Electronic Structure Calculations on Thousands of CPUs and GPUs

J. Bernholc, W. Lu, M. Hodak and Emil Briggs

Department of Physics, North Carolina State University

Raleigh, NC, USA

The latest generation of supercomputers is capable of multi-petaflop peak performance. These results are achieved by using thousands of multi-core CPU's, often coupled with thousands of GPU accelerators. However, efficient utilization of this computing power for electronic structure calculations presents significant challenges. Previous parallelization schemes based on one MPI process per CPU core have worked well on prior generations of supercomputers, but current state-of-the-art machines typically have hundreds of thousands of CPU cores distributed over thousands of network nodes. Using one MPI process per core in these situations leads to an unacceptably large performance degradation and also affects scalability.

We describe an alternative parallelization strategy adopted in the Real-Space Multigrid (RMG) code, which dramatically improves scalability and performance. Specifically, we employ a hybrid approach with one MPI process per node, rather than per core, and achieve intra-node parallelization by using POSIX threads and OpenMP. The advantages of this approach are: (i) the number of MPI processes is reduced by an order of magnitude, and (ii) shared memory is used within a node, dramatically reducing the ratio of communication to computation.

Turning to supercomputers with GPU accelerators, a large fraction of their peak performance is usually provided by the accelerators. However, adapting a code originally designed for CPU's to GPU accelerators is not always straightforward. GPU's perform well on vector-type operations, but they have much lower scalar performance than CPU's and a markedly different programming model. Furthermore, the current generation of accelerators utilizes a separate memory space and data transfer between the CPU and GPU memories is costly.

In our case, optimizing RMG for GPU utilization required significant restructuring of code. When combined with our MPI/Posix threads/OpenMP parallelization, we have obtained large performance increases and excellent scalability to over 100,000 CPU cores on the Cray XE6, as well as more than threefold performance improvement on the Cray XK7 system, in which the XE6 nodes with two multi-core CPUs are replaced with XK7 (CPU-GPU) nodes.