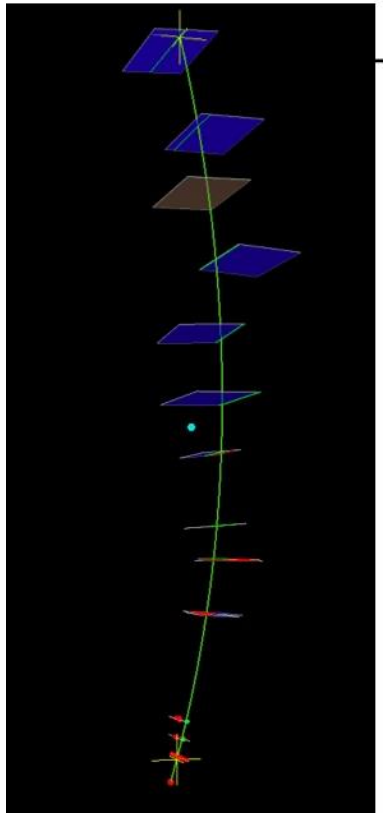
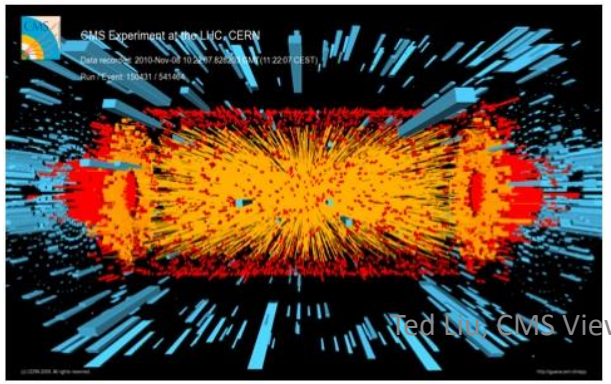
CMS Views for the *Off-Detector* L1 Tracking Trigger Electronics

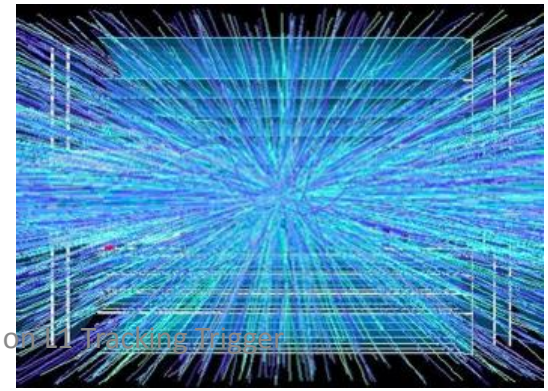
Common ATLAS CMS Electronics Workshop

Ted Liu (FNAL)

March. 19, 2014







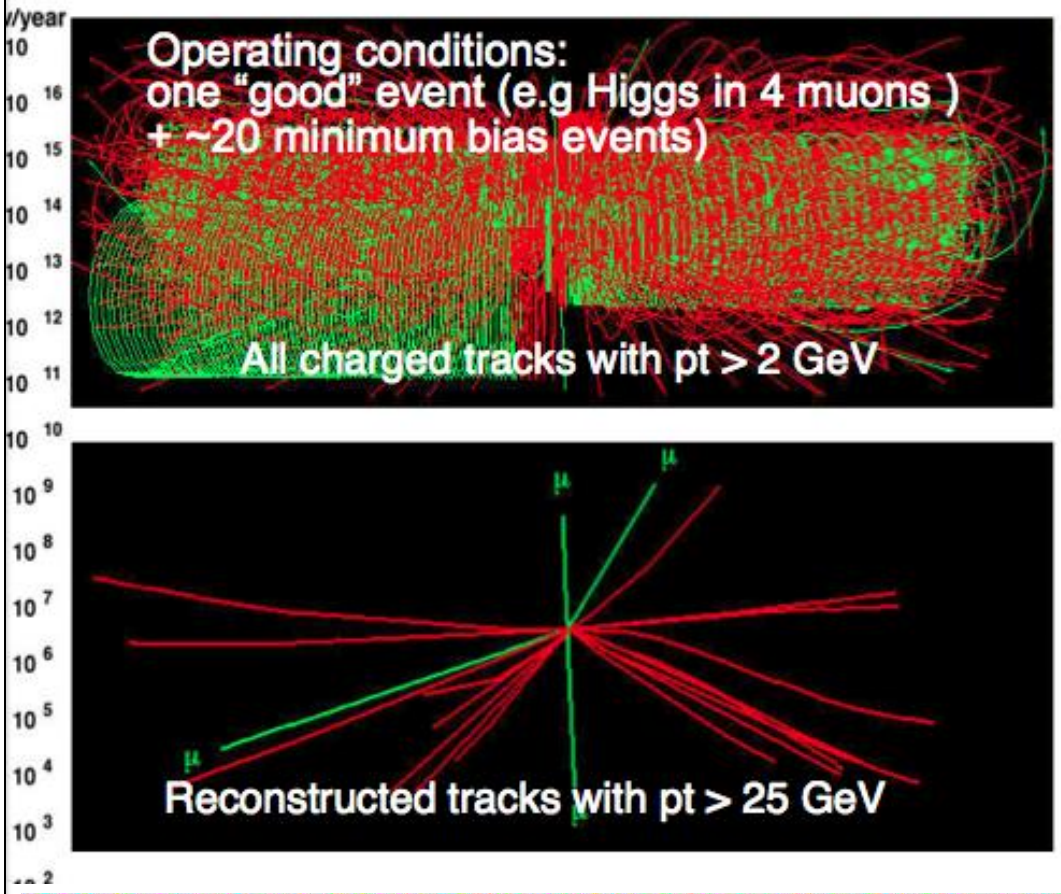
CMS L1 Tracking Trigger:

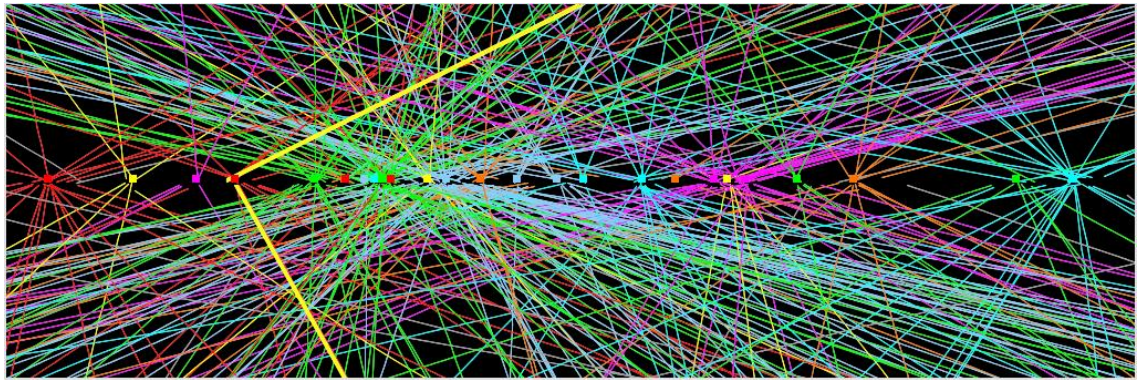
Will need to reconstruct charged particle trajectories "on-the-fly" for every beam crossing (25 ns, or 40 Million beam crossings per second), from an ocean of input data (bandwidth required to transfer up to ~ 50-100Tb/s)

This requires extremely fast high bandwidth data communication as well as massive pattern recognition power,
with lots known patterns to be compared against the multiple input data streams simultaneously with near zero latency (~ few μs)

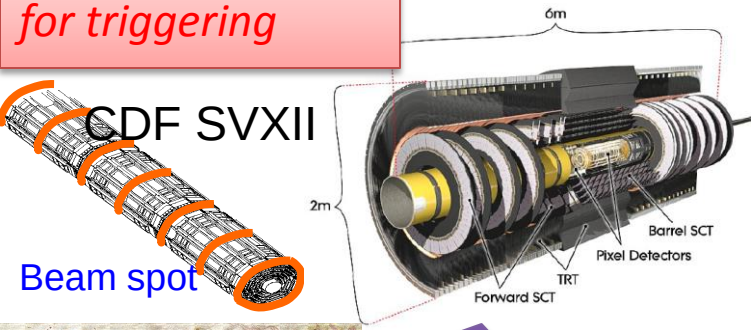
This is challenging!

Pileup at HL-HC: > ~ 140 (only 20 shown here)





Detector design for triggering



Silicon Based Tracking Trigger at Hadron Colliders

Partition detector into trigger towers/sectors

Data transfer

Data formatting

Pick your favorite method:

Associative Memory (AM) Approach (proven approach from CDF/SVT)

Hough Transformation

tracklet-based

Adaptive Pattern Recognition

Biology Inspired …

your choice here…

Pattern Recognition

Finer pattern recognition

Track Fitting

- *In this talk (focus on off-detector)*

(1) Pattern Recognition + Track Fitting options

-- AM + linearized track fitting (FPGA): traditional

-- AM + Hough transform (FPGA): new, being studied

-- Tracklet-base approach (FPGA): new, being studied

FPGA vs GPU vs CPU

HLT

(2) Data Formatting and System Architecture options:

-- FPGA based Full-mesh enabled ATCA approach (Pulsar 2)
(common electronics developed for Atlas FTK and CMS L1 Track Trigger R&D)

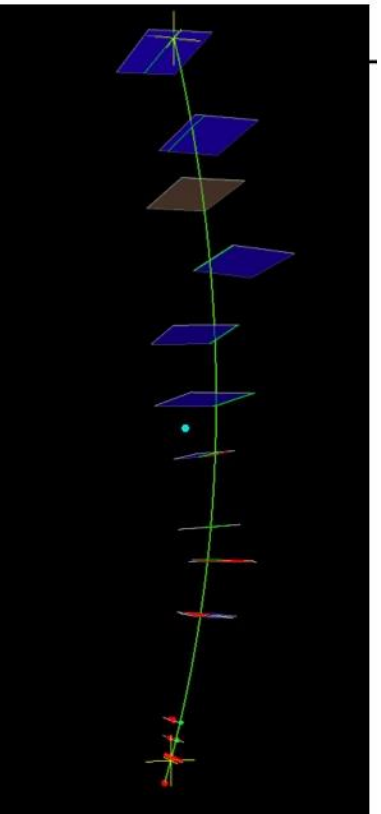
-- MicroTCA based (MP7): electronics developed for CMS L1 CAL trigger upgrade

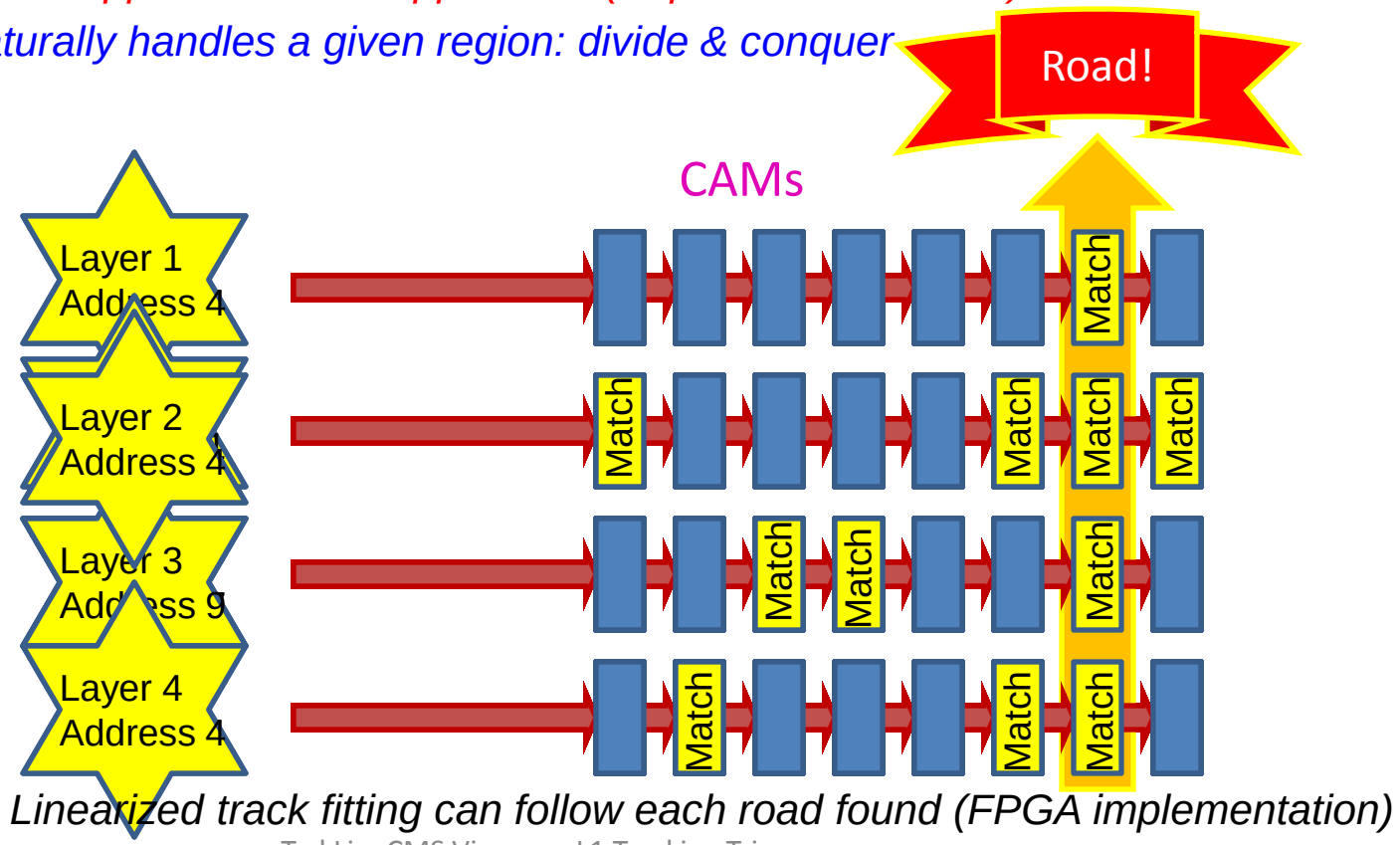
(3) Summary (Vertical Slice Demonstration)

The AM approach

- Pattern Recognition Associative Memory
  - Based on *CAM cells to match and majority logic to associate* hits in different detector layers to a set of pre-determined hit patterns (simple working unit, yet massively parallel)
  - *Pattern Recognition finishes right after all hits arrive (fast data delivery important)*
  - *Potentially good approach for L1 application (require custom ASIC)*

A PR engine naturally handles a given region: divide & conquer

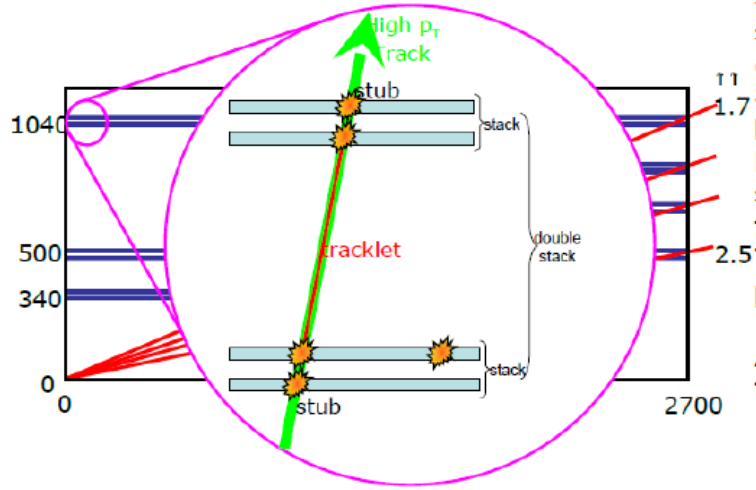




Linearized track fitting can follow each road found (FPGA implementation)

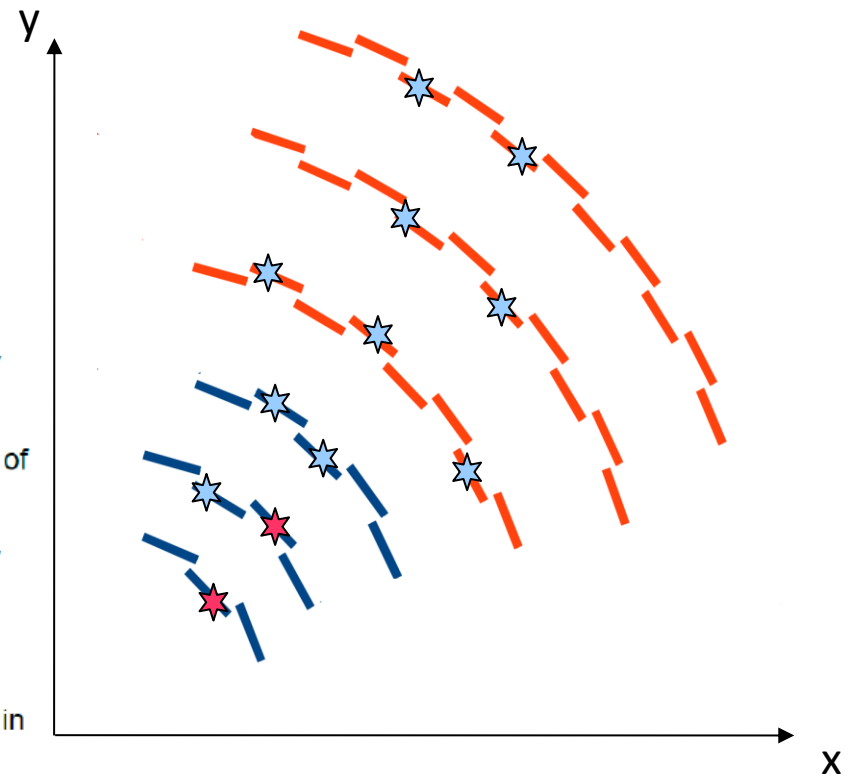
Tracklet Based Track Finding → new concept for L1 tracking trigger

- Form track seeds, tracklets, from pairs of stubs in neighboring layers



Stack:pair of closely spaced sensors (~1mm)
Stub:correlated pair of hits in stub
Double stack:Two stacks separated by few cm.
Tracklet A matched pair of stubs.

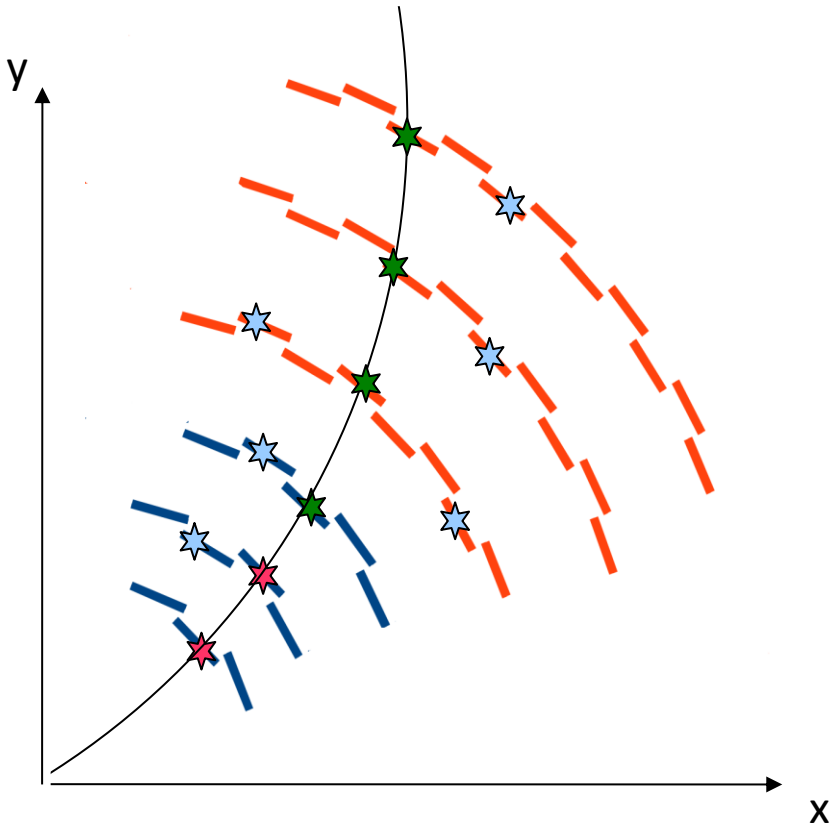
A layer is one stack in this talk.



Original concept → Being explored for current geometry

Tracklet Based Track Finding → new concept, being explored at CMS

- Form track seeds, tracklets, from pairs of stubs in neighboring layers
- Match stubs on road defined by tracklet and IP constraint
- Fit the hits matched to the tracklet using a linearized fit
- Seeding is done in parallel in different layers
- Duplicate tracks are removed if they share 2 or more stubs



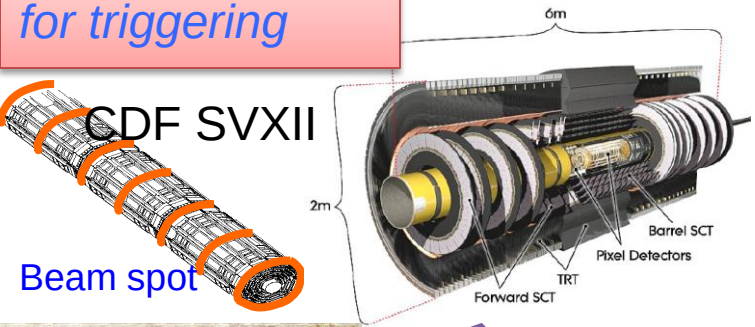
Slide from Anders Ryd (Cornell/USA)

Comparison of the two approaches

	AM + TF approach	Tracklet +TF approach
advantages	• Proven approach for silicon based track finding • AM pattern recognition algorithm: simple, fast and flexible	• New approach for *hardware* silicon based track finding • Software simulation promising • *Can be implemented in FPGA in principle: no need for custom designed chips*
challenges	• *Requires custom ASIC: high performance AMchip* • Track Fitting in FPGA to be demonstrated for L1 • New architecture (see below)	• *It is new* • *Feasibility to be demonstrated in hardware (FPGA)*
	***Common: Fast** Data delivery/sharing to Pattern Recognition Engines*	

The rest of the talk …

Detector design for triggering



Silicon Based Tracking Trigger at Hadron Colliders

Partition detector into trigger towers/sectors

Data transfer

Data formatting

Pick your favorite method:

Associative Memory (AM) Approach (proven approach from CDF/SVT)

Hough Transformation

tracklet-based

Adaptive Pattern Recognition

Biology Inspired …

your choice here…

Pattern Recognition

Finer pattern recognition

Track Fitting

FPGA vs GPU vs CPU

HLT

- *In this talk (focus on off-detector)*

(1) Pattern Recognition + Track Fitting options

-- AM + linearized track fitting (FPGA): traditional

-- AM + Hough transform (FPGA): new, being studied

-- Tracklet-base approach (FPGA): new, being studied

(2) Data Formatting and System Architecture options:

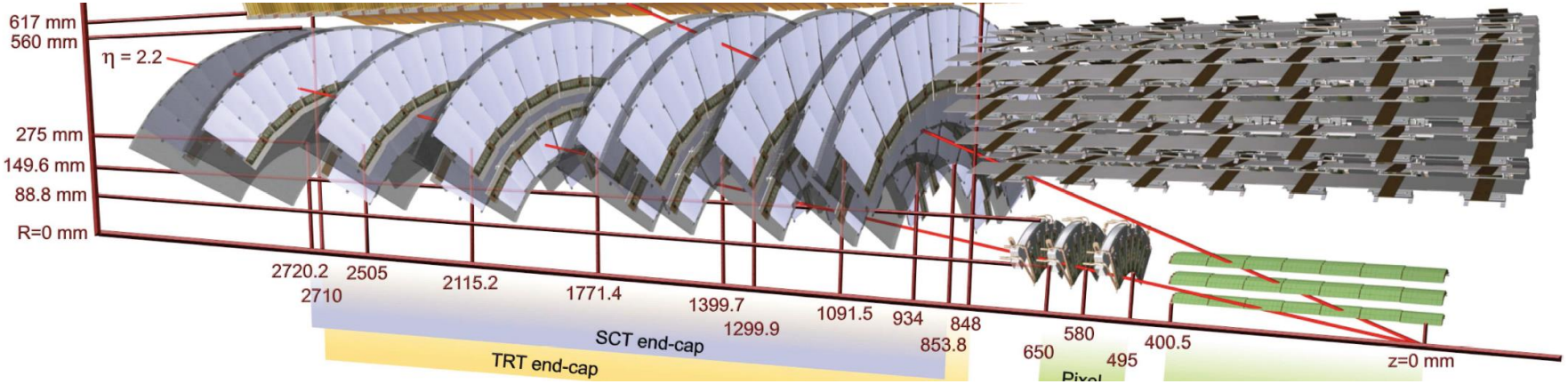
-- FPGA based Full-mesh enabled ATCA approach (Pulsar 2)

*(**common** electronics developed for Atlas FTK and CMS L1 Track Trigger R&D)*

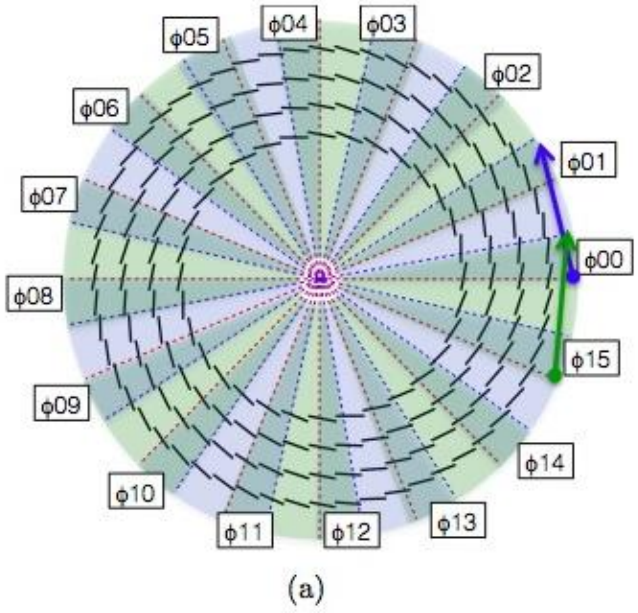
-- MicroTCA based (MP7): electronics developed for CMS L1 CAL trigger upgrade

(3) Summary (Vertical Slice Demonstration)

First take a look at: Data formatting challenges for Atlas FTK at L2

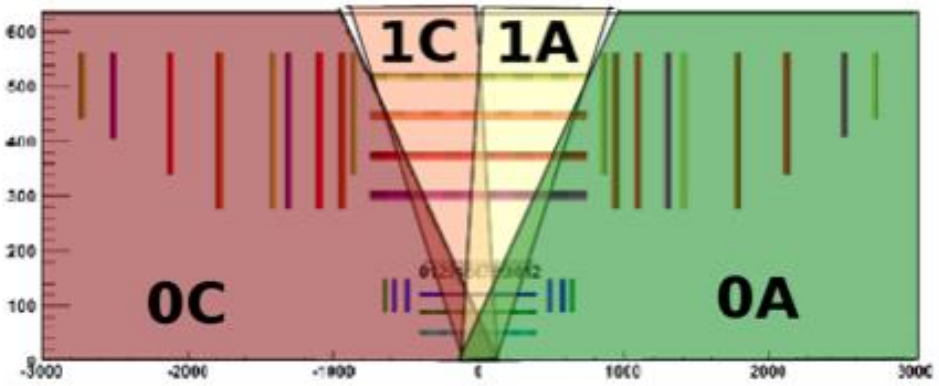


Input data from all silicon detector modules has to be formatted into 64 η-φ trigger towers after reformatting and sharing, ready for downstream pattern recognition

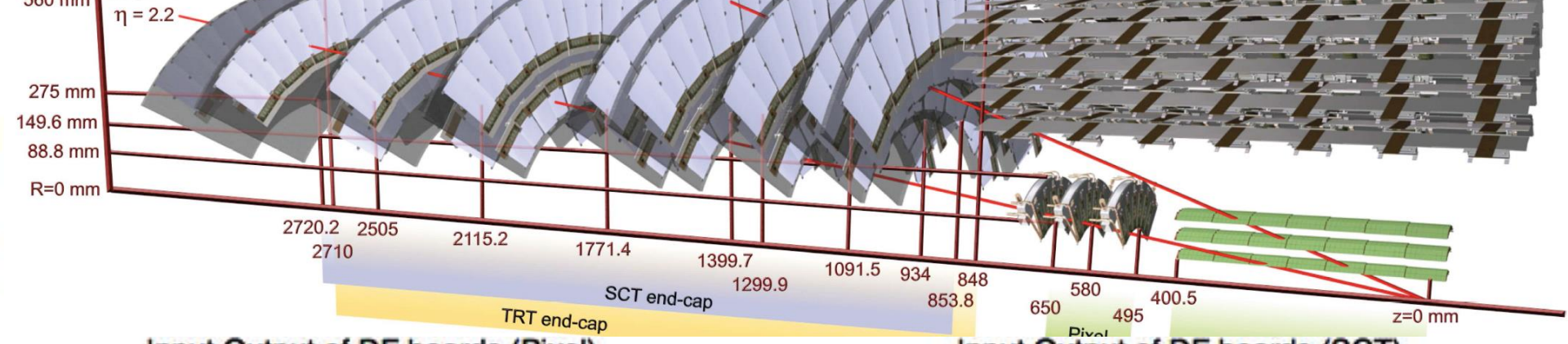


(a)

4 in η

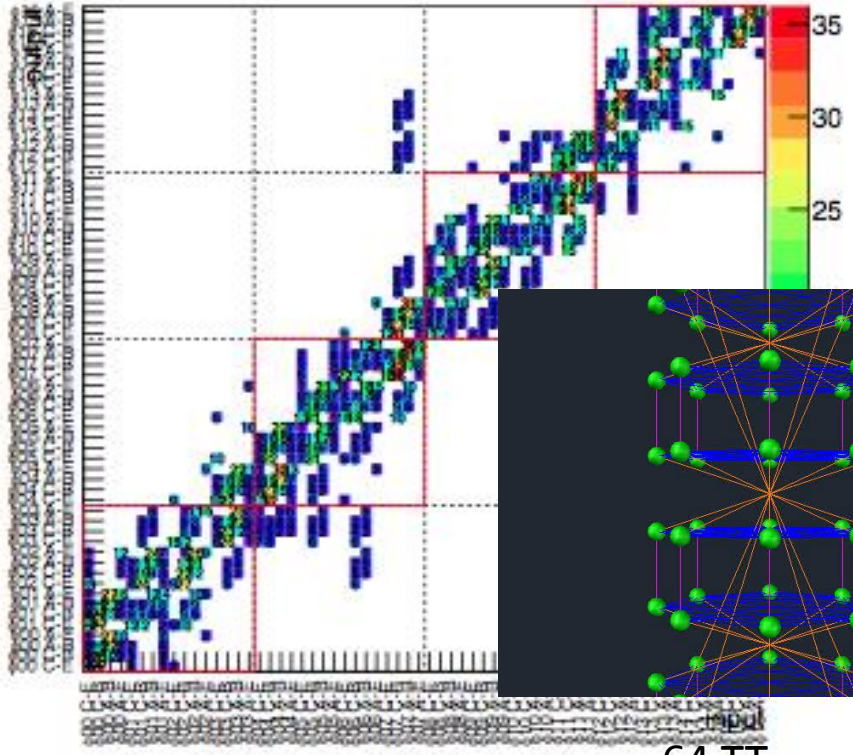


ATLAS FTK: see Alberto Annovi's FTK talk



64 TT

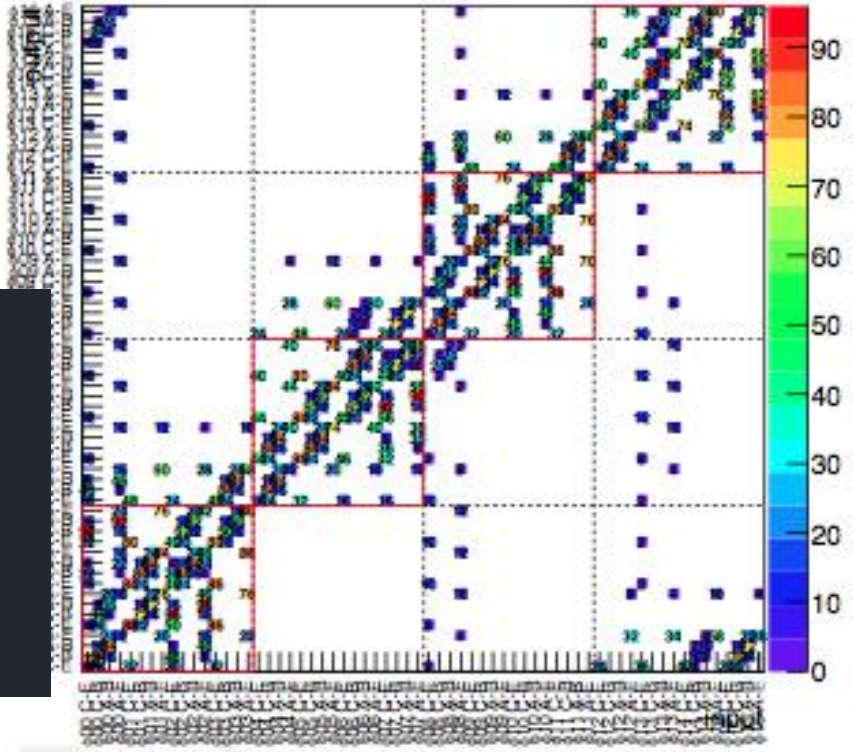
Input Output of DF boards (Pixel)



64 TT

(a)

Input Output of DF boards (SCT)

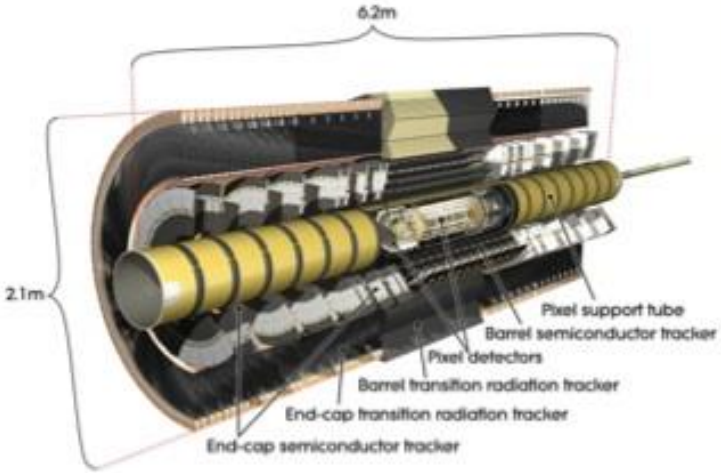


(b)

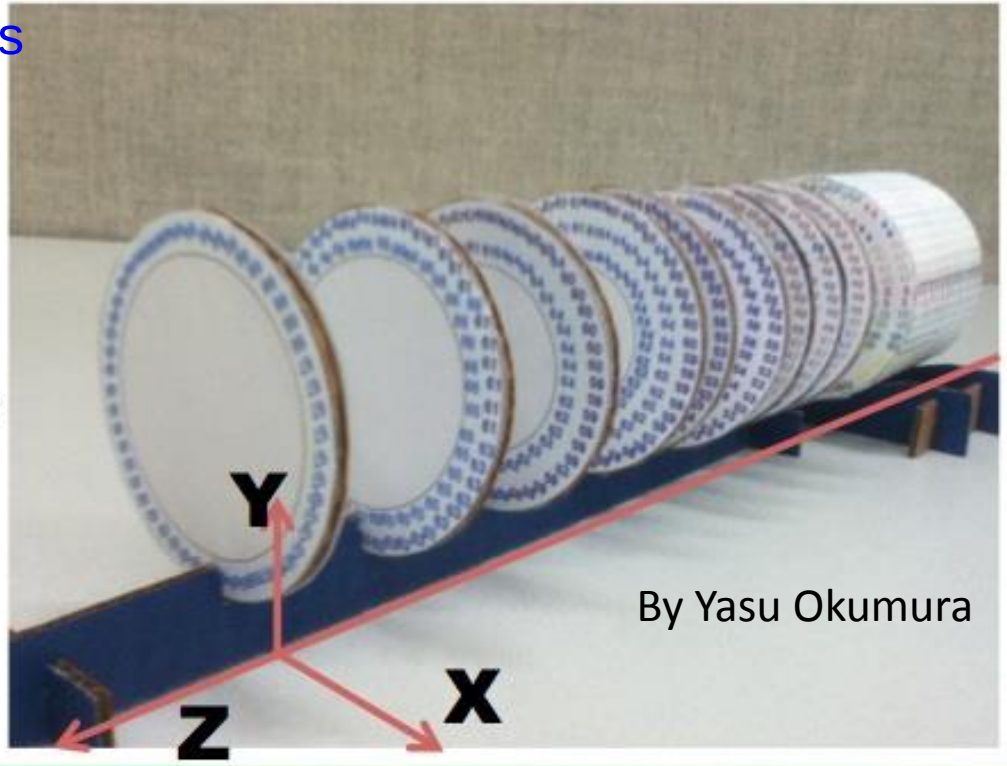
Detailed beam data analysis *using exact module/ROD cable mapping to trigger towers*

From “Data Formatter Design Specification”, Fermilab-TM-2553-E-PPD.

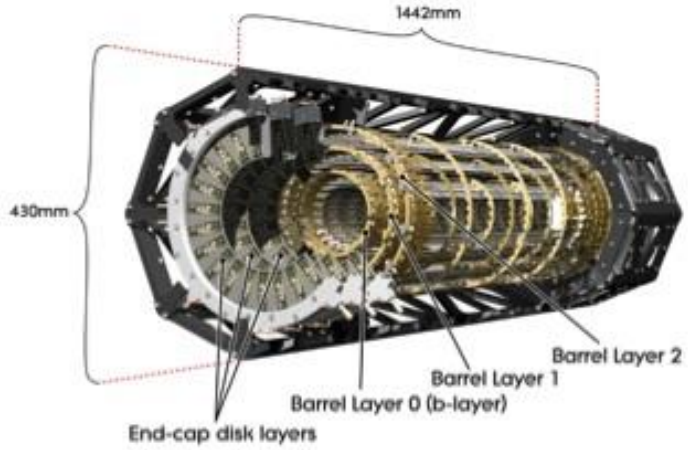
FTK Data Formatting Challenges

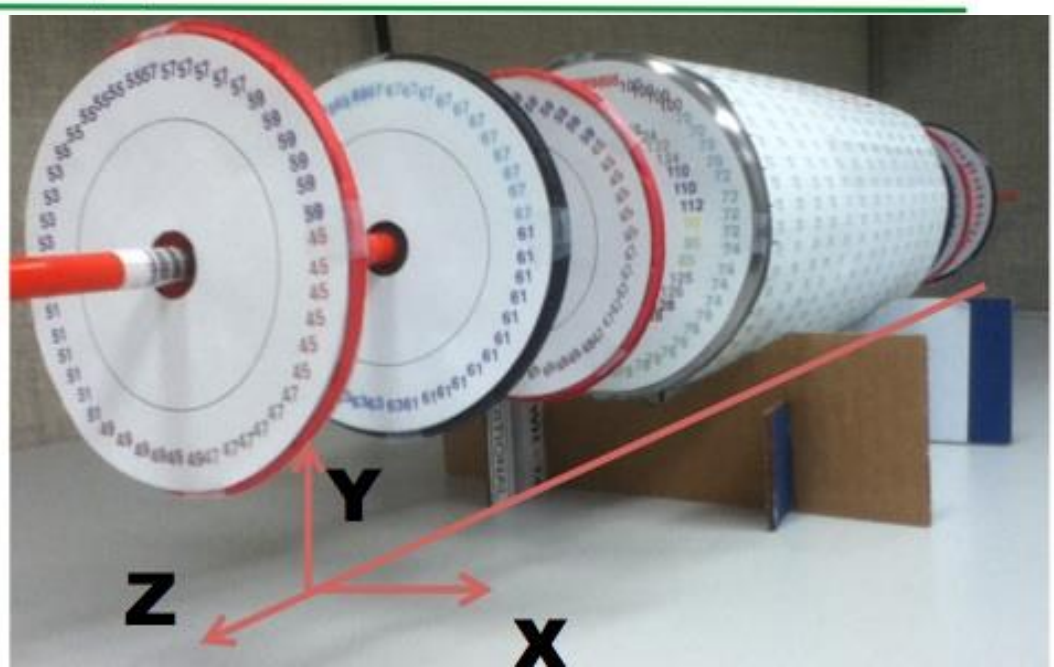


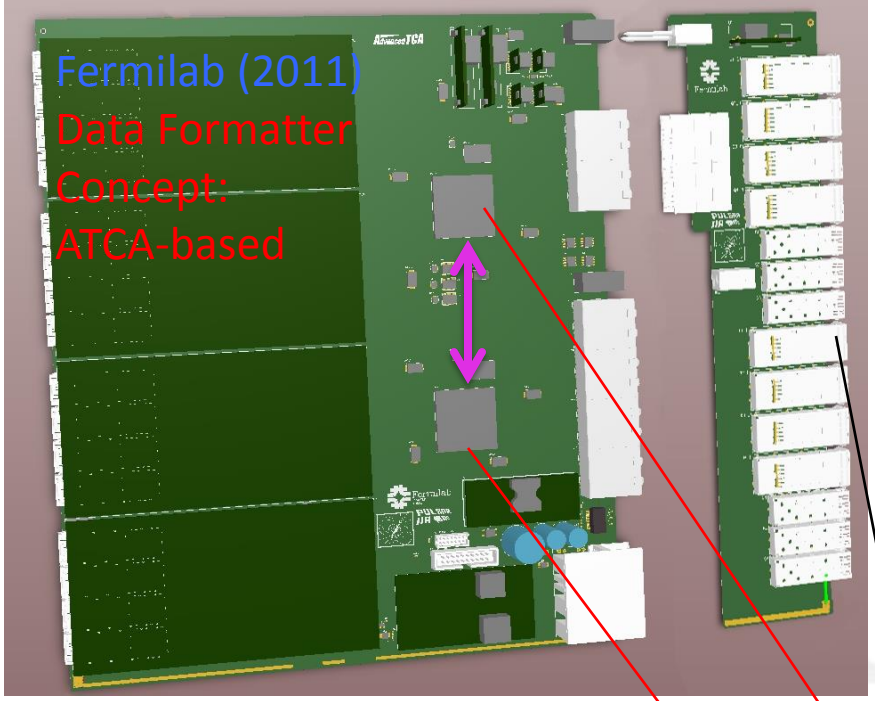
3D paper models at Fermilab with detailed ROD ID mapping, to help us understand the issues



By Yasu Okumura

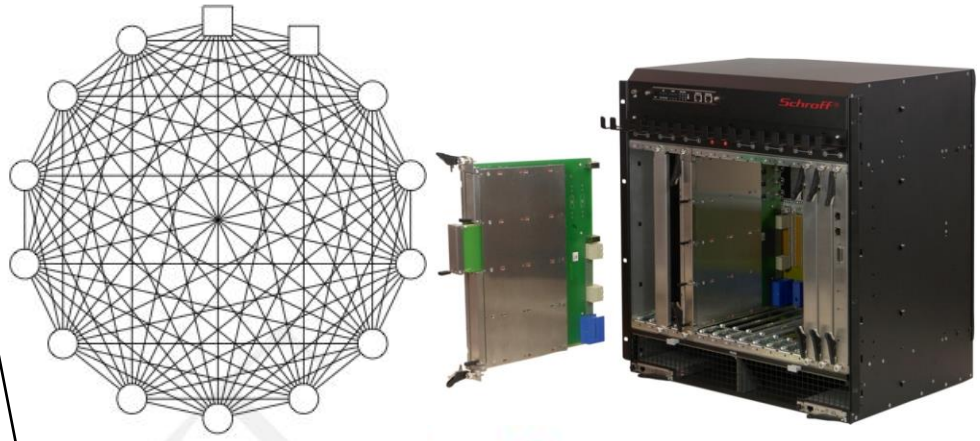


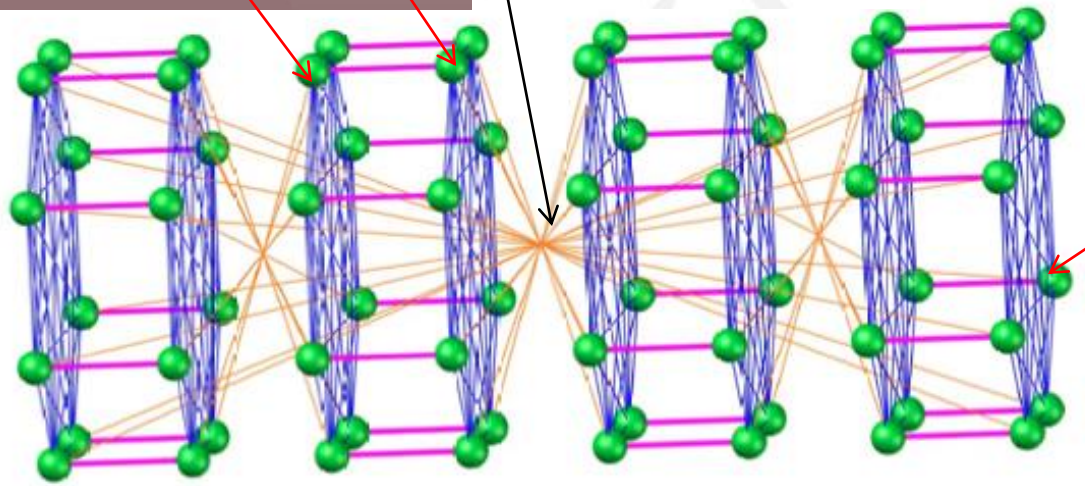


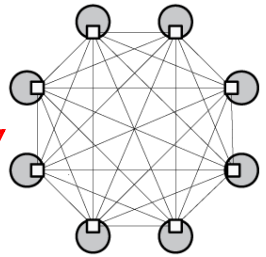


How to use ATCA backplane for data sharing

(1 FPGA per trigger tower, 2 trigger towers per board, 32 boards needed over 4 ATCA shelves)







ATCA Full-mesh backplane

Only data sharing links shown, not inputs/outputs

Figure 8: A 3D representation of FPGA interconnects in the Data Formatter system. 64 FPGAs (green) are connected through the ATCA backplane Fabric Interface (blue), local buses (purple) and inter-shelf links (orange). Each FPGA uses one inter-shelf link. This

Appendix P Unconstrained Data Volume Study

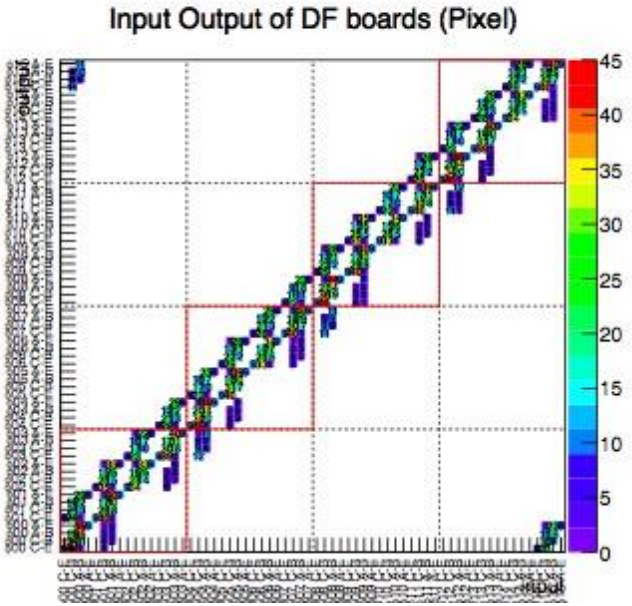
As previously mentioned the inner detector readout system was not originally designed for a track trigger. Modules were connected to RODs to minimize data rates and balance bandwidth. In this section we consider Data Formatter performance assuming an idealized module-ROD and ROD-DF mapping.

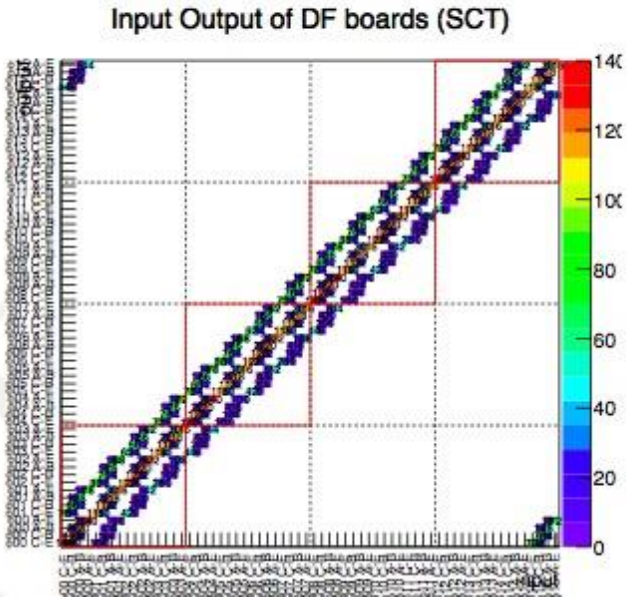
The band would be much cleaner without ROD constraints:

P.1 Data Sharing

The cabling was done to optimize for DAQ readout, not trigger

Refer to Figure 15 to compare these idealized results with the "real world" module-ROD cabling constraints.



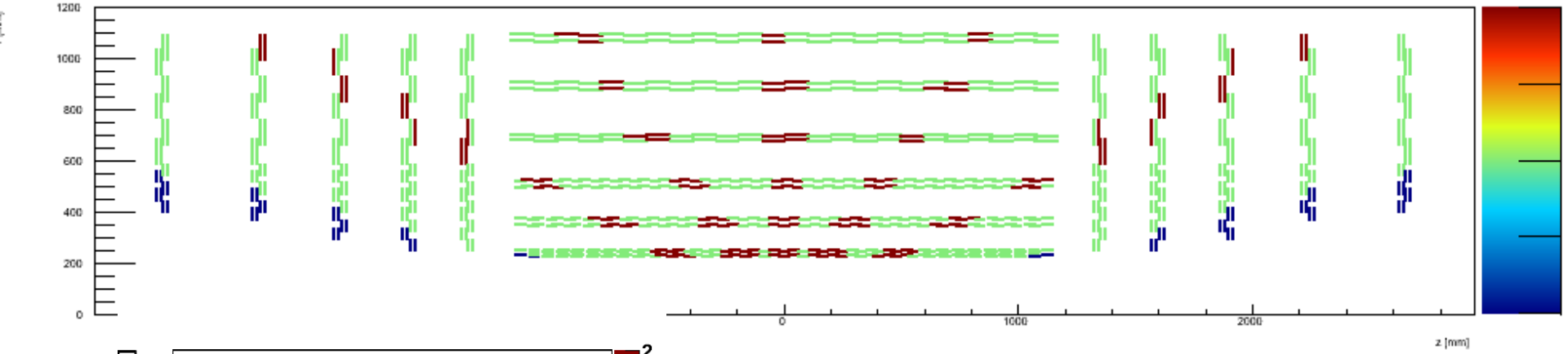


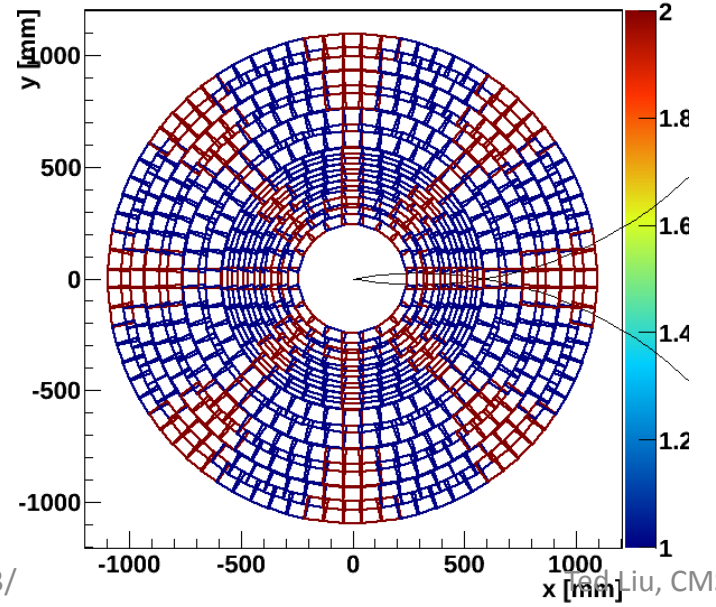
From "Data Formatter Design Specification", Fermilab-TM-2553-E-PPD, page 78.
Available at: http://www-ppd.fnal.gov/EEDOffice-w/Projects/ATCA/
(Pulsar IIa design spec)

CMS Tracker Layout and Trigger Tower (6 in eta x 8 in phi)

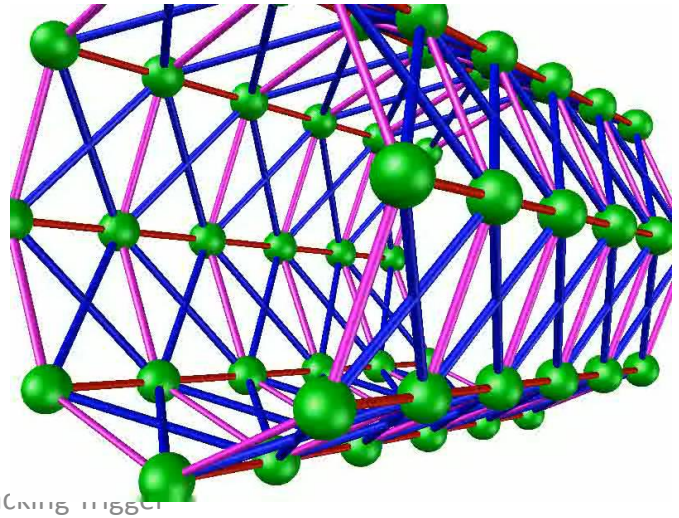
- 15K modules (see talk by Stefano Mersi this morning)

What we learned from FTK data formatting helps to understand CMS L1 case





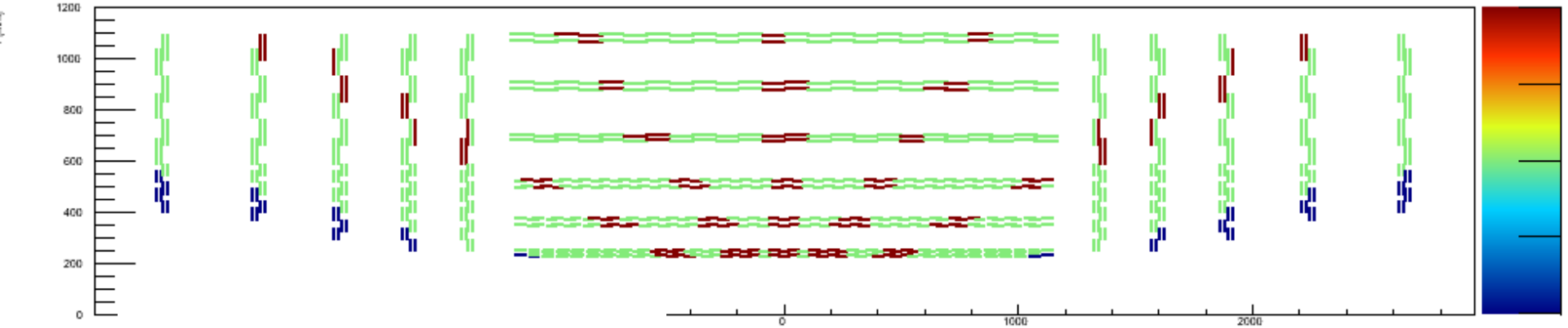
Animation by Jamieson Olsen (FNAL engineer)

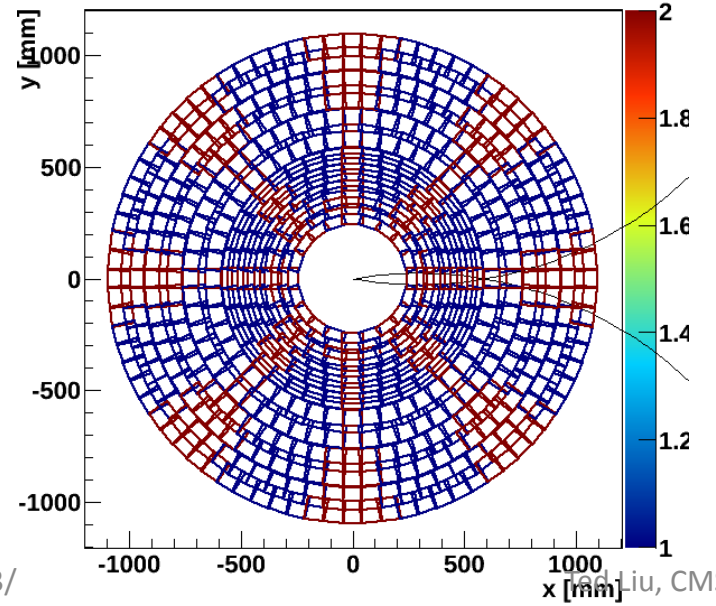


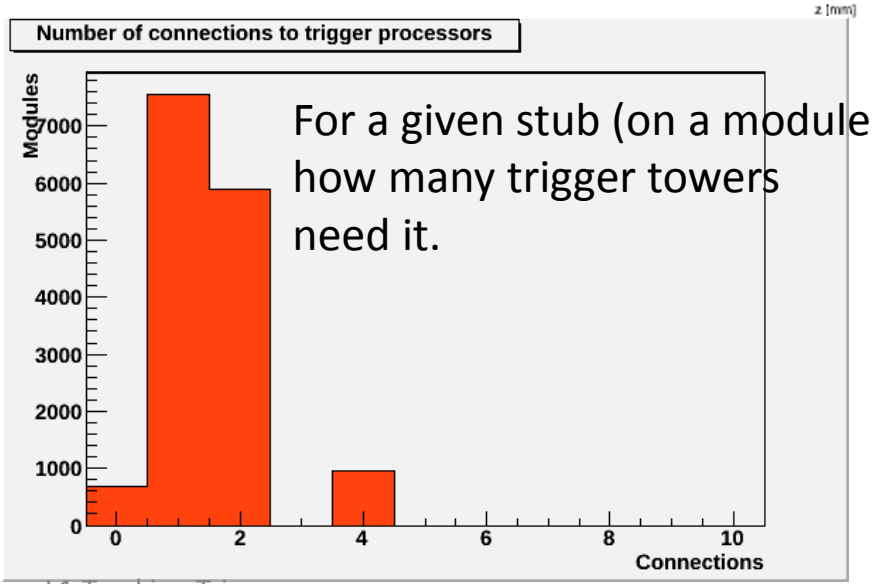
CMS Tracker Layout and Trigger Tower (6 in eta x 8 in phi)

- 15K modules (see talk by Stefano Mersi this morning)

What we learned from FTK data formatting helps to understand CMS L1 case

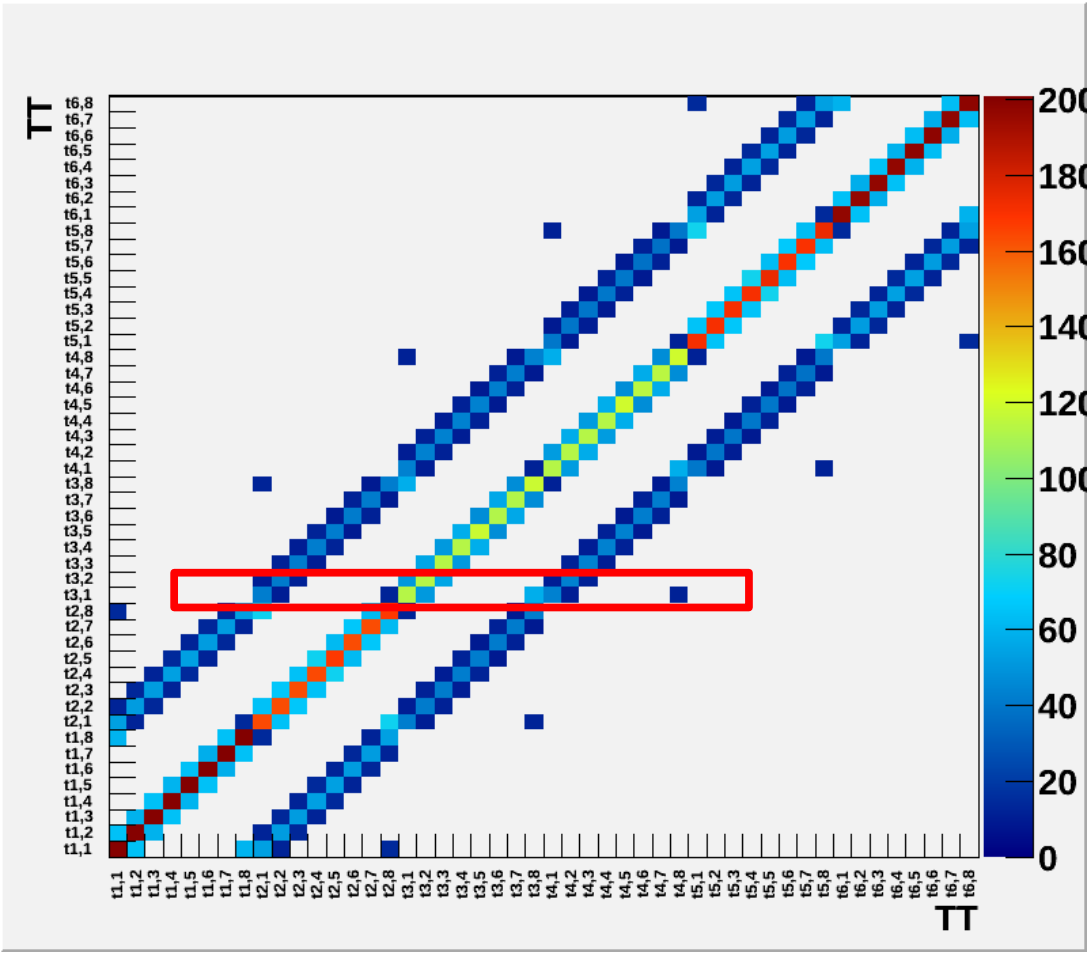






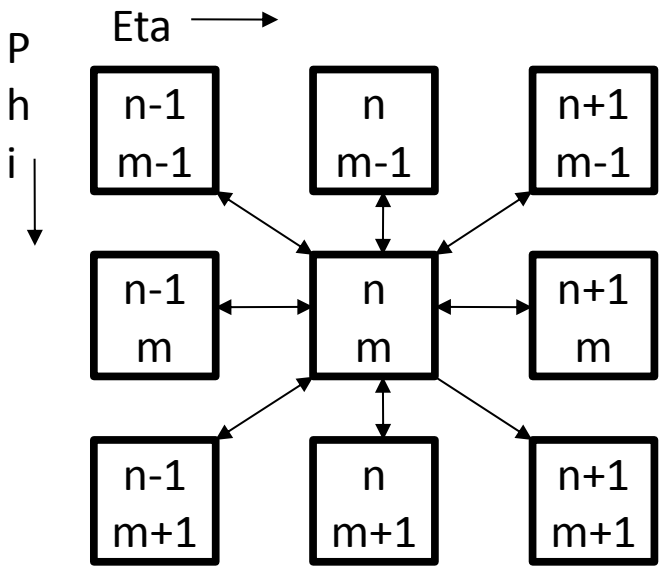
For a given stub (on a module), how many trigger towers need it.

CMS Trigger Tower Data Sharing

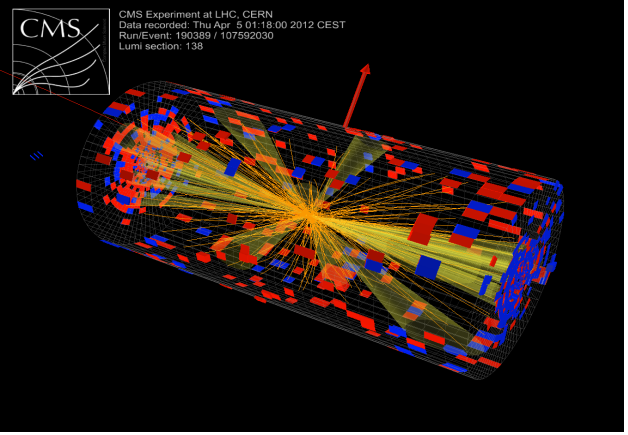


Study done by Giovanni Bianchi (CERN)

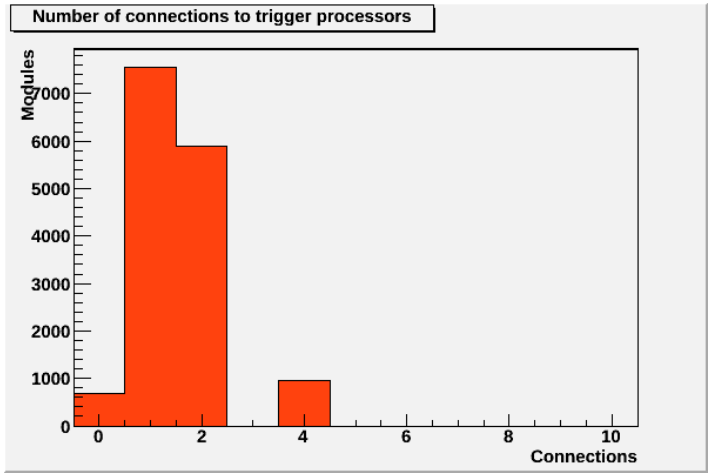
- In other words,
  A Trigger Tower shares at most with 8 other Trigger Towers

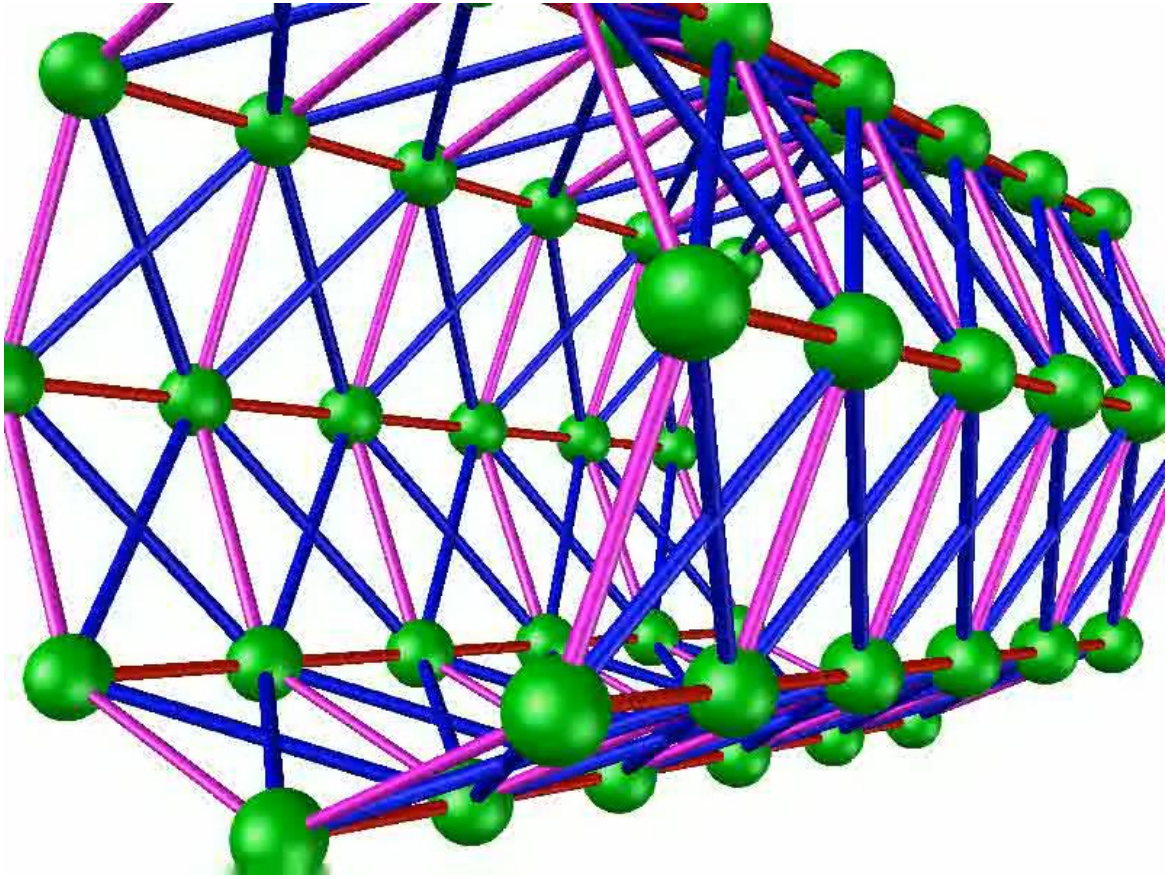


- *In fact,*
  Only with immediate neighbors



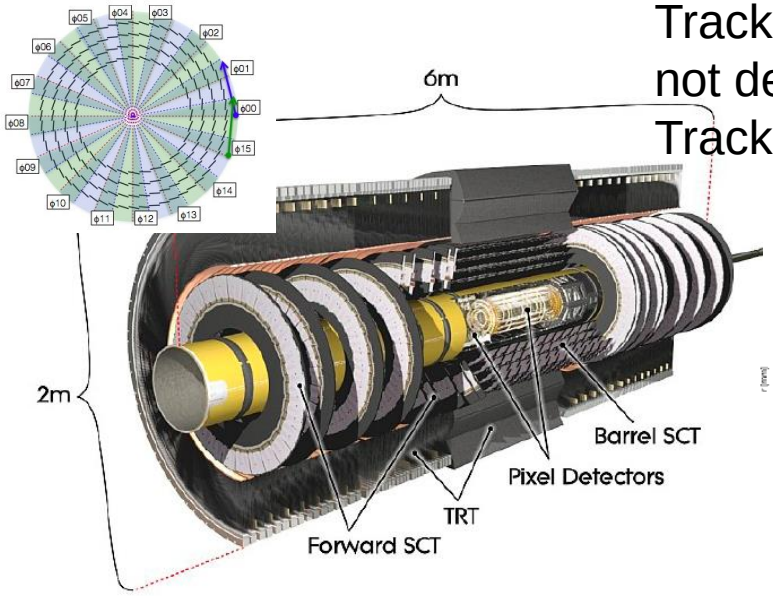
CMS L1 tracking trigger for Phase II: 6 (in eta) x8 (in phi) = 48 Trigger towers & their interconnections





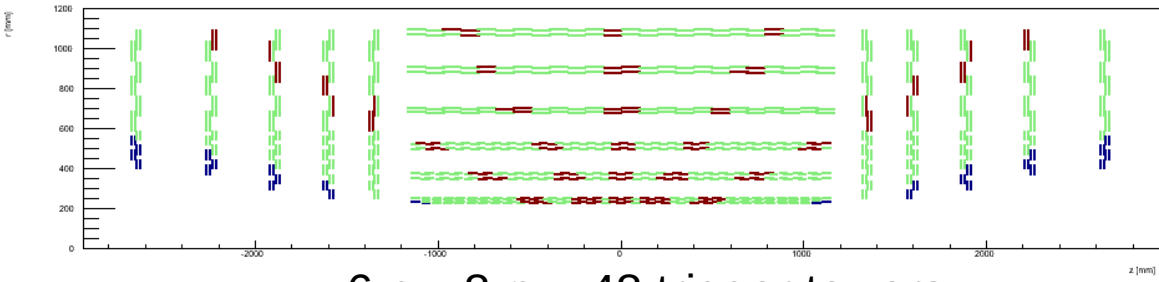
Data coming from a given trigger tower may need to be delivered to multiple trigger towers.
This happens,
when a stub comes from a detector element is close to the border Between trigger towers, due to the finite curvature of charged particles in the magnetic field and finite size of the beam luminous region along the beam axis.

Comparison: ATLAS L2 FTK and CMS L1 Track Trigger



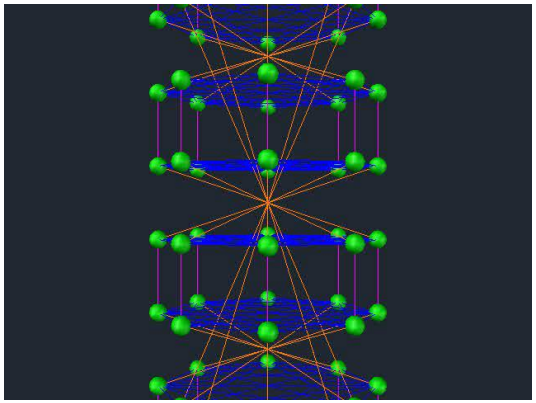
Tracker/cabling not designed for Tracking Trigger

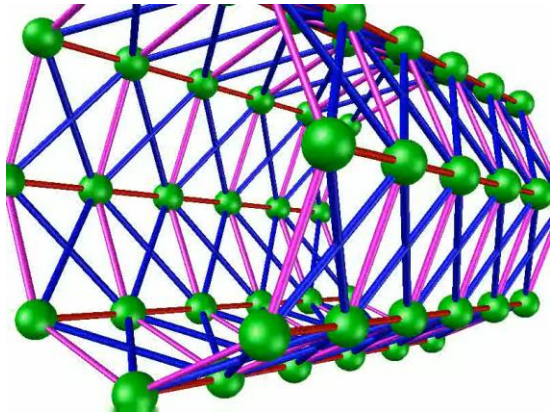
Tracker/Cabling designed for Tracking Trigger



4 η x 16 φ = 64 trigger towers

6 η x 8 φ = 48 trigger towers

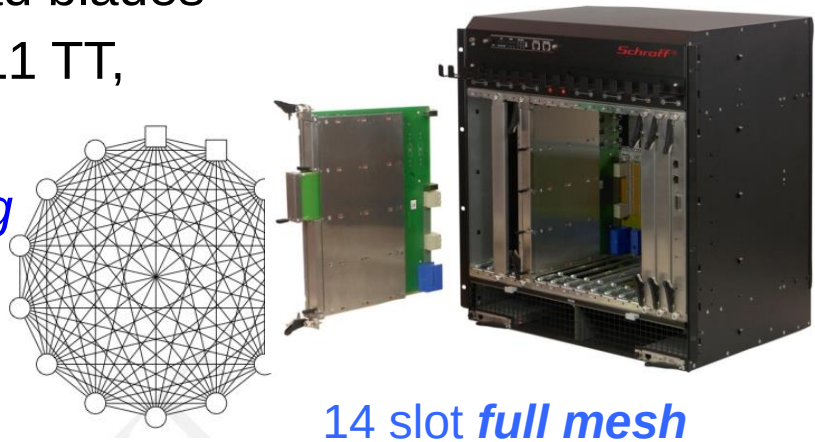




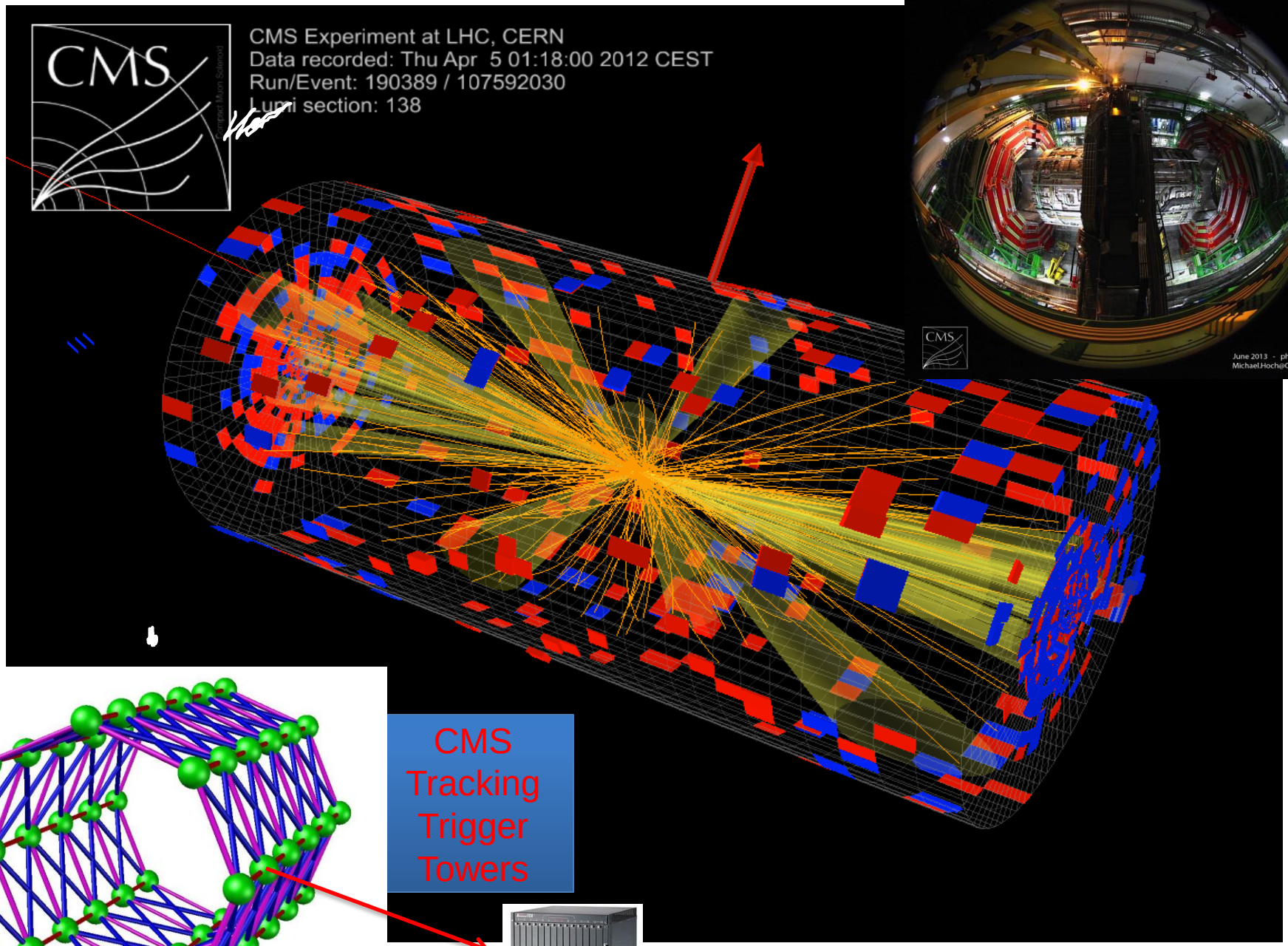
This leads to new Architecture….

General considerations for the tower processor Platform for silicon based tracking trigger system

- The tower processor platform must support large numbers of fiber transceivers, used for receiving input links and data sharing
- A flexible, high bandwidth backplane is desirable to quickly transfer data between boards
- The boards should be large enough to support pattern recognition engines and fiber connections, in a comfortable way
- ***A Full Mesh***, 14 slot ATCA shelf is a natural fit as the platform with 12 slots available for processor or payload blades
- This applies to both Atlas FTK and CMS L1 TT, but *architecturally they are very different:*
  Atlas FTK: full-mesh used for data sharing
  CMS L1 TT: full-mesh mostly used for time-multiplexing



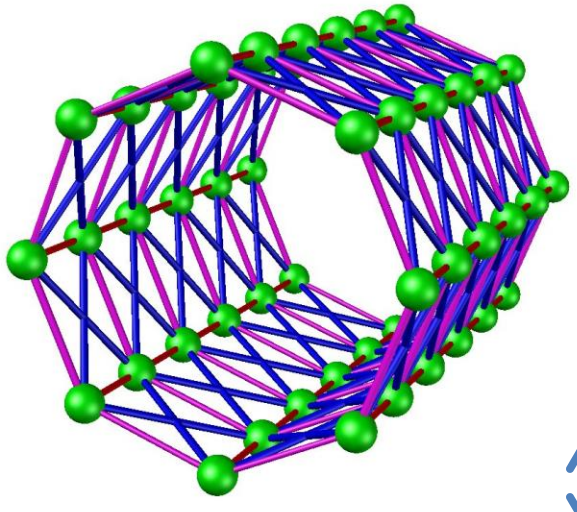
14 slot ***full mesh ATCA backplane***:

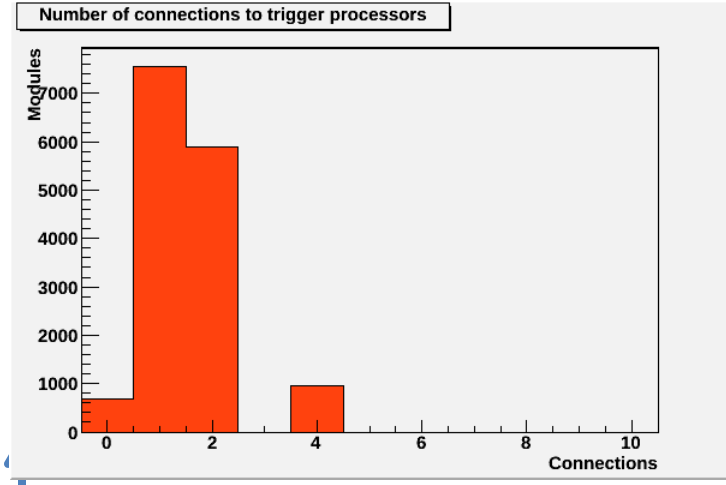


CMS Tracking Trigger Towers

ATCA

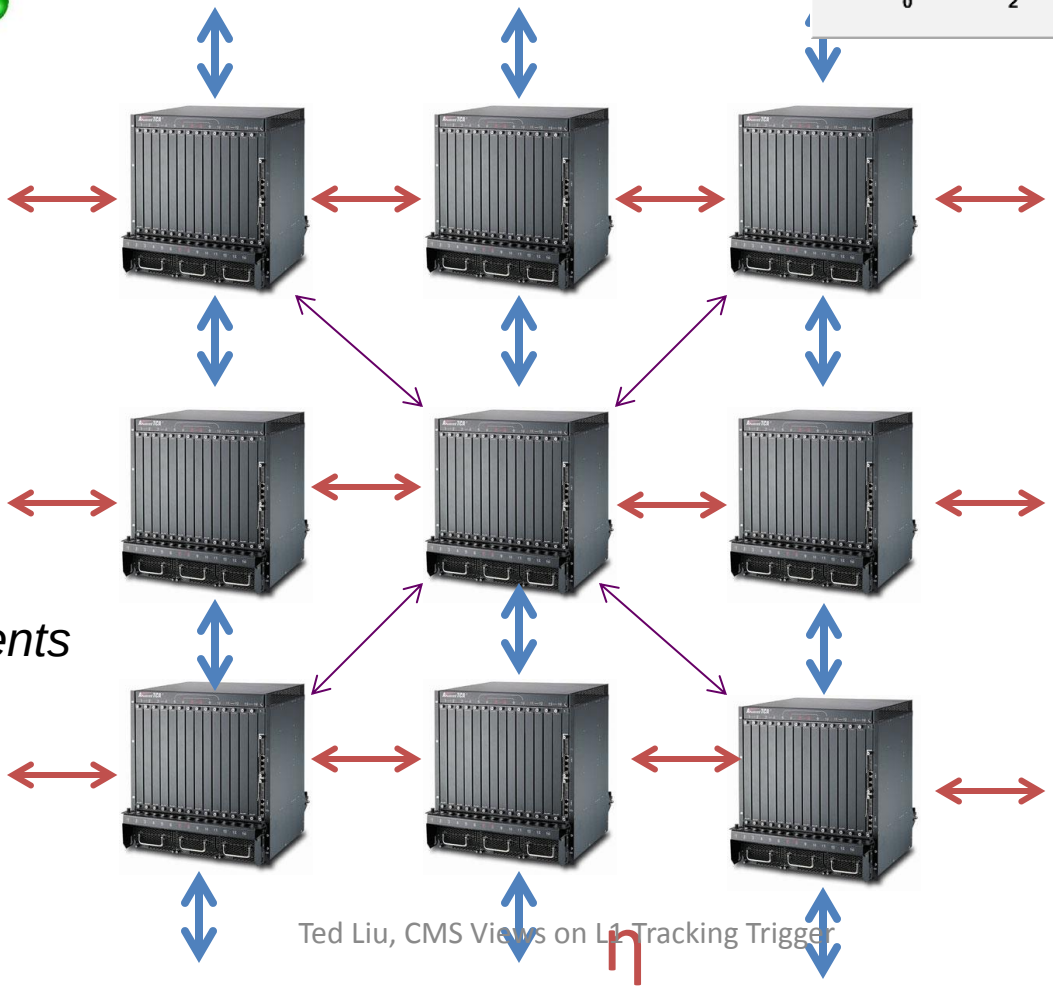
For simplicity, let's assume one crate is assigned to one trigger tower

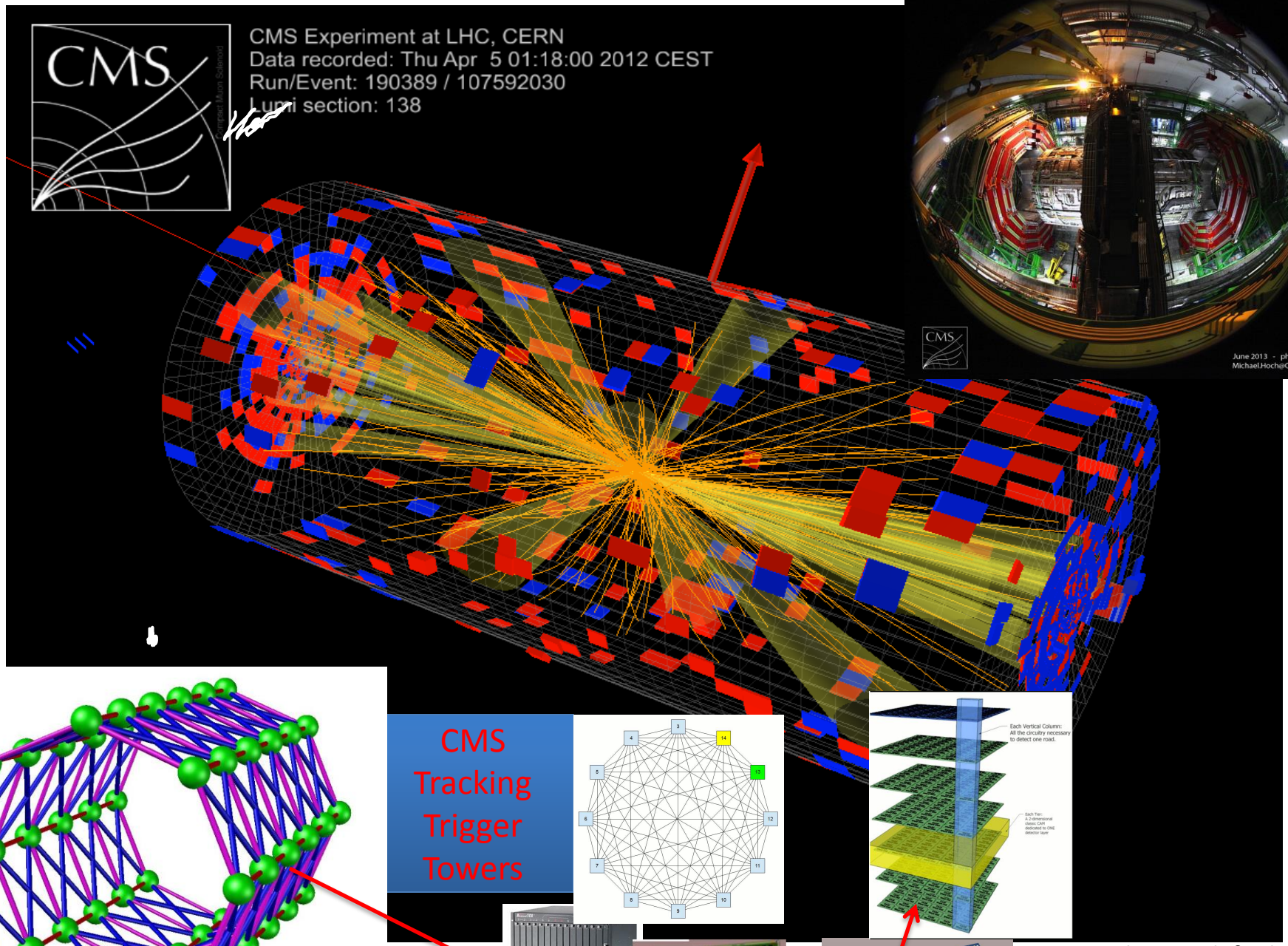




Simple
Trigger Tower
Interconnections.

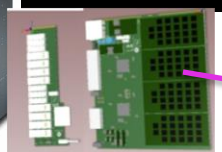
Each box represents a trigger tower.

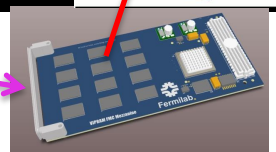




CMS Tracking Trigger Towers

ATCA





AM or other track finding approaches implemented on mezzanine (PR engine)

Commericial HUB boards

One simple configuration as an example

4 Pattern Recognition Boards

ATCA Full-Mesh Backplane

8 Data Input Boards

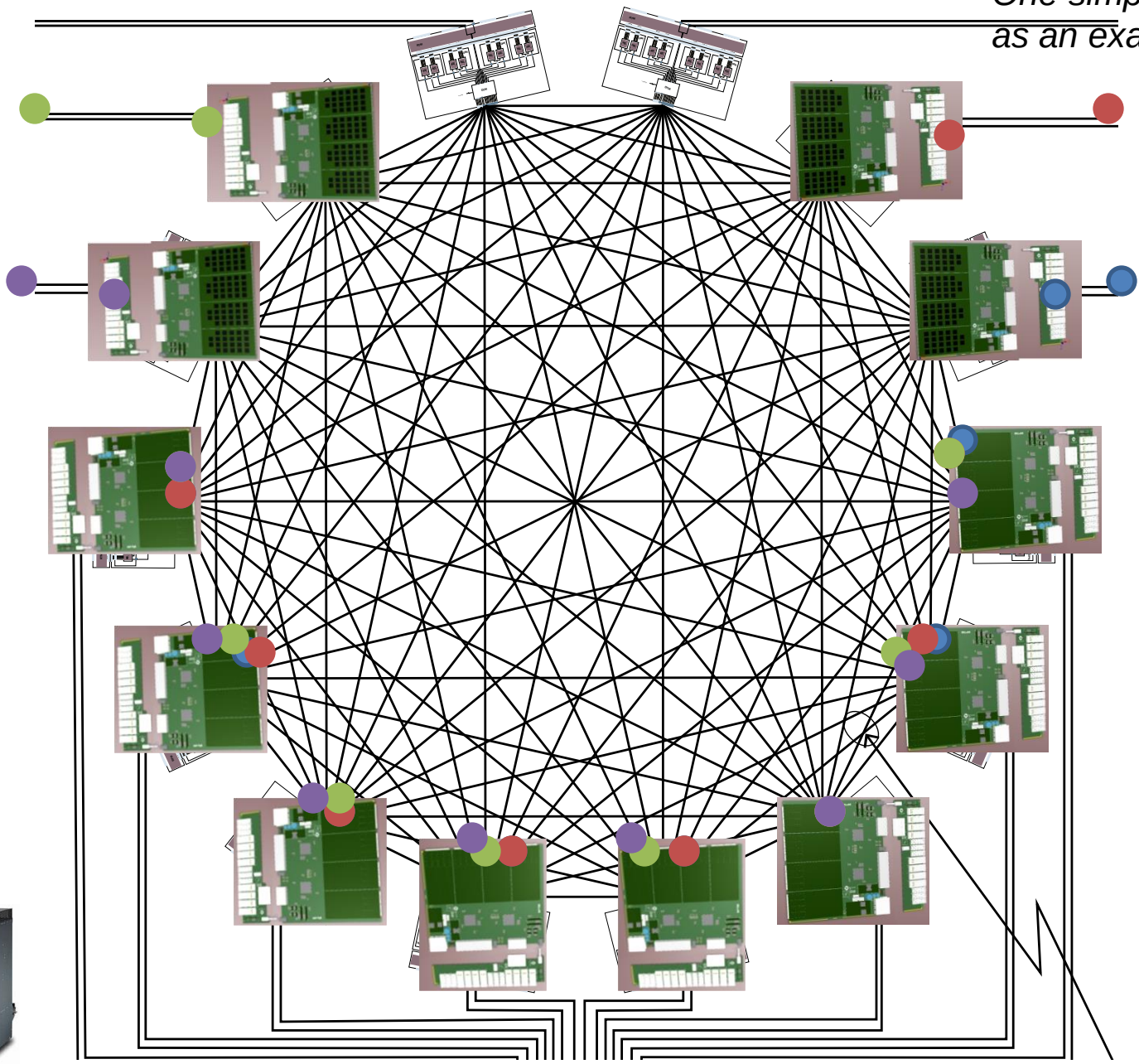
What's inside one trigger tower/crate

Fibers from upstream

40Gbps

Commericial HUB boards

One simple configuration as an example



4 Pattern Recognition Boards (PRB)

8 Data Input Boards

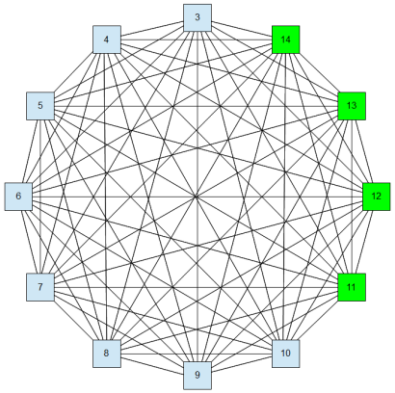
Fibers from upstream

Commericial HUB boards

One simple configuration as an example

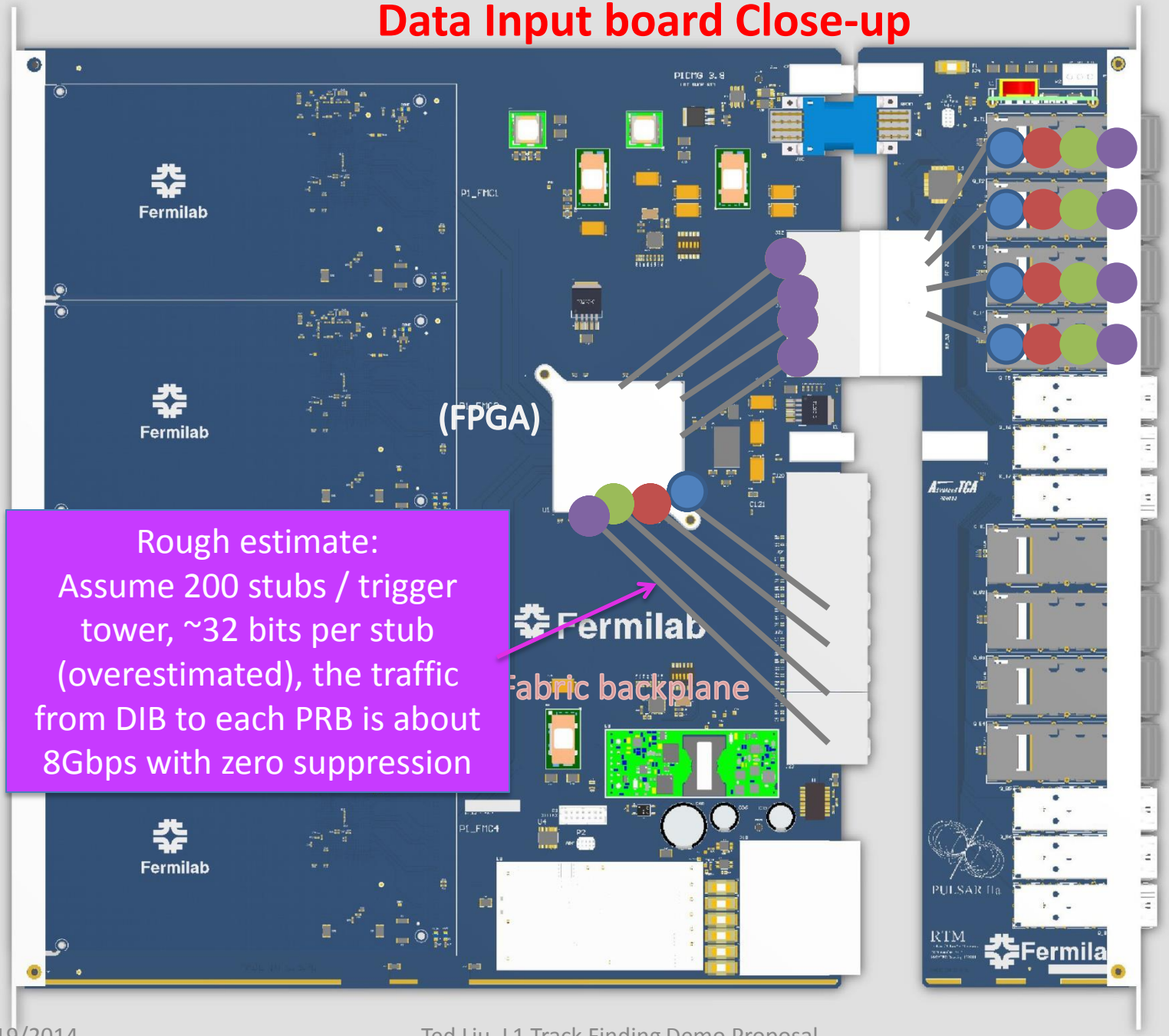
4 Pattern Recognition Boards

8 Data Input boards



Fibers from upstream

Data Input board Close-up

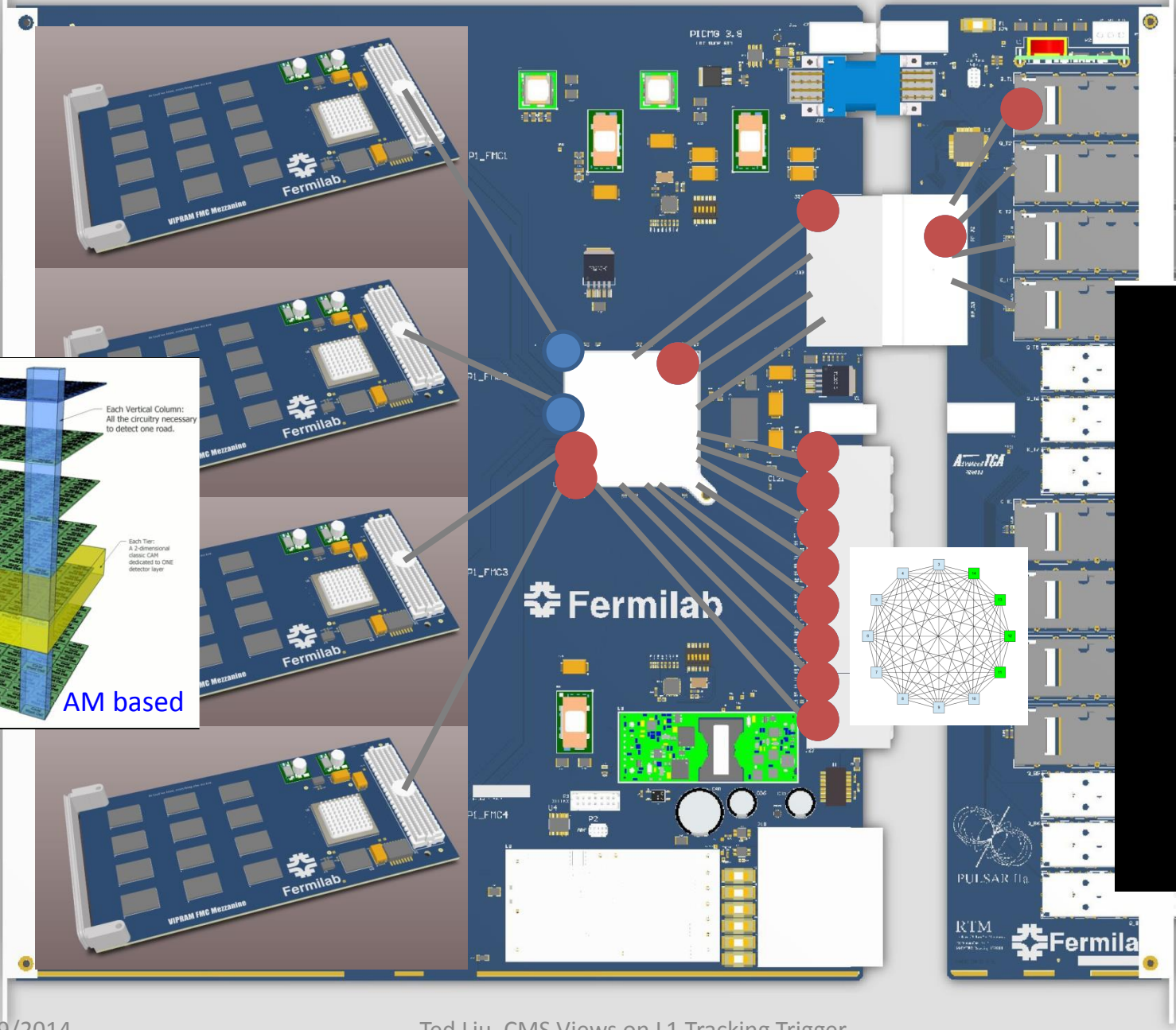


RTM:

Fibers from upstream

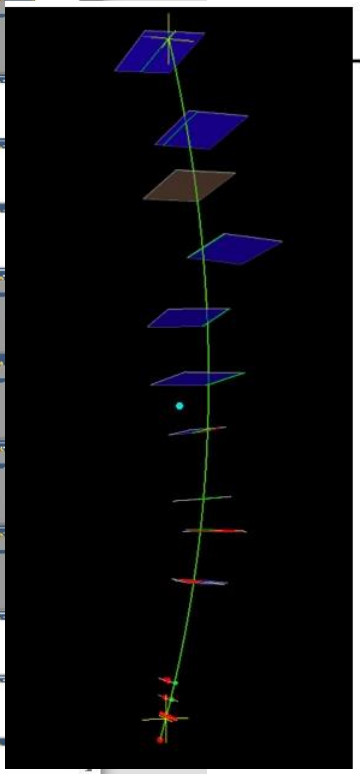
Rough estimate:
Assume 200 stubs / trigger tower, ~32 bits per stub (overestimated), the traffic from DIB to each PRB is about 8Gbps with zero suppression

Pattern Recognition Board (PRB) data flow



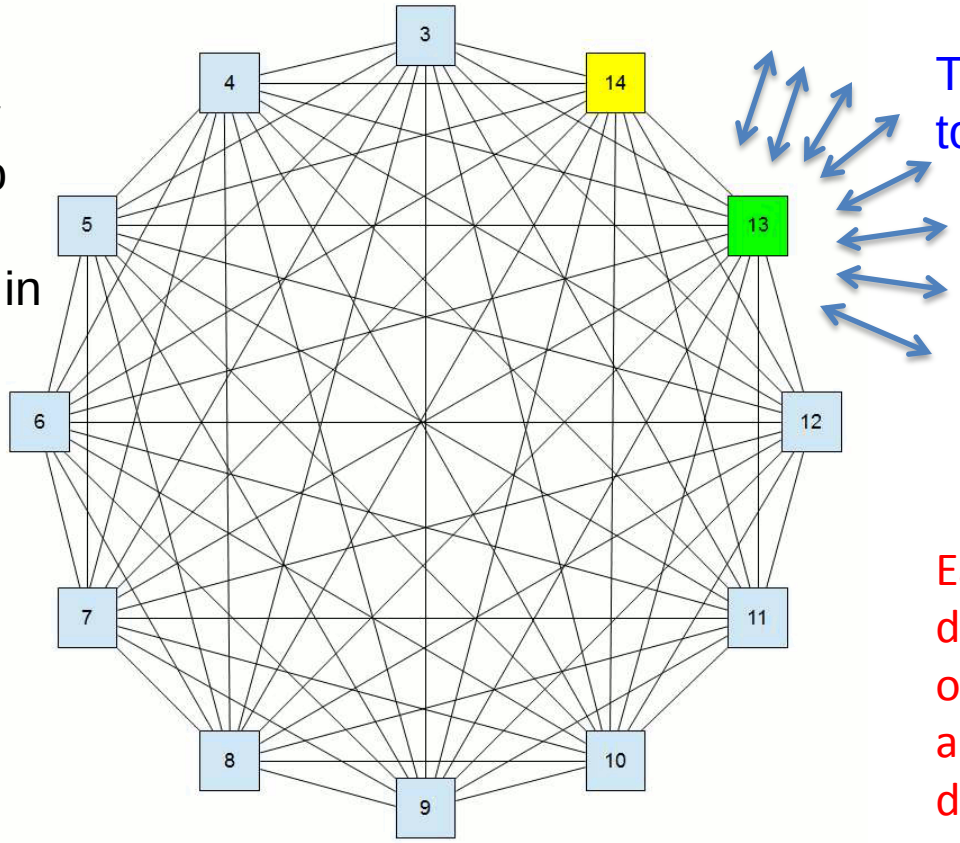
Neighbor data sharing to/from another Trgger Tower

AM based



More advanced configuration

Ten Processors and the Gateway send the event to the target Processor Blade in a round robin scheme.



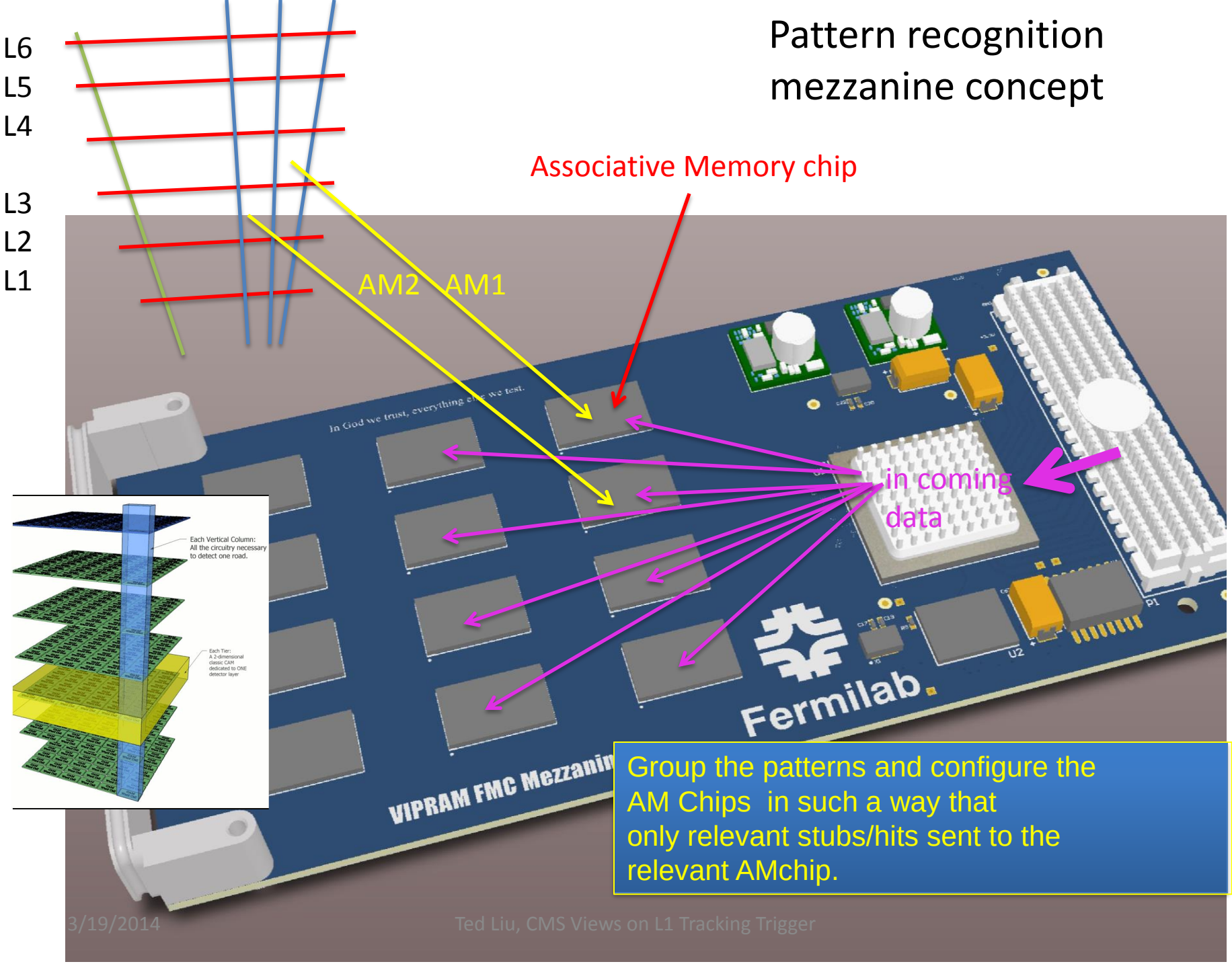
To/from eight neighbor towers

Each processor receives data directly from upstream on RTM,
and then take turn to handle different event for each trigger tower

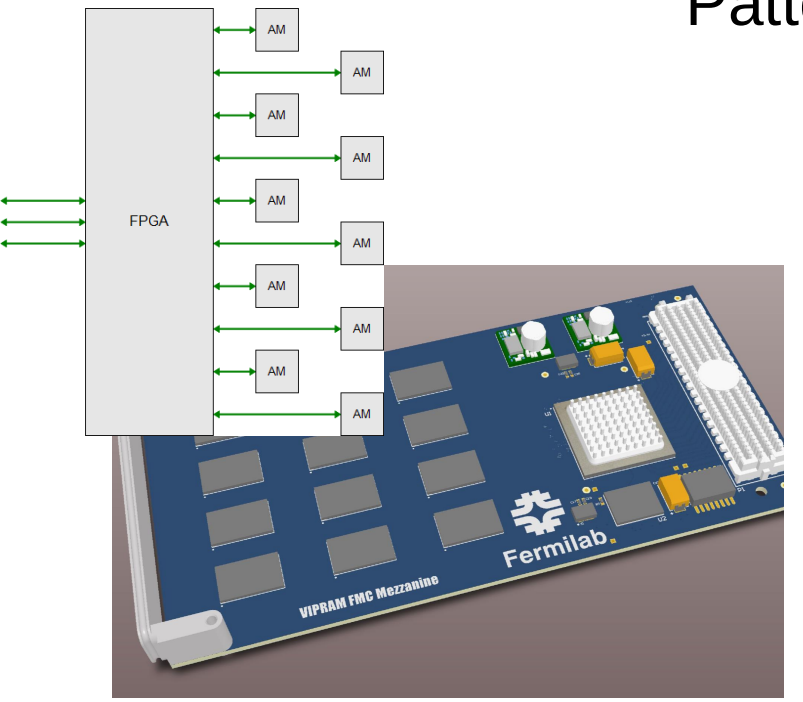
The full mesh based architecture is highly flexible.
Many performance and bandwidth bottlenecks can be solved/avoided/relaxed simply by better configurations.

This also makes an early technical demonstration feasible using today's technology.
The flexible architecture is a good platform for a vertical slice demonstration and beyond.

Pattern recognition mezzanine concept



Pattern Recognition Mezzanine (PRM)



System size shrinks with better AMchip performance:

If 2X more AM pattern density,
or 2X higher AM speed,
→ 2 x less system size
(48 crates → 24 crates)

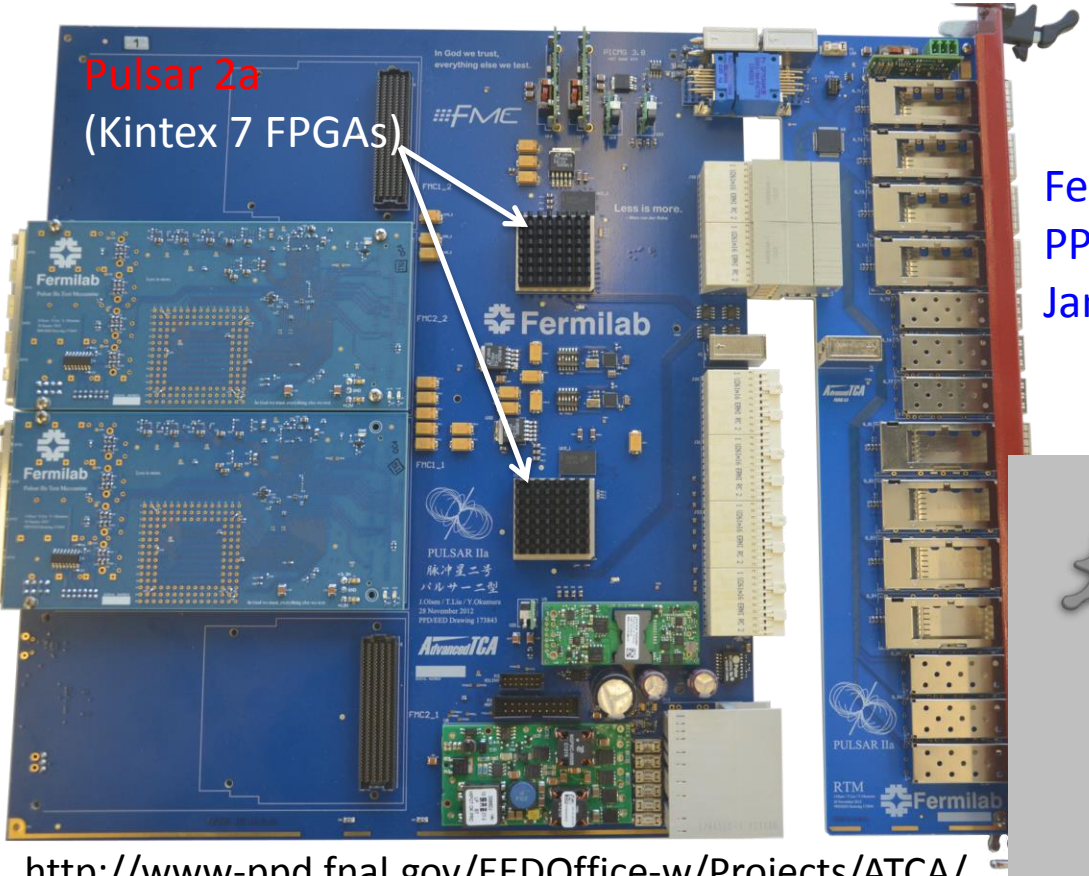
Relaxed Performance Requirements (in the case of 10 PRBs with ~40 PRMs):

- 40MHz input handled by 40 PRM mezzanines in round robin, each handles ~1MHz event input rate
- Event Processing >= 1MHz (out of 40MHz)
- Input BW >= 16Gbps

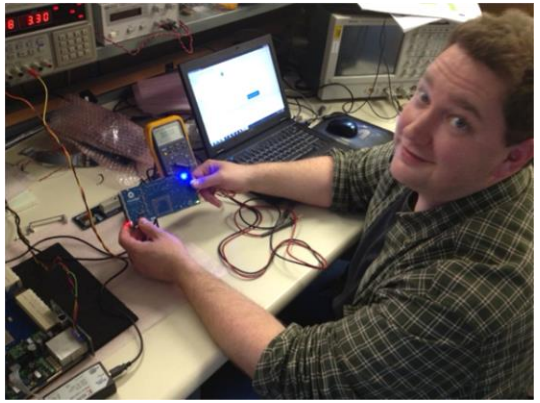
- In the case of AM approach:
  - ~10 AM chips / PRM
  - ~200k patterns / AMchip
  - ~ 2M patterns / tower
  - (2M x 48 towers ~ 100M patterns)

The relaxed performance requirement would make early technical demonstration easier for different track finding approaches.

First Pulsar 2a prototypes work well: "plug & play" (summer 2013)

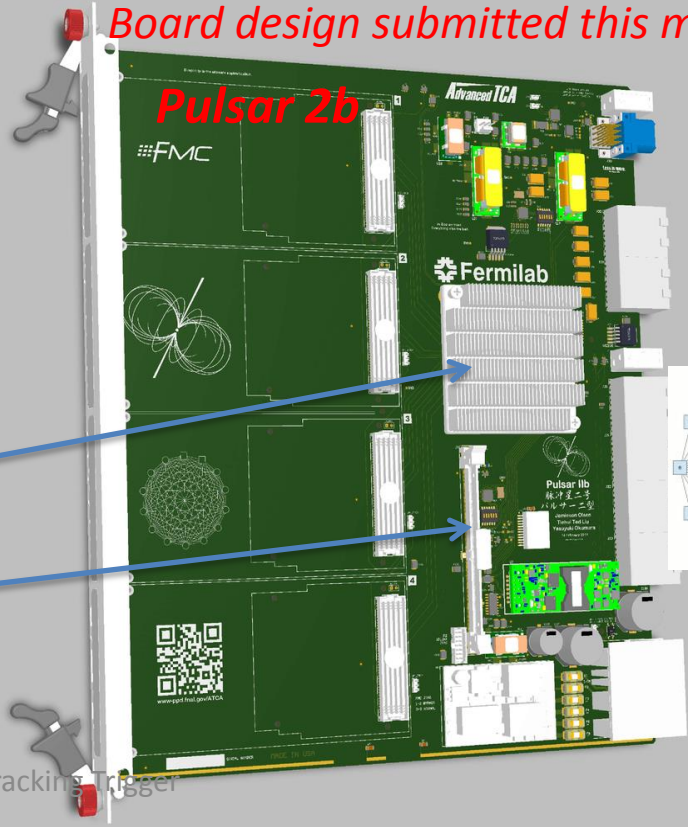


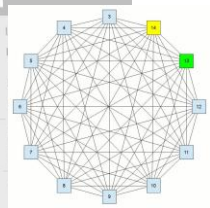
Fermilab
PPD engineer
Jamieson Olsen



http://www-ppd.fnal.gov/EEDOffice-w/Projects/ATCA/

Board design submitted this month



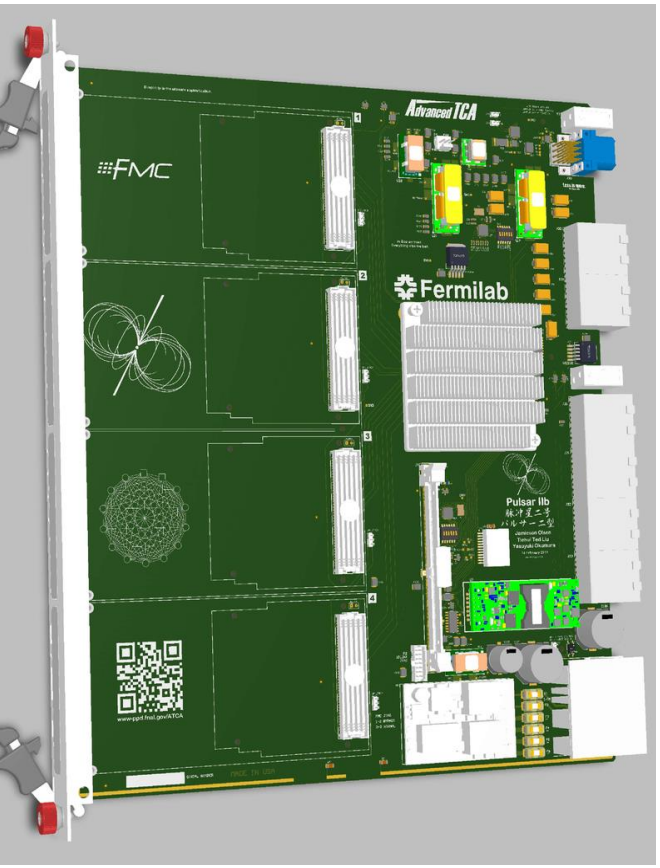


Pulsar 2b:

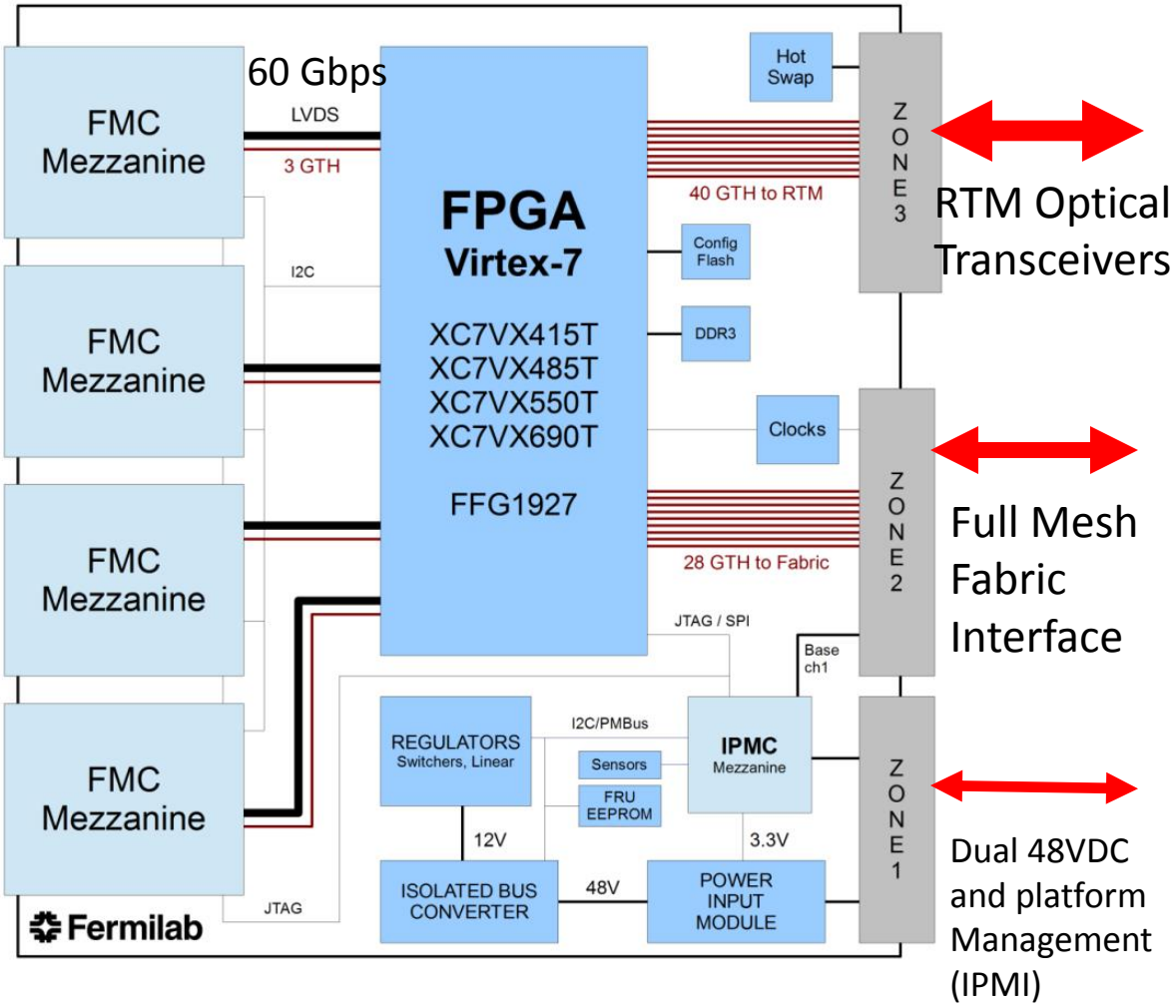
- ***Vertex 7 FPGA*** (XC7VX690T)
- ***80 GTH lines***
- Compatible with LAPP IPMC module
- FMC TTC compatible, backplane clock dist.
- Plan to use CMS IPBus user interface
- General purpose design
- I/O ~ 1 Tbps

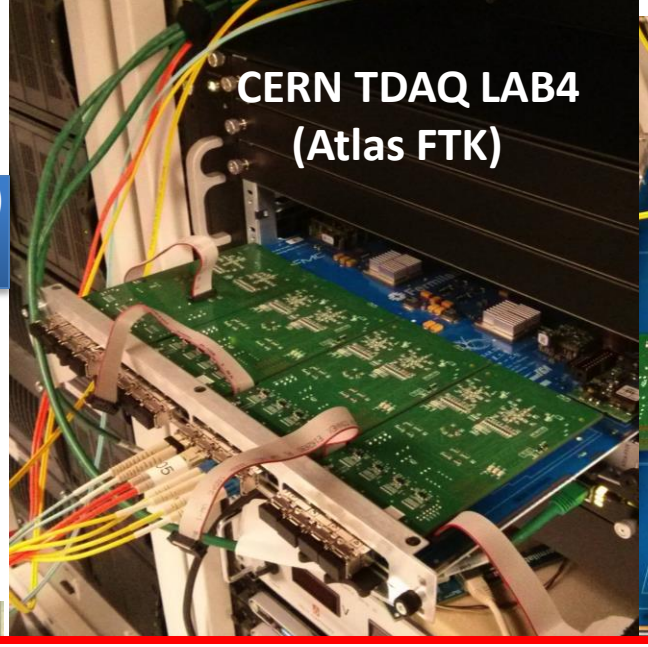
How full-mesh used in CMS L1 TT case

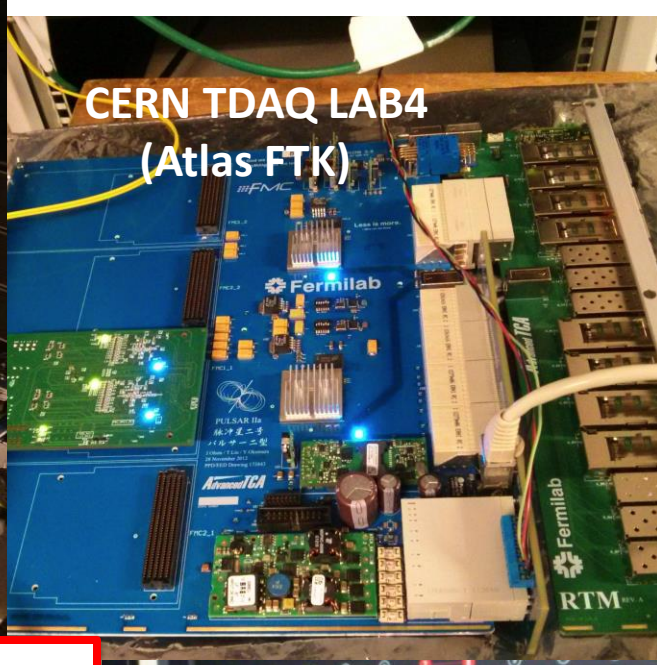
Pulsar 2b Block Diagram



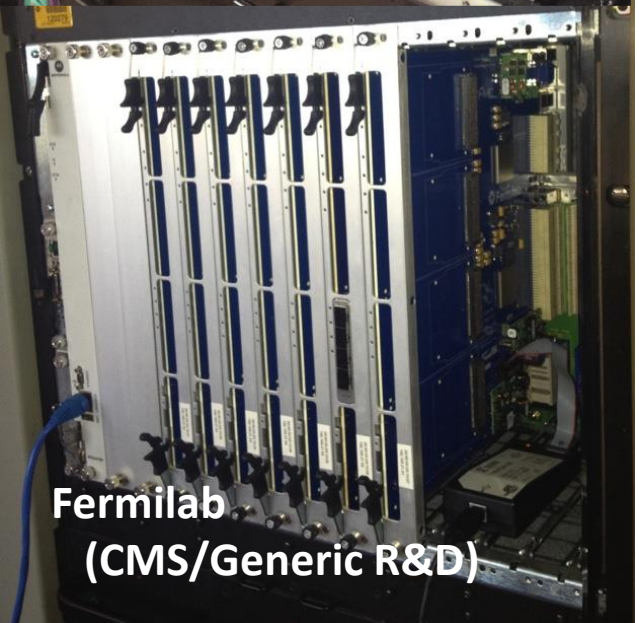
I/O: ~ 1 Tbps

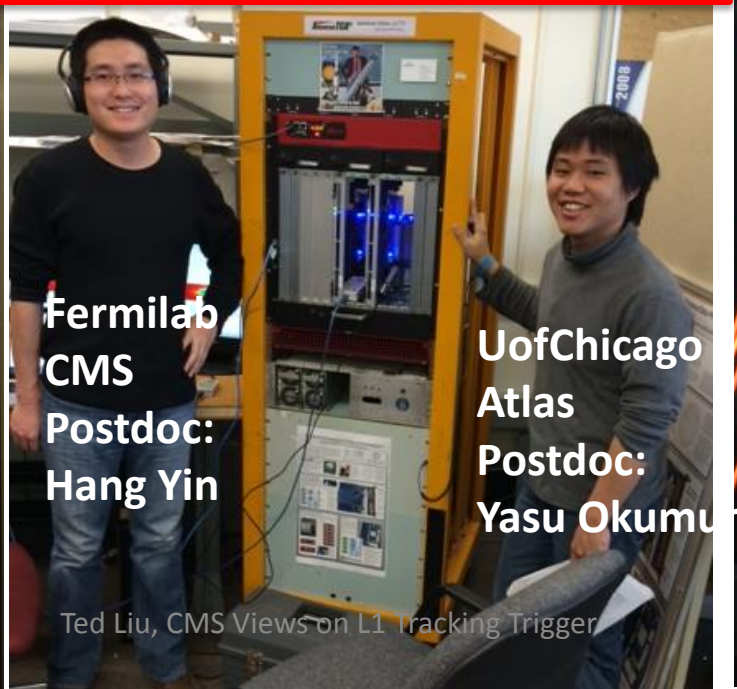


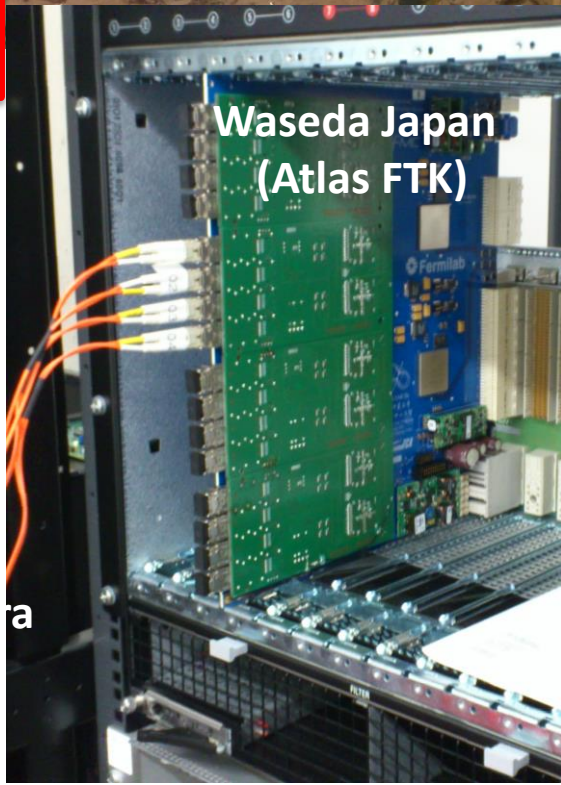




Existing Pulsar2 ATCA Teststands
(Atlas/CMS Common Electronics)

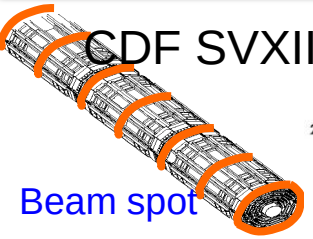


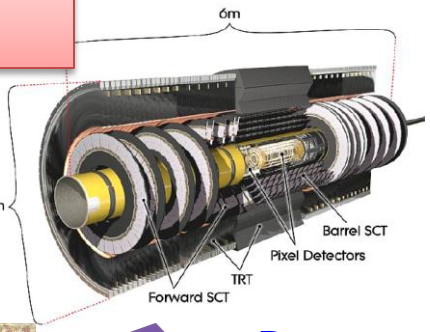




Detector design for triggering

Silicon Based Tracking Trigger at Hadron Colliders





Data transfer

Data formatting

Partition detector into trigger towers/sectors

Pick your favorite method:

Associative Memory (AM) Approach (proven approach from CDF/SVT)

Hough Transformation

tracklet-based

Adaptive Pattern Recognition

Biology Inspired …

your choice here…

Pattern Recognition

Finer pattern recognition

Track Fitting

FPGA vs GPU vs CPU

HLT

- *In this talk (focus on off-detector)*

(1) Pattern Recognition + Track Fitting options

-- AM + linearized track fitting (FPGA): traditional

-- AM + Hough transform (FPGA): new, being studied

-- Tracklet-base approach (FPGA): new, being studied

(2) Data Formatting and System Architecture options:

-- FPGA based Full-mesh enabled ATCA approach (Pulsar 2)
(common electronics developed for Atlas FTK and CMS L1 Track Trigger R&D)

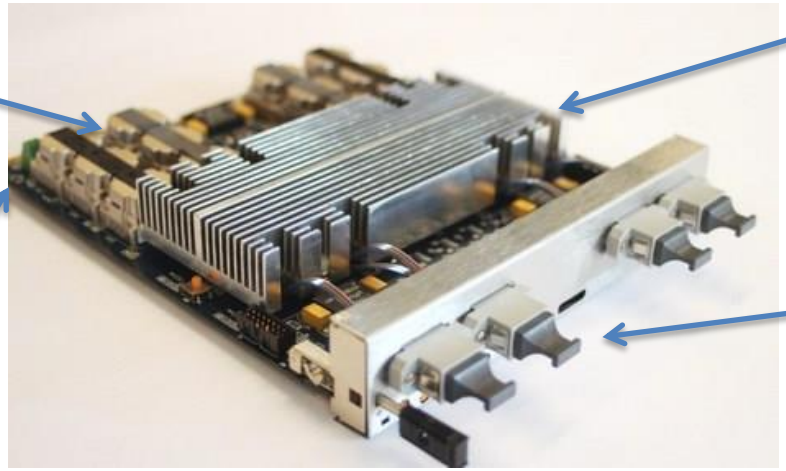
-- MicroTCA based (MP7): electronics developed for CMS L1 CAL trigger upgrade

(3) Summary (Vertical Slice Demonstration)

Slide from Mark Pesaresi

Alternative demonstrator for TM track trigger:

takes advantage of hardware & expertise developed for the L1 global calorimeter trigger upgrade: ***MP7 processing card***



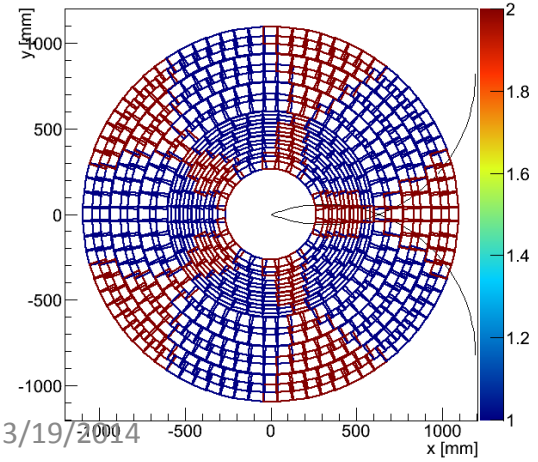
miniPOD 12 channel parallel optics up to 11.3Gbps per link

uTCA compatible (GbE, CMS DAQ, TTC, SATA, PCIe/SRIO)

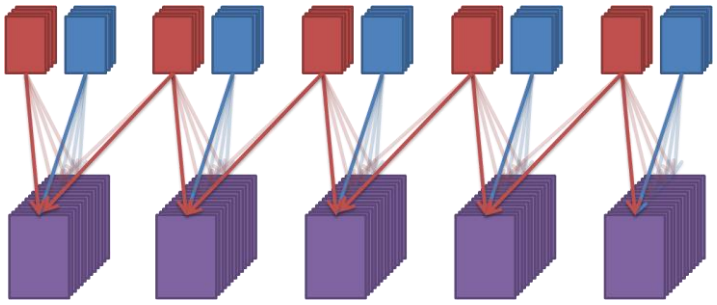
Xilinx Virtex 7 690T FPGA mature system & link firmware, ready for drop in algos

72 input / 72 output optical links => 0.9Tb/s total bandwidth

example implementation:
divide tracker up into 5 regions in phi



processors (purple) could build tracks in the FPGA, or data can be forwarded to AM ASICs



~230 FEDs
- input data from tracker
- output trigger data is formatted & time multiplexed

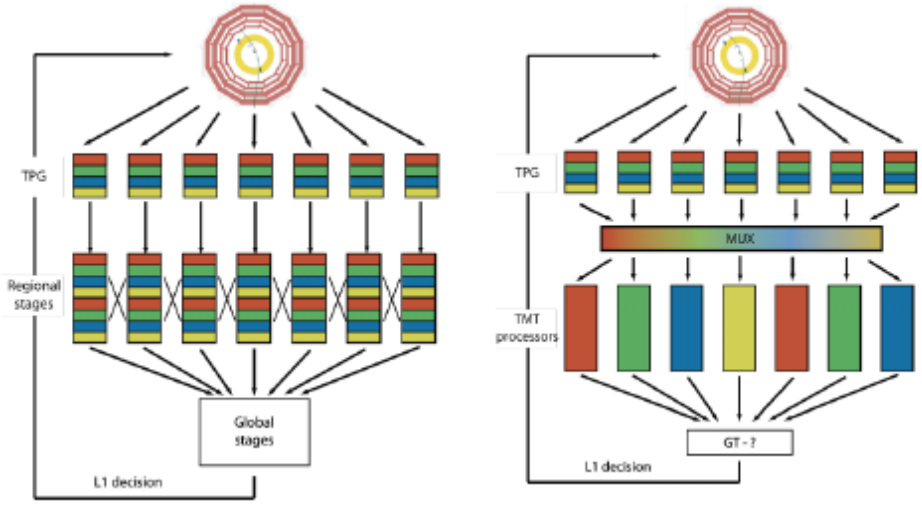
120 Processors
- each receive data over 24BX
- each processes one phi sector per event

More details see talk by Andrew Rose tomorrow morning.

Slide from Mark Pesaresi

architecture option

*make full use of the **time multiplexing technique (TMT)**, extending an architecture choice currently being implemented in the L1 calorimeter trigger*

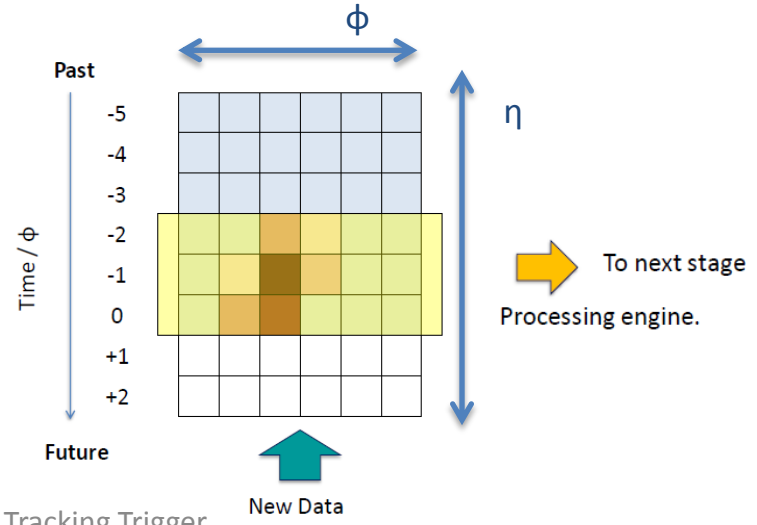


design motivations

*- **simplifies data flow** problems, no need for track finding/fitting algorithms to share data across regional boundaries, no merging/ removal of duplicates required, no reduction of data volume by stages (maximises efficiency)*

*- allows **spatial pipelining** of data: essential for designing algorithms implementable in FPGAs & for optimising build times by reducing combinatorial logic requirements*

- future L1 trigger may also be time multiplexed, which could simplify input of tracking data (tracks) into the global calorimeter & muon triggers



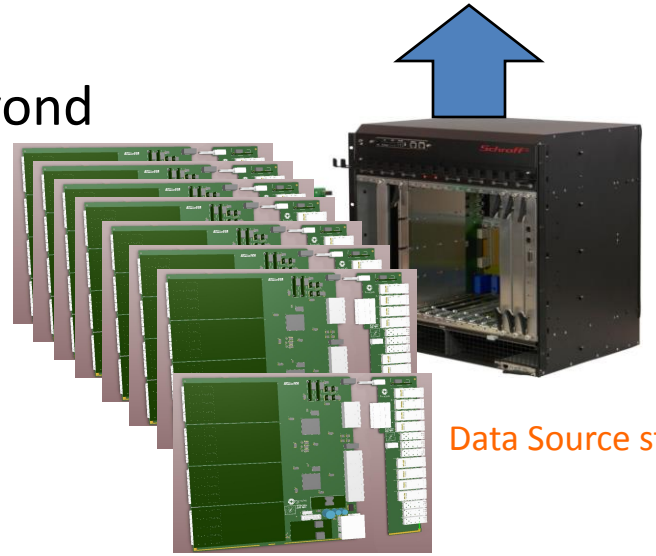
Vertical Slice System Demonstration over next few years

Core trigger tower

Can and will be Implemented in stages:
mezzanine, board, crate and multi crate level
(ATCA & uTCA)

With the goals:

- Performance study (latency, efficiency etc)
- Identify issues/bottlenecks
- Guide future R&D, find solutions
- *A common platform to explore new ideas/algorithm/approaches*
- An important step towards TDR and beyond
- *A major undertaking* !



Data Source stage

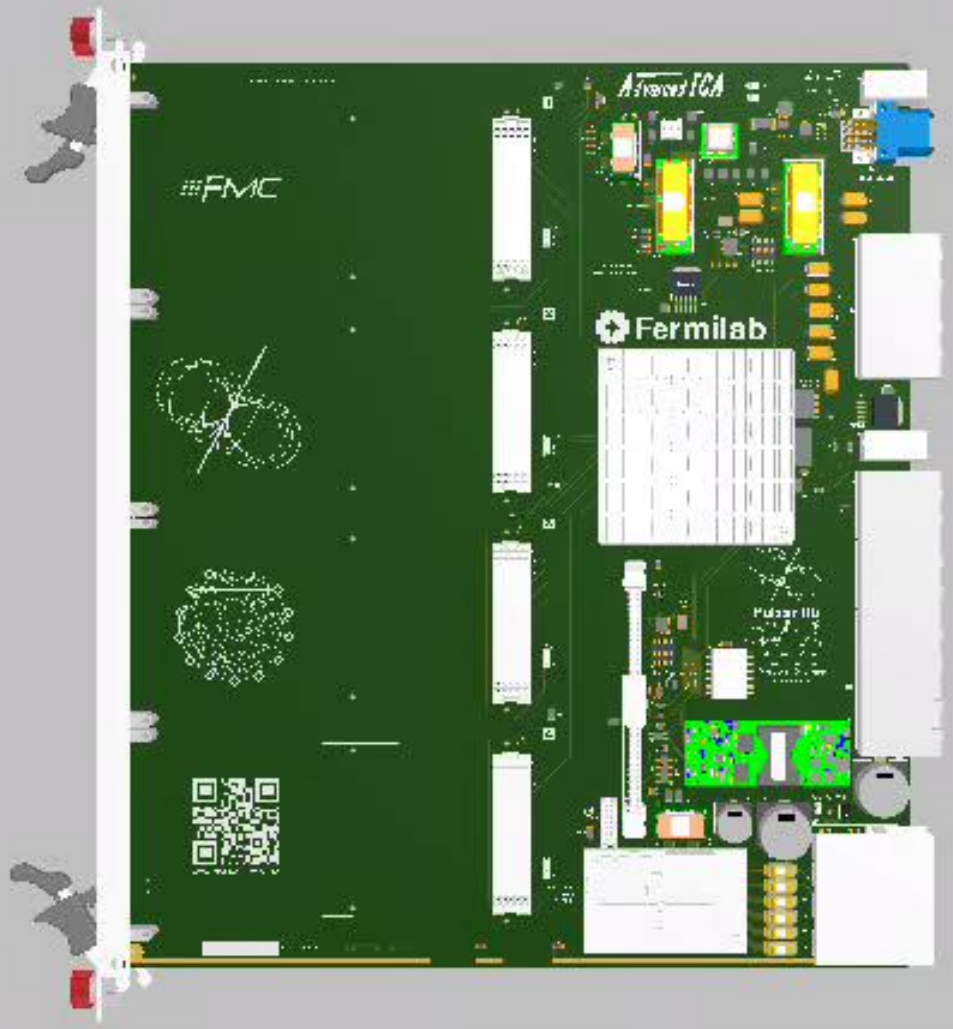
CMS people involved:
Lyon/INFN/Cornell/Northwestern/
Florida/Purdue/KIT/UK/CERN/FNAL …

With some Atlas & CMS common electronics …

Backup

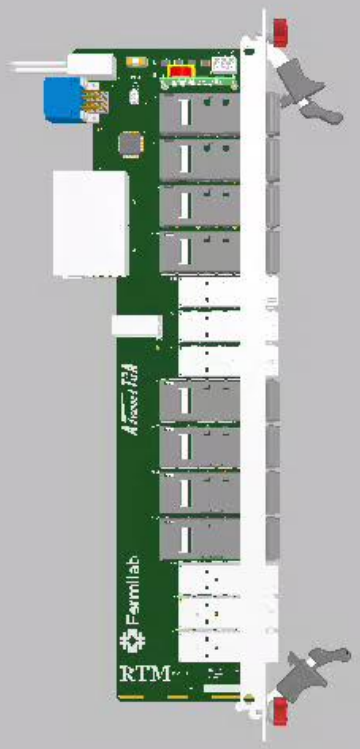
- Pulsar 2b related
- Some background materials on Fermilab VIPRAM project (in case asked).

http://www-ppd.fnal.gov/EEDOffice-w/Projects/ATCA/



Design philosophy: Less is more …

Rear Transition Module (RTM)

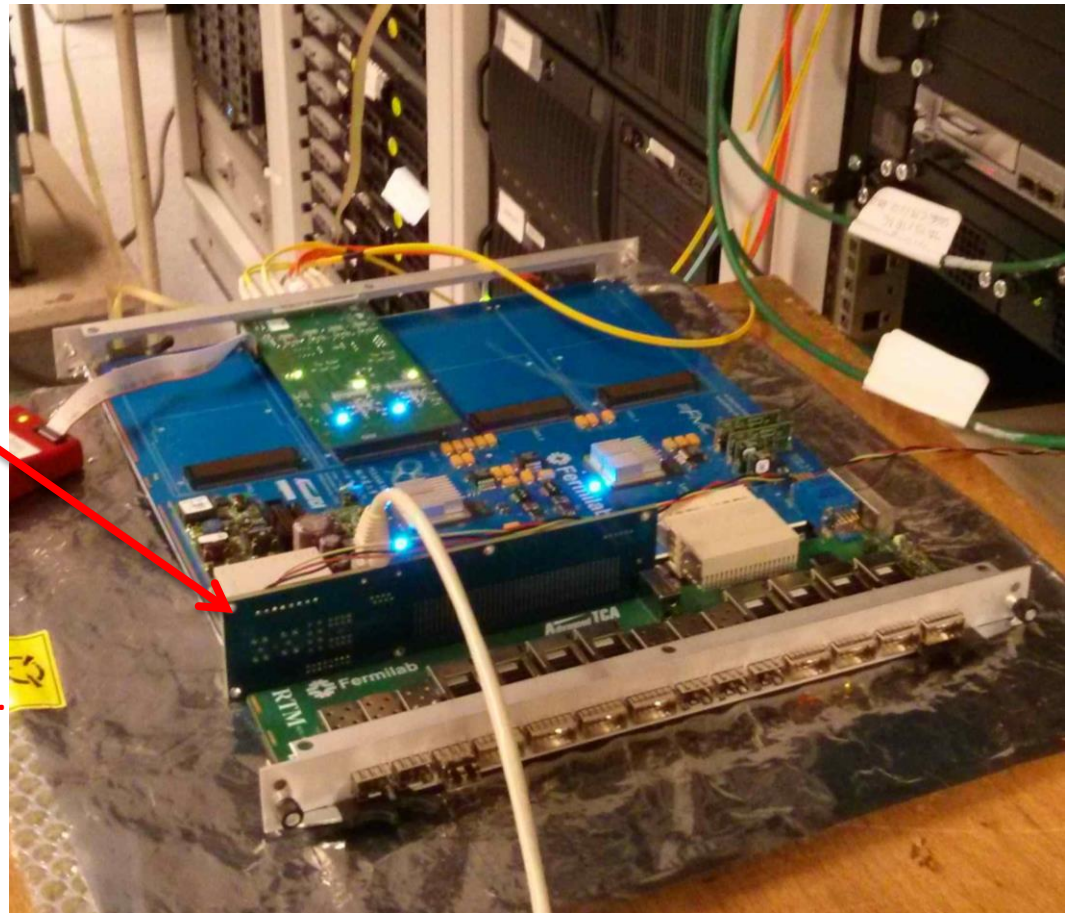


- Fiber optic transceivers
  - 8 QSFP+
  - 6 SFP+
- Up to 380 Gbps bi-directional
- Hot Swap
- PICMG 3.8 “Zone-3A” compliant

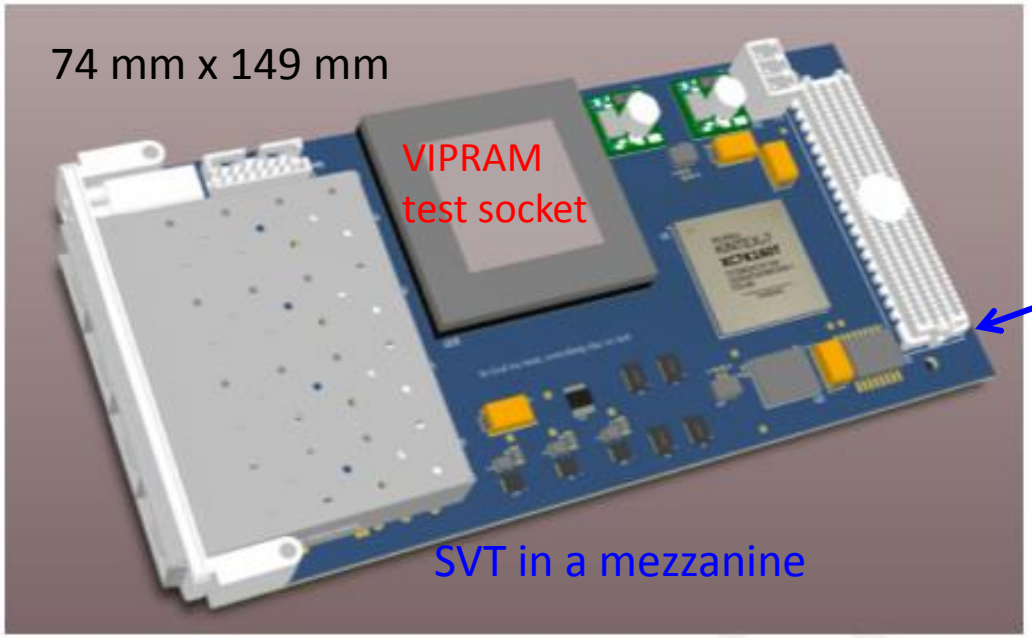
Pulsar II ATCA Mini power adaptor

The Mini Backplane can power an ATCA board on the bench:
A 48VDC power supply is required.
The Base Interface Ethernet port is brought out to an RJ45 connector.
The I2C IPMB bus signals are brought out to a terminal block.

full-mesh backplane connector with ALL connections in loop-back mode (Tx → Rx).



Used extensively for table-top testing/debug/firmware development

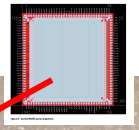


Test mezzanine card

design

actual card

protoVIPRAM chip



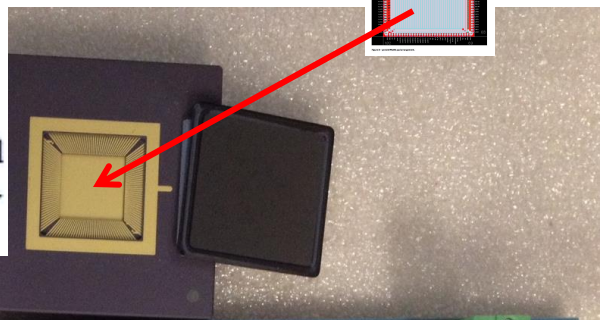
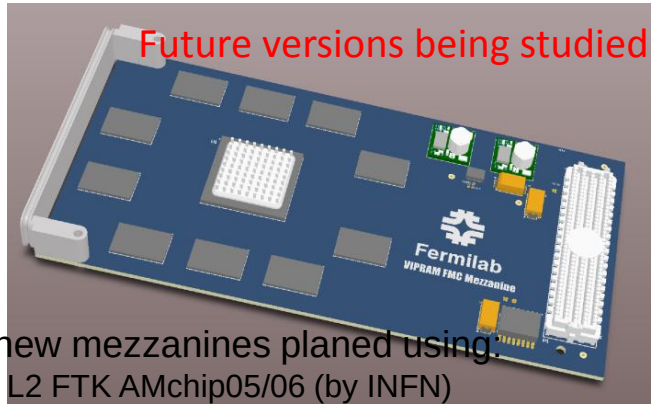


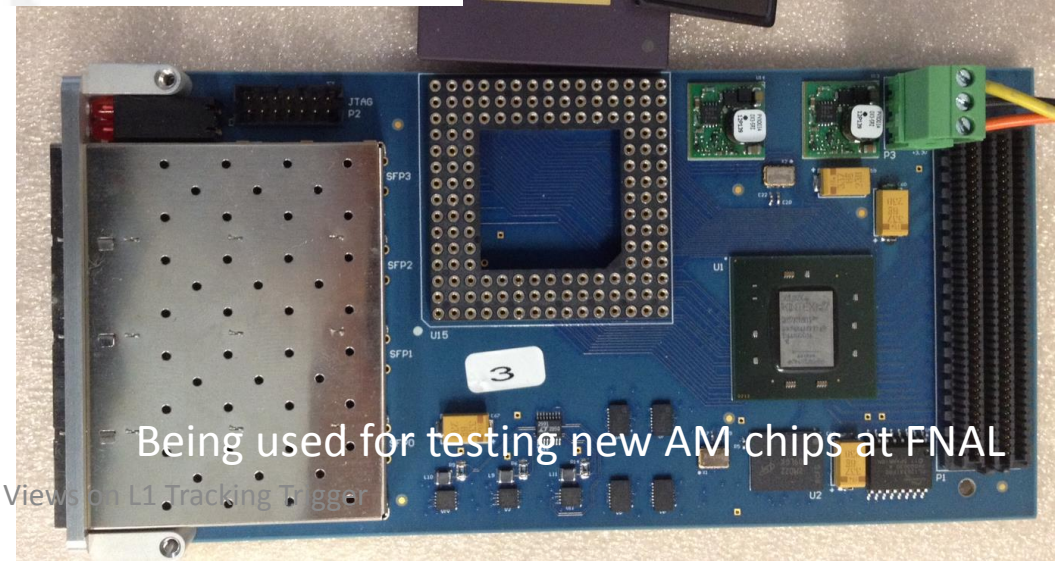
Figure 13: A test Mezzanine Card. This design features four SFP+ pluggable serial transceivers, a small Kintex FPGA, configuration flash memory, DDR3 memory, power supplies, local oscillators, a test socket and FMC connector.



Two new mezzanines planed using:

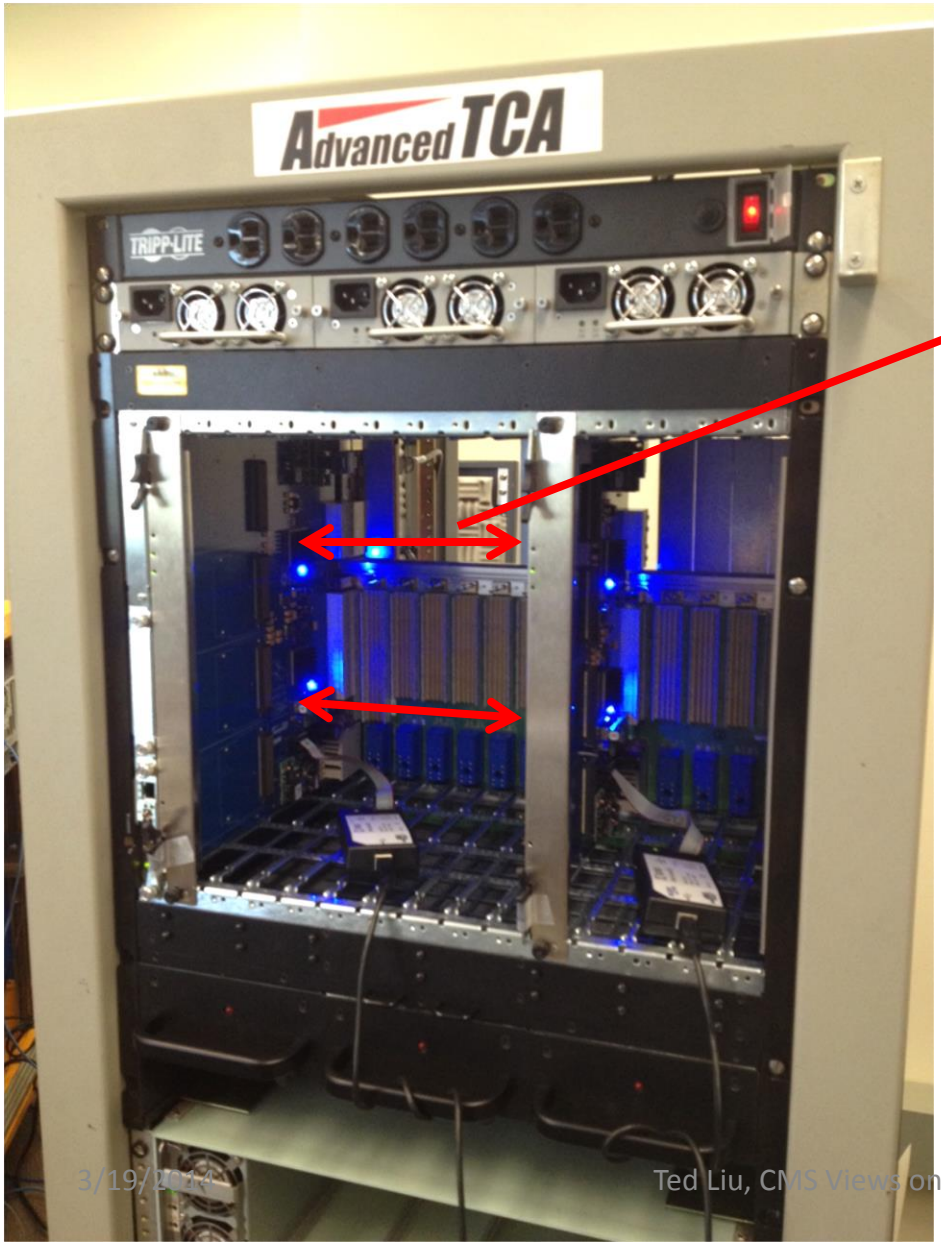
INFN L2 FTK AMchip05/06 (by INFN)

FNAL CMS AMchip dedicated for L1 (later)



Being used for testing new AM chips at FNAL

Test stand at FNAL



Achieved maximum speed:

Local bus: 10 Gb/s

RTM : 10 Gb/s

Fabric BP: 6.25Gb/s (limited by this backplane)

Upper limit of error rate (zero error):

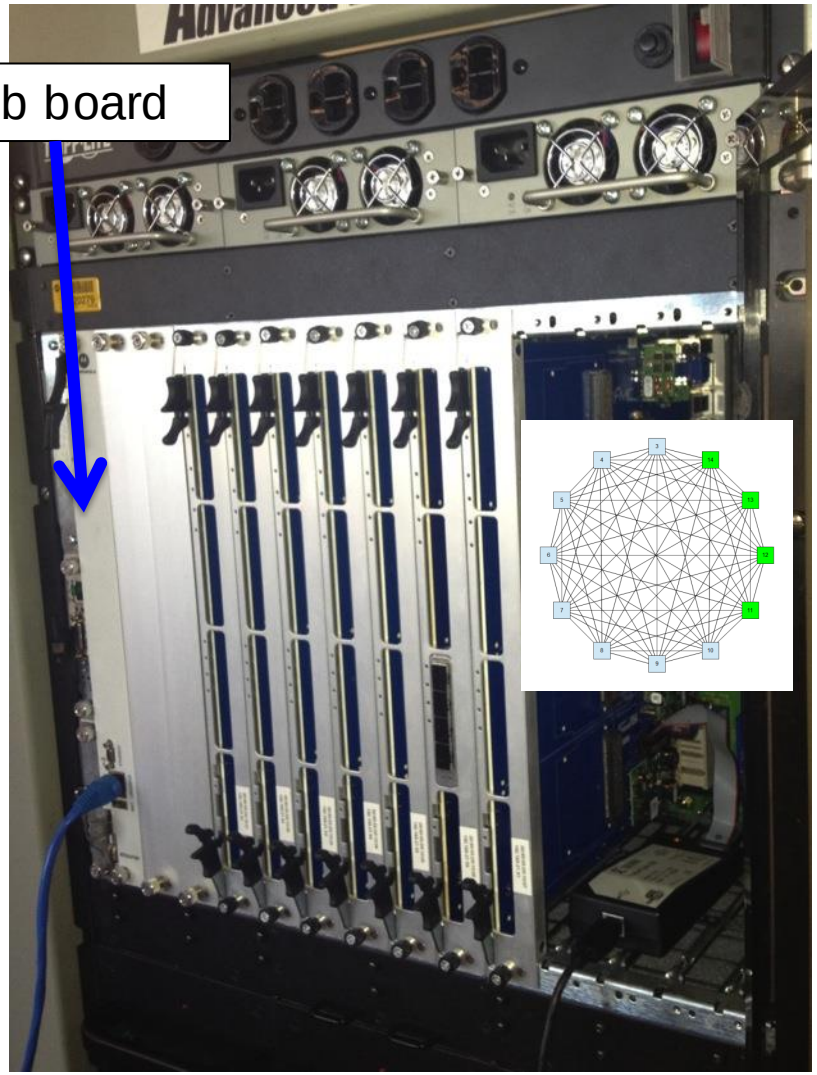
Local bus : 1.4E-15 (~ 10 hours)

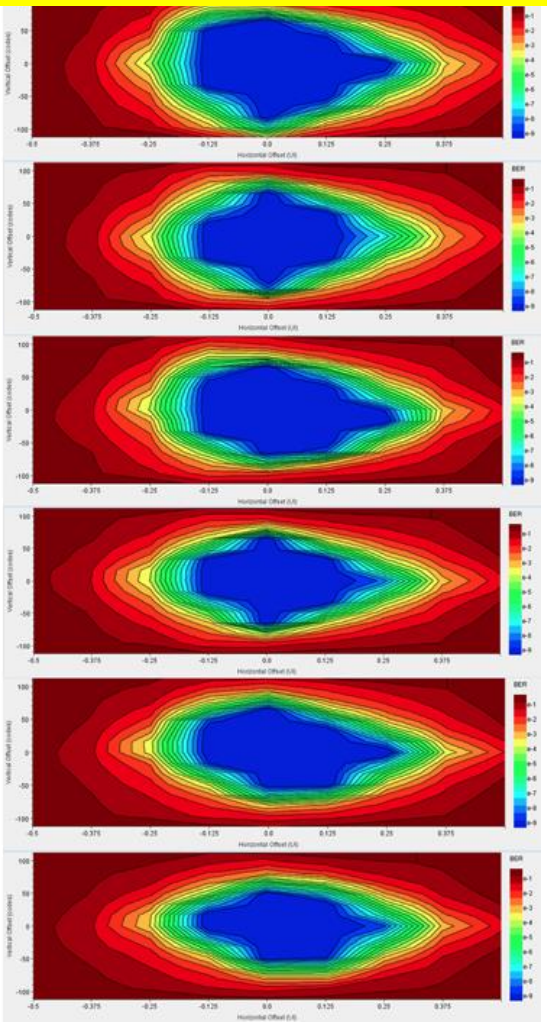
RTM : 1.9E-16

Fabric BP : 4.2E-17 (1 week)

Pulsar IIa Full crate testing at Fermilab

activate all the GTX at 6.25 Gb/s on 7 boards on full-mesh backplane



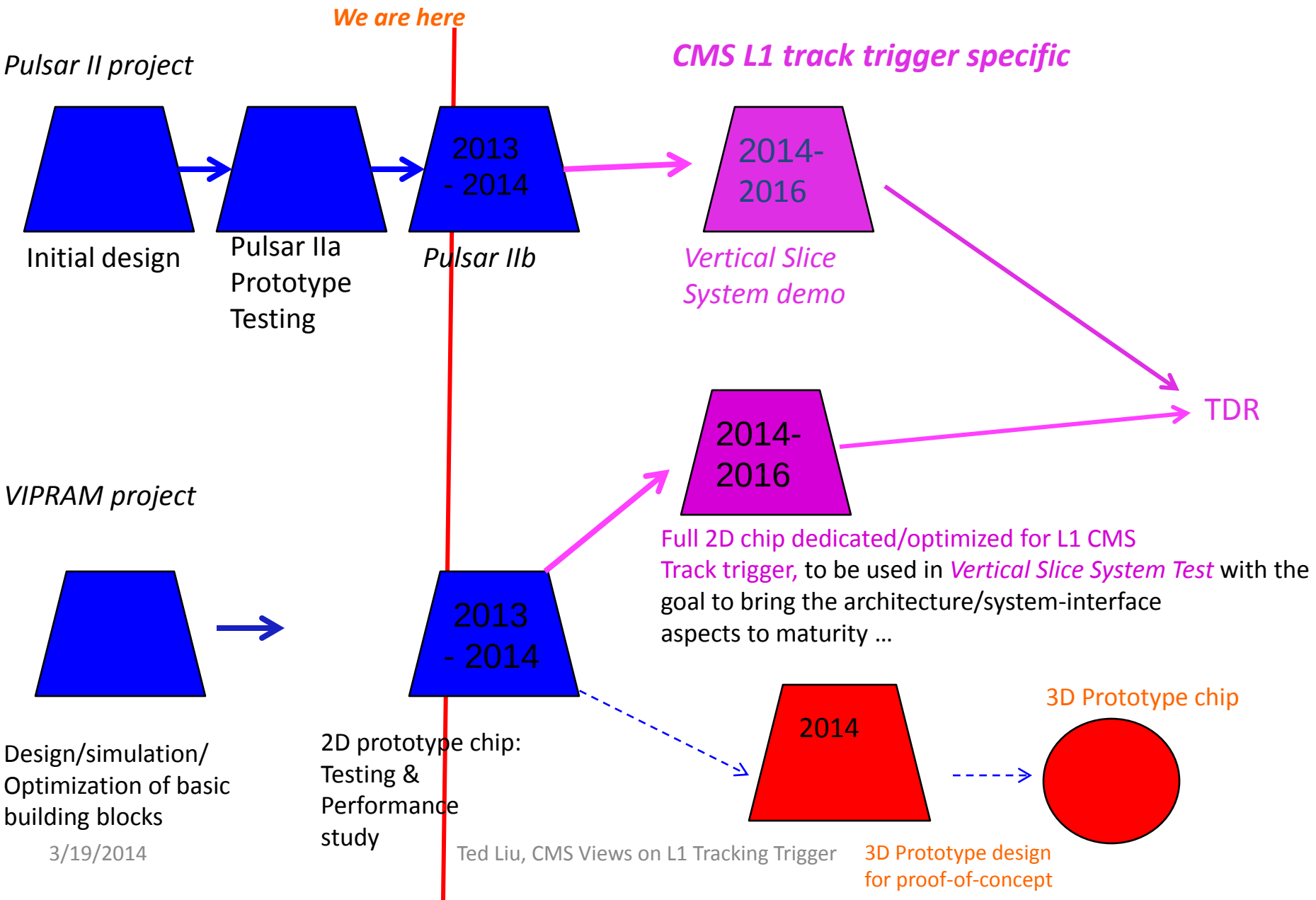


Achieved maximum speed:

Local bus: 10 Gb/s

RTM : 10 Gb/s

Fermilab Tracking Trigger R&D Roadmap:
→from generic R&D to system demonstration

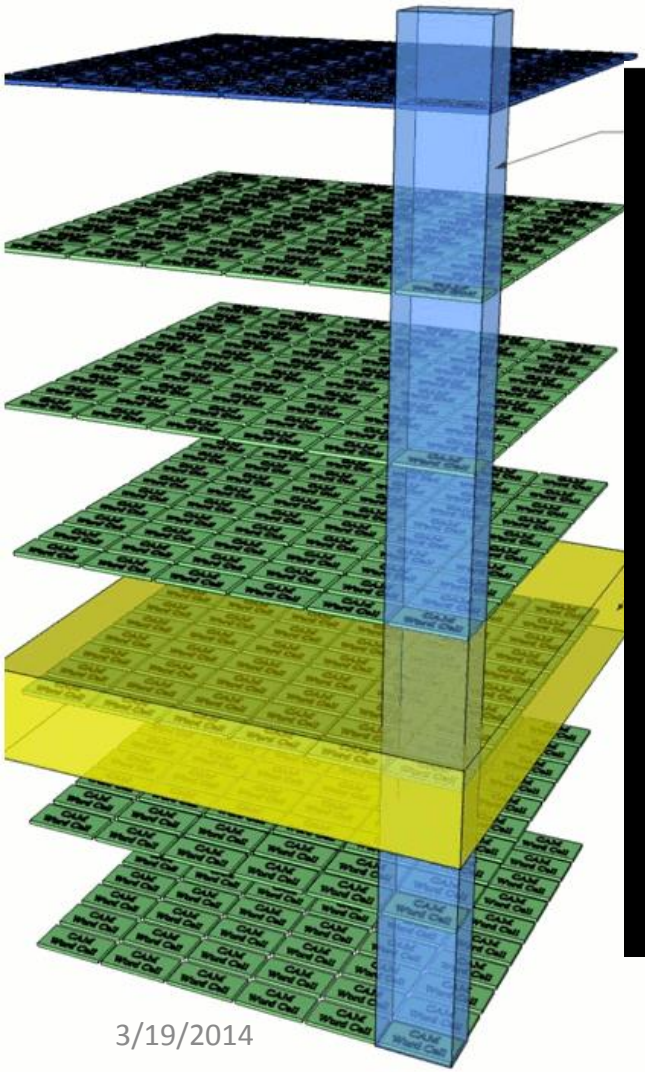


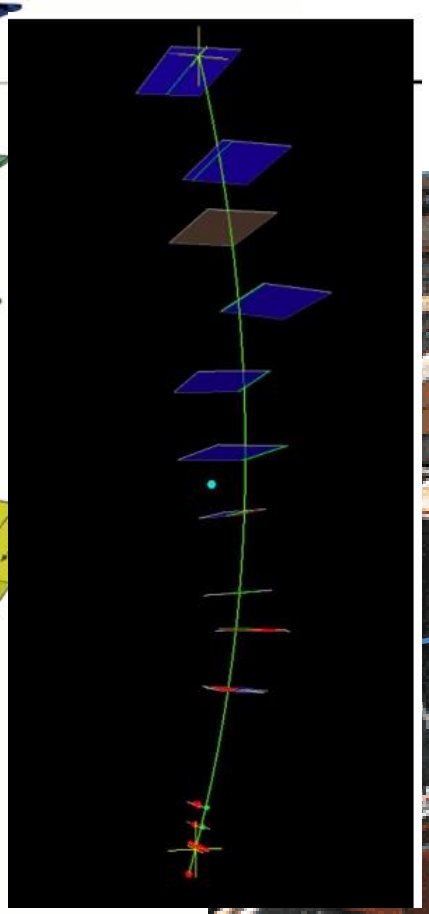
"A New Concept of Vertically Integrated Pattern Recognition Associative Memory"

TIPP 2011 Proceedings

http://www.sciencedirect.com/science/article/pii/S1875389212019165

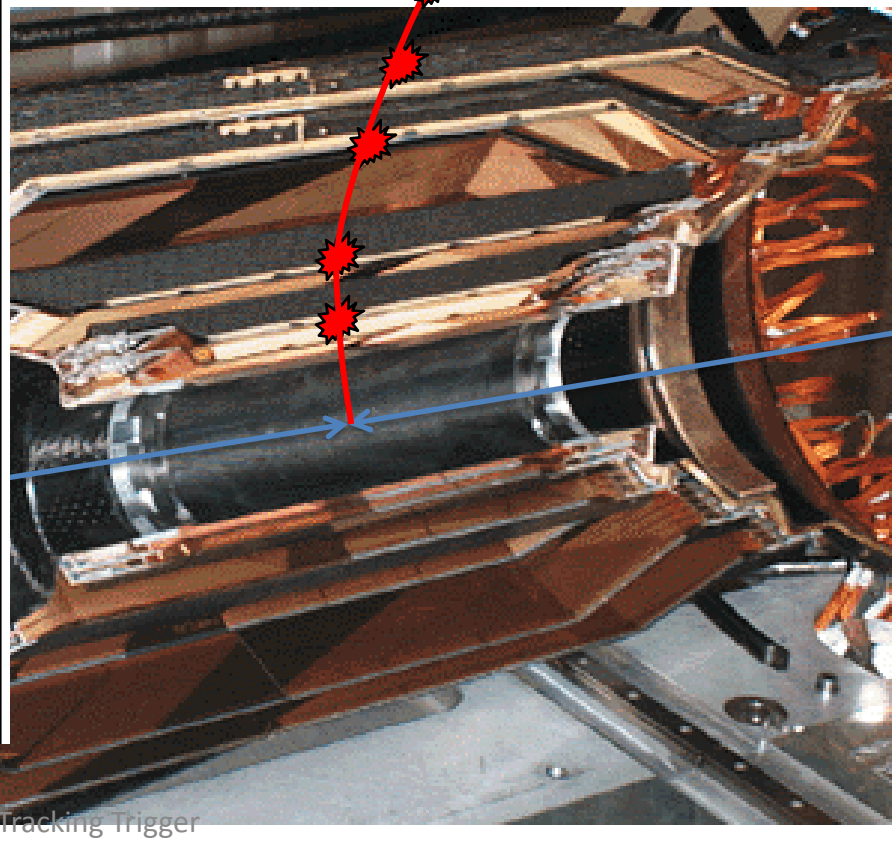
fired road





Pattern recognition for tracking is naturally a task in 3D

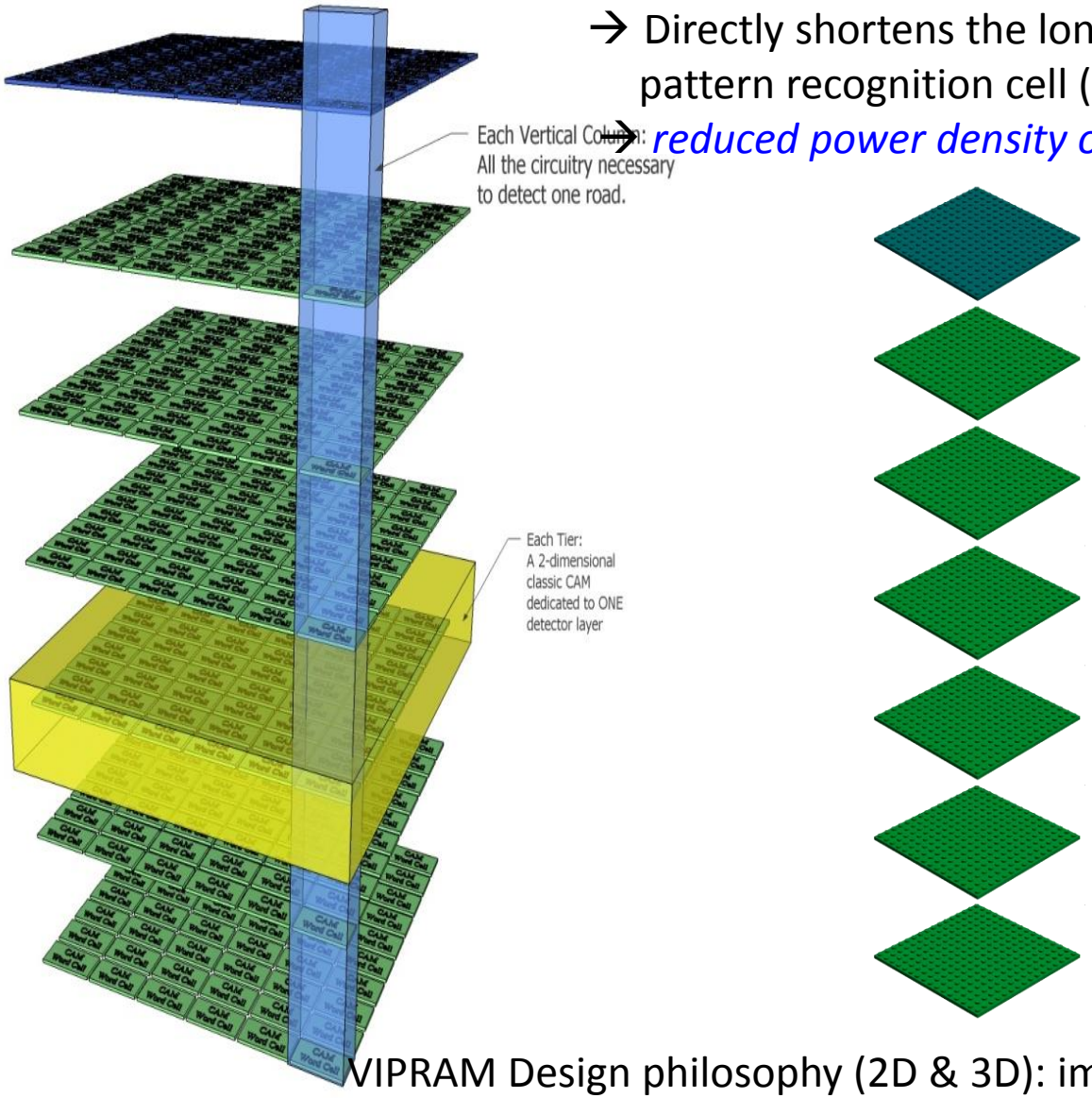
track



The 3D stack is very thin (< ~ 10 um per tier)

→ Directly shortens the longest of the driving lines in the pattern recognition cell (address match lines).

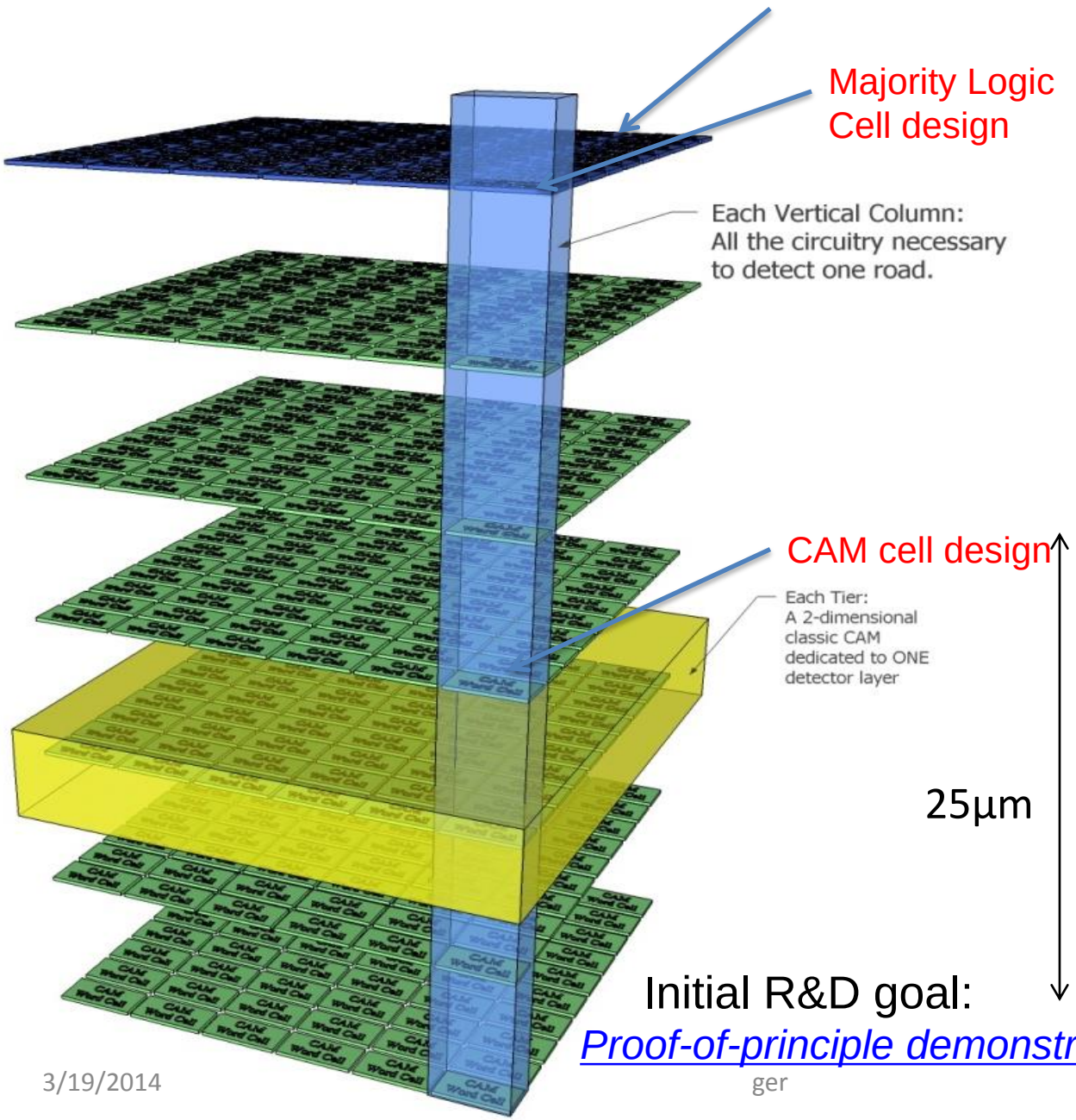
→ *reduced power density or increased speed*

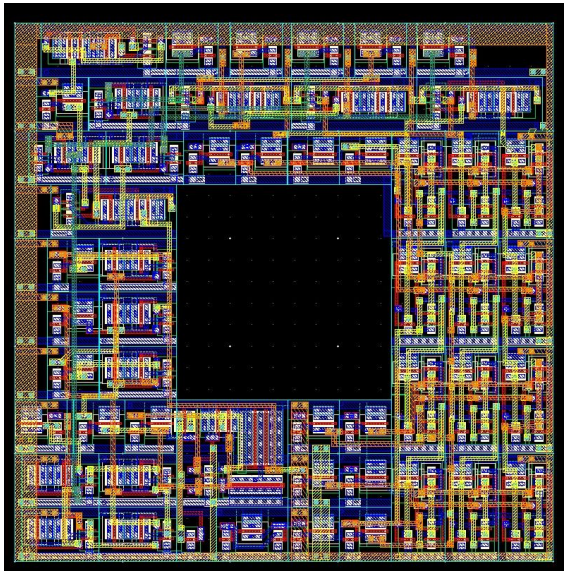


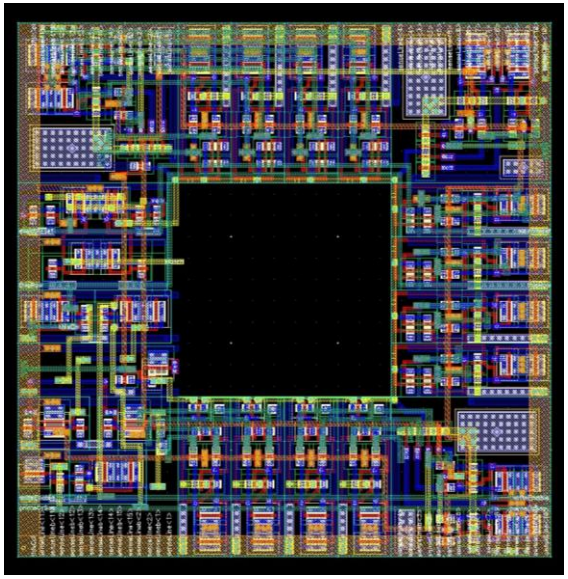
VIPRAM Design philosophy (2D & 3D): implement the core of AM engine, leave most of the control and interface logic to FPGA.

Design involved

In 130nm







Initial R&D goal:

Proof-of-principle demonstration

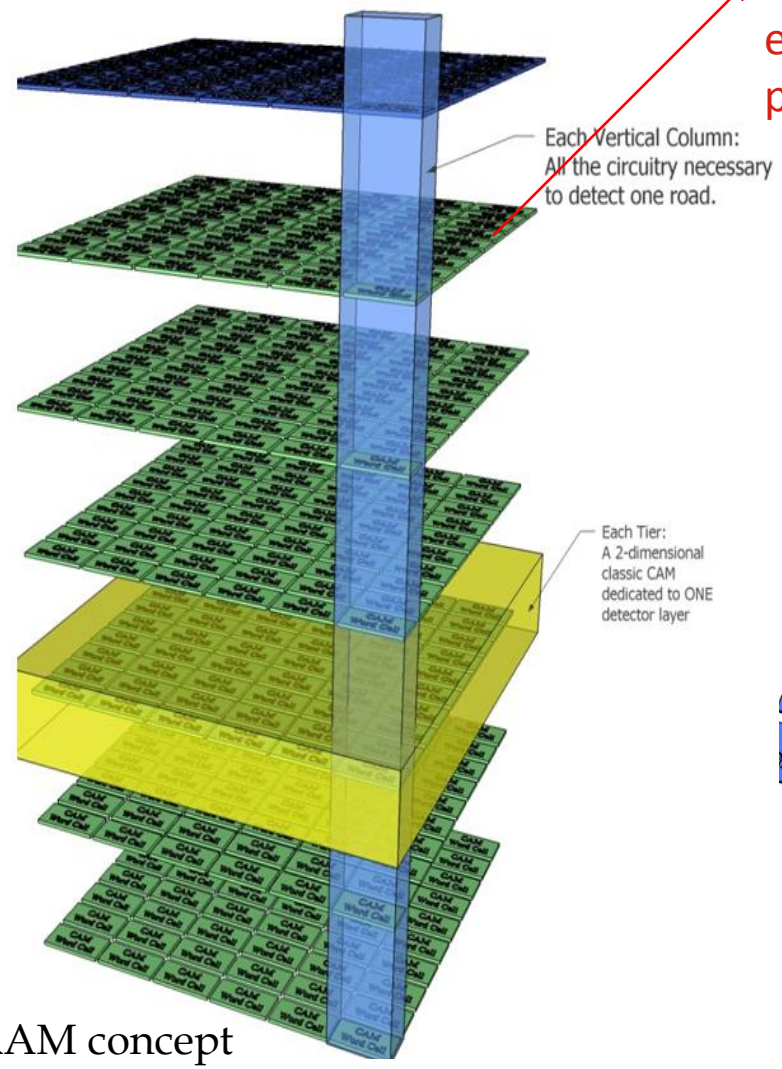
3D Architecture intrinsically open/flexible:

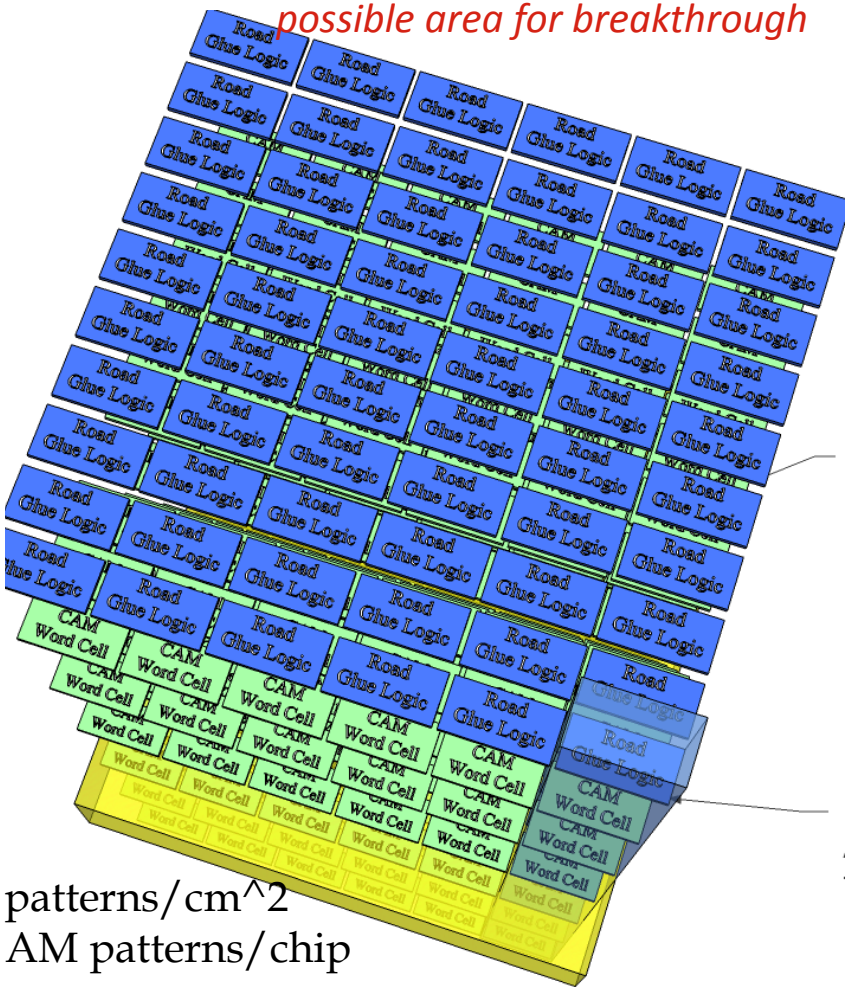
"CAM cell" doesn't have to be CAM

e.g. simple algorithm or even NN …

platform to implement more advanced/fast pattern recognition algorithm…

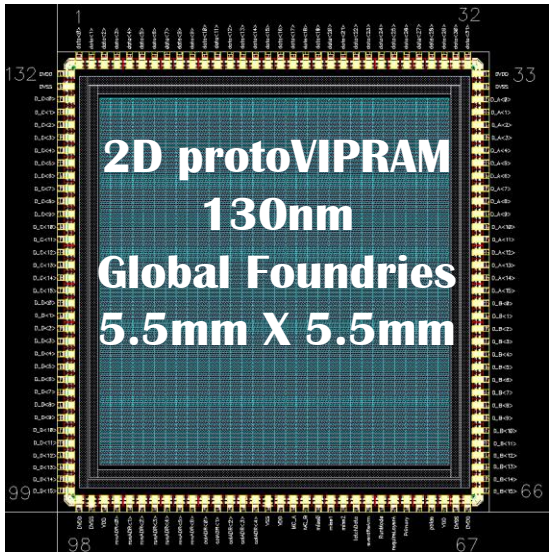
possible area for breakthrough

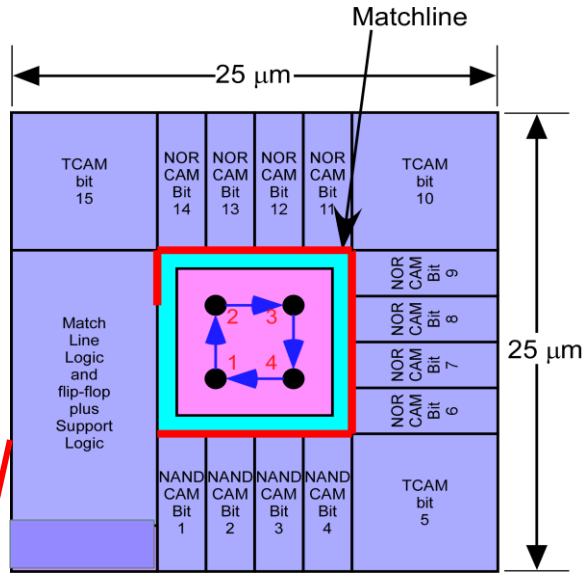




VIPRAM concept

- In 3D, with 130 nm, VIRPAM ~160K AM patterns/cm^2
- CDF AMchip03: 180 nm, standard cell, 5K AM patterns/chip

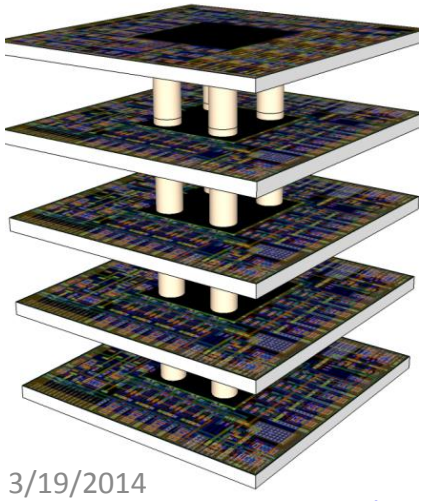




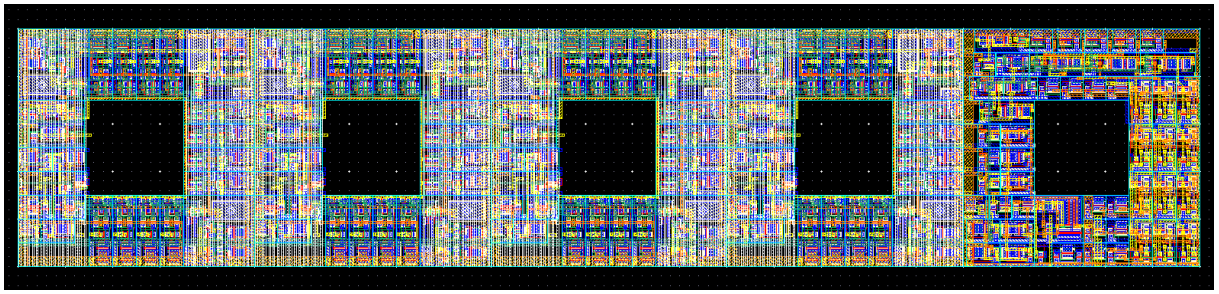
3D building blocks are implemented and tested in 2D:

important design & test methodology

Much of the design work gone into optimization in 2D

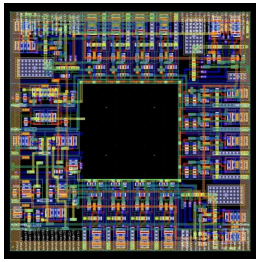


CAM CAM CAM CAM ML

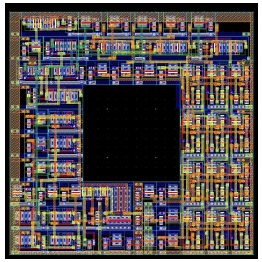


Every single transistor in the core pattern can be tested in 2D

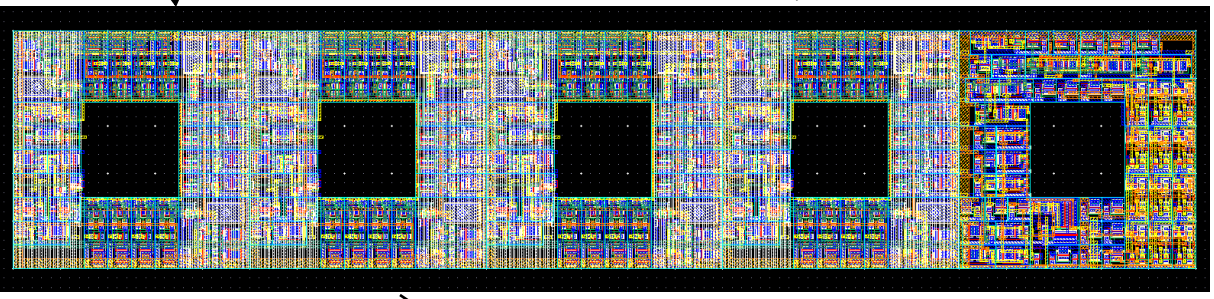
Different level of design simulation and optimization process



CAM cell

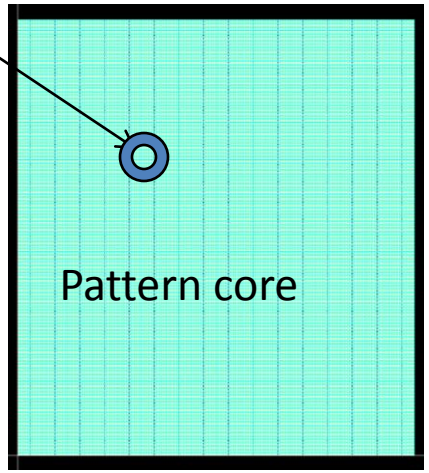


Majority Logic cell



Single Pattern

Power delivery
Signal timing …



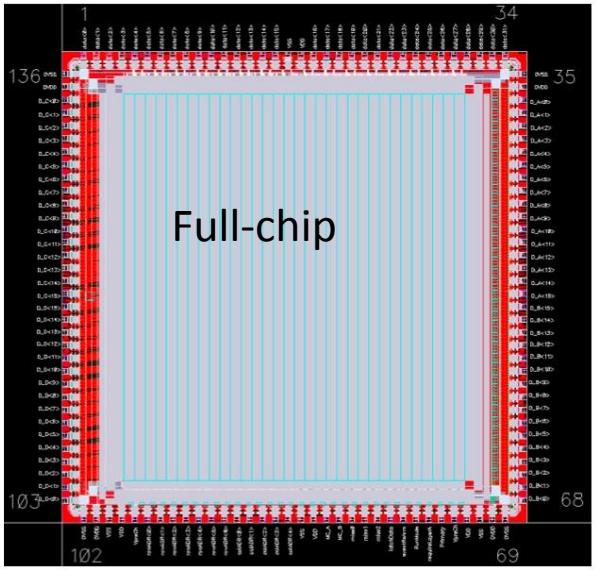


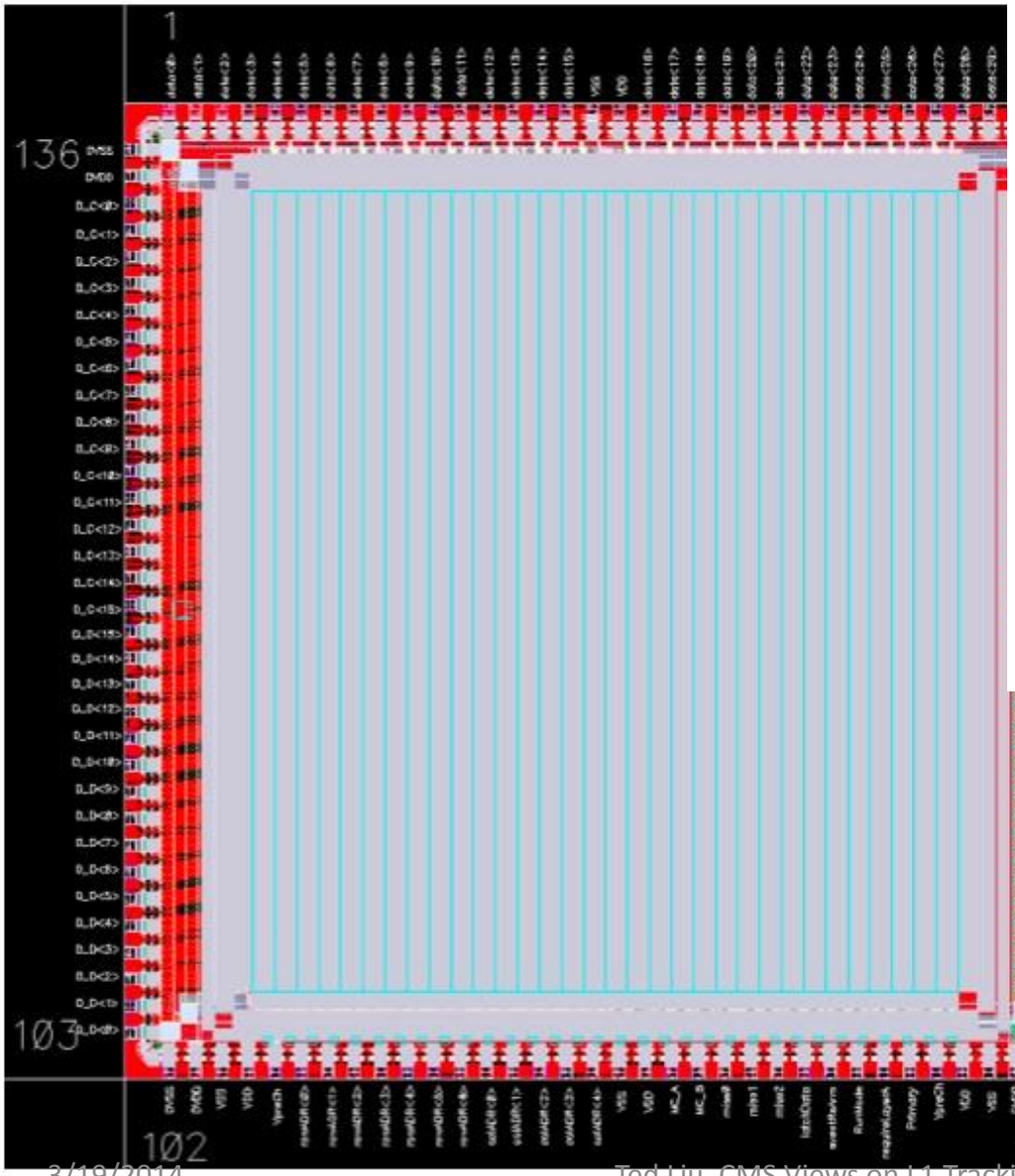
Figure 8 – protoVIPRAM pad arrangement.

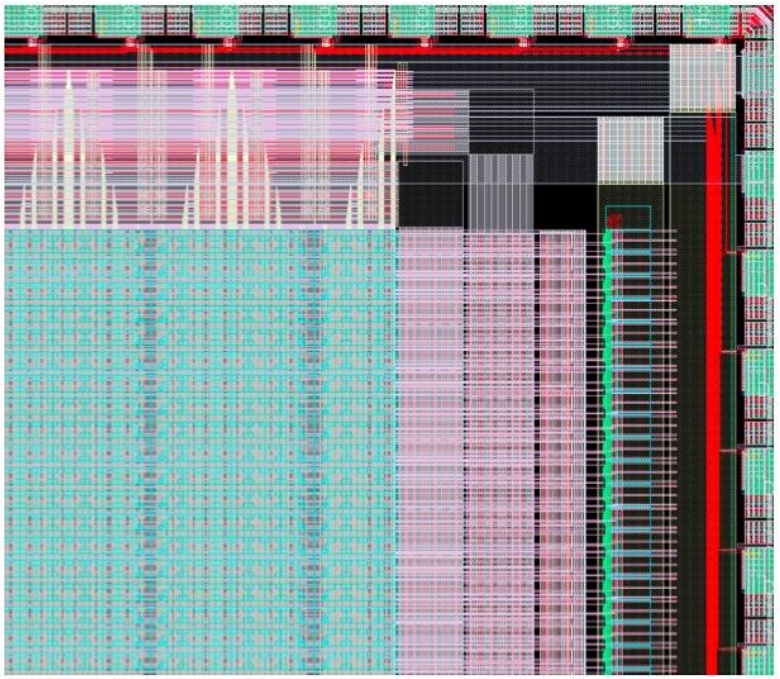
People involved so far from Fermilab:

Engineers: G. Deptuch, J. Hoff, S. Joshi, J. Olsen, M. Trimpl,

Physicists: S. Jindariani, T. Liu, N. Tran

Power/thermal analysis by W. Xia, P. Gui (SMU EE)





First protoVIPRAM in 2D

Chip delivered Jan 2014

Basic functional testing done

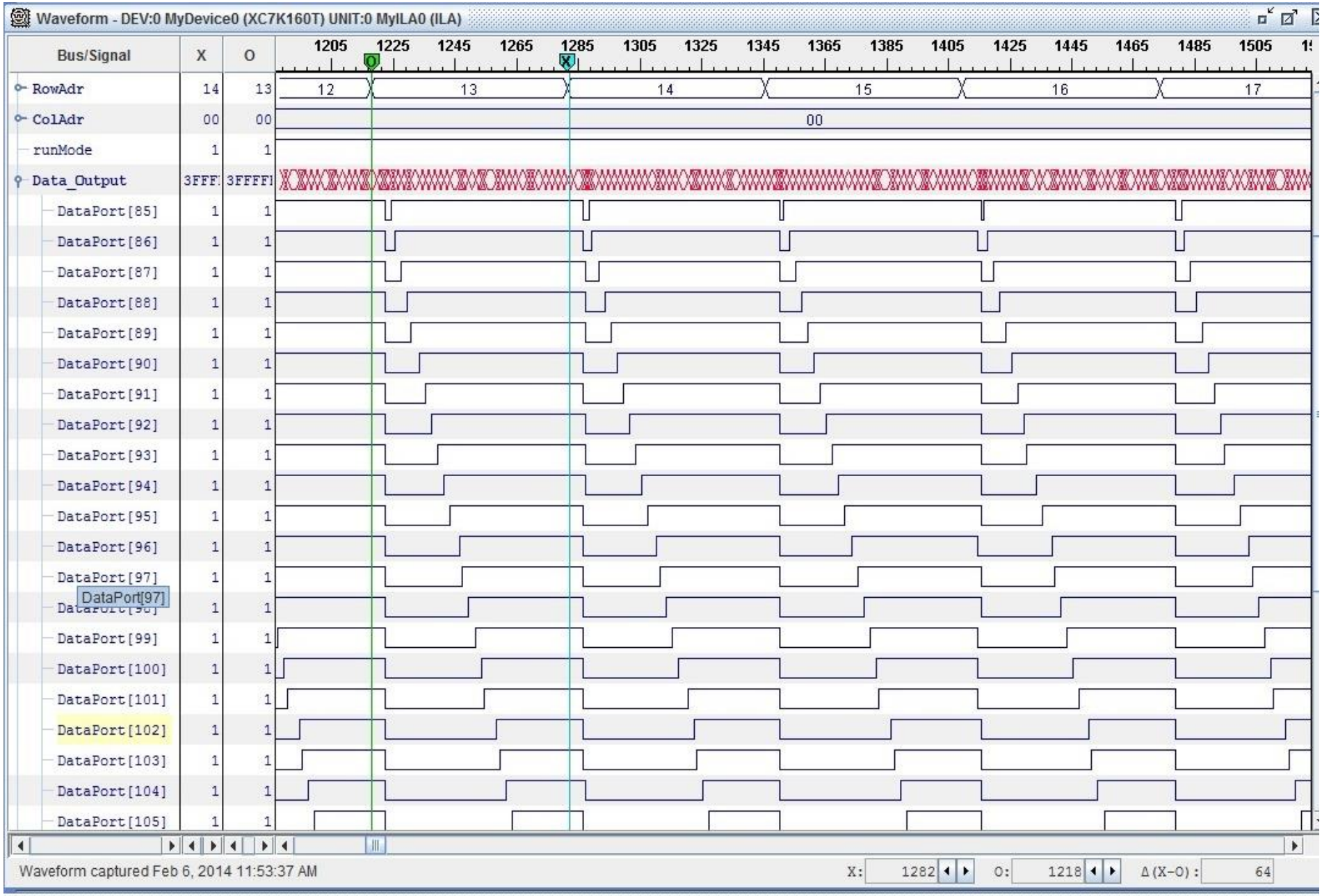
No problem found.

What's next:

Performance scan

3D stacking …

Figure 8 – protoVIPRAM pad arrangement.

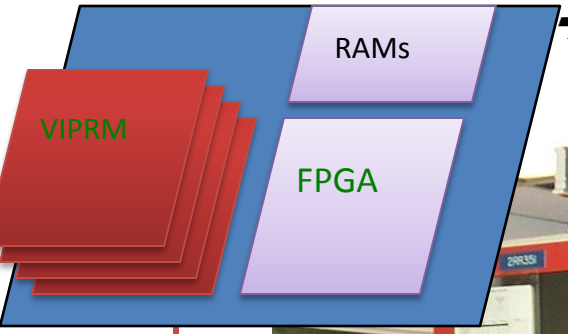


Recent Full Chip Test example (chipscope sampling results): pattern scanning.

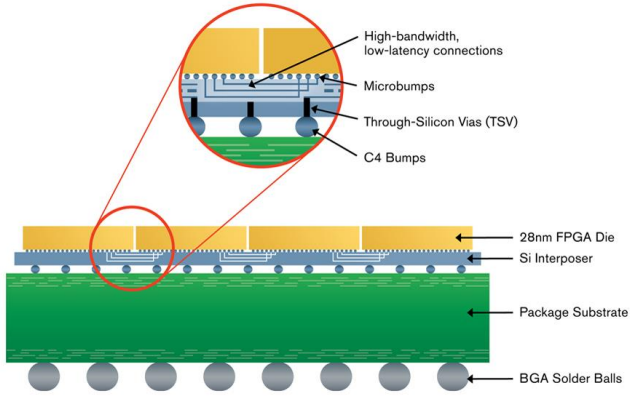
Plan to present full results at TWEPP 2014

SVT in one chip: 2nd phase of VIPRAM project

➔ *general purpose pattern recognition*



New 2.5/3D technology



2 meters

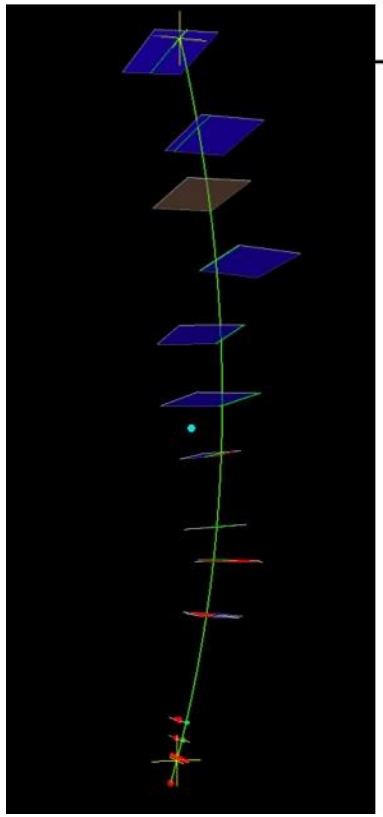
Original SVT system had ~400K patterns total

Aim to reach ~500K patterns per chip for VIPRAM (long term goal)…

High Performance Computing

→from US "Report to the President and Congress" by President's Council of Advisors on Science and Technology, Dec. 2010 (page 65)

- Compute-intensive
  - massively parallel computation involving *very large number of processing elements*;
- Communication-intensive
  - *high-speed transfer of data* among processing elements;
- Data-intensive
  - *high-speed manipulation of* very large quantities of data



HL-LHC L1 Tracking Trigger is High Performance Computing
(Non-von Neumann approach)
but with very Low Latency and in Real Time
HL-LHC requires the most advanced Real Time processing technology