Intrinsic fluctuations in stochastic delay systems: Theoretical description and application to a simple model of gene regulation

Tobias Galla*
Theoretical Physics, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, United Kingdom
(Received 21 January 2009; published 12 August 2009)

The effects of intrinsic noise on stochastic delay systems is studied within an expansion in the inverse system size. We show that the stochastic nature of the underlying dynamics may induce oscillatory behavior in parameter ranges where the deterministic system does not sustain cycles, and compute the power spectra of these stochastic oscillations analytically, in good agreement with simulations. The theory is developed in the context of a simple one-dimensional toy model, but is applicable more generally. Gene regulatory systems in particular often contain only a small number of molecules, leading to significant fluctuations in messenger RNA (mRNA) and protein concentrations. As an application we therefore study a minimalistic model of the expression levels of *hes*1 mRNA and Hes1 protein, representing the simple motif of an autoinhibitory feedback loop and motivated by its relevance to somite segmentation.

DOI: 10.1103/PhysRevE.80.021909 PACS number(s): 87.18.Tt, 02.50.Ey, 87.18.Cf, 05.10.−a

I. INTRODUCTION

Most processes in biology are intrinsically stochastic, due to the random fashion in which molecules interact. In order for a biochemical reaction to occur, for example, all reagents must be sufficiently close in space, and due to thermal or other types of stochasticity the execution of reaction is fundamentally a stochastic process. This type of randomness has, until recently, mostly been neglected in attempts to model biochemical reaction systems, and deterministic rate equations have often been used to describe the dynamics of such systems. Noise has here often been assumed to have only a minor effect on the dynamics so that it could safely be ignored. The use of deterministic approaches implies the assumption of large formally infinite system sizes; only in this limit can the law of large numbers be applied to show that the resulting mean-field equations give an accurate description of the dynamics of the system. Additionally, it is frequently assumed implicitly that the reactor in which the chemical dynamics takes place is well mixed so that spatial variation in concentrations of the interacting chemicals can be ignored.

The reason for the popularity of such approaches undoubtedly rests in their relative mathematical simplicity: while the methods with which to analyze sets of nonlinear deterministic differential equation are fully developed (see e.g., [1] or similar textbooks), a theory for the corresponding stochastic systems is far less advanced. If the number of reacting molecules in the system is small, the stochastic effects can no longer be ignored. An important example are messenger RNA (mRNA) molecules in gene expression [2–5], where only a small number of molecules is involved in the reaction dynamics. Unsurprisingly, deviations from the mean-field dynamics are here to be expected, and stochastic rather than deterministic modeling approaches to such systems in molecular biology are appropriate. Only in recent years have analytical and more systematic studies of such systems been undertaken, and substantial differences between the behavior of stochastic systems and their deterministic counterparts have been found in different model systems. In particular so-called demographic noise [6] due to the discreteness of the dynamics may change the structure of the attractor of a given system fundamentally. References relevant for the present work can be found, e.g., in [3–5,7–9].

Stochastic approaches to biochemical reaction systems typically start from a master equation describing the microscopic dynamics; the deterministic mean-field dynamics can then formally be derived to lowest order within a van Kampen expansion in the inverse system size [10]. Taking into account next-to-leading order, finite-size corrections can alter the dynamics considerably. Predator-prey systems with a fixed point on the deterministic level can for example be seen to exhibit coherent sustained oscillations at finite sizes [7]. The spectrum of these cycles can be obtained analytically to striking precision within the system-size expansion. Similar oscillations have been found in a variety of other systems, including models of epidemiology, opinion dynamics, and biochemical reaction networks [8,9].

In addition to the discreteness of the dynamics and the resulting intrinsic stochasticity, processes in gene regulatory systems are typically subject to considerable delays induced by the underlying biochemical reactions, i.e., processes such as transcription and translation do not occur instantaneously, but generate their reaction products only well after the reaction has been triggered [3–5]. The aim of the present paper is therefore to extend the theoretical tools developed in [7–9] to the case of stochastic delay systems, and to use them to study a simple model of gene regulation. As we will see the dynamics of such systems may well exhibit stochastic coherent oscillations at finite system size in ranges of the reaction rates where the deterministic infinite delay system has no cycles, but approaches a fixed point instead. Such oscillations in delay stochastic systems have been reported in [3–5], but to our knowledge a theoretical computation of correlations and power spectra of these cycles within a systematic expansion in the inverse system size has not been attempted in the context of such models. Theoretical approaches based on generating functions have however been discussed in [5].

*tobias.galla@manchester.ac.uk

In this paper we will follow a complementary approach, and in particular we will describe how the method of the system-size expansion applies to delay systems, and how a linear delay Langevin equation can be derived for fluctuations about the trajectory of the deterministic mean-field system. From this Langevin equation the power spectra of these stochastic cycles can then be obtained analytically. Our analysis therefore offers a theoretical characterization of results from simulations reported, e.g., in [4], and an alternative to the analytical approaches discussed in [5].

II. TOY MODEL

A. Definition and deterministic description

In order to develop the formalism we start with a simple model of delay stochastic processes, and consider a system in which there is only one reacting substance X. The dynamics are given by the following reactions:

$$A \to A + X, \tag{1}$$

$$B + X \to B, \tag{2}$$

$$X + C \Rightarrow C. \tag{3}$$

Note that neither reaction affects the number molecules of substances A, B, and C in the system, so that these reactants are mere "dummy" variables, and their only role is to set the relative rates with which the three reactions occur. While the first two reactions are assumed to occur instantaneously, the third reaction involves a delay, we indicate this by the double arrow in Eq. (3), i.e., if a reaction between a molecule of type C and a molecule of type X is triggered at time t, then one molecule of type X is removed from the system at a later time $t+\tau$ (provided there is at least one X molecule in the system at this later time). For simplicity we will assume that τ is a constant delay period, but variable delay times, drawn, e.g., from some probability distribution, can be in principle considered as well [3]. The model in this setup has previously been discussed and studied within an alternative approach in [5].

On the mean-field level the concentration of X molecules in the system is described by the delay differential equation

$$\dot{x}(t) = a - bx(t) - cx(t-\tau), \tag{4}$$

where a, b, and c are non-negative rate constants related to the (constant) number of molecules of types A, B, and C in the system, respectively (further details will be discussed below).

Equation (4) is linear and its asymptotic behavior can be computed straightforwardly. In particular a linear stability analysis has been carried out in [5], and a phase diagram was obtained in terms of the parameters a, b, c, and τ. The only fixed point of Eq. (4) is $x^*=a/(b+c)$, and at fixed a it is found to be unstable at large τc or small τb, respectively. In such circumstances oscillations grow indefinitely. Below the line marking the Hopf bifurcation in the $(\tau b, \tau c)$ plane, the fixed point is stable [5]. We will focus on this regime in the following.

B. Stochastic dynamics, van Kampen expansion, and spectrum of fluctuations

In order to model the dynamics on the microscopic level, let us assume the reactor in which the various reactions take place contains $a\Omega$ molecules of type A, $b\Omega$ molecules of type B, $c\Omega$ molecules of type C, and n particles of type X. Since the number of A, B, and C molecules remains unchanged by the reactions given above, the only dynamical variable in the system is $n(t)$ (which will be of order Ω as are the numbers of the other particles in the system). The first reaction, Eq. (1), then occurs with a rate $a\Omega$, the second reaction with rate nb, and the third with rate nc, and the resulting stochastic process is described by the master equation [5]

$$\frac{d}{dt}P(n,t) = a\Omega(\mathbb{E}^{-1}-1)P(n,t) + b(\mathbb{E}-1)[nP(n,t)] + c\sum_{m=0}^{\infty} m(\mathbb{E}-1)[\Theta(n)P(n,t;m,t-\tau)] \tag{5}$$

for the probability $P(n,t)$ of finding the system in state n at time t. $\mathbb{E}$ is here an operator acting on a function of n via $\mathbb{E}f(n)=f(n+1)$, and not to be confused with an expectation value of some kind. $\mathbb{E}^{-1}$ stands for the inverse operation. $\Theta(n)$ is the step function, i.e., $\Theta(n>0)=1$ and $\Theta(n=0)=0$, which ensures that the delayed removal of X molecules only occurs provided there is at least one molecule present in the system at the time at which the removal is due to take place. Note that Eq. (5) is not closed on the level of one-time quantities, as $P(n,t;m,t-\tau)$ describes the joint probability distribution of finding n X molecules at time t and m X molecules at time $t-\tau$.

Following [10] and anticipating that n will be of order Ω with fluctuations of order $\Omega^{1/2}$ we now introduce continuous degrees of freedom, and write

$$\frac{n(t)}{\Omega} = x(t) + \frac{\xi(t)}{\Omega^{1/2}}. \tag{6}$$

The above master equation for $P(n,t)$ can then be written in terms of the distribution $\Pi(\xi,t)$, and within an expansion in powers of $\Omega^{-1/2}$ we have similar to [10]

$$\begin{aligned}\partial_t\Pi(\xi,t) - \Omega^{1/2}\frac{\partial\Pi(\xi,t)}{\partial\xi}\dot{x} &= a\Omega\left[-\Omega^{-1/2}\frac{\partial}{\partial\xi} + \Omega^{-1}\frac{1}{2}\frac{\partial^2}{\partial\xi^2}\right]\Pi(\xi,t) \\ &+ b\left[\Omega^{-1/2}\frac{\partial}{\partial\xi} + \Omega^{-1}\frac{1}{2}\frac{\partial^2}{\partial\xi^2}\right] \\ &\times[(\Omega x(t) + \Omega^{1/2}\xi)\Pi(\xi,t)] \\ &+ c\int d\eta\left\{\left[\Omega^{-1/2}\frac{\partial}{\partial\xi} + \Omega^{-1}\frac{1}{2}\frac{\partial^2}{\partial\xi^2}\right]\right. \\ &\left.\times[(\Omega x(t-\tau) + \Omega^{1/2}\eta)\Pi(\xi,t;\eta,t-\tau)]\right\},\end{aligned} \tag{7}$$

where we introduce η by writing $n(t-\tau)/\Omega = x(t-\tau)$

$+\eta/\Omega^{1/2}$, and where we have ignored higher-order terms. Anticipating that we will take the limit of large systems eventually and that trajectories at which $n(t)=0$ at any time will not contribute in this limit we have ignored the factor $\Theta(n)$ in the last term of Eq. (5). This is common in the context of the van Kampen expansion, which is usually unable to capture features such as absorbing states at the boundaries of configuration space. Collecting terms of order $\Omega^{1/2}$ in Eq. (7) one finds

$$-\Omega^{1/2}\frac{\partial\Pi}{\partial\xi}\dot{x}=-a\Omega^{1/2}\frac{\partial\Pi}{\partial\xi}+bx(t)\Omega^{1/2}\frac{\partial\Pi}{\partial\xi}+cx(t-\tau)\Omega^{1/2}\frac{\partial\Pi}{\partial\xi} \tag{8}$$

in the lowest order of the van Kampen expansion. We have written Π here as a shorthand for $\Pi(\xi,t)$ and used the identity $\Pi(\xi,t)=\int d\eta\Pi(\xi,t;\eta,t-\tau)$. From Eq. (8) one has

$$\dot{x}(t)=a-bx(t)-cx(t-\tau), \tag{9}$$

i.e., one recovers the above deterministic Eq. (4). To next-leading order [i.e., collecting $\mathcal{O}(\Omega^0)$ terms] one finds

$$\begin{aligned}\partial_t\Pi(\xi,t)=&\frac{1}{2}a\frac{\partial^2}{\partial\xi^2}\Pi(\xi,t)+b\frac{\partial}{\partial\xi}[\xi\Pi(\xi,t)]+\frac{1}{2}bx(t)\frac{\partial^2}{\partial\xi^2}\Pi(\xi,t)\\&+c\int d\eta\left\{\frac{\partial}{\partial\xi}[\eta\Pi(\xi,t;\eta,t-\tau)]\right.\\&\left.+\frac{1}{2}x(t-\tau)\frac{\partial^2}{\partial\xi^2}\Pi(\xi,t;\eta,t-\tau)\right\}.\end{aligned} \tag{10}$$

Integrating out η in the last term one then has

$$\begin{aligned}\partial_t\Pi(\xi,t)=&\frac{1}{2}a\frac{\partial^2}{\partial\xi^2}\Pi(\xi,t)+b\frac{\partial}{\partial\xi}[\xi\Pi(\xi,t)]+\frac{1}{2}bx(t)\frac{\partial^2}{\partial\xi^2}\Pi(\xi,t)\\&+c\int d\eta\left\{\frac{\partial}{\partial\xi}[\eta\Pi(\xi,t;\eta,t-\tau)]\right\}\\&+c\frac{1}{2}x(t-\tau)\frac{\partial^2}{\partial\xi^2}\Pi(\xi,t).\end{aligned} \tag{11}$$

At asymptotic times t the mean-field trajectory approaches its fixed point (our analysis is restricted to the stable phase, for similar studies in nondelay systems with a limit cycle see [11]). We therefore replace $x(t)$ and $x(t-\tau)$ in Eq. (11) by $x^*=a/(b+c)$. With this substitution Eq. (11) then describes a delayed Langevin dynamics of the form

$$\dot{\xi}=-b\xi(t)-c\xi(t-\tau)+\zeta(t), \tag{12}$$

where $\zeta(t)$ is Gaussian white noise of zero mean and with variance

$$\langle\zeta(t)\zeta(t')\rangle=(a+bx^*+cx^*)\delta(t-t'). \tag{13}$$

The equivalence of delay Langevin equations of type (12) and generalized Fokker-Planck equations of type (11) is established by Frank *et al.* [12,13] using the "method of steps." These considerations are based on the formulation of the one-dimensional delay Langevin Eq. (12) as a multidimensional Langevin equation on the time interval $[0,\tau]$. This multidimensional Langevin equation involves couplings between the different coordinates, where each coordinate represents a time interval of length τ of the original process, and where the coupling reflects the delay. The multivariate problem itself is then local in time and does not have delay terms. A corresponding multivariate Fokker-Planck equation can hence be derived straightforwardly. Equations of type (11) are then obtained using a suitable projection. We will here not discuss the further details, but refer to [12,13]. Second it is worth pointing out that a Langevin description can be obtained without explicitly assuming that the deterministic system has reached a fixed point. We have made this assumption early on in the calculation mainly to keep the mathematics simple. A more elaborate approach is taken for example in [11], where we have studied stochastic processes which in the deterministic limit tend to limit cycles. A Langevin equation can then still be derived, but the subsequent analysis is more intricate as drift and diffusion coefficients may come out as time dependent. We will not discuss such cases further in this paper, as we are mostly interested in systems which approach a fixed point in the deterministic description.

Equation (12) is linear and can be solved in Fourier space (similar approaches to delay Langevin equations have been discussed in [14]). In particular one has

$$[i\omega+b+ce^{-i\omega\tau}]\tilde{\xi}(\omega)=\tilde{\zeta}(\omega), \tag{14}$$

where $\tilde{\xi}(\omega)$ and $\tilde{\zeta}(\omega)$ indicate the Fourier transforms of $\xi(t)$ and $\zeta(t)$, respectively. From Eq. (14) one directly reads off the power spectrum of $\xi(t)$ and finds

$$\begin{aligned}P(\omega)\equiv&\langle|\tilde{\xi}(\omega)|^2\rangle\\=&\frac{a+bx^*+cx^*}{[b+c\cos(\omega\tau)]^2+[\omega-c\sin(\omega\tau)]^2}\\=&\frac{2a}{[b+c\cos(\omega\tau)]^2+[\omega-c\sin(\omega\tau)]^2}.\end{aligned} \tag{15}$$

We have here used the relation $\langle\tilde{\zeta}(\omega)\tilde{\zeta}(\omega')\rangle=(a+bx^*+cx^*)\delta(\omega+\omega')$, where $\langle\cdots\rangle$ denotes an average over the stochastic process described by Fokker-Planck Eq. (11), or equivalently over realizations of Langevin Eq. (12).

C. Test against simulations

Microscopic simulations of the processes defined by reactions (1)–(3) can be carried out using the algorithm originally proposed by Gillespie [16], suitably modified to take into account the delayed reactions. Details of such modified Gillespie schemes have been discussed for example in [17], but for completeness we reiterate them here. Essentially the simulations follow that of the classic Gillespie algorithm [16], and whenever a delayed reaction is triggered it is added to a list of delay reactions to be executed at a later time (τ units of time after the reaction is initiated). This list is constantly updated, and delay reactions are executed (and removed from the list) in a manner consistent with the probabilistic description in terms of the above master equation. Specifically the simulations of our toy model dynamics proceed according to the following algorithm:

(1) Initialize. Set model parameters a, b, and c and the system size Ω. Set the initial number n of molecules X, and set $t=0$. Set list of scheduled delay reaction to an empty list.

(2) Calculate the propensity functions $a_1=a\Omega$, $a_2=bn$, and $a_3=cn$.

(3) Compute $a_0=a_1+a_2+a_3$.

(4) Generate an independent random number r from a uniform distribution over (0,1], and set $\Delta=-\ln(r)/a_0$.

(5) If there is a delayed reaction scheduled to occur during the interval $[t,t+\Delta)$ then

(a) Identify next delayed reaction scheduled, and, provided $n>0$ execute it, i.e., reduce n by one. If $n=0$ before the reaction, then do not execute the update (otherwise n would go negative). In either case remove the reaction from the list of scheduled reactions.

(b) Update t to the time for which this reaction was scheduled.

(c) Go to 2.

(6) If there is no delayed reaction scheduled for the interval $[t,t+\Delta)$ then

(a) Generate an independent random number r from a uniform distribution over (0,1], and find $\mu\in\{1,2,3\}$ such that

$$\sum_{k=1}^{\mu-1} a_k < r' \leq \sum_{k=1}^{\mu} a_k.$$

(b) If $\mu=1$ or $\mu=2$ and then execute the corresponding reaction (not a delay reaction), and increment t by Δ. Go to 2.

(c) If $\mu=3$, schedule a reaction of type $X+C\Rightarrow C$ to be executed at later time $t+\tau$, i.e., amend list of scheduled reactions accordingly. Increment t by Δ, and go to 2.

Each run of the simulation generates a time series $n(t)$ from which $\xi(t)=\sqrt{N}[n(t)/\Omega-x^*]$ can be obtained (after a suitable equilibration time), where x^* is the asymptotic fixed-point value of the deterministic dynamics given by Eq. (9), i.e., $x^*=a/(b+c)$, or equivalently the long-time average of $n(t)/\Omega$. From these time series $\xi(t)$ a numerical measurement of the power spectrum $P(\omega)$ is obtained by subsequent Fourier transform, and finally results are averaged over a sufficiently large number of independent runs.

Results of stochastic simulations of this system are shown in Figs. 1 and 2. The first figure depicts a single run of the stochastic system and shows that coherent oscillations are sustained in parameter regimes in which the deterministic equations approach a fixed point. The mechanism by which these oscillations are generated is the following: the deterministic system approaches its fixed point in an oscillatory manner (i.e., the Jacobian at the fixed point has eigenvalues with nonzero imaginary parts), and the stochasticity of the finite system results in persistent perturbations away from this fixed point, so that both features together result in an overall oscillatory effect. This has been seen in a variety of different systems [7,9], but it is worth pointing out that in nondelay systems a minimum of two dimensions is necessary to allow for a complex eigenvalue of the Jacobian. In delay systems one degree of freedom is sufficient [5], so that even the one-dimensional toy model discussed in this section

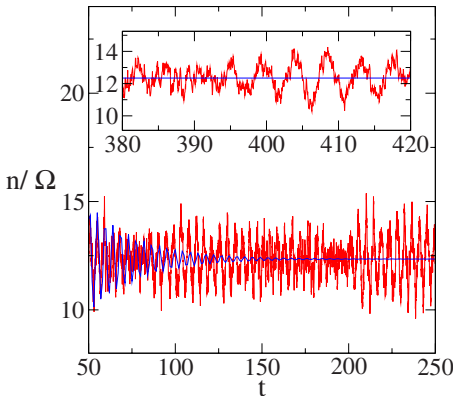

FIG. 1. (Color online). Dynamics of the toy model at $a=100$, $b=4.1$, $c=4$, and $\tau=2$. The dark line with decaying oscillations in the main panel shows the behavior of the concentration of X molecules in the deterministic system, Eq. (9) [15]. The noisy line with persistent oscillations represents one simulation run of the stochastic dynamics at $\Omega=100$. The inset shows a zoom at large times in the equilibrated regime, the horizontal line is the mean-field fixed point, and the stochastic system shows persistent cycles.

is able to produce demographic oscillations about the mean-field fixed point. Figure 2 demonstrates that the analytically obtained spectrum, Eq. (15), agrees very well with simulations; we attribute the remaining small discrepancies to finite-size or equilibration effects. A similar figure was obtained by different methods (based on generating functions) in [5].

III. SIMPLE MODEL OF GENE REGULATION

The second system we will be studying is a simple model of gene regulation. We here chose a system that represents one of the most common motifs in gene regulatory networks, namely, a model of a single gene-protein synthesis with negative delayed feedback [5]. Specifically we address a model previously discussed in [4], describing the coupled

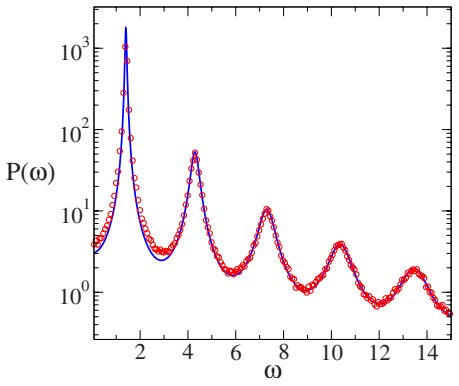

FIG. 2. (Color online) Power spectrum $P(\omega)=\langle|\tilde{\xi}(\omega)|^2\rangle$ of the fluctuations about the mean-field fixed point of the toy model. Parameters a, b, and c are as in Fig. 1. The line shows the analytically obtained spectrum of Eq. (15), symbols represent results from simulations at $\Omega=200$ (averaged over 341 samples, measurements start at $t=100$ to allow for some equilibration period).

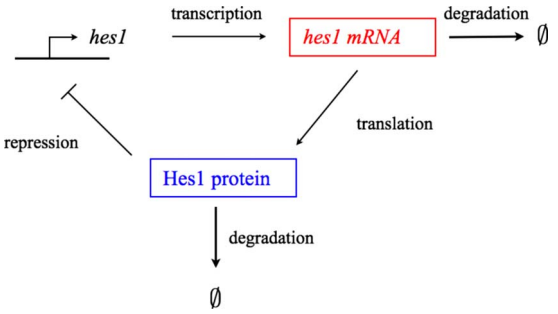

FIG. 3. (Color online) Schematic illustration of the Hes1 regulatory system. See [22] for a similar picture. The first process (transcription) includes the elongation, splicing, processing and export from the nucleus of primary gene transcript. The synthesis of Hes1 protein occurs by translation of *hes*1 mRNA. Any translational delay is here absorbed into the transcriptional delay time τ. The Hes1 protein finally represses the transcription through the binding to the promoter. Both, the mRNA and the protein are subject to degradation.

time behavior of the expression levels of so-called *hes*1 messenger RNA (mRNA) and Hes1 protein. Following the notation in biology literature we will italicize and use lower case when referring to mRNA, and will use nonitalicized font with the first letter in upper case when referring to the protein [18,4]. Hes1 here is a Notch-signaling molecule, where the so-called Delta-Notch signaling process is a mechanism for cell-cell communication and underlies cell differentiation for example in vertebrates [3,19,20]. Cycles with a time period of approximately two hours have been reported for the concentrations of *hes*1 mRNA and Hes1 protein in mice [4,18]. These oscillatory processes, also referred to as the somite-segmentation clock [20], are linked to the formation of somites, i.e., the emergence of blocks of cells which determine the future positions of skeletal muscles or vertebrae [3,20]. Spatial segmentation in the body here can be understood as a reflexion of temporal oscillations in gene expression [3,19–21]. The underlying molecular mechanism producing the oscillations of mRNA and protein are hence of great interest, and several mathematical models have been proposed, among them [4,5,21–23]. See also [2,24] for stochastic effects in models of gene regulation.

We will not discuss the details of the biochemical mechanisms in this paper, but will only present a brief abstraction of the reactions necessary to define the mathematical model we will study here. Further details on modeling genetic circuits can be found in [25] or in similar textbooks. In essence the model describes the concentrations and interactions of two types of substances, *hes*1 mRNA and Hes1 protein, as illustrated in Fig. 3. mRNA molecules are produced by transcription of DNA. This involves several biochemical processes (e.g., elongation and splicing) which we will neglect in our description. The rate at which mRNA molecules are produced depends on the concentration of protein through a negative feedback mechanism. Transcription is here associated with a delay time τ, so that it is the protein concentration at time $t-\tau$ which affects the production rate of mRNA at time t. This will be explained further below. mRNA is also subject to degradation (i.e., removal from the system) at a constant rate μ_m. In a process subsequent to transcription *hes*1 mRNA molecules are translated into Hes1 protein (the mRNA molecule is not used up in this process). Translation may involve another delay, which for simplicity can be absorbed into the transcriptional delay [5]. Protein molecules finally are subject to a degradation process at rate μ_p. Crucially, a negative feedback is induced by a repressatory effect of Hes1 protein on the transcription process. Hes1 dimers may bind to the relevant promoter regions in the DNA, and reduce the transcription rate at which mRNA is generated (following [4] dimerization is not discussed as an explicit step in our work, but see [5] for models taking this into account). The transcriptional repressor Hes1 thus negatively affects its own expression [18]. Mathematically this is modeled by a transcription rate which depends on the concentration of protein via a decreasing function, as we will now explain.

We will focus on the model proposed in [4,22]. In its deterministic form it is given by the differential equations

$$\frac{d}{dt}M(t) = \alpha_m f(P(t-\tau)) - \mu_m M(t), \tag{16}$$

$$\frac{d}{dt}P(t) = \alpha_p M(t) - \mu_p P(t). \tag{17}$$

$M(t)$ here labels the concentration of *hes*1 mRNA, and $P(t)$ that of the Hes1 protein. μ_m and μ_p are degradation rates for the mRNA and the protein, respectively, and α_m is the mRNA transcription rate in the absence of protein expression [$f(P)$ is still to be defined, but we will have $f(0)=1$]. α_p is the translation rate. $f(P)$ finally is a monotonically decreasing Hill function representing the suppression of mRNA production through the binding of Hes1 protein dimers into the promotion region. In the model it takes the form [4,25]

$$f(P(t)) = \frac{1}{1+[P(t)/P_0]^h} \tag{18}$$

with h the so-called Hill coefficient. P_0 is the concentration of protein at which $f(P=P_0)=1/2$. Equations (16) and (17) are the deterministic abstraction of an underlying microscopic stochastic model defined by the following four reactions [4]:

$$M \xrightarrow{\mu_m} \emptyset, \tag{19}$$

$$P \xrightarrow{\mu_p} \emptyset, \tag{20}$$

$$M \xrightarrow{\alpha_p} M + P, \tag{21}$$

$$\emptyset \xRightarrow{\alpha_m f} M. \tag{22}$$

These dynamics may be described by a stochastic process for the numbers n_m of mRNA molecules and n_p of protein molecules in the system. For later convenience we will write $\mathbf{n}=(n_m,n_p)$. The first two reactions here describe the degra-

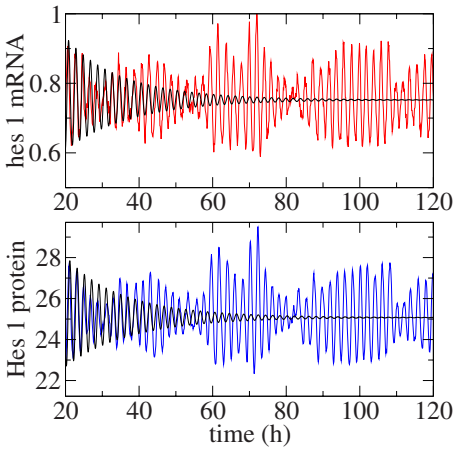

FIG. 4. (Color online) Time series of the concentrations of mRNA and protein concentrations respectively. Solid curves, decaying toward a fixed point, are from a numerical integration of the deterministic dynamics, Eqs. (16) and (17). Curves with persistent oscillations represent a single run of the stochastic dynamics at $\Omega=1000$. Model parameters are as some of the examples in [4]: $P_0=10$, $h=4.1$, $\tau=18.7$, $\alpha_m=\alpha_p=1$, and $\mu_m=\mu_p=0.03$. Units of α_p, α_m and of μ_m, μ_p are min^{-1}, τ is measured in minutes [4].

dation of mRNA and protein, respectively. The third reaction captures the translation of mRNA into protein. The last reaction finally corresponds to the production of *hes*1 mRNA via transcription. Note that DNA is not part of the dynamical model (its concentration is constant), which is why M appears to be produced out of the void in the fourth reaction. In absence of protein ($n_p=0$) this reaction occurs at a rate α_m, but is suppressed by the Hes1 protein, and in total the rate at which mRNA is produced at time t is hence $\alpha_m f(n_p(t-\tau)/\Omega)$, where Ω is a measure of the system size. The time lag τ in the argument of f models the delayed repression of *hes*1 mRNA production by the protein. The rate of production of mRNA at time t is therefore negatively regulated by the concentration of protein at time $t-\tau$. A typical run of the stochastic system is shown and compared with the deterministic system in Fig. 4. Model parameters are chosen as in [4], in particular delay times estimated from real-world experiments are typically in the range of 15–20 min, sometimes ranging up to 30 min [4,21,22]. As seen in the figure, the deterministic system approaches a stable fixed point asymptotically, with a complex eigenvalue as indicated by the decaying modulations. The stochastic system at finite size remains in an oscillatory state as previously observed in [4]. We will now proceed to characterize these oscillations analytically, applying the formalism developed in the previous section.

The master equation describing processes (19)–(22) then takes the form

$$\frac{d}{dt}P(\mathbf{n};t) = \mu_m(\mathbb{E}_M - 1)[n_m P(\mathbf{n};t)] + \mu_p(\mathbb{E}_P - 1)[n_p P(\mathbf{n};t)] + \alpha_p(\mathbb{E}_P^{-1} - 1)[n_m P(\mathbf{n};t)] + \alpha_m \Omega \sum_{\mathbf{n}'} f(n_p'/\Omega) \times (\mathbb{E}_M^{-1} - 1)[P(\mathbf{n},t;\mathbf{n}',t-\tau)], \tag{23}$$

where $P(\mathbf{n},t)$ is the probability of finding the system in state $\mathbf{n}$ at time t, and $P(\mathbf{n},t;\mathbf{n}',t')$ is the probability for the system to be in state $\mathbf{n}$ at t and in state $\mathbf{n}'$ at time t'. $\mathbb{E}_M$ and $\mathbb{E}_P$ are raising operators acting on functions of n_m, n_p via $\mathbb{E}_M g(n_m,n_p)=g(n_m+1,n_p)$ and $\mathbb{E}_P g(n_m,n_p)=g(n_m,n_p+1)$. In the case of two-time quantities, e.g., $P(\mathbf{n},t;\mathbf{n}',t')$, the raising or lowering occurs in the first (later-time) argument $\mathbf{n}$. Note that all terms on the right-hand side of Eq. (23) are of order Ω. This overall factor could in principle be absorbed into a re-scaling of time, even though we will not do so here.

The further analysis proceeds along the lines of what was discussed for the toy model, and we will not report all intermediate steps in all detail. First one writes

$$\frac{n_m(t)}{\Omega} = M(t) + \frac{\xi_m(t)}{\Omega^{1/2}}, \tag{24}$$

$$\frac{n_p(t)}{\Omega} = P(t) + \frac{\xi_p(t)}{\Omega^{1/2}}, \tag{25}$$

and then systematically expands the above master equation in powers of $\Omega^{-1/2}$. To lowest order one recovers mean-field Eqs. (16) and (17), and in next-to-leading order one finds Langevin equations for the fluctuations about the mean-field trajectory. In the fixed-point regime of the mean-field dynamics (i.e., at large times t) these equations read

$$\dot{\xi}_m(t) = -\mu_m \xi_m(t) + \alpha_m f'(P^*)\xi_p(t-\tau) + \zeta_m(t), \tag{26}$$

$$\dot{\xi}_p(t) = \alpha_p \xi_m(t) - \mu_p \xi_p(t) + \zeta_p(t), \tag{27}$$

where (M^*,P^*) is the mean-field fixed point, $f'(P)=df(P)/dP=-hP_0^{-1}(1+P/P_0)^{-(h+1)}$. $\zeta_m(t)$ and $\zeta_p(t)$ are Gaussian noise terms of zero mean and in the limit of large t,t' (when the mean-field trajectory has reached its fixed point) they have covariances

$$\langle \zeta_m(t)\zeta_m(t')\rangle = \delta(t-t')[\mu_m M^* + \alpha_m f(P^*)], \tag{28}$$

$$\langle \zeta_p(t)\zeta_p(t')\rangle = \delta(t-t')[\mu_p P^* + \alpha_p M^*], \tag{29}$$

$$\langle \zeta_m(t)\zeta_m(t')\rangle = 0. \tag{30}$$

Inverting in Fourier space one then finds after some algebraic manipulations

$$\langle |\tilde{\xi}_m(\omega)|^2\rangle = \frac{(\omega^2+\mu_p^2)[\alpha_m f(P^*)+\mu_m M^*] + [\alpha_m f'(P^*)]^2(\alpha_p M^* + \mu_p P^*)}{[-\omega^2 + \mu_m\mu_p - \alpha_m\alpha_p f'(P^*)\cos(\omega\tau)]^2 + [(\mu_m+\mu_p)\omega + \alpha_m\alpha_p f'(P^*)\sin(\omega\tau)]^2}, \tag{31}$$

$$\langle|\tilde{\xi}_p(\omega)|^2\rangle = \frac{\alpha_p^2[\alpha_m f(P^*) + \mu_m M^*] + (\omega^2 + \mu_m^2)(\alpha_p M^* + \mu_p P^*)}{[-\omega^2 + \mu_m\mu_p - \alpha_m\alpha_p f'(P^*)\cos(\omega\tau)]^2 + [(\mu_m + \mu_p)\omega + \alpha_m\alpha_p f'(P^*)\sin(\omega\tau)]^2}. \tag{32}$$

The resulting power spectra for a given set of parameters used, e.g., in [4], are shown in Fig. 5, and as seen in the figure direct simulations based on a modified Gillespie algorithm, similar to the one described in the section of the toy model, agree well with the theoretical predictions. Remaining discrepancies are presumably due to finite-size and equilibration effects. We note that the peak of the spectra shown in Fig. 5 occurs at an angular frequency of roughly $\omega \approx 0.05$ in units of 1/min, corresponding to a time period of $T \approx 125$ min, i.e., approximately two hours and therefore close to the results from experiments reported, e.g., by Hirata *et al.* [18]. This agreement is of course a consequence of the specific choice of parameters, but it demonstrates that the oscillatory behavior of mRNA and protein concentrations can be observed in the stochastic models at values of the model parameters, for which the deterministic system does not sustain oscillations. This enlarges the range of permissible model parameters, and our analysis as well as that of [4,5] may therefore provide a starting point for further more elaborate stochastic models of the somite-segmentation clock. One should here stress that the agreement with data from real-world experiments is far from perfect however. The amplitude of oscillations in the protein levels for example is much lower than those observed in experiments [18], so that further improvements of the model are required, potentially based on the analytical method discussed in the present work. Our theoretical analysis may also be used in order to test the robustness of the model without performing costly stochastic simulations throughout a large range of parameter values. In Fig. 6 for example we depict the frequency at which the spectrum of mRNA fluctuations has its maximum in dependence on the time delay τ. This data is to be compared with Fig. 8 of [4], where quantitatively very similar results were obtained from actual stochastic simulations. Care needs to be taken though in interpreting the maximum of the power spectra as the frequency at which the system oscillates. The peaks in the spectra can be broad and hence several modes contribute. Also, of course, finite systems at small sizes may show deviations from the curves obtained from the system-size expansion, as the latter curves, even though they represent the next-to-leading order in $\Omega^{-1/2}$, can only be expected to be accurate at large system sizes. Still, Fig. 6 provides a theoretical confirmation of the findings of [4], and suggests that two-hour cycles are found at values of the delay time of about $\tau \approx 5-10$ min at $h=3$ and at slightly larger values of $\tau \approx 15-17$ min at $h=4$. It should be noted however that the variation in the observed time period is rather small as τ and h are varied in Fig. 6 so that other parameter ranges are not ruled out by the experimentally observed 2 hr period. Still, with theoretical approaches available along the lines discussed in the present paper or along those of [5], the most efficient way of identifying parameter values compatible with measurements in real-world experiments might be to use analytical expressions of the type presented in Eqs. (31) and (32) first to narrow down the range of possible parameter values. Such closed-form expressions can be evaluated relatively quickly and this preselection of model parameters based on analytical results is hence much less costly than performing parameter scans in stochastic simulations. Once suitable parameters

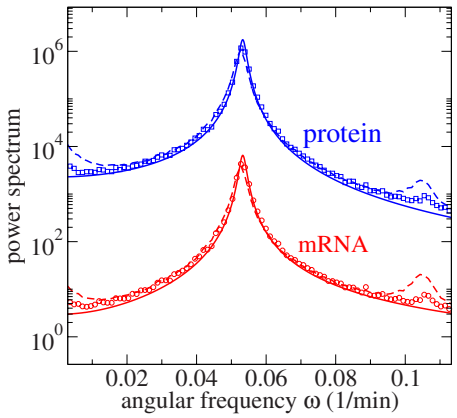

FIG. 5. (Color online) Power spectra of the fluctuations of mRNA and protein concentrations, respectively. Solid lines are the analytical expressions of Eqs. (31) and (32). Markers are from simulations at system size $\Omega=5000$, and dashed lines from simulations at $\Omega=500$. Averages over more than 700 independent samples are taken in the simulations. Measurements start at $t=3000$ min in order to allow for equilibration. Model parameters are $P_0=10$, $h=4.1$, $\tau=18.7$, $\alpha_m=\alpha_p=1$, and $\mu_m=\mu_p=0.03$ [4].

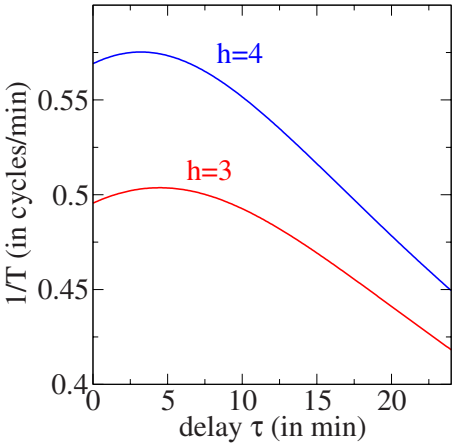

FIG. 6. (Color online) Frequency $1/T$ at which the power spectrum of mRNA fluctuations has its maximum. Results are from Eq. (31), evaluated at the fixed point of the mean-field dynamics. The latter is obtained by integrating the deterministic Eqs. (16) and (17) using an Euler-forward scheme ($\Delta t=0.1$). Model parameters are $P_0=100$, $\alpha_m=\alpha_p=1$, and $\mu_m=\mu_p=0.03$. The figure is to be compared with Fig. 8 of [4].

have been identified from the theory, subsequent stochastic simulations in a much smaller range of parameters can then be carried out to confirm whether or not the experimentally observed behavior is indeed found in the stochastic system.

IV. DISCUSSION AND CONCLUDING REMARKS

In summary we have successfully extended recent analyses of the effects of intrinsic noise to chemical systems with delay. In particular we have shown that the picture of coherent oscillations, induced by the discreteness of the microscopic dynamics, applies to systems with delay as well. Stochastic self-sustained oscillations can here be found in *finite* systems at choices of the model parameters for which the *infinite* system, described by deterministic mean-field equations, does not exhibit cycles. This observation has important implications for the modeling of oscillatory biological systems with delay, as reaction rates are often not known experimentally, but are instead tuned in theoretical approaches, in order to ensure that simple model systems reproduce the experimentally observed oscillatory behavior. Our analysis shows that confining model parameters to ranges in which the deterministic model shows oscillations may be unnecessarily restrictive, as the stochastic dynamics at finite system size may well exhibit oscillatory behavior outside these ranges of the model parameters.

While this observation as such has been made previously, e.g., in [4,5], the main contribution of the present work is the extension of van Kampen expansion techniques to the case of stochastic delay systems, and based on the resulting Langevin equation the analytical calculation of the power spectra of fluctuations about mean-field fixed points in delay systems. To our knowledge this has not been attempted before, even though previous work on Kramers-Moyal expansions in delay systems can be found in [26]. Our approach is here complementary to that of [5], which have used generating function techniques to study master equations of delay systems, and to computer power spectra in good agreement with simulations, but which, in our understanding, have not carried out a systematic expansion in the inverse system size. Since the two approaches each rely on a series of approximations and simplifications analytical expressions derived in the two formalisms may not be fully equivalent. The spectra derived in our work are however in excellent agreement with simulations, confirming the validity of the procedure carried out here. We have here first developed the general theory in the context of a simple one-dimensional system. Generalization to more complex models with a higher number of degrees of freedom is possible however, and as a further example we have addressed a basic model of gene regulation. In particular we have studied a delay system describing the regulatory processes underlying the expression of *hes*1 mRNA and Hes1 protein. Simulational work has here recently been reported by Barrio *et al.* [4], and our work complements these mostly computational studies by an analytical computation of the spectra of the observed oscillations in mRNA and protein expression levels. See again also [5] for related models. Based on our analytical results further characterization of the behavior of the model is possible, without the need to perform computationally expensive simulations in a wide range of parameters. This could serve as a starting point for the development of more realistic stochastic models of the somite-segmentation clock, with a view of reducing the remaining discrepancies with experimental data [27]. Our theoretical approach is furthermore applicable more generally, and can be expected to be useful for the theoretical understanding of the behavior more intricate stochastic delay systems.

ACKNOWLEDGMENTS

T.G. acknowledges support from the RCUK (RCUK Reference No. EP/E500048/1), and would like to thank R. Schlicht for useful discussions on stochastic delay systems, and A. J. McKane and R. P. Boland for earlier collaboration on systems with intrinsic fluctuations.

[1] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (Perseus Book Publishing, Cambridge, Massachusetts, 1994).
[2] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Rev. Genet. **6**, 451 (2005).
[3] R. Schlicht and G. Winkler, J. Math. Biol. **57**, 613 (2008).
[4] M. Barrio, K. Burrage, A. Leier, and T. Tian, PLOS Comput. Biol. **2**, e117 (2006).
[5] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, Proc. Natl. Acad. Sci. U.S.A. **102**, 14593 (2005).
[6] R. Nisbet and W. Gurney, *Modelling Fluctuating Populations* (Wiley, New York, 1982).
[7] A. J. McKane and T. J. Newman, Phys. Rev. Lett. **94**, 218102 (2005).
[8] A. J. McKane, J. D. Nagy, T. J. Newman, and M. O. Stefanini, J. Stat. Phys. **128**, 165 (2007).
[9] T. Reichenbach, M. Mobilia, and E. Frey, Phys. Rev. E **74**, 051907 (2006); M. Pineda-Krch, H. J. Blok, U. Dieckmann, and M. Doebeli, Oikos **116**, 53 (2007); R. Kuske, L. F. Gordillo, and P. Greenwood, J. Theor. Biol. **245**, 459 (2007); D. Alonso, A. J. McKane, and M. Pascual, J. R. Soc., Interface **4**, 575 (2007); M. Simoes, M. M. Telo da Gama, and A. Nunes, *ibid.* **5**, 555 (2008); F. Di Patti and D. Fanelli, J. Stat. Mech. (2009) P01004.
[10] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 2nd ed. (Elsevier Science, New York, 1997).
[11] R. P. Boland, T. Galla, and A. J. McKane, J. Stat. Mech. (2008) P09001; Phys. Rev. E **79**, 051131 (2009).
[12] T. D. Frank, P. J. Beek, and R. Friedrich, Phys. Rev. E **68**, 021912 (2003).
[13] T. D. Frank, Phys. Rev. E **66**, 011914 (2002).
[14] S. Guillouzic, I. L'Heureux, and A. Longtin, Phys. Rev. E **59**, 3970 (1999).
[15] The numerical integration of the deterministic equation is performed with a simple Euler-forward scheme. The initial oscil-

lation amplitude of the mean-field trajectory may here depend slightly on the chosen time-stepping. We have typically used $\Delta t=0.1$ or smaller. The asymptotic fixed point is not affected.
[16] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976); J. Phys. Chem. **81**, 2340 (1977).
[17] D. F. Anderson, J. Chem. Phys. **127**, 214107 (2007).
[18] H. Hirata, S. Yoshiura, T. Ohtsuka, Y. Bessho, T. Harada, K. Yoshikawa, and R. Kageyama, Science **298**, 840 (2002).
[19] Y.-J. Jiang, B. L. Aerne, L. Smithers, C. Haddon, D. Ish-Horowicz and J. Lewis, Nature (London) **408**, 475 (2000).
[20] Y. Saga and H. Takeda, Nat. Rev. Genet. **2**, 835 (2001).
[21] J. Lewis, Curr. Biol. **13**, 1398 (2003).
[22] N. A. M. Monk, Curr. Biol. **13**, 1409 (2003).
[23] M. H. Jensen, K. Sneppen, and G. Tiana, FEBS Lett. **541**, 176 (2003).
[24] M. Scott, B. Ingalls, and M. Kaern, Chaos **16**, 026107 (2006).
[25] U. Alon, *An Introduction to Systems Biology—Design Principles of Biological Circuits* (Chapman and Hall, London, 2007).
[26] T. D. Frank, Phys. Lett. A **360**, 552 (2007).
[27] H. Momiji and N. A. M. Monk, J. Theor. Biol. **254**, 784 (2008).