Productivity and Performance of XcalableACC Language for Acceralated Clusters

Masahiro Nakao[1], Hitoshi Murai[1], Hidetoshi Iwashita[1], Akihiro Tabuchi[2], Taisuke Boku[2], Mitsuhisa Sato[1]
1. RIKEN AICS 2.University of Tsukuba

XcalableACC

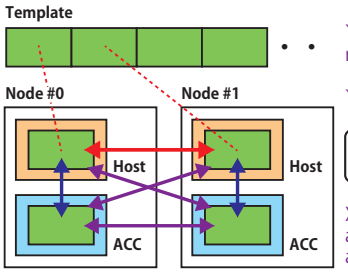
Overview

- XcalableACC (XACC) is a hybrid model of XcalableMP (XMP) and OpenACC for accelerated clusters
- XACC is a directive-based language extension of C and Fortran
- XACC enables programmers to develop applications on acceralated clusters with ease

Components

- XMP for distributed-memory parallelism
  XcalableMP XMP is a directive-based language extension of C and Fortran for cluster systems
- OpenACC for offloading works for accelerators
  OpenACC DIRECTIVES FOR ACCELERATORS OpenACC is another directive-based language extension for heterogeneous CPU/Accelerator systems
- XACC extensions for communication of data on accelerators

Memory Model



``Template'' is a dummy array that represents a global index space.

``Node'' is an execution unit.

XACC directive transfers data between accelerator memories, and between accelerator memory and host memory.

Omni XcalableACC Compiler

- *http://omni-compiler.org*
- Developed by RIKEN AICS and Center for Computational Sciences in University of Tsukuba

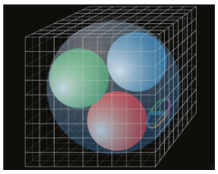
Evaluation of Lattice QCD

What is Lattice QCD ?

- Solve the quantum chromodynamics (QCD) theory of quarks and gluons

$$D_{x,y} = \delta_{x,y} - \kappa \sum_{\mu=1}^{4} \{(1-\gamma_\mu)U_\mu(x)\delta_{x+\hat{\mu},y} + (1+\gamma_\mu)U_\mu^\dagger(x-\hat{\mu})\delta_{x-\hat{\mu},y})\}$$

- The four-dimensional space-time continuum is replaced by a four-dimensional hypercubic lattice
- Lattice QCD simulates dynamics of a "quark," which is the smallest unit of a substance, and a "gluon" that connects the quarks



- We have developed XACC version Lattice QCD based on Bridge++, which is a real world application (http://bridge.kek.jp/Lattice-code/)
- A quark is also represented by a "color" that has three color charges, and a "spinor" that has four colors. A gluon field that exerts influence on the quark interactions is defined as a 3 × 3 complex matrix in each direction

A part of code

A programmer adds XMP and OpenACC directives into the serial Lattice QCD code.

```
QCDSpinor_t v[NT][NZ][NY][NX];     // Quark
QCDMatrix_t u[4][NT][NZ][NY][NX];  // Gluon
...
#pragma xmp align v[i][j][*][*] with t[i][j]
#pragma xmp align u[*][i][j][*][*] with t[i][j]
#pragma xmp shadow v[1][1][0][0]
#pragma xmp shadow u[0][1][1][0][0]
#pragma acc enter data copyin(v, u)
...
#pragma xmp reflect (v) width(/periodic/1:1,/periodic/1:1,0,0) acc
#pragma xmp reflect (u) width(0,/periodic/1:0,/periodic/1:0,0,0) acc
...
#pragma acc parallel loop present(v, u) ···
#pragma xmp loop (it,iz) on t[it][iz]
  for(it = 0; it < NT; it++){
    for(iz = 0; iz < NZ; iz++){
      for(iy = 0; iy < NY; iy++){
        for(ix = 0; ix < NX; ix++){
          ···
```

Define XMP distributed arrays

Exchange halo region

Parallelize loop

OpenACC directive parallelizes the loop statement parallelized by XMP directive

Evaluation

http://ccs.tsukuba.ac.jp/eng/research-activities/projects/ha-pacs/

- Ivy Bridge E5-2680v2, 10Cores x 2 Sockets
- DDR3 128GB (59.7GB/s x 2, NUMA)
- NVIDIA K20X (D.P. 1.31TFlops) x 4 GPUs
- GDDR5 6GB (250GB/s)
- InfiniBand 4xQDR x 2rails, 8GB/s
- MVAPICH2 2.1, Intel 16.0,2, CUDA 7.5.18

Productivity (Comparison with MPI+CUDA and MPI+OpenACC)

	Serial	MPI+CUDA	MPI+OpenACC	XcalableACC
SLOC(*)	842	1,091	1,002	996
Delta SLOC(**)	-	832	223	160
Added	-	322	160	154
Deleted	-	73	0	0
Modified	-	437	63	6

(*) Source Lines of Code

(**) How the SLOC change from the serial code

Performance (Time required to solve one CG iteration)