# 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

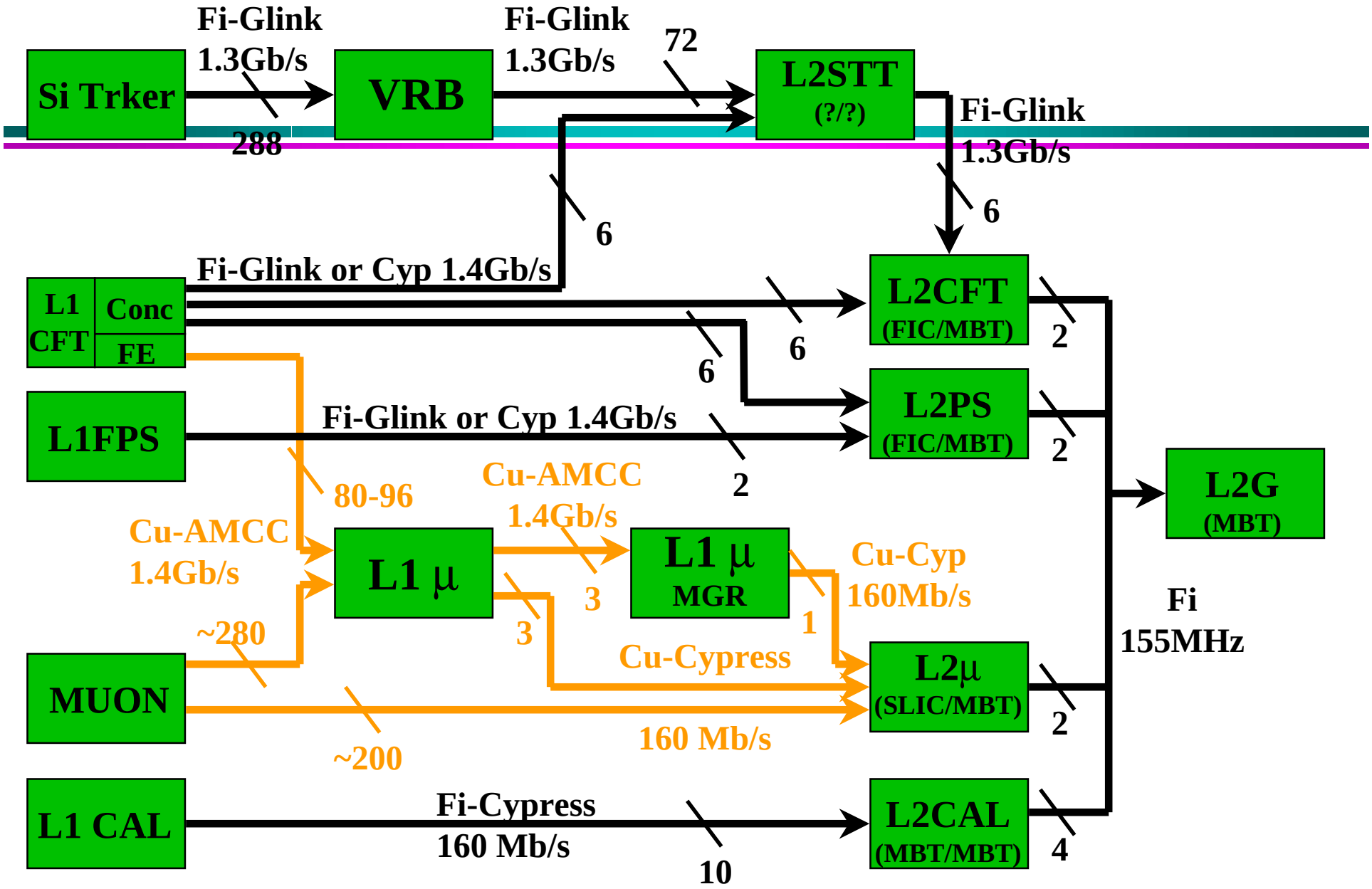
  Preshower: Grannis, Bhattachardee

  Mu: Evans, Gershtein

  MBT/ CFT: Baden, Bard, Giganti, Toback

  FIC/SFO: Le Du, Renardy, Bernard

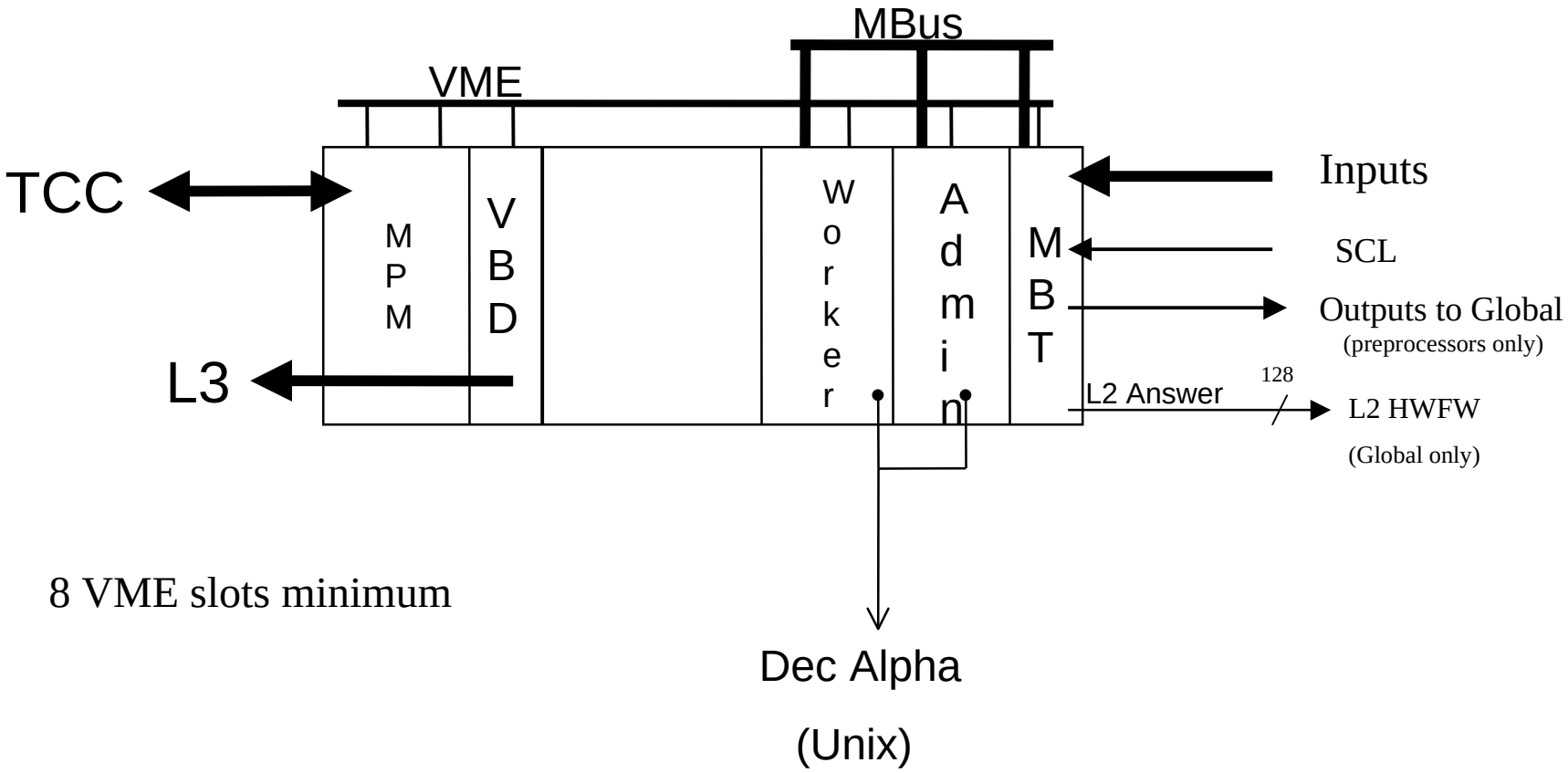
  – will soon start needing grad students!!!!


Si Trker
Fi-Glink
1.3Gb/s
288
VRB
Fi-Glink
1.3Gb/s
72
L2STT
(?/?)
Fi-Glink
1.3Gb/s
6
6
Fi-Glink or Cyp 1.4Gb/s
L1
CFT
Conc
FE
L2CFT
(FIC/MBT)
2
6
6
L2PS
(FIC/MBT)
2
L1FPS
Fi-Glink or Cyp 1.4Gb/s
2
L2G
(MBT)
80-96
Cu-AMCC
1.4Gb/s
Cu-AMCC
1.4Gb/s
L1 μ
L1 μ
MGR
3
3
1
Cu-Cyp
160Mb/s
Fi
155MHz
~280
Cu-Cypress
MUON
L2μ
(SLIC/MBT)
2
160 Mb/s
~200
L1 CAL
Fi-Cypress
160 Mb/s
10
L2CAL
(MBT/MBT)
4

# 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




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"*

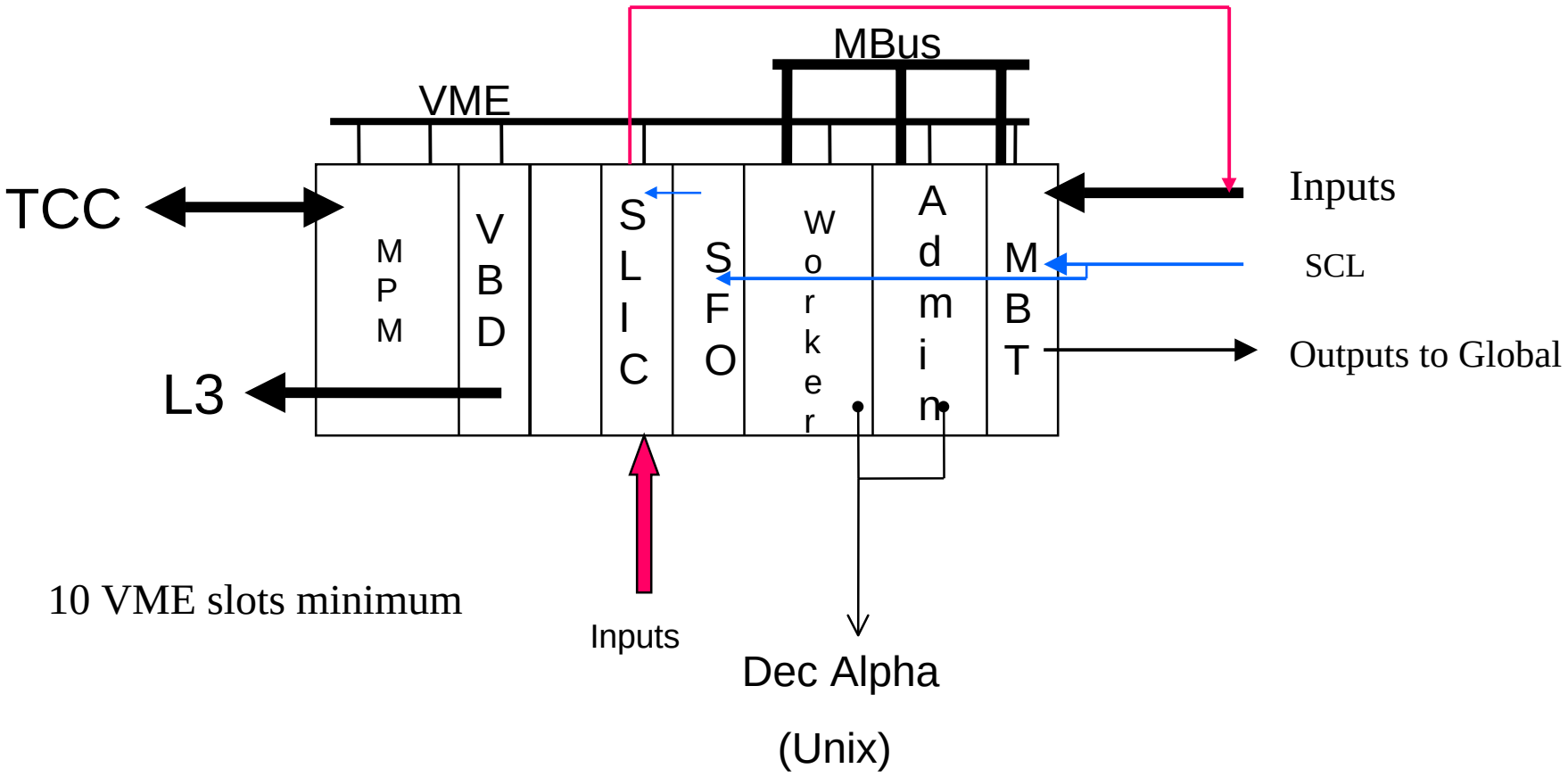
# SLIC:
# Serial Link Input Card

- 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

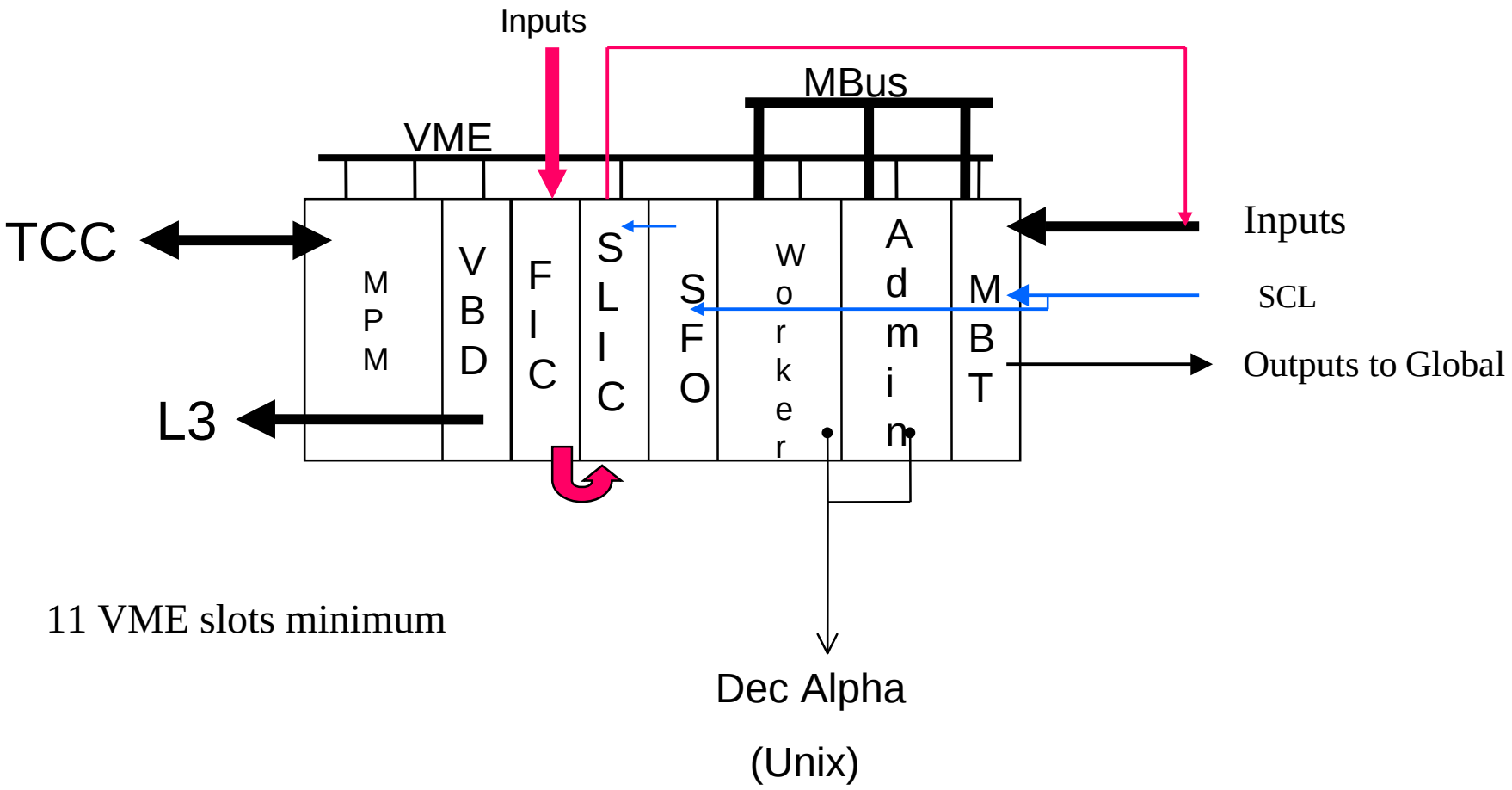



JTL, MSU 12/18/97

# Fiber Input Converter (FIC)

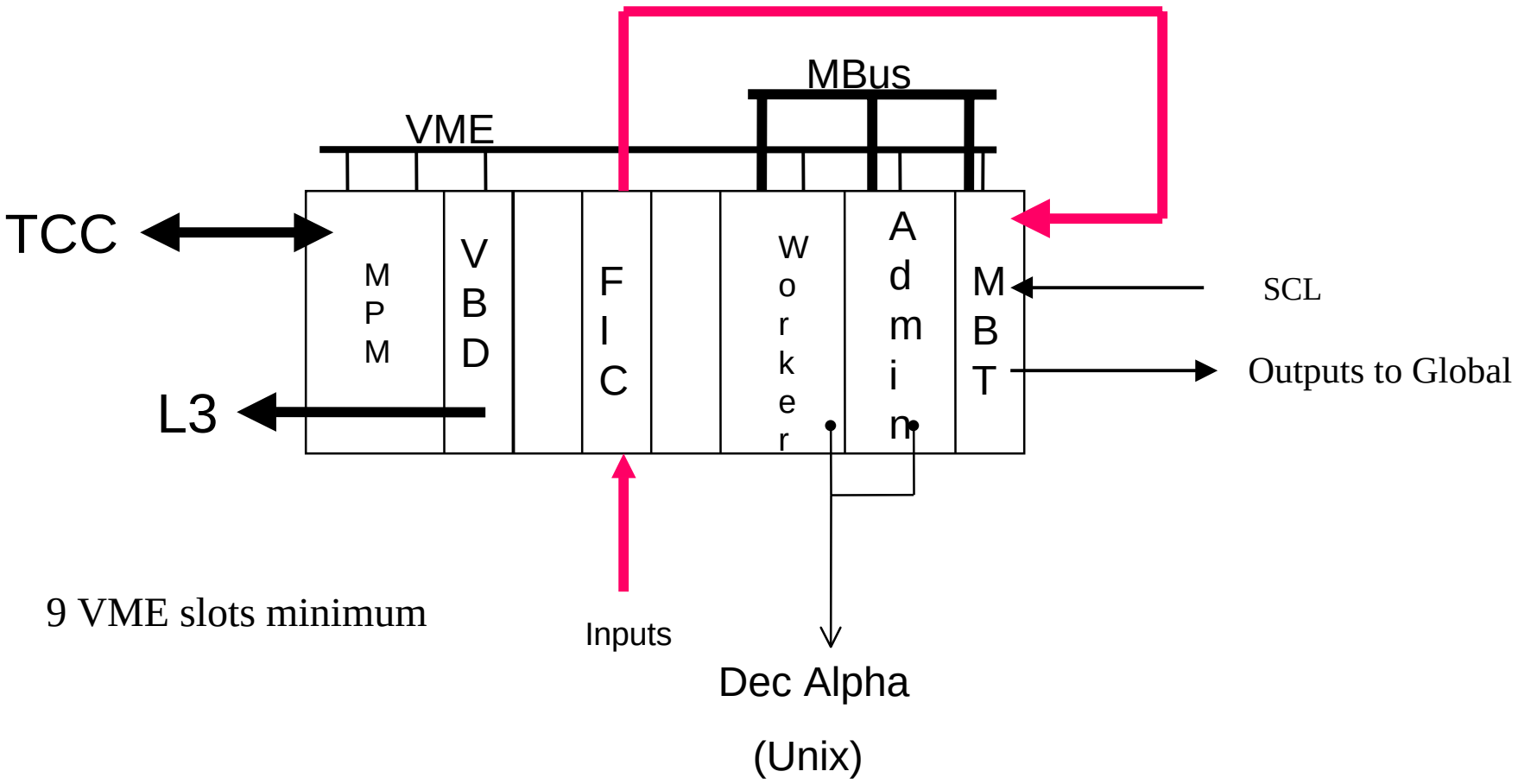
- Convert Fiber Input to Cu Cypress Hotlink
  - What Cypress speed? 160 or 320?
  - What Speed Fiber? LED or Laser?
- Front end to *either* SLIC or MBT
  - avoids variants of complex card
- No VME needed (need not live in VME crate)
- Need if inputs are long haul from platform ?
  - (vs. transformers?)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

# Standard Crate with FIC to SLIC




JTL, MSU 12/18/97

# Standard Crate with FIC to MBT




JTL, MSU 12/18/97

# SCL Fanout Questions

- Modest project, small production run
- Needed only by SLIC's
- 11channels 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 L1formatting + 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

# Development System, II

- 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!

| ***L2 Parts Count*** | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***12/18/97 18:43*** | | | | | | | | | | | | | | | |
| | PC | Alpha | MBT | SLIC | SCL Fa | Fiber | Bit3 MPM | VBD | Cables | Crate | Mbus | Power | Cooling | Cost | |
| Unit Cost | 3000 | 10000 | 5000 | 10000 | 5000 | 10000 | 5000 | 0 | 100 | 3300 | 1500 | 4000 | 2000 | | |
| ***Count of Standard Parts*** | | | | | | | | | | | | | | | |
| Global | 1 | 2 | 2 | 0 | 0 | 0 | 1 | 1 | 16 | 1 | 1 | 1 | 0.5 | 49400 | |
| Cal | 0 | 3 | 1 | 0 | 0 | 0 | 1 | 1 | 3 | 1 | 1 | 1 | 0.5 | 50100 | |
| Mu/Tracking | 0 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 0.5 | 75100 | |
| Data | 0 | 2 | 2 | 0 | 0 | 0 | 1 | 0 | 10 | 1 | 1 | 1 | 0.5 | 45800 | |
| Less Cal Development | -1 | -2 | -2 | 0 | 0 | 0 | -1 | 0 | -3 | -1 | -1 | -1 | 0 | -47100 | |
| Spares | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 5 | 1 | 1 | 1 | 1 | 74300 | |
| **Test System/Spares** | 1 | 8 | 5 | 3 | 3 | 3 | 4 | 4 | 34 | 4 | 4 | 4 | 3 | 173300 | |
| | | | | | | | | | | | | | | | |
| System | 1 | 2 | 2 | 0 | 0 | 0 | 1 | 1 | 16 | 1 | 1 | 1 | 1 | 50400 | |
| Development | 0 | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 5 | 1 | 1 | 1 | 0 | 39300 | MSU |
| **Global** | 1 | 4 | 3 | 0 | 0 | 0 | 2 | 1 | 21 | 2 | 2 | 2 | 1 | 89700 | |
| | | | | | | | | | | | | | | | |
| System | 0 | 4 | 2 | 0 | 0 | 0 | 1 | 1 | 16 | 1 | 1 | 1 | 1 | 67400 | |
| Development | 1 | 2 | 2 | 0 | 0 | 0 | 1 | 0 | 3 | 1 | 1 | 1 | 0 | 47100 | UIC to D0 |
| **Cal** | 1 | 6 | 4 | 0 | 0 | 0 | 2 | 1 | 19 | 2 | 2 | 2 | 1 | 114500 | |
| | | | | | | | | | | | | | | | |
| System | 0 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 20 | 1 | 1 | 1 | 1 | 82800 | |
| Development | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 3 | 1 | 1 | 1 | 0 | 42100 | UMD |
| **CFT** | 1 | 4 | 3 | 2 | 1 | 1 | 2 | 1 | 23 | 2 | 2 | 2 | 1 | 124900 | |
| | | | | | | | | | | | | | | | |
| System | 0 | 4 | 4 | 16 | 2 | 0 | 2 | 2 | 250 | 2 | 2 | 2 | 1 | 284600 | |
| Development | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3000 | NIU to D0 |
| **Mu** | 1 | 4 | 4 | 16 | 2 | 0 | 2 | 2 | 250 | 2 | 2 | 2 | 1 | 287600 | |
| | | | | | | | | | | | | | | | |
| System | 0 | 4 | 4 | 4 | 2 | 2 | 2 | 2 | 100 | 2 | 2 | 2 | 1 | 169600 | |
| Development | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3000 | SB |
| **Preshower** | 1 | 4 | 4 | 4 | 2 | 2 | 2 | 2 | 100 | 2 | 2 | 2 | 1 | 172600 | |
| | | | | | | | | | | | | | | | |
| **Totals for Parts** | 6 | 30 | 23 | 25 | 8 | 6 | 14 | 11 | 447 | 14 | 14 | 14 | 8 | 962600 | |
| | | | | | | | | | | | | | | | |
| System | 0 | 2 | 2 | 0 | 0 | 0 | 1 | 1 | 16 | 1 | 1 | 1 | 1 | 47400 | |
| Development | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 5 | 1 | 1 | 1 | 0 | 42300 | BU? |
| **STT** | 1 | 4 | 3 | 0 | 0 | 0 | 2 | 1 | 21 | 2 | 2 | 2 | 1 | 89700 | |
| **Less STT Devel in Test** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **Totals Parts (w/ STT)** | 7 | 34 | 26 | 25 | 8 | 6 | 16 | 12 | 468 | 16 | 16 | 16 | 9 | 1656700 | |

# 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)

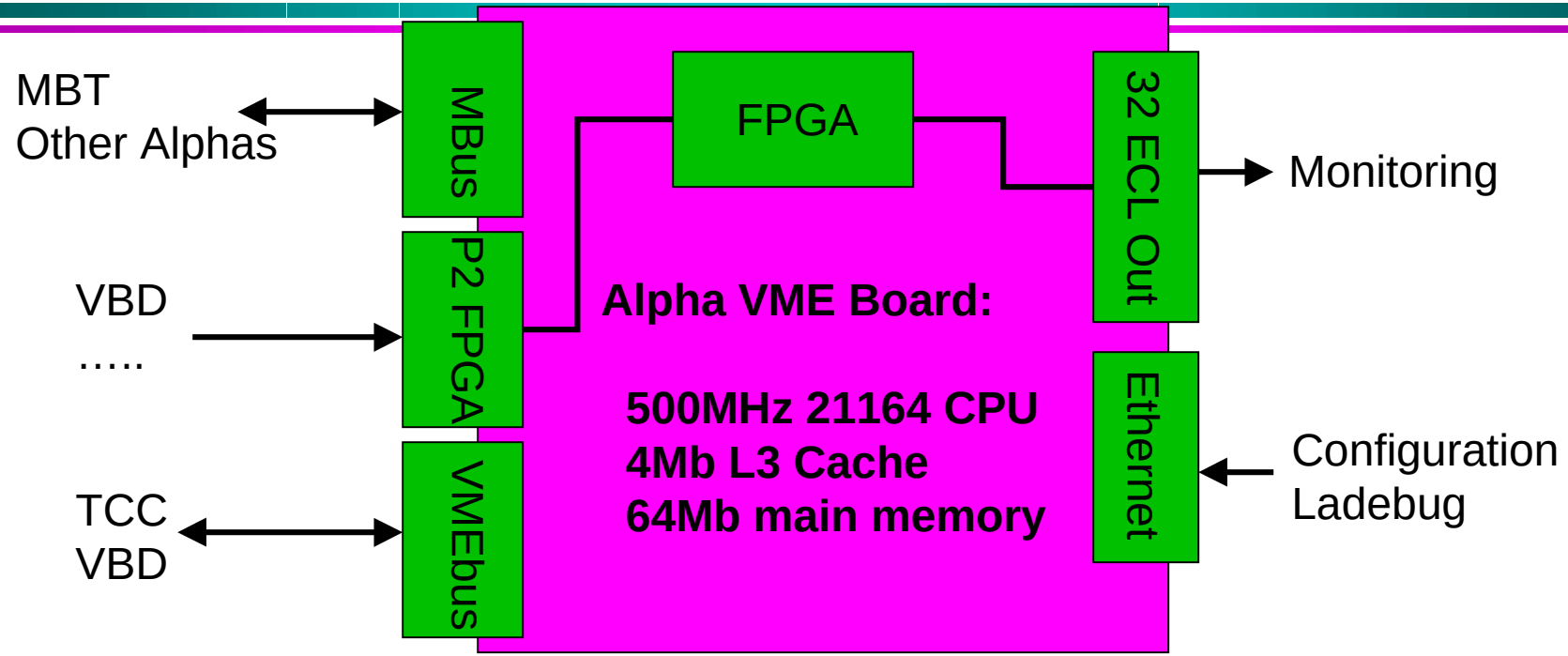
  – useable for STT also?
- MBT  U Md

  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

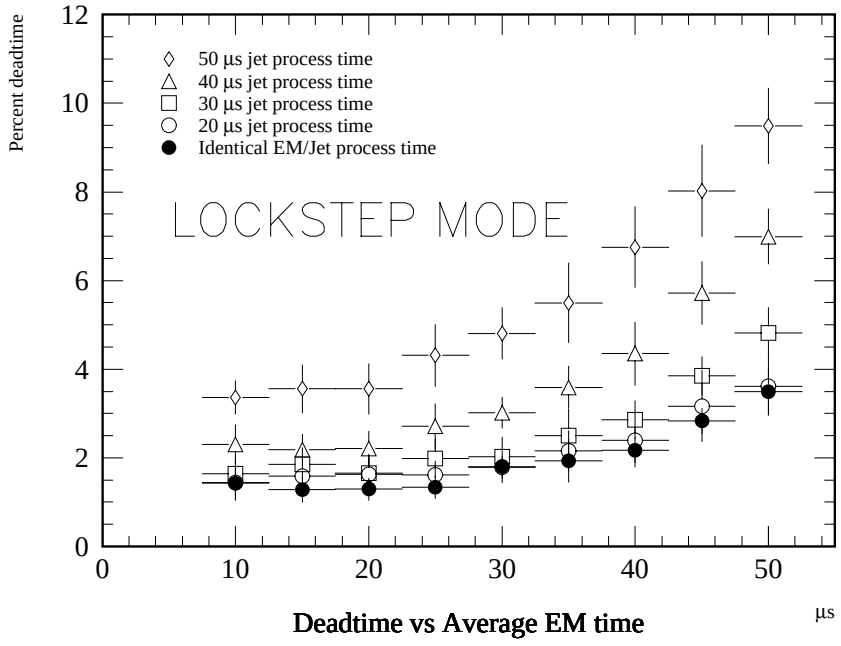
| Source | Destination | Communication | Medium | Notes |
|---|---|---|---|---|
| Administrator | Worker | Allocate buffers and start processing next event | MBus | |
| Administrator | Worker | Stop everything and reset yourself | MBus(?) | Requires interrupt |
| Administrator | Worker | Disconnect from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

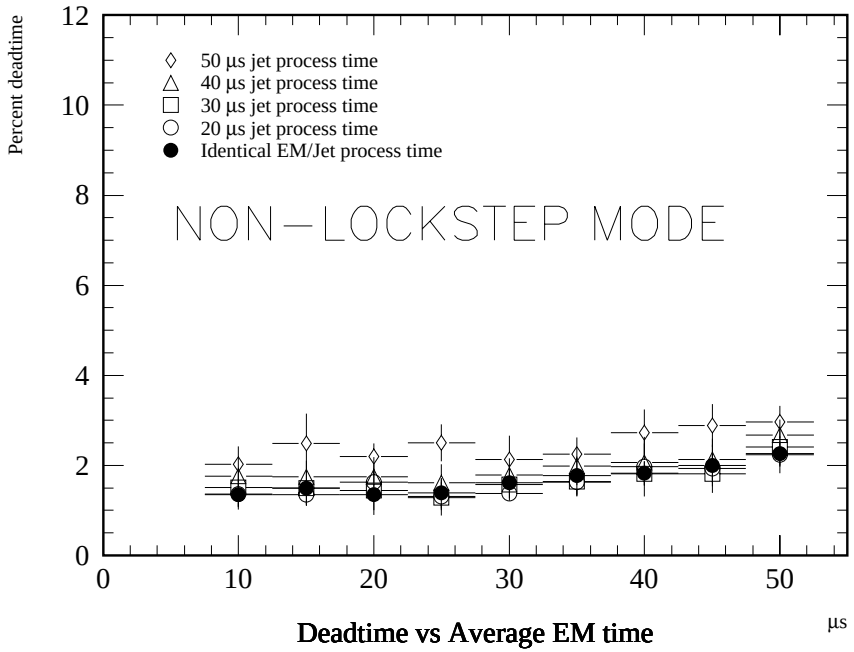
# L2CalPP Control Issues

- 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)

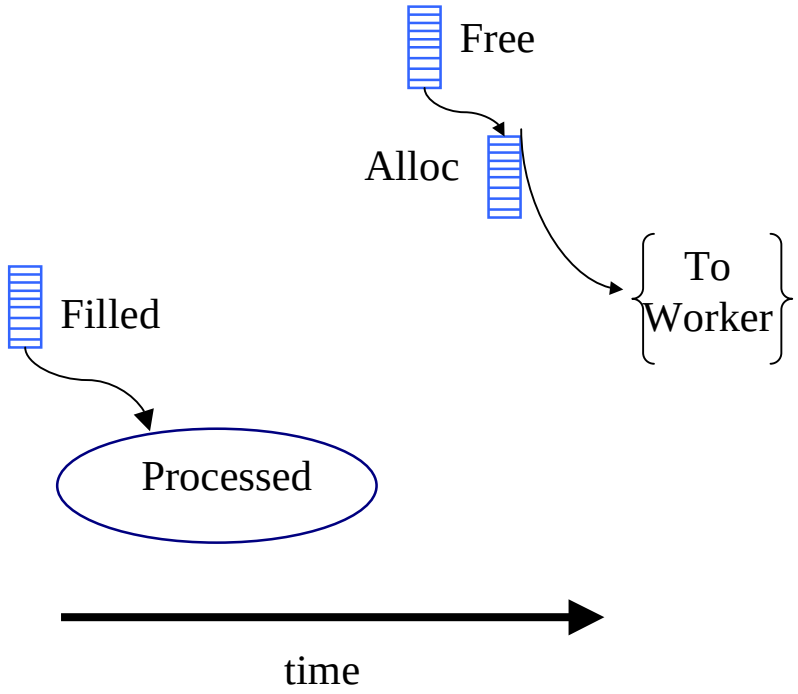
***Use non-lockstep mode in L2CalPP***

# 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.

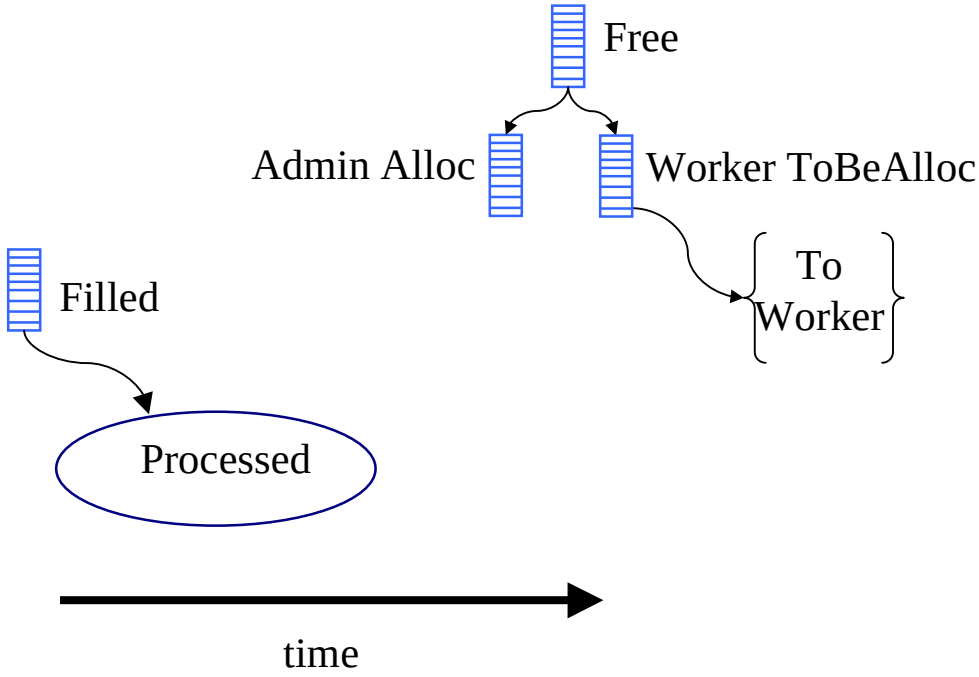
# Admin Event Completion, Single Worker System

- T(F) ⟶ {P}
- H(R) ⟶ T(A)
- [T(A)] sent in reply to Worker

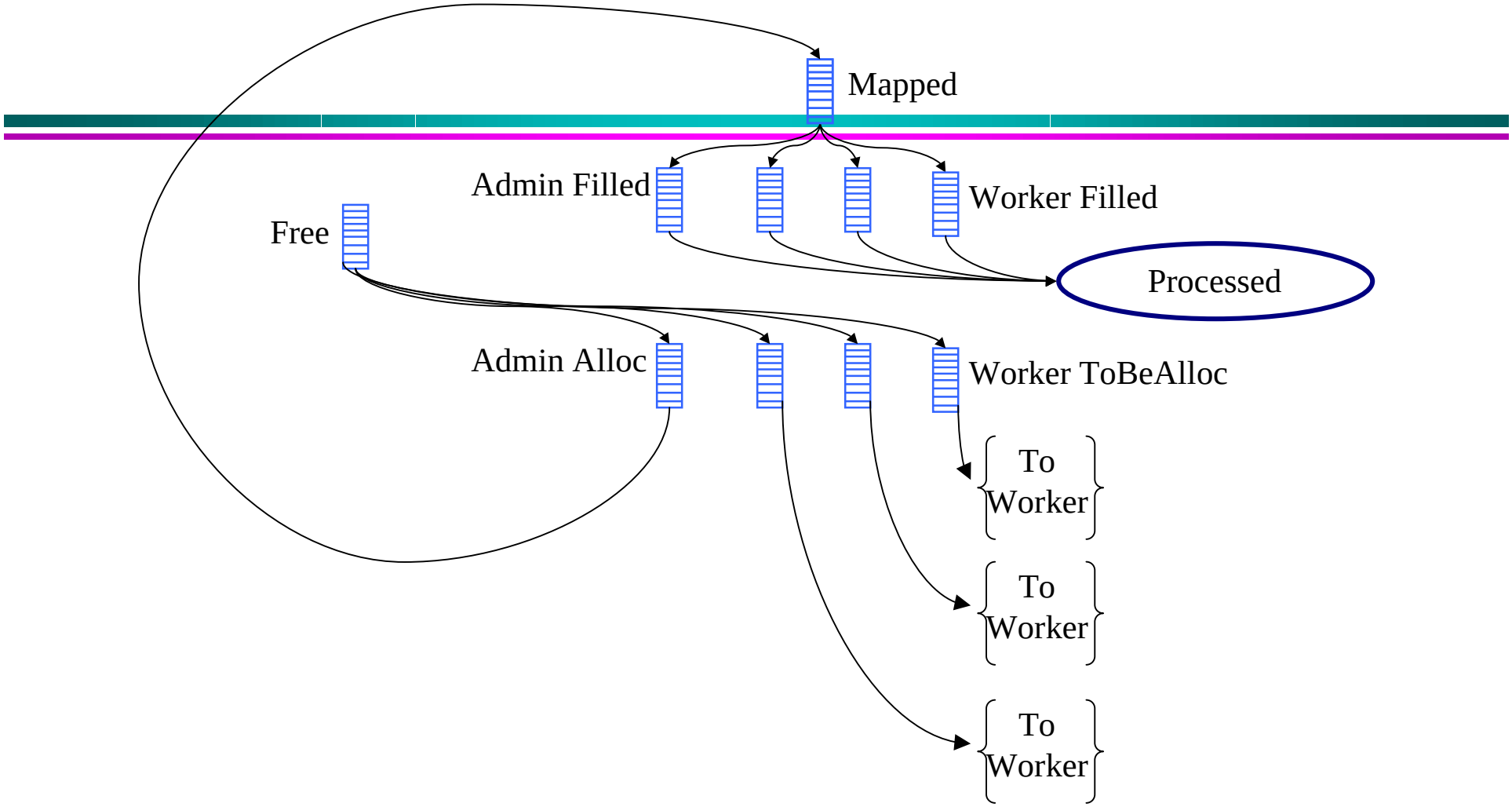


# Admin Event Completion, Single Worker System, II

- T(F) ⟶ {P}
- H(R) ⟶ 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



# Admin Event Completion, Multiple Worker System



# 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!!!