MULTISCALE HYBRID SIMULATION OF BRITTLE FRACTURE

MULTISCALE HYBRID SIMULATION OF BRITTLE FRACTURE

James Richard Kermode

PEMBROKE COLLEGE

UNIVERSITY OF CAMBRIDGE

THIS DISSERTATION IS SUBMITTED FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

NOVEMBER 2007

Contents

List of Figures

Summary

Multiscale Hybrid Simulation of Brittle Fracture

James Kermode
Pembroke College

When a brittle material is loaded to the limit of its strength, it fails by nucleation and propagation of a crack. The conditions for crack propagation are created by the concentration of a long-range stress field at an atomically sharp crack tip, creating a complex and strongly coupled multiscale system.

This thesis reports the results of multiscale simulations of the brittle fracture of silicon on the (111) cleavage plane. The simulations are made possible by combining a quantum mechanical description of the processes taking place near the crack tip with a classical atomistic model that captures the long-range elastic relaxation. The 'Learn on The Fly' technique is used to couple the quantum and classical models, allowing accurate quantum forces to be combined with classical forces using a simple adjustable potential to give stable dynamics.

The simulations predict that fracture is unstable on the (111) plane at low speeds; conventionally this has been thought of as the most stable crack plane. The instability is caused by a crack tip reconstruction which triggers a positive feedback 'sinking' mechanism leading to macroscopic, experimentally observable corrugations. Recent experiments have observed crack surface features consistent with these predictions. The instability is the first example in a crystalline material of a fracture instability which onsets below a critical velocity, and shows how subtle atomistic details at the crack tip can control the qualitative macroscopic fracture behaviour.

Preface

This dissertation describes work done between October 2004 and November 2007 at the Cavendish Laboratory, Cambridge, under the supervision Prof. M. C. Payne.

I have contributed to three publications during the course of my PhD:

- G. Csányi, G. Moras, J. R. Kermode, M. C. Payne, A. Mainwood and A. de Vita, Multiscale Modeling of Defects in Semiconductors: A Novel Molecular Dynamics Scheme. In D. A. Drabold and S. K. Estreicher, editors, *Theory of Defects in Semiconductors* (Springer, 2007)
- G. Csányi, S. Winfield, J. R. Kermode, A. Comisso, A. de Vita, N. Bernstein and M. C. Payne, Expressive Programming for Computational Physics in Fortran 95+, *IoP Comput. Phys. Newsletter*, Spring 2007
- J. R. Kermode, T. Albaret, D. Sherman, N. Bernstein, P. Gumbsch, M. C. Payne, G. Csányi and A. de Vita, Low speed fracture instabilities in a brittle crystal, Submitted to *Nature*

The latter paper reports the outcome of my work on the brittle fracture of silicon. Material from this paper appears in this dissertation.

This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and acknowledgements. It has not been submitted in whole or in part for any degree or diploma at this or any other university, and does not exceed 60 000 words, including tables, footnotes, bibliography and appendices, but excluding photographs and diagrams.

James Kermode
November 2007

Acknowledgements

I would like to thank my supervisor Mike Payne for providing me with the opportunity to work in the TCM group and to benefit from all it has to offer, and for his generosity in supplying many hours of both computer and proof-reading time. I have been priveleged to work with Gábor Csányi, who has been a continuous source of inspiration, enthusiasm and encouragement. It has been a great pleasure working in the unique environment of TCM — completing my PhD here has been an enjoyable as well as rewarding experience.

I have been fortunate to collaborate with many people happy to share their knowledge and ideas, both within the LOTF project and more widely, including Noam Bernstein, Alessandro de Vita, Steve Winfield, Gianpetro Moras, Tristan Albaret, Peter Gumbsch, and John Willis. I am grateful to Dov Sherman for collaborating with us with respect to the low crack speed experimental results discussed in Chapter 8. I acknowledge financial support from EPSRC during the course of my PhD and for several conferences and workshops. Computational resources were provided by the Cambridge University HPC service.

I am sure that TCM would soon grind to a halt without Tracey: acting as the common sense for over fifty physicists is no mean feat! Thanks are due to Michael Rutter for making the computers work, and for imparting information on topics as diverse as computer architecture and Thatcherism, all supplied with a healthy dose of sarcasm. I have been very fortunate to share an office with Danny Cole for three years, and could not have hoped for a better office mate. I am grateful to my friends in TCM, CHaOS, Pembroke, the Strange Blue ultimate frisbee club and elsewhere who kept me sane during my PhD. In particular, life in the Cavendish would have been far more uncomfortable without Tea, so I thank all those who believe in it, and with whom many I have enjoyed so many half-hours of pointless conversation.

Finally, I would like to thank my family for their support, and Selina for keeping me supplied with encouragement and ice cream.

Chapter 1

Introduction

Atomistic simulation is a discipline based on numerically solving the fundamental equations of classical and quantum physics in order to directly predict the trajectories of the atoms which make up a physical system. That this can be done in principle has been appreciated for some time — centuries in the case of classical mechanics — but it is only with the advent of computers capable of carrying out this procedure accurately and reliably that simulation has become an indispensable research tool, complementary both to experimental and theoretical approaches. In Chapter 2, I review the background to two of the main methods currently used to model physical systems at the atomic level: classical molecular dynamics and first principles quantum mechanical calculations. The former allow simulations of millions of atoms to be carried out on a nanosecond timescale [1, 2, 3, 4], but the accuracy is limited by the requirement to use simple parameterisations as interatomic potentials [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. If the scientific question of interest can be effectively answered by considering the behaviour of a very small number of atoms, up to around a hundred, then *ab initio* approaches allow this limitation to be overcome [19, 20, 21, 22, 23, 24].

In many cases we can extract enough information from these accurate quantum mechanical calculations to parametrise less transferable, but far less expensive, models and use them on a larger length scale [25]. For some systems however, it is impossible to separate the behaviour on the various length scales, since the coupling between them is strong and bidirectional. Then the only option is to carry out a *hybrid* simulation, where some parts of the system are treated at a higher level of accuracy [26, 27, 28, 29]. An overview of existing hybrid schemes is given in Chapter 3.

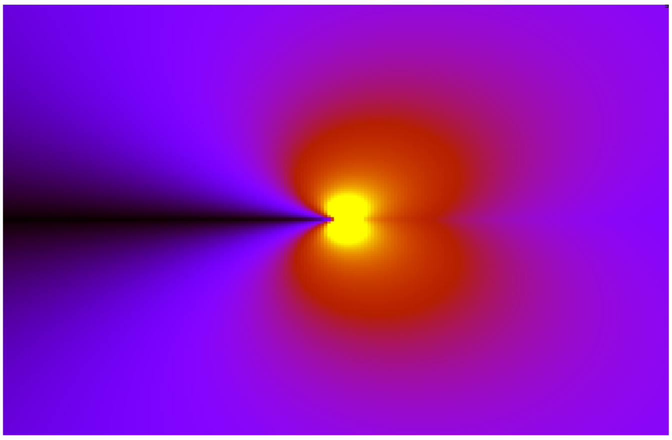


Figure 1.1: Maximum principal stress near the the tip of a crack under uniaxial tension in the opening mode, from the linear elastic solution (discussed in detail in Section 5.2). Black areas are the least stressed and yellow the most.

The best known example of such a multiscale system is the fracture of brittle materials, which forms the subject of Chapter 5. The conditions for crack propagation are created by stress concentration at the crack tip, and depend on macroscopic parameters such as the loading geometry and dimensions of the specimen [30, 31, 32, 33]. In real materials, however, the detailed crack propagation dynamics, are entirely determined by atomic scale phenomena since brittle crack tips are atomically sharp and propagate by breaking bonds, one at a time, at each point along the crack front [34, 35]. This means the tip region is primarily a one dimensional line, perpendicular to the direction of propagation, and so it should be possible to define a contiguous embedding region to be treated with a more accurate model in a hybrid simulation. There is a constant interplay between the length scales because the opening crack gives rise to a stress field with a singularity at the tip [36], as illustrated in Fig. 1.1, and in turn it is this singular stress field which breaks the bonds that advance the crack. Only by including the tens of thousands of atoms that contribute significantly to the elastic relaxation of this stress field can we hope to accurately model the fracture system, and thus a multiscale approach is essential. This thesis is concerned with the application of hybrid methods to the brittle fracture of silicon, in particular, I consider low-speed fracture on the (111) cleavage plane.

The major difficulty in hybrid simulation is dealing with the boundary between the two different descriptions of material; often unphysical forces arise here when two incompatible models are combined [37, 38, 39]. The 'Learn on the Fly' (LOTF) hybrid method [40, 41, 42, 43, 44], described in Chapter 4, is a recently introduced strategy to solve the boundary problem, at least in materials that can be described using short range forces. LOTF couples the quantum and classical dynamics in such a way that the system behaves instantaneously as if the entire specimen is being treated at a quantum mechanical level of detail. This is achieved by combining accurate quantum forces for the region of interest obtained from *ab initio* calculations with qualitatively correct classical forces further away, using an adjustable potential to give stable dynamics.

Chapter 6 sets out the methodology for the fracture simulations and describes a series of validation tests, whilst the main results of this work are presented in Chapter 7. I find that fracture is unstable on the (111) plane at low speeds; conventionally this has been thought of as the most stable crack plane. The instability I report in this thesis is caused by a crack tip reconstruction which triggers a positive feedback 'sinking' mechanism leading to macroscopic, experimentally observable, corrugations. As discussed in Chapter 8, recent experiments have observed surface features consistent with these predictions [45, 46].

Chapter 2

Classical and Quantum Simulation

2.1 Introduction

This aim of this chapter is to review the necessary theoretical background to this thesis. Since the range of topics to be covered is quite large, each has been described as briefly as possible, with references to sources of further information. The work contained within this thesis applies ideas from both classical and quantum physics.

I start with a review of the background necessary to carry out classical molecular dynamics simulations on large numbers of atoms, and then look at the interatomic potentials used for these simulations in a little more detail. The next section focuses on a much smaller length scale: we discuss the *ab initio* quantum mechanical approach to electronic structure calculation. Finally, I shall look at the tight binding method, often thought of as a way to bridge the gap between quantum and classical simulations.

2.2 Classical Methods

Classical Molecular Dynamics

Molecular dynamics (MD) is a tool to simulate the motion of a system of classical particles based on Newton's Laws of Motion, to facilitate atomically precise computer-based experiments. In this section I give only a brief introduction to some relevant areas of classical MD, see Haile [1] and Rapaport [2] for further details. The particle trajectories are the immediate outcome of a simulation but, in addition, these trajectories can be used to compute a variety of structural and thermodynamic properties.

Classical MD is most useful when applied to very complex structures containing millions of atoms, where we have no idea of the microscopic processes taking place before carrying out the simulation. It is particularity useful for exploring the phase space of systems such as liquids and amorphous solids where it is very difficult to find a unique ground state structure using simple geometry optimisation techniques.

Newtonian Equations of Motion

For an isolated system of N particles with masses $\{m_i\}$, the total energy $E = T + V$ is conserved. We can write the Hamiltonian for this system as

$$H = T(\{\mathbf{p}_i\}) + V(\{\mathbf{r}_i\}) \tag{2.1}$$

where $\mathbf{r}_i$ and $\mathbf{p}_i$ are the position and momentum of the i^{th} particle, $V(\{\mathbf{r}_i\})$ is the potential energy and $T(\{\mathbf{p}_i\})$ is the kinetic energy, given by

$$T(\{\mathbf{p}_i\}) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2\, m_i}. \tag{2.2}$$

Molecular dynamics is essentially independent of how the potential energy V is calculated, whether it be from a simple interatomic potential (see Section 2.3), a more complex quantum mechanical method such as tight binding (Section 2.5) or even a fully *ab initio* calculation (Section 2.4).

Using the Lagrangian formulation of dynamics and applying the Euler-Lagrange equation, this Hamiltonian gives rise to the well known Newtonian equation of motion for each particle

$$\mathbf{F}_i = m_i\, \ddot{\mathbf{r}}_i \tag{2.3}$$

where $\mathbf{F}_i$ is the total force on the i^{th} particle caused by all the other particles, and is given by $\mathbf{F}_i = -\nabla_i V$.

Integrating the Equations of Motion

To perform an MD simulation, we need to numerically integrate Eq. 2.3. Taylor expanding the trajectory to third order in Δt gives

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_t \Delta t + \frac{1}{2}\left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_t \Delta t^2 + \frac{1}{6}\left.\frac{\mathrm{d}^3\mathbf{r}}{\mathrm{d}t^3}\right|_t \Delta t^3 + O(\Delta t^4). \tag{2.4}$$

This inspires the simplest possible integration scheme, the Euler method:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\,\Delta t + O(\Delta t^2), \tag{2.5}$$

$$\dot{\mathbf{r}}(t+\Delta t) = \dot{\mathbf{r}}(t) + \frac{\mathbf{F(t)}}{m}\,\Delta t + O(\Delta t^2). \tag{2.6}$$

While it has the advantage of being extremely simple, the Euler method is not widely used because of its limited accuracy. In fact for no more work, we can do much better by summing the Taylor series of Eq. 2.4 with a similar expansion for $t-\Delta t$ to eliminate the terms of odd order:

$$\mathbf{r}(t+\Delta t) = 2\,\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \frac{\mathbf{F(t)}}{m}\,\Delta t^2 + O(\Delta t^4). \tag{2.7}$$

This is the widely used Verlet scheme [3]. Due to cancellation of errors, it has third-order accuracy even though it contains no third-order derivatives. It also has time-reversal symmetry, which gives the algorithm excellent numerical stability, even for moderately large time steps.

A disadvantage of the method is that it treats velocities as less important than positions; the velocities are not directly integrated but calculated from the atomic positions using a first-order central difference method:

$$\dot{\mathbf{r}}(t) = \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t-\Delta t)}{2\,\Delta t} + O(\Delta t^2). \tag{2.8}$$

Although there is no accumulation of error since the velocities are recalculated at each time step, subtracting numbers of similar magnitude like this can lead to computational round-off error, and so the basic Verlet method is seldom used when accurate velocities are required. This disadvantage is overcome by the Velocity-Verlet scheme [4], where the positions and velocities are treated on an equal footing.

In the Velocity-Verlet algorithm, Eqs. 2.7 and 2.8 are replaced by

$$\begin{aligned}\mathbf{r}(t+\Delta t) &= \mathbf{r}(t)+\dot{\mathbf{r}}(t)\,\Delta t+\frac{1}{2}\frac{\mathbf{F}(t)}{m}\,\Delta t^2+O(\Delta t^4), \qquad (2.9)\\ \dot{\mathbf{r}}(t+\Delta t) &= \dot{\mathbf{r}}(t)+\frac{1}{2}\left[\frac{\mathbf{F}(t)}{m}+\frac{\mathbf{F}(t+\Delta t)}{m}\right]\Delta t+O(\Delta t^4). \qquad (2.10)\end{aligned}$$

Velocity-Verlet is also known as the ‘leapfrog’ method. It has all the desirable properties of the basic Verlet scheme, with the added advantage that the velocities are also accurate to third-order. The algorithm ensures that the area of phase-space occupied remains constant as the simulation evolves, which guarantees time-reversibility and energy conservation. It is important to ensure that energy is being numerically conserved in an MD simulation to verify that the time step selected is not too large; the Velocity-Verlet will often appear to give reasonable looking trajectories for large time steps that give a considerable drift in the total energy.

Thermodynamic Properties

To extract macroscopic thermodynamic properties from an MD simulation, we need to have good sampling of the phase space of the system. This is the *ergodic hypothesis*: that each microstate is accessed with equal probability. We can then take time averages of observables from thousands of steps in an MD trajectory:

$$\langle A\rangle = \lim_{t\to\infty}\frac{1}{t}\int_{t_0}^{t_0+t} A(\tau)\,\mathrm{d}\tau. \qquad (2.11)$$

In the limit of long run times, these will be the same as statistical mechanics ensemble averages, so we can use them to measure various properties, for example the average temperature, pressure or kinetic energy.

Geometry Optimisation

Given a system of particles and a potential energy function, we can ask the question: which arrangements of atoms correspond to local minima in the total energy? This is the geometry optimisation problem, and many different approaches to solving it have been proposed.

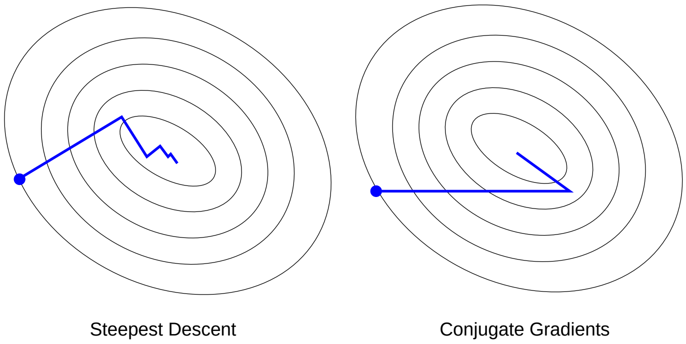


Figure 2.1: Comparison of the steepest descent and conjugate gradient minimisation methods. Steepest descent takes many more steps to reach the minimum since it has to zig-zag along the elliptical valley.

Direct Minimisation

If we reshape the atomic positions into a $3N$ dimensional vector $\mathbf{R}$, then geometry optimisation can be reformulated as a straight-forward minimisation problem. The simplest way to proceed is by steepest descent where we simply follow the gradient of the potential energy downhill

$$\mathbf{R}_{n+1} = \mathbf{R}_n - \gamma \nabla V(\mathbf{R}) \tag{2.12}$$

where the step size γ is chosen by a line search in the ∇V direction. This approach can be very inefficient because the algorithm can take many iterations to converge to a local minimum if the curvature of the energy surface is very different in different directions.

The *conjugate gradient* method is a much improved minimisation approach which uses conjugate directions instead of the local gradient to go downhill; this is particularly advantageous if there is a long, narrow valley near to the minimum, as illustrated in Fig. 2.1.

An even more efficient method is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [47]. This is one of a large number of quasi-Newton methods which build up knowledge of the Hessian matrix of second derivatives of the energy function without having to evaluate the matrix explicitly at each step. BFGS combines this with a more sophisticated line search algorithm to find the optimal step size.

It is possible to allow the lattice parameters of the system under consideration to vary during the geometry optimisation process in response to the total stress.

These variable-cell methods make it easy to determine properties such as the lattice constant for a system at the same time as relaxing the internal atomic configuration.

Damped Molecular Dynamics

It is possible to perform geometry optimisation using molecular dynamics by adding a fictitious frictional force to drive the system downhill in energy. The total force on the i^{th} particle becomes

$$\mathbf{F}_i = -\nabla_i V - \gamma\,\dot{\mathbf{r}}_i. \tag{2.13}$$

This is known as *damped molecular dynamics*, and can be extended for systems with multiple local minima in an approach known as *simulated annealing*.

It is important to choose the damping coefficient γ appropriately to be close to the critical damping factor for the system to proceed to the minimum as efficiently as possible; this can be done for each atom independently. A procedure for doing this was proposed by Probert [48], where it is reported that the resulting damped MD can be comparable in efficiency to the BFGS algorithm.

Fast Inertial Relaxation Engine

FIRE (Fast Inertial Relaxation Engine) is a recently proposed optimisation algorithm [49]. It is based on damped MD but with some modifications that speed up convergence, such as an adaptive time-step. The authors describe FIRE as the route that a blind skier would take when descending a mountain, following the equation of motion

$$\dot{\mathbf{v}}(t) = \frac{\mathbf{F}(t)}{m} - \gamma(t)|\mathbf{v}(t)|\left(\hat{\mathbf{v}}(t) - \hat{\mathbf{F}}(t)\right), \tag{2.14}$$

with the mass m, velocity $\mathbf{v} = \dot{\mathbf{r}}$, force $\mathbf{F}$ and hats denoting unit vectors. The function $\gamma(t)$ is chosen to introduce acceleration in a steeper downhill direction when the power $\mathbf{P} = \mathbf{F} \cdot \mathbf{v}$ is positive. FIRE is particularity well suited to pathological situations such as a spiral shaped potential energy surface where conjugate gradient based algorithms are held back by the inefficiency of the line minimisation, but it has also been shown to give performance comparable with BFGS in more realistic situations.

Ensembles and Thermostats

A standard molecular dynamics simulation using the Velocity-Verlet algorithm of Eq. 2.9 conserves the total energy, so it generates the microcanonical ensemble NVE. Here the number of particles N, volume V and total energy E are kept constant. In this approach, the temperature can be set by rescaling velocities until the system equilibrates at the desired temperature, and monitoring the total energy E to ensure that the microcanonical ensemble is sampled correctly.

In 1980 Andersen noted the possibility of generating other ensembles for the first time [50], with the aim of more closely matching experimental conditions. In particular it is desirable to generate the canonical ensemble NVT, where the temperature T is kept constant instead of the total energy. For a very large system containing many thousands of atoms, we approach the thermodynamic limit, the system acts as its own heat bath, and the NVE and NVT ensembles become equivalent. In order to generate the canonical ensemble in a smaller system we need a thermostat which adds or removes heat from the system, maintaining a constant temperature by simulating the coupling of the system to a heat bath.

Simple Thermostats

Many different thermostats have been proposed. Andersen's first method consisted of choosing a particle at random and choosing a new velocity for it from the correct Maxwell distribution. Whilst this generates the correct ensemble, it proved unpopular since the impulses applied to particles significantly affect the dynamics.

An even simpler approach is to periodically rescale the velocities directly to the target temperature. However this can have a pronounced effect on atomic trajectories and still does not generate the correct ensemble. A widely used improvement is the Berendsen thermostat [51], which smoothly rescales the kinetic energy towards the target value, with a driving force equal to the difference between the instantaneous and target kinetic energies. This results in a stable and easy to implement method, but it is still not guaranteed to generate the correct ensemble.

Nosé-Hoover Thermostat

This popular thermostat was first proposed by Nosé in 1984 [52], and subsequently by Hoover in an equivalent form [53]. A major attraction is the existence of a conserved quantity in the extended dynamics of the combined system and heat bath. It guarantees sampling from the isothermal ensemble, however, Nosé-Hoover type thermostats can behave non-ergodically, since there is no stochastic component. This shortcoming can be overcome, to an extent, by chaining several thermostats together; this is now common practice. This results in a more chaotic scheme, but it is still fundamentally deterministic and can result in long correlation times.

Langevin Thermostat

The stochastic nature of a heat bath can be modelled using Langevin dynamics to give a straightforward thermostat that maintains ergodic sampling and samples the canonical ensemble correctly, in an approach that has become known as a Langevin thermostat [54]. The Newtonian equations of motion are modified by the addition of a dissipative term representing viscous damping due to fictitious heat bath particles, and a random force to represent the effect of collisions with those particles:

$$\mathbf{F}_i = -\nabla_i V - \gamma\,\dot{\mathbf{r}}_i + \boldsymbol{\xi}_i. \tag{2.15}$$

The random forces $\boldsymbol{\xi}_i$ are drawn from a Gaussian distribution with zero mean and unit variance scaled according to the Stokes-Einstein fluctuation-dissipation relation for the diffusion coefficient, *i.e.* by a factor

$$\sqrt{\frac{2mk_BT\gamma}{\Delta t}}. \tag{2.16}$$

where m is the particle mass, T the simulation temperature and Δt the molecular dynamics time-step. This approach can be readily integrated into the Velocity-Verlet scheme [55]. We can identify the damping coefficient γ with $1/\tau$, where τ is the characteristic time scale over which the thermostat acts.

Deterministic and stochastic thermostats affect the trajectories in different ways. A strongly deterministic thermostat does not exhibit ergodicity, and so the region of phase space explored by the simulation is restricted. Conversely, a very strongly stochastic thermostat makes the MD simulation equivalent to Monte Carlo: this is useless if the aim is to understand the microscopic dynamical processes. The approach used in this work is to use a stochastic Langevin thermostat with τ chosen to be as large as possible while still maintaining a constant temperature, so that the effect on the dynamics is as small as possible.

Calculation of Elastic Properties

The elastic constants of a material place constraints on crystal structure stability and they determine the response of a crystal to external forces. In order to compare the accuracy of the description of a material provided by different interatomic potentials and quantum mechanical methods, and to compare with experimental measurements, we need to be able to calculate values for the elastic constants predicted by a particular model. For simplicity, here we restrict the discussion to the particular case of cubic crystals, but the method described can be extended to other crystal structures.

The elastic constant tensor c_{ijkl} relates the applied stress tensor σ to the resulting strain tensor ϵ, and is defined by $\sigma_{ij} = c_{ijkl}\,\epsilon_{kl}$. Symmetries of ϵ and σ, and the requirement for the strain energy density to be positive definite, reduce the fourth order tensor c_{ijkl} to a 6×6 matrix C_{ij}. For a cubic cell, there are only three independent elements, so the equation relating stress and strain becomes

$$\begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{xy} \\ \sigma_{xz} \\ \sigma_{yz} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ \epsilon_{xy} \\ \epsilon_{xz} \\ \epsilon_{yz} \end{pmatrix}. \qquad (2.17)$$

From the three non-zero elements of this matrix, C_{11}, C_{12} and C_{44}, it is possible to calculate all other elastic constants, for example the bulk modulus is

$$B = \frac{C_{11} + 2\,C_{12}}{3}. \qquad (2.18)$$

The Young's modulus and Poisson ratio of a cubic material can be calculated from the C_{ij} as a function of crystallographic orientation [56]. There are two different C_{44} constants in common use: C_{44} is the elastic response when the internal degrees of freedom of the cell are allowed to relax, whilst for C_{44}^0 the atoms are fixed. C_{44}^0 is not accessible to experiment, and so was largely ignored in the early literature on elastic forces, but modern *ab initio* methods that can predict elastic constants within a few percent of experiment have greatly increased its utility.

For cubic cell materials such as silicon, we have seen there are only three independent elastic constants so we can evaluate them by calculating the values of any three linearly independent combinations. It is common to choose the physically important combinations B, $C_{11} - C_{12}$ and C_{44}. This is because a cubic material will not be mechanically stable unless all of these are positive. Firstly, we need to calculate the bulk modulus, defined by

$$B = -V\frac{\partial p}{\partial V} = V\frac{\partial^2 E}{\partial V^2}. \tag{2.19}$$

An estimate for the bulk modulus of a cubic crystal can therefore be obtained by expanding and compressing the cell and measuring the curvature of the energy. The minimum of this curve corresponds to the equilibrium lattice constant a_0. To calculate the other two combinations, we define a reduced strain tensor that transforms the primitive cubic cell lattice vectors $\mathbf{a}_i$ to new vectors via the relation

$$\mathbf{a}_i^{'} = \mathbf{a}_i\,(I + \epsilon) \tag{2.20}$$

where I is the 3×3 identity matrix. For non-rotating strains ϵ is a symmetric tensor with six independent components:

$$\epsilon = \begin{pmatrix} e_1 & \frac{1}{2}e_6 & \frac{1}{2}e_5 \\ \frac{1}{2}e_6 & e_2 & \frac{1}{2}e_4 \\ \frac{1}{2}e_5 & \frac{1}{2}e_4 & e_3 \end{pmatrix}. \tag{2.21}$$

If we apply this strain to a lattice, then by Hooke's Law the total energy of the system changes by an amount

$$\Delta E(\epsilon) = -P(V)\Delta V + \frac{1}{2}V\sum_{i=1}^{6}\sum_{j=1}^{6} e_i\,C_{ij}\,e_j + O(\epsilon^3) \tag{2.22}$$

where V is the volume of the undistorted lattice, $P(V)$ the pressure of the undistorted lattice and ΔV is the change in volume of the lattice due to the strain ϵ.

The combination $C_{11} - C_{12}$ can be calculated using the volume preserving orthorhombic strain $e_1 = -e_2 = x$, $e_3 = x^2/(1 - x^2)$, $e_4 = e_5 = e_6 = 0$, which yields

$$\Delta E(x) = V\,(C_{11} - C_{12})x^2 + O(x^4). \tag{2.23}$$

The C_{44} constants can be calculated using a volume conserving monoclinic strain

$e_6 = x$, $e_3 = x^2/(4-x^2)$, $e_1 = e_2 = e_4 = e_5 = 0$, giving

$$\Delta E(x) = \frac{1}{2} V\, C_{44}\, x^2 + O(x^4). \tag{2.24}$$

For C_{44}, we carry out the internal relaxation using a geometry optimisation technique to minimise the forces on all atoms for each point on the $E(x)$ curve. For the unrelaxed modulus C_{44}^0 we simply omit the minimisation step.

For a crystal structure to be stable, there are a series of Cauchy-Born relationships [57] between elastic constants that must be satisfied. Harrison [58] has developed these relationships for the diamond structure, and they have recently been further extended to include the unrelaxed shear modulus C_{44}^0 by Bazant *et al.* [13]. The final relation is

$$C_{44}^0 - C_{44} = \frac{(C_{11} + 8C_{12})^2}{9\,(7C_{11} + 2C_{12})}. \tag{2.25}$$

Any empirical model that does not obey this condition will not give a good description of elastic properties in the diamond phase.

2.3 Interatomic Potentials

So far we have not discussed the form of the potential energy function $V(\{\mathbf{r}_i\})$ in Eq. 2.1 from which the forces used for molecular dynamics are derived. An interatomic potential is constructed by choosing a functional form and then fitting the parameters of this potential to reproduce observables such as the lattice constant, cohesive energy and other elastic constants. The fit target can either be experimental results, in which case the potential is referred to as an *empirical interatomic potential*, or the results of *ab initio* calculations.

Any interatomic potential can be decomposed into a sum of one-body, two-body, three-body, etc. contributions

$$V(\{\mathbf{r}_i\}) = \sum_i V_1(\mathbf{r}_i) + \sum_{\substack{\text{bonds}\\ i,j\\ i<j}} V_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{\substack{\text{angles}\\ i,j,k\\ i<j<k}} V_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \cdots + V_N(\mathbf{r}_1, \ldots, \mathbf{r}_N) \tag{2.26}$$

where the conditional sums avoid double-counting. For this expression to be useful, the V_n functions must converge quickly to zero with increasing n. The one-body term V_1 describes an external force applied to the system so is often absent. The simplest interatomic potentials consist of only two-body contributions. The two-body term is commonly a function only of the bond length $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$, and the three-body term of the bond lengths r_{ij} and r_{ik} and the angle θ_{jik} between the bonds ij and ik, defined by

$$\cos\theta_{jik} = \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{ik}}{|\mathbf{r}_{ij}|\,|\mathbf{r}_{ik}|}. \tag{2.27}$$

A general two- and three-body interatomic potential can therefore be written

$$V(\{\mathbf{r}_i\}) = \sum_{\substack{i,j \\ i<j}} V_2(r_{ij}) + \sum_{\substack{i,j,k \\ i<j<k}} V_3(r_{ij}, r_{ik}, \theta_{jik}). \tag{2.28}$$

Lennard-Jones Potential

One of the simplest two-body potentials was proposed by Lennard-Jones [5] in 1924. It combines a short-range repulsive term to prevent the substance from collapsing in on itself with a longer range, attractive, dispersion term which prevents the substance from disintegrating. A popular version of the potential chooses to make the short-range term proportional to $1/r^{12}$ and the dispersion term proportional to $1/r^6$, the latter having some theoretical justification as it matches the long range behaviour of van der Waals forces [59]. The potential is thus given by

$$V_2(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{2.29}$$

where σ is the distance to the zero in $V_2(r)$ and ϵ is the energy at the minimum as depicted in Fig. 2.2 The Lennard-Jones potential is widely used for ionic and metallic systems, but it is too simplistic to accurately model the behaviour of atoms in a covalent material.

Stillinger-Weber Potential for Silicon

The Stillinger-Weber (SW) potential [6] was one of the first attempts at modelling a semiconductor material with an interatomic potential. Since it was proposed in 1985 it has been very widely used for a great variety of applications including lattice dynamics [60], point defects [61], surfaces [62], and the liquid state [63]. It consists

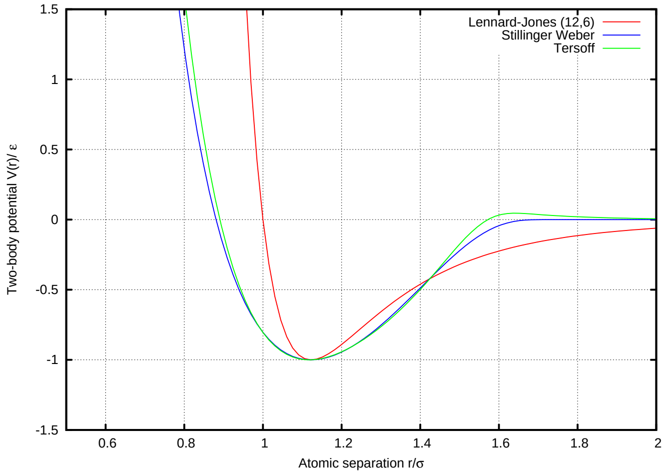


Figure 2.2: The Lennard-Jones $(12, 6)$, Stillinger-Weber and Tersoff two-body potentials in terms of the reduced units r/σ and V_2/ϵ. The Tersoff bond-order potential $B(\zeta)$ has been evaluated for a representative atom in a diamond bulk structure with the cohesive energy scaled to match the Lennard-Jones well depth.

of two- and three-body potentials of the form

$$V_2(r) = \epsilon\, f\left(\frac{r}{\sigma}\right) \tag{2.30}$$

$$V_3(r_{ij}, r_{ik}, \theta_{jik}) = \epsilon\, g\left(\frac{r_{ij}}{\sigma}\right)\, g\left(\frac{r_{ik}}{\sigma}\right)\, h\left(\theta_{jik}\right) \tag{2.31}$$

where ϵ sets the energy scale and σ the distance scale.

The functional forms of $f(r)$ and $h(\theta)$ are

$$f(r) = \left(\frac{A}{r^4} - B\right) f_{\mathrm{cut}}(r) \tag{2.32}$$

$$h(\theta) = \lambda \left(\cos\theta - \cos\theta_0\right)^2 \tag{2.33}$$

where $f_{\mathrm{cut}}(r)$ is a cutoff function that is zero for $r > \sigma r_c$, and A, B and λ are parameters. $f(r)$ is shown in Fig. 2.2 for comparison with the Lennard-Jones potential. Its range is just below the second neighbour distance in the diamond lattice and there is a deep well at the first neighbour distance to represent the restoring force of stretched sp^3 bonds. The value of $\cos\theta_0$ is set to $-\frac{1}{3}$ so that $h(\theta)$ penalises angles away from the ideal tetrahedral bond angle; this is a simple way of describing the directional sp^3 bonding. The radial parts $g(r)$ of the three-body term decrease its effect when bonds are stretched. $g(r)$ contains a similar cutoff function to $f(r)$, with

	Expt	DFT	SW	Tersoff	EDIP
C_{11}	166	159	151	122	175
C_{12}	64	60	76	86	62
C_{44}	80	85	56	10	71
C_{44}^0	-	111	109	92	112

Table 2.1: Elastic constants of diamond structure silicon, in units of GPa. Experimental values are from Harrison [58]. DFT calculations were performed using CASTEP [24] with the GGA approximation and a cutoff energy of 400 eV. SW is the Stillinger-Weber potential [6]. The Tersoff potential is the version proposed in Ref. 7 and the EDIP values are from Justo *et al.* [14].

the same range.

The parameters in the Stillinger-Weber potential were determined by fitting to experimental data, with the requirement that the diamond structure be the most stable at low pressure and that the melting point agrees reasonably with experiment. The SW potential is qualitatively successful at representing the condensed phases of silicon, and does a reasonable job for the liquid phase.

It is possible to calculate the elastic constants for the SW potential analytically [64], and this gives excellent agreement with the procedure outlined in Section 2.2. The potential gives a very good description of the elastic properties of silicon in the diamond structure, despite not having been fit to these. This is because the potential is constrained by its simple functional form to nearly obey the Harrison condition of Eq. 2.25, regardless of the values of the empirical parameters.

The SW potential has been extended to model the interaction of hydrogen with silicon surfaces [65] by introducing additional two- and three-body parameters for the Si-H interactions. Parameters to describe the diamond polytype of carbon have been determined using *ab initio* calculations [66]; since these are not derived from experiment this version of the potential cannot really be considered to be empirical, but this makes little practical difference. For the theoretical reasons outlined above, the carbon parametrisation of the SW potential gives a good description of the elastic properties of diamond.

Tersoff and Brenner Potentials

The Tersoff potential [7, 8, 9] is another very popular empirical model for silicon, based on the *bond order* concept, which is that the strength of bonds is affected by the presence of surrounding atoms. Like the SW potential, the Tersoff potential has also been widely applied, for example to lattice dynamics [60], point defects [8, 9] and the liquid and amorphous phases [67].

The Tersoff potential is the sum of a repulsive pair potential $\phi_R(r_{ij})$ and an attractive potential $\phi_A(r_{ij})$ modified by a bond order function $B(\zeta_{ij})$:

$$V = \sum_{\substack{i,j \\ i<j}} \phi_R(r_{ij}) + B(\zeta_{ij})\phi_A(r_{ij}) \tag{2.34}$$

ζ_{ij} is essentially a weighted measure of the number of bonds competing with the bond ij, and is formed from a sum of three-body terms similar to Eq. 2.31, but without a fixed minimum in the angular part $h(\theta)$ at the tetrahedral angle. Whilst the potential resembles a two-body model, the functional form includes many-body interactions in the bond order term. The bond-order potential is approximately given by $B(\zeta_{ij}) = \zeta_{ij}^{-\delta}$, with δ a parameter to be fitted.

The original version [7] of the potential had only six adjustable parameters, while subsequent versions have added a further seven to improve the elastic properties. The parameters were fitted to *ab initio* calculations on the various polytypes of silicon. A representative plot of the effective two-body Tersoff interaction when the atomic neighbourhood is a perfect bulk crystal is shown in Fig. 2.2 — the potential very closely resembles the SW potential in this case. As can be seen in Table 2.3, despite its greater theoretical justification due to the use of the bond order, the Tersoff potential does not give as good as description of the elastic properties of bulk silicon as the SW potential; this is because it does not satisfy the Harrison condition (Eq. 2.25) as closely as the SW potential.

Soon after the Tersoff potential was developed, an extension to allow simulation of hydrocarbon materials was proposed by Brenner [10]. In its initial form the Brenner potential was a simple extension of the Tersoff model to correct the overbinding of radicals which results in a non-physical description of the bonding process in many common situations, for example vacancy formation. This can be achieved by rewriting the Tersoff potential in the form

$$V = \sum_{\substack{i,j \\ i<j}} \phi_R(r_{ij}) + \bar{B}_{ij}\phi_A(r_{ij}) \tag{2.35}$$

where the Tersoff bond order factor $B(\zeta_{ij})$ has been replaced by

$$\bar{B}_{ij} = \frac{B(\zeta_{ij}) + B(\zeta_{ji})}{2} + F_{ij}. \tag{2.36}$$

It is formed by averaging B_{ij} and B_{ji} and adding a correction F_{ij} for bonds between pairs of atoms with different coordinations.

A second generation Brenner potential was proposed in 2000 [11] and further developed in 2002 [12]. This potential describes atomic hybridisations, allowing for covalent bond breaking and formation, and thus providing a better description of the elastic, defect and surface properties of diamond. The revised potential is primarily intended, however, for the simulation of hydrocarbon molecules and the additional complexity of the potential and number of fit parameters required is considerable. For many applications, particularly those focused on the solid state, the simplicity of the earlier potential outweighs the gains in transferability of the newer potential, and so the 1990 potential continues to be widely used. Recent applications of the Brenner potential include simulations of the mechanical properties of nanotubes [68], modelling of amorphous carbon [69] and calculation of the elastic properties of graphene [70].

Environment Dependant Interatomic Potential

The Environment Dependant Interaction Potential (EDIP) [13, 14] is a development of the SW potential which incorporates some of the environment dependent ideas of Tersoff-like potentials, through an effective coordination number for each atom. The potential has more adjustable parameters than either SW or Tersoff, and was fit to *ab initio* calculations of bulk phases and defect structures. The simple functional form guarantees an accurate description of the elastic properties of silicon in the bulk phase, and the coordination dependent extensions include the essential features of chemical bonding well enough to describe various bulk defects, and the liquid and amorphous phases.

The added complications make EDIP more difficult to implement than the SW potential, and more computationally expensive to evaluate. Its real benefit is an improved description of the energetics of defects, since the bulk elastic behaviour is essentially unchanged from that of the SW model. EDIP has been widely applied, for example to compute thermodynamic properties [71], elastic properties of the amorphous phase [72] and recently to model nanoparticle surface interactions [73].

EDIP has recently been generalised to allow it to be used for carbon [15], with additional coordination functions to describe dihedral rotation, non-bonded π-repulsion and fractional coordination. This gives a transferable potential with good elastic constants and an accurate description of the molten state. However, even more so than for silicon, the generalisations make the potential more difficult to implement and significantly more computationally expensive, and thus it has not been as widely applied as the simpler but less transferable Brenner potential.

Implementation of Interatomic Potentials

The most time consuming part of a molecular dynamics simulation is evaluating the forces on all the atoms. For practical purposes, potentials are truncated at a finite cutoff radius r_c in order to reduce the computational effort required to evaluate them. This means that we only need to evaluate a small subset of the $\frac{1}{2}N(N-1)$ unique interatomic distances, since most pairs of atoms (i, j) will be separated by a distance $r_{ij} > r_c$ and will not contribute to the potential.

The force evaluation can be made more efficient by maintaining a list of all the neighbours $\{j\}$ of each atom i together with their separations $\{r_{ij}\}$. We include all atoms in this neighbour list that lie within a distance r_C of the atom, with $r_C > r_c$. The *neighbour crust*, equal to $r_C - r_c$, should be chosen so that atoms cannot cross it in one time step, *i.e.* $r_C - r_c > \langle \dot{\mathbf{r}} \rangle \Delta t$.

Limitations of Classical Potentials

Despite the huge amount of effort expended on the development of classical interatomic potentials in recent decades, in many circumstances they do not provide sufficient accuracy for molecular dynamics simulations. This may be because they are being used outside the regime in which they were fitted, or because they are too simple to describe the physical processes taking place. Often this occurs when bonds break during chemical reactions: a classical potential cannot be used to model the entire system unless it can be made accurate enough to describe the full range of chemical reactions that may take place.

Recently, there has been much work to develop *reactive* classical potentials capable of high accuracy for particular materials and specific classes of reactions. These go some way to overcoming the limitations of traditional classical potentials. The revised version of the Brenner hydrocarbon potential [11] discussed above is an example of the large amount effort required to do this. Other examples include the ReaxFF [16] force field and many potentials based on the bond order idea [17, 18].

The effort required to create and test a reactive potential is enormous. These potentials attempt to be transferable, but the fitting process still has to be repeated for new chemical environments. Moreover, the potential is not demonstrably correct apart from where it has been fitted, in contrast to the *ab initio* approach discussed below. For many systems, reactive potentials are either too difficult to produce or insufficiently robust to be useful. In these cases we must use a first principles quantum mechanical method, as described in the following section.

2.4 The *Ab Initio* Method

Over the last twenty years, first principles methods have developed to the extent that it is now possible to solve the many-body Schrödinger equation approximately for systems containing hundreds of atoms with a sufficient level of detail to allow the electronic structure of a system to be calculated. Since no empirical parameters are required, these calculations do not suffer from the transferability problems of interatomic potential, and, to a lesser extent, tight binding based approaches. In this section we give only a brief overview of the *ab initio* method; more details can be found in, for example, Payne *et al.* [23].

The time-independent, non-relativistic, many-body Schrödinger equation for a system of N interacting electrons and M nuclei is

$$H|\Psi\rangle = E|\Psi\rangle \tag{2.37}$$

where E is the energy eigenvalue of the many-body wavefunction $|\Psi\rangle$ and the Hamiltonian H is given by

$$\begin{aligned} H \;=\; & -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \frac{1}{2}\sum_{\substack{i,j\\ i<j}}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|} - \sum_{i,I}\frac{Z_I}{|\mathbf{r}_i-\mathbf{R}_I|} \\ & -\frac{\hbar^2}{2m_I}\sum_I \nabla_I^2 + \frac{1}{2}\sum_{\substack{I,J\\ I<J}}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|} \end{aligned} \tag{2.38}$$

in atomic units. The indices i and j are summed over electrons, and I and J over nuclei, which have positions $\{\mathbf{r}_i\}$ and $\{\mathbf{R}_I\}$ respectively. Z_I and m_I are the atomic number and mass of the I^{th} nucleus. The first term is the electronic kinetic energy, and the fourth the ionic kinetic energy. Electron-electron interaction is described by the second term, electron-ion interaction by the third and the last term represents the ion-ion interactions.

In principle, solving Eq. 2.37 to get all the eigenstates and eigenvalues of the Hamiltonian would allow all the physical properties of the system to be calculated. However, an analytical solution cannot be obtained for systems containing more than one electron: we clearly need to approximate in order to proceed.

The approximations required include the Born-Oppenheimer approximation to simplify the Hamiltonian, density functional theory to model the electron-electron interactions and pseudopotential theory to model the electron-ion interactions. Each of these is described briefly in the following sections.

The Born-Oppenheimer Approximation

An essential step along the road to a practicable method for approximately solving the many-body Schrödinger equation is the well known Born-Oppenheimer approximation [74], proposed in 1927. Since the mass of the electron is several orders of magnitude smaller than that of the nuclei, they move on much faster time scales. We can therefore assume that the as the nuclei move, the electrons immediately relax to their instantaneous eigenstate. In this case, the motion of the electrons can be decoupled adiabatically from that of the ions. Eq. 2.38 reduces to the electronic Hamiltonian

$$H = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \frac{1}{2}\sum_{\substack{i,j\\ i<j}}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}. \tag{2.39}$$

This approximation applies to high accuracy in most solid state systems. It breaks down when there is significant *vibronic coupling* between the nuclear and electronic degrees of freedom. This is common in linear molecules where symmetry gives rise to degeneracy of the electronic states. The first calculation to go beyond the adiabatic approximation and to include extra terms to take account of these coupling effects was performed by Renner [75] in 1934 for the case of an electronically excited Π-state in CO_2.

Density Functional Theory

Density functional theory (DFT), developed by Hohenberg, Kohn and Sham in the mid 1960s, proves that the ground state energy of an electron gas is a unique and universal functional of the electron density [19], and that it is possible to replace the many-electron problem by an equivalent set of self-consistent one effective particle problems [20]. Within this framework the energy E of a set of N electrons in an external potential $V_{\mathrm{ext}}(\mathbf{r})$ is

$$E\left[n(\mathbf{r})\right] = F_{HK}\left[n(\mathbf{r})\right] + \int V_{\mathrm{ext}}(\mathbf{r})\, n(\mathbf{r})\; \mathrm{d}^3\mathbf{r} \tag{2.40}$$

where square brackets denote functionals, $n(\mathbf{r})$ is the electron density and F_{HK} is the unique Hohnberg-Kohn functional. Variational minimisation of this equation leads to the ground-state density $n_0(\mathbf{r})$ and ground-state energy E_0. The second central principle of DFT is now applied: we rewrite the density in terms of a set of orthonormal non-interacting single particle functions $|\phi_i\rangle$:

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2 \tag{2.41}$$

We separate the unknown functional $F_{HK}\left[n(\mathbf{r})\right]$ into a sum of three contributions: the kinetic energy of these non-interacting electrons, the Hartree electron-electron interaction energy and a final term which captures the effects of electron exchange and correlation:

$$F\left[n(\mathbf{r})\right] = T_S\left[n(\mathbf{r})\right] + E_H\left[n(\mathbf{r})\right] + E_{XC}\left[n(\mathbf{r})\right]. \tag{2.42}$$

The kinetic energy T_S and the Hartree energy E_H can be written exactly:

$$\begin{aligned} T_S\left[n(\mathbf{r})\right] &= -\frac{1}{2}\sum_i \int \phi_i^*(\mathbf{r})\nabla^2\phi_i(\mathbf{r})\; \mathrm{d}^3\mathbf{r} \\ E_H\left[n(\mathbf{r})\right] &= \frac{1}{2}\int\int \frac{n(\mathbf{r})\, n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\; \mathrm{d}^3\mathbf{r}\, \mathrm{d}^3\mathbf{r}' \end{aligned}$$

Minimising this energy functional with respect to each of the $|\phi_i\rangle$ leads to a set of N single particle Schödinger equations, called the Kohn-Sham equations

$$\left(-\frac{1}{2}\nabla^2 + \underbrace{V_H(\mathbf{r}) + V_{\text{ext}}(\mathbf{r}) + V_{XC}(\mathbf{r})}_{=V_{KS}(\mathbf{r})} \right) \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \tag{2.43}$$

where V_H and V_{XC} are obtained by taking functional derivatives of E_H and E_{XC}, and V_{KS} is the Kohn-Sham effective potential.

These equations are exact, but to make progress we must approximate E_{XC}. The choice of this functional is the only uncontrolled approximation in the DFT formalism and the development of improved approximations to the functional has been an active topic of research since the method was first proposed. Since we've now separated out all the dominant long-range contributions to the energy, E_{XC} can be approximated, either as a local functional of the density in the Local Density Approximation (LDA) or of the density and its gradient in the Generalised Gradient Approximation (GGA). The GGA is more robust in systems where the electron density varies rapidly.

Both LDA and GGA are parametrised using accurate but computationally very expensive quantum Monte Carlo (QMC) calculations on the electron gas at a range of densities. Some GGA parametrisations are additionally fit to experimental data for a large number of atomic and molecular systems; these are referred to as hybrid functionals, with the BLYP functional being the most widely used example [76]. In this work I shall use the popular Perdew, Burke and Ernzerhof (PBE) GGA functional which has no semi-empirical parameters apart from those of the local spin density approximation from which it is derived [21].

Basis Set and $\mathbf{k}$-point Sampling

To solve the Kohn-Sham equations numerically, we need a basis set in which to represent the wavefunctions. Possible basis sets fall into two broad categories: those that are atom centred, *e.g.* gaussians or linear combinations of atomic orbitals (LCAO), and basis sets derived from planewaves that have the same accuracy throughout the system. Localised basis sets allow an efficient treatment of molecular systems, since the number of states required will often be far fewer than that required to give the same accuracy with a planewave calculation (however the operations are much more expensive per basis function). Thus they are popular in quantum chemistry and for biomolecular calculations. For periodic systems, planewaves are often preferred as they provide a mathematically straightforward representation, and only have to be converged with respect to a single parameter, the total number of plane waves in the basis set, usually represented in terms of a cutoff energy for the plane waves retained in the set. Since they treat all areas of the system equally, plane waves are free of the bias that can afflict calculations using localised basis sets. In this work I will use a planewave basis set, as implemented in the CASTEP package [24].

Bloch's theorem tells us that the eigenstates of a one-electron Hamiltonian in a periodic potential, with $V(\mathbf{r}+\mathbf{R}) = V(\mathbf{r})$ for all Bravais lattice vectors $\mathbf{R}$, are of the form

$$\phi_{i,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_i(\mathbf{r}) \tag{2.44}$$

where the $\{u_i\}$ are periodic Bloch functions. We can expand these functions as a sum over the reciprocal lattice vectors $\mathbf{G}$ of the periodic system, giving

$$\phi_{i,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{i,\mathbf{k}+\mathbf{G}}\, e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{2.45}$$

where the $c_{i,\mathbf{k}+\mathbf{G}}$ are the expansion coefficients for the i^{th} band. Since these coefficients are typically smaller the larger the planewave kinetic energy, we can truncate this sum at some cutoff energy radius, with $\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 \leq E_{\text{cut}}$, and then converge the basis set by increasing E_{cut} until the final total energy of the system does not change.

In principle, once we have solved the Kohn-Sham equations, we can calculate physical properties such as the total energy by integrating over all $\mathbf{k}$ in the first Brillouin zone. In practice it turns out to be sufficient to use a finite $\mathbf{k}$-point sampling scheme, since the Kohn-Sham eigenstates, eigenvalues, and electron density are all smoothly varying functions of $\mathbf{k}$. In this work I will use the Monkhorst-Pack sampling scheme [77]. This consists of a uniform grid of $\mathbf{k}$-points that follow the shape of the Brillouin zone. Symmetries of the supercell can be used to further reduce the number of $\mathbf{k}$-points required. If a system is very large, then the Brillouin zone volume will be very small, so a smaller number of $\mathbf{k}$-points are required to adequately sample it. In the limit of a very large system, we can consider only a single point, often the Γ point for which $\mathbf{k} = 0$.

Since the Bloch formalism only applies to periodic systems, if we wish to study non-periodic systems such as molecules, defects or surfaces then we have to simulate

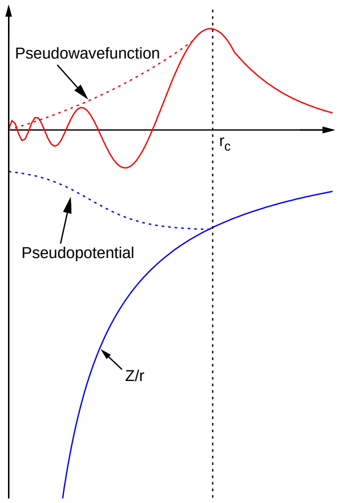


Figure 2.3: Schematic illustration of the all-electron potential and the pseudopotential, and the corresponding wavefunctions.

non-periodicity using the supercell approximation: our periodic system is a regular array of widely spaced copies of the non-periodic system, with enough vacuum-padding to prevent spurious interaction between images.

The Pseudopotential Approximation

Close to the nucleus, the requirement for the valence and core electron wavefunctions to be orthogonal causes rapid variations in both wavefunctions and oscillations in the valence wavefunction. A very large number of planewaves would be required to accurately describe these oscillations. We can avoid this by replacing the true all-electron potential with a much weaker, smoother, pseudopotential in this region, which takes into account the combined effect of the nucleus and the core electrons, removing these oscillations. This is illustrated in Fig. 2.3. The resulting ‘pseudowavefunction’ can be represented to the same accuracy by a basis containing far fewer planewaves, so the computational cost of the calculation is reduced.

Pseudopotentials are normally constructed from DFT calculations on isolated atoms. The valence wavefunctions are modified in the core region to remove the rapid oscillations, then the Schrödinger equation is inverted to find the corresponding pseudopotential. It is common to require that the core region of the all-electron and pseudowavefunctions must enclose the same charge; this leads to what are known as *norm-conserving pseudopotentials* [78]. Providing the pseudo- and all-electron wavefunctions match outside the core region, norm-conserving pseudopotentials result in the correct first-order variation of phase shifts with energy, giving an accurate description of scattering over a wide range of energies.

Relaxing the norm-conservation constraints leads to an alternative class of pseudopotentials, called *ultrasoft pseudopotentials*, in a procedure first proposed by Vanderbilt [79]. They are smoother than norm-conserving pseudopotentials, so require even fewer planewaves to describe the pseudowavefunctions; typically the cutoff energy E_{cut} can be reduced by about a factor of two. A charge augmentation function acting within the core region is used to compensate for the relaxation of the norm-conservation constraint. Ultrasoft pseudopotentials no longer predict the correct dependence of phase shift on energy, so it is common to interpolate between pseudopotentials parameterised at several reference energies.

Minimisation of the Total Energy

The Kohn-Sham equation (Eq. 2.43) must be solved self-consistently to obtain the ground-state electron density and hence the ground-state total energy. The original approach was to do this directly by matrix diagonalisation, however this is very time consuming and inefficient since the computational cost of diagonalisation scales as the cube of the number of plane waves but only a small number of the solutions require occupied Kohn-Sham orbitals.

The Car-Parrinello method [80] improves upon this by iteratively improving a trial wavefunction using fictitious electron dynamics to minimise the total energy. This process can be greatly sped up by instead using a preconditioned conjugate gradient method to minimise the energy function [22]. The overall scaling of the calculation is still cubic in both cases since maintaining orthonormalisation of the wavefunctions requires an integral with cost proportional to N to be evaluated for each of the $\frac{1}{2}N(N-1)$ orbital pairs. The conjugate gradient approach is taken in the CASTEP package used in this work.

The Hellman-Feynman theorem

In order to perform molecular dynamics or carry out geometry optimisation using the *ab initio* method, we need to evaluate the quantum mechanical forces acting on the ions. The Hellman-Feynman theorem [81, 82] simplifies the evaluation of these forces; it shows that the partial derivative of the total energy with respect to the Kohn-Sham wavefunctions vanishes when we sum over all bands, so the force on the I^{th} nucleus at position $\mathbf{R}_I$ is simply given by

$$\begin{aligned}\mathbf{F}_I &= \frac{\partial E_{KS}}{\partial \mathbf{R}_I} \\ &= \sum_i \langle \phi_i | \frac{\partial H_{KS}}{\partial \mathbf{R}_I} | \phi_i \rangle\end{aligned}$$

Once these forces have been calculated, geometry optimisation or molecular dynamics can be carried out in a straightforward fashion. This approach is in contrast to the Car-Parinello Molecular Dynamics (CPMD) method, where the fictitious electronic and real ionic degrees of freedom are integrated simultaneously. This requires a far smaller timestep but it is not necessary to solve the electronic problem to full self-consistency at each step.

2.5 Tight Binding

Originally proposed by Slater and Koster in 1954 [83], the Tight Binding (TB) method [84] lies in between fast but inaccurate empirical potential modelling and very expensive, very accurate *ab initio* techniques. Since the TB method retains essentially a quantum mechanical description of bonding, it is useful where quantum mechanics is needed but a large system size means *ab initio* calculations are not feasible.

Semi-Empirical Orthogonal Tight Binding

The TB method works by writing eigenstates of the Hamiltonian in an atomic-like basis set, and replacing the exact many-body Hamiltonian with a parametrised Hamiltonian matrix H_{TB}. The total electronic energy is a sum of the energy eigenvalues of the filled states:

$$E_{\text{band}} = \sum_{i=1}^{N} \langle \psi_i | H_{TB} | \psi_i \rangle = \sum_{i=1}^{N} \epsilon_i. \tag{2.46}$$

To apply tight binding as a molecular dynamics technique, we must choose a continuous functional form for the distance dependence of the matrix elements of H_{TB}.

We also add a repulsive energy term to account for ion-ion repulsion and the over counting of the electron-electron repulsion in E_{band}. Typically this term is formed from a sum of pair potentials. Adding this repulsive term to the ionic kinetic energy, we arrive at the TB molecular dynamics Hamiltonian:

$$H = \sum_{I=1}^{M} \frac{\mathbf{p}_I^2}{2m_I} + E_{\text{band}} + E_{\text{rep}} \tag{2.47}$$

where I is summed over the M ions. In the orthogonal tight binding method, the basis functions $\{\psi_i(\mathbf{r})\}$ are formed from the non-orthogonal but short-ranged atomic orbitals in such a way as to preserve their symmetry properties; the resulting basis functions are known as Löwdin functions. We assume that these functions are also short-ranged. The eigenvalues ϵ_i and eigenvectors of H_{TB} can be evaluated using exact matrix diagonalisation which means the scaling with the number of atoms is $O(N^3)$ and the method is practical for around a thousand atoms.

Once we have diagonalised H_{TB}, we calculate the forces on the ions by evaluating the density matrix ρ in either real space or basis space and taking the derivative with respect to an ionic position $\mathbf{R}_I$. For a real space density matrix the energy and force are

$$\begin{aligned} E_{\text{band}} &= 2\,\mathrm{Tr}\left[\rho H_{TB}\right] \\ \mathbf{F}_I &= -2\,\mathrm{Tr}\left[\rho \frac{\partial H_{TB}}{\partial \mathbf{R}_I}\right] \end{aligned}$$

Early TB implementations [85] used simple repulsive pair potentials proportional to $1/r^2$, and fitted to bulk elastic properties. Atomic energies and hopping matrix elements were fit to band structures derived from *ab initio* calculations. This gives an accurate description of crystalline systems but is not very transferable. Goodwin *et al.* [86] suggested using short-range scaling functions to control the range of the TB interactions in order to make calculations feasible for larger systems. Further developments to these scaling functions gave rise to more broadly applicable TB potentials [87, 88] for silicon. However transferability is still limited by the need to fit to well defined chemical surroundings.

Extensions to Orthogonal Tight Binding

Self-consistency

Commonly TB is used non self-consistently: the Hamiltonian matrix does not depend on the distribution of electrons. This can cause problems when net charge transfer between atoms is important, for example in materials with a large degree of ionic bonding, or when spurious charge transfer — that should be prevented by Coulombic screening — arises. The simplest approach to overcome this problem is Local Charge Neutrality (LCN), where the on-site energies are adjusted by a local offset chosen iteratively so that the number of electrons on each atom balances the ionic charge. LCN works well for metallic systems where the atoms do not carry significant charge because the screening length is typically less than the interatomic spacing. If there is a greater amount of charge transfer then an extra correction term can be added to the on site energies in the Hamiltonian. The correction to the energy is

$$\delta E_I = U(Q_I - Z_I) \tag{2.48}$$

where Q_I is the number of electrons on the atom (evaluated as a sum of the occupation factors of each eigenstate), Z_I is the nuclear charge, and U is the strength of the correction, referred to as the Hubbard U. This introduces an energy cost for charging an atom, which acts to prevent unphysical charge transfer.

Non-orthogonal Tight Binding

The orthogonal TB method outlined above has many advantages and is a very popular simulation technique, allowing systems containing thousands of atoms to be modelled in a way that includes quantum many-body effects which are totally inaccessible to an empirical interatomic potential. However a major drawback of the method, in common with empirical potentials, is its lack of transferability. For example, a parametrisation developed for crystalline silicon is unlikely to accurately describe the liquid phase without expanding the fitting data but if this were tried, the large dataset required would result in a dramatic drop in overall accuracy.

One obvious improvement consists of incorporating the overlap integral between basis functions, by solving for the eigenstates ψ_i and energies E_i of the generalised eigenvalue problem

$$(H - E_i\, S)\, \psi_i = 0 \tag{2.49}$$

where H and S are the Hamiltonian and overlap matrices in the atomic basis set representation. This approach is used in a variety of non-orthogonal tight binding methods, which have enjoyed considerable success in improving the accuracy of TB calculations for semiconductor systems, for example in silicon clusters [89].

Density Functional Tight Binding

Density functional tight binding (DFTB) is based on a more sophisticated second-order expansion of the Kohn-Sham energy with respect to fluctuations in charge density [90, 91]. However, it retains many aspects of the traditional TB formalism, such as tabulated Hamiltonians and the two-centre approximation to the hopping integrals.

An atomic basis set $\{\phi_i(\mathbf{r})\}$ is used to solve a simplified Kohn-Sham equation (c.f. Eq. 2.43) with a pseudo-atom potential consisting of a nuclear potential, the Hartree potential V_H and the exchange correlation potential V_{XC}

$$\left[-\frac{1}{2}\nabla^2 + V_H(\mathbf{r}) + V_{\mathrm{nuc}}(\mathbf{r}) + V_{XC}(\mathbf{r}) + \left(\frac{r}{r_0}\right)^N\right]\phi_i = \epsilon_i \phi_i(\mathbf{r})\,. \tag{2.50}$$

The purpose of the last term in the potential is to concentrate charge density closer to the nucleus to make bulk calculations easier. The solutions for the $\{\phi_i\}$ are used to tabulate Hamiltonian matrix elements. To complete the model the repulsive pair potential is calculated for each atomic configuration by taking the energy difference between the band energy given by this method and that of a full self-consistent solution to the modified Kohn-Sham equation.

Despite its relative simplicity, the DFTB method does a good job of producing transferable tight-binding parametrisations. Since the Kohn-Sham energy also includes the Coulomb interactions between charge fluctuations long range electrostatic forces are handled correctly. By modifying the DFTB matrix elements to redistribute Mulliken charges, it is possible to include charge self-consistency effects into the model [92]. This further improves the method's transferability and allows it to be applied to organic and biological systems where charge transfer and long-range Coulomb interactions are important so traditional non-self-consistent TB cannot be used.

2.6 Summary

This chapter has outlined the major ingredients that are fundamental to this thesis: classical molecular dynamics and accurate quantum mechanical calculation of forces. In the next chapter I will look at how these can be combined to create a multiscale hybrid model, before describing the 'Learn on the Fly scheme' in Chapter 4 and then going into more detail on the specific problem to be attacked with this method in Chapter 5.

Chapter 3

Hybrid Simulation

3.1 Introduction

Over the last twenty years, the *ab initio* methods described in the previous chapter have made modelling of simple semiconductor systems reliable, accurate and routine. This is partly due to the significant increase in capacity and speed of available computers and partly to the development of high quality codes that make effective use of these resources. As a result, attention is now focusing on modelling larger scale, more complex systems. Representative examples from fields as disparate as

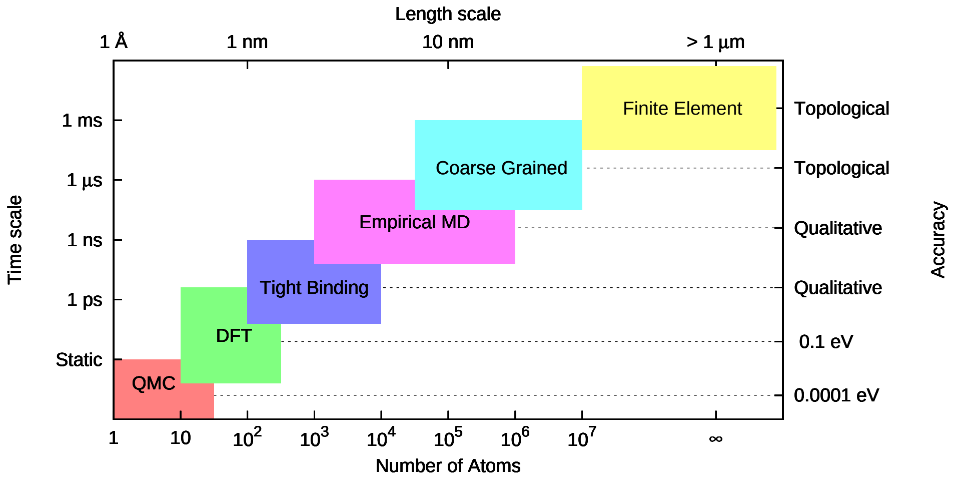


Figure 3.1: Schematic representation of the range of length- and time-scales accessible to a variety of modelling methods, from quantum Monte Carlo (QMC) for very accurate, very expensive static calculations through to approximate methods such as finite-element modelling.

biology and materials science include enzyme catalysis [28] and defect migration in semiconductors; both of these systems are described in more detail below. Ideally, we would simulate such systems using entirely first principles methods, free of empirical parameters and the accuracy and transferability problems associated with them. However, *ab initio* molecular dynamics is limited to simulating a few hundred atoms for up to a few picoseconds. Using more approximate methods (*e.g.* tight binding), the number of atoms can be extended to perhaps thousands, or the time period increased by a few orders of magnitude, but for many problems this is still insufficient. Fig. 3.1 illustrates the approximate range of application of various modelling techniques and makes clear the challenges we face if we wish to model complex problems with high accuracy.

Hierarchical Modelling

In recent years, there has been a great deal of work on multiscale methods that attempt to apply accurate quantum calculations to larger systems in one way or another. Most commonly, such methods are examples of *hierarchical multiscale modelling*, where the results of a calculation at one scale are used to parameterise less accurate calculations at a larger scale, making bigger systems or longer simulation times possible [1].

For many biological processes, the relevant timescale is of the order of milliseconds, well beyond the capability of traditional molecular dynamics. To make progress we can either form a hierarchical multiscale model by coarse-graining the system and considering the dynamics of the aggregate particles, or we can try to extract activation energies and reaction pathways from static calculations or short MD runs to parameterise Monte Carlo models. For a review of hierarchical multiscale methods and examples of their application, see Nieminen [25].

Simultaneous Modelling

There is a large class of problems where the physical processes on the various length scales are strongly coupled and cannot be separated into a series of independent calculations; often this is because the nanoscale phenomena is driven by forces determined at least partially on the macroscopc scale. Simulation of such systems

---

[1] We have seen this approach already in Section 2.3 — for example the Tersoff potential is a classical potential fitted to *ab initio* data, capturing aspects of the quantum mechanical bonding.

requires *simultaneous* coupling of length scales. Over the last ten years there has been much effort to devise schemes, referred to as *hybrid* or *embedded* methods, that combine a range of modelling techniques into a single simulation.

Occasionally, the large scale processes are so simple that they can be simulated very easily, as an example Martoňák *et al.* [93] added a classical pressure reservoir of soft spheres to an *ab initio* simulation of a small molecule. Usually, however, the large scale behaviour requires a more complex model to accurately capture the physics; this will be assumed to be the case for the remainder of the work discussed in this chapter.

Multiscale Applications in the Solid State

Stress induced defect processes in metals and semiconductors often give rise to strongly coupled multiscale phenomena. Examples include point-defect diffusion, dislocation motion, grain boundaries and, of course, the prototypical multiscale modelling problem: fracture.

Point-Defect Diffusion The stability and migration of point defects in semiconductors is affected both by local chemical interactions and long range strain fields. An example long range effect is the strain field resulting from the lattice mismatch between epitaxial layers in semiconductors. Although the quantum mechanical treatment of the bonding rearrangement around a defect requires only a few hundred atoms, we would need to include thousands more atoms to accurately represent the inhomogeneous strain environment, particularly if we are to model interactions between multiple defects.

Dislocation Motion The strength of many materials is dominated by the behaviour of their dislocations. The core of a dislocation is a 1D region in which the bonding is significantly distorted. Dislocations in covalent materials move by the formation of kinks in the dislocation, where the bonding is very highly distorted. As the kink moves, so does this region of distortion. This motion requires bond breaking and reformation, therefore this region should be modelled by a highly accurate quantum mechanical technique.

Grain Boundaries It is not always possible to assume perfect single crystal structure and ignore the effect of grain boundaries when studying the physical and electronic properties of semiconductors. This is true for many materials of growing technological relevance, for example gallium nitride, silicon carbide and diamond [94]. Grain boundaries change the crystal structure on two length scales: they introduce long-range elastic distortion and local bonding disorder. They also act as sinks and sources for dislocations and traps for dopants, electrons and holes, further increasing the local chemical complexity. A multiscale description is needed to describe these systems since empirical interatomic potentials describe the long range-interactions adequately but local rebonding requires quantum mechanical accuracy.

Brittle Fracture As discussed in the introduction to this thesis, fracture is perhaps the best example of a multiscale materials process. It is the prototypical problem that has spurred many recent advances in the field of hybrid modelling of materials systems, and thus reviewing these developments is most easily accomplished by briefly considering the history of multiscale fracture simulation. We will delay a full discussion of fracture to Chapter 5, and in this chapter concentrate only on its multiscale nature.

3.2 Coupling Continuum and Atomistic Systems

In this thesis, I concentrate on hybrid schemes which link quantum mechanical and classical modelling, but we shall first look briefly at a larger length scale. The pioneering hybrid simulations of materials systems were performed by Kohlhoff *et al.* [26], where classical atomistic and continuum elastic models were coupled to successfully model the directional cleavage anisotropy of a BCC crystal. This approach has been developed in the *quasicontinuum* (QC) method of Tadmor *et al.* [27].

The key problem with coupling atomistic and continuum models of matter is finding ways to connect these conceptually very different descriptions. Atomic positions need to be mapped onto a continuous displacement field, and energy calculations from interatomic potentials in the atomistic region and constitutive laws in the continuum region need to be harmonised. In the QC approach, a small subset of the atoms that would appear in a fully atomistic model are selected to represent the system as a whole, with a higher sampling density in highly deformed regions. The system is divided into cells, with one representative atom in each cell, as illustrated in Fig. 3.2. We assume that the energy of all the atoms in each cell is the same as that of the representative atom. The energies of these representative atoms are

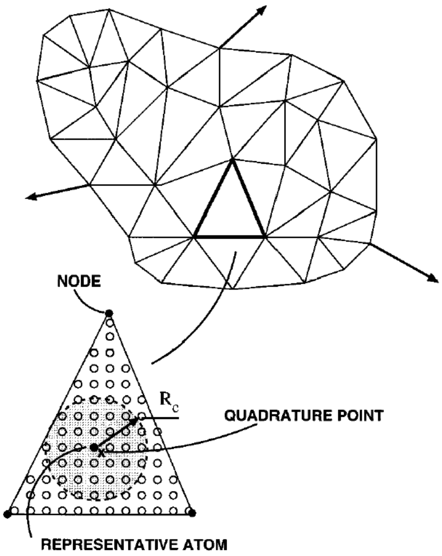


Figure 3.2: Schematic illustration of the finite element discretisation of a solid in the quasicontinuum method. The lower panel shows the representative atom for a particular triangular element. Reproduced from Tadmor *et al.* [27]

computed from the local environment, either from constitutive laws in areas that are nearly homogeneously deformed, or fully atomistically for non-uniformly deformed regions.

The atomistic and continuum methods are not completely compatible: non-physical forces arise on the continuum side of the boundary since it looks like an artificial surface. The atomistic interactions used in QC are limited to be nearest neighbour models so there are no artificial forces in the atomistic region. In a refinement of the QC method, Shenoy *et al.* [95] removed these ghost forces in what they called the dead load approximation. The QC method has been applied to many systems, for example to study the interaction of dislocations with grain boundaries [96] and the effect of grain orientation on fracture [97].

3.3 Coupling Two Classical Atomistic Systems

As a prelude to looking at the difficulties posed when attempting to couple quantum and classical systems, let's consider how two classical atomistic models could be combined. Providing the models are both short ranged, a straightforward treatment of the boundary is possible. We allow the regions to overlap as shown in Fig. 3.3, then evaluate the energy for the two regions separately, with a buffer region for

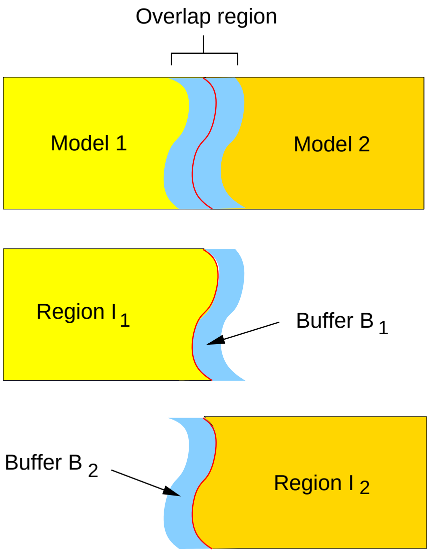


Figure 3.3: Schematic showing how overlap buffers can be used to solve the boundary problem in a classical/classical embedding scheme, where all interactions are short ranged.

both calculations. The locality of the classical potentials means that each of these energies is a sum of local energies ϵ_i for each atom, so it is easy to separate the energy into a contribution due to the interior atoms and one due to the buffer atoms. For example, the energy for Model 1 of Fig. 3.3 can be decomposed as

$$E^{(1)} = \sum_i \epsilon_i = \underbrace{\sum_{i \in I_1} \epsilon_i}_{E_I^{(1)}} + \underbrace{\sum_{i \in B_1} \epsilon_i}_{E_B^{(1)}} \tag{3.1}$$

where I_1 and B_1 denote the interior and buffer sections of region one, as shown in Fig. 3.3. The same decomposition can be applied to give $E^{(2)}$ for Model 2. The total hybrid energy is then obtained by summing the contributions from the two interior regions, neglecting the buffers:

$$E_{\text{hybrid}} = E_I^{(1)} + E_I^{(2)} \tag{3.2}$$

The artificial surfaces created at the boundary will be much more of a problem when we come to consider embedding a non-local quantum system which is described in the next section.

3.4 Coupling Quantum and Classical Systems

Coupling quantum and classical systems poses significantly greater challenges than combining two classical descriptions of matter. As quantum mechanics is non-local the simple partitioning scheme described above will not work. To overcome this problem we need to provide appropriate boundary conditions for the quantum calculations and find a way to spatially localise their effects.

The quantum mechanical model is assumed to be accurate enough to describe the physics of the region of interest correctly, perhaps using tight binding or an *ab initio* approach. The classical model needs only to correctly capture the basic topology of bonding and give the correct response to small elastic deformations, while remaining inexpensive to compute: empirical interatomic potentials are ideal for this purpose. Furthermore, since we shall use the quantum model anywhere we suspect the classical model will be unreliable, we prefer that the classical model be robust and inexpensive rather than being highly transferable. There has been a great deal of effort in recent years to produce potentials which attempt to model complex processes such as defect formation — generally we have found that such potentials are not useful in a hybrid simulation. We prefer simple potentials such as the Stillinger-Weber model to more complex ones such as EDIP.

The widely used assumption, upon which all quantum/classical hybrid schemes rely, is that the physics is local so that observables can be computed locally, taking into account only atoms which lie within some finite distance of the region of interest. Equivalently, we require that distant trajectories are instantaneously independent. Providing the quantum region is large enough, the trajectories that are important are not affected by the fact that, far away, the system is treated classically. However, it is also necessary, from a practical point of view, that the quantum trajectories can be computed accurately using a small quantum region. Both these conditions are satisfied by the *strong locality* condition:

$$\frac{\partial^n}{\partial \mathbf{r}_j^n}\frac{\partial E_{\text{total}}}{\partial \mathbf{r}_i} \to 0 \ \text{ as } \ |\mathbf{r}_i - \mathbf{r}_j| \to \infty \ \ \forall\, n \in \mathbb{N}, i \neq j \tag{3.3}$$

where $\mathbf{r}_i$ and $\mathbf{r}_j$ are the positions of atoms i and j. This spatial localisation of observables is a stronger requirement than that the density matrix be sparse so that its elements decay rapidly as the separation between two atoms increases. The strong locality assumption can be tested for a particular system by testing the rate of convergence of the force on the central atom of a cluster as the cluster radius is

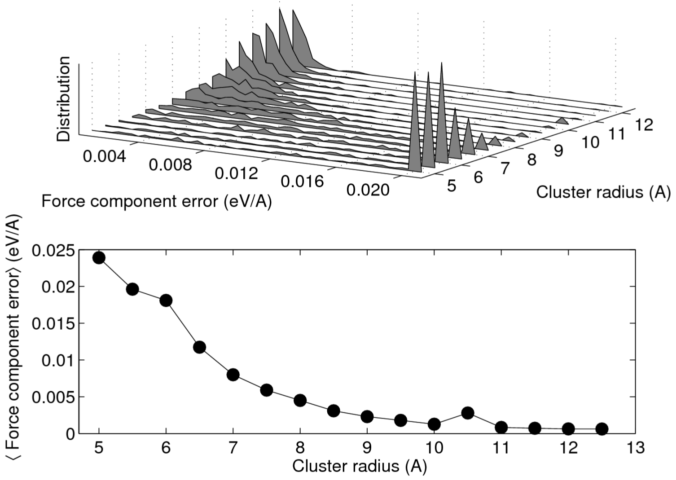


Figure 3.4: Top: distribution of force component errors at the centre of a finite cluster for a range of cluster sizes. The peaks on the right represent the integrated distribution of errors which are larger than 0.02 eV Å^{-1}. The system is a silicon self-interstitial in a 512-atom cubic cell with periodic boundary conditions, equilibrated at 1400 K. Bottom: the mean of the absolute error in the force components as a function of cluster radius. The peak at 10.5 Å is due to the periodicity of the underlying system. Clusters were terminated with hydrogen atoms. The force model is tight binding, from Bowler *et al.* [88]. This figure is reproduced from Csányi *et al.* [40].

increased. Fig. 3.4 shows an example of a test of strong locality for silicon using a tight binding force model. Most quantum systems either obey strong locality, or at least the parts of the Hamiltonian that do not, such as long range Coulomb and van der Waals interactions, can be dealt with in a purely classically manner.

Before we consider the details of the coupling strategy, it is appropriate to ask what we want from an ideal hybrid simulation. It is not feasible for the atoms in the quantum region to move as if the whole system were described quantum mechanically, since the classical atoms still move along classical trajectories, and the quantum atoms will respond to the new positions of the classical atoms. Hence, the best we can aim for is for the quantum atoms to behave *instantaneously* as if they are embedded in a fully quantum system.

3.5 The QM/MM Approach

The earliest quantum/classical hybrid simulation was performed by Warshel and Levitt [28] in 1976 in which they model the reaction of the enzyme lysozyme. Enzyme catalysis is often controlled by large scale motion of macromolecules, with a small active site at which the chemical reaction takes place. Warshel and Levitt noted the need to describe the active site at a quantum mechanical level of detail to give an accurate description of hydrophobic interactions and hydrogen bonding during catalysis. The electrostatic environment at the active site is determined by the configuration of the entire system of enzyme, substrate and solvent. The long range nature of electrostatic forces in such systems means that atoms far from the active site respond to the presence of a substrate, which in turn causes a change in the local electrostatic environment.

Hybrid methods of this kind, where the dominant interaction is electrostatic, have become known as quantum mechanical/molecular mechanical (abbreviated QM/MM) methods. They have become very popular in the biological and biochemical modelling communities in recent decades. All the fundamental aspects of modern QM/MM techniques were contained in Warshel's pioneering work: the quantum region was chosen very carefully by hand, and the boundary atoms were terminated with a frozen hybrid orbital.

In this section I will give only a brief overview of the QM/MM method; for reviews of recent developments in the QM/MM field see Lin and Truhlar [98] and Ruiz-Lopez [99]. Within the QM/MM framework, the total energy is the sum of three contributions: the quantum mechanical energy of the quantum region, the classical energy of the rest of the system, and a term representing the interaction between the two. There are two distinct approaches to performing a QM/MM calculation, which differ in their treatment of the interaction between regions: *mechanical embedding* and *electrostatic embedding*. I describe each of these below.

Mechanical Embedding

Mechanical embedding schemes perform quantum calculations for the QM region in the absence of the MM region, treating the interactions between the regions classically. The simplest mechanical embedding scheme is the two-layer ONIOM method [29], illustrated in Fig. 3.5. Here the total energy is obtained from three

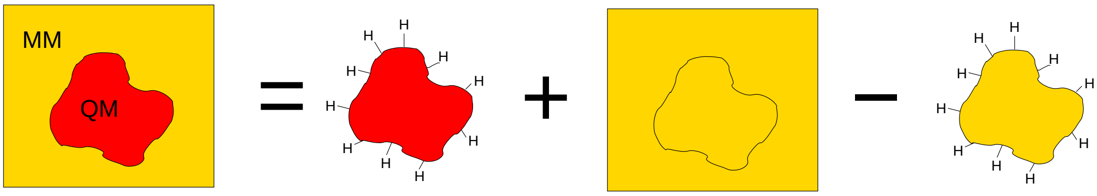


Figure 3.5: The two-layer ONIOM-style QM/MM scheme. Regions coloured red are treated with QM and those coloured yellow with MM. The termination atoms indicated as H could in fact be a more complex pseudoatom termination.

independent calculations:

$$E_{QM/MM} = E_{QM}(QM) + E_{MM}(QM + MM) - E_{MM}(QM) \tag{3.4}$$

where the subscripts denote the energy model and the function arguments indicate the parts of the system to be included in each calculation. The MM system contains all the atoms and the quantum system contains the atoms of quantum mechanical interest plus *link atoms* used to cap dangling covalent bonds. ONIOM relies on cancellation of errors between the two surface energies.

There are two major drawbacks to the mechanical embedding approach. Firstly it is not always possible to obtain an accurate set of electrostatic MM parameters for atoms in the QM region; this is a particular problem since it is often the unavailability of such parameters which motivates the desire to treat this region quantum mechanically in the first place. Secondly, the scheme ignores perturbations in the electronic structure of the QM region caused by the charge distribution of the MM region. The three layer ONIOM method [100, 101] goes some way to solving these problems by introducing an intermediate layer, treated at semi-empirical quantum mechanical accuracy. This allows a consistent treatment of the polarisation of the active site.

Electrostatic Embedding

In an electrostatic embedding scheme the QM calculation is carried out in the presence of the classical charge distribution by adding terms that describe the electrostatic interaction between regions to the QM Hamiltonian. Normally atom centred partial point charges are used, but more advanced techniques employ a multipole expansion of the electric field for increased accuracy. Bonded and van der Waals interactions between the regions are still treated classically.

One problem of the standard electrostatic embedding approach is that classical atoms just outside the quantum region appear as bare coulomb charges in the quantum calculation. There is a tendency for electron density to unphysically ‘spill-out’ onto these atoms to neutralise these charges. Laio *et al.* [102] have developed an efficient implementation of an electrostatic embedding scheme that addresses this issue by dealing with the short and long range electrostatic interactions differently to avoid spill-out.

Termination of Covalent Bonds

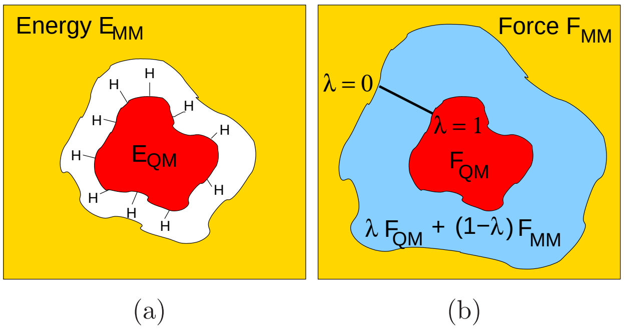
The QM/MM method has been applied fairly extensively to multiscale solid state systems of the types described at the beginning of this chapter. Electrostatic screening is very effective in metals and small band gap insulators, so mechanical embedding schemes are widely used for such systems. Bonded interactions between the QM and MM regions are much more of a problem in the solid state, since for a typical spherical QM region the number of covalent bonds that have to be cut to generate the QM cluster is of the same order as the number of atoms in the region.

To incorporate a quantum mechanical calculation of a subsystem into the total Hamiltonian, these artificially cut bonds must be terminated. There are various methods for doing this, usually based on using hydrogen link atoms or parameterised semiempirical ‘pseudoatoms’ that attempt to mimic the electronic effect of the region outside the subsystem that has been removed. A localised orbital parameterised with calculations on small model systems can be used to provide a quantum mechanical description of the charge distribution around the QM/MM boundary [103] [2].

We have seen in Section 3.2 that these termination strategies are sufficient to give accurate classical forces, since the classical description of covalent bonding is very near sighted. Quantum mechanics, however, is not a nearest neighbour model, so atoms close to the terminated boundary of the QM subsystem feel an artificial environment, no matter how complex the passivation scheme employed. Moreover, it is impossible to exclude the contribution of the termination atoms to the total quantum mechanical energy of the subsystem. In a typical covalent system the length scale for strong locality of the electronic energy is the order of a nanometre, so any termination method that merely replaces a broken bond with a single atom

---

[2] This approach is less widely used since is not possible to include these hybrid orbitals in a plane wave *ab initio* code.



(a) (b)

Figure 3.6: Comparison of (a) QM/MM and (b) force mixing approaches to the electronic termination problem.

cannot hope to give accurate forces at the boundary, due to the non-local nature of the quantum mechanical forces.

Force Mixing

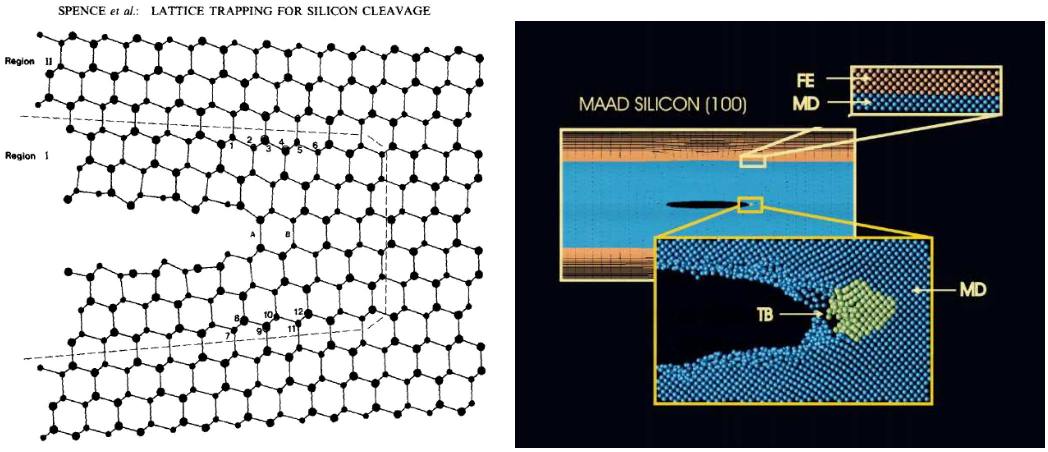
An alternative to the standard QM/MM termination method is to move smoothly from quantum to classical forces over a transition region. This is the *force mixing* technique, where the forces used for the dynamics are interpolated, commonly linearly according to

$$\mathbf{F} = \lambda \mathbf{F}_{\mathrm{QM}} + (1 - \lambda) \mathbf{F}_{\mathrm{classical}} \tag{3.5}$$

with λ varying from zero on the classical edge of the transition zone to one at the QM edge. Higher order interpolation is also possible. A comparison of traditional QM/MM termination and force mixing is illustrated in Fig. 3.6.

Compared to the link atom method, force mixing slightly reduces the effect of inaccurate forces on atoms near to the edge of the QM region, since they are reduced in weight and mixed with classical forces. However, since the strong locality length scale is large, the transition zone must be very wide for this to have much of an effect, so large quantum mechanical zones are required.

A major disadvantage of force mixing is that since the forces no longer come from a single Hamiltonian neither energy nor momentum are conserved. The resulting dynamics can be unphysical. The action-reaction principle is not obeyed so, for example, the forces on a dimer spanning the boundary do not sum to zero. This creates a mechanical incompatibility across the boundary, which can lead to



(a) (b)

Figure 3.7: Early hybrid fracture simulation techniques. (a) Relaxed 324 atom structure obtained using flexible boundary conditions. Region I (120 atoms) was treated with an *ab initio* method and Region II with an interatomic potential. Reproduced from Spence *et al.* [37]. (b) Snapshot from a MAAD simulation of fracture, showing the decomposition of the simulation into finite element (FE), molecular dynamics (MD) and tight binding (TB) regions. Reproduced from Broughton *et al.* [38].

instabilities in the dynamics. Nevertheless, force mixing continues to be the most widely used approach for hybrid simulation of solid state systems.

The earliest quantum/classical multiscale fracture simulations were published in 1993 by Spence *et al.* [37]. They describe their approach as a flexible boundary condition for an *ab initio* calculation, but it is effectively a force mixing embedding scheme. Alternate relaxations of the two regions illustrated in Fig. 3.7a were performed, with an overlap buffer to ensure self consistency. Some years later, Broughton *et al.* [38] proposed the MAAD (macroatomistic *ab initio* dynamics) method which couples finite elements, molecular dynamics and semiempirical tight binding in a QM/MM approach to model crack propagation; a snapshot of the dynamics is shown in Fig. 3.7b. Pseudoatom terminator 'silogens' designed to behave like monovalent silicon atoms were used to terminate the tight binding region and a force mixing embedding approach was used.

Ogata's group has applied the ONIOM method to the simulation of cracks [104], surface oxidation [105], and more recently they have investigated the effect of water on the initiation of corrosion induced cracks [106]. The group uses an improved version of ONIOM called the buffered-cluster method [107]. The QM region is cut out as normal, but then buffer atoms are added to terminate broken covalent bonds. The buffer is then relaxed using the classical force model, resulting in a relaxed buffered cluster which gives better surface error cancellation since it is closer to the equilibrium bulk structure.

Multiple Layer Termination

In 2001, Bernstein [39] proposed a modified treatment of the quantum zone boundary that addresses the electronic termination problem. He used a Green's function technique to create a *transition zone* with a thickness of several atomic layers which is included in the quantum mechanical calculation. Forces from this zone are not included in the dynamics. This method was later employed in a hybrid classical and tight binding simulation [108], referred to as the DCET (dynamic coupling of empirical potential and tight binding) method. The combination of transition zone and force mixing gives accurate quantum mechanical forces and allowed the QM region to be moved during a simulation for the first time. However, the force mixing technique requires a large QM region, making the method difficult to scale up to a full *ab initio* calculation.

There is an alternative, more straightforward, termination strategy: we can obtain accurate *forces* for all atoms in the quantum region by using a wider buffer region. If we include a thick enough shell of nominally classical atoms in the quantum calculation, then the forces on the QM atoms themselves will be accurate. Since these forces are local quantities, we can easily discard the contaminated termination region and keep only the forces on the original QM atoms. This finite buffer scheme, illustrated in Fig. 3.8, is a major ingredient of the 'Learn on The Fly' hybrid method which is discussed in detail in Chapter 4.

These multiple layer termination approaches can solve the electronic termination problem, but there will still be a mechanical incompatibility across the boundary. If we used forces from the finite buffer scheme to do molecular dynamics, the resulting trajectories could be unstable, exactly as in the force mixing approach discussed above.

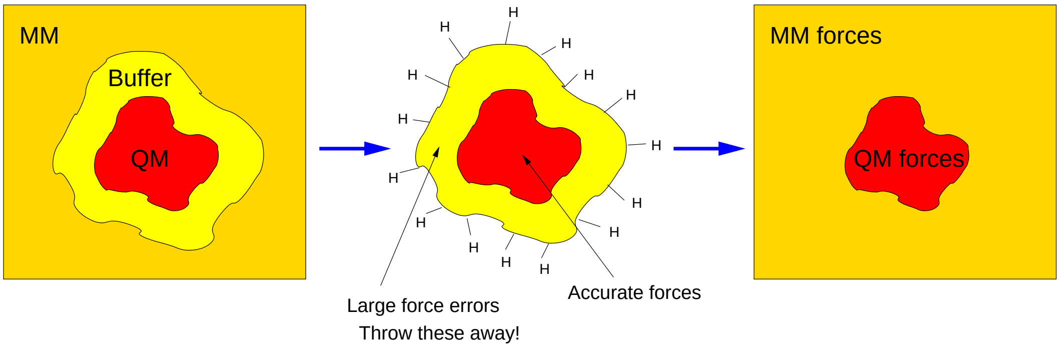


Figure 3.8: The finite buffer termination strategy. Forces on the atoms in the buffer region (yellow) are discarded to give accurate hybrid forces on all atoms (right hand side).

3.6 Summary

This chapter has introduced hybrid modelling techniques and reviewed a number of approaches which allow simultaneous simulation of coupled quantum and classical systems. We have seen that the fundamental difficulty of constructing such a hybrid modelling scheme lies in finding an effective treatment of the boundary. This is a particular problem in solid state systems, where many covalent bonds have to be cut to form the QM cluster and this has restricted the application of the QM/MM method. The problem can be divided into the electronic termination problem, which can be solved by discarding the inaccurate forces in a buffer zone at the edge of the QM region, and the mechanical matching problem. A solution to this mechanical mismatch forms the basis of the 'Learn on The Fly' method, discussed in Chapter 4.

Chapter 4

The ‘Learn on the Fly’ Method

4.1 Introduction

This chapter begins with an overview of the ‘Learn on the Fly’ (LOTF) method after which various aspects of the scheme will be discussed in more detail. In particular, I will concentrate on developments that have been made during the course of my PhD. I shall primarily consider the method in the context of the application that forms the remainder of this thesis: the brittle fracture of silicon.

4.2 Overview of the LOTF Method

As discussed in detail in Chapter 3, the simulation of hybrid quantum and classical systems has two main difficulties, both related to the boundary between the quantum and classical regions. Firstly the non-locality of quantum mechanics makes it difficult to evaluate accurate forces for atoms at the edge of the QM zone. This is the electronic termination problem, and we have seen that it is possible to solve this using a finite buffer provided we are willing to forgo conservation of the total energy. Secondly, if forces from two different models are combined then a mechanical incompatibility arises at the interface as illustrated in Fig. 4.1.

The ‘Learn on the Fly’ (LOTF) method, proposed by De Vita and Car [41] in 1998, is essentially a way to deal with this mismatch to allow the QM/MM framework to be extended to solid state systems where many covalent bonds are cut by the insertion of a boundary between the quantum and classical subsystems [40, 42]. In contrast to QM/MM approaches such as ONIOM, there is no attempt to build a combined Hamiltonian from the separate classical and quantum Hamiltonians:

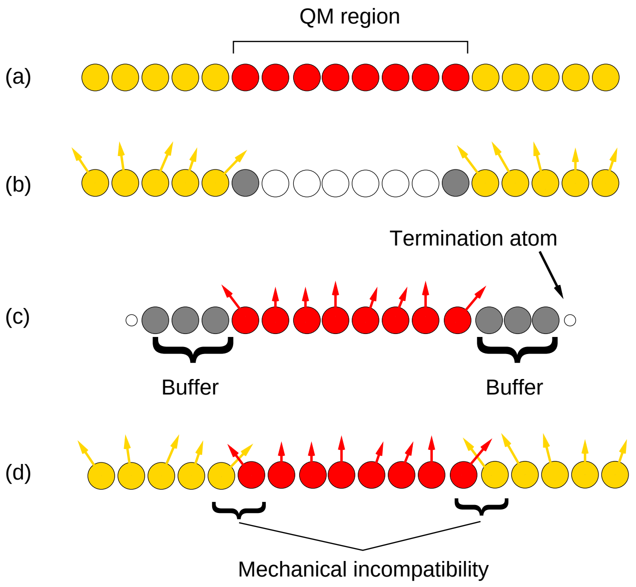


Figure 4.1: Schematic illustration of the mechanical matching problem at the boundary between classical and quantum regions (a) Definition of QM and classical regions (b) Evaluation of classical forces. The grey atoms are required to compute the forces on atoms at the edge of the classical region (c) QM force computation using a finite buffer. Forces on the grey atoms are thrown away to give accurate forces on the red atoms (d) Combination of classical and quantum forces to give the hybrid forces. Note the mechanical incompatibility at the boundary.

rather we focus on the forces on each atom since these are purely local quantities. Essentially, we introduce a simple adjustable model — this is a set of springs between pairs of atoms, with different spring constants for each spring — in addition to the classical model. We then tune the spring constants to reproduce the correct forces on all the atoms, *i.e.* we seek to minimise the functional

$$\mathcal{F} = \sum_{\text{all atoms}} \left|(\mathbf{F}_{\text{QM}} - \mathbf{F}_{\text{classical}}) - \mathbf{F}_{\text{adj}}\right|^2 \tag{4.1}$$

where $\mathbf{F}_{\text{QM}}$ is the quantum mechanical force on an atom, $\mathbf{F}_{\text{classical}}$ the classical force and $\mathbf{F}_{\text{adj}}$ the force from the simple adjustable model. If a good fit can be obtained, the result of the optimisation is that

$$\mathbf{F}_{\{\text{QM,classical}\}} \simeq \mathbf{F}_{\text{classical}} + \mathbf{F}_{\text{adj}} \tag{4.2}$$

$\mathbf{F}_{\{\mathrm{QM,classical}\}}$ is defined by

$$\mathbf{F}^{(i)}_{\{\mathrm{QM,classical}\}} = \begin{cases} \mathbf{F}^{(i)}_{\mathrm{QM}} & \text{Atom } i \in \text{QM region,} \\ \mathbf{F}^{(i)}_{\mathrm{classical}} & \text{Atom } i \in \text{Classical region.} \end{cases} \tag{4.3}$$

We then use the sum of classical and adjustable forces to propagate the molecular dynamics, so the total potential and force are given by

$$V_{\mathrm{LOTF}}(\mathbf{R}, \boldsymbol{\alpha}) = V_{\mathrm{classical}}(\mathbf{R}) + V_{\mathrm{adj}}(\mathbf{R}, \boldsymbol{\alpha}), \tag{4.4}$$

$$\mathbf{F}_{\mathrm{LOTF}} = -\nabla V_{\mathrm{LOTF}} = \mathbf{F}_{\mathrm{classical}} + \mathbf{F}_{\mathrm{adj}}. \tag{4.5}$$

$V_{\mathrm{classical}}$ is the classical interatomic potential and V_{adj} the adjustable potential. $\mathbf{R}$ represents the coordinates of all atoms and $\boldsymbol{\alpha}$ the parameters of the adjustable potential. Since the forces are the derivative of a single total energy function, the action-reaction principle is restored and there are no mechanical inconsistencies at the boundary. The mechanical mismatch has been traded for small force errors, smeared out over the correcting region. Thus we have found a single Hamiltonian which instantaneously yields forces that are closest to the desired forces.

After moving the atoms according to these forces, we recompute the quantum forces and retune the adjustable potential parameters to reproduce these new forces. Combining this fitting strategy with the use of a buffer region around the QM zone allows us to throw away the unphysical surface region of each QM calculation. The LOTF scheme is illustrated in Fig. 4.2.

In practice, a number of simplifications to the procedure are possible. The target quantum mechanical forces do not have to be recomputed at every step, since the adjustable potential parameters can be used unchanged for a few time steps without causing a significant deviation from the true hybrid trajectory. Secondly, far away from the quantum region, the classical potential describes the dynamics of the system adequately, and so we can restrict the range of application of the corrective potential to a *fit region* centred on the quantum region and extending typically over a few nanometres away from this region.

The following sequence of steps make up the LOTF scheme. I describe various parts of the algorithm in more detail in the following sections.

1. **Initialisation.** Define QM zone, then grow this by bond hopping (described in Section 4.6) to form the fit region. Create springs between pairs of atoms in fit zone. The adjustable potential is then initialised by performing classical

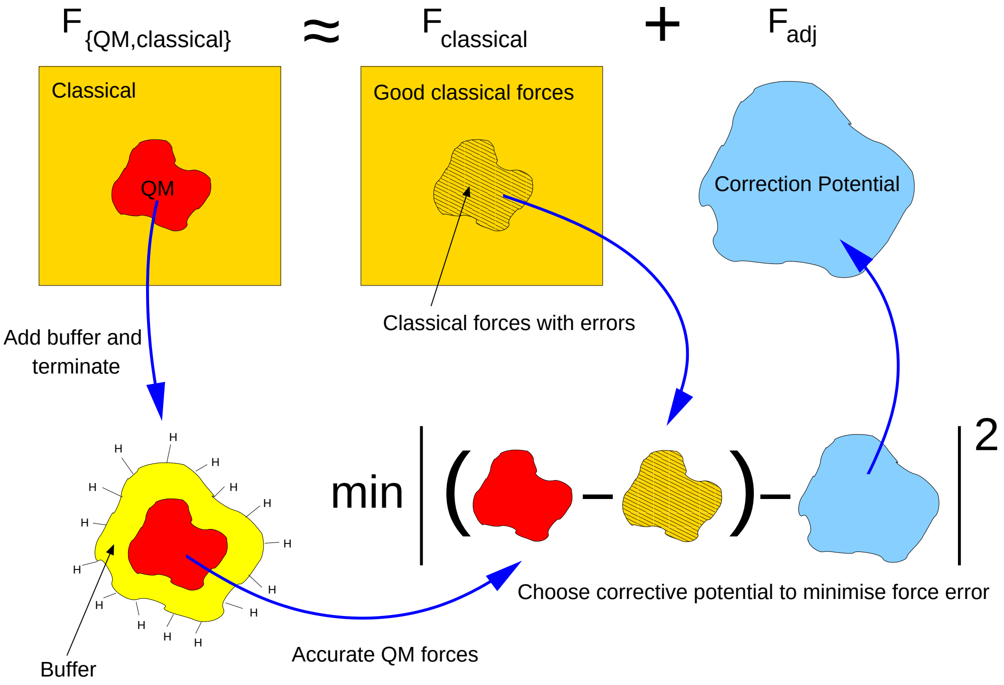


Figure 4.2: Schematic overview of the LOTF scheme, showing QM force computation using a buffer and optimisation of the corrective potential to minimise the force difference between classical and quantum forces. The final forces used for MD are the sum of the classical and corrective contributions.

and QM calculations and optimising its parameters to reproduce the initial force differences.

2. **Extrapolation.** Use the sum of classical and adjustable potentials with fixed parameters to propagate the dynamics forwards for N_{interp} steps using the velocity Verlet algorithm.

3. **Selection.** Identify atoms to be described to QM accuracy. This cannot be done in a fully general way because it depends on details of the particular application, although specific examples will be described in Section 4.6. The fit region is determined from the QM region by bond hopping, also described in Section 4.6.

4. **Force Computation.** Compute accurate forces on all flagged atoms, either with individual atom centred clusters or in one large calculation using the methods described in Section 4.7.

5. **Optimisation.** Minimise Eq. 4.1 by varying the adjustable potential parameters. The optimisation is performed using either conjugate gradients or SVD, depending on the form of adjustable potential used. The adjustable potential is described in Section 4.8.

6. **Interpolation.** Return the state of the system to that before the extrapolation and rerun the dynamics, interpolating the adjustable potential parameters between the old and new values, to achieve a smooth evaluation of the parameters in time. Interpolation is described in detail in Section 4.5.

7. **Return to 2.**

4.3 Validation and Applications

The LOTF method has been validated on a number of solid state systems which are small enough to allow direct comparison with fully quantum calculations. I will mention some of these briefly here; see the references given for full details.

The LOTF scheme has been used to measure the diffusion constant of vacancies [42] and hydrogen atoms [43] in crystalline silicon. In both cases, the method correctly reproduces the diffusion constant measured in a fully quantum calculation. LOTF has also been used to measure the stress *vs.* strain behaviour of a silicon nanobar subjected to uniaxial tension [40]. In this case the LOTF simulation gives the same critical loading as the fully quantum mechanical model, in contrast to a classical simulation which overestimates this loading by around 50%.

In an example of an application which is too large for a fully quantum calculation, LOTF has been used to study the glide of partial dislocations, again in silicon [42]. Two partial dislocations are observed to move together and eventually annhilate. Moving from semi-empirical tight binding to DFTB allows the interaction of the dislocations with dopants to be modelled [44].

4.4 Development of the LOTF95 code

The LOTF scheme was originally proposed by Alessandro de Vita in 1998 [41]. It was first implemented in Fortran 77, and hereafter this version is referred to as LOTF77. This version of the code was fairly basic, partly due to the limitations imposed by the choice of programming language. Consequently, Gábor Csányi developed a new implementation, this time using the C++ programming language. This version is referred to as C-LOTF. It was this version of the code that was used for most of the validation work published in Csányi *et al.* [42] and Csányi *et al.* [40].

During the course of my PhD, the LOTF method has once again been completely reimplemented, this time in Fortran 95. This implementation uses the new features of modern Fortran to combine excellent numerical efficiency with aspects of Object Oriented (OO) programming inspired by languages such as C++. The resulting code is referred to as LOTF95, and is based on the *expressive programming* philosophy, in which the structure and language of the top level code is as close as possible to the abstract algorithm we are trying to implement. In the context of a molecular dynamics simulation, achieving this goal while maintaining reasonable code speed means having objects of type `Atoms` and variables like `Atoms%positions` and `Atoms%velocities` so that the code which implements a time step integrator can resemble the corresponding mathematical formula.

During the implementation of the LOTF95 code, it was decided to split the project into a general purpose, reusable, molecular dynamics library and a part specific to the LOTF method. The former has developed into the LIBATOMS project [109]. LIBATOMS is designed to scale well up to hundreds of thousands of atoms. More details of the expressive programming approach and the implementation of LIBATOMS have been published in Csányi *et al.* [110].

I have been heavily involved in all stages of the development of this new code, from design and specification through to implementation, testing, profiling, debugging and writing of documentation [109]. Almost all of the results that appear in this thesis have been produced using LOTF95, and in the discussion below this is the version to which I refer, unless otherwise stated.

4.5 Parameter Interpolation

Within the LOTF framework, the evolution of the atomic trajectories can be written as a function of the current atomic positions and the adjustable potential parameters

$$\dot{\mathbf{R}} = f(\mathbf{R}, \boldsymbol{\alpha}) \tag{4.6}$$

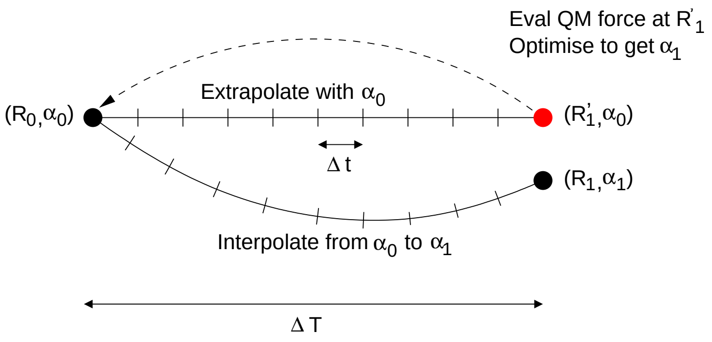


Figure 4.3: Predictor-corrector parameter evolution as used in the LOTF scheme.

The parameters $\boldsymbol{\alpha} = \{\alpha\}$ can be computed at any point in the phase space $\mathbf{R}$ of the hybrid system by evaluating the quantum forces and optimising the adjustable potential. This is an expensive process, but formally $\boldsymbol{\alpha}$ is simply a function of the current atomic positions, so can be written as $\boldsymbol{\alpha}(\mathbf{R})$. The trick then lies in choosing the functional form of the adjustable potential to make $\boldsymbol{\alpha}(\mathbf{R})$ a slowly varying function so that we don't have to evaluate it at every time step. For this to be true, $\boldsymbol{\alpha}$ should change slowly with respect to the atomic positions. This will be the case providing

$$\left|\frac{\dot{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}}\right| \ll \left|\frac{\dot{\mathbf{R}}}{\mathbf{R}}\right|. \tag{4.7}$$

If this equation holds, then we can use the same values of the adjustable potential parameters for a number of time steps, the number depending on how much more slowly the components of $\boldsymbol{\alpha}$ evolve compared to the evolution of the atomic positions.

The procedures for determining the trajectory and updating $\boldsymbol{\alpha}$ are formalised using a predictor-corrector scheme, as shown in Fig. 4.3. We start at a point $\mathbf{R}_0$ in phase space with parameters $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}(\mathbf{R}_0)$. We shall denote this point $(\mathbf{R}_0, \boldsymbol{\alpha}_0)$. We then take N_{interp} steps of size Δt with fixed parameters to arrive at the point $(\mathbf{R}'_1, \boldsymbol{\alpha}_0)$. Here we compute new quantum forces and obtain a new set of parameters $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}(\mathbf{R}'_1)$. This completes the predictor part of the cycle. Rather than continuing the simulation from this point with the new Hamiltonian parameters, which would give an abrupt discontinuity in the second derivative of the atomic trajectories, we return to the point $(\mathbf{R}_0, \boldsymbol{\alpha}_0)$ and redo the dynamics, interpolating the parameters between $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$. At the end of the interpolation we arrive at a slightly different point in phase space, $(\mathbf{R}_1, \boldsymbol{\alpha}_1)$. This is the end of the corrector stage.

The cycle could be repeated by computing new quantum mechanical forces at $\mathbf{R}_1$, getting new parameters and going back to $(\mathbf{R}_0, \boldsymbol{\alpha}_0)$ to re-interpolate. In practice, for sensible choices of N_{interp} and Δt, $\mathbf{R}_1'$ and $\mathbf{R}_1$ are very close and the region of validity of the parameters is large enough that a single cycle is sufficient. The choice of the optimal interpolation scheme is a matter for further investigation. Initially, a cubic function with the derivatives fixed to be zero at the beginning and end of the interpolation was used:

$$\boldsymbol{\alpha}(\lambda) = \boldsymbol{\alpha}_0 + 6\left(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}_1\right)\left(\frac{\lambda^3}{3} - \frac{\lambda^2}{2}\right) \tag{4.8}$$

where λ varies from zero at the beginning of the interpolation to one at the end.

However, this choice results in unnecessary acceleration and deceleration of the parameters. Recently, we have in preference adopted a linear interpolation scheme, where the parameters evolve according to

$$\boldsymbol{\alpha}(\lambda) = (1-\lambda)\,\boldsymbol{\alpha}_0 + \lambda\boldsymbol{\alpha}_1 \tag{4.9}$$

This is actually the optimal choice if we have no information about the time derivatives of the parameters. In practice, as will be shown in Section 4.8 the choice of interpolation order makes little difference. The parameter derivatives could be evaluated by taking a small step in parameter space and using the method of finite differences. The increase in the computational cost of interpolation that this would cause would have only a minimal impact in the time for one complete step, since LOTF is generally used in a regime where the overall cost is dominated by the QM force evaluation.

4.6 QM Selection

In common with all hybrid schemes, the largest approximation in the LOTF approach is the classical treatment of the majority of the atoms in the simulation. This is a controllable approximation, since we are free to choose exactly which atoms to include in the higher accuracy calculation.

In traditional QM/MM, it is not possible to move the QM region during a calculation, or for atoms to move in or out of the QM region. Adding a new atom to a cluster causes a discontinuous change in the electronic structure, so that the resulting forces and thus trajectories of atoms near the edge of the QM region also change discontinuously. The buffer region allows this fundamental limitation to be overcome. If we move the quantum region in space together with its buffer then the Hamiltonian in the QM region changes smoothly. The use of a buffer region rather than force mixing also means that the QM region can be kept small, greatly increasing the speed of force evaluation. This is particular important in view of the typical $O(N^3)$ scaling of the computational cost of tight binding and *ab initio* calculations with the total number of atoms in the QM calculation.

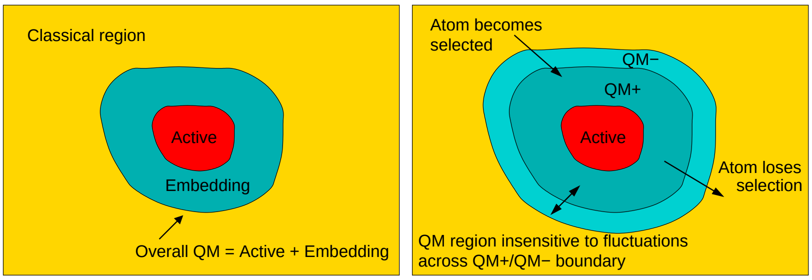
Using a small and mobile QM region allows us to continually select which atoms to treat more accurately during the simulation. We therefore require robust algorithms to determine where the QM region should be. It is impossible to design a completely general algorithm to do this, since the selection criteria clearly depend on the nature of the scientific question we are asking. In general, it is not possible to correct classical trajectories everywhere that they go wrong: even at modest temperatures, some atoms are displaced sufficiently far from equilibrium that classical models predict forces that disagree with quantum mechanical forces by more than 1 eV/Å.

Rapid oscillations of the set of atoms which make up the QM region cause a number of problems. Atoms that move in and out of the QM region can generate heat as they do so, since the two energy models are each trying to drive the system towards a different configuration. From a practical point of view, such changes in the number of QM atoms are inefficient since *ab initio* calculations can be sped up immensely by reusing the electronic density and wavefunctions from the previous timestep (see Section 4.7 below). Thus unnecessary fluctuations in the QM region hugely decrease the computational efficiency of the method.

Active and Embedding Regions

The selection of the QM region therefore needs to be based on geometrical and topological criteria. It is useful to calculate time-averaged atomic positions, to filter out the effect of fast optical phonons. These are defined by:

$$\langle \mathbf{r} \rangle (t) = \frac{1}{1/2 + \tau/\Delta t} \sum_{n=0}^{\infty} \mathrm{e}^{-n\Delta t/\tau} \mathbf{r}(t - n\Delta t) \tag{4.10}$$



(a) (b)

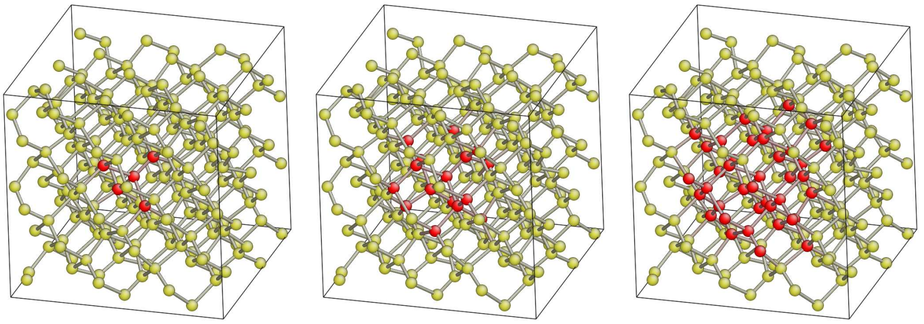
Figure 4.4: QM selection algorithms. (a) active (red) and embedding (cyan) regions (b) the hysteretic selection algorithm, showing the active region and the pair of embedding regions, labelled as QM+ and QM-.

where Δt is the MD time step and τ the averaging time constant, typically equal to a few times the period of the highest frequency phonon mode. These coordinates are very useful for determining topological properties, such as the nearest-neighbour count of each atom, which can be used to locate point defects and breaking bonds. These atoms can be flagged as *active*, and surrounded by a shell of *embedding* atoms to make up the entire QM region, as shown in Fig. 4.4a. The thickness of the embedding shell should be of the order of the strong locality length scale of the system under consideration. This approach to determining the QM region was used in the LOTF77 and C-LOTF codes but in LOTF95 this has been replaced by bond hopping, as described in the following section.

Bond Hopping

The early LOTF implementations constructed the embedding region as the union of a set of spheres of fixed radius centred on each of the active atoms. This can lead to undesirable fluctuations in the QM region for atoms near to the edge of one of these spheres.

In LOTF95, this has been replaced by a *bond hopping* approach. Each hop is a breadth-first step on the graph of atomic connectivity, as illustrated in Fig. 4.5. Once the connectivity of the system has been computed in the manner described in Section 2.3, this approach is straight-forward to implement. We have to define what we mean when we say two atoms are nearest neighbours. This is done using a tolerance factor f_{nn}: atoms i and j are considered to be nearest neighbours if they



(a) (b) (c)

Figure 4.5: Illustration of bond hopping in bulk silicon. The central atom is used to form the QM selection region (red) by bond hopping. The panels show the resulting QM region after (a) 1 hop (b) 2 hops (c) 3 hops.

satisfy

$$r_{ij} < f_{\mathrm{nn}} \left[r_{\mathrm{cov}} \left(Z_i \right) + r_{\mathrm{cov}} \left(Z_j \right) \right] \tag{4.11}$$

where Z_i and Z_j are the atomic numbers and $r_{\mathrm{cov}} \left(Z_i \right)$ and $r_{\mathrm{cov}} \left(Z_j \right)$ are the associated covalent radii. Typically a tolerance of around 1.3 gives good results. Bond hopping is only useful when combined with the time-averaged atomic coordinates of Eq. 4.10: otherwise pairs of atoms separated by close to the threshold distance would appear to bond and debond as they oscillate in and out of the range of nearest neighbour bonding according to Eq. 4.11.

Hysteretic Selection Algorithm

In many circumstances it is not possible to completely avoid undesirable fluctuations in the QM region using the techniques described above. Difficulties occur in situations where it is impossible to choose the averaging time τ to be large enough to exclude atomic oscillations, but not so large as to miss rapid bond breaking events. This is particularly true for the case of a fast-moving crack: it has been found that τ needs to be below about 100 fs to correctly follow a moving crack in silicon, but this is of the same order as the thermal oscillation period of a single Si—Si bond. Therefore atoms at the edge of the QM region tend to pop in and out as they vibrate backwards and forwards.

To solve this problem, I have developed a selection algorithm with hysteresis. We flag atoms as active using topological criteria based on the time-averaged coordinates as before, for example this could be the atoms which have changed coordination since the start of the simulation. The embedding region is replaced with a pair of shells, as depicted in Fig. 4.4b. Atoms must come within the inner shell to be added to the QM region, but they will continue to be treated quantum mechanically while they are inside the outer shell. The composition of the two shells QM_- and QM_+ can be written

$$QM_- = \bigcup_{i \in A} S(i, R_-) \tag{4.12}$$

$$QM_+ = \bigcup_{i \in A} S(i, R_+) \tag{4.13}$$

where A is the set of active atoms and $S(i, R)$ is the set of atoms contained within a sphere of radius R centred on atom i. R_- and R_+ are the radii of the two embedding shells, with $R_- < R_+$. The width of the hysteresis zone $R_+ - R_-$ can be tuned to remove rapid fluctuations of atoms between the quantum and classical regions. Further stability can be obtained by constructing the embedding shells from the active zone by bond hopping rather than placing them at fixed distances from the active atoms. In systems where there are several areas of interest the active region can be formed simply by taking the union of the individual active regions.

4.7 Force Evaluation

Force evaluation can be neatly separated into classical and quantum mechanical force evaluation. LOTF95 has been designed to allow both of these to be obtained from 'black boxes' that take atomic coordinates and lattice parameters and produce forces. This allows new force models to be added quickly and easily, in contrast to multi-Hamiltonian schemes such as QM/MM which are generally built around a single quantum package.

Classical Force Models

The classical interatomic potentials used in this work have been outlined in Section 2.3: in this thesis I use the Stillinger-Weber potential for silicon and the Brenner potential for carbon. These simple potentials have been implemented within the libAtoms framework. To ensure that a potential has been correctly implemented, it is important to test that the forces that result from the analytical derivative of the potential with respect to atomic coordinates agree accurately with the gradient of the potential evaluated numerically by finite differences. This simple test, commonly referred to as the gradient test, catches many programming errors.

To prevent the classical force evaluation from becoming a bottleneck for very large systems, it has been necessary to parallelise this part of the code. Since the potentials are local this can be done in a very straight-forward fashion using the Message Passing Interface (MPI) library. The atoms are shared evenly among all available processors. Each processor then computes the forces on the atoms it has been assigned. The resulting partial force arrays are then combined to give the overall force.

Interfaces to the popular classical molecular dynamics codes DL_POLY [111] and AMBER [112] have been developed within the LOTF95 code. These programs provide ready access to features such as the Ewald algorithm, essential for applications involving long range Coulomb interactions. The systems studied in this work do not contain any significant long range interactions, so these features were not required for the application described in Chapters 6 and 7.

Elastic Matching

It is important that there is no discontinuity in elastic behaviour at the quantum/classical boundary or phonon waves would be unphysically reflected from the artificial interface. Provided that the elastic constants of the two models are in good agreement, we have demonstrated that LOTF gives no reflection of sound waves at the boundary between models.

Ideally to avoid the boundary reflection problem, one would refit the classical potential to exactly reproduce all the elastic constants of the quantum mechanical model; unfortunately this is not always possible to implement in practice. Of all the relevant properties, the lattice constant and bulk modulus are the most important quantities to match between the QM and classical models. The unmodified Stillinger-Weber potential does a surprisingly good job of describing the elastic behaviour of silicon in the diamond phase, for the theoretical reasons discussed in Section 2.3. By rescaling the distance scale σ and refitting the two body parameters A and B, we can exactly reproduce the lattice constant and bulk modulus of a given quantum model. This must be done separately for each QM package that we wish to interface into LOTF95. For the case of CASTEP, this procedure gives a Stillinger-Weber potential that provides a reasonably good description of the other elastic constants. The largest error is in C_{44} and is around 30%. It is possible to improve the fit to the cubic elastic constants by varying the λ parameter that controls the stiffness of the three body interaction. However, doing this drastically alters other physical properties of the Stillinger-Weber potential such as the melting point. In fact, as we shall see in Section 5.5, doubling the value of λ changes the phenomenology of fracture predicted by the SW model from ductile to brittle [113]. In order to avoid unexpected side effects like this, I have chosen not to attempt to further improve the elastic constant fit. Since I shall be carrying out crack simulations under uniaxial tension, an accurate C_{11} constant is the most important to describe accurately and the shear moduli C_{12} and C_{44} are of secondary concern.

Quantum Mechanical Force Evaluation

Quantum mechanical force evaluation within the LOTF scheme is a more elaborate process than obtaining the classical forces. Since the evaluation of the QM forces is so expensive, we are willing to take on extra complexity in order to gain speed. For example, passivating the QM region with hydrogen atoms adds to the complexity, but allows the buffer thickness to be decreased for the same force accuracy. Typically, QM calculations scale as $O(N^3)$ in the number of atoms in the cluster so this makes a very significant difference to the computational efficiency of the method. Two different QM force evaluation methods are employed in LOTF: we can either use multiple small clusters or a single cluster containing all the quantum mechanical atoms.

Multiple Small Clusters In this approach we carve a cluster around each of the atoms that we wish to treat with QM accuracy as illustrated in Fig. 4.6a. In C-LOTF this was done with a spherical cutoff, but in LOTF95 we have moved over to carving clusters entirely using bond hopping to reduce fluctuations in the clusters from one time step to the next. The cluster is then terminated chemically by replacing atoms outside the cluster with hydrogen atoms, positioned along the cut bond at a distance

$$r_{\mathrm{term}} = r_{ij} \cdot \frac{r_{\mathrm{cov}}(Z_i) + r_{\mathrm{cov}}(H)}{r_{\mathrm{cov}}(Z_i) + r_{\mathrm{cov}}(Z_j)}. \tag{4.14}$$

The force on the central atom in the cluster has to be converged with the radius of the cluster and, for *ab initio* methods, with the amount of vacuum surrounding the cluster. This approach is trivially parallelisable: the QM force evaluation for each

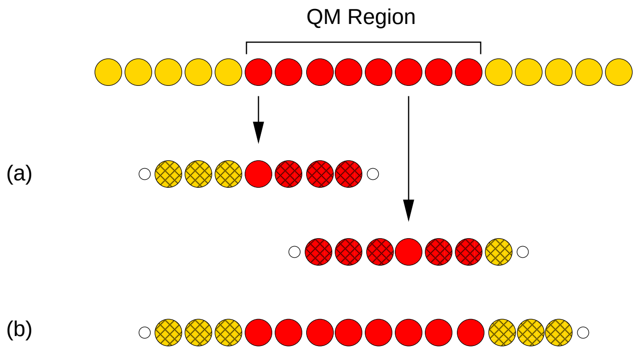


Figure 4.6: Quantum mechanical force evaluation in LOTF, using: (a) small clusters centred on each QM atom (b) one large cluster containing all QM atoms. Atoms are coloured according to their original region: red for QM and yellow for classical. Cross hatching indicates the buffer atoms in each cluster calculation. The small white circles represent hydrogen termination atoms.

small cluster is an independent calculation that can be delegated to a processor. In this way, excellent scaling up to very large numbers of processors is possible since very little communication between nodes is required. These small clusters typically contain between fifty and a hundred atoms.

Single Large Cluster In this alternative method, a single large cluster is carved, including all the QM atoms plus a buffer region as illustrated in Fig. 4.6b. This cluster is terminated in the same way as the individual clusters described above. Only the forces on the original QM atoms are kept. In this case, the calculation has to be converged with respect to the thickness of the buffer zone, rather than with respect to the cluster radius as is done in the multiple small clusters approach described above. In this case we rely on the parallelisation of the QM black box to share the work of force evaluation between multiple processors. For the simulations reported in this thesis, the typical cluster size in this approach is around 200 atoms.

Which approach is better depends on the scaling performance of the QM package used as well as the geometry of the system. An example in which the second approach can be more efficient is the case of a slab geometry, where the system is very thin in one direction. We can then carve clusters that are periodic in this direction, to take advantage of the excellent parallelisation over **k**-points in modern DFT codes. As a result of their relatively high startup overheads and excellent scaling behaviour, linear scaling DFT techniques such as ONETEP [114] would be also be better suited to the single large cluster scheme.

Tight Binding

In the initial validation and application of LOTF, a variety of empirical tight binding methods [87, 88] were used. For many systems, these methods strike a good balance between computational efficiency and accuracy. The matrix diagonalisation algorithms used to solve the tight binding Hamiltonian do not parallelise well, so the small cluster method is generally preferred when using LOTF with tight binding force models. The Hellman-Feynman force is computed after a direct diagonalisation of the cluster system. Forces on the central atom of small clusters converge quickly with cluster radius (see Fig. 3.4), and are found to be relatively insensitive to the details of the termination strategy. Since only these forces are kept, the atomic trajectories are not affected by the representation of the boundary.

LOTF95 also has an interface to the DFTB quantum mechanical package, which implements the density functional tight binding method described in Section 2.5. This gives a more accurate description of the underlying physics, especially when used with charge self consistency, although at a greatly increased computational cost.

Density Functional Theory: An Interface to CASTEP

In many cases, tight binding does not provide a sufficiently accurate description of the physics we wish to study and we are forced to use a more expensive method such as DFT. However given the high computational cost of *ab initio* calculations, a sensible methodology is to broadly explore the system of interest using a series of fully classical calculations, since these are very cheap in comparison with the hybrid method. This allows us to converge the qualitative behaviour with respect to system size and time scale parameters that do not depend on the precise atomistic details. We can then move to tight binding dynamical simulations to assess the effect of treating the region of interest quantum mechanically and modify the simulation parameters accordingly. The final accurate calculations can then be performed using a very accurate *ab initio* method, hopefully using a single simulation.

Earlier versions of LOTF have been interfaced to the SIESTA package [115], which uses a local orbital basis set, but in this work I have used the CASTEP plane-wave code for all DFT calculations. As a result of the 'black box' design principle, the interface between LOTF95 and CASTEP is fairly straightforward. The combination

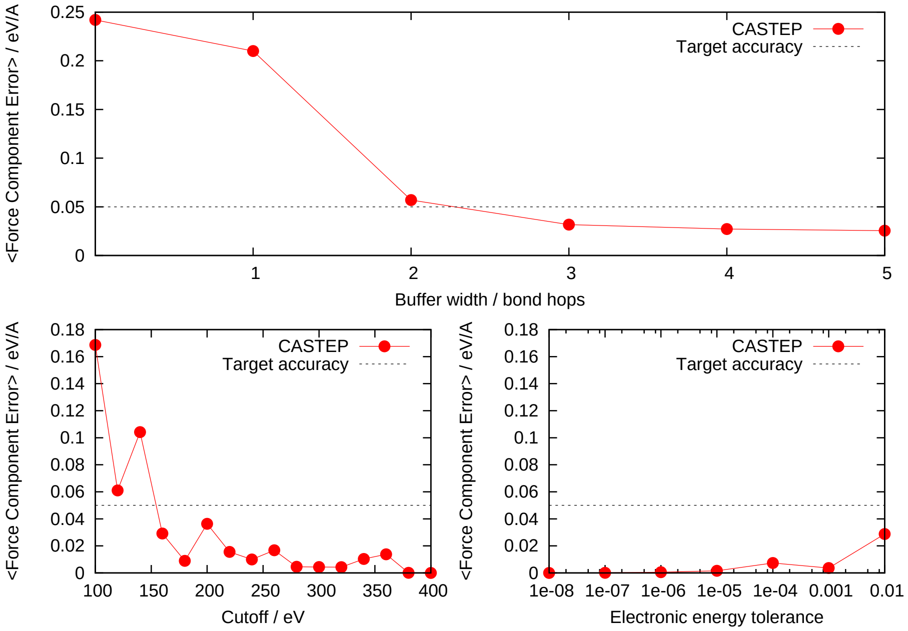


Figure 4.7: Upper plot: Mean force component errors as a function of the buffer thickness, compared to a fully periodic calculation, using CASTEP with clusters carved from a 72 atom slab of bulk silicon containing a vacancy. The force error converges rapidly: three hops are sufficient to set the target accuracy for the CASTEP calculation to 0.05 eV/Å. Below: for a fixed cluster of 36 atoms, we consider convergence of the force error on the central atom with: the cutoff energy, with electronic energy tolerance fixed at 10^{-4} (lower left) and the electronic energy tolerance, with cutoff fixed at 240 eV (lower right).

of the relatively high overhead for each CASTEP calculation and its good parallel performance — especially for systems with multiple **k**-points — mean that we invariably use the single cluster strategy of Fig. 4.6b.

Hellman-Feynman forces are obtained from a single point calculation on the cluster in which the electronic degrees of freedom are fully relaxed. The resulting ground state electronic density $n(\mathbf{r})$ is saved and used as an initial guess for the following time step, hugely reducing the time taken for each calculation.

CASTEP is a plane wave code, so it uses periodic boundary conditions. To simulate an isolated cluster, we must add sufficient vacuum to simulate eliminate interactions with periodic images. The forces on the interior atoms must be converged with respect to the buffer thickness and the amount of vacuum added. The error caused by using a finite buffer zone sets the target accuracy for the CASTEP calculation: it is a waste of time converging forces more accurately than this. We therefore choose the planewave cutoff energy and electronic minimisation tolerance accordingly. This convergence process is illustrated in Fig. 4.7 for the example case of a slab of bulk silicon containing a vacancy. These calculations indicate a planewave cutoff of 150 eV and electronic energy tolerance of 10^{-2} eV give sufficient accuracy for the forces. These tolerances are much weaker than those typically required for *ab initio* molecular dynamics. The additional accuracy is not needed because the adjustable potential (and to a lesser extent the thermostat) maintain the dynamical and numerical stability of the atomic trajectories.

Hydrogen passivation when using CASTEP is almost identical to the general case described at the beginning of this section, with one small change. To avoid the factor of two overhead of performing a spin polarised calculation, we require the number of up- and down-spin electrons in the system to be equal. It may be necessary to omit a termination atom to enforce this constraint: if so we choose a buffer atom that has a maximal number of passivation hydrogens attached to it and remove one of these. Providing we have converged with respect to the buffer thickness this will not affect the final QM forces. To remove any systematic error associated with the precise details of the buffer and termination, we slightly randomise the positions of the outermost layer of buffer atoms.

4.8 The Adjustable Potential

The choice of the form of the adjustable potential $V_{\mathrm{adj}}(\mathbf{R})$ is crucial to the success of the LOTF scheme. This has been the focus of considerable recent development. In this section I chart the evolution of the adjustable potential from the initially proposed modification of the parameters of an existing classical potential to the current, simpler form which is based upon a set of linear springs.

Modified Stillinger-Weber Potential

The LOTF77 and C-LOTF implementations of LOTF used the classical potential itself as the adjustable potential, rather than adding a corrective potential in the way outlined in Section 4.2. This is an appealing viewpoint when the corrections to be made are small: we regard the classical Hamiltonian as universal, and improve its accuracy in certain places in space and time by computing new information in those places using the quantum model. We can incorporate this information into the classical trajectories by slightly adjusting the parameters of the potential in the relevant region.

This was accomplished for silicon systems using a version of the Stillinger-Weber potential with variable parameters. Equations 2.32 and 2.33 for the two- and three-body parts of the potential were replaced by

$$\begin{aligned} f(r_{ij}) &= \left(\frac{A_{ij}}{r_{ij}^4} - B_{ij}\right) f_{\text{cut}}(r_{ij}) && (4.15) \\ h(\theta_{jik}) &= \lambda \left(\cos\theta - \cos\theta_{jik}^0\right)^2 && (4.16) \end{aligned}$$

where A_{ij}, B_{ij} and θ_{jik}^0 are no longer constant parameters, but instead can take different values for each bond and angle in the system. With this form of the corrective potential, the objective functional used to optimise the parameters is simply given by the sum of the squared differences between the classical and quantum forces, so Eq. 4.1 is replaced by

$$\mathcal{F} = \sum_{\text{all atoms}} \left|\mathbf{F}_{QM} - \mathbf{F}_{\text{classical}}\right|^2 . \qquad (4.17)$$

The optimisation was limited to a fit region extending 12 Å from the QM region containing N_{fit} atoms. The functional $\mathcal{F}$ is minimised using the conjugate gradients algorithm; for a typical silicon system about 30 steps in parameter space are sufficient to achieve a converged fit with a maximum deviation from the target forces of around 0.01 eV/Å [42].

A special reformulation of the classical potential is needed to make it computationally viable to evaluate the potential and its derivatives for fixed atomic positions $\mathbf{R}$ thousands of times during each optimisation. This is done by recasting it into the form

$$V_{\text{classical}}(\mathbf{R}, \boldsymbol{\alpha}) = \mathbf{W}_0 \cdot \mathbf{1} + \mathbf{W}_1 \cdot \boldsymbol{\alpha} + \mathbf{W}_2 \cdot \boldsymbol{\alpha}^2 \qquad (4.18)$$

where only the vectors $\mathbf{W}_n$ depend on the atomic coordinates and can be computed once at the start of the optimisation. The force and the derivatives of the potential with respect to the parameters can be rewritten in a similar way.

Spline Potential

Modifying the parameters of a universal potential does indeed work; this approach was used for all the validation tests described above. However, this viewpoint can be misleading if the changes required to the potential are not small, as is often

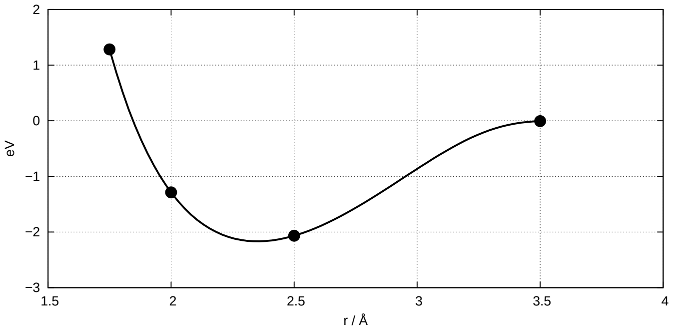


Figure 4.8: Example of a two body adjustable spline potential for silicon. The large dots are the spline knot points, the centre two of which can move during optimisation. They are constrained to lie below zero so that the function is guaranteed to have a minimum. The derivative and value are set to zero at 3.5 Å, and fixed to match the quantum model at 1.75 Å.

the case in practice. Furthermore, the work associated with converting each new classical potential into the form of Eq. 4.18 makes it very time consuming to apply this method to new systems. A more flexible approach is to choose a functional form for the adjustable potential that offers a good compromise between expressive power and robustness. This method has been adopted in the LOTF95 implementation.

We have now returned to the 'new' way of looking at the LOTF method outlined in Section 4.2 above: we want to add a simple adjustable potential to an existing classical model, and tune the parameters of this simple potential to reproduce the correct quantum mechanical forces. The question is: what form should this simple potential take? Since we will be changing the potential parameters on the fly, it is desirable for it to vary smoothly with these parameters. In contrast to the modified SW potential, we choose to construct our new adjustable potential so that its gradient with respect to the parameters is easy to evaluate, since this will have to be done many times when optimising the parameter values. A cubic spline functional form fulfils these requirements. Keeping the bond lengths and bond angles as the fundamental coordinates, we can create cubic splines for the two- and three-body parts of the potential: For example, for silicon, the two body term could be represented by a spline with four knots at $r = \{1.75, 2.0, 2.5, 3.5\}$ Å, as shown in Fig. 4.8. For other elements, the knot points would be rescaled appropriately by the atomic radius. A three body potential could also be constructed in a similar way, using a cubic spline of $\cos\theta$ to allow for an asymmetric angular dependence.

The spline adjustable potential would then have this form:

$$V_{\text{spline}}(\mathbf{R},\boldsymbol{\alpha}) = \sum_{\substack{\text{bonds}\\ i,j\\ i<j}} s_1(r_{ij},\boldsymbol{\alpha}) + \sum_{\substack{\text{angles}\\ i,j,k\\ i<j<k}} s_2(\cos\theta_{jik},\boldsymbol{\alpha}) \tag{4.19}$$

The spline functions s_1 and s_2 are linear functions of the values at the knot points — our free parameters — so evaluating the splines and their derivatives for fixed atomic positions in the fitting process is straightforward and efficient [116]. We can do this by reshaping the components of the adjustable force into a $3N_{\text{fit}}$ dimensional vector $\mathbf{F}_{\text{spline}}$. This can be written

$$\mathbf{F}_{\text{spline}} = \mathbf{a} + B\,\boldsymbol{\alpha} \tag{4.20}$$

where the $3N_{\text{fit}}$ dimensional vector $\mathbf{a}$ and $N_{\text{param}} \times 3N_{\text{fit}}$ dimensional matrix B depend only on atomic coordinates and so need to be computed only once before the optimisation. The objective function for the optimisation and its derivative with respect to $\boldsymbol{\alpha}$ are then given by

$$\begin{aligned} \epsilon &= \left|\mathbf{F}_{\text{target}} - \mathbf{F}_{\text{spline}}\right|^2 && (4.21)\\ \nabla\epsilon &= -2\left(\mathbf{F}_{\text{target}} - \mathbf{F}_{\text{spline}}\right) B && (4.22) \end{aligned}$$

where $\mathbf{F}_{\text{target}}$ is a $3N_{\text{fit}}$ dimensional vector of the force differences between classical and quantum force models. For atoms in the fit region but not in the QM region the target force difference is zero.

In practice we find that it is possible to get very good parameter fits using only two-body spline potentials, providing that the splines are long enough to connect second neighbours. This observation begs the question: is a physically motivated pair potential a useful starting point from which to construct the adjustable potential? I shall address this question in the following section.

Linear Spring Potential

Our experience suggests that in regions where the classical potential provides a poor description of the forces on the QM atoms, there is no correlation between the classical and accurate quantum forces, and that that the correction forces required can be as large or larger than the classical forces themselves. The requirements for a good adjustable potential are twofold: firstly we want to be able to apply these large corrective forces in all directions, and secondly we want the parameters of our potential to vary slowly in comparison with the atomic trajectories, in the sense of Eq. 4.7, so that we can use the predictor-corrector algorithm.

A sensible methodology is to deal with the first of these requirements by constructing a very simple potential to correct the force errors, and then assess its interpolation performance in comparison with the spline and Stillinger-Weber based adjustable potentials. A suitable potential for doing this is a set of linear springs that presupposes nothing about the functional form required:

$$V_{\text{spring}}(\mathbf{R}, \boldsymbol{\alpha}) = \sum_{\substack{\text{bonds} \\ i,j \\ i<j}} \alpha_{ij} r_{ij} + \sum_{\substack{\text{angles} \\ i,j,k \\ i<j<k}} \beta_{jik} \cos\theta_{jik} \tag{4.23}$$

The first term describes a set of two body springs, linear in the interatomic distance r_{ij} and the second a set of three body springs linear in $\cos\theta_{jik}$.

What motivates this choice of functional form? Firstly, we haven't quite abandoned all physical intuition: we are assuming that bond lengths and angles will give rise to coordinates that move more slowly than the atomic coordinates themselves. Secondly, the linearity of the potential has important practical consequences. We can write the force from this potential as a single linear operation on the parameters:

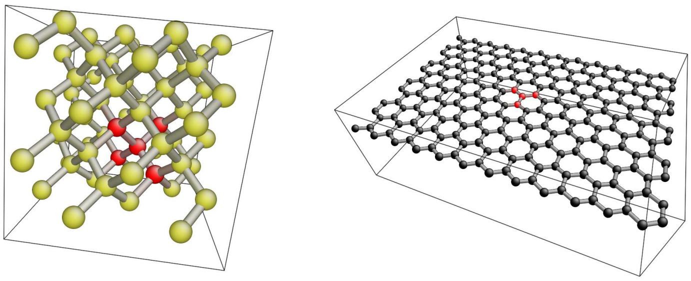
$$\mathbf{F}_{\text{spring}} = C\,\boldsymbol{\alpha} \tag{4.24}$$

where C is a matrix of size $N_{\text{param}} \times 3N_{\text{fit}}$ that once again only depends on the atomic positions. Each two body spring contributes six elements to C and each three body term contributes nine elements.

We seek the parameters $\boldsymbol{\alpha} = \{\alpha_{ij}, \beta_{ij}\}$ that make $\mathbf{F}_{\text{spring}}$ as close as possible to the target force differences $\mathbf{F}_{\text{target}}$. The singular value decomposition (SVD) of the matrix C can be used to find the least squares solution for the parameters $\boldsymbol{\alpha}$. This solution is defined by

$$\min |\mathbf{F}_{\text{target}} - C\,\boldsymbol{\alpha}|^2 \tag{4.25}$$

Compared to the conjugate gradient approach to minimisation that we were forced to adopt for the modified Stillinger-Weber and spline potentials, SVD has two advantages. Firstly, we are guaranteed to find the least squares minimum, with no chance that the optimisation process will get stuck in local minima. Secondly, we



(a) (b)

Figure 4.9: Systems used to test the various adjustable potentials: (a) 64 atom silicon bulk cube with 5 QM atoms (b) 240 atom graphene sheet with 4 QM atoms. The atoms to be treated with quantum mechanical accuracy are shown in red in both cases.

can take advantage of extremely fast linear algebra routines to perform the SVD calculation, such as those available in the LAPACK libraries.

Tests using the simple systems shown in Fig. 4.9 show that the linear spring potential produces fits which are at least as good as previous versions of the adjustable potentials, with the maximum force error after fitting being typically less than 0.01 eV/Å.

Choosing the Springs

Initially, splines or springs were added between all pairs of atoms separated by less than a critical distance, in exactly the same way as a cutoff distance is used to limit the range of classical potential. Unlike traditional classical potentials, neither the value nor the derivative of the spring potential are zero at this cutoff, therefore as atoms oscillate in and out of range, springs get turned on and off, affecting the stability of potential energy and forces. Two changes have been made to avoid this.

Firstly, we replace the cutoff sphere with a bond hopping approach. The fit region can be produced by growing the QM region by some number of bond hops in all directions. We then choose our spring set by considering each atom in the fit region in turn and adding springs by hopping along bonds. The length of the springs produced can be controlled with a parameter indicating the number of bonds to hop along. Fig. 4.10 shows the effect of varying the size of the fit region and the maximum spring length on the transferability of the spring potential. For a spherical fit region, the total number of springs increases approximately as the cube of the maximum

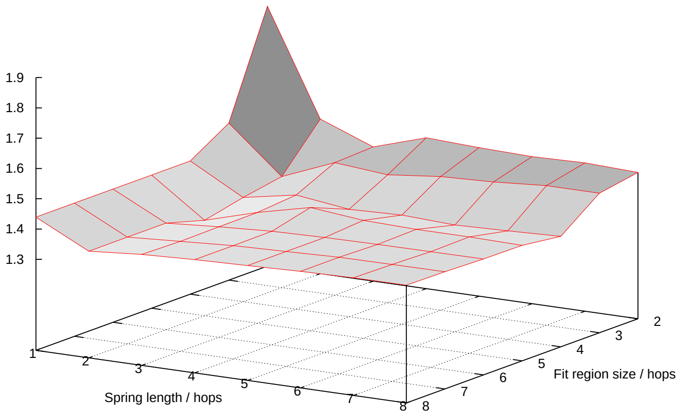


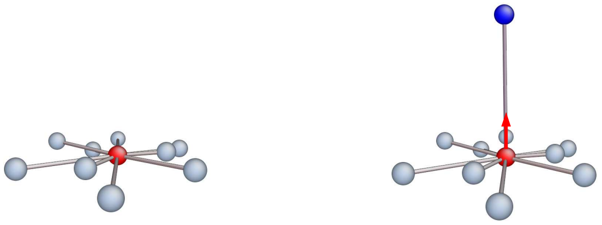
Figure 4.10: Maximum error in hybrid force (in eV/Å) at the end of 10 fs extrapolation with fixed spring parameters for the graphene system of Fig. 4.9b, averaged over many independent trajectories, as a function of fit region size and maximum spring length.

spring length. Since the SVD algorithm is effective at identifying degenerate degrees of freedom which might otherwise introduce numerical instabilities in the fitting procedure, introducing large numbers of springs does not pose too great a problem. However, it is obviously inefficient to use vastly more springs than are necessary to achieve a good fit. For the system shown in Fig. 4.10, we can conclude a fit region of four hops and springs of maximum length two hops gives a good compromise between optimisation speed and fit accuracy.

Secondly, there is no reason to allow the set of springs to change every time we want to re-optimise the potential. We only need to determine which springs to use whenever the QM region moves.

Spring Directionality Criteria

Sometimes pathological atomic configurations lead to large force errors after the adjustable potential optimisation. An example is when the neighbours of an atom are almost coplanar, as shown in Fig. 4.11a. If springs are attached only to these atoms, then there is no way to apply a corrective force perpendicular to this common plane. This obviously applies particularity strongly for planar systems such as the graphene sheet illustrated in Fig. 4.9b, but it still causes problems in far less extreme cases.



(a) (b)

Figure 4.11: (a) The set of springs (shown as bonds) connected to the central (red) atom are almost coplanar. (b) A new spring is added to the blue atom, in the direction of the smallest eigenvector of the directionality ellipsoid (red arrow). This is sufficient to give a good spanning of 3D space.

We can solve this problem by adding extra springs in the missing directions for these difficult atoms. This is done by first adding default springs as described above, then using these springs to compute a directionality ellipsoid for each atom in the fit region, represented by a 3 × 3 matrix:

$$E_i = \sum_{\text{springs } ij} \hat{\mathbf{r}}_{ij} \, \hat{\mathbf{r}}_{ij}^T \tag{4.26}$$

where $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/r_{ij}$. E_i is the average over all springs that connect atom i to other atoms of the outer products of the normalised spring vectors with themselves. The eigenvectors and eigenvalues of this matrix define the axes of the directionality ellipsoid. If the springs are isotropically distributed then the three eigenvalues will be of similar magnitude. One large eigenvalue suggests roughly linearly clusters, while two of similar value and one smaller eigenvalue suggests a planar distribution.

Once the difficult atoms have been identified, we can add more springs by looking for atoms which lie close to the line defined by the eigenvector corresponding to the smallest eigenvalue of E_i, as shown in Fig. 4.11b. We can repeat this procedure until the distribution of springs is sufficiently close to spherical, choosing the shortest possible new spring to add at each step.

For the planar graphene system, we note that periodic boundary conditions imply we are really simulating an array of stacked graphene sheets, well-separated to prevent interactions between separate sheets. The directionality condition can be met by allowing springs to span the periodic boundary and connect to images of other atoms. Springs are forbidden from joining an atom to its own periodic image since this would violate conservation of momentum. With these directionality modifications, the linear spring potential allows a good fit to arbitrary forces to be achieved.

Measuring Transferability

The three functional forms discussed above give broadly equivalent fit accuracy in the QM forces across a wide range of test systems. However, it is possible that they could give rise to different interpolation behaviour: we might expect that more physically based potentials would result in slower varying parameters and consequently better interpolation performance. This can be tested by evaluating the full hybrid force at each small time step and calculating the force error:

$$\mathbf{F}_{\mathrm{error}} = \mathbf{F}_{\{\mathrm{QM,classical}\}} - \mathbf{F}_{\mathrm{LOTF}}. \tag{4.27}$$

This has been done for the two systems shown in Fig.4.9 which were chosen to be representative structures for the multiscale systems of interest. The graphene system is not relevant to the simulations reported in this thesis, but is included here as an illustration of the different interpolation behaviour exhibited by different physical systems.

For each of these systems, 1000 independent trajectories were generated by randomising the initial atomic positions and velocities. The force errors during one predictor-corrector cycle were recorded and then averaged over the trajectories to produce distributions of RMS and maximum force errors as a function of time during the extrapolation and interpolation. Figs. 4.12 and 4.13 show examples of these distributions.

The force errors rise approximately linearly during the extrapolation part of the cycle, as we move away from the point in phase space where the potential parameters were fitted. Providing the range of validity of the new parameters fitted at the end of the extrapolation is large enough, the force errors should remain small throughout the interpolation. We must choose the number of interpolation steps N_{interp} appropriately to ensure this is the case The appropriate scale is set by the accuracy to which the forces have been computed using the finite buffer method; for CASTEP this is around 0.05 eV/Å. For the silicon test system, we can see from Fig. 4.12 that $N_{\mathrm{interp}} = 10$ gives acceptable accuracy: the RMS deviation typically

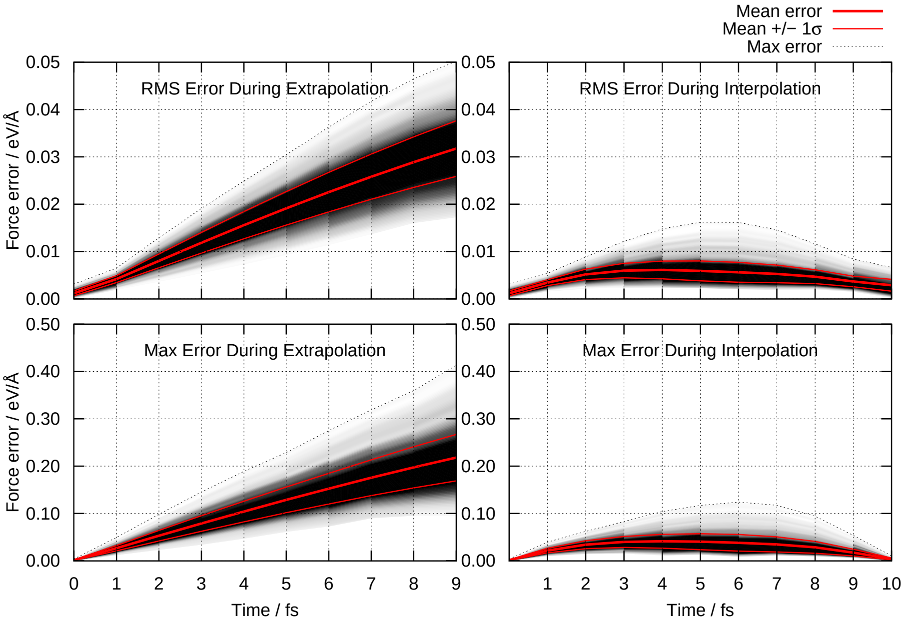


Figure 4.12: Force errors during one predictor corrector cycle, averaged over many independent trajectories. Test system is the silicon bulk cube of Fig. 4.9a at a temperature of 2000 K, with MD timestep $\Delta t = 1$ fs and $N_{\mathrm{interp}} = 10$. The gray-scale colouring shows the density of force errors as a function of time, with black corresponding to high and white to low densities. The mean and standard deviation of the distribution are indicated by the red lines, and the extreme values by the black dashed lines. The linear spring adjustable potential with two body springs only was used. QM force model is tight binding, from Kwon *et al.* [87].

remains smaller than 0.01 eV/Å, and the maximum force error is typically below 0.1 eV/Å.

Interpolation is not always possible: for the graphene system force errors of 0.5 eV/Å are typical with $N_{\mathrm{interp}} = 10$ (see Fig. 4.13). The MD timestep has been decreased from 1 fs to 0.64 fs to take account of carbon's lower atomic mass. Reducing the number of steps reduces these errors, but not enough to make the interpolation accurate, even for N_{interp} as low as two. Even when the directionality criteria described above are used to add extra springs, the initial fit accuracy is lower for the graphene system than for silicon. The poorer fits are caused partly by limitations on where springs can be placed imposed by the geometry of the system, and partly by the artificial stiffness of the classical potential. The fit of the classical potential elastic constants to the *ab initio* values ensures a good description of the equilibrium properties, but even at modest temperatures the displacements

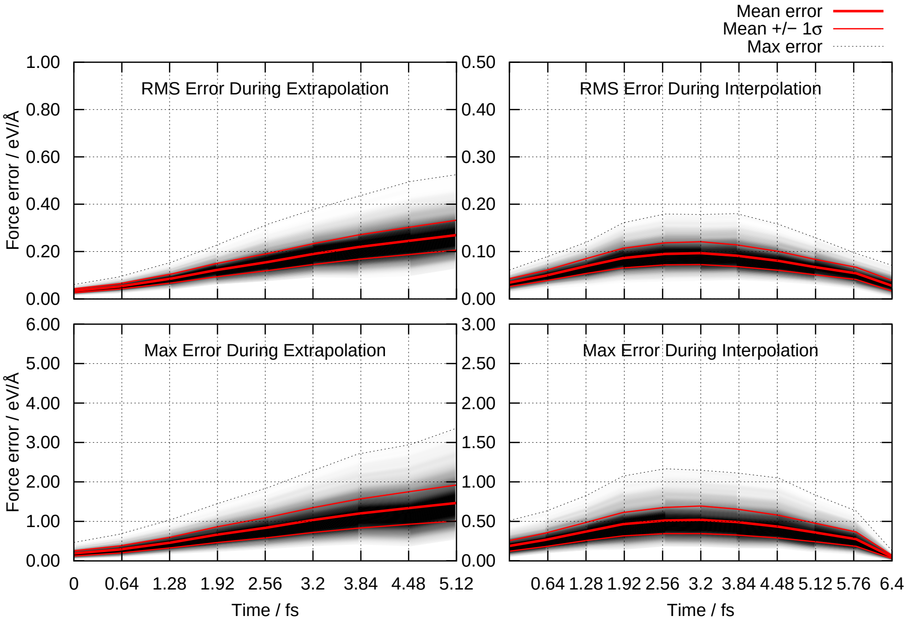


Figure 4.13: As in Fig. 4.12, but the test system is that shown in Fig. 4.9b, again at a temperature of 2000 K, with MD timestep $\Delta t = 0.64$ fs and $N_{\text{interp}} = 10$. The linear spring adjustable potential with two body springs only was used. The QM force model was the DFTB method of Frauenheim *et al.* [91].

from equilibrium are large enough to give rise to large force errors. Furthermore, this stiffness causes large changes in the classical force for very small atom movements. Therefore the adjustable potential parameters need to vary rapidly in order to correct for these erroneous forces, making interpolation difficult.

Three Body Springs and Interpolation Order

The transferability tests were repeated with each of the three adjustable potentials. There was no significant difference in the distribution of force errors produced by the different potentials for either the silicon or carbon system. Moreover, the force error distribution is essentially unchanged when three body springs are included. Since we expect the speed with which the adjustable potential parameters change to limit the possible interpolation time, this suggests that parameters based on bond lengths do not change any faster than those based on angles. The choice of linear or cubic parameter interpolation makes only a very small difference to the measured force errors, as does repeating the predictor corrector cycle two or more times. These conclusions are illustrated in Fig. 4.14.

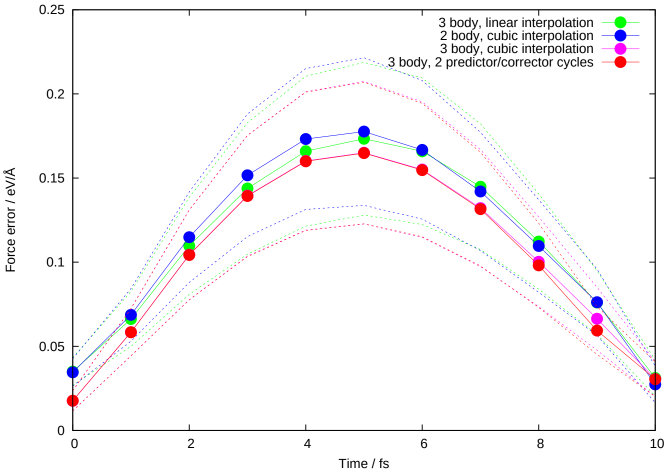


Figure 4.14: RMS force error distributions during interpolation obtained from averaging 1000 independent trajectories of the graphene test system of Fig. 4.9b. The solid lines are the mean μ, and the dashed lines are $\mu \pm 1\sigma$. There is no statistically significant difference in the distributions obtained with two and three body springs and linear interpolation (green), two body springs only and cubic interpolation (blue), two and three body springs and cubic interpolation (pink) or two and three body springs and two iterations of the predictor corrector cycle.

Summary of Adjustable Potential Results

At first sight it appears that we must lose something by throwing away all the physical intuition that lies behind the functional form of traditional pair potentials. However, such intuition becomes meaningless when the potentials have to be stretched to second neighbour distance to achieve a good fit, as is the case for both the modified SW and spline adjustable potentials.

Given the equivalent fit accuracy and interpolation performance, the linear potential is preferred on the grounds of simplicity and computational efficiency. The results presented in this thesis have been obtained using a combination of the spring and spline methods described above. Two body springs have been used for all the simulations reported in Chapter 7.

4.9 Energy Conservation

If the Hamiltonian for a system is time dependant, as is the case with LOTF, then energy will not be conserved. This is inevitable if the QM region moves, since then the set of springs will change and the new Hamiltonian will be incompatible with the

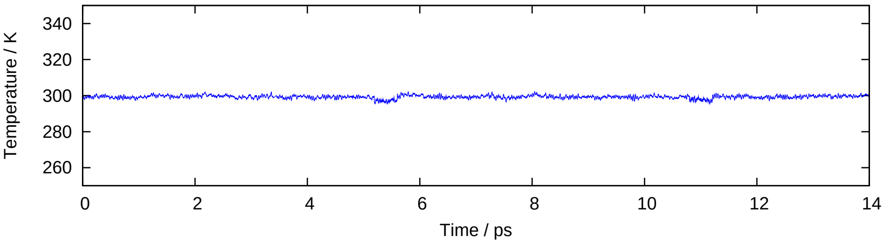


Figure 4.15: Temperature evolution during a LOTF hybrid simulation of silicon fracture containing approximately 100 000 atoms. A weak Langevin thermostat with a time constant $\tau = 0.5$ ps is adequate to adsorb the energy drift caused by the time-dependent Hamiltonian.

old one. If we consider a fixed QM region, then the condition for energy conservation is that the overall potential is conservative, *i.e.* motion in small loops does no work.

The time dependent Hamiltonian gives rise to a series of different potential energy surfaces. For energy to be conserved, small loops in phase space must return to the same starting point, with no discontinuous jumps. This requires the surface normals of the consecutive potential energy surfaces to match at the boundary. However, this means that the forces produced by the potentials will also match at the boundary. It is therefore not possible both to reproduce the new QM force and to conserve energy, so if the force matching is good we should expect to see a temperature drift.

In practice, small energy drifts can be absorbed using a thermostat to maintain constant temperature, as described in Section 2.2. The Langevin thermostat is ideal for this purpose since it couples to each mode of the system independently. As shown in Fig. 4.15, we typically require only a weak thermostat to maintain a constant temperature, so the dynamics of the system should not be adversely affected. The NVT ensemble that we obtain using a thermostat will approach the microcanonical ensemble for very large systems; in fact such large systems are self thermostating.

Conservation of the Extended Energy

If the QM region does not move, there is a conserved quantity in a LOTF simulation, although it is not the total energy: we can modify Eq. 2.1 to create a time independent Hamiltonian:

$$\widetilde{H} = T(\{\mathbf{p}_i\}) + V(\{\mathbf{r}_i\}, t) + \epsilon \tag{4.28}$$

where ϵ is a new dynamical variable conjugate to the time t. Applying Hamilton's equations of motion for t and ϵ yields

$$\dot{t} = 1 \tag{4.29}$$

$$\dot{\epsilon} = -\frac{\partial V}{\partial t} \tag{4.30}$$

Eq. 4.30 can be numerically integrated to keep track of ϵ during the dynamics. The sum of kinetic energy, potential energy and ϵ is a constant of the motion, referred to as the *extended energy*.

4.10 Summary

In this chapter, I have introduced the 'Learn on the Fly' method for coupling quantum and classical molecular dynamics simulations and described in detail how the scheme works. Compared to previous QM/MM approaches, the most important feature of LOTF is that we do not attempt to form a single Hamiltonian for the hybrid system, but instead concentrate on local forces. Lifting this restriction allows accurate quantum forces to be combined with classical forces using a simple adjustable potential to give stable dynamics. I have described recent changes to the method, in particular a series of simplifications to the adjustable potential and an interface to the planewave DFT code CASTEP.

The aim of this thesis is to use the LOTF method to study a problem which cannot be attacked in any other way: the prototypical multiscale system, brittle fracture. To this end, in the following chapter we shall consider fracture in more detail, before going on to describe how the LOTF method can be applied to the brittle fracture of silicon systems in Chapter 6.

Chapter 5

Fracture of Brittle Materials

5.1 Introduction

In this chapter, I first give an overview of the continuum mechanics techniques that have been used to analyse brittle fracture since the beginning of the 20th century. I then consider the limitations of this approach when considering real materials which contain a discrete lattice of atoms. Taking silicon as a case study, I review existing experimental, theoretical and computational studies of brittle fracture, considering in particular the failure of classical interatomic potentials to accurately describe the process. This failure motivates the use of multiscale hybrid techniques to model brittle fracture. The chapter concludes with an overview of previous hybrid simulations of fracture.

5.2 Continuum Fracture Mechanics

Until relatively recently, brittle fracture has been analysed exclusively by continuum mechanics techniques. In this section I outline these techniques and review some relevant results from the field. For more details, see Lawn [34], Broberg [30] and Freund [31].

Continuum fracture mechanics was pioneered by Inglis and Griffith at the beginning of the 20th century. Inglis [32] noted in 1913 that the local stress at the end of an elliptical cavity can be several times larger than the applied stress. Stress concentration is greatest where the radius of curvature is a minimum (indicated by an arrow in Fig. 5.1). Griffith [33] took the limiting case of an infinitely narrow elliptical cavity, and modelled this as a reversible thermodynamic system. He noted

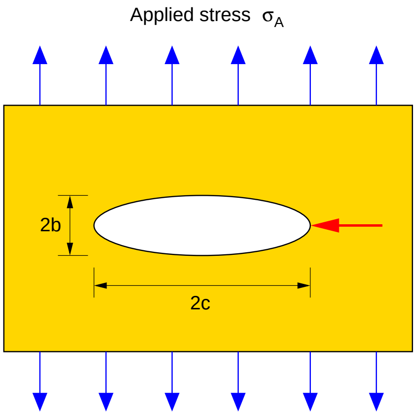


Figure 5.1: Plate containing an elliptical cavity as analysed by Inglis. The semi-axes of the ellipse are b and c. Uniform tension σ_A is applied, resulting in stress concentration at the notch tips, one of which is indicated with the red arrow.

that a crack is in equilibrium when the elastic energy U_E of the strained slab is balanced by the surface energy U_S of the crack. For unit width of crack front, these energies are

$$U_E = \frac{\pi c^2 \sigma_A^2}{E'} \tag{5.1}$$
$$U_S = 4c\gamma \tag{5.2}$$

where σ_A is the applied stress, and γ the surface energy density. E' is the effective Young's modulus, and is different depending on whether we have plane stress (thin plate) or plane strain (thick plate) conditions:

$$E' = \begin{cases} E & \text{Plane stress} \\ E/\left(1-\nu^2\right) & \text{Plane strain} \end{cases} \tag{5.3}$$

where E is the Young's modulus and ν is Poisson's ratio for the material. Equating U_E and U_S and rearranging gives the critical loading:

$$\sigma_c = \sqrt{\frac{2E'\gamma}{\pi c}} \tag{5.4}$$

Since $U_E \sim c^2$ and $U_E \sim c$ this critical loading is length dependent. For $\sigma_A < \sigma_c$ the crack closes up, while for $\sigma_A > \sigma_c$ it grows spontaneously. Conversely, for a given loading σ_A there is a critical size of crack c_0 below which no propagation will occur.

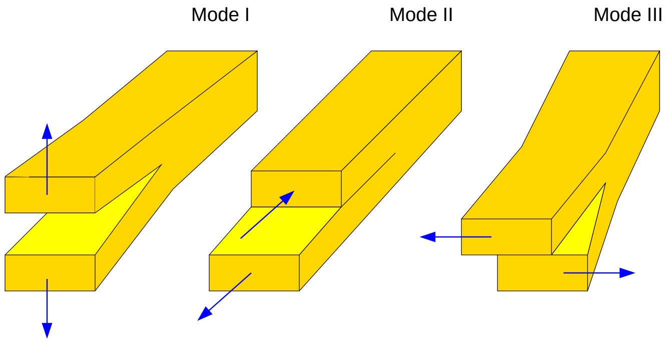


Figure 5.2: The three modes of fracture: Mode I, the opening mode, Mode II, the sliding mode and Mode III, the tearing mode.

This length is given by

$$c_0 = \frac{2E'\gamma}{\pi\sigma_A^2} \tag{5.5}$$

This has very important engineering consequences: for example when inspecting an aeroplane for defects, cracks smaller than c_0 can safely be ignored since they will not grow.

Stress Intensity Factor K

There are three basic modes of crack surface displacement, as shown in Fig. 5.2. In this thesis I shall focus entirely on the tensile opening mode, Mode I, which is by far the most relevant mode when considering the fracture of a highly brittle solid. Irwin [36] used the constitutive equations of linear elasticity to derive solutions for the stress and displacement fields near to the tip of an infinitesimally narrow slit crack. The resulting stress field is of the form

$$\sigma_{ij} = \frac{K_I}{\sqrt{2\pi r}} f_{ij}^{I}(\theta) + \frac{K_{II}}{\sqrt{2\pi r}} f_{ij}^{II}(\theta) + \frac{K_{III}}{\sqrt{2\pi r}} f_{ij}^{III}(\theta) \tag{5.6}$$

where the universal functions $f_{ij}(\theta)$ contain all the angular dependence of the fields. The coordinate system is as shown in Fig. 5.3. K_I, K_{II} and K_{III} are the *stress intensity factors* for each mode: all the details of the applied loading and the specimen geometry depend only these factors.

Hereafter we shall consider only pure Mode I loading, *i.e.* $K_{II} = K_{III} = 0$, and we shall refer to K_I simply as K, the formal definition of which is

$$K = \lim_{r \to 0+} \left[\sqrt{2\pi r}\, \sigma_{yy}(r, \theta = 0) \right] \tag{5.7}$$

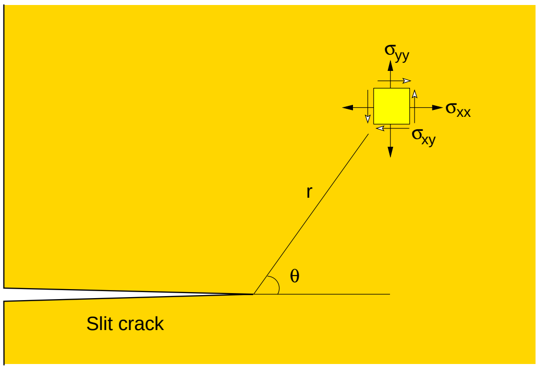


Figure 5.3: Geometry and coordinate system for Irwin infinitesimal slit crack. The Cartesian components of the stress on an element at a position (r, θ) relative to the crack tip are shown.

The Irwin solutions result in a parabolic crack profile, but due to the divergence of the stress and the non-linearity of real materials near to the crack tip, this is not physically realistic. The treatment of the crack as an infinitesimal slit gives rise to a singularity at $r = 0$, and higher order terms need to be added to these solutions to match the outer boundary conditions. Therefore, the Irwin results cannot be applied either very close to or very far from the crack tip: they are consequently referred to as the *near-field* solution. Fig. 5.4 illustrates these solutions for an infinite system. In addition to the K-field, there is also a constant contribution to σ which is unrelated to K. Depending on the geometry this can be as large or larger than the asymptotic K-field. We can therefore write the total Mode I stress field as

$$\sigma_{ij} = \frac{K}{\sqrt{2\pi r}} f_{ij}(\theta) + \sigma_{ij}^{(1)} \tag{5.8}$$

where $\sigma_{ij}^{(1)}$ is the stress term of $O(1)$.

The stress intensity field provides an adequate characterisation of fracture despite the breakdown of linear elasticity near to the crack tip providing that the principle of *small scale yielding* applies. This arises when the potentially very large stresses near the tip are relieved by some inelastic process throughout a region that is small compared to the crack length and specimen dimensions.

The *fracture toughness* of a material, commonly denoted by K_c, is the stress intensity factor at which a crack is observed to propagate in a laboratory sample. Providing the sample is representative of the behaviour of the material outside the laboratory, K_c provides an excellent single parameter characterisation of the fracture

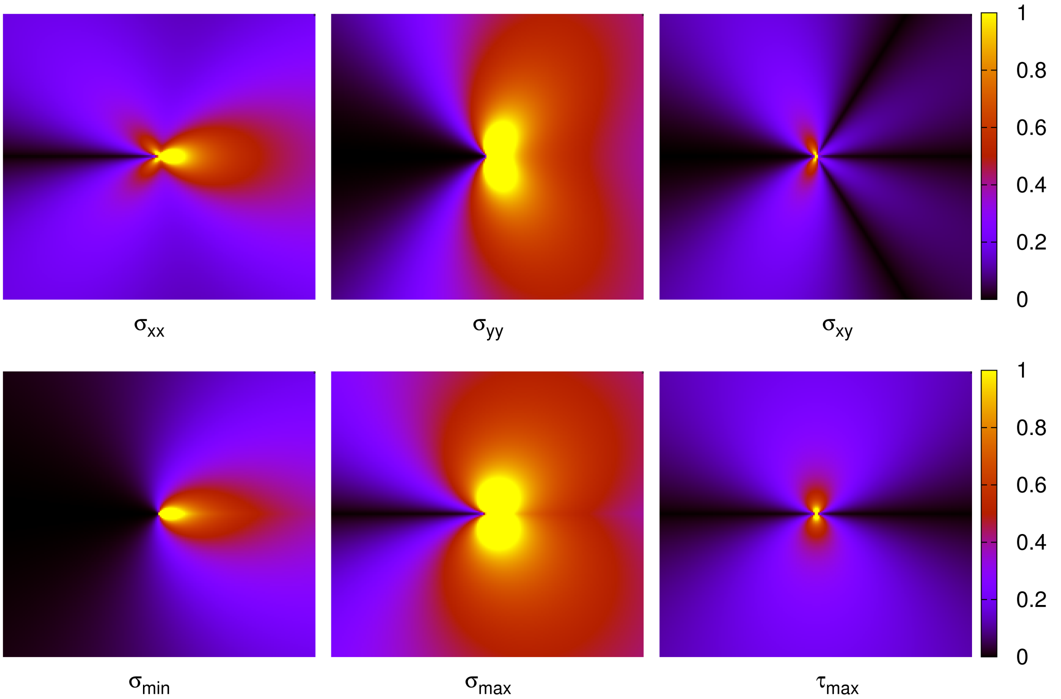


Figure 5.4: Mode I near-field stress solutions for an infinitely narrow crack. The crack slit runs from the middle of the left hand edge to the centre of each plot. The absolute values of σ_{xx}, σ_{yy} and σ_{xy} are shown to the same scale in the top row; σ_{zz} is zero for plane stress and equal to $\nu(\sigma_{xx}+\sigma_{yy})$ for plane strain. The bottom row shows the minimum and maximum principal stresses and the principal shear stress.

process. If the applied load results in a stress intensity factor $K > K_c$ then a crack will propagate: this is the Irwin fracture criterion. The universality of the Irwin solution within the small scale yielding approximation reduces the analysis required when considering a new fracture system to calculating K as a function of geometry dependent parameters.

Energy release rate G

The energy release rate G is the release of elastic energy per unit crack advancement, and is defined by

$$G = -\frac{\partial U_E}{\partial c} \tag{5.9}$$

per unit width of crack front, where c is the crack length. This definition is independent of the way in which the loading is applied. Often G is easier to calculate

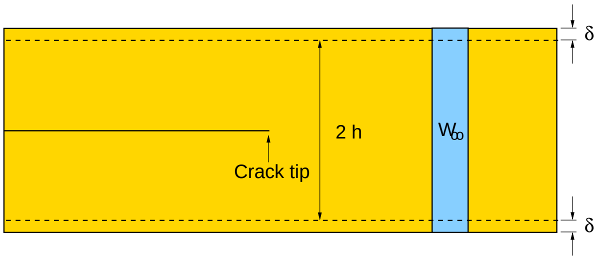


Figure 5.5: Thin strip loading geometry. The edges of the strip are clamped and displaced vertically by an amount δ. The energy density far ahead of the crack tip is W_∞.

than K since energy balance arguments can be made. K and G can be shown to be related by integrating the strain energy density over the interfacial crack area for an infinitesimal extension of the crack. For a Mode I crack the result is

$$G = \frac{K^2}{E'} \tag{5.10}$$

where E' is the effective Young's modulus defined in Eq. 5.3. See Lawn [34, p. 29] for a derivation of this result. Balancing the energy flow to the crack tip with the cost of creating two new fracture surfaces, we can generalise the Griffith criterion of Eq. 5.4 into the form:

$$G_c = 2\gamma \tag{5.11}$$

where γ is the surface energy density of the newly opened surfaces, and G_c is the critical energy release rate. For $G > G_c$ the crack advances, for $G < G_c$ it retracts.

Strip Geometry

Situations for which the stress intensity factor can be computed exactly are rare. A notable exception is the case of steady crack growth in the thin strip geometry illustrated in Fig. 5.5, where the edges of the crack are clamped and displaced vertically. This configuration has been widely used since it is both practical for experiment (see, for example, Ref. 117) and amenable to analysis.

In this geometry, the stress intensity factor does not depend on the crack length, and can be derived entirely by considering the energetics of an advancing crack using an approach first suggested by Rivlin and Thomas [118] in 1952. The horizontal edges of the strip are given a uniform normal displacement δ, so the applied strain is $\epsilon_0 = \delta/h$. Far ahead of the crack, the strip is in uniaxial tension: $\epsilon_{yy} \to \epsilon_0$ as

$x \to \infty$. The stress far ahead of the crack is given by $\sigma_0 = E'\epsilon_0$, and therefore the elastic energy per unit length and per unit thickness far ahead of the crack tip is

$$W_\infty = \frac{1}{2}E'\epsilon_0^2 \cdot 2h = \frac{\delta^2 E'}{h} \tag{5.12}$$

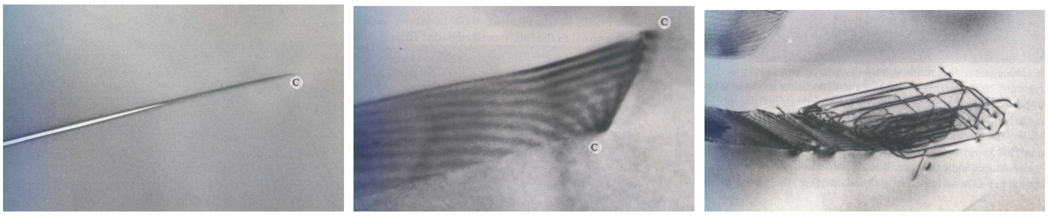
where E' is the effective Young's modulus and ν Poisson's ratio. Far behind the tip, the energy density is zero. Since no energy disappears through the clamped edges, if the crack is to advance by unit distance, a vertical strip of material with energy density W_∞ is effectively replaced by a strip with energy density zero. The energy supplied to the crack tip is therefore equal to W_∞, so the energy release rate is simply

$$G = W_\infty = \frac{\delta^2 E'}{h} \tag{5.13}$$

While this calculation is exact and independent of the crack length, the stress fields that result from substituting this stress intensity factor into the Irwin solutions are only valid in the vicinity of the crack tip. Further away, the constant order stress terms $\sigma_{ij}^{(1)}$ cannot be neglected in comparison with the asymptotic $1/\sqrt{r}$ term. The Irwin K-field stress components all tend to zero far ahead of the crack tip, rather than converging to the externally applied stress as they do in the strip geometry. It is possible to solve this linear elastic problem fully for the case of a semi-infinite crack in an infinite strip, to obtain the stress and strain fields at any point in the strip. Knauss [119] did this in 1966 using the Wiener-Hopf technique to solve the linear elastic equations subject to the appropriate boundary conditions. The resulting expressions have to be integrated numerically to give the full stress field solution: details can be found in Appendix A and will be compared with the results of atomistic calculations in Section 6.5.

5.3 Fracture of Real Materials

Real materials differ markedly from the linear elastic continuum model: they are made of atoms. There are two major problems involved in applying continuum mechanics to atomic crystals. Firstly, the Irwin solutions neglect the effects of the anisotropy of elastic constants and secondly, they ignore the discreteness of the atomic lattice. Neglecting the discreteness of the atomic lattice gives rise to a phenomena known as lattice trapping which is discussed in Section 5.4 below.



(a) (b) (c)

Figure 5.6: Transmission electron micrographs of fracture in silicon (a) Crack formed at 25°C, width of field is 1 μm. (b) The same crack, tilted slightly to reveal the crack front line C—C. The crack is atomically sharp, with no evidence of dislocation activity in the region of the crack tip. (c) A similar crack, formed at $\sim 500^\circ$C, above the brittle-ductile transition. Reproduced from Lawn [34].

Crystal anisotropy

The orientation of a crack in a crystalline material is defined by a plane and a line, written in the form $(ijk)[lmn]$. The plane with Miller indices (ijk) is the surface opened as the crack propagates and the line $[lmn]$ defines the crack front.

The effects of crystal anisotropy on elastic constants can be incorporated in one of two ways. If the orientation of the slab is unknown, then we should determine the Young's modulus and Poisson ratio by averaging the elastic constants over all possible orientations of the crystal axes. This is called the Voigt average, and details can be found in Hirth and Lothe [120, p. 430]. If we do know the crystallographic orientation, then the appropriate values for E and ν can be calculated from the cubic elastics constants C_{11}, C_{12} and C_{44} using the formulæ given by Brantley [56]; this will be the case for the crack systems considered in this work.

Brittle Fracture of Silicon

In this work, I focus entirely on the brittle fracture of silicon. The Griffith energy-balance approach has been shown to provide a good description of the fracture process for straight cracks in silicon in both static and dynamic regimes [31]. In 1986, Lin and Thomson [35] extended the approach to describe dislocation emission and cleavage in crystalline materials and found that the cleavage process is still well described by the Griffith criteria.

An essential requirement for a real material to be described by the Griffith theory is that it undergoes brittle rather than ductile fracture. Brittle cracks are

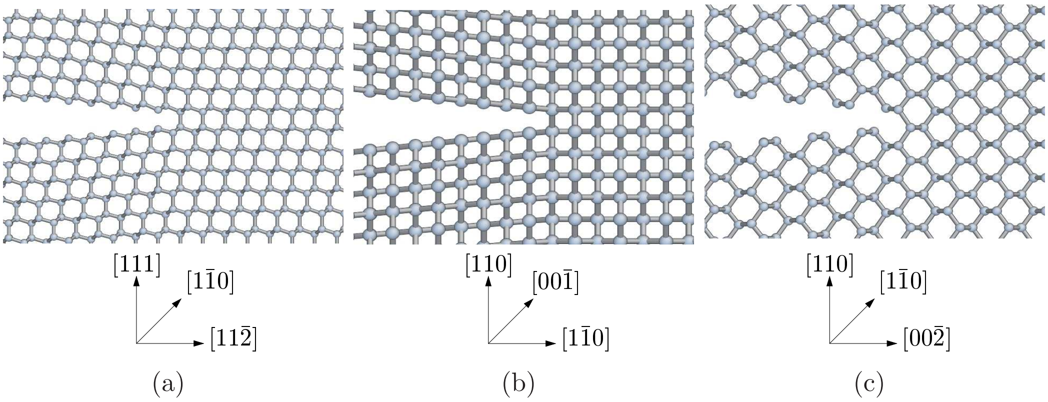


Figure 5.7: Three seed cracks in silicon with different crystallographic orientations (a) (111)[1$\bar{1}$0], easy cleavage plane (b) (110)[00$\bar{1}$], hard cleavage plane, easy direction (c) (110)[1$\bar{1}$0], hard cleavage plane, hard direction

atomically sharp and propagate by breaking bonds one at a time at each point along the moving crack front. Figs. 5.6a and 5.6b show transmission electron microscope (TEM) images of a crack in silicon at room temperature: it can be directly seen that the failure is brittle. Silicon undergoes a brittle-ductile transition (BDT) at around 800 K [121], above which fracture progresses by the emission of dislocations as shown in Fig. 5.6c. In this work I consider only brittle behaviour, so I restrict all studies to temperatures well below the transition temperature.

Experiments with straight cracks on carefully prepared single crystal silicon have shown good agreement with the predictions of the continuum approach. For a review of the experimental techniques used to study brittle fracture, see Ref. 122. TEM results indicate that the fracture surfaces remain atomically smooth for loadings several times larger than G_c [117].

The three most studied crack systems in silicon crystals are (111)[1$\bar{1}$0], (110)[1$\bar{1}$0] and (110)[00$\bar{1}$]; these orientations are illustrated in Fig. 5.7. The (111) plane is the easy cleavage plane since it has the lowest surface energy density, making (111)[1$\bar{1}$0] the most commonly occurring crack system in practical applications [123]. In this work I shall exclusively consider fracture in this orientation, hereafter denoted simply as Si(111).

Surface energy and Griffith critical load

The Griffith critical loading G_c is equal to twice the surface energy density. Computing G_c therefore requires knowledge of the details of the surface reconstruction

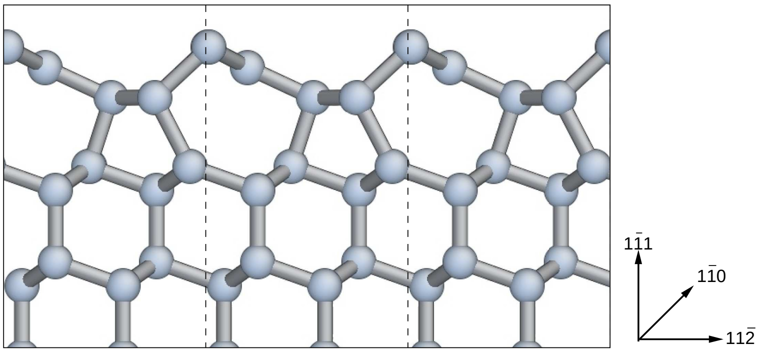


Figure 5.8: Pandey 2×1 reconstruction of the Si(111) surface. Atomic coordinates are those obtained from a geometry optimisation using CASTEP. Three unit cells are shown along the $[11\bar{2}]$ direction, separated by the dashed lines. The reconstruction is also periodic in the $[1\bar{1}0]$ direction (into the paper).

as this will determine the surface energy. For open Si(111) surfaces, this is generally accepted to be the 2×1 Pandey π-bonded chain [124], which consists of alternate 5– and 7– membered rings along the $[11\bar{2}]$ direction as shown in Fig. 5.8. This reconstruction is consistent with a variety of experimental surface science techniques, but *ab initio* calculations show that the 7×7 surface is 60 meV per atom lower in energy [125, 126]. However the annealing process by which the extremely complex 7×7 surface forms takes thousands of seconds at room temperature [127]. The barrier to the formation of the Pandey surface from the unreconstructed surface is less than 0.03 eV/atom, so it is expected to be the equilibrium surface configuration except at very low temperatures [128] and certainly after crack propagation.

The energy density of this surface within a particular computational model can be computed in a straightforward fashion: a silicon slab, periodic in x and z with open surfaces at the top and bottom is initialised with the approximate atomic coordinates for the reconstruction, then geometry optimised to find the local energy minimum. The surface energy can then be obtained by comparing this energy with that of a similar fully periodic system.

Surface energies computed using a variety of methods are compared with experimental measurements in Table 5.1. *Ab initio* calculations show good agreement with experiment, in contrast to the tight binding results. In particular, it is to be noted that different TB models give very different values for the surface energy; this is a consequence of the semi-empirical nature of the method. The SW potential agrees fairly well with both *ab initio* and experiment. Experimental surface energies formed part of the dataset to which its parameters were fit, so this should not

Description	Reference	γ / J/m^2	G_c / J/m^2
Literature results			
Direct experiment	[129]	1.24	2.48
Estimated from sublimation energy	[130]	1.46	2.92
Ab initio (LDA)	[37]	1.34	2.68
Tight binding (NRL)	[108]	1.0	2.0
This work			
Stillinger-Weber		1.35	2.70
Tight binding (Kwon)		1.89	3.78
Ab initio (GGA, CASTEP)		1.44	2.88

Table 5.1: Surface energy density and corresponding Griffith critical energy release rates for Si(111). G_c has been obtained using $G_c = 2\gamma$ in all cases. The values for the SW potential and both tight binding models are for the unreconstructed surface since those models do not predict spontaneous reconstruction after fracture, while the *ab initio* result of Spence *et al.* and the CASTEP calculation performed for this work are for the Pandey 2×1 reconstruction.

be taken as an indication that the potential has superior predictive power to tight binding in general.

Velocity Gap and High Speed Instabilities

Above the critical load G_c, dynamic linear elasticity theory predicts that the velocity v of a running crack should increase with increasing energy flow to the crack tip, according to

$$\frac{v}{c_R} = 1 - \frac{G_c}{G} \tag{5.14}$$

For a derivation of this result see Freund [31]. c_R is the theoretical maximum crack propagation speed, the Rayleigh wave speed, equal to the velocity of acoustic surface waves. For Si(111) the Rayleigh speed is $c_R = 4.68$ km/s [131]. Thus all speeds $0 < v < c_R$ are predicted to be accessible by the continuum model. However, in recent years doubts have been raised about both limits of this inequality.

Holland and Marder [113] observed a velocity band gap at the onset of fracture, with the velocity of the crack tip rising almost instantaneously from zero to about 2 km s^{-1}, half of its terminal velocity. This effect arises in a lattice model proposed by Marder and Gross [132] in 1995, in which they argue that below a certain speed the crack does not have sufficient kinetic energy to jump over the energy barrier separating one lattice site from the next. The observation of a velocity gap is disputed by Be'ery *et al.* [133] who have used three-point bending experiments to produce cracks moving at less than 1% of the Rayleigh speed, with no evidence of a velocity gap. Empirical potential studies by Gumbsch *et al.* [134] in 1997 and Hauch *et al.* [117] in 1999 provide evidence in favour of the velocity gap, but the applicability of such methods has been called into question by their poor description of the onset of fracture: see Section 5.4 below.

Recent experiments indicate that speeds as high as c_R are unachievable. The first observation of a high speed instability was in the amorphous material PMMA, where Fineberg *et al.* [135] discovered an instability at around $0.6c_R$, above which the crack breaks up into a large number of 'microbranches'. More recently, a similar instability has been observed in crystalline silicon, at speeds of between $0.7c_R$ and $0.9c_R$ [136]. At intermediate fracture energies a faceted crack front arises to dissipate the additional energy that flows to the crack tip. At these loads, sub-micron surface features can be seen using AFM (atomic force microscope) imaging techniques [117].

5.4 Lattice Trapping

Experimentally, the onset of fracture on the (111) plane in silicon is found to be of the order of 10–20% above G_c [117, 136]. This discrepancy is explained by the phenomena of *lattice trapping*: the atomic nature of real materials gives rise to a periodic potential at the crack tip that manifests itself in a higher resistance to fracture than that predicted by Griffith's criteria. According to the continuum description, there is a unique critical stress intensity K_c for which the crack is in stationary equilibrium. For loads above this the crack advances, and below it retracts. When the discrete nature of the lattice is taken into account, the crack becomes trapped for a range of stress intensity factors $K_- < K_c < K_+$, as illustrated in Fig. 5.9. Lattice trapping of cracks is analogous to the trapping of dislocations by the Peierls barrier [120]. It can lead to slow crack growth [137], and anisotropy with respect to propagation along different crystallographic orientations [138]. The range of lattice trapping can be quantified using the ratio

$$R = \frac{K_+}{K_c} = \sqrt{\frac{G_+}{G_c}} \tag{5.15}$$

R can be measured quasi-statically by initialising a test system close to K_c and increasing the load in small steps, fully relaxing the system after each increase. The

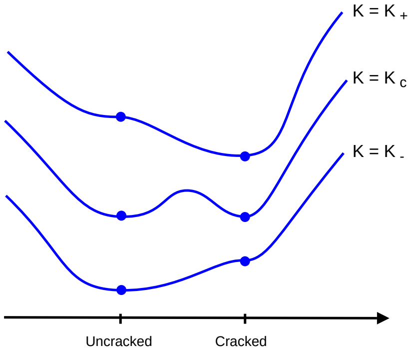


Figure 5.9: Schematic illustration of lattice trapping, showing total elastic energy as a function of crack tip displacement for various stress levels. There are minima at the uncracked and cracked states. For $K < K_-$ the uncracked state is globally stable, so the crack closes. For $K > K_+$ the cracked state is globally stable and the crack opens. At K_c the energies are equal, but separated by a finite barrier.

barrier to fracture goes to zero at K_+. In an atomistic simulation, this is typically defined as the load at which bond rearrangements at the crack tip first occur.

Even very simple atomistic models give rise to lattice trapping. The first use of the term lattice trapping dates back to 1971, when Thomson *et al.* [139] formulated two analytically solvable lattice models of fracture, both of which predict large lattice trapping, of the same order as the Griffith load. This result called into question the whole idea of a single critical stress. In an extensive study Sinclair [140] investigated the effects of lattice trapping in silicon for a large range of interatomic potentials. Flexible boundary conditions designed to match the continuum elasticity solutions were used. It was found that short-ranged force laws led to much greater lattice trapping than those with long attractive tails. Later, Curtin [141] developed an analytical connection between the range of lattice trapping and the range of the interatomic potential. With simple classical potentials, he found the effect of lattice trapping to be small, but noted that it decreased non-monotonically with the range of the force law. In the first example of a hybrid *ab initio* and classical simulation of fracture, Spence *et al.* [37] found significant lattice trapping. The approach used for this simulation has been described in Section 3.5. They concluded that the barrier to crack advance was sufficiently large to prevent thermal activation of fracture in silicon.

An *ab initio* quasi-static simulation carried out by Perez and Gumbsch [123] in 2000 explained the directional anisotropy in silicon fracture on the (110) plane by a difference in the extent of lattice trapping in the two different cleavage directions.

Description	Reference	$R = K_+/K_c$
Experimental		
(111) cleavage	[117]	∼1.1
(110) cleavage	[136]	∼1.2
Ab initio		
(111) cleavage	[123]	1.25
(110) cleavage	[123]	1.25–1.35
(111) cleavage	[37]	1.31
Empirical potentials		
Various short ranged potentials	[140]	1.13–7
EDIP	[108]	1.65–1.73
Brittle SW	[113]	1.53
MEAM	[142]	1.20
Stillinger-Weber (extrapolated)	[143]	1.51
Stillinger-Weber	(this work)	1.50
Hybrid Simulation		
DCET	[108]	1.09–1.16
ReaxFF	[131]	1.48

Table 5.2: The extent of lattice trapping measured in a variety of experimental and theoretical studies. The *ab initio* and empirical potential results are obtained quasi-statically.

They found that the $[1\bar{1}0]$ direction was preferred, with the $[00\bar{1}]$ direction being associated with larger lattice trapping. On the (111) plane, they found significant lattice trapping, in good agreement with Spence *et al.* Table 5.2 summaries various measurements of the lattice trapping range.

5.5 Empirical Potential Molecular Dynamics

Classical interatomic potentials tend to dramatically overestimate the extent of lattice trapping in silicon, resulting in crack arrest until the load is well above G_c. When fracture finally occurs, the excess elastic energy is typically dissipated by the creation of rough fracture surfaces, bluntening of the crack tip and even repeated crack arrest, requiring the load to be further increased: these are all characteristics of ductile fracture. The Stillinger-Weber, EDIP and Tersoff potentials all predict ductile fracture [108, 113, 144]; Fig. 5.10 shows a snapshot from a classical MD simulation carried out with the SW potential. The failure of empirical potentials to describe the fracture of silicon is in contrast to metallic systems, where embed-

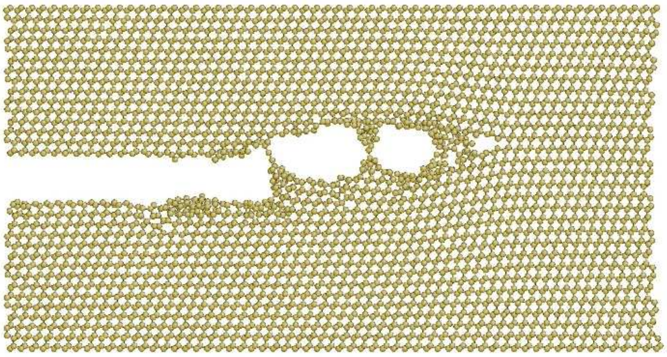


Figure 5.10: Snapshot from an MD simulation of fracture using the Stillinger-Weber potential. Size of system is $200 \times 100 \times 7.68$ Å^3. The description of fracture is ductile: note the rough surfaces and blunt tip.

ded atom potentials often provide a good description of fracture and deformation processes [145].

It is possible to create empirical potentials for silicon which do exhibit brittle fracture. Holland and Marder [113] used a modified version of the SW potential, with the λ angle parameter doubled. This change has a marked effect on the properties of the model, increasing the melting point from 1683 K to 3500 K, so the modified potential cannot be regarded as a satisfactory model of silicon [146]. The resulting lattice trapping is as large as that of the unmodified potential (see Table 5.2). A measurement of the lattice trapping range of the SW potential by extrapolation of the barrier heights at lower loads gives 1.51 [143], in good agreement with the value computed quasi-statically in this work. Swadener *et al.* [142] have conducted a simulation using the Modified Embedded Atom Method (MEAM) potential, which does exhibit brittle fracture. Their simulation also gave fracture above the Griffith criteria, suggesting significant lattice trapping.

At this point it is appropriate to consider why empirical interatomic potentials which accurately describe many of the physical properties of silicon do such a bad job of modelling fracture: in many cases far worse even than the simple Griffith energy-balance argument. We have seen that near to the crack tip, the continuum stress field diverges. Bonds in this region will therefore be subject to large, anharmonic displacements, and eventually bond rupture will take place. Classical potentials cannot hope to model such complex chemical processes accurately; such processes can only be reliably described by quantum mechanics.

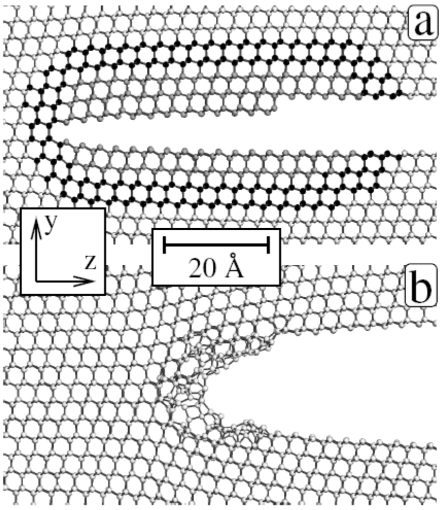


Figure 5.11: Snapshot 2 ps into fracture simulations using: (a) the DCET hybrid method, and (b) the Stillinger-Weber empirical potential. Reproduced from Bernstein and Hess [108].

5.6 Hybrid Simulation of Fracture

A fully quantum mechanical simulation of fracture is unfortunately not feasible: as has been discussed extensively in Chapter 3, fracture is an example of a strongly coupled multiscale system, since it is the long range stress field caused by the opening crack which breaks the bonds that advance it. Tens of thousands of atoms contribute significantly to the elastic relaxation of this stress field, and it is currently impossible to perform a fully QM simulation for such a large system. *Ab initio* simulations such as those of Perez and Gumbsch [123] make significant approximations of the near crack tip displacement field, since the outer boundaries, set by the Irwin continuum field, are not sufficiently far away from the process zone where rebonding is taking place to reach the asymptotic limit. The best we can do is to construct a hybrid model of the fracture process, with a quantum mechanical model centred on the crack tip to accurately describe the bond stretching and breaking processes coupled to a much larger classical region which captures the long range stress field.

Reactive potentials such as ReaxFF go some way to addressing the failings of empirical potentials. In 2006, Buehler *et al.* [147] reported a force-mixing hybrid simulation using thousands of ReaxFF atoms near the crack tip and the Tersoff potential elsewhere which gives brittle fracture, but they do not attempt to measure the critical load, concentrating instead on high speed phenomena such as crack branching. In a very recent paper, Buehler *et al.* [131] report large lattice trapping ($R = 1.48$) using the same approach, which they attribute to a crack tip reconstruction. I shall postpone a detailed discussion of the rôle of crack tip reconstructions to Chapter 7, remarking here only that there are some problems with the approach taken by Buehler *et al.*

The best hybrid fracture results to date were obtained by Bernstein and Hess [108], who used a force-mixing approach to combine a classical empirical potential with a semi-empirical tight binding model. This approach, known as DCET, yields brittle fracture of silicon on the (111) plane (see Fig. 5.11). They report R in the range 1.09–1.16, and suggest that two length scales are important in determining the lattice trapping barrier: one is set by bond breaking and the other by elastic relaxation. In the DCET hybrid simulation, the bond breaking length scale is larger and the elastic relaxation distance smaller than empirical potential simulations, resulting in reduced lattice trapping.

5.7 Summary

In this chapter I have reviewed a range of approaches to understanding the fracture of brittle materials, from the continuum theories of Griffith and Irwin to modern computational simulations using a variety of techniques. I have considered the crucial rôle of lattice trapping in determining the critical load for a particular system.

In recent years it has become clear that the extent of lattice trapping is very sensitive to the detailed nature of atomic interaction, and that it may be very different for different cleavage planes. The latest experimental and theoretical results seem to indicate a preference for a small but significant amount of lattice trapping. However, given the sensitive dependence on the choice and range of interatomic interaction and the limited precision and large variance in the experimental measurements, it is impossible to be sure of this.

The essential multiscale nature of brittle fracture has been elucidated: only by modelling a large enough system to include the effects of the long stress range field can we hope to accurately describe the fracture process. I have also explained why a quantum mechanical description of the bond breaking processes taking place at the tip is necessary. In the next chapter, I will describe how the LOTF hybrid method introduced in Chapter 4 can be applied to model the brittle fracture of silicon.

Chapter 6

Simulation Approach and Methodology

6.1 Introduction

We are now in a position to draw together the tools and methodologies described in the preceding chapters and apply them to a real problem. The scientific problem of interest is the brittle fracture of Si(111) outlined in Chapter 5, and the central method is the 'Learn on the Fly' scheme described in Chapter 4. In this chapter I describe the application of LOTF to this problem, and include details of preliminary tests and validation work; the central results of this thesis will be presented in the following chapter.

We have seen that a hybrid approach is needed to model brittle fracture accurately. LOTF is particularly well suited to this task since the adjustable potential and buffer region allow the quantum region to be small and mobile, so we can follow the crack tip as it moves. This makes it feasible to use full *ab initio* calculations in the QM region for the first time, in contrast to force-mixing schemes such as DCET [108] where a large QM region is required for stability and thus fully first principles calculations are too costly to incorporate in the scheme. It is necessary to describe the processes taking place at the crack tip at an accurate QM level of detail: for example the semi-empirical tight binding potential used in DCET does not give the experimentally observed Pandey surface reconstruction.

6.2 The CRACK code

During the course of this work, I have developed a general purpose code for conducting fracture simulations with the LOTF scheme, appropriately named CRACK. This program was originally written in C with the C-LOTF code, but has been converted to Fortran 95 in parallel with the development of the LOTF95 code. As LOTF95 and CRACK matured together, the fracture application provided the first large-scale test of the new implementation of LOTF.

The essential simulation methodology is straightforward. A crack system is prepared in the thin strip geometry described in Section 5.2, periodic in the crack-front direction and with vacuum surrounding the in-plane edges. Since the cell thickness in the periodic direction is fixed, plane strain conditions apply. We begin by relaxing the system, then the load is increased in a number of small steps until the crack starts to propagate.

The CPU time required for a hybrid fracture simulation with *ab initio* QM calculations is extremely large: for around 100 QM atoms the CASTEP force computation takes around ten minutes on 32 nodes of the Cambridge supercomputer,[1] and this must be done at least every 10 fs during the molecular dynamics. This means that simulations take of the order of 48 hours per picosecond of MD, so it is essential that the code be fully restartable. This is achieved by writing a checkpoint file containing all the dynamical and state dependent variables every few time steps.

The following sections provide further details regarding important aspects of the implementation of the CRACK code.

Hybrid Structural Relaxation

Classical geometry optimisation can be performed rapidly using the conjugate gradients algorithm, and this provides a good first approximation to the hybrid relaxed geometry, which in turn approximates the atomic configuration that would be found by a fully quantum mechanical optimisation of the entire system. Relaxation using the hybrid forces is slightly more involved. Whenever forces are needed, a QM calculation is performed and the adjustable potential parameters are optimised to reproduce the QM forces. The forces used for the geometry optimisation are the sum of the classical and adjustable potential forces: as for MD, this ensures that there is no mechanical incompatibility at the boundary between classical and quantum regions.

---

[1] The Cambridge high performance computer *Darwin* comprises 2340 × 3.0 GHz Intel Woodcrest cores with 8 GB of memory per node. The cluster has a peak Linpack Performance of 27 Tflop/s and a sustained value of 18.27 Tflop/s. The Infiniband interconnect provides 900 MB/s bandwidth with a 1.9 μs latency.

The standard formulation of the conjugate gradients algorithm requires both an objective function and its derivative. For geometry optimisation, the former is the total energy. In the LOTF hybrid scheme, there is no well defined total energy, so this approach is not possible. There are several alternative ways to proceed. Damped molecular dynamics methods only require forces but are generally slower than direct minimisation approaches unless sophisticated coupling schemes are used as discussed in Section 2.2. The FIRE method, which is based on damped MD, also does not require a total energy function and so is a viable approach to hybrid structural relaxation. It is also possible to modify the line minimisation step of the conjugate gradient algorithm to work with derivative information only. This is done by extrapolating the projection of the derivative along the search direction to zero. To avoid errors associated with large extrapolation lengths, a maximum step size is specified and then the procedure is iterated until the projected derivative is zero.

All three of these methods have been implemented within the LOTF95 framework. Conjugate gradients with derivative extrapolation gives the best combination of speed and robustness, usually converging faster than FIRE for the fracture systems considered in this work, where there are initially very large forces near the tip as well as a very long range stress field to be relaxed.

Displacement Loading Scheme

The external load on the model crack must be applied in a way that is equivalent to the 'fixed grips' displacement boundary conditions described in Section 5.2. To exactly mimic experiment, the loading would be applied simply by displacing the top and bottom rows of atoms, then fixing these atoms and allowing the system to relax. However the time between loadings in experiments is of the order of several minutes which is completely inaccessible to simulation. We therefore need to perform the loading more smoothly to reduce the time required for equilibration. We also need to insert a seed crack into the system, so that the crack tip is far from the edges of the simulated system and the thin strip approximation of Eq. 5.13 applies.

Both of these requirements can be met by considering the equilibrium stress distribution in a slab containing a stationary crack, then choosing simple initial conditions that approximate this configuration: far behind the tip the strain is zero, and far ahead it approaches the applied loading. Fig. 6.1 illustrates these conditions, and shows the changes that take place in a classical geometry optimisation as a result of this stress distribution. The optimisation is quick since the initial conditions are

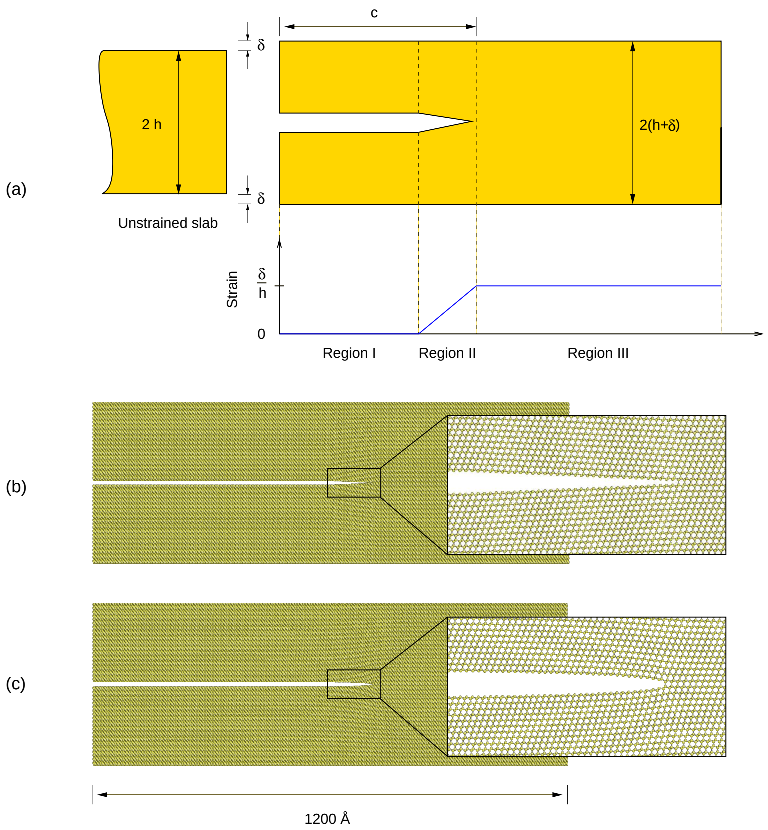


Figure 6.1: Displacement loading scheme. (a) An initial load is applied to an unstrained slab. The atoms in Region I are vertically shifted by δ since there is zero strain behind the tip of a relaxed crack. In Region III the strain is constant and equal to the far field value of δ/h. The seed crack length is c. Across Region II the strain increases linearly. The lower panels show the atomistic configuration of a $1200 \times 400 \times 3.68$ Å^3 slab containing 90720 atoms with a seed crack of length 600 Å (b) before and (c) after classical relaxation, with the insets showing Region II in more detail.

a reasonable approximation to the relaxed strain field. Prior to straining the slab, care has to be taken to align the vertical coordinate origin with the centre of an Si—Si bond to ensure that the crack opens cleanly. In a slight variation of the fixed grip boundary conditions, the top and bottom rows of atoms are not rigidly fixed but are constrained to move in planes of constant y using the RATTLE algorithm [148]. This prevents the free vertical surfaces at each end of the slab from bowing in at their centres due to effects associated with the non-zero Poisson ratio. An alternative approach not pursued in this work is to use the Irwin displacement field to apply the slab boundary conditions. In contrast to the simplicity of the thin strip geometry, in this case the energy release rate is no longer constant for a steadily propagating crack since it depends on the length of the crack.

Once the initial configuration of the model system has been generated, the load is incremented in one of two ways. The first approach, used in my early simulations, is to simply rescale the entire slab vertically by a small amount:

$$\mathbf{r}_i' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1+\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{r}_i \tag{6.1}$$

Providing the strain rate α is small this will not deform the system too far from its relaxed configuration. The amount of equilibration required after each loading can be reduced by instead performing classical relaxations at strains ϵ and $\epsilon + \alpha$ then computing the displacement field as

$$\mathbf{u}_i = \mathbf{r}_i^{(\epsilon+\alpha)} - \mathbf{r}_i^{(\epsilon)} \tag{6.2}$$

where $\mathbf{r}_i^{(\epsilon)}$ denotes the relaxed position of atom i under an external strain of ϵ. Eq. 6.1 is then replaced by

$$\mathbf{r}_i' = \mathbf{r}_i + \frac{\beta}{\alpha}\mathbf{u}_i \tag{6.3}$$

where β is the desired load increment. The $\{\mathbf{u}_i\}$ only need to be evaluated once for a given crack configuration.

To prevent accumulation of numerical error when combining small strains, the energy release rate G is directly calculated using Eq. 5.13 by measuring the total displacement δ relative to the unstrained slab height. It should be stressed that these two loading methods are equivalent after elastic relaxation has taken place; the second scheme merely decreases the amount of time required to perform this relaxation.

Molecular Dynamics Methodology

In this section I describe details of the fracture molecular dynamics simulations carried out in this work. We start with a relaxed configuration at a load G below the critical Griffith load G_c. Depending on the range of lattice trapping, it should be possible to find a load that satisfies $G_- < G < G_c$ so the relaxed crack is lattice trapped and will not move until the load is increased above G_+.

Molecular dynamics is then carried out at a temperature of 300 K, with a weak

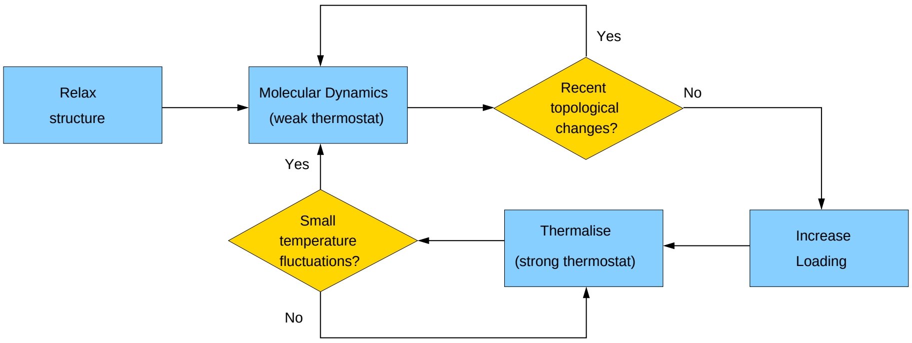


Figure 6.2: Flow chart illustrating molecular dynamics methodology used for the fracture simulations. See text for a description of each state and the conditions that have to met for transitions to take place.

Langevin thermostat to correct for the energy drift caused by the time dependence of the LOTF Hamiltonian. The dynamics are accelerated using the predictor-corrector scheme outlined in Section 4.5. Extrapolation for 10 steps of $\Delta t = 1$ fs is possible for pure silicon systems. In some simulations the Si—Si bonds cut by the insertion of the seed crack were terminated with hydrogen atoms to prevent the crack closing up at low loads: the lower atomic mass of hydrogen requires the MD timestep to be decreased to $\Delta t = 0.1$ fs; however, tests show the number of extrapolation steps can be increased to 20 in this case without the maximum force errors during interpolation becoming larger than 0.1 eV/Å.

Unfortunately it is not possible to run dynamics for long enough to fully explore the environment at each load and to cross barriers which the real system would have sufficient thermal energy to pass over at 300 K. Instead we allow the dynamics to proceed for some fixed amount of time T_{MD}, and then periodically check for rebonding near the crack tip using the time-averaged coordinates (calculated using Eq. 4.10 with an averaging time of $\tau = 50$ fs). We approach realistic timescales from below, with hybrid *ab initio* simulations carried out with $T_{\mathrm{MD}} = 0.2$ ps, 1 ps and 5 ps — the results observed in each case are discussed in the following chapter.

Fig. 6.2 schematically illustrates the details of the molecular dynamics methodology used for the fracture simulations. If no rebonding occurs for some time then we increase the load using either Eq. 6.1 or Eq. 6.3 as described above. After each loading increment there is a thermalisation period in which a stronger thermostat is used to dissipate the energy produced by the rescaling. The thermalisation continues until the fluctuations in temperature are small, defined by the inequality

$$\frac{T - \langle T \rangle}{T} < \frac{1}{\sqrt{N}} \tag{6.4}$$

where T and $\langle T \rangle$ are the instantaneous and average temperatures and N is the total number of atoms in the simulation. Once this condition is satisfied the thermostat is turned down and we return to almost microcanonical molecular dynamics. After the crack has started to move, the rebonding is automatically detected and the load is not increased further.

QM Selection Algorithm

A major advantage of LOTF is that it allows us to keep the QM region small. This requires a robust selection algorithm to follow the crack tip as it moves and identify the atoms that need to be treated with quantum mechanical accuracy. As discussed in Section 4.6, this is a difficult problem since the timescales of thermal vibration and crack motion are not well separated. The hysteretic selection algorithm described in that section provides an effective solution to this problem; indeed this is the very purpose for which the algorithm was developed.

We identify atoms as active when they change their bonding topology, and then construct embedding ellipses around each active atom. The set of active atoms is seeded with a few atoms near to the crack tip at the start of the simulation. Ellipses are used rather than the spheres of Section 4.6 to allow the embedding region to be biased forwards so that the QM region always extends ahead of the crack tip. Fig. 6.3 illustrates how the algorithm works in a simple case with only two active atoms — in reality there could be several hundred. Ellipses with different radii are used to define inner and outer selection regions, and then the hysteretic algorithm ensures that atoms near the edges of the QM region do not oscillate in and out of the active region.

As the crack moves on, we can stop treating atoms behind the crack tip quantum mechanically. We cap the size of the QM region at some limit N_{quantum} based on our computational capability — this can be several hundred atoms for a tight binding simulation, or of the order of a hundred for an *ab initio* simulation. By keeping track of the order in which atoms became active, we can remove them from the QM region in a consistent fashion. An additional condition prevents atoms further than

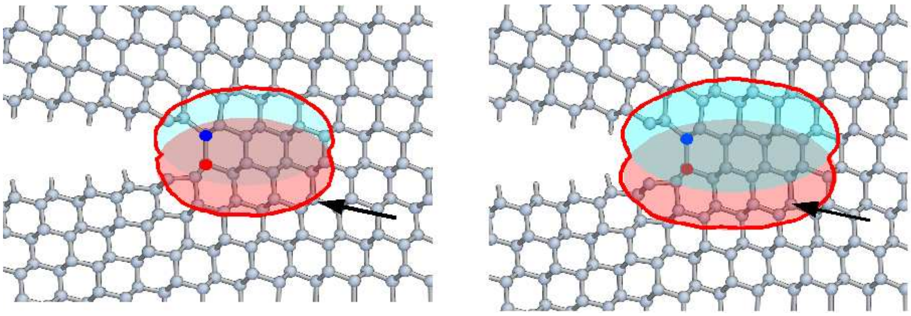


Figure 6.3: Hysteretic QM selection algorithm applied to crack tip region. The red and blue atoms are considered 'active', and are used to define inner (left panel) and outer (right panel) selection regions. The atom indicated with the black arrow remains selected despite oscillating in and out of the inner region providing that it stays inside the outer region.

a threshold distance away from the centre of mass of the current QM region from becoming active.

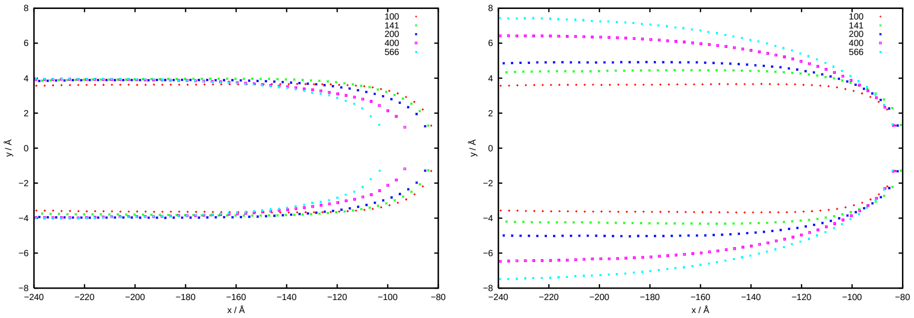
6.3 Classical Fracture Calculations

Comparing molecular dynamics simulations with experiment is very difficult since the time and distance scales accessible to the two techniques are separated by many orders of magnitude. Ideally we would converge the observables that are the outcome of the simulation with system size and with simulation time. For example, the critical load for fracture measured in a simulation should approach that found in experiment. Unfortunately, finite computational resources mean this is not always practical.

In this section I present some preliminary studies used to test the CRACK code and determine reasonable simulation parameters. The large-scale elastic and thermodynamic properties that affect the overall scaling of the outcome of a fracture simulation with system size should be adequately characterised by the classical interatomic potential. Hybrid simulations are too computationally expensive to carry out this convergence testing.

Crack Tip Curvature

First, I shall describe tests to verify that the energy release rate G provides a size invariant measurement of the applied loading suitable for direct comparison between



(a)

(b)

Figure 6.4: Classically optimised crack profile as a function of slab height when (a) the displacement δ is kept constant, and (b) the energy release rate G is fixed at G_c. Each colour is a different slab height, measured in Å.

experiment and simulation. The force required to break the critical crack tip bond must be constant, independent of the specimen size, and therefore so is the stress concentration as quantified by the stress intensity factor K. Since G is directly related to K by Eq. 5.10 the stress concentration is the same for different size systems with the same energy release rate.

Figure 6.4 shows crack profiles extracted from classically optimised slabs of varying heights containing stationary cracks, using either a constant displacement δ or constant energy release rate equal to the Griffith critical value of G_c. Returning to the elliptical crack analysed by Inglis and illustrated in Fig. 5.1, we recall that stress concentration is directly related to the curvature of the crack profile near to the tip. The figure clearly shows a good match in the curvature around the crack tip over a wide range of slab heights providing that G is kept constant. If instead the displacement δ is fixed as the slab height varies, then the resulting stress intensity factors and curvatures are very different. The large range of heights means that for the larger slabs the constant δ condition results in loads with $G < G_-$ so the crack starts to close up on minimisation.

Classical Fracture Toughness

As well as MD, the CRACK code can perform quasi-static simulations where the system is fully relaxed after each increment of load. This approach can be used to estimate the fracture toughness G_+, the load at which the lattice trapping barrier first falls to zero. In the thin strip regime the energy balance argument given in

Force tolerance / eV/Å	Fracture toughness / J/m^2
0.001	15.6
0.01	15.7
0.1	18.0
0.5	23.1
1.0	31.1
2.0	54.5

Table 6.1: Classical fracture toughness G_{100} as a function of the force tolerance. System geometry is $1200 \times 400 \times 3.68$ Å^3 with a seed crack extending half way through the slab.

Section 5.2 indicates that G should be independent of the crack length and slab height. We can use this criterion to determine a suitable simulation size that adequately captures the effects of the long range stress field by verifying that G_+ does not depend on geometric parameters.

Since the Stillinger-Weber classical potential used here hugely overestimates the lattice trapping range and does not predict a brittle mode of fracture, the value G_+ will be much larger than the experimentally observed critical load. For the purpose of measuring the lattice trapping range R of an interatomic potential, G_+ is usually defined as the load at which rebonding first occurs at the crack tip. Conducting quasi-static classical simulations with the SW potential and this definition leads to $G_+ \sim 6$ J/m^2 and $R \sim 1.5$ as reported in Table 5.2. However, since the description of fracture provided by the potential is ductile, the load has to be increased considerably further before the crack tip advances significantly. Since the classical potential does not provide an accurate description of the microscopic processes taking place at the crack tip, we shall make an alternative definition of the fracture toughness purely for the purposes of testing size convergence: define G_{100} to be the load required for the crack tip to advance by a distance of 100 Å from its initial position. This definition is adequate for establishing suitable parameters for the system size and simulation time.

A series of quasi-static simulations was performed with a fixed size slab to identify the level of relaxation required to achieve a converged value of G_{100}. The results shown in Table 6.1 indicate that an RMS force tolerance of 0.01 eV/Å is sufficient. Inspection of the atomic trajectories obtained confirms that G_{100} is a reasonable definition, since the crack goes from being heavily lattice trapped to relatively free

Seed crack	Slab height		
fraction	100 Å	200 Å	400 Å
0.2	14.6	15.0	17.9
0.3	14.6	15.0	15.6
0.4	13.5	15.0	15.7
0.5	13.5	15.0	15.7

Table 6.2: Classical fracture toughness G_{100} in units of J/m^2 for various geometries. In each case the slab width is three times its height and the quasi-static force tolerance is 0.01 eV/Å.

(although still ductile) propagation at this loading.

Table 6.2 shows the values of G_{100} for a variety of slab geometries with different slab heights and seed crack lengths. These results indicate that G_{100} is constant providing there is of the order of 200 Å of material surrounding the crack tip on all sides. This provides an independent estimate of the size of the region that contributes significantly to the elastic relaxation of the stress field surrounding the crack tip. A region of this size contains of the order of 20 000 atoms, confirming that a direct fully quantum mechanical simulation is significantly beyond the capability of even the latest linear scaling *ab initio* codes. Visualising the atomistic stress fields directly provides an alternative technique to arrive at this figure and will be discussed in Section 6.4 below.

Classical molecular dynamics simulations at a temperature of 300 K should give a lower estimate for G_{100} since it is possible for the system to thermally excite its way over the lattice trapping barrier, providing that we allow sufficient time before increasing the load. Table 6.3 shows that the fracture toughness measured in this way is highly sensitive to the minimum time T_{MD} between successive loadings. A time of $T_{\mathrm{MD}} = 5$ ps gives values of G_{100} consistent with those obtained quasi-statically above. This time is equivalent to the sound travel time from crack tip to the edge of the system which sets the approximate time scale for elastic relaxation.

6.4 Linear Elastic Stress Distribution

An excellent way to verify that the simulation cell is large enough to provide an accurate description of fracture is to calculate the stress distribution of the relaxed system and from this estimate the size of the region that makes a significant contribution to this stress field. This procedure also allows comparison between the stress

Seed crack fraction	T_{MD} 1 ps	T_{MD} 5 ps
0.2	28.9	13.4
0.3	32.1	13.4
0.4	32.1	13.4

Table 6.3: Critical loading G_{100} estimated using classical MD simulations for a $1200\times400\times3.68$ Å^3 slab with two different minimum times T_{MD} between loading.

distributions predicted by continuum analysis, classical geometry optimisation and hybrid simulation.

In order to compare the stress distributions obtained in different fracture simulations with one another and with the analytical continuum elastic results discussed in Chapter 5, we need to calculate the stress and strain on each atom. A truly atomistic definition of stress is hard to provide: the virial theorem is commonly used to define local stresses but this can result in unphysical oscillations since it should only be used to calculate macroscopic properties, averaging both temporally and spatially over large numbers of atoms [149, 150]. An alternative method proposed by Hardy [151], in which the stress fields are evaluated at a fixed spatial point uses a localisation function to perform local spatial averaging of the density, momentum and potential energy around each atom, has recently been shown to predict stresses which are at least as accurate as those calculated with the virial theorem [150].

In this work we avoid the problems associated with defining a local stress by instead considering the local strain of each atom. This can be measured by comparing the atom's position to that it would have in an unstrained crystal. We can then calculate the stress from this strain using linear elasticity theory. We have developed a method to do this for four-fold coordinated atoms in tetragonally bonded structures. The neighbours of each atom are used to define a local set of cubic axes, and the deformations along each of these axes are combined into a matrix E describing the local deformation:

$$E = \begin{pmatrix} | & | & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ | & | & | \end{pmatrix} \tag{6.5}$$

where, for example $\mathbf{e}_1$ is the relative deformation along the first cubic axis. To compute the local strain of the atom, we need to separate this deformation into a contribution due to rotation and one due to strain. This can be done by finding the polar decomposition of E, by writing E in the form $E = SR$ with R a pure rotation and S a symmetric matrix. Diagonalising the product EE^T allows R and S to be calculated. The strain components ϵ_{xx}, ϵ_{yy}, ϵ_{zz}, ϵ_{xy}, ϵ_{xz} and ϵ_{yz} can then be calculated by rotating S to align the local cubic axes with the Cartesian axes:

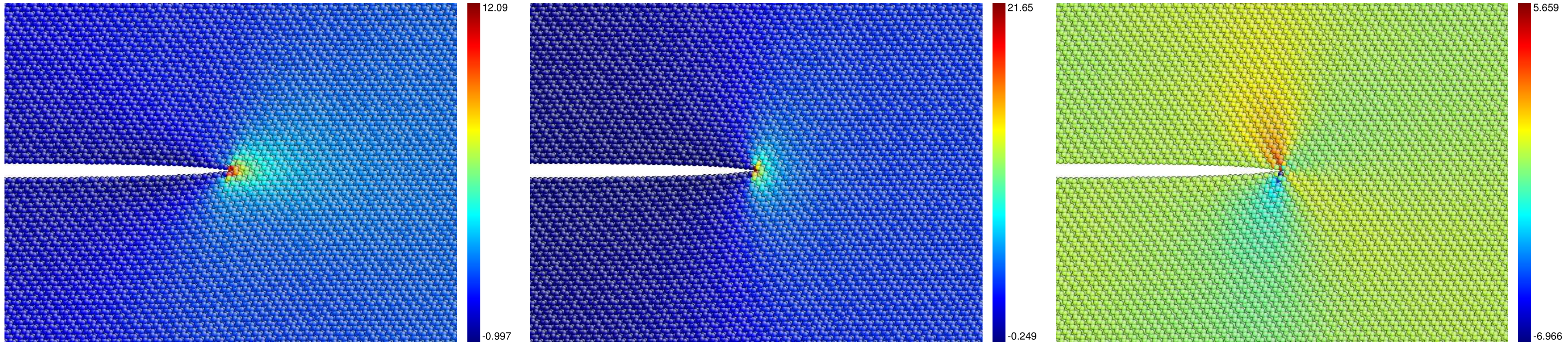
$$R^T S R = I + \epsilon = \begin{pmatrix} 1+\epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{xy} & 1+\epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{xz} & \epsilon_{yz} & 1+\epsilon_{zz} \end{pmatrix}. \tag{6.6}$$

Finally, the atomistic stress can be computed simply as $\sigma = C\epsilon$ where C is the 6×6 matrix of elastic constants defined in Eq. 2.17; for cubic crystals the only independent elastic constants are C_{11}, C_{12} and C_{44} so these make up the only non-zero elements of C.

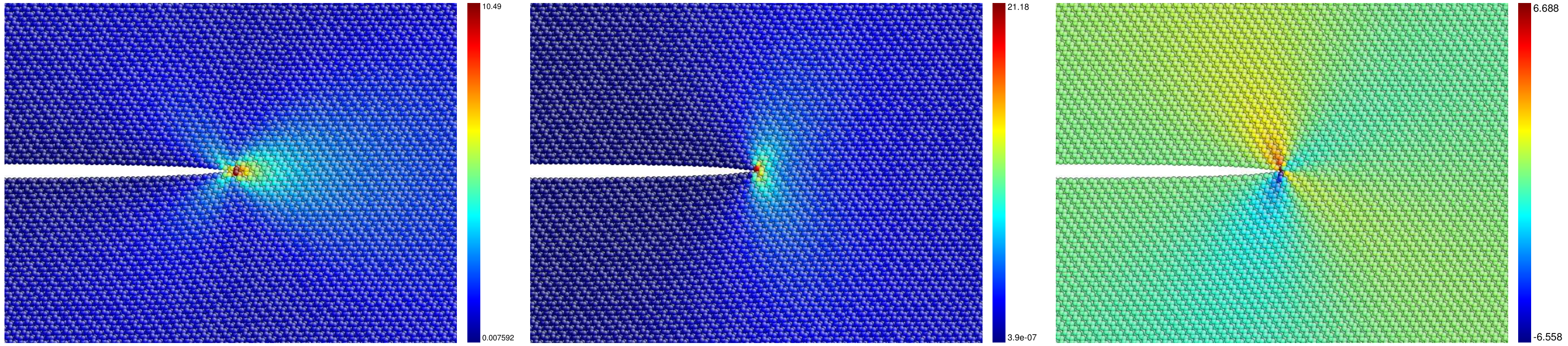
Comparison of Atomistic and Irwin Stress Fields

The method outlined above allows the stress distribution to be calculated for any atomistic configuration. This might be a snapshot from an MD simulation or the outcome of a classical or hybrid geometry optimisation. It is most useful to do this with a very well relaxed geometry to remove fluctuations caused by kinetic effects: this is essential in order to make comparisons with the continuum predictions of the Irwin K-field.

The Irwin solutions are derived for the case of an infinitesimal slit crack, where the crack tip is a mathematically exact point. To compare with an atomistic model, we have to choose where to centre the stress field. This can be done by minimising the RMS difference between the atomistic and continuum fields along the line ahead of the crack tip, as shown in Fig. 6.7. All the stress components apart from σ_{xx} give rise to clear minima in the error at around $x = 107$ Å which is within 0.2 Å of the bond at the crack tip. This fitting procedure is not ideal since the Irwin field only applies in the near-field: in the thin strip geometry the stress components far ahead of the tip do not tend to zero, so there are large discrepancies between the fields in this region. This may explain why the minimum in the stress field error for σ_{xx} does not coincide with the other minima. The agreement between measured atomistic and continuum stress fields is good in the near field, especially for the σ_{yy} component, as can be seen from the stress maps shown in Fig 6.5.



(a) Atomistic σ_{xx} (b) Atomistic σ_{yy} (c) Atomistic σ_{xy}



(d) K-field σ_{xx} (e) K-field σ_{yy} (f) K-field σ_{xy}

Figure 6.5: Comparison of classical atomistic and K-field stress distributions at $G = 2.16$ J/m^2. The first stress map in each column is computed from a relaxed SW configuration as described in the text, the second is from the Irwin K-field near field solutions for the same loading.

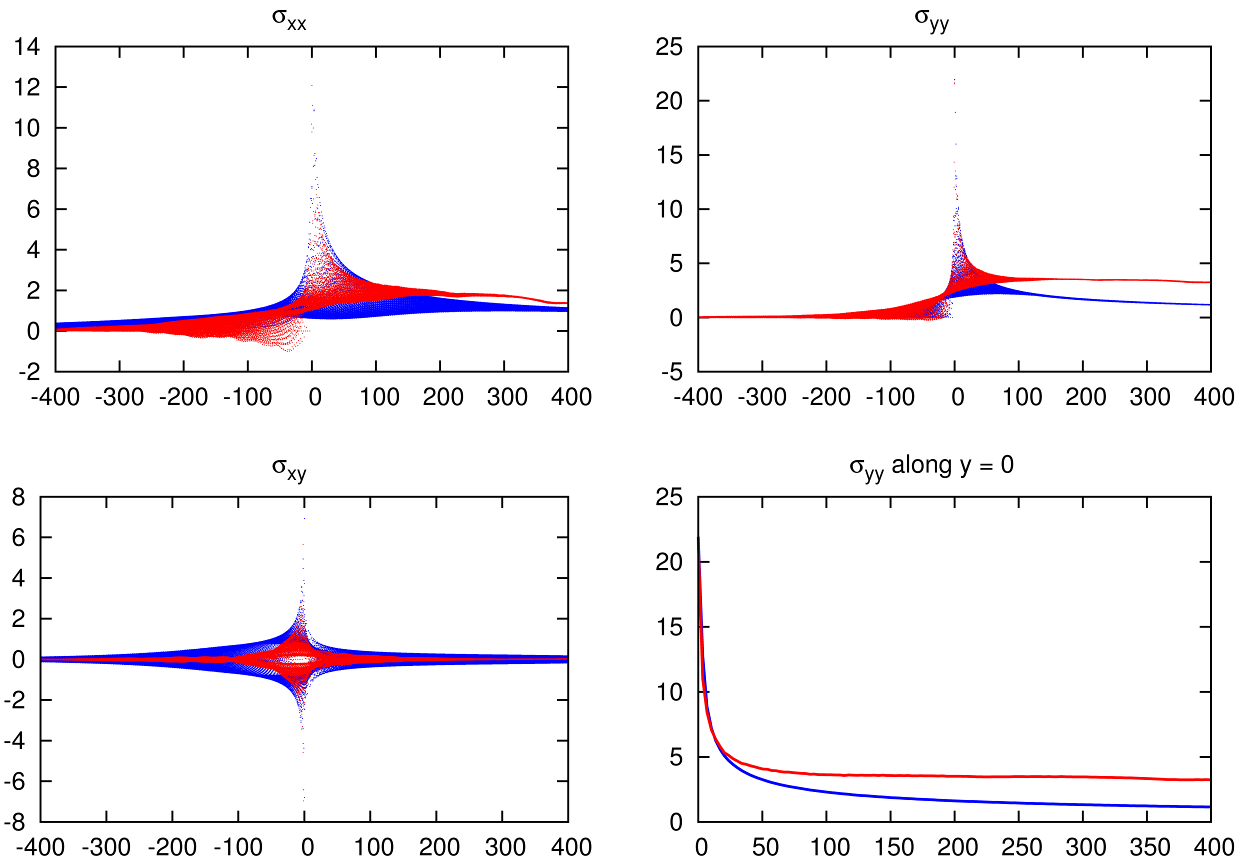


Figure 6.6: (a) Comparison of the stress distributions for $1200 \times 400 \times 3.68$ Å^3 crack slab optimised with Stillinger-Weber potential (red) and the analytical K-field solution (blue). The lower right plot shows σ_{yy} on the line $y = 0$ ahead of the crack tip. All atoms apart from those within 50 Å of the edges are shown. The x axis is centred on the crack tip, and units are Å. Units of stress are GPa. The load is $G = 2.16$ J m^2, compared to the Griffith critical load of $G_c = 2.7$ J m^2.

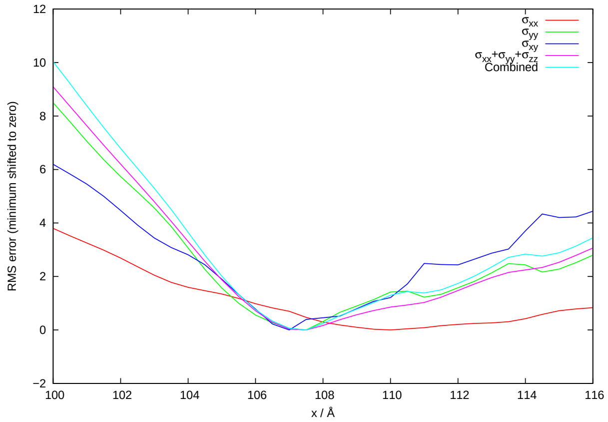


Figure 6.7: RMS difference between atomistic and K-field stress components as a function of the position at which the K-field is centred. The curve labelled 'Combined' is the quadrature mean of the σ_{xx}, σ_{yy} and σ_{zz} curves. The curves have been shifted vertically so their minima lie at $y = 0$ to allow comparison.

Planar projections of these fields are shown in Fig. 6.6, allowing direct comparison between the atomistic and Irwin values. The K-field solution agrees well with the atomistic calculation even close to the tip. This good agreement indicates the length scale for small scale yielding predicted by the Stillinger-Weber potential is small.

The length scale over which the stress fields decay to zero behind the crack tip and to the external load ahead of the tip enables us to estimate the size of the simulation cell required for our simulations. Fig. 6.6 indicates that the region over which the stress fields are significant extends approximately 200 Å either side of the crack tip, in good agreement with the estimate from fracture toughness calculations presented above.

As can be seen from the plot of σ_{yy} along the line $y = 0$ ahead of the crack tip shown in Fig. 6.6, at large distances ahead of the crack, the atomistic field tends to the external load whereas the near field analytical solution tends to zero. This discrepancy arises because the K-field solutions are exact only for a crack in an infinite medium. In most cases this approximation is sufficient but here the boundary conditions are prescribed close to the crack, so there is a significant difference in the resulting stress fields.

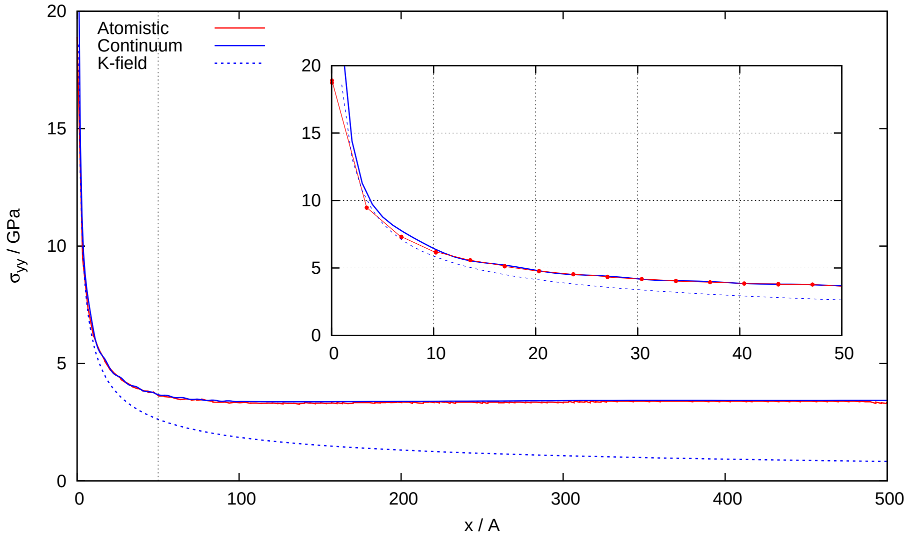


Figure 6.8: σ_{yy} along the line $y = 0$ ahead of a stationary crack tip located at the origin. The energy release rate is $G = 1.75$ J/m^2. The inset shows the region $x < 50$ Å, indicated by the vertical dashed line on the main plot.

6.5 Stress Fields of a Semi-Infinite Crack in an Infinite Strip

The solution to the significant stress difference between models is to use Knauss' continuum results for the stress field in an infinite strip containing a semi-infinite crack. Appendix A presents an outline of Knauss' approach and describes the combination of analytical and numerical integration techniques employed to calculate the stress on the line $y = 0$ ahead of a crack tip. The atomistic stress distribution obtained from a classical geometry optimisation with the Stillinger-Weber potential is shown in Fig. 6.8, along with the results of the integration calculated using Eq. A.23. To facilitate comparison between the two models, the elastic constants used for the the numerical calculation were those predicted by the Stillinger-Weber potential, rather than those of real silicon. The asymptotic K-field contribution to σ_{yy} is also shown in the figure; it is the term in the stress proportional to $1/\sqrt{x}$.

Agreement between the continuum and atomistic approaches is very good for $x > 10$ Å, with the two curves lying exactly on top of one another. Knauss' solution is only exact for the case of a semi-infinite crack in an infinite strip but we can see from this calculation that the system size used for the atomistic calculation is large enough to approach this limit very closely. In contrast, the Irwin K-field solution is approximately correct in an annular region extending from 10 Å out to of the order of 100 Å from the crack tip; this explains the reasonable near-field agreement seen in the stress maps of Fig. 6.5 and the deviation at larger distances seen in the plots in Fig. 6.6. The breakdown in the agreement between the continuum and atomistic descriptions for very small x indicates that the length scale for small scale yielding predicted by the Stillinger-Weber model is of the order of 10 Å— this is the size of the region within which non-linear processes contribute significantly to the stress distribution. This calculation provides reassurance that the classical atomistic model chosen here does an excellent job of describing the elastic relaxation of the crack system, even though it fails to describe the bond breaking processes near to the tip accurately.

It should be noted that these results are obtained from purely classical geometry optimisations. We will describe the effect of introducing quantum mechanics in the vicinity of the crack tip on the stress distribution in Chapter 7. The fact that the length scale for small scale yielding appears to be so small is encouraging: in hybrid fracture simulations the region to be treated with quantum mechanical accuracy will extend sufficiently far from the tip to include all of this region.

6.6 Summary

In this chapter I have discussed the development of a general purpose computer code to model brittle fracture within the LOTF scheme and have highlighted important aspects of the code such as the need for a robust algorithm to track the crack tip as it moves. The thin strip loading geometry discussed in the previous chapter suggests a straightforward loading scheme that facilitates direct comparison between simulation and experiment. I have also reviewed a number of classical calculations made using the code which justify the choice of various parameters such as the overall size of the simulation cell.

An atomistic approach to the calculation of stress has been introduced: this is a powerful tool which will be of utility in understanding the important differences between the classical and quantum mechanical descriptions of fracture that will be described in the following chapter. An exact linear elastic solution for an equivalent geometry to that used in the simulations gives a stress distribution along the line ahead of the crack tip that agrees very well with that obtained from geometry optimisation with the classical Stillinger-Weber potential, indicating that the potential describes the long-range elastic behaviour of silicon to high accuracy.

We conclude that the crack slab needs to be around 1200 Å in length to adequately capture the effects of the long range stress field that drives the fracture process; since a cell of this size contains of the order of 100 000 atoms a multi-scale simulation is indeed the only feasible approach to carrying out an *ab initio* simulation of the fracture process.

Chapter 7

Results

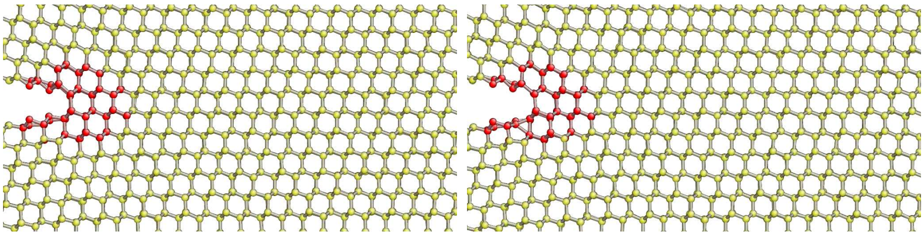
7.1 Introduction

This chapter presents the main results obtained for the brittle fracture of the Si(111) system. We begin by describing hybrid simulations conducted using a variety of tight binding potentials. These simulations indicate that a crack tip reconstruction plays a central rôle in the fracture process. Further information about this reconstruction is then obtained from calculations on crack tip clusters extracted from the tight binding dynamical simulations. Finally, I report results obtained from fully *ab initio* hybrid simulations of fracture. The results of these simulations confirm the indications provided by the exploratory tight binding calculations.

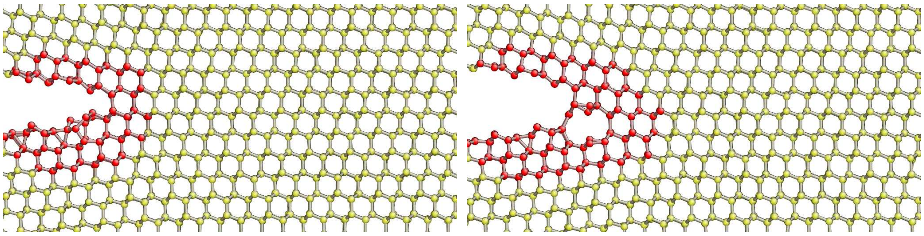
7.2 Tight Binding Hybrid Results

Using a tight binding quantum mechanical method within the LOTF hybrid scheme provides an improved description of the bond breaking processes in comparison to purely classical simulations, while remaining computationally inexpensive relative to *ab initio* calculations. Hybrid TB simulations are thus ideal for exploring the behaviour of the silicon fracture system. The results presented in this thesis were initially observed in TB simulations, and then confirmed with *ab initio* calculations.

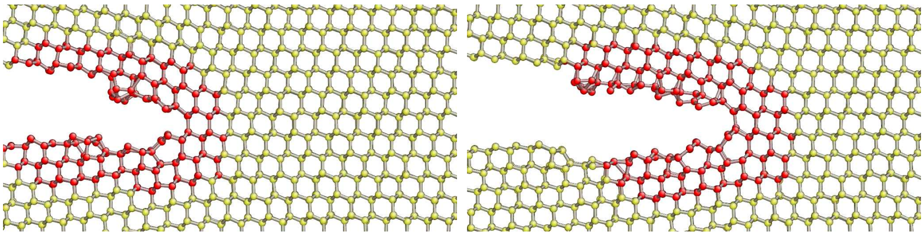
Hybrid fracture simulations have been carried out using a variety of different TB potentials. In each case brittle fracture is observed, with continuous propagation above a critical load. The Kwon [87] and Bowler [88] semi-empirical TB potentials and the more accurate DFTB [91] method give qualitatively similar results, both for room temperature simulations ($T = 300$ K) and at low temperature



(a) $t = 37.5$ fs, $G = 2.5$ J/m^2 (b) $t = 50.0$ fs, $G = 2.5$ J/m^2



(c) $t = 62.5$ fs, $G = 4.0$ J/m^2 (d) $t = 75.0$ fs, $G = 5.8$ J/m^2



(e) $t = 87.5$ fs, $G = 5.8$ J/m^2 (f) $t = 100.0$ fs, $G = 7.9$ J/m^2

Figure 7.1: Snapshots from LOTF tight binding MD simulation at a temperature of 300 K using the Kwon *et al.* [87] tight binding model. The red atoms are treated quantum mechanically and the yellow atoms with the classical potential. All atomistic visualisations were produced with AtomEye [152]. See text for a detailed description.

($T = 10$ K). In all cases the most striking feature of the fracture dynamics is the formation of a crack-tip reconstruction, reminiscent of the Pandey reconstruction of the Si(111) surface. The reconstruction forms spontaneously at loads less than the Griffith critical load G_c via a bond rotation, resulting in 5– and 7–membered rings. Formation of this reconstruction relieves the stress along the main load axis, thus inducing additional lattice trapping. An example hybrid trajectory generated in a simulation using the Kwon TB potential is illustrated in Fig. 7.1. Panel (a) shows the initial configuration, with an unreconstructed crack tip as predicted by the Irwin displacement field. After just 50 fs of dynamics, the reconstruction starts to form. Above the critical load for fracture, the slowly moving crack occasionally visits the reconstructed tip configuration as it propagates, for example as seen in

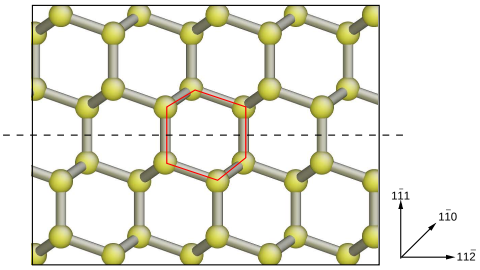
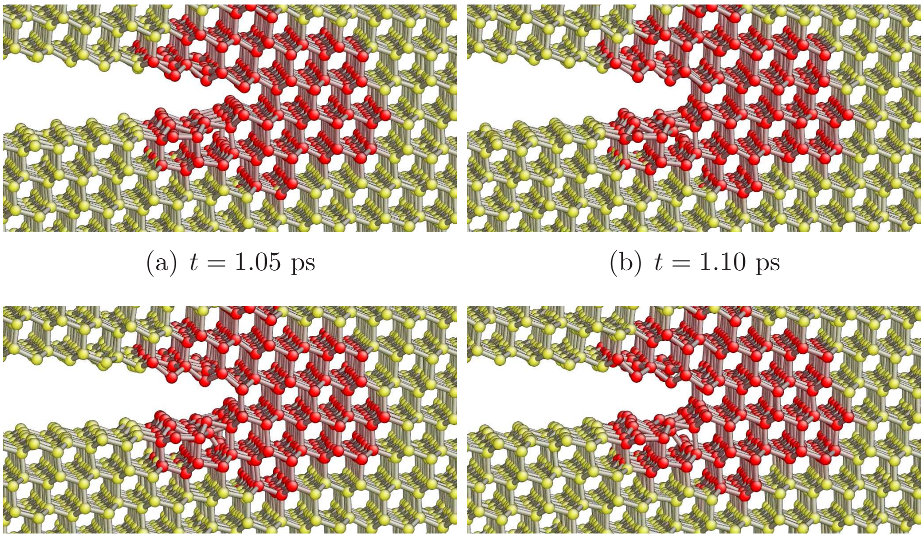


Figure 7.2: Broken mirror symmetry of the $(111)[1\bar{1}0]$ orientation of a bulk silicon crystal. The highlighted 6–membered ring clearly indicates that there is an asymmetry with respect to reflection about the black dashed line.

panels (d) and (f). In panel (d) it can be seen that the 5–membered ring has migrated backwards via a further bond rotation to form a 5–6–7 reconstruction. For sufficiently slow propagation, the [111] orientated bonds one atomic layer below the initial crack plane and further along the propagation direction can break before the bridging bonds associated with the 7–membered ring, since stress is mostly concentrated at these points after the reconstruction. Every reconstruction event causes a downward [111] step in the crack path as seen in panel (e). All these steps are in the same direction due to the broken mirror symmetry of the crystal depicted in Fig. 7.2. The crack tip reconstruction always forms on the lower crack face for the system as oriented in Fig. 7.1. The simulations therefore predict a systematic deflection of the crack path from the (111) plane at sufficiently low velocities, with fast cracks being dynamically steered away from the instability.

Truly three dimensional simulations are prohibitively computationally expensive, but it has been verified that increasing the depth of the system in the periodic direction to as many as eight layers (a depth of 29.44 Å) does not change the tip reconstruction behaviour. Fig. 7.3 shows snapshots from an eight layer simulation using DFTB quantum calculations; the 5–7 reconstruction forms across almost the full width of the system before the onset of fracture. It is prevented from spanning the system completely by the strongly alternating up-down Hanneman reconstruction predicted by the DFTB model. Enforcing charge self-consistency during the DFTB calculation to prevent unphysical charge transfer would reduce the extent of this Hanneman reconstruction, but unfortunately this makes the calculation an order of magnitude slower and is thus not feasible. Since all the simulations are conducted at relatively low temperature and the system is highly driven, thermal activation and entropic effects, which could cause qualitatively different behaviour in three-dimensional systems, are unlikely to play a significant rôle.



(a) $t = 1.05$ ps (b) $t = 1.10$ ps

(c) $t = 1.15$ ps (d) $t = 1.20$ ps

Figure 7.3: Snapshots from DFTB hybrid simulation at a temperature of 300 K and a load of $G = 2.5$ J/m^2 for an 8-layer crack system. Geometry of slab is $1200 \times 400 \times 29.44$ Å^3, containing 725 760 atoms. The formation of the 5–7 crack tip reconstruction across seven of the eight layers can be seen.

7.3 Crack Tip Reconstruction

The results outlined above are already encouraging since many different TB methods predict the same basic behaviour. To be truly convincing, more detailed studies of the crack tip reconstruction process must be made with accurate *ab initio* methods.

Let's consider a simple spring model to describe the origin of the crack tip reconstruction, illustrated in Fig. 7.4. For the unreconstructed tip, the critical bond (indicated with an arrow in the figure) can be modelled as a single spring. If we assume constant strain as the reconstruction forms, then the reconstructed system can be modelled by three springs in series, each extended less than the single spring that represents the unreconstructed tip. The extension of this spring is reduced since the load is now shared out between three bonds. The overall stress concentration is therefore reduced.

Ab Initio Crack Tip Cluster Calculations

The local chemical energy cost of the bond rotation associated with the formation of the tip reconstruction can be evaluated by calculating the energies of crack tip clusters extracted from dynamical simulations using an accurate first principles method.

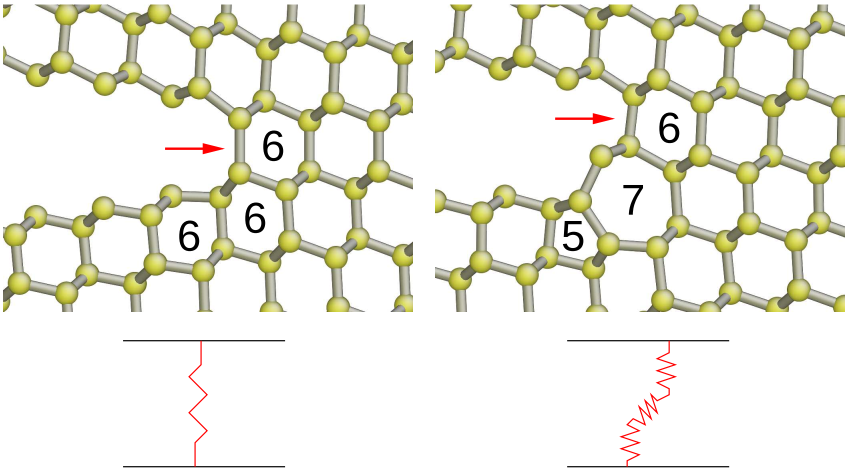
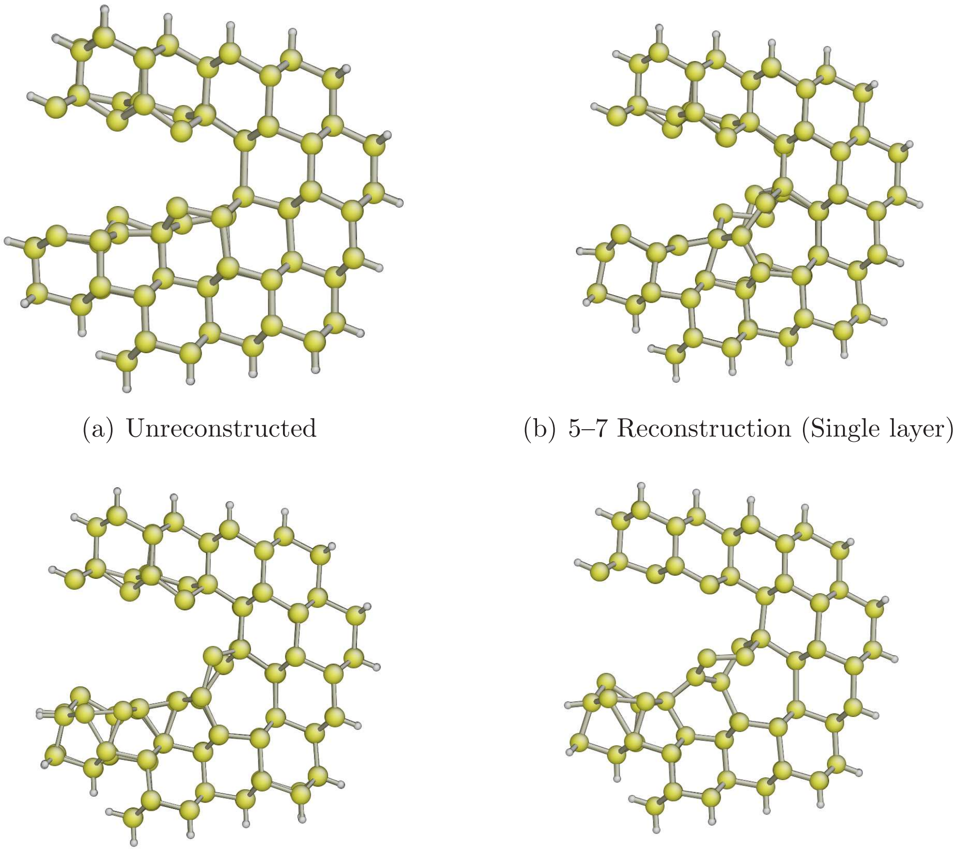


Figure 7.4: Simple spring model for crack tip reconstruction process. The left hand side shows the unreconstructed tip, and the right the tip after the 5–7 reconstruction has taken place. See text for description.

This analysis was carried out for three variations of the tip reconstruction observed in the hybrid simulations with the TB force models. For technical reasons, these early simulations were carried out with a depth of two unit cells (7.68 Å) in the periodic direction. This leads to two different forms of the 5–7 reconstruction: one in which it forms in a single layer, and another where it forms across both layers. A further variant of the basic reconstruction occurs when a subsequent bond rotation causes the 5–membered ring to migrate backwards with respect to the propagation direction, resulting in a 5–6–7 structure.

Configurations representative of each of these reconstructions and containing the same set of atoms were extracted from hybrid TB simulations and rescaled to the same sub-critical ($G < G_c$) load. These clusters are illustrated in Fig. 7.5. Each cluster contains approximately 100 silicon atoms plus terminating hydrogen atoms. The outer ring of silicon atoms and the termination atoms were fixed in place and then the interior atoms were allowed to relax. The resulting energies of these clusters using the Kwon TB potential and DFT methods are shown in Table 7.1. All the structures are stable in both the TB and DFT descriptions, with geometry optimisation carried out until the RMS force was below 0.01 eV/Å. However the energy of the reconstructed structures is higher than that of the unreconstructed tip, indicating as expected that locally the reconstructions cost energy to create. The two 5–7 reconstructions are similar in energy, suggesting that structural variations in the periodic direction are not very significant. Since the 5–6–7 structure results



(a) Unreconstructed

(b) 5–7 Reconstruction (Single layer)

(c) 5–7 Reconstruction (Both layers)

(d) 5–6–7 Reconstruction

Figure 7.5: Crack tip clusters obtained by CASTEP geometry optimisation of structures extracted from TB hybrid simulations.

Structure	RMS position difference /Å	Tight Binding		DFT (CASTEP)	
		E / eV	ΔE / eV	E / eV	ΔE / eV
Unreconstructed	0.000	-811.0	0.0	-11534.82	0.00
5–7 (Single layer)	0.038	-804.6	6.4	-11533.69	1.13
5–7 (Both layers)	0.041	-805.4	5.6	-11533.14	1.68
5–6–7	0.119	-803.9	7.1	-11532.26	2.56

Table 7.1: Energies of the relaxed crack tip cluster structures illustrated in Fig. 7.5. The tight binding results give energies resulting from geometry optimisation using the Kwon *et al.* [87] model, and the DFT results results were obtained using CASTEP. ΔE is the energy difference from the unreconstructed tip in each case.

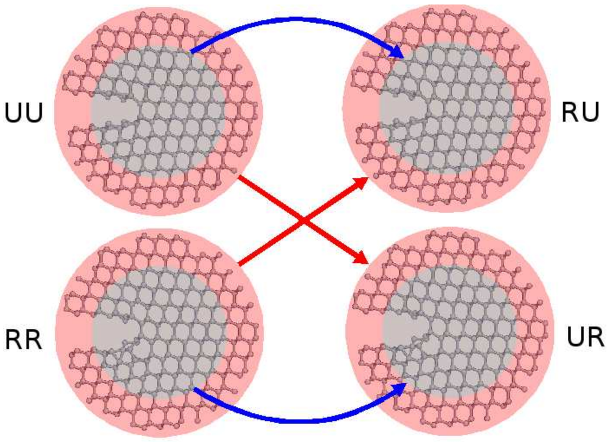


Figure 7.6: Clusters made from crossing over the crack tip boundary conditions. UU = unreconstructed outer, unreconstructed inner; RU = reconstructed inner, unreconstructed outer; RR = reconstructed inner, reconstructed outer; UR = unreconstructed inner, reconstructed outer.

in more under- and over-coordinated atoms, it is not surprising that it is higher in energy. The RMS difference in the positions of the atoms in each cluster from those of the unreconstructed tip can be used to quantify the degree of rearrangement. Again the two 5–7 cases can be seen to be very similar. The energy cost ΔE increases with increasing RMS position difference. From now on we shall consider only the unreconstructed tip and the full 5–7 reconstruction, the latter being the most commonly observed structure in the dynamical simulations.

The boundary conditions applied to the unreconstructed and reconstructed crack tip clusters are not the same, since the atoms at the edges of the cluster were fixed at the the positions extracted from hybrid simulations in much larger systems. Interchanging the boundary conditions between the two systems in the manner illustrated by the red and blue arrows in Fig. 7.6 and then re-relaxing the clusters with CASTEP confirms that it is the outer boundary conditions that determine the energy ordering. This indicates that the naïve spring model discussed above is too simple, since in reality there is no fixed boundary close to the crack tip — all atoms are free to move. There is a tradeoff between the local energy cost associated with the formation of the reconstruction and elastic energy relief over a large region.

In principle, we could increase the radius of the clusters used for these calculations until we include sufficient atoms that the elastic energy reduction balances the chemical cost of the bond rotation and $\Delta E = 0$; in practice this would require many thousands of atoms to be included in the simulation which is beyond the capability of DFT. Atomistic stress calculation as described in Section 6.4 provides a direct way to visualise this energy reduction; we shall shall return to this with hybrid *ab initio* calculations in Section 7.4 below.

These results explain why previous *ab initio* studies of crack tips, for example the work of Perez and Gumbsch [123], have not detected these reconstructions: in these works the edges of the crack tip cell are constrained to match the Irwin continuum displacement field. Firstly, the simulation cells used are not large enough to include enough of the stress relieved region to make the reconstructed configuration lower in energy, and secondly the fixed boundary conditions do not give sufficient scope for internal rearrangement to allow tip reconstructions to form. It is only with a hybrid technique that we can hope to observe and explain these kinds of phenomena.

Hybrid Force Integration

We have seen that cluster calculations are not suitable for determining the energetics of the crack tip system since the choice of boundary makes a difference, therefore, we need to compare relaxed total energies of the entire system in the unreconstructed and reconstructed cases. Each structure contains over 90 000 atoms so it is clearly not possible to evaluate the quantum mechanical energy from first principles. Within the LOTF scheme, there is not a meaningful total energy for the hybrid classical/quantum system; the method matches forces between the QM and classical regions, rather than energies, so the solution is to use these accurate forces to evaluate the energy difference between two configurations by force integration:

$$\Delta E = \int_\gamma \mathbf{F}\left(\mathbf{R}\right) \cdot \mathrm{d}\mathbf{R} \tag{7.1}$$

where $\mathbf{R}$ denotes all atomic positions and the integration contour γ can be any path between the two configurations of interest. The start and end configurations are obtained by hybrid minimisation as described in Section 6.2, resulting in $\mathbf{F} = 0$ at both integration limits.

For the crack tip reconstruction process, there is no obvious reaction coordinate, since the bond rotation which forms the 5– and 7–membered rings involves the correlated motion of many atoms. I have therefore chosen the simplest contour connecting the two minima: linear interpolation between the relaxed unreconstructed state with atomic coordinates $\mathbf{R}_U$ and the reconstructed state with coordinates $\mathbf{R}_R$. The QM region is fixed during the integration process. The forces $\mathbf{F}(\mathbf{R})$ are cal-

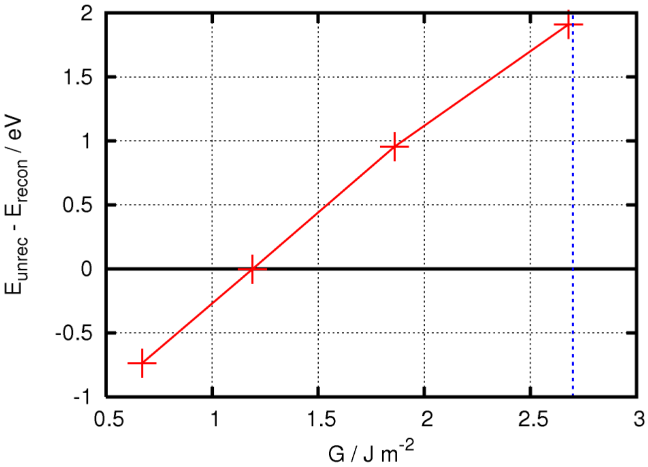


Figure 7.7: Energy difference $\Delta E = E_U - E_R$ obtained using hybrid force integration as a function of loading G. The blue dashed line shows the Griffith critical loading $G_c = 2\gamma$. QM force model was that of Kwon *et al.* [87].

culated in the standard LOTF fashion: classical and quantum forces are evaluated for the two regions, then the adjustable potential is optimised to reproduce the quantum forces. The forces used for the integration are the sum of the classical and corrective forces to ensure that there is no mechanical mismatch at the boundary. The integration path is discretised into $N+1$ samples according to

$$\Delta \mathbf{R} = \frac{1}{N}\left(\mathbf{R}_R - \mathbf{R}_U\right) \tag{7.2}$$

$$\mathbf{F}_i = \mathbf{F}(\mathbf{R}_U + i\,\Delta\mathbf{R}), \quad 0 \leq i \leq N \tag{7.3}$$

and then Eq. 7.1 can be evaluated using Simpson's Rule:

$$\Delta E \approx \frac{\Delta \mathbf{R}}{3} \cdot \left[\mathbf{F}_0 + 2\sum_{j=1}^{N/2-1} \mathbf{F}_{2j} + 4\sum_{j=1}^{N/2} \mathbf{F}_{2j-1} + \mathbf{F}_N\right] \tag{7.4}$$

The step-size $\Delta\mathbf{R}$ required for accurate integration of the energy difference was calibrated using force integration with the classical potential alone, where the energy difference can be calculated exactly — typically $N = 20$ gives good results. The method was validated by confirming that perturbing the integration path does not affect the value of ΔE obtained.

This procedure for calculating the energy difference has been carried out with the Kwon TB potential as the QM force model. The unreconstructed and reconstructed configurations were extracted from hybrid dynamical simulations, then rescaled to the same load and relaxed within the hybrid scheme. Damped molecular dynamics minimisation is sufficiently fast to carry out the relaxation efficiently. Repeating this process at a number of different loads gives the data shown in Fig. 7.7. The range of lattice trapping predicted by the TB model used here is relatively large, meaning that there is a local minimum corresponding to a stationary crack tip in both the reconstructed and unreconstructed states for loadings as low as 0.67 J/m^2. At this low load, the unreconstructed configuration is 0.74 eV lower in energy than the reconstructed structure. As the loading is increased, the energy difference $\Delta E = E_U - E_R$ increases, with the reconstruction becoming globally stable at $G \approx 1.2$ J/m^2, well below the Griffith critical loading G_c. For larger loads the preference becomes stronger still.

This force integration procedure is equivalent to the thermodynamic integration method [153] commonly used to calculate free energy barriers, but carried out at zero temperature. The absence of a good reaction coordinate means I have not been able to calculate the energy barrier to formation of the reconstruction. Nevertheless, the observation that the reconstruction forms spontaneously in dynamical simulations at temperatures as low as 10 K means that this barrier must be small, at least within the tight binding approximation.

Summary of Tight Binding Results

The conclusion from tight binding hybrid simulations is that the 5–7 reconstructed crack tip becomes the equilibrium configuration well below the critical load, and that this reconstruction causes significant stress relief, enhancing the lattice trapping effect and delaying the onset of fracture. The next question is whether this remains the case when a more accurate quantum mechanical method is used to describe the crack tip region.

7.4 *Ab Initio* Hybrid Results

Whilst the tight binding results discussed above are very encouraging and consistent across a range of different TB parametrisations, the semi-empirical nature of all tight binding models mean that results obtained using these techniques are not generally accepted as convincing evidence for the existence of particular effects.

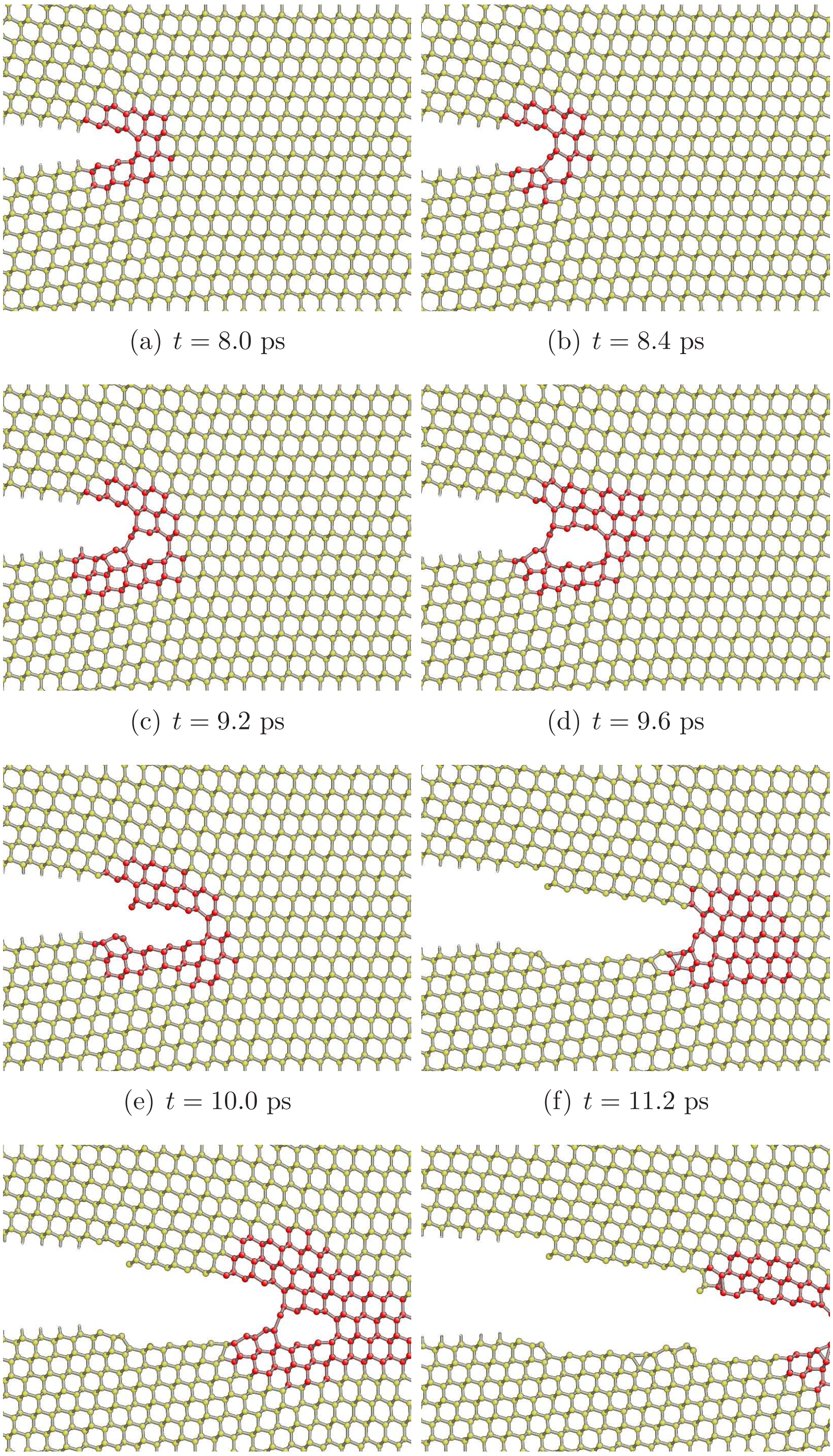
Other studies of Si(111) fracture using different tight binding models have reported different outcomes depending on the choice of TB potential: for example the DCET hybrid model predicts brittle fracture with smooth, unreconstructed surfaces [108], whereas a linear-scaling tight binding study on systems containing around 10 000 atoms predicts both upward and downward steps in the fracture surface [154]. The work described above reveals that an accurate description of the energetic ordering of crack tip defects is essential.

The only solution is to carry out fully *ab initio* hybrid fracture simulations using a DFT code to remove all empirical assumptions from the quantum calculation. As has been mentioned, this is a very computationally expensive procedure; to make progress it is essential to keep the quantum mechanical region as small as possible. For this reason the simulation cell was reduced in depth from the double cell used for TB simulations to a single unit cell with a periodic lattice parameter of 3.68 Å. [1] In an attempt to prevent the crack from closing up at low loads, the open seed crack surfaces were terminated with hydrogen atoms. This required additional terms to be added to the Stillinger-Weber potential to describe the Si–H interactions; the parametrisation of Kohen *et al.* [65] was used for this purpose. The addition of hydrogen atoms also necessitated a reduction of the molecular dynamics time-step to 0.1 fs, in order to correctly integrate the equations of motion for the lightweight hydrogen atoms. Ultimately, this hydrogen termination was found to be unnecessary, since the lattice trapping induced by the tip reconstruction is even greater in *ab initio* hybrid simulations than it was with TB, so fracture is suppressed until the loading is significantly above G_c.

Fig. 7.8 shows snapshots from a hybrid simulation using CASTEP as the QM force model which has a minimum time between loadings of $T_{\mathrm{MD}} = 0.2$ ps, the same as that used for the TB simulations. This loading rate significantly overestimates the critical loading, since the system does have sufficient time to get over the lattice trapping barrier at loads that would cause fracture in reality. Unfortunately, realistic time scales are inaccessible to simulation, so we are forced to strike a balance between

---

[1] Originally, a technical limitation prevented the cell from being smaller in any direction than the neighbour cutoff distance used for the classical potential, typically 5 Å. Initially this was worked around by considering the full system to consist of a real and an image plane of atoms, stacked in the periodic lattice direction. QM forces are evaluated only for the the real atoms, and copied to each of the images. Classical forces and atomic positions are symmetrised at each time step. Recently the LIBATOMS code has been modified to remove this limitation, allowing arbitrarily thin slabs to be modelled. This is done by allowing individual atoms to appear multiple times in neighbour lists, providing each occurrence has a different periodic shift.



(a) $t = 8.0$ ps (b) $t = 8.4$ ps

(c) $t = 9.2$ ps (d) $t = 9.6$ ps

(e) $t = 10.0$ ps (f) $t = 11.2$ ps

(g) $t = 11.6$ ps (h) $t = 12.4$ ps

Figure 7.8: Snapshots from LOTF/CASTEP MD simulation at 300 K. Note the formation of the reconstruction at $t = 8.4$ ps, the subsequent 'microcracking' and stepping of the crack. The reconstruction forms again at $t = 11.2$ ps.

T_{MD} / ps	G_{recon} / J/m^2	
	Tight Binding	*Ab Initio*
0.2	5.5	19.4
1.0	3.5	8.5
5.0	2.5	6.0
∞	1.2	2.1

Table 7.2: Energy release rate G_{recon} at which 5–7 crack tip reconstruction forms dynamically as observed in simulation as a function of the minimum time T_{MD} between loadings; $T_{\mathrm{MD}} = \infty$ corresponds to the lowest load at which the reconstruction is found to be globally stable by hybrid minimisation. Below these loads the TB crack is stable in the unrelaxed configuration, whereas the *ab initio* crack closes up.

computational expense and simulation time. Static calculations at lower loads giving further insight into the load at which the reconstruction becomes stable will be discussed below. The fracture dynamics are qualitatively similar to those seen with tight binding. The reconstruction forms before the onset of fracture (panel (b) in the figure). The stress relief and consequent increase in lattice trapping induced by the reconstruction require the load to be significantly further increased before the crack can propagate. As in the TB simulations, the next bonds to break are located ahead of the bridging bond associated with the 7–membered ring and one atomic layer below the initial crack plane, as shown in panels (c) and (d). We see in panel (e) that, eventually, the bridging bond breaks, whereupon the crack takes a downward [111] step. The crack occasionally revisits the reconstructed configuration as it moves; see panels (f) and (g). Every time it does so there is a downward step in the crack path, exactly as seen in tight binding hybrid simulations such as the one shown in Fig. 7.1.

Repeating the hybrid dynamics with lower loading rates (larger values of T_{MD}) gives qualitatively the same behaviour, with the tip reconstruction forming before the onset of fracture, but the load at which the reconstruction forms and at which fracture initiates is vastly reduced. The upper section of Table 7.2 shows the loads at which formation of the tip reconstruction has been observed in dynamical simulations. The *ab initio* hybrid fracture dynamics are similar to those seen with TB, but the effect of the reconstruction is even greater: the lattice trapping induced results in a large overestimate of the fracture toughness.

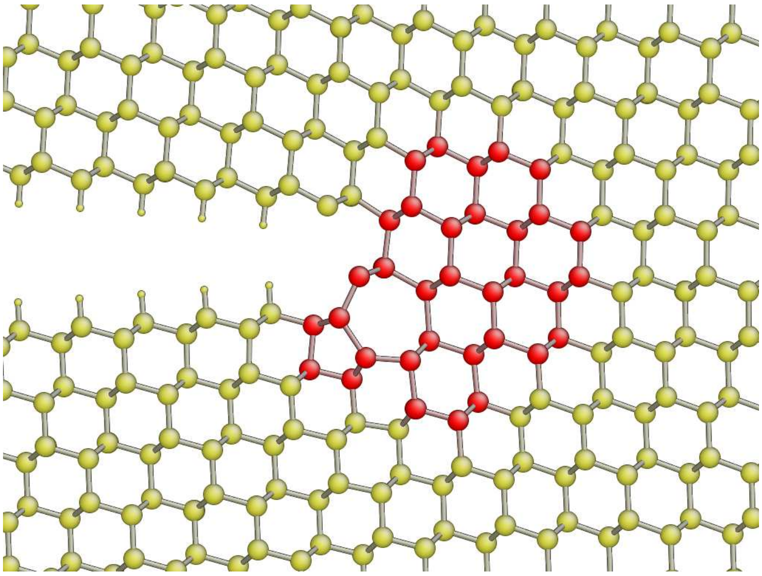


Figure 7.9: Ground state structure obtained using hybrid geometry optimisation with CASTEP at an applied load of $G = 2.1$ J/m^2, the lowest load which does not close up on relaxation.

Ab Initio Hybrid Relaxation

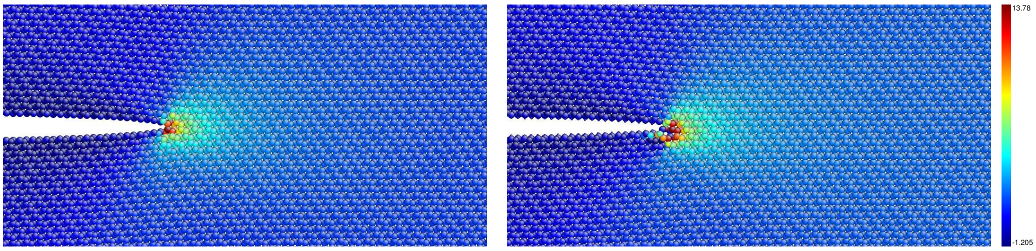
To show that the reconstructed geometry is lower in energy than the crack structure predicted by the Irwin solutions, it is desirable to repeat the force integration procedure conducted above with the crack tip region treated at an *ab initio* level of accuracy. The procedure is as before: unreconstructed and reconstructed configurations are extracted from the hybrid dynamical simulations and rescaled to the same applied load. We then relax the entire structure using the hybrid forces. The force calculations are now much more expensive than in the TB case, so the derivative extrapolation adaptation of the conjugate gradient method described in Section 6.2 is essential to carry out these geometry optimisations efficiently.

Surprisingly the unreconstructed crack tip displacement field predicted by continuum mechanics is never stable with the *ab initio* force model. Below a load of $G = 2.1$ J/m^2, much lower than the *ab initio* Griffith critical load of 2.9 J/m^2, the crack closes up on minimisation. This indicates that this load is less than the lower lattice trapping threshold G_-. Above this load, the 5–7 crack tip reconstruction forms spontaneously. It is therefore not possible to carry out the force integration, since there is no local minimum at the unreconstructed state. Fig. 7.9 shows a detail of the crack tip configuration obtained at the lowest load which prevents the crack closing up, $G = 2.1$ J/m^2.

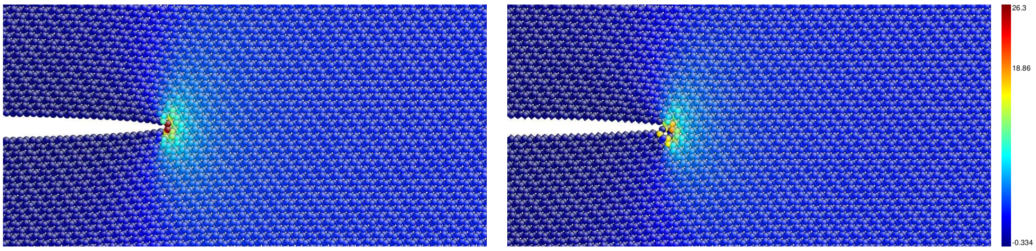
Stress Relief due to Reconstruction

Although it has not been possible to quantify the energy difference between the unreconstructed and reconstructed states at an *ab initio* level of detail, we can compare the hybrid relaxed configuration illustrated in Fig. 7.9 with a classically optimised geometry at the same load. Since we know the latter agrees very well with the continuum predictions of linear elasticity (Section 6.5), this is equivalent to comparing the quantum mechanical predictions to those of continuum theory. Fig. 7.10 shows stress maps comparing the classical and quantum mechanical atomistic stress distributions at a loading of $G = 2.1$ J/m^2. The effect of the reconstruction on the resulting stress distribution is considerable. In particular the planar projection of σ_{yy} in panel (g) illustrates that, far in front of the crack, the σ_{yy} fields match, but closer to the tip the reconstruction gives rise to significant stress relaxation.

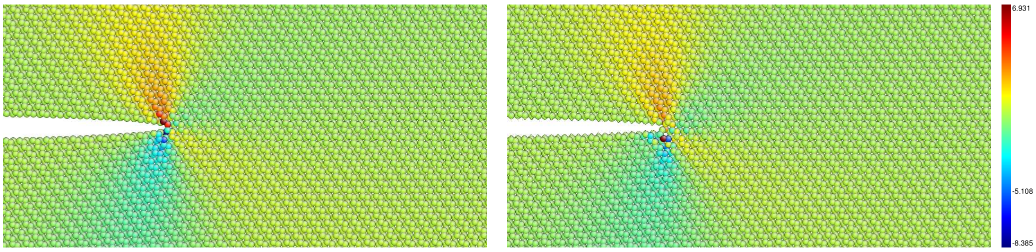
To see this more clearly, I have averaged the classical and QM σ_{yy} distributions onto a 20×20 Å^2 grid. The contour maps in Fig. 7.11 illustrate these data. It can be seen that the reconstruction causes a stress relief over a very large region: the contour labelled zero at which the fields become equal is almost 400 Å in front of the crack tip. At this loading the elastic energy density is 60 meV/atom. These calculations indicate that the formation of the reconstruction results in an average reduction in energy of 0.3 meV/atom for around 10 000 atoms, giving an energy gain of 3 eV. This is sufficient to overcome the local energy cost of creating the reconstruction, which the cluster calculations reported in Table 7.1 suggest to be slightly over 1 eV.



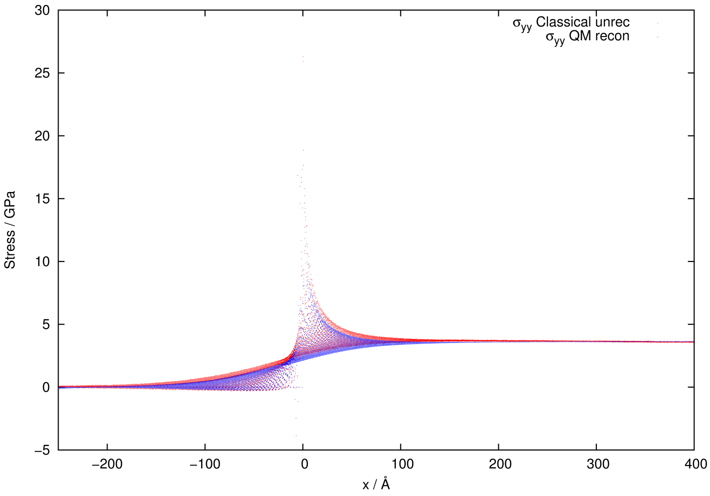
(a) Classical σ_{xx} (b) QM σ_{xx}



(c) Classical σ_{yy} (d) QM σ_{yy}

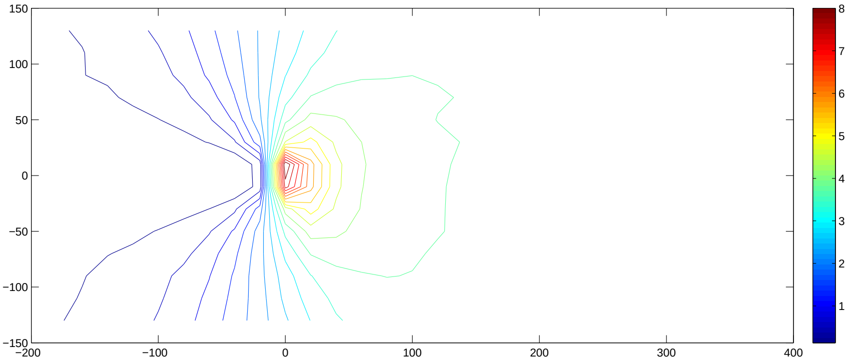


(e) Classical σ_{xy} (f) QM σ_{xy}

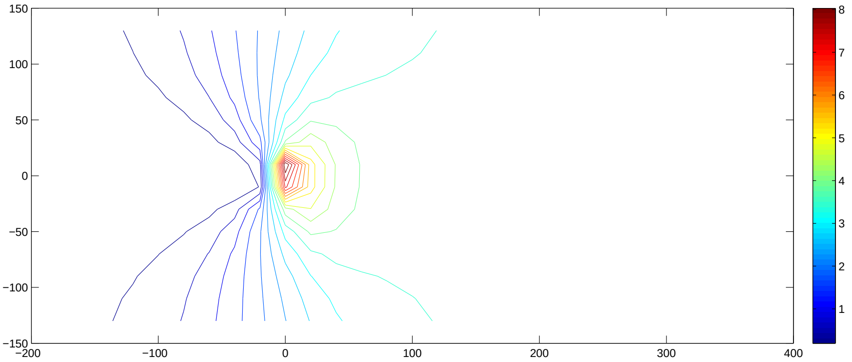


(g)

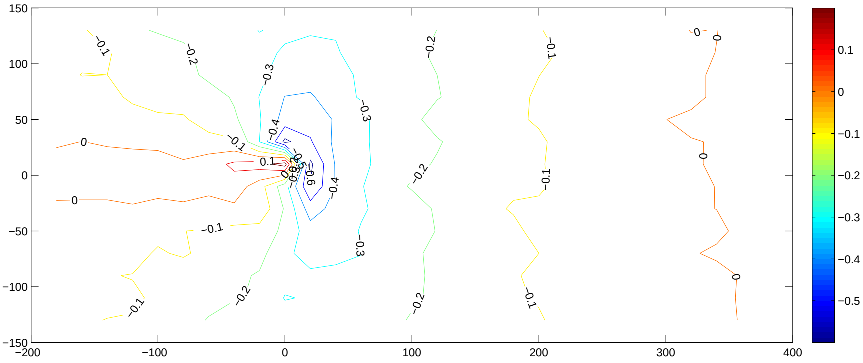
Figure 7.10: (a)–(f) Comparison of classical and quantum mechanical stress fields for relaxed configurations at $G = 2.1$ J/m^2. The 5–7 crack tip reconstruction is present in the hybrid structure. (g) Planar projection of σ_{yy} with classical stress shown in red and hybrid in blue. Stress units are GPa throughout.



(a) Classical



(b) Quantum Mechanical



(c) Difference (QM - Classical)

Figure 7.11: Contour maps of the distribution of σ_{yy} at $G = 2.1$ J/m^2 for (a) classically relaxed and (b) LOTF hybrid relaxation with 5–7 crack tip reconstruction. Panel (c) is a difference plot of the stress maps in (a) and (b), showing that the reconstructions relieve stress as far as ~ 400 Å from the crack tip. Stress units are GPa. Vertical axis is inverted with respect to the atomistic views that appear elsewhere in this chapter.

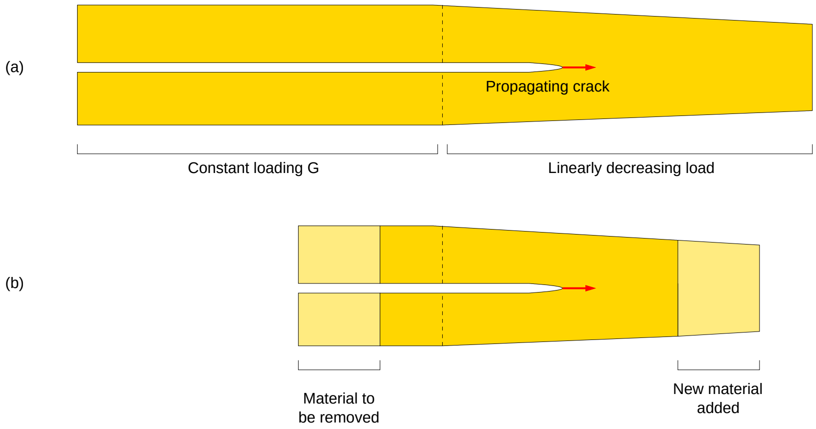


Figure 7.12: Possible approach to measure the fracture toughness by unloading a moving crack. (a) The entire cell must be sufficiently long for the crack to accelerate to its terminal velocity, and then the load can be slowly decreased until the crack arrests. (b) Suggested implementation in which material is removed from behind the crack and added in front as it propagates.

Measuring Fracture Toughness

The $T_{\mathrm{MD}} = 5$ ps simulation is at the upper limit of what is computationally feasible to run,[2] so it is not possible to increase this time far enough to converge the fracture toughness observed in simulation with that measured in experiment. This is a consequence of the large lattice trapping associated with the crack tip reconstruction. Obtaining the fracture toughness for a number of relaxation times up to about 20 ps and extrapolating to $T_{\mathrm{MD}} = \infty$ is possible but would be extremely computationally expensive and is likely to be inaccurate. An alternative approach would be to gradually unload a running crack and see when it stops propagating. This would require a simulation cell long enough for the crack to accelerate to its terminal velocity and then equilibrate, before gradually decreasing the applied loading, perhaps in the manner shown in Fig. 7.12a. This would require major changes to the code to introduce data parallelisation, since the 90 000 atom simulations performed in this work occupy almost all of the memory available on each node of a typical high performance computer.[3] A possible solution, developed for classical fracture simulations by Holland and Marder [113], would be to periodically remove material from behind the crack and add new material in front of it, as depicted in Fig. 7.12b. This scheme is complicated slightly by the requirement to pre-equilibrate the new material. Such a simulation would need to run for several hundreds of picoseconds,

---

[2] The 22 ps simulation conducted required just over 67 000 CPU hours

[3] This is 2 GB per CPU on the Cambridge high performance cluster *Darwin*

far longer than the fracture simulations presented in this work which have all been halted shortly after the initiation of fracture, with no attempt to allow the cracks to accelerate to their final velocities.

7.5 Summary

The central results of this thesis have been presented in this chapter. Applying the ‘Learn on the Fly’ technique to the brittle fracture of Si(111) reveals a previously unknown low speed instability caused by a crack tip reconstruction. The simulations predict that low speed propagation should be unstable on the (111) plane, conventionally regarded as the most stable cleavage plane. In the following chapter I will discuss the low speed instability induced by this crack tip reconstruction in more detail, in light of recent experimental results.

Chapter 8

Discussion of the Si(111) Low Speed Fracture Instability

8.1 Introduction

The hybrid fracture simulations described in the previous chapter and summarised in Fig. 8.1 predict systematic deflection of the crack path from the (111) plane at low velocities, with fast cracks dynamically steered away from this instability. The deflection process may start with a reconstruction event induced by the slowing down of the crack front segment due to a local velocity fluctuation or an encounter with a crystal defect. Irrespective of how the process starts, the initial reconstruction will initiate a positive feedback 'sinking' mechanism, since each reconstruction event will further trap and slow down the crack, leading to ever more reconstructions.

The reconstruction observed in hybrid simulations is associated with an atomistic length scale of around 1 Å, whereas AFM imaging cannot resolve fracture surfaces at a resolution much above 0.1 μm. In order to connect simulations on the nanoscale with experiment, a mesoscopic model is required to span the three orders of magnitude that separates these length scales. In the following section, a model is discussed which predicts observable surface features to arise from the microscopic instability.

Recently this instability has been observed experimentally for the first time at a range of low speeds, using a variety of novel techniques. I shall discuss these experiments and compare the observations of the low speed instability with predictions made with hybrid simulations and with the mesoscopic model.

Whilst the low speed instability predicted by my results has not been previously

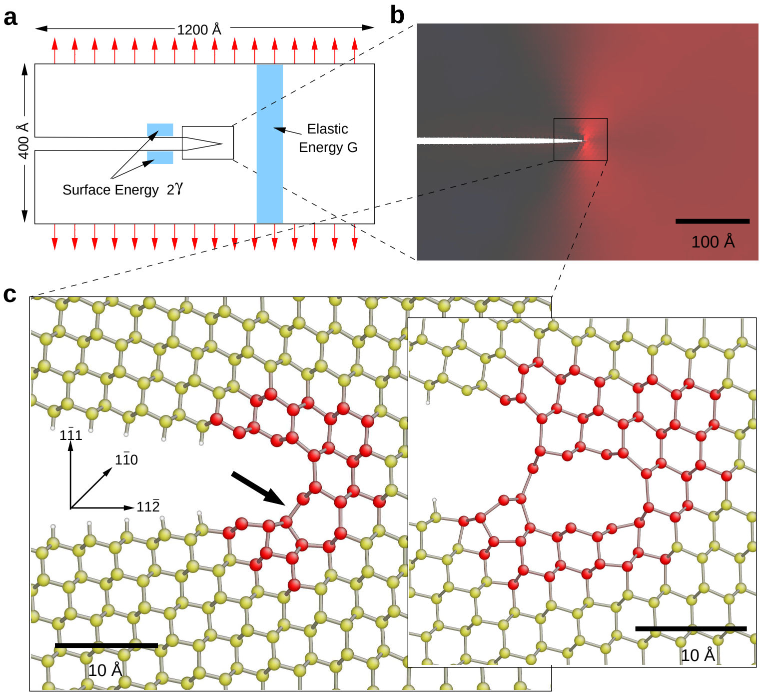


Figure 8.1: Overview of crack tip reconstruction results. (a) Thin strip geometry of the simulation. (b) Map of the σ_{yy} stress component near the crack tip, with red corresponding to highly stressed and black to unstressed. (c) Ground state reconstruction geometry of the crack tip under load, created by a rotation of the bond indicated by the arrow. The first few bonds that break after the reconstruction are in 6–membered rings adjacent to the 7–member ring of the reconstruction. The inset shows the crack tip structure after reconstruction and subsequent bond breaking. Further increasing the load finally breaks the bridging bond and advances the open crack surface yielding a downward step in the crack plane.

reported, a very recent study by Buehler *et al.* [131] using the ReaxFF potential to describe atoms near the crack tip and the Tersoff classical potential elsewhere also exhibits a 5–7 crack tip reconstruction. They interpret the reconstruction as a high speed instability that only arises at speeds above $\frac{1}{3}c_R$, and attempt to use it to explain the origin of the velocity gap. This is contrary to my results which show the reconstruction to be the equilibrium crack tip configuration for any load for which the crack does not close up. The paper by Buehler *et al.* gives me a number of causes for concern. Firstly the choice of simulation cell is unconventional, consisting of an infinite array of periodic surface cracks, making comparison with experiment difficult. Secondly, all data are rescaled relative to the observed molecular dynamics critical load $G_{0,\mathrm{MD}}$ making it impossible to determine the extent of lattice trapping.

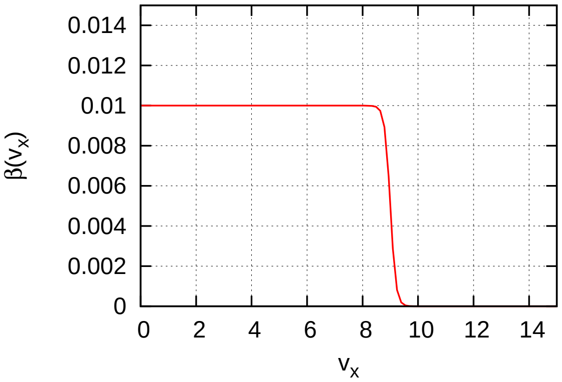


Figure 8.2: Rate of crack tip reconstruction $\beta(v_x)$ in the mesoscopic model as a function of velocity. Parameter values are $v_c = 10.0$, $A = 0.01$, $B = 0.9$ and $C = 0.01$. The rate of reconstruction rises smoothly for $v_x < v_c$.

Finally it is unclear whether the observed fracture is really brittle since there are many unexplained surface features.

8.2 Mesoscopic Model of Crack Tip Reconstruction

A mesoscopic model of the crack tip reconstruction process in which the crack front is represented by a line of particles connected by springs was developed in collaboration with Noam Bernstein and Gábor Csányi. The springs provide line tension and the crack front initially accelerates under a constant load until a terminal speed v_m, set by a velocity-dependant drag force, is reached. The instability is initiated by a perturbation which slows a point on the crack front below a critical speed v_c, with $v_c < v_m$. In the macroscopic system, this instability could be initiated by fluctuations in crack speed or by an encounter with a defect. Below this critical speed, crack tip reconstruction occurs at a velocity-dependant rate $\beta(v_x)$, with v_x the horizontal velocity component, which rises smoothly as the speed of the crack decreases further below v_c. $\beta(v_x)$ is defined by

$$\beta(v_x) = A\left[1 + \exp\left(\frac{v_x - Bv_c}{Cv_c}\right)\right]^{-1} \tag{8.1}$$

where A, B and C are parameters. Fig. 8.2 illustrates $\beta(v_x)$ with the chosen model parameters.

The dynamics of the particles in the mesoscopic model are obtained by inte-

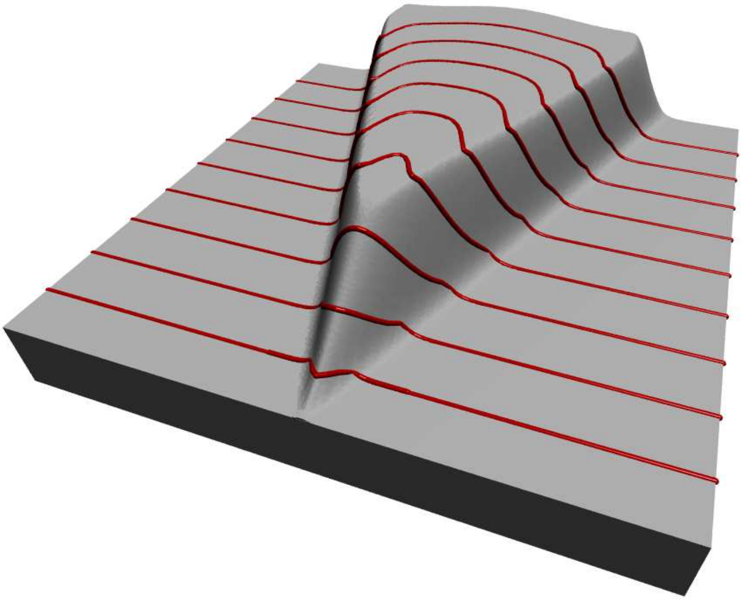


Figure 8.3: Evolution of the crack front with the ridge shape as predicted by a mesoscopic model based on the atomistic reconstruction mechanism. The red lines represent snapshots of the propagating crack front yielding the interpolated crack surface.

grating the equations of motion numerically using the Verlet algorithm. The model incorporates the observation that the crack front sinks upon reconstruction by modifying the vertical positions at each time step according to the equation

$$z(t + \Delta t) = z(t) + \beta\left(v_x(t)\right)\Delta x \tag{8.2}$$

Slowing of the crack upon sinking due to lattice trapping induced by the reconstruction is described by the equation

$$v_x(t + \Delta t) = v_x(t)\left[1 - \beta(v_x(t))\right] \tag{8.3}$$

Finally, stress enhancement is included in the model by increasing the line tension on the parts of the crack front left behind by reconstruction. Fig. 8.3 shows the evolution of the crack front predicted by this mesoscopic model. The crack tip reconstruction and ensuing feedback process cause the atomistic perturbation to evolve into a macroscopic ridge.

8.3 Single Crystal Experiments

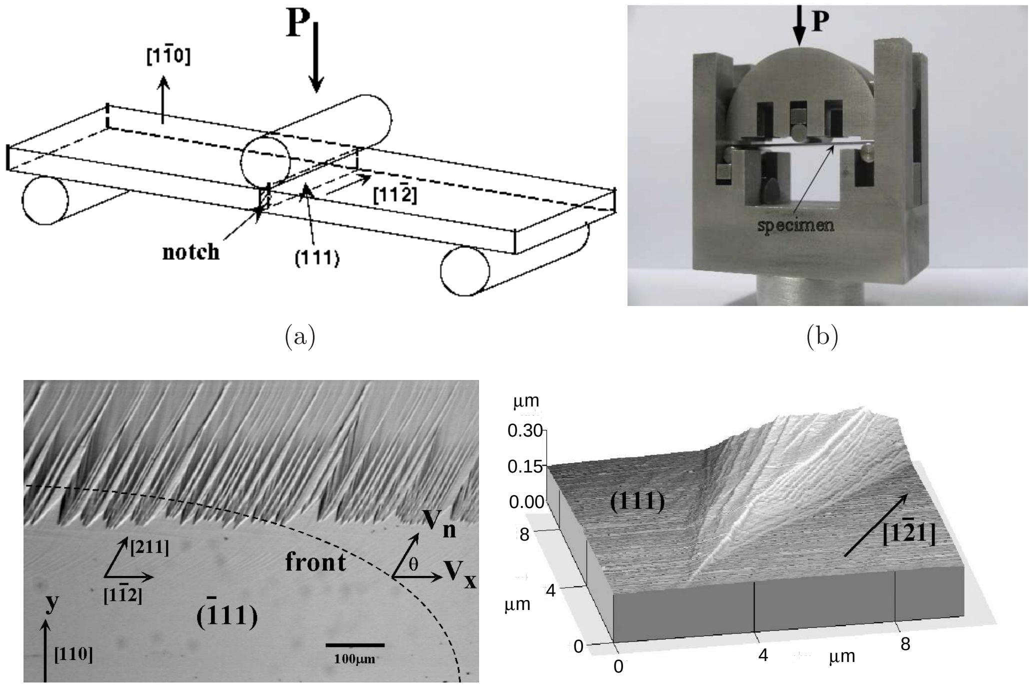
Following discussions with Peter Gumbsch concerning the results of the hybrid simulations, a collaboration with an experimental group lead by Dov Sherman was initiated. In this section I report recent experiments they have carried out in the low crack-speed regime using two different loading configurations, both of which exhibit the low speed instability predicted by my hybrid simulations and the mesoscopic model described in the previous section.

Three point bending

Sherman *et al.* [45] have very recently reported experimental studies using the three bond bending loading configuration shown in Figs. 8.4a and 8.4b. Cleaving of 525 μm thick silicon specimens of 20×43 mm^2 lateral dimensions was performed under three point bending. The specimens were notched with a 150 μm thick diamond saw to a length of 0.1–1.5 mm.

With reference to Fig. 8.4c, the crack initiates at the bottom left point of the specimen. It starts to propagate as a quarter circular crack, becoming elliptical as it reaches the steady state. A wide range of speeds are realised along the curved crack front, from thousands of ms^{-1} at the bottom surface to several tens of ms^{-1} at the top portion of the fracture surface [155]. The curved crack front (dashed black line in the figure) propagates from left to right so that the normal crack speed V_n becomes progressively smaller in higher regions of the sample. Below a critical speed of about 1000 ms^{-1} instabilities develop, which are seen as ridges coming out of the crack plane for y larger than a critical value. Fig. 8.4d shows an AFM micrograph of one of these triangular ridges. The crystallographic direction of the deviation is the same as the reconstruction-induced steps in the hybrid atomistic simulation described in the previous chapter, and the shape of the ridges is qualitatively in agreement with the results of the mesoscopic model depicted in Fig. 8.3.

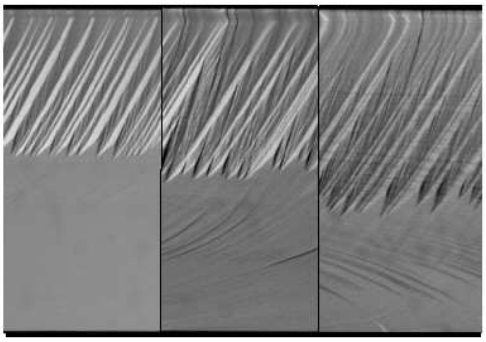
In this loading geometry, varying the initial crack length changes the critical loading and hence the crack speed. Fig. 8.5 shows that when this is done, the instability onsets at a different vertical distance y up the specimen, and analysis shows that this occurs in just such a way that the onset corresponds to a crack speed of approximately 1000 ms^{-1}.



(a) (b)

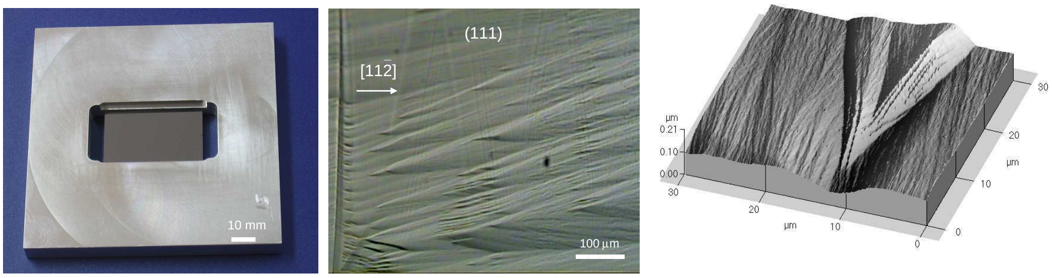
(c) (d)

Figure 8.4: Three point bending experimental results, showing ridges formed by low-speed instabilities on the (111) crack plane. Panels (a) and (b) show schematic and photographic views of the apparatus respectively. P is the applied load. (c) optical micrograph of the (111) crack system. (d) AFM micrograph close-up of a (111) ridge. All images from Dov Sherman.



a	0.15 mm	0.5 mm	0.7 mm
V_x	3440 m/sec	3100 m/sec	2280 m/sec
y/h	0.528	0.476	0.354
V_{nc}	1000 m/sec	1050 m/sec	1085 m/sec

Figure 8.5: Optical micrographs showing onset of low speed instability for three different initial crack lengths. In each case the instability onsets at a crack speed of roughly 1000 ms^{-1}. Image from Dov Sherman.



(a) (b) (c)

Figure 8.6: Tensile loading experimental results, showing ridges formed by low-speed instabilities on the (111) crack plane. (a) The fracture apparatus, showing the silicon specimen glued to the aluminium frame (b) optical micrograph of (111) crack system. The ridges initiate at a variety of crack speeds. (c) AFM micrograph close up of one of the ridges. All images from Dov Sherman.

Tensile Loading

Even more recently, Sherman *et. al.* have carried out tensile studies which also show the low speed instability. Cleavage of 525 µm thick silicon specimens of 42×26 mm^2 lateral dimensions, cut from a [110] silicon wafer, was performed using a novel technique to apply very small but steady and well controlled tensile loads. The specimens were notched with a 150 µm thick diamond saw to a length of 1.5 mm, heated to 200°C and quenched in water to generate a pre-crack of the required length. They were then glued to an aluminium loading frame using a 150 µm thick film of epoxy resin. The load is applied by taking advantage of the thermal expansivity mismatch between the sample and the aluminium loading frame (see Fig. 8.6a). Micrographs of the resulting (111) fracture surface are shown in Figs 8.6b and 8.6c. The crack advances from left to right as it accelerates up to a speed of about 800 ms^{-1}, as determined from the faintly visible Wallner lines [156].

Triangular ridges, all deviating in the same direction from the fracture surface, form at a range of low crack speeds from about 400 to about 800 ms^{-1}. At higher speeds, about 2000 ms^{-1}, the surface is mirror smooth and no ridges are present. As in the three point bending experiments, the crystallographic direction of the deviation matches that predicted by the instabilities produced by the atomistic crack tip reconstruction, and the shape of the ridges is qualitatively in agreement with the mesoscopic model.

8.4 Summary

In this chapter I have discussed the macroscopic consequences of the reconstruction-induced low speed instability on the (111) cleavage plane of silicon reported in the preceeding chapter. The hybrid simulation results are used to construct a mesoscopic model in which the crack tip reconstruction leads, via a positive feedback mechanism, to observable surface features.

Ridges in exactly the same crystallographic direction and of qualitatively the same shape as those predicted by the mesoscopic model have very recently been observed in low crack-speed fracture experiments. The results of the tensile loading experiments described above and the (111) cleavage results presented in this thesis have been written up as in a joint experimental and theoretical paper and submitted for publication [46].

Chapter 9

Conclusions

There are many avenues open to extend the work carried out in this thesis, both within silicon and to other materials. In this section I give an overview of these problems, for some of which work is already underway.

The crack tip reconstruction observed in hybrid simulations suggest significant enhancement of the lattice trapping barrier to brittle fracture. Coupled with the fundamental limitations on the timescale accessible to molecular dynamics simulations, this large barrier means that we cannot hope to measure the fracture toughness, and hence the extent of this lattice trapping, directly. An indirect approach based on unloading a running crack has been proposed in this work. Extending the inquiry to look at other cleavage orientations in silicon would be straightforward. In particular, it will be interesting to see if the crack tip reconstruction arises for cracks running on the (111) plane in the perpendicular $[00\bar{1}]$ direction.

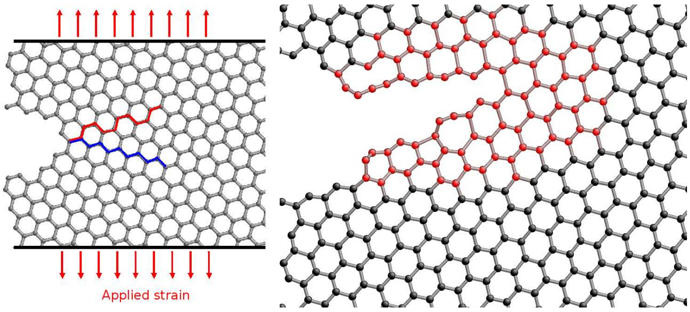
Silicon undergoes a brittle to ductile transition at around 800 K, above which fracture proceeds by dislocation emission. It has been argued that the lattice trapping barrier plays an important rôle in this transition, via an interplay between the barriers to fracture and to dislocation nucleation and migration [108]. Further simulations to investigate the temperature dependence of the lattice trapping barrier could provide insight on the origin of the brittle to ductile transition.

Using *ab initio* modelling at the crack tip removes all empirical assumptions concerning the chemical processes taking place there. Extending the method to model more complex chemical environments therefore requires only the addition of a classical potential which provides a reasonable description of the new species. An example application is stress corrosion cracking, where samples unexpectedly fail when subjected to constant or cyclic tensile loadings below the nominal fracture toughness, due to the effect of a corrosive environment, typically consisting of water and oxygen. Stress corrosion induced cracks at the Si/SiO_2 interface in nanoscale electronic devices are of particular technological importance. This could be modelled using a combination of *ab initio* accuracy at the crack tip and a recently developed classical potential which reproduces the structure of the natively oxidised silicon surface and its interaction with water solutions [157].

The cleavage of diamond shows a strong anisotropy between different cleavage planes, with (111) strongly preferred. Existing calculations have not been able to explain this anisotropy, since there is a large barrier to the formation of the 2×1 Pandey reconstructed surface from the unreconstructed (111) surface. The existence of a crack tip reconstruction similar to that reported in this thesis could explain how this barrier is overcome. A lattice trapping model based on *ab initio* calculations has been proposed by Ruben Perez and Pablo Pou Bell, but the boundary conditions used preclude the formation of a crack tip reconstruction. In diamond, the possibility for sp^2 hybridisation and graphitisation hugely extends the space of possible crack tip morphologies, making the necessity of multiscale modelling even stronger.

Finally, hybrid simulations are currently underway to study the fracture of graphene, a truly two dimensional form of carbon. It behaves as a metal, with a very complex electronic structure, in which electronic edge states are thought to be very important [158]. These states are strongly dependent on the geometry at the edge of the sheet, which can be either armchair, zigzag or a mixture of the two [159]. Graphene is produced by tearing sections from the top layer of a graphite sample; an open question is whether the fracture that forms the edges of these sheets results in a preference for either armchair or zigzag edges. Surface energy calculations indicate a preference for armchair edges, but STM experiments are as yet inconclusive [159]. Embedding a quantum mechanical description of the crack tip within a classical reparameterised to match the *ab initio* elastic constants allows the fracture surface morphology to be investigated. The preliminary results illustrated in Fig. 9.1 suggest a preference for the formation of zig-zag edges.

The outlook for hybrid simulation of materials systems is bright, with no shortage of potential applications. In biology, where such methods are standard since they provide a way to understand the underlying physical mechanisms behind processes such as enzyme catalysis. In the years to come, the LOTF scheme has great promise in this arena, where traditional QM/MM is currently dominant despite the boundary problems associated with it, but there are many further difficulties to be overcome



(a) (b)

Figure 9.1: Preliminary simulation of graphene fracture. (a) The starting point is a sheet of graphene with a notch defect, oriented mid-way between zig-zag (blue) and armchair (red) edges. (b) Snapshot from a hybrid fracture simulation, where the red atoms are treated with CASTEP and the black atoms with the Brenner potential, refitted to match the *ab initio* elastic constants. The opened surface corresponds to a zig-zag edge, at a steep angle.

in applying the technique once the comparative simplicity of covalently bonded crystalline systems of the type studied in this thesis is left behind.

In conclusion, this thesis has described the application of hybrid quantum mechanical and classical molecular dynamics simulations to the brittle fracture of silicon on the (111) cleavage plane. The conditions for crack propagation are created by the concentration of a long-range stress field at an atomically sharp crack tip, creating a complex and strongly coupled multiscale system. The simulations are made possible by combining a quantum mechanical description of the processes taking place near to the crack tip with a classical atomistic model to capture the long-range elastic relaxation. In this way, direct simulation of the fracture process at the DFT level of accuracy is possible.

Hybrid simulations predict (111) fracture to be unstable at low speeds due to a crack tip reconstruction which triggers a ‘sinking’ mechanism leading to macroscopic, experimentally observable, corrugations. Recent experiments have observed surface features consistent with these predictions. For crystalline materials, this is the first example of a fracture instability that sets in only below a critical velocity. Subtle atomistic details at the crack tip control the qualitative macroscopic fracture behaviour. It is suggested that this could be the case in general for brittle fracture in crystalline materials, and that all covalent brittle crystals in all cleavage orientations should be investigated to determine whether similar crack tip reconstructions are present. This thesis is only the start of this process.

Appendix A

Linear Elastic Thin Strip Solution

In this Appendix, I first present a brief sketch of Knauss' approach to the linear elastic problem of a semi-infinite crack in an infinite strip under fixed grip boundary conditions; for full details consult Ref. 119. I then describe the approach I have taken to numerically integrate Knauss' solution.

The boundary conditions used by Knauss are the same as in the discussion of Section 5.2: the top and bottom edges are clamped and displaced vertically by an amount δ. We begin by subtracting the field of a uniformly stressed strip:

$$\sigma^0_{yy} = E'\delta \tag{A.1}$$

$$\sigma^0_{xx} = \nu E'\delta \tag{A.2}$$

$$\sigma^0_{xy} = 0 \tag{A.3}$$

where as usual E' is the effective Young's modulus of the material (this differs depending on whether plane stress or plane strain conditions apply) and ν is Poisson's ratio. Knauss chooses to express the problem in terms of the two-dimensional Airy stress function $\phi(x, y)$, which is related to the stresses by

$$\sigma_{yy} = \frac{\partial^2\phi}{\partial x^2} \tag{A.4}$$

$$\sigma_{xx} = \frac{\partial^2\phi}{\partial y^2} \tag{A.5}$$

$$\sigma_{xy} = -\frac{\partial^2\phi}{\partial x\partial y} \tag{A.6}$$

The linear elastic stress-strain relations can be re-expressed in terms of ϕ providing the compatibility equation is satisfied:

$$\nabla^4\phi = \frac{\partial^4\phi}{\partial x^4} + 2\frac{\partial^4\phi}{\partial x^2\partial y^2} + \frac{\partial^4\phi}{\partial y^4} = 0 \tag{A.7}$$

The problem reduces to integrating this equation subject to the appropriate boundary conditions. We can do this by Fourier transforming the equation and boundary conditions with respect to x, taking ω as the conjugate variable. This allows all the fields to be expressed in terms of the transformed stress $\hat{\sigma}_+(\omega)$ along the line ahead of the crack tip and the transformed displacement $\hat{v}_-(\omega)$ along the line behind the crack tip. These functions are shown to be related by

$$\hat{\sigma}_+(\omega) + E'F(\omega)\hat{v}_-(\omega) = \frac{1}{i\omega}E'\delta \tag{A.8}$$

where $F(\omega)$ is a known function. Knauss then uses the Wiener-Hopf technique to write $F(\omega)$ as a product of the two functions $F_+(\omega)$ and $F_-(\omega)$, where F_+ has neither poles nor zeroes in a upper half plane and F_- has neither poles nor zeroes in a lower half plane. These requirements can be satisfied by choosing

$$F_-(\omega) = F_+(-\omega) = \frac{1-\frac{\omega}{\omega_0}}{1-\frac{\omega}{z_0}} \prod_{n=1}^{\infty} \frac{\left(1-\frac{\omega}{\omega_n}\right)\left(1+\frac{\omega}{\bar{\omega}_n}\right)}{\left(1-\frac{\omega}{z_n}\right)\left(1+\frac{\omega}{\bar{z}_n}\right)} \tag{A.9}$$

where the complex numbers ω_n and z_n are the roots of the equations

$$(3-\nu)\sinh^2\omega_n + (1+\nu)\omega_n^2 + \frac{4}{1+\nu} = 0 \tag{A.10}$$

$$(3-\nu)\sinh z_n \cosh z_n - (1+\nu)\, z_n = 0 \tag{A.11}$$

Eq. A.8 can now be rewritten

$$\frac{i\omega\hat{\sigma}_+(\omega) - E'\delta}{F_+(\omega)} = -i\omega E'F_-(\omega)\hat{v}_-(\omega). \tag{A.12}$$

By considering the asymptotic behaviour of $\hat{v}_-$ and σ_+ required to obey the boundary conditions, the half-planes in which F_+ and F_- are analytic can be shown to overlap. Appealing to Liouville's theorem proves that the two sides of Eq. A.12 must be equal to a constant in the overlap strip. This argument yields

$$\hat{\sigma_+}(\omega) = \frac{iE'\delta\,(F_+(\omega)-1)}{\omega}. \tag{A.13}$$

The stress distribution along the line $y = 0$ can be obtained by taking the inverse Fourier transform of this equation and adding the contribution from the uniform field:

$$\sigma_{yy}(x,0) = \frac{1}{2\pi}\int_{\gamma} \hat{\sigma}_{+}(\omega) e^{-i\omega x} \mathrm{d}\omega + \sigma_{yy}^{0}. \tag{A.14}$$

The integration contour γ runs along the real axis from $-\infty$ to ∞ with a semi-circular deformation of radius ϵ around the origin to avoid the singularity in $\hat{\sigma}_{+}$.

Numerical Evaluation of $\sigma_{yy}(x,0)$

In principle, Knauss' method can be used to evaluate all stress components at any position in the strip by the application of Eqs. A.4– A.6, but here I consider only $\sigma_{yy}(x,0)$. This calculation is sufficient to enable useful comparisons with my atomistic calculations.

The roots of Eqs. A.10 and A.11 must be found numerically, and the infinite product of Eq. A.9 has to be truncated after a finite number of terms, denoted here by N. The integral can then be evaluated by breaking it down into a sum of contributions:

$$\sigma_{yy}(x,0) = \frac{1}{2\pi}\left\{\int_{-\infty}^{-R} + \int_{-R}^{\epsilon} + \int_{C} + \int_{\epsilon}^{R} + \int_{R}^{\infty}\right\} + \sigma_{yy}^{0} \tag{A.15}$$

where the contour C represents a semi-circle of radius ϵ around the origin in the lower half-plane. We denote the five integrals by the symbols I_1, I_2, I_3, I_4 and I_5 respectively. Let's consider I_5 first. Knauss gives an asymptotic form for the function $F_{+}(\omega)$:

$$\lim_{\omega\to\infty} F_{+}(\omega) = \frac{\theta}{\Gamma^{2}\left(\frac{5}{4}\right)}\left(\frac{i\omega}{e\pi}\right)^{\frac{1}{2}} \tag{A.16}$$

where θ is a constant that depends on the ω_n and z_n. In this limit F_{+} behaves like $A\sqrt{\omega}$. Substituting this asymptotic form into Eq. A.15 and transforming $x \to x - i\eta$ to ensure that the integrand tends to zero as ω tends to infinity allows us to write

$$I_5(x,R) = \lim_{\eta\to 0+} \frac{iE'}{2\pi}\int_{R}^{\infty} \frac{(A\sqrt{\omega}-1)}{\omega} e^{-i\omega x-\omega\eta}\mathrm{d}\omega \tag{A.17}$$

This expression can be integrated analytically: the first term can be transformed into a Gaussian integral by the substitution $s = \sqrt{\omega}$, and the second can be expressed in terms of the incomplete Gamma function. Taking the limit $\eta \to 0$, we recover the value of the original integral:

$$I_5(x,R) = \frac{E'\delta}{2\pi}\left\{\frac{\theta\,(1+i)}{\Gamma^2\left(\frac{5}{4}\right)\sqrt{2eix}}\left(1-\operatorname{erf}\sqrt{-iRx}\right) - i\Gamma(0,iRx)\right\} \tag{A.18}$$

where the error function $\operatorname{erf}(z)$ and the incomplete Gamma function $\Gamma(a,z)$ are defined as follows and can be evaluated using series representations.

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-s^2}\mathrm{d}s \tag{A.19}$$

$$\Gamma(a,z) = \int_z^\infty s^{a-1}e^{-s}\mathrm{d}s \tag{A.20}$$

Repeating this procedure for I_1 with the substitution $x \to x+i\eta$ so that the integrand vanishes as $x \to -\infty$ yields the simple result

$$I_1(x,R) = I_5(x,R)^* \tag{A.21}$$

I_2 and I_4 have to be evaluated numerically, but again we can take advantage of symmetry to show that

$$I_2(x,\epsilon,R) = I_4(x,\epsilon,R)^* \tag{A.22}$$

and so only one of the integrals needs to be evaluated. I_3 can either be determined numerically or by considering the contribution made by the residue of $\hat{\sigma}_+$ at the origin. From a practical point of view this term is not significant to the overall value of the integral. Putting all these results together, we have:

$$\sigma_{yy}(x,0) = 2\,\mathbb{R}\mathrm{e}\left[I_1(x,R)\right] + 2\,\mathbb{R}\mathrm{e}\left[I_2(x,\epsilon,R)\right] + I_3(x,\epsilon) + \sigma^0_{yy} \tag{A.23}$$

The symbol $\mathbb{R}\mathrm{e}$ denotes taking the real part; we make use of Eqs. A.21 and A.22 to reduce the numerical computation required. The result has to be converged with respect to the number of terms N used to approximate the infinite product in F_+, and the parameters ϵ and R. Since the contribution made by I_3 is negligible, the final stress is relatively insensitive to the choice of ϵ. R can be chosen to reduce the contribution of the asymptotic integrals I_1 and I_5 below some threshold. Good results have been achieved with $N = 100$, $\epsilon = 0.1$ and $R = 10^4$.

Bibliography

[1] J. M. Haile. *Molecular Dynamics Simulation: Elementary Methods* (Wiley, New York, USA, 1992).

[2] D. C. Rapaport. *The Art of Molecular Dynamics Simulation — Second Edition* (Camb. Univ. Press, Cambridge, UK, 2004).

[3] L. Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, **159**, 98 (1967).

[4] W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, **76**, 637 (1982).

[5] J. Lennard-Jones. The Determination of Molecular Fields. I. From the Variation of the Viscosity of a Gas with Temperature. *Proc. Roy. Soc. (Lond.)*, **106A**, 441 (1924).

[6] F. H. Stillinger and T. A. Weber. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B.*, **31**, 5262 (1985).

[7] J. Tersoff. New empirical model for the structural properties of silicon. *Phys. Rev. Lett.*, **56**, 632 (1986).

[8] J. Tersoff. New empirical approach for the structure and energy of covalent systems. *Phys. Rev. B*, **37**, 6991 (1988).

[9] J. Tersoff. Empirical interatomic potential for silicon with improved elastic properties. *Phys. Rev. B*, **38**, 9902 (1988).

[10] D. W. Brenner. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. *Phys. Rev. B*, **42**, 9458 (1990).

[11] D. Brenner. The art and science of an analytic potential. *Phys. Stat. Sol. (b)*, **217**, 23 (2000).

[12] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni and S. B. Sinnott. A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. *J. Phys.: Cond. Mat.*, **14**, 783 (2002).

[13] M. Z. Bazant, E. Kaxiras and J. F. Justo. Environment-dependent interatomic potential for bulk silicon. *Phys. Rev. B*, **56**, 8542 (1997).

[14] J. F. Justo, M. Z. Bazant, E. Kaxiras, V. V. Bulatov and S. Yip. Interatomic potential for silicon defects and disordered phases. *Phys. Rev. B*, **58**, 2539 (1998).

[15] N. A. Marks. Generalizing the environment-dependent interaction potential for carbon. *Phys. Rev. B*, **63**, 035401 (2000).

[16] A. van Duin, S. Dasgupta, F. Lorant and W. Goddard. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A*, **105**, 9396 (2001).

[17] A. P. Horsfield, A. M. Bratkovsky, M. Fearn, D. G. Pettifor and M. Aoki. Bond-order potentials: Theory and implementation. *Phys. Rev. B*, **53**, 12694 (1996).

[18] S. J. Stuart, A. B. Tutein and J. A. Harrison. A reactive potential for hydrocarbons with intermolecular interactions. *J. Chem. Phys.*, **112**, 6472 (2000).

[19] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, **136**, B864 (1964).

[20] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, **140**, A1133 (1965).

[21] J. P. Perdew, K. Burke and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, **77**, 3865 (1996).

[22] M. P. Teter, M. C. Payne and D. C. Allan. Solution of Schrödinger's equation for large systems. *Phys. Rev. B*, **40**, 12255 (1989).

[23] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias and J. D. Joannopoulos. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, **64**, 1045 (1992).

[24] M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark and M. C. Payne. First-principles simulation: ideas, illustrations and the CASTEP code. *J. Phys.: Cond Mat*, **14 (11)**, 2717 (2002).

[25] R. M. Nieminen. From atomistic simulation towards multiscale modelling of materials. *J. Phys.: Cond. Mat.*, **14**, 2859 (2002).

[26] S. Kohlhoff, P. Gumbsch and H. F. Fischmeister. Crack propagation in b.c.c. crystals studied with a combined finite-element and atomistic model. *Phil. Mag. A*, **64**, 851 (1991).

[27] E. B. Tadmor, R. Phillips and M. Ortiz. Mixed atomistic and continuum models of deformation in solids. *Langmuir*, **12**, 4529 (1996).

[28] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **103**, 227 (1976).

[29] F. Maseras and K. Morokuma. IMOMM: A new integrated *ab initio* + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.*, **16**, 1170 (1995).

[30] K. B. Broberg. *Cracks and Fracture* (Academic Press, San Diego, CA, 1999).

[31] L. B. Freund. *Dynamic Fracture Mechanics* (Cambridge Univ. Press, Cambridge, UK, 1990).

[32] C. E. Inglis. Stresses in a plate due to the prescence of cracks and sharp corners. *Trans. Inst. Naval. Archit.*, **55**, 219 (1913).

[33] A. A. Griffith. The Phenomena of Rupture and Flow in Solids. *Philos. Trans. R. Soc. London A*, **221**, 163 (1921).

[34] B. Lawn. *Fracture of Brittle Solids — Second Edition* (Cambridge Univ. Press, Cambridge, UK, 1993).

[35] I. H. Lin and R. Thomson. Cleavage, Dislocation Emission, and Shielding for Cracks Under General Loading. *Acta Metall.*, **34**, 187 (1986).

[36] G. R. Irwin. *Fracturing of Metals*, pages 147–166 (American Society for Metals, Cleveland, OH, 1948).

[37] J. C. H. Spence, Y. M. Huang and O. Sankey. Lattice trapping and surface reconstruction for silicon cleavage on (111). Ab-initio quantum molecular dynamics calculations. *Acta Metall. Mater.*, **41**, 2815 (1993).

[38] J. Q. Broughton, F. F. Abraham, N. Bernstein and E. Kaxiras. Concurrent coupling of length scales: Methodology and application. *Phys. Rev. B*, **60**, 2391 (1999).

[39] N. Bernstein. Linear scaling nonorthogonal tight-binding molecular dynamics for nonperiodic systems. *Europhys. Lett.*, **55**, 52 (2001).

[40] G. Csányi, T. Albaret, G. Moras, M. C. Payne and A. De Vita. Multiscale hybrid simulation methods for material systems. *J. Phys.: Cond Mat*, **17**, R691 (2005).

[41] A. De Vita and R. Car. *Mater. Res. Soc. Symp. Proc*, **491**, 473 (1998).

[42] G. Csányi, T. Albaret, M. C. Payne and A. De Vita. “Learn on the Fly”: A Hybrid Classical and Quantum-Mechanical Molecular Dynamics Simulation. *Phys. Rev. Lett.*, **93**, 175503 (2004).

[43] G. Moras, G. Csanyi, M. Payne and A. De Vita. A novel molecular dynamics approach to large semiconductor systems. *Physica B*, **376-377**, 936 (2006).

[44] G. Csányi, G. Moras, J. R. Kermode, M. C. Payne, A. Mainwood and A. de Vita. Multiscale modeling of defects in semiconductors: A novel molecular dynamics scheme. In D. A. Drabold and S. K. Estreicher, editors, *Theory of Defects in Semiconductors*, pages 193–212 (Springer, 2007).

[45] D. Sherman, M. Markovitz and O. Barkai. Dynamic instabilities in {111} silicon. *J. Mech. Phys. Solids* (2007). In press, doi:10.1016/j.jmps.2007.05.010.

[46] J. R. Kermode, T. Albaret, D. Sherman, N. Bernstein, P. Gumbsch, M. C. Payne, G. Csányi and A. D. Vita. Low speed fracture instabilities in a brittle crystal (2007). Submitted to *Nature*.

[47] C. G. Broyden. The convergence of a class of double-rank minimization algorithms. I: General considerations. *J. Inst. Maths. Applics.*, **6**, 76 (1970).

[48] M. I. J. Probert. Improved algorithm for geometry optimisation using damped molecular dynamics. *J. Comput. Phys.*, **191**, 130 (2003).

[49] E. Bitzek, P. Koskinen, F. Gahler, M. Moseler and P. Gumbsch. Structural relaxation made simple. *Phys. Rev. Lett.*, **97**, 170201 (2006).

[50] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, **72**, 2384 (1980).

[51] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684 (1984).

[52] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, **81**, 511 (1984).

[53] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, **31**, 1695 (1985).

[54] S. A. Adelman and J. D. Doll. Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *J. Chem. Phys.*, **64**, 2375 (1976).

[55] D. Quigley and M. I. J. Probert. Langevin dynamics in constant pressure extended systems. *J. Chem. Phys.*, **120**, 11432 (2004).

[56] W. Brantley. Calculated elastic constants for stress problems associated with semiconductor devices. *J. Appl. Phys.*, **44**, 534 (1973).

[57] A. E. H. Love. *Mathematical Theory of Elasticity — Fourth edition* (Camb. Univ. Press, 1927).

[58] W. A. Harrison. *Electronic Structure and the Properties of Solids* (Dover, 1980).

[59] F. London. Properties and Applications of Molecular Forces. *Zeit. Physik. Chem. B*, **11**, 222 (1930).

[60] X.-P. Li, G. Chen, P. B. Allen and J. Q. Broughton. Energy and vibrational spectrum of the Si(111) (7×7) surface from empirical potentials. *Phys. Rev. B*, **38**, 3331 (1988).

[61] I. P. Batra, F. F. Abraham and S. Ciraci. Molecular-dynamics study of self-interstitials in silicon. *Phys. Rev. B*, **35**, 9552 (1987).

[62] T. W. Poon, S. Yip, P. S. Ho and F. F. Abraham. Equilibrium structures of Si(100) stepped surfaces. *Phys. Rev. Lett.*, **65**, 2161 (1990).

[63] M. Ishimaru, K. Yoshida and T. Motooka. Application of empirical interatomic potentials to liquid Si. *Phys. Rev. B*, **53**, 7176 (1996).

[64] E. R. Cowley. Lattice Dynamics of Silicon with Empirical Many-Body Potentials. *Phys. Rev. Lett.*, **60**, 2379 (1988).

[65] D. Kohen, J. C. Tully and F. H. Stillinger. Modeling the interaction of hydrogen with silicon surfaces. *Surface Science*, **397**, 225 (1998).

[66] S. Barnard and S. P. Russo. Development of an improved Stillinger-Weber potential for tetrahedral carbon using *ab initio* (Hartree-Fock and MP2) methods. *Molecular Physics*, **100**, 1517 (2002).

[67] P. C. Kelires and J. Tersoff. Glassy Quasithermal Distribution of Local Geometries and Defects in Quenched Amorphous Silicon. *Phys. Rev. Lett.*, **61**, 562 (1988).

[68] Z. H. Xia, P. R. Guduru and W. A. Curtin. Enhancing Mechanical Properties of Multiwall Carbon Nanotubes via sp^3 Interwall Bridging. *Phys. Rev. Lett.*, **98**, 245501 (2007).

[69] O. K. Steven J Stuart, M Todd Knippenberg and P. S. Krstic. Simulation of amorphous carbon with a bond-order potential. *Physica Scripta*, **T124**, 58 (2006).

[70] Y. Huang, J. Wu and K. C. Hwang. Thickness of graphene and single-wall carbon nanotubes. *Phys. Rev. B.*, **74**, 245413 (2006).

[71] P. Keblinski, M. Z. Bazant, R. K. Dash and M. M. Treacy. Thermodynamic behavior of a model covalent material described by the environment-dependent interatomic potential. *Phys. Rev. B*, **66**, 064104 (2002).

[72] C. L. Allred, X. Yuan, M. Z. Bazant and L. W. Hobbs. Elastic constants of defected and amorphous silicon with the environment-dependent interatomic potential. *Phys. Rev. B.*, **70**, 134113 (2004).

[73] P. Valentini and T. Dumitrica. Microscopic theory for nanoparticle-surface collisions in crystalline silicon. *Phys. Rev. B.*, **75**, 224106 (2007).

[74] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, **389**, 457 (1927).

[75] R. Renner. Zur Theorie der Wechselwirkung zwischen Elektronen- und Kernbewegung bei dreiatomigen, stabförmigen Molekülen. *Z. Phys.*, **92**, 172 (1934).

[76] C. Lee, W. Yang and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, **37**, 785 (1988).

[77] H. J. Monkhorst and J. D. Pack. Special points for Brillouin-zone integrations. *Phys. Rev. B*, **13**, 5188 (1976).

[78] N. Troullier and J. L. Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, **43**, 1993 (1991).

[79] D. Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, **41**, 7892 (1990).

[80] R. Car and M. Parrinello. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.*, **55**, 2471 (1985).

[81] H. Hellmann. Zur Rolle der kinetischen Elektronenenergie für die zwischenatomaren Kräfte. *Zeitschrift für Physik A Hadrons and Nuclei*, **85**, 180 (1933).

[82] R. P. Feynman. Forces in Molecules. *Phys. Rev.*, **56**, 340 (1939).

[83] J. Slater and G. Koster. Simplified LCAO Method for the Periodic Potential Problem. *Phys. Rev.*, **94**, 1498 (1954).

[84] C. M. Goringe, D. R. Bowler, and E. Hernández. Tight-binding modelling of materials. *Rep. Prog. Phys.*, **60**, 1447 (1997).

[85] D. J. Chadi. Theoretical study of the atomic structure of silicon (211), (311), and (331) surfaces. *Phys. Rev. B*, **29**, 785 (1984).

[86] L. Goodwin, A. Skinner and D. Pettifor. Generating Transferable Tight-Binding Parameters: Application to Silicon. *Europhys. Lett.*, **9**, 701 (1989).

[87] I. Kwon, R. Biswas, C. Wang, K. Ho and C. Soukoulis. Transferable tight-binding models for silicon. *Phys. Rev. B.*, **49**, 7242 (1994).

[88] D. R. Bowler, M. Fearn, C. M. Goringe, A. P. Horsfield and D. G. Pettifor. Hydrogen diffusion on Si(001) studied with the local density approximation and tight binding. *J. Phys.: Cond Mat*, **10**, 3719 (1998).

[89] M. Menon and K. R. Subbaswamy. Transferable nonorthogonal tight-binding scheme for silicon. *Phys. Rev. B*, **50**, 11577 (1994).

[90] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert and R. Kaschner. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B*, **51**, 12947 (1995).

[91] T. Frauenheim, G. Seifert, M. Elsterner, Z. Hajnal, G. Jungnickel, D. Porezag, S. Suhai and R. Scholz. A Self-Consistent Charge Density-Functional Based Tight-Binding Method for Predictive Materials Simulations in Physics, Chemistry and Biology. *Phys. Status Solidi B*, **217**, 41 (2000).

[92] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, **58**, 7260 (1998).

[93] R. Martoňák, C. Molteni and M. Parrinello. *Ab Initio* molecular dynamics with a classical pressure reservoir: Simulation of pressure-induced amorphization in a $\mathrm{Si_{35}H_{36}}$ cluster. *Phys. Rev. Lett.*, **84**, 682 (2000).

[94] M. A. Angadi, T. Watanabe, A. Bodapati, X. Xiao, O. Auciello, J. A. Carlisle, J. A. Eastman, P. Keblinski, P. K. Schelling and S. R. Phillpot. Thermal transport and grain boundary conductance in ultrananocrystalline diamond thin films. *J. Appl. Phys.*, **99**, 114301 (2006).

[95] V. B. Shenoy, R. Miller, E. B. Tadmor, D. Rodney, R. Phillips and M. Ortiz. An adaptive finite element approach to atomic-scale mechanics — the quasicontinuum method. *J. Mech. Phys. Solids*, **47**, 611 (1999).

[96] V. B. Shenoy, R. Miller, E. B. Tadmor, R. Phillips and M. Ortiz. Quasicontinuum models of interfacial structure and deformation. *Phys. Rev. Lett.*, **80**, 742 (1998).

[97] R. Miller, M. Ortiz, R. Phillips, V. B. Shenoy and E. B. Tadmor. Quasicontinuum models of fracture and plasticity. *Eng. Fract. Mech.*, **61**, 427 (1998).

[98] H. Lin and D. G. Truhlar. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.*, **117**, 185 (2007).

[99] M. F. Ruiz-Lopez. Combined QM/MM calculations in chemistry and biochemistry. *J. Mol. Struct.: THEOCHEM*, **632**, ix (2003).

[100] M. Svensson, S. Humbel, R. Froese, T. Matsubara, S. Sieber and K. Morokuma. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy predictions. A Test for Diels-Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *J. Phys. Chem.*, **100**, 19357 (1996).

[101] T. Vreven, K. Morokuma, O. Farkas, H. B. Schlegel and M. J. Frisch. Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *J. Comput. Chem.*, **24**, 760 (2003).

[102] A. Laio, J. VandeVondele and U. Rothlisberger. A hamiltonian electrostatic coupling scheme for hybrid car–parrinello molecular dynamics simulations. *J. Chem. Phys.*, **116**, 6941 (2002).

[103] N. Reuter, A. Dejaegere, B. Maigret and M. Karplus. Frontier bonds in qm/mm methods: A comparison of different approaches. *J. Phys. Chem. A*, **104**, 1720 (2000).

[104] S. Ogata, F. Shimojo, R. K. Kalia, A. Nakano and P. Vashishta. Hybrid quantum mechanical/molecular dynamics simulation on parallel computers: density functional theory on real-space multigrids. *Comput. Phys. Commun.*, **149**, 30 (2002).

[105] S. Ogata, E. Lidorikis, F. Shimojo, A. Nakano, P. Vashishta and R. K. Kalia. Hybrid finite-element/molecular-dynamics/electronic-density-functional approach to materials simulations on parallel computers. *Comput. Phys. Commun.*, **138**, 143 (2001).

[106] S. Ogata, F. Shimojo, R. K. Kalia, A. Nakano and P. Vashishta. Environmental effects of H_2O on fracture initiation in silicon: A hybrid electronic-density-functional/molecular-dynamics study. *J. Appl. Phys.*, **95**, 5316 (2004).

[107] S. Ogata. Buffered-cluster method for hybridization of density-functional theory and classical molecular dynamics: Application to stress-dependent reaction of H_2O on nanostructured Si. *Phys. Rev. B.*, **72**, 045348 (2005).

[108] N. Bernstein and D. W. Hess. Lattice Trapping Barriers to Brittle Fracture. *Phys. Rev. Lett.*, **91**, 025501 (2003).

[109] G. Csányi, S. Winfield, J. Kermode, A. Comisso, A. De Vita, N. Bernstein and M. Payne (2007). http://www.libatoms.org.

[110] G. Csányi, S. Winfield, J. Kermode, A. Comisso, A. De Vita, N. Bernstein and M. Payne. Expressive Programming for Computational Physics in Fortran 95+. *IoP Comput. Phys. Newsletter*, page Spring 2007 (2007). http://www.iop.org/activity/groups/subject/comp/News/file_21759.pdf.

[111] I. Todorov and W. Smith. DL_POLY_3: the CCP5 national UK code for molecular-dynamics simulations. *Phil. Trans. R. Soc. Lond. A*, **362**, 1835 (2004).

[112] D. A. Case, T. E. C. III, T. Darden, H. Gohlke, R. Luo, K. M. M. Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods. The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668 (2005).

[113] D. Holland and M. Marder. Ideal Brittle Fracture of Silicon Studied with Molecular Dynamics. *Phys. Rev. Lett.*, **80**, 746 (1998).

[114] P. D. Haynes, A. A. Mostofi, C.-K. Skylaris and M. C. Payne. ONETEP: linear-scaling density-functional theory with plane waves. *J. Phys.: Conf. Ser.*, **26**, 143 (2006).

[115] D. Sanchez-Portal, P. Ordejón, E. Artacho and J. M. Soler. Density-functional method for very large systems with LCAO basis sets. *Int. J. Quantum Chem.*, **65**, 453 (1997).

[116] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling. *Numerical Recipes: The Art of Scientific Computing* (Camb. Univ. Press, Cambridge (UK), 1992), 2nd edition.

[117] J. A. Hauch, D. Holland, M. P. Marder and H. L. Swinney. Dynamic Fracture in Single Crystal Silicon. *Phys. Rev. Lett.*, **82**, 3823 (1999).

[118] R. S. Rivlin and A. G. Thomas. Rupture of rubber: I. Characteristic energy for tearing. *J. Pol. Sci.*, **10**, 251 (1952).

[119] W. G. Knauss. Stresses in an infinite strip containing a semi-infinite crack. *J. Appl. Mech.*, **33**, 356 (1966).

[120] J. P. Hirth and J. Lothe. *Theory of Dislocations — Second Edition* (Wiley, 1982).

[121] P. B. Hirsch and S. G. Roberts. The brittle-ductile transition in silicon. *Phil. Mag. A*, **64**, 55 (1991).

[122] J. E. Field. Brittle Fracture: its Study and Application. *Contemporary Physics*, **12**, 1 (1971).

[123] R. Perez and P. Gumbsch. Directional Anisotropy in the Cleavage Fracture of Silicon. *Phys. Rev. Lett.*, **84**, 5347 (2000).

[124] K. C. Pandey. New π-bonded chain model for Si(111)-(2×1) surface. *Phys. Rev. Lett.*, **47**, 1913 (1981).

[125] F. Ancilotto, W. Andreoni, A. Selloni, R. Car and M. Parrinello. Structural, electronic, and vibrational properties of Si(111)-2×1 from *ab initio* molecular dynamics. *Phys. Rev. Lett.*, **65**, 3148 (1990).

[126] K. D. Brommer, M. Needels, B. Larson and J. D. Joannopoulos. Ab initio theory of the si$(111)-(7\times7)$ surface reconstruction: A challenge for massively parallel computation. *Phys. Rev. Lett.*, **68**, 1355 (1992).

[127] W. Shimada and H. Tochihara. Formation mechanism of the si(111)7×7 reconstruction studied by scanning tunneling microscopy: Zipper-like restructuring in the sequential size changes of isolated single faulted-halves. *Surface Science*, **526**, 219 (2003).

[128] J. E. Northrup and M. L. Cohen. Reconstruction mechanism and surface-state dispersion for Si(111)-2×1. *Phys. Rev. Lett.*, **49**, 1349 (1982).

[129] J. J. Gilman. Direct measurements of the surface energies of crystals. *J. Appl. Phys.*, **31**, 2208 (1960).

[130] J. E. Sinclair and B. R. Lawn. An atomistic study of cracks in diamond-structure crystals. *Proc. R. Soc. Lond.*, **A329**, 83 (1972).

[131] M. J. Buehler, H. Tang, A. C. T. van Duin and W. A. Goddard III. Threshold crack speed controls dynamical fracture of silicon single crystals. *Phys. Rev. Lett.*, **99**, 165502 (2007).

[132] M. Marder and S. Gross. Origin of crack tip instabilities. *J. Mech. Phys. Solids*, **43**, 1 (1995).

[133] I. Be'ery, U. Lev and D. Sherman. On the lower limiting velocity of a dynamic crack in brittle solids. *J. Appl. Phys.*, **93**, 2429 (2003).

[134] P. Gumbsch, S. J. Zhou and B. L. Holian. Molecular dynamics investigation of dynamic crack stability. *Phys. Rev. B*, **55**, 3445 (1997).

[135] J. Fineberg, S. P. Gross, M. Marder and H. L. Swinney. Instability in the propagation of fast cracks. *Phys. Rev. B*, **45**, 5146 (1992).

[136] T. Cramer, A. Wanner and P. Gumbsch. Energy Dissipation and Path Instabilities in Dynamic Fracture of Silicon Single Crystals. *Phys. Rev. Lett.*, **85**, 788 (2000).

[137] R. F. Cook and E. G. Linegar. Kinetics of Indentation Cracking in Glass. *J. Am. Ceram. Soc*, **76**, 1096 (1993).

[138] J. Riedle, P. Gumbsch and H. F. Fischmeister. Cleavage Anisotropy in Tungsten Single Crystals. *Phys. Rev. Lett.*, **76**, 3594 (1996).

[139] R. Thomson, C. Hsieh and V. Rana. Lattice Trapping of Fracture Cracks. *J. Appl. Phys.*, **42**, 3154 (1971).

[140] J. E. Sinclair. The Influence of the Interatomic Force Law and of Kinks on the Propagation of Brittle Cracks. *Phil. Mag.*, **31**, 647 (1975).

[141] W. Curtin. On lattice trapping of cracks. *J. Mater. Res.*, **5**, 1549 (1990).

[142] J. G. Swadener, M. I. Baskes and M. Nastasi. Molecular dynamics simulation of brittle fracture in silicon. *Phys. Rev. Lett.*, **89**, 085503 (2002).

[143] T. Zhu, J. Li and S. Yip. Atomistic configurations and energetics of crack extension in silicon. *Phys. Rev. Lett.*, **93**, 205504 (2004).

[144] D. Holland and M. Marder. Cracks and Atoms. *Adv. Mater.*, **11**, 793 (1999).

[145] M. S. Daw and M. I. Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B*, **29**, 6443 (1984).

[146] D. Holland and M. Marder. Erratum: Ideal Brittle Fracture of Silicon Studied with Molecular Dynamics [Phys. Rev. Lett. **80**, 746 (1998)]. *Phys. Rev. Lett.*, **81**, 4029 (1998).

[147] M. J. Buehler, A. C. T. van Duin and W. A. Goddard III. Multiparadigm modeling of dynamical crack propagation in silicon using a reactive force field. *Phys. Rev. Lett.*, **96**, 095505 (2006).

[148] H. C. Andersen. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comp. Phys.*, **52**, 24 (1983).

[149] G. Marc and W. G. McMillan. The virial theorem. *Adv. Chem. Phys.*, **58**, 209 (1985).

[150] J. A. Zimmerman, E. B. Webb, J. J. Hoyt, R. E. Jones, P. A. Klein and D. J. Bammann. Calculation of stress in atomistic simulation. *Modelling Simul. Mater. Sci. Eng.*, **12**, S319 (2004).

[151] R. J. Hardy. Formulas for determining local properties in molecular-dynamics simulations: Shock waves. *J. Chem. Phys.*, **76**, 622 (1982).

[152] J. Li. AtomEye: an efficient atomistic configuration viewer. *Modelling Simul. Mater. Sci. Eng.*, **11**, 173 (2003).

[153] W. K. den Otter. Thermodynamic integration of the free energy along a reaction coordinate in cartesian coordinates. *J. Chem. Phys.*, **112**, 7283 (2000).

[154] T. Hoshi, T. Igushi and T. Fujiwara. Nanoscale structures formed in silicon cleavage studied with large-scale electronic structure calculations: Surface reconstruction, steps, and bending. *Phys. Rev. B*, **72**, 075323 (2005).

[155] D. Sherman. Energy considerations in crack deflection phenomenon in single crystal silicon. *Int. J. Fract.*, **140**, 125 (2006).

[156] E. Sharon, G. Cohen and J. Fineberg. Propagating solitary waves along a rapidly moving crack front. *Nature*, **410**, 68 (2001).

[157] D. J. Cole, M. C. Payne, G. Csányi, S. M. Spearing and L. C. Ciacchi. Development of a classical force field for the oxidized si surface: Application to hydrophilic wafer bonding. *J. Chem. Phys.*, **127**, 204704 (2007).

[158] P. Avouris, Z. Chen and V. Perebeinos. Carbon-based electronics. *Nature Nanotechnology*, **2**, 605 (2007).

[159] Y. Kobayashi, K. ichi Fukui, T. Enoki and K. Kusakabe. Edge state on hydrogen-terminated graphite edges investigated by scanning tunneling microscopy. *Phys. Rev. B*, **73**, 125415 (2006).