NSF Graduate Research Fellowship Proposal
Bill March

Computer simulations are widely used in biology and chemistry. *Add a better opening sentence. Discuss some compelling applications here.*

Implementing Simulations. All of these simulations require a tradeoff between detail and scale. Detailed quantum mechanics methods can require quadratic space and $O(N^4)$ time. On the other hand, simplified models can easily lose the details necessary to capture interesting behavior. Therefore, one of the primary challenges of developing practical computer simulations is to find the correct simplifications. In quantum mechanical methods, sets of functions, called basis sets, are combined to approximate the wavefunction. Basis sets are carefully optimized for both speed and accuracy and often are specialized for particular systems. In classical simulations on proteins, a set of parameters are fixed, called a force field. These included atomic charges which are developed to accurately reproduce experiments. Determining a force field is extremely time consuming, and force fields rarely generalize to other systems. New algorithms are necessary to allow larger and more accurate simulations and reduce the need for scientists to spend time developing elaborate parameter sets.

Many-Body Interactions. One of the key (and largely unexamined) simplifying assumptions in nearly all methods is that all interactions between particles can be computed through sums of interactions between pairs. Since computing the interaction between all k-tuples scales as $O(N^k)$, higher-order interactions are generally considered intractable. While this assumption is reasonable for many systems, it fails for some. For example, three-body terms in liquid argon account for about 10% of the total energy (cite). These terms have received little attention, not because they are physically insignificant, but because of the difficulties with computing them.

Efficient Many-Body Potentials. I propose a general framework for making many-body potentials computationally feasible for large systems. It applies methods from statistics and machine learning to generalize the Fast Multipole Method to many-body potentials. I will also release code for the algorithm, enabling widespread scientific investigations of many-body potentials in general.

Fast Two-Body Computations. In general, two-body potentials scale as $O(N^2)$. For many potentials, it is possible to truncate such computations for distant particles. However, for long-range effects, such as the Coulomb electrostatic potential, this simple method is too inaccurate, so one must find a better approximation for distant interactions.

FMM. The Fast Multipole Method [3] makes large-scale electrostatic computations feasible. The FMM groups charged particles in the nodes of a space-partitioning tree, and computes interactions between pairs of particles by considering pairs of nodes. Interactions between nearby groups of particles can be computed exhaustively, as in the truncation method. When two nodes are distant, the algorithm approximates their interactions with a single computation. The potential due to points in a node can be approximated with a Taylor series. This *multipole expansion* is what makes the FMM effective.

N-body Problems in Statistics. The FMM is effective for the Coulomb potential between pairs of particles. I will apply ideas from statistics in order to extend the FMM to include more general potentials involving many particles. Gray and Moore [2] have suggested a framework for accelerating more general computations between pairs of points. This framework has been

applied to computing kernel functions [4], which, like many potentials, depend on distances between points.

In order to compute all pairs interactions between two sets of points, each set is organized in a binary space partitioning tree, like in the FMM. Consider the computation between two nodes, one from each tree. If the nodes are distant enough to make a valid approximation, it is possible to *prune* further computation. Additionally, it is possible to bound the error incurred by this approximation. Otherwise, recurse on both nodes simultaneously, and consider the four pairs of children. This dual-tree recursion can easily be extended to computations involving k-tuples. Instead of employing two trees, use k, and recurse in each simultaneously, pruning as in the dual-tree case.

Many-Body Potentials. In order to apply the multi-tree framework to many-body potentials, I will implement the multi-tree recursion algorithm. Then, each potential requires a series expansion in order to approximate groups of points. Also, the algorithm should include remainder terms from the series expansion to bound the error of our approximation.

Initially, I will develop this framework for the Axilrod-Teller three-body potential [1]. We will develop multipole expansions for this potential, then build a multi-tree framework to handle the recursive approximations. We also intend to test this algorithm on systems of chemical interest in order to demonstrate its effectiveness.

Progress. At Georgia Tech, I have worked with Alex Gray on developing and implementing dual- and multi-tree algorithms. His guidance is important to understanding and expanding these methods. Also, I am collaborating with David Sherrill in chemistry. He has helped identify the importance of many-body potentials. Also, his support and advice will be crucial in ensuring that my algorithms are useful and relevant to computational chemists.

My collaborators and I have developed a simplified monopole expansion using only the constant terms from the Taylor series. This allows some progress in designing the multi-tree algorithm. However, the multipole expansion is still necessary for scientific accuracy.

Extensions. These techniques (meaning series expansions and trees) can be extended to treat many other many-body potentials. We are also considering extensions of this framework to other problems in computational chemistry. As mentioned above, quantum mechanics simulations utilize basis functions. The rate-limiting step in the fundamental Hartree-Fock method is computing electron repulsion integrals (ERI's) among sets of four basis functions. *Give the formula here.* This computation is very similar to a four-body potential, and we believe that our methods can be extended to accelerate this and other methods.

References

[1] B. M. Axilrod and E. Teller. Interaction of the van der waals type between three atoms. *J. Chem. Phys.*, 11(299), 1943.

[2] A. Gray and A. W. Moore. N-body problems in statistical learning. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems 13 (December 2000)*. MIT Press, 2001.

[3] L. Greengard and V. Rokhlin. A Fast Algorithm for Particle Simulations. *Journal of Computational Physics*, 73:325–248, 1987.

[4] D. Lee, A. Gray, and A. Moore. Dual-tree fast gauss transforms. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 747–754. MIT Press, Cambridge, MA, 2006.