



Research article

A comparative analysis of noise properties of stochastic binary models for a self-repressing and for an externally regulating gene

Guilherme Giovanini¹, Alan U. Sabino², Luciana R. C. Barros³ and Alexandre F. Ramos^{1,2,*}

¹ Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Av. Arlindo Béttio, 1000, 03828-000, São Paulo, Brazil

² Instituto do Câncer do Estado de São Paulo, Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Arnaldo, 251, 01246-000, São Paulo, Brazil

³ Fundação Faculdade de Medicina, Faculdade de Medicina, Universidade de São Paulo, Av. Rebouças, 381, 05401-000, São Paulo, Brazil

* **Correspondence:** Email: alex.ramos@usp.br; Tel: +551130911235.

Abstract: This manuscript presents a comparison of noise properties exhibited by two stochastic binary models for: (i) a self-repressing gene; (ii) a repressed or activated externally regulating one. The stochastic models describe the dynamics of probability distributions governing two random variables, namely, protein numbers and the gene state as ON or OFF. In a previous work, we quantify noise in protein numbers by means of its Fano factor and write this quantity as a function of the covariance between the two random variables. Then we show that distributions governing the number of gene products can be super-Fano, Fano or sub-Fano if the covariance is, respectively, positive, null or negative. The latter condition is exclusive for the self-repressing gene and our analysis shows the conditions for which the Fano factor is a sufficient classifier of fluctuations in gene expression. In this work, we present the conditions for which the noise on the number of gene products generated from a self-repressing gene or an externally regulating one are quantitatively similar. That is important for inference of gene regulation from noise in gene expression quantitative data. Our results contribute to a classification of noise function in biological systems by theoretically demonstrating the mechanisms underpinning the higher precision in expression of a self-repressing gene in comparison with an externally regulated one.

Keywords: noise in gene regulatory models; stochastic gene regulation; binary gene model; fluctuations on gene expression

1. Introduction

Gene regulation is a main feature of multicellular living organisms to ensure that proper amounts of gene products are expressed at proper time and positions. That precision is achieved after the molecular machinery surpassing the major challenge posed by unavoidable random fluctuations of intracellular environment. They are caused by reactants being present in low copy numbers [1], and have been detected experimentally by application of advanced fluorescence and microscopy techniques to investigate the dynamics of molecular biological processes occurring in prokaryotic and eukaryotic cells [2–5]. Noise on the amounts of gene products enables one to determine gene regulation [6] and, because of its unavoidability, it is fair to conjecture the converse, that is, that the randomness of many distinct gene networks is actively regulated to ensure their capability of performing a myriad of specific cellular functions [7–10]. Indeed, when normal regulation of gene expression in metazoans is disrupted the collective dynamics of the cells of specific tissues leads to non-lethal (for the cell) deleterious (for the organism) behavior [11,12]. Hence, a classification of the gene regulatory strategies enabling control of both the amounts of gene products and their noise levels can be a key for deepening our understanding of living organisms and for speeding up synthetic biology.

In face of the diversity of dynamical regimes of gene expression such as the self-activating or self-repressing genes or the feed-forward loop, existing classificatory efforts are useful starting points [13]. E.g., feed-forward loops controlled by multiple transcription factors confer stronger stability in response to environmental stimulus [14] while feedback loops permit faster responses [15]. The negative feedback loops are employed on the control of circadian cycles and response to DNA damage [16]. The simplest negative feedback loop is the self-repressing gene, which enables noise reduction in gene expression [17–20]. But even the well studied self-repressing gene, may have unexpected regimes of arbitrarily reduced noise [21] or regimes emerging only from the stochastic nature of the intracellular biochemistry [22, 23]. Another widely recurring motif of gene networks is the externally regulated gene. The WNT/beta-catenin pathway is an example of an external regulating gene highly relevant for developmental biology because its conservation in all metazoans. Upon WNT activation, beta-catenin is released, migrate to the cell nucleus where it interacts with transcriptional factors, as TCF-4, and induces the expression of WNT target genes. When WNT signal is OFF, beta-catenin is degraded [24]. This pathway is altered in many cancers cells in a tumor-specific manner, it is used as a prognostic factor and demands further studies to explore its potential therapeutic impacts [25].

Those observations indicate the existence of general principles underlying the functioning of biological systems. That encourages the application of the theoretical machinery of stochastic processes and Lie symmetries as an additional guide for the classification of noise regulation and gene networks function. Such an analysis is important as it enables the classification of noise properties related with specific gene regulatory strategies [6, 10, 26–28]. As we will show in this manuscript, distinct gene regulatory strategies may generate equivalent noise on the number of products and determining those regimes may be useful for parameter inference determining a given gene's regulatory strategy. In this manuscript we present a comparative analysis of noise behavior in the context of the exactly solvable stochastic binary models for the self-repressing and the externally regulating gene [29–32]. The steady state solutions are analyzed comparatively and interpretation of the physical behavior of the two gene regulatory strategies is presented. Self-repression enables

protein synthesis at lower fluctuation levels than the externally regulated gene even if both strategies generate the same average number of proteins. We also analyze the conditions for which noise in protein numbers are equivalent or similar for both gene strategies. Our results contribute for the classification of functional noise and on the construction of synthetic circuits, being helpful for the design of cancer therapy or diagnostics based on regulation of gene expression [25, 33]; to the understanding of emergence of precision on the control of gene expression randomness [19, 20]; for the inference of the topology of gene networks based on gene expression noise [34, 35]; to the establishment of a classification of regulation of noise in gene expression accordingly with the function of its products in a cell or network [36]. That sets the question of whether a given gene regulatory mechanism was favoured for regulating noise in the duecourse of evolution [37–40], and sets a “chicken and egg” type of question of what came first, noise regulation for functional performance or the reverse?

This manuscript is organized as follows. In section 2 we briefly discuss the regulation of gene expression and show the well-known exact solutions for the stochastic binary models for a self-repressing gene (SRG) [30] and for an externally regulated gene (ERG) [29] that can be either repressed or activated by transcription factors encoded in another gene. Section 3 shows the multiple qualitatively distinctive probability distributions and corresponding moments that govern the steady state protein numbers on both the SRG and ERG. A discussion of our results is given at section 4 and demonstrations of our calculations are presented in the Appendix A.

2. Models and methods

Gene expression is the process by which the information encoded on DNA is converted into functional units, as RNA and proteins. Though gene transcription and mRNA translation are key processes of gene expression, the normal functioning of a cell requires different genes to be expressed at proper quantities and instants. This regulation can be pre and post transcription, and even post translation by acceleration of protein degradation for the control of cellular processes such as cell cycle [41]. Specific DNA sequences interact with regulatory proteins to activate the binding of RNAPolIII complex to the promoter site [42, 43] and transcription is initiated. Once complete, mRNAs are transported to the cytoplasm to be translated. Translation can also be regulated by miRNAs-mRNA interaction [44]. Such a multi-stages regulation of gene expression can be formulated effectively in terms of their coupling whenever possible. At first, this approximated description of gene expression appears to be very limited, but the wide diversity of qualitatively distinct gene expression regimes enables its use on the interpretation of particular experimental designs. Furthermore, that phenomenological approximation drives the proposition of exactly solvable mathematical models to compute gene expression randomness.

The simplest stochastic models for gene regulation assume transcription and translation as coupled processes and regulation of protein synthesis done by the spontaneous switching of the promoter of the gene [29]. Self-regulation can be introduced if one of the switching rates is written as a function of the protein numbers. As a first approximation, one may assume that the protein numbers do not change because of the promoter switching [30]. Biologically, that condition may correspond to the expressed protein acting as an enzyme that catalyzes the production of the actual repressor or, if it is the actual repressor, its numbers are sufficiently large to turn negligible the effect of the binding of one protein.

Alternatively, one may consider the change on the protein numbers by binding to the regulatory site in the stochastic binary model for the SRG [45] whose noisiness properties are fully investigated in terms of its complete exact solutions [46]. The probability distributions obtained using these two families of stochastic models may be very different [47, 48] and searching for the most appropriate approaches for modeling stochastic gene expression is a task to be performed with the help of experimental data [46]. For example, one may show that these simple models have the often detected bursting regime [49] as a limit for very high synthesis rate happening at very short time interval [50] or explicitly add bursty synthesis assuming that it is sufficiently fast in comparison with the remaining time scales involved in protein synthesis [51]. Alternatively, one might approach the multiple nascent RNAs that give rise to transcriptional burst by modeling the multiple RNAPol bound to the DNA as a traffic/random walk problem on a 1D grid [52–54]. That enables the calculation of probability distributions governing transcriptional bursts and the fluorescence intensities of probes of nascent RNAs [54]. An alternative route for modeling stochastic gene expression is to consider transcription and translation as separate processes without [55] or with promoter switching between ON and OFF states [56, 57].

In this manuscript we focus on the simplest class of stochastic binary models for regulation of gene expression that are fully solvable [29–32] and provide a brief comparison with *i.* the approach in which binding of the regulatory protein changes their numbers [45, 46]; *ii.* the bursty models [51]; *iii.* the functioning of the linear approximations [58] for self-regulating genes based on the spontaneous switching model [29]. Despite their simplicity, these models are useful building blocks [59] for the understanding of smaller experiments and for the construction of incrementally more complex models incorporating additional effects [49, 54]. The building blocks enable the biologists to gain a deeper intuition on the physical behavior of the molecular components participating in regulation of gene expression while their composition enables an integrative perspective about the dynamics of biological processes. Additionally, they may motivate the design of experiments useful for further development of more useful stochastic models for gene expression.

2.1. A qualitative description of regulation of gene expression

In its simplest description, regulation of gene expression may be considered as a simple ON and OFF switching of the promoter as induced by regulatory proteins [60]. We also consider transcription and translation as coupled processes by assuming no post-transcriptional regulation and a sufficiently small variation on the number of proteins produced from one mRNA. Considering the randomness of the gene expression process, we set the state of this system using two random variables (m, n) , where $m = \alpha$ or $m = \beta$ denotes the state of the promoter as ON or OFF, and $n = 0, 1, \dots$ indicates the number of proteins.

The cascades of chemical reactions that comprise protein production are summarized as a minimal set of effective reactions. The protein synthesis happens at constant rate denoted by k (or 0) when the promoter of the gene is ON (or OFF) while ρ indicates the rate of protein loss of function. The promoter switches from OFF to ON state at a constant rate denoted by f . Our choice for the rate of the ON to OFF switch indicates the gene regulation strategy as ERG or SRG. We denote the ON to OFF switching rate of the ERG by h_2 . The SRG has the ON to OFF switching rate is denoted by h_1 and it is a function of the number of proteins n . In its simplest approximation, we consider a linear dependence such that we write $h_1 n$. Figure 1 shows a cartoon describing both genes.

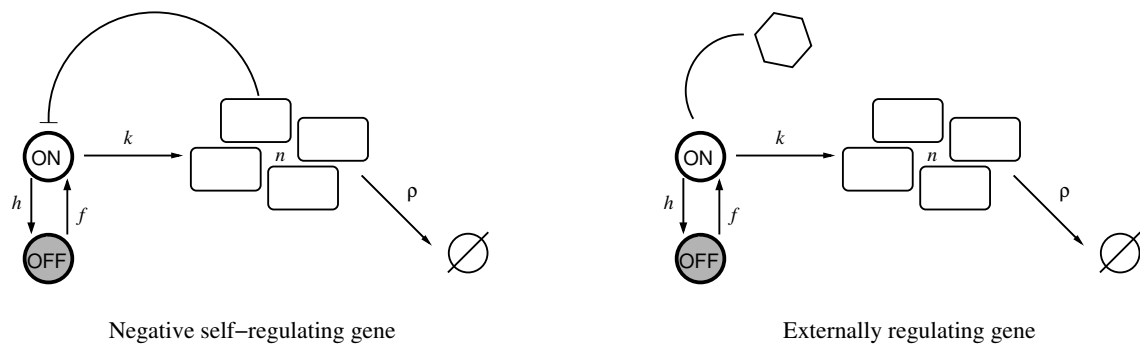
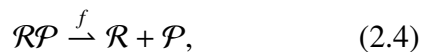


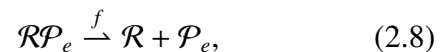
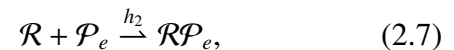
Figure 1. The left cartoon is a representation of a SRG while the right cartoon represents an ERG.

The cartoon shown in Figure 1 corresponds to a set of effective chemical reactions. The left and right effective reactions correspond to the SRG and ERG. For the SRG we denote a protein by \mathcal{P} . The regulatory region of the gene is denoted by \mathcal{R} and the gene state is determined by the binding of \mathcal{P} to the regulatory region. The regulatory protein of the ERG is denoted by \mathcal{P}_e . The symbols for reaction rates appear on top of arrows indicating the reactants and products of the effective reactions.

Self-repressing gene



Externally regulating gene



Equations (2.1) and (2.5) indicate protein synthesis when the promoter state is ON while protein degradation is indicated by Eqs (2.2) and (2.6). The gene switching from ON to OFF state is indicated by Eqs (2.3) and (2.7) while the opposite transition is presented in Eqs (2.4) and (2.8).

2.2. Fluctuations and their quantification

Usually, the fluctuations on the amounts of gene products at a time instant t are quantified by the Fano factor. This quantity is denoted by \mathcal{F} , and defined as the ratio of the variance to the average protein number, namely:

$$\mathcal{F} = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle}. \quad (2.9)$$

We can compute \mathcal{F} by means of the time-dependent or steady state probabilities governing n or by estimating expression levels in a sample of genes if, for example, a probe's fluorescence intensity can be measured [61]. The Fano factor is a measure of how different from a Poissonian a probability distribution is. One may nominate a probability distribution accordingly with the value of its Fano factor, such that, a Fano (or Poissonian) distribution has $\mathcal{F} = 1$, a quasi-Fano (or quasi-Poissonian) distribution has $\mathcal{F} \approx 1$, a super-Fano (or super-Poissonian) has $\mathcal{F} > 1$, a sub-Fano (or sub-Poissonian) has $\mathcal{F} < 1$, and an infra-Fano (or infra-Poissonian) has $\mathcal{F} \ll 1$ [21].

The Fano factor provides a criterion to evaluate fluctuations in gene expression levels and the moments of a distribution that must be quantified for its full characterization. For example, the Poissonian and sub-Poissonian distributions are peaked around the average such that first and second order moments are sufficient to roughly estimate the numbers of gene products that are governed by them, since n is mostly within the interval $[\langle n \rangle - \sqrt{\mathcal{F}\langle n \rangle}, \langle n \rangle + \sqrt{\mathcal{F}\langle n \rangle}]$. That may not be the case of the super-Poissonian distributions that can have a multiplicity of shapes, for instance, the negative binomial, bimodal, table shaped, and so on. These distributions are spread widely and the computation of the first and second order moments is not sufficient for a rough estimate of the numbers of gene products at a given instant. Evaluating super-Poissonian noise in gene expression and their origins is a harder task and the use of exactly solvable mathematical models providing probability distributions can produce further insights on understanding the mechanisms of gene regulation.

2.3. A stochastic model for regulation of gene expression

Here we consider a well-known stochastic model for protein synthesis from a two state promoter [29, 30]. Since the random variables of the model are n and m , the system state is characterized by a probability distribution

$$\Pi(\alpha_n(t), \beta_n(t)), \quad (2.10)$$

where the probability of finding n proteins at time t when the gene is ON, or OFF, is indicated by $\alpha_n(t)$, or $\beta_n(t)$. The master equation governing the dynamics of the probabilities is:

$$\frac{d\alpha_n(t)}{dt} = k(\alpha_{n-1}(t) - \alpha_n(t)) + \rho[(n+1)\alpha_{n+1}(t) - n\alpha_n(t)] - (h_1n + h_2)\alpha_n(t) + f\beta_n(t), \quad (2.11)$$

$$\frac{d\beta_n(t)}{dt} = \rho[(n+1)\beta_{n+1}(t) - n\beta_n(t)] + (h_1n + h_2)\alpha_n(t) - f\beta_n(t). \quad (2.12)$$

We assume the rates k , ρ , f , h_1 , and h_2 to be constants during the time interval of integration of Eqs (2.11) and (2.12). The SRG is modeled considering $h_1 \neq 0$ and $h_2 = 0$, such that the switching rate from ON to OFF state depends on n . We assumed that n do not change with the interaction between the protein and the gene.

The contrary condition, $h_1 = 0$ and $h_2 \neq 0$, results in a model for the ERG which switching constants indicate a random spontaneous switching between ON and OFF states. However, one may also write the switching constants as functions of the concentrations of the regulatory transcription factors encoded in another gene to phenomenologically indicate repression or activation. The value of h_2 is dependent on the instantaneous amounts of \mathcal{P}_e and we approximate it to be constant under the following assumptions: The probabilities governing the amounts of \mathcal{P}_e reached the steady state regime; the average amounts of \mathcal{P}_e are large enough to ensure that small fluctuations would not change its binding probability to the regulatory site of the gene; the fluctuations on the amounts of \mathcal{P}_e are within a sufficiently small range. In case of significant changes on the amounts of \mathcal{P}_e leading to a distinguishable regulatory landscape, we might use a piecewise integration of Eqs (2.11) and (2.12). Hence, the rates are assumed as constant during a time interval in which the changes of regulatory environment are negligible [33, 62].

2.4. Marginal distributions and moments

The exact solutions for both the steady state [29, 30] and the time-dependent regime [31, 59] have already been presented elsewhere. Here we apply the steady state solutions to compute the full and the marginal probability distributions and their related moments for a comparative analysis of the noise on the protein numbers produced from an ERG or a SRG. The marginal probabilities of finding the promoter at ON or OFF state independently of the amount of proteins inside the cell are denoted by p_α and p_β , respectively, such that:

$$p_\alpha = \sum_{n=0}^{\infty} \alpha_n \quad \text{and} \quad p_\beta = \sum_{n=0}^{\infty} \beta_n. \quad (2.13)$$

The marginal probability of finding n proteins inside the cell independently of the promoter state being ON or OFF, denoted by ϕ_n , is given by

$$\phi_n = \alpha_n + \beta_n. \quad (2.14)$$

The marginal and the auxiliary p -th order moments of protein numbers n , denoted by $\langle n^p \rangle$, $\langle n_\alpha^p \rangle$, $\langle n_\beta^p \rangle$ are, respectively,

$$\langle n^p \rangle = \sum_{n=0}^{\infty} \phi_n n^p, \quad \langle n_\alpha^p \rangle = \sum_{n=0}^{\infty} \alpha_n n^p, \quad \langle n_\beta^p \rangle = \sum_{n=0}^{\infty} \beta_n n^p. \quad (2.15)$$

These quantities will be useful on the calculations of the covariance between the two random variables of the model.

2.5. Exact solutions of the model

The exact solutions of Eqs (2.11) and (2.12) were obtained recently using the generating function technique. The generating functions of the time-dependent probabilities for the model for a SRG are given in terms of the confluent Heun functions [32] while for the ERG these solutions are given in terms of the confluent hypergeometric functions of the first kind [31] also known as KummerM or WhittakerM functions ([63], page 503). Here we focus on the steady state solutions which are also given by KummerM functions for both the SRG [30] and the ERG [56] models. The KummerM function is denoted by $M(A, B, \eta)$, see [63], and written as a power series:

$$M(A, B, \eta) = \sum_{n=0}^{\infty} \frac{(A)_n \eta^n}{(B)_n n!}, \quad (2.16)$$

where the $(A)_n = A(A+1)\dots(A+n-1)$ is the Pochhammer symbol. Note that for $A = B$ the power series defining the KummerM function coincides with that for the exponential, hence:

$$M(A, A, \eta) = e^\eta. \quad (2.17)$$

The next Subsections present the steady state solutions for both the ERG and SRG models, which probabilities and parameters are distinguished by superscripts e and s , respectively. The parameters that are the same for both models have no superscripts.

2.5.1. Externally regulating gene

The distribution $\Pi(\alpha_n^e, \beta_n^e)$ for the ERG has three parameters, a, N, b^e , defined in terms of the reaction rate constants as [56]:

$$a = \frac{f}{\rho}, \quad N = \frac{k}{\rho}, \quad b^e = \frac{f + h_2}{\rho}. \quad (2.18)$$

Here, the protein half life, $\approx 1/\rho$, sets the time scale of model. The parameter a is the ratio of the OFF to ON switching rate to the decaying rate and N is the steady state average number of proteins if the promoter remains exclusively ON. The parameter b^e is the ratio of the decaying rate of the promoter switching to steady state to the protein degradation rate.

The probabilities α_n^e, β_n^e , and ϕ_n^e are:

$$\alpha_n^e = \frac{a}{b^e} \frac{N^n}{n!} \frac{(a+1)_n}{(b^e+1)_n} \mathbf{M}(a+n+1, b^e+n+1, -N), \quad (2.19)$$

$$\beta_n^e = \frac{b^e - a}{b^e} \frac{N^n}{n!} \frac{(a)_n}{(b^e+1)_n} \mathbf{M}(a+n, b^e+n+1, -N), \quad (2.20)$$

$$\phi_n^e = \frac{N^n}{n!} \frac{(a)_n}{(b^e)_n} \mathbf{M}(a+n, b^e+n, -N), \quad (2.21)$$

were obtained using their respective generating functions found in [29, 56, 59]. The marginal probabilities of finding the gene ON or OFF are

$$p_\alpha^e = \frac{a}{b^e}, \quad p_\beta^e = \frac{b^e - a}{b^e}. \quad (2.22)$$

The probability for the promoter to be ON is proportional to the switching rate from OFF to ON state and it indicates the average fraction of time for the promoter being ON during an interval. Note, $0 \leq p_\alpha^e \leq 1$ implies a $0 \leq a \leq b^e$.

2.5.2. Self-repressing gene

The distribution $\Pi(\alpha_n^s, \beta_n^s)$ for the SRG has the parameters a, N , as for the ERG in Eq (2.18), and the parameters b^s and z_0 , where:

$$a = \frac{f}{\rho}, \quad N = \frac{k}{\rho}, \quad z_0 = \frac{\rho}{\rho + h_1}, \quad b^s = \frac{f}{\rho + h_1} + \frac{N\rho h_1}{(\rho + h_1)^2} = az_0 + Nz_0(1 - z_0). \quad (2.23)$$

These parameters are defined in terms of the protein removal rate from cytoplasm because of protein degradation or protein binding to the regulatory region of the gene $\propto 1/(\rho + h_1)$ [23].

The probabilities α_n^s, β_n^s , and ϕ_n^s , are written using the generating functions presented in [22] with replacement of that N by Nz_0 and $\chi = 0$, such that:

$$\alpha_n^s = \frac{a}{b^s} \frac{z_0}{c} \frac{(a+1)_n}{(b^s+1)_n} \frac{(Nz_0)^n}{n!} \mathbf{M}(a+n+1, b^s+n+1, -Nz_0^2), \quad (2.24)$$

$$\phi_n^s = \frac{1}{c} \frac{(a)_n}{(b^s)_n} \frac{(Nz_0)^n}{n!} \mathbf{M}(a+n, b^s+n, -Nz_0^2), \quad (2.25)$$

$$\beta_n^s = \phi_n^s - \alpha_n^s. \quad (2.26)$$

$c = M(a, b^s, Nz_0(1 - z_0))$ is the normalization constant of the distribution. The marginal probabilities of finding the gene ON or OFF are:

$$p_\alpha^s = \frac{a}{b^s} \frac{z_0}{c} M(a + 1, b^s + 1, Nz_0(1 - z_0)), \quad \text{or} \quad p_\beta^s = 1 - p_\alpha^s. \quad (2.27)$$

Since $Nz_0(1 - z_0)$ is non-null and only for limiting cases of $z_0 \rightarrow 1$ or $z_0 \rightarrow 0$ it approaches zero, it is not possible to write p_α^s as a simple function of a and b^s . That happens because the ON to OFF switching rate depends of n .

Despite the convenience of writing Eqs (2.24)–(2.27) in terms of four parameters (a, b^s, N, z_0) , those probabilities have only three free parameters, as it happens with the ERG. In both models, the fourth parameter is obtained as a decaying rate to steady state regime [31, 32]. That is because the probabilities of the models for the SRG or ERG are computed using the KummerM function, which has two parameters, respectively, (a, b^s) or (a, b^e) , and the argument depending on one additional parameter: N in the ERG model and a combination of two parameters in the SRG, namely Nz_0^2 . Therefore, one must set the constraints among a , b^s , N , and z_0 given by Eq (2.23). We choose N to be written as a function of the remaining parameters on the following basis. In previous studies, we demonstrated the Lie symmetries underlying the existence of the analytical solutions of the stochastic models for the SRG [22, 23] and the ERG [59]. Since the invariants of the $\mathfrak{so}(2, 1)$ algebra satisfied by the two models are b^s and b^e , they are natural candidates to be set as constants. In the model for the SRG, the maximum steady state protein numbers N will be given in terms of b^s . The switching of the promoter of the SRG can be characterized in terms of the parameters a and z_0 with the former being proportional to f and the latter being inversely proportional to h_1 . We use Eq (2.23) to obtain

$$N = \frac{b^s - az_0}{z_0(1 - z_0)} > 0 \Rightarrow 0 < a < b^s/z_0, \quad (2.28)$$

where the left inequality is imposed by the requirement of positivity of N and the right hand inequalities are because $0 < z_0 < 1$, see Eq (2.23).

2.6. Covariance

The model described here has two random variables, the promoter state, being ON or OFF, and the protein number, a non-negative integer. To compute the covariance between the two coupled processes we consider the phenomenology of the model to propose a random variable v associated with the promoter states. If the system reaches a stationary regime with the promoter being exclusively ON (or OFF) the corresponding number of proteins is governed by a Poissonian distribution of average N (or 0). In the timescale of a protein's lifetime, those averages are the rate of synthesis of proteins when the promoter is ON (or OFF). Therefore, one may describe the ON and OFF states of the promoter by means of the steady state average rates of protein synthesis and we write our auxiliary random variable as $v_m = N \mathbb{I}_\alpha(m)$, where $\mathbb{I}_\alpha(m)$ is the indicator function defined as $\mathbb{I}_\alpha(m) = \begin{cases} 1, & \text{if } m=\alpha, \\ 0, & \text{if } m=\beta. \end{cases}$. The probability of $v_\alpha = N$ and $v_\beta = 0$ is, respectively, p_α and p_β . We denote the covariance between v_m and n by ξ such that:

$$\xi = \langle v_m n \rangle - \langle v_m \rangle \langle n \rangle = \sum_{n=0}^{\infty} (v_\alpha \alpha_n + v_\beta \beta_n) n - (v_\alpha p_\alpha + v_\beta p_\beta) \left(\sum_{n=0}^{\infty} n \phi_n \right). \quad (2.29)$$

Because of the phenomenology of v_m , here on we interpret ξ as the covariance between the promoter state and the protein numbers. Higher order mixed moments, $\langle v_m^p n^q \rangle$, are given by $\sum_{n=0}^{\infty} (v_\alpha^p \alpha_n + v_\beta^p \beta_n) n^q$, where $p, q \in \{1, 2, \dots\}$.

2.7. Quantifying fluctuations on gene expression

In a recent publication, we demonstrated that the steady state Fano factor for the model for a SRG can be written in terms of its average protein numbers and the covariance between v_m and n [21]. We can extend that relation for the model for an ERG, such that the Eq (2.9) becomes:

$$\mathcal{F} = 1 + \frac{\xi}{\langle n \rangle}, \quad (2.30)$$

where the covariance is:

$$\xi = N (\langle n_\alpha \rangle - \langle n \rangle p_\alpha). \quad (2.31)$$

The demonstration of these equations is included in the Appendix A.1 and it is similar to that presented in [21].

The calculation of $\langle n_\alpha \rangle$ can be done using the exact solutions of Eqs (2.11) and (2.12). Hence, the form of Eq (2.30) enables the classification of its related probability distribution accordingly with the covariance ξ . The genes having a positive covariance between its protein numbers and promoter state are in a super-Fano regime while a sub-Fano regime depends on negative covariance. The Fano distributions are characterized by null covariance.

2.7.1. Externally regulating gene

In [56] it was shown that the average protein number for the ERG, denoted by $\langle n^e \rangle$, is:

$$\langle n^e \rangle = p_\alpha^e N. \quad (2.32)$$

Note that for $p_\alpha^e = 1$ the average protein number is N , which corroborates the biological interpretation given in Eq (2.18).

Equation (2.31) is used to compute the covariance between variables (v_m, n) in the ERG such that:

$$\xi^e = \frac{N^2}{b^e + 1} p_\alpha^e (1 - p_\alpha^e). \quad (2.33)$$

This expression is obtained using Eq (2.22) and $\langle n_\alpha^e \rangle = \frac{a}{b^e} \frac{1+a}{1+b^e} N$ which calculation is given in the Appendix A.2. Now we can write the Fano factor for the ERG as

$$\mathcal{F}^e = 1 + \frac{N}{b^e + 1} (1 - p_\alpha^e), \quad (2.34)$$

which is a linear function of the probability for the gene to be in the OFF state. The covariance for the ERG is always non-negative and its probability distributions governing n are only Fano and super-Fano.

2.7.2. Self-repressing gene

The average protein number expressed from the SRG is denoted by $\langle n^s \rangle$, and it is [21]:

$$\langle n^s \rangle = p_\alpha^s N, \quad (2.35)$$

as can be noticed employing Eqs (A.2) in the Appendix and (2.27).

To compute the covariance between variables (ν_m, n) we use Eq (2.31) which results:

$$\xi^s = az_0 \frac{N - \langle n^s \rangle}{1 - z_0} - \langle n^s \rangle^2. \quad (2.36)$$

This expression has been presented in [21] and here we present a more detailed demonstration in the Appendix A.3. The Fano factor for the SRG is then written as:

$$\mathcal{F}^s = 1 + \left[\frac{az_0}{\langle n^s \rangle} \frac{N - \langle n^s \rangle}{1 - z_0} - \langle n^s \rangle \right], \quad (2.37)$$

which is a non-linear function of the parameters of the model. In the model for the SRG, the covariance can be positive, null, and negative which results on probability distributions being super-Fano, Fano, and also sub-Fano [21]. The later regime shows why the self-repression enables noise reduction in gene expression.

3. Results

3.1. Correlations for external regulation and self-repression

Figure 2 shows the Pearson correlation between n and ν_m for both gene regulatory strategies. Graph A shows the Pearson correlation, denoted by $P^e(p_\alpha)$, of the model for the ERG as function of the probability for the promoter to be ON, given by

$$P^e(p_\alpha) = \frac{1}{\sqrt{1+b^e}} \sqrt{\frac{N(1-p_\alpha^e)}{1+b^e+N(1-p_\alpha^e)}}. \quad (3.1)$$

For a pair of fixed (N, b^e) we have $P^e(p_\alpha)$ approaching 0 as $p_\alpha^e \rightarrow 1$ while it asymptotically approaches $\frac{1}{\sqrt{1+b^e}} \sqrt{\frac{N}{1+b^e+N}}$ as $p_\alpha \rightarrow 0$. The curve $P^e(p_\alpha)$ has a reversed sigmoidal shape with asymptotic maximum value denoted by P_m^e and half height occurring at \hat{p}^e , which are

$$P_m^e = \frac{1}{\sqrt{1+b^e}} \sqrt{\frac{N}{1+b^e+N}}, \quad \text{and} \quad \hat{p}^e = \frac{3}{4 - \frac{N}{1+b^e+N}}. \quad (3.2)$$

Hence, we write

$$\frac{3}{4} < \hat{p}^e < 1 \Rightarrow P^e(\hat{p}^e) = \frac{P_m^e}{2}. \quad (3.3)$$

Equation (3.2) indicates that P_m^e is bounded above at $1/\sqrt{1+b^e}$ such that for $b^e \gg 1$ we have $P_m^e \rightarrow 0$ even if $N \gg b^e$ and it tends to one when $b^e \rightarrow 0$ and $N \gg b^e$. Graphs B and C show, respectively,

the correlation as function of the probability for the promoter to be ON and as function of the average protein number for the SRG. Lines of the same color indicate the same value for b^s on both graphs. The correlation value depends on the relation between parameters a and b^s : It is zero for $a = b^s$, it is positive for $a < b^s$ and negative otherwise. The correlation behaves almost linearly as a function of the ON state probability. The condition $a = b^s$ results on $p_\alpha^s = z_0$. Graph C shows that the SRG enables the existence of two probability distributions having the same $\langle n \rangle$ and opposite signals correlations, implying on the possibility of protein synthesis with large or small noise while keeping a fixed average. For a given value of z_0 the maximum average protein numbers happens for $a = b^s$ and $\langle n^s \rangle = Nz_0$. For $b^s \geq 1$ the absolute value of the maximal or minimal Pearson correlations is reduced and tends to 0. On both Graphs B and C the correlation approaches a linear behavior for lower values of b^s (see red, blue, and green curves) while non-linear effects start to appear as we increase b^s (see cyan and magenta curves).

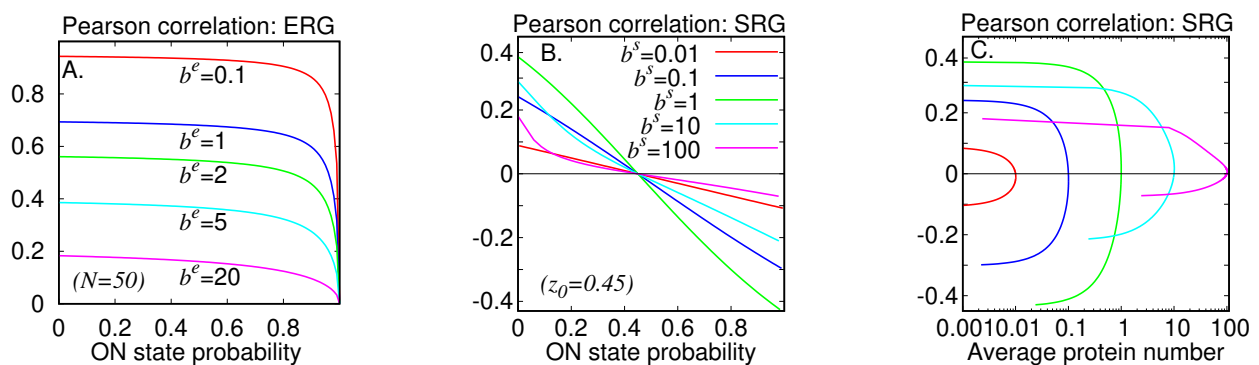


Figure 2. Graph A. shows the Pearson correlation as a function of probability of ON state. Pearson correlation is defined as the ratio of the covariance (see Eq (2.33)) to the standard deviation of n , denoted by σ , and to standard deviation of v_m , which results: $\frac{\xi^e}{\sigma \sqrt{N^2 p_\alpha (1-p_\alpha)}}$. Graph B. shows the correlation versus the probability for the SRG being ON. Graph C. shows the correlation versus the average protein number using the same parameter values of Graph B, where lines of the same colors have the same value for b^s .

Graph A shows that the Pearson correlation for the ERG is $\sim P_m^e$ for most values of p_α . Indeed, the correlation is half of its maximum values when $p_\alpha = \hat{p}^e$ and obeys Eq (3.3). Hence, the reversed sigmoidal shape of the correlation starts decaying when the probability for the promoter to be ON is closer to one. However, one must notice that the maximum value of the correlation is upper bounded by $(1 + b^e)^{-1/2}$, and for $b^e \gg 1$ the maximal correlation approaches zero. That corresponds to a promoter that rapidly switches between ON and OFF states during the lifetime of proteins whose presence would not be a readout of the state of the promoter of the gene. Besides, that may also correspond to a quasi-Poisson regime in which $\mathcal{F}^e \rightarrow 1$ for the case of $b^e \gg N(1 - p_\alpha^e)$ (see Eq (2.34)). Alternatively, for the case of very efficient protein synthesis and small probability for the gene to be ON, the gene would be expressed in a bursty fashion [50], a regime widely detected experimentally [49]. On the other hand, as $b^e \ll 1$ we have a slow switching promoter that stands ON or OFF for a time interval that is larger than the proteins lifetime. In that case, there will be a finite number of proteins or mostly none proteins when the promoter is ON or OFF, respectively. That result is in agreement with our calculations using information theory where we show that binary distributions enable enhanced reliability of information

flow in gene networks [36]. That regime corresponds to a weaker coupling between the two stochastic processes of the binary model for gene expression.

Graph B shows that as p_α^s increases, the correlation for the SRG decays from a positive value to 0 when $p_\alpha^s \rightarrow z_0$ (and $a \rightarrow b^s$). As we keep increasing the p_α^s , the correlation keeps diminishing and the reduced noise regime is established, a behavior that is exclusive for the SRG. Note that on the model for the SRG, the maximum value of the correlation is smaller than that for the ERG and it happens for $b^s = 1$. As discussed in a previous publication [23], the biochemical interpretation of b^s is the same of that for b^e . Therefore, when the two time scales of the binary model for expression of a SRG are the same, the correlation between the promoter states and the protein numbers can reach larger absolute values. Note, however, that in the ERG, the correlation is larger when $b^e \rightarrow 1$. The stronger coupling of the OFF switching rate caused by dependence of the number of proteins causes this weaker correlation in comparison with the model for the ERG. Graph C unveils the degeneracy resulting from the symmetries of the stochastic model for a SRG [22, 23], a limit that is not established when we consider only the deterministic model for a SRG [23].

3.2. Fano and quasi-Fano probability distributions

Fano and quasi-Fano probability distributions governing protein numbers obtained from the stochastic model for a SRG and an ERG are shown in Figure 3. Graph A shows Fano distributions governing protein numbers expressed from a SRG, which happens when $a = b^s$. At this limit, Eq (2.27) becomes $p_\alpha^s = z_0$ because the KummerM function for $a = b$ is equal to the exponential, see Eq (2.17), and $M(A, B, \eta) = M(1 + A, 1 + B, \eta) = e^\eta$. The probabilities of finding n proteins in Eq (2.25) become $\phi_n^s = \frac{(Nz_0)^n}{n!} \exp(-Nz_0)$ which corresponds to the Poissonian distribution of average $\langle n^s \rangle = Nz_0 = b^s$, because $N = \frac{b^s}{z_0}$ as verifiable in Eq (2.28). Alternatively, one may use those relations in Eq (2.37) to verify that $\mathcal{F}^s = 1$. Graph B (SRG) presents quasi-Fano probability distributions obtained by fixing b^s and taking $a \approx b^s$. We consider a being both larger and smaller than b^s such that the covariance will be, respectively, negative and positive. The average protein number is reduced as we increase a because the parameter of the ‘‘Poisson’’ distribution is $\frac{b^s - az_0}{z_0(1 - z_0)}$. Since the distributions are single peaked, the reduction on average protein numbers implies a left displacement of the peak of the probability distributions. Graph C (ERG) also presents quasi-Fano distributions for large b^e and $a \approx b^e$ restricted to $a < b^e$. The case $a = b^e \Rightarrow h_2 = 0$ is also shown by the magenta solid circles. It corresponds to a non-switching promoter, i.e., it models a constitutive gene whose protein numbers are governed by a Poissonian distribution with parameter N . The remaining distributions are quasi-Fano with positive definite covariance and the average number of proteins increasing linearly with a . That is because of the linear dependence between the average protein number produced from an ERG and the probability for the promoter being ON given by Eq (2.32).

Graph A shows that the noise on the protein numbers synthesized from an SRG is the same as if they were synthesized from a non-regulated constitutive gene, a regime that is possible for the SRG but not for the ERG. This happens when $a = Nz_0 = \langle n^s \rangle$ as derived from Eqs (2.23), (2.27) and (2.35), which in terms of the biochemical parameters, becomes

$$\langle n^s \rangle = \frac{f}{\rho} = \frac{k}{\rho + h_1}. \quad (3.4)$$

Hence, the switching SRG behaves effectively as a constitutive gene with synthesis and degradation

rates given by, respectively, k and $\rho + h_1$. Furthermore, it corresponds to a “resonant” regime in which the ON-OFF-ON average switching rate of the SRG defined as $f + kh_1/(\rho + h_1)$ in [23], is equal to the synthesis rate k as we can verify after using the relation between the reaction rates given by Eq (3.4).

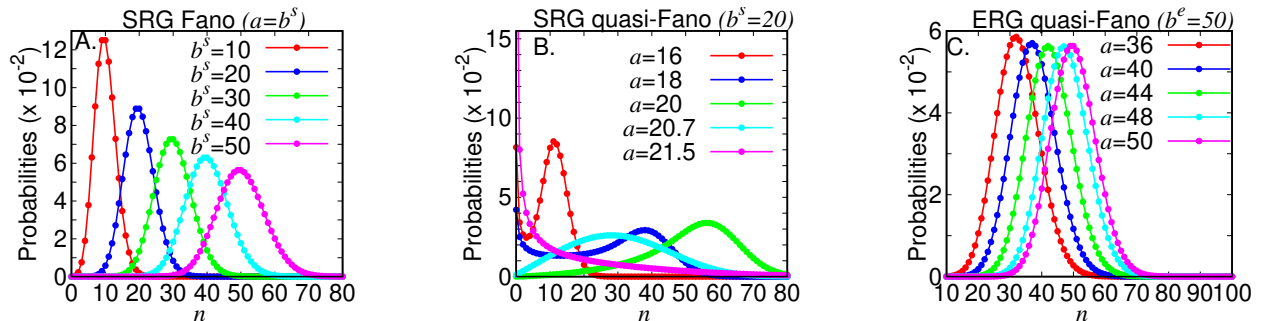


Figure 3. We consider probability distributions denoted Fano ($\mathcal{F} = 1$) or quasi-Fano ($\mathcal{F} \approx 1$). The graphs for the SRG have $z_0 = 0.9$. Graph A has strictly Fano distributions and for the condition $a = b^s$ the moments of the distributions are $(\mathcal{F}^s, \xi^s, \langle n^s \rangle, p_\alpha^s) = (1, 0, b, z_0)$. Graph B has quasi-Fano distributions for SRG which moments are presented in increasing order of a as shown in legend: $\mathcal{F}^s \approx (2.8, 8.7, 3.6, 6.7, 25.9)$; $\xi^s \approx (17.7, 207.7, 133.5, 173.3, 352.6)$; $\langle n^s \rangle \approx (9.6, 26.8, 51.4, 30.3, 14.2)$; $p_\alpha^s \approx (0.76, 0.62, 0.81, 0.33, 0.065)$. Graph C has quasi-Fano distributions for ERG with $N = 50$. The moments of the distributions are presented in increasing order of a as shown in legend: $\mathcal{F}^e \approx (1.3, 1.2, 1.1, 1.0, 1)$; $\xi^e \approx (9.9, 7.8, 5.2, 1.9, 0)$; $\langle n^e \rangle \approx (36, 40, 44, 48, 50)$; $p_\alpha^e \approx (0.72, 0.80, 0.88, 0.96, 1)$.

Graphs B and C shows quasi-Fano probability distributions as expected if we have $a \sim b$. The model for the ERG also has the condition in which the synthesis and the switching rates are the same, namely $f + h_2 = k$, implying on $b^e = N$ (see Eq (2.18)). However, the Fano factor for the probability distributions at this limit is $\mathcal{F}^e = 1 + \frac{N}{1+N}(1 - p_\alpha^e)$ which, for $N \gg 1$ results on $\mathcal{F}^e \rightarrow 2 - p_\alpha^e$. Hence, the ERG gene has no resonant limit resulting in an effectively constitutive gene expression regime as it happens for the SRG.

3.3. Super-Fano probability distributions

Super-Fano probability distributions for both regulatory strategies are presented in Figure 4. We select parameter values resulting on widespread probability distributions, such as the table shaped and bimodal ones. Graph A shows probability distributions for the SRG where we take a fixed value for b^s . The choices of a or z_0 help to understand their effect on probability distributions. The values of z_0 are defined such that the degradation rate is the main cause of protein removal from cytoplasm. Graph B shows a second family of super-Fano distributions obtained for the SRG. We take two values for z_0 and vary (a, b) . Graph C shows probability distributions for the ERG, with fixed value of N and multiple values of a and b^e .

Figure 4, Graph A has five probability distributions. The distributions displayed in magenta and cyan are bimodal. The green line shows a distribution approximately like a table. The blue and red lines indicate single modal widespread distributions because of the significant probability of finding

from zero to ≈ 100 proteins while their averages are ≈ 50 . Figure 4, Graph B shows two bimodal probability distributions, shown in blue and red. The distribution in green is table-shaped and the one in cyan is a single modal and in both there are significant probabilities of finding from 0 to ≈ 130 proteins. The magenta distribution can be approximated as a negative binomial distribution which is a typical distribution when one has transcriptional or translational bursts [50, 57]. Figure 4, Graph C shows super-Fano distributions for the ERG. We assume a fixed value for N and change values of b^e considering some values for a that permit us to generate probability distributions that are widespread. The red and blue distributions are bimodal with modes centered around 50 and 0. The green distribution is table shaped such that there is a significant probability of finding from zero to a bit beyond N proteins. The blue distribution is one modal and it is widespread, similarly to the table-shaped distribution. Finally we have the magenta distribution which is approximated by a negative binomial.

The parameter values generating the probability distributions of Graphs A–C are rewritten in terms of the kinetic constants in Eqs (2.1)–(2.8). We assume $\rho = 0.01 \text{ min}^{-1}$ and include the approximate values of the kinetic parameters in the Table 1 following the order of the parameter values in each caption of their corresponding Graphs. Where k , f , and h are given in min^{-1} .

Table 1. Approximated values of the kinetic parameters on Figure 4.

Graph A (N, k, f, h)	Graph B (N, k, f, h)	Graph C (k, f, h)
(80, 0.8, 10^{-2} , 10^{-4})	(10, 0.1, 4×10^{-3} , 10^{-4})	(0.5, 3×10^{-3} , 2×10^{-3})
(100, 1, 10^{-2} , 10^{-4})	(40, 0.4, 6×10^{-3} , 10^{-4})	(0.5, 5×10^{-3} , 5×10^{-3})
(100, 1, 8×10^{-3} , 10^{-4})	(60, 0.6, 2×10^{-2} , 10^{-4})	(0.5, 10^{-2} , 10^{-2})
(60, 0.6, 8×10^{-3} , 2×10^{-4})	(90, 0.9, 2×10^{-2} , 10^{-3})	(0.5, 1.5×10^{-2} , 1.5×10^{-2})
(40, 0.4, 8×10^{-3} , 3×10^{-4})	(200, 2, 3×10^{-2} , 10^{-3})	(0.5, 0.017, 0.183)

The super-Fano distributions of Graph A have the OFF to ON switching rate of the order of 10^{-2} min^{-1} which is the same order of the effective ON to OFF switching rate, $\frac{hk}{\rho+h}$ [23] and of the degradation rate of the proteins. That causes the probability distributions to be bi-peaked as the biochemical processes are slower. The super-Fano distributions of Graph B have the effective ON to OFF switching rate given in min^{-1} approximated by (10^{-3} , 4×10^{-3} , 6×10^{-3} , 9×10^{-2} , 0.2) presented in increasing order of N . For the maximal N the effective rate of ON to OFF switching is almost a hundred times larger than the corresponding OFF to ON switching rate. Note that the synthesis rate a hundred times larger than the degradation, which can be considered as a very efficient protein production process. At that limit the promoter remains mostly OFF and the distribution approaches a negative binomial as occurs in bursty protein synthesis [50]. The super-Fano distributions of Graph C show that the stochastic binary model for the ERG having kinetic parameters of the same order as that for a SRG will generate proteins governed by similar probability distributions. That similarity may become a confusing element on the inference of gene regulation based on noise in gene expression. Perhaps, this problem is only theoretical significance, if the differences on the probability distributions generate fluctuations on protein numbers lying within the detection limits of biological systems. About 40% of *E. coli* genes are self-repressed [30] and it would be interesting to determine

whether this is because of the wider possibility of ranges of stochastic gene expression regimes enabled by the SRG or if a downstream regulated gene can detect the particularities on the fluctuations of the protein numbers generated by either the ERG or SRG.

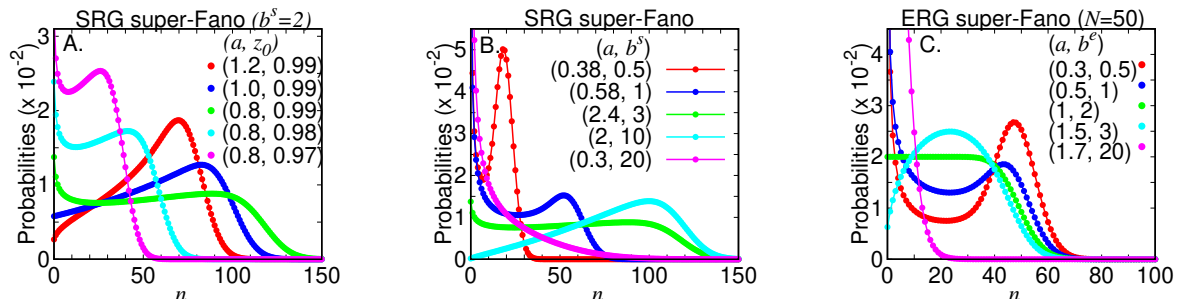


Figure 4. We consider super-Fano probability distributions ($a < b$). In Graph A (SRG) the parameters are indicated inside graph and the moments of the distributions are given in the same order of the values for pairs (a, z_0) within the graph: $\mathcal{F}^s \approx (9.8, 14.7, 21.3, 11.2, 7.8)$; $\xi^s \approx (472.3, 808.4, 1225.9, 310.5, 139.8)$; $\langle n^s \rangle \approx (53.7, 58.9, 60.4, 30.5, 20.5)$; $p_\alpha^s \approx (0.66, 0.58, 0.50, 0.49, 0.49)$. In Graph B (SRG) z_0 is 0.99 for the three smallest values of b^s and for the remaining two values of b^s we have $z_0 = 0.9$. The moments are presented in increasing order of b^s : $\mathcal{F}^s \approx (1.4, 1.1, 1, 0.98, 0.98)$; $\xi^s \approx (17.5, 4.7, 0, -0.3, -0.2)$; $\langle n^s \rangle \approx (47, 35, 20, 14, 7)$; $p_\alpha^s \approx (0.75, 0.82, 0.9, 0.93, 0.97)$. Graph C has super-Fano distributions for ERG with $N = 50$. The moments of the distribution are presented in increasing order of a : $\mathcal{F}^e \approx (14.3, 13.5, 9.3, 7.3, 3.2)$; $\xi^e \approx (400, 312.5, 208.3, 156.3, 9.3)$; $\langle n^e \rangle \approx (30, 25, 25, 25, 4.3)$; $p_\alpha^e \approx (0.6, 0.5, 0.5, 0.5, 0.085)$.

3.4. Sub-Fano probability distributions

Sub-Fano and Fano probability distributions for protein numbers produced from the SRG and Fano factors versus the probability for the gene being ON or versus average protein numbers are presented in Figure 5, Graph A shows sub-Fano probability distributions having Fano factor ≈ 0.5 . To enable a comparison we superpose each sub-Fano with a Fano distribution having the same average protein number with thinner and equally colored lines. The difference on the Fano factor of the two distributions increases with $\langle n \rangle / 2$ and turns the curves more distinguishable. In the Figure 5, Graphs B and C were constructed using the same parameters and share the same key code. The graphs show that a 0.5 Fano factor is an asymptotic noise regime for the SRG which can be achieved for an arbitrary average protein number and with p_α^s approaching zero. Note that the Fano factor approaches one as p_α^s increases, as expected for a gene which is at ON state the majority of time. Then we have the infra-Fano regime, for an average protein number equals to one and having p_α^s approaching zero as the Fano factor gets arbitrarily lower [21].

In the Table 2, the parameter values generating the probability distributions of Graph A are rewritten in terms of the kinetic constants in Eqs (2.1)–(2.4). A fixed $z_0 = 5 \times 10^{-4}$ implies a $h \sim 20 \text{ min}^{-1}$ if we assume $\rho = 0.01 \text{ min}^{-1}$. For each pair of sub-Fano and Fano probability distributions having the same average, a pair (a, b^s) results on the following approximate values of the kinetic parameters, where k and f are given in min^{-1} and K indicates the equilibrium binding affinity of the regulatory protein.

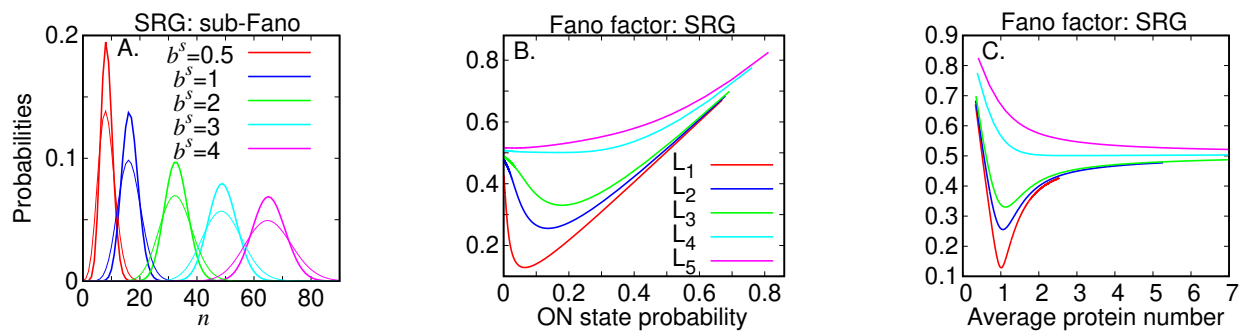


Figure 5. Graph A are Fano and sub-Fano probability distributions. The Fano distributions have thinner lines and we considered distributions having the same average protein numbers to establish a comparison. The values of a in increasing order of b^s are $\approx (833, 1667, 3333, 5000, 6667)$ and all distributions have $z_0 = 5 \times 10^{-4}$. The moments of the sub-Fano distributions in increasing order of b^s are: $\mathcal{F}^s \approx (0.50, 0.51, 0.51, 0.51, 0.52)$; $\xi^s \approx (-4.2, -8.1, -15.9, -23.8, -31.6)$; $\langle n^s \rangle \approx (8.4, 16.5, 32.8, 49.0, 65.3)$; $p_\alpha^s \approx (0.050, 0.049, 0.049, 0.049, 0.048)$; for the Fano distributions we have $\mathcal{F}^s = 1$, $\xi^s = 0$, $p_\alpha^s = 5 \times 10^{-4}$ and the same average protein numbers as for the sub-Fano distributions to which they are superposed. Graphs B and C, respectively, show the Fano factor as function of the probability for the SRG being ON and of the average protein number. The parameters (a, b^s) have the same values on both graphs such that for L_1, L_2, L_3, L_4 , and L_5 we respectively have $(500, 0.01), (500, 0.05), (500, 0.1), (500, 0.5),$ and $(500, 2)$.

Table 2. Values of the kinetic parameters on Figure 5A.

b^s	Sub-Fano ($N, k, f, K = f/h$)	Fano ($N, k, f, K = f/h$)
0.5	(167.1, 1.7, 8.3, 2.4)	($10^3, 10.0, 5 \times 10^{-3}, 4 \times 10^3$)
1	(333.2, 3.3, 16.7, 1.2)	($2 \times 10^3, 20.0, 10^{-2}, 2 \times 10^3$)
2	(667.3, 6.7, 33.3, 0.6)	($4 \times 10^3, 30.0, 2 \times 10^{-2}, 10^3$)
3	(1000.5, 10.0, 50.0, 0.4)	($6 \times 10^3, 60.0, 3 \times 10^{-2}, 7 \times 10^2$)
4	(1333.7, 13.3, 66.7, 0.3)	($8 \times 10^3, 80.0, 4 \times 10^{-2}, 5 \times 10^2$)

The order of K for the sub-Fano distributions are $K \sim 1$ while they are of the order of 10^3 on the Fano ones. The OFF to ON switching rates on the sub-Fano distributions are typically a hundred times faster than those of the Fano ones. On the other hand, the effective OFF to ON switching rate, $\frac{hk}{\rho+h}$, given in min^{-1} for the sub-Fano and the Fano distributions are approximately (1.7, 3.3, 6.7, 10, 13.3) and (10., 20., 40., 60., 80.), respectively, for each value of b^s given in crescent order accordingly with the captions. Therefore, the effective ON to OFF switching rate is almost ten times faster on the Fano distributions. When we compare effective switching rates between the two promoter states, we note that the sub-Fano processes have an $\sim 10\times$ faster OFF to ON transition while it is $\sim 10^{-3}\times$ slower for Fano ones. That causes the ON state probabilities in the sub-Fano regime, and hence the duration of the ON state, to be about a hundred times longer than in the Fano one. The synthesis rate values provide the compensatory effect to ensure that the two noise regimes will have the same average number of

products. Graph C shows that the asymptotic Fano factor for the sub-Fano regime happens for a wide range of values of average protein numbers. Graph B shows that the asymptotic Fano factor happens for smaller ON state probabilities in combination with larger values of the synthesis rate. This parameter regime of SRG is qualitatively distinct from the ERG, as the latter would be in the bursty limit of gene expression [50].

4. Discussions and conclusions

Theoretical models furnish a useful machinery for biologists on the description and understanding of intracellular phenomena. Exhaustive search based on experiments are labor-intensive and expensive, as occurs in analysis of gene expression pathways [64, 65], and enables the construction of knowledge whose gaps can be fulfilled by the use of quantitative models. For instance, the conclusion that the TGF-beta pathway is a negative feedback loop was reached after the model analysis indicate consistence with observed data [66]. Besides, those experiments generate a plethora of experimental data which analysis demands bioinformatics tools to build gene networks based on gene expression, previously known networks, and correlations [34]. Re-analysis of these results applying stochastic models for gene expression could lead us to more accurate information about the dynamics and the topology of gene networks.

Gene switching have been characterized as an important source of noise in quantities of gene products when, for example, transcription or translation occurs in bursts [4, 5, 67, 68]. Transcriptional bursts occur when a gene switches between ON and OFF states standing shortly ON while it is efficiently transcribed and longly OFF when transcripts decay [4, 5, 50, 57]. Translational bursts are the result of efficient translation of small numbers of mRNA copies among other effects [69]. Equation (2.30) and graphs of Figure 2 are showing that the coupling between two originally Poissonian stochastic processes may result in super-Fano distributions governing protein numbers. That condition occurs when one has a positive covariance ξ . Figure 2A shows this as the condition for all parameter values of the stochastic model for the ERG despite the occurrence of quasi-Fano distributions for the limits $p_{\alpha}^e \rightarrow 1$ and $b^e \gg N(1 - p_{\alpha}^e)$. The SRG, however, has a null covariance for $a = b^s$, corresponding to the “resonant” condition in which the protein synthesis and the promoter switching rates are the same. Hence, one may have a SRG that behaves effectively as a constitutive gene despite the switching of its promoter between ON and OFF states. At this limit, $p_{\alpha}^s = z_0$ implies on the average duration of the ON state of the promoter to be equals to the fraction of proteins removed from cytoplasm by degradation [23]. Hence, one may design (or find) a switching gene behaving as a constitutive one, that is, a gene which protein numbers would behave equivalently while performing their function. However, since the SRG is switching, one may change, for example, the degradation rates of proteins or their binding affinity to the regulatory site of the gene to promote a wider range of epigenetic variation. That approach may have practical implications, for example, on the design of therapeutic strategies based on regulation of gene expression [33] and developing a strategy to characterize the regulation of a gene would be important.

The model for the SRG also shows the possibility of an additional noise regime taking place when the covariance assumes negative values. Here we show that $\xi^s < 0$ for $a > b$, as indicated at Figure 2B or in previous papers [21–23]. The occurrence of a low noise regime occurs only for the SRG as observed experimentally, e.g., [17, 19, 20]. That is a surprising result as we have a coupling between

stochastic processes that enables noise reduction instead of amplification [70].

One may also compare Figures 2B,C, and verify that when we deal with a multi state gene, evaluation of only average protein number of the SRG is not sufficient to characterize its dynamics. The calculation of the Fano factor also helps on the characterization of the gene's switching dynamics. For example, the classification of the noise on the protein numbers as super-Fano, a Fano or a sub-Fano gives us an indication about the relation between the probability for the promoter of the gene to be ON and z_0 , the fraction of proteins removed by degradation. The Pearson correlation is negative for $p_\alpha^s > z_0$; it is null for $p_\alpha^s = z_0$; it is positive for $p_\alpha^s < z_0$ and the corresponding distributions governing the protein numbers will be, respectively, sub-Fano; Fano; super-Fano.

Although Eq (2.30) helps on understanding noise reduction mathematically, a biological picture is necessary for interpreting the conditions satisfied by the fast switching SRG in a sub-Fano regime. For example, to obtain cyan colored sub-Fano probability distribution of Figure 5A we used the following parameter values: $b^s = 3$, $a = 5 \times 10^3$, $z_0 = 5 \times 10^{-4}$. For simplicity, we define $\rho = 1$ which implies the time scale of the model to be roughly given in terms of the protein mean life time. The remaining kinetic parameter values can be obtained by algebraic calculations and they are $k = 10^3$, $f = 5 \times 10^3$, and $h_1 = 2 \times 10^3$. Note that the gene will perform thousands of switches during the mean life time of a protein. Furthermore, the proportion of time during which the gene remains ON is about 5% as given on captions of Figure 5. Therefore, during the mean life of a protein there will be an average net synthesis of only ≈ 50 proteins with a plethora of fast gene switching between ON and OFF states, and slow protein degradation. That is analogous to conditions recently observed in the context of transcription [71] and analysis of the symmetries of the model for the SRG reveal this as a non-degenerated regime, that is, it is not possible to have a super-Fano distribution for those parameter values [23].

The quantification of noise in gene expression by means of the Fano factor enables one to infer the behavior of the protein source. For example, constitutive genes can be represented as Poissonian birth-death processes which would generate Fano probability distributions on the number of gene products. Previous works dedicated to analysis of transcriptional burst have used Fano factor to infer burst size [55, 57] as the result of the gene switching between a short duration ON state associated with very efficient transcription followed by a long duration OFF state [5, 50]. However, as we show in Figure 4, the Fano factor may not be sufficient for characterizing a probability distribution. The super-Fano regime enables a diversity of shapes for the probability distributions. Furthermore, in case of a SRG even Fano processes may still be caused by gene switching instead of constitutive expression such that measuring promoter activity in terms of its probabilities of being ON and OFF is important for characterization of its expression dynamics.

The exclusive occurrence of the sub-Fano, and specially the infra-Fano, regimes on the stochastic model for the SRG is a theoretical demonstration of how this strategy enables a more precise control of gene expression [19–21]. The sub-Fano regime, having for example $\mathcal{F} \approx 0.5$ may occur for a wide range of average protein numbers as shown in Figure 5. The difference on how spread is this distribution in comparison with the Fano one poses the question on how susceptible biological systems are to random fluctuations. Inspection indicates that the difference between Fano and sub-Fano distributions with the same average protein number is small. Hence, we may ask whether this difference is sufficiently big to be detected by genes being regulated by those proteins and, eventually, drive random fluctuations on cellular phenotypes.

The noise reduction enabled by self-repression, as indicated by the exclusivity of the sub-Fano regime for the SRG has practical implications. For example, let us consider the search of approximate solutions for the SRG based on the stochastic model for the ERG [58]. In a previous article an alternative model for the SRG was proposed [45] and its complete steady state exact solutions presented later [46]. If one rephrases the parameters of the alternative model for the SRG as $a = \sigma_u$, $b = \frac{\sigma_u}{1+\sigma_b} + \frac{\sigma_b}{1+\sigma_b} \frac{\rho_u}{1+\sigma_b}$ and $z_0 = \frac{1}{1+\sigma_b}$, it is possible to write $\rho_u = \frac{b-az_0}{z_0(1-z_0)}$, where the parameters are written in the time scale of the protein degradation rate. The synthesis at the ON state, the promoter switching from ON to OFF state, and from OFF to ON state are denoted by, respectively, ρ_u , σ_b , and σ_u and the remaining parameters are set as $\rho_b = \theta = 0$. Then, if we set $a > b + \epsilon$ with ϵ being a small number, the alternative SRG will also generate sub-Fano distributions which may not be well approximated by the qualitatively different super-Fano distributions generated by the model for the ERG. That happens if the differences between the distributions become biologically relevant, i.e., if the differences are detectable by the biological systems being affected by those proteins. An experiment to detect such a precision limits would be welcome. A similar analysis can be carried out for the stochastic binary bursty models for the SRG [51] and their linear approximations [58]. In a bursty model, the probabilities governing the protein numbers will be mostly super-Fano unless the bursting size b approaches 1, which would result on the non-bursty model discussed here. In a bursty model, the noise reduction is interpreted in terms of the capacity of a given gene regulatory strategy to reduce the Fano factor (or variance, relative deviation, or any other quantity) in relation to its value whether protein synthesis were carried out as unregulated bursts. The latter implies a Fano factor given by $1 + b$. One may show that for the bursty SRG, the Fano factor may be $< 1 + b$ and, at that limit, the linear approximation for the SRG is not satisfactory because its Fano factor will be only $> 1 + b$. Again, it would be interesting to design an experiment to investigate the conditions for the biological systems to distinguish between those two limits.

The infra-Fano regime revisited here in Figure 5 also points to an additional unexpected result: The possibility of an arbitrarily low noise regime even when only few molecules are interacting. Such a regime appears when the ON to OFF state switching rate is so big that even a small number of proteins is enough to induce the switching to the OFF state. That has been approached in the context of a prokaryotic cell previously [21] and an effective reproduction of these conditions in an eukaryotic cell would be challenging. Further research to determine the mechanisms necessary for such a reliable epigenetic silencing of a gene might require the addition of effective reactions such as mediator molecules [72, 73]. A theoretical formalization of the occurrence of the sub-Fano regime would also be welcome and is beyond the scope of this manuscript.

The exact solutions for Eqs (2.11) and (2.12) are given in terms of special functions, either Heun or Kummer, which are the “parents” of several special functions. Applications of the Heun functions are still rare and it is fair to consider them as the state of the art on the field of exact solutions for physical models [74]. Therefore, we may state that the quantitative description of more complex biological systems will motivate the development of advanced machinery to complement the toolbox of Theoretical Biology. That will deepen our understanding of the functioning of living systems and provide insights for further developments of synthetic biology. One example is the analysis of the self-repressing system when cooperative binding takes place [39, 75, 76]. In that case, numerical solutions given by recursive relations are used [75, 76] because of the increased order of the ODEs governing the steady state generating functions of the SRG. Higher order ODEs are harder to be solved, but finding

analytical closed forms solving these models is not hopeless. For example, [76] indicates that in case of self-repression by a dimer we would have $h_1 n \rightarrow h_1 n(n - 1)$ in Eqs (2.11) and (2.12). Application of the generating function techniques as in [30] would lead to a third order ODE which we would attempt to solve using special cases of solutions for third order ODEs based on the hypergeometric functions [77].

Acknowledgments

We would like to thank the three anonymous reviewers whose comments contributed for increasing the quality of our article. We also would like to thank Prof. John Reinitz for invaluable discussions about the biochemical interpretation of the stochastic models for gene expression. This project had support from Dean of Research of the University of São Paulo for support from the “Use of Intelligent Systems” call (18.5.245.86.7), CAPES (88881.062174/2014-01) and FAPESP (2012/24962-3).

Conflict of interest

The authors declare no conflict of interest.

References

1. M. Delbrück, Statistical fluctuations in autocatalytic reactions, *J. Chem. Phys.*, **8** (1940), 120–124.
2. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell, *Science*, **297** (2002), 1183–1186.
3. W. J. Blake, M. Kærn, C. R. Cantor, J. J. Collins, Noise in eukaryotic gene expression, *Nature*, **422** (2003), 633–637.
4. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, Stochastic mrna synthesis in mammalian cells, *PLOS Biol.*, **4** (2006), e309.
5. D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, F. Naef, Mammalian genes are transcribed with widely different bursting kinetics, *Science*, **332** (2011), 472–474.
6. B. Munsky, G. Neuert, A. van Oudenaarden, Using gene expression noise to understand gene regulation, *Science*, **336** (2012), 183–187.
7. J. M. Raser, E. K. O’Shea, Noise in gene expression: origins, consequences, and control, *Science*, **309** (2005), 2010–2013.
8. A. M. Arias, P. Hayward, Filtering transcriptional noise during development: concepts and mechanisms, *Nat. Rev. Genet.*, **7** (2006), 34–44.
9. G. Chalancon, C. N. Ravarani, S. Balaji, A. Martinez-Arias, L. Aravind, R. Jothi, et al., Interplay between gene expression noise and regulatory network architecture, *Trends Genet.*, **28** (2012), 221–232.
10. A. Sanchez, I. Golding, Genetic determinants and cellular constraints in noisy gene expression, *Science*, **342** (2013), 1188–1193.

11. J. Ansel, H. Bottin, C. Rodriguez-Beltran, C. Damon, M. Nagarajan, S. Fehrmann, et al., Cell-to-cell stochastic variation in gene expression is a complex genetic trait, *PLoS Genet.*, **4** (2008), e1000049.
12. A. Brock, S. Krause, D. E. Ingber, Control of cancer formation by intrinsic genetic noise and microenvironmental cues, *Nat. Rev. Cancer*, **15** (2015), 499–509.
13. S. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of escherichia coli, *Nat. Genet.*, **31** (2002), 64–68.
14. G. T. Reeves, The engineering principles of combining a transcriptional incoherent feedforward loop with negative feedback, *J. Biol. Eng.*, **13** (2019).
15. N. Rosenfeld, M. B. Elowitz, U. Alon, Negative autoregulation speeds the response times of transcription networks, *J. Mol. Biol.*, **323** (2002), 785–793.
16. A. Sancar, L. A. Lindsey-Boltz, T.-H. Kang, J. T. Reardon, J. H. Lee, N. Ozturk, Circadian clock control of the cellular response to DNA damage, *FEBS Lett.*, **584** (2010), 2618–2625.
17. M. A. Savageau, Comparison of classical and autogenous systems of regulations in inducible operons, *Nature*, **252** (1974), 546–549.
18. S. Hooshangi, R. Weiss, The effect of negative feedback on noise propagation in transcriptional gene networks, *Chaos*, **16** (2006), 026108.
19. D. Nevozhay, R. M. Adams, K. F. Murphy, K. Josic, G. Balázsi, Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression., *Proc. Natl. Acad. Sci. U.S.A.*, **106** (2009), 5123–5128.
20. A. Becskei, L. Serrano, Engineering stability in gene networks by autoregulation., *Nature*, **405** (2000), 590–593.
21. A. F. Ramos, J. E. M. Hornos, J. Reinitz, Gene regulation and noise reduction by coupling of stochastic processes, *Phys. Rev. E*, **91** (2015), 020701(R).
22. A. F. Ramos, J. E. M. Hornos, Symmetry and stochastic gene regulation., *Phys. Rev. Lett.*, **99** (2007), 108103.
23. A. F. Ramos, J. Reinitz, Physical implications of $so(2, 1)$ symmetry in exact solutions for a self-repressing gene, *J. Chem. Phys.*, **151** (2019), 041101.
24. J. N. Anastas, R. T. Moon, Wnt signalling pathways as therapeutic targets in cancer, *Nat. Rev. Cancer*, **13** (2013), 11–26.
25. C. K. Mirabelli, R. Nusse, D. A. Tuveson, B. O. Williams, Perspectives on the role of wnt biology in cancer, *Sci. Signal.*, **12** (2019), eaay4494.
26. G. Balázsi, A. van Oudenaarden, J. J. Collins, Cellular decision making and biological noise: from microbes to mammals, *Cell*, **144** (2011), 910–925.
27. L. S. Tsimring, Noise in biology, *Rep. Prog. Phys.*, **77** (2014), 026601.
28. K. Sneppen, Models of life: epigenetics, diversity and cycles, *Rep. Prog. Phys.*, **80** (2017), 042601.
29. J. Peccoud, B. Ycart, Markovian modelling of gene product synthesis, *Theor. Popul. Biol.*, **48** (1995), 222–234.

30. J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, et al., Self-regulating gene: an exact solution., *Phys. Rev. E*, **72** (2005), 051907.
31. S. Iyer-Biswas, F. Hayot, C. Jayaprakash, Stochasticity of gene products from transcriptional pulsing, *Phys. Rev. E*, **79** (2009), 031911.
32. A. F. Ramos, G. C. P. Innocentini, J. E. M. Hornos, Exact time-dependent solutions for a self-regulating gene, *Phys. Rev. E*, **83** (2011), 062902.
33. A. F. Ramos, L. R. Gama, M. C. C. Morais, P. C. M. Martins, Chapter 14: Stochastic modeling for investigation of regulation of transcription of RKIP gene, *Prognostic and Therapeutic Applications of RKIP in Cancer*, Academic Press, (2020), 257–276.
34. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, et al., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U.S.A.*, **102** (2005), 15545–15550.
35. K. Park, T. Prüstel, Y. Lu, J. S. Tsang, Machine learning of stochastic gene network phenotypes, preprint, bioRxiv:825943.
36. L. R. Gama, G. Giovanini, G. Balázsi, A. F. Ramos, Binary expression enhances reliability of messaging in gene networks, *Entropy*, **22** (2020), 479.
37. A. Bakk, R. Metzler, K. Sneppen, Sensitivity of OR in phage λ , *Biophys. J.*, **86** (2004), 58–66.
38. A. Grönlund, P. Lötstedt, J. Elf, Transcription factor binding kinetics constrain noise suppression via negative feedback, *Nat. Commun.*, **4** (2013), 1864.
39. D. C. Marciano, R. C. Lua, C. Herman, O. Lichtarge, Cooperativity of negative autoregulation confers increased mutational robustness, *Phys. Rev. Lett.*, **116** (2016), 258104.
40. K. S. Farquhar, D. A. Charlebois, M. Szenk, J. Cohen, D. Nevozhay, G. Balázsi, Role of network-mediated stochasticity in mammalian drug resistance, *Nat. Commun.*, **10** (2019), 2766.
41. G. M. Cooper, *The Cell: A Molecular Approach*, 2nd edition, Sinauer Associates Inc, Sunderland (MA), 2000.
42. D. K. Hawley, W. R. McClure, Mechanism of activation of transcription initiation from the lambda-PRM promoter, *J. Mol. Biol.*, **157** (1982), 493–525.
43. A. R. Kim, C. Martinez, J. Ionides, A. F. Ramos, M. Z. Ludwig, N. Ogawa, et al., Rearrangements of 2.5 kilobases of noncoding dna from the drosophila even-skipped locus define predictive rules of genomic cis-regulatory logic, *PLoS Genet.*, **9** (2013), e1003243.
44. M. R. Fabian, N. Sonenberg, W. Filipowicz, Regulation of mrna translation and stability by micromnas, *Annu. Rev. Biochem.*, **79** (2010), 351–79.
45. R. Grima, D. R. Schmidt, T. J. Newman, Steady-state fluctuations of a genetic feedback loop: An exact solution, *J. Chem. Phys.*, **137** (2012), 035104.
46. G. C. P. Innocentini, A. F. Ramos, J. E. M. Hornos, Comment on “steady-state fluctuations of a genetic feedback loop: an exact solution” [J. Chem. Phys. 137, 035104 (2012)], *J. Chem. Phys.*, **142** (2015), 027101.
47. J. Holehouse, Z. Cao, R. Grima, Stochastic modeling of autoregulatory genetic feedback loops: A review and comparative study, *Biophys. J.*, **118** (2020), 1517–1525.

48. C. Jia, R. Grima, Small protein number effects in stochastic models of autoregulated bursty gene expression, *J. Chem. Phys.*, **152** (2020), 084115.
49. Z. Cao, R. Grima, Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells, *Proc. Natl. Acad. Sci. U.S.A.*, **117** (2020), 4682–4692.
50. R. Yvinec, L. G. S. da Silva, G. N. Prata, J. Reinitz, A. F. Ramos, Bursting on a two state stochastic model for gene transcription in drosophila embryos, preprint, bioRxiv:107979.
51. N. Kumar, T. Platini, R. V. Kulkarni, Exact distributions for stochastic gene expression models with bursting and feedback, *Phys. Rev. Lett.*, **113** (2014), 268105.
52. T. Tripathi, D. Chowdhury, Interacting RNA polymerase motors on a DNA track: Effects of traffic congestion and intrinsic noise on RNA synthesis, *Phys. Rev. E*, **77** (2008), 011921.
53. S. Choubey, J. Kondev, A. Sanchez, Deciphering transcriptional dynamics in vivo by counting nascent rna molecules, *PLoS Comput. Biol.*, **11** (2015), e1004345.
54. H. Xu, S. O. Skinner, A. M. Sokac, I. Golding, Stochastic kinetics of nascent rna, *Phys. Rev. Lett.*, **117** (2016), 128101.
55. M. Thattai, A. van Oudenaarden, Intrinsic noise in gene regulatory networks, *Proc. Natl. Acad. Sci. USA*, **98** (2001), 8614–8619.
56. G. C. P. Innocentini, J. E. M. Hornos, Modeling stochastic gene expression under repression., *J. Math. Biol.*, **55** (2007), 413–431.
57. V. Shahrezaei, P. S. Swain, Analytical distributions for stochastic gene expression, *Proc. Natl. Acad. Sci. USA*, **105** (2008), 17256–17261.
58. Z. Cao, R. Grima, Linear mapping approximation of gene regulatory networks with stochastic dynamics, *Nat. Commun.*, **9** (2018), 3305.
59. A. F. Ramos, G. C. P. Innocentini, F. M. Forger, J. E. M. Hornos, Symmetry in biology: from genetic code to stochastic gene regulation, *IET Syst. Biol.*, **4** (2010), 311–329.
60. R. Andersson, A. Sandelin, Determinants of enhancer and promoter activities of regulatory elements, *Nat. Rev. Genet.*, **21** (2019), 71–87.
61. N. Rosenfeld, T. J. Perkins, U. Alon, M. B. Elowitz, P. S. Swain, A fluctuation method to quantify in vivo fluorescence data, *Biophys. J.*, **91** (2006), 759–766.
62. G. N. Prata, J. E. M. Hornos, A. F. Ramos, Stochastic model for gene transcription on *drosophila melanogaster* embryos, *Phys. Rev. E*, **93** (2016), 022403.
63. M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, National Bureau of Standards Applied Mathematics Series vol. 55*, 10th edition, U.S. Department of Commerce, Washington (DC), 1972.
64. O. Hallikas, K. Palin, N. Sinjushina, R. Rautiainen, J. Partanen, E. Ukkonen, et al., Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity, *Cell*, **124** (2006), 47–59.
65. C. P. Fulco, J. Nasser, T. R. Jones, G. Munson, D. T. Bergman, V. Subramanian, et al., Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations, *Nat. Genet.*, **51** (2019), 1664–1669.

66. I. Heemskerk, K. Burt, M. Miller, S. Chhabra, M. C. Guerra, L. Liu, et al., Rapid changes in morphogen concentration control self-organized patterning in human embryonic stem cells, *eLife*, **8** (2019), e40526.
67. A. Paré, D. Lemons, D. Kosman, W. Beaver, Y. Freund, W. McGinnis, Visualization of individual scr mrnas during drosophila embryogenesis yields evidence for transcriptional bursting, *Curr. Biol.*, **19** (2009), 2037–2042.
68. A. Crudu, A. Debussche, O. Radulescu, Hybrid stochastic simplifications for multiscale gene networks, *BMC Syst. Biol.*, **3** (2009), 89:1–89:25.
69. H. Kuwahara, S. T. Arold, X. Gao, Beyond initiation-limited translational bursting: the effects of burst size distributions on the stability of gene expression, *Integr. Biol.*, **7** (2015), 1622–1632.
70. J. M. Pedraza, A. van Oudenaarden, Noise propagation in gene networks, *Science*, **307** (2005), 1965–1969.
71. L. A. Sepúlveda, H. Xu, J. Zhang, M. Wang, I. Golding, Measurement of gene regulation in individual cells reveals rapid switching between promoter states, *Science*, **351** (2016), 1218–1222.
72. D. Chetverina, M. Fujioka, M. Erokhin, P. Georgiev, J. B. Jaynes, P. Schedl, Boundaries of loop domains (insulators): Determinants of chromosome form and function in multicellular eukaryotes, *BioEssays*, **39** (2017), 1600233.
73. J. Mozziconacci, M. Merle, A. Lesne, The 3d genome shapes the regulatory code of developmental genes, *J. Mol. Biol.*, **432** (2020), 712–723.
74. P. P. Fiziev, The heun functions as a modern powerful tool for research in different scientific domains, preprint, arXiv:1512.04025.
75. T. Fournier, J. Gabriel, C. Mazza, J. Pasquier, J. Galbete, N. Mermod, Steady-state expression of self-regulated genes, *Bioinformatics*, **23** (2007), 3185–3192.
76. D. Lepzelter, H. Feng, J. Wang, Oscillation, cooperativity, and intermediates in the self-repressing gene, *Chem. Phys. Lett.*, **490** (2010), 216–220.
77. E. S. Cheb-Terrab, A. D. Roche, Hypergeometric solutions for third order linear odes, preprint, arXiv:0803.3474.

A. Appendix: demonstrations

A.1. Demonstration of Eq (2.30)

To demonstrate that

$$\mathcal{F} = 1 + \frac{\xi}{\langle n \rangle},$$

we need to express $\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = \langle n \rangle + \xi$, with ξ being the covariance between $v_m = N \mathbb{I}_1(m)$ and n .

We start summing Eqs (2.11) and (2.12) and dividing the resulting expression by ρ such that at the steady state limit we have:

$$0 = N(\alpha_{n-1} - \alpha_n) + [(n+1)\phi_{n+1} - n\phi_n], \quad (\text{A.1})$$

where we used $N = k/\rho$ and $\phi_n = \alpha_n + \beta_n$ as given by Eqs (2.18) and (2.14).

Now we multiply Eq (A.1) by n , and apply the operator $\sum_{n=0}^{\infty}$, which results on:

$$\langle n \rangle = p_{\alpha} N, \quad (\text{A.2})$$

obtained using Eqs (2.13) and (2.15) for $p = 1$. Then we multiply Eq (A.1) by n^2 , and apply the operator $\sum_{n=0}^{\infty}$, which results on:

$$\langle n^2 \rangle = \langle n \rangle + N \langle n_{\alpha} \rangle, \quad (\text{A.3})$$

obtained using Eqs (2.13) and (2.15). To calculate $\langle n^2 \rangle$ we need to obtain an expression for $N \langle n_{\alpha} \rangle$ which can be done by means of the covariance ξ .

The covariance is $\xi = \langle v_m n \rangle - \langle v_m \rangle \langle n \rangle$ and we may compute $\langle v_m \rangle$ as:

$$\langle v_m \rangle = v_{\alpha} p_{\alpha} + v_{\beta} p_{\beta} = N p_{\alpha} = \langle n \rangle,$$

where we used $v_{\alpha} = N$ and $v_{\beta} = 0$ and Eq (A.2). Next, we compute

$$\langle v_m n \rangle = \sum_{n=0}^{\infty} n (v_{\alpha} \alpha_n + v_{\beta} \beta_n) = N \sum_{n=0}^{\infty} n \alpha_n = N \langle n_{\alpha} \rangle,$$

where we used Eq (2.29) for $p = q = 1$, $v_{\alpha} = N$, $v_{\beta} = 0$, and Eq (2.15). Hence, if we use Eq (A.3) we write that

$$N \langle n_{\alpha} \rangle = \xi + \langle n \rangle^2 \Rightarrow \langle n^2 \rangle - \langle n \rangle^2 = \sigma^2 = \langle n \rangle + \xi.$$

A.2. Computation of $\langle n_{\alpha}^e \rangle$

In [59] we indicated how to compute the moments of the distribution in terms of the generating functions of the probabilities α_n^e , β_n^e , and ϕ_n^e but we did not compute the closed expressions for the auxiliary moments $\langle n_{\alpha}^e \rangle$ and $\langle n_{\beta}^e \rangle$ in Eq (2.15). The generating function for the probabilities α_n^e [29, 56, 59] can be written as:

$$\alpha^e(z) = \frac{a}{b^e} M(a+1, b^e+1, N(z-1)). \quad (\text{A.4})$$

The auxiliary moment $\langle n_{\alpha}^e \rangle$ is obtained using the first derivative of $\alpha^e(z)$ which results:

$$\langle n_{\alpha}^e \rangle = \left. \frac{d\alpha^e(z)}{dz} \right|_{z=1} = \frac{a}{b^e} \frac{1+a}{1+b^e} N. \quad (\text{A.5})$$

Here we used the identity $\frac{d}{d\eta} M(A, B, \eta) = \frac{A}{B} M(A+1, B+1, \eta)$ (see formula 13.4.8 in [63]).

A.3. Computation of ξ^s

The procedure for computation of the moments of the distribution for the SRG from the generating functions of the probabilities α_n^s , β_n^s , and ϕ_n^s is the same of the ERG [59]. In our previous publications [21–23] we did not compute the closed expressions for the auxiliary moments $\langle n_{\alpha}^s \rangle$ and $\langle n_{\beta}^s \rangle$ defined in Eq (2.15). The generating function for the probabilities α_n^s [21–23, 30] can be written as:

$$\alpha^s(z) = \frac{a}{b^s} \frac{z_0}{c} M(a+1, b^s+1, N z_0 (z - z_0)). \quad (\text{A.6})$$

As in subsection A.2, the auxiliary moment $\langle n_\alpha^s \rangle$ is obtained using the first derivative of $\alpha^s(z)$ evaluated at $z = 1$, which results:

$$\langle n_\alpha^s \rangle = \frac{a}{b^s} \frac{z_0}{c} \frac{1}{1 - z_0} [b^s M(a, b^s, Nz_0(1 - z_0)) + (b^s - Nz_0(1 - z_0)) M(a + 1, b^s + 1, Nz_0(1 - z_0))]. \quad (\text{A.7})$$

Here we used the identity $\frac{d}{d\eta} M(A + 1, B + 1, \eta) = \frac{1}{\eta} [BM(A, B, \eta) + (\eta - B)M(A + 1, B + 1, \eta)]$ with $\eta = Nz_0(z - z_0)$ (see formula 13.4.14 in [63]). Recall that $c = M(a, b^s, Nz_0(1 - z_0))$ and using Eqs (2.23) and (2.27) one may rewrite the Eq (A.7), after some manipulation, as follows:

$$\langle n_\alpha^s \rangle = \frac{az_0}{1 - z_0} (1 - p_\alpha^s). \quad (\text{A.8})$$

Now we can write ξ from Eq (2.31) for the SRG as $\xi^s = N \left(\frac{az_0}{1 - z_0} (1 - p_\alpha^s) - \langle n \rangle p_\alpha^s \right)$ which becomes Eq (2.36) after a rearrangement.



AIMS Press

© 2020 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)