Symmetry and Stochastic Gene Regulation

Alexandre F. Ramos* and José E. M. Hornos

Instituto de Fisica de S. Carlos, Universidade de S. Paulo, Caixa Postal 369, BR-13560-970, S. Carlos, S. P, Brazil

(Received 17 April 2007; published 7 September 2007)

Lorentz-like noncompact Lie symmetry SO(2,1) is found in a spin-boson stochastic model for gene expression. The invariant of the algebra characterizes the switch decay to equilibrium. The azimuthal eigenvalue describes the affinity between the regulatory protein and the gene operator site. Raising and lowering operators are constructed and their actions increase or decrease the affinity parameter. The classification of the noise regime of the gene arises from the group theoretical numbers.

DOI: 10.1103/PhysRevLett.99.108103 PACS numbers: 87.16.Ac, 02.20.Sv, 02.50.Ey, 87.10.+e

Despite the huge amount of data collected during the genomic era and the elucidation of some basic mechanisms of gene regulation in λ phage and *E. coli*, the understanding of gene expression remains an important open problem. Growth, development, and other cellular functions are controlled by cascades of genes, ordered expressed in time and space [1,2]. Such a web of gene interactions shows a remarkable complexity even in a simple prokaryotic organism [3]. A robust strategy to handle these systems is the decomposition of the entire network in elementary building blocks related to specific biological functions [4,5]. In recent years a series of important experimental developments have provided a large amount of data on regulation chains in bacteria, yeast, and *Drosophila* which renders feasible the modeling of those elementary circuits [6–9].

The presence of noise, caused by a frequently small number of molecular species in a cell [10,11], has been observed in several experiments involving the monitoring of fluorescence microscopy [12]. Consequently, reliable data on protein concentrations were obtained and significant fluctuations around the mean value were observed. Two different approaches have been employed in the theoretical treatment of noise. In the first, ordinary nonlinear phenomenological equations are written for the molecular concentrations, followed by the introduction of noise by the Langevin mechanism [13]. Microscopic master equations for the protein distributions are emphasized in the second approach [14,15]. Mean values, square root deviation, etc., are then calculated from these probabilities. In this context, a remarkable spin-boson or binary model has been proposed [16,17] and solved analytically for an auto interacting gene [18] and also for noise induced by external repression [19].

In this Letter we show that the equations for this model exhibit a continuous Lorentz-like noncompact Lie symmetry SO(2,1). The general angular momentum or Casimir operator of the algebra has a simple biological implication: it characterizes how the system approaches the stationary state. The raising and lowering operators are related to the affinity of repressing proteins to the gene operator site. The azimuthal and total angular momentum are arbitrary real numbers, in contrast to the SO(3) operators, and are related to the cellular noise.

In the spin-boson model protein synthesis is described by a Markov process for the number n of free proteins in the cytoplasm produced by the corresponding gene. The system has two states, in the first the operator site is occupied by a repressive protein while in the second the operator site is vacant. The probability to find the gene without a repressing protein and having around n proteins is α_n, otherwise the probability is β_n. The master equations are

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

$$\frac{d\beta_n}{dt} = \chi k[\beta_{n-1} - \beta_n] + \rho[(n+1)\beta_{n+1} - n\beta_n] + hn\alpha_n - \tilde{f}\beta_n, \tag{2}$$

where the parameter k is the rate of protein synthesis while ρ is the protein decay rate. The parameters h and $\tilde{f}$ regulate the binding and unbinding rate, respectively. The model also allows a repressed production with a small rate χk, where $\chi < 1$.

The model was solved by the celebrated generating function method in which the differential-difference equations above are replaced by partial differential equations introducing holomorphic functions defined by

$$\alpha(z,t) = \sum_{n=0}^{\infty} \alpha_n(t) z^n, \qquad \beta(z,t) = \sum_{n=0}^{\infty} \beta_n(t) z^n; \tag{3}$$

the resulting equations are

$$\frac{\partial \alpha}{\partial \tau} = (z-1)\left(N\alpha - z_0 \frac{\partial \alpha}{\partial z}\right) + (z_0 - 1) z \frac{\partial \alpha}{\partial z} + f\beta, \tag{4}$$

$$\frac{\partial \beta}{\partial \tau} = (z-1)\left(N\chi\beta - z_0 \frac{\partial \beta}{\partial z}\right) - (z_0 - 1) z \frac{\partial \alpha}{\partial z} - f\beta, \tag{5}$$

where

$$\tau = (\rho + h)t, \qquad N = \frac{k}{\rho + h},$$
$$f = \frac{\tilde{f}}{\rho + h}, \qquad z_0 = \frac{\rho}{\rho + h}. \tag{6}$$

A second-order differential equation can be written by solving the stationary Eq. (4) for β and replacing it in Eq. (5),

$$p\frac{d^2\alpha}{dz} + q\frac{d\alpha}{dz} + r\alpha = 0, \tag{7}$$

where

$$p = \frac{z - z_0}{N(1 - \chi/z0)},$$
$$q = \frac{1 + b - N(1 + \chi/z_0)(z - z_0)}{N(1 - \chi/z0)}, \tag{8}$$

and

$$r = \frac{2N\chi(z - z_0) - (z_0 + \chi)(1 + b)}{2(z_0 - \chi)} - a - \frac{1 - b}{2}. \tag{9}$$

The constants a and b are given by

$$a = f(1 - \chi)/(z_0 - \chi) \quad \text{and} \quad b = f + (1 - z_0)N. \tag{10}$$

The Eq. (7) has a simple pole at $z = z_0$ and an irregular singularity at infinity, suggesting a solution in terms of the confluent hypergeometric functions [20]. In fact, the stationary solutions are

$$\alpha(z) = c^{-1}\frac{a}{b}\frac{z_0 - \chi}{1 - \chi}\exp[N\chi/z_0(z - 1)] \times M(1 + a, 1 + b, N(1 - \chi/z_0)(z - z_0)), \tag{11}$$

$$\beta(z) = c^{-1}\exp[N\chi/z_0(z - 1)] \times M(a, b, N(1 - \chi/z_0)(z - z_0)) - \alpha(z), \tag{12}$$

$$\phi(z) = \alpha(z) + \beta(z) = c^{-1}\exp[N\chi/z_0(z - 1)]M(a, b, N(1 - \chi/z_0) \times (z - z_0)), \tag{13}$$

where,

$$c = M(a, b, N(1 - \chi/z_0)(1 - z_0)). \tag{14}$$

The probabilities ϕ_n can be recovered calculating

$$\phi_n = \frac{1}{n!}\frac{d^n\phi}{dz^n}(z)|_{z=0}. \tag{15}$$

For sake of simplicity, from here, we restrict ourselves to the case $\chi = 0$. The total probabilities are easily calculated and are

$$\phi_n = \frac{1}{n!}\frac{N^n}{c}\frac{(a)_n}{(b)_n}M(a + n, b + n, -Nz_0). \tag{16}$$

The symmetry emerges if we consider the operator

$$L_z = \frac{z - z_0}{N}\frac{d^2}{dz^2} + \frac{1 + b - N(z - z_0)}{N}\frac{d}{dz} - \frac{1 + b}{2}; \tag{17}$$

comparing with Eq. (7) we see that stationary solution obeys

$$L_z\alpha(z) = m\alpha(z), \tag{18}$$

with $m = a + (1 - b)/2$. Stating the system as an eigenvalue problem we can consider the Cartan companions of this operator, namely,

$$L_+ = \frac{z - z_0}{N}\frac{d^2}{dz^2} + \frac{1 + b}{N}\frac{d}{dz}, \tag{19}$$

and

$$L_- = \frac{z - z_0}{N}\frac{d^2}{dz^2} + \frac{1 + b - 2N(z - z_0)}{N}\frac{d}{dz} + N(z - z_0) - 1 - b. \tag{20}$$

Calculating the commutators we obtain

$$[L_z, L_\pm] = \pm L_\pm, \qquad [L_+, L_-] = -2L_z, \tag{21}$$

which is the rotational algebra with the "wrong sign", that is the $so(2, 1)$ algebra. The invariant operator is well known,

$$\mathbb{C} = L_z^2 - L_+L_-/2 - L_-L_+/2, \tag{22}$$

and whose eigenvalue is

$$C = l(l + 1), \tag{23}$$

where $l = -(1 + b)/2$. Since the quantities a and b are real, the representation given here is unitary and also unbounded above and below [21].

The biological meaning of the invariant can be understood considering the time dependence of the probability generating functions. The Eqs. (4) and (5) have the general form

$$\frac{d\Psi}{dt} = \mathcal{H}\Psi, \tag{24}$$

where $\mathcal{H}$ is a two-by-two matrix constructed from the Eqs. (4) and (5) and $\Psi = [\alpha\beta]^T$. The solution of the Eq. (24) is given by expanding Ψ over the basis of the eigenfunctions of the operator $\mathcal{H}$. Although the eigenfunctions of $\mathcal{H}$ cannot be expressed in terms of hypergeometric functions, the eigenvalues can be easily calculated under the requirement of analyticity. We search

for analytical solutions around z_0

$$\alpha(z,t) = e^{\lambda t}(z - z_0)^j \tilde{\alpha}(z), \tag{25}$$

where $\tilde{\alpha}(z)$ is analytical. The eigenvalues λ are

$$\lambda_1^j = -(b + j)(\rho + h), \qquad j = 0, 1, 2, \ldots, \tag{26}$$

where the analyticity of $\alpha(z)$ imposes j to be integer and we see that b is the relative frequency for the smallest eigenvalue which controls the approach of the switch to equilibrium.

The action of the raising and lowering operators on the generating function can be calculated using the Kummer relations:

$$L_- M(1 + a, 1 + b, N(z - z_0)) = (a - b)M(a, 1 + b, N(z - z_0)), \tag{27}$$

$$L_+ M(a, 1 + b, N(z - z_0)) = aM(1 + a, 1 + b, N(z - z_0)); \tag{28}$$

therefore we see that the parameter $a = f/\rho$ is increased (decreased) by L_+ (L_-) increasing the affinity factor f.

The biological quantities of the model were calculated in terms of the group theoretical numbers b, a, and z_0 by solving the Eqs. (10) for f and N,

$$f = az_0, \qquad N = \frac{b - az_0}{1 - z_0}, \tag{29}$$

where we considered $\chi = 0$. Since $N \geq 0$, the relation $b - az_0$ is greater or equal to zero. Considering that a vector of an irreducible representation is fixed by b and a, the possible values of z_0 are in the interval (0,1) if $a \leq b$ or in the interval $(0, b/a]$ for $a > b$. Note that z_0 cannot be zero or one, because in these limits b should go to zero or a, respectively.

Finally, we calculate the noise as function of the group theoretical numbers, b and a. The noise, expressed in terms of the Fano factor $\sigma = (\bar{n}^2 - \bar{n})/\bar{n}$, is shown in Figs. 1(a) and 1(b). Each curve in these figures corresponds to different values of a for fixed b. Figure 1(a) corresponds to a fast switch, which reaches rapidly equilibrium while Fig. 1(b) describes a slow switch. The curves in the bottom of the plot have a values greater than b. The top curves correspond to $a < b$ and we also can see a squared single point for $a = b$. The parameter b labels an irreducible representation of the group SO(2,1) and a denotes a vector in the representation space. In the case $a > b$ the Fano factor is smaller than 1, a sub-Fano stochastic process, while for $a = b$ we have a Fano process. The super-Fano behavior occurs for $a < b$. In all cases we see that if the representation label number a and b are fixed the mean protein number will range in a defined interval. In the sub-Fano regime the interval is $0 \leq \bar{n} < \bar{n}_0$ and it is $\bar{n}_0 < \bar{n} < \infty$ for super Fano. $\bar{n}_0 = aM(a + 1, b + 1, b)/M(a, b, b)$ is the protein mean number in cytoplasm for z_0 going to zero and the corresponding noise is $1 + ab/\bar{n} - \bar{n}$. The Fano process is degenerate, since in this case the mean number of proteins is equal to b, and the noise is equal to one for any value of z_0. Biologically, this means that if the relative switching time is fixed and also the affinity, the possible mean values are restricted.

In Fig. 2 we show the probability distributions. Each plot has fixed a and b, and the lines are for different z_0. Sub-

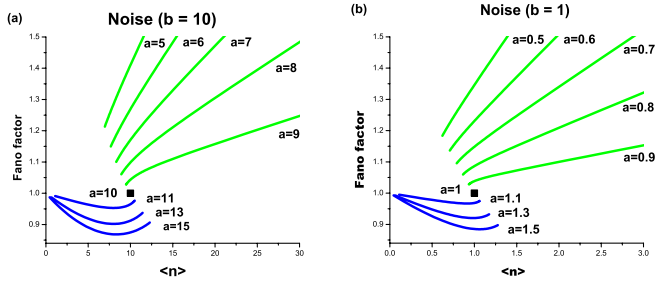

FIG. 1 (color online). Fano factor, $\sigma = \bar{n}^2/\bar{n} - \bar{n}$, versus mean number of synthesized proteins, $\bar{n}$, for a fast (a) and a slow (b) switch state. For $a \leq b$, we choose z_0 in the interval (0,1), in the contrary case, $a > b$, the interval is: $(0, b/a]$. For $a < b$, the noise is greater than one, characterizing a super-Fano regime. Fano states, with noise equal to unity, are obtained when $b = a$ and $\bar{n} = b$. The sub-Fano processes occur when $a > b$.

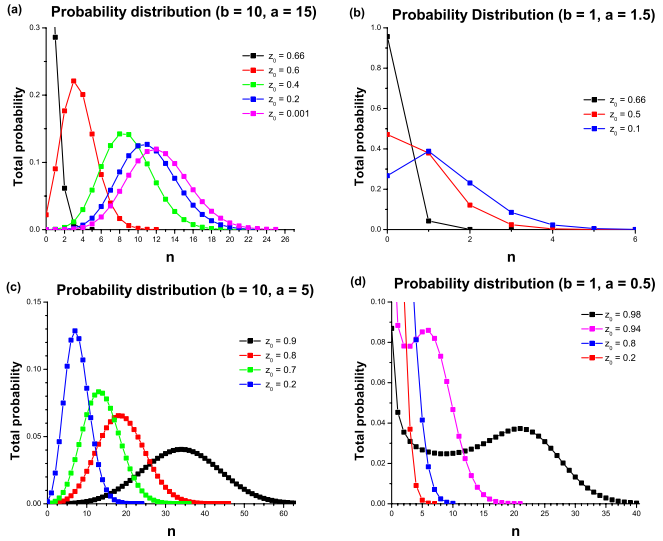

FIG. 2 (color online). Probability distributions, ϕ_n. Sub-Fano probabilities for fixed a and b are shown for fast and slow switches in (a) and (b), respectively, whose mean number of proteins are $\{12.3, 11.8, 11.1, 6.9, 0.4\}$ and $\{1.2, 0.7, 0.04\}$, in crescent order of z_0. Fast and slow switches with super-Fano behavior are displayed in (c). The mean number of proteins and the correspondent noise of each curve are given by $\{7.7, 13.8, 18.9, 34.0\}$ and $\{1.3, 1.6, 1.9, 2.7\}$, in crescent order of z_0. In (d) the values for the mean number of proteins and noise are $\{0.7, 1.7, 3.8, 9.7\}$ and $\{1.2, 1.6, 2.4, 4.7\}$. This plot also shows the presence of two peaks in the probability distribution.

Fano probability distributions for fast and slow switch states are shown in Figs. 2(a) and 2(b), respectively. In both, the diminishing of z_0 leads to a right displacement of the top of the probability distributions. In Figs. 2(c) and 2(d), super-Fano probability distributions are given for fast and slow switch states, respectively. Here, the maximum probability is displaced to right with the growing of z_0. Probability distributions of a super-Fano slow switch state present two peaks for low values of a and z_0 close to 1, as we see in Fig. 2(d). Probability distributions of fast switch states have their peaks centered around greater values of n than slow switch states as we see by comparing Figs. 2(a) and 2(c) with Figs. 2(b) and 2(d). Anyway, super-Fano with low b are noisier than high b states. This suggests that a gene in a fast switch state is more copious and less fluctuating than a gene in a slow switch state. Moreover, the existence of two peaked probability distributions indicate that the bistable behavior of the gene occurs for slow switch states, since there is no similar regime for high values of b. Another biological interpretation for the group theoretical label a can be obtained comparing our results with [18]. The adiabatic parameter defined in our previous work is here the label a of a vector in the representation b. Using this nomenclature we see that if the relative switching time decay b is bigger than the adiabatic parameter the noise regime is super-Fano otherwise it is sub-Fano.

We conclude by summarizing our results. The notions of symmetry and invariance have been introduced in the field of stochastic gene expression by rephrasing the differential-difference master equation in the language of differential operators allowing the set up of the Lie algebra theory. We found a SO(2,1) Lie symmetry, invisible in the traditional form of the master equation, explicitly constructed with differential operators. The biological meaning of the symmetry was revealed showing that the Lie algebra invariant is the decay time of the switch. The group theoretical representation labels are connected with the nature of the system noise measured by the Fano factor. Representation vectors for which $m > -l$ correspond to a sub-Fano noise while super-Fano behavior occurs in the opposite case. Symmetry should be considered as a starting point followed by the inclusion of the symmetry breaking terms such as gene interactions, dimmer formation [22], etc. Although there are few instances of the application of symmetry groups to biological theory [23], the group theoretical machinery has been shown to be a powerful tool in mathematics and physics and can provide here a composition principle for understanding more complex systems.

We thank Professor John Weiner for his reading and suggestions on the writing of the paper, and CNPq and FAPESP for financial support.

*aramos@ifsc.usp.br

[1] A. B. Oppenheim, O. Kobiler, J. Stavans, D. L. Court, and S. Adhya, Annu. Rev. Genet. **39**, 409 (2005).
[2] M. Ptashne *A Genetic Switch: Phage λ and Higher Organisms* (Cell Press and Blackwell Science, Cambridge, MA, 1992), 2nd ed..
[3] A. Arkin, J. Ross, H. H. McAdams, Genetics **149**, 1633 (1998).
[4] N. Rosenfeld, M. Elowitz, and U. Alon, J. Mol. Biol. **323**, 785 (2002).
[5] F. Isaacs, J. Hasty, C. R. Cantor, and J. J. Collins, Proc. Natl. Acad. Sci. U.S.A. **100**, 7714 (2003).
[6] V. V. Gursky, J. Jaeger, K. N. Kozlov, J. Reinitz, and A. M. Samsonov, Physica (Amsterdam) **D197**, 286 (2004).
[7] U. S. Bhalla and R. Iyengar, Science **283**, 381 (1999).
[8] G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell, Nature (London) **406**, 188 (2000).
[9] J. M. Raser and E. K. O'Shea, Science **309**, 2010 (2005).
[10] M. Delbrück, J. Chem. Phys. **8**, 120 (1940).
[11] O. G. Berg, J. Theor. Biol. **71**, 587 (1978).
[12] A. Becksei and L. Serrano, Nature (London) **405**, 590 (2000).
[13] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
[14] J. Paulsson, O. G. Berg, and M. Ehrenberg, Proc. Natl. Acad. Sci. U.S.A. **97** 7148 (2000).
[15] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).
[16] M. Sasai, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **100**, 2374 (2003).
[17] R. Metzler and P. G. Wolynes, Chem. Phys. **284**, 469 (2002).
[18] 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).
[19] G. Innocentini and J. E. M. Hornos, J. Math. Biol. **55**, 413 (2007).
[20] *Handbook of Mathematical Functions*, Nat. Bur. Stand. Appl. Math. Series Vol. 55, edited by M. Abramowitz and I. A. Stegun (U.S. GPO, Washington, D.C., 1972).
[21] B. G. Wybourne, *Classical Groups for Physicists* (John Wiley, New York, 1974).
[22] K. Y. Kim, D. Lepzelter, and J. Wang, J. Chem. Phys. **126**, 034702 (2007).
[23] J. E. M. Hornos and Y. M. M. Hornos, Phys. Rev. Lett. **71**, 4401 (1993).