CERN/ACES Workshop

A new proposal for the construction of high speed, massively parallel, ATCA based Data Acquisition Systems

Michael Huffer, mehsys@slac.stanford.edu
Stanford Linear Accelerator Center
March, 3-4, 2009

Representing:
Mark Freytag
Gunther Haller
Ryan Herbst
Chris O'Grady
Amedeo Perazzo
Leonid Sapozhnikov
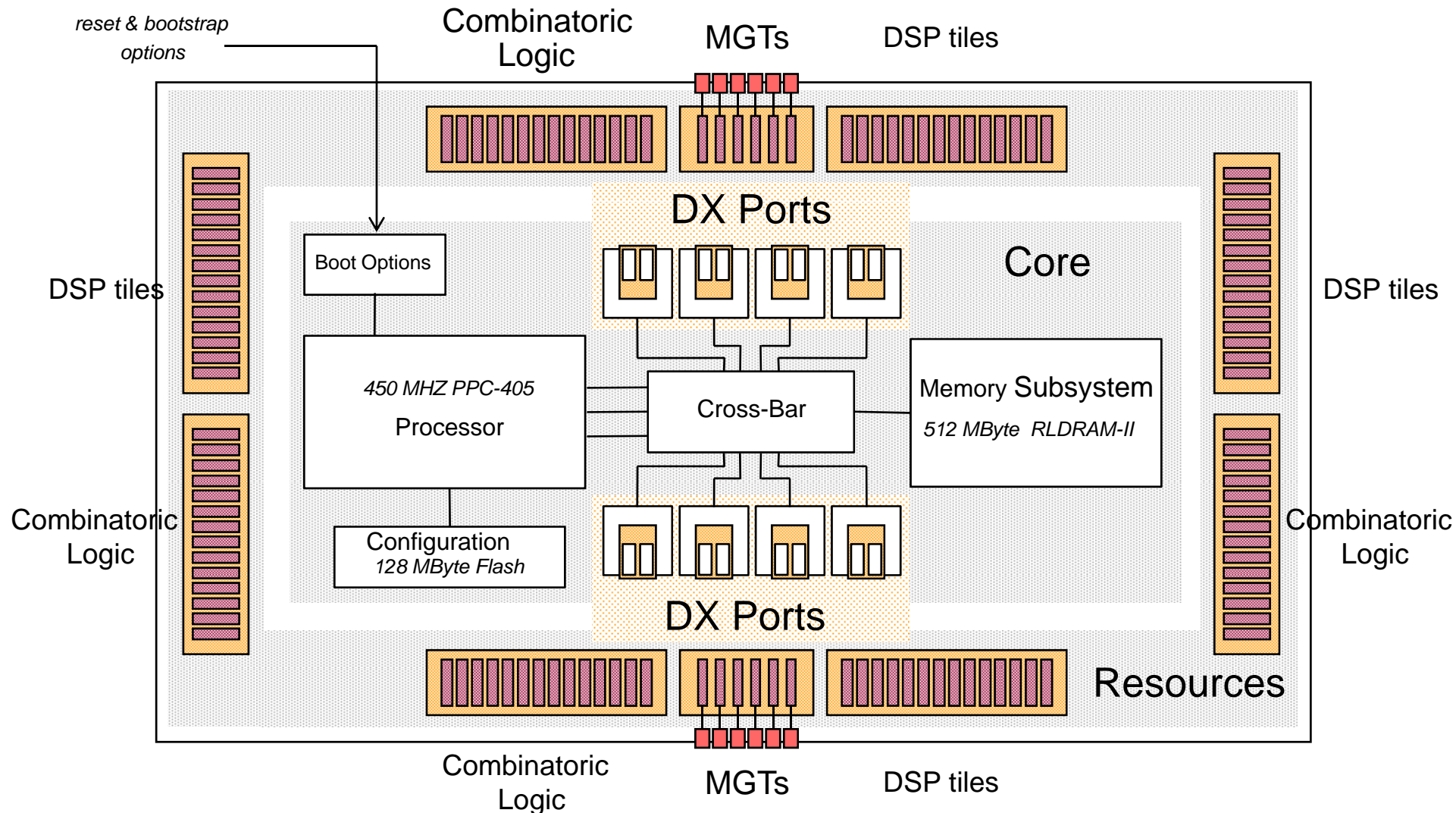*Eric Siskind*
Matt Weaver

Outline

- *DAQ/trigger "technology"* for next generation HEP experiments…
  - **Result of ongoing SLAC R & D project**
    - "Survey the requirements and capture their commonality"
  - **Intended to leverage recent industry innovation**
  - **Technology, "one size does <u>not</u> fit all" (Ubiquitous building blocks)**
    - The *(Reconfigurable) Cluster Element* (RCE)
    - The *Cluster Interconnect* (CI)
    - Industry standard packaging (*ATCA*)
  - **Technology evaluation & demonstration "hardware"**
    - The RCE & CI boards
- **Use this technology to explore alternate (ATLAS) TDAQ architectures**
- **For example (one such case study)…**
  - <u>Common</u> (subsystem independent) ROD *platform*
  - **Combine abstract functionality of <u>bot</u>h ROD + ROS**
  - **Provide <u>both</u> horizontal & *vertical* connectivity (peer-to-peer)**
  - **Hardware architecture for such a scheme has three elements…**
    - ROM (Read-Out-Module)
    - CIM (Cluster-Interconnect-Module)
    - ROC (Read-Out-Crate)

Three building block concepts

- Computational elements
    - must be low-cost
        - $$$
        - footprint
        - power
    - must support a variety of computational models
    - must have both flexible and performanent I/O
- Mechanism to connect together these elements
    - must be low-cost
    - must provide low-latency/high-bandwidth I/O
    - must be based on a commodity (industry) protocol
    - must support a variety of interconnect topologies
        - hierarchical
        - peer-to-peer
        - fan-In & fan-Out
- Packaging solution for both element & interconnect
    - must provide High Availability
    - must allow scaling
    - must support different physical I/O interfaces
    - preferably based on a commercial standard

- The Reconfigurable Cluster Element (RCE) based on:
    - System-On-Chip technology (SOC)
        - *Virtex*-4 & 5

- The Cluster Interconnect (CI)
    - based on 10-GE *Ethernet* switching

ATCA
    - *Advanced Telecommunication Computing Architecture*
    - crate based, serial backplane
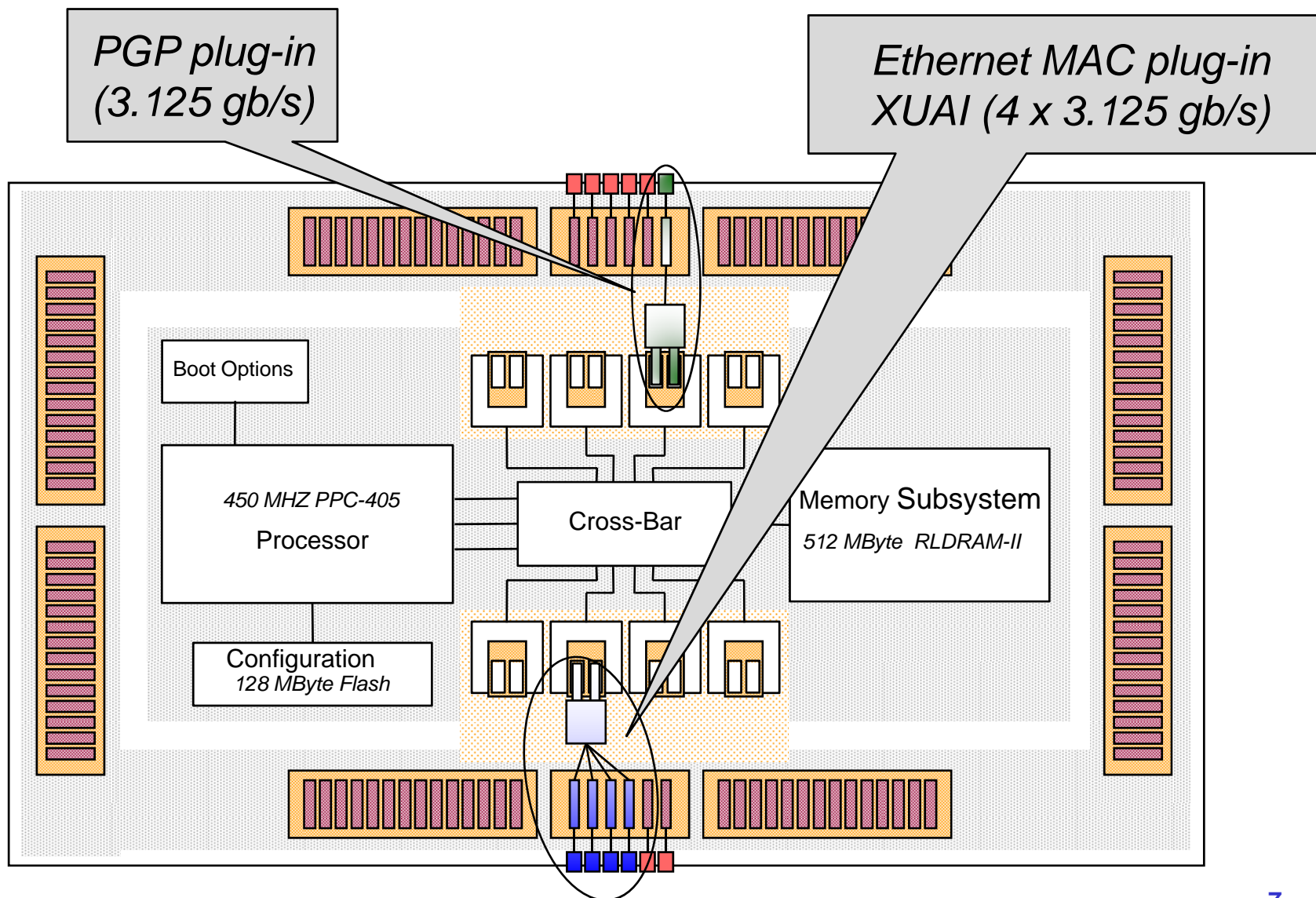
3

(Reconfigurable) Cluster Element (RCE)

reset & bootstrap options

Combinatoric Logic

MGTs

DSP tiles

Boot Options

DX Ports

Core

DSP tiles

DSP tiles

450 MHZ PPC-405 Processor

Cross-Bar

Memory Subsystem *512 MByte RLDRAM-II*

Configuration *128 MByte Flash*

DX Ports

Combinatoric Logic

Combinatoric Logic

Combinatoric Logic

MGTs

DSP tiles

Resources

Software & development

- **Cross-development…**
  - **GNU cross-development environment (C & C++)**
  - **remote (network) GDB debugger**
  - **network console**
- **Operating system support…**
  - **Bootstrap loader**
  - **Open Source Real-Time kernel (RTEMS)**
    - **POSIX compliant interfaces**
    - **Standard I/P network stack**
  - **Exception handling support**
- **Object-Oriented emphasis:**
  - **Class libraries (C++)**
    - **DEI support**
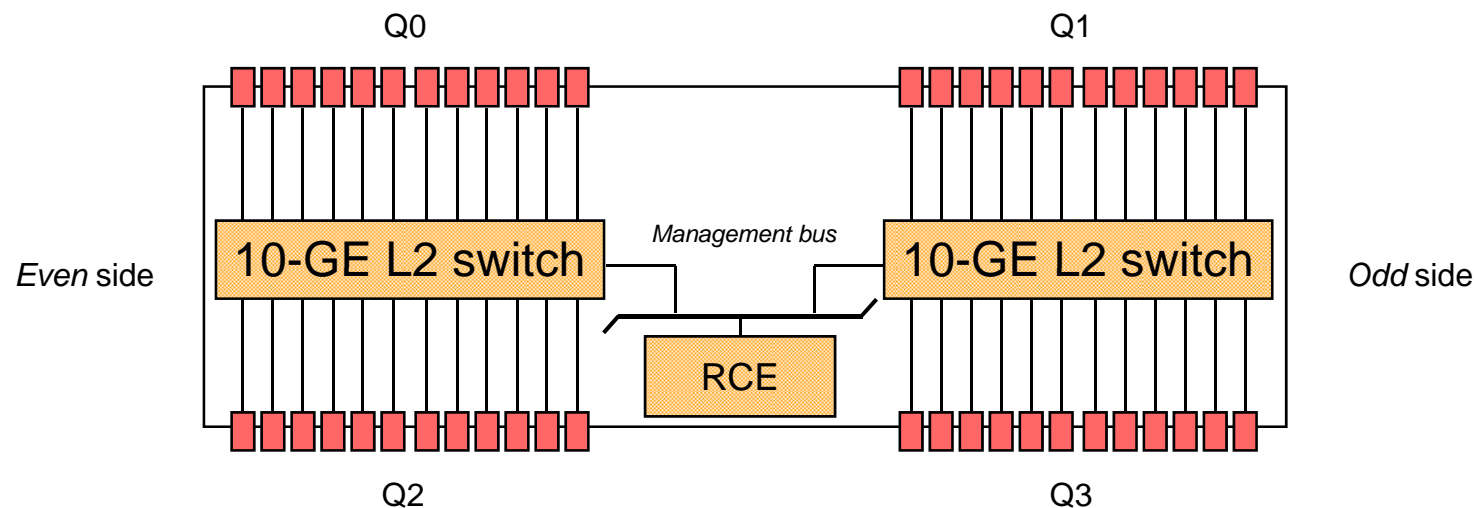   - **Configuration Interface**

Resources

- **Multi-Gigabit Transceivers (MGTs)**
  - up to 12 channels of:
    - SER/DES
    - input/output buffering
    - clock recovery
    - 8b/10b encoder/decoder
    - 64b/66b encoder/decoder
  - each channel can operate up to 6.5 gb/s
  - channels may be bound together for greater aggregate speed
- Combinatoric logic
    - gates
    - flip-flops (block RAM)
    - I/O pins
- DSP support
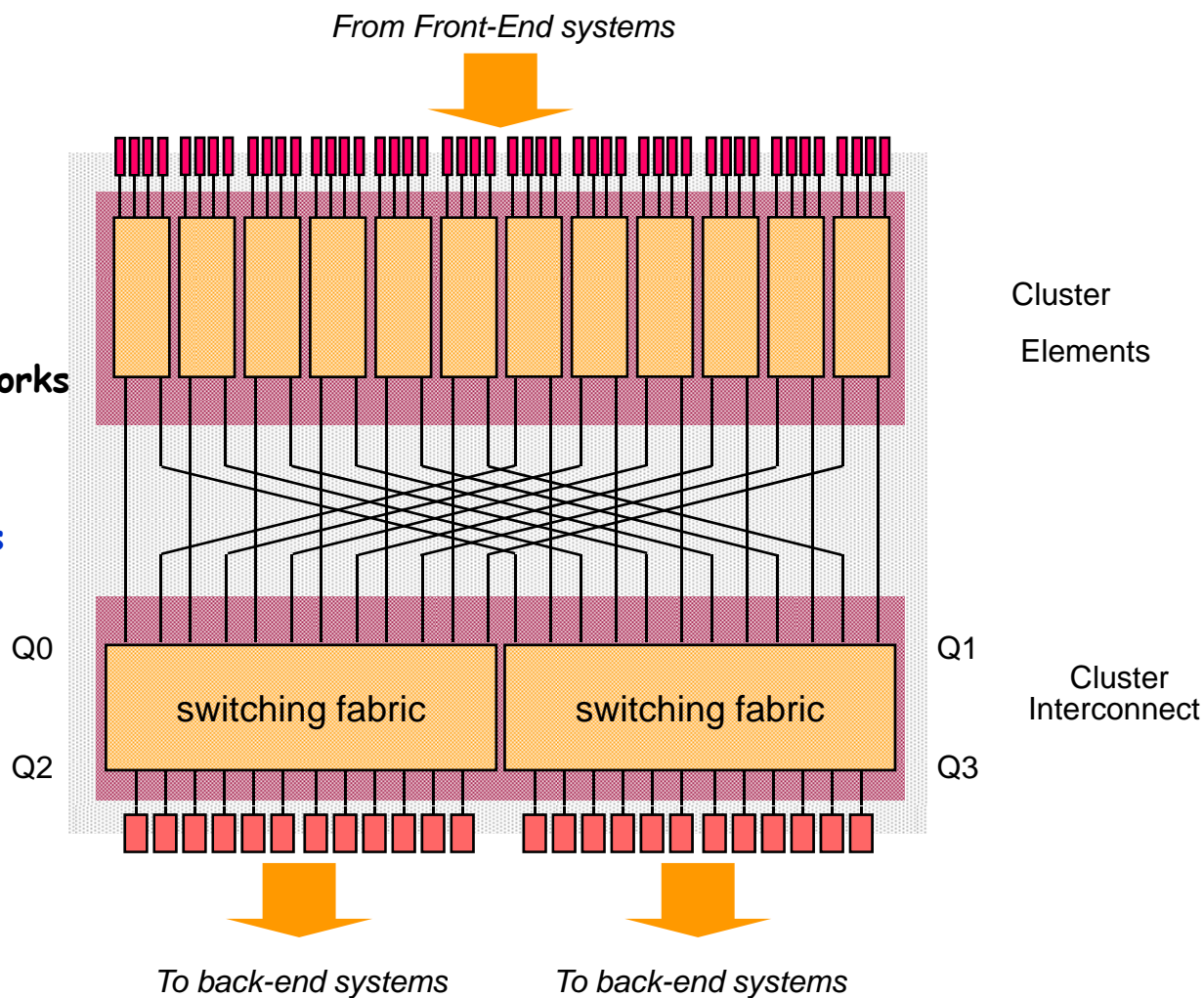 - contains up 192 Multiple-Accumulate-Add (MAC) units

DX "Plug-Ins"

PGP plug-in
(3.125 gb/s)

Ethernet MAC plug-in
XUAI (4 x 3.125 gb/s)

Boot Options

450 MHZ PPC-405
Processor

Cross-Bar

Memory Subsystem
512 MByte RLDRAM-II

Configuration
128 MByte Flash

The Cluster Interconnect (CI)

- **Based on two *Fulcrum* FM224s**
  - **24 port 10-GE switch**
  - **is an *ASIC* (packaging in 1433-ball BGA)**
  - **10-GE XAUI interface, however, supports multiple speeds…**
    - **100-BaseT, 1-GE & 2.5 gb/s**
  - **less then 24 watts at full capacity**
  - **cut-through architecture (packet ingress/egress < 200 ns)**
  - **full Layer-2 functionality (VLAN, multiple spanning tree etc..)**
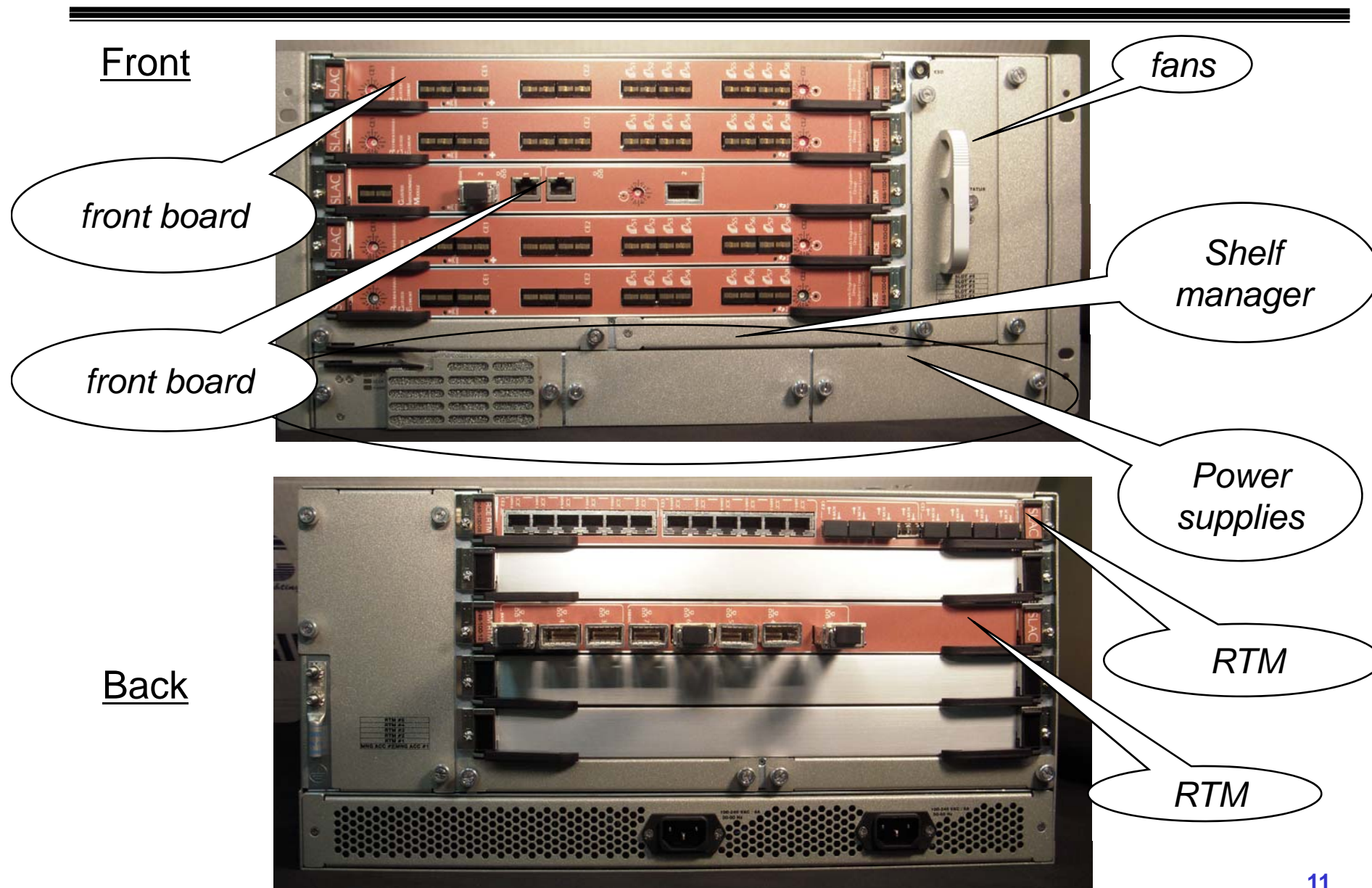 - **configuration can be managed or unmanaged**

A *cluster* of 12 elements

- **Two 12 node (element) networks**
- **Each *element*:**
  - **manages 4 inputs**
  - **connects to 2 networks**
- **Each CI *network***
  - **switches 12 elements**
  - **has symmetric output**

From Front-End systems

Cluster

Elements

Q0

switching fabric

switching fabric

Q1

Cluster
Interconnect

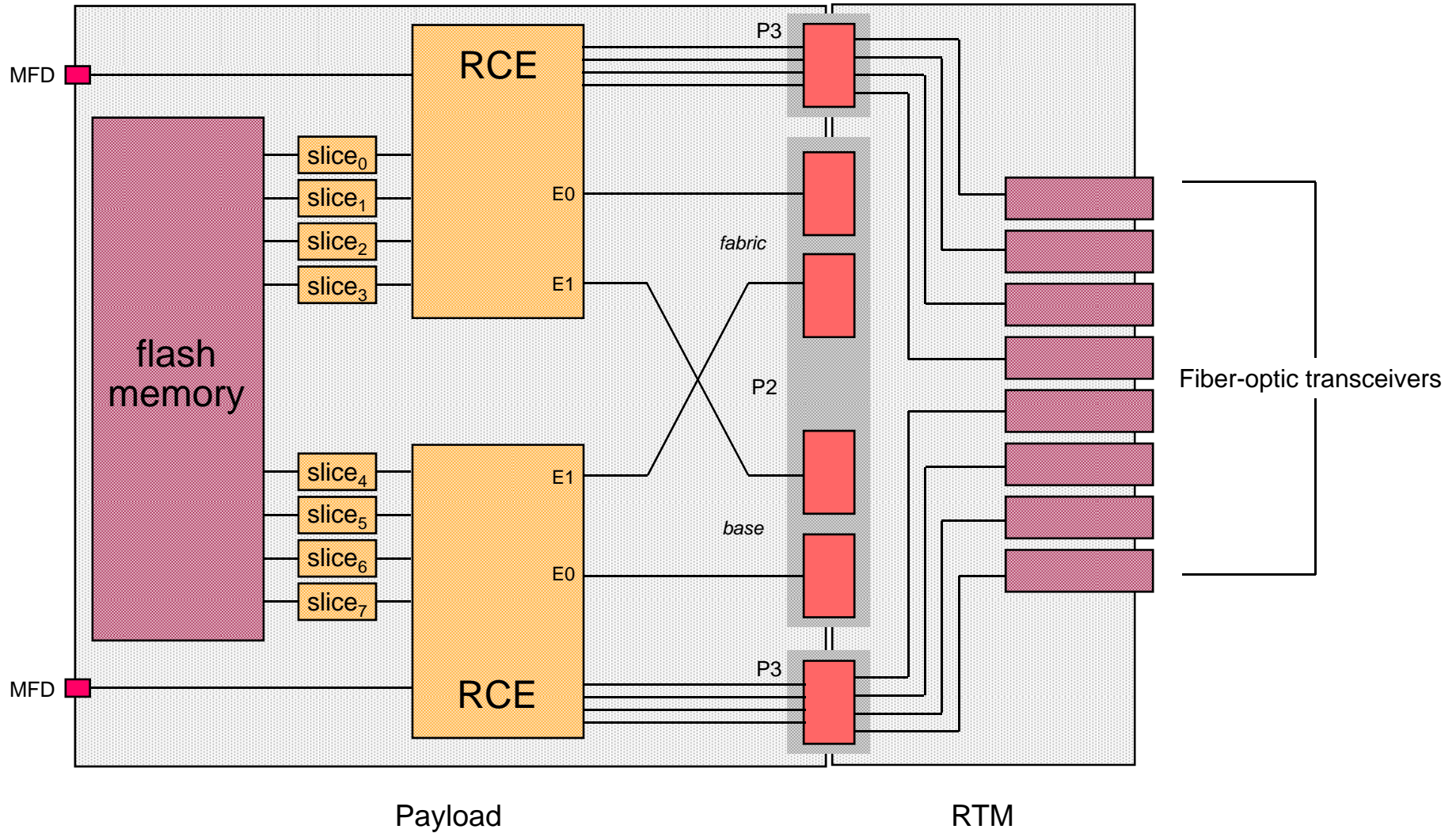Q2

Q3

To back-end systems

To back-end systems

Why ATCA as a packaging standard?

- An emerging *telecom* standard…
  - see previous talk given at ATLAS Upgrade week
- Its attractive features:
  - backplane & packaging available as a *commercial* solution
  - (relatively) generous form factor
    - 8U x 1.2" pitch
  - emphasis on High Availability
    - lots of redundancy
    - hot swap capability
    - well-defined environmental monitoring & control (IPM)
      - pervasive industry use
  - external power input is low voltage DC
    - allows for rack aggregation of power
- Its <u>very</u> attractive features (as substrate for RCE & CI):
  - the concept of a Rear Transition Module (RTM)
    - allows all cabling on rear (module removal without interruption of cable plant)
    - allows separation of data interface from the mechanism used to process that data
  - high speed serial backplane
    - protocol agnostic
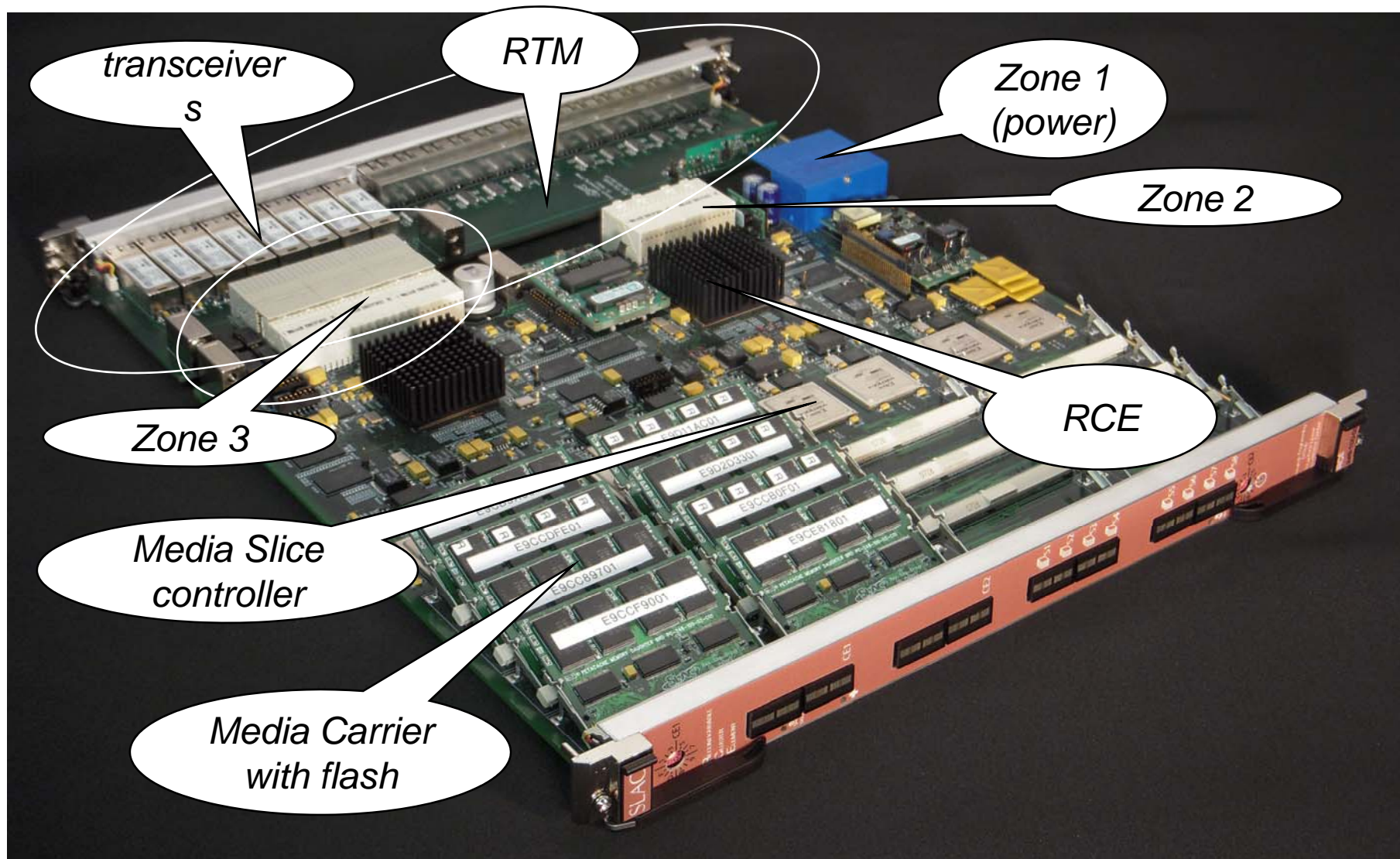   - provision for different interconnect topologies
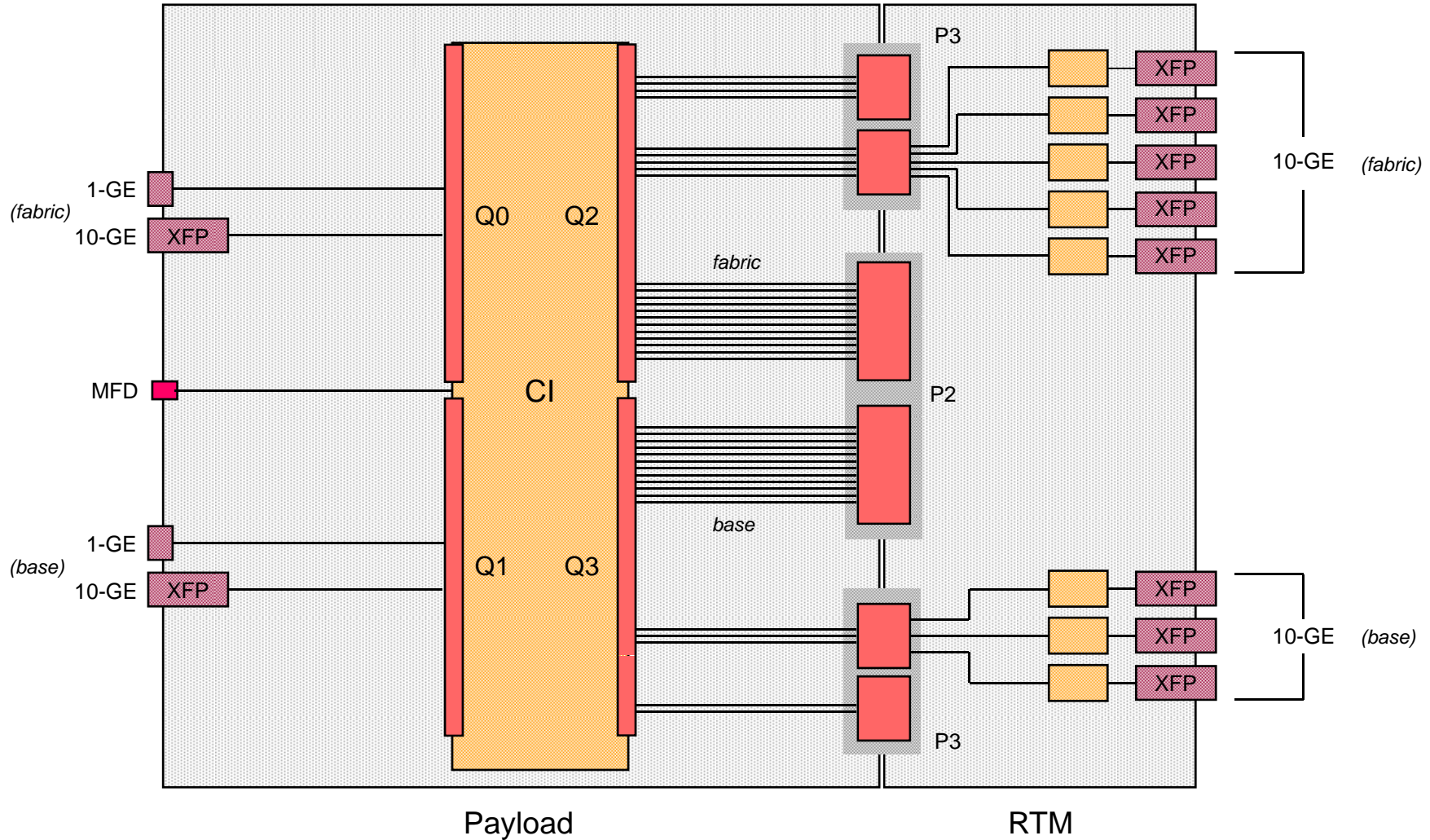
Typical (5 slot) shelf

Front

fans

front board

Shelf manager

front board

Power supplies

Back

RTM

RTM

RCE board + RTM (Block diagram)

Payload

RTM

RCE board + RTM

transceivers

RTM

Zone 1 (power)

Zone 2

Zone 3

RCE
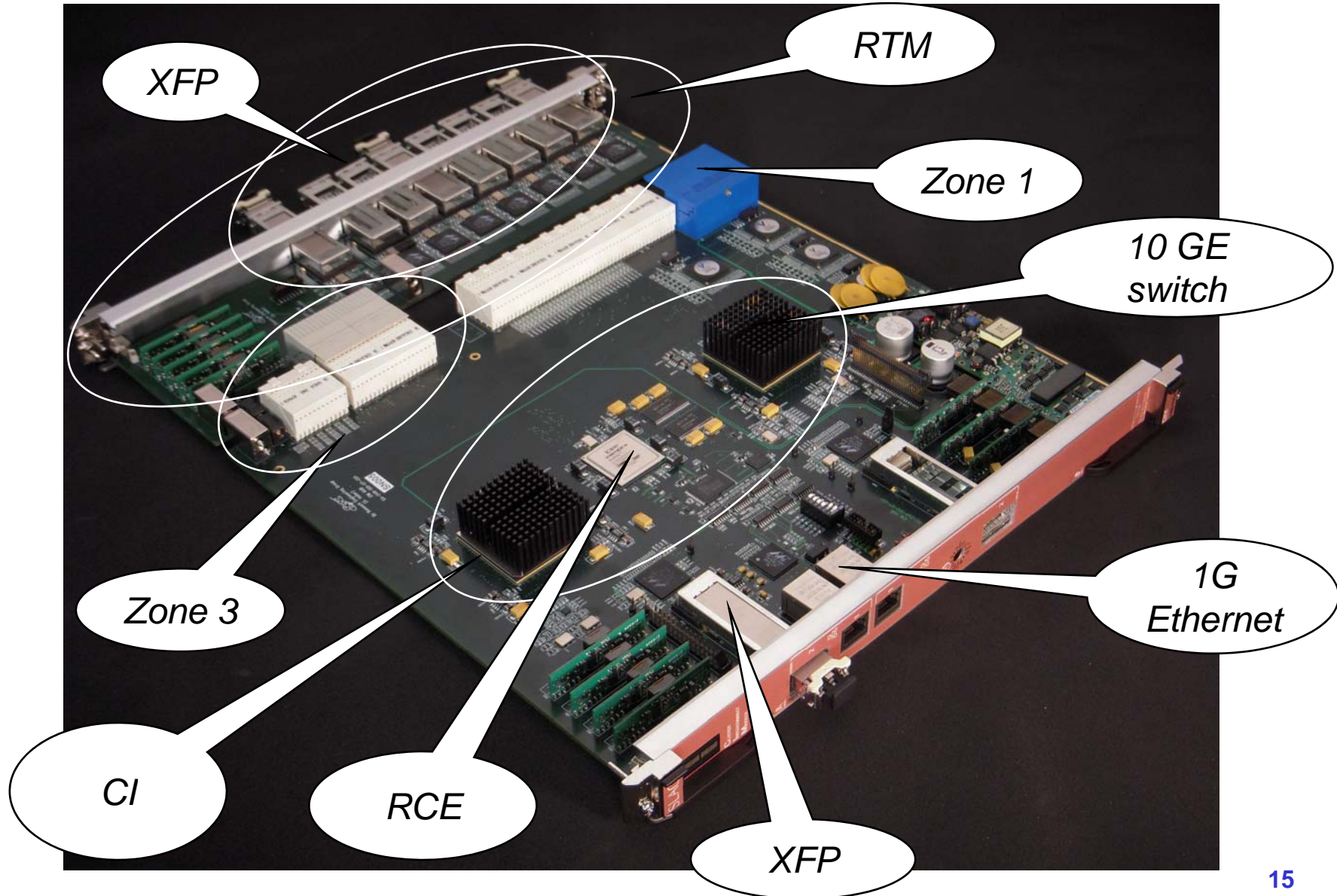
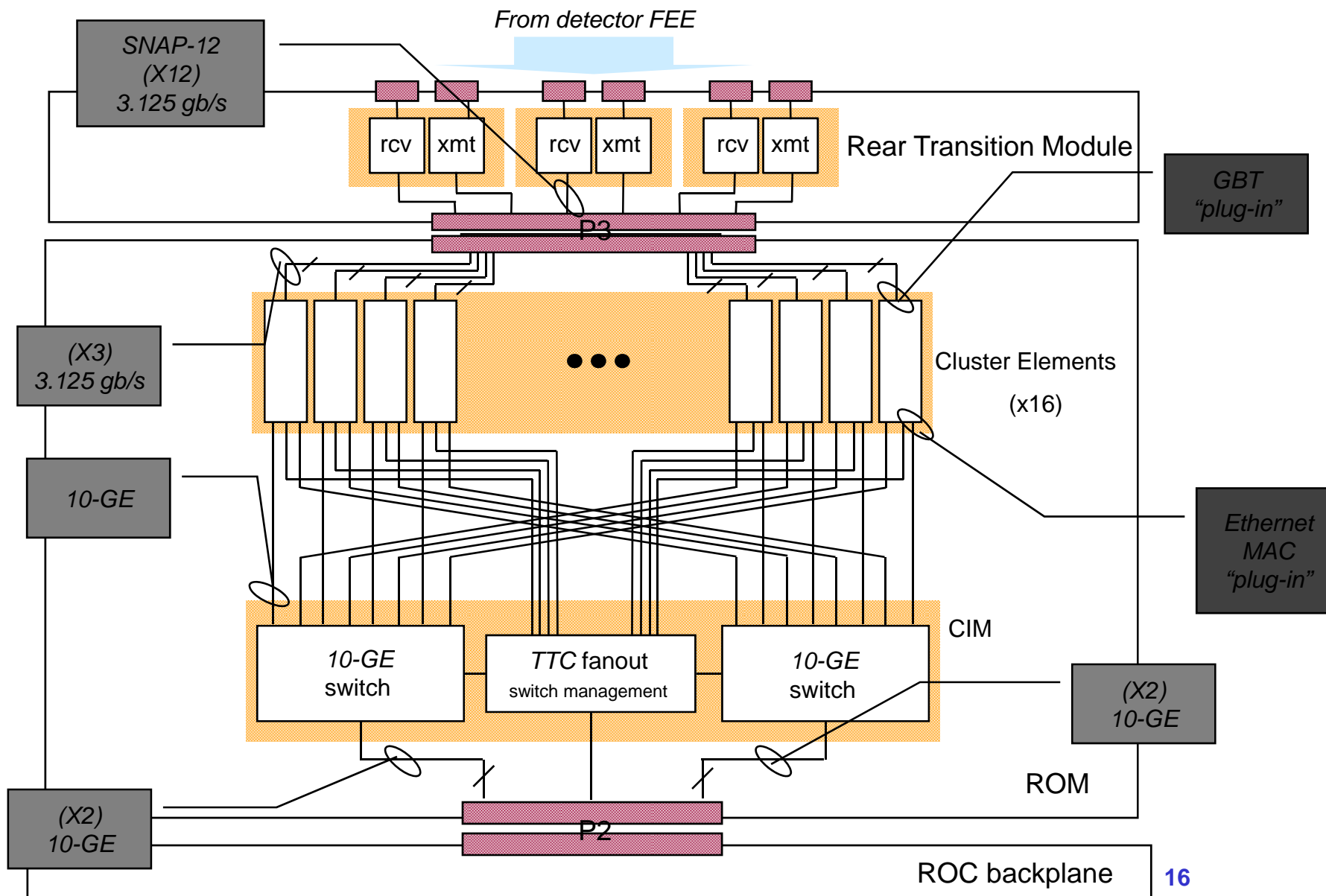Media Slice controller

Media Carrier with flash

Cluster Interconnect board + RTM (Block diagram)
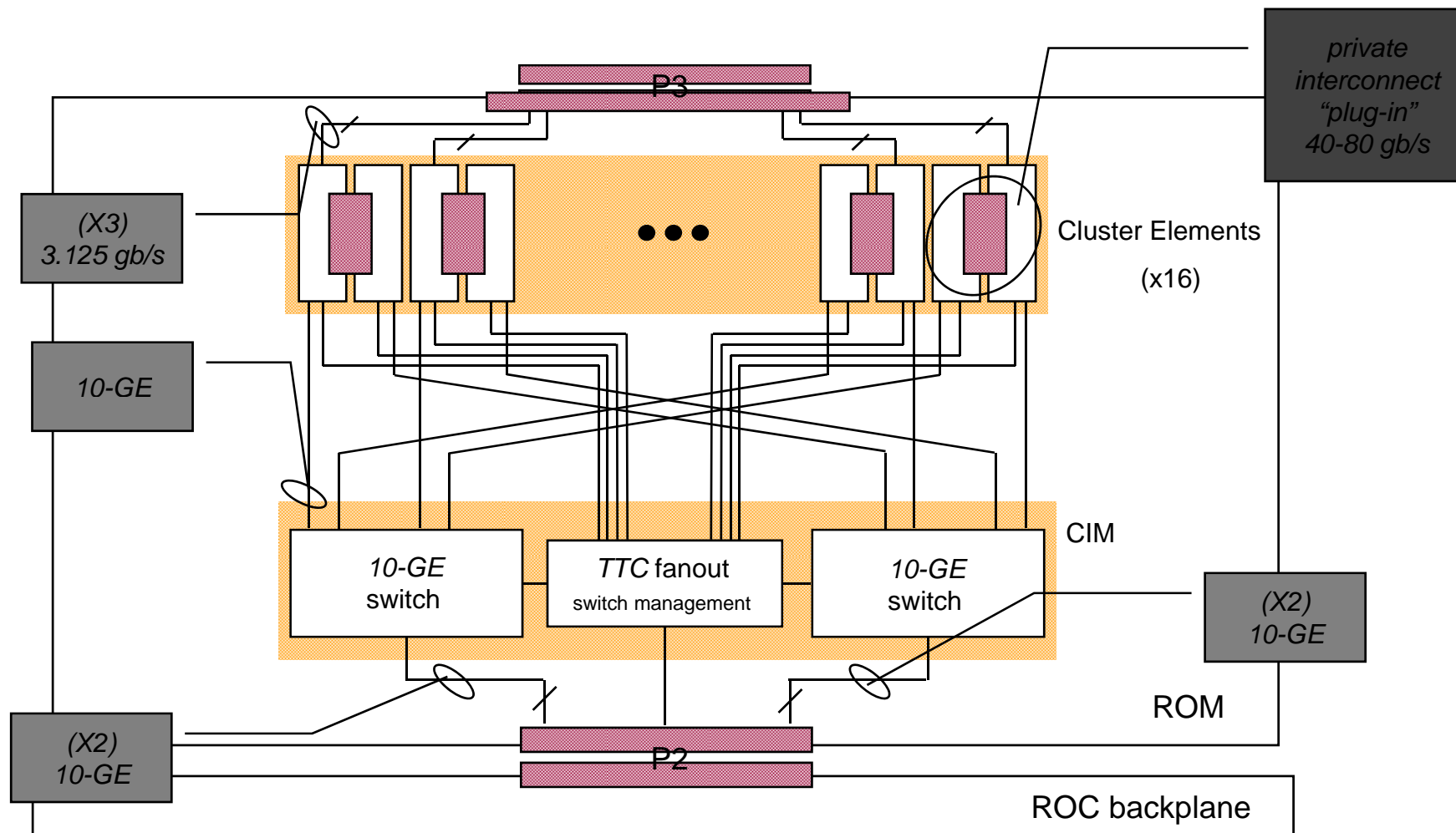

Payload

RTM

Cluster Interconnect board + RTM

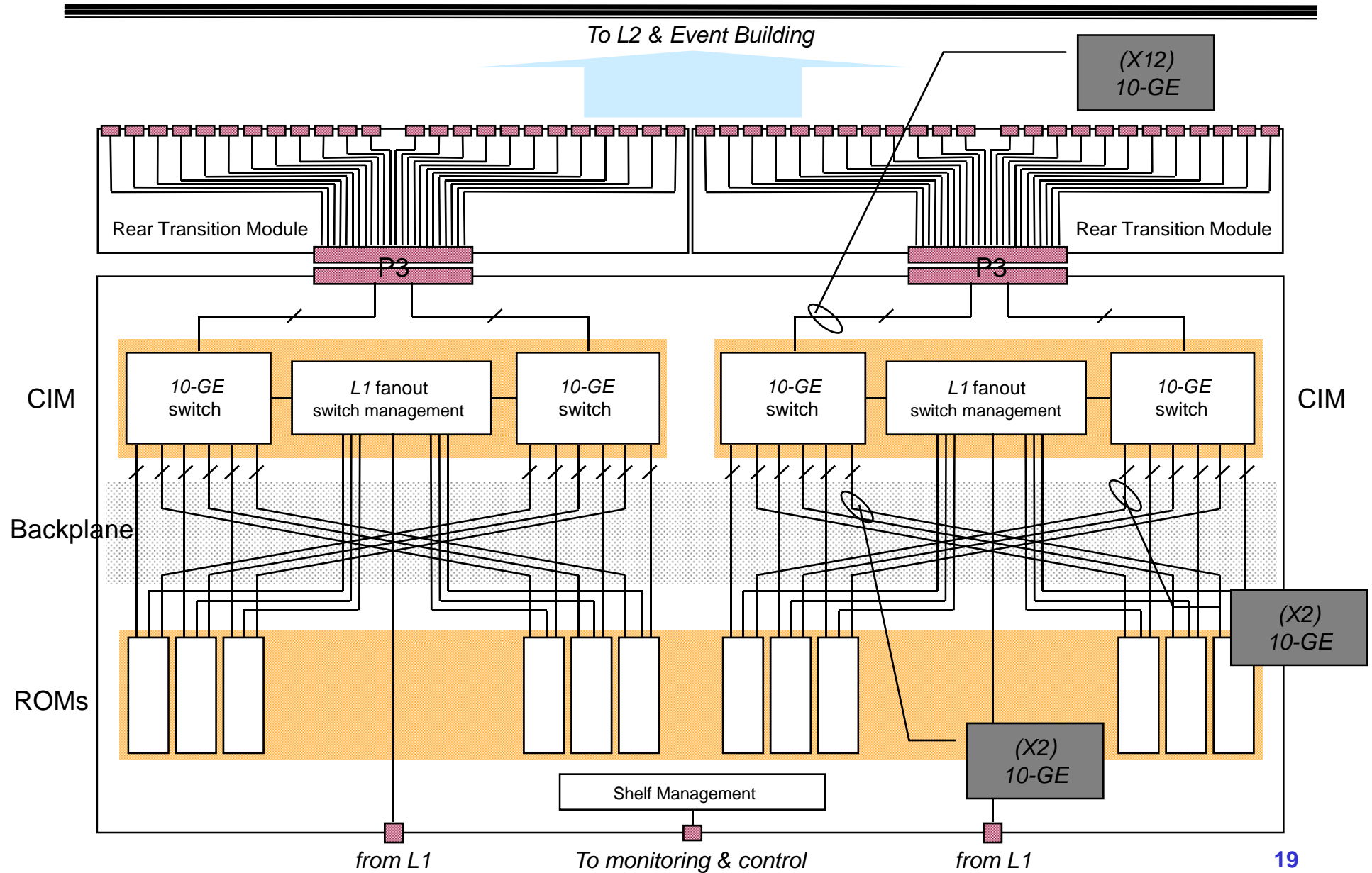"Tight" coupling (48-channel) ROM (Read-Out-Module)

"Loose" coupling (24-channel) ROM

"Tight" versus "Loose" coupling

- **Commonalities:**
  - **Two physically, disjoint networks (1 for subsystem, 1 for TDAQ)**
  - **Board output is 4 GBytes/sec (2 for subsystem, 2 for TDAQ)**
    - (at some cost) could be doubled with 10 gb/s backplane
  - **Same GBT & Ethernet MAC "plug-ins"**
  - **ROD function manages 3 channels/element**
- **Differences:**
  - **"Tight"**
    - One element shares <u>both</u> ROD & ROS functionality
    - Coupling is implemented through a <u>software</u> interface
    - One element connects to <u>both</u> (Subsystem & TDAQ) networks
  - **"Loose"**
    - One element for ROD functions & one element for ROS functions
    - Coupling is implemented through a <u>hardware</u> interface
    - One element connects to <u>one</u> (Subsystem or TDAQ) network
- **While attractive *loose* "costs" <u>twice</u> as much as *tight*...**
  - **$$$**
  - **footprint**
  - **power**

Read-Out-Crate (ROC)

To L2 & Event Building

(X12)
10-GE

Rear Transition Module

Rear Transition Module

P3

P3

CIM

CIM

10-GE switch

L1 fanout switch management

10-GE switch

10-GE switch

L1 fanout switch management

10-GE switch

Backplane

(X2)
10-GE

ROMs

(X2)
10-GE

Shelf Management

from L1

To monitoring & control

from L1

Summary

- SLAC is positioning itself for a new generation of DAQ…
  - strategy is based on the idea of modular building blocks
    - inexpensive computational elements (the RCE)
    - interconnect mechanism (the CI)
    - industry standard packaging (ATCA)
  - architecture is now relatively mature
    - both demo boards (& corresponding RTMs) are functional
    - RTEMS ported & operating
    - network stack fully tested and functional
  - performance and scaling meet expectations
  - costs have been established (engineering scales):
    - ~$1K/RCE (goal is less then $750)
    - ~$1K/CI (goal is less then $750)
  - documentation is a "work-in-progress"
  - public release is pending Stanford University "licensing issues"
- This technology strongly leverages off recent industry innovation, including:
  - System-On-Chip (SOC)
  - High speed serial transmission
  - low cost, small footprint, high-speed switching (10-GE)
  - packaging standardization (serial backplanes and RTM)
- Gained experience with these innovations will itself be valuable…

- This technology offers a ready-today vehicle to explore both alternate architectures & different performance regimes

- Thanks for the many valuable discussions:
  - Benedetto Gorini, Andreas Kugel, Louis Tremblet, Jos Vermeulen, Mimmo della Volpe (ROS)
  - Fred Wickens (RAL)
  - Bob Blair, Jinlong Zhang (Argonne)
  - Rainer Bartoldus & Su Dong (SLAC)