CLASSICAL MOLECULAR SIMULATIONS OF COMPLEX, INDUSTRIALLY-IMPORTANT SYSTEMS ON THE INTEL PARAGON

H. D. Cochran and P. F. LoCascio
Chemical Technology Division
Oak Ridge National Laboratory*
Oak Ridge, TN 37831-6224

P. T. Cummings
Department of Chemical Engineering
University of Tennessee
Knoxville, TN 37996-2200

Presented and published:
Intel Supercomputer User Group
(ISUG '96), Knoxville, Tennessee
June 19-22, 1996

Submitted to:
J. Comp. & Math
June, 1996

The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-96OR22464. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

*Managed by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy under contract number DE-AC05-96OR22464.
Classical Molecular Simulations of Complex, Industrially-Important Systems on the Intel Paragon

H. D. Cochran and P. F. LoCascio†
Chemical Technology Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6224

P. T. Cummings
Chemical Engineering Department
University of Tennessee
Knoxville, TN 37996-2200

Advances in parallel supercomputing now make possible molecular-based engineering and science calculations that will soon revolutionize many technologies, such as those involving polymers and those involving aqueous electrolytes. We have developed a suite of message-passing codes for classical molecular simulation of such complex fluids and amorphous materials and have completed a number of demonstration calculations of problems of scientific and technological importance with each. In this overview paper we will outline the techniques for classical molecular simulation of these industrially-important systems on the Intel Paragon and we will summarize some of the important scientific and technical results of the varied applications, including the following: 1) Parallel codes for quaternion dynamics using techniques for handling long-range Coulombic forces allow study of ion pairing in supercritical aqueous electrolyte solutions.1–3 Ion pairing lies at the heart of technological problems with corrosion and solids deposition in industrial processes utilizing high temperature water. 2) Non-equilibrium, multiple time step molecular dynamics lets us investigate the rheology of molecular fluids.4–7 Such calculations enable the molecular-based design of new synthetic lubricants of importance in the automotive engines of the future. 3) Chain molecule Monte Carlo simulations in the Gibbs ensemble8–10 now permit calculation of phase equilibrium of long-chain molecular systems. With complementary equilibrium molecular dynamics (with multiple time steps) we have been able to gain fundamental insight into the technologically-important problem of liquid-liquid phase separation in polymer blends.11

Computational Techniques

Figure 1 is a cartoon showing a trajectory through the phase space. The abscissa represents the 3N-dimensional position coordinates of the N particles (say, atoms) in a system. The ordinate represents the 3N-dimensional momentum coordinates of the particles. The trajectory, then, traces the evolution with time of the positions and momenta under the influence of the total potential energy of the system which consists of the potential energy of each particle due to its position in any external fields plus the potential energy of interaction of the particles with one another, including both bonded and non-bonded interactions. Classical molecular dynamics simulation solves Newton’s equations of motion for such a system for a small portion of the trajectory. Application of the laws of statistical mechanics, then, allows calculation of the bulk properties of the substance from the positions and momenta of the particles. An alternative approach to classical molecular simulation samples randomly (rather than deterministically) from the points in the configuration space of the system according to their probability distribution; this technique is called Monte Carlo simulation. The general techniques of classical molecular simulation on sequential computers are described by Allen and Tildesley.12 Usually, to make the computation tractable, a number of approximations are employed; interactions among particles are treated pairwise, the range of such interactions is truncated, and small systems O(102 – 105 particles) are surrounded on all sides by periodic images of themselves to eliminate surface effects. The great bulk of the computational work is in calculating the pairwise interactions. The use of neighbor lists and/or link cells can reduce this to an O(N) problem. Even with these simplifications, classical molecular simulations are challenging computations, and until the advent of massively parallel supercomputers such as the Intel Paragons, calculations such as those described here, on many industrially-important systems, were beyond the horizon of practicality.

Figure 2 shows two approaches to parallel molecular simulation calculations (for clarity of presentation a 2-dimensional system is shown). Figure 2a shows a domain decomposition approach in which the area (volume in 3-D) containing the sample of particles constituting the simulated system is divided into subareas (subvolumes). In a molecular simulation performed using domain decomposition, the particles within each subdomain (and their positions and momenta) are assigned to a separate processor which calculates the interactions among particles within its subdomain. Then, movement of particles between subdomains and interactions which extend into neighboring subdomains are communicated each timestep. For homogeneous systems (single phase in the absence of external fields), this domain decomposition is intrinsically load-balanced. Pinches et al.,13 presented a domain decomposition code for classical molecular dynamics of simple atomic systems that showed good scaling. As might be expected, domain decomposition approaches permit the simulation of very large systems, spreading the memory required for the positions and momenta of the many particles over the local memory of a number of processors. However, the computation of the pairwise interactions of the particles in the subdomain of each processor is performed only at the speed of a desktop workstation; so, large but not long long simulations are possible using domain decomposition.

† currently Center for Computational Sciences, ORNL.
Figure 2b shows a replicated data approach to molecular simulation. In this approach identical representations of the entire system are replicated in the local memory of a number of processors. Then, the computational work of evaluating the pairwise interactions is divided among the processors and the local results are summed. The interaction is evaluated within a double loop that compares uniquely every particle with every other particle. The computational work is divided up among processors by assigning a portion of the range of the outer loop to each processor. These calculations can be done in parallel, and the simplicity of the division permits some straightforward load balancing to take place. Improved load balancing divides the work for a set of P processors, as if we had 2P available nodes. One then assigns iterations from opposite ends of the outer loop, in effect, giving each processor a large and a small portion to work on. Clearly, this replicated data approach allows small systems (so that all data fit in local memory) to be simulated for long times. The latency of global communication (in summing the local contributions to the total interaction) limits the ultimate scaling of the replicated data approach. Current implementations on the Paragon limit the scaling to the range 10x to 100x.

Figure 3 illustrates the complementary advantages and disadvantages of the two approaches. Unfortunately, for real physical systems, larger system size requires simulation for longer times. For example, one part of a system can communicate its influence to other parts of the system no faster than the speed of sound so large systems must be simulated for long enough time for sound to propagate from one extremity to another, else the parts could be simulated independently. In fact, most large systems require much longer simulations than this to encompass the range of timescales intrinsic to the problem. For example, in a large protein molecule the time constant of the interatomic interactions (which ultimately determine all of the dynamics) is of order femtoseconds while the time constant for folding of the entire protein molecule (which determines its biological function) may be of order seconds or longer. Substantial improvements in both parallel supercomputer performance and molecular simulation programs will be required before the dynamics of protein folding can be simulated atomistically.

Because calculations of the pairwise interaction constitute the bulk of the computational cost in all cases, we have invested in making that part of the calculation as efficient as possible. The first step was to rewrite the source code so that it vectorizes to the greatest extent possible and so that the cache is utilized as effectively as possible. Then, assembler routines for the kernel calculations were written. These steps have achieved speed-up ranging from 4x to 10x on our various calculations. We have managed to achieve 60 MFLOPS on each node (30 MFLOPS/processor) in some of these kernel calculations.

Another massively parallel supercomputers utilize microprocessors for which far better compilers have been written than are available for the i860; so, such specialized coding might not be needed and might not yield such significant performance improvements.

**Industrially-Important Applications**

We describe briefly some of the significant scientific and technological results from classical molecular simulations of industrially-important systems on the Intel Paragons. These applications are described more fully elsewhere. Additional computational techniques have been required to meet the demands of these specific applications; these more specialized techniques are described briefly with the appropriate application.

**Ion Pairing**—Ion pairing lies at the heart of technological problems with corrosion and solids deposition in industrial processes utilizing high temperature water; because of severe corrosion problems, experimental study has not been possible for many important cases. In our molecular simulations, water has been modeled as a small rigid rotator with imbedded charges. We have employed quaternion coordinates to avoid the difficulty of trigonometric functions (which go to zero at some angle) in the denominator of the usual equations of Newtonian dynamics. The long range of Coulombic forces requires special treatment. We have employed two approaches. The more efficient technique involves truncating the forces at some distance and assuming a uniform dielectric beyond the cutoff (the reaction field technique). A more rigorous but more costly technique uses Ewald summation for the forces beyond the cutoff. We have developed both replicated data and domain decomposition codes for such simulations. The domain decomposition approach has allowed study of the very large systems which are required with dilute electrolytes.

Figure 4 shows the probability of finding an oxygen atom at a given distance from another oxygen atom normalized to the bulk average probability (the intermolecular oxygen-oxygen pair correlation function) from a simulation of 42,592 water molecules at a supercritical state. Also shown are intermolecular hydrogen-hydrogen and oxygen-hydrogen pair correlation functions. The very good statistics from such a large simulation yield smooth curves that can be precisely compared with the results of neutron scattering experiments. Both our simulations and the experiments of a European group have been improved through efforts to quantify the degree of hydrogen-bonding in supercritical water which can be inferred from such results. Figure 5 shows the pair correlation function between sodium ions and chloride ions in a dilute, high-temperature aqueous solution. Because the solution is dilute, the results show considerable statistical uncertainty even in the largest aqueous electrolyte simulation ever performed. The sodium chloride dissociation constant, calculated from these results, is in good agreement with extrapolations form experiments performed at less severe conditions.
Rheology of Alkanes—The variation of the viscosity of lubricants with temperature and strain rate (rheology) is a technologically important problem for which existing scientific understanding cannot give needed quantitative predictions. To study rheology of lubricant molecules (alkanes) by molecular simulation we have used the techniques of non-equilibrium molecular dynamics with both replicated data and domain decomposition codes. Figure 6 shows the two approaches (deforming-cell and sliding-brick) we have employed to handle the periodic boundary conditions when a velocity profile is imposed. The deforming-cell approach facilitates communications in domain decomposition codes, and the sliding-brick approach is simplest in replicated data codes. Chain molecules like the n-alkanes have internal degrees of freedom (bond vibration, bond-angle bending, and torsion about the dihedral angle) which occur on a faster timescale than the overall molecular motions. A Trotter factorization of the Liouvillian of the motion permits the use of small time steps in the integration of the fast motion and larger time steps in the integration of the slower motion. The combination of all of these techniques has enabled calculation of the rheology of lubricants on the Intel Paragons.

Figure 7 shows the viscosity of the normal alkanes with 1, 10, 16, and 24 carbon atoms versus shear strain rate. The observed changes in the transition from Newtonian to non-Newtonian behavior can be understood in terms of the modes of intramolecular and intermolecular motion which can be visualized from the simulation results. The smallest molecules show behavior in agreement with mode-mode coupling theory, and the longer chains show behavior consistent with Rouse dynamics. Studies in progress of branched alkanes will enable lubricant manufacturers understand how molecular structure affects properties and to design improved synthetic lubricants for the high-speed, high-temperature automobile engines of the future.

Phase Separation in Polymer Blends—By mixing together two polymers it is possible to make new materials (polymer blends) with new properties but without requiring new synthesis. However, such systems have a tendency to phase separate. Theoretical and computational approaches to studying this problem are actively sought. The most direct molecular simulation method to studying phase equilibrium employs Monte Carlo simulation in the Gibbs ensemble. Experimental studies often employ small angle neutron scattering to determine the $\chi$ parameter, which is a measure of the interaction between chain segments. Behavior not predicted by theory, composition dependence of the $\chi$ parameter, which has been observed experimentally can be studied by molecular dynamics employing the same technique to determine $\chi$ is as is used in analyzing experimental data, and the unexpected result can be understood. Both techniques involve simulations of long-chain molecules; Figure 8 shows how we can handle such molecules in domain decomposition codes. The part of the molecule within the subdomain of each processor is treated as an independent molecule, and its bonded interaction with its part in the subdomain of another processor is handled just like the interaction between any pair of particles, one in each subdomain. We have implemented several approaches to parallel Monte Carlo simulations. A domain decomposition approach employs the Time Warp algorithm for parallel discrete event simulation in which each processor calculates independently until it receives a message denoting interaction from another processor. Events must be kept in correct sequence; so, if the receiving processor has progressed in sequence beyond the point of the received interacting event, it rolls back its sequence, inserts the interacting event, and begins calculations anew from that point. A replicated data approach attempts $P$ ambitious Monte Carlo trial moves on $P$ processors in parallel. The most favorable move is tested for acceptance, the bias being corrected for in the acceptance criterion. The maximum speed-up in the replicated data approach is limited by the global communication latency just as in the molecular dynamics case.

Figure 9 shows the results of vapor-liquid phase equilibrium calculations for model chains of 32, 100, and 500 segments, results which are well beyond those previously attainable. Figure 10 shows the $\chi$ parameter as a function of composition for two model polymer blends. From the simulations we can determine the relative contributions of effects of 1) isothermal compressibility, 2) excess volume of mixing, and 3) composition fluctuations to the observed composition variation of $\chi$ which is not predicted by Flory theory.

These calculations have been performed on Intel Paragon supercomputers at Oak Ridge National Laboratory. Based upon these demonstrated capabilities, we have established an ORNL/UT Joint Institute for Molecular-Based Engineering and Science to make these new computational capabilities available for academic, government, and industrial research and development. Further information is available on the Web at http://flory.engr.utk.edu/ltdr/project/ltdr-annual95.html and http://flory.engr.utk.edu/ltdr/w.html.

Acknowledgments

We are pleased to acknowledge the many contributions of A. A. Chialvo, S. T. Cui, R. K. Bhupathiraju, S. A. Gupta, F. A. Escobedo, M. J. Greger, and J. J. de Pablo, upon whom outstanding work this paper is based.

The work of HDC has been supported by the Division of Chemical Sciences of the U. S. Department of Energy (DOE) at Oak Ridge National Laboratory (ORNL). This work was sponsored by the Laboratory Directed Research and Development Program of ORNL. The authors acknowledge the use of the Intel Paragon supercomputers in the Center for Computational Sciences at ORNL, funded by the DOE’s Mathematical, Information, and Computational Sciences Division. ORNL is managed by Lockheed Martin Energy Research Corp. for the DOE under Contract No. DE-AC05-96OR22464.
References


Figure 1. Trajectory through phase space.

Figure 3. Advantages and disadvantages.

Figure 5. Ion pairing of NaCl.

Figure 2. a) Domain decomposition. b) Replicated data.

Figure 4. Supercritical water.

Figure 6. Deforming-cell and sliding-brick BCs.
Figure 7. Rheology of some n-alkanes.

Figure 8. Chain spanning three subdomains.

Figure 9. Vapor-liquid phase equilibrium of chain molecules.

Figure 10. Composition dependence of χ parameter.

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.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.