J Stat Phys (2008) 131: 153–165
DOI 10.1007/s10955-008-9484-4

# Multiple Stochastic Point Processes in Gene Expression

**Rajamanickam Murugan**

Received: 21 July 2006 / Accepted: 10 January 2008 / Published online: 25 January 2008


**Abstract** We generalize the idea of multiple-stochasticity in chemical reaction systems to gene expression. Using Chemical Langevin Equation approach we investigate how this multiple-stochasticity can influence the overall molecular number fluctuations. We show that the main sources of this multiple-stochasticity in gene expression could be the randomness in transcription and translation initiation times which in turn originates from the underlying bio-macromolecular recognition processes such as the site-specific DNA-protein interactions and therefore can be internally regulated by the supra-molecular structural factors such as the condensation/super-coiling of DNA. Our theory predicts that (1) in case of gene expression system, the variances ($\varphi$) introduced by the randomness in transcription and translation initiation-times approximately scales with the degree of condensation ($s$) of DNA or mRNA as $\varphi \propto s^{-6}$. From the theoretical analysis of the Fano factor as well as coefficient of variation associated with the protein number fluctuations we predict that (2) unlike the singly-stochastic case where the Fano factor has been shown to be a monotonous function of translation rate, in case of multiple-stochastic gene expression the Fano factor is a turn over function with a definite minimum. This in turn suggests that the multiple-stochastic processes can also be well tuned to behave like a singly-stochastic point processes by adjusting the rate parameters.

**Keywords** Multiple stochastic point processes · Gene expression · Time dependent rates

R. Murugan (✉)
Kreiman Lab, Department of Ophthalmology and Neurology, Children's Hospital Boston Harvard Medical School, Boston, MA 02115, USA
e-mail: rajamanickam.murugan@childrens.harvard.edu

R. Murugan
Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400005, India
e-mail: rmurugan@gmail.com

## 1 Introduction

Multiple stochastic point processes play critical roles in the generalization of the mass action kinetics in biology and chemistry [1–5]. For example, the simple first order irreversible chemical reaction $x \xrightarrow{k} y$ can be well described by the deterministic mass action rate law as $d_t x = -kx$ where $x$ and $y$ are the concentrations or number of reactant and the product molecules respectively and $k$ is the rate constant associated with this first order transition. However, instead of producing a deterministic trajectory for the initial condition. $x|_{t=0} = x_0$ as $x = x_0 e^{-kt}$ or $y = x_0(1 - e^{-kt})$, any experimental observation of this system will always produce any one of the realizations from the family of trajectories corresponding to the integral solution of the stochastic differential equation (SDE) or Chemical Langevin Equation (CLE) $d_t x = -kx + \sqrt{kx}\eta_{x,t}$. Here $\eta_{x,t}$ is the delta correlated Gaussian white noise with mean $\langle \eta_{x,t} \rangle = 0$ and the variance obeying the fluctuation dissipation theorem as $\langle \eta_{x,t} \eta_{x,t'} \rangle = \delta(t - t')$. The described system is a typical example of a singly-stochastic Poisson point processes. Here we implicitly assumed that the free-energy barrier that separates the reactant $x$ from the product $y$ is an invariant with respect to the time scales associated with the first order reaction dynamics and therefore $k$ is a constant quantity. However, in most of the biological systems, particularly in the fundamental bio-macromolecular recognition processes such as DNA-protein [6–10], DNA-DNA [8], protein-protein and protein-ligand interactions [11] this assumption has been shown to be invalid. Because the rate coefficients as well as the diffusion terms associated with such processes will be a time dependent fluctuating quantities whose probability distribution functions are in turn determined by the higher order structural features of the interacting bio-macromolecules such as super-coiling/condensation of DNA [9] in case of DNA-protein and DNA-DNA interactions [6–9] and the tertiary structure of protein molecule in case of protein-protein and protein-ligand [11] interactions. These are typical examples of multiple-stochastic point processes.

The presence of these multiple stochastic sources eventually increases the overall fluctuations in the number of molecules involved in the deterministic transition which in turn drives the system towards higher disordered states. Therefore an external free energy input is necessary to maintain the system in any ordered state or to reduce the number fluctuations. For example, in case of site-specific DNA-protein interactions [6–9], the protein molecules non-specifically binds with DNA lattice in the first step and then searches for the specific site via one dimensional diffusion dynamics along DNA lattice. Since the non-specific contact can form anywhere on DNA lattice, the rate coefficient associated with the protein molecule to find its specific-site on DNA lattice is a random quantity. However, this randomness can be reduced when the protein molecule which is undergoing a one-dimensional diffusion dynamics along DNA lattice is energetically driven toward its specific-site. Interestingly, studies [6–9] showed that when the protein molecule searches for the specific-site via unbiased one-dimensional random jump dynamics with certain critical jump sizes as $k_c \approx 2N^{2/3}$ where $N$ is the size of DNA lattice under consideration in base-pairs (bps), then there is no necessity for the existence of such external free energy inputs. The overall fluctuations in the rate coefficients associated with the site-specific DNA-protein interactions can be reduced by introducing a static as well as dynamical disorder in the diffusion term associated with the one-dimensional searching of the non-specifically bound protein molecule on DNA [6–10] lattice. Recently we have shown that by including the fluctuations in transcriptional initiation times which actually stem from the randomness in the site-specific interactions of RNA polymerase (RNAP) enzyme with the promoter sequences on DNA lattice, the singly-stochastic transcription process can be generalized as a doubly-stochastic point process [10]. In this article, we generalize the idea of entire stochastic gene expression as a

multiple-stochastic system and we will investigate how the multiple-stochasticity affects the overall molecular number fluctuations and how the higher order molecular structural factors controls such enhanced levels of fluctuations in gene expression. The organization of this paper is as follows. First we derive the statistical measures associated with the fluctuations in various components of gene expression system with singly stochastic noise sources and then we discuss about the origin of various sources of the multiple-stochasticity in gene expression. Then we derive the statistical measures associated with the fluctuations in various components of gene expression system in the presence of multiple-stochasticity. Finally we investigate how the overall noise level is influenced by the multiple-stochasticity and how the supra molecular structural factors associated with DNA or mRNA controls such multiple-stochastic effects.

## 2 Singly Stochastic Gene Expression

The entire process of gene expression that includes transcription of the genomic DNA into mRNA by RNA polymerase enzyme (RNAP) and the subsequent translation of mRNA into the protein polypeptide chain by ribosome can be well described [12, 13] by the following chemical reaction scheme

$$\text{DNA} \underset{\text{RNAP}}{\overset{m_r}{\rightarrow}} [\text{Transcription-Initiation}] \overset{\tau_r}{\rightarrow} \text{mRNA} \underset{\text{RIBOSIME}}{\overset{m_p}{\rightarrow}} [\text{Translation-Initiation}] \overset{\tau_p}{\rightarrow} \text{Protein}.$$

Here $m_r$ is the average transcription initiation time, $m_p$ is the average translation initiation time, $\tau_r$ and $\tau_p$ are the total times associated with the formation of a complete transcript of mRNA and the corresponding complete protein polypeptide chain respectively. The birth-death master equation that describes the time evolution of the probability $P_{m,p,t}$ of observing $m$ number of mRNA molecules and $p$ number of protein molecules at time $t$ can be written as follows:

$$\begin{aligned}\partial_t P_{m,p,t} &= k_r P_{m-1,p,t} + \gamma_r (m+1) P_{m+1,p,t} + k_p m P_{m,p-1,t} + \gamma_p (p+1) P_{m,p+1,t} \\ &\quad - (k_r + \gamma_r m + k_p m + \gamma_p p) P_{m,p,t}. \end{aligned} \tag{1}$$

Here $k_r = (m_r + \tau_r)^{-1}$ is the overall transcription rate coefficient and $k_p = (m_p + \tau_p)^{-1}$ is the overall translation rate coefficient. Since the rate coefficients are invariant with respect to the time variable, the differential difference equation (1) describes the dynamics of a coupled two singly-stochastic systems where $\gamma_r$ and $\gamma_p$ are the decay rate constants associated with mRNA and the protein molecules respectively. The Fokker-Plank approximation corresponding to the master equation (1) with the initial condition $P_{m,p,t|m_0 \cdot p_0,0} = \delta(m - m_0)\delta(p - p_0)$ can be given as follows [14, 15]:

$$\begin{aligned}\partial_t P_{m,p,t} &= -\partial_m [(k_r - \gamma_r m) P_{m,p,t}] - \partial_p [(k_p m - \gamma_p p) P_{m,p,t}] \\ &\quad + 2^{-1} \partial_m^2 [(k_r + \gamma_r m) P_{m,p,t}] + 2^{-1} \partial_p^2 [(k_p m + \gamma_p p) P_{m,p,t}]. \end{aligned} \tag{2}$$

The Fokker-Plank approximate equation (2) is equivalent to the following set of nonlinear stochastic differential equations (SDE) or Chemical Langevin Equations (CLE)

$$\begin{aligned} d_t m &= k_r - \gamma_r m + \sqrt{k_r + \gamma_r m}\, \eta_{r,t}, \\ d_t p &= k_p m - \gamma_p p + \sqrt{k_p m + \gamma_p p}\, \eta_{p,t}. \end{aligned} \tag{3}$$

Here $\eta_{r,t}$ and $\eta_{p,t}$ are delta-correlated Gaussian white noise processes. When the dynamical system described by (3) is close to the steady state which is particularly true under *in vivo* biological conditions, we find that $k_r \to \gamma_r \langle m \rangle_\infty$ and $k_p \langle m \rangle_\infty \to \gamma_p \langle p \rangle_\infty$ and therefore we can use the following system of linear-approximations:

$$\begin{aligned} d_t m &\approx k_r - \gamma_r m + \sqrt{q_R}\eta_{r,t}, \\ d_t p &\approx k_p m - \gamma_p p + \sqrt{q_P}\eta_{p,t}. \end{aligned} \tag{4}$$

Here we have redefined $q_R = 2k_r$ and $q_p = 2k_p \langle m \rangle_\infty$ where $\langle m \rangle_\infty$ and $\langle p \rangle_\infty$ are the number of mRNA and protein molecules in the stationary state respectively. The random term $\eta_{r,t}$ has the mean $\langle \eta_{r,t} \rangle = 0$ and variance $\langle \eta_{r,t}\eta_{r,t'} \rangle = \delta(t - t')$. Similarly the random term $\eta_{p,t}$ associated with the number of protein molecules is such that $\langle \eta_{p,t} \rangle = \langle \eta_{r,t}\eta_{p,t} \rangle = 0$ and $\langle \eta_{p,t}\eta_{p,t'} \rangle = \delta(t - t')$. The general integral solutions to (4) can be given as follows:

$$\begin{aligned} m &= m_0 e^{-\gamma_r t} + e^{-\gamma_r t} \int_0^t (k_r + \eta_{m,s}) e^{\gamma_r s} ds, \\ p &= p_0 e^{-\gamma_p t} + e^{-\gamma_p t} \int_0^t (k_p m_s + \eta_{p,s}) e^{\gamma_p s} ds. \end{aligned} \tag{5}$$

Using Fourier transform methods, we can easily compute the stationary state means and variance associated with the number fluctuations in mRNA and protein molecules (e.g., Ref. [14], p. 42). Defining the Fourier transforms of the dynamical variables $p$ and $m$ in (4) as $\tilde{m}_w = \int_{-\infty}^{+\infty} m e^{-iwt} dt$ and $\tilde{p}_w = \int_{-\infty}^{+\infty} p e^{-iwt} dt$, (4) can be rewritten in the frequency domain $t \to w$ as follows:

$$\begin{aligned} \tilde{m}_w &= 2\pi k_r (\gamma_r + iw)^{-1} (\delta(w) + \eta_{r,w}), \\ \tilde{p}_w &= (\gamma_p + iw)^{-1} (k_p \tilde{m}_w + \eta_{p,w}) \\ &= (\gamma_p + iw)^{-1} (k_p 2\pi k_r (\gamma_r + iw)^{-1} (\delta(w) + \eta_{r,w}) + \eta_{p,w}). \end{aligned} \tag{6}$$

Here the noise terms are defined as $\eta_{r,w} = \int_{-\infty}^{+\infty} \eta_{r,t} e^{-iwt} dt$ and $\eta_{p,w} = \int_{-\infty}^{+\infty} \eta_{p,t} e^{-iwt} dt$, and we also have [14] the relationships $\langle \eta_{r,w} \rangle = \langle \eta_{p,w} \rangle = 0$, and $\langle \eta_{r,w}\eta_{r,w'} \rangle = \delta(w - w')$ and $\langle \eta_{p,w}\eta_{p,w'} \rangle = \delta(w - w')$. Using (6) we can derive the stationary state mean values associated with the fluctuations in the number of mRNA and protein molecules as follows:

$$\begin{aligned} \langle m \rangle_\infty &= k_r \int_{-\infty}^{+\infty} (\gamma_r + iw)^{-1} \delta(w) e^{iwt} dw = k_r \gamma_r^{-1}, \\ \langle p \rangle_\infty &= k_p \int_{-\infty}^{+\infty} (\gamma_p + iw)^{-1} (k_r (\gamma_r + iw)^{-1} \delta(w)) e^{iwt} dw = k_p k_r (\gamma_p \gamma_r)^{-1}. \end{aligned} \tag{7}$$

Similarly from (6) one can derive the following relationships:

$$\langle \tilde{m}_w \tilde{m}_{w'} \rangle = 4\pi^2 k_r^2 (\gamma_r + iw)^{-1} (\gamma_r + iw')^{-1} (\delta(w)\delta(w') + q_R \delta(w - w')), \tag{8}$$

$$\langle \tilde{p}_w \tilde{p}_{w'} \rangle = (\gamma_p + iw)^{-1} (\gamma_p + iw')^{-1} (q_P \delta(w - w') + k_p^2 \langle \tilde{m}_w \tilde{m}_{w'} \rangle). \tag{9}$$

From (8) and (9) one can derive the stationary-state zero-centered auto-covariance associated with the number of mRNA and protein molecules in the limit $w \to w'$ as fol-

lows:

$$\begin{aligned}\langle m^2\rangle_\infty &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(\gamma_r+iw)^{-1}(\gamma_r+iw')^{-1}(k_r^2\delta(w)\delta(w')\\ &\quad + q_R\delta(w-w'))e^{iwt}e^{iw't}dwdw'\\ &= k_r^2\gamma_r^{-2}+q_R(2\gamma_r)^{-1}, \qquad (10)\\ \langle p^2\rangle_\infty &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(\gamma_p+iw)^{-1}(\gamma_p+iw')^{-1}(k_p^2\langle\tilde{m}_w\tilde{m}_{w'}\rangle+q_P\delta(w-w'))dwdw'\\ &= k_p^2k_r^2\gamma_p^{-2}\gamma_r^{-2}+q_Rk_p^2(2\gamma_p\gamma_r)^{-1}(\gamma_p+\gamma_r)^{-1}+q_P(2\gamma_p)^{-1}. \qquad (11)\end{aligned}$$

From (7), (10–11) we finally obtain the stationary state mean centered auto-covariance associated with the fluctuations in the number of mRNA and protein molecules as follows:

$$\left.\begin{aligned}\langle m^2\rangle_\infty-\langle m\rangle_\infty^2 &= q_R(2\gamma_r)^{-1},\\ \langle p^2\rangle_\infty-\langle p\rangle_\infty^2 &= q_P(2\gamma_p)^{-1}+k_p^2q_R(2\gamma_r\gamma_p(\gamma_p+\gamma_r))^{-1}.\end{aligned}\right\} \qquad (12)$$

Now using the previously defined quantities $q_R = 2k_r$ and $q_p = 2k_p\langle m_\infty\rangle$ one can easily show that the stationary state Fano factor associated with the fluctuations in the number of mRNA and protein molecules are given as follows:

$$\left.\begin{aligned}F_{m,\infty} &= \langle m\rangle_\infty^{-1}(\langle m^2\rangle_\infty-\langle m\rangle_\infty^2)=1,\\ F_{p,\infty} &= \langle p\rangle_\infty^{-1}(\langle p^2\rangle_\infty-\langle p\rangle_\infty^2)=1+k_p(\gamma_p+\gamma_r)^{-1}.\end{aligned}\right\} \qquad (13)$$

Equation (13) clearly states that under singly-stochastic conditions the Fano factor associated with the fluctuations in the number of mRNA molecules will be equal to one. Now by defining translational efficiency as the ratio $\varepsilon = k_p\gamma_r^{-1}$ and noting the fact that $1 \gg (\gamma_p\gamma_r^{-1})$ i.e., the life times $\gamma_r^{-1}$ of mRNA transcript is much less than that of the life times $\gamma_p^{-1}$ of the protein molecule, one can easily obtain the relationship $F_{p,\infty} \approx 1+\varepsilon$ which indicates that the statistics of the fluctuations in the number of protein molecules significantly deviates from the Poisson as translational efficiency increases. However here one should note that the fluctuations in the number of mRNA molecules still obey the Poisson statistics since the Fano factor is $F_{m,\infty} = 1$. Similarly the coefficient of variation that is simply the ratio between the standard deviation and the mean, which is a statistical measure of the strength of the number fluctuations associated with various components of the gene expression system can be given as follows:

$$\left.\begin{aligned}C_{m,\infty} &= \sqrt{\gamma_rk_r^{-1}} = \langle m\rangle_\infty^{-1/2},\\ C_{p,\infty} &= \sqrt{\gamma_r\gamma_p(k_rk_p)^{-1}+\gamma_r\gamma_p(k_r(\gamma_p+\gamma_r))^{-1}} = \langle p\rangle_\infty^{-1/2}\sqrt{1+k_p(\gamma_p+\gamma_r)^{-1}}.\end{aligned}\right\} \qquad (14)$$

So far we have derived the statistical measures associated with the fluctuations in the number of various components in the singly-stochastic gene expression system. In the flowing sections we will show that this singly-stochastic assumption is not valid and we enumerate the various possibilities associated with the origin of such multiple-stochasticity in gene expression.

## 3 Sources of Multiple-Stochasticity in Gene Expression

Both transcription and translation consist of at least three sub processes viz. initiation, elongation and termination. Here, elongation and termination are somewhat thermodynamically driven processes and therefore the corresponding fluctuations are efficiently controlled. However, the initiation step is purely a stochastic process and therefore prone to significant amount of fluctuations. At the molecular level, transcription or translation initiation processes involves the recognition of the promoter sequences on DNA lattice by RNA polymerase enzyme (RNAP) or ribosome binding sites (RBS) on mRNA lattice by ribosome [16]. Here one should note that the process of initiation of transcription and translation can be modeled as a random walk with random step size where RNAP searches for the promoter sequences on DNA lattice in case of transcription and the small subunit of ribosome searches for the ribosome binding site (RBS) on mRNA lattice in the case of translation. In both cases, a protein molecule (RNAP or small subunit of ribosome) searches for its recognition site on the nucleic acid (DNA/mRNA) lattice starting from a random non-specific initial position. Therefore the time that is taken by the protein molecule to locate its specific site on the nucleic acid template will be a random quantity. Since the overall transcription process also includes the random initiation step, we can conclude that the overall transcription rate should be a time dependent random quantity. In the following sections we compute the central measure properties such as mean and the variance associated with these random transcription and translation rates.

Let us start with considering a stretch of linear DNA lattice of $N$ base-pairs in length containing the promoter, coding sequence and the terminator which are the minimal requirements of the simple house-keeping genes. Let us assume that the promoter which is situated at the lattice position $N_g$ such that $0 < N_g < N$ where the set of lattice points $\{0, N\}$ constitutes the reflecting [see Fig. 1 for details on the various boundary conditions discussed in this section] helical ends [7, 10] and the RNA-polymerase (RNAP) enzyme molecule is

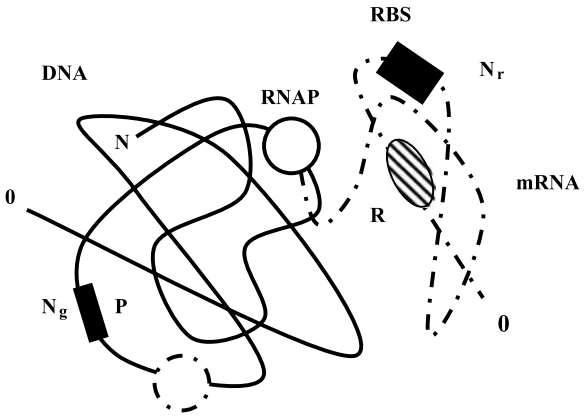


**Fig. 1** Various boundary conditions used in deriving the statistical measures associated with the fluctuations in transcription and translation initiation times. Here RNAP denotes the RNA polymerase enzyme, $P$ denotes the promoter sequences whose starting position is $N_g$ on DNA lattice, RBS denotes the ribosome binding site on mRNA lattice whose position is $N_r$ on mRNA lattice and $R$ denotes the ribosome. Here both RNAP and $R$ first non-specifically binds with the respective nucleic acid lattices and then search for their corresponding specific sites via one dimensional unbiased random jump dynamics along the respective nucleic acid lattices. Here the jump size associated with the dynamics is directly proportional to the degree of super-coiling/condensation of the respective nucleic acid templates

non-specifically bound at the lattice position $x = 0$ at time $t = 0$ and currently performing a one dimensional unbiased random walk search along DNA lattice to locate the promoter sequence in order to initiate transcription event which is then followed by the synthesis of mRNA transcript from DNA template. The probability of observing the non-specifically bound RNAP at an arbitrary position $x$ on DNA lattice at time $t > 0$ can be given as follows:

$$P(x,t|0,0) = 2N_g^{-1} \sum_{n=1}^{\infty} e^{-(2n-1)^2 8^{-1} N_g^{-2} \pi^2 D_g t} \cos((2n-1)\pi 2^{-1} N_g^{-1} x). \tag{15}$$

Equation (15) can be obtained by solving the one-dimensional Fokker-Plank equation $\partial_t P(x,t|0,0) = 2^{-1} D_g \partial_x^2 P(x,t|0,0)$ for the initial condition $P(x,0|0,0) = \delta(x)$ and the boundary conditions $\partial_x P_x(x,t|0,0)|_{x=0} = P(x,t|0,0)|_{x=N_g} = 0$, using bi-orthogonal eigenfunction expansion method [14, 15] in the interval $0 < x < N_g$ where $x = 0$ is reflecting boundary and the lattice point $x = N_g$ is the absorbing boundary and, $D_g = k_{dg} 3^{-1} s_r (s_r + 1)(2s_r + 1) \approx 2k_{dg} 3^{-1} s_r^3$ is the one dimensional phenomenological diffusion coefficient associated with the dynamics of the non-specifically bound RNA polymerase on DNA lattice. Here $s_r$ is the jump size, and $k_{dg}$ is the maximum three dimensional diffusion controlled collision rate associated with the dynamics of RNAP under solution conditions. Here one should note that the jump size $s_r$ can be positively correlated with the degree of super-coiling/condensation of the template DNA under consideration. Now, the first passage time $T$ taken by the RNAP to find the promoter for the first time starting from the DNA position $x = 0$ at time $t = 0$ can be shown [10] to be distributed $\omega(T|0)$ as follows:

$$\omega(T|0) = 2^{-1} \pi D_g N_g^{-2} \sum_{n=1}^{\infty} (-1)^{n+1} (2n-1) e^{-(2n-1)^2 8^{-1} N_g^{-2} \pi^2 D_g T}. \tag{16}$$

Equation (16) can be obtained [14] from the relationship $\omega(T|0) = -\int_0^{N_g} \partial_T P(x,T|0,0) dx$. Now the mean $m_r$ associated with the first passage time $T$ (MFPT) can be given as follows:

$$\langle T \rangle = \int_0^{\infty} T \omega(T|0) dT = N_g^2 D_g^{-1} = m_r. \tag{17}$$

Here one should note that the probability distribution function $\omega(T|0)$ given by (16) is strongly dependent on the jump size $s_r$ associated with the dynamics of the RNAP on DNA lattice in the process of searching for the promoter sequences since we have $D_g \approx 2k_{dg} 3^{-1} s_r^3$. Higher jump sizes will eventually narrow down the probability distribution function given by (16). This is a typical case where the fluctuations in the jump sizes can in turn reduce the fluctuations in the initiation times. Now, one can easily derive the expression for the variance associated with transcription initiation times as follows:

$$\langle T^2 \rangle - \langle T \rangle^2 = \varphi_r = 2N_g^4 3^{-1} D_g^{-2}. \tag{18}$$

From the relationship $D_g \approx 2k_{dg} 3^{-1} s_r^3$ and (18), it is clear that the variance associated with transcription initiation times decreases with the jump sizes $s_r$ associated with the dynamics of RNAP on DNA lattice as $\varphi_r \propto s_r^{-6}$. Here we should recall the fact that $T$ is the time required for the initiation of transcription which is then followed by the synthesis of mRNA transcript. If the time that is required to make a complete mRNA transcript by the RNAP machinery after transcription initiation event is $\tau_r$, then the average overall transcription rate is simply given as $k_r = \langle (T + \tau_r) \rangle^{-1} = (m_r + \tau_r)^{-1}$. Due to the

presence of in-homogeneities in the initiation times associated with transcription, $k_r$ is always a time dependent random quantity with definite mean and variance i.e., one can write $k_r \to (k_{r,t} = \mu_r + \xi_{r,t})$ so that the mean value is given as $\langle k_{r,t}\rangle = \mu_r + \langle \xi_{r,t}\rangle = \mu_r$. Though it is obvious to note that $k_r = \mu_r$, we have used different symbol just to discriminate the mean value of the time-dependent transcription rate $\mu_r$ from the normal time-independent transcription rate $k_r$. Here $\xi_{r,t}$ is the delta correlated Gaussian white noise process with mean $\langle \xi_{r,t}\rangle = 0$ and the variance $\langle \xi_{r,t}\xi_{r,t}\rangle = \phi_R$. Since we assumed a constant time for transcription elongation step, the variance associated with the fluctuations in the overall transcription times $(T + \tau_r)$ can be simply given as $\langle (T+\tau_r)^2\rangle - \langle (T+\tau_r)\rangle^2 = \varphi_r$. Apart from this, the variance associated with the fluctuations in the overall transcriptional rate $k_{r,t}$ should also be a function of $\varphi_r$ as $\langle \xi_{r,t}\xi_{r,t}\rangle = \phi_R = \Omega(\varphi_r)$. Noting the fact that $\mu_r \propto T^{-1}$ one can guess the functional relationship as $\phi_R \propto \varphi_r^{-1}$. To evaluate $\phi_R$ one need to transform the distribution function given by (16) as $\omega(T|0) \to \omega(\mu_r|0)$ using the variable transformation as $T \to \mu_r^{-1}$. The transformed distribution function $\omega(\mu_r|0)$ can be given [14] as $\omega(\mu_r|0) \propto 4\pi^{-1}\sum_{n=1}^{\infty}(-1)^{n+1}\mu_r^{-2}(2n-1)^{-1}e^{-(2n-1)^2 8^{-1} N_g^{-2}\pi^2 D_g \mu_r^{-1}}$. However, since the integrals $\int_0^\infty \mu_r \omega(\mu_r|0) d\mu_r \to \infty$ and $\int_0^\infty \mu_r^2 \omega(\mu_r|0) d\mu_r \to \infty$ diverge, we cannot use these transformations for computing the mean and variance associated with the rate constant variable $\mu_r$. Using the following simple arguments, one can show that the functional $\Omega(\varphi_r)$ is approximately a linear type when transcription elongation time $\tau_r$ is much higher than that of transcription initiation times which is true when the length of the transcripts is much higher. Let us assume that the lowest as well as highest values of the overall transcription times are given as $(T_{\min} + \tau_r)$ and $(T_{\max} + \tau_r)$ where $T_{\min}$ and $T_{\max}$ are chosen in such a way that the corresponding probabilities satisfying the inequalities $\omega(T_{\min}|0) \geq p_{\text{cutoff}}$ and $\omega(T_{\max}|0) \geq p_{\text{cutoff}}$. Here we should note that the cutoff probability $p_{\text{cutoff}}$ can be arbitrarily chosen. Now the effective range that is associated with transcription initiation times can be given as $\delta_T = (T_{\max} - T_{\min})$. On the other hand the corresponding range in the overall transcription rate for the same probability cutoff $p_{\text{cutoff}}$ can be given as $(T_{\min} + \tau_r)^{-1} - (T_{\max} + \tau_r)^{-1}$. When the elongation time $\tau_r$ is such that $T_{\min} \ll \tau_r$ and $T_{\max} \ll \tau_r$ then it is obvious to note that the range of the overall transcription rates can be approximately given as $\delta_\mu = \tau_r^{-2}(T_{\max} - T_{\min})$ which is simply the linear transformation of the range of overall transcription initiation times as $\delta_\mu = \tau_r^{-2}\delta_T$. Since the effective range of the probability distribution function is directly proportional to the variance as $\delta_T \propto \varphi_r$ and, therefore $\delta_\mu \propto \delta_T$ and $\delta_\mu \propto \varphi_R$, we can conclude that the functional $\Omega(\varphi_r)$ is correlated linearly with $\varphi_R$ as $\phi_R \propto \tau_r^{-2}\varphi_r$ when $T_{\min} \ll \tau_r$ and $T_{\max} \ll \tau_r$.

Now in case of translation, the small subunit of ribosome nonspecifically binds to mRNA transcript and then searches for the ribosome binding sites (RBS) via one-dimensional diffusion dynamics along mRNA lattice. Here the position $N_r$ of the RBS on mRNA lattice is such that $0 < N_r < N_m$, where $N_m$ is the total size of mRNA under consideration and the small subunit of the ribosome searches for the RBS (ribosome binding site) which acts as the absorbing boundary and the set of lattice points $\{0, N_m\}$ constitutes the reflecting boundaries. If the time that is necessary to make one fully translated polypeptide chain is $\tau_p$, then the overall translation rate is given as $k_p = (m_p + \tau_p)^{-1}$ where $m_p = N_r^2 D_r^{-1}$ is the mean time associated with the ribosome to find the ribosome binding site (RBS), starting from the position $x = 0$ on mRNA lattice at time $t = 0$ and $D_r = k_{dr} 3^{-1} s_p (s_p + 1)(2s_p + 1) \approx 2k_{dr} 3^{-1} s_p^3$ is the one-dimensional phenomenological diffusion coefficient associated with the dynamics of the ribosome. Here $k_{dr}$ is the three dimensional diffusion-controlled rate associated with the dynamics of the ribosome small subunit (RBS) in solution condition and $s_p$ is the jump size associated with the one dimensional diffusion of the ribosome on mRNA lattice. When we consider the in-homogeneities in translation initiation times, we obtain the time

dependent overall translation rate as $k_p \to (k_{p,t} = \mu_p + \xi_{p,t})$ with a mean value such that $\langle k_{p,t}\rangle = \mu_p$ and the variance such that $\langle k_{p,t}k_{p,t}\rangle = \langle \xi_{p,t}\xi_{p,t}\rangle$, where $\xi_{p,t}$ is a delta correlated Gaussian white noise with mean $\langle \xi_{p,t}\rangle = 0$ and variance $\langle \xi_{p,t}\xi_{p,t}\rangle = \phi_P$. Though it is obvious to note that $k_p = \mu_p$, we have used different symbol just to differentiate the mean value of time-dependent translation rate $\mu_p$ from the normal time-independent translation rate $k_p$. Here we should note that $\varphi_p = 2N_r^4 3^{-1} D_r^{-2}$ is the variance associated with translation initiation times and using the same arguments as in case of transcription we can show that $\langle \xi_{p,t}\xi_{p,t}\rangle = \phi_P = \Omega(\varphi_p)$, where $\Omega(\varphi_p)$ is approximately a linear function as $\phi_P \propto \tau_p^{-2}\varphi_p$ when translation elongation time $\tau_p$ is much higher than that of translation initiation times.

## 4 Multiple-Stochasticity in Gene Expression

With this background, the set of stochastic differential equations that describe the time evolution of the number of mRNA and protein molecules in the presence of in-homogeneities in the corresponding initiation times can be written as follows:

$$\begin{aligned} d_t m &= \mu_r + \xi_{r,t} - \gamma_r m + \eta_{r,t}, \\ d_t p &= \mu_p m + \xi_{p,t} m - \gamma_p p + \eta_{p,t}. \end{aligned} \tag{19}$$

Equation (19) describes a coupled two multiple-stochastic point processes where $\langle \xi_{r,t}\eta_{r,t}\rangle = 0$ and $\langle \xi_{p,t}\eta_{p,t}\rangle = 0$, and also $\langle \xi_{r,t}\xi_{p,t}\rangle = 0$. The general integral solution to (19) can be written as follows:

$$\begin{aligned} m &= m_0 e^{-\mu_m t} + e^{-\mu_r t}\int_0^t (\xi_{r,s} - \gamma_r m_s + \eta_{r,s}) e^{\mu_r s} ds, \\ p &= p_0 e^{-\mu_p t} + e^{-\mu_p t}\int_0^t (\mu_p m_s + \xi_{p,s} m_s - \gamma_p p_s + \eta_{p,s}) e^{\mu_p s} ds. \end{aligned} \tag{20}$$

Since we are interested in the stationary state means and the variances, we can easily obtain them from (19) by Fourier transform methods. Defining the Fourier transforms as we have done in the previous sections as $\tilde{m}_w = \int_{-\infty}^{+\infty} m e^{-iwt} dt$ and $\tilde{p}_w = \int_{-\infty}^{+\infty} p e^{-iwt} dt$ (19) will take the form as follows:

$$\tilde{m}_w = (\gamma_r + iw)^{-1}(2\pi\mu_r \delta(w) + \xi_{r,w} + \eta_{r,w}), \tag{21}$$

$$\tilde{p}_w = (\gamma_p + iw)^{-1}(\mu_p \tilde{m}_w + (\xi_{p,w} \otimes \tilde{m}_w) + \eta_{p,w}). \tag{22}$$

Now using the convolution theorem $(F(w) \otimes G(w)) = \int_{-\infty}^{+\infty} F(y)G(w-y)dy$, the convolution term $(\xi_{p,w} \otimes \tilde{m}_w)$ in (22) can be expanded as follows:

$$(\xi_{p,w} \otimes \tilde{m}_w) = 2\pi\mu_r\gamma_r^{-1}\xi_{p,w} + \int_{-\infty}^{+\infty} (\xi_{p,y}\xi_{m,w-y} + \xi_{p,y}\eta_{r,w-y})(\gamma_r + i(w-y))^{-1} dy. \tag{23}$$

Using (21–23) we can derive the stationary state mean values associated with the number of mRNA and protein molecules in the presence of in-homogeneities in transcription and translation rates as follows:

$$\begin{aligned} \langle m\rangle_\infty &= \mu_r \int_{-\infty}^{+\infty} \delta(w)(\gamma_r + iw)^{-1} e^{iwt} dw = \mu_r \gamma_r^{-1}, \\ \langle p\rangle_\infty &= \mu_p \int_{-\infty}^{+\infty} (\gamma_p + iw)^{-1} \tilde{m}_w e^{iwt} dw = \mu_p \mu_r (\gamma_p \gamma_r)^{-1}. \end{aligned} \tag{24}$$

Equation (24) clearly states that the mean values associated with the number of mRNA and protein molecules are not indeed affected by the presence of in-homogeneities in translation and transcription rates. From (21) and (22) we can drive the following relationships:

$$\begin{aligned}\langle \tilde{m}_w \tilde{m}_{w'} \rangle &= 4\pi^2 \mu_r^2 (\gamma_r + iw)^{-1} (\gamma_r + iw')^{-1} (\delta(w)\delta(w') \\ &\quad + q_R \delta(w - w') + \phi_R \delta(w - w')), \qquad (25)\\ \langle \tilde{p}_w \tilde{p}_{w'} \rangle &= (\gamma_p + iw)^{-1} (\gamma_p + iw')^{-1} (q_P \delta(w - w') + \mu_r \gamma_r^{-1} \phi_R + \mu_p^2 \langle \tilde{m}_w \tilde{m}_{w'} \rangle). \qquad (26)\end{aligned}$$

From (25) in the limit $w \to w'$, one can drive the stationary state zero centered auto co-variances associated with the number of mRNA as follows:

$$\begin{aligned}\langle m^2 \rangle_\infty &= \mu_r^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\gamma_m + iw)^{-1} (\gamma_m + iw')^{-1} \delta(w) e^{iwt} \delta(w') e^{iw't} dw dw' \\ &\quad + (q_R + \phi_R) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\gamma_m + iw)^{-1} (\gamma_m + iw')^{-1} \delta(w - w') e^{iwt} e^{iw't} dw dw' \\ &= k_r^2 \gamma_r^{-2} + (2\gamma_r)^{-1} (q_R + \phi_R). \qquad (27)\end{aligned}$$

Similarly from (26) in the limit $w \to w'$ one can drive the stationary state zero centered auto co-variance associated with the protein number fluctuations as follows:

$$\begin{aligned}\langle p^2 \rangle_\infty &= \mu_p^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\gamma_p + iw)^{-1} (\gamma_p + iw')^{-1} \langle \tilde{m}_w \tilde{m}_w \rangle e^{iwt} e^{iw't} dw dw' \\ &\quad + (q_P + \mu_r \phi_R \gamma_r^{-1}) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\gamma_p + iw)^{-1} (\gamma_p + iw')^{-1} \\ &\quad \times \delta(w - w') e^{iwt} e^{iw't} dw dw' \\ &= \mu_p^2 \mu_r^2 \gamma_p^{-2} \gamma_r^{-2} + \mu_p^2 (\phi_P + q_R)(2\gamma_p \gamma_r (\gamma_p + \gamma_r))^{-1} \\ &\quad + (2\gamma_p)^{-1} (q_P + \mu_r \phi_R \gamma_r^{-1}). \qquad (28)\end{aligned}$$

From (24–28) we finally obtain the stationary state mean centered auto co-variances associated with the fluctuations in the number of mRNA and protein molecules in the presence of in-homogeneities in both transcription and translation initiation times as follows:

$$\left.\begin{aligned}\langle m^2 \rangle_\infty - \langle m \rangle_\infty^2 &= q_R (2\gamma_r)^{-1} + \phi_R (2\gamma_r)^{-1}, \\ \langle p^2 \rangle_\infty - \langle p \rangle_\infty^2 &= q_P (2\gamma_p)^{-1} + \mu_p^2 q_R (2\gamma_p \gamma_r (\gamma_p + \gamma_r))^{-1} + \mu_p^2 \phi_P (2\gamma_p \gamma_r (\gamma_p + \gamma_r))^{-1} \\ &\quad + (2\gamma_p \gamma_r)^{-1} \mu_r \phi_R.\end{aligned}\right\} \qquad (29)$$

Equation (29) clearly states that unlike the mean values, the stationary state variances associated with the fluctuations in the number of mRNA and protein molecules are strongly influenced by the in-homogeneities in the initiation times of transcription and translation. As we have discussed in the introduction section, the multiple stochasticity eventually increases the overall disorder of the system.

## 5 Results and Discussion

Noting that $q_P = 2\mu_p\mu_r\gamma_r^{-1}$ and $q_R = 2\mu_r$, from (29) we can calculate the stationary state Fano factors $F_{m,\infty}$ and $F_{p,\infty}$ associated with the stationary state fluctuations in the number of mRNA and protein molecules as follows:

$$\begin{aligned} F_{m,\infty} &= 1 + \phi_R(2\mu_r)^{-1}, \\ F_{p,\infty} &= 1 + \mu_p(\gamma_p + \gamma_r)^{-1} + \mu_p\phi_P(2\mu_r(\gamma_p + \gamma_r))^{-1} + \phi_R(2\mu_p)^{-1}. \end{aligned} \tag{30}$$

Equations (29) and (30) are the central results of this paper which clearly states that the Fano factor associated with the fluctuations in the number of protein molecules is strongly influenced by the in-homogeneities in transcription as well as translation initiation times. From the comparison of (13) with (30) we predict that unlike singly stochastic gene expression, in case of multiple-stochastic gene expression, the Fano factor is *nonlinearly correlated with translation rate* $\mu_p$. Moreover upon equating the first derivative of $F_{p,\infty}$ with respect to $\mu_p$ as $\partial_{\mu_p} F_{p,\infty} = 0$ one finds that the Fano factor is a minimum at $\mu_p = \mu_{p,\min}$ where $\mu_{p,\min}$ is defined as $\mu_{p,\min} = \phi_R\mu_r(\gamma_p + \gamma_r)(2\mu_r + \phi_P)^{-1}$. This clearly indicates that by changing the rate constant parameters of the reaction system, the multiple-stochastic processes can be tuned to behave like the singly-stochastic processes. Figure 2 clearly demonstrates the behavior of the $F_{p,\infty}$ as the function of both $\mu_p$ and $\mu_r$. Similarly the coefficients variation $C_{m,\infty}$ and $C_{p,\infty}$ of the number fluctuations associated with the various components of the gene expression in the presence of in-homogeneities in the initiation times can be given as follows:

$$\begin{aligned} C_{m,\infty} &= \sqrt{\mu_r^{-1}\gamma_r}\sqrt{1 + \phi_R(2\mu_r)^{-1}}, \\ C_{p,\infty} &= \sqrt{\gamma_r\gamma_p(\mu_r\mu_p)^{-1}} \\ &\quad \times \sqrt{1 + \mu_p(\gamma_p + \gamma_r)^{-1} + \mu_p\phi_P(2\mu_r(\gamma_p + \gamma_r))^{-1} + \phi_R(2\mu_p)^{-1}}. \end{aligned} \tag{31}$$

Equation (31) clearly states that similar to the Fano factor, the coefficient of variation which is the measure of the overall fluctuations in the number of mRNA and the protein molecules also significantly increases with the in-homogeneities in transcription and translation

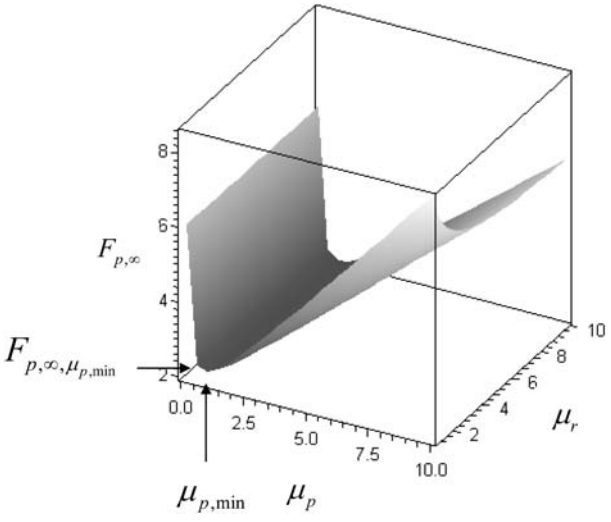


**Fig. 2** Nonlinear behavior of the stationary state Fano factor $F_{p,\infty}$ associated with fluctuations in the number of protein molecules with respect to translational rate and transcriptional rate. Equation (30) was used for plotting purpose. Here the assumed parametric values are $\phi_P = \phi_R = 1$ and $\gamma_p = \gamma_r = 1$

initiation times. Now we will show that the super coiling of DNA as well as mRNA lattice can reduce the increased level of fluctuations which is caused by the multiple-stochasticity in gene expression. RNA polymerase or the ribosome non-specifically binds with DNA or mRNA lattice in the first step and then searches for the corresponding promoter or ribosome binding sites via one-dimensional diffusion along DNA or mRNA lattice. Here one should note that the corresponding one-dimensional diffusion coefficient is connected with the jump sizes associated with the dynamics of RNAP or ribosome as $D_g \approx 2k_{dg}3^{-1}s_r^3$ or $D_r \approx 2k_{dr}3^{-1}s_p^3$ where $s_r$ and $s_p$ are the jump sizes of RNAP and ribosome respectively. When transcription as well as translation elongation times are much higher than that of the corresponding transcription $\bar{T}_r$ and translation $\bar{T}_p$ initiation times as $\bar{T}_r \ll \tau_r$ and $\bar{T}_p \ll \tau_p$ then variance associated with the overall translation and transcription rates $\phi_P = \Omega(\varphi_p)$ and $\phi_R = \Omega(\varphi_r)$ are approximately the linear functionals of the corresponding variances associated with the initiation times as $\phi_R \propto \tau_r^{-2}\varphi_r$ and $\phi_P \propto \tau_p^{-2}\varphi_p$, and therefore one can derive the scaling laws as $\phi_P \propto s_p^{-6}$ and $\phi_R \propto s_r^{-6}$ from the definitions of the variances of transcription and translation initiation times, i.e., $\varphi_r = 2N_g^4 3^{-1} D_g^{-2}$ and $\varphi_p = 2N_r^4 3^{-1} D_r^{-2}$. These scaling laws indirectly indicate that the super-coiling of DNA or mRNA lattice can in turn significantly reduce the fluctuations introduced by the multiple-stochasticity in gene expression since the jump size associated with such one-dimensional diffusion dynamics of RNAP or ribosome is in turn directly proportional to the degree of super coiling/condensation of the corresponding DNA or mRNA lattice. However, here one should note that the intrinsic fluctuations introduced by the stochastic forces $\eta_{r,t}$ and $\eta_{p,t}$ cannot be controlled by these supra molecular structural factors.

## 6 Conclusions

In summary, using the gene expression as a model system, we have shown that the multiple-stochasticity eventually increases the overall molecular number fluctuations in the chemical reaction systems. We showed that the main sources of this multiple-stochasticity could be the randomness in transcription and translation initiation times which in turn originate from the fluctuations in the underlying fundamental processes such as site-specific DNA-protein interactions and therefore these multiple-stochastic effects can be internally controlled by the supra molecular structural factors such as the super coiling of DNA lattice as in case of DNA-protein interactions. Analysis of the Fano factor as well as the coefficient of variation associated with the protein number fluctuations revealed that unlike the singly-stochastic case where the Fano factor is a monotonous function of translation rate, in case of multiple-stochastic gene expression the Fano factor is a turn-over function of translation rate with a definite minimum. This indicates that the multiple-stochastic processes can be well tuned to behave like the singly-stochastic point processes by adjusting the rate parameters.

**Acknowledgements** The author thanks Professor Gabriel Kreiman for offering a research position and financial support and the referee for the constructive comments which improved this manuscript a lot.

## References

1. Teich, M.C.: Fractal character of the auditory neural spike train. IEEE Trans. Biomed. Eng. **36**, 150–160 (1989)
2. Turcott, R.G., Lowen, S.B., Li, E., Johnson, D.H., Tsuchitani, C., Teich, M.C.: A nonstationary Poisson point process describes the sequence of action potentials over long time scales in lateral-superior-olive auditory neurons. Biol. Cybern. **70**, 209–217 (1994)

3. Lowen, S.B., Teich, M.C.: Doubly stochastic Poisson point process driven by fractal shot noise. Phys. Rev. A. **43**, 4192 (1991)
4. Lowen, S.B., Teich, M.C.: The periodogram and Allan variance reveal fractal exponents greater than unity in auditory-nerve spike trains. J. Acoust. Soc. Am. **99**, 3585–3591 (1996)
5. Lowen, S.B., Teich, M.C.: Fractal-Based Point Processes. Wiley-Interscience, New York (2005)
6. Murugan, R.: Critical jump sizes in DNA protein interactions. Biophys. Chem. **120**, 143–148 (2006)
7. Murugan, R.: DNA-protein interactions under random jump conditions. Phys. Rev. E **69**, 011911 (2004)
8. Murugan, R.: Effect of external fluctuations on the affinity-specificity negative correlation in DNA probe interactions. Phys. Rev. E **73**, 051915 (2006)
9. Murugan, R.: Generalized theory of site-specific DNA-protein interactions. Phys. Rev. E **76**, 011901 (2007)
10. Murugan, R.: Stochastic transcription initiation: time dependent transcription rates. Biophys. Chem. **121**, 51–56 (2006)
11. Zwanzig, R.: Rate processes with dynamical disorder. Acc. Chem. Res. **23**, 148–152 (1990)
12. Raser, J.M., O'Shea, E.K.: Noise in gene expression: origins, consequences, and control. Nature **309**, 2010–2013 (2005)
13. Berg, O.G.: A model for the statistical fluctuations of protein numbers in a microbial population. J. Theor. Biol. **71**, 587–603 (1978)
14. Risken, H.: Fokker–Plank Equations. Springer, Berlin (1992)
15. Gardiner, C.W.: Handbook of Stochastic Methods. Springer, Berlin (2002)
16. Lewin, B.: Genes VIII. Oxford University Press, London (2004)