Uncertainty and anharmonicity in thermally activated dynamics

T D Swinburne

Version of Record: https://www.sciencedirect.com/science/article/pii/S0927025620307473
Manuscript_b8298f5133fafdb9155401783010d17d

Uncertainty and anharmonicity in thermally activated dynamics

Thomas D Swinburne*

Aix-Marseille Université, CNRS, CINaM UMR 7325,

Campus de Luminy, 13288 Marseille, France

Abstract

At the atomic level, many phenomena in materials science reduce to long periods of thermal vibration interspersed by transitions between local free energy minima. The resultant rare event dynamics are exponentially sensitive to the catalog of available transitions, meaning incomplete models can make catastrophically erroneous predictions. This review summarises some recent efforts towards quantifying this uncertainty. I show that Bayesian methods can rigorously measure sampling incompleteness, be propagated to yield a quantified prediction uncertainty and autonomously manage massively parallel simulations. These methods allow uncertainty-controlled investigation of complex atomistic processes with minimal end-user supervision, facilitating high-throughput workflows. For individual transitions rates, I also show how the activation free energy can be evaluated with full treatment of anharmonic thermal vibrations. The developed methods, all freely available, are demonstrated on a wide range of challenging materials science problems.

* swinburne@cinam.univ-mrs.fr

I. INTRODUCTION

The ability to directly simulate the dynamics of hundreds to millions of atoms has allowed unprecedented insight into how materials transform and evolve. However, materials phenomena as fundamental as diffusion, creep or corrosion require *thermal activation*, meaning structural changes (such as defect migration or nanoparticle transformations) occur as a series of rare events between long periods of thermal vibration. In this regime, brute force generation of a single simulation trajectory by molecular dynamics is typically insufficient[1]. It has long been recognised that rare event dynamics can be mapped to a jump process between discrete states[2], with an error decays exponentially with the expected time spent in each state[3, 4]. States can be identified by performing an energy minimisation from finite temperature, with the minimized structure processed to produce some state label or index[5, 6]. The probability per unit time (transition rate) of a particular jump depends only on the currently occupied state and the magnitude of the free energy barrier for the jump[7] and can be calculated using well known transition state theory methods[8].
With a complete set of states and transition rates, often known as a kinetic transition network (KTN), one can construct essentially exact state-to-state trajectories. A popular method to generate trajectories is kinetic Monte Carlo[2, 5, 9], though a KTN can also be treated as a Markov chain, allowing quantities such as diffusion constants or first passage time distributions to be calculated using linear algebra techniques[10–13]. The topology of the KTN also allows more general insight into system properties, particularly when visualised using disconnectivity graphs[10, 14–16].

A recurrent issue[10, 11, 13, 17–19] when building rare event models is *incompleteness*-atomic mechanisms are too complex to identify *a priori*, meaning states and transitions must be discovered via simulation. As all models are incomplete their predictions are always uncertain, but this source of potentially unbounded error is rarely quantified.

This review briefly summarises our efforts over the last three years to quantify and propagate sampling uncertainty during model construction[10, 11, 13]. A key consequence is the ability to use uncertainty measures to *autonomously* decide where sampling should be performed to reduce prediction uncertainty[10, 20]. This has clear implications for the rapidly growing discipline of high-throughput simulation[21, 22], where end-user supervision must

be minimial[23, 24], to treat ever more ambitious simulation tasks. I describe initial applications to complex diffusion pathways[13] and nanoparticle transformations[11], providing rare measures of convergence unavailable in traditional approaches. Future work will also augment the training of machine learning interatomic potentials[25–29].

Whilst this review does not consider interatomic potential error, I highlight a method[7] that allows transition rate calculations to go beyond the harmonic vibration approximation[30, 31]. Application to large-scale simulations of dislocation migration reveal that anharmonic effects can be significant at surprisingly low temperatures.

II. STATE-WISE INCOMPLETENESS: THE UNKNOWN RATE

The most direct method to find transitions from a given state is to simply run dynamics, typically at an elevated temperature $\mathrm{T_H}$ where transitions are more frequent[17, 32]. There are also a variety of methods[33–35] that follow curvatures on the potential energy surface to discover transition pathways, which can be more efficient in some cases[36]. However, dynamical sampling has the unique advantage that incompleteness will always decrease with additional computational effort, which is essential for robust uncertainty quantification- the probability of not observing a transition of rate k after a time τ in some basin is the decay $\exp(-\mathrm{k}\tau)$[3, 10, 17]. After some period of computational effort, a total time τ_i has been spent in a state i, during which a set of *first passage times* τ_{ji} to states j have been observed, for which transition state theory (TST) can be used to calculate rates k_{ji}.

The `TAMMBER` code (available at `github.com/tomswinburne/tammber`) employs a variable-temperature accelerated dynamics (TAD) method[32], accumulating sampling information in parallel and producing $(\tau_i, \{\tau_{ji}, \mathrm{k}_{ji}\})$ over a range of possible $\mathrm{T_H}$ down to the lowest application temperature $\mathrm{T_L}$. The acceleration temperature Details on how sampling data at variable T_H is collated can be found in Ref. [10].

The current implementation calculates rates using harmonic TST, though future work will relax this assumption using the `PAFI` method[7], detailed below. Sampling incompleteness in a state i can then be formally quantified as the *unknown rate* k_i^u, such that the total escape rate from i writes $\mathrm{k}_i^t = \mathrm{k}_i^u + \mathrm{k}_i^o$, where $\mathrm{k}_i^o = \sum_j \mathrm{k}_{ji}$ is the total observed escape rate.

Using Poissonian statistics, one can derive the likelihood of observing our simulation data for any postulated value of k_i^t, ~~yeilding~~yielding a Bayesian posterior distribution

$\pi(\mathrm{k}_i^t|\tau_i, \{\tau_{ji}, \mathrm{k}_{ji}\})$ and an expected unknown rate

$$\langle \mathrm{k}_i^u \rangle = \int_{\mathrm{k}_i^o}^{\infty} \mathrm{k}\pi(\mathrm{k}|\tau_i, \{\tau_{ji}, \mathrm{k}_{ji}\})\mathrm{dk} - \mathrm{k}_i^o, \tag{1}$$

as illustrated in Figure 1[10]. Estimators for $\langle \mathrm{k}_i^u \rangle$ have been proposed in pioneering earlier work to address this issue[17–19]. However, the Bayesian approach described here has the key property of monotonic decay with computational effort, even under the discovery of new transitions. This is essential for autonomous convergence analysis in a high-throughput setting; one can show unless this property is satisfied, inferred low temperature prediction timescales can be exponentially volatile[10]. In addition, by specifying the computational cost of elemenary sampling operations (thermalization and transition rate calculation), it is possible to determine the derivative $\langle \mathrm{dk}_i^u(\mathrm{T_L})/\mathrm{d}c_i(\mathrm{T_H}) \rangle$, the expected change in the low temperature unknown rate $\mathrm{k}_i^u(\mathrm{T_L})$ with respect to high temperature computational work $c_i(\mathrm{T_H})$. In this manner $\mathrm{T_H}$ can be chosen such that $\mathrm{k}_i^u(\mathrm{T_L})$ decreases as fast as possible[10], but most importantly allows for autonomous allocation of simulation tasks.

III. GLOBAL INCOMPLETENESS : THE RESIDENCE TIME

Whilst rare event models are typically built from a set of transition rates $\{\mathrm{k}_{ji}\}$, the presence of unknown rates $\{\langle \mathrm{k}_i^u \rangle\}$ allows trajectories to leave the known state space, corresponding to the discovery of some unknown event[37, 38]. This ~~yields~~yields a global measure of model quality, the expected *residence time* over which model predictions can be trusted, reading

$$\langle \tau_{res} \rangle = -\mathbf{1}[\mathbf{Q}]^{-1}\mathbf{P}_0, \quad [\mathbf{Q}]_{ji} = \mathrm{k}_{ji} - \delta_{ji}\left(\langle \mathrm{k}_i^u \rangle + \mathrm{k}_i^o\right), \tag{2}$$

where $\mathbf{P}_0$ is some distribution over the known states and $\mathbf{1}$ is a row vector of ones. It is simple to show that $\langle \tau_{res} \rangle \to \infty$ as $\sum_i \langle \mathrm{k}_i^u \rangle \to 0$, whilst if $\mathbf{P}_0$ is concentrated on some initial state $\langle \tau_{res} \rangle$ is monotonic with sampling effort, a key consequence of our estimation approach. To determine how much sampling effort should be allocated to each state, one can then use results discussed above to evaluate $\mathrm{d}\langle \tau_{res} \rangle/\mathrm{d}c_i$, the change in the residence time with additional computational work in i, giving an optimal work distribution[7] where states that are of greater importance in a global sense are autonomously allocated more resources. This is the essential workflow of `TAMMBER`, illustrated in Figure 1. An example application to the interaction of a self-interstitial atom with a stable C15[15] defect is shown, with

the discovered KTN displayed as a disconnectivity graph[14]. The tree structure shows the minimum activation energy to reach one state from any other, allowing a succinct overview of the energy landscape. I show two disconnectivity graphs, one where each found state has a unique label and a much smaller graph which collates states invariant under exchange, translation and space group symmetries. Sampling in this 'compressed' space ~~yeilds~~yields large sampling efficiencies, returning large prediction timescales and allowing a quantitative assesment of the ability of the model to capture e.g. the breakup mechanism of these defects[7, 11]. This compressed space is particular advantageous for the diffusion of isolated defects, due to the high symmetry of their configurations.

IV. AUTONOMOUS CONVERGENCE OF DEFECT DIFFUSIVITIES

The diffusion of crystal defects is a key process in numerous areas of materials science. The motion of surface islands or interstitial clusters is often highly correlated[39], with many 'flickering' transitions between states before a net translation is made[40].

In recent work, `TAMMBER` has been used to efficiently and autonomously evaluate a 'generating function' for defect migration, from which mean squared displacements over $\langle\tau_{res}\rangle$ and thus estimates of diffusion constants D_α, $\alpha = 1, 2, 3$ can be extracted[11]. Using the estimates of $\{\langle \mathrm{k}_i^u \rangle\}$ and enforcing detailed balance, a convergence measure for D_α is derived from the Kullback-Leibler divergence[41] across diffusion processes consistent with the sampling uncertainty. It can be shown that the convergence decays as $1/\langle\tau_{res}\rangle$, meaning the optimal distribution of sampling tasks is unchanged. An example of this fully autonomous methodology is shown in figure 2 for trimer diffusion on the (110) surface of tungsten, which exhibits significant correlation effects. The ability to converge these highly nontrivial observables with minimal end-user analysis is currently being applied to high-throughput simulations.

V. SENSITIVITY AND CONVERGENCE UNDER DOUBLE-ENDED PATH SEARCHES

Whilst `TAMMBER` constructs a KTN through so-called 'open-ended' searches from some initial state, finding nearby connections, an alternative approach is to perform 'double-ended' searches for connections between possibly distant state pairs, using e.g. the doubly nudged elastic band method (`DNEB`)[42]. A successful search then returns a pathway of multiple

states, some or all of which were previously undiscovered. This is the main form of KTN construction in discrete path sampling simulations[43], which are often used to determine transition probabilities or mean first passage times between regions of state space.

Figure 3 shows a disconnectivity graph for the 38-atom Lennard-Jones cluster LJ_{38}. A 'double-funnel' landscape can be seen, with two low energy basins of cuboctahedral or icosahedral structure. The chosen observable is the 'committor' probability $C_{\mathcal{B}}^{\mathcal{A}}$, which gives the probability of a trajectory reaching $\mathcal{A}$ before returning to $\mathcal{B}$, a key quantity when calculating reaction rates[8].

For large networks at low temperature, observables such as $C_{\mathcal{B}}^{\mathcal{A}}$ can be so small that typical linear algebra routines fail due to floating point error. This has motivated the development of specialized renormalization methods[44], specifically designed to allow the exact evaluation of branching probabilities and mean first passage times. The underlying algorithms have recently been implemented, along with a variety of related methods in a `Python` package (`PyGT` module, `pygt.readthedocs.io`).

These techniques allow the production of a smaller network involving only the particular states of a given pathway, with renormalized transition rates such that branching probabilities are exact. One can then analytically evaluate the sensitivity to discovery of new pairwise connections[11] between any state pair.

As non-dynamic search methods have no rigourous uncertainty measures, consider a range of possible sensitivities, allowing an assesment of error based on their *consensus*. This is similar in approach to some machine learning measures of uncertainty[29].

A key observation is that real networks are typically sparse, meaning sensitivities can be tightened via an estimate of the sparsity ξ. Figure 3 shows three error measures, all of which converge at different rates: the bounds $\xi\sigma_{\pm}^{tot}$ sum all perturbations that increase or decrease $C_{\mathcal{B}}^{\mathcal{A}}$, multiplied by ξ. This typically produces a large uncertainty that decays slowly with sampling effort. The bounds $\sigma_{\pm}^{1}$ take the largest single sensitivity, whilst $\sigma_{\pm}^{\xi}$ multiplies the average sensitivity by the estimated number of remaining connections.

I note that all bounds are highly asymmetric due to the possibility of a direct connection with a low barrier. In the Bayesian analysis of dynamical trajectories discussed above this quickly has a vanishing probability. Furthermore, the lack of any rigorous error means the sensitivity bounds are non-monotonic, leaving large 'spikes' in the convergence plots that are quickly annuled once sampling is performed there. However, the use of double-ended

methods is extremely advantageous for the study of complex transition processes; this has motivated ongoing work to combine these consensus approaches with the rigourous state-wise uncertainty quantification employed by `TAMMBER`.

VI. VIBRATIONAL ANHARMONICITY IN REACTION RATES

The final section of this review focusses on a single thermally activated process, the fundamental ingredient of the rare event models discussed above. The transition rate between two states is given by the famous transition state theory expression[8]

$$\mathrm{k} = \omega_{ji} \exp(-\beta \Delta\mathcal{F}) \simeq \omega_{ji} \exp(-\beta \Delta\mathcal{F}_{harm}), \tag{3}$$

where $\beta = 1/\mathrm{k_B T}$, ω_{ji} is the attempt frequency, $\Delta\mathcal{F}$ is the maximum change in free energy during the transition and $\Delta\mathcal{F}_{harm} = \Delta\mathcal{U} - \mathrm{T}\Delta\mathcal{S}_{harm}$ is the widely used harmonic approximation of this change. $\Delta\mathcal{U}$ is the maximum energy change over the minimum energy pathway (MEP), which can be found using methods such as the nudged elastic band (`NEB`) method[45]. $\Delta\mathcal{S}_{harm}$ can be found by diagonalizing the Hessian matrix of second derivatives at the minimum and maximum of the MEP to find real vibrational frequencies $\{\nu^0\}$ and $\{\nu^\dagger\}$, then using the harmonic oscilator result $\mathcal{S} = \mathcal{S}_0 - \mathrm{k_B} \ln|\nu|$. In addition to making the uncontrolled assumption that thermal vibrations are harmonic oscilations, unlikely to hold at elevated temperature, the diagonalization scales cubically with the number of atoms in the system. This complicates application to the large systems routinely found when investigating linear or areal defects in materials[46], motivating the development of linear-scaling methods[47, 48] to evaluate harmonic TST prefactors by approximation of Hessian spectral densities.

The `PAFI` code is a linear-scaling method to evaluate $\Delta\mathcal{F}$ with no assumption on the nature of thermal vibrations. Starting from a converged `NEB` calculation, constrained sampling is performed on hyperplanes perpendicular to the MEP[10], as shown in figure 4. This constraint is available as part of the open-source `LAMMPS` package[49], designed to be used with the `PAFI` code (available at `github.com/tomswinburne/pafi`). One central innovation is to evaluate the inner product of the MFEP and MEP tangents, even though the MFEP is not known, allowing the MEP tangent force to be related to the MFEP tangent force. Providing these two paths are not orthogonal the true free energy gradient can then be ex-

tracted then integrated to produce $\Delta\mathcal{F}$. The computational cost of a single `PAFI` calculation is around $10^5 - 10^6$ force calls, depending on the degree of convergence and system under study. This is comparable to linear-scaling methods, which require around 10^6 force calls for well converged results[47]. Both methods have excellent parallel scalability, meaning this effort can be distributed across hundreds or thousands of processors.

Figure 4 shows the free energy barrier for the thermally activated migration of a $a/2\langle 111\rangle$ screw dislocation in tungsten, the prototypical process in bcc plasticity[46, 50]. To converge long-ranged elastic fields a large simulation size of over 10^5 atoms is required, where evaluation of $\Delta\mathcal{F}_{harm}$ is only possible with advanced shared memory linear algebra routines[46]. As can be seen, the harmonic and anharmonic predictions diverge at only 200K, around 10% of the melting temperature. The harmonic prediction of the resultant glide velocity (approximately given by the transition rate) is therefore many orders of magnitude too slow at realistic application temperatures. Due to the system size, this application used an empirical interatomic potential; current work is using advanced sampling schemes to implement more accurate force models.

VII. OUTLOOK

Quantification of sampling uncertainty in atomistic simulation is in many ways a nascent field, whose growth is strongly aligned with the recent explosion in machine learning methods applied to materials problems. There are many interesting directions in which unsupervised methods can be guided by uncertainty assesment algorithms, with many exciting opportunities for theoretical research.

This review summarised an approach that exploits the proof-of-work inherent in dynamical sampling to derive rigorous and robust uncertainty measures. I also summarized an approach that uses multiple less robust measures to build a consensus on convergence in double-ended search methods that are able to treat problems inaccessible to dynamic sampling. Current work is combining these two flavours of uncertainty quantification.

An important open question is how to quantify exploration of the vast atomic configuration space beyond measures of discovered novelty[23, 24, 28, 29]. This will certainly require a deeper understanding of the energy landscape of real materials, which may lead the way to new methods for computational discovery.

VIII. HIGHLIGHTED REFERENCES

[24] ** Uses active learning to build interatomic potentials on-the-fly for the accurate modeling of rare events.

[20] ** Uses speculation to asses optimal strategies for parallel trajectory generation.

[28] ** A powerful technique to gauge novelty in high dimensional materials data, with many implications for unsupervised simulation.

There are few published developments in this nacscent field, hence two self-highlights:

[7] ** Rigorous uncertainty quantification and propagation which is uses to build rare event models for materials.

[13] ** Autonomous scheme to evaluate of transport properties of defects with a rigorous convergence measure.

IX. ACKNOWLEDGEMENTS

It is a pleasure to warmly thank my collaborators in the work presented above, namely Danny Perez, Mihai-Cosmin Marinica and David Wales, for many stimulating discussions. I gratefully recognize support from the Agence Nationale de Recherche, via the MEMOPAS project ANR-19-CE46-0006-1. This work was granted access to the HPC resources of IDRIS under the allocation AP010910718 and A0070910965 attributed by GENCI. This work has been carried out within the framework of the EUROfusion consortium and has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission. Work at Los Alamos National Laboratory was supported by the U. S. Department of Energy, Office of Nuclear Energy and Office of Science, Office of Advanced Scientific Computing Research through the Scientific Discovery through Advanced Computing (SciDAC) project on Fission Gas Behavior. Los Alamos National Laboratory is operated by Triad National Security LLC, for the National Nuclear Security administration

of the U.S. DOE under Contract No. 89233218CNA0000001.

[1] B. P. Uberuaga and D. Perez, Handbook of Materials Modeling: Methods: Theory and Modeling (2020).

[2] A. F. Voter, in Radiation Effects in Solids (Springer, 2007) pp. 1–23.

[3] C. Le Bris, T. Lelievre, M. Luskin, and D. Perez, Monte Carlo Methods and Applications **18**, 119 (2012).

[4] T. Lelièvre, Handbook of Materials Modeling: Methods: Theory and Modeling (2020).

[5] L. K. Béland, P. Brommer, F. El-Mellouhi, J.-F. Joly, and N. Mousseau, Physical Review E **84**, 046704 (2011).

[6] D. Perez, E. D. Cubuk, A. Waterland, E. Kaxiras, and A. F. Voter, Journal of chemical theory and computation **12**, 18 (2015).

[7] T. D. Swinburne and M.-C. Marinica, Phys. Rev. Lett. **120**, 135503 (2018).

[8] P. Hänggi, P. Talkner, and M. Borkovec, Reviews of Modern Physics **62**, 251 (1990).

[9] G. Henkelman, Annual Review of Materials Research (2017).

[10] T. D. Swinburne and D. Perez, Phys. Rev. Materials **2**, 053802 (2018).

[11] T. D. Swinburne and D. J. Wales, Journal of Chemical Theory and Computation **16**, 2661 (2020).

[12] T. D. Swinburne, D. Kannan, D. J. Sharpe, and D. J. Wales, The Journal of Chemical Physics **153**, 134115 (2020).

[13] T. D. Swinburne and D. Perez, NPJ Computational Materials **6**, 190 (2020).

[14] D. J. Wales, Energy Landscapes, edited by C. U. Press (Cambridge, 2003).

[15] M.-C. Marinica, F. Willaime, and J.-P. Crocombette, Phys. Rev. Lett. **108**, 025501 (2012).

[16] K. Röder, J. A. Joseph, B. E. Husic, and D. J. Wales, Advanced Theory and Simulations **2**, 1800175 (2019).

[17] S. T. Chill and G. Henkelman, The Journal of chemical physics **140**, 214110 (2014).

[18] A. Chatterjee and S. Bhattacharya, The Journal of Chemical Physics **143**, 114109 (2015).

[19] D. Aristoff, S. Chill, and G. Simpson, Communications in Applied Mathematics and Computational Science **11**, 171 (2016).

[20] A. Garmon and D. Perez, Modelling and Simulation in Materials Science and Engineering **28**, 065015 (2020).

[21] N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti, G. Pizzi, et al., Nature nanotechnology **13**, 246 (2018).

[22] S. P. Ong, Computational Materials Science **161**, 143 (2019).

[23] N. Bernstein, G. Csányi, and V. L. Deringer, npj Computational Materials **5**, 1 (2019).

[24] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, npj Computational Materials **6**, 1 (2020).

[25] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Science advances **3**, e1701816 (2017).

[26] A. M. Goryaeva, J.-B. Maillet, and M.-C. Marinica, Computational Materials Science **166**, 200 (2019).

[27] A. V. Shapeev, Multiscale Modeling & Simulation **14**, 1153 (2016).

[28] A. M. Goryaeva, C. Lapointe, C. Dai, J. Dérès, J.-B. Maillet, and M.-C. Marinica, Nature Communications **11**, 1 (2020).

[29] M. Wen and E. B. Tadmor, npj Computational Materials **6**, 1 (2020).

[30] A. Glensk, B. Grabowski, T. Hickel, and J. Neugebauer, Physical Review X **4**, 011018 (2014).

[31] T. D. Swinburne, J. Janssen, M. Todorova, G. Simpson, P. Plechac, M. Luskin, and J. Neugebauer, Physical Review B **102**, 100101 (2020).

[32] M. Sorensen and A. Voter, The Journal of Chemical Physics **112**, 9599 (2000).

[33] L. J. Munro and D. J. Wales, Physical Review B **59**, 3969 (1999).

[34] N. Mousseau and G. T. Barkema, Physical Review E **57**, 2419,2424 (1998).

[35] G. Henkelman and H. Jónsson, The Journal of Chemical Physics **111**, 7010 (1999).

[36] S. T. Chill, J. Stevenson, V. Ruehle, C. Shang, P. Xiao, J. D. Farrell, D. J. Wales, and G. Henkelman, Journal of chemical theory and computation **10**, 5476 (2014).

[37] G. C. Boulougouris and D. Frenkel, Journal of chemical theory and computation **1**, 389 (2005).

[38] G. C. Boulougouris and D. N. Theodorou, The Journal of chemical physics **127**, 084903 (2007).

[39] D. R. Trinkle, Physical review letters **121**, 235901 (2018).

[40] M. A. Novotny, Phys. Rev. Lett. **74**, 1 (1995).

[41] T. M. Cover and J. A. Thomas, Elements of information theory (John Wiley & Sons, 2012).

[42] E. F. Koslover and D. J. Wales, J. Chem. Phys. **127**, 134102 (2007).

[43] D. J. Wales, Mol. Phys. **100**, 3285 (2002).

[44] S. A. Trygubenko and D. J. Wales, J. Chem. Phys. **124**, 234110 (2006).

[45] G. Henkelman, B. P. Uberuaga, and H. Jonsson, The Journal of Chemical Physics **113**, 9901 (2000).

[46] L. Proville, D. Rodney, and M. C. Marinica, Nature Materials (2012).

[47] C. Huang, A. F. Voter, and D. Perez, Physical Review B **87**, 214106 (2013).

[48] L. Lin, Y. Saad, and C. Yang, SIAM review **58**, 34 (2016).

[49] S. Plimpton, Journal Computational Physics **117**, 1 (1995).

[50] T. Swinburne and S. Dudarev, Physical Review Materials **2**, 073608 (2018).

[51] A. Stukowski, Modelling and Simulation in Materials Science and Engineering **18**, 015012 (2010).

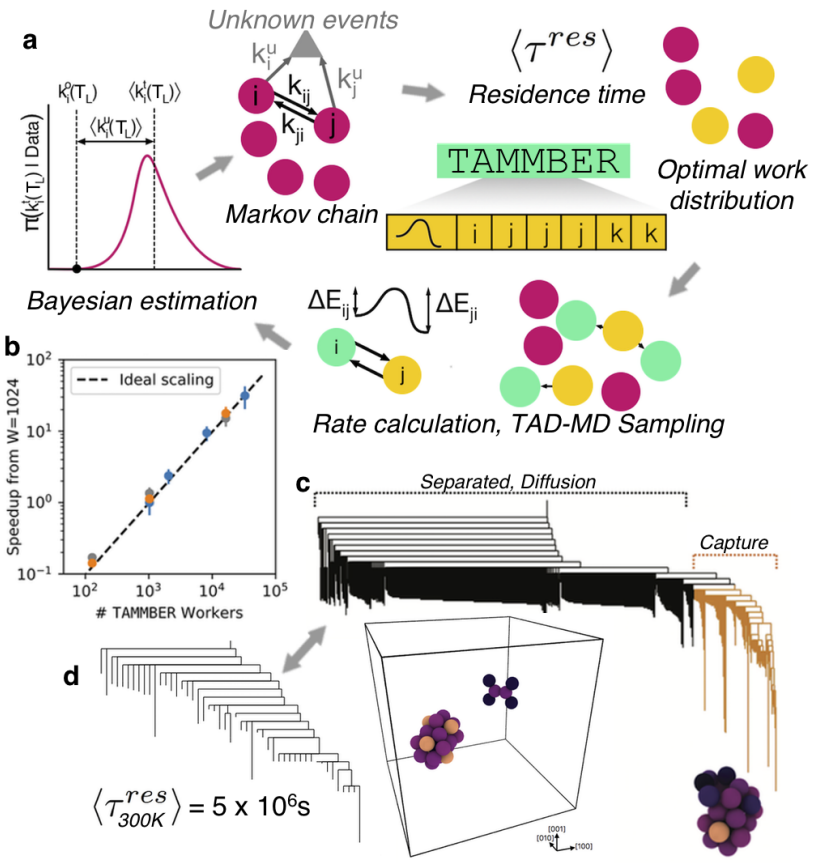

FIG. 1. Outline of the TAMMBER code. a) Sampling data is used to estimate unknown rates $\{k_i^u\}$ and barrier calculations determine known rates $\{k_{ji}\}$. An absorbing Markov chain is built, ~~yeilding~~yielding a residence time in the known state space and an updated work distribution. b) Tests show efficient management of $10^2 - 10^5$ simultaneous sampling tasks. c) Application to the interaction of interstitial clusters in Fe[7, 15] reveals a complex energy landscape[14] (main text). d) This sampling can be accelerated by collating states symmetric under exchange and space symmetries, ~~yeilding~~yielding large prediction timescales for the complex break-up process of these clusters. Adapted from [10] and [13].

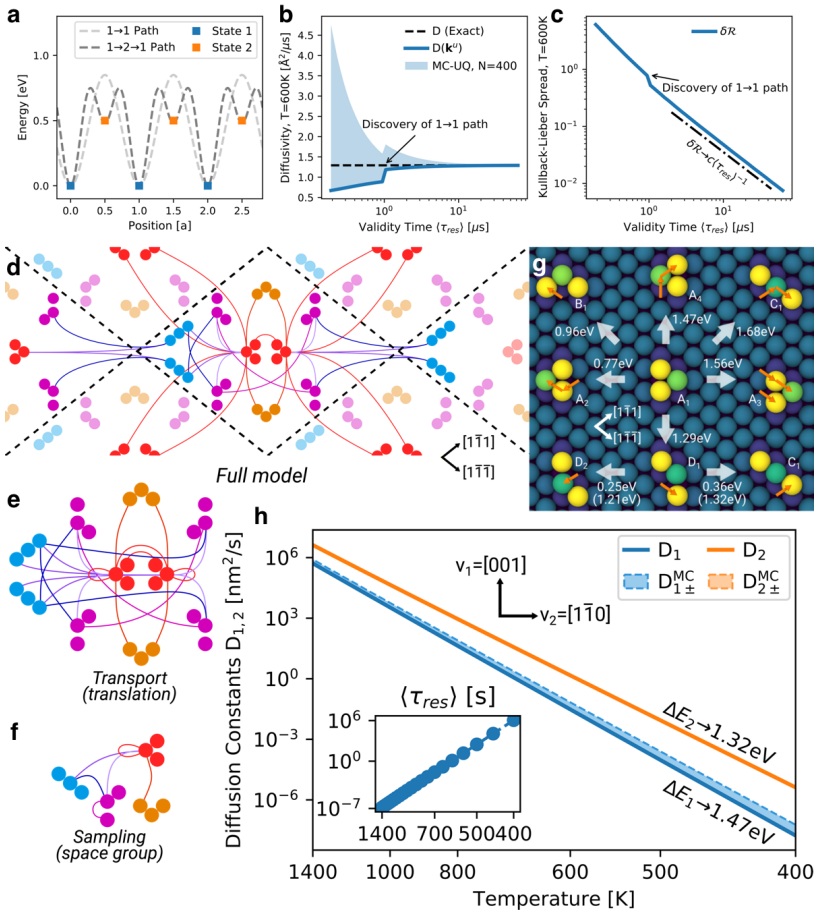

FIG. 2. Convergence of diffusion constants in a simple one dimension example (a-c) and with TAMMBER applied to trimer diffusion on W(110) (d-h). a) The one dimensional toy model has a direct and indirect transition. b) The estimated diffusivity with sampling time with Monte Carlo bounds. c) The Kullback-Lieber convergence measure, which is monotonically decreasing. d) Cartoon of the full akMC state space and those irreducible under space and translation symmetries. Isomorphic states have the same color. e) One primitive unit cell with self transitions (closed loops) is sufficient to build a Markov model for transport, whilst f) the set of states irreducible under all space group symmetries is optimal for discovery. g) The lowest energy irreducible states and found transition mechanisms. Atoms colored by centrosymmetry[51]. h) Diffusion tensor eigenvalues D_1, D_2 over a range of temperatures with bounds $D_{1\pm}^{MC}$, $D_{2\pm}^{MC}$. Effective Arrhenius slopes at low temperature are given. Adapted from [13].

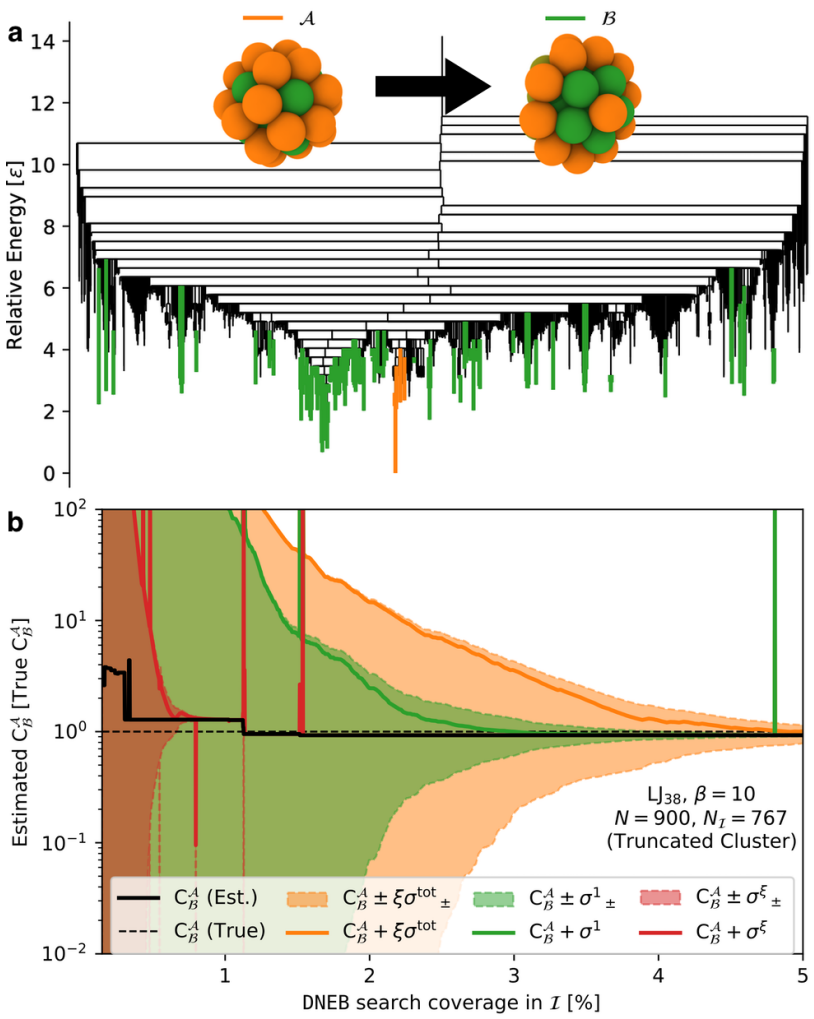

FIG. 3. Convergence of nanoparticle transformation paths under pairwise doubly nudged elastic band (`DNEB`) searches. Above: Disconnectivity graph[14] for LJ_{38}, truncated to 900 states. The cuboctahedral ($\mathcal{A}$) and icosahedral ($\mathcal{B}$) states are coloured orange and green, respectively. Minimal free-energy configurations from each basin are shown. Below: Convergence of the $\mathcal{B} \to \mathcal{A}$ committor probability $C_{\mathcal{B}}^{\mathcal{A}}$ under a variety of measures detailed in the main text. Whilst no one measure is as robust as the dynamical method of `TAMMBER`, their consensus gives a useful measure of convergence. Adapted from [11].

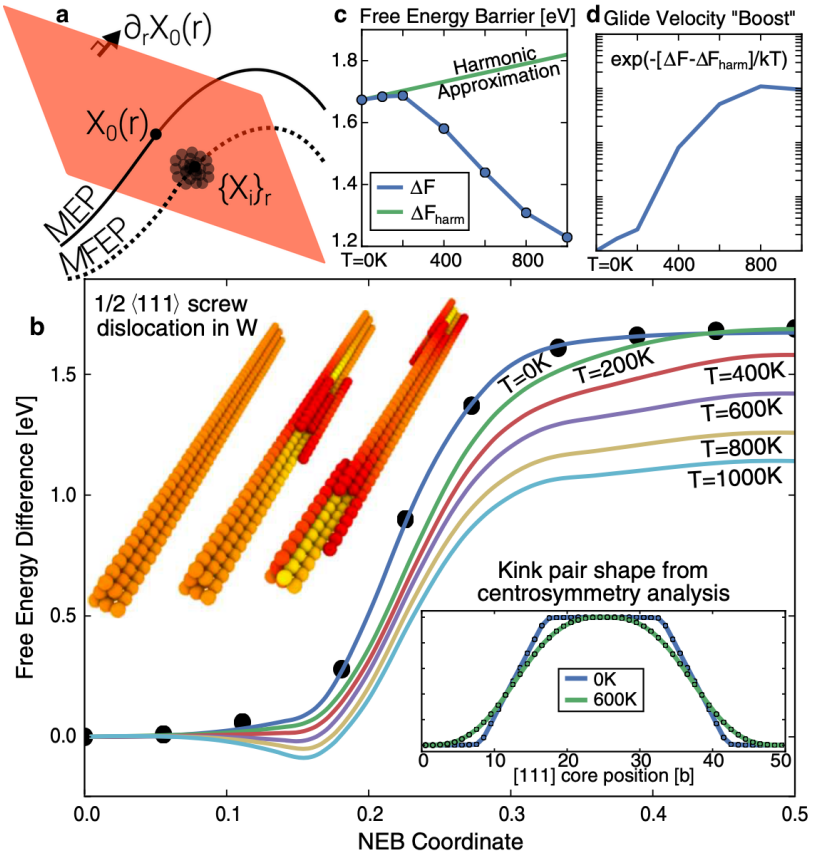

FIG. 4. Activation free energy profile for $1/2\langle 111\rangle$ screw dislocation migration in an empirical model of tungsten, as evaluated by `PAFI`. a) Cartoon of the constrained sampling methodology. b) The free energy difference $\Delta \mathrm{F}(r) = \mathrm{F}(r) - \mathrm{F}(0)$ for a range of temperatures. Inset: the double kink profile at $\mathrm{T} = 0, 600\mathrm{K}$, showing how the pathway changes with temperature. c) The free energy barrier shows significant anharmonic effects at only 200K, with d) considerable influence on the dislocation velocity. Adapted from [7].