GPU & MIC for Lattice Field Theory

Steven Gottlieb
Indiana University

(MILC Collaboration & USQCD)

NERSC HEP Requirements Review
Rockville, MD
November 27-28, 2012

Introduction

- GPU computing has been embraced by the lattice QCD community over the past four to five years
- Used more for analysis runs than for configuration generation
- JLab and Fermilab both have production GPU clusters for USQCD
  - some 20 projects allocated time
- We are still developing code and learning how to best configure machines
- Rapidly changing GPU environment is both an opportunity and a challenge

Code development

- “Lattice QCD as a video game”, G. Egri *et al,* hep-lat/0611022, predates CUDA
- “Blasting through lattice calculations using CUDA”, K. Barros *et al*, hep-lat:0810.5365, starts the development of QUDA, QCD software based on CUDA
- BU team (Barros, Babich, Brower, Clark, Rebbi) concentrated on Wilson type quarks
- Staggered development took off when Guochun Shi at NCSA started working with me during my sabbatical there (2009)
- There are additional developers (see below) working on domain wall quarks, other theories, etc.

QUDA Status

- Code distributed from http://lattice.github.com/quda
- Current author list: Ron Babich, Kipton Barros, Rich Brower, Mike Clark, Justin Foley, Joel Giedt, Steve Gottlieb, Bálint Joó, Claudio Rebbi, Guochun Shi, Alexei Strelchenko
  - one of my concerns is that key developers are moving to industry: Babich and Clark are now at NVIDIA, but can still participate; Shi left NCSA for Google
- Solvers for Wilson/Clover and staggered quarks
- HISQ/asqtad link fattening
- Gauge force term
- HISQ/asqtad fermion force

Production Running

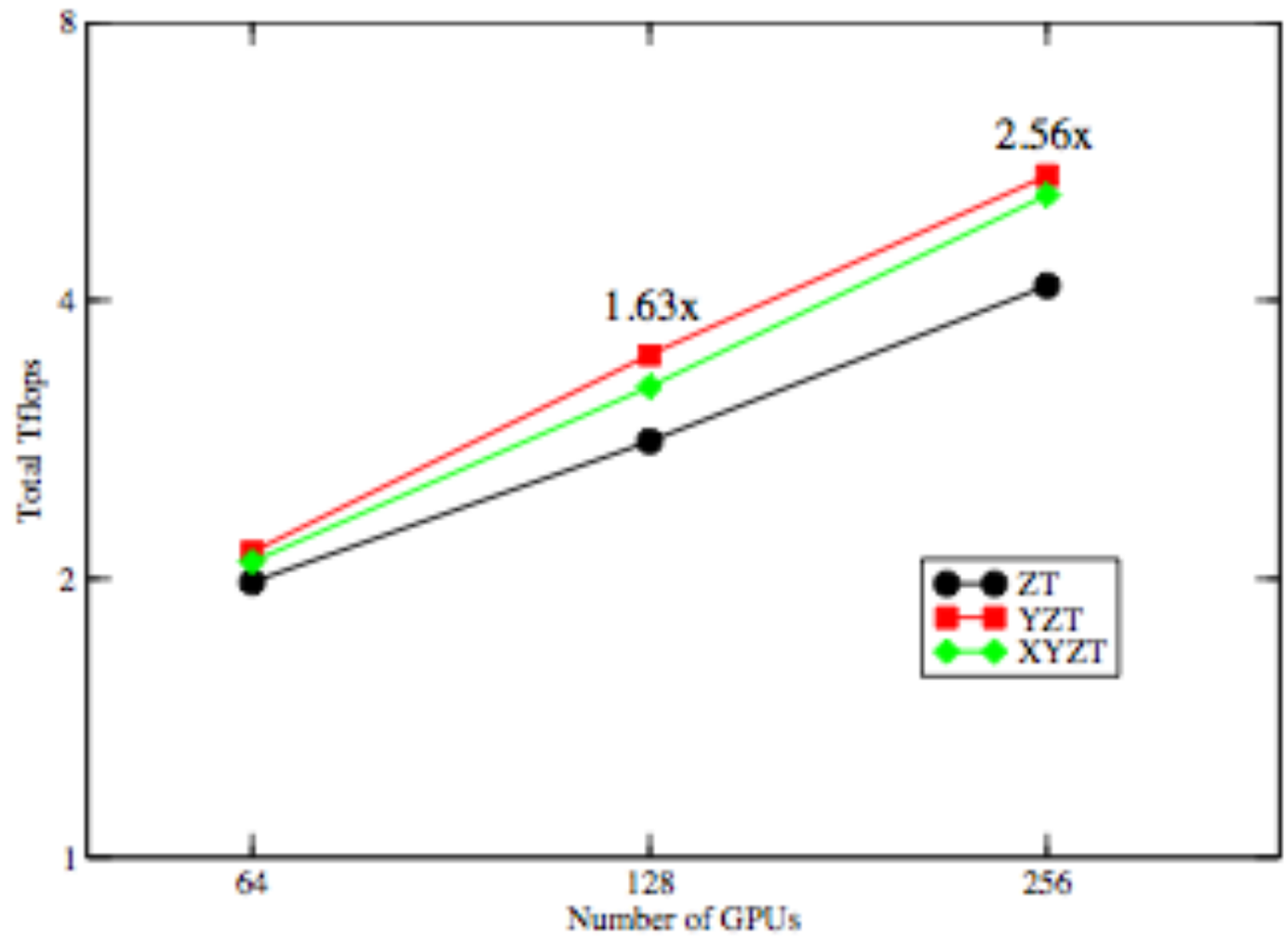
- MILC has used GPU clusters primarily for its electromagnetic corrections project
- GPU clusters include: longhorn (TACC); greenstreet & forge (NCSA)
  - getting started on Keeneland (64 GPUs for $64^3 \times 192$)
- At Jlab, Chroma & QUDA have been used extensively for hadron spectrum calculations of excited states
- HotQCD using LBL clusters for T>0 lattice generation
- Benchmarking is currently in progress on TitanDev and Blue Waters
- Should be possible to get figures for all USQCD projects allocated to GPU clusters

Performance

- Most of the production running on USQCD clusters uses 32 or fewer GPUs
- Staggered benchmarks on the next page were run on Edge computer at LLNL
  - two NVIDIA Tesla M2050 GPUs per node
  - dual socket 6-core Intel X5660 Westmere at 2.8 GHz
  - QDR Infiniband
  - ECC enabled on GPU
- Wilson benchmarks on following page courtesy of B. Joó on ORNL Cray XK6 TitanDev system

Scaling with Staggered Quarks

- 64^3 X 192 lattice
- Mixed precision multi-mass solver
- Achieving 5.5 TFlops on 256 GPUs
- From Babich et al., Supercomputing '11, Lattice '11



Architecture Aware Algorithms

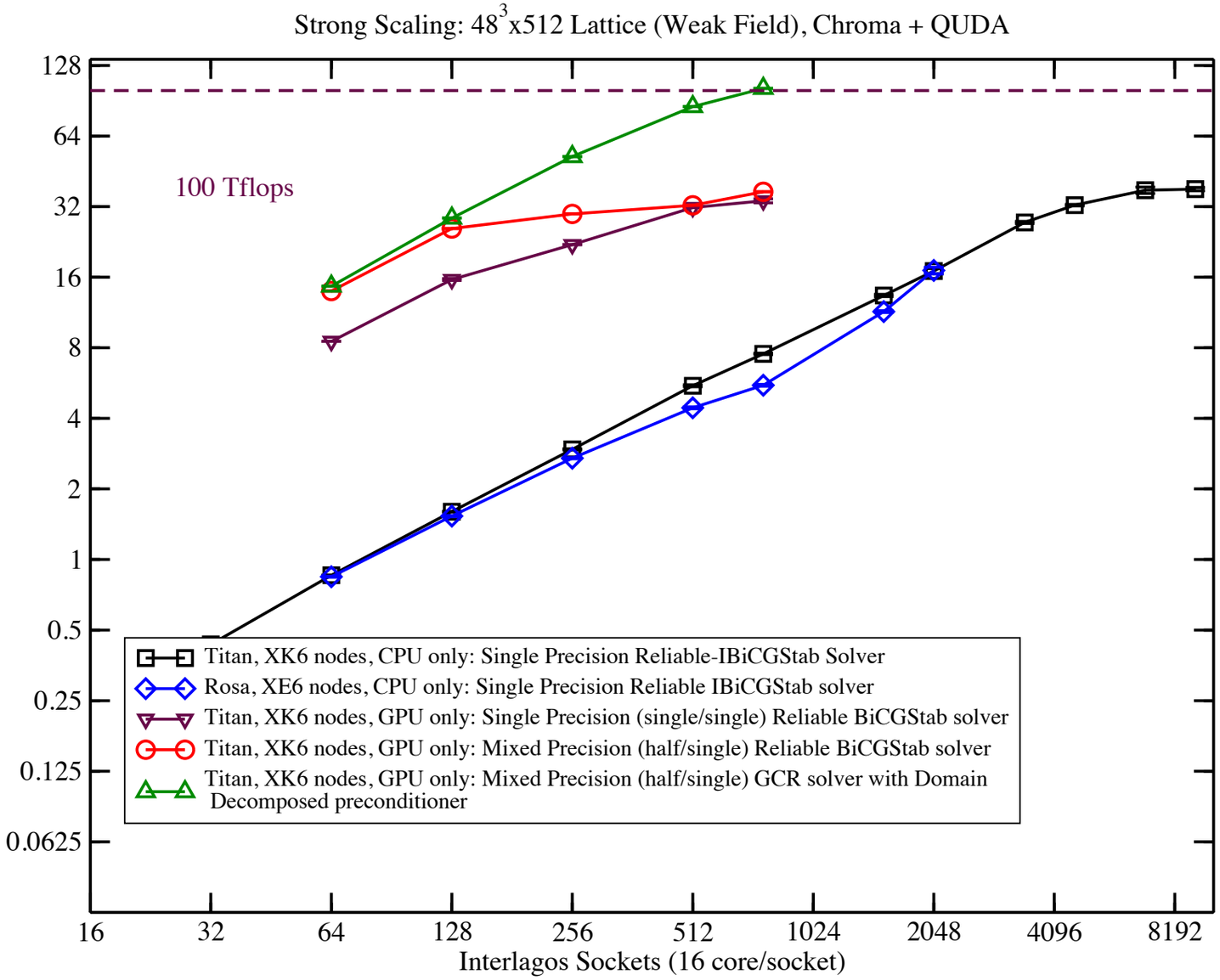


Image Courtesy of Oak Ridge Leadership Computing Facility (OLCF), Oak Ridge National Laboratory

Our work with strong scaling targets the newly installed Cray XK7 Titan System at Oak Ridge Leadership Computing Facility (OLCF) (pictured above) and other large scale GPU based systems such as NCSA BlueWaters, Keeneland and others

- A domain decomposed preconditioner combined with a GCR solver
  - reduced communication needs in the Linear Solver
  - strong scaled to 768 nodes on the TitanDev Cray XK6 system (Fermi Tesla GPUs) at the OLCF

R. Babich, M. Clark, B. Joo, G. Shi, R. Brower, S. Gottlieb, SC'11, Seattle

Intel MIC (Xeon Phi)

- ✦ Unfortunately, this work is still under NDA
- ✦ Chroma work has been going on at Jlab where there is some hardware
- ✦ I have a grad student working on BEACON project at NICS
  - I have been planning to get involved with the benchmarking and porting, but have been traveling too much!
  - Recently introduced to the Intel staff member who has been working with Joo on Chroma port
- ✦ Frankly, I am not that enamored of having to deal with three levels of parallelism
- ✦ Wish I could tell you more...

Future

- GPUs have proven value for analysis jobs
- There is much development to be done on the code
  - multigrid solvers
  - additional work on domain decomposition to improve scaling
  - persistent objects to reduce data transfer between host & GPU
  - GPUDirect testing and optimization
  - Kepler optimization
  - vectorization on Xeon Phi
- Need to see how well we can do configuration generation on large volumes with various actions
  - Currently useful for finite temperature calculations
- Need more direct comparisons of cost-effectiveness
- We live in interesting times