

Tracking Trigger R&D for High Luminosity LHC

Ted Liu
Fermilab

Aug. 09, 2012
LBL Seminar

An old story from my graduate school days at CLEO

- Long ago, a young theoretical physicist had real trouble finding a girlfriend for a *long* time. Very frustrated, he complained to Hans Bethe
- Hans's advice (with his strong German accent):

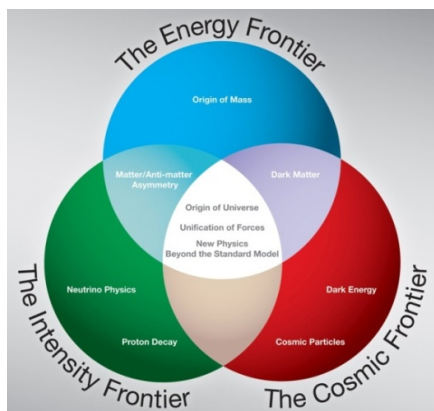
*Young man, if the cross section is soooo low,
increase the luminosity !*

→ good advice for students who were interested in charm and beauty

$$\text{Rate} = \sigma L$$

σ - cross-section

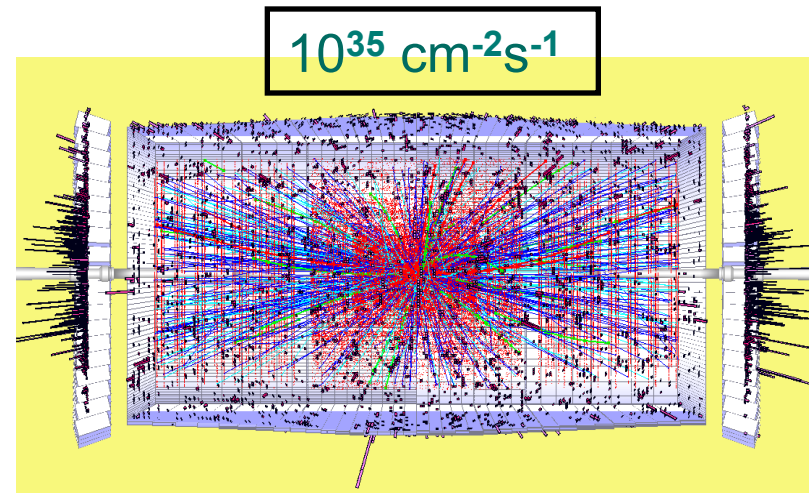
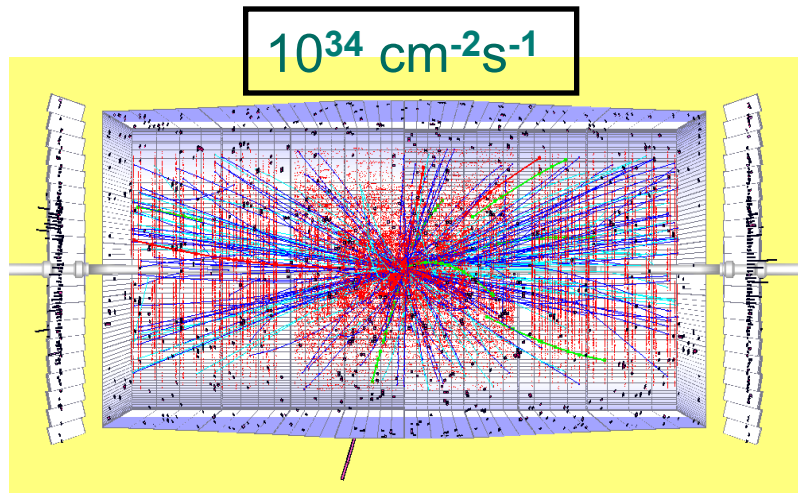
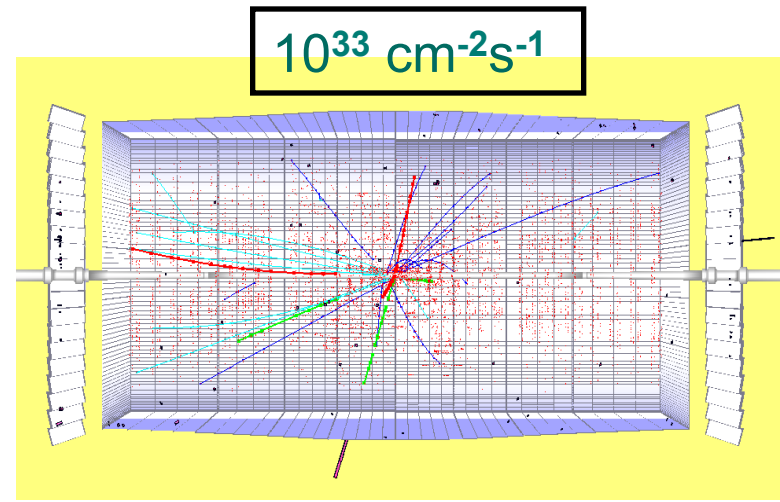
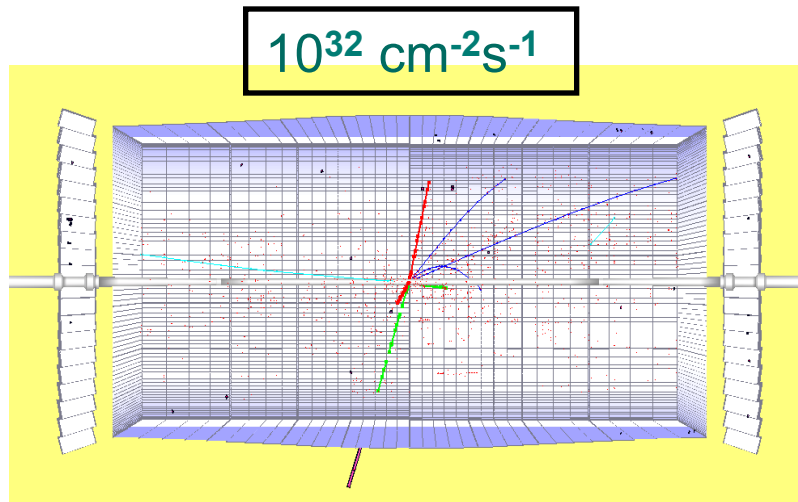
probability that an interaction will occur



*For the experimentalists, we have a similar problem,
on all frontiers ... increasing the luminosity is one
way to go... esp. when cannot increase cross section*

For hadron collider: increasing luminosity
could get one into deep trouble ...

- $H \rightarrow ZZ \rightarrow \mu\mu ee$, $M_H = 300$ GeV for different luminosities in CMS



Paper in 1982: concerns with High Luminosity

A PRIMER ON DETECTORS IN HIGH LUMINOSITY ENVIRONMENT

R. Huson, L. M. Lederman and R. Schwitters
Fermi National Accelerator Laboratory*
Batavia, Illinois 60510

I. History

The following remarks are relevant to the problem of balancing luminosity versus energy in new HEP construction.

In a 1973 Isabelle Summer study,¹ it was stated that the only experiment that would succeed at a luminosity of $10^{33} \text{cm}^{-2} \text{sec}^{-1}$ was one in which the apparatus was shielded from the collision region by massive quantity of steel. In 1981, this opinion was confirmed by an authority no less than S.C.C. Ting.² It may be instructive to review the progress of collider detectors over the past decade. In 1973, the time resolution or, better, the integrating time of tracking detectors was $\sim 100 \text{ ns}$. In 1982, this time has remained the same since PWC's are still the fastest tracking devices available. The fundamental limit is the saturated drift velocity of electrons in gases. Better resolution and three dimensional properties have led to the choice of drift chambers and TPC's which have considerably longer integration times. A new characteristic of 1982 detectors is the increasing pervasiveness of calorimeters which have become indispensable devices for measurement of electromagnetic and hadronic energy, especially at momenta where magnetic measurements become imprecise. Calorimeters, because of their innate geometric dimensions set by the nuclear mean free path and their distance from the interaction point have integration times of $\sim 200\text{--}1000 \text{ ns}$. Of course this is the present state of the art which depends on the properties of BBQ, gas chambers, liquid argon, lead glass, etc.

The conclusion is that things have only gotten worse since 1973.

II. Integration Time - Tracking

What are the implications of long integration times? We are facing collision energies so high that the charged and neutral multiplicities N_{ch} and N_{ch}

tracking efficiency; there is in fact a fair likelihood that these high multiplicities will render any of the tracking devices, as we now understand them, inoperable. PWC's have operated at ambient singles rates of 10 Mcps with fairly simple track configurations. However, experience with 20-30 tracks, e.g., at the ISR's Split Field Magnet or at various multiparticle spectrometers suggest a CDC 7600 CPU analysis time per event of hundreds of milliseconds up to $\sim 5 \text{ sec!}$ To contemplate the functioning of a track chamber with several hundreds of tracks, many of low and "curling" energies (even given scintillation tagging) clearly requires a major advance. As a dramatic example, look at Fig. 1 and imagine superposing 2, 3 or 5 such events in a single trigger.

We should note that before one can reject tracks for pointing incorrectly one must be able to do the pattern recognition. A more quantitative tabulation of the influence of finite integrating time is presented in Tables I and II.

III. Calorimetry

To this tale of woe we must add the problem of the calorimeters. Now we have ~ 30 charged and 30 neutral particles incident upon the calorimeter which has an optimistic integrating time of $\pm 200 \text{ ns}$. This is at $\sim 1 \text{ TeV}$. Multiplicities will about double at 10 TeV . It is true that a typical event may add negligibly to a (say) 100 GeV/c transverse momentum trigger. Some fraction of good events would be confused by the integration, but it is also clear that a large enough number of random accumulations of 10 or 20 minimum bias events can generate fake physics. These may provide a background for a large fraction of the anticipated physics signatures. During the interval between real 100 GeV/c jets say (at the rate of 10 per day) there would be $\sim 5 \times 10^{11}$ accumulations of twenty random events! If each charged particle

ab

LBL-15973
April 1983

PROCEEDINGS
OF THE
1983 DPF WORKSHOP
ON
**Collider Detectors:
Present Capabilities
And
Future Possibilities**

February 28 - March 4, 1983

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

Edited By
Stewart C. Loken and Peter Nemethy

or the U.S. Department of Energy under Contract DE-AC03-76SF00098 and for the
Physics Section of the National Science Foundation.

FOREWORD

The "Workshop on Collider Detectors: Present Capabilities and Future Possibilities" was sponsored by the Division of Particles and Fields of the APS and hosted by Lawrence Berkeley Laboratory. It was held at LBL from February 28th to March 4th, 1983.

The organizing committee consisted of A.K. Mann (Chairman), C. Baltay, R. Diebold, H. Gordon, D. Hartill, P. Nemethy, D. Ritson and R. Schwitters. The local organizing committee was R. Cahn, S. Loken and P. Nemethy.

The workshop focused on the problems posed by high luminosities at hadron colliders, considering luminosities on a continuous range from 10^{29} to $10^{34} \text{ cm}^{-2} \text{ sec}^{-1}$, picking two specific center-of-mass energies, 1 TeV and 20 TeV. The participants divided into the five working groups tabulated below.

These proceedings contain three sections. Section I consists of input to the workshop, the introductory comments of the organizing committee chairman (A.K. Mann); two out of our three invited talks (W.J. Willis, M. Banner, C. Rubbia) on collider experience; finally two documents, which were invaluable in getting the workshop started, theoretical estimates of relevant cross sections (R. Cahn) and of high P_{\perp} jet behavior (F. Paige).

Section II contains the working group summary reports from the five working groups; this is the meat of the workshop. Section III is a rich mix of contributed papers relevant to the workshop.

I want to thank Jeanne Miller, our workshop secretary, and Peggy Little, the LBL Conference Coordinator, for all their help; Donna Vercelli, Judy Davenport and Loretta Lizama for their work on the proceedings. I also thank our working group leaders and scientific secretaries for their dedication and all the participants for a lively and spirited workshop. Support for this workshop was provided by the Department of Energy and the National Science Foundation.

Peter Nemethy
Workshop Organizer

WORKING GROUP	GROUP LEADER	SCIENTIFIC SECRETARY
Tracking Detectors	Don Hartill	David Herrup
Calorimetry	Bernie Pope	Melissa Franklin
Triggers	Mel Shochet	Mike Ronan
Particle Identification	Dave Nygren	Rem Van Tyen
Detector Systems	Barry Barish	Mark Nelson

Some colliders in the past

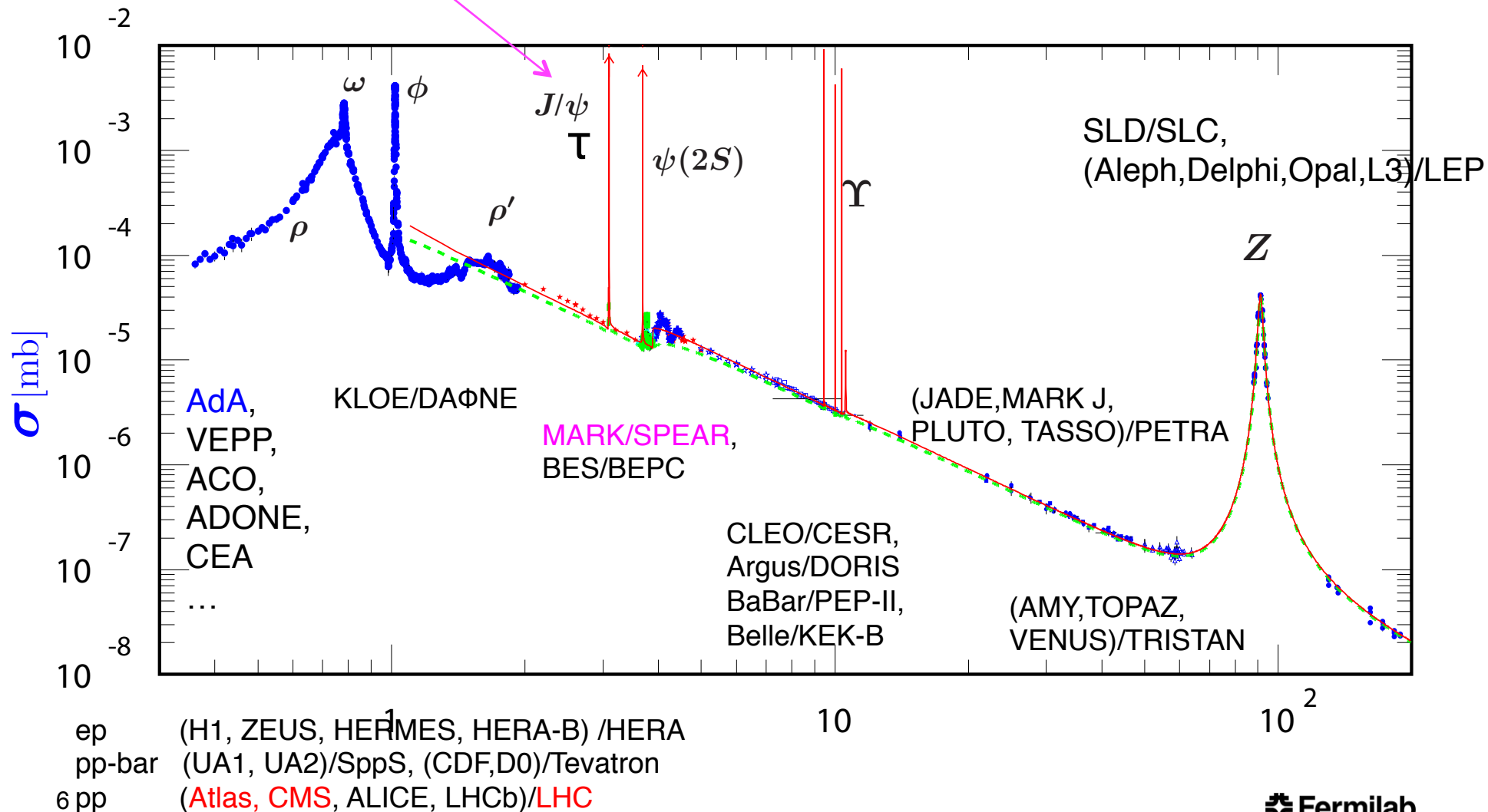
Trigger challenges at high lumi:

e^+e^- , ep : beam background

pp -bar, pp : pile up & pile up

“unauthorized particles on an unauthorized machine”

σ and R in e^+e^- Collisions



Tracking Trigger R&D - Ted Liu

B Factory case: $\text{Rate} = \sigma L$

- Some basics:

- ↘ *Luminosity*: units of $\text{cm}^{-2}\text{s}^{-1}$ or $\text{barn}^{-1} \text{sec}^{-1}$

- Often in $\sim 10^{34} \text{cm}^{-2}\text{s}^{-1} = 10 \text{ nb}^{-1}\text{s}^{-1} = 10^{-5} \text{fb}^{-1}\text{s}^{-1}$

- This was roughly B Factory luminosity

- A few physics processes at B Factory

- ↘ $\sigma(e^+e^- \rightarrow e^+e^-)$ at BF $\sim 72 \text{ nb} \rightarrow 720 \text{ Hz}$

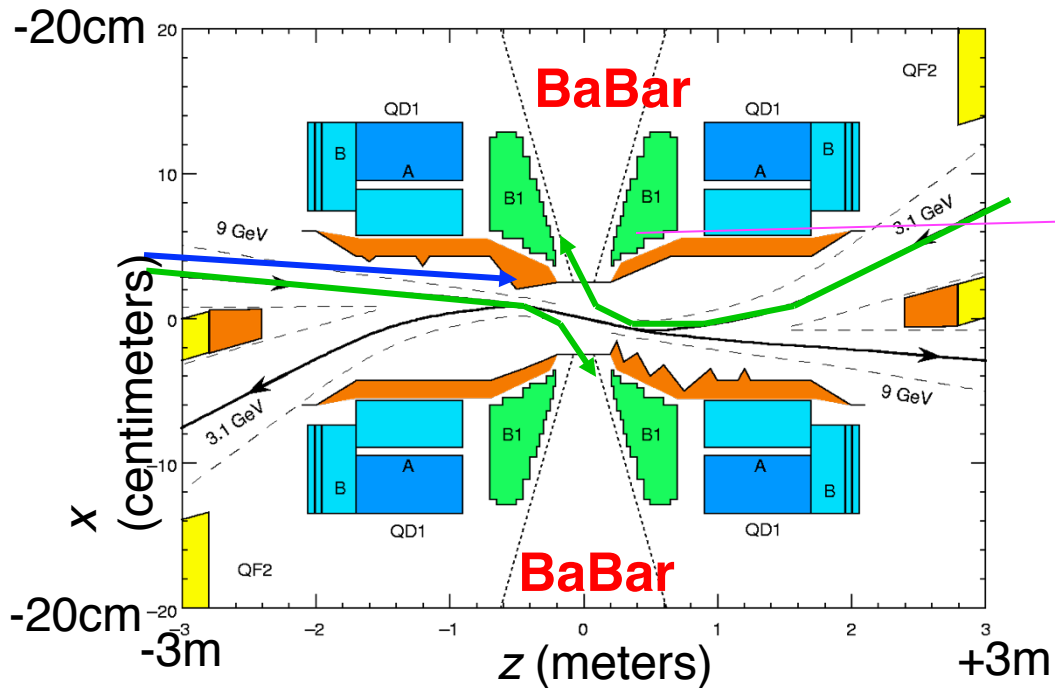
- ↘ $\sigma(e^+e^- \rightarrow \text{hadrons})$ at BF $\sim 4 \text{ nb} \rightarrow 40 \text{ Hz}$

- ↘ $\sigma(e^+e^- \rightarrow \text{total})$ at BF $\sim 84 \text{ nb} \rightarrow 840 \text{ Hz}$

- ↘ Final events to tape $\sim 100 \text{ Hz} \rightarrow$ need to pre-scale

One can get a sense for B Factory trigger,
though there were interesting challenges...

Interaction Region and High Background



beams passing through bending magnets
generate a “fan” of X-rays in bending plane

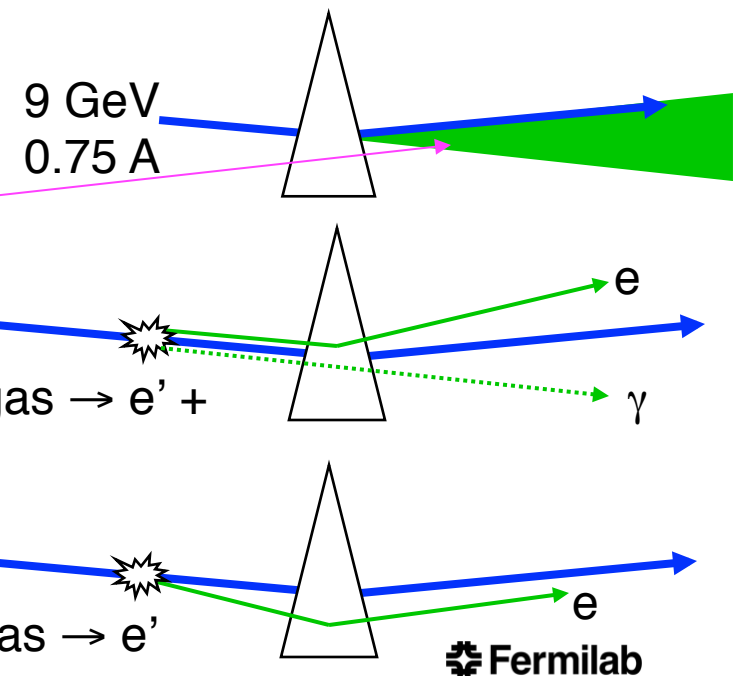
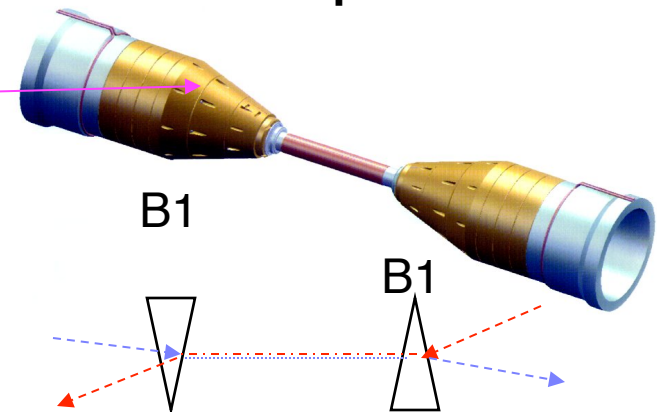
Bremsstrahlung & Coulomb interactions
produce energetic e, γ along beam line

High Luminosity \Rightarrow **High Background**

Beam Crossing @ **238 MHz**

Head-on collision

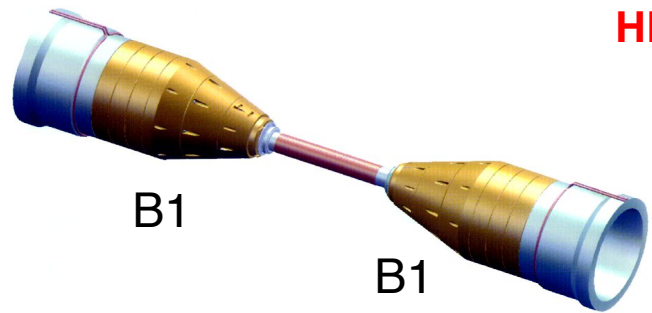
Need to separate the beam



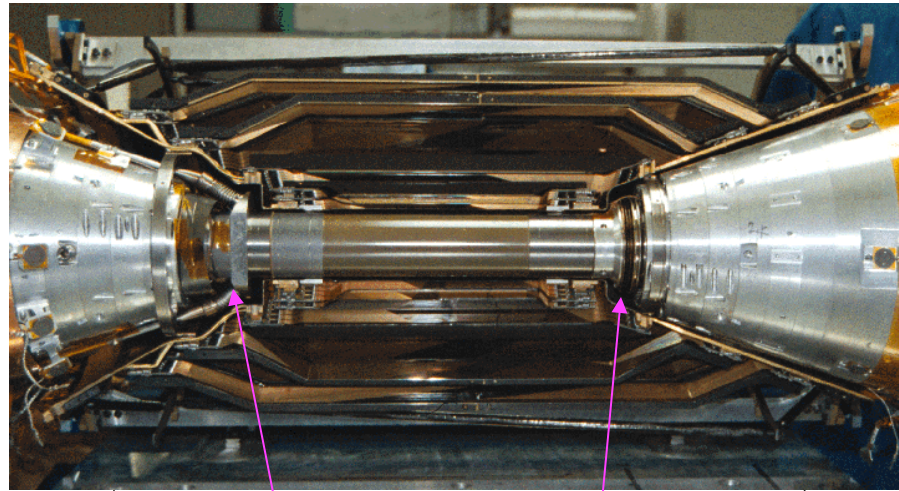
Fermilab

PEP II B factory

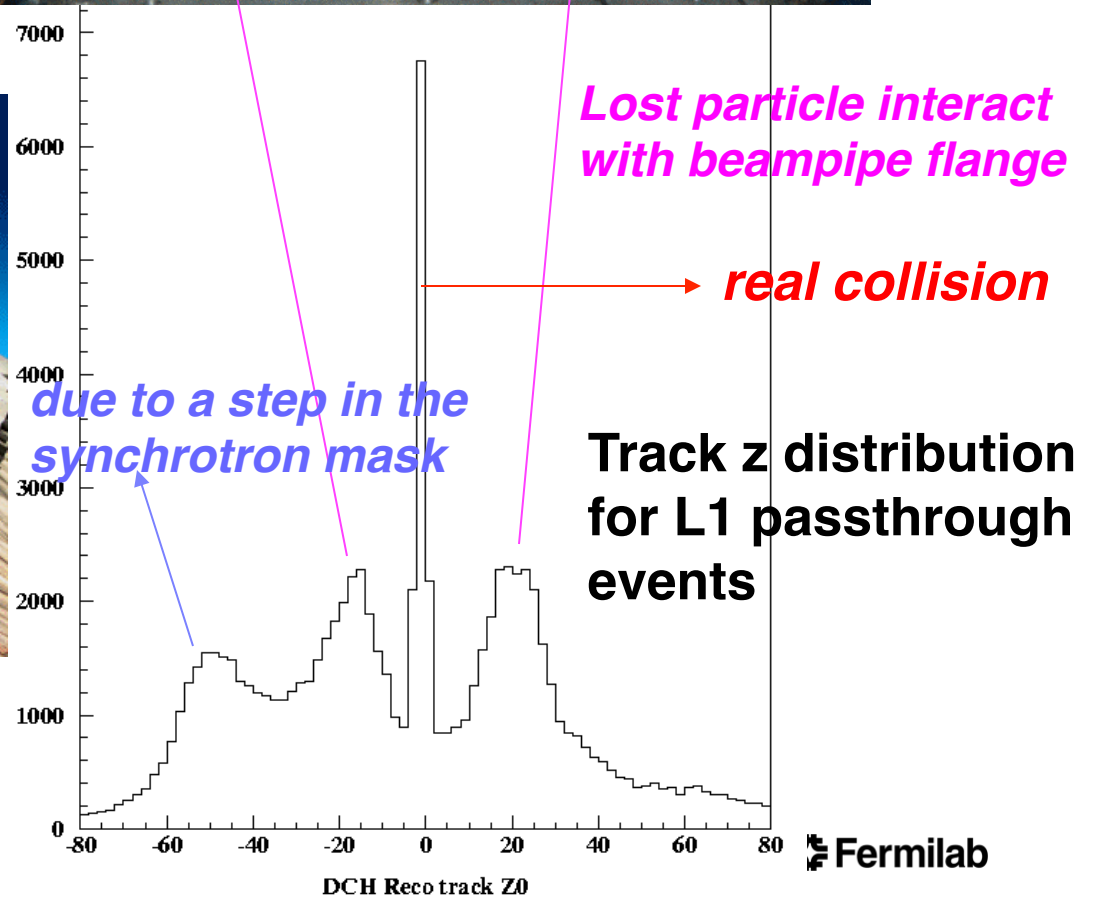
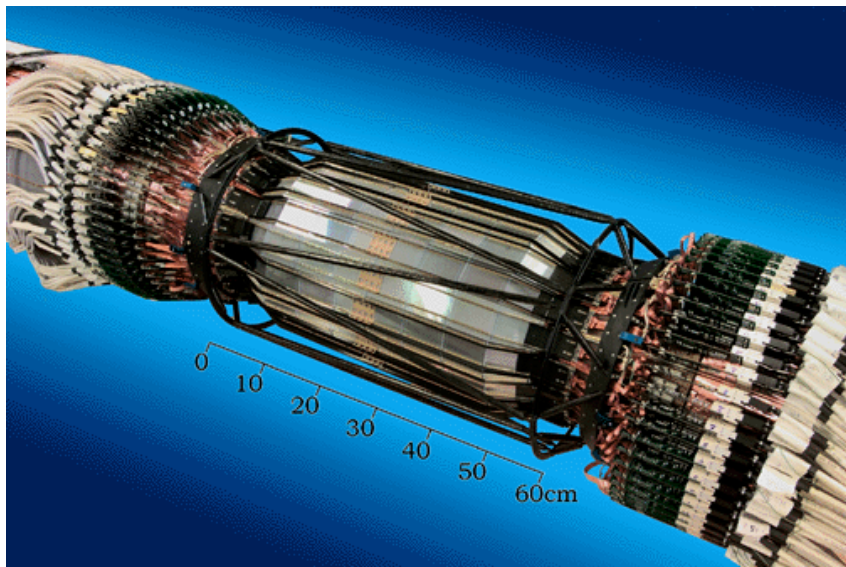
High Luminosity → High Background



HER



LER

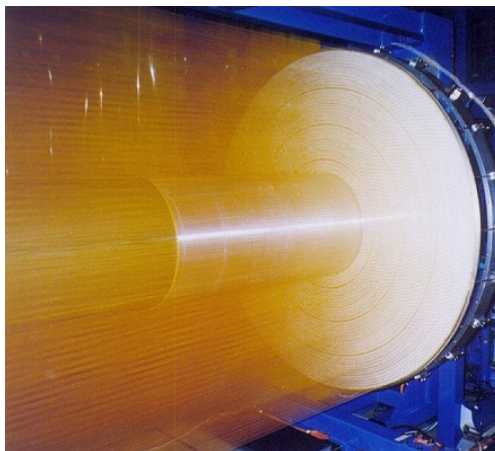
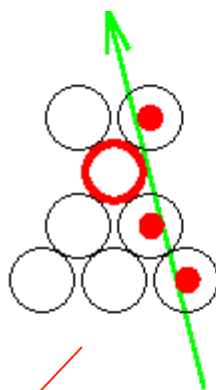


Babar Drift Chamber (tracking) Trigger (LBNL)

The heart of DCT is the **Track Segment Finder**

TSF → continuously live image processor →

The method: using both occupancy and drift-time information, to find track segments continuously with:
time resolution of ~ 100 ns,
spatial resolution ~ 1 mm

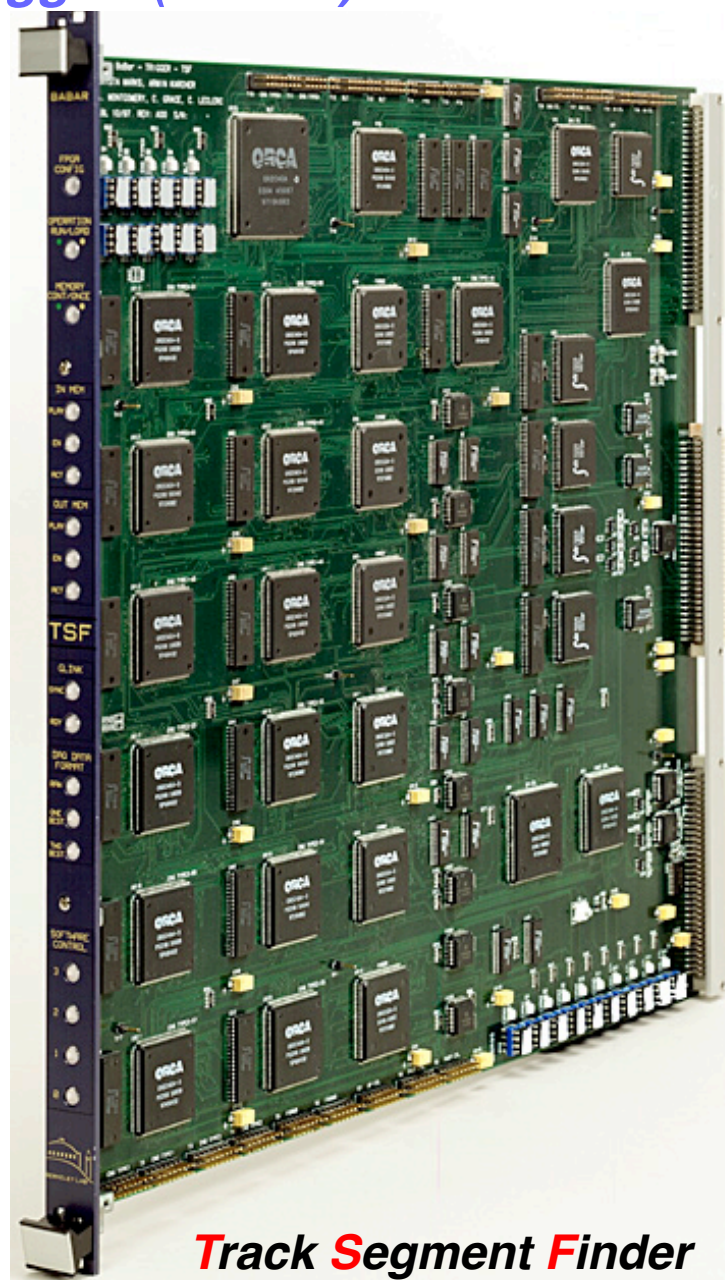


Drift Chamber



Look-Up-Table address

position and time



Track Segment Finder

Tevatron case: $\text{Rate} = \sigma L$

- Basics:

- *Luminosity*: units of $\text{cm}^{-2}\text{s}^{-1}$ or $\text{barn}^{-1} \text{sec}^{-1}$
 - Often in $\sim 10^{32} \text{cm}^{-2}\text{s}^{-1} = 0.1 \text{ nb}^{-1}\text{s}^{-1} = 10^{-4} \text{ pb}^{-1}\text{s}^{-1} =$
 - This was roughly Tevatron RunII luminosity

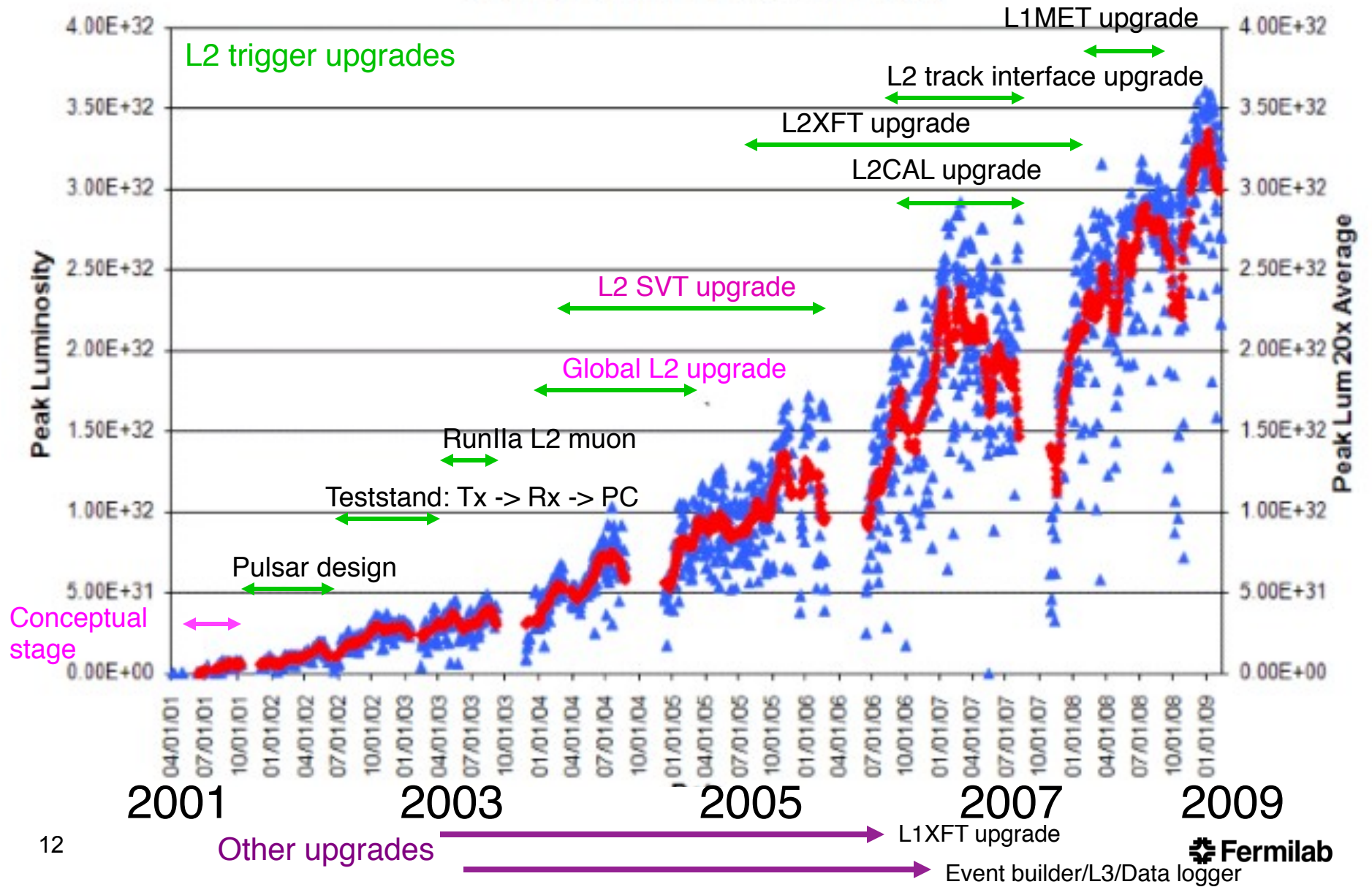
- A few physics processes at Tevatron

- σ (top pair production) $\sim 7 \text{ pb} \rightarrow 0.0007 \text{ Hz}$ or $\sim 60/\text{day}$
- σ (inelastic scattering) at Tevatron $\sim 50 \text{ mb} \rightarrow 5 \text{ MHz}$
- Actual rate is limited by bunch crossing at $\sim 1.7 \text{ MHz}$
 - Multiple interaction per beam crossing ...
- Final events to tape $\sim 100 \text{ Hz} \rightarrow$ challenge

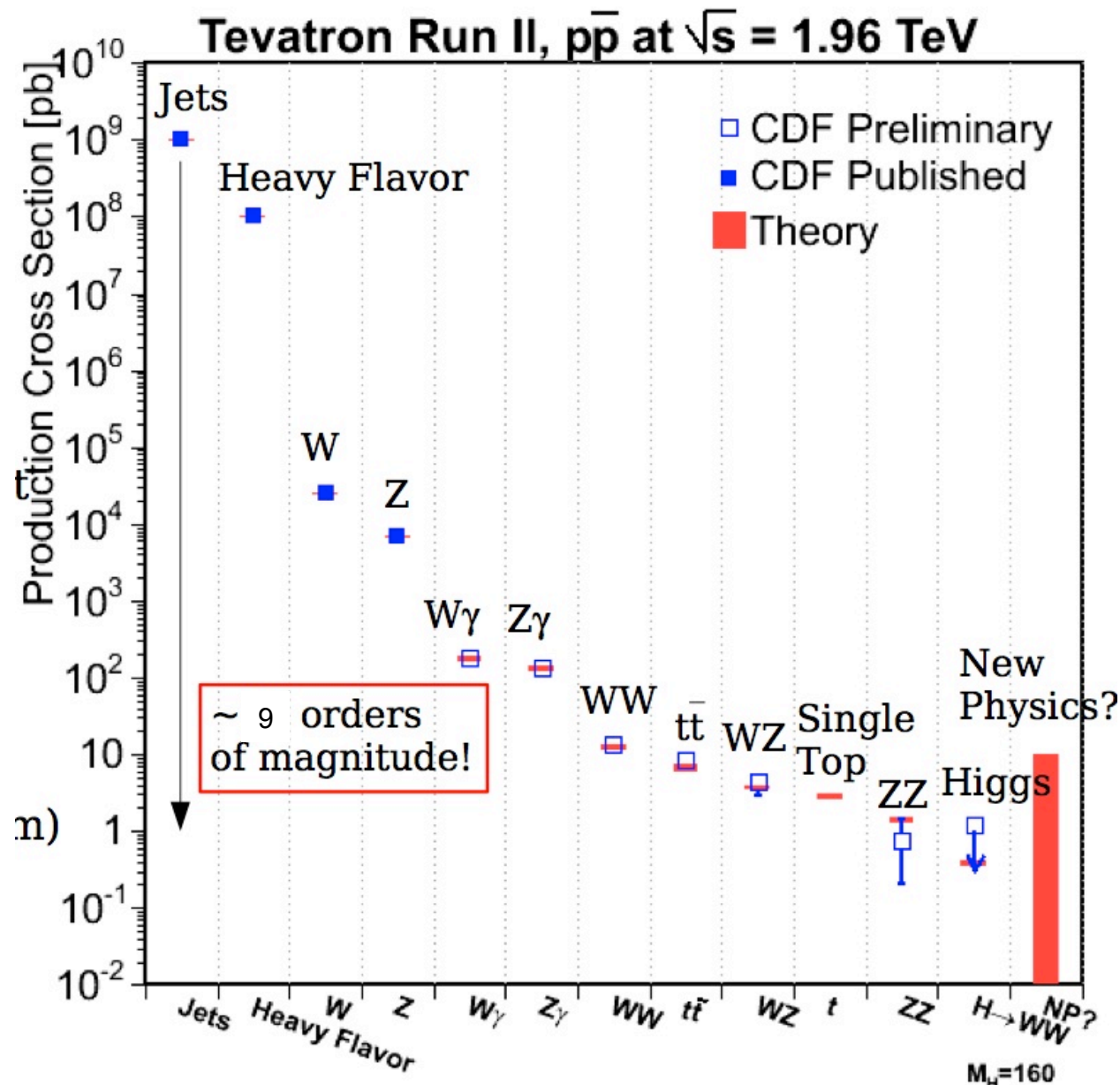
Trigger is crucial for Hadron Collider ...

Tevatron RunII Peak Luminosity

SVT Gigafitter upgrade



Evolution of CDF's RunII Physics Program



Trigger improvements have been crucial

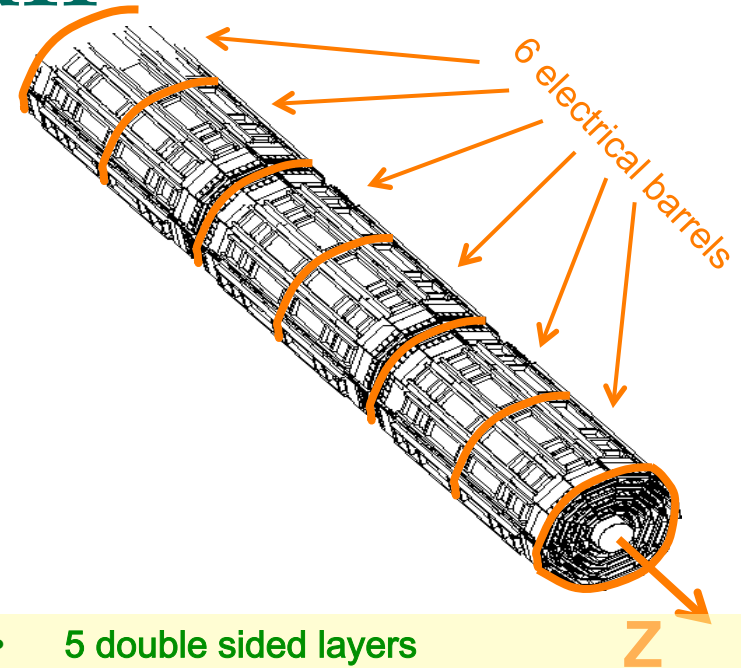
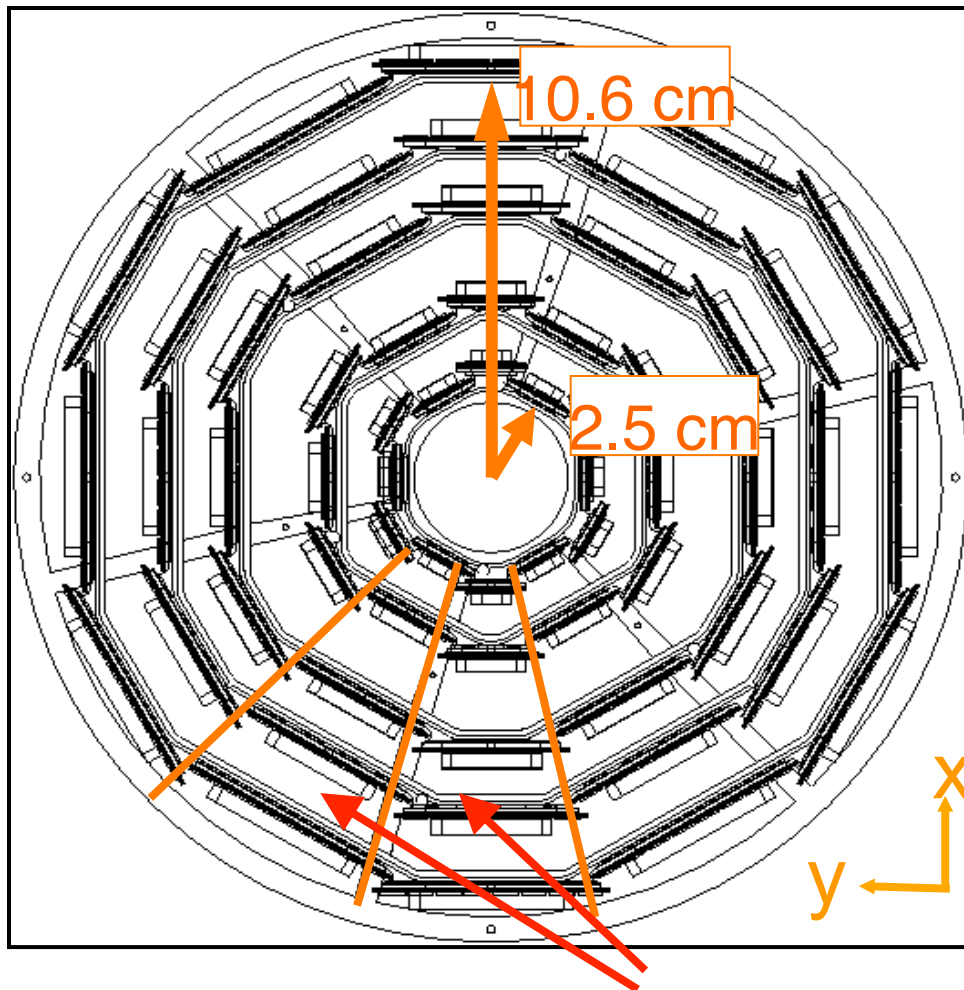
Interesting history of silicon detector and silicon based tracking trigger at hadron collider

- Stories on some technological innovations at CDF in the 1980s-1990s
 - the first silicon detector (SVX) at CDF
 - the first Silicon Vertex Trigger (SVT) at CDF

Initially many people didn't think it was useful or it could have ever worked....

APS Panofsky Prize to Aldo Menzione and Luciano Ristori

CDF SVXII



- **5 double sided layers**
 - 5 axial + 3 x 90°, 2 x 1.2°
- **Very compact**
- **Tight alignment tolerances**
 - For the trigger
- **Very symmetric**
 - 12 fold in Φ
 - 6 barrels in Z

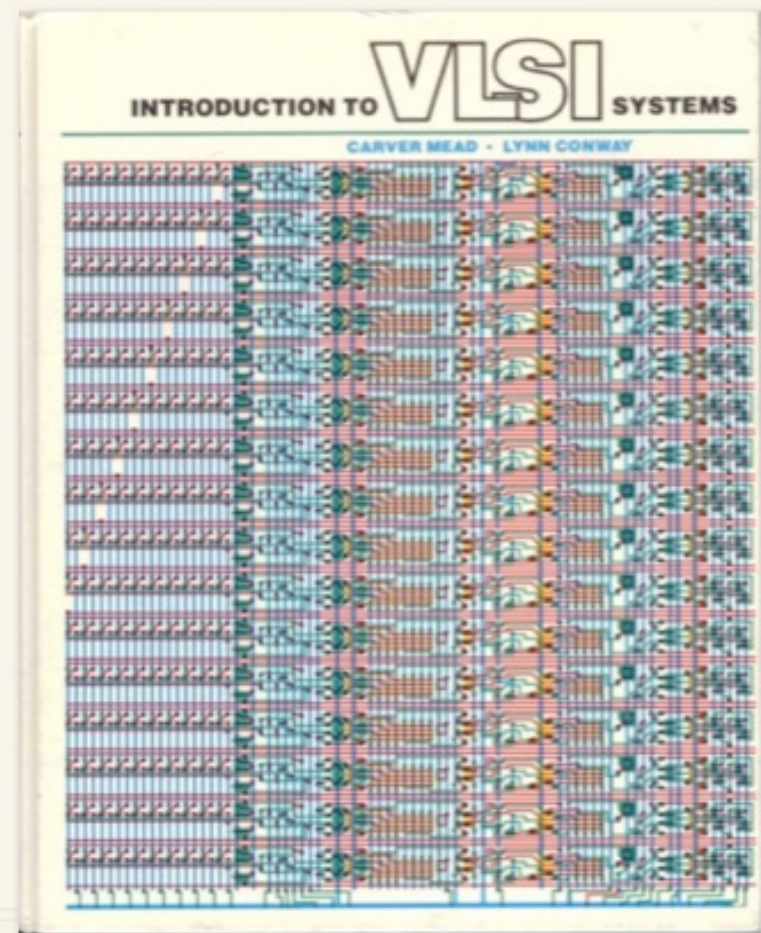
Note “wedge” symmetry

Very Large Scale Integration the revolution

Carver Mead & Lynn Conway

in the '80s the technology of
VLSI design becomes
available to the universities
and to small research
projects

A slide from Luciano Ristori
at TIPP 2011 conference



October 24, 1988

VLSI STRUCTURES FOR TRACK FINDING

Mauro DELL'ORSO

Dipartimento di Fisica, Università di Pisa, Piazza Torricelli 2, 56100 Pisa, Italy

Luciano RISTORI

INFN Sezione di Pisa, Via Vecchia Livornese 582a, 56010 S. Piero a Grado (PI), Italy

Received 24 October 1988

We discuss the architecture of a device based on the concept of *associative memory* designed to solve the track finding problem, typical of high energy physics experiments, in a time span of a few microseconds even for very high multiplicity events. This "machine" is implemented as a large array of custom VLSI chips. All the chips are equal and each of them stores a number of "patterns". All the patterns in all the chips are compared in parallel to the data coming from the detector while the detector is being read out.

1. Introduction

The quality of results from present and future high energy physics experiments depends to some extent on the implementation of fast and efficient track finding algorithms. The detection of *heavy flavor* production, for example, depends on the reconstruction of secondary vertices generated by the decay of long lived particles, which in turn requires the reconstruction of the majority of the tracks in every event.

Particularly appealing is the possibility of having detailed tracking information available at trigger level even for high multiplicity events. This information could be used to select events based on impact parameter or secondary vertices. If we could do this in a sufficiently short time we would significantly enrich the sample of events containing heavy flavors.

Typical events feature up to several tens of tracks each of them traversing a few position sensitive detector layers. Each layer detects many hits and we must correctly correlate hits belonging to the same track on different layers before we can compute the parameters

2. The detector

In this discussion we will assume that our detector consists of a number of layers, each layer being segmented into a number of bins. When charged particles cross the detector they hit one bin per layer. No particular assumption is made on the shape of trajectories: they could be straight or curved. Also the detector layers need not be parallel nor flat. This abstraction is meant to represent a whole class of real detectors (drift chambers, silicon microstrip detectors etc.). In the real world the coordinate of each hit will actually be the result of some computation performed on "raw" data: it could be the center of gravity of a cluster or a charge division interpolation or a drift-time to space conversion depending on the particular class of detector we are considering. We assume that all these operations are performed upstream and that the resulting coordinates are "binned" in some way before being transmitted to our device.

M. Dell'Orso, L. Ristori / VLSI structures for track finding

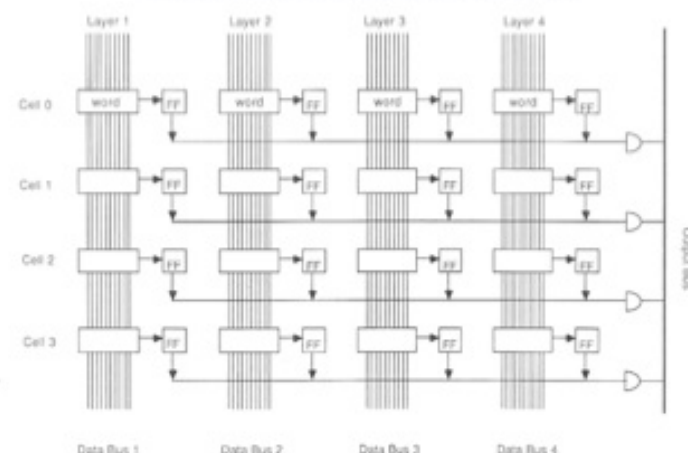


Fig. 3. Associative memory architecture.

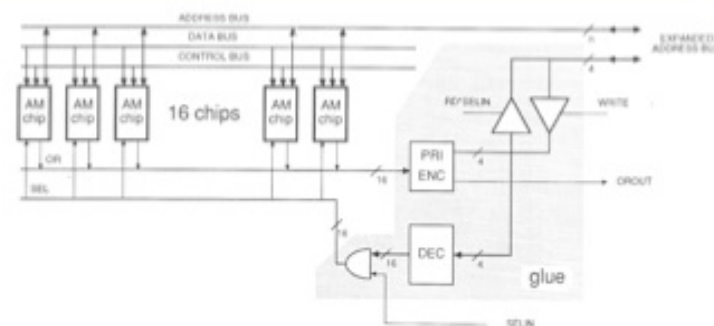


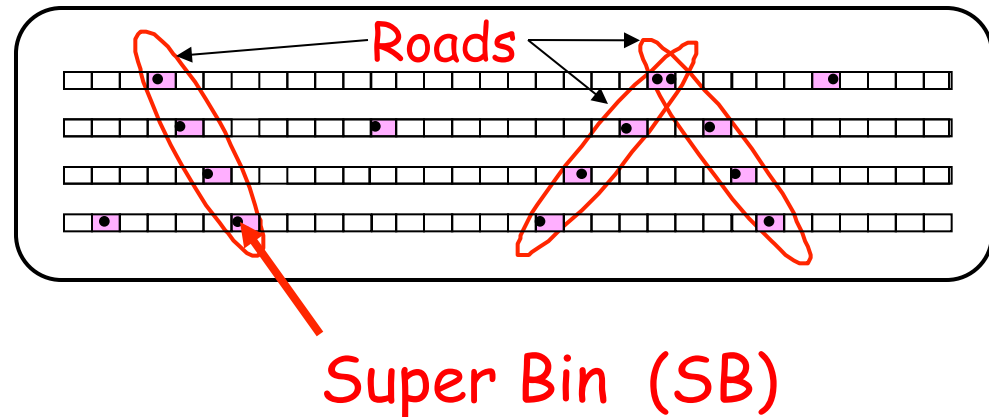
Fig. 5. 16 AM chips tied by the "glue".

We discuss the architecture of a device based on the concept of *associative memory* designed to solve the track finding problem, typical of high energy physics experiments, in a time span of a few microseconds even for very high multiplicity events. This "machine" is implemented as a large array of custom VLSI chips. All the chips are equal and each of them stores a number of "patterns". All the patterns in all the chips are compared in parallel to the data coming from the detector while the detector is being read out.

Tracking in 2 steps

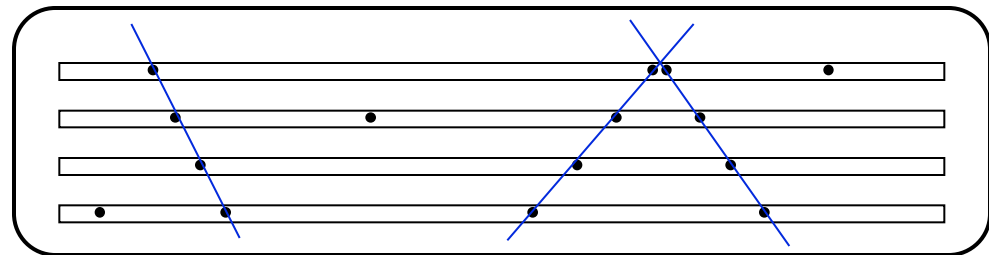
- Pattern recognition and track fitting done separately and pipelined

1. Find low resolution track candidates called "roads".



2. Then fit tracks inside roads.

a linear approximation
gives near ideal precision

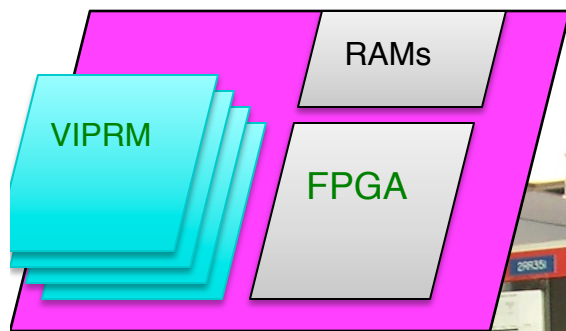


A very successful approach at CDF for RunII, based on Associative Memory, which in turn, based on CAM

CDF original SVT system had ~400K patterns total ...
128 patterns per AMchip -- commissioned around ~2001.



The rest, is history ... (the rich physics program out of CDF)



*Can we put entire SVT in one chip?
(see later in the talk)*

2 meters



Original SVT system had ~400K patterns total

LHC case: $\text{Rate} = \sigma L$

- Some basics:

- ↘ *Luminosity*: units of $\text{cm}^{-2}\text{s}^{-1}$ or $\text{barn}^{-1} \text{sec}^{-1}$

- Often in $\sim 10^{34} \text{cm}^{-2}\text{s}^{-1} = 10 \text{ nb}^{-1}\text{s}^{-1} = 10^{-2} \text{pb}^{-1}\text{s}^{-1}$

- This is LHC design luminosity

- A few physics processes at LHC at 14 TeV

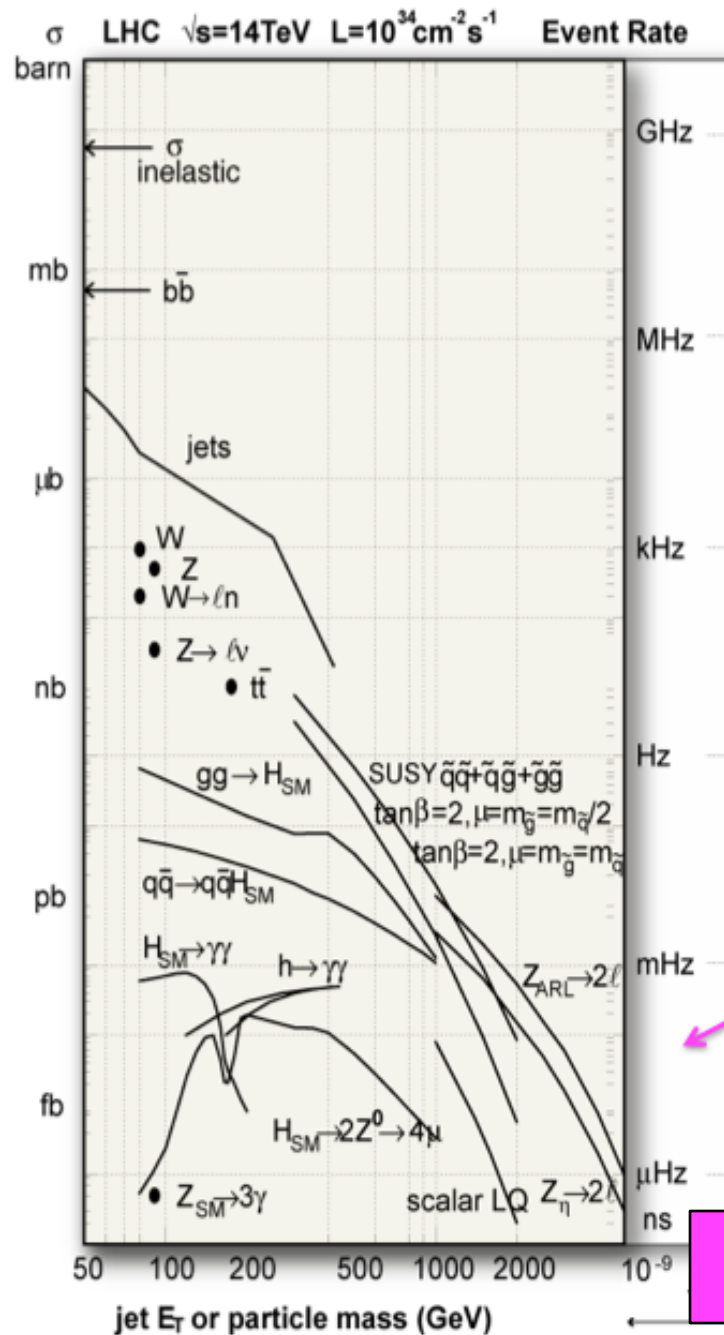
- ↘ σ (top pair production) $\sim 700 \text{ pb} \rightarrow 7 \text{ Hz}$ or $\sim 600\text{K/day}$

- ↘ σ (inelastic scattering) $\sim 70 \text{ mb} \rightarrow 700 \text{ MHz}$ (interaction)

- ↘ Actual event rate is limited by bunch crossing at 20/40MHz

- Multiple interaction per beam crossing ...

Why go to such high luminosity with all the trouble?



$$\text{Rate} = \sigma L$$

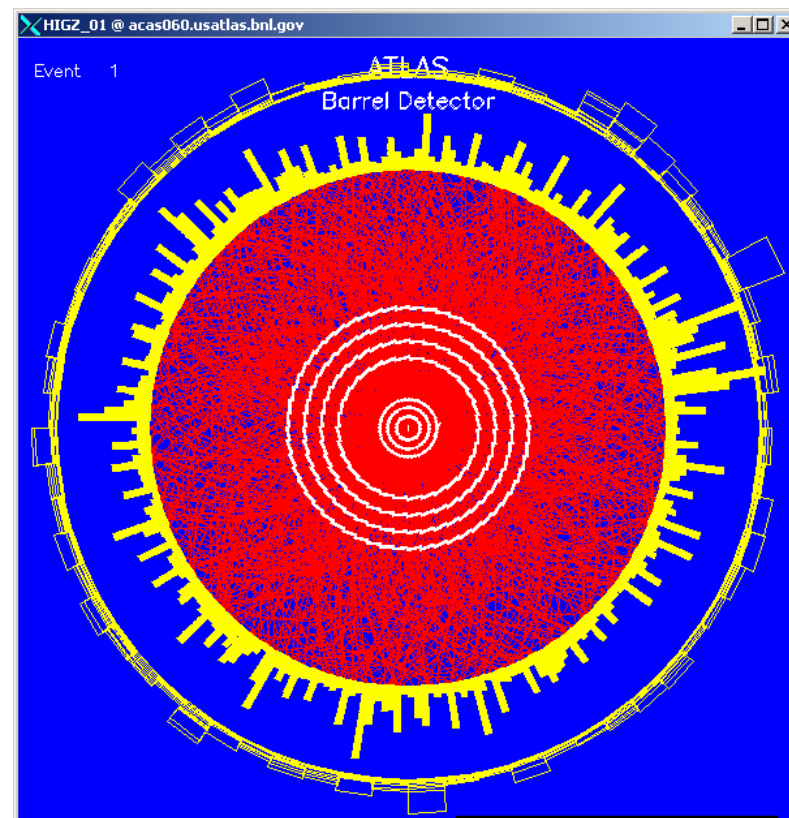
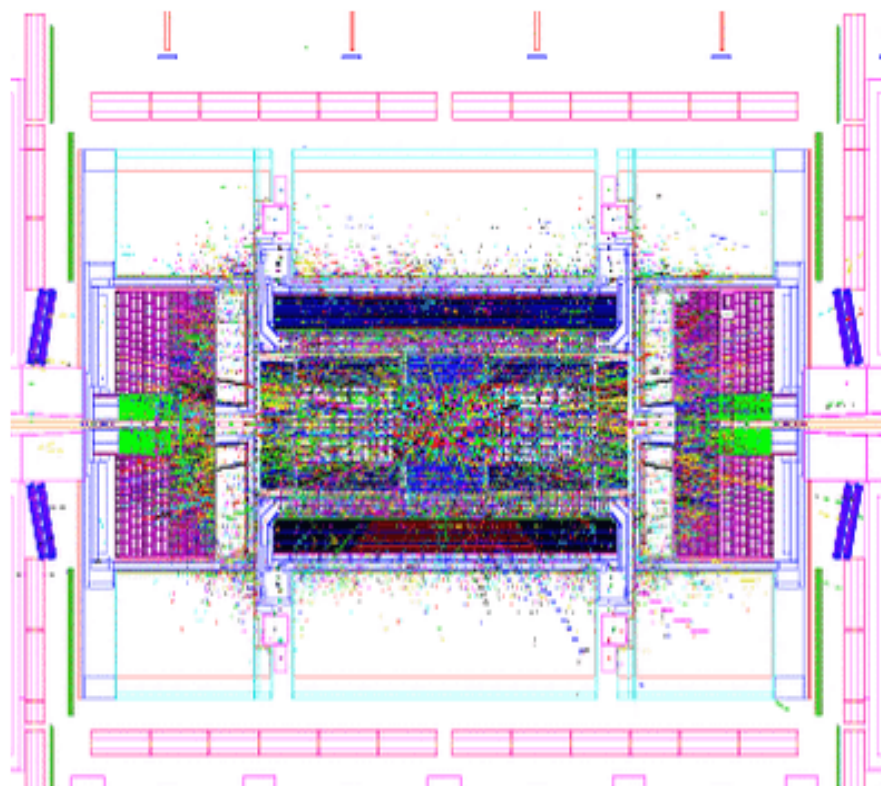
Trigger & reconstruction eff.

Hunting for new particle with
cross section down to
 $\sim 1 \text{ fb}$

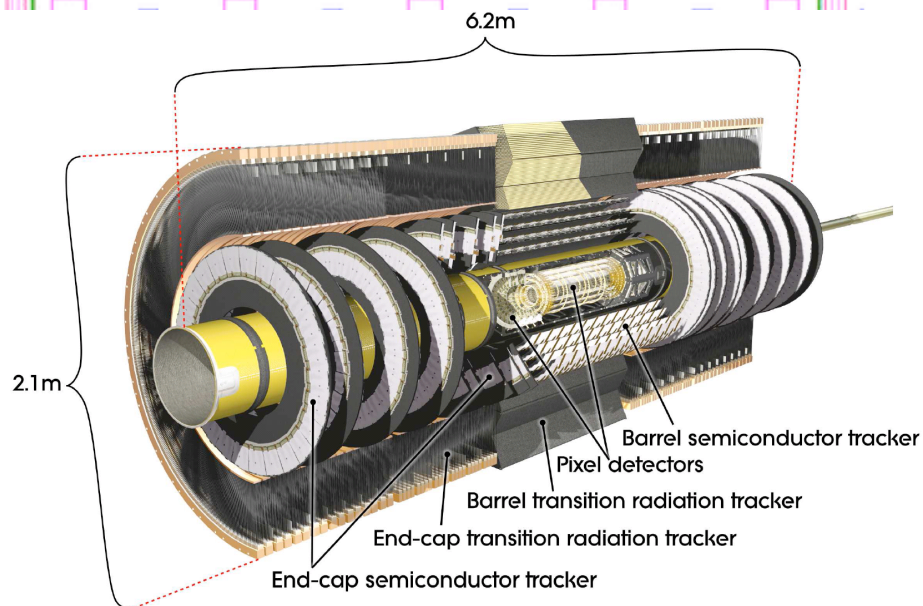
$$100 \text{ fb}^{-1} \text{ per year} \sim 10^{-5} \text{ fb}^{-1}\text{s}^{-1} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$$

The need for Higher LHC Luminosity

Expected Pile-up at High Lumi LHC in ATLAS at 10^{35}

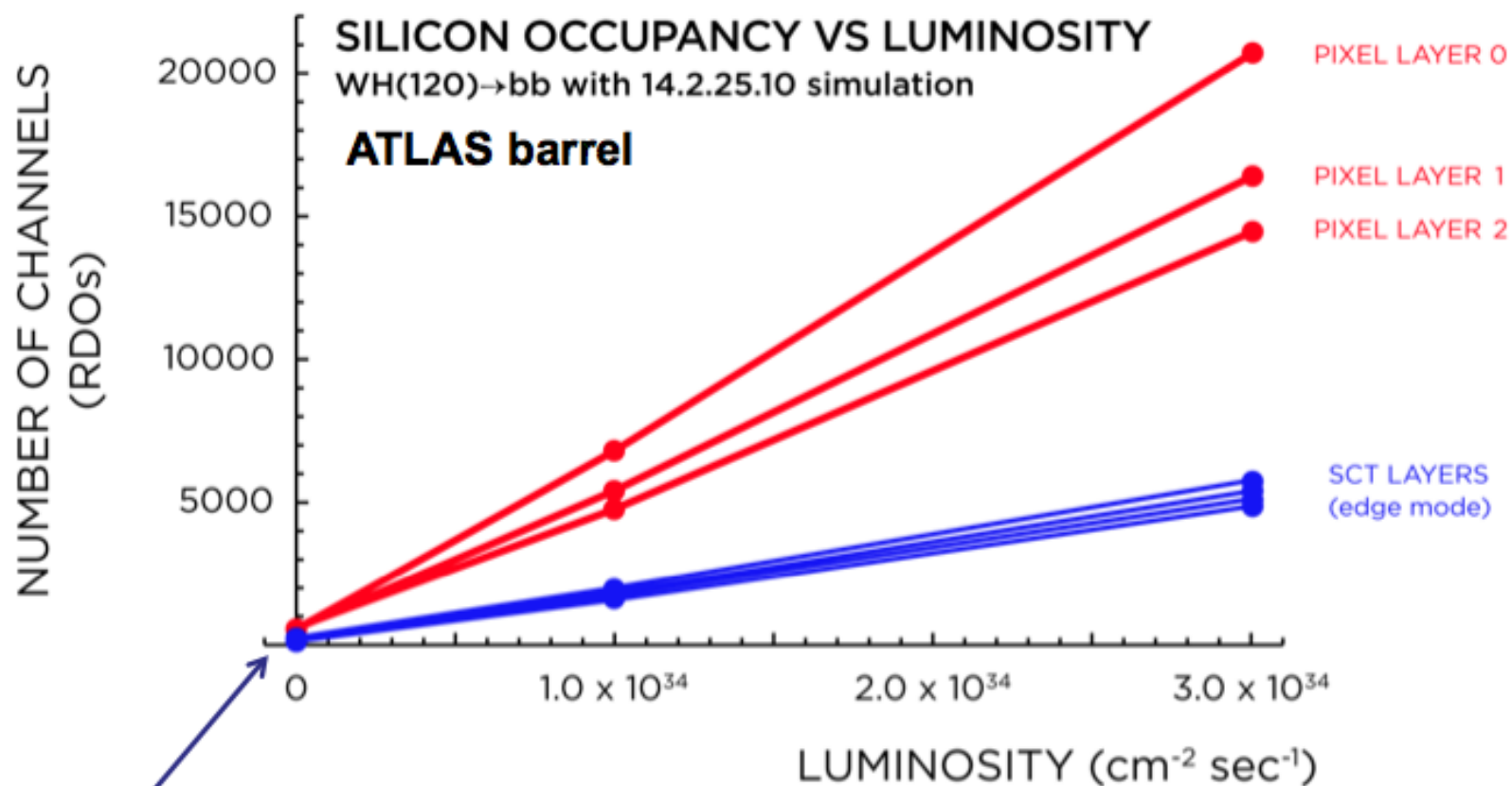


$N_{ch}(|y| \leq 0.5)$



The effect of pile-up on silicon detector occupancy

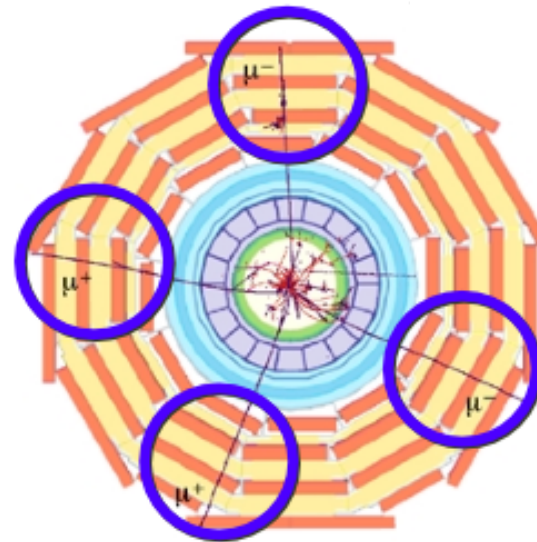
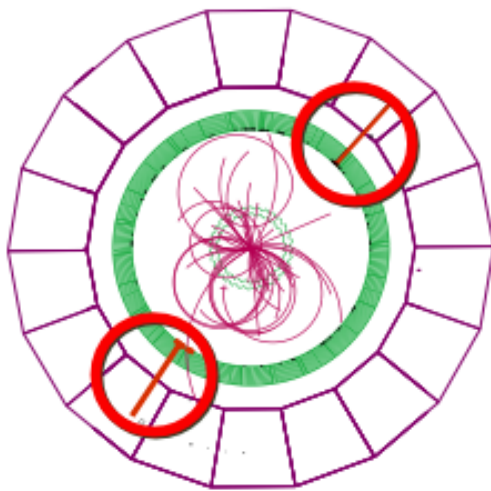
From Atlas FTK Technical Proposal



y intercept: all of the tracks from the primary interaction!

ATLAS and CMS only have calorimeter and muon trigger at Level 1

Pattern recognition much easier on calo & muon:

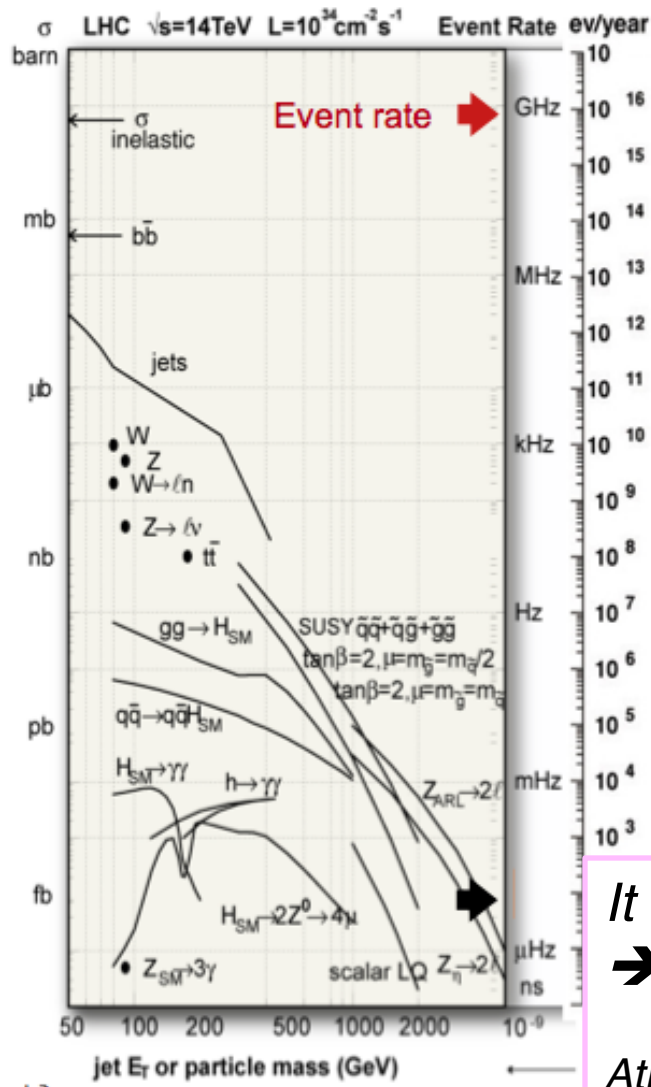


Complexity
handled in
software on
CPUs

at high level
trigger

The approach works well at low luminosity ...

Collisions (p-p) at LHC



Operating conditions:
one "good" event (e.g Higgs in 4 muons)
+ ~20 minimum bias events)

All charged tracks with $p_t > 2\text{ GeV}$

Reconstructed tracks with $p_t > 25\text{ GeV}$

*It is crucial to have tracking at early trigger stages
→ enormous challenges*

Atlas FTK (Fast TrackEr) is the first step in this direction at LHC, and uses all silicon tracker info at L2

The LHC case

Frontier physics at LHC in the future:

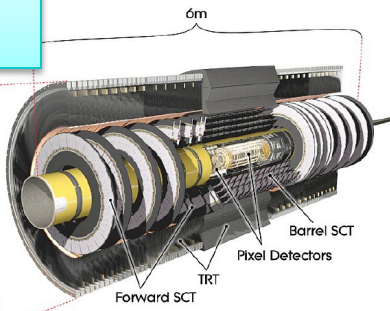
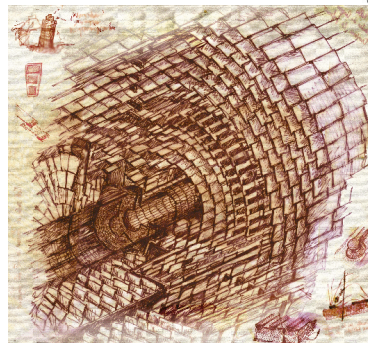
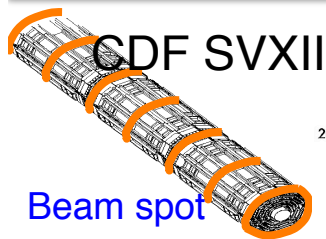
Go to Higher Energy: needs aggressive R&D in Magnet
OR

Go to Higher Luminosity: needs aggressive Tracking Trigger R&D

- *It has become clear that track based trigger capability will be crucial to the frontier physics reach at LHC in the future, as luminosity increases.*
- *The current technology using fiber data transfer, FPGAs, custom chips and modern PCs cannot be scaled in a simple manner to accommodate all the tracking trigger demands.*
- *Significant improvements, or breakthroughs, will be needed. In other words, aggressive R&D efforts will be required.*

→ How to take full advantages of modern technology

Detector design for triggering



Data transfer

Data
formatting

Tracking Trigger Issues at hadron colliders (L1 & L2/HLT)

Pattern
Recognition

Associative Memory
Others ...

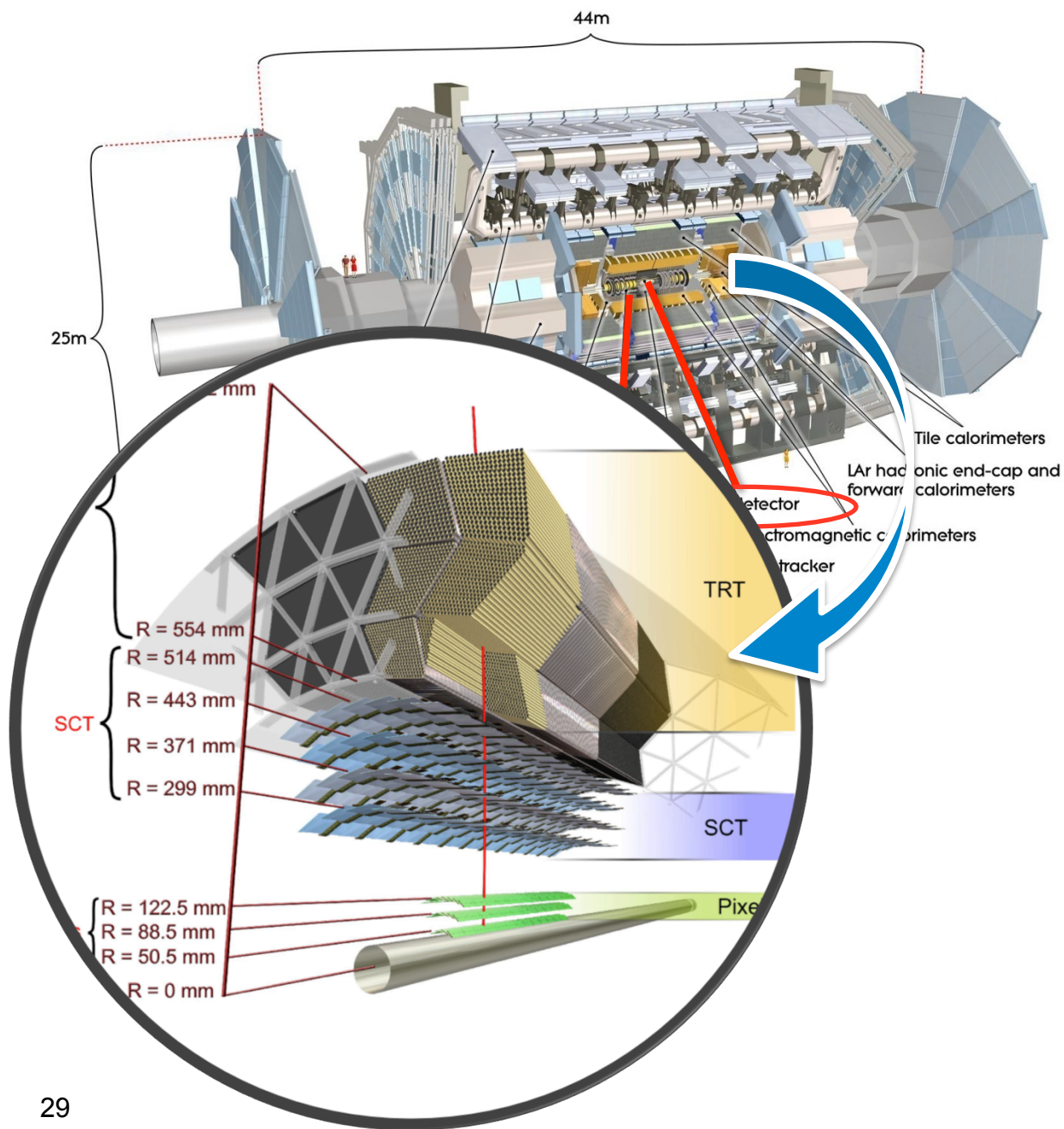
Track
Fitting

FPGA vs CPU vs GPU

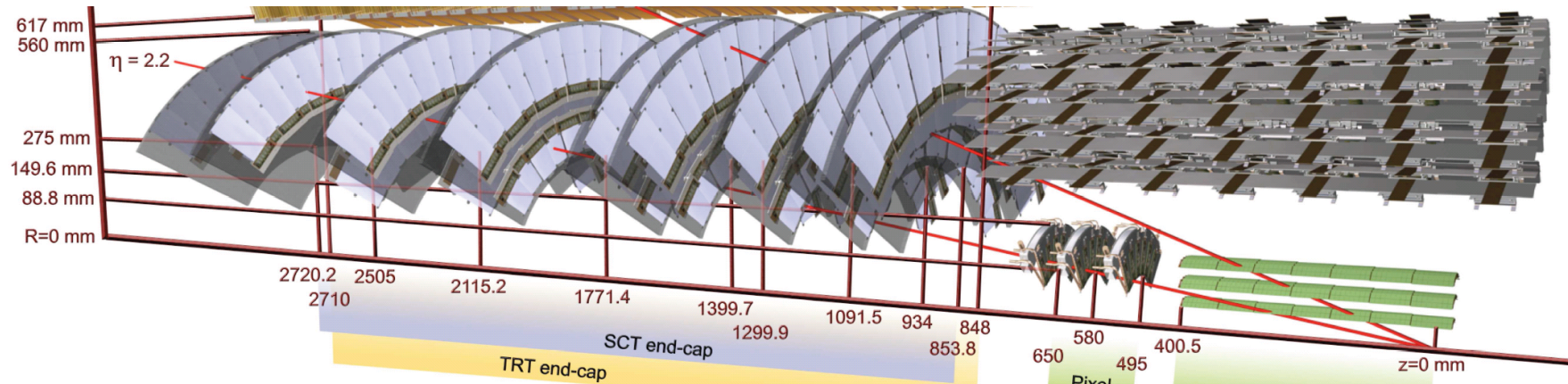
HLT

Selected R&D topics

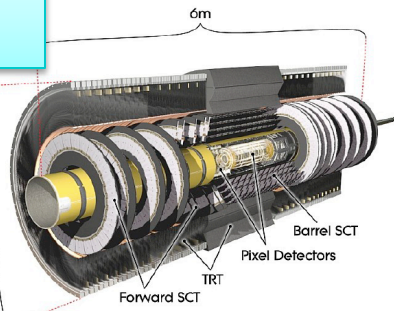
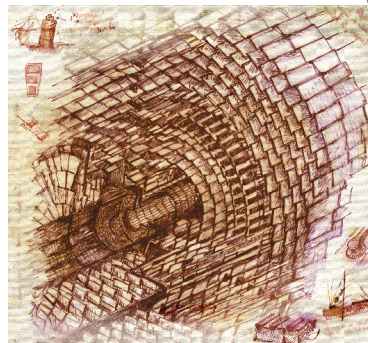
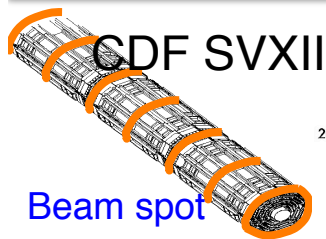
- (1) Data Reduction at detector/sensor stage
- (2) Data Transfer challenges
- (3) Data Formatter using ATCA technology
- (4) Pattern Recognition: 3D AM R&D: VIPRAM
= "Vertically Integrated Pattern Recognition Associative Memory"
- (5) Track Fitting: GPU study for low latency tracking



A sense of scale: Atlas Silicon Tracker vs CDF SVX II



Detector design for triggering



Data transfer

Data
formatting

Tracking Trigger Issues at hadron colliders (L1 & L2/HLT)

Pattern
Recognition

Associative Memory
Others ...

Track
Fitting

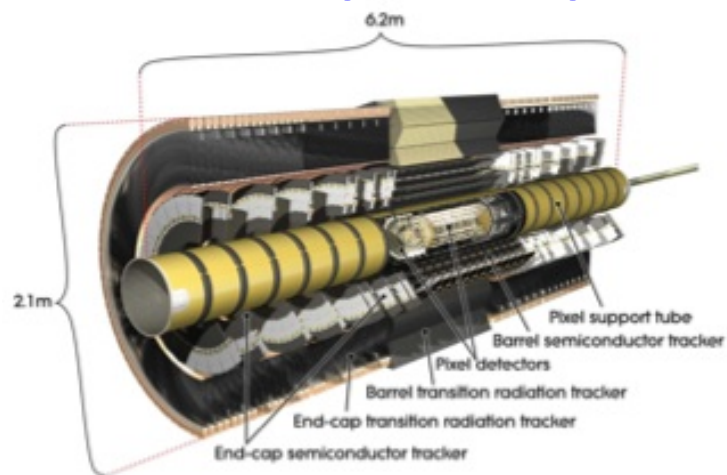
FPGA vs CPU vs GPU

HLT

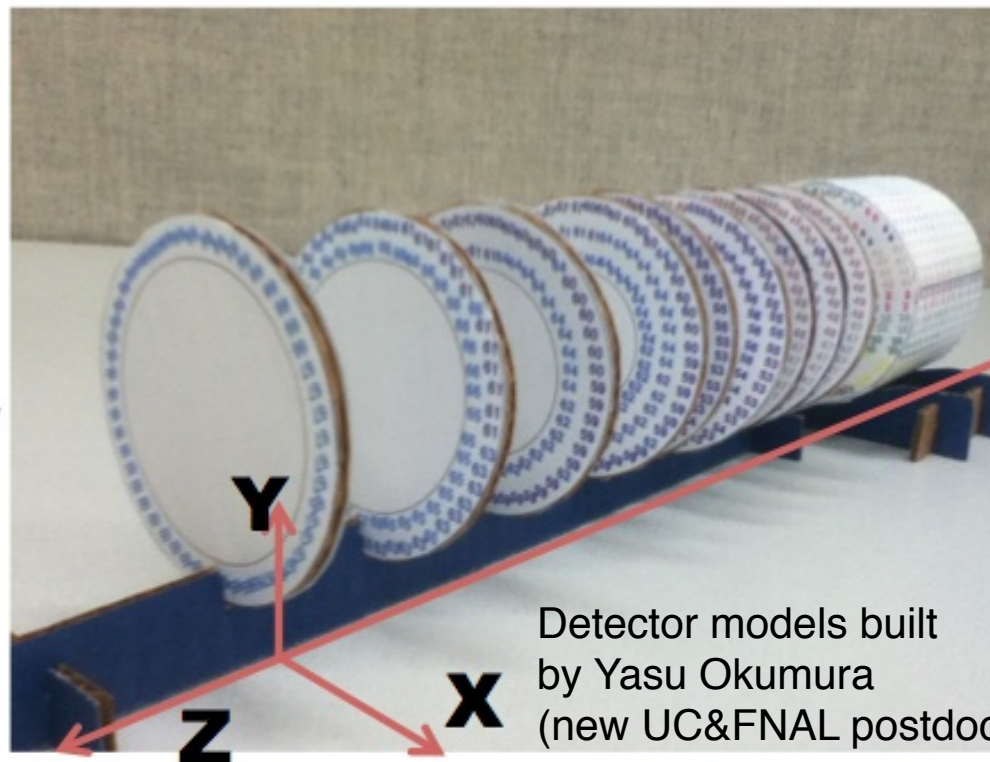
Selected R&D topics

- (1) Data Reduction at detector/sensor stage
- (2) Data Transfer challenges
- (3) Data Formatter using ATCA technology
- (4) Pattern Recognition: 3D AM R&D: VIPRAM
= "Vertically Integrated Pattern Recognition Associative Memory"
- (5) Track Fitting: GPU study for low latency tracking

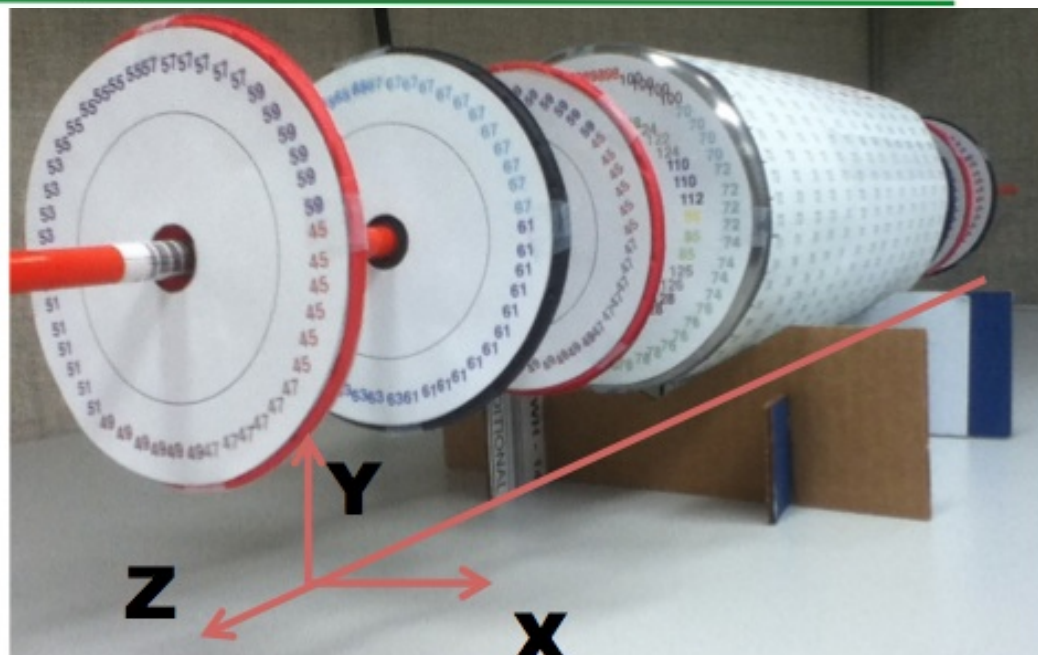
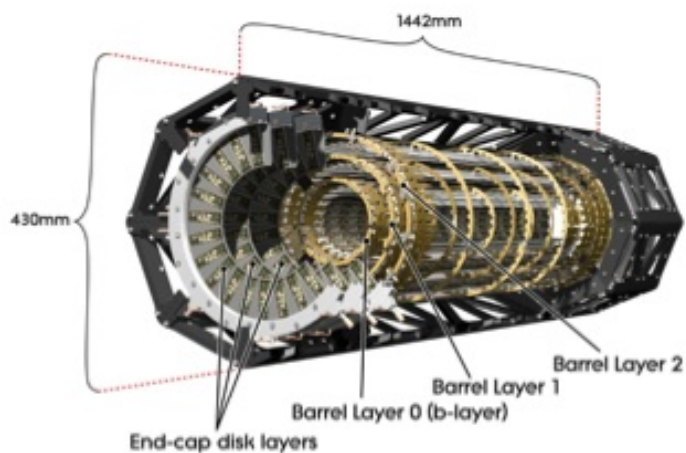
Data Formatting Challenges



Massive amount of data need to be shared for pattern recognition for tracking trigger



Detector models built by Yasu Okumura (new UC&FNAL postdoc)



Data Formatting

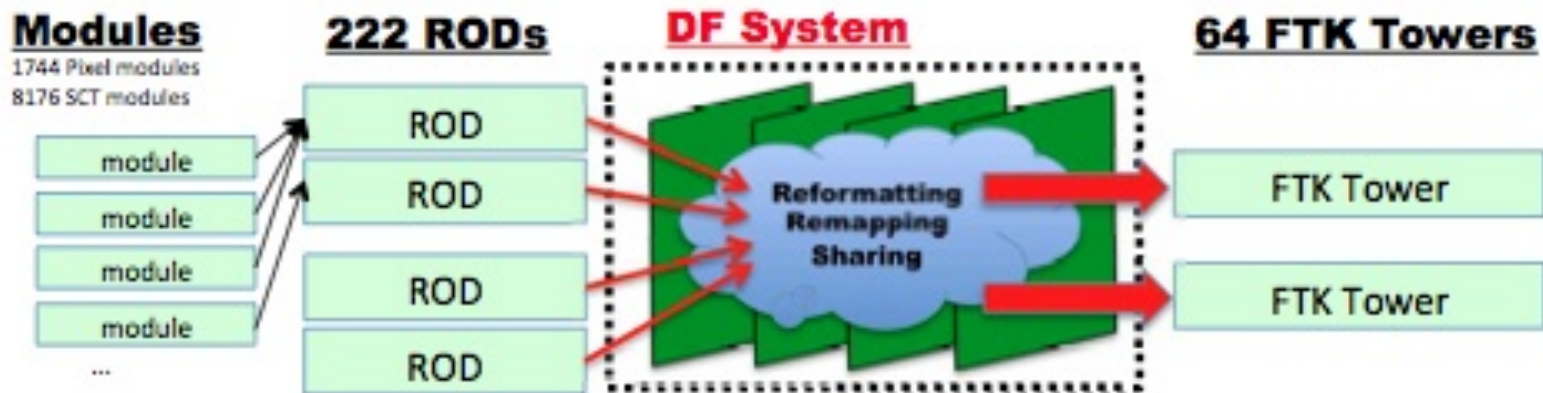
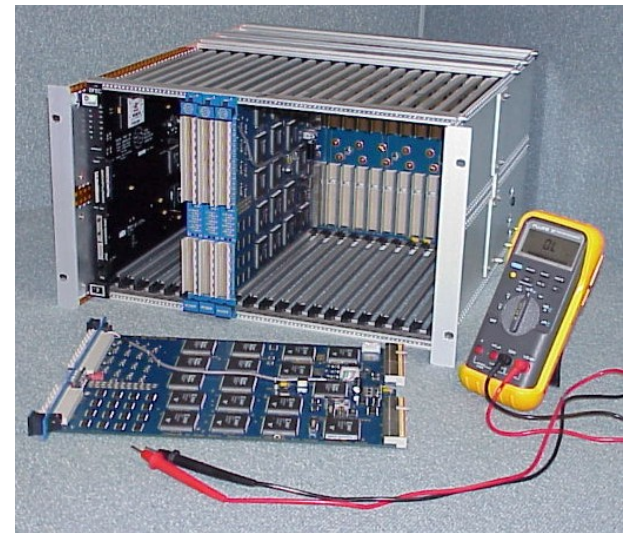
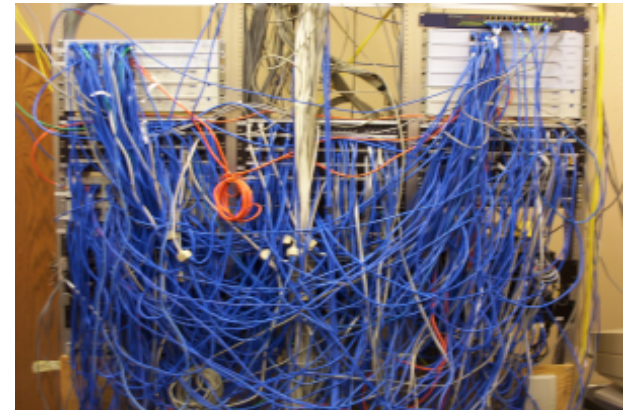


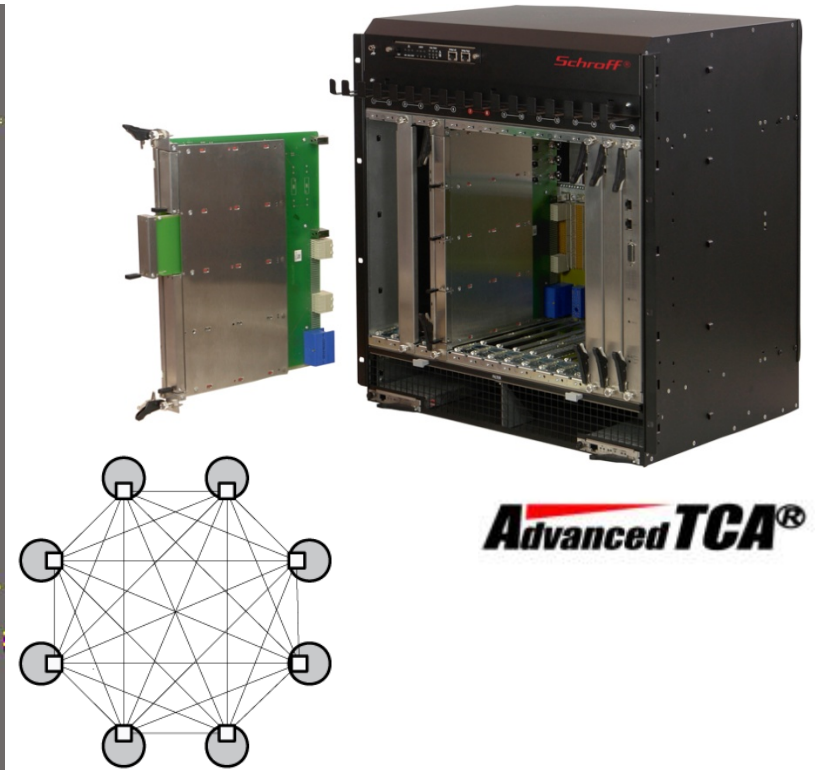
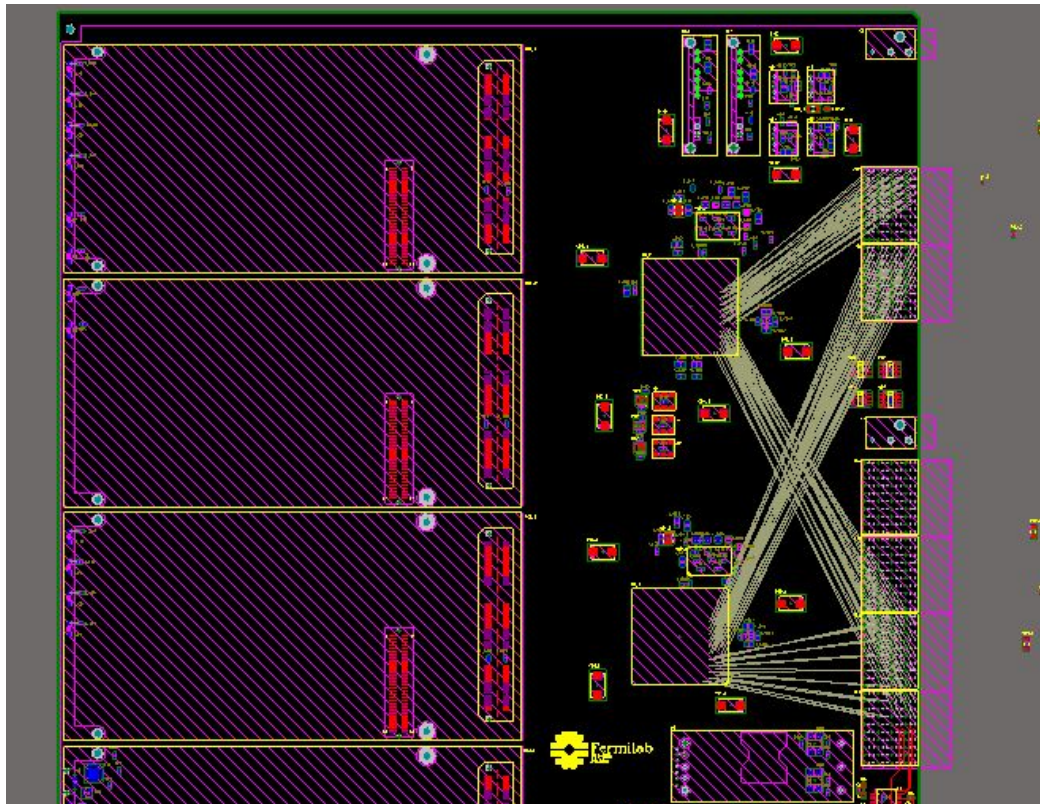
Figure 2: Data from 222 inputs to the Data Formatter and are sent to downstream 64 FTK towers after remapping, reformatting, and sharing.

Old Data Sharing techniques

-- a slide from Jamieson Olsen (FNAL engineer)

- Jumper cables
 - Flexible, but ugly and difficult to maintain
 - Still requires custom backplane
- Dedicated traces on the backplane
 - Custom backplane
 - Each crate may be different
 - Inflexible design
- Another option?
 - Modern ATCA with full-mesh





Data Formatter ATCA board (Pulsar-IIa) design at Fermilab with full-mesh backplane for data sharing (one FPGA per trigger tower, 64 trigger towers total)

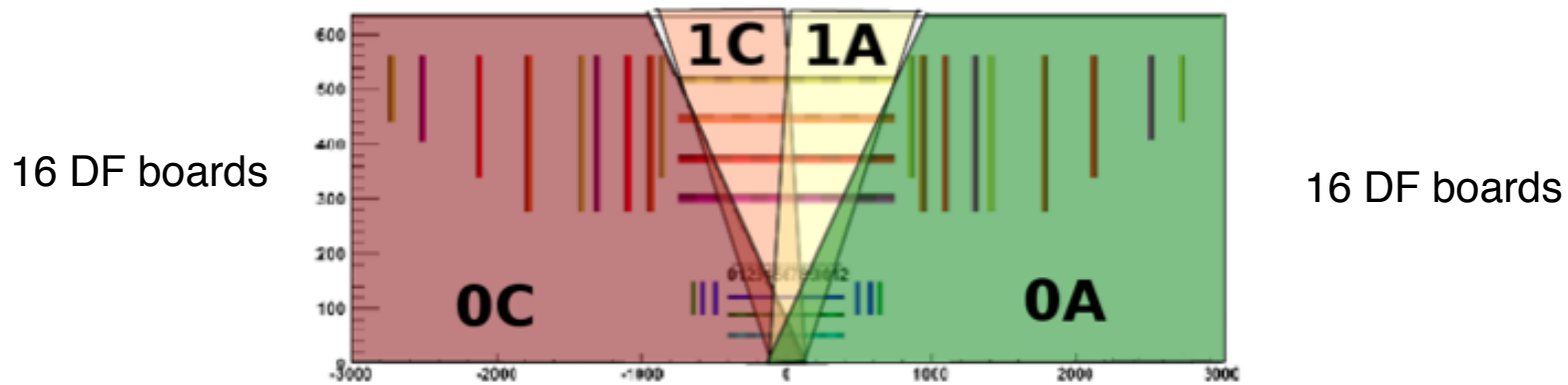
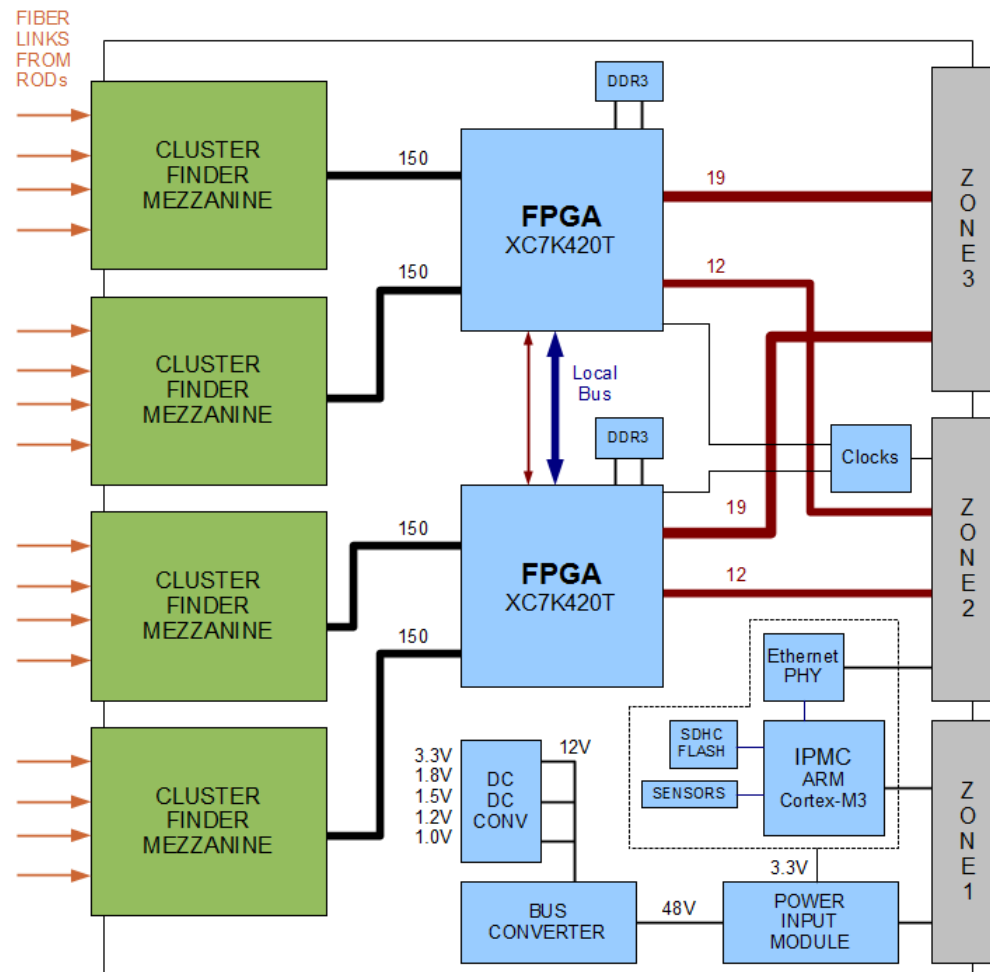


Figure 2: The Four FTK η regions. Note the significant overlap in the high occupancy central barrel regions.

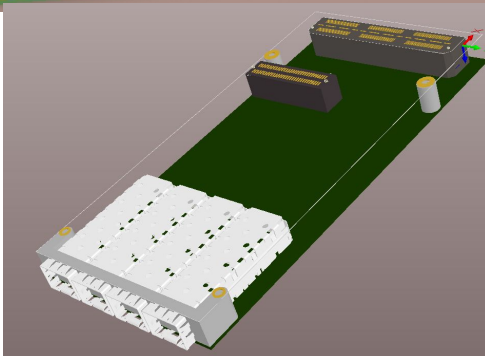
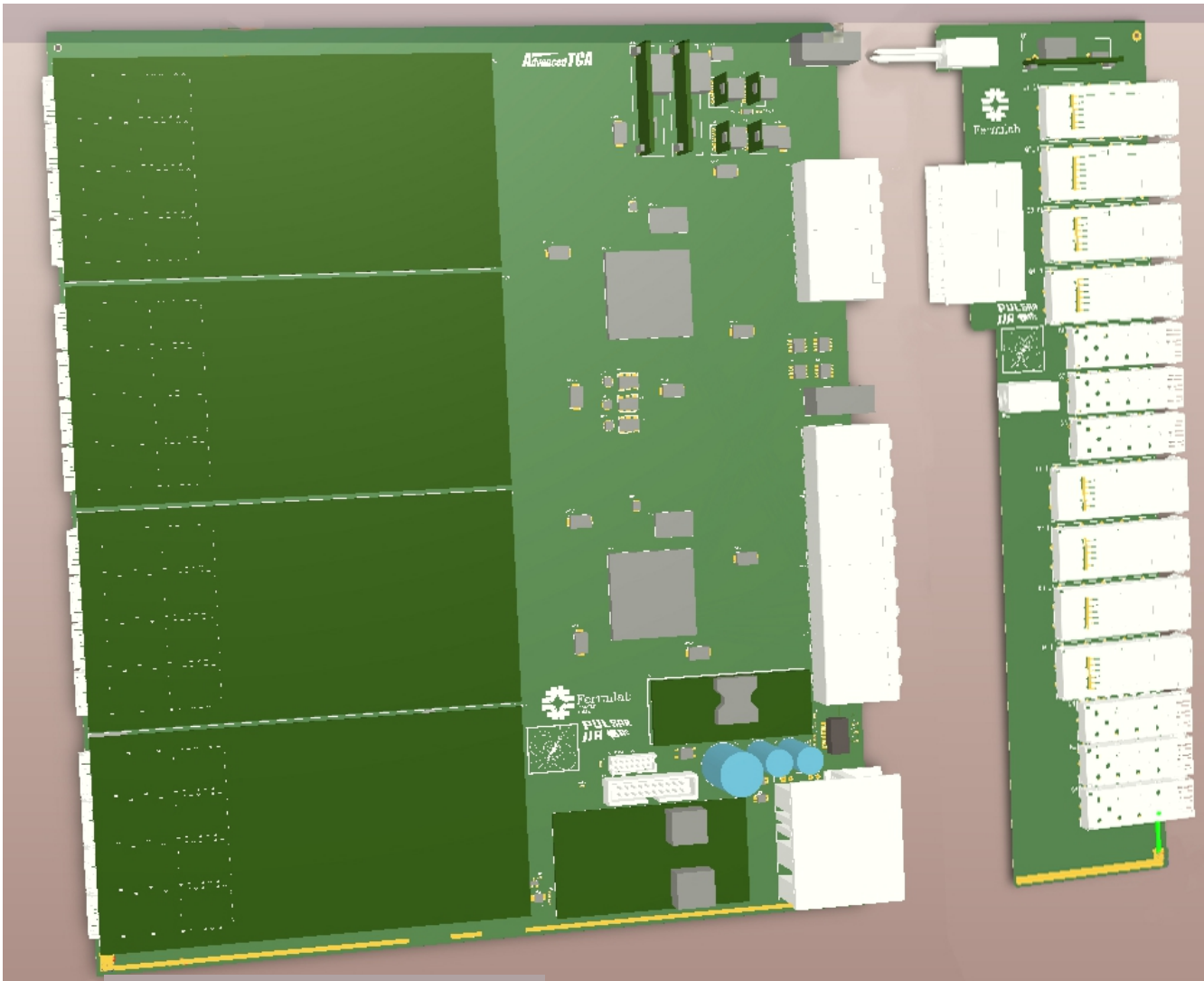
Block Diagram



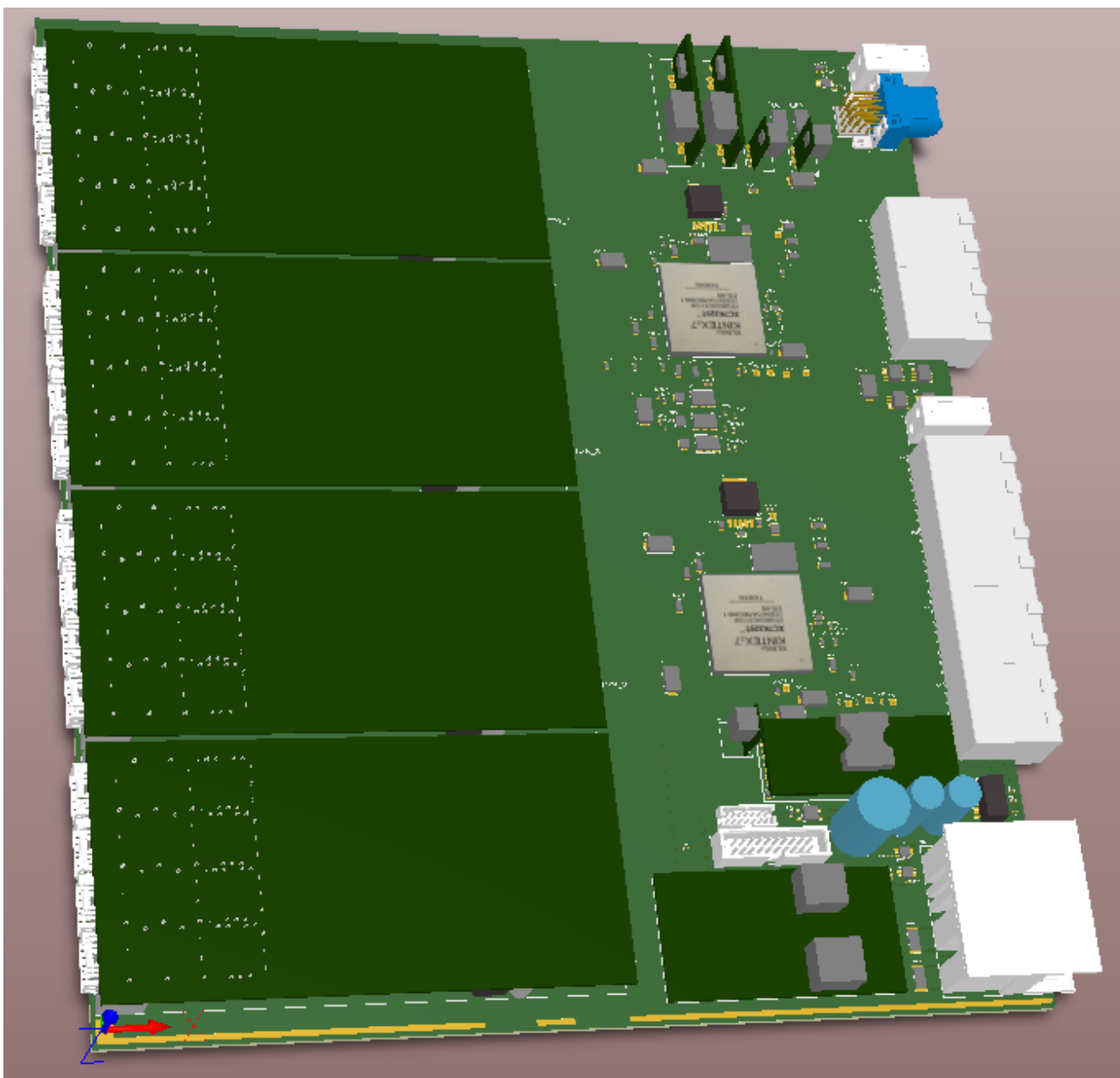
Pulsar-IIa

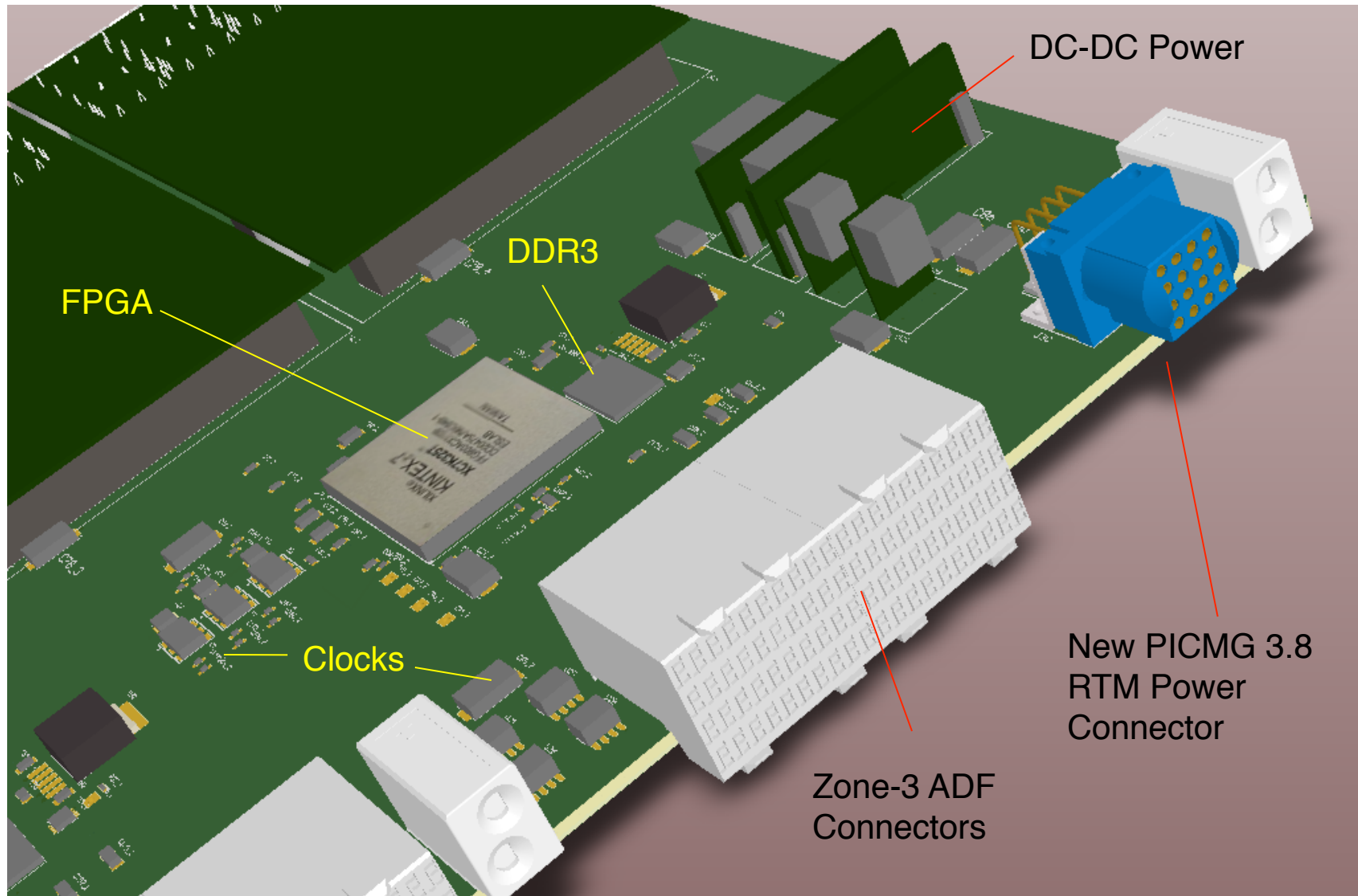
I/O bandwidth
approaching:

~ 1 Tbps



*Challenging Implementation
Developed for Atlas FTK (L2) at Fermilab,
potentially useful for L1 track trigger in the future*







Data Formatter System for FTK

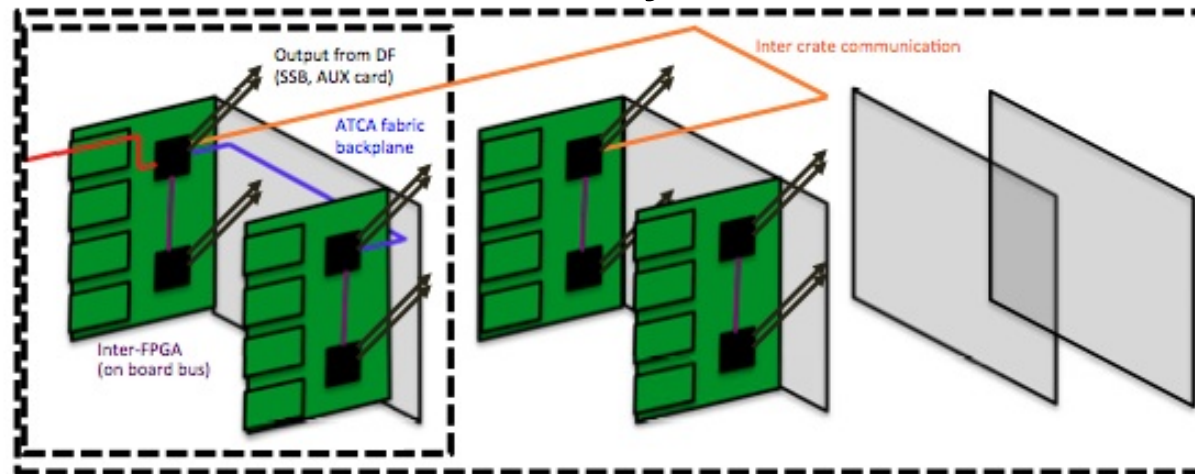
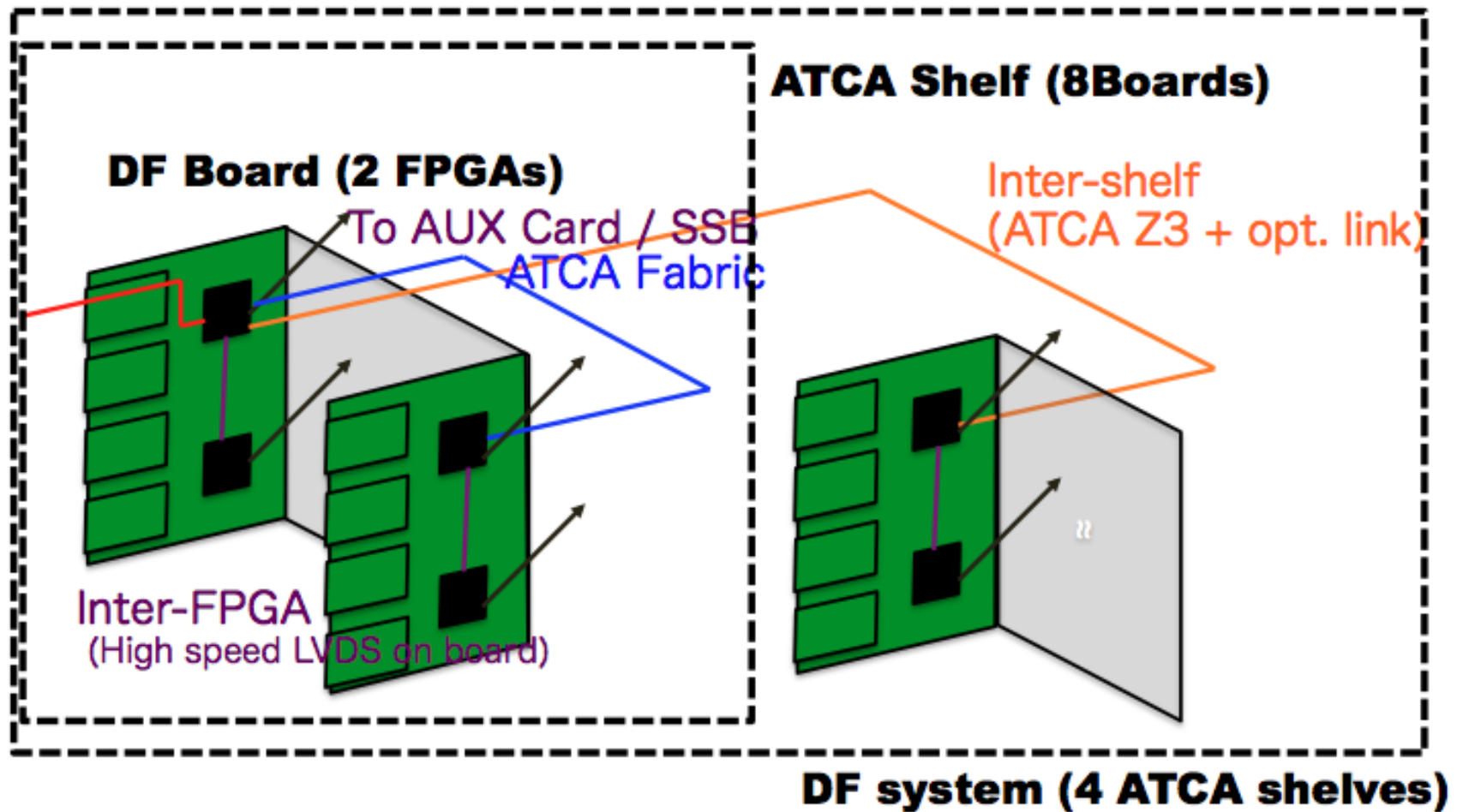
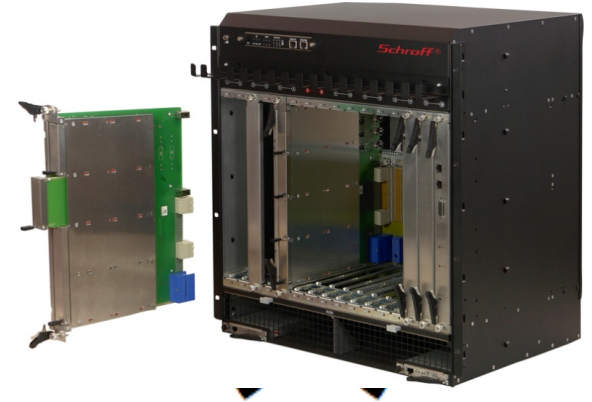


Figure 7: How to share the data inside of the Data Formatter system. The Data Formatter system consists of 4 ATCA shelves, 32 boards (8 boards/crate), and 64 FPGAs (2 FPGAs/board). The data are input to a FPGA and shared in Data Formatter system via (1) inter FPGA on board bus (purple lines), (2) ATCA fabric backplane (blue lines), and (3) inter-crate fibers (orange lines). Eventually the data will be sent to downstream for main track finding and fitting processes.

Extensive bandwidth requirements
study done using real beam data
with optimized (detailed) cabling
assignments (no IBL yet)



Shelf-board-FPGA-Trigger Tower assignments

Shelf1	Phi Start		0					
	Board0	Board1	Board2	Board3	Board4	Board5	Board6	Board7
FPGA0	Phi0 C-Endcap	Phi0 C-Barrel	Phi0 A-Barrel	Phi0 A-Endcap	Phi2 C-Endcap	Phi2 C-Barrel	Phi2 A-Barrel	Phi2 A-Endcap
FPGA1	Phi1 C-Endcap	Phi1 C-Barrel	Phi1 A-Barrel	Phi1 A-Endcap	Phi3 C-Endcap	Phi3 C-Barrel	Phi3 A-Barrel	Phi3 A-Endcap

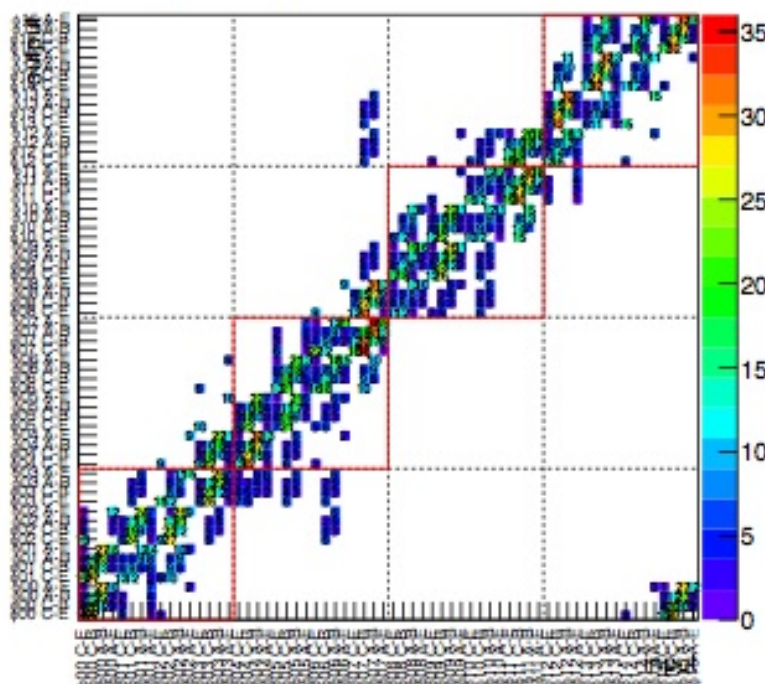
Shelf2	Phi Start		4					
	Board0	Board1	Board2	Board3	Board4	Board5	Board6	Board7
FPGA0	Phi4 C-Endcap	Phi4 C-Barrel	Phi4 A-Barrel	Phi4 A-Endcap	Phi6 C-Endcap	Phi6 C-Barrel	Phi6 A-Barrel	Phi6 A-Endcap
FPGA1	Phi5 C-Endcap	Phi5 C-Barrel	Phi5 A-Barrel	Phi5 A-Endcap	Phi7 C-Endcap	Phi7 C-Barrel	Phi7 A-Barrel	Phi7 A-Endcap

Shelf3	Phi Start		8					
	Board0	Board1	Board2	Board3	Board4	Board5	Board6	Board7
FPGA0	Phi8 C-Endcap	Phi8 C-Barrel	Phi8 A-Barrel	Phi8 A-Endcap	Phi10 C-Endcap	Phi10 C-Barrel	Phi10 A-Barrel	Phi10 A-Endcap
FPGA1	Phi9 C-Endcap	Phi9 C-Barrel	Phi9 A-Barrel	Phi9 A-Endcap	Phi11 C-Endcap	Phi11 C-Barrel	Phi11 A-Barrel	Phi11 A-Endcap

Shelf4	Phi Start		12					
	Board0	Board1	Board2	Board3	Board4	Board5	Board6	Board7
FPGA0	Phi12 C-Endcap	Phi12 C-Barrel	Phi12 A-Barrel	Phi12 A-Endcap	Phi14 C-Endcap	Phi14 C-Barrel	Phi14 A-Barrel	Phi14 A-Endcap
FPGA1	Phi13 C-Endcap	Phi13 C-Barrel	Phi13 A-Barrel	Phi13 A-Endcap	Phi15 C-Endcap	Phi15 C-Barrel	Phi15 A-Barrel	Phi15 A-Endcap

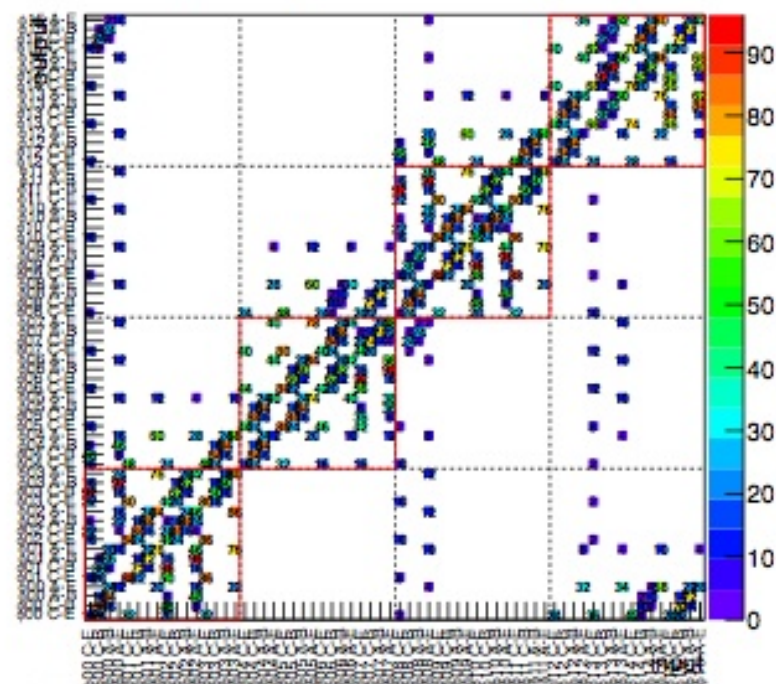
Data Sharing

Input Output of DF boards (Pixel)



(a)

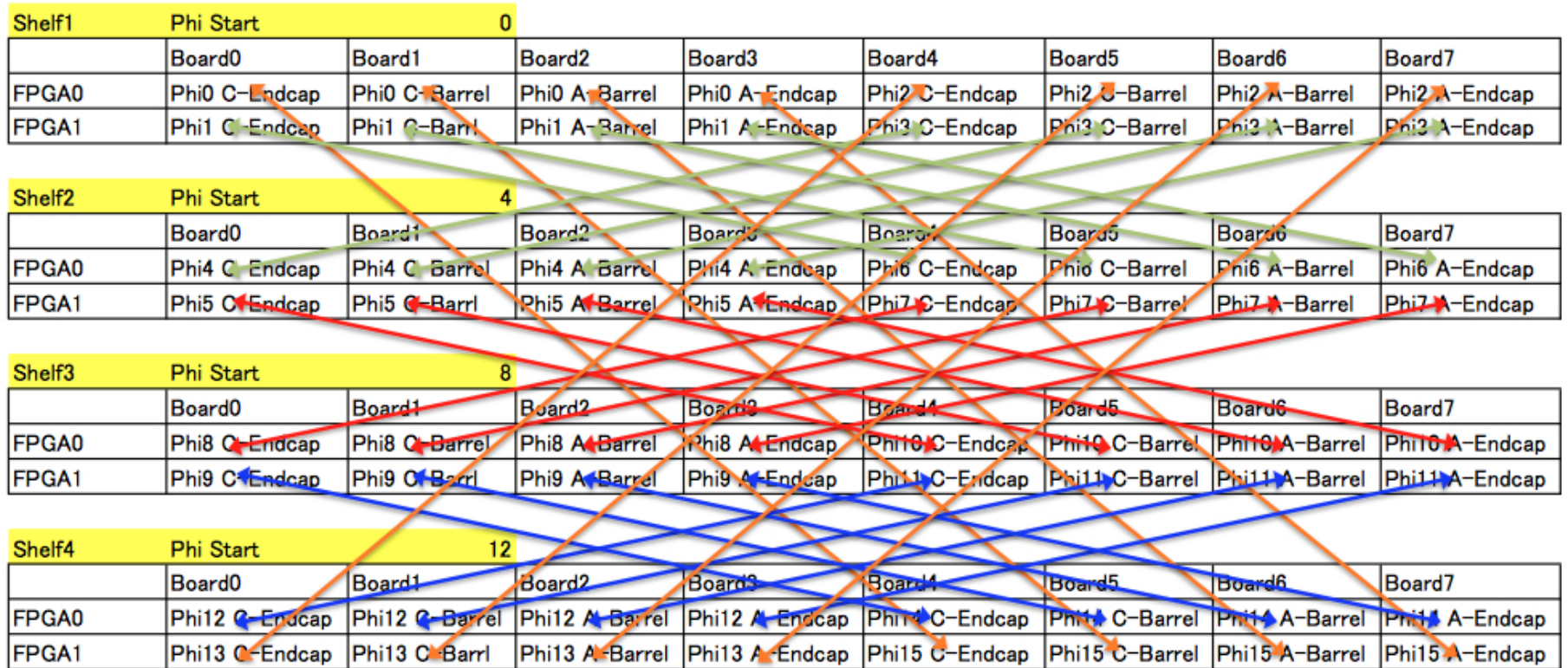
Input Output of DF boards (SCT)



(b)

Input - Output

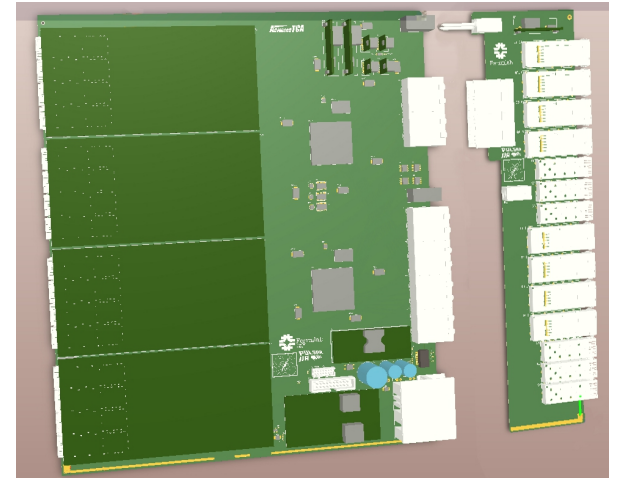
Inter-crate communications (a fraction of the connections needed for FTK)



Data Volume Analysis Summary

(by Yasu Okumura)

- Analysis uses real beam data
 - 7 TeV with $\langle\mu\rangle \sim 10$
 - Extrapolated to 14 TeV and $\langle\mu\rangle \sim 80$
- Analyzed data volume on
 - Output to AUX
 - Output to SSB
 - Inter-FPGA local bus
 - ATCA Backplane Fabric Interface channels
 - Inter-Shelf fiber links
- Worse case shown below (conservative, assume no clustering in pixel)

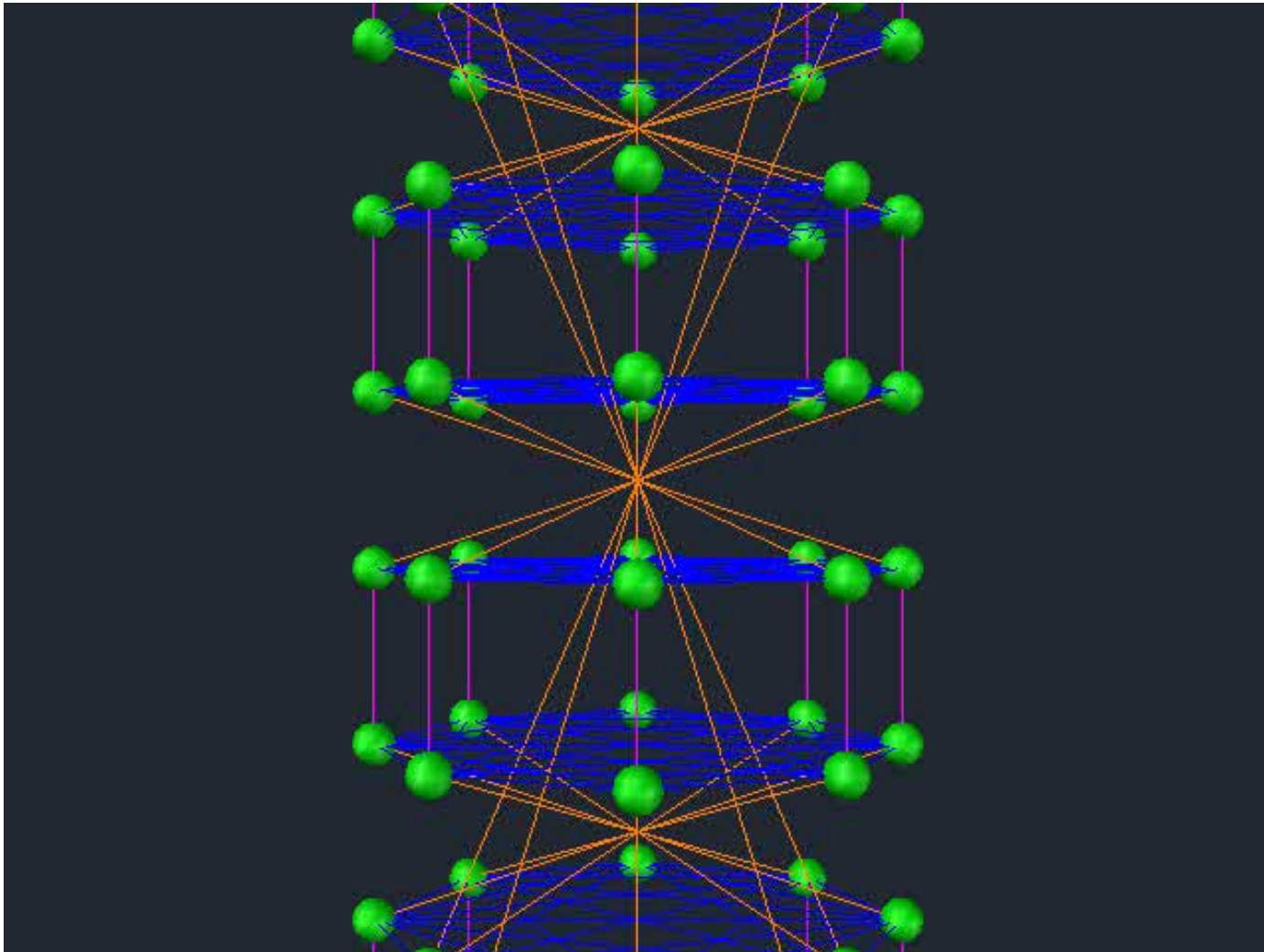


	7 TeV $\langle\mu\rangle = 10$	14 TeV $\langle\mu\rangle = 80$
AUX card	2.9	23.2
SSB	0.6	5.1
Inter-FPGA	2.4	19.2
ATCA Fabric	0.6	5.1
Inter-Shelf	0.7	5.8

Units are Gbps

Data Formatter FPGA Constellation for FTK

Animation by Jamieson Olsen



Routing Firmware

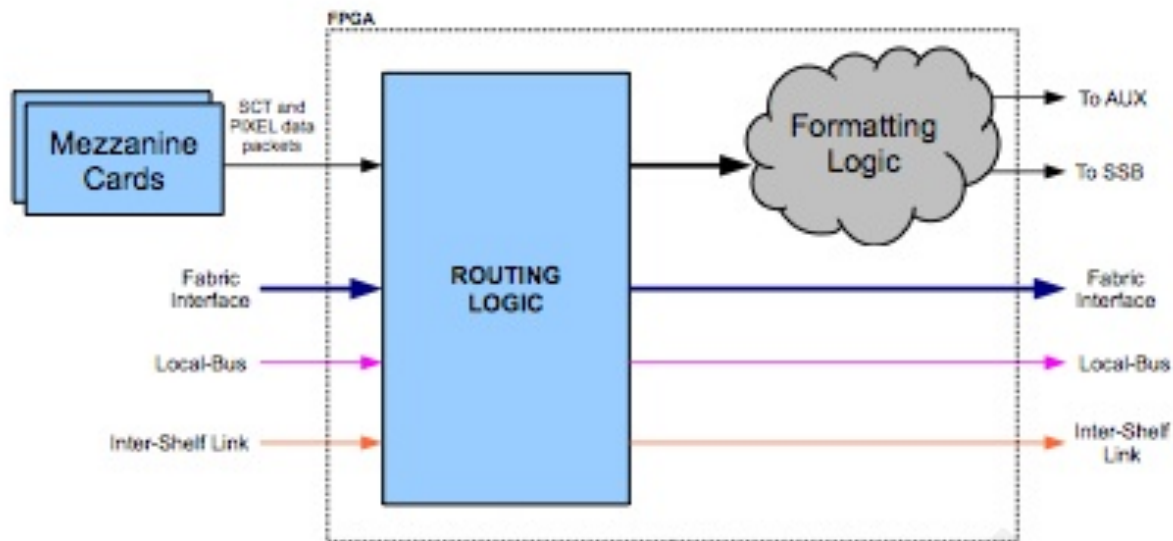
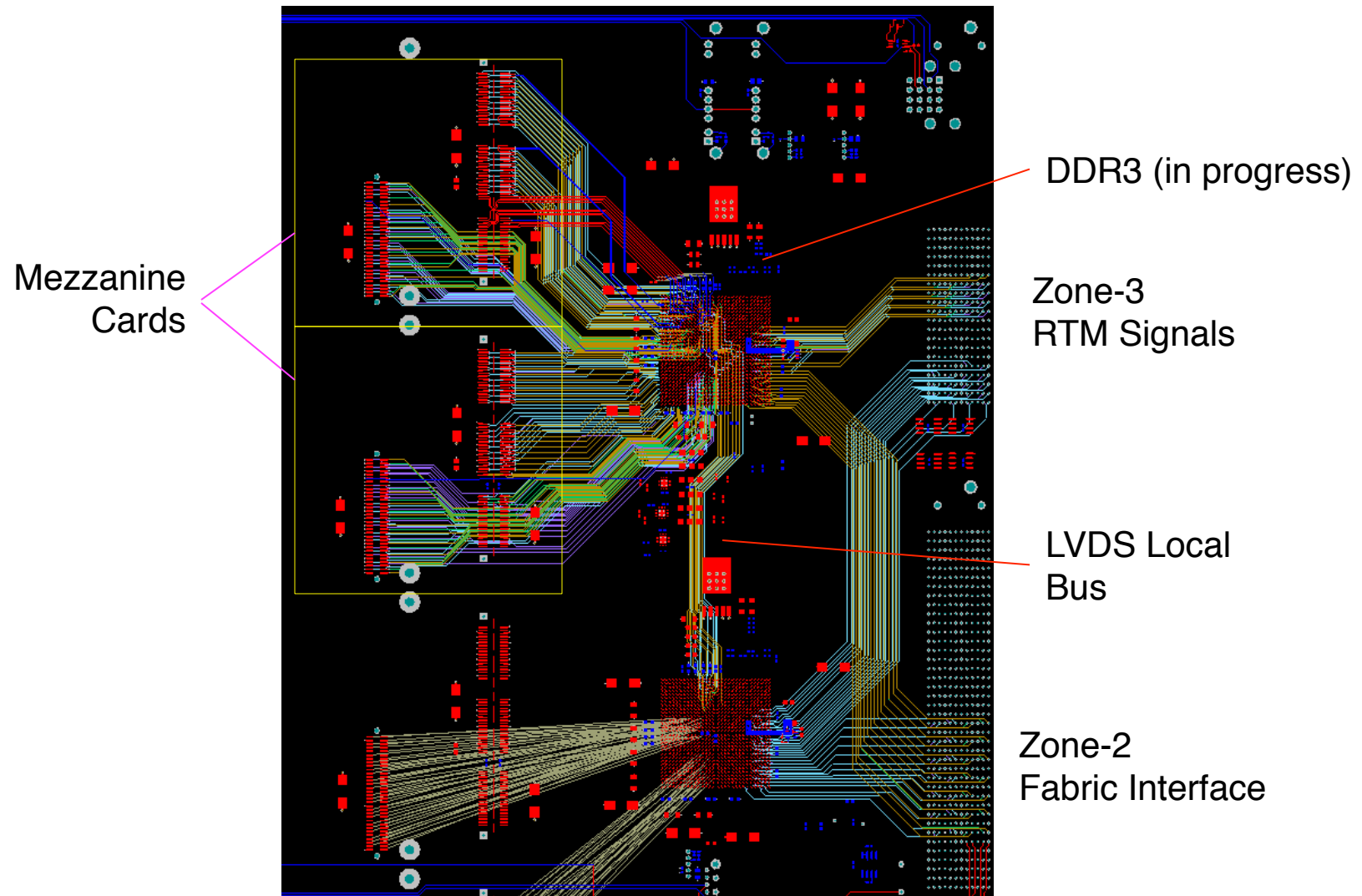
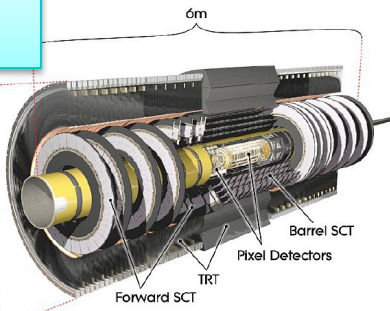
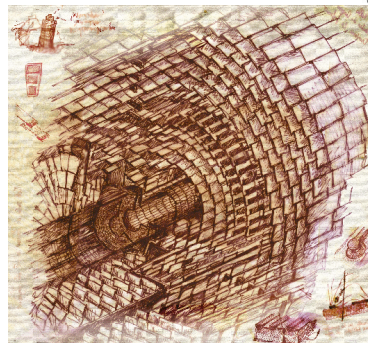
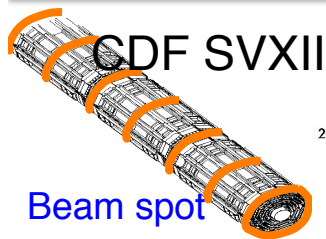


Figure 31: Overview of the FPGA routing firmware. Data packets containing SCT and Pixel hits and clusters are output from a pair of mezzanine cards. Data packets also arrive on the Fabric Interface channels, the local-bus link, and inter-shelf link. The firmware supports “route through” which enables an incoming packet to be retransmitted over any output.

Prototype Board layout in progress (FNAL engineer: Jamieson Olsen)



Detector design for triggering



Data transfer

Data
formatting

Tracking Trigger Issues at hadron colliders (L1 & L2/HLT)

Pattern
Recognition

Associative Memory
Others ...

Track
Fitting

FPGA vs CPU vs GPU

HLT

• Selected R&D topics

- (1) Data Reduction at detector/sensor stage
- (2) Data Transfer challenges
- (3) Data Formatter using ATCA technology
- (4) Pattern Recognition: 3D AM R&D: VIPRAM
= "Vertically Integrated Pattern Recognition Associative Memory"
- (5) Track Fitting: GPU study for low latency tracking

CDF original SVT system had ~400K patterns total ...
128 patterns per AMchip -- commissioned around ~2001.



Question: Can we put the entire SVT system into one chip?
~400K patterns per chip → few x 1000 chips → Billion patterns ...

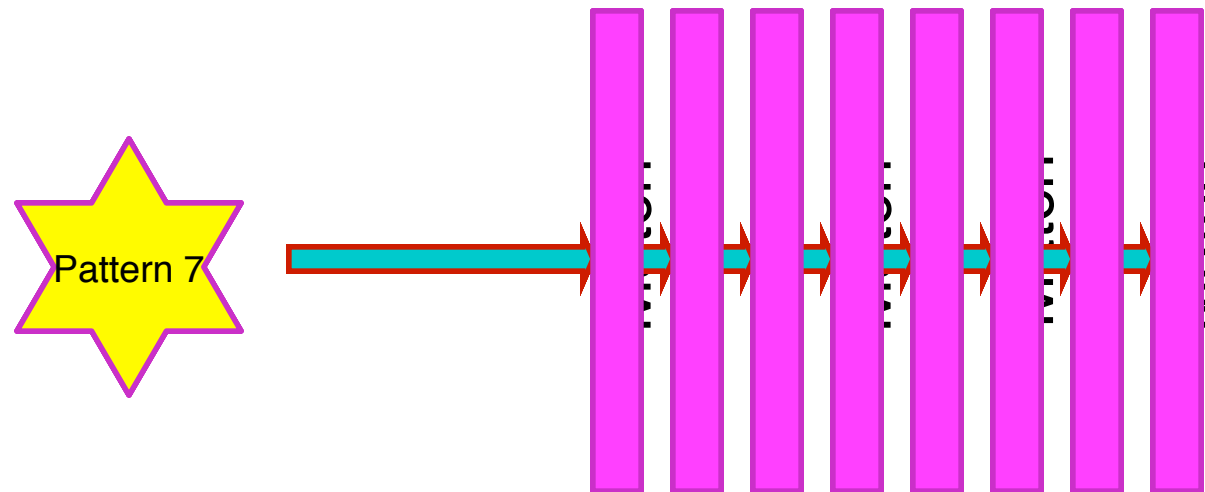
RAM vs CAM

- RAM (Random-Access-Memory): user supplies a memory address and it returns the data word stored at that address
- CAM (Content-Addressable-Memory): user supplies a data word and it searches its entire memory to see if that data word is stored anywhere in it. In other words, it is accessed by virtue of its contents, not its location.
 - If the data word is found/matched, the CAM returns the storage addresses where the word was found
 - search its entire memory in a single operation, much faster than RAM
- In essence, CAM ~ “Inverse RAM”
- AM or PRAM: Pattern Recognition Associative Memory
 - CAM based

Content Addressable Memory (CAM)

- CAM: inverse of RAM

- user supplies a data word and it searches its *entire* memory *in a single operation* to see if that data word is stored anywhere in it



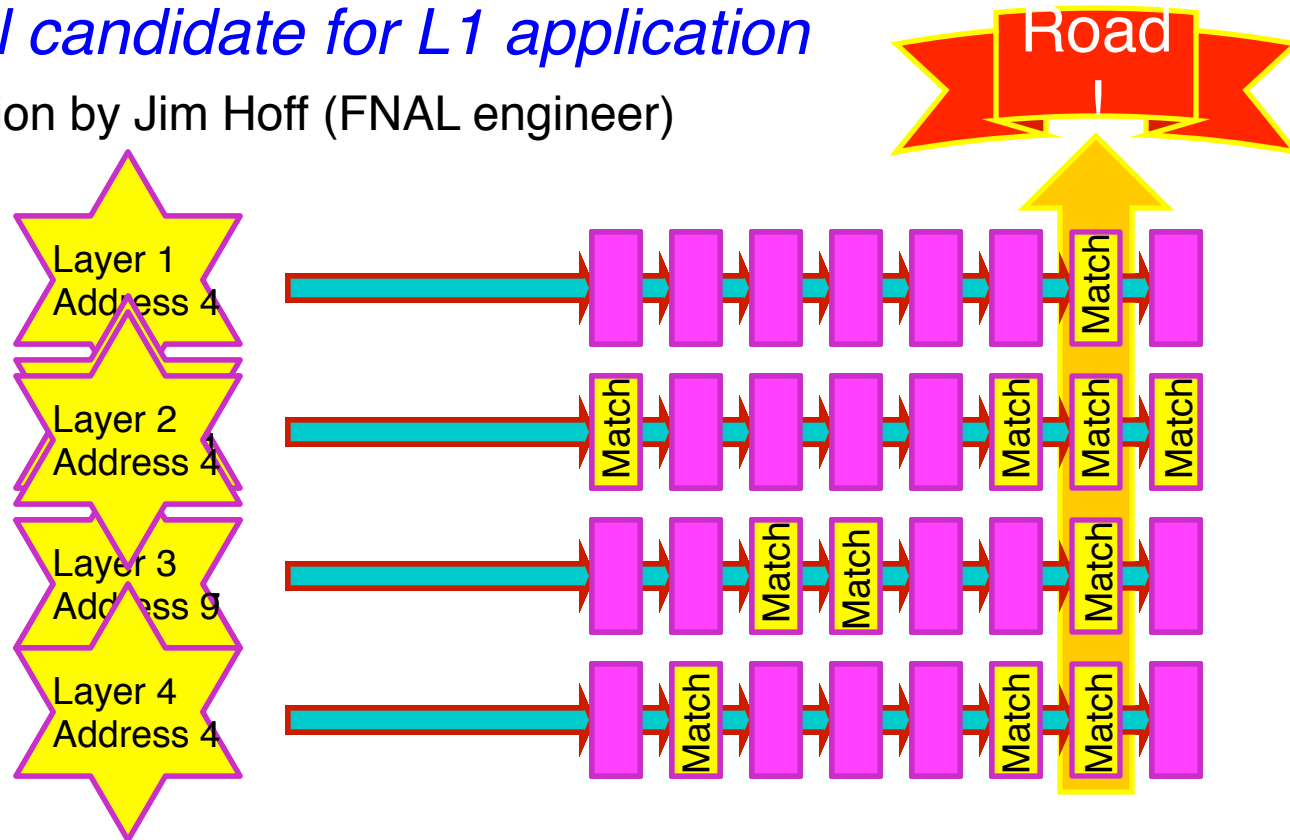
- One incoming pattern/hit at a time
- There is no memory of previous matches

How PRAM works

- Pattern Recognition Associative Memory (PRAM)

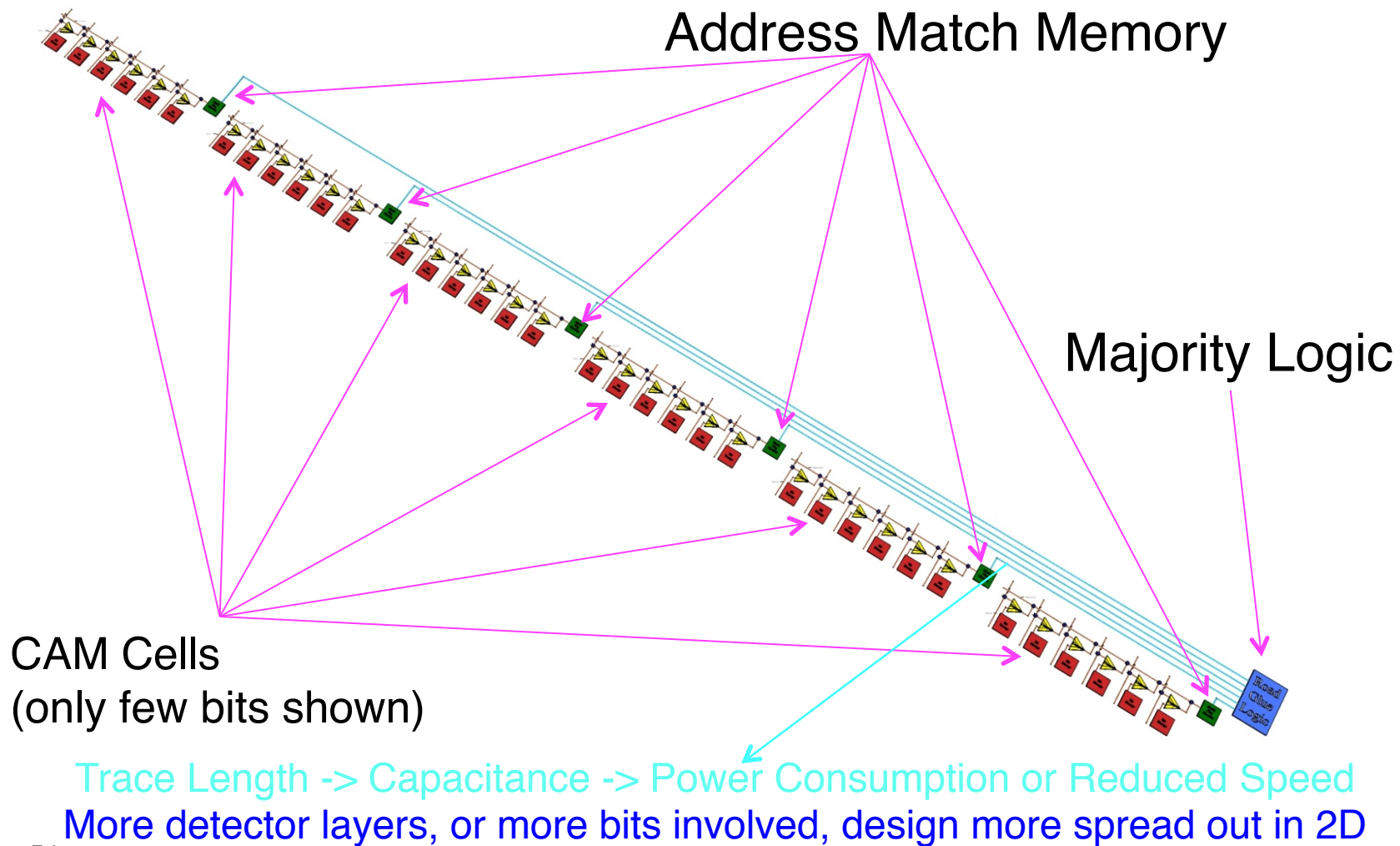
- *Pattern recognition finishes as soon as all hits arrive*
- *Potential candidate for L1 application*

Animation by Jim Hoff (FNAL engineer)



Anatomy of a PRAM

(Pattern Recognition Associative Memory)



Comments on 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
- Critical figures of merit for an AM based system:
(higher) pattern density & speed and (lower) power density
 - However, at chip level, more detector layers means more CAM cells are needed for a given pattern, the layout are more spread out in two dimensions (for a given technology node) resulting in decreasing pattern density and increasing driving load capacitance or power consumption, which in turn reduces the maximal speed of operation
 - Performance fundamentally limited by Moore's Law
- *This is the main limitation of an otherwise very powerful and proven approach for its future applications within and beyond HEP.*

The Challenge of future AM design

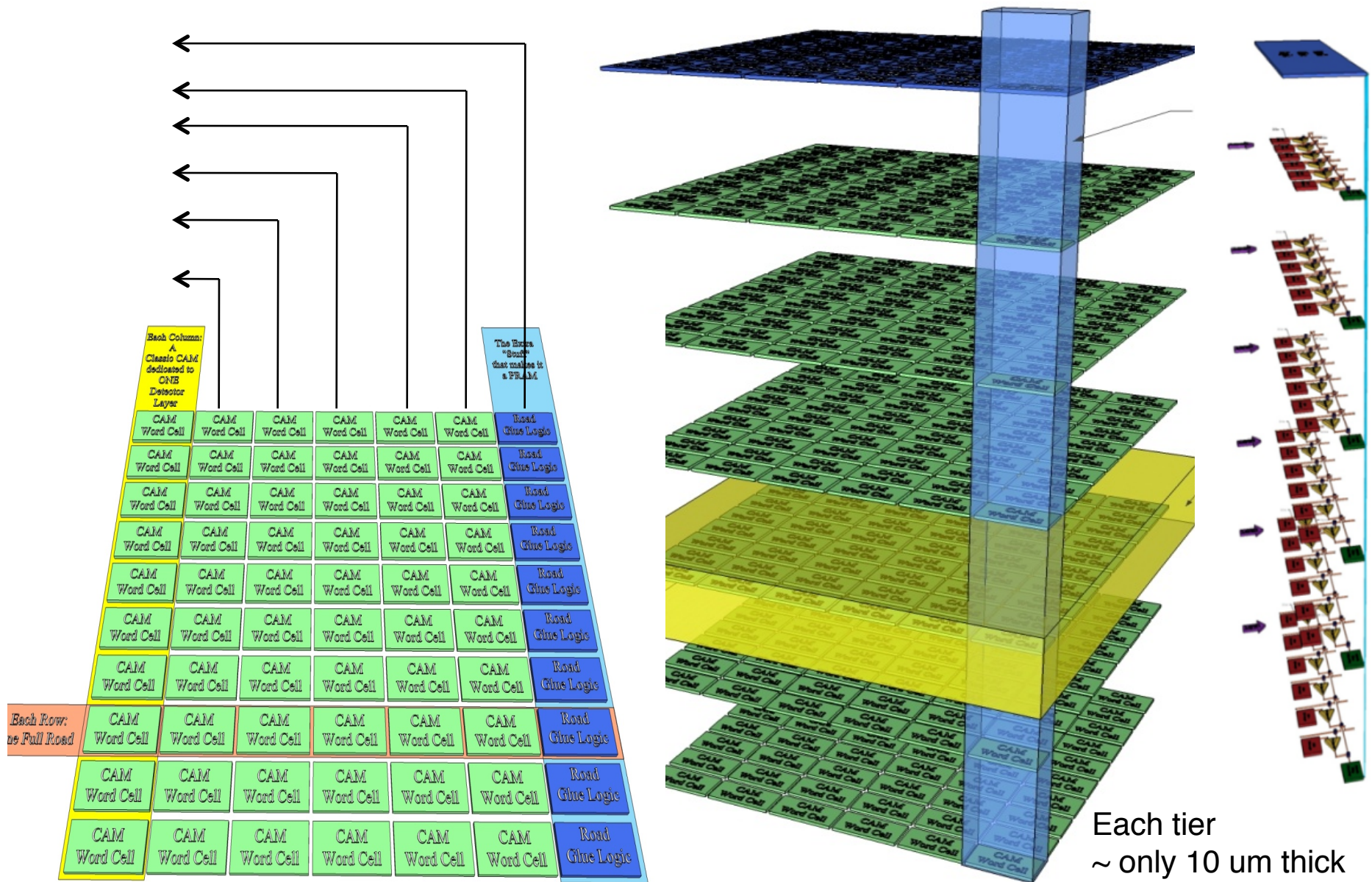
Increase the patterns density by 2 orders of magnitude;
and
increase the speed by a factor of $>\sim 3$,
while
keeping the power consumption more or less the same

*Much higher Patten Density & higher Speed
Yet much less Power Density
almost too good to be true*

*New idea: could go to “extra dimension” to achieve this
→ generic R&D effort at Fermilab*

From 2D to 3D

VIPRAM (Vertically Integrated Pattern Recognition Associative Memory)

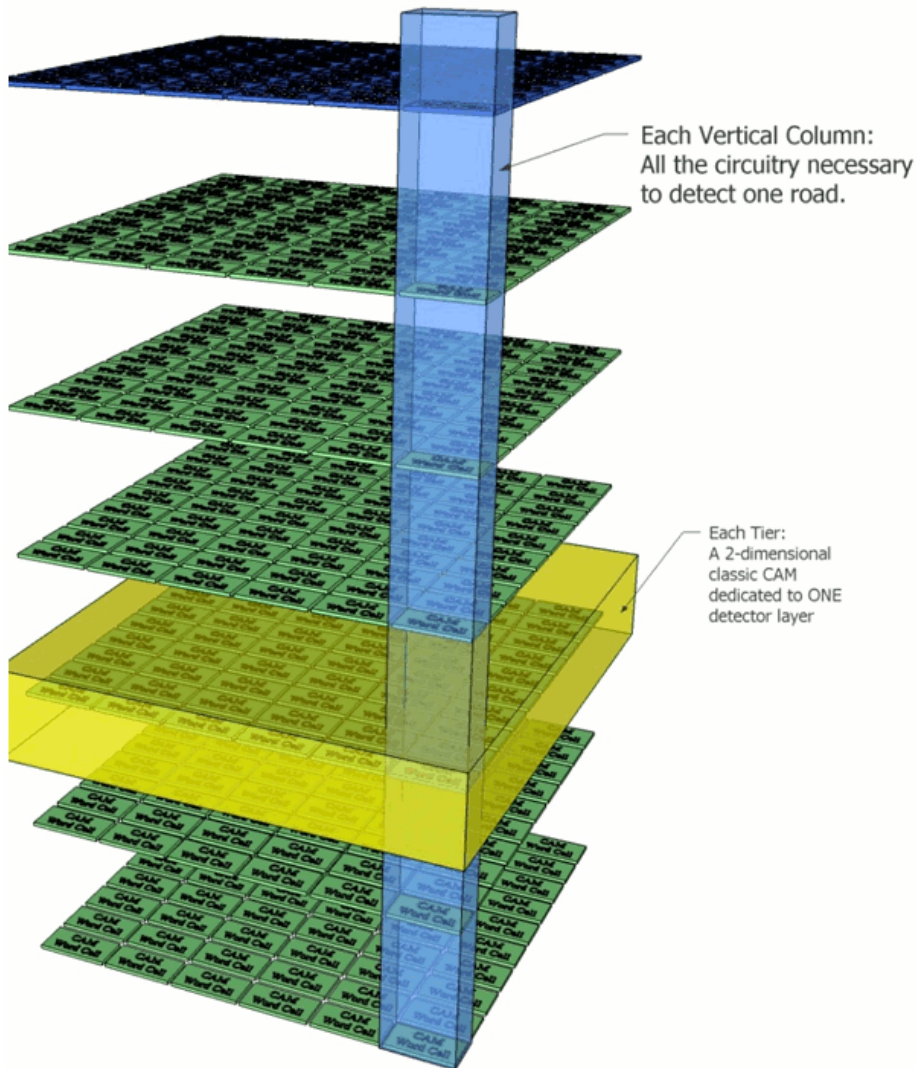


VIPRAM

(Vertically Integrated Pattern Recognition Associative Memory)

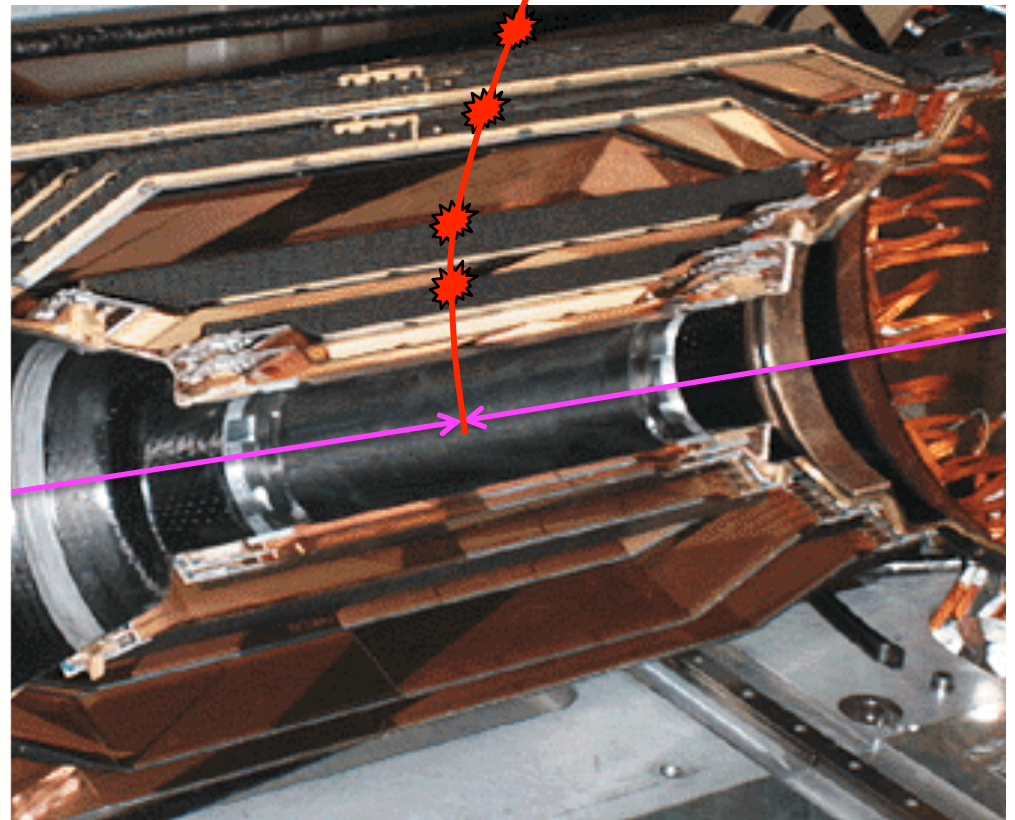
http://hep.uchicago.edu/~thliu/projects/VIPRAM/TIPP2011_VIPRAM_Paper.V11.preprint.pdf

fired road



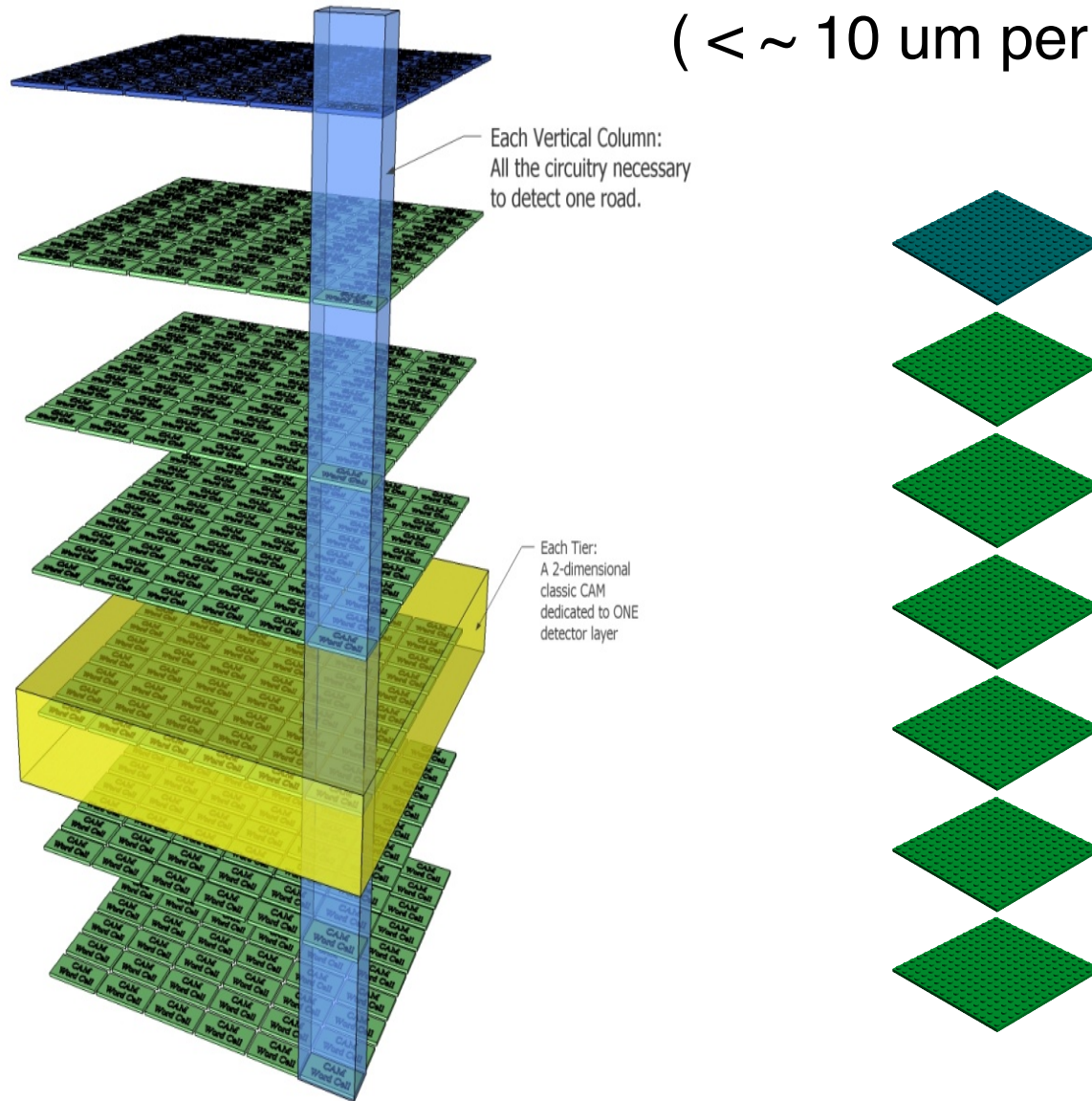
Pattern recognition for tracking
is naturally a task in 3D

track



ger R&D - Ted Liu

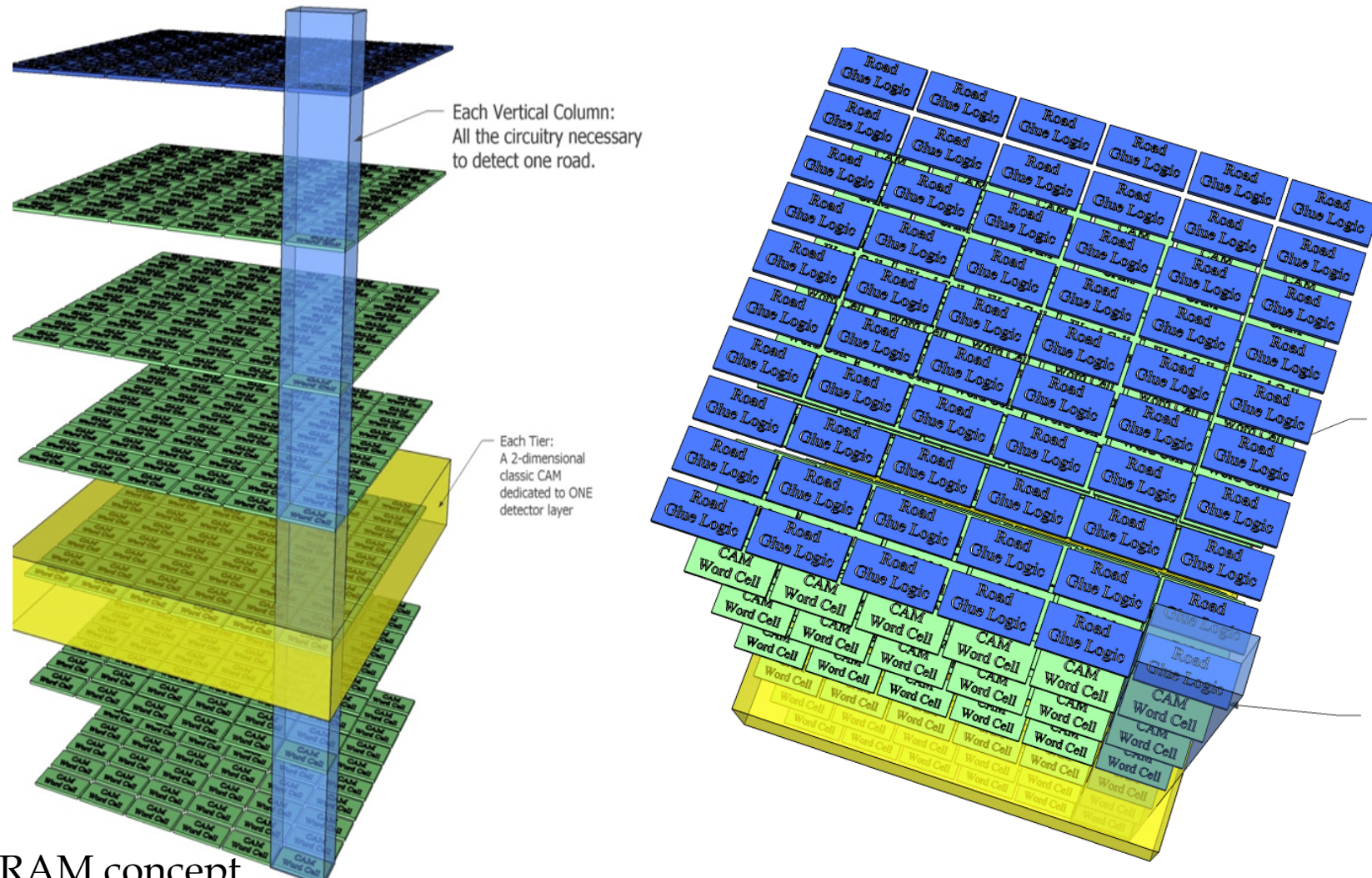
The 3D stack is actually very thin
($< \sim 10$ μm per tier)



Side view

Top view

R&D for future track trigger applications (L1/L2):

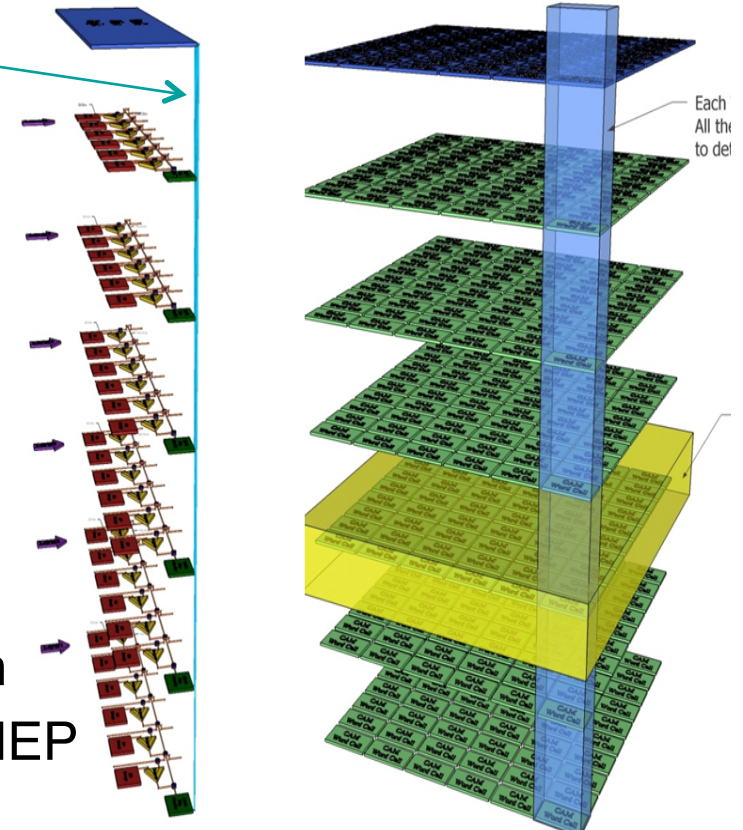


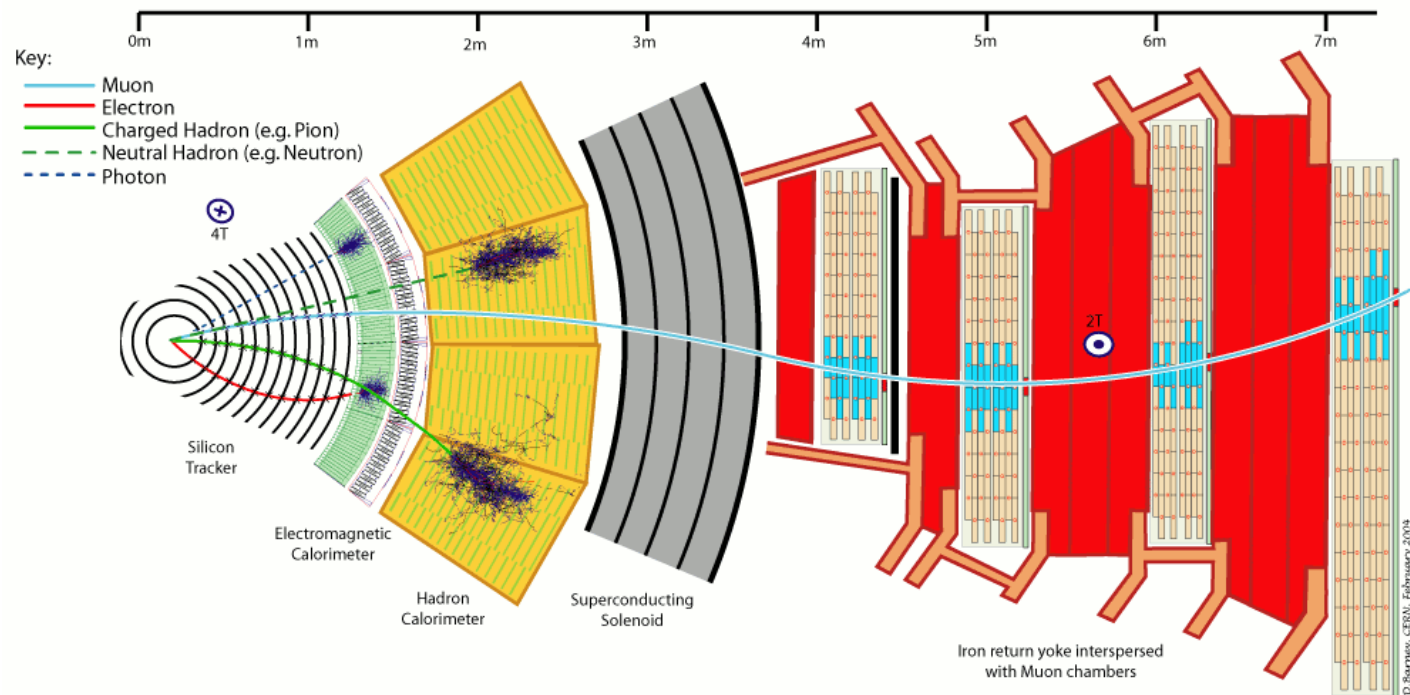
VIPRAM concept

- In 3D, with 130 nm, VIPRAM could reach $\sim 200\text{K patterns/cm}^2$
- In 2D, even with 65 nm, could only reach $\sim 50\text{K patterns/cm}^2$ (AMchip04)

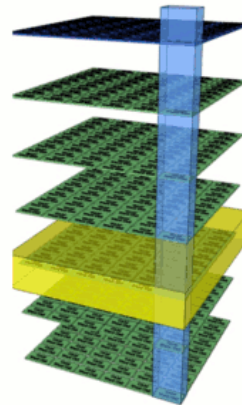
VIPRAM is almost an ideal case for the application of 3D vertical integration technology

- A VIPRAM cell can process n layers of a road pattern in about the size of just one CAM cell (*pattern density increased by $\sim n$*)
- Directly shortens the longest of the driving lines in the pattern recognition cell (*address match lines*).
(*reduced power density or higher speed*)
- Layout of the CAM cells, Majority Logic cells, input/output busses simpler, more efficient. More freedom in layout.
- Could add CAL/Muon info to extra tiers, to ID electron/muon objects directly (L1)
- VIPRAM architecture is inherently open and flexible, making possible the design of more general purpose fast pattern recognition devices far beyond the original AM used for HEP

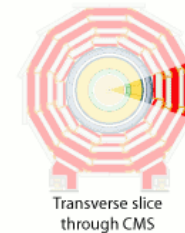




Animation by
high school
summer intern:
Tom Klonowski
62



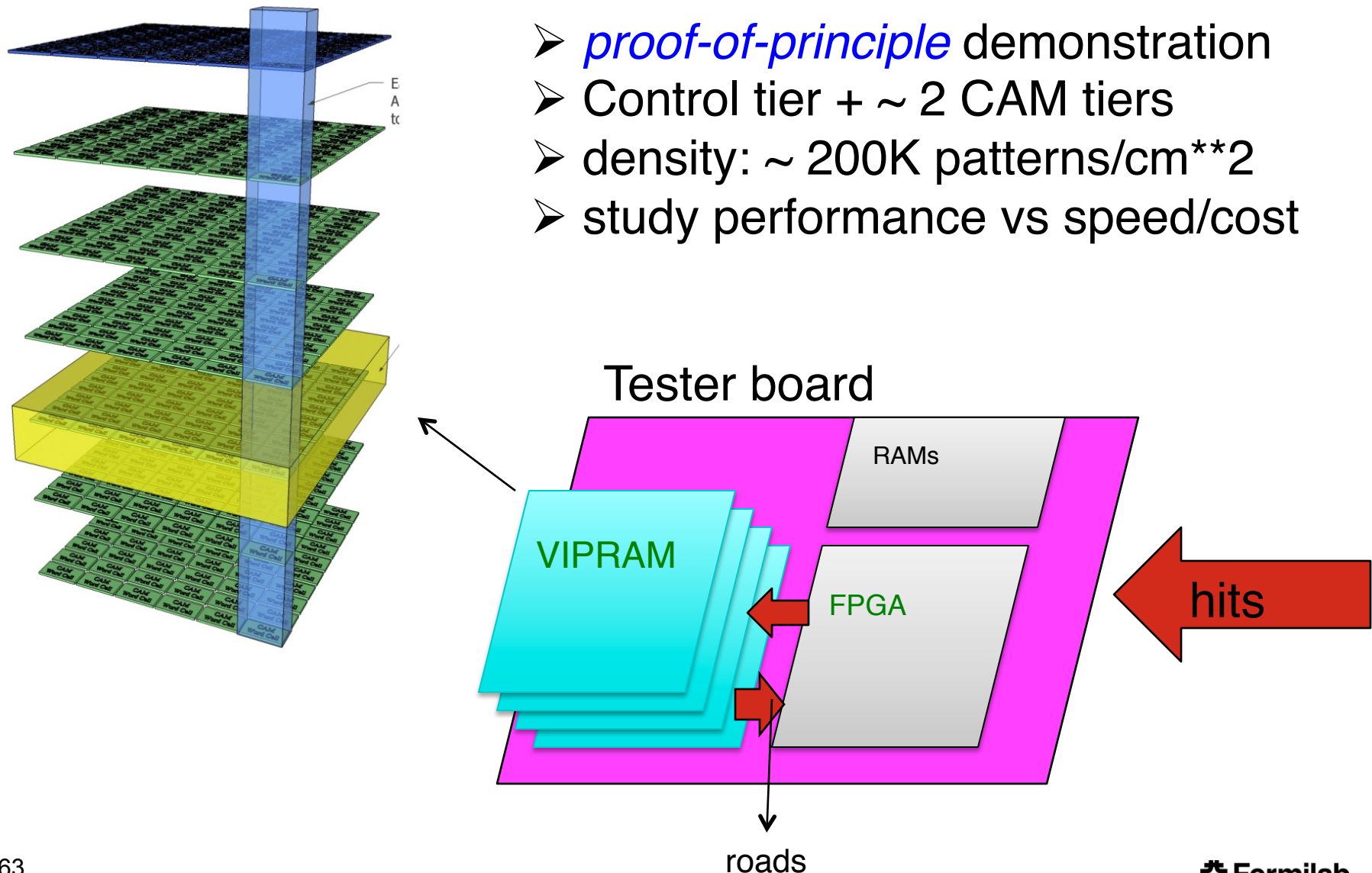
Potentially useful to directly
identify physics objects at L1,
Many advantages this way.



Fermilab

Initial Goal of R&D

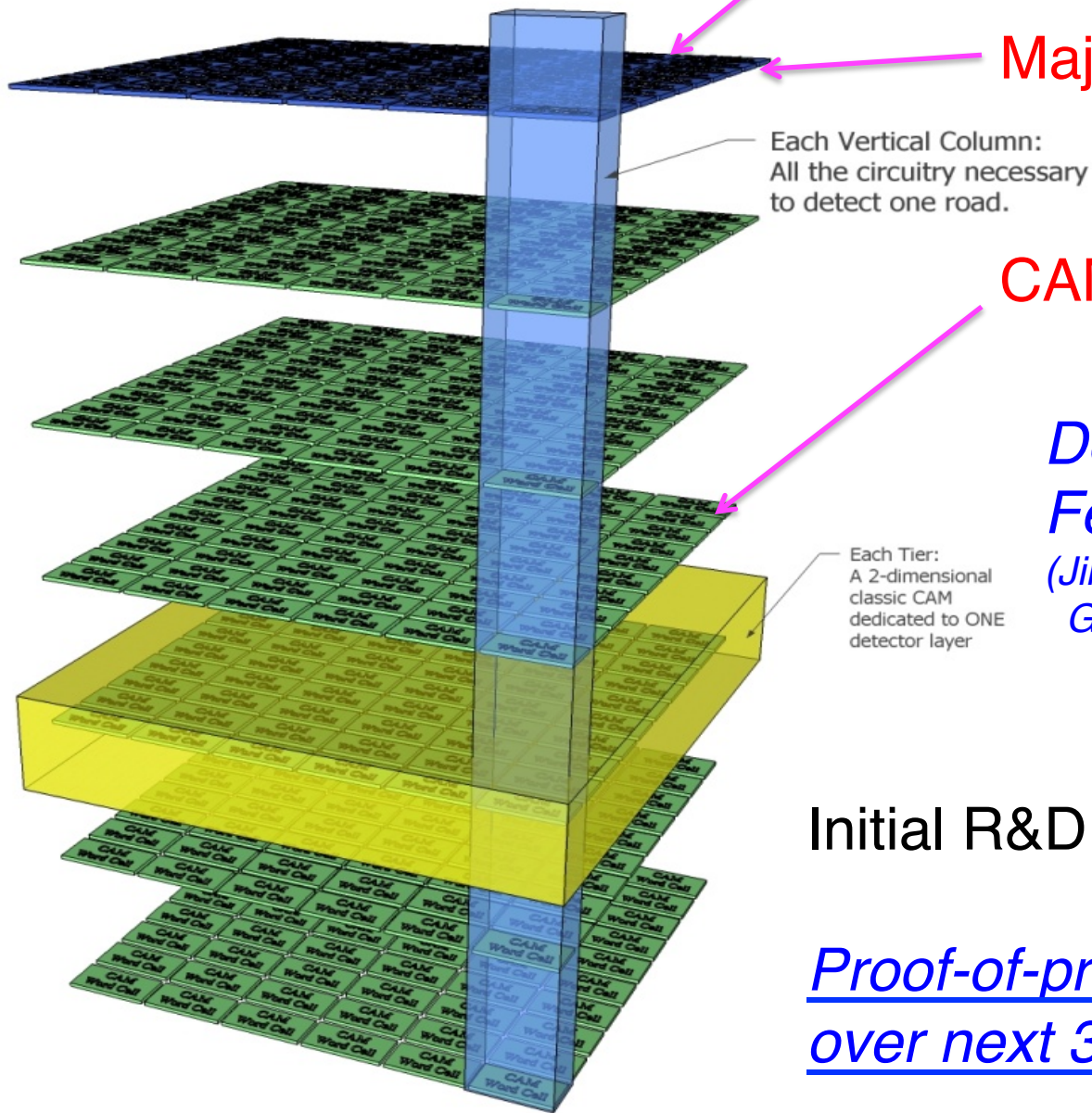
- *proof-of-principle* demonstration
- Control tier + ~ 2 CAM tiers
- density: $\sim 200\text{K patterns/cm}^2$
- study performance vs speed/cost



Basic Assumptions for R&D

- 130nm Global Foundries CMOS
- Tezzaron's 3D process
- 18 bits in the CAM cell (like AMchip04)
- design for up to 8 detector layers (like AMchip04)
 - ↘ This can be achieved with either 4 or 8 CAM tiers
 - ↘ Only stack 1 Control + ~2 CAM tiers for proof-of-principle
- 4 μm center-to-center TSV spacing for compatibility with current Tezzaron's 3D process.
- If a PRAM cell size can be about $\sim 20\mu\text{m} \times \sim 20\mu\text{m}$
 - ➔ expect up to $\sim 250\text{K}$ patterns per cm^2 (in 130 nm).
aim for $\sim 200\text{K}$ for proof-of-principle

Design Work involved



Control/interface/readout design

Majority Logic cell design

CAM cell design

*Design work by
Fermilab ASIC group
(Jim Hoff, Marcel Trimpl,
Gregory Deptuch)*

Initial R&D goal:

*Proof-of-principle demonstration
over next 3 years*

Testing 3D building blocks in 2D first

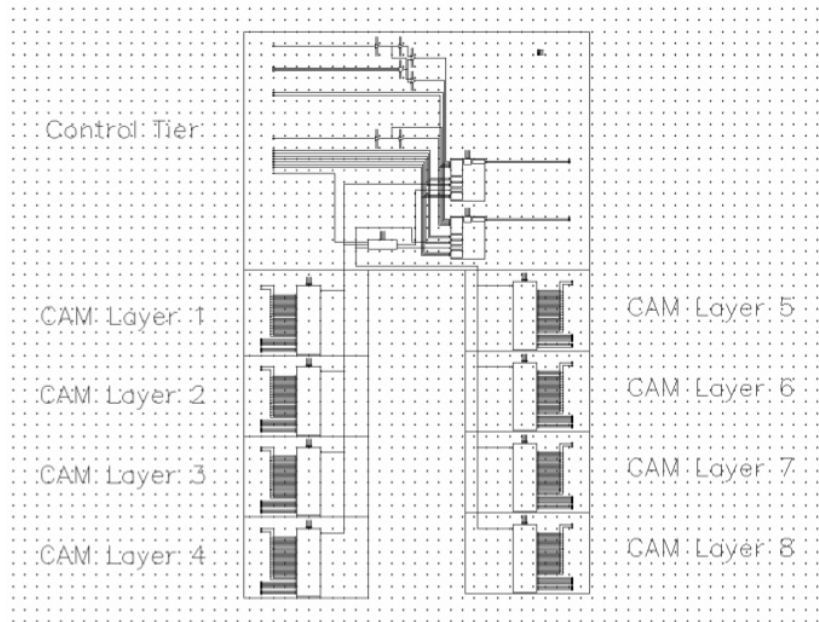


Figure 1 – protoPants

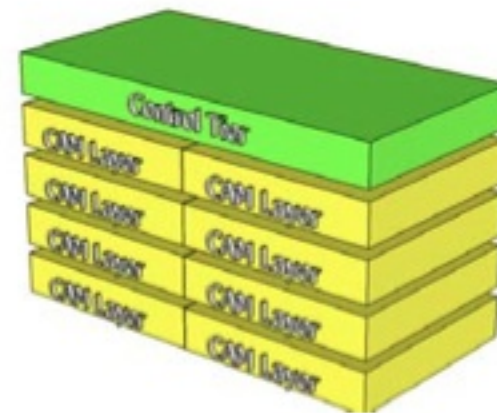
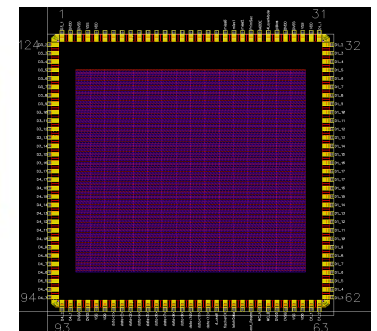


Figure 3 - A Block layout of the protoPants.

First implement in 2D for initial testing (work in progress)



Figure 3 – protoPants



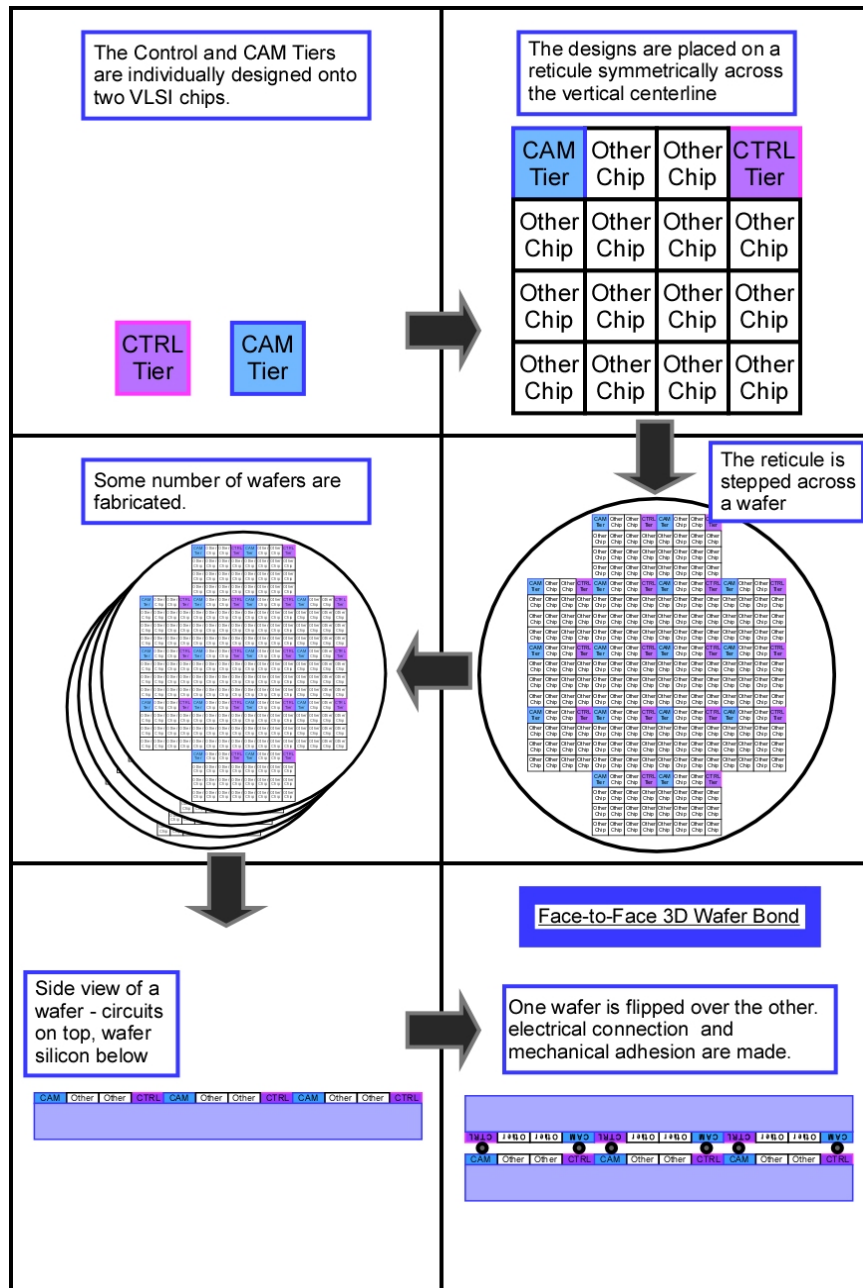


Figure 5 - A two-tier, Single Mask Set 3D MPW process

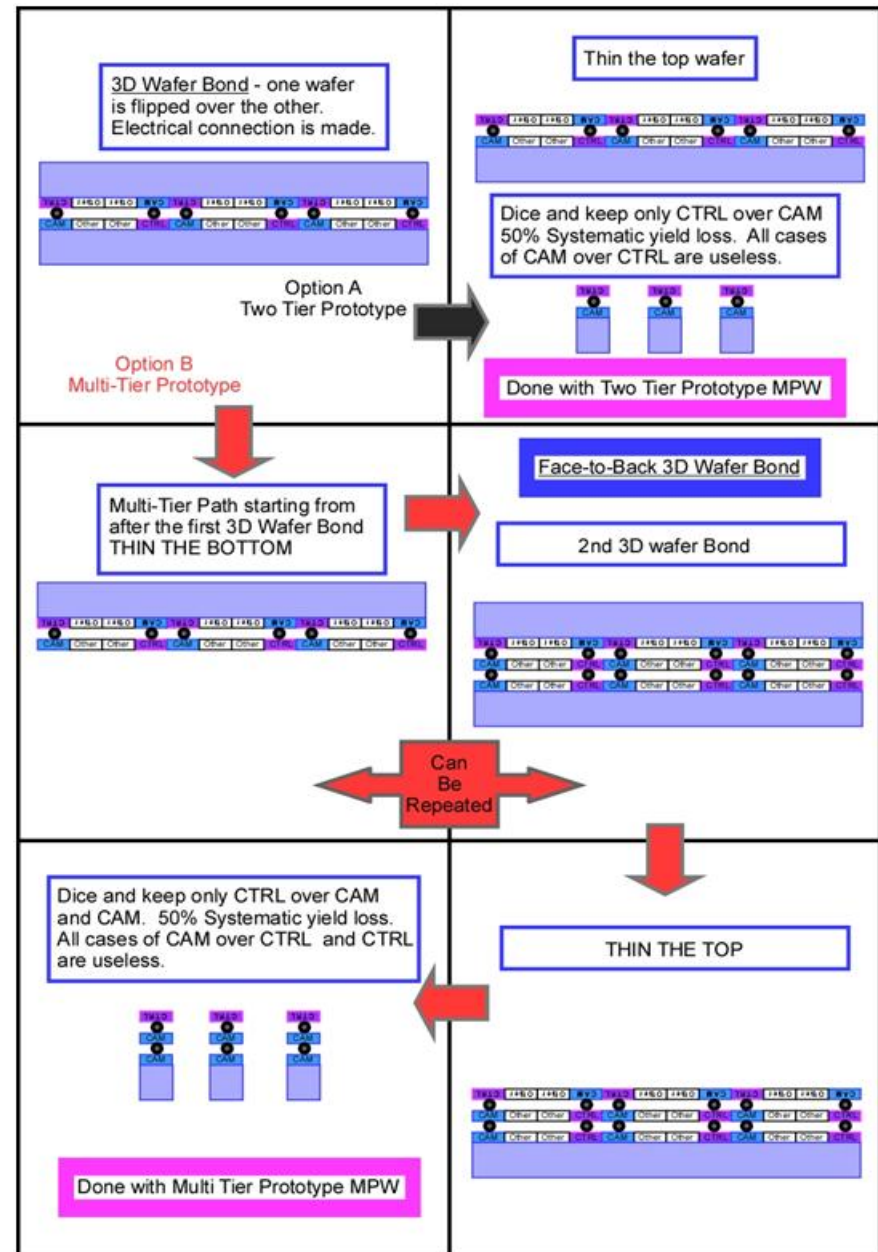
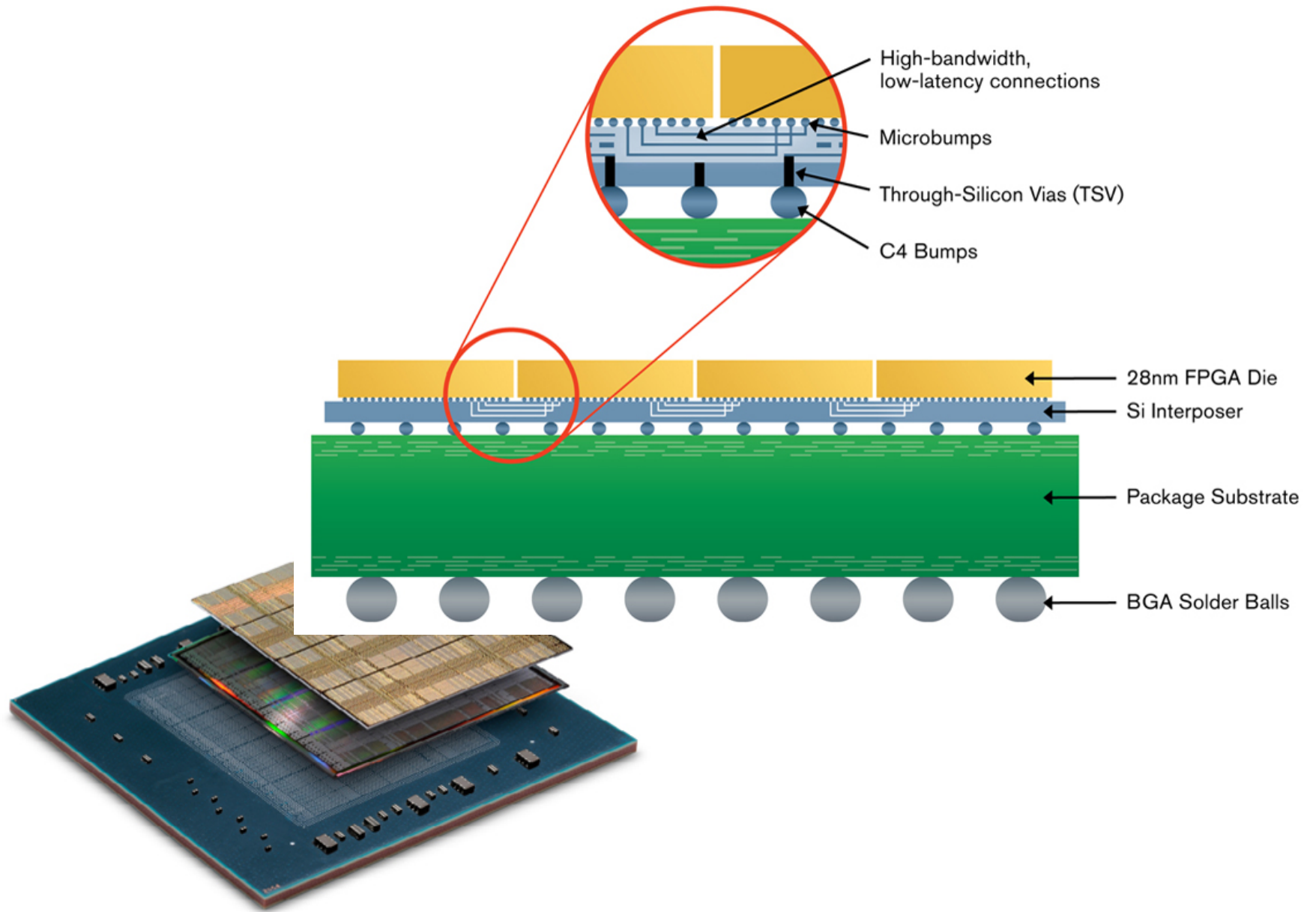


Figure 6 - The conclusion of a typical 3D MPW process OR an alternate process available to the VIPRAM.

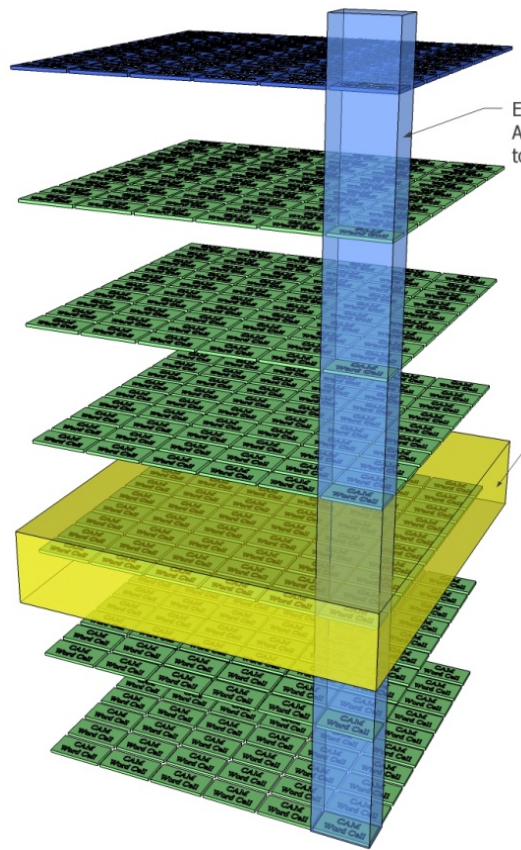
Integrate AM and TF/FPGA stages into one chip

- Bandwidth between AM stage and Track Fitting stage could be another challenge
 - needs to transfer large number of fired roads and associated full resolution hits into the TF stage
 - The larger the AM pattern size per chip, the more demand
 - *Highly desirable if the two stages can be integrated*
 - High speed serial I/O on FPGA can be used for input data IO
 - *Board & system level design could be much simplified*
- 3D Technology could help here (in the future)
 - Example: silicon interposer approach for Xilinx Virtex-7 FPGA
 - *Would make the chip much more flexible (within & outside HEP)*

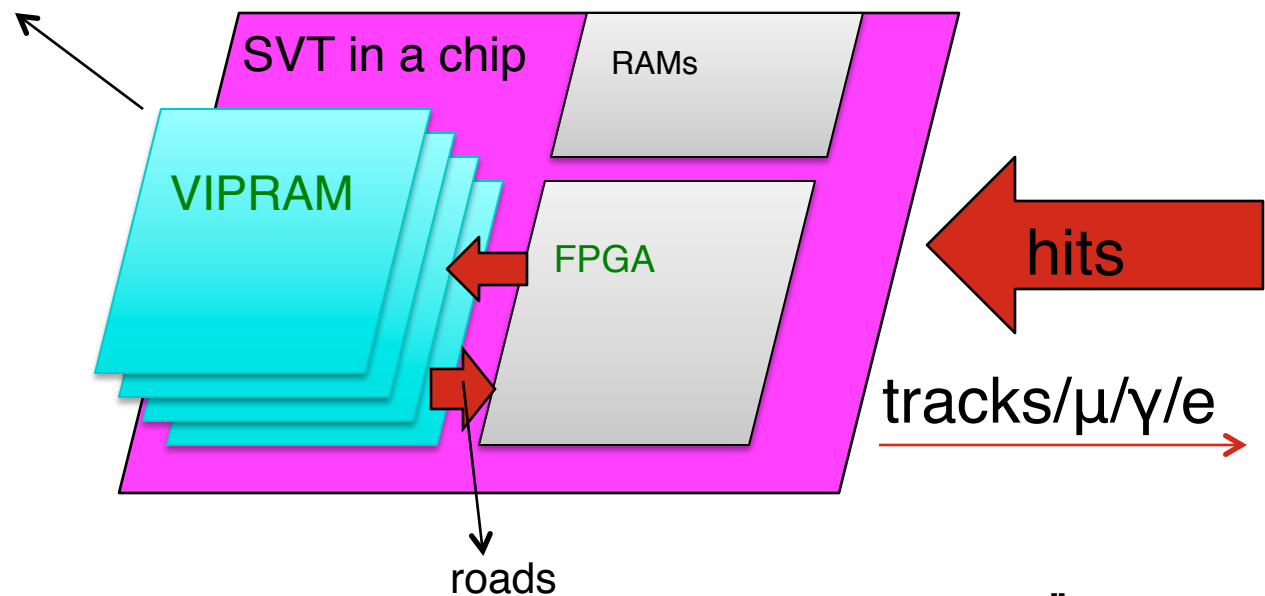


Virtex-7 2000T FPGA Utilizing
Stacked Silicon Interconnect Technology^{ed Liu}

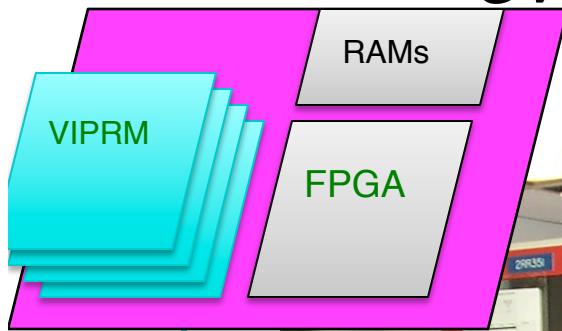
Long term goal of R&D



- $\sim 500\text{K patterns/cm}^{**2}$
- Running with $> 100\text{ MHz}$ input rate
- N CAM tiers + Control tier
- integrated with FPGA/RAM
(*general purpose pattern recognition*)



*SVT in one chip: 2nd phase of VIPRAM project
→ general purpose chip*



2 meters



Original SVT system had ~400K patterns total

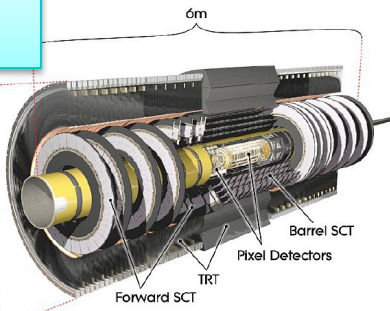
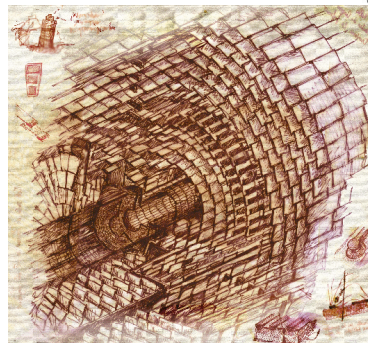
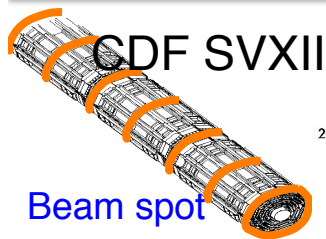
*Aim to reach ~500K per cm**2 for VIPRAM chip ...*



Comments on future tracking trigger applications of VIPRAM

- VIPRAM project is “Generic R&D” at this stage, and current focus is the “proof-of-principle”.
- It was motivated by FTK simulation studies and is based on AMchip design (the core), as such,
 - Should be useful for future L2-like applications
 - “SVT in one chip” approach could simplify board and system design
- For L1 tracking trigger
 - The *ultimate goal* is to design it for L1 application
 - *VIPRAM architecture is inherently flexible & open (highly desirable)*
 - need to work out system level design vs chip level
 - need extensive simulation studies with physics cases for guidance
 - *Inputs and collaboration are welcome ...*

Detector design for triggering



Data transfer

Data formatting

Tracking Trigger Issues at hadron colliders (L1 & L2/HLT)

Pattern Recognition

Associative Memory
Others ...

Track Fitting

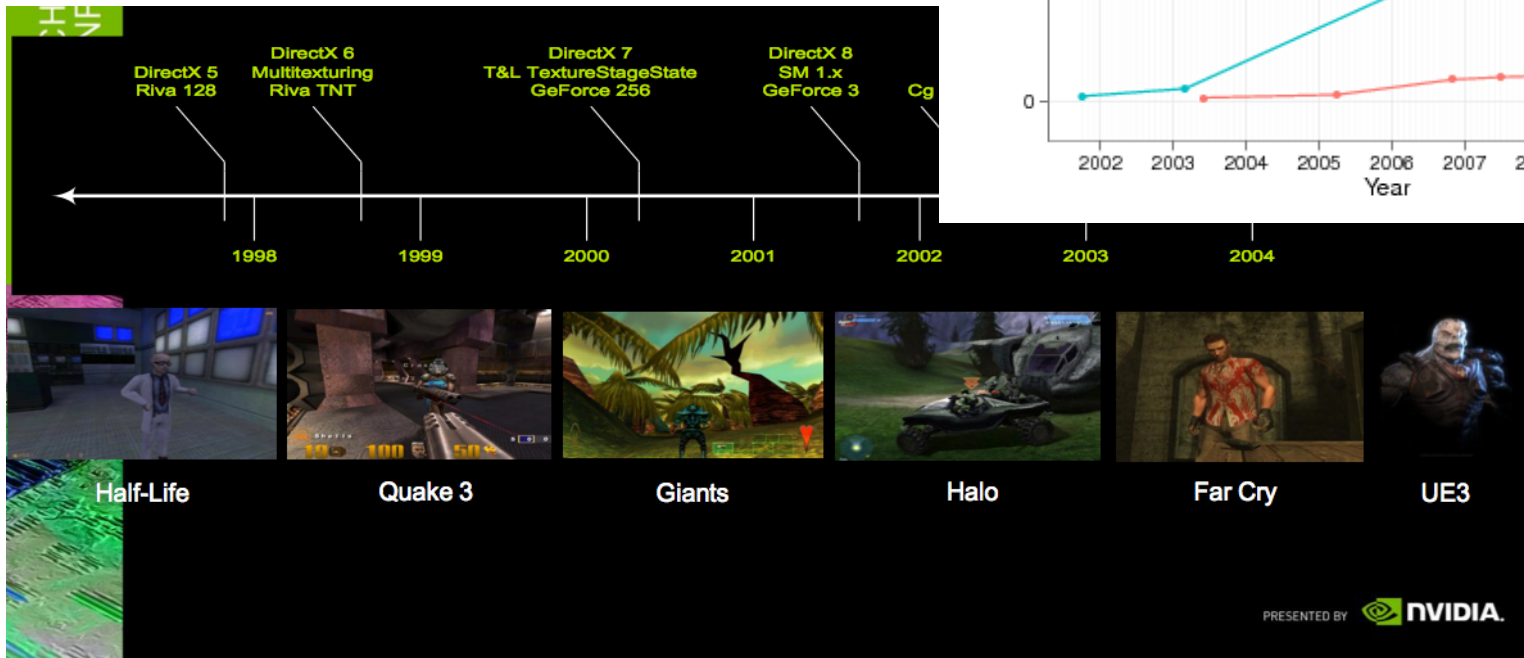
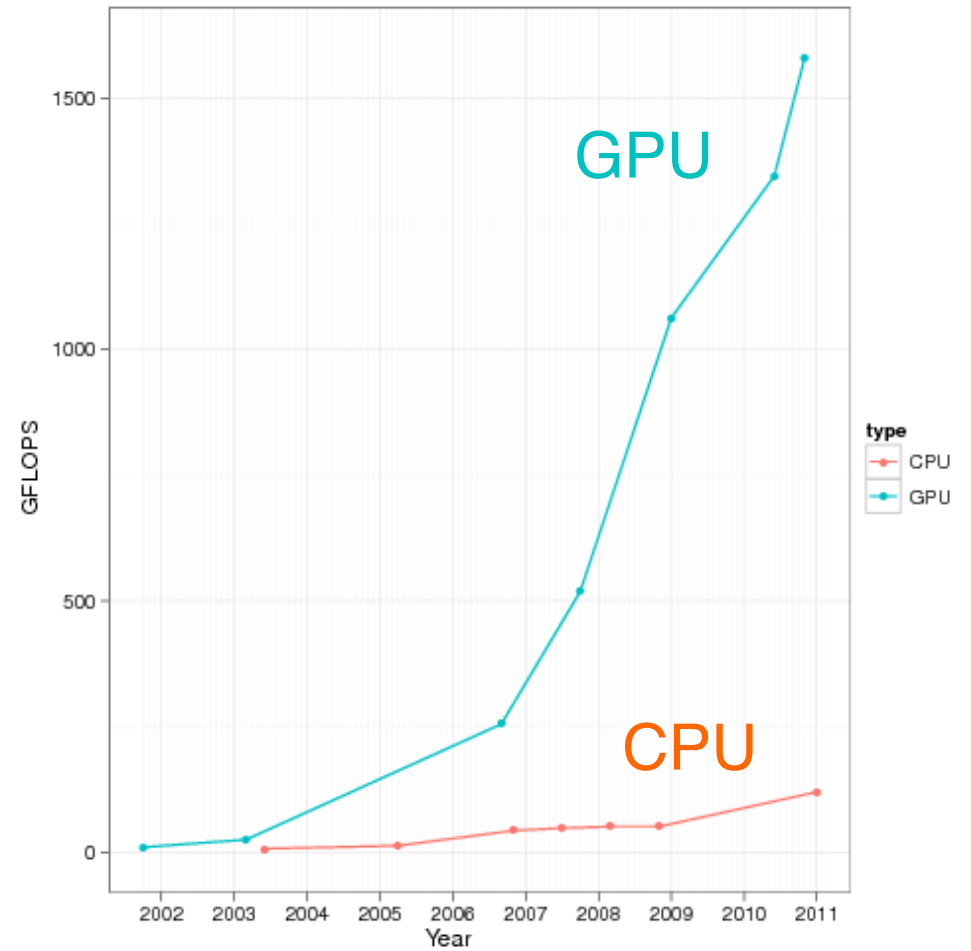
FPGA vs CPU vs GPU

HLT

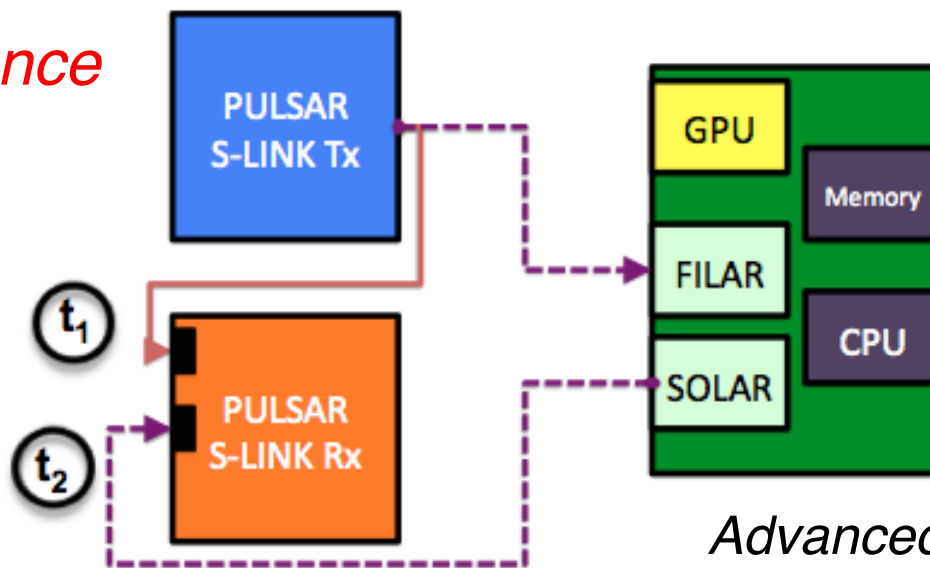
Selected R&D topics

- (1) Data Reduction at detector/sensor stage
- (2) Data Transfer
- (3) Data Formatter using ATCA technology
- (4) Pattern Recognition: 3D AM R&D: VIPRAM
= "Vertically Integrated Pattern Recognition Associative Memory"
- (5) Track Fitting: GPU study for low latency tracking

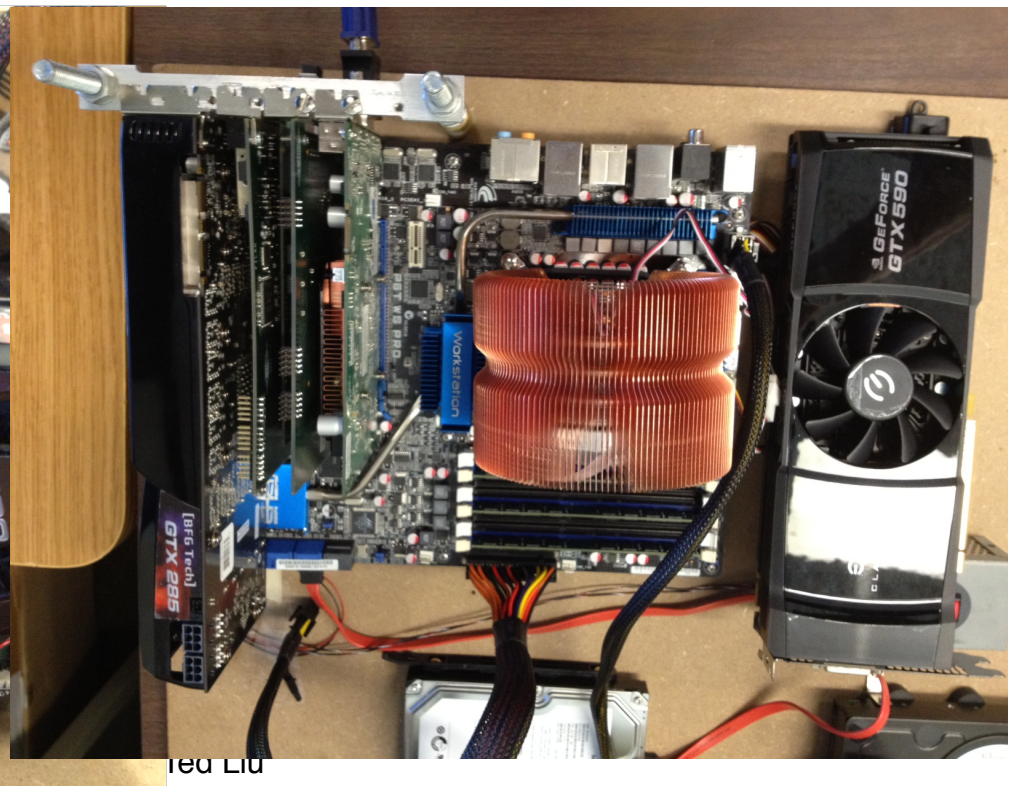
GPU vs CPU evolution



*GPU performance
Study for low
latency
track fitting*



Advanced GPU lab at FNAL



GPUs: highly parallel, multi-threaded, multicore processors with remarkable computational power and high memory bandwidth: promising candidate for fast track fitting at high luminosity for L2 or HLT

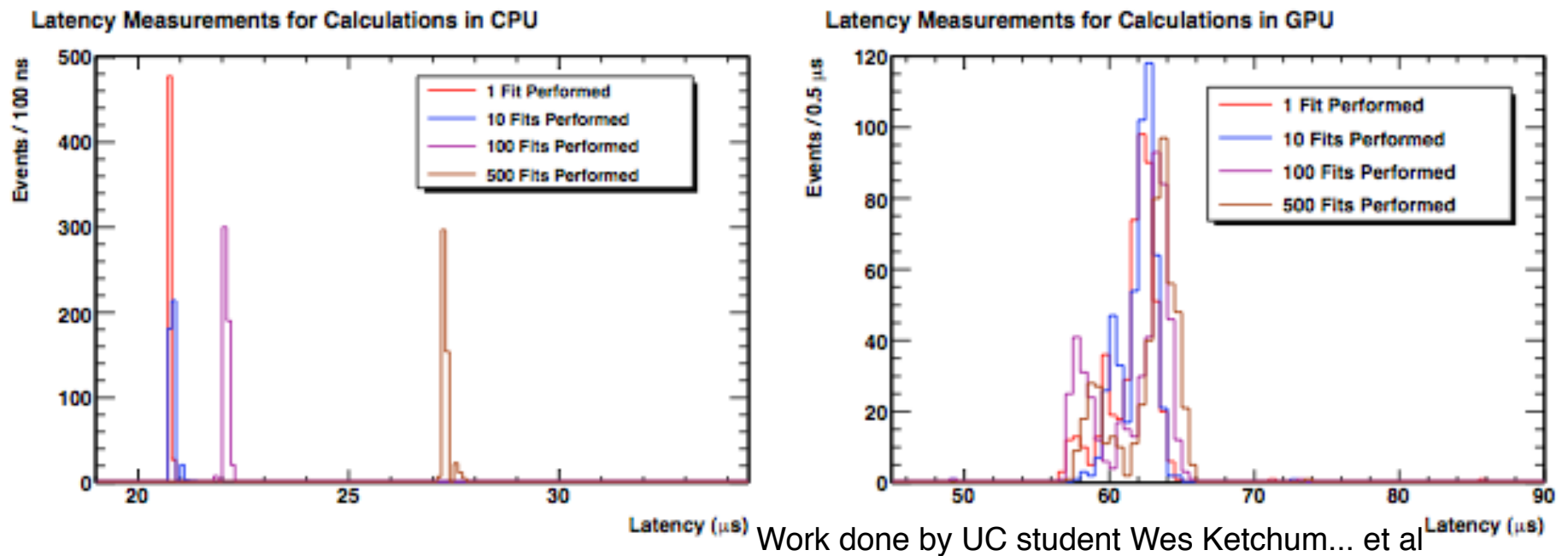
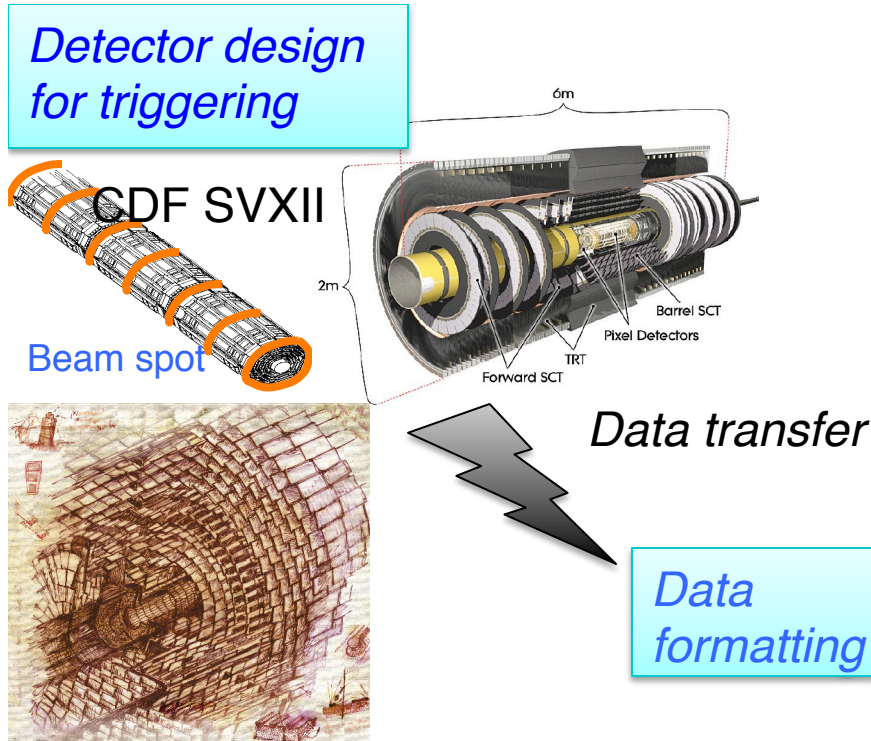


Figure 4: Latency when performing our track-fitting algorithm in the CPU (left) and GPU (right), varying the number of fits performed.

Initial results promising even with an old low-end gaming GPU from Best-Buy

http://hep.uchicago.edu/~thliu/projects/TriggerRD/Processing-Power/TIPP_Proceedings_GPU_preprint.pdf

Tracking Trigger Issues



ATCA based Data Formatting

Pattern Recognition

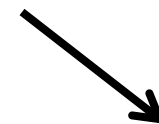
3D VIPRAM / SVT in one chip

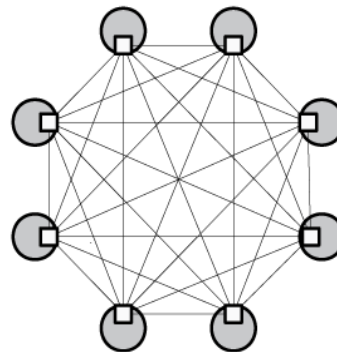
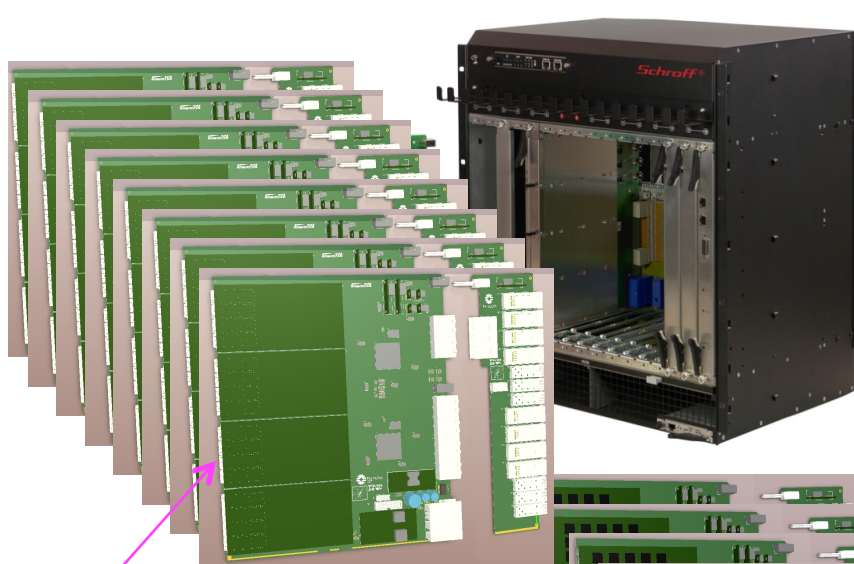
Track Fitting

GPU

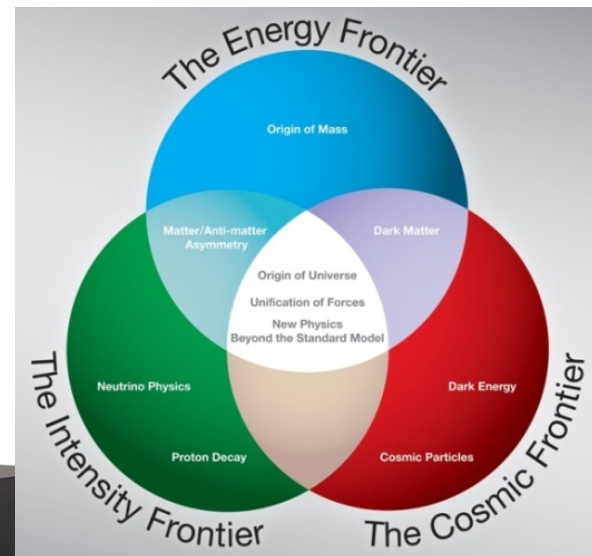
HLT

Combining the *three* developments, one could have a very powerful, general purpose modern trigger/pattern recognition/DAQ system ... far beyond the HEP tracking trigger applications...





Advanced TCA[®]



Fiber inputs

A bank of
VIPRAM or
SVT in a chip

*General Purpose
Trigger/DAQ System
with MASSIVE parallel
processing & pattern
recognition power*

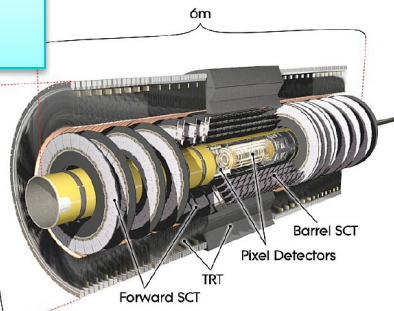
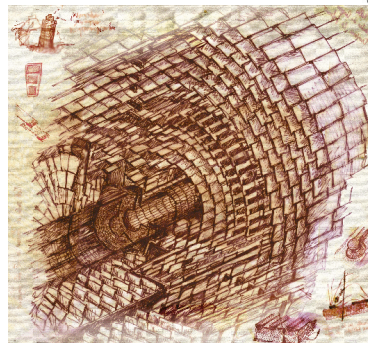
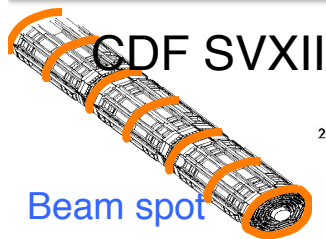
Patterns ~ Billion / crate/shelf

GPU farm



Fermilab

Detector design for triggering



Data transfer

Data
formatting

ATCA based general purpose board

Pattern
Recognition

3D VIPRAM /
SVT in one chip

Track
Fitting

GPU

HLT

Fermilab

Frontier Physics reach at high luminosity LHC to hunt for new physics require: higher detector precision and higher selection capabilities including higher sophistication in the selection algorithms thus the most advanced *Real Time* processing technology

LHC developments in this respect are also a “playground” for developing and experiencing novel *Real Time* field.

Summary