectra for the various chemical agents.

The reactions which comprise the Sel'kov model of a key stage (phosphorylation of F6P by the PFK1 enzyme) in glycolysis are shown schematically in
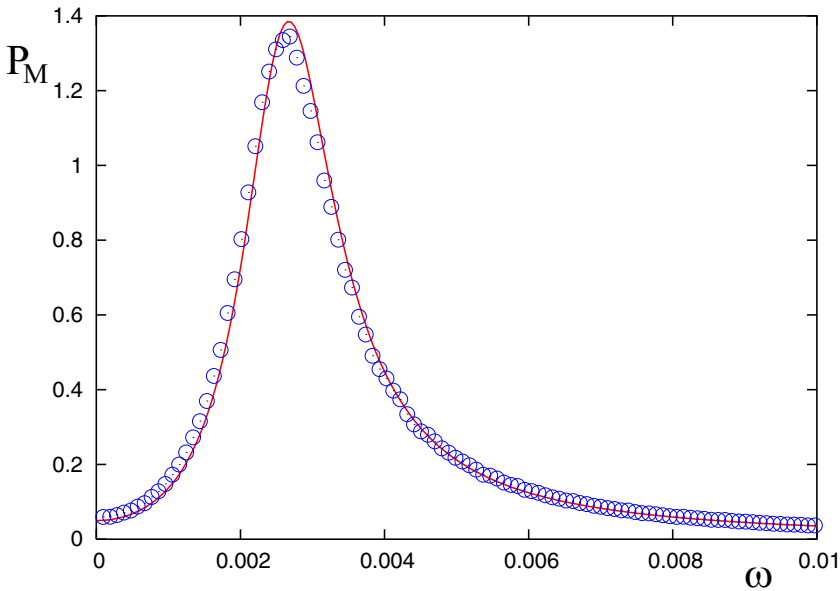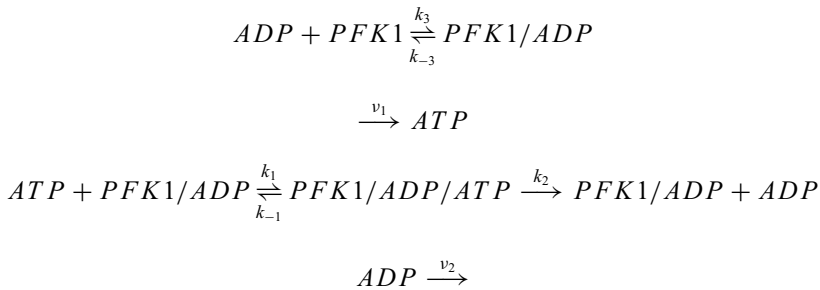
Fig. 6. The unnormalized power spectrum for mRNA fluctuations, as a function of frequency (for parameter values, refer to Fig. 4c). The solid line is the exact result (25), while the circles correspond to the power spectrum computed from the discrete Fourier transforms of 10,000 mRNA time series generated from Gillespie's algorithm.

Fig. 1c and d. Writing them out in a slightly different format, they read

$$ADP + PFK1 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} PFK1/ADP$$

$$\xrightarrow{v_1} ATP$$

$$ATP + PFK1/ADP \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} PFK1/ADP/ATP \xrightarrow{k_2} PFK1/ADP + ADP$$

$$ADP \xrightarrow{v_2}$$

Once again we will introduce two baths.

For notational simplicity we will denote ATP by S_1, ADP by S_2, PFK1 by E_2, PFK1/ADP by A and PFK1/ADP/ATP by B. The first two will be placed in bath number 1, and the last three in bath number 2 (Fig. 3b). It will also be necessary to introduce "nulls" into bath 1, which we will denote by E_1. These represent the aqueous space which can potentially be filled by ATP or ADP. With this notation A is defined to be $E_2 S_2$ and B is defined to be $S_1 E_2 S_2$ and the five reactions are

1. $E_1 \xrightarrow{\nu_1} S_1$
2. $S_2 \xrightarrow{\nu_2} E_1$
3. $S_1 + A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B + E_1$
4. $B + E_1 \xrightarrow{k_2} A + S_2$
5. $S_2 + E_2 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} A + E_1$.

Let us denote the number of molecules of the various kinds as follows: in bath 1, the number of molecules of S_α, $\alpha = 1, 2$ is m_α, and in bath 2, the number of molecules of A is n_1 and of B is n_2. The state of the system is then denoted by the four numbers (m_1, m_2, n_1, n_2), which we will again write as $\underline{n}$ when we simply want to refer to the general state of the system. If, as before, we denote the total number of constituents in bath 1 by N_1 and that in bath 2 by N_2, then the number of E_1 and E_2 constituents is simply $N_1 - m_1 - m_2$ and $N_2 = n_1 - n_2$, respectively.

The transition rates for this model are:

1. $\underline{n}' = (m_1 + 1, m_2, n_1, n_2)$.

$$T(\underline{n}'|\underline{n}) = \nu_1 \frac{(N_1 - m_1 - m_2)}{N_1}.$$

2. $\underline{n}' = (m_1, m_2 - 1, n_1, n_2)$.

$$T(\underline{n}'|\underline{n}) = \nu_2 \frac{m_2}{N_1}.$$

3. Forward reaction: $\underline{n}' = (m_1 - 1, m_2, n_1 - 1, n_2 + 1)$.

$$T(\underline{n}'|\underline{n}) = k_1 \frac{m_1}{N_1} \frac{n_1}{N_2}.$$

  Backward reaction: $\underline{n}' = (m_1 + 1, m_2, n_1 + 1, n_2 - 1)$.

$$T(\underline{n}'|\underline{n}) = k_{-1} \frac{(N_1 - m_1 - m_2)}{N_1} \frac{n_2}{N_2}.$$

4. $\underline{n}' = (m_1, m_2 + 1, n_1 + 1, n_2 - 1)$.

$$T(\underline{n}'|\underline{n}) = k_2 \frac{(N_1 - m_1 - m_2)}{N_1} \frac{n_2}{N_2}.$$

5. Forward reaction: $\underline{n}' = (m_1, m_2 - 1, n_1 + 1, n_2)$.

$$T(\underline{n}'|\underline{n}) = k_3 \frac{m_2}{N_1} \frac{(N_2 - n_1 - n_2)}{N_2}.$$

  Backward reaction: $\underline{n}' = (m_1, m_2 + 1, n_1 - 1, n_2)$.

$$T(\underline{n}'|\underline{n}) = k_{-3} \frac{(N_1 - m_1 - m_2)}{N_1} \frac{n_1}{N_2}.$$

To find the deterministic equation we define $s_\alpha = m_\alpha/N_1, a = n_1/N_2$ and $b = n_2/N_2$ and scale the time by introducing $\tau = t/N_1$. This gives the deterministic equations, corresponding to the individual based stochastic model we have defined, to be

$$\frac{ds_1}{d\tau} = v_1(1 - s_1 - s_2) - k_1 s_1 a + k_{-1}(1 - s_1 - s_2)b, \tag{31}$$

$$\frac{ds_2}{d\tau} = -v_2 s_2 + k_2(1 - s_1 - s_2)b - k_3 s_2(1 - a - b)$$
$$+ k_{-3}(1 - s_1 - s_2)a, \tag{32}$$

$$\sigma^{-1}\frac{da}{d\tau} = -k_1 s_1 a + k_{-1}(1 - s_1 - s_2)b + k_2(1 - s_1 - s_2)b$$
$$+ k_3 s_2(1 - a - b) - k_{-3}(1 - s_1 - s_2)a, \tag{33}$$

$$\sigma^{-1}\frac{db}{d\tau} = k_1 s_1 a - k_{-1}(1 - s_1 - s_2)b - k_2(1 - s_1 - s_2)b, \tag{34}$$

where $\sigma = N_1/N_2$. The Eqs. (31)–(34) are Sel'kov's equations[8,24] except for the additional factors of $(1 - s_1 - s_2)$ which arise from the introduction of the nulls E_1.

To investigate the fixed point structure, let us first note that if we replace the term $(1 - s_1 - s_2)$ by Ω in the Eqs. (31)–(34), it will enable us to examine the Sel'kov model ($\Omega = 1$) and our deterministic model $\Omega = 1 - s_1 - s_2$, in tandem. From Eqs. (31) and (34) we see that the fixed point has either to have $\Omega^* = 0$ or $v_1 = k_2 b^*$ (all fixed point values are again denoted by asterisks). The first condition cannot hold in the original Sel'kov model, but can in our version of the model: in this case we see from Eq. (32) that $s_2^* = 0$, and so $s_1^* = 1$. This implies $a^* = 0$ and b^* is indeterminate. If $\Omega^* \neq 0$, remarkably a unique fixed point is found in both forms of the model, and moreover it takes on a reasonably simple form. The specific results are:

- *Fixed points of the original Sel'kov model*

$$s_1^* = \frac{[v_2 k_{-3} + v_1 k_3]}{k_1 k_2 k_3}(k_2 + k_{-1})\left(1 - \frac{v_1}{k_2}\right)^{-1}, \quad s_2^* = \frac{v_1}{v_2},$$

$$a^* = \frac{v_1 k_3}{[v_2 k_{-3} + v_1 k_3]}\left(1 - \frac{v_1}{k_2}\right), \quad b^* = \frac{v_1}{k_2}. \tag{35}$$

- *Fixed points of the modified Sel'kov model*
  The trivial fixed point: $s_1^* = 1, \quad s_2^* = 0, \quad a^* = 1.$

The non-trivial fixed point:

$$s_1^* = \frac{[v_2 k_{-3} + v_1 k_3] (k_2 + k_{-1})}{\Delta}, \quad s_2^* = \frac{v_1}{v_2} \frac{k_1 k_2 k_3}{\Delta} \left(1 - \frac{v_1}{k_2} \right),$$

$$a^* = \frac{v_1 k_3}{[v_2 k_{-3} + v_1 k_3]} \left(1 - \frac{v_1}{k_2} \right), \quad b^* = \frac{v_1}{k_2}, \tag{36}$$

where

$$\Delta \equiv k_1 k_2 k_3 \left(1 - \frac{v_1}{k_2} \right) \left(1 + \frac{v_1}{v_2} \right) + [v_2 k_{-3} + v_1 k_3] (k_2 + k_{-1}).$$

To make contact with the discussion in earlier sections, let us introduce the notation $\phi_1 = s_1, \phi_2 = s_2, \phi_3 = a, \phi_4 = b$. Then the mean-field Eqs. (31)–(34) take the form $d\phi_i/d\tau = f_i(\phi)$ where $i = 1, \ldots, 4$. Performing a linear stability analysis gives the matrix M_{ij}. The explicit forms for the entries of this matrix for the case when ϕ^* is the fixed point (36) are given in the Appendix.

To study the fluctuations we introduce new continuous variables x_1, x_2, x_3, x_4 in place of the previously used discrete variables m_1, m_2, n_1, n_2, to describe the probability distribution. The explicit form of the replacements are

$$\frac{m_1}{N_1} = \phi_1 + \frac{x_1}{\sqrt{N_1}}, \quad \frac{m_2}{N_1} = \phi_2 + \frac{x_2}{\sqrt{N_1}},$$

$$\frac{n_1}{N_2} = \phi_3 + \frac{x_3}{\sqrt{N_2}}, \quad \frac{n_2}{N_2} = \phi_4 + \frac{x_4}{\sqrt{N_2}}. \tag{37}$$

Proceeding as in Sec. 4 we carry out a system-size expansion. We find the deterministic Eqs. (31)–(34) to leading order and the Fokker-Planck equation (16) at next-to-leading order, with $A_i(\underline{x})$ given by Eq. (17), except now that j runs from 1 to 4. So in this case, the probability distribution at next-to-leading order, $\Pi(\underline{x}, \tau)$, is completely determined by two 4×4 matrices M and B. The stability matrix for the fixed point (36) is given in the Appendix, where we also give the entries of the matrix B. The analysis of the Fokker-Planck equation, the conversion to a set of Langevin equations and the subsequent Fourier analysis is now exactly as in Sec. 4, but with $i, j = 1, \ldots, 4$.

We have calculated the power spectra for the ATP and ADP concentration fluctuations exactly. We find that these power spectra have very large and sharp peaks for a wide range of parameter values (Fig. 7c and d)—indicating that the concentrations of these molecules will undergo significant oscillatory behavior within a small-system setting. Explicit stochastic simulations of the reaction network show that indeed large amplitude cycling occurs (Fig. 7a and b). The constant dashed lines are the predictions from the corresponding deterministic theory—showing a complete absence of cycling, as expected. The power spectrum peaks
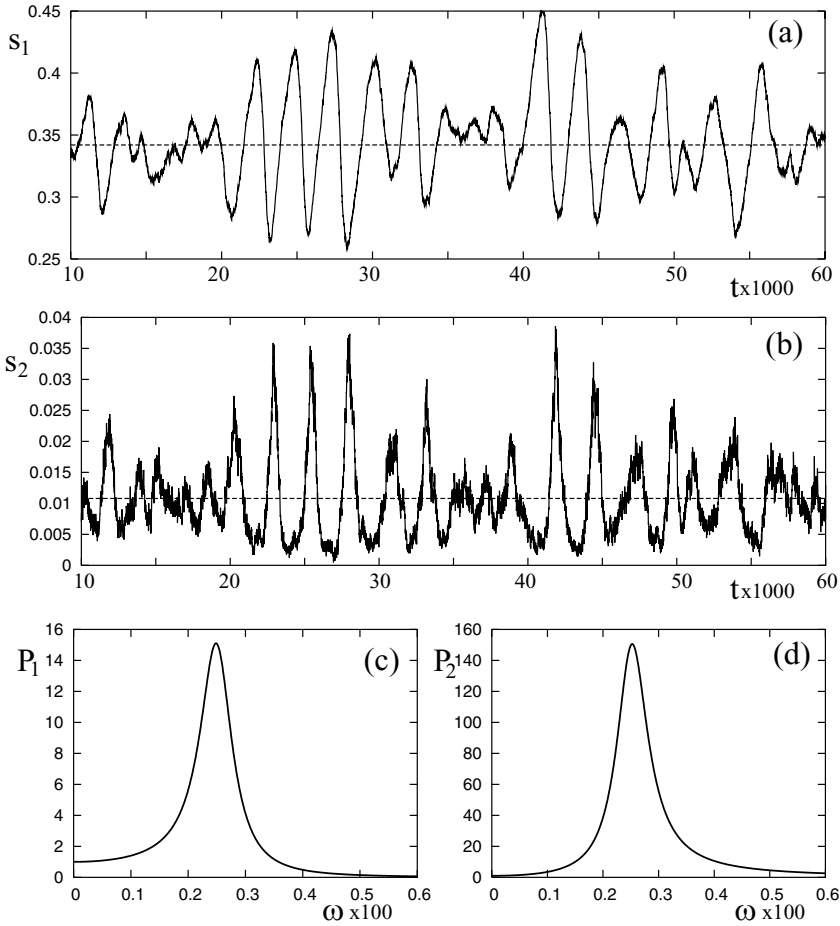
Fig. 7. Dynamics of ATP and ADP concentrations in the stochastic reformulation of Sel'kov's reaction scheme. The rate constants are $k_1 = 1.0$, $k_{-1} = 0.5$, $k_2 = 0.2$, $k_3 = 0.2$, $k_{-3} = 1.0$, $v_1 = 0.0005$, $v_2 = 0.03$, and the bath sizes are $N_1 = N_2 = 4096$. Panels (a) and (b) show time series of the relative concentrations of ATP (s_1) and ADP (s_2) respectively. The constant dashed lines are the predictions from the deterministic theory. Panels (c) and (d) show the normalized power spectra associated with the fluctuations in the ATP and ADP concentrations respectively. The sharp resonance peaks are the signature of this cycling mechanism. The relative amplification of fluctuations is $R = 15.1$ for ATP, and $R = 150.4$ for ADP.

found in this example are rather large—of the order of 150 for ADP. This indicates that intracellular environments composed of up to 25,000 ADP molecules operating according to this reaction scheme will show pronounced oscillations, caused by amplification of the underlying stochasticity of the reaction kinetics.

6. DISCUSSION AND CONCLUSIONS

We have described a mechanism by which a microscopic biochemical system can show an amplification of stochastic fluctuations in the reaction kinetics of its constituents. This mechanism leads to sizable oscillations in the concentrations of the reagents for reaction schemes, which when modeled using rate equations, show no cycling behavior for any values of their rate constants. We have illustrated the effect in two very different biological examples—self-regulation of a gene, relevant to the study of circadian rhythms, and the dynamics of ADP, ATP, and PFK1 concentrations during glycolysis.

We stress here that the system-size expansion always leads to linear equations for the fluctuations, with coefficients related to the steady-state concentrations predicted from the first-order theory (i.e. the deterministic rate equations). Thus, the evaluation of the power spectra is simply an exercise in linear algebra. The signature of the amplification mechanism is the existence of a peak in the power spectra. The existence of a peak is guaranteed if the stability matrix of the deterministic rate equations has complex eigenvalues, with negative real parts to ensure stability. This provides a simple test for the existence of the mechanism directly from the deterministic theory. Crudely speaking, if the approach to the deterministic steady-state occurs via damped oscillations, then the inclusion of second-order fluctuations will lead to the amplification of sustained oscillations. It is also important to point out that the mechanism described here requires no external tuning of rate constants. This is because the underlying stochasticity has a flat spectrum in frequency space (i.e. white noise), and this automatically excites the resonant frequencies of the system. The oscillations are also robust—in the two examples given in this paper, cycles are present over a very broad range of parameter values.

The examples of circadian rhythms and glycolysis given here are very simple and designed to illustrate the amplification mechanism in two well-known areas of cell biology. It will be important to analyze more realistic (and hence, more complex) reaction networks using the same techniques. There is intense current interest in analyzing the dynamical properties of such networks, especially with regard to robustness.[25] Deterministic rate equations are used for these analyzes, but this may need to be revisited on the basis of the results of this paper. For a network involving M different chemical reagents, the analysis of fluctuations in the system size expansion reduces to linear algebra with matrices of rank M. If oscillatory behavior is found, the "special frequencies" will be closely related to the M normal modes of the system. This implies an interesting analogy with the theory of small oscillations in mechanics,[26] whereby complex oscillatory motion can be decomposed into normal modes, with the lowest frequency modes almost completely characterizing the dynamics. As we have stressed, this mechanism only operates within a microscopic system, such as a cell, or internal region of a cell. There is, however, a means by which this mechanism can lead to macroscopic

oscillations in a population of cells; namely, synchronization of the individual cellular cycles through weak intercellular interactions.

The amplification mechanism described here may be considered as a possible cause for oscillatory phenomena in more complicated biological situations. The hypothesis is theoretically elegant in that it does not require *ad hoc* nonlinearities postulated simply to "generate" cycles. For example, endogenous circadian rhythms in eucaryotes appear to be controlled by a complex transcriptional feedback loop involving the genes *clock, cry, Bmal1* and members of the *per* gene family.[27,28] Oscillations in the activities of these genes may be explained by this new mechanism if the number of mRNAs, translational complexes, and product proteins involved are small enough to allow the amplification term R to dominate. This is not unreasonable given that the number of mRNAs (for a given gene) in a eucaryotic cell ranges from $10^1 - 10^4$.[29]

Although substantial evidence implicates nonlinear feedbacks between downstream intermediates and upstream effector enzymes as the cause of *in vivo* glycolytic oscillations,[30] other causes cannot be ruled out. The amplification mechanism described here might also cause or modify *in vivo* oscillations. The system size of glycolysis, especially if compartmentalized in a eucaryotic cell, approaches that required for noise amplification to have a pronounced effect. Intracellular F6P concentration in the yeast *Saccharomyces cerevisiae* is approximately 0.1 mM.[31] Therefore, a cell of 10 μm diameter contains on the order of 10^6 to 10^7 F6P molecules. In mammalian cells, the concentration of PFK1 appears to vary between 0.1 and 1 micromolar,[32] yielding on the order of 10^3 to 10^4 PFK tetramers per cell. Most eucaryotic cells have on the order of 10^9 ATP molecules, but some orders of magnitude less ATP participates directly in glycolysis. There is also the possibility of an interplay between limit cycles and amplified stochastic oscillations since, for systems which allow limit cycles, these two mechanisms sit either side of the bifurcation point. If physiological parameter values shift over time, it is possible that the mechanism responsible for oscillations may shift from one form to another.

In conclusion, the condition for this effect of amplified oscillations is simply the existence of damped oscillations in the corresponding rate equation model (*i.e.* complex eigenvalues with negative real part in the stability matrix). Thus, this mechanism is likely to be widespread in biochemical networks, implying that dynamics in the cytoplasmic environment may be far richer than formerly supposed on the basis of chemical rate equations.

A. EXPLICIT FORMS OF THE MATRICES M AND B

As explained in Sec. 5 of the main text, the matrix M, for the set of deterministic equations $d\phi_i/d\tau = f_i(\underline{\phi})$, is defined by $M_{ij} = \partial f_i/\partial \phi_j|_{\text{FP}}$. For the system

(31)–(34) the entries of the matrix are

$$M_{11} = -v_1 - k_1 a - k_{-1} b, \quad M_{12} = -v_1 - k_{-1} b,$$

$$M_{13} = -\sigma^{1/2} [k_1 s_1], \quad M_{14} = \sigma^{1/2} [\Omega k_{-1}],$$

$$M_{21} = -k_2 b - k_{-3} a, \quad M_{22} = -(v_2 + k_3) + (k_3 - k_{-3})a + (k_3 - k_2)b,$$

$$M_{23} = \sigma^{1/2} [k_{-3}\Omega + k_3 s_2], \quad M_{24} = \sigma^{1/2} [k_2 \Omega + k_3 s_2],$$

$$M_{31} = \sigma^{1/2} [(k_{-3} - k_1)a - (k_2 + k_{-1})b],$$

$$M_{32} = \sigma^{1/2} [k_3 + (k_{-3} - k_3)a - (k_{-1} + k_2 + k_3)b],$$

$$M_{33} = \sigma [-k_1 s_1 - k_3 s_2 - k_{-3}\Omega], \quad M_{34} = \sigma [(k_{-1} + k_2)\Omega - k_3 s_2],$$

$$M_{41} = \sigma^{1/2} [k_1 a + k_{-1} b + k_2 b], \quad M_{42} = \sigma^{1/2} [(k_{-1} + k_2)b],$$

$$M_{43} = \sigma [k_1 s_1], \quad M_{44} = \sigma [-(k_{-1} + k_2)\Omega], \tag{A.1}$$

where $\Omega = 1 - s_1 - s_2$. The a, b, s_1 and s_2 in these matrix elements are all assumed to be evaluated at the fixed point (36).

The noise-correlation matrix, B_{ij}, is symmetric and given by

$$B_{11} = v_1(1 - s_1 - s_2) + k_1 s_1 a + k_{-1}(1 - s_1 - s_2)b,$$

$$B_{12} = 0, \quad B_{13} = \sigma^{1/2} [k_1 s_1 a + k_{-1}(1 - s_1 - s_2)b], \quad B_{14} = -B_{13},$$

$$B_{22} = v_2 s_2 + k_2(1 - s_1 - s_2)b + k_3 s_2(1 - a - b) + k_{-3}(1 - s_1 - s_2)a,$$

$$B_{23} = \sigma^{1/2} [k_2(1 - s_1 - s_2)b - k_3 s_2(1 - a - b) - k_{-3}(1 - s_1 - s_2)a],$$

$$B_{24} = \sigma^{1/2} [-k_2(1 - s_1 - s_2)b],$$

$$B_{33} = \sigma [k_1 s_1 a + k_{-1}(1 - s_1 - s_2)b + k_2(1 - s_1 - s_2)b].$$
$$\qquad + k_3 s_2(1 - a - b) + k_{-3}(1 - s_1 - s_2)a],$$

$$B_{34} = \sigma [-k_1 s_1 a - k_{-1}(1 - s_1 - s_2)b - k_2(1 - s_1 - s_2)b],$$

$$B_{44} = \sigma [k_1 s_1 a + k_{-1}(1 - s_1 - s_2)b + k_2(1 - s_1 - s_2)b]. \tag{A.2}$$

ACKNOWLEDGMENTS

REFERENCES

1. B. Hess and A. Boiteux, *Ann. Rev. Biochem.* **40**:237 (1971).
2. A. Goldbeter and S. R. Caplan, *Ann Rev. Biophys. Bioeng.* **5**:449 (1976).

3. A. Goldbeter, *Biochemical Oscillations and Cellular Rhythms* (Cambridge University Press, 1996).
4. J. J. Tyson, in *Computational Cell Biology*, C. P. Fall, E. S. Marland, J. M. Wagner, J. J. Tyson, Eds. (Springer-Verlag, 2002) Ch. 9.
5. P. L. Lakin-Thomas and S. Brody, *Ann. Rev. Microbiol.* **58**:489 (2004).
6. M. Nakajima, *et al.*, *Science* **308**:414 (2005).
7. P. Smolen, *J. Theor. Biol.* **174**:137 (1995).
8. J. Keener and J. Sneyde, *Mathematical Physiology* (Springer-Verlag, 1998).
9. L. Meinhold and L. Schimansky-Geier, *Phys. Rev. E* **66**:050901(R) (2002).
10. Z. Hou and H. Xin, *J. Chem. Phys.* **119**:11508 (2003).
11. M. Falcke, *Biophys. J.* **84**:42 (2003).
12. M. Falcke, *Adv. Phys.* **53**:255 (2004).
13. H. Li, Z. Hou and H. Xin, *Phys. Rev. E* **71**:061916 (2005).
14. U. Kummer, *et al.*, *Biophys. J.* **89**:1603 (2005).
15. M. S. Bartlett, *J. Roy. Stat. Soc. A* **120**:37 (1957).
16. R. M. Nisbet and W. S. C. Gurney, *Modelling Fluctuating Populations* (Wiley, New York, 1982).
17. N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 1981).
18. A. J. McKane and T. J. Newman, *Phys. Rev. E* **70**:041902 (2004).
19. A. J. McKane and T. J. Newman, *Phys. Rev. Lett.* **94**:218102 (2005).
20. G. Kurosawa, A. Mochizuki and Y. Iwasa, *J. Theor. Biol.* **216**:193 (2002).
21. D. T. Gillespie, *J. Comp. Phys.* **22**:403 (1976).
22. E. Renshaw, *Modelling biological populations in space and time* (Cambridge University Press, Cambridge, 1991).
23. S. H. Strogatz, *Nonlinear Dynamics and Chaos* (Addison Wesley, 1994).
24. E. E. Sel'kov, *Eur. J. Biochem.* **4**:79 (1968).
25. G. von Dassow, E. Meir, E. M. Munro and G. M. Odell, *Nature* **406**:188 (2000).
26. H. Goldstein, *Classical Mechanics* (Addison Wesley, Reading, MA, ed. 2, 1980).
27. D. Bell-Pedersen, *et al.*, *Nature Rev. Genetics* **6**:544 (2005).
28. M. Akashi and T. Takumi, *Nature Struct. Biol.* **12**:441 (2005).
29. B. Alberts, *et al.*, *Molecular Biology of the Cell* (Garland Publishing, New York, ed. 4, 2002).
30. M. F. Madsen, S. Dano and P. G. Sorensen, *FEBS J.* **272**:2648 (2005).
31. B. Teusink, *et al.*, *Eur. J. Biochem.* **267**:1 (2000).
32. J. B. Hansen and C. M. Veneziale, *J. Lab. Clin. Med.* **95**:133 (1980).