# The QCDOC software: Benchmarks, User Environment

Chulwoo Jung, Robert Mawhinney

(Brookhaven National Laboratory/Columbia University)

For QCDOC collaboration

Mar 26th 2004, Brookhaven National Laboratory

- Programming Environment
- Operating System (QOS)
- Performance
- Future Activities

# Programming Environment

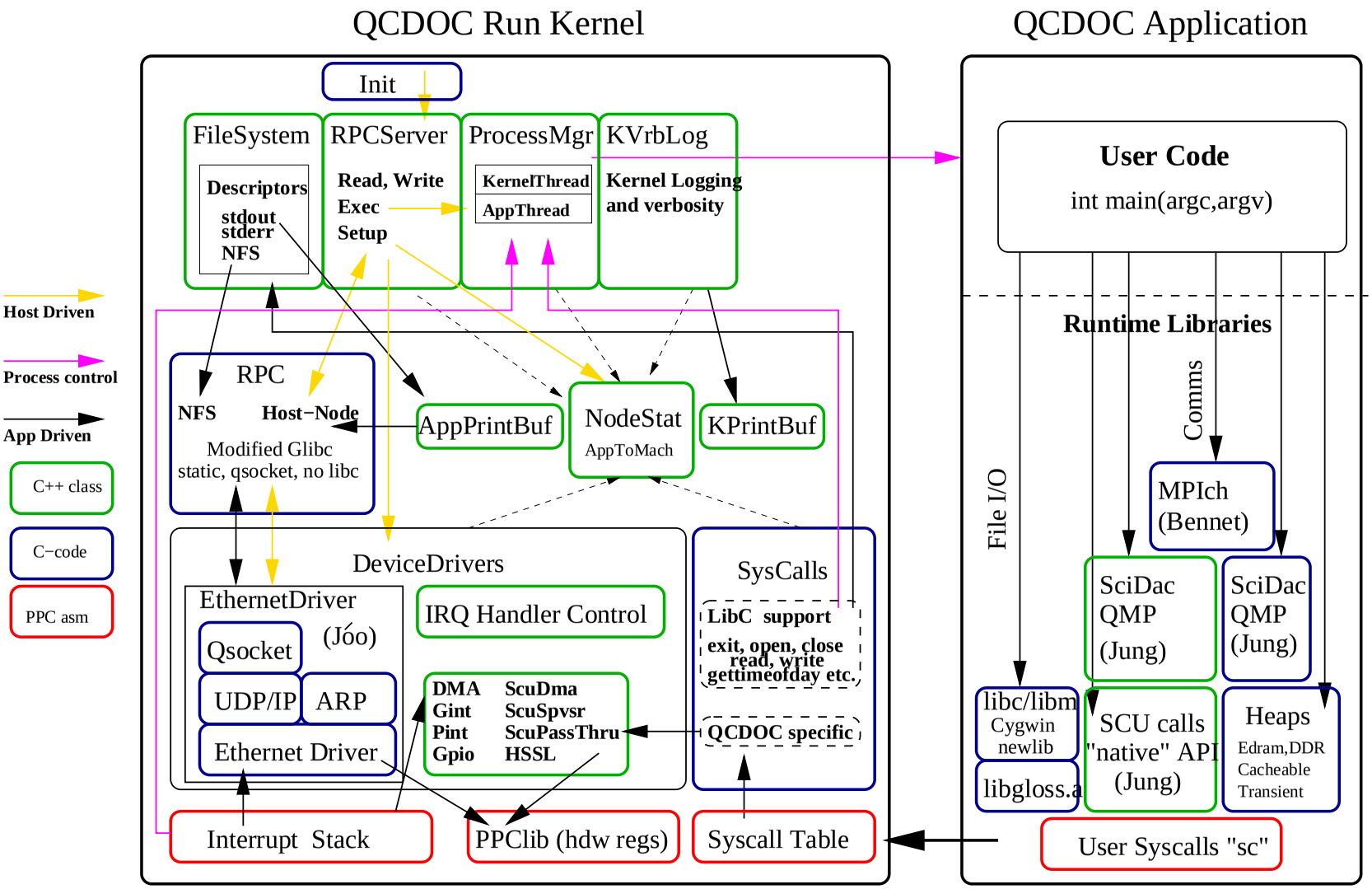
- QOS - Custom-designed Operating system (P. Boyle)
  - Boot (Ethernet/JTAG)
  - Load/Run Apps (Ethernet)
  - Hardware diagnostics
  - Command line argument passing
  - File I/O - NFS heavily tested (currently to the front end)
  - Shell support (qcsh : tcsh + QCDOC-specific commands) (K. Petrov/ P. Boyle)
  - Profiling, Source level debugging
- Standard GNU tools (Compiler, Linker, Binutils...) at the front end
- Commercial PowerPC tools
  - High-performance compilers - XLC
  - Multi-processor debugger - Riscwatch

- SciDAC software
  - QMP - Optimized Message Passing
  - MILC, QDP/C , Chroma, QDP/C++ compiled and tested with QMP
  - Hardware-optimized routines (Level 3 QCD-API)
    * Integrated in CPS
    * Being tested with MILC, Chroma

# Operating system

- Features:
  - Janus OS - one face to the machine, another to the Unix SMP front-end
  - POSIX Unix SMP front end - linux, solaris, (possibly AIX)
  - Software partitioning
  - Communication remapping from 6 dimensional sublattices to smaller dimesionalities - transparent to users
  - Serve several users and tens of thousands processors
  - Aggressive multi-threading on host side
- Qdaemon - Interface to the QCDOC back-end
  - RPC over multiple gigE links
  - PBS or Qcsh
- Boot Kernel
  - Controlled boot sequence via Ethernet/JTAG, Ethernet

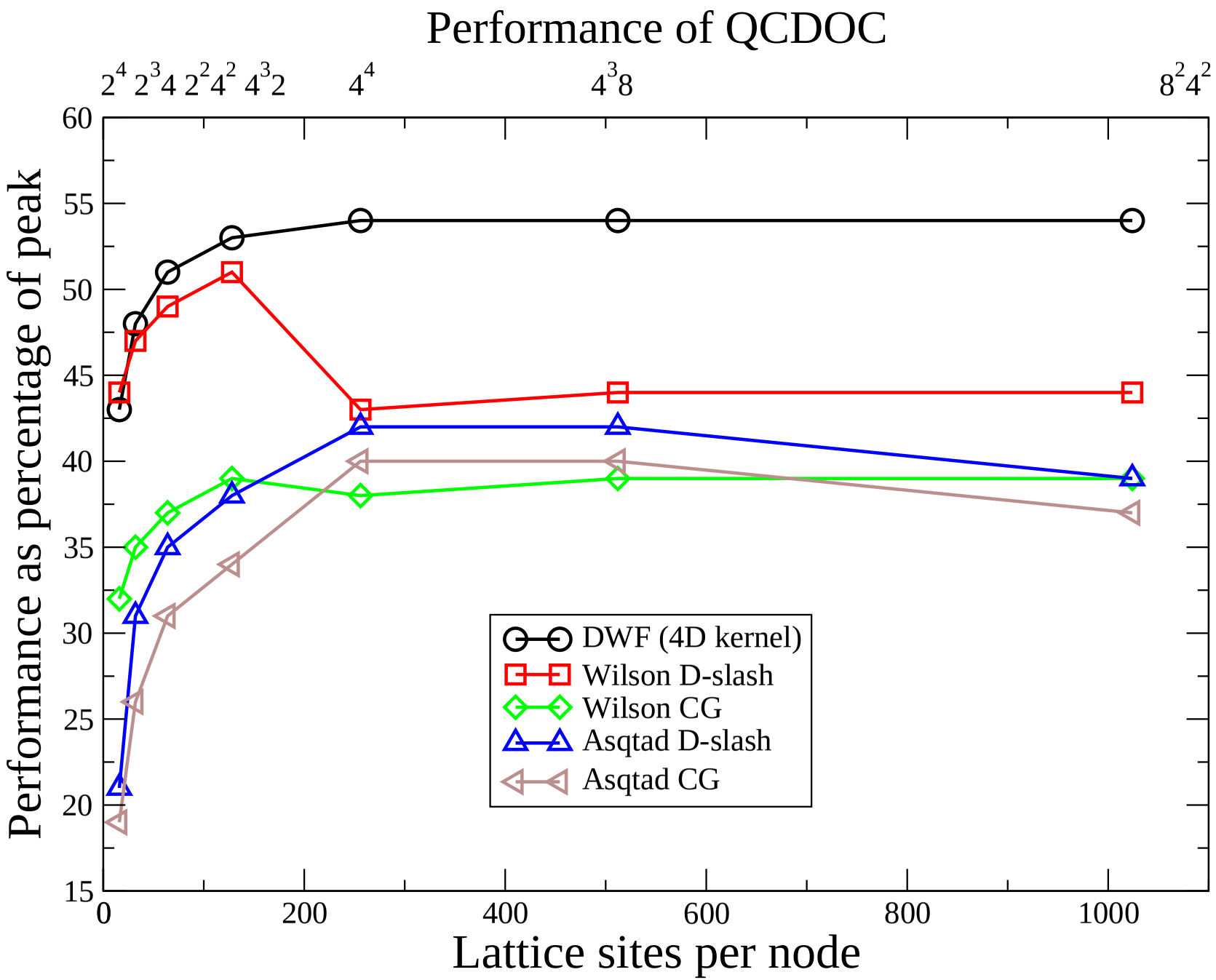
- **Run kernel:**
  - Run one process and run it *well*
  - Memory protection but not translation
  - No scheduling
  - Circular print buffers per each node
  - Service hardware/software interrupts
  - Service communications and I/O requests via System Calls - high degree of separation between user app and run kernel


QCDOC Run Kernel
QCDOC Application
Init
FileSystem
Descriptors
stdout
stderr
NFS
RPCServer
Read, Write
Exec
Setup
ProcessMgr
KernelThread
AppThread
KVrbLog
Kernel Logging
and verbosity
RPC
NFS
Host−Node
Modified Glibc
static, qsocket, no libc
AppPrintBuf
NodeStat
AppToMach
KPrintBuf
DeviceDrivers
EthernetDriver
(Jóo)
Qsocket
UDP/IP
ARP
Ethernet Driver
IRQ Handler Control
DMA
Gint
Pint
Gpio
ScuDma
ScuSpvsr
ScuPassThru
HSSL
SysCalls
LibC support
exit, open, close
read, write
gettimeofday etc.
QCDOC specific
Interrupt Stack
PPClib (hdw regs)
Syscall Table
Host Driven
Process control
App Driven
C++ class
C−code
PPC asm
User Code
int main(argc,argv)
Runtime Libraries
Comms
File I/O
MPIch
(Bennet)
SciDac
QMP
(Jung)
SciDac
QMP
(Jung)
libc/libm
Cygwin
newlib
libgloss.a
SCU calls
"native" API
(Jung)
Heaps
Edram,DDR
Cacheable
Transient
User Syscalls "sc"

# Performance of QCDOC

Tested maximum speed = 450Mhz, Maximum Flops = 900MFlops/s

| | Sites/node | Optimized | | |
|---|---|---|---|---|
| Application | | D | D | CG |
| Machine size | | Sim | 128 | 128 |
| Wilson | $2^4$ | 47% | 44% | 32% |
| | $4^4$ | | 43% | 38% |
| | $6^4$ | | 44% | 39% |
| | $8^4$ | | 29% | |
| Asqtad | $2^4$ | | 22% | 19% |
| | $4^4$ | 43% | 42% | 40% |
| | $6^4$ | | 35% | 33% |
| | $8^4$ | | 29% | 28% |


Performance of QCDOC
$2^4$ $2^3 4$ $2^2 4^2$ $4^3 2$
$4^4$
$4^3 8$
$8^2 4^2$
Performance as percentage of peak
60
55
50
45
40
35
30
25
20
15
0
200
400
600
800
1000
Lattice sites per node
DWF (4D kernel)
Wilson D-slash
Wilson CG
Asqtad D-slash
Asqtad CG

# Performance of QCDOC(extended)

Tested maximum speed = 450Mhz, Maximum Flops = 900MFlops/s

| | Sites/node | Optimized | | | MILC | | | MILC/QDP |
|---|---|---|---|---|---|---|---|---|
| Precision | | Double | | | Single | | Double | Single |
| Application | | D | D | CG | D | CG | | CG |
| Machine size | | Sim | 128 | 128 | Sim | 16 | 16 | 16 |
| Wilson | $2^4$ | 47% | 44% | 32% | | | | |
| | $4^4$ | | 43% | 38% | | | | |
| | $6^4$ | | 44% | 39% | | | | |
| | $8^4$ | | 29% | | | | | |
| Asqtad | $2^4$ | | 22% | 19% | | | | 2% |
| | $4^4$ | 43% | 42% | 40% | 15% | 7% | 6% | 15% |
| | $6^4$ | | 35% | 33% | | 9% | 6% | 19% |
| | $8^4$ | | 29% | 28% | | 7% | 6% | |
| DWF 4D kernel | $2^4$ | | 43% | | | | | |
| | $4^4$ | | 54% | | | | | |
| | $6^4$ | | 55% | | | | | |

# Planned activities

- QMP & MPI
- Batch system(PBS)
- Parallel file system
- Support for xlC (xcoff)