

Molecular Dynamics Saddle Search Adaptive Kinetic Monte Carlo

Samuel T. Chill and Graeme Henkelman

*Department of Chemistry and the Institute for Computational Engineering and Sciences,
The University of Texas at Austin, Austin, TX 78712-0165*

(Dated: May 7, 2014)

A method for accelerating molecular dynamics simulations in rare event systems is described. From each new state visited, high temperature molecular dynamics trajectories are used to discover the set of escape mechanisms and rates. This event table is provided to the adaptive kinetic Monte Carlo algorithm to model the evolution of the system from state to state. Importantly, an estimator for the completeness of the calculated rate table in each state is derived. The method is applied to three model systems: adatom diffusion on Al(100); island diffusion on Pt(111); and vacancy cluster ripening in bulk Fe. Connections to the closely-related temperature accelerated dynamics method of Voter and coworkers is discussed.

I. INTRODUCTION

Adaptive kinetic Monte Carlo (AKMC) is a method which applies dynamically-constructed rate tables to kinetic Monte Carlo (KMC) simulations.[1] For each unique state that the system visits, searches are performed on the potential energy surface (PES) to find low-energy first-order saddle points leading to adjacent states. Saddle searches have been carried out with minimum mode (min-mode) following algorithms such as the dimer method [2], Raleigh-Ritz minimization [3] as in the hybrid eigenvector following method [4], or the Lanczos method as in the activation relaxation technique (ART) nouveau [5]. Given the geometry of the saddle point, rates can be efficiently calculated for the forward and backward reactions using the harmonic approximation to transition state theory (HTST). In this paper, we compare the efficiency of min-mode following saddle searches to high temperature molecular dynamics (MD) saddle searches.

When using a min-mode following method, initial configurations are generated by displacing away from a minimum energy configuration. The choice of which degrees of freedom to displace (*e.g.* under-coordinated atoms) and the distribution of the displacement (*e.g.* a Gaussian distribution with a predetermined variance) need to be determined for each system under investigation. These parameters not only effect the computational efficiency of the algorithm, but also the accuracy of the resulting KMC simulation, as the distribution of initial configurations and the shape of the potential energy surface determine the probability that a particular saddle will be found. This makes it difficult to calculate the confidence that all of the important reactive events that are relevant at the simulation temperature have been found.

In a MD saddle search, the trajectory is confined to the initial potential energy basin by detecting when it escapes the basin and restarting it within the basin. The escape events can be detected by periodically performing a geometry optimization to determine if the trajectory is still in the initial energy basin. If it has exited, the trajectory is terminated and a nudged elastic band (NEB) [6, 7] and/or min-mode following calculation is performed

to locate the saddle point between the initial and final state basins.

An important advantage of using MD over min-mode following methods to find saddle points is that the probability of finding escape mechanisms with MD is directly proportional to their rates and their relative importance in the AKMC event table. At elevated MD temperatures, high entropy processes are overrepresented as compared to the temperature of interest, but this bias can be corrected within the HTST approximation. This is the strategy used by Voter and coworkers in their temperature accelerated dynamics (TAD) method [8], where escape events are found with high temperature MD trajectories and the escape times at the low temperature are determined from an Arrhenius extrapolation.

This work closely follows the TAD procedure for sampling possible escape pathways with high temperature MD. The difference is that we are not aiming to find just the first escape event at low temperature, instead, we want to find the entire set of escape pathways and rates that are accessible at the low temperature for use in AKMC. Key to the effective use of MD saddle search with AKMC (MDSS-AKMC) is an estimator for the completeness of the rate table. In past work, a confidence in the rate table found with min-mode following methods was based upon an assumed distribution for discovering saddle points, such as a uniform distribution.[9] This assumption can be a poor one. Even in well understood systems where the chosen initial displacement size and direction are close to optimal, it can be hundreds of times more likely to find one saddle than another, even when the two events have a similar rate.[2] With MD saddle searches, however, an unbiased error in the KMC rate catalog can be determined. Recently, Bhute and Chatterjee have shown how this can be done using a maximum likelihood estimation of the total escape rate.[10, 11] Here, we use the HTST expression as in TAD to derive an estimator for the error in the total rate, and use this as a criterion for sufficient discovery of the rate catalogue to escape a state. In this way, we show how MDSS based AKMC can be done with higher accuracy and sometimes even more efficiently than when based upon min-mode following saddle searches. All numerical calculation were

performed with the EON software.[12]

II. ERROR IN THE ESCAPE RATE DUE TO AN INCOMPLETE RATE CATALOG

The complete rate catalog \mathbf{C} is the set of escape rate constants of all N possible escape processes from a potential energy basin. Technically this is a multiset as different processes may have the same rate constant. Initially, when dynamically building the rate catalog, none of the processes are known. As they are identified, they are added to the set of found events \mathbf{F} and removed from the complementary set of missing events \mathbf{M} . The total escape rate, K , is defined as the sum of the escape rates, k_i , of each process at the low temperature of interest

$$K = \sum_{i=1}^N k_{i,T_{\text{low}}}. \quad (1)$$

In a KMC simulation, the probability of picking an event is proportional to its rate. Thus an appropriate error measure E for the rate catalogue \mathbf{F} is the probability of picking one of the missing processes (in \mathbf{M}) in a KMC step based upon the complete catalogue \mathbf{C} ,

$$E(\mathbf{F}) = 1 - \frac{1}{K} \sum_{k_i \in \mathbf{F}} k_{i,T_{\text{low}}} = \frac{1}{K} \sum_{k_i \in \mathbf{M}} k_{i,T_{\text{low}}}. \quad (2)$$

Under our assumption of first-order kinetics, the mean-first-escape-time for each process, τ_i , is exponentially distributed according to the rate k_i . Integrating the distribution up to time t yields the probability that the process has occurred by time t

$$p(t; k_i) = \int_0^t k_i \exp(-k_i \tau_i) d\tau_i = 1 - \exp(-k_i t). \quad (3)$$

The probability of having found a particular set \mathbf{F} of processes by time t in the high temperature MD simulation is

$$P(\mathbf{F}) = \prod_{k_i \in \mathbf{F}} p(t; k_{i,T_{\text{high}}}) \prod_{k_i \in \mathbf{M}} 1 - p(t; k_{i,T_{\text{high}}}). \quad (4)$$

Here, $k_{i,T_{\text{high}}}$ are the rate constants at the high temperature. $P(\mathbf{F})$ represents the joint probability of having independently found the events in \mathbf{F} and having not yet found the events in \mathbf{M} . Now we may express the average error at time t by averaging over all possible sets of processes that may be found

$$\overline{E}(\mathbf{C}) = \sum_{\mathbf{F} \in \mathcal{P}(\mathbf{C})} P(\mathbf{F}) E(\mathbf{F}) \quad (5)$$

$$= 1 - \frac{1}{K} \sum_{i=1}^N p(t; k_{i,T_{\text{high}}}) k_{i,T_{\text{low}}}, \quad (6)$$

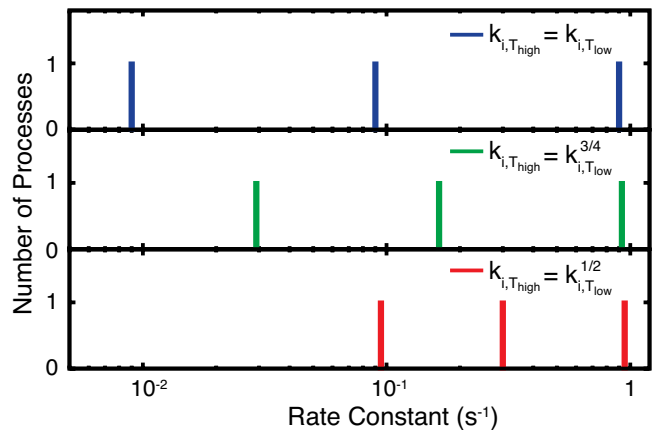


FIG. 1. Distribution of rate constants $k_{i,T_{\text{low}}}$ in a model system. As the temperature is raised, the rate constants $k_{i,T_{\text{high}}}$ become closer.

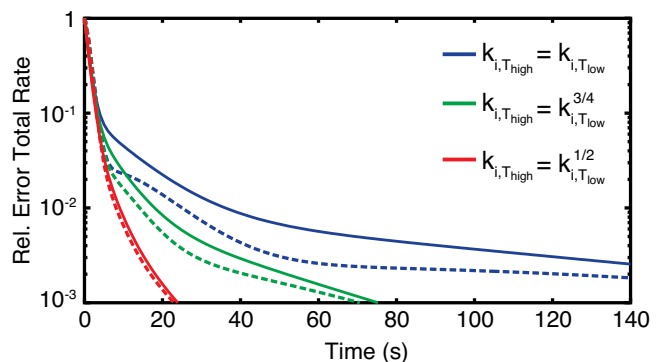


FIG. 2. Comparison of the average error in the rate catalog $\overline{E}(\mathbf{C})$ (solid line) and the proposed estimator of the error $\overline{X}(\mathbf{F})$ (dashed line) for the three distributions of high-temperature rates shown in Fig. 1.

where the $\mathcal{P}(\mathbf{C})$ represents the power set (the set of all subsets) of \mathbf{C} . A derivation of Eq. 6 is given in Appendix A. In the case that $k_{1,T_{\text{high}}} = k_{2,T_{\text{high}}} = \dots = k_{N,T_{\text{high}}} = k_{T_{\text{high}}}$, Eq. 6 reduces to the simple form

$$\overline{E}(\mathbf{C}) = 1 - p(t; k_{T_{\text{high}}}) = \exp(-k_{T_{\text{high}}} t). \quad (7)$$

Note that there is no dependence upon N in this last expression. This means that if all of high temperature rate constants are equal then the uncertainty in the rate table can be expressed exactly using only that rate.

III. ESTIMATOR OF THE ESCAPE RATE ERROR

In an AKMC simulation, only the set of found processes \mathbf{F} are known. The total rate K is unknown and therefore Eqs. 2 and 5 cannot be evaluated directly. Instead, we can construct an estimator for the average error using information from the set \mathbf{F} and the MD time used

to discover \mathbf{F} ,

$$X(\mathbf{F}) = 1 - \sum_{k_i \in \mathbf{F}} p(t; k_i, T_{\text{high}}) k_{i, T_{\text{low}}} / \sum_{k_i \in \mathbf{F}} k_{i, T_{\text{low}}}. \quad (8)$$

The assumption made in Eq. 8 is that the *average error* at time t from the known set of events \mathbf{F} is a good estimator for the *error* from the complete set \mathbf{C} . Another way of stating this approximation is that the events in \mathbf{F} are characteristic of those in \mathbf{C} .

The estimator $X(\mathbf{F})$ asymptotically approaches the average error, $\overline{E}(\mathbf{C})$, in two cases. The first is as t approaches infinity (*i.e.* when all processes have been found), where it reduces to Eq. 6. The second is as the MD temperatures approaches infinity for systems where each process has the same entropic prefactor (*i.e.* when $k_{1, T_{\text{high}}} = k_{2, T_{\text{high}}} = \dots = k_{N, T_{\text{high}}}$), where it reduces to Eq. 7.

In order to demonstrate the behavior of the estimator, we have chosen a simple model system with three processes: $k_{1, T_{\text{low}}} = 0.009$, $k_{2, T_{\text{low}}} = 0.09$, and $k_{3, T_{\text{low}}} = 0.9 \text{ s}^{-1}$. Three cases will be examined for the high temperature MD rates: $k_{i, T_{\text{high}}} = k_{i, T_{\text{low}}}$, $k_{i, T_{\text{high}}} = k_{i, T_{\text{low}}}^{3/4}$, $k_{i, T_{\text{high}}} = k_{i, T_{\text{low}}}^{1/2}$. These cases correspond to performing the MD sampling directly at the temperature of interest, at a 33% increase, and at a 100% increase in temperature. In this model, the prefactor of each process is considered to be the same. The rate constants for each case are shown in Fig. 1.

The quality of the estimator $X(\mathbf{F})$ is determined by comparing its average, $\overline{X}(\mathbf{C})$, to the exact average error $\overline{E}(\mathbf{C})$. $\overline{X}(\mathbf{C})$ is obtained in a similar manner to Eq. 5,

$$\overline{X}(\mathbf{C}) = \sum_{\mathbf{F} \in \mathcal{P}(\mathbf{C})} P(\mathbf{F}) X(\mathbf{F}). \quad (9)$$

Results from analytic evaluation of Eqs. 5 and 9 for the model system are shown in Fig. 2. In the case that $k_{i, T_{\text{high}}} = k_{i, T_{\text{low}}}$, the MD rates are separated by an order of magnitude, which is far from the equal-rate case where the estimator is exact. At short time this leads to significant underestimation of the error because the fast process, which is found first, is not characteristic of the entire set, violating the assumption of the estimator. As the temperature is raised, the rate of finding the processes increases. Importantly, the spread between the high-temperature rates also decreases so that the error is accurately modeled by the estimator at all times.

IV. VACANCY CLUSTER FORMATION IN IRON

Systems that have been modeled using long time scale dynamics include materials which have been damaged by radiation. One such model system that has been used to compare long time scale methods is vacancy cluster formation in body centered cubic (bcc) Fe. The system

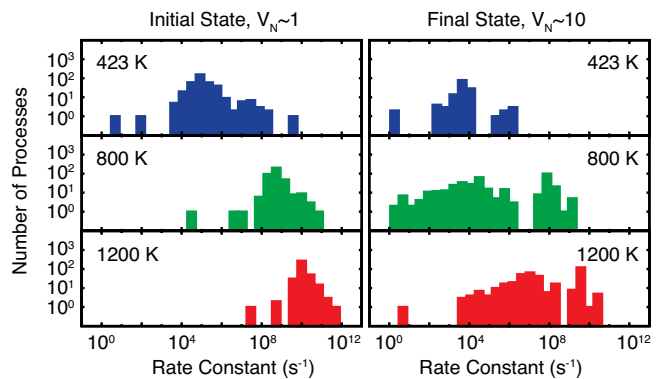


FIG. 3. The effect of temperature on the HTST rate constants in a bcc Fe lattice with vacancies.

was introduced by Fan et. al to demonstrate their autonomous basin climbing (ABC) algorithm.[13] In their calculation, the coalescence of vacancies into nano-voids was determined to occur on the time scale of hours at an initial temperature of 50°C and a heating rate of 0.01 K/s. Interestingly, a similar calculation was done by Brommer et. al using the kinetic activation relaxation technique (k-ART) [14] who calculated a time scale of milliseconds for the coalescence – a difference of eight orders of magnitude. This remarkable disagreement provides a strong motivation for developing benchmarks that can be used to compare the accuracy of different long timescale dynamics methods. As such, we define such a benchmark which is close to these previous calculations, which we then also use to test our error estimator in MDSS-AKMC.

The initial configuration for the benchmark has 50 randomly placed vacancies in a $10 \times 10 \times 10 a_0$ supercell of bcc Fe, where $a_0 = 2.87 \text{ \AA}$ is taken as the experimental lattice constant. Initially, the average vacancy cluster size, V_n is unity (or very close to unity). The state-to-state evolution of the system is followed in time based upon rate constants calculated using HTST with a fixed entropic prefactor of $5 \times 10^{12} \text{ s}^{-1}$ at a temperature of 423 K. A fixed prefactor was chosen to focus the benchmark to the efficiency of saddle point determination. States are defined as the set of points that minimize to the same geometry. The potential energy is evaluated with an embedded atom method (EAM) model, as parameterized by Ackland et. al.[15] The requirement of the benchmark is to determine the average time for the potential energy to decrease below -7763.5 eV. This final energy corresponds to an average vacancy cluster size $V_n > 9$.

The choice of T_{high} is important for the efficiency of MDSS-AKMC. Increasing T_{high} increases the rate at which processes are found. Too high, however, and a systematic error is introduced in $X(\mathbf{F})$ due to anharmonic corrections to the HTST rates in Eq. 8 and the loss of first-order kinetics. The second issue can be addressed by reaching local equilibrium before running high temperature dynamics as in modified TAD.[16] Fig. 3 shows

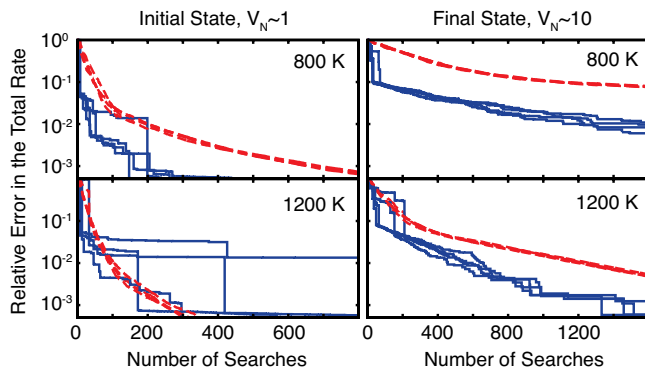


FIG. 4. The relative error in the rate, $E(\mathbf{F})$, (blue solid lines) compared with the estimated error, $X(\mathbf{F})$, (red dashed lines) at 800 K (top) and 1200 K (bottom). Four realizations are superimposed to show the random variation between simulations.

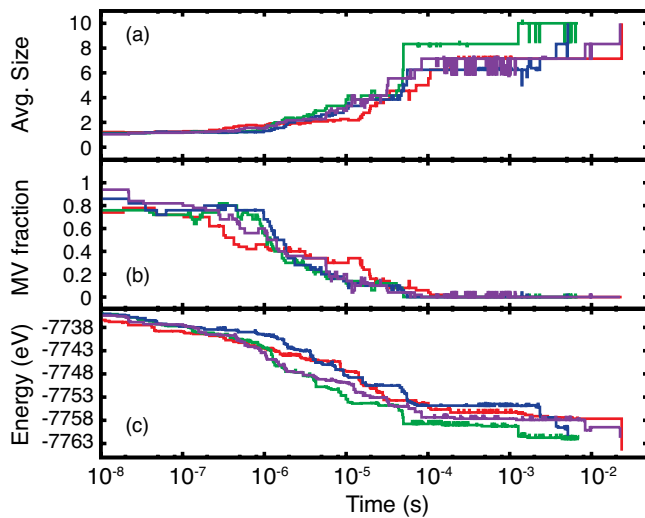


FIG. 5. Four AKMC trajectories of 50 randomly distributed vacancies in a $10 \times 10 \times 10$ supercell of bcc Fe. (a) Average vacancy cluster size; (b) fraction of monovacancies (MV); and (c) potential energy of the minima along the trajectories.

the spectrum of rate constants for an initial configuration ($V_n \sim 1$) and a final configuration ($V_n \sim 10$) at 423 K (T_{low}), 800 K, and 1200 K. In both states, MD at T_{low} cannot be used to sample transitions on the picosecond time scale of our saddle searches. At 800 K the temperature is sufficient to overcome the relatively low barriers of vacancy diffusion in the initial state. In the final state, however, when the vacancies have clustered, a higher temperature of 1200 K is necessary.

The accuracy of the estimator is shown in Fig. 4 as a function of T_{high} . In the initial state, $X(\mathbf{F})$ is a good (and safe) estimator of the error in the rate catalogue at 800 K. At 1200 K the harmonic approximation starts to break down, and the estimator loses accuracy. In the final state 1200 K is appropriate and the estimator is accurate. In order to overcome the high barriers at the

end of the simulation, we choose T_{high} to be 1200 K. Tuning T_{high} appropriately for different states would be a natural improvement to the method.

Four independent MDSS-AKMC simulations were run with MD saddle searches performed in each state at 1200 K, until $X(\mathbf{F}) < 0.01$, which corresponds to a 99% confidence in the total escape rate. The average time taken to reach a potential energy of less than -7763.5 eV was calculated as 12 ± 5 ms. Figure 5 shows the average vacancy cluster size, fraction of defects that are monovacancies, and the potential energy as functions of time for each trajectory. While it is not possible to directly compare this time to previously reported ABC and k-ART simulations because of differences in the temperature profile, our calculated time scale for vacancy cluster formation are in much better agreement with k-ART than ABC.

V. COMPARING THE EFFICIENCY OF MD AND DIMER SADDLE SEARCHES

Saddle searches based on MD have the advantage of allowing an error estimator of the escape rate from a state. This does not necessarily mean, however, that MD is a computationally efficient way of finding saddle points. A numerical comparison of MD and dimer saddle searches is done for three systems: Al adatom diffusion on an Al(100) surface modeled with an embedded atom model developed by Voter and Chen [17]; the motion of a compact Pt heptamer island on a Pt(111) surface modeled with a Morse potential [18]; and vacancy cluster formation in bcc Fe, as described in Sec. IV.

Our metric for comparing the saddle search methods is the average relative error in the total escape rate. Here the escape rate is defined as the rate to exit from the initial potential energy basin and is obtained by averaging Eq. 2 over 50 runs vs. the number of potential energy (force) evaluations. All rates are calculated using HTST, with a constant vibrational prefactor of $5 \times 10^{12} \text{ s}^{-1}$, at 300 K for the Al and Fe systems and at 700 K for the Pt system. The total escape rate K in Eq. 2 was evaluated by first running 20,000 high temperature MD saddle searches to obtain a rate catalog that was considered complete.

The initial distribution of configurations for the dimer saddle searches was tuned for each system based upon chemical intuition. In this way, *a priori* knowledge of likely reaction mechanisms can be used to reduce the computational effort. In each case, searches were initiated with displacements from the reactant minimum drawn from a Gaussian distribution in a subset of the Cartesian degrees of freedom. In the Al system, the adatom and its first coordination shell (15 degrees of freedom) were displaced with a standard deviation of $\sigma = 0.2 \text{ \AA}$. In the Pt system, all seven island atoms (21 degrees of freedom) were displaced by $\sigma = 0.1 \text{ \AA}$. In the Fe system a random Fe atom with coordination number

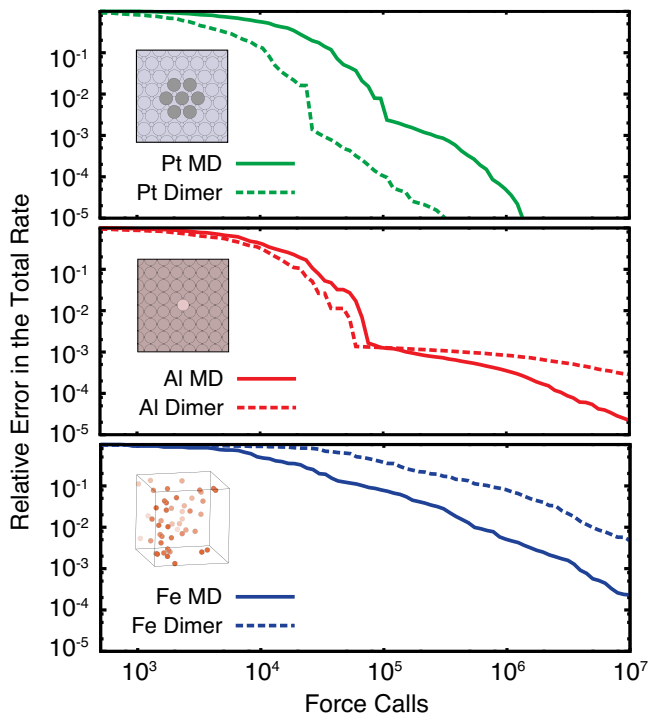


FIG. 6. An efficiency comparison of dimer and MD saddle searches for three different systems: Pt heptamer diffusion; Al adatom diffusion; and vacancy cluster ripening in bcc Fe. The inset for the Fe system shows the initial distribution of vacancies.

less than eight was displaced, as well as all neighbors within 6 \AA , by $\sigma = 0.2 \text{ \AA}$. For the MD saddle searches only the T_{high} parameter is required; temperatures of 1000, 1200, and 2000 K were chosen for the Al, Fe, and Pt systems respectively.

Differences in efficiency are shown for the three systems in Fig. 6. In the Pt system, the dimer searches significantly outperform the MD searches. This is due to the localized displacement scheme that effectively targets the most important mechanism of island sliding.

In the Al system, diffusion mechanisms involve both the adatom and surface atoms. This makes it more difficult to construct an effective distribution for dimer search displacements. The performance of the two methods here is similar unless a highly accurate rate catalog is required. In this low-error regime, high-energy long range events involving many atoms must be found and these are hard to find with the dimer initiated with displacements localized around the adatom.

In the initial state of the Fe system, there are 337 atoms that neighbor the 50 vacancies. These under-coordinated atoms are targeted by the dimer searches. While each of the vacancies can diffuse, the escape rate is dominated by a single fast process involving two nearby vacancies. This outlier can be seen in the initial state spectrum of rates at 423 K in Fig. 3. Since a random selection of the correct under-coordinated atom to displace has a small

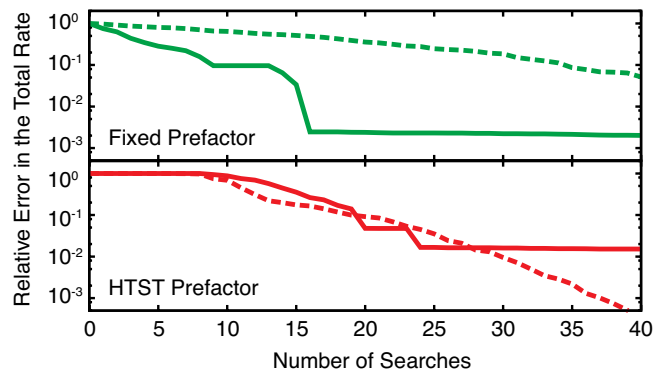


FIG. 7. The average relative error in the rate for Al adatom diffusion on Al(100) at 300 K with MDSS performed at 800 K, $E(\mathbf{C})$, (solid lines) compared with the average estimated error, $\bar{X}(\mathbf{C})$, (dashed lines) with a fixed harmonic prefactor of $5 \times 10^{12} \text{ s}^{-1}$ (top) and the Vineyard prefactor (bottom).

probability, the MD search strategy is more efficient because it automatically finds the fast event with a high probability.

Each MD saddle search takes on average several times more force calls to find a saddle. In the Al, Fe, and Pt systems the MD saddle searches were six, two, and four times more expensive, respectively. Despite the increased cost per search, the MD method can still outperform dimer searches. There are a few reasons for this. First, dimer searches can wander to configurations of high energy where they are terminated, or to saddles which do not connect back to the initial state minimum (by steepest descent). Second, while dimer searches can be localized to active regions of configuration space, intuition may not be good enough to target the part of the system with the highest rates (as in the bcc Fe case). While MD has a higher overhead, per search, it is more likely to find relevant saddles of high rate which are connected to the initial state. Combined with a good error estimation, MD searches can be preferable, particularly in cases where high accuracy is desired.

VI. DISCUSSION

In order to evaluate the estimator $X(\mathbf{F})$ the high temperature rates $k_{i,T_{\text{high}}}$ must be known. In the Fe vacancy cluster formation simulation, the estimator, when evaluated with these HTST rates, was found to be accurate enough to be useful as a stopping criterion. However, it is not always the case that the HTST rate is a good estimate of the high temperature escape rate.

For example, the three fastest diffusion events in the Al adatom on Al(100) system at 300 K are a 2-atom exchange mechanism, where the adatom pushes a substrate atom up onto the surface; a hop mechanism, where the adatom moves directly to a neighboring site; and a 4-atom exchange mechanism where the adatom pushes

TABLE I. The three fastest events at 300 K for Al adatom diffusion on a Al(100) surface.

Event	Prefactor (s^{-1})	Barrier (eV)	HTST Rate (s^{-1})		MD Rate (s^{-1})
			300 K	800 K	800 K
2-Atom Exchange	1.4×10^{13}	0.206	5.0×10^9	7.3×10^{11}	$(4.2 \pm 0.3) \times 10^{11}$
Hop	5.8×10^{12}	0.377	2.7×10^6	2.4×10^{10}	$(7.7 \pm 0.8) \times 10^{10}$
4-Atom Exchange	2.0×10^{14}	0.396	4.6×10^7	6.6×10^{11}	$(2.9 \pm 0.5) \times 10^{10}$

three substrate atoms so that one surfaces three sites away. The rates of these events as calculated by HTST and MD are shown in Table I. The HTST rate of the 4-atom exchange mechanism is 22 times greater than what is observed in a direct MD simulation. The high HTST rate means that the estimator underestimates the error on the characteristic timescale of this event. Figure 7 shows the average value of the error estimator compared to the true average error. Two cases are considered: a constant prefactor for all events of $5 \times 10^{12} s^{-1}$ and a Vineyard harmonic prefactor calculated by diagonalizing the dynamical matrix, which in turn is calculated by finite difference. [19] In the case of the constant prefactor the predicted high temperature rates significantly underestimate the true rate, which results in an overly conservative estimator. With the harmonic prefactor, the error estimator is accurate until the relative error reaches the contribution of the 4-atom exchange, at about 1% of the total rate. At this point the estimator diverges from the true error. A promising future direction is to obtain a more accurate true rate from the statistics of the high temperature MD trajectory.

MDSS-AKMC is similar to Voter’s TAD method, but there are significant differences. In TAD, high temperature MD is used to find escape events from a state and the time at which that event would have occurred at a lower temperature of interest is extrapolated from HTST. Once confidence has been reached that the first event at low temperature has been found, the transition is taken and the processes is repeated in the new state. One might then ask why one should do the additional work in MDSS-AKMC to reach confidence for the rate catalogue. First, an advantage of doing AKMC with the rate catalogue is that it is based upon rates calculated at the low temperature, and does not rely on an extrapolation based upon the HTST approximation at the high tem-

perature, as in TAD. In principle the AKMC rates can be made as accurate as desired, for example using dynamical corrections to TST. Second, MDSS-AKMC can be augmented with computational strategies that efficiently recover known reaction mechanisms, including saddle recycling [9] and the kinetic database [20]. Third, AKMC allows for efficient coarse-graining of fast rates through the Monte Carlo with absorbing Markov chains approach [21]. The relative strengths of TAD and MDSS-AKMC, as well as the possibility of a hybrid approach, will be the subject of future studies.

VII. CONCLUSION

We have described a method to determine the events that go into an AKMC rate catalog using high temperature MD saddle searches. In simulations of surface and bulk diffusion, this MDSS-AKMC method is shown to be efficient for the calculation of long time scale dynamics in comparison to AKMC based upon dimer saddle searches.

VIII. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under grant CHE-1152342 and the Welch Foundation under grant F-1841. We gratefully acknowledge computational resources at the Texas Advanced Computing Center and the generous people who contributed computing time through the EON project [22], and Gideon Simpson for helpful discussions.

Appendix A: Derivation of Equation 6

Here we give the details of how Eq. 6 is derived from the definition given in Eq. 5.

Proof.

$$\bar{E}(\mathbf{C}) = \sum_{\mathbf{F} \in \mathcal{P}(\mathbf{C})} P(\mathbf{F})E(\mathbf{F}) \quad (5)$$

$$= \frac{1}{K} \sum_{\mathbf{F} \in \mathcal{P}(\mathbf{C})} \sum_{k_i \in \mathbf{M}} k_{i,T_{\text{low}}} \prod_{k_j \in \mathbf{F}} p(t; k_j, T_{\text{high}}) \prod_{k_j \in \mathbf{M}} 1 - p(t; k_j, T_{\text{high}}) \quad (A1)$$

$$= \frac{1}{K} \sum_{n_1=0}^1 \cdots \sum_{n_N=0}^1 \sum_{i=1}^N (1 - n_i) k_{i,T_{\text{low}}} \prod_{j=1}^N p(t; k_j, T_{\text{high}})^{n_j} (1 - p(t; k_j, T_{\text{high}}))^{1-n_j} \quad (A2)$$

$$= \frac{1}{K} \left[k_{N,T_{\text{low}}} (1 - p(t; k_N, T_{\text{high}})) \sum_{n_1=0}^1 \cdots \sum_{n_{N-1}=0}^1 \prod_{j=1}^{N-1} p(t; k_j, T_{\text{high}})^{n_j} (1 - p(t; k_j, T_{\text{high}}))^{1-n_j} \right. \\ \left. + \sum_{n_1=0}^1 \cdots \sum_{n_{N-1}=0}^1 \sum_{i=1}^{N-1} k_i n_i \prod_{j=1}^{N-1} p(t; k_j, T_{\text{high}})^{n_j} (1 - p(t; k_j, T_{\text{high}}))^{1-n_j} \right] \quad (A3)$$

$$= \frac{1}{K} k_{N,T_{\text{low}}} (1 - p(t; k_N, T_{\text{high}})) + \bar{E}(\{k_1, \dots, k_{N-1}\}) \quad (A4)$$

$$= \frac{1}{K} \sum_{i=1}^N (1 - p(t; k_i, T_{\text{high}})) k_{i,T_{\text{low}}} \quad (A5)$$

$$= 1 - \frac{1}{K} \sum_{i=1}^N p(t; k_i, T_{\text{high}}) k_{i,T_{\text{low}}} \quad (6)$$

Eq. A1 follows from the definitions of $E(\mathbf{F})$ and $P(\mathbf{F})$. In Eq. A2, the sum over the power set of \mathbf{C} has been rewritten as N sums over indicator variables (n_i) in order to enumerate all subsets of \mathbf{C} . In Eq. A3, the N^{th} sum has been explicitly evaluated for $n_N = 0$ and $n_N = 1$.

The factor $\sum_{n_1=0}^1 \cdots \sum_{n_{N-1}=0}^1 \prod_{j=1}^{N-1} p(t; k_j, T_{\text{high}})^{n_j} (1 - p(t; k_j, T_{\text{high}}))^{1-n_j}$ is equal to one as it represents the sum of probabilities of all the ways to find any subset of $\{k_1, \dots, k_{N-1}\}$. Eq. A4 is a recursion relation for $\bar{E}(\mathbf{C})$ used to give Eq. A5, which is equivalent to Eq. 6. \square

-
- [1] G. Henkelman and H. Jónsson, *J. Chem. Phys.* **115**, 9657 (2001).
- [2] G. Henkelman and H. Jónsson, *J. Chem. Phys.* **111**, 7010 (1999).
- [3] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985).
- [4] L. J. Munro and D. J. Wales, *Phys. Rev. B* **59**, 3969 (1999).
- [5] R. Malek and N. Mousseau, *Phys. Rev. E* **62**, 7723 (2000).
- [6] G. Henkelman and H. Jónsson, *J. Chem. Phys.* **113**, 9978 (2000).
- [7] G. Henkelman, B. P. Uberuaga, and H. Jónsson, *J. Chem. Phys.* **113**, 9901 (2000).
- [8] M. R. Sørensen and A. F. Voter, *J. Chem. Phys.* **112**, 9599 (2000).
- [9] L. Xu and G. Henkelman, *J. Chem. Phys.* **129**, 114104 (2008).
- [10] V. J. Bhute and A. Chatterjee, *J. Chem. Phys.* **138**, 084103 (2013).
- [11] V. J. Bhute and A. Chatterjee, *J. Chem. Phys.* **138**, 244112 (2013).
- [12] S. T. Chill, M. Welborn, R. Terrell, L. Zhang, J.-C. Berthet, A. Pedersen, H. Jónsson, and G. Henkelman, *Model. Simul. Mater. Sci. Eng.* (in press, 2012).
- [13] Y. Fan, A. Kushima, S. Yip, and B. Yildiz, *Phys. Rev. Lett.* **106**, 125501 (2011).
- [14] P. Brommer and N. Mousseau, *Phys. Rev. Lett.* **108**, 219601 (2012).
- [15] G. J. Ackland, M. I. Mendelev, D. J. Srolovitz, S. Han, and A. V. Barashev, *J. Phys.: Condens. Matter* **16**, S2629 (2004).
- [16] D. Aristoff and T. Lelièvre, *Multiscale Model. Simul.* **12**, 290 (2014).
- [17] A. F. Voter and S. P. Chen, *Mat. Res. Soc. Symp. Proc.* **82**, 175 (1987).
- [18] D. W. Bassett and P. R. Webber, *Surf. Sci.* **70**, 520 (1978).
- [19] G. H. Vineyard, *J. Phys. Chem. Solids* **3**, 121 (1957).
- [20] R. Terrell, M. Welborn, S. T. Chill, and G. Henkelman, *J. Chem. Phys.* **137**, 015105 (2012).
- [21] M. A. Novotny, *Phys. Rev. Lett.* **74**, 1 (1995).
- [22] The EON distributed computing project can be found at <http://eon.ices.utexas.edu/>.