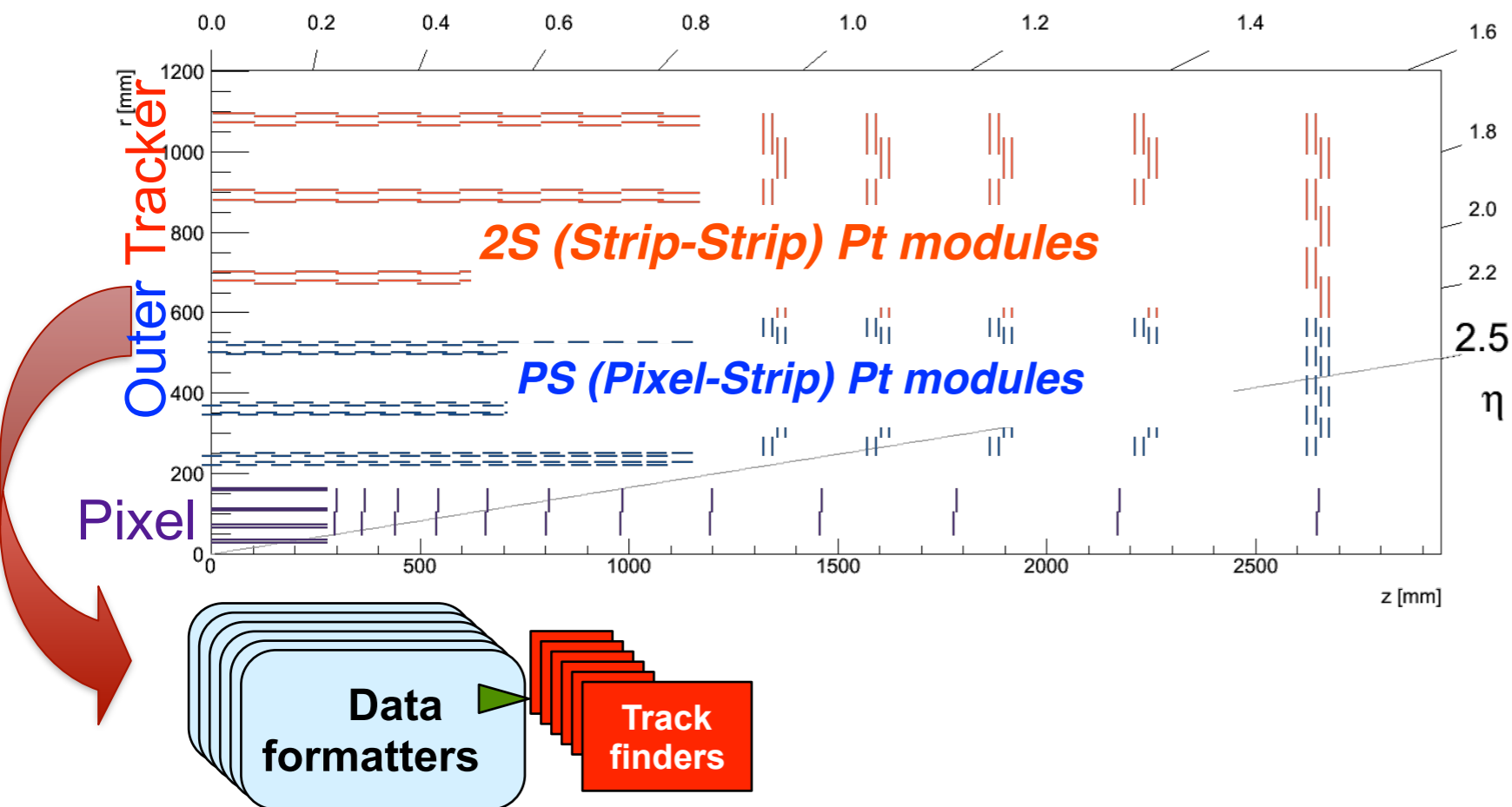




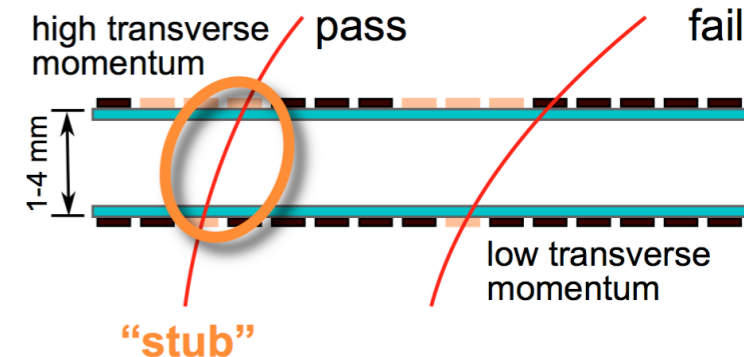
# AM + FPGA based L1 Track Trigger for CMS and INFN developments

*Fabrizio Palla*  
*INFN Pisa*

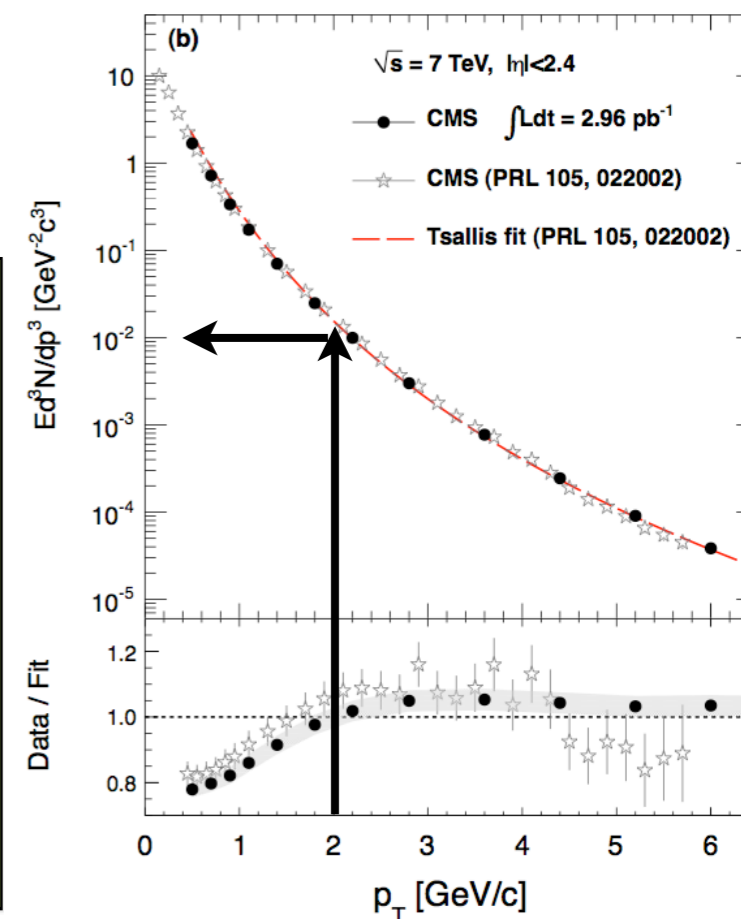
8<sup>th</sup> INFIERI Workshop  
FNAL - October 19, 2016



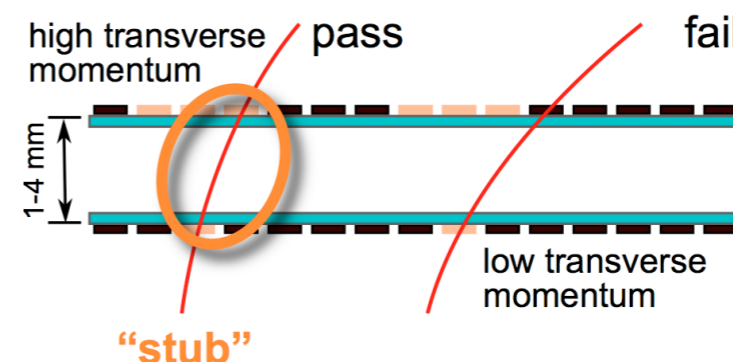
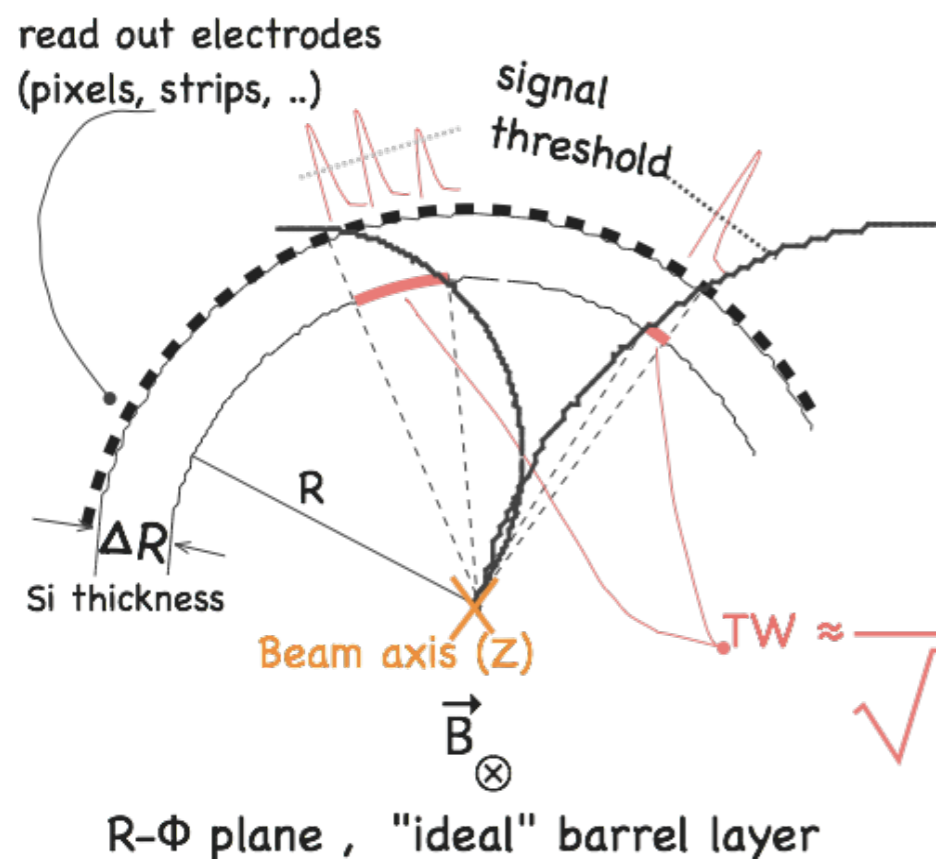
On detector data reduction  
~20, mainly limited by nuclear interactions, conversions and loopers



- Baseline L1 latency  $12.5 \mu s$
- Track Trigger latency  $< 5 \mu s$
- Trigger providing tracks with  $p_T > 2 \text{ GeV}$  at 40 MHz
  - Need time-multiplexing to be able to process  $\sim 50 \text{ Tb/s}$  incoming data
  - Expect up to 300 candidate L1 tracks to Global Trigger



Select “high- $p_T$ ” tracks ( $>2$  GeV) by correlating hits in 2 nearby sensors (stub)

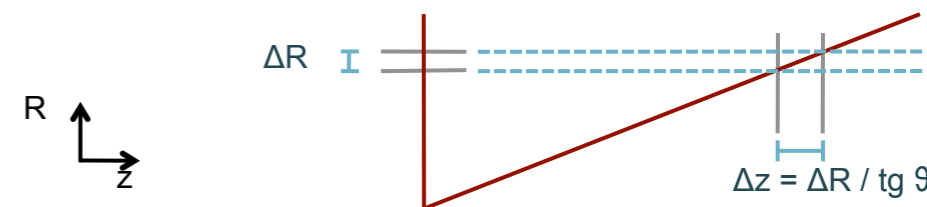


F. Palla, G. Parrini, PoS VERTEX2007 (2007) 034, [http://pos.sissa.it/archive/conferences/057/034/Vertex%202007\\_034.pdf](http://pos.sissa.it/archive/conferences/057/034/Vertex%202007_034.pdf)

J. Jones, A. Rose, C. Foudas, G. Hall, <http://arxiv.org/pdf/physics/0510228v1.pdf>

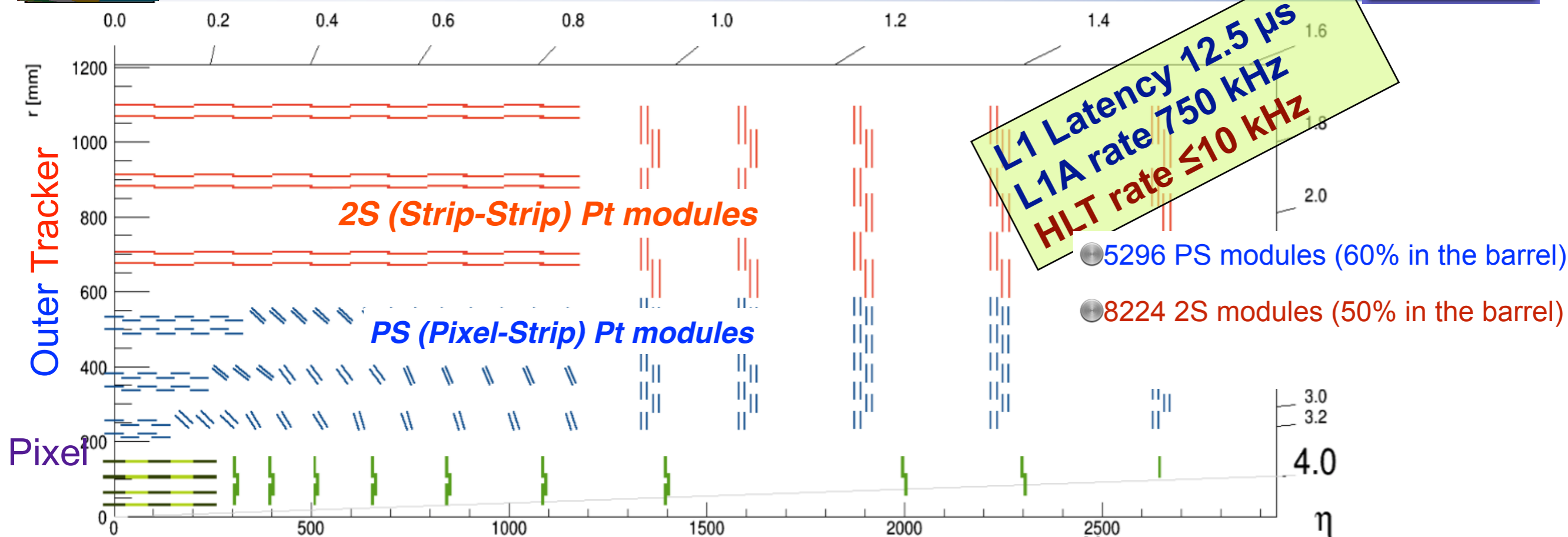
**Large B field of CMS beneficial!**

- In the barrel,  $\Delta R$  is given directly by the sensors spacing
- In the end-cap, it depends on the location of the detector
- ➔ End-cap configuration typically requires wider spacing (up to  $\sim 4$  mm)



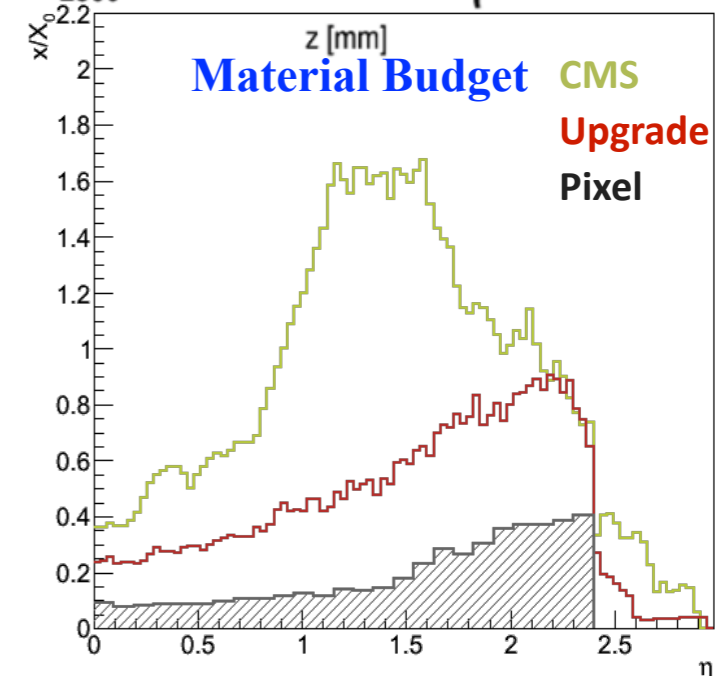
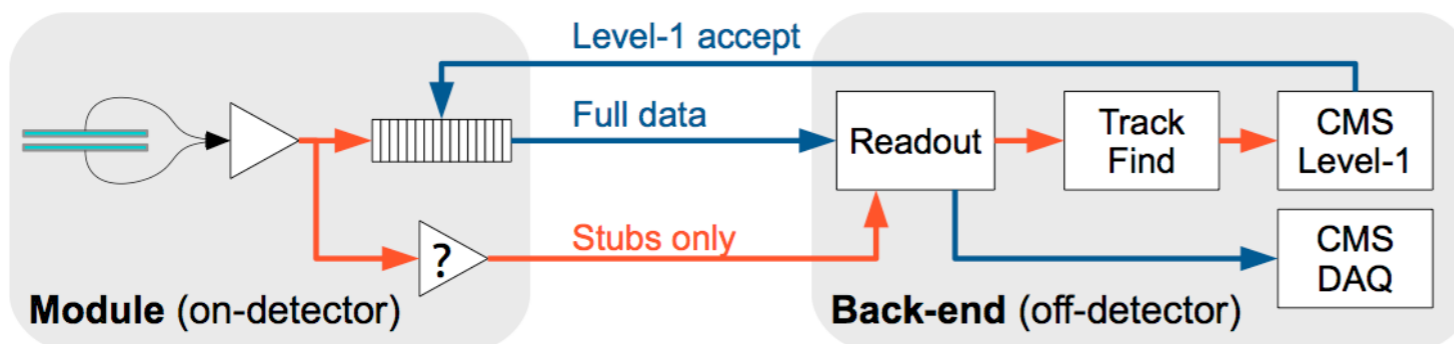


# CMS Upgraded Tracker



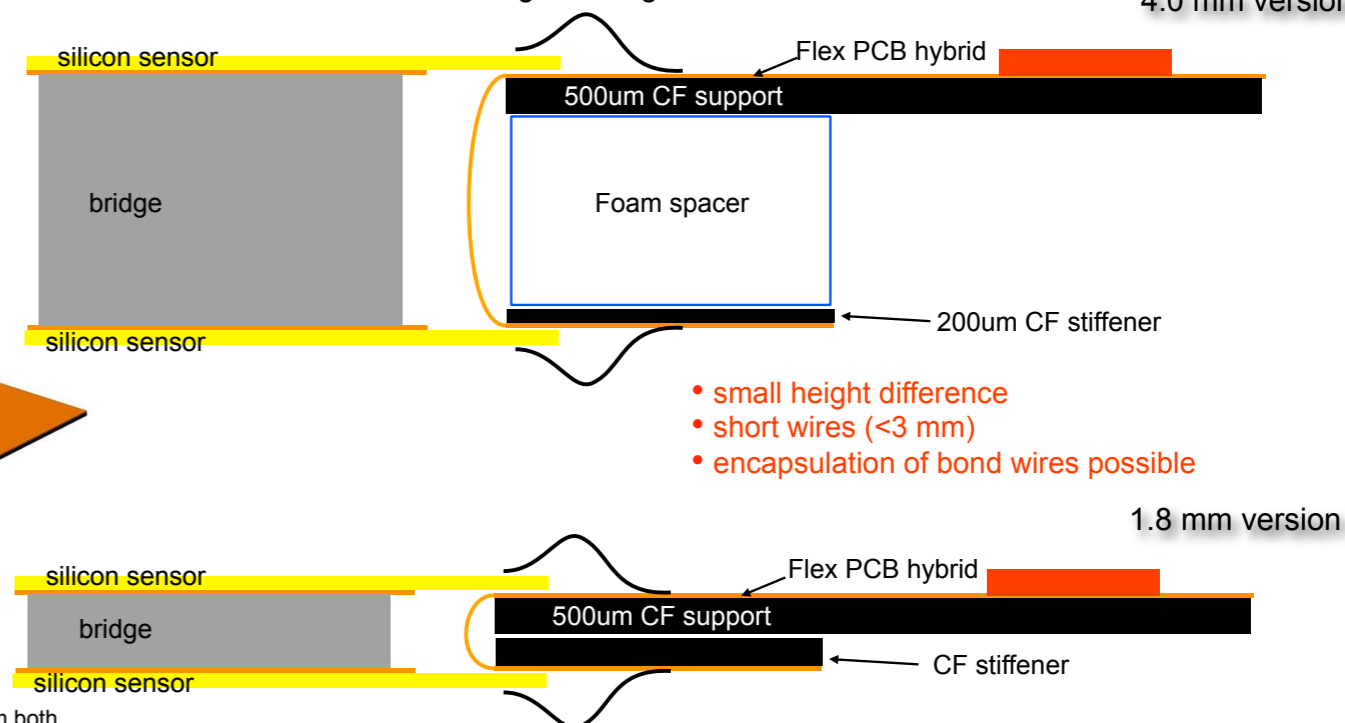
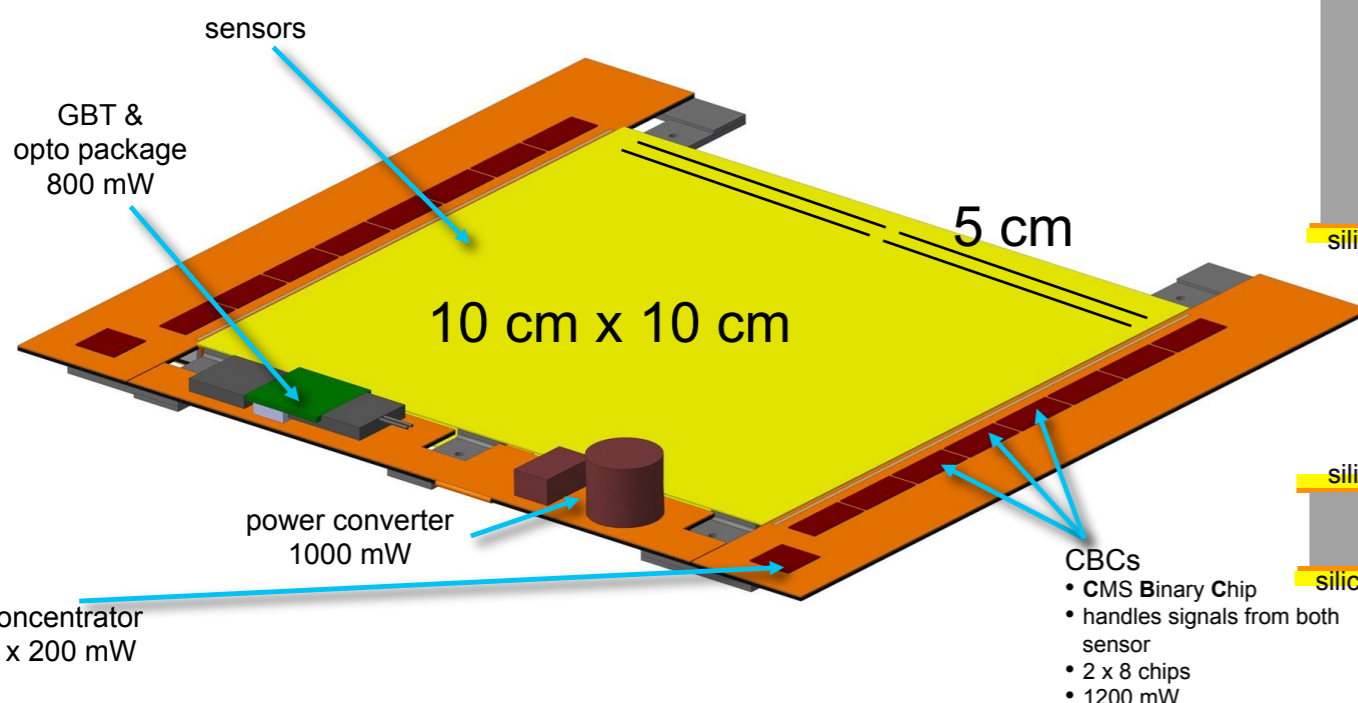
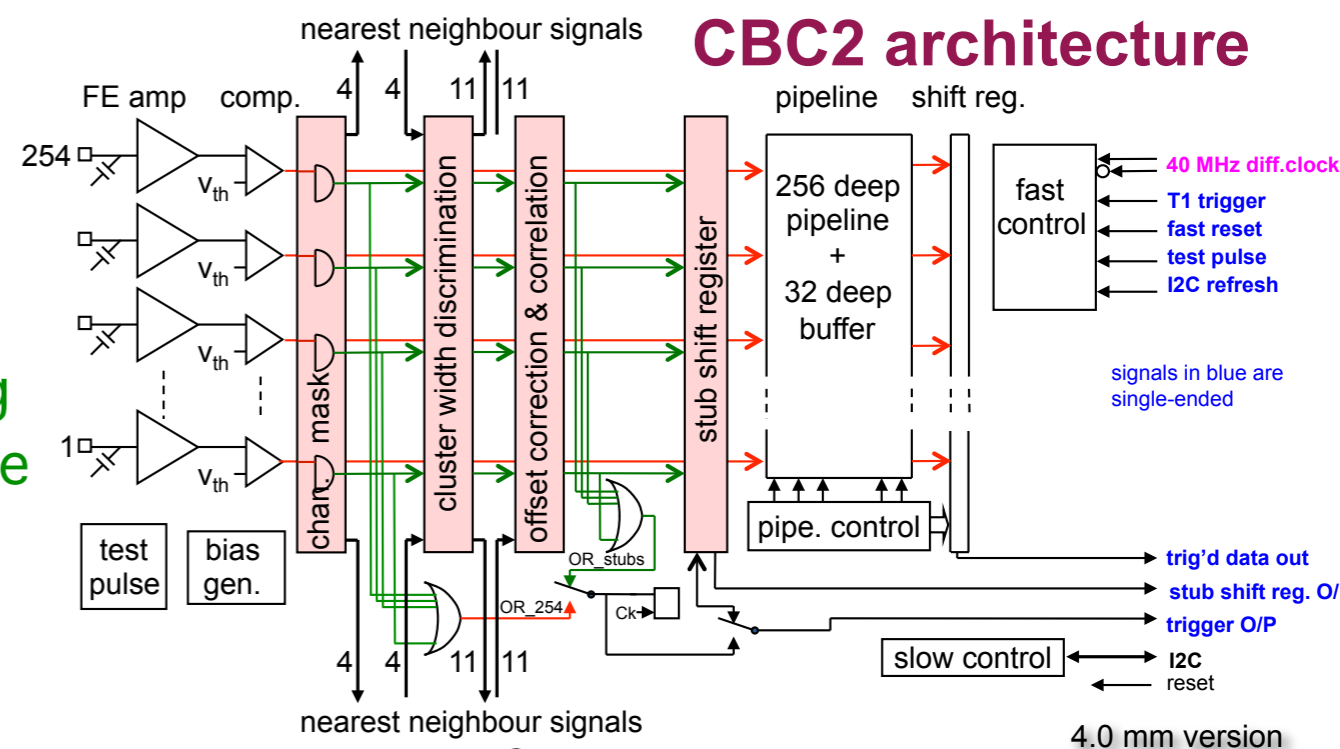
Better  $p_T$  resolution and lighter than current tracker

## Readout and Trigger schematics



## 2S(trip) sensors modules

- 100  $\mu\text{m}$  x 5 cm long strips on both sensors
- readout by 8 CBC on either sides
  - First discriminates signals by rejecting large clusters; then form a coincidence between the two sensor planes
  - Concentrator chip sends data from 8 chips to GBT



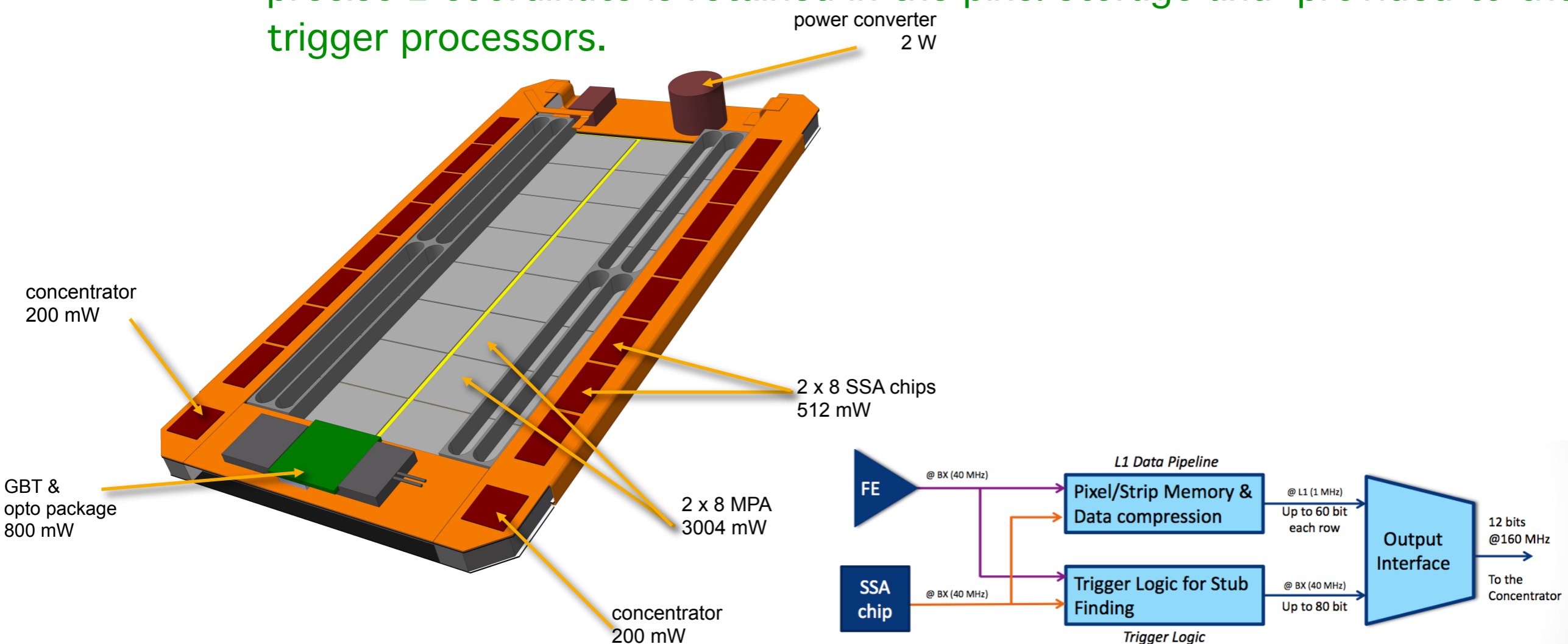
- small height difference
- short wires (<3 mm)
- encapsulation of bond wires possible

## P(ixel)S(strip) module

strips =  $100\ \mu\text{m} \times 2.4\ \text{cm}$




pixels =  $100\ \mu\text{m} \times 1.5\ \text{mm}$

- Pixels are logically OR-ed for finding coincidence in the  $r$ - $\phi$  plane, and the precise  $z$ -coordinate is retained in the pixel storage and provided to the trigger processors.






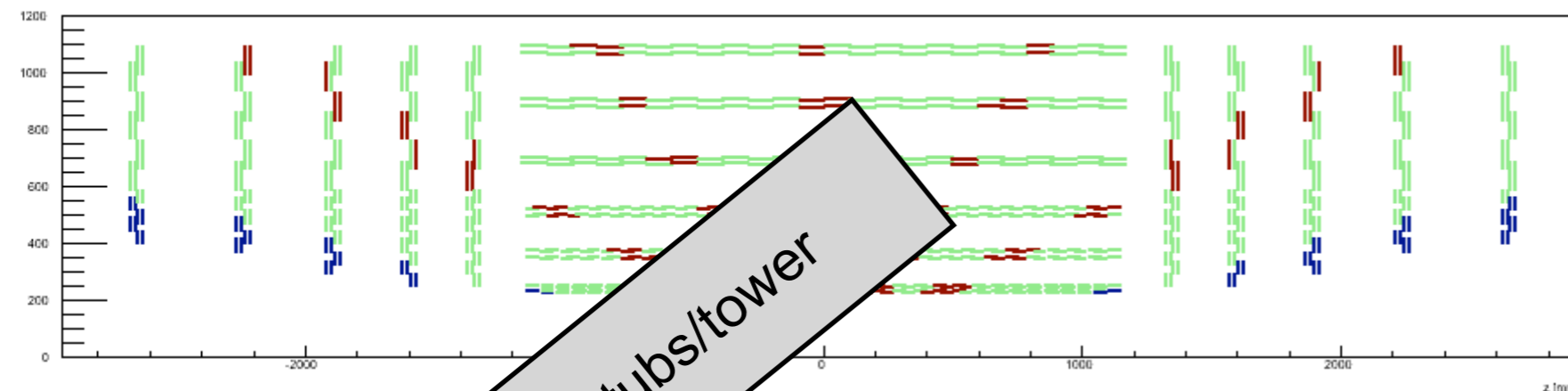


## Goal the demonstrators

-  Develop an hardware system capable to (efficiently) reconstruct tracks with  $p_T > 3$  GeV within the latency of  $4 \mu\text{s}$ , using the current state of the art technology, validating the simulation studies
-  Use the simulation and emulation to then dimension the system with technology available for HL-LHC
-  Evaluate the costs of the final system

## Three demonstrators under development

-  AM + FPGA (this + Sergo's talks)
-  Tracklet (Jorge's talk)
-  TMT (Ian + Davide + Luigi's talks)



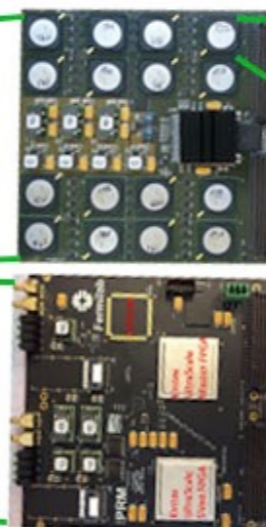
6 ( $\eta$ ) x 8 ( $\phi$ ) sectors  
 $\sim 1$  M patterns/sector  
 x20 time multiplexed  
 (eventually 10)  
**Total of  $\sim 1$  Billion patterns**



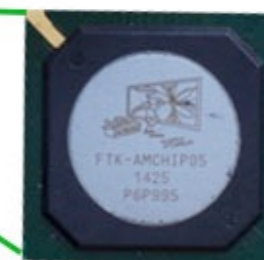
ATCA



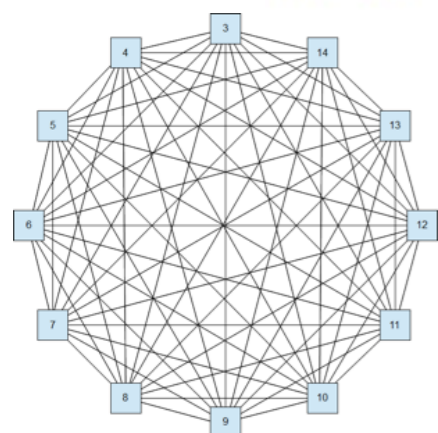
PulsarIIb



PRM

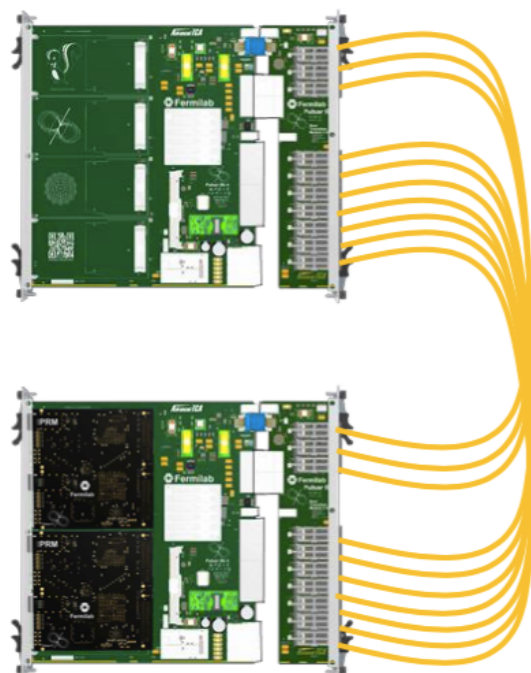


AM Chip



**Send data to Pattern Recognition Mezzanine in each ATCA blade**

- Data distributed to Pulsar boards in time multiplexed mode in round robin
- Perform pattern recognition using several AM chips
- Track fit with FPGA inside the Mezzanine ( $\sim 1$  ns/fit)
- Latency  $< 4 \mu\text{s}$  (out of  $12.5 \mu\text{s}$ )

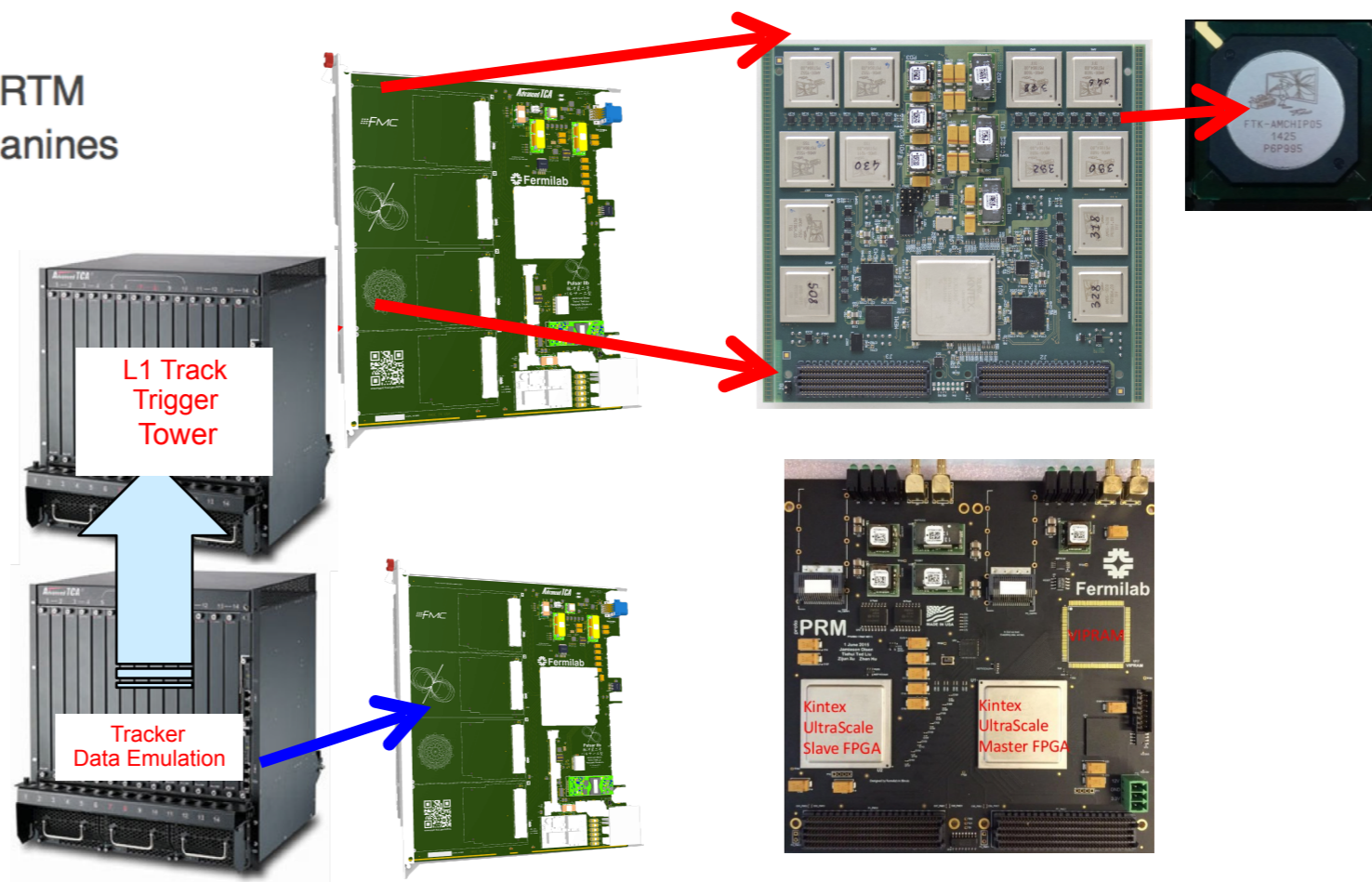


## Data Source Board (DSB) shelf

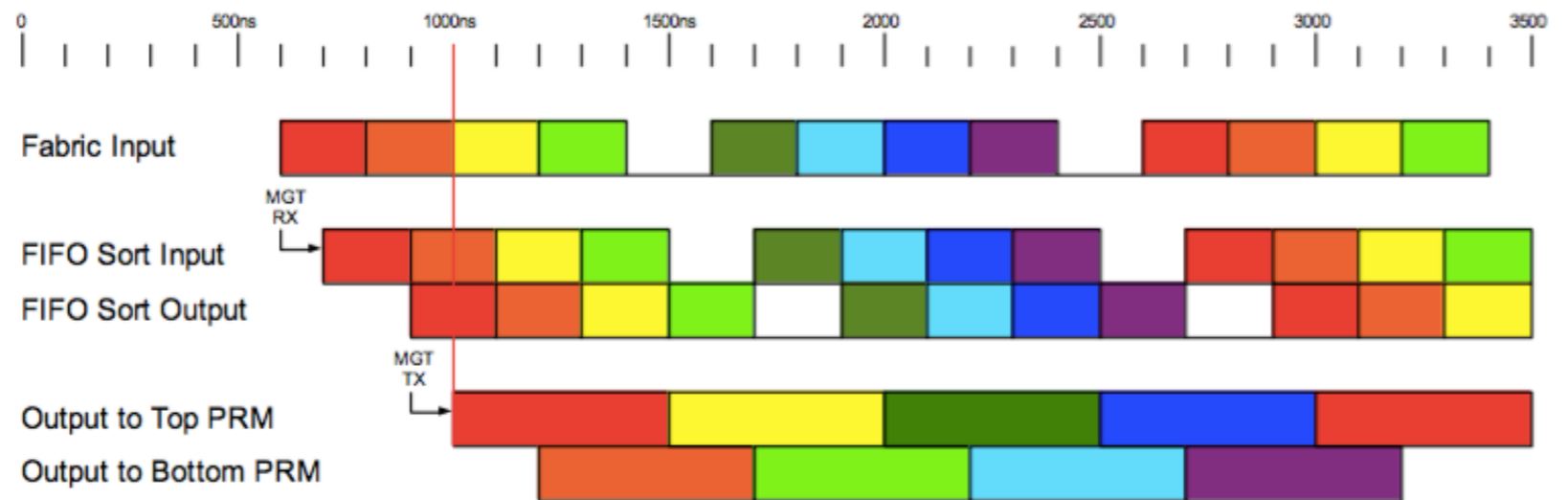
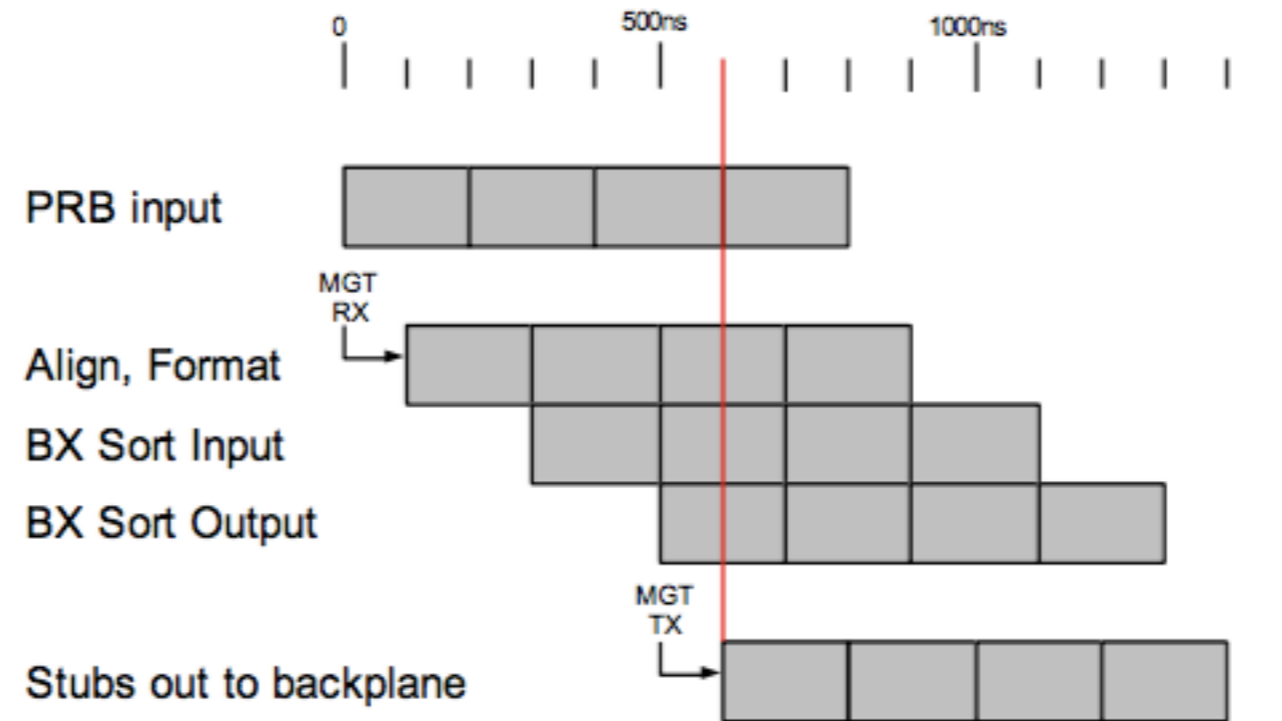
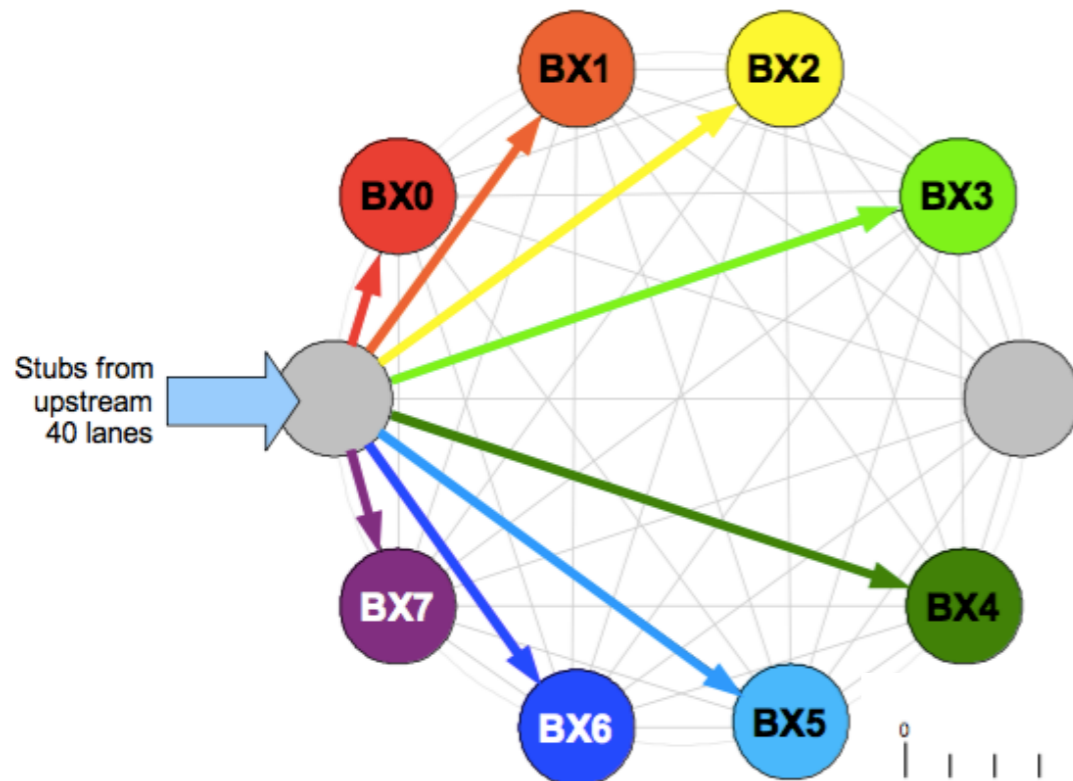
- Emulates the output of ~400 modules
- 10 x Pulsar2b with RTM
- 100 QSFP+ fibers
  - 400 lanes @ 10Gbps

## Pattern Recognition Board (PRB) shelf

- One Trigger Tower
- 10 x Pulsar2b with RTM
- Various PRM Mezzanines



Since events from LHC arrive every 25 ns and the time to reconstruct tracks is  $4\mu\text{s}$  one needs to divide event processing in time (“time multiplexing”)

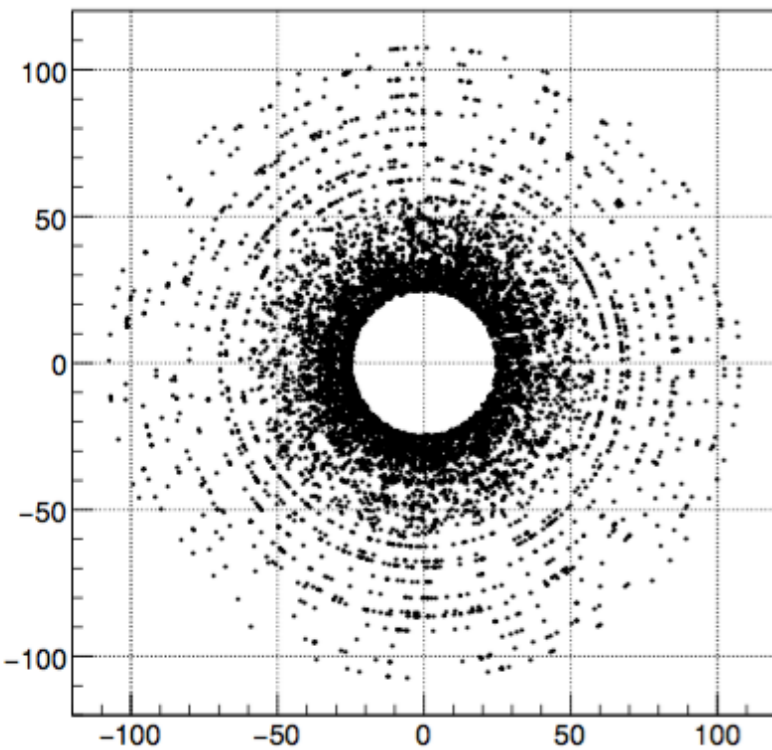


**Stubs**

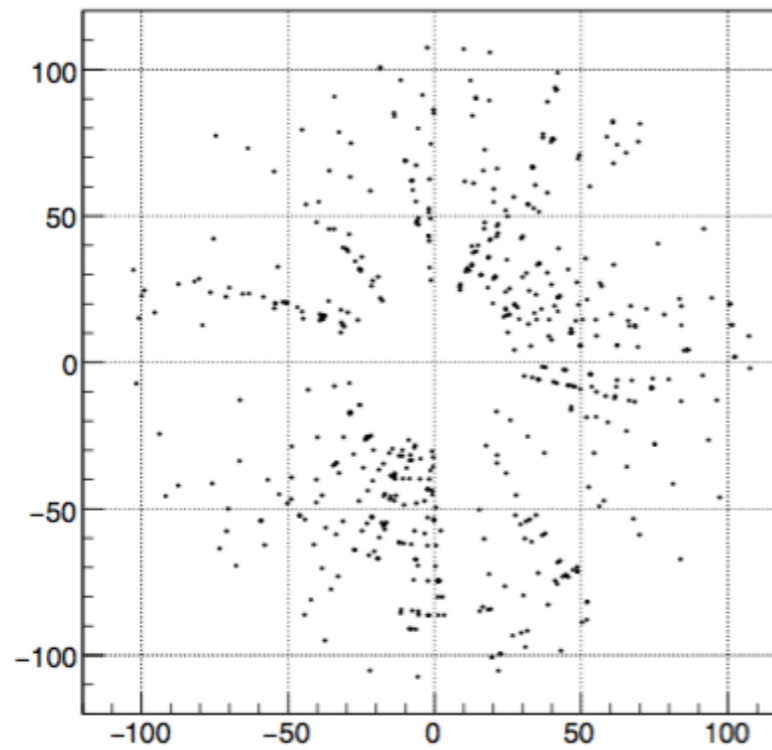
**Matched  
roads**

**Tracks**

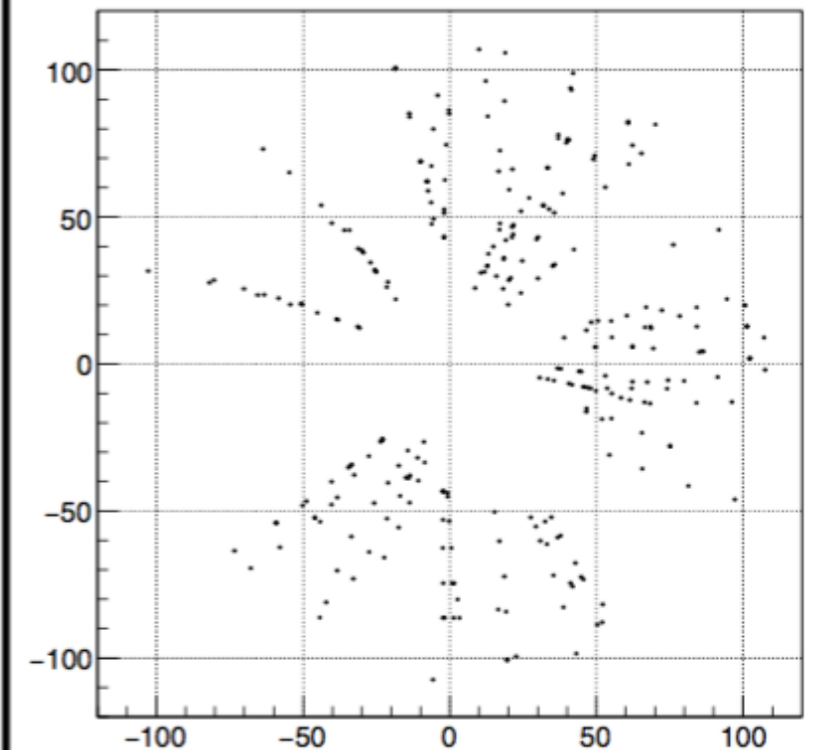
**RAW**



**AM chips**



**FPGA**

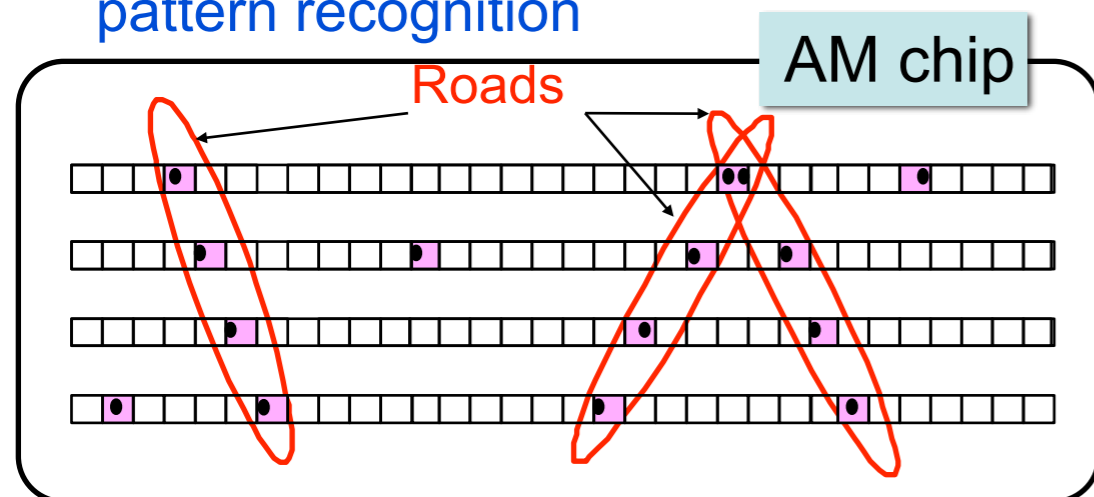


**8240 stubs**

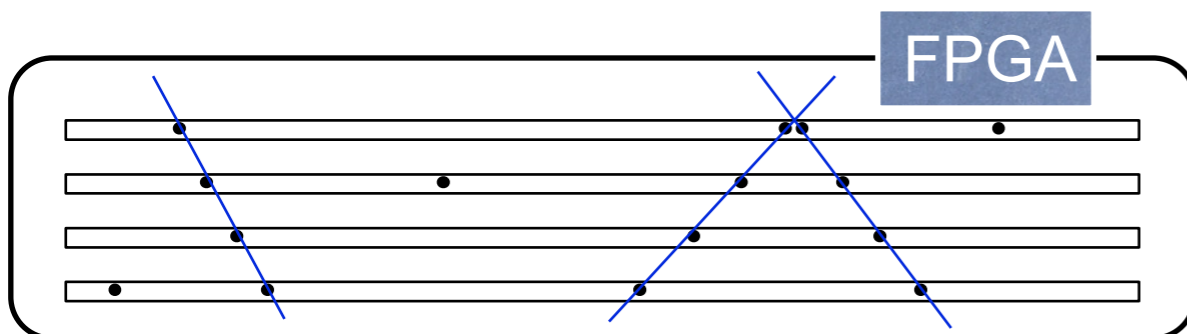
**532 stubs and  
proto track  
candidates**

**281 stubs and  
fitted tracks**

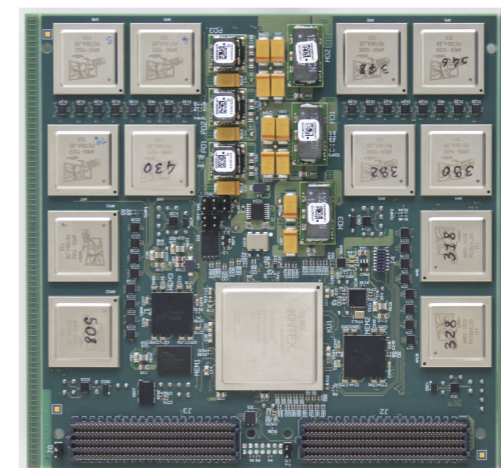
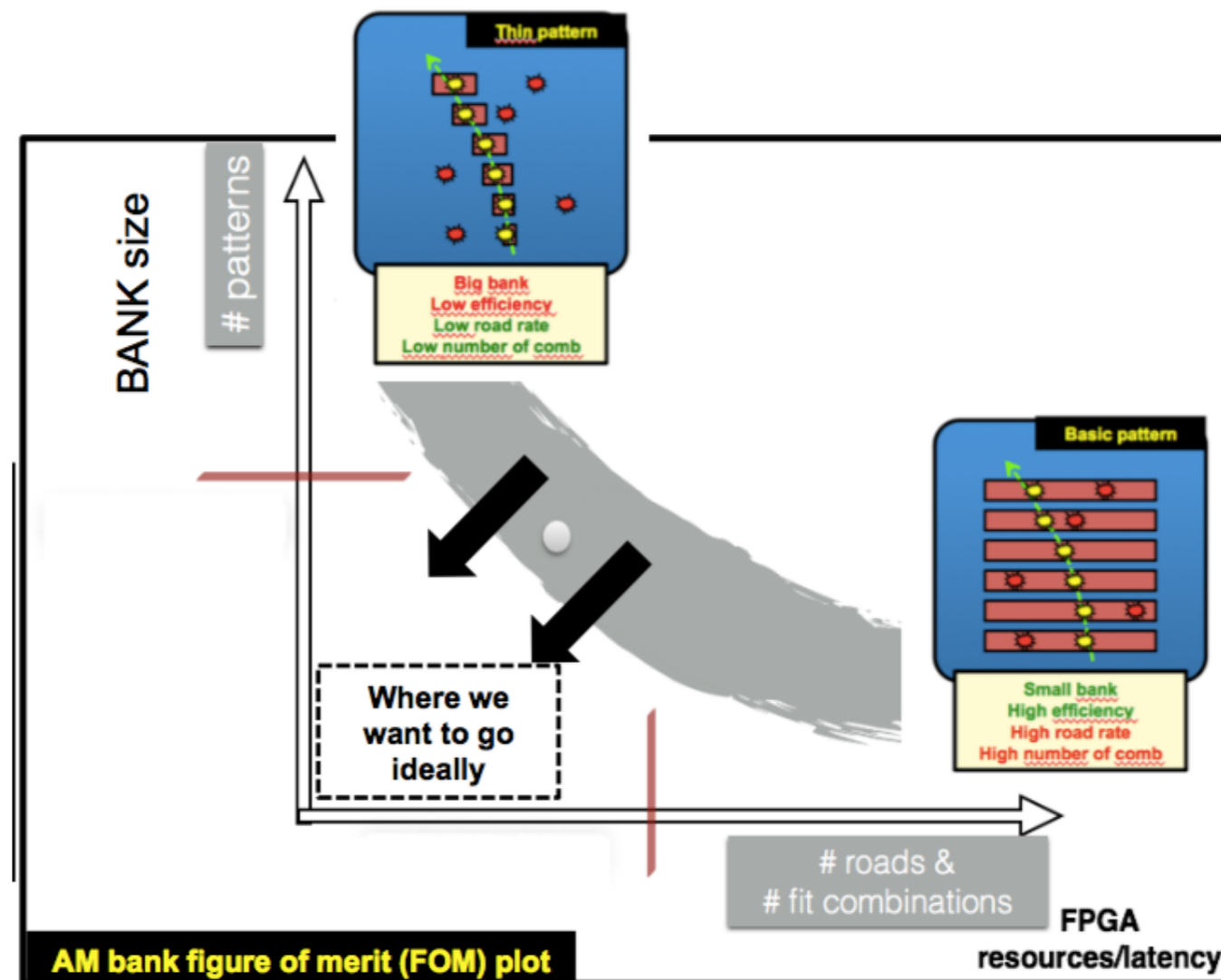
1. Find low resolution track candidates called **"roads"**. Solve most of the pattern recognition



2. Then fit tracks inside roads.  
Thanks to 1<sup>st</sup> step it is much easier



AM chip + FPGA



## Number of patterns/Tower

~ 0.5 M for the Barrel and Forward towers

~1 M for the intermediate  $\eta$  region (Hybrid) towers

4 (or 8) AM06 chips (128k patterns - today's technology)

## Matched roads & combinations in $\langle PU \rangle 200 + t\bar{t}$

~20 matched roads on average (~60@95% percentile)

~90 combinations on average (~250@95% percentile)

## Fitting with "Principal Component Analysis"

- Over narrow regions of the detector, equations linear in the local hit coordinates give resolutions on track parameters nearly as good as time-consuming helical fit

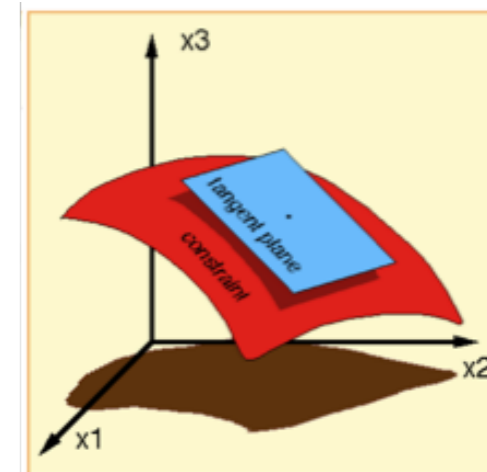
$p_i$ 's are the track parameters

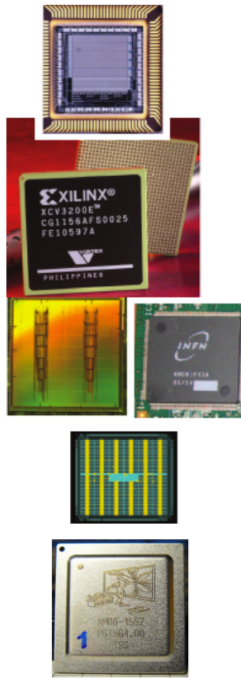
$x_j$ 's are the hit coordinates in the silicon layers.

$A_{ij}$  &  $B_i$  are pre-stored constants determined from full simulation or real data tracks.

Several ways to implement: ~20K constants

$$p_i = \sum A_{ij} x_j + B_i$$





Version	Year	Design	Tech.	Area (cm <sup>2</sup> )	Patterns	Frequency (MHz)	Power (W)
AM03	2004	Std. cells	180 nm	1	5k	40	1,26
AM04	2012	Std. cells+ Custom	65 nm	0,12	8k	100	3,70
AM05	2013	Std. cells+ Custom+ IP	65 nm	0,12	1k+2k	100	<1
AM06	2014	Std. cells+ Custom+ IP	65 nm	1,7	128k	100	2-3
<b>AM07</b>	<b>2016</b>	<b>Std. cells+ Custom</b>	<b>28 nm</b>	<b>0,1</b>	<b>16k</b>	<b>200</b>	<b>0,1</b>

AM05: technology testing chip

AM06: production chip, used in FTK ATLAS track trigger

**AM07: technology testing chip, under design**

## AM07 ARCHITETURAL FEATURES

Area: 10 mm<sup>2</sup>

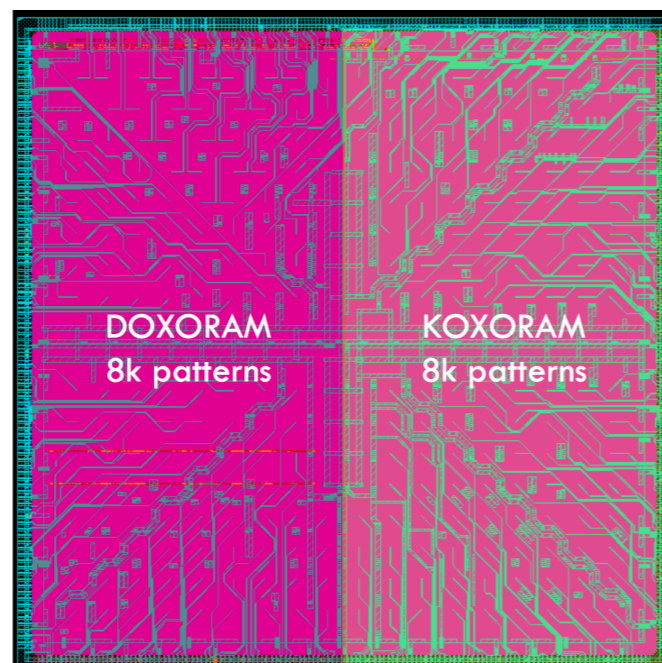
Memory depth: 16 kpatterns

400 bumps

4 independent cores

LVDS or LVCMOS interface

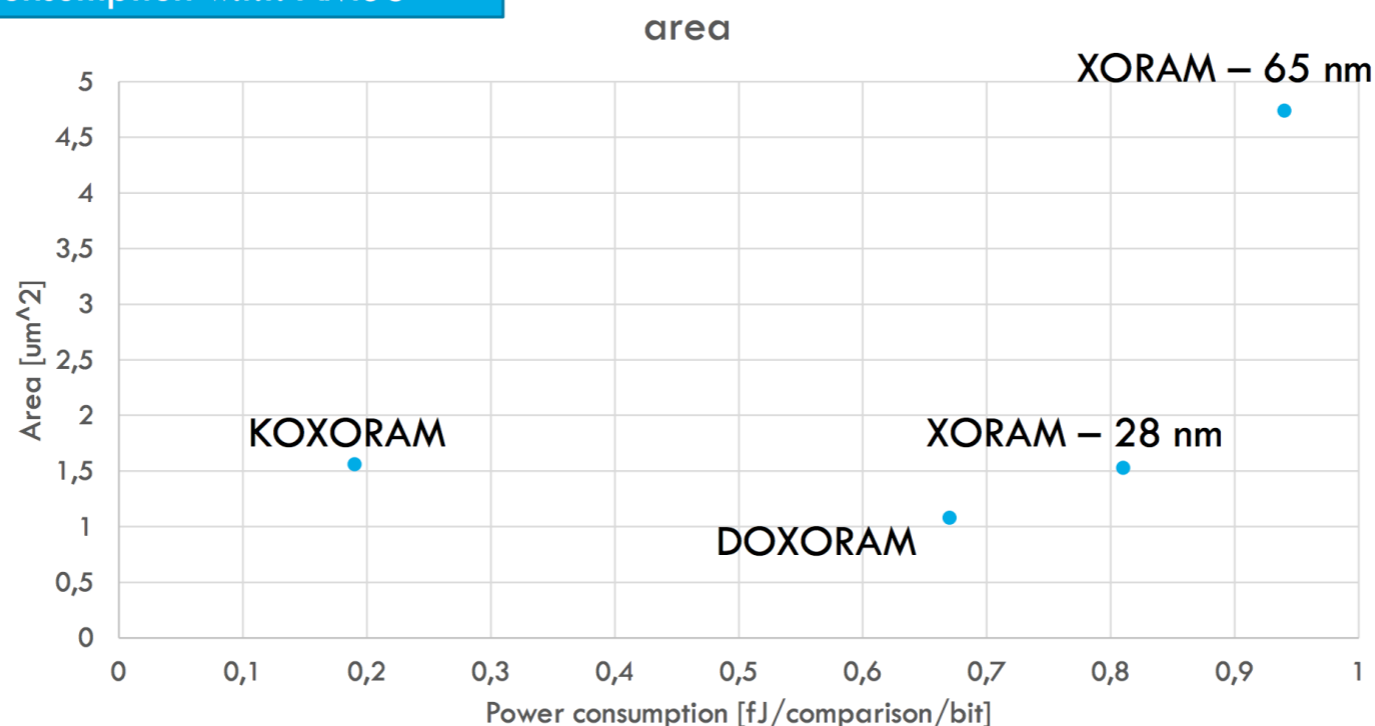
Working frequency: 200 MHz

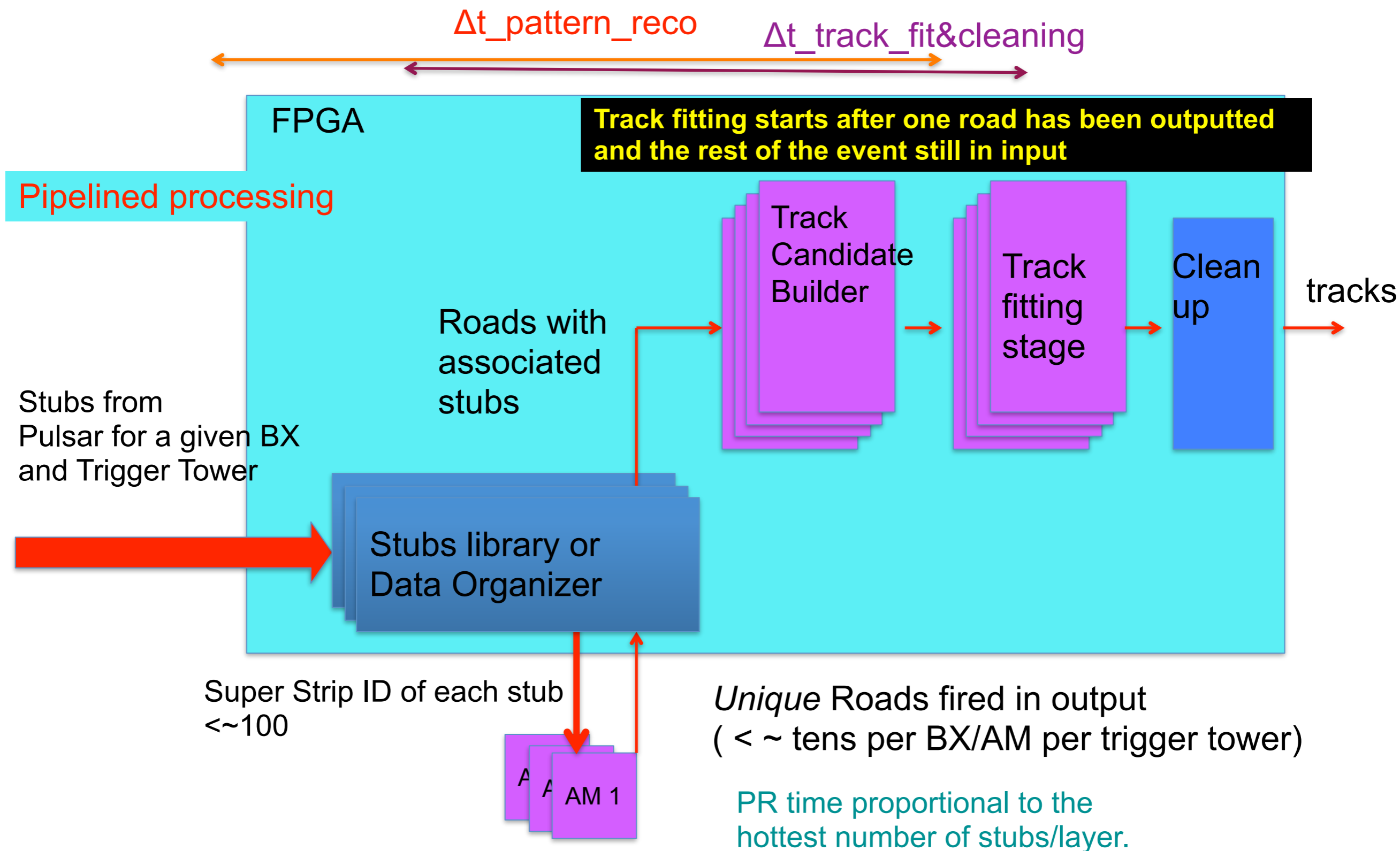


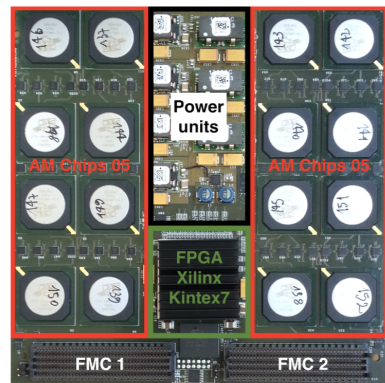
1/4 area and power consumption w.r.t. AM06

## AM07 goals

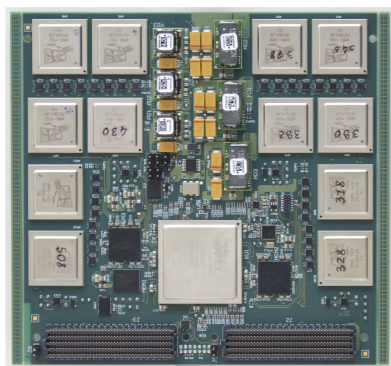
- Provide a working AM chip at 28 nm
- Test two different CAM design
- Aim for pattern/unit area 4x AM06
- Design for 200+ MHz clock speed
- Lower energy/comparison/bit
- Include and validate LVDS I/O





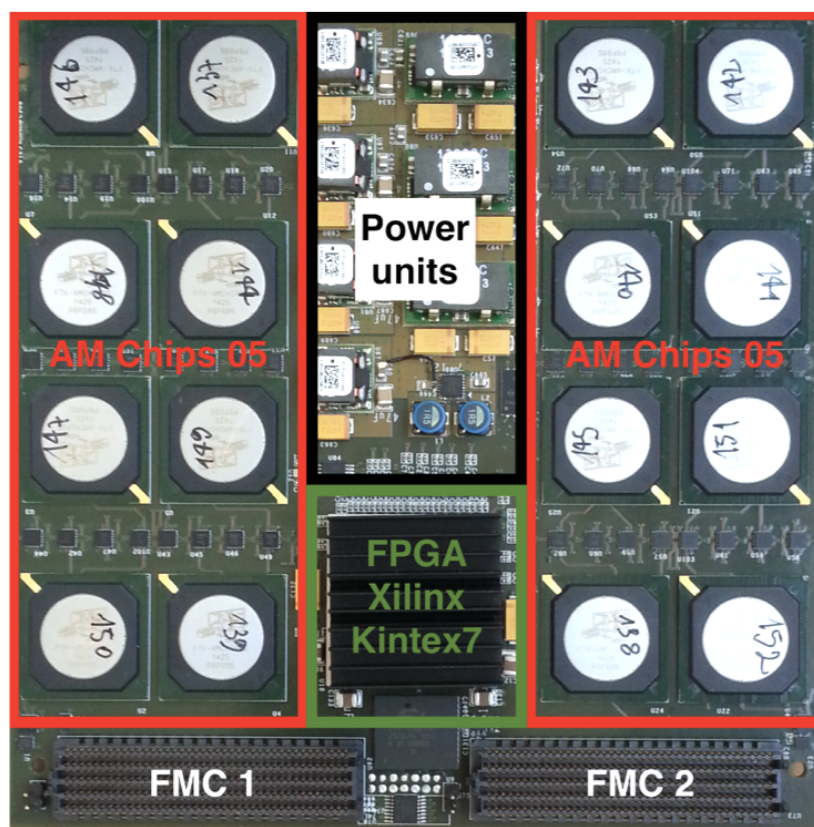


**PRM05:** first PRM developed for CMS L1 track trigger, designed as pilot board with technology-test AM05 chip. We have been developing and testing the reconstruction FW in this board before porting it to PRM06.

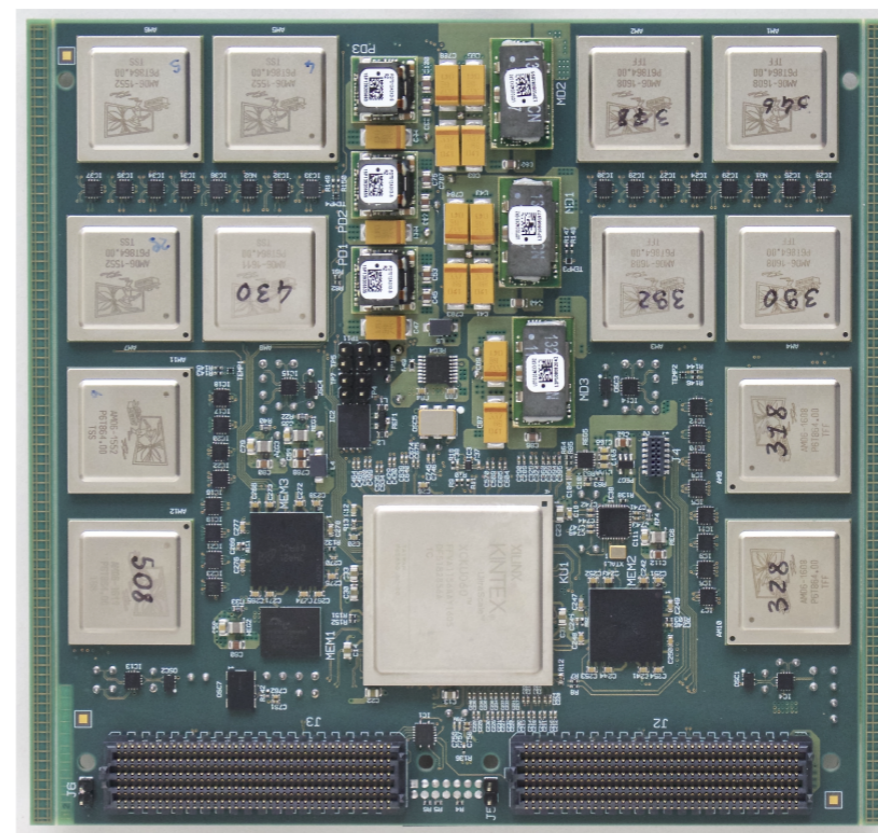


**PRM06:** designed to be used in the demonstrator, once AM06 chip be available (Spring 2016), with its 1.5 million pattern bank **can cover a full trigger tower** pattern bank (0.5-1M patterns). Profits of a Ultrascale FPGA

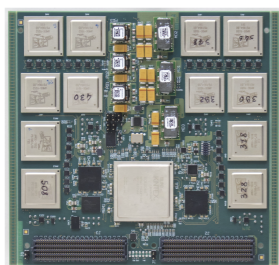
## INFN PRM05



## INFN PRM06



Logic blocks	Number of resources (PRM05)	Number of resources (PRM06)
AM Patterns	32 kpatterns	1,5 Mpatterns
External Memory	18 Mbit	1,1 Gbit
Logic Cells	356 k	726 k
Block RAM	25 Mb	38 Mb
DSP Slices	1444	2760
Transceivers	24 @ 8 Gbps	28 @ 10.3 Gbps
I/O Pins	300	104, 416



Prototype delivery date: June 2016  
 Prototype validation date: July 2016  
 2 additional PRM06s: July 2016  
 Pulsar-PRM06 testing date: September 2016

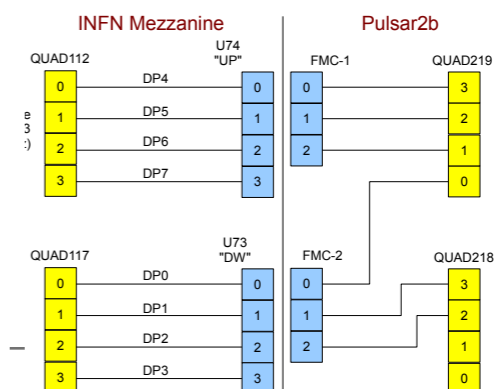
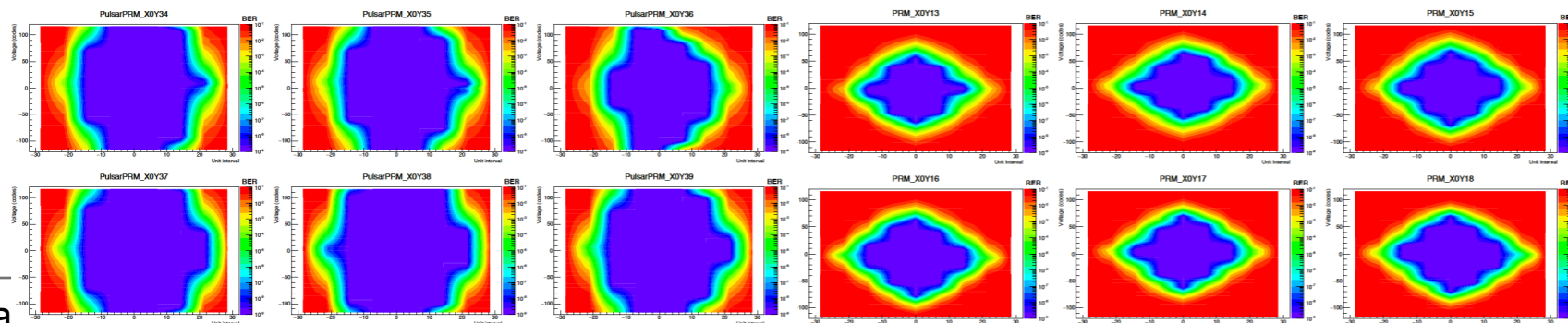
**PRM06 hardware tested and working as expected**

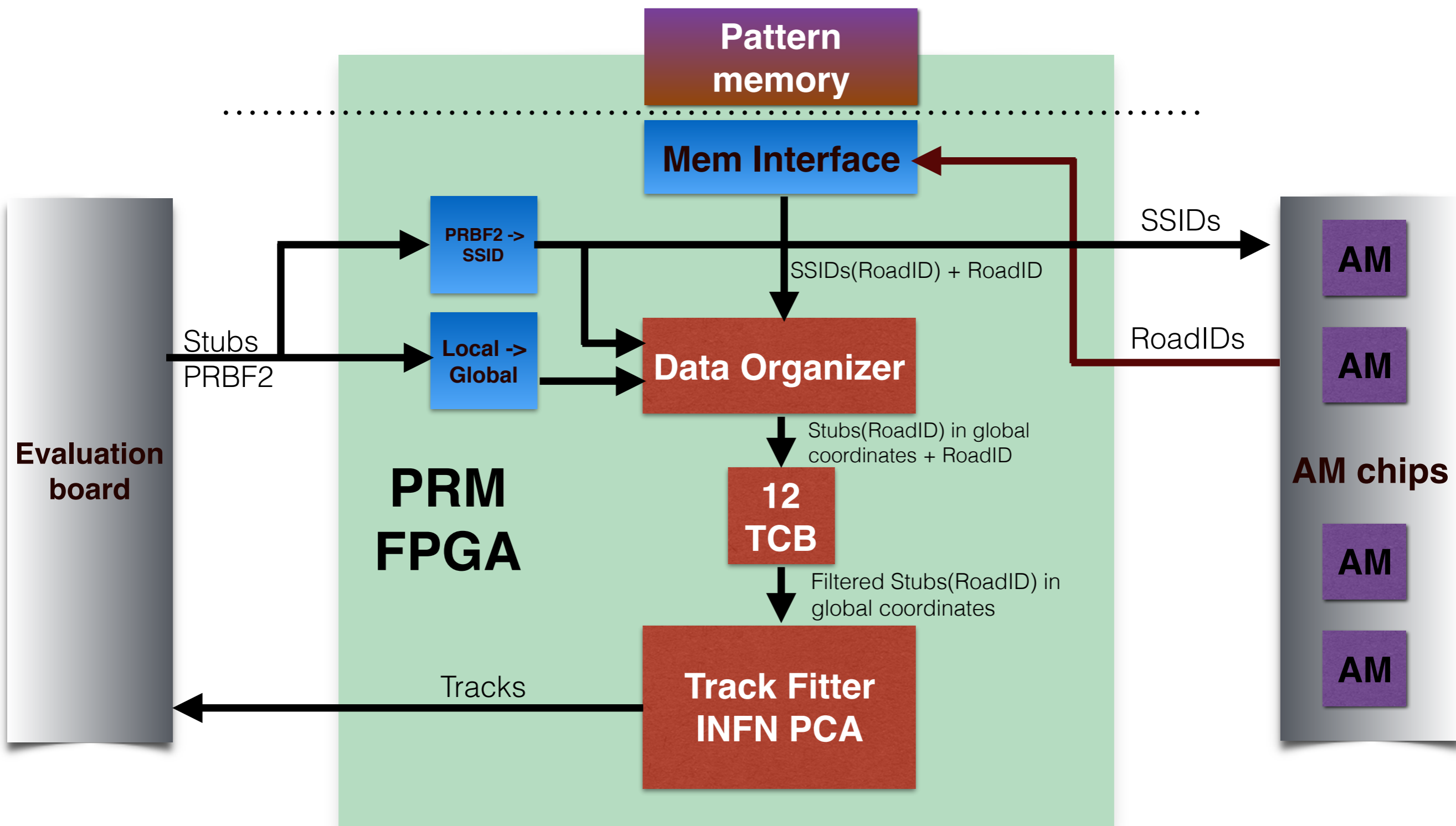
- **GTH links:** IBERT PRBS7 on all links
- **AM06 communication and configuration:** JTAG communication (each AM06 tested before to be mounted on PCB: serdes' and bank memory with built-in test)
- **Serdes links:** PRBS on all links
- **LVDS links going to FMC connector:** loopback on evaluation board, static test
  - Update w.r.t. May: we tested all the links with the Pulsar and they are all ok
- **RLD3RAM:** using Xilinx tools, checked the reading and writing
- **External flash memory:** using OpenCores IP we tried basic communication
- Test of the **GTH links between PulsarIIb and PRM06**

## IBERT eyescans @10Gbps

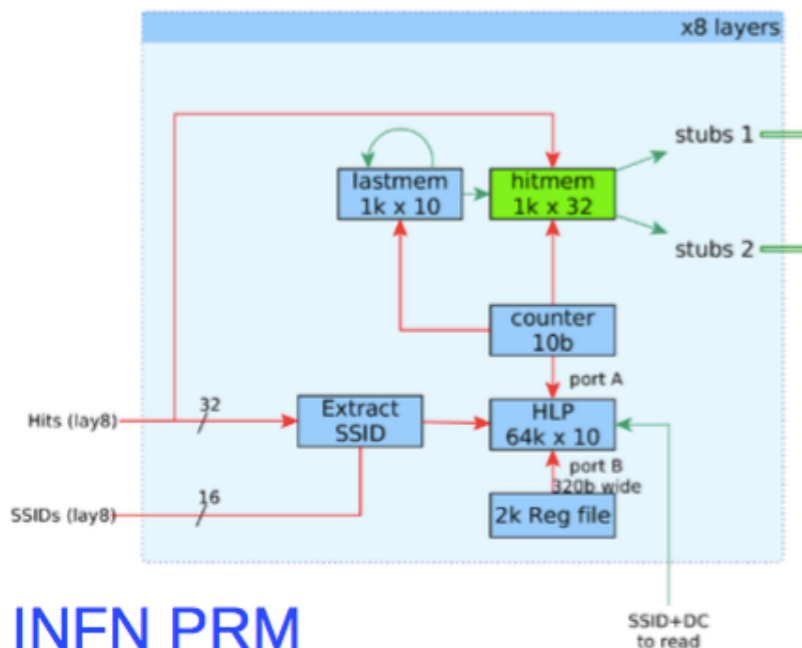
Pulsar:

PRM:



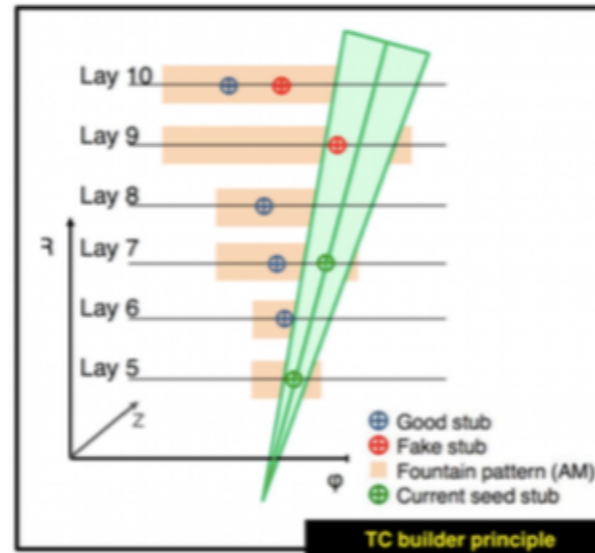


## Data Organizer



## INFN PRM

- Stores up to 1k stubs/layer/event
- No fixed max on stubs per SS
- Supports Dont Care Bits & missing layers
- Ping-pong operation w/ 2 instances
- 45 cycle read latency (first SSID in, first stub out), many stubs output per cycle, currently 450 MHz max

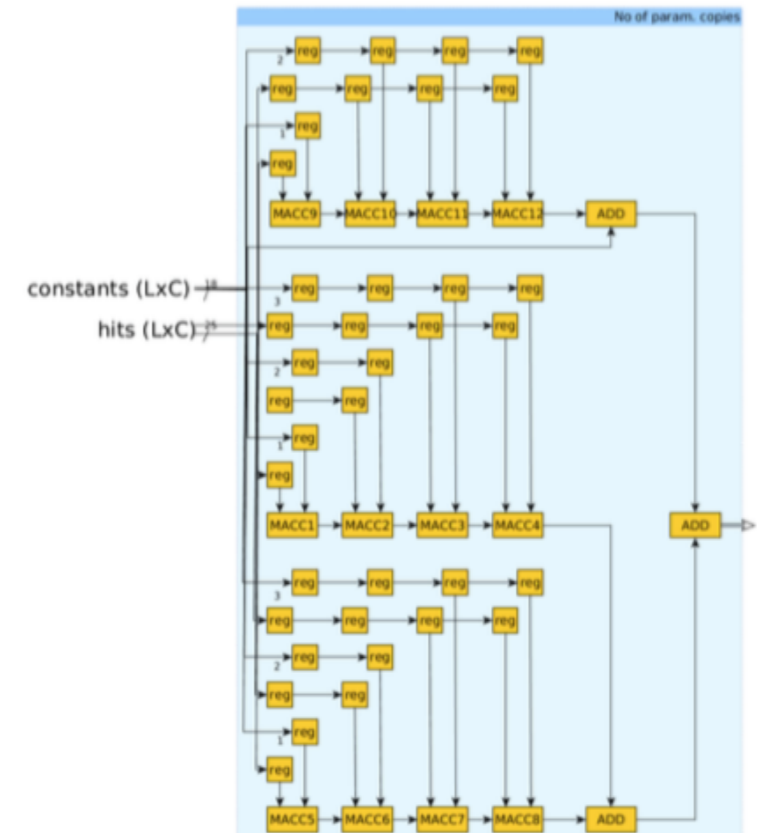


## INFN PRM

- Seed using pairs (<12) of PS stubs
- Project into 2S, send out only closest projected combination for given road
- Helps tail events w/ many matched roads w/ many combinations
- ~120 - ~200 cycle latency, 200-300 MHz, can include track estimation
- More resource/latency intensive, but lightens load on downstream TF

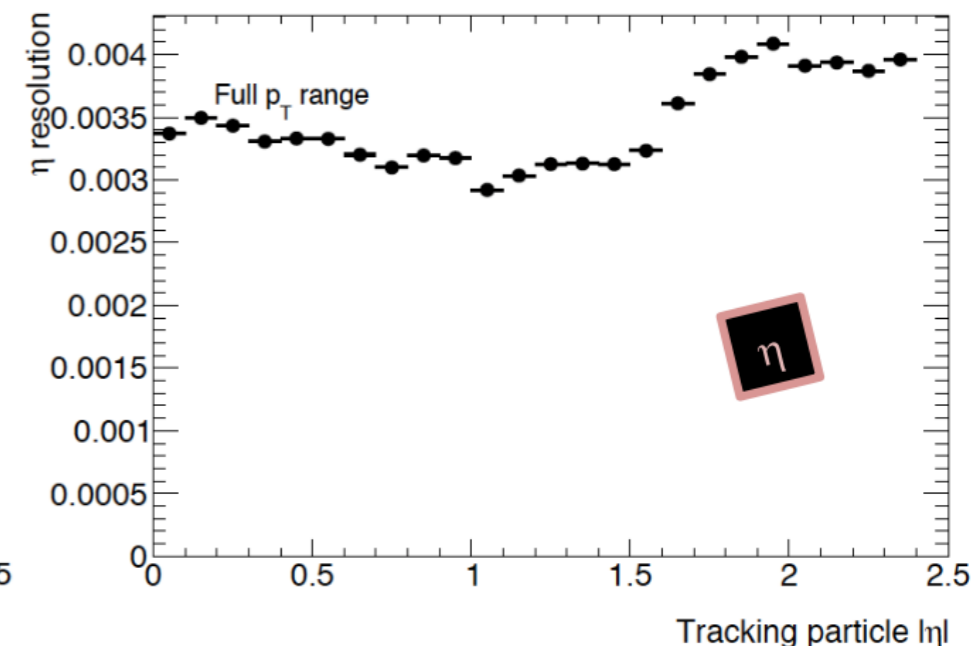
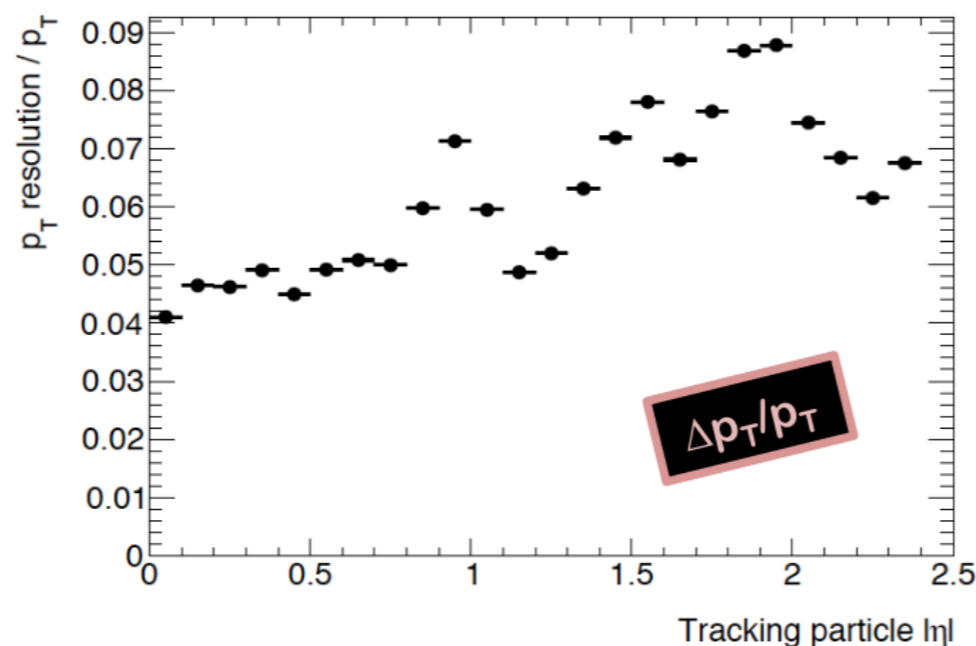
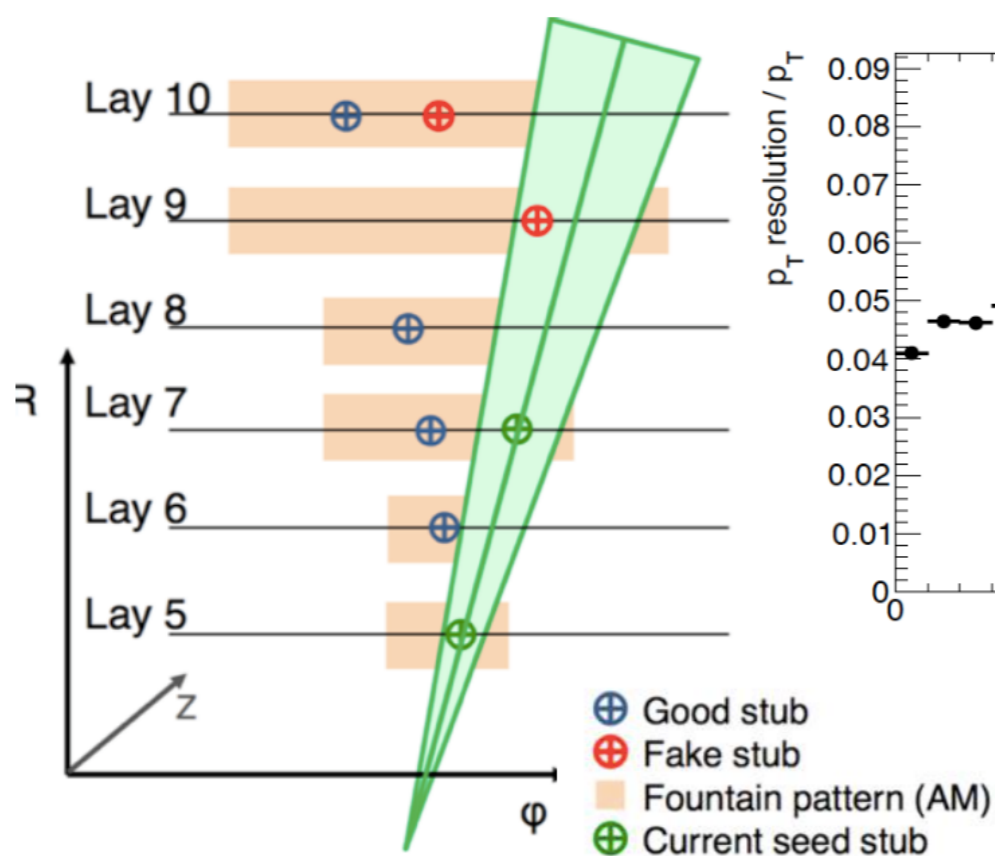
## Track Candidate Builder

## Track Fitter



- Fmax of 500MHz+
  - clock cycle of 2ns or less
- Latency of 47 clock cycles
  - 94ns @500MHz
- One fit/clock cycle after that initial latency
  - 2ns/fit @500MHz

- It selects one combination for each set of stubs of a matched road
- Use the innermost PS modules to build seeds, which are then extrapolated to the outer layers, where compatible stubs are searched for and the one closest to the extrapolation are retained
- Very efficient (in  $t\bar{t}$ +PU200 is 98.5%), it provides excellent track parameters determination, used in the PCA track fitter
- After TCB, only 5% of the original stubs are retained, 70% of which belong to a primary particle



## Principal component analysis

Track parameters:  $p_i = \sum A_{ij} x_j + B_i$

## Fit separately $r$ - $\phi$ and $r$ - $z$ views

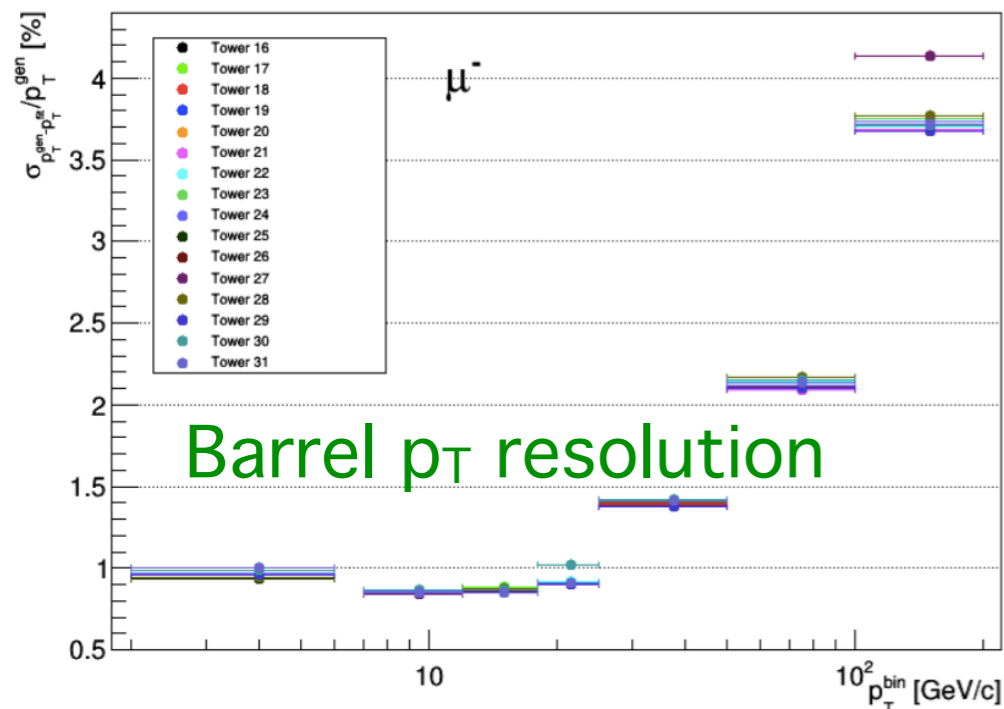
R- $z$ : 20 bins in  $\eta$  (size of 0.05) - only use precise PS modules information

- 20 set of constants (also including 2/3)

- $z_0$  resolution better than 1 mm

R- $\phi$ : 2(charge) x 7( $p_T$ ) bins (will be more) - from the TCB

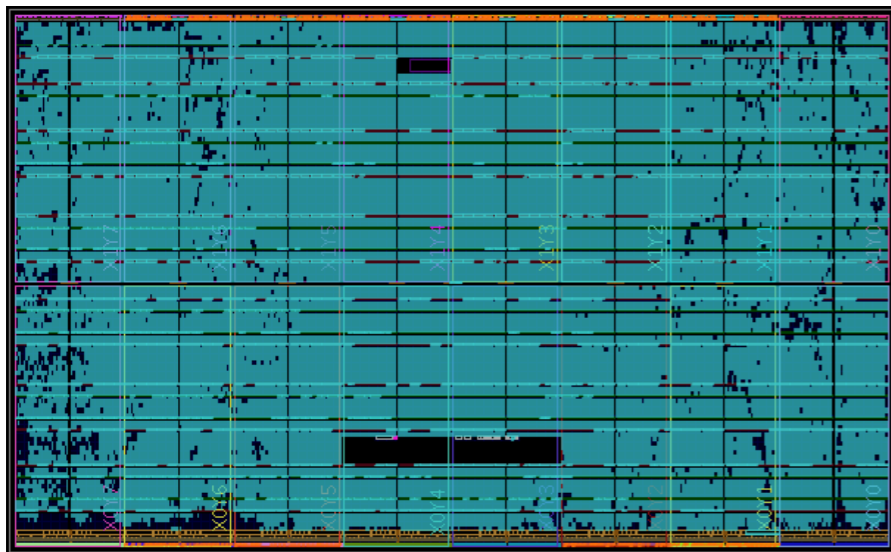
- 98 set of constants (also including 5/6)



rz plane 6/6	
	20 bins in $\eta$
$\Delta\eta$	0.0024
$\Delta z_0$ cm	0.089
r $\phi$ plane 6/6	
	7 bins in $p_T$
$\Delta\phi$ rad	0.00022 to 0.0018
$\Delta c/p_T$	0.8% to 4.1% 0.8% to 3.7% (without tower 27)

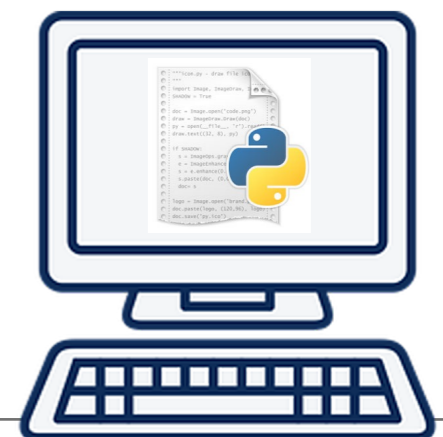
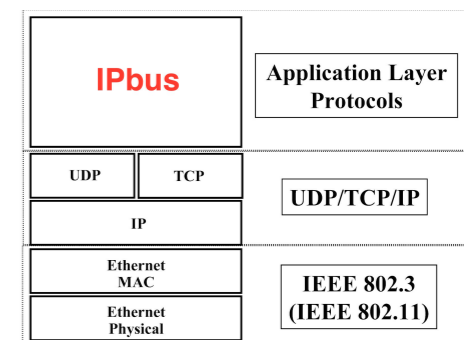
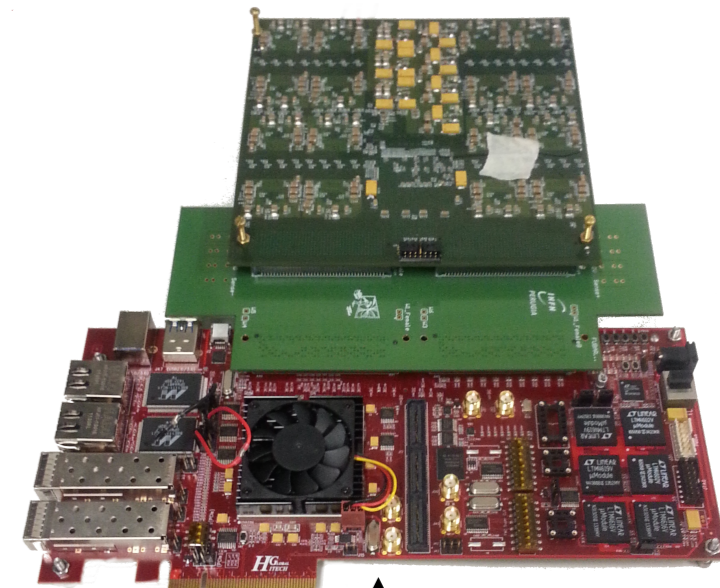
rz plane 5/6	
	20 bins in $\eta$
$\Delta\eta$	0.0024 (6) to 0.0045 (7)
$\Delta z_0$ cm	0.090 (6) to 0.17 (5)
r $\phi$ plane 5/6	
	7 bins in $p_T$
$\Delta\phi$ rad	0.00024 to 0.0018 (9) 0.00031 to 0.0019 (10)
$\Delta c/p_T$	0.8% to 4.2% (9) 1.2% to 6.4% (10)

- **Includes entire reconstruction chain (all blocks integrated)**
- **Operates with 16 AM05 chips.**
- **Features:**
  - Multiple clock domains: DO and TF @ 200 MHz, TCB @ 100 MHz
  - Per layer DC bits
  - Partial trigger tower coverage (32k patterns - though covering ~30% of the full tower)
  - Missing layer (5/6) handling
- **Resources:** about 50% of the Kintex7 resources have been used. Expect to have better performance once ported to PRM06 (Kintex Ultrascale 060)
- Power consumption with the loaded FW and AM chip configured: less than 50 W



Resource	Utilization	Available	Utilization %
LUT	157413	298600	52.72
LUTRAM	23745	108600	21.86
FF	277107	597200	46.40
BRAM	669	955	70.05
DSP	503	1920	26.20
IO	88	380	23.16
GT	24	28	85.71
BUFG	13	32	40.63
MMCM	1	8	12.50

- Tested using a Virtex6 evaluation board (limitations on GTX speed and FMC connectors 1 HPC, 1 LPC)
- **Bank file**: 32k patterns of the most probable patterns out of 0.5M patterns for barrel TT (18)
- **Event root files**: single muon, and complex events
- Simulated **track parameters are in agreement** with what we obtain from the actual implementation of the hardware. Small differences between software simulation and FW results (**level of %**) due to integer-float representation differences between the simulated and implemented TF.
- Validation of the firmware and simulations is ongoing.



- Processing time , from the first road out  
AM05 chips to the last track out:  $\sim 2.1 \mu\text{s}$ ,  
reducible with faster clocks**

# 1 busy $t\bar{t}$ +PU140 event, $\max(\text{stubs/layer})=79$ , matched roads from multiple chips





# Processing time in ModelSim simulation



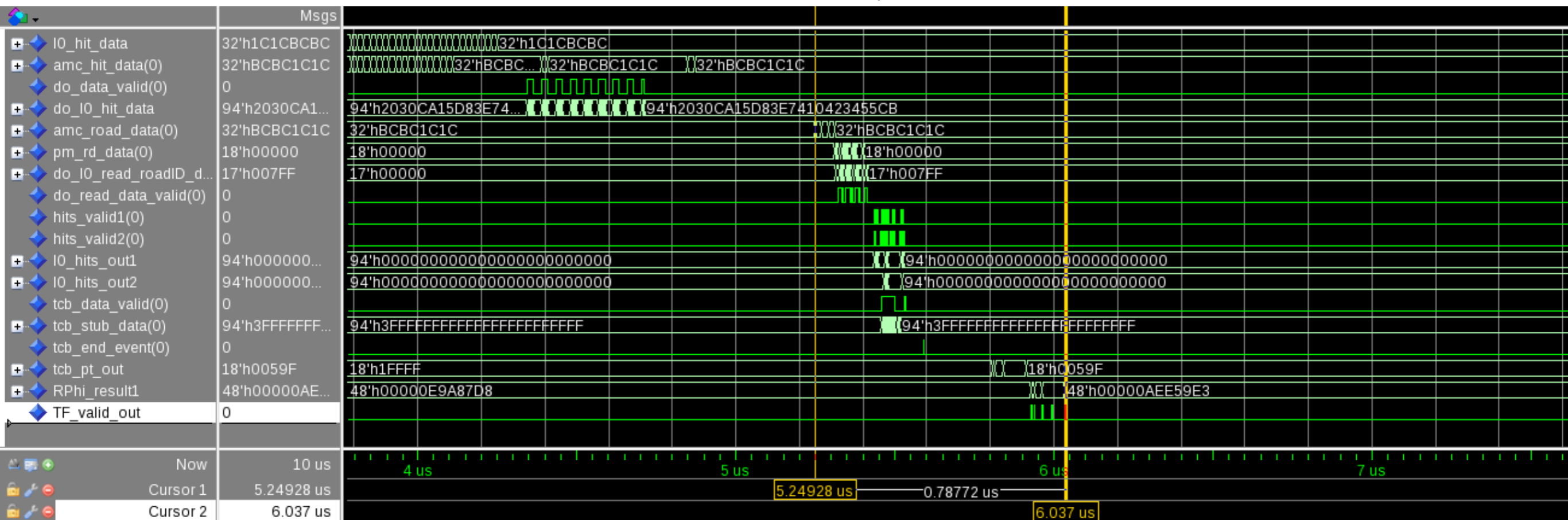
- Processing time measured for a complex events in simulation (ModelSim) with the same banks and chip configuration
- Extrapolation of the processing time in case of faster clocks. Case study:
  - AM chip: same frequency of AM05
  - Data Organizer: 400 MHz
  - TCB: 300 MHz
  - Track Fitter: 500 MHz
- The same event of previous slide has been sent to simulated FW/HW

Processing time , from the first road out AM05 chips to the last track out reduce to:  
**0.8  $\mu$ s**

**1 busy  $t\bar{t}$ +PU140 event, max(stubs/layer)=79, matched roads from multiple chips**

First road  
out AM05 chips

Last track  
out TF

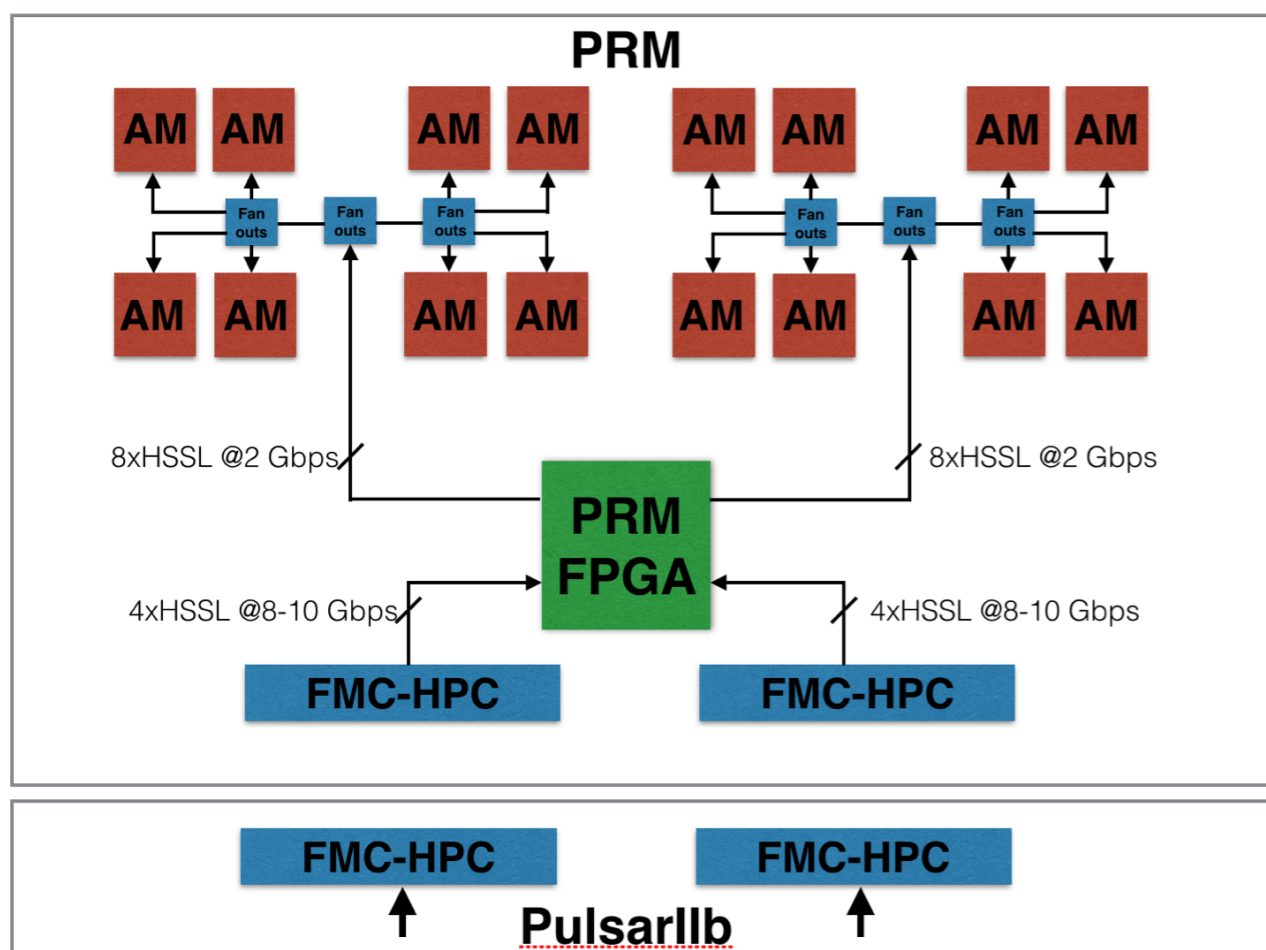


- The Associative Memory + FPGA based demonstrator for the Level-1 Track Trigger of CMS is based on partitioning the tracker into  $6(\eta) \times 8(\phi)$  towers, with a factor 20 time-multiplexing
  - Each tower requires between 500k to 1M patterns, corresponding to between 4 to 8 AM06 chips
  - Pattern Recognition Mezzanines have been developed by INFN aiming to demonstrate the ability to reconstruct tracks with full tower number of patterns with state of the art technology (AM06 chip and KU060 FPGA)
  - The full FW has been developed, integrated and tested with 32k patterns and is being ported to 1.5M patterns mezzanine and well on track for the demonstration
  - The latency measured in the HW indicate that the target of  $4\mu s$  reconstruction is well within reach
- The next month will be crucial for demonstrating the full size pattern behaviour

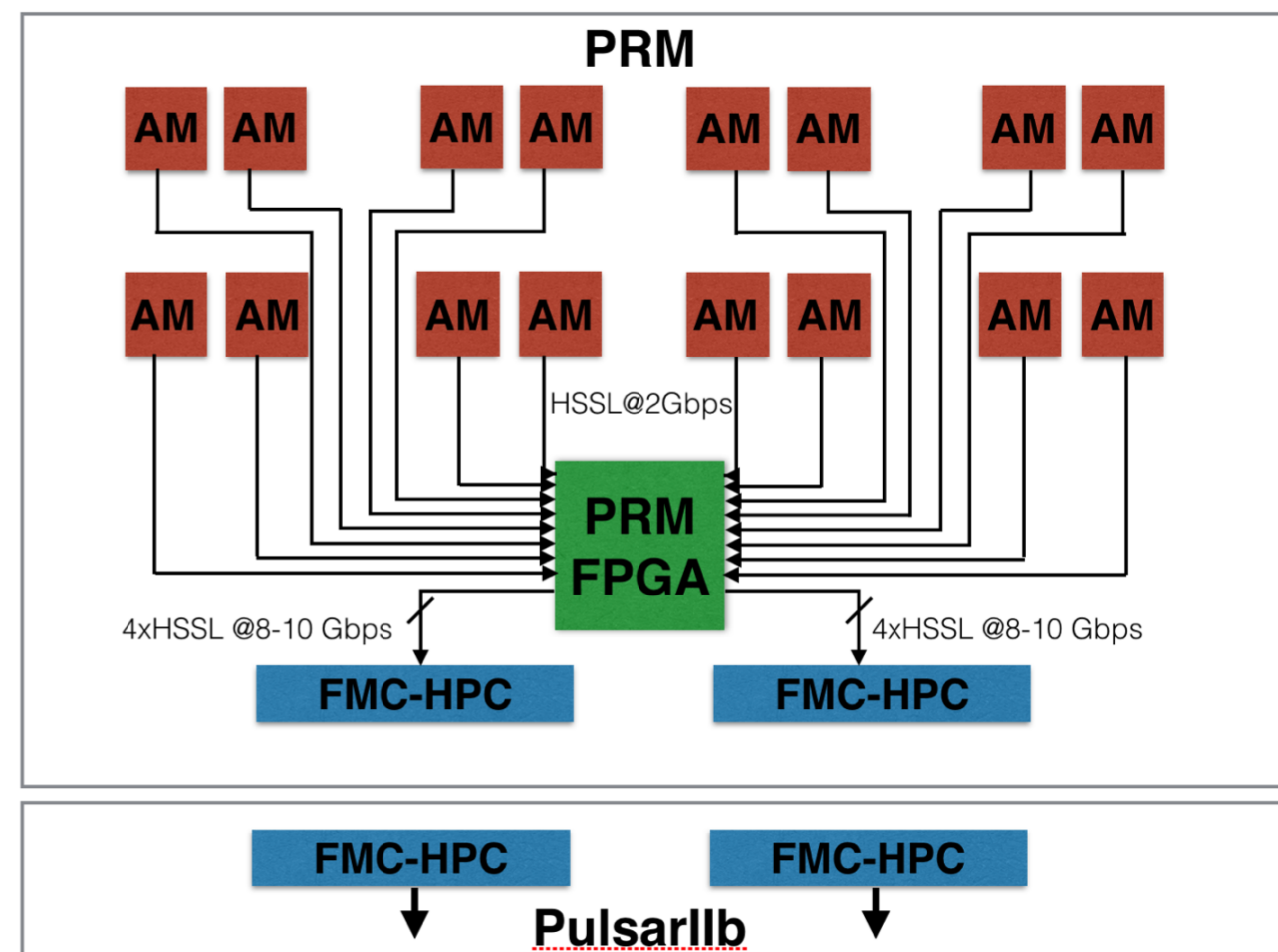
*Backup material*

- **Basic firmware with basic interfaces is ready**
  - Serial links (GTH)
  - Basic memory interface
  - I2C and Flash interfaces
- **Ongoing developments:**
  - Data Organizer  $\Leftrightarrow$  External-memory interface
  - PRM06 communication using IPBus: master sitting on Evaluation board FPGA and slave registers in the PRM FPGA. It is currently working fine on PRM05.
- **Firmware migration (DO, TCB, TF, FSMs,...):**
  - Move the current modules with the internal pattern memory (limited number of patterns: 32k patterns)
  - Use the external memory (full trigger-tower bank coverage)
  - Increase the frequency of the clock domains. Target freq. in ref. [1]

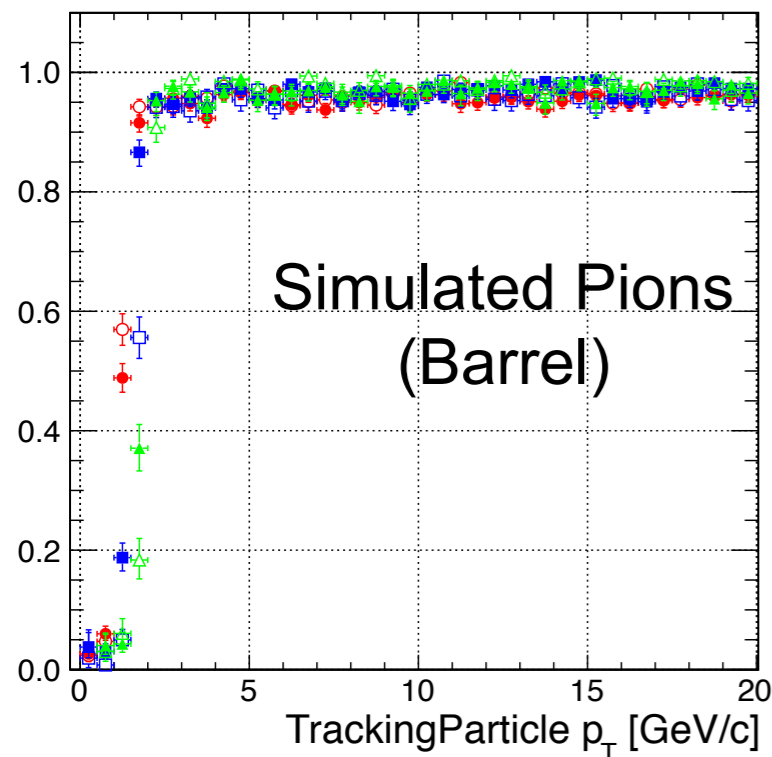
## Data input



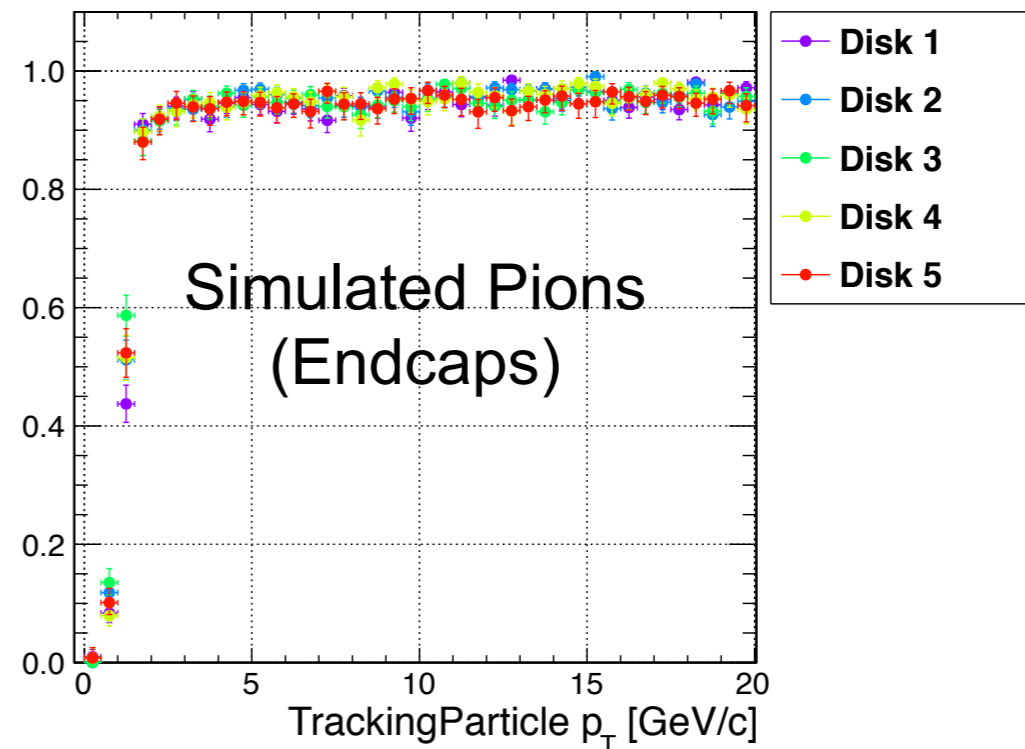
## Data output



Efficiency



Efficiency



Prototype module test beam at DESY (2-4 GeV  $e^+$ )

