Cracks and Atoms**

By Dominic Holland* and Michael Marder

1. Introduction

Many materials scientists and engineers are, with some justification, suspicious of theoretical and numerical studies ascending from the atomic scale on the mechanical response of materials. On the one hand, there is a reluctance to believe that the invisible atomic scale is important for macroscopic mechanical deformation. Out of sight, out of mind. On the other hand, many large scale computer simulations that produce brightly colored pictures with gobs of toy atoms, and sometimes even impressive statistics on processing efficiency, seem simply to avoid questions on how to compare computation with either theory or experiment. For in fact, a calculation involving ten billion atoms, necessarily with questionable effective atomic interactions, would exceed the powers of the world’s largest computers, and yet describe only a cube of matter no more than half a micrometer along each side. And even when computers become large enough to store and manipulate the coordinates of this many particles, it will not be possible to follow their behavior for much more than a nanosecond, thus making comparison with experiment seem as remote as a manned flight to Pluto.

These simple observations lie behind the dominance of continuum mechanics in most studies of mechanical behavior of materials. Obviously, so the argument goes, it is an enormous waste of effort to calculate the motion of every atom when all information of interest is contained in continuous fields that are most sensibly studied by other means. Hence the feeling, widely held but seldom expressed, that “real materials are not made of atoms”.

The point of this article is to show that this feeling is wrong. Materials constantly betray their atomic underpinnings. When this happens, it should come as no surprise that the continuum theory breaks down, since it requires a great deal of cleverness indeed to apply continuum elastic theory to phenomena that are neither continuous nor elastic. We will discuss properties of materials for which atomic features are essential to even a qualitative understanding, and show how to design studies at the atomic scale in an efficient manner, studies which permit direct comparison with experiment.

The mechanical response of materials is an enormous and varied subject. We will therefore focus on one particular case that makes it possible to examine the relationship between atomic and macroscopic scales in detail: the process of brittle fracture.

Fracture is important because it determines the ultimate strength of a wide range of materials. Fracture fundamentally has to do with the severing of inter-atomic bonds: this points theoretical investigations toward atomic-scale studies. As gem-cutters know, cracks tend to run along crystal planes, showing that the process is sensitive to atomic detail. Nevertheless, most fracture research is carried out in the context of continuum elasticity through an elegant framework that bypasses most of the questions arising at the atomic scale. Our aim is to identify the questions that the continuum approach cannot address, and to show how a combination of theoretical insight and numerical computation can be employed to answer them. The ability to compare directly with experiment will then provide a strong test of the correctness of the underlying interatomic potentials used in simulations.

2. Continuum Elastic Fracture Mechanics

Drop a metal spoon on the floor and it may only bend a little. Drop a glass and it will shatter. Why? Continuum mechanics cannot resolve this matter, but it has a great deal to say about the rapid fracture taking place when something breaks.

In particular, using continuum elastic fracture mechanics one can derive, with some effort, an equation of motion for a straight, clean, dynamic crack. Let \( l(t) \) be the length of a dynamic crack in a two-dimensional large plate under tensile loading of strength \( \sigma_t \) applied only to the crack faces. Let \( G(v) \) be the amount of mechanical strain energy required for the crack to advance by unit area while traveling at velocity \( v \); \( G \) is called the fracture energy. The equation of motion is shown in Equation 1.\(^{[1-7]} \) where \( v \) is Poisson’s ratio, \( E \) is Young’s modulus, and \( c_R \) is the Rayleigh wave speed (the speed of sound on a free surface, about 0.9 times the shear wave speed). Equation 1 relates the stress fields near the crack tip with the total flow of elastic energy in the entire medium to the crack tip energy sink (assumed to be the only mechanical dissipative region in the system). The effects of bulk material properties are incorporated in the elastic constants (\( E \) and \( v \)), and the resistance of the material to crack extension is represented by \( G \).

** This work was supported by the National Science Foundation (DMR-9531187), the Texas Advanced Computing Center, and the National Partnership for Advanced Computational Infrastructure. Thanks to Jens Hauch and Harry Swinney.
Continuum elastic fracture mechanics by itself makes no prediction about how the crack velocity should vary as the fracture energy changes. The fact of the matter is that there is an unknown functional relationship in Equation 1: \( G(u) \) is unknown \((u, l, \text{and } \sigma_0 \text{ are interrelated}).\) Equation 1 does say that the Rayleigh wave speed is an upper bound on the possibilities for the crack speed \( u \) (otherwise if \( u \) exceeded \( c_R \), the energy required for unit area crack extension would be negative); but there is no requirement that \( u \) should approach \( c_R \) as the energy available for fracture increases. If \( G \) is bounded, however, then \( u \) must approach \( c_R \), since \( l \) on the right hand side of Equation 1 can become arbitrarily large. Intuitively, one would not expect the energy required to advance the crack by unit area at some speed \( u \) \((<c_R)\) to become unboundedly large. From a first approximation, an advancing crack involves only the severing of bonds, which would lead to a constant \( G(u) \); but even allowing for the possibility that sound waves also will be generated by the advancing tip, one would not expect these waves to be capable of carrying off ever-increasing amounts of energy.

The form of \( G(u) \) cannot be settled using continuum elastic fracture mechanics, for to do so would require a detailed knowledge of what goes on in the process zone, the region near the crack tip where the material is furthest from equilibrium. To make progress, one must either assume a form for \( G(u) \), estimate it numerically, or measure it experimentally. For dynamic cracks, continuum elastic fracture mechanics then predicts the crack velocity given the loading \( \sigma_0 \). The simplest assumption is that \( G(u) \) is a constant, and taking this to be true, Equation 1 is easily put in the form\(^{[5,8]}\) shown in Equation 2, where \( l_0 \) is the crack length at which the crack first receives enough energy to begin moving.

\[
v = c_R \left( 1 - \frac{l_0}{l} \right)
\]  

(2)

The results of Equation 2, alas, are not in accord with experiment. The equation predicts that the speed of a crack should approach the Rayleigh wave speed as the crack lengthens. Experimentally, however, the limiting speed is usually between 20 and 80 % of this value.\(^{[8–12]}\)

We will present a detailed description of the nature of these limiting speeds in later sections, but a sketch of the main idea is as follows: As a crack approaches the Rayleigh wave speed, continuum theory predicts that the elastic region about the crack tip must store greater and greater amounts of elastic energy in the form of rapidly varying stress and strain fields. Because real materials are not prone to handling an ever-increasing energy flux in a controlled manner, the continuum description of a single straight crack must break...
Thomson and coworkers\cite{13}, who uncovered a property they such models were studied in the early 1970s by Robb springs that snap after undergoing a small extension. The first at lattice sites connected to nearest neighbors with Hooke one and two dimensions. The models are made from particles provided by lattice theory. It is built around lattice models in the continuum point of view. The most complete description of dynamic fracture is provided by lattice theory. It is built around lattice models in one and two dimensions. The models are made from particles at lattice sites connected to nearest neighbors with Hooke springs that snap after undergoing a small extension. The first such models were studied in the early 1970s by Robb Thomson and coworkers\cite{14-16} who uncovered a property they called lattice trapping, which is the inability of a crack to advance in a crystal even though, from an energy argument alone, it should do so. Then in the early 1980s, Leonid Slepyan took things mathematically much further, studying dynamic cracks in an infinite square lattice, which he thought of as modeling inhomogeneous solids such as concrete. However, it is more natural to consider the lattice points to be atoms, and long thin lattice strips have been studied extensively with this point of view in the background. These strip lattice models are completely analytically solvable, and have the following four properties: 1. When the fracture energy is moderate, i.e., for moderate strains, dynamic cracks are attracted to steady states. In other words, they want to be well-behaved; 2. When in steady state, there is an intimate connection between the speed of the crack and the wave properties of the medium; 3. There is a large minimum velocity, below which a crack will not travel, i.e., there is a velocity gap or forbidden band of velocities. This is related to the lattice trapping earlier found by Thomson; 4. When the fracture energy exceeds some threshold, the crack will become unstable, it will no longer be in a steady state, and its dynamics rapidly become very complex.

The last property, instability, is well-established experimentally for brittle amorphous materials, where it is clearly responsible for the inability of cracks to reach the Rayleigh wave speed\cite{14-16}. But the other three properties are difficult to observe experimentally, and have not conclusively been observed in the crystals, where they are most confidently predicted. These four properties together define an ideal brittle universality class.

It is only proper to ask whether the lattice theory of fracture has much bearing on reality; in particular, does there exist in nature an ideal brittle universality class of materials? After all, there does not exist any two-dimensional material that is rigid as the crack proceeds in a steady fashion, the energy dissipated by crack motion per unit area crack advance must then be given by Equation 4.

\[
\frac{\varepsilon_{\text{right}}}{A} = \frac{1}{2} \frac{\delta^2 E}{L} \quad (3)
\]

Far to the left of the crack, the material is completely relaxed, and therefore if the boundaries of the system are rigid as the crack proceeds in a steady fashion, the energy dissipated by crack motion per unit area crack advance must then be given by Equation 4.

\[
G = \frac{1}{2} \frac{\delta^2 E}{L} \quad (4)
\]

The reason for this assertion is that when the crack moves ahead in a steady fashion the stress and strain fields around it do not change. As shown in Figure 1, the fields translate in a few kilometers per second, bifurcating and changing direction, or when dislocation loops nucleate, propagate, and entangle. That is, it is difficult to get a theoretical handle on much of reality. So between theory (or lack of it) and experiment, then, is a large, unexplored territory. This is where realistic large-scale molecular dynamics simulations can be a powerful investigative tool.

How realistic is realistic? Although it depends on what “is” is, the answer is not very. This important matter is discussed more fully below. How large is large? Again, not very. But does it matter? On a large supercomputer, large and super being datable modifiers, one can follow on the order of ten million particles for ten nanoseconds, with reasonably sophisticated interactions between the particles, in three dimensions. There are fifteen or so orders of magnitude to go to reach the macroscopic scale studied in the laboratory. Molecular dynamics simulations are not going to get there in a brute-force way. It must be borne in mind that it is important to be able to compare with experiment: simulations can be extraordinarily complex, and are necessarily crude approximations of reality. Thus, when one is making claims to realism, one had better be able to compare with experiment, the arbiter at the end of the day, and not get lost in a vast virtual phase space. Toward this end, with molecular dynamics modeling of dynamic fracture, it is imperative that simulations be designed around a scaling argument that will take one through the fifteen or so orders of magnitude from virus-sized systems up to the macroscopic scale, and do so in a manner that is well-founded at the atomic level.

4. The Scaling Argument

The scaling argument and the thin strip geometry go hand in hand. What one needs control over, after all, is the energy flux to the crack tip. For a long crack moving at steady velocity in a strip whose upper and lower boundaries are held rigid, one can deduce the energy consumed by the fracture in a trivial way. Consider the crack moving in the thin strip depicted in Figure 1. Far to the right of the crack tip the material is under tension, and it stores an elastic energy \(E_{\text{right}}\) per area \(A\) (Eq. 3).

\[
\frac{\varepsilon_{\text{right}}}{A} = \frac{1}{2} \frac{\delta^2 E}{L} \quad (3)
\]

Far to the left of the crack, the material is completely relaxed, and therefore if the boundaries of the system are rigid as the crack proceeds in a steady fashion, the energy dissipated by crack motion per unit area crack advance must then be given by Equation 4.

\[
G = \frac{1}{2} \frac{\delta^2 E}{L} \quad (4)
\]
the direction of crack motion, and the change in energy comes from the transformation of stressed material ahead of the crack to the unstressed material behind.

Fig. 1. The upper surface of a strip of height \( L \) and thickness \( w \) is rigidly displaced upwards by distance \( l \). A crack is cut through the center of the strip, and relieves all stresses in its wake. When the crack moves distance \( d \) from (A) to (B), the net effect is to transfer length \( d \) of strained material into a length \( d \) of unstrained material.

This conclusion rests upon symmetry, and does not even demand that strains ahead of the crack tip be so small that linear mechanics be applicable. Steady velocity \( u \) and a corresponding energy flux \( G \) are achieved in the long time limit, the natural scale in experiments as well as in analytical calculations, where the crack tip reaches dynamic equilibrium with waves reflecting from top and bottom boundaries.

That is, the return of acoustic waves from the system boundaries actually simplifies the task of understanding energy balance in the thin strip geometry. According to fracture mechanics, the relationship between energy flowing to a crack tip and crack velocity is, for a given lattice direction, universal. Having obtained the relation in a strip, one knows it for any of the vast range of geometries to which fracture mechanics is applicable, such as a long crack in a large plate.\(^{[17]}\)

In order to relate samples of different size to one another, let \( G_c \) be the Griffith energy density\(^{[5,18]}\) that is, twice the crack surface energy density, a lower bound on the energy per unit area required for a perfectly efficient crack to propagate along a certain plane. One can then define a dimensionless measure of loading \( \Delta \) as in Equation 5, where \( G \) is the fracture energy density already introduced, i.e., the elastic strain energy stored per unit area (in the fracture plane) ahead of the crack.

\[
\Delta = \sqrt{G/G_c} \tag{5}
\]

Analytical solutions for the ideal brittle solid show that the relationship between \( \Delta \) and crack velocity becomes independent of the height of the strip (number of planes stacked vertically) for surprisingly small strips\(^{[19]}\), a strip 80 atoms high has, for all practical purposes, reached the infinite limit, Figure 2. Guidance for conducting computationally expensive, moderately realistic, molecular dynamics simulations can be had from this simple result: the very rapid convergence of the main quantity of physical interest allows one to obtain physically meaningful results from simulations that are considerably smaller than many that are currently being carried out.\(^{[20–24]}\) This approach can also be viewed as an alternative to methods that join together atoms and continua.\(^{[25–29]}\)

If computational resources were infinite, one would still need to properly choose the sample geometry in order to compare and inform experiment and theory. With limited computational resources, choice of sample geometry and scale are of paramount importance: when performing molecular dynamics simulations of fracture, one should not ask how large a system one can work with, but rather what is the smallest system that will give results scaleable to the macroscopic level. A good scaling argument can transform a “grand challenge” problem into something that’s better focused and less of a challenge. Size does matter: a smaller spatial scale allows one to follow the evolution of systems for longer times; and with a strip geometry one can then fully analyze steady state behavior and connect theory and experiment.

5. Molecular Dynamics

The molecular dynamics method applies to any system of particles with some prescribed inter-particle potential. It consists of integrating Newton’s equations of motion for all particles in lock step over a series of time steps, the size of the chosen step being small enough to give converged dynamics. Time step integration for all results reported here was achieved using the Verlet algorithm\(^{[30–32]}\).

Lattice theory makes predictions that are hard to observe in experiment, and one of the reasons for doing computer simulations is to relate the two. The simulations to be described later were carried out for silicon. Silicon is
extremely brittle, and high-quality macroscopic single crystal wafers are cheap. It is therefore an excellent candidate for laboratory fracture experiments. Silicon is also of great technological importance and, as a result, is one of the most studied materials. One measure of this is that there are over thirty effective (actually rather ineffective) interatomic potentials for silicon in the literature.\(^{33,34}\) Furthermore, transmission electron micrographs of cracks in silicon wafers reveal atomically sharp crack tips.\(^{35,36}\) Silicon therefore is an obvious candidate for molecular dynamics investigations of dynamic fracture, and an appropriate setting for testing lattice theory.

**5.1. Interatomic Potentials**

The equation underlying materials physics is not in doubt. It is Schrödinger’s equation. This equation can be solved analytically for the hydrogen atom, numerically for the helium atom, and with reliable approximate methods of quantum chemistry for small molecules. For any solid of interest in the study of materials, all hope of controlled approximations must be abandoned, and the Schrödinger equation is brutally reduced to tractable form in a way that is refined by comparison with experiment. Quantitative methods that employ such approximations from the beginning are called ab initio.

Despite their origin, ab initio methods are the most reliable techniques available for numerical treatment of materials. However, they are restricted to perfect crystals where the unit cell is not much bigger than 1000 atoms, and these atoms can be followed for only a few tens of picoseconds. The main motivation, then, for constructing empirical or effective “classical” interatomic potentials is speed of computation and the ability to work with relatively large numbers of particles. With effective potentials, it is possible to follow \(~10^7\) atoms for a few tens of nanoseconds. This difference in computational scales becomes important in the modeling of processes that require a minimum of \(~10^5\) atoms to capture just some of the complex underlying physics: processes involving fracture, dislocation loops, grain boundaries, or amorphous-to-crystal transitions, for example.

Another motivation for constructing effective interatomic potentials is that they make the complex physics of what are fundamentally quantum mechanical phenomena more physically intuitive, so that one may interpret the results of atomistic simulations in terms of the simple principles of chemical bonding.

**5.2. Realistic Potentials for Silicon?**

Solid silicon is composed of covalently bonded atoms and has the open diamond crystal structure. If only two-body potentials operated among the atoms, one would expect the crystal to collapse in on itself to form a close-packed structure, thereby reducing its energy. Covalent systems, however, are characterized by restoring forces between pairs of contiguous interatomic bonds. That is, pairs of bonds with an atom in common want to maintain a preferred angle between them. These extra forces are what stabilize the open diamond structure in silicon, carbon, grey-tin, and germanium, for example. The lowest-order way of capturing this property with effective interatomic potentials is to go beyond binary bonding and include a three-body term in the system’s Hamiltonian.

The Stillinger–Weber (SW) (two- and three-body) potential\(^{37}\) has proved to be very popular and durable in the literature. It gives excellent elastic properties, and captures well the nonlinear physics involved in heating and melting. It is therefore a reasonable starting point for conducting molecular dynamics fracture simulations in silicon. Unfortunately, the potential will not yield fracture along the experimentally preferred fracture planes (111) and (110). At low or moderate strains, what happens is that two dislocations open up at the crack tip, blunting it and preventing it from advancing (Fig. 3). One can play around with giving a transverse opening velocity to a select few atoms around the tip. But to no avail. The crack simply will not crack. At very high strains, the crack tip region melts.

![Fig. 3. Crack tip blunting in SW silicon: two dislocations open up at the tip, preventing it from advancing. The crack is pointing in the direction (110) in the plane (110), and the system is loaded with a strain parameter \(\lambda = 1.6\).](image)

SW does give a type of fracture along the (100) plane which is quite rough on the atomic scale (Fig. 4). Abraham et al. call this brittle fracture.\(^{23}\) There is as yet little consensus on a precise definition of brittleness. The experimental results of Lawn and Hockey\(^{38,36}\) for fracture along (111) in silicon show, however, that it is possible to have atomically sharp fracture, i.e., where the newly created fracture surfaces are atomically flat. The experimental evidence for fracture along (100), on the other hand, is scant and inconclusive,\(^{38}\) and that SW yields fracture along this plane, albeit in a rough
manner, might even be yet another indication of the potential’s shortcomings.

It is possible to get cracks going in an ideal brittle manner along (111) and (110) with the SW potential by increasing the restoring forces between pairs of bonds; i.e., by increasing the stability of the tetrahedra in the diamond lattice, and thus making the crystal more brittle. This can be done by scaling a parameter, \( \lambda \), the coupling constant in the three-body term. Originally, Stillinger and Weber set \( \lambda = 21 \) (dimensionless). However, by doubling this, one can obtain fast brittle fracture. With a fast crack running, if one quasistatically decreases \( \lambda \), the crack arrests well before one reaches \( \lambda = 21 \) (Fig. 5). Although \( \lambda = 42 \) gives fracture phenomenology in reasonable accord with experiment, it has the adverse effect of raising the melt temperature above 3500 K, whereas experimentally the melt happens at 1685 K. The Young moduli also get shifted. Moduli results from tensile tests on small samples, for three lattice directions, are given in Table 1 for both the original and modified SW potentials, along with the corresponding experimental values.

Table 1. Elastic constants of silicon, comparing SW, modified SW potentials, and experiment.

<table>
<thead>
<tr>
<th></th>
<th>Silicon</th>
<th>Original SW</th>
<th>Modified SW</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_{33} ) GPa</td>
<td>150</td>
<td>114</td>
<td>172</td>
</tr>
<tr>
<td>( E_{11} ) GPa</td>
<td>169</td>
<td>139</td>
<td>189</td>
</tr>
<tr>
<td>( E_{11} ) GPa</td>
<td>188</td>
<td>151</td>
<td>201</td>
</tr>
</tbody>
</table>

A more recent, and also much more sophisticated, potential for silicon is the environment-dependent interatomic potential, EDIP, by Bazant et al.\[34,39–41\] It has much functional similarity with the SW potential, but its novelty and greater complexity lie in its sophisticated environment dependence. Silicon is tetravalent, and can form a number of covalent bonding configurations, corresponding to different bonding orbital hybridizations. Different hybridizations kick in for different coordination numbers. The method of EDIP is to provide an accurate potential for each of the different preferred integer coordinations, and to have smooth interpolation between these configurations; \( sp^2 \) hybrids around three-fold coordinated atoms in a hexagonal plane (the theoretical graphitic structure), \( sp^3 \) hybrids with four-fold coordination arranged in tetrahedra (the diamond structure), six-fold coordination for simple cubic, and eight-fold coordination for bcc, for example. The potential was developed for bulk materials—explicitly excluding asymmetric distributions of neighbors—which occur for under-coordinated structures like surfaces and small clusters. A scalar representation of the environment, the coordination number, will therefore suffice; it varies continuously as hybridizations change. It is reasonable to assume that changing hybridization is important at an advancing crack tip. Unfortunately, EDIP is even more resistant to dynamic fracture than SW. Again, dislocations will open up at the seed crack tip, blunting it and keeping it in arrest.

As with SW, EDIP also captures reasonably well the nonlinear physics of heating and melting. But both potentials fail rather dramatically when applied to the nonlinear processes of stretching and rupturing interatomic bonds involved in fracture. A key to understanding why comes from density functional theory, or rather an empirical fit to DFT cohesive energy curves, viz. the universal energy relation of Rose et al.,\[42\] which characterizes materials strained far from...
equilibrium. The universal energy relation (Eq. 6) gives the cohesive energy as a function of a uniformly strained lattice, where 

\[ s = (1 - a/a_0)c, \]

with \( a \) being the scaled nearest neighbor atomic separation and \( a_0 \) the equilibrium atomic separation (2.35 Å in Si). \( b = 4.64 \) and \( c = 4.88 \) for silicon. \( E(s) \) is in eV.

\[ E(s) = b[\exp(s)](-1 + s + 0.05s^3) \quad (6) \]

This equation gives remarkably good fits (with appropriate values of \( b \) and \( c \)) to experimental data and density functional calculations for all materials it has been applied to, accurately modeling moderate uniform compression and also moderate uniform expansion (where experimental data are lacking). Figure 6 (upper) is a comparison plot of the universal energy relation applied to silicon, Equation 6, showing an interaction cutoff at \( \approx 5.5 \) Å, and the cohesive energy curves for SW and EDIP, the latter two having cutoffs less than the second-nearest neighbor distance, <3.84 Å. Figure 6 (lower) shows the corresponding restoring forces. The salient point is that the nearest neighbor potentials are too short in range, and must rise from the bottom of the cohesive well and go to zero rapidly as the system is dilated. This results in a large gradient before cutoff, which means there is an unreasonably large force of attraction before rupture. This is more pronounced in the shorter-ranged EDIP. These large forces inhibit crack propagation.

5.3. Remarks on Potentials for Silicon

Most of the potentials for silicon in the literature have a cutoff restricted to nearest neighbors, examples of which, along with SW and EDIP, are the Dodson potential \([43]\) and the three Tersoff potentials \([44–46]\). Two longer range potentials are those due to Biswas and Hamann, \([44]\) and Pearson, Takai, Halicioglu, and Tiller (PTHT), \([46]\) with interactions up to the third (cutoff at 5 Å) and seventh (cutoff at 7.3 Å) shells, respectively. The elastic properties of these longer range potentials are very poor—the bottoms of the wells are too narrow and steep. These potentials are also too steep for all compression strain values. The short range potentials, on the other hand, describe compression well, and expansion up to a nearest neighbor separation \( r \approx 2.8 \) Å. The PTHT potential gives a bulk cohesive energy for diamond of –5.45 eV, whereas SW and EDIP give –4.34 and –4.65 eV, respectively, in close agreement with density functional theory. The Biswas–Hamann potential and Tersoff’s third potential \([46]\) overestimate the melt temperature, giving 2900 K and 3000 K, respectively (not too far from the modified SW potential). None of the potentials is able to model the various \((1 \times 1, 2 \times 1, 7 \times 7)\) reconstructions of the (111) surface.

The potential by Bolding and Andersen \([49]\) has over 30 adjustable parameters, fit to a wide range of structures, and is claimed to describe accurately bulk phases, defects, surfaces, and small clusters.

6. Fracture Simulations

In this section, we will describe molecular dynamics simulations of fracture in silicon. The scaling argument of Section 4 will be used in determining the appropriate size and geometry of the sample, the boundary conditions, and also the rate at which the sample is strained, i.e., the rate at which the fracture energy \( G \) changes in time. All of the
results presented are from extensive computational runs performed using the modified SW potential as described above in Section 5.2, i.e., where the restoring force between pairs of contiguous bonds, the three-body term, is doubled in strength: $\lambda = 42$ is used instead of $\lambda = 21$.

Simulations were run to investigate the fracture properties of silicon for a number of crystal planes, and for different directions in those planes. Different strain rates were used, but these were generally quasistatic. The temperature dependence of crack properties was also investigated. And finally, room temperature fracture runs were carried out and directly compared with experiment.

The main physical quantity of interest in what follows is $\Delta(u)$ (or $G(u)$). To determine $u$, the crack tip was located and recorded every few time steps. A simple opening criterion was used. A sample of tip location versus time is in Figure 7. Each bond breaking event is clearly distinguished. Velocity can then be measured as accurately as possible.

6.1. Design of Simulations

In the simulations depicted in Figure 8, the crack runs along $x$, exposing either (111) or (110) planes. Three separate boundary conditions are employed:

- **$x$–$y$ planes**: Two layers of atoms at the top and bottom of the strip are held rigid during the simulation. By pulling them apart, elastic energy of any desired amount can be stored ahead of the crack tip. Sometimes the distance between top and bottom layers is held fixed during the whole simulation, while other times it is increased or decreased quasistatically.

- **$y$–$z$ planes**: To model a crack in an infinitely long strip, cutting and pasting on a conveyor belt is used. First, whenever the crack tip approaches within 150 Å of the right-hand boundary, new crystal is pasted onto the right-hand side, and broken crystal cut from the left. Second, to prevent elastic waves from informing the crack tip that it lives in a strip of finite extent along $x$, there are energy absorbing regions 20 Å thick at both the left and right ends—though this is not important for the region ahead of the crack tip, which remains practically undisturbed until the crack becomes unstable.

- **$x$–$z$ planes**: These boundaries are periodic. This choice enables one to describe correctly the flux of energy to the crack tip in a macroscopic sample, at least while the crack is stable. Nakamura and Parks have shown that in a macroscopic plate of thickness $d$, at distances from the crack tip much smaller than $d$, the appropriate elastic solutions are found to be those with such periodic boundary conditions.

The initial equilibrium sample is $614 \times 20 \times 153 \, \text{Å}^3$. The length, 614 Å along the $x$-axis, is sufficient so that the crack tip will not feel the effect of the damped head and tail boundaries, and also that the waves emanating from the crack tip can be characterized well and analyzed. The thickness, 20 Å along the $y$-axis, is three unit cells deep, which is more than generous for this periodic axis. The primary dimension informed by the scaling argument, the height, is 153 Å along the $z$-axis, a value which, like the thickness, errs on the generous side. It is sufficiently large to give scaling behavior for all values of $G$ up to instability, i.e., where the steady state analysis breaks down. A system twice as large gives identical results, although after a longer propagation time and a much longer central processing unit (CPU) time.

With the above dimensions, the number of atoms involved at any stage of the simulation is approximately 94,000, although after numerous cuts and pastes at left and right
hand boundaries tens of millions of atoms are cumulatively involved. This number is small enough to enable one to follow crack motion for the times (~10 ns) needed for the cracks to proceed through a succession of steady states between arrest and instability.

The top and bottom layers of atoms are pulled apart so that the load on the system that drives the crack is given by the strain parameter $\Delta$ defined in Equation 5. The two most explored fracture planes were (110) and (111), for which the Griffith energy densities are $G_c = 3.3$ J/m$^2$ and $G_c = 2.7$ J/m$^2$, respectively.

To initiate the simulation, the two $x$-$y$ boundaries are pulled apart so that $\Delta$ in Equation 5 is 1.6; a narrow seed crack is inserted running half the sample length; an initial velocity is given to a few atoms near the crack tip; and it is let rip, i.e., Newtonian mechanics takes over. The time step used is 4 fs. There are about 30 time steps in the smallest period of vibration in the system, giving very good energy conservation. Decreasing the time step by a factor of ten shows no change in the steady state dynamics. At very high strains, there is much more energetic particle motion, and a time step of ~0.4 fs is required.

6.2. Steady States

Steady states are most directly manifested by the properties of the waves emanating from the crack tip and running along the newly created crack surfaces. To investigate the existence and character of steady states, it is desirable to be able to monitor the motions of an arbitrary number of particles along the fracture plane. In a steady state, it might be that the atoms at every equivalent lattice site along the crack line behave in the same way, or perhaps only every second or third atom, etc. This can be checked by monitoring atoms at successive equivalent lattice sites along the crack line.

Particles are found up ahead of the crack and tracked as the crack advances, hits them, and on until they reach the tail where they get removed by the advancing conveyor belt mechanism, Figure 9. For silicon, the only steady states found were those where a particle at every successive lattice site along the fracture plane behaved in the same way.

7. Results of Zero Kelvin Calculations in Silicon

Questions to be answered: (1) Are there loads $\Delta$ where cracks are attracted to steady states? (2) Do cracks emit phonons at the predicted frequencies? (3) Do cracks refuse to travel below a minimum velocity $v_1 > 0$?, and (4) Do they go unstable above an upper load $\Delta_c$? The answer to all questions is yes.

Figure 9 shows the time history of two different atoms on an upper (111) fracture surface for $\Delta = 1.6$ after approximately 50 ps. The atoms are not yet behaving identically. This is because the advancing crack tip has not yet reached equilibrium with the waves it sends out that reflect from the upper and lower rigid boundaries. Figure 10, on the other hand, shows atomic motions after the crack has been traveling for over 0.24 ns. As anticipated by the theory of ideal brittle fracture, the crack has reached steady state with velocity $v = 3460$ m/s, which means that the vertical displacement $z_{n, R}$ of an atom originally at crystal location $R$ is related to the vertical displacement $z_R(t + n\delta t)$ of an atom $n$ lattice spacings $a = a\delta t$ to the right by Equation 7.

$$z_{n, R} = z_R(t + n \delta t)$$

For a range of loads $\Delta$, Equation 7 applies for any pair of atoms, whatever their separation along the crack surface. In order to obtain the perfect periodicity shown in Figure 10, the crack was allowed to run first for 60,000 time steps so as to come into equilibrium with the waves it sends towards top and bottom boundaries.

The longer and/or higher the system the longer it will take to reach a steady state. The 240 ps required to reach the steady state depicted in Figure 10 is about an order of magnitude longer than the duration of most large-scale molecular dynamics simulations of fracture. Using a reduced spatial system size, as validated by the scaling argument, is what makes this possible.
when proceeding through a sequence of steady states, which requires quasistatic changes in the loading, \( \Delta \) or \( G \). To obtain a full set of results, like Figure 12, a crack actually will travel tens of micrometers, or for times on the order of a tenth of a microsecond. To achieve this, one needs not only efficient code and a high performance computer, but also a physically motivated smallest computational cell: a minimum thin strip on a conveyor belt.

Under steady state conditions as described by Equation 7, the radiation far from the crack tip obeys Equation 8, so that the crack excites all surface phonons whose frequency \( \omega(k) \) and wave number \( k \) in the extended zone scheme obey the Cherenkov condition (Eq. 9), where \( k' \) is restricted to the first Brillouin zone and \( K \) is a reciprocal lattice vector, so that crack velocity \( \bar{u} \) equals phonon phase velocity.

\[
\exp[i \cdot k \cdot (R + na) - i \omega(k)(t + na/\bar{v})] = \exp[i \cdot k \cdot R - i \omega(k)t] \quad (8)
\]

\[
\omega(k) = v \cdot k = \omega(k') = v \cdot (k' + K) \quad (9)
\]

Thus, the Cherenkov condition is equivalent to demanding that a propagating wave obey Equation 7, so Figure 10 also shows that Equation 9 is satisfied. A close-up snapshot of a steady-state crack in silicon is in Figure 11.

In order to find fracture speed \( u \) as a function of the loading parameter \( \Delta \), the separation between the \( x-y \) boundaries is decreased adiabatically while allowing the simulation to run. The crack tip is precisely located every second time step, showing clearly every single bond-breaking event, (Fig. 7). In order to estimate just how slow the strain rate must be to achieve the adiabatic limit, numerical simulations of the analytically solvable models were carried out (Fig. 2), and the results compared with analytical results, leading to the criterion that the dimensionless strain rate \( \dot{\varepsilon} / c \) should be much less than one, where \( \dot{\varepsilon} \) is the strain rate, \( c = 5500 \text{ m/s} \) is a sound speed, and \( h_z = 153 \text{ Å} \) is the height of the sample along \( z \). For the simulations described here, low rates \( \dot{\varepsilon} < 100 \text{ μs}^{-1} \), or \( \dot{\varepsilon}h_z/c < 10^{-4} \) were employed. Prior molecular dynamics simulations of fracture have been carried out with strain rates of order >10⁴ times greater, for which steady states are unattainable and the crack very rapidly goes unstable. In laboratory experiments,[11] it is possible to have \( \dot{\varepsilon}h_z/c \sim 10^{-8} \).

### 7.1. Along (110)

The relation between velocity \( \bar{u} \) and load \( \Delta \) for cracks exposing (110) and traveling along [1\( \bar{T} \)0] is shown in Figure 12. The crack velocity smoothly decreases as \( \Delta \) decreases, until at \( v = 2256 \text{ m/s} \) and \( \Delta = 1.258 \), the crack abruptly comes to a halt. Raising \( \Delta \) again, the crack does not begin to move until \( \Delta = 1.366 \), a value that is sensitive to residual vibrations in the crystal, but the rising curve then perfectly overlaps the descending one. Crack speed continues to rise smoothly until \( v = 3586 \text{ m/s} \), \( \Delta \approx 2.2 \), at which point steady state motion becomes unstable. When a crack becomes unstable, complicated phenomena such as formation of small branches, emission of dislocations, and changes in the plane of propagation can occur, and intermittency where the crack makes repeated attempts at branching.

![Fig. 12. Relation between crack speed \( \bar{u} \) and load \( \Delta \) for crack along (110) at 0 K. As \( \Delta \) descends, velocity drops abruptly to zero at a lower critical value, and as \( \Delta \) ascends resumption of crack motion is hysteretic. For convergence check, system size was doubled along \( x \) and \( z \), and \( v(\Delta) \) measured. Cf. Figure 2.](image)

### 7.2. Along (111): Crackons

The relation between velocity \( \bar{u} \) and load \( \Delta \) for cracks along (111) and traveling along [01\( \bar{T} \)] is shown in Figure 13. For \( 1.44 < \Delta < 2.2 \), the crack has stable steady states, and for \( \Delta > 2.2 \) it goes unstable in a similar manner to cracks along (110). However, for \( 1.175 < \Delta < 1.44 \) the dynamics of the crack exhibit a number of interesting features that have not been seen previously, and for which there is not yet a complete theoretical description. There is a variety of different dynamic states available for each value of \( \Delta \), where
the crack travels at different speeds. Each of these states corresponds to a plateau in \( u(D) \); \( D \) can change by as much as one fifth of the amount needed to go from arrest to instability and the crack velocity does not alter within numerical resolution. When the crack finally decides to accelerate out of the plateau, it may jump by over 1 km/s and reach an upper plateau to within a few m/s. On cyclical loading the same plateaus are always reached. All of these transitions are hysteretic, as depicted in Figure 13. The different states emit noticeably different phonons; on a given plateau, the phonon frequencies appear fixed and their amplitude changes, while between plateaus the frequencies change in accord with Equation 9. This is crackon behavior. All these phenomena are easily disguised if strain rates are too high. Resolving all the fine structure visible in Figure 13 required \( e < 8 \) μs\(^{-2} \), or \( \frac{\varepsilon}{\nu^2} < 1 \times 10^{-5} \).

7.3. Other Lattice Directions

A crack running along (111) but in a direction \([21\bar{1}]\), perpendicular to that shown in Figure 13, was also simulated. Its \( u(D) \) relation is in Figure 14. It has less plateau structure, a larger velocity gap, and higher lattice trapping. Since crack velocity and phonon phase velocity are directly related, and sound velocities depend on the direction of propagation, it follows that different \( u(D) \) relations will result for a single fracture plane, with well-defined Griffith point, but with the cracks traveling in different directions in the plane.

Some crystal planes are not fracture planes at all, and some planes barely allow crack propagation. Figure 15 shows a snapshot of two such situations. In Figure 15a, a seed crack is oriented along \([211] \) \( [1\bar{1}1] \) (same as \( [2\bar{1}\bar{1}] \) \( [\bar{1}1\bar{1}] \)) at \( D = 1.6 \). However, the crack immediately takes off in a diagonal along \([111] \) \( [21\bar{1}] \). This can be guessed from looking at a ball-and-stick model of silicon.

Figure 15b shows a crack reluctantly traveling along \([001] \) \( [100] \) at \( D = 1.2 \). This plane will only permit unstable crack propagation in a narrow window around \( D = 1.2 \). The \( [100] \) plane is the plane of propagation only in an average sense, as the crack keeps jumping about.

8. Implementing Temperature

The cracking crystal is maintained at a particular temperature by having a strip region running the height of...
the system far ahead of the crack act as a heat bath at the desired temperature. In the heat bath, the fluctuation–dissipation theorem \[54\] is used to set the temperature. That is, at each time step, the atoms are given random forces or kicks within a certain amplitude and their motion undergoes a corresponding damping. The square amplitude of the random force \(F^2\) is given by Equation 10, where \(b\) is the applied damping, \(\langle p^2 \rangle\) is the average momentum squared of a particle at a time step, and \(\Delta t\) is the size of the time step. Random forces are correlated over a time step, and uncorrelated from time step to time step.

\[
\langle F^2 \rangle = 2b \langle p^2 \rangle / \Delta t
\]  

(10)

Taking into account Equation 11, where \(m\) is the mass of an atom, \(k\) is Boltzmann’s constant, and \(T\) is the temperature, the amplitude of the random force is given by Equation 12.

\[
\langle p^2 \rangle / 2m = \frac{3kT}{2}
\]

(11)

\[
F_A = \sqrt{6bmkT/\Delta t}
\]

(12)

Note that if there are different atomic species in the crystal with different masses, there will be a different random force amplitude for each. The random forces and damping are implemented at the start of each time step. For each dimension (\(x\), \(y\), and \(z\)), the random forces come from a square distribution of amplitude \(F_A\) centered on zero. The central limit theorem \[55\] shows that the actual shape of the random distribution is unimportant; one will still end up with a Boltzmann distribution of particle velocities corresponding to the desired temperature. Thus, for example, at the start of some time step, particle \(i\) will have its force along the \(x\)-axis initialized to \(f_{ix}\), see Equation 13, where \(p_{ix}\) is the particle’s present \(x\)-momentum, and \(w\) is a random number, \(-1 \leq w \leq 1\).

\[
f_{ix} = -bp_{ix} + wF_A \quad (b > 0)
\]

(13)

9. Crack Behavior at Non-Zero Temperatures in Silicon

9.1. Vanishing Velocity Gap

Temperature implies energy fluctuations in time, so that if the crack gets trapped due to an energy fluctuation that reduces the fracture energy, further kinetic fluctuations may subsequently enable the crack to move on, giving rise to the possibility of creep. This suggests that the strength of lattice trapping should be a function of temperature. A series of simulations at 50 K intervals were carried out to investigate the effect of temperature on lattice trapping along (111) \([01\bar{1}]\). The results are shown in Figure 16. The lattice trapping/forbidden band of velocities hysteresis loop narrows while the lowest velocity before arrest remains the same as temperature increases, and eventually the velocity gap vanishes above about 200 K. It might therefore be necessary to go to liquid nitrogen temperatures in order to observe lattice trapping and a velocity gap. Lattice trapping has been postulated since the early 1970s \[13\] but has never been seen in experiment. Careful fracture experiments have not yet been carried out at low temperatures.

![Fig. 16. Velocity gap vanishes near 200 K. Dotted lines indicate forbidden velocities. All low-lying velocity states exist at room temperature. Cracking is along (111) \([01\bar{1}]\).](image)

Two further sets of runs were carried out at 300 K where the strain was quasistatically increased up to where the cracks became unstable. The first of these (Fig. 17) is a continuation of the 300 K segment in Figure 16. Notice that although all low-lying velocity states now exist, there still exists some dynamic forbidden band of velocities and hysteresis at higher velocities and strain.

![Fig. 17. Crack velocity profile with respect to quasistatic loading \(\Delta\) along (111) \([01\bar{1}]\) in silicon at 300 K. All low-lying velocity states exist.](image)
With all low-lying velocity states accessible at room temperature, as just mentioned, it might then be possible for the crack to creep. Whether a crack can creep for values of $\lambda$ all the way down to 1, or as the fracture energy density approaches the Griffith point, cannot be settled with molecular dynamics; the time scales required for thermal fluctuations simply are beyond the reach of computers.

Note also in Figure 17 that as $\lambda$ is decreased below 1 the crack actually heals and travels backwards. This is perfectly reasonable as there has not been any non-uniform surface damage, and no oxide layer has formed.

9.2. Direct Comparison with Experiment

The second full set of room-temperature runs was performed along (111) $\{\bar{1}1\bar{2}\}$ in order to allow direct comparison with experiment$^{[57]}$ (the zero Kelvin runs for this lattice direction are in Fig. 14). The results are in Figure 18, and were obtained for a thin strip of size $532 \times 15 \times 154 \AA^2$, periodic along the thin axis. As before, new material was added ahead of the crack tip and old material lopped off at the tail every time the crack advanced to within 200 Å of the forward end of the strip. In this fashion, the crack traveled 7 μm during the course of the simulations as $G$ was varied between 5 and 14 J/m². The room-temperature experiments for fracture in silicon wafers along (111) $\{\bar{1}1\bar{2}\}$ are described in the literature.$^{[57]}$ These experiments are difficult to perform because of the high Young’s modulus and brittleness of silicon.

The highest experimental and numerical crack velocities shown in Figure 18 are reasonably close, but the minimum fracture energies at which a crack propagates differ: 2.3 J/m² in the experiments and 5.2 J/m² in the simulations. Since the scale of crack velocities in a material is bounded by sound speeds,$^{[5]}$ which are given correctly by the SW potential, it is not surprising that the experimental and computational crack velocity scales agree. Furthermore, the potential gives the correct cohesive energy of silicon (but an inaccurate potential cannot be expected to correctly reproduce experimental cohesive energy curve, see Section 5.2), leading to the agreement in numerical and experimental energy scales. However, the nonlinear parts of the potential involved in stretching and rupturing bonds play an important role in determining the actual fracture energies and crack velocities, in particular where the crack arrests and what its highest velocity will be. The quantitative disagreements shown here point to a shortcoming of the nonlinear parts of the potential, which have not received much attention. The modified SW potential cannot be expected to correctly reproduce experimental data. But it highlights the control over brittleness in the three-body term, and the inadequacy of the potential tails in the two-body term. These are complicated matters, and yet perhaps only hint at greater difficulties on the road to a better potential.

The lowest fracture energy density at which a crack propagated in the experiments was close to the Griffith energy density for a (111) plane, 2.2 J/m².$^{[58]}$ Since a crack cannot travel with less energy, there must be a narrow range of fracture energy over which the crack velocity rises rapidly from zero to the lowest value measured, $\approx 2 \text{ km/s}$. This phenomenon is also seen in glass and polymers.$^{[14]}$ Because of the extreme precision required at the boundaries, experiments with silicon are not yet capable of settling the matter of whether this sharp velocity rise signals a velocity gap. However, as shown in Figure 16, in numerical silicon, the velocity gap is temperature dependent, vanishing above 200 K. At 300 K there do exist “steady states” at all velocities between 0 and 3 km/s. For glass and Plexiglas, carefully controlled crack arrest experiments rule out the existence of a velocity gap.$^{[59]}$

The silicon experiments covered a range of fracture energy densities, 2–16 J/m², in which the cracks produced very smooth surfaces. Thus, cracks in silicon can dissipate large amounts of energy, more than seven times the amount needed to create a clean cleavage through the whole crystal, without leaving behind any large scale damage on the fracture surfaces. Investigation by atomic force microscopy shows that for low fracture energies the fracture surfaces are flat on the nanometer scale, while at higher energies the surfaces have pronounced features. These features, however, are smooth on the micrometer scale, and account for height variations on the order of 30 nm over an area of 16 μm². The roughness gives an area increase of only ~0.1 % above that of a flat cleaved surface. This extra surface cannot account for the sevenfold increase in dissipated energy. The simulations, however, indicate that most of the energy can be carried off in lattice vibrations.

10. End Notes

It is possible to perform molecular dynamics simulations of fracture in a way that is directly comparable with laboratory experiments. A simple scaling argument demonstrates this. Every indication from such simulations is that the predic-
tions of lattice theory are correct. Molecular dynamics simulations, however, enable one to work with sophisticated interparticle potentials and crystal structures that are difficult to analyze analytically in the manner of lattice theory. More dynamic possibilities then open up.

Evolution has a long way to go in the zoo of interatomic potentials. How things will develop for silicon, at least, can be measured in part by the strong new test of comparing molecular dynamics simulations of fracture and laboratory experiments. For many years it has been computationally desirable to use potentials with a short cutoff. Such nearest neighbor potentials can describe fairly well quasi-equilibrium phenomena, like elastic constants and bulk defects, and perhaps even fortuitously or by design give a reasonable melt temperature. What has been demonstrated in this article, however, is that in order to characterize properly non-equilibrium phenomena, better longer-range potentials are required, in particular when modeling the severing of bonds involved in fracture. Although our discussion has centered on silicon, the methodology of comparing simulations with actual experiments should apply to a broad range of brittle materials. For any particular material, then, with a better potential for fracture, one which necessarily will reproduce the full density functional theory cohesive energy curve, and not just the quasi-equilibrium deep-well part, one should have a more transferable and correct potential than has hitherto existed.

Finding these corrected potentials has a relevance going beyond the particular details of how cracks behave in a material; indeed, the niceties of dynamic fracture are themselves usually of no practical importance whatsoever. More significant is the broad hope that materials of the future will be designed from the atoms up on computers. This hope is more widespread among those who use computers than among those who make materials. If it is to be realized, it must be based on physics that is correct and has been tested in difficult cases. Brittle fracture provides a test case that is simple enough to be tractable, yet demanding enough that attempts at quantitative prediction have so far been defeated. Ultimately, it will be necessary to crack this difficult matter in order to design and uncover the full potential of advanced materials.

Received: March 1, 1999
Final version: April 26, 1999