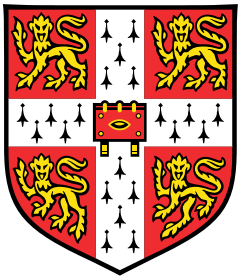
Developing an off-lattice kinetic Monte Carlo model to simulate hydrogen-crystal defect interactions in α–iron



Conor John Williams

Supervisor: Dr E I Galindo-Nava

MPhil in Scientific Computing
University of Cambridge

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Masters of Philosophy

Churchill College August 2020

Abstract

The project aims to implement an off-lattice kinetic Monte Carlo model, based on the dimer saddle-point finding method, suitable for simulating the interaction of hydrogen with crystal defects in iron, such as vacancies, grain-boundaries and dislocations, over timescales not achievable with classical methods including molecular dynamics. The framework is used to interrogate the mechanisms of hydrogen embrittlement, causing severe loss in ductility in steels with complicated microstructures, focusing on hydrogen-vacancy systems.

Acknowledgements

I would like to thank my supervisor Dr Enrique Galindo-Nava, for his support and guidance throughout this project, as well as his consistently-excellent ideas and criticisms of the research.

Table of contents

List of figures

Nomenclature

Roman Symbols

C	Curvature
D	Diffusivity
D_0	Diffusivity prefactor
E	Energy
$\boldsymbol{H}$	Hessian matrix
$\overline{\boldsymbol{H}}$	Inverse Hessian matrix
$\boldsymbol{I}$	Identity matrix
$\mathbb{N}$	Natural numbers, including zero
$\mathbb{R}$	Real numbers
U	Potential energy

Greek Symbols

Γ_{ij}	Rate constant linking states $i \rightarrow j$
κ	Isopotential curvature
ρ	Density

Acronyms / Abbreviations

AIDE	Adsorption-induced decohesion
Al	Aluminium

BCC	Body-centred cubic
BFGS	Broyden–Fletcher–Goldfarb–Shanno
CG	Conjugate gradient
CPU	Central processing unit
DFT	Density functional theory
EAM	Embedded atom model
FCC	Face-centred cubic
Fe	Iron
GB	Grain boundary
GPU	Graphics processing unit
HCP	Hexagonally closed packed
HE	Hydrogen embrittlement
HELP	Hydrogen-enhanced local plasticity
HESIV	Hydrogen-enhanced strain-induced vacancy formation
H	Hydrogen
HID	Hydrogen-induced decohesion
HIPT	Hydrogen-induced phase transformation
HTST	Harmonic transition state theory
KMC	Kinetic Monte Carlo
L-BFGS	Limited-memory Broyden–Fletcher–Goldfarb–Shanno
LCL	Linked cell list
LCR	Linked cell reordering
LEPS	London–Eyring–Polany–Sat

MC	Monte Carlo
MD	Molecular dynamics
MH	Metropolis–Hastings
NEB	Nudged elastic band
nn	Nearest neighbour
OLKMC	Off-lattice kinetic Monte Carlo
PES	Potential energy surface
PR	Polak–Ribière
PS	Projection sorting
RDF	Radial distribution function
SP	Saddle point
V	Vacancy

Chapter 1

Introduction

The objective of this project is to develop and implement, from the ground up, a tool capable of modelling the complex interactions between crystal defects and hydrogen (H) in iron (Fe), into the timescales required to study hydrogen embrittlement (HE). Hydrogen embrittlement is the long-studied phenomena wherein, metals exposed to hydrogen undergo a catastrophic ductile to brittle transformation, causing failure below the yield stress. Despite years of study, the full mechanisms of HE are not yet fully understood [6].

The simulation technique chosen is the off-lattice kinetic Monte Carlo (OLKMC) model, this is an extension to the traditional kinetic Monte Carlo (KMC) algorithm, described in Section 1.2. KMC is the lowest level of theory that can currently be used to model non-equilibrium processes at the atomistic scale into the timescales required to investigate HE, up to the order of seconds. OLKMC extends KMC to a continuous description of the system, this massively expands KMC to the generality of molecular dynamics (MD) and frees the method from the need to know all possible mechanisms *a priori*. This is particularly important as there are numerous cases when the atomistic mechanism of some process has been incorrectly identified [23, 20], as mechanisms can significantly deviate from the simplest pathway. OLKMC operates on a set of atoms in 3D space and uses a force field to describe the interaction between atoms. This force field is the limiting level of theory in OLKMC; we choose to use a semi-empirical force field, placing our model at the same theoretical baseline as MD.

The decision to work from the ground up rather than building on other implementations [68, 52] was twofold. Firstly, the techniques employed in OLKMC – fully explored in Chapter 3 – have rapidly evolved over the last twenty years, they are reviewed in Chapter 2. Hence, older implementations lack critical optimisations, reworking them at this stage would be more complex than restarting. Secondly, a full understanding of the method, required to develop

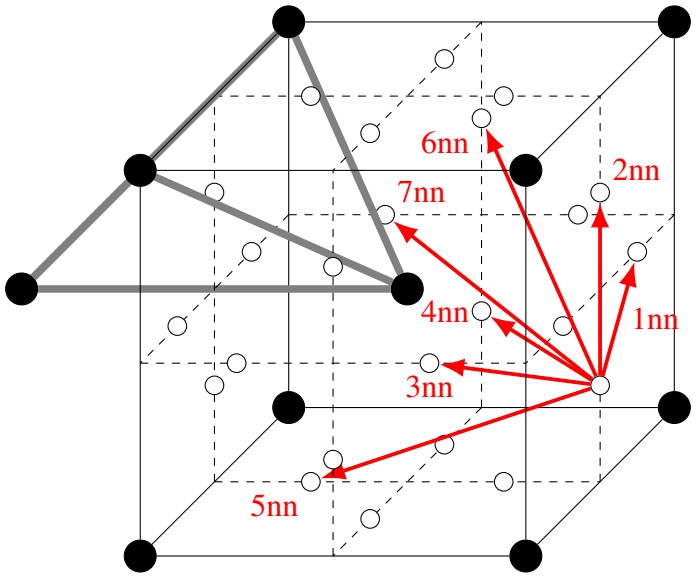


Fig. 1.1 Diagram of the BCC unit cell of Fe atoms (black) alongside the tetrahedral interstices marked (unfilled circles). The lattice parameter is 2.8557Å. The thick grey lines act as a guide to the eye to highlight the tetrahedral shape. Hydrogen atoms can occupy the tetrahedral interstices; marked with red arrows are the seven nearest-neighbour (nn) tetrahedral-tetrahedral adjacencies. The corresponding nn separations are: 1.01, 1.43, 1.75, 2.02, 2.26, 2.47 and 2.67Å, respectively.

the theory in future work, can only be achieved by exploring every subroutine and recursively understanding them and their limitations.

The remainder of the report is structured as follows: Section 1.1 is a brief summary of the Fe-H system and HE; Section 1.2 summarises the long-established theory of KMC as a primer to introduce OLKMC; Chapter 2, the literature review, explores the progression of the OLKMC and current state-of-the-art methods; Chapter 3 details the theory and implementation of all the numerical methods employed in this work surrounding OLKMC; in Chapter 4 we verify our implementation and methodology across a series of well explored systems; in Chapter 5 we present the results and discussion of modelling hydrogen, vacancy (V) and V-H complex diffusion in the α-Fe lattice using OLKMC and finally, in Chapter 6 we present our conclusions and suggestions for future work.

1.1 The Fe-H system

Iron and its alloys are particularly important engineering materials therefore, the focus shall be HE in the α-Fe allotrope which is the primary phase of mild steels. α-Fe, or ferrite, has a body-centred cubic (BCC) lattice structure with two atoms per unit cell, presented diagrammatically in Fig. 1.1. There exist two types of interstitial sites in the BCC lattice, tetrahedral and octahedral.

Experimental measurement and density functional theory (DFT) calculations have shown that hydrogen preferably dissolves in the tetrahedral sites with a dissolution energy of approximately 0.3eV [75, 42]. There are 12 tetrahedral sites per unit cell, they are marked in Fig. 1.1.

Hydrogen has a low solubility in iron but high diffusivity [29] but, it is understood crystal defects can act as traps for diffusing hydrogen. Hydrogen can bind strongly to vacancies, dislocations and grain boundaries (GB), this reduces hydrogen mobility/diffusivity. It is therefore common to distinguish between between freely *diffusible hydrogen* in the lattice and *trapped hydrogen* bound to crystal defects. There exists an entire theoretical understanding of hydrogen trapping in equilibrium such as Oriani's trapping theory [64].

1.1.1 Hydrogen embrittlement

Many mechanisms for HE have been proposed, almost all of them revolve around the interactions between hydrogen and crystal defects, which have long been understood to play a critical role in the deformation of metals. For a full reference to all of the following mechanisms see the review by Barrera et al. [6]. In extreme brief, the predominant HE mechanism theories fall into several categories:

- Hydrogen-enhanced localised plasticity (HELP) postulates that, possibly through the formation of Cottrell-like atmospheres, hydrogen increases the mobility of dislocations causing increased plasticity, localised to regions containing hydrogen.
- Hydrogen-enhanced strain-induced vacancy (HESIV) formation explains HE as a consequence of hydrogen causing a superabundance of vacancies, which coalesce to form nanovoids/microvoids, offering low-energy pathways for crack propagation.
- Hydrogen-induced decohesion (HID) is the one of the oldest ideas [70] and suggests that the presence of hydrogen in the α-Fe lattice weakens the Fe-Fe bonding and reduces the strain at which decohesion occurs, e.g. at grain boundaries.
- Adsorption-induced decohesion (AIDE) likens HE to stress corrosion cracking, implying the adsorption of hydrogen at the crack tip facilitates dislocation emission and is responsible for hydrogen-induced cracking.
- Hydrogen-induced phase transformation (HIPT) predicts high hydrogen concentrations triggers embrittling phase transformations. In particular, martensitic transformations of the face-centred cubic (FCC) γ phase to hexagonally closed packed (HCP) or BCC phases.

The bodies of evidence supporting many of these ideas makes it probable that a full description of HE contains a combination of two/three of these mechanisms (alongside some yet undiscovered) as few are orthogonal to each other.

The richness in mechanisms is vast – due to the breadth of hydrogen interactions with crystal defects – and we will restrict this work to explore V-H interactions as a case-study for validating our method. Vacancies, as point-defects, are the simplest of crystal defects, so are a natural starting point. Nevertheless, the ideas of the HESIV theory suggests their behaviour are no less significant to understanding HE.

1.2 Kinetic Monte Carlo

Interesting events/long-timescale dynamics are dominated by rare events which are often outside the reach of MD. By ignoring atomic vibrations and representing the time evolution of a system as a series of discrete transitions between local minima or *basins*, KMC overcomes this hurdle. Kinetic Monte Carlo is part of a wider collection of stochastic Monte Carlo (MC) methods that use random numbers to walk Markov chains through state space and approximate a distribution of interest. At the cost of losing the explicit trajectory of atoms we can simulate over much longer timescales than MD, up to seconds for some systems.

The connection between basins in the KMC framework is the set of rate constants Γ_{ij} that represent the probability per unit time or frequency of making the transition from state $i \rightarrow j$. These rate constants are independent of the systems history, thus forming a Markov chain. It is the use of rates linking states (as opposed to energy changes, used in the Metropolis–Hastings (MH) algorithm[36]) that enable KMC to probe non-equilibrium phenomena.

Under the kinetic Monte Carlo framework, using the rejection-free n-fold way algorithm [13], with the system in state i, the next state k is selected as the solution to:

$$\sum_{j=1}^{k-1} \Gamma_{ij} < \rho_1 \sum_{j=1}^{n} \Gamma_{ij} \leq \sum_{j=1}^{k} \Gamma_{ij} \tag{1.1}$$

where $\rho_1 \in (0, 1]$ is a uniform random number and $j, k \in \{1, 2, \ldots, n\}$. This ensures the probability of transitioning to a new state is proportional to its rate.

The KMC walk models a Poisson process and therefore the time elapsed during a single MC step is [12]:

$$\Delta t = \frac{-\ln(\rho_2)}{\sum_{j=1}^{n} \Gamma_{ij}} \tag{1.2}$$

where $\rho_2 \in (0, 1]$ is a second uniform random number. Both equations require knowledge of the complete set of non-zero rate constants connected to a given state.

1.2.1 Harmonic transition state theory

Within the approximation of harmonic transition state theory we can obtain the set of rate constants from the potential energy surface (PES); the rate constant connecting $i \rightarrow j$ via a single, first-order, saddle-point $\ddagger$ is described by the Arrhenius equation [82]:

$$\Gamma_{ij}^{\mathrm{TST}} = \tilde{\nu}_{ij} e^{-\beta\left(E^{\ddagger} - E^{i}\right)} \tag{1.3}$$

where $\beta = \frac{1}{k_B T}$, $\tilde{\nu}_{ij}$ is the attempt frequency and $E^{\ddagger}$, E^{i} are the energies of the system at the saddle point and state i, respectively. This approximation holds well for metals away from their melting point as the atoms are rigidly held in place and the basins well approximated by a quadratic expansion [82]. Furthermore, assuming the PES is quadratic near the saddle point then:

$$\tilde{\nu}_{ij} = \frac{\prod_{k=1}^{N} \nu_k^i}{\prod_{k=1}^{N-1} \nu_k^{\ddagger}} \tag{1.4}$$

where $\nu_k^{\ddagger}$, ν_k^i are the real normal-mode frequencies at the saddle point and state i respectively. It is common to set $\tilde{\nu}_{ij} = \nu$, a constant [52, 81] due to the high cost to compute ν_{ij}, the fact it varies little between mechanisms in the same metal and the exponential influence of $\left(E^{\ddagger} - E^{i}\right)$ on $\Gamma_{ij}^{\mathrm{TST}}$ having a much larger affect than the linear multiplier $\tilde{\nu}_{ij}$.

Therefore, we have reduced the problem of finding all non-zero rate constants to finding the energies of all the saddle points adjacent to some initial basin that connects it to surrounding basins. This is the fundamental idea behind OLKMC; we no longer require any kind of lattice to discretise the possible transition that can occur, only a method for finding saddle-points of the potential energy.

Chapter 2

Literature review

As we developed our OLKMC implementation from the ground up, a broad literature review was conducted. We explored not only the saddle-point (SP) search methods that OLKMC is directly built upon, but also the broader collection of numerical methods for function optimisation and empirical force field calculations. We shall now summarise the key developments in the literature, in a approximately chronological order, that this is work is built-upon.

2.1 Off-lattice kinetic Monte Carlo methods

Henkelman and Jónsson's [38] (2001) seminal paper is the first to describe the exciting OLKMC method, extending KMC simulations to off-lattice environments with no preconceived notion of transition pathways/mechanisms. This is achieved through a search of the potential energy landscape, using their efficient *dimer method* [37], to locate saddle points adjacent to the occupied local minima. The harmonic transition state theory (HTST) approximation is then applied to find the corresponding transition rates; then time progressed with the traditional KMC algorithm and the system is progressed to an adjacent basin. This process is summarised in Fig. 2.1 in a model potential. Henkelman and Jónsson report finding several, novel 3/4 atom processes for adatom descent on the Al(100) surface, that are rate competitive with the standard 2 atom process. Many potential optimisations such as SP caching and active volume partitioning have yet to be explored.

El-Mellouhi, Mousseau, and Lewis's [52] (2008) crucial contribution to the off-lattice KMC literature is the introduction of local topological classification, based on the `nauty` [50] graph library. This allows for reconstruction of saddle points from those previously discovered, significantly reducing the computational workload by avoiding many SP searches. Although the topological classification may not exploit all geometric information available (and relies on

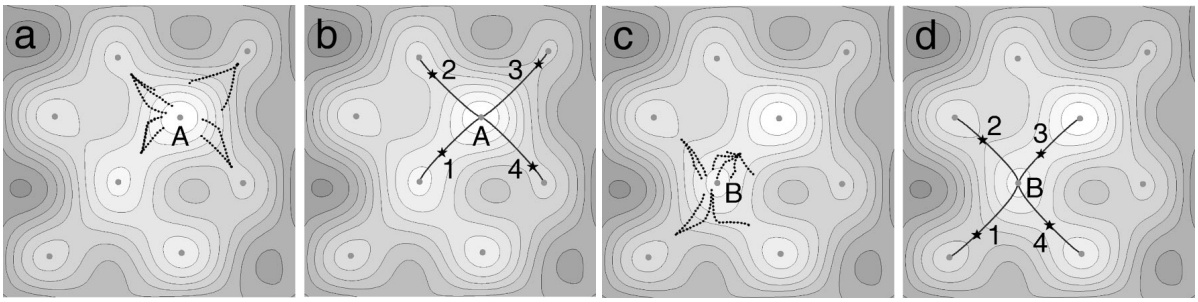


Fig. 2.1 Diagram of the OLKMC method, adapted from Henkelman and Jónsson [38], using a model two-dimensional potential. The simulation starts in basin A in Fig. 2.1a, several SP searches are launched (dotted lines), converging onto saddle points (stars). In Fig. 2.1b the pathways/mechanisms to adjacent basins via the saddle-points are highlighted. The KMC algorithm is applied between Fig. 2.1b and Fig. 2.1c, choosing the mechanisms which moves the system to basin B. The process then repeats, identifying the mechanisms out of basin B in Fig. 2.1d.

heavy graph algorithms), it produces an impressive speed up, enabling OLKMC to compete with traditional KMC for highly symmetric problems. Local topological classification can completely elide SP searches after translationally symmetric mechanisms. This is significant, as previously OLKMC was only superior to traditional KMC when the problem was sufficiently complex that the mechanisms and rate constants were unknown or too diverse to categorise.

OLKMC is particularly promising for exploring HE, as it requires no *a priori* knowledge of the mechanisms, meaning it introduces no human bias into the search. Furthermore, it is capable of reaching HE timescales and removes the constraint of hand-coding a lattice, appropriate for every crystal defect; a recent comprehensive review by Trochet et al. [81] fully explores OLKMC. We shall now discuss some of the papers expanding on the dimer method.

2.1.1 The dimer method

Henkelman and Jónsson [37] (1999) presented an efficiently scaling method to determine the saddle-points surrounding a potential-energy well, named the dimer method. The dimer is an object, which explores the potential energy surface (PES) and is constructed from a pair of images of the system displaced by a small amount. Successive rotations, minimising the total energy, and translations, following an effective force, orient the dimer with lowest curvature mode and move the dimer towards a saddle point. Figure 2.2 summarises how the dimer converges to a SP. By varying the initial orientation of the dimer and perturbing the initial state, different saddle points can be found. The authors are able to apply the dimer method to find saddle points for reaction pathways on the Al(100) surface using an embedded atom model (EAM) potential.

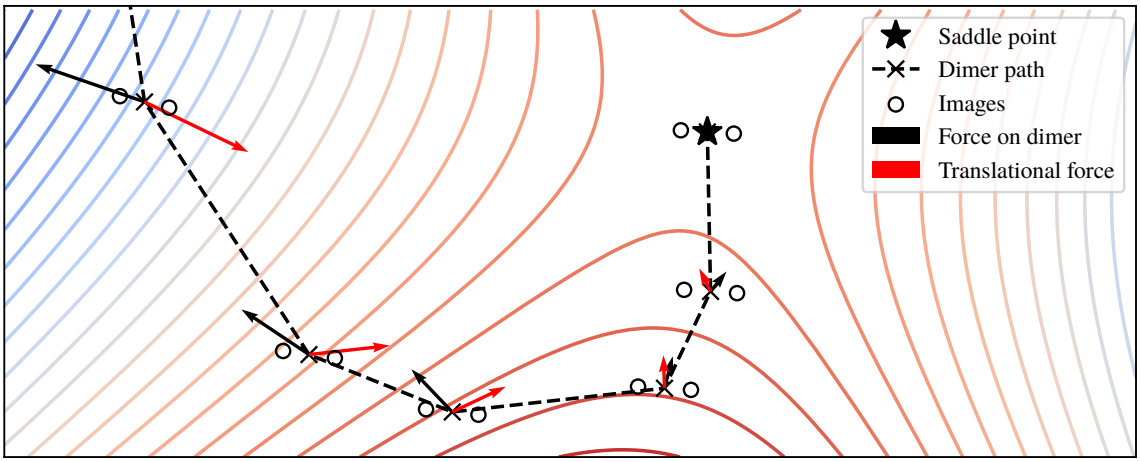


Fig. 2.2 Diagram of a dimer converging to a SP in a model two-dimensional potential. The dimer is rotated between each step to minimise its rotational potential energy, aligning its axis with the minimum-mode. The displacement between image-pairs for each dimer has been exaggerated for clarity. The force and the effective translational force on the dimer are marked for each step.

This improves over previous minimum-mode following algorithms [14] that require calculation (involving second order derivatives) and inversion of the Hessian matrix, a computationally expensive process. The dimer method requires only the initial basin to locate saddle points, unlike the nudged elastic band (NEB) [43] method, which also requires the target basin. The dimer method is particularly suited to the Fe-H system as good, fast semi-empirical potentials exist [75, 80].

Heyden, Bell, and Keil [39] (2005) constructed the *improved dimer method*, building on the original dimer method [37], which halves the number of force calculations during dimer translations and reduces the number of force calculations during rotation from six to four. This is achieved by working with the potential gradients at the centre and one end of the dimer (as opposed to each end), such that the central gradient can be reused during rotations. Furthermore, through a Fourier series expansion of the curvature and approximation of its second derivative, a better guess for the dimer rotation can be made and a rapid convergence to the minimum mode is achieved. In all, these are an effortless set of optimisations to be included.

Kästner and Sherwood [45] (2008) further optimised the improved dimer method [39]. This is achieved by exploiting the planarity of the rotated dimer to extrapolate one of the gradients at the dimer end, thus saving a force call per rotation. Additionally, the limited-memory formulation of the Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) optimiser [49, 62] is employed to determine the plane of rotation and dimer translation step. This is demonstrated to converge faster than conjugate gradient (CG) methods and removes an energy calculation

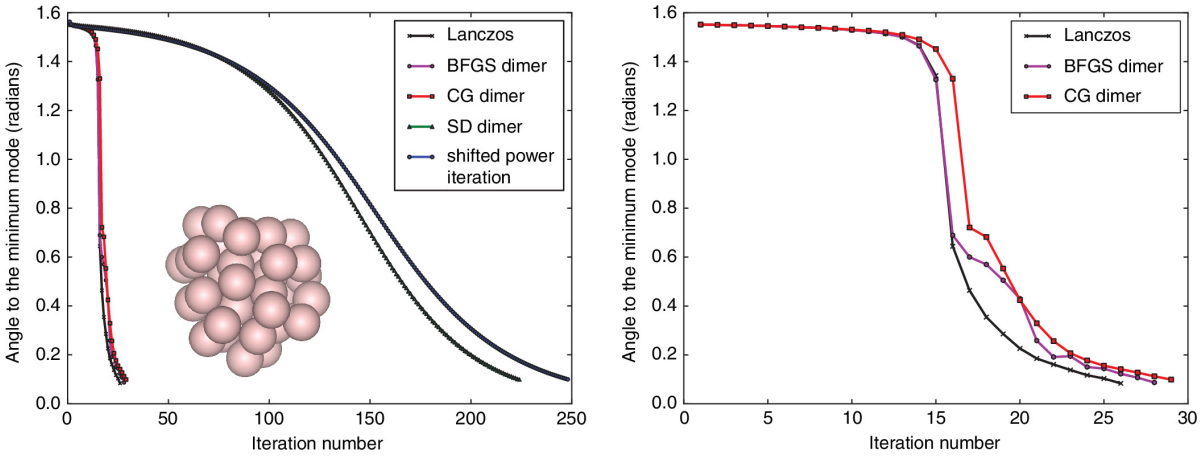


Fig. 2.3 A comparison of several minimum-mode finding algorithms, taken from Zeng, Xiao, and Henkelman [88]. To the right is a zoom, on the region of relevance, for the fastest three methods.

per translation. The convergence of the dimer to a saddle point is shown to scale superlinearly with the number of degrees of freedom and is thus referred to as the *superlinear dimer method*. The authors do not fully describe the coupling of their dimer method to the L-BFGS optimiser however, their results clearly demonstrate improvement.

Xiao, Wu, and Henkelman [86] (2014) began with the original dimer method [37] and attempt to avoid locating saddle points disconnected from the initial local minima. A common problem that occurs in OLKMC SP searching, that results in useless saddle points. This is achieved by weighting the parallel and perpendicular components of the effective translational forces applied to the dimer, by functions of the local isopotential curvature, κ. The isopotential curvature is shown to typically have positive sign in regions near the true ridge between basins of the potential energy thus, the dimer can be prevented from crossing the true ridge. The *κ-dimer method* is shown to find connected saddles $> 99\%$ of the time (compared to $\approx 70\%$ for the traditional dimer method) at the cost of twice as many force calls.

Finally, **Zeng, Xiao, and Henkelman [88] (2014)** unified several minimum-mode following algorithms (including the dimer method) under one mathematical framework. They prove all investigated methods are bounded in efficiency by the Lanczos method [48]. In numerical tests the L-BFGS dimer method was shown to converge almost as fast as the Lanczos method but require fewer force evaluations, their results are reproduced in Fig. 2.3. This is the basis for our choice to uses the dimer method.

2.2 Empirical potentials for the Fe-H system

As discussed in Chapter 1 the limiting level of theory in OLKMC is the underlying potential. As we wish to simulate HE timescales an *ab initio* approach, such as DFT, is well out of reach due to the poor quantum scaling. Therefore, an accurate fast semi-empirical potential is required; ideally, well tested and incorporating fitting targets from quantum calculations and experiment. Accurately reproducing DFT results for crystal defect structures/interaction is a long standing challenge for semi-empirical potentials and is pivotal for exploring HE as defect-H interactions are understood to play such an important role.

Finnis and Sinclair's [25] (1984) *embedded atom model* (EAM) is an empirical potential for describing the metallic bonding in transition metals. The EAM potential form includes a traditional pairwise repulsive component alongside a new embedding term. It approximately takes into account the band character, via the second-moment approximation to the tight-binding model thus, scaling the cohesive energy with the atomic coordination, z, as $\sqrt{z}$. The authors fit their model to a selection of experimental data for several transition metals and are able to successfully describe the vacancy formation energy, something not previously possible with pairwise potentials due to their lack of generality and intrinsic mathematical limitations. Crucially, EAM potentials are sufficiently short range to enable fast MD simulations and have therefore become well established in the literature.

Ramasubramaniam, Itakura, and Carter [75] (2009) presented a set of modified EAM potential for the Fe-H system, the EAM embedding function is generalised to obtain better fits to first-principles (DFT) measurements and experimental data. The wide variety of the fitting targets (perfect crystal, point defect, surface measurements, etc) provide good reproduction of several crystal defect structures, including the sixfold symmetry of screw dislocation cores. The authors investigate the V-H and screw-H interaction energies and demonstrate the potentials provide good matches to DFT studies. The results show that semi-empirical potentials can be used to study defect-H interactions, to probe HE mechanisms.

Finally, **Song and Curtin [80] (2012)** built upon the potentials of Ramasubramaniam, Itakura, and Carter [75], introducing an additional H-H repulsive component to reduce the observed tendency for hydrogen clustering in the original potentials. The new potentials are used in a MD crack-tip simulation in α-Fe; HE via suppression of dislocation emission is observed. The work highlights the power of semi-empirical potentials for investigating HE mechanisms as well as the need for further improving said potentials. Furthermore, the results demonstrate that, although empirical potentials are not yet perfect, they can be used to study HE and obtain insightful results.

2.3 Numerical methods

Liu and Nocedal [49] (1989) introduce a limited-memory version of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimiser named the L-BFGS method. This method is an extension of CG methods – storing a fixed (user controlled) amount of search history – that approximates the inverse Hessian and accelerates convergence. Furthermore, scaling methods are presented to allow for good initial estimates of the step size thus, eliding many line searches. The L-BFGS method is shown to converge rapidly for large problems and outperform other quasi-Newton methods. This is the optimiser Kästner and Sherwood employed in the superlinear dimer method; care must be taken during the line-search procedure as Liu and Nocedal make no provision for the failure to uphold the strong Wolfe conditions [84, 85].

Welling and Germano [83] (2011) compared a number of linked cell list (LCL) algorithms that efficiently determining the neighbours within some cut of radius r_{cut} of an atom, to accelerate force calculations. Key considerations for their algorithms include the cell size to r_{cut} ratio, cyclic boundary implementation, cache locality and projection sorting (PS) methods. They conclude the linked cell reordering algorithm (LCR) is the fastest of the non-PS methods. Due to the high complexity of PS methods, but relatively small performance gains reported, we conclude the LCR methods are the superior choice. This paper forms the basis for our EAM force field implementation.

2.4 Atomistics of hydrogen in metals

Here we summarise some works modelling HE and using OLKMC to appreciate current, cutting-edge methods and establish the breadth of applicability of OLKMC.

Katzarov, Pashov, and Paxton [46] (2017) used a discretised line-model of the α-Fe $\frac{1}{2}[111]$ screw dislocation, embedded in a hexagonal lattice and the traditional KMC algorithm to study its time evolution. The transition rates are parametrised on experimental measurements and DFT simulations with and without hydrogen, as well as empirical fits to MD simulations.

The simulations show that screw dislocation diffusivity is strongly dependant on hydrogen concentration, applied stress and temperature. In low concentrations, hydrogen enhances dislocation mobility, but inhibits mobility at high concentrations. Some transitions/mechanisms such as edge segment climb, as well as hydrogen interaction with the surrounding lattice, have been omitted for simplicity. This, alongside the $\frac{1}{2}[111]$ screw dislocation specific nature of the paper, suggest it would be challenging to extend this kind of model to include multiple kinds of

crystal defects and richer hydrogen interactions but, the potential for atomistic simulations to probe HE is clear.

Benediktsson et al. [9] (2013) explored the energy landscape of V-H complexes in aluminium (Al) through first principles calculations and minimum-mode following methods. Several mechanism/pathways for V-H dissociation are observed including multi-atom transitions. The energy barriers for vacancy diffusion are observed to increase with increasing hydrogen saturation. The barrier with a single hydrogen atom is similar to the bulk, but two hydrogen atoms incurs a $\approx$ 20% increase while four hydrogen atoms causes the complex to become immobile. These sessile V-H complexes are proposed as nucleation sites for extended crystal defects.

Momida et al. [56] (2013) investigated the interactions between hydrogen, vacancies and $\Sigma 3$ grain boundaries in iron through first-principles calculations. Hydrogen and V-H complexes are observed to be stabilised by grain boundaries. V-H complexes are shown to have a stronger embrittling effect on the GB than H. Long time scale migration of V-H complexes to grain boundaries is proposed as a mechanism for delayed HE. This highlights the need for a better understanding of how hydrogen impacts vacancy diffusion.

Finally, **Pedersen and Jonsson [67] (2009)** used an OLKMC model to study hydrogen diffusion in aluminium. The paper demonstrates two key advancements in OLKMC: the modelling of large-scale crystal defects, in this case the $\Sigma 5$ twist/tilt grain boundaries, over timescales of tens of µs and the interaction of interstitial hydrogen atoms with the partially disordered GB. The minimum-mode following method is used to find saddle points and the minimum mode is found through the dimer method [37]. The simulations suggest grain boundaries in general impede diffusion and offer explanations for the observed variation in hydrogen diffusivity with grain size. Crucially, only a single H atom was simulated diffusing in the model and therefore H-H interactions have been omitted, this can clearly be improved upon.

2.5 Summary

Off-lattice KMC is a developing method, for which multiple improvements and optimisations have been found since its inception. It seems perfectly suited to studying the complex phenomena of HE as it captures mechanisms at the atom level, can reach KMC timescales and does not require *a priori* knowledge of mechanisms. The need for a from-the-ground-up implementation is clear, to fully integrate new methods and lay solid foundations for tackling the hurdles that are sure to arise when applying OLKMC to the Fe-H system.

Chapter 3

Numerical methods

The following chapter explores in detail the numerical methods employed in this work as well as discussing their implementation. Beginning with Section 3.1 we describe the form of the EAM used in this work and the strategies used to efficiently implement them. Section 3.2 reviews the state-of-the-art methods for computing sums over atom pairs, including the linked cell reordering (LCR) method. The classical L-BFGS optimiser is reviewed in Section 3.3 to understand how it can be implemented in problems were the potential is not known explicitly. Section 3.4 describes the dimer method for saddle-point finding, tying together several advancements in the literature since its inception. Additionally, modifications to the dimer translational stage are discussed to improve the line-search when using the L-BFGS optimiser. Section 3.5 details the system employed for caching saddle-point searches by analysing the local topology around a mechanisms centre and reconstructing future mechanisms from a topological key. Finally, the mean rate method extension to KMC is explored to solve the low-barrier problem and refined to fit the requirements of OLKMC and topological caching.

3.1 EAM potentials

Before we can begin to systematically discover saddle points (required for finding mechanisms and rate constants) we need a description of the potential energy surface for a collection of atoms. Although *ab initio* methods such as DFT exist that could fulfil this role, due to their high computational cost, we must resort to semi-empirical potentials. The embedded atom method, originally developed by Finnis and Sinclair [25], describes the potential energy of a collection

of atoms at positions $\{\boldsymbol{r}^\alpha\}$:

$$U = U_P + U_N \tag{3.1}$$

The potential is separated into a classical pairwise potential U_P as well as U_N the N-body or *embedding* potential. The generalised form of the embedding potential is:

$$U_N = \sum_\beta F^\beta \left(\rho^\beta\right) \tag{3.2}$$

$$\rho^\beta = \sum_{\alpha \neq \beta} \phi^{\alpha\beta} \left(r^{\alpha\beta}\right) \tag{3.3}$$

where we have adopted superscripts with Greek letters to refer to the atom index; $r^{\alpha\beta} = \left\|\boldsymbol{r}^{\alpha\beta}\right\| = \left\|\boldsymbol{r}^\beta - \boldsymbol{r}^\alpha\right\|$ being the atomic separations; F^β the embedding function of the β^{th} atom and $\phi^{\alpha\beta}$ the electron density of the α^{th} atom acting at the β^{th} atom. Strictly speaking, the embedding and potential functions are empirically chosen and do not represent physical electron densities [25].

The generalised pairwise potential takes the familiar form:

$$U_P = \frac{1}{2} \sum_{\substack{\alpha,\beta \\ \alpha \neq \beta}} V^{\alpha\beta} \left(r^{\alpha\beta}\right) \tag{3.4}$$

with $V^{\alpha\beta}$ a symmetric pair-potential term such that $V^{\alpha\beta} = V^{\beta\alpha}$.

The choice of a non-trivial embedding function means the EAM potential is no longer expressible as the sum of a pair potential over all atom pairs; this generalisation is required to prevent the fulfilment of many of the Cauchy relations [22] on the elastic constants and match experimental observations. Choosing $F(\rho) = -A\sqrt{\rho}$ as motivated by the second-moment approximation to tight binding reduces Eq. (3.2) to that originally described by Finnis and Sinclair [25]. Although this is physically motivated, better fitting can be achieved by relaxing this to a more general function, this is the approach taken in the potentials used in this work [75, 1, 54]. If electron density is independent of the sync atom such $\phi^{\alpha\beta} = \phi^\alpha$, the EAM potential reduces to the Daw and Baskes form [17, 18, 26]. Finally, the functional forms of V and ϕ are chosen such that $\forall r^{\alpha\beta} > r_{\text{cut}}, V\left(r^{\alpha\beta}\right) = \phi\left(r^{\alpha\beta}\right) = F(0) = 0$ in this way the potentials are described as *local*.

The forces acting on each atom are required for MD simulations and saddle point finding methods; noting (see Appendix A.1 for a derivation):

$$\frac{\partial r^{\alpha\beta}}{\partial \boldsymbol{r}^{\gamma}} = \left(\delta^{\gamma\beta} - \delta^{\gamma\alpha}\right) \hat{\boldsymbol{r}}^{\alpha\beta} \tag{3.5}$$

the forces from the pairwise potential U_P take the conventional form:

$$\boldsymbol{f}_P^{\gamma} = -\frac{\partial U_P}{\partial \boldsymbol{r}^{\gamma}} = -\frac{1}{2} \sum_{\substack{\alpha,\beta \\ \alpha\neq\beta}} V'^{\alpha\beta} \left(\delta^{\gamma\beta} - \delta^{\gamma\alpha}\right) \hat{\boldsymbol{r}}^{\alpha\beta} = -\sum_{\alpha\neq\gamma} V'^{\alpha\gamma} \hat{\boldsymbol{r}}^{\alpha\gamma} \tag{3.6}$$

where we have dropped the explicit function arguments for compactness such that:

$$V'^{\alpha\beta} = \left.\frac{\partial V^{\alpha\beta}}{\partial r}\right|_{r=r^{\alpha\beta}} \tag{3.7}$$

The forces from the embedding potential take the more complex form:

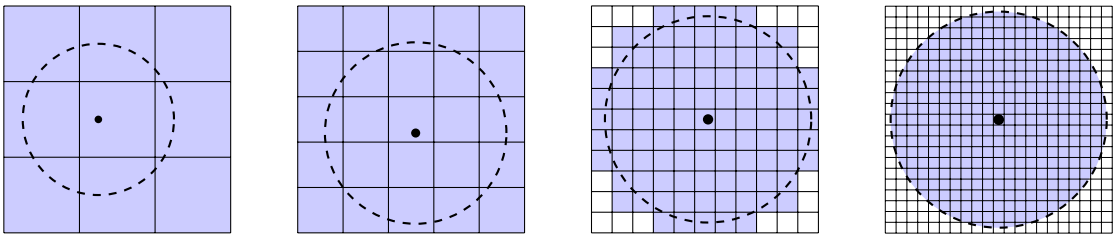
$$\boldsymbol{f}_N^{\gamma} = -\frac{\partial U_N}{\partial \boldsymbol{r}^{\gamma}} = -\sum_{\beta} F'^{\beta} \frac{\partial \rho^{\beta}}{\partial \boldsymbol{r}^{\gamma}} = -\sum_{\substack{\alpha,\beta \\ \alpha\neq\beta}} F'^{\beta} \phi'^{\alpha\beta} \frac{\partial r^{\alpha\beta}}{\partial \boldsymbol{r}^{\gamma}} \tag{3.8}$$

where:

$$F'^{\beta} = \left.\frac{\partial F^{\beta}}{\partial \rho}\right|_{\rho=\rho^{\beta}} \quad \text{and} \quad \phi'^{\alpha\beta} = \left.\frac{\partial \phi^{\alpha\beta}}{\partial r}\right|_{r=r^{\alpha\beta}} \tag{3.9}$$

Applying Eq. (3.5) to Eq. (3.8) we then obtain:

$$\begin{aligned}
\boldsymbol{f}_N^{\gamma} &= -\sum_{\substack{\alpha,\beta \\ \alpha\neq\beta}} F'^{\beta} \phi'^{\alpha\beta} \left(\delta^{\gamma\beta} - \delta^{\gamma\alpha}\right) \hat{\boldsymbol{r}}^{\alpha\beta} \\
&= -\sum_{\alpha\neq\gamma} F'^{\gamma} \phi'^{\alpha\gamma} \hat{\boldsymbol{r}}^{\alpha\gamma} + \sum_{\beta\neq\gamma} F'^{\beta} \phi'^{\gamma\beta} \hat{\boldsymbol{r}}^{\gamma\beta} \\
&= -\sum_{\alpha\neq\gamma} F'^{\gamma} \phi'^{\alpha\gamma} \hat{\boldsymbol{r}}^{\alpha\gamma} - \sum_{\alpha\neq\gamma} F'^{\alpha} \phi'^{\gamma\alpha} \hat{\boldsymbol{r}}^{\alpha\gamma} \\
&= -\sum_{\alpha\neq\gamma} \left(F'^{\gamma} \phi'^{\alpha\gamma} + F'^{\alpha} \phi'^{\gamma\alpha}\right) \hat{\boldsymbol{r}}^{\alpha\gamma}
\end{aligned} \tag{3.10}$$



(a) **(b)** **(c)** **(d)**

Fig. 3.1 Two-dimensional representation of the cells (blue) that contain atoms that could be within r_{cut} of the central atom (black dot). The cut-off radius is marked with a dashed circle. The cell sizes in Fig. 3.1a, Fig. 3.1b, Fig. 3.1c and Fig. 3.1d are r_{cut}, $\frac{r_{\text{cut}}}{2}$, $\frac{r_{\text{cut}}}{5}$ and $\frac{r_{\text{cut}}}{10}$, respectively.

where we have used the antisymmetry of $\boldsymbol{r}^{\alpha\beta}$ in the second line. The forces still act along the line between each atom as in Eq. (3.6) however, an additional summation over all atoms is required for evaluating ρ^{α} for every atom. The first term of Eq. (3.10) describes the force due to the change in the electron density at the γ^{th} atom, while the second term describes the force due to the change in electron density at the α^{th} atom due to the movement of the γ^{th} atom. The total force is thus:

$$\boldsymbol{f}^{\gamma} = \boldsymbol{f}_P^{\gamma} + \boldsymbol{f}_N^{\gamma} = -\sum_{\alpha\neq\gamma} \left(V'^{\alpha\gamma} + F'^{\gamma}\phi'^{\alpha\gamma} + F'^{\alpha}\phi'^{\gamma\alpha}\right)\hat{\boldsymbol{r}}^{\alpha\gamma} \tag{3.11}$$

Due to the complexity of the analytic forms of the fitted functions and the high cost of evaluating them we pre-compute and tabulate the 9 EAM functions. During the simulation we pre-compute and store the natural cubic spline coefficients [2] from the tabulated values, then reconstruct the smooth functions and their differentials appropriately.

Finally, the observation is made that during the calculation of the force on all atoms, the evaluation of ρ^{α} required for F'^{α} in Eq. (3.11) requires a sum over second nearest neighbours. To avoid the requirement of tracking second nearest neighbours or recursion in the implementation, F'^{α} is first computed and stored for all atoms in a single pass then Eq. (3.11) is evaluated for each atom. This affords an approximate 20 times speed-up over the naïve implementation by eliminating repeated calculations of ρ^{α} for shared neighbours.

3.2 Cell methods

A naïve approach to computing the forces on all atoms with Eq. (3.11) requires $\mathcal{O}\left(n^3\right)$ calculations, this can be vastly accelerated by exploiting the local nature of the potential. If the *neighbours* within the cut-off radius r_{cut} of each atom are known then, the sum over α in

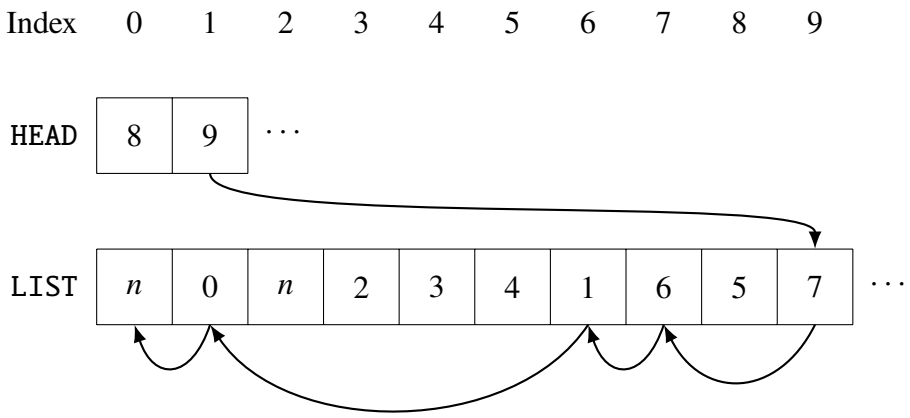


Fig. 3.2 A graphical representation of the LCL data structure. For a simulation with M cells and n atoms `HEAD` is a length M array and `LIST` is a length n array. Here cell C_2, corresponding to `HEAD[2]`, contains atoms indexed $7, 6, 1$ and 0. Following the C-style 0-based array indexing; n is the end of array sentinel, indicating no more atoms in a cell.

Eq. (3.11) only needs run over neighbours of γ and the sum to calculate ρ^α needs only run over the neighbours of α. Thus, assuming an approximately uniform atom density such that each atom has n_p neighbours, the number of calculations is reduced to $\mathcal{O}\left(n \cdot n_p^2\right)$.

The cost of finding the neighbours of all the atoms, when taking the simple approach of comparing all the atomic separations, scales as $\mathcal{O}\left(n^2\right)$. This is undesirable as the goal is to match the complexity of evaluating the force. Fortunately, by binning the atoms into small cells, with edge length $L = r_{\text{cut}}$, we only need to investigate the atoms in adjacent cells. This reduces the complexity of finding the neighbours of all atoms to $\mathcal{O}\left(n \cdot 3^3 n_c\right)$ with n_c the average number of atoms per cell.

The cell size can be smaller than r_{cut} but this requires iterating over more cells to find all neighbours. Figure 3.1 highlights the relationship between cell size and cells required to search. With a uniform atom density, the number of cells to iterate over scales as $\mathcal{O}\left(\frac{r_{\text{cut}}}{L}^3\right)$. For $L = r_{\text{cut}}$ we expect only 16% of tested atoms to fall within the cut-off radius. For a ratio of $2L = r_{\text{cut}}$ this improves to 27% but requires 4.5× more cells to be inspected, this process could be continued until each cell holds only one atom but the cost of iterating over all the cell becomes prohibitive. Based on the empirical results of Welling and Germano [83], we choose the simplest ratio $L = r_{\text{cut}}$ which performed well in all their testing. Finally, we consider using a projection sorting method [33] between adjacent cells, which has been shown to increase the probability of a tested atom being a neighbour to 59%. However, due to the complexity and overhead required, we choose not to use this method.

3.2.1 Linked cell list

Algorithm 3.1 Linked cell list construction.

procedure MakeCellList(HEAD, LIST)
Assuming a periodic simulation box with side lengths 0 to L_i in the i^{th} axis.
HEAD$[0, 1, \ldots, M-1] \leftarrow n$ ▷ Fill with sentinels
for $\alpha \leftarrow 0, 1, \ldots, n-1$ **do**
$i \leftarrow \texttt{int}(r_x^\alpha \cdot m_x / L_x)$ ▷ Rounds toward zero
$j \leftarrow \texttt{int}(r_y^\alpha \cdot m_y / L_y)$
$k \leftarrow \texttt{int}(r_z^\alpha \cdot m_z / L_z)$
$\lambda \leftarrow \Lambda(i, j, k)$ ▷ Section 3.2.2; $\lambda \in \{0, 1, \ldots, M-1\}$
LIST$[\alpha] \leftarrow$ HEAD$[\Lambda]$ ▷ Possibly another atom in C_λ
HEAD$[\Lambda] \leftarrow \alpha$

An efficient data structure for storing the atoms in each cell is a linked cell list (LCL) [3, p. 151] consisting of two arrays HEAD and LIST of length $M = m_x \cdot m_y \cdot m_z$ and n respectively, with m_x, m_y, m_z the number of cells along the x, y, z axis, respectively. Each element of array HEAD represents a cell C_i with $i \in \{0, 1, \ldots, M-1\}$ and contains the index α of an atom in C_i. Additionally, LIST$[\alpha]$ holds the index of the next atom in C_i. This structure is represented graphically in Fig. 3.2. The LCL can be constructed rapidly in $\mathcal{O}(n)$ time using Algorithm 3.1 and the atoms in a cell can be retrieved using Algorithm 3.2.

If the use of $\mathcal{O}(M)$ memory is too high then the array HEAD could be replaced with a hash table. This would reduce the memory required to $\mathcal{O}(n)$ at the cost of evaluating a hash function at each random access. Because there should be multiple atoms per cell we anticipate $M < n$ therefore, $\mathcal{O}(M)$ arrays are preferred.

3.2.2 Cell indexing

A key requirement of the LCL algorithms described is a method for mapping the three-tuple coordinate to an integer cell index: $\mathbb{R}^3 \mapsto \mathbb{N}^3 \mapsto \mathbb{N}$. The first transformation can be achieved trivially through a `float` $\rightarrow$ `int` truncating cast. The second transformation, which we shall name $\Lambda\colon \mathbb{N}^3 \rightarrow \mathbb{N}$, must be injective and ideally surjective onto the set $\{0, 1, \ldots, M-1\}$ such that we minimise the memory required for HEAD. The simplest choice for Λ is the trivial bijection used for array indexing in many computing languages:

$$\Lambda(i, j, k) = i + m_x \cdot j + m_x m_y \cdot k \tag{3.12}$$

Algorithm 3.2 Find indexes of all atoms in a cell.

function FINDINCELL(HEAD, LIST, i, j, k)
    ATMS $\leftarrow$ []
    $\alpha \leftarrow$ HEAD[$\Lambda(i, j, k)$] ▷ Section 3.2.2.
    loop
        if $\alpha = n$ **then**
            return ATMS
        else
            Append atom α to ATMS
        $\alpha \leftarrow$ LIST[α]

which is very fast to evaluate but has poor *locality* (i.e points close in 3-space mapping to points close in 1-space; Eq. (3.12) preserves locality only in i). The locality preserving property of Λ is advantageous as we periodically sort the atoms by their Λ ordering—as Λ maps to natural numbers we can use an integer sorting algorithm to achieve $\mathcal{O}(n)$ complexity, see Section A.2—this enhancement is named the linked cell reordering [53] (LCR) method. LCR accelerates performance by reducing cache misses in three ways:

1. When constructing the LCL memory read/writes to HEAD will be sequential (with gaps).
2. When accessing atoms in the same cell they will be continuous in memory.
3. When accessing neighbours in adjacent cells, if Λ is strongly locality preserving, they are more likely to be close in memory to the initial cell atoms.

Better choices for Λ could therefore be the space filling, Hilbert [40] or Z-order/Morton [60] curves. The Hilbert curve in particular has been suggested [4, 83] as a good candidate for its strong locality preserving properties [58] and ability to sort into Hilbert-order without computing Hilbert-indices [41]. Using these curves for Λ for both cell indexing and ordering would be beneficial (albeit take longer per Λ evaluation) however, the recursive definitions of both curves puts a power of two constraint on the number of cells in each axis. Using a space filling curve for the atom ordering and Eq. (3.12) for the indexing would reduce cache misses via mechanism 3 above but weaken 1 and 2 to 'approximately' sequential and 'approximately' continuous, respectively. Preliminary testing of using Hilbert indexing for Λ showed the overhead of computing the Hilbert indices and the relaxing of points 1 and 2 outweighed optimisation 3 compared to using Eq. (3.12).

3.2.3 Boundary conditions

We must introduce cyclic boundary conditions into the simulation cell. This can be done on a per-norm calculation level but requires the introduction of conditionals or modulo operations into each norm calculation. This is expensive (conditionals preventing vectorisation) and wasteful as most atom pairs need no boundary corrections. Alternatively, corrections can be pre-computed at the cell level with correction vectors $\boldsymbol{q}_{ij}$ for each cell pair (C_i, C_j) such that:

$$r^{\alpha\beta} = \left\| \boldsymbol{r}^{\beta} - \boldsymbol{r}^{\alpha} - \boldsymbol{q}_{ij} \right\| \tag{3.13}$$

with $\boldsymbol{r}^{\beta}$, $\boldsymbol{r}^{\alpha}$ in cells C_i, C_j respectively. This only require three extra floating-point additions per calculation but requires an additional $\mathcal{O}(M^2)$ memory to store the correction vectors. Finally, as described by Rapaport [77, p. 18], a set of *ghost atoms* can be introduced outside the simulation cell by copying the atoms within r_c of a face to the outside of the opposite face. The forces on the atoms inside the cell can then be computed without application of boundary conditions thus, eliminating all branching. This requires $\mathcal{O}(n^{\frac{2}{3}})$ extra work and space to find/copy and store the ghost atoms but completely removes any additional per-norm overhead. As n grows this cost will become negligible (compared to the force calculation itself) therefore, we choose this method for including cyclic boundary conditions.

3.3 L-BFGS method

As a subroutine for later algorithms we shall require a fast and robust general optimiser. The limited-memory formulation of the Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm is such an optimiser for solving the problem:

$$\min U(\boldsymbol{x}) \quad \text{with} \quad U\colon \mathbb{R}^N \to \mathbb{R}$$

Where the function U has associated gradient $\boldsymbol{g} = \boldsymbol{\nabla} U$ and Hessian $(\boldsymbol{H})_{ij} = \frac{\delta^2 U}{\delta x_i \delta x_j}$. It has been shown to produce superlinear convergence for similar problems [45]. We detail the algorithm here in order fully understand how it can be best applied to our problem.

The Newton step towards a local minima (assuming $\boldsymbol{H}$ positive definite) is then [63, p. 22]:

$$\boldsymbol{p} = -\overline{\boldsymbol{H}}\boldsymbol{g} \tag{3.14}$$

where we have used the shorthand for the inverse Hessian $\overline{\boldsymbol{H}} = \boldsymbol{H}^{-1}$. This would require the explicit calculation and inversion of the Hessian at each iteration, an expensive set of operations. The BFGS method approximates the inverse Hessian with knowledge of the curvature obtained as the algorithm explores the function landscape. Introducing the following notation for the k^{th} iteration:

$$\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \quad \text{and} \quad \boldsymbol{y}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k \tag{3.15}$$

$$\rho_k = \left(\boldsymbol{y}_k^\mathsf{T} \boldsymbol{s}_k\right)^{-1} \tag{3.16}$$

The BFGS update formula for the $k+1$ approximation to the inverse Hessian is then [49]:

$$\overline{\boldsymbol{H}}_{k+1} = (\boldsymbol{I} - \rho_k \boldsymbol{s}_k \boldsymbol{y}_k^\mathsf{T}) \overline{\boldsymbol{H}}_k (\boldsymbol{I} - \rho_k \boldsymbol{y}_k \boldsymbol{s}_k^\mathsf{T}) + \rho_k \boldsymbol{s}_k \boldsymbol{s}_k^\mathsf{T} \tag{3.17}$$

and the update formula for the position is:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \overline{\boldsymbol{H}}_k \boldsymbol{g}_k \tag{3.18}$$

where α_k is a step length chosen to satisfy the Wolfe conditions and guarantee convergence. For details on the Wolfe conditions used see Appendix A.3. Such a step length can be found via an appropriate line search method (such as that by Moré and Thuente [59]) however, $\alpha_k = 1$ should always be tested for acceptance first as in the quadratic region this leads to superlinear convergence. The BFGS update step preserves the positive definiteness of $\overline{\boldsymbol{H}}$ provided $\boldsymbol{y}_k^\mathsf{T} \boldsymbol{s}_k > 0$ [62], which is satisfied as a corollary of the Wolfie curvature condition Eq. (A.5).

BFGS still requires expensive matrix multiplications and more importantly the storage of the entire inverse Hessian, requiring $\mathcal{O}(N^2)$ memory. The L-BFGS method overcomes these limitations by implicitly storing a modified $\overline{\boldsymbol{H}}$, instead caching the previous m values of $\boldsymbol{s}, \boldsymbol{y}, \rho$ and computing the product in Eq. (3.14) without explicitly forming the inverse Hessian. The

modified inverse Hessian approximation is [49]:

$$\begin{aligned}\overline{\boldsymbol{H}}_k = &\left(\boldsymbol{V}_{k-1}^{\mathsf{T}} \dots \boldsymbol{V}_{k-m}^{\mathsf{T}}\right) \overline{\boldsymbol{H}}_k^0 \left(\boldsymbol{V}_{k-m} \dots \boldsymbol{V}_{k-1}\right) \\ &+ \rho_{k-m} \left(\boldsymbol{V}_{k-1}^{\mathsf{T}} \dots \boldsymbol{V}_{k-m+1}^{\mathsf{T}}\right) \boldsymbol{s}_{k-m} \boldsymbol{s}_{k-m}^{\mathsf{T}} \left(\boldsymbol{V}_{k-m+1} \dots \boldsymbol{V}_{k-1}\right) \\ &+ \rho_{k-m+1} \left(\boldsymbol{V}_{k-1}^{\mathsf{T}} \dots \boldsymbol{V}_{k-m+2}^{\mathsf{T}}\right) \boldsymbol{s}_{k-m+1} \boldsymbol{s}_{k-m+1}^{\mathsf{T}} \left(\boldsymbol{V}_{k-m+2} \dots \boldsymbol{V}_{k-1}\right) \\ &\vdots \\ &+ \rho_{k-1} \boldsymbol{s}_{k-1} \boldsymbol{s}_{k-1}^{\mathsf{T}}\end{aligned} \tag{3.19}$$

where:

$$\boldsymbol{V}_k = \boldsymbol{I} - \rho_k \boldsymbol{y}_k \boldsymbol{s}_k^{\mathsf{T}} \tag{3.20}$$

which then makes Eq. (3.19) the application of Eq. (3.17) to the m most recent iterations starting from some estimate of the inverse Hessian, $\overline{\boldsymbol{H}}_k^0$. There is an efficient algorithm, due to Nocedal [62], for computing the product $\overline{\boldsymbol{H}}\boldsymbol{g}$ in Eq. (3.14) using the inverse Hessian from Eq. (3.19) called the *two-loop recursion* detailed in Algorithm 3.3.

Algorithm 3.3 L-BFGS two-loop recursion [62].

function TwoLoop(k, m)
  $\boldsymbol{q} \leftarrow \boldsymbol{g}_k$
  for $i \leftarrow k-1, k-2, \dots, k-m$ **do**
    Store $a_i \leftarrow \rho_i \boldsymbol{s}_i^{\mathsf{T}} \boldsymbol{q}$
    $\boldsymbol{q} \leftarrow \boldsymbol{q} - a_i \boldsymbol{y}_i$
  $\boldsymbol{q} \leftarrow \overline{\boldsymbol{H}}_k^0 \boldsymbol{q}$
  for $i \leftarrow k-m, k-m+1, \dots, k-1$ **do**
    $b \leftarrow \rho_i \boldsymbol{y}_i^{\mathsf{T}} \boldsymbol{q}$
    $\boldsymbol{q} \leftarrow \boldsymbol{q} + (a_i - b)\, \boldsymbol{s}_i$
  return $\boldsymbol{q}$ ▷ Where $\boldsymbol{q} = \overline{\boldsymbol{H}}_k \boldsymbol{g}_k$.

Choosing $\overline{\boldsymbol{H}}_k^0 = \gamma_k \boldsymbol{I}$ in Algorithm 3.3 makes for an efficient implementation; the scaling factor γ_k used is [49]:

$$\gamma_k = \left(\rho_{k-1} \boldsymbol{y}_{k-1}^{\mathsf{T}} \boldsymbol{y}_{k-1}\right)^{-1} \tag{3.21}$$

which helps to ensure the step length $a_k = 1$ satisfies the Wolfe conditions [49]. If there is no initial information of the Hessian or inverse Hessian then $\overline{\boldsymbol{H}}_0^0 = \boldsymbol{I}$ is typically used. The full L-BFGS method is specifed in Algorithm 3.4.

Algorithm 3.4 L-BFGS method.

function L_BFGS($\boldsymbol{x}_0$, m)
  $k \leftarrow 0$
  repeat
    $\boldsymbol{p}_k \leftarrow -\overline{\boldsymbol{H}}_k \boldsymbol{g}_k$ ▷ Algorithm 3.3.
    $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$ ▷ Where α_k satisfies Wolfe conditions.
    if $k > m$ **then**
      Discard $\boldsymbol{s}_{k-m}$ and $\boldsymbol{y}_{k-m}$
    Store $\boldsymbol{s}_k \leftarrow \alpha_k \boldsymbol{p}_k$
    Store $\boldsymbol{y}_k \leftarrow \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$
    $k \leftarrow k + 1$
  until converged
  return $\boldsymbol{x}_k$

3.4 Dimer method

The original dimer method was developed by Henkelman and Jónsson [37] as a method for efficiently finding local saddle points in a function of many variables, such as the potential energy:

$$U : \mathbb{R}^{3n} \to \mathbb{R}$$

The dimer method is a minimum-mode following method, originally described by Cerjan and Miller [14], that climbs the potential energy surface (PES) from a local minima to an adjacent saddle point. The dimer method is superior to earlier methods as it avoids computing the full Hessian matrix. This avoids expensive matrix operation that scale as $\mathcal{O}(n^3)$ as well as only requiring the first derivatives of the potential energy. We shall follow the formulation of the dimer method presented by Kästner and Sherwood [45] which contains several optimisations over the original method.

The dimer method manipulates two images of the system $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ separated by a small distance $2\Delta R$ along the unit vector $\hat{\boldsymbol{N}}$ in $3n$ dimensional space, with n the number of atoms in

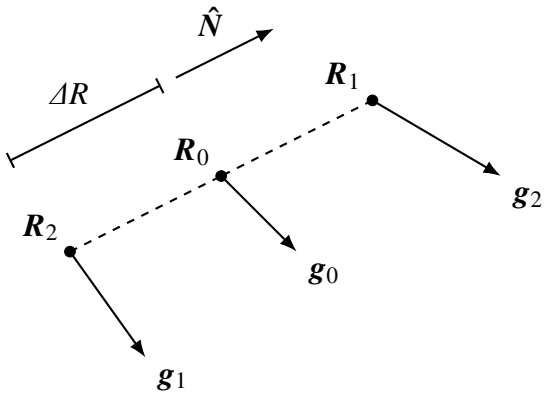


Fig. 3.3 Definition of the dimer images, $\boldsymbol{R}_x$, in $3n$ dimensional space and the corresponding potential gradients at each image, $\boldsymbol{g}_x$.

the system. This pair of images are referred to as “the dimer" and represented in Fig. 3.3. At $\boldsymbol{R}_x$ the gradient of the potential energy is denoted $\boldsymbol{g}_x$ and the potential energy E_x. The dimer midpoint is denoted $\boldsymbol{R}_0$ and related as:

$$\boldsymbol{R}_1 = \boldsymbol{R}_0 + \Delta R\hat{\boldsymbol{N}} \tag{3.22}$$

$$\boldsymbol{R}_2 = \boldsymbol{R}_0 - \Delta R\hat{\boldsymbol{N}} \tag{3.23}$$

The curvature, C, of the potential along the dimer is calculated with the quadratic finite difference formula:

$$C_N = \hat{\boldsymbol{N}}^{\mathsf{T}}\boldsymbol{H}\hat{\boldsymbol{N}} \approx \frac{(\boldsymbol{g}_1 - \boldsymbol{g}_2)^{\mathsf{T}}\,\hat{\boldsymbol{N}}}{2\Delta R} \approx \frac{E_1 + E_2 - 2E_0}{\Delta R^2} \approx \frac{E - 2E_0}{\Delta R^2} \tag{3.24}$$

where $E = E_1 + E_2$ is the dimer energy. The gradient $\boldsymbol{g}_2$ can be linearly interpolated:

$$\boldsymbol{g}_2 \approx 2\boldsymbol{g}_0 - \boldsymbol{g}_1 \tag{3.25}$$

this allows the reuse of $\boldsymbol{g}_0$ as the dimer rotates, saving a gradient evaluation per rotation, but the accuracy of the curvature approximation goes from $\mathcal{O}(\Delta R^2)$ to $\mathcal{O}(\Delta R)$:

$$C_N \approx \frac{(\boldsymbol{g}_1 - \boldsymbol{g}_0)^{\mathsf{T}}\,\hat{\boldsymbol{N}}}{\Delta R} \tag{3.26}$$

The saddle point search has two steps: rotation of the dimer to align it with the minimum curvature mode of the Hessian and translation of the dimer along the PES in towards a local

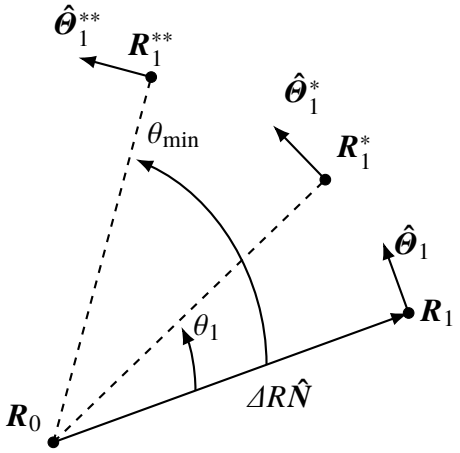


Fig. 3.4 Definition of the dimer images during a rotation in $3n$ dimensional space, all vectors are in the plane of the paper/rotation. The image at $\boldsymbol{R}_2$ has been omitted from this figure for clarity.

saddle point. The former is described in Section 3.4.1 and the latter in Section 3.4.2, a single iteration of the dimer method is an application of both of these steps.

3.4.1 Rotation of the dimer

Rotation plane

As the midpoint of the dimer remains fixed during a rotation we see from Eq. (3.24) that minimising the dimer energy is equivalent to minimising the curvature along the dimer. This assumes a sufficiently small ΔR such that the potential varies quadratically, therefore Eq. (3.24) holds. Let the plane of rotation be spanned by $\hat{\boldsymbol{N}}$ and a unit vector $\hat{\boldsymbol{\theta}}$. The steepest ascent direction for rotation of the dimer maximises $\frac{\partial E}{\partial \theta}\big|_{\theta=0}$ hence, expanding the partial derivative:

$$\frac{\partial E}{\partial \theta} = \frac{\partial E_1}{\partial \theta} + \frac{\partial E_2}{\partial \theta} = \boldsymbol{g}_1^{\mathsf{T}} \frac{\partial \boldsymbol{R}_1}{\partial \theta} + \boldsymbol{g}_2^{\mathsf{T}} \frac{\partial \boldsymbol{R}_2}{\partial \theta} \tag{3.27}$$

and applying:

$$\boldsymbol{R}_1 = \boldsymbol{R}_0 + \Delta R \left(\hat{\boldsymbol{N}} \cos\theta + \hat{\boldsymbol{\theta}} \sin\theta \right) \tag{3.28}$$

$$\boldsymbol{R}_2 = \boldsymbol{R}_0 - \Delta R \left(\hat{\boldsymbol{N}} \cos\theta + \hat{\boldsymbol{\theta}} \sin\theta \right) \tag{3.29}$$

we see:

$$\frac{\partial E}{\partial \theta} = \Delta R \left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right)^{\mathsf{T}} \left(-\hat{\boldsymbol{N}} \sin\theta + \hat{\boldsymbol{\theta}} \cos\theta\right) \tag{3.30}$$

$$\left.\frac{\partial E}{\partial \theta}\right|_{\theta=0} = \Delta R \left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right)^{\mathsf{T}} \hat{\boldsymbol{\theta}} \tag{3.31}$$

hence the optimal direction for steepest descent is parallel to $-\left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right)$ and the rotation plane of steepest descent is spanned by $\left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right)$ and $\hat{\boldsymbol{N}}$. To preserve orthogonality to $\hat{\boldsymbol{N}}$ we choose $\hat{\boldsymbol{\theta}}$ parallel to the applied rotational force:

$$\begin{aligned} \boldsymbol{F}_r &= -\left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right) + \left[\left(\boldsymbol{g}_1 - \boldsymbol{g}_2\right)^{\mathsf{T}} \hat{\boldsymbol{N}}\right] \hat{\boldsymbol{N}} \\ &\approx -2\left(\boldsymbol{g}_1 - \boldsymbol{g}_0\right) + 2\left[\left(\boldsymbol{g}_1 - \boldsymbol{g}_0\right)^{\mathsf{T}} \hat{\boldsymbol{N}}\right] \hat{\boldsymbol{N}} \end{aligned} \tag{3.32}$$

where we have used the linear interpolation Eq. (3.25). Finally, normalising $\hat{\boldsymbol{\theta}} = \boldsymbol{F}_r / \|\boldsymbol{F}_r\|$. This specifies the plane of rotation of steepest descent with a pair orthonormal vectors and is visualised in Fig. 3.4.

As an enhancement to the rotation algorithm, instead of choosing $\hat{\boldsymbol{\theta}}$ in the plane of steepest descent we use the L-BFGS method to select the plane of rotation; see Section 3.3 for a full description of the L-BFGS method. The dimer axis, $\hat{\boldsymbol{N}}$, is used as the coordinate input to L-BFGS whilst the associated gradient is $\boldsymbol{g}_1 - \boldsymbol{g}_2 \propto \boldsymbol{g}_1 - \boldsymbol{g}_0$. The L-BFGS descent direction is then orthonormalised to the dimer axis to find $\hat{\boldsymbol{\theta}}$. The L-BFGS history is reset after each dimer translation and we start from the unit Hessian.

Rotation angle

To determine the value of $\theta = \theta_{\min}$ that minimises the curvature in the plane of rotation we expand the curvature along $\boldsymbol{x} = \hat{\boldsymbol{N}} \cos\theta + \hat{\boldsymbol{\theta}} \sin\theta$, a unit vector in the plane of rotation:

$$\begin{aligned} C_x &= \boldsymbol{x}^{\mathsf{T}} \boldsymbol{H} \boldsymbol{x} \\ &= \hat{\boldsymbol{N}}^{\mathsf{T}} \boldsymbol{H} \hat{\boldsymbol{N}} \cos^2\theta + (\hat{\boldsymbol{N}}^{\mathsf{T}} \boldsymbol{H} \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{N}}^{\mathsf{T}} \boldsymbol{H} \hat{\boldsymbol{\theta}}) \sin\theta \cos\theta + \hat{\boldsymbol{\theta}}^{\mathsf{T}} \boldsymbol{H} \hat{\boldsymbol{\theta}} \sin^2\theta \\ &= \cos^2\theta \sum_i \lambda_i \left(\hat{\boldsymbol{N}}^{\mathsf{T}} \boldsymbol{u}_i\right)^2 + 2 \sin\theta \cos\theta \sum_i \lambda_i \left(\hat{\boldsymbol{N}}^{\mathsf{T}} \boldsymbol{u}_i\right)\left(\hat{\boldsymbol{\theta}}^{\mathsf{T}} \boldsymbol{u}_i\right) + \sin^2\theta \sum_i \lambda_i \left(\hat{\boldsymbol{\theta}}^{\mathsf{T}} \boldsymbol{u}_i\right)^2 \\ &= \frac{a_0}{2} + a_1 \cos(2\theta) + b_1 \sin(2\theta) \end{aligned} \tag{3.33}$$

therefore:

$$\frac{\partial C_x}{\partial \theta} = -2a_1 \sin(2\theta) + 2b_1 \cos(2\theta) \tag{3.34}$$

where we have expanded the Hessian in terms of its eigenvectors, $\boldsymbol{u}_i$, and corresponding eigenvalues, λ_i, through the eigendecomposition relation for real symmetric matrices [11, p. 80]:

$$\boldsymbol{H} = \sum_i \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^{\mathsf{T}} \tag{3.35}$$

θ_{min} then satisfies:

$$\left.\frac{\partial C_x}{\partial \theta}\right|_{\theta=\theta_{\text{min}}} = 0 \tag{3.36}$$

therefore:

$$\theta_{\text{min}} = \frac{1}{2} \arctan\left(\frac{b_1}{a_1}\right) \tag{3.37}$$

This may be a minimum or maximum of the curvature and can be checked via substitution back into Eq. (3.33), if it is a maximum (i.e the curvature increases) then $\frac{\pi}{2}$ is added to θ_{min} to bring us to a minimum. This $\frac{\pi}{2}$ offset is due to the π periodicity of Eq. (3.33).

To approximate a_1 and b_1 we differentiate Eq. (3.24) with respect to θ and substitute in Eq. (3.30):

$$\frac{\partial C_x}{\partial \theta} \approx \frac{2\left(\boldsymbol{g}_1 - \boldsymbol{g}_0\right)^{\mathsf{T}} \left(-\hat{\boldsymbol{N}} \sin\theta + \hat{\boldsymbol{\theta}} \cos\theta\right)}{\Delta R} \tag{3.38}$$

where we have used the linear interpolation Eq. (3.25). Evaluating Eq. (3.38) and Eq. (3.34) at $\theta = 0$ we obtain:

$$b_1 = \frac{1}{2} \left.\frac{\partial C_x}{\partial \theta}\right|_{\theta=0} \approx \frac{\left(\boldsymbol{g}_1 - \boldsymbol{g}_0\right)^{\mathsf{T}} \hat{\boldsymbol{\theta}}}{\Delta R} \tag{3.39}$$

Finally after rotating the dimer through a trial angle θ_1 and substituting into Eq. (3.33):

$$a_1 = \frac{C_x|_{\theta=0} - C_x|_{\theta=\theta_1} + b_1 \sin(2\theta_1)}{1 - \cos(2\theta_1)} \tag{3.40}$$

where $C_x|_{\theta=0} = C_N$ and $C_x|_{\theta=\theta_1}$ is computed using the new dimer axis, $\hat{\boldsymbol{N}}^*$, and gradient at the new dimer end point, $\boldsymbol{g}_1^*$, see Fig. 3.4 for definitions.

An estimate for $\theta_{\min}$ can be obtained without a trial rotation of the dimer (avoiding a gradient calculation at $\boldsymbol{g}_1^*$) by approximating the second derivative of the curvature along the dimer axis [39], this estimate should be used as the trial angle:

$$\theta_1 = -\frac{1}{2}\arctan\left(\frac{b_1}{|C_x|_{\theta=0}|}\right) \tag{3.41}$$

Should either $|\theta_{\min}|$ or $|\theta_1|$ be less than a specified θ_{tol} then the rotation is considered converged and no further rotation is carried out.

After the dimer has been rotated by $\theta_{\min}$, the gradient of the potential at the new end point $\boldsymbol{g}_1^{**}$ must be calculated as it becomes $\boldsymbol{g}_1$ in the next iteration. However, $\boldsymbol{R}_0$, $\boldsymbol{R}_1$ and $\boldsymbol{R}_1^*$ are coplanar therefore, we can linearly interpolate between them to obtain $\boldsymbol{g}_1^{**}$ with one less gradient calculation [45]:

$$\boldsymbol{g}_1^{**} = \frac{\sin(\theta_1 - \theta_{\min})}{\sin\theta_1}\boldsymbol{g}_1 + \frac{\sin\theta_{\min}}{\sin\theta_1}\boldsymbol{g}_1^* + \left(1 - \cos\theta_{\min} - \sin(\theta_{\min})\tan\left(\frac{\theta_1}{2}\right)\right)\boldsymbol{g}_0 \tag{3.42}$$

The dimer rotation procedure is summarised in Algorithm 3.5. We set a limit on the number of dimer rotations per iteration ($I_{\max}^R = 15$ unless otherwise specified), after which the rotation is considered sufficiently converged. This allows us to use a very small θ_{tol} that is only fully resolved during the latter part of the dimer translation whilst converging onto the saddle point. During the initial translation exact rotational convergence is not required, here small errors in the translation force's direction are handled by the L-BFGS optimiser.

3.4.2 Translation of the dimer

After rotation of the dimer to align with the minimum mode the modified translational force acting on the dimer is defined as:

$$\boldsymbol{F}_t = -\boldsymbol{g}_0 + 2\hat{\boldsymbol{N}}\boldsymbol{g}_0^{\mathsf{T}}\hat{\boldsymbol{N}} \tag{3.43}$$

This is the net translational force acting on the dimer with the component parallel to the dimer axis (ideally aligned with the minimum mode) reversed. Translating along this force maximises the dimer energy along the minimum mode and minimises the energy along all other modes. Therefore, assuming appropriate rotational convergence at each step, translating along $\boldsymbol{F}_t$ converges to a local saddle point [45].

Algorithm 3.5 Rotation procedure to align the dimer axis with the minimum mode.

procedure RotateDimer($\boldsymbol{R}_0, \boldsymbol{g}_0, \hat{\boldsymbol{N}}$) ▷ Orients $\hat{\boldsymbol{N}}$ along the minimum mode.
  $\boldsymbol{g}_1 \leftarrow \boldsymbol{\nabla} U(\boldsymbol{R}_0 + \Delta R \hat{\boldsymbol{N}})$
  for $k \leftarrow 1, 2, \ldots, I^R_{\max}$ **do**
    Calculate $\hat{\boldsymbol{\theta}}(\boldsymbol{g}_0, \boldsymbol{g}_1, \hat{\boldsymbol{N}})$ using L-BFGS ▷ Section 3.3.
    $b_1 \leftarrow b_1(\boldsymbol{g}_0, \boldsymbol{g}_1, \hat{\boldsymbol{\theta}})$ ▷ Equation (3.39).
    $C_x|_{\theta=0} \leftarrow C_N(\boldsymbol{g}_0, \boldsymbol{g}_1, \hat{\boldsymbol{N}})$ ▷ Equation (3.26).
    $\theta_1 \leftarrow \theta_1(b_1, C_x|_{\theta=0})$ ▷ Equation (3.41).
    if $|\theta_1| < \theta_{\text{tol}}$ **then**
      return
    else
      $\hat{\boldsymbol{N}}^* \leftarrow \hat{\boldsymbol{N}} \cos\theta_1 + \hat{\boldsymbol{\theta}} \sin\theta_1$ ▷ Trial rotation.
      $\boldsymbol{g}_1^* \leftarrow \boldsymbol{\nabla} U(\boldsymbol{R}_0 + \Delta R \hat{\boldsymbol{N}}^*)$
      $C_x|_{\theta=\theta_1} \leftarrow C_N(\boldsymbol{g}_0, \boldsymbol{g}_1^*, \hat{\boldsymbol{N}}^*)$ ▷ Equation (3.26).
      $a_1 \leftarrow a_1(C_x|_{\theta=0}, C_x|_{\theta=\theta_1}, b_1, \theta_1)$ ▷ Equation (3.40).
      $\theta_{\min} \leftarrow \theta_{\min}(a_1, b_1)$ ▷ Equation (3.37).
      if $C_x|_{\theta=\theta_{\min}} > C_x|_{\theta=0}$ **then** ▷ Equation (3.33).
        $\theta_{\min} \leftarrow \theta_{\min} + \frac{\pi}{2}$
      $\hat{\boldsymbol{N}} \leftarrow \hat{\boldsymbol{N}} \cos\theta_{\min} + \hat{\boldsymbol{\theta}} \sin\theta_{\min}$ ▷ Optimal rotation.
      $\mathbf{g}_1 \leftarrow \mathbf{g}_1^{**}(\theta_1, \theta_{\min}, \boldsymbol{g}_0, \boldsymbol{g}_1, \boldsymbol{g}_1^*)$ ▷ Equation (3.42).
      if $|\theta_{\min}| < \theta_{\text{tol}}$ **then**
        return

Diverging from the formulation of Kästner and Sherwood [45] we treat the dimer translation differently in the *convex regions* of the potential energy, where all modes have positive curvature and *nonconvex regions*, where at least one mode has negative curvature. This is determined from the sign of the minimum curvature mode, as computed during the dimer rotation. The local minima reside in convex regions whilst the saddle points in nonconvex regions.

In the nonconvex regions, similar to the dimer rotation, we use the L-BFGS algorithm for determining the translation direction and step size. No line search was used for determining the step size in the L-FBGS algorithm as no *effective* potential of $\boldsymbol{F}_t$ can be formed (necessary to evaluate the Wolfe sufficient decrease condition Eq. (A.4)), instead the $\alpha_k = 1$ parameter is used with a scaled unit Hessian and a maximum step size, s_{trust}. This additionally saves energy/gradient calculations that would be required during a line search. The dimer is considered

converged when $\|\boldsymbol{F}_t\|$ is below F_{tol}. The maximum step size is scaled according to the success of the previous steps; the projection of the effective gradient on the search direction is calculated after a step:

$$P = -\boldsymbol{F}_t^{\mathsf{T}} \boldsymbol{p} \tag{3.44}$$

where $\boldsymbol{p}$ is the approximate Newton step, computed using the L-BFGS method. An ideal step length would have $P = 0$, while an oversized step results in $P > 0$ and an undersized step $P < 0$. Consequently, we increase the the maximum step size by some multiplayer m when $P < -P_{\text{tol}}$ and divide by m when $P > P_{\text{tol}}$. Typically, we choose $m = 2$ and bound s_{trust} such that $s_{\min} < s_{\text{trust}} < s_{\max}$. An attempt was made to introduce a line search that satisfied the Wolfie curvature condition Eq. (A.5) (which does not require the effective potential) however, this required alignment of the dimer axis at every trial step-length which requires additional gradient calculations and did not significantly accelerate convergence.

It is not suitable to use the L-BFGS method without a line search in the convex regions of the potential as, while the dimer climbs out of the convex minima, the product $\boldsymbol{s}_k^{\mathsf{T}} \boldsymbol{y}_k$ frequently becomes negative. This no longer guarantees positive definiteness of the inverse Hessian which can result in the L-BFGS method producing non-descent directions; this causes the dimer to converge back to the starting minima for most trials. The connection between $\boldsymbol{s}_k^{\mathsf{T}} \boldsymbol{y}_k$ and minima/saddle-points of the potential U becomes clearer if we define $\boldsymbol{H}^{\text{eff}}$ the Hessian of the effective potential of Eq. (3.43), denoted U^{eff}. Then for sufficiently small steps:

$$\begin{aligned}
\boldsymbol{s}_k^{\mathsf{T}} \boldsymbol{y}_k &= (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)^{\mathsf{T}} \left(\boldsymbol{\nabla} U_{k+1}^{\text{eff}} - \boldsymbol{\nabla} U_k^{\text{eff}} \right) \\
&\approx (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)^{\mathsf{T}} \left(\left(\boldsymbol{\nabla} U_k^{\text{eff}} + \boldsymbol{H}_k^{\text{eff}} (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) \right) - \boldsymbol{\nabla} U_k^{\text{eff}} \right) \\
&\approx (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)^{\mathsf{T}} \boldsymbol{H}_k^{\text{eff}} (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) \\
&\approx \boldsymbol{p}_k^{\mathsf{T}} \boldsymbol{H}_k^{\text{eff}} \boldsymbol{p}_k
\end{aligned} \tag{3.45}$$

which has the same sign as the curvature of the effective potential along $\boldsymbol{p}$. We can therefore conclude in nonconvex regions near saddle points, which are minima of the effective potential and thus convex regions of U^{eff}, that $\boldsymbol{s}_k^{\mathsf{T}} \boldsymbol{y}_k > 0$ for all small steps. Thus, no line search is required close to saddle points. This also explains why normally $\boldsymbol{s}_k^{\mathsf{T}} \boldsymbol{y}_k < 0$ for small steps near minima of U, they are nonconvex regions of U^{eff} (probably concave regions) so some step directions have negative curvature and as a result $\boldsymbol{s}_k^{\mathsf{T}} \boldsymbol{y}_k$ is negative.

Algorithm 3.6 Dimer method, local saddle point search.

function DIMERSEARCH($\boldsymbol{R}_0, \hat{\boldsymbol{N}}$) ▷ Move $\boldsymbol{R}_0$ to a local SP and orient $\hat{\boldsymbol{N}}$ along the corresponding minimum mode.
  $s_{\text{trust}} \leftarrow s_{\text{min}}$
  $\boldsymbol{g}_0 \leftarrow \boldsymbol{\nabla} U(\boldsymbol{R}_0)$
  ROTATEDIMER($\boldsymbol{R}_0, \boldsymbol{g}_0, \hat{\boldsymbol{N}}$) ▷ Algorithm 3.5.
  $\boldsymbol{F}_t \leftarrow \boldsymbol{F}_t(\boldsymbol{g}_0, \hat{\boldsymbol{N}})$ ▷ Equation (3.43).
  Escape the convex region following directions PR^+ conjugate to $\boldsymbol{F}_t$ ▷ Section 3.4.2
  for $k \leftarrow 1, 2, \ldots, I^{\mathsf{T}}_{\text{max}}$ **do**
    if $\|\boldsymbol{F}_t\| < F_{\text{tol}}$ **then**
      return Success ▷ Converged to SP.
    Calculate descent step $\boldsymbol{p}(\boldsymbol{F}_t, \boldsymbol{R}_0)$ using L-BFGS ▷ Section 3.3.
    $\boldsymbol{R}_0 \leftarrow \boldsymbol{R}_0 + \min\left(1, \frac{s_{\text{trust}}}{\|\boldsymbol{p}\|}\right)\boldsymbol{p}$
    $\boldsymbol{g}_0 \leftarrow \boldsymbol{\nabla} U(\boldsymbol{R}_0)$
    ROTATEDIMER($\boldsymbol{R}_0, \boldsymbol{g}_0, \hat{\boldsymbol{N}}$) ▷ Algorithm 3.5.
    $\boldsymbol{F}_t \leftarrow \boldsymbol{F}_t(\boldsymbol{g}_0, \hat{\boldsymbol{N}})$ ▷ Equation (3.43).
    $P \leftarrow -\boldsymbol{F}_t^{\mathsf{T}}\boldsymbol{p}$ ▷ Equation (3.44)
    if $P > P_{\text{tol}}$ and $s_{\text{trust}} > s_{\text{min}}$ **then**
      $s_{\text{trust}} \leftarrow \frac{s_{\text{trust}}}{m}$ ▷ m is a constant multiplier.
    else if $P < -P_{\text{tol}}$ and $s_{\text{trust}} < s_{\text{max}}$ **then**
      $s_{\text{trust}} \leftarrow m s_{\text{trust}}$
  return Fail ▷ Failed to converge to SP.

To rapidly escape the convex regions of the potential U we begin by translating the dimer small, fixed, distances along directions PR^+ (Polak–Ribière) [30, 73] conjugate to the translational force, until the dimer is in a nonconvex region.

A single dimer method iteration consists of rotating the dimer to align with the minimum mode and a single translation, this process is summarised in Algorithm 3.6. We set a limit on the total number of iterations ($I^{\mathsf{T}}_{\text{max}} = 200$ unless otherwise specified), above which the saddle point search is considered to have failed.

3.5 Caching saddle-point searches

By far the most time intensive subroutine of the OLKMC method is the SP search procedure however, most of these SP searches are unnecessary. This is because, due to the local nature

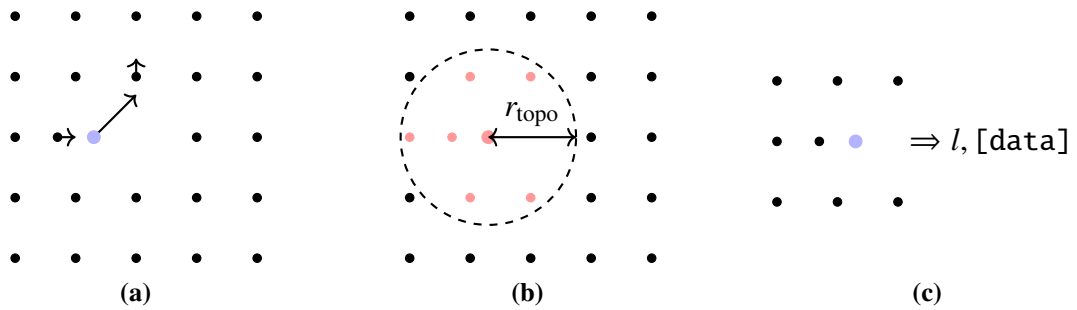


Fig. 3.5 Summary of the mechanism caching process. In Fig. 3.5a the basin → basin displacement vectors of a (hypothetical) mechanism are marked; an H atom (blue dots) bound to a vacancy in the α-Fe lattice (black dots) migrates to an adjacent site. Fig. 3.5b highlights in red the atoms within r_{topo} of the maximally displaced atom. Fig. 3.5c shows the information used to produce the topological key l and reconstruction data stored in the catalogue.

of the perturbations, most successful SP searches are rediscovering mechanisms (explicitly the basin → SP → basin pathway) that are identical to others in the current time-step, subject only to a Euclidean transformation. Additionally, there is the potential to reuse saddle-points discovered at previous time steps or between simulation runs and even refine saddle-points that are commonly encountered using a higher level of theory (i.e. DFT + nudged elastic band method [43]).

To reuse previously discovered saddle-points instead of running SP searches we devise the following generalised method (also sketched in Fig. 3.5) to find and store new saddle points:

1. Beginning with the relaxed set of atoms, A, find a connected SP linking the current basin to an adjacent basin.

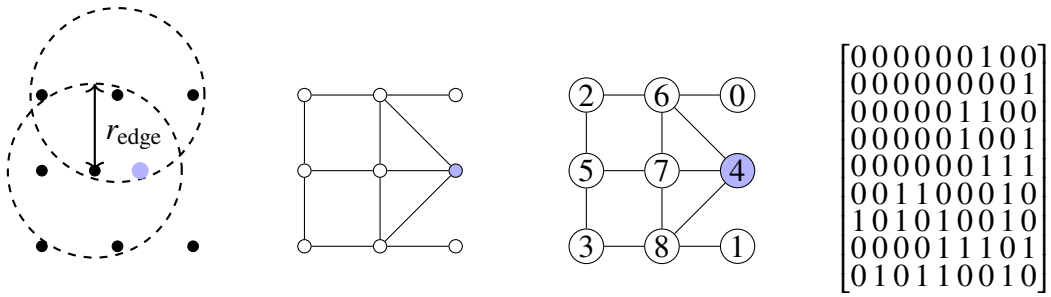
2. Identify the 'central' atom $\omega \in A$ of the mechanism, we choose the atom that undergoes the largest displacement.

3. Identify the subset of atoms within r_{topo} of ω: $A' = \left\{\alpha \in A \mid \|\boldsymbol{r}^{\alpha} - \boldsymbol{r}^{\omega}\| < r_{\text{topo}}\right\}$.

4. Produce some discrete label, l, as a function of the positions of the atoms in A', invariant under Euclidean transformations, that identifies the local structure around ω.

5. Store in some associative container (called the *catalogue*) using l as the key: the rate constant of the reaction, displacement vectors during the mechanism of the atoms $\alpha \in A'$ and any additional information required to reconstruct the mechanism.

Before launching any future SP searches, the label for each atom (as a function of its local environment) can be calculated and their SP/mechanisms reconstructed from the data stored



(a) **(b)** **(c)** **(d)**

Fig. 3.6 Simplified topological classification of the set of atoms identified in Fig. 3.5 using a fixed r_{edge}. The nodes in Fig. 3.6b represent the atoms in Fig. 3.6a, atoms within r_{edge} are connected by an edge. Figure 3.6c is the canonically labelled graph and Fig. 3.6d is the canonical adjacency matrix which, alongside the colour vector, forms part of the topological label of the geometry.

in the catalogue. This method decouples the perturbation applied to the initial system during the SP search, from the centre of the resulting mechanism discovered. Furthermore, as the rate constants are stored in the catalogue the KMC algorithm can be applied first and only the successful mechanism reconstructed, further accelerating the simulation. Once a mechanism has been reconstructed from the catalogue the full set of atoms is fully relaxed to account for any small long-range effects that may accumulate over many MC steps.

A key assumption about the nature of the mechanisms required for the above method to hold is that all atoms $\alpha \in A \setminus A'$ displace little or not at all during the transition. This assumption has practically been observed to hold true for all our tests when using a sufficiently large r_{topo}, typically chosen as $r_{\text{topo}} = r_{\text{cut}} = 6\text{Å}$.

3.5.1 Topological classification of the local geometry

The non-trivial aspect of the method is the procedure for producing a discrete label from the local structure around an atom. A balance must be struck between: being too general, thus mapping distinct local structures with different mechanisms to the same label (a *topology collision*) and too intolerant of tiny perturbations in the positions of the atoms in A' due to floating-point error and small long-range effects.

To ensure the labelling is invariant under Euclidean transformations it is easiest to construct a function of invariants rather than a function of the atomic coordinates $\{\boldsymbol{r}^\alpha \mid \alpha \in A'\}$. The most natural set of invariants under the Euclidean transformation $\boldsymbol{r} \mapsto \boldsymbol{O}\boldsymbol{r} + \boldsymbol{c}$, with $\boldsymbol{O}$ an orthogonal matrix, are the set of $\frac{1}{2}n(n-1)$ distances between atoms: $\left\{r^{\alpha\beta} = \left\|\boldsymbol{r}^\alpha - \boldsymbol{r}^\beta\right\| \mid \alpha, \beta \in A'\right\}$.

Our attempts to define an invariant function, using the full $\left\{r^{\alpha\beta}\right\}$ to construct a discretised radial distribution failed, as they were too sensitive to small variations in $r^{\alpha\beta}$ at the borders

of radial bins causing them to fall into different bins non-deterministically. To overcome this problem we choose a second radius r_{edge} and define the adjacency matrix:

$$G_{\alpha\beta} = \begin{cases} 1, & r^{\alpha\beta} \leq r_{\text{edge}} \\ 0, & \text{otherwise} \end{cases} \tag{3.46}$$

By choosing r_{edge} between the first and second nearest-neighbour distance in the BCC lattice the topology of the graph defined by $\boldsymbol{G}$ becomes insensitive to small perturbations in the atomic positions.

The question of determining whether two sets of atoms have the same label now becomes a question of whether their graphs are isomorphic, a problem in its own complexity class GI $\in$ NP which is not known to be in either P or NP-complete [27]. Fortunately, there exits graph isomorphism testing algorithms that have polynomial time complexity for most practical graphs, such as the highly optimised `nauty`[1] library by McKay and Piperno [50].

We assign a colour to each atom and use `nauty` to produce a canonical labelling of the atoms in A' and the canonised adjacency matrix $\boldsymbol{C}_G$, stored in a compacted bit field. This labelling process is represented graphically in Fig. 3.6. The topological label for the graph is then the pair $(\boldsymbol{C}_G, \boldsymbol{k}_G)$, where $\boldsymbol{k}_G$ is a vector of colours in canonical order. If two pairs compare equal (all elements compare equal) then the two graphs are isomorphic and we assume the geometry of the atoms in each set is the same.

Initially we coloured the nodes in the graph according to atomic number of the atoms. Unfortunately, this method did not incorporate enough information about the local geometry to prevent topology collisions between complex systems with small atomic displacements, such as the small displacement of a tetrahedral hydrogen as an adjacent tetrahedral site becomes occupied. Furthermore, as r_{edge} is fixed, atoms near the surface of the (approximate) ball of atoms making up A' have fewer neighbours, thus fewer edges in $\boldsymbol{G}$ (as can be seen in Fig. 3.6c with nodes 0 & 1 having only a single edge) and therefore their geometry is less well specified. To defeat both these problems we generalise r_{edge} allowing it to vary for each atom. Now to colour each atom α:

1. An initial r^{α}_{edge} is chosen such that no other atoms are within r^{α}_{edge} of α.

2. $r^{\alpha}_{\text{edge}} \leftarrow r^{\alpha}_{\text{edge}} + \delta r$, until k atoms are within r^{α}_{edge} of α.

[1] https://pallini.di.uniroma1.it/

3. Redefine $\boldsymbol{G}$:

$$G_{\alpha\beta} = \begin{cases} 1, & r^{\alpha\beta} \leq r^{\alpha}_{\text{edge}} \\ 0, & \text{otherwise} \end{cases} \tag{3.47}$$

4. The colour of the α^{th} node, in the new digraph defined by $\boldsymbol{G}$, is given by pair (Z^{α}, s^{α}) with Z^{α} the atomic number and $s^{\alpha} = \left\lfloor c \sum_{\beta} G_{\beta\alpha} r^{\alpha\beta} \right\rfloor$.

This procedure encodes much more of the information within $\left\{r^{\alpha\beta}\right\}$ into the digraph, dramatically reducing the possibility of topology collisions without any modifications required to the graph canonization/labelling. Typically, k is set to the number of spatial dimensions in the simulation and c is a fraction of the smallest atomic separation.

The use of $G_{\beta\alpha}$ instead of $G_{\alpha\beta}$ when calculating s^{α} is significant as often mechanism in the Fe-H system will involve symmetric rearrangements of an H atom, this only significantly alters $r^{\alpha\beta}$ for atoms directly connected to the H atom. If $G_{\alpha\beta}$ was used then these changes would only be visible in s^{hydrogen} and can be lost during the summation. This is avoided by summing over $G_{\beta\alpha}$, which distributes the information of a single H atom moving onto all its neighbours.

The validity of the topological classification depends on the existence of a one-to-one correspondence between topologies and geometries. Choosing a larger r_{topo} or k reduces the probability of a topology collision. During the simulation, in order to constantly validate the existence of the one-to-one correspondence, we store a second catalogue associating topological keys to a reference geometry. Every time a topology is re-encountered the geometry of the atoms are confirmed to match the stored reference. If a match fails then k is dynamically adjusted until the topologies are distinguishable.

All topological catalogues are written to a file on disc at the end of each simulation. This enables the implementation to 'learn' between runs, as a complete catalogue of events that can occur in each system is slowly built. This catalogue file can also be shared between computers, to bypass the learning phase at the beginning of simple systems.

3.5.2 Mechanism reconstruction

To reconstruct a mechanism for the set of atoms B', from a mechanism found at a set of atoms B that has the same topological label, we define an orthonormal basis sets in B' and B:

1. Order the atoms into their canonical orders using `nauty`.

2. Taking the first atom as the origin and the second atom as the direction of the first basis vector $\boldsymbol{e}^1 = \boldsymbol{r}^1 - \boldsymbol{r}^0$.

3. Taking the next atom α noncollinear to atoms $0, 1$ as the direction of the second basis vector $\boldsymbol{e}^2 = \boldsymbol{r}^\alpha - \boldsymbol{r}^0$.

4. Choosing the next atom β noncoplanar to atoms $0, 1, \alpha$ as the final basis vector $\boldsymbol{e}^3 = \boldsymbol{r}^\beta - \boldsymbol{r}^0$.

5. Orthonormalizing $\boldsymbol{e}^1, \boldsymbol{e}^2, \boldsymbol{e}^3$ using the Gram–Schmidt process [5, p. 516].

The orthogonal transformation matrices $\boldsymbol{L'}$ and $\boldsymbol{L}$ are then constructed using the basis vectors as columns, e.g. $L_{ij} = \left(\boldsymbol{e}^j\right)_i$. The displacement of the α^{th} atom in B' is then computed from the stored displacement vector $\delta\boldsymbol{r}^\alpha$ in B as:

$$\delta\boldsymbol{r'}^\alpha = \boldsymbol{L'}\boldsymbol{L}^\mathsf{T}\delta\boldsymbol{r}^\alpha \tag{3.48}$$

To avoid having to store $\boldsymbol{L}$ in the catalogue in our implementation we store $\left\{\boldsymbol{L}^\mathsf{T}\delta\boldsymbol{r}^\alpha \mid \alpha \in B\right\}$ directly.

As an additional check that no topological collisions or reconstruction failures have occurred we store in the catalogue of mechanisms the total energy change of the system during the mechanism. If the energy change after the reconstruction is too different or the reconstructed basin is too far from a true minima the reconstruction is considered to have failed. In this case r_{topo} and/or k can be adjusted.

3.6 The low barrier problem

A common issue encountered during KMC simulations is the so called, *low barrier problem* [52, 81], which occurs when a collection of basins are connected by a series of low energy barriers. This collection of basins – often called a *superbasin* – takes many MC steps to escape from. During this period the simulation clock advances very little as internal transition are very fast. Numerous solutions have been proposed including, exact analytical treatments of the superbasins [74, 24, 8] that compute the mean superbasin escape times as well as simpler TABU-like [32] methods that ban recent transitions [52, 15].

We initially tested the TABU-like method as maintaining a memory kernel of the n most recent events is far less overhead than the matrix inversions required for the analytical solutions. We modify the rate constants according to:

$$\Gamma_{ij} \leftarrow \begin{cases} 0 & \text{if } i \to j \text{ in the memory kernel} \\ \Gamma_{ij}^{\text{TST}} & \text{otherwise} \end{cases} \tag{3.49}$$

which has been shown to be thermodynamically sound providing $n \ll$ the total number of MC steps [15]. However, during testing of the vacancy-hydrogen complex diffusion (see Section 5.3) this method produced large biases in the results. Therefore, we resolved to implement the full analytical superbasin solution.

3.6.1 Superbasin accelerated off-lattice kinetic Monte Carlo

To overcome the low-barrier we begin by following Puchala, Falk, and Garikipati's [74] *mean rate method*, the system is divided into two the sets of states: those states inside the current superbasin called the *transient* states and *absorbing* states outside the superbasin. As we cannot know which states are transient or absorbing before the simulation begins (or even what the states are), we must make this distinction as they are discovered. We define the two states to be part of the same superbasin if there exists a low-barrier mechanism linking them. A low-barrier mechanism is one with forward and reverse energy-barriers below some energy threshold, e.g. $E^{\ddagger} - E^{i} < E^{\text{tol}}$ and $E^{\ddagger} - E^{f} < E^{\text{tol}}$. Unfortunately, partitioning the states into transient/absorbing using this definition would require exploring all states inside a superbasin in order to apply the mean rate method, which could become unnecessarily expensive. This is because each new state may require a full set of SP searches to identify its set of mechanisms. Furthermore, just entering each state during the exploration requires reconstructing each mechanism from the catalogue.

The heart of the problem lies in the mean rate method focusing on differentiating states, while the fundamental object in OLKMC are mechanisms. Therefore, we change the frame of reference in the mean rate method such that we do not require knowledge of the absorbing states, by further partition the mechanisms into transient and absorbing categories. We denote a rate constant connecting states $i \rightarrow j$ via a transient mechanism as Γ^{*}_{ij}, whilst the rate constant for an equivalent absorbing mechanism remains Γ_{ij}.

Beginning from an empty set we append a *basin object*, which includes the current state of the system as well as the set of mechanisms leaving the state. As the state of the system after any of these mechanisms occurs is currently unknown, they are all marked as absorbing mechanisms. This set of basins is the superbasin and will be constructed on-the-fly.

First, the traditional KMC algorithm is used to select a mechanism and move/jump the system into a new state, s. If the selected mechanism is a high-barrier mechanism, then the superbasin is cleared and the process resets. However, if it was a low-barrier mechanism then it is marked as a transient mechanism and a new basin (constructed from s) is appended to the

superbasin. The transient transition probability matrix, $\boldsymbol{T}$, is then formed:

$$T_{ji} = \frac{\Gamma^*_{ij}}{\sum_j \Gamma^*_{ij} + \sum_k \Gamma_{ik}} = \tau_i^1 \Gamma^*_{ij} \tag{3.50}$$

where the indices i, j run over the states in the superbasin (these are the transient states), the index k runs over all absorbing mechanism leaving state i and τ_i^1 is the mean residence time in state i each time it is occupied. This describes the total probability of moving from a transient state to another transient state via a transient mechanism.

Now we introduce the transient occupation probability vector, $\boldsymbol{O}_m$, which encodes the probability of being in each transient state after m transient jumps and before jump $m+1$. Hence, with the system in transient state s, $(\boldsymbol{O}_0)_i = \delta_{is}$. Using the relationship $\boldsymbol{O}_m = \boldsymbol{T}^m \boldsymbol{O}_0$, the total number of times a transient state is visited before a jump via an absorbing mechanism occurs is:

$$\boldsymbol{O}^{\text{Tot}} = \sum_{m=0}^{\infty} \boldsymbol{O}_m = \sum_{m=0}^{\infty} \boldsymbol{T}^m \boldsymbol{O}_0 = \left(\sum_{m=0}^{\infty} \boldsymbol{T}^m \right) \boldsymbol{O}_0$$

solving the Neumann series, similar to the mean rate method [74]:

$$\boldsymbol{O}^{\text{Tot}} = (\boldsymbol{I} - \boldsymbol{T})^{-1} \boldsymbol{O}_0 \tag{3.51}$$

The mean total-time spent in each transient state is therefore $\tau_i = \tau_i^1 O_i^{\text{Tot}}$. We can now compute the mean rate of exiting from transient state i to some state l via an absorbing mechanism:

$$\langle \Gamma_{il} \rangle = \frac{\tau_i}{\sum_j \tau_j} \Gamma_{il} \tag{3.52}$$

Finally, we use the KMC algorithm on this set of absorbing mean-rates to select the next mechanism and jump to the new state. Several possibilities now exist: if the chosen mechanism leads to a transient state then it is marked as a transient mechanism, $\boldsymbol{T}$ is recalculated and the process repeats. Else, if the mechanism was a low-barrier mechanism it is marked as a transient mechanism, a new basin object is constructed and appended to the superbasin, $\boldsymbol{T}$ is recalculated and the process repeats. Otherwise, the mechanism was a high-barrier mechanism therefore, the superbasin is cleared and the process resets.

This method reduces to the mean rate method if the simulation explores every internal mechanism inside a superbasin but *may* not require exploring the entire superbasin. Additionally, three further optimisations are implemented:

- Upon entering a new superbasin state for the first time all exiting mechanism will be absorbing mechanisms, this will reduce: $\boldsymbol{O}^{\mathrm{Tot}} = \boldsymbol{O}_0$, hence $\langle \Gamma_{il} \rangle = \delta_{is} \Gamma_{il}$ and the matrix inversion can be skipped.

- When resting the method, instead of discarding the superbasin it is cached into a circular buffer. Following a future high-barrier absorbing mechanism the superbasin can be reloaded from this cache if it has been encountered before. This is particularly critical as it avoids re-exploring a superbasin that is re-entered immediately after it has been escaped.

- If the superbasin grows beyond some maximum size then the algorithm reverts to standard KMC, this prevents overfilling the memory of the host computer and slowdowns as the matrix inversions scales polynomially with the number of basins in the superbasin. The maximum size must be set larger than the number of states in the superbasin of interest e.g. 6 states in a vacancy-hydrogen complex.

Chapter 4

Validation

This chapter focuses on verifying our implementations of the numerical methods discussed in Chapter 3. Section 4.1.1 tests the saddle-point search algorithm in a well-studied 2D potential and explores an alternative translational force. Section 4.1.2 combines our EAM force implementation with the dimer SP search to interrogate a simple and relevant system, a vacancy in the α-Fe lattice. Finally, in Section 4.2 we combine the kinetic Monte Carlo algorithm and topological caching into the model and we verify the kinetic results using a asingle hydrogen atom diffusing in the perfect lattice as a simplest test model.

4.1 Dimer method

4.1.1 Two-dimensional model potential

A simple two-dimensional model potential was the first test for the dimer-method implementation as it provides easy visualisation and rapid testing. The model chosen was a London–Eyring–Polany–Sat (LEPS) potential [72] between three collinear atoms coupled to a harmonic oscillator, the analytical form is detailed by Jónsson, Mills, and Jacobsen [43]. Additionally, the Gaussian modifications made by Henkelman and Jónsson [37] are included to increase the number of saddle points, this potential has been used in previous studies to investigate the effectiveness of saddle point search algorithms [37, 86, 21]. Key features of the potential include four saddle points and two local maxima surrounding the initial basin (all highlighted in Fig. 4.1), centred at $(0.77, -0.07)$. The slowest ascent direction, towards $-\hat{\boldsymbol{y}}$, does not lead to a saddle point and there are no saddle points in the $-\hat{\boldsymbol{x}}$ direction $\left(\lim_{\|\boldsymbol{r}\|\to\infty} U^{\mathrm{LEPS}}(\boldsymbol{r}) \to \infty\right)$.

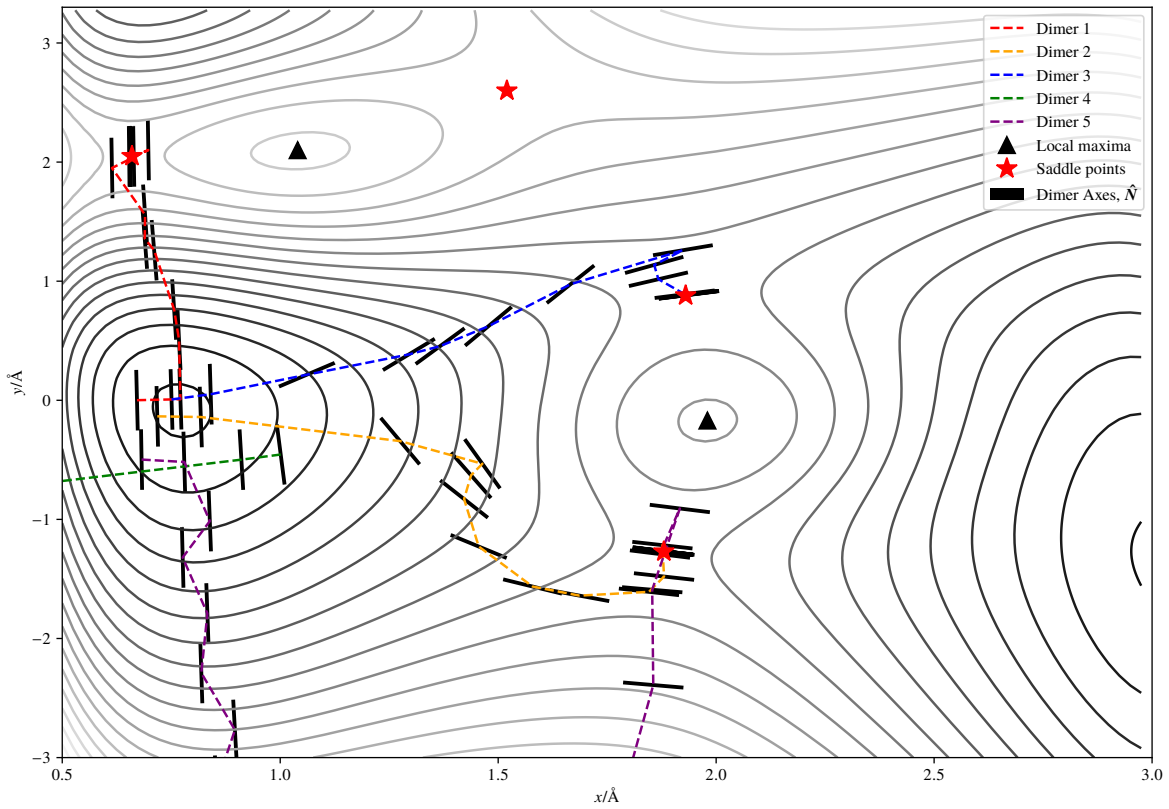


Fig. 4.1 Dimer search trajectories in a two-dimensional LEPS potential coupled to an harmonic oscillator [43]. The dimers are initialised with random Gaussian distributed positions centred on the basin at $(0.77, -0.07)$, with a standard deviation of 0.1Å along each axis. The dimer axes are marked with black bars and should be aligned with the minimum curvature mode at the location of the dimer. The dimer separations are too small to resolve. Dimers 1, 2 and 3 converge to saddle points in 14, 17 and 15 iterations respectively. Dimer 4 moves to $x \approx -4$Å before the search is aborted due to not finding a nonconvex region within 50 iterations. Dimer 5 follows the softest mode aligned with the y-axis until it reaches $y \approx -10$Å then, realigns with the x-axis aligned mode and returns to a saddle point.

Dimer searches were carried out starting the system from the initial basin with small Gaussian displacements along each coordinate. In total ten dimer searches were conducted, trajectories for five of the dimer searches are displayed in Fig. 4.1.

Dimers 1, 2 and 3 rapidly converge to saddle points connected to the initial basin. A sharp change of direction in dimer path 2 is noticeable, this is the point the dimer switches from the conjugate gradient method in the convex region to the L-BFGS method in the nonconvex region. Upon switching the L-BFGS algorithm has no knowledge of the history of the path so initially follows the steepest descent direction and thus the change in direction.

None of the dimer searches converged to the upper right saddle point located at $(1.51, 2.60)$, any dimers moving in this direction are instead attracted to the lower energy saddle at $(1.93, 0.88)$.

This behaviour is observed in previous works [37], thus is not an implementation issue. Rather, this is a systematic exclusion by the dimer method, that can only be overcome with larger initial displacements or biasing the dimer away from previously discovered saddle points. However, the former can increase the chance of finding disconnected saddle points (which could be mitigated at a higher computational cost with the κ-dimer method [86]) and the latter could still exclude saddle points 'shielded' by saddle points closer to the initial basin. Fortunately, the saddle point excluded is on the highest energy pathway out of the basin as well as being on a very curved reaction pathway. This means the rate constant along this pathway is likely to be extremely low and excluding it should not have too large an effect on the KMC simulation.

Three of the dimer searches did not converge to a saddle point. Instead, like dimer 4 in Fig. 4.1, they move in the $-\hat{\boldsymbol{x}}$ direction where there are no saddle points until some iteration limit. To mitigate this effect the alternative translational force [37]:

$$\boldsymbol{F}_t = \hat{\boldsymbol{N}}\boldsymbol{g}_0^{\top}\hat{\boldsymbol{N}} \tag{4.1}$$

was investigated in the convex region. This resulted in far fewer non-converging dimers but severely biased the dimer search to the saddle points at $(0.66, 2.04)$ and $(1.87, -1.29)$. This is because the minimum mode near the centre of the basin is oriented along the y-axis, resulting in many dimers initially moving along the $\pm\hat{\boldsymbol{y}}$ directions. This can be avoided by orthogonalising successive dimer-search axes but, this comes with additional computational costs and reverts to the initial problem of searching in directions with no saddle points. The decision was made not to use the modified translational force due to this bias.

4.1.2 Vacancies in α-Fe

The first 3-dimensional test for the dimer method implementation was a simple α–Fe lattice using a cubic supercell of 685 atoms, containing a single vacancy, under full periodic boundary conditions. Two EAM potentials (denoted A & B) of the form described in Section 3.1 were investigated, the full analytical functions are given by Ramasubramaniam, Itakura, and Carter [75]. Gaussian displacements along each coordinate axis for each atom close to the vacancy were made and a saddle-point search carried out; the maximum translational step size was $s_{\max} = 0.5$Å. Providing the SP search converged the system was then displaced ± 0.1Å along the dimer axis and the total systems energy minimised. If one of the final positions was sufficiently close to the initial position/minima the SP was considered *connected* and the SP search successful. The norm of the $3n$-dimensional force/effective-force was converged to 10^{-5}eV Å^{-1} and the dimer axis converged to within 1° of the minimum mode. The program

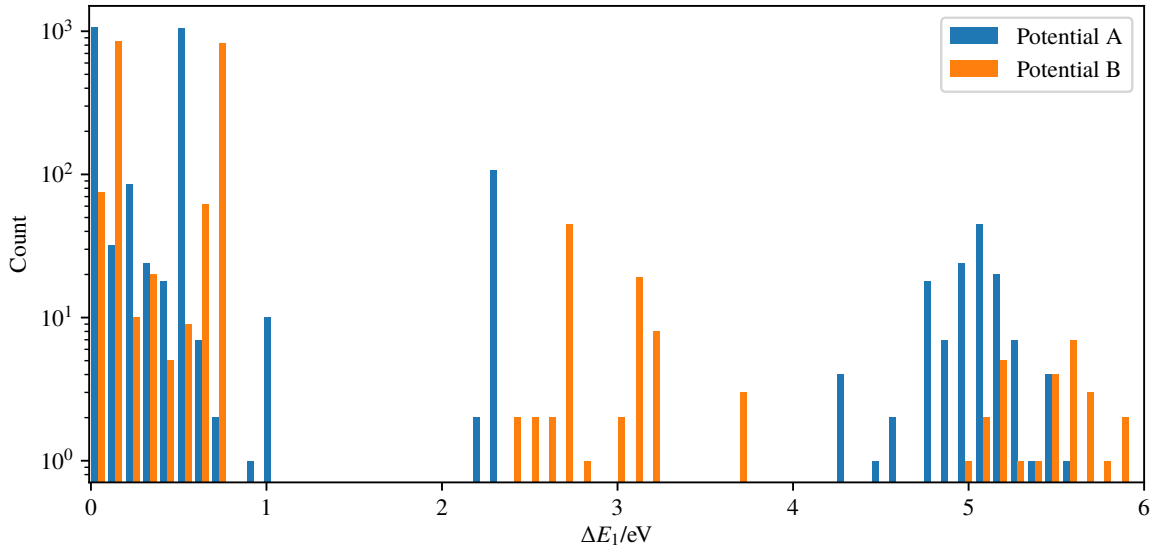


Fig. 4.2 Histogram of the occurrence count of each mechanism found during 5000 iterations of SP searches classified by their activation energy, ΔE_1. Potential A successfully converged to 2534 connected saddles whilst potential B converged to 1976. The amplitude of the Gaussian displacements was 0.3Å. The mechanisms with activation energies (potential A) of 0.53, 1.06 and 2.34Å are sketched in Fig. 4.3. In total 43 mechanisms were found with unique activation energies.

carried out SP searches until successful, the mechanism found was then carried out (regardless of the energy barrier) and the process repeated. This method was chosen in the hope of discovering the maximum variety of mechanisms.

Two energies were recorded for each transition, the activation energy, ΔE_1, equal to the energy change between the SP and the initial state and, ΔE_2, the energy change between the final relaxed state and the SP. Figure 4.2 displays a histogram of the activation energies found during 5000 iterations for both potentials. The mechanism found can be summarised into four groups, correspondingly sketched in Fig. 4.3:

- A vacancy corner atom (A) transitioning into a shallow minima on the cusp of moving to the vacancy centre. This accounts for the two most frequently found mechanisms with activation energies 0.53, and 0.04eV corresponding to moving into and out-of the shallow minima, respectively. A pair of these mechanisms constitutes a $\frac{1}{2}\langle 111\rangle$ translation of the vacancy. This pair of transitions collectively account for $\approx \frac{4}{5}$ of the successful SP searches.

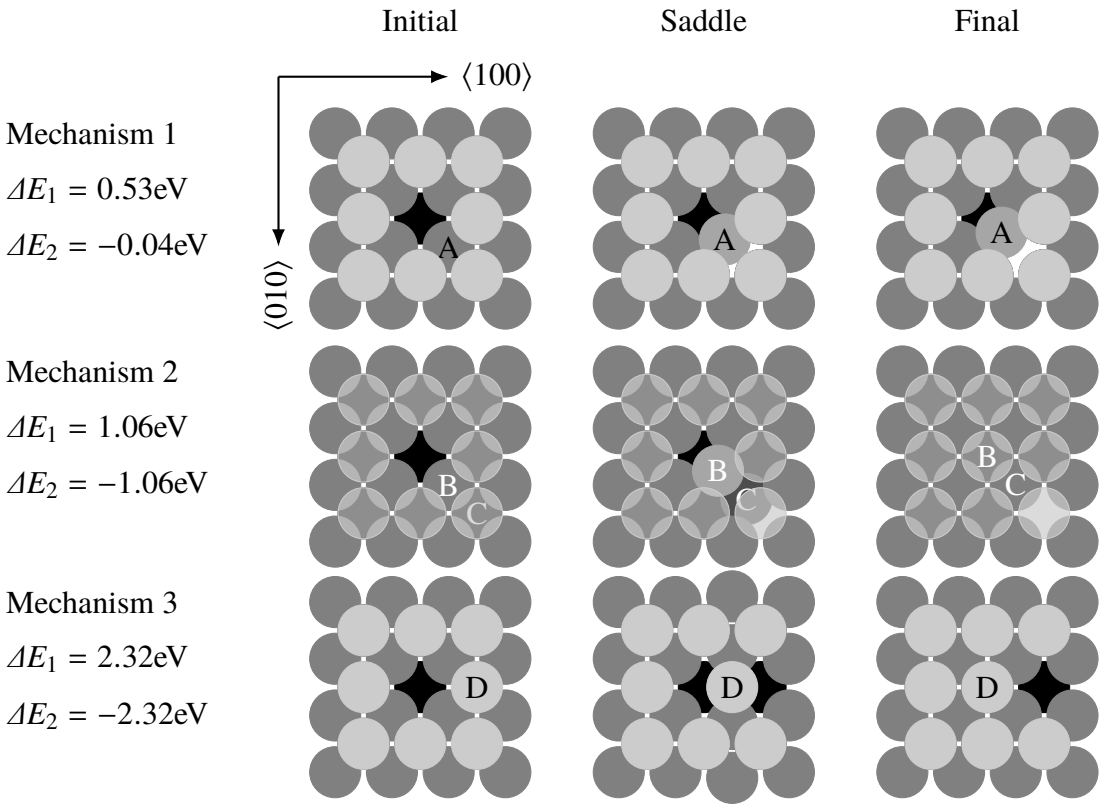


Fig. 4.3 A selection of the lowest energy mechanisms found for vacancy migration in α–iron using the dimer method saddle-point search and potential A [75]. ΔE_1 is the energy change for the initial→saddle transition while ΔE_2 is the corresponding saddle→final energy change. All the mechanism have a corresponding reverse mechanism such that $\Delta E_1 \leftarrow -\Delta E_2$ and $\Delta E_2 \leftarrow -\Delta E_1$.

- A rarely found ≈ 1eV mechanism, similar to mechanism 1, with two aligned corner atoms (B & C) simultaneously moving resulting in the vacancy moving two lattice sites i.e a ⟨111⟩ vacancy migration.
- An atom adjacent to a vacancy (D) pushing through the under-coordinated corner atoms resulting in the ⟨100⟩ vacancy migration. The energy barrier is four times higher than the $\frac{1}{2}\langle 111\rangle$ migration explaining why this mechanism is not experimentally observed.
- Finally, a clustering of high energy ≈ 5eV mechanisms, involving several atoms, forming interstitial Fe atoms or rearrangements of the lattice away from the vacancy.

These mechanisms are in good agreement with the literature [16] and are the most immediately physically reasonable. Furthermore, the activation energy for vacancy migrations for both potentials are in good agreement with the experimental range of 0.55–1.3eV [65].

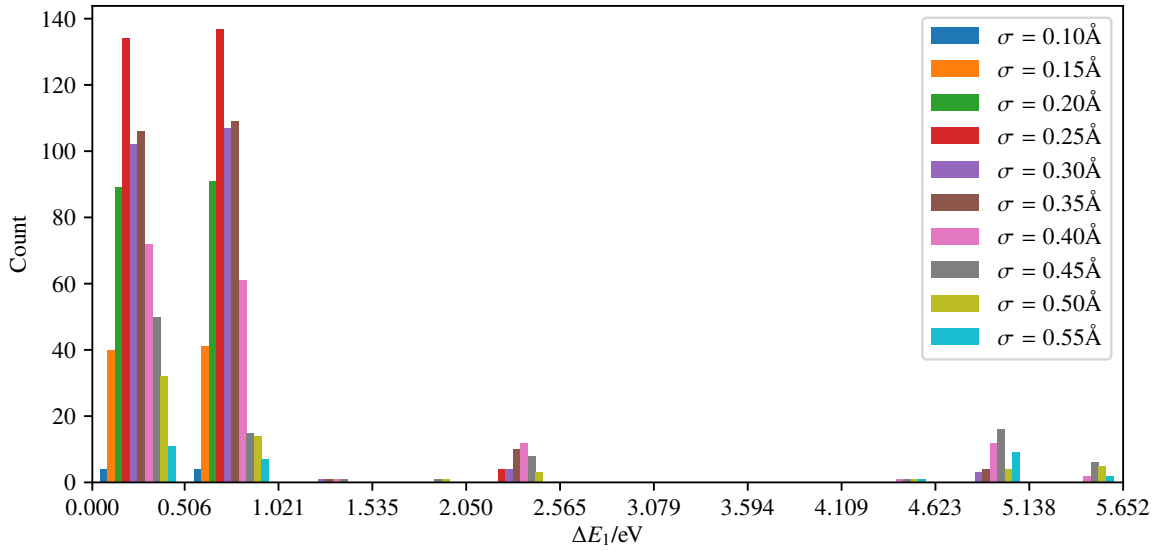


Fig. 4.4 Histogram of activation energies, ΔE_1, found during 500 iterations of SP searches using potential A for a selection of standard deviations of the initial Gaussian displacements. Coarse binning (11 bins) has been used to highlight the distribution of energies rather than show each distinct mechanism. Figure 4.5 presents the corresponding success rate of finding a SP at each standard deviation.

The mechanisms discovered were the same for both potentials however the activation energy reported for potential B was consistently higher than potential A. This is an unfortunate discrepancy as it shows there is still disagreement between potentials. We can draw reassurance however, that the mechanisms describe the same rearrangement of atoms and the relative activation energies are consistent. Hence with appropriate choices of the Arrhenius prefactor (see Section 1.2.1 for a detailed discussion) the potentials should produce similar kinetic results.

The large skew in the distribution of mechanism discovery suggests a high number of SP searches will be required per local arrangement of atoms, in order to build a complete catalogue of saddle points required for accurate KMC. To investigate the relationship between the distribution of mechanisms and the amplitude of the initial Gaussian displacements a series of shorter runs using potential A were performed, the results are presented in Fig. 4.4 and Fig. 4.5.

From Fig. 4.5 it is clear that the amplitude of the initial displacements has a strong effect on the success rate of SP searches. Large standard deviations are expected to converge to disconnected saddle points thus decreasing the success rate, this behaviour is seen in Fig. 4.5 when $\sigma > 0.325$Å. However, below $\sigma = 0.25$Å the success rate also drops, almost zero at $\sigma = 0.1$Å. This is due to the smaller displacements increasing the probability of starting in a convex region of the effective potential which then increases the probability of getting stuck in

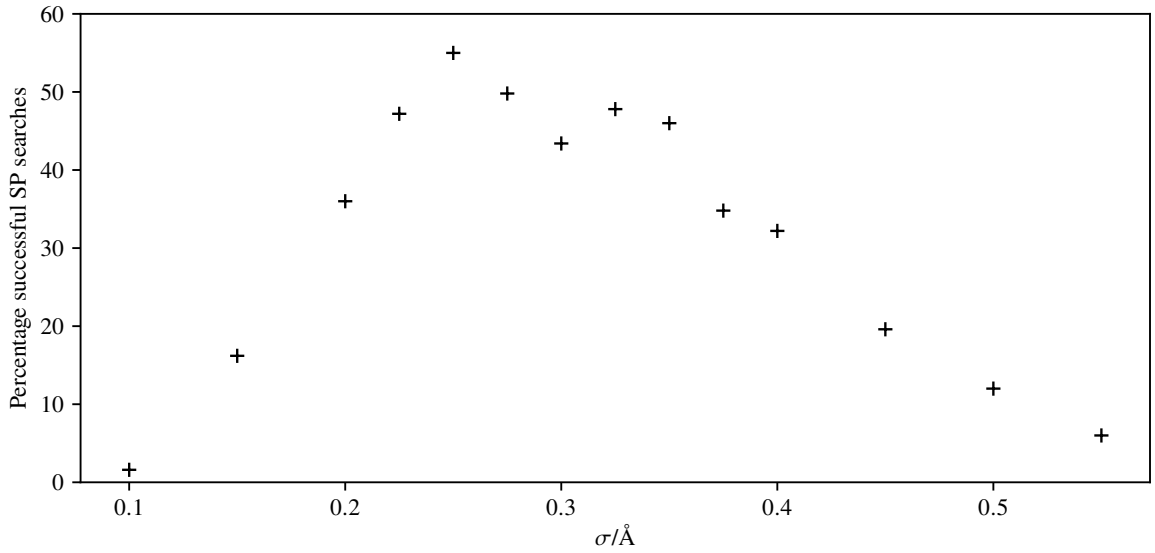


Fig. 4.5 Percentage success rate (probability of finding a connected SP) during 500 iterations of SP searches using potential A for a range of standard deviations, σ, of the initial Gaussian displacements. Corresponding distributions of activation energies are available in Fig. 4.4.

a *soft* mode. This is when the dimer axis displaces all atom uniformly in the same direction, effectively translating the supercell. Due to the periodic boundary conditions this has no effect on the total energy thus all potential energy derivatives are zero along this axis, including the second derivatives therefore, the curvature along the minimum mode is zero. This results in the SP search not converging.

The small peak around $\sigma = 0.325$Å in Fig. 4.5 is likely due to previously shielded SP becoming accessible at the larger displacements. This is supported by Fig. 4.4 when we see high energy saddle points ($\Delta E_1 > 2.5$eV), corresponding to forming interstitial Fe atoms away from the vacancy, only being discovered when $\sigma \geq 0.3$Å.

Figure 4.4 reveals that smaller displacements biases the SP search to finding lower energy saddle points. The two largest peaks corresponding to the lowest energy mechanisms discovered: the reverse and forward of mechanism 1 in Fig. 4.3, respectively. We expect lower energy saddle points to be closer to the initial basin so this behaviour is natural. Consequently, the solution to picking the best σ is not as simple as choosing the σ that produces the highest SP success, which for this test seems to be $\sigma = 0.25$Å, as this may result in requiring more SP searches to find the complete set of mechanisms. This is supported by Fig. 4.4 where $\sigma = 0.25$Å finds no mechanisms with $\Delta E_1 > 2.5$eV. Conversely picking a σ too high will bias the search toward, less kinetically relevant, high energy saddle points. This bias has not invalidated the mechanisms found previously, as a higher $\sigma = 0.3$Å was used in Fig. 4.2 and many more SP

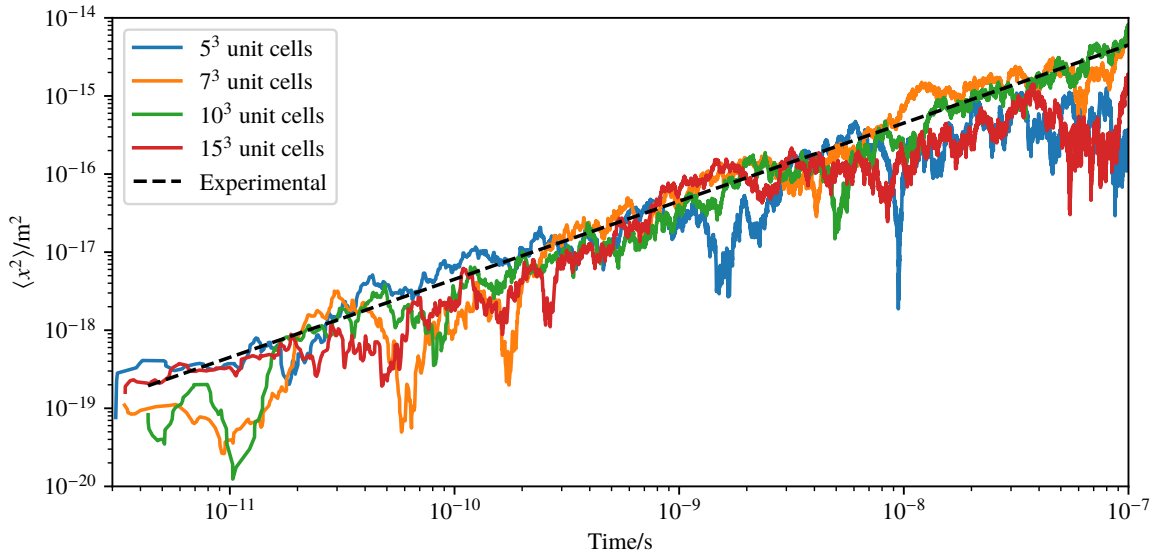


Fig. 4.6 Simulated displacement of a single H atom freely diffusing in the perfect α-Fe lattice, for a selection of supercell volumes, at 300K and with an Arrhenius pre-factor of 5.12×10^{12}Hz. The simulations typically required 475,000 Monte Carlo steps and approximately 120 hours of total run-time on a quad-core Intel® Xeon® at 3.10GHz. Plotted for reference is the expectation squared displacement using the experimental value of diffusivity, $D = 7.5 \times 10^{-9} \mathrm{m}^2\,\mathrm{s}^{-1}$ [35] and the Einstein diffusion equation.

searches were performed. This is evidenced by the large number of high energy mechanism in Fig. 4.2.

In conclusion, the optimal value of σ is likely to be problem dependant and should be chosen to minimise the number of SP searches to find all mechanisms of kinetic relevance. A better method of perturbing the initial atoms may be equidistributed on the surface of a hypersphere [34] or an adaptive method; first using small displacements to find the most important low energy mechanisms then large displacements to find the higher energy mechanisms.

4.2 A single hydrogen atom diffusing in the α-Fe lattice

To verify the topological caching, scaling and kinetic behaviour of the model; hydrogen diffusion in the perfect α-Fe lattice was simulated over a series of supercells. A single H atom was placed in a 5^3, 7^3, 10^3 and 15^3 unit-cell supercell, at 300K, with an Arrhenius pre-factor of 5.12×10^{12}Hz and allowed to diffuse for 10^{-7}s. The conditions were chosen to replicate those used by Bombač et al. [12] in their traditional KMC experiment. The results are presented in

Table 4.1 Summary of the results for the simulation presented in Fig. 4.6.

Supercell size/unit-cells	Number of Fe atoms	ΔE_a/eV	$D \times 10^9$/m^2 s^{-1}
5^3	250	0.0378	7.9 ± 0.7
7^3	686	0.0377	7.4 ± 0.6
10^3	2000	0.0376	8.4 ± 0.7
15^3	6750	0.0376	7.6 ± 0.7

Fig. 4.6 and Table 4.1; diffusivity, D, in Table 4.1 is computed via the following equation [66]:

$$D = \frac{1}{N} \sum_{i=1}^{N} \frac{R_i^2}{6\Delta t_i} \tag{4.2}$$

where the trajectory of the H atom is broken into N time segments Δt_i of approximately equal length Δt and R_i^2 is the squared displacement of the atom in the corresponding time segment. The resulting diffusivity is weakly dependant on Δt and is chosen to be as large as possible such that D converges to the asymptotic limit however, there must be enough samples to reduce the standard error.

The results for the activation energy of the hydrogen tetrahedral $\rightarrow$ tetrahedral transition, ΔE_a – which quickly converges to 0.0376eV and becomes independent of system size – confirmed the energy, force and dimer method subroutines scale correctly with the system. Additionally, the absolute value of 0.0376eV, exactly matches the value measured using the LAMMPS[1] software package [71], as well as being almost identical to the diffusion barrier of 0.039eV [75] reported by the authors of the potential. This gives a high level of confidence that the energy barriers reported are correct within the bounds of the underlying potential.

Figure 4.6 demonstrates clear diffusive behaviour over the 10^{-7}s timescale, confirmed by the fitted diffusivity values, which are all within error of the experimentally measured value of 7.5×10^{-9}m^2 s^{-1} [35] and close to replicating the KMC prediction of 8.981×10^{-9}m^2 s^{-1} [12]. However, the error in the diffusivity fit is quite high and the diffusivity does not converge to a constant, as observed by Bombač et al. [12]. This is due to the use of a single H atom, as opposed to the 5–30 used by Bombač et al., producing poor statistics as we are unable to average the displacement over multiple H atoms. Furthermore, Bombač et al. do not report the diffusivity converging to a constant value until simulation times beyond 3×10^{-6}s hence, this non-convergence is expected.

[1] http://lammps.sandia.gov

We can conclude the model and potential reproduce the expected kinetic results and that size of the supercell, for supercells greater than 5^3 unit cells, has very little impact on the activation energy or measured diffusivity. This is to be expected as the presence of the interstitial hydrogen has only a very small effect on the positions of its neighbouring Fe atoms and the smallest supercells edge length was still twice the cut of radius (6Å) of the EAM potential. Therefore, we can safety use smaller supercells to reduce the simulation run time.

Now that we have validated the methodologies and implementation in the simplest environments, next step is to begin to explore more complex configurations that are out of reach of traditional KMC methods, as well as continuing to make comparisons when possible to other experiments and modelling.

Chapter 5

Results & Discussion

In this chapter the complexity of the simulated systems is expanded. During Section 5.1 we increase the number of hydrogen atoms in the supercell and contrast the results to traditional KMC methods, as well as exploring the limitations of the underlying EAM potential. In Section 5.2 we explore the richer set of mechanism discovered through saddle-point searches for two and three vacancy clusters and the effects they have on the diffusivity of these point defects. Finally, Section 5.3 combines the progress made in modelling vacancies and hydrogen to study the diffusion of V-H complexes and comparison to Oriani's trapping theory [64] are made.

5.1 Multiple hydrogen atoms diffusing in the α-Fe lattice

As an extension of the single hydrogen diffusion experiments in Section 4.2, equivalent runs were performed in the 7^3 unit-cell supercell, this time varying the concentration of hydrogen, the results are presented in Table 5.1. We see the measured diffusivity remains largely independent of the concentration of hydrogen. This agrees with similar traditional KMC results [12] and is

Table 5.1 Summary of results for multiple H atoms diffusing in the perfect α-Fe lattice as the hydrogen concentration varies. The simulations required approximately 170 hours of total run-time on a quad-core Intel® Xeon® at 3.10GHz.

Number of H atoms	Concentration/at.%	Simulated time/s	$D \times 10^9/\mathrm{m}^2\ \mathrm{s}^{-1}$
1	0.15	10^{-7}	7.4 ± 0.7
2	0.29	10^{-6}	8.9 ± 0.4
3	0.44	10^{-9}	6.7 ± 0.9
4	0.58	10^{-9}	8.0 ± 1.1

expected as, even at its highest, the hydrogen concentration was low enough that the obstruction to free diffusion should have been minimal. This is confirmed by the average energy barrier remaining the same as the single hydrogen case.

The larger errors in the measured diffusivity for the 3H and 4H runs are due to a significantly reduced simulated times for these experiments. This occurred due to an implementation constraint on the size of the catalogue during the simulation which has since been overcome; due to time constraints, the decision was made not to rerun the experiments. The increased number of H atoms of these later runs did partially make up for the reduced run times.

To investigate the distribution of the H atoms, the normalised radial distribution function (RDF) was calculated according:

$$g(r) = \frac{V}{4\pi t_{\text{tot}} N(N-1) r^2 \delta r} \sum_{i}^{\text{frames}} \Delta t_i \sum_{\substack{\alpha,\beta \\ \alpha \neq \beta}}^{\text{H atoms}} \delta(r - r_i^{\alpha\beta}) \tag{5.1}$$

with V the volume of the supercell, N the number of H atoms, Δt_i the time elapsed during the i^{th} frame (calculated using Eq. (1.2)), $r_i^{\alpha\beta}$ the distance between the α^{th} and β^{th} H atom during frame i and $t_{\text{tot}} = \sum_i^{\text{frames}} \Delta t_i$ the total elapsed time. This is equivalent to the standard form of the radial distribution function however, each frame of the simulation is weighted by the time elapsed during that frame instead of unity. The results are presented in Fig. 5.1, alongside a comparison to the expectation RDF for uniformly distributed hydrogen in the ideal BCC lattice under the minimum image convention in an equivalent supercell.

In Fig. 5.1 we see strong agreement between the RDFs across the three simulations, this further justifies the shorter simulated times of the 3H and 4H runs. Two clear deviations from the expectation RDF are evident. Firstly, at small r the RDF is much higher than expected with a large peak centred on 2.13Å. This peak is a combination of the 4nn peak at 2.02Å (displaced slightly apart due to the finite size of the H atoms) and the 5nn peak at 2.26Å (pulled closer together by H-H attraction). Even as a composite g is still much larger than expected and therefore implies the H atoms clustered together. This is supported by the clear depletion of H, relative to the expectation, beyond $r = 6$Å caused by this clustering. Similar clustering has been observed by others [80] using these EAM potentials [75] therefore, this unphysical behaviour is attributed as an artefact of the potential. It does not appear to have introduced any bias into the diffusivity measurements however, adopting an additional H-H repulsive component into the potential [80] seems a sensible improvement for future experiments.

Secondly, none of the H atoms approach closer than the 4nn position during any of the simulations. This can be attributed to a high energy barrier due to strong H-H repulsion at

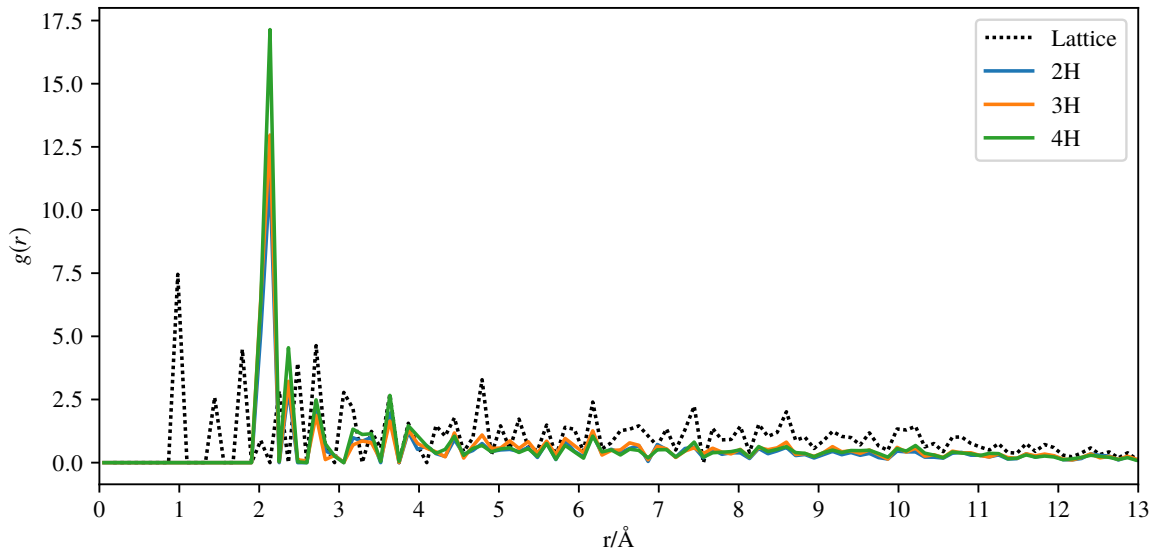


Fig. 5.1 Radial distribution functions for H atoms diffusing in the perfect lattice, calculated using Eq. (5.1) and the data gathered in the stimulations from Table 5.1. For reference the expectation radial distribution function for a uniform distribution of H atoms occupying tetrahedral interstices in the BCC lattice is plotted with a dashed black line.

small separations. This is in perfect agreement with the Switendick criterion which states the minimum H-H separation is 2.1Å [76].

Beyond $r = 6$Å the simulated RDF matches the expectations shape very closely, the only difference is depletion due to the clustering discussed above. This is because the cut-off radius for H-H interaction is 6Å hence, there is no attraction between the hydrogen atoms beyond this separation and (as the perturbation to the α-Fe lattice by hydrogen is small and local) so the H atoms adopt the expectation separations.

5.2 Vacancy diffusion in the α-Fe lattice

The motion of point defects plays a role in the deformation of crystalline materials and simulating their behaviour is the starting point for modelling extended crystalline defects. The effects of hydrogen on point defect mobility is likely to play a role in HE, the first step to understanding this interaction is to investigate the hydrogen-free mobility of di-vacancy clusters to establish a baseline. Off-lattice kinetic Monte Carlo is uniquely suited to this task due to the increasing mechanism diversity that develops for larger clusters that is difficult to predict *ab initio*.

Identification of the vacancy/vacancies location during a simulation is more complex than tracking a particle whose position is explicitly tracked hence, the process was split in two stages.

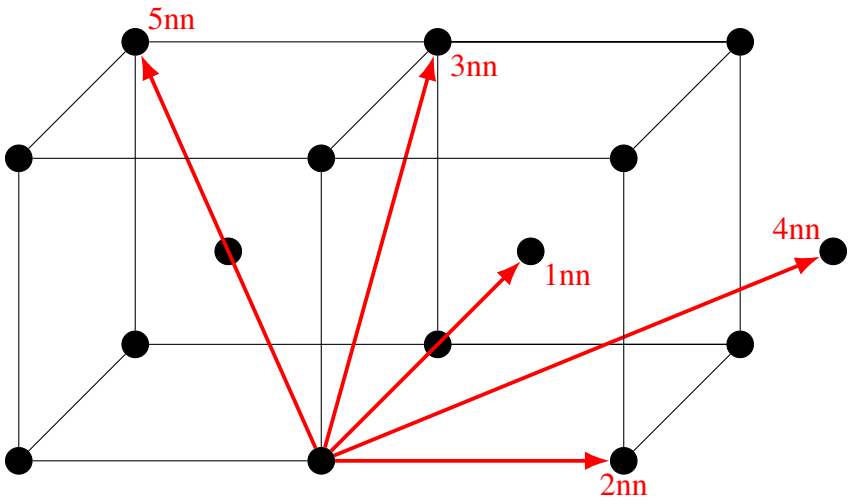


Fig. 5.2 Section of the BCC lattice with the first five nearest-neighbours (nn) marked with red arrows. A di-vacancy cluster with vacancies at either end of the 2nn arrow would be denoted 2nn di-vacancy. Vacancies predominantly migrate through $\frac{1}{2}\langle 111\rangle$ transitions, e.g moving along the 1nn arrow. The α-Fe lattice parameter is 2.8557Å. The first five nearest-neighbour separations are: 2.47, 2.86, 4.04, 4.74 & 4.95Å, respectively.

First, all uncoordinated Fe atoms (less than 8 neighbours) are identified using the LCR method. Second, the under-coordinated atoms positions were passed through the k-means clustering algorithm, converging onto a predetermined number of vacancies. The k-means algorithm is a well-established algorithm for clustering arbitrary dimensional data, for a full description see Peña, Lozano, and Larrañaga [69]. In brief, the k-means algorithm is a rapidly converging, iterative process to partition a set of points into k sets, which minimises the sum of the squared Euclidean norm between each point and the centroid of its set. To reduce the risk of relabelling vacancies between steps, the initial position of the k-means was taken as the final position at the previous frame.

5.2.1 Mono-vacancy

We begin with the simplest case, the mono-vacancy cluster, whose mechanism and diffusion barriers have been discussed in Section 4.1.2. The results of the vacancy diffusing for 20s are presented in Fig. 5.3. The motion of the vacancy during the simulation is entirely via the $\frac{1}{2}\langle 111\rangle$ mechanism with an activation energy of 0.534eV. The resulting mean diffusivity was $D = (2.3 \pm 0.1) \times 10^{-16}\mathrm{m}^2\,\mathrm{s}^{-1}$ calculated using Eq. (4.2). This corresponds to a diffusivity prefactor of $D_0 = (2.15 \pm 0.09) \times 10^{-7}\mathrm{m}^2\,\mathrm{s}^{-1}$ where D_0 is defined using the familiar Arrhenius

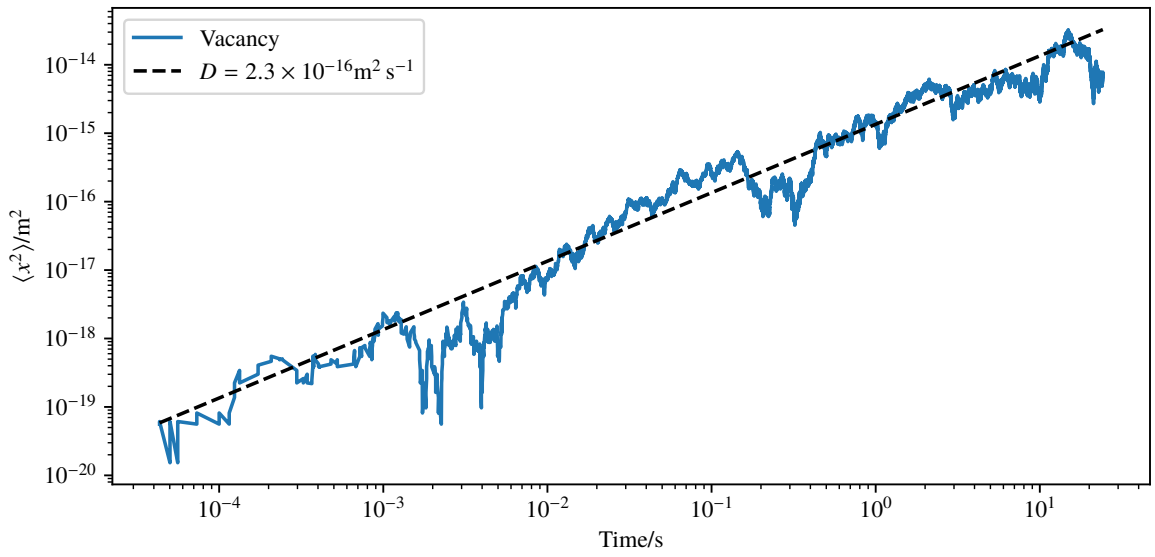


Fig. 5.3 Mono-vacancy cluster diffusing in the perfect 7^3 unit-cell lattice at 300K with an Arrhenius pre-factor of 5.12×10^{12}Hz over approximately 2 million Monte Carlo steps. This required approximately 80 hours of total run-time on a quad-core Intel® Xeon® at 3.10GHz. Dashed line is the expectation squared displacement using the mean simulated diffusivity of $D = 2.3 \times 10^{-16}$m^2 s^{-1} and the Einstein diffusion equation.

relationship:

$$D = D_0 e^{\frac{-E_a}{k_B T}} \tag{5.2}$$

This is close to the value reported by Messina et al. of $D_0 = 4.73 \times 10^{-7}$m^2 s^{-1} [55].

Off-lattice KMC offered no advantage here over traditional KMC, as there was only one dominant mechanism however, it is positive to see a method with the generality of MD achieving the simulation timescales of KMC, of the order of seconds. These longer timescales were accessible due to the topological caching; the periodic boundary conditions mean the supercell is in a topologically equivalent state after a $\frac{1}{2}\langle 111 \rangle$ vacancy migration. Therefore, SP searches were only required at the very beginning of the simulation, massively accelerating the simulation speed. As such, these results highlight the need for topology caching in any OLKMC hoping to reach traditional KMC timescales.

As is to be expected, due to the much higher activation energy, the diffusivity of vacancies is orders of magnitude slower than hydrogen. This presents an issue if we wish to simulate vacancies moving in the presence of hydrogen as almost all CPU time would be spent tracking the (less interesting) motion of the hydrogen atoms.

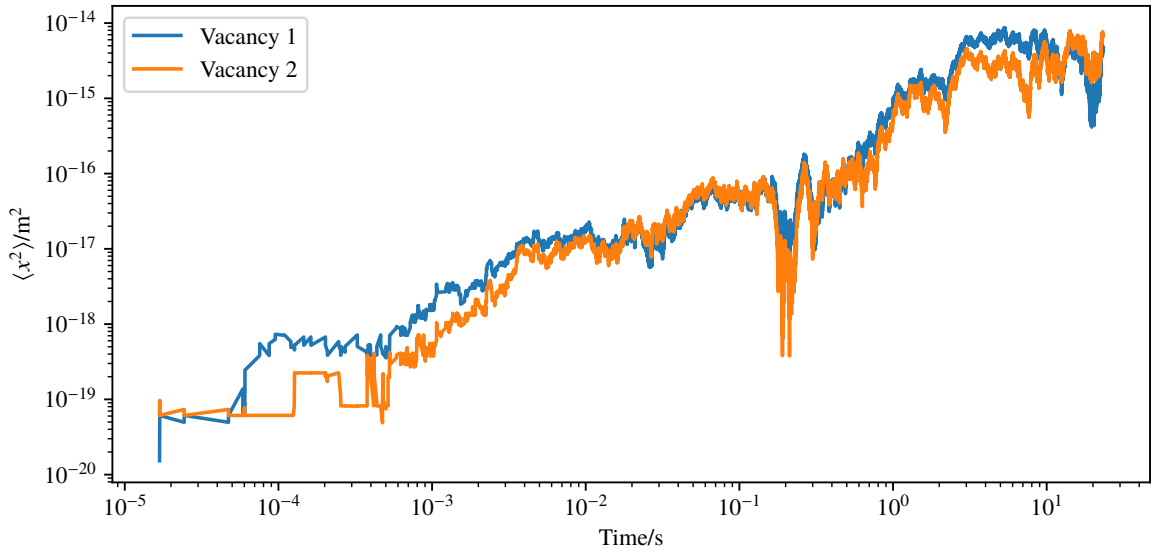


Fig. 5.4 Di-vacancy cluster diffusing in the perfect 7^3 unit-cell lattice at 300K with an Arrhenius pre-factor of 5.12×10^{12}Hz over approximately 4 million Monte Carlo steps. This required approximately 160 hours of total run-time on a quad-core Intel® Xeon® at 3.10GHz. As the two vacancies were not bound into a di-vacancy during the entire simulation, no expectation displacement is plotted.

5.2.2 Di-vacancy

The second experiment was the more interesting case of di-vacancy diffusing in the pure α-Fe lattice. The most stable configuration for a di-vacancy is the 2nn pair followed by the 1nn then 4nn orientations [20]; see Fig. 5.2 for a diagram of these configurations.

The simulation was started with the vacancies in the 2nn position. The displacements of the two vacancies are presented in Fig. 5.4. The diffusivity was again calculated using Eq. (4.2) however, only data points where the vacancies were close enough to be a considered a di-vacancy (5nn or closer) were used; the resulting diffusivity was $D = (1.3 \pm 0.1) \times 10^{-16} \mathrm{m}^2\,\mathrm{s}^{-1}$.

The predominant diffusion mechanism observed was the 2nn to 4nn transition, with an energy barrier of 0.531eV. This is counter intuitive as one may expect the transition to the lower energy 1nn di-vacancy to have a lower energy barrier however, the 2nn to 1nn transition has an energy barrier of 0.611eV thus occurring infrequently. Therefore, the resulting diffusivity prefactor is $D_0 = (1.1 \pm 0.1) \times 10^{-7} \mathrm{m}^2\,\mathrm{s}^{-1}$.

The diffusion barrier is very close to the diffusion barrier for the single vacancy, this may be an artefact of the EAM potential as *ab initio* studies typically predict an energy barrier 0.05–0.11eV [19, 28] lower for the di-vacancy. Nevertheless, the resulting diffusivity

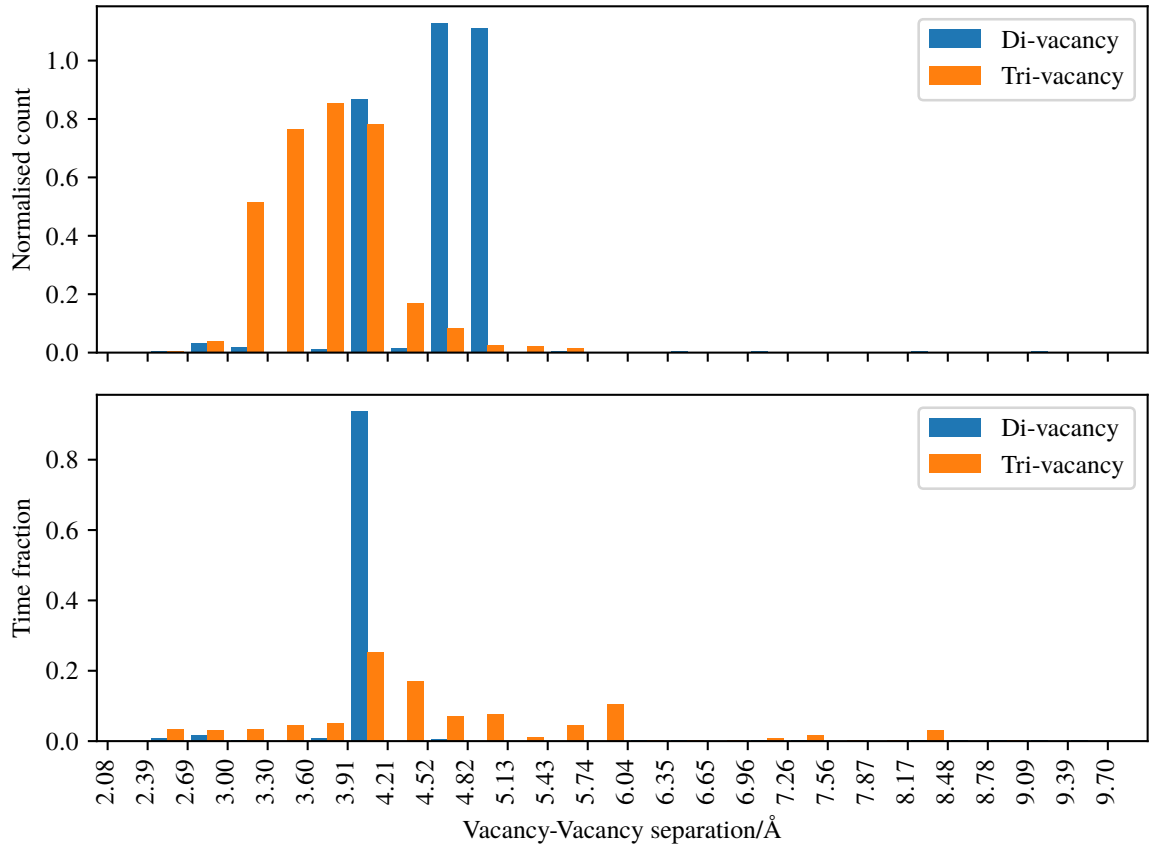


Fig. 5.5 Occurrence count and occupation time of the vacancy-vacancy (minimum image) separation during the di-vacancy and tri-vacancy simulation runs shown in Fig. 5.4 and Fig. 5.7, respectively. See the text for relation between measured vacancy-vacancy separation and nearest-neighbour positions in Fig. 5.2.

prefactor is in agreement with the literature, which ranges from $D_0 = 1.1 \times 10^{-7} \mathrm{m}^2\,\mathrm{s}^{-1}$ to $D_0 = 4.8 \times 10^{-7} \mathrm{m}^2\,\mathrm{s}^{-1}$ [20].

Figure 5.5 presents a histogram of the occurrence count and occupancy time of di-vacancy states defined by the separation between the vacancies. Due to using the k-means algorithm to locate the vacancies, the vacancy-vacancy separation does not directly correspond to the nearest-neighbour (nn) separations in Fig. 5.2: the small peak at 2.7Å is due to 1nn di-vacancies; the peak at 4Å corresponds to 2nn di-vacancies and the two peaks at 4.6Å and 4.9Å are due to 4nn di-vacancies and a transition state from 2nn to 4nn. We see that the most commonly occurring di-vacancies are 4nn and 2nn, while the count of 1nn di-vacancies is very low. This confirms that di-vacancies predominately diffuse by oscillating between the 4nn and 2nn states. Most strikingly in Fig. 5.5, the time spent in the 4nn state is many orders of magnitude lower, this is because the energy barrier for the transition from 4nn to 2nn is only 0.359eV thus, the

4nn di-vacancy is a relatively short-lived state. This is to be expected as 4nn di-vacancy is higher energy than the 1nn and 2nn vacancies and – obeying Boltzmann's statistics – must be less probable in equilibrium.

The reason for the reduced diffusivity compared to the single vacancy, even though the diffusion barriers are almost identical, is due to the di-vacancy first requiring a transition from 2nn to 4nn with almost the same probability as the mono-vacancy followed by the reverse 4nn to 2nn transition. Even though this second transition occurs almost instantly compared to the rate of the first it has a $\approx 50\%$ chance of returning the di-vacancy to the original position and a $\approx 50\%$ chance of moving the di-vacancy along the $\langle 111 \rangle$ direction. Therefore, the di-vacancy diffuses approximately twice as slowly.

Off-lattice KMC has correctly identified the counter-intuitive diffusion mechanisms of the di-vacancy. Although, this could have been captured with traditional KMC and careful DFT analysis, it arises naturally here without any special modifications or care. It is worth highlighting that simplified models for predicting energy barriers in KMC such as [20, 7]:

$$E_a = E_0 + \frac{E_f - E_i}{2} \tag{5.3}$$

with E_0 a constant and E_f, E_i the energy of the final and initial states, respectively, would have incorrectly caused a traditional KMC simulation to predict diffusion predominantly via oscillating between the 2nn and 4nn positions. Additionally, for larger and more complex systems, DFT run times may be out of reach for predicting energy barriers *ab initio*. This increases the risk of identifying incorrect mechanisms for traditional KMC. We are limited here in the accuracy of the measured diffusivity but, this is a limitation of the potentials used and will be overcome as Fe-H potentials are developed.

5.2.3 Tri-vacancy

The tri-vacancy was the largest vacancy cluster investigated. Figure 5.7 shows the results of three vacancies, initially started with a pair in the 2nn orientation and the third a few lattice vectors away from the pair, diffusing in the perfect lattice. During the first 10^{-3}s a few higher energy (0.4–0.5eV) barriers are crossed, resulting in the formation of a tri-vacancy. The lowest energy conformation of the three vacancies is sketched in Fig. 5.6a. This matches theoretical predictions [28] and is the conformation that minimises the distance between all three vacancies thus, allowing the lattice the largest volume to relax into. From this point the tri-vacancy diffuses through the lattice without disassociating for the remainder of the simulation.

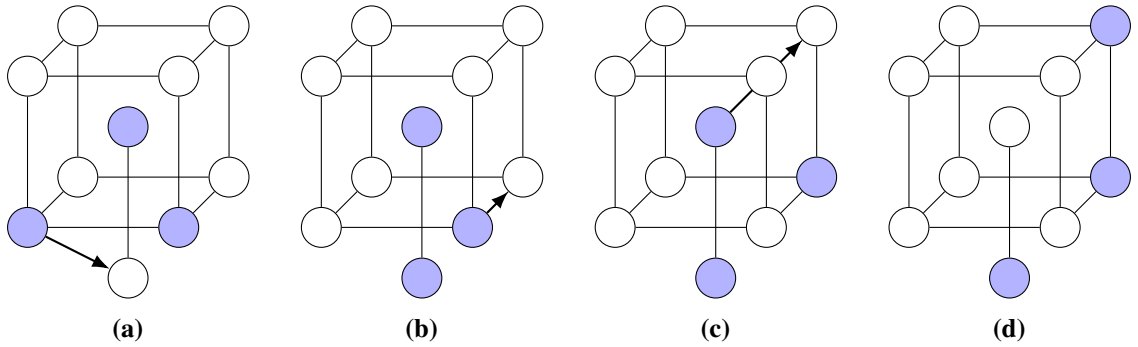


Fig. 5.6 The three most common mechanism encountered during the simulated tri-vacancy diffusion discovered with the dimer method during the run. Blue circles mark a vacant lattice site; white circle represent an occupied lattice site. Perturbations to the lattice have been omitted for clarity. Each pair of diagrams represent a beginning and end state (left to right) with the arrow in the initial state marking the path of the vacancy during the mechanism. The energy barriers for the three mechanisms are: Fig. 5.6a to Fig. 5.6b, $E_a = 0.301$eV; Fig. 5.6b to Fig. 5.6c, $E_a = 0.323$eV and Fig. 5.6c to Fig. 5.6d, $E_a = 0.466$eV.

The predominant diffusion mechanism is sketched in Fig. 5.6a & Fig. 5.6b and has a diffusion barrier of 0.301eV. This is significantly lower than the di-vacancy and mono-vacancy diffusion barriers and quantitatively matches DFT predictions of 0.34eV [28] but is often not captured in other adaptive kinetic Monte Carlo models [55].

The calculated diffusivity, $D = (1.8 \pm 0.2) \times 10^{-13}\text{m}^2\,\text{s}^{-1}$, is three orders of magnitude higher than di-vacancies and mono-vacancies. This corresponds to a diffusivity prefactor of $D_0 = (2.0 \pm 0.5) \times 10^{-8}\text{m}^2\,\text{s}^{-1}$ which is within the wide range of values presented in the literature of $0.2 \times 10^{-8}\text{m}^2\,\text{s}^{-1}$ to $11.2 \times 10^{-8}\text{m}^2\,\text{s}^{-1}$ [55, 19, 66]. It is surprising to find that larger vacancy clusters are more mobile than their smaller equivalents as typically larger objects diffuse slower. Other studies have shown this trend reverse (to the expectation) for larger vacancy clusters; the energy barrier for five and six vacancy clusters correspondingly increasing to 0.88eV and 1.11eV, respectively [19].

Additionally, the tri-vacancy occasionally undergoes a mechanism with a slightly higher energy barrier of 0.323eV, detailed in Fig. 5.6b & Fig. 5.6c. This mechanism is absent from other KMC studies, tending to focus on only the lowest energy barrier mechanisms, and could explain why the measured diffusivity prefactor is toward the higher end of the range reported in the literature.

The steps visible in Fig. 5.7 are due to a final, rarely occurring, mechanism with an energy barrier of 0.466eV. This mechanism is detailed in Fig. 5.6c & Fig. 5.6d and is the only mechanism observed to move the three vacancies out of their lowest energy configuration. When this occurs the tri-vacancy becomes pinned, unable to diffuse via the fast conservative

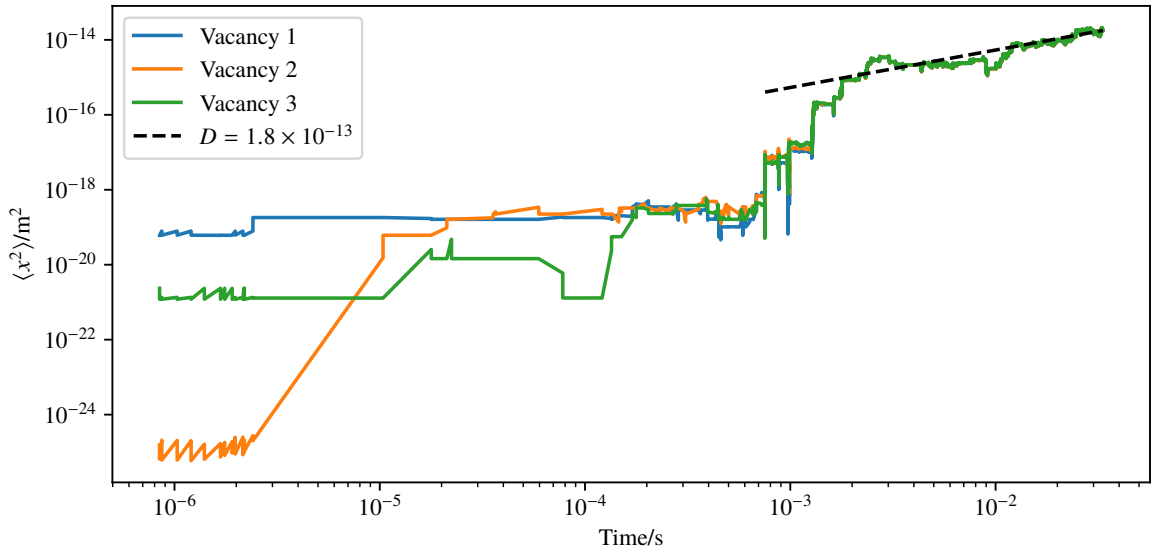


Fig. 5.7 Tri-vacancy cluster diffusing in the perfect 7^3 unit-cell lattice at 300K with an Arrhenius pre-factor of 5.12×10^{12}Hz over approximately 3 million Monte Carlo steps. This required approximately 90 hours of total run-time on a quad-core Intel® Xeon® at 3.10GHz. Dashed line is the expectation squared displacement using the mean simulated diffusivity of $D = 1.8 \times 10^{-16} \mathrm{m}^2\,\mathrm{s}^{-1}$ and the Einstein diffusion equation over the range where the three vacancies were bound into a tri-vacancy.

0.301eV or 0.323eV mechanisms, until the lowest energy configuration reforms. This causes the steps in Fig. 5.7 while the tri-vacancy remains stationary.

Figure 5.5 additionally contains the occurrence count and occupancy time of the tri-vacancy states, as a consequence of the increased number of under-coordinated atoms the k-means algorithm struggles to cleanly identify individual vacancy locations. This, unlike the di-vacancy, manifests as a broadening of the possible vacancy-vacancy separations to occupy a more continuous range of values. The distribution of separations centred around 3.8Å in the occurrence count histogram correspond to the separations in the lowest energy tri-vacancy state. Similar to the di-vacancy case these are under-represented in the time fraction histogram due to the pinned state (which has larger vacancy-vacancy separations) occupying a larger fraction of the simulated time each time it occurs. The k-means algorithm was less capable of distinguishing the vacancies position as the size of the vacancy cluster increased, this would need to be improved to investigate larger vacancy clusters.

Once again OLKMC has correctly identified the primary diffusion mechanism, as well as correctly predicting a significantly lower energy barrier than single and di-vacancies. Furthermore, OLKMC has identified two additional mechanisms not typically accounted for in traditional KMC experiments, one of which helps to explain the elevated diffusivity compared

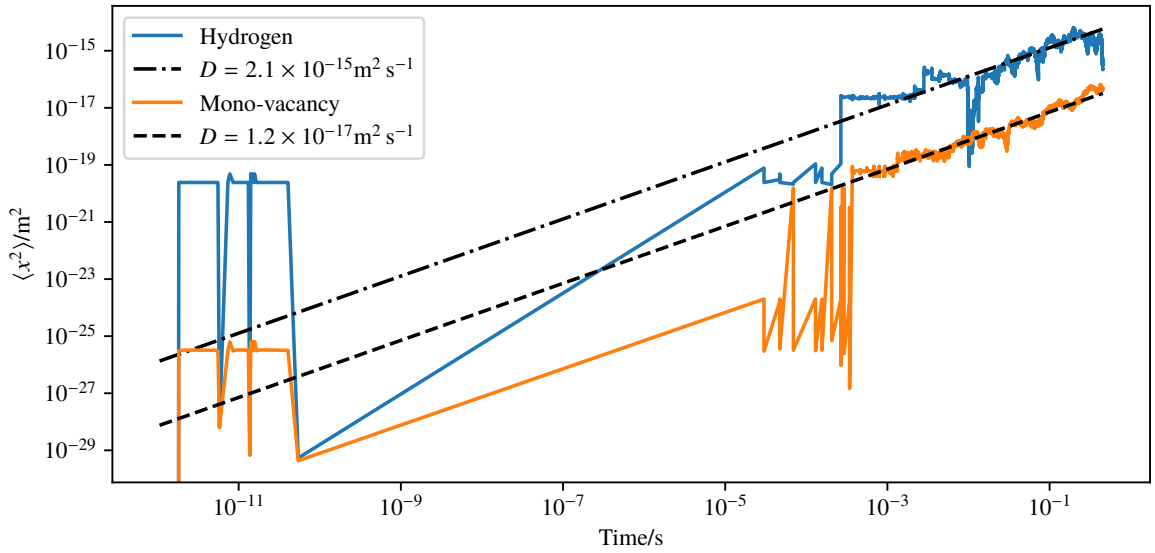


Fig. 5.8 Mono-vacancy cluster and a single H atom diffusing in the the perfect 7^3 unit-cell lattice at 300K with an Arrhenius pre-factor of 5.12×10^{12}Hz over approximately 1 million Monte Carlo steps. This required approximately 75 hours of total run-time on a quad-core Intel® Xeon® at 3.10GHz. The KMC algorithm was accelerated with the superbasin dynamics algorithm, detailed in Section 3.6. Dashed lines are the expectation squared displacement using the mean simulated diffusivities and the Einstein diffusion equation.

to the literature. The absence of these mechanisms in the literature highlights the difficulty in predicting mechanisms using intuition and DFT + NEB methods. However, we must be wary of the small probability these mechanisms and/or their energy barriers are an artefact of the EAM potential; verification with DFT would be a sensible future extension. Finally, if we were to continue to larger vacancy clusters, is reasonable to assume that the mechanisms would follow the pattern of diversifying, rapidly reaching a point outside the reach of traditional KMC and DFT.

5.3 Vacancy-hydrogen complexes

To simulate the interaction between vacancies and hydrogen into the long timescale and avoid the low barrier problem, we use the superbasin KMC algorithm discussed in Section 3.6. We repeated the simulations from Fig. 5.3 (mono-vacancy) and Fig. 5.4 (di-vacancy) with a single hydrogen atom added to the supercell, the results are presented in Fig. 5.8 and Fig. 5.9, respectively. In both simulations the H atom was initialised in a trapped state near the centre of the defect.

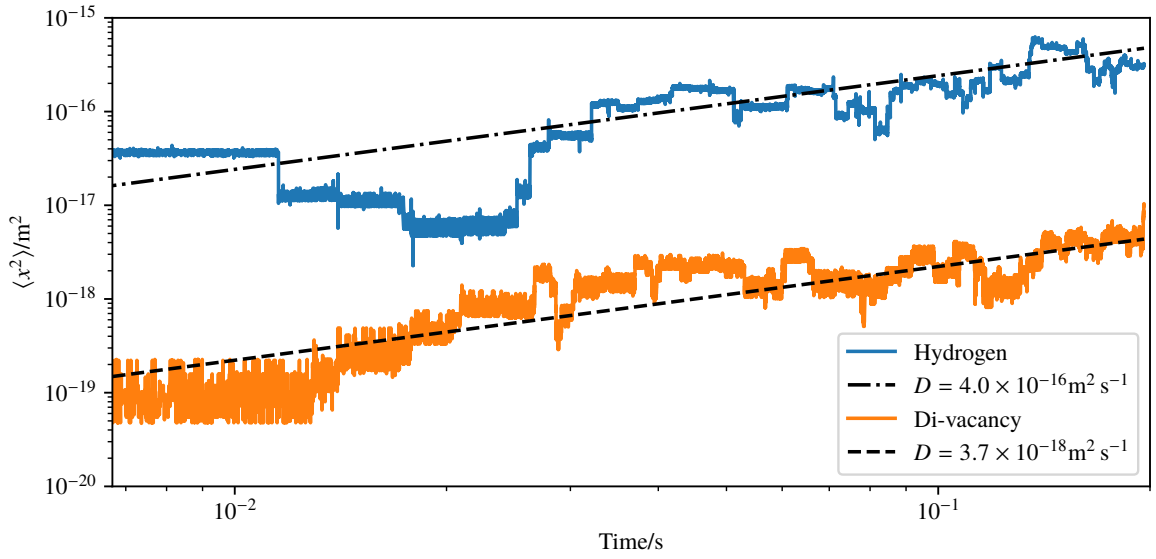


Fig. 5.9 Di-vacancy cluster and a single H atom diffusing in the the perfect 7^3 unit-cell lattice using the same conditions as Fig. 5.8 over approximately 1 million Monte Carlo steps. This required approximately 105 hours of total run-time on a quad-core Intel® Xeon® at 3.10GHz. Dashed lines are the expectation squared displacement using the mean simulated diffusivities and the Einstein diffusion equation. The first several hundred frames (in which the simulation clock increased by less than 10^{-10}s) have been omitted for clarity.

Figure 5.8 highlights the huge range of timescales accessible with this method. Initially, the H atom explores the 6 trapping states and 24 low-barrier transitions inside the vacancy then crosses several high energy barriers, attempting to escape the superbasin, but continually falls back in. Eventually, the H atom escapes the vacancy and diffuses through the lattice until it is trapped at another vacancy and the cycle repeats; this causes the steps in Fig. 5.8. Occasionally the escape mechanism includes a $\frac{1}{2}\langle 111 \rangle$ vacancy migration, causing the vacancy to diffuse as well as the H atom. The vacancy and hydrogen diffusivities calculated as in Section 5.2.1 are presented in Table 5.2.

Similar events occur in Fig. 5.9 however, in the 2nn di-vacancy superbasin there exists 14 trapping states and many low-barrier transitions. Furthermore, the 1nn and 4nn di-vacancies have corresponding superbasins associated with them; the increased size and diversity of the superbasins causes the even larger stepping in Fig. 5.9. Again, the vacancy and hydrogen diffusivities calculated as in Section 5.2.2 are presented in Table 5.2.

From Table 5.2 we see the presence of the hydrogen has reduced the diffusivity of both the mono and di-vacancy, both have fallen by approximately a factor of 25. To understand this,

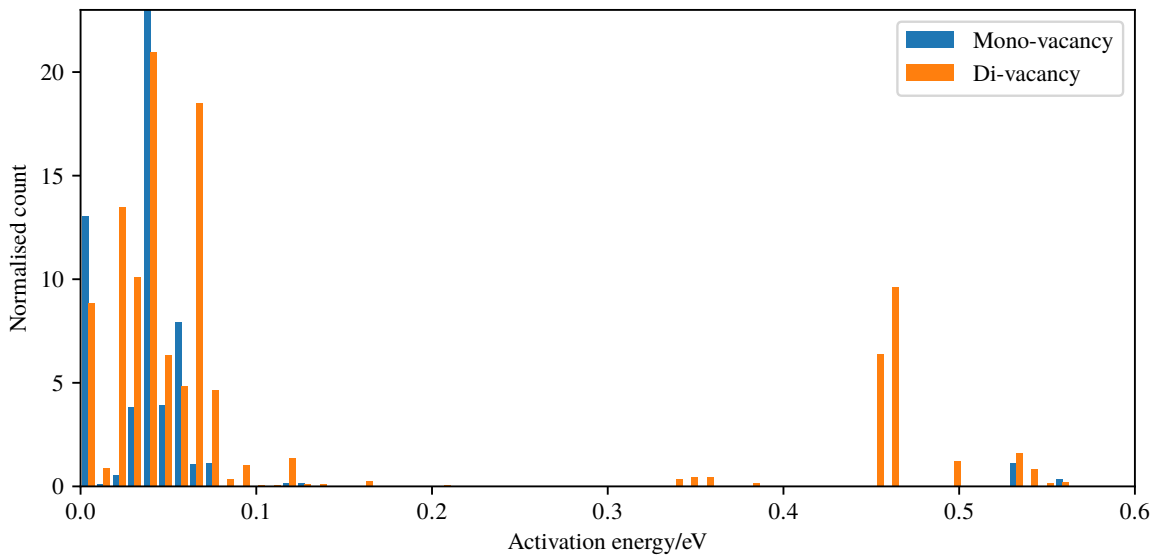


Fig. 5.10 Histogram of energy barriers encountered during the simulated vacancy-hydrogen complex diffusion in Fig. 5.8 and Fig. 5.9. The graph has been truncated along the y-axis such that the lower occurrence bins are visible, the truncated bin at 0.038eV extends to a normalised count of 80.

we analyse the distribution of energy barriers encountered during the simulation presented in Fig. 5.10.

For the mono-vacancy the highest peak (truncated) in Fig. 5.10 is the 0.038eV bin, corresponding to hydrogen diffusing in the lattice. The other significant peaks of interest are: a sharp peak in the 0.002eV bin corresponding to a low barrier flicker between two H atom positions just outside the vacancy; the 0.062eV bin corresponding to the barrier between trapping states; the peak at 0.53eV is the detrapping barrier as the H atom moves from inside the vacancy to an adjacent tetrahedral site and finally, the peak at 0.56eV is the barrier for a $\frac{1}{2}\langle 111\rangle$ vacancy migration. The remaining visible peaks correspond to hydrogen tetrahedral $\rightarrow$ tetrahedral transitions with modified energy barriers near the vacancy.

For the di-vacancy the peaks in Fig. 5.10 correspond to similar mechanisms however, there is now a much greater diversity of mechanisms both inside and immediately outside the superbasin. The new peaks of interest are: the two peaks at 0.45 and 0.46eV, which are the detrapping barriers as the H atom escapes the 2nn di-vacancy, while the peaks from 0.54–0.56eV are detrapping barriers from other di-vacancy configurations; the three peaks between 0.34–0.36eV are 4nn $\rightarrow$ 2nn di-vacancy migrations when the moving vacancy does not contain the H atom; the tiny peak at 0.38eV is the 4nn $\rightarrow$ 2nn di-vacancy migration when the moving vacancy does contain the H atom and finally, the peak at 0.53eV is the 2nn $\rightarrow$ 4nn di-vacancy migration. In all 4nn $\rightarrow$ 2nn migrations the H atom remains trapped in the vacancy that did not move.

Table 5.2 Summary of defect and hydrogen diffusivities during the simulations in Fig. 5.8 and Fig. 5.9. Additionally. the theoretical effective hydrogen diffusivities calculated using Oriani's trapping theory [64] and the data in Table 5.3 are presented for comparison.

Defect	Diffusivity, $D/\mathrm{m}^2\,\mathrm{s}^{-1}$			
	Defect without H	Defect with H	D_{eff} of H	$D_{\mathrm{eff}}^{\mathrm{Or.}}$ of H
Mono-vacancy	2.3×10^{-16}	1.2×10^{-17}	2.1×10^{-15}	5.4×10^{-15}
Di-vacancy	1.3×10^{-16}	3.7×10^{-18}	4.0×10^{-16}	4.9×10^{-16}

We can now clearly see the reason for the fall in diffusivity of the mono-vacancy is the increase in the energy barrier of the $\frac{1}{2}\langle 111\rangle$ vacancy migration. The case for the di-vacany is more complex as the 2nn $\rightarrow$ 4nn migration barrier remains the same as in the hydrogen free simulation. Instead, the increase in the 4nn $\rightarrow$ 2nn barrier, being only in the case when the migrating vacancy contains the hydrogen, biases the vacancy to reform into its original 2nn configuration rather than forming a 2nn di-vacancy displaced by $\frac{1}{2}\langle 111\rangle$. This in turn reduces the diffusivity of the di-vacancy.

The most extreme results in Fig. 5.9 are the effective hydrogen diffusivities which have fallen by seven orders of magnitude, compared to the calculated diffusivity in the perfect lattice of $7.4 \times 10^{-9}\mathrm{m}^2\,\mathrm{s}^{-1}$. This may seem like an unphysical reduction however, we must remember the defect concentration is exceptionally high, around $200\mathrm{mol}\,\mathrm{m}^{-3}$. We can compare our results with Oriani's trapping theory which states [64]:

$$\frac{\theta_x}{1-\theta_x} = \frac{\theta_L}{1-\theta_L} e^{-\frac{E_B}{K_B T}} \tag{5.4}$$

where E_B is the trap binding energy and θ_x, θ_L are the fractional occupancy of available trap and regular lattice hydrogen sites, respectively. In this case we assume the trap binding energy is equivalent to the hydrogen detrapping barrier, this is true within a few multiples of the lattice hydrogen diffusion barrier. Furthermore, the effective diffusivity of hydrogen in the presence of traps, $D_{\mathrm{eff}}^{\mathrm{Or.}}$, is predicted to obey [64]:

$$D_{\mathrm{eff}}^{\mathrm{Or.}} = D \frac{N_L \theta_L}{N_L \theta_L + N_x \theta_x (1-\theta_x)} \tag{5.5}$$

with N_x, N_L the number of trap and regular lattice sites per unit volume and D the diffusivity of hydrogen in the perfect lattice. To obtain the average fractional occupancy of trap and regular lattice sites over the course of the entire simulation we partition the frames into two sets: frames

Table 5.3 Summary of the time the H atom in Fig. 5.8 and Fig. 5.9 spent in trap/lattice sites during the simulations, used to calculate the effective diffusivity alongside Oriani's trapping theory [64]. Additionally, the range of detrapping barriers from Fig. 5.10 are provided for comparison.

Defect	t_x/s	t_L/s	E_B/eV	$E_B^{\text{Or.}}$/eV
Mono-vacancy	4.535×10^{-1}	2.749×10^{-7}	0.53	0.54
Di-vacancy	1.952×10^{-1}	1.210×10^{-8}	0.45–0.56	0.58

with the defect-H separation less than 2 or 5Å for the mono and di-vacancy respectively and *vice versa*. The total time spent in each partition, t_x and t_L , are presented in Table 5.3 and relate directly to θ and N via[1]:

$$\frac{t_x}{t_x + t_L} = N_x \theta_x \quad \text{and} \quad \frac{t_L}{t_x + t_L} = N_L \theta_L \tag{5.6}$$

Using this we can compute the expectation trapping energy $E_B^{\text{Or.}}$ presented in Table 5.3, as well as the theoretical effective diffusivity $D_{\text{eff}}^{\text{Or.}}$ in Table 5.2.

The agreement between Oriani's theory and our simulated results in Table 5.2 is excellent. For both crystal defects Oriani's estimate is marginally higher but, given the change between D_{eff} and D is orders of magnitude, this is to be expected. The agreement for the detrapping barrier/trap binding energy in Table 5.3 is also excellent for the mono-vacancy. However, for the di-vacany $E_B^{\text{Or.}}$ is significantly higher. This is probably because Oriani's theory is a model for homogeneous trapping however, during the simulation the di-vacancy transforms between the 1nn, 2nn and 4nn states all of which have different trapping energies. Therefore, the di-vacancy simulation is a heterogeneous trapping problem and we see Oriani's theory beginning to break down. Furthermore, Oriani's theory is based on thermodynamic equilibrium which may not have been reached in these simulations. This is reason to be careful when applying Oriani's theory to real steels which have a huge variety of crystal defects thus are clearly heterogeneous trapping problems.

The small increase in the mono-vacancy migration energy barrier and corresponding decrease in diffusivity mirrors Benediktsson et al. [9] results for hydrogen in aluminium. If higher hydrogen loading of vacancies causes them to become pinned, as Benediktsson et al. observed in aluminium, then this could act as a possible mechanism for HE. Sessile vacancies could act as nucleation sites for extended crystal defects or nucleation sites for nanovoids [9]. The former could imply an increased localised plasticity and contribute towards the HELP

[1]For simplicity, as there is only a single H atom, the dimensional-correcting number density of hydrogen has been omitted.

mechanism, should additional dislocations be generated. The latter, especially in combination with the observed superabundance of vacancies in hydrogen charged steels [78, 51, 61], helps to explain nanovoid formation and coalescence and contributes towards the HESIV mechanism.

In summary, Figure 5.10 has again highlighted the increase in mechanism diversity as the crystal defect structure becomes more complex. Attempting to model this with traditional KMC would require simplification, of the possible mechanisms and states, likely resulting in a failure to capture the full systems behaviour. Furthermore, the timescales reached in these simulations are out of reach of current MD, making superbasin-accelerated OLKMC the best and most general alternative.

Chapter 6

Conclusions & Future work

6.1 Conclusions

We have built, from the ground up, an implementation of the off-lattice kinetic Monte Carlo model, that is capable of simulating the Fe-H system. Many optimisations and enhancements to the OLKMC method, which have been developed over the last 20 years, have been incorporated and extended. We have validated our model, simulating diffusion of vacancies and hydrogen in the α-Fe lattice, as well as probed more involved vacancy-hydrogen complexes. Our key results include:

- We demonstrated OLKMC can reproduce the simulated diffusivity of single and multiple H atoms, with a final calculated value of $D = (7.4 \pm 0.7) \times 10^{-9} \mathrm{m}^2\,\mathrm{s}^{-1}$ at 300K.
- We used OLKMC to simulate the kinetics of: single vacancy, di-vacancy and tri-vacancy clusters. Alongside, calculating their diffusivity prefactors and finding them within the range of the literature values. These experiments had no pre-conceived notion of the diffusion pathways the defects follow, yet were able to accurately predict the predominant diffusion mechanisms for all cases. This includes the counter-intuitive case of the di-vacancy, primarily migrating by oscillating between 2nn and 4nn configurations. Furthermore, the sharp-drop in the energy barrier to migration for the tri-vacancy, compared to the di-vacancy, was accurately predicted, falling from 0.53eV to 0.30eV, respectively.
- Simulating hydrogen diffusion, including trapping and detrapping, in the presence of a vacancy and a di-vacancy, into timescales of the order of seconds. The effective diffusivity of hydrogen in these environments was shown to be orders of magnitude less than the

perfect lattice, a strong trapping effect. Comparisons to Oriani's trapping theory [64] showed excellent agreement, validating the method.

- Simulating the effect of a single H atom on the diffusivity of a vacancy and a di-vacancy, into timescales of the order of seconds. In both cases hydrogen reduced their diffusivity by a factor of approximately 25. Furthermore, we were able to probe the exact hydrogen-induced mechanisms causing this reduction, which were different for each defect. These results were linked back to HE, potentially contributing to the HESIV mechanism.

Some of these results, particularly the effects of hydrogen on the kinetics of vacancies, are not possible to achieve with any other current method, while capturing the same diversity of mechanisms and reaching equivalent timescales. This is extremely promising for the future of modelling defect-H interactions. Moreover, it should be possible to extend OLKMC to simulate extended crystal defects such as dislocations and grain boundaries. This should give more insights into the HELP and other mechanisms of HE.

Importantly, our results have been made possible by our contributions to the OLKMC methodology, specifically:

- Bringing together enhancements to the dimer method, from across the literature [37, 39, 45, 88], as well as refining the line-search method used in conjunction with L-BFGS optimiser.

- Expanding El-Mellouhi, Mousseau, and Lewis's topological caching method to incorporate more information about the local geometry, allowing us to perform topological caching of saddle-points in the Fe-H system. This is a more difficult system to implement topological caching, as the small size of hydrogen atoms, induce only small perturbations to the surrounding Fe atoms positions.

- Re-phrasing the mean rate method, by additionally partitioning mechanisms into transient and absorbing categories, such that it can be efficiently applied to an OLKMC which uses topological SP caching, allowing us to overcome the low-barrier problem while simulating V-H complexes.

The limiting factor underlying our work was the semi-empirical potentials however, the correctness of our method is independent from the underlying potentials. Hence, the accuracy of OLKMC results will only improve, as better Fe-H potentials are developed. Nevertheless, our results illustrate, even with today's potentials, OLKMC can be used to push forward our understanding of the defect-H interactions, that underpin hydrogen embrittlement.

6.2 Future work

Many enhancements to the OLKMC method and implementation, which could significantly reduce run times and allow the simulation of many more atoms or improve the accuracy of results, have yet to be fully explored:

- We already make (and verify) the assumption that mechanisms are sufficiently local to be captured within r_{topo} of the maximally displaced atom. It would be a straightforward extension to define a second (larger) radius that limits the atoms that can take part in each SP search. This would make SP searches scale as a constant with the number of atoms in the simulation, instead of linearly.

- If the initial local environment/topology has any symmetries, it is likely any mechanism from this state has corresponding mechanisms, related by those same symmetries. These symmetries (rotations, reflections) are already captured in the graph representation (by construction) and during topological caching, we already obtain the automorphism group of the graph. These automorphisms could be used to calculate the transformations required to map the displacement vectors for each mechanism onto its corresponding symmetrically-related mechanisms. This would reduce the number of SP searches required for highly symmetric topologies.

- As discussed in Section 1.2.1, the Arrhenius prefactor could be calculated for each mechanism, this would improve the accuracy of the model. Furthermore, if the normal-modes of the initial state and saddle-points could be associated with their topological keys, much of the introduced computational overhead could be avoided by caching normal-modes.

- As we create a catalogue of encountered topologies, it becomes possible to identify those that occur frequently. This could be used to run SP searches with better force fields (e.g. *ab initio* DFT forces) and refine energy barriers for the most important (frequent) topologies on-the-fly.

- Currently SP searches incorporate no knowledge of other SP searches performed from the same basin, they are completely independent. This results in many SP searches converging to the same saddle points, wasting compute cycles. One possibility could be searching for multiple saddle-points simultaneously, akin to a particle swarm optimiser [44], with a repulsive component between dimers.

- Due to the complete independence between SP searches we currently trivially-employ parallelism via the shared-memory paradigm. To scale onto multi-node supercomputers the distributed-memory programming model must be adopted. However, the limitation of not being able to scale onto more CPU's than SP searches launched per topology, would still exist. This could be overcome with multi-level parallelism, distributing the work of a single SP search over multiple CPU's on single node, this would require more synchronisation primitives.

- Currently the initial displacements for the SP searches are random, others have demonstrated improved results using more complex techniques, such as choosing points equidistributed on the surface of a hypersphere [34]. This, alongside exploring possible pre-conditioners [57, 10], are strong avenues for investigation.

- Finally, others have demonstrated GPU accelerated EAM force calculations for MD [87]. This could be used during SP searches (or perhaps SP searches could be conducted entirely on the GPU) to fully utilise available hardware.

Due to the off-lattice nature of the OLKMC framework, it extends naturally to strained systems. Therefore, exploring the kinetics of strained systems (relevant to HE), historically dominated by MD simulations, is the direction we shall take in the immediate future. Specifically, the effect of hydrogen on the strained lattice surrounding a crack tip and vacancies interacting with grain-boundaries strain-fields, are key targets.

In conclusion, the ceiling for OLKMC modelling is high above the current best. Long-timescale simulations of defect-H systems, critical to understanding hydrogen embrittlement, are almost within reach.

References

[1] G J Ackland et al. “Development of an interatomic potential for phosphorus impurities in -iron”. In: *Journal of Physics: Condensed Matter* 16.27 (June 2004), S2629–S2642. DOI: 10.1088/0953-8984/16/27/003.

[2] J. H. Ahlberg. *The Theory of Splines and Their Applications : Mathematics in Science and Engineering: A Series of Monographs and Textbooks, Vol. 38*. Saint Louis: Elsevier Science, 2016. ISBN: 9781483222950.

[3] M. P. Allen. *Computer simulation of liquids*. Oxford England New York: Clarendon Press Oxford University Press, 1987. ISBN: 0-19-855375-7.

[4] Joshua A. Anderson, Chris D. Lorenz, and A. Travesset. “General purpose molecular dynamics simulations fully implemented on graphics processing units”. In: *Journal of Computational Physics* 227.10 (May 2008), pp. 5342–5359. DOI: 10.1016/j.jcp.2008.01.047.

[5] George Arfken. *Mathematical Methods for Physicists*. Third. San Diego: Academic Press, Inc., 1985.

[6] O. Barrera et al. “Understanding and mitigating hydrogen embrittlement of steels: a review of experimental, modelling and design progress from atomistic to continuum”. In: *Journal of Materials Science* 53.9 (Feb. 2018), pp. 6251–6290. DOI: 10.1007/s10853-017-1978-5.

[7] C. S. Becquart and C. Domain. “Introducing chemistry in atomistic kinetic Monte Carlo simulations of Fe alloys under irradiation”. In: *physica status solidi (b)* 247.1 (Jan. 2010), pp. 9–22. DOI: 10.1002/pssb.200945251.

[8] Laurent Karim Béland et al. “Kinetic activation-relaxation technique”. In: *Physical Review E* 84.4 (Oct. 2011). DOI: 10.1103/physreve.84.046704.

[9] Magnús Þ Benediktsson et al. “Stability and mobility of vacancy–H complexes in Al”. In: *Journal of Physics: Condensed Matter* 25.37 (Aug. 2013), p. 375401. DOI: 10.1088/0953-8984/25/37/375401.

[10] Michele Benzi and Andrew J. Wathen. “Some Preconditioning Techniques for Saddle Point Problems”. In: *Mathematics in Industry*. Springer Berlin Heidelberg, 2008, pp. 195–211. DOI: 10.1007/978-3-540-78841-6_10.

[11] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN: 0387310738.

[12] D. Bombač et al. “Theoretical evaluation of the role of crystal defects on local equilibrium and effective diffusivity of hydrogen in iron”. In: *Materials Science and Technology* 33.13 (Apr. 2017), pp. 1505–1514. DOI: 10.1080/02670836.2017.1310417.

[13] A.B. Bortz, M.H. Kalos, and J.L. Lebowitz. "A new algorithm for Monte Carlo simulation of Ising spin systems". In: *Journal of Computational Physics* 17.1 (Jan. 1975), pp. 10–18. DOI: 10.1016/0021-9991(75)90060-1.

[14] Charles J. Cerjan and William H. Miller. "On finding transition states". In: *The Journal of Chemical Physics* 75.6 (Sept. 1981), pp. 2800–2806. DOI: 10.1063/1.442352.

[15] Mykyta V. Chubynsky et al. "Exploiting memory in event-based simulations". In: *Journal of Non-Crystalline Solids* 352.42-49 (Nov. 2006), pp. 4424–4429. DOI: 10.1016/j.jnoncrysol.2006.01.117.

[16] C. P. Chui et al. "Molecular Dynamics Simulation of Iron — A Review". In: *SPIN* 05.04 (Dec. 2015), p. 1540007. DOI: 10.1142/s201032471540007x.

[17] Murray S. Daw and M. I. Baskes. "Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals". In: *Physical Review B* 29.12 (June 1984), pp. 6443–6453. DOI: 10.1103/physrevb.29.6443.

[18] Murray S. Daw and M. I. Baskes. "Semiempirical, Quantum Mechanical Calculation of Hydrogen Embrittlement in Metals". In: *Physical Review Letters* 50.17 (Apr. 1983), pp. 1285–1288. DOI: 10.1103/physrevlett.50.1285.

[19] F.G. Djurabekova et al. "Stability and mobility of small vacancy and copper-vacancy clusters in bcc-Fe: An atomistic kinetic Monte Carlo study". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 255.1 (Feb. 2007), pp. 47–51. DOI: 10.1016/j.nimb.2006.11.009.

[20] F. Djurabekova et al. "Kinetics versus thermodynamics in materials modeling: The case of the di-vacancy in iron". In: *Philosophical Magazine* 90.19 (July 2010), pp. 2585–2595. DOI: 10.1080/14786431003662515.

[21] Juliana Duncan et al. "Biased gradient squared descent saddle point finding method". In: *The Journal of Chemical Physics* 140.19 (May 2014), p. 194102. DOI: 10.1063/1.4875477.

[22] Luis Elcoro and Jesús Etxebarria. "Common misconceptions about the dynamical theory of crystal lattices: Cauchy relations, lattice potentials and infinite crystals". In: *European Journal of Physics* 32.1 (Nov. 2010), pp. 25–35. DOI: 10.1088/0143-0807/32/1/003.

[23] Peter J. Feibelman. "Diffusion path for an Al adatom on Al(001)". In: *Physical Review Letters* 65.6 (Aug. 1990), pp. 729–732. DOI: 10.1103/physrevlett.65.729.

[24] Kristen A. Fichthorn and Yangzheng Lin. "A local superbasin kinetic Monte Carlo method". In: *The Journal of Chemical Physics* 138.16 (Apr. 2013), p. 164104. DOI: 10.1063/1.4801869.

[25] M. W. Finnis and J. E. Sinclair. "A simple empirical N-body potential for transition metals". In: *Philosophical Magazine A* 50.1 (July 1984), pp. 45–55. DOI: 10.1080/01418618408244210.

[26] S. M. Foiles, M. I. Baskes, and M. S. Daw. "Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys". In: *Physical Review B* 33.12 (June 1986), pp. 7983–7991. DOI: 10.1103/physrevb.33.7983.

[27] Scott Fortin. "The Graph Isomorphism Problem". In: (1996). DOI: 10.7939/R3SX64C5K.

[28] Chu-Chun Fu et al. "Multiscale modelling of defect kinetics in irradiated iron". In: *Nature Materials* 4.1 (Dec. 2004), pp. 68–74. DOI: 10.1038/nmat1286.

[29] Yuh Fukai. *The metal-hydrogen system: basic bulk properties*. Berlin New York: Springer, 2005. ISBN: 978-3-540-28883-1.

[30] Jean Charles Gilbert and Jorge Nocedal. "Global Convergence Properties of Conjugate Gradient Methods for Optimization". In: *SIAM Journal on Optimization* 2.1 (Feb. 1992), pp. 21–42. DOI: 10.1137/0802003.

[31] William F. Gilreath. *Hash sort: A linear time complexity multiple-dimensional sort algorithm*. 2004. arXiv: cs/0408040 [cs.DS].

[32] Fred Glover and Manuel Laguna. "Tabu Search". In: *Handbook of Combinatorial Optimization*. Springer US, 1998, pp. 2093–2229. DOI: 10.1007/978-1-4613-0303-9_33.

[33] Pedro Gonnet. "A simple algorithm to accelerate the computation of non-bonded interactions in cell-based molecular dynamics simulations". In: *Journal of Computational Chemistry* 28.2 (2006), pp. 570–573. DOI: 10.1002/jcc.20563.

[34] Manuel Plasencia Gutiérrez, Carlos Argáez, and Hannes Jónsson. "Improved Minimum Mode Following Method for Finding First Order Saddle Points". In: *Journal of Chemical Theory and Computation* 13.1 (Dec. 2016), pp. 125–134. DOI: 10.1021/acs.jctc.5b01216.

[35] Hideki Hagi, Yasunori Hayashi, and Namio Ohtani. "Diffusion Coefficient of Hydrogen in Pure Iron between 230 and 300 K". In: *Transactions of the Japan Institute of Metals* 20.7 (1979), pp. 349–357. DOI: 10.2320/matertrans1960.20.349.

[36] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (Apr. 1970), pp. 97–109. DOI: 10.1093/biomet/57.1.97.

[37] Graeme Henkelman and Hannes Jónsson. "A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives". In: *The Journal of Chemical Physics* 111.15 (Oct. 1999), pp. 7010–7022. DOI: 10.1063/1.480097.

[38] Graeme Henkelman and Hannes Jónsson. "Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table". In: *The Journal of Chemical Physics* 115.21 (Dec. 2001), pp. 9657–9666. DOI: 10.1063/1.1415500.

[39] Andreas Heyden, Alexis T. Bell, and Frerich J. Keil. "Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method". In: *The Journal of Chemical Physics* 123.22 (Dec. 2005), p. 224101. DOI: 10.1063/1.2104507.

[40] David Hubert. "Ueber die stetige Abbildung einer Linie auf ein Flächenstück". ger. In: *Mathematische Annalen* 38 (1891), pp. 459–460.

[41] Yasunobu Imamura et al. "Fast Hilbert Sort Algorithm Without Using Hilbert Indices". In: *Similarity Search and Applications*. Springer International Publishing, 2016, pp. 259–267. DOI: 10.1007/978-3-319-46759-7_20.

[42] D. E. Jiang and Emily A. Carter. "Diffusion of interstitial hydrogen into and through bcc Fe from first principles". In: *Physical Review B* 70.6 (Aug. 2004). DOI: 10.1103/physrevb.70.064102.

[43] Hannes Jónsson, Greg Mills, and Karsten W. Jacobsen. "Nudged elastic band method for finding minimum energy paths of transitions". In: *Classical and Quantum Dynamics in Condensed Phase Simulations*. WORLD SCIENTIFIC, June 1998. DOI: 10.1142/9789812839664_0016.

[44] Mudita Juneja and S. K. Nagar. "Particle swarm optimization algorithm and its parameters: A review". In: *2016 International Conference on Control, Computing, Communication and Materials (ICCCCM)*. IEEE, Oct. 2016. DOI: 10.1109/icccm.2016.7918233.

[45] Johannes Kästner and Paul Sherwood. "Superlinearly converging dimer method for transition state search". In: *The Journal of Chemical Physics* 128.1 (Jan. 2008), p. 014106. DOI: 10.1063/1.2815812.

[46] Ivaylo H. Katzarov, Dimitar L. Pashov, and Anthony T. Paxton. "Hydrogen embrittlement I. Analysis of hydrogen-enhanced localized plasticity: Effect of hydrogen on the velocity of screw dislocations in α–Fe". In: *Physical Review Materials* 1.3 (Aug. 2017). DOI: 10.1103/physrevmaterials.1.033602.

[47] Donald E. Knuth. *The Art of Computer Programming, Volume 3: (2nd Ed.) Sorting and Searching*. USA: Addison Wesley Longman Publishing Co., Inc., 1998. ISBN: 0201896850.

[48] C. Lanczos. "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators". In: *Journal of Research of the National Bureau of Standards* 45.4 (Oct. 1950), p. 255. DOI: 10.6028/jres.045.026.

[49] Dong C. Liu and Jorge Nocedal. "On the limited memory BFGS method for large scale optimization". In: *Mathematical Programming* 45.1-3 (Aug. 1989), pp. 503–528. DOI: 10.1007/bf01589116.

[50] Brendan D. McKay and Adolfo Piperno. "Practical graph isomorphism, II". In: *Journal of Symbolic Computation* 60 (Jan. 2014), pp. 94–112. DOI: 10.1016/j.jsc.2013.09.003.

[51] R.B. McLellan and Z.R. Xu. "Hydrogen-induced vacancies in the iron lattice". In: *Scripta Materialia* 36.10 (May 1997), pp. 1201–1205. DOI: 10.1016/s1359-6462(97)00015-8.

[52] Fedwa El-Mellouhi, Normand Mousseau, and Laurent J. Lewis. "Kinetic activation-relaxation technique: An off-lattice self-learning kinetic Monte Carlo algorithm". In: *Physical Review B* 78.15 (Oct. 2008). DOI: 10.1103/physrevb.78.153202.

[53] Simone Meloni, Mario Rosati, and Luciano Colombo. "Efficient particle labeling in atomistic simulations". In: *The Journal of Chemical Physics* 126.12 (Mar. 2007), p. 121102. DOI: 10.1063/1.2719690.

[54] M. I. Mendelev et al. "Development of new interatomic potentials appropriate for crystalline and liquid iron". In: *Philosophical Magazine* 83.35 (Dec. 2003), pp. 3977–3994. DOI: 10.1080/14786430310001613264.

[55] Luca Messina, Lorenzo Malerba, and Pär Olsson. "Stability and mobility of small vacancy–solute complexes in Fe–MnNi and dilute Fe–X alloys: A kinetic Monte Carlo study". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 352 (June 2015), pp. 61–66. DOI: 10.1016/j.nimb.2014.12.032.

[56] Hiroyoshi Momida et al. "Hydrogen-enhanced vacancy embrittlement of grain boundaries in iron". In: *Physical Review B* 88.14 (Oct. 2013). DOI: 10.1103/physrevb.88.144107.

[57] Letif Mones, Christoph Ortner, and Gábor Csányi. "Preconditioners for the geometry optimisation and saddle point search of molecular systems". In: *Scientific Reports* 8.1 (Sept. 2018). DOI: 10.1038/s41598-018-32105-x.

[58] B. Moon et al. “Analysis of the clustering properties of the Hilbert space-filling curve”. In: *IEEE Transactions on Knowledge and Data Engineering* 13.1 (2001), pp. 124–141. DOI: 10.1109/69.908985.

[59] Jorge J. Moré and David J. Thuente. “Line search algorithms with guaranteed sufficient decrease”. In: *ACM Transactions on Mathematical Software (TOMS)* 20.3 (Sept. 1994), pp. 286–307. DOI: 10.1145/192115.192132.

[60] G. M. Morton. *A computer oriented geodetic data base and a new technique in file sequencing*. Tech. rep. IBM Ltd, Mar. 1966.

[61] R. Nazarov, T. Hickel, and J. Neugebauer. “Ab initiostudy of H-vacancy interactions in fcc metals: Implications for the formation of superabundant vacancies”. In: *Physical Review B* 89.14 (Apr. 2014). DOI: 10.1103/physrevb.89.144108.

[62] Jorge Nocedal. “Updating quasi-Newton matrices with limited storage”. In: *Mathematics of Computation* 35.151 (Sept. 1980), pp. 773–773. DOI: 10.1090/s0025-5718-1980-0572855-7.

[63] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. 2nd ed. Springer New York, 2006. DOI: 10.1007/978-0-387-40065-5.

[64] R.A Oriani. “The diffusion and trapping of hydrogen in steel”. In: *Acta Metallurgica* 18.1 (Jan. 1970), pp. 147–157. DOI: 10.1016/0001-6160(70)90078-7.

[65] Yu. N. Osetsky, A. G. Mikhin, and A. Serra. “Study of copper precipitates in α–iron by computer simulation I. Interatomic potentials and properties of Fe and Cu”. In: *Philosophical Magazine A* 72.2 (Aug. 1995), pp. 361–381. DOI: 10.1080/01418619508239930.

[66] M.I. Pascuet et al. “Stability and mobility of Cu–vacancy clusters in Fe–Cu alloys: A computational study based on the use of artificial neural networks for energy barrier calculations”. In: *Journal of Nuclear Materials* 412.1 (May 2011), pp. 106–115. DOI: 10.1016/j.jnucmat.2011.02.038.

[67] Andreas Pedersen and Hannes Jonsson. “Simulations of hydrogen diffusion at grain boundaries in aluminum”. In: *Acta Materialia* 57.14 (Aug. 2009), pp. 4036–4045. DOI: 10.1016/j.actamat.2009.04.057.

[68] Andreas Pedersen and Hannes Jónsson. “Distributed implementation of the adaptive kinetic Monte Carlo method”. In: *Mathematics and Computers in Simulation* 80.7 (Mar. 2010), pp. 1487–1498. DOI: 10.1016/j.matcom.2009.02.010.

[69] J.M Peña, J.A Lozano, and P Larrañaga. “An empirical comparison of four initialization methods for the K-Means algorithm”. In: *Pattern Recognition Letters* 20.10 (Oct. 1999), pp. 1027–1040. DOI: 10.1016/s0167-8655(99)00069-0.

[70] Leonard Bessemer Pfeil. “The effect of occluded hydrogen on the tensile strength of iron”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 112.760 (Aug. 1926), pp. 182–195. DOI: 10.1098/rspa.1926.0103.

[71] Steve Plimpton. “Fast Parallel Algorithms for Short-Range Molecular Dynamics”. In: *Journal of Computational Physics* 117.1 (Mar. 1995), pp. 1–19. DOI: 10.1006/jcph.1995.1039.

[72] J. C. Polanyi and W. H. Wong. “Location of Energy Barriers. I. Effect on the Dynamics of Reactions A + BC”. In: *The Journal of Chemical Physics* 51.4 (Aug. 1969), pp. 1439–1450. DOI: 10.1063/1.1672194.

[73] B.T. Polyak. "The conjugate gradient method in extremal problems". In: *USSR Computational Mathematics and Mathematical Physics* 9.4 (Jan. 1969), pp. 94–112. DOI: 10.1016/0041-5553(69)90035-4.

[74] Brian Puchala, Michael L. Falk, and Krishna Garikipati. "An energy basin finding algorithm for kinetic Monte Carlo acceleration". In: *The Journal of Chemical Physics* 132.13 (Apr. 2010), p. 134104. DOI: 10.1063/1.3369627.

[75] Ashwin Ramasubramaniam, Mitsuhiro Itakura, and Emily A. Carter. "Interatomic potentials for hydrogen inα–iron based on density functional theory". In: *Physical Review B* 79.17 (May 2009). DOI: 10.1103/physrevb.79.174101.

[76] B. K. Rao and P. Jena. "Switendick criterion for stable hydrides". In: *Physical Review B* 31.10 (May 1985), pp. 6726–6730. DOI: 10.1103/physrevb.31.6726.

[77] D.C. Rapaport. "Large-scale molecular dynamics simulation using vector and parallel computers". In: *Computer Physics Reports* 9.1 (Dec. 1988), pp. 1–53. DOI: 10.1016/0167-7977(88)90014-7.

[78] K SAKAKI et al. "The effect of hydrogen on vacancy generation in iron by plastic deformation". In: *Scripta Materialia* 55.11 (Dec. 2006), pp. 1031–1034. DOI: 10.1016/j.scriptamat.2006.08.030.

[79] Robert Sedgewick and Michael Schidlowsky. *Algorithms in Java, Third Edition, Parts 1-4: Fundamentals, Data Structures, Sorting, Searching*. 3rd. USA: Addison-Wesley Longman Publishing Co., Inc., 1998. ISBN: 0201361205.

[80] Jun Song and W. A. Curtin. "Atomic mechanism and prediction of hydrogen embrittlement in iron". In: *Nature Materials* 12.2 (Nov. 2012), pp. 145–151. DOI: 10.1038/nmat3479.

[81] Mickaël Trochet et al. "Off-Lattice Kinetic Monte Carlo Methods". In: *Handbook of Materials Modeling*. Springer International Publishing, 2018, pp. 1–29. DOI: 10.1007/978-3-319-42913-7_29-1.

[82] George H. Vineyard. "Frequency factors and isotope effects in solid state rate processes". In: *Journal of Physics and Chemistry of Solids* 3.1-2 (Jan. 1957), pp. 121–127. DOI: 10.1016/0022-3697(57)90059-8.

[83] Ulrich Welling and Guido Germano. "Efficiency of linked cell algorithms". In: *Computer Physics Communications* 182.3 (Mar. 2011), pp. 611–615. DOI: 10.1016/j.cpc.2010.11.002.

[84] Philip Wolfe. "Convergence Conditions for Ascent Methods". In: *SIAM Review* 11.2 (Apr. 1969), pp. 226–235. DOI: 10.1137/1011036.

[85] Philip Wolfe. "Convergence Conditions for Ascent Methods. II: Some Corrections". In: *SIAM Review* 13.2 (Apr. 1971), pp. 185–188. DOI: 10.1137/1013035.

[86] Penghao Xiao, Qiliang Wu, and Graeme Henkelman. "Basin constrained κ-dimer method for saddle point finding". In: *The Journal of Chemical Physics* 141.16 (Oct. 2014), p. 164111. DOI: 10.1063/1.4898664.

[87] Lin Yang et al. "Implementation of metal-friendly EAM/FS-type semi-empirical potentials in HOOMD-blue: A GPU-accelerated molecular dynamics software". In: *Journal of Computational Physics* 359 (Apr. 2018), pp. 352–360. DOI: 10.1016/j.jcp.2018.01.015.

[88] Yi Zeng, Penghao Xiao, and Graeme Henkelman. "Unification of algorithms for minimum mode optimization". In: *The Journal of Chemical Physics* 140.4 (Jan. 2014), p. 044115. DOI: 10.1063/1.4862410.

Appendices

A.1 Partial vector-derivative of the line between atoms

Here we derive Eq. (3.5), repeated below for convenience:

$$\frac{\partial r^{\alpha\beta}}{\partial \boldsymbol{r}^{\gamma}} = \left(\delta^{\gamma\beta} - \delta^{\gamma\alpha}\right) \hat{\boldsymbol{r}}^{\alpha\beta} \tag{A.1}$$

where $\boldsymbol{r}^{\alpha}$ is the position vector of the α^{th} atom, $r^{\alpha\beta} = \left\|\boldsymbol{r}^{\alpha\beta}\right\| = \left\|\boldsymbol{r}^{\beta} - \boldsymbol{r}^{\alpha}\right\|$ and δ is the Kronecker Delta. Through the next section we use: Greek superscripts to denote the atom index without implied Einstein summation; Roman subscripts to denote the component of a vector with implicit Einstein summation over repeated indices. With this notation:

$$\frac{\partial r^{\alpha\beta}}{\partial r_i^{\gamma}} = \frac{\partial}{\partial r_i^{\gamma}} \sqrt{r_j^{\alpha\beta} r_j^{\alpha\beta}} = \frac{1}{2\sqrt{r_j^{\alpha\beta} r_j^{\alpha\beta}}} \frac{\partial}{\partial r_i^{\gamma}} \left(r_j^{\alpha\beta} r_j^{\alpha\beta}\right) = \frac{1}{r^{\alpha\beta}} \left(r_j^{\alpha\beta} \frac{\partial}{\partial r_i^{\gamma}} r_j^{\alpha\beta}\right) \tag{A.2}$$

then expanding $r_j^{\alpha\beta} = r_j^{\beta} - r_j^{\alpha}$:

$$\frac{\partial r^{\alpha\beta}}{\partial r_i^{\gamma}} = \frac{r_j^{\alpha\beta}}{r^{\alpha\beta}} \left(\frac{\partial r_j^{\beta}}{\partial r_i^{\gamma}} - \frac{\partial r_j^{\alpha}}{\partial r_i^{\gamma}}\right) = \frac{r_j^{\alpha\beta}}{r^{\alpha\beta}} \left(\delta^{\gamma\beta}\delta_{ij} - \delta^{\gamma\alpha}\delta_{ij}\right) = \frac{r_i^{\alpha\beta}}{r^{\alpha\beta}} \left(\delta^{\gamma\beta} - \delta^{\gamma\alpha}\right) \tag{A.3}$$

which is the component form of Eq. (3.5).

A.2 An integer sorting algorithm for the LCR method

Algorithm A.1 In-place counting sort.

```
procedure SORT(list, key, M)                    ▷ key: list[i] → {0, 1, ..., M − 1}.
    offset[0, 1, ..., M − 1] ← 0;
    for each x in list do                       ▷ Count occurrences of each value.
        v ← key(x)
        offset[v] ← offset[v] + 1
    s ← 0                                       ▷ Holds partial sum.
    for i ← 0, 1, ..., M − 1 do                 ▷ Compute offsets, cumulative sum of counts.
        t ← offset[i]                           ▷ t is a temporary to shift cumulative sum.
        counts[i] ← s
        s ← s + t
    next ← offset                               ▷ Makes copy.
    s ← 0                                       ▷ Now counts how many swaps we have performed
    while s ≠ length of list do
        for g ← M − 1, M − 2, ..., 1 do
            for i ← offset[g] − 1, offset[g] − 2, ..., next[g − 1] do
                swap(list[i], list[next[key(list[i])]++])
                s ← s + 1
```

As discussed in Section 3.2 the LCR method requires sorting an array of atoms into their Λ ordering. A standard integer sorting algorithm such as a radix [47, p. 168] or hash [31] sorting could be used however, there are domain specific constraints to be exploited. Firstly, the number of atoms is much greater than the number of cells $n \approx 25M$ such that, the range of possible values is much less than the number of items to sort. Secondly, we expect the atoms to only move a small distance between each sort hence, the list will be approximately sorted most of the time.

We devise an in-place variation of a counting sort [79, p. 312], detailed in Algorithm A.1, that achieves $\mathcal{O}(n)$ time complexity with only $\mathcal{O}(M)$ additional memory and exhibits particularly good performance for nearly sorted data. Performance analysis confirmed our algorithm outperformed built-in sorting functions for our (specific) use-case.

A.3 Wolfe conditions on line-searches

The strong Wolfe conditions for an inexact line search of the objective function f are [84, 85]:

$$f\left(\mathbf{x}_k + \alpha_k \mathbf{p}_k\right) \leq f\left(\mathbf{x}_k\right) + c_1 \alpha_k \nabla f(\mathbf{x}_k)^T \mathbf{p}_k \tag{A.4}$$

$$\left|\nabla f\left(\mathbf{x}_k + \alpha_k \mathbf{p}_k\right)^T \mathbf{p}_k\right| \leq c_2 \left|\nabla f(\mathbf{x}_k)^T \mathbf{p}_k\right| \tag{A.5}$$

where the constants c_1, c_2 satisfy $0 < c_1 < c_2 < 1$. Equation (A.4) is called the *sufficient decrease* condition and ensures the step suitably reduces the objective function. Eq. (A.5) is called the *curvature condition* and ensures the step is not too small. Together with a reasonable choice of decent direction the Wolfie conditions can ensure a gradient descent algorithm converges in a finite number of steps [59]. The choices of the constants c_1 and c_2 used in this work were 10^{-4} and 0.9 respectively.