Cellular growth and division in the Gillespie algorithm

T. Lu, D. Volfson, L. Tsimring and J. Hasty

Abstract: Recent experimental studies elucidating the importance of noise in gene regulation have ignited widespread interest in Gillespie's stochastic simulation technique for biochemical networks. We formulate modifications to the Gillespie algorithm which are necessary to correctly simulate chemical reactions with time-dependent reaction rates. We concentrate on time dependence of kinetic rates arising from the periodic process of growth and division of the cellular volume, and demonstrate that a careful re-derivation of the Gillespie algorithm is important when all stochastically simulated reactions have rates slower or comparable to the cellular growth rate. For an unregulated single-gene system, we illustrate our findings using recently proposed hybrid simulation techniques, and systematically compare our algorithm with analytic results obtained from the chemical master equation.

1 Introduction

Successful completion of the various genome projects has led to the realisation that effective models for predicting cellular behaviour must take into account the network interactions that dynamically mediate gene regulation. Since behaviour arising from these complex interactions is difficult to predict without quantitative models, there is a need for experimentally validated computational modelling approaches that can be used to understand the complexities of gene regulation. Such model approaches will be invaluable in the generation of logically consistent hypotheses and will provide a framework for the systematic comparison of data across multiple experiments.

There is strong experimental evidence that the level of expression from the same gene varies significantly from one cell to another within a genetically-identical colony [1–8]. Such variations are routinely observed in the cells of organisms ranging in complexity from bacteria to mammals. Theoretically, with mRNA numbers that are often less than ten, the stochastic nature of the underlying biochemical reactions must lead to large fluctuations. While the importance of fluctuations in this context was stressed over thirty years ago [9], the recent experimental evidence has revived interest in the utilisation of stochastic simulation techniques to model gene regulatory networks [4, 10–14].

When fluctuations arise from the small number of reactant molecules, the stochastic simulation algorithm developed by Gillespie is considered the 'gold standard' for modelling [15, 16]. The advantage of this algorithm is that it generates an ensemble of trajectories with correct statistics for a set of biochemical reactions. However, both the original algorithm and later developments have focused on systems with fixed volumes, and have not systematically considered the effect of volume changes on the simulation routine [17–19]. In the context of genetic regulatory and protein-protein interaction networks, the processes of cellular growth and division can play an important role since the cell volume is inversely related to the association rate for a bimolecular reaction (Fig. 1). Therefore, the extension of the Gillespie algorithm to account for volume growth is an important and necessary developmental step. Here we extend the Gillespie simulation routine to account for the deterministic time dependence of rate constants arising from cellular growth and division. Although our results are general for simulations involving time-dependent rate constants, we focus on the derivation of an algorithm for cellular systems with changing volume. Using an analytically tractable model system, we compare our generalised algorithm with a standard Gillespie approach, and find that the proposed modifications are significant when all stochastically simulated reactions have rates slower or comparable to the cellular growth rate.

2 Background

When the average cellular size is small (e.g. a bacterium), it is commonly assumed that the cell acts as a well-mixed bioreactor. While there are specific gene regulatory processes where spatial compartmentalisation is the first-order consideration, there is no evidence to date that such spatial effects dominate in a generic sense in small cells. Small numbers of molecules naturally lead to large fluctuations of the reactant species in a typical set of biochemical reactions. In such cases, the most complete description of the stochastic system is given by the corresponding chemical master equation (see [20], for example).

Consider a well-stirred mixture of N chemical species $(X_1, ..X_N) = Y$ which may react through $\mu = 1..M$

Systems Biology online no. 20045016

doi: 10.1049/sb:20045016

Paper first received 31st March and in revised form 18th May 2004

T. Lu is with the Department of Physics, University of California, San Diego, California, USA

D. Volfson and L. Tsimring are with the Institute for Nonlinear Science, University of California, San Diego, California, USA

J. Hasty is with the Department of Bioengineering, University of California, San Diego, California, USA, email: hasty@ucsd.edu

D. Volfson is also with the Department of Bioengineering, University of California, San Diego, California, USA

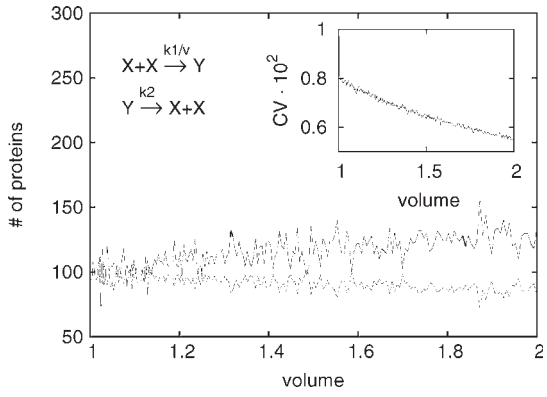

Fig. 1 *Gillespie simulation of a simple dimerisation reaction illustrating the importance of volume growth on the evolution of protein number and noise. The number of monomers (x, solid line) and dimers (y, dashed line), along with the coefficient of variation (CV, inset), are plotted versus the cellular volume (scaled by the initial volume). The forward bimolecular reaction rate scales as inverse volume, and when the reactions are fast, the system adiabatically follows the quasi-equilibrium relation* $(k1/v(t))x^2 = k2y$. *The mean and variance of x are both proportional to* $\sqrt{v}$, *so that the CV (variance/mean*^{2}*) scales as* $1/\sqrt{v}$ *(inset)*

elementary reaction channels R_μ within a volume V. Here, X_i is the number of molecules of type i in the system. For each reaction channel R_μ assume that there exists a scalar rate constant c_μ such that $c_\mu dt + o(dt)$ is the probability that a random combination of molecules from channel R_μ selected at the moment t will react in the interval $[t, t+dt)$. Biological examples of such reactions are protein-protein or protein-DNA binding, as well as the processes of transcription and translation. Let $h_\mu(Y)$ be the total number of possible distinct combinations of molecules for a channel R_μ when the system is in state Y, and $\alpha_\mu = (\alpha_{\mu 1}..\alpha_{\mu N})$ a constant stoichiometric vector prescribing the change in the state of the system after the reaction μ has occurred. The chemical master equation describing the evolution of the probability of obtaining the state Y at time t, given the initial state (Y_0, t_0), is

$$\frac{\partial}{\partial t}P(Y,t|Y_0,t_0) = \sum_{\mu}^{M}[c_\mu h_\mu(Y-\alpha_\mu) \times P(Y-\alpha_\mu,t|Y_0,t_0) - c_\mu h_\mu(Y)P(Y,t|Y_0,t_0)] \quad (1)$$

The specific form of the coefficients of the master equation (1) follows from the types of elementary reactions which occur in the system. For example, for second order reactions, $A + B \xrightarrow{c} AB$, the rate constant c should be inversely proportional to the volume of the system. This stems from the fact that the probability of a binary collision in a volume V is inversely proportional to V. When the volume changes deterministically it leads to the notion of time-dependent rate $c(t) = c'/V(t)$, where now c' is a 'true' constant, formally corresponding to the rate measured at a reference volume $V = 1$. For a third-order reaction, the time-dependent rate will have the form $c(t) = c'/V^2(t)$, and so on for higher orders.

The master equation framework becomes unmanageable even for a moderate number of possible states of the system. An alternative to solving for the probability of obtaining a given state is the stochastic simulation approach devised by Gillespie [16]. This approach entails the generation of an ensemble of sample trajectories of the system with statistics which asymptotically converges to the solution of the corresponding master equation. In the Gillespie algorithm, the state of the system is updated by determining (i) the time to the next reaction, and (ii) which reaction will occur next. The first question is addressed by considering the distribution of the time intervals τ until the next reaction

$$P(\tau) = a_\mu \exp\left[-\tau \sum_\mu a_\mu\right] \quad (2)$$

where $a_\mu = c_\mu h_\mu$ is called the propensity of the channel R_μ. The choice of the next reaction is made based upon the discrete distribution

$$P(\mu = \mu') = a_{\mu'} \Big/ \sum_\mu a_\mu \quad (3)$$

The aim of this paper is to formulate modifications to the Gillespie algorithm which are necessary to correctly simulate cellular systems with deterministic volume growth and division. While the processes of volume growth and division are likely to be stochastic in nature, we restrict our focus to deterministic growth and division in order to provide a framework for future modifications to the algorithm.

For ease of comparison with our modified algorithm, we provide here the original Gillespie recipe [16]:

(1) Input values for c_μ, $\mu = 1,\ldots,M$ and initial state $(x_1,\ldots,x_N)$, set $t = 0$.
(2) Compute propensities $a_\mu = h_\mu c_\mu$, $\mu = 1..M$.
(3) Generate uniform random numbers u_1, $u_2 \in [0,1)$.
(4) Compute the time interval τ until the next reaction according to distribution (2), viz, $\tau = -(\ln u_1)/\sum_\mu a_\mu$.
(5) Find the channel of the next reaction μ from distribution (3), viz, take μ to be the integer for which $\sum_{v=1}^{\mu-1} a_v < u_2 A \leq \sum_{v=1}^{\mu} a_v$, where a_0 is the total propensity, $A = \sum_{v=1}^{M} a_v$.
(6) Update time $t \to t+\tau$, and adjust Y in accordance with the particular reaction R_μ, viz, update $Y \to Y + \alpha_\mu$. Proceed to step 2.

This algorithm provides a systematic method for obtaining a sample of trajectories that are consistent with the underlying master equation. Crucially, working directly with the master equation is not typically feasible, and this method provides a practical alternative for generating the evolution of the probability distribution [Note 1].

When the chemical rates are known functions of time, then the simple and seemingly intuitive way to augment the Gillespie algorithm is to assume that the propensities in (2) and (3) are (known) functions of time [10, 21]. However, this approach leads to an approximate rather than an exact stochastic algorithm. In the next Section, we systematically derive an exact Gillespie-type algorithm which incorporates cellular growth and division.

3 Modified Gillespie algorithm for reactions in variable volume

We now consider the necessary modifications to the Gillespie algorithm to account for cellular growth and division. Suppose the volume $V(t)$ contains a mixture of X_i, $i = 1..N$ species which may interact through the reaction

Note 1: In this work, we consider the so-called Direct Gillespie algorithm. Another exact stochastic algorithm introduced by Gillespie [15] is the First Reaction method. In this method, putative times are computed *for every* reaction channel and the reaction with the smallest putative time is carried out. Since the probability distributions used to perform updates of chemical system are the same, the modifications described in this paper are also relevant for the First Reaction method.

channels R_μ, $\mu = 1..M$. Next assume that a subset of these channels is characterised by the time-dependent propensities $a_s(t) = a'_s/V(t)$, $s = 1..S$, while the propensities of the remaining channels do not depend on time; denote those a_q, $q = S+1..M$. We normalise time so that the volume of the cell doubles in a unit time interval, after which the cell divides. While the process of cell division is not deterministic, in order to illustrate our approach we will make the simplifying assumption that the cell division time is constant, and that at division all molecules are evenly divided among the daughter cells. Following Gillespie [16] we introduce the following probabilities:

- $P(\tau, \mu|Y,t)d\tau$ – probability that, given the state $Y = (X_1, \ldots, X_N)$ at time t, the next reaction will occur in the infinitesimal time interval $(t+\tau, t+\tau+d\tau)$, and it will be reaction R_μ
- $a_\mu(t)dt$ – probability that, given the state $Y = (X_1, \ldots, X_N)$ at time t, reaction R_μ will occur within the interval $(t, t+dt)$.

We compute $P(\tau, \mu|Y,t)d\tau$ as a product of the probabilities that no reaction will occur within $(t, t+\tau)$ times the probability that R_μ will occur within the subsequent interval $(t+\tau, t+\tau+d\tau)$:

$$P(\tau, \mu|Y,t)d\tau = P_0(\tau|Y,t)a_\mu(t+\tau)d\tau \tag{4}$$

To find $P_0(\tau|Y,t)$ we note that $\left[1 - d\tau \sum_\mu a_\mu(t+\tau)\right]$ is the probability that no reaction will occur during $d\tau$, hence

$$\begin{aligned} P_0(\tau+d\tau|Y,t) &= P_0(\tau|Y,t)\left[1 - d\tau \sum_\mu a_\mu(t+\tau)\right] \\ &= P_0(\tau|Y,t)\left[1 - d\tau \sum_s a_s(t+\tau) - d\tau \sum_q a_q\right] \end{aligned} \tag{5}$$

Using the initial condition $P_0(\tau = 0|Y,t) = 1$, we solve the differential equation (5) to find

$$P_0(\tau|Y,t) = \exp\left[-\sum_s \int_t^{t+\tau} d\tau' a_s(t+\tau')\right] \exp\left[-\tau \sum_q a_q\right] \tag{6}$$

Next we combine (4) and (6):

$$\begin{aligned} P(\tau, \mu|Y,t) = a_\mu(t+\tau) &\exp\left[-\sum_s \int_t^{t+\tau} d\tau' a_s(t+\tau')\right] \\ &\times \exp\left[-\tau \sum_q a_q\right] \end{aligned} \tag{7}$$

Now we specify the dependence of a_s on time explicitly, assuming that $V(t+\tau) = V(t)\exp[c\tau]$ with $c = \ln(2)$.

Performing the integration in (7) we find

$$P(\tau, \mu|Y,t) = \begin{cases} a_s(t)\exp[-c\tau]\exp[-f_c(\tau)A_s - \tau A_q] & s = 1..S \\ a_q \exp[-f_c(\tau)A_s - \tau A_q] & q = S+1..M \end{cases} \tag{8}$$

with

$$\begin{aligned} A_s &= \sum_s a_s(t) \\ A_q &= \sum_q a_q \\ f_c(\tau) &= (1 - \exp[-c\tau])/c \end{aligned} \tag{9}$$

It is easy to compute the probability of any reaction occuring between time t and $t+T$. Integrating (8) over time and summing over all channels, we find

$$\int_0^T \sum_\mu d\tau P(\tau, \mu|Y,t) = 1 - \exp[-A_s f_c(T) - TA_q] \tag{10}$$

The limit of this equation when $T \to \infty$ yields the probability that any reaction will occur after time t:

$$P_\infty = \begin{cases} 1, & A_q \neq 0 \\ 1 - \exp[-A_s/c], & A_q = 0 \end{cases} \tag{11}$$

This probability is 1 when at least one time-independent channel is present ($A_q \neq 0$). However if $A_q = 0$, this probability is less than one because there is a finite probability $\exp[-A_s/c]$ that no reaction will ever occur due to the exponentially decaying propensity of all of the reactions.

When all channels are time-independent, $S = 0$ (8) is reduced to the standard formula derived by Gillespie [16]

$$P(\tau, \mu|Y,t) = a_\mu \exp\left[-\tau \sum_\mu a_\mu\right] \tag{12}$$

where the summation over μ now includes all channels. We note that the same result is recovered in the formal limit $c \to 0$ and $\lim_{c\to 0} f_c(\tau) = \tau$, which corresponds to time-independent volume.

We should note that in fact the cell volume grows exponentially only until it doubles from its original value $V_0 = 1$, after which the cell divides and the volume is reset to V_0. So we may only use (7) for times $t + \tau < t_n$, the next cell division time (or equivalently, for $\{t\} + \tau < 1$, where $\{t\} = t \bmod 1$).

In order to implement the Gillespie algorithm we must determine (i) the time of the next reaction, and (ii) which reaction occurs.

Next reaction time. It is convenient to distinguish three possible scenarios, depending on the presence of time-dependent channels.

Case 1: $A_q \neq 0$, $A_s \neq 0$. Here we have both types of reactions and use (8). To find the time of the next reaction, $t+\tau$, we use the inversion method [22]. According to this method, in order to map the uniform random number (URN) $u_1 \in [0,1)$ to a number from a distribution with a given probability density function (PDF), one has to obtain a cumulative distribution from this PDF, draw a URN, set the cumulative distribution function equal to this number, and invert the equation. Using (8), and summing $P(\tau, \mu|Y,t)$ over all channels and integrating over time up to $\tau' = \tau$, we find the cumulative distribution function

$$\begin{aligned} F(\tau, |Y,t) &= \int_0^\tau d\tau' \sum_\mu P(\tau', \mu|Y,t) \\ &= 1 - \exp[-f_c(\tau)A_s - A_q\tau] \end{aligned} \tag{13}$$

The time interval until the next reaction is given by the solution of the transcendental equation ($1 - u_1$ is also a URN)

$$\exp[-f_c(\tau)A_s - A_q\tau] = u_1 \tag{14}$$

which may be expressed using a Lambert function [Note 2] $W(x)$ as

Note 2: 'The Lambert function $W(x)$, is a solution of the equation $W\exp[W] = x$ [23]'.

$$c\tau = W(\alpha \exp[\alpha + \beta]) - \alpha - \beta$$
$$\text{with} \quad \alpha = A_s/A_q, \quad \beta = c\ln(u_1)/A_q \qquad (15)$$

If the time of the next reaction $t+\tau$ with τ obtained from (15) is less than the time of the next cell division t_n, the time is advanced to $t+\tau$ and one of the reactions is selected (see below). However, the time of the next reaction calculated using τ from (15) may exceed the time of the next cell division. In this case we simply advance time to the next cell division time t_n, divide the volume and the numbers of proteins by two (more generally, this division may be non-even and chosen from a binomial distribution or partitioned in accordance with some empirically determined distribution), and select τ again.

Case 2: $A_q = 0$, $A_s \neq 0$. In this case only time-dependent channels are present and we use (8) with $a_q = 0, \quad q = S+1\,..\,M$. As a result we obtain

$$\tau = c^{-1} \ln\left[\frac{A_s}{A_s + c\ln(1-u_1)}\right] \qquad (16)$$

This case is special because there is a finite possibility that no reaction will happen in the interval $\tau = [0, \infty)$.

Formally, (16) has no solution τ for $u_1 > F_\infty = 1 - \exp[-A_s/c]$ (see Fig. 2). In this case, as in the case when the solution τ exists but $\{t+\tau\} > 1$, no reaction is implemented but time is advanced to the time of the next cell division (t_n), and the volume and number of proteins are reset.

Case 3: $A_q \neq 0$, $A_s = 0$. This is the standard situation covered by the Gillespie algorithm, and the time to the next reaction is given by

$$\tau = -\frac{1}{A_q}\ln u_1 \qquad (17)$$

Which reaction to choose. This step is similar to the standard Gillespie algorithm. The only difference comes from the fact that in the interval $[t, t+\tau]$, the time-dependent propensities change due to cell growth, and thus one has to choose which reaction will occur based on the propensities at time $t+\tau$. Using a second URN u_2 we find the channel $\nu = \mu$:

$$\sum_{\nu=1}^{\mu-1} a_\mu(t+\tau) < u_2(A_q + A_s(t+\tau)) \leq \sum_{\nu=1}^{\mu} a_\mu(t+\tau), \quad 1 \leq \nu \leq M \qquad (18)$$

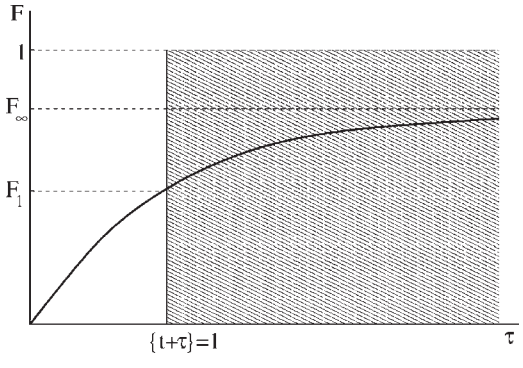

Fig. 2 *Cumulative distribution function of times until the next reaction for case 2 when no time-independent channels are present. If URN u_1 is greater than F_1 (given by (13) with $\tau = 1 - \{t\}$), instead of selecting the next reaction, the cell division at time $t - \{t\} + 1$ is performed*

To summarise, the modified Direct Gillespie algorithm contains the following steps:

(1) Input values for c_μ, $\mu = 1, \ldots, M$ and initial state $(x_1, \ldots, x_N)$, set $t = 0$.
(2) Compute $a_\mu = h_\mu c_\mu$, along with $A_s = \sum_s a_s(t)$, $A_q = \sum_q a_q$, $s = 1\,..\,S$, $q = S+1\,..\,M$.
(3) Generate uniform random numbers u_1, u_2.
(4) Check if A_q is zero; if it is then use (16) to compute the time interval τ until the next reaction, otherwise compute τ according to (15).
(5) Check whether $\{t\} + \tau < 1$. If yes, go to the next step. If no, update time $t \to [t]+1$, reset volume $V \to V_0$ and the number of proteins of each species in the cell $x_i \to x_i/2$, and return to step 2.
(6) Find the channel of the next reaction μ using (18).
(7) Update time $t \to t+\tau$, and adjust x_i in accordance with the particular reaction R_μ. Proceed to step 2.

We now discuss when our exact time-dependent Gillespie (ETG) algorithm is expected to yield results that differ from an 'adiabatic' generalisation of the standard Gillespie routine, which we will denote the adiabatic time-dependent Gillespie (ATG) approach. This adiabatic approach arises from using the standard Gillespie algorithm with time-dependent rate constants that are updated after each reaction event using the current (at the moment of the reaction) value of the cellular volume. As we show below, this approach is quite accurate when the volume changes slowly as compared with the timescale of the fastest chemical reaction. As an illustration, consider the simple case with a single time-dependent channel, and compare the distributions for the time interval until the next reaction

$$P_1(\tau|t) = a(t)\exp[-f_c(\tau)a(t) - c\tau], \qquad \text{(ETG)} \qquad (19)$$

$$P_0(\tau|t) = a(t)\exp[-a(t)\tau], \qquad \text{(ATG)} \qquad (20)$$

In both cases the distribution functions decay exponentially at a rate which is a linear function of $a(t)$. When $a(t) \gg 1$, in both cases the mean value of τ is much smaller than one. Additionally, $c\tau$ may be neglected as compared with the much larger term $f_c(\tau)a(t) = a(t)(\tau + \mathcal{O}(\tau^2)) \approx a(t)\tau$. Therefore, for large a and small τ, the distributions (19), (20) are nearly the same, and the two algorithms should yield similar results. On the other hand, when $a(t)$ is small, the probability that a reaction will occur during the cell doubling time approaches $a/(2\ln 2)$ according to the ETG algorithm, whereas it is a according to the ATG algorithm. Thus, the difference between these two algorithms for small a can be significant. Since in the context of gene regulation, protein multimerisation and DNA binding are many orders of magnitude faster than the timescale of cellular growth, the ATG algorithm is reliable when all of the reactions are simulated with a Gillespie scheme. However, in realistic settings it may be necessary to utilise hybrid techniques [17, 24], whereby the fast reactions are simulated with deterministic or Langevin equations, and the slow reactions simulated with Gillespie. In these situations, if the average time between random 'Gillespie events' is comparable with the cell division time, then a generalisation of the Gillespie algorithm is essential.

4 Example: transcriptional regulation without feedback

In order to illustrate the difference between our algorithm and the adiabatic approach, we now turn to a simple yet non-trivial example where numerical results can be

compared with analytical expressions obtained from the master equation. The example involves a single gene which fluctuates between two states S_0 and S_1. The transition $S_0 \to S_1$ occurs when a regulator protein is bound to the gene's promoter, and so this transition probability is inversely proportional to the cell volume V. Upon division of the cell, both its volume and the number of all proteins are halved. For simplicity, we assume that the number of regulator proteins quickly reaches a steady state, which is maintained throughout the cell cycle. We neglect the fast relaxation of the number of regulator proteins, and assume that an effective propensity for the transition $S_0 \to S_1$ is inversely proportional to the volume at any instance of time. The propensity of the reverse process $S_1 \to S_0$ is assumed independent of the cell volume. The transcription of gene S leads to the production of the protein X at a rate α_0 when its promoter is in state S_0, and α_1 when it is in state S_1 (for definiteness, we assume $\alpha_0 < \alpha_1$). The biochemical reactions describing this model system are summarised in Table 1.

Table 1: Biochemical reactions for a simple system describing a constitutive promoter

μ	Reaction	a_μ
1	$S_0 \xrightarrow{k_1/v} S_1$	$k_1/v\, s_0$
2	$S_1 \xrightarrow{k_{-1}} S_0$	$k_{-1}\, s_1$
3	$S_0 \xrightarrow{\alpha_0} S_0 + X$	$\alpha_0\, s_0$
4	$S_1 \xrightarrow{\alpha_1} S_1 + X$	$\alpha_1\, s_1$
5	$X \xrightarrow{k_x}$	$k_x\, x$

We assume that the volume of the cell grows exponentially during the cell cycle, $V(t) = \exp[\ln 2\ t]$ (without loss of generality we assume $V(0) = 1$). At time $t = n$, the volume V is halved, $V \to V/2$, and the number of proteins is halved, $X \to X/2$. In the case of constant volume, this single-gene constitutive model was explored by Kepler and Elston [25] using a master equation for the time-dependent probability p_x^s of having both the promoter in the state $s = [0, 1]$ and x proteins. Using a similar approach for growing and dividing cells, we obtain equations for the dynamics of the partial moments $M_q^s \equiv \langle x^q \rangle_s = \sum_x x^q p_x^s$ (see Appendix,

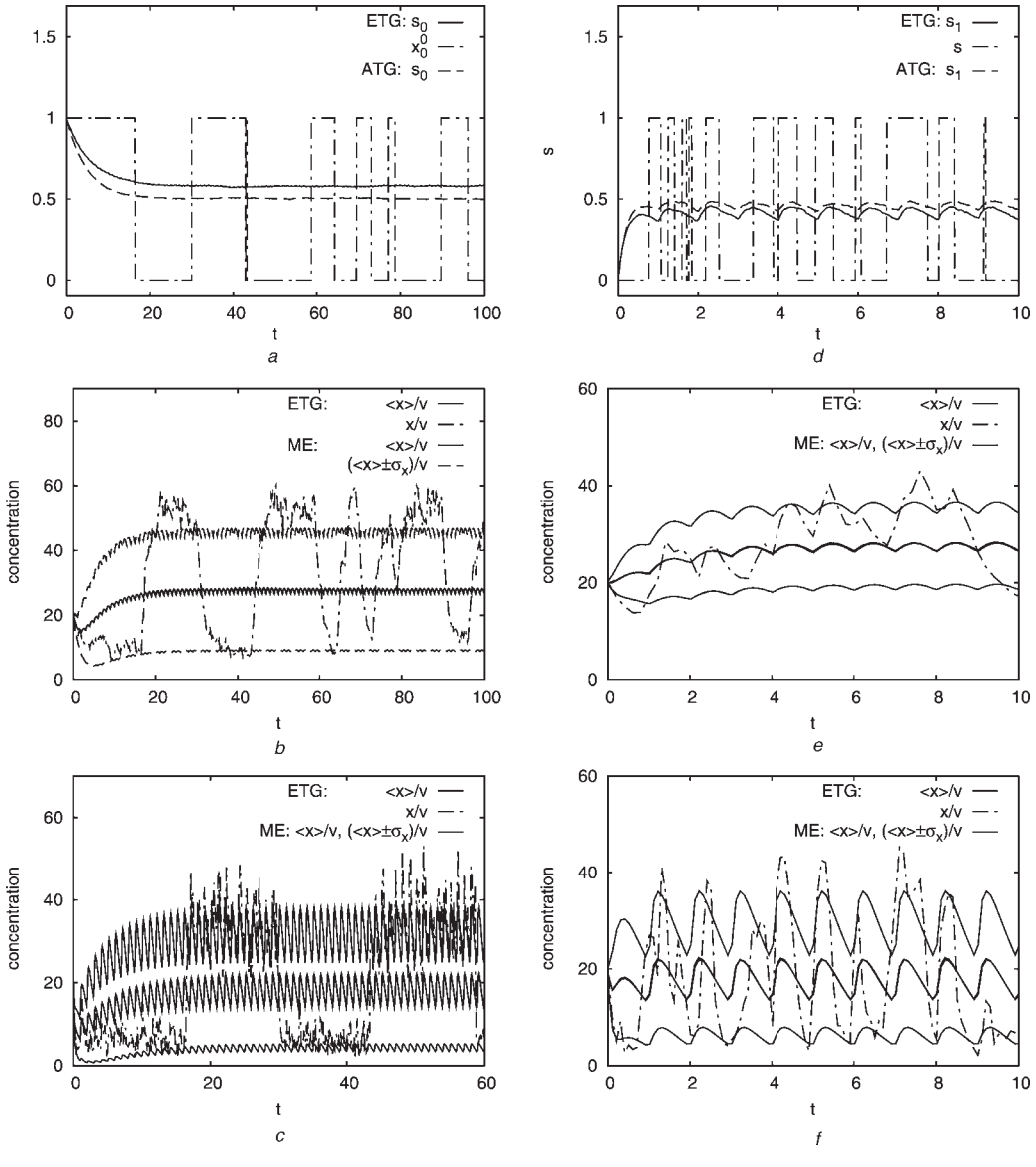

Fig. 3 *Time series of the probability of state S_1 of the promoter, s_1, (a, d) and the concentration of proteins $\langle x \rangle / v$ (b, c, e, f) obtained with the ATG and ETG algorithms, and theoretically using the master equation approach for different parameter values. Left column: $k_1 = k_{-1} = 0.1$, right column: $k_1 = k_{-1} = 10$, second row: $k_x = 0.01$, $\alpha_0 = 10$, $\alpha_1 = 50$, third row: $k_x = 10$, $\alpha_0 = 100$, $\alpha_1 = 500$. Dashed lines in panels b, c, e, and f show the $(\langle x \rangle \pm \sigma_x)/v$ range of concentration fluctuations. Dash-dotted lines show a single typical trajectory of the stochastic system. The ETG results virtually coincide with the theoretical curves*

Section 8). The *zero*th moments M_0^s represent the probabilities for the promoter to be in the sth state. The sum of the first moments $M_1 = M_1^0 + M_1^1$ is the average number of proteins $\langle x \rangle$, and $Var = M_2^0 + M_2^1 - (M_1^0 + M_1^1)^2$ is the variance of the number of proteins. In the case of constant volume, these moments reach a steady state [25], but when the cell grows and divides, the moments asymptotically approach an oscillatory regime as one might expect (Fig. 3). The time-averaged probability of finding the promoter in an unbound (bound) state M_0^0 (M_0^1) can be accurately approximated by the formulas $M_0^0 = k_{-1}/(k_{-1} + k_1/2 \ln 2)$ and $M_0^1 = 1 - M_0^0$.

Let us now describe the results of the simulations of the reactions listed in Table 1. In this simple example, the fast reactions 3–5 are computationally expensive in the Gillespie algorithm, since the average time step between these reactions is very small compared with the cell division time. In order to make progress on realistic problems involving many such fast reactions, a more natural approach is a hybrid simulation technique [17, 24], where the dynamics of the fast subset are modelled either deterministically or using Langevin equations, while the slow reactions are modelled with the Gillespie technique. We adopt this approach and simulate the dynamics of the proteins X using a Langevin equation

$$\dot{x} = \alpha_0(1-s) + \alpha_1 s - k_x x + \sqrt{\alpha_0(1-s) + \alpha_1 s + k_x x}, w(t) \tag{21}$$

where $w(t)$ is a Gaussian white noise process. This equation was integrated using the Euler-Murayama method [26].

The differences between the use of the standard and modified Gillespie algorithms for the slow reaction are evident in the distributions of residence times, $p_0(t_r)$, and of phases of transitions from S_0 to S_1 within the cell-doubling interval, $p(\theta)$ (here $\theta = \{t\}$). Since the propensity of the inverse reaction is independent of volume, the residence time distribution for the bound state is simply described by Poisson statistics. For small k_1, step 2 of the algorithm typically yields a negative result since the randomly selected time $t + \tau$ exceeds the next cell division time. In this case, time is advanced to the next cell division time, and another draw is performed. Therefore, most of the transitions from S_0 to S_1 will occur after a draw at cell division (after V is reset to one). Then it is easy to see from [19, 20] that the phase distributions for the two versions of Gillespie algorithm are

$$p_0(\theta) \approx 1 \quad \text{(ATG)}, \quad p_1(\theta) \approx 2^{1-\theta} \ln 2 \quad \text{(ETG)} \tag{22}$$

Accordingly, the distribution of residence times at small k_1 is mostly determined by the number of unit time intervals during which the transition $S_0 \to S_1$ does *not* occur. The probability of the transition to not occur per unit time is $\exp(-k_1)$ for the ATG algorithm and $\exp(-k_1/2 \ln 2)$ for the ETG algorithm, and thus for long times, the same distribution scales as

$$p_\theta(t_r) \propto e^{-k_1 t_r} \quad \text{(ATG)}, \quad p_1(t_r) \propto e^{(-k_1/2 \ln 2) t_r} \quad \text{(ETG)} \tag{23}$$

In Fig. 3 we compare the time dependencies of ensemble-averaged probability of the promoter in state S_1 (a, d), and protein concentration $\langle x \rangle / v$ (b, c, e, f) obtained numerically using exact and adiabatic Gillespie algorithms with theoretical curves generated using the master equation, see Appendix. In the same plots we depict single realisations of s and x. As seen for the plots, the exact Gillespie algorithm yields results that are virtually indistinguishable from the theoretical curves. For comparison, Figs. 3*a* and 3*d* also show the mean values of s_1 as a function of time obtained using the adiabatic Gillespie algorithm based on instantaneous volume. In this case, visible deviations from the theoretical dependencies are obtained.

Figure 4 shows the time-averaged characteristics $\overline{s_0}$, $\overline{\langle x \rangle}$, $\overline{\sigma_x} \equiv (\overline{Var})^{1/2}$ as functions of the forward propensity k_1 for $\alpha_0 = 100$, $\alpha_1 = 500$, $k_x = 10$, $k_{-1} = 1$. As can be seen, the ETG simulations are in excellent agreement with master equation analysis, whereas the ATG simulations show systematic deviations.

Figure 5*a* illustrates the distributions of phases θ of 'forward' reaction $S_0 \to S_1$ within a single cell cycle and Fig. 5*b* shows the distribution of residence times t_r in S_0 state. These distributions obtained numerically with ATG and ETG algorithms are in good agreement with theoretical predictions [22, 24] for small k_1.

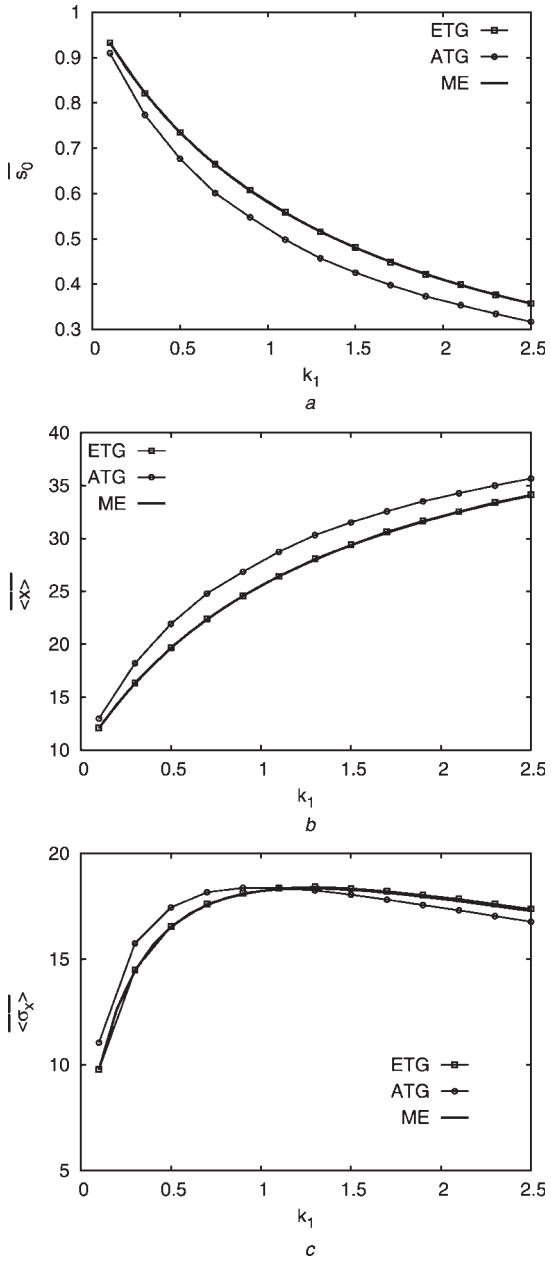

Fig. 4 *Time-averaged values of s_0, $\langle x \rangle$, and standard deviation σ_x as a function of 'forward' rate k_1 for $\alpha_0 = 100$, $\alpha_1 = 500$, $k_x = 10$, $k_{-1} = 1$*

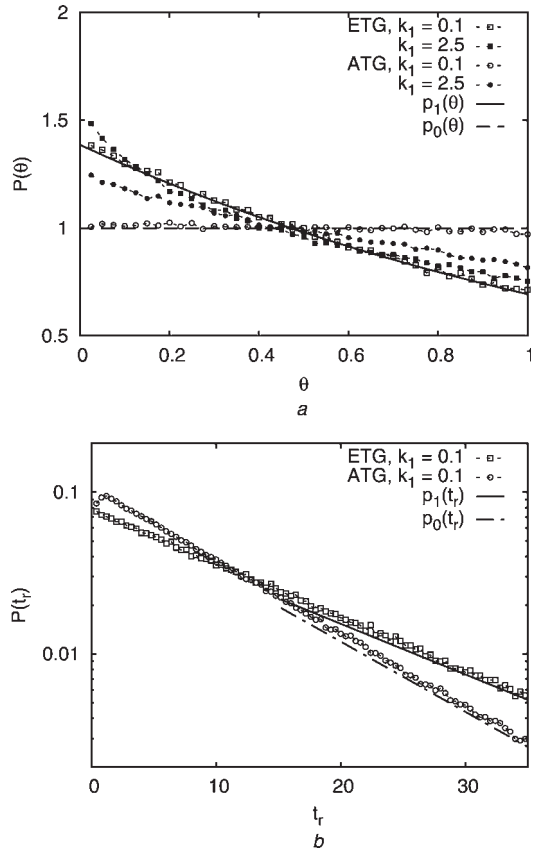

Fig. 5

a Distribution of phases θ of 'forward' reactions $S_0 \rightarrow S_1$ within a single cell cycle for $k_1 = k_{-1} = 0.1$; 2.5

b Distribution of residence times t_r in S_0 state for $k_1 = k_{-1} = 0.1$. Symbols show ATG and ETG simulations, and lines are plotted using theoretical formulas (22) and (23)

5 Conclusions

In this work, we have derived a generalisation of the Gillespie algorithm to account for cellular growth and division, and compared this algorithm with an adiabatic Gillespie routine where the volume is simply updated as time progresses. While the adiabatic routine was shown to be quite accurate if any of the chemical reactions are fast compared with the growth rate, it is typically not feasible in realistic settings to simulate all reactions with a pure Gillespie-type routine. We therefore focused our comparison on a hybrid simulation technique [17, 24], whereby the fast reactions were simulated with Langevin equations, and the slow reactions simulated with Gillespie method. In these simulations, the average time between random events may be significant as compared with the cell division time, and here we were able to demonstrate the necessity of using our generalisation of the Gillespie algorithm. Importantly, the generalised algorithm does not significantly increase the computational expense, so this derived method is preferred regardless of accuracy considerations.

Our method specifies how a periodic deterministic event, namely cell division, can be incorporated into the Gillespie routine. While in order to illustrate our approach, we considered deterministic volume growth and division, future work could focus on additional sources of noise, such as a stochastic growth rate, fluctuations arising from unequal partitioning of molecules at cell division, or variations in the DNA replication time before division. As experiments begin to discriminate between the sources of noise in the cellular environment [3], simulation routines that correctly incorporate individual noise sources will be increasingly useful.

6 Acknowledgments

We thank Peter Wolynes for stimulating discussions and suggestions. This work was supported by the NSF Division of Molecular and Cellular Bioscience (LT, DV, JH), the Alfred P. Sloan Foundation (JH), the NSF Division of Bioengineering (JH), and the La Jolla Interfaces in Science Predoctoral Fellowship (TL).

7 References

1 Becskei, A., and Serrano, L.: 'Engineering stability in gene networks by autoregulation', *Nature*, 2000, **405**, p. 590
2 Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and van Oudenaarden, A.: 'Regulation of noise in the expression of a single gene', *Nat. Gen.*, 2002, **31**, p. 69
3 Elowitz, M., Levine, A., Siggia, E., and Swain, P.: 'Stochastic gene expression in a single cell', *Science*, 2002, **297**, pp. 1183–1186
4 Isaacs, F.J., Hasty, J., Cantor, C.R., and Collins, J.J.: 'Prediction and measurement of an autoregulatory genetic module', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, pp. 7714–7719
5 Blake, W.J., Kaern, M., Cantor, C.R., and Collins, J.J.: 'Noise in eukaryotic gene expression', *Nature*, 2003, **422**, pp. 633–637
6 Gardner, T.S., Cantor, C.R., and Collins, J.J.: 'Construction of a genetic toggle switch in *Escherichia coli*', *Nature*, 2000, **403**, pp. 339–342
7 Elowitz, M.B., and Leibler, S.: 'A synthetic oscillatory network of transcriptional regulators', *Nature*, 2000, **403**, p. 335
8 Becskei, A., Séraphin, B., and Serrano, L.: 'Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion', *EMBO J.*, 2001, **20**, p. 2528
9 Stuart, R.N., and Branscomb, E.W.: 'Quantitative theory of *in vivo lac* regulation: Significance of repressor packaging', *J. Theor. Biol.*, 1971, **31**, pp. 313–329
10 Arkin, A., Ross, J., and McAdams, H.H.: 'Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells', *Genetics*, 1998, **149**, p. 1633
11 Hasty, J., Pradines, J., Dolnik, M., and Collins, J.J.: 'Noise-based switches and amplifiers for gene expression', *Proc. Natl. Acad. Sci. USA*, 2000, **97**, p. 2075
12 Swain, P., Elowitz, M., and Siggia, E.: 'Intrinsic and extrinsic contributions to stochasticity in gene expression', *Proc. Natl. Acad. Sci. USA*, 2002, **99**, pp. 12795–12800
13 Rao, C., Wolf, D.M., and Arkin, A.P.: 'Control, exploitation and tolerance of intracellular noise', *Nature*, 2002, **420**, pp. 231–237
14 Simpson, M.L., Cox, C.D., and Sayler, G.S.: 'Frequency domain analysis of noise in autoregulated gene circuits', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, (8), pp. 4551–4556, (Available online http://www.pnas.org/cgi/content/abstract/100/8/4551)
15 Gillespie, D.: 'General method for numerically simulating stochastic time evolution of coupled chemical-reactions', *J. Comput. Phys.*, 1976, **22**, pp. 403–434
16 Gillespie, D.T.: 'Exact stochastic simulation of coupled chemical reactions', *J. Phys. Chem.*, 1977, **81**, (25), pp. 2340–2361
17 Haseltine, E.J., and Rawlings, J.B.: 'Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics', *J. Chem. Phys.*, 2002, **117**, (15), pp. 6959–6969
18 Rao, C.V., and Arkin, A.P.: 'Stochastic chemical kinetics and the quasi steady-state assumption: Application to the Gillespie algorithm', *J. Chem. Phys.*, 2003, **118**, pp. 4999–5010
19 Gibson, M.A., and Bruck, J.: 'Efficient exact stochastic simulation of chemical systems with many species and many channels', *J. Phys. Chem.*, 2000, **104**, pp. 1876–1889
20 Kampen, N.G.V.: Stochastic Processes in Physics and Chemistry, (North-Holland, Amsterdam, 1992)
21 Kierzek, A.: 'Stocks: Stochastic kinetic simulations of biochemical systems with Gillespie algorithm', *Bioinformatics*, 2002, **18**, (3), pp. 470–481
22 Gillespie, D.T.: 'Markov Processes: An Introduction for Physical Scientists' (Academic Press, 1992)
23 Corless, R.M., *et al.*: 'On The Lambert W Function', *Adv. Comp. Math.*, 1996, **5**, pp. 329–359
24 Adalsteinsson, D., McMillen, D., and Elston, T.: 'Biochemical network stochastic simulator (bionets): Software for stochastic modeling of biochemical networks', *BMC Bioinformatics*, 2004, **5**, p. 24
25 Kepler, T., and Elston, T.: 'Stochasticity in transcriptional regulation: origins, consequences and mathematical representations', *Biophys, J.*, 2001, **81**, p. 3116
26 Gard, T.C.: 'Introduction to Stochastic Differential Equations' (Marcel Dekker, New York, 1988)

8 Appendix: moment equations for the single gene – no feedback system

Similar to [25], we introduce the time-dependent probability p_x^s to have the promoter in state $s = [0, 1]$ and x proteins. The evolution of this probability between cell division times is described by the two master equations

$$\dot{p}_x^0 = \alpha_0\left(p_{x-1}^0 - p_x^0\right) + k_x\left[(x+1)p_{x+1}^0 - xp_x^0\right] + k_{-1}p_x^1 - k_1V^{-1}p_x^0, \tag{24}$$

$$\dot{p}_x^1 = \alpha_1\left(p_{x-1}^1 - p_x^1\right) + k_x\left[(x+1)p_{x+1}^1 - xp_x^1\right] + k_1V^{-1}p_x^0 - k_{-1}p_x^1, \tag{25}$$

At cell division time t_n, the volume V is halved, and also the number of the proteins in the cell, so $p_x(t_n+) = p_{2x}(t_n-)$.

From (25) and (26) we can derive the equations for the partial moments of the distribution of the number of proteins, defined as

$$\langle x^q\rangle_{0,1} \equiv \sum_x x^q p_x^s \tag{26}$$

The *zero*th moments $s_{0,1} = \langle x^0\rangle_{0,1}$ give the marginal probabilities of the promoter to be in states 0,1, respectively. The equations for $s_{0,1}$ read

$$\dot{s}_0 = -\frac{k_1}{V(t)}s_0 + k_{-1}s_1 \tag{27}$$

$$\dot{s}_1 = \frac{k_1}{V(t)}s_0 - k_{-1}s_1 \tag{28}$$

The equations for the first moments read

$$\langle\dot{x}\rangle_0 = \alpha_0 s_0 - k_x\langle x\rangle_0 - \frac{k_1}{V(t)}\langle x\rangle_0 + k_{-1}\langle x\rangle_1 \tag{29}$$

$$\langle\dot{x}\rangle_1 = \alpha_1 s_1 - k_x\langle x\rangle_1 + \frac{k_1}{V(t)}\langle x\rangle_0 - k_{-1}\langle x\rangle_1 \tag{30}$$

and for the second moments

$$\langle\dot{x}^2\rangle_0 = \alpha_0 s_0 + 2\alpha_0\langle x\rangle_0 + k_x(\langle x\rangle_0 - 2\langle x^2\rangle_0) - \frac{k_1}{V(t)}\langle x^2\rangle_0 + k_{-1}\langle x^2\rangle_1 \tag{31}$$

$$\langle\dot{x}^2\rangle_1 = \alpha_1 s_1 + 2\alpha_1\langle x\rangle_1 + k_x(\langle x\rangle_1 - 2\langle x^2\rangle_1) + \frac{k_1}{V(t)}\langle x^2\rangle_0 - k_{-1}\langle x^2\rangle_1 \tag{32}$$

Here, at cell division times, the values of $\langle x\rangle_{0,1}$ and $\langle x^2\rangle_{0,1}$ have to be reset, $\langle x\rangle_{0,1}(t_n+) = \langle x\rangle_{0,1}(t_n-)/2$ and $\langle x^2\rangle_{0,1}(t_n+) = \langle x^2\rangle_{0,1}(t_n-)/4$.

The asymptotic solution for s_0 at large time t can be written in the form

$$s_0(t) = k_{-1}\int_{-\infty}^{t} e^{-\int_{t'}^{t}\frac{k_1}{V(y)}+k_{-1}dy}dt' \tag{33}$$

The mean values of $s_{0,1}$ can be approximated with good accuracy by the formulas

$$\overline{s_0(t)} = \frac{k_{-1}}{k_{-1} + k_1\overline{V^{-1}}} \tag{34}$$

$$\overline{s_1(t)} = \frac{k_1\overline{V-1}}{k_{-1} + k_1\overline{V^{-1}}} \tag{35}$$

For small decay rates $k_x \ll 1$, the mean value of the number of proteins $\overline{\langle x\rangle} = \overline{\langle x\rangle_0} + \overline{\langle x\rangle_1}$ can be found from (29) assuming that the number of proteins doubles during the cell division time. For small decay rates, this simply leads to

$$\overline{\langle x\rangle} = \frac{3}{2}\frac{k_{-1}\alpha_1 + k_1\overline{V(t)^{-1}}\alpha_0}{k_{-1} + k_1\overline{V(t)^{-1}}} \tag{36}$$

Similarly, we can obtain the mean variance of the number of proteins at large t, $\overline{\langle x^2\rangle - \langle x\rangle^2}$ (it has to increase four times between consecutive cell divisions).