Miniapps for Enabling Architecture-Application Co-design for Exascale Supercomputing

Naoya Maruyama
nmaruyama at riken.jp
RIKEN Advanced Institute for Computational Science

19th Workshop on Sustained Simulation Performance
March 28, 2014

HPCI Application Feasibility Study

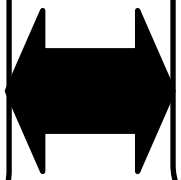
Objective

1. Identify social and scientific challenges for next 5 to 10 years → Computational Science Roadmap
   - Based on the 2011 Computational Science Roadmap
   - Involves a wide range of domains with particular focus on cross-cutting issues
2. Present requirement and tools for evaluating architectures
   - Social and scientific challenges as computational problems
   - Effectiveness to solve the social and scientific challenges

Organization

Roadmap Development:

- Led by major computational science researchers in Japanese universities and national labs



System Evaluation:

- Led by RIKEN AICS and Tokyo Institute of Technology in partnering with major HPC centers in Japan

HPCI Application Feasibility Study

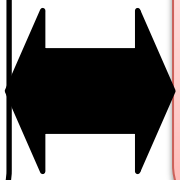
Objective

1. Identify social and scientific challenges for next 5 to 10 years → Computational Science Roadmap
   - Based on the 2011 Computational Science Roadmap
   - Involves a wide range of domains with particular focus on cross-cutting issues
2. Present requirement and tools for evaluating architectures
   - Social and scientific challenges as computational problems
   - Effectiveness to solve the social and scientific challenges

Organization

Roadmap Development:

- Led by major computational science researchers in Japanese universities and national labs



System Evaluation:

- Led by RIKEN AICS and Tokyo Institute of Technology in partnering with major HPC centers in Japan

Miniapp?

A tool for application and architecture codesign

- *Small* applications
  - Lines of code < 1,000
  - Simplified program organization
- *Not too small* applications
  - Retain essential characteristics of the original applications
  - What is essential?
- Simple process to "Download → Compile → Execute"
  - Open-source licenses
  - Easy compilation steps
  - Documentation of program execution input and parameters
  - Packaging of necessary input files

Existing Miniapp Projects

- Mantevo
  - US Sandia NL
  - http://mantevo.org
  - PDE, MD, etc.
- ExaCT
  - US DoE Codesign center
  - http://exactcodesign.org/
  - Combustion miniapp
- LULESH
  - US Lawrence Livermore NL
  - Hydrodynamics
  - Many variants: Serial, MPI, OpenMP, CUDA, OpenACC, Chapel, Charm++, Listz
- New projects at EU

FiBER Mini-App Suite

- A suite of miniapps derived from the full-scale applications for the future computational science challenges
  - Supported by the HPCI Application FS
- Originally developed and used on high-end machines such as the K supercomputers
- Mainly developed at AICS with collaboration with the full-app developers

		処理量 /1PE	処理量 /1チップ	1stepの時間 （演算の時間）	実効効率	性能 /1チップ	性能 /1グループ	ホストへの データ送受信 (双方向の計)
1億原子	構成A （12.3Tflops）	原子数6 （1セル）	原子数24,576 （16x16x16）	0.75ms (0.61ms)	39.7%	4.88 Tflops	20.0 Pflops	0.18ms
	構成B （8.2Tflops）	原子数12 （1セル）	原子数24,576 （16x16x8）	1.02ms (0.86ms)	41.0%	3.36 Tflops	13.8 Pflops	0.18ms
2.5億原子	構成A （12.3Tflops）	原子数15 （1セル）	原子数61,440 （16x16x16）	2.02ms (1.79ms)	43.4%	5.33 Tflops	21.8 Pflops	0.44ms
	構成B （8.2Tflops）	原子数30 （2セル）	原子数61,440 （16x16x16）	3.04ms (2.68ms)	43.2%	3.53 Tflops	14.5 Pflops	0.44ms

E.g., Evaluation results of a MD miniapp at the TSUKUBA FS

Development & Usage

- Call for full-apps
  - 17 full-apps submitted to RIKEN
  - Being converted to mini-apps at RIKEN
- Call for mini-apps
  - 8 mini-app-like small apps submitted to RIKEN
  - Packaged as mini-apps at RIKEN
- Provided to the System Architecture FS for evaluating their systems
  - University of Tohoku, University of Tsukuba, University of Tokyo

Submitted Applications（1/3）

feram	Nishimatsu	Ferroelectrics MD	OpenMP (with MPI for parameter survey), 3D FFT
MARBLE	Ikekuguchi	MD(PME)	MPI+OpenMP, 3D FFT
SMASH (para-TCCI)	Ishimura	Hartree-Fock	MPI+OpenMP, Sequential diagonalization of dense matrices
FFVC	Ono	Thermal-Fluid Analysis	MPI+OpenMP, SOR or GMRES
pSpatiocyte	Iwamoto	Signal propagation	MPI+OpenMP
NEURON_K+	Kazawa	Neural circuit simulation (Modified NEURON)	MPI+OpenMP, many ALL_GATHER, Translated to C from modeling language
GT5D	Idomura	5D plasma turbulence (5D FDM+2D FEM)	MPI+OpenMP, CG, 1D FFT
MODYLAS	Ando	MD(FMM)	MPI+OpenMP
STATE	Inagaki	First-Principles MD (DFT)	MPI+OpenMP (Replica paralle, k-point parallel, band or plane-wave parallel), FFT, eignevalue problem (RMM)
FrontFlow/blue	Yamade	Thermal-Fluid Analysis (Irregular mesh、FEM)	MPI+auto parallelization, BiCGSTAB
SiGN-L1	Tamada	Neural network (L1 regularization)	MPI+OpenMP, Bottleneck at file output
NTChem/RI-MP2	Katouda	Electron correlation	MPI+OpenMP, DGEMM, Some sequential computation, Memory usage $O(N^3)$

Submitted Applications（2/3）

OpenFMO	Inadomi	Hartree-Fock FMO	MPI+OpenMP, Dynamic load balancing
CONQUEST	Miyazaki	First-principles MD, O(n) method	MPI+OpenMP SpMV, FFT
NGS Analyzer	Tamada	Genome sequence analysis	MPI I/O bound
DCPAM	Nishizawa	Climate model with spectral method	MPI+OpenMP
RSGDX	Hyodo	Earthquake simulation	

Submitted Applications（3/3）

Simplified versions submitted to the RIKEN FS

fft_check, fft_check_mpi	Nishimatsu	3D FFT benchmarking
ALPS/looper	Todo	Quantum monte carlo，linked lists, integer ops MPI+OpenMP
CCS QCD Solver Benchmark test program	Ishikawa	Lattice QCD benchmark，BiCGStab, MPI+OpenMP
ZZ-EFSI	Sugiyama	Fluid Structure Integration MPI+OpenMP
rmcsm bench nocore	Shimizu	Monte Carlo Nuclear Shell Model MPI+OpenMP
GCEED	Nobusada	DFT
NICAM-DC	Yashiro	Climate model (NICAM), dynamics，FVM MPI
mVMC	Imada	Strongly correlated matter

FiBER Mini-App Current Status

CCS QCD	Lattice QCD	→Tokyo FS, Tsukuba FS
MARBLE	MD(PME)	
MODYLAS	MD(FMM)	→Tokyo FS, Tsukuba FS
FFVC	Thermal-Fluid Analysis (Cartesian, FDM)	→Tohoku FS
NGS Analyzer	Genome Sequence Analysis	→Tokyo FS, Tsukuba FS
ALPS/looper	Quantum Monte Carlo	(Almost done)
CONQUEST	First-Principles MD (O(N))	(Almost done)
NICAM-DC	Climate	(Almost done)
FrontFlow/ blue	Thermal-Fluid Analysis (Irregular mesh, FEM)	(Under development)
mVMC	Variational Monte Carlo	(Under development)

Molecular Dynamics

- Two alternative algorithms for solving equivalent problems
  - Particle Mesh Ewald
    - Bottlenecked by all-to-all communications at scale
    - Example implementation: MARBLE (Ikeguchi et al.)
  - Fast Multipole Method
    - Tree-based problem formulation with no all-to-all communications
    - Example implementation: MODYLAS (Okazaki et al.)
- Allows algorithmic comparisons
  - FFT vs. FMM?

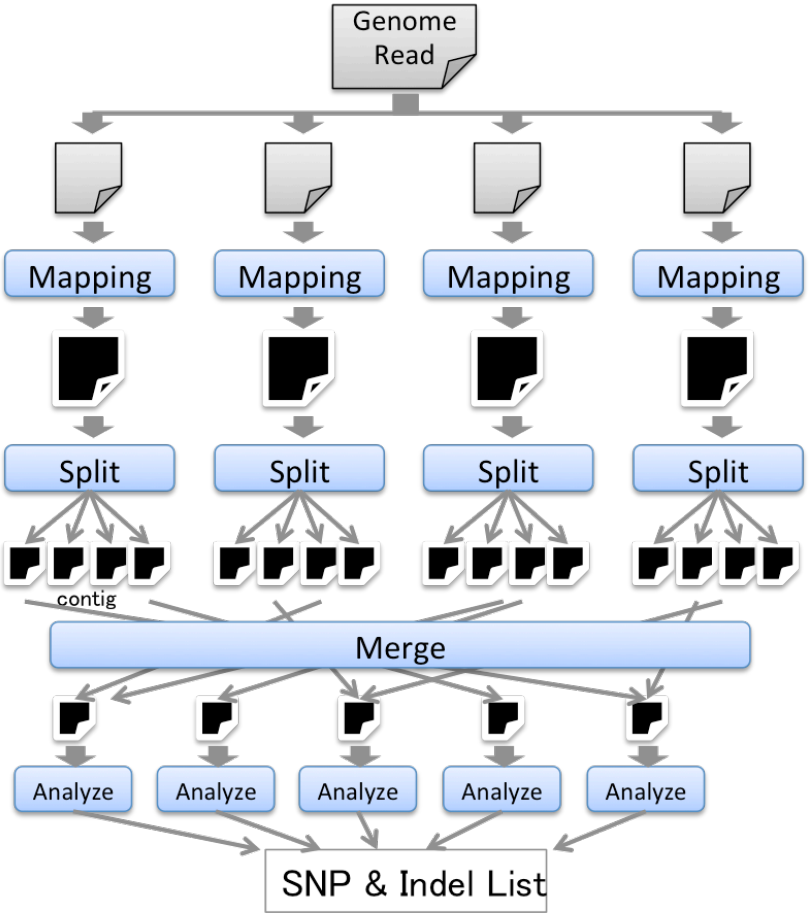
Molecular Dynamics

- Simplified problem settings
  - Only simulates water molecules in the NVE setting
  - Can reduce the codebase significantly
  - Easier to create input data sets of different scales
  - Whether it's sufficient is still under discussions
- Kernels: Pairwise force calculation + Long-range updates (FFT or FMM)
- Two reference implementations to study performance implications by algorithmic differences
  - MARBLE (12K SLOC)
  - MODYLAS (11K SLOC)

NGS Analyzer

- Genome analysis tool for cancer cell's mutation detection
  - Read the output genome data generated by a next-generation genome sequencer
    - Current: 500GB/human
    - In 2020: 100TB/human
  - The analysis pipeline consists of widely used genome analysis software, and performs sequence mapping, genotyping, etc.
    - BWA sequence mapping
    - SAM/BAM formatting software
  - Runs on K computer in parallel

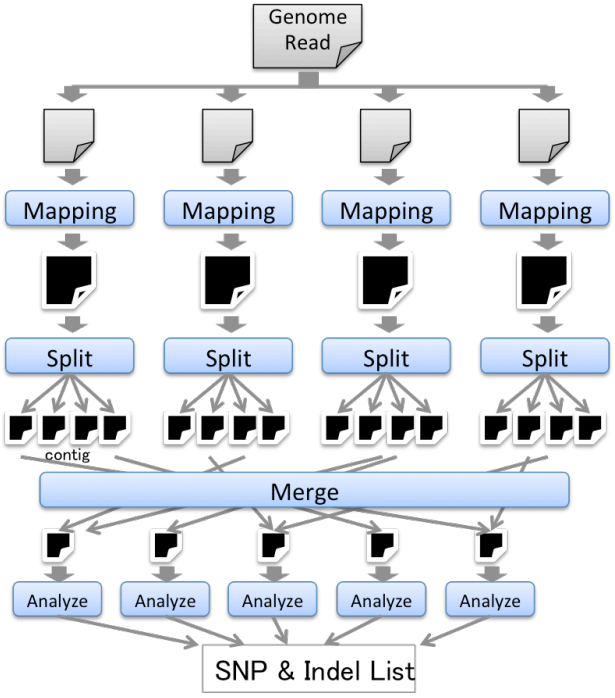
Work Flow of NGS Analyzer



- Workflow
  - Split the input genome read, perform mapping on each of them and then split the results based on contigs
  - Merge mapping results of each contig and run analysis process on each merged data
- The original NGS Analyzer runs this workflow by executing five separated jobs in turn

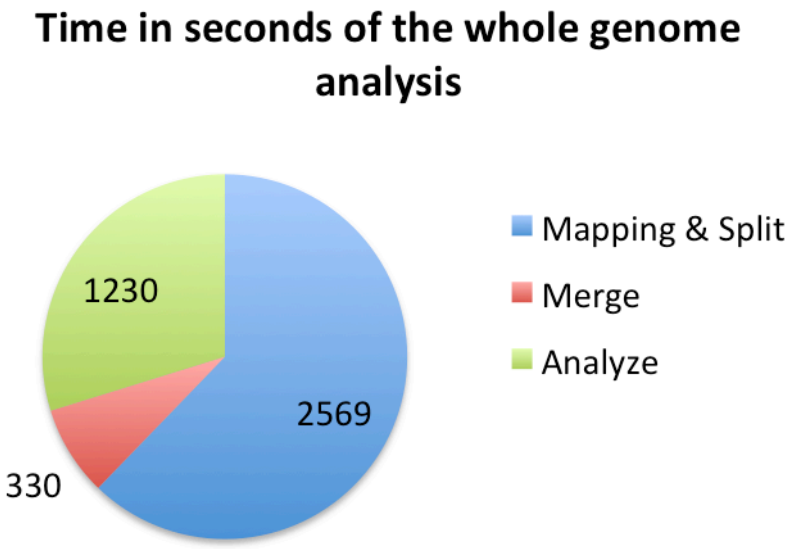
NGS Analyzer IO

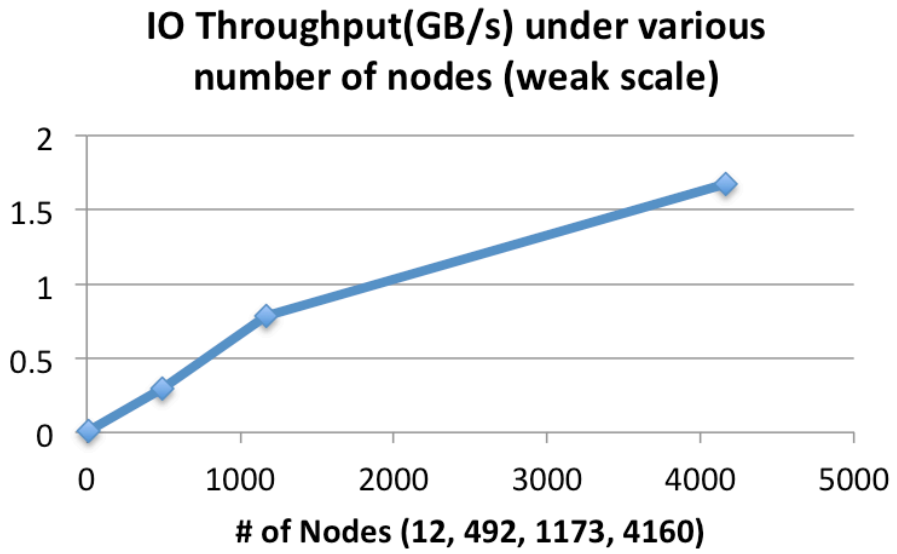
- Input
  - Output of a next-generation sequencer
  - Reference genome
  - e.g.) whole genome of a Japanese individual
    - 490 GB of genome read,
      6.3GB of reference genome
- Output
  - Analysis results of each config
  - e.g.) Output of the above input: 874 MB
- Intermediate output
  - unsplit mapping results, results that contain duplicate sequence, etc.
  - e.g.) Intermediate output of the above input: 6 TB
- Though IO of “Mapping”, “Split” and “Analyze” can be local, IO of “Merge” is global where all nodes exchange data
  - e.g.) Exchanged data of the above input: 617 GB



NGS Analyzer Performance

- Execution time for analyzing the whole genome of a Japanese individual : 4,129 sec
  - IO throughput is 1.6 GB/s
    (490GB + 874MB + 6TB) / 4129
  - Measured on the K computer using 4,160 nodes





NGS Analyzer Mini App

- Three serial programs that run individual steps of the workflow to measure both computational and IO performance of each step
  - Sequence mapping
  - Duplicate removal of the mapping results
  - Mutation detection
- Single program that runs the entire workflow in parallel to measure the overall performance

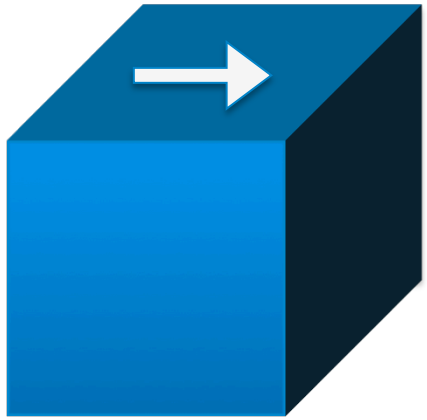
FFVC (FrontFlow/Violet Cartesian)

- 3D unsteady incompressible thermal fluid solver developed at University of Tokyo
- Regular Cartesian grids, Finite volume, Fractional Step method
- Mini-app version
  - 3D cavity flow cavity flow model
    - Outer boundary condition
    - Retain the original control flow

FFVC mini

- Algorithm (fixed)
  - Time step: Euler explicit method
  - Pressure Poisson: Red-black SOR
- Simplified program execution
  - No configuration nor input file necessary
- Code size

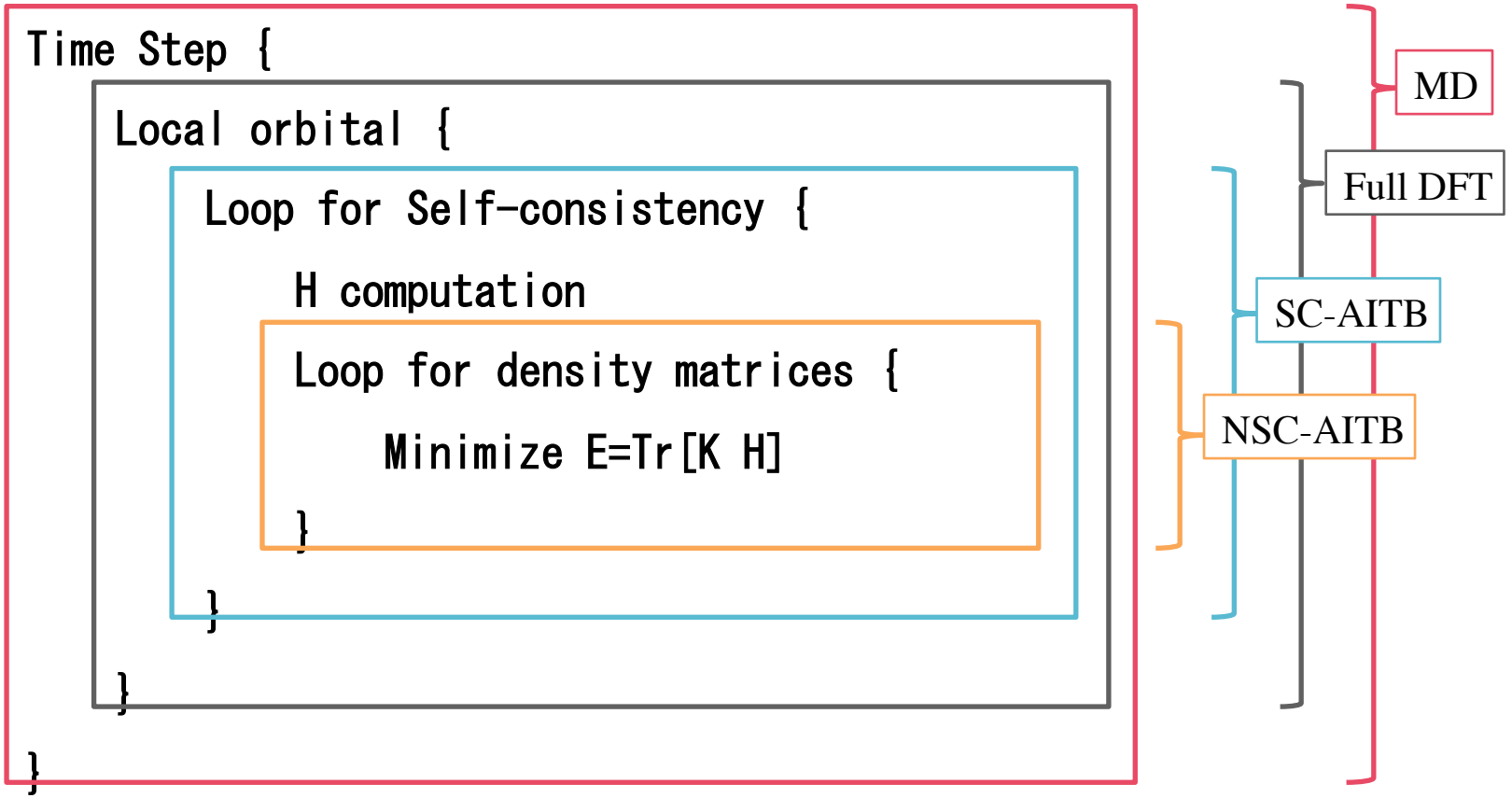
  100K SLOC → 10K SLOC (sloccount)



CONQUEST

- Domain and method:
  - First-Principles computation with O(n) method
  - Structural optimization, molecular dynamics
- Author:
  - Tsuyoshi Miyazaki（National Institute for Materials Science）, et al.
  - URL: http://www.order-n.org/
- Programming language:
  - Fortran90
- Program size:
  - 115K SLOC
- Target problem size:
  - Ensemble simulations with 100K to 1M atoms
  - Simulation of 100M atoms

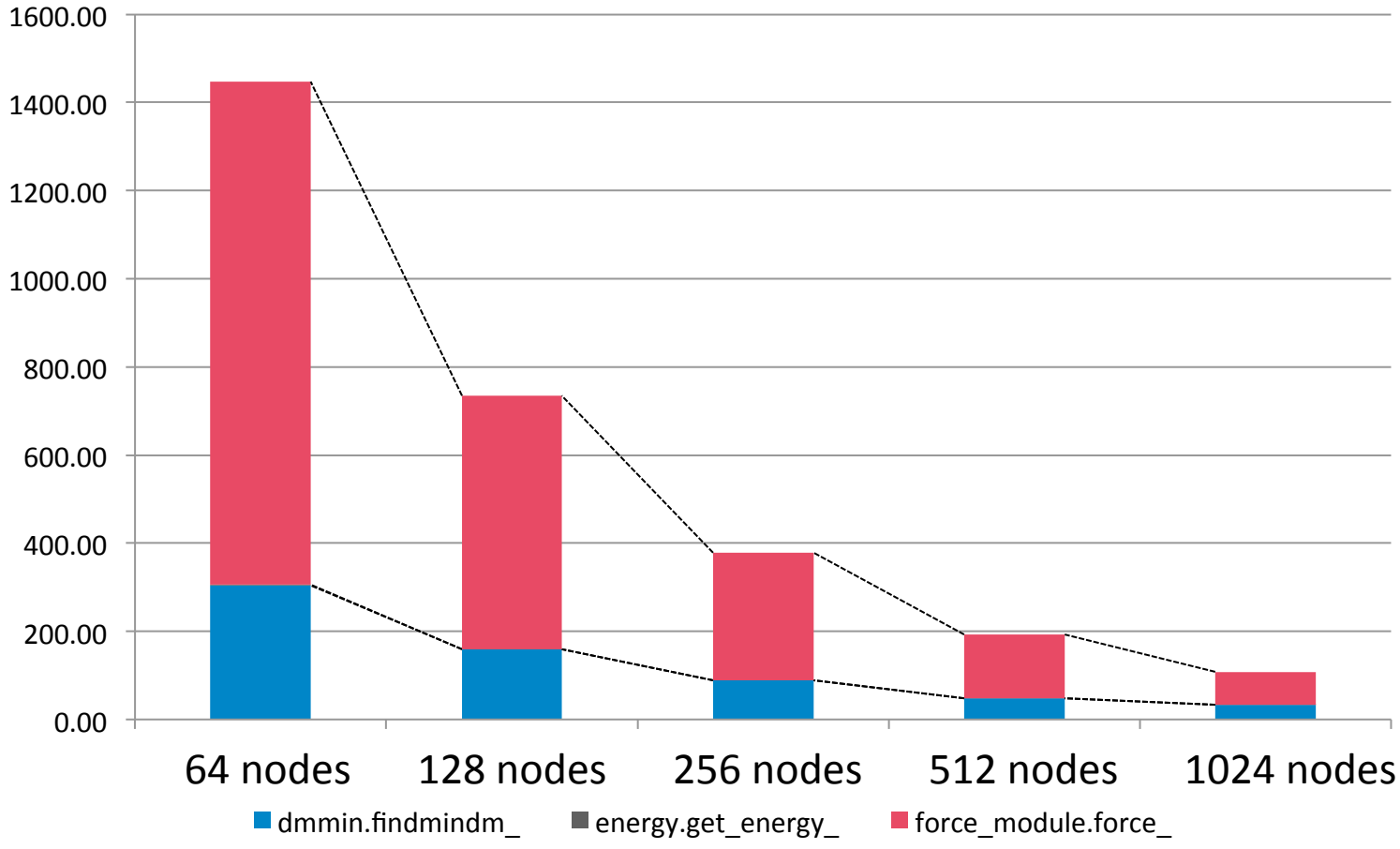
Program Loop Structure and Mini-App



- Compute NSC-AITB only as benchmark mode
  - The time for full DFT can be easily estimated from the time for NSC-AITB
- Approximately 25K SLOC

Strong Scaling (Atom Si32768)

- FX10 (Oakleaf @ U-Tokyo)



Future Plan

- Need to be continuously developed and maintained as the Application Roadmap progresses
- Will be continued as part of the Exa-scale project at RIKEN AICS
  - Positions available! http://bit.ly/1nxI22L
- Wider coverage of application domains
- Optimization, porting, performance modeling
- First release
  - http://github.com/fiber-miniapp
  - miniapp at riken.jp
- Usage of miniapps
  - M. Kondo et al., “Evaluation of power dissipation of HPC systems using miniapps,” SDHPC10