

Reliable kinetic Monte Carlo simulation based on random set sampling

Yan Wang

© Springer-Verlag Berlin Heidelberg 2013

Abstract Kinetic Monte Carlo (KMC) method has been widely used in simulating rare events such as chemical reactions or phase transitions. Yet lack of complete knowledge of transitions and the associated rates is one major challenge for accurate KMC predictions. In this paper, a reliable KMC (R-KMC) mechanism is proposed in which sampling is based on random sets instead of random numbers to improve the robustness of KMC results. In R-KMC, rates or propensities are interval estimates instead of precise numbers. A multi-event algorithm based on generalized interval probability is developed. The weak convergence of the multi-event algorithm towards the traditional KMC is demonstrated with a generalized Chapman–Kolmogorov equation.

Keywords Kinetic Monte Carlo · Uncertainty · Interval probability · Differential Chapman–Kolmogorov equation

1 Introduction

Compared to molecular dynamics (MD), atomic scale Kinetic Monte Carlo (KMC) (Chatterjee and Vlachos 2007) is more efficient in simulating the infrequent transition processes with longer times than thermal vibration. To simulate the rare events of transitions or reactions, several improvements of MD have been proposed to bridge

the gap of time scale and accelerate the simulation speed of rare events, such as by running multiple trajectories (Voter 1998), introducing bias potentials (Voter 1997), or increasing temperatures (Sörensen and Voter 2000). However, the inherent inefficiency of MD is that computational time is spent on trajectory prediction, which is not important for rare event simulations.

KMC does not simulate a system based on its continuous evolution along time as in MD. Instead, it defines a discrete set of states of the system (i.e. all possible configurations). It simulates state transitions between states which are triggered by events (also called processes in chemistry-oriented literature) that cause state changes. For instance, in the vapor deposition process of crystal growth shown in Fig. 1a, each “snapshot” of the system at a time where atoms are located in the space is a state of the KMC model. In Fig. 1b, the major events that change the system’s state include adsorption (particles in vapor are attracted to solid surface) with rate a_1 (rates indicate the frequencies of events), desorption (particles previously absorbed on the solid surface escape and are vaporized) with rate a_7 , surface diffusion (particles on the surface move to a different location) with rates a_2 to a_5 , and vertical diffusion (particles on the surface move into the solid) with rate a_6 . In chemical reactions, the numbers of different species naturally form the states of KMC models. Reactions themselves are the events.

During the KMC simulation, events are randomly selected to occur sequentially based on their respective probabilities of occurrence, which are exponential distributions in general. When an event occurs (or it is usually said that the event is fired), the system’s state is updated according to the nature of the event. At the same time, the system’s clock that keeps track of the current time of the system needs to be advanced by a certain period, which is

Communicated by V. Kreinovich.

Y. Wang (✉)
Woodruff School of Mechanical Engineering,
Georgia Institute of Technology, 813 Ferst Drive NW,
Atlanta, GA 30332-0405, USA
e-mail: yan.wang@me.gatech.edu
URL: <http://www.me.gatech.edu/~ywang/>

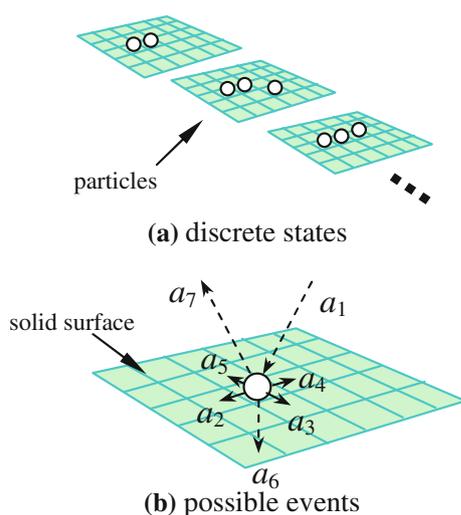


Fig. 1 Example states and events in vapor deposition process simulation

also randomly generated based on probabilistic principles. The clock provides information of how long the process lasts. Therefore, KMC essentially simulates state transitions or reactions. Clock advancement is not based on a fixed time period but directly determined by when the next event occurs. Trajectories of atoms are not simulated. Time does not evolve continuously. This discrete-event approach allows us to simulate much larger atomistic systems for much longer times than those MD can provide.

Although the importance of KMC in rare-event simulation has been widely recognized, there are three major challenges in KMC simulation. The first challenge of accuracy for KMC is that ideally all states and events have to be known a priori and listed in the event catalog so that the dynamics of physical processes can be simulated accurately. That is, we should know all possible events and the associated probabilities of occurrence under all configurations when building KMC models. However, this is not realistic because the lack of full knowledge of physical systems does not allow us to know all possible states and transition pathways.

The second challenge of robustness is that during simulation the occurring probabilities of events are assumed to be accurate and fixed. In the real world, this is not true. Uncertainties are always involved in estimating the transition rates thus the probabilities of events. When the rates are estimated by physical experiments, measurement errors are unavoidable. When rates are estimated from first-principles calculation, numerical setup and approximations for computability bring unintentional uncertainties such as the exchange–correlation treatment in density functional theory (DFT) and sampling limitation in quantum Monte Carlo simulation, both of which are quantum mechanical simulation methods to study electronic structures. Model

uncertainty is also inevitable when mathematical models are used to describe physical phenomena. In addition, the transition rates are dynamically changing over time. For instance, in diffusions, when the material structure is under dynamic mechanical load, the stress will change the diffusion rate. Furthermore, bulk diffusion or inter-layer mass transport is very sensitive to temperature change compared to surface diffusion or intra-layer mass transport. Temperature variation will affect the accuracy of KMC prediction. In biochemical processes, the crowding effect, where macromolecules block reaction paths, changes kinetic constants of reactions.

The third challenge of efficiency for KMC is that the time scales of the events vary significantly. Thus it is possible that the frequencies or probabilities of those events are in very different scales. For instance, in simulating of diamond growth in chemical vapor deposition, surface diffusion could be several orders faster than adsorption and desorption. Computational time is not optimized to focus more on slower but critical events.

In this paper, we propose a reliable KMC (R-KMC) simulation framework to address the second challenge of KMC, which is to improve the robustness of KMC prediction under input uncertainty. The need to quantify variability and incertitude separately has been well-recognized. Variability or aleatory uncertainty is the inherent randomness in the system because of fluctuation and perturbation, whereas incertitude or epistemic uncertainty is due to lack of perfect knowledge about the system. The sources of epistemic uncertainty in modeling and simulation include lack of data or missing data, conflicting information, conflicting beliefs, lack of introspection, measurement and numerical errors, and lack of information about dependency.

Here, the new R-KMC mechanism is to perform sampling based on imprecise probability. Interval-valued rates and probabilities are used as a result of the lack of perfect knowledge about transitions. The interval width captures the epistemic portion of uncertainty. Statistical sampling is based on random sets instead of random numbers. Thus the two types of uncertainties are represented and used explicitly to increase the confidence of simulation results. In the remainder of the paper, in Sect. 2 we give a brief overview of self-learning and adaptive KMC which are to address the first challenge of KMC, some approaches to address the uncertain rate challenge, simulation acceleration to address the third challenge, and lastly imprecise probability. In Sect. 3, we introduce the proposed R-KMC mechanism. The multi-event algorithm as the core part of R-KMC is described. In Sect. 4, the R-KMC is demonstrated by examples. And finally in Sect. 5, the convergence of R-KMC towards traditional KMC is studied.

2 Background

The major concept of the proposed R-KMC mechanism is that since the lack of complete knowledge on the underlying physics of reaction or transition processes poses challenges on the accuracy and robustness of KMC prediction, reliable KMC simulation should be conducted such that more information about imprecise parameters can be provided without significantly increasing computational load and the robustness can be improved. The three major KMC challenges are interrelated to each other. Some approaches have been proposed to address those challenges.

2.1 The stochastic simulation algorithm (SSA)

In KMC, how often event j occurs or the probability that event j is fired is determined by its associated rate a_j (also called propensity in chemistry-oriented literature) with respect to other events. Among all M possible transitions leaving the current state, the probability p_j of the transition from the current state to state j is the probability of firing event j , calculated as $p_j = a_j / \sum_{i=1}^M a_i$. The time between transition events X_j is exponentially distributed with the parameter or rate a_j . Thus the probability that the inter-arrival time of event j is the minimum among M independent events is $\Pr[X_j = \min(X_1, \dots, X_M)] = a_j / (a_1 + \dots + a_M)$. Thus, for each KMC step, an event is randomly selected based on the proportions $(a_1 / \sum_{i=1}^M a_i), \dots, (a_M / \sum_{i=1}^M a_i)$. In addition, once an event is fired, the clock of the system is advanced by a second random value t that is exponentially distributed with the rate $\sum_{i=1}^M a_i$, which is the earliest occurring time for any event out of M independent ones that are exponentially distributed. This is the most used so-called direct method or SSA (Gillespie 1976), among others (Chatterjee and Vlachos 2007), which was first proposed by Gillespie. The event selection process in the original algorithm has the linear time complexity. More efficient selection algorithm with data structures of two-level binning (Maksym 1988), binary tree (Blue et al. 1995), and distinct rate list (Schulze 2002) have been developed.

2.2 Self-learning and adaptive KMC

To address the first challenge of KMC, self-learning or adaptive KMC methods have been proposed to accumulate the knowledge of events and build on-the-fly while running KMC, including searching saddle points on the potential energy surface (PES) as part of the simulation, such as by the dimmer method (Henkelman and Jónsson 2001; Mei et al. 2009) and the drag method (Trushin et al. 2005). However, in these adaptive KMC methods where saddle

points are searched on-the-fly, it is assumed that the PES estimated by DFT is accurate. This assumption makes the KMC simulation not robust, since DFT itself has various approximations. Particularly, approximations of exchange and correlation energies in DFT are the major sources of errors.

2.3 Uncertain rate

The uncertainty associated with probabilities in KMC simulation has been recognized. It was shown that the kinetic rate in biochemical reactions varies over time instead of being a constant, because of the macromolecular crowding effect (Berry 2002; Schnell and Turner 2004). It causes mutual impenetrability of solute molecules and phase separation, thus increased folding and refolding rates or reduced diffusion rates. To model the so-called spatial challenge and time dependency of kinetic rates, fractal and Zipf–Mandelbrot relationships were used. A combination of spatial lattice diffusion and chemical reaction simulations was proposed to model the actual crowding effect. In addition, delays associated with transcription and translation are essential in modeling cellular pathways and regulatory networks, where biochemical processes such as diffusion, translocation, protein synthesis and folding do not occur instantaneously and often affected by spatial homogeneities. Delay SSA (Bratsun 2005; Burrage et al. 2007) were proposed to capture delays in temporal models, where the quantities of reactants and products are updated separately with a delayed period of time.

2.4 Simulation acceleration

To improve the time efficiency of KMC, a τ -leap method (Gillespie 2001; Rathinam et al. 2003) was proposed to sample the number of reactions instead of reaction time based on the intrinsic relation between the exponential and Poisson distributions. To alleviate the problem that negative number of species may appear because a large number of reactions are fired at once, improved methods such as leap based on binomial distributions (Chatterjee et al. 2005; Tian and Burrage 2004) and multinomial distributions (Auger et al. 2006; Cai and Xu 2007) have been proposed. The coarse-grained KMC (Katsoulakis et al. 2003) enables simulation of larger length and time scales by grouping lattice sites into coarse cells as the basic spatial unit in simulation. A partitioning approach (Haseltine and Rawlings 2002) categorizes events based on speeds so that fast events are simulated by leaping or even continuous methods to achieve acceleration. Nested SSA (E et al. 2007) simulates fast reactions in the inner loop and slow reactions in the outer loop. Accelerated superbasin KMC (Chatterjee and Voter 2010) dynamically introduces

biases for frequent events so that rare transitions can happen more frequently.

2.5 Imprecise probability

The imprecise probability is a generalization of the traditional precise probability where probability values are no longer precise numbers. Instead of a precise value of the probability $P(E) = p$ associated with an event E , a pair of lower and upper probabilities $P(E) = [\underline{p}, \bar{p}]$ are used to include a set of probabilities and quantify the uncertainty. Imprecise probability differentiates epistemic and aleatory uncertainties both qualitatively and quantitatively, which is the alternative to the traditional sensitivity analysis in probabilistic reasoning to model indeterminacy and imprecision. The range of the interval $[\underline{p}, \bar{p}]$ captures the epistemic uncertainty component. When $\underline{p} = \bar{p}$, the degenerated interval probability becomes a precise one.

Several theories and representations of imprecise probability have been proposed. The Dempster–Shafer theory (Dempster 1967; Shafer 1976) characterizes evidence with discrete probability masses associated with a power set of values. The theory of coherent lower previsions (Walley 1991) models uncertainties with the lower and upper previsions with behavioral interpretations. The possibility theory (Dubois and Prade 1988) represents uncertainties with Necessity–Possibility pairs. Probability bound analysis (Ferson et al. 2003) captures uncertain information with pairs of lower and upper distribution functions or p-boxes. Interval probability (Kuznetsov 1995) characterizes statistical properties as intervals. F-probability (Weichselberger 2000) represents an interval probability as a set of probabilities which maintain the Kolmogorov properties. A random set (Molchanov 2005) is a multi-valued mapping from the probability space to the value space. Fuzzy probability (Möller and Beer 2004) considers probability distributions with fuzzy parameters. A cloud (Neumaier 2004) is a combination of fuzzy sets, intervals, and probability distributions. Recently a generalized interval probability (Wang 2010) based on the generalized interval (Dimitrova et al. 1992; Gardeñes et al. 2001) was proposed.

A generalized interval $\mathbf{x} := [\underline{x}, \bar{x}]$ is defined as a pair of numbers. In contrast, the classical set-based interval is defined as $[[\underline{x}, \bar{x}]] := \{x \in \mathbb{R} | \underline{x} \leq x \leq \bar{x}\}$. $[\underline{x}, \bar{x}]$ is called proper if $\underline{x} \leq \bar{x}$, and called improper if $\underline{x} \geq \bar{x}$. The introduction of improper intervals greatly simplifies the calculus structure of interval probability and makes it very similar to the precise probability in the traditional probability theory. The calculation of generalized intervals is based on the Kaucher interval arithmetic

(Kaucher 1980). Particularly, for two intervals $[\underline{x}, \bar{x}]$ and $[\underline{y}, \bar{y}]$, addition and subtraction are defined as $[\underline{x}, \bar{x}] + [\underline{y}, \bar{y}] = [\underline{x} + \underline{y}, \bar{x} + \bar{y}]$ and $[\underline{x}, \bar{x}] - [\underline{y}, \bar{y}] = [\underline{x} - \bar{y}, \bar{x} - \underline{y}]$ respectively. The width of interval $\mathbf{x} = [\underline{x}, \bar{x}]$ is defined as

$$\text{width} := |\bar{x} - \underline{x}| \quad (1)$$

The relationship between proper and improper intervals is established by a dual operator defined as

$$\text{dual}[\underline{x}, \bar{x}] := [\bar{x}, \underline{x}] \quad (2)$$

In generalized interval probability, the probability measure has the value of generalized interval. The most important property of the generalized interval probability is the logic coherence constraint (LCC). It is stated that for a mutually disjoint event partition $\bigcup_{i=1}^n E_i = \Omega$, the sum of interval probabilities is always one, i.e. $\sum_{i=1}^n \mathbf{p}(E_i) = 1$. The LCC ensures that the imprecise probabilities are logically coherent with precise probabilities. For instance, for a system's working status with $\mathbf{p}(\text{down}) = [0.2, 0.3]$, $\mathbf{p}(\text{idle}) = [0.3, 0.5]$, and $\mathbf{p}(\text{busy}) = 1 - \mathbf{p}(\text{down}) - \mathbf{p}(\text{idle}) = [0.5, 0.2]$. We can interpret it as

$$(\forall p_1 \in [[0.2, 0.3]]) (\forall p_2 \in [[0.3, 0.5]]) (\exists p_3 \in [[0.2, 0.5]]) (p_1 + p_2 + p_3 = 1)$$

according to the interpretability theorems (Gardeñes et al. 2001) of generalized interval.

Based on the logic interpretation, we differentiate non-focal events from focal events. An event E is focal if the associated semantics for $\mathbf{p}(E)$ is universal (\forall). Otherwise, it is non-focal if the semantics is existential (\exists). In the above example, 'down' and 'idle' are focal events, whereas 'busy' is non-focal. The selection of focal and non-focal events is problem-specific and dependent on the analyst's interest. If the epistemic component of uncertainty associated with an event is critical to the analyst, it is treated as a focal event. Otherwise, it is a non-focal event. The semantics of \forall is critical, absolute, and uncontrollable, whereas that of \exists is complementary, balancing, and derived. Obviously, if the analyst is more concerned with 'busy' and 'down' states in the above example, the probability assignments can become $\mathbf{p}(\text{down}) = [0.2, 0.3]$, $\mathbf{p}(\text{busy}) = [0.2, 0.5]$, and $\mathbf{p}(\text{idle}) = 1 - \mathbf{p}(\text{busy}) - \mathbf{p}(\text{down}) = [0.6, 0.2]$. The differentiation between proper and improper interval probabilities and thus focal and non-focal events is to make sure LCC is satisfied. All imprecise probability assignments should satisfy certain conditions to meet rationality criteria. LCC is similar to, but more restrictive than, the coherence condition in Walley's theory of coherent lower previsions. See Wang (2010) for more information of generalized interval probability.

3 Proposed reliable kinetic Monte Carlo (R-KMC) simulation

Without loss of generality, we use the vapor deposition process in Fig. 1 to illustrate the R-KMC. The vapor deposition process is characterized by several events such as adsorption (with rate \mathbf{a}_1), intra-layer surface diffusion (with rates $\mathbf{a}_2, \dots, \mathbf{a}_5$), inter-layer diffusion (with rate \mathbf{a}_6), and desorption (with rate \mathbf{a}_7) associated with adatoms, as well as others. Each of these rates is an interval, i.e. $\mathbf{a}_j = [\underline{a}_j, \bar{a}_j]$ for all j . All boldface symbols in this paper are intervals. Here, all interval rates are proper intervals with $\underline{a}_j \leq \bar{a}_j$ for all j . If a null-event KMC algorithm without differentiating different locations is applied, the imprecise probability that event j is chosen to be the next event is $\mathbf{p}_j = [\underline{a}_j, \bar{a}_j]/a_0$ where a_0 is the upper bound of the sum of all rates. When choosing which event to fire, we may select either one or two events because of the imprecise probability. We call this multi-event algorithm.

Here a simplified example is used to illustrate the multi-event algorithm. Suppose that there are three events A, B , and C that may occur, which have the rates of $\mathbf{a}_1 = [1, 3]$, $\mathbf{a}_2 = [1, 3]$ and $\mathbf{a}_3 = [4, 5]$ respectively. First, we sort the rates based on the uncertainty levels, i.e. the widths of the intervals as defined in Eq. (1), in an ascending order. Next, the lower and upper bounds of intervals are “flipped” in an alternating pattern. Now the sequence of rates is $\mathbf{a}'_3 = [4, 5]$, $\mathbf{a}'_1 = [3, 1]$, $\mathbf{a}'_2 = [1, 3]$. Then, a null event N is introduced, e.g. $\mathbf{a}'_N = [1, 0]$. In contrast to other events, when a null event is selected to fire, there is no change of the system’s state. The general principle of introducing null events is to make sure that the interval sum $a_0 = \mathbf{a}'_3 + \mathbf{a}'_1 + \mathbf{a}'_2 + \mathbf{a}'_N$ degenerates to a precise number ($a_0 = [9, 9]$ in this example). This is based on the LCC mentioned in Sect. 2.5. An empirical cumulative distribution function (c.d.f.) then can be constructed from $\mathbf{a}'_3, \mathbf{a}'_1, \mathbf{a}'_2$, and \mathbf{a}'_N , as shown in Fig. 2. Based on the empirical interval c.d.f., the events are chosen to fire by an inverse transform method that is similar to the one in the traditional KMC. That is, a random number u which is uniformly distributed between 0 and 1 is generated. The evaluation of the inverse c.d.f. is used to select the random sets that u corresponds to. For instance, in Fig. 2, if u_1 is generated, the next event is $\{C\}$. If u_2 is generated, the next events are $\{A, B\}$. That is, a random set of events are chosen. The probability that an event is selected is proportional to the value of the associated rate in the inverse transform method. That is, the probability that event A is selected is at least \underline{a}_1/a_0 and at most \bar{a}_1/a_0 . The probabilities for other events are similar. Notice that the goal of the sorting and flipping procedure for the rates is to make the constructed interval c.d.f.’s always have the same zig-zag pattern for

the lower and upper bounds of accumulated probabilities as in Fig. 2, which is necessary to ensure that the lower and upper probabilities of event selections exactly correspond to the originally specified interval rates. A rigorous derivation will be shown in Sect. 3.1.

Once a random set of events are fired, the time should be advanced. In the traditional KMC, the time increment is exponentially distributed with the rate $\sum_j a_j$ and sampling of time increment is based on the inverse function of exponential c.d.f. In R-KMC, the time to fire a random set of events is an interval. We choose the interval-valued average time among all firing events, which are the earliest and latest possible times that multiple events occur.

In the traditional KMC where epistemic uncertainty of input parameters is ignored, the samplings for event selection and clock advancement are based on precise probability with certitude. In R-KMC, the uncertainty associated with the rates leads to the imprecise probability that an event is likely to be chosen to fire. The details of event selection and clock advancement in the multi-event algorithm are described in the following subsections.

3.1 Event selection

For M possible reaction or transition events, each event is characterized by an interval-valued rate or propensity function \mathbf{a}_j ($j = 1, \dots, M$). First, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M$ are sorted based on the widths of the intervals in the ascending order to $\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots, \mathbf{a}^{(M)}$, where $\text{wid}(\mathbf{a}^{(1)}) \leq \text{wid}(\mathbf{a}^{(2)}) \leq \dots \leq \text{wid}(\mathbf{a}^{(M)})$. Further, we define a *-sum, denoted by \sum^* , recursively as

$$\sum_{j=1}^{*J} \mathbf{a}^{(j)} := \text{dual} \left(\sum_{j=1}^{*J-1} \mathbf{a}^{(j)} \right) + \mathbf{a}^{(J)} \quad (J = 2, \dots, M) \tag{3}$$

with $\sum_{j=1}^{*1} \mathbf{a}^{(j)} = \mathbf{a}^{(1)}$. The original intent of interval rates or propensities to specify the lower and upper probabilities is maintained during event selection, shown as follows.

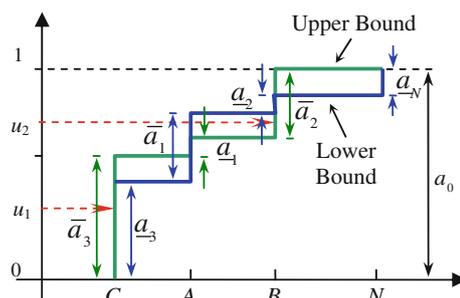


Fig. 2 An illustration of random set sampling based on the empirical interval c.d.f

Based on operator “+” of the Kaucher interval arithmetic, if $\text{wid} \mathbf{x} \leq \text{wid} \mathbf{y}$ for any two intervals \mathbf{x} and \mathbf{y} that are not improper (i.e. proper intervals or point-wise real numbers), then $\text{dual} \mathbf{x} + \mathbf{y}$ is not improper and $\text{wid}(\text{dual} \mathbf{x} + \mathbf{y}) \leq \text{wid} \mathbf{y}$. Therefore,

$$\text{wid} \left(\sum_{j=1}^{*J-1} \mathbf{a}^{(j)} \right) \leq \text{wida}^{(J-1)} \leq \text{wida}^{(J)} \quad (J = 2, \dots, M) \tag{4}$$

Figure 3 illustrates how the cumulative *-sum of the sorted interval propensities is constructed. Because of Eq. (4), the lower and upper bounds of $\sum_{j=1}^{*J} \mathbf{a}^{(j)}$ are always calculated from the upper and lower bounds of $\sum_{j=1}^{*J-1} \mathbf{a}^{(j)}$ respectively. At the top of the accumulative *-sum, a null event is introduced with the propensity of $\mathbf{a}_{null} = [0, \text{wid}(\sum_{j=1}^{*M} \mathbf{a}^{(j)})]$. Then we define the real-valued upper limit of propensity sum as

$$a_0 = \text{dual} \left(\sum_{j=1}^{*M} \mathbf{a}^{(j)} \right) + [0, \text{wid}(\sum_{j=1}^{*M} \mathbf{a}^{(j)})] \tag{5}$$

Thus when a random fraction ua_0 is generated, either one or two events will be selected to fire, depending on the random number u . If the projected line of ua_0 simultaneously intersects with two propensities (J and $J + 1$ shown in Fig. 3), the corresponding two events will be fired. This process ensures that the probability that event J is selected is $[\underline{a}^{(J)}/a_0, \bar{a}^{(J)}/a_0]$, which is the intent of interval propensities. Therefore, the multi-event algorithm based on the total propensity calculated by the *-sum as in Eq. (5) faithfully chooses events according to their corresponding minimum and maximum probabilities, whereas a naïve attempt of simple sum of the respective lower and upper propensities does not. Again, the introduction of null event is to satisfy LCC so that the sum is up to a real value. When event M and the null event are selected, only event M is fired whereas the null event does not change the state of the system.

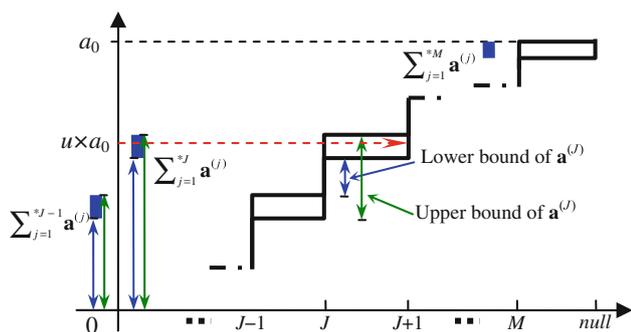


Fig. 3 The cumulative *-sum to choose events randomly based on interval propensities

The pseudo-code of the multi-event algorithm is listed in Table 1. In each iteration of R-KMC, the cumulative *-sum is updated first. With a random number generated, the corresponding random set of events is selected. These events then are fired and the system’s state is updated.

3.2 Clock advancement

In the classical SSA, the simulation clock is advanced by the earliest time approach. That is, M events are assumed to be independent, where the time between events T_j are exponentially distributed, i.e. $P(T_j \leq \tau) = 1 - \exp(-a_j \tau)$ for $j = 1, \dots, M$. The earliest time $T^{(1)} = \min(T_1, \dots, T_M)$ of any event occurs is also exponentially distributed, as in $P(T^{(1)} \leq \tau) = 1 - P(T^{(1)} > \tau) = 1 - \prod_{j=1}^M P(T_j > \tau) = 1 - \exp(-(\sum_{j=1}^M a_j) \tau)$, with the rate of $\sum_{j=1}^M a_j$. Therefore, the clock is advanced by the random variant $T = -\ln \rho / \sum_{j=1}^M a_j$, where ρ is a random number sampled from the standard uniform distribution between 0 and 1.

In the multi-event algorithm, a random set of multiple events are chosen and fired at a time. Uncertainties are associated with the clock advancement for firing the random set of events. That is, how much time is needed for these random events to occur varies and includes both components of epistemic and aleatory uncertainties. To estimate the epistemic uncertainty component, we use an interval approach. The worst and best scenarios need to be given. The least time for the set of events to fire is when all of them occur at the same time. Therefore, the lower bound of clock advancement is

$$T_L = -\ln \rho / \sum_{j=1}^M \bar{a}_j. \tag{6}$$

The longest possible time for the set of events to fire is when the events occur consecutively one by one. Suppose that the respective rates of the firing sequence of n out of N events from a random set are $[\underline{a}^{(1)}, \bar{a}^{(1)}], \dots, [\underline{a}^{(n)}, \bar{a}^{(n)}]$. In the worst-case scenario, the total elapsed time for the sequence of n events is $T^{(n)} = X^{(1)} + X^{(2)} + \dots + X^{(n)}$, where inter-arrival times $X^{(1)}, X^{(2)}, \dots, X^{(n)}$ are exponentially distributed with the respective rates of $\underline{\alpha}_0, \underline{\alpha}_1, \dots, \underline{\alpha}_{n-1}$ where

$$\begin{cases} \underline{\alpha}_0 = \sum_{i=1}^N a_i \\ \underline{\alpha}_1 = \sum_{i=1}^N a_i - \underline{a}^{(1)} \\ \dots \\ \underline{\alpha}_{n-1} = \sum_{i=1}^N a_i - \sum_{j=1}^{n-1} \underline{a}^{(j)} \end{cases} \tag{7}$$

The rates are changing because after each event is fired, the earliest time for any of the left events to occur depends on the number of events that are left. Then the c.d.f. of $T^{(n)}$ is

Table 1 The pseudo-code of the multi-event algorithm in each R-KMC iteration

```

INPUT: The list of propensity_interval
OUTPUT: The list of events to fire
Sort propensity_interval in the ascending order based on the
widths of intervals;
// calculate the *sum of interval propensities to build the
cumulative distributive function
sum_interval[1]=propensity_interval[1];
FOR i=2:numEvents
    sum_interval[i]=dual(sum_interval[i-1])
    +propensity_interval[i];
END
sum_interval[numEvents+1]=
dual(sum_interval[numEvents])+propensityOfNullEvent;

// sampling a random set of events
Generate a random number  $u \sim \text{Uniform}(0,1)$ ;
Fraction= $u \times \text{sum\_interval}[\text{numEvents}+1]$ 
FOR i=1:numEvents
    Find the smallest  $i$  as  $\text{index\_L}$  such that
    Fraction $\leq \text{sum\_interval}[\text{index\_L}].\text{UBound}$ ;
END
FOR i=numEvents:1
    Find the largest  $i$  as  $\text{index\_U}$  such that
    Fraction $\geq \text{sum\_interval}[\text{index\_U}-1].\text{LBound}$ ;
END
RETURN the random set of events corresponding to
propensity_interval[ $\text{index\_L}:\text{index\_U}$ ]
    
```

4 Implementation and demonstration

The proposed R-KMC based on random set sampling is implemented in C++ and integrated with SPPARKS (2009), which is an open-source KMC toolbox developed at the Sandia National Laboratories. We demonstrate R-KMC by two examples, an *E. coli* reaction network and the microbial fuel cell (MFC) simulation.

4.1 *E. coli* reaction network

The R-KMC is first demonstrated by the classical model of the *E. coli* reaction network from Kierzek (2002), as shown in Fig. 4, with real-valued reaction rates. In R-KMC, the rates are intervals because of uncertainties involved. With the interval rates, interval propensities and probabilities are calculated. Within each iteration, a random set of reaction events are selected based on the algorithm in Sect. 3.1. The interval time advancement is sampled based on the method in Sect. 3.2.

At $t = 0$, the initial counts of species are one for PLac, 35 for RNAP, and 350 for Ribosome. All others are zeros. Based on the nominal values of rates in Fig. 4, interval rates are used. We chose interval rates for ± 1 and $\pm 10\%$ of the nominal values in the respective experiments and compare the results with that from the traditional KMC. For instance, the rates of the reaction $PLacRNAP \rightarrow TrLacZI$ at the right-bottom corner of Fig. 4 become [0.99, 1.01] and [0.9, 1.1] respectively. For each of the three scenarios (original real-valued rates, $\pm 1\%$ interval rates, and $\pm 10\%$ interval rates), a sample size of 20 is used. Figure 5 compares the sample average values of some species over time, including *product*, *Ribosome*, *dgrLacY*, and *LacZ*. In the charts, red curves represent the average values in the real case. Blue dotted and dash curves are the lower and upper bounds of those for the $\pm 1\%$ interval case. For the lower-bound curves, the upper-bound time increment Δt_U is used in the plot, whereas the Δt_L is used for the upper-bound curves. Similarly, green dotted and dash curves are the lower and upper bounds of those for the $\pm 10\%$ interval case. It is observed that the $\pm 10\%$ interval case enclose the real case better than the $\pm 1\%$ interval case.

4.2 Microbial fuel cell (MFC)

The MFC is a type of fuel cell that converts the chemical energy contained in organic matter to electricity using microorganisms (bacteria) as a biocatalyst from organic waste and renewable biomass. A MFC typically consists of two chambers, an anaerobic anode chamber and an aerobic cathode chamber separated by a proton exchange membrane (PEM). For scaling up the MFC technology to the commercial scale, fine-grained multi-physics simulation of

$$\begin{aligned}
 P(T^{(n)} \leq \tau) &= 1 - P(T^{(n)} > \tau) \\
 &= 1 - \sum_{j=0}^{n-1} A_j \exp(-\alpha_j \tau)
 \end{aligned}
 \tag{8}$$

where $A_j = \prod_{i=0, i \neq j}^{n-1} \frac{\alpha_i}{\alpha_i - \alpha_j}$. Instead of sampling based on Eq.

(8), it is easier to sample each of $X^{(j)}$'s as exponential distributions and the sum of the samples will be a sample of $T^{(n)}$. Therefore, the upper bound of clock advancement is

$$T_U = -\ln \rho \left[\sum_{j=0}^{n-1} 1/\alpha_j \right]
 \tag{9}$$

where ρ is the same random number as in Eq. (6) and α_j 's are defined as in Eq. (7).

With the interval time increment $[T_L, T_U]$, the current system time $[t_L, t_U]$ in the simulation clock as an interval is updated to $[t_L + T_L, t_U + T_U]$. The purpose of keeping track of an interval time is to estimate the best-case and worst-case scenarios that how fast a system evolves. The interval clock indirectly provides an estimation of lower and upper bounds of system states, which forms the basis of R-KMC to accommodate the uncertain simulation parameters, particularly transition rates.

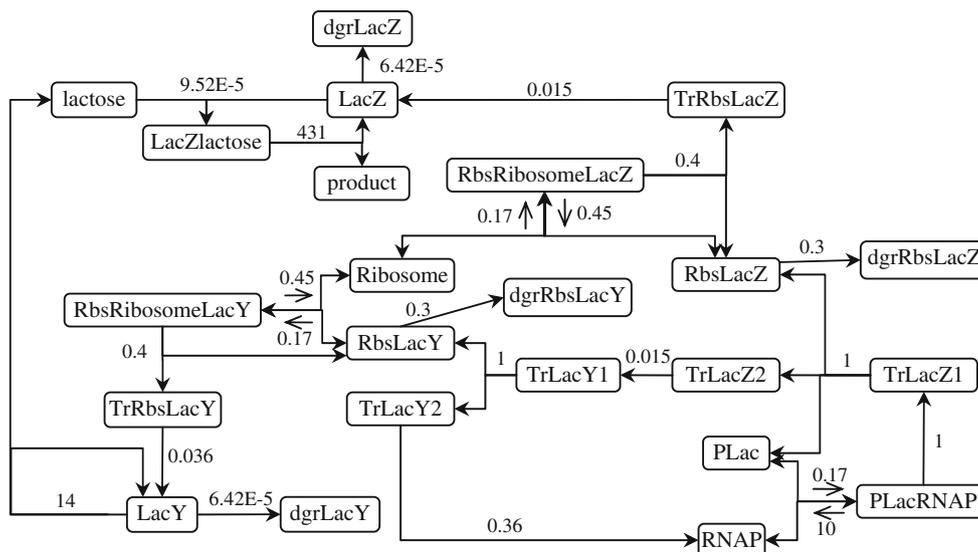


Fig. 4 The reaction channels of LacZ and LacY proteins in *E. coli* (Kierzek 2002)

electrochemical processes for the entire fuel cell system is necessary for better understanding of the system. Here we demonstrate the R-KMC mechanism by a simplified example of MFC reaction networks adopted from Refs. (Picioreanu et al. 2010; Zeng et al. 2010) as in Table 2, where the estimated rates used in the R-KMC simulation are also listed. The first nine reactions (R1–R9) occur in the anode chamber whereas the last four (R10–R13) in the cathode chamber. Two example outputs are shown in Fig. 6. The engineering design procedure involves the optimization of geometries or chemical compositions of electrode, PEM, reduction-oxidization mediator, etc for desirable rates.

5 Weak convergence

The random set based R-KMC can be viewed as a generalization of the traditional KMC. In this section, we show that the multi-event algorithm in R-KMC converges to the traditional SSA in KMC smoothly with respect to both distributions and expected values as the widths of interval transition rates reduce towards zeros, when epistemic uncertainty vanishes.

Suppose that a system consisting of N species $\{S_1, S_2, \dots, S_N\}$ has the state variable $X = (X_1, \dots, X_N) \in \mathbb{Z}_+^N$ where $\mathbb{Z}_+ = N \cup \{0\}$ is the set of nonnegative integers. There are a total of M reaction channels R_j 's ($j = 1, \dots, M$), each of which is characterized by an interval-valued propensity function $\mathbf{a}_j(x)$ given the current state $X = x$, where $\mathbf{a}_j: \mathbb{Z}_+^N \rightarrow \mathbb{R}_+^2$, and its state change vector $v_j = (v_{j1}, \dots, v_{jN})$. From the sum of propensity function $a_0(x)$ defined in Eq. (5), the probability of firing R_j is $P(T_j = \min(T_1, \dots, T_M)) = [a_j/a_0, \bar{a}_j/a_0]$.

As a special case of the generalized Chapman–Kolmogorov equation [Eq. (22) in “Appendix”], the interval master equation that describes the evolution of the distribution of the system states is

$$\frac{d}{dt} \mathbf{p}(x, t|x_0, t_0) = \sum_{j=1}^M \mathbf{W}_j(x - v_j) \mathbf{p}(x - v_j, t|x_0, t_0) - \text{dual} \left[\sum_{j=1}^M \mathbf{W}_j(x) \mathbf{p}(x, t|x_0, t_0) \right] \tag{10}$$

where $\mathbf{p}(x, t|x_0, t_0)$ is the probability of $X(t) = x$ given the initial distribution $X(t_0) = x_0$, and $\mathbf{W}_j(x) = \mathbf{a}_j(x)/a_0(x)$ is the transition rate. Because \mathbf{a}_j 's are interval propensities, the solution of Eq. (10) is a set of probability evolution paths.

In the multi-event algorithm in Sect. 3, the probability density function of time between R_j is fired at the current state x and next R_j is

$$\mathbf{f}_j(x, t) = \mathbf{W}_j(x) \cdot a_0(x) \exp(-a_0(x)t) = \mathbf{a}_j(x) \exp(-a_0(x)t) \tag{11}$$

which converges to the one in the traditional SSA algorithm as the widths of the interval propensities \mathbf{a}_j 's reduce to zeros.

A different view of the system is the interval-valued state variable with precise time. The path-wise representation of the Poisson process with the integral form is

$$\mathbf{X}(t) = \mathbf{x}_0 + \sum_{j=1}^M P_j \left(\int_0^t \mathbf{a}_j(\tau) d\tau \right) \mathbf{v}_j \tag{12}$$

where $P_j(t) = P_j(\int_0^t \mathbf{a}_j(\tau) d\tau)$ is a generalized Poisson or counting process of the number of events to fire R_j and

Fig. 5 Comparisons of the numbers of species over time in the *E. coli* reaction network between the traditional KMC and R-KMC simulations, where interval reaction rates are ± 1 and $\pm 10\%$

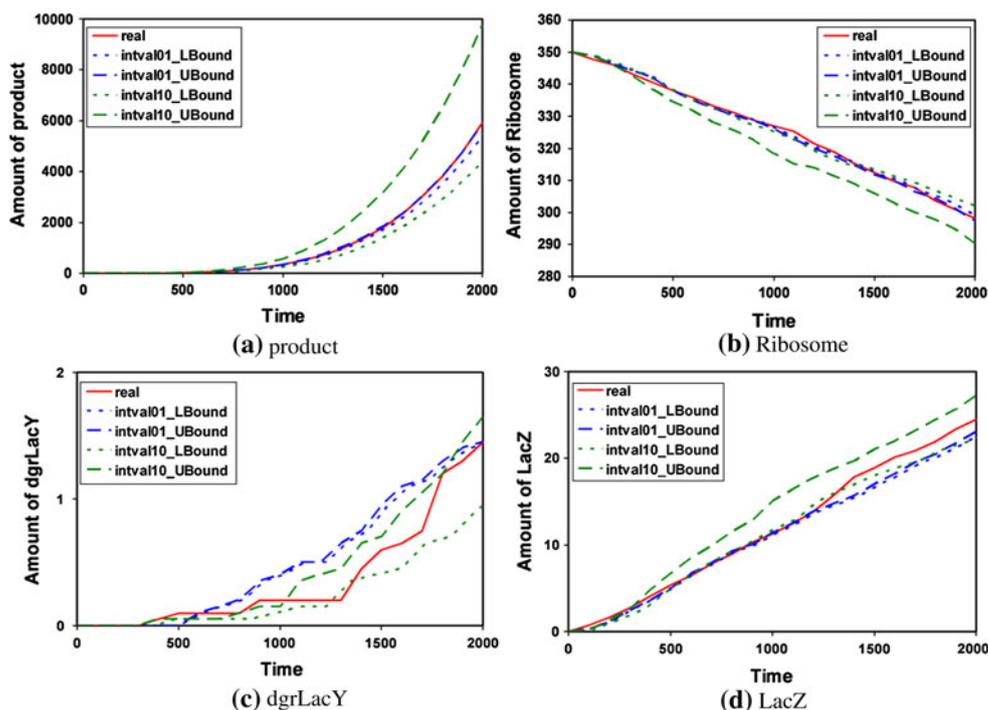


Table 2 The reactions of a two-chamber microbial fuel cell used in the R-KMC model

Number of sites involved	Reaction/transition event	Rate constant
R1: water dissociation	$H_2O \leftrightarrow OH^- + H^+$	10^1
R2: carbonic acid dissociation	$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$	10^1
R3: acetic acid dissociation	$AcH \leftrightarrow Ac^- + H^+$	10^1
R4: reduced thionine first dissociation	$MH_3^+ \leftrightarrow MH_2 + H^+$	10^1
R5: reduced thionine second dissociation	$MH_4^{2+} \leftrightarrow MH_3^+ + H^+$	10^1
R6: acetate with oxidized mediator	$Ac^- + MH^+ + NH_4^+ + H_2O \rightarrow X_{Ac} + MH_3^+ + HCO_3^- + H^+$	10^1
R7: oxidation double protonated mediator	$MH_4^{2+} \rightarrow MH^+ + 3H^+ + 2e^-$	10^1
R8: oxidation single protonated mediator	$MH_3^+ \rightarrow MH^+ + 2H^+ + 2e^-$	10^1
R9: oxidation neutral mediator	$MH_2 \rightarrow MH^+ + H^+ + 2e^-$	10^1
R10: proton diffusion through PEM	$H^+ \rightarrow H_-^+$	10^{-2}
R11: electron transport from anode to cathode	$e^- \rightarrow e_-^-$	10^{-2}
R12: reduction of oxygen with current generated	$2H_-^+ + 1/2O_{2-} + 2e_-^- \rightarrow H_2O_-$	10^5
R13: reduction of oxygen with current generated	$O_{2-} + 4e_-^- + 2H_2O_- \rightarrow 4OH_-^-$	10^3

$\int_0^t \mathbf{a}_j(\tau) d\tau$ is the mean firing rate of time t . $\mathbf{X}(t)$ is a random interval with discrete integer values at time t . The expectation is

$$\begin{aligned}
 E(\mathbf{X}(t)) &= x_0 + \sum_{j=1}^M \mathbf{v}_j \int_0^t \mathbf{a}_j(\tau) d\tau \\
 &= \left[x_0 + \sum_{j=1}^M \mathbf{v}_j \int_0^t \underline{\mathbf{a}}_j(\tau) d\tau, x_0 + \sum_{j=1}^M \mathbf{v}_j \int_0^t \bar{\mathbf{a}}_j(\tau) d\tau \right] \quad (13)
 \end{aligned}$$

Therefore, when $\underline{\mathbf{a}}_j(t) = \bar{\mathbf{a}}_j(t)$ for all t , $P_j(t)$'s are degenerated to regular Poisson processes in the traditional

KMC, and the expected state values in R-KMC are the same as the ones in KMC.

The reliable KMC is to improve the robustness of the simulation given input uncertainties. During simulation, the imprecise rates in input are converted to imprecise times to reach certain states in output. We propose that the robustness is measured by the probability of time enclosure as

$$P_e = P(\underline{X}^{-1}(x) \leq X^{-1}(x) \leq \bar{X}^{-1}(x)) \quad (14)$$

for a particular state x , where random variables $\underline{X}^{-1}(\cdot)$, $X^{-1}(\cdot)$, and $\bar{X}^{-1}(\cdot)$ are the times that state x is first reached

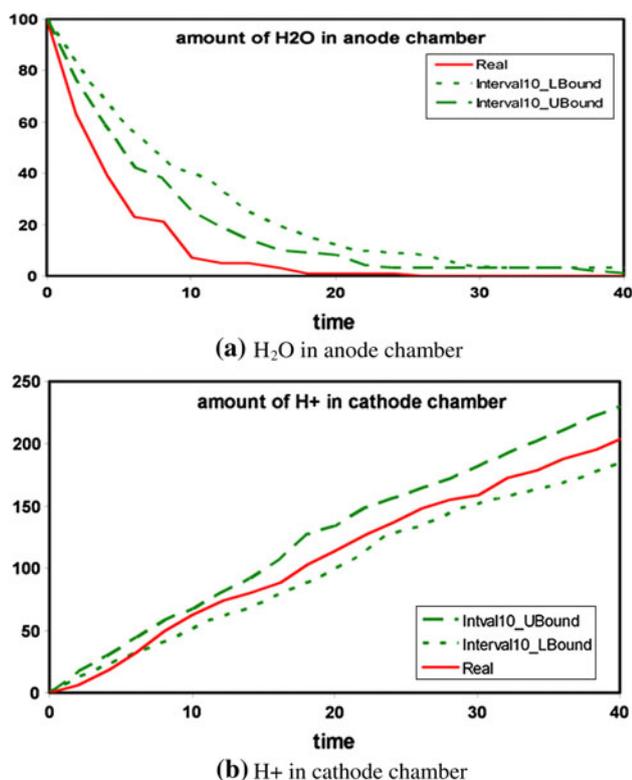


Fig. 6 Comparisons of the numbers of species over time in the MFC reaction network between the traditional KMC and R-KMC simulations, where interval reaction rate is $\pm 10\%$

for the lower-bound, real-valued, and upper-bound scenarios respectively. The probabilistic measure in Eq. (14) indicates that the predicted time interval by R-KMC does not necessarily always enclose the time predicted by KMC. For instance, with the selected number of runs in the *E. coli* example, the first-passage time when the amount of Ribosome reaches 325 in KMC is not enclosed by the one in R-KMC, as shown in Fig. 5b. Similarly in Fig. 6a, the amount of H₂O for the first transient period at the beginning of MFC simulation is not enclosed.

The probability of the first-passage time to reach certain state can be estimated, since all three time variables follow the Erlang distribution with c.d.f.'s as

$$P(X_i^{-1}(\cdot) \leq t) = 1 - \sum_{k=0}^{n_i-1} e^{-a_0 t} (a_0 t)^k / k!$$

$$(X_i^{-1} = \underline{X}^{-1}, X^{-1}, \bar{X}^{-1}; n_i = \underline{n}, n, \bar{n})$$

where \underline{n} , n , and \bar{n} are the minimal, real-valued, and maximal numbers of events that occur.

6 Concluding remarks

In this paper, we propose an R-KMC mechanism based on random-set sampling to improve the simulation robustness

where uncertainty associated with transition rates and probabilities in simulation is considered. The new R-KMC mechanism considers a range of possible state updates for each simulation step based on a proposed multi-event algorithm. The multi-event algorithm uses generalized intervals and strictly follows the semantics of interval probability so that the sampling process is a generalization of the SSA. The system can be viewed either as interval-valued state variables with precise time, or as precise state with interval-valued time. The convergence of the multi-event algorithm towards the traditional SSA in KMC is demonstrated.

For simple chemical reactions, the uncertainty effect is captured by an interval system time, which indicates the earliest and latest possible times required for the system reach a possible state, as demonstrated by the examples. For more complex systems where complete chemical and physical processes at different locations need to be kept track of, the uncertainty effect can be similarly captured as the earliest and latest times that a number of certain species at a particularly location is observed.

In this paper, only the worst- and best-case of time advancement is considered in the multi-event algorithm. Future work will include the extension to incorporate generalized intervals so that interval uncertainty of time can be reduced to zeros by calculating with improper intervals. For instance, when measurements are taken at certain time, they can serve as a reference point and precise numbers of species are given without interval uncertainty. This generalizes the procedure of R-KMC simulation and validation.

Appendix: Derivation of the generalized Chapman–Kolmogorov equation

Following the definition of interval derivative by Markov (1979), the derivative of a generalized interval function

$\mathbf{f}(t) = [\underline{f}(t), \bar{f}(t)]$ is defined as

$$d\mathbf{f}(t)/dt := \lim_{\Delta t \rightarrow 0} (\mathbf{f}(t + \Delta t) - \text{dual}\mathbf{f}(t))/\Delta t$$

$$= \left[\lim_{\Delta t \rightarrow 0} \frac{\underline{f}(t + \Delta t) - \underline{f}(t)}{\Delta t}, \lim_{\Delta t \rightarrow 0} \frac{\bar{f}(t + \Delta t) - \bar{f}(t)}{\Delta t} \right]$$

where dual is defined as in Eq. (2). Note that all boldface symbols in this paper are generalized intervals. We define the derivative of generalized interval probability $\mathbf{p}(x, t|y, t')$ with respect to time t as follows. With state variables $x, y \in \mathbf{R}^n$,

$$\frac{\partial}{\partial t} \mathbf{p}(x, t|y, t') := \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left\{ \begin{array}{l} \mathbf{p}(x, t + \Delta t|y, t') \\ -\text{dual}\mathbf{p}(x, t|y, t') \end{array} \right\} \quad (15)$$

Because of the logic coherent constraint in generalized interval probability, we have

$$\int dz \mathbf{p}(z, t + \Delta t | x, t) = 1 \tag{16}$$

where state variable $z \in R^n$. With the Markovian property, this leads to

$$\mathbf{p}(x, t + \Delta t | y, t') = \int dz \mathbf{p}(x, t + \Delta t | z, t) \mathbf{p}(z, t | y, t') \tag{17}$$

Replace the first term in Eq. (15) with Eq. (17) and multiply the second term by Eq. (16). Eq. (15) now becomes

$$\begin{aligned} & \frac{\partial}{\partial t} \mathbf{p}(x, t | y, t') \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left\{ \int dz \left[\begin{array}{l} \mathbf{p}(x, t + \Delta t | z, t) \mathbf{p}(z, t | y, t') \\ - \text{dual} \mathbf{p}(z, t + \Delta t | x, t) \mathbf{p}(x, t | y, t') \end{array} \right] \right\} \end{aligned} \tag{18}$$

Consider two subdomains $\|x - z\| \leq \varepsilon$ and $\|x - z\| > \varepsilon$ separately where ε is small enough. Eq. (18) can be regarded as

$$\frac{\partial}{\partial t} \mathbf{p}(x, t | y, t') = \mathbf{I}_{\|x-z\| \leq \varepsilon} + \mathbf{I}_{\|x-z\| > \varepsilon}$$

where $\mathbf{I}_{\|x-z\| \leq \varepsilon}$ is the right-side integral in Eq. (18) within the small neighborhood that captures continuous diffusion processes, whereas $\mathbf{I}_{\|x-z\| > \varepsilon}$ is the one outside the neighborhood that represents jump processes.

Within the small neighborhood $\|x - z\| \leq \varepsilon$, let $x - z = h$ and define $\mathbf{f}(x, h) := \mathbf{p}(x + h, t + \Delta t | x, t) \mathbf{p}(x, t | y, t')$. Then $\mathbf{p}(z, t + \Delta t | x, t) \mathbf{p}(x, t | y, t') = \mathbf{f}(x, -h)$

and

$$\begin{aligned} & \mathbf{p}(x, t + \Delta t | z, t) \mathbf{p}(z, t | y, t') \\ &= \mathbf{p}(x, t + \Delta t | x - h, t) \mathbf{p}(x - h, t | y, t') = \mathbf{f}(x - h, h). \end{aligned}$$

Apply the Taylor alike expansion

$$\begin{aligned} \mathbf{f}(x - h, h) &= \mathbf{f}(x, h) - \text{dual} \sum_{i=1}^n \frac{\partial \mathbf{f}}{\partial x_i} h_i \\ &+ \frac{1}{2} \sum_i \sum_j \frac{\partial^2 \mathbf{f}}{\partial x_i \partial x_j} h_i h_j + O(\varepsilon^3) \end{aligned}$$

where

$$\frac{\partial \mathbf{f}(x, h)}{\partial x_i} := \lim_{\Delta x_i \rightarrow 0} \left\{ \frac{\mathbf{f}(x + \Delta x_i, h) - \text{dual} \mathbf{f}(x, h)}{\Delta x_i} \right\}$$

and similarly.

$$\begin{aligned} & \frac{\partial^2 \mathbf{f}(x, h)}{\partial x_i \partial x_j} := \\ & \lim_{\substack{\Delta x_i \rightarrow 0 \\ \Delta x_j \rightarrow 0}} \left\{ \frac{\mathbf{f}(x + \Delta x_i + \Delta x_j, h) - \text{dual} \mathbf{f}(x + \Delta x_i, h) - \text{dual} \mathbf{f}(x + \Delta x_j, h) + \mathbf{f}(x, h)}{\Delta x_i \Delta x_j} \right\} \end{aligned}$$

Then

$$\begin{aligned} & \mathbf{I}_{\|x-z\| \leq \varepsilon} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left\{ \int_{\|x-z\| \leq \varepsilon} dz \left[\begin{array}{l} \mathbf{p}(x, t + \Delta t | z, t) \mathbf{p}(z, t | y, t') \\ - \text{dual} \mathbf{p}(z, t + \Delta t | x, t) \mathbf{p}(x, t | y, t') \end{array} \right] \right\} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left\{ \int_{\|h\| \leq \varepsilon} dh \left[\begin{array}{l} \mathbf{f}(x, h) - \text{dual} \sum_{i=1}^n \frac{\partial \mathbf{f}}{\partial x_i} h_i \\ + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 \mathbf{f}}{\partial x_i \partial x_j} h_i h_j \\ - \text{dual} \mathbf{f}(x, -h) + O(\varepsilon^3) \end{array} \right] \right\} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left[\int_{\|h\| \leq \varepsilon} dh \mathbf{f}(x, h) - \text{dual} \int_{\|h\| \leq \varepsilon} dh \mathbf{f}(x, -h) \right] \\ &\quad - \text{dual} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} dh \sum_{i=1}^n \frac{\partial \mathbf{f}}{\partial x_i} h_i \\ &\quad + \frac{1}{2} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} dh \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 \mathbf{f}}{\partial x_i \partial x_j} h_i h_j \\ &\quad + \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} O(\varepsilon^4) \\ &= -\text{dual} \sum_{i=1}^n \left[\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} dh \frac{\partial \mathbf{f}}{\partial x_i} h_i \right] \\ &\quad + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left[\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} dh \frac{\partial^2 \mathbf{f}}{\partial x_i \partial x_j} h_i h_j \right] \\ &\quad + \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} O(\varepsilon^4) \end{aligned}$$

since $\int_{\|h\| \leq \varepsilon} dh \mathbf{f}(x, h) = \int_{\|h\| \leq \varepsilon} dh \mathbf{f}(x, -h)$. Furthermore,

$$\begin{aligned} & \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} dh \frac{\partial \mathbf{f}}{\partial x_i} h_i \\ &= \frac{\partial}{\partial x_i} \left[\frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} h_i \mathbf{f} dh \right] \\ &= \frac{\partial}{\partial x_i} \left[\mathbf{p}(x, t | y, t') \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} h_i dh \mathbf{p}(x + h, t + \Delta t | x, t) \right] \end{aligned}$$

Similarly,

$$\begin{aligned} & \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} dh \frac{\partial^2 \mathbf{f}}{\partial x_i \partial x_j} h_i h_j \\ &= \frac{\partial^2}{\partial x_i \partial x_j} \left[\mathbf{p}(x, t|y, t') \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} h_i h_j dh \mathbf{p}(x+h, t+\Delta t|x, t) \right] \end{aligned}$$

In addition, $\lim_{\Delta t \rightarrow 0} O(\varepsilon^4)/\Delta t = \lim_{\Delta t \rightarrow 0} O(\varepsilon^3)/1 = O(\varepsilon^3)$
 We define

$$\mathbf{A}_i(x, t) := \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} h_i dh \mathbf{p}(x+h, t+\Delta t|x, t) \quad (19)$$

$$\mathbf{B}_{ij}(x, t) := \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{\|h\| \leq \varepsilon} h_i h_j dh \mathbf{p}(x+h, t+\Delta t|x, t) \quad (20)$$

If $X(t) = (x_1(t), \dots, x_n(t))$ represents the state of the system at time t , then

$$\begin{aligned} & \int_{\|h\| \leq \varepsilon} h_i dh \mathbf{p}(x+h, t+\Delta t|x, t) \\ &= E[x_i(t+\Delta t) - x_i(t)|X(t)] \end{aligned}$$

is the expected state value change in the i th direction. Therefore, Eq. (19) is interpreted as the state value change rate along one direction, known as the drift vector. Similarly,

$$\begin{aligned} & \int_{\|h\| \leq \varepsilon} h_i h_j dh \mathbf{p}(x+h, t+\Delta t|x, t) \\ &= E[(x_i(t+\Delta t) - x_i(t))(x_j(t+\Delta t) - x_j(t))|X(t)]. \end{aligned}$$

Equation (20) is the combined area change rate in two directions, known as the diffusion matrix.

We now have

$$\begin{aligned} & \mathbf{I}_{\|x-z\| \leq \varepsilon} \\ &= -\text{dual} \sum_{i=1}^n \frac{\partial}{\partial x_i} [\mathbf{A}_i(x, t) \mathbf{p}(x, t|y, t')] \\ &+ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} [\mathbf{B}_{ij}(x, t) \mathbf{p}(x, t|y, t')] + O(\varepsilon^3) \end{aligned}$$

for the diffusion process.

For the jump process

$$\begin{aligned} & \mathbf{I}_{\|x-z\| > \varepsilon} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left\{ \int_{\|x-z\| > \varepsilon} dz \left[\mathbf{p}(x, t+\Delta t|z, t) \mathbf{p}(z, t|y, t') \right] \right. \\ &= \int_{\|x-z\| > \varepsilon} dz \left[\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \{ \mathbf{p}(x, t+\Delta t|z, t) \} \mathbf{p}(z, t|y, t') \right] \\ & \quad \left. - \text{dual} \int_{\|x-z\| > \varepsilon} dz \left[\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \{ \mathbf{p}(z, t+\Delta t|x, t) \} \mathbf{p}(x, t|y, t') \right] \right\} \end{aligned}$$

We define

$$\mathbf{W}(y|x, t) := \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \{ \mathbf{p}(y, t+\Delta t|x, t) \} \quad (21)$$

as the rate of transition from x to y . Then

$$\begin{aligned} \mathbf{I}_{\|x-z\| > \varepsilon} &= \int_{\|x-z\| > \varepsilon} dz [\mathbf{W}(x|z, t) \mathbf{p}(z, t|y, t')] \\ & \quad - \text{dual} \int_{\|x-z\| > \varepsilon} dz [\mathbf{W}(z|x, t) \mathbf{p}(x, t|y, t')] \end{aligned}$$

As $\varepsilon \rightarrow 0$, the subdomain $\|x-z\| > \varepsilon$ becomes the complete domain.

Therefore, with the consideration of both $\|x-z\| \leq \varepsilon$ and $\|x-z\| > \varepsilon$, the generalized differential Chapman-Kolmogorov equation is

$$\begin{aligned} & \frac{\partial}{\partial t} \mathbf{p}(x, t|y, t') \\ &= -\text{dual} \sum_{i=1}^n \frac{\partial}{\partial x_i} [\mathbf{A}_i(x, t) \mathbf{p}(x, t|y, t')] \\ &+ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} [\mathbf{B}_{ij}(x, t) \mathbf{p}(x, t|y, t')] \\ &+ \int dz \mathbf{W}(x|z, t) \mathbf{p}(z, t|y, t') \\ & \quad - \text{dual} \int dz \mathbf{W}(z|x, t) \mathbf{p}(x, t|y, t') \end{aligned} \quad (22)$$

References

Auger A, Chatelain P, Koumoutsakos P (2006) R-leaping: accelerating the stochastic simulation algorithm by reaction leaps. *J Chem Phys* 125(8):084103. doi:10.1063/1.2218339
 Berry H (2002) Monte Carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation. *Biophys J* 83(4):1891–1901. doi:10.1016/s0006-3495(02)73953-2

- Blue J, Beichl I, Sullivan F (1995) Faster Monte Carlo simulations. *Phys Rev E* 51(2):R867–R868. doi:[10.1103/PhysRevE.51.R867](https://doi.org/10.1103/PhysRevE.51.R867)
- Bratsun D (2005) Delay-induced stochastic oscillations in gene regulation. *Proc Natl Acad Sci* 102(41):14593–14598. doi:[10.1073/pnas.0503858102](https://doi.org/10.1073/pnas.0503858102)
- Burrage K, Hancock J, Leier A, Nicolau DV (2007) Modelling and simulation techniques for membrane biology. *Brief Bioinform* 8(4):234–244. doi:[10.1093/bib/bbm033](https://doi.org/10.1093/bib/bbm033)
- Cai X, Xu Z (2007) K-leap method for accelerating stochastic simulation of coupled chemical reactions. *J Chem Phys* 126(7):074102. doi:[10.1063/1.2436869](https://doi.org/10.1063/1.2436869)
- Chatterjee A, Vlachos DG (2007) An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. *J Comput Aid Mater Des* 14(2):253–308. doi:[10.1007/s10820-006-9042-9](https://doi.org/10.1007/s10820-006-9042-9)
- Chatterjee A, Voter AF (2010) Accurate acceleration of kinetic Monte Carlo simulations through the modification of rate constants. *J Chem Phys* 132(19):194101. doi:[10.1063/1.3409606](https://doi.org/10.1063/1.3409606)
- Chatterjee A, Vlachos DG, Katsoulakis MA (2005) Binomial distribution based τ -leap accelerated stochastic simulation. *J Chem Phys* 122(2):024112. doi:[10.1063/1.1833357](https://doi.org/10.1063/1.1833357)
- Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. *Ann Math Stat* 38(2):325–339. doi:[10.1214/aoms/1177698950](https://doi.org/10.1214/aoms/1177698950)
- Dimitrova NS, Markov SM, Popova ED (1992) Extended interval arithmetics: new results and applications. In: Atanassova L, Herzberger J (eds) *Computer arithmetic and enclosure methods*. North-holland, Amsterdam, pp 225–232
- Dubois D, Prade HM (1988) *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, New York
- E W, Liu D, Vanden-Eijnden E (2007) Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales. *J Comput Phys* 221(1):158–180. doi:[10.1016/j.jcp.2006.06.019](https://doi.org/10.1016/j.jcp.2006.06.019)
- Ferson S, Kreinovich V, Ginzburg L, Myers DS, Sentz F (2003) *Constructing probability boxes and Dempster-Shafer structures*. Sandia National Laboratories, Albuquerque (SAND2002-4015)
- Gardeñes E, Sainz MÁ, Jorba L, Calm R, Estela R, Mielgo H, Trepát A (2001) *Reliab Comput* 7(2):77–111. doi:[10.1023/a:1011465930178](https://doi.org/10.1023/a:1011465930178)
- Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22(4):403–434. doi:[10.1016/0021-9991\(76\)90041-3](https://doi.org/10.1016/0021-9991(76)90041-3)
- Gillespie DT (2001) Approximate accelerated stochastic simulation of chemically reacting systems. *J Chem Phys* 115(4):1716. doi:[10.1063/1.1378322](https://doi.org/10.1063/1.1378322)
- Haseltine EL, Rawlings JB (2002) Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J Chem Phys* 117(15):6959. doi:[10.1063/1.1505860](https://doi.org/10.1063/1.1505860)
- Henkelman G, Jónsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. *J Chem Phys* 115(21):9657. doi:[10.1063/1.1415500](https://doi.org/10.1063/1.1415500)
- Katsoulakis MA, Majda AJ, Vlachos DG (2003) Coarse-grained stochastic processes for microscopic lattice systems. *Proc Natl Acad Sci* 100(3):782–787. doi:[10.1073/pnas.242741499](https://doi.org/10.1073/pnas.242741499)
- Kaucher E (1980) Interval analysis in the extended interval space IR. In: *Computing supplementa*, vol 2. pp 33–49
- Kierzek AM (2002) STOCKS: stochastic kinetic simulations of biochemical systems with Gillespie algorithm. *Bioinformatics* 18(3):470–481. doi:[10.1093/bioinformatics/18.3.470](https://doi.org/10.1093/bioinformatics/18.3.470)
- Kuznetsov VP (1995) Interval methods for processing statistical characteristics. In: *International workshop on applications of interval computations*, El Paso
- Maksym PA (1988) Fast Monte Carlo simulation of MBE growth. *Semicond Sci Technol* 3(6):594–596. doi:[10.1088/0268-1242/3/6/014](https://doi.org/10.1088/0268-1242/3/6/014)
- Markov S (1979) Calculus for interval functions of a real variable. *Computing* 22(4):325–337. doi:[10.1007/bf02265313](https://doi.org/10.1007/bf02265313)
- Mei D, Xu L, Henkelman G (2009) Potential energy surface of methanol decomposition on Cu(110). *J Phys Chem C* 113(11):4522–4537. doi:[10.1021/jp808211q](https://doi.org/10.1021/jp808211q)
- Molchanov IS (2005) *Theory of random sets. Probability and its applications*. Springer, London
- Möller B, Beer M (2004) *Fuzzy randomness: uncertainty in civil engineering and computational mechanics*. Springer, Berlin
- Neumaier A (2004) Clouds, fuzzy sets, and probability intervals. *Reliab Comput* 10(4):249–272. doi:[10.1023/B:REOM.0000032114.08705.cd](https://doi.org/10.1023/B:REOM.0000032114.08705.cd)
- Picioreanu C, van Loosdrecht MCM, Curtis TP, Scott K (2010) Model based evaluation of the effect of pH and electrode geometry on microbial fuel cell performance. *Bioelectrochemistry* 78(1):8–24. doi:[10.1016/j.bioelechem.2009.04.009](https://doi.org/10.1016/j.bioelechem.2009.04.009)
- Rathinam M, Petzold LR, Cao Y, Gillespie DT (2003) Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method. *J Chem Phys* 119(24):12784. doi:[10.1063/1.1627296](https://doi.org/10.1063/1.1627296)
- Schnell S, Turner TE (2004) Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Prog Biophys Mol Biol* 85(2–3):235–260. doi:[10.1016/j.pbiomolbio.2004.01.012](https://doi.org/10.1016/j.pbiomolbio.2004.01.012)
- Schulze T (2002) Kinetic Monte Carlo simulations with minimal searching. *Phys Rev E* 65(3):036704. doi:[10.1103/PhysRevE.65.036704](https://doi.org/10.1103/PhysRevE.65.036704)
- Shafer G (1976) *A mathematical theory of evidence*. Princeton University Press, Princeton
- Sörensen MR, Voter AF (2000) Temperature-accelerated dynamics for simulation of infrequent events. *J Chem Phys* 112(21):9599. doi:[10.1063/1.481576](https://doi.org/10.1063/1.481576)
- SPPARKS (2009) Kinetic Monte Carlo simulator. <http://www.cs.sandia.gov/~sjlimp/spparks.html>
- Tian T, Burrage K (2004) Binomial leap methods for simulating stochastic chemical kinetics. *J Chem Phys* 121(21):10356. doi:[10.1063/1.1810475](https://doi.org/10.1063/1.1810475)
- Trushin O, Karim A, Kara A, Rahman T (2005) Self-learning kinetic Monte Carlo method: application to Cu(111). *Phys Rev B* 72(11) doi:[10.1103/PhysRevB.72.115401](https://doi.org/10.1103/PhysRevB.72.115401)
- Voter A (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. *Phys Rev Lett* 78(20):3908–3911. doi:[10.1103/PhysRevLett.78.3908](https://doi.org/10.1103/PhysRevLett.78.3908)
- Voter A (1998) Parallel replica method for dynamics of infrequent events. *Phys Rev B* 57(22):R13985–R13988. doi:[10.1103/PhysRevB.57.R13985](https://doi.org/10.1103/PhysRevB.57.R13985)
- Walley P (1991) *Statistical reasoning with imprecise probabilities*. In: *Monographs on statistics and applied probability*, 1st edn. Chapman and Hall, London
- Wang Y (2010) Imprecise probabilities based on generalised intervals for system reliability assessment. *Int J Reliab Saf* 4(4):319. doi:[10.1504/ijrs.2010.035572](https://doi.org/10.1504/ijrs.2010.035572)
- Weichselberger K (2000) The theory of interval-probability as a unifying concept for uncertainty. *Int J Approx Reason* 24(2–3):149–170. doi:[10.1016/s0888-613x\(00\)00032-3](https://doi.org/10.1016/s0888-613x(00)00032-3)
- Zeng Y, Choo YF, Kim B-H, Wu P (2010) Modelling and simulation of two-chamber microbial fuel cell. *J Pow Sources* 195(1):79–89. doi:[10.1016/j.jpowsour.2009.06.101](https://doi.org/10.1016/j.jpowsour.2009.06.101)