Particle Detectors – Trigger/DAQ

Frank Winklmeier
University of Oregon

CERN Academic Training Lecture Programme
https://indico.cern.ch/event/526768

12th May 2016

Introduction

- Trigger/DAQ is a huge area
  - Specific experiments have very specific Trigger/DAQ needs
  - Concentrating on Trigger/DAQ in the 4 LHC experiments and their upgrades
    - But cannot describe all 4 DAQ systems in detail
  - Should give a good overview of current and future technologies

- For a more pedagogical introduction to Trigger/DAQ
  - W. Vandelli, CERN Summer student lectures, “Electronics, DAQ & Trigger” [1] [2] [3]
  - ISOTDAQ 2016 - International School of Trigger & Data Acquisition

- For a complete overview of Trigger/DAQ at the LHC
  - DAQ@LHC 2016 - Workshop on Trigger/DAQ at the LHC experiments

- Much of the material inspired by similar talks from
  - Su Dong, Frans Meijers, Andrea Negri, Niko Neufeld, Francesca Pastore, Brian Petersen, Gerhard Raven, Brigitte Vachon, Wainer Vandelli, …

Outline

- Introduction to Trigger and DAQ
- Trigger Algorithms
- Trigger/DAQ upgrades for Run-2
- Future upgrades and technology evolution

- Disclaimer
  - I have been working on ATLAS Trigger/DAQ the last 10 years
  - Any bias towards ATLAS and mistakes in other areas is due to that

What is a DAQ System?

Google

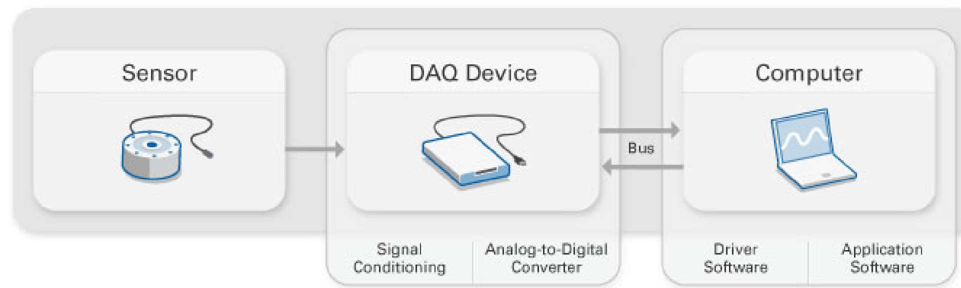
what is a daq system

All Images Shopping Videos News More Search tools

About 620.000 results (0,26 seconds)

Data acquisition (**DAQ**) is the process of measuring an electrical or physical phenomenon such as voltage, current, temperature, pressure, or sound with a computer. A **DAQ system** consists of sensors, **DAQ** measurement hardware, and a computer with programmable software.

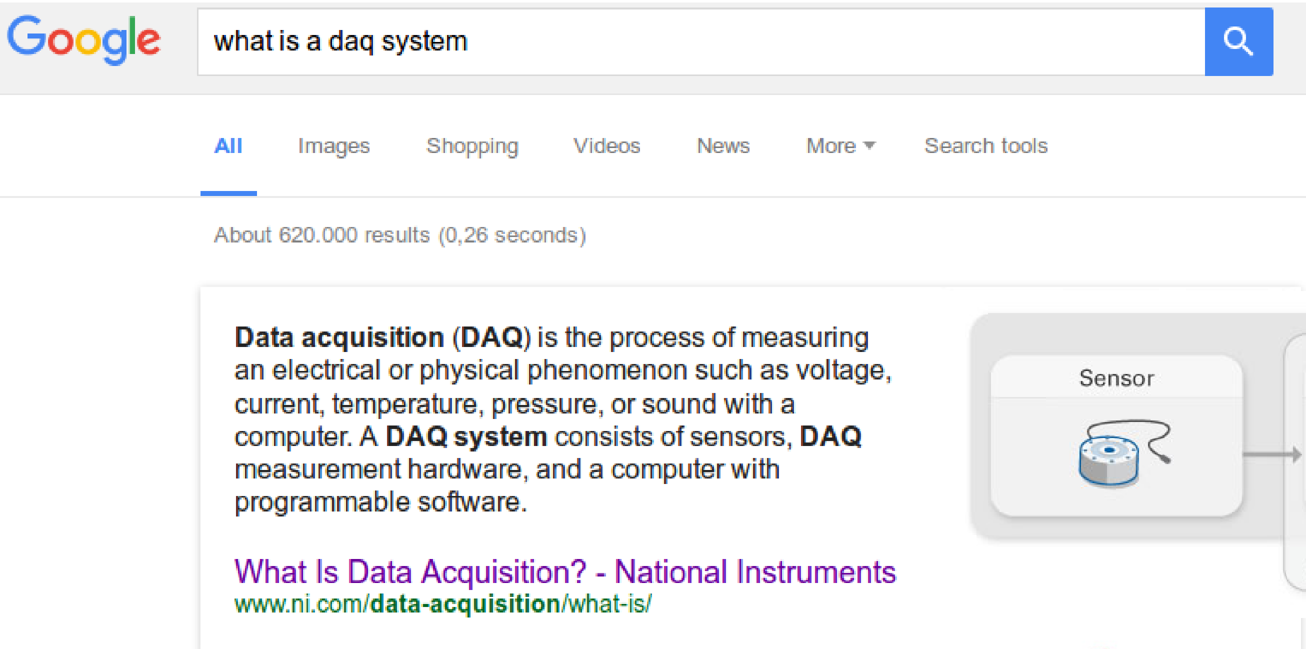
What Is Data Acquisition? - National Instruments
www.ni.com/**data-acquisition**/what-is/

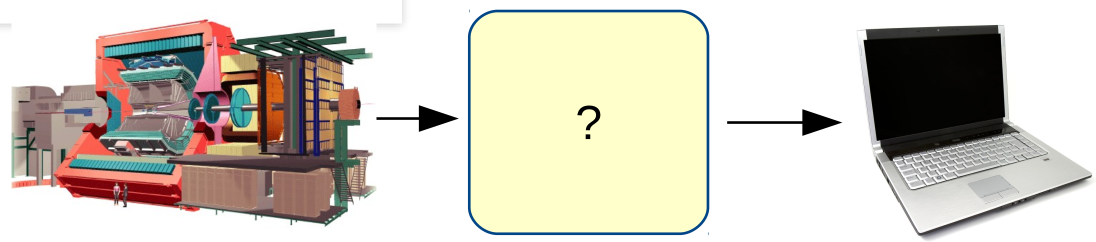


About this result • Feedback

- Main keywords
  - measure electrical phenomenon
  - computer
  - programmable software

What is a DAQ System?

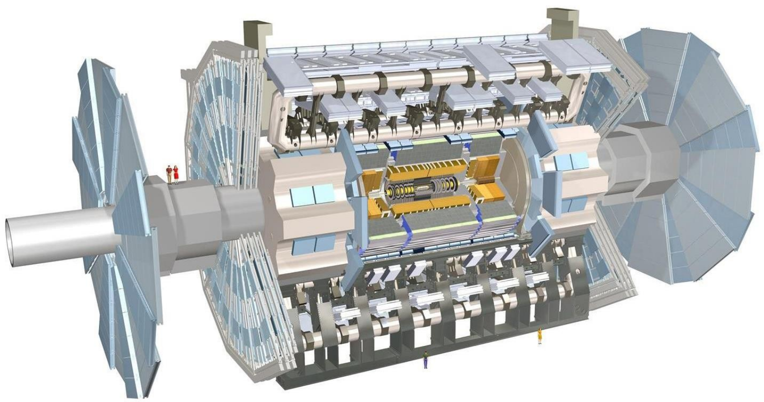




- Main keywords
  - measure electrical phenomenon
  - computer
  - programmable software

The problem is….

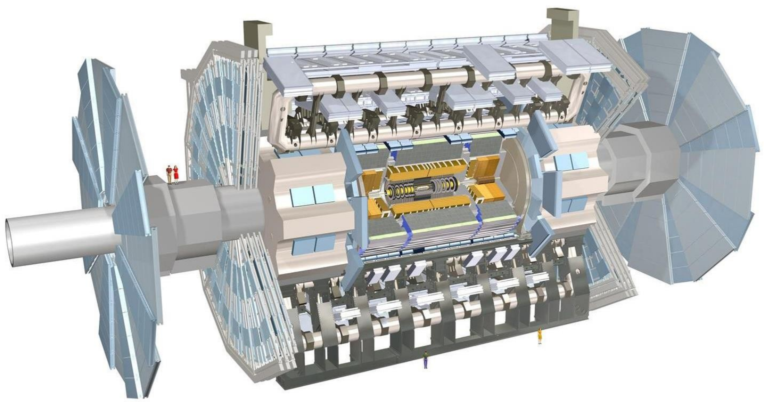
- … our sensor is really BIG



- ~100 million channels
- ~1 MB of RAW data per measurement

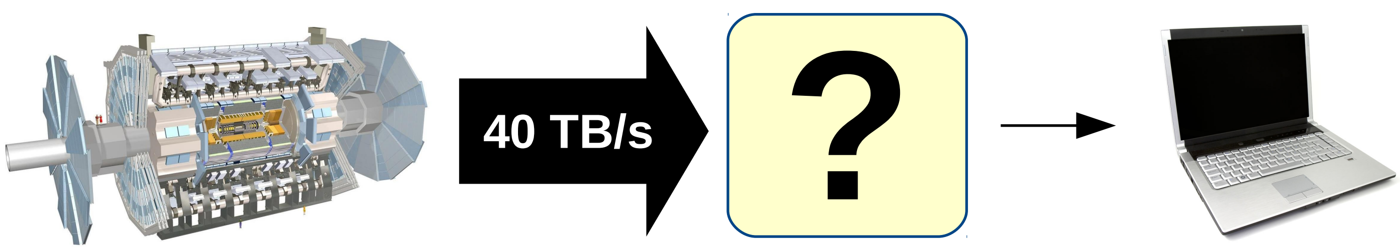
The problem is….

- … our sensor is really BIG



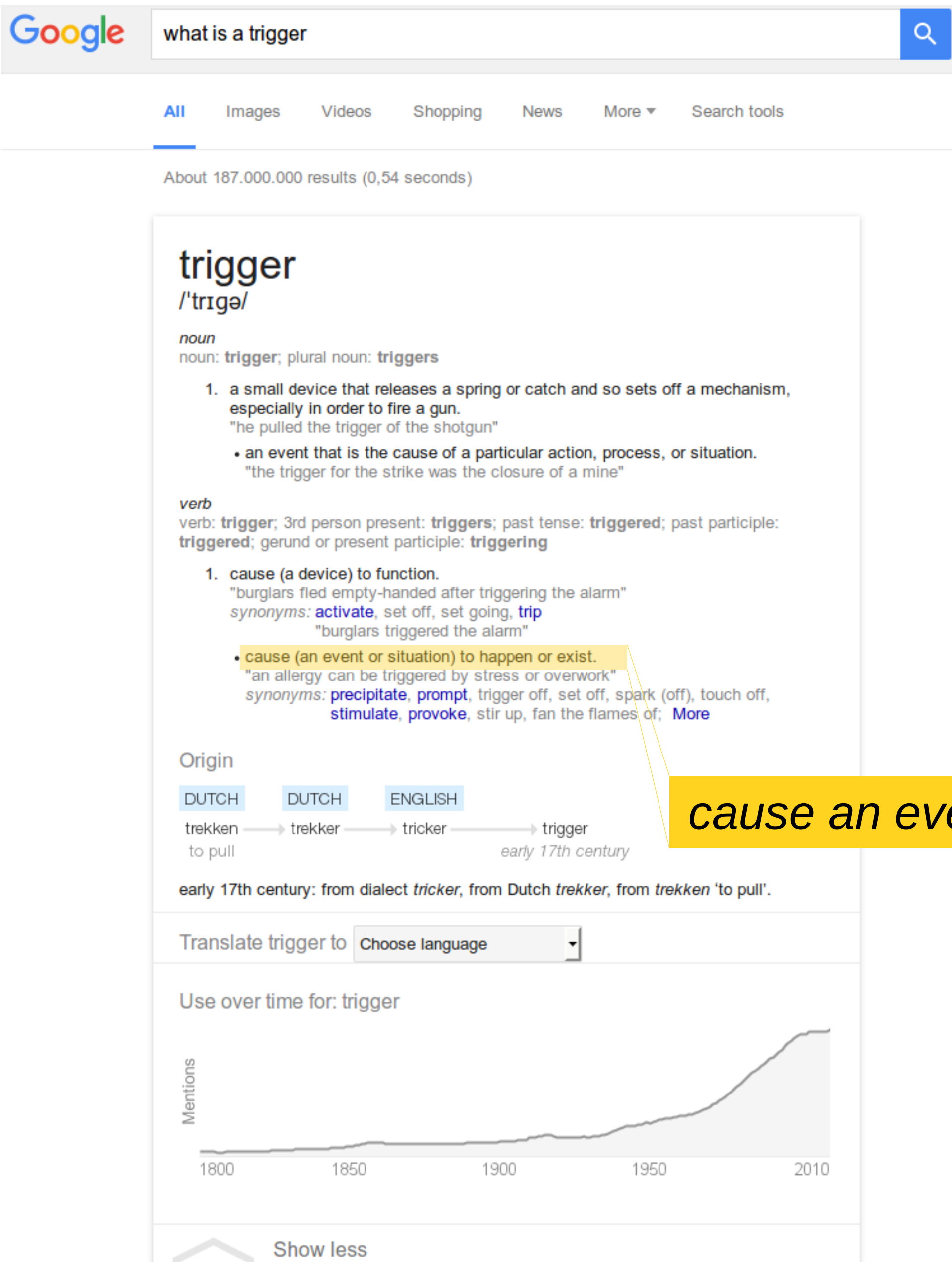
- ~100 million channels
- ~1 MB of RAW data per measurement

- … and really FAST
  - ~40 MHz measurement rate (every 25ns)



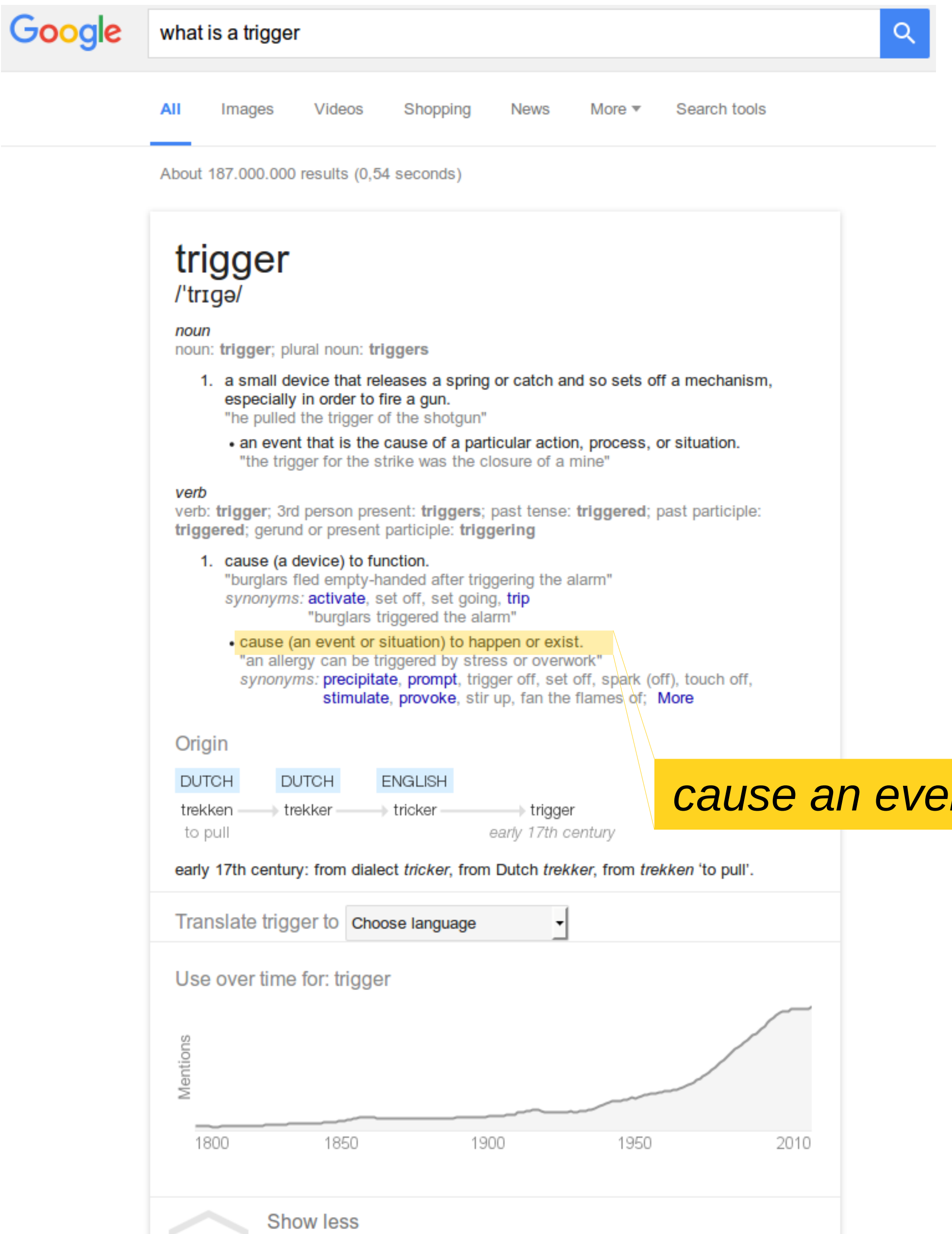
What was this lecture about?

What is a trigger?



cause an event to happen or exist

What is a trigger?



Trigger (particle physics)

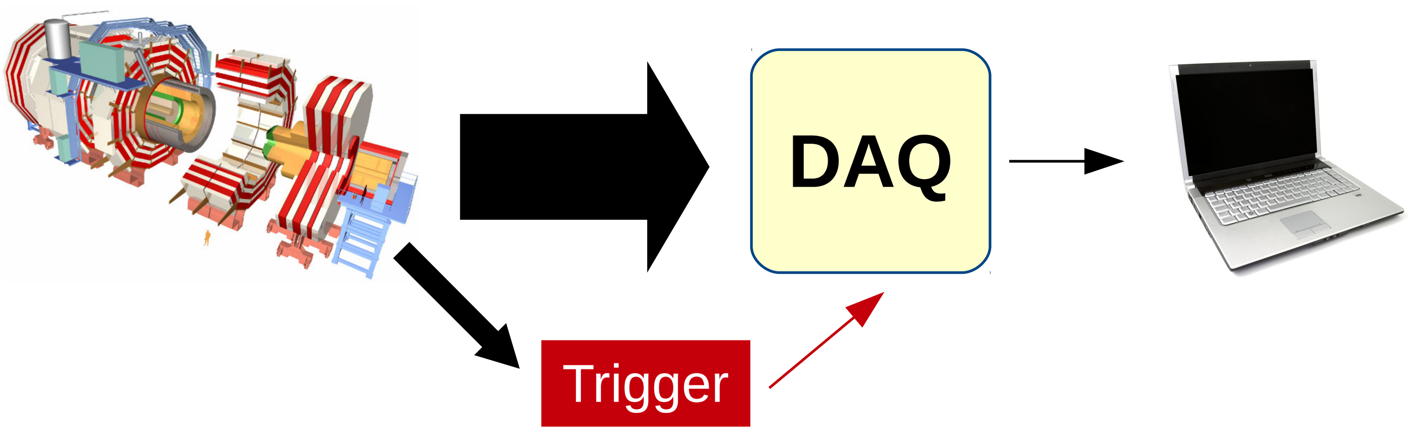
From Wikipedia, the free encyclopedia

In particle physics, a **trigger** is a system that uses criteria to rapidly decide which events in a particle detector to keep when only a small fraction of the total can be recorded. Trigger systems are necessary due to real-world limitations in computing power, data storage capacity and rates. Since experiments are typically searching for "interesting" events (such as decays of rare particles) that occur at a relatively low rate, trigger systems are used to identify the events that should be recorded for later analysis. Current accelerators have event rates greater than 1 MHz and trigger rates that can be below 10 Hz. The ratio of the trigger rate to the event rate is referred to as the selectivity of the trigger. For example, the Large Hadron Collider (LHC) has an event rate of 40 MHz ($4 \cdot 10^7$ Hz), and the Higgs boson is expected to be produced there at a rate of roughly 1 Hz. The LHC detectors can manage to permanently store a few hundred events per second. Therefore the minimum selectivity required is 10^{-5}, with much stricter requirements for the data analysis afterwards.[1]

- Main keywords
  - rapidly decide
  - which events
  - small fraction

cause an event to happen or exist

Putting it together: Trigger and DAQ



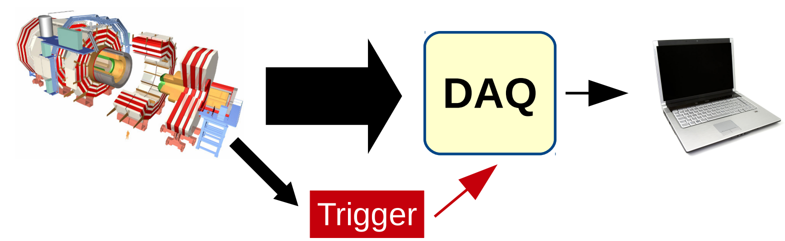
DAQ is responsible for collecting data from detector systems, digital conversion and recording them to mass storage for offline analysis.

Trigger is responsible for real-time selection of the subset of data to be recorded.

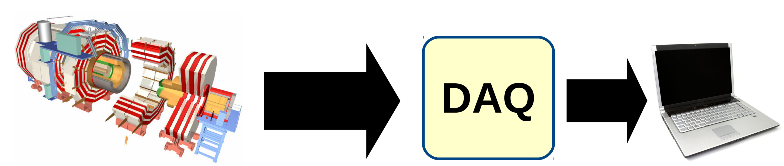
At collider experiments, the combined system of Trigger/DAQ is often referred to as **TDAQ**.

[To Trigger or not to Trigger]

- *Triggered* DAQ
  - Standard for all LHC experiments
  - Single or multi-level triggers in hardware and/or software
  - Always introduces some inefficiencies
    - Which may or may not be relevant for the Physics program

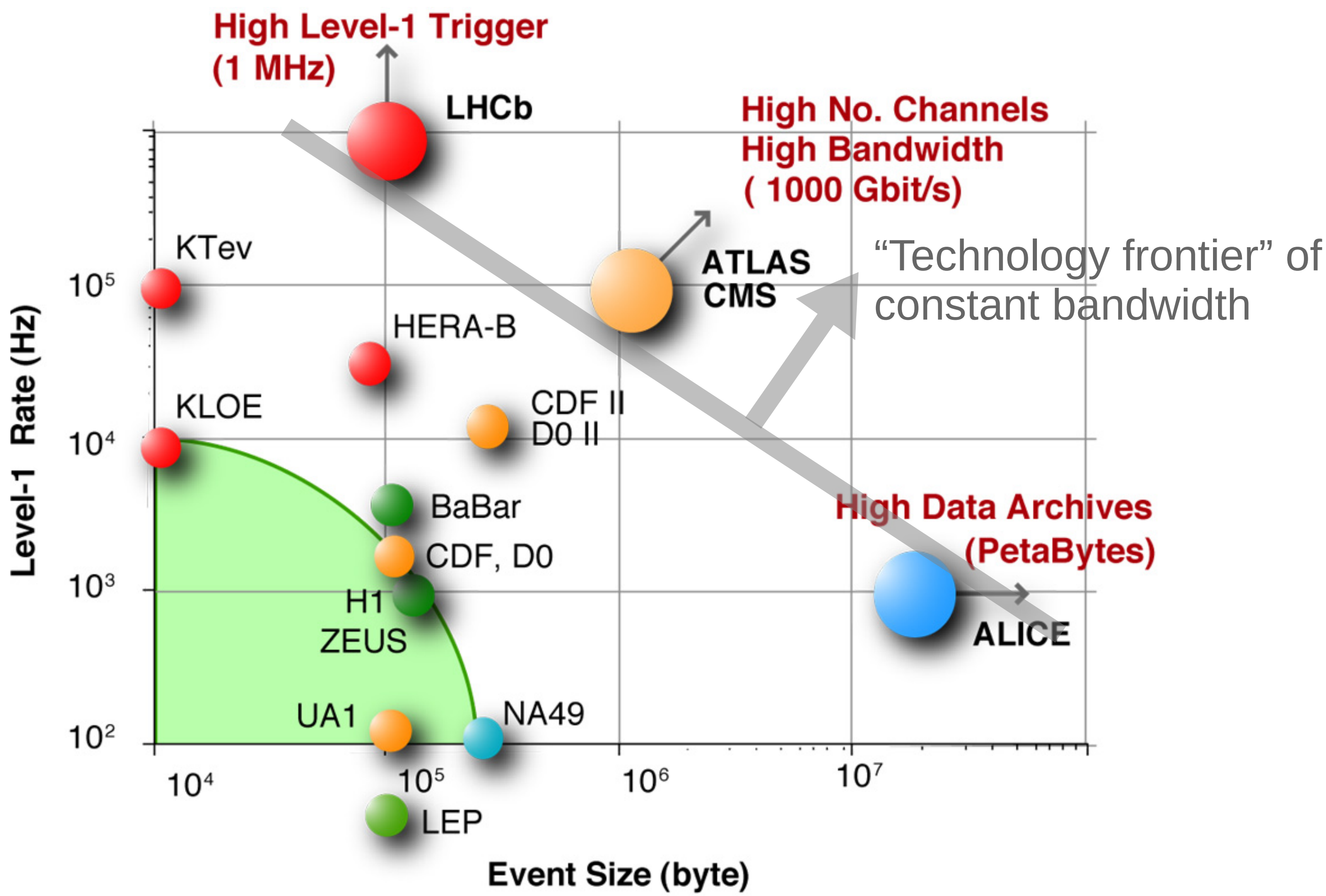


- *Triggerless* DAQ
  - Usually in the sense of hardware trigger
    - No or trivial external trigger
    - In most cases a software-based event filter is still used
  - Simpler and less custom build electronics
  - Not used in LHC experiments due to large event size
    - But LHCb/ALICE moving into that direction (see later)
  - non-LHC examples
    - LSST ~ 3 GB/s
    - mu2E ~ 30 GB/s
    - DUNE ~ 1 TB/s (in 2020+)
  - Will become standard for many experiments
    - Commercial technology enables this



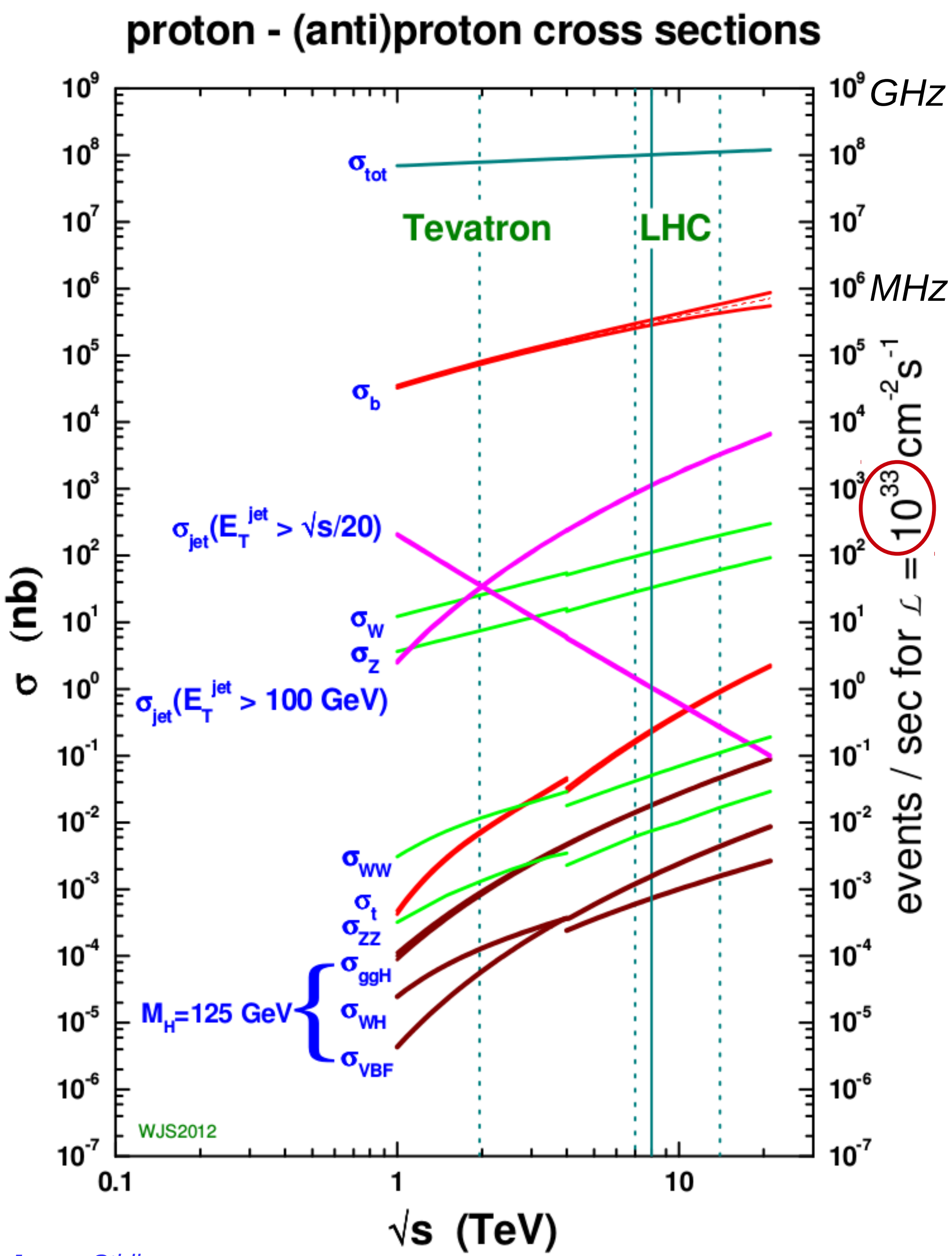
Experiment Landscape

- LHC experiments are pushing the limits
  - Bandwidth = (L1 trigger rate) x (Event size)
  - Not a coincidence that all 4 experiments are operating roughly at the same bandwidth
    - A result of what is technically possible and affordable



based on S. Cittolin, LHC DAQ Systems, 8th Workshop on Electronics for LHC Experiments (2002)

What to Trigger on (at hadron colliders) ?



James Stirling

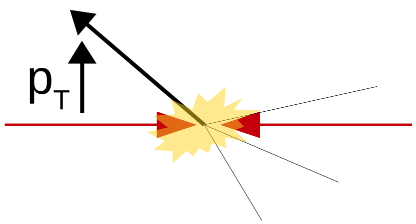
- At hadron colliders a typical collision is rather boring
  - 1 GHz @ 10^{33} cm^{-2} s^{-1}
- Interesting physics is 6-8 orders of magnitude rarer
  - Electro-weak (W/Z) and Top Physics
- Physics, the LHC was built to explore even more rare
  - Higgs produced in about 1 / 10^9 collisions
    - Detection rate is even lower

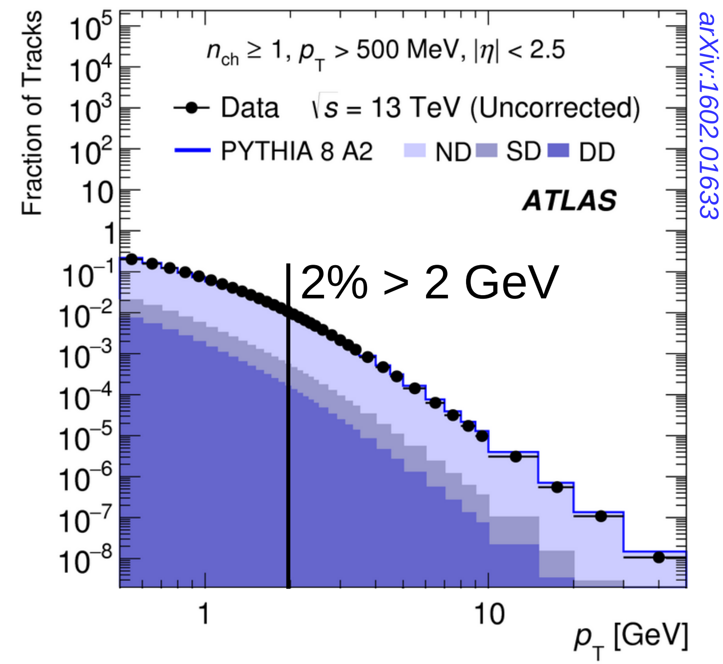
Even if we could, saving all events at <u>hadron colliders</u> is not useful.

not necessarily true at lepton colliders

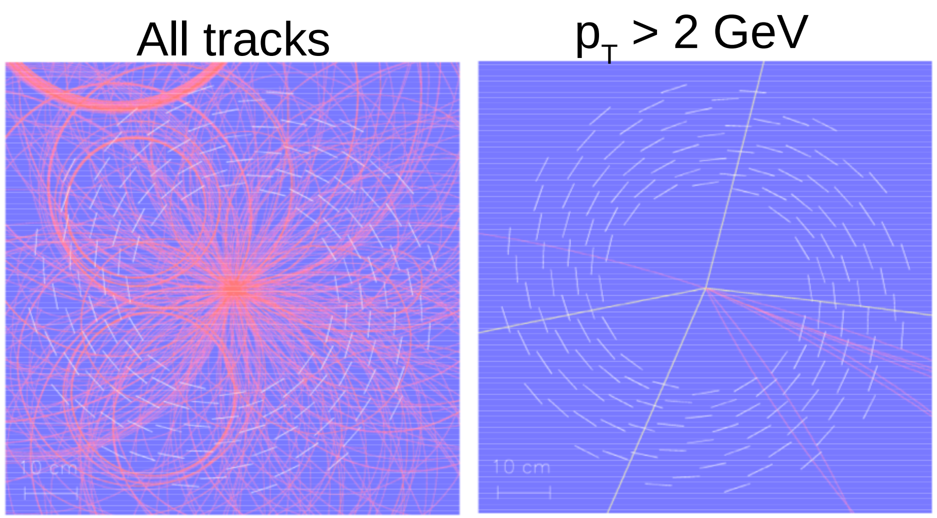
How to identify the “interesting” events

- Proton collisions produce mainly hadrons with low transverse momentum
  - Only 2% of all tracks have p_T > 2GeV





- “Interesting” physics is usually high-pt
  - H → γγ, $p_T(\gamma)$ ~ 50-60 GeV
  - W → eν, $p_T(e)$ ~ 30-40 GeV
  - Obvious signature to use in trigger

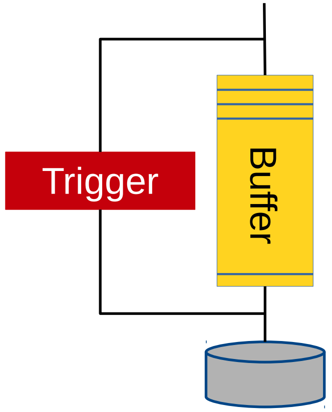


Simulated $H \rightarrow 4\mu$ + 17 minbias events

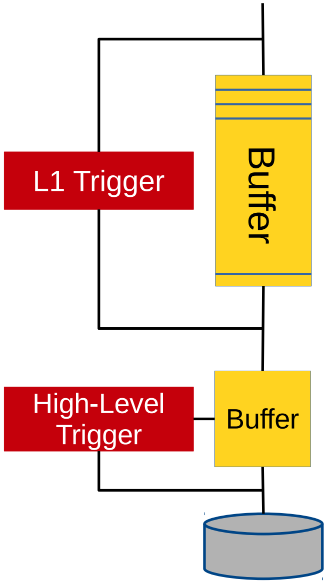
- [What if new physics is “soft” ?]
  - That's where triggering becomes very challenging...

How fast do we need to Trigger?

- While the Trigger decision is taken the data must be buffered
  - Can be done on-detector (front-ends) or off-detector
    - Buffering of either analogue or digital signals
    - Usually implemented as FIFO pipelines
    - Off-detector allows for bigger buffer sizes (but higher readout bandwidth)
  - The buffer size defines the maximum trigger latency
    - 100 event buffer @ 40 MHz → 2.5 μs trigger latency
    - Buffers also serve as derandomizers to smooth out fluctuations
  - Once the latency (buffer size) is defined, very difficult to increase
    - Requires replacement of all front-end electronics
    - Big issue for the multi-decade LHC experiments

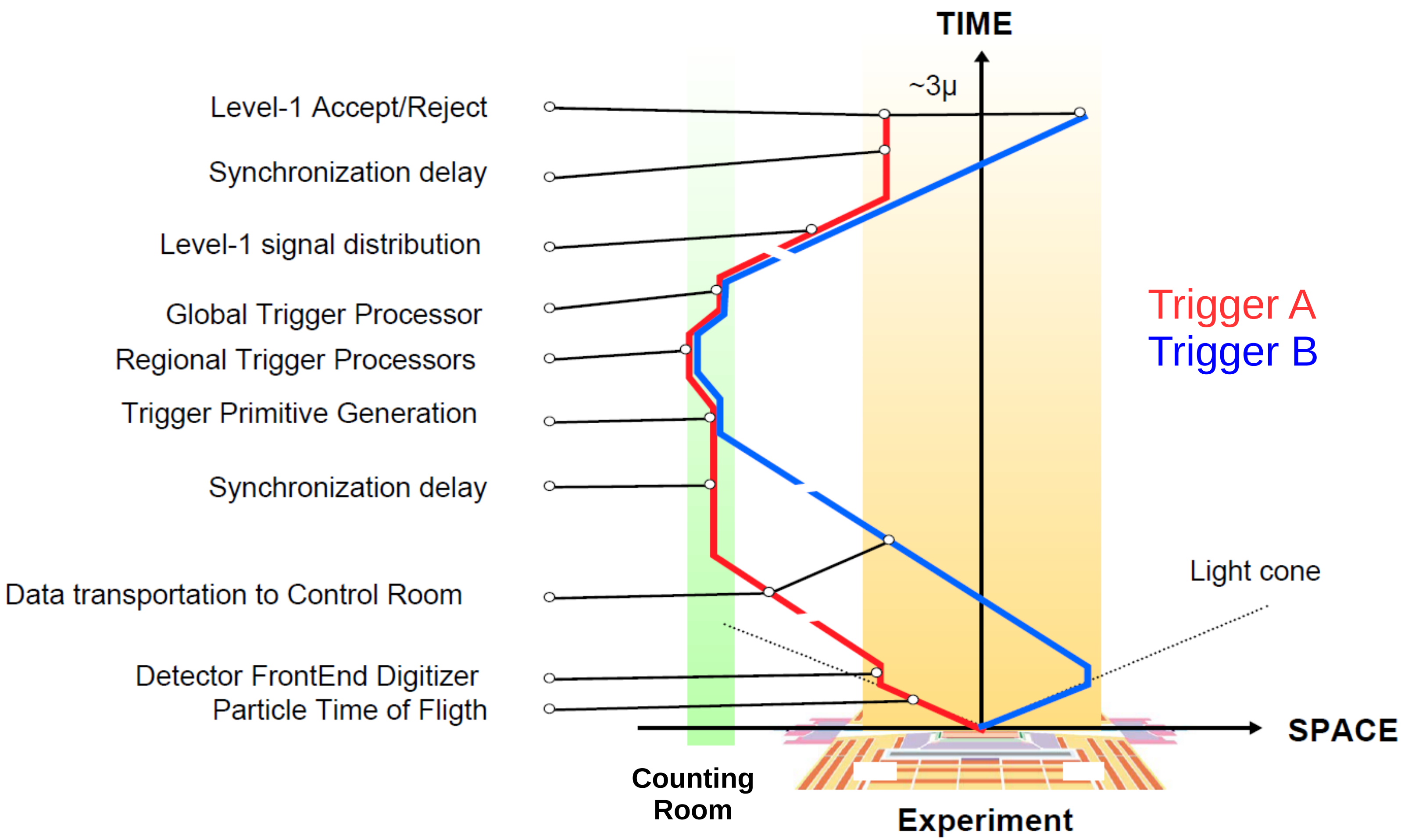


- Multi-level trigger systems
  - Reducing rates in stages allows for longer latencies
    - Fast electronics-based First-level trigger (L1) with O(μs) latency
    - Software-based Event Filter or High-Level Trigger (HLT) with O(s) latency
  - Standard for essentially all current collider experiments



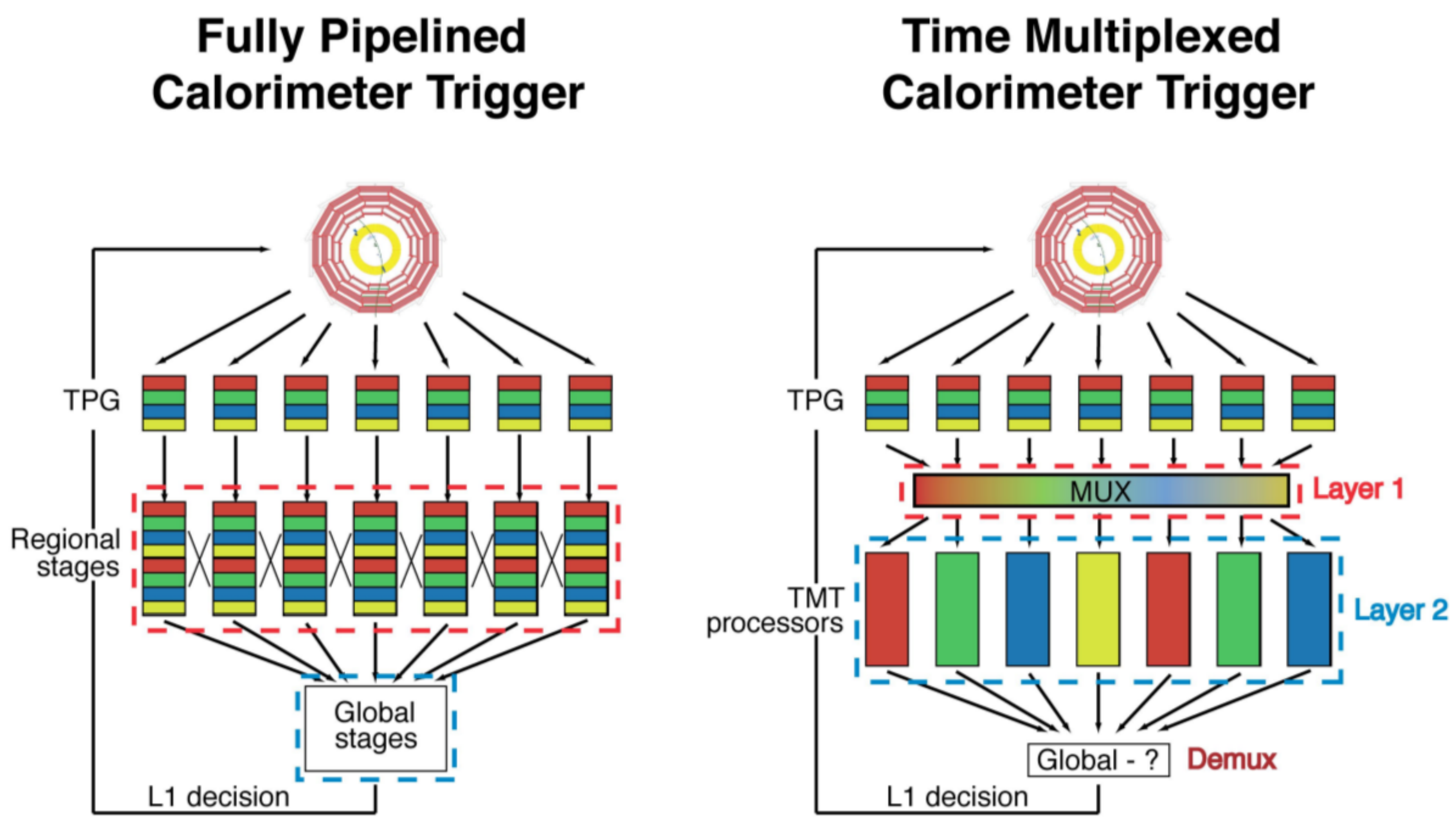
L1 Latency

- Time between collision and arrival of L1 decision
  - Sum of many contributions (here CMS)
  - Significant time spent in transmitting signals (10m/c=33ns)
  - Synchronization delays needed for signals from/to different parts of the detector
    - Due to physical location, cable length and processing time on front-ends



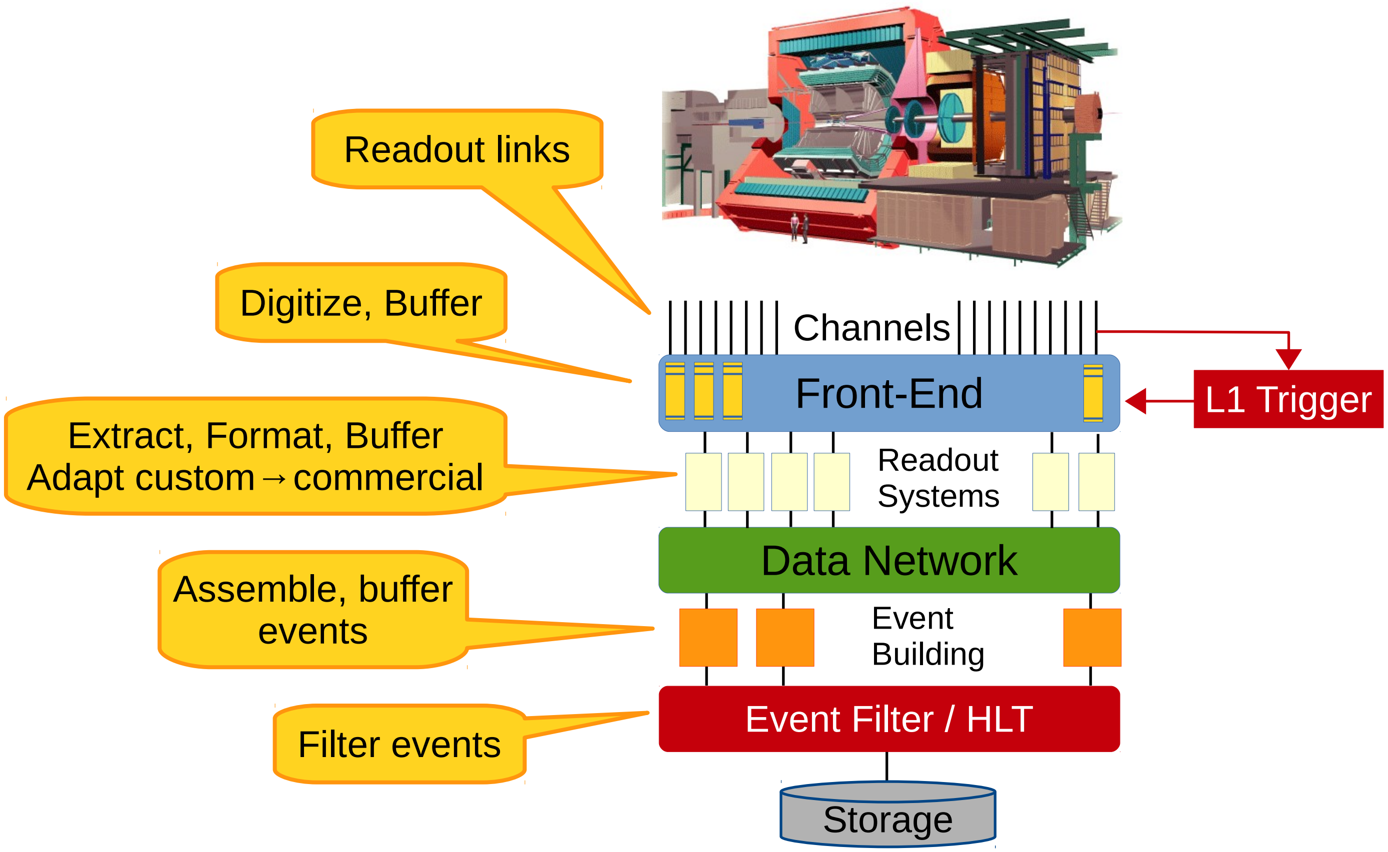
Pipelining and Multiplexing

- Need to process multiple events in parallel
  - Pipelined: Step-wise processing and split event into regions
  - Time Multiplexed: one event per processor that performs all steps

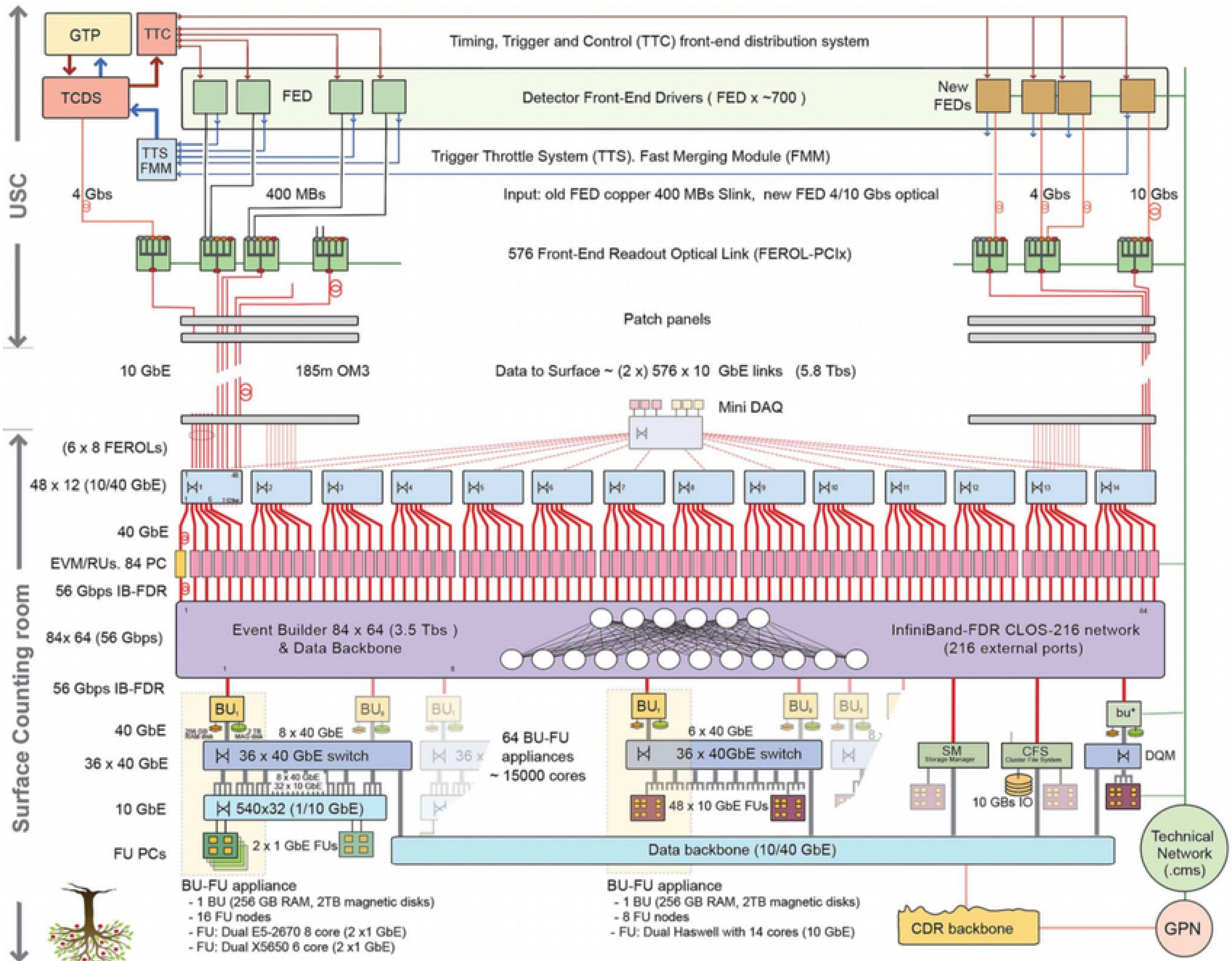


CMS-CR-2012-300

A generic Trigger/DAQ system



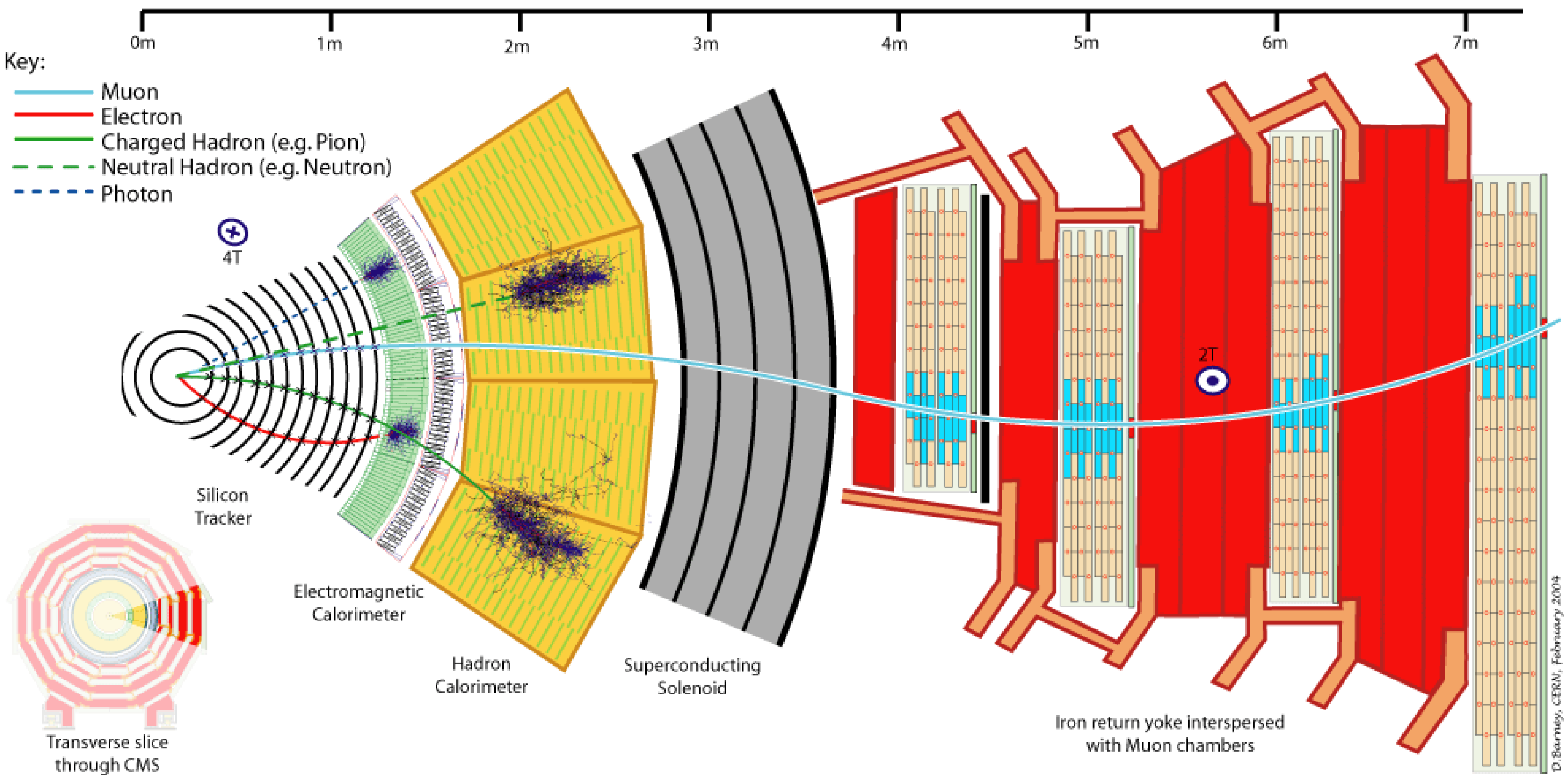
And in reality... (CMS)



Trigger Algorithms: L1 and HLT

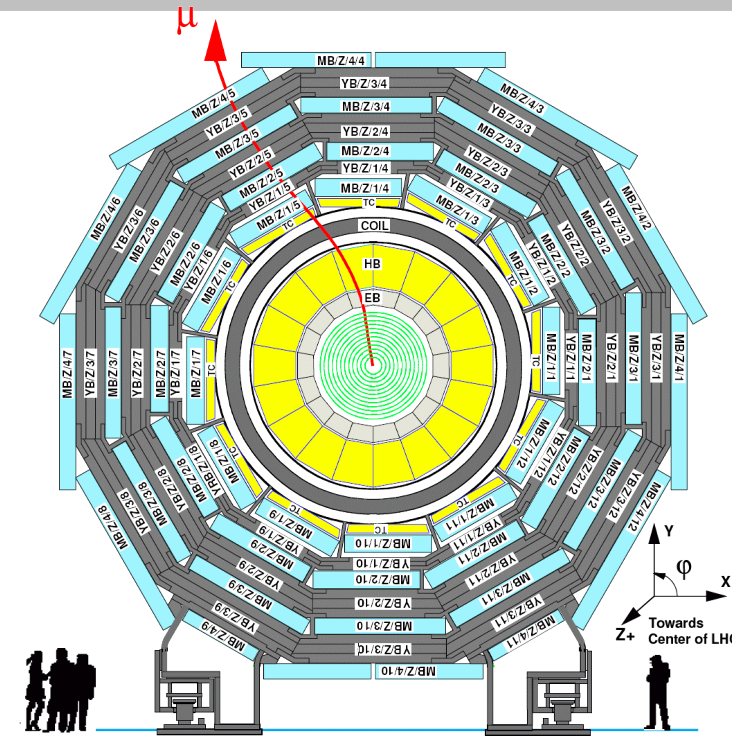
Which objects to Trigger on?

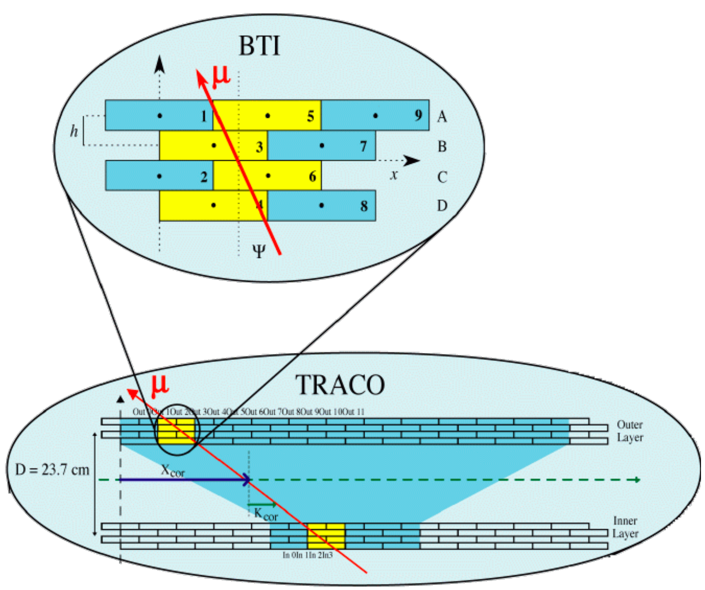
- High-pT objects leave very distinct traces in our detectors
  - Electron, photons, hadrons (jets), muons
  - In addition can use global event quantities
    - Total energy
    - Missing transverse momentum/energy

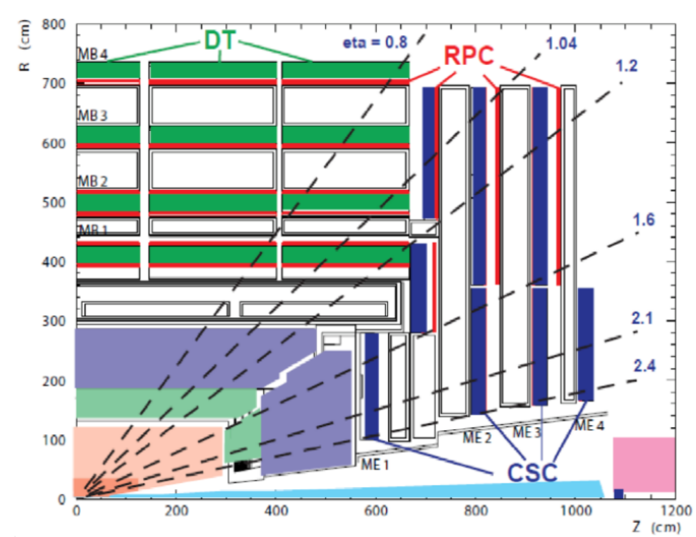


L1 Muon Trigger (CMS)

- Mostly muons traverse the calorimeter
    - Build coincidence stations and measure momentum in magnetic field
    - Reconstruct local segments on chambers using ASICs
    - Combine segments to tracks using lookup tables
    - p_T resolution typically not very good (~20%)
- Dedicated trigger chambers in use
    - For triggering need “fast” detectors (e.g. RPC, CSC)

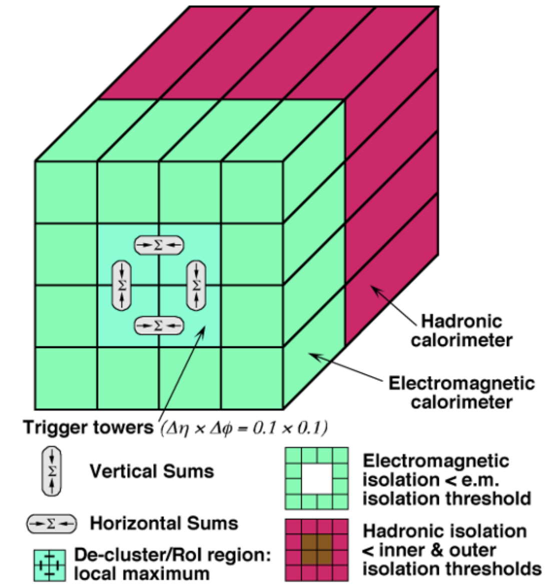






L1 Calorimeter Trigger (ATLAS)

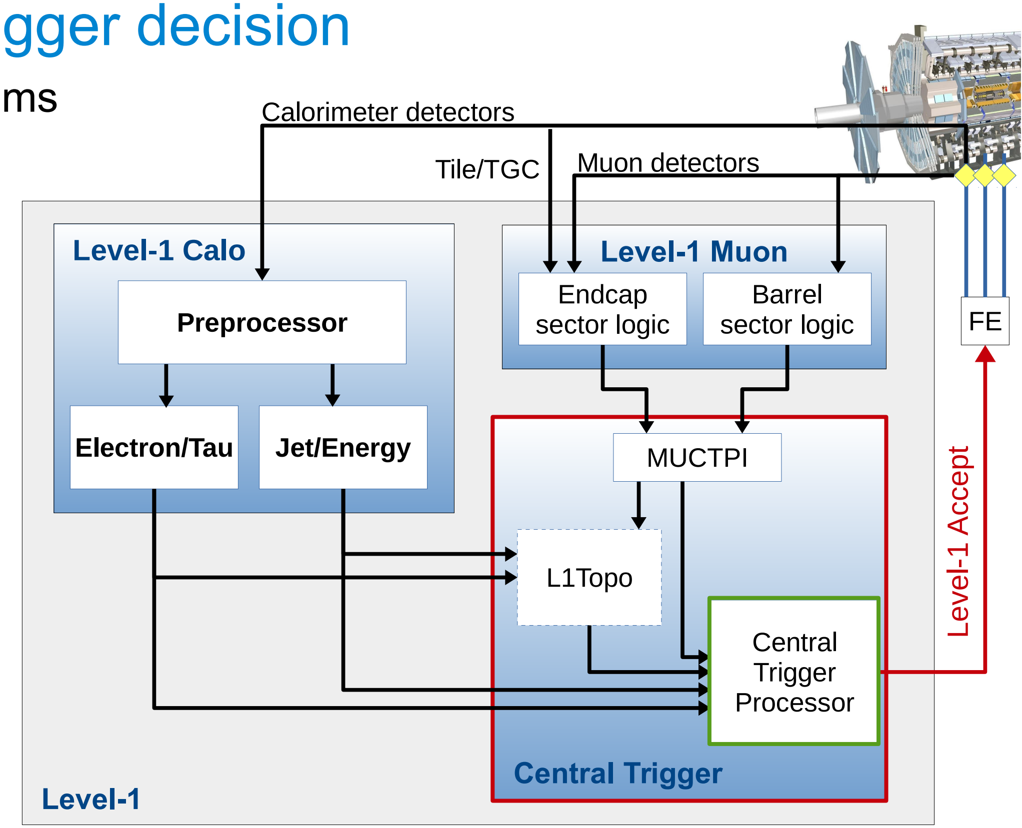
- For triggering on any EM objects
  - Electrons, Photons, Jets, Taus
  - Global event quantities
- Pre-processing
  - Several calorimeter cells are summed into trigger towers
    - Either analogue (ATLAS) or digital (CMS)
  - Resulting in towers of reduced granularity, e.g. $\eta \times \phi = 0.1 \times 0.1$
    - In ATLAS ~7000 calorimeter trigger towers
- Object reconstruction
  - Find local maximum via sliding window algorithm
    - Apply energy selection based on sum in towers
  - Window size depending on object
    - Electron/Photon 0.2 x 0.2
    - Jets 0.4 x 0.4
  - Can apply additional selections
    - EM Isolation (ring around core)
    - Hadronic isolation (no activity in had. layer)



Analogue trigger cables (ATLAS L1Calo)

Global/Central Trigger

- Responsible for making the final L1 trigger decision
  - Taking the inputs from the various L1 sub-systems
    - Time-align the signals
  - Apply prescales
  - Apply multiplicities
    - e.g. 2 muons with p_T > 20 GeV
  - Apply logical selections
    - e.g. 1 tau AND 1 electron with p_T > 10 GeV
  - Apply topological selections
    - e.g. two jets with $|\Delta\eta|$ > 1.5
    - e.g. two muons with 2.8 < $M(\mu\mu)$ < 3.2



ATLAS CTP

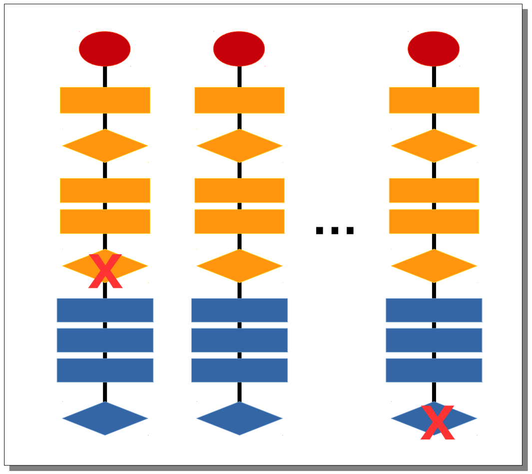
Event Filtering (HLT)

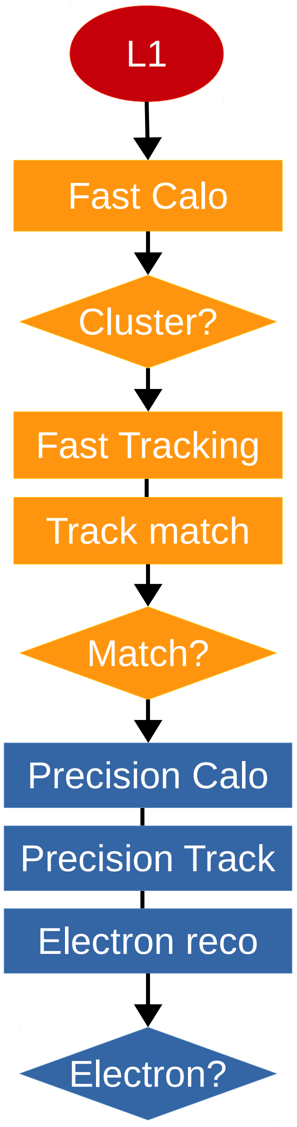
- Software running on large commercial PC farms
  - Typically running one filtering application per physical CPU core
    - Required number cores = Input rate * <processing time>
    - Example: 100 kHz * 200ms = 20.000 cores
    - Peaks are absorbed by appropriate buffers
- Events can be processed independently
  - “Embarrassingly parallel”
  - Bottle-neck is RAM/application, i.e. with trend to many-core CPUs
    - Use memory sharing techniques
    - Multi-threading becoming increasingly important
- Networking based on commercial technologies
  - Ethernet
  - InfiniBand

Typical HLT node:
2x12-core Intel Xeon Haswell
→ 96 cores/box
48 GB RAM, 10Gb Ethernet
4 motherboards in 2U box

Event Filter Algorithms

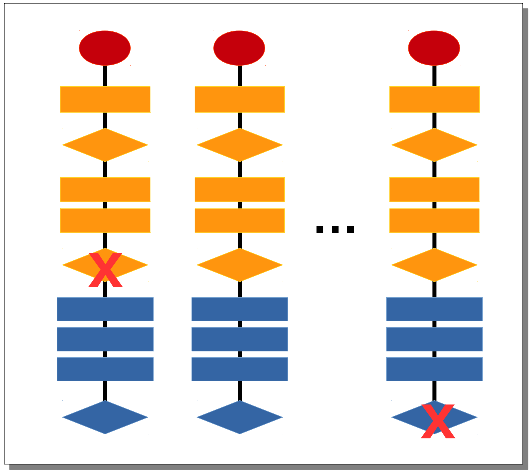
- Offline reconstruction too slow to be used directly
  - Takes >10s per event but HLT usually needs << 1s
- Requires step-wise processing with early rejection
  1) Fast reconstruction
     - Trigger-specific or special configurations of offline algorithms
     - L1-guided regional reconstruction
  2) Precision reconstruction
     - Offline (or very close to) algorithms
     - Full detector data available
  - Stop processing as soon as one step fails
- Streaming
  - Event gets accepted if any trigger passes
  - Events can be written to different streams depending on which trigger passes





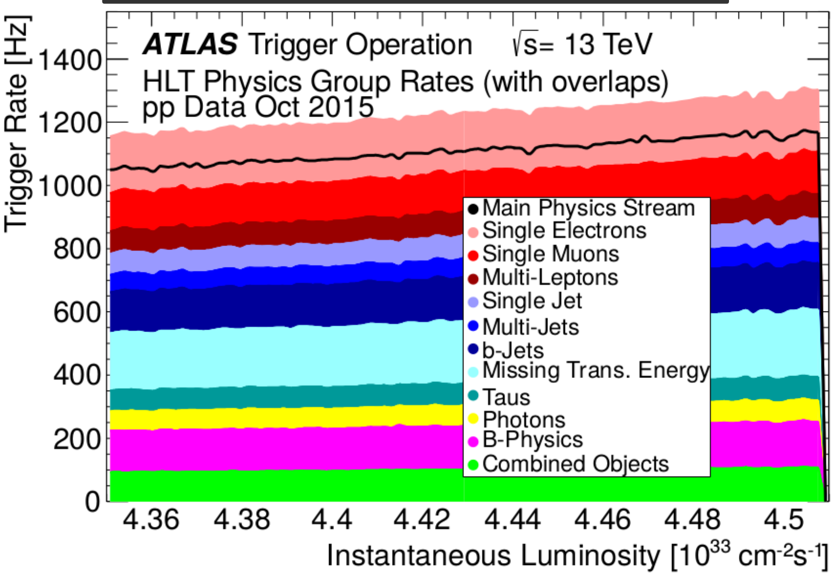
Trigger Menu / Table / List

- Defines the Physics program/reach of the experiment
  - Each physics signature defines one or more trigger “lines”
  - Collection of trigger “lines” is the trigger menu / table / list
  - In addition to primary physics triggers, contains
    - support triggers (e.g. for efficiency measurements)
    - triggers for detector calibration and monitoring
  - Usually menu contains several 100 trigger lines
    - ATLAS run-2 menu currently contains ~1800 trigger lines



- Trigger menu varies with luminosity and time
  - Constantly fine-tuned according to running conditions
- Trigger Menu design driven by
  - Physics priorities
  - Rate limitations at all trigger levels
  - Online resources (CPU, bandwidth)

ATLAS HLT Trigger Rates by group



Example: Main ATLAS Physics Triggers

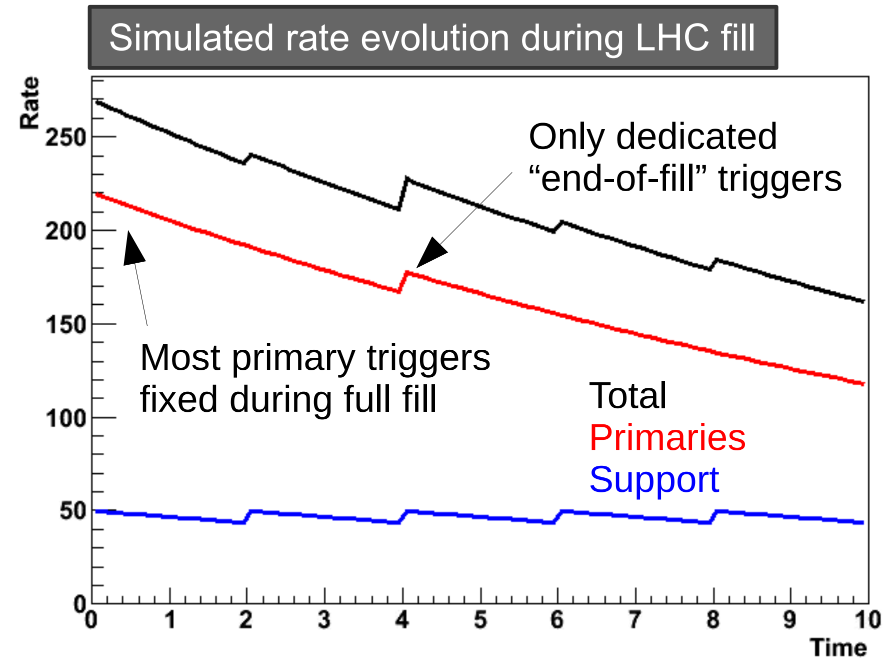
Trigger	Typical offline selection	Trigger Selection		Level-1 Rate [kHz]	HLT Rate [Hz]
		Level-1 [GeV]	HLT [GeV]	$L = 5 \times 10^{33}$ cm^{-2}s^{-1}	
Single leptons	Single iso μ, $p_T > 21$ GeV	15	20	7	130
	Single e, $p_T > 25$ GeV	20	24	18	139
	Single μ, $p_T > 42$ GeV	20	40	5	33
	Single τ, $p_T > 90$ GeV	60	80	2	41
Two leptons	Two μ's, each $p_T > 11$ GeV	2×10	2×10	0.8	19
	Two μ's, $p_T > 19, 10$ GeV	15	18, 8	7	18
	Two loose e's, each $p_T > 15$ GeV	2×10	2×12	10	5
	One e & one μ, $p_T > 10,26$ GeV	20 (μ)	7, 24	5	1
	One loose e & one μ, $p_T > 19,15$ GeV	15, 10	17, 14	0.4	2
	Two τ's, $p_T > 40,30$ GeV	20, 12	35, 25	2	22
	One τ, one μ, $p_T > 30, 15$ GeV	12, 10 (+jets)	25, 14	0.5	10
	One τ, one e, $p_T > 30, 19$ GeV	12, 15 (+jets)	25, 17	1	3.9
Three leptons	Three loose e's, $p_T > 19, 11, 11$ GeV	$15, 2 \times 7$	$17, 2 \times 9$	3	< 0.1
	Three μ's, each $p_T > 8$ GeV	3×6	3×6	< 0.1	4
	Three μ's, $p_T > 19, 2 \times 6$ GeV	15	$18, 2 \times 4$	7	2
	Two μ's & one e, $p_T > 2 \times 11, 14$ GeV	2×10 (μ's)	$2 \times 10, 12$	0.8	0.2
	Two loose e's & one μ, $p_T > 2 \times 11, 11$ GeV	$2 \times 8, 10$	$2 \times 12, 10$	0.3	< 0.1
One photon	One γ, $p_T > 125$ GeV	22	120	8	20
Two photons	Two loose γ's, $p_T > 40, 30$ GeV	2×15	35, 25	1.5	12
	Two tight γ's, $p_T > 25,25$ GeV	2×15	2×20	1.5	7
Single jet	Jet ($R = 0.4$), $p_T > 400$ GeV	100	360	0.9	18
	Jet ($R = 1.0$), $p_T > 400$ GeV	100	360	0.9	23
E_T^{miss}	$E_T^{miss} > 180$ GeV	50	70	0.7	55
Multi-jets	Four jets, each $p_T > 95$ GeV	3×40	4×85	0.3	20
	Five jets, each $p_T > 70$ GeV	4×20	5×60	0.4	15
	Six jets, each $p_T > 55$ GeV	4×15	6×45	1.0	12
b–jets	One loose b, $p_T > 235$ GeV	100	225	0.9	35
	Two medium b's, $p_T > 160,60$ GeV	100	150,50	0.9	9
	One b & three jets, each $p_T > 75$ GeV	3×25	4×65	0.9	11
	Two b & two jets, each $p_T > 45$ GeV	3×25	4×35	0.9	9
B–physics	Two μ's, $p_T > 6,4$ GeV plus dedicated J/ψ-physics selection	6,4	6,4	8	52
Total				70	1400

ATL-DAQ-PUB-2016-001

ATLAS Run-2 trigger menu for 5×10^{33} cm^{-2}s^{-1}, full menu contains >1800 items

Prescaled Triggers

- Not all triggers need to run at full rate
  - Rate might be just too high
  - Often just a subsample is enough (support triggers)
  - Adding triggers when luminosity drops to make optimal use of resources
- Prescales are used to reduce rate
  - Prescale of N (e.g. N=10)
    - Only accept 1 out of N events
  - Useful to have fractional prescales (e.g. N=1.5)
- Prescales can be applied at any trigger level
  - To save resources should be done as early as possible
  - Usually done in global trigger logic



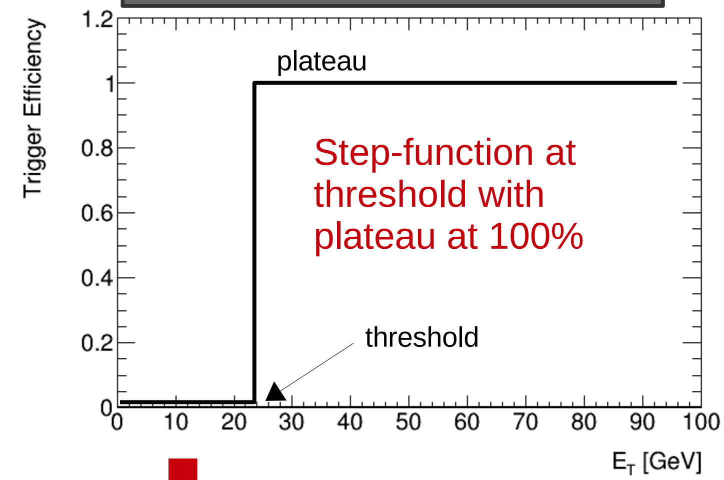
Trigger Efficiency

- Efficiency for triggering/recording an event?
  - Trigger is just another “cut” in the physics analysis event selection
  - Very important to measure efficiency for cross-section measurements, etc.
- Definition
  - Trigger efficiency usually measured w.r.t. offline reconstruction
    - e.g. # triggered electrons vs # offline electrons

$$\epsilon_{trigger} = \frac{N_{trigger}}{N_{offline}}$$

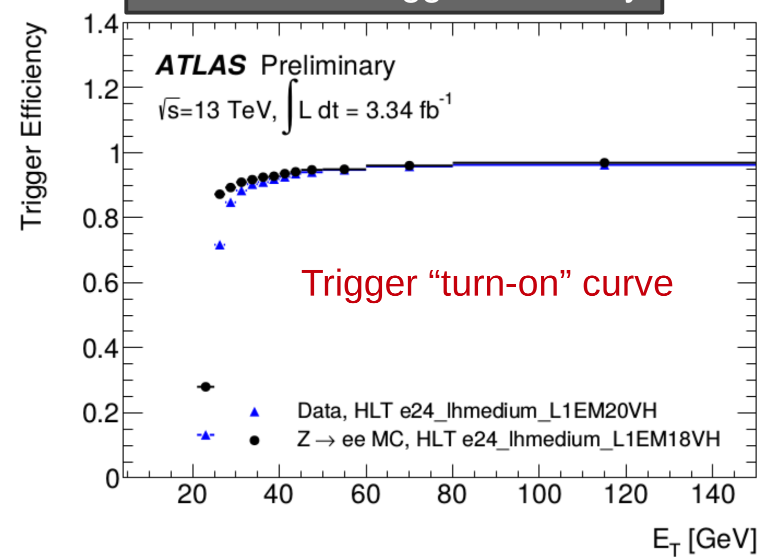
- Measurement via
  - Tag-and-probe
    - Trigger on particle from resonance ($Z \rightarrow \mu\mu$) and measure how often second particle (probe) passes trigger selection
  - Boot-strap
    - Use looser (prescaled) trigger (e.g. 40 GeV jet to measure 60 GeV trigger eff)
  - Orthogonal trigger
    - Trigger on one physics signature, measure a different one
  - Simulation

Ideal 24 GeV trigger efficiency



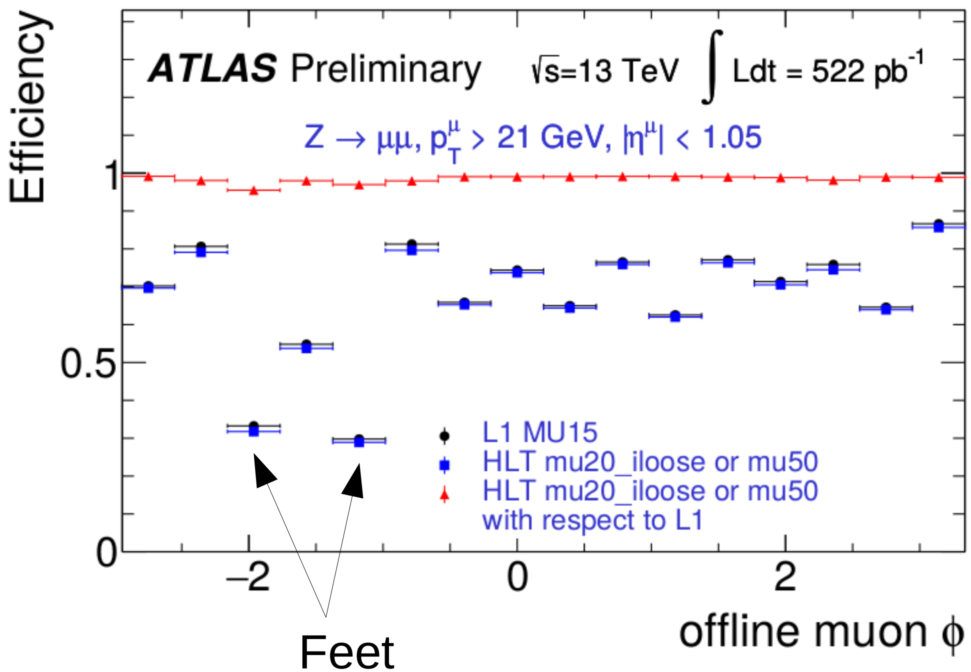
Resolution
Inefficiencies
Trigger/offline differences

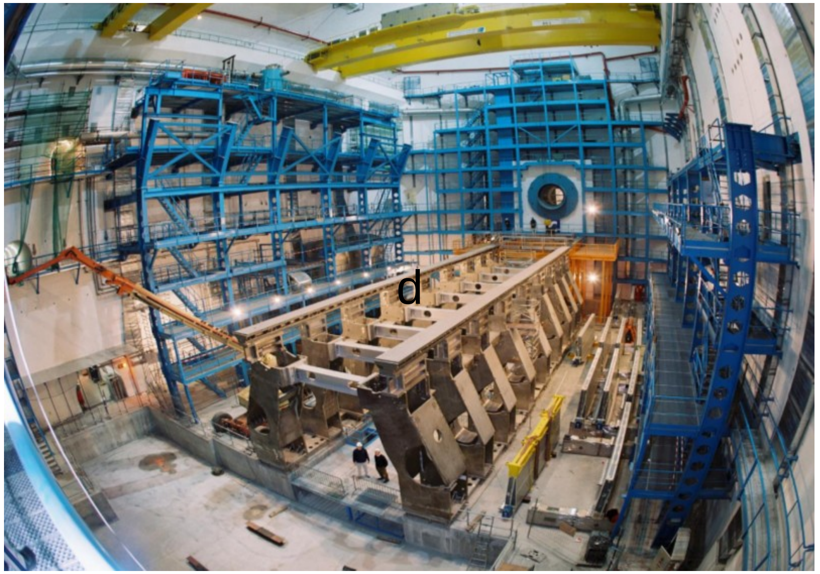
Real 24 GeV trigger efficiency



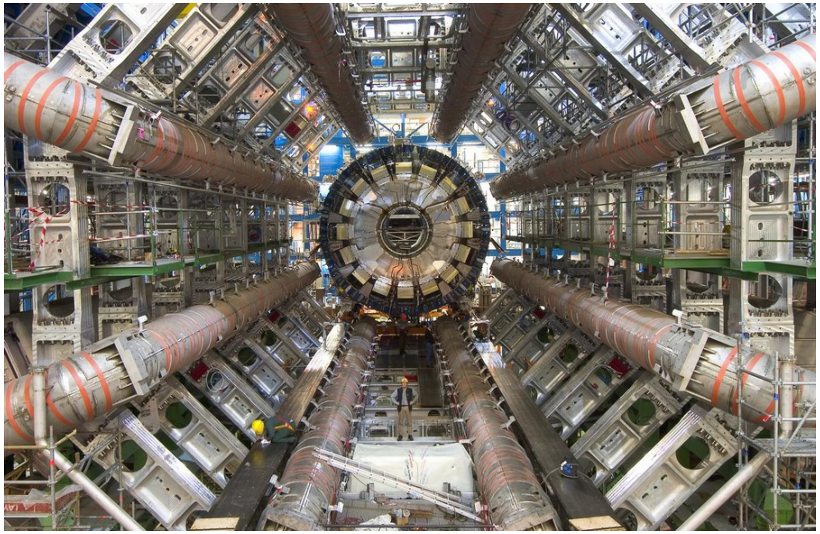
Trigger Efficiency

- Trigger efficiency is not constant
  - Can also depend on geometry
  - Example: ATLAS barrel muon trigger
    - Support structure prevents having trigger chambers uniformly installed
    - Important to have a good detector simulation





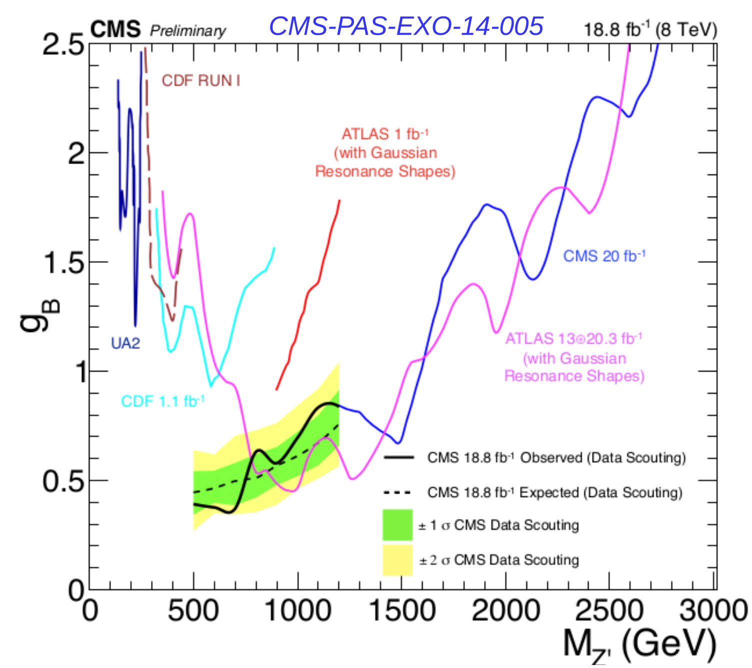
The feet of ATLAS (2004)



ATLAS support structure (2005)

Physics analysis @ HLT

- HLT output rate limited by offline storage/computing
  - Some analyses are limited by the thresholds used in the trigger
  - But the HLT “sees” many more events
  - Can we do physics analysis directly at the HLT?
    - HLT very similar to offline reconstruction already
  - Only write reconstructed objects instead of full event
    - Data Scouting (CMS)
    - Turbo Stream (LHCb)
    - Trigger-Level Analysis (ATLAS)
    - (ALICE does not save RAW data at all)
- Example: CMS di-jet resonance search
  - Limited by jet trigger threshold
  - Using Data Scouting allows to achieve a significant lower limit in di-jet invariant masses
    - Reduces the lower limit on Z' from 1.2 → 0.5 GeV



- Next logical step
  - Combine trigger and offline (see later)

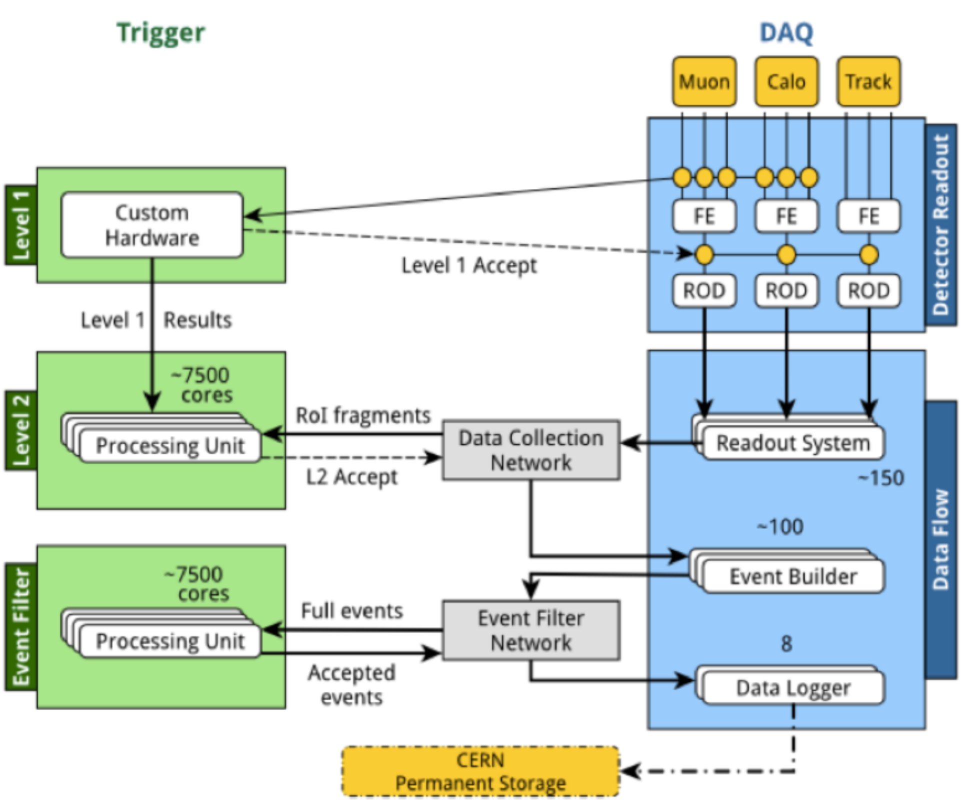
TDAQ Changes for Run-2

- Only showing main architectural changes during LS1
- Huge number of improvements by all experiments

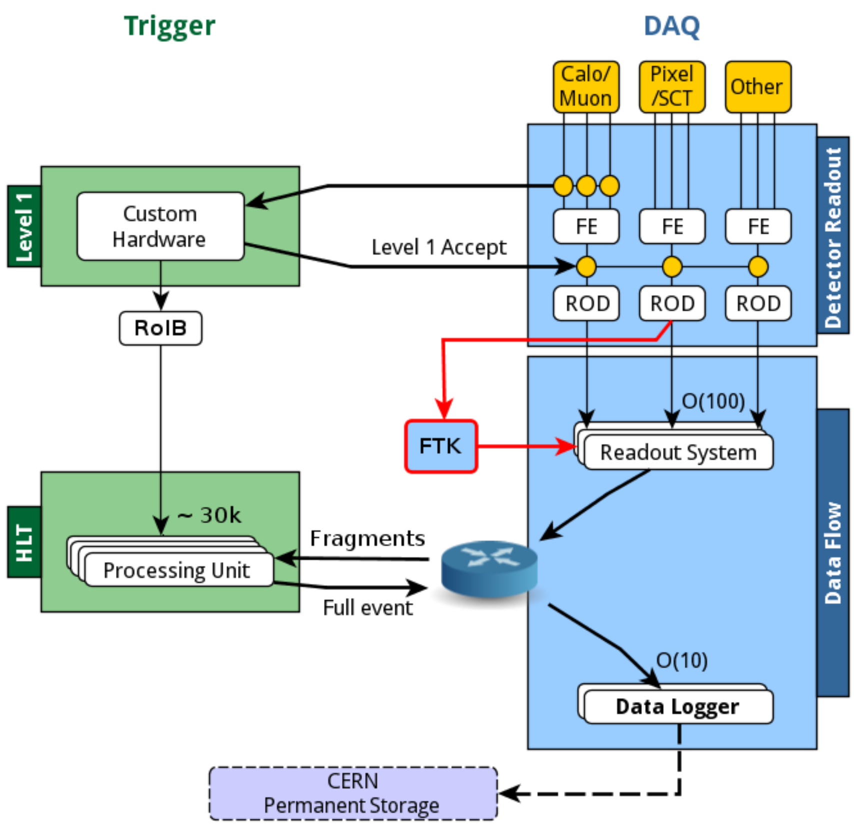
ATLAS TDAQ system evolution

- Merge of two-level HLT system during LS1
  - Following CMS single-level HLT design
  - Huge simplification and less resource limits for HLT algorithms
    - No longer a fixed L2 output rate limit of ~6 kHz
  - But keeping Region-Of-Interest based approach for data requests

ATLAS Run-1

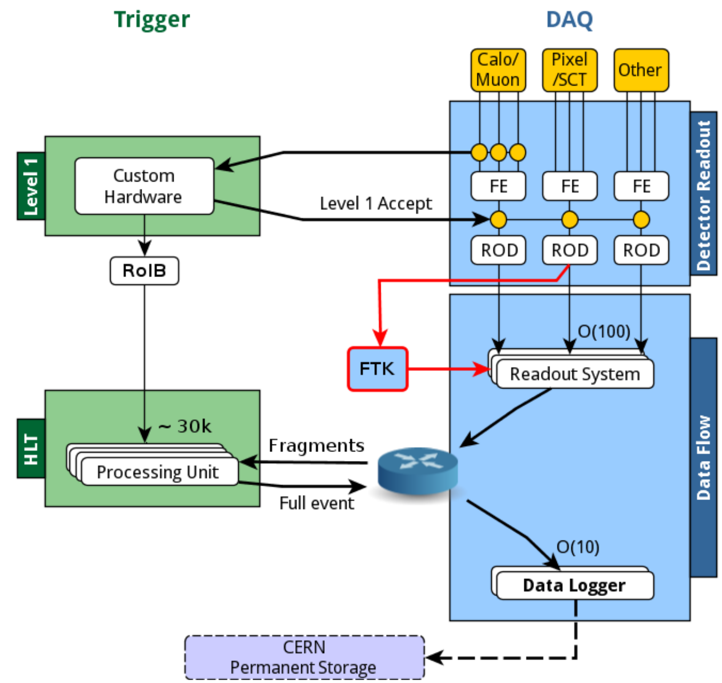


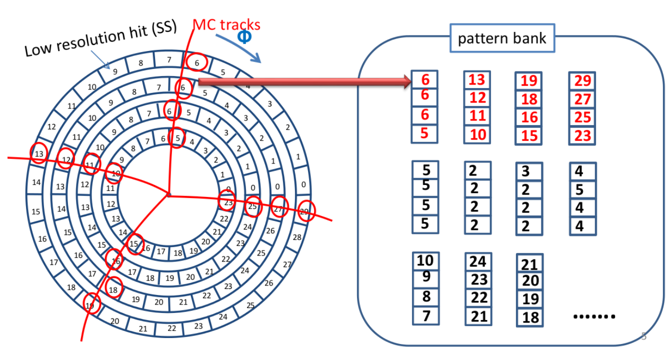
ATLAS Run-2

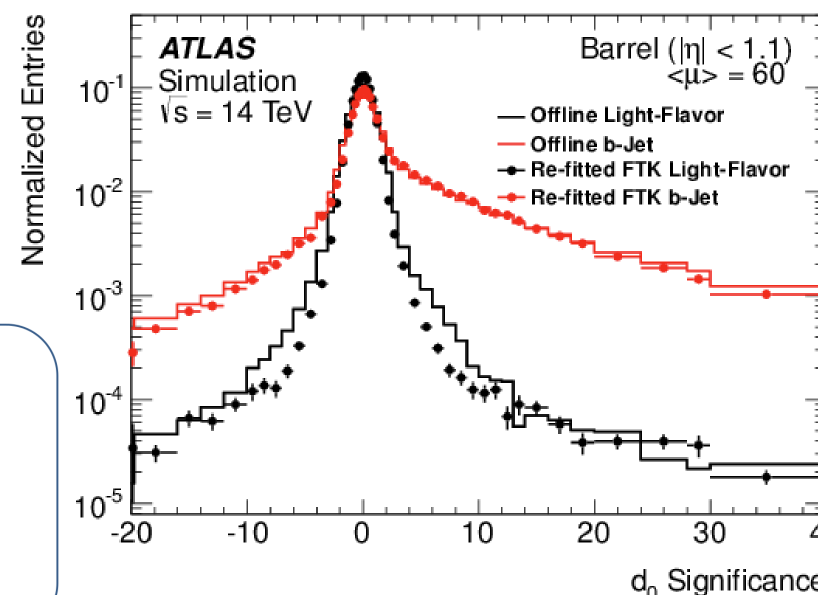


ATLAS Fast TracKer (FTK)

- A co-processor for the ATLAS HLT
  - Based on CDF's Silicon Vertex Tracker (SVT)
  - Not a trigger, performs tracking at 100 kHz (p_T > 1 GeV)
    - O(100µs) compared to O(100ms) in software
  - Tracks for full event available to HLT
  - Fully installed (up to µ=40) by end of 2016
- Expected improvements:
  - Resolve the topology of b- and tau-jets
  - Determine the number and position of the primary vertex
  - Improvement for jets and MET in high pileup events
- Two stages
  - Pattern matching
    - FTK tests 1 billion patterns
    - Patterns pre-loaded on associate memory chips
  - Track fit on FPGAs

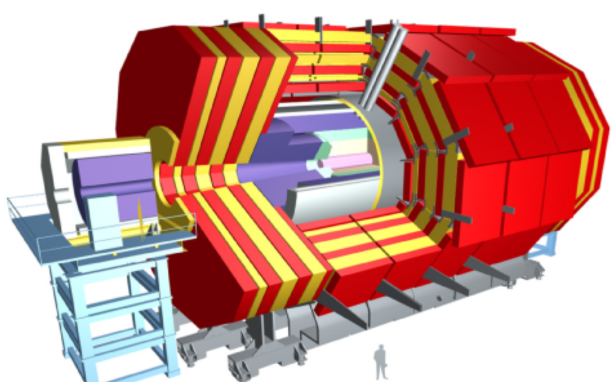






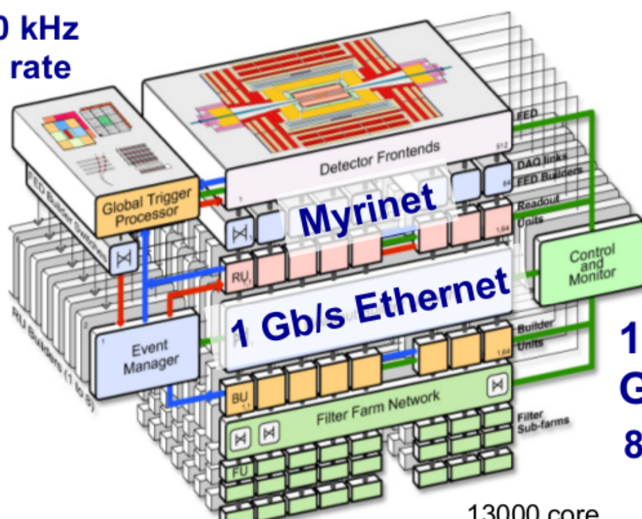
CMS TDAQ system evolution

- Take advantage of current network technologies
  - Remove the need to slice the system



Event size up to 1MB

100 kHz L1 rate



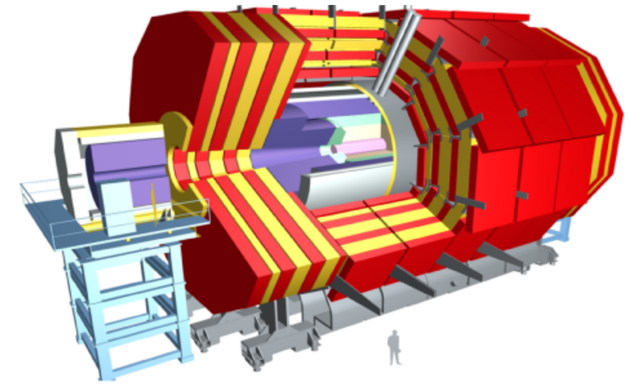
100 GB/s

8 slices

CMS DAQ 1

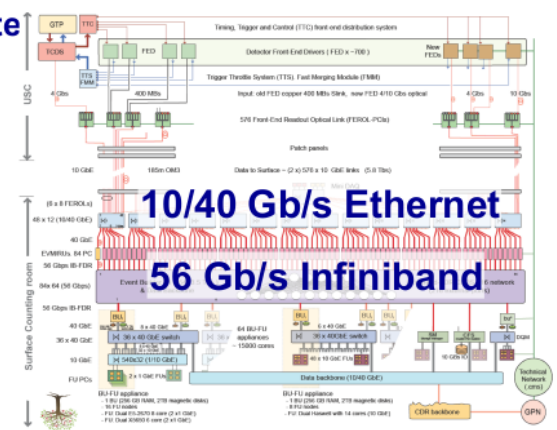
13000 core, 1260 host filter farm

max. 1.2 GB/s to storage



Event size up to 2MB

100 kHz L1 rate



~200 GB/s

1 slice

CMS DAQ 2

16000+ core, 900 host filter farm

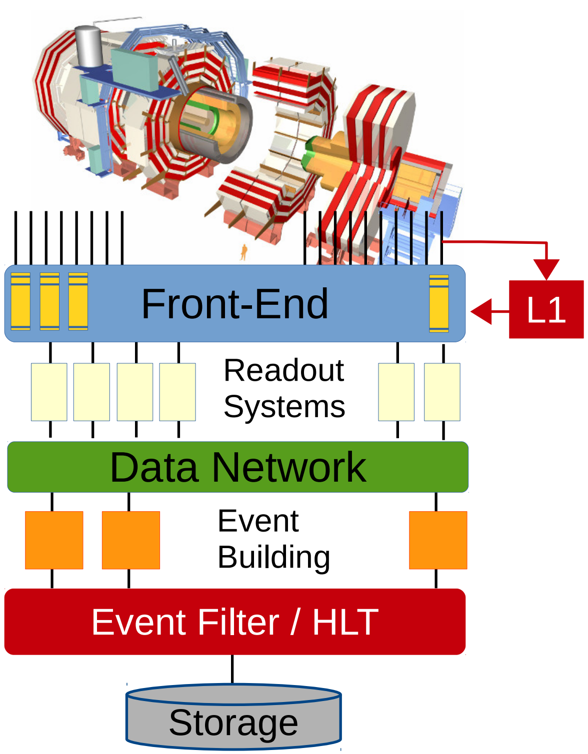
~ 3-6 GB/s to storage

Srecko Morovic, DAQ@LHC 2016

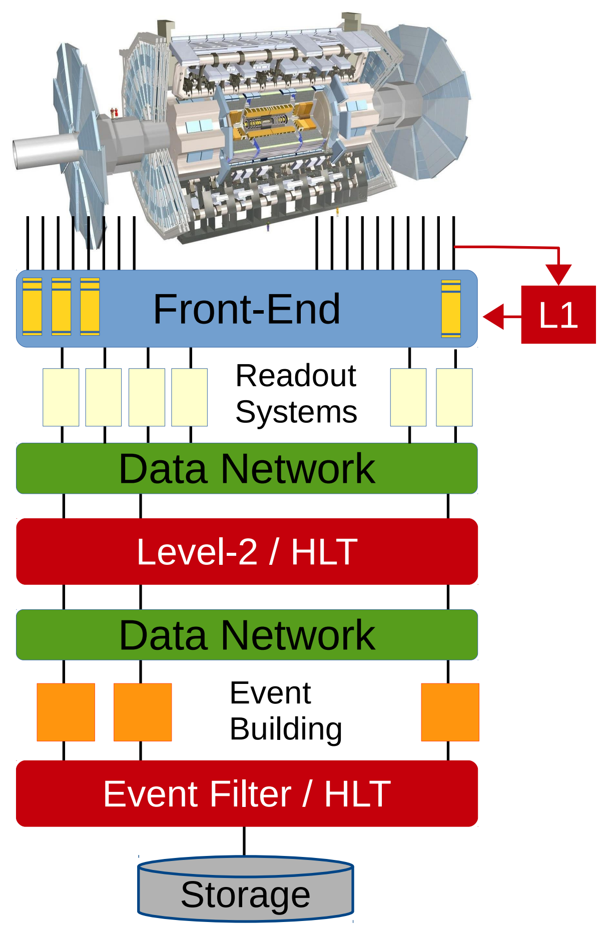
Case-Study: Event Building CMS/ATLAS

- ATLAS/CMS chose different EB strategies in the initial design
  - CMS: “Technology-aggressive” (at the time) full EB
  - ATLAS: Conservative two-level HLT design

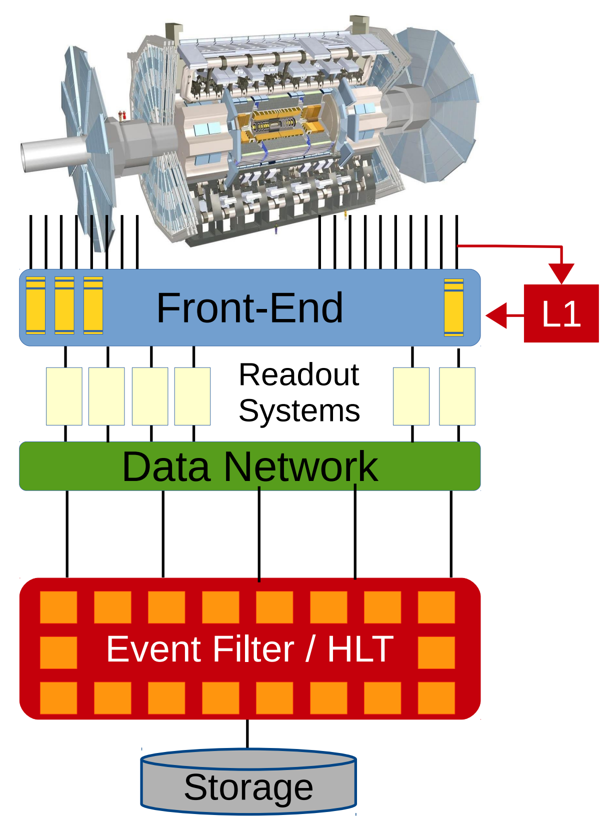
CMS Run-1&2
- 100 kHz EB
- Dedicated EB farm



ATLAS Run-1
- 5 kHz EB after L2
- Dedicated EB farm

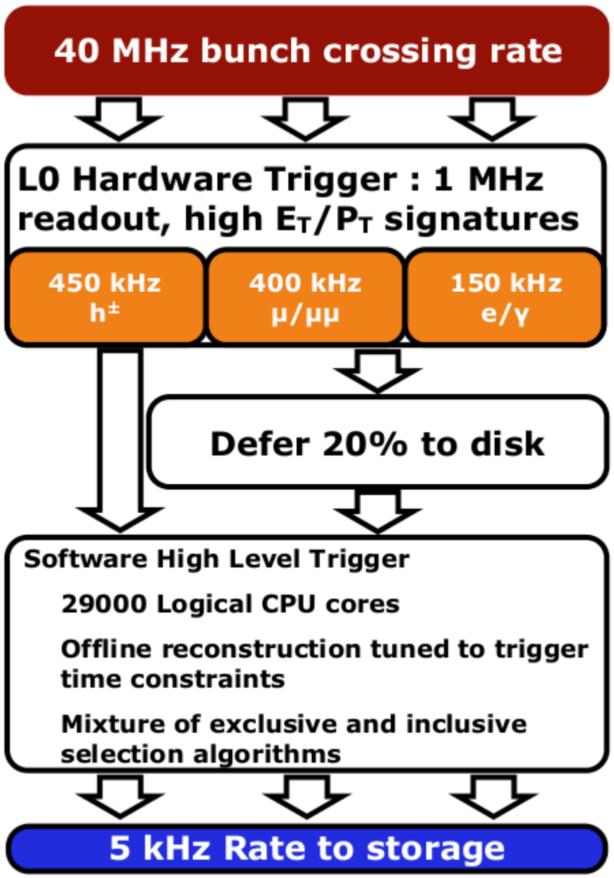


ATLAS Run-2
- incremental EB
- EB on HLT nodes



LHCb TDAQ system evolution

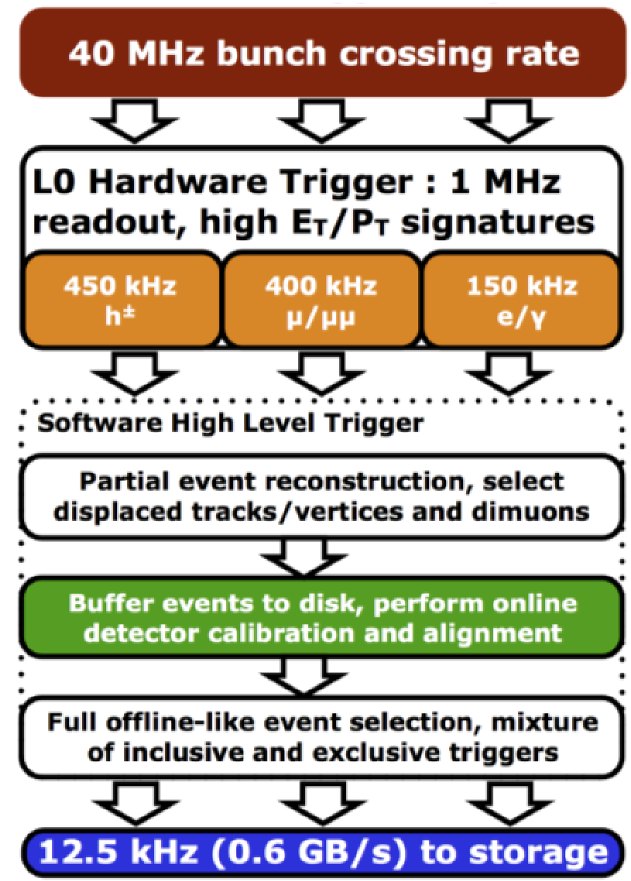
LHCb Run-1



1 HW trigger
1 SW trigger

- HLT farm size doubled
  - 27.000 cores
- Split HLT in two levels
  - Events buffered after HLT1
    - 150 kHz output rate
    - 10 PB or ~13h of buffer
  - Allows time for offline-quality calibration and alignment
  - HLT2 runs offline-like event selection

LHCb Run-2

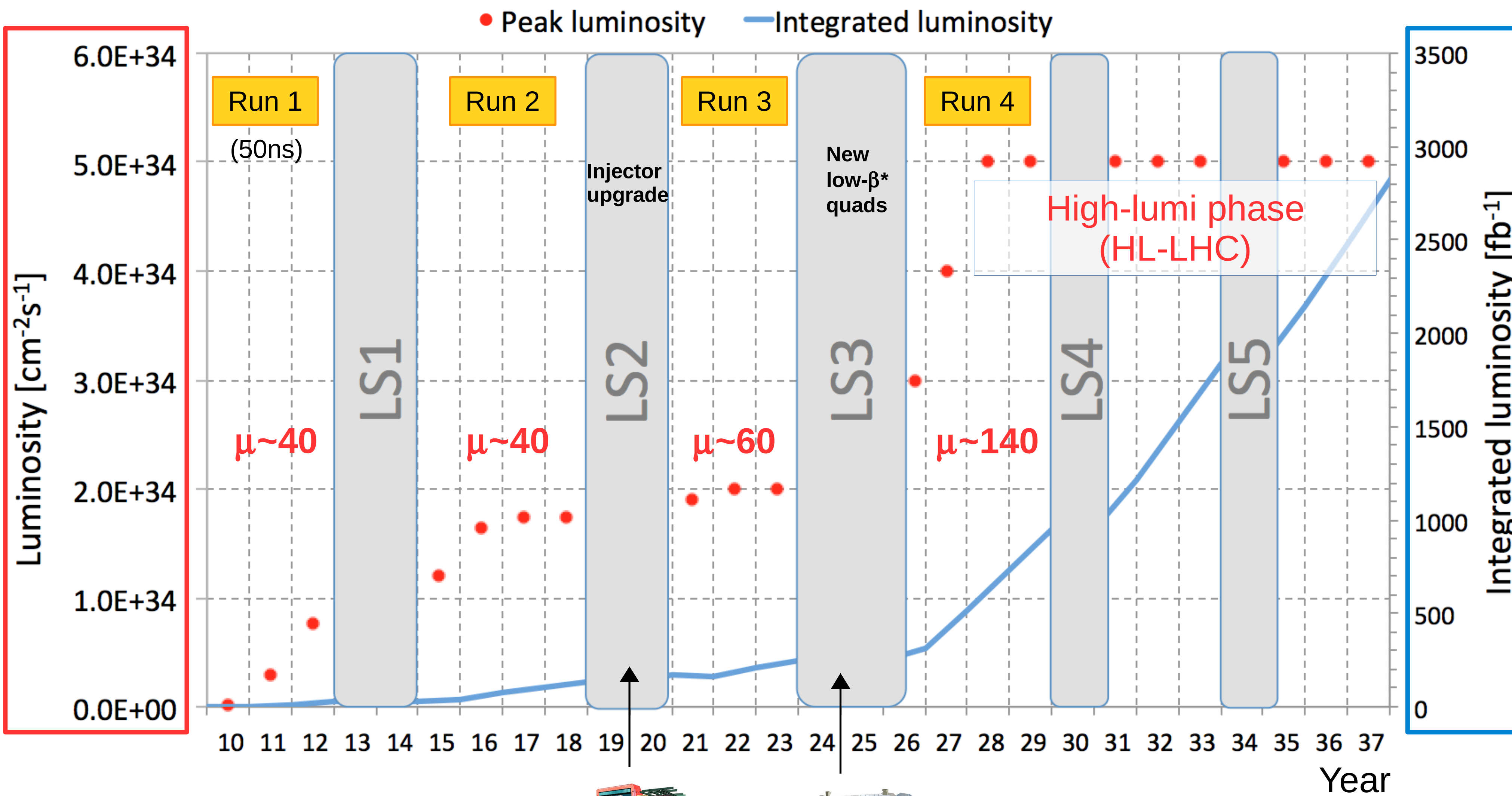


1 HW trigger
2 SW trigger

Roel Aaij, DAQ@LHC 2016

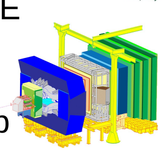
The LHC Experiment Upgrades

LHC Schedule and Performance Projection

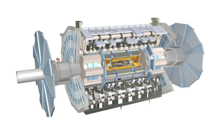


Major upgrades:

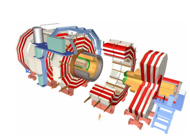
ALICE



LHCb



ATLAS



CMS

Oliver Bruning, ACES 2016

The upgraded LHC experiments (LS2 and LS3)

- ALICE
  - Continuous readout at TPC limit (~50 kHz)
  - Merge of online and offline computing farm
- LHCb
  - No HW trigger → 40(30) MHz to HLT
- ATLAS/CMS
  - Increase HW trigger output rate to ~ 1 MHz
  - Replacement of the majority of FE electronics
  - New inner trackers incl. HW-based track triggers
  - Details of TDAQ systems still very much under discussion

		# Trigger Levels HW	SW	Accept rate		Event size	Event building	Permanent Storage
ALICE (Pb-Pb)	Run-3	0	1	50 kHz		60 MB	† 0.5 TB/s	† 90 GB/s
LHCb	Run-3	0	1	30 MHz	20 kHz	0.1 MB	4 TB/s	2 GB/s
ATLAS	Run-4	1 (or 2)‡	1	0.4(1) MHz	10 kHz	5 MB	2(5) TB/s	50 GB/s
CMS	Run-4	1	1	0.75 MHz	7.5 kHz	5 MB	4 TB/s	40 GB/s

† Alice: event compression (factor~6) and only storing reconstructed objects
‡ Atlas: One or two-level HW trigger under discussion

The upgraded LHC experiments (LS2 and LS3)

- ALICE
  - Continuous readout at TPC limit (~50 kHz)
  - Merge of online and offline computing farm
- LHCb
  - No HW trigger → 40(30) MHz to HLT
- ATLAS/CMS
  - Increase HW trigger output rate to ~ 1 MHz
  - Replacement of the majority of FE electronics
  - New inner trackers incl. HW-based track triggers
  - Details of TDAQ systems still very much under discussion

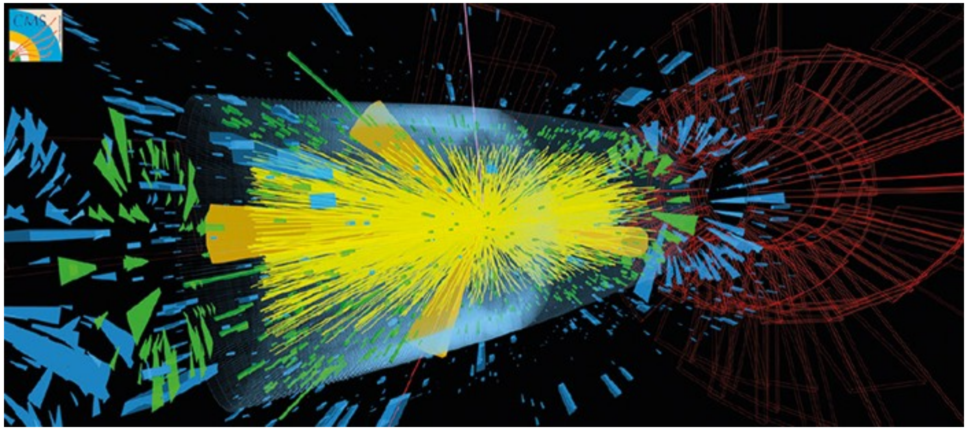
Approximate increase compared to Run-2

		# Trigger Levels HW	SW	Accept rate		Event size	Event building	Permanent Storage
ALICE (Pb-Pb)	Run-3	**0**	1	**x100**		60 MB	† 0.5 TB/s	† **x15**
LHCb	Run-3	**0**	1	**x30**	**x2**	**x2**	**x60**	**x4**
ATLAS	Run-4	1 (or 2)‡	1	0.4(1) MHz **x4-10**	10 kHz **x10**	5 MB **x5**	2(5) TB/s **x50**	50 GB/s **x50**
CMS	Run-4	1	1	0.75 MHz	7.5 kHz	5 MB	4 TB/s	40 GB/s

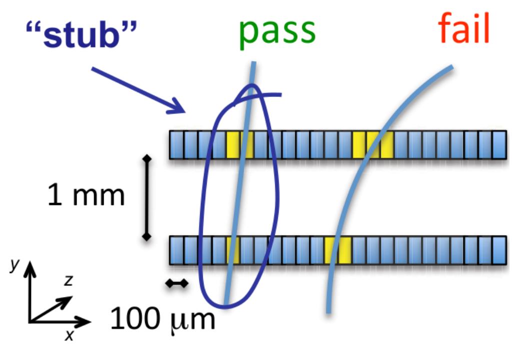
† Alice: event compression (factor~6) and only storing reconstructed objects
‡ Atlas: One or two-level HW trigger under discussion

Track Trigger at the LHC

- Challenges for Track Trigger at LHC
  - Transmitting all data at 40 MHz would require huge amount of electrical power
  - Need to reduce rate via some "pre-trigger"
- Physics-case for track triggers
  - Separate overlapping hadronic interactions
  - Rate reduction due to cluster-track matching
  - Track-based isolation
- CMS: self-seeded
  - Rely on doublet construction of new inner detector
  - Build high-pt track stubs directly on modules
    - Filter all tracks with $p_T > 2$ GeV
  - Time-multiplexed track fitting on FPGAs
- ATLAS: RoI-based
  - Rely on regional information from (Level-0) Calo/Muon trigger
  - Data request to tracker only for these regions (~1MHz)
  - Using similar technology as FTK



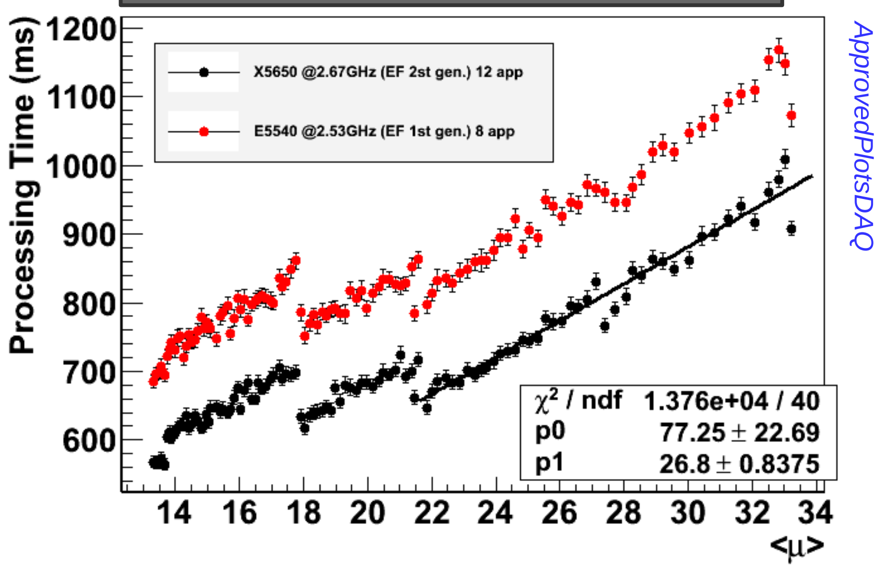
VBF Higgs production on top of 200 pile-up collisions (CMS)



Technology requirements for the Upgrades

- Networking
  - Increased bandwidth requirements by all experiments
- Storage
  - Increase in permanent storage capacity (up to 4 PB/day)
  - Large intermediate buffers planned by all experiments
- CPU
  - No or more sophisticated L1 triggers increase CPU needs at HLT
    - Rejection previously done at HLT is ported to L1 → HLT needs more complex algorithms
  - For ATLAS/CMS, the additional effect of pileup will be a major problem
    - Processing time increases at least linear with pileup (dominated by tracking)

ATLAS EF Processing Time vs Pileup



Much of this is not possible/affordable with current technology

Technology Evolution

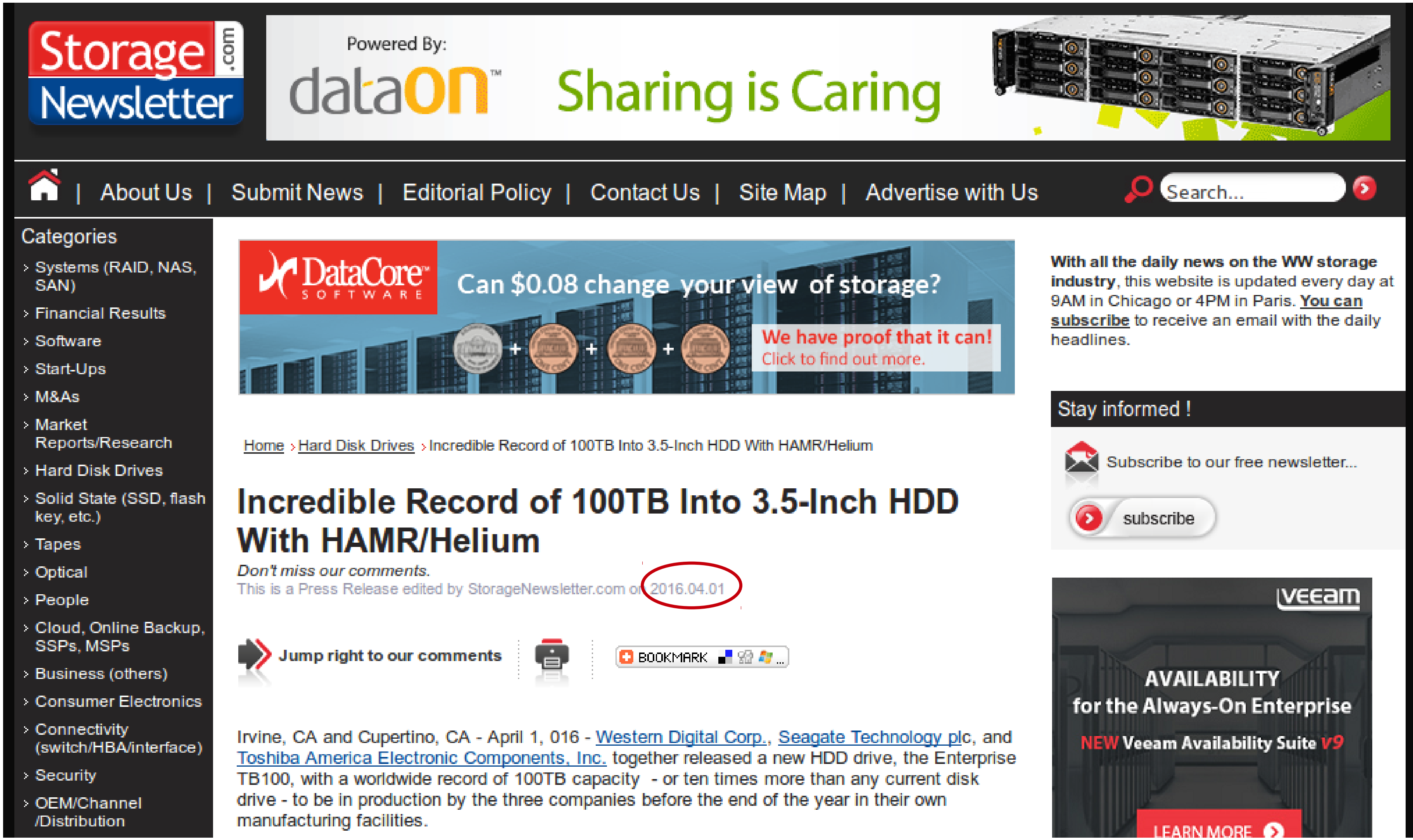
… and how it will enable the experiment upgrades

Where does the Technology go from here?

- HEP is no longer at the fore-front of computing
    - But TDAQ systems still have very specific challenges
- Technology used by HEP driven by
    - Hyper-scale / cloud computing
    - Telecommunications
- Technology develops according to user/market needs
    - Deep learning and AI
        - Massive compute power → GPU
    - Cloud storage
        - Massive storage needs
    - Telecommunication
        - Massive bandwidth needs
- Can we leverage these technologies?

Storage

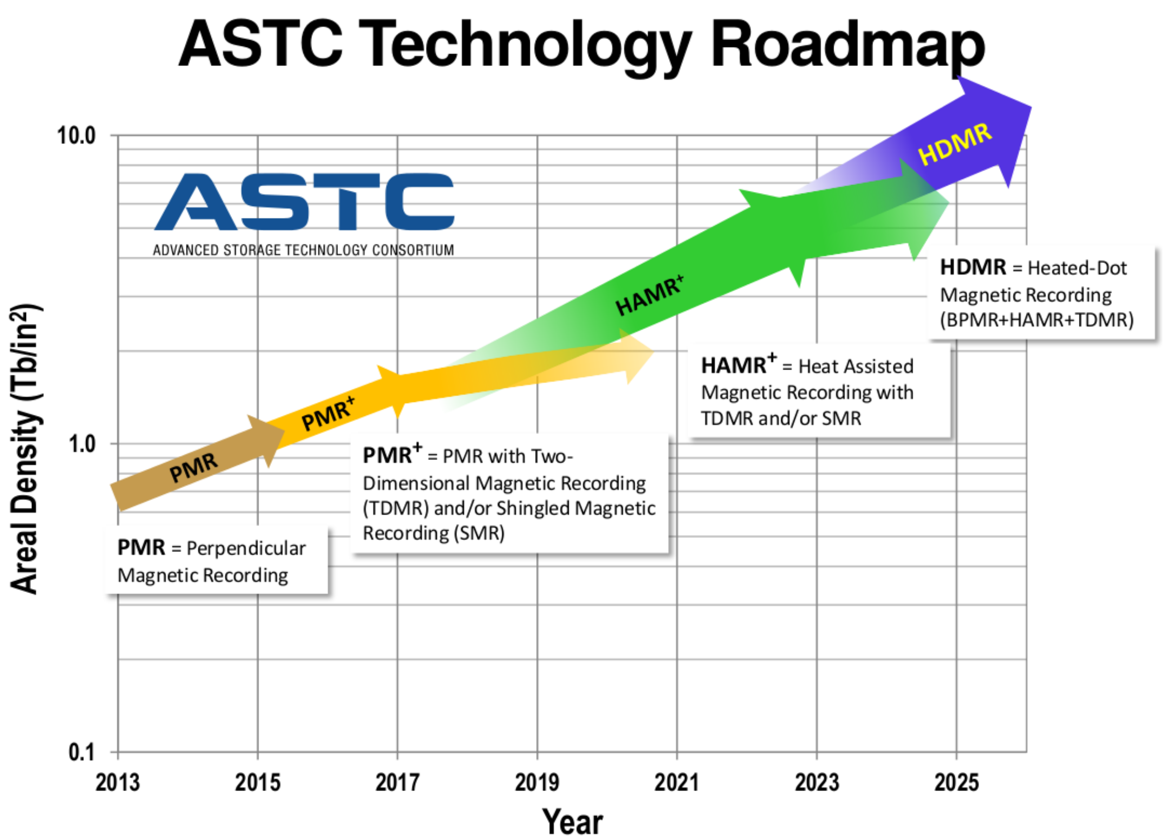
- The future is here...



http://www.storagenewsletter.com/rubriques/hard-disk-drives/incredible-record-of-100tb-into-3-5-inch-hdd-with-hamrhelium/

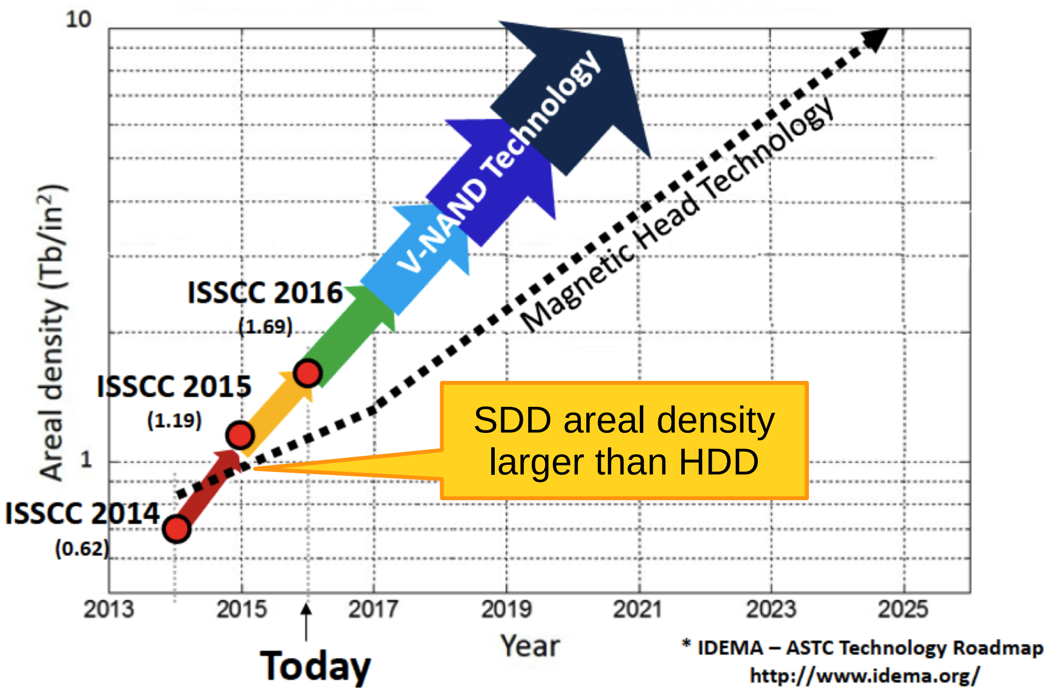
Storage

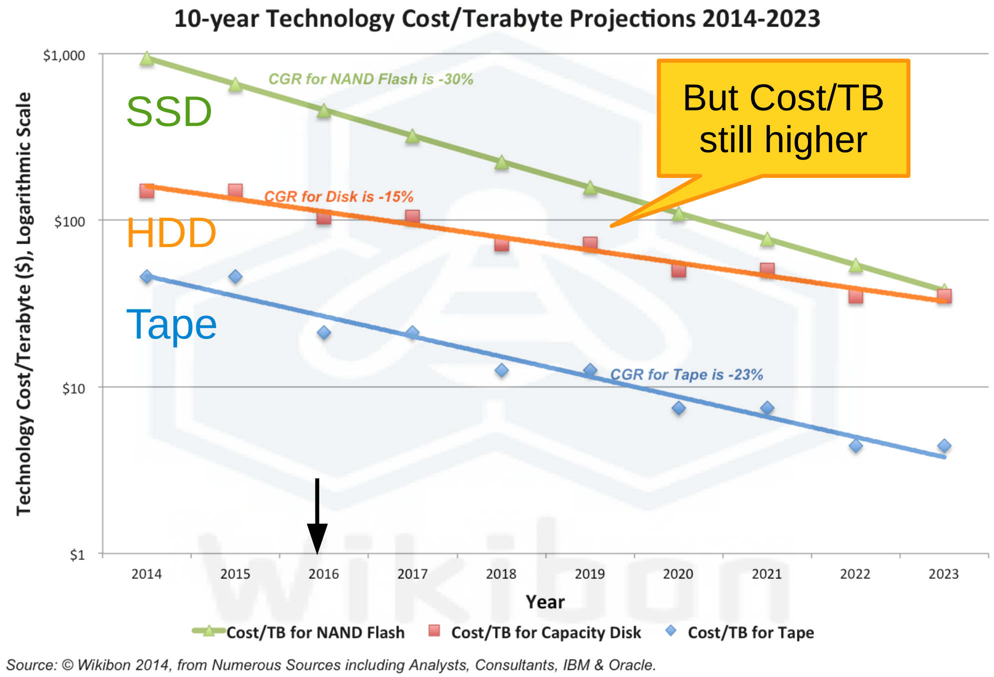
- 100TB HDD will no longer be a joke on April 1st 2025
  - Areal density forecast to increase by factor 10 within next decade
  - Current biggest drive of ~10 TB will be ~100 TB in 10 years
  - Enabled by new recording technologies

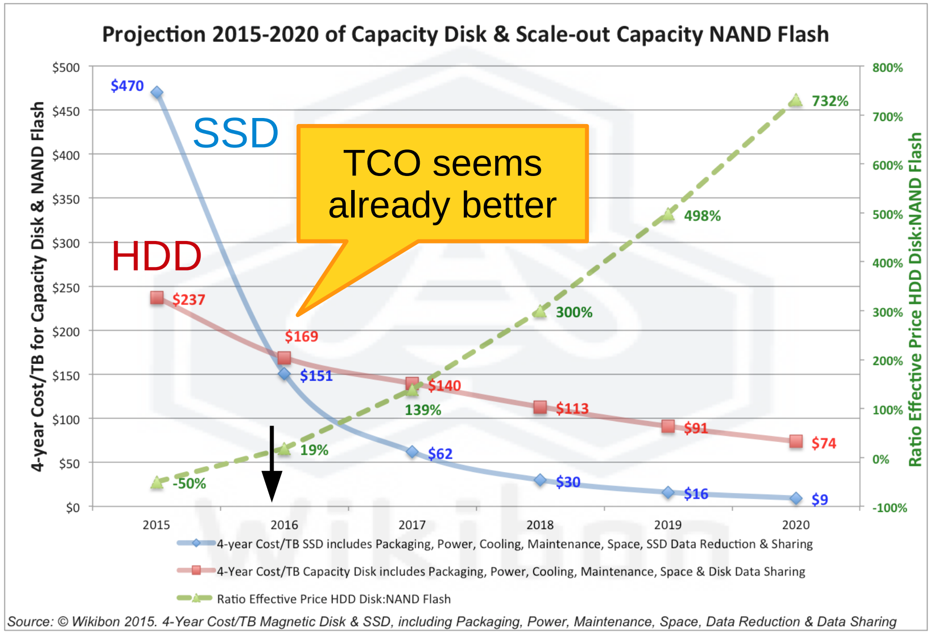


ASTC

What about SSD?



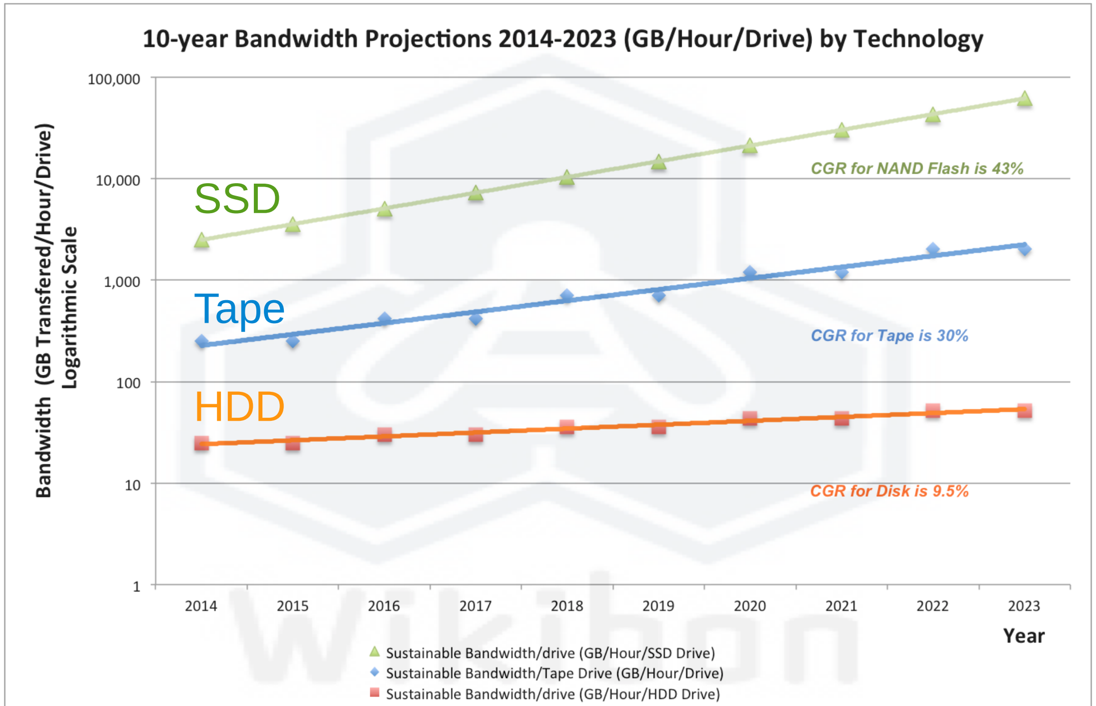




TCO = Total Cost of Ownership
incl. power, space, maintenance, …
(depends on exact assumptions)

OK, so we will have big drives...

- But how to get the data on/off?



Source: © Wikibon 2014, from Numerous Sources including Analysts, Consultants, IBM & Oracle.

Sustainable Bandwidth increase over 10 years :

	Factor
SSD	~35
Tape	~14
HDD	~2.5

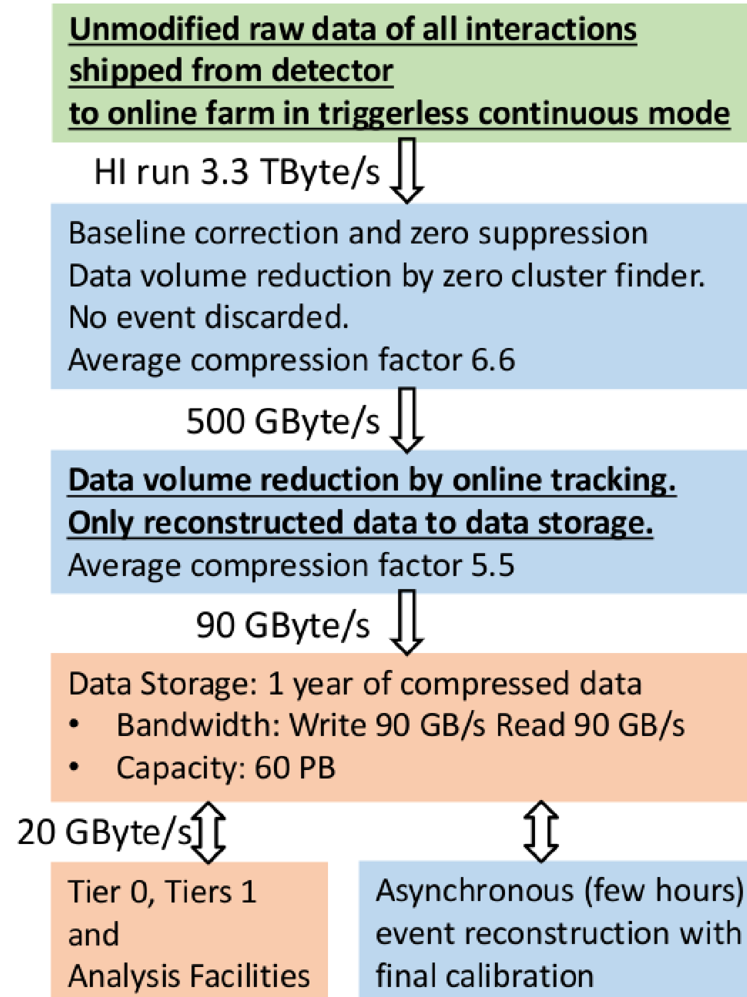
Large uncertainties

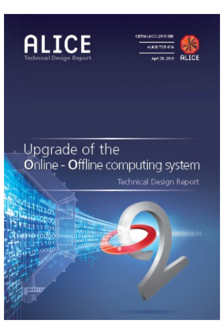
Rotation speed ~fixed. Increase only from linear density increase.

- Take-away points
  - Both SSD and HDD will provide >100 TB capacities/drive in 10 years
  - HDD read/write bandwidth will not keep up with capacity increase (factor 4-5 gap)
  - SSD read/write bandwidth likely to scale with capacity
    - ~30 GB/s by 2025 if above evolution holds
    - But how to attach this drive?

ALICE O²: Online – Offline Computing System

- Physics goals of ALICE upgrade (Run-3)
  - Rare processes
  - Very small signal over background ratio
  - Needs large statistics of reconstructed events
  - Triggering techniques very inefficient if not impossible
- New computing system
  - Read-out the data of all interactions (50 kHz)
    - Limited by drift time in TPC
  - Compress these data intelligently by online reconstruction
  - One common online-offline computing system: O²
    - Resources naturally shared between online/offline and Grid
- Storage needs
  - 60 PB/year with 90 GB/s read/write
    - already possible today



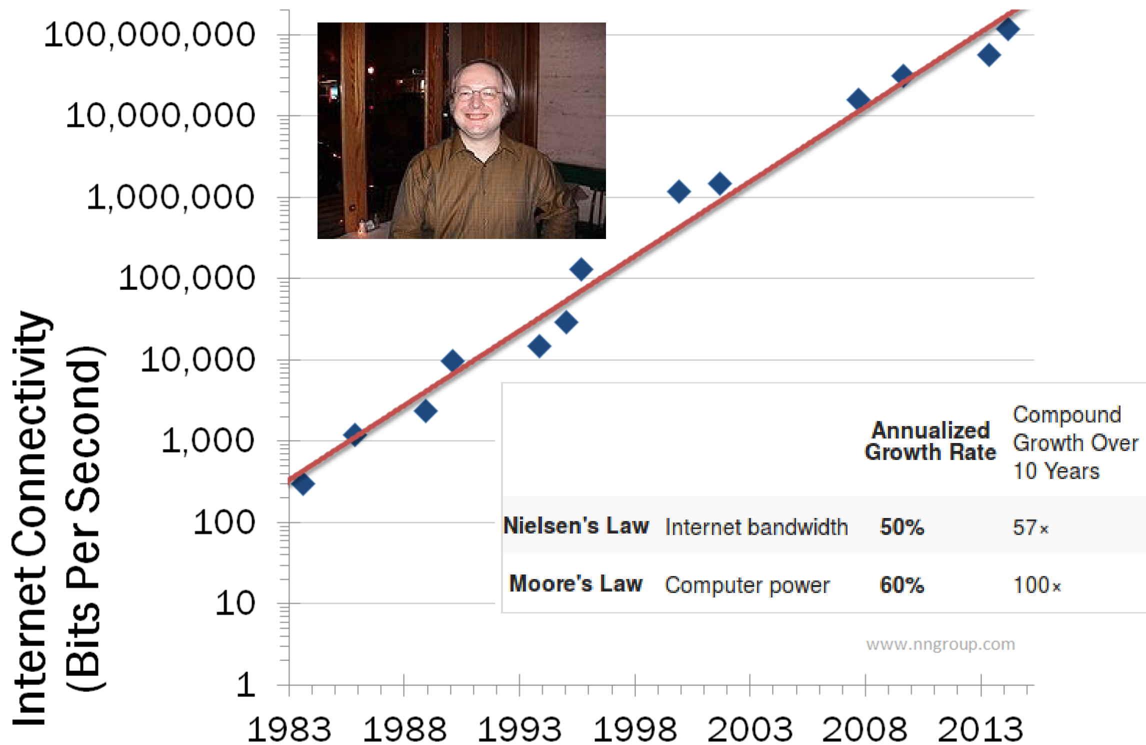


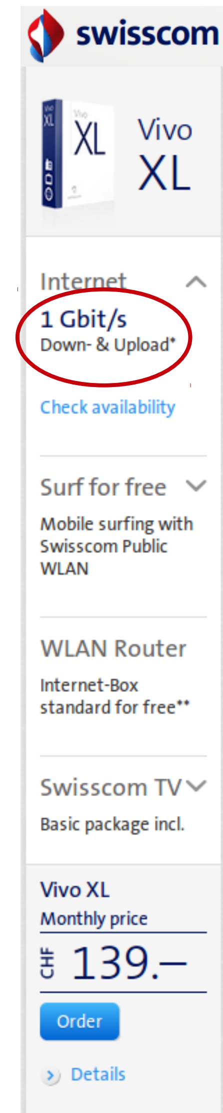
ALICE-TDR-019

Pierre Vande Vyvre, DAQ@LHC 2016

Network technology evolution

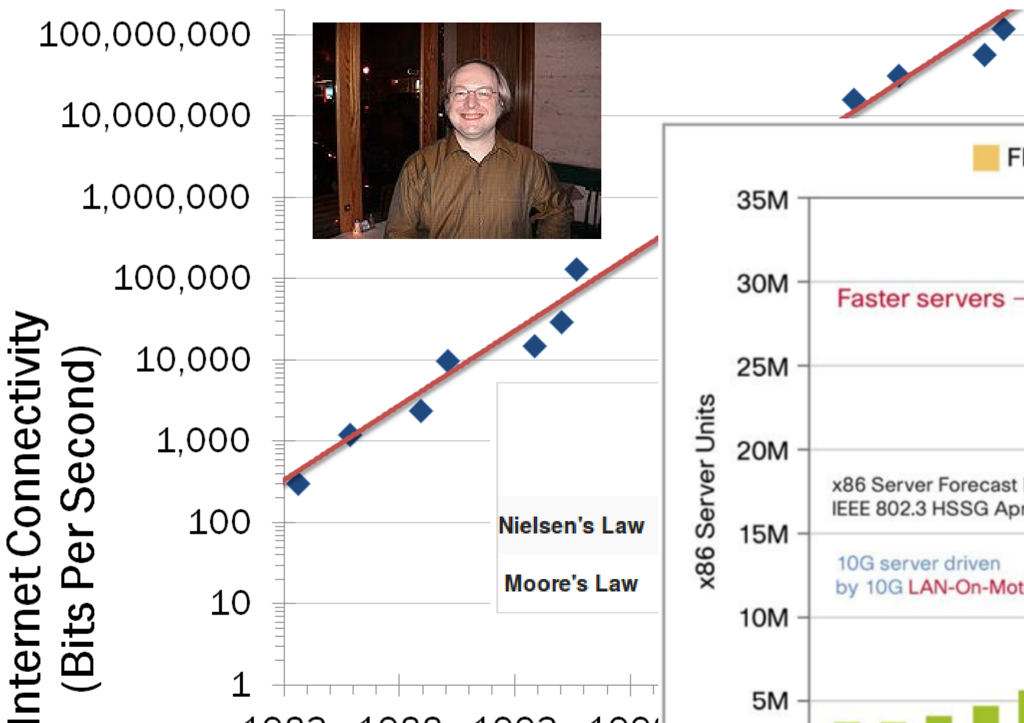
- Nielsen's Law
  - Users' bandwidth grows by 50% per year
    - Compared to 60% p.a. for Moores's Law (factor ~2 gap in 10 years)
  - Telecom backbone switches will need to keep up
    - That is the technology of interest to us!

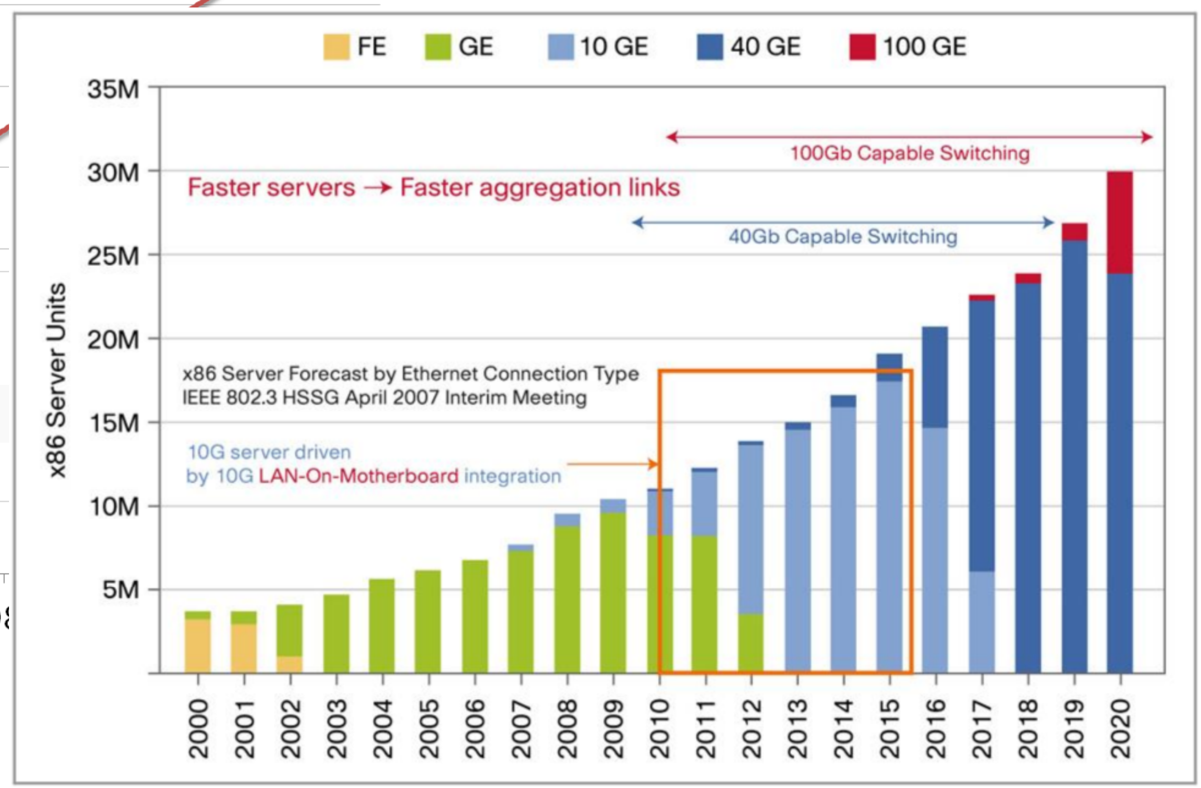




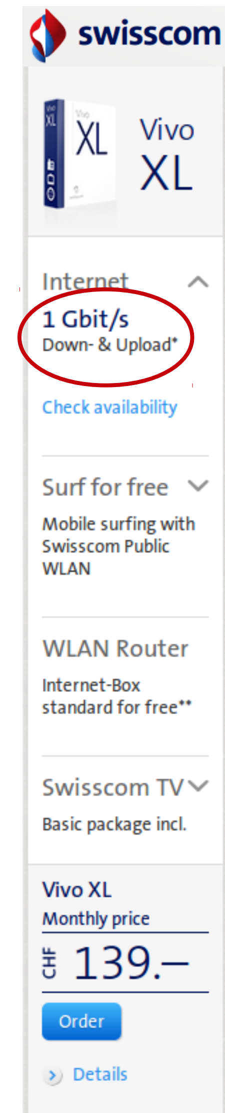
Network technology evolution

- Nielsen's Law
  - Users' bandwidth grows by 50% per year
    - Compared to 60% p.a. for Moores's Law (factor ~2 gap in 10 years)
  - Telecom backbone switches will need to keep up
    - That is the technology of interest to us!



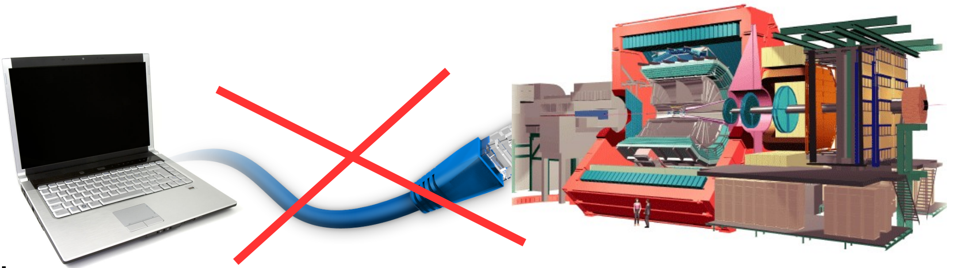


The Market Need for 40 Gigabit Ethernet, Cisco (2014)



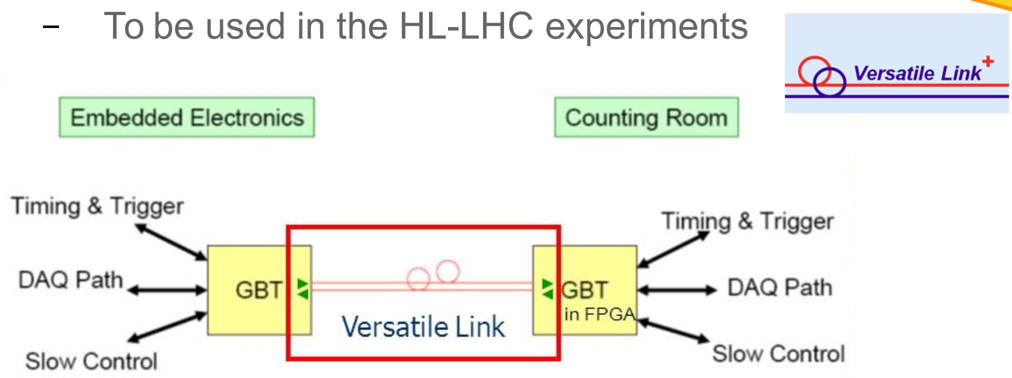
Ethernet on the Detector?

- The "first mile" problem
  - Detector front-ends are in radiation environment
  - Requires radiation-hard links
    - Commercially practically non-existent
  - Typical COTS components withstand O(0.01) Mrad
  - Detector front-end links require O(100) Mrad!
    - … and link length of at least 100m (to surface)



- GBT / Versatile Link project at CERN
  - 4.8 Gb/s optical link physical layer for use in upgraded LHC detectors
    - Currently used custom links support ~1.3 Gb/s (DDL, SLINK, Glink)
  - Versatile Link PLUS aiming at 5-10 Gb/s
    - To be used in the HL-LHC experiments

10-40 times less
than non-radhard
commercial technology

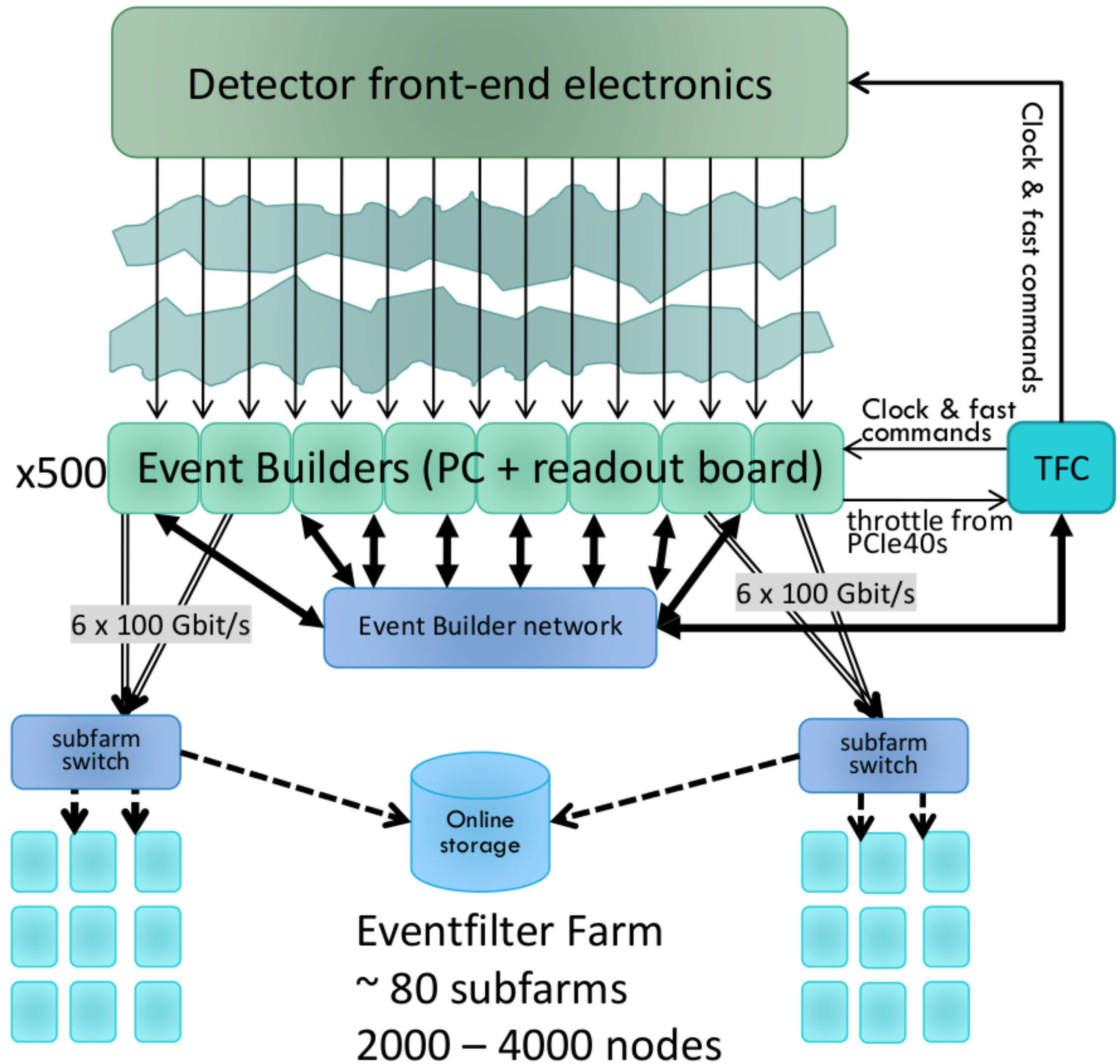


LHCb networking for Run-3 (2020)

- “The highest-throughput DAQ system”
  - Triggerless readout at 40 MHz
  - 32 Tb/s aggregate bandwidth

~10000 custom links
- detector → surface (350m)
- up to 4.8 Gb/s

Event builder network
- Leverage 100Gb Ethernet



- R&D for ATLAS/CMS
  - Similar event building requirements for Run-4/2025 (~32 Tb/s)
    - Should be “trivial” by then

Niko Neufeld, DAQ@LHC 2016

Triggerless DAQ for ATLAS/CMS ? (this is not planned !)

- Assumptions for triggerless DAQ in 10 years
  - Assume 10 Gb (rad-hard) GBT link and 400 Gb Ethernet
  - 100 TB SSD drives with very optimistic 10GB/s/drive
  - Buffer data for N minutes until HLT has processed events

40 MHz * 5 MB/event
= 200 TB/s = 1600 Tb/s

>100 times #links as of today

Impossible!

160.000 x 10 Gb/s links (VL)

10 times more links than LHCb is planning for Run-3

Maybe feasible

4000 x 400 Gb/s Ethernet

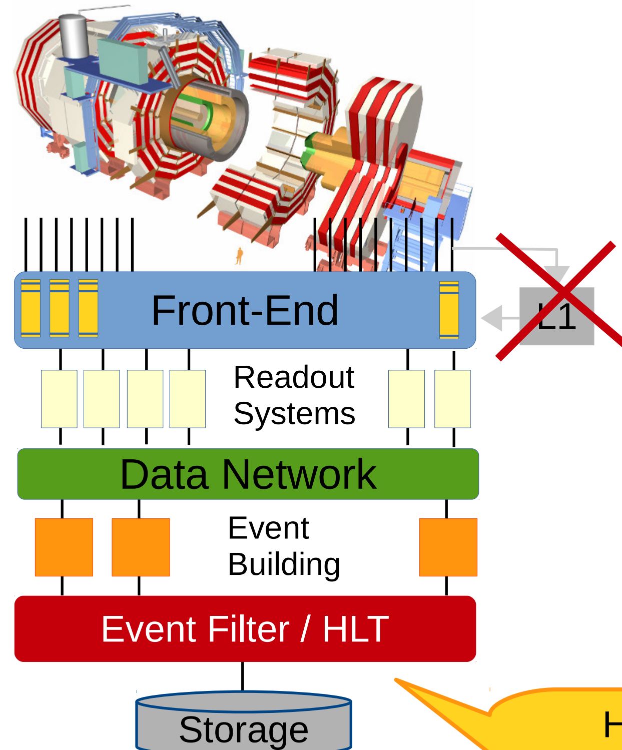
12 PB/min buffer

120 drives per minute of buffer...

No problem

…but to achieve the throughput with 10GB/s/drive → 20.000 drives

Huge discrepancy of storage vs bandwidth!

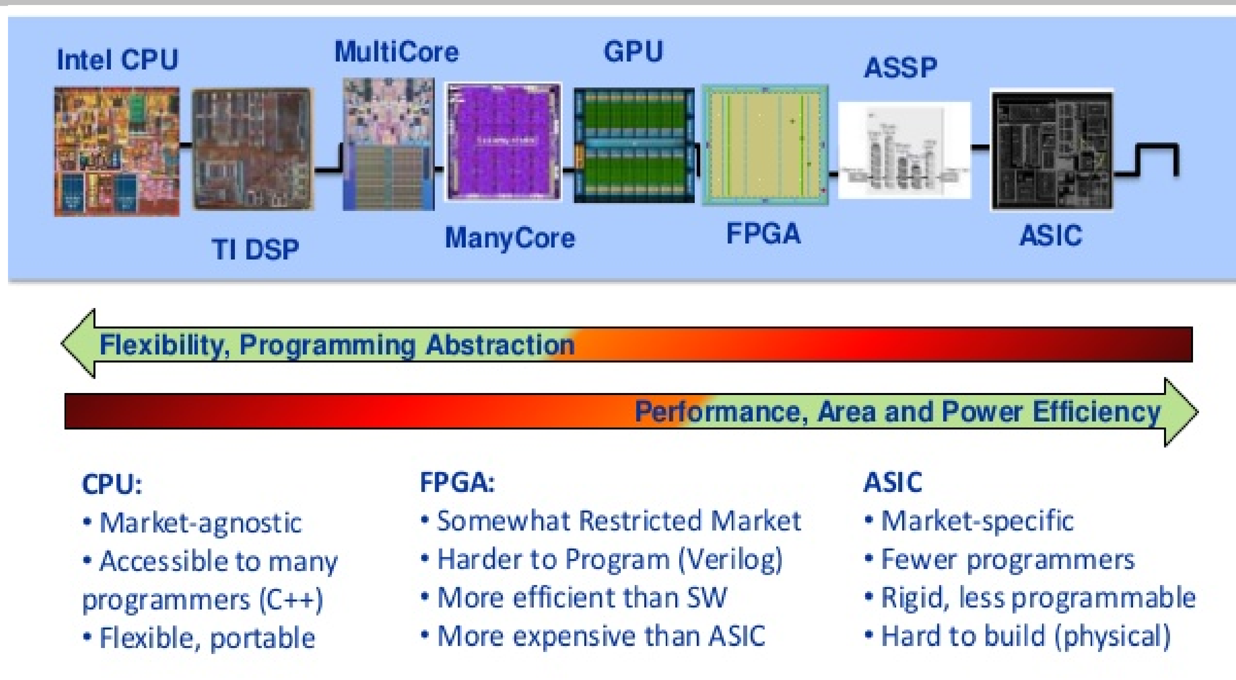


HLT CPU:
400(rate)*3(pileup)
~1000x today

From Moore ~100x

Maybe with GPU/FPGA

Programmable devices: ASIC, FPGA, GPU, CPU



ALTERA.

http://www.slideshare.net/TomSpyrou/tau-2015-spyrou-fpga-timing-45995712

- Typical use-cases in LHC experiments
  - Driven by latency constraints
  - Radiation hardness is an important factor for detector front-ends
  - Where possible FPGAs replace ASICs

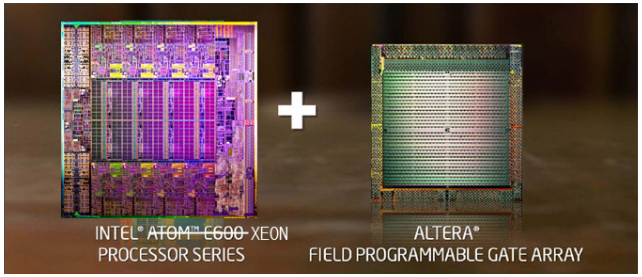
	ASIC	FPGA	GPU	CPU
Front-end	✔	(✔)		
L1	✔	✔	?	
HLT			(✔)	✔
Offline/Grid		$	$	✔

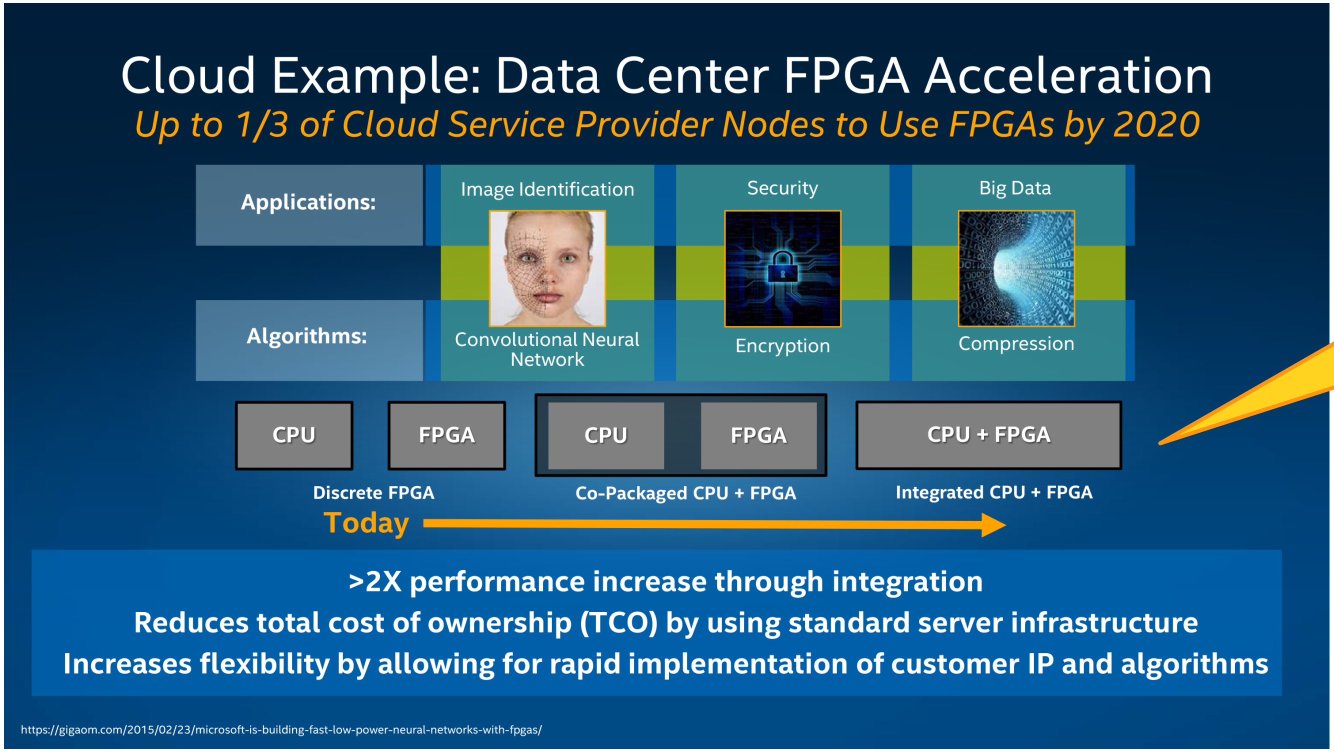
✔ standard
(✔) increasingly
? near future
$ market driven

FPGA becoming Mainstream?

- 2010: Intel announces Atom + FPGA processor
  - Mainly for embedded market (in competition with ARM)
- 2015: Intel acquires Altera
  - Altera is the second largeste FPGA vendor (after Xilinx)
  - Intel announces Xeon processor with FPGA
  - Speedup of specific algorithms in data centers (Google, Facebook, etc.)

now part of Intel

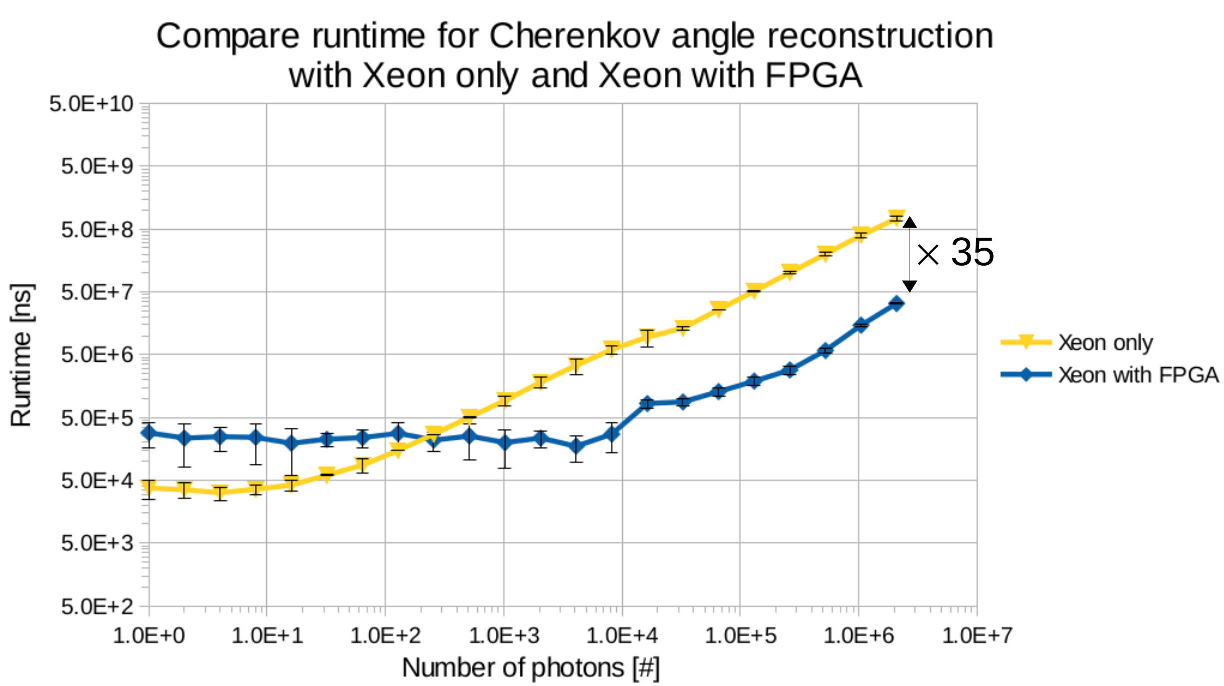




Can we make use of these FPGAs ?

Study: FPGA in LHCb

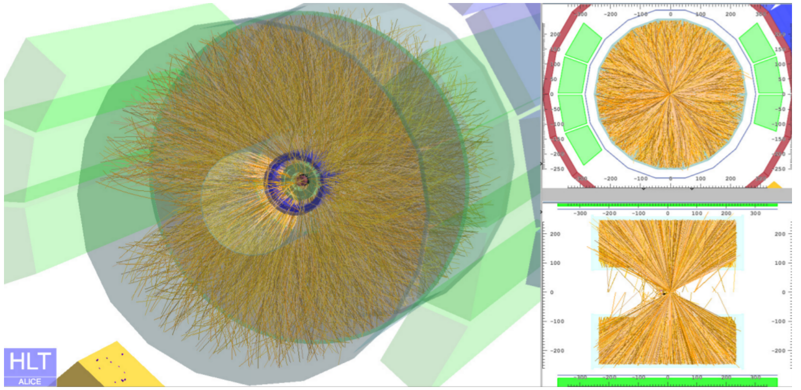
- High Throughput Computing Collaboration (HTCC)
  - Members from Intel, CERN OpenLab and LHCb
- Particle identification in LHCb
  - Calculate Cherenkov angle in RICH detector
  - Currently cannot be done for every event as too CPU expensive
- Xeon + FPGA
  - Acceleration of factor up to 35 with Xeon/FPGA
  - Theoretical limit of photon pipeline: factor 64
  - Bottleneck: Data transfer bandwidth to FPGA



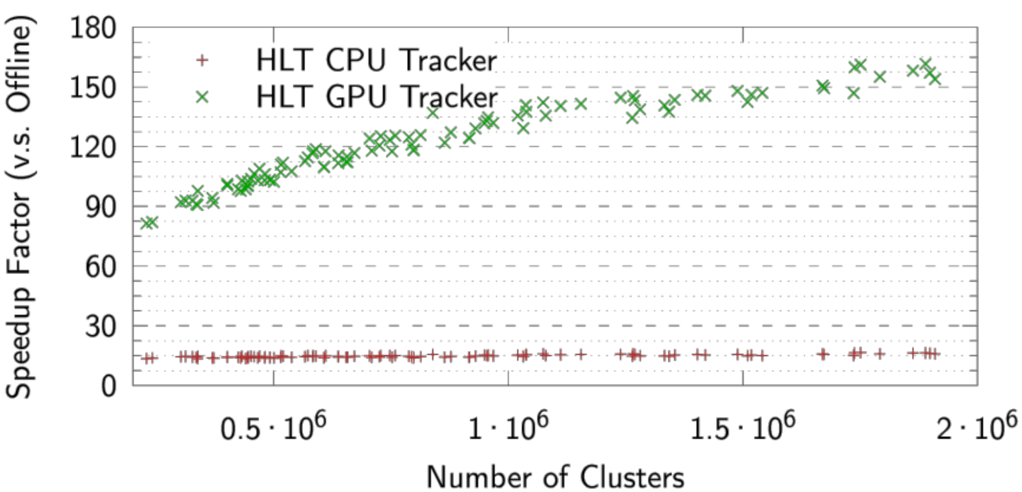
Christian Faerber, DAQ@LHC 2016

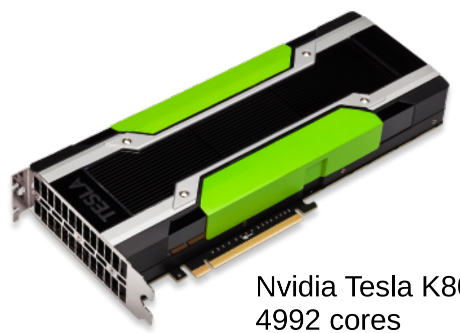
GPU usage in LHC experiments

- GPUs currently only used in ALICE HLT
  - CPU time dominate entirely by (TPC) tracking
  - Ideal for offloading to GPU
  - Other experiments have more heterogenic compute loads



- ALICE HLT
  - 180 compute nodes with GPUs (AMD FirePro S9000)
  - Factor 20 speedup compared to HLT CPU tracking
  - Cost saving of several 100k CHF for online farm



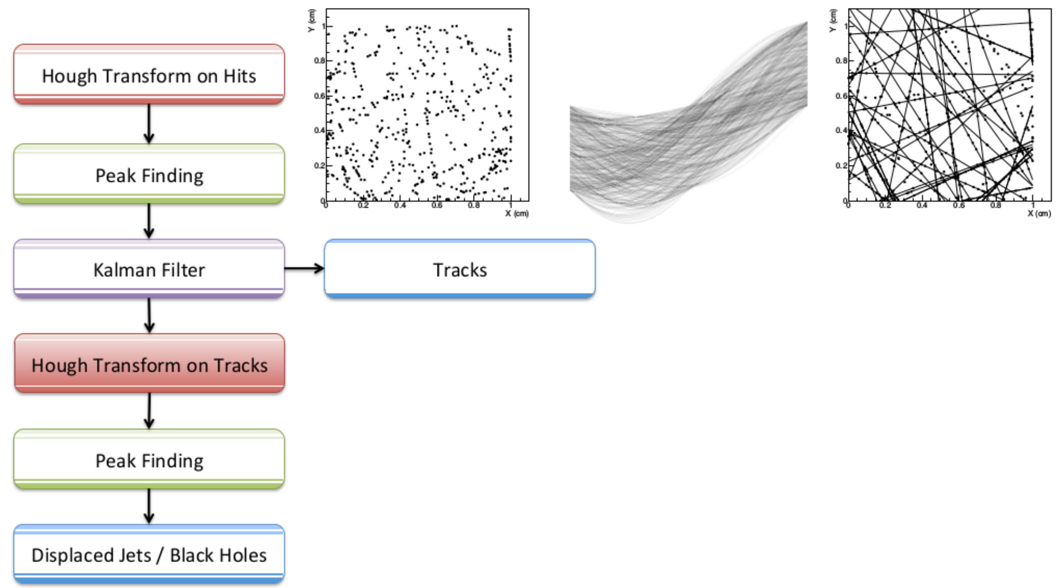


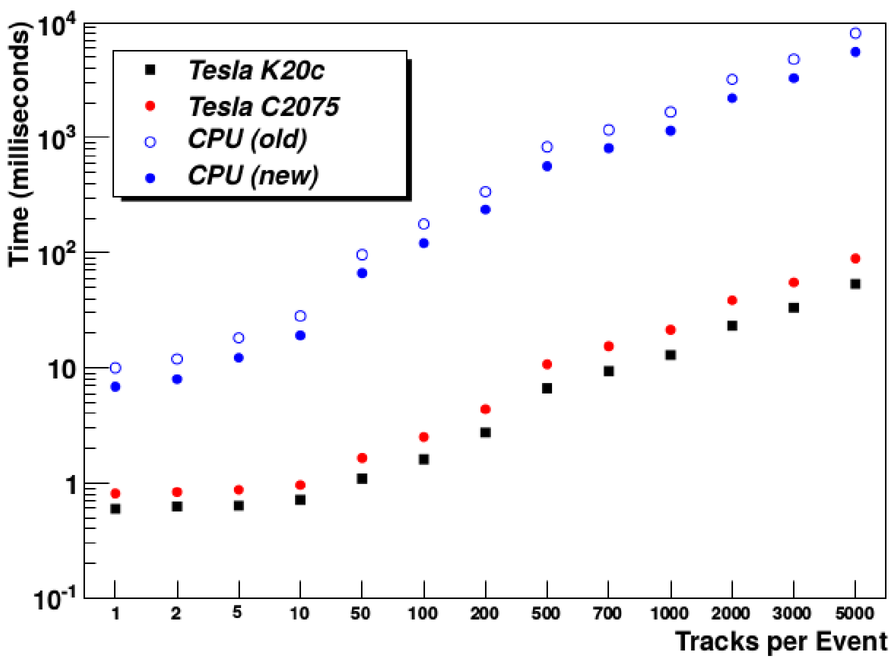
Nvidia Tesla K80
4992 cores
24 GB memory

- Other experiments in evaluation stage
  - LHCb study for use in VELO tracking
  - ATLAS/CMS study for use in tracking

Study of GPU use in CMS

- Tracking based on Hough transforms
  - On a CPU typically slower than traditional pattern recognition
  - But can be efficiently parallelized
  - Factor 10-60 speedups measured



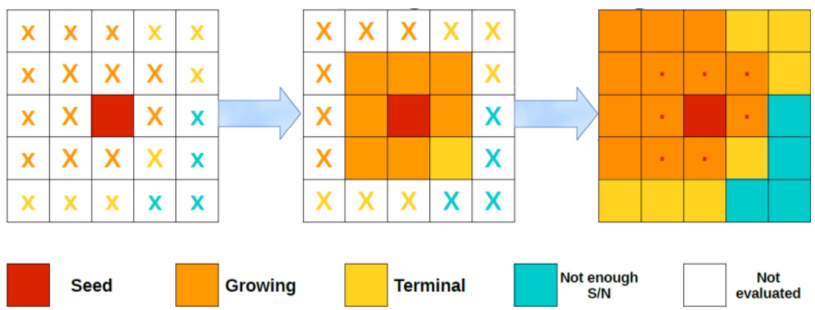


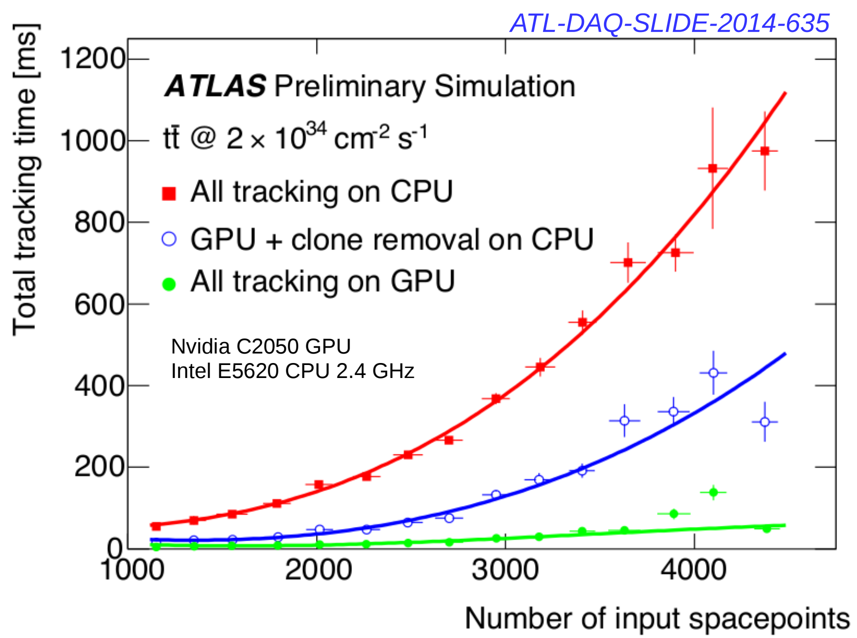
CPU implementation before and after optimization (on Core i7-3770) vs GPU

arXiv:1309.6275

Study of GPU use in ATLAS

- GPU for tracking
  - Speedup of factor 12 for the whole tracking chain
    - GPU vs one CPU core
- GPU for calorimeter clustering
  - Use GPUs to "grow" clusters from cells
  - Algorithm implemented, performance being evaluated





- Extended prototype being worked on
  - Inner Detector Tracking
  - Calorimeter Clustering
  - Muon tracking based on Hough transforms

Common challenges of FPGA and GPU use

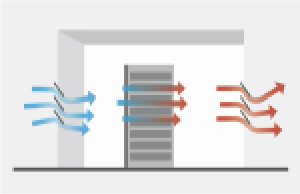
- Very useful for specific algorithms
  - Ideal if reconstruction time is dominated by a single algorithm (e.g. ALICE tracking)
- Most useful for online applications
  - Dedicated compute farms
  - Under full control of the experiments
- Sharing of FPGA/GPU
  - Typical compute node has several CPU cores
  - How to efficiently share the single FPGA/GPU between the cores?
  - Currently I/O from CPU to FPGA/GPU are limiting factors
- Cost/benefit calculation not always easy
  - HLT farms typically heterogenic
  - Manpower "cost" for developing new algorithms

Where to put all this new Hardware?

- Experiment upgrades will require additional computing
  - LHCb and ALICE require 2000-4000 compute nodes with ~2MW power consumption
  - Existing buildings are not sufficient
    - Cooling and power
    - Rack space
    - Weight limits on floors
  - New buildings expensive and long lead times
    - Also not very flexible

Current CMS datacenter at Point-5

- Container-ized data-centers
  - Deployable within 3-6 months
  - Minimal site requirements
  - Scalable
  - Re-usable (at different location)
  - Typically fit 20 racks/container
  - Can be cooled mostly by airflow in Geneva
    - Inlet temperature <35^{o} most of the year



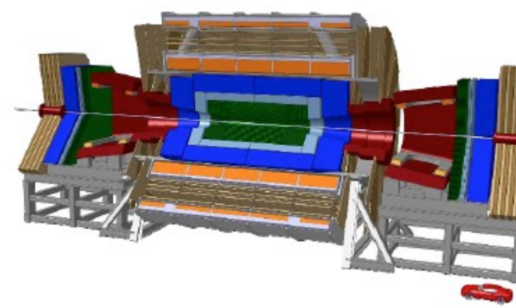
Next datacenter ?

Heinrich Schindler, DAQ@LHC 2016

Trigger/DAQ for future experiments?

- Storage and networking evolution change the Trigger/DAQ landscape
  - Many experiments are or will be moving to triggerless systems
    - e.g. ALICE/LHCb in Run-3
  - HLT/Event Filter (if needed) becomes highly asynchronous
    - enabled by large (many hours) event buffers
    - no longer real-time → more similar to offline batch systems

- Trigger/DAQ at FCC-hh (~2035)
  - Pileup of 850 (170) with 25 (5) ns bunch spacing
  - Detector: Scaled up version of CMS
    - Estimated 10 times[1] higher RAW data rates → 2000 TB/s
  - Triggerless design very unlikely
    - Multiply our "triggerless ATLAS/CMS" from earlier by factor 10
    - Main challenge is again the number of rad-hard FE links
  - Large on-detector buffers could significantly increase the L1 trigger latency
    - Sequential (pipelined) readout
    - First level trigger could be implemented in "fast" software (FPGA/GPU)
    - Then follow ALICE/LHCb model of large off-detector buffers and run offline reconstruction



[1] Dave Newborn, *Future Trigger and DAQ developments*, *FCC Week 2016*

Summary

- Trigger/DAQ at hadrons colliders remains challenging
  - Technology evolution pushes the boundaries of Physics we can do with these detectors
- Move to commercial hardware where possible
  - Network and storage evolution allows for some triggerless systems
  - Reduces cost and effort for specialized L1 trigger hardware
    - But also shifts R&D away from HW to SW (learning process also for funding agencies)
  - Need to adapt to whatever the market gives us
    - Many-core CPUs, GPU, FPGA
  - May need drastically new approaches to benefit from it
- LHC experiment upgrades underway
  - ALICE/LHCb will push the throughput to new levels in Run-3
  - ATLAS/CMS will need complex L1/HLT triggers for HL-LHC in Run-4 and beyond
  - This is crucial R&D for the next generation experiments

References

- ACES 2016
  - https://indico.cern.ch/event/468486
- DAQ@LHC 2016
  - https://indico.cern.ch/event/471309
- CPAD Meetings
  - 2013: https://indico.fnal.gov/conferenceTimeTable.py?confId=6050#20130109
  - 2015: https://indico.hep.anl.gov/indico/conferenceDisplay.py?confId=625