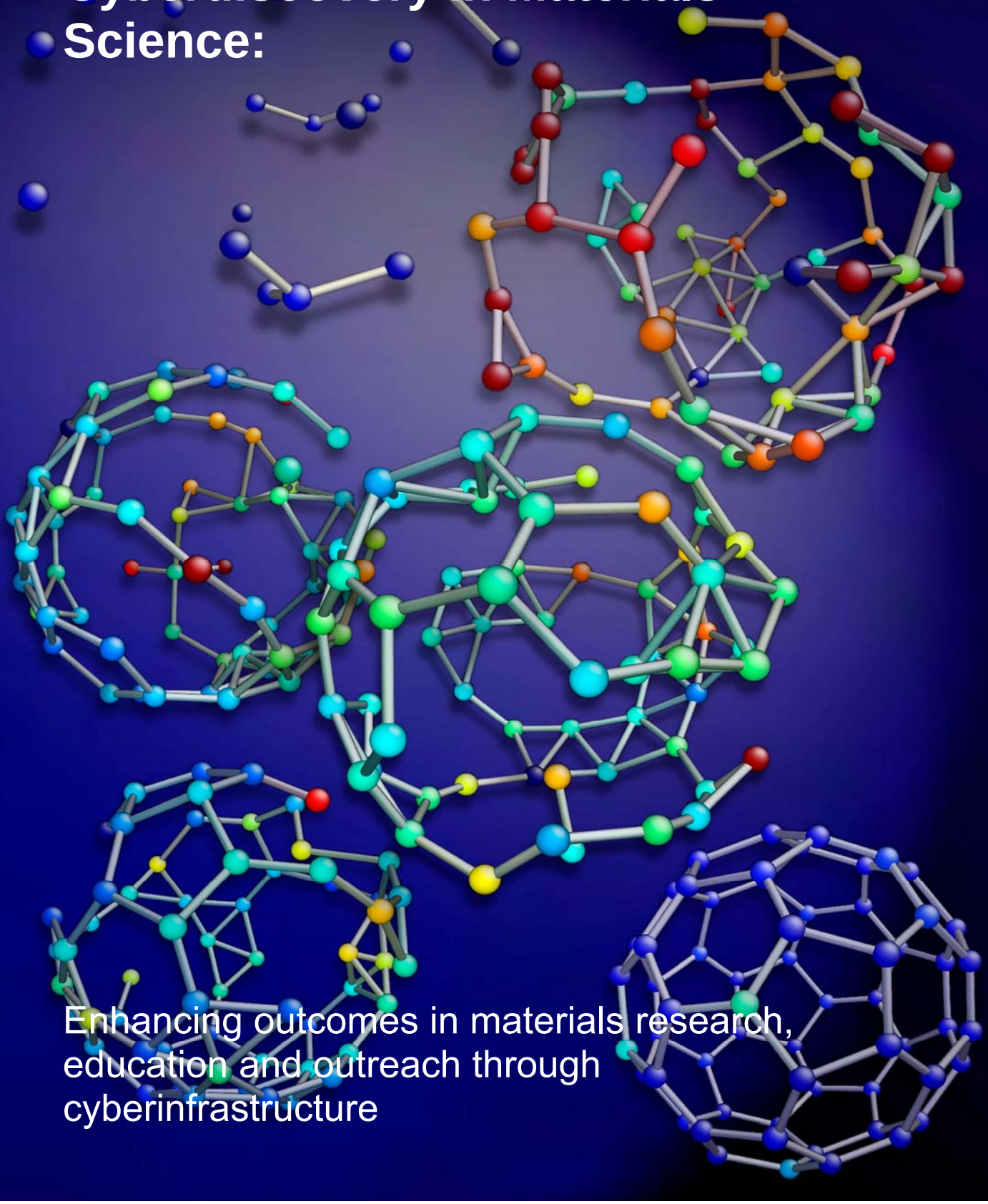
From Cyberinfrastructure to Cyberdiscovery in Materials Science:



Enhancing outcomes in materials research, education and outreach through cyberinfrastructure

From Cyberinfrastructure to Cyberdiscovery in Materials Science: Enhancing outcomes in materials research, education and outreach

Report from a workshop held in Arlington, Virginia
August 3rd- 5th, 2006

Sponsored by the National Science Foundation

Professor Simon J. L. Billinge
Department of Physics and Astronomy, Michigan State University

Professor Krishna Rajan
Department of Materials Science and Engineering, Iowa State University

Professor Susan B. Sinnott
Department of Materials Science and Engineering, University of Florida

Steering Committee

Prof. Simon Billinge
(coChair, Michigan State U)

Dr. Ernest Fontes
(Cornell/CHESS)

Prof. Mark Novotny
(Mississippi State U)

Prof. Krishna Rajan
(coChair, Iowa State U)

Prof. Bruce Robinson
(U. Washington)

Prof. Fred Sachs
(SUNY Buffalo),

Prof. Susan Sinnott
(coChair, U. Florida)

Prof. Henning Winter
(U Mass, Amherst)

Table of Contents

1. Preamble

This report is the result of presentations and discussions that took place between the attendees of the workshop on Cyberinfrastructure in Materials Research that was sponsored by the Division of Materials Research of the National Science Foundation and held at the Arlington Hilton and at NSF over the 2½ days during August 5-8th, 2005. This meeting brought together leading researchers in Materials Science, including experts in the application of cyberinfrastructure to materials research. It also included some experts from other scientific domains who have pioneered the use of cyberinfrastructure in their research. These experts represented areas where we project that cyberinfrastructure issues will impact materials research in the upcoming years, such as the creation and exploitation of databases and the curation and analysis of very large datasets. We would like to thank everyone who graciously donated their time to contribute to the lively discussions at the workshop, as well as those who helped with the myriad details involved in the organization of such a workshop.

Many materials researchers were not able to be included in the workshop. Whilst we hope the workshop is representative of the broader needs of the materials community, we welcome and seek feedback from the broader community through a number of processes. The results of the workshop and this report will be presented at a number of "town hall" meetings at major materials conferences during the fall and winter of 2006-2007: MRS meeting in December, the TMS meeting in February and the APS March meeting. Finally, the content of this report is posted online at an interactive "wiki" where discussions and contributions from the wider community can be deposited. We look forward to follow-up discussions on this topic that will be as dynamic, exciting and inspiring as were the discussions at the original workshop.

Simon J. L. Billinge

Krishna Rajan

Susan B. Sinnott

2. Executive summary and recommendations

Materials science research has played a leading role in the development of current cyberinfrastructure; indeed it is the enabling technology underpinning all of cyberinfrastructure developments. There is absolutely no doubt that future evolutionary (improvements in current technologies) and revolutionary (development of novel technologies) developments in CI will be closely coupled to breakthroughs in basic materials research. DMR has a unique role and responsibility to continue to fund research in this area. We note the historical importance of "blue-sky" and fundamental research in this process. We also note the great importance of funding developments in tools and methodologies that underpin this research, especially in the domain of the nanosciences. This report describes a number of research frontiers that may result in CI revolutions in architecture and CI technology. It is not clear which of these will succeed and there is a place for funding these as broadly as possible. At the very least, the intellectual challenges will result in important and unexpected by-products. There is a place for funding early stage applied research, but the CI economy is sufficiently large that downstream research and development can be supported in large part by the private sector.

The impact of cyberinfrastructure on materials science research is expected to be considerable but is, as yet, under-exploited. There is therefore enormous potential for the application CI tools in creative ways. The greatest impacts are likely to be unexpected ones. However, a number of aspects of materials research will clearly benefit from judicious investments in CI.

- Materials theory remains at the cutting edge of CI exploitation and will continue to yield critical insights with continued investment.
- National user facilities are now of unprecedented power and represent a significant financial investment that is not matched by investments in software and other CIs to exploit them. CI investments in this domain are highly leveraged if they successfully broaden and extend the quantity and scope of scientific knowledge coming from these precious sources. CI will also play a critical role in broadening research access to facilities in a sustainable way through remote access. It also has the potential to inspire the next generation of materials scientists by bringing frontier research at the facilities into the graduate, undergraduate and K-12 classroom in an interactive way.
- CI has a special role in building community. It is revolutionizing the way we interact with each other in everyday life. It also has great potential for bringing together geographically, scientifically and demographically diverse groups of people and providing an interface and remote communication tools that lead to rich, but sustainable and affordable, interactions. Historically, materials research is a science at the interface between traditional disciplines and it has had a leadership role in pioneering approaches to build meaningful functional interdisciplinary groups of researchers. Interdisciplinarity will grow in importance as we tackle increasingly complex scientific problems. CI is an important enabling technology in this process.
- Software developments will lead to new science, better science and more science. Physical scientists played a leading role in early computer science developments, for example, the FORTRAN computer language: physical science was an important computer science driver. This is much less true

now, and there is a disconnect between state of the art software capabilities and practices and their application in the physical sciences. This presents a special opportunity to the community: there is great potential for leveraging software engineering practices into tools allowing scientists to extend their research. The community has been slow to realize and embrace this vision but this is set to change. Investment here should have significant payoffs, though lessons learned from spectacular failed software development projects in the private sector need to be closely heeded. In general, software developed with DMR funding should be broadly accessible and long lived. Modern software engineering practices enable this.

- Algorithm development will always lead to scientific breakthroughs and should be funded. A distinction should be made between theoretical and computational algorithms. Both are important. The former result in scientific breakthroughs, for example, the development of density functional theory. The latter are implementational and generally change the size scaling of a particular problem; a classic example is the fast Fourier transform (fft). Computational algorithms are often developed in the applied math community, but they can be very problem specific and collaborations between applied mathematicians and materials scientists are to be encouraged.
- Materials databases, such as phase diagrams and structural databases have always played an important role in materials science. Full exploitation of these resources using electronic data-mining methods is at a fledgling stage and scientific advances and discoveries can be expected from application of electronic data exploitation methods in materials science. These approaches are relatively well developed in the biological sciences, for example in genomics and bioinformatics and this knowledge can be leveraged. A serious issue is the creation and curation of high quality materials databases. Funding mechanisms for these two activities need to be explored.

The meeting identified a number of critical issues that need to be addressed for the greatest scientific return on CI investments to be realized:

- Rewards
- Standards
- Sustainability of data, databases, software
- Effective sharing
- Education, training, career paths
- Access to computational resources on different scales

3. Cyberinfrastructure-revolution through materials evolution

3.1 Introduction

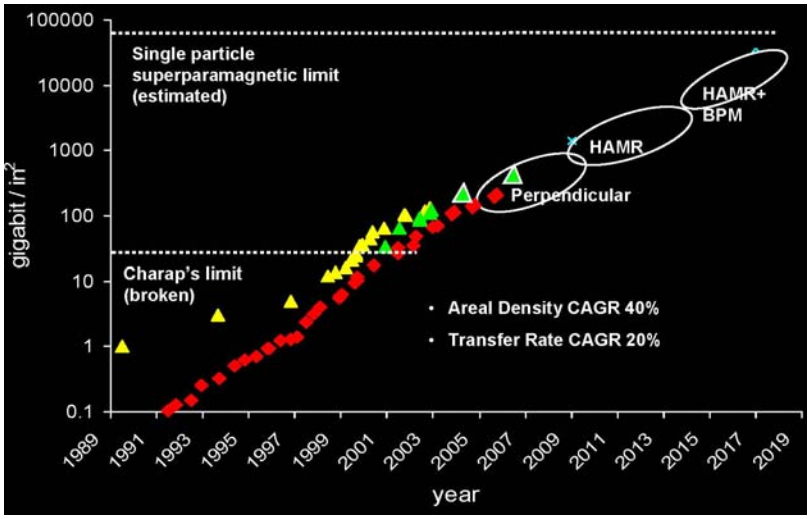
The invention of the solid-state transistor led not only to a Nobel prize for its inventors, but to a revolution in computing. Throughout the history of computing in this solid-state era, materials science has been the foundation on which advances have been built. The original transistor was 2 inches high. The cpu chips in a modern laptop computer have 10^8 active devices in an area 10 mm x 10 mm. Such a chip made out of the 1948 devices would fill a modern football stadium. The importance of materials science doesn't stop with the processor chip but is ubiquitous in all areas of cyberinfrastructure development from high areal density magnetic storage, to high frequency applications enabling cell phones. Basic research in materials continues to play a fundamental role in surmounting the current barriers to growth in processing power, memory and bandwidth that have fueled this computer revolution over the past 50 years.

The Division of Materials Research is in a unique position within the NSF with respect to the cyber revolution. Only due to research on materials and devices, research which falls under the DMR mandate, has CI progressed to its current value as a tool for science, engineering, and business. Examples abound to justify this position: from early work on transistors, to integrated circuits, to materials for magnetic recording, to materials for flat-panel displays, to allowing stability of smaller-scale features by alleviating atomistic migration problems, to the current revolution in recording migration that uses blue lasers. These are only instances that directly affect recording and computing portions of cyberinfrastructure. DMR has a responsibility in the future to fund research that will affect not only cyberinfrastructure but also many other areas of importance to modern life; from electrochemical research, which impacts batteries and fuel cells, to plastics research, which make up laptop frames to the shrink-wrap of the delivered products of the web economy, to powdered metal technology, which is used in computer frames. Furthermore, DMR will continue to have this novel connection to future revolutionary CI technologies; for example, photonic materials, Bose-Einstein Condensate materials, superconducting materials and so on. This future work in materials includes extending the evolutionary Moore's-law behavior for current CI technologies, as well as enabling the revolutionary CI developments such as spintronics, quantum computing and encryption, and technologies based on coupling biomaterials and materials devices.

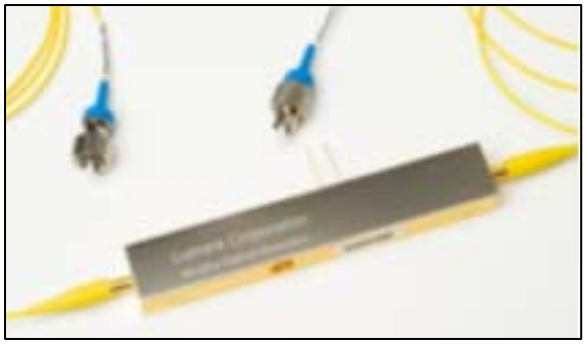
This section will provide some examples where DMR-type research is impacting CI, and will also provide "blue-sky" views of how DMR-type research may impact future CI revolutions.

3.2 Extrapolating Moore's law by materials research

In 1965 Gordon Moore (co-founder of Intel) predicted that the density of transistors on semiconductor chips would double approximately every 18 months leading to exponential growth. Arguably the major reason for the cyber-revolution and the emerging cyber-economy is because such doubling has continued. Such doubling has also occurred in other areas of CI technology such as processor speed, the bandwidth and speed of communications inside and between computers, and the density of storage in computer memories. The figure below gives an example of the recent past and predicted future of the density of magnetic recording. Although a smooth Moore's law-type of exponential growth is apparent, and is obvious to computer purchasers, it has only been through both evolutionary and revolutionary progress in materials and materials/device engineering that such growth has and may continue. For example, before 2005 almost no magnetic recording devices used perpendicular recording methods, whereas now almost all drives do. Another key materials breakthrough was the discovery of the giant magnetoresistance (GMR) effect in the 1980. This discovery was a surprise to the scientific community. However, subsequent experimental and theoretical research has led to GMR (and GMR/spin-valve) read heads being in virtually every disk-drive sold today. This is an example where a basic, unexpected scientific discovery has transformed a multibillion dollar industry within a few decades. As illustrated in the figure, further materials advances, including self-organized or patterned magnetic nanomaterials, and combined heat/optical read/write functions, will be needed to continue this Moore's law-type of growth to the ultimate limit imposed by nature of one bit recorded per spin degree of freedom (estimated in the figure). Such advances in materials also translate into other cyberinfrastructure areas, such as magnetic random-access-memory (MRAM).



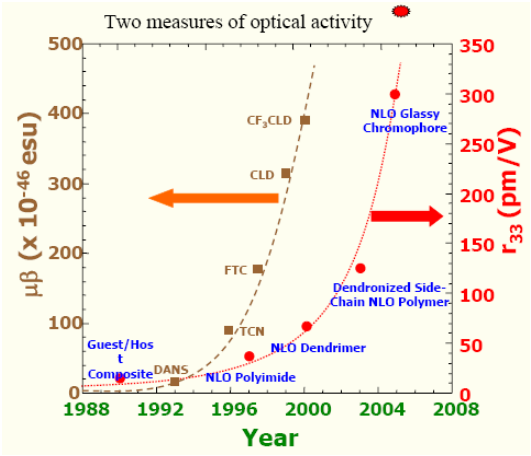
All optical switching and electro-optic modulation devices



The increase in bandwidth, and back-plane interconnections in clustered computers, has been enabled in part by advances in materials such as optical fibers. Modeling of data transmission protocols has been used to understand

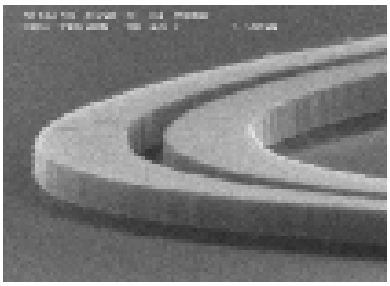
and design data transfer methodologies and protocols. Advances in electro-optic (EO) materials have been incorporated into signal transduction to put cyber information on and off the web and allowed fast, efficient connections (ports) between computers and the web. Shown are some of the newest devices that will soon replace existing electro-optic modulators. These have increasing bandwidths and data rates, and may well lead to terahertz data-transfer rates. This is an example of one of the newest generation of commercially available EO modulators that enables control of light, and conversion of light to voltages, at rates exceeding 100 GHz. It requires less than two volts to operate.

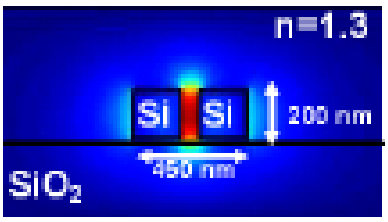
Future electrooptic computing applications will require qualitatively *new* materials. Candidate materials are based on organic non-linear optical chromophores. The figure to the right shows the dramatic improvement in performance of these materials as research progresses. Two important performance indicators are shown. Both indicators show that an exponential Moore's law type improvement in performance holds. The development of these materials has been funded by DMR.



Power delivery for the next generation of cyberdevices

Cyberinfrastructure is nothing but a paperweight if it does not have an adequate source of energy. The energy costs associated with supercomputers are staggering. Power use scales with the fourth power of processor clock-speed whereas processor speed scales only linearly. Power supply and power dissipation are limiting factors in high-performance computing. Further research into materials and device properties are desperately needed. Small gains in energy density in batteries, energy efficiency of computers and light bulbs, or transmission efficiency of energy have the possibility to both make significant cost and environmental savings while simultaneously enabling the next generations of cyberinfrastructure and other electronic devices. Some possibilities where research has the possibility of near-term payoffs include spintronic devices, which alleviate ohmic heating of devices, higher energy density batteries, better catalysts for fuel cells, better electrodes for rechargeable batteries, thermoelectric materials for active cooling, and photonic devices. Some possibilities for long-term payoffs include applications of superconductors as interconnects, photonic devices, quasi-one-dimensional conductors, flexible and efficient light-emitting arrays, multiferroic materials, materials based on Bose-Einstein Condensates, materials at the nano, bio interfaces, and so on.





New electro-optic (EO) materials have been developed to integrate existing silicon based technologies with optical methods, a critical step in their widespread adoption. In particular research teams [1] have shown that new electro-active materials can be directly deposited on silicon devices and allow the silicon based

signal switching to effect optical responses (and vice versa). The interfacing is done by cutting a narrow groove between two blocks of silicon, which is filled with the EO material. This allows the computational (silicon-based) component to directly modulate the light that is confined in the EO material. The figure above shows the physical device and the simulation of the confinement of the beam to the region filled with the EO material.

Devices of this sort have already been demonstrated as a basis for optical switching and solving the problem of connecting silicon-based computational results with optical-based output and data transmission. These devices currently have bandwidths that exceed 200 GHz. As a result of these breakthroughs, the possibilities now exist for terahertz optical signal process and all optical computing.

[1] See for example work of the Dalton, Lifson, Steier, and Scherer groups

3.3 Revolutions in computing through non traditional architectures and algorithms enabled by materials

Moore's law is rapidly approaching a number of physical limits, such as the ability to dissipate heat, and the atomic scale of miniaturization, using current technologies. Yet, our insatiable desire for greater computing power remains unabated. To continue to enjoy further rapid improvements in computing power, revolutionary changes in basic computer technology and architecture will be needed. These include quantum computing, optical computing, spintronics and other novel developments. As was the case with more conventional silicon based electronics, these revolutionary developments are carried forward on the back of innovative materials research.

DNA-based computation

DNA-based computation is a discipline whose initial emphasis was to use the vast parallelism of molecules to solve NP-complete programs through methods closely related to combinatorial chemistry. The first successful experimental demonstration was Len Adleman's solution of a small Hamiltonian path problem in 1994.

In the succeeding years, many of the laboratories involved in this effort have recognized that it is unlikely that the technical problems associated with the field will eventually be overcome to the extent that DNA-based computation will be competitive with silicon in the near future. However, many investigators have re-directed their efforts towards bringing logic to the field of DNA nanotechnology. DNA nanotechnology is a field that uses the predictability of base pairing to build objects, devices and periodic lattices. There are many examples of each of these. Many workers in DNA-based computation have applied (and are continuing to apply) the computational logic of Wang tile assembly to DNA nanotechnology, so as to produce aperiodic structures, such as cumulative XOR calculations or Sierpinski triangles.

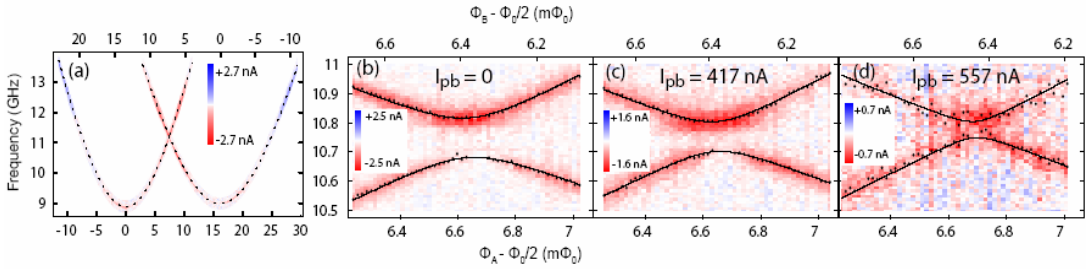
Quantum computing and encryption

Quantum computing has the potential to provide exponentially faster computation in certain applications – for example, the factorization of very large integers – than today's classical computer [1]. Whereas the classical computer is based on devices with

just two states, 0 and 1, a future quantum computer would involve a superposition of two quantum eigenstates of the form $\alpha|0> + \beta|1>$, where the coefficients α and β can assume an infinite number of values. Thousands of two-level systems (quantum bits or "qubits") would be coupled together, and would have to allow for the separate preparation of their initial states, the subsequent entanglement of these states, and the readout of the final states, all with a very high degree of quantum coherence. Qubits have been demonstrated with cavity quantum electrodynamics, ion traps, and large numbers of identical nuclear spins, all of which have high coherence, but which present extreme difficulties in scaling up to large numbers. Various kinds of solid state qubits have also been demonstrated, for example, with electron spins, nuclear spins and superconducting circuits. Such qubits are highly scalable, but have short decoherence times because of their interaction with the environment. Progress in this area over the last 6 years has been truly remarkable, largely due to progress in the materials science of the devices.

Recent advances enable a quantum dot in the semiconductor GaAs to confine a single electron. In fact, it is now possible to fabricate two nearby quantum dots each of which contains a single electron, thus forming an artificial H_2 molecule [2]. Because each electron "sees" roughly 10^6 nuclei, its decoherence time is short, about 10 ns. With the aid of spin-echo pulses, however, it is possible to extend the decoherence time to about 1 μs. Another approach to the control of single-electron spins involves nitrogen-vacancy centers in high-purity diamond [3]. The electron spin couples coherently to a single ^{13}C nuclear spin, so that in an appropriate magnetic field one can control the coupled electron-nuclear spin system. In particular, it is possible to manipulate individual isolated nuclei coherently via a nearby electron spin, thereby taking advantage of the long coherence time of isolated nuclear spins. Both experiments – which were supported in part by the NSF – illustrate that control of individual spins in semiconductor or insulating host materials is now a reality. Both experiments also demonstrate the vital role played by the advances in materials science that make them possible.

The most advanced experiments with solid state qubits, however, have been performed with superconducting qubits, which consist of circuits fabricated with photolithography or electron-beam lithography. Broadly speaking, there are three kinds [4] of superconducting qubit, involving electron charge, the phase difference across a Josephson junction and magnetic flux. In each case, microwave manipulation of the quantum states has resulted in the Rabi oscillations, Ramsey fringes and echoes long familiar in atomic physics and nuclear magnetic resonance. The charge qubit consists of a tiny superconducting island coupled to a superconducting reservoir via a Josephson junction and to a voltage-controlled gate via a capacitance. The two quantum states are $|n>$ and $|n+1>$, where n, the number of Cooper pairs on the island, is controlled by the gate voltage. Coherence times as long as 0.5 μs have been achieved [5]. The coherent superposition of a charge qubit and a single photon has been demonstrated [6]. The phase qubit is based on a single Josephson junction, and the two-level system consists of the ground and first excited states in the zero-voltage regime. Tomography of the two coupled-phase qubits was recently demonstrated, with strikingly good agreement with theory [7]. The mostly widely used flux qubit consists of a superconducting loop interrupted by three Josephson junctions [8]. The two quantum states are anticlockwise and clockwise supercurrents, corresponding to "spin-up" and "spin-down" magnetic flux. The loop can be 50 μm or more in size and contain as many as 10^{12} atoms. Coherence times of 2-3 μs have been achieved. In a very recent experiment [9], it was shown that



Frequency versus flux Φ_A (lower abscissa) and Φ_B (upper abscissa) for coupled flux qubits A and B. (a) Spectra of qubits A and B. Each spectrum arises from transitions from the ground state to the first excited state. Color bar indicates peak heights. (b) - (d) Spectrum of (a) expanded to reveal the anticrossing of the energy levels of the coupled qubits. The lower spectrum corresponds to transitions from the ground state |0> to the first excited state |1>, and the upper from the ground state |0> to the second excited state |2>. A SQUID prebias current I_{pb} was applied during the microwave excitation. Splitting is (b) 135 ± 2 MHz, (c) 103 ± 2 MHz and (d) 56 ± 5 MHz.

the coupling energy between two flux qubits could be controlled – and even made to reverse its sign – by means of a current applied to the Superconducting QUantum Interference Device (SQUID) used to read out the quantum states of the coupled qubits (see the figure above). This technique makes it straightforward to couple and decouple qubits to perform quantum computing algorithms.

These various kinds of superconducting qubit could be scaled to large numbers using currently available lithographic patterning techniques. The relatively short coherence times (~1 μs), however, present a serious obstacle to such scale-up. These short times result from either extrinsic sources of decoherence, for example, imperfect filtering against radio and television stations, which can be eliminated, or intrinsic sources of decoherence, which are generally believed to be due to defect states in the qubit or its substrate that behave as two-level systems. For example, the hopping of an electron in and out of a trap can produce fluctuations in the critical current of a Josephson junction if it is within the tunnel junction and in the charge on a superconducting island if it is in the substrate near the island. An ensemble of such independent processes leads to a noise power spectrum that scales approximately as l/f at low frequencies f. The existence of intrinsic l/f magnetic flux noise has been known for two decades, but is still not understood. In addition, these two-level systems can exchange energy with the superconducting qubit, giving rise to substantial levels of decoherence.

Significant progress has been made in reducing the density of defect states in deposited dielectrics, resulting in much longer coherence times [10]. It is evident, however, that further progress in superconducting qubits – and indeed all solid state qubits – depends critically on understanding the nature of defect states in materials and in finding methods to reduce their density by several orders of magnitude. This is *the* outstanding issue in improving the performance of solid state qubits, and one that deserves the focused attention of the materials community.

[1] D.P. DiVencenzo, *Science* **270**, 255 (1995).
[2] J.R. Petta *et al.*, *Science* **309**, 2180 (2005).
[3] L. Childress *et al.*, *Science* **314**, 281 (2006).
[4] Y. Makhlin, G. Schön and A. Shnirman, *Rev. Mod. Phys.* **73**, 357 (2001).
[5] D. Vion *et al.*, *Science* **296**, 886 (2002).

[6] A Wallraff *et al*., *Nature* **431,** 162 (2004).
[7] M. Steffen *et al.*, *Science* **313**, 1423 (2006).
[8] J.E. Mooij *et al.*, *Science* **285**, 1036 (1999).
[9] T. Hime *et al.*, *Science* **314**, xxxx (2006).
[10] J.M. Martinis *et al.*, *Phys. Rev. Lett.* **95**, 210503 (2005).

Bose-Einstein condensation of magnons in nanostructures

In a recent article on quantum computing entitled "*One qubit at a time*", the London magazine, The Economist: (5-6-2006, Vol. 379 Issue 8476, p79-80), suggested that to build a quantum computer, one should use "Condensed thinking":

> "Clinging to tried and trusted methods, though, may not be the right approach. ... Developing existing technology for use in quantum computers might prove equally mistaken. In this context, a relatively newly discovered form of matter called a Bose-Einstein condensate may point the way ahead."

The occupation of a single quantum state by a large fraction of bosons at low temperatures was predicted by Bose and Einstein in the 1920s. The quest for Bose-Einstein condensation (BEC) in a dilute atomic gas was achieved in 1995 using laser-cooling to reach ultra-cold temperatures of 10-7 K. BEC of dilute atomic gases, now regularly created in a number of laboratories around the world, have led to a wide range of unanticipated applications. Especially exciting is the effort to use BEC for the manipulation of quantum information, entanglement, and topological order. A promising extension of the atomic gas BECs is to magnons—spin-wave quanta that behave as bosonic quasiparticles—in magnetic nanoparticles. This system has unique characteristics creating the potential for a whole new variety of interesting behaviors and applications that include high-temperature Bose condensation (at tens or possibly even hundreds of Kelvin) and novel nanomagnetic devices.

The discovery of Bose-Einstein condensation of magnons in nanostructures [1,2] through measurements of magnetization and magnetic aftereffect set the stage for designing a quantum computer using these magnons. The most important point is that a magnon propagates spatially all over the magnet. By the propagation, quantum coherence is established between spatially separated points. Therefore by exciting a macroscopic number of magnons, one can easily construct states with huge entanglement. (See [3])

Not only would the development of a quantum computer greatly benefit CI, but CI will be essential in moving from discovery to implementation. For example, to fully understand this new phenomenon, it will be necessary to carry out quantum Monte Carlo and electronic band structure calculations on nanoparticles containing the millions of atoms needed for the BEC. This will require more powerful computers than are now available.

[1] E. Della Torre, L.H. Bennett, and R.E. Watson, "Extension of the Bloch $T^{3/2}$ law to magnetic nanostructures: Bose-Einstein condensation" *Phys. Rev. Lett.* **94**, 147210 (2005)
[2] S. Rao, E. Della Torre, L.H. Bennett, H.M. Seyoum, and R.E. Watson, "Temperature variation of the fluctuation field in Co/Pt", *J. Appl. Phys.* **97**, 10N113 (2005)
[3] T. Morimae, A. Sugita, and A. Shimizu, "Macroscopic entanglement of many-magnon states", *Phys. Rev. A* **71**, 032317 (2005)).

Spintronics

Spintronics is an emergent technology that exploits the spin of the electron (see Figure 2 below)(a). Electron spin has been known about since the early days of quantum mechanics; however, before 1980 no devices even tried to make use of the electron spin. That is now changing, leading to what some hope will be a "spintronic revolution". The commercially successful poster-child of spintronics to date is the spin valve, used in GMR magnetic recording read heads. However, future spintronic-based transistors (see Figure 2(b)), and ultimately complete integrated circuits, could lead to much higher clock speeds, smaller devices, and much less energy dissipation than current charge-based devices. Research is currently at an early stage, but breakthroughs in materials research are needed for economical room-temperature spintronic integrated circuits that are stable over many years to become a reality.

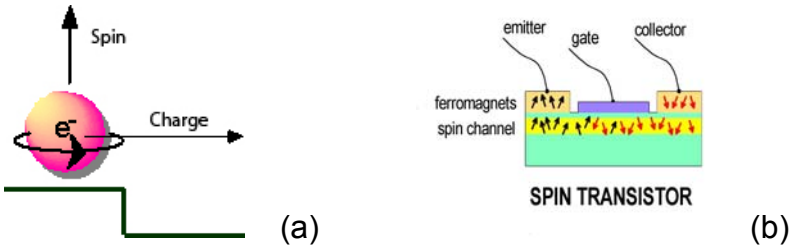


Figure 2: (a) An illustration that an electron has both spin and charge. Currently all devices utilize current flow, but there is no theoretical reason why flow of spin can not make comparable devices. (b) A schematic of a spin-transistor.

[1] "Persistent sourcing of coherent spins for multifunctional semiconductor spintronics", I. Malajovich et al, *Nature*, **411**, p. 770 (2001).

[2] "Spin Polarization Dependence of Carrier Effective Mass in Semiconductor Structures: Spintronic Effective Mass", Y. Zhang and S. Das Sarma, Physical Review Letters, **95**, 256603 (2005).

Molecular electronics

Molecular electronics is one approach to miniaturization of electronic devices by making *atomic-scale* active components: a feature problem in molecular non-equilibrium science. The simplest problem is the molecular transport junction – that is, the current/voltage characteristics of a single molecule assembled between two electrodes, either in the presence of the absence or a third gate electrode. Good experimental work on this has now been ongoing for about a decade, and entirely different behaviors have been observed than are found in larger scale organic electronics. This presents great opportunities, but also great experimental challenges because at these scales we really have no precise experimental structural probes. Molecular electronics is an activity at the very limit of our experimental capabilities and it pushes the boundaries of what is possible in materials science. It also shows promise as a technology, though it is not clear when control of fabrication will allow economical scale-up to make integrated circuits necessary for real commercial products. Molecular electronics, if commercialized, would represent a revolution in electronics, but will take a revolution in materials science research to bring about.

Computer modeling is a crucial part of the study of molecular electronics. The most prominent cyberinfrastructure for so doing is centered at Purdue University, site of the NSF supported National Center for Nanotechnology (NCN) cyber infrastructure initiative, under the leadership of Professor Mark Lundstrom. Collaborating with several other Universities, the Purdue PUNCH site, (in particular its Nano Hub component, see

Section 4.4), has been the most useful national resource for modeling in this area. PUNCH is now morphing into Vigo, a more advanced site. PUNCH/Nano Hub had provided a complete set of tools for molecular electronics. These include: tutorials, modules for learning, real time lectures from leaders in the field, and programs (both tutorial and advanced) for actual modeling of molecular conductance spectra. Progress is being made on more advanced programs for dealing with particular effects such as vibronic coupling and Coulomb Blockade regimes, (much of it from Purdue's Supriyo Data) and on a database for reporting on transport on specific junctions.

This problem has become a very "hot" one, and many calculations are reported monthly. Linking different programs in different centers worldwide, the Purdue site has proven tremendously valuable to the entire community, as a model case of how cyber infrastructure can serve to bring together researchers, students, educators and industry. This is permitting the worldwide community to pursue a fundamental science/engineering challenge in cyberinfrastructure development, using an appropriate existing cyberinfrastructure.

Photonic crystals

Photonic crystal micro- and nano-devices are expected to shape the future in optical computing (all-optical switches, micro-transistor and memories) but also to contribute fundamental hardware for linear optical quantum computing and quantum cryptography: novel architectures with the potential for radical departures in computing.

The synergetic interplay between material science and theoretical computational analysis is a driving force in the field of photonic crystals and photonic band gap materials. Recent advances in micro-structuring technology have allowed the realization and controlled engineering of three-dimensional photonic crystal at the near IR and visible frequency spectrum of the electromagnetic spectrum. Cyber-infrastructure enabled theoretical description of photonic crystals has advanced to the point where it provides a reliable predictive tool to both material synthesis and spectroscopic analysis of these semiconductors for light. This allows the computer-design of novel photonic crystal architectures. This is an example of closing the loop: CI methods aid in the development of new CI materials.

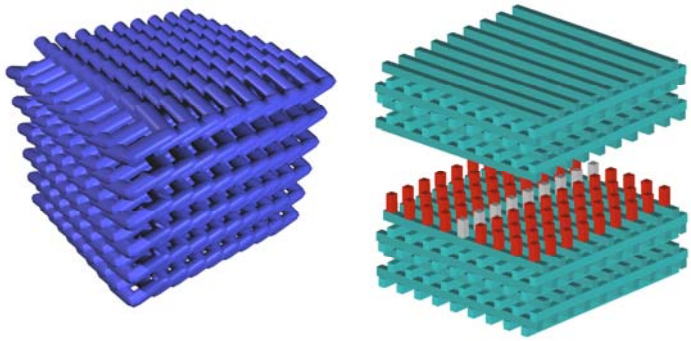


Figure 3: Right panel: Tetragonal square spiral photonic crystal structure discovered by Toader and John (Science 292, 1133, 2001). Left panel: Photonic crystal heterostructure for implementation of a deterministic single photon source (Florescu et al. , EPL 69, 945, 2005).

4. Materials revolution through cyberinfrastructure-evolution

Materials science has enabled the cyber-revolution. How can recent developments in cyberinfrastructure revolutionize materials science? As computer power continues to evolve, the size of systems that can be tackled computationally increases to the point where, for the first time, it is possible to carry out realistic atomic simulations of real physical systems that are under study in the laboratory. Thus, we can begin to study at the atomic level what happens at the tip of a propagating crack, or at the interface between two surfaces sliding over one another. We can also start to study problems at the frontier of complexity, such as strongly interacting systems, bio-materials, excited states, non-equilibrium systems, and terabyte data-sets. Cyberinfrastructure will also revolutionize the way we do materials science by enabling remote collaborations, remote access to instrumentation, education and training, and allowing unprecedented understanding to be uncovered from databases and through visualization.

4.1 Materials by design

The emergent role of atomic-scale simulations in materials research

The ambitious goal of creating “Materials by Design” requires the capability of making predictions about materials before they are synthesized, and thus relies heavily on computational modeling. In order to be predictive, theoretical models should incorporate as little as possible—ideally not any—input from experiment.

This requirement naturally leads to the use of microscopic models, and in particular atomic-scale molecular dynamics (MD) simulations, in which macroscopic behavior is entirely defined by microscopic interatomic forces. Interatomic interactions can in turn be defined in terms of adjustable parameters fitted to a limited set of experiments (classical molecular dynamics), or derived from first-principles using quantum mechanical electronic structure computations (first-principles molecular dynamics).

Both approaches have been extremely useful in the study of materials properties, and depend on the availability of high-performance computers. Classical molecular dynamics (MD) has been extensively optimized for use of large-scale parallel platforms and is currently capable of describing the behavior of billions of atoms over lengthy (on atomic time-scales) durations of the order of a microsecond.

First-principles simulations are more computationally demanding since they require the computation of the electronic structure of the material under study. This usually leads to the use of large supercomputers. Following the dramatic increase in available computing power during the past decade, first-principles simulations are now playing an increasingly important role, both as a tool for computational discovery and to help in the interpretation of experimental results. Optimized first-principles MD (FPMD) codes are now routinely used on large parallel computers and are capable of simulating systems comprising a few hundred atoms for durations of tens of picoseconds.

First-principles simulations are particularly useful in nanoscience applications, for which empirical models can often be unreliable. For example, in a recent investigation of the growth mechanisms of carbon nanotubes, first-principles molecular dynamics was used to simulate the early stages of nanotube growth on an iron catalyst [1]. This application was a prime example in which the first-principles approach provided a detailed atomic-scale description of the growth mechanism, a phenomenon that is too fast to be observed experimentally. The simulation was repeated using another catalyst (gold) on which the nanotube did not grow, in agreement with experimental observations. This kind of computational investigation required several months of CPU time on a medium-size cluster (approx. 64 CPUs). Furthermore, the results obtained suggested numerous other simulations that would each require similar computational resources. This example is one of many applications of quantum simulations to materials science, and illustrates the growing need for computational resources in the area of quantum simulations of materials.

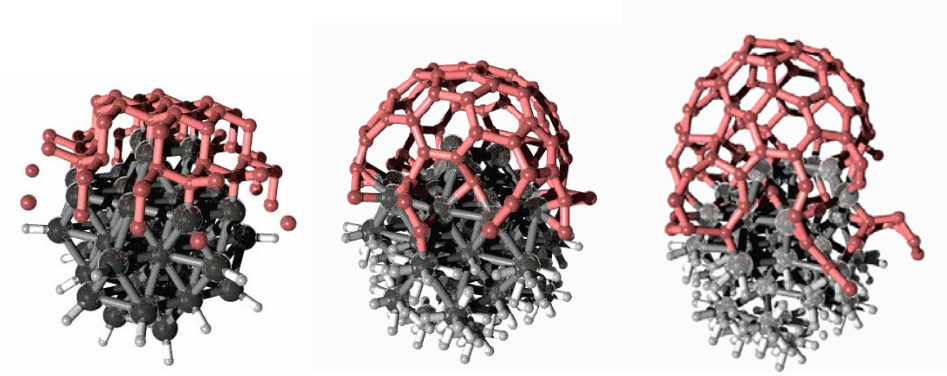


Figure 4: Early stages of the growth of a single-walled carbon nanotube on an iron catalyst nanoparticle simulated by first-principles molecular dynamics. The simulation shows the formation of C-C bonds, the appearance of pentagons and hexagons on the carbon surface, and the subsequent growth of the carbon nanotube [1].

The computational complexity of first-principles methods requires the development of advanced simulation codes. Software development for first-principles simulations has become increasingly challenging during the past few years due to the changing architecture of new supercomputers. Careful code optimization is needed in order to obtain a significant fraction of the computer peak performance. In the coming five to ten years, petascale computers will have the potential to enable quantum simulations of unprecedented size and accuracy. However, existing simulation software is not expected to run unchanged on such large computers. New computational algorithms will have to be developed in order to exploit petascale platforms efficiently. In particular, new developments in linear-scaling electronic structure methods will be needed to reduce the amount of inter-processor communication taking place in a simulation. Based on experience accumulated on terascale computers, it is likely that a 3-5 year effort in software development will be needed in order to produce efficient codes for 1-10 petaflop class computers. As a consequence, it is important that software development be initiated well before petascale platforms are installed in order to ensure efficient use of hardware resources as soon as they appear.

First-principles simulations depend on cyberinfrastructure in several ways. Beyond the need for high-performance computers, quantum simulations generate a large amount of data that in turn requires important resources in data storage and network bandwidth. Publicly accessible simulation data repositories will be needed and are expected to

greatly facilitate computational investigations and collaborations. Improved accessibility of simulation data will also facilitate the training of future computational materials scientists.

Because of the versatility of atomic-scale simulation methods, it must be expected that the demand for large-scale simulations will sharply increase in the years to come. An adequate growth in cyberinfrastructure will allow the materials science community to meet this challenge and, as a consequence, dramatically accelerate the scientific discovery process.

[1] J.-Y. Raty, F. Gygi, G. Galli, "*Growth of Carbon Nanotubes on Metal Nanoparticles: Microscopic Mechanism from Ab Initio Molecular Dynamics Simulations"* Phys. Rev. Lett. **95,** 096103 (2005).
[2] NNSA Announces Record Performance on IBM Blue Gene/L http://www.hpcwire.com/hpc/701664.html

Materials science simulations are among the leading applications for scientific supercomputing

Large-scale simulations in materials science have captured the Gordon Bell Prize for Peak Performance three times in the history of the award, including the two most recent prizes, which represent the highest sustained performance ever archivally documented on a full-fledged scientific application run on a general-purpose scientific computer. The titles and authors of the prize papers are listed below:

1998 *High-performance First-principles Method for Complex Magnetic Properties*, B. Ujfalussy, X. Wang, X. Zhuang, D. M. C. Nicholson, W. A. Shelton, G. M Stocks, A. Canning, Y. Wang, and B. L. Gyorffy (1.02 Tflop/s on 1536 processors of a T3E-1200)

An *O(N)* Local Density Approximation (LDA) density functional theory computation using locally self-consistent multiple scattering (LSMS)

2005 *100+ Tflop Solidification Simulations on BlueGene/L*, F. H. Streitz, J. N. Glosli, M. V. Patel, B. Chan, R. K Yates, B. R. de Supinski, J. Sexton, and J. A. Gunnels (101.7 Tflop/s on 131,072 processors of a BG/L)

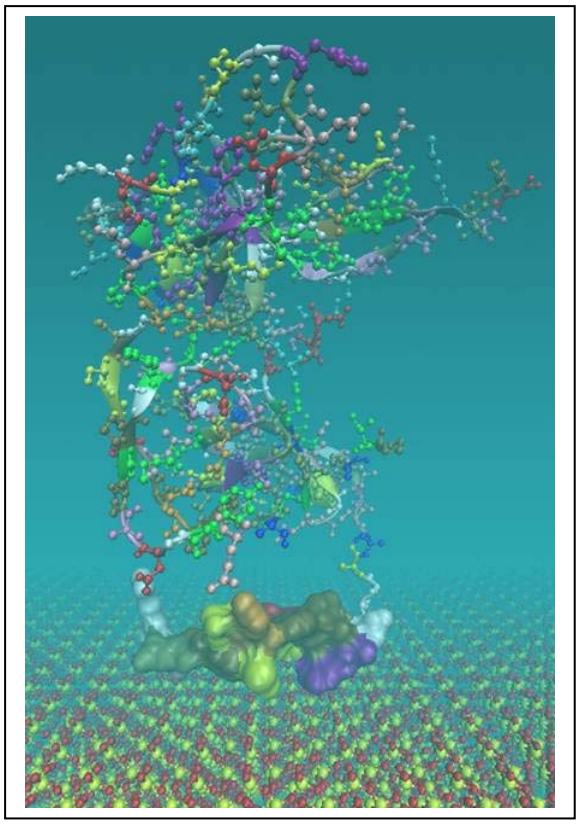
A Modified Generalized Potential Theory (MGPT) molecular dynamics simulation

2006 *Large-scale Electronic Structure Calculations on High-Z Metals on the BlueGene/L Platform*, F. Gygi, E. W. Draeger, M. Schulz, B. R. de Supinski, J. A. Gunnels, V. Austel, J. C. Sexton, F. Franchetti, S. Kral, C. W. Ueberhueber, J. Lorenz (207.3 Tflop/s on 131,072 processors of a BG/L)

A first-principles molecular dynamics simulation

Maximum power at minimal cost: Designed organic photovoltaics

At a time of increasing geopolitical instability and declining energy reserves, the development of a primary renewable energy source - solar electricity - is of paramount importance. The current challenge is to produce solar electricity at sufficiently low cost, in the range of $0.40/Wp, to enable widespread deployment in the energy infrastructure



both in the United States and globally. “Plastic” solar cell technology based on organic molecular, polymeric, and nanostructured materials, offers one of the most promising routes towards realizing this objective. The unique ease of processing of organic based materials, offers the intriguing possibility of developing inexpensive “solar paint” that can be deployed over large non-uniform areas for maximum solar utilization. [1]

The materials challenge is to increase the efficiency of organic photovoltaics from their current 2-5% range to the 20-50% range, i.e. by a factor of 10. This order of magnitude increase can only be achieved by a fundamental materials redesign to optimize structural, optical, and electronic properties for photovoltaic energy conversion over the full solar spectrum [2]. However, unlike in their inorganic counterparts, the basic physical process of solar energy conversion in organic photovoltaics remains poorly understood. Computational approaches are needed to gain microscopic insight into the mechanism of exciton generation, charge separation, diffusion, and recombination. This requires establishing a basic cyberinfrastructure of theoretical models and algorithms, ranging from scalable methods for large-scale quantum-mechanical simulation, to theories beyond density functional theory for excited states, and multi-scale and embedding techniques that can bridge the molecular to organic-inorganic interface length scales. These capabilities will serve not only guide the interpretation of experiment and provide a detailed model for the energy conversion process, but form the initial step to solving the “inverse” problem of virtual organic photovoltaic materials design.

[1] Polymer Photovoltaic Cells: Enhanced Efficiencies via a Network of Internal Donor-acceptor Heterojunctions, G. Yu, J. Gao, J.C. Hummelen, F. Wudl, and A.J. Heeger, Science **270** 1789 (1995)

[2] Next Generation Photovoltaics: High Efficiency through Full Spectrum Utilization. A. Marti and A. Luque (Eds.), Institute of Physics, Bristol, UK (2004)

Modeling nano/bio interface at the molecular scale:

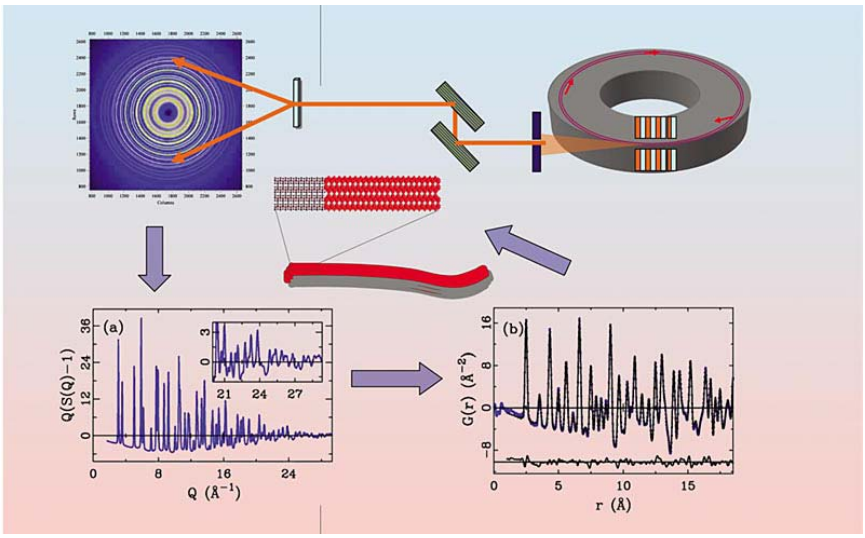
A revolution in materials is taking place at the bio/nano interface where biology and materials science truly meet [1-4]. The realization of biology-inspired materials depends on designing biomacromolecules that interact predictably with materials are addressable. For example, protein interactions at solid surfaces play key roles in hard tissue regeneration, nanomedicine, and bioenabled photonics and electronics [1-4]. Genetically engineered peptides for inorganics (GEPI) are short (7-14) amino acid sequences that are selected via combinatorial mutagenesis for their affinity to specific materials such as metals, oxides, minerals, and semiconductors [3,4]. The peptide/material interaction may be studied through appropriate modeling at the peptide-

inorganic interface. The figure on the previous page gives an example of a modeling study in which a quartz-binding peptide (QBP1) is displayed in its natural condition. The structure of the construct was predicted using the RAMP Software [7]. RAMP combines multiple algorithms for regression including simulated annealing, genetic algorithms, graph theory, and semi-exhaustive searches, with move sets derived from distributions of amino acid φ/ψ preferences, and scoring functions consisting of all-atom based pairwise preferences, hydrophobicity indices, secondary structure preferences, hydrogen bonding, and the degree of clustering exhibited by the end points of different trajectories. The interaction of the GEPI with the quartz surface (i.e., with the ordered atoms of silicon and oxygen) is not currently possible but awaits advances in our knowledge of the interaction parameters between these atoms and the atoms in the binding residues of this GEPI.

[1.] H. Colfen and S. Mann, *Angew Chem. Intl. Ed*., **42**, 2350 (2003).
[2.] S. Zhang, *Nature Biotech*, **21**, 1171 (2003).
[3.] N. Seeman and A. M. Belcher, *PNAS*, **99**, 6452 (2002).
[3.] M. Sarikaya *et al.*, *Nature-Mater.*, **2**, 577 (2003).
[4.] E. E. Oren *et al*., Unpublished Research (2006).
[5.] http://www.pdb.org/
[7.] http://software.compbio.washington.edu/ramp/

4.2 Nanostructured materials

The nanostructure problem



Emerging materials are increasingly nanostructured. The nano-meter length-scale is perfect for engineering complexity into a material to give it a directed functionality for some application. This is the origin of the excitement and the potential of nanoscience and nanotechnology. By definition, nanostructures are not infinitely periodic, long-range ordered, crystalline structures. The crystallographic methods that are the *sine qua non* of structure determination no longer work on the nanoscale and there is no robust way of solving the structure of nanostructured materials. Since knowledge of structure is a prerequisite to a full understanding of materials properties, and accurate atomic scale characterization of a material is a requirement in any coherent materials design strategy, the inability to solve nanostructures accurately presents a serious hurdle in the nanorevolution. Many techniques exist for probing nanostructured materials. Some are inherently local probes such as TEM and STM. Others are average probes that are sensitive to local structure such as the atomic pair distribution function (PDF) method or extended x-ray absorption

fine structure analysis (XAFS). The principal difficulty with the application of these methods to solving the nanostructure problem is that, in general, any one technique does not contain sufficient information to constrain a unique structural solution. What is required is a coherent strategy for combining data from these techniques self-consistently in a global local-structure optimization scheme. This is a challenging but fundamentally important problem whose solution will require close collaboration between materials scientists, physicists, chemists, applied mathematicians and computer scientists at the very least. It is clearly an area where CI has the potential to result in a materials science revolution.

Nanoengineering design

The unique properties of nanostructured materials are intimately related to the significantly increased surface to volume ratio at the nanometer scale. For nanoengineering design, a detailed atomistic understanding of the interface structures becomes essential. Interface-driven phenomena span many areas of materials research, from nanotribology to loss of catalytically active surface area during use of fuel cells to interactions of nanostructures with organic molecules. Many of these phenomena are driven by physics and chemistry at multiple length- and time-scales. For example, energy dissipation mechanisms underlying the tribological response of nanostructured materials depend on the grain size of the material, phonon spectrum, the interplay between surface chemistry and adhesion, etc. In the current state of knowledge we are not able to predict the atomic coefficient of friction, even if the bulk and surface properties of a material are known. In order to address this and many other exciting problems emerging in nano-engineering design, there is an urgent need to bring together in a seamless fashion simulation tools spanning various levels of accuracy and length-scales. Effective multiscale modeling requires CI development in a number of different areas including algorithms and code integration as well as the availability of high performance computing resources.

Increasing computer power and hence increasing size of simulated structures, will bridge the existing length-scale disparity between many simulation models and experiments. Computer simulations involving many millions of atoms have already brought an invaluable insight into the nanostructure-property relation (Fig. 5), however seeking patterns and extracting information from these massive data sets present a non-trivial and increasing challenge. For example, a nanostructured ceramic can contain thousands of randomly oriented crystalline grains surrounded by the intergranular region with various levels of topological disorder. The grains' crystallographic structure and chemical ordering depend on many variables, e.g., internal pressure. The density and type of topological defects are another important factor determining a material's properties. Tracking deformations in a stand-alone amorphous material (which lacks a medium- and long-range order) presents a challenge

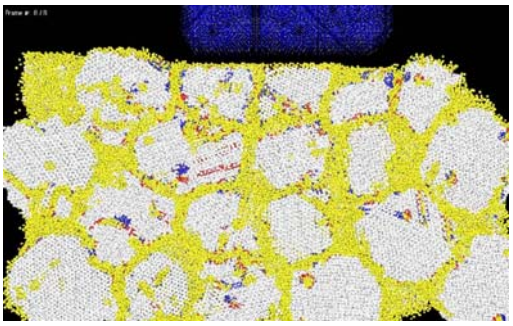


Figure 5: Atomic structure of nanocrystalline SiC. Atoms are color-coded by their topological similarity to the perfect lattice. Crystalline grains are shown in white, amorphous grain boundaries in yellow. Mechanical response of the material is driven by competition between ordering and disordering trends.

in itself, let alone as a part of a complex nanostructured material. Until now seeking patterns in such structures has required an extensive hands-on analysis. Data mining algorithms and automatic search for topological defects based on rigorously defined materials science criteria will be invaluable to the modeling community.

4.3 Materials out of equilibrium

New materials with improved corrosive and mechanical properties

The engineering of materials with improved anti-corrosive and mechanical properties is a primary challenge for materials science [1,2]. In a major federal study commissioned in 1998, materials failure resulting from corrosion was estimated to be in the range of 276 billion USD per year, or 3.1% of GDP in direct costs, with at least a further equal amount estimated for indirect costs resulting from lost time, delays, and litigation [3].

Computational approaches have the possibility to revolutionize the traditional Edisonian materials science approach to materials discovery through advances in materials simulation and high performance computing. To correctly simulate the structural corrosion induced mechanical failure requires incorporating the relevant length and time-scales of the problem from the level of the atoms via the mesoscopic scales associated with interacting dislocations and grain microstructures and diffusion, all the way up to the continuum level. A comprehensive cyberinfrastructure of models and algorithms is needed not only to capture each of the physical levels of theory including quantum mechanical electronic structure, interatomic potentials, long time-scale molecular dynamics algorithms, and finite-element modeling, but also to provide the necessary feedback and links between these levels.

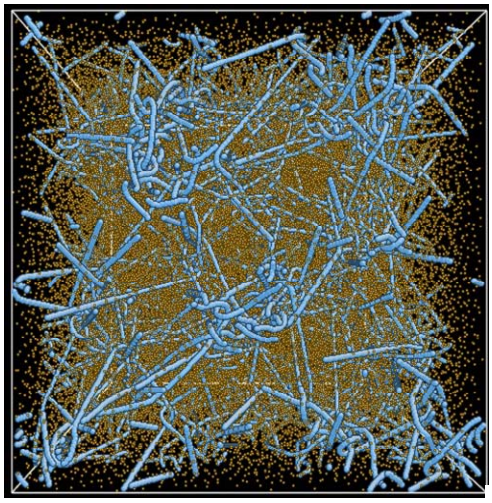


Figure. 6: Orange chains represent fine-grained "pearl necklace" chains simulated by molecular dynamics simulations, while thick blue chains are the "primitive paths" representing the shortest paths that respect the entanglements with surrounding chains. The upper-right inset shows the network of primitive paths for a box containing all chains in the simulation, while individual chains and their primitive paths are illustrated along the side and bottom. (from Zhou and Larson 2005)

[1] Why stainless steel corrodes, MP Ryan, DE Williams, RJ Chater, BM Hutton, DS McPhail, Nature **415** 770 (2002)
[2] Evolution of nanoporsity in dealloying, J Erlebacher MJ Aziz, A Karma, N Dimitrov, K Sieradzki, Nature **410** 450 (2001)
[3] Corrosion Cost and Preventive Strategies in the United States, GH Koch, MPH Brongers and NG Thompson and Y Paul Virmani and JH Payer, Technical Report, Office of Infrastructure Research and Development, Federal Highway Administration, FHWA-RD-01-156 (2001)

Polymers

One example material that has been investigated in this way is polymers. The relationship between polymer statistics, such as molecular weight or long-chain branching structure, and rheology is not well understood, despite significant computational efforts. Advanced computational methods

now allow linkages to be built between fine-grain and coarse-grain models of the entanglement structure of dense polymer melts as illustrated in Figure 6. Since the entanglement structure controls the rate of polymer relaxation, such computations allow development, testing, and the improvement of mesoscopic "tube" models for polymer melt dynamics, near or far from the equilibrium state.

Large-scale computation and networking with groups worldwide who are carrying out similar analyses is crucial to the refinement of these advanced methods. Large sets of polymer chain configurations generated through molecular dynamics simulations need to be shared with groups worldwide to determine the optimal methods of analysis and development of multi-scale methods for determining linear and nonlinear flow properties of polymer melts. Cyberinfrastructure, including readily accessible terascale computing and high-speed networking, is important to the development of this field, especially for enhancing collaborations between industry and academia worldwide. The "flow diagram" shown below illustrates the need for modeling tools developed at multiple institutions to be integrated on a common platform, to allow polymer synthesis and processing to be designed synergistically.

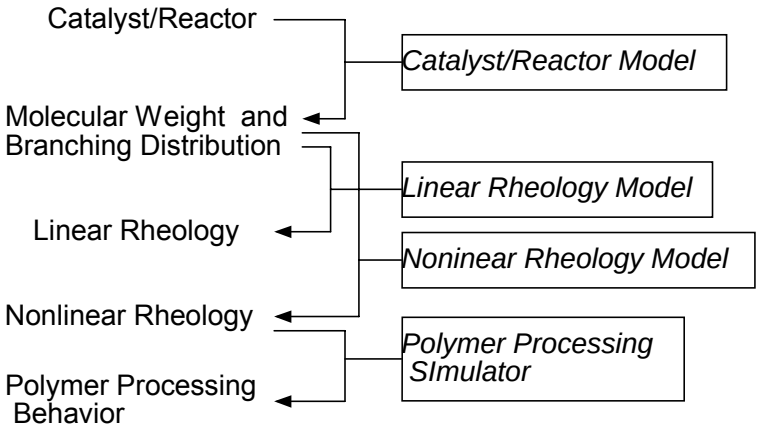
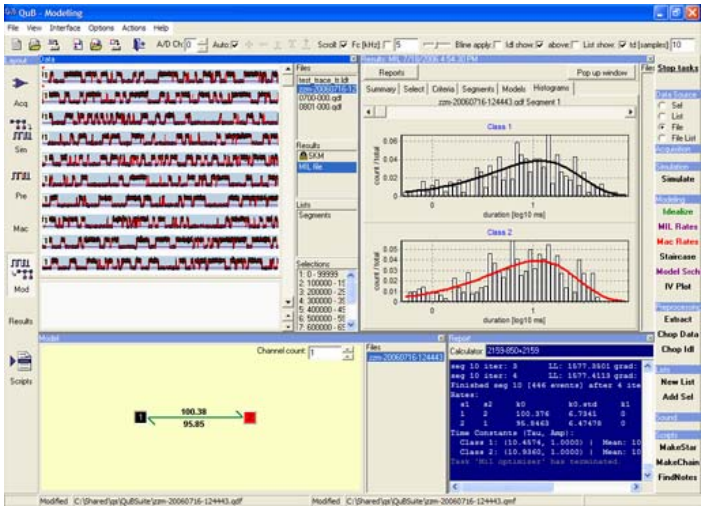


Fig. 7: "Flow diagram" showing the interrelationship among models for polymer synthesis, rheology, and processing.

[1] Q. Zhou and R.G. Larson, Macromolecules, Macromolecules 38:5761-5765 **2005**, "Primitive Path Identification in Molecular Dynamics Simulations of Entangled Polymer Melts."

[2] S. Shanbhag and R.G. Larson, Phys. Rev. Lett., Art. No. 076001, **2005**, "The Chain Retraction Potential in a Fixed Entanglement Network."

Transition states in biological enzymes



A related class of materials that have benefited from computational advancements are biological. For example, the discovery that the transition states of biological enzymes are a broad ensemble of states rather than the intersection of two parabolas (Auerbach, 2005) required the availability of algorithms (Feng et al., 1997) and a cyber infrastructure that can dissect the kinetics of single molecules. Freely available software resulted from these developments. A screenshot from the Qub application is shown below. This resulted from collaborations across disciplines and the application of novel algorithms to this problem, coupled with experimental developments. Without the support of CI involving interdisciplinary collaborations the fundamental new insights made possible by experimental advances would not have been possible. These breakthroughs did not rely on high performance computing but innovation, software developments and cross-pollination coming from working at the interface between disciplines.

[1] Auerbach, A. 2004. Gating of acetylcholine receptor channels: brownian motion across a broad transition state. *Proc. Natl. Acad. Sci. U. S. A* 102:1408-1412.

[2] Feng, Q., A. Auerbach, and F. Sachs. 1997. Maximum likelihood estimation of aggregated Markov processes. *Proc R Soc Lond B Biol Sci* 264:375-383.

4.4 Building research and learning communities

Cyberinfrastructure has a special role to play in building and integrating research communities, education communities, and the wider world. This is particularly important, and particularly challenging, in the field of materials science whose community is notable among scientific disciplines for its breadth and diversity. Scientific breakthroughs often happen at the interface between disciplines. Materials science resides precisely at the interface of physics, chemistry, biology, applied mathematics, basic science and engineering. Facilitating these interactions therefore adds significant value to the materials science enterprise, often in unanticipated ways. As well as bridging disciplines, well designed web portals lower barriers to participation by facilitating access to software applications and other resources and allowing remote participation at low cost. Web services allow access to sophisticated high-end computing, and physical resources such as transmission electron microscopes, from little more than a web-enabled laptop pc, located anywhere in the world. Such capabilities have been

demonstrated but are the exception rather than the rule. Their wider development will result in a paradigm shift in the utilization of shared facilities in the future, both increasing the scientific output per dollar of these facilities and at the same time democratizing their use. Software development is a vital component to make this work.

Global CI initiative for interactive reology

The Cyber Infrastructure for Rheology [1] has come together as a collaboration of researchers from 9 countries (Australia, Belgium, Germany, Israel, Italy, Japan, Switzerland, UK, USA). With the CI, the user is able to infuse experimental data with cutting-edge theories (including multiscale models and simulation) of these globally distributed expert groups. The project started in 1987 by standardizing diverse rheological data and by introducing reliable and robust data analysis methods. From that point on new capabilities (expert modules) were added step by step with the objective of translating the most recent advances of rheology into classroom teaching and into research. An integral feature of the CI platform is its rapid access to data and knowledge. Imagine being able to instantly pursue a question on rheology even while away from your desk. You only need your laptop and a wireless connection. Now, simply connect to your data source, plot the data, choose from several theories to predict rheological properties, relate the predictions to the experimentally observed rheology, explore variations in molecular architecture, and connect into an application. Within about 30 minutes or less, you will generate synergy between rheological experiments and theory. Today's CI tools seamlessly merge experimental data with theoretical predictions and modeling calculations in rheology. Eventually, seemingly disparate theories and experimental observations will be linked and taken to their limits, thereby leading to unexpected results and new questions. Polymer properties can not studied without accounting for engineering aspects. The next steps in CI development will include engineering topics such as reaction engineering and polymer processing. Models on structure–property relations are less advanced but should be included as much as they become available. With that extended CI tool set, a researcher or a student can span the entire range of a polymer. He/she can predict the molecular architecture of a polymer produced under specified reaction conditions, predict the rheology for the molecular architecture (polymer dynamics calculation), predict the preprocessing behavior and, finally, the resulting properties. In this multi-scale CI, new ideas can be explored rapidly and comprehensively, with the objective of discovering and providing novel materials that are important for society.

[1] Winter HH, Mours M (2006) The cyber infrastructure initiative for rheology. Rheol Acta 45:331-338

The Biology Workbench—A Possible Template for a Materials Science Workbench

The Biology WorkBench (Hhttp://workbench.sdsc.edu/H) is an example of a tool that integrates data sources, analysis tools, and visualization tools in a common portal that enables serious computational biology research, computational augmentation of experimental work, and use of research tools in education. It has expanded to a stage where almost a 100,000 individuals have accounts on the Workbench. Its original development was supported by the National Science Foundation through the National Center for Supercomputing Applications. It is currently housed at the San Diego Supercomputing Center, where its continuing enhancement is supported by the National Institutes of Health. It is used by many researchers around the world. Using biology as a model discipline, one can conceive of a "Materials Science Workbench" a portal that would provide integrated access to data sources, relevant literature, analysis tools, simulation tools, visualization tools, and interpretative materials to bridge the gaps between computational researchers, experimental researchers, materials design engineers, educators, and students. This can serve as a cyberinfrastructure for materials informatics education and research The web portal will not only showcase the work of students but will also be used as a test platform for new models of data management, semantic web technologies, ontology development, information sharing, computational grid development and algorithm development. It can serve as the focal point for integrating research and education. Some of its research and educational functions may include:

- Create and test an open architecture to compute materials properties.
- Incorporate intelligent tools to read and extract the molecular data from literature based information, which can be turned into a validated markup language such as MatML (Materials Markup Language) or CML (Chemical Markup Language).
- Establish and test a web service model that permits seamless links between experiment, computation and data mining tools.

Based on the experience of the Biology Workbench,

- It is essential to provide appropriate interdisciplinary environments with access to high end computational resources for the development of a Materials Science Workbench and the components that will go into it.
- Interoperability is essential to creating a Materials Science Workbench. Achieving that will require flexibility, good communications, and a sense of community among the participants, in addition to computational expertise.
- Cyberinfrastructure must be evolvable; i.e., must be maintained while also adapting, and occasionally completely restructuring, in response to evolving needs of the community it serves.
- Cyberinfrastructure can enable authentic problem-centered learning, which is badly needed in our educational institutions.

Cyberinfrastructure opportunities for the promotion of minorities

CI gives us a great opportunity of involving underrepresented people in Materials Science research and education. Through the effective use of communication tools and the generation of a multidisciplinary material networks where students can get access to literature, user-friendly software, simulation and visualization tools and equipment and virtual facilities from remote locations, underrepresented students and faculty will be able to establish collaboration links and substantially improve their ability to carry out cutting-edge research. For example in the use of remote equipment facilities, at UPRM and in

collaboration of NSF-IMI-COSMIC, the SEM has been effectively introduced as an important characterization tool to 25-30 students per session. The “hands-on” operation was possible due to a combination of remote SEM operation through Internet and “smart board facilities. Students were able to change magnification, voltage, position and carried out analysis of images by pushing virtual bottoms at the white board. The participation of the students was an interactive-cinema type of experience where self-discovery was carried out by all the student group (25-30 students) at one time.

Faster communication, enhance collaboration and availability of remote control in state-of–the-art instrumentation will allow us to effectively introduce these techniques in the classroom and even in remote K-7-K12 school locations.



NanoHUB

The Network for Computational Nanotechnology (NCN) has created an excellent example of a cyberinfrastructure for nanoscience research and education, embodied by the web site nanoHUB.org. In 2005, more than 12,000 users accessed nanoHUB to view a collection of seminars, tutorials, animations, publications, and simulation tools submitted by more than 250 contributors from all over the world. But the nanoHUB is more than just a repository. It is a place where researchers and educators can meet and accomplish real work. The nanoHUB offers integrated, online web meetings via Macromedia Breeze, source code collaboration through its nanoFORGE area, events calendars, and many other services designed to connect researchers and build community. Most importantly, the nanoHUB connects users to the simulation tools they need for research and education. Users can access more than 40 interactive, graphical tools, and not only launch jobs, but also visualize and analyze the results, all via an ordinary web browser. The NCN’s emphasis on usability has produced a clean interface that makes it easy to use powerful research tools. Simulation jobs can be dispatched on national Grid resources, including the NSF TeraGrid and the Open Science Grid. The nanoHUB middleware hides much of the complexity of Grid computing, handling authentication, authorization, file transfer, and visualization, and letting the researcher focus on research. This approach also helps educators bring these tools to the classroom, letting them bypass the mire of Grid computing and focus instead on physics.

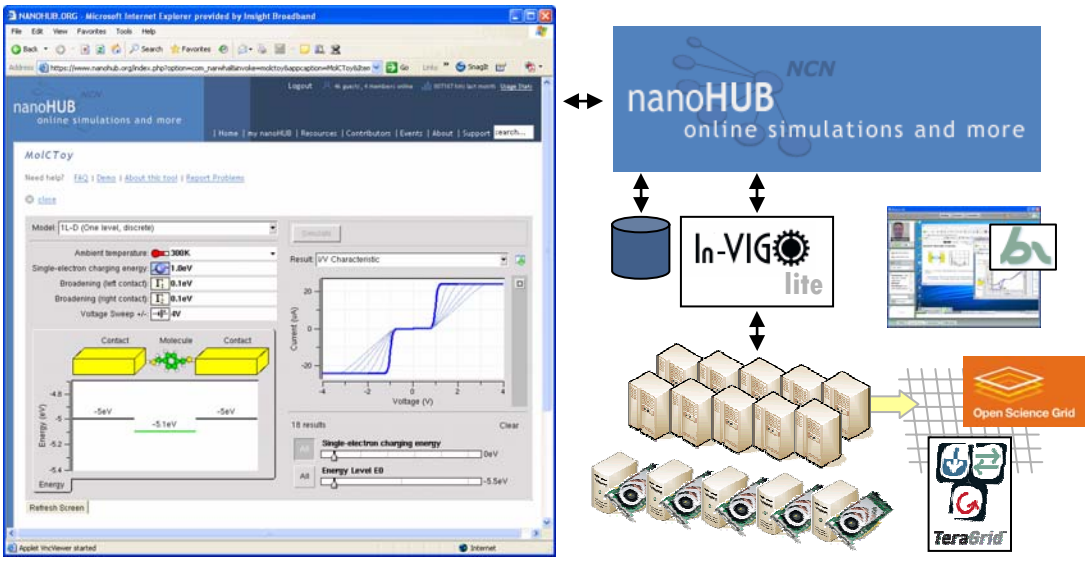
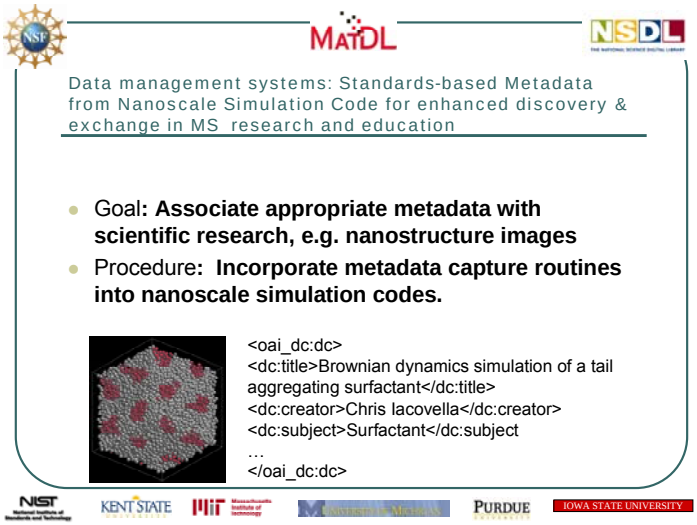


Figure 8: nanoHUB.org provides web-based access to more than 40 simulation tools running on clusters and national grid resources.

Materials science research groups, digital libraries, & education: Metadata from nanoscale simulation code for enhanced discovery and exchange

Associating appropriate and standards-based metadata with scientific resources is essential for their interpretation, especially when they are shared with collaborators, students, or the broader scientific community. As part of the National Science Digital Library (NSDL) the Materials Digital Library Pathway (MatDL), information scientists and materials scientists are collaborating to capture, in Dublin Core (DC) XML format, an optimal description of nanoscale computer simulation output as research codes are executed. Metadata capture routines have been incorporated into the simulation codes of research groups. Input parameters and associated values that determine the output of nanoscale simulations are recorded in order to capture the metadata material scientists consider important. This strategy has the advantage of creating more granular, consistent, and accurate metadata from input file parameters and values that can be associated immediately with simulation output. This approach also eliminates duplicate effort and possible recording errors that could occur if the metadata were produced by a separate mechanism at a later time. Ultimately, streamlining the process of submitting these resources to a digital library and eliminating the necessity for hand-generating metadata, should increase the likelihood of submissions to digital libraries for broader dissemination and preservation of scientific data. Capturing and associating metadata with research data as it is generated will enable the data, including non-print materials, to be more easily located and interpreted within a lab, a group of labs collaborating through a distributed network, as well as for use in the classroom and for long term preservation within a digital library.



In the future, distributed information infrastructures, such as the NSDL, its Materials Digital Library Pathway, and Fedora (flexible, extensible, digital object repository architecture, middleware software) unobtrusively integrated into the workflow of MS research groups, can help facilitate communication and collaboration in research and education as well as interactions between the two. Automatically incorporating standards, such as Dublin Core metadata and its extensions for materials science as MS research is generated, can help ensure the rapid, easy, and rich transfer of data, including annotations and comments among research groups for exchange as well as among teaching groups for classroom use. Ultimately and in partnership with others, such as national user facilities, such as CHESS, the "repository-ready" data and associated research output, such as simulations, codes, and preprints, can be easily and accurately archived supporting open access, reuse, and preservation of the scientific record.

Remote learning, remote participation

Cyberinfrastructure presents a unique opportunity to engage people in scientific learning and discovery. Simulation tools and supporting resources can be distributed via the web to reach under-represented minorities, elementary grades K-12, and other groups that traditionally have not had access to these resources. Workshops can help introduce teachers to new materials. However, the real opportunity of cyberinfrastructure is to keep participants engaged between workshops, and to reach many more who could not attend. Seminars delivered during workshops can be packaged and distributed on the web, providing a library of reference materials for educational development.

Cutting edge research in the classroom

New curricula are needed to teach Materials Science concepts at all levels, pushing the latest research into the classroom at the graduate level, and trickling down to undergraduate and K-12 levels, as well as lifelong education. Lessons should be supported by virtual lab experiments which access remote instruments and simulations, teaching concepts through hands-on activities. Here, cyberinfrastructure acts as a conduit, providing these experiences for all students, including those who would not have access otherwise.

Research engagement at an early, and influential, level of educational development is a vital element for recruiting the young people into scientific careers that are needed to sustain the future competitiveness of the United States. This is virtually absent in the current science education process. Cyberinfrastructure has the potential to reverse this trend to make early scientific education inspiring and effective and to reach a broad segment of society. This issue transcends scientific disciplines, but Materials Science has a special role because of the human-scale and physical nature of the problems

being studied and the direct relationship to the technologies that populate our everyday existence.

CI for teaching advanced topics in materials science

CI tools can fundamentally change the teaching of advanced topics. In the classroom and for homework, students will explore science questions by accessing metadata and by matching these with the most advanced theories from anywhere in the world as soon as they become available. With the right CI tools at hand, students will be able to integrate results from several theory groups. For instance, a student will be able to virtually synthesize a macromolecular material under specified reaction conditions, predict the molecular architecture of a polymer produced, predict its properties through molecular dynamics theory, model their processing dynamics, and predict the resulting morphology and end-use properties. Thus, a student can span the entire range of a polymeric material ("material by design") and predict final product properties without ever having made the material or done the physical testing. Within a class session (45 minutes or less), the student can generate synergy between experimental data and theory. Eventually, seemingly disparate theories and experimental observations will be linked and taken to their limits, thereby leading to unexpected results and new questions. New ideas can be explored rapidly and comprehensively, with the objective of discovering and providing novel materials that are important for society. New teaching methods will be required for efficiently using such powerful tools.

Students will particularly benefit for the following reasons:

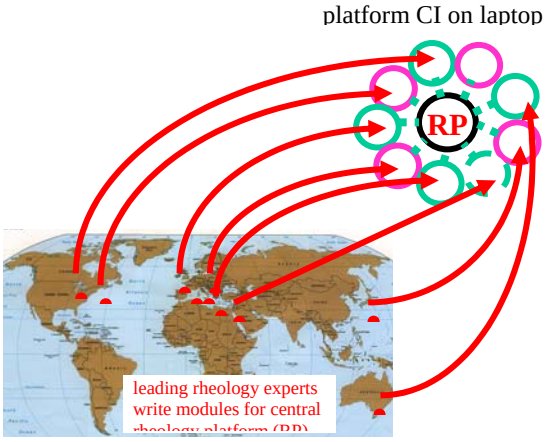
1 Career Acceleration and making the field accessible to larger numbers of students: CI removes barriers by allowing students early in the education to explore real experiments and perform studies with theories which would ordinarily be inaccessible to them because of lack of expertise, equipment, and time.

2. Creativity: CI stimulates creativity by allowing students to interact with the best minds around the world

3. Active learning: the program allows further independent study. Exchanging ideas with peers by sharing metafiles in the classroom or when working on take-home projects

3. Discovery: working with real theories and real data in a new environment will potentially allow the student to discover a novel phenomenon or a novel material. Instant feedback conveys the excitement of the research process: having an idea and being able to test it and see if it is correct.

4. Global interaction: puts student in touch with groups around the world that specifically work in his/her chosen field of interest, affecting the choice of future places for advanced degrees and employment



For the specific materials science area of polymer rheology, Winter and Mours (2006) developed such CI tools and used them in classroom teaching, (see the figure on the left). Typical features are ease-of-use, fast response, interactivity, metafile production and sharing. It is straightforwardly operated on a laptop-PC as typically used by students.

With CI, a student is able to infuse experimental data with cutting-edge theories (including multiscale models and simulation) of globally distributed expert groups in nine countries (Australia, Belgium, Germany, Israel, Italy, Japan, Switzerland, UK, USA). Future extensions of the polymer CI will have the potential of directly connecting variations in molecular architecture with changes in processing and end-use properties. Merging experts through a CI platform is generic and can be extended to other topics and other material groups. It also will gain from integration into grid technology and other middleware tools.

Winter HH, Mours M (2006) The cyber infrastructure initiative for rheology. Rheol Acta 45:331-338; DOI 10.1007/s00397-005-0041

5. Materials cyberinfrastructure imperatives

The development of quantum simulation codes also relies on the availability of a supporting software infrastructure, e.g. in the form of efficient parallel linear algebra software. All these aspects of cyberinfrastructure are essential to the success of quantum simulations.

The materials community needs:

5.1 Accessible, reusable, maintainable software for research, education and outreach

Reusable extensible software for data analysis and modeling

In the early days of computing, physical scientists were at the forefront of computing developments. For example, David Sayre, a crystallographer, was involved in the early development of the hugely successful FORTRAN language. The physical sciences are still influential in CI advances (for example, "earth simulator" is the name of the, until recently, most powerful unclassified computer in the world). However, enormous strides are being made in CI innovation in the commercial sector driven by gaming, business, medicine, finance and so on. This is certainly evident in software engineering where modern practices result in high-quality software that is modular, maintainable, extensible and reusable. These are not characteristics of much of the current materials science software. A closer collaboration between materials science, applied mathematics and computer science, together with funding for software construction of a professional standard will certainly yield scientific dividends. The software should be developed to professional standards of coding, testing and documentation and be in the public domain and widely available. Core libraries and modules (in a modular programming paradigm) should be maintained, extended and curated to meet quality assurance standards. This will result in significant reuse and provide basic computing tools that will allow programming knowledgeable scientists to give free reign to their creativity in the scientific domain. These tools should also be used to create applications in response to the community's needs for the non-programmers. One potential model for this kind of activity that satisfies the needs outlined above would be a software user facility, modeled on traditional user facilities. Necessary core infrastructure (the accelerator in a synchrotron source, the computers in a software user facility) is maintained by expert engineers, and developed by expert scientists (accelerator physics/computer scientists). Beamlines at the facilities are developed, maintained and run by domain scientists with technical and engineering help. In the computing analogy, scientist/programmers would develop scientific software in response to the needs of their users, ensuring it is science driven software, with help from programmers. In the future, integration of different techniques (e.g., combining scattering and spectroscopy), and of theory and experiment, will be important for solving the complexity of problems we are studying. Having a centralized structure to the facility will aid the development of standards and common data-structures, helping the integration process. It will also help scientific cross-pollination, as happens at the facilities, as people work together to make progress on complex problems. Team development of software is also widely acknowledged to lead

to higher quality, more long-lived software. Any such data/software user facility would benefit from centralized organization and interchange of information within and between teams. However, CI would allow it to be geographically distributed in principle. There is one imperative for such a facility: that it have a transparent and fluid interface with the scientist users. The software development must be driven by scientific need.

Reusable software for first-principles simulations

First-principles simulation (e.g., DFT, QMC, etc.) is an example of software technology that has required substantial investments for its development. A handful of first-principles codes currently dominate the market and are available under various licensing agreements (both commercial and open source). It is an important goal to ensure that the materials research community has access to state-of-the-art simulation codes without having to duplicate the considerable investments in software development made over the past twenty years. However, the few simulation codes available today face a maintenance crisis. As first-principles simulation codes have become larger and more complex over the years, their maintenance has become correspondingly more burdensome. The process of porting and optimizing a code for a new platform can involve months of work even for expert developers, even though that activity is not traditionally considered a research activity. As a consequence, it has often been difficult to support code maintainers financially through traditional funding channels. The availability of simulation codes today is essentially the result of the work of a few dedicated individuals, who typically have insufficient time to set aside for code verification. It is desirable to improve upon the current situation by encouraging code development and maintenance through specific, dedicated funding streams. It should be recognized that the existence of multiple simulation codes is essential for successful code cross-verification and to ensure a healthy level of competition in the search for highest performance. A similar situation exists in other areas of scientific software such as data analysis and regression codes such as structure solution and refinement and these arguments apply equally well in this broader domain.

Rich software libraries

Software development can be greatly accelerated through the use of libraries and toolkits. This approach is clearly evident in the area of graphical user interfaces, where developers combine buttons, menus, and other standard toolkit components to create new applications. Using a toolkit not only reduces work and accelerates development, but also improves uniformity and usability of the resulting applications. A cyberinfrastructure for Materials Science research should include the development of libraries and toolkits for numerical solution algorithms, database access, visualization, and user interfaces.

One example of a toolkit supporting scientific research is Rappture, the Rapid APPlication infrastrucTURE toolkit, being developed by the Network for Computational Nanotechnology (NCN). Rappture is available as open source at http://www.rappture.org, and is currently being used by more than 70 nanoscience simulation tools under development at universities throughout the country. Rappture accelerates development, deployment, and maintenance of software by taking the grunt work out of user interface development. It analyzes a description of the inputs and outputs for a simulator and generates a corresponding graphical user interface (GUI) automatically. The result is a cross-platform GUI that runs in Windows, Linux, and Macintosh environments, and can also be deployed over the web via the middleware at

nanoHUB.org. Rappture supports a variety of programming languages, so the underlying simulator can be written in C/C++, Fortran, Python, Perl, Matlab, Octave, or Tcl. The Materials Science community should take advantage of such toolkits, extend their functionality as needed in this domain, and promote the creation of new toolkits.

5.2 Tools for remote collaboration

Enabling remote collaboration with CI

With both the environmental and dollar cost of fuels at a premium, yet the need to work together to solve scientific, technological and societal problems also paramount, the need for effective tools for remote collaboration has never been greater. CI is playing an enormously important role in this domain. Of all the computational parameters such as processing power, dynamic memory, and magnetic memory, bandwidth is scaling at a faster rate than any other. This is really revolutionizing the possibilities for remote collaboration and building virtual communities of remote researchers. Online communities have become an important social phenomenon, especially among the younger generation, enabled by CI developments. These are beginning to be used effectively by scientists for sharing knowledge, data and ideas over large geographical distances. It is now possible to share a computer desktop from across the world, see and hear a talk that is being given overseas, or maybe even was given at a different time. Email is ubiquitous, but also chat, voice over internet protocol communications, streaming video, whiteboard software for drawing, blogs, wikis, file-sharing, web services, portals and so on, are revolutionizing how we interact with each other scientifically. With these capabilities becoming available now through cell-phones and hand-held computers, this mode of operating is only set to grow. Face to face meetings will always have an important place in our activities, but highly effective remote collaboration and, even student mentoring, is now becoming a real possibility. These developments are also highly important for expanding participation, especially to less well funded institutions and underrepresented groups in the US and overseas at relatively modest cost. This kind of virtual research network is set to grow tremendously in the upcoming years, with modest investment.

CI tools for collaborations at national user facilities

With active programs for visiting students and scientists, the national user facilities are ideally positioned to become “centers of collaboration” and provide test beds for CI tools aimed at creating collaboratories and diverse communities. Faculty mentors, or remote collaborators, might be able to see and interact with virtual interfaces to experiments and data analysis software, all while conversing with graduate students stationed at the facility instruments. Well ahead of its time, the UARC project discussed in the Atkins report demonstrated how cyberinfrastructure (often invented by the researchers at the time) was able not only to erase the problem of distance between collaborators, but the tool also created a new paradigm for teaching by allowing students to “converse” with scientists across the globe as well as review a history of conversation via a complete electronic log of all conversation that had taken place previously. Since that project, many new cyber tools have emerged, both commercially and from the open-source development communities, which allow greatly improved web-based communications, database interface, and portal development. NUFs need to explore these new tools to enable and energize collaborations centered on remote data collection and data analysis.

CI tools to lower barriers to use

CI tools could lower the barriers to access for many underrepresented communities by providing efficient, real-time access to instruments at national user facilities and other shared facilities via remote control and data collection via web tools. K-12 outreach programs could motivate young scientists with hands-on lessons about atoms and molecules using remote-controlled electron or x-ray microscopes.

5.3 Materials informatics

Cyberinfrastructure for materials informatics

Ultimately the "processing-structure-properties" paradigm that forms the core of materials development is based on understanding multivariate correlations and their interpretation in terms of the fundamental physics, chemistry and engineering of materials. The field of materials informatics can advance that paradigm in a significant manner. A few critical questions may be helpful to keep in mind in building the informatics infrastructure for materials science:

a) How can data mining/machine learning best be used to discover what attributes (or combination of attributes in a material may govern specific properties Using information from different databases, we can compare and search for associations and patterns that can lead to ways of relating information among these different datasets.
b) What are the most interesting patterns that can be extracted from the existing material science data? Such a pattern search process can potentially yield associations between seemingly disparate data sets as well as establish possible correlations between parameters that are not easily studied experimentally in a coupled manner.
c) How can we use mined associations from large volumes of data to guide future experiments and simulations? How does one select from a materials library, which compounds are most likely to have the desired properties? Data mining methods should be incorporated as part of design and testing methodologies to increase the efficiency of material application process. For instance a possible test bed for materials discovery can involve the use of massive databases on crystal structure, electronic structure and thermochemistry. Each of these databases by themselves can provide information on over hundreds of binaries, ternary and multicomponent systems. Coupled to electronic structure and thermochemical calculations one can enlarge this library to permit a wide array of simulations for over thousands of combinations of materials chemistries. Such a massively parallel approach in generation new "virtual" data would be a daunting if not impossible were it not for data mining tools interfaced with a mechanism of accessing data libraries that are distributed and heterogeneous; and hence a cyber infrastructure for cyber discovery.

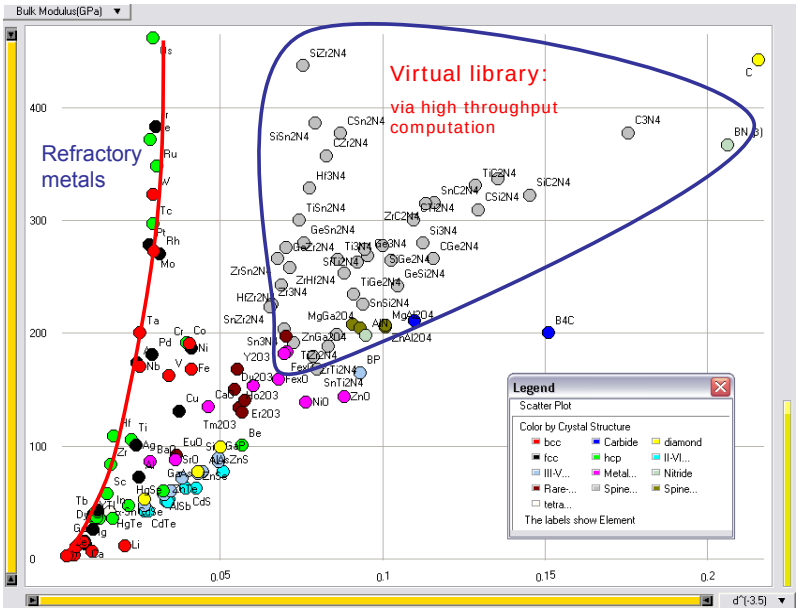


Figure 9: A "Structure–Property" map combining theoretical and experimental data with data-mining tools to predict properties of yet to be characterized crystal chemistries ("virtual materials"). Using data mining tools we have been able to reproduce known data which helps tests the robustness of our predictions. The informatics approach when coupled to access to diverse and distributed data (through a cyber infrastructure) can help to significantly accelerate materials discovery. [*Virtual Screening and QSAR Formulations for Crystal Chemistry*: C. Suh and K. Rajan ; Journal of QSAR and Combinatorial Sciences Vol 24, 114 (2005)]

Databases to improve the graduate student experience and speed education

Initially, graduate students have little knowledge about their first projects. As a first step to tackle their given problems, they try to get familiar with materials by spending time searching for the information about the materials and properties that are already known. Despite an enormous amount of time they put in, the results are usually not successful. Part of the reason is due to the lack of their knowledge in the field but mostly due to the lack of a good, reliable materials database and a search engine.

As a second step, they try to understand what their experimental data really means. Most of the time, they discuss their data with their advisers. However, to be a successful researcher, graduate students should have an ability to correctly interpret data on their own. One possible way is through comparison to comparable studies conducted on similar systems (materials with similar properties, structures, or compositions). However, this takes an enormous amount of time. They do not know exactly which materials to search. Even if they find a paper for a related material, running around to several libraries and surfing tons of web pages often fails to yield the specific information they are looking for. With a good reliable materials database and search engine, graduate students can spend much more time on real scientific problems.

As a third step, they would get their own new ideas based on their results. However, it is very difficult to test out the ideas. It requires synthesizing other materials.

It is impossible to know which materials to look at because we are looking for a previously unknown correlation. A good reliable materials database and search engine can be a graduate students' virtual laboratory to test out their ideas quickly; look for all the materials with a certain structure, sort them out and see if any of the materials shows a certain property we are interested in. It is very important that materials database must be high quality, reliable, well maintained, and with clear standards. Frequent users of such databases will be graduate students who are new in the field and have not developed good judgment on good and bad data.

A good reliable, and well maintained materials database and search engine

1. can accelerate graduate students collecting pre-knowledge
2. can accelerate graduate students in analyzing their data
3. can help graduate students avoid going in the wrong direction on their research
4. can accelerate graduate students finishing their projects
5. can be a virtual laboratory of graduate students where they can quickly test their new ideas

Data interchange standards

The appearance of a new data management infrastructure (e.g., materials-oriented web portals) will greatly accelerate the exchange of scientific information, both for research and education purposes. Data standards in materials science will greatly facilitate this enterprise. This should not be limited to data generated in a physical experiment. Among the many types of information being exchanged, *simulation data* is rapidly growing in size and importance. The results of computer simulations are used to help interpret experiments, but also to validate theoretical models. It is therefore highly desirable to make them available to the scientific community. Modern experiments and simulations are often generated using expensive facilities and high performance computers computer time. They should be stored and reused in order to avoid wasting resources. Sharing data naturally raises the issue of adopting common data representation standards. The difficulty in rallying a community around a common standard has been recognized in many scientific areas, and materials science is no exception. Some efforts have been initiated in the first-principles simulation community to adopt XML-based syntax for the representation of simulation data [1,2] This leads to the possibility of automatic data translation between the representations used by various codes. Similar discussions are taking place in the worldwide neutron scattering community, with the development of NeXuS data standards [3]. Such efforts are still in their infancy, and can involve complex technological and sociological challenges, but have the potential to accelerate enormously the exchange of data and bring forward greater integration of experiment and theory.

[1] qmcpack, QMC XML Schemas, http://mccweb1.mcc.uiuc.edu/software/display/qmcpack/QMC+XML+Schema
[2] Web standards for quantum simulations, http://www.quantum-simulation.org
[3] http://www.nexus.anl.gov/

5.4 Shared facilities

Using CI to leverage shared and national user facilities

Shared user facilities (SUFs) are multi-user, cross-disciplinary laboratories and centers that provide access to unique capabilities for materials science research and training to a very large community of scientists and students each year. SUFs range is size from

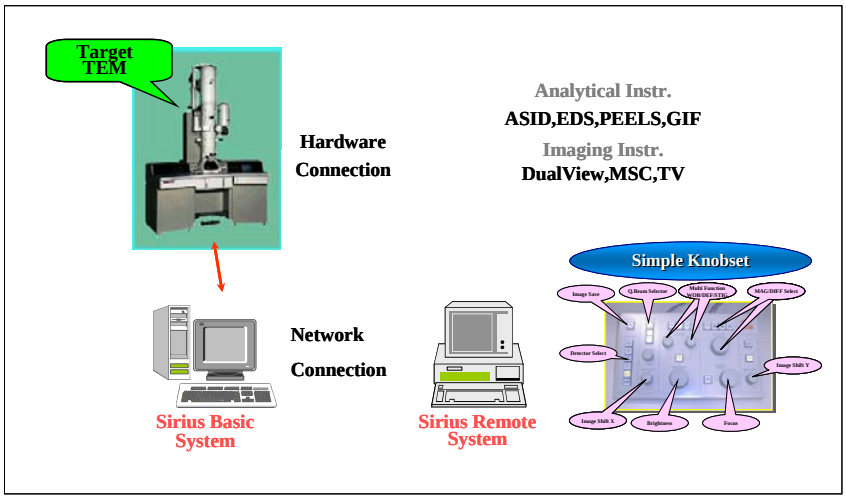
building- or campus-based instruments shared by two or more research groups all the way up to National User Facilities (NUFs) hosting hundreds or thousands of visitors each year. It is estimated that more than 20,000 scientists use NUFs each year typically for between one and seven days. This number does not include the vast cadre of research collaborators who do not visit the facility but still depend upon data for analysis or modeling. Since nearly half of the scientists visiting NUFs are graduate students or post-doctoral associates, the US should expect increased demand and usage in any foreseeable future.

SUF construction, maintenance and support involves many government agencies: NSF, DOE, NIH, and branches of the military and industry. NSF-DMR National Facilities include the NIST Center for High Resolution Neutron Scattering, the Low Energy Neutron Scattering Facility at Indiana University, the Cornell High Energy Synchrotron Source, the Synchrotron Radiation Center at Wisconsin, and the High Magnetic Field Laboratory at Florida State University. Many other NSF-supported materials research centers are part of the NUF/SUF portfolio, including the National Nanofabrication Infrastructure Network (NNIN) and Materials Research Science and Engineering Centers (MRSEC) found at many universities across the US.

Research at many SUFs, and certainly all the NUFs, is often markedly different from home laboratories. Scientists are away from home and the types of data they collect, the high data collection rates and the extraordinary large volume of data present unique challenges. They are often constrained to use only portable tools and specimens, and incur extraordinary research expenses for travel and support. Usually, the visiting scientists represent only part of the larger research group involved in a research program. Because all NUFs are distant from home laboratories, quite often scientists visit many different facilities during the course of an experimental program. Because of this, compatibility of remote data sets, storage and transport are all difficult issues. Improving CI could have an enormous positive impact on the scientific output, educational missions and technological dissemination of SUFs. Today, the software used at national user facilities is often purpose-built and unique. In the future, national user facilities could expand their roles to become scientific data analysis centers as well as data collection centers. The tools needed to analyze and visualize data from large instruments could be developed through collaborative efforts, possibly following an open-source model with global scope, and distributed freely with data, further expanding the roles of the NUFs as centers for collaboration, education and dissemination of new technologies. Alternatively, this activity could take place in a distinct software user facility described in §5.1.

Remote Utilization of Instruments through CI:

Advanced instruments for materials research often are not available in every research laboratory because of cost, maintenance, and other issues. CI enabled remote access is required to better utilize the equipment and to lower the barrier to access to the instruments. An example of CI enabled remote access is to electron microscopes. Here, a remote control system is set in the satellite institution's laboratory connected, through the CI, to the hub institution's instrument. A sample is inserted into the TEM at the by a technician at the hub but then control of the instrument is handed to an operator at the remote location. Current implementations allow all modes of operation of the instrument. Multiplexing amplifies the benefits of this system. The remote handling CI allows the remote institution to access instruments at several different hubs, as well as the hub



serving multiple satellites. This leverages major investments in instrumentation, leads to novel scientific discoveries, and broadens participation. The educational possibilities are also apparent, especially if remote controls can be made portable so that they can be taken to the classroom. We recommend expanding the scope of this kind of remote control to different facilities, including nation user facilities.

5.5 Better algorithms

The ultimate goal in materials simulations is to be able to perform finite-temperature *ab initio* relativistic quantum calculations on Avogadro's number of atoms at a chemical accuracy of 10^{-7} K for time scales of hundreds of years. The community is extremely far, some would argue impossibly far, from this goal. Furthermore, it is clear that no foreseeable increase in computing power can achieve these goals. It is only through the advancement of the underlying theory and associated efficient algorithms that there is any hope of coming close to achieving these goals. Furthermore, because we have not achieved this ultimate goal, there is a need for course-grained theories and models, together with multiscale modeling advances that allow predictive and exploratory studies for the study of specific materials.

Faster than real time for bridging disparate time scales

Consider the time scales involved in magnetic materials and for magnetic recording. The basic underlying time scale is given by the inverse phonon frequency (say 10^{-13} seconds), the time scale for writing a bit should be comparable to processor clock speeds (say 10^{-9} seconds), the time scale over which you want your data to remain intact is somewhere between a human lifetime (say 10^{2} years or 10^{9} seconds), and geological time scales (say the time scale for magnetic reversals of the Earth's field which are measured using geomagnetism, say 10^{5} years or 10^{13} seconds). Given that the time scales involved are much faster than today's processors and much longer than a human lifetime, studying magnetic materials requires faster-than-real-time dynamic simulations. The ***only*** way to perform such simulations is through the use of faster-than-real-time algorithms --- no future cyberinfrastructure can hope to bridge these disparate time scales. One current example, from dynamic Monte Carlo simulations of discrete magnetic models, of a faster-than-real time algorithms is projective dynamics. Without changing the underlying physical dynamic, it has enabled simulations of 10^{60} algorithmic time steps (roughly the square of the age of the universe in femtoseconds) [1,2,3]. Further work on faster-than-real-time algorithms are needed to apply to materials systems with continuum degrees of freedom. Related faster-than-real-time approaches along these lines that are useful for realistic rare-event simulations for a few hundred

atom systems include algorithms developed by Voter and coworkers such as the accelerated dynamics for molecular dynamics [4], self-consistent mean-field Langevin/molecular dynamics methods [5]. Furthermore, faster-than-real-time algorithms and CI-enabled experiments are important to simulate, validate, understand and alleviate the rare events that cause failure of materials and devices or the rare events that lead to nucleation of crystals. It is only through cross-disciplinary interactions with computer scientists, mathematicians, and theorists that such algorithms have any possibility to allowing materials research to approach the ultimate goal.

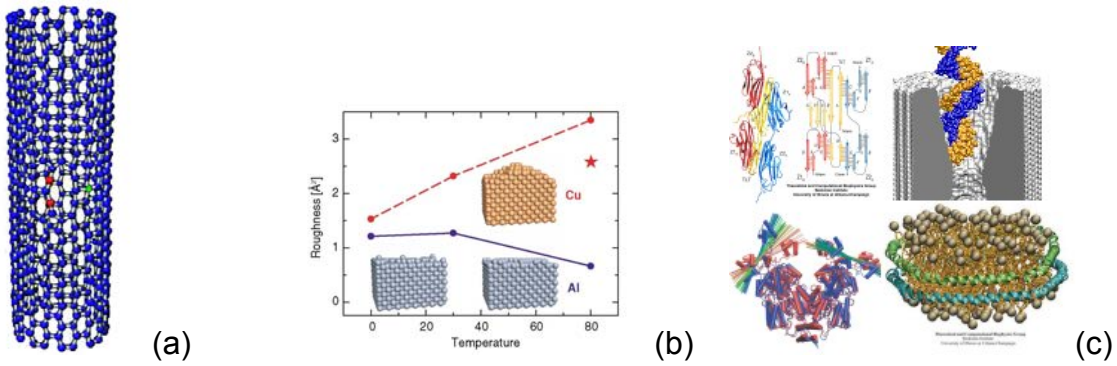


Figure 10: Examples of materials simulations enabled by advanced atomistic algorithms. (a) Defects in a carbon nanotube that is undergoing strain. (b) From reference [6] (c) An example of materials at the bio- nano- chemical- overlap of subfields.

[1] "A projection method for statics and dynamics of lattice spin systems", M.Kolesik, M.A. Novotny, and P.A. Rikvold, Physical Review Letters, **80**, pages 3384-3387 (1998).

[2] "Monte Carlo with absorbing Markov chains: fast local algorithms for slow dynamics", M.A. Novotny, Physical Review Letters, pages 1-5 (1995), erratum vol. **75**, p.1424 (1995).

[3] "Extreme long-time dynamic Monte Carlo simulations for metastable decay in the d=3 Ising ferromagnet", M. Kolesik, M.A. Novotny, and P.A. Rikvold, International Journal of Modern Physics C, **14**, p. 121-131 (2003).

[4] "A method for accelerating the molecular dynamics simulation of infrequent events", A.F. Voter, Journal of Chemical Physics, **106**, p. 4665 (1997).

[5] "Self-Consistent mean-field model based on molecular dynamics: application to lipid-cholesterol bilayers", G.A. Khelashvili, S.A. Pandit, and H.L. Scott, Journal of Chemical Physics, **123**, p. 34910 (2005).

[6] "Multiple time scale simulation of metal crystal growth reveal the importance of multiatom surface processes", G. Henkelman and H. Jónsson, Physical Review Letters, **90**, 116101 [4 pages] (2003).

Embedding theories that bridge multiple length scales

At the heart of the challenge of complex materials modeling is the need to bridge multiple length and time-scales. Consider, for example, a complex surface reaction such as the corrosion of plutonium that is the dominant degradation pathway governing the long-term stability of the nation's strategic nuclear deterrent. Corrosion begins at the molecular scale with attack of water at surface sites to form hydrated plutonium oxides, progresses onto mesoscopic scales with diffusion of oxygen into the interior and the formation and propagation of defects, and proceeds to macroscopic scales with crack propagation as well as bulk thermodynamic transformations between the competing structural phases.

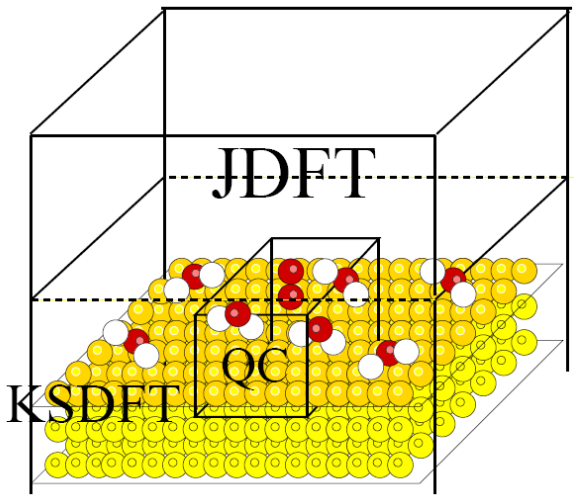


Figure 11: Embedding theories for a surface chemical reaction such as corrosion: High level quantum chemical methods for the surface active site (QC) are embedded within a Kohn-Sham Density Functional (KSDFT) description of the bulk material and first solvation layer, which is further interfaced with a continuum theory of the bulk solvent (JDFT) .

At each of these different length scales, there is an appropriate level of theory and modeling. For the correlated electrons of plutonium and their oxidation chemistry, we need high level quantum chemical modeling and even fundamental new theories of correlated electronic structure. For mesoscale phenomena associated, for example, with surface reconstruction and defect formation, density functional methods or inter-atomic potentials together with molecular dynamics are appropriate. Finally, at longer length scales, we can use continuum descriptions or finite element representations of the stress and strain of the material (see Figure 11). The greatest challenge, however, lies not with the difficulties in modeling each of the scales separately, but in the fact that we must consider the coupling *between* the different scales, which is characteristic of a true multiscale problem. Thus shorter scales feed into longer scales, for example, via the dependence of interatomic potential parameters on the underlying quantum mechanical electronic structure, while longer scales feed back into shorter scales, for example, via screening of charges by bulk polarization and the dielectric constant. To describe these types of coupled multiscale materials questions, we must develop new algorithms that will allow us to bridge disparate length scales and embed different models and theories within each other. Such infrastructure will take the form of two research directions:

(1) *New embedding formalisms and renormalization group procedures*: At a theoretical level, we must specify new formalisms by which descriptions employing physical variables on different scales can be embedded within each other in a rigorous manner. These will include purely quantum mechanical embedding e.g. to link correlated electronic structure and quantum chemistry methods with density functional theory; quantum-classical interface methods, to join density functional and molecular dynamics theories, and purely classical embedding, to bridge atomistic descriptions within continuum environments. In addition, Renormalization Group procedures must be devised to provide a first-principles link between coarse-grained theories and those at finer levels.

(2) *Interoperable software components and scalable multilevel algorithms*: At the software level, it will be necessary to identify the common set of physical variables and parameters that are to be used to connect different scales. Common input and output formats as well as applications programming interfaces must be developed and documented for popular program packages at each of the modeling scales. Such efforts will necessarily involve multi-language software implementations and the investigation of scripting frameworks, e.g., via Python-callable modules should be encouraged. Finally, new algorithms to address the challenging problem of mapping hierarchically organized multiscale algorithms and embedding theories onto massively parallel petascale architectures with tens or hundreds of thousands of processors, must be developed.

Comparison of results from different computational methods and theory-driven experiments needed for ***validation*** and ***verification*** (figure 12 is an example of combined experimental and computational study of the structure of a ceramic grain boundary). Cyberinfrastructure developments are important to facilitate connections among computational methods and between computation and experiment.

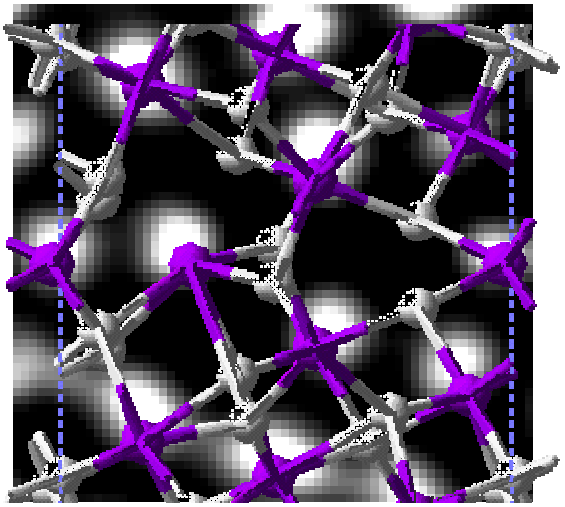


Figure 12: Computational density functional theory calculated model of a ZrO_2 tilt grain boundary overlaid on a transmission electron microscopy image of the grain boundary. The computational study provided information about the positions of the oxygen atoms and the nature of chemical bonding at the grain boundary not easily available from experiment. At the same time, the experimental data was crucial in focusing the computational study. From "Ab-initio Calculations of Pristine and Doped Σ5 (310)/[001] ZrO_2 Grain Boundaries", Z. Mao, S.B. Sinnott, and E.C. Dickey, *Journal of the American Ceramic Society* **85**, 1594-1600 (2002).

Simulations with energy resolution at the ambient

Creating "materials by design" is a holy grail of Materials Science. However, the computational simulations used in the design have energy precisions much worse than thermal fluctuations under ambient conditions. For example, with modern density functional theory (DFT), the energy precision is currently of order 1000 K, a factor of 10 too high. Methods with much higher precision are possible in principle (e.g., with quantum Monte-Carlo techniques), though are not computationally feasible on systems

of an interesting size. The development of algorithms that solve this problem will revolutionize Materials Science and materials design. A major CI initiative should therefore be directed toward the development of efficient algorithms that go beyond DFT that achieve a factor of ten better accuracy.

Algorithms for efficient use of massively parallel computers

One of the most difficult problems to solve in CI is the problem of efficient use of massively parallel computers. Even on the PC level, as dual-core PCs morph into multiprocessor PCs, this difficulty will be encountered at ever-lower levels of computing power. With ITR funding from NSF DMR, researchers have been able to use ideas from materials theory where multi-particle systems are treated through the use of statistical mechanics and non-equilibrium surface growth to address parallelization issues. In particular, they have been able to show that a large class of non-trivial parallelization problems, namely short-ranged discrete event simulations, can be made *perfectly scalable* given a particular computer architecture (which may be heterogeneous, but must include small-world connections between processors), software, and algorithms [1-4].

[1] "From Massively Parallel Algorithms and Fluctuating Time Horizons to Non-equilibrium Surface Growth", G. Korniss, Z. Toroczkai, M.A. Novotny, and P.A. Rikvold, Physical Review Letters, volume 84, pages 1351-1354 (2000).

[2] "Suppressing Roughness of Virtual Times in Parallel Discrete-Event Simulations", G. Korniss, M.A. Novotny, H. Guclu, Z. Toroczkai, and P.A. Rikvold, Science, volume 299, p. 677-679 (2003).

[3] "Fully scalable computer architecture", U.S. patent number 6,996,504, M.A. Novotny and G. Korniss, inventors, issued February 7, 2005.

[4] "Synchronization landscapes in small-world-connected computer networks", H. Guclu, G. Korniss, M.A. Novotny, Z. Toroczkai, and Z. Rácz, Physical Review E, volume 73, article number 066115 [20 pages] (2006).

Investment in fundamental computational infrastructure

The DOE's program of Scientific Discovery through Advanced Computing (SciDAC) has focused attention and resources on a scientific software infrastructure to support multiple applications. Abstract mathematical tasks such as solving a linear or nonlinear systems, integrating a stiff system of differential equations, generating a mesh for a complex domain, adapting a mesh to an evolving solution (given a user-supplied error estimation criterion), and visualizing multiple grid functions in three-dimensional space-time, are common to a vast range of applications beyond materials science, from astrophysics to molecular biology. Traditionally, applications scientists selected from a vast array of software libraries designed to run on a single CPU, or wrote their own specialized routines. Such libraries are few and far between for applications that must scale to 10^5 processors and run acceptably efficiently on machines costing $100M at purchase and millions per month in operating expenses. The DOE's investment in the development and maintenance (including porting to new architectures) of such libraries in the SciDAC program is an excellent start. However, there are holes in the existing software portfolio in areas of relevant to materials science (for instance parallel FFTs) and there is a shortage of human expertise to carry the SciDAC software packages into all of the applications groups that could benefit from adopting them.

Since most simulation time is spent solving linear equations, quality assured, efficient open source linear algebra libraries, including codes for parallel architectures, are crucial for efficient computations and are highly leveraged as they get used in multiple codes.

In addition we need smart compilers for parallel computers that can properly load balance and link intraprocedural processes without detailed intervention of the programmer. Investments are needed in these basic tools. Another fundamental area that deserves attention is the need to find better global optimization methodologies. Investment in these areas is highly leveraged because of the wide and repeated use of such libraries.

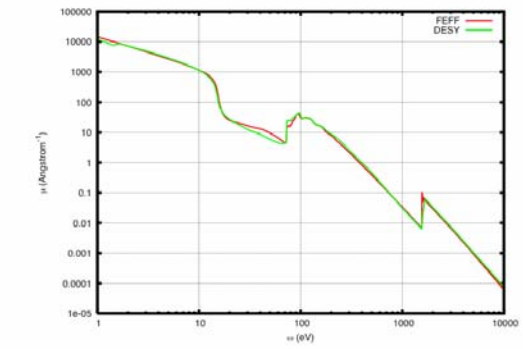


Figure 12: X-ray absorption spectra of Al vs photon energy in eV from the UV to hard-x-ray energies.

Excited states: theoretical tools and algorithms for optical response

A major frontier of modern Materials Research is an understanding of excited states, and their frequency and time-dependent response over a broad range of frequencies from DC through optical to x-ray energies. These response functions are important both in characterizing new materials (e.g., using synchrotron photon sources) and in creating new devices (e.g. in photonics). Theoretical calculations of such properties are challenging but are becoming more quantitative. A notable success story has been the development of *ab initio* codes for x-ray absorption and related spectra. For example the U. Washington FEFF codes [1] are widely used to determine the structure of complex materials at synchrotron radiation sources throughout the world. The development of these codes required advances both in theoretical algorithms and CI. For example, FEFF is based on a real space Green's function approach that is complementary to the wave-function, reciprocal space, approach typically used in electronic band structure codes for condensed matter research. Currently this approach is applicable from the UV to x-ray energies, as illustrated below.

There has also been a considerable effort by many groups to develop tools for optical response in the visible and below [2]. This effort also required the development of new algorithms. However, first principles methods such as the Bethe-Salpeter equation (BSE) or traditional quantum-chemistry techniques are not currently feasible in

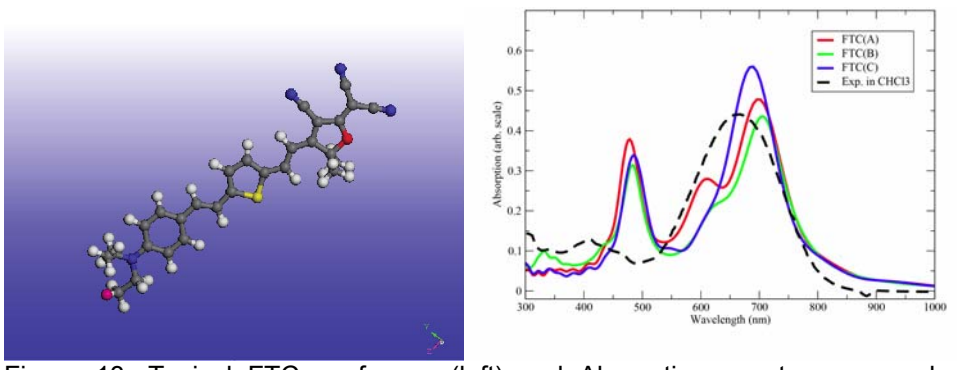


Figure 13: Typical FTC conformer (left) and Absorption spectra vs wavelength for various FTC conformers calculated by the RT-TDDFT code and compared to experiment (FTC in chloroform) (right)

complex materials with current computational resources since they scale badly with system size. Thus there has been an effort to develop other approaches. A very promising method is time-dependent-density functional theory (TDDFT), which is a generalization of ground state DFT to excited states. A number of codes are now being developed for implementing TDDFT. One of the most efficient is a real-time extension of the SIESTA code for both linear and non-linear optical response. For example with the RT-TDDFT code, the optical response of FTC has been calculated.

Advances like these need to be funded. We believe a major advance could now be achieved by combining these approaches in a way to make them generally accessible to materials scientists. In addition advances are needed to achieve more efficient treatments both of linear and non-linear optical response (e.g., with parallelization algorithms to treat larger systems with solvent effects). A successful development would lead, for example, to predictive capability for determining optical properties and response functions. We believe that the use of modern CI tools like graphical interfaces (GUIs), standardized input-output, and combinations of codes should be an important goal. This would contribute to a Materials Science Toolkit. Such a toolkit would combine ground state electronic structure and MD codes for structure with excited state codes for optical response and time-dependent optical response. Such tools could greatly speed the development of materials that would lead to advances in CI, for example, of fully realistic photonic devices including solvent effects and device geometry

[1] J. J. Rehr and R. C. Albers, Rev. Mod. Phys. 72, 621 (2000).
[2] G. Onida et al. Rev. Mod. Phys. 74, 601 (2002).

5.6 Integration and interoperability of software and of communities

Developing libraries and toolkits for Materials Science software will greatly help the integration and interoperability of software. But there is also a social component to this problem. In addition to integrating software, we must also integrate and coordinate the efforts of the researchers creating the software. The same cyberinfrastructure used by the Open Source community can be leveraged in the Materials Science community. Codes should be collected into repositories with source code control, supported by bug-tracking systems, as exemplified by http://www.sourceforge.net. Codes should leverage standard toolkits to avoid duplication of effort. Developers and Universities should be encouraged to release their codes as open source, making their work accessible to the rest of the community. Reviews of funding proposals should look favorably on projects conducted in this manner. An NSF endorsed open-source license could be valuable in this regard. Funding solicitations should also require descriptions of the software development methodology used by any projects which create software deliverables—particularly in the areas of code verification (correctness of algorithms and implementation) and model validation (correctness of theory compared with experiment). Efforts should be initiated to create community databases containing both experimental and simulated results, thereby establishing the benchmarks needed for tool comparison and validation across the entire community.

Community e-laboratories for materials by design

Discovering new materials by design will require an innovative computational system that can combine multi-scale models (atomistic to meso-scale to macro-scale) with experiment to predict structure and process parameters to meet multifunctional performance requirements. Such an E-Lab should be a sustainable, shared community resource which utilizes local and wide-area distributed computing and storage systems, and manages integration, extension and interoperability challenges. Multidisciplinary collaboration between material scientists, computers scientists and applied mathematicians will be required to determine new scalable algorithms for automated verification, validation, information retrieval and mining, and the predictive analyses of the combinatorially large space of process and structure parameters and properties obtained from multi-scale simulations and experiment.

MATCASE (http://www.matcase.psu.edu) is an example of a framework for predicting macro-scale properties of alloys given structure and processing parameters, thus encapsulating the multi-scale simulation aspects of future systems for materials by design. MATCASE combines atomistic, meso-scale and macro-scale models, information bases, and finite-element analysis through an extensible software system, as shown below.

[1.] Z.-K. Liu, L.-Q. Chen, P. Raghavan, Q. Du, J. O. Sofo, S. A. Langer and C. Wolverton, "An integrated framework for multi-scale materials simulation and design," *J. Comput-Aided Mater. Des.,* Vol.11, 2004, 183–199.

5.7 To inspire future generations of materials scientists

High school research outside the classroom

How do you excite high-school students about science and engineering? One of the best ways is to get them into cutting-edge research at an early stage. A relatively small cost (minimum wage is paid) is required to have high school students involved in fundamental research. They should not be doing just make-work studies, but to convey the excitement of science they should venture into an area where simultaneously the research requires only their level of sophistication and is leading-edge in that the results are not known beforehand. One example using three high-school juniors and one high-school senior was a project involving magnetic small-world nanomaterials, where the students built model systems and took statistics of physical networks they built [1].

[1] "Magnetic Small-world Nanomaterials: Physical Small World Networks", M.A. Novotny, X. Zhang, J. Yancey, T. Dubreus, M.L. Cook, S.G. Gill, I.T. Norwood, A.M. Novotny, and G. Korniss, Journal of Applied Physics, **97**, 10B309 (2005).

Lifelong learning

CI has a special role to play in lifelong learning. Most of the advantages that CI brings to education in the college

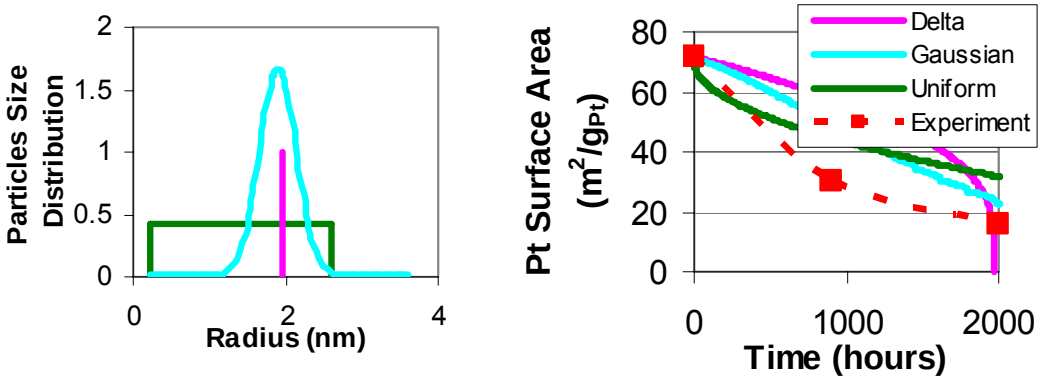
and pre-college domain are transferable to the sphere of lifelong learning. In this case, however, the ability to introduce distance learning and flexible course schedules is particularly relevant as many lifelong learners have to work their learning schedule around busy work and family obligations. CI can have an enormous impact in this regard.

5.8 To bring together experiment and theory

Matching up of experiment and theory has great potential for synergy. With the backing of theory, experimentalists can ask "what if" questions and develop a strategy for the next required experiment. Theoreticians gain from the "reality check" when comparing to an experiment and from predicting material properties that are not accessible to experiment. Important is that there is some overlap between the variables that get measured and the ones that get predicted. Such overlap is not achieved automatically. Model experiments specifically have the purpose of facilitating the overlap with theory. Vice versa, theory gains from including material parameters and variables that are accessible to experiment. CI needs to be developed to bring experimental data and theory onto the same CI platform.

Self consistent use of theory in constrained optimization problems

It is often a non-trivial step to go from an experimental measurement to scientific understanding. In general this will require data processing and modeling and in many cases it requires the solution of an inverse problem; for example, given this scattering data, what was the structure of the material. This inverse problem may require inductive reasoning since there is no direct computation that takes you from the data to the desired knowledge. An example of this is solving structure using crystallography where crucial phase information is lost from the scattered radiation in the measurement. In this case, the structure can be built up using global optimization techniques, where trial structures are tried, the scattering calculated and the process iterated in such a way as to find an optimal structure given the data. Such optimization problems are only well behaved when the amount of information in the data exceeds that needed to specify the



For a nanoparticle Pt catalyst, different particle size distributions (left) can lead to very different surface area loss in fuel cell applications (right). (D. Morgan, University of Wisconsin – Madison, MS&E Dept.)

model; i.e., the optimization problem is well conditioned. Poorly conditioned problems can be regulated by the addition of additional data constraints, or by inputting prior

knowledge from theoretical calculations. We are increasingly studying more complex systems and that fall into this category of being ill conditioned. New algorithms and theoretical understanding are needed and more robust CI that can facilitate the combination of theory and experiment, and different experimental probes.

Simulations to understand properties

Theory, experiments, and computer simulations are different tools to address common scientific and engineering problems and these tools can work most effectively if there is a mutual feedback between them. Examples include solving electrochemical rate equations for the loss of surface area of a catalyst during fuel cycle or solving diffusion equation for transport through complex microstructures. In both of these problems it is often difficult to identify the most relevant mechanisms driving the phenomenon of interest and the experimental parameters for the rate equations are often inferred from indirect measurements. Simulations can bring an invaluable contribution to such problems, by breaking down the process into small sub-problems and determining more accurate parameters as a function of the nano- or micro-structure. An example of the interplay between experiment and simulations at multiple scales can be seen in the challenging problem of understanding loss of Pt catalysts during fuel cell use. Rate equation parameters can be determined from *ab initio* calculations and fitting to experiment, and the kinetic model can then be used to help interpret experiments, e.g., understanding the role of the nanoparticle size distribution on Pt loss (see Figure 16)

Recent development of high-resolution experimental imaging techniques (e.g., atom-probe, TEM) has opened up opportunity for connecting experiments and simulations at yet another level. Namely, the experimental images can be digitized and fed as an input to simulation models (an example is shown in Fig. 16). Simulation can be then performed to predict materials behavior in altered conditions (e.g., pressure).

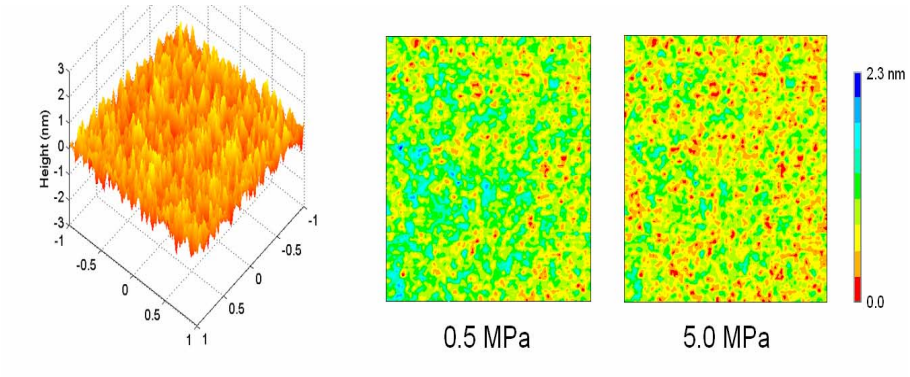


Figure 16: Digitized AFM image (left) is fed into the finite element simulations for prediction of interfacial gap under applied pressure (middle and right) (courtesy of K. Turner, University of Wisconsin –

5.9 Effective scientific data management and curation

Today materials science employs a multidisciplinary and multi-technique approach that requires investigators to combine data from a large complex group of experimental and computational sources for investigating complex material systems and phenomena. The material science community lacks the scientific data management systems to support the complex workflow and burgeoning datasets generated. Computational tools for

materials science must include multiscale simulation in support of experimental planning and analysis; real-time reconstruction and processing of high-throughput data; and data-mining of large and complex datasets. Simulation tools must be further developed to produce results expressed as observables in experiments. The multidisciplinary character requires team approaches with real-time interaction during experiments and calculations where telepresence and telecollaboratory is critical.

In material sciences, various complementary measurements and techniques are required to fully determine the relevant information. For example, the complete structural and dynamical analysis of a material may require data from x-ray and neutron scattering; optical, vibrational and NMR spectroscopies; scanning probe techniques; and electron microscopies. Database comparison with known structures and properties would be a crucial factor in defining the modeling. Computational tools including electronic structure and molecular dynamics calculations would then be employed to compute appropriate response and structural behaviors, to test the consistency of the structure/function scheme. This creates a large and complex workflow with streams of information from multiple tools, techniques, calculations, analyses, coming either from within a common facility or from multiple institutions. The material science community lacks the scientific data management systems needed to support the complex workflow and burgeoning datasets generated. Much of the basic technology does exist to attack this problem, but the cyberinfrastructure tools for scientific database management have yet to be implemented for the material science community. Commercials tools, such as content management systems, relational databases, and portal applications, need to be customized and further develop to support scientific data sets. Appropriately developed scientific data management systems would enable data mining and effective searching; provide controls for provenance, pedigree, versioning, and workflow; provide the means for multilevel or federated linking of data, databases, and schemas; enhance access through portals; address security, interface, reliability, and capacity issues; and address the issues of long-term archival maintenance and access. In addition to development of common cyberinfrastructure tools to support scientific database management systems, development of common metadata systems within and across relevant disciplines in and related to material science will be critical. Individual national user facilities are already facing this challenge. We are already in the situation where far more data can be produced than we can effectively manage, thus losing a significant portion of the value created by the generation of data at these facilities.

5.10 Scalable real time data analysis

More powerful experimental probes with computerized control and data collection result in much more rapid and voluminous data acquisition. In many cases the rate-limiting step in experiments is our ability to process and understand the data. This is expected, but it presents a problem in data collection strategy where information from initial measurements is needed to guide later measurements. If the preliminary data cannot be processed, there is inadequate information to make good experimental decisions during data collection leading to inefficiencies and sub-optimal data. This problem is particularly acute at large-scale national user facilities where experimentalists generally only have access to precious beamtime for a short fixed period of time. Such facilities often generate large volumes of data, and future generations with wide solid-angle detector coverage will be generating Gbytes of data per minute making the data

processing problem acute. Investment in cyberinfrastructure to yield real-time data processing at these high rates will be critical for getting the most science out of these instruments and to avoid costly wastage of beam-resources.

5.11 Access to advanced computing resources at different scales

Use of computers is now ubiquitous in materials research. The diversity of problems studied call for a similar diversity of computation solutions, from laptops and hand-held devices to high end supercomputers. Funding is needed for all sectors of computing to ensure access to the appropriate level of computer for a given problem. A sector that is growing in importance is the use of commodity clusters of fewer than 100 nodes. These clusters are becoming inexpensive but are extremely powerful for many sophisticated computations. Software investments are key to make the most of these resources. Such clusters can be sadly underutilized, or dominated by a few individuals, because of the technical barrier to academics of porting software to run on a parallel architecture. Access is then determined not based on scientific potential of a calculation but on the programming sophistication of the users. A small investment in computer science technical expertise to port academic codes is a worthwhile investment.

The software issues for high-end computing are different but also challenging. These high-end systems generally have novel architectures but short design lives. Investing in software development in advance of machine commissioning is essential to ensure the greatest scientific utilization of computer.

Appendix: Workshop contributors

Prof. Meigan Aronson
(U. of Michigan)

Prof. Laura Bartolo
(Kent State U)

Prof. Lawrence H. Bennett
(George Washington U)

Prof. Simon Billinge
(Chair, Michigan State U)

Prof. Pablo Caceres-Valencia
(U. of Puerto Rico at Mayaguez)

Prof. David Ceperley
(U. of Illinois at Urbana-Champaign)

Prof. Garnet Chan
(Cornell U)

Prof. Robert Chang
(Northwestern U)

Prof. John Clarke
(U.C. Berkeley)

Mr. Bryce Devine
(U. of Florida)

Dr. Marian Florescu
(NASA Jet Propulsion Laboratory)

Dr. Ernest Fontes
(Cornell/CHESS)

Prof. Sharon Glotzer
(U. of Michigan)

Prof. Sarah Graves
(U. of Alabama in Huntsville)

Prof. Francois Gygi
(U. of California - Davis)

Prof. Eric Jakobsson
(U. of Illinois)

Prof. David Keyes
(Columbia U)

Ms. HyunJeong Kim
(Michigan State U)

Dr. Mark Kryder
(Seagate Technology)

Prof. Vipin Kumar
(U. of Minnesota)

Prof. Ronald Larson
(U. of Michigan)

Prof. Mark Novotny
(Mississippi State U)

Dr. Derrick Mancini
(Argonne National Laboratory)

Prof. Michael McLennan
(Purdue)

Prof. Michael Naughton
(Boston College)

Prof. Padma Raghavan
(Pennsylvania State U)

Prof. Krishna Rajan
(coChair, Iowa State U)

Prof. Mark Ratner
(Northwestern U)

Prof. John Rehr
(U. of Washington)

Prof. Bruce Robinson
(U. Washington)

Prof. Anthony Rollett
(Carnegie Mellon U)

Dr. John Rumble
(Information International Associates)

Prof. Thomas Russell
(U. of Massachusetts - Amherst)

Prof. Fred Sachs
(SUNY Buffalo)

Prof. Greg Salamo
(U. of Arkansas)

Prof. Nadrian Seeman
(New York U)

Prof. Susan Sinnott
(coChair, U. Florida)

Prof. Jayanta Sircar
(Harvard U)

Prof. Michael Stopa
(Harvard U)

Dr. Changwon Suh
(Iowa State U)

Prof. Alex Szalay
(Johns Hopkins U)

Prof. Izabela Szlufarska
(U. of Wisconsin)

Prof. Henning Winter
(U. Mass, Amherst)

Ms Amy Young
(U. Illinois, Urbana-Champaign)