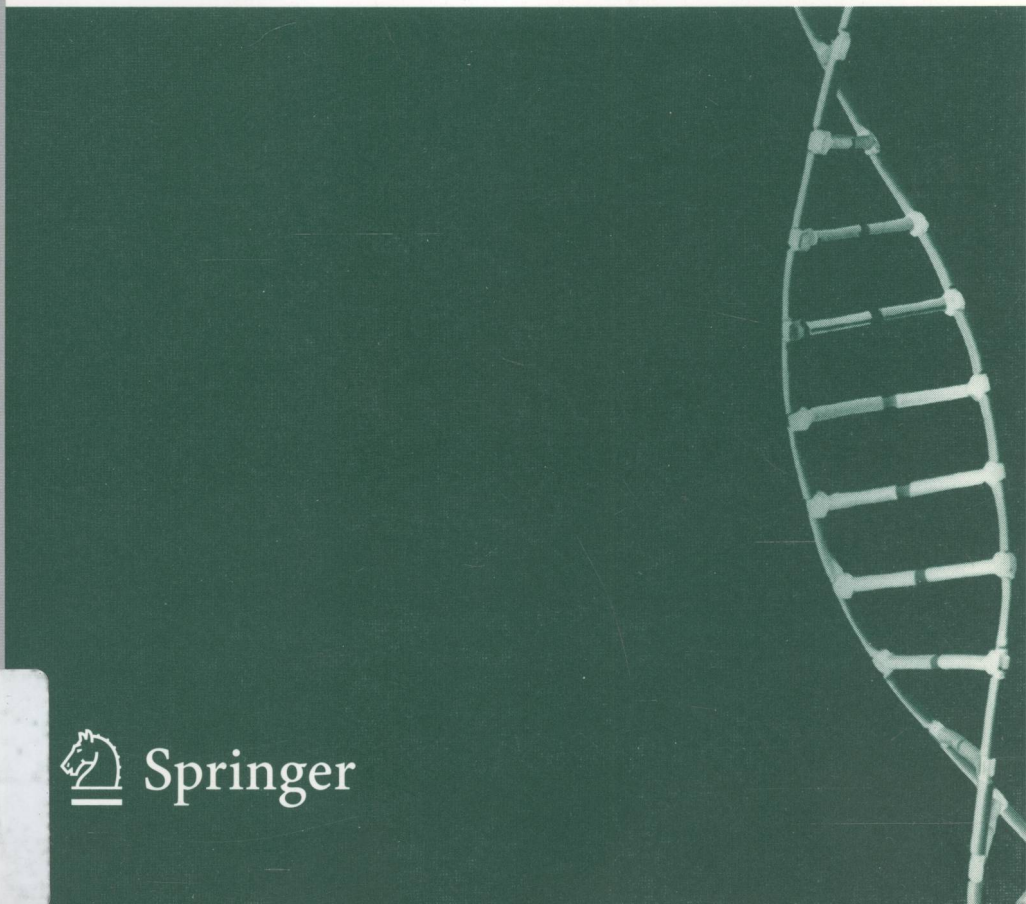
Muffy Calder
Stephen Gilmore (Eds.)

LNBI 4695

Computational Methods in Systems Biology

International Conference CMSB 2007
Edinburgh, Scotland, September 2007
Proceedings



Springer

Muffy Calder Stephen Gilmore (Eds.)

Computational Methods in Systems Biology

International Conference CMSB 2007
Edinburgh, Scotland, September 20-21, 2007
Proceedings

Springer

Volume Editors

Muffy Calder
Department of Computing Science
The University of Glasgow
Glasgow, Scotland
E-mail: muffy@dcs.gla.ac.uk

Stephen Gilmore
Laboratory for Foundations of Computer Science
The University of Edinburgh
Edinburgh, Scotland
E-mail: stg@inf.ed.ac.uk

Library of Congress Control Number: 2007934798

CR Subject Classification (1998): I.6, D.2.4, J.3, H.2.8, F.1.1

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743
ISBN-10 3-540-75139-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-75139-7 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12162642 06/3180 5 4 3 2 1 0

Lecture Notes in Bioinformatics 4695

Edited by S. Istrail, P. Pevzner, and M. Waterman

Subseries of Lecture Notes in Computer Science

Preface

Systems biology is an exciting new field bringing together life scientists, mathematicians, computer scientists and engineers to explore a new and deeper understanding of biological systems. Computational models and methods of analysis are essential components of the systems biology programme, not only reflecting, but also driving wet lab experimentation and the formation of new hypotheses about system behaviour.

This volume contains the proceedings of the fifth meeting of the international conference on Computational Methods in Systems Biology. The first conference was in Trento, Italy in 2003. The second meeting was in Paris in 2004, and in 2005 the conference came to Edinburgh for the first time. Last year's meeting was again in Trento and this year the conference was again in Edinburgh.

This year the conference attracted over 60 paper submissions. Sixteen of these were selected for presentation at the conference. In choosing the 16 best papers, the conference Chairs received wonderful support from the Programme Committee, who delivered thorough and insightful reviews of all papers in a very short time scale. We thank all of the members of the Programme Committee and their sub-referees for their industriousness. We also thank the authors for responding swiftly to the comments of the referees and revising their papers to address these comments earnestly.

The electronic submission of papers, refereeing and Programme Committee work were made possible by the excellent EasyChair free conference management system. EasyChair managed all of the aspects of the review process from submission to review and discussion, through to sending decisions by e-mail to authors. EasyChair compiled the list of referees which appears in this front matter. We give hearty thanks to Andrei Voronkov for providing this wonderful service to the scientific community.

The conference received financial support this year from the e-Science Institute, the Centre for Systems Biology in Edinburgh, and Microsoft Research, Cambridge. In addition, the Engineering and Physical Sciences Research Council supported the conference and contributed to the student bursaries, which we distributed to PhD students to allow them to attend the conference free of charge.

The conference this year was held in the e-Science Institute, Edinburgh. Lee Callaghan and the administrative team at the e-Science Institute provided excellent support for all of the organisational aspects of the conference, allowing us to concentrate on the technical aspects. We received additional support from the administrative staff in our respective departments, assisting with the preparation of this volume, and planning the associated opening reception and conference dinner.

We were very fortunate this year to have two outstanding invited speakers in Daniel T. Gillespie and Mark Girolami.

July 2007

Muffy Calder
Stephen Gilmore

Organization

The organisers and Co-chairs of the CMSB 2007 conference are Muffy Calder of the University of Glasgow and Stephen Gilmore of the University of Edinburgh.

Steering Committee

Finn Drabløs, Norwegian University of Science and Technology, Trondheim (Norway)
Monika Heiner, TU Cottbus (Germany)
Patrick Lincoln, Stanford Research International (USA)
Satoru Miyano, University of Tokyo (Japan)
Gordon Plotkin, University of Edinburgh (UK)
Corrado Priami, The Microsoft Research – University of Trento Centre for Computational and Systems Biology (Italy)
Magali Roux-Rouquié, CNRS-UPMC (France)
Vincent Schachter, Genoscope, Evry (France)
Adelinde Uhrmacher, University of Rostock (Germany)

Programme Committee

Alexander Bockmayr, Freie Universität Berlin (Germany)
Muffy Calder (Co-chair), University of Glasgow (UK)
Luca Cardelli, Microsoft Research Cambridge (UK)
Vincent Danos, CNRS, Université Denis Diderot (France)
Pierpaolo Degano, Università di Pisa (Italy)
Finn Drabløs, Norwegian University of Science and Technology, Trondheim (Norway)
François Fages, INRIA Rocquencourt (France)
Anthony Finkelstein, University College London (UK)
Stephen Gilmore (Co-chair), University of Edinburgh (UK)
David Harel, Weizmann Institute (Israel)
Monika Heiner, TU Cottbus (Germany)
Walter Kolch, Beatson Institute for Cancer Research (UK)
Ina Koch, Technische Fachhochschule Berlin (Germany)
Gethin Norman, University of Birmingham (UK)
Corrado Priami, The Microsoft Research – University of Trento Centre for Computational and Systems Biology (Italy)
Stephen Ramsey, Institute for Systems Biology, Seattle (USA)
Adelinde Uhrmacher, University of Rostock (Germany)

Referees

Joerg Ackermann
Paolo Ballarini
Roberto Barbuti
Maurice ter Beek
Andrea Bracciali
Linda Brodo
Muffy Calder
Enrico Cataldo
Matteo Cavaliere
Davide Chiarugi
Federica Ciocchetta
Pierpaolo Degano
Lorenzo Dematte
Ross Duncan
Jerome Feret
Anthony Finkelstein
Stephen Gilmore
Maria Luisa Guerriero
Jane Hillston
Bjorn Junker
Michal Kiwkowitz
Gunnar W. Klau
Sandy Klemm
Walter Kolch
Sriram Krishnamachari
Jean Krivine
Abdelhalim Larhlimi
Paola Lecca
Sebastian Lehrack
Roberto Maranngoni
Vikas Marda
Radu Mardare
Thierry Martinez
Patrick May
Paolo Milazzo
Orianne Mozemondet
Ivan Mura
David Parker
Dusko Pavlovic
Nadia Pisanti
Razvan Popescu
Heike Pospisil
Davide Prandi
Corrado Priami
Stephen Ramsey
Ronny Richter
Aurélien Rizk
Alessandro Romanel
Mehrnoosh Sadrzadeh
Yvonne Schmitz
Martin Schwarick
Heike Siebert
Sylvain Soliman
Lin Uhrmacher
Vlad Vyshemirsky
Malcolm Walkinshaw

Lecture Notes in Bioinformatics

Vol. 4751: G. Tesler, D. Durand (Eds.), Comparative Genomics. IX, 193 pages. 2007.

Vol. 4695: M. Calder, S. Gilmore (Eds.), Computational Methods in Systems Biology. X, 249 pages. 2007.

Vol. 4689: K. Li, X. Li, G.W. Irwin, G. He (Eds.), Life System Modeling and Simulation. XIX, 561 pages. 2007.

Vol. 4645: R. Giancarlo, S. Hannenhalli (Eds.), Algorithms in Bioinformatics. XIII, 432 pages. 2007.

Vol. 4643: M.-F. Sagot, M.E.M.T. Walter (Eds.), Advances in Bioinformatics and Computational Biology. XII, 177 pages. 2007.

Vol. 4544: S. Cohen-Boulakia, V. Tannen (Eds.), Data Integration in the Life Sciences. XI, 282 pages. 2007.

Vol. 4532: T. Ideker, V. Bafna (Eds.), Systems Biology and Computational Proteomics. IX, 131 pages. 2007.

Vol. 4463: I.I. Măndoiu, A. Zelikovsky (Eds.), Bioinformatics Research and Applications. XV, 653 pages. 2007.

Vol. 4453: T. Speed, H. Huang (Eds.), Research in Computational Molecular Biology. XVI, 550 pages. 2007.

Vol. 4414: S. Hochreiter, R. Wagner (Eds.), Bioinformatics Research and Development. XVI, 482 pages. 2007.

Vol. 4366: K. Tuyls, R.L. Westra, Y. Saeys, A. Nowé (Eds.), Knowledge Discovery and Emergent Complexity in Bioinformatics. IX, 183 pages. 2007.

Vol. 4360: W. Dubitzky, A. Schuster, P.M.A. Sloot, M. Schröder, M. Romberg (Eds.), Distributed, High-Performance and Grid Computing in Computational Biology. X, 192 pages. 2007.

Vol. 4345: N. Maglaveras, I. Chouvarda, V. Koutkias, R. Brause (Eds.), Biological and Medical Data Analysis. XIII, 496 pages. 2006.

Vol. 4316: M.M. Dalkilic, S. Kim, J. Yang (Eds.), Data Mining and Bioinformatics. VIII, 197 pages. 2006.

Vol. 4230: C. Priami, A. Ingólfsdóttir, B. Mishra, H. Riis Nielson (Eds.), Transactions on Computational Systems Biology VII. VII, 185 pages. 2006.

Vol. 4220: C. Priami, G. Plotkin (Eds.), Transactions on Computational Systems Biology VI. VII, 247 pages. 2006.

Vol. 4216: M. R. Berthold, R.C. Glen, I. Fischer (Eds.), Computational Life Sciences II. XIII, 269 pages. 2006.

Vol. 4210: C. Priami (Ed.), Computational Methods in Systems Biology. X, 323 pages. 2006.

Vol. 4205: G. Bourque, N. El-Mabrouk (Eds.), Comparative Genomics. X, 231 pages. 2006.

Vol. 4175: P. Bücher, B.M.E. Moret (Eds.), Algorithms in Bioinformatics. XII, 402 pages. 2006.

Vol. 4146: J.C. Rajapakse, L. Wong, R. Acharya (Eds.), Pattern Recognition in Bioinformatics. XIV, 186 pages. 2006.

Vol. 4115: D.-S. Huang, K. Li, G.W. Irwin (Eds.), Computational Intelligence and Bioinformatics, Part III. XXI, 803 pages. 2006.

Vol. 4075: U. Leser, F. Naumann, B. Eckman (Eds.), Data Integration in the Life Sciences. XI, 298 pages. 2006.

Vol. 4070: C. Priami, X. Hu, Y. Pan, T.Y. Lin (Eds.), Transactions on Computational Systems Biology V. IX, 129 pages. 2006.

Vol. 4023: E. Eskin, T. Ideker, B. Raphael, C. Workman (Eds.), Systems Biology and Regulatory Genomics. X, 259 pages. 2007.

Vol. 3939: C. Priami, L. Cardelli, S. Emmott (Eds.), Transactions on Computational Systems Biology IV. VII, 141 pages. 2006.

Vol. 3916: J. Li, Q. Yang, A.-H. Tan (Eds.), Data Mining for Biomedical Applications. VIII, 155 pages. 2006.

Vol. 3909: A. Apostolico, C. Guerra, S. Istrail, P. Pevzner, M. Waterman (Eds.), Research in Computational Molecular Biology. XVII, 612 pages. 2006.

Vol. 3886: E.G. Bremer, J. Hakenberg, E.-H.(S.) Han, D. Berrar, W. Dubitzky (Eds.), Knowledge Discovery in Life Science Literature. XIV, 147 pages. 2006.

Vol. 3745: J.L. Oliveira, V. Maojo, F. Martín-Sánchez, A.S. Pereira (Eds.), Biological and Medical Data Analysis. XII, 422 pages. 2005.

Vol. 3737: C. Priami, E. Merelli, P. Gonzalez, A. Omicini (Eds.), Transactions on Computational Systems Biology III. VII, 169 pages. 2005.

Vol. 3695: M. R. Berthold, R.C. Glen, K. Diederichs, O. Kohlbacher, I. Fischer (Eds.), Computational Life Sciences. XI, 277 pages. 2005.

Vol. 3692: R. Casadio, G. Myers (Eds.), Algorithms in Bioinformatics. X, 436 pages. 2005.

Vol. 3680: C. Priami, A. Zelikovsky (Eds.), Transactions on Computational Systems Biology II. IX, 153 pages. 2005.

Vol. 3678: A. McLysaght, D.H. Huson (Eds.), Comparative Genomics. VIII, 167 pages. 2005.

Vol. 3615: B. Ludäscher, L. Raschid (Eds.), Data Integration in the Life Sciences. XII, 344 pages. 2005.

Vol. 3594: J.C. Setubal, S. Verjovski-Almeida (Eds.), Advances in Bioinformatics and Computational Biology. XIV, 258 pages. 2005.

Vol. 3500: S. Miyano, J. Mesirov, S. Kasif, S. Istrail, P. Pevzner, M. Waterman (Eds.), Research in Computational Molecular Biology. XVII, 632 pages. 2005.

Vol. 3388: J. Lagergren (Ed.), Comparative Genomics. VII, 133 pages. 2005.

Vol. 3380: C. Priami (Ed.), Transactions on Computational Systems Biology I. IX, 111 pages. 2005.

Vol. 3370: A. Konagaya, K. Satou (Eds.), Grid Computing in Life Science. X, 188 pages. 2005.

Vol. 3318: E. Eskin, C. Workman (Eds.), Regulatory Genomics. VII, 115 pages. 2005.

Vol. 3240: I. Jonassen, J. Kim (Eds.), Algorithms in Bioinformatics. IX, 476 pages. 2004.

Vol. 3082: V. Danos, V. Schachter (Eds.), Computational Methods in Systems Biology. IX, 280 pages. 2005.

Vol. 2994: E. Rahm (Ed.), Data Integration in the Life Sciences. X, 221 pages. 2004.

Vol. 2983: S. Istrail, M.S. Waterman, A. Clark (Eds.), Computational Methods for SNPs and Haplotype Inference. IX, 153 pages. 2004.

Vol. 2812: G. Benson, R.D.M. Page (Eds.), Algorithms in Bioinformatics. X, 528 pages. 2003.

Vol. 2666: C. Guerra, S. Istrail (Eds.), Mathematical Methods for Protein Structure Analysis and Design. XI, 157 pages. 2003.

Table of Contents

Chemical Master Equation and Langevin Regimes for a Gene Transcription Model

Raya Khanin[1] and Desmond J. Higham[2]

[1] University of Glasgow, Glasgow, G12 8QQ , UK
[2] University of Strathclyde, Glasgow, G1 1XH, UK

Abstract. Gene transcription models must take account of intrinsic stochasticity. The Chemical Master Equation framework is based on modelling assumptions that are highly appropriate for this context, and the Stochastic Simulation Algorithm (also known as Gillespie's algorithm) allows for practical simulations to be performed. However, for large networks and/or fast reactions, such computations can be prohibitatively expensive. The Chemical Langevin regime replaces the massive ordinary differential equation system with a small stochastic differential equation system that is more amenable to computation. Although the transition from Chemical Master Equation to Chemical Langevin Equation can be heuristically justified, there is very little guidance available about how closely the two models match. Here, we consider a transcription model from the recent literature and show that it is possible to compare first and second moments in the two stochastic settings. To analyse the Chemical Master Equation we use some recent work of Gadgil, Lee and Othmer, and to analyse the Chemical Langevin Equation we use Ito's Lemma. We find that there is a perfect match—both modelling regimes give the same means, variances and correlations for all components in the system. The model that we analyse involves 'unimolecular reactions', and we finish with some numerical simulations involving dimerization to show that the means and variances in the two regimes can also be close when more general 'bimolecular reactions' are involved.

1 Background

Several experimental techniques are now available to measure gene expression, even at the single cell level [1,2,3]. In parallel, mathematical models and simulation algorithms have been developed to explain these observations and make new predictions [4,5,6,7,8,9,10]. Key modeling and simulation challenges in this area are that (a) some components may be present in relatively small quantities, (b) there can be a wide range of natural time scales in operation, and (c) on the level at which observations are made, the process is inherently stochastic. A Markov process, or *Chemical Master Equation* (CME) framework is highly appropriate in this context, and is now widely used. The CME methodology and an accompanying simulation algorithm can be traced back to the work of Gillespie in the chemical kinetics literature [11,12]. Recent overviews can be found

M. Calder and S. Gilmore (Eds.): CMSB 2007, LNBI 4695, pp. 1–14, 2007.

in [6,13,14] and we note that there are close connections to Petri nets, discrete event simulation and birth-and-death processes [15].

Because the CME framework takes account of every reaction, for many realistic models it is too computationally expensive to be useful. The *Chemical Langevin Equation* (CLE) provides an alternative model that retains some of the main features of the CME whilst making simulations more feasible. The CLE, which takes the form of an Ito stochastic differential equation (SDE), can be derived from the CME via a series of reasonable modeling assumptions [16], and under the extreme case where fluctuations in the CLE are ignored, we recover the traditional deterministic *Reaction Rate Equation* (or *Law of Mass Action*). Many authors are now developing *multi-scale* simulation methods that automatically operate in the cheapest modeling regime that captures the appropriate behaviour [17,18]. For this reason it is important to have an understanding of how the different modelling regimes compare. This motivates the work here, where the means and variances of the CME and CLE are compared for a recent gene transcription model. To analyse the CME we make use of the general first-order reaction theory of Gadgil et al. [19] and to analyse the CLE we perform what appears to be the first application of Ito's lemma in this context.

The article is organised as follows. In the next section we give a very simple example that illustrates the main concepts involved in our work. Then in section 3 we set up the general specification of the CME and CLE and introduce Ito's lemma. The gene regulation model is described in section 4 and moments for the CME and CLE are derived analytically in sections 5 and 6 respectively. A numerical experiment involving dimerization is given in section 7 to show that similar behaviour can also arise when we leave the first-order realm.

2 Illustrative Example: Unimolecular Decay

To illustrate the ideas in this work, we begin with the simplest possible type of reaction; unimolecular decay. We suppose that there is only one species, S, in our system, and the only event that can take place at any time is that one molecule of S may decay. We could write the system symbolically as

$$S \xrightarrow{c} \emptyset \tag{1}$$

Here, $c > 0$ is a constant that relates to the propensity of the decay process.

We suppose that initially, at time $t = 0$, the number of molecules of S is known to be N. The state of the system at time t is fully described by a non-negative integer $X(t)$, representing the number of molecules of S present. So $X(t)$ may take any of the values $N, N-1, N-2, \ldots, 1, 0$. In the CME setting we regard $X(t)$ as a discrete-valued random variable at each point in time, and work in terms of the probability $p_i(t)$ that $X(t) = i$, arriving at the ordinary differential equation (ODE) system

$$\frac{d}{dt}p_N(t) = -cNp_N(t), \tag{2}$$

$$\frac{d}{dt}p_i(t) = c \cdot (i+1) \cdot p_{i+1}(t) - c \cdot i \cdot p_i(t), \quad \text{for } i = N-1, N-2, \ldots, 0. \tag{3}$$

The general ODE (3) has a natural interpretation. The rate of change of $p_i(t)$ has a positive contribution $c \cdot (i+1) \cdot p_{i+1}(t)$, which corresponds to the fact that we arrive at state i via one decay from state $i+1$. Conversely, there is a negative contribution $-c \cdot i \cdot p_i(t)$ due to the fact that, when in state i, we leave that state when a decay takes place.

The system (2)–(3) is readily solved to give

$$p_i(t) = \frac{N!}{i!(N-i)!} e^{-cit} \left(1 - e^{-ct}\right)^{N-i}, \quad \text{for } i = 0, 1, 2, \ldots, N. \tag{4}$$

Using $\mathbb{E}[\cdot]$ and $\mathsf{Var}[\cdot]$ to denote the mean and variance, respectively, it follows that

$$\mathbb{E}\left[X(t)\right] = Ne^{-ct} \quad \text{and} \quad \mathsf{Var}\left[X(t)\right] = Ne^{-ct}\left(1 - e^{-ct}\right). \tag{5}$$

Details can be found, for example, in [20] by observing that this system corresponds to a pure death process in the context of stochastic population modelling.

In the CLE setting, we represent the amount of species S present at time t by the real-valued stochastic process $Y(t)$. In other words, at each time t, $Y(t)$ is a continuous-valued random variable. The CLE is then the Ito SDE [21,22]

$$dY(t) = -cY(t)\,dt - \sqrt{cY(t)}\,dW(t), \quad Y(0) = N. \tag{6}$$

Because the drift coefficient $-cY(t)$ is linear, it follows immediately that $\mathbb{E}\left[Y(t)\right]$ satisfies the ODE that arises when the noise is switched off, giving

$$\mathbb{E}\left[Y(t)\right] = Ne^{-ct}. \tag{7}$$

To find the second moment, we may apply Ito's lemma, as described in section 3.2, to get

$$\frac{d}{dt}\mathbb{E}\left[Y(t)^2\right] = -2\,c\,\mathbb{E}\left[Y(t)^2\right] + c\,\mathbb{E}\left[Y(t)\right].$$

Using the expression (7), this solves to give $\mathbb{E}\left[Y(t)^2\right] = Ne^{-ct}$, so that

$$\mathsf{Var}\left[Y(t)\right] = Ne^{-ct}\left(1 - e^{-ct}\right). \tag{8}$$

Comparing (7) and (8) with (5), we see that the models give precisely the same expressions for the mean and variance of S. This happens despite the fact that one uses the discrete, integer-valued state vector $X(t)$ and the other uses the real-valued $Y(t)$.

For completeness, we mention that the law of mass action, or reaction rate equation, formulation for the system (1) has the form of a scalar ODE $dz(t)/dt = -cz(t)$, where $z(t)$ is a deterministic real-valued quantity representing the amount of S present at time t. This is precisely the ODE for the mean in the CLE, and hence $z(t) = \mathbb{E}[Y(t)] = Ne^{-ct}$.

Two features of the CLE (6) for this simple model are generic.

1 The diffusion coefficient is nonlinear.
2 The description of the problem involves a square root, and hence the problem is only well defined if the solution remains non-negative.

With regard to the second point, the particular CLE (6) is a special case of a *square root process*. These SDEs are widely used as interest rate models in mathematical finance, and it can be shown that the solution in (6) maintains non-negativity with probability one [22]. However, we note that the issue of negative solutions seems to be open for general CLEs. In this work, we will always assume that the CLE has a well-defined, unique solution.

The main result in this article is that the coincidence of CME and CLE mean and variance in the simple model (1) carries through to a gene transcription model.

3 Stoichiometric Formalization

3.1 Chemical Master Equation

Suppose that there are N chemical species, $S_1, S_2, \ldots, S_N$ taking part in M different chemical reactions. In the CME formulation, we have a state vector $\mathbf{X}(t) \in \mathbb{R}^N$ whose ith component, $X_i(t)$, denotes the number of molecules of S_i present at time t. Hence, each $X_i(t)$ is a non-negative integer. For each $1 \leq j \leq M$ we have a *stoichiometric vector* $\boldsymbol{\nu}_j \in \mathbb{R}^N$, and *propensity function* $a_j(\mathbf{X}(t))$, such that the jth reaction takes place over the infinitesimal interval $[t, t+dt)$ with probability $a_j(\mathbf{X}(t))\,dt$ and causes the change $\mathbf{X}(t) \mapsto \mathbf{X}(t) + \boldsymbol{\nu}_j$ to the state vector.

Letting $P(\mathbf{x}, t)$ denote the probability that $\mathbf{X}(t) = \mathbf{x}$, the CME is the ODE system

$$\frac{dP(\mathbf{x},t)}{dt} = \sum_{j=1}^{M} \left(a_j(\mathbf{x}-\boldsymbol{\nu}_j)P(\mathbf{x}-\boldsymbol{\nu}_j,t) - a_j(\mathbf{x})P(\mathbf{x},t)\right). \tag{9}$$

Generally, the CME cannot be solved analytically in any useful way, although Gillespie's *Stochastic Simulation Algorithm* (SSA) [11,12] gives a way to compute realisations of $\{t, \mathbf{X}(t)\}$ that respect the CME. However, in the case where all reactions are unimolecular (or first-order), detailed analysis is possible, both for the first and second moments [19] and the general distributions [23]. In this work we will show that, at least for a specific gene regulation model, useful analytical results can also be derived for the CLE formulation described in the next subsection.

3.2 Chemical Langevin Equation

The CLE uses a real-valued random variable $\mathbf{Y}(t) \in \mathbb{R}^N$ to describe the state of the system at time t. The jth component $Y_j(t)$ represents the amount of species j. In moving from the CME to the CLE we (typically) make a dramatic

reduction in the number of components, but pay the price that each component is a real-valued random variable, rather than a non-negative integer. The CLE takes the form of an Ito SDE [21,22]

$$d\,\mathbf{Y}(t) = \sum_{j=1}^{M} \boldsymbol{\nu}_j\, a_j(\mathbf{Y}(t))\, dt + \sum_{j=1}^{M} \boldsymbol{\nu}_j \sqrt{a_j(\mathbf{Y}(t))}\, dW_j(t), \tag{10}$$

where the $\{W_j(t)\}_{j=1}^{M}$ are independent Brownian motions.

As background for the SDE analysis in section 6, we now state the relevant part of Ito's lemma; see, for example, [22]. For the general Ito SDE system with n components and d independent Brownian motions

$$dY_i(t) = b_i\left(\mathbf{Y}(t)\right) dt + \sum_{j=1}^{d} \sigma_{ij}\left(\mathbf{Y}(t)\right) dW_j(t), \quad 1 \le i \le n, \tag{11}$$

we let

$$a\left(\mathbf{Y}(t)\right) := \boldsymbol{\sigma}\left(\mathbf{Y}(t)\right) \boldsymbol{\sigma}\left(\mathbf{Y}(t)\right)^T \in \mathbb{R}^{n\times n}. \tag{12}$$

Then for any function $f : \mathbb{R}^n \to \mathbb{R}$ that is twice continuously differentiable, Ito's lemma tells us that

$$\begin{aligned} d\,f\left(\mathbf{Y}(t)\right) = &\left(\sum_{i=1}^{n} \frac{\partial f\left(\mathbf{Y}(t)\right)}{\partial x_i} b_i\left(\mathbf{Y}(t)\right) + \tfrac{1}{2}\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f(\mathbf{Y}(t))}{\partial x_i \partial x_j} a_{ij}\left(\mathbf{Y}(t)\right)\right) dt \\ &+ \text{mart.}, \end{aligned} \tag{13}$$

where "mart." denotes a martingale whose precise form is not relevant to our work. We will use two particular cases of f. When $f\left(\mathbf{Y}\right) = Y_k^2$, (13) becomes

$$d\left(Y_k^2\right) = \left(2\, Y_k\, b_k\left(\mathbf{Y}(t)\right) + a_{kk}\left(\mathbf{Y}(t)\right)\right) dt + \text{mart.} \tag{14}$$

and when $f\left(\mathbf{Y}\right) = Y_k Y_l$, for $k \neq l$, it becomes

$$d\left(Y_k Y_l\right) = \left(Y_l b_k\left(\mathbf{Y}(t)\right) + Y_k b_l\left(\mathbf{Y}(t)\right) + \tfrac{1}{2} a_{kl}\left(\mathbf{Y}(t)\right) + \tfrac{1}{2} a_{lk}\left(\mathbf{Y}(t)\right)\right) dt + \text{mart.} \tag{15}$$

4 Gene Regulation Model

We now consider a model of eukaryotic gene regulation, originally proposed in [24]. This model incorporates two states of promoters: an inactive state, D, not permissive of transcription, and an active state $D^\star$ that is competent for transcription. Transition between the two states of promoter is reversible and the total number of promoters is conserved, i.e. $D + D^\star = D_T$. Transcription takes place from the active state $D^\star$ with the linear rate k_r, resulting in production of messenger RNA (mRNA) molecules that decay with rate γ_r. Proteins P are translated from mRNA molecules with linear rate k_p and they decay with rate γ_p.