High-Performance Computational Chemistry and Physics: Perspectives from Both Sides of the Table

Robert A. DiStasio Jr.

Department of Chemistry, Princeton University, Princeton, NJ 08544 USA

<u>Without current intervention</u>, the (albeit near) future of high-performance computing (HPC) will provide a *select group of users* access to exascale computational resources—state-of-the-art resources which are inevitably destined to be *grossly underutilized* in the fields of computational chemistry and physics. Although this statement may be viewed as too disparaging, please allow me to play the infamous role of devil's advocate and attempt to justify this stance herein, and in the process, identify some of the questions that logically follow from this line of reasoning. In doing so, I hope to address some of the most important challenges (and potential dilemmas) that lie ahead for the future of high-performance computational chemistry and physics, and in particular, the issues that I believe are most pertinent to the initiative and central mission of the NSF Sustainable Software Innovation Institute for Computational Chemistry and Materials Modeling ($(\text{SICM})^2$).

Access to HPC Resources

By a *select group of users*, I am simply alluding to the fact that access to currently available HPC resources is already quite restrictive: substantial allocation requests are reserved for fairly specialized (mostly academic) research groups who compete for a share of computer time on relatively few available HPC architectures via a proposal-based selection process. Even though the computational resources associated with each of the HPC centers in the USA has been steadily increasing (and is forecast to continue doing so), the average number of users of the current computational chemistry and physics software packages has literally exploded over the last decade, a trend that can arguably be attributed to the success of density functional theory (DFT) in treating a wide range of systems of interest in chemistry, physics, materials science, and many other scientific disciplines. As a result, most of the users of the current software packages in computational chemistry and physics do not actually have access to the HPC resources that are available today (never mind the HPC resources of tomorrow). In this regard, one immediately arrives at the following questions:

- How can we realistically increase access to the current (and future) HPC resources in the computational chemistry and physics communities?

- Does it actually make sense to invest the financial and human resources required to develop and maintain computational chemistry and physics software packages that can effectively utilize the computational resources afforded by the current (and future) HPC architectures, if the access to said resources is only granted to a relatively small subset of the computational chemistry and physics communities?

From a business and/or economical point of view, the lack of large-scale access to HPC resources resulting from this pyramidal segregation of the computational chemistry and physics communities is an inhibitor of (any kind of) substantial horizontal growth; hence, the commercial computational chemistry and physics software industry simply does not have the financial impetus to be the <u>sole</u> driving force in the development and maintenance of software packages that can effectively utilize the HPC resources that are available even today (again never mind the HPC resources of tomorrow). In the same breath, however, it would be imprudent and misguided for the community as a whole to disregard the tremendous contributions of the commercial software sector to the fields of computational chemistry and physics; the existence of effectively optimal serial algorithms in several of the leading commercial software packages (particularly in the computational chemistry community) have essentially paved the way for high-performance computational chemistry and physics today. Hence, one is again lead to another set of questions:

- Without large-scale access to HPC resources, will the future of high-performance computational chemistry and physics be delegated to the (primarily academic) open-source software community?

- On the other hand, would large-scale access to HPC resources allow for (or even foster) a continued collaborative effort between industry and the open-source community with the goal of attaining truly high-performance computational chemistry and physics?

Efficient Utilization of HPC Resources

For the sake of argument, I would also contend that there is not a single computational chemistry or computational physics software package in existence today that can *maximally* utilize the currently available HPC resources (once again never mind the HPC resources of tomorrow). In order to provide a computational physics point of view on this matter, the current discussion will be restricted to applications of DFT-based *ab initio* molecular dynamics (AIMD) in condensed phased systems (*e.g.*, periodic solids and liquids) utilizing a pseudopotential-based plane wave (originless and delocalized) basis for the expansion of the requisite quantities (*e.g.*, wavefunctions, charge densities, etc.). DFT-based AIMD simulations represent a majority of the calculations performed by computational physicists today and are therefore very representative of this community.

Within the framework of DFT-based AIMD simulations, there are two apparent and immediately obvious branches in the development of software capable of truly large-scale massively-parallel applications:

- **Accessing Large Time Scales in Standard DFT-AIMD Simulations.** Out of necessity, computational chemists and physicists have become quite adept at solving fundamental problems of interest using systems which contain approximately 100-1000 atoms. In this size regime, "standard" density functional approximations for the exchange-correlation potential, such as the local density approximation (LDA) and the generalized gradient approximation (GGA), have become the computational workhorses for attacking problems of interest to the computational physics community. The simulation times that are actually accessible during such calculations are relatively short and range from approximately 1-100 picoseconds (ps); with simulations times of this order, it is often difficult to converge even static equilibrium properties, which makes accurate determinations of dynamical properties most often completely out of reach. The main parallelization scheme utilized to perform AIMD simulations using the aforementioned functionals partitions the real-space grid along a given axis (like the principal z-axis), thereby allowing for the use of the fast Fourier transform (FFT) algorithm on individual slabs (or the respective x-y planes). For systems in this size regime (and typical choices for the plane wave cutoff parameter), the number of points in the real-space grid is on the order of $(100\text{-}1000)^3$, thereby limiting the extent of truly scalable parallelization to 100-1000 processors. The real time cost associated with this class of simulations can be further reduced using shared memory multiprocessing (*via* OpenMP or Pthreads) and task groups (which further distribute the workload associated with each band/orbital), but the computational savings afforded by these schemes quickly pan out and such approaches are not amenable to large-scale massively-parallel (super)computer architectures. In this branch of development, alternative FFT parallelization schemes might be a potential solution, but one is nevertheless forced to ask if the plane wave framework itself is a fundamental limitation in the effective utilization of HPC resources...

- **Enabling AIMD Simulations Using Hydrid Functionals (and Beyond).** For problems of interest in the computational physics community that require a more sophisticated treatment of the underlying electronic structure theory, hybrid exchange-correlation functionals represent the next frontier in DFT-based AIMD simulations. In this regard, the combined use of sparsity (*via* orbital localization) and an efficient MPI-OpenMP algorithmic implementation can allow for linear scaling hybrid DFT calculations that utilize 10,000-100,000 processors on the BlueGene/Q (Mira) computer architecture at the Argonne Light Computing Facility (ALCF). As a result of these efforts (carried out at Princeton University in conjunction with the open-source Quantum ESPRESSO (QE) software package), the overall computational cost of hybrid DFT calculations has been reduced by 1-2 orders of magnitude and are now only 2-3 times more expensive (in real time) than standard LDA- or GGA-DFT calculations. This parallelization scheme makes highly accurate hybrid DFT calculations on condensed phase systems *feasible* by attempting to efficiently utilize the HPC resources available today. However, the actual computational cost associated with a hybrid DFT-based AIMD simulation is still quite substantial; as an example, a 20 ps simulation of liquid water (represented here by 64 water molecules in a periodic cubic box) at the hybrid DFT level of theory requires approximately 10-15 million core hours, an allocation time which is not accessible to most of the computational physics community. In this branch of development, hybrid DFT functionals are really only the tip of the iceberg, as many applications of interest to the computational physics community will require an even more sophisti-

cated treatment of electron correlation effects. In this regard, the methods of many-body physics, which have actually (and somewhat ironically) been the focus of intense development in the quantum chemistry community for several decades now, will be the primary choice for most computational physicists that need to venture beyond DFT. With the proper support and an open line of communication, these two communities will have a unique opportunity here to essentially coalesce and collectively attack the frontiers of theoretical chemistry and physics *via* the efficient utilization of truly large-scale high-performance computation.