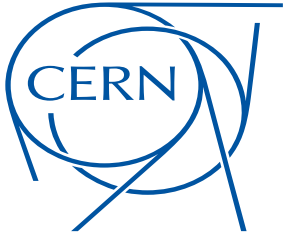
CMS at HL-LHC

Online Challenges

Emilio Meschi – CERN EP/CMD
The CMSDAQ group



SUMMARY

- CMS Phase-II Upgrades: Trigger/DAQ
- Challenges of the Baseline Design
- New directions and Challenges

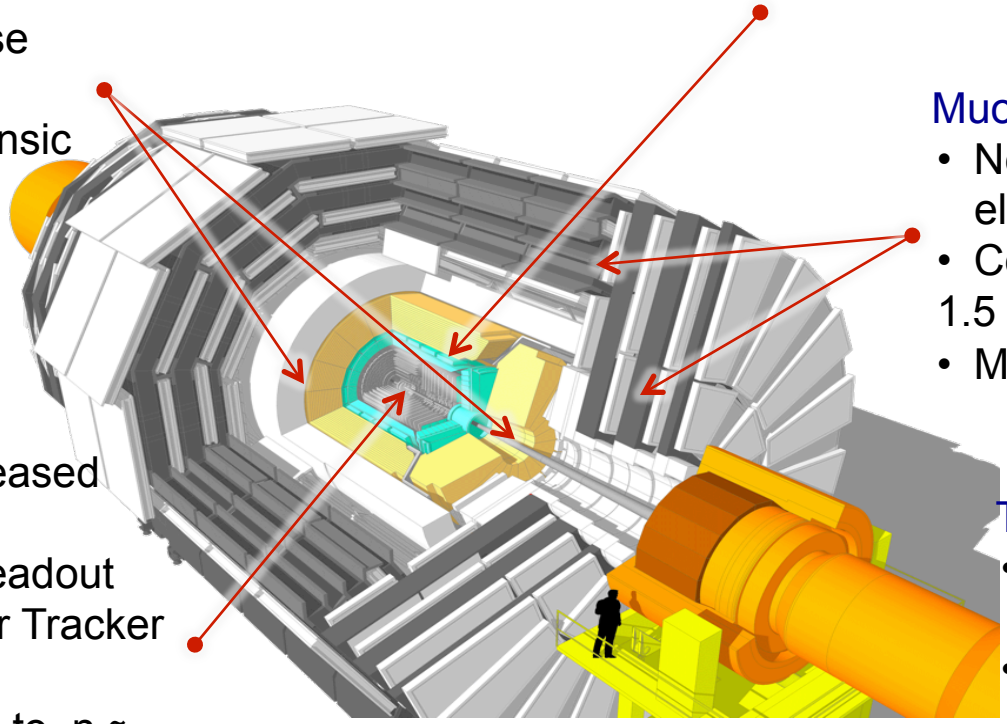
CMS Upgrades for Phase-II

New Endcap Calorimeters

- Rad. tolerant - increased transverse and longitudinal segmentation - intrinsic precise timing capability

Barrel EM&HAD calorimeter

- New FE/BE electronics



Muon systems

- New DT & CSC FE/BE electronics
- Complete RPC coverage $1.5 < \eta < 2.4$
- Muon tagging $2.4 < \eta < 3$

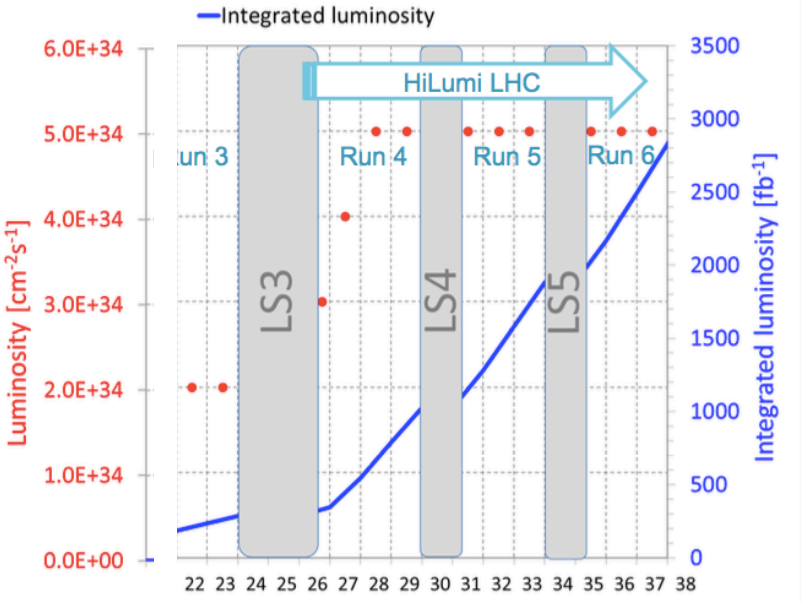
New Tracker

- Rad. tolerant - increased granularity - lighter
- 40 MHz selective readout ($Pt \geq 2$ GeV) in Outer Tracker for Level-1Trigger
- Extended coverage to $\eta \approx 3.8$

Trigger/HLT/DAQ

- Track information in Trigger (hardware)
- Trigger latency 12.5 μs L1 output rate 750 kHz
- HLT output 7.5 kHz

Machine and Experiment



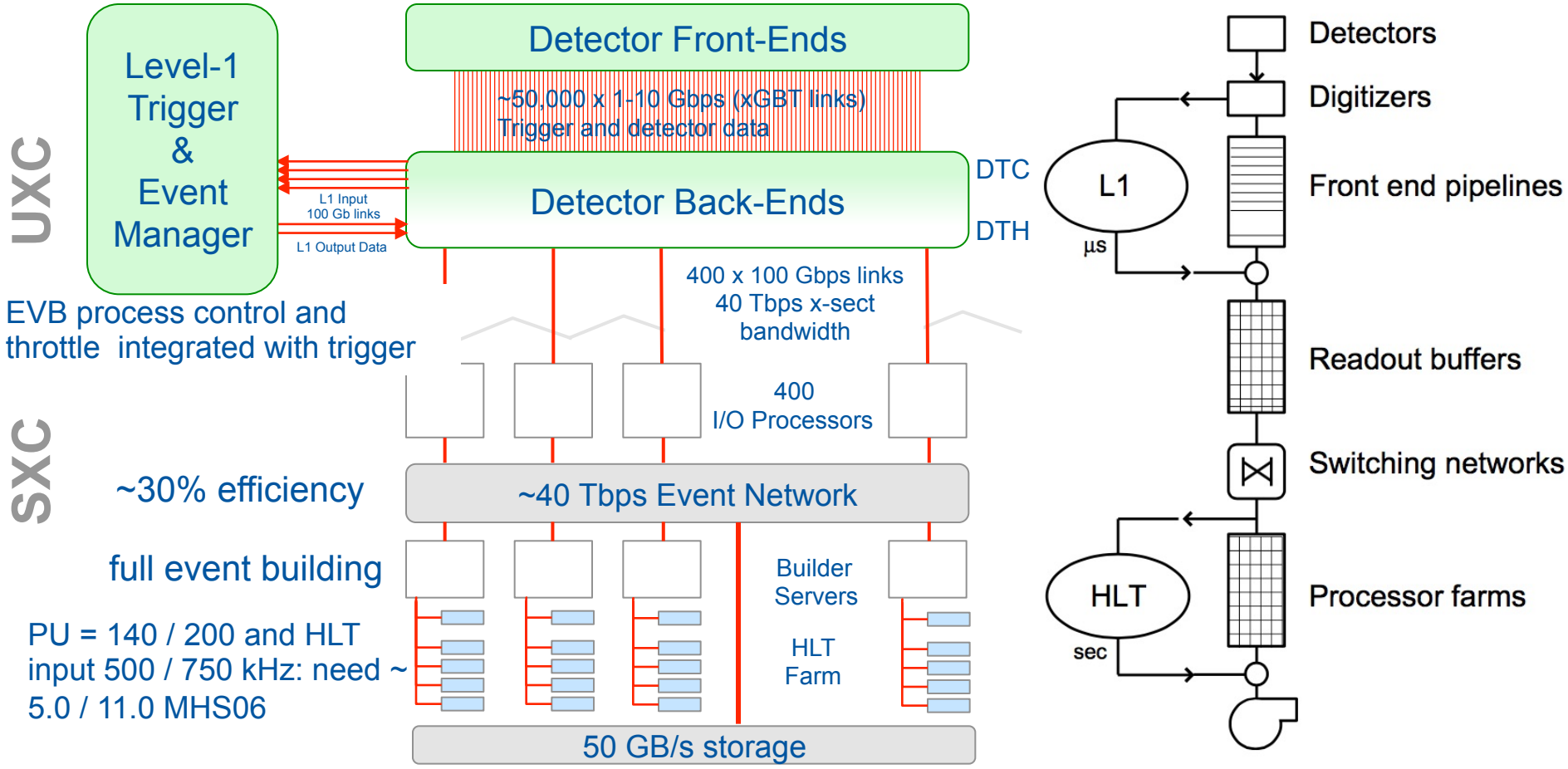
- About 50k point-to-point bidirectional optical links (lpGBT) on-to-off-detector with varying fractions devoted to trigger data
- 750 kHz max L1 accept rate and 7.5 kHz max HLT accept rate
- Typical event size 5MB

At least barrel calorimetry and muon systems read out in streaming (untriggered) mode

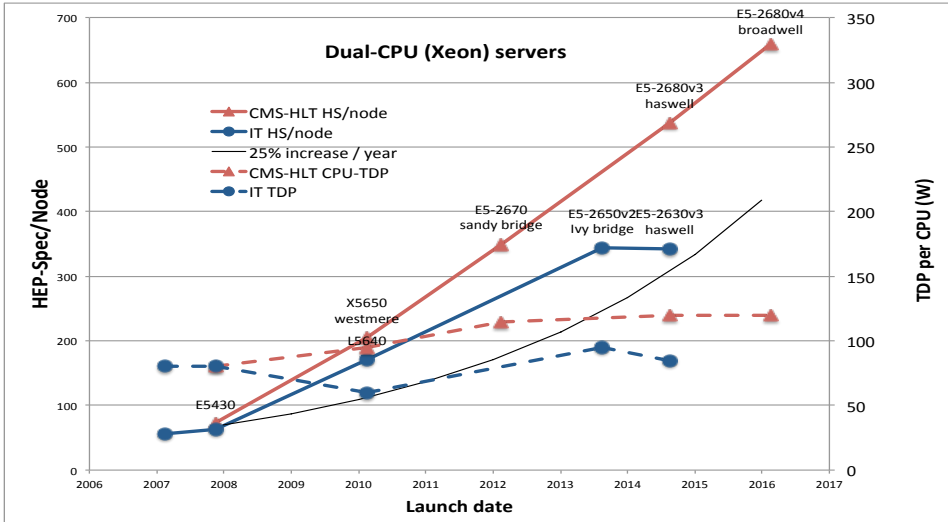
- 5 (7.5) 10^{34} $cm^{-2}s^{-1}$ level (peak) lumi
- 140 (200) baseline (ultimate) PU with ~1.2 PU/mm
- 250 (>300) fb^{-1}/y and 3 (4) ab^{-1} of total integrated luminosity

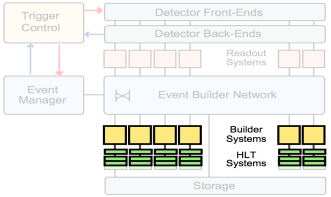
	HL-LHC Phase-II upgr.	
Energy	13 TeV	
Peak Pile Up (Av./crossing)	140	200
Level-1 accept rate (maximum)	500 kHz	750 kHz
Event size (design value)	4.5 MB	5.0 MB
HLT accept rate	5 kHz	7.5 kHz
HLT computing power	5.0 MHS06	11 MHS06
Storage throughput (design value)	27 GB/s	42 GB/s

L1 / DAQ / HLT: Baseline



High Level Trigger: CPU





Guestimate based on Current detector, PU dependence, assuming current algorithms scale

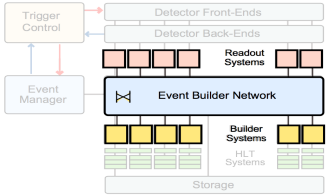
- Possible gain by using L1 Track trigger
- PU = 140 / 200, L1 Rate 500 / 750 kHz:
  ~ 5.0 / 11.0 MHS06
- Compare LHCb: need 3.3 MHS06 in 2021

Assumed performance Increase of server	Exponential 25% / year [WLCG 2014]	Exponential 12.5 % / year	Linear 73 HS/year
11 years	12	3.7	2.2
#servers in Q1-27	1431	4562	7860
Total power Q1-27	0.5 MW	1.6 MW	3 MW

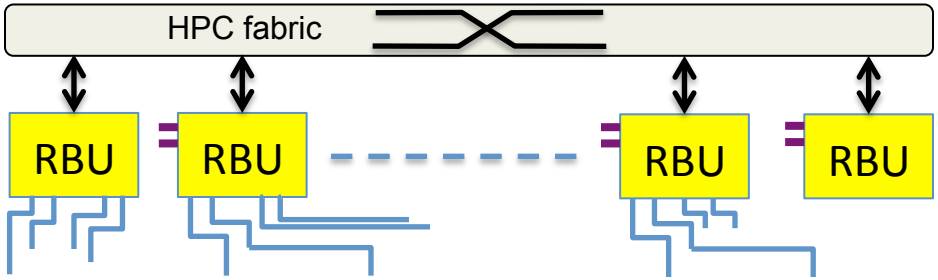
Baseline: Challenges

- Baseline DAQ architecture for Run4 feasible with readily available technology
- Main Challenge: Limited Budget
- Directions:
  - Data to Surface: efficient concentration and transition to asynchronous/reliable protocol
  - Event Builder: Reduce size and complexity
    - I/O processors
    - Choice of Network
  - Size/cost of HLT farm
    - Understand actual evolution of hardware
    - Use of heterogeneous architectures
    - Evolution of reconstruction software

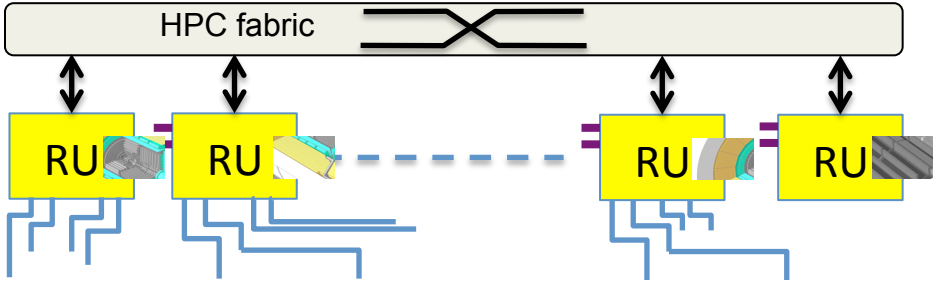
From Event Builder to Event Network



Exploit bi-directional links at > 50%

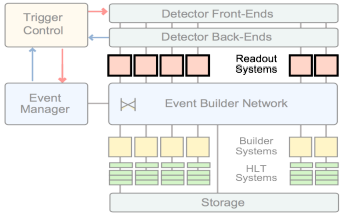


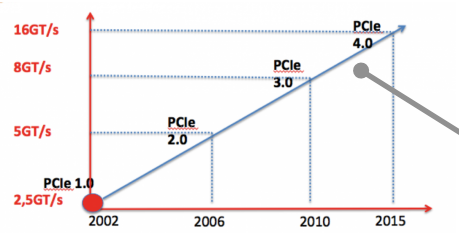
One-to-one correspondence of geographical and network address: events accessed directly by the HLT process through the HPC fabric (on demand if sw supports it)



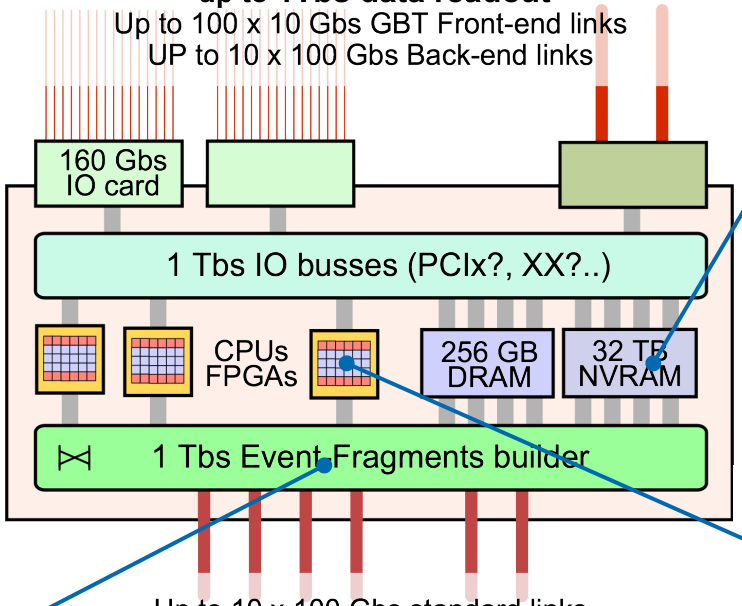
- HPC fabrics: IB, omnipath, ..
  - on chip ?
- 100 Gbps (4x25 Gbps lanes) to 200 Gbps (4x50 Gbps lanes) or more
- Optimizing use of network
  - Efficient use of HPC fabric
  - Optimization:
  - folded EVB -> non-building

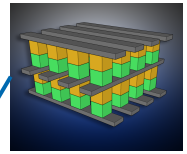
I/O, Buffer and Process (extrapolation)





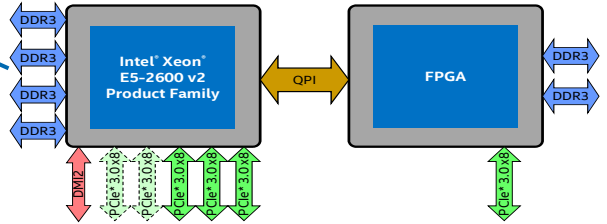
up to 1Tbs data readout
Up to 100 x 10 Gbs GBT Front-end links
UP to 10 x 100 Gbs Back-end links



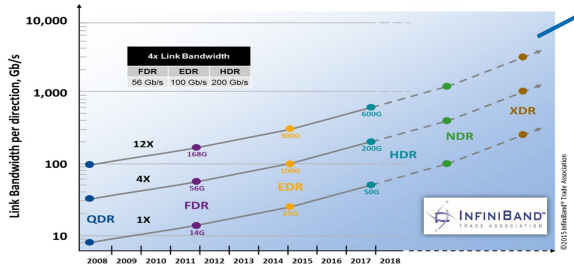


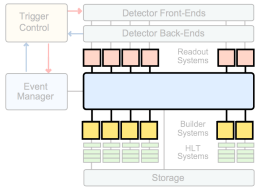
Readout Server
PC, ATCA crates etc..

CPU power for:
- Link protocols
- Detector hit builder
- Event fragment builder
- Large latency buffer
- Detetcor data monitor

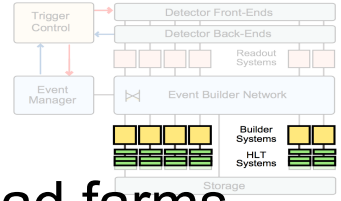


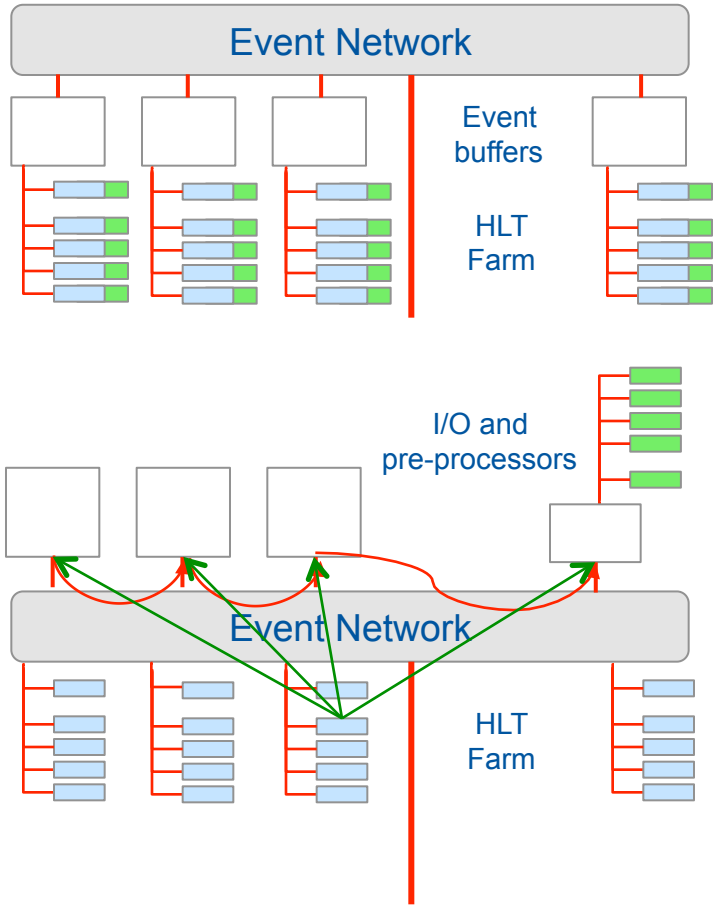
Up to 10 x 100 Gbs standard links
up to 1 Tbs Event-Fragments output





HLT: harness efficient CPU power





- Coprocessors and offload farms
  - CUDA, OpenCL…
  - Preemptive local reconstruction with ad-hoc software/hardware
- Truly distributed local processing (exploiting high-performance fabric)
- FPGA-assisted regional reconstruction
- Early classification – real-time indices
- Container / query programming style leveraging large NV memory

Example

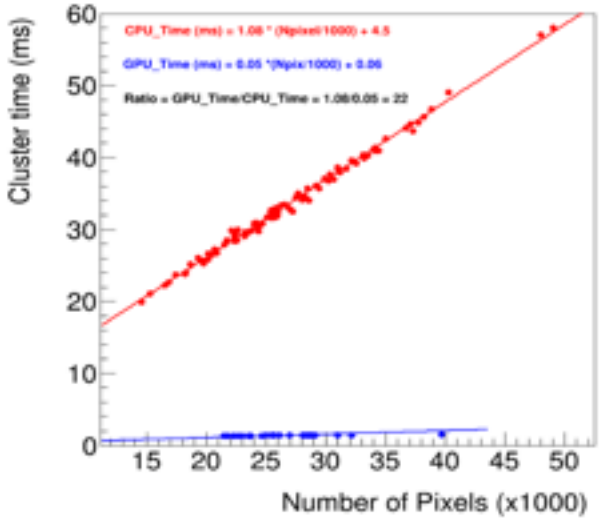
Timing performance ttbar 50 pileup

Events with PU50 are not getting even close to saturate the GPU

- Only 2-5% of the GPU busy
- ~100MB GPU DRAM used per event
- This allows us to offload many events on the same GPU by many threads

	time per event CPU (ms)	time per event GPU (ms)
Triplet propagation	66.3	N/A
CA	22	1.6 (15.2)

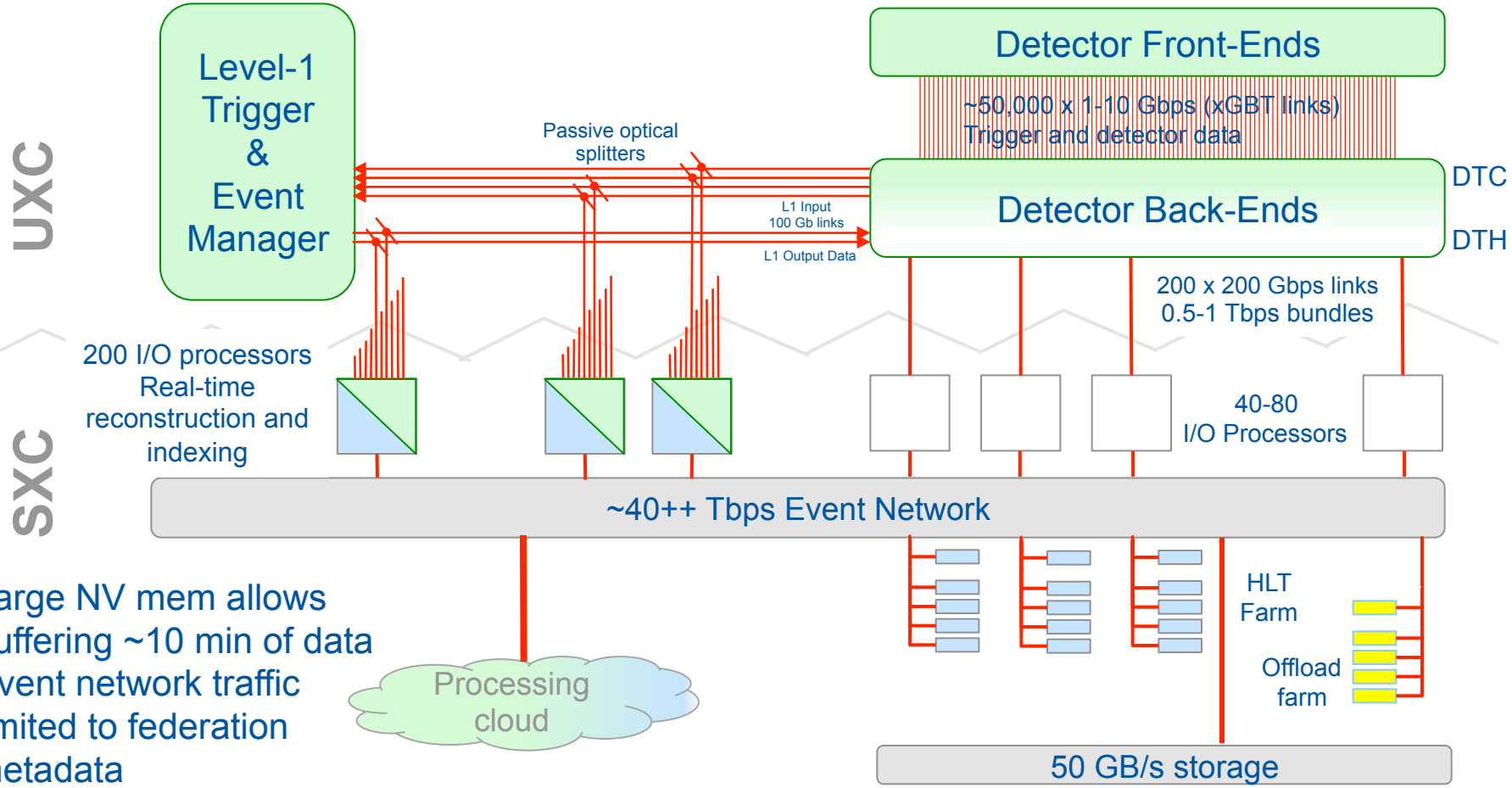
- Hardware used:
  - CPU Intel 4771K
  - GPU NVIDIA K40



Cursory conclusion

- Reduction of complexity and cost may result from tracking and late adoption of maturing technologies
  - Multi - 100 Gb links, on chip fabric
  - High bandwidth and large NV memory I/O servers
- Design options and exploratory work
  - Event Network: Non-building
  - HLT: programming styles more suitable for truly distributed systems with large NV memory
  - HLT: coprocessors and offload engines

Phase II DAQ and “Scouting” Option



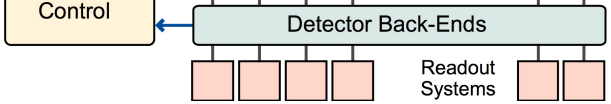
Large NV mem allows buffering ~10 min of data
Event network traffic limited to federation metadata

Final Remarks

- Baseline DAQ architecture for Run4 feasible with readily available technology
  - Main Challenge: reduction of complexity and cost
- Paradigm shift required to harness heterogeneous computing architecture
  - H”B”C -> HPC: not “embarrassingly” parallel
- “Scouting” at 40 MHz may offer interesting physics opportunity
  - And poses new computing challenges

Backup

Overview



ATCA

Datacenter hw + network

FE

BE

Trigger

Sync

DAQ

Async

I/O Buf PP

HLT

Event Network

Store

Global Trigger

TTC/TTS

CLK L1A

BSY

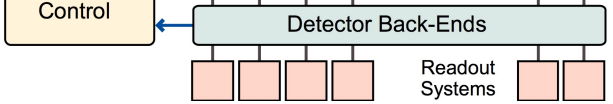
UCX Data to surface (D2S)

SCX

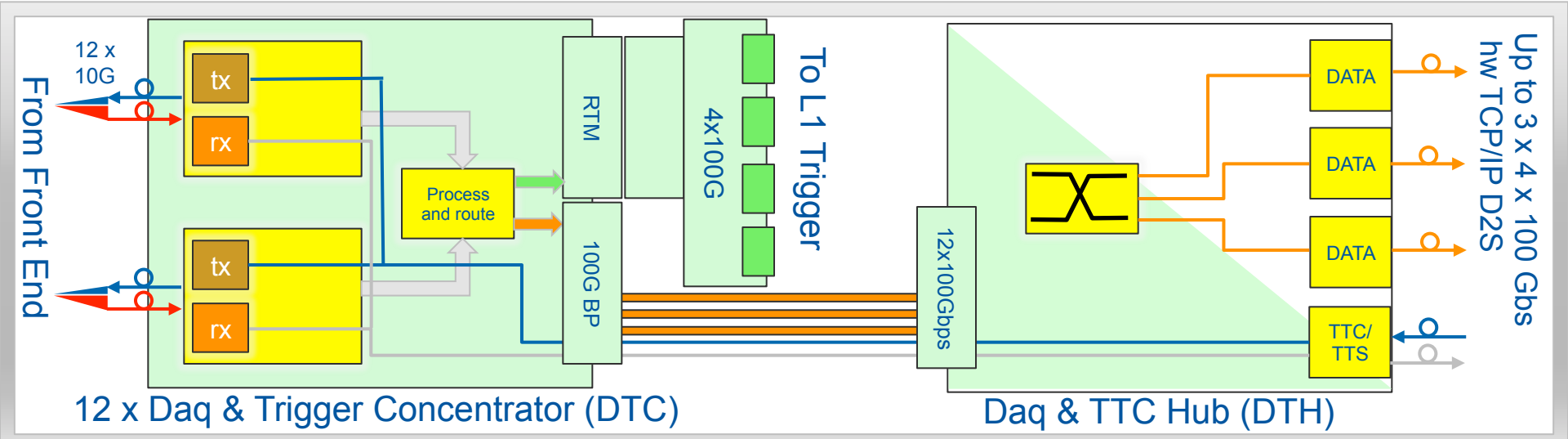
Synchronous (40 MHz clock driven)

Asynchronous (event driven)

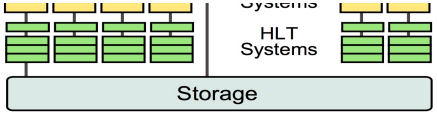
DAQ&TTC Hub (DTH)



- Data aggregation from leaf cards in crate
- Timing + control
- Convert to commercial network+protocol
  - Most likely Ethernet with 25 Gbps or 50Gbps lanes and TCP/IP (need fast buffer memory)
- Emulation / data generation for testing purposes
- Monitoring of data and TTC/TTS functions
- Software stack for local processing

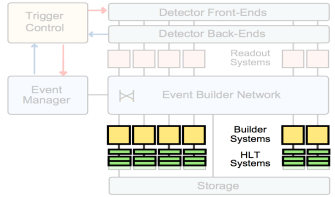


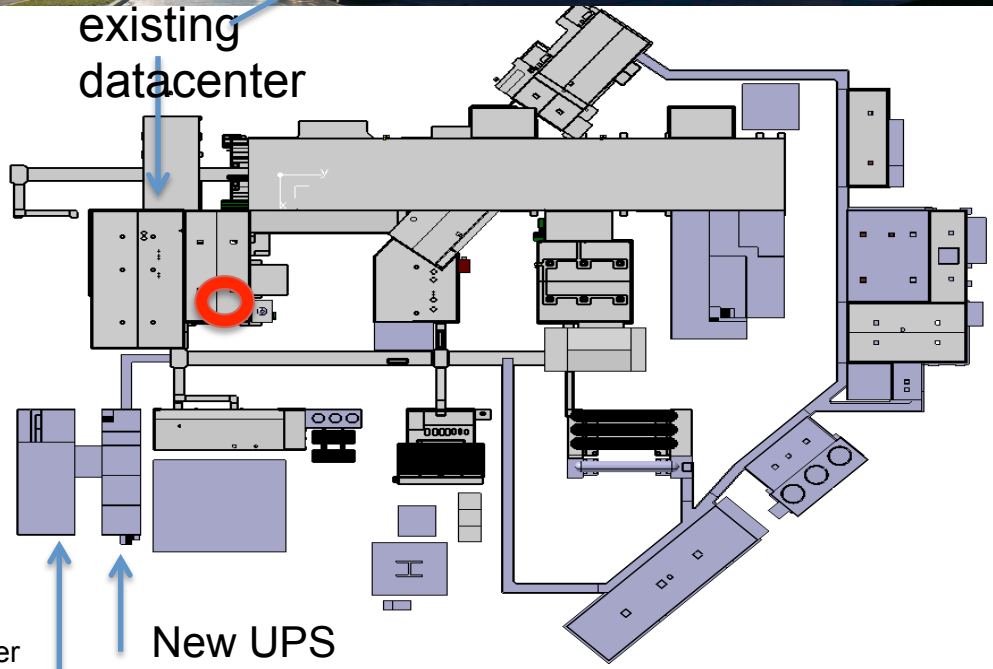
HLT: local or remote ?



- Observation
  - Need new infrastructure on site for data centre >1 MW
  - 25 Gbps lanes over MM fibre go only up to 100 m, next hop is 10 km
- Suppose CERN provided a high bandwidth link from Pt5 to a central data-center (eg Prevessin)
  - 5 MB@750kHz = 4 TB/s, need 50 Tbps
- EVB at Pt5
  - Built full events, for convenience built lumi-sections (~10 s)
  - Store EVB output in filesystem (as today done on ram-disk)
  - Plenty of Space and BW with 3D-xpoint or SSDs
- Transfer files to remote datacenter to run HLT
  - “EOS++” on SDD for buffering
  - Condor batch for running jobs
- Clean separation online / offline
  - Flexibility for waiting for calibration, multi-pass, etc ..
- Note:For run 2,3: need 200 GB/s so 40x40 Gbps links (have 4x40 Gbps now)

HLT: local vs. remote





- Existing datacenter: racks, cooling to be redone if >1 MW
- New datacenter @P5: building+infrastructure
- Remote datacenter: need 5000 GB/s so 50x100 Gbps links over > 10 Km (have 4x40 Gbps now)

Event Scouting in CMS (HLT scouting)

- Generic signatures with acceptable rates at L1, but no easy additional features for L1 or HLT…
  - … which would require too high b/w to disk or too much CPU to select at HLT…(e.g. PFjets)
- Examples
  - Multiple, relatively low-pt jets (e.g. dijet bumps)
  - Low-pt muons (e.g. low-mass resonances with suppressed production)
- Technique:
  - Do analysis using HLT objects (**reduced accuracy**!!! E.g. “CaloJets”), store minimal information (as opposed to parking which stores full events but postpones analysis)
  - Look for interesting features (e.g. jj mass bump) – if found, adapt HLT to select…and discover

Higgs Rare Decays

Main limitation: small number of events after selection (~3 at HL-LHC)

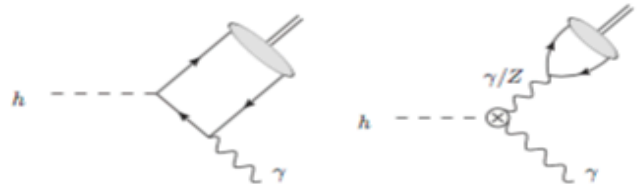
Exclusive modes $BR(H \to V\gamma) \sim 10^{-6}$ (V = vector meson)
allow the extraction of Yukawa couplings to first 2 quark generations

[Bodwin et al. PRD 88 (2013) 053003; Kagan et al. PRL 114 (2015) 101802]

$H \to J/\psi\,\gamma \longrightarrow y_c$

$H \to \phi\gamma \longrightarrow y_s$

$H \to \rho\gamma \longrightarrow y_{u,d}$



Limits on $J/\psi\,\gamma$:

Current (LHC 8 TeV 20 fb^{-1}): $< 1.5 \times 10^{-3}$

LHC (14 TeV 300 fb^{-1}): $\lesssim 150 \times 10^{-6}$

HL-LHC (14 TeV 3 ab^{-1}): $\lesssim 45 \times 10^{-6}$

Main limitation: small number of events after selection (~3 at HL-LHC)

Can we recover signal efficiency?

Limits on triggering - photon spectrum, photon eta distribution

Can we use electron decays of quarkonia? Can we push HL-LHC down to the SM?

http://arxiv.org/abs/1501.03276

$$B_{SM}(Z \to J/\psi + \gamma) = (9.96 \pm 1.86) \times 10^{-8},$$
$$B_{SM}(Z \to \Upsilon(1S) + \gamma) = (4.93 \pm 0.51) \times 10^{-8},$$
$$B_{SM}(Z \to \phi + \gamma) = (1.17 \pm 0.08) \times 10^{-8}.$$

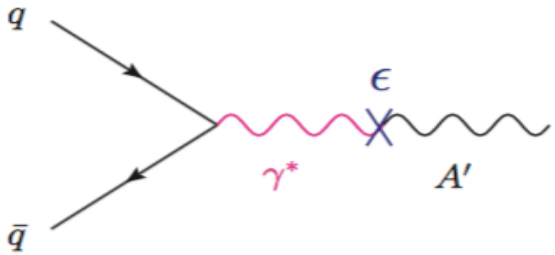
$$\mathcal{B}_{\rm SM}(H \to J/\psi + \gamma) = 2.79^{+0.16}_{-0.15} \times 10^{-6},$$
$$\mathcal{B}_{\rm SM}[H \to \Upsilon(1S) + \gamma] = 6.11^{+17.41}_{-6.11} \times 10^{-10},$$
$$\mathcal{B}_{\rm SM}[H \to \Upsilon(2S) + \gamma] = 2.02^{+1.86}_{-1.28} \times 10^{-9},$$
$$\mathcal{B}_{\rm SM}[H \to \Upsilon(3S) + \gamma] = 2.44^{+1.75}_{-1.30} \times 10^{-9}.$$

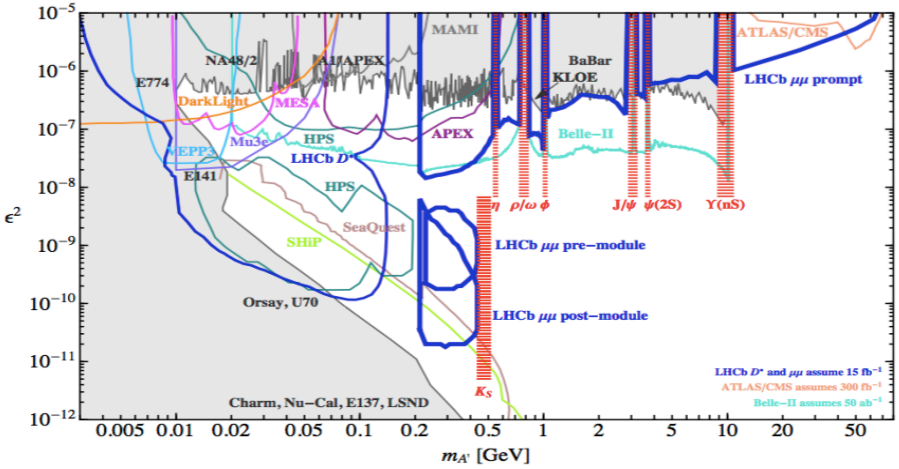
Similar for Z boson http://arxiv.org/abs/1407.6695

Dark Sector

Photon Portal:

$$\mathcal{L}_{\text{mix}} = \frac{\epsilon}{2} F^{\text{QED}}_{\mu\nu} F^{\mu\nu}_{\text{Dark}}$$





1. two opposite-sign muons with η(μ±) ∈ [2,5], p(μ±) > 10 GeV, and pT (μ±) > 0.5 GeV
2. a reconstructed A′ → μ+μ− candidate with η(A′) ∈ [2, 5], pT (A′) > 1 GeV, and passing the isolation criterion for mA′ > mφ
3. an A′ → μ+μ− decay topology consistent with either a prompt or displaced A′ decay

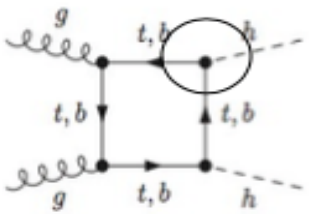
Look for e+e- decays?
Extend current searches to lower masses…

http://arxiv.org/pdf/1603.08926v1.pdf

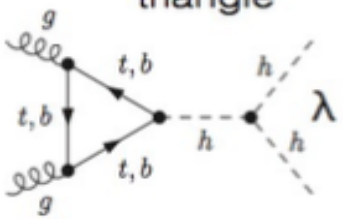
Higgs Pair Production

https://cds.cern.ch/record/2063038/files/FTR-15-002-pas.pdf

box



triangle



$\sqrt{s}$ [TeV]	σ^{NLO} [fb]
8	8.2
14	33.9

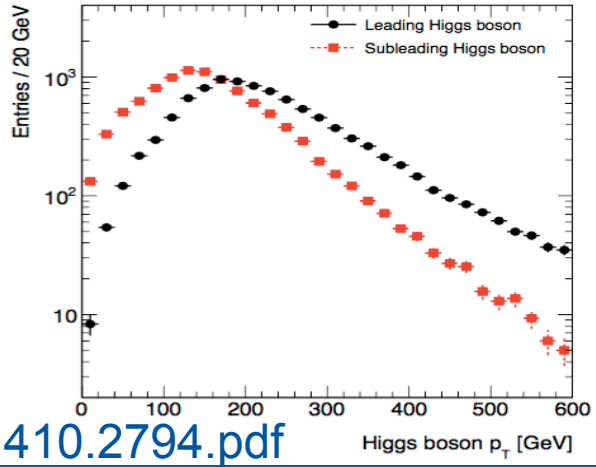
less interesting:
tth, bbh coupling

Interesting:
self-coupling

	# Higgs pairs to bbγγ
LHC: 14TeV 300fb^{-1}	36
HL-LHC: 14TeV 3ab^{-1}	360

bbγγ, bbττ and bbWW considered
How about bbbb channel ? Can we do something with it ?

$$HH \rightarrow b\bar{b}b\bar{b}$$



$\sigma \times BR$ [pb] 1.16×10^{-2}

"Standard" L1-oriented cuts will reduce this to few hundred events
Jets at threshold...
Formidable bkg, very hard analysis
B tagging? Jet Substructure ?
Profit from new techniques…

http://arxiv.org/pdf/1410.2794.pdf