STATUS OF THE QCDOC PROJECT

- The Collaboration
- Hardware
  - Architecture
  - Status
- Software
  - Strategy
  - Status
  - Bench marks
- Performance/Cost/Schedule

QCDOC COLLABORATION

Columbia:	Norman Christ
	Calin Cristian
	Zhihua Dong
	Changhoan Kim
	Ludmilia Levkova
	Xiaodong Liao
	Guofeng Liu
	Robert Mawhinney
	Azusa Yamaguchi
IBM:	Dong Chen
	Alan Gara
KEK/RBRC:	Shigemi Ohta
UKQCD:	Peter Boyle
	Balint Joo
Yale/RBRC:	Tilo Wettig

TECHNOLOGY STRATEGY: BUILD UPON *QCDSP*

- Focus maximum computing resources on a single problem: many processor with fast/low-latency communications.
  - As with QCDSP, design for 20K node machines.
  - $16^3 \times 32$ lattice on a 8K node, 4 Tflops (sustained) partition (2^4 local volume).
- Homogeneous/modular approach:
  - Mesh communications, no central communications switch.
  - Add two extra dimensions for software partitioning: 4-D $\rightarrow$ 6-D
- Target a small, cool processor
  - Fastest, hottest processor does not provide optimum cost performance.
  - Cooling, packaging and reliability are essential for a large machine.

QCDOC ARCHITECTURE

- **IBM-fabricated, single-chip node.**
  [50 million transistors, 1-2 Watt, $\approx$1cm$\times$1cm die]

- **PowerPC 32-bit processor, 1 Gflops, 64-bit IEEE FPU.**

- **4 Mbyte on-chip memory.**

- **Extra 2.0 Gbyte/node with DIMM card.**

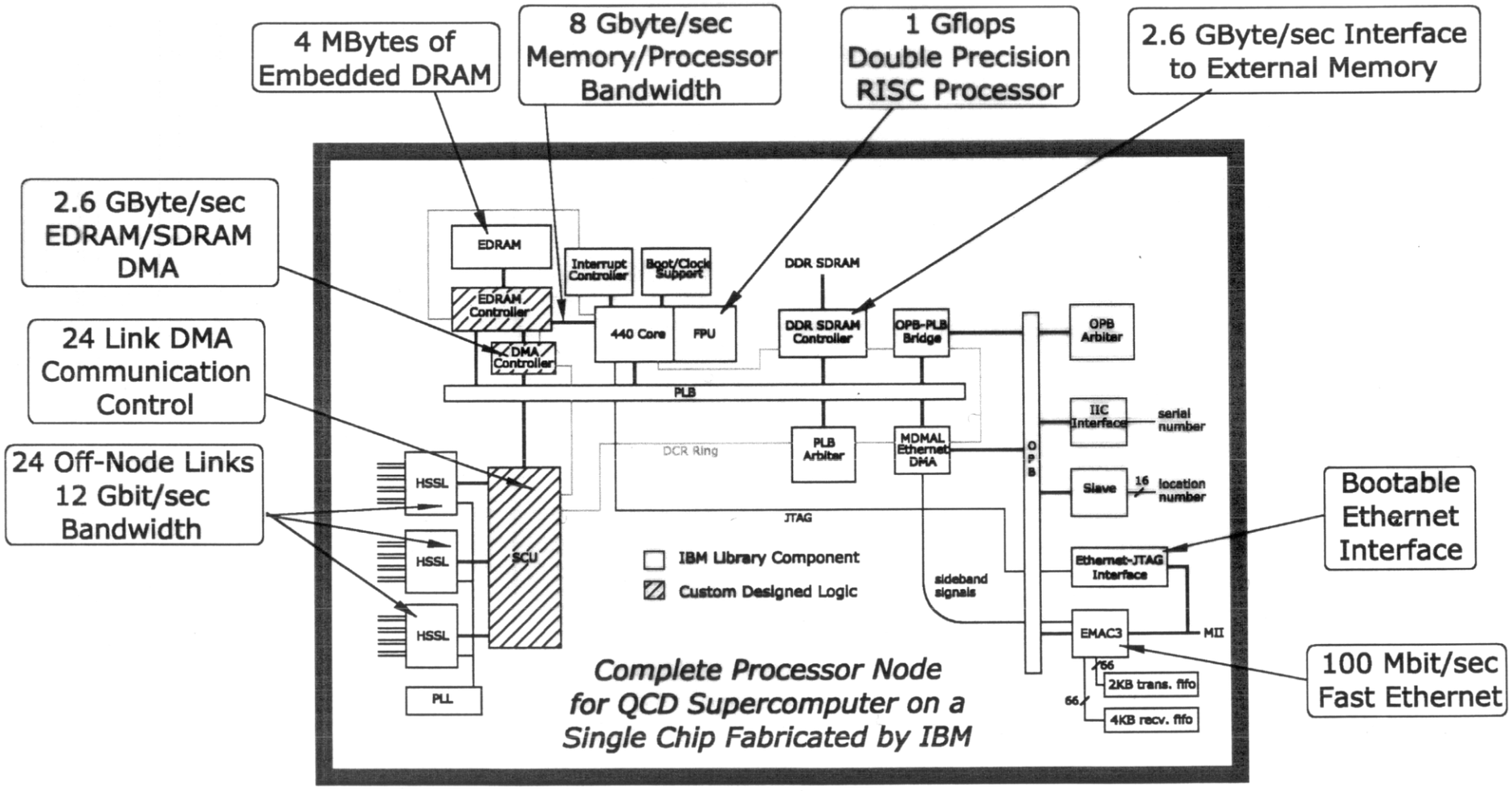
- **500 MHz serial communications: 6-D, LVDS, bidirectional.**

- **100 Mbit/sec, Fast Ethernet diagnostic-boot-I/O network.**

HARDWARE STATUS

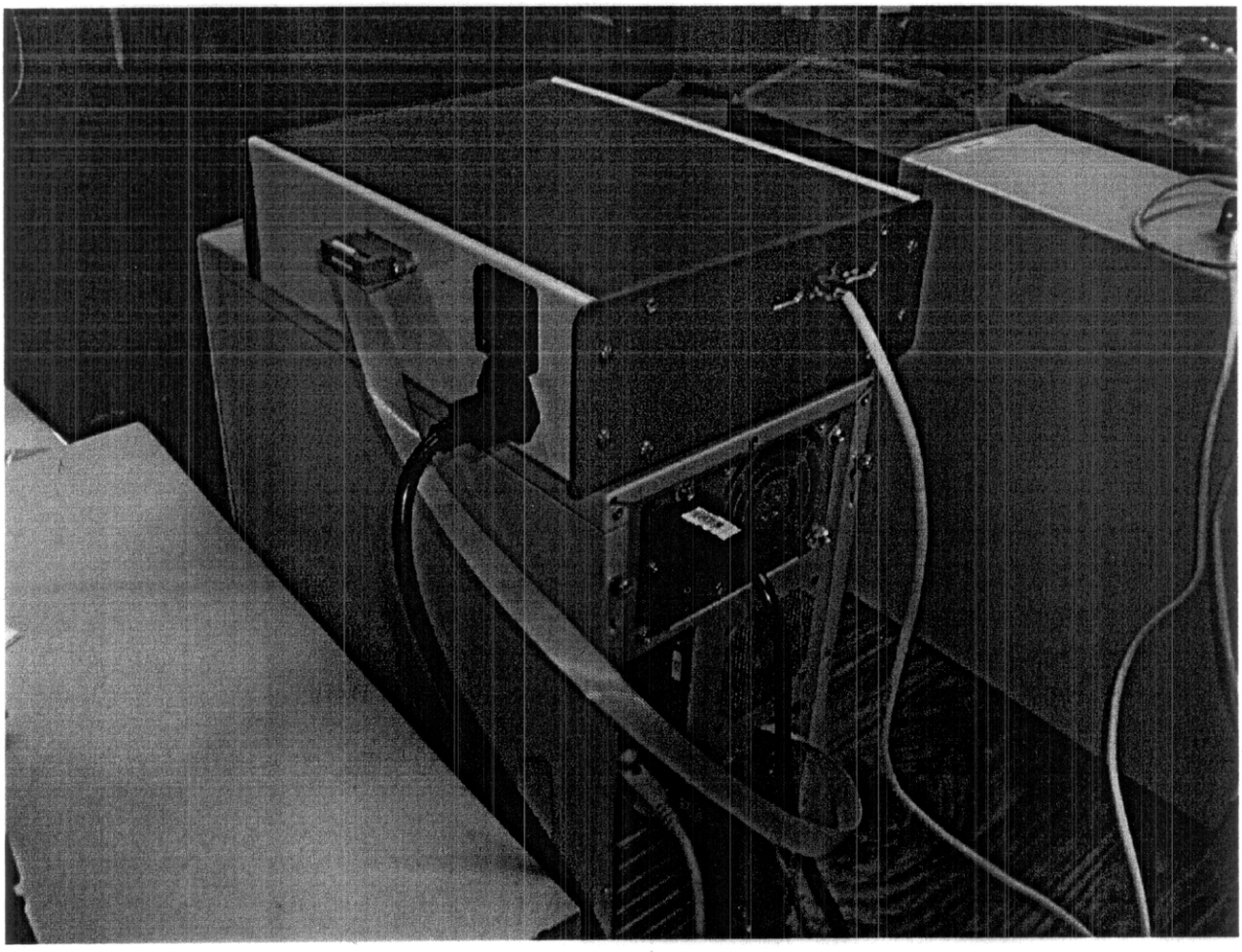
- **Present focus: Complete ASIC design.**
  - **Many items integrated, tested (preliminary netlist sent to IBM):**
    - 440 core with FPU
    - EDRAM and prefetching controller
    - DDR controller with memory model
    - Ethernet/DMA controller
  - **Ethernet/JTAG controller just working.**
  - **Serial communications unit 2/3 done.**
  - **DMA unit 1/2 designed by IBM/Rochester.**
  - **Final IBM design kit to be released August or early September.**
  - **All items synthesize without timing constraints.**
- **PC board layout, cabinet design will start this fall.**

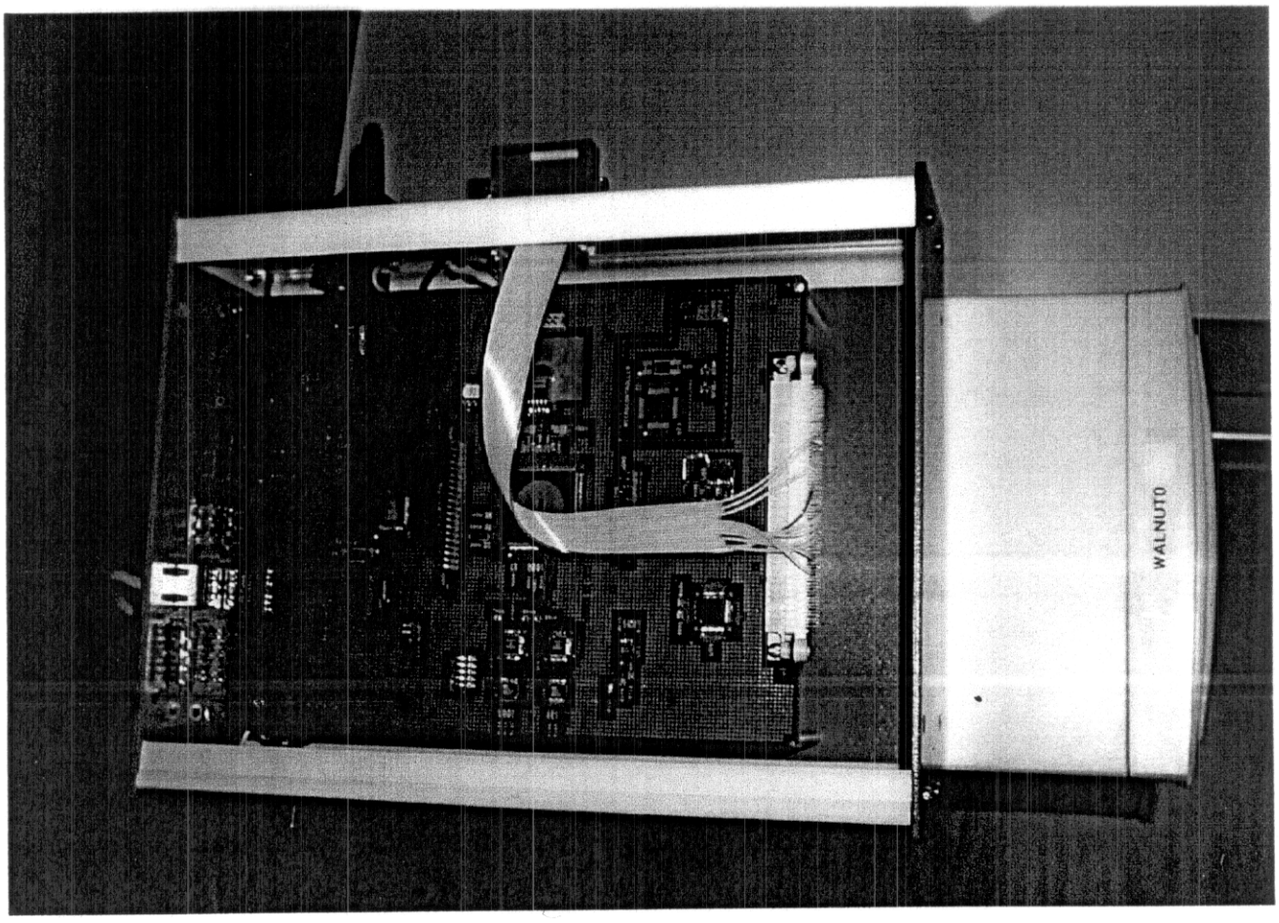
QCDOC ASIC DESIGN



DUCATION VERSION - NOT FOR COMMERCIAL USE

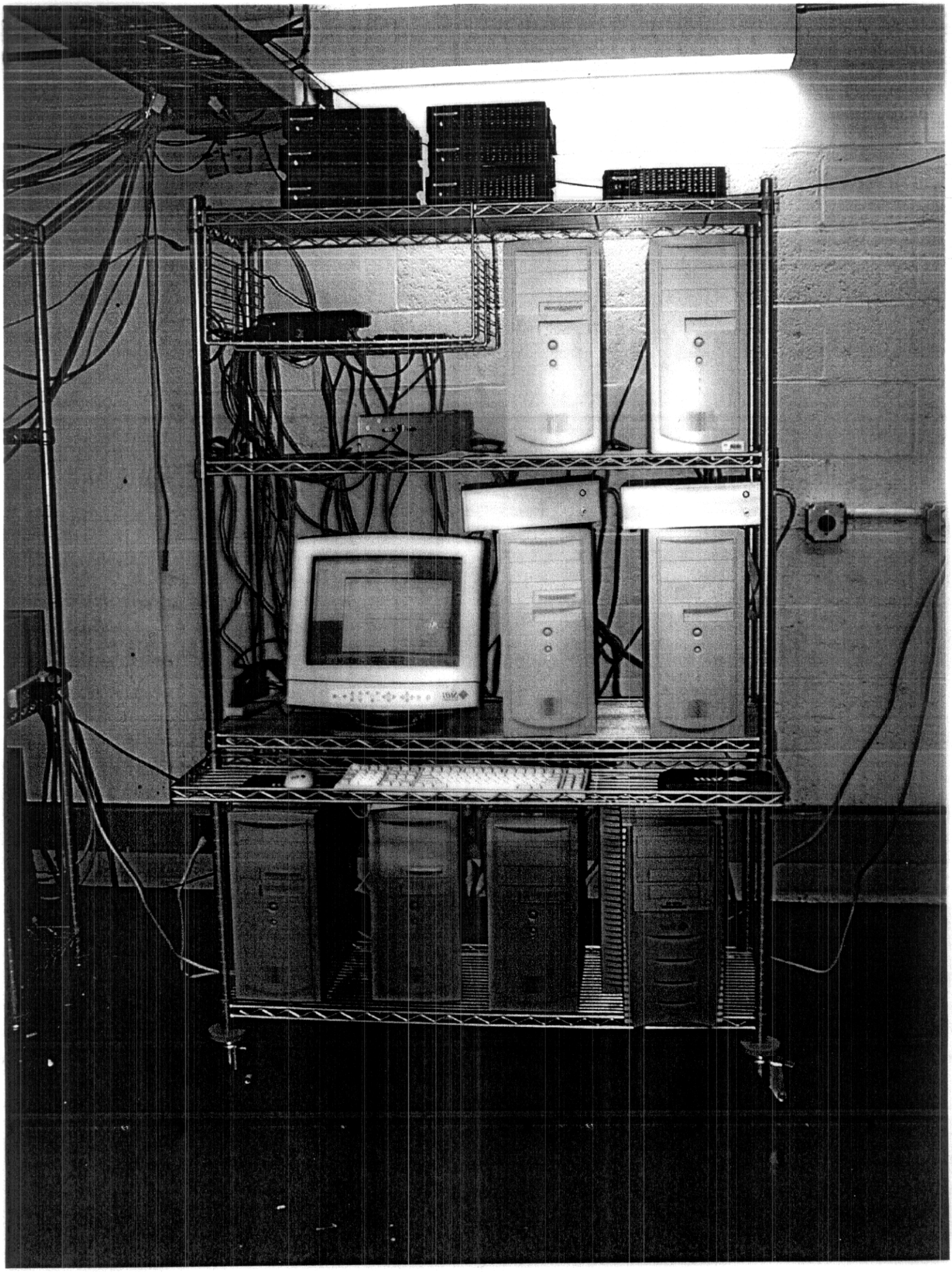
Ethernet/JTAG Test Hardware

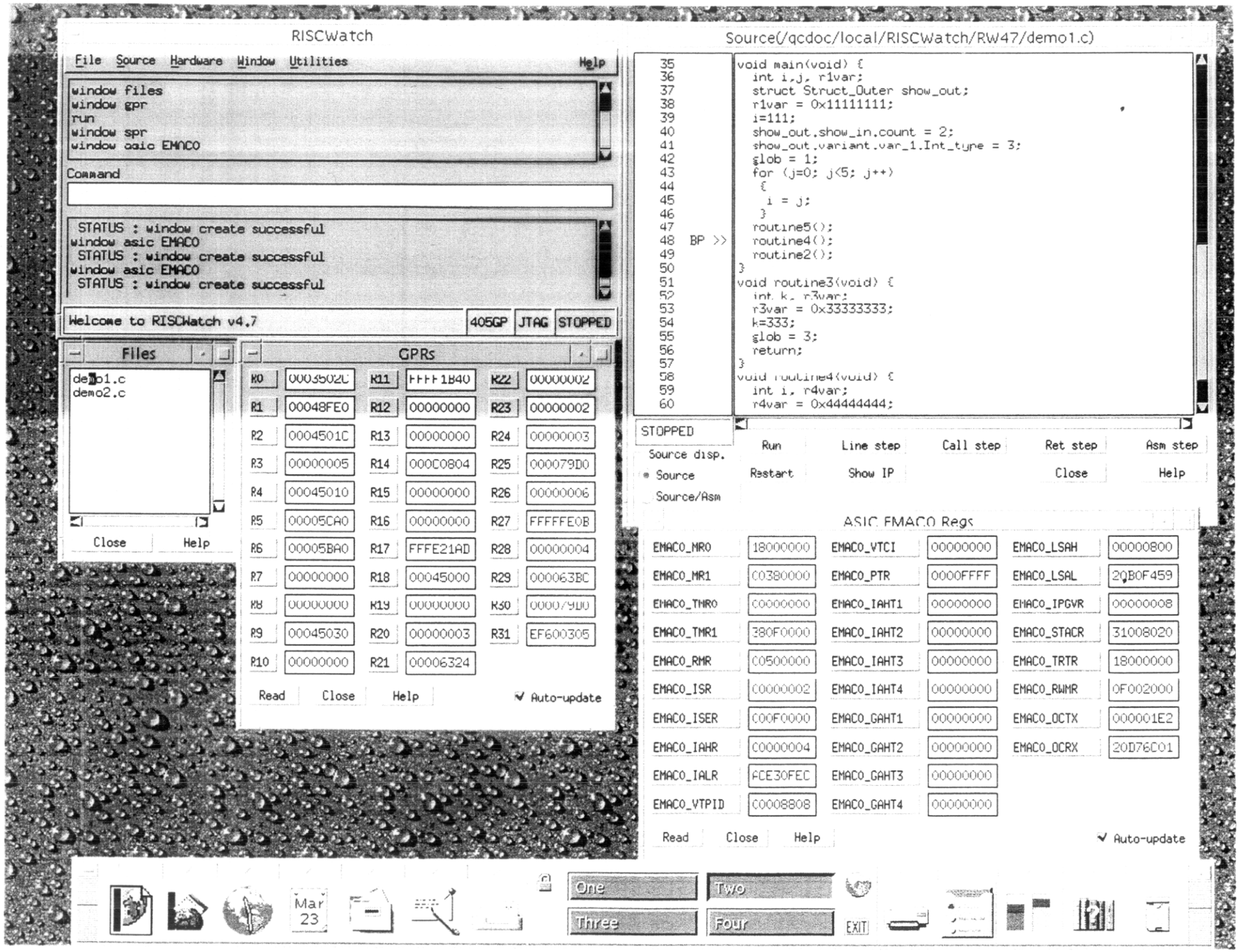




SOFTWARE STRATEGY

- **Industry standard, "RISC" processor.**
  - **GNU $C++$ compiler.**
  - **Memory management provided.**
  - **LINUX on every node possible.**
- **Minimum QCDSP functionality planned:**
  - **Low level diagnostic and boot kernel.**
  - **Host-side OS controlled by**
    - C-shell extension from console window.
    - C-shell script.
    - PERL through a socket interface.
  - **C-like routines called from application code running on the machine providing host system services: fopen(), printf()....**
  - **Nearest neighbor communications subroutines called from applications code.**
- **UKQCD/EPCC partnership in application development.**
  - **Start with present $C++$ QCDSP code, *CPS*.**
  - **Remove non-ANSI constructs.**
  - **Provide MPI backend.**
- **U.S. SciDAC support for design of common software infrastructure: QCDAPI.**




RISCWatch
File Source Hardware Window Utilities Help
window files
window gpr
run
window spr
window asic EMAC0
Command
STATUS : window create successful
window asic EMAC0
STATUS : window create successful
window asic EMAC0
STATUS : window create successful
Welcome to RISCWatch v4.7
405GP JTAG STOPPED
Files
demo1.c
demo2.c
Close Help
GPRs
R0 00035020 R11 FFFF1B40 R22 00000002
R1 00048FE0 R12 00000000 R23 00000002
R2 0004501C R13 00000000 R24 00000003
R3 00000005 R14 000C0804 R25 000079D0
R4 00045010 R15 00000000 R26 00000006
R5 00005CA0 R16 00000000 R27 FFFFFE0B
R6 00005BA0 R17 FFFE21AD R28 00000004
R7 00000000 R18 00045000 R29 000063BC
R8 00000000 R19 00000000 R30 0000790D
R9 00045030 R20 00000003 R31 EF600305
R10 00000000 R21 00006324
Read Close Help Auto-update
Source(/qcdoc/local/RISCWatch/RW47/demo1.c)
35 void main(void) {
36 int i,j, r1var;
37 struct Struct_Outer show_out;
38 r1var = 0x11111111;
39 i=111;
40 show_out.show_in.count = 2;
41 show_out.variant.var_1.Int_type = 3;
42 glob = 1;
43 for (j=0; j<5; j++)
44 {
45 i = j;
46 }
47 routine5();
48 BP >> routine4();
49 routine2();
50 }
51 void routine3(void) {
52 int k, r3var;
53 r3var = 0x33333333;
54 k=333;
55 glob = 3;
56 return;
57 }
58 void routine4(void) {
59 int i, r4var;
60 r4var = 0x44444444;
STOPPED
Run Line step Call step Ret step Asm step
Source disp.
Source
Source/Asm
Restart Show IP Close Help
ASIC EMAC0 Regs
EMAC0_MR0 18000000 EMAC0_VTCI 00000000 EMAC0_LSAH 00000800
EMAC0_MR1 C0380000 EMAC0_PTR 0000FFFF EMAC0_LSAL 2QB0F459
EMAC0_TMR0 C0000000 EMAC0_IAHT1 00000000 EMAC0_IPGVR 00000008
EMAC0_TMR1 380F0000 EMAC0_IAHT2 00000000 EMAC0_STACR 31008020
EMAC0_RMR C0500000 EMAC0_IAHT3 00000000 EMAC0_TRTR 18000000
EMAC0_ISR C0000002 EMAC0_IAHT4 00000000 EMAC0_RWMR 0F002000
EMAC0_ISER C00F0000 EMAC0_GAHT1 00000000 EMAC0_OCTX 000001E2
EMAC0_IAHR C0000004 EMAC0_GAHT2 00000000 EMAC0_OCRX 20D76C01
EMAC0_IALR FCE30FEC EMAC0_GAHT3 00000000
EMAC0_VTPID C0008808 EMAC0_GAHT4 00000000
Read Close Help Auto-update
Mar 23
One Two Three Four EXIT

Software Status and Benchmarks

- **Reasonable software environment for compiling and running code on the simulator.**
- **Working OS software:**
  - **Low-level Ethernet/JTAG software.**
  - **JTAG boot software.**
  - **Node Ethernet driver.**
  - **Host-node read/write/execute OS kernel.**
- **RISCWatch debugger modified to work through our Ethernet/JTAG interface by IBM/Raleigh.**
- **Benchmarks:**
  - **Maximum sustained EDRAM/440 bandwidth:**
    read: 3.2 Gbytes/sec write: 2.0 Gbytes/sec
    (50% of this sustained to external memory).
  - $SU(3)$ **link** $\times$ **Wilson spinor:**
    - 92% maximum
    - 84% data in cache
    - 78% data in EDRAM

PERFORMANCE/COST/SCHEDULE

- **We expect to sustain $\approx 50\%$ of peak, even for small volumes/node.**

- **Cost performance below \$1/Mflops:** $10\times$ **enhancement over QCDSP.**

- **Schedule** (peak speeds)**:**

Prototype hardware	03/02
UKQCD/Columbia/U.S pilot machine (7 Tflops)	12/02
UKQCD in Edinburgh (10 Tflops)	10/03
RBRC (10 Tflops)	
BNL LGT Center (20 Tflops)	