

***Construction of high speed,
massively parallel, ATCA
based Data Acquisition
Systems using modular
components***

Gregg Thayer

Representing:

Richard Claus, Gunther Haller, Ryan Herbst, Michael Huffer,
Chris O'Grady, Jim Panetta, Amedeo Perazzo, Steve Tether,
Matt Weaver

Introduction

- Develop technology to construct DAQ/Trigger systems for HEP experiments from generic, inexpensive, scalable components
 - No commercial solutions exist which satisfy requirements
 - Large amount of I/O required between computational elements
 - Varied and often unique I/O required at interfaces
 - HEP DAQ systems are almost universally custom built, but share a great deal of architectural commonality
 - Part of an ongoing SLAC detector R&D project
- Phase 0 (Design)
 - Survey of requirements and capture commonality
 - Leverage recent industry innovation
 - Design system elements
- Phase 1
 - Technology and hardware demonstration
- Phase 2
 - Useful, sustainable architecture
- Phase 3
 - Upgrade to achieve performance goals

Phase 0: Three Building Blocks

Requirements:

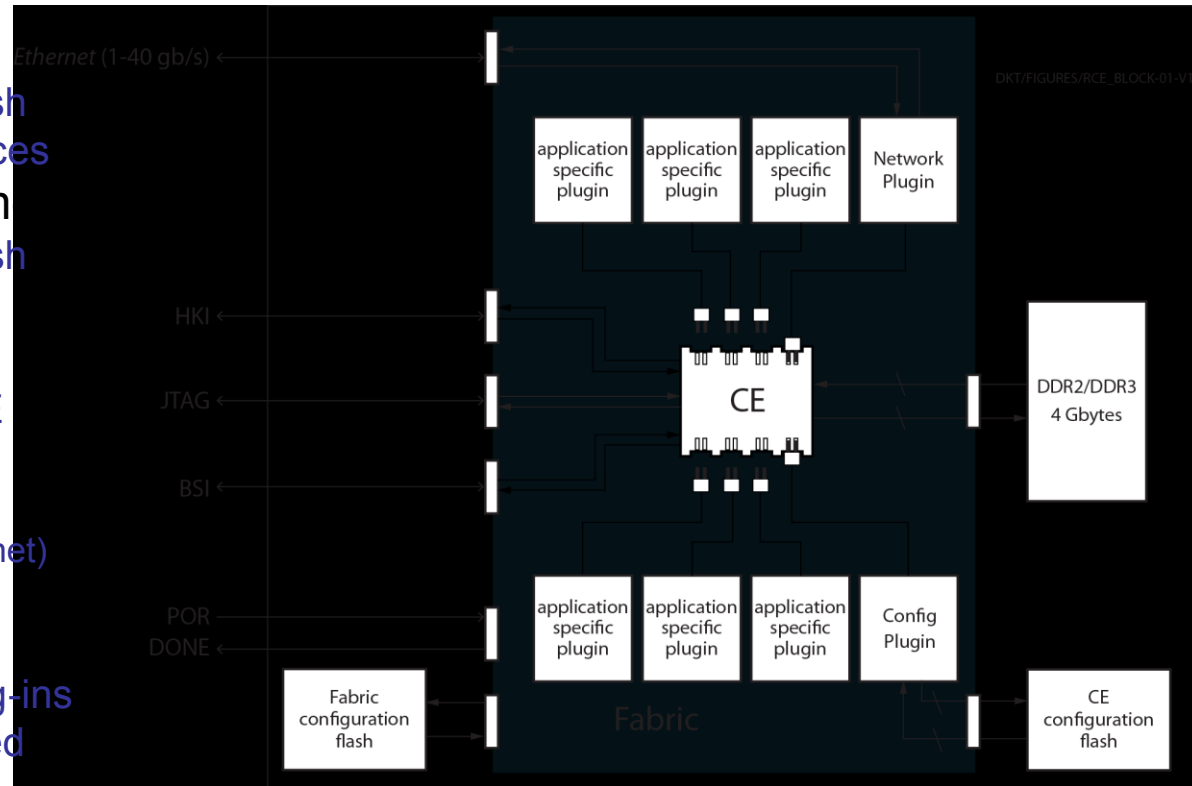
- Computational Elements
 - Low cost
 - Variety of computational models
 - Flexible and permanent I/O
- Element Interconnect
 - Low cost
 - Low-latency/high-bandwidth I/O
 - Based on industry protocol
 - Variety of interconnect topologies
- Packaging Solution
 - High availability
 - Scaling
 - Different physical I/O interfaces
 - Prefer commercial standard

Solutions:

- Reconfigurable Cluster Element (RCE)
 - Xilinx Virtex and Zynq System-on-chip
- Cluster Interconnect (CI)
 - Based on 10 G Ethernet switching
 - Fulcrum Tahoe and eventually Alta
- AdvancedTCA
 - Emphasis on High Availability
 - High speed serial backplane
 - Rear Transition Module (RTM)

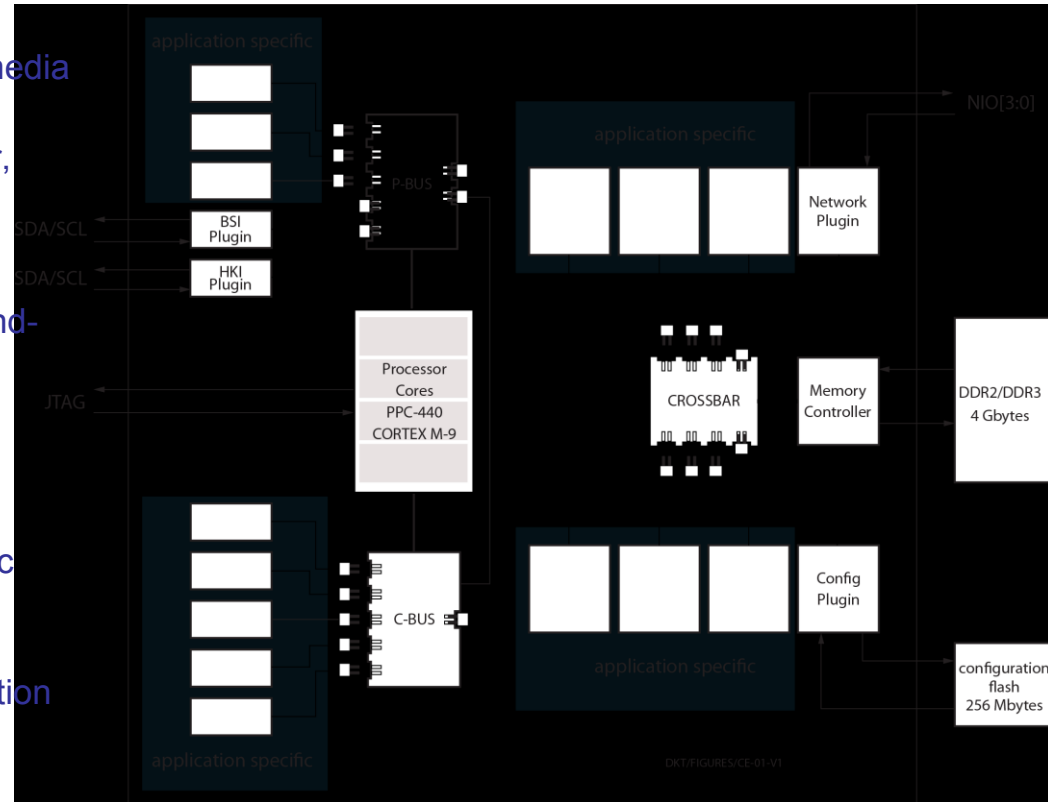
The RCE

- Programmable Fabric
 - Paired with configuration flash
 - Soft and Hard silicon resources
- Programmable Cluster Element
 - Paired with configuration flash
 - RAM DDR3 up to 4 GB
- Plug-ins
 - Glues fabric resources to CE
 - Built from fabric resources
 - Prebuilt
 - Network (1-40 Gbit/s Ethernet)
 - Interface to CE flash
- CE supports eight plug-ins
 - Two prewired to prebuilt plug-ins
 - Six can be application defined
- Hardened resources
 - DSP tiles nearly 1 GMACS
 - SerDes (+ support) up to 10.3 Gbit/s



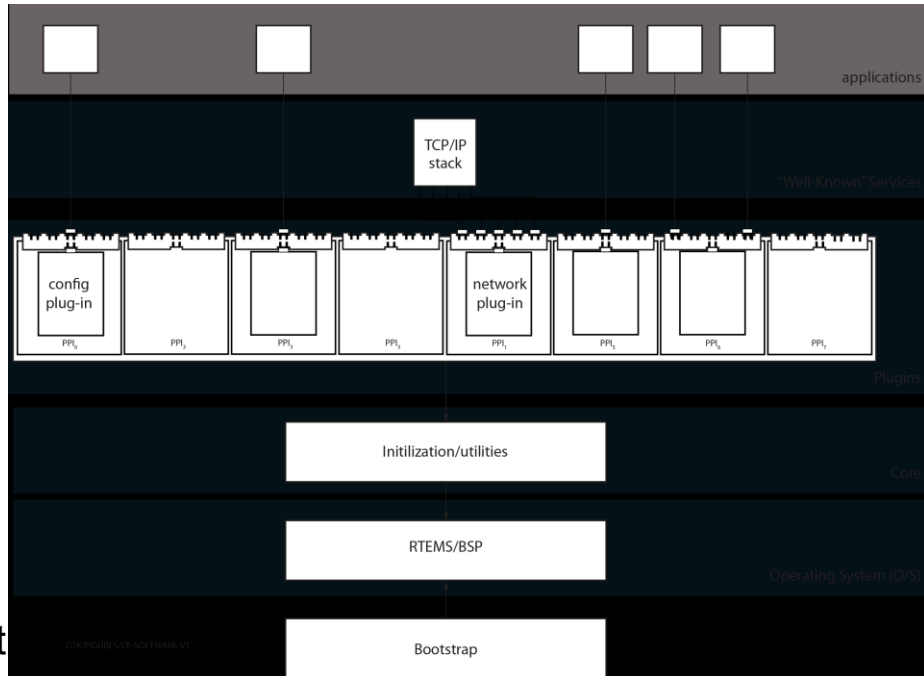
The CE

- Processor (multi-core)
 - Up to 5000 DMIPS
 - Code and Data stored on configuration media
- Crossbar
 - Provides connections between processor, memory, and plug-ins
 - Over 8 GB/s of switching bandwidth
- Peripheral Bus (P-BUS)
 - With BSI and bootstrap provides “Plug-and-Play” support for plug-ins
- Command Bus (C-BUS)
 - Allows processor interaction with plug-in concurrent with its memory transfers
 - Extends the processors instruction set
 - Provides for user application specific logic
- Bootstrap Interface (BSI)
 - I²C slave interface
 - Allows external control and parameterization of boot process
- Housekeeping Interface (HKI)
 - I²C master interface
 - Allows external configuration, control, and monitoring of “slow” devices



CE as a Software System

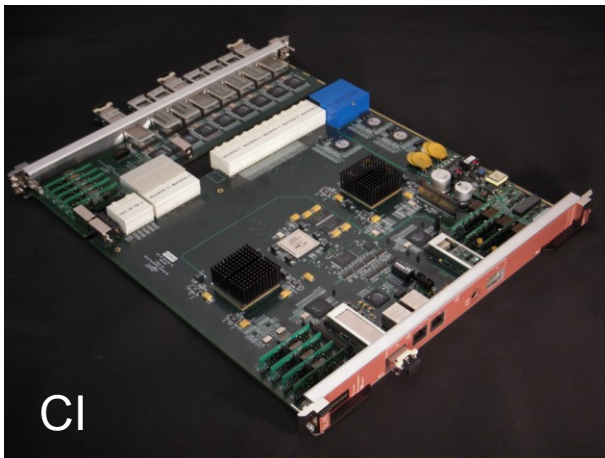
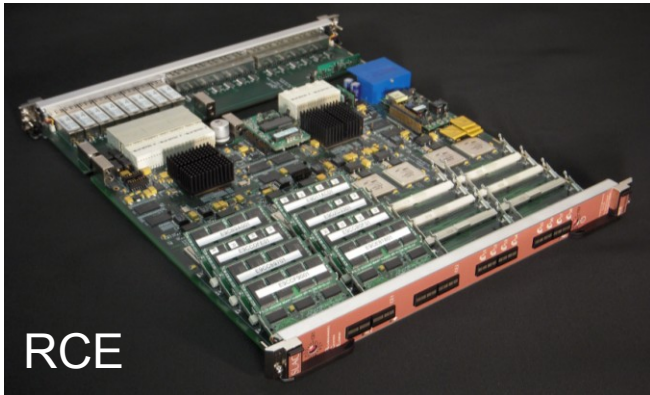
- Primary Bootstrap
 - Operating System agnostic
 - Driven by BSI
- Operating System and BSP
 - Open-Source Real-Time kernel (RTEMS)
- Core
 - CE initialization
 - Plug-in support
 - Utilities
- Plug-ins
 - Driven by the set of physical plug-ins present
- Well Known Services (WKS)
 - Standard BSD Network (IP) stack
 - Telnet, GDB stub, NFS, etc...
 - Customizable
- Standard GNU cross-development environment
 - Includes remote GDB debugger
 - All software is object-oriented with C and C++ support



The CI and Cluster

- The CI consists of
 - 96 channel 10G-Ethernet switch
 - Partitioned into 24 ports of 4 channels
 - Each port can be configured as
 - 1-4 10G (KR)
 - 1 40G (KR4)
 - 1 4x3.125G (XAUI)
 - 1 SGMII (10, 100, 1000 Base-T)
 - Cut through, industry lowest (200-300 ns) hop latency
 - Full Layer 2 and Layer 3 switching
 - One RCE
 - Manages switch (supports all management protocol)
 - Has fixed connection to one port of the switch
- The Cluster consists of
 - One CI
 - One or more (up to 96) RCEs

Phase 1: Hardware Demonstration



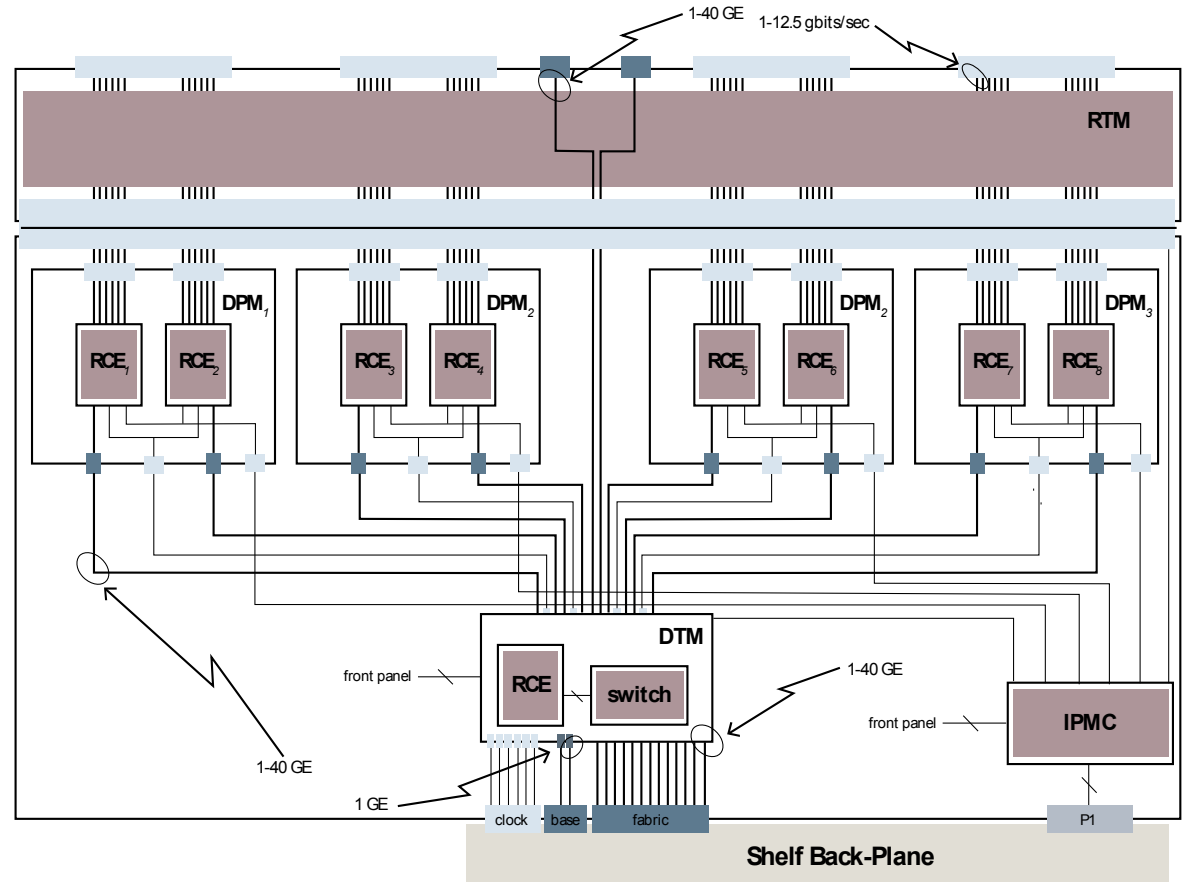
Phase 1 RCE and CI boards are in use at Linac Coherent Light Source

Phase 2: Cluster on Board (COB)

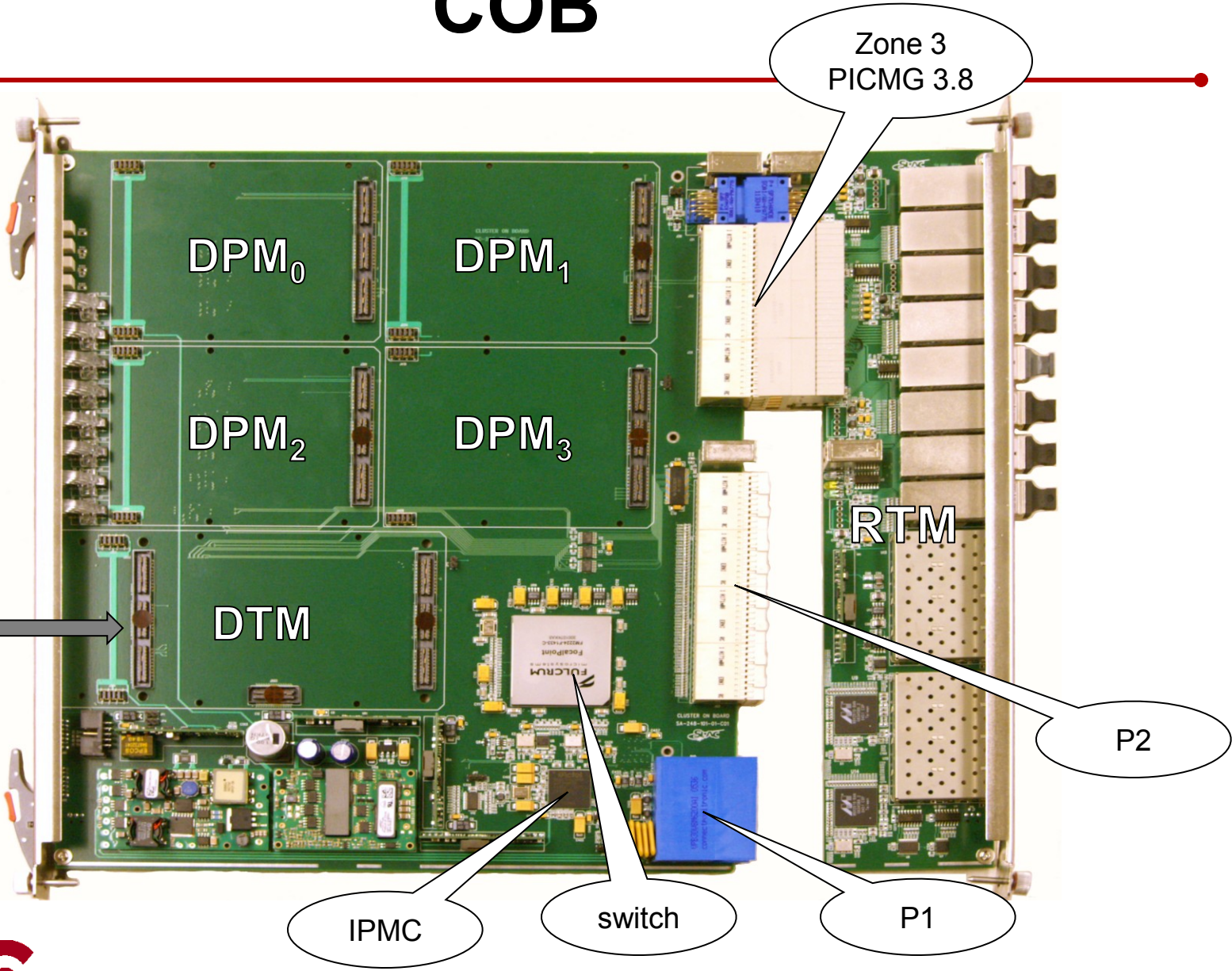
- Redesign the RCE and CI
 - Re-factor the RCE and CI boards as mezzanine boards
 - Decouple the RCE and CI from choice of packaging standard
 - Ensure adequate room for user logic in RCE FPGA fabric
 - Improve support for multiple outstanding transactions in protocol plug-ins
- Create a new ATCA Front Board (COB)
 - Carrier for new RCE and CI mezzanine boards
 - Fully implements IPMI (configuration and monitoring)
 - Full Mesh backplane support
 - Applications now require only one type of board
 - Interoperability with any backplane topology
 - AdvancedTCA Rear Transition Module standard (PICMG 3.8)
 - From the ATCA for Physics working group
 - Complete interoperability with any type of ATCA board
 - Generic, synchronous Timing and Trigger signal distribution

COB Block Diagram

- Four Data Processing Modules (DPMs)
 - Process data from RTM
 - Can contain up to 8 RCEs
- One Data Transport Module (DTM)
 - Distributes 10G Ethernet Fabric and timing signals to DPMs
 - Contains CI
 - Mezzanine extends to front panel (unlike DPMs)
- IPM controller to monitor and configure the five zones and RTM



COB



Phase 3: Looking Ahead

- Hardware adjustments to achieve design performance
 - Introduce Zynq (ARM) based RCE mezzanine
 - Upgrade CI to 40G capable switch
 - Move to 10G/lane capable backplane
- The result is that one 14-slot shelf is capable of
 - Providing 112 RCEs
 - 112 10G Ethernet nodes capable of operating IP protocols
 - Nearly 500,000 DMIPS of generic processing
 - Over 100 TMACS of DSP processing
 - Nearly 500 GB of data buffering
 - Over 700 channels of detector input
 - Over 7 Tbit/s input bandwidth
 - 112 channels of 10G Ethernet output
 - Over 1 Tbit/s output bandwidth
 - 14 COB-internal Ethernets
 - Nearly 14 Tbit/s COB-internal switching bandwidth
 - One Full Mesh backbone connecting networks
 - Over 7 Tbit/s switching bandwidth between networks