Self-regulatory gene: an exact solution for the gene gate model

Yves Vandecan, Ralf Blossey

Self-regulatory gene: An exact solution for the gene gate model

Yves Vandecan and Ralf Blossey

Interdisciplinary Research Institute USR 3078 CNRS and Université de Sciences et de Technologies de Lille, Parc de la Haute Borne,
50 Avenue de Halley, 59658 Villeneuve d'Ascq, France

The stochastic dynamics of gene expression is often described by highly abstract models involving only the key molecular actors DNA, RNA, and protein, neglecting all further details of the transcription and translation processes. One example of such models is the "gene gate model," which contains a minimal set of actors and kinetic parameters, which allows us to describe the regulation of a gene by both repression and activation. Based on this approach, we formulate a master equation for the case of a single gene regulated by its own product—a transcription factor—and solve it exactly. The obtained gene product distributions display features of mono- and bimodality, depending on the choice of parameters. We discuss our model in the perspective of other models in the literature.

I. INTRODUCTION

Gene expression is a stochastic but tightly regulated process [1–3]. From a physics point of view, the transcription of a gene can be represented in an abstract way as a biochemical network involving DNA, mRNA, and proteins, while ignoring all details of the transcription and translation process: polymerase and ribosome activities, as many other processes, are assumed to be taken into account by a proper choice of reaction rates. Then, denoting, e.g., the biomolecular actors as a gene (G), RNA (R), and proteins (P), the stochastic dynamics of a gene is described by the time evolution of the triplet (G, R, P), in which G can assume two states only (e.g., on, off), while R, P are variables with integer ranges of molecule numbers (m, n).

The literature on this general family of models is becoming increasingly hard to overlook, given the many variants of possible interactions that can be considered (binding of transcription factors to promoter sites, dimerization, etc.); an exemplary and still not exhaustive list is Refs. [4–28]. Obviously, an increase of the number of interactions will enlarge parameter space and allow for an increasingly rich stochastic dynamics. The question then arises as to what the modeling criterion is: as the models are highly reduced to begin with, (a) *what can be considered as essential* in them and (b) *according to what criterion?*

Undoubtedly, one of the first models addressing stochastic gene regulation by a master equation approach is the model by Peccoud and Ycart (in what follows PYM) [4]. One can easily characterize the complexity of this model by the number of its variables (3) and parameters (reaction rates, 4). A deficit of the PYM model is that a gene cannot be regulated: hence there is, in particular, no feedback interaction of a gene product on a gene. The "gene gate model" (GGM) we base our work on here is of the same complexity as the PYM in the above sense, but it allows for the interaction with transcription factors and hence in particular for a feedback interaction with its own product; one may thus consider it as a "minimal" model for a regulated gene. The GGM was introduced originally in Ref. [5] and further developed in Refs. [6,7]; it has so far been used for Gillespie-based simulations of small gene circuits, up to the level of the repressilator, and its deterministic limit has also been studied.

In the nomenclature proposed in Ref. [19], in its simplest version the gene gate model is an example of a "two-state" model; i.e., it only contains the gene and a single product variable, which can be interpreted as either an RNA-molecule or a protein; we here choose to call the output protein for definiteness. The completion of the model to include both RNA and protein is trivial and hence ignored here. The main modeling issue we consider in this paper is the treatment of the proteins in terms of their feedback interactions with one isolated gene, i.e., the regulation of a single gene by its own product, assumed to be a transcription factor, and the obtention of an exact solution for this model. This is not a technical feat by itself, as exact solutions have already been given for more detailed models (but in general, the GGM does not reduce to one of them). The specific questions we address and answer here for the self-regulated gene are: what does such a maximally reduced model produce as output and hence what will additional interactions modify? The latter point is addressed by putting our model in the context of some of the recently discussed other models at the end of the paper. In that manner we contribute to the discussion of auto-regulation of a single gene.

II. THE GENE GATE MODEL FOR THE SELF-REGULATED GENE

We begin by introducing the rate equations of the application of the GGM to the self-regulated gene. It comes in two variants describing an auto-repressed and an auto-activated gene, respectively. In the repressed case (R), the gene G can be constitutively transcribing at a rate ε,

$$G \to_\varepsilon G + P, \tag{1}$$

which is considered the "on" state of the gene. Proteins degrade at a rate δ,

$$P \to_\delta 0. \tag{2}$$

The feedback of a protein P on the gene can affect gene activity by putting the gene into its second state. This is described by the reaction

$$G + P \to_r G' + P. \tag{3}$$
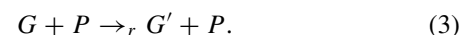
FIG. 1. Schematic representation of gene-protein interaction, making use of a Feynman-graph-like notation. Left, the gene goes from state 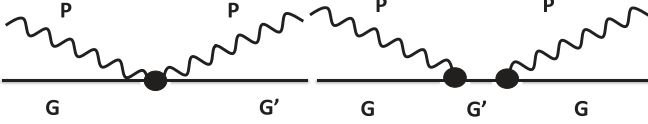G to G' after "scattering" with a protein P; right, the gene "absorbs" the protein, turns from state G to G', and then returns to state G while "emitting" a protein.

In the repressed case, the gene state G' is not productive and simply relaxes back to G at a rate η,

$$G' \to_\eta G. \tag{4}$$

G' is thus the "off" state of the gene. In the activated case (A), the gene is productive in state G' via

$$G' \to_\eta G + P, \tag{5}$$

hence, it relaxes back to gene state G while producing protein P, provided the rate $\eta > \varepsilon$. Here, G and G' change roles: G is now the "off" state (although never fully off) and G' is the "on" state. We note that in the case $\eta < \varepsilon$, the gene is actually repressed but left with protein production at a finite rate η. This case (not shown) does not differ qualitatively from the case (R).

It is worth commenting on reaction Eq. (3), which in conjunction with Eqs. (4) and (5) is key to the GGM. The GGM does not have a state of a protein bound to the promoter (a transcription factor-DNA complex), in contrast to most (but not all) other models. One can compare the two cases by alluding to the analogy of the presence of "bound state" as an "absorbing" state of the gene, while the GGM in this terminology only has a "scattering" state; see the illustration in Fig. 1. Here, the gene "propagates" in time in state G when at a time t a protein interacts and puts it instantaneously into state G'. In the absorbing state, G' describes a gene "propagating" with the transcription factor bound to its promoter site. The neglect of the bound state in the GGM is a crucial simplifying assumption of the model, made in order to make the model as minimal as possible and keeping it regulated at the same time.

We stress again that this simple set of reactions with variables G, G', P and parameters $(\varepsilon, \delta, r, \eta)$ corresponds in terms of its complexity *exactly* to the PYM. The major difference between the two is that a gene in the PYM never interacts with proteins, and hence reaction Eq. (3) occurs without P intervening. Further, the reaction Eq. (5) of case (A) does not apply. The main difference between the GGM and the PYM lies, in fact, precisely in this reaction Eq. (5), as it leads in the master equation to state changes involving both the gene state and protein number at the same time, and this difference requires us to obtain a (novel) exact solution of the corresponding master equation of the GGM for the single gene.

III. THE MASTER EQUATION OF THE SELF-REGULATING GENE AND ITS EXACT STEADY-STATE SOLUTION

The master equations for the two versions of the GGM are readily established along the same line as for the PYM. We

denote the basal state of the gene by 0, its second state by 1, and count protein number by n, so that we can introduce time-dependent probability distributions $p_{i,n}(t)$, with $i = 0, 1$. In case (R), Eq. (4) applies and the master equations read as

$$\forall n \geqslant 0: \partial_t p_{1,n} = nr p_{0,n} - \eta p_{1,n}$$
$$+ \delta[(n+1)p_{1,n+1} - n p_{1,n}]$$
$$\partial_t p_{0,0} = \eta p_{1,0} - \varepsilon p_{0,0} + \delta p_{0,1} \tag{6}$$
$$\forall n \geqslant 1: \partial_t p_{0,n} = \eta p_{1,n} + \varepsilon p_{0,n-1} - (\varepsilon + nr)p_{0,n}$$
$$+ \delta[(n+1)p_{0,n+1} - n p_{0,n}],$$

while in case (A), for which Eq. (5) applies, and we have

$$\partial_t p_{0,0} = -\varepsilon p_{0,0} + \delta p_{0,1}$$
$$\forall n \geqslant 1: \partial_t p_{0,n} = -(\varepsilon + nr)p_{0,n} + \varepsilon p_{0,n-1} + \eta p_{1,n-1}$$
$$+ \delta[(n+1)p_{0,n+1} - n p_{0,n}]$$
$$\forall n \geqslant 0: \partial_t p_{1,n} = -\eta p_{1,n} + nr p_{0,n}$$
$$+ \delta[(n+1)p_{1,n+1} - n p_{1,n}]. \tag{7}$$

Introducing the generating functions $G_0(z,t) = \sum_{n=0}^\infty p_{0,n}(t)z^n$ and $G_1(z,t) = \sum_{n=0}^\infty p_{1,n}(t)z^n$, the master equations above are transformed into first-order differential equations in t and z. For case (R), we have

$$\partial_t G_0 = \eta G_1 + \varepsilon(z-1)G_0 - (\delta(z-1) + rz)\partial_z G_0$$
$$\partial_t G_1 = -\eta G_1 - \delta(z-1)\partial_z G_1 + rz\partial_z G_0, \tag{8}$$

while in case (A), we have

$$\partial_t G_0 = \eta z G_1 + \varepsilon(z-1)G_0 - (\delta(z-1) + rz)\partial_z G_0$$
$$\partial_t G_1 = -\eta G_1 - \delta(z-1)\partial_z G_1 + rz\partial_z G_0. \tag{9}$$

At this point it is instructive to compare the differential equations for the GGM to those of the PYM. The reaction Eq. (3), which describes the state change of the gene from G to G', is protein-number dependent and leads to a change from a term $\sim rG_0$ to a term $\sim z\partial_z G_0$; otherwise, the equations remain unchanged for case (R). In case (A), the notable difference is the presence of the term $\sim \eta z G_1$ in the equation for G_0, which results from reaction Eq. (5). Note that the corresponding η-dependent term in the equation for G_1 does not depend on z: the symmetry between the equations is broken.

We begin by solving the equations for case (A). As in Ref. [24], instead of solving for G_0 and G_1 separately, we introduce the total generating function $G_T = G_0 + G_1$ with straightforward biological interpretation, allowing us to determine the mean value of the protein number and its fluctuations.

In case (A), the term $(z-1)$ can be placed in front in the equation for G_T, i.e.,

$$\partial_t G_T = (z-1)[-\delta \partial_z G_T + \varepsilon(G_T - G_1) + \eta G_1]. \tag{10}$$

Obviously, a system singularity arises at $z = 1$. The equality Eq. (10) is trivially satisfied at $z = 1$, because $G_T(1,t) = 1 \, \forall t$. A stationary solution requires $\partial_t G_T(z,t) = 0 \, \forall z, t$. For $z \neq 1$ this implies

$$-\delta \partial_z G_T + \varepsilon G_T + (\eta - \varepsilon)G_1 = 0. \tag{11}$$

Using Eq. (11), we can express G_1 and $\partial_z G_1$ in terms of $G_T(z,t)$. For the stationary solution, expression Eq. (9) can be

rewritten in a second-order differential equation

$$\{\delta[\delta - z(r+\delta)]\}\partial_z^2 g_T(z) + (-\delta\varepsilon + \delta z\varepsilon + rz\eta - \eta\delta)\partial_z g_T(z)$$
$$+ \eta\varepsilon g_T(z) = 0, \tag{12}$$

with $g_T(z) \equiv \lim_{t\to\infty} G_T(z,t)$. Substitution of $x = \delta[\delta - z(r+\delta)]$ in Eq. (12) leads to the familiar Kummer equation,

$$x\frac{\partial^2}{\partial x^2}\tilde{g}_T(x) + (a+bx)\frac{\partial}{\partial x}\tilde{g}_T(x) + c\tilde{g}_T(x) = 0. \tag{13}$$

Here, $a = (\varepsilon r + \delta\eta)/(r+\delta)^2$, $b = (\delta\varepsilon + r\eta)/[\delta^2(r+\delta)^2]$, $c = \eta\varepsilon/[\delta^2(r+\delta)^2]$, and $g_T(z) = \tilde{g}_T(x)$. Since $x = 0$ is a weak singular point of the differential Eq. (13), according to the method of Frobenius, a power series solution can be found around the singular point, i.e.,

$$\tilde{g}_T(x) = C_1 \sum_{n=0}^{\infty} a_n x^n + C_2 x^{1-a} \sum_{n=0}^{\infty} b_n x^n. \tag{14}$$

The substitution of the first term, i.e., the power series $\sum_{n=0}^{\infty} a_n x^n$, into Eq. (13), leads to the KummerM or the hypergeometric function $a_0 \, {}_1F_1(c/b,a;-bx)$. The second term corresponds to the KummerU function, which can be discarded because the probability of n proteins does not tend to zero for $n \to \infty$. Consequently, the appropriate generating function $g_T(z)$ reads

$$g_T(z) = C_1' \, {}_1F_1\{c/b,a;b\delta[z(r+\delta) - \delta]\}. \tag{15}$$

The coefficient C_1' is determined by the condition $\lim_{z\to 1} g_T(z) = g_T(1) = 1$ due to the continuity of the hypergeometric function ${}_1F_1$, so that $C_1' = {}_1F_1(c/b,a;b\delta r)$. The asymptotic solution in all the parameters equals

$$g_T(z) = {}_1F_1\left\{\frac{\eta\varepsilon}{\delta\varepsilon + r\eta}, \frac{\varepsilon r + \eta\delta}{(r+\delta)^2}; \frac{(\delta\varepsilon + r\eta)[z(r+\delta) - \delta]}{\delta(r+\delta)^2}\right\}$$
$$\bigg/ {}_1F_1\left\{\frac{\eta\varepsilon}{\delta\varepsilon + r\eta}, \frac{\varepsilon r + \eta\delta}{(r+\delta)^2}; \frac{(\delta\varepsilon + r\eta)r}{\delta(r+\delta)^2}\right\}. \tag{16}$$

From the total generating function Eqs. (11) and (15), we can also separately derive the asymptotic solutions $g_0(z) = \lim_{t\to\infty} G_0(z,t)$ and $g_1(z) = \lim_{t\to\infty} G_1(z,t)$. The full expressions of the asymptotic solutions are

$$g_0(z) = C_1'\bigg[{}_1F_1\{c/b,a;b\delta[z(r+\delta) - \delta]\}$$
$$+ \frac{1}{\eta - \varepsilon}\bigg(\varepsilon \, {}_1F_1\{c/b,a;b\delta[z(r+\delta) - \delta]\}$$
$$- \delta^2 c \frac{r+\delta}{a} {}_1F_1\{c/b+1,a+1;b\delta[z(r+\delta) - \delta]\}\bigg)\bigg] \tag{17}$$

$$g_1(z) = C_1' \frac{1}{\eta - \varepsilon}\bigg(-\varepsilon \, {}_1F_1\{c/b,a;b\delta[z(r+\delta) - \delta]\}$$
$$+ \delta^2 c \frac{r+\delta}{a} {}_1F_1\{c/b+1,a+1;b\delta[z(r+\delta) - \delta]\}\bigg). \tag{18}$$

The probability for measuring n proteins, $p_{0,n} + p_{1,n}$, is obtained from $(1/n!)\partial_z^n g_T(z=0)$. We obtain an analytical expression for the stationary protein distribution $P(n)$

$$P(n) = \frac{1}{n!}\left[\frac{\delta\varepsilon + r\eta}{\delta(r+\delta)}\right]^n \frac{\left(\frac{\eta\varepsilon}{\delta\varepsilon + r\eta}\right)_n}{\left[\frac{\varepsilon r + \eta\delta}{(r+\delta)^2}\right]_n}$$
$$\times \frac{{}_1F_1\left[\frac{\eta\varepsilon}{\delta\varepsilon + r\eta} + n, \frac{\varepsilon r + \eta\delta}{(r+\delta)^2} + n; \frac{-\delta(\delta\varepsilon + r\eta)}{\delta(r+\delta)^2}\right]}{{}_1F_1\left[\frac{\eta\varepsilon}{\delta\varepsilon + r\eta}, \frac{\varepsilon r + \eta\delta}{(r+\delta)^2}; \frac{(\delta\varepsilon + r\eta)r}{\delta(r+\delta)^2}\right]}, \tag{19}$$

with $(\cdot)_n$ being the Pochhammer symbol.

The mean protein number E and its fluctuations V are easily obtained from the first and second derivative of the generating function $g_T(z)$ with respect to z, evaluated at $z = 1$. In the activated case the resulting expressions are given by

$$E = C_1' \frac{c\delta(r+\delta)}{a} {}_1F_1(c/b+1,a+1;b\delta r) \tag{20}$$

and

$$V = C_1' \frac{c(c/b+1)b\delta^2(r+\delta)^2}{a(a+1)} {}_1F_1(c/b+2,a+2;b\delta r)$$
$$+ C_1' \frac{c\delta(r+\delta)}{a} {}_1F_1(c/b+1,a+1;b\delta r)$$
$$- \left[C_1' \frac{c\delta(r+\delta)}{a} {}_1F_1(c/b+1,a+1;b\delta r)\right]^2. \tag{21}$$

In case (R), the set of Eqs. (8) corresponds to the master equations of the gene model studied before by Hornos *et al.* [11,24] when identifying $G_0 \equiv \alpha$, $G_1 \equiv \beta$, $\varepsilon \equiv g_\alpha$, $r \equiv h$, $\delta \equiv k$, $\eta \equiv f$, and putting $g_\beta = 0$. In terms of our terminology, we can proceed as before and obtain the following asymptotic solution for $g_T(z) = g_0(z) + g_1(z)$,

$$g_T(z) = {}_1F_1\left\{\frac{\eta}{\delta}, \frac{(\varepsilon + \eta)r + \delta\eta}{(r+\delta)^2}; \frac{\varepsilon[z(r+\delta) - \delta]}{(r+\delta)^2}\right\}$$
$$\bigg/ {}_1F_1\left\{\frac{\eta}{\delta}, \frac{(\varepsilon + \eta)r + \delta\eta}{(r+\delta)^2}; \frac{\varepsilon r}{(r+\delta)^2}\right\}, \tag{22}$$

from which the probability distribution of the protein number, mean, and variance can be derived.

Figure 2 shows the resulting characteristic probability distributions for the activated (A) and the repressed gene (R). If protein degradation dominates over protein production, the protein probability is peaked at $n = 0$ [shown for (A), but likewise for (R)] (a). An increase of protein production lets a peak at finite n grow out of the peak at $n = 0$ (b). At a sufficiently high protein production, the distribution turns into a boundary bimodal when tuning down the basal rate (c). To be more specific, in function of decreasing protein number n, the distribution curve lowers monotonically in value from the peak at finite n (due to the sufficiently high protein production rate) to $n = 1$, followed by a sudden increase, a jump at $n = 0$. In the limit of the basal rate to zero, $n = 0$ becomes an absorbing state [23]. Concerning the repressed gene gate model, a boundary bimodal is generated with a more gradual change in derivative toward the $n = 0$ state, i.e., from the peak at finite n, the probability is first lowered, after which it gradually rises to $n = 0$ (d). As a last characteristic probability distribution, a clear monomodal protein distribution appears with a peak at finite protein number for (A) if we retain the same parameter
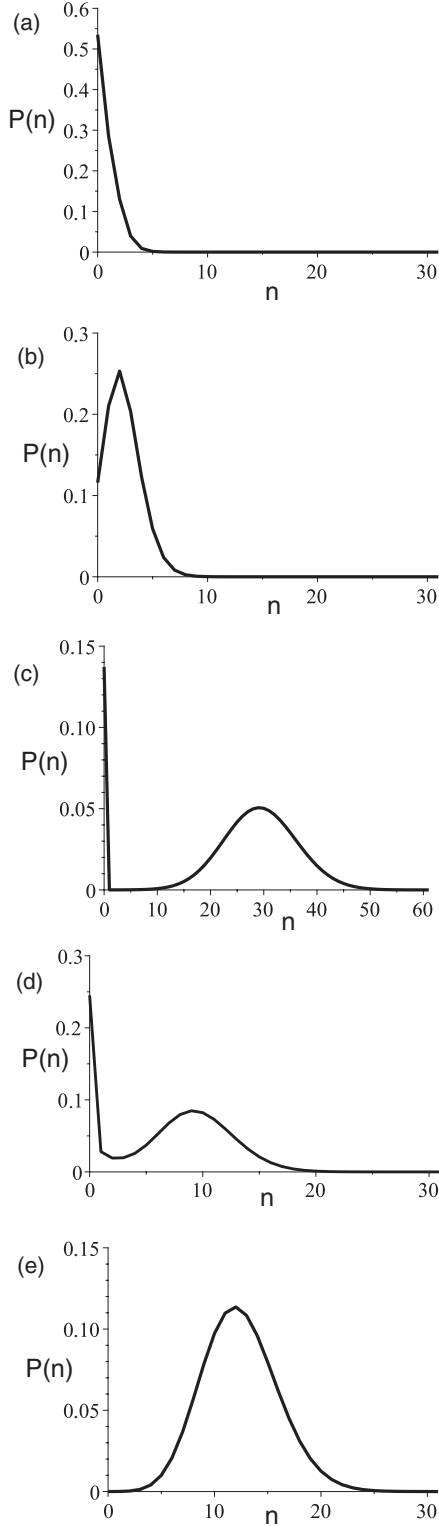
FIG. 2. The probability distribution $P(n)$. Models and parameters from (a) to (e): (a) (A), $\varepsilon = 0.01$; $\eta = 0.1$; $r = 1.0$; $\delta = 0.1$; (b) (A), $\varepsilon = 0.01$; $\eta = 0.1$; $r = 1.0$; $\delta = 0.04$; (c) (A), $\varepsilon = 0.0001$; $\eta = 30$; $r = 1$; $\delta = 0.5$; (d) (R), $\varepsilon = 10$; $\eta = 0.1$; $r = 0.005$; $\delta = 1$; (e) (A), $\varepsilon = 0.01$; $\eta = 0.1$; $r = 1.0$; $\delta = 0.008$.

values as in (b) but lower the degradation rate even further (e). Our analytical results of the gene gate models predict accurately the results of the stochastic dynamics simulations

of these simple gates. For the parameters $r = 1$, $\varepsilon = 0.01$, $\eta = 0.1$, and $\delta = 0.001$, our Eq. (19) reduces approximately to a Poisson distribution,

$$P(n) \approx \frac{\lambda^n}{n!} e^{-\lambda}, \tag{23}$$

with a mean $\lambda = (\delta\varepsilon + r\eta)/[\delta(r + \delta)] \approx 99$ proteins in steady-state, in accordance with the simulations for the auto-activated gene gate [5].

IV. THE TIME-DEPENDENT PROBLEM

In this section, we show, for completeness, that the master equation of the gene gate model of the self-regulated gene can be solved also in the time-dependent case, for both the activated and repressed gene, using the approach developed in Ref. [24].

Activated case. The time-dependent equations for the generating function $G_0(z,t)$ and $G_1(z,t)$ are

$$\frac{\partial}{\partial t} G_0(z,t) = \varepsilon(z - 1)G_0(z,t) + \delta(1 - z)\frac{\partial}{\partial z} G_0(z,t)$$

$$- rz\frac{\partial}{\partial z} G_0(z,t) + \eta z G_1(z,t) \tag{24}$$

$$\frac{\partial}{\partial t} G_1(z,t) = \delta(1 - z)\frac{\partial}{\partial z} G_1(z,t)$$

$$+ rz\frac{\partial}{\partial z} G_0(z,t) - \eta G_1(z,t). \tag{25}$$

Summing up Eqs. (24) and (25) yields

$$\frac{1}{z - 1}\frac{\partial G_T(z,t)}{\partial t} = -\delta\frac{G_T(z,t)}{\partial z} + \varepsilon G_T(z,t)$$

$$+ (\eta - \varepsilon)G_1(z,t). \tag{26}$$

Consequently, we can express $G_1(z,t)$ in terms of the total generating function $G_T(z,t) = G_0(z,t) + G_1(z,t)$, which, after substitution in Eq. (25), transforms the latter into a second-order partial differential equation for $G_T(z,t)$, i.e.,

$$\frac{\partial^2 G_T(z,t)}{\partial t^2} = [\delta(\delta + r)(z - 1)^2 + r\delta(z - 1)]\frac{\partial^2 G_T(z,t)}{\partial z^2}$$

$$+ [(2\delta + r)(z - 1) + r]\frac{\partial^2 G_T(z,t)}{\partial t \partial z}$$

$$+ [-\varepsilon(z - 1) - \frac{r}{z - 1} - r + \eta - \delta]\frac{\partial G_T(z,t)}{\partial t}$$

$$+ [-(\delta\varepsilon + \eta r)(z - 1)^2 + \eta(1 - r)\delta(z - 1)]$$

$$\times \frac{\partial G_T(z,t)}{\partial z} - \eta\varepsilon(z - 1)G_T(z,t). \tag{27}$$

The substitution of the variables $\nu = [z(\delta + r) - \delta]/r$ and $\mu = \nu e^{-(\delta+r)t}$ makes Eq. (27) separable in these variables. This substitution brings the singularities at $z = \delta/(\delta + r)$ and $z = 1$ to $\nu = 0$ and $\nu = 1$, respectively. The new form of the partial equation in terms of these variables is

$$\left[\frac{\partial^2}{\partial \nu^2} + \mu P_a \frac{\partial^2}{\partial \nu \partial \mu} + \mu Q_a \frac{\partial}{\partial \mu} + R_a \frac{\partial}{\partial \nu} + S_a\right]G_T(\mu,\nu) = 0 \tag{28}$$

The coefficients P_a, Q_a, R_a, and S_a are given by

$$P_a = \frac{1}{\nu} - \frac{\delta + r}{(\nu - 1)\delta}, \tag{29}$$

$$R_a = \frac{-r^2 \eta}{(\delta + r)^2 \delta} + \frac{-\varepsilon r}{(\delta + r)^2} + \frac{\varepsilon r + \delta \eta}{\nu(\delta + r)^2}, \tag{30}$$

$$Q_a = \frac{\delta \eta + \varepsilon r - r^2 - \delta^2 - 2\delta r}{\nu^2 (\delta + r)^2}$$
$$+ \frac{\delta^2 \eta + 2\delta \eta r + r^2 \varepsilon}{\nu(\delta + r)^2 \delta} - \frac{\eta}{(\nu - 1)\delta} + \frac{\delta + r}{(\nu - 1)^2 \delta}, \tag{31}$$

$$S_a = \frac{-r\varepsilon \eta}{(\delta + r)^2 \nu \delta}. \tag{32}$$

Now we search for special solutions of the type $\mu^\lambda H_\lambda(\nu)$, with λ the separation constant. The equation for the function $H_\lambda(\nu)$ corresponds to a confluent Heun equation in most general form,

$$\left[\frac{\partial^2}{\partial \nu^2} + (R_a + \lambda P_a) \frac{\partial}{\partial \nu} + S_a + \lambda Q_a \right] H_\lambda(\nu) = 0. \tag{33}$$

This differential equation is easily manipulated by using the symbolic computational software package Maple. The general solution for the time-dependent total generating function $G(z,t)$ is given by

$$(\nu - 1)e^{c_a \nu} \left[\sum_{j=0}^{\infty} C_j^1 e^{-j\delta t} H_C \left(c_a, \theta_a^1, 1 - j, \delta_a^1, \eta_a^1, \nu \right) \right.$$
$$\left. + \sum_{j=0}^{\infty} C_j^2 \nu^{j+1} e^{-(\delta + r)(j + b_a)t} H_C \left(c_a, j + 1, \sigma_a^2, \delta_a^2, \eta_a^2, \nu \right) \right]. \tag{34}$$

Here, H_C is the confluent Heun function according to the definition of Maple, $\nu = [z(\delta + r) - \delta]/r$ as already mentioned, and j is an integer running from zero to infinity. The explanations of the symbols in terms of the model parameters are shown in Appendix. The coefficients C_j^1 and C_j^2 depend on the initial conditions. We use a similar notation as in Ref. [24], in order to be able to compare the two results easily.

The $j = 0$ term of the first series of confluent Heun functions is independent of the time t, reduces to the confluent hypergeometric function $_1F_1$, and corresponds to the asymptotic solution in our paper. The remaining terms are exponentially decaying functions of time, which reflect the transient behavior of the system. It is interesting to note that for the activated gene gate, the gene switching time $1/[b_a(\delta + r)]$ is small with respect to the protein degradation time $1/\delta$, so that a monomodal behavior would be expected. However, because in the limit of vanishing basal production rate ε, $n = 0$ becomes an absorbing state and hence a boundary bimodal is found. In the repressed case, the gene switching time is large with respect to the degradation time, and a "proper" boundary bimodal is found. The gene switching and protein degradation rates for our cases are shown in Table I.

Repressed case. The self-repressed gene coincides with the analysis of Ref. [24] with no output in the repressed state ($\chi = 0$ in Ref. [24]). The time-dependent equations for the

TABLE I. The protein degradation rate and gene switching rate for the five cases described in the legends of Figs. 2(a)–2(e). The formula for b_a is found in the Appendix; note that in the repressed case, b applies instead of b_a.

	(1)	(2)	(3)	(4)	(5)
δ	0.1	0.04	0.5	1	0.008
$b_a(\delta + r)$	0.018	0.013	10	0.15	0.011

generating function $G_0(z,t)$ and $G_1(z,t)$ are

$$\frac{\partial}{\partial t} G_0(z,t) = \varepsilon(z - 1)G_0(z,t) + \delta(1 - z)\frac{\partial}{\partial z} G_0(z,t)$$
$$- rz \frac{\partial}{\partial z} G_0(z,t) + \eta G_1(z,t), \tag{35}$$

$$\frac{\partial}{\partial t} G_1(z,t) = \delta(1 - z)\frac{\partial}{\partial z} G_1(z,t) + rz \frac{\partial}{\partial z} G_0(z,t)$$
$$- \eta G_1(z,t). \tag{36}$$

Summing up Eqs. (35) and (36) yields

$$\frac{1}{z - 1} \frac{\partial G_T(z,t)}{\partial t} = \varepsilon G_0(z,t) - \delta \frac{\partial G_T(z,t)}{\partial z}. \tag{37}$$

We are then able to express $G_0(z,t)$ in terms of the total generating function $G_T(z,t) = G_0(z,t) + G_1(z,t)$, which, after substitution in Eq. (35), transforms the latter into a second-order partial differential equation for $G_T(z,t)$; i.e.,

$$\frac{\partial^2 G_T(z,t)}{\partial t^2} = [\delta^2(z-1)^2 + r\delta(z-1)^2 + r\delta(z-1)]\frac{\partial^2 G_T(z,t)}{\partial z^2}$$
$$+ [(2\delta + r)(z - 1) + r]\frac{\partial^2 G_T(z,t)}{\partial t \partial z}$$
$$+ \left[-\varepsilon(z-1) - \frac{r}{z-1} - r + \eta - \delta \right]\frac{\partial G_T(z,t)}{\partial t}$$
$$+ [-\delta\varepsilon(z-1)^2 + \eta\delta(z-1)]\frac{\partial G_T(z,t)}{\partial z}$$
$$- \eta\varepsilon(z-1)G_T(z,t), \tag{38}$$

$$P = \frac{1}{\nu} - \frac{\rho + r}{(\nu - 1)\delta}, \tag{39}$$

$$R = \frac{-\varepsilon r}{(\delta + r)^2} + \frac{\delta \eta + \varepsilon r + r\eta}{\nu(r + \delta)^2}, \tag{40}$$

$$Q = \frac{(\delta + r)}{(\nu - 1)^2 \delta} + \frac{\delta \eta + \varepsilon r - r^2 - \delta^2 - 2\delta r + r\eta}{\nu^2(\delta + r)^2}$$
$$- \frac{\eta}{(\nu - 1)\delta} + \frac{\delta^2 \eta + 2\delta \eta r + r^2 \eta + r^2 \varepsilon}{\nu(\delta + r)^2 \delta}, \tag{41}$$

$$S = \frac{-r\varepsilon \eta}{(\delta + r)^2 \nu \delta}. \tag{42}$$

After similar computations, as in the activated case, we obtain the general time-dependent solution for $G(z,t)$, i.e., a product of the term $(\nu - 1)e^{\varepsilon r(\nu - 1)/(\delta + r)^2}$ with two families of confluent

Heun functions:

$$\left[\sum_{j=0}^{\infty} C_j^1 e^{-j\delta t} H_C(c, \theta^1, 1 - j, \delta^1, \eta^1, \nu) \right.$$
$$\left. + \sum_{j=0}^{\infty} C_j^2 \nu^{j+1} e^{-(\delta+r)(j+b)t} H_C(c, j+1, \sigma^2, \delta^2, \eta^2, \nu) \right]. \quad (43)$$

This form is identical to the model of Ref. [24]. For the definition of the symbols, we refer to Appendix.

V. DISCUSSION

It is worth putting our findings in the context of other, more complex models; obviously, we cannot do so for all models in the literature (e.g., those cited above), and therefore choose three representatives: (a) the model discussed by Karmakar [22]; (b) the models discussed by Hornos *et al.* [11], Schultz *et al.* [17], and Grima *et al.* [27], the latter of which critisizes Ref. [11]; (c) the model by Friedman *et al.* [13].

(a) The model by Karmakar [22] considers a gene switching between three states (activated, basal, and repressed) without interactions mediated by the gene product. Mathematically it thus corresponds to an extension of the PYM [4], with four variables (three gene states, the protein) and six reaction rates. While the Peccoud-Ycart model generates only monomodal protein distributions, the extra state introduces the necessary variability to exclusively generate boundary bimodals. The model is clearly more complex than the GGM, as it uses more variables and parameters. An auto-regulated stochastic model, however, displays always a bimodal protein distribution under proper circumstances [19].

(b) The model by Hornos *et al.* [11] and Schultz *et al.* [17], e.g., go beyond the GGM by considering additional transitions between the model states and, in particular, allow for the presence of bound proteins. As can been easily checked from the Hornos *et al.* model [11], with five parameters in the repressed case, two conditions are generally required in the region where degradation does not dominate in order to obtain two distinct maxima at finite protein numbers. First, the rate from the repressed state to the normal state must be high enough in relation to the reverse rate, so that a depopulation from the low-state peak realizes a second maximum at a higher protein number. A second requirement corresponds to a nonzero production rate in the repressed state. A zero production rate (a reduction of one model parameter) in the repressed state leads to a boundary bimodal, as in our case. Our model of the self-repressed gene gate, in fact, reduces to the model of Ref. [11] when the repressed gene has no output. The self-activated GGM produces at most boundary bimodality when, at a sufficiently high production rate, the basal rate is strongly reduced. In the limit of a zero basal rate, the $n = 0$ state becomes an adsorbing state [23]. If more interaction parameters are included, like in Refs. [11,17,27], a bimodal with two distinct peaks can be generated. An interesting remark is that the nature of boundary bimodality in the self-activated case model originates from an absorbing state and not from a slow gene switching time as in the repressed case. Moreover, as explained above, in the self-activated case we see a sudden jump from the $n = 1$ to the $n = 0$ state, while

in the self-repressed case a more gradual increase toward the $n = 0$ state is found.

(c) The model of Friedman *et al.* goes beyond the master equation approach in two ways, such that a simple counting of variables and parameters is not useful anymore. Friedman *et al.* introduce a continuous master equation in which the autoregulation of the gene is modeled by a response function $c(x)$ and a burst size distribution $\nu(x)$ [13]. The response function $c(x)$ in the Friedman model replaces the protein-dependent feedback in our activated gene gate model. But, in our model and in most simple Markovian models (i.e., without added distributions), there is only one specific rate from the n to the $n + 1$ protein state, while an exponential burst size distribution includes transitions from n to $n + 1$, $n + 2, \ldots$ with all different weighting numbers. Consequently, although the Friedman model seems a very simple model too, the employed exponential burst size distribution includes an infinite number of parameters in the continuous master equation compared to the gene gate model for a self-regulating gene. Further, if the response function is characterized by a Hill coefficient $H < 0$, values smaller than -1 corresponding to cooperative effects need to be chosen in order to obtain proper bimodality rather than boundary bimodality in this model.

In conclusion, we believe that the issue of the physical soundness of the class of "abstract" models for gene expression (notably those based on master equations) is, to a considerable extent, interpretation-dependent. This whole model class needs to be taken as a highly abstract description, which on every level of detail included has "unphysical" (or, maybe more specifically, nonbiological) features. In the end, in order to describe the *qualitative behavior* of a gene circuit, what counts more than the level of descriptive detail that is included in the model are the characteristic features of the probability distribution it can generate. One might consider it in fact useful to classify all models into five categories *according to the distributions they generate*: "bimodal" (two distinct peaks), "boundary bimodal" (two peaks with a peak at zero protein number), "monomodal" (one peak), and finally "monotonically decaying" (monotonically decaying distribution as function of increasing protein number n). Our minimal transcriptional regulation model Ð has the least number of variables (3) and parameters (4 rate constants) and is the only auto-activating model that produces a boundary bimodal with a discrete jump from the $n = 1$ to $n = 0$ state. Adding one more parameter, like the Schultz and Hornos model, generates a proper bimodal with two lobes. For the time-independent problem, the stationary probability protein distribution, as well as the mean and its fluctuations, can be expressed in the confluent hypergeometric function. The time-dependent problem is also completely integrable in terms of the well-known confluent Heun functions. The gene gate model, as a minimal transcriptional model for a regulating gene, serves, therefore, as a basic theory to build more complex systems and gives insight into what the additional interactions modify.

APPENDIX: THE PARAMETERS OF THE CONFLUENT HEUN EQUATION

Activated case. In order to simplify notation, we use similar auxiliary expressions as in Ref. [24]:

$$a_a = \frac{\eta}{\delta}, \quad b_a = \frac{\eta}{\delta + r} + \frac{\varepsilon r - \eta r}{(\delta + r)^2}.$$

The Heun parameters are

$$c_a = \frac{r(\eta r + \delta \varepsilon)}{(\delta + r)^2 \delta}, \quad \theta_a^1 = b_a - 1 - j\frac{\delta}{\delta + r}$$

$$\delta_a^1 = -\frac{c_a}{2}\left(2a_a - b_a - j\frac{r}{\delta + r}\right) + \frac{a_a r^2}{(\delta + r)^2}(a_a - j)$$

$$\eta_a^1 = \frac{1}{2}\left[1 - b_a(c_a + j) + 2a_a c_a + j(c_a + j)\frac{\delta}{\delta + r}\right]$$
$$\qquad - \frac{a_a r^2}{(\delta + r)^2}\left(a_a + \frac{\delta}{r}j\right)$$

$$\sigma_a^2 = 1 - (b_a + j)\frac{\delta + r}{\delta}$$

$$\delta_a^2 = \frac{c_a}{2}\left[b_a - 2a_a + (b_a + j)\frac{r}{\delta}\right]$$
$$\qquad + a_a\left[-\frac{r}{\delta}c_a + a_a\frac{r}{\delta}\frac{r^2}{(r + \delta)^2} + c_a\frac{r}{r + \delta} - \frac{r^2}{\delta(\delta + r)}j\right]$$

$$\eta_a^2 = \frac{1}{2}\left[1 + (2a_a + j)c_a + j(b_a + j)\frac{\delta + r}{\delta}\right]$$
$$\qquad - c_a a_a\frac{r}{\delta + r} - a_a^2\frac{(r^2\delta + \delta^2 r)}{(\delta + r)^3}$$

Repressed case. The model parameters for the negative gene gate are shown below. The auxiliary expressions are

$$a = \frac{\eta}{\delta}, \quad b = \frac{\eta}{\delta + r} + \frac{\varepsilon r}{(\delta + r)^2}.$$

The Heun parameters are

$$c = \frac{r\varepsilon}{(\delta + r)^2}, \quad \theta^1 = b - 1 - j\frac{\delta}{\delta + r}$$

$$\delta^1 = -\frac{c_a}{2}\left(2a - b - j\frac{r}{\delta + r}\right)$$

$$\eta^1 = \frac{1}{2}\left[1 - b(c + j) + 2ac + j(c + j)\frac{\delta}{\delta + r}\right]$$

$$\sigma^2 = 1 - (b + j)\frac{\delta + r}{\delta}, \quad \delta^2 = \frac{c}{2}\left[b - 2a + (b + j)\frac{r}{\delta}\right]$$

$$\eta^2 = \frac{1}{2}\left[1 + (2a + j)c + j(b + j)\frac{\delta + r}{\delta}\right]$$

[1] M. Delbrück, J. Chem. Phys. **8**, 120 (1940).
[2] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. USA **98**, 8614 (2001).
[3] W. J. Blake, C. R. Cantor, and J. J. Collins, Nature (London) **422**, 633 (2003).
[4] J. Peccoud and B. Ycart, Theor. Popul. Biol. **48**, 222 (1995).
[5] R. Blossey, L. Cardelli, and A. Phillips, T. Comp. Sys. Biol. IV **3939**, 99 (2006).
[6] R. Blossey, L. Cardelli, and A. Philips, HFSP J. **2**, 17 (2008).
[7] R. Blossey and C. V. Giuraniuc, Phys. Rev. E **78**, 031909 (2008).
[8] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).
[9] R. Metzler and P. G. Wolynes, Chem. Phys. **284**, 469 (2002).
[10] R. Bundschuh, F. Hayot, and C. Jayaprakash, Biophys. J. **84**, 1606 (2003).
[11] 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).
[12] J. Paulsson, Phys. Life Rev. **2**, 157 (2005).
[13] N. Friedman, L. Cai, and X. S. Xie, Phys. Rev. Lett. **97**, 168302 (2006).
[14] G. C. P. Innocenti and J. E. M. Hornos, J. Math. Biol. **55**, 413 (2007).
[15] T. Fournier, J. P. Gabriel, C. Mazza, J. Pasquier, J. L. Galbete, and N. Mermod, Bioinformatics **23**, 3185 (2007).
[16] R. Karmakar and I. Bose, Phys. Biol. **4**, 29 (2007).
[17] D. Schultz, J. N. Onuchic, and P. G. Wolynes, J. Chem. Phys. **126**, 245102 (2007).
[18] P. Paszek, Bull. Math. Biol. **69**, 1567 (2007).
[19] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. USA **105**, 17256 (2008).
[20] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Phys. Rev. E **79**, 031911 (2009).
[21] T. L. To and N. Maheshri, Science **327**, 1142 (2010).
[22] R. Karmakar, Phys. Rev. E **81**, 021905 (2010).
[23] R. Hermsen, D. W. Erickson, and T. Hwa, PLoS Comp. Biol. **7**, e1002265 (2011).
[24] A. F. Ramos, G. C. P. lnnocentini, and J. E. M. Hornos, Phys. Rev. E **83**, 062902 (2011).
[25] J. Ohkubo, Phys. Rev. E **83**, 041915 (2011).
[26] P. Bokes, J. R. King, A. T. A. Wood, and M. Loose, J. Math. Biol. **64**, 829 (2012).
[27] R. Grima, D. R. Schmidt, and T. J. Newman, J. Chem. Phys. **137**, 035104 (2012).
[28] G. Tkacik, A. M. Walczak, and W. Bialek, Phys. Rev. E **85**, 041903 (2012).