Materials Research—Evolving Science Dedicated to Improving Products

Daniel P. Stevens DMD, Ph.D., Eric Renner, B.S., Roberto Gomperts, Ph.D., Martin Hilgeman, Ph.D.

Table of Contents

Successful development of new materials over the centuries has enhanced the quality of life by improving products from a variety of industries such as plastics, microprocessors, glass, metal, paint, aerospace, and automobile. Initially a trial-by-error process, advances in laboratory chemistry techniques helped material research evolve into a rigorous experimental science. Today, significant advancements in software development, the ability to perform complex data analysis, and the ability to predict the properties of new materials *in silico* before any laboratory effort, have brought enormous efficiency to the materials research community. Clearly, the field of computational materials science (CMS) has become a key driver in materials innovation by facilitating effective modeling and analysis of both empirical and virtual chemical entities. Due to these advancements in understanding the physics and chemistry of materials, CMS has become one of the fastest growing areas within the field of materials research. However, for the sustained growth and success of CMS, significant resources are required, including:

1. Sophisticated and validated software, for each field of research, that supports a predictable and rational product development cycle.
2. Advanced high-performance data processing and analysis systems that provide appropriate resources for the computationally intensive modeling algorithms.
3. Interactive visualization systems that provide real-time, 3-dimensional, accurate rendering of the system(s) modeled.

While the availability of powerful computer systems has enabled the integration of computational tools into modern materials product development, the recent introduction of high-volume, lower cost, and high performance microprocessors have made the fundamental understanding of new materials commonplace.

Still, today's R&D organizations face new productivity challenges that require optimized performance to maintain a competitive

advantage. In today's global marketplace, with inexpensive personnel resources, efficient distributed manufacturing and distribution, and aggressive margin-depleting pricing, development groups are required to produce differentiated products more efficiently and more effectively. To survive in this environment, development teams must use production-oriented computational tools for materials research that allow for comprehensive product development pipelines that extend through rational design, identification, characterization, manufacturing, and testing of new materials.

Materials Research—An Intersection of Solutions, Research, and Products

Computational materials science comprises a number of research focus areas, each consisting of various theories for software solution development. Clearly, the fields that study the macroscopic properties of materials are well established. These include simulations of stiffness, flow, and viscosity, which are successfully accomplished through finite element analysis, computational fluid dynamics, and impact simulations. However, the techniques which examine a material's basic components, its electrons, atoms, molecules, and functional groups, provide a more fundamental understanding of its basic components and properties. Knowledge of these factors can lead to a state of predictable and rational design and ultimately translate this knowledge into building a superior final product faster.

Categories of Molecular Computational Solutions

Simulations of these basic molecular components *in silico* is achieved through a number of techniques, each having unique capabilities to address a research question. The major focus areas can be summarized by the following categories:

Quantum Mechanics – Quantum Mechanics (QM) methods estimate molecular properties from the interactions of electrons and, because they are based on a fundamental equation of quantum chemistry, they are much more accurate than the models developed by Molecular Dynamics techniques. Additionally, QM methods are required if the processes under study involve the breaking and forming of chemical bonds. The breakage and creation of chemical bonds is a crucial step in the formation of every chemical compound. Due to the extremely complex analysis used in QM, the main disadvantage of this technique is that long computation times can be required to model a simple system. Significant computational resources should be accessible for meaningful calculations and the size of the system that can be studied has typically been limited by this requirement.

Molecular Dynamics – Molecular Dynamics (MD) calculates the motions of atoms or molecules and, in these simulations, atoms and molecules are seen as compact entities without any substructure. Therefore, unlike in QM simulations, electrons are not "visible" or analyzed in MD. Based on these analyses, a variety of molecular attributes can be estimated. For MD computations, larger groups of atoms can be modeled compared to QM, since it works at a lower level of complexity.

Recently, the advent of technologies have either mixed the QM and MD theory levels or expanded upon them. For instance, there are now Quantum Mechanics/Molecular Mechanics (QM/MM) techniques that provide the benefits of both techniques by treating certain small but interesting regions of a model at the QM level and other regions at the MM level. The Periodic Boundary Conditions (PBC) technique permits the modeling of extensive material regions that are based on the regular repetition of units. Historically, these techniques were only available with MM methods but are now also possible with QM approaches. There are various software solutions on the market today which combine QM and MD theory to study the properties of large molecules.

Mesoscale Dynamics – A broad term, mesoscale dynamics refers to the simulation of more than a few thousand atoms but less than billions. In mesoscale calculations, subunits that are larger than atoms, such as each monomer in a polymer, are used as the basis of calculation. Mesoscale dynamics simulations often are used to study chemical reactions, kinetics, interface dynamics, and crystal structures for solid materials, fluids, proteins and gases. Because they use larger subunits than the molecular models, simulations can be developed over longer time intervals thus allowing for the study of complex materials including liquids and polymers on the nanometer to micron scale.

Fields of Materials Research

Appropriate application of these technologies into R&D efforts can dramatically improve product design efforts. When applied to the following areas of research, significant improvements to both product quality and development productivity can be achieved.

- Optimization of chemical reactions and production
- Development of crystalline structures from organic molecules
- Analysis of solid materials (e.g., crystallography)
- Understanding of the structure of plastic polymers
- Understanding of suspensions of solids in liquids (colloidal systems)
- Design of new solid materials and understanding of surface chemistry
- Solid state and surface chemistry
- Catalysis

SGI Solution Leads to Productive and Reliable Computational Materials Research

To build an infrastructure that effectively supports the rational development of new materials, it is critical that a flexible and enabling high-performance computational platform be available. Beyond simply increasing the speed of data analysis, the infrastructure must offer scientists the ability to not only perform the tasks fast and accurately, but also allow them to ask new questions of the data and, ultimately, bring their research forward into new previously unimaginable areas. Ultimately, the type of questions researchers need to ask should not be limited by the infrastructure they depend upon. Today's reality is that computing solutions are typically restricted to systems based on high-volume microprocessors combined with the now-standard open source Linux operating system. Within this area, there are numerous choices and a careful assessment of various parameters should be considered before making a computational system decision. These include:

- The size and nature of the problem
- The number of users working on the problem or on the research team
- Size of data and how data is accessed (e.g. disk access vs. resident in memory)
- Size of primary, intermediate, and final data (scratch data is often much larger than primary and final data)
- Available data management solutions for the platform
- Type of computational power needed to provide useful turnaround of workload
  - Type of math needed in calculations (e.g. integer vs. floating point)
  - Application performance on 32-bit vs. 64-bit environment
  - Speed of connection between components (e.g., processors, memory, visualization, other peripherals)
- Requirement for and availability of scalable visualization for the platform
- Development environment to support custom algorithms where needed

Based on the Intel® Itanium® 2 processor, the SGI® Altix® family of servers offers some key advantages for running computational materials research applications. It satisfies the requirements for a standard microprocessor-based system running Linux while offering unique capabilities to support state-of-the-art R&D. Most of the popular applications used for the various materials research fields are ported to the Altix and many have been optimized for superior performance on the system.

The Intel Itanium 2 processor is based on the Explicitly Parallel Instruction Computing or EPIC architecture. Since the instruction set is explicitly parallel, the architecture attains high levels of parallelism inside the processor (so called "instruction level parallelism"). Leveraging the EPIC architecture of the

SGI Altix family of servers

Intel® Itanium® 2 processors, the Altix adds parallel system architecture, a massive set of internal resources, and high bandwidth connection to I/O and memory, delivering industry leading performance to scientific computing applications. This means more jobs can be run and more users serviced in less time. And with the large shared address space provided by the Altix NUMA architecture, programming is dramatically simplified.

In addition to the raw microprocessor performance offered by the Intel Itanium 2 processor the 64-bit Altix offers scientists virtually unlimited memory capability in a tightly coupled system. The ultra-high speed NUMAlink™ communication environment connects the Altix system's microprocessors, memory, visualization and other peripherals and leads the industry in performance. With this capability, the Altix shared-memory platform is able to perform complex simulations with turnaround times that accelerate, rather than stall, the R&D process. This is in stark contrast to the barriers presented by other cluster platforms in terms of job turnaround, ease of programming, and versatility.

"With the chemistry applications and some of the CFD codes developed by our researchers, we had jobs that required us to hold the entire data set in memory. Faculty were reporting that many of these jobs required more memory than the cluster could provide, and they simply wouldn't finish at all on our older systems. To ensure that we could address the needs of our user base for the foreseeable future, we really needed to push our environment to the next step, and we quickly concluded that the [Intel Itanium 2-based] Altix was the best solution for us."

David Roach, Director of the Mississippi Center for Supercomputing Research (MCSR).

"The SGI Altix is very well suited for electronic structure calculations, because it has not only a fast [Intel Itanium 2] processor, but also very good inter processor communication. An additional advantage is the high memory capacity of the system, which allowed me to complete two research projects that were previously too big for our systems to handle."

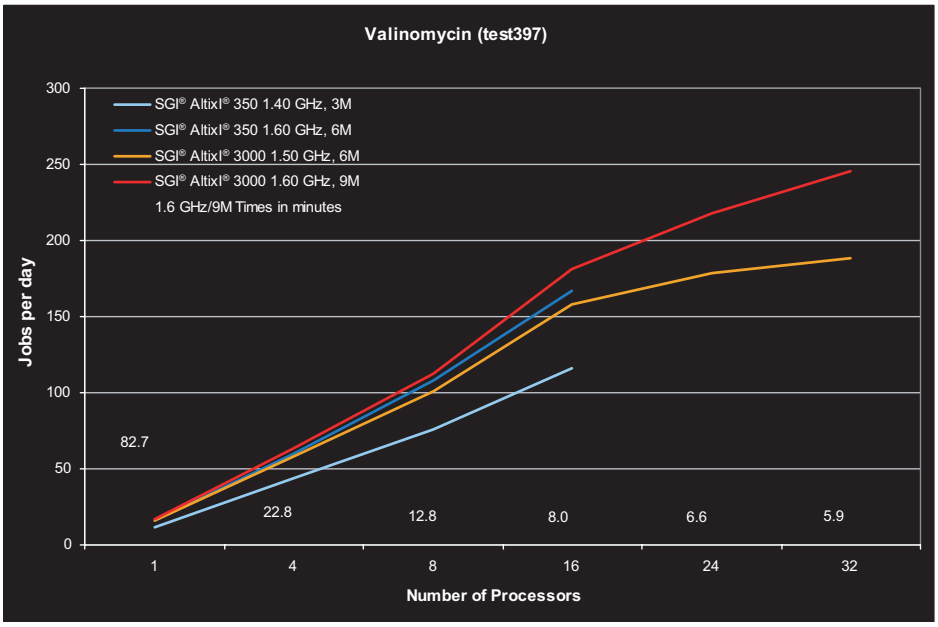
Dario Alfè, Reader in Physics and Royal Society University Research Fellow at the Department of Earth Sciences and Department of Physics and Astronomy, University College London.

SGI provides comprehensive solutions for materials research with the Altix family of servers, the SGI® InfiniteStorage data management systems, and the new Silicon Graphics Prism™ visualization platform. In addition to building the technical infrastructure, SGI engineers work closely with the materials research software developers so that the key applications are optimized for performance and scalability on SGI's Altix, Prism, and InfiniteStorage systems. Since the list of materials research applications is extensive and constantly evolving, performance for a number of key applications will be reported here. An updated list of Altix ports can be found at: *http://www.sgi.com/industries/sciences/chembio/comp_chem.html*

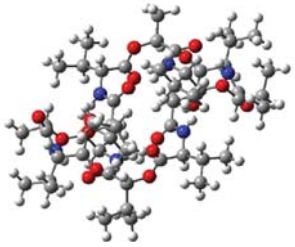
In the field of Quantum Mechanics, SGI Altix provides the required large memory, significant processing power, and high-speed connections to all system components. Beyond this, SGI offers a complete High Performance Computing solution for data management and visualization. SGI also offers highly tuned mathematical and parallel software libraries to achieve outstanding numerical and scalable performance for applications on the Altix platform. SGI software engineering specialists constantly work with application providers to assure the algorithms are tuned for optimal performance on the Altix platform.

A few key quantum chemistry applications include, but are not limited to, Gaussian, GAMESS, NWChem, VASP, Wien2K, ADF, CPMD, and CASTEP. Representative of this group of applications, their performance is demonstrated in the following graphs.

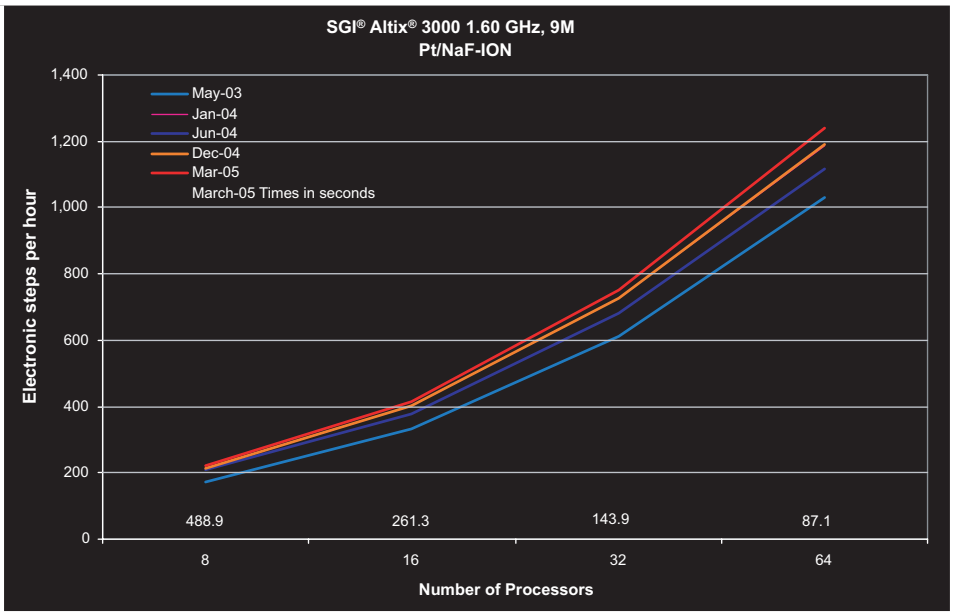
Gaussian 03 – Performance on Intel® Itanium® 2-based SGI Altix 3000 and SGI Altix 350



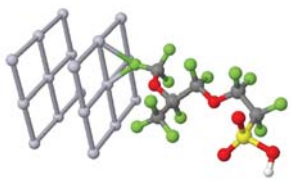
Valinomycine: $C_{54}H_{90}N_6O_{18}$
C1 symmetry
882 basis functions
RB3LYP/3-21g
Force calculation



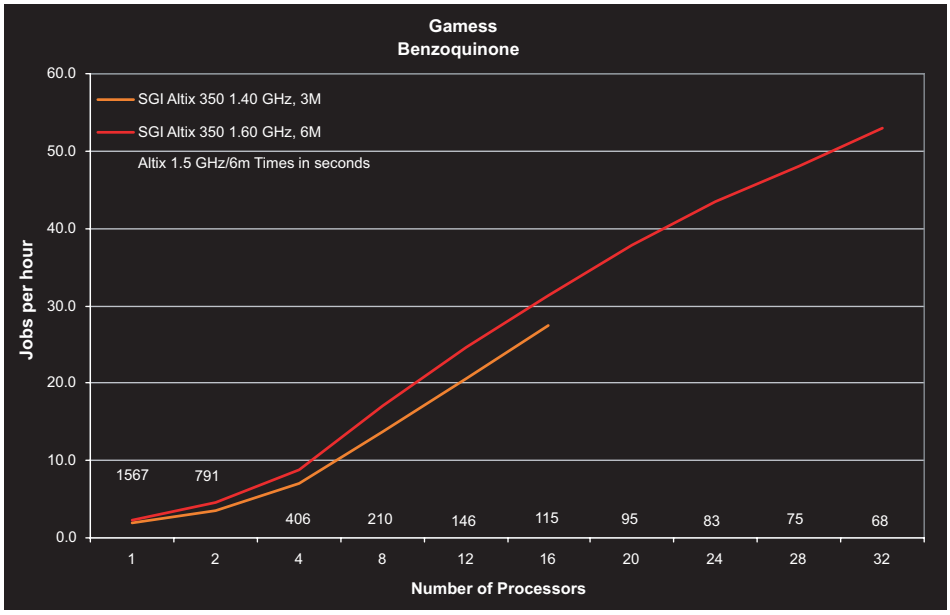
VASP – Performance on Intel® Itanium® 2-based SGI Altix 3000



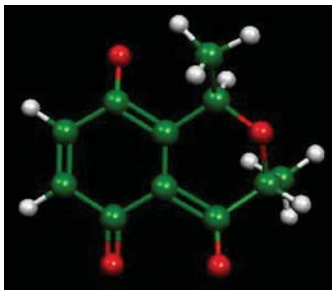
Pt/NAFION
44 ions system,
199 bands,
ionic relaxation,
3 ionic steps,
10 electronic steps,
1 k-point, 612360 plane waves,
convergence window 1E-03 eV,
3 x 3 x 2 cell with 6 layers



GAMESS – Performance on Intel® Itanium® 2-based SGI Altix 3000

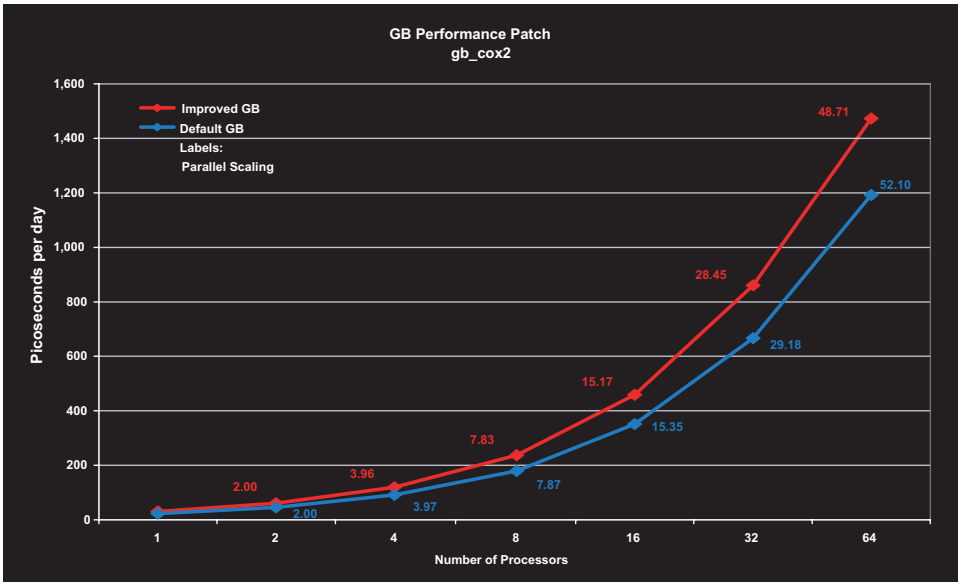


Benzoquinone: $C_{11}H_{10}O_4$
In-core
MP2 gradients
6-31g(d)
245 basis functions



Molecular Dynamics applications are used to model larger molecules. To achieve maximum performance and productivity, SGI engineers work with software developers to optimize the key applications used to support this research. Some of the popular molecular dynamics applications include Amber and CHARMm. Their Altix performance, as demonstrated in the following graphs, is representative of the overall performance one can expect from many of the molecular dynamics applications on this server platform.

Amber 8 – Performance on Intel® Itanium® 2-based SGI Altix



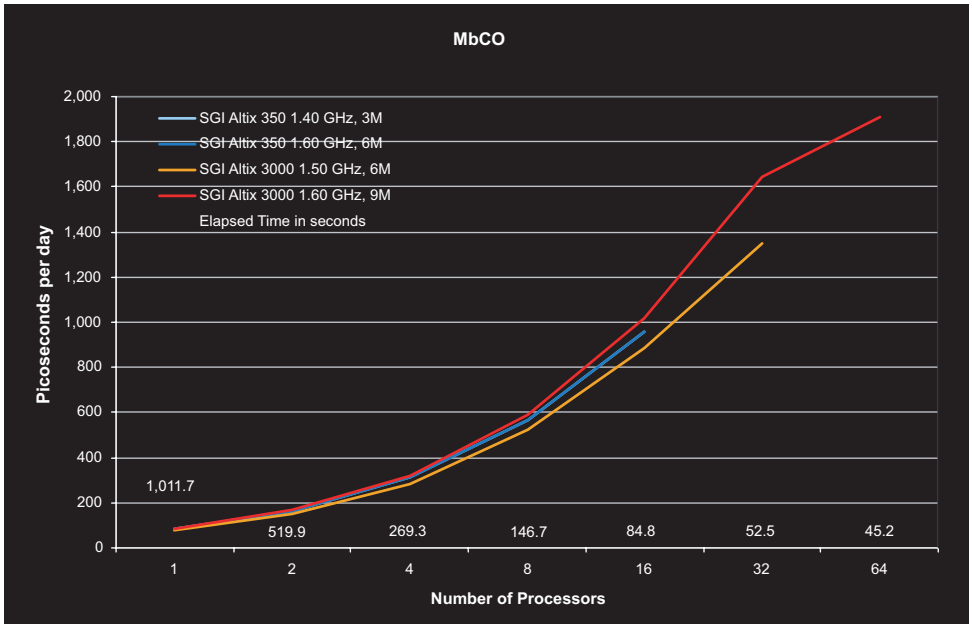
Gb_cox2 (Generalized Born Approximation):

Thermalization
Number of atoms: 18,056
Generalized Born simulation with cutoff of 12 Ang., salt concentration 0.2 M, nrespa=4

Test performed on SGI Altix 3700 Bx2, 1.6GHz/9M Intel Itanium 2 processors

See: http://amber.scripps.edu/

CHARMm c30b2 – Performance on Intel® Itanium® 2-based SGI Altix



Myoglobin:
Protein DHFR, solvated with TIP3 water, in a periodic box. There are 23,558 total atoms, and PME used with a direct space cutoff of 9 Ang.

Materials Research Case Study—SGI Solutions Required to Get Results

The use of computational chemistry to understand the fundamental nature of materials can contribute to the efficient development of new products and the understanding of properties of the natural environment. To illustrate how SGI hardware and engineering solutions can obtain results for large computational chemistry research interests, a case study was developed by Dario Alfè, Reader in Physics and Royal Society University Research Fellow at University College London.

Background: Knowledge about the inner structure of the Earth is of fundamental importance to understanding all the living geological processes, like plate tectonics, volcanism, and earthquakes. The compound $MgSiO_3$ is one of the components that is present in the Earth's lower mantle. It is believed that the relevance of the melting temperature of $MgSiO_3$ at the pressures of the mantle is related to our knowledge of the formation of the Earth. Additionally, at the bottom of the mantle there is a so-called 'ultra-low velocity zone' (ULVZ) which has anomalous seismic velocities that could be due to the presence of partially melted material.

Advances in first principles electronic structure techniques and the availability of suitable computational resources have made it possible to study the thermodynamic and elastic properties of such systems in detail. The result is that the physical properties of the elements contained in the Earth's core can be studied using first principles methods and *ab initio* computational quantum chemistry software.

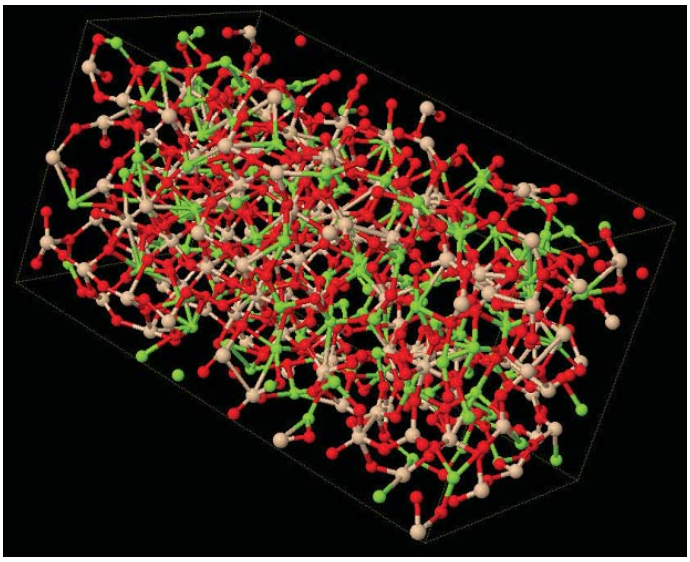
The Problem: To find the presence of partially melted material using *ab initio* quantum mechanical molecular dynamics methods, one has to simulate a system of a sufficient number of ions over a period that is long enough, so that the equilibrium temperature between solid and liquid is preserved. In terms of a MD simulation, that means that at least 10-20 thousand steps must be performed to give a point on the melting curve of the compound. Further, the accuracy of a MD calculation is greatly determined by both the amount of ionic steps that controls the movement of the ions and the number of electronic self-consistency (SC) iterations needed for the total energy of the system to be between certain error criteria. For this case study, a number of 10,000 ionic steps and an energy cutoff of 10^{-5} eV for the electronic SC phase were required.

The Solution: For this project, VASP, a leading software package for performing *ab initio* quantum mechanical molecular dynamics (MD) calculations using pseudo potentials and a plane wave basis set, was used to perform a MD simulation on a system of 720 atoms of $MgSiO_3$. Initial measurements revealed that one ionic step would take approximately 21 minutes on an SGI Altix 3300 system with eight Intel® Itanium® 2 1.3 GHz processors, which means it would take 150 days to complete 10,000 steps. While the most obvious method to reduce the time to solution would be to add more processors to the system, this solution assumes continuous performance improves with each resource adding. However, with VASP, like many other *ab initio* computational chemistry applications, scalability to more than 64 processors can be limited due to requirements like:

1. Data redistribution at each electronic step: after each step, the microprocessors send distinct results to the other microprocessors, where each message can have a different size and displacement. This kind of message passing, where each processor has to simultaneously send and receive data from other processors, greatly benefits from a high bandwidth and low latency parallel infrastructure. For this to be achieved efficiently, a system with a large shared memory core, such as that offered by the SGI Altix, is required. Without a single large memory core and the high bandwidth and low latency it provides between system components, this stepwise data distribution requirement could quickly overburden the computing system.
2. At each electronic step, a serial matrix diagonalization is performed. This requirement becomes a factor of importance for MD calculations with hundreds of ions, and impacts scalability when more processors are added to the system.

To allow researchers to take full advantage of the performance resources offered by the Altix and to provide ultimate performance scalability, SGI applications engineers have refined the data distribution routine in VASP so that it takes full advantage of the shared memory and low latency infrastructure offered by Altix. Additionally, the ScaLAPACK routines of the SGI SCSL scientific library have been used to perform the time-consuming matrix diagonalization steps in parallel. The net result of this work is that one ionic step takes only 75 seconds on a 256-processor, 1.3 GHz Intel Itanium 2-based Altix 3700 system. This translates to the reduction in time to solution from 150 days on an eight-processor system to 9 days on a 256-processor system. This performance result could be achieved only by the technologies offered by the Altix server coordinated with the application tuning performed by the SGI applications engineering team. Further, to demonstrate how the constantly evolving Altix technology can bring more value to research scientists, additional measurements have shown that the latest SGI Altix 3700 Bx2 server with the higher bandwidth NUMAlink™ 4 interconnect and 1.6 GHz Intel Itanium 2 processors would double the performance of the aforementioned 256-processor Altix 3700 system. Thus, using a current SGI Altix, one would see results in approximately 4.5 days, or half the time!

Figure 1. A structural representation of the 720 ion $MgSiO_3$ system



Lessons Learned

Improved techniques in materials research provide tangible benefits to today's consumers. Successful materials investigation relies on having access to the tools needed to achieve analysis in numerous scientific computational areas.

- Rational material design is a key component in understanding their basic properties and is critical to their development in a predictable manner. QM, MD, and mesoscale dynamics are scientific fields that are used in the rational design of new materials.
- The combination of the Intel® Itanium® 2 processor and the shared-memory NUMAlink architecture of the Altix offers a uniquely parallel architecture for computationally demanding materials research applications.
- Working with the developers of materials research software, SGI engineers develop application ports that are optimized for both performance and scalability on the Altix family of servers.
- The SGI Altix family of servers with Intel Itanium 2 processors provides materials research organizations a computing solution that delivers a clear roadmap for cost-effective productivity and growth. Together with the applications ported and tuned to this architecture, the Altix provides the high performance computational resources that allow scientists to focus on their research, instead of having to focus on the research tools

3834 [10.2005] J14976