# Development in DAQ and Triggering

**Jinlong Zhang**

**Argonne National Laboratory**

**TIPP'17, May 22-26, 2017, Beijing**

# Disclaimer

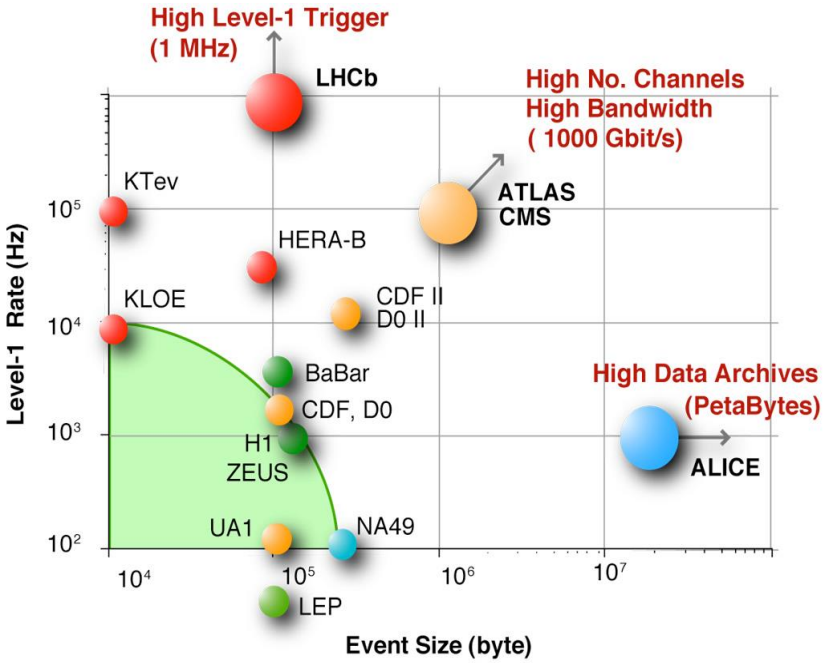
- This may not be the presentation you originally expected, not even the conference originally planned
- The theme is LHC experiment abundant
- Some statements are my personal bias
- I was not able to cite the references properly for the materials included

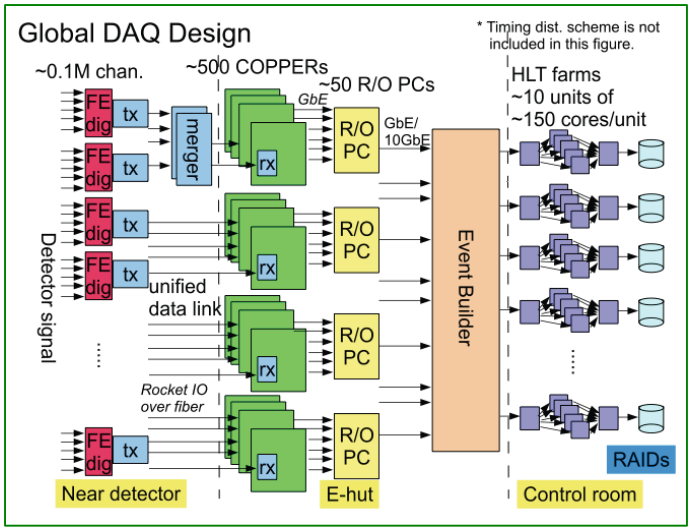
**My apology**

# Introduction

- Trigger/DAQ system overview
- Developing in trigger
  - Triggerless scheme
  - Specific aspects (track, global, timing)
- Developing in DAQ
  - Accessing commodity (PCIe)
  - Storage evolution
- Trends
  - Accelerator (GPU, CPU+FPGA)
  - Common platform
- Summary

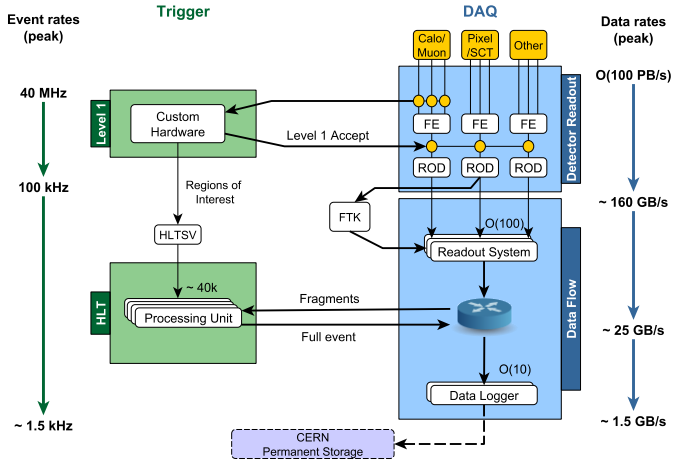
# Collider Experiment Examples

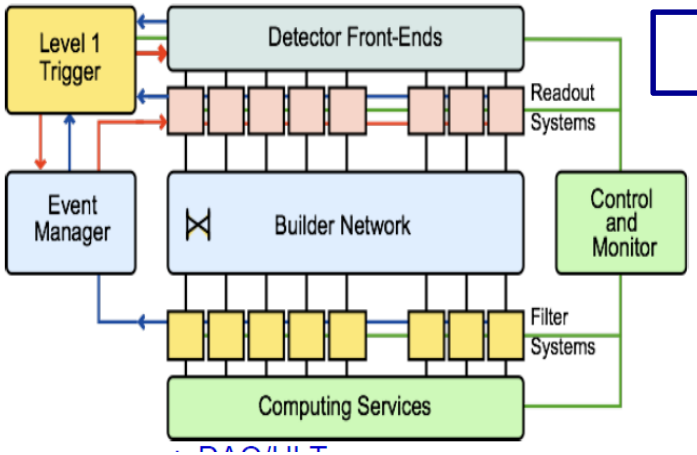



**BELLE II**

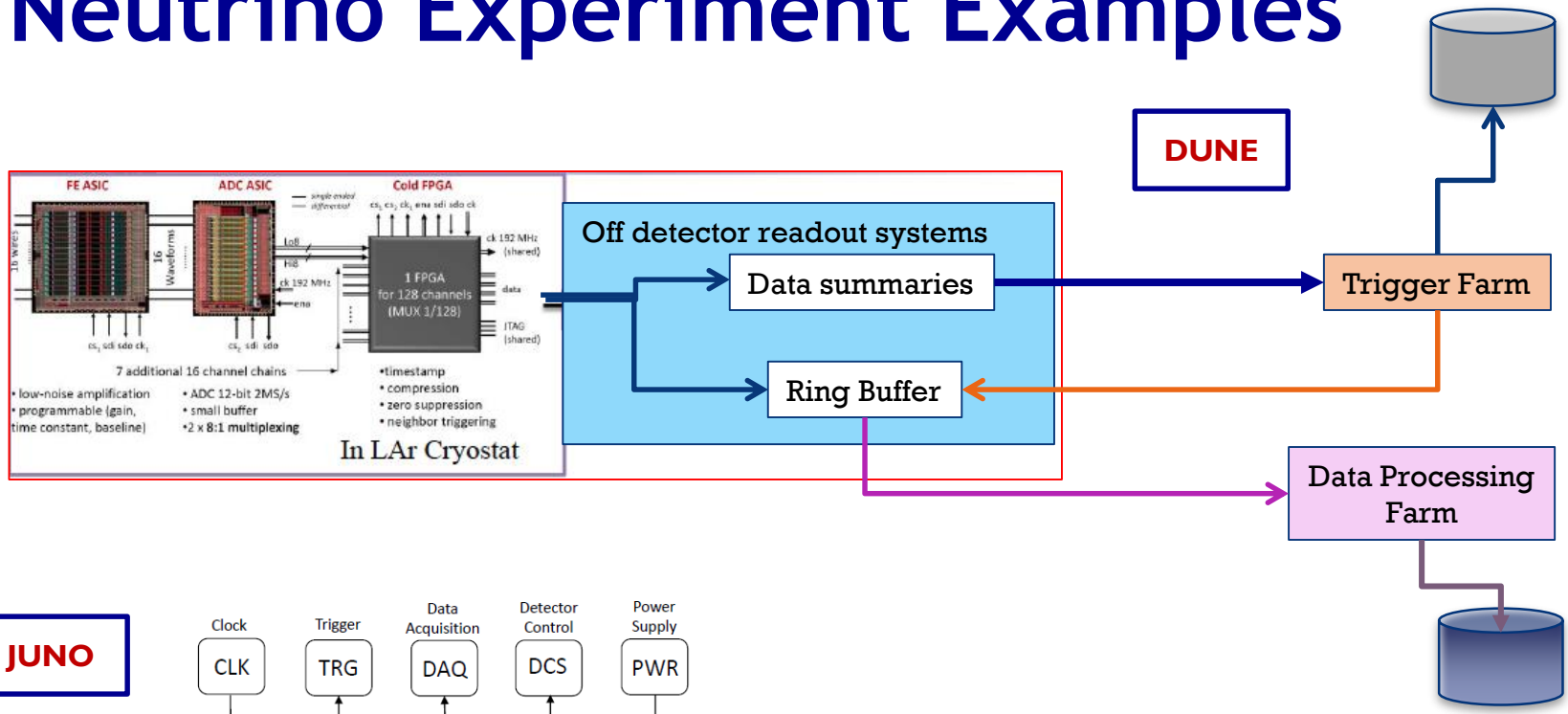
**ATLAS**

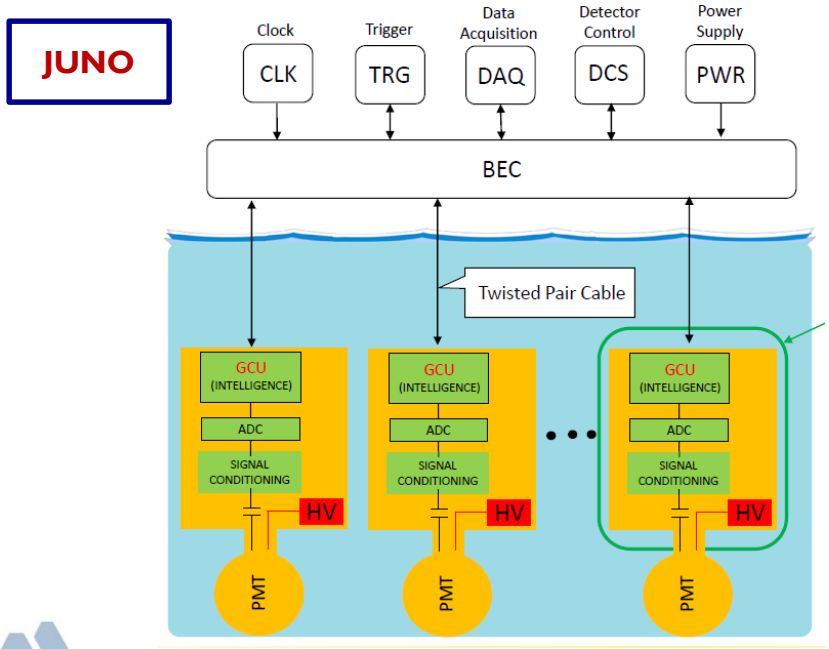



**CMS**

# Neutrino Experiment Examples

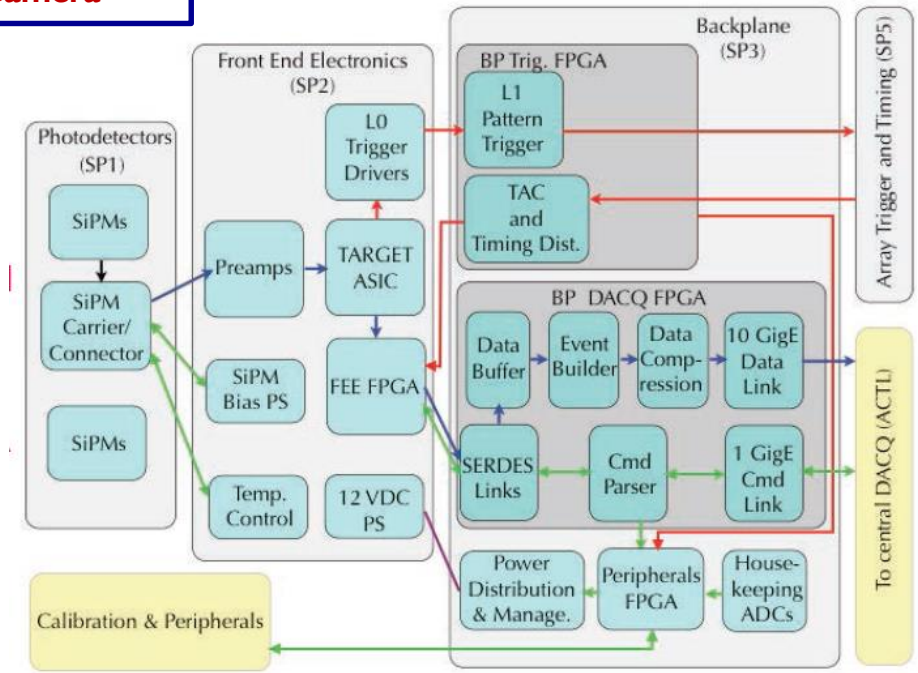


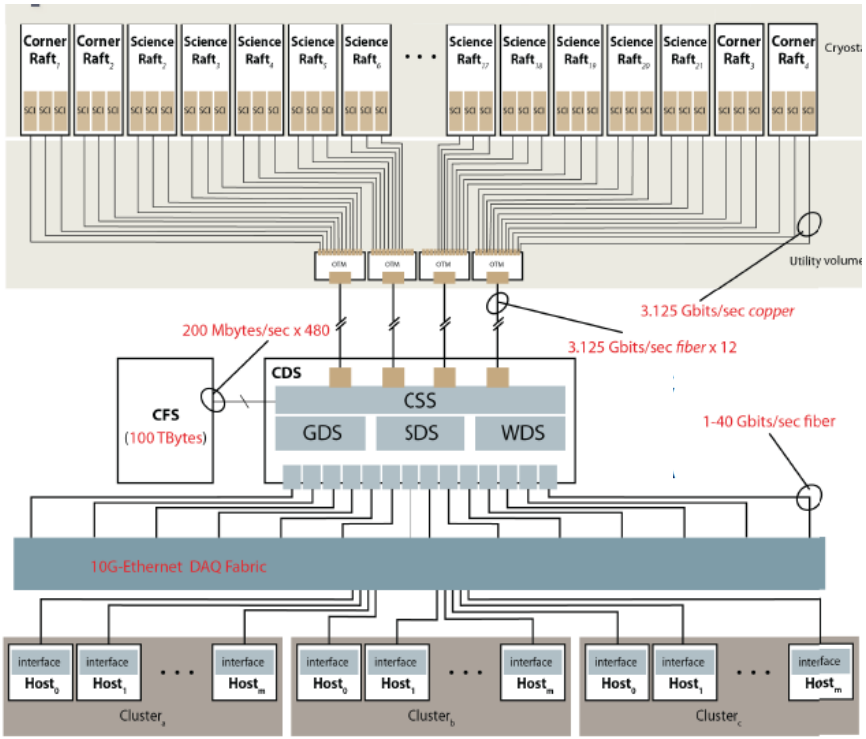

**Trigger not necessarily as complicated as collider experiments, but data throughput comparable**

# Cosmologic Instrument Examples

**CTA Camera**




**LSST 4000 3.2Gbps**




**Data throughput comparable to LHC experiments, or even larger**

# Over-simplified Requirements

- Customized ASICs to handle the detector signals (FE electronics) in the upstream of the Trigger/DAQ
- **Powerful hardware (FPGA based, GPU, CPUs and/or combinations) and software algorithms to perform data reduction (trigger)**
- **High speed links, huge computing capacity and storage space to handle the event data (DAQ)**
- **Enabled by**
  - **Moore's Law (CPUs, also FPGAs and GPUs)**
  - **Link technology (transceivers, networking)**
  - **Storage technology**

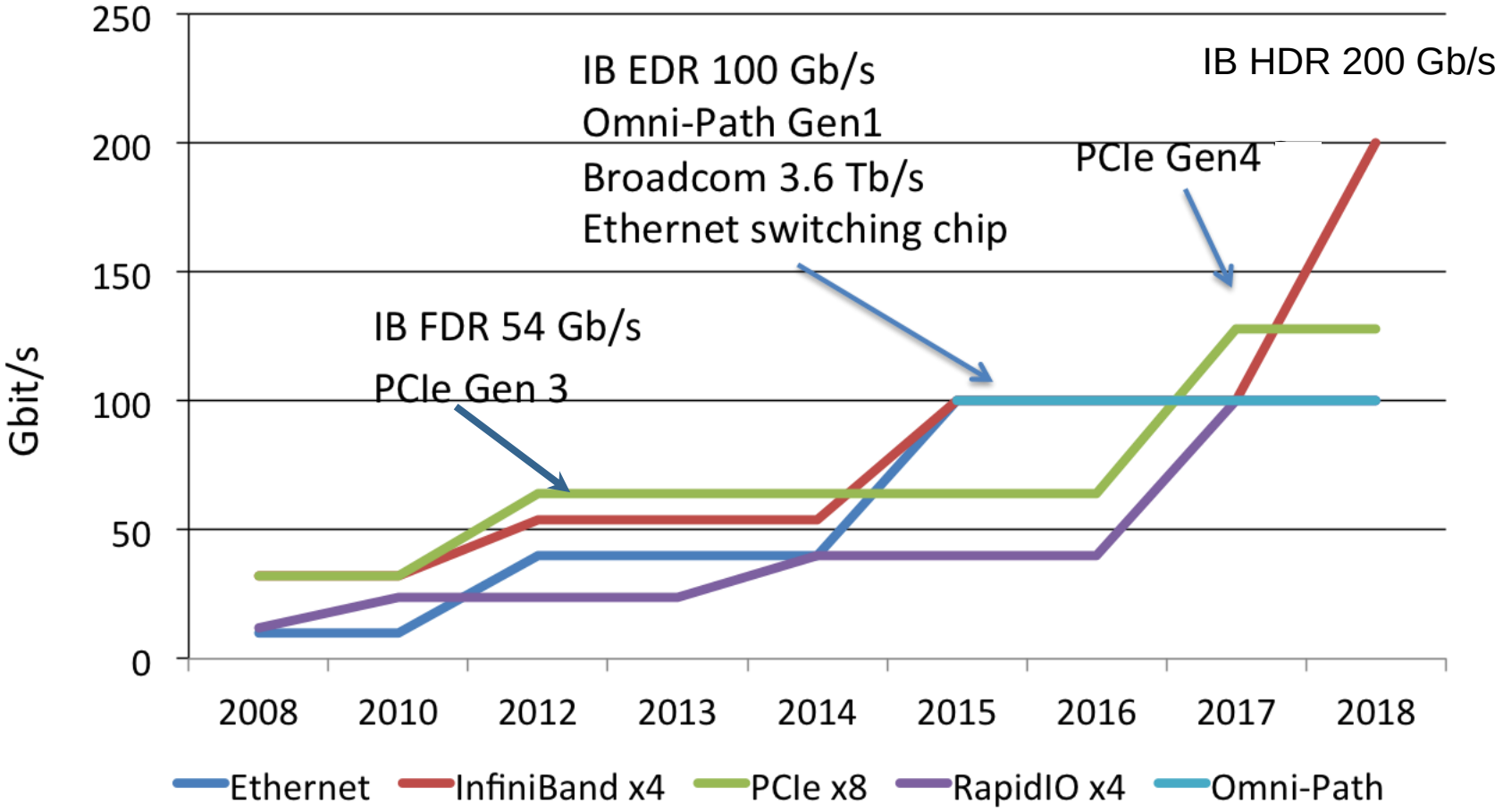
# Link to Upstream

http://www.xilinx.com

| | Type | Max Performance[1] | Max Transceivers | Peak Bandwidth[2] |
|---|---|---|---|---|
| Virtex UltraScale+ | GTY | 32.75 | 128 | 8,384 Gb/s |
| Kintex UltraScale+ | GTH/GTY | 16.3/32.75 | 44/32 | 3,268 Gb/s |
| Virtex UltraScale | GTH/GTY | 16.3/30.5 | 60/60 | 5,616 Gb/s |
| Kintex UltraScale | GTH/GTY | 16.3/16.3 | 64 | 2,086 Gb/s |
| Virtex-7 | GTX/GTH/GTZ | 12.5/13.1/28.05 | 56/96/16[3] | 2,784 Gb/s |
| Kintex-7 | GTX | 12.5 | 32 | 800 Gb/s |
| Artix-7 | GTP | 6.6 | 16 | 211 Gb/s |
| Zynq UltraScale+ | GTR/GTH/GTY | 6.0/16.3/32.75 | 4/44/28 | 3,268 Gb/s |
| Zynq-7000 | GTX | 12.5 | 16 | 400 Gb/s |

- **Readout system will utilize these serDes speeds or faster, so**
- **High speed radiation hard link need be developed**
  - **lpGBT modest**

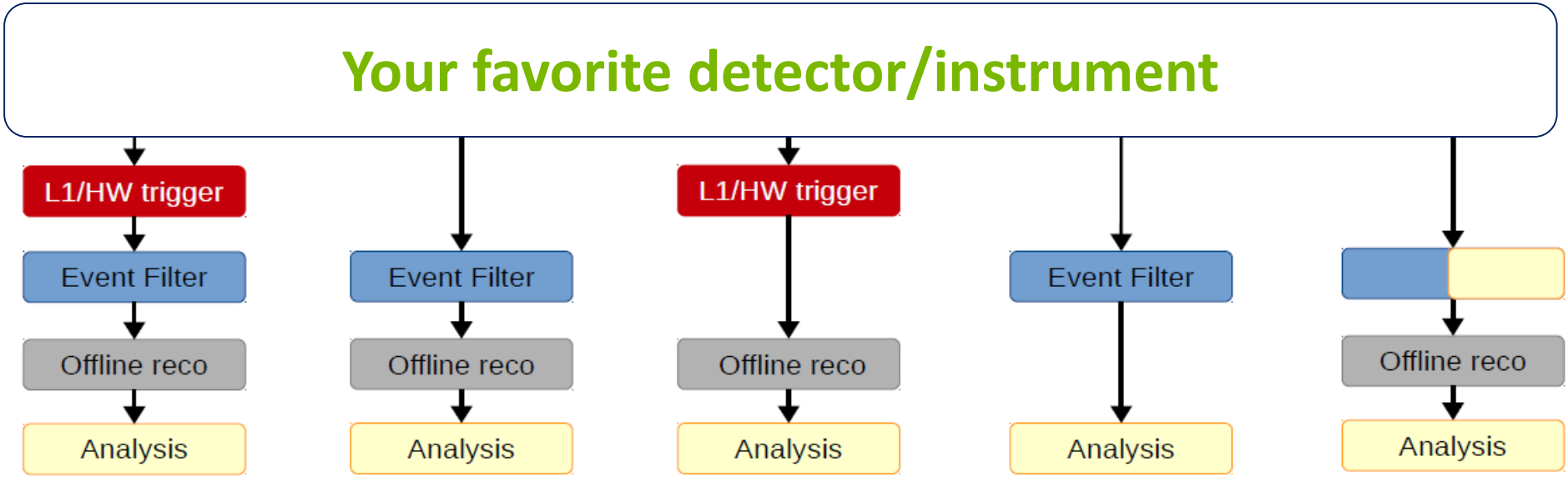
# Link in Downstream




- **Network for hundreds of 100 GBE links not a problem soon**
- **PCIe Gen4 expected in later 2017**

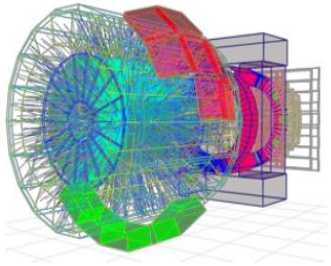
# Changing Paradigms

- No trigger (triggerless) or less trigger levels
- Online Offline fusion
- Better physics performance or enhancing physics capability
- Less/common effort



# Go Triggerless

## ALICE in Run 3



3.6 TByte/s into PC farm

**$O^2$ (Online Offline) System**

Partial calibration and reconstruction online, replacing the original raw data with compressed data

**STORAGE**

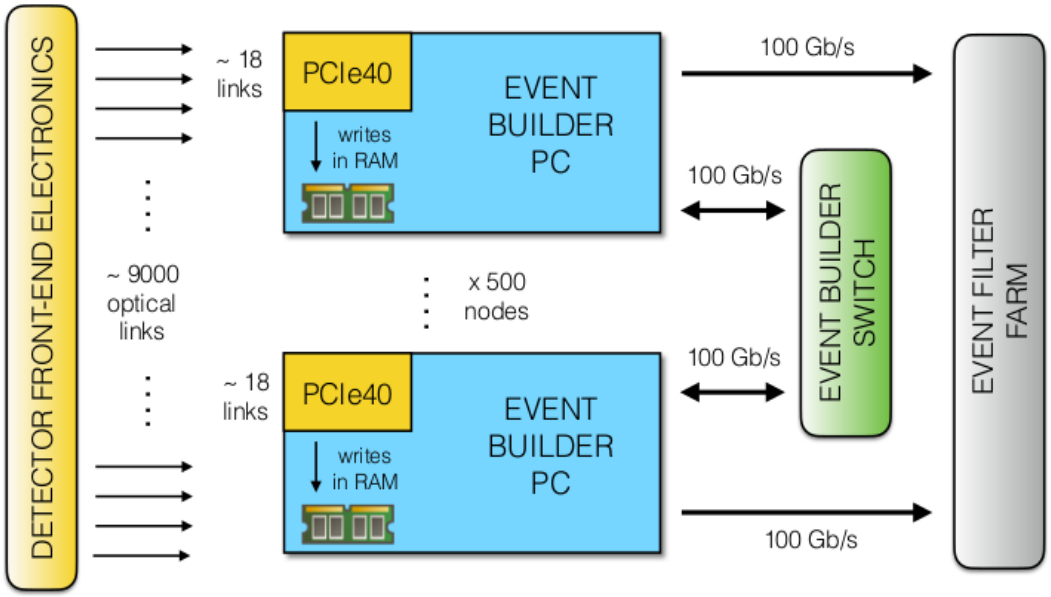
**90 GB/s**

Acq rate:
**Pb-Pb 50 kHz**
**pp and p-Pb up to 200 kHz**

Complete change in detector readout
- continuous
- triggered

New DAQ - HLT - OFFLINE systems.

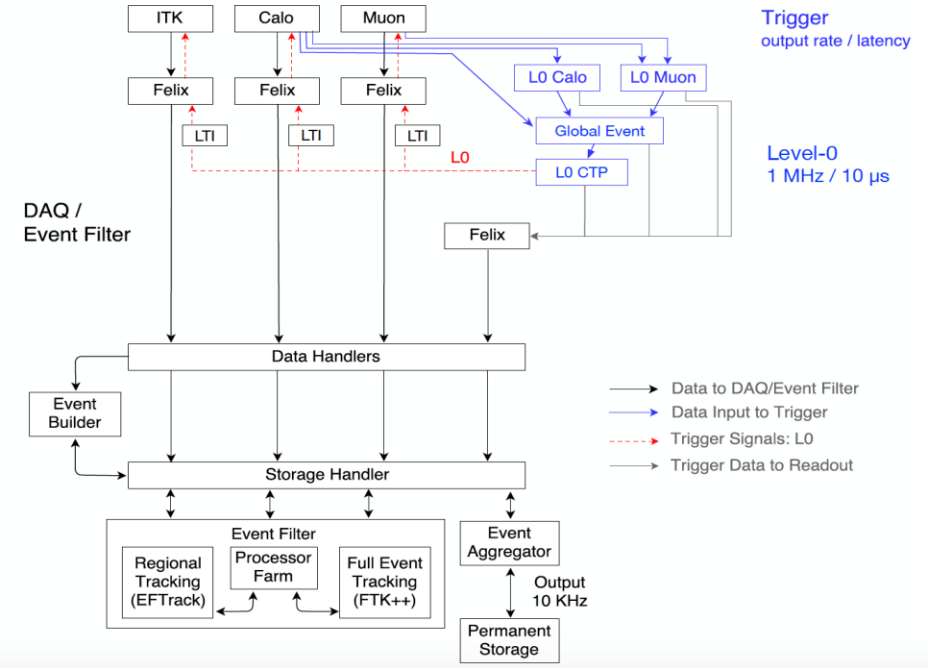
## LHCb in Run 3



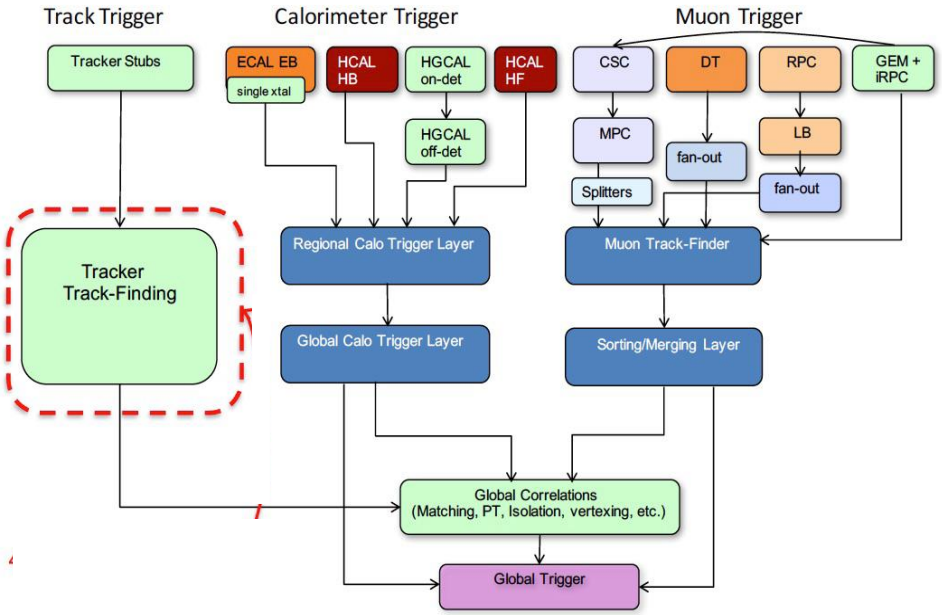

| | 2015 | | 2018 |
|---|---|---|---|
| Event size | 65 KB | → | 100 KB |
| Event rate | 1 MHz | → | 40 MHz |
| Aggregate bandwidth | 520 Gb/s | → | 32 Tb/s |

# Triggerless Not Yet Possible

**ATLAS in Run4**

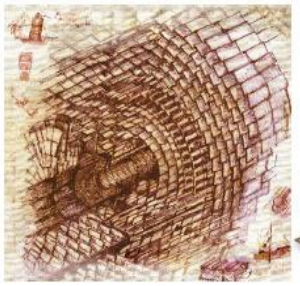



**CMS in Run 4**




- **Track trigger**
  - **ATLAS seeded, regional @ 1 MHz and full event @ 100 kHz**
  - **CMS self seeded, @ 40 MHz and latency of ~4 µs**
- **Global Event Processing with more/finer input**
- **Possible to use precise timing**

# Track Trigger



Data transfer(50-100 Tbs)

Partition detector into trigger regions
Bring data from each region to the corresponding processing engine

**Associative Memory (AM) or FPGA based**

Data formatting

Fit hits in FPGA to determine track parameters

Δt1

Pattern Recognition

Δt2

***Total processing latency Δt ?***

Track fitting and duplicate removal

Δt3

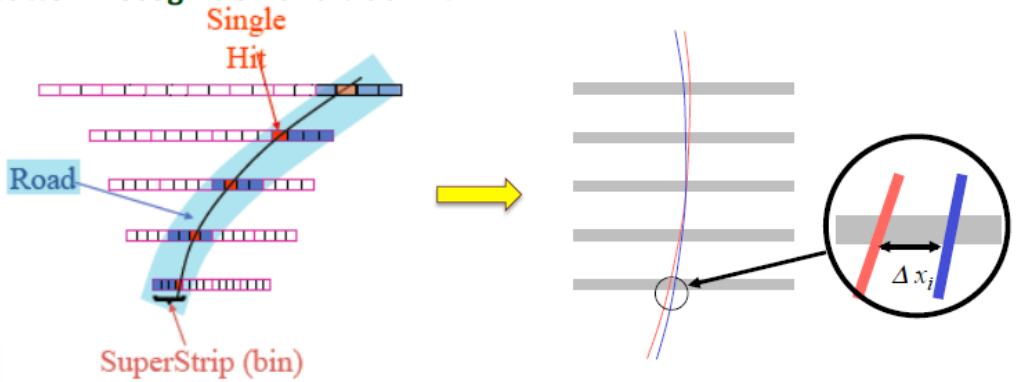
Tracks out

**Also others ways to perform tracking: HEP track finding with Micron Automata Processor, Artifical Retina processor, … …**

**And the buzzword: Deep learning**

# AM Approach

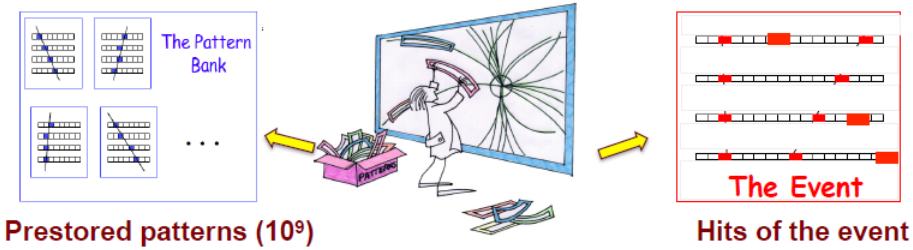
Use hardware to perform the global tracking in two steps
pattern recognition and track fit




Pattern recognition in coarse resolution (superstrip➔road)

Track fit in full resolution (hits in a road)

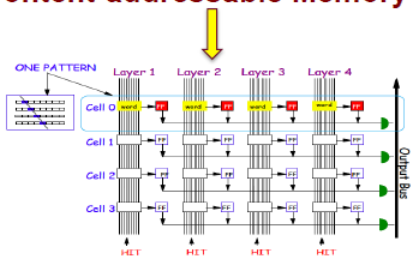
$F(x_1, x_2, x_3, \ldots) \sim a_0 + a_1\Delta x_1 + a_2\Delta x_2 + a_3\Delta x_3 + \ldots = 0$




Prestored patterns ($10^9$)

Hits of the event

Content-addressable Memory (CAM)

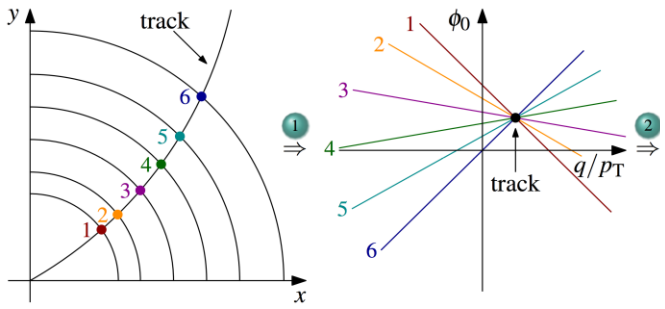


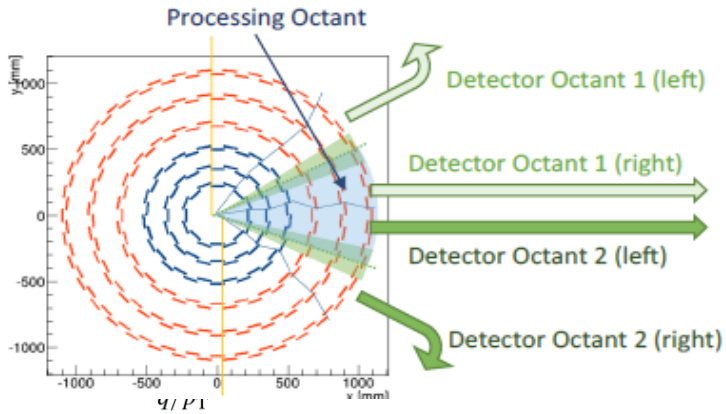

Fast pattern recognition

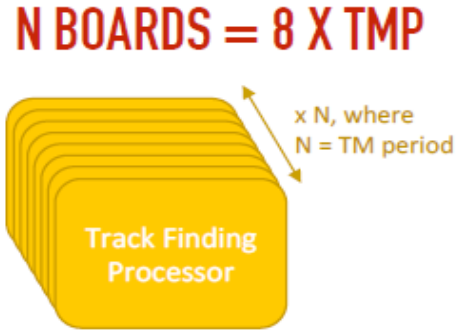
- ATLAS FTK (Phase-I upgrade) with
  - <1 billion of patterns
  - AMChip06 (~128K pattern)
  - <100 μs
- ATLAS hardware trigger in Phase-II upgrade
  - ~10 billion of patterns
  - ~512k patterns per chip

# FPGA Approach

- **The reference option for CMS Phase-II Upgrade**
  - **Hough Transform and Kalman Filter in FPGA**

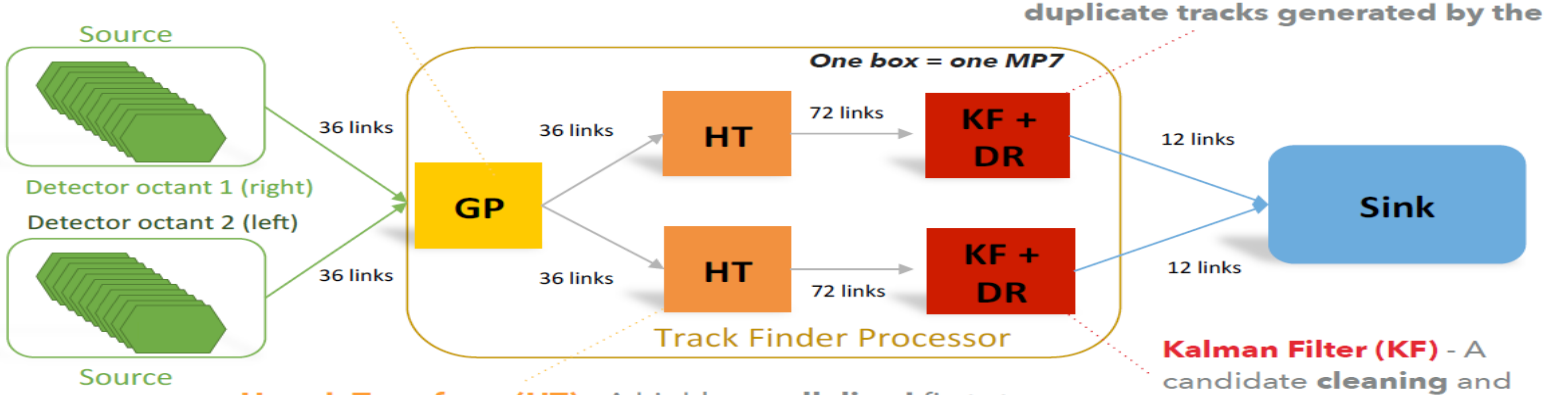



**Geometric Processor (GP)** - **pre-processes stub data**, and **divides the octant** into 36 **finer sub-sectors**

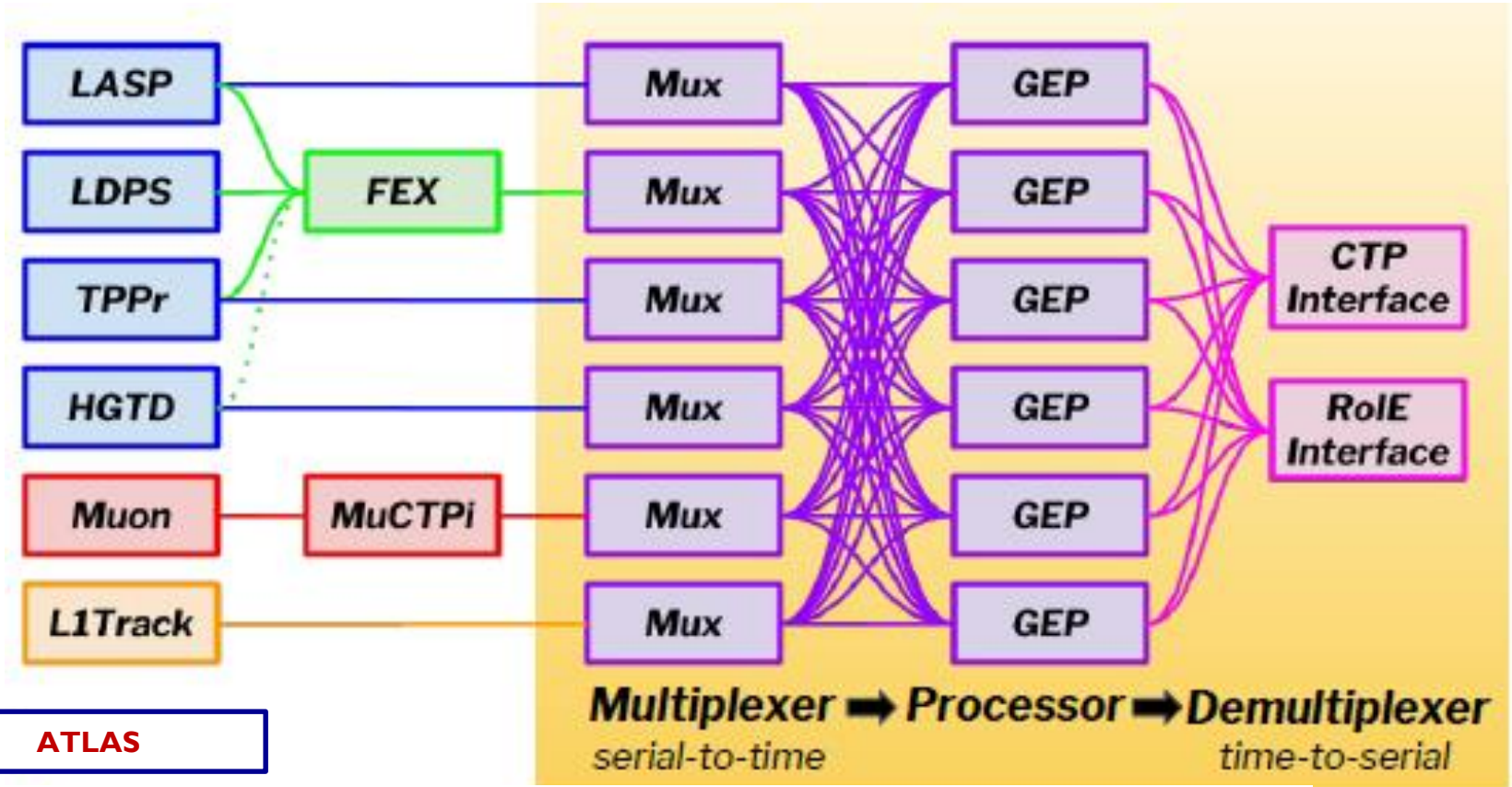
**Duplicate Removal (DR)** - Uses precise fit information to **remove duplicate tracks generated by the HT**




**Hough Transform (HT)** - A highly **parallelised** first stage **track-finder** that identifies **groups of stubs consistent with a track in the r-Φ plane**

**Kalman Filter (KF)** - A candidate **cleaning** and **precision fitting** algorithm

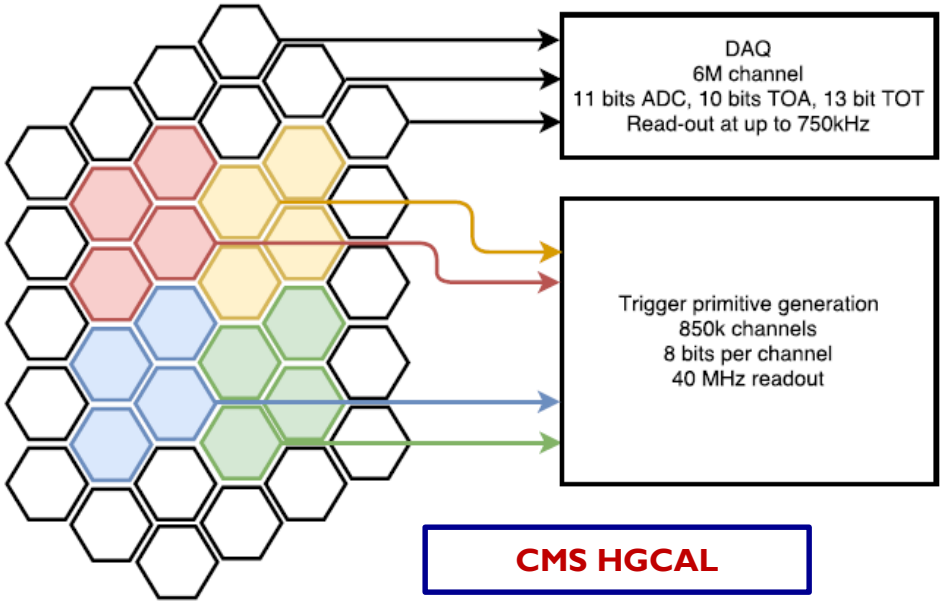
# Global Event Processing



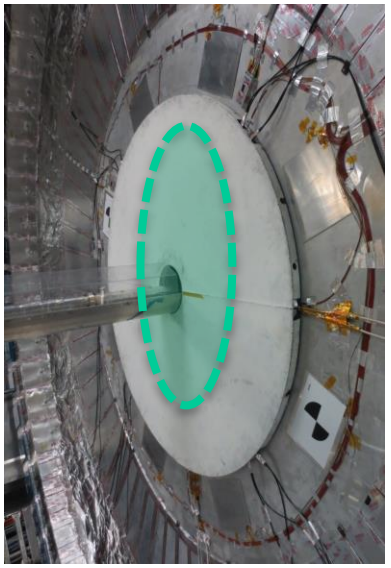

ATLAS

- Data transfers are time multiplexed within the system
  - Increases flexibility
  - Simplifies evolution
  - Maximizes physics

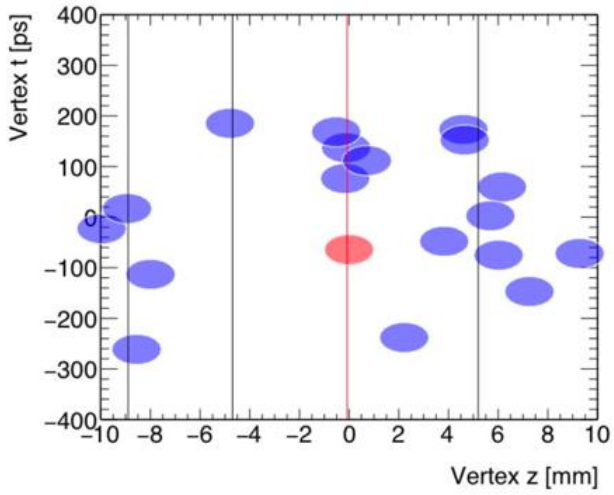
# Precise Timing




CMS HGCAL



ATLAS HGTD

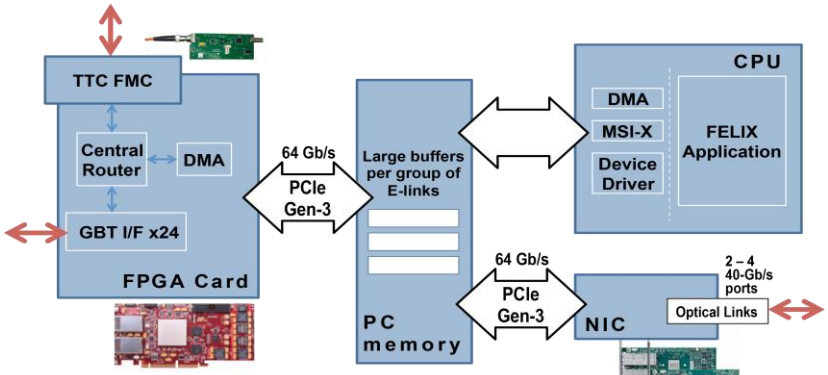



- Using precise timing information in trigger for pileup rejection
- Challenging to achieve the time resolution as a sizeable detector
- Huge data throughput (pixel detector after all)

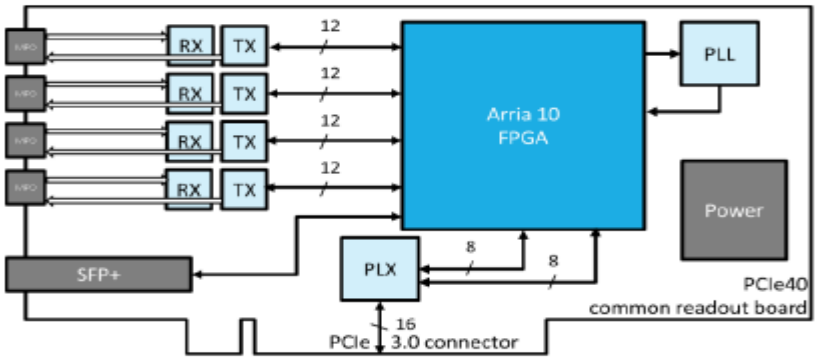
# DAQ in General

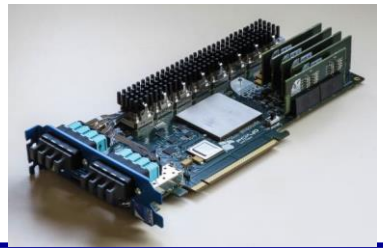
- PC-based data aggregation
  - Ethernet or InfiniBand
  - PCIe
- Network bandwidth becoming very affordable
  - Revisiting the philosophy of “move minimal amount of data”
  - Capability for high event building rate (even decouple from event filtering or other data processing)
- Heterogeneous computing resource (ASIC/FPGAs, GPGPUs, … )
- Tight integration with offline
  - From the blur boundary to the full fusion
  - Better utilization of (online) resources

# I/O Card Utilizing PCIe




**ATLAS FELIX**



**LHCb PCIe40**

CaRIBou as a multi-chip modular DAQ system

System-on-Chip borard (ZC706)

FMC cable (optional)

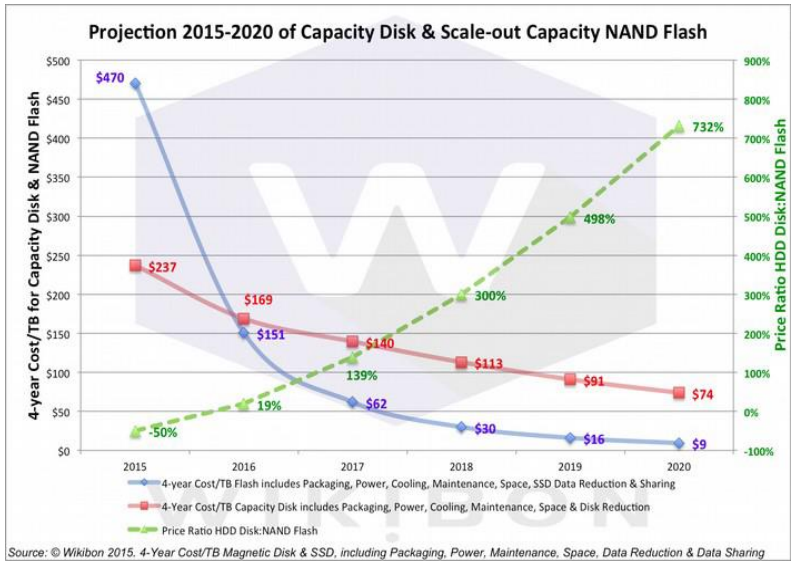
interface board (CaR)

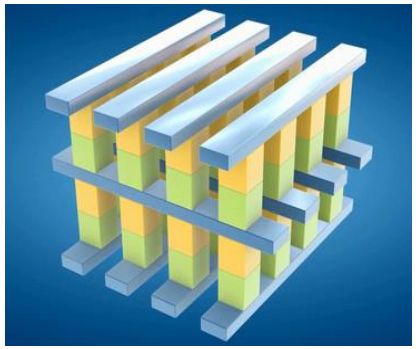
application specific chipboard

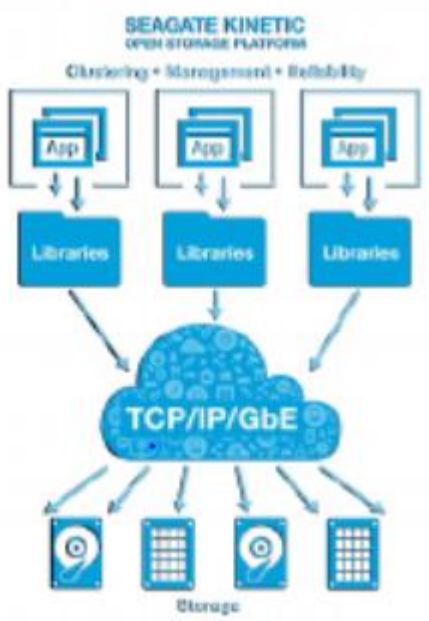
- **COTS hardware may require tweaks to be used, but still could be game changer, particularly for small experiments**

# Storage Evolution

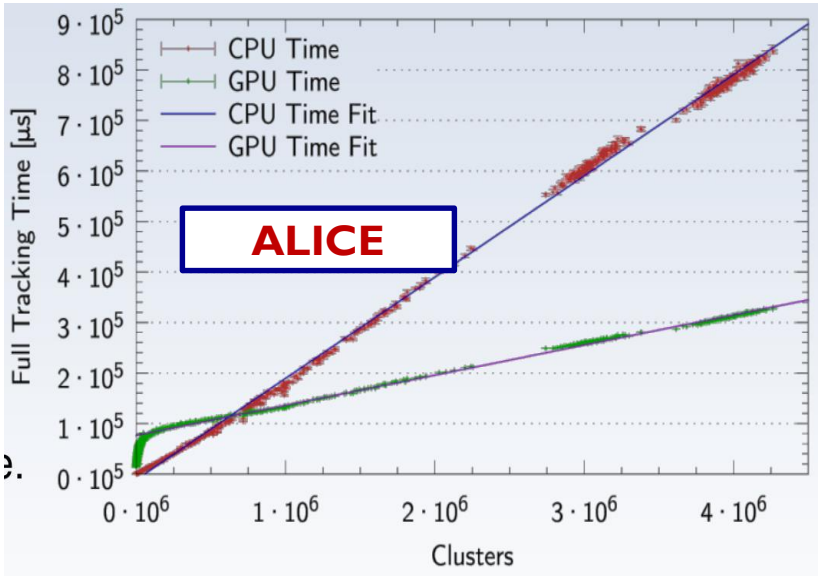
- **Throughput is the real challenge**
- **Real world example exists with current Technology for a system with capacity of ~50PB and throughput of ~5 TB/s**
- **We should look at storage technologies 10 years from now**
- **Evolution of existing technologies**
  - **Consumer NAND drive getting cheaper than spinning drive**
  - **Lustre and GPFS**
- **New technologies**
  - **3D XPoint**
- **Innovations in the storage stack**
  - **Seagate Kinetic, …**



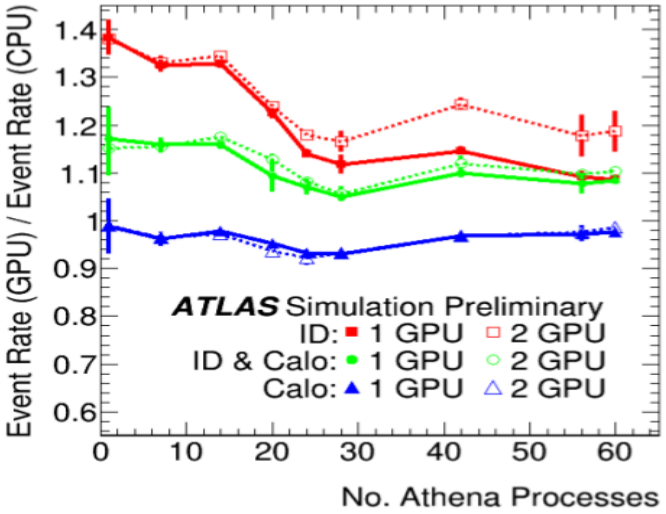





# GPU




ALICE TPC track reconstruction got a factor 2-3 speedup and saved 0.5M USD during Run 1

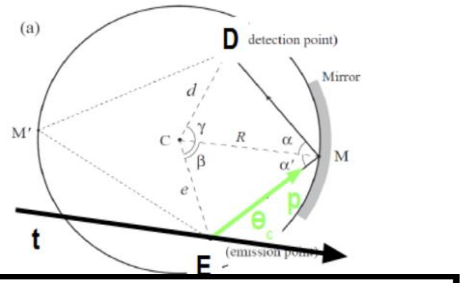



ATLAS HLT 20-40% higher throughput so not yet compelling

- Performance highly dependent on workload
- Could also integrate with other components (NIC) for serious data processing
- Comparison need consider hardware, power, cooling, and effort,… …

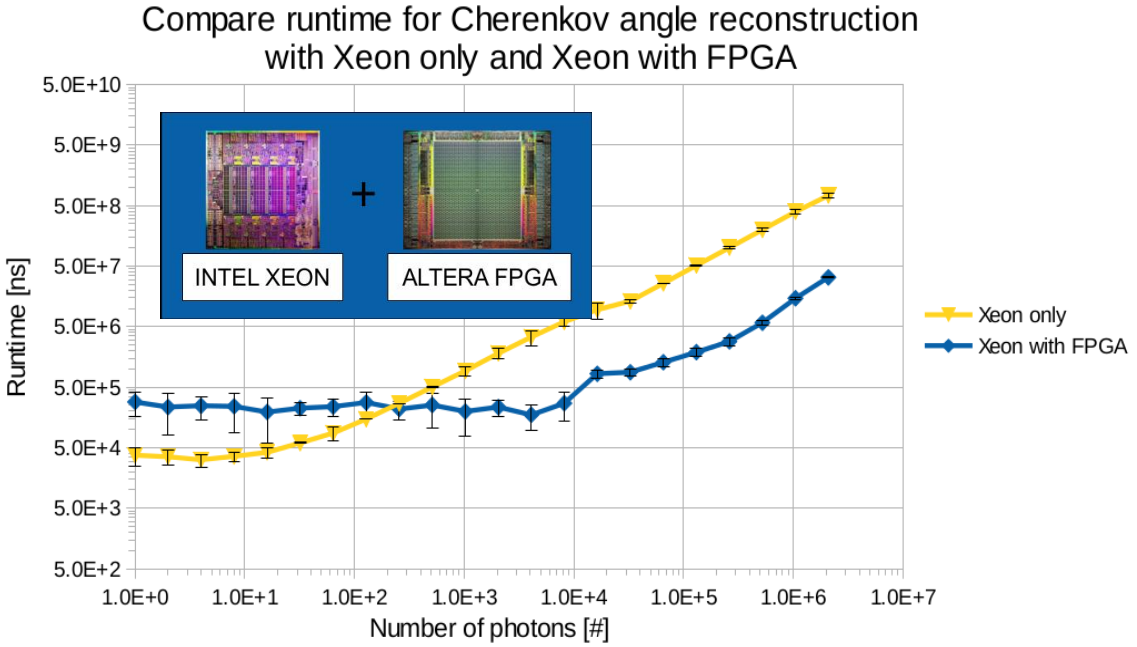
# CPU + FPGA

## LHCb RICH PID algorithm




Calculations:

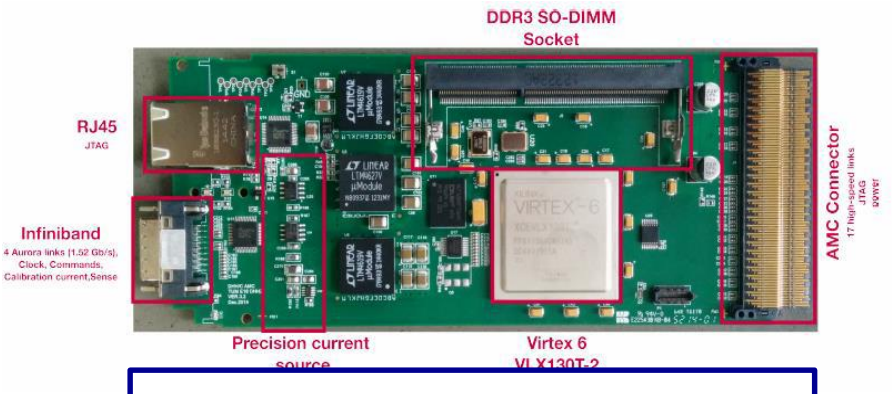
- solve quartic equation
- cube root
- complex square root
- rotation matrix
- scalar/cross products



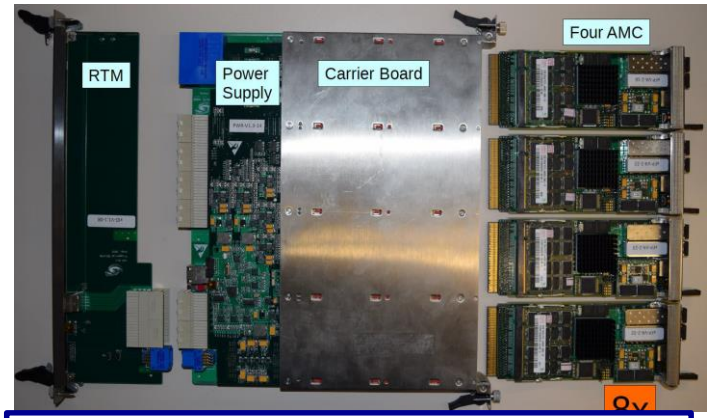

- **Acceleration of factor up to 35 with Intel® Xeon®-FPGA with respect to single Intel® Xeon® thread**
- **Theoretical limit of photon pipeline: a factor 64 for Stratix V FPGA, for Arria 10 FPGA a factor ~ 300**
- **Bottleneck: Data transfer bandwidth to FPGA**

# Common Platform

- Sharing a hardware unit with powerful FPGA(s) and high speed links
  - ATCA/xTCA, PCIe, etc
- Leaving the intelligence differences for firmware and software




**IFDAQ hardware evnt builder**

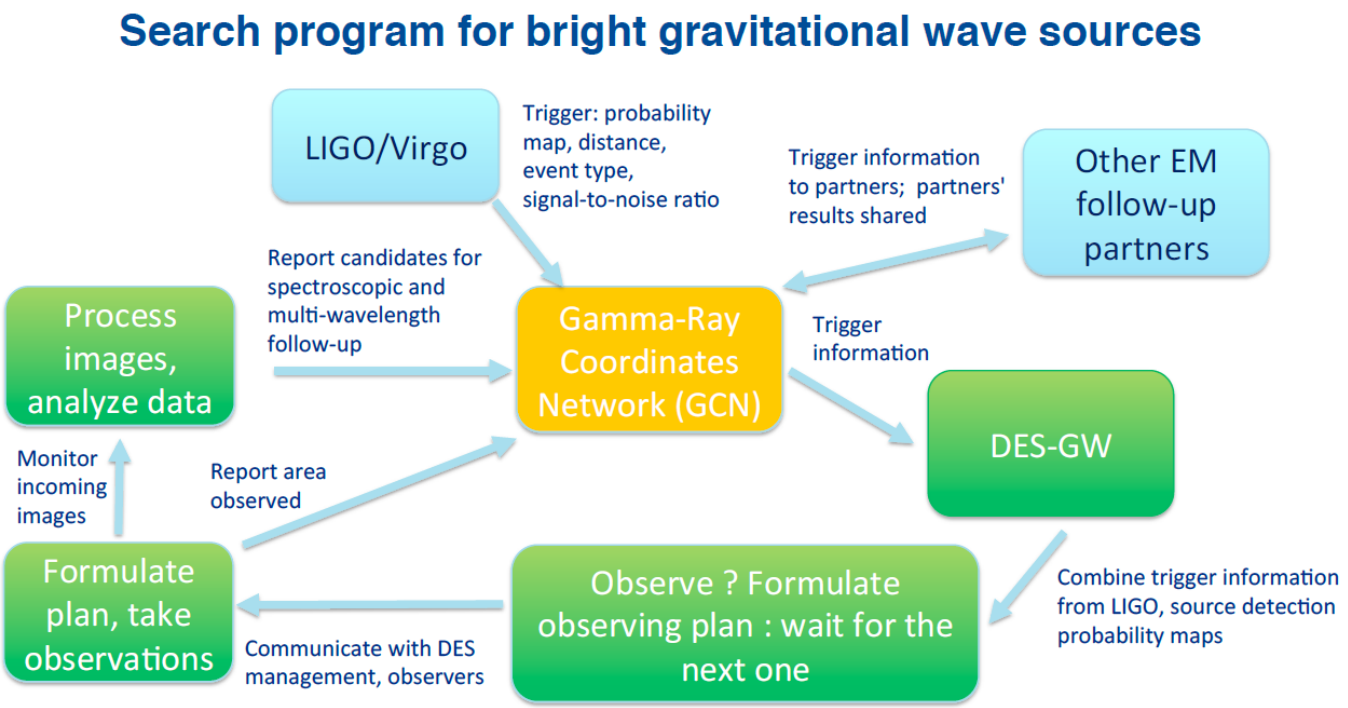



**BELLE II Compute Node**

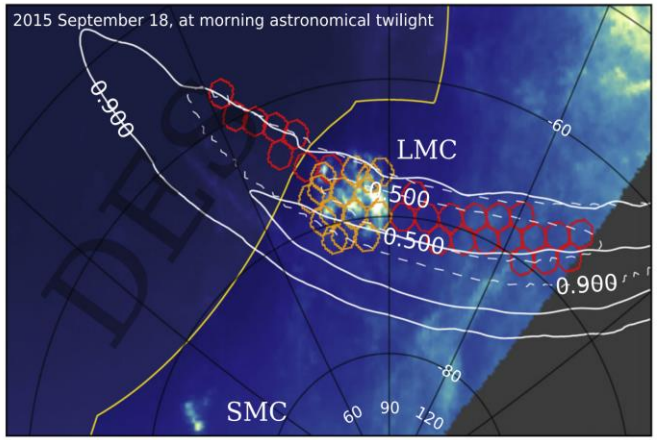
**State of the Art: ATLAS gFEX**

- 30 layer PCB
- 3 Virtex Ultrascale+
- 1 Zynq Ultrascale+
- 35 minPODs

# Trigger(ed by) Others




**DES-GW program using DECam at Chile to perform optical followup of gravitational wave signals from LIGO/Virgo**



# Recommendations From CPAD

- Encourage the development of high-bandwidth radiation hard optical links (>10Gb/s)
- Encourage the development of scalable DAQ system to enable the transition from custom hardware to commodity networking and computing as early as possible
- Encourage the development in hybrid CPU-FPGA, GPGPU, storage, high speed optical and electrical communication
- Encourage studies of the impact of timing information in the trigger at ATLAS/CMS
- Encourage focus on emerging technologies such as photonics and wireless communication