A New Approximate Method for the Stochastic Simulation of Chemical Systems:

The Representative Reaction Approach

Shantanu Kadam† and Kumar Vanka‡*

†Physical Chemistry Division, National Chemical Laboratory, Dr. Homi Bhabha Road,
Pashan, Pune, Maharashtra – 411 008, India

*Corresponding author. E-mail: k.vanka@ncl.res.in

Abstract:

We have developed two new approximate methods for stochastically simulating chemical systems. The methods are based on the idea of representing all the reactions in the chemical system by a single reaction, i.e., by the “representative reaction approach” (RRA). Discussed in the paper are the concepts underlying the new methods along with flowchart with all the steps required for their implementation. It is shown that the two RRA methods{with the reaction $2A \rightarrow B$ as the representative reaction (RR)} perform creditably with regard to accuracy and computational efficiency, in comparison to the exact stochastic simulation algorithm (SSA) developed by Daniel Gillespie and are able to successfully reproduce at least the first two moments of the probability distribution of each species in the systems studied. As such, the RRA methods represent a promising new approach for stochastically simulating chemical systems.
Introduction

The study of the kinetics of chemical systems has traditionally involved a Master Equation\textsuperscript{12}: a set of coupled ordinary differential equations (reaction rate equations) describing the time evolution of the concentration of the different chemical species reacting in the system. In order to solve such differential equations numerically, with a given set of rate constants and initial concentrations, some integrators\textsuperscript{3} have been used to calculate the time profiles of the concentrations. An alternative to this deterministic approach has been put forward by Doob\textsuperscript{4,5} and Gillespie\textsuperscript{6,7} whose seminal work focused on a stochastic approach, more specifically, the Kinetic Monte Carlo\textsuperscript{8} based stochastic simulation algorithm (SSA). Unlike the numerical algorithms, the SSA does not approximate the time increments by specific finite time steps and also takes into account the fluctuations in the system.

Since its introduction, the SSA\textsuperscript{4,6} has become popular for the study of chemical kinetics of different systems, especially biological systems that involve genetic regulatory networks and cellular processes.\textsuperscript{9-12} However, due to the occurrence of a single reaction event in each time increment, the practical application of the SSA is severely limited with respect to the time scale and the molecular populations of the chemical systems that it can effectively simulate. Several approximate methods have therefore been proposed so as to speed up the SSA. These include the Poisson $\tau$ -leap method,\textsuperscript{13} the midpoint $\tau$ -leap method,\textsuperscript{13} the implicit $\tau$ -leap method,\textsuperscript{14} the Poisson Runge-Kutta methods\textsuperscript{15}, the multinomial $\tau$ -leap method\textsuperscript{16}, efficient step size selection for the tau leaping\textsuperscript{17} and the binomial $\tau$ -leap methods.\textsuperscript{18,19} Apart from these, other attempts have also been made to reduce the computational load of the SSA: He et. al have used a hybrid Monte Carlo
algorithm for polymerization reaction kinetics;\textsuperscript{20} Gibson and Bruck have modified the first reaction method (similar to the SSA) such that unused reaction times could be refined for reuse;\textsuperscript{21} Rao and Arkin applied the quasi steady state assumption to the subset of fast reactions in the system in order to reduce the computational time to numerically simulate the systems\textsuperscript{22} and Haseltine and Rawlings have tried to improve the computational efficiency by partitioning the system into subsets of fast and slow reactions.\textsuperscript{23}

Of these, the most appealing in terms of (relative) simplicity and effectiveness is Gillespie’s approximate Poisson $\tau$ -leap method\textsuperscript{13} (henceforth referred to in the manuscript as GASA: Gillespie’s approximate stochastic algorithm). The principal idea behind this method is that instead of executing a single reaction in every time increment and changing the molecular population accordingly (as in the SSA), a larger leap “$\tau$” in time is taken, during which all of the reactions in the system are allowed to occur, the number of occurrences of each being determined with the aid of random numbers selected from the Poisson Distribution. The size of the $\tau$ leap is determined from the “Leap Condition” which Gillespie defines as the necessary requirement that the propensity functions (the product of the rate constants with the number of reactant combinations) of the reactions not change appreciably in value as a result of the leap.

GASA has proved successful in accelerating the simulations of several different types of chemical systems, in comparison to the SSA, while also replicating at least the first two moments of the probability distribution of the species with time. However, for chemical systems where the species amounts fluctuate significantly, or cases where the species amounts vary by different orders of magnitude, GASA has been found wanting in
terms of accurately replicating the changes in species amounts with time. Modifications have been proposed to the GASA method in order to improve its reliability, such as the Gillespie-Petzold (G-P) method\textsuperscript{28}, but, to date, most attempts at improvement have also led to increasing complexity of the algorithms employed. A method that can retain the simplicity of GASA, whilst also being able to reliably and efficiently simulate complex and challenging chemical systems, would therefore be highly desirable. Our objective in this manuscript is to propose just such a method.

This paper is organized as follows: (i) first, we have discussed the theoretical basis for our proposed method, which we have termed the representative reaction approach (RRA), where we have discussed the leap condition for the RRA as well as the choice of the most appropriate RR, along with a flowchart describing the steps in the algorithm; then we have tested our two proposed RRA methods, RRA-$\tau$ and RRA-$N$, on (ii) a model system of chemical reactions, on (iii) a more complicated system with chemical oscillations, on (iv) a system having reacting species fluctuating by different orders of magnitude, and finally (v) we have provided an evaluation of our method and presented conclusions.

**Methodology**

**Chief concepts and notations**

This subsection discusses in brief some of the notations and concepts that have been employed by Gillespie\textsuperscript{13} and co-workers\textsuperscript{18,19,24,25} in developing the approximate stochastic simulation methods. In order to study the evolution of molecular numbers, a well-mixed reacting system of $N$ molecular species $S_1, S_2, \ldots, S_N$ is considered. Let
$X_i(t)$ denote the number of species of $S_i$ at time $t$. This entire mixture of chemical species interacts inside some fixed volume $\Omega$ at a constant temperature through reaction channels $R_1, R_2, \ldots, R_M$. For each reaction channel $R_i (i = 1, \ldots, M)$, a propensity function $a_i(x)$ is defined which, along with $a_i(x)dt^{2.426}$, gives us the probability that the $R_i$ reaction will take place in the infinitesimal time interval $t, t + dt$. This propensity function $a_i(x)$ is the product of the rate constant with the number of reactant combinations for the given reaction. For a reaction of the type $X_1 + X_2 \rightarrow 2X_1$, $a_i(x) = c_jx_1x_2$, where $c_j$ is the “specific reaction probability rate constant” for the reaction, being algebraically related to the conventional deterministic rate constant $k_j$ by $c_j = k_j/(N_A * \Omega)$; where $k_j$ are in the units of mole inverse second inverse and $\Omega$ in litres. $N_A$ is the Avogadro’s number. $x_1$ and $x_2$ are the amounts of the reactants $X_1$ and $X_2$. For a reaction of the type $2X_1 \rightarrow 2X_2$, $a_i(x) = c_2x_1(x_1-1)/2$, where $c_2$ is the specific reaction probability rate constant for the reaction, with $c_2 = \left(\frac{2*k_2}{N_A * \Omega}\right)$.

**Our Approach**

For any given step $j$ during the simulation, our approach is to first determine the expected number of reactions that would take place for that step $j$ for the entire chemical system comprising all of the individual reactions. *This is achieved by representing the entire chemical system as a single, representative reaction, and then determining the expected number of reactions for that representative reaction (RR).* Our method is therefore termed the representative reaction approach (RRA). The propensity
function of the RR is taken to be $a_0(x)$: the sum of the propensity functions of all the individual reactions, while $C_0$—the specific reaction probability rate constant for RR, is determined as $C_0 = \sum_{i=1}^{N} \left( \frac{a_i(x)}{a_0(x)} \right) c_i$ : the weighted average of all the $c_i$s in the system. For any individual reaction, assuming a Poisson Distribution for the possible number of occurrences of that reaction, the expected number of occurrences, $n_i$, in any given step is equal to $a_i(x)r$.\textsuperscript{13} Now, one can either determine the value of $\tau$ for that particular step by (i) evaluating $\tau$ for the representative reaction, or alternatively, (ii) one can evaluate $N_0$—the total number of reactions taking place for that step for the representative reaction and then determine the value of $\tau$ for that step from $\tau = N_0 / a_0(x)$.

Once $\tau$ is determined for that particular step, the expected number of occurrences for the $i$th reaction is determined from $n_i = a_i(x)r$.\textsuperscript{13} This value $a_i(x)r$ is the variable used in “poisev” the computer algorithm used to generate the corresponding Poisson random number\textsuperscript{27} $k_i$. $k_i$ is the number of times the $i$th reaction will occur in that particular step $j$. Thus, the values of $k_i$ corresponding to all the reactions can be determined for this $j$th step and the amounts of reactants for each reaction adjusted accordingly to afford the propensity functions $a_i(x)$ for the next, $j+1$th, step, where the process of determining the $k$s is carried out in exactly the same fashion.

**The Leap Condition for the RRA**

Detailed in the previous subsection are the two possible ways, indicated as (i) and (ii), for determining the $k$s for each of the reactions for a specific step, based on the approach of
considering the entire system as being represented by a single reaction. However, the question that has not been answered yet is: how does one determine either the \( \tau \) for the method (i) or \( N_0 \) for the method indicated in (ii)? We propose that \( \tau \) or \( N_0 \) be chosen in a manner such that it satisfies the “Leap Condition”\(^{13}\) for RR: that the propensity (in this case, \( a_0(x) \)) of RR is not altered by an appreciable extent by the size of the jump, i.e., by the value chosen for \( \tau \) or \( N_0 \).

For method (i), henceforth, referred to as the RRA-\( \tau \) method, one can employ the Leap Condition criteria established by Gillespie\(^{13}\) to determine the change in \( \tau \) for the single reaction chosen as the representative reaction. Henceforth, this approach will be called the RRA-\( \tau \) method. For single reactions, the Gillespie Leap Condition criteria for choosing \( \tau \) becomes a simple expression and thus easy to implement. Examples of such expressions will be shown in the next sub-section, when discussing the choice of the most appropriate RR.

For method (ii), henceforth, referred to as the RRA-\( N \) method, the approach employed will be to bound the change in the value of \( a_0(x) \) – the propensity function for the RR, as follows:

\[
|a_0(x + \lambda) - a_0(x)| \leq \varepsilon a_0(x)
\]

or

\[
|\Delta a_0(x)| \leq \varepsilon a_0(x)
\]

where, \( \Delta a_0(x) = a_0(x + \lambda) - a_0(x) \)
Here, $\varepsilon$ is a parameter that would remain constant throughout the simulation and $\lambda$ is the amount by which we change the state of the system. Now, it has been shown in the past that the term $\Delta a_0(x)$ can be approximated by a first order Taylor series expansion as:

$$\Delta a_0(x) = \lambda \nabla a_0(x)$$

Hence, $\lambda \nabla a_0(x) \leq \varepsilon a_0(x)$

A further approximation is added at this point to further accelerate the system. This can be done by multiplying the function on the right of the inequality above ($\varepsilon a_0(x)$) by a factor. This is the equivalent of dividing $\Delta a_0(x)$ by the same factor, and is therefore justified in that it will lead to a reduction in the $\Delta a_0(x)$ value and thus further comply with the Leap Condition. This factor is chosen to be 16, so that we can define a new parameter $\varepsilon'$, where $\varepsilon' = 16 \varepsilon$. The somewhat arbitrary nature of the choice of the value of “16” for the factor will be understood in the context of the essential tunability of the values of $\varepsilon$ and $\varepsilon'$ that will be discussed in the next section: it is found that the algorithm produces the best results in terms of accuracy and speed when such a factor is employed.

Hence, after incorporating the factor of 16, we get the new equation:

$$\lambda \nabla a_0(x) \leq \varepsilon'a_0(x)$$

where $\varepsilon' = 16 \varepsilon$.

By using the eq. (1) above, and using the expression for $\nabla a_0(x)$ for the chosen RR one can determine the value of $N_0$, and thereby $\tau$, for a given step for that specific RR.
Examples of the values of $N_0$ thus obtained for different RR cases will be discussed in the next subsection.

**What is the most appropriate representative reaction (RR)?**

Described in this subsection are the expressions that can be derived for the methods (i) RRA-$\tau$ and (ii) RRA-$N$ for different representative reactions (RRs).

**$A \rightarrow B$ as the **Representative Reaction (RR):**

The unimolecular reaction $A \rightarrow B$ is the simplest possible choice for the RR, and therefore the one that is considered first.

(i) The RRA-$\tau$ method: As discussed in the previous subsection, for this method, the value of $\tau$ is determined from the Leap Condition for the RR (in this case $A \rightarrow B$), as established in Gillespie’s Approximate Stochastic Algorithm (GASA)$^{13}$. For $A \rightarrow B$, the expression for $\tau$ is:

$$\tau = \frac{\varepsilon}{C_0}$$  \hspace{1cm} (2)

$\varepsilon$ is a parameter that is kept constant throughout the simulation. The value of $C_0$ for the RR is determined, as discussed earlier from the weighted average of all the $c_i$ of the different reactions in the system: $C_0 = \sum_{i=1}^{M} \left( \frac{a_i(x)}{a_0(x)} \right) c_i$. This expression is indeed quite simple, but it suffers from the drawback that it only depends on $C_0$ and not on “$x_0$”; the amount of the hypothetical species “$A$”. As will be seen for the subsequent examples, $x_0$
appears in the denominator of the expression for $\tau$ for other RRs, and thus serves to modulate and reduce the value of $\tau$. With the absence of $x_0$ in eq. (2), the value of $\tau$ tends to be somewhat high, especially for the beginning few steps of the simulation. Thus the RR $A \rightarrow B$, despite its simplicity, is not a good choice for doing simulations with the RRA-$\tau$ method.

(ii) The RRA-$N$ method: For the RR $A \rightarrow B$, the expression for $\nabla a_0(x)$ is: $C_0$. Therefore, using the eq. (1), one gets the following expression for $N_0$:

$$N_0 = \frac{\epsilon a_0(x)}{C_0} \quad (3)$$

Like for the RRA-$\tau$ method, this expression suffers from the absence of “$x_0$”. For other RRs, as will be shown below, $x_0$ appears in the denominator in the right hand side of the eq. (3), and serves to regulate the value of $N_0$. The absence of $x_0$ in eq. (3) leads to somewhat large changes in $N_0$ for the beginning few steps of the simulations for any given chemical system. Thus, as for the RRA-$\tau$ method, $A \rightarrow B$ is not a good choice as the RR for the RRA-$N$ method.

$2A \rightarrow B$ as the Representative Reaction (RR):

(i) The RRA-$\tau$ method: Since, in case of this RR the propensity function is given by

$$a_0 = \left(\frac{x_0(x_0 - 1)}{2!}\right) C_0 \quad (4)$$
to get an expression for $x_0$ we solve the resulting quadratic equation which yields two roots for $x_0$, of which we choose the root:

$$x_0 = \frac{C_0 + \sqrt{C_0^2 + 8a_0C_0}}{2C_0}$$

in order to avoid the possibility of negative values for $x_0$.

Here, the value of $\tau$, determined from the GASA Leap Condition, is:

$$\tau = \frac{\varepsilon}{\left| C_0(2x_0 - 1) \right|}$$

This expression gives rise to acceptable values of $\tau$, because of the presence of $x_0$ in the denominator.

As will be shown in the examples discussed in the next section, the use of the RRA-$\tau$ method with $2A \rightarrow B$ as the RR provides results which equal or improve upon the accuracy and reliability of the other approximate methods.

(ii) The RRA-$N$ method: The value of $N_0$, determined from this method, using the eq. (1), is:

$$N_0 = \frac{\varepsilon a_0(x)}{C_0(2x_0 - 1)}$$
where, the $x_0$ is calculated as discussed earlier. As with the RRA-$\tau$ method, this value of $N_0$ provides results that are quite acceptable, as evidenced by the results for the examples discussed in the subsequent sections of the manuscript.

3$A$ $\rightarrow$ $B$ as the Representative Reaction (RR):

(i) The RRA-$\tau$ method:

Now, as before, considering “$x_0$” as the number of reactant molecules for the hypothetical reactant species “$A$” for the RR 3$A$ $\rightarrow$ $B$, we have the propensity,

$$\alpha_0 = \left( \frac{x_0(x_0-1)(x_0-2)}{3!} \right) C_0$$

So to get an expression for $x_0$ for this case, one could use a subroutine to calculate the cubic roots but it makes our algorithm computationally more complicated, losing the simplicity which is one of the criteria for choosing the RR. Instead of that, we can neglect the lower order terms for $x_0$ in eq. (8) to get a final expression for $x_0$. The expression of $\tau$ for this RR is derived to be:

$$\tau = \frac{\epsilon}{-3C_0(3x_0^2 - 6x_0 + 2) / 6}$$

This expression for $\tau$ leads to a problem that is the opposite of the one faced when using $A$ $\rightarrow$ $B$: here, the quadratic dependence of $\tau$ on $x_0$ in the denominator leads to values that are too small in size as compared to values obtained for the RR 2$A$ $\rightarrow$ $B$. 
In other words, not much acceleration over the exact stochastic simulation method (SSA) is observed in this case. The reliability of the algorithm with this RR is thus not in doubt, but the essential purpose of making an accelerated algorithm is lost for this case.

A similar problem occurs for the cases $4A \rightarrow B$, $5A \rightarrow B$ and other higher order versions of this type of RR. The term $x_0$ begins to appear in higher and higher orders in the denominator, thereby making the size of the jumps smaller and smaller, and reducing the efficiency of the accelerated algorithm.

(ii) The RRA-$N$ method: The value of $N_0$, determined from this method, using the eq.(1), is:

$$N_0 = \frac{2e \, a_0(x)}{C_0(3x_0^2 - 6x_0 + 2)}$$

where, the $x_0$ is calculated as discussed in the earlier method. In this case too, as for the RRA-$\tau$ method, the presence of $x_0^2$ in the denominator leads to values of $N_0$ that are smaller than those obtained from the RR, $2A \rightarrow B$. Hence the efficiency of the method is reduced. Again, as for the for the RRA-$\tau$ method, higher powers of $x_0$ appear in the denominator for the cases $4A \rightarrow B$, $5A \rightarrow B$ and so on, thereby making the algorithms even less efficient, and thus defeating the purpose of making an accelerated algorithm.

$A + B \rightarrow C$ as the Representative Reaction (RR):

In case of this particular RR, we have “$x_0$” as the number of the reactant molecules for the hypothetical reactant species “A” and “$y_0$” as the number of the reactant
molecules for the hypothetical reactant species “B”. Thus, the corresponding propensity function is,

\[ a_0 = C_0 x_0 y_0 \]  \hspace{1cm} (11)

Even though the expression looks simple, this is a case of a single equation with two unknown variables, namely \( x_0 \) and \( y_0 \). Therefore, the values of \( x_0 \) and \( y_0 \) cannot be determined independently in terms of the known \( a_0(x) \) and \( C_0 \) values.

Naturally, a similar problem also occurs if we take our RR to be \( A + 2B \rightarrow C \) or \( 2A + B \rightarrow C \) or \( A + B + C \rightarrow D \) or any other variant of a bimolecular or tri-molecular or any other multi-molecular reaction.

Our analysis thus indicates that the most appropriate RR for doing the simulation, the one that combines simplicity and efficiency most effectively, is the RR \( 2A \rightarrow B \). This is therefore our chosen reaction for the representative reaction approach. Its efficiency will be revealed in the three illustrative examples that are described later in the paper.

**Steps for the Implementation of the RRA-\( \tau \) and RRA-N methods**

Based on the discussion above, the implementation steps for the RRA methods, for the RR: \( 2A \rightarrow B \), are outlined as follows:

Step 1: input the initial number of species, rate constants of the constituent reactions;

initialize the counters and the random number generators to a seed value.

Step 2: calculate the propensity functions : \( a_1, a_2, \ldots, a_m \)
the sum of the propensity functions: \( a_0(x) = \sum_{i=1}^{M} a_i(x) \)

the weighted rate constant: \( C_0 = \sum_{j=1}^{M} \left( \frac{a_j(x)}{a_0(x)} \right) c_j \)

Step 3: calculate the total number of species present: \( x_0 = \frac{C_0 + \sqrt{C_0^2 + 8a_0C_0}}{2C_0} \)

Step 4: calculate the tau step:

(i) For RRA-\( \tau \) method: \( \tau = \frac{\varepsilon}{|C_0(2x_0 - 1)|} \)

if the tau step is less than or equal to \( \frac{2}{a_0(x)} \) then perform the SSA

otherwise continue with the RRA-\( \tau \) method

(ii) For RRA-N method: \( \tau = \frac{N_0}{a_0(x)} \)

where, the total number of reactions, \( N_0 = \frac{\varepsilon a_0(x)}{C_0(2x_0 - 1)} \)

if \( N_0 \) is less than or equal to 1 then perform the SSA otherwise continue with

RRA-N method

Step 5: calculate the expected number of occurrences, the \( n_i \)s, for the individual reactions \( n_i = a_i \tau \)
Step 6: use the Poisson Random Number Generator to find the $k_i$s, the actual number of occurrences for the individual reactions. $k_i = poidev(n_i, iseed)$

Step 7: make the necessary changes in the species population using the appropriate stoichiometric parameters and reaction numbers

Step 8: go to Step 2

This algorithm is further outlined in the Flowchart below:

Flowchart
Initialise: $t = 0$, all counters = 0
RNG to a seed value

Give: Initial values for the species
Rate constants of reactions

Calculate: the propensity functions
$a_i (i = 1, ..., M)$
total propensity: $a(x) = \sum_{i=1}^{M} a_i (x)$
weighted rate constant:
$C_i = \sum_{i=1}^{M} \frac{a_i (x)}{a(x)} c_i$

Calculate: hypothetical reactant species
$s_i = C_i + \sqrt{C_i (2s_i - 1)}$

Calculate: the tau step
For $RRA-\tau$ method: $\tau = \frac{\epsilon}{[-C_i (2s_i - 1)]}$
For $RRA-\mathcal{N}$ method: $\tau = \frac{N_i}{a_i (x)}$
with $N_i = \frac{\epsilon a_i (x)}{C_i (2s_i - 1)}$

Check if
For $RRA-\tau$ method: $\tau \leq \frac{2}{a_i (x)}$
For $RRA-\mathcal{N}$ method: $N_i \leq 1$

YES
Perform SSA

NO
Calculate: mean number of reactions $n_j$
Find: reactions from Poisson RNG $k_j$
Make time increment $t = t + \tau$

Make changes in species population
The Fortran 95 RRA-τ and the RRA-N codes that we have employed for the simulation of the “Four Reaction Model” example (discussed below) are provided. Also provided are the Fortran 95 codes of the GASA and G-P methods, as well as for the SSA.

**Optimization of the Values of the Parameters \( \varepsilon \) and \( \varepsilon' \)**

It was determined in the last section that the most appropriate choice of RR for both the RRA-τ and the RRA-N approaches is the reaction \( 2A \rightarrow B \). For this RR, the expressions derived for \( \tau \) and \( N_0 \) for the two methods depend on the parameters \( \varepsilon \) and \( \varepsilon' \) respectively. Discussed in this subsection is the choice of the most appropriate values of \( \varepsilon \) and \( \varepsilon' \) for the two methods. To this end, the two methods were tested on a “Four Reaction Model”: a chemical system that has been demonstrated to be successfully simulated by GASA. The system comprises of the following reactions:

\[
\begin{align*}
R_1 : X_1 & \xrightarrow{c_1} 0 \\
R_2 : X_1 + X_1 & \xrightarrow{c_2} X_2 \\
R_3 : X_2 & \xrightarrow{c_3} X_1 + X_1 \\
R_4 : X_2 & \xrightarrow{c_4} X_3
\end{align*}
\]

\[ (12) \]

Where \( X_1, X_2, X_3 \) are the species participating in this reaction system, with \( c_1, c_2, c_3, c_4 \) as the rate constants of the reactions in (12).

Table S11 given in the Supplementary Information provides the values of the different model parameters used in the simulation of the reaction system (12).

The RRA-τ and the RRA-N methods were tested for this Four Reaction Model system for different values of \( \varepsilon \) and \( \varepsilon' \) respectively. The results are discussed below.
(i) The RRA-$\tau$ method: Shown in the five figures S1 to S5 in the Supplementary Information file are the results obtained by the application of the RRA-$\tau$ algorithm on the Four Reaction Model, as well as the comparison to the results obtained by the exact stochastic simulation algorithm (SSA). The simulations were run ten times for both the SSA and the RRA, using different seed values for the random number generators, and the mean values of the chemical species were calculated. The mean values obtained for the species $X_1$ and $X_2$ with time, using the SSA and the RRA-$\tau$, as well as the values of the coefficient of variation (CV) for the species, for both the SSA and the RRA-$\tau$ are shown in the five figures. The values of $\varepsilon$ employed are as follows: 0.03, 0.09, 0.15, 0.2 and 0.3 for the curves shown in the figures S1, S2, S3, S4 and S5 respectively. The comparison of the mean and the CV of the results from the RRA-$\tau$ method with the SSA indicates that the alteration of the value of $\varepsilon$ from 0.09 to 0.3 leads to almost identical results and, as compared to the values obtained from SSA, to fairly accurate results. The accuracy of the results is seen to be marginally more accurate for $\varepsilon = 0.03$, but this comes at the cost of a slower simulation. Hence, keeping the balance of accuracy and efficiency in mind, the value chosen for $\varepsilon$ in evaluating the RRA-$\tau$ method and comparing it to SSA and the other approximate methods GASA and Gillespie-Petzold (G-P)$^{28}$ is 0.2.

(ii) The RRA-N method: The figures S6 to S10 in the Supplementary Information file show the results of the simulations done for the Four Reaction Model with different values of $\varepsilon'$ employed for the RRA-N method. The values of $\varepsilon'$ employed are as follows: 0.48, 0.80, 1.28, 1.6 and 3.2 for the curves shown in the figures S6, S7, S8, S9 and S10 respectively. As for the RRA-$\tau$ method, both the mean and the CV of the probability
distributions for the species $X_1$ and $X_2$ were plotted for the different cases, and the values are compared to the SSA. The figures indicate that there is little difference in the curves upon changing the value of $\varepsilon^\prime$ from 0.48 to 3.2. However, what is affected is the size of the first jump: from $\varepsilon^\prime = 1.6$ onwards, it was seen that the change in the values of $X_1$ and $X_2$ was quite significant for the first step. Hence, in order to balance efficiency and reliability of the method, the optimal value of $\varepsilon^\prime$ for the RRA-N method was chosen to be 1.28.

Hence the optimized parameters for the two methods are: $\varepsilon = 0.2$ for the RRA-\(\tau\) method and $\varepsilon^\prime = 0.08$ for the RRA-N method. The efficiency and reliability for the two RRA methods with these optimized values of $\varepsilon$ and $\varepsilon^\prime$ will be evaluated for three separate examples discussed in the next few sections.

**The Four Reaction Model: Comparison of RRA-\(\tau\) and RRA-N to other Accelerated Methods**

The Four Reaction Model system was discussed in the previous section in the context of the optimization of the values of $\varepsilon$ and $\varepsilon^\prime$ for the RRA-\(\tau\) and the RRA-N methods respectively. Discussed in this section is the comparison, for the same Four Reaction Model, of the RRA-\(\tau\) and the RRA-N methods with two other accelerated methods: Gillespie’s Approximate Stochastic Algorithm (GASA) and the Gillespie-Petzold (G-P) method. It is noted here that both GASA and G-P also employ the parameter $\varepsilon$, and the value that the methods recommend is 0.03. For all the cases discussed here onwards, this standard value of $\varepsilon = 0.03$ has been employed for all the simulations using these two approximate methods.
Figure 1 below shows the comparison of the mean and the CV of the probability distributions for the species $X_1$ and $X_2$ obtained from running simulations with the four different methods: the exact SSA (shown in blue), GASA (shown in green), G-P (shown in magenta) and the RRA-$\tau$ (shown in red). As the figure indicates, all three accelerated methods work quite well in predicting the mean of the probability distributions for the two species $X_1$ and $X_2$. However, there is a distinct difference in performance when it comes to the second moment: it is seen from the CV curves Figures 1c-d that GASA provides results that are less accurate than the RRA-$\tau$ method and for the G-P. A similar result is obtained when comparing the RRA-$N$ method to GASA and G-P as Figures 2a-d indicate. However, a comparison of the average time taken to run the simulations for the three different algorithms indicates that GASA and G-P hold a distinct advantage over the two RRA methods. The average CPU time taken when employing the different methods is shown in Table 1 (which also collects the corresponding information for the other examples that have been studied). For the case of the Four Reaction Model, it was determined that the average time taken to do a GASA simulation was 0.036 seconds, the G-P: 0.053 seconds, the RRA-$N$: 0.300 seconds, the RRA-$\tau$: 1.622 seconds and the SSA: 11.439 seconds. Thus, while the two RRA methods do provide a distinct acceleration in comparison to the SSA, as evidenced by the CPU time taken, they are not as efficient as GASA and G-P. The explanation for this is the fact that, while the two RRA methods do accelerate the system, they take smaller jumps in time than GASA and the G-P methods. This is because of the regulating variable “$x_0$” in the denominator for the RR $2A\rightarrow B$, which serves to decrease the size of the jump for the RRA methods. While this appears to be a disadvantage, in terms of efficiency, for relatively simple systems such as the Four
Reaction Model considered here, this is actually a significant advantage for more complicated systems, two examples of which will be discussed in the next two sections. As will be seen, the sacrifice of a minor loss in efficiency because of smaller step size is more than compensated by the gain in the accuracy of the simulations of the more complicated systems.

**Figure 1.** The trajectories of the means [(a) and (b)] and the CVs [(c) and (d)] for the probability distributions of the species $X(1)$ and $X(2)$ using SSA (blue curve), GASA (green curve), G-P (magenta curve) and RRA-τ (red curve) for the case of the Four Reaction Model.
Figure 2. The trajectories of the means [(a) and (b)] and the CVs [(c) and (d)] for the probability distributions of the species X(1) and X(2) using SSA (blue curve), GASA (green curve), G-P (magenta curve) and RRA-N (red curve) for the case of the Four Reaction Model.

The Model of Oscillatory Reactions: The Oregonator Model

We show in this section the application of our RRA algorithms to the more complicated case of an oscillatory chemical system\(^7\). To correctly simulate oscillatory chemical systems is a significant challenge for an approximate accelerated method, because of the fluctuations in the species amounts with small changes in time. The model discussed here is the “Oregonator” which is a theoretical model for autocatalytic reactions. The system consists of the following set of reactions:
\[ \begin{align*}
R_1: \overline{X}_1 + Y_2 & \xrightarrow{c_1} Y_1 \\
R_2: Y_1 + Y_2 & \xrightarrow{c_2} Z_1 \\
R_3: \overline{X}_2 + Y_1 & \xrightarrow{c_3} 2Y_1 + Y_3 \\
R_4: 2Y_1 & \xrightarrow{c_4} Z_2 \\
R_5: \overline{X}_3 + Y_3 & \xrightarrow{c_5} Y_2
\end{align*} \] (13)

Where \( Y_1, Y_2, Y_3 \) are the species participating in this reaction system, while \( \overline{X}_1, \overline{X}_2, \overline{X}_3 \) indicates that the molecular population level of these species is assumed to remain constant with \( c_1, c_2, c_3, c_4, c_5 \) as the rate constants of the reactions in (13).

Table S12 given in the Supplementary Information provides the values of the different model parameters used in the simulation of the reaction system (13).

The Oregonator has been investigated by the SSA, GASA, G-P and the two RRA methods. Shown in the Figures 3a-c is the comparison of the mean values obtained for the species \( Y_1, Y_2 \) and \( Y_3 \) with time, using SSA, GASA, G-P and RRA-\( \tau \). The CV values obtained for the \( Y_1, Y_2 \) and \( Y_3 \) are shown in Figure 4. Likewise, the Figures 5 and 6 show the mean and CV values respectively for SSA, GASA, G-P and RRA-\( N \).
Figure 3. The trajectories of the means [(a), (b), (c)] for the probability distributions of the species Y(1), Y(2) and Y(3) using SSA (blue curve), GASA (green curve), G-P (magenta curve) and RRA-τ (red curve) for the case of the Oregonator Model.
Figure 4. The trajectories of the CVs [(a), (b), (c)] for the probability distributions of the species Y(1), Y(2) and Y(3) using SSA (blue curve), GASAA (green curve), G-P (magenta curve) and RRA-τ (red curve) for the case of the Oregonator Model.

Figure 5. The trajectories of the means [(a), (b), (c)] for the probability distributions of the species Y(1), Y(2) and Y(3) using SSA (blue curve), GASAA (green curve), G-P (magenta curve) and RRA-N (red curve) for the case of the Oregonator Model.
Figure 6. The trajectories of the CVs [(a), (b), (c)] for the probability distributions of the species Y(1), Y(2) and Y(3) using SSA (blue curve), GASA (green curve), G-P (magenta curve) and RRA-N (red curve) for the case of the Oregonator Model.

As is clear from the four figures, the two RRA methods don’t give an exact correspondence to the SSA values, which is indeed very difficult to achieve for the challenging oscillatory systems, but it is also clear that the two RRA methods perform significantly better than GASA and G-P, not only in replicating the mean values but even in replicating the second moment (CV) of the probability distributions for the three species in question. In the case of GASA, it is found that the curves showing the mean and CV of the probability distributions for the different species appear to be laterally displaced from the other curves. This is because it was found that the first $\tau$ value obtained from the GASA simulations was very large: 5.6969 units, as compared to the
first step for the other methods: G-P: 1.2079E-2 units, RRA-$\tau$: 2.7177E-5 units, RRA-$N$: 1.7393E-4 units and SSA: 1.5315E-6 units. Hence, though the subsequent steps for GASA are smaller in size, the first, large jump violates the Leap Condition and makes GASA inappropriate for simulating this class of chemical systems.

With regard to the average time taken for the simulations, the values in seconds are GASA: 0.020, G-P: 1.494, RRA-$N$: 1.492, RRA-$\tau$: 8.893, and SSA: 52.524 (see Table 1). GASA, of course, is quite fast in comparison to the others, but, as discussed above, it is also completely unreliable. Among the rest, it is seen that the RRA-$N$ method performs as efficiently as G-P in terms of time taken, and as the curves indicate, is significantly more reliable in replicating both the first and second moments of the probability distributions for the different species.

Overall, the comparison of efficiency and accuracy among the different algorithms for the challenging oscillatory model example provides definite evidence of the efficiency and reliability of the RRA methods, especially the RRA-$N$ method.

**Example Involving Species Fluctuating by Different Orders of Magnitude**

In this section, we discuss the simulation of a chemical system where the chemical species fluctuate by different orders. As mentioned in the Introduction, this belongs to the class of problems where accelerated methods have been found wanting in terms of combining accuracy and efficiency. The system discussed in this section has been studied and discussed previously by Haseltine et al. They observed that the three principal species: “template”, “genome” and “struct” in the system varied by different orders of
magnitude as the reactions proceeded over time: the amount of the species template fluctuates between 5 to 25, genome between 0 to 200, and struct between 100 to 12000.

For this system, the simulations have been done with the SSA, GASA, G-P as well as with the two RRA methods. For each case, the simulations have been repeated ten times, and the values for the mean as well as for the coefficient of variation, CV, have been calculated.

This model, the reactions for which are discussed below, shows the mechanism of the infection of a cell by a virus. The reactions are:

\[
\begin{align*}
R_1 &: \text{nucleotides} \xrightarrow{\text{template}} \text{genome} \\
R_2 &: \text{nucleotides} + \text{genome} \xrightarrow{\text{template}} \text{struct} \\
R_3 &: \text{nucleotides} + \text{aminoacids} \xrightarrow{\text{template}} \text{struct} \\
R_4 &: \text{template} \xrightarrow{} \text{degraded} \\
R_5 &: \text{struct} \xrightarrow{} \text{secreted} / \text{degraded} \\
R_6 &: \text{genome} + \text{struct} \xrightarrow{} \text{secreted}
\end{align*}
\]

(14)

where “genome” and “template” are the nucleic acids and “struct” is the viral structural protein. In this system, the molecular population number of the nucleotides and amino acids are assumed to remain constant and the template is considered to act as a catalyst for the reactions \(R_1\) and \(R_3\).

Table S13 given in the Supplementary Information provides the values of the different model parameters used in the simulation of the reaction system (14).

The Figures 7a-c show the change in the mean values for the three species: template, genome and struct with change in time. Each figure has four curves,
corresponding to the SSA (blue), GAS (green), the G-P (magenta) and the RRA-$\tau$ (red).

Figure 8 shows the corresponding CV values for the three different species as obtained from the four methods. Likewise, the Figures 9 and 10 show the comparison of the SSA, GAS and G-P with the RRA-N method.

**Figure 7.** The trajectories of the means [(a), (b), (c)] for the probability distributions of the species Template, Genome and Struct using SSA (blue curve), GAS (green curve), G-P (magenta curve) and RRA-$\tau$ (red curve) for the case of the Viral Infection Model.
Figure 8. The trajectories of the CVs [(a), (b), (c)] for the probability distributions of the species Template, Genome and Struct using SSA (blue curve), GASA (green curve), G-P (magenta curve) and RRA- $\tau$ (red curve) for the case of the Viral Infection Model.
Figure 9. The trajectories of the means [(a), (b), (c)] for the probability distributions of the species Template, Genome and Struct using SSA (blue curve), GASA (green curve), G-P (magenta curve) and RRA-N (red curve) for the case of the Viral Infection Model.

![Figure 9](image)

Figure 10. The trajectories of the CVs [(a), (b), (c)] for the probability distributions of the species Template, Genome and Struct using SSA (blue curve), GASA (green curve), G-P (magenta curve) and RRA-N (red curve) for the case of the Viral Infection Model.

As the figures indicate, the two RRA methods outstrip GASA and G-P in replicating both the mean and the CV of the three species in question. With regard to the average time taken, the values in seconds are as follows: GASA: 0.074, G-P: 0.549, RRA-N: 72.509, SSA: 140.090 and RRA-τ: 147.750. Hence, while RRA-τ does well to replicate the results of the SSA, it is unable to accelerate the system. However, GASA and G-P, while being much faster than the RRA methods, provide results for both the mean and the CV for the three species that are quite inaccurate. It is only the RRA-N method that succeeds in providing accuracy while also managing to accelerate the simulation of the system at
least by a factor of 2. Hence, for this difficult and very relevant set of problems, it is demonstrated that the RRA methods, especially the RRA-\( N \) method, can be effective and reliable substitutes for the exact SSA algorithm.

The following table provides the averaged values of the CPU time (in seconds) taken by different simulation methods (SSA, GASA, GP, RRA-\( \tau \) and RRA-\( N \) ) for the chemical systems (12), (13) and (14) discussed above.

<table>
<thead>
<tr>
<th>Chemical Systems</th>
<th>SSA</th>
<th>GASA</th>
<th>GP</th>
<th>RRA-( \text{TAU} )</th>
<th>RRA-( N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Reaction Model</td>
<td>11.439</td>
<td>0.036</td>
<td>0.053</td>
<td>1.622</td>
<td>0.300</td>
</tr>
<tr>
<td>Oregonator Model</td>
<td>52.524</td>
<td>0.020</td>
<td>1.494</td>
<td>8.893</td>
<td>1.492</td>
</tr>
<tr>
<td>Intracellular Viral Infection Model</td>
<td>140.090</td>
<td>0.074</td>
<td>0.549</td>
<td>147.750</td>
<td>72.509</td>
</tr>
</tbody>
</table>

**Table 1.** The averaged values of the CPU time (in seconds) taken by different simulation methods (SSA, GASA, GP, RRA-\( \tau \) and RRA-\( N \) ) for the different chemical systems discussed in the text.

Overall, what the application of the two RRA methods to the three examples indicates is that the RRA methods, in general, take more steps, and thus slightly longer times, in comparison to the other accelerated methods discussed here: GASA and G-P. However, this, in turn, leads to a considerable improvement in the accuracy and reliability of the methods. The RRA method that combines the qualities of reliability and efficiency the
best is found to be the RRA-N method. This, combined with the overall simplicity of the
approach and the resultant algorithm, makes it a promising method for stochastically
simulating different types of chemical systems.

Conclusions

In the work described in this manuscript, we have endeavored to develop new
approximate methods for conducting stochastic simulations on chemical systems. The
new methods are based on the concept of treating $2A \rightarrow B$ as a single representative
reaction for the system: the “representative reaction approach” (RRA). Two methods
have been proposed based on this approach: RRA-$\tau$ and the RRA-N, and the application
of these methods to three different chemical systems indicates that the two methods,
especially the RRA-N, perform creditably in combining accuracy and efficiency in
simulating the mean and the CV of the probability distributions of the different species,
in comparison to other approximate methods such as Gillespie’s Approximate Stochastic
Algorithm (GASA) and the Gillespie-Petzold method (G-P). It is to be noted that the new
approximate methods take smaller jumps in time than the other approximate methods that
have been proposed. However the subsequent loss in efficiency is more than
compensated for by the concurrent increase in accuracy of the simulations. This is
especially relevant when one wants to simulate more challenging and complicated
systems, as demonstrated in the manuscript for an oscillatory model system as well as for
a system where the species concentrations vary by different orders of magnitude.

The methods have the added virtue of giving rise to very simple and straightforward
algorithms. Indeed, as can be observed from the flowchart provided for the
implementation of the RRA methods, they are simpler algorithms than GASA, for which it is necessary to calculate all the partial derivatives of the propensity functions with respect to the different species present in the system and then to make a stoichiometric matrix to calculate the denominators corresponding to each reaction involved in the system;\textsuperscript{13} such calculations are unnecessary in our method.

In terms of potential drawbacks for our method, it should be mentioned that for reactions where species approach molecular populations close to zero, the RRA methods, like other approximate accelerated methods\textsuperscript{13,14,22,23} may begin to provide negative molecular populations. Work is currently in progress in order to address this issue.

Overall, it is clear that the RRA methods, especially the RRA-N method, provides a simple, easy way to successfully simulate a wide variety of different chemical systems over long periods of time, and, as such, should find wide-ranging applicability in the fields of chemistry and bio-chemistry.

**Supplementary Information Available:** The figures indicating the mean and CV values for the species distributions for the Four Reaction Model example, for different values of $\varepsilon$ and $\varepsilon'$ for both the RRA methods, are provided in the Supplementary Information file. Also provided are the Tables containing the values of parameters used for different simulation methods.

**Acknowledgements**
The authors acknowledge financial support from Department of Science and Technology (DST), India. We wish to thank the Centre of Excellence in Scientific Computing (COESC) at the National Chemical Laboratory (NCL), Pune, for providing computational facilities. We also wish to thank the reviewers of the paper for providing valuable suggestions that have improved the quality of the paper. We also wish to thank Dr. Chetan Gadgil for helpful discussions.

References