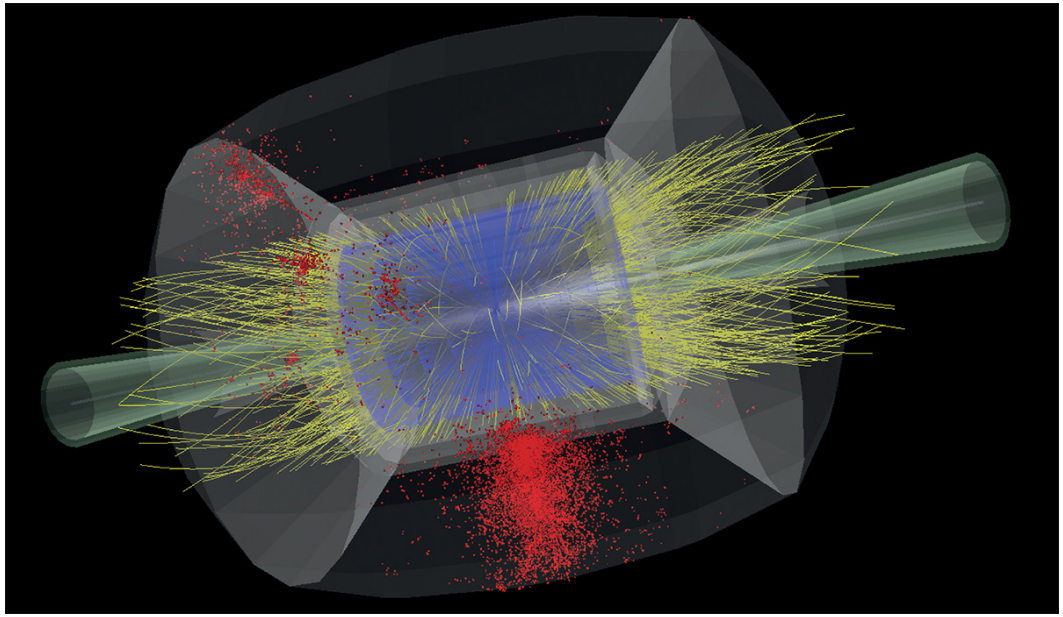


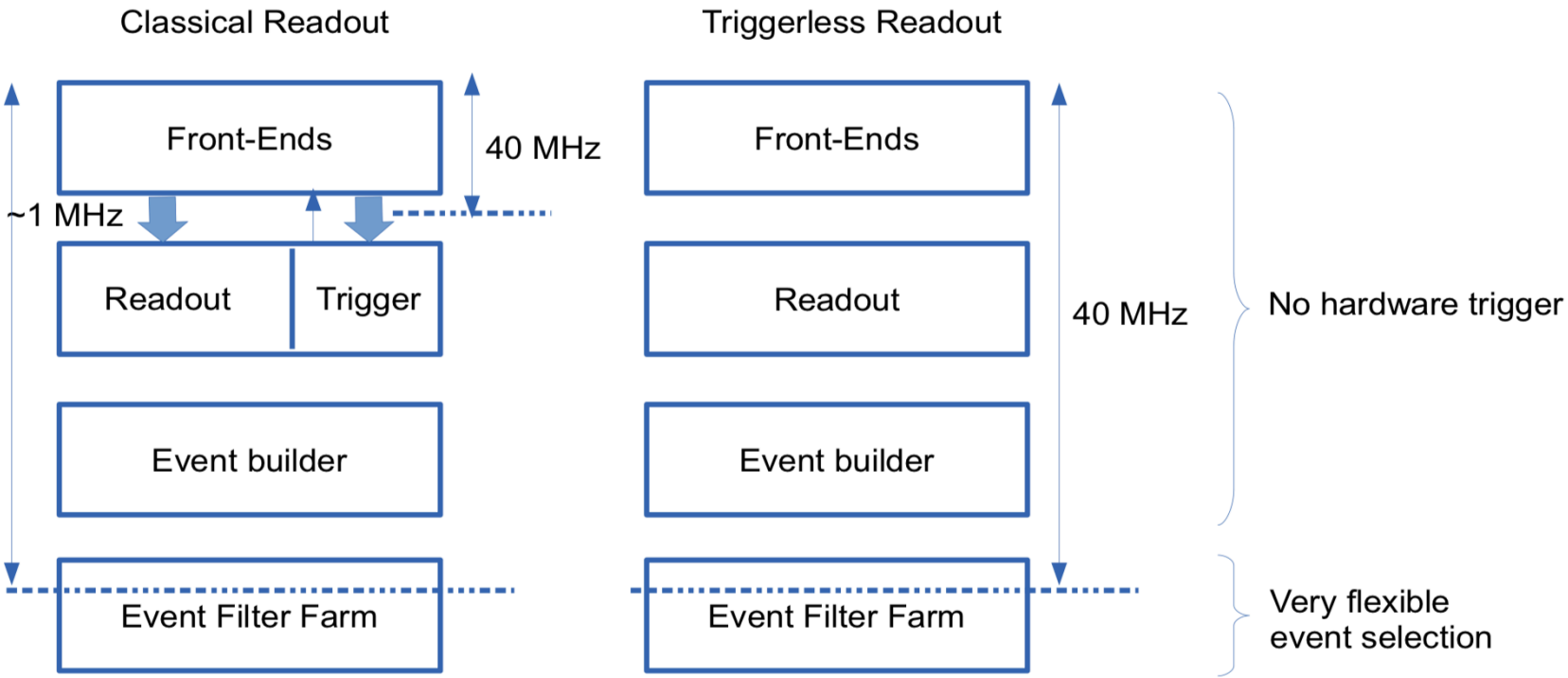
TDAQ Considerations

Sergo Jindariani (Fermilab)

With input from many…

May 18th, 2021

Collider Landscape



Fermilab

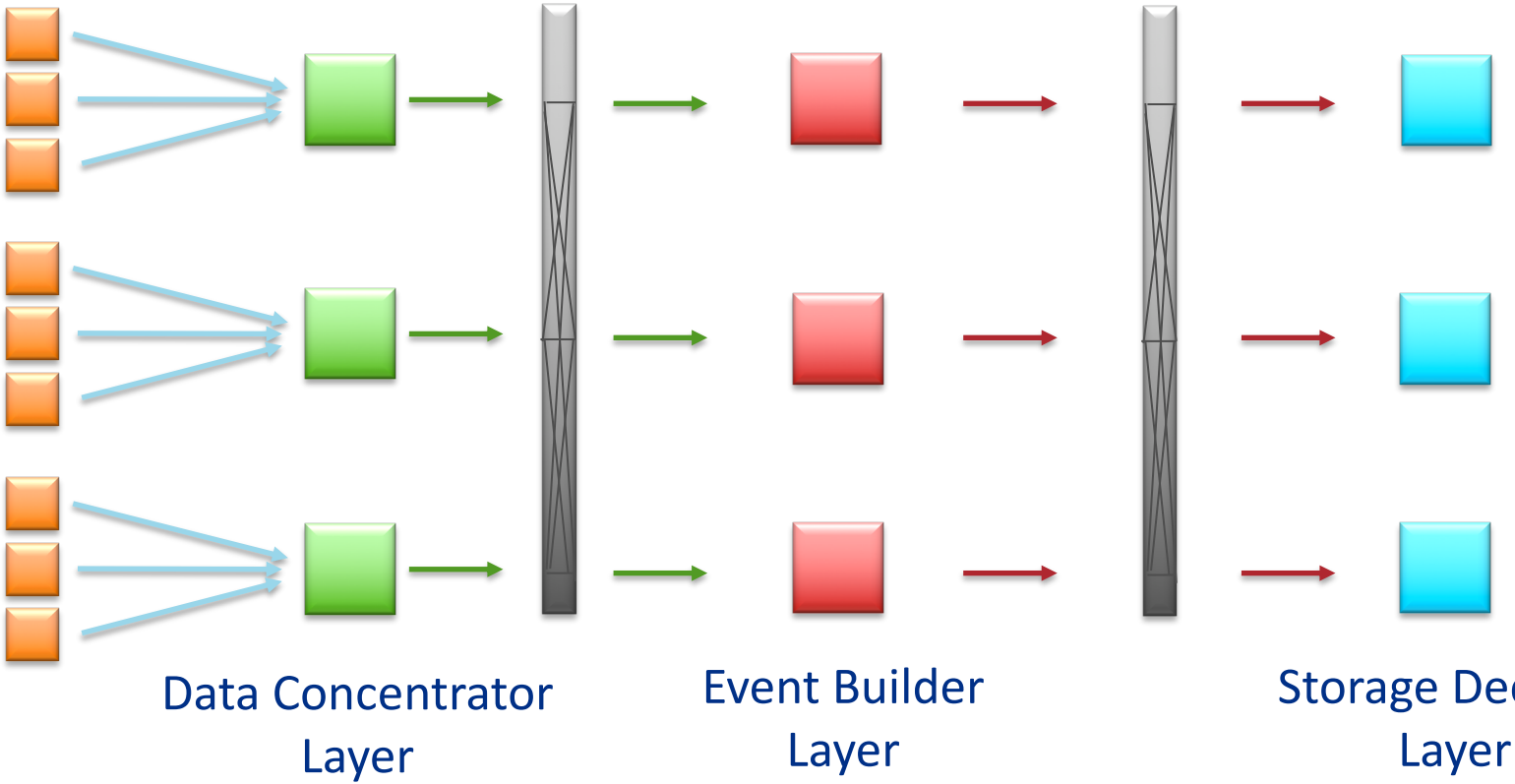
Hardware or Streaming Trigger

- Streaming systems should be explored where possible
  - Remember – discarded data is lost forever
  - Get most data from the detector. More information allows for a better decision
  - No need for long buffers in the detector front-end
  - Software trigger more flexible and easier to maintain
  - Sleeker overall design

- Can you ship the data at BX rate from the detector? Can you process within reasonable time?
  - ATLAS and CMS have to have a hardware trigger at Level-1
  - LHCb opted for a streaming system for Run-3

Generic Readout Data Topology

Illustration by R. Rivera



Fermilab

30,000 feet view

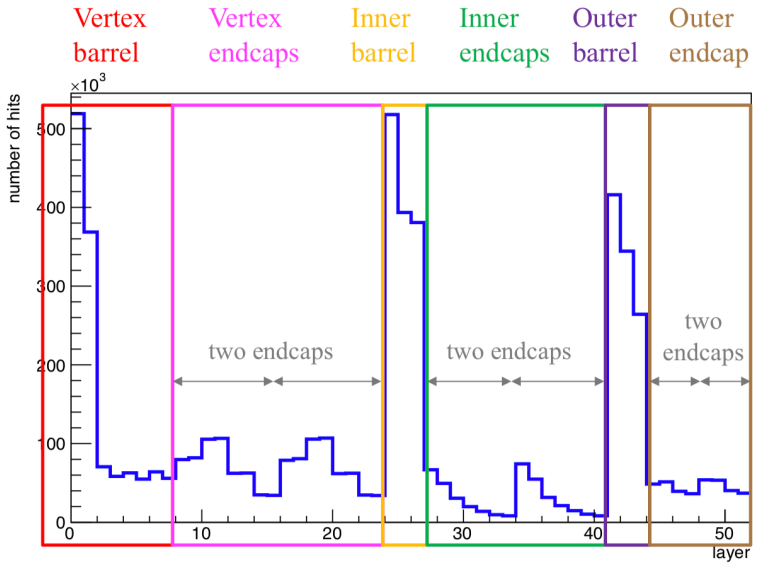
Few Simple Numbers

- HLT Farm Input:
  - LHCb ~ 100kB (event size) * 30 MHz (assuming 75% accelerator duty cycle)
  - CMS now ~ 2MB (event size) * 100 kHz
  - CMS in HL-LHC ~ 10 MB * 750 kHz
- HLT Farm Output:
  - LHCb ~ 100 kB * 20 kHz
  - CMS now ~ 2MB * 1kHz
  - CMS in HL-LHC ~ 7MB * 7.5 kHz
- Let's estimate these numbers for the Muon Collider detector
  - Input event rate 100 kHz
  - What is the event size ?

Event Size - Tracker

Input from Hannsjoerg

- Plot made with 1ns upper bound, but the numbers do not change too much with 2ns:
- Assume 50x50, 50x1000, 50x10000 (macro-) pixel sizes
  - The hit energy threshold is 1.8 keV.
  - From simulation - total number of hits in the entire tracker is 5 Million



Fermilab

Event Size - Tracker

Input from Simone

- Simone has shown that realistic digitization creates more hits than #SimHits
  - The ~ x2-3 in each dimension for 25x25 pixels
  - Since the smallest pixels I assume are 50x50, I will assume a factor of 2
  - 32 bits per hit to encode position, charge, timing, etc

- Event Size: 5M (hits) * 2 (digi factor) * 32 bits ~ 40 MB
- Tracker data rate: 40MB * 100 kHz ~ 30 Tbps

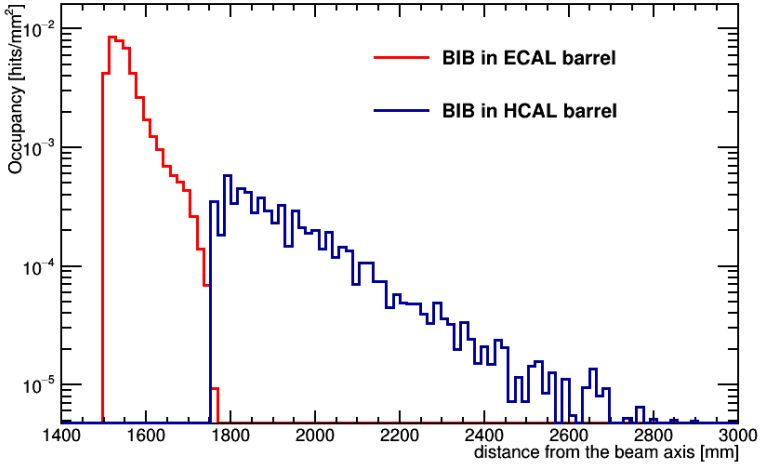
- Note I am not using any data compression or pT-module based filtering
- Time window of 2ns would push these numbers closer to 50 MB/event and ~40 Tbps data rates, ~ 20% increase

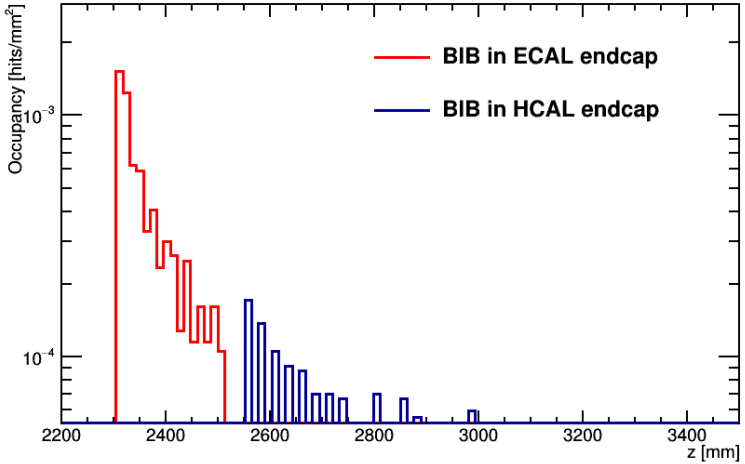
Event Size - Calorimeter

Input from Lorenzo

- Plots made before applying time window of 250 ps
- Cell minimum energy 0.2 MeV

	#channels	Occupancy (hits/mm2)	Hits
ECAL Barrel	64M	$5*10^{-3}$	8M
ECAL Endcap	28M	10^{-4}	70k
HCAL Barrel	6.2M	$5*10^{-4}$	155k
HCAL Endcap	4.4M	$5*10^{-5}$	5k





Event Size - Calorimeter

- Event size dominated by ECAL Barrel
  - 64M channels (for comparison CMS HGCal has 6M)

- Assume on average 20 bits per hit based on HGCal HLT path

- Event size: 8.5M (hits) * 20 bits = 20 MB
- Calorimeter data rate: 20MB * 100 kHz ~ 15 Tbps

- Recall my tracker estimate was 30 Tbps => total 45 -50 Tbps

Data to Storage

- HLT Farm Input:
  - Muon Collider ~ 50 MB * 100 kHz

- What about output?
  - If I assume that HLT output is 75 GB/s (~ CMS in HL-LHC) => 1.5 kHz output rate
  - This would correspond to ~4 PB of storage needed per day of running
  - This assumes that you keep all the raw data for the stored events for future reprocessing
  - Alternative strategy = store more events but with partially reconstructed data

- Is 1.5 kHz enough for interesting physics?
- For comparison:
  - Higgs production rate at 10 TeV and 10^{35} cm^{-2} s^{-1} is <0.1 Hz
  - VBF WW production is ~1Hz

Data to Storage

- HLT Farm Input:
  - LHCb ~ 100kB (event size) * 30 MHz (assuming 75% accelerator duty cycle)
  - CMS now ~ 2MB (event size) * 100 kHz
  - CMS in HL-LHC ~ 10 MB * 750 kHz
  - Muon Collider ~ 50 MB * 100 kHz

- HLT Farm Output:
  - LHCb ~ 100 kB * 20 kHz
  - CMS now ~ 2MB * 1kHz
  - CMS in HL-LHC ~ 7MB * 7.5 kHz
  - Muon Collider ~ 50 MB * 1.5 kHz

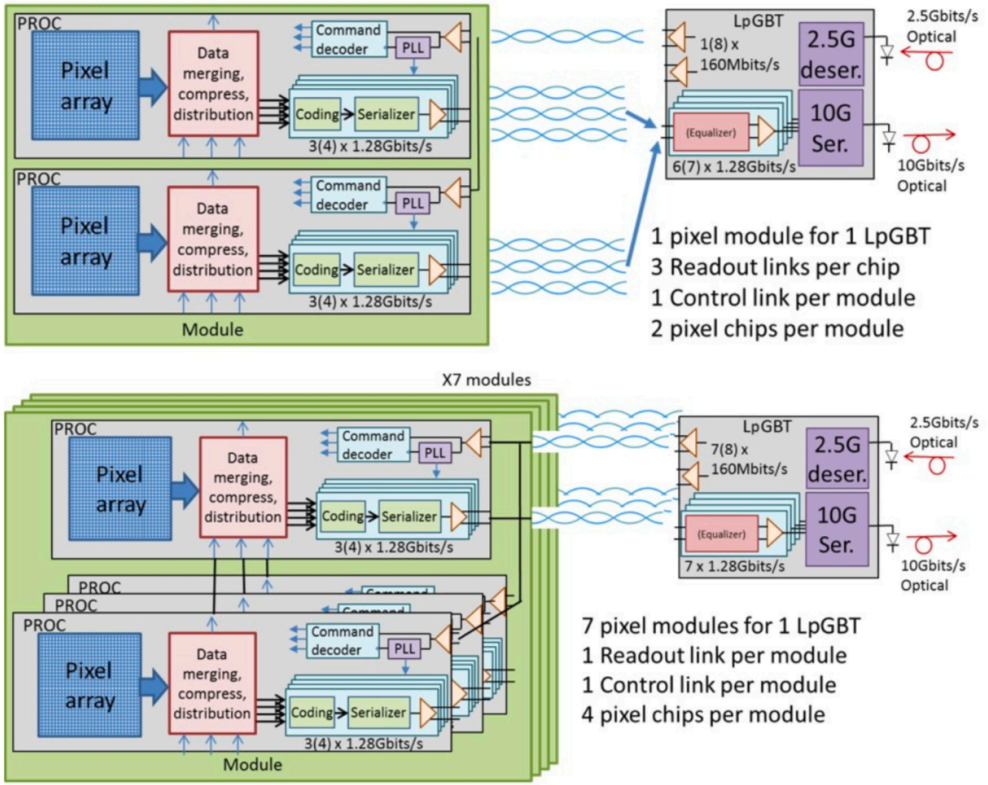
Zoom in a little

CMS Phase-2 Pixel Module

Input from CMS Tracker TDR

- Pixels readout by an array (2-4) of ROCs
- Gigabit transceivers not rad hard – run copper differential pairs from the ROC to the optical module
- RD53 chip has 1-4 E-links running at 1.28 Gbps => up to 5 Gbps
- HL-LHC lpGBT sends data with speeds up to 10 Gbps to the Back End

- In CMS Phase-2 pixels 1(inner)-7(outer) pixel modules are served by a single GBT
- Reasonable to assume that both E-link and GBT bandwidth can be pushed by x2 in the next 20 years ?

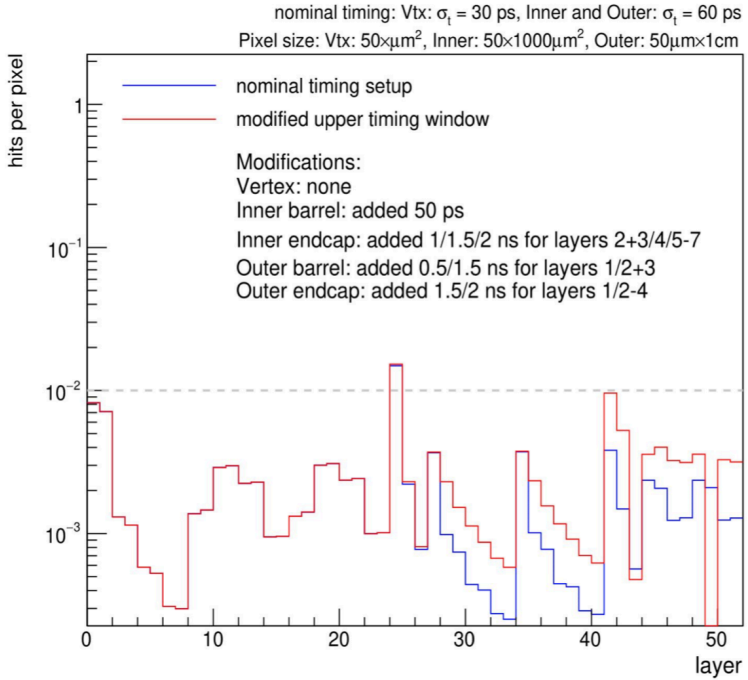


Fermilab

Tracker Modules

Input from Hannsjoerg, Massimo, Nazar

- Assume readout chips with 2.5x2.5 cm^2 area
- Vertex detector modules:
  - ~10 cm2 - 2 chips
  - Per chip: 100kHz * 250k (pixels per chip) *0.01 (occupancy) * 32bits ~ 7.5 Gbps
  - Per module 15 Gbps
- Inner Tracker modules:
  - ~20 cm2 - 4 chips
  - Per chip: 100kHz * 12500 (macro pixel per chip) *0.01 (occupancy) * 32bits ~ 0.4 Gbps
  - Per module 2 Gbps
- Outer Tracker modules:
  - ~ 40 cm2 - 8 chips
  - Per chip: 100kHz * 1250 (macro pixel per chip) *0.01 (occupancy) * 32bits < 0.1 Gbps
  - Per module 1Gbps



Fermilab

Tracker Modules

Input from Hannsjoerg, Massimo, Nazar

- Number of modules calculated by using “area” and adjusting for module size
- 1 module per GBT in the Vertex detector
- 10 (20) modules can be ganged into a single GBT for the IT (OT)

	Modules	Links
VXDB	400	400
ITB	4500	450 (10 modules/link)
OTB	16000	800 (20 moduls/link)
Total Barrel	20900	1650

- I did not have numbers for the endcap but assuming same as barrel we endcap with 3300 links

Calorimeter Modules

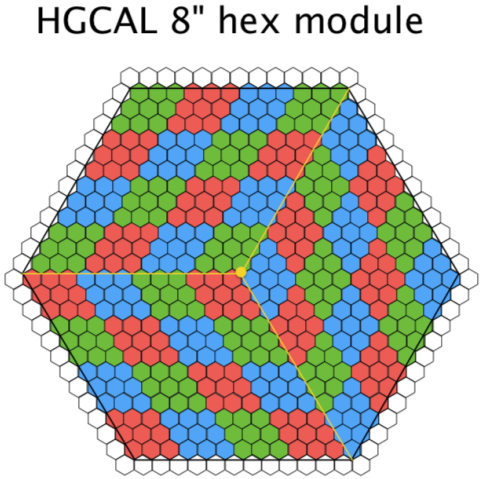
Input from Lorenzo and HGCal TDR

- HGCal: cell size 0.5-1 cm^2 with 200 channels per module

- Muon Collider: cell size 0.25 cm^2, 800 channels per module (assuming same physical module size)

  - Per module: 100kHz * 800 (channels) * 0.005 (occupancy) * 20 bits ~ 0.5 Gbps

  - Gang 20 modules per GBT
  - 127k modules => 6400 links running at 10 Gbps

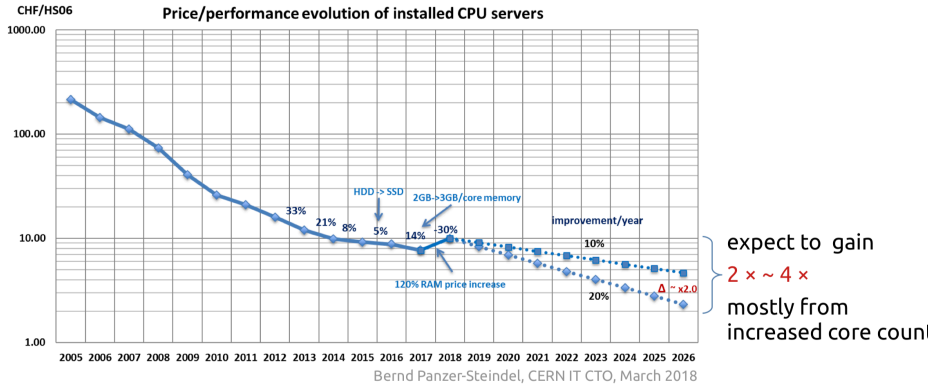


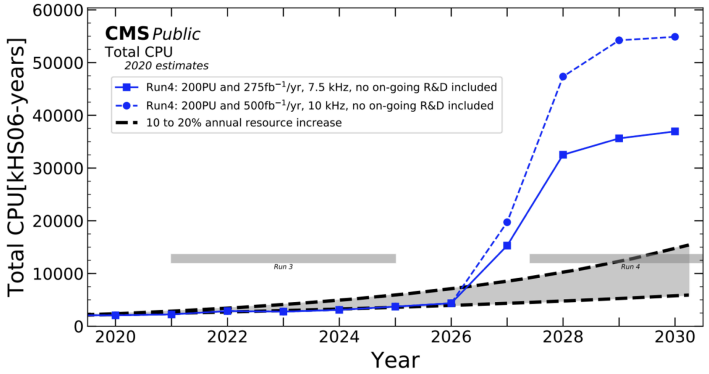
Fermilab

Processing Times

Input from LHCb and CMS HLT TDRs

- LHCb farm uses CPU+GPU cores, CMS HLT farm 30,000 CPU cores
- Maximum processing time per event 12ms/event in LHCb, CMS ~300ms
  - For HL-LHC projected 1s/event with tails up to ~minute

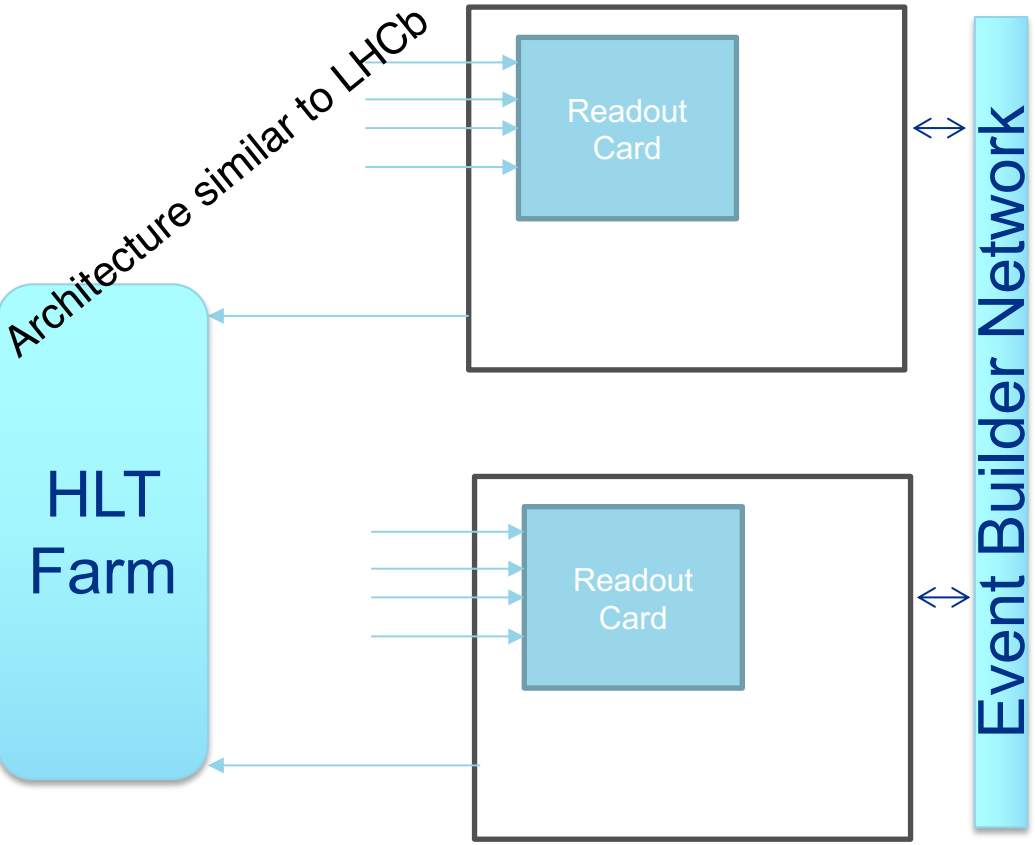




- x4 increase in the number of cores by 2026… Let's assume another x2 by 2040 => 200,000 cores
- This puts us in average projected latency of few seconds per event
- Current reconstruction takes days

Fermilab

Putting All Together



- 10,000 links at 10-20 Gbps
- PCIe Gen 5 allows 400Gbps
- 500 boards (20 links per board)
- 100 Gbps LAN for Event Building
- 50 Tbps aggregate bandwidth

Summary and Outlook

- Presented first attempt to outline a potential trigger strategy
  - This should not be taken as a real design
- Estimated data rates at event and module level
- From this preliminary look, it appears that from the bandwidth perspective a trigger-less readout is a possible
  - R&D is needed to advance technology by a factor of x2-3 in various places
  - Additional handles can be explored
- Reconstruction time appears to be a major bottleneck:
  - We need to understand how we can speed it up
  - Tiered reconstruction?
  - Accelerators (GPU, FPGA) can help