Lattice QCD experiences on TitanDEV

Bálint Joó, Jefferson Lab
Mike Clark, NVIDIA
USQCD Collaboration

Accelerating Computational Sciences Symposium

Washington D.C.

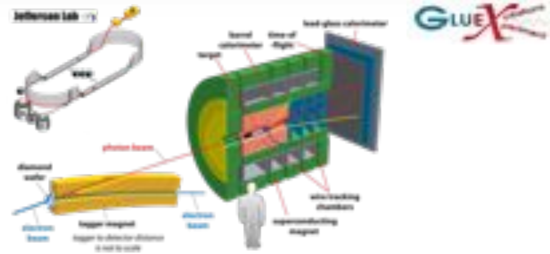
March 29, 2012

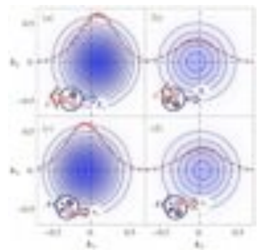
Contents

- Motivation: Lattice QCD Calculations
- LQCD on GPUs
- Accommodating the Accelerated Architecture
- Scaling Benchmarks
- Prospects for Nuclear Physics calculations on the full Titan system
  - Software
  - Physics

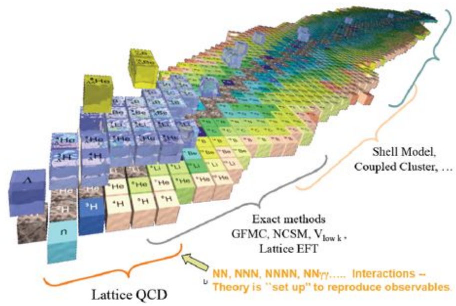
QCD In Nuclear Physics

- Can QCD predict the spectrum of hadrons ?
  - what is the role of the gluons?
  - what about exotic matter?
- How do quarks and gluons make nucleons?
  - what are the distribution of quarks, gluons, spin, etc ?
- QCD must explain nuclear interactions
  - ab initio calculations for simple systems
  - bridges to higher level effective theories
- QCD phase structure, equation of state
  - input to higher level models (e.g hydrodynamics)
  - experiments (e.g. RHIC), astrophysics (early universe)
- The USQCD Collaboration is engaged in studies of QCD in both Nuclear and High Energy Physics as well as 'Beyond the Standard Model' physics using the methodology of lattice gauge theory.



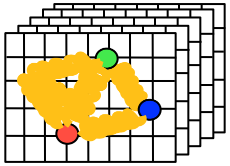


Hägler, Musch, Negele, Schäfer, EPL ***88*** *61001*



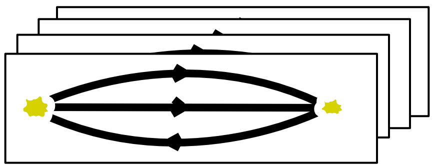
Large Scale LQCD Simulations Today

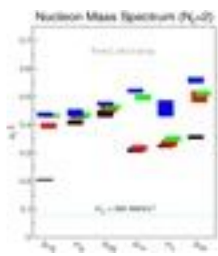
- Stage 1: Generate Configurations
  - snapshots of QCD vacuum
  - configurations generated in sequence
  - capability computing needed for large lattices and light quarks



- Stage 2a: Compute quark propagators
  - task parallelizable (per configuration)
  - capacity workload (but can also use capability h/w)

- Stage 2b: Contract propagators into Correlation Functions
  - determines the physics you'll see
  - complicated multi-index tensor contractions





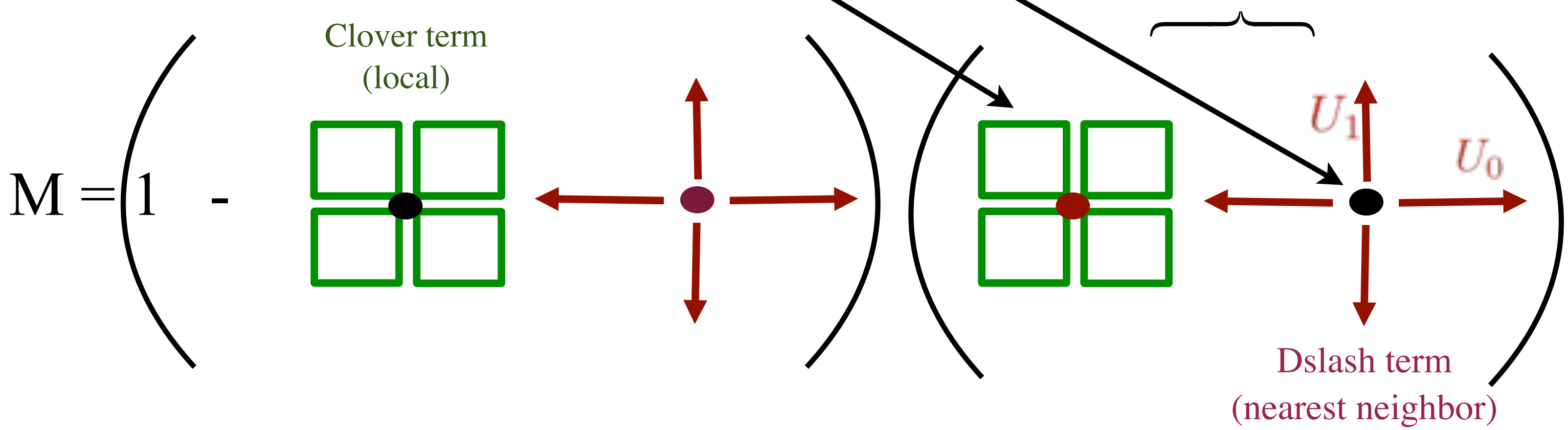
- Stage 3: Extract Physics
  - on workstations, small cluster partitions

The Wilson-Clover Fermion Matrix

After even-odd (red-black) preconditioning (Schur style):

$$M = 1 - A_{oo}^{-1} D_{oe} A_{ee}^{-1} D_{eo}$$

total: 1824 flops,
408 words in + 24 words out
FLOP/Byte: 1.06 (SP), 0.53 (DP)



SU(3) matrix

permutes spin components, flips signs

'get nearest neighbour' from forward μ direction

$$D_{x,y} = \frac{1}{2} \sum_{\mu=0}^{4} U_\mu(x) \otimes (1 - \gamma_\mu) \otimes \delta_{x+\hat{\mu},y} + U_\mu^\dagger(x - \hat{\mu}) \otimes (1 + \gamma_\mu) \otimes \delta_{x-\hat{\mu},y}$$

Enter QUDA

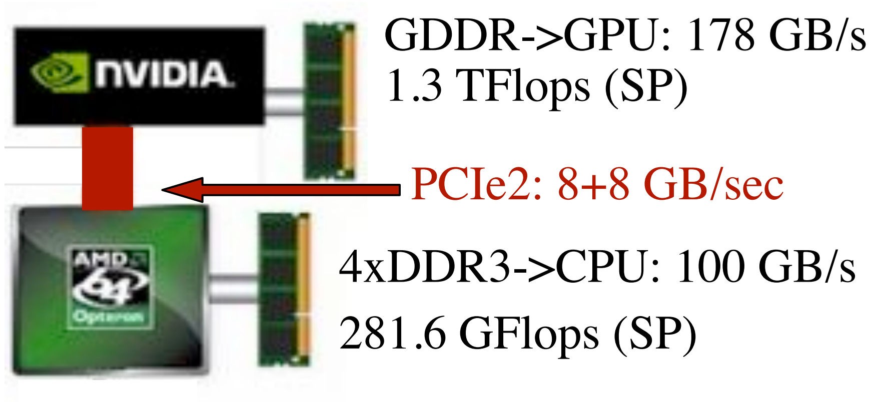
- QUDA is a library of solvers for lattice QCD on CUDA GPUs
  - *Clark, et. al., Comp. Phys. Commun. 181:1517-1528, 2010*
  - Supports: Wilson-Clover, Improved Staggered fermions
  - Domain Wall fermion support is ‘in development’
  - ‘Standard’ Krylov Solvers for QCD: CG(NE), BiCGStab
- Key Optimizations
  - Memory Coalescing Friendly Data Layout
  - Memory Bandwidth reducing ‘tricks’
    - Mixed Precision (16 bit, 32 bit, 64 bit) solvers
    - Field Compression
    - Dirac Basis (save loading half of t-neighbours)
    - Solve in Axial Gauge (save loading t-links)

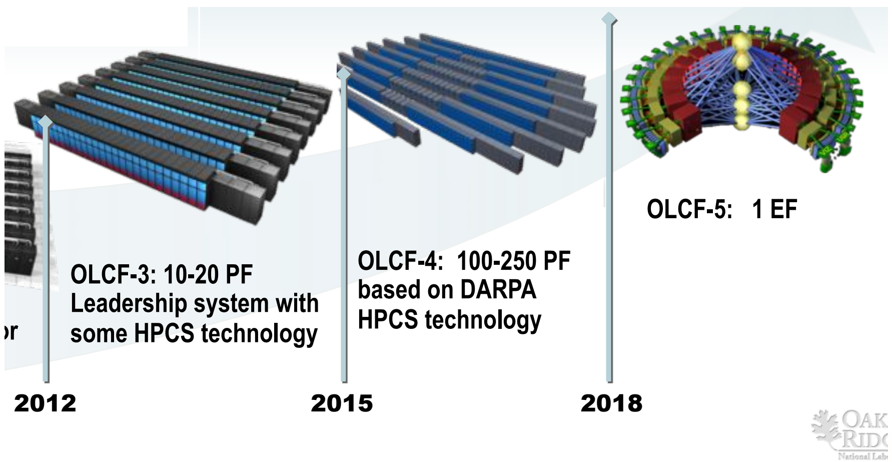
QUDA Community

- Integrated with Application Codes - enlarge user base
  - Chroma & MILC
- A group of interested developers coalesced around QUDA
  - Mike Clark (NVIDIA), Ron Babich (NVIDIA) - QUDA leads
  - Bálint Joó (Jefferson Lab) - Chroma integration
  - Guochun Shi (NCSA), Justin Foley (U. Utah) - MILC integration
  - Will Detmold, Joel Giedt, Alexei Strelchenko, Frank Winter, Chris Schroeder, Rich Brower, Steve Gottlieb
- Source Code Openly available from GitHub
  - http://github.com/lattice/quda

What about Scalability?

- Accelerated Capability Machines are ~~on the way~~ here:
  - TitanDev@OLCF
  - BlueWaters
  - Edge
  - Keeneland
  - Tianhe 1-A...
- PCIe Gen 2: B/W bottleneck
- Can QCD Scale to a really large number of GPUs ?





Domain Decomposed GCR Algorithm

Apply Preconditioner: reduced precision inner solve **LOCAL**

(Re)Start

$r_0 = b - Mx$

$\hat{r}_0 = r_0$

$\hat{x}_0 = 0$

$k = 0$

Generate Subspace

$\hat{p}_k = \hat{K}^{-1}\hat{r}_k$

$\hat{z}_k = \hat{M}\hat{p}_k$

Reduced Precision M v

$\beta_{i,k} = (\hat{z}_i, \hat{z}_k)$

Orthogonalize $\hat{z}$-s

$\gamma_k = ||\hat{z}_k||$

normalize $\hat{z}_k$

$\alpha_k = (\hat{z}_k, \hat{r}_k)$

$\hat{r}_{k+1} = \hat{r}_k - \alpha_k \hat{z}_k$

$k = k + 1$

repeat for all k or until residuum drops enough or convergence

Update Solution

Solve for χ_l l=k,k-1,...,0:

$$\gamma_l \chi_l \sum_{i=l+1}^{k} \beta_{l,i}\chi_l = \alpha_l$$

Compute correction for x:

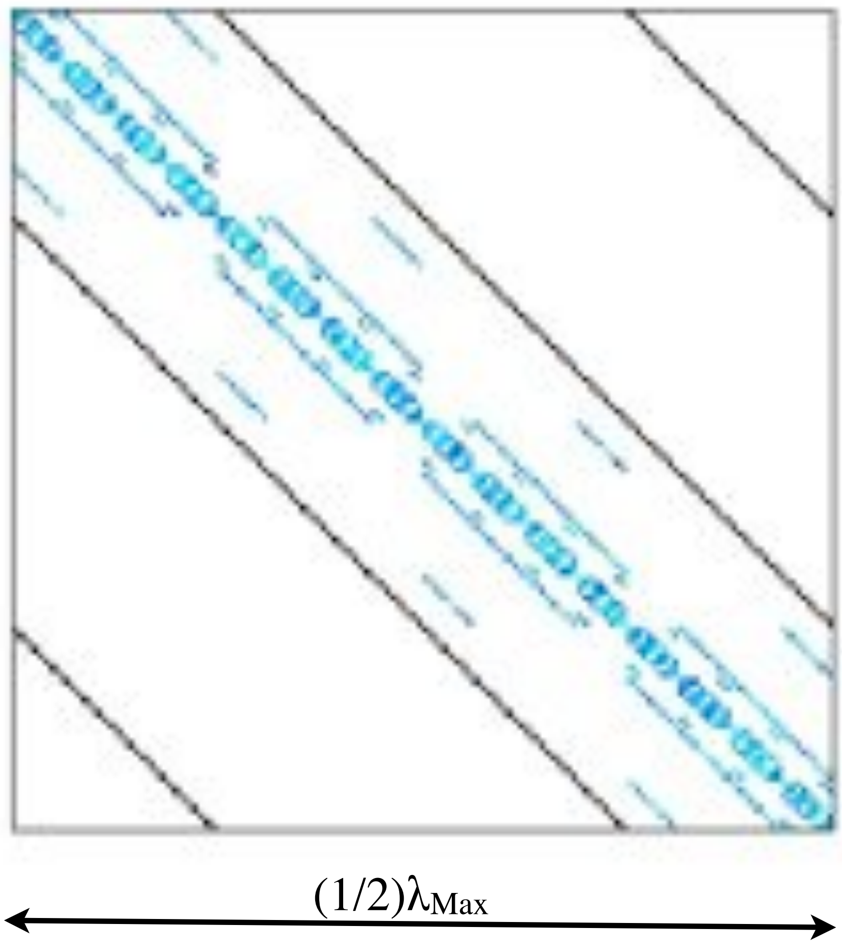
$$\hat{x} = \sum_{i=0}^{k-1} \chi_i p_i$$

$$x = x + \hat{x}$$

Full precision restart if not converged

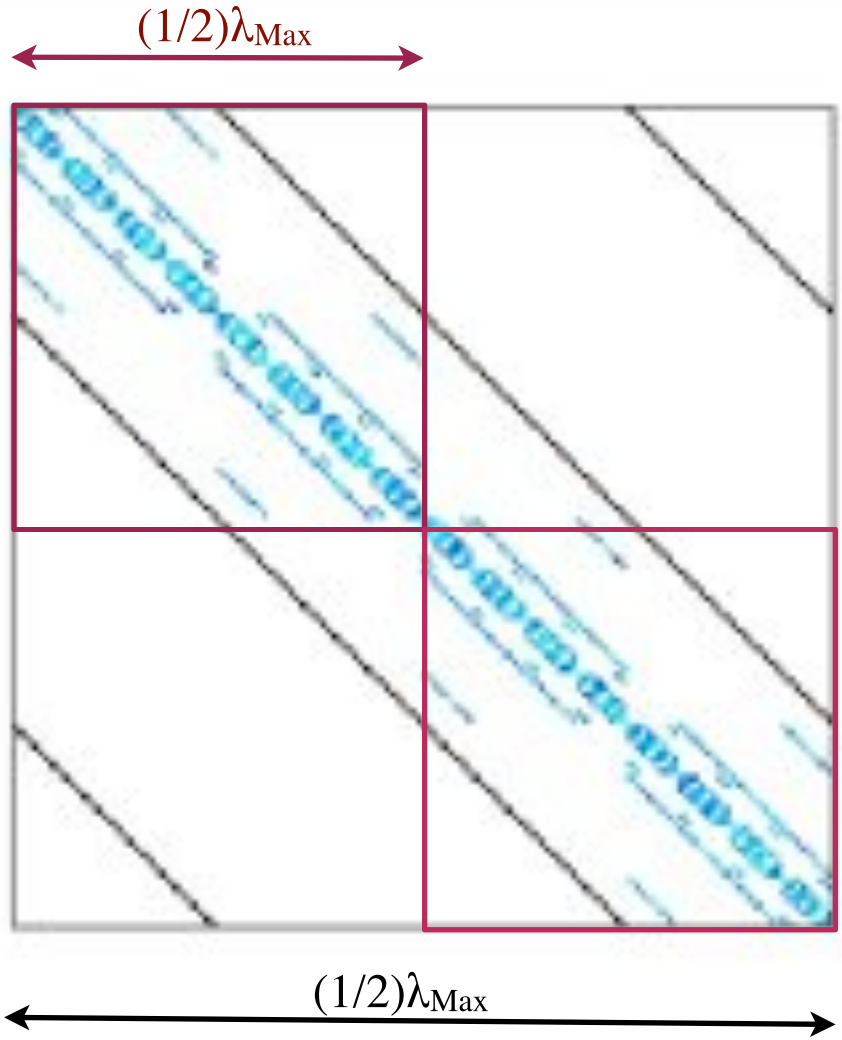
Quantities with ^ are in reduced precision

Size Matters



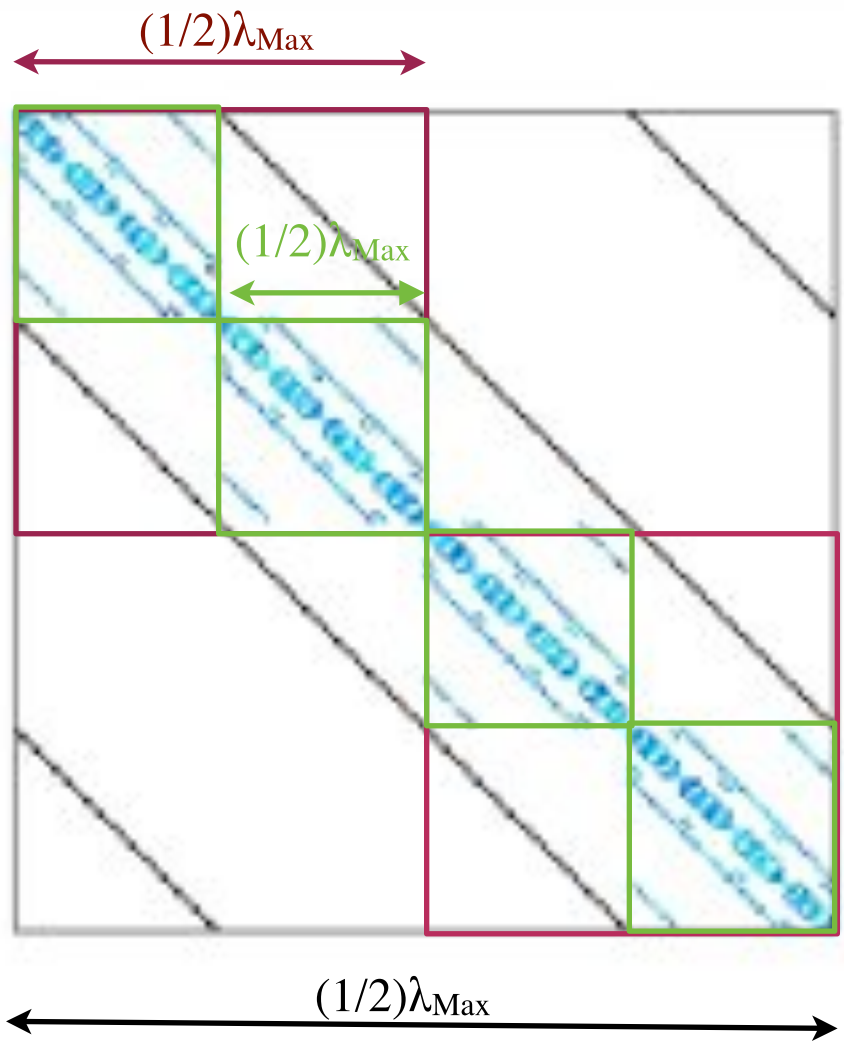
- *No communications* between domains
  - Block Diagonal Preconditioner
- Blocks impose λ cutoff
- Finer Blocks
  - lose long wavelength/low energy modes
- Heuristically (& from Lüscher)
  - keep wavelengths of ~ $O(\Lambda_{QCD}^{-1})$
  - $\Lambda_{QCD}^{-1} \sim 1\text{fm}$
  - Anisotropic Grid: (a_s=0.125fm, a_t=0.035fm)
    - Our case: 8^3x32 blocks are ideal
  - 48^3x512 lattice:
    - 8^3x32 blocks: 3456 GPUs
    - 8^3x16 blocks: 6912 GPUs (at a pinch)
  - 64^3x512 lattice:
    - 8^3x32 blocks: 8192 GPUs
    - 8^3x16 blocks: 16384 GPUs (at a pinch)

Size Matters



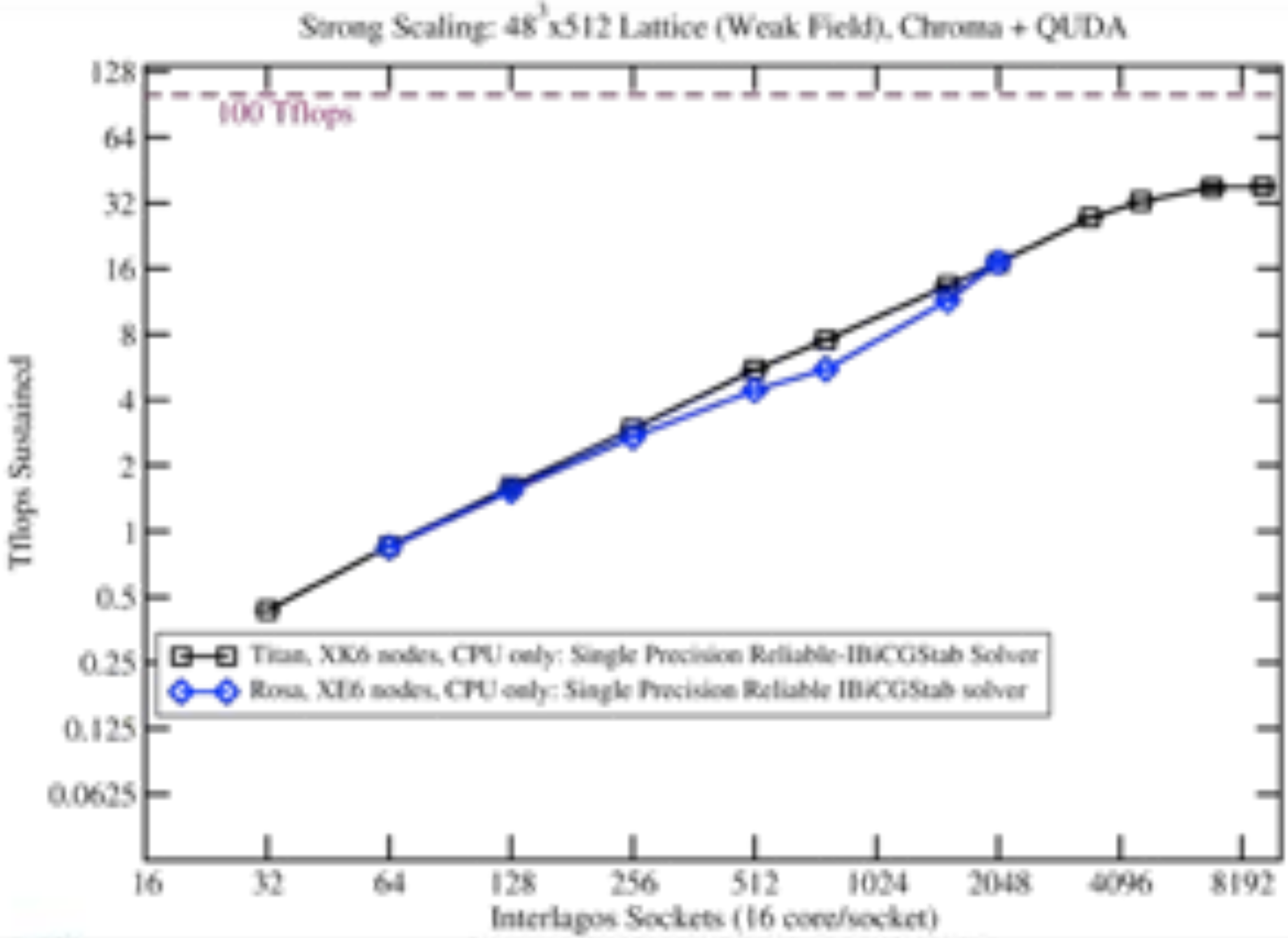
- *No communications* between domains
  - Block Diagonal Preconditioner
- Blocks impose λ cutoff
- Finer Blocks
  - lose long wavelength/low energy modes
- Heuristically (& from Lüscher)
  - keep wavelengths of ~ $O(\Lambda_{QCD}^{-1})$
  - $\Lambda_{QCD}^{-1} \sim 1\text{fm}$
  - Anisotropic Grid: (a_s=0.125fm, a_t=0.035fm)
    - Our case: 8^3x32 blocks are ideal
  - 48^3x512 lattice:
    - 8^3x32 blocks: 3456 GPUs
    - 8^3x16 blocks: 6912 GPUs (at a pinch)
  - 64^3x512 lattice:
    - 8^3x32 blocks: 8192 GPUs
    - 8^3x16 blocks: 16384 GPUs (at a pinch)

Size Matters

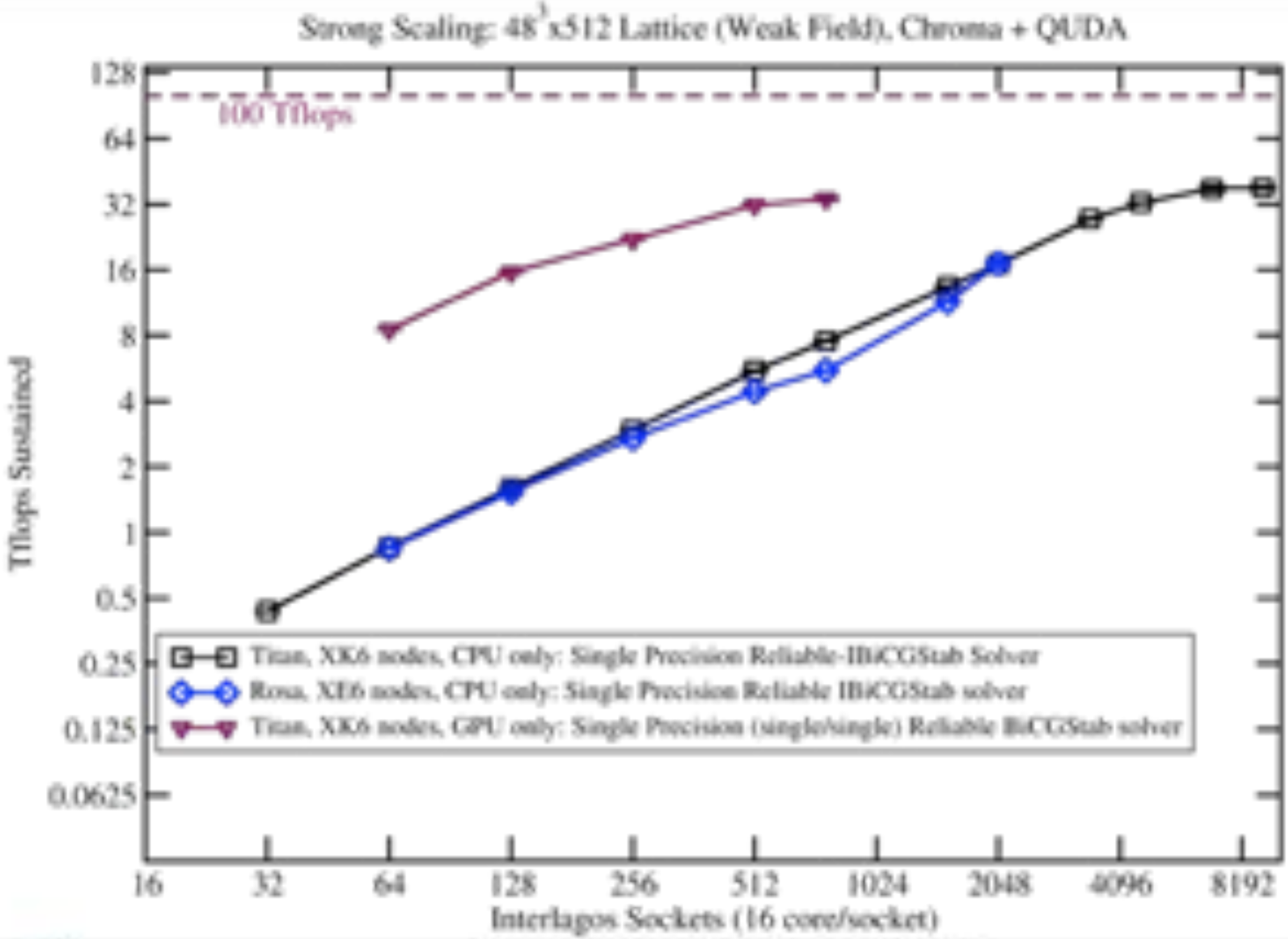


- ***No communications*** between domains
  - Block Diagonal Preconditioner
- Blocks impose λ cutoff
- Finer Blocks
  - lose long wavelength/low energy modes
- Heuristically (& from Lüscher)
  - keep wavelengths of ~ $O(\Lambda_{QCD}^{-1})$
  - $\Lambda_{QCD}^{-1} \sim 1\text{fm}$
  - Anisotropic Grid: (a_s=0.125fm, a_t=0.035fm)
    - Our case: 8^3x32 blocks are ideal
  - 48^3x512 lattice:
    - 8^3x32 blocks: 3456 GPUs
    - 8^3x16 blocks: 6912 GPUs (at a pinch)
  - 64^3x512 lattice:
    - 8^3x32 blocks: 8192 GPUs
    - 8^3x16 blocks: 16384 GPUs (at a pinch)

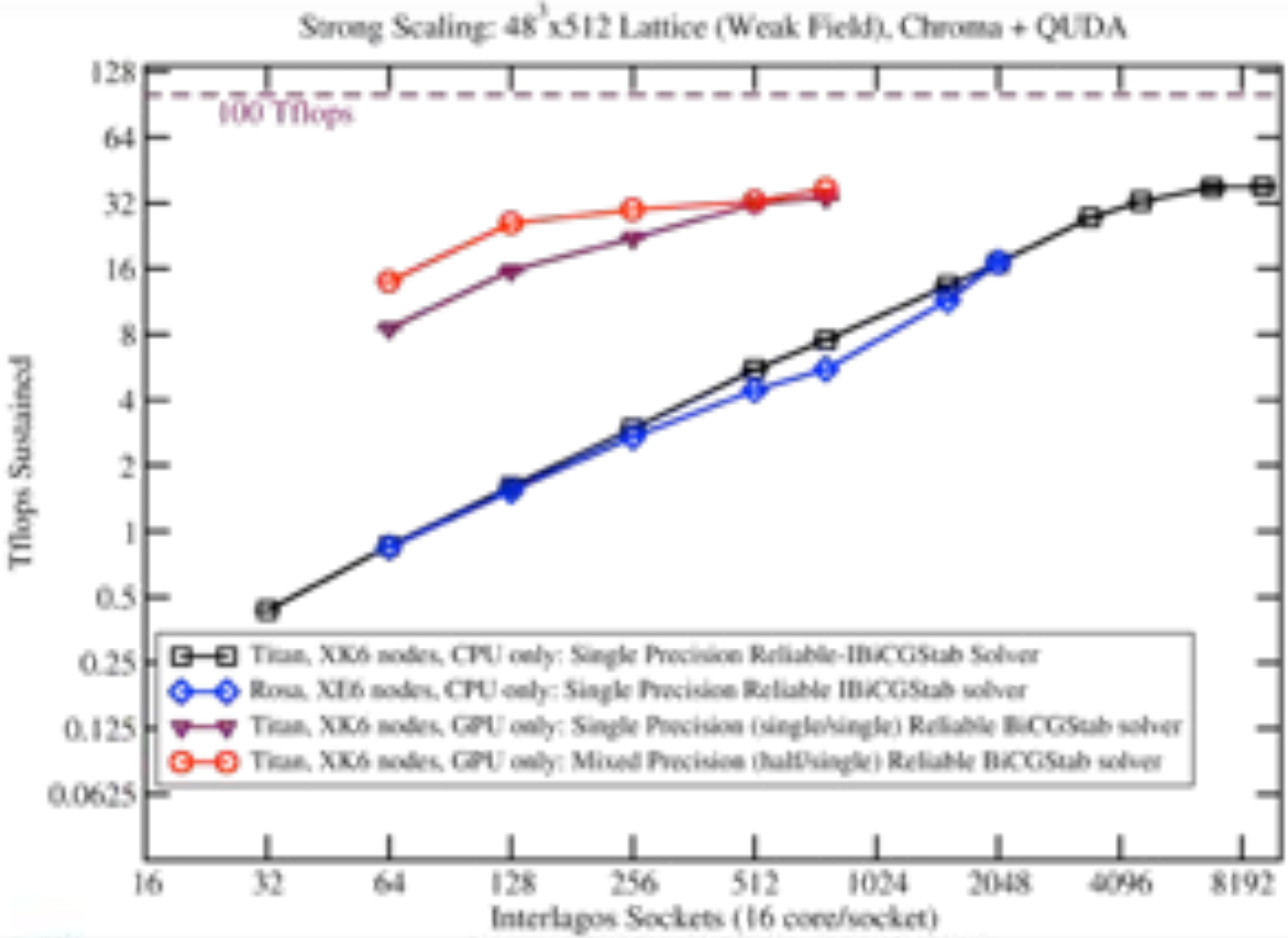
Benchmarks: CPU only, BiCGStab



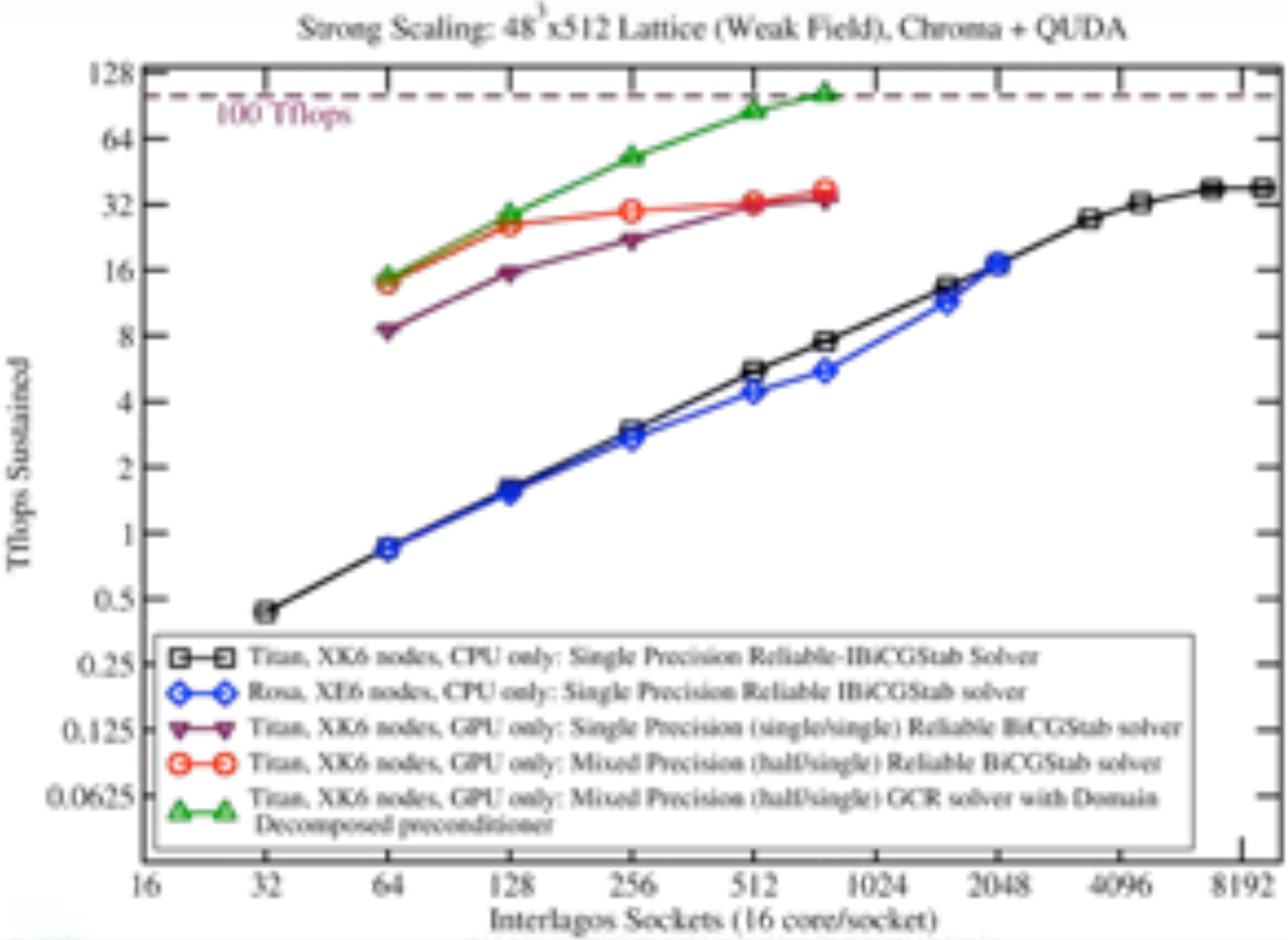
Benchmarks: + GPU (BiCGStab)



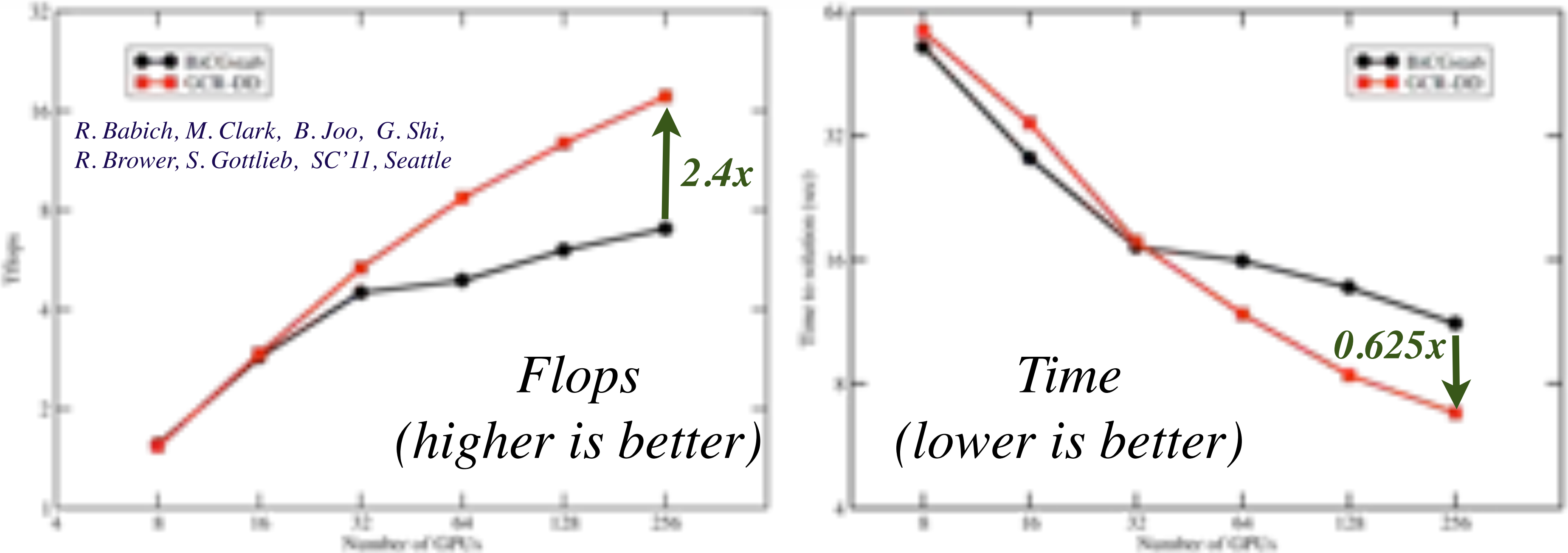
Benchmarks: + GPU (BiCGStab)



Benchmarks: + GPU (DD+GCR)



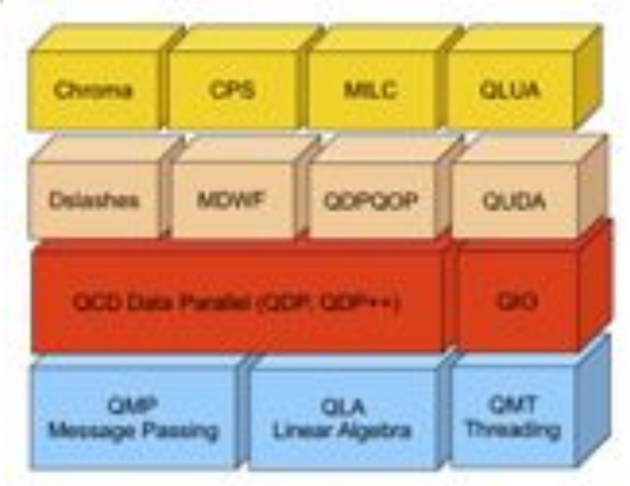
Algorithmic differences



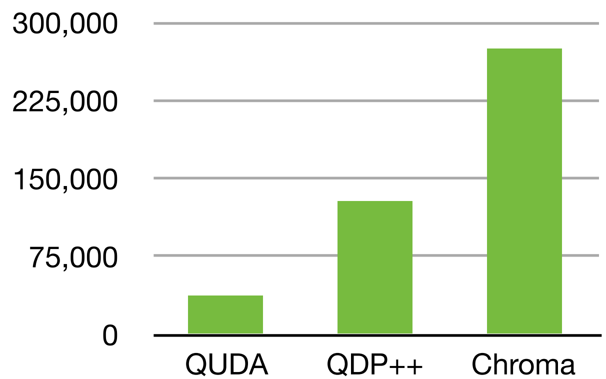
- SC'11 result (Edge Cluster, LLNL), 32^3x256 production lattices
- DD+GCR gets 2.4x BiCGStab flops, but only 1.6x gain in wall-time
  - Conservative factor: 1 DD-GCR flop ~ 1.5 BiCGStab flop
  - To Do: measure factor on 48^3x512 lattices (need thermalized cfgs)

Further Code Development

- Immediate Next Steps: Full Gauge Generation using GPU Nodes
- Step 1: GPUs for solvers only, rest on CPU
- Step 2: Port full Chroma Gauge Generation application to GPUs
  - QDP-JIT ports QDP++ to CUDA
    - Frank Winter (U. Edinburgh)
  - Convert Chroma kernels which don't use QDP++ by hand
- USQCD software development is funded through the DoE SciDAC and SciDAC-2 programs through the Office of Nuclear Physics, the Office of High Energy Physics and the Office of Advanced Scientific Computing Research.

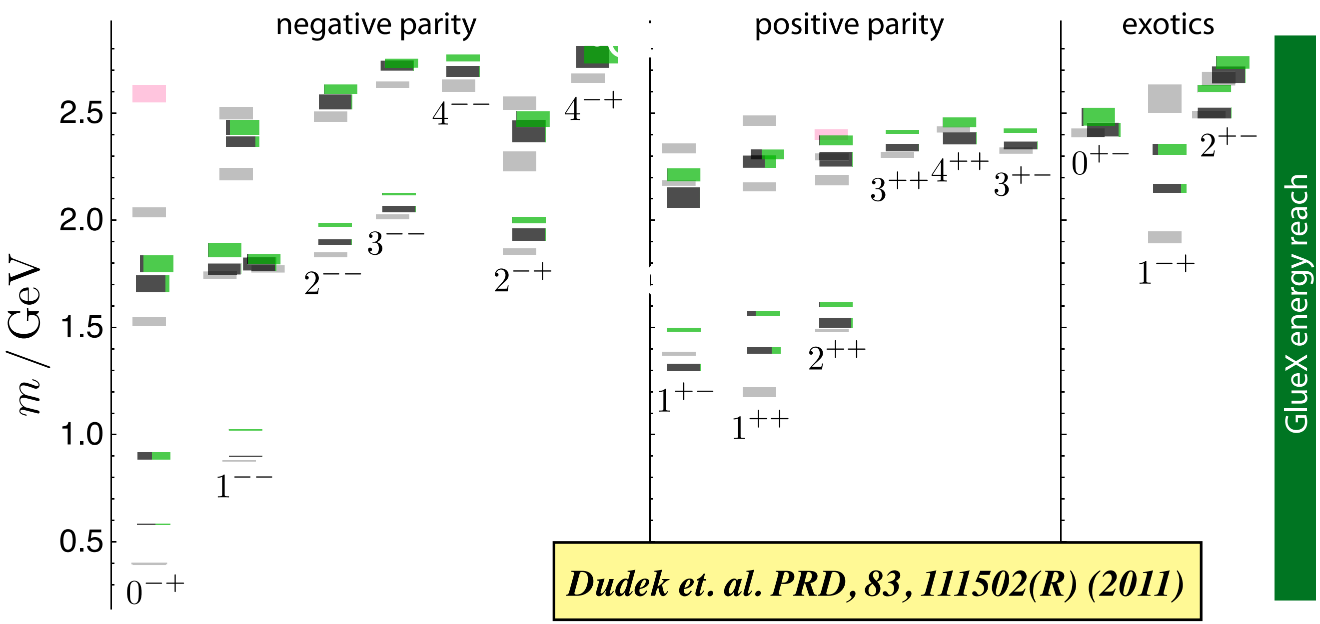


USQCD Software Stack
Developed with DoE SciDAC funding

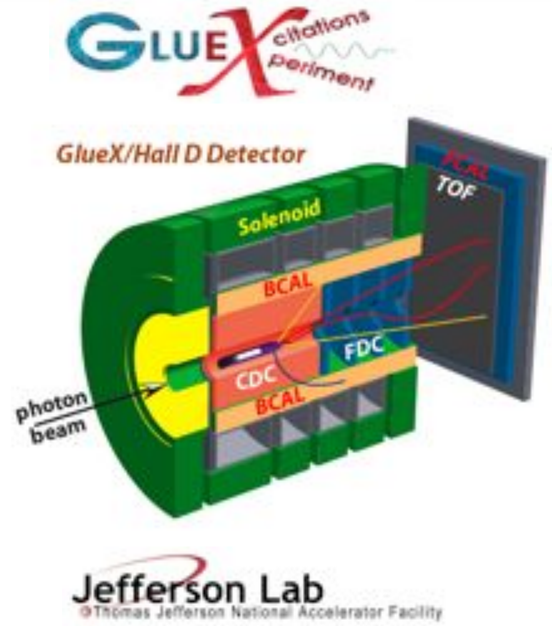


Lines of C/C++ Code per package measured on May 11, 2011, using CLOC *http://cloc.sourceforge.net/*

Prospects for Scientific Program



Dudek et. al. PRD, 83, 111502(R) (2011)



- Goal:
  - Determine if exotic meson states exist, ***before*** the Glue-X experiment (JLab@12GeV Hall-D) turns on in 2015.
  - using quarks with the physical values of the quark masses.
- Production of configurations has begun on BlueWaters ESS System

The scale of the problem

- Physical mass calculations require large lattices
  - need $m_\pi L > \sim 4.5$ for algorithm stability/avoid size effects
  - prefer $m_\pi L > \sim 10$ to suppress thermal effects
- Need multiple volumes for analyzing decays/multi particle states
- At m_π=140MeV (physical pions) we need
  - 48^3x512 for $m_\pi L \sim 4.2$, $m_\pi T \sim 12.5$ (up to 6912 GPUs)
  - 64^3x512 for $m_\pi L \sim 5.6$, $m_\pi T \sim 12.5$ (up to 16384 GPUs)
- Rough idea of cost: going from 32^3x256 with $V^{5/4}$ scaling
  - 48^3x512 : 10.9x (INCITE 2010 + 2011) from V growth alone
  - 64^3x512: 32x (INCITE 2010 + 2011) from V growth alone
  - assume cost growth with reduced quark mass is mild
- Running at scale on accelerated h/w is needed for timely progress

Conclusions

- Lattice QCD can usefully utilize 768 GPUs (over 3/4 of TitanDev)
- Waiting for more hardware: 1024 GPUs is next partition size up
- DD + GCR Solver reached 100TFlops on 768 Fermi GPUs
  - expect further speed up from Kepler
- We are close to running full gauge generation on accelerated nodes
  - Initially using just the solvers from QUDA
  - Ultimately to moving the whole application to the GPU
- Estimate being able to scale to 3456-6912 GPUs using our current global volume of 48^3x512 sites, and the DD+GCR solver, maybe more on bigger lattices.
- Using such a machine, given sufficient resource, we aim to compute the properties of exotic mesons prior to their planned production in Glue-X as part of the 12GeV upgrade

Acknowledgements

- We'd like to thank
  - OLCF for access to TitanDev
  - CSCS for access to Monte Rosa
- B. Joó gratefully acknowledges funding through the following grants:
  - SciDAC-2 program "USQCD: The Secret Life of Quarks" through DOE Grants:
    - DE-FC02-06ER41440,
    - DE-FC02-06ER41449
  - DE-AC05-060R23177 under which Jefferson Science Associates LLC manages and operates the Jefferson Lab.