National Software Infrastructure for Lattice Gauge Theory

Software Co-ordinating Committee

R. Brower (BU), C. DeTar (Utah), R. Edwards (Jlab), D. Holmgren (FNAL), B. Mawhinney (Columbia), W. Watson, III (JLab) and Y. Zhang (North Carolina)

(Go to `http://usqcd.org/` for link to Software & Documentation)

Project

Software Infrastructure Goals: *Create a unified software environment that will enable the US lattice community to achieve very high efficiency on diverse multi-terascale hardware: QCDOC, QCD clusters and Commercial Platforms.*

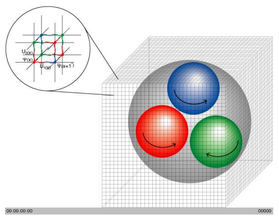
TASKS:		LIBRARIES:
I.	Message Passing API	QMP
II.	Linear Algebra API	QLA
III.	Data Parallel API	QDP
IV.	I/O Data Files and Data Grid	QIO
V.	Level 3 Opt "Inverters"	CPS/MILC/Chroma
VI.	Execution Environment	unify BNL/FNAL/JLab

Major Participants in Software Project

Arizona	Doug Toussaint	MIT	Andrew Pochinsky
	Dru Renner *(Eric Gregory)*	North Carolina	Daniel Reed *(Celso Mendes)*
BU	Rich Brower *		Ying Zhang *
	James Osborn *(Hartmut Neff)*	JLab	Robert Edwards *
BNL	Chulwoo Jung		Chip Watson *
	Kostantin Petrov		Jie Chen
Columbia	Bob Mawhinney *		Walt Akers
FNAL	Don Holmgren *	Utah	Carleton DeTar *
	Jim Simone		Francesca Maresca
	Eric Neilsen	UK	Peter Boyle
	Amitoj Singh		Balint Joo (Mike Clark)

* Software Coordinating Committee

QCD — extremely regular

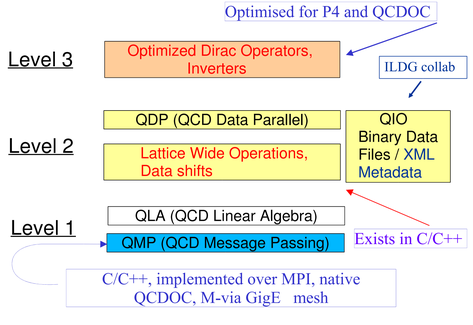


- (Perfect) Load Balancing: Uniform periodic lattice & identical sublattice per processor.
- (Complete) Latency hiding:) overlap computation/communication.
- Data Parallel: operations on small 3x3 complex matrices per link.
- Critical kernl: Dirac Solver is ∼ 90% of Flop/s.

Lattice Operator:

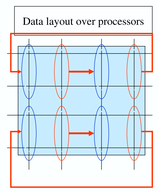
$$[D\Psi]^i_{\alpha} = \frac{1}{2a}\sum_{x,\mu}\gamma^{\alpha\beta}_{\mu}[U^{ij}(x)\Psi^j_{\beta}(x+\hat{\mu}) - \text{h.c.}]\cdots$$

SciDAC QCD API



Level 1 Message Passing and Algebra

QMP: Overlapping communications and computations



- C(x)=A(x) * shift(B, + mu):
  1. Send face forward non-blocking to neighboring node.
  2. Receive face into pre-allocated buffer.
  3. Meanwhile do A*B on interior sites.
  4. "Wait" on receive to perform A*B on the face.
- Lazy Evaluation (C style):
  1. Shift(tmp, B, + mu);
  2. Mult(C, A, tmp);

Level 1:Linear Algebra Implementations

```
// Lattice operation
A = adj(B) + 2 * C;

// Lattice temporaries
t1 = 2 * C;
t2 = adj(B);
t3 = t2 + t1;
A = t3;

// Merged Lattice loop
for (i = ... ; ... ; ...) {
   A[i] = adj(B[i]) + 2 * C[i];
}
```

- C implementation 24,000 functions generated with Perl and test scripts
- C++ Operators overloading and eliminate lattice temps - PETE
- Allows further combining of operations (*adj(B)*C*)
- Assemble code in SSE2 for Pentium, using BEGEL for QCDOC et al.

Level 2 Data Parallel API

Data-parallel Operations

- *Unary and binary:* -a; a-b; …
- *Unary functions:* adj(a), cos(a), sin(a), …
- *Random numbers:* // platform independent random(a), gaussian(a)
- *Comparisons (booleans)* a <= b, …
- *Broadcasts:* a = 0, …
- *Reductions:* sum(a), …

- Fields have various types (indices):

Lattice: A(x), Color: $U^{ij}(x)$, Spin: $\Gamma_{\alpha\beta}$, $\psi^i_\alpha(x)$, $Q^{ij}_{\alpha\beta}(x)$

Example of QDP++ Expression

Typical for Dirac Operator:

$$\psi^i_\alpha(x) = U^{ij}_\mu(x)\chi^j_\alpha(x+\mu)+2\phi^i_\alpha(x) \quad \forall\, i,\alpha,x \in \text{even}$$

- QDP/C++ code:

```
multi1d<LatticeColorMatrix> u[Nd] ;
LaticeDiracFermion psi, chi, phi ;
int mu;
psi[even] = u[mu]*shift[chi,mu] + 2*phi ;
```

- Portable Expression Template Engine (PETE)

Temporaries eliminated, expressions optimized

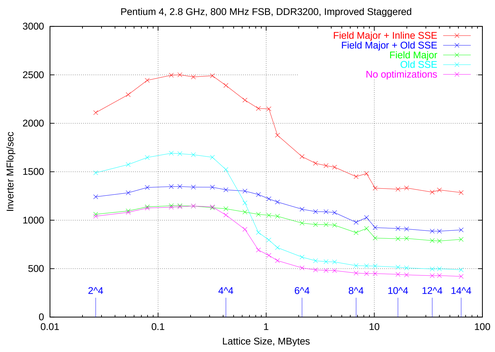
Level 3 Performance

Level 3 on QCDOC

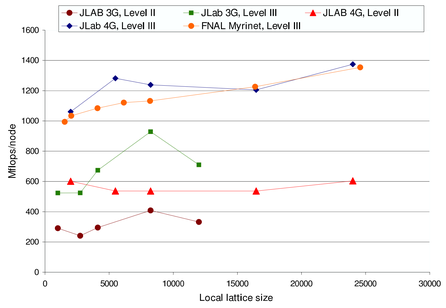
Quark Action	Local Volume	Mflop/s per node
Wilson	2^4	32%
Wilson	4^4	38%
Wilson–Clover	2^4	32%
Wilson–Clover	4^4	47.5%
Domain Wall	2^4	32%
Domain Wall	4^4	42%
Asqtad	2^4	19%
Asqtad	4^4	42%

Performance of the level 3 QCDOC assembly code for four Dirac CG inverters as a percentage of the peak rate of 840 Mflop/s running at 420 MHz on 2^4 and 4^4 sublattices.

Asqtad Fermion Inverter



Domain Wall Fermion Inverter



I/0 for International Data Grid

Data Archives and Data Grid

- File formats and header
  - Build ILDG on successful NERSC QCD archive
  - Extend to include lattice sets, propagators, etc.
- XML for ascii headers
  - Control I/O for data files
  - Search user data using SQL to find locations.
- Lattice Portal
  - Replicate data (multi-site), global tree structure.
  - SQL-like data base for storing data and retrieving
- Web based computing
  - batch system and uniform scripting tool.

QCDOC at BNL

Pentium 4 Infiniban Cluster at FNAL

Pentium 4 GigE Cluster at JLab

Future Directions

- **Integration:**
  Optimization, Testing and Hardening of components.
  New Level 3 and application routines.
  Executables for BG/L, X1E, XT3, Power 3, PNNL cluster, APEmille…
- **Algorithm Research:**
  Increased performance (leverage TOPS & NSF IRT).
  New physics (Scattering, resonances, SUSY, Fermion signs, etc?)
- **Data Grid and Meta Facility:**
  Data, Code, Job control as 3 Lab Metafacility (leverage PPDG)
  Uniform Interfaces, Uniform Execution Environment for 3 Labs.
  Middleware for data sharing with ILDG.