# Exact time-dependent solutions for a self-regulating gene

A. F. Ramos*
*Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Avenida Arlindo Béttio 1000, 03828-00, São Paulo, SP, Brazil*

G. C. P. Innocentini† and J. E. M. Hornos‡
*Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, CEP 13560-970, São Carlos, SP, Brazil*
(Received 15 October 2010; published 27 June 2011)

The exact time-dependent solution for the stochastic equations governing the behavior of a binary self-regulating gene is presented. Using the generating function technique to rephrase the master equations in terms of partial differential equations, we show that the model is totally integrable and the analytical solutions are the celebrated confluent Heun functions. Self-regulation plays a major role in the control of gene expression, and it is remarkable that such a microscopic model is completely integrable in terms of well-known complex functions.

DOI: 10.1103/PhysRevE.83.062902 PACS number(s): 87.16.Yc, 87.10.Ed, 87.10.Mn

Understanding the role of randomness that occurs inside the cell and its apparent contrast with, for example, precise formation of temporal and spatial patterns of gene expression during biological processes such as embryogeneses is a central open problem of the postgenomic era [1,2]. The significance of statistical fluctuations in the cellular environment was recognized by Delbruck since the infancy of molecular biology [3]. The fluctuations are caused by the presence of a frequently small number of reacting molecules in a cell. The mechanism of gene expression, which involves a web of chemical interactions, is an example where stochasticity plays an important role, as recently verified in experiments using fluorescent protein techniques [4,5].

Simulations of the gene regulation process in *E. coli*, $\lambda$ phage, and other simple systems have been performed to explain the experimental data [6]. Also, the description of the concentration dynamics of gene products has been explored by using nonlinear differential equations, followed by the introduction of noise by the Langevin mechanism [7,8]. Although the mentioned efforts have been successfully employed in the phenomenological control of the experiments, the understanding of the biological and biophysical meaning of noise in gene networks requires, as usual in science and particularly in physics, a soluble model. Therefore, we adopt here a different strategy, searching for exactly soluble models to be used as building blocks to decompose complex networks. Specifically, we target microscopic soluble models to understand the basic stochastic properties of simple gene expression mechanisms [9–15]. The full knowledge of the stochastic properties encoded in probability distributions, instead of mean values and noise, is a requirement since several systems show many stable configurations that cannot be explained by the first moments [16].

Recently, a binary stochastic model for gene expression was proposed [17], inspired by the spin-boson models used in many body theory [9,10,12]. Exact solutions for the stationary regime of the single gene spin-boson model were already obtained [17,18]. The underlying symmetries responsible for the integrability of this model were identified [19,20], and several applications of the model have been discussed in the literature [21,22]. In this Brief Report we solve completely the model exhibiting analytical solutions for the time-dependent stochastic process in terms of Heun functions [23–26].

We assume that the transcription and translation are combined, resulting in an effective process that is appropriate for prokaryotic cells, where translation follows, almost simultaneously, transcription. We focus our attention on the self-regulation process, not only because of the importance of this phenomenon [5] but also because it is a challenge due to its mathematic complexity.

Our stochastic variable $n$ is the number of free proteins expressed by a two state gene in the cell. The probability to find the operator site free and the gene in full operation is $\alpha_n(t)$, while $\beta_n(t)$ is the probability to find the gene inactivated or in a basal level of expression. The model is based on the coupling of two Markov processes where a parameter $\rho$ describes the protein degradation and $k$ and $\chi k$ ($\chi < 1$) denote the production rate in the activated and repressed modes, respectively. The master equations are coupled by two parameters: $h$ describes the binding of the regulatory protein in the operator site and $f$ is the release. The equations for the probabilities are

$$\frac{d\alpha_n}{dt} = k[\alpha_{n-1} - \alpha_n] + \rho[(n+1)\alpha_{n+1} - n\alpha_n] - nh\alpha_n + f\beta_n, \quad (1)$$

$$\frac{d\beta_n}{dt} = k\chi[\beta_{n-1} - \beta_n] + \rho[(n+1)\beta_{n+1} - n\beta_n] + nh\alpha_n - f\beta_n. \quad (2)$$

Self-regulation is characterized by the linear dependence of the binding rate on $n$. Large values of $n$ enhance the binding rate, while the unbinding is independent of the protein number.

Instead of equations involving time derivatives and differences as in the master equations, we rephrase the problem in terms of partial equations by introducing the functions $\alpha(z,t) = \sum_{n=0}^{\infty} \alpha_n(t) z^n$ and $\beta(z,t) = \sum_{n=0}^{\infty} \beta_n(t) z^n$. The

*alex.ramos@usp.br
†innocentini@ursa.ifsc.usp.br
‡hornos@ifsc.usp.br

 

corresponding partial differential equations coupling $\alpha(z,t)$ and $\beta(z,t)$ are

$$\frac{\partial\alpha}{\partial t} = (z-1)\left[k\alpha - \rho\frac{\partial\alpha}{\partial z}\right] - hz\frac{\partial\alpha}{\partial z} + f\beta, \tag{3}$$

$$\frac{\partial\beta}{\partial t} = (z-1)\left[\chi k\beta - \rho\frac{\partial\beta}{\partial z}\right] + hz\frac{\partial\alpha}{\partial z} - f\beta, \tag{4}$$

and looking for analytical solutions, we ensure a well-behaved probability distribution.

In order to understand the topology of this system of equations it is useful to consider the variable $z$ in the complex plane. The coefficient of the derivative with respect to $z$ in Eq. (3), $-\rho(z-1)-hz$, vanishes at $z=\rho/(\rho+h)$ and is a singular point of the system. A second singularity at $z=1$ comes from Eq. (4), and at infinity we have the third singularity. One might expect the solutions to be expressed in terms of hypergeometric functions; however, the irregularity of the singular point at infinity can only be taken into account by the use of the Heun confluent functions.

Instead of the pair $\alpha(z,t)$ and $\beta(z,t)$ we introduce the total probability $\phi(z,t)=\alpha(z,t)+\beta(z,t)$, which has a straightforward biological meaning. The mean value of proteins and also the fluctuations, which are commonly measured in experiments, are directly obtained by taking derivatives of $\phi(z,t)$ with respect to $z$. The new system of equations is

$$\frac{\partial\alpha}{\partial t} = (z-1)\left[k\alpha - \rho\frac{\partial\alpha}{\partial z}\right] - hz\frac{\partial\alpha}{\partial z} + f(\phi-\alpha), \tag{5}$$

$$\frac{\partial\phi}{\partial t} = (z-1)\left[k\alpha - \chi k\alpha + \chi k\phi - \rho\frac{\partial\phi}{\partial z}\right], \tag{6}$$

where Eq. (6) is obtained by summing Eqs. (3) and (4), introducing $\phi(z,t)$, and eliminating $\beta(z,t)$.

This system of first order partial equations [Eqs. (5) and (6)] generates a second order equation for $\phi(z,t)$ by solving Eq. (6) for $\alpha(z,t)$ and replacing it in Eq. (5), and the resulting equation is shown as Eq. (1) of the supplemental material [27]. The equation is not separable in the variables $z$ and $t$; however, new variables can be found to rewrite the equations in a canonical form, namely, $\nu=\frac{z(\rho+h)-\rho}{h}$ and $\mu=\frac{z(\rho+h)-\rho}{h}\exp\left[-(\rho+h)t\right]$. The motivation for the introduction of the variable $\nu$ is to bring the singular points at $z=\rho/(\rho+h)$ and $z=1$ to $\nu=0$ and $\nu=1$, respectively. The variable $\mu$ was an ansatz aiming at separability in the new variables. Asymptotically, $t\to\infty$, the variable $\mu$ vanishes, and the partial equation is reduced to the Kummer equation, as we have discussed in [17]. At the initial state, $t=0$, the variables coincide, $\nu=\mu$. The new form of the partial equation in terms of these variables is

$$\left[\frac{\partial^2}{\partial\nu^2} + \mu P\frac{\partial^2}{\partial\nu\partial\mu} + \mu Q\frac{\partial}{\partial\mu} + R\frac{\partial}{\partial\nu} + S\right]\phi(\mu,\nu) = 0. \tag{7}$$

The coefficients of this equation have a simple linear dependence on $\mu$, as we can see inspecting by Eq. (7), and the coefficients $P$, $Q$, $R$, and $S$ are rational functions depending only on the variable $\nu$ and have poles on the singular points; for details, see [27]. The irregularity at infinity can be seen in the usual way by making the transformation $\nu\to 1/\nu$. The appearance of $\mu$ always together with the derivatives with respect to $\mu$ in the coefficients of Eq. (7) guarantees separability, making possible a solution in the form $\phi(\mu,\nu)=\mu^{\lambda}H_{\lambda}(\nu)$, where $\lambda$ is the separation constant. The equation for the function $H_{\lambda}(\nu)$ can be reduced to the celebrated confluent Heun equation,

$$\left[\frac{d^2}{d\nu^2} + (R+\lambda P)\frac{d}{d\nu} + (S+\lambda Q)\right]H_{\lambda}(\nu) = 0. \tag{8}$$

This equation has recently been exhaustively studied, and solutions of several physical phenomena, from general relativity to condensed matter, can be expressed by these functions [24,26]. It can be easily manipulated by using symbolic computational software packages such as MAPLE. The general solution is a product of a term, $(\nu-1)\exp\left[kh(\nu-1)/(\rho+h)^2\right]$, and a superposition of the two families of the Heun confluent $H_C$ functions:

$$C_j^1\, e^{-j\rho t} H_C(c,\theta^1,1-j,\delta^1,\eta^1,\nu), \tag{9}$$

$$C_j^2\, \nu^{j+1}\, e^{-(\rho+h)(j+b)t} H_C(c,j+1,\sigma^2,\delta^2,\eta^2,\nu), \tag{10}$$

where $j$ is an integer running from zero to infinity. The other parameters of the Heun functions are related to the model parameters and are shown in the supplemental material [27]. The coefficients $C_j^1$ and $C_j^2$ depend on the initial conditions. Once the generating functions are calculated, the probabilities $\alpha_n(t)$ and $\phi_n(t)$ are recovered by taking derivatives and evaluating them at $z=0$.

The biological picture emerging from the analysis of the solutions shows that the protein synthesis approaches equilibrium by combining two processes with different time scales. The distribution for the transcription-translation probability is composed by the linear combination of Heun states given in Eqs. (9) and (10). The first set of solutions characterizes the protein deactivation as a transcription factor and depends on the parameter $\rho$ describing the death rate in the master equation. The second is a function of the gene switching parameters $h$ and $f$, which control the regulation of the gene. In the beginning, the probability distribution is a linear combination of Heun states weighted by the constants $C_1^j$ and $C_2^j$. The information about the initial states is lost during the probability dynamics following the hierarchy of decaying times. If the switching time $1/T_s=b(\rho+h)$ is smaller than the protein deactivation time (fast switching gene), $1/\rho$, the second family of states decays faster than the first and the equilibrium is approached after the controlling components are already stable. Here, the probability distribution shows only one Poissonian-like peak. On the other hand, when the switching time is large, the final state can show two-peaked distributions (bimodality). This is, of course, in agreement with the symmetry analysis performed in Ref. [19] and in light of the adiabatic analysis done formerly [17]. We emphasize that our model gives an effective phenomenological description of the combined transcription-translation processes, and its parameters are selected to reflect globally the results of the chemical reactions responsible for the deactivation of the regulatory protein and the gene state control. The first situation will correspond to systems in which the regulation of the control site is simple and the protein deactivation depends primarily on the chemical degradation of the transcription

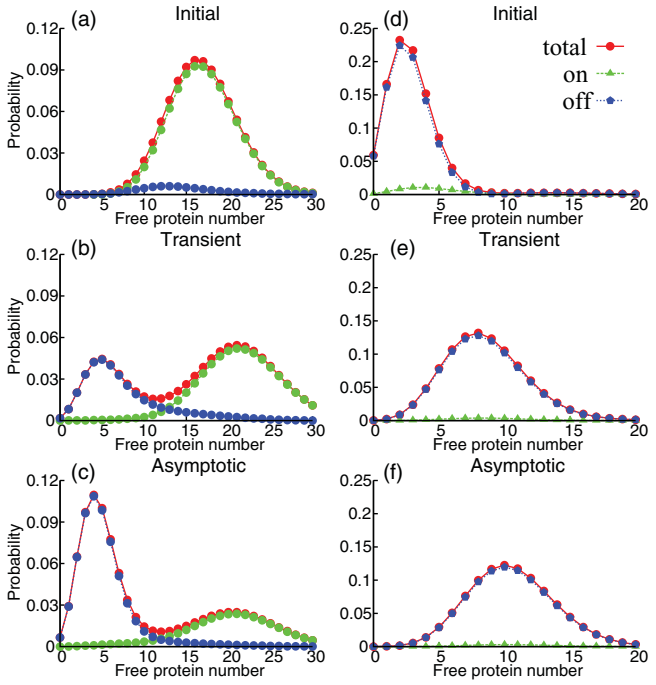


FIG. 1. (Color online) We display the time evolution of the total probability, $\phi_n(t) = \alpha_n(t) + \beta_n(t)$, and of the on and off modes, as indicated. (a)–(c) Parameters values of $k = 22$, $f = 0.05$, $h = 0.005$, $\rho = 1$, and $\chi = 1/5$; (d) and (e) $k = 44$, $f = 0.8$, $h = 3$, $\rho = 1$, and $\chi = 1/5$.

factor. However, one should keep in mind the use of the model to understand effectively more complex regulatory machinery.

In order to illustrate our results we have selected two set of values for the parameters and initial state to be propagated. In Figs. 1(a) and 1(d) we show the starting distributions by plotting the partial and total probabilities versus $n$. The families of constants $C_j^1$ and $C_j^2$ have been adjusted to produce a Poisson-like initial state for the total probability. The time is scaled by fixing the protein degradation rate as 1. Figure 1(a) shows a state of low repression probability, i.e., the gene is fully active, while the opposite is shown in Fig. 1(d). Figure 1(b) corresponds to a transient configuration in which repression has increased, forming a peak around $n = 5$, and the mode of full activity of the gene has been damped, producing another peak around $n = 20$. The stationary state is show in Fig. 1(c). The pronounced peak at $n = 5$ means that the proteins are also expressed in the repressed mode. In Fig. 1(c) we see that the strong initial peak has decreased by a third and has displaced slightly to the right. The life time of the nonequilibrium regime is dominated by the gene switching decay rate $b(\rho + h) \sim 0.16$. The probability for the gene to be repressed increases with time and, when the system is at steady state, is of the same order as the probability for the gene to be active.

The transient configuration in Fig. 1(e) shows that the maximum probability has displaced from $n = 2$ to $n = 8$. The steady state configuration is still a one-peaked distribution centered around $n = 10$, as shown in Fig. 1(f). Inspection is enough to verify that $\beta_n \approx \phi_n$ and that the total probability of the gene to be in the active state, $p_\alpha = \sum_{n=0}^{\infty} \alpha_n$, is almost zero, indicating that the gene remains in the repressed state the majority of time, even in the equilibrium regime. The protein deactivation dominates the nonequilibrium regime since the $b(\rho + h) = 33.8$ contribution rapidly falls off.

We discuss the meaning of the switching constant $b$ by comparing the two steady state probability distributions that we show. In Figs. 1(a)–1(c), we have $b \approx 0.16$, which indicates that the gene switching is slower than the protein degradation. In terms of average life times, it means that the gene stands in a state long enough that a protein is synthesized and loses its functionality before a transition to the other gene state. This explains the existence of two-peaked probability distributions for low values of $b$, independently of the relation between $f$ and $h$. In Figs. 1(d) and 1(e) we have selected the switching constant $b = 8.45$, indicating that the gene switching performs several transitions between its states during the mean life time of a protein.

Two points deserve a brief comment: bimodality and positive regulation. The master equations for the spin-boson model are equivalent to a linear partial differential equation for the time-dependent generating function and consequently present a single equilibrium regime reached by all initial conditions. This is in contrast to dynamical systems of nonlinear differential equations, which can present more than one equilibrium state, with each one reached from a different set of initial conditions. The appearance of single- or double-peaked distributions cannot be confused with the bistability of a dynamical system. The binary spin-boson model allows distributions centered around one or two points and also table-shaped unlocalized probabilities, depending on the region of the parameter space chosen. This welcome freedom may be used depending on the phenomena selected for the investigation. In this Brief Report we consider negative self-regulation by allowing a linear dependence on $n$ in the binding rate and $\chi < 1$, enforcing that the $\alpha$ mode is a copious transcription state. Positive regulation can be obtained by changing $n$ dependence to the unbinding constant or considering $\chi > 1$ in the solutions presented here. Those are minor changes and not a severe limitation of the model.

Bimodality is caused by slow transitions between two expression states and not by the class of regulation that the gene is submitted to (positive or negative, auto or external). Thus, a gene operating in $N$ states (or $N$ different rates of synthesis) and with switching constant and very slow would express proteins in terms of $N$-modal probability distributions. Since bimodality has been attributed to positive feedback [28], mainly based on experimental data, it would be interesting to verify the bimodal behavior in simple wild or synthetic negatively regulated systems. In the case of the nonexistence of binary behavior in wild negative self-regulating genes, it is a task to understand, in an evolutionary framework, how and why nature evolved such that only positive feedback is present.

In summary, we have shown here and in [17] that the binary stochastic model for gene regulation is completely integrable, both in the stationary state and in the dynamical regime. The integrability, as usual, is a consequence of symmetries, as we have shown in [19]. We expect such a soluble time-dependent model for the expression of a single gene can be used as a building block for more complex systems, taking advantage of the analytical solution in terms of well-known functions in

closed forms and easily manipulating the available calculation techniques or mathematical machinery.

The model presented here has a major limitation, which is the nature of the Markov process selected at the beginning, which couples the birth and death stochastic processes by the linear coupling of the binding probabilities, although dimer-like coupling is feasible [29,30]. In principle, we can expect that the solutions obtained here may be used as an approximating procedure. A second limitation is the existence of only two states corresponding to whether an operator is occupied or not. Even the simple lac operon exhibits more than two regulation levels. The control of the process of the *Drosophila* segmentation is made by many different enhancers and transcription factors. Of course, we can approximate the phenomena by means of major classes of "on"and "off"; however, a more realistic approach requires the inclusion of higher spins, which are under investigation in our group.


[1] Manu, S. Surkova, A. V. Sirov, V. V. Gursky, H. Janssens, A. R. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz, Plos Computational Biology **5**, 1000303 (2009).

[2] A. N. Boettiger and M. Levine, Science **325**, 471 (2009).

[3] M. Delbrück, J. Chem. Phys. **8**, 120 (1940).

[4] J. M. Raser and E. K. O'Shea, Science **309**, 2010 (2005).

[5] R. M. Adams, K. F. Murphy, K. Josic, and G. Balázsi, Proc. Natl. Acad. Sci. U.S.A. **106**, 5123 (2009).

[6] H. H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. U.S.A. **94**, 814 (1997).

[7] E. Aurell and K. Sneppen, Phys. Rev. Lett. **88**, 048101 (2002).

[8] M. Santillán and M. C. Mackey, Biophys. J. **86**, 1282 (2004).

[9] J. Peccoud and B. Ycart, Theor. Popul. Biol. **48**, 222 (1995).

[10] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).

[11] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. U.S.A. **98**, 8614 (2001).

[12] M. Sasai and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **100**, 2374 (2003).

[13] J. Paulsson, Phys. Life Rev. **2**, 157 (2005).

[14] T. Shen, M. R. Bennett, P. G. Wolynes, and J. Hasty, Proc. Natl. Acad. Sci. U.S.A. **104**, 18982 (2007).

[15] A. Mugler and C. H. Wiggins, Proc. Natl. Acad. Sci. U.S.A. **106**, 6529 (2009).

[16] R. Guantes and J. F. Poyatos, PLoS Comput. Biol. **4**, e1000235 (2008).

[17] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Phys. Rev. E **72**, 051907 (2005).

[18] G. C. Innocentini and J. E. Hornos, J. Math. Biol. **55**, 413 (2007).

[19] A. F. Ramos and J. E. M. Hornos, Phys. Rev. Lett. **99**, 108103 (2007).

[20] A. F. Ramos, G. C. P. Innocentini, F. M. Forger, and J. E. M. Hornos, IET Syst. Biol. **4**, 311 (2010).

[21] D. B. Saakian, Phys. Rev. E **78**, 061920 (2008).

[22] L. M. Bishop and H. Qian, Biophys. J. **98**, 1 (2010).

[23] *Heun's Differential Equations*, edited by A. Ronveaux (Oxford University Press, Oxford, 1995).

[24] A. A. Stahlhofen, J. Phys. A **37**, 10129 (2004).

[25] E. S. Cheb-Terrab, J. Phys. A **37**, 9923 (2004).

[26] P. P. Fiziev, Classical and Quantum Gravity **27**, 135001 (2010).

[27] See supplemental material at [http://link.aps.org/supplemental/10.1103/PhysRevE.83.062902] for explicit formulas.

[28] T.-L. To and N. Maheshri, Science **327**, 1142 (2010).

[29] D. Schultz, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **105**, 19165 (2008).

[30] K-Y. Kim, D. Lepzelter, and J. Wang, J. Chem. Phys. **126**, 034702 (2007).