Building a kinetic Monte Carlo model with a chosen accuracy

Vijesh J. Bhute, and Abhijit Chatterjee

Citation: The Journal of Chemical Physics **138**, 244112 (2013); doi: 10.1063/1.4812319
View online: https://doi.org/10.1063/1.4812319
View Table of Contents: http://aip.scitation.org/toc/jcp/138/24
Published by the American Institute of Physics

Building a kinetic Monte Carlo model with a chosen accuracy

Vijesh J. Bhute[1] and Abhijit Chatterjee[1,2,a)]
[1]*Department of Chemical Engineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India*
[2]*Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India*

(Received 1 April 2013; accepted 13 June 2013; published online 28 June 2013)

The kinetic Monte Carlo (KMC) method is a popular modeling approach for reaching large materials length and time scales. The KMC dynamics is erroneous when atomic processes that are relevant to the dynamics are missing from the KMC model. Recently, we had developed for the first time an error measure for KMC in Bhute and Chatterjee [J. Chem. Phys. **138**, 084103 (2013)]. The error measure, which is given in terms of the probability that a missing process will be selected in the correct dynamics, requires estimation of the missing rate. In this work, we present an improved procedure for estimating the missing rate. The estimate found using the new procedure is within an order of magnitude of the correct missing rate, unlike our previous approach where the estimate was larger by orders of magnitude. This enables one to find the error in the KMC model more accurately. In addition, we find the time for which the KMC model can be used before a maximum error in the dynamics has been reached. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4812319]

I. INTRODUCTION

The kinetic Monte Carlo (KMC) method[1–6] is well suited for studying activated processes that are rarely observed with the computationally expensive molecular dynamics (MD) technique.[7] Instead of studying the vibrational motion of atoms and molecules, as is done in MD, KMC can reach long timescales by randomly selecting a process from the catalog of processes for each state visited by the system during its state-to-state evolution. Hence, KMC can be used to bridge the large gap between atomic and experimental scales in several material systems with modest computational resources. At the heart of a KMC calculation is a "kinetic-map" of the underlying potential energy surface[8] (PES) that provides the catalog of the pathways (or processes) connecting one state (or energy basin) to other states along with the associated rate constants. This kinetic map is the KMC model (also known as the Markov state model). The KMC dynamics is accurate as long as all atomic processes and their rates are correctly known and the underlying assumption that the atomic processes are independent Poisson processes is satisfied. However, in many situations atomic processes can be missing from the KMC model. This is a source of error in the KMC dynamics.

In order to address this problem, recently, attempts to construct an "accurate" KMC model for different materials have been made by starting with an interatomic potential or a quantum mechanical representation and using basin escape pathway search (BEPS) techniques to find atomic processes from different basins visited by the system. Atomic processes can be sought either using molecular statics BEPS techniques such as minimum energy path and mode following methods,[9–15] or using dynamical BEPS approaches such as MD[16–19] and accelerated MD[20–24] methods that follow the actual dynamics of the system. Unfortunately, processes that occur at timescales larger than those accessible to BEPS calculations cannot be sought using dynamical BEPS. A similar situation can arise in molecular statics BEPS because of limited sampling. Processes that are missing from the KMC model can be sought by performing additional BEPS calculations, however, it is not straightforward to guess how long these BEPS calculations need to performed.

Recently, we developed a computational procedure in Ref. 1 to address two aspects, namely, (i) finding the error in the KMC dynamics when an incomplete catalog of processes is being used and (ii) finding the time after which additional BEPS calculations are required. The underlying philosophy of our approach is that not all processes are *relevant* to the dynamics and that a KMC model should be created such that it contains all the relevant processes. The relevance of a process from a particular state is given by the probability of observing the process in the correct dynamics. As it will become evident later, this probability depends on the process rates and the timescales that are being accessed. Missing processes that are unlikely to be observed in the dynamics will not affect the accuracy significantly. On the other hand, the KMC dynamics can be incorrect when relevant processes are missing. A process can become more relevant at longer times where the probability of observing it is higher. The accuracy of a catalog can be maintained by seeking missing processes and ensuring that the probability of observing the missing processes from a state in the correct dynamics remains small. As discussed later, this introduces a timescale, called the validity time for a catalog, for which a fixed catalog of known processes from a state can be used with a chosen accuracy. The validity time for a catalog depends on the sum of missing rates from the catalog, which unfortunately is not known to us. The correctness of the error measure and the catalog validity time depends largely on our ability to accurately estimate

a)Author to whom correspondence should be addressed. Electronic mail: abhijit@che.iitb.ac.in

the missing rate. The estimate developed in Ref. 1 was found to be larger than the correct missing rate by orders of magnitude. In this work, we present an improved estimate that is within an order of magnitude of the correct missing rate. This enables us to find the error associated with dynamical BEPS-based KMC models as shown in this work and the error associated with standard KMC and self-learning local environment KMC models[15,25–27] as will be demonstrated in future publications.

The paper has been divided into the following sections. In Sec. II, mathematical expressions to obtain the relevance of a process and the error associated with the KMC dynamics are derived. In Sec. III, we present our improved procedure to estimate the missing rate in a catalog. The estimate is assessed by working with test catalogs that are completely known to us at outset. In Sec. IV, we study a test catalog where process timescales are overlapping. This represents a more realistic picture of KMC models used in literature. Finally, conclusions are provided in Sec. V.

II. ERROR IN A KMC MODEL

In this section, we describe the procedure for building a KMC model with a chosen accuracy. First, we discuss about the relevance of atomic processes in the KMC dynamics and accuracy of a catalog of processes from a basin. Next, we discuss the error in a KMC model comprising of several basins. We describe our original method in Ref. 1 to find the error associated with a catalog that has been constructed using dynamical BEPS.

A. Relevance of missing processes in the correct dynamics

Consider a basin B in the PES. Some of the processes from B might be already known to us. Alternatively, we can create a catalog of processes using BEPS as described later in Sec. II C. Let C_K denote the catalog of *known* processes from B. The catalog of *missing* or *unknown* processes is denoted C_U. The complete catalog of pathways for B is given by $C_C = C_K \cup C_U$. Similarly, catalogs of known processes for other basins in the PES can be created. The KMC model, as shown schematically in Fig. 1, contains a list of states and their process catalogs. Next, we focus on the error associated with the catalog C_K for basin B.

The correct dynamics is obtained when the complete catalog C_C is used with the KMC method. Assuming the atomic processes to be independent Poisson processes, the probability density associated with the first escape involving a process from the catalog C_U in the correct dynamics is given by[3,28]

$$p_U(\tau) = k_U \exp(-k_U \tau). \quad (1)$$

Here, k_U denotes the missing rate, i.e., the sum of rates in C_U. The probability P_U that at least one of the processes from C_U will be observed during time τ_B is obtained by integrating Eq. (1) and is given by

$$P_U(\tau_B) = 1 - e^{-k_U \tau_B}. \quad (2)$$

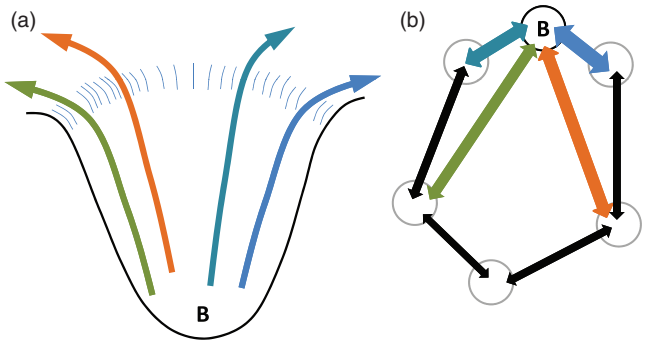

FIG. 1. (a) Schematic of a basin denoted B in the potential energy surface (PES). Consider the case where four atomic processes from the basin B are found by performing basin escape pathway search (BEPS). (b) By repeating this procedure for other basins, we obtain a "kinetic map" of the PES, which we will term the KMC model. This KMC model consists of a list of states and atomic processes found. In addition, the process rates, the time for which the system resides in each basin, and the number of times each process has been witnessed in the BEPS calculations are also stored.

Equation (2) suggests that the probability P_U is small when $k_U \tau_B \ll 1$. Therefore, processes in C_U are not relevant to the dynamics at KMC time τ_B as these processes have a low probability of being selected. In such a situation, the catalog C_K is deemed to contain all relevant processes. The probability P_U can be significant when $k_U \tau_B$ is large. Such situations can be avoided by ensuring that the value of P_U remains less than a maximum error δ. From Eq. (2) we require that $\tau_B \leq -\ln(1 - \delta)/k_U$. The time for which C_K can be employed with KMC while the error is less than δ is called the validity time τ_V for catalog C_K, i.e.,

$$\tau_V = -\frac{\ln(1-\delta)}{k_U}. \quad (3)$$

For instance, when δ is chosen to be 0.01, the catalog validity time $\tau_V \approx 0.01/k_U$. The validity time is small when one or more unknown processes have large rate constants, i.e., k_U is large. The average time required to observe the unknown processes is $1/k_U$. The numerator $-\ln(1 - \delta)$ in Eq. (3) accounts for the fact that an unknown process can be selected in the correct dynamics at times smaller than $1/k_U$.

It should be noted that the error δ and the validity time depend only on the missing rate k_U and not the individual rates in C_U. Unfortunately, k_U is not known to us. We will find an estimate for k_U given by $\tilde{k}_U$, such that $\tilde{k}_U \geq k_U$. Using the rate estimate in Eq. (3) we obtain

$$\tilde{\tau}_V = -\frac{\ln(1-\delta)}{\tilde{k}_U}. \quad (4)$$

Since the validity time $\tilde{\tau}_V \leq \tau_V$, the error associated with C_K will be less than δ.

So far we have derived expressions for the error (Eq. (2)) and the validity time (Eq. (4)) associated with the catalog C_K. Next, we find the error associated with the KMC model comprising of several states, catalogs of known processes, and validity times.

B. Error in the KMC model

Consider a dynamical path denoted as $\pi = \{B_1, B_2, \ldots, B_q\}$ connecting two basins B_1 and B_q in the PES. The path can be constructed such that the system visits a basin more than once. We know from Eq. (2) that the probability of observing at least once a process with rate constant k_i in the dynamical path in time τ_i is given by $1 - \exp(-k_i\tau_i)$. Since the basin-to-basin transitions are independent, the probability that the escapes occur within the times $\{\tau_1, \tau_2, \ldots, \tau_{q-1}\}$ is given by $\prod_{j=1}^{q-1} 1 - \exp(-k_j\tau_j)$. When $k_j\tau_j \gg 1, j = 1, 2, \ldots, q - 1$, the probability associated with the sequence of escapes is close to 1, i.e., the path is relevant to the dynamics. One the other hand, the probability is less than δ as long as the time τ_j is smaller than the validity time $\tau_{V,j}$ for all basins B_j in the path and the dynamical path comprises of one or more missing processes. As long as the maximum error in the catalog of known processes for each basin is δ, all dynamically relevant paths at the current timescales can be constructed using the KMC model. The error in the KMC model is bounded by δ. The KMC model needs to be updated by seeking missing processes when longer timescales are accessed. It is possible that several states might be missing from the KMC model. Many of these states are not relevant to the KMC model at the current timescales as the system will not visit these states. However, it is possible that some states are added to the KMC model as it is being updated. Such states have zero validity time when the catalog of known processes is empty, i.e., BEPS calculations will be required to move from these states to other states of the system.

The systematic procedure described so far forms the mathematical basis for maintaining the accuracy of a KMC model. Clearly, our approach is different from the standard approaches used in literature. Most KMC models assume a fixed catalog of processes. According to our procedure the error in the KMC model is determined by the largest catalog error for any of the states relevant to the dynamics. We expect this error to be significant in many situations. Even though our approach is computationally expensive, it represents a major advance over standard KMC models. One can conceive ways in which dynamical trajectories from BEPS techniques are used to build a KMC model. Using our approach the error associated with the model is obtained. Once enough validity time has been accrued, the KMC model can be reused to generate multiple dynamical trajectories with low computational overhead. Readers familiar with local environment KMC models[15,25–27] would realize that our approach will be particularly promising for accurately generating such models. However, this will be the subject of a future publication. Next, we describe how one can construct a KMC model using information collected from dynamical BEPS.

C. Building a KMC model using dynamical BEPS

We again consider the basin B in the PES. None of the processes from B are known to us when the basin is visited for the first time. A BEPS technique such as MD with a thermostat can be employed to obtain a sequence of n_{esc} escapes from the current basin B. Here, n_{esc} is an integer value. Each time an escape occurs from B the system is returned to the basin and a new escape is sought. The goal is to obtain from this sequence a KMC catalog and the associated error. When a process is observed for the first time, the new state of the system and the rate constant is recorded in C_K. In addition, the number of times the process is observed with BEPS is also recorded, as this will be used to estimate the missing rate.

Let t_B denote the total time elapsed over all dynamical BEPS calculations performed in the basin, i.e., t_B is the residence time in B. One expects that k_U will become smaller as time t_B increases. The randomness inherent in BEPS implies that two catalog generation attempts for the same basin can result in different values of k_U even though the time t_B might be same for the two catalogs. An advantage of dynamical BEPS methods, like MD and accelerated MD, is that they follow the correct escape times from the basin. Generally, processes with large rates will be observed first with dynamical BEPS, while slower processes will be observed later. Hence, dynamical BEPS techniques provide a systematic way of searching for processes based on their rates. We consider this to be an advantage of dynamical BEPS techniques. The orders of magnitude computational speed-up of accelerated MD techniques over standard MD renders them ideal for generating a sequence of escapes from the basin. In the rest of this work, we focus on dynamical BEPS based KMC models. Hereafter, the term BEPS is used to refer to dynamical BEPS techniques.

A question that arises is how the validity time $\tilde{\tau}_V$ is related to the BEPS time t_B. Although one would expect that $\tilde{\tau}_V < t_B$, in order to answer this question we need to find how $\tilde{k}_U$ depends on t_B. In Ref. 1, we obtained $\tilde{k}_U$ as follows. Since $k_U = k_C - k_K$, where k_C and k_K are the sum of rates in the complete and known catalogs, respectively, we estimated the value of k_C as m'/t'_B using maximum likelihood estimation, such that the estimate $\tilde{k}_C \geq k_C$. Here, m' and t'_B are related to the number of escapes from B and the time t_B, respectively (see Ref. 1 for more details). Unfortunately, we found that the maximum likelihood estimate converges slowly towards correct value of k_C as the number of escapes from the basin increases. As a result, $\tilde{k}_U = \tilde{k}_C - k_K$ can be as large as k_C, even though k_U might be orders of magnitude smaller. The calculated validity time is found to be much smaller than it should be. Next, we develop an improved estimate for the missing rate which enables better estimation of the error and the catalog validity time.

III. IMPROVED ESTIMATE FOR THE MISSING RATE

In this section, we develop an improved estimate for the missing rate. The schematic in Fig. 2 is useful for understanding the procedure. The x-axis denotes the time spent in the basin. Several escapes might have been observed during this time. Known processes are grouped along x-axis into spectral bands according to their average escape times, i.e., inverse of rate constants. The bars denote spectral bands. Processes are numbered based on when they were first observed. Fast processes (e.g., process index 1, 2, and 3 in the blue vertical bar), which occur at shorter timescales, are observed many more times than slower processes (e.g., process 7 and 8 in the green and brown bars, respectively). When the width of the bar is

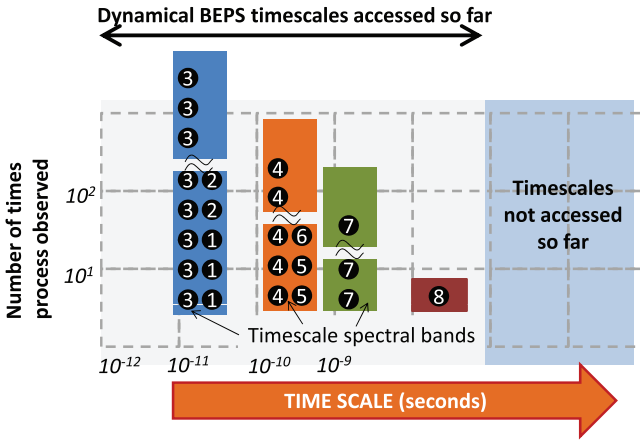

FIG. 2. Schematic of number of times processes are observed when BEPS calculations are performed in a particular basin. The x-axis denotes the time t_B spent in the basin with BEPS. Numbers inside the colored vertical bands denote the process index; multiple sightings of a process are represented by multiple circles. It is likely that processes belonging to the inaccessible timescales and some of the processes from the accessible timescales are missing from the catalog.

small, processes belonging to a band have similar average escape times and we expect them to be observed similar number of times with BEPS. Most spectral bands lie within the total time t_B elapsed in the BEPS calculation for the basin. Few processes with average escape times greater than t_B might be observed with BEPS. Such processes will lie in the shaded region corresponding to times that have not been accessed so far. The spectral bands for such processes are not shown in Fig. 2.

The contributions to the missing rate from the accessible and inaccessible timescales are estimated separately, i.e.,

$$\tilde{k}_U = \tilde{k}_{U,\text{accessible}} + \tilde{k}_{U,\text{inaccessible}}. \tag{5}$$

Processes can be missing from a particular band due to inherent randomness in BEPS even though BEPS is accessing time scales where these processes would occur. The estimate for these missing processes is given by $\tilde{k}_{U,\text{accessible}}$. Processes that occur at timescales that have not been accessed by BEPS are likely to be missing. The estimate for such processes is given by $\tilde{k}_{U,\text{inaccessible}}$.

In Secs. III A and III B, we develop a procedure to obtain $\tilde{k}_{U,\text{inaccessible}}$ and $\tilde{k}_{U,\text{accessible}}$, respectively. We shall employ KMC as a BEPS technique to select processes from a complete catalog C_C that is known to us *a priori*. This will allow us to construct an incomplete catalog C_K and assess the estimate for the missing rate. A basin escape occurs in KMC-based BEPS by randomly selecting a process from the catalog C_C with a probability proportional to its rate constant and advancing the time t_B by

$$\Delta t_B = -\frac{\ln \xi}{k_C}. \tag{6}$$

Here, ξ is a uniform random deviate.

Before we proceed to derive the rate estimate we describe how the accessible timescales are partitioned into spectral bands. As we shall show later $\tilde{k}_{U,\text{accessible}}$ is obtained using the approximation that all rates in a spectral band are identical. When the width of a band is large, the large variation in the rates in the band causes this approximation to become invalid. On the other hand, when the width is small, few processes will be present in each band and a large number of bands will be present. This can result in accumulation of error due to the noise present in the BEPS data. We define the width w of the bth spectral band in terms of the largest rate $k_{\max,b}$ and smallest rate $k_{\min,b}$ in band b, such that

$$\frac{k_{\max,b}}{k_{\min,b}} \leq w. \tag{7}$$

One can partition known processes into bands by starting with the largest rate constant $k_{\max,1}$ in C_K in the first band, finding $k_{\min,1}$ using Eq. (7), and then proceeding to other spectral bands with smaller rate constants. More details of the algorithm are given in Sec. IV. In this work, we have chosen $w = 5$. The averaged rate constant associated with the band is given by

$$k_b = \frac{\sum_{\mu \in b} k_\mu}{n_{pb}}. \tag{8}$$

Here, k_μ is the rate constant for a known process μ in the bth band and n_{pb} denotes the number of known processes from the band. The process rates in a band are replaced by the average rate k_b while estimating $\tilde{k}_{U,\text{accessible}}$.

A. Rate estimate for inaccessible timescales

Suppose all missing processes belonged to the inaccessible timescales. The probability P'_U that no process from C_U has been observed during time t_B from Eq. (1) is given by

$$P'_U = e^{-k_U t_B}. \tag{9}$$

The probability is independent of the individual process rates in C_U and depends only on k_U. P'_U is close to one when $k_U t_B \ll 1$. On the other hand, P'_U is small when $k_U t_B \gg 1$, i.e., it is likely that one of the processes from C_U would have been observed in BEPS during time t_B. In such a case, the catalog C_U can no longer contain only unknown processes, which is contrary to the original definition of C_U. We require that the probability P'_U of not observing processes from catalog C_U to be significant, such that it is greater than α, i.e.,

$$\alpha \leq e^{-k_U t_B}. \tag{10}$$

From Eq. (10), the upper bound for the missing rate is given by

$$\tilde{k}_{U,\text{inaccesible}} = -\frac{\ln \alpha}{t_B}. \tag{11}$$

In this work we have employed $\alpha = 4.54 \times 10^{-5}$, i.e., $\tilde{k}_{U,\text{inaccesible}} = 10/t_B$. The small value of α accounts for the rare situation some of the processes should have been observed by now, but they are still missing.

This point will become clear as we study a catalog C_C with one process with rate $k = 10^9\ s^{-1}$ (see Fig. 3). The result from one of the catalog generation calculations is shown by the black line. When $k t_B \ll 1$, Eq. (11) provides an estimate $\tilde{k}_{U,\text{inaccessible}} \gg k$. This suggests that C_U can have one or more missing processes such that the missing rate is greater

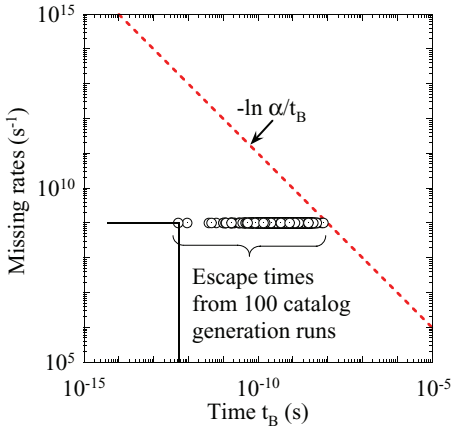

FIG. 3. Missing rate from a catalog C_K for a basin that contains one process with rate constant 10^9 s^{-1}. The catalog C_K is initially empty. Processes from the basin are found by sampling escape pathways from the basin. The rate estimate for the inaccessible timescales (dashed line; Eq. (11)) decreases as the time t_B spent in the basin increases. Circles denote the time at which the process was first observed (shown for 100 independent catalog generation calculations). The solid line denotes the correct unknown rate for one such catalog.

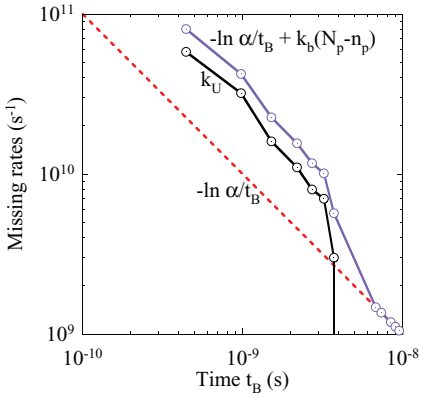

FIG. 4. Missing rate for a basin with $N_p = 100$ processes. Each process has a rate constant 10^9 s^{-1}. Missing processes are observed as time t_B elapsed in the basin increases. The rate estimate from the inaccessible timescales (dashed line; Eq. (11)) is smaller than the correct unknown rate k_U. When contributions from the accessible timescale spectral band are added to the ones from the inaccessible timescales the rate estimate (blue line) is very close to k_U. Symbols denote the times when the rates were computed. Here, n_p denotes the number of known processes.

than k (in this case we already know $k_U = k$). After some-time the process is observed with BEPS and the value of k_U becomes 0. This is shown by the step change in the value of k_U in Fig. 3. The escape occurs at dimensionless time $kt_B = 5 \times 10^{-3}$ even though the probability of observing the process at this time is small. As we see in Fig. 3, Eq. (11) still provides an estimate $\tilde{k}_U \gg k$. The circles in Fig. 3 denote the time at which the process is observed for the first time with 99 other catalog generation calculations. In few cases the escape occurs after the average escape time 1/k yet $\tilde{k}_{U,\text{inaccesible}} > k$ because of the chosen value of α in Eq. (11).

B. Rate estimate for a spectral band

Once a process from a spectral band is known, we focus on the missing processes from the spectral band. We begin by considering an example of a catalog C_C with $N_p = 100$ processes. All processes have a rate constant $k = 10^9$ s^{-1}. The average time required to observe one of the N_p processes is given by $1/kN_p = 10^{-11}$ s. Figure 4 shows results from a catalog generation calculation. The value of k_U is shown by the black line. The symbols denote the rate after every $n_{esc} = 50$ escapes from the basin. The estimate from Eq. (11) is shown by the dashed line. At short times, none of the processes have been observed and Eq. (11) can be used. As time t_B increases we begin observing processes with BEPS. After 50 escapes, $n_p = 42$ processes are known, the time $t_B = 4.4 \times 10^{-10}$ s, and the estimate from Eq. (11) is smaller than the correct value of k_U by a factor of 2.57. Clearly, $\tilde{k}_{U,\text{inaccessible}}$ does not account for the total missing rate. We find missing processes even after 200 escapes. The last process is observed at time $t_B = 5.3 \times 10^{-9}$ s which is greater than the average time for a particular process given by $1/k = 10^{-10}$ s.

The rationale behind the estimate $\tilde{k}_{U,\text{accessible}}$ is as follows. Suppose n_p processes of the N_p processes belonging to the accessible band have been observed with BEPS. We can write

$$\tilde{k}_U = -\frac{\ln \alpha}{t_B} + k_b(N_p - n_p), \tag{12}$$

where $k_b = k$ for this catalog. Equation (12) is plotted in Fig. 4 (blue line). The contribution from the accessible band is larger than the one from the inaccessible band. Obviously, Eq. (12) overestimates the missing rate by $-\ln \alpha/t_B$. As we have already witnessed in Sec. III A, the term $-\ln \alpha/t_B$ would have been important to account for the missing processes in the inaccessible timescales. Once all processes have been observed, $\tilde{k}_U$ becomes $-\ln \alpha/t_B$ in Eq. (12). Although this example has illustrated the importance of the contributions from the accessible band to k_U, the value of N_p is not known to us in the first place. Next, we develop a procedure to estimate N_p.

Let m_t denote the total number of escapes observed with BEPS. Assuming that the rates in a band are identical, the probability of observing a particular process i is given by $p_i = 1/N_p$. When N_p is large the probability of selecting a particular process becomes small. The probability of observing m_i escapes with the ith process, $i = 1,2,\ldots, N_p$, is given by the multinomial distribution

$$P(m_1, m_2, ..., m_{Np}; m_t) = \frac{m_t!}{\prod_{i \in C_C} m_i!} \prod_{i \in C_C} p_i^{m_i}. \tag{13}$$

Equation (13) involves a constraint on $\{m_i\}$, $i \in C_C$, namely,

$$\sum_{i \in C_C} m_i = m_t. \tag{14}$$

Assuming that the number of escapes for a particular process i is independent of other processes, i.e., Eq. (14) is not required, the distribution for the number of escapes m_i is given by the binomial distribution

$$P(m_i; m_t) = \frac{m_t!}{m_i!(m_t - m_i)!} p_i^{m_i}(1 - p_i)^{m_t - m_i}. \tag{15}$$

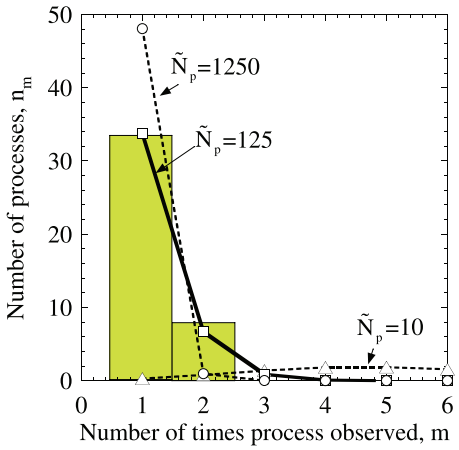

FIG. 5. Number of processes observed m times when the catalog was generated in Fig. 4 after a total of $m_t = 50$ escapes from the basin. The value of $\tilde{N}_p$ that results in the least sum of squared error with respect to the data from the histogram is regarded as an estimate for the number of processes in a spectral band.

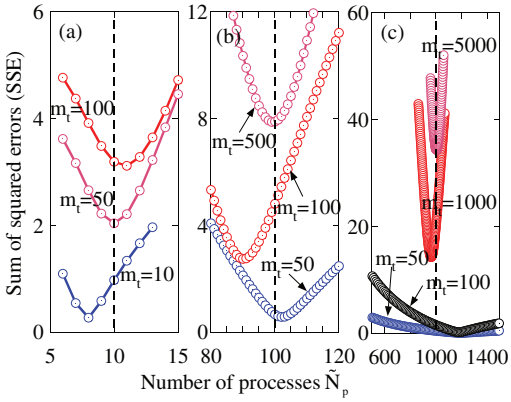

FIG. 6. Sum of squared errors (SSE) obtained from a catalog being generated. Three complete catalogs with the number of processes N_p being (a) 10, (b) 100, and (c) 1000 processes were considered (N_p is indicated by the dashed vertical line). All processes have identical rate constant given by k = 10^9 s^{-1}. SSE is plotted for different values of the parameter $\tilde{N}_p$ in Eq. (17). The value of $\tilde{N}_p$ that gives the smallest SSE is the best estimate for the number of processes in the spectral band.

Since $p_i = 1/N_p$, Eq. (15) is rewritten as

$$P(m_i; m_t, N_p) = \frac{m_t!}{m_i!(m_t - m_i)!}\left(\frac{1}{N_p}\right)^{m_i}\left(1 - \frac{1}{N_p}\right)^{m_t - m_i}. \tag{16}$$

Rewriting Eq. (16) in terms of an estimate for number of processes in the band $\tilde{N}_p$, we obtain

$$p(m; m_t, \tilde{N}_p) = \frac{m_t!}{m!(m_t - m)!}\left(\frac{1}{\tilde{N}_p}\right)^{m}\left(1 - \frac{1}{\tilde{N}_p}\right)^{m_t - m}. \tag{17}$$

Note that the subscript i in m_i has been ignored as it applies to all processes in the catalog.

Figure 5 shows a histogram for the number of processes n_m observed m times from the catalog generation calculation in Fig. 4 after $m_t = 50$ escapes. From Eq. (17) the number of processes with m escapes is given by

$$\tilde{n}_m = \tilde{N}_p p(m; m_t, \tilde{N}_p),\ m = 1, 2, \ldots. \tag{18}$$

The parameter $\tilde{N}_p$ is determined using least squared estimation, i.e., by finding the value of $\tilde{N}_p$ that minimizes the sum of squared error (SSE):

$$\min_{\tilde{N}_p} \sum_m \left(n_m - \tilde{N}_p \frac{m_t!}{m!(m_t - m)!}\left(\frac{1}{\tilde{N}_p}\right)^{m}\left(1 - \frac{1}{\tilde{N}_p}\right)^{m_t - m}\right)^2. \tag{19}$$

Note that the sum is performed over the number of escapes m. In order to perform a numerical fit we require at least two values of m for which $n_m > 0$. The plot for $\tilde{n}_m$ for three values of $\tilde{N}_p$, namely, 10, 125, and 1250, is shown in Fig. 5. The binomial distributions with $\tilde{N}_p = 10$ and 1250 are in poor agreement with the BEPS data. Despite the noise present in the BEPS data we find that the best estimate ($\tilde{N}_p = 125$) is close to the correct value ($N_p = 100$) with only $m_t = 50$ escapes. This indicates that ignoring Eq. (14) might be a reasonable approximation.

Figure 6 shows the plot for SSE in terms of $\tilde{N}_p$ for three different complete catalogs. All processes have rate constants k = 10^9 s^{-1} but the number of processes in the catalogs is given by N_p = 10, 100, and 1000 for panels (a), (b), and (c), respectively. The value of $\tilde{N}_p$ that gives in the least value of SSE after a chosen number of escapes m_t is the best estimate. In Figs. 6(a)–6(c) we find once again that despite the noise present in n_m and the approximations involved in Eq. (19), $\tilde{N}_p$ is close to correct value of N_p. In Fig. 6(a) where N_p = 10, only n_p = 6 processes were observed with m_t = 10, yet the estimate $\tilde{N}_p$ is found to be 8. $\tilde{N}_p$ is 11 and 10 for m_t = 50 and 100, respectively. A better estimate for $\tilde{N}_p$ is obtained when m_t is large. When N_p = 100, the optimum value for $\tilde{N}_p$ is found to be 18 with m_t = 10 escapes (not shown). We obtain $\tilde{N}_p$ = 102 with m_t = 50 although only 40 processes were observed. Similarly when m_t = 100, 62 processes have been observed and $\tilde{N}_p$ = 90. When m_t = 500, all the processes in the band were observed and $\tilde{N}_p$ = 99. Based on these observations we choose n_{esc} = 50, i.e., least squared estimation of $\tilde{N}_p$ is performed after every 50 escapes. A more extensive study of estimation of $\tilde{N}_p$ for the complete catalog in Fig. 6(b) is performed in Fig. 7. The histogram was obtained from 1000 catalogs C_K generated from the complete catalog after m_t escapes. Although in some situations $\tilde{N}_p$ is significantly different from N_p, we find that the peak of the histogram lies at $\tilde{N}_p$ = 100. It is clear that the noise in n_m plays a role in the value of $\tilde{N}_p$. It is observed that the width of the histogram decreases as m_t increases.

When a single band is present, Eq. (12) is rewritten as

$$\tilde{k}_U = -\frac{\ln \alpha}{t_B} + k_b(\max(\tilde{N}_p, n_p) - n_p). \tag{20}$$

The contribution from the accessible timescales will be zero in Eq. (20) when $\tilde{N}_p \le n_p$.The average rate constant for the band k_b from Eq. (8) is employed so that the numerical scheme can be used when the rate constants vary within a spectral band.

We expect that as processes from the band are observed more often, beyond a certain time t_b there should be no missing process from the band. The probability of missing a process from the band in BEPS is given by $\exp(-k_b t_b)$. We

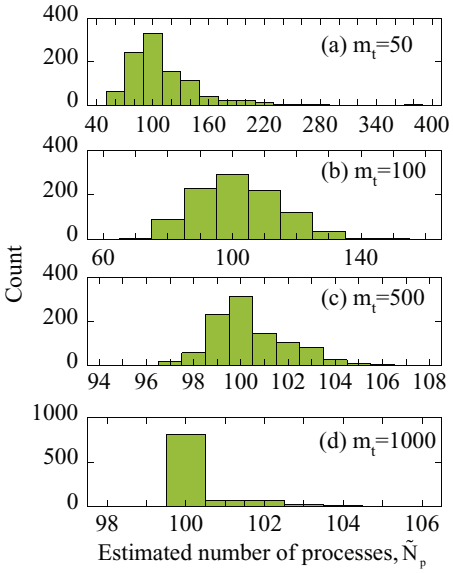

FIG. 7. Histogram for $\tilde{N}_p$ from different catalogs generated for a basin that contains $N_p = 100$ processes, each having a rate constant $k = 10^9$ s^{-1}. Histogram was obtained after m_t number of escapes occurred.

require this probability to be smaller than α. The time t_b after which the band b stops contributing to the missing rate is given by

$$t_b = -\frac{\ln \alpha}{k_b}. \tag{21}$$

Thus, all processes are deemed to be known once $t_B \geq t_b$. In other words,

$$\tilde{k}_U = \begin{cases} -\ln\alpha/t_B + k_b(\max(\tilde{N}_p, n_p) - n_p), & t_B \leq -\ln\alpha/k_b \\ -\ln\alpha/t_B, & t_B > -\ln\alpha/k_b \end{cases}. \tag{22}$$

In Fig. 8(a) we again study the catalog C_K in Fig. 4. The estimate from Eq. (20) is shown in blue circles after every 50 escapes. Note $\tilde{N}_p$ is estimated each time $\tilde{k}_U$ needs to be computed. The noise present in $\tilde{N}_p$ causes $\tilde{k}_U$ to fluctuate. Since $\tilde{k}_U$ is used to obtain $\tilde{\tau}_V$, the new value of $\tilde{k}_U$ is accepted only when it is smaller than the current value so that $\tilde{\tau}_V$ never decreases. This results in step increase in the validity time in Fig. 8(c). It is observed that Eq. (20) can reasonably estimate the missing rate even though the value of $\tilde{N}_p$ was less than 100 in Fig. 6(a) in some cases. All processes in the catalog C_C were observed within 500 escapes, however, $\tilde{k}_U$ remains non-zero because of contributions from the inaccessible timescales.

We study a complete catalog with $N_p = 1000$ processes in Fig. 8(b). Once again we find that Eq. (22) can provide a reliable estimate of the missing rate. The number of missing processes at 10^{-9} s (which corresponds to the average escape time for a process) is 32 and 394, for Figs. 8(a) and 8(b), respectively. It is also observed that in some cases $\tilde{k}_U$ is slightly less than k_U. Equation (22) is used with Eq. (4) to obtain the validity time for a catalog. The validity time for catalogs with 100 and 1000 processes can be completely different as shown in Fig. 8(c). The validity time of the catalog in Fig. 8(a) is larger than the validity time for the catalog in Fig. 8(b) because fewer processes are missing in the former catalog. After Eq. (21) is satisfied, $\tilde{k}_U$ contains contributions only from the inaccessible timescales and both catalogs have the same validity time.

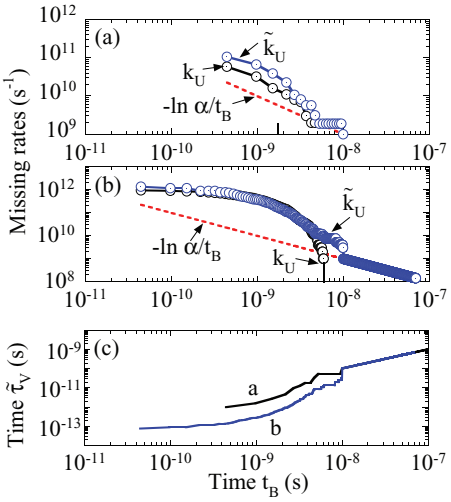

FIG. 8. Estimate for the missing rate for a catalog generated with BEPS. The catalog contains (a) $N_p = 100$ and (b) 1000 processes. Each process has a rate constant $k = 10^9$ s^{-1}. Black line in panels (a) and (b) denotes the correct unknown rate k_U. Red dashed and blue solid lines denote rate estimate from Eqs. (11) and (22), respectively. Symbols denote the time when the rate estimate was obtained. The estimate is non-zero even though k_U becomes zero after some time. (c) Validity times for catalog generated in panels (a) and (b) using $\delta = 0.1$.

C. Extension to multiple spectral bands

Next, we study complete catalogs where two spectral bands are present. Each band b in the catalog contains $N_{pb} = 100$ processes. The catalog is chosen such that all processes in first band have a rate constant given by k_1, while processes in the second band have a rate k_2. Figure 9(a) shows results from a catalog generation attempt with $k_1 = 10^9$ s^{-1} and $k_2 = 10^6$ s^{-1}. A large spectral gap is present between the two bands. Processes with rates k_1 are observed first with BEPS. The number of processes in the first band $\tilde{N}_{p1}$ are estimated using least squared estimation. Once $t_B > -\ln\alpha/k_1$, the contribution from the first band becomes zero, however, the $\tilde{k}_{U,\text{inaccessible}}$ contributes to $\tilde{k}_U$. Eventually, a process from the second band is observed at time $t_B = 1.38 \times 10^{-8}$ s indicating the presence of a second spectral band. Equation (22) is rewritten as

$$\tilde{k}_U = -\frac{\ln\alpha}{t_B} + \sum_{b=1}^{N_b} k_b(\max(\tilde{N}_{pb}, n_{pb}) - n_{pb})(1 - \Theta(t_B + \ln\alpha/k_b)) \tag{23}$$

to account for N_b number of spectral bands. The term with the Heaviside function $\Theta(t_B + \ln\alpha/k_b)$ indicates that the contribution from a spectral band is included as long as $t_B < -\ln\alpha/k_b$. Using Eqs. (4) and (23), the validity time for the catalog C_K is given by

$$\tilde{\tau}_V = \frac{-\ln(1-\delta)}{-\frac{\ln\alpha}{t_B} + \sum_{b=1}^{N_b} k_b(\max(\tilde{N}_{pb}, n_{pb}) - n_{pb})(1 - \Theta(t_B + \ln\alpha/k_b))}. \tag{24}$$

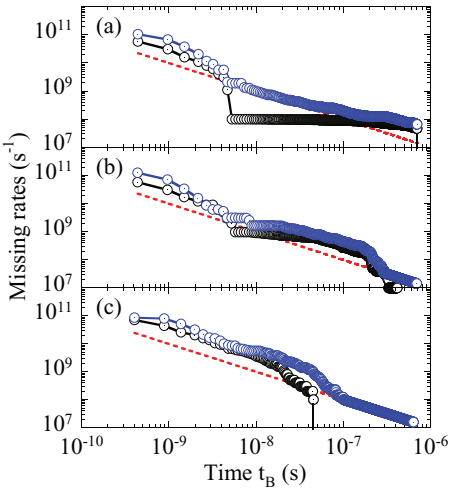

FIG. 9. Estimate for missing rate from a catalog C_K generated for a basin using BEPS. Two spectral bands are present. The first band contains $N_{p1} = 100$ processes with individual rate constant $k_1 = 10^9$ s^{-1}, while the second band contains $N_{p2} = 100$ processes with rate constant (a) $k_2 = 10^6$ s^{-1}, (b) $k_2 = 10^7$ s^{-1}1, and (c) $k_2 = 10^8$ s^{-1}. Black line denotes the correct unknown rate. Dashed red and solid blue lines denote estimate from Eqs. (11) and (23). Symbols denote the times where the unknown rate was computed.

As in Sec. III B the number of processes $\tilde{N}_{pb}$ from each band can be estimated using the least squared estimator in Eq. (19). A faster way of estimation is possible by realizing that in the limit where the number of escapes $n_b \to \infty$, one obtains

$$k_b \tilde{N}_{pb} = \lim_{t_B \to \infty} \frac{n_b}{t_B}, b = 1, 2, \ldots. \quad (25)$$

Here, n_b is the number of escapes from the bth band. The left-hand side gives the total rate from the bth band. Using Eq. (25) we obtain

$$\frac{n_b}{n_1} = \frac{k_b \tilde{N}_{pb}}{k_1 \tilde{N}_{p1}}, b > 1. \quad (26)$$

Here, $\tilde{N}_{pb}$ is the estimated number of processes in the band b. From Eq. (26) one can write

$$\tilde{N}_{pb} = \frac{k_1 n_b}{k_b n_1} \tilde{N}_{p1}. \quad (27)$$

When all processes from the first band are known, the number of processes from the bth band is found using Eq. (27). This procedure has been used for bands $b > 1$ and as shown in Fig. 9 it is found to work very well.

IV. ALGORITHM FOR GENERATING A KMC MODEL OF CHOSEN ACCURACY

Based on the discussion in Sec. III, we now present the algorithm for generating the validity time for a catalog of known processes from a basin that can be used to decide when additional BEPS calculations are required. The main steps involved in building a catalog C_K for a particular basin B are (i) seek processes from the basin using BEPS, (ii) categorize the known processes according to their rates to form the spectral bands, (iii) estimate missing rate for the current catalog C_K, and (iv) obtain the validity time for the catalog C_K based on the chosen accuracy. The details of the steps are as follows.

Step 1: Initialization step: Recover the catalog C_K containing the following escape information: known processes from B, their rate constants, number of times each process has been observed during past visits to the basin, and the time t_B spent in the basin. When the basin is visited for the first time, the catalog is empty and the validity time is zero.

Step 2: Search step: Perform n_{esc} BEPS calculations in the basin. Analyze the escapes that have occurred. Update the catalog C_K and total time spent in the basin t_B.

Step 3: Create spectral bands: Known processes are categorized into spectral bands based on their rates. Initially the number of bands N_b is zero. The bands are created in starting with the largest rates. First band to be created is denoted as $b = 1$. Following steps are required:

a. Select the fastest rate that has not been included in any band as $k_{max,b}$ for the new band b. The smallest rate $k_{min,b}$ in the band is found using Eq. (7).
b. Rates between $k_{max,b}$ and $k_{min,b}$ belong to the spectral band b. Count the number of processes n_{pb} belonging to band b.
c. Compute the average rate k_b using Eq. (8).
d. Create the next spectral band (Step 3a) if there are processes which have not been included in any band.

Step 4: Estimate missing rate from the accessible timescale: Initialize rate $\tilde{k}_U = 0$ and proceed with following steps for $b = 1,\ldots,N_b$, where N_b is the number of spectral bands.

a. If $t_B < -\ln \alpha / k_b$, go to step 4b. Otherwise, analyze the next band.
b. Estimate the total number of processes $\tilde{N}_{pb}$ from the band (using Eqs. (19) or (27)).
c. Increment $\tilde{k}_U$ by $k_b(\max(\tilde{N}_{pb}, n_{pb}) - n_{pb})$.

Step 5: Estimate missing rate from the inaccessible timescale: Increment $\tilde{k}_U$ by $-\ln \alpha / t_B$.

Step 6: Retain the previous value for $\tilde{k}_U$ in case the new estimate is smaller than the previous one for the basin.

Step 7: Compute the validity time $\tilde{\tau}_V$ for the catalog C_K using Eq. (4).

V. RATE ESTIMATE WHEN PROCESS TIME SCALES OVERLAP

Typically, KMC is used for material systems that are large in size and involve a large number of processes from each state of the system. The slowest processes in the system are less relevant as they occur at very large timescales. In such cases, some processes will always be missing from the catalog C_K and the situation in Sec. III A where only inaccessible timescales contribute to the missing rates will not arise. Furthermore, the rate constants often overlap with each other, i.e., there is no separation of process timescales. We shall demonstrate using a catalog with overlapping rate constants that the

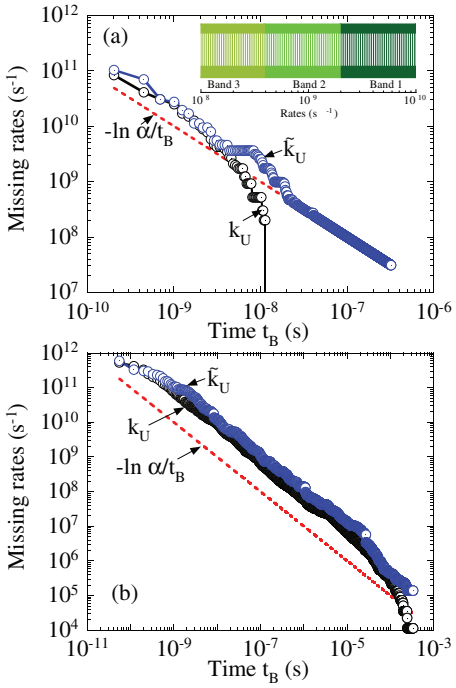

FIG. 10. Estimate for missing rate for a particular basin with (a) $N_p = 100$ and (b) $N_p = 1000$ processes. Rate constants and the three spectral bands to which the rates belong for panel (a) are shown in the inset by the vertical lines and the shaded area, respectively. Black line denotes the correct unknown rate. Dashed red and solid blue lines denote estimate from Eqs. (11) and (23).

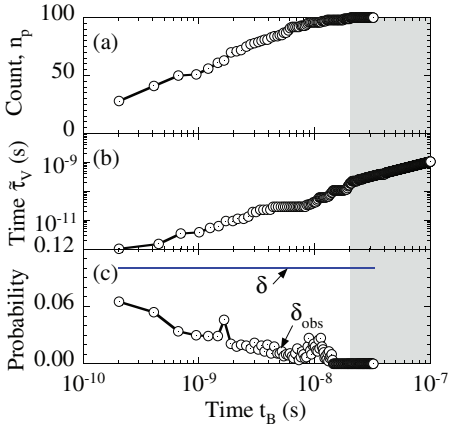

FIG. 11. (a) Number of processes observed for catalog in Fig. 10 with time t_B spent in the basin. All processes known once times in the shaded area are reached. (b) Validity time for the catalog as it is being generated. (c) Average observed error δ_{obs} is less than the target error $\delta = 0.1$. Averaging was performed over 100 independent KMC runs.

estimated rate constant shall remain within an order of magnitude of the correct missing rate.

We assess our procedure by studying a complete catalog C_C that contains 100 processes with rates given by a geometric series, such that the largest and the smallest rates in the complete catalog are 10^{10} and 10^8 s^{-1}, respectively. The process rates are shown in the inset of Fig. 10(a). Three spectral bands shown by the shaded regions in inset are found using $w = 5$ as the width of the spectral band. The first and second bands contain 35 processes. The third band contains the remaining processes. As evident, no clear-cut separation of timescales is present. Figure 10(a) shows estimate $\tilde{k}_U$ (blue line) for a catalog generated from C_C. The black line denotes the correct unknown rate k_U. It is observed that $\tilde{k}_U$ is greater than k_U. After the first 50 escapes have occurred the first band contains process rates in the range 10^{10} to 2×10^9 s^{-1}, while the second band contains only 5 processes in the range 1.96×10^9 to 3.92×10^8 s^{-1}. The number of processes in the catalog C_K changes with time as shown in Fig. 11(a). Figure 11(b) shows the validity time obtained for the catalog at different times t_B with 90% accuracy, i.e., $\delta = 0.1$. Eventually, all 100 processes have been found once the times shown in the gray shaded area are reached in Fig. 11. We believe that in realistic systems such a situation will never arise as the number of processes will be large and processes will span timescales that are not relevant to the dynamics. Examples of such behavior were demonstrated in our recent work on local environment dependence of rate constants in metal systems.[29] Generally, the contribution from the accessible timescales will dominate over the contributions from the inaccessible timescales. Hence, a proper assessment of the rate estimate can be made when contributions from the accessible timescales are present. As in Figs. 8 and 9, we find that the estimate is within one order of magnitude of the correct missing rate.

Figure 10(b) shows another example of a complete catalog C_C that contains 1000 processes with rates given by a geometric series; the largest and the smallest being 10^{10} and 10^4 s^{-1}, respectively. The gap between the rates is smaller than the one present in the rate catalog of Fig. 10(a). It is observed in Fig. 10(b) that once again the estimated missing rate is remarkably close to the correct missing rate. Next, we investigate in more detail the maximum error from the complete catalog in Fig. 10(a).

In order to verify that the maximum error is indeed given by δ we performed the following tests. The catalog C_K is deemed to be successful when all processes observed in the correct dynamics are already present in it. Otherwise, the catalog C_K is said to have failed. The probability of failure, given by Eq. (2), is less than δ as long as the KMC time is less than the catalog validity time. In Fig. 11(c), we perform numerical calculations to obtain the observed fraction of failures δ_{obs} associated with C_K. We performed 1000 catalog generation calculations with the complete catalog to find the number of times the generated catalog C_K failed for different validity times. It is found that δ_{obs} was less than $\delta = 0.1$. The value of δ_{obs} decreases as t_B increases because the contribution from the inaccessible band remains significant even though the correct missing rate is much smaller.

The observed fraction of failures δ_{obs} for 100 catalog generation attempts with three different target errors, namely, $\delta = 0.001$, 0.01, and 0.1, is shown in Fig. 12. The catalogs are generated after 100 escapes are observed with BEPS. For each catalog that was generated, 1000 KMC calculations performed with the complete catalog till the catalog validity time was reached. Figure 12 shows the histogram for δ_{obs}. We find that in some cases δ_{obs} is slightly greater than δ, which can be attributed to the noise in the BEPS data. The majority of times the catalogs were found to be safe, i.e., δ_{obs}, was less than δ. The number of catalogs (out of 100 catalogs) that resulted in

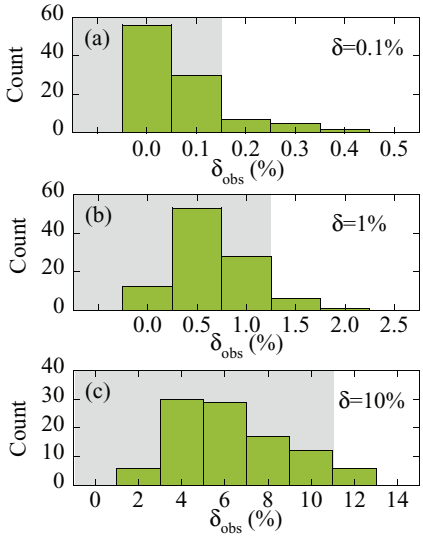

FIG. 12. Histogram for average observed error δ_{obs} for different values of target error δ (a) 0.001, (b) 0.01, and (c) 0.1 (shown as percentage in figure) for the catalog in Fig. 10. The shaded area shows cases where the error does not exceed δ. Out of 100 different catalogs generated, the number of catalogs that resulted in more than δ was 14, 7, and 6 times for panels (a), (b), and (c), respectively.

$\delta_{obs} > \delta$ was 14, 7, and 6 for $\delta = 0.001$, 0.01, and 0.1, respectively. Similar results were obtained for the complete catalog in Fig. 10(b). It is found that the maximum observed error (number of catalogs out of 100 catalogs that resulted in $\delta_{obs} > \delta$) is 0.004 (10), 0.015 (12), and 0.11 (4) for $\delta = 0.001$, 0.01, and 0.1, respectively. The catalogs were generated after 1000 escapes in this case. These results demonstrate that the procedure outlined in this work can reliably estimate the error associated with a catalog of processes generated for a basin using dynamical BEPS.

VI. CONCLUSIONS

We have presented an improved procedure for building a KMC model with a chosen accuracy. The KMC model can be prepared by cataloguing processes found from states in the PES using BEPS calculations. The catalog of processes from a state can be incomplete as some of the processes, which can be observed in the correct dynamics, might be missing in the catalog. This introduces an error in the dynamics when the catalog is employed with the KMC method. In this work, we have developed a mathematical framework to ascertain this error. The error associated with the process catalog is obtained in terms of the probability that a process that is missing from the catalog will not be selected in the correct dynamics. Further, the introduction of a validity time associated with the catalog ensures that we can specify maximum error associated with a catalog. We show that the missing rate can be estimated within an order of magnitude of the correct missing rate in most realistic situations where the contribution from the accessible timescales to the missing rate will dominate over the contribution from the inaccessible timescales. Thus, our approach can be used to find both the error and validity time of a KMC model even though all states and processes in the PES might not be known to us *a priori*. Besides laying a mathematical foundation for finding error associated with KMC, our approach provides a way of deciding when to stop seeking for missing processes when self-learning KMC models are being generated.

ACKNOWLEDGMENTS

We acknowledge helpful discussions with A. F. Voter. A.C. acknowledges support from BRNS Young Scientist Award from the Department of Atomic Energy (DAE-BRNS) No. 2011/36/43-BRNS/1975.

[1] V. J. Bhute and A. Chatterjee, J. Chem. Phys. **138**, 084103 (2013).
[2] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. **17**, 10–18 (1975).
[3] D. T. Gillespie, J. Comput. Phys. **22**, 403–434 (1976).
[4] K. A. Fichthorn and W. H. Weinberg, J. Chem. Phys. **95**(2), 1090–1096 (1991).
[5] A. F. Voter, in *Radiation Effects in Solids*, edited by K. E. Sickafus, E. A. Kotomin, and B. P. Uberuaga (Springer, NATO Publishing Unit, Dordrecht, 2006).
[6] A. Chatterjee and D. G. Vlachos, J. Comput.-Aided Mater. Des. **14**(2), 253–308 (2007).
[7] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford Science Publications, Oxford, 1989).
[8] D. J. Wales, Int. Rev. Phys. Chem. **25**(1-2), 237–282 (2006).
[9] G. Henkelman, B. P. Uberuaga, and H. Jónsson, J. Chem. Phys. **113**(22), 9901–9904 (2000).
[10] W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).
[11] G. Henkelman and H. Jonsson, J. Chem. Phys. **115**, 9657 (2001).
[12] N. Mousseau and G. T. Barkema, Phys. Rev. E **57**, 2419–2424 (1998).
[13] J. Rogal, K. Reuter, and M. Scheffler, Phys. Rev. Lett. **98**, 046101 (2007).
[14] T. Rehman, M. Jaipal, and A. Chatterjee, J. Comput. Phys. **243**, 244–259 (2013).
[15] D. Konwar, V. J. Bhute, and A. Chatterjee, J. Chem. Phys. **135**, 174103 (2011).
[16] G. R. Bowman, X. Huang, and V. S. Pande, Cell Res. **20**, 622–630 (2010).
[17] G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande, J. Chem. Phys. **131**, 124101 (2009).
[18] G. H. Gilmer, H. C. Huang, T. D. de la Rubia, J. Dalla Torre, and F. Baumann, Thin Solid Films **365**(2), 189–200 (2000).
[19] R. Elber, Curr. Opin. Struct. Biol. **15**(2), 151–156 (2005).
[20] A. F. Voter, Phys. Rev. B **57**, R13985–R13988 (1998).
[21] A. F. Voter, J. Chem. Phys. **106**(11), 4665–4677 (1997).
[22] R. Miron and K. A. Fichthorn, J. Chem. Phys. **119**(12), 6210–6216 (2003).
[23] M. R. Sorenson and A. F. Voter, J. Chem. Phys. **112**(21), 9599–9606 (2000).
[24] R. A. Miron and K. A. Fichthorn, Phys. Rev. Lett. **93**(12), 128301 (2004).
[25] L. Xu and G. Henkelman, J. Chem. Phys. **129**, 114104 (2008).
[26] O. Trushin, A. Karim, A. Kara, and T. S. Rahman, Phys. Rev. B **72**, 115401–115409 (2005).
[27] L. K. Béland, P. Brommer, F. El-Mellouhi, J.-F. Joly, and N. Mousseau, Phys. Rev. E **84**(4), 046704 (2011).
[28] D. T. Gillespie, J. Phys. Chem. **81**, 2340–2361 (1977).
[29] S. Verma, T. Rehman, and A. Chatterjee, Surf. Sci. **613**, 114–125 (2013).