

# CHALLENGES AND STATE OF THE ART IN SIMULATION OF CHEMO-MECHANICAL PROCESSES

S. B. Trickey  
Quantum Theory Project  
University of Florida  
Gainesville, FL 32611-8345

P.A. Deymier  
Department of Materials Science and Engineering  
University of Arizona  
Tucson AZ 85721-0012

## Abstract

Planarization in semiconductor device fabrication is, from the perspective of materials physics and chemistry, an almost perfect example of a chemo-mechanical process. Such processes depend critically upon the interplay of chemistry and mechanical stress.

Atomistic simulations of chemo-mechanical processes with real predictive power are not yet available. An obvious barrier is computational cost. Underlying those costs is the inherently multi-scale nature of the problem. The challenges of straight-forward multi-scale simulation, while large, are reasonably well-understood, with major progress made already. The current big step is inclusion of chemical reactivity. A less obvious challenge is to treat the multiple time and length scales on an equal footing. Current multi-scale approaches start with a successful method developed at one primary length and/or time scale. One scale thus has priority. Overcoming this built-in bias to phenomena on a particular scale is the challenge of *scale parity*. Doing so will require seamless merging of quantum chemistry, surface physics, thermal physics, classical mechanics, and fluid dynamics to yield real materials science (not just solutions to model problems).

Here we summarize the state of the art at a tutorial level and try to give a useful assessment of prospects.

## INTRODUCTION

### Perspective

From the vantage point of computational materials scientists, chemo-mechanical processes are an intellectual gold mine of fascinating problems. But the challenges are many and immense, hence the perceptible intellectual horror expressed by such workers upon encountering these problems: quantum mechanics and classical mechanics both are important in the same problem. Ordinarily we can keep to the one or the other but not here. Another challenge is that chemical reactivity must be treated with both fluid and solid dynamics included. In contrast, most computational progress on reactions has been via calculations that correspond to the dilute gas phase. Another challenge is breaking of bonds, a characteristic that suggests strongly that models using clever potentials may not be adequate.

This category of problems also is both fascinating and scary for computationalists because of the great breadth of diverse phenomena it comprises. These include chemo-mechanical polishing and surface finishing of materials, stress corrosion etching of defects and dislocations in materials, slow crack growth and premature failure



of solids, stress-driven electro-chemical corrosion and weathering, pressure-driven chemistry (e.g. explosive detonation), radiation-induced failure and chemical reactivity, hydrogen embrittlement, and others. The common feature is an intertwining of mechanics, chemistry, surface physics, and thermodynamics through the underlying atomistic dynamics.

Looking on the optimistic side, progress in atomistic simulations, computational quantum chemistry, and surface electronic structure has brought us to being able to make a meaningful beginning on first-principles modeling of technologically important chemo-mechanical processes. Therefore we will deal with the challenges by describing what one would like to do, what pieces of that puzzle exist today, what is missing, and, where possible, what the plausible routes to success may be. We hope to describe those issues that, from our side of the endeavor, characterize the divide between where we in the simulation, modeling, and predictive computation community are today and what we believe engineers would like us to deliver.

Two recent semi-popular papers may be helpful collateral reading. Abraham et al. (1) have given a good picture of the state-of-the-art up to two years ago for simulation of brittle fracture and Bernholc (2) has summarized progress at the quantum mechanical end of the effort up to about the same time.

### Assumptions and Requirements

We assume that the objective is an *inclusive* simulation. It would address the following problem: Given a comprehensive description of a specific chemo-mechanical process, provide reliable computed predictions of the key outcomes of that process *before* it is tried in the laboratory. We assume that the comprehensive description would provide such information as the size ranges, composition, and state of the solid elements, the chemical and hydrodynamical nature of fluids in the interface, and the range of speed of motions and their duration. Even the ability to distinguish which processes have high or low prospects for meeting a specified set of needs would be a success. We assume therefore that the simulations must be *predictive*. Obviously it is valuable to parameterize data in models that can be used as design tools but that is not our focus.

Predictive simulations must be *atomistic* and from *first principles*. *Atomistic* calculations are those for which the inputs are determined by natural properties at the atomic level. They are inescapable because materials properties eventually trace to those of the constituent atom. *First principles* calculations are those in which any physically or chemically significant parameters are determined independently of the simulation and prior to it. There are no free parameters to adjust in order to improve the comparison with experiment.

One of the requirements is that inclusive simulations not be computational grand challenges! That is, the simulation must be affordable on at most a Tier-3 supercomputer (roughly, a 16 to 32 node message passing machine using superscalar processors). A better situation would be a two to four processor workstation. Affordable means that the time to run any single simulation is no more than a few days. Simulations that take months may give valuable insights but hardly are practical as technological development tools. We also require that the inclusive simulation software not need a team of experts to operate. At the moment that requirement is far from met.

## FUNDAMENTALS - INGREDIENT METHODS

Since it must start at atoms and end up with predictions about material samples of ponderable size (e.g. a wafer), the inclusive simulation clearly is involved with the



issue of length scales from the start. The issue becomes particularly evident upon considering the methods that must be ingredients in the inclusive scheme. Multiple time scales are not quite as obvious, but they are there.

### Finite Element

Start at the sample level, planarization or polishing for example. At these length scales atoms are invisible as such and continuum mechanics treatments are the tool of choice. Key system and material parameters - density, viscosity, bulk modulus, etc., develop as a function of time according to differential equations for mass motion, fluid flow, concentration and so forth, subject to appropriate boundary conditions. The time scales involved may range from milliseconds to minutes or longer. The way this domain is treated varies but the underlying “finite element” scheme is common to all approaches.

Focus on the kinetic and elastic energy densities for the continuum. Both are functions of the displacement field and therefore for all times  $t$ , the system energy is calculable as an integral over the continuum volume. The integral is handled numerically by introducing a mesh of points, called nodes, throughout the volume. Assume that the displacements  $\mathbf{u}_i$  and velocities  $\dot{\mathbf{u}}_i$  are known at all the nodes  $i$ . The nodes define irregular tetrahedra that fill the system volume (to whatever accuracy is desired or affordable). Within each tetrahedron, the required energy integrals can be approximated by introducing a scheme to interpolate from the values of variables on the nodes to all of the interior of the tetrahedron. The result is a discretized expression for the energy of the whole object

$$E = \frac{1}{2} \sum_{i=1}^{N_{\text{cells}}} \sum_{\mu\nu} \{ u_{\mu,i} \mathcal{K}_{\mu\nu}^i u_{\nu,i} + \dot{u}_{\mu,i} \mathcal{M}_{\mu\nu}^i \dot{u}_{\nu,i} \} \quad [1]$$

The matrices  $\mathcal{K}^i$  and  $\mathcal{M}^i$  are the “local stiffness” and “local mass” matrices that characterize the particular cell as a result of the tetrahedral integration.

From a computational point of view this discretization is much more than conversion of an integral into a numerically manageable finite sum because it introduces a particulate character to the continuum problem. Once we have these macroscopic pseudo-particles (for lack of a better term), the energy expression can be treated as a Hamiltonian for those pseudo-particles and it follows that we can derive dynamical equations for them. That opens the opportunity for coupling the dynamics of the pseudo-particles to the dynamics of atomic-scale particles. We defer discussing how such coupling is done.

### Molecular Dynamics

Below the resolution of discretized continuum mechanics in the finite element calculation, the continuum behavior is determined by atomistic interactions and dynamics. Fortunately, these are relevant only in particular regions. In a solid-solid polishing process, the special region obviously is the rubbing interface. In brittle fracture, it is the crack tip. In such regions the coordinated motions of many individual atoms are critically important. One of the great successes of statistical physics is the insight that, for the most part, those motions can be treated by classical mechanics. So, suppose there are  $N_a$  atoms, typically between a few hundred and a million. In some convenient coordinate system, atom  $i$  has a position  $\mathbf{R}_i$  and velocity  $\dot{\mathbf{R}}_i$  or, collectively  $\mathbf{R} = \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \dots, \mathbf{R}_{N_a}$  and  $\dot{\mathbf{R}} = \dot{\mathbf{R}}_1, \dot{\mathbf{R}}_2, \dot{\mathbf{R}}_3, \dots, \dot{\mathbf{R}}_{N_a}$ .



At any particular configuration, the atoms have an interaction energy  $V(\mathbf{R})$ . Then Newton's second principle is

$$\mathbf{F} = m_i \ddot{\mathbf{R}}_i = -\nabla_i V(\mathbf{R}) \quad [2]$$

Given the initial and boundary conditions, this is a well-determined problem involving a set of  $3N_a$  partial differential equations.

All the materials physics and chemistry are in the interaction energy  $V(\mathbf{R})$ , so clearly it is crucial. Since even the general three-body problem is insoluble in Newtonian mechanics, computing is inescapable. Inevitably this means conversion of the time variable to discrete steps. Obvious issues include the boundary conditions. Perhaps less obvious but highly important are matters of sample size (number of "atoms" in the sample), number and size of time steps, and fast, accurate algorithms.

First consider discrete time steps: the time derivatives are replaced by finite difference scheme which steps the particles of the system forward (or backward) in time along their trajectories. A variety of time integrators has been developed to solve for the particle trajectories, with the algorithms designed in terms of a compromise between speed and accuracy at each step (3). Time integration steps  $\delta t$  typically are on the order of  $10^{-16}$  to  $10^{-15}$  sec. The number of time steps is closely linked with the number of particles in the sample, the characteristic time  $\Delta$  of the properties being investigated, and the available computational resources. If  $\tau$  is the compute-time per step per particle, then the total run time is

$$T = N_a \times (\Delta / \delta t) \tau \quad [3]$$

With the advent of high-performance computing platforms, MD simulations now can be done on systems with several hundred up to a few tens of millions of interacting particles. Evidently even the largest systems of these is small compared to Avogadro's number, hence does not comprise a sample of macroscopic dimensions. At present it is fairly typical to run a few thousand to a few tens of thousand time steps, corresponding to a range of treatable physical time intervals of about 1 – 10 femtosecond. In many instances, that interval is the dominant consideration. Calculation of fluid transport, for example, requires long simulations to obtain good statistical convergence. Required duration of physical time may even be a barrier to any applicability of MD. Slow relaxation processes, such as those characterizing the glassy state of matter, diffusion in macromolecular systems, or the spontaneous formation of spatially organized structures, are examples of problems with temporal scales many orders of magnitude larger than those currently treatable via MD.

Faster computers obviously help but to expect machine speedups by many orders of magnitude is unrealistic. The challenge is illustrated by the same considerations. Suppose that the problem of interest were to need a physical duration of 500 fs and a sample of at least  $2 \times 10^6$  particles and that the goal is to be able to run at least one simulation every 24 hours. Then the entire simulation algorithm must run at  $8.7 \times 10^{-8}$  seconds per time step. On a 1 Teraflop machine this rate caps the *entire* code at a budget of 87,000 floating point operations and/or their equivalent in loads and stores, a ridiculously small budget of operations compared to any current computer code. Thus, the issues of MD sample size and their interplay with physical duration limitations raise many unresolved questions concerning the size-scale and time-scale parity of a multi-scale description of chemo-mechanical processes.

The boundary conditions deserve brief mention. To simulate bulk material one wants to eliminate effects specific to the boundary surface. Periodic boundary condi-



tions do that by having a particle re-enter from, say the left, just as it exits from the right of the simulation cell. The choice of boundary conditions for the special regions of an inclusive simulation can be a harder problem; see “handshaking” below.

Finally, MD simulations can be conducted in a variety of statistical thermodynamical ensembles. Constant temperature and constant stress ensembles are particularly relevant to chemo-mechanical processes. In addition to equilibrium MD, non-equilibrium MD can provide information on the response of a simulated system to some external stimulus. Both stimulus and response can be mechanical in nature but cross field stimuli and responses are possible. Particular attention has to be paid, however, to the energy transferred to the simulated system by the stimulus and often, therefore, algorithmic thermostats are included in the model to maintain the system under stable conditions.

### Electronic Structure vs. Potentials

An obvious determiner of the quality of the MD calculations is the potential  $V(\mathbf{R})$ . What information it must contain and how that information is introduced are questions that can be rephrased usefully but seemingly opaquely: how much detail of the structure of the atoms must be included and with what accuracy?

To respond requires examination of at least a sample of the processes in chemo-mechanical processing. Organic additives are used during the electrodeposition process to produce void-free copper-filled trenches. These additives are growth inhibitors for the electrodeposition process. However, the effect of residual additives (in the ppm) on the grain structure/grain growth of the copper is unknown. During the chemo-mechanical planarization itself, the relevant processes include slurry chemistry, stress-corrosion cracking, and stress-induced dissolution and recrystallization. Even the most cursory look at this arbitrarily abbreviated list makes clear that in the region where the action is taking place the chemistry must be handled correctly. For the great majority of chemical phenomena, electron behavior is the determiner. Thus we are led to the treatment of many-electron systems and, implicitly, an intrinsically quantum mechanical region of the inclusive simulation.

In the quantum mechanics of many-electron systems, a powerful and physically sensible simplification is the Born-Oppenheimer approximation. Essentially it says that the nuclei in a system are so much more massive than the electrons that it is an excellent approximation to assume that the electrons respond instantaneously and perfectly to the nuclear motion. Under that assumption, the Schrödinger equation for the electrons depends on the set of nuclear coordinates  $\mathbf{R}$  as parameters:

$$H(\mathbf{R}; \mathbf{r})\Psi_j(\mathbf{R}; \mathbf{r}) = E_j(\mathbf{R})\Psi_j(\mathbf{R}; \mathbf{r}) \quad [4]$$

Here the  $\mathbf{r}$  are all of the electron coordinates and all the quantum numbers are bundled into a generic index  $j$ .

To make the inclusive simulation problem manageable, it is most common to assume that everything takes place in the ground state of the electrons,  $j = 0$ . Then the tremendous simplification of the Born-Oppenheimer approximation is that the potential  $V(\mathbf{R})$  used in the MD eq. (2) is just the electrons' ground state energy for the current configuration of nuclei  $\mathbf{R}$ :

$$V(\mathbf{R}) \approx E_0(\mathbf{R}) \quad [5]$$

In this way of looking at the problem, the choice is how to generate the ground state energy. Retain the electrons explicitly or hide them in a sum of potential functions?



Retaining them means that somehow we must do the quantum mechanical problem in eq. (3), which is hard. Hiding them means that we must have a valid way of generating the potential functions, which turns out to be harder than it sounds.

### Explicit Treatment of the Electrons

Electrons interact by the Coulomb potential  $V_{coul}(|\mathbf{r}_1 - \mathbf{r}_2|) = ke^2 / r_{12}$  and their states must obey the Pauli principle,  $\Psi_j(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i \dots \mathbf{r}_j \dots \mathbf{r}_N) = -\Psi_j(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_j \dots \mathbf{r}_i \dots \mathbf{r}_N)$ , that is, be antisymmetric under interchange of any two electrons. This set of requirements makes the solution of the many-electron Schrödinger equation a tough business: even a single H<sub>2</sub> molecule requires an approximate treatment. Development of high-quality, systematically improvable approximations is the area of electronic structure (surface, solid physics) and quantum chemistry (molecules, clusters).

Ideally, the inclusive simulation would do one MD step  $\delta t$ , thereby giving the new configuration  $\mathbf{R}(t + \delta t)$ . Then an electronic Schrödinger equation module would solve for  $E_0[\mathbf{R}(t + \delta t)]$  and its gradients  $\nabla_i E_0[\mathbf{R}(t + \delta t)]$  and that information would go back to the MD module to advance the next step. MD of this sort, Born-Oppenheimer MD (BOMD), does exist (3), with a key *proviso* that the electronic structure problem is not solved exactly; see discussion below. It is easy to understand why. The time to solution of most electronic structure methods scales as some relatively high power  $\mu$  of the number of electrons  $N_e$ . If programmed as a standard generalized matrix eigenvalue problem, density functional theory (“DFT”) goes as  $\mu = 3$ , Hartree-Fock (“scf”) as  $\mu = 4$ , and very-high-accuracy methods such as many-body perturbation theory, configuration interaction, and coupled-cluster theory (5) go as even higher powers. (Hartree-Fock omits the detailed correlation of electron motion, so is largely irrelevant.)

Furthermore, the very-high-accuracy methods are applicable today only to moderately small molecules (up to say a dozen nuclei and a few hundred electrons) and not to surfaces and solids. It is not atypical for modern electronic structure codes to require run times from multiple seconds for simple systems up to days for very complex ones. Even the optimistic assumption of a total of one second run time for the combined electronic structure module (to produce each solution  $\nabla_i E_0[\mathbf{R}(t + \delta t)]$ ) and the MD code to step particle  $i$  forward leads to a prohibitive result. On a 1 Teraflop computer, a simulation including only 100,000 particles for 4,000 time steps would require  $4 \times 10^8$  seconds or 12.6 years. To echo the previous section, if the time-step is  $\delta t = 1 \times 10^{-15}$  s (a fairly reasonable value), then 4000 time steps would yield only  $1 \times 10^{-12}$  s of the physical history of the system. Since the electronic structure calculation for a new configuration requires far more time than the MD calculation of the next configuration, attention focuses on making the electronic structure calculation faster or avoiding it altogether.

Present thinking is that the only long-term route to using quantum mechanics explicitly within standard MD is to achieve much faster performance for small  $N_e$  with an algorithm that also scales linearly with  $N_e$  i.e.  $\mu = 1$ . Another jargon term for linear-scaling is order-N methods. The first step to achieving order-N scaling is to reduce or reformulate the quantum mechanical problem to an equivalent effective one-electron problem. In such approaches, there is an effective one-electron Schrödinger equation

$$\left[ -\nabla^2 + V_{eff} \right] \phi_j(\mathbf{r}) = \varepsilon_j \phi_j(\mathbf{r}) \quad [6]$$



whose effective potential and eigenfunctions (“orbitals”) are related in a method-dependent way to the ground state of the system.

Several such schemes exist. One, Density Functional Theory, is exact in principle; see Görling et al. (6) for a review and original literature. The basic theorems of DFT are that the ground state energy of an  $N_e$  electron system can be determined variationally with respect to the electron density without explicit knowledge of the corresponding Schrödinger state:

$$\delta E_{HK}[n] = 0 \Rightarrow n = n_0, E_{HK}[n_0] = E_0 \quad [7]$$

with  $n$  the number density given by  $n(\mathbf{r}) = N_e \int d\mathbf{r}_2 \dots d\mathbf{r}_N \Psi_0^*(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \Psi_0(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N)$   $E_{HK}[n]$  is the Hohenberg-Kohn density functional. DFT is converted into a one-electron effective Schrödinger equation scheme via the Kohn-Sham procedure. It assumes there is a system of “non-interacting electrons” whose number density is the same as the real system. Then the effective potential in eq. [6] becomes

$$V_{eff}(\mathbf{r}) = V_{ne}(\mathbf{r}) + V_{ee}(\mathbf{r}) + V_{xc}(\mathbf{r}) \quad [8]$$

and it can be proved that the DFT variation principle corresponds to the statement that the nuclear-electron, electron-electron coulomb repulsion, and electron-electron exchange-correlation potentials are functional derivatives of the corresponding terms in the DFT energy expression:  $V_{ne}(\mathbf{r}) = \delta E_{ne} / \delta n$ ,  $V_{ee}(\mathbf{r}) = \delta E_{ee} / \delta n$ ,  $V_{xc}(\mathbf{r}) = \delta E_{xc} / \delta n$ . With such potentials eq. [6] is the Kohn-Sham equation.

The expressions for  $E_{ne}, V_{ne}, E_{ee}, V_{ee}$  are familiar from classical physics. Of course there is no free lunch. DFT escapes from having to construct the Schrödinger  $\Psi_j$  at a cost: the theorems only establish the existence of  $E_{xc}$ . They do not provide a route to its construction. Therefore much effort has gone into finding workable, reliable approximations. While not a settled issue, by now there are reasonable ways to get chemically useful solutions.

It is relatively straightforward to implement high-precision solutions of the KS equation with order  $N^3$  scaling. Several such implementations actually scale better for relatively small numbers of electrons but are not fast enough in the absolute sense (too slow for small, simple systems.) Development of order-N methods is current research. At present, there are two routes to implement DFT directly in MD. The plane-wave plus pseudopotential approach expands the KS orbitals in plane waves and removes the core electrons (which have little effect on the chemical behavior of the system) from consideration via norm-conserving pseudopotentials. With these approximations, MD can be implemented so as to follow the resulting ground electronic state energy  $E_0(\mathbf{R})$  exactly. This is the BOMD scheme mentioned above.

An earlier (and presently more popular) scheme that follows the ground state electronic energy (again in the pseudopotential approximation) approximately is due to Car and Parrinello (7). Their approach arose when BOMD was not understood to be practical. By converting the Kohn-Sham problem into a fictitious electron dynamics coupled with real atomic dynamics (all at non-zero temperature), their scheme transcribes the problem of finding the DFT  $E_0(\mathbf{R})$  into an MD simulation. Specifically they obtain a set of coupled Newtonian equations of motion for both the KS eigenfunctions and the atomic coordinates. These equations of motion are derived from a Lagrangian written as a functional of the KS eigenfunctions and the nuclear positions as dynamical variables. A fictitious electron mass is introduced in order to define an artificial kinetic energy associated with the dynamics of the KS eigenfunction



degrees of freedom. Of course, the Lagrangian also incorporates the kinetic energy of motion of the nuclei. The interaction between electronic and atomic degrees of freedom is embedded in a potential energy derived from DFT. The equations of motion of the wave functions and the atoms are solved numerically. The dynamics of the electronic degrees of freedom are fictitious; the MD scheme is used to sample configurational space for the nuclei and variational degrees of freedom for the KS eigenfunctions. The scheme gains speed over conventional electronic structure methods because, in contrast with them, it does not use matrix diagonalization or self-consistent iterations. The advantage is that the ion-motion evolution is built right in. The price is that one must choose the fictitious masses for the KS degrees of freedom to be appropriately small so as to ensure that the Born-Oppenheimer adiabatic approximation is not violated significantly. To ensure that is the case it also is necessary to quench the system repeatedly onto the ground electronic energy surface. Both are matters of non-trivial computational art.

An alternative scheme is tight-binding. In it the one-electron equation is converted into a matrix eigenvalue problem for the real system orbitals in terms of linear combinations of atomic orbitals. The matrices themselves are integrals of the original Hamiltonian operator with respect to products of the atomic orbitals, the “matrix elements.” These are fit, as a function of interatomic separation, to results for perfect and defective crystals. As a result, the effective one-electron eigenvalues are calculable, approximately, as a function of the current nuclear configuration  $\varepsilon_j = \varepsilon_j(\mathbf{R})$ . Then the potential is completed by writing

$$V_{TB}(\mathbf{R}) = \sum_j \varepsilon_j(\mathbf{R}) + \sum_{\mathbf{R}_{ij}} V_{rept}(\mathbf{R}_{ij}) \quad [9]$$

where the repulsive potential that corrects the eigenvalue sum also is fit to a database. Tight-binding has been used a great deal, for example Ref. (1). A related scheme that is a considerable step beyond tight binding and that has been quite successful in predicting structure of very large molecules and has a lot of promise for driving MD simulations is semi-empirical. In the typical semi-empirical method (8), properties of the quantum mechanical overlap matrix (between basis functions) are used to determine the relative importance of the matrix elements and to give controlled approximations to them, with certain subsets calibrated to experimental data for electron affinities, ionization energies, and related atomic properties.

### Implicit Treatment of the Electrons: Potentials

If one were unaware of electrons but knew that atoms and molecules interact, the obvious way to do MD simulations on assemblies of particles would be purely classical mechanical. In this approach, one simply assumes there is a potential  $V(\mathbf{R})$  that corresponds to those interactions. Such interatomic (and intermolecular) potentials are a venerable topic (9) that continues to the present. The total potential energy may be written in the general form:

$$V(\mathbf{R}) = V_1(\mathbf{R}) + V_2(\mathbf{R}) + V_3(\mathbf{R}) + \dots \quad [10]$$

where the first term corresponds to an external force field, the second is a sum of pairwise interactions, the third is the sum three-body interactions, etc.

Some additional simplifying approximations often are made. In the absence of the external potential, the potential function is often limited to the sum of pair interatomic potentials,  $V_2$ . Potentials in that family range from the well-known Lennard-Jones (L-J) potential for rare gases, to the long-ranged Coulombic potential augmented by a



short ranged repulsive potential such as the Born-Mayer potential (10) for interactions between ions. Many materials systems with covalent bonding cannot be modeled by simple pair potentials and higher order terms in the potential energy function have to be included such as three-body interactions. This is the case for instance of the Stillinger-Weber potential for Silicon (11). Metals have received a great deal of attention. Potentials for metals range from simple empirical pair potentials such as L-J potentials or Morse potentials to more realistic pair potentials based on pseudopotential and perturbation theories for simple metals (12) to empirical many-body potentials derived using the Embedded-Atom method (EAM) (13). EAM potentials have had some success in the simulation of transition metals such as Cu, Ni,

Empirical potentials generally are parametric functions of positions of the atoms. The parameters are obtained by fitting some set of calculated physical properties of some materials to experimentally measured values. This approach, however, does not ensure the transferability of the interatomic potentials to conditions differing significantly from those under which the parameters were obtained. It is therefore greatly desirable to obtain potential functions from first-principle quantum calculations or to perform potential-less MD simulations from first-principles as discussed in the preceding sub-section.

## FUNDAMENTALS - SCALE PARITY

### Sequential and Concurrent Multi-scale Simulations

In reverse order of scale, the logic so far goes as follows. All the chemical activity in the chemo-mechanical process takes place in some bounded zone. (Determination of that zone in a predictive way is itself a challenge.) Within that zone we handle the chemical reactivity either by doing an appropriate electronic structure calculation or by replacing it with a suitably chosen approximate potential. In a much larger but still not macroscopic zone, we run an MD calculation. It provides a classical mechanical picture of the sub-microscopic particle motion. With suitable assumptions it also provides thermodynamic insights. With the MD information as input in critical regions, we build a finite element scheme that approximates the continuum scale of system behavior.

Calculations of this sort have been and are being done. If linked by some systematic set of assumptions, they usually are called multi-scale simulations (1). Multi-scaling recently has received much attention in several branches of physical science. In materials, a large part of the work is devoted to modern simulation methods and coupling of (mainly) length scales. Currently favored approaches can be characterized as serial or concurrent.

What we have described so far falls mostly in the concurrent category. Such methods rely on different computational methodologies applied to different regions of the material. For example, crack propagation is a problem that was tackled early on (and still is) by atomic simulation techniques near the tip of a crack where large deformations (even bond breakage) occur plus continuum approaches at large distances from the crack tip. This approach is implemented by partitioning the space into two regions: the tip of the crack (where the material is treated atomistically with molecular dynamics) and the rest of the material (modeled as a continuum via the finite element method) (14, 15).

In serial methods a set of calculations at a fundamental level (small length scale) is used to evaluate parameters for use in a more phenomenological model that describes a phenomenon of interest at longer length scales. For example, atomistic simulations can be used to obtain the constitutive behavior of finite elements, which are then used to simulate larger scale problems. Several research groups are presently working productively on such methods, and several applications can be found (16).



As already discussed, sometimes the same methods are used in both types of simulation. For example, DFT calculations can appear both as an integral part of a concurrent simulation or as a way to parameterize the potentials used in MD.

Identification of the regions to be treated at different scales in concurrent multi-scale modeling requires prior knowledge of the important factors controlling the process under investigation. Consider, for instance, the simulation of crack propagation. Here, we may use the magnitude of the displacement as a criterion for delimiting the various regions. It is known that the large atomic displacements (i.e. bond breakage) take place at the tip of the crack, leading to the need for a quantum description of that region. Dislocations are often emitted at the tip of the crack, hence an atomistic (MD simulation) area needs to be applied to the model beyond the quantum region. Where displacements are small and vary smoothly, a continuum region may be sufficient. Concurrent multi-scale models must address the problem of the interface between the different regions by devising some appropriate boundary conditions (e.g. a condition that allows for dislocation motion across the interface between the atomistic and continuum regions). Even with appropriate boundary conditions, a criterion based on displacement magnitude, however, may be insufficient in that it may not allow inclusion of the complete physics of the process such as cooperative phenomena with long-range spatial correlation.

Once the regions are identified it remains to link the methods appropriate to each across the interfaces between them. Jargon for this task is handshaking or embedding. In the sketch of Finite Element (FE) methodology we hinted at the possible handshake to an MD region. For example (1), the FE tetrahedra can be reduced in size as the boundary is approached so that at the interface their edge lengths are appropriate to bond lengths. Real particles on the MD side that are nearest neighbors to FE nodes link via weighted potentials: the FE algorithms see an external driving force from the MD side and conversely. At the interface with the electronic structure region the methods are similar, with the caveat that dangling bonds have to be terminated in an unbiased way.

Multi-scale approaches are demanding, so it is no surprise that the majority of today's concurrent schemes rely on evolutionary refinement of single-scale approaches. Such calculations have a problem which can be called scale parity. Historically, each method was developed to treat one set of phenomena and their corresponding critical length and time scales. Electronic structure approximations evolved largely independently of MD, as did MD with respect to FE. When multi-scale problems are addressed, it usually is the case that one method predominates (perhaps because of researchers' expertise). In such cases, behaviors at other length and time scales are omitted or approximated crudely. Thus multi-scale approaches today are distinguished from one another mostly by the particular scheme they adopt for interfacing the primary and secondary methods.

### Scale Parity

These approaches clearly are immensely valuable. They cannot, however address materials problems in which mechanical and chemical processes are connected inextricably. Those problems demand a simulation approach that fully integrates quantum mechanics (to handle reactivity and surface phenomena), ensemble simulations (to get large-scale atomic arrangements and thermal effects right), and mechanics (to get the human-scale properties right) without any primary-secondary method distinction.

The importance of the scale parity concept can be illustrated via a hypothetical one-dimensional function that represents the spatial variation of some property,  $E$  (e.g. energy). Call that variation the map of  $E$ . For this exercise it is obtained from



some model, henceforth model “1”. Imagine that this map represents a physical system with three characteristic scales, fine, intermediate and coarse. The inset in Figure A is a representation of this hypothetical map obtained by the linear superposition of three sinusoidal functions with three periods, small, intermediate, and large.

Now discretize the spatial axis (X) into 512 points and use wavelet analysis to identify the relevant scales for the map of E. A general goal of wavelet analysis is to describe graphs (signals, images) as a superposition of elementary functions each with specific scaling characteristics. A wavelet transform therefore is characterized by the values of the coefficients that weight the elementary functions. The wavelet transform of the map of E produces 512 coefficients. The first 256 coefficients correspond to the decomposition of the energy map at the finest scale, that we call  $n = 9$ . This scale is associated with features in the map with dimensions on the order of  $1/512$ . The next 128, 64, 32, 16, 8, 4, 2, and 1 entries are for the scales 8, 7, 6, 5, 4, 3, 2, and 1. Scale 1 is therefore characteristic of features in the map with spatial dimensions  $\sim 512$ .

The value of energy,  $E^*$ , associated with each particular scale,  $n$ , can be calculated by performing an inverse wavelet transform with only the specific set of wavelet coefficients associated with that scale. Thus use of only the second set of 128 wavelet coefficients yields the contribution of scale 8 to the energy map. Summing all the values of energy from this limited map yields  $E^*(n=8)$ . This approach allows us to identify which structural features in the map of E contribute most to the total energy. The total energy versus scale for that system is illustrated in Fig. A. It exhibits three maxima, at scales  $n=8, 6$ , and  $3$ . They represent the amount of energy associated with features at the fine, intermediate and coarse scale, respectively. We now may consider an approximate model “2” that describes the physical system over only the intermediate and coarse scales. Such a choice might correspond to the results from a finite element calculation. An energy map produced by this model (not shown in Fig. A) may be obtained by the superposition of only two sinusoidal functions with the intermediate and large periods. This second model, also, represents the physical system with the same spatial dimensions as the first one. Space, however, for this second model is discretized into only 128 points, thus the wavelet transform produces 128 coefficients. Out of the 128 coefficients, 64, 32, 16, ..., 2, and 1 represent the phenomenon at scales  $n=7, 6, 5, \dots, 2$ , and 1. Calculation of E at the different scales allows us to report the total energy as a function of scale for that second model in Fig. A. Its energy shows two peaks at the intermediate scale, 6, and the coarse scale, 3. This model is unable to represent the physical system at fine scales 8 and 9. Now consider model “3”. It models the system over a region of space half the size of the other two. Space is discretized into 256 points. The energy map is the sum of the two sine functions with the fine and intermediate periods. This model might correspond, for instance, to the result of a molecular dynamics simulation over a selected region. The fact that this model is half the size of the other ones then would correspond to limitations in computer capabilities that precluded MD simulation of large molecular systems. Model “3” produces 256 wavelet coefficients. The finer scale of model “3” is equivalent to the finer scale of model “1”, so the first group of 128 wavelet coefficients is associated with scale 9, the next group of 64 with scale 8, ... , with the last coefficient corresponding to scale 2. Scale 1 is not defined for model “3”. The total energy versus scale has two maxima at  $n=6$  (intermediate scale) and  $n=8$  (fine scale). We can now use the fact that the approximate models “2” and “3” overlap in scale to construct a compounded model statistically equivalent to model “1”. The intermediate scales (around  $n=6$ ) are common to both models “2” and “3”; this overlapping is essential for bridging the two models and to establish scale parity.

To get scale parity will require, *at the least*, use of quantum mechanical methods to simulate the chemical physics in the presence of mechanics (both ensemble and



continuum) and hydrodynamics. In turn this means a requirement for balanced, seamless merging of quantum chemistry and surface physics with classical mechanics and fluid dynamics. Implicitly the focus is on surfaces and interfaces of solid materials in an environment that can be gaseous, liquid or solid.

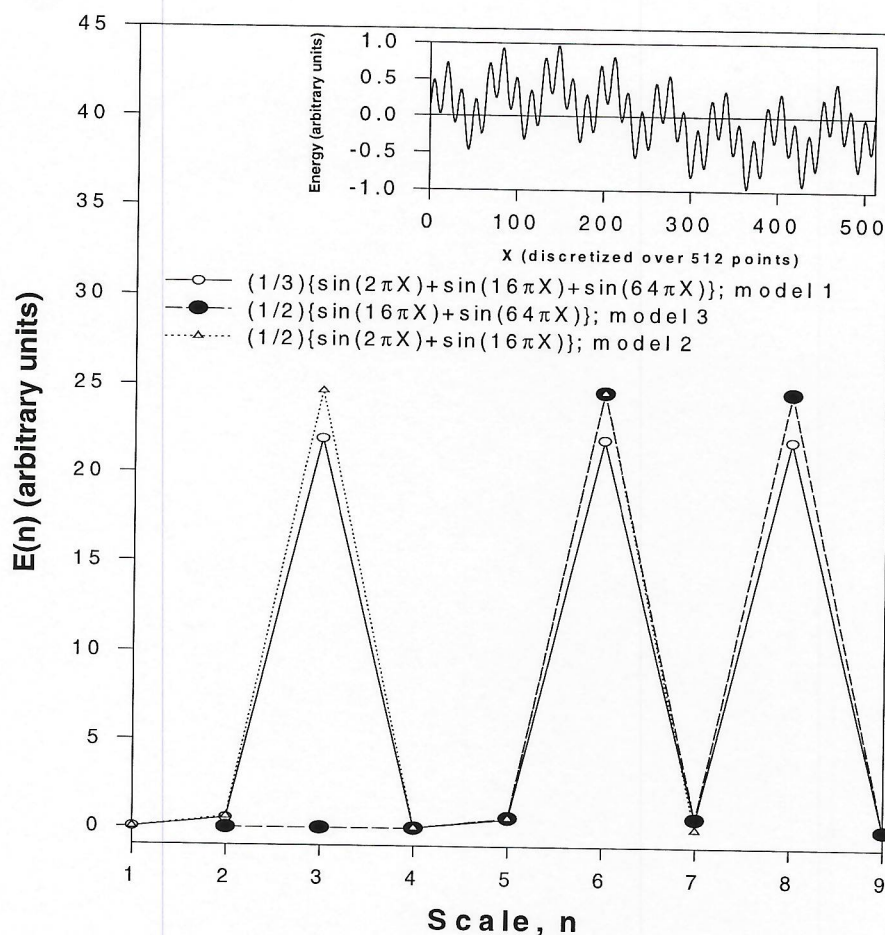


Fig. A: Illustration of scale parity through hypothetical 1-D energy maps. Total energy  $E$  as a function of scale is shown for three separate models. The inset shows the map of model 1, where the horizontal axis denotes space,  $X$ , discretized at 512 points. The vertical axis of the inset is “energy” given by the superposition of three sine functions with different periods. Model 2 contains only the intermediate and large periods and uses a discretized space of 128 points. Model 3 contains only the intermediate and small periods, defined on a region  $X$ , half of that of model 1 and 2 and with 256 points.

### FUNDAMENTALS - MISSING PIECES

At least some of the pieces missing from current simulation efforts are obvious. First, the great majority of quantum mechanical treatments of chemical reactivity are for dilute gas phase interactions. Dilute is an understatement, since most calculations are for as few formula units as the basic reaction equation requires: calculation of a diatomic reaction  $A + B \rightarrow C$  is done with one  $A$  atom interacting with one  $B$ . The calculations are purely quantum mechanical, not quantum statistical mechanical, therefore implicitly are at  $T = 0$  K. What is omitted are concentration effects and solvent effects. Some progress has been made on solvent effects using simple dielectric cavity models and similar schemes (17,18). These approaches seem to be adequate to predict molecular structures, rank reactive sites, etc., but whether they are realistic



enough and simultaneously can be made fast enough to support an inclusive simulation is, we think, an entirely open question.

State-of-the-art surface electronic structure calculations have two of the limitations of molecular calculations:  $T = 0$  K and either no solvents or simplified solvents. Additionally, relatively little has been done on the surfaces of disordered systems; ordered systems are the norm.

Third, almost nothing is known about chemically realistic calculation of solvent effects that include solvent dynamics, that is, fluid flow. MD simulations of fluid motion and continuum treatments by computational fluid dynamics (CFD) today are specialties rather different from the simulation and modeling sketched above. Another area of simulation that today is a separate specialty and in which relatively little has been done is atomic-level calculations of bare (i.e. no solvent or reactivity contributions) frictional effects. So far as we are aware, there is no work at the electronic structure level on this problem; it is not even obvious how to formulate the problem.

Next there are confinement effects on reactivity. Steric effects long have been a staple of chemistry but only recently has experimental evidence developed that shows that rates and reactivities can be modified quite dramatically when one or more reactants are confined. A reasonably accessible summary is in Ref. (19). The existence of such confinement effects may be problematic for potentials and parameterized electronic structure schemes.

We have noted the interplay between run-time and duration of physical phenomena already. Although MD simulations can be parallelized to help cope with ever larger systems (to date,  $\sim 10^6 - 10^8$  atoms), little has been done in terms of time scale-up. Finding a solution to that issue is particular pressing for processes involving infrequent events. Several approaches have been proposed to extend the length of time achievable with molecular dynamics simulations. Among those we can cite the hybrid molecular dynamics method (which does periodic Monte Carlo tests to allow time acceleration) (20), or parallel MD of replica to obtain long-time transition-time distributions (21), or temperature-accelerated dynamics (22). In addition to the issue of time acceleration, the problem of time scaling is intricately related to spatial scaling. Small MD systems do not exhibit the same relaxation time dependencies as larger systems. This results from the neglect of long-range cooperative processes in small systems. For example, consider the problem of normal grain growth in polycrystalline materials. The kinetics, with anisotropic grain boundary energy (i.e. low-angle and high-angle grain boundaries), is affected by the existence of clusters of grains separated by low-energy grain boundaries. To account for the spatial variability of grain boundary energies, the microstructure possesses a new length scale that characterizes the cluster size. This cluster length scale, larger than the mean grain size, is at the origin of a slower kinetics (23). MD simulation of grain growth in systems with dimensions exceeding the mean grain size but smaller than the cluster length scale would exhibit an incorrect kinetics. This problem would be particularly acute when treating a system with a concurrent multi-scale method that may not capture the interplay between the atomic scale processes and the cooperative processes.

## BENCHMARK PROBLEMS

In most areas of simulation and modeling, there are benchmark problems that are valuable for interpretation, helpful as screens against invalid approximations, and useful tools for timing competing methods. Often these are for “toy” problems, problems too simple to be of inherent interest but important as test cases. For MD there is “Lennard-Jonesium”, i.e. a fluid or a solid made up of spheres that interact via the L-J pair potential. Computed results can be compared against values measured



on liquid or solid Argon. For the quantum chemistry of molecules, the light diatomic molecules, such as  $H_2$ ,  $O_2$  are used as quick test cases, because results for them are known to high precision and their symmetry enables various predictions to be found "by hand." For the quantum dynamics of molecules, simple homonuclear rearrangement reactions, e.g. the simplest  $H + H_2 = H_2 + H$  often are used. For metallic crystals and surfaces there is the homogeneous electron gas model ("jellium").

Note that such "toy problems" have at least two properties: they are easy to understand conceptually and they are quick and easy to treat computationally. So far as we know there are no equivalent, no clear-cut model problems in chemo-mechanical processing that could be used as benchmarks for development of integrated computational models. There is a clear and pressing need for detailed, thoughtful cooperation between engineers and computational materials scientists to select and specify such models. They must be just barely realistic enough to provide useful insights to the working technologists and engineers, yet simplified enough to be useful as test cases for the simulationists. One of the challenges will be to include characteristic time and length scales corresponding at least roughly to a real chemo-mechanical process.

Settling on some well-defined, respected, standard problems also might provide a useful way to deal with the non-trivial barriers of professional organization and sociology. Standard funding mechanisms are specialized to academic disciplines. Institutional reward and incentive structures are built on those disciplines. Industry focuses with great urgency on the short-term solution of some particular, usually complicated, problem. A set of benchmarks recognized and respected across disciplines and in both industry and academe might help lower these barriers.

## OPPORTUNITIES

Having focused on missing pieces and barriers, we conclude with a few remarks on opportunities. Begin with the aforementioned limitation to dilute gas phase and  $T=0$  K for computational quantum chemistry and surface electronic structure. *Despite* those limitations, the two have proven very valuable in the pharmaceutical and semiconductor industries, respectively. That history is encouraging.

MD simulations likewise have a history of utility for prediction and characterization of meaningful materials problems. There is substantial sustained progress in speed and accuracy of both these sets of methods. Meanwhile computer potency is growing while costs decline. A good deal has been learned about recasting serial algorithms into forms useful on message-passing parallel machines, though a considerable amount of work needs to be done. Integration of the various codes required for different time and length scales into a coherent software package seems likely to require significant application of software analysis, management, and structural relationship methodologies: techniques rooted in Universal Modeling Language are under study in parts of the community.

So far as the materials physics and chemistry is concerned, the inclusion of reactivity as a logical next step is now underway. Progress on solvent effects makes the inclusion of solvent dynamics a logical second-next step. Confinement effects and treatment of relatively simple fluid dynamical contributions (e.g. viscosity corrections) lie beyond that stage. At least one can see some ways to address these problems that have a rational hope of success.

Beyond that come the really complex phenomena. Several of these are characterized by being slow and/or both quantum mechanical and yet large scale. We do not pretend at this point to know to include all of them. For example, post-planarization cleaning using ultra-sound introduces a highly complex interplay between the sound field and the chemistry of the cleaning solvent (water, ammonia, hydrogen peroxide), including microcavitation and hence the role of dissolved gases. At the start of the



section on electronic structure treatments, we also mentioned the slurry chemistry, stress-corrosion cracking, and stress-induced dissolution and recrystallization. Not mentioned were the effects of extended defects (e.g. stacking faults, dislocations, grain boundaries) and interaction between these defects and stress/chemistry. Some of these phenomena will be treatable within a straightforward BOMD but those with macroscopic length scales may be a serious challenge. Finally there are the issues of grain growth at low and room temperatures as the metal is put into the trenches and afterwards. These relate to the discussion above about the physical duration that the simulation must cover.

#### ACKNOWLEDGMENTS

SBT was supported in part by the U.S National Science Foundation (KDI-9980015). Thanks to George Frantziskonis, J.H. Simmons, F.E. Harris, H-P. Cheng, and R.J. Bartlett for valuable discussions and suggestions.

#### REFERENCES

1. F.F. Abraham, J.Q. Broughton, N. Bernstein, and E. Kaxiras, *Computers in Phys.* **12**, 538 (1998); also see J.Q. Broughton, F.F. Abraham, N. Bernstein, and E. Kaxiras, *Phys. Rev. B* **60**, 2391 (1999).
2. J. Bernholc, *Physics Today*, **52**, 30 (1999).
3. D.C. Rapaport, *The Art of Molecular Dynamics Simulation*, Cambridge University press, Cambridge (1995).
4. H-P. Cheng, R.B. Barnett, and U. Landman, *Chem. Phys. Lett.* **237**, 161 (1995)
5. R.J. Bartlett and J.F. Stanton, in *Reviews in Computational Chemistry*, K.B. Lipkowitz and D.B. Boyd, Editors, vol. 5, p 65, VCH Publishers, NY (1994); R.J. Bartlett in *Modern Electronic Structure Theory*, D.R. Yarkony Editor, p 1047, World Scientific, Singapore (1995).
6. A. Görling, S.B. Trickey, P. Gisdakis, and N. Rösch, in *Topics in Organometallic Chemistry*, P. Hofmann and J.M. Brown, Editors vol. 4, p. 109, Springer Verlag, Heidelberg (1999).
7. R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
8. T. O'Brien, P. Rack, P. Holloway, and M.C. Zerner, *J. Luminescence*, **78**, 245 (1998); A.V. Voityuk, M.C. Zerner, and N. Rösch, *J. Chem. Phys.* **103**, 4553 (1999).
9. J.O. Hirschfelder, C. F. Curtiss and R. B. Bird, *Molecular Theory of Gases and Liquids*, New York, Wiley, (1954).
10. F.G. Fumi and M.P. Tosi, *J. Phys. Chem. Solids* **25**, 31 (1964).
11. F.H. Stillinger and T.A. Weber, *Phys. Rev. B* **31**, 5262 (1985).
12. W.A. Harrison, *Pseudopotentials in the Theory of Metals*, Benjamin, NY (1966).
13. M.S. Daw and M.I. Baskes, *Phys. Rev. B* **29**, 6443 (1984).
14. M. Mullins and M.A. Dokainish, *Phil. Mag. A* **46**, 771 (1982).
15. H. Kitagawa, A. Nakatami, and Y. Sibutani, *Mat. Sci. Engr. A* **176**, 263 (1994).
16. E.B. Tadmor, R. Phillips and M. Ortiz, *Langmuir*, **12**, 4529 (1996).
17. A. Broo, G. Pearl, and M.C. Zerner, *J. Phys. Chem. A* **101**, 2478 (1997).
18. J. Andzelm, C. Kömel, and A. Klamt, *J. Chem. Phys.* **103**, 9312 (1995)
19. A.M. Rouhi, *Chemical and Engineering News* **78**, 40 (Aug. 21, 2000)
20. S. Duane, A.D. Kennedy, B.S. Pendleton and D. Roweth, *Phys. Lett. B* **195**, 216 (1987); D.W. Heerman, P. Nielaba and M. Rovere, *Comput. Physics Commun.* **60**, 311 (1990)
21. A.F. Voter, *Phys. Rev. B* **57**, R13985 (1998).
22. M.R. Sorensen and A.F. Voter, *J. Chem. Phys.* **112**, 9599 (2000).
23. P.A. Deymier, J.O. Vasseur and L. Dobrzynski, *Phys. Rev. B* **55**, 1 (1997).