Twisted-Mass Lattice QCD using OpenCL
An Update

Matthias Bach

Frankfurt Institute for Advanced Studies
Institut für Informatik
in collaboration with: C. Pinke, O. Philipsen, V. Lindenstruth

31st International Symposium on Lattice Field Theory
Mainz, 30 July 2013
LOEWE-CSC

- Frankfurt University
- 786 GPU Nodes
- Node:
  - 2 AMD Magny-Cours
  - AMD Radeon HD 5870
- QDR Infiniband
- 285 TFLOPS (#22)
- 740.78 MFLOPS/W
- Green500 #8 Nov 10
Sanam

- Coop:
  - FIAS (Frankfurt)
  - KACST (Saudi-Arabia)
- 300 Nodes
- Node:
  - 2 Intel Xeon E5-2650
  - 2 AMD FirePro S10000
- FDR Infiniband
- 532 TFLOPS (#52)
- 2351 MFLOPS/W
- Green500 #2 Nov 12
OpenCL and CL²QCD

OpenCL
- Open Standard
- Like CUDA Driver API
  - Kernels
  - Threads
  - C-API
- Wide hardware support
  - CPUs
  - AMD GPUs
  - NVIDIA GPUs
  - Xeon Phi
  - Mobile Devices

CL²QCD
- Twisted Mass / Pure Wilson
- All calculations in OpenCL
- Layered
  - Hardware
  - Kernels
  - QCD-Types
  - Algorithms
- Separation of Concerns
  - Hardware Specifics
  - Optimization
  - Application Logic

Matthias Bach
Twisted-Mass Lattice QCD using OpenCL
Double-Precision $\mathcal{D}$ Performance

![Graph showing performance comparison between various GPUs for different lattice sizes.]

CG Performance – Weak Scaling

![Graph showing CG Performance – Weak Scaling with constant lattice size per GPU for two different configurations: one with a lattice size of $32^3 \times 16$ and another with $24^3 \times 32$. The graph plots GFLOPS against GPUs, showing linear scaling as the number of GPUs increases.]
CG Performance – Hard Scaling

Varying lattice size per GPU

Matthias Bach
Twisted-Mass Lattice QCD using OpenCL
Average Power Consumption

**CPU System:**
- 2 AMD Bulldozer
  8 Cores, 3 GHz

- *tmlqcd*
  

**GPU System:**
- 2 AMD Magny-Cours
  12 Cores, 2 GHz

- 1 to 2 AMD Radeon HD7970

- *CL²QCD*
Energy-Efficiency

HMC Setup:
- $32^3 \times 12$
- $N_f = 2$ Twisted-Mass Wilson Fermions
- $m_{\pi} \approx 270$ MeV

Multiple GPUs:
- 2nd GPU: Separate HMC Chain
- Projection to 8 GPUs: 3.9 to 4.25 trajectories per kWh
Total Cost of Acquisition in General Purpose Systems

HMC in Sanam:

- GPU Throughput: 16 times CPU
- GPU Cost: \(\approx 25\%\)
- GPU Cost Advantage: \(\approx 12\)

Cost for Inversions:

- Kraken (x86): 1.14 $/MFLOPS
- Sequoia (BlueGene/Q): 0.040 $/MFLOPS
- Sanam (x86 + GPU): 0.038 $/MFLOPS

Cost on Kraken and Sequoia estimated based on publications.
Total Cost of Acquisition in General Purpose Systems

HMC in Sanam:
- GPU Throughput: 16 times CPU
- GPU Cost: ≈ 25%
- GPU Cost Advantage: ≈ 12

Cost for Inversions:
- Kraken (x86) 1.14 $/MFLOPS
- Sequoia (BlueGene/Q) 0.040 $/MFLOPS
- Sanam (x86 + GPU) 0.038 $/MFLOPS

Cost on Kraken and Sequoia estimated based on publications.
Total Cost of Acquisition in Dedicated Systems

- Radeon HD 7970
- FirePro S10000
- Tesla K20
- Xeon Phi 5110P

Based on retail prices

Comparision should use HMC trajectories!
Total Cost of Acquisition in Dedicated Systems

<table>
<thead>
<tr>
<th></th>
<th>$/MFLOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radeon HD 7970</td>
<td>0.02</td>
</tr>
<tr>
<td>FirePro S10000</td>
<td>0.03</td>
</tr>
<tr>
<td>Tesla K20</td>
<td>0.02</td>
</tr>
<tr>
<td>Xeon Phi 5110P</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Based on retail prices

Comparison should use HMC trajectories!

- Utilized LOEWE-CSC and Sanam
**Outlook**

Achieved:
- 100 GFLOPS DP $\emptyset$ on AMD GPU
- Works on NVIDIA / Intel
- Full HMC on GPU
- Scaling on GPUs within Node
- Four Times as Energy-Efficient as CPU
- Cost-Efficient

In Progress / TODO:
- Staggered Fermions
- Mixed-Precision Solvers
- MPI
- NVIDIA Optimizations
- Intel Optimizations