Fermilab

Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

ProtoPRM: An FPGA-Based High Performance Pattern Recognition Memory Track Finder Mezzanine

J. Olsen[1], T. Liu[1], J. Wu[1], Z. Hu[1], Z. Xu[2]

[1] Fermi National Accelerator Laboratory, Batavia, Illinois U.S.A.
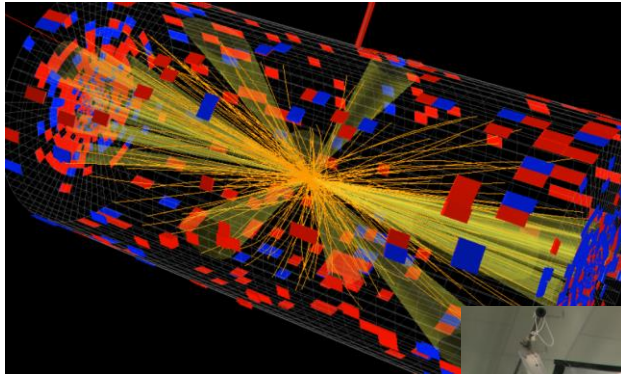[2] Peking University, Beijing, CHINA
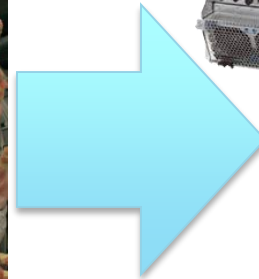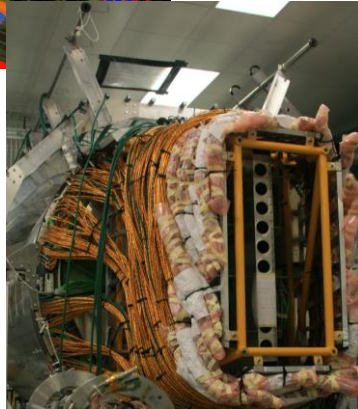
29 September 2016

Outline

- L1 Tracking Trigger Introduction and Challenges
- Pattern Recognition with Associative Memory
- Pulsar II Hardware Components
- Demonstration System
- Data Delivery
- ProtoPRM Mezzanine Firmware
  - Data Organizer
  - PRAM
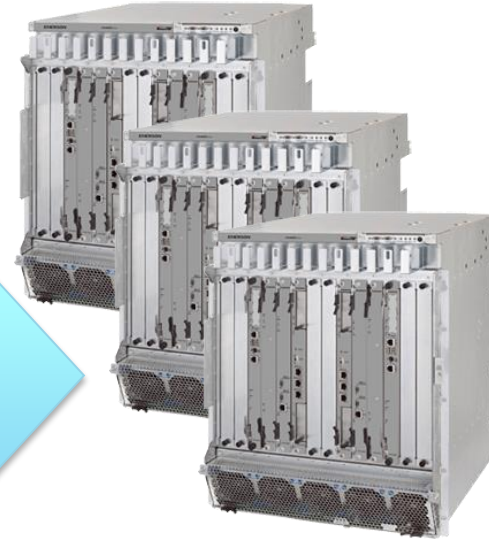- Conclusion

Level-1 Track Trigger Challenges

4μs

Data Source

- 15k modules/fibers
- 100 Tbps

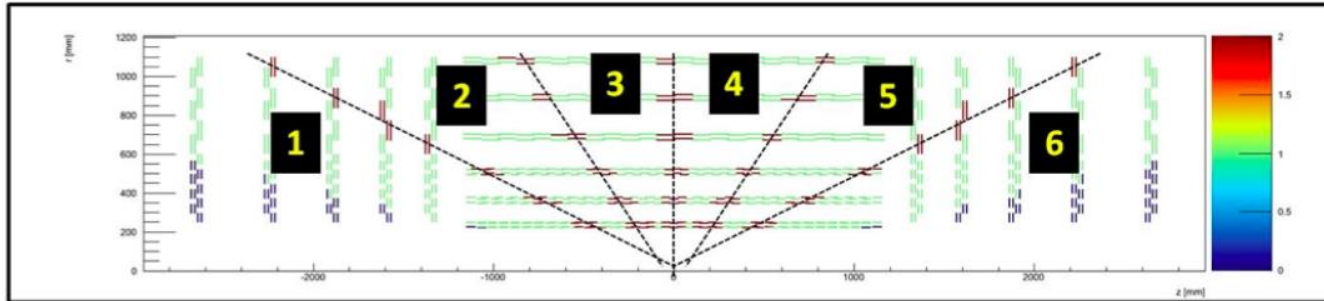**Data Formatting and Delivery**

- Partitioning

Data Processing

- Pattern Recognition (AM)
- Track Fitting (FPGA)

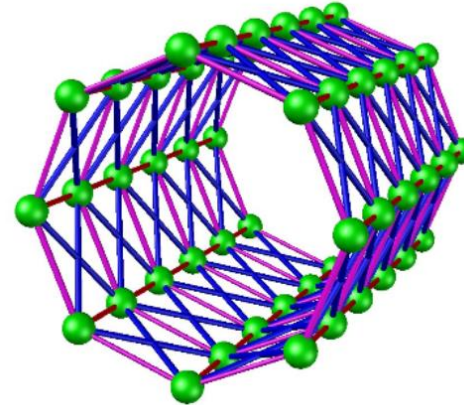Track Parameters
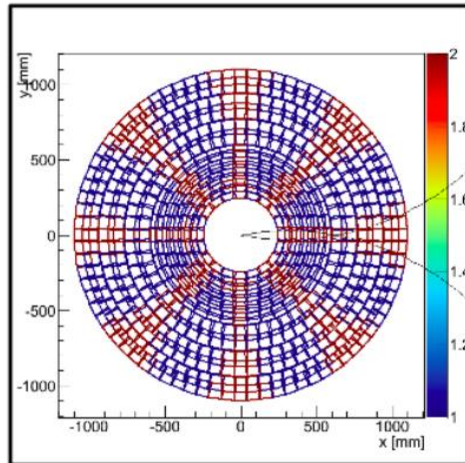
We use the divide and conquer approach...

🐾 Fermilab

Detector Partitioning

Use regional multiplexing => divide the detector into 6x8=48 trigger towers

6 regions in η

8 regions in r-phi

- 48 Trigger Towers
- ~400 front end modules/tower
- For demonstration purposes assume one processor shelf per tower

🔁 **Fermilab**

Time Multiplexing

- A trigger tower processor consists of an array of independent engines
  - Pattern Recognition (AM)
  - Track Fitting (FPGA)
- High speed, low latency, non-blocking communication channels for efficient data delivery
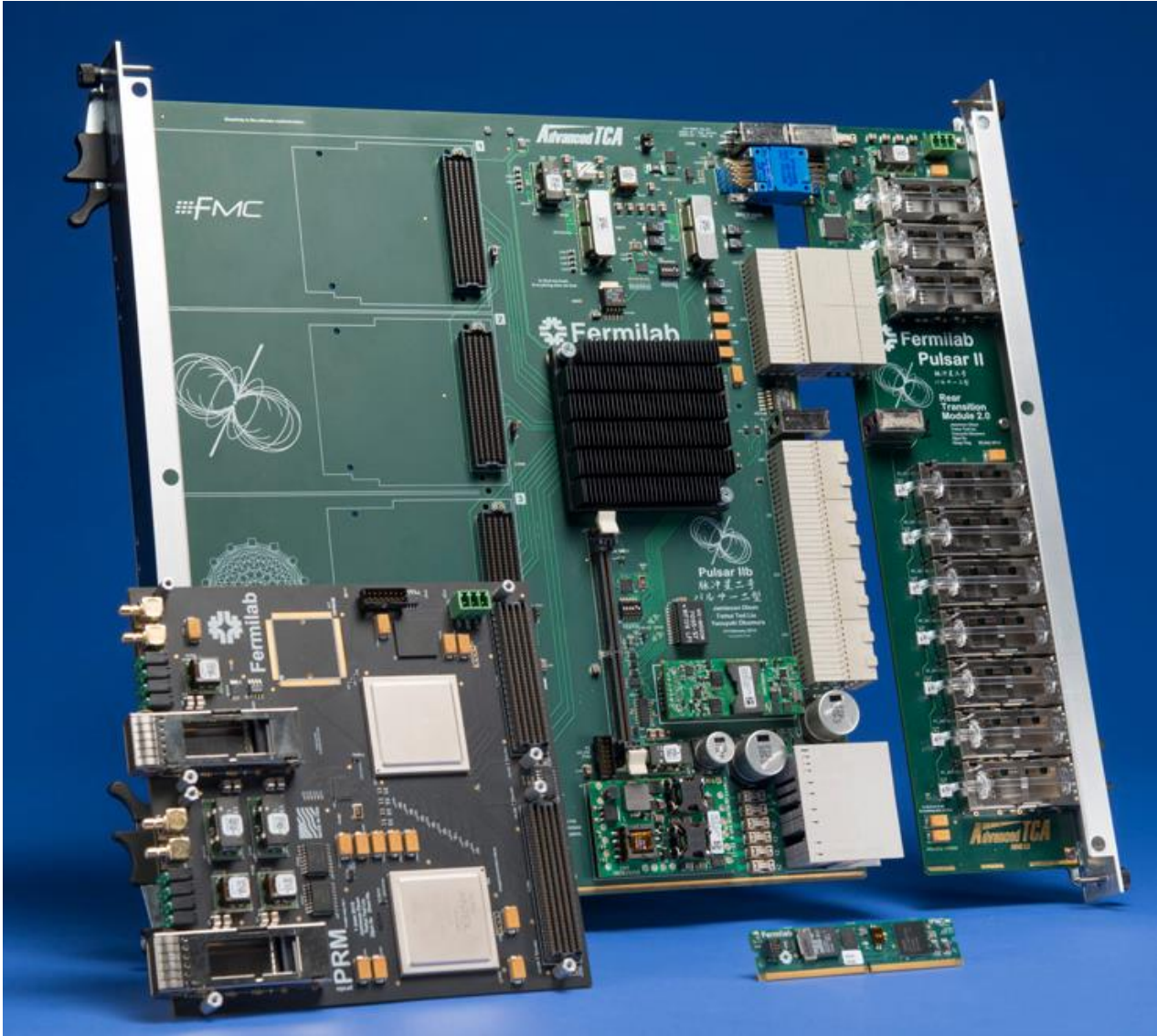- 20x time multiplexed
- Event rate 25ns → 500ns
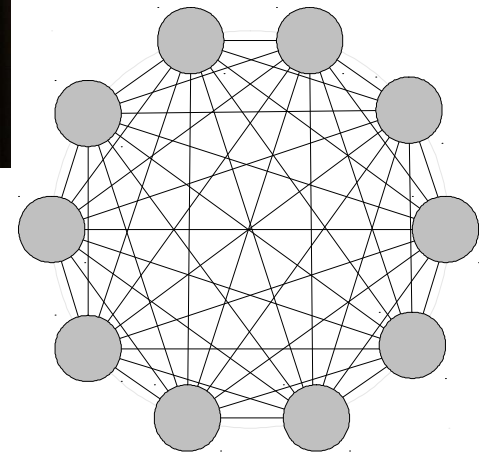
Fermilab

Pattern Recognition Associative Memory (PRAM)
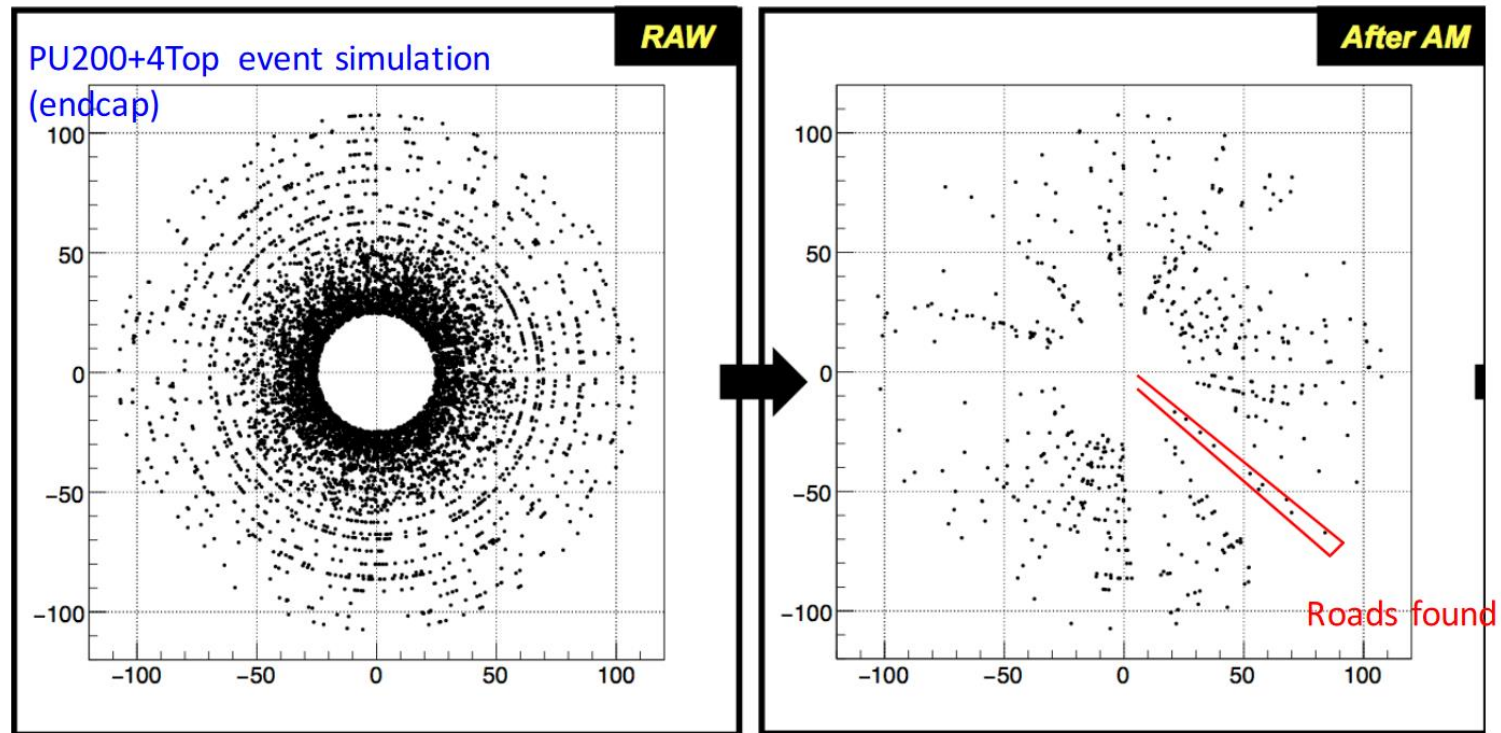
- Factor of ~10x occupancy reduction
- More importantly, the hits/stubs are organized in found roads ("hits of interest") which makes the track fitting easier

🔷 **Fermilab**

Pulsar II Hardware

Pulsar IIb Front Board

- Xilinx Virtex 7 FPGA
  - XC7VX690T -2 FFG1927 C
- 80 GTH serial transceivers
  - up to 11.3 Gbps (-2)
  - 40 for RTM
  - 28 for Full Mesh Fabric
  - 12 for Mezzanines
- Four FMC Mezzanine Cards
  - High Pin Count (HPC)
  - 35W, up to 60W possible
  - 34 pair LVDS/slot
  - 3 GTH lanes/slot
- Intelligent RTM / PICMG 3.8
- IPMC Mezzanine Card
- TTC timing and control over ATCA backplane
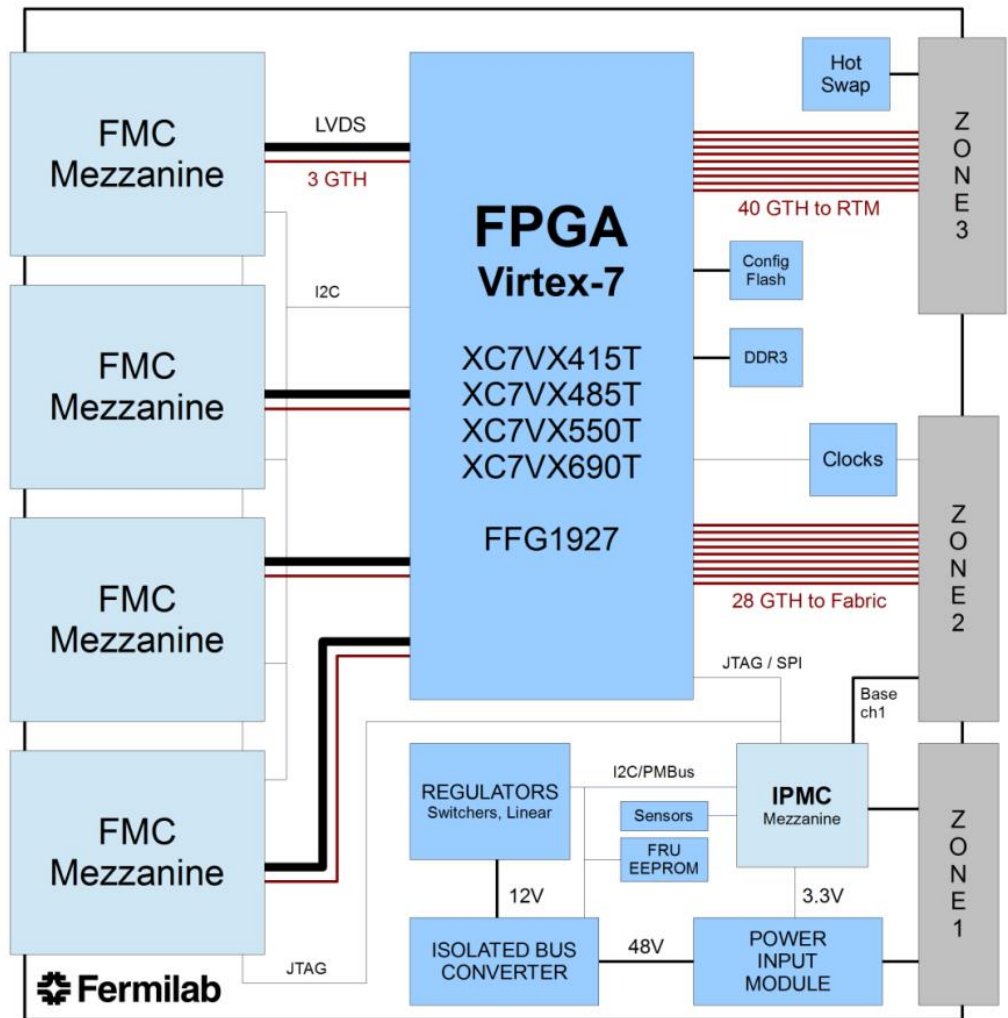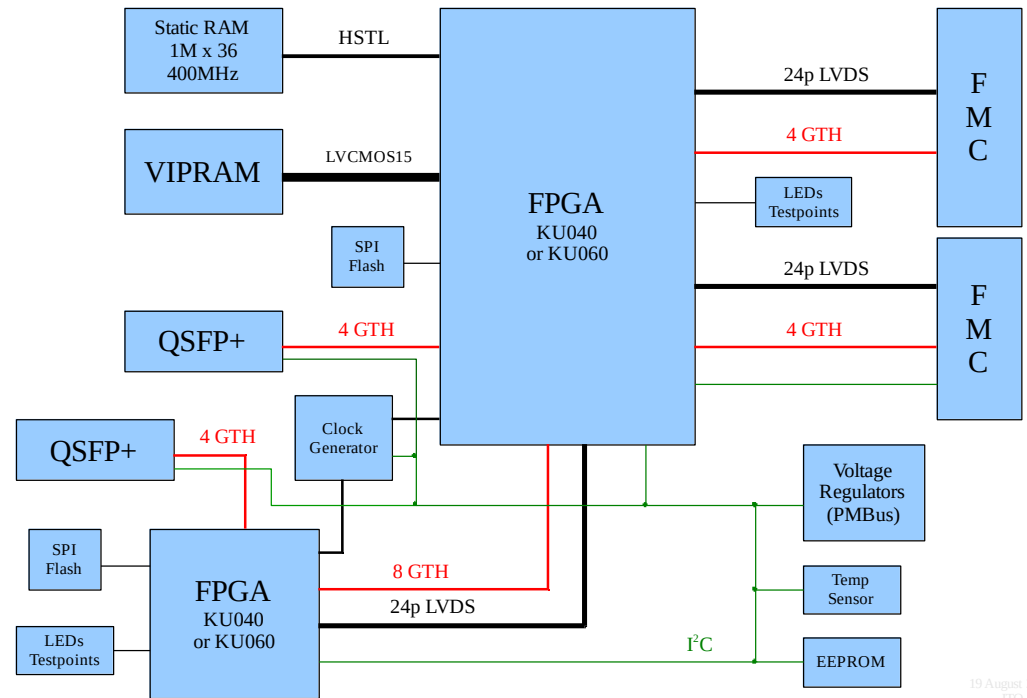
Rear Transition Module (RTM)

- 10 QSFP+ transceivers
- 400 Gbps full duplex
- ATCA/PICMG 3.8 spec
- MMC is a ARM Cortex-M3 micro
  - Read sensors, access QSFP registers
  - Basic IPMI functionality: hot swap, LEDs, handle, etc.

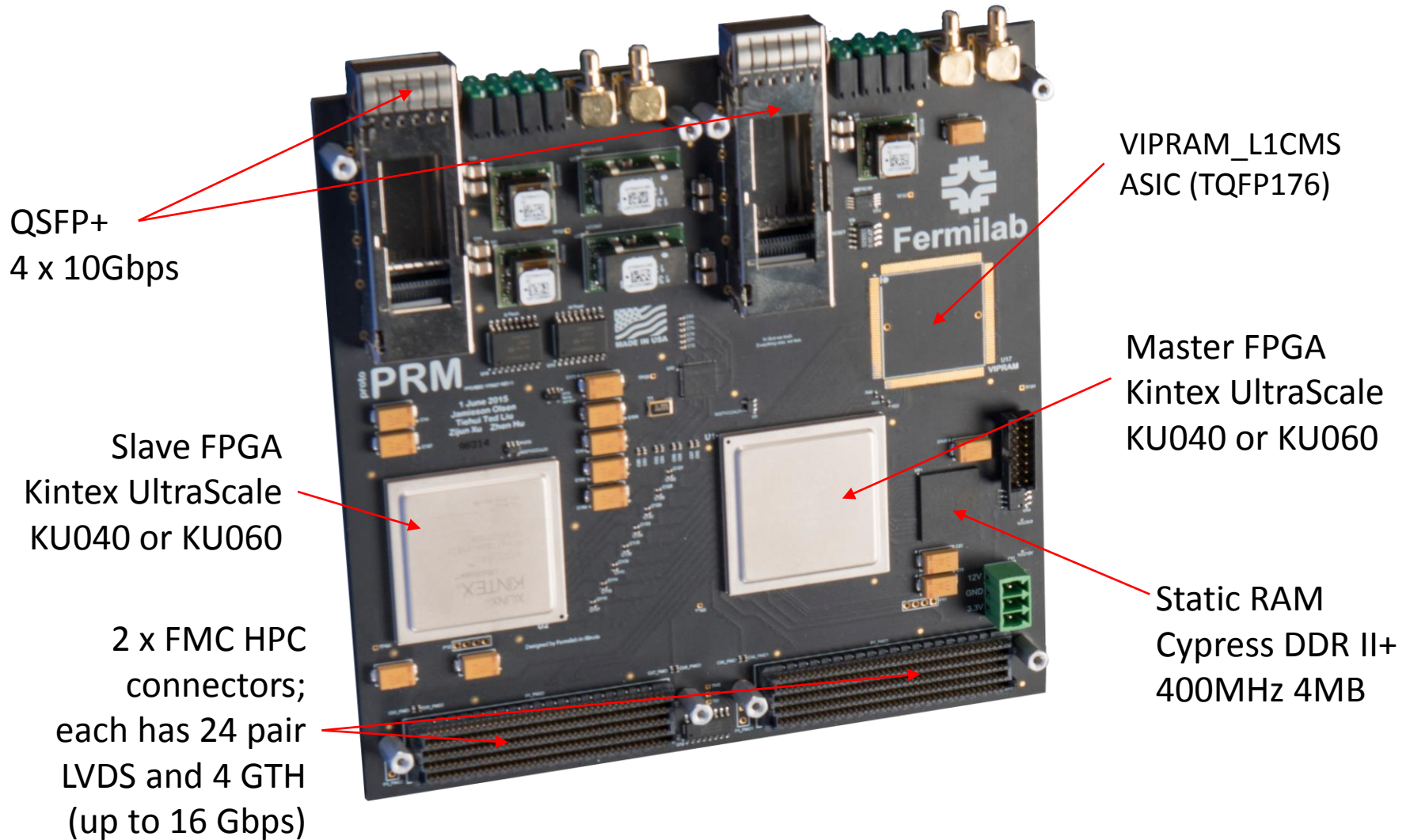**Fermilab**

Pattern Recognition Mezzanine (PRM)

- Designed to explore high performance and low latency PRAM architectures
- Single PRAM channel, pipelined readout
- Kintex UltraScale KU060 FPGAs
- Master FPGA
  - Formatting, Data Organizer, Combiner, and Track Fitters
- Slave FPGA
  - PRAM emulation
  - 1k to 4k patterns
  - Develop new high speed FPGA-PRAM interfaces
  - Local bus is LVDS + 8 x 16 Gbps lanes
- AM ASIC (VIPRAM_L1CMS)

VIPRAM_L1CMS (130nm, two-tier)
ASIC wafers in 3D processing now

🔷 Fermilab

ProtoPRM Board

QSFP+
4 x 10Gbps

VIPRAM_L1CMS
ASIC (TQFP176)

Master FPGA
Kintex UltraScale
KU040 or KU060

Slave FPGA
Kintex UltraScale
KU040 or KU060

Static RAM
Cypress DDR II+
400MHz 4MB

2 x FMC HPC
connectors;
each has 24 pair
LVDS and 4 GTH
(up to 16 Gbps)

🍃 Fermilab

Demonstration System

- One trigger tower = 1 ATCA shelf
- This demonstration system supports different PRMs
  - FNAL ProtoPRM
  - INFN AM05/AM06 PRM
- Time multiplexed transfers to track finder engines which process one event
- Up to 20 PRMs per shelf
- This system architecture utilizes the ATCA full mesh backplane to fullest extent…

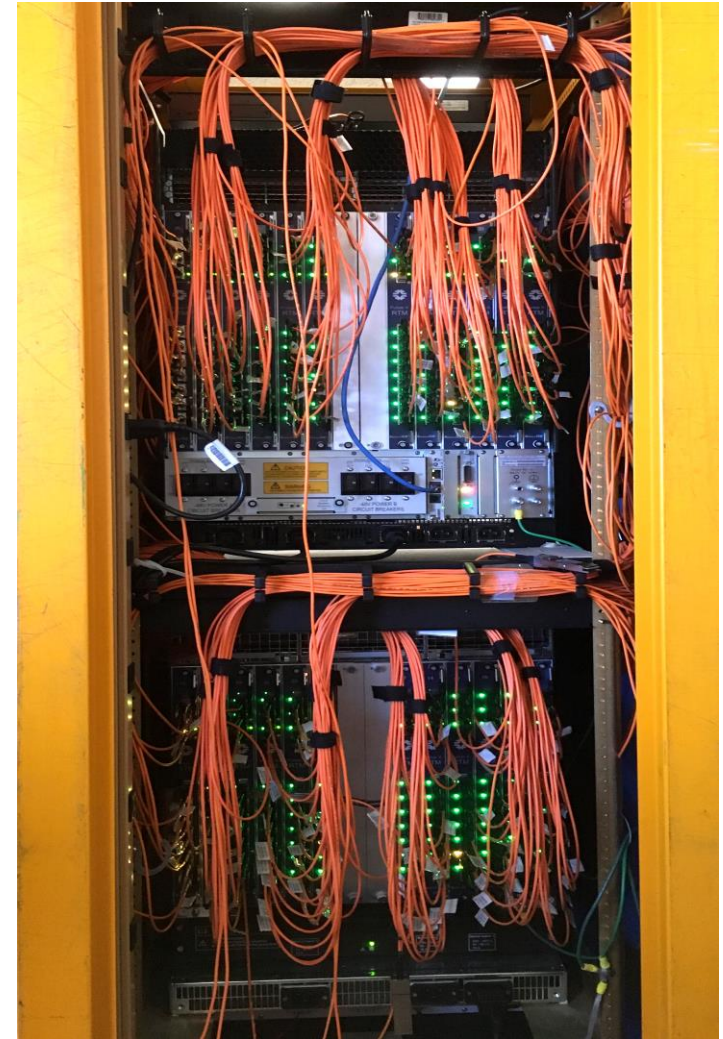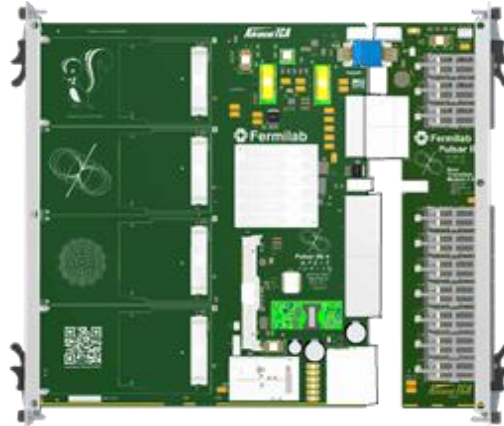**🐝 Fermilab**

Demonstration System Data Flow

Pattern Recognition Board (PRB) shelf

- One Trigger Tower
- 10 Pulsar IIb
- Some boards with PRM Mezzanines
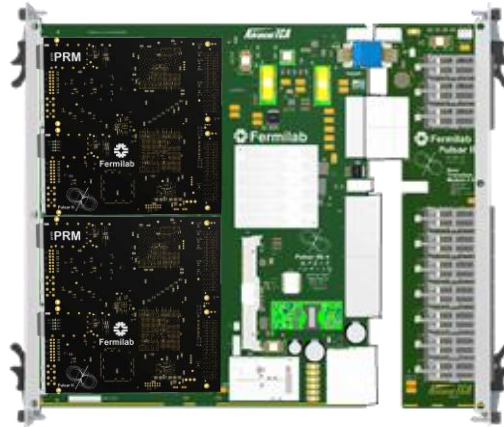
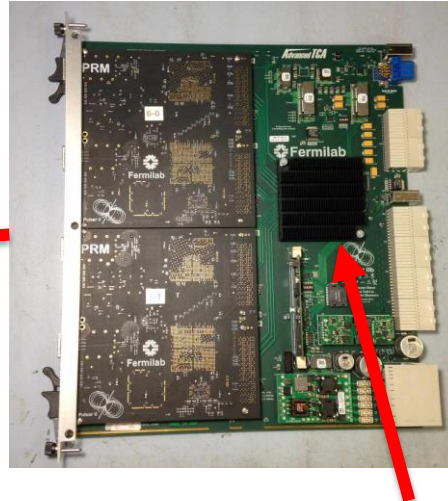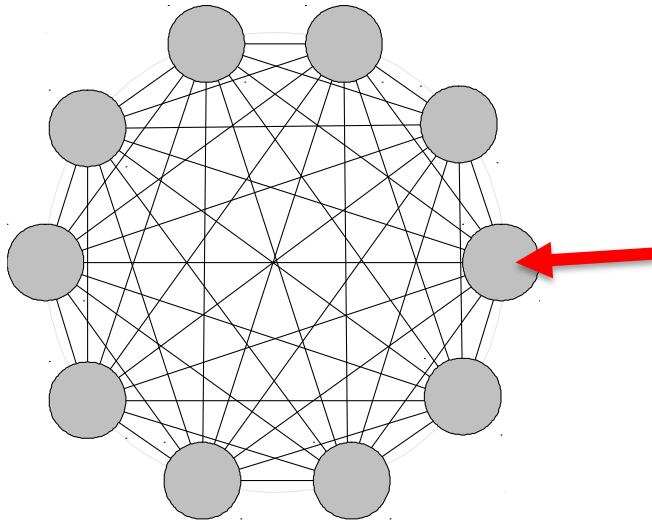**Data Source Board (DSB) shelf**

- Emulates the output of ~400 modules
- 10 Pulsar IIb
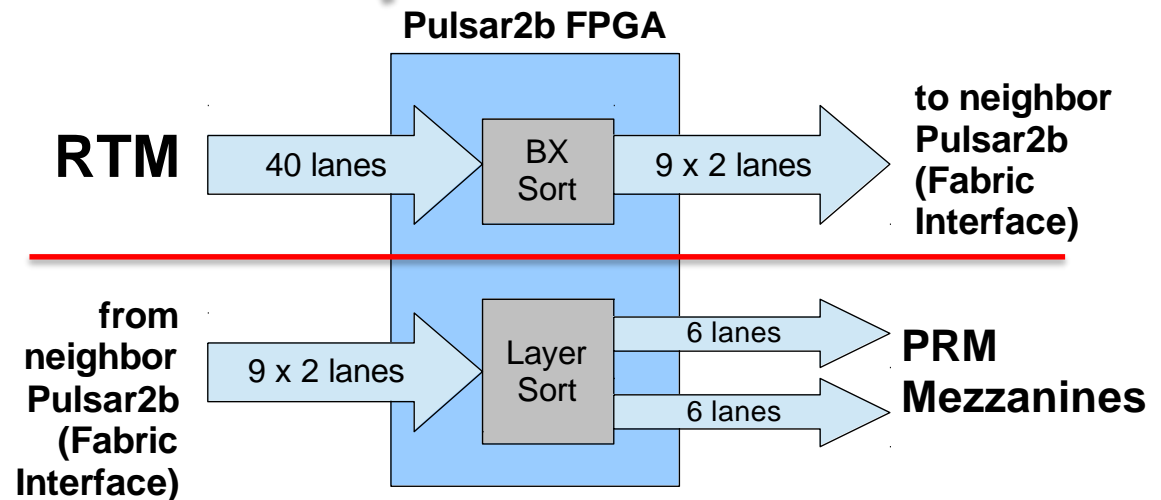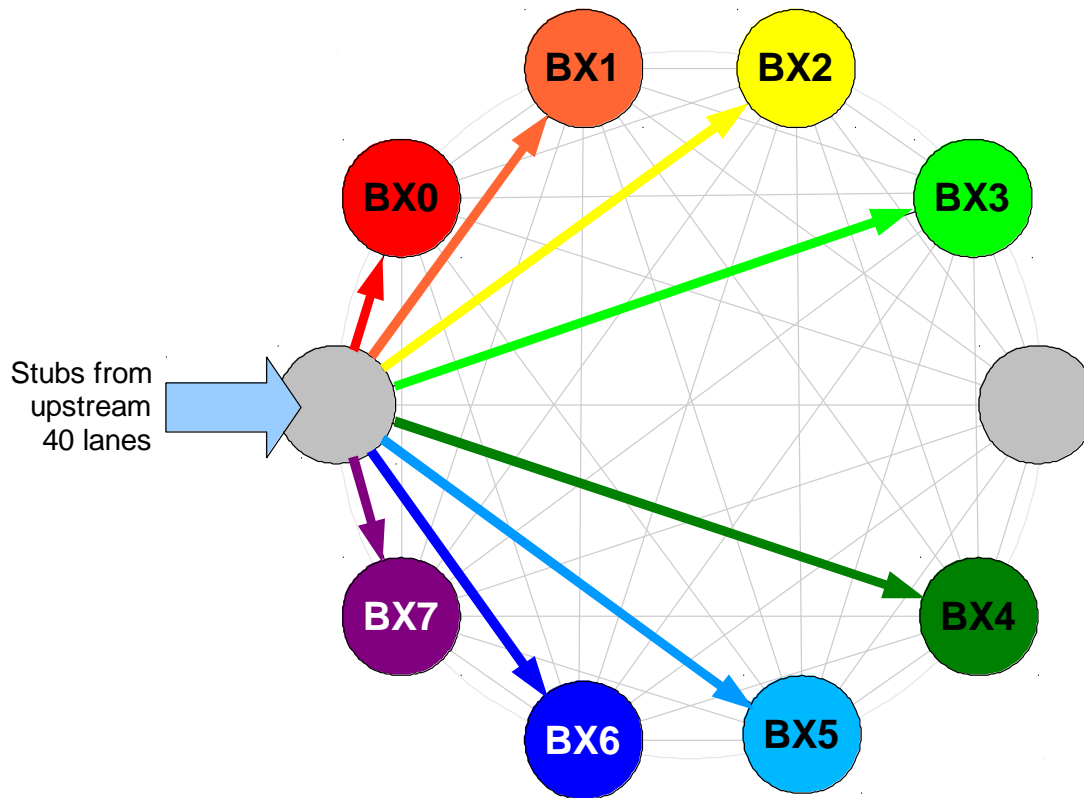- 100 QSFP+ fibers
- > 4Tbps

Fermilab

Data Delivery

Pattern Recognition Board (PRB) is a Pulsar2b with one or two Pattern Recognition Mezzanine (PRM) boards installed. The FPGA on the PRB is tasked with delivering stubs to the proper PRM.

Pulsar2b FPGA

Stubs make **TWO** passes through the PRB before they arrive at the destination PRM boards.

RTM → 40 lanes → BX Sort → 9 x 2 lanes → **to neighbor Pulsar2b (Fabric Interface)**

from neighbor Pulsar2b (Fabric Interface) → 9 x 2 lanes → Layer Sort → 6 lanes → **PRM Mezzanines**, 6 lanes

Fermilab

Data Transfers on the Full Mesh Backplane

Stubs from upstream 40 lanes

All ten boards do this continuously.
First stubs to ProtoPRM by 1.5μs.

- Each Pulsar2b receives stubs on 40 links
- Stubs arrive in a "train" which contains stubs for up to 8 BX
- New train every 200ns
- Pulsar2b FPGA sorts stubs by BX and sends to 7 or 8 neighbors over the backplane
- Backplane transfers must complete in 200ns
- Each board can send up to ~100 stubs to each neighbor
- Full mesh channels are 2 x 10 Gbps, non-blocking

**Fermilab**

ProtoPRM Firmware Overview
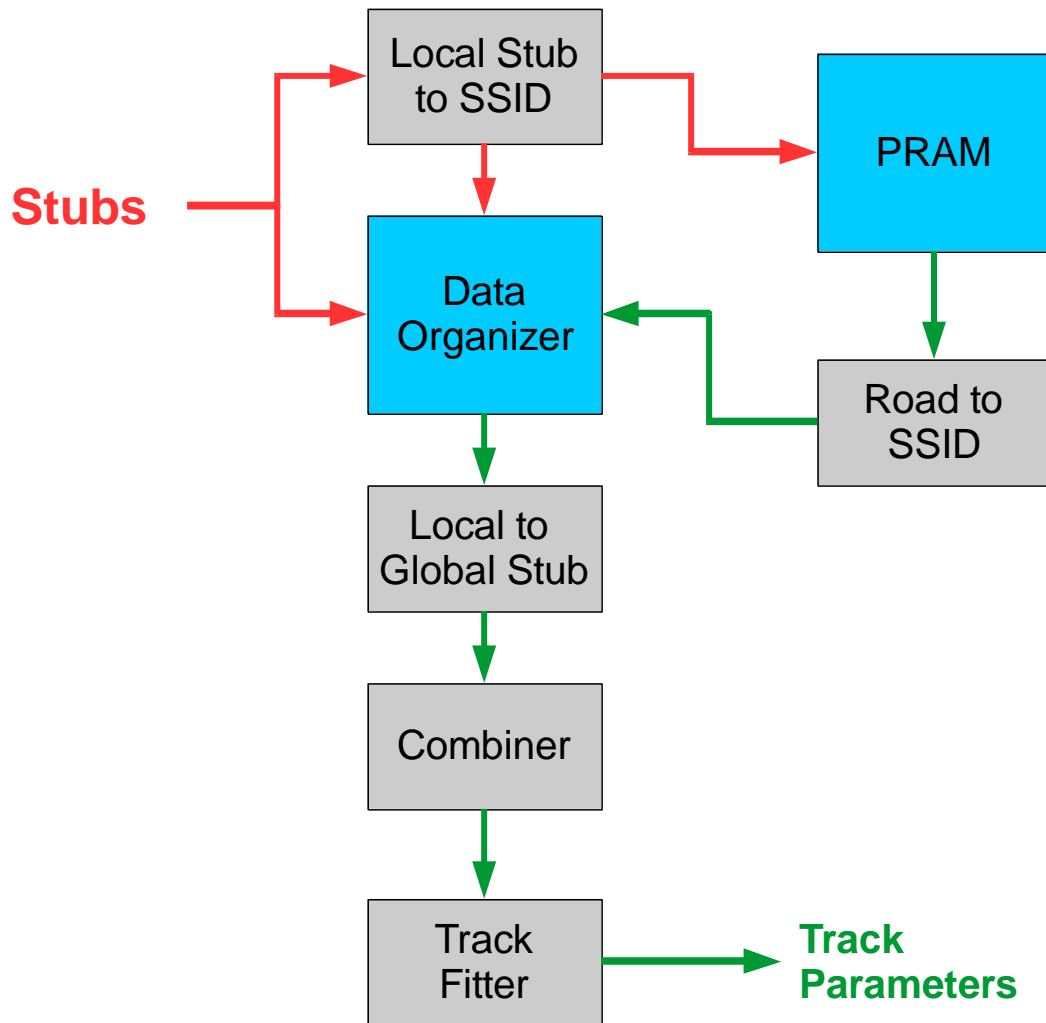
- Conversion/Lookup Functions
  - Local stub to SSID
  - Road-ID to SSID
  - Local stub to global stub
- PRAM Bank
  - 6 layer
  - Pipelined Readout
  - ASIC and FPGA emulation
- Data Organizer
  - Pipelined to match AM
  - Stores stubs at address pointer by SSID
  - Stores multiple stubs per SSID
  - Writes like FIFO
  - Reads like RAM
- Combiner generates multiple stub combinations
- Track Fitter

춘 Fermilab

Data Organizer: Overview

- "smart database" stores stubs at the address pointed to by the SSID
  - SSID = 12 bits → 4k memory locations
  - Store up to **4** stubs per SSID
- The DO architecture is fundamentally geared towards **read-modify-write** operations
- A redesign of the DO was needed because:
  - Our VIPRAM/PRAM readout is pipelined
  - The data organizer must concurrently store stubs for event N while recalling stubs for event N-1
  - DO must "ping pong" dual RAM banks
- RAM "scrubbing" functions are implemented
  - Periodic clearing of the RAM is done with writes (no global reset)
  - Prevent stubs from old events from being read out ("masking")

Data Organizer: Operation
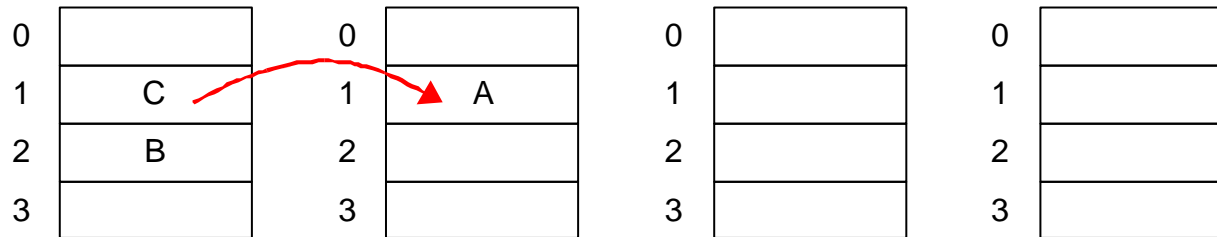
- New design **eliminates read-modify-write cycles**

- Use 7-Series/UltraScale BlockRAM "read first mode"

  – As data is written into BlockRAM the *previous data* at that location is pushed to the output

- Four cascaded RAMs hold the stub information

- Simple, fast, and efficient configuration resembles an "array of FIFOs"

- Read latency is very fast, just like reading BlockRAM, stubs output in parallel

- Dual port BlockRAMs simplify the "ping pong" mechanism

  – Port A is used for writing event N stubs

  – Port B is used for reading event N-1 stubs

Fermilab

Data Organizer: Design

11 bit SSID = 32 BlockRAMs (3% KU060)
12 bit SSID = 64 BlockRAMs (6% KU060)
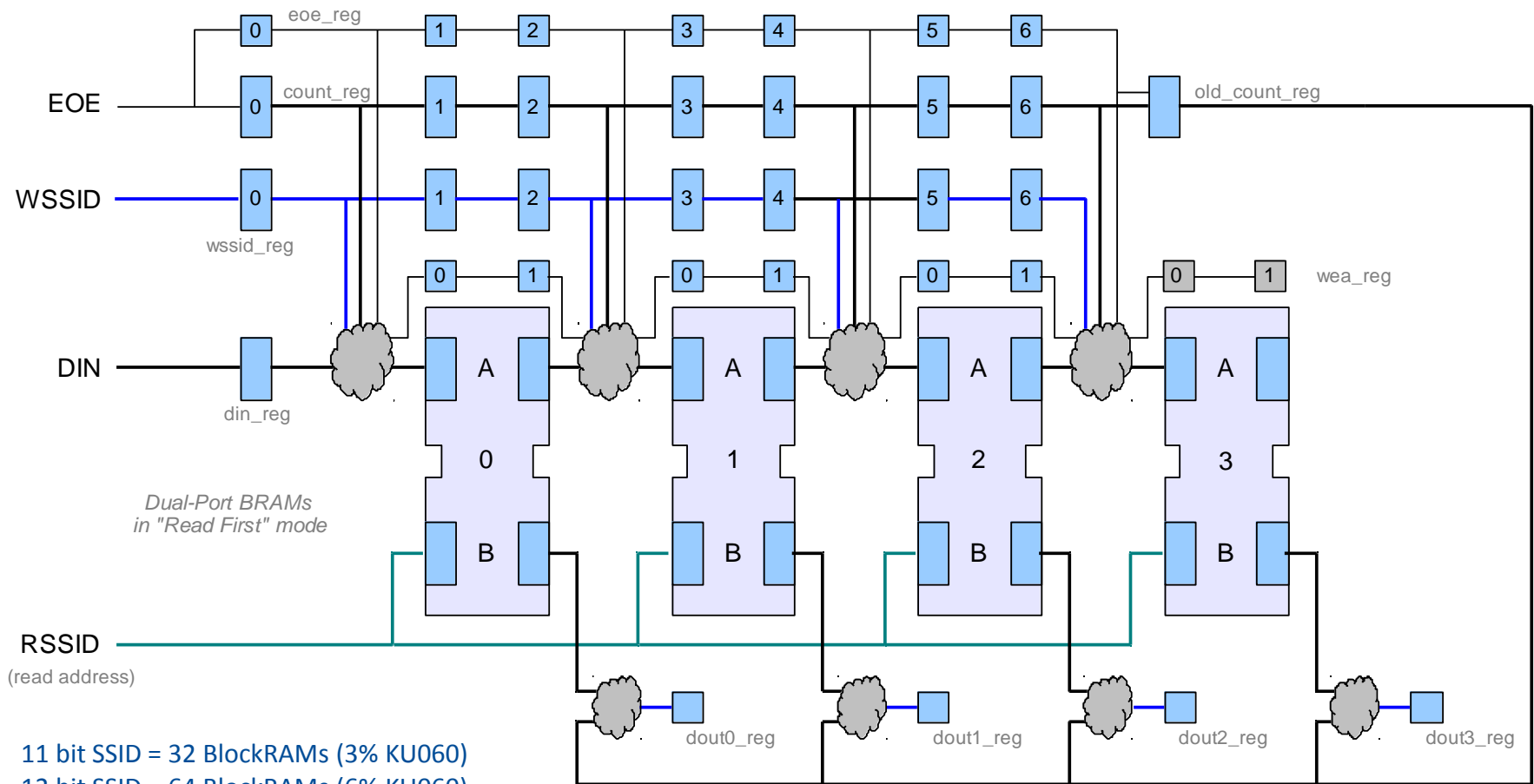
One DO per layer is required. BlockRAMs are wide enough to store global stubs.

As shown, DO can store up to 4 stubs per SSID. This can be increased easily by adding additional BlockRAMs.

Fermilab

PRAM in FPGA

- Fully synthesizable VHDL model of current VIPRAM_L1CMS
  - Multi-tier pipelined readout
  - "CAM tier" processes stubs for the current event
  - "I/O tier" captures road flags and outputs road addresses for previous event
- Design optimized for 7-Series/UltraScale architecture
- Fairly close to "cycle accurate" timing
- 6 input layer buses, 12 bits per layer
- 1k to 4k patterns, fully programmable
- Option for "don't care bits"

🟦 **Fermilab**
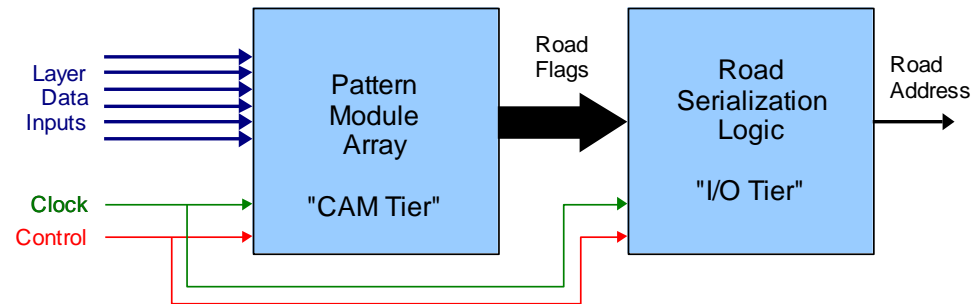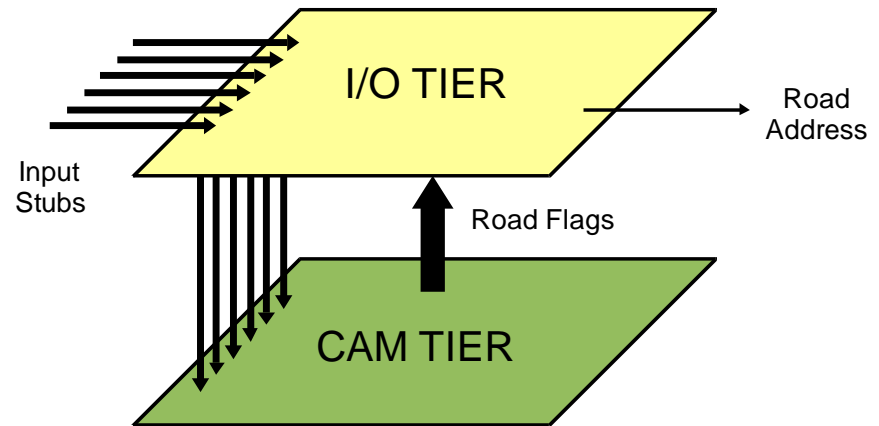
PRAM in FPGA: Road Serialization Logic
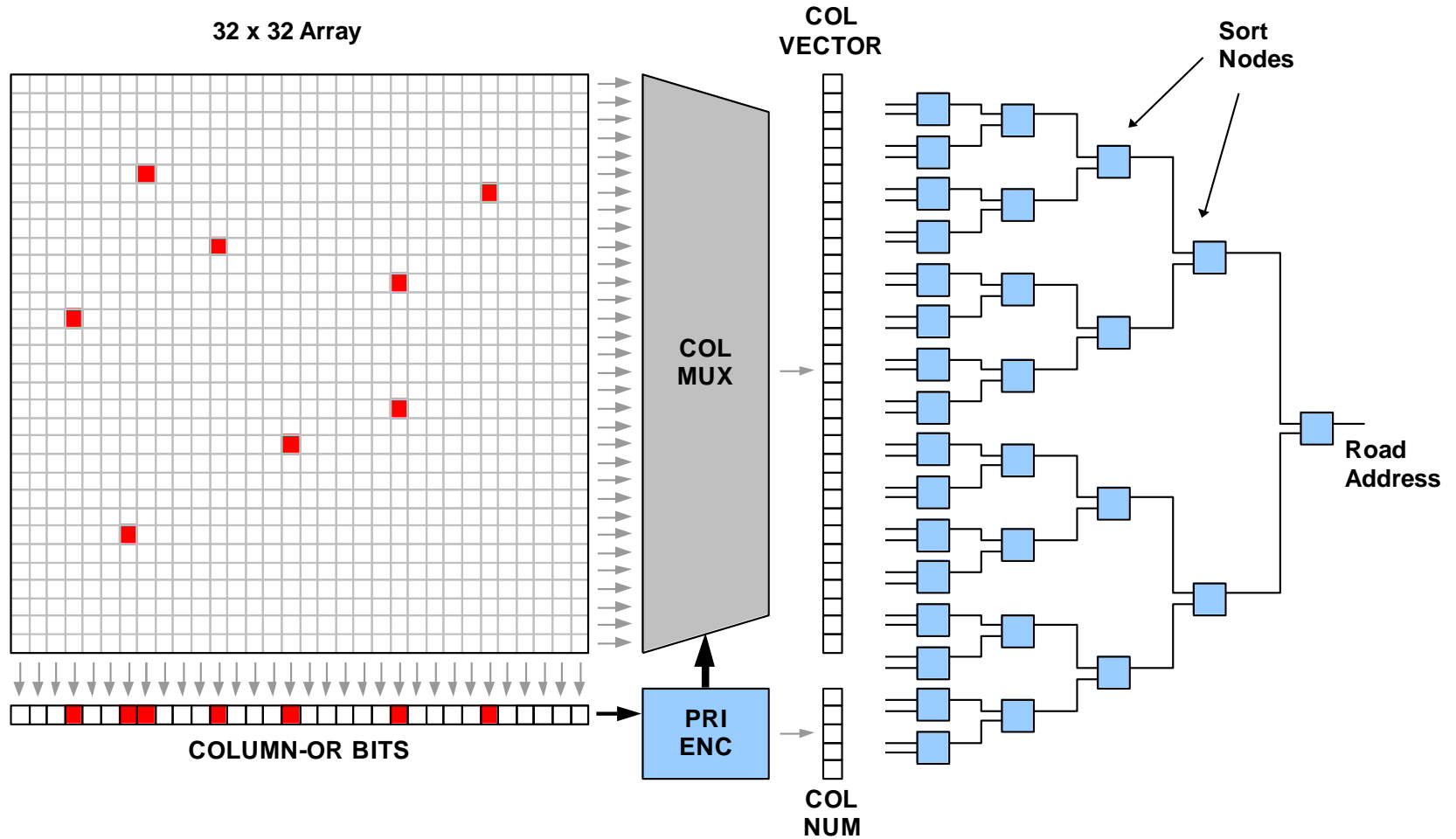
PRAM in FPGA: Interface

- The ProtoPRM Master-Slave local bus consists of:
  - 8 GTH lanes, up to 16.3 Gbps / lane
  - 24 LVDS pairs, up to 1 Gbps / pair
- We plan to use this local bus to develop high performance interfaces for future PRAM ASICs
- Low latency is critical for this path
  - Roads must get back to the Data Organizer before next event arrives
  - GTH transceiver latency is a bit too high (~150ns)
  - Source synchronous DDR/serial has the lowest latency (~12ns with 240MHz clock)
- This interface is similar to VIPRAM_L1CMS ASIC

🟦 **Fermilab**

ProtoPRM timing
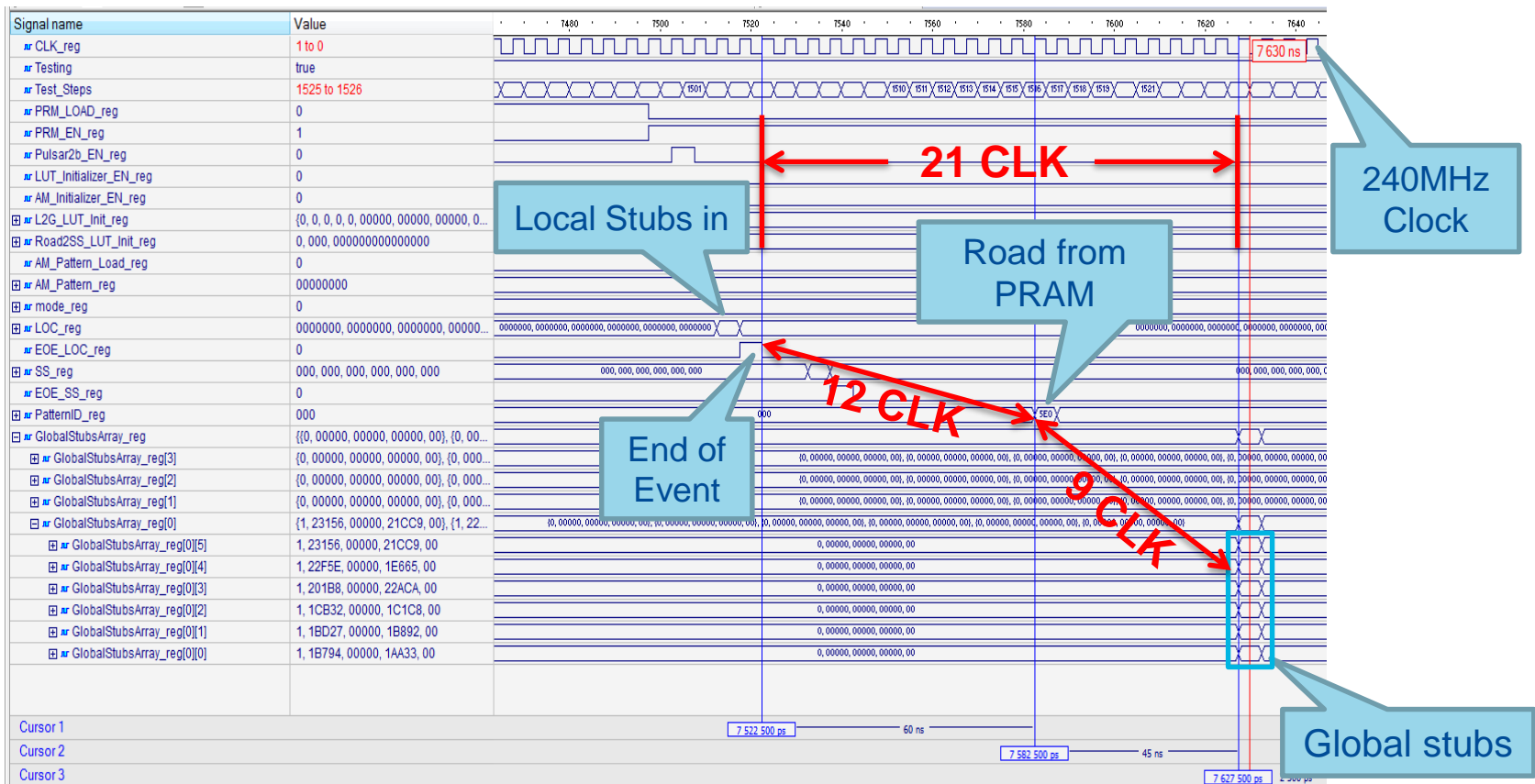
The overhead of conversion functions, lookup tables, data organizer, and PRAM is on the order of 100ns.

ProtoPRM Latency Estimate
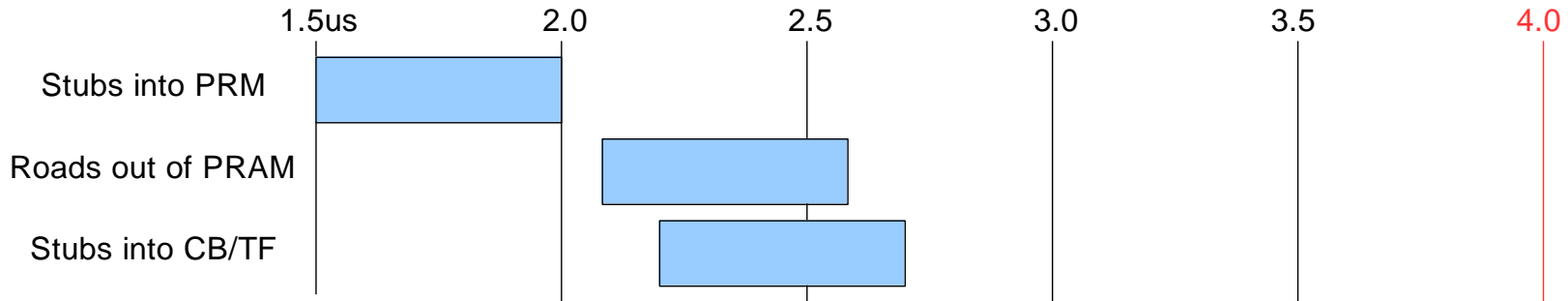
- Target: total latency =< 4µs

- Data delivery = 1.5µs

    – First stubs arrive at ProtoPRM input

- Stubs into Combiner/Track fitter @ 2.2µs

- Combiner/Track fitter latency on the order of 1µs

Summary

- Our L1CMS track trigger demonstration system is based around time multiplexed data transfers over the ATCA full mesh backplane

- Pulsar IIb front boards make up the backbone of the system

- ProtoPRM mezzanine boards are the track finder engines

- New firmware designs have been optimized for high performance, pipelined, low latency operation
  - Data Organizer
  - PRAM in FPGA

- Demonstration system integration is underway

- We look forward to sharing new results at TWEPP-17!

🎇 Fermilab

Backup Slides

Fermilab

# Link Performance: Full Mesh Backplane

- *Many* different full mesh backplanes were tested at Fermilab

- In late 2014, we purchased the next generation 100G Air-/-Plane backplane from COMTEL

- ALL of the 56 bidirectional links among 8 Pulsar2b boards were tested at *10.0 Gbps (PRBS7)*
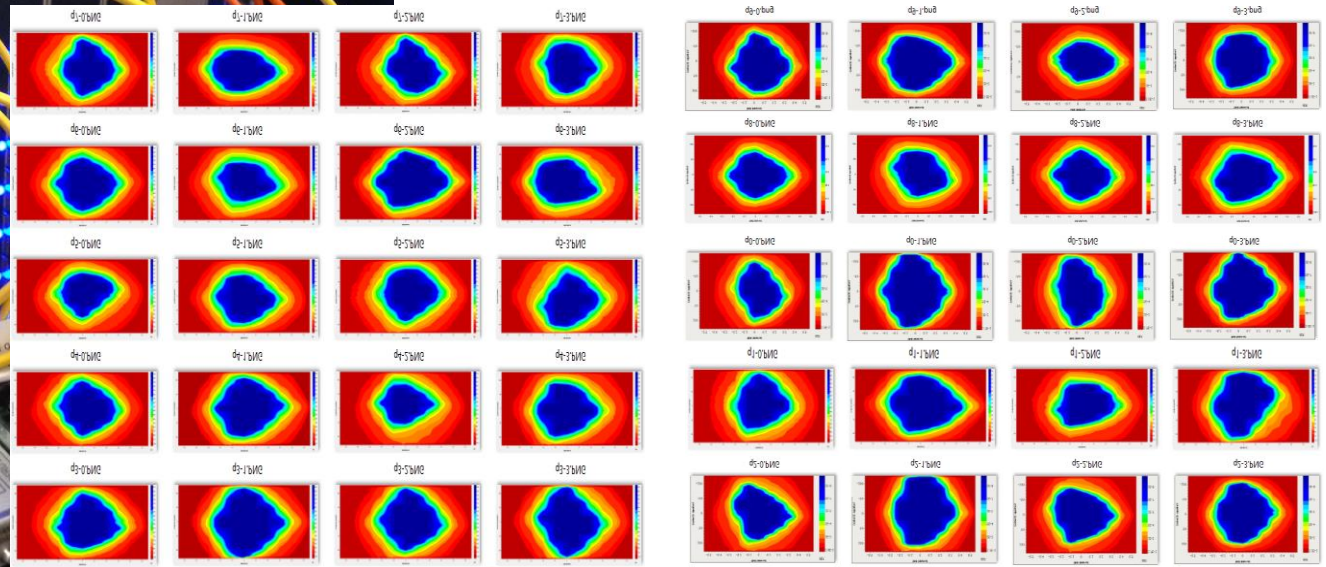
- The best and most consistent link performance to date

Air-/-Plane

🐾 **Fermilab**
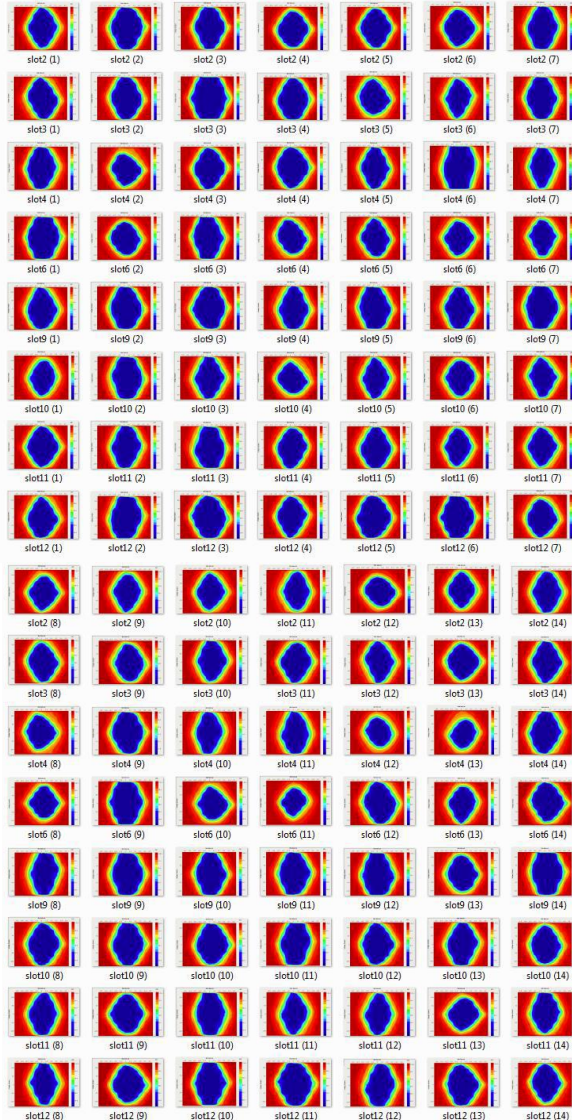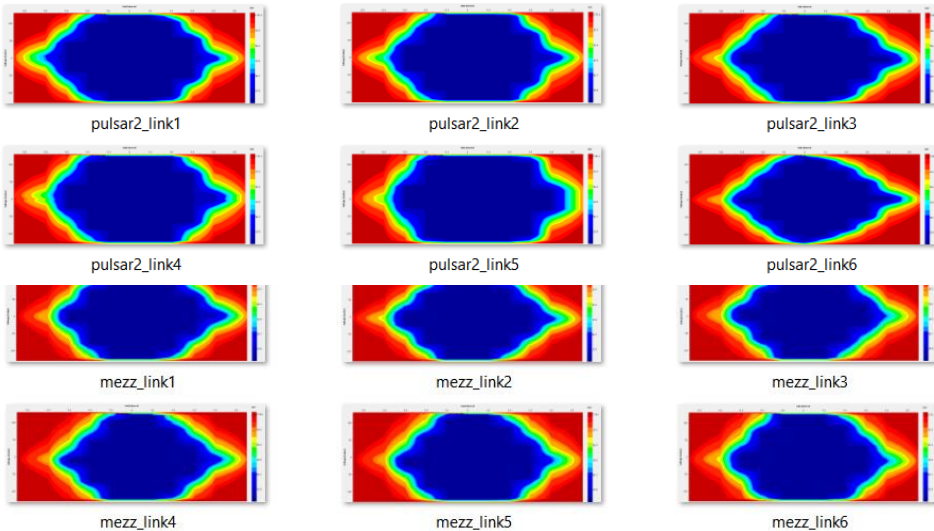
Link Performance: RTM
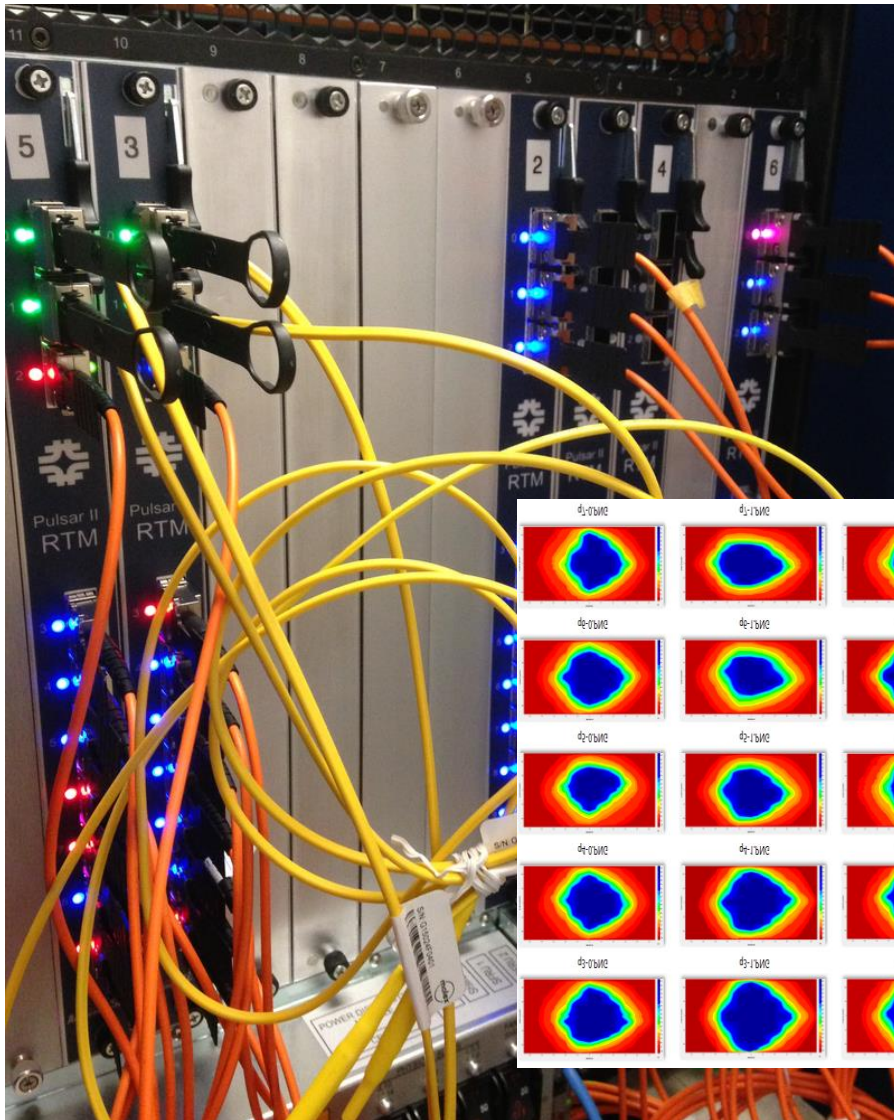
- 10 QSFP+ modules in RTM
  - Design goal:
    10 Gbps per lane
  - 400 Gbps to Pulsar IIb
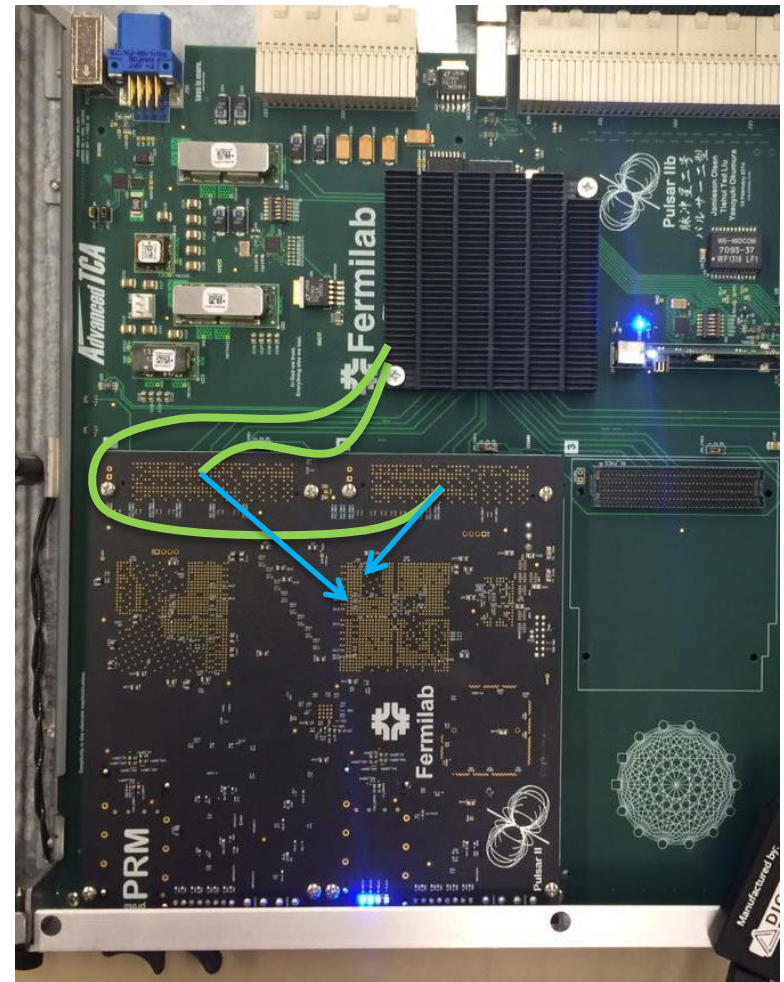- Links @ 10.0Gbps/PRBS7:

Link Performance: Pulsar IIb to protoPRM
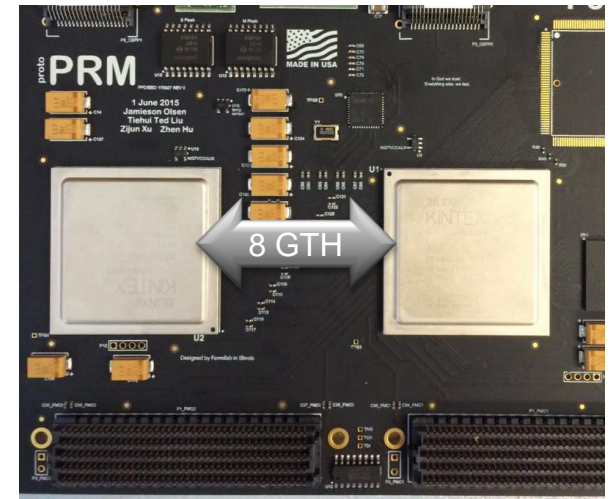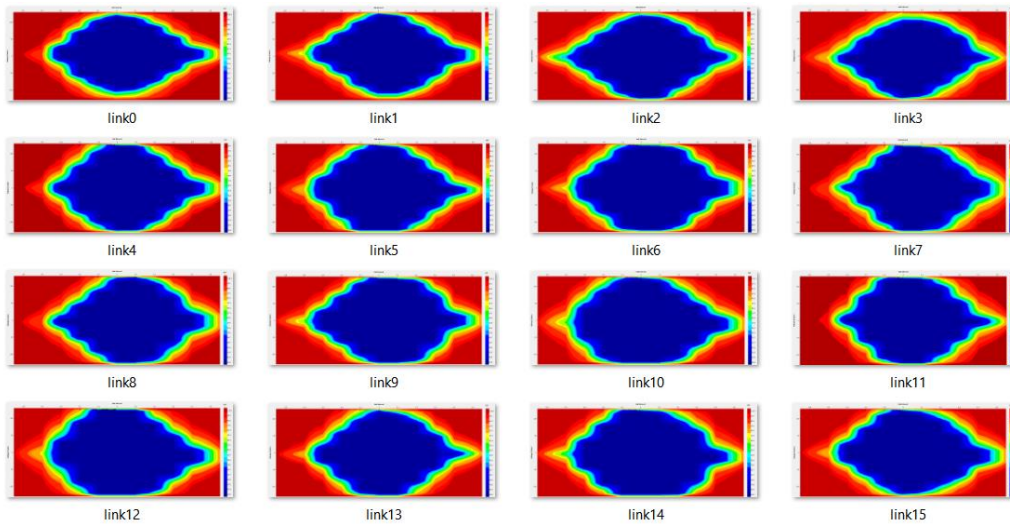
- ## 10.0 Gb/s: BER<1e-14

- 6.25Gb/s and 8.0 Gb/s also tested

- All 6 channels are error free

- Pulsar2b assigns 3 GTH to each FMC, while protoPRM has 4 available

- 10Gb/s is limited by FMC connector

Link Performance: protoPRM Local Bus

- ## 16.3 Gb/s, PRBS7: BER <1e-14

- # 6.25, 8.0, 10.0 and 12.5 Gb/s also tested
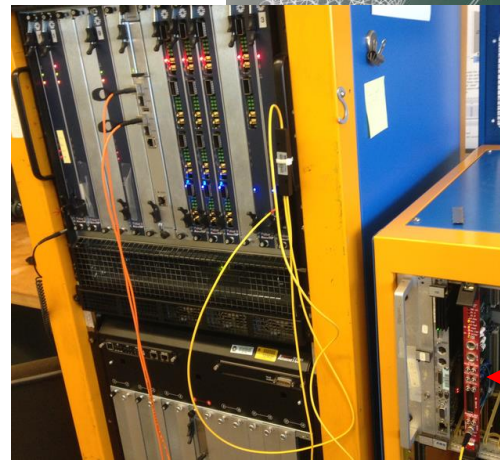- # all 8 channels are error free
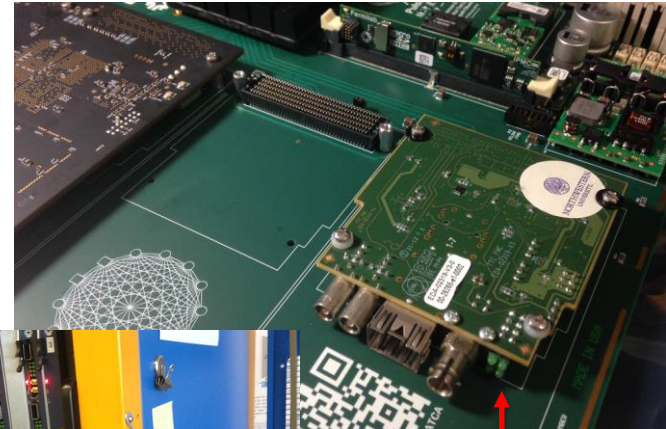
🔷 Fermilab

System Synchronization

Intra-Shelf

- User clocks on ATCA backplane connect to Pulsar2b FPGA
- Any Pulsar2b board can be master
- 40MHz LHC clock
- TTC A/B Channel Data
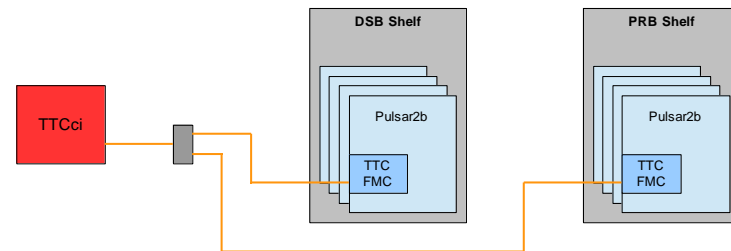- M-LVDS tested to 100MHz (Northwestern Univ.)

Inter-Shelf

- TTC receiver FMC mezzanines installed on Pulsar2b boards
- Source is TTCci VME board
- Passive fiber splitter

TTC FMC

TTCci

Pulsar2b Backplane Synchronization Logic