

Tractable Quantum Chemistry Computations

Keywords. *Hartree-Fock, higher-order recursion, drug design*

I am investigating more efficient methods for modeling chemical systems at the quantum level. I intend to apply sophisticated algorithmic techniques from computer science to a fundamental calculation in quantum-level molecular simulations.

Applications such as drug and materials design require detailed simulations of the behavior of molecular systems. However, due to computational constraints, some accuracy must be sacrificed. At the most detailed level, quantum mechanical methods model electrons and nuclei separately. Larger systems can only be handled approximately, with atoms, molecules, or entire proteins treated as interacting units. While these approximations greatly increase the size of systems that can be modeled, they are simply too inaccurate for the most important applications such as protein folding and drug discovery.

The basis for all first-principles quantum mechanical simulations is the Hartree-Fock (HF) method [7]. All other, more sophisticated methods such as coupled cluster build on the HF solution [1]. In quantum mechanics, the *wavefunction* governs the behavior of electrons. Hartree-Fock estimates the wavefunction as a linear combination of specified, usually Gaussian, *basis functions*. The HF algorithm determines the correct coefficients in this linear combination. The algorithm can be decomposed into two parts: the formation of the Fock matrix F_{ab} , and the solution to an eigenvalue problem using it.

$$F_{ab} = H_{ab}^{\text{core}} + \sum_c \sum_d P_{cd} \int \int \phi_a(r_1) \phi_b(r_1) ||r_1 - r_2||^{-1} \phi_c(r_2) \phi_d(r_2) dr_1 dr_2$$

Since a, b, c and d range over the basis functions, this computation is naïvely $O(N^4)$, where N is the number of basis functions, thus making it the rate-limiting step. By using Gaussian basis functions, the integrals here have closed-form solutions, which can be represented as Gaussians. I will focus on computing this matrix tractably while maintaining very high accuracy and bounded error.

Since the integrals above are the quantum analog of Columbic interactions, the Fast Multipole Method has been applied to this problem [4]. For pairwise interactions, which are naïvely $O(N^2)$, the FMM computes them in $O(N)$ time. In HF, the Continuous FMM [9] requires truncations and other approximations which lose the rigorous error bounds of the original FMM. Additionally, the CFMM still does not directly attack the $O(N^4)$ computation, since the FMM cannot deal with four-way interactions. Therefore, the FMM has not permitted quantum simulations on large systems.

There is a method, which can be viewed as a generalization of the FMM, that extends to k -way interactions [2, 3]. These approaches arise from a computer science viewpoint and are a form of higher-order divide-and-conquer. This line of work has tackled statistical problems, such as the N -point correlation (which involves a k -way interaction) [6] and kernel density estimation [5]. This recursive method can automatically guarantee rigorous error bounds and can use adaptive space-partitioning trees which are more sensitive to structured data than the oct-trees of the FMM.

For a k -way computation, the algorithm considers interactions between k nodes in the tree at a time. Using the bounding boxes of the nodes, it determines bounds on the possible values of the computation between points. In previous applications,

multipole expansions similar to those in the FMM have been used to improve these bounds. When these bounds indicate the computation will be within a specified relative error tolerance, further computation is *pruned* and the computation between each k -tuple of points in the k -nodes is approximated by the computation between the nodes. Otherwise, the algorithm recursively considers interactions between the children of the nodes.

I am investigating a fast algorithm for computing the Fock matrix using this algorithmic framework. This involves deriving the bounds on computations mentioned above and determining the best tree to organize the basis functions. Once the algorithm is designed, I will implement it in software and test against existing methods. With the support of David Sherrill and his research group, I test my algorithm on systems of real chemical interest. I will also integrate my algorithm with his open-source quantum chemistry code [8], thus providing a useful alternative to the current commercialized programs.

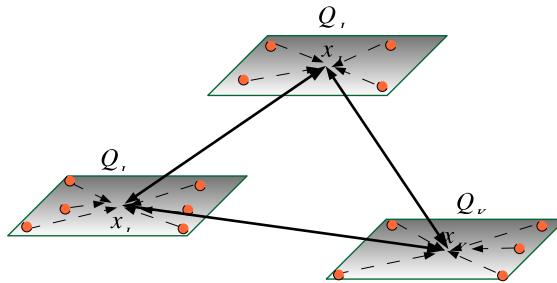


Figure 1: Three nodes Q_I, Q_J, Q_K with bounding boxes. The arrows represent the approximated interactions.

Given the past success of algorithms employing this framework on similar problems, I believe this work will provide a tremendous speedup over existing methods. This will allow simulations on much larger systems than were previously feasible at this level of detail. Additionally, detailed simulations are often used to determine appropriate approximations for simpler ones, so my algorithm will help expand the accuracy of these models as well. By allowing new simulations, and improving existing ones, it should be possible for researchers to make great progress in understanding protein interactions, leading to better drug design.

Originality. This proposal represents my own work and ideas.

- [1] J. Cizek. On the correlation problem in atomic and molecular systems. *J. Chem. Phys.*, 45(11):4256–4266, 1966.
- [2] A. Gray and A. Moore. N-body problems in statistical learning. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *NIPS 13*. MIT Press, 2001.
- [3] A. Gray and R. Riegel. Generalized n-body problems: A mathematical theory for scalable learning. Technical report, Georgia Institute of Technology, 2007.
- [4] L. Greengard and V. Rokhlin. A Fast Algorithm for Particle Simulations. *Journal of Computational Physics*, 73:325–248, 1987.
- [5] 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.
- [6] A. Moore, A. Connolly, C. Genovese, A. Gray, L. Grone, N. Kanidoris, R. Nichol, J. Schneider, A. Szalay, I. Szapudi, and L. Wasserman. Fast algorithms and efficient statistics: N-point correlation functions. In *Proceedings of MPA/MPE/ESO Conference Mining the Sky, July 31–August 4, Garching, Germany*, 2000.
- [7] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23(2):69–89, Apr 1951.
- [8] T. Crawford *et al.* Psi3: an open-source *Ab Initio* electronic structure package. *J. Comput. Chem.*, 28(9):1610–1616, 2007.
- [9] C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon. The continuous fast multipole method. *Chemical Physics Letters*, 230(1-2):8–16, November 1994.