L2 Status

James T. Linnemann
Michigan State University
DØ Collaboration Meeting
April 3, 1998
History (Since Bloomington)

- September: Beaune IEEE, present 1st cut design
- October: NIU workshop; Standard Crate
- December: FNAL workshop: L1CFT for STT
- January: Lehman
- February: L2 Global TDR; Saclay joins
- March
  - U Md Workshop: FIC and MBT
  - CDF/DØ L2 workshop (Alpha proto)
  - STT review
- April:
  - L2 Global Review
  - UIC workshop coming (components, STT)
Money and Manpower

- 500K$ MRI grant (NIU, MSU, Stony Brook)
- Continual ramp-up since IU
  - Cal: Varelas, Adams, Hirosky, Martin, Di Loretto
  - Global: Moore
  - Preshower: Grannis, Bhattachardee
  - Mu: Evans, Gershtein
  - MBT/ CFT: Baden, Bard, Giganti, Toback
  - FIC/SFO: Le Du, Renardy, Bernard
    - will soon start needing grad students!!!!
L2 Trigger

- 10 KHz L1 out to 1 KHz L2 out
  128 L2 decision bits, 1:1 with L1
  few % deadtime
- Global Processor selects events
  threshold for object
  matching objects from different detectors
  cuts on quality
  kinematic variables (but Zv=0)
- Objects from single-detector preprocessors
Standard Crate

- TCC
- L3
- 8 VME slots minimum
- Inputs: SCL
- Outputs to Global (preprocessors only)
- L2 HWFW (Global only)

Diagram:
- VME
- MBus
- MPM
- VBD
- Worker
- Admin
- M BT
- L2 Answer
- Dec Alpha
- (Unix)

JTL, MSU 12/18/97
Bit3 MPM

- PCI Card for PC, cable, and VME master
- Add Multiport Memory Module
- Perform general VME I/O, generate interrupts
- Download parameters for run
- Run begin/end commands
- Collect Monitoring information
  preferably, already placed in MPM by Administrator Alpha
  If necessary, can collect from other modules
VBD

- VME Master to read out to L3
- Not interruptable during Readout
- Probably 10-20 MB/s effective
- Must read from SAME set of VME addresses every event
  - some of wordcounts may be zero
  - faster if fewer addresses
- intent is readout from Worker Alpha
Alphas

- Up to 1 GIP Alpha 21164 on VME card
  - small local disk for bootup
  - Enet to Dec Unix Alpha for user .EXE, debugging
- All Mbus I/O via MBT card
  - Mbus DMA input  80-100 MB/s
  - Mbus bidirectional programmed I/O  20 MB/s?
- 64b parallel I/O
- 2 per crate
  - Worker formatting, Output to Global
  - Administrator housekeeping, L3 R/O
MBT
Magic Bus Transceiver

- Vme slave; Mbus Master and slave
  Administrator controls card(s)
- 7-8 Cypress Hotlink inputs
  160 or 320 MB/s in Copper Cables
  broadcast to Alphas (Workers & Admin) on Mbus
  normal data Input path
- 2 Cypress Outputs
  Preprocessor output to L2 Global input MBT’s
MBT, continued

● Serial Command Link (SCL) Receiver
  broadcast L1 to Alphas on Mbus
    – synchronization check
    – L1 Qualifiers
  Queue L2 for Administrator Mbus reads

● 128 b Parallel I/O
  Global uses to send L2 decision to L2 HWFW
  Misc communication/control signals (VBD?)
Standard Crate Uses

- Global JUST Standard Crate described so far
- Cal: more workers
- Standard Crate can also be used with non-Alpha, non-MBus pre-preprocessor
  Cypress inputs to Worker via MBT
  - format, massage data for Global handle L2, L3 buffering & I/O, most of monitoring
  Completely standard data movement software
  - User code testable once data structure fixed
Penalty: extra latency (lose a buffer)
  - “pre-preprocessor”
SLIC: Serial Link Input Card

- 16 Cypress serial inputs
  - VME slave card (single slot?)
- 4 TI DSP’s, up to 2 GIPS each
- more inputs, CPU / slot than Alpha
- output via Hotlink to MBT
- Readout via Worker Alpha via MBT
  - Acts as pre-preprocessor
- test registers on all inputs (eg. SCL)
SFO: SCL Fanout

- Receives L1 SCL information
- Fans out as Cypress output to 16 SLIC cards
  event synchronization
  L1 Qualifiers
- functional blocks all from MBT
- No VME interface required
  except for testing?
  need not be in VME crate?
Standard Crate with SLIC

10 VME slots minimum

Dec Alpha
(Unix)
Fiber Input Converter (FIC)

- Convert Fiber Input to Cu Cypress Hotlink
  - What Cypress speed? 160 or 320?
  - What Speed Fiber? LED or Laser?
- Front end to either SLIC or MBT
  - avoids variants of complex card
- No VME needed (need not live in VME crate)
- Need if inputs are long haul from platform? (vs. transformers?)
- Harder (more expensive, fewer channels) if full-speed g-link conversion needed
Standard Crate with FIC to SLIC

11 VME slots minimum

Inputs to Global

VME

Dec Alpha
(Unix)

Michigan State University 4/10/98
Standard Crate with FIC to MBT

9 VME slots minimum

Outputs to Global

Dec Alpha
(Unix)

VME

MPM

L3

MBus

SCL

TCC
SCL Fanout Questions

- Modest project, small production run
- Needed only by SLIC’s
- 11 channels for crate filled with SLIC’s
- When? Only by Commissioning
  no trigger framework: fake SCL on SLIC
- Who?
  MBT designer, in series?
  SLIC designer or someone else?
  – after relevant MBT blocks designed
FIC: L2CFT from L1 CFT trigger

- Presently, plan g-link 1.3Gb/s = 100MB/s
  L1CFT: 100B (50 tracks)/fiber to STT in 1 \( \mu s \)
  - L1CFT plans to send fixed length, pad w/ trailing zeros

- 4 g-link inputs per card max

- 8 fibers = 2 cards for L2CFT

- Advantage of g-link FIC:
  could accept raw data (e.g. for CPS)

- 320MB/s Cu Cypress + transformer???
  only if lower to 24 tracks, and time budget to 2 \( \mu s \)
  cheaper, 8 inputs, single card for L2CFT
  - no buffering needed?
FIC: Raw Data Input

- Split of raw data fiber requires 1.3 Gb/s g-link
- needed if do CPS
  - no cable count yet
  - use as part of STT?
    - More likely, recycle part of VRB input
MBT Simplifications: are all sources intelligent?

- Enforce padding to 16 B? No?
  probably can’t if accepting raw data

- Enforce maximum event size? Try.
  Input FIFOs hold 16 worst-case M+P events
  – need definition from EVERY know source
  Truncate if overflow anyway (no marker added!)
  – In-band marker makes assumptions about data formats!
  – OK if processors can recognize w/o extra work
    - OK for L2-formatted inputs (trailers broken)
    - what about raw fiber data?

- SAME issues for SLIC inputs
MBT Testing Questions

- VME OR MBus
  - Control/Setup
  - Fake data for inputs, outputs
  - Loopback test of output(s) to inputs at full speed
    - VME readback of filled FIFO’s needed

- MBus only: need MBus, Alphas
  - Broadcast input test
  - Parallel I/O test
  - Mbus Control/Setup

- SCL Test Jig?
  - SCL L1 formatting + standard input
  - SCL L2: need Alpha?
  - Check with SCL designers: Walter Knopf in Barsotti group
Development System Questions

- Digital Unix Alpha required for debugging compile, link at any Alpha; serve disk anywhere?
- Most user software needs only simulator with correct data format and buffer structure should build into simulator
- Data movement software from Global & Cal MINOR modifications
  - specific qualifiers needed
How long do which systems stay at home?

Current estimate is 50K for a Standard Crate

Attempt communication with Global before commissioning--requires extra development crate

Timing may force production of Alpha cards early

– lose potential for later speedup?
Test Stand at Fermi

- Global, Cal-like, Mu/Track-like, Data Source
- Incomplete system--
  no HWFW
  not enough parts for full code of any/all crates
  – except maybe full playback for Global
  – could reconfigure if need be--painful!
<table>
<thead>
<tr>
<th>L2 Parts Count</th>
<th>12/18/97 18:43</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PC</strong> Alpha MBT SLIC SCL Fa Fiber Bit3 MPM VBD Cables Crate Mbus Power Cooling Cost</td>
<td>3000</td>
</tr>
<tr>
<td><strong>Unit Cost</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Count of Standard Parts</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Global</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Cal</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Mu/Tracking</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Less Cal Development</strong></td>
<td>-1</td>
</tr>
<tr>
<td><strong>Spares</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Test System/Spares</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>System</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Development</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Global</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>System</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Development</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Cal</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>System</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Development</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>CFT</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>System</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Development</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Mu</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>System</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Development</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Preshower</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Totals for Parts</strong></td>
<td>6</td>
</tr>
<tr>
<td><strong>System</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Development</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>STT</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Less STT Devel in Test</strong></td>
<td>0</td>
</tr>
<tr>
<td><strong>Totals Parts (w/ STT)</strong></td>
<td>7</td>
</tr>
</tbody>
</table>
Low Level Software

● with PC164 board:
  boot code review
    – specifics to VME Alpha board probably only in user code
  interrupt routines written
  code timer (instruction cycles)
  realtime clock interrupts
  studying interaction with debugger
  memory map under study
    – (avoiding cache trashing)
Higher Level Software

- C and C++ downloaded
- timing C++ a bit better(!) on simple codes (e.g. an implementation of FIFO)
  - writing other base data structures, facilities
    - circular buffer, time-stamp, state machine, error message
- Design in progress (TDR)
  - 2-processor communication protocol
    - for L2 Global (with 1 or more workers)
    - for L2 Preprocessor with multiple worker(s)
  - handling for 16 input buffers and 8 output buffers

L2 Global Script Runner Prototypes in C and C++
Current Status

- Alpha final spec negotiation with U Mich
- SLIC = Second Level Interface Card
  - under design at Nevis (Evans, Gara)
    - useable for STT also?
- MBT U Md
  - design under way; iterating specs
- FIC Saclay
  - inputs to both MBT and FIC
  - Standardize on 212 MHz Cypress Fiber??
Status of Alpha VME Board

- Due to go to production in ~2 weeks
- L3 Cache now increased to 4Mb as opposed to original 1Mb
- Reset register to be added to PCI addressable through VME to allow TCC to reset board
Status of Alpha VME Board

- P2 connector defined:
  26 pins of rows A/C connected (2 used for CDF PECL clock)
  all connected to Xilinx FPGA acting as PCI slave (but capable of generating PCI interrupts)
  compatible with D0 VME crate since A/C rows not used or bussed
- Digital I/O lines added for monitoring and VBD status
  VBD lines connect to TTL pins on P2 connector
  32 channels ECL out on front panel (not yet confirmed) for hardware monitoring (CDF configuration of 16 in/16 out LVDS possible instead if anyone needs it!)
  can add more channels if needed using a transition board attached to P2 connector to drive ECL/TTL/…. from TTL inputs
# L2 Communication

<table>
<thead>
<tr>
<th>Source</th>
<th>Destination</th>
<th>Communication</th>
<th>Medium</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrator</td>
<td>Worker</td>
<td>Allocate buffers and start processing next event</td>
<td>MBus</td>
<td></td>
</tr>
<tr>
<td>Administrator</td>
<td>Worker</td>
<td>Stop everything and reset yourself</td>
<td>MBus(?)</td>
<td>Requires interrupt</td>
</tr>
<tr>
<td>Administrator</td>
<td>Worker</td>
<td>Disconnect from data broadcast</td>
<td>MBus(?)</td>
<td>Needed if we want shadow nodes</td>
</tr>
<tr>
<td>Administrator</td>
<td>MBT</td>
<td>Ready for next event</td>
<td>MBus</td>
<td>All workers must also be ready for event</td>
</tr>
<tr>
<td>Administrator</td>
<td>MBT</td>
<td>L2 global decision for event</td>
<td>MBus</td>
<td></td>
</tr>
<tr>
<td>Administrator</td>
<td>VBD</td>
<td>Start L3 readout</td>
<td>VME</td>
<td>locks VME for several milliseconds</td>
</tr>
<tr>
<td>Administrator</td>
<td>TCC(Bit3)</td>
<td>Monitor data ready</td>
<td>VME</td>
<td></td>
</tr>
<tr>
<td>Worker</td>
<td>Administrator</td>
<td>Finished processing event, tell me what to do next</td>
<td>MBus</td>
<td></td>
</tr>
<tr>
<td>MBT</td>
<td>Administrator</td>
<td>HWFW Event Decision</td>
<td>MBus</td>
<td></td>
</tr>
<tr>
<td>MBT</td>
<td>All active alpha CPU cards</td>
<td>Event data</td>
<td>MBus</td>
<td>MBus broadcast to all cards</td>
</tr>
<tr>
<td>VBD</td>
<td>Administrator</td>
<td>Status of L3 readout</td>
<td>FRED port(?)</td>
<td></td>
</tr>
</tbody>
</table>
L2CalPP Control Issues

- Lockstep vs non-lockstep/asynchronous processing
- Lockstep mode = Event start time the same for all workers. First worker to finish must wait for slowest one.
- Non-lockstep mode = Worker starts processing next event as without regard to state of other workers.
RESQ Simulations

- Use Jay Wightman’s “realistic L2” set up
- 1 Missing E_T Worker, fixed time 45 µs
- EM/Jet independently vary by Hyperexponential dist
- Solid points requires EM/Jet identical
- All processing times listed are for algorithm only, data movement and control are separate parts of simulation
RESQ--The Upshot

- lockstep very sensitive to processing time (over almost all acceptable times)
- Within reason, processing time irrelevant in non-lockstep mode (times < 50 µs)

*Use non-lockstep mode in L2CalPP*
L2CalPP Event Loop

- Non-lockstep event loop conceptually more difficult than lockstep
- In principle, normal event processing portion of event loop is a solved problem
- Still many open issues re: monitor/ing event processing in non-lockstep mode.
Admin Event Completion, Single Worker System

- $T(F) \rightarrow \{P\}$
- $H(R) \rightarrow T(A)$

$[T(A)]$ sent in reply to Worker.
Admin Event Completion, Single Worker System, II

- $T(F) \rightarrow \{P\}$
- $H(R) \rightarrow T(A-Admin), T(TBA-Worker)$
- Only difference between $T(A-Admin)$ and $T(TBA-Worker)$ is the label
- $[T(A-Worker)]$ sent in reply to Worker

Diagram:

- Free
- Admin Alloc
- Worker ToBeAlloc
- To Worker
- Filled
- Processed
- time
Admin Event Completion, Multiple Worker System

Free → Mapped

Admin Filled → Worker Filled

Admin Alloc → Worker ToBeAlloc

To Worker

Processed
Simulation (Sigh)

- L2 Global script runner prototypes under way
  C and C++ versions for timing ("self-simulating")
  fixed allocation at initialization
  script generation still under discussion
- No L2 preprocessor simulation of L2G inputs
- No L2G output simulation for inputs to L3
- No L1 simulation to provide inputs to L2
  Unlike L2, these are "extra work"
- We NEED these simulations linked together!!!