Challenges and Solutions for Peta- and Exa-Scale Programming

Daisuke Takahashi
University of Tsukuba, Japan

When Exa-scale system will come?

- 1964: First Megaflops system
  - CDC 6600 (Roots of RISC)
- 1983: First Gigaflops system
  - NEC SX-2 (Vector)
- Jun. 1997: First LINPACK Teraflops system
  - Intel ASCI Red (Parallel)
- Nov. 2008: First LINPACK Petaflops system
  - IBM Roadrunner (Heterogeneous)
- Nov. 2018?: First LINPACK Exaflops system

Issues on Peta- and Exa-scale programming

- Nov. 2008 TOP500 Supercomputing Sites
  - Roadrunner: 1,105.00 TFlops (129,600 Cores)
  - Jaguar (Cray XT5 QC 2.3GHz): 1,059.00 TFlops (150,152 Cores)
  - The number of cores will be increasing.
- Systematic issues
  - Communications (esp. latency)
  - By reducing the number of MPI processes required, the scaling of application programs should be improved.
- Algorithmic issues
  - Degree of parallelism
  - Accuracy (quad-precision arithmetic?)

Programming style

- The combination of MPI and OpenMP will remain in wide use through the decade of the 2010's.
- The MPI has been successful as a portable way to program HPC applications.
- MPI + OpenMP programs remain portable while exploiting hardware capabilities.
- Most application programmers are too conservative.
  - Fortran has been in use for more than fifty years!

My challenges

- The fast Fourier transform (FFT) for Peta- and Exa-scale system.
- A typical decomposition for performing a parallel 3-D FFT is slabwise.
  - A 3-D array $x(N_1, N_2, N_3)$ is distributed along the third dimension N_3.
  - N_3 must be greater than or equal to the number of MPI processes.
- This becomes an issue with very large node counts for a massively parallel cluster of multi-core processors.