Title: Scalable Quantum Mechanical Simulation of Large Polymer Systems

Author(s): S. Goedecker
A. Hoisie
J. Kress
O. Lubeck
H. Wasserman

Submitted to: Supercomputer '97
Mannheim, Germany
June 19-21, 1997

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Los Alamos National Laboratory

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. The Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Form 836 (10/96)
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
Scalable Quantum Mechanical Simulation of Large Polymer Systems

S. Goedecker,
Max-Planck Institute for Solid State Research, Stuttgart, Germany
A. Hoisie(1), J. Kress(2), O. Lubeck(1), H. Wasserman(1)
Los Alamos National Laboratory, USA
(1) Computer, Information and Communications Division
(2) Theoretical Division

Abstract

We describe a program for quantum mechanical calculations of very large hydrocarbon polymer systems. It is based on a new algorithmic approach to the quantum mechanical tight binding equations that naturally leads to a very efficient parallel implementation and that scales linearly with respect to the number of atoms. We get both very high single node performance as well as a significant parallel speedup on the SGI Origin 2000 parallel computer.

Introduction

The understanding of molecular systems and bulk materials on an atomistic level is one of the fundamental tasks of contemporary physics and chemistry. The role of computational methods in this challenging task is steadily increasing as result of both the rapid progress in computer performance and the algorithmic advances in the field.

The equations governing the behavior of matter are well known. They are the equations of quantum mechanics which were found at the beginning of this century by Heisenberg and Schrödinger. In principle, all questions in chemistry and materials science are therefore problems in computational science. If it were possible to solve these equations for any molecular system, all possible properties of this system such as mechanical, electronic and optical properties would be accessible. In practice however, it is not possible to solve the many-electron equations of quantum mechanics exactly.

Three classes of methods are available today that allow modeling at the atomic scale. These methods approximate the quantum mechanical equations with different levels of accuracy. Not surprisingly, less accurate methods are less demanding in their computational requirements. The least accurate method is the classical force field. It basically ignores the quantum mechanical origin of the forces acting among the atoms. These forces
are replaced with classical potentials that were fitted to reproduce the geometries of a set of molecules used in this fitting procedure. Therefore, by definition, classical force fields can not describe electronic and optical properties. Even the geometric and mechanical properties are frequently wrong, as there are instances when purely quantum mechanical effects have a strong influence on the geometry.

At the other end of the spectrum there are the very accurate density functional and quantum chemistry methods. Unfortunately they are computationally so expensive that only systems with at most a few hundred atoms can be treated.

Tight Binding (TB) methods fall in between. They are based on quantum mechanics and can therefore capture many of the features that are inaccessible to force field methods. At the same time they are much less costly computationally compared with the above mentioned methods. For certain classes of materials, such as hydrocarbons, very reliable TB parameters have recently been developed at Oxford University\textsuperscript{1}, which give accuracies that are very close to those obtained from density functional calculations. In contrast to previous TB methods, this new scheme incorporates self-consistency corrections, which probably explains its higher accuracy.

However, using a tight binding scheme alone does not yet allow one to simulate very large atomic systems. Traditional methods solving for quantum mechanical equations scale as the cube of the number of atoms. Therefore, in this framework, it is not possible to treat systems containing more than 500 atoms, even when using tight binding schemes.

So-called O(N) algorithms that scale linearly with respect to the number of atoms have recently been developed to overcome these limitations. In this work we use a projection O(N) algorithm\textsuperscript{2} that has the very important additional property of being intrinsically parallel. This means that this algorithm allows us to split up the big computational problem into loosely-coupled subproblems based on our understanding of molecular systems at the physical level. This approach ensures extremely good scalability since the communication requirements are very modest. This program will be used in the framework of the Advanced Strategic Computing Initiative on Los Alamos National Laboratory’s ASCI Blue Mountain SGI supercomputer to solve problems in polymer science. At the same time the code will be made available to other researchers in electronic structure and molecular dynamics. On machines of Teraflop capabilities, we anticipate doing both static and dynamic (i.e. molecular dynamics) quantum mechanical calculations for systems containing up to 100,000 atoms. These systems are then two orders of magnitude larger in size than what was previously possible.

**Brief Description of the Algorithm**

The basic quantity in this algorithm is the finite temperature density matrix $F$. From this matrix, all quantities of interest such as the total energy and the atomic forces\textsuperscript{3} can be obtained. The matrix $F$ is calculated as a Chebychev series

$$F = \sum_j T_j(H)$$

where $T_j$ is a Chebychev matrix polynomial and $H$ the Hamiltonian matrix of the system. For large atomic systems most of the elements of $F$ are negligible and $F$ is thus effectively
a banded matrix. This means that the total number of matrix elements that need to be calculated is proportional to the dimension of the matrix. Since the dimension of the matrix is proportional to the size of the system, the implication is that the computational effort scales linearly with the system size. Each row of the matrix $F$ is called a localized orbital and it is essentially an atomic orbital that is modified by the scattering properties of the neighboring atoms.

The parallelization scheme follows naturally. Each processor calculates one or a few localized orbitals. To do this it only needs to know the atomic positions of the atoms contained in the localization region. This localization region is the volume of space where the amplitude of the localized orbital is sufficiently large and it is much smaller than the whole molecular volume. Each processor then calculates that part of the Hamiltonian matrix that describes its localization region and then from Eq. (1) its part of the matrix $F$. The calculation of the Hamiltonian matrix elements is computationally negligible. The most costly operation is the Chebychev matrix recursion which involves sparse matrix-vector multiplications between the matrix representing the local Hamiltonian and the vector of the orbital that is being built up. Another time consuming step is to sum all the Chebychev matrices of different order. These two basic operations take more than 90 percent of the CPU time. In the communication steps, we have to send out the atomic positions to all processors and to receive all the individual forces that are needed to update the positions in a molecular dynamics simulation. Another communication step derives from the need to find a global potential that ensures the local charge neutrality.

Optimization and Performance of the Main Kernels

An important concern in the design of this program was data locality to minimize interprocessor memory traffic and to have optimal cache reuse on a single node. The dramatic decrease in performance of the R10000 processor on the Origin 2000 machine once the data sets are larger than the level-2 cache (4Mbyte) is shown in Fig.1.

Several arrays are used in the main kernels and all these arrays have variable length and shape during a molecular dynamics simulation. Therefore pointers were used to pack all relevant data in a dense data structure which fits into the level-2 cache. The pointers are recalculated in each molecular dynamics step taking into account the changes in data structure due to the changes in the atomic positions. The pointer structure was however hidden to the compiler in the basic kernel routines to avoid aliasing problems that might inhibit optimization. The performance of the sparse matrix-vector multiply depends strongly on the physical context. In the case of carbon-carbon interactions the corresponding sparse matrix-vector multiplication was organized into block-form to minimize indirect indexing. Each symmetric 4-by-4 block contains the Hamiltonian matrix elements of the 4 carbon electrons located on different atoms. This resulted in a performance of 150 Mflops on this kernel, which compares favorably with the vendor optimized dense symmetric matrix-vector multiplication DSYMV running at 200 Mflops. In the case of hydrogen-hydrogen interactions there is only one electron per atomic site, therefore these optimizations are not possible and the performance is reduced to approximately 13 Mflops. The mixed carbon-hydrogen interactions fall in-between with a performance
of 35 MFlops. Fortunately the number of operations for these slow kernels is much lower than for the fast part. The hydrogen-hydrogen component accounts for only 9 percent of the CPU time compared with 30 percent for the carbon-carbon part and 13 percent for the mixed interaction kernel. The summation of the different Chebychev matrices takes another 30 percent of the CPU time. It has a structure similar to a DAXPY and we get a performance of 70 MFlops, which is actually somewhat higher than the performance of the vendor supplied DAXPY for the relevant loop lengths.

Optimization and Performance of the Communication Part

The intrinsically parallel structure of the algorithm results in a fairly small number of communication steps. For computations on up to 50 processors done on the ASCI Origin 2000 machine in Los Alamos, not more than 1 percent of the whole CPU time went into communication. Since the messages are very short, latency rather than bandwidth is the limiting factor. The fact that communication is not the critical component allowed us to use the message passing paradigm with the standard MPI library. The program is thus portable to nearly any other parallel computer, as well as to clusters of workstations. A predecessor of this program won the Gordon Bell Award in 1994, in the price/performance category, for its excellent performance on a loosely-coupled cluster of workstations.

Parallel Performance: Results and Timings

The ASCI machine at Los Alamos National Laboratory will have more than 4,000 processors in its final full configuration. At the present time, it is under construction and we were not able to do test runs on more than 50 processors. Since we felt that demonstrating weak scalability (i.e., proportionally increasing both the problem size and the number
of processors) on this very small number of processors would not be very convincing, we choose to demonstrate the more difficult strong scalability (i.e., constant problem size solved on increasing number of processors).

The results of these test runs are shown in Fig. 2. The physical system used for this test run was the polymer $C_{1536}H_{3072}$. The size of this polymer was chosen such that, from a memory requirement point of view, the sequential computation was still feasible on a single node. Letting the system evolve for 45 steps took roughly 15 minutes in serial mode, and about 20 seconds on the 50 processor configuration.

In a realistic physical system it is practically impossible to have a perfect load balancing. Since a polymer has a nonregular structure, not all of the localization regions have the same size. For instance the localization region for an atom at the end of a polymer chain is just half as big as the localization region of an atom in the middle of the chain. But even for atoms farther from the end of the chain, the localization regions can vary in size due to the thermal motion of the chain. Load imbalance arises because the number of operations necessary to calculate a localized orbital is proportional to the size of the localization region. In addition, the number of carbon and hydrogen atoms assigned to each processor is not exactly equal unless the total number of both carbon and hydrogen atoms is a multiple of the number of processors, which is not possible in most cases. These two factors lead to a load imbalance which causes deviation from linear speedup. In the case of the 50 processor run, the load imbalance was of the order of 10 percent, which correlates well with the observed deviation from perfect speedup. As was already mentioned, the communication part was negligible even for the 50 processor run, it was only one percent.

Given the fact that for larger polymer simulations the load balancing will be better, we are therefore very confident that the envisioned 100,000 atom simulations will be possible once the machine is fully configured.

Figure 2: Strong scalability of the code on the SGI Origin 2000 system
Conclusions

We have developed a high performance Tight Binding program using a novel intrinsically parallel algorithm that scales linearly with respect to the number of atoms in the system. For up to 50 nodes we obtained very good speedup for a polymer containing 4610 atoms. The performance analysis presented indicates an even better parallel performance for much larger polymer simulations, such as those planned for the fully configured ASCI machine at the Los Alamos National Laboratory.

Acknowledgements

Large parts of this program were developed by S.G. during a stay at Los Alamos National Laboratory. S.G. thanks LANL for its hospitality. Interesting discussions with Art Voter are acknowledged. We thank A. Horsfield for providing us with a detailed data base of the new Tight Binding scheme before its publication.

Special thanks to Mark Dalton with SGI in Los Alamos for his support during the benchmarks.

The authors wish to acknowledge the Advanced Computing Laboratory of Los Alamos National Laboratory, Los Alamos, NM 87545. This work was performed on computing resources located at this facility.

This work was performed under the auspices of the U.S. Dept of Energy by Los Alamos National Laboratory under Contract W-7405-ENG-36.

References