

Production-Passage-Time Approximation: A New Approximation Method to Accelerate the Simulation Process of Enzymatic Reactions*

Hiroyuki Kuwahara¹ and Chris Myers²

¹ School of Computing, University of Utah
Salt Lake City, UT 84112, U.S.A.
kuwahara@cs.utah.edu

² Department of Electrical and Computer Engineering, University of Utah
Salt Lake City, UT 84112, U.S.A.
myers@ece.utah.edu

Abstract. Given the substantial computational requirements of stochastic simulation, approximation is essential for efficient analysis of any realistic biochemical system. This paper introduces a new approximation method to reduce the computational cost of stochastic simulations of an enzymatic reaction scheme which in biochemical systems often includes rapidly changing fast reactions with enzyme and enzyme-substrate complex molecules present in very small counts. Our new method removes the substrate dissociation reaction by approximating the passage time of the formation of each enzyme-substrate complex molecule which is destined to a production reaction. This approach skips the firings of unimportant yet expensive reaction events, resulting in a substantial acceleration in the stochastic simulations of enzymatic reactions. Additionally, since all the parameters used in our new approach can be derived by the Michaelis-Menten parameters which can actually be measured from experimental data, applications of this approximation can be practical even without having full knowledge of the underlying enzymatic reaction. Furthermore, since our approach does not require a customized simulation procedure for enzymatic reactions, it allows biochemical systems that include such reactions to still take advantage of standard stochastic simulation tools. Here, we apply this new method to various enzymatic reaction systems, resulting in a speedup of orders of magnitude in temporal behavior analysis without any significant loss in accuracy.

1 Introduction

This paper considers a well-stirred chemically reacting system with the following enzymatic reaction scheme:



* This material is based upon work supported by the National Science Foundation under Grant No. 0331270.

where E, S, C, and P represent an enzyme, a substrate, an enzyme-substrate complex, and a product, respectively, and k_1 , k_{-1} , and k_2 represent non-zero rate constants for the reaction channels, R_1 , R_{-1} and R_2 , respectively. This enzymatic reaction scheme specifies the transformation of S into P catalyzed by E where E has one active site to which S can bind to form C. These types of enzymatic reactions can be found in many biochemical pathways such as metabolic pathways, and therefore, abstracting away low-level details found in enzymatic reaction schemes may have a significant computational benefits.

Traditionally, biochemical systems—including the enzymatic reaction system that is considered in this paper—are modeled and analyzed typically within the continuous-deterministic, *classical chemical kinetics* (CCK) framework based on the *law of mass action* where the dynamics of a well-stirred system is described by a set of ordinary differential equations (ODEs). However, the limitations of the CCK analysis have been broadly accepted [1–4]. In particular, given the same initial condition, the CCK analysis of biochemical systems always produces the same results as it neglects fluctuations. Such treatment, nevertheless, can be justified when the molecular populations are very large, and hence a CCK analysis may provide the most efficient approach to determine the time evolution of a system in such cases. However, many regulatory components in biological systems are often present in amounts too small to simply neglect the effects of inherent fluctuations [5–8]. Moreover, if a system being analyzed has multiple steady states, the traditional ODE approach may not be able to provide an accurate time evolution of a system since it cannot capture spontaneous transitions between steady states [9, 10].

In order to more accurately predict the temporal behavior of biochemical systems without acquiring more information on a biological system such as the positions and the velocities of every molecule, the *stochastic chemical kinetics* (SCK) framework can be used [11]. SCK describes the time evolution of a well-stirred biochemical system as a discrete-state jump Markov process that is analytically governed by the *chemical master equation* (CME) [12]. The CME is derived from the *state-change vector*, specifying the change in each molecular species population for each reaction, and a *propensity function* for each reaction. For example, the enzymatic reaction scheme (1) contains the following propensity functions for each reaction R_i :

$$R_1 : a_1(\mathbf{x}) = k_1 x_E x_S, \quad R_{-1} : a_{-1}(\mathbf{x}) = k_{-1} x_C, \quad R_2 : a_2(\mathbf{x}) = k_2 x_C$$

where $\mathbf{x} = (x_E, x_S, x_C, x_P)$, and each x_* is the value of random variable $X_*(t)$ representing the molecular population of the species subscripted. Thus, the vector of these random variables: $\mathbf{X}(t) = (X_E(t), X_S(t), X_C(t), X_P(t))$ represents the system state at time t . Assuming that the system is spatially homogeneous, this SCK approach describes the time evolution of a biochemical system at the individual reaction level by exactly tracking the quantities of each molecular species and by treating each reaction as a separate random event. However, directly obtaining the solution of the CME of any realistic system, either analytically or numerically, is not feasible due to its intrinsic complexity. Note that, though

it is possible to numerically solve the CME of the enzymatic reaction scheme (1) as the system state is bounded albeit with potentially substantial computational demands, if systems also contain other reactions and species as is the case for many realistic biological systems, then the space complexity of CMEs of such systems often inevitably becomes too large to be tractable, making the numerical solutions of such CMEs infeasible. Consequently, the time evolution of moments $\langle \mathbf{X}^n(t) \rangle$ is also generally infeasible to compute from the CME.

To overcome this, several methods have been introduced to approximate the time evolution of moments of the process $\mathbf{X}(t)$ without solving the CME [9, 13]. Such approximations are very useful to efficiently understand the mean behavior, standard deviation, skewness, etc. of $\mathbf{X}(t)$, as well as to potentially characterize the time evolution of the asymptotic probability distribution of the system states. However, utilizing such methods alone may encounter difficulties in quantitative analyses of some biologically relevant properties based on stochastic competition such as probabilistic analysis of lysis/lysogeny developmental pathways in bacteriophage λ -infected *Escherichia coli* [1]. Furthermore, since the complexity of the moment evolution equations may significantly increase as that of a system increases [9], such approach may be unwieldy for a large-scale biological system.

Instead of attempting to solve the CME, exact discrete-stochastic numerical realizations of a system via Gillespie’s *stochastic simulation algorithm* (SSA) [14], which is derived from the same premise as the CME, are often used to infer the temporal system behavior with a much smaller memory footprint. This Monte Carlo simulation approach is useful to intuitively observe the trend of system dynamics, which may be possible with as few as tens of numerical realizations. Furthermore, *in silico* experiments via Monte Carlo simulation come with potentially unlimited controlling capabilities and abilities to capture virtually any dynamical properties of the system, making a number of qualitative and quantitative analyses which cannot be done in *wet-lab* experiments possible. Unfortunately, the computational requirements of the SSA—even with the Gibson and Bruck optimization [15], which, among other things, reduces the generations of the random numbers by reusing them—can be substantial. This is due largely to the fact that it not only requires a potentially large number of simulation runs in order to estimate the system behavior at a reasonable degree of statistical confidence, but it also requires every single reaction event to be simulated one at a time. For example, if $k_2 \ll k_{-1}$ in the enzymatic reaction scheme (1), then C dissociates into S much more often than into P, and thus, much of the computation time is allocated for this *substrate-complex loop*.

Several approximation methods have been proposed to accelerate the simulation process of the SSA by sacrificing exactness. For example, the explicit τ -leaping method approximates the number of firings of each reaction in a pre-defined interval rather than executing each reaction individually [16]. While this and similar methods [17–19] are very promising, they may not perform well for an enzymatic reaction which bears rapidly changing fast reactions driven by the enzyme and enzyme-substrate complex molecules present in very small counts because the *leaping condition* may not be satisfied in such situations.

Some acceleration methods for the stochastic simulations of enzymatic reactions have been proposed that perform well even when the enzyme is present in very small amounts by eliminating the undesirable substrate-complex loop in the enzymatic reaction scheme. For example, Rao and Arkin have performed model abstraction by using biochemical insight in combination with the *quasi-steady-state approximation* (QSSA) to remove the expensive substrate-complex loop, and then applied a modified version of SSA to the simplified model [20]. Cao et al. [21] have demonstrated how the substrate-complex loop can be removed by applying the enzyme substrate reaction system to their slow-scale SSA approach which explicitly simulates the firings of only the slow reaction events [22]. Both approximation methods require the use of special simulation procedures. Thus, there might be cases where one finds the use of these approximation methods inconvenient when it comes to the analysis of a system containing enzymatic reactions along with other types of reactions. Such cases occur, for example, when a biochemical system is represented in the *Systems Biology Markup Language* (SBML), the emerging standard format to represent models of biochemical reaction networks [23]. SBML level 2 version 1 contains reactions only in the generic type, and it cannot specify any specific reaction types without using proprietary annotations. Thus, in order for SBML compliant SSA tools to know when to use a specially tailored Monte Carlo simulation procedure for enzymatic reactions, the tools must either understand the semantics of proprietary fields that specify enzymatic reactions or perform structural analysis to find enzymatic reactions.

This paper introduces a new approximation approach to accelerate the process of stochastic simulations of enzymatic reactions. Our new approach, which we call *production-passage-time approximation* (PPTA), approximates the passage time of the complex C which is destined to turn into the product P, and only tracks such instances of C. Thus, this approach eliminates the substrate-complex loop by removing R_{-1} , allowing a substantial acceleration in stochastic simulations of enzymatic reactions. Furthermore, since our approach does not require a customized simulation procedure for enzymatic reactions, it allows a biochemical system comprising the PPTA reactions along with other types of reactions to still be modeled using a SBML modeling tool such as PathwayBuilder from BioSPICE [24], and analyzed by using any SBML compliant SSA tools.

This paper describes the PPTA method in Section 2. Section 3 demonstrates how our approach can help analyze the temporal behaviors of enzymatic one-substrate reaction models efficiently with reasonable accuracy. This is shown by applying our new approximation method to various systems and comparing the full models with the corresponding PPTA models in terms of the accuracy—by calculating means and standard deviations—as well as runtime. Finally, this paper concludes in Section 4 by discussing the benefits gained by the PPTA.

2 Production-Passage-Time Approximation

To describe the PPTA method, the enzymatic reaction scheme (1) is first considered to have the initial condition: $\mathbf{X}(t_0) = \mathbf{x}_{t_0}$, where $\mathbf{x}_{t_0} = (e_{tot}, s_{tot}, 0, 0)$,

$e_{tot} \geq 1$, and $s_{tot} \geq 1$. Let $\mathbf{x}_\infty = (e_{tot}, 0, 0, s_{tot})$, then the probability that $\mathbf{X}(t) = \mathbf{x}_\infty$ given $\mathbf{X}(t_0) = \mathbf{x}_{t_0}$ approaches 1, as $t \rightarrow \infty$. In other words, in any simulation runs, the enzymatic reaction process always reaches \mathbf{x}_∞ eventually. In order for each numerical realization of $\mathbf{X}(t)$ to transition from \mathbf{x}_{t_0} to \mathbf{x}_∞ , S must be transformed into C at least s_{tot} times and C must be converted into P exactly s_{tot} times. Thus, let $\mathbf{x}^{(i)}(t)$ be the i -th sample trajectory of $\mathbf{X}(t)$ given that $\mathbf{X}(t_0) = \mathbf{x}_{t_0}$ and \mathbf{T}_i be a set of time instances such that each time instance t_j^i represents the time point where the j -th reaction event occurs in $\mathbf{x}^{(i)}(t)$. Then, the statement $\forall i. |\mathbf{T}_i| \in [2s_{tot}, \infty)$ must be true. Intuitively, if $k_{-1} \ll k_2$, then C tends to be consumed by R_2 rather than R_{-1} , making the size of each \mathbf{T}_i close to the lower bound $2s_{tot}$. On the other hand, if $k_{-1} \gg k_2$, then C is more likely to be consumed by R_{-1} , and in consequence each $|\mathbf{T}_i|$ is very likely much greater than $2s_{tot}$, making the computational cost of simulations significantly higher.

Our new PPTA approach minimizes the number of reaction events that fire through the passage of each $\mathbf{x}^{(i)}(t)$ to \mathbf{x}_∞ by preventing each $\mathbf{x}^{(i)}(t)$ from revisiting the same state. Thus, it guarantees that $\forall i. |\mathbf{T}_i| = 2s_{tot}$. This is achieved by eliminating R_{-1} and approximating transitions of each $\mathbf{x}^{(i)}(t)$ using only complex-formation and production reactions. In other words, the PPTA approximates the passage time of the formation of each C molecule which leads to a production of P, and only keeps track of such instances of formation of C, rather than explicitly also simulating the formation of C molecules that are destined to dissociate into E and S molecules. Therefore, the PPTA can accelerate the stochastic simulations of the enzymatic reaction scheme (1), especially when $k_{-1} \gg k_2$ where the reduction in each $|\mathbf{T}_i|$ by this new approach is substantial.

Let us first consider the special case where the total molecular count of the enzyme is 1 (i.e., $e_{tot} = 1$), and describe the derivation of the PPTA model. This section then extends this special case to more general cases where the total molecular count of the enzyme is greater than 1 (i.e., $e_{tot} > 1$).

When e_{tot} is 1, the enzyme state for all $t \geq 0$ is defined by $X_E(t) = 1 - X_C(t)$. Also, R_1 is only enabled when E is active (i.e., $X_E(t) = 1$), and R_{-1} and R_2 are only enabled when C is active (i.e., $X_C(t) = 1$). In this case, $\mathbf{X}(t)$ can be seen as a temporal-homogeneous birth-death Markov process $\mathbf{Y}(t)$ with $2s_{tot} + 1$ states as shown in Figure 1. Each state $s \in [0, 2s_{tot}]$ of $\mathbf{Y}(t)$ can then be mapped onto a system state \mathbf{x}_s of $\mathbf{X}(t)$ by the relationship: $\mathbf{x}_s \equiv ((s + 1) \bmod 2, s_{tot} - \lceil s/2 \rceil, s \bmod 2, \lfloor s/2 \rfloor)$. Thus, for all $t > t_0$, the probability that $\mathbf{Y}(t) = s$ given that $\mathbf{Y}(t_0) = 0$ is the same as the probability that $\mathbf{X}(t) = \mathbf{x}_s$ given that $\mathbf{X}(t_0) = \mathbf{x}_{t_0}$, and with the initial condition $\mathbf{X}(t_0) = \mathbf{x}_{t_0}$, each simulation run of $\mathbf{Y}(t)$ starts in state 0, and eventually ends up in state $2s_{tot}$. Since E is active only in even number states in this process, R_1 can fire only in these states except in state $2s_{tot}$. Similarly, C is active only in odd number states, so R_{-1} and R_2 can fire in these states. Thus, let \mathbf{S}_e be a set of even number states $\{2m \mid 0 \leq m \leq s_{tot}\}$, $\mathbf{S}_{e'}$ be a set of states $\mathbf{S}_e \setminus \{2s_{tot}\}$, and \mathbf{S}_o be a set of odd number states $\{2m + 1 \mid 0 \leq m < s_{tot}\}$. Then, the $s \rightarrow s + 1$ transition rate λ_s is $a_1(\mathbf{x}_s)$ if $s \in \mathbf{S}_{e'}$, and $a_2(\mathbf{x}_s)$ if $s \in \mathbf{S}_o$, whereas the $s \rightarrow s - 1$ transition rate μ_s is $a_{-1}(\mathbf{x}_s)$ if $s \in \mathbf{S}_o$ and 0 if $s \in \mathbf{S}_e$.

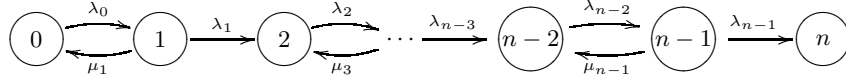


Fig. 1. The state graph of the birth-death process of the enzymatic reaction scheme (1) when $e_{tot} = 1$. This birth-death process has $n + 1$ states where $n = 2s_{tot}$, and each state s can be mapped onto a system state of $\mathbf{X}(t)$ by the relationship $\mathbf{x}_s \equiv ((s + 1) \bmod 2, s_{tot} - \lceil s/2 \rceil, s \bmod 2, \lfloor s/2 \rfloor)$. Transition rate λ_s is $a_1(\mathbf{x}_s)$ if s is an even number, and $a_2(\mathbf{x}_s)$ if s is an odd number. Transition rate μ_s is $a_{-1}(\mathbf{x}_s)$ if s is an odd number and 0 otherwise.

Suppose $\mathbf{Y}(t)$ starts in state s_0 where $s_0 \in \mathbf{S}_{e'}$. Then, the average waiting time that $\mathbf{Y}(t)$ spends in states s_0 and $s_0 + 1$ for each simulation run are equivalent to $t(s_0; s_0 \rightarrow s_0 + 2)$ and $t(s_0 + 1; s_0 \rightarrow s_0 + 2)$, respectively, where $t(s_j; s_i \rightarrow s_k)$ is the mean time that $\mathbf{Y}(t)$ spends in state s_j in the course of a (first) passage from s_i to s_k . In other words, using the variable $t(s_j; s_i \rightarrow s_k)$,

$$\begin{aligned} t(s_0; s_0 \rightarrow s_0 + 2) &\equiv t(s_0; 0 \rightarrow 2s_{tot}), \\ t(s_0 + 1; s_0 \rightarrow s_0 + 2) &\equiv t(s_0 + 1; 0 \rightarrow 2s_{tot}), \end{aligned}$$

since the transitions: $s_0 \rightarrow s_0 - 1$ and $s_0 + 2 \rightarrow s_0 + 1$ are not allowed in $\mathbf{Y}(t)$. To find out the mean waiting times in states s_0 and $s_0 + 1$ using the *pedestrian approach* [9], then, variables: $v(s)$ and $v_+(s)$ are defined. The variable $v(s)$ is defined as the average number of visits by $\mathbf{Y}(t)$ to state s in the course of a first passage from state 0 to state $2s_{tot}$ while $v_+(s)$ is defined as the average number of transitions $s \rightarrow s + 1$ taken by $\mathbf{Y}(t)$ in the course of a first passage from state 0 to state $2s_{tot}$. Using these variables, the probability that $\mathbf{Y}(t)$ moves to state $s_0 + 2$ from state $s_0 + 1$ at the very next jump can be expressed as $v_+(s_0 + 1)/v(s_0 + 1)$. Since this probability can also be expressed as $\lambda_{s_0+1}/(\lambda_{s_0+1} + \mu_{s_0+1})$, and since $v_+(s_0 + 1)$ is 1, we can say

$$v(s_0 + 1) = \frac{(\lambda_{s_0+1} + \mu_{s_0+1})}{\lambda_{s_0+1}}.$$

Because state $s_0 + 1$ can only be visited from state s_0 in $\mathbf{Y}(t)$, $v_+(s_0)$ must be equal to $v(s_0 + 1)$. Furthermore, since the transition from state s_0 to state $s_0 - 1$ cannot occur in $\mathbf{Y}(t)$, $v(s_0)$ must be equivalent to $v(s_0 + 1)$. Therefore,

$$v(s_0) = \frac{(\lambda_{s_0+1} + \mu_{s_0+1})}{\lambda_{s_0+1}}.$$

Now, let $T(s)$ be a random variable which represents the pausing time in state s in $\mathbf{Y}(t)$. Then, since $\mathbf{Y}(t)$ is a temporally homogeneous birth-death Markov process, $T(s)$ must be a random variable which is necessarily exponentially distributed with parameter $(\lambda_s + \mu_s)$. Then, the mean pausing times in states s_0

and $s_0 + 1$ can be expressed, respectively, as:

$$\begin{aligned}\langle T(s_0) \rangle &= \int_0^\infty t \lambda_{s_0} \exp(-\lambda_{s_0} t) dt = \frac{1}{\lambda_{s_0}}, \\ \langle T(s_0 + 1) \rangle &= \int_0^\infty t (\lambda_{s_0+1} + \mu_{s_0+1}) \exp(-(\lambda_{s_0+1} + \mu_{s_0+1}) t) dt \\ &= \frac{1}{\lambda_{s_0+1} + \mu_{s_0+1}}.\end{aligned}$$

Since $t(s_j; s_i \rightarrow s_k)$ can be formulated as the product of $\langle T(s_j) \rangle$ and $v(s_j)$, the mean waiting times that $\mathbf{Y}(t)$ spends in states s_0 and $s_0 + 1$ can be expressed as:

$$\begin{aligned}t(s_0; 0 \rightarrow 2s_{tot}) &= \frac{\lambda_{s_0+1} + \mu_{s_0+1}}{\lambda_{s_0+1} \lambda_{s_0}} = \frac{a_2(\mathbf{x}_{s_0+1}) + a_{-1}(\mathbf{x}_{s_0+1})}{a_2(\mathbf{x}_{s_0+1}) a_1(\mathbf{x}_{s_0})}, \\ t(s_0 + 1; 0 \rightarrow 2s_{tot}) &= \frac{1}{\lambda_{s_0+1}} = \frac{1}{a_2(\mathbf{x}_{s_0+1})}.\end{aligned}$$

Using this information, $\mathbf{Y}(t)$ can be approximated by creating a temporally homogeneous birth Markov process $\mathbf{Y}'(t)$ with the same state space where the mean waiting time in each state s is $t(s; 0 \rightarrow 2s_{tot})$ derived from $\mathbf{Y}(t)$. Figure 2 shows the state graph of $\mathbf{Y}'(t)$. Since the waiting time in each state s in $\mathbf{Y}'(t)$ is exponentially distributed, the $s \rightarrow s + 1$ transition rate λ'_s is the reciprocal of $t(s; 0 \rightarrow 2s_{tot})$. Thus, λ'_s is $a_1(\mathbf{x}_s) a_2(\mathbf{x}_{s+1}) / (a_{-1}(\mathbf{x}_{s+1}) + a_2(\mathbf{x}_{s+1}))$ if $s \in \mathbf{S}_o$ and $a_2(\mathbf{x}_s)$ if $s \in \mathbf{S}_{e'}$. Therefore, using the PPTA, the enzymatic reaction scheme (1) with e_{tot} being 1 is approximated by a new reaction scheme:



where $k_{1'} = k_1 k_2 / (k_{-1} + k_2)$.

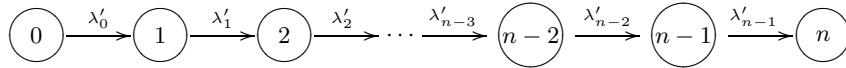
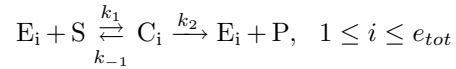
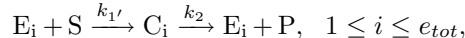


Fig. 2. The state graph of the pure birth process of the PPTA model when $e_{tot} = 1$. This birth process has the same state space as the birth-death process in Figure 1. Transition rate λ'_s is $a_1(\mathbf{x}_s) a_2(\mathbf{x}_{s+1}) / (a_{-1}(\mathbf{x}_{s+1}) + a_2(\mathbf{x}_{s+1}))$ if $s \in \mathbf{S}_o$ and $a_2(\mathbf{x}_s)$ if $s \in \mathbf{S}_{e'}$.

When $e_{tot} > 1$, the enzymatic reaction scheme (1) is considered as a set of the enzymatic reactions as follows:



where $X_{E_i}(t_0) = 1$ and $X_{C_i}(t_0) = 0$ for each i . Although simulations of this process is definitely slower than that of $\mathbf{X}(t)$, this transformation itself does not require any approximation as, when $\mathbf{X}(t) = \mathbf{x}$, $k_1 x_E x_S \equiv \sum_{i=1}^{e_{tot}} k_1 x_{E_i} x_S$, $k_{-1} x_C \equiv \sum_{i=1}^{e_{tot}} k_{-1} x_{C_i}$, and $k_2 x_C \equiv \sum_{i=1}^{e_{tot}} k_2 x_{C_i}$. Thus, by applying the PPTA to each of the transformed enzymatic reactions, the enzymatic reaction scheme (1) can be approximated by



which can now be represented using reaction scheme (2). This implies that the accuracy of the PPTA of the $e_{tot} > 1$ case is based on that of the PPTA of the $e_{tot} = 1$ case, and that the PPTA model provides the most accurate results if $e_{tot} = 1$.

The two parameters in a PPTA model: k_1' and k_2 can be derived from K_M , and V_{max} , the maximal reaction rate. Unlike the parameters: k_1 and k_{-1} , the parameters K_M and V_{max} can actually be measured experimentally. Thus, a PPTA model can be constructed and simulated even when full knowledge of the underlying enzymatic reaction is not available and the enzymatic reaction cannot be analyzed quantitatively at that level of detail. This is also true for a QSSA model as its MM form only requires K_M and V_{max} parameters; however, since a PPTA model does not assume that the intermediate species is in quasi-steady state, a PPTA model may perform better than a QSSA model in terms of accuracy, especially in the pre-steady state phase. For example, suppose the enzymatic reaction scheme (1) has the conditions:

$$\mathbf{x}_{t_0} = (25, 50, 0, 0), \quad k_1 = 100.0, \quad k_{-1} = 10.0, \quad k_2 = 0.1.$$

Then, since e_{tot} is arguably much smaller than s_{tot} , the QSSA *could* be applied to safely approximate the temporal behavior of the underlying enzymatic reaction. However, in this system, the propagation effects of the pre-steady state dynamics are rather important, making any QSSA-based models unable to describe the temporal behavior well. Therefore, as shown in Figure 3, the estimated mean time evolution of this enzymatic reaction model is captured more accurately by the PPTA model than by the QSSA model.

3 Case Studies

This section describes the benefits gained by the PPTA method by applying it to various systems containing enzymatic reaction scheme. This section first considers two model of the enzymatic reaction scheme (1) which are used to help illustrate the application of the slow-scale SSA in [21]. It then considers the *enzymatic futile cycle* motif which can be ubiquitously seen in biological systems including GTPase cycles, mitogen-activated protein kinase cascades, and glucose mobilization [4]. Finally, it considers a more complex competitive enzymatic reaction. Each model is encoded in SBML [23] and simulated for 1,000 runs using the same stochastic simulator, an optimized SSA implementation within

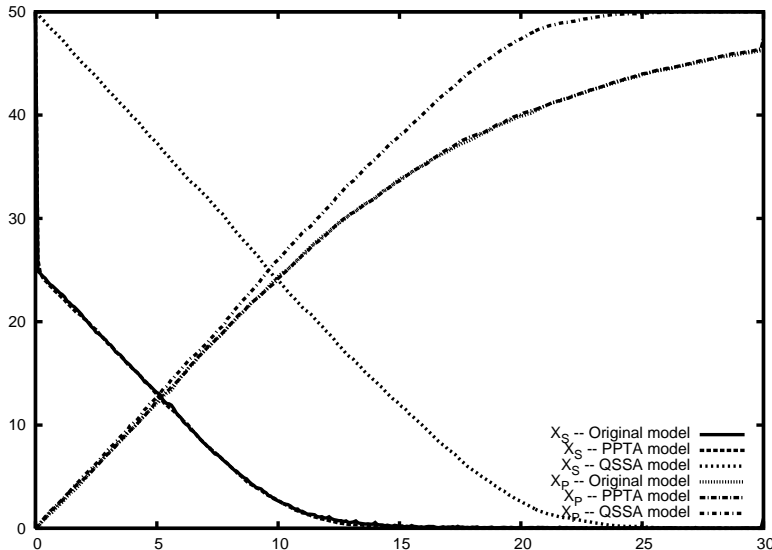


Fig. 3. Comparison of the average time evolutions of the original enzymatic reaction model, its PPTA model, and its QSSA model with initial conditions: $\mathbf{x}_{t_0} = (25, 50, 0, 0)$ and the rate constants: $k_1 = 100.0$, $k_{-1} = 10.0$, $k_2 = 0.01$.

our modeling and analysis tool `reb2sac` [25]. Accuracy of a PPTA model is measured by comparing the time evolution of means and standard deviations.

3.1 Single Enzymatic Reaction

The first model of the enzymatic reaction scheme (1) has the following initial condition and the reaction rate constants:

$$\mathbf{x}_{t_0} = (220, 3000, 0, 0), \quad k_1 = 0.01, \quad k_{-1} = 100.0, \quad k_2 = 0.01.$$

This system is simulated for 20,000 time units and each data point is plotted every 100 time units. Figure 4 shows the results from the original model and the PPTA model of this system. The estimated means and standard deviations of X_S and X_P are shown in Figures 4(a) and (b), respectively.

The results from the PPTA model are in a very close agreement with those from the original model, yet the speedup gained by the PPTA model is significant. While the entire simulation of the original model takes 68.58 hours, that of the PPTA model only takes 22.8 seconds, achieving more than 10,800 times speedup. Furthermore, since the speedup gained by the slow-scale SSA is about 950 on this system, the PPTA method is able to outperform the slow-scale SSA by an order of magnitude.

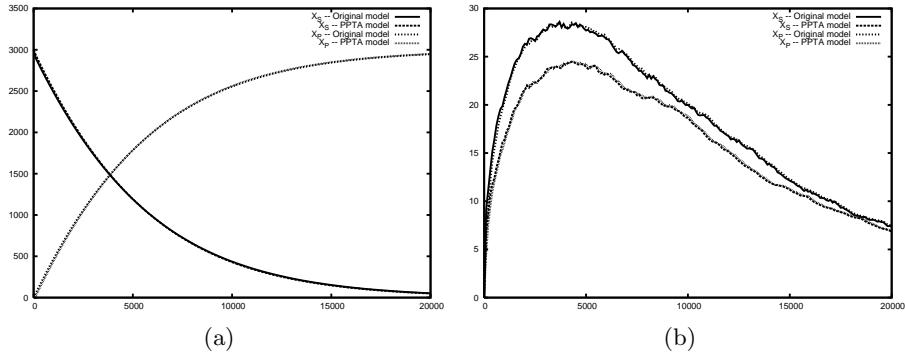


Fig. 4. Comparison of the original enzymatic reaction model and its PPTA model with initial conditions: $\mathbf{x}_{t_0} = (220, 3000, 0, 0)$ and the rate constants: $k_1 = 0.01$, $k_{-1} = 100.0$, $k_2 = 0.01$. (a) Means of X_S and X_P . (b) Standard deviations of X_S and X_P .

The second enzymatic reaction system has the following initial conditions and the specification of the reaction rate constants:

$$\mathbf{x}_{t_0} = (10, 3000, 0, 0), \quad k_1 = 0.01, \quad k_{-1} = 600.0, \quad k_2 = 0.1.$$

This system illustrates a case where the average of $X_C(t)$ remains less than 1 as the maximum reaction rate of R_1 (i.e., $k_1 e_{tot} s_{tot}$) is less than k_{-1} . This system is simulated for 80,000 time units and each data point is again plotted every 100 time units. Figure 5(a) shows the estimated means of X_S and X_P , and Figure 5(b) shows the estimated standard deviations.

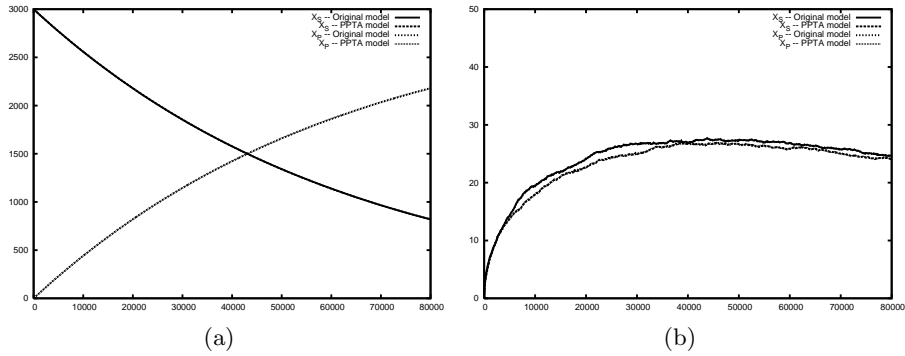
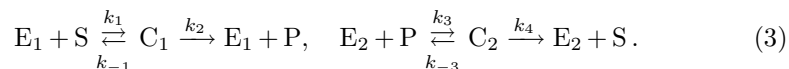


Fig. 5. Comparison of the original enzymatic reaction model and its PPTA model with initial conditions: $\mathbf{x}_{t_0} = (10, 3000, 0, 0)$ and the rate constants: $k_1 = 0.01$, $k_{-1} = 600.0$, $k_2 = 0.1$. (a) Means of X_S and X_P . (b) Standard deviations of X_S and X_P .

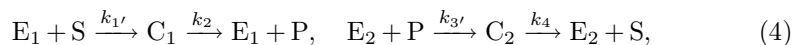
Both the means and the standard deviations from the PPTA model track those from the original model very well while, at the same time, the simulation time of the PPTA is substantially reduced compared with that of the original model. Whereas the simulation of the original model takes 27.63 hours, that of the PPTA model only takes 17.9 seconds, improving the computation performance by a factor of more than 5,500. Since the speedup gained by the slow-scale SSA is 400 on this system, the PPTA method is able to outperform the slow-scale SSA by an order of magnitude.

3.2 Enzymatic Futile Cycle

The enzymatic futile cycle motif consists of two instances of the enzymatic reaction scheme (1) as follows:



One is to transform S into P catalyzed by E_1 , and the other one is to transform P into S catalyzed by E_2 . This motif is found in many biological systems [4], abstracting away low-level detail of the motif such as unproductive substrate-complex cycles may provide a significant improvement in performance of the overall system behavior analysis. With the PPTA method, unproductive dissociation reactions are removed, transforming the enzymatic futile cycle model into the following PPTA model:



where $k_{3'} = k_3 k_4 / (k_{-3} + k_4)$.

The original enzymatic futile cycle model and its PPTA model are simulated for 300 time units with one time unit plot-interval to analyze the accuracy as well as the performance gain of the PPTA model with the initial conditions:

$$(X_S(0), X_P(0), X_{E_1}(0), X_{E_2}(0), X_{C_1}(0), X_{C_2}(0)) = (0, 100, 10, 20, 0, 0),$$

and the rate constants:

$$k_1 = 10^3; k_{-1} = 1.5 \times 10^3; k_2 = 2; k_3 = 10^3; k_{-3} = 5 \times 10^2; \text{ and } k_4 = 1.$$

Since each numerical simulation of the two models starts with no copies of S and 10 copies of E_1 , this system illustrates a case where substrate is initially lower than the catalyzing enzyme. Furthermore, since $X_S(t) + X_P(t) + X_{C_1}(t) + X_{C_2}(t)$ is fixed at 100 for all $t \geq 0$, this enzymatic futile cycle system illustrates an applicability of the PPTA model when the numbers of both substrate and enzyme molecules are very low.

Figure 6 shows the results from the original model and the PPTA model of this enzymatic futile cycle system. The time evolutions of the estimated means

and standard deviations of X_S and X_P are shown in Figures 6(a) and (b), respectively. From these figures, it is clear that both the means and the standard deviations of X_S and X_P from the PPTA model approximate those from the original model very well. The simulation of the original enzymatic futile cycle model takes 17.73 hours while that of the PPTA model only takes 87.51 seconds which represents a speedup of more than 729 times. Furthermore, this demonstrates that the PPTA can be applicable to systems with a low number of substrate molecules.

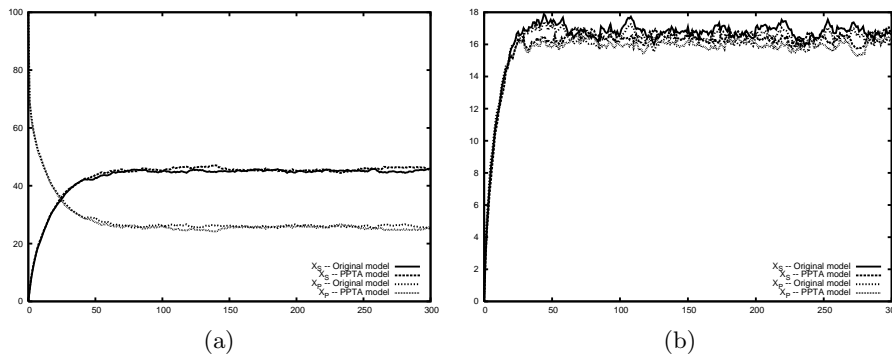
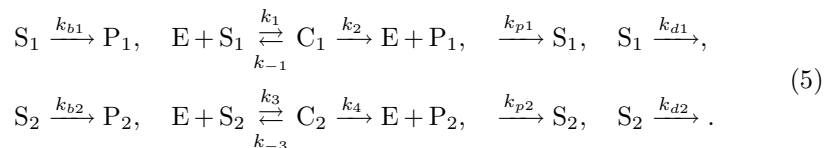


Fig. 6. Comparison of the original enzymatic futile cycle model (3) and its PPTA model (4) with initial conditions: $X_S(0) = 0$, $X_P(0) = 100$, $X_{E_1}(0) = 10$, $X_{E_2}(0) = 20$, $X_{C_1}(0) = 0$, $X_{C_2}(0) = 0$, and the rate constants: $k_1 = 1.0 \times 10^3$, $k_{-1} = 1.5 \times 10^3$, $k_2 = 2.0$, $k_3 = 1.0 \times 10^3$, $k_{-3} = 5.0 \times 10^2$, $k_4 = 1.0$. (a) Means of X_S and X_P . (b) Standard deviations of X_S and X_P .

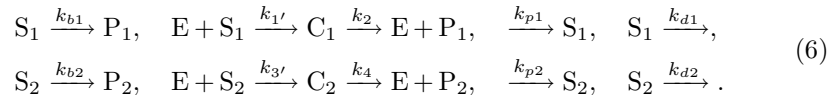
3.3 Competitive Enzymatic Reaction

To further demonstrate the usefulness of the PPTA, the following competitive enzymatic reaction scheme is considered:



In this scheme, both S_1 and S_2 compete to bind to E to produce P_1 and P_2 , respectively. Also, this scheme contains basal reactions to transform S_1 and S_2 into P_1 and P_2 , respectively, without being catalyzed by E . Moreover, since substrates S_1 and S_2 are often produced and consumed via various reactions, reaction scheme (5) also contains reactions to model productions and consumptions of S_1 and S_2 .

The PPTA model of the competitive enzymatic reaction model (5) removes the substrate-dissociation reactions from C_1 and C_2 , resulting in the following model:



To analyze the accuracy of this PPTA model, the following initial conditions:

$$(X_E(0), X_{S_1}(0), X_{S_2}(0), X_{P_1}(0), X_{P_2}(0), X_{C_1}(0), X_{C_2}(0)) = (10, 0, 0, 0, 0, 0, 0),$$

and the rate constants:

$$\begin{aligned} k_{b1} &= 2 \cdot 10^{-5}; k_1 = 10^2; k_{-1} = 10^2; k_2 = 0.1; k_{p1} = 10; k_{d1} = 0.2; \\ k_{b2} &= 10^{-5}; k_3 = 200; k_{-3} = 10^2; k_4 = 0.15; k_{p2} = 10; \text{ and } k_{d2} = 0.2, \end{aligned}$$

are used for the simulations. The values of rate constants for the productions and consumptions of S_1 and S_2 are chosen so that the consumption rate constants are relatively high to capture isolation of substrates from binding to the enzyme and that both substrates are present in low counts throughout the simulations (i.e., $\forall t \geq 0. \langle X_{S_1}(t) \rangle \leq 100 \wedge \langle X_{S_2}(t) \rangle \leq 100$). The values of basal transformation rate constants k_{b1} and k_{b2} are chosen so that basal transformation rates are much smaller than those from the catalyzed reactions when the substrates are present in low counts (i.e., $k_2 \cdot e_{tot} \gg 100k_{b1}$ and $k_4 \cdot e_{tot} \gg 100k_{b2}$).

Figure 7 shows the results from the simulations of the two models. The estimated means of X_{S_1} and X_{S_2} are shown in Figure 7(a), while the estimated standard deviations of X_{S_1} and X_{S_2} are shown in Figure 7(b). Once again, both the means and the standard deviations of X_{S_1} and X_{S_2} from the PPTA model track those from the original model very well with a substantial improvement in simulation time. While the simulation of the original model takes 65.16 minutes, that of the PPTA model only takes 35.78 seconds, achieving more than 109 times speedup.

4 Conclusions

This paper introduces a new model abstraction method, *production-passage-time approximation* (PPTA), that can significantly improve the temporal behavior analysis time of enzymatic reaction systems. As a case study, we have applied the PPTA method to various systems, and compared the accuracy as well as the runtime between the original model and the PPTA model. The preliminary results are promising. This paper has shown that the PPTA model can make stochastic simulations orders of magnitude faster without any significant loss in accuracy. This paper has also shown that the PPTA method achieves an acceleration of an order of magnitude over the slow-scale SSA for the two enzymatic reaction systems from [21]. Moreover, this paper has demonstrated that the PPTA can be

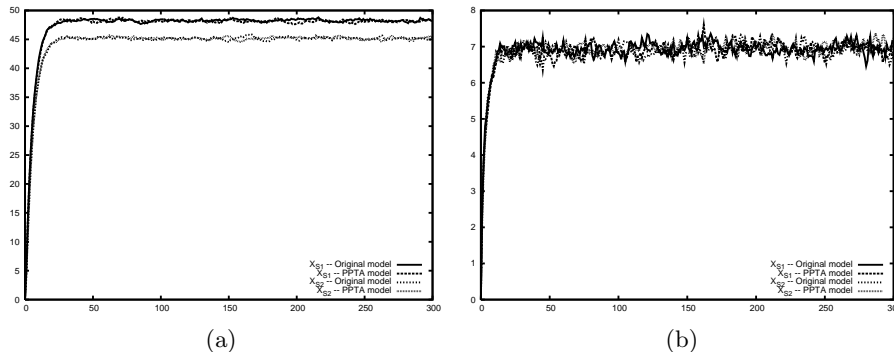


Fig. 7. Comparison of the original model of competitive enzymatic reaction model (5) and its PPTA model (6) with initial conditions: $X_E(0) = 10$ and 0 molecule for the rest of the species, and the rate constants: $k_{b1} = 2 \cdot 10^{-5}$, $k_1 = 10^2$, $k_{-1} = 10^2$, $k_2 = 0.1$, $k_{p1} = 10$, $k_{d1} = 0.2$, $k_{b2} = 10^{-5}$, $k_3 = 200$, $k_{-3} = 10^2$, $k_4 = 0.15$, $k_{p2} = 10$, $k_{d2} = 0.2$. (a) Means of X_{S1} and X_{S2} . (b) Standard deviations of X_{S1} and X_{S2} .

utilized to efficiently approximate more complex systems, exemplified here using an enzymatic futile cycle model and a competitive enzymatic reaction model. Additionally, our approach can also be used within a continuous, deterministic framework to remove the stiff condition often found in enzymatic reactions with $k_{-1} \gg k_2$ that gives significant computational challenges. Furthermore, since our approach does not require a customized simulation procedure for enzymatic reactions, it allows biochemical systems comprising such reactions along with other types of reactions to still take advantage of utilizing general stochastic simulation tools for the standard Gillespie stochastic simulation algorithm.

Future work includes comprehensive analysis of the PPTA errors under various conditions and analysis of efficiency-versus-accuracy among the PPTA and other approximations such as the QSSA based on the initial conditions, that is, a static and systematic approach to determine when to use the PPTA.

References

1. Arkin, A., Ross, J., McAdams, H.: Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected escherichia coli cells. *Genetics* **149** (1998) 1633–1648
2. Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S.: Stochastic gene expression in a single cell. *Science* **297** (2002) 1183–1186
3. Rao, C.V., Wolf, D.M., Arkin, A.P.: Control, exploitation and tolerance of intracellular noise. *Nature* **420** (2002) 231–238
4. Samoilov, M., Plyasunov, S., Arkin, A.P.: Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. *Proceedings of the National Academy of Sciences US* **102**(7) (2005) 2310–5

5. McAdams, H.H., Arkin, A.: Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences USA* **94**(3) (1997) 814–819
6. Pedraza, J.M., van Oudenaarden, A.: Noise Propagation in Gene Networks. *Science* **307**(5717) (2005) 1965–1969
7. Cai, L., Friedman, N., Xie, X.S.: Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**(7082) (2006) 358–362
8. Newman, J.R.S., Ghaemmighami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., Weissman, J.S.: Single-cell proteomic analysis of *s. cerevisiae* reveals the architecture of biological noise. *Nature* **441**(7095) (2006) 840–846
9. Gillespie, D.T.: *Markov Processes An Introduction for Physical Scientists*. Academic Press, Inc. (1992)
10. Gillespie, D.: The chemical langevin equation. *J. Chem. Phys.* **113**(1) (2000)
11. Gillespie, D.T.: 5.11. In: *Handbook of Materials Modeling*. Springer (2005) 1735–1752
12. Gillespie, D.T.: A rigorous derivation of the chemical master equation. *Physica A* **188** (1992) 404–425
13. Achimescu, S., Lipan, O.: Signal propagation in nonlinear stochastic gene regulatory networks. *IEE Proc. Sys. Bio.* **153** (2006) 120–134
14. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **22** (1976) 403–434
15. Gibson, M., Bruck, J.: Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* **104** (2000) 1876–1889
16. Gillespie, D.T.: Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **115**(4) (2001) 1716–1733
17. Rathinam, M., Cao, Y., Petzold, L., Gillespie, D.: Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *J. Chem. Phys.* **119** (2003) p12784–94
18. Gillespie, D., Petzold, L.: Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.* **119** (2003)
19. Cao, Y., Gillespie, D., Petzold, L.: Avoiding negative populations in explicit tau leaping. *J. Chem. Phys.* **123** (2005)
20. Rao, C.V., Arkin, A.P.: Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm. *J. Phys. Chem.* **118**(11) (2003)
21. Cao, Y., Gillespie, D., Petzold, L.: Accelerated stochastic simulation of the stiff enzyme-substrate reaction. *J. Phys. Chem.* **123** (2005)
22. Cao, Y., Gillespie, D., Petzold, L.: The slow-scale stochastic simulation algorithm. *J. Chem. Phys.* **122** (2005)
23. Finney, A., Hucka, M.: *Systems Biology Markup Language (SBML) level 2: Structures and facilities for model definitions* (2003)
24. : BioSPICE. (<http://www.biospice.org/>)
25. Kuwahara, H., Myers, C., Barker, N., Samoilov, M., Arkin, A.: Automated abstraction methodology for genetic regulatory networks. *Trans. on Comput. Syst. Biol.* VI (2006)