

Dynamic Partitioning of Large Discrete Event Biological Systems for Hybrid Simulation and Analysis

Kaustubh R. Joshi¹, Natasha Neogi¹, and William H. Sanders¹

Coordinated Science Laboratory,
University of Illinois at Urbana-Champaign,
Urbana IL 61801

Abstract. Biological systems involving genetic reactions are large discrete event systems, and often contain certain species that occur in small quantities, and others that occur in large quantities, leading to a difficulty in modelling and simulation. Small populations inhibit the usefulness of utilizing differential equations to represent the system, while the large populations cause stochastic discrete event simulation to become computationally intensive. This paper presents an algorithmic approach for the dynamic partitioning and stochastic hybrid simulation of biological systems. The algorithm uses a Poisson approximation for discrete event generation and ordinary differential equations to model continuous behavior. The populations are dynamically partitioned so that some populations are simulated in a discrete stochastic fashion, while others are simulated by continuous differential equations, and this partition between discrete and continuous behavior is updated. The hybrid model of a biological toggle switch yields promising results.

1 Introduction

The advancement of experimental techniques and rapid accumulation of genetic information have opened a new frontier in biomedical engineering. The ability to engineer artificial gene regulatory networks with sophisticated computational and functional abilities grows ever closer due to the availability of well-characterized components from natural gene networks. Hence, the construction, analysis, and interpretation of qualitative and quantitative models acts to drive the field forward [10]. There are several complementary approaches involving discrete and stochastic simulation that can be used to model gene networks; they are surveyed in this paper. The goal is to use such descriptions to accurately predict the properties and function of modules connected into networks, and to make *in silico* suggestions for optimal design strategies prior to implementation *in vivo*. However, as the models become progressively more complex, the algorithms become computationally expensive [11].

For systems in which species fluctuate by varying orders of magnitude, the largest fluctuating species require the most time to simulate stochastically because exact stochastic simulation techniques scale with the number of reaction

events [9]. The use of stochastic partitioning, whereby the state of the reaction is modelled using extents of reactions rather than molecules of species, thereby leading to the ability to separate the state into subsets of fast and slow reactions, allows for the reduction in the computational burden for simulation. The fast reactions can then be approximated deterministically as ordinary differential equations (or stochastically by the Langevin Equations), while the slow reactions can be treated as stochastic events with time-varying reaction rates. However, the partition must be evaluated periodically, as reactions can change their behavior as time progresses.

In the next section, an informal discussion of the hybrid nature of large, discrete event biological systems is conducted. A survey of simulation and analysis techniques to date is performed, and the potential limitations of these methods are discussed. Section 3 focuses on the various modeling techniques used to encapsulate biological systems. Deterministic differential equations, the Langevin approximation, and discrete stochastic simulation models are developed. A Poisson process is used to define the discrete reaction rate for continuous mass action models in Section 4. Section 5 discusses several strategies to partition reactions in subsets that can be modeled continuously or discretely. A method for sampling from a non-homogeneous Poisson process in order to determine the next reaction and time to next reaction is outlined in Section 6. Section 7 contains the discussion of a model of a toggle switch that is simulated using dynamic partitioning of a hybrid model, where the population partitioning is used to separate fast and slow reactions. The fast reactions are simulated using ordinary continuous differential equations, and the slow reactions are simulated using a non-homogeneous Poisson process. The simulation algorithm is outlined, and implementation issues are discussed.

2 Problems with Large Discrete Event Biological Systems

In many systems, the probability of occurrence of any event in an infinitesimal time interval dt will depend only on the number of individuals at time t and on parameters that might depend on t , that is, the future evolution of the system depends only on its present state, and not on the system's history. Such processes can be represented as Markov processes [8]. These systems are memoryless, and do not need to keep a state history to determine their evolution.

Systems that are Markov in nature and have exponential event distributions can be modeled using Monte Carlo simulation. The notion of Monte Carlo simulation comes from the solution of the inverse problem, in which differential equations are approximated by stochastic jumps obtained by the Monte Carlo method. This simulation is a simple realization of the Markov process, and is often called the *Feller process* [3].

If the population size is large, then in the limiting case, the fractions of the total population represented by each species are the relevant quantities to be considered. This notion is called the *Mass-Action Law*, and is commonly accepted in physical chemistry [2].

2.1 Current Simulation and Analysis Techniques

Solutions to the stochastic formulation of coupled chemical reactions can be computed using the Monte Carlo procedure specified by Gillespie [7]. The algorithm calculates a stochastic description of the temporal behavior of the coupled reactions. This description can be shown to have a more rigorous physical basis than conventional kinetic equation formulations based on the assumption that changes in the chemical reaction system over time are both continuous and deterministic. This assumption of continuous and deterministic behavior is invalid for low concentrations of reactant species or sufficiently slow reactant rates. The Direct and First Reaction methods [7], outlined by Gillespie, are simulation algorithms that calculate the probabilistic outcome of each discrete chemical event and the resulting changes in the number of each molecular species. It should be noted that, in the limiting case of a large number of reactant molecules, these methods are entirely equivalent to the solution of the traditional kinetic differential equations derived from the *Law of Mass Action* [8].

As models become progressively more complex, however, these algorithms often become expensive computationally. Several techniques have been employed to reduce the computational burden. Employment of a deterministic equilibrium assumption on polymerization reaction kinetics has yielded a decrease in computational complexity, as shown by He, Zhang, Chen, and Yang [?]. Gibson and Bruck have refined Gillespie’s First Reaction algorithm to reduce the required number of pseudo-random numbers that must be generated, a technique that works best for systems in which some reactions occur much more frequently than others [5]. Rao and Arkin have demonstrated how to numerically simulate systems reduced by the quasi-steady-state assumption [?]. This work expands upon ideas by Janssen [?] and Vlad and Pop [?], who examined the adiabatic elimination of fast relaxing variables in stochastic chemical kinetics. Resat, Wiley, and Dixon applied a probability-weighted Monte Carlo approach to address systems with reaction rates that vary by several orders of magnitude, but this method increases error in species fluctuations [?]. Haseltine and Rawlings [9] expanded upon the idea of a partitioned system and simulation via Gillespie’s direct method to construct approximations that reduce the computational burden for simulation. Their work sets up the framework for partitioning a system into discrete and continuous reactions, but addresses only static partitioning.

Recently, Solari et al. developed an approximation to stochastic population dynamics based on almost independent Poisson processes whose parameters obey a set of coupled ordinary differential equations [14, 15, 1]. Error bounds for the moment-generating function have been developed. For large populations, the Poisson approximation becomes a discrete integration of the Langevin approximation [12]. A implementation of this simulation method has been outlined, and several improvements to improve its efficiency are discussed in the relevant literature [14].

Unfortunately, for large-scale biological systems that have complex macromolecules (i.e., proteins), long and short time scales, and interactions of several

different molecule types, no one approach has yet to yield satisfactory results in all the arenas of accuracy, computational complexity, scalability, and coverage.

3 Modeling

A coupled system of p biochemical reactions can be modeled by equations of the form:



where one molecule each of n different reactants react to form one molecule each of m different products P_i , $k = 1 \dots p$. If it is assumed that enough molecules of each species are present such that the number of molecules can be approximated as a continuous quantity that varies deterministically over time, the concentration of each species can be written in terms of the concentrations $[R_i]$, $[P_i]$ of all other species via a set of differential equations. However, when the concentrations of certain species drop below a given level, the assumption that the number of molecules can be approximated as a continuously varying deterministic function is no longer valid.

Alternatively, it can be assumed that non-reactive collisions occur far more often than reactive collisions, and thus that the fast dynamics of motion can be neglected [5]. Thus, the system may be represented best using the number of each kind of molecule. Using this approach, it can be concluded that the probability that a certain reaction μ will take place in the next instant of time dt is given by $a_\mu dt + o(dt)$, where a_μ is independent of dt and $o(dt)$ represents terms that are negligible for small dt . However, a_μ may depend on the current number of molecules of each kind, and the current time. Furthermore, it may depend on quantities such as temperature and volume, which may change with time. The state of the system in the stochastic framework is defined by the number of molecules of each species and changes discretely whenever one of the reactions occurs. Formally, we can define the probability that the state S of a system changes to S' via the occurrence of the reaction as [5]:

$$P(S', t + dt | S, t) = a_\mu dt + o(dt) \quad (2)$$

Note that since a_μ , and thus the transition probability, are dependent only on the current state and not on previous states, the underlying process is Markov in nature. If an individual probability variable is created for each possible state of the system, then Equation 2 can be used to write out the system of coupled differential equations that defines the system, such as:

$$\frac{d}{dt} P(S_j, t + dt | S_i, t) = f(a_{\mu_{ij}}, P(S_i, t)) \quad (3)$$

for all reactions μ_{ij} that bring the system state from S_i to S_j . This set of differential equations has probabilities as variables and is referred to as a *Master Equation*. It is a direct consequence of the satisfaction of the forward Kolmogorov

equation by the probabilities [6]. For systems with very few states, the entire set of Master Equations may be written out and explicitly solved. For large systems, however, this approach quickly becomes intractable [5].

The Master Equation, can, however, be approximated by the stochastic Langevin equations, which are differential equations, and therefore scale more easily with respect to population size. If the characteristic size of the system is defined by Ω , the Master Equation is recast in terms of intensive variables (concentrations), and a Kramers-Moyal expansion is performed, the Master equation results in a system size expansion in Ω . In the limiting case as Ω becomes large, the discrete Master Equation can be approximated by its first two differential moments with the continuous Fokker-Planck equation, and has an Ito solution of the form [5]:

$$d[\mu_i] = b_i(\mu)dt + \sqrt{b_i(\mu)}dW_i \quad (4)$$

where:

$$b_i(\mu) = \sum_{i=0}^{all\mu} a_i P(S, t) \quad (5)$$

and dW_i is a Weiner process. Equation (5) is a stochastic differential equation that specifies the evolution of the trajectories of the system state [7]. The error induced by this approximation is directly related to the size of the system Ω . Even if the system size Ω is large, the Langevin approximation will most likely be valid for only a subset of the reactions [6].

We wish to create a process model for the large, discrete event system that, in the limiting case of large population sizes, captures the Langevin equations, and in the case of small population sizes, can be represented using stochastic discrete interactions.

4 Poisson Processes for Mass-Action Models

Intuitively, the mass action law formalizes the notion that if the size of a system (both in populations and the environment) increases, then the rate of actions also increases. Mathematically, if the rate of the i^{th} reaction is $\lambda_i(X)$, which is a function of the population size X , then the mass action law enforces the limiting property [8]:

$$\lim_{X \rightarrow \infty} \lambda_i(X) = k\lambda_i\left(\frac{X}{k}\right) \quad (6)$$

More specifically, in the context of chemical reactions, the mass action law states that the rate of actions is proportional to the product of the concentration of the dependent populations. Consider a system with M substances (populations), in which r reactions occur. Let $X_i, i = 1, 2, \dots, M$ be the quantities of each substance. Each of the r reactions is represented as follows:

$$R_r = \sum_{i=1}^M \alpha_{i,r} X_i \xrightarrow{k_r} \sum_{i=1}^M \beta_{i,r} X_i \quad (7)$$

where k_r is the kinetic reaction constant associated with each reaction.

The system specified by Equation (7) can then be modeled using a Continuous Time Markov Chain (CTMC), where each state $\vec{X} = (X_i = v | i = 1 \dots M, v \in R)$ corresponds to a distinct valuation for each X_i . In a state, each reaction fires after an exponentially distributed random delay whose rate is given by:

$$\lambda_r(X) = \frac{k_r}{V^{(\sum_{i=1}^M \alpha_{i,r})-1}} \prod_{i=1}^M \frac{X_i!}{(X_i - \alpha_{i,r})!} \quad (8)$$

As the population size increases, that is, $X_i \rightarrow \infty, i = 1 \dots M$, the limiting behavior of the above CTMC can be described by the following differential equations:

$$\lim_{X_1, \dots, X_M \rightarrow \infty} \frac{dX_i}{dt} = \sum_r (\beta_{i,r} - \alpha_{i,r}) k_r \prod_{i=1}^M [X_i]^{\alpha_{i,r}} \quad (9)$$

where

$$[X_i] = \frac{X_i}{V} \quad (10)$$

and $[X_i]$ is the concentration of the substance X_i in the limit. A first order correction for stochastic departure can be introduced to obtain the corresponding Langevin Equations. For the purposes of this paper, we shall instead approximate the continuous behavior using deterministic ordinary differential equations. In future work, we shall investigate the implementation of stochastic differential equations.

We wish to create an approach to simulating large, discrete event systems that, in the limiting case of large population sizes, can be modeled by continuous (and possibly stochastic) differential equations, and in the case of small population sizes, can be represented using event-by-event realizations of the Monte-Carlo simulation method. We hope to leverage advantages inherent meshing the Monte-Carlo and continuous approaches in order to achieve either computational savings or an increase accuracy over an approach that employs a single simulation method.

5 Partitioning Techniques: Continuous versus Discrete Behavior

Our basic approach is to partition the set R of reactions into a set of *continuous* reactions and a set of *discrete* reactions. The continuous reactions evolve using the differential equations, while the discrete reactions evolve using standard

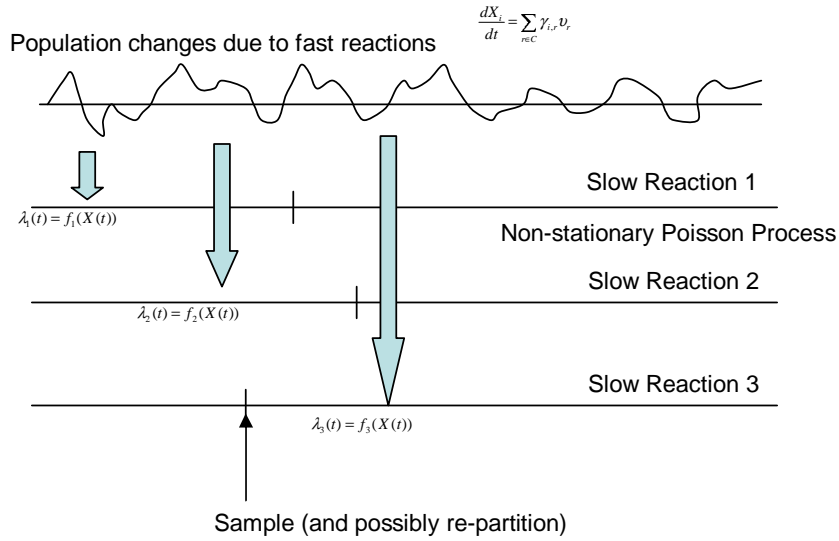


Fig. 1. Partitioning of State Variables into Continuous and Discrete Classes for Simulation

CTMC simulation. During the occurrence of two consecutive discrete reactions, the system state changes due to the continuous reactions. Since these continuous reactions could change the quantities of substances on which the rates of the discrete reactions depend, the discrete reactions have to be represented by non-homogeneous Poisson processes (rate is time varying). Between two consecutive discrete events, progress of the continuous reactions is simulated using standard differential equations simulation.

5.1 Dynamic Partitioning

Although the determination of which reactions exhibit deterministic behavior versus which exhibit stochastic behavior could be done statically based on prior knowledge about the system, the classification into stochastic or deterministic is intrinsically dependent on the system state, which evolves over time. Therefore, more flexibility (and accuracy) can be obtained by deciding the partition as the system evolves (Figure 1). The partition is based on the rates of reaction (as given by Equation 8 and 9, and re-partitioning can be performed either only after a discrete transition, or as and when needed (i.e. even between discrete events). The earlier approach, even though easy to implement, has the drawback that it is possible for the state of the system to change such that the assumptions made by the continuous regime are no longer valid.

Re-partitioning only after discrete events is easy to implement, but has the significant drawback of

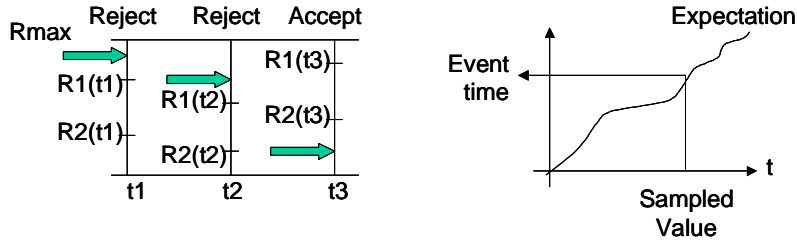
The approaches to perform such a partitioning are given below. They each have their advantages and their drawbacks. Note that the dynamic partitioning described here is performed after a state change due to a discrete event. This ensures that all continuous variables will be updated by the effects of the discrete event instantaneously.

Population Partitioning In the population partitioning approach, the basic idea is to partition the various populations X_i in the system into a set of continuous populations, and a set of discrete populations. Let $P = X_i | i = 1, \dots, M$ be the set of populations in the system. For some state \vec{X} , $C(\vec{X}) \subset P$ is the set of continuous populations, and $D(\vec{X}) = P - C(\vec{X})$ is the set of discrete populations. The set of reactions R is also partitioned into two sets; those which modify any discrete population, and those that do not.

After the occurrence of each event, we reconstruct $C(\vec{X})$ and $D(\vec{X})$ if needed using the rule: if $X_i > k$ then $X_i \in C(\vec{X})$ for some threshold k . Initially, the threshold k would be a user definable parameter based on experimental determination of the magnitude at which the discrete behavior of the species in question is reasonably approximated by continuous equations for the given environmental conditions. If the performance of the simulation algorithm is sensitive to disturbances in k , it becomes necessary to define an *error bound* on k , and to reformulate the algorithm so that it adaptively selects k to be the lowest value that gives a bounded error for any propagation of the simulation within one discrete time step. This technique is similar to the technique used by adaptive meshing algorithms that are used to pick the step size in the numerical solution of differential equations [6]. The error bounds derived on the simulation of non-homogeneous Poisson processes [14] would be useful in this process to limit the highest value selected for k .

Partitioning based on population size is intuitive and reactions may change from discrete to continuous only at discrete events, while reactions may change from continuous to discrete at any point the population size drops below the threshold k . However, only those reactions that do not modify any discrete populations can be simulated using differential equations. The remaining reactions must be simulated using CTMC simulation. The examples in Section 7 utilize this method.

Action Partitioning In this hypothetical approach, the reactions are partitioned directly without partitioning the populations as well. This partitioning is done strictly based on the rate of reaction. Hence, some limiting condition that provides an estimate of the error between the stochastic model and the deterministic model for a reaction with a given rate must be derived to act as an upper bound. It is unclear at present what exact metric will best serve this purpose. All fast reactions can then be simulated using differential equations, resulting in a great computational savings, thereby greatly increasing the size of system which can be modelled. However, reactions may change from discrete to continuous at any time, and hence the partitions may have to be updated at



(a) Accept/Reject Process for Next Event (b) Expected Value of Next Event Time

Fig. 2. (a) Next event generation: Arrow represents the maximum rate for the reactions R1 and R2 (b) "Time to Next Event" Process Generation

every differential equation time step. Stochastically simulating reactions which depend on both continuous and discrete variables is more computationally intensive, due to the increase in Poisson parameters λ_i in the stochastic process. All the populations can change between two discrete events. Hence the number of continuous parameters to the continuous Poisson process increases, thus decreasing its efficiency. At this point, it is unclear how to exactly implement this approach, and it will be investigated more thoroughly in future work.

6 Sampling from a Non-Homogeneous Poisson Process

Using the partitioning approaches described in Section 5, the simulation problem reduces to one of simulating a non-homogeneous Poisson process, i.e. one in which the rate changes continuously with time. Such a process can be simulated using the technique of *thinning* [3]. Thinning is essentially a uniformization technique, whereby the variable rate of the Poisson process is uniformized using a constant rate λ_{max} that upper bounds the varying rate of the process between updates. An exponentially distributed random number with rate λ_{max} is then generated to represent a tentative next arrival time for the non homogeneous Poisson process. Let the current time be t and the tentative next arrival time be Δt . The actual rate $\lambda(t + \Delta t)$ of the non-homogeneous Poisson process is computed at time $t + \Delta t$. Then, the tentative arrival time is accepted as the next arrival time of the non-homogeneous Poisson process with probability (Figure 2):

$$\frac{\lambda(t + \Delta t)}{\lambda_{max}} \quad (11)$$

and rejected with probability

$$1 - \frac{\lambda(t + \Delta t)}{\lambda_{max}} \quad (12)$$

If the arrival time is rejected, the process of generating tentative arrival times can be repeated with the current time set to $t + \Delta t$, due to the memoryless nature

of the Poisson process, until an arrival time is finally accepted. Let the random variable N represent the number of arrival events (all but the last one being tentative) that need to be generated to simulate one discrete event. If Δt_i is the i^{th} tentative arrival time, then the probabilistic moment function for N is given by:

$$P[N = n] = \prod_{i=1}^{n-1} \left(1 - \frac{\lambda(t + \sum_{j=1}^i \Delta t_j)}{\lambda_{max}} \right) \frac{\lambda(t + \sum_{j=1}^n \Delta t_j)}{\lambda_{max}} \quad (13)$$

Let:

$$0 \leq r(i) = \frac{\lambda(t + \sum_{j=1}^i \Delta t_j)}{\lambda_{max}} \leq 1 \quad (14)$$

be the acceptance probability for the i^{th} sample, and let:

$$r_{max} = \max_i \{r(i)\} \quad (15)$$

be the maximum possible value for $r(n)$. Then, the average value of N can be upper bounded by the average value of a geometric random variable with success probability r_{max} . That is:

$$E[N] \leq \frac{1}{r_{max}} \quad (16)$$

Clearly, bringing this ratio as close to 1 as possible will result in the generation of fewer event arrival time and improve the running time of the algorithm. Two techniques can be used to do this task. First, using the memoryless nature of the Poisson process, we can make more informed estimates of λ_{max} at the time of each tentative event rejection. Such a technique is similar to adaptive uniformization. The second technique is to use the law of conservation of mass to compute an upper limit. However, this approach is clearly conservative because the law of conservation of mass provides an upper limit on the amount of substance that can ever be generated in the given system. A better approach might be to optimistically assume a smaller upper bound $\lambda_{max}(\textit{optimistic})$, relying on the fact that most of the generated arrival times will not be too far away. Then, while integrating, if the actual rates ever become higher than $\lambda_{max}(\textit{optimistic})$, then we redo the computation from the last event with a larger $\lambda_{max}(\textit{optimistic})$. The efficiency of this computation can be calculated based on the exponential distribution function. Clearly, the more number of times we have to recompute, the less gains we obtain. Also, it is possible that this technique biases the arrival rate distribution.

Algorithm 1: Hybrid Simulation Algorithm

```

HybridSim( $t_1, t_2, X, R, \alpha, \beta$ )
 $t \leftarrow t_1$ 
while  $t < t_2$ 
  if reclassify
     $C_s \leftarrow D_s \leftarrow C_r \leftarrow C_d \leftarrow \phi$ 
    foreach  $x_i \in X$ , if  $|x_i| < \text{threshold}_i$  then  $D_s \leftarrow C_s \cup x_i$ 
    else  $C_s \leftarrow C_s \cup x_i$ 
    foreach  $r_i \in R$ , if  $\text{REACTANTS}(r) \cap D_s \neq \phi$  then  $D_r \leftarrow D_r \cup r$ 
    else  $C_r \leftarrow C_r \cup r$ 
    Stoichiometric matrix  $\gamma_{i,r} \leftarrow \begin{cases} \beta_{i,r} - \alpha_{i,r} & \text{if } r \in C_r \\ 0 & \text{otherwise} \end{cases}$ 
     $\forall r \in D_r$  compute  $\lambda_{max}(i)$ 
     $\lambda_{max} \leftarrow \sum_{r \in D_r} \lambda_{max}(i)$ 
     $\Delta t \leftarrow \text{RANDOM}(\text{Exp}(\lambda_{max}))$ ,  $\Delta \leftarrow \min(\Delta t, t_2)$ 
     $\Delta t' \leftarrow \text{ODESOLVE}(dx_i/dt = \sum_r \gamma_{i,r} k_r \prod_{i=1}^M [|x_i|]^{\alpha_{i,r}}$ , over  $[t, t + \Delta]$ )
    if  $\Delta t' < \Delta t$  (threshold crossed in OdeSolve) then reclassify  $\leftarrow$  true
    else
      foreach  $r \in D_r$   $\lambda(r) = k_r/V \sum_{i=1}^M \alpha_{i,r}^{-1} \prod_{i=1}^M |x_i|! / (|x_i| - \alpha_{i,r})!$ 
      if ( $u \leftarrow \text{RANDOM}(U[0, 1])$ )  $< \sum_{r \in D_r} \lambda(r) / \lambda_{max}$ 
         $m \leftarrow \min r \in D_r$ , s.t.  $\left\{ u < \sum_{j=1}^r \lambda(j) / \lambda_{max} \right\}$ 
        FIRE(Reaction  $m$ )
        if threshold crossed during firing then reclassify  $\leftarrow$  true
       $t \leftarrow t + \Delta t'$ 

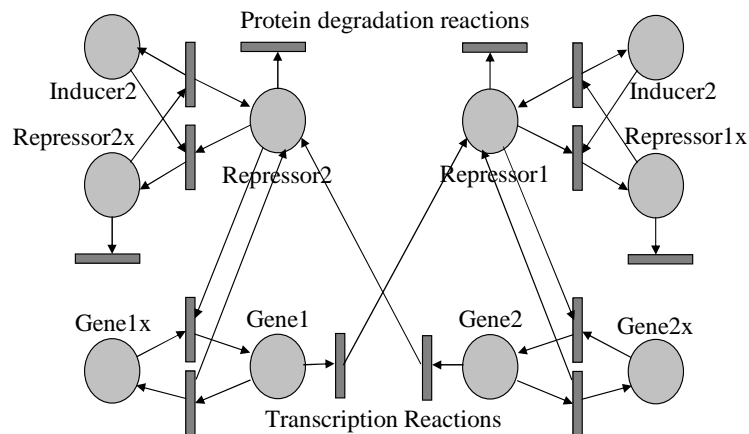
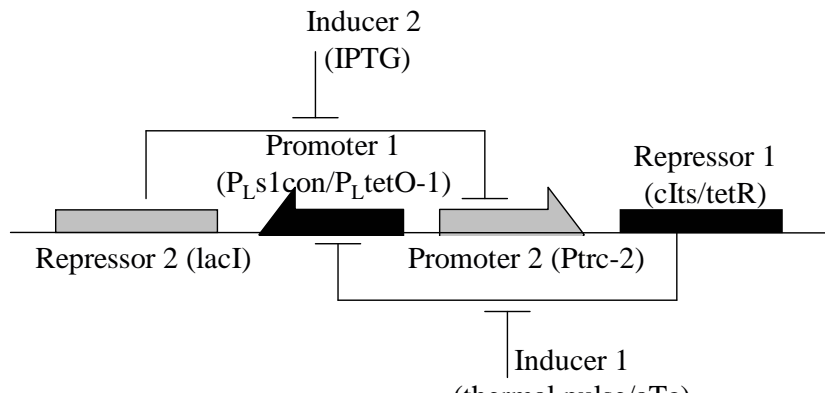
```

7 Simulator Implementation and Examples

7.1 Simulator Implementation

7.2 A Genetic Toggle Switch

Gardner *et al.* [23] describe a genetic switch that toggles between stable transcription from either of two promoters in response to external signals. It is constructed from two promoters and their cognate repressors, arranged so that each promoter can be inhibited by the repressor transcribed by the other promoter. The circuit has two stable states: the "high" state with the promoter P2 on, and the "low" state with the promoter P1 on, given comparable promoter kinetics. Gardner et al. [4] demonstrate how this inherently bi-stable relation can be engineered by using the *IPTG* inducible *lac* promoter (*Plac*), and the thermally unstable *CI* repressor, whereby the state of the switch is sensed using a reporter gene coding for green fluorescent protein (*GFP*) directly downstream from the *CI* gene. The "high" state occurs with the introduction of inducer 1 (I1), *IPTG*, which turns the promoter P2 (*Plac*) on, and produces repressor 2 (R2) *TetR* and the reporter *GFP*. The "low" state occurs when inducer 2 (I2),



(b) Stochastic Petri-Net Model for the Toggle Switch

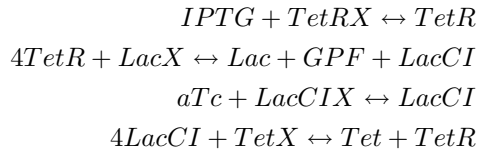
aTc, is introduced, and the promoter P1, *Ptet*, is turned on, and *LacCI* is made (Figure 3(a)). Gardner et al. found that all cells grown in colonies for six hours in the presence of *IPTG* under conditions for *CI* stability remained in the P2 on state after the removal of *IPTG*. Transiently increasing the temperature so that *CI* was unstable for 7 hours resulted in the toggling of the system state to P1 on.

Histogram of Reporter Concentration Distribution over Time for 1000 Simulation Runs for Toggle Switch Example

For the purpose of simulation, using Poisson event generation for CTMC processes and the Langevin approximation, the reporter distribution was simulated for an initial setting of the toggle switch as being in the "high" state, that is, the *IPTG* inducer signal (30 ng/ml) is present, and *Plac* is expressed. Then,

the signal is removed at 500 seconds, and the *aTc* inducer signal (30 ng/ml) is applied at 3000 seconds. The kinetic reaction parameters are taken from the literature [23,24]. The concentration of the reporter distribution with time over 1000 runs is shown in Figure 4. The simulation concentration of *GFP* starts at approximately 15 ng/ml, and ends at an average of 0 ng/ml over 1000 runs.

The main equations that govern this system can be expressed (Figure 3(a)):



where *geneX* is the unexpressed form of the promoter or repressor gene. Dynamic partitioning via population was performed, and thresholds for continuous behavior were set for concentrations at 0.75 of the predicted normalized final value. Thus, the first two reactions in [?] start as continuous reactions, and the latter are simulated discretely. At approximately 6000 seconds, the first two equations are reclassified as discrete, and by 8000 seconds the latter two equations become continuous in simulation. The results compare favourably with those found in experimental literature, in that the behavior is correct [13]. The computation time taken for this example is marginally better than that achieved by statically partitioned hybrid simulation [9]. It is hoped that greater computational saving will be evinced in the simulation of a more complex example, such as the yeast Glucose-Galactose system.

Yeast Glucose-Galactose Model [?]

Presently, the simulation of the yeast Glucose-Galactose switch, a larger example, which possesses a signal transduction mechanism and a single metabolic pathway with Protein-Protein interactions and regulated gene expression is being attempted. This system has been extensively described [?,?] and possesses a large number of mutants and phenotypes, as seen in Figure 5. Unfortunately, upon simulation, the results did not properly match the experimental findings. An elementary analysis, in which an invariant regarding the mass conservation law was formulated to check the maximal concentration values of all reactants, revealed the unregulated overproduction of a single product in the early stages of the simulation results, leading to the discovery of a mistake in the entry of a reaction in the simulation model using a single simulation run. This allowed for the debugging of the model (which consists of numerous reactions) in an online fashion, at a comparatively cheap computational cost, as usually the simulation would be forced to completion (generally hundreds of runs) before the results would be synthesized and a conclusion made.

Even after the model was extensively debugged, the results were still inaccurate, and the major divergence between experimental and simulated results occurred in the over-production of glucose. For a large number of simultaneously

occurring biological reactions which have common products and reactants, such as the yeast Glucose-Galactose switch, even if the reactions can be abstracted into a class of fast and slow reactions, experimentally the fast reactions often reach a bottleneck. This may be due in part to the coupled nature of the reactions: the slow reactions may not produce enough products, or release enough catalytic converters (such as UDP, see Figure 5) to enable some of the fast reactions to remain fast [9]. The bottleneck can lead to an aperiodic switching of reactions between the set of fast reactions and the set of slow reactions. Certain reactions then play a key role in system dynamics: the reactants behave according to their original partitioning, until some limiting resource is encountered, whereby a subset of the fast reactions is then suddenly inhibited, until the resource becomes plentiful once again. It is postulated that certain driving reactions play a critical role in complex, coupled systems. The critical reactants/products and enabling catalysts which drive the entire process forwards must be identified a priori. This would help to identify the kinetic reaction constants which dominate each reaction regime, that would in turn lead to a better dynamic partitioning scheme.

8 Conclusions and Future Work

Quantitative models and simulation techniques for the analysis of complex biological and genetic processes are indispensable in the construction and evaluation of these systems.

The ability to dynamically approximate and allocate the behavior of fast reactions using continuous dynamics such as the Langevin Equations, and slow dynamics through non-homogeneous Poisson processes in continuous time Markov chains has proved to be a successful technique for modeling these systems, Future efforts will focus on creating a hierarchical representation of the biological models which will enable the user to ascertain whether the model has all of the correct reactions entered. As well, this will hopefully aid in the identification of key reactions, that, in complex models, may create bottlenecks in efficiency in the dynamic partitioning of fast and slow reactants, thereby allowing for more complex examples to be accurately simulated.

9 Acknowledgements

This work was supported, in part, by a grant from the National Science Foundation under the Embedded and Hybrid System Project Grant Number: CCR-0311616, and by a gift from the Pioneer Hi-Bred International Incorporation. The authors would like to thank Helen Gill of the NSF, and Jean Peccoud of Pioneer for all of their help.

References

1. Juan P. Aparicio and H. G. Solari. Population dynamics: Poisson approximation and its relation to the langevin process. *Physical Review Letter*, 86(18), April 2001.

2. T.C. Gard. *Introduction to Stochastic Differential Equations*. Marcel Dekker, New York, 1998.
3. C.W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer-Verlag, Berlin, 2nd ed. edition, 1990.
4. T.J. Gardner, C.R. Cantor, and J.J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342, 2000.
5. M.A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, 104:1876–1889, 2000.
6. M.S. Gibson and E. Mjolsness. *Computational Methods for Modelling Biochemical Networks*. MIT Press, Cambridge, MA, in press.
7. D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81:2340–2361, 1977.
8. D.T. Gillespie. *Markov Processes: An Introduction for Physical Scientists*. Academic Press, New York, 1992.
9. E. L. Haseltine and J.B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *Journal of Chemical Physics*, 117(15):6959–6969, October 2002.
10. Mads Kaern, William J. Blake, and J.J. Collins. The engineering of gene regulatory networks. *Annu. Rev. Biomed. Eng.*, 5:179–206, 2003.
11. Jason Kastner, Jerry Solomon, and Scott Fraser. Modeling a hox gene network in silico using a stochastic simulation algorithm. *Developmental Biology*, 246, Elsevier Science, 246:122–131, 2002.
12. G.T. Kurtz. *Stochastic Nonlinear Systems in Physics, Chemistry and Biology*, chapter Approximation of Discontinuous Processes by Continuous Processes. Springer-Verlag, Berlin, 1981.
13. J. Peccoud and Kent Vander Velden. Molecular networks: Structure, dynamics and applications. In *International Multiconference on Measurement, Modelling and Evaluation of Computer Communication Systems*, Urbana, IL, September 2003.
14. H.G. Solari and M.A. Natiello. Poisson approximation to density dependent stochastic processes: A numerical implementation and test. *Physical Review*, 2003. in press.
15. H.G. Solari and M.A. Natiello. Stochastic population dynamics: the poisson approximation. *Physical Review E*, 67(031918), 2003.