Stability of stochastic gene regulatory networks using entropy methods *

M. Pájaro * A. A. Alonso ** J. A. Carrillo *** C. Vázquez ****

* Process Engineering Group, IIM-CSIC, Spanish Council for Scientific Research. Eduardo Cabello 6, 36208 Vigo, Spain (e-mail: mpajaro@iim.csic.es).
** Process Engineering Group, IIM-CSIC, Spanish Council for Scientific Research. Eduardo Cabello 6, 36208 Vigo, Spain (e-mail: antonio@iim.csic.es)
*** Department of Mathematics, Imperial College London, SW7 2AZ, UK (e-mail: carrillo@imperial.ac.uk)
**** Departament of Mathematics, Faculty of Informatics. Campus de Elviña, s/n 15071 - A Coruña, Spain (e-mail: carlosv@udc.es)

Abstract: The study of self regulated gene expression networks must be modelled using chemical master equations. However, its solution is not available in the most cases. In this work, we derive a partial integral differential model as the continuous counterpart of one master equation with jump process. This model allows us to reproduce numerically the dynamic behaviour of the protein distribution whose steady state admits an analytical solution. To study the convergence to the equilibrium, we test the applicability of entropy methods. Using these techniques we find numerical evidences of exponential stability. The derivation and methods presented can be of the help to extend the applicability of this model to more complex gene regulatory networks including more than one protein.

© 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Gene circuits, mathematical models, stochastic systems, entropy methods, stability, dynamics.

1. INTRODUCTION

The study of the DNA expression (transcription into messenger RNA and translation into proteins) and their regulation becomes essential to predict the response of cells to environmental signals. The regulatory mechanism normally takes place under the union of proteins to the DNA binding sites that inhibit or activate its expression. Typically, the number of molecules involved in the regulation mechanism is small, thus making gene expression a truly stochastic process (Gillespie, 2007; Kepler and Elston, 2001).

The chemical master equation (CME) is at the basis of dynamic reaction network modelling (Kepler and Elston, 2001; Paulsson, 2005; Mackey et al., 2011; Sherman and Cohen, 2014) as the method which incorporates the underlying stochastic behaviour. However, the CME solution cannot be obtained in most cases, due to the large (even infinite) number of coupled equations. Although computationally very involved, extensive stochastic simulations via SSA (Gillespie, 1976) are typically the approach adopted to reproduce the CME dynamics. Alternatives are CME approximations, such as, moment methods (Engblom, 2006), finite state projection (Munsky and Khammash, 2006) or hybrid models (Jahnke, 2011). Unfortunately, those methods are only able to approximate the CME solution in quite particular situations.

In case of gene self regulatory networks, the obstacles to the solution of the CME can be overcome by the 1D partial integral differential equation (PIDE) model proposed by Friedman et al. (2006). To the best of our knowledge, a rigorous deduction of the PIDE model from the CME has not been reported yet. Here we show that, under protein production in bursts (Friedman et al., 2006; Shahrezaei and Swain, 2008; Dar et al., 2012), the PIDE model can be deduced as the continuous counterpart of one CME with jump processes. Using this 1D PIDE model, we can both reproduce the dynamics of one protein distribution and obtain an analytical solution for the steady state.

In addition, we make use of an entropy method (Michel et al., 2005; Cáceres et al., 2011; Carrillo et al., 2011) to study stability of steady state solutions of the PIDE system. In particular, we test the applicability of entropy methods show asymptotic stability and give numerical evidences of the exponential rate of convergence.

The contribution is structured as follows: In Section 2 we discuss the gene regulatory network and its corresponding CME and PIDE dynamic descriptions. The entropy methods together with results on asymptotic and exponential stability are presented in Section 3. We end up with some conclusions and future work.
DNA off $\xrightarrow{k_{on}}$ mRNA $\xrightarrow{k_{m}}$ DNA on $\xrightarrow{k_{e}}$ X $\xrightarrow{\gamma_{m}}$ 0 $\xrightarrow{\gamma_{e}}$ 0

Fig. 1. Schematic representation of the transcription-translation mechanism under study. The promoter associated with the gene of interest is assumed to switch between active (DNA on) and inactive (DNA off) states, with rate constants $k_{on}$ and $k_{off}$ per unit time, respectively. In this study, the transition is assumed to be controlled by a feedback mechanism induced by the binding/unbinding of a given number of X-protein molecules, what makes the network self-regulated. Transcription of messenger RNA (mRNA) from the active DNA form, and translation into protein X are assumed to occur at rates (per unit time) $k_{m}$ and $k_{e}$, respectively. $k_{e}$ is the rate constant associated with transcriptional leakage. Both mRNA and X-protein degradation are assumed to occur by first order processes with rate constants $\gamma_{m}$ and $\gamma_{e}$, respectively.

2. SELF REGULATORY GENETIC SYSTEMS

The genetic system under study consists of a transcription-translation network involving a single gene that expresses a protein X which regulates its own production. The representative biochemical steps, including protein and mRNA degradation, are depicted in Fig. 1. We represent also a basal transcription level from the inactive promoter which gains at a rate constant $k_{e}$ lower than $k_{m}$, (Friedman et al., 2006; Ochab-Marcinek and Tabaka, 2015; Pájaro et al., 2015).

Typically, the self regulation mechanism is described by one input function of the form (Friedman et al., 2006; Ochab-Marcinek and Tabaka, 2015; Pájaro et al., 2015):

\[ c(x) = [1 - \rho(x)] + \rho(x)\varepsilon, \]

with $x$ representing protein level, $\varepsilon = \frac{k_{e}}{k_{m}} \in (0, 1)$ the transcriptional leakage constant and $\rho(x)$ a Hill type function (Alon, 2007) that relates $x$ to the fraction of DNA off:

\[ \rho(x) = \frac{x^{H}}{x^{H} + K^{H}}. \]

where $K = \frac{k_{e}x}{k_{m}}$ is an equilibrium constant and $H$ the Hill coefficient, proportional to the number of protein molecules bonded to the promoter. Its values can be positive or negative depending on whether the circuit represses or promotes protein production, thus resulting into a negative or positive feedback, respectively.

2.1 Continuous formulation deduction

In the following, we consider gene self regulatory networks where the degradation rate of mRNA is much faster than the corresponding to protein, so that $\gamma_{m}/\gamma_{e} \gg 1$. Such condition is verified in many gene regulatory networks, both in prokaryotic and eukaryotic organisms (Shahrezaei and Swain, 2008; Dar et al., 2012), and results in protein being produced in bursts. As suggested in Friedman et al. (2006); Elgart et al. (2011), the burst size (denoted by $b = \frac{x}{x_{c}}$) is typically modelled by an exponential distribution. The conditional probability for protein level to jump from a state $y$ to $x$ after a burst is proportional to:

\[ \omega(x - y) = \frac{1}{b} \exp \left[ -\frac{(x - y)}{b} \right] \]

This burst behaviour in protein production can be modelled by the superposition of jumps from lower states as it is depicted in Fig. 2. We define $g_{n}^{i} : \mathbb{N} \rightarrow [0, 1]$ as the transition probability for a jump going from a lower state $i$ into a state $n$, assuming that the size of the jump follows the expression (3). Furthermore, the transition probability is proportional to the messenger RNA production rate, so that, $g_{n}^{i}$ is defined as:

\[ g_{n}^{i} := k_{m}c(i)\omega(n - i). \]

Let $P : \mathbb{R}_{+} \times \mathbb{N} \rightarrow [0, 1]$, be the probability of having $n$ proteins at time $t$. The time evolution of $P(t, n)$ is given by a chemical master equation (Gardiner, 2009; Van Kampen, 2007) with jumps that reads:

\[ \frac{\partial P(t, n)}{\partial t} = \sum_{i=0}^{n-1} g_{i}^{n} P(t, i) - \sum_{i=n+1}^{\infty} g_{i}^{n} P(t, n) \]

\[ +r_{n+1} P(t, n + 1) - r_{n} P(t, n), \]

where $r_{n} = \gamma_{e} n$ represents the degradation transition probability. In order to obtain a continuous version of (5) we define $p : \mathbb{R}_{+} \times \mathbb{N} \rightarrow \mathbb{R}_{+}$, as the continuous protein probability distribution, and add and subtract $g_{n}^{i} P(t, n)$ at the right hand side of (5) to get:

\[ \sum_{i=0}^{n-1} g_{i}^{n} P(t, i) - \sum_{i=n}^{\infty} g_{i}^{n} P(t, n) = \sum_{i=0}^{n} g_{i}^{n} P(t, i) - \sum_{i=n}^{\infty} g_{i}^{n} P(t, n). \]

Next, approximating the summations at the right hand side of the last equation by integrals and substituting in (5) we obtain:

\[ \frac{\partial p(t, x)}{\partial t} = \int_{0}^{x} g_{y}^{n} p(t, y) \, dy - \int_{x}^{\infty} g_{y}^{n} p(t, x) \, dy \]

\[ +r_{x+1} p(t, x + 1) - r_{x} p(t, x), \]

where the integer indexes $n$ and $i$ are substituted by real $x$ and $y$ respectively. Note that the second term at right hand side in (7) reduces to:

\[ \int_{x}^{\infty} g_{y}^{n} p(t, y) \, dy = k_{m} c(x) p(t, x) \int_{x}^{\infty} \omega(y - x) \, dy, \]

with $\int_{x}^{\infty} \omega(y - x) \, dy = 1$. Employing the Taylor theorem to approximate the third term at right hand side in (7) to the first order, we also get:

\[ r_{x+1} p(t, x + 1) \approx r_{x} p(t, x) + \frac{\partial r_{x} p(t, x)}{\partial x}. \]

Finally, replacing the last expressions (8)-(9) in (7) and using a dimensionless time, $\tau = \gamma_{e} t$, associated with the time scale of protein degradation, we obtain the temporal evolution of the probability distribution $p(\tau, x)$, which reads as:
Fig. 2. Jump process representation of one protein produced in bursts, where one state \( n \) can be reached from lower states \( 0 \leq i < n \) with different transition probability functions \( g_{i}^{n} \). Analogously, from the state \( n \) the protein number can jump to higher states \( i \) with transition probability function \( g_{n}^{i} \). The degradation follows a one step process.

\[
\frac{\partial p(\tau, x)}{\partial \tau} = \frac{\partial [xp(\tau, x)]}{\partial x} - ac(x)p(\tau, x) + a \int_{0}^{x} \omega(x - y)c(y)p(\tau, y)dy,
\]

where \( a = k_{m}/\gamma_{x} \) is the dimensionless rate constant for transcription, which relates to the mean number of bursts produced per cell cycle (burst frequency). Note that (10) is equivalent to the equation proposed in Friedman et al. (2006).

3. ENTROPY METHODS

Let us introduce \( P_{\infty}(x) \) as the stationary solution of the equation (10), which verifies:

\[
\frac{\partial [xp(\infty, x)]}{\partial x} = ac(x)P_{\infty}(x) - a \int_{0}^{x} \omega(x - y)c(y)P_{\infty}(y)dy.
\]

Typically, the convex function \((u - 1)^{2}\) is used to explore the properties of the entropy functional (Michel et al., 2005; Cáceres et al., 2011; Carrillo et al., 2011), which is defined as:

\[
G_{2}(u) := \int_{0}^{\infty} \left( \frac{p}{P_{\infty}} - 1 \right)^{2} P_{\infty}dx = \int_{0}^{\infty} u^{2}P_{\infty}dx - 1
\]

where \( u(x) := \frac{p}{P_{\infty}}(x) \). As we will show, this functional is decreasing in time, what will allow us to conclude convergence. Before, let us consider the following result:

**Lemma 1.** The following equality holds:

\[
xp(x)\frac{\partial u(x)}{\partial x} = u(x)\frac{\partial [xp(x)]}{\partial x} - u^{2}(x)\frac{\partial [xP_{\infty}(x)]}{\partial x}
\]

**Proof.** First, we note that:

\[
u(x)\frac{\partial [xp(x)]}{\partial x} = u(x)p(x) + u(x)x\frac{\partial p(x)}{\partial x},
\]

\[
u^{2}(x)\frac{\partial [xP_{\infty}(x)]}{\partial x} = u^{2}(x)P_{\infty}(x) + u^{2}(x)x\frac{\partial P_{\infty}(x)}{\partial x}
\]

\[
xp(x)\frac{\partial u(x)}{\partial x} = xu(x)\frac{\partial p(x)}{\partial x} - xu^{2}(x)\frac{\partial P_{\infty}(x)}{\partial x}.
\]

Substituting the first two expressions at the right hand side of the third we get:

\[
xp(x)\frac{\partial u(x)}{\partial x} = u(x)\frac{\partial [xp(x)]}{\partial x} - u(x)p(x)
\]

\[-u^{2}(x)\frac{\partial [xP_{\infty}(x)]}{\partial x} + u^{2}(x)P_{\infty}(x),
\]

which is equivalent to (13) since \( u^{2}(x)P_{\infty}(x) = u(x)p(x) \).

**Proposition 2.** Let \( G_{2}(u) \) be the entropy functional, then the following equality is verified

\[
\frac{dG_{2}(u)}{d\tau} = -D_{2}(u),
\]

where

\[
D_{2}(u) = a \int_{0}^{\infty} \int_{y}^{\infty} \omega(x - y)[u(x) - u(y)]^{2}P_{\infty}(y)dx dy.
\]

**Proof.** The function \( p(\tau, x) \) is a probability mass distribution, so that the mass is conserved and the following integral vanishes:

\[
\int_{0}^{\infty} \frac{\partial p(\tau, x)}{\partial \tau}dx = \frac{\partial}{\partial \tau} \int_{0}^{\infty} p(\tau, x)dx = 0
\]

Using this property and the definition (12) we obtain the following equality:

\[
\frac{dG_{2}(u)}{d\tau} = 2 \int_{0}^{\infty} \left( \frac{p}{P_{\infty}} - 1 \right) \frac{\partial p}{\partial \tau}dx = 2 \int_{0}^{\infty} u(x) \frac{\partial p}{\partial \tau}dx
\]

Replacing the expression (10) for the time derivative of \( p \) in the last equation, we obtain:

\[
\frac{dG_{2}(u)}{d\tau} = 2 \int_{0}^{\infty} u(x) \left( \frac{\partial [xp(x)]}{\partial x} - ac(x)p(x) \right)dx + 2 \int_{0}^{\infty} u(x) \left( a \int_{0}^{x} \omega(x - y)c(y)p(x)dy \right)dx
\]

Integrating by parts the first term in the right hand side of (22) reads:

\[
2 \int_{0}^{\infty} u(x) \frac{\partial [xp(x)]}{\partial x}dx = -2 \int_{0}^{\infty} xp(x) \frac{\partial u(x)}{\partial x}dx
\]

and applying lemma 1 we get that:

\[
2 \int_{0}^{\infty} u(x) \frac{\partial [xp(x)]}{\partial x}dx = -2 \int_{0}^{\infty} \left( u(x) \frac{\partial [xp(x)]}{\partial x} - u^{2}(x) \frac{\partial [xP_{\infty}(x)]}{\partial x} \right)dx
\]

Substituting the last equality in (22) and replacing the partial derivatives in the integral by their expressions in (10) and (11) we obtain:

\[
\frac{dG_{2}(u)}{d\tau} = -2 \int_{0}^{\infty} u(x) \frac{\partial p}{\partial \tau}dx - 2a \int_{0}^{\infty} u(x)c(x)p(x)dx + 2u(x) \int_{0}^{\infty} x\omega(x - y)c(y)p(y)dy dx
\]

\[-2a \int_{0}^{\infty} u^{2}(x) \int_{0}^{\infty} x\omega(x - y)c(y)P_{\infty}(y)dy dx + 2a \int_{0}^{\infty} u^{2}(x)c(x)P_{\infty}(x)dx
\]

Note that, using the equality (21) and the fact that \( p(x) = u(x)P_{\infty}(x) \), the above expression simplifies into:
\[
\frac{dG_2(u)}{d\tau} = -a \int_0^\infty u^2(x)c(x)P_\infty(x)dx \\
+2a \int_0^\infty u(x) \int_0^x \omega(x-y)c(y)u(y)P_\infty(y)dydx \\
-a \int_0^\infty u^2(x) \int_0^x \omega(x-y)c(y)P_\infty(y)dydx 
\]

Changing the order of integration in (26) and reordering terms we arrive at:
\[
\frac{dG_2(u)}{d\tau} = a \int_0^\infty \int_y^\infty \left(2 \int_y^\infty \omega(x-y)u(x)dx\right) c(y)P_\infty(y)dy \\
- \int_y^\infty \omega(x-y)u^2(x)dx - u^2(y) c(y)P_\infty(y)dy.
\]

In addition, the equality \( \int_y^\infty \omega(x-y)dx = 1 \) is verified, so that, multiplying the term \( u^2(y) \) in (27) by this integral and taking common factor the last expression simplifies to:
\[
\frac{dG_2(u)}{d\tau} = a \int_0^\infty \int_y^\infty \left[2u(x)u(y) - u^2(x) - u^2(y)\right] \\
\times \omega(x-y)c(y)P_\infty(y)dxdy.
\]

Finally, \( 2u(x)u(y) - u^2(x) - u^2(y) = -(u(x) - u(y))^2 \) which concludes the proof.

The decay of the relative entropy in Proposition 2 can be generalized to a wider class of entropies. In fact, all convex functions of \( u \) can be used instead of the quadratic function. This leads to some consequences (Michel et al., 2005; Carrillo et al., 2011) we summarize next:

**Corollary 3.** If the initial condition for (10), \( p_0 \), is nonnegative, then its solution is nonnegative.

**Corollary 4.** Given any solution \( p \) with normalized initial data \( p_0 \) to equation (10), then the following properties hold:

(i) Contraction principle:
\[
\int_0^\infty |p(\tau,x)|dx \leq \int_0^\infty |p_0(x)|dx. \tag{29}
\]

(ii) \( L^q \) bounds, \( 1 < q < \infty \):
\[
\int_0^\infty P_\infty(x)|u(x)|^qdx \leq \int_0^\infty P_\infty(x)|u_0(x)|^qdx, \tag{30}
\]

with \( u_0(x) := \frac{p_0(x)}{P_\infty(x)} \).

(iii) Pointwise estimates:
\[
\inf_{x \in (0, \infty)} u_0(x) \leq u(x) \leq \sup_{x \in (0, \infty)} u_0(x). \tag{31}
\]

### 3.1 Numerical evidence of exponential stability

The PIDE model (10) is numerically solved by a semi-lagrangian scheme to obtain the transient solution for different times. The analytical solution for the steady state of (10) reads:
\[
P_\infty(x) := C(\rho(x))^{\frac{\alpha(u(x))}{\mu(x)}} x^{-(1-\alpha)} e^{-x}, \tag{32}
\]

where \( C \) is an integration constant that normalizes the corresponding probability distribution function. Thus, for each set of parameters, the expressions for \( G_2(u) \) and \( D_2(u) \) can be approximated in a straightforward manner by a composed trapezoidal quadrature formula. We select two examples representing negative and positive feedback (Pájaro et al., 2015) to show exponential convergence. Fig. 3 A depicts a unimodal stationary protein distribution obtained under negative feedback. For positive feedback, we choose one set of parameters for which bimodal behaviour emerges (Fig. 4 A). Fig. 3 B and Fig. 4 B represent \( G_2(u) \) and \( D_2(u) \) in a semi-logarithmic scale, what turns functions into straight lines, thus suggesting a exponential stability rate for both cases.

### 4. CONCLUSION AND FUTURE WORK

In this contribution we propose a continuous approximation of the CME with jumps events that leads to 1D PIDE model for gene self regulatory networks. In addition, we verify the applicability of entropy methods to study stability of equilibrium solutions, and give some numerical evidences of exponential convergence.

Future research directions will include modelling and stability of complex gene regulatory networks which include protein interactions under cross regulation. On the one hand, to extend the continuous PIDE approximation. On the other, to adapt entropy-based stability methods to the analysis of multiprotein systems.

**ACKNOWLEDGEMENTS**

This work has been partially supported by Ministerio de Economía y Competitividad AGL2012-39951-C02-01. M. Pájaro acknowledge support from grants BES-2013-063112 and EEIBB-1-16-10540.
REFERENCES

Alon, U. (2007). *An Introduction to Systems Biology. Design Principles of Biological Circuits.* Chapman & Hall/ CRC, London.

Cáceres, M.J., Cañizo, J.A., and Mischler, S. (2011). Rate of convergence to an asymptotic profile for the self-similar fragmentation and growth-fragmentation equations. *J. Math. Pures Appl.*, 96(4), 334–362.

Carrillo, J.A., Cordier, S., and Mancini, S. (2011). A decision-making fokker-planck model in computational neuroscience. *J. Math. Biol.*, 63(5), 801–830.

Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U.S.A.*, 109(43), 17454–17459.

Elgart, V., Jia, T., Fenley, A.T., and Kulkarni, R. (2011). Connecting protein and mRNA burst distributions for stochastic models of gene expression. *Phys. Biol.*, 8, 046001.

Engblom, S. (2006). Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comput.*, 180(2), 498–515.

Friedman, N., Cai, L., and Xie, X.S. (2006). Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys. Rev. Lett.*, 97(16), 168302.

Gardiner, C. (2009). *Stochastic Methods. A Handbook for the Natural and Social Sciences.* Springer, Berlin, fourth edition.

Gillespie, D.T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22(4), 403–434.

Gillespie, D.T. (2007). Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58, 35–55.

Huang, L., Yuan, Z., Liu, P., and Zhou, T. (2015). Effects of promoter leakage on dynamics of gene expression. *BMC Syst. Biol.*, 9(8).

Jahnke, T. (2011). On reduced models for the chemical master equation. *Multiscale Model. Simul.*, 9(4), 1646–1676.

Kepler, T.B. and Elston, T.C. (2001). Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.*, 81(6), 3116–3136.

López-Caamal, F. and Márquez-Lago, T.T. (2014). Order reduction of the chemical master equation via balanced realisation. *PLoS One*, 9(8).

Mackey, M.C., Tyran-Kamiska, M., and Yvinec, R. (2011). Molecular distributions in gene regulatory dynamics. *J. Theor. Biol.*, 274(1), 84–96.

Michel, P., Mischler, S., and Perthame, B. (2005). General relative entropy inequality: An illustration on growth models. *J. Math. Pures Appl.*, 84(9), 1235–1260.

Munsky, B. and Khammash, M. (2006). The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124(4), 1–12.

Ochab-Marcinek, A. and Tabaka, M. (2015). Transcriptional leakage versus noise: A simple mechanism of conversion between binary and graded response in auto-regulated genes. *Phys. Rev. E*, 91(1), 012704.

Pájaro, M., Alonso, A.A., and Vázquez, C. (2015). Shaping protein distributions in stochastic self-regulated gene expression networks. *Phys. Rev. E*, 92(3), 032712.

Paulsson, J. (2005). Models of stochastic gene expression. *Phys. Life Rev.*, 2(2), 157–175.

Shahrezaei, V. and Swain, P.S. (2008). Analytical distributions for stochastic gene expressions. *Proc. Natl. Acad. Sci. U.S.A.*, 105(45), 17256–17261.

Sherman, M.S. and Cohen, B.A. (2014). A computational framework for analyzing stochasticity in gene expression. *PLoS Comput. Biol.*, 10(5), 1003596.

Van Kampen, N.G. (2007). *Stochastic Processes in Physics and Chemistry.* Elsevier, Netherlands, third edition.