Electronics, Trigger and Data Acquisition

part 3

Summer Student Programme 2016, CERN

July 13, 2016

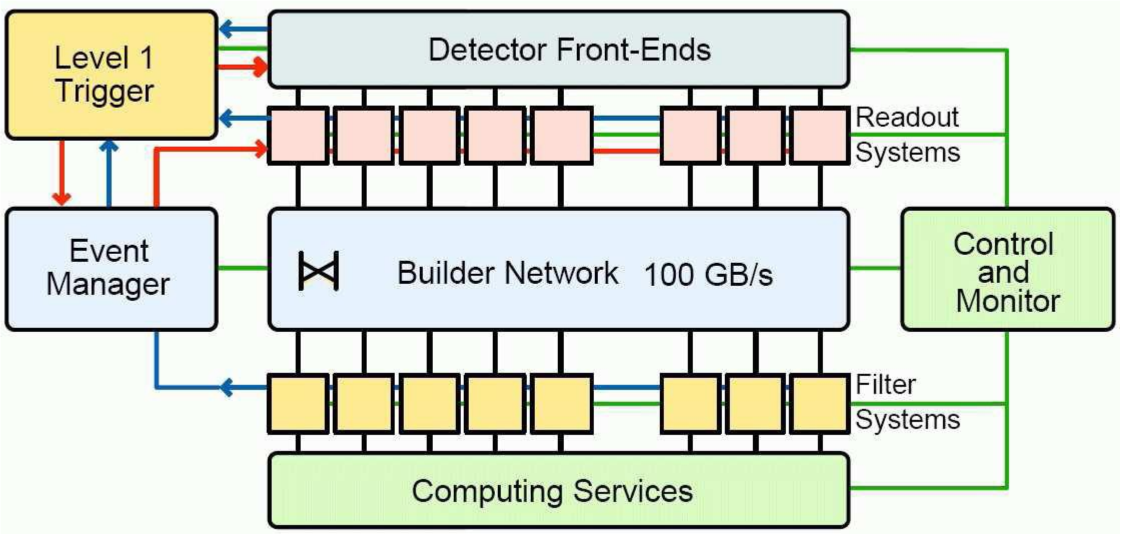
Roberto Ferrari
Instituto Nazionale di Fisica Nucleare

roberto.ferrari@pv.infn.it

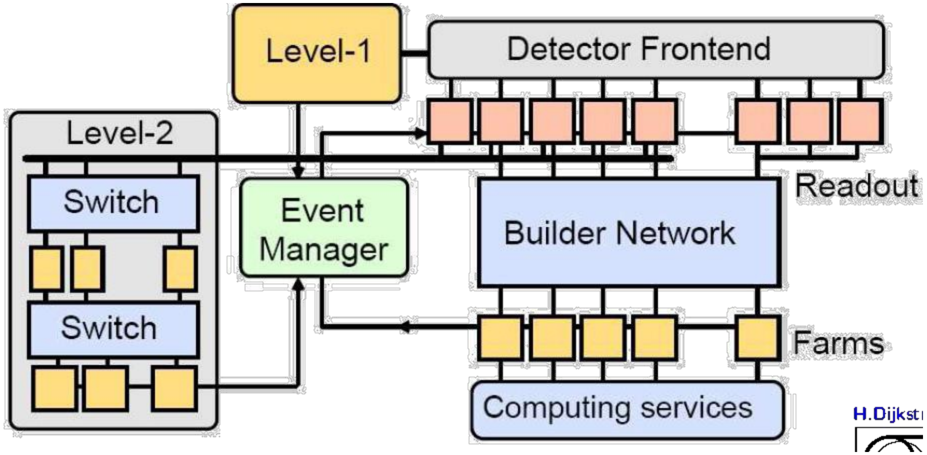
Event Building

Two Philosophies @ LHC

- Send everything, ask questions later (ALICE, CMS, LHCb)

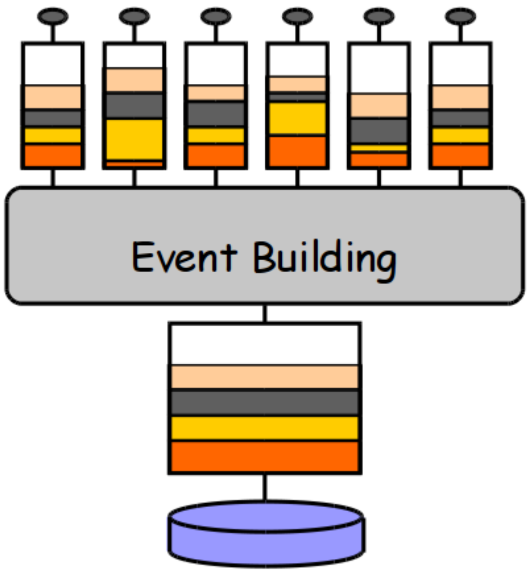


- Send a part first, get better question → send everything only if interesting (ATLAS)

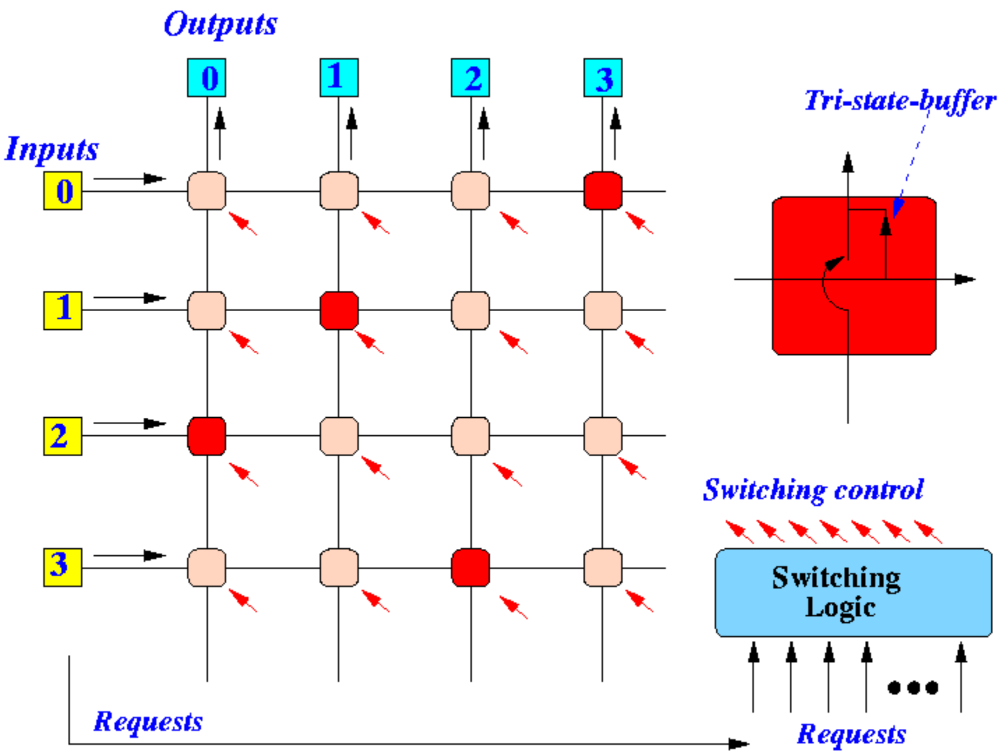


Event Building: Network Perspective

- Event Building: collection and formatting of all data elements of an event into a single unit
  - normally last step before high-level trigger or storage
  - can be implemented on buses, can use custom interconnects, can be based on (Ethernet) **network**
- Network-based EB is choice of all LHC experiments and case study for networking in DAQ

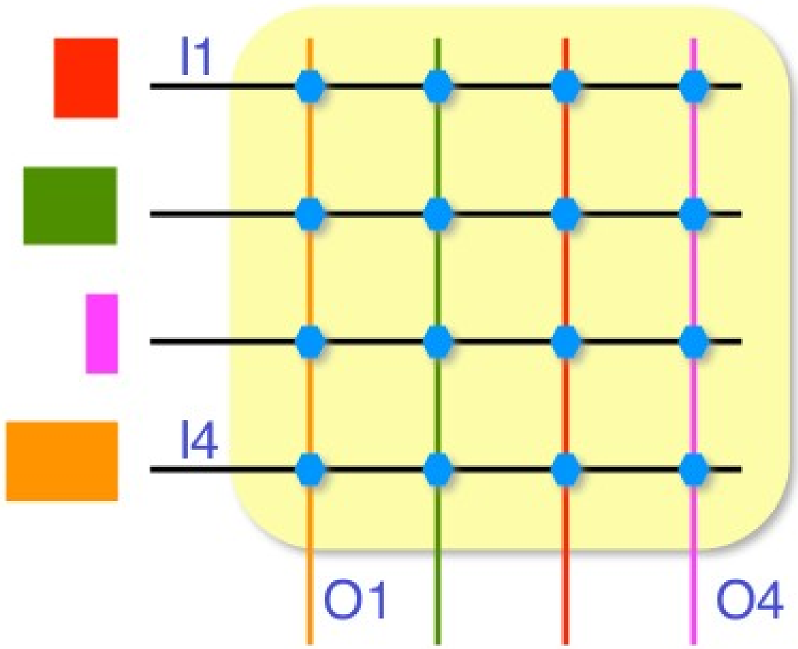


Network Switch: Crossbar



- Each input port can potentially be connected to each output port
- At any given time, only one input port can be connected to a given output port
- Different output ports can be reached concurrently by different input ports

Network Switch: Crossbar

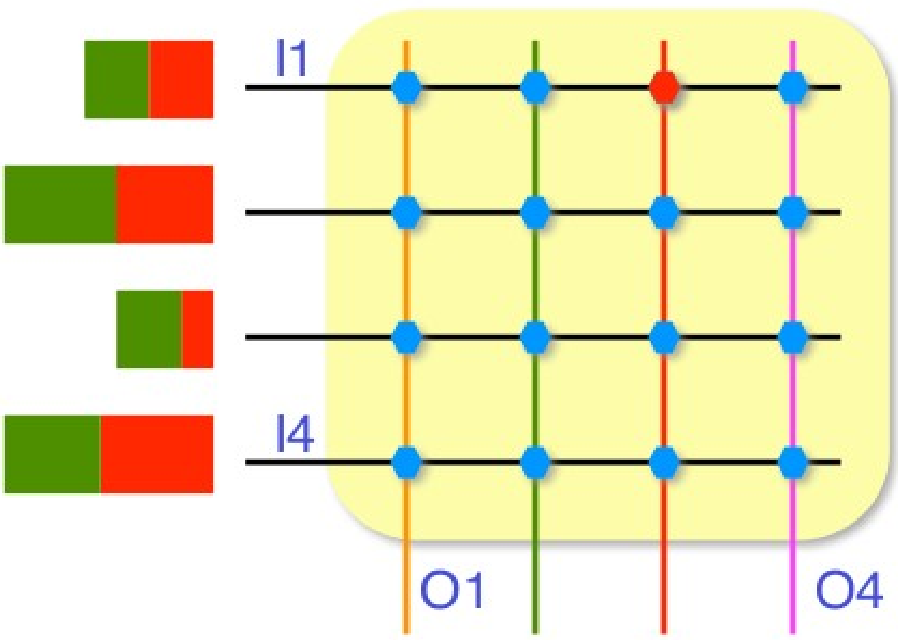


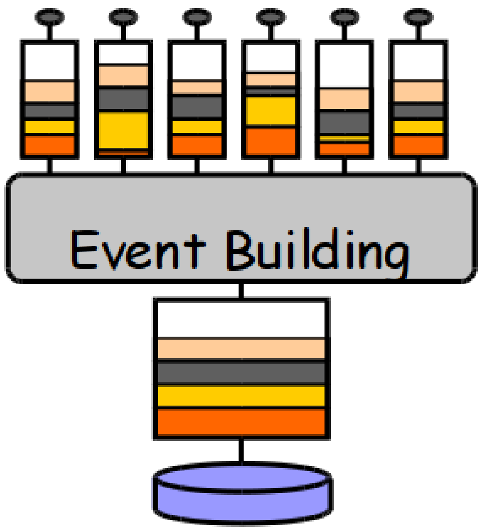
- Ideal situation → all inputs send data to different outputs

No interference (Congestion)

All input ports send data concurrently

Crossbar Switch: Event Building



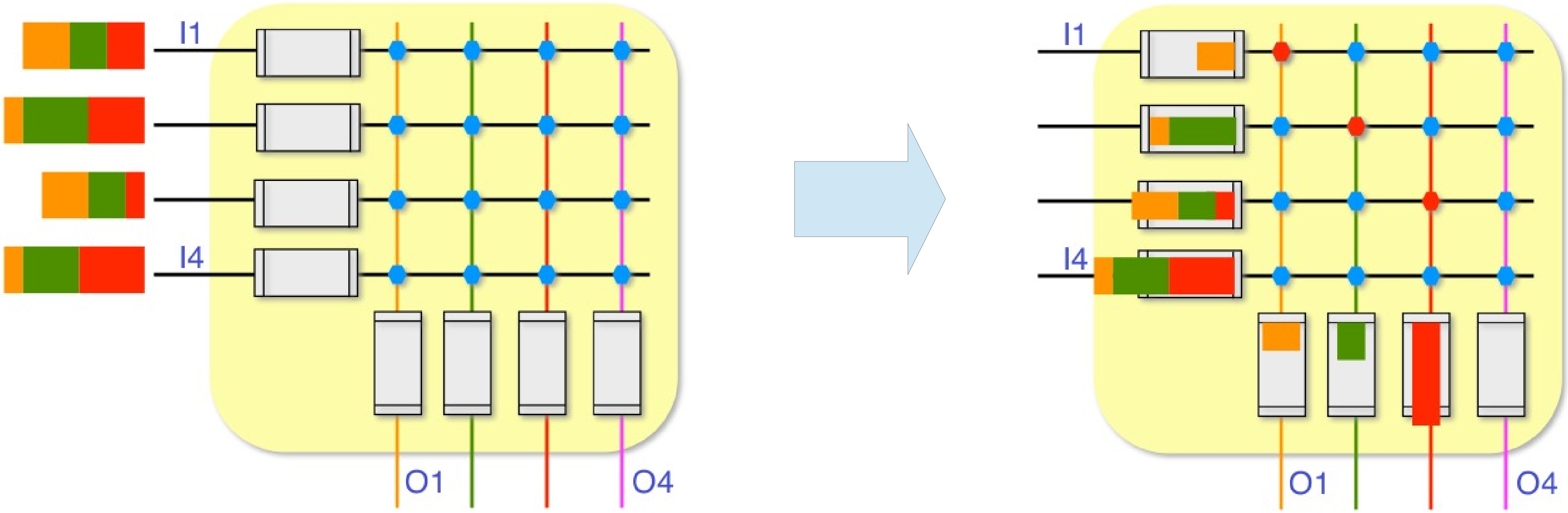


- EB workload implies converging data flow
  - all inputs want to send to same destination **at the same time**
- "Head of line blocking"
  - congestion

Congestion

- Well know phenomena ..
  - in Geneva and other cities
- Differently from road traffic, Ethernet HW allowed to “drop” vehicles (packets)
  - Higher level protocols have to take care of re-sending
  - Possibly important performance impacts

Queueing



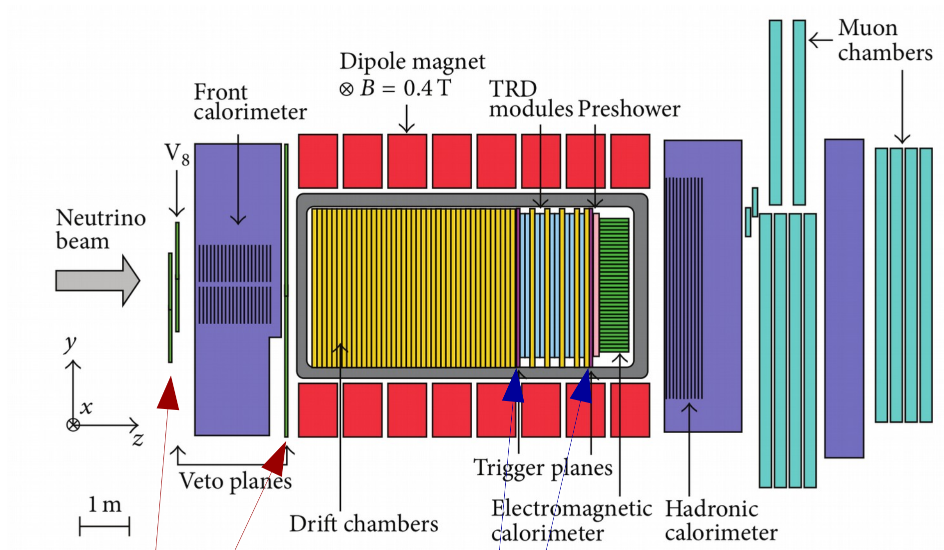
- Adding input and output FIFO dramatically improves EB pattern handling
- EB workload anyway problematic
  - limited FIFO size, variable data size
  - limited internal switching speed

traffic shaping
or
network oversizing

Use Cases

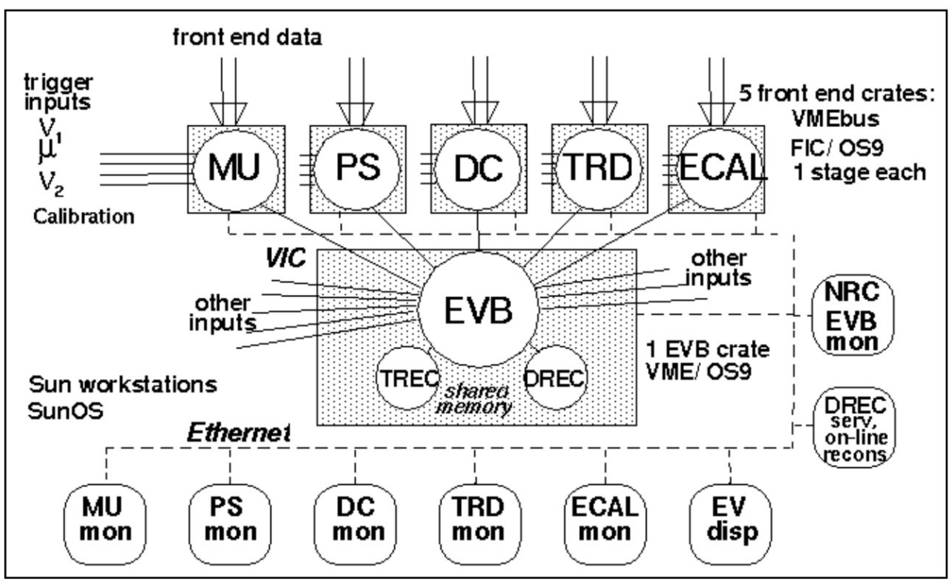
NOMAD (1995-1998)

- Search for $\nu_\mu \rightarrow \nu_\tau$ oscillations at the CERN WA neutrino facility (WANF)
- $2.4 \times 2.4\ m^2$ fiducial (beam) area
- two 4ms spills with 1.8×10^{13} P.o.T. each (ν spills)
- a (2s) slow-extraction spill (μ spill)
- 14.4s cycle duration



veto counters

trigger counters



→ DAQ layout

WANF - SPS SuperCycle

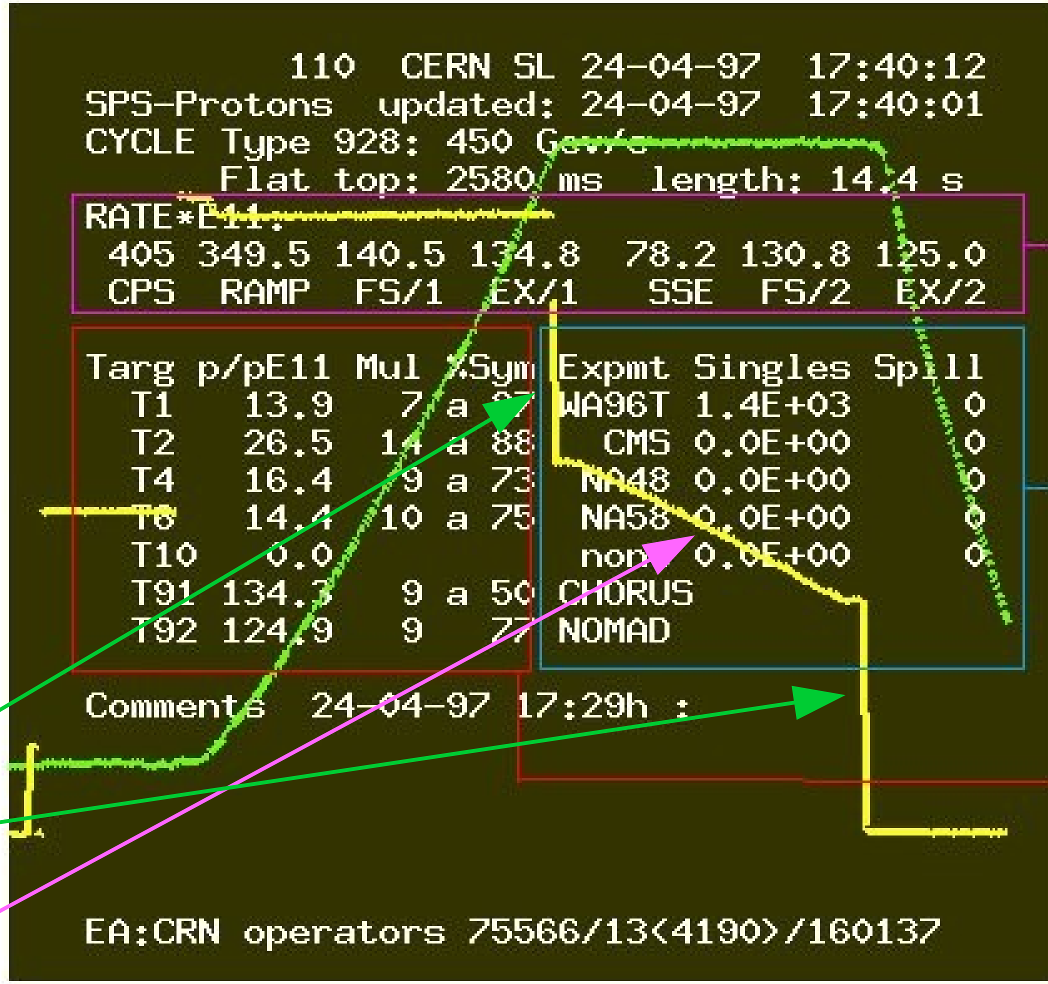
14.4 s cycle length

2 × 4 ms neutrino spills (f/s extractions)

1 × 2 s muon spill (slow extraction)

f/s extractions

slow extraction



Intensities in the SPS

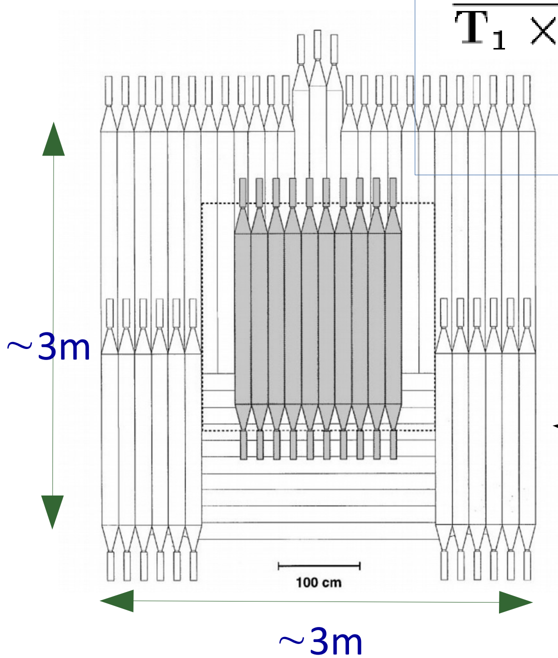
Data from experiments

Steering on targets

Triggering once more ...

menu for NOMADs:

ν-spill triggers	μ-spill triggers
$\overline{V} \times T_1 \times T_2$	$V \times T_1 \times T_2$
$\overline{V_8} \times FCAL$	$V_8 \times T_2$
$\overline{V_8} \times FCAL' \times T_1 \times T_2$	$V_8 \times T_1$
$\overline{T_1 \times T_2} \times ECAL, \overline{V_8} \times ECAL$	$V_8 \times T_1 \times T_2 \times FCAL'$
RANDOM	$V \times T_1 \times T_2 \times ECAL$

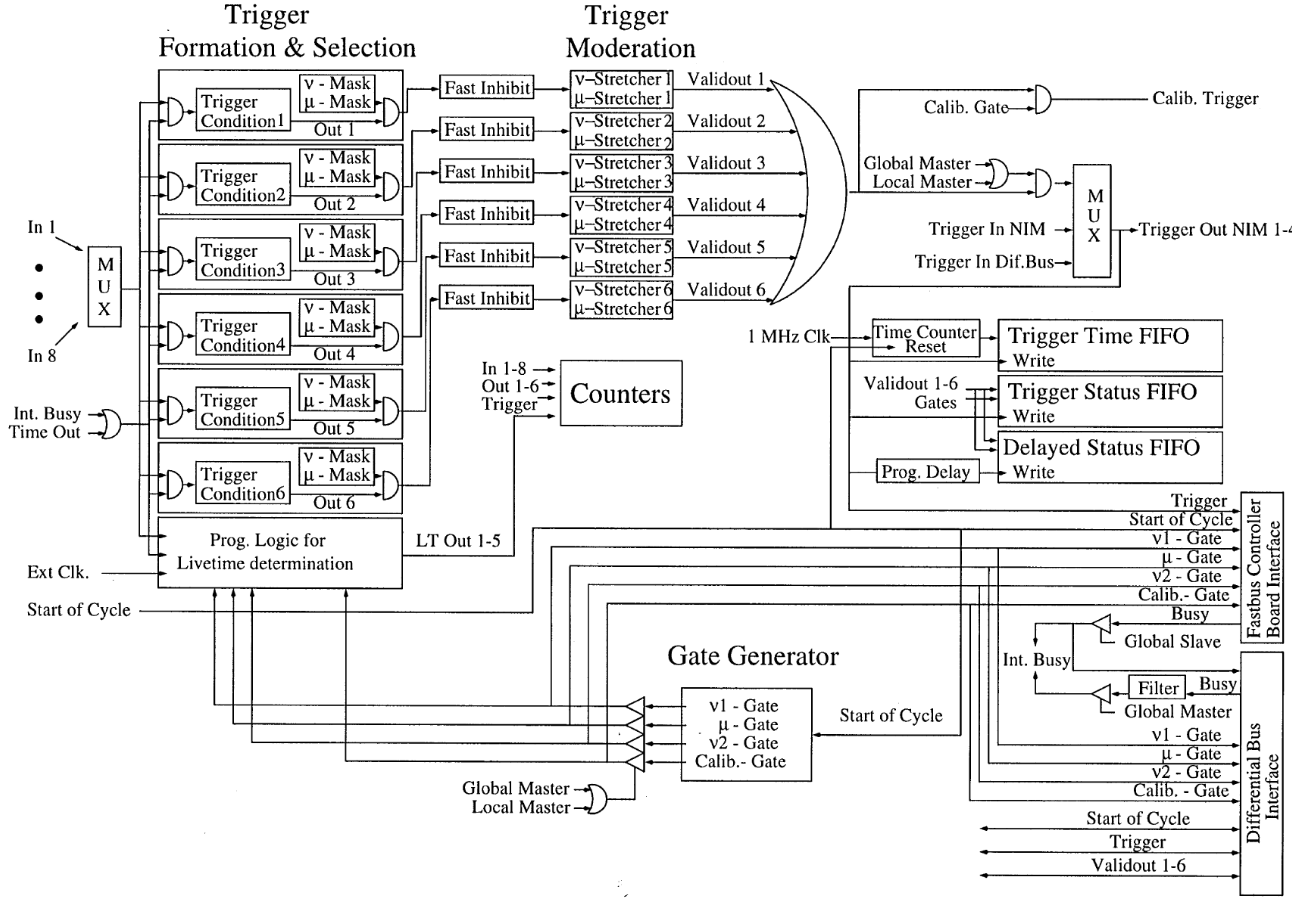


veto counters (central shaded area is V8)

Triggering → FPGA.s at work

MOdular TRIgger for NOmad (MOTRINO):

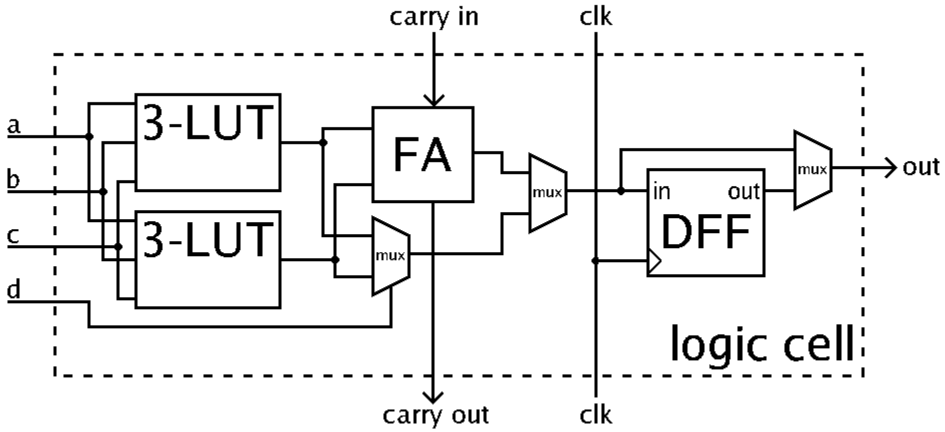
6 VME boards providing local and global trigger generation and propagation



FPGA: Field-Programmable Gate Array

array of configurable logic blocks providing tons (thousands → millions) of dynamically connected logic units, made of:

- look-up tables (to perform complex combinatorial functions)
- flip-flops (to synchronously store results)



Incredibly competitive wrt. ASIC until you don't need millions of chips

With O(50-100 €) you may start up your own project in few days

http://www.fpga4fun.com:

"you can design a circuit on your computer and have it running on your desk in minutes"

NOMAD DAQ

- $\sim$30(?) (64 or 96 channel) Fastbus xDC boards [x = Q, P, T]
- Typically:
    - $\sim$15 evts in each 4ms spill (neutrino triggers)
    - $\sim$60 evts in each 2s spill (muon triggers)
    - 256-events in off-spill calibration cycles (calibration triggers)
  - On spill(cycle): on-board buffering of up to 256 events (no way to read event-by-event)
  - End of spill(cycle): block transfer to 5 VME PU.s (motorola 68040 FIC8234 board, OS9 real-time system)
  - Event building and storage on another VME PU
  - Monitoring and control on SunOs/Solaris workstations

$\rightarrow$ on-board buffering

$\rightarrow$ data processing done off-spill (i.e. off-beam)

DREAM (2006→): a Testbeam Case

R&D on dual-readout calorimetry, setup:

- Crystals
- Scintillating/cherenkov fibers in lead/copper matrices
- Scintillator arrays as shower leakage counters
- Trigger/veto/muon counters
- Precision chamber hodoscope

… always evolving

Acquiring: waveforms, total charge, time information

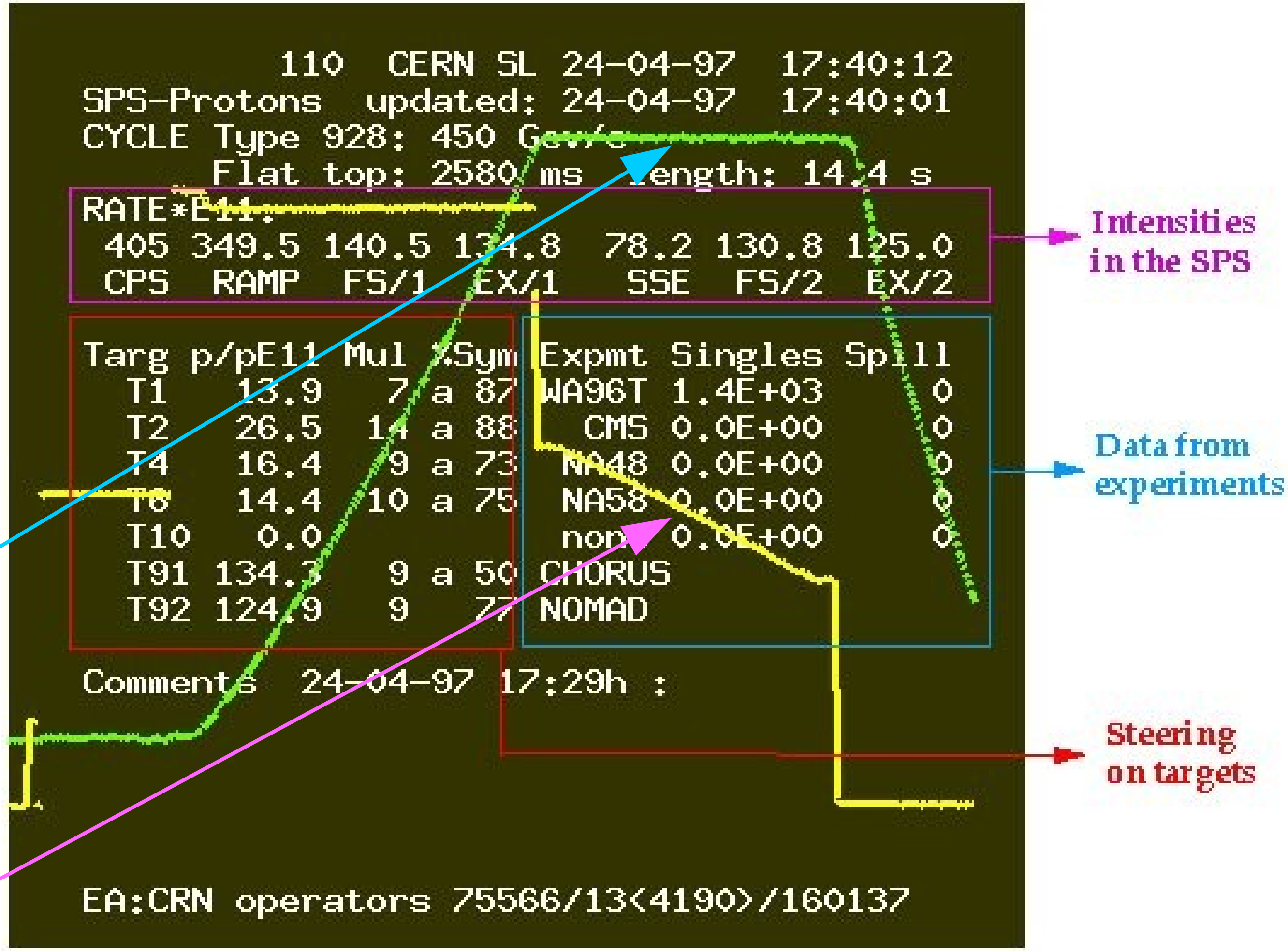
DREAM (2006→): a Testbeam Case

a possible SPS cycle

beam:

~2 s / 14.4 s

(flat top)



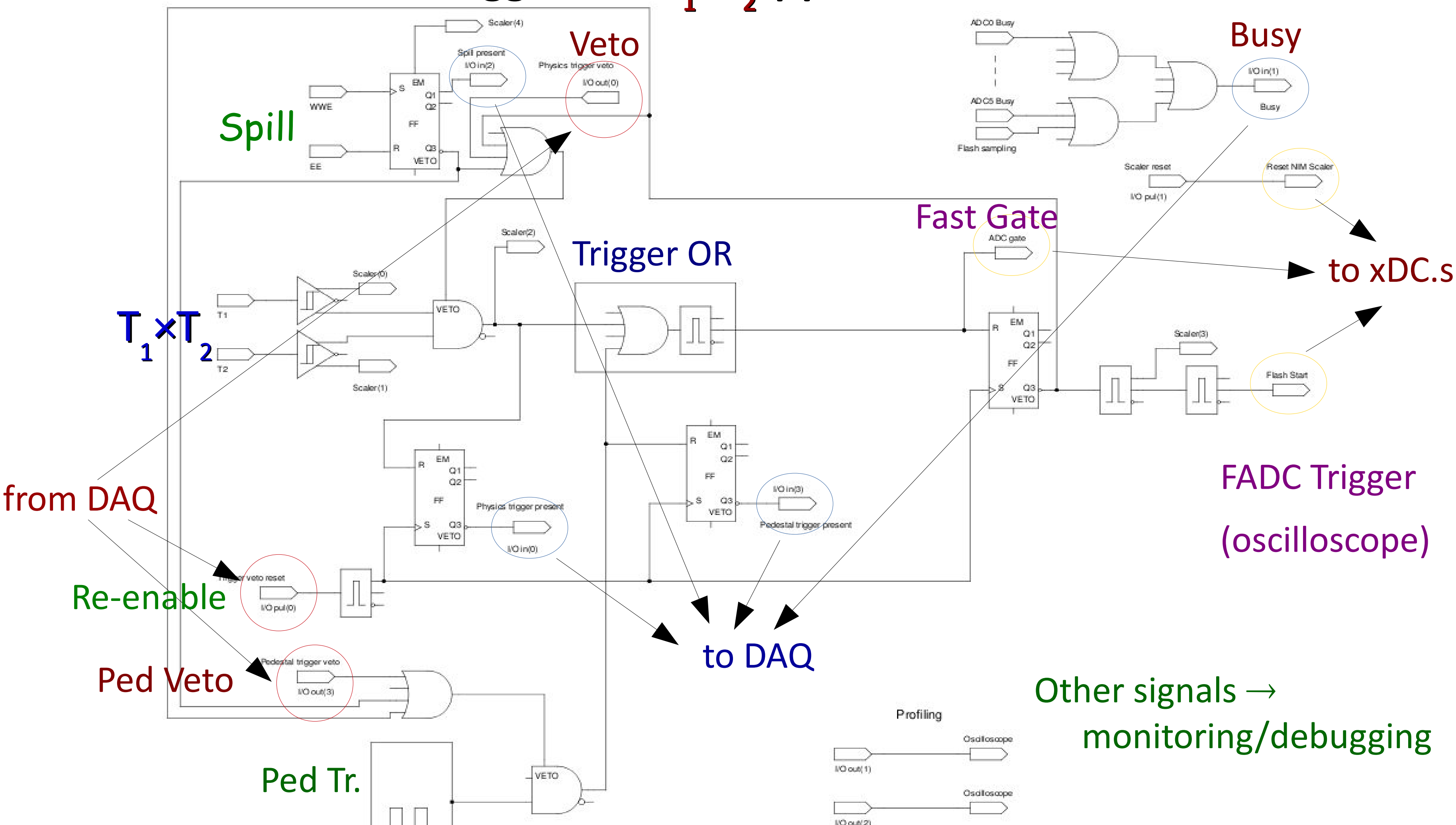
flat top

slow extraction

Trigger = ($\overline{V}\times T_1\times T_2$ | ped) → easy !

Spill-driven (asynchronous) Trigger

Trigger = $\overline{V} \times T_1 \times T_2$ | ped



First discrete, then FPGA (Xilinx Spartan 3AN evaluation board)

DREAM DAQ

1 PC → 2 VME crates (access via CAEN optical interfaces) + 1 PC → storage
6 x 32 ch xDC.s (x = Q, T : CAEN V792, V862, V775)
1 x 34 ch (CAEN V1742) 5Gs/s Digitizer (single event: ~34x1024x12bit)
1 x 4 ch Tektronix TDS7254B 20 Gs/s oscilloscope
… few VME I/O & discriminator boards

DAQ logic spill-driven (no “real time”, scientific linux desktop)

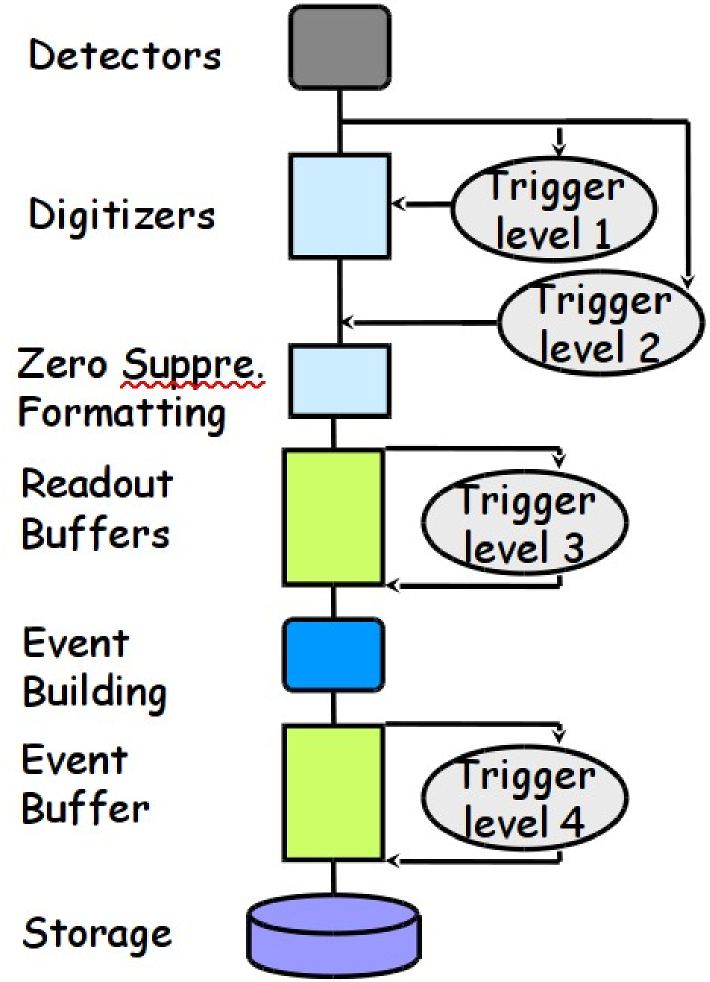
- in-spill (slow extraction)
  - a) poll trigger signal … if trigger present:
    - b) read all VME boards (w/ DMA, whenever possible)
    - c) format & store on a large buffer (FIFO over RAM)
    - d) re-enable trigger
- out-of-spill
  - a) read scope (in case) → size is fixed at run start
  - b.1) monitor data (produce root files)
  - b.2) store on disk files (beam and pedestal files) over network

rate ~ O(1 kHz)

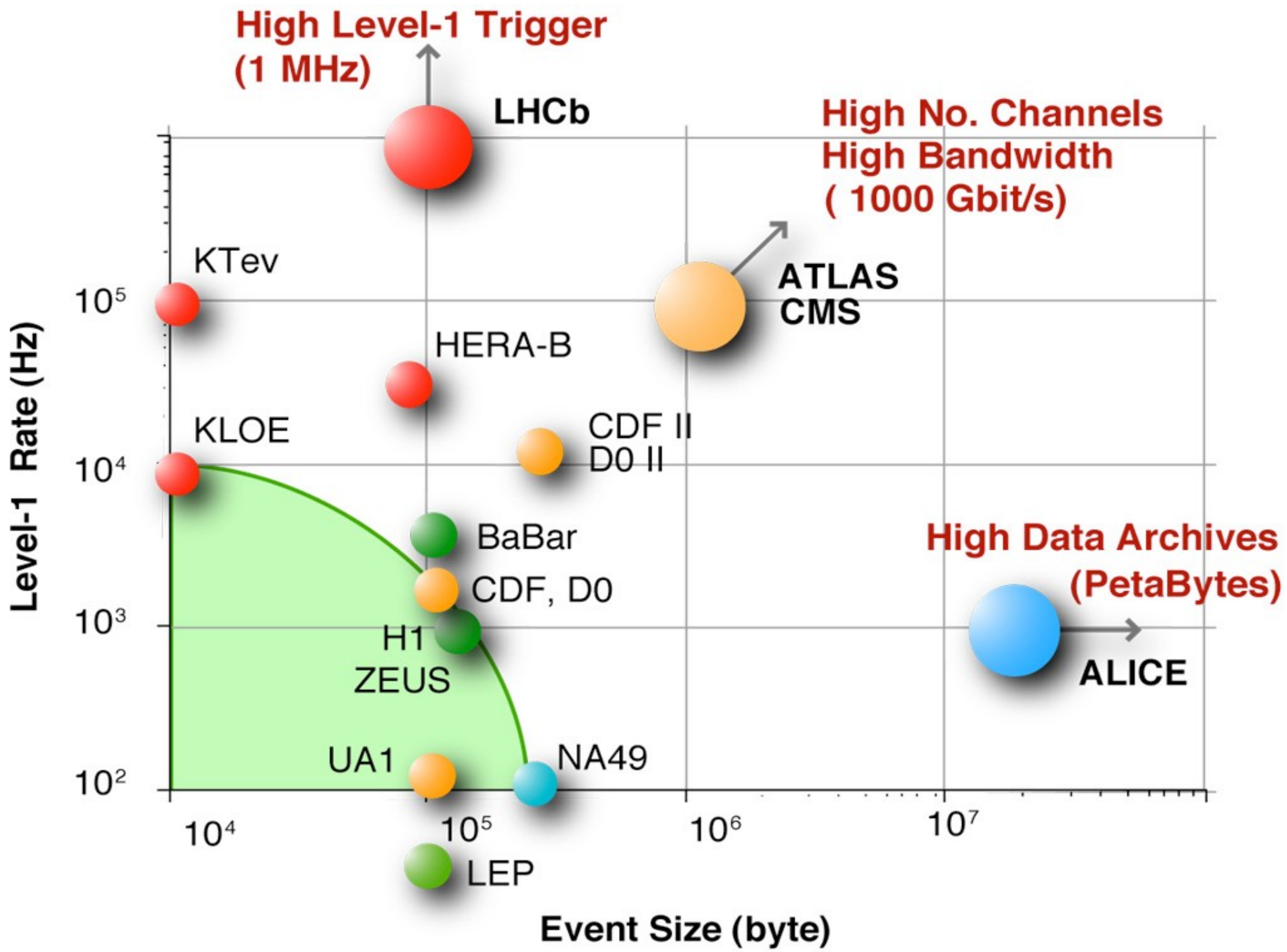
LHC Experiments

Multi-Level Trigger Systems

- Often impossible to take a proper decision in a single place
  - too long decision time
  - too far
  - too many inputs
- Distribute decision burden in a hierarchical structure
  - usuall $\tau_{N+1} \gg \tau_N$, $f_{N+1} \ll f_N$
- At DAQ level, proper buffering must be provided for every trigger level
  - absorb latency
  - de-randomise



HEP DAQ Phase-Space



Trigger & DAQ Challenges @ LHC

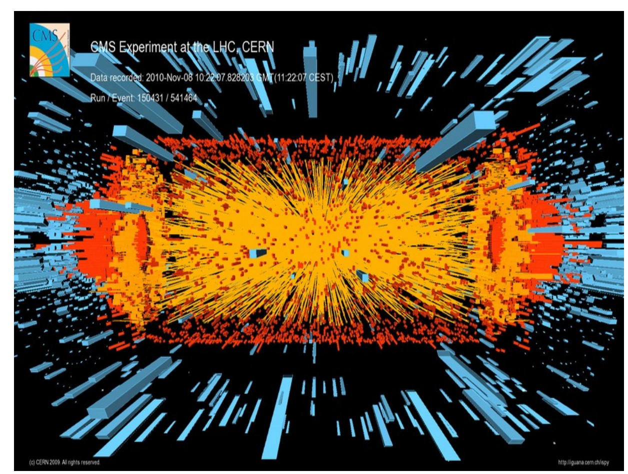
- Experiments with O(10^7-10^8) channels operating at 40 MHz (25 ns) → ~**40 TB/s**
- Searching for extremely rare phenomena:

$$\sigma_H / \sigma_{Tot} \sim O(10^{-13})$$

… find the needle in a haystack:

- Pretty complex events
  - significant number of overlapping collisions (pile-up)
- Large dimensions/distances (O(10 m))

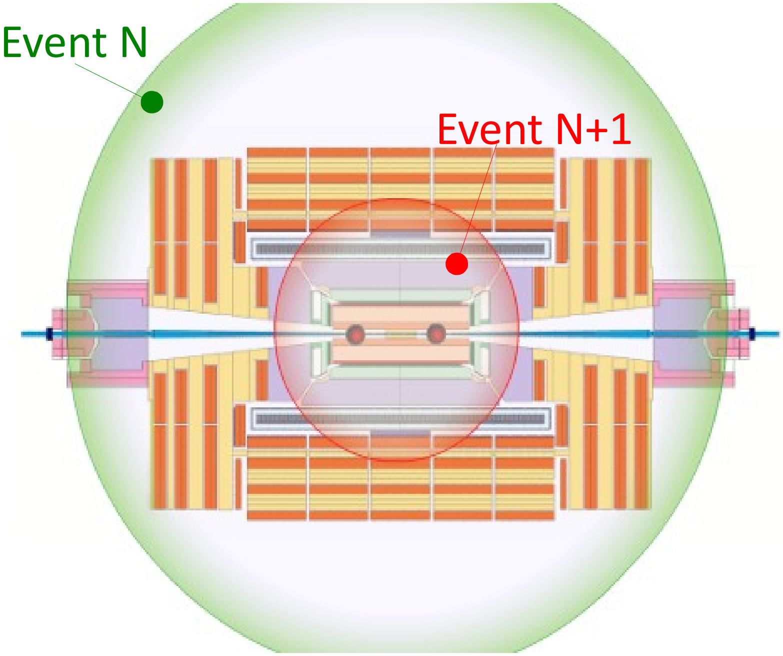
Multi-level trigger system and …



Level-1 Trigger & FE Electronics

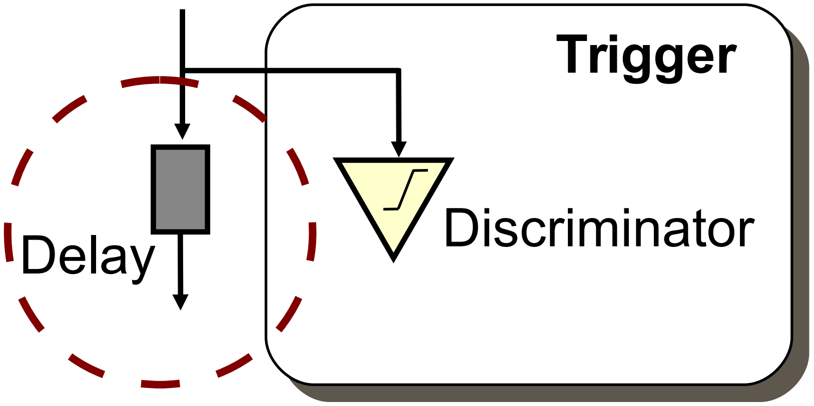
- Particle time of flight >> 25 ns
- Cable delays >> 25 ns

Dedicated synchronization, timing and signal distribution facilities



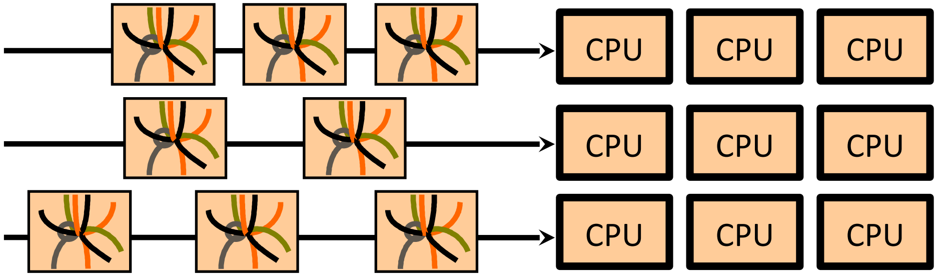
- Typical Level-1 decision latency: O(μs)
  - dominated by signal propagation in cables

Digital/analog custom front-end pipelines store information during L1 trigger decision



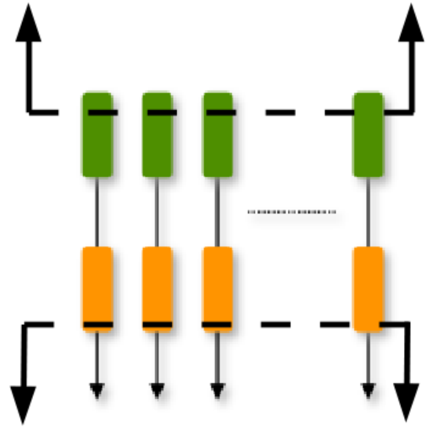
After Level-1 ?

- Custom hardware for Level-1 trigger and front-end electronics followed by network-based **High-Level Trigger** farm(s)
  - commercially available HW organized in a farm
    - <u>events are independent</u>



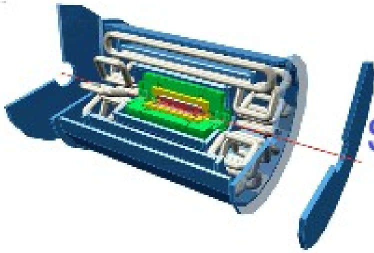
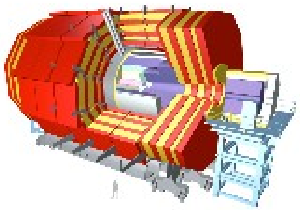
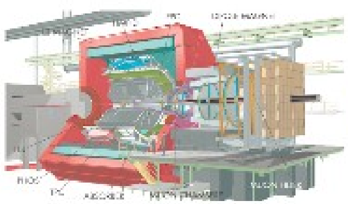
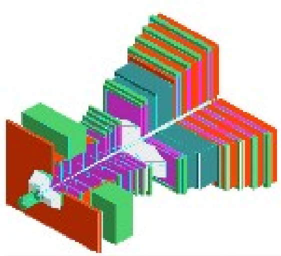
- Connection between via dedicated HW and point-to-point connectivity
  - electrical or optical, standard or custom

Custom electronics

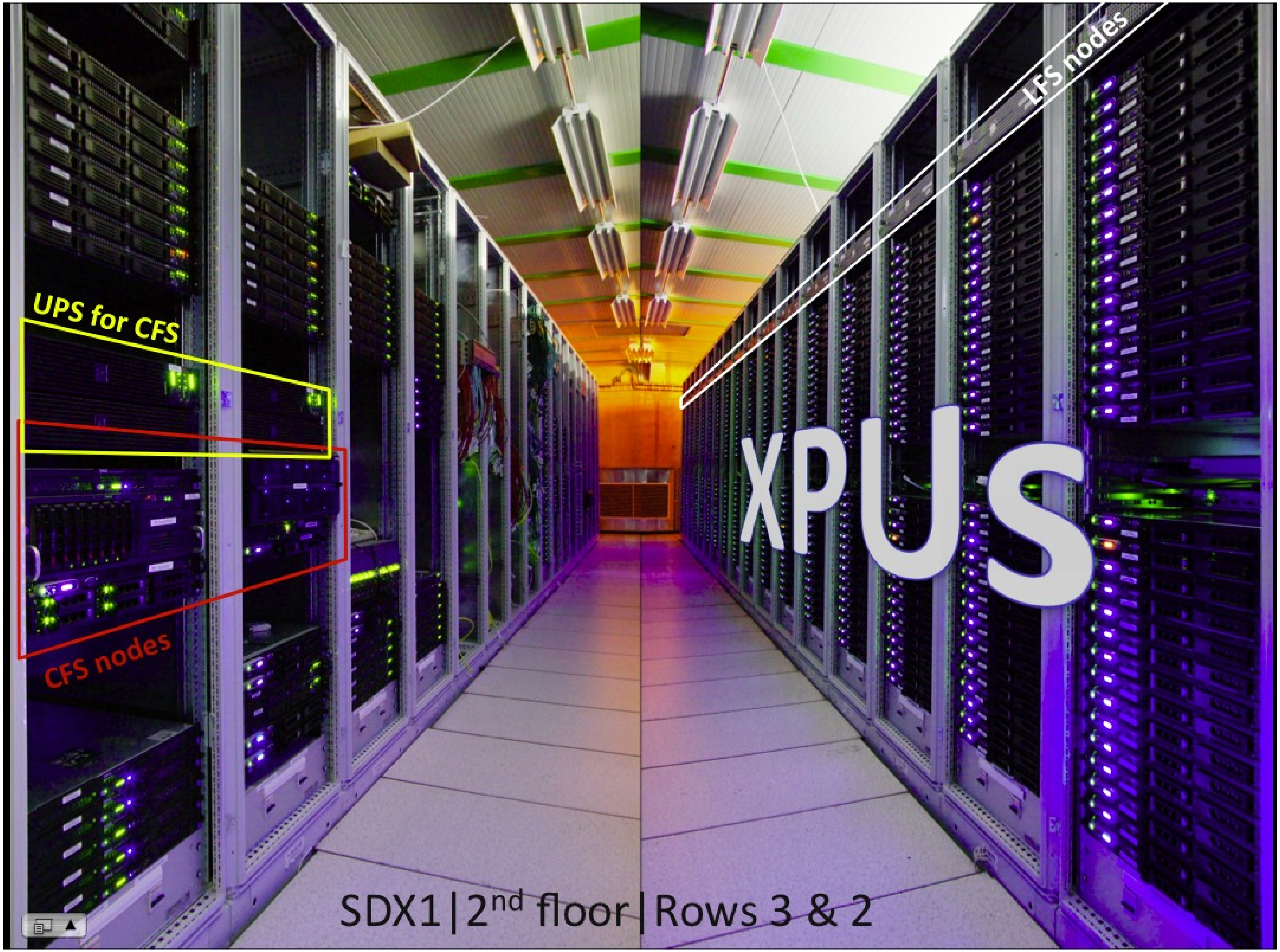


Network based

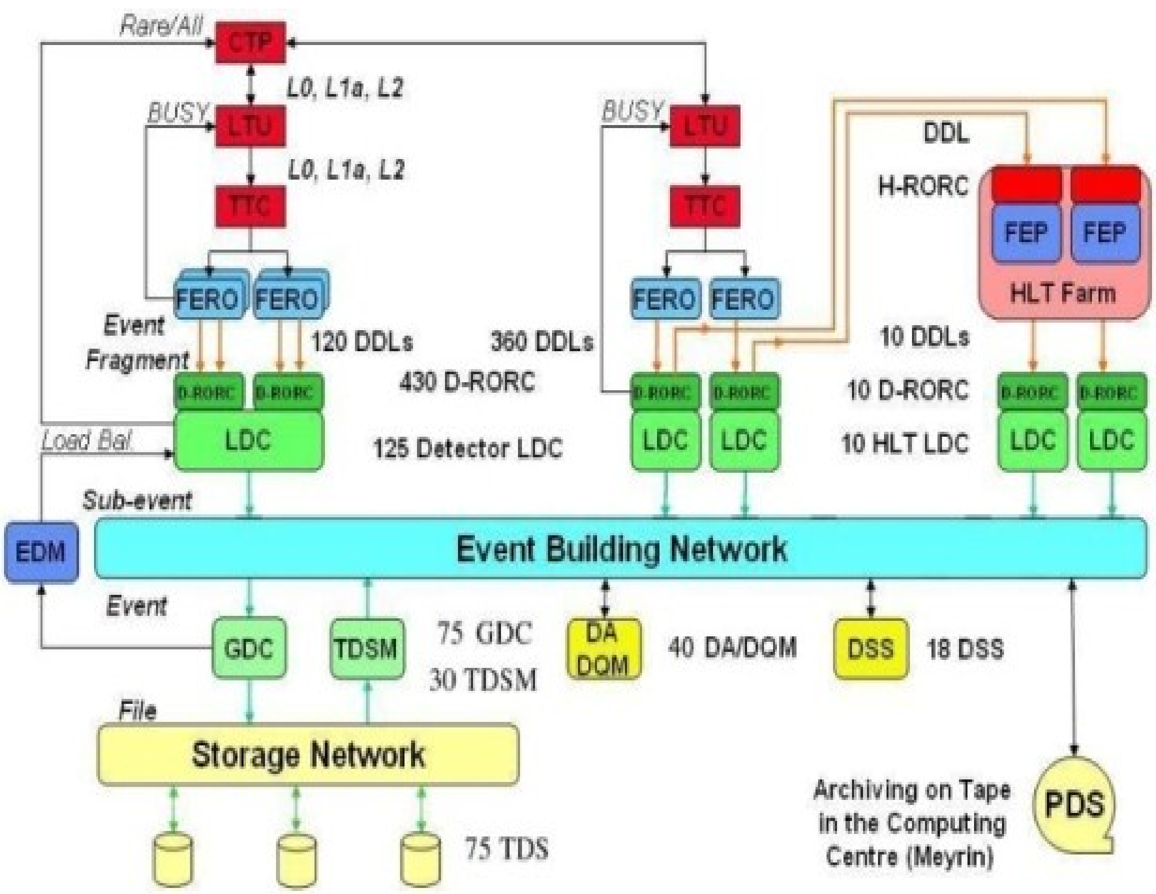
Read-Out Links (in LHC Run 1)

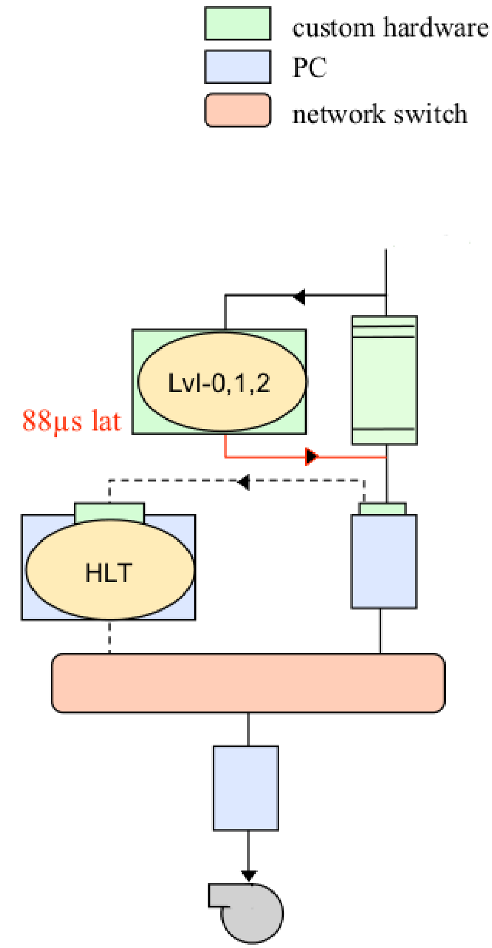
			Flow Control
	SLINK	Optical: 160 MB/s ≈ 1600 Links Receiver card interfaces to PC.	Yes
	SLINK 64	LVDS: 400 MB/s (max. 15m) ≈ 500 links (FE on average: 200 MB/s to readout buffer) Receiver card interfaces to commercial NIC (Network Interface Card)	yes
	DDL	Optical 200 MB/s ≈ 500 links Half duplex: Controls FE (commands, Pedestals,Calibration data) Receiver card interfaces to PC	yes
	TELL-1 & GbE Link	Copper quad GbE Link ≈ 400 links Protocol: IPv4 (direct connection to GbE switch) Forms "Multi Event Fragments" Implements readout buffer	no

ATLAS HLT Farm

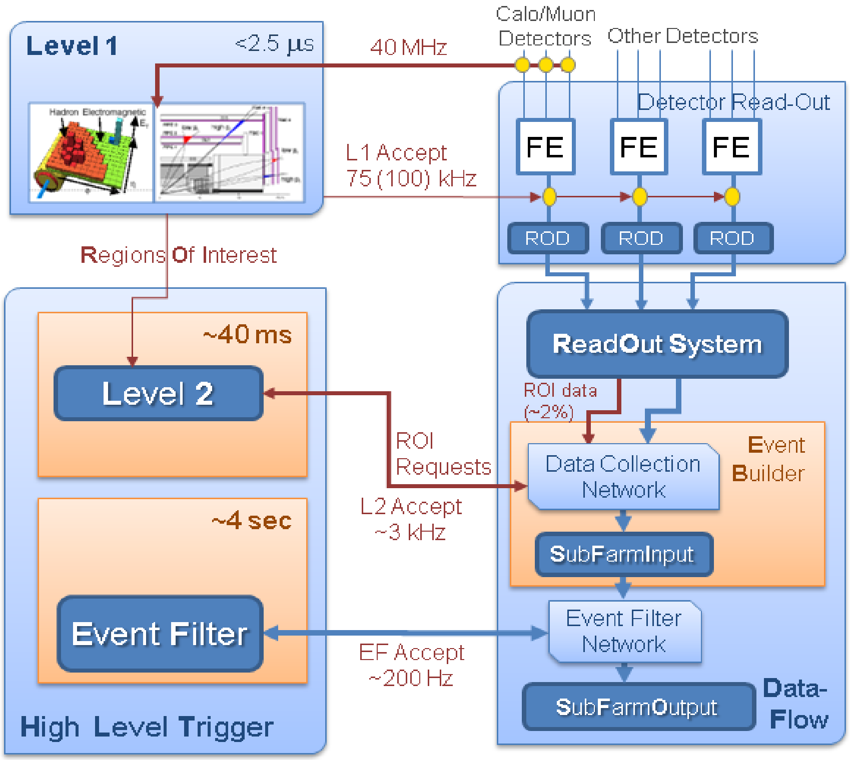


ALICE

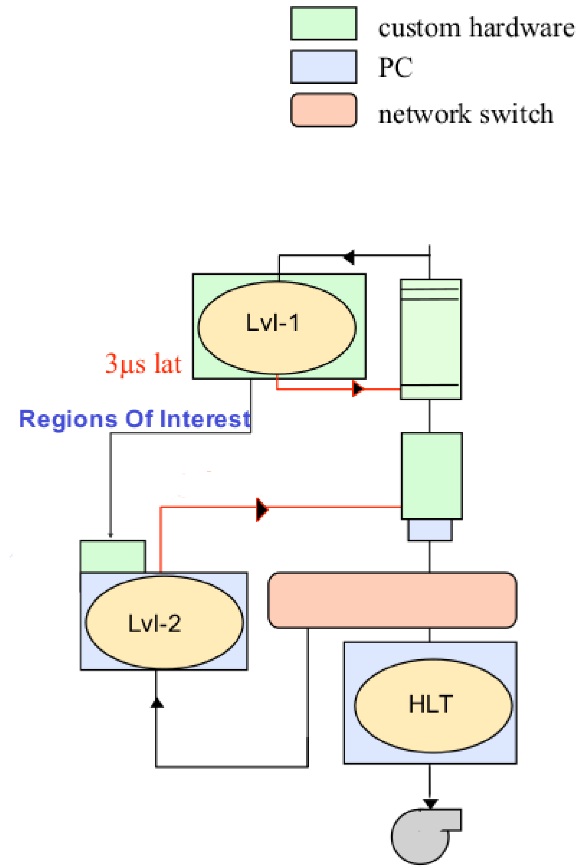

Rare/All
CTP
L0, L1a, L2
BUSY
LTU
L0, L1a, L2
TTC
FERO
FERO
Event Fragment
120 DDLs
360 DDLs
430 D-RORC
D-RORC
LDC
Load Bal
125 Detector LDC
DDL
H-RORC
FEP
FEP
HLT Farm
10 DDLs
10 D-RORC
10 HLT LDC
Sub-event
EDM
Event Building Network
Event
GDC
TDSM
75 GDC
30 TDSM
DA DQM
40 DA/DQM
DSS
18 DSS
File
Storage Network
75 TDS
Archiving on Tape in the Computing Centre (Meyrin)
PDS


custom hardware
PC
network switch
Lvl-0,1,2
88µs lat
HLT

ATLAS

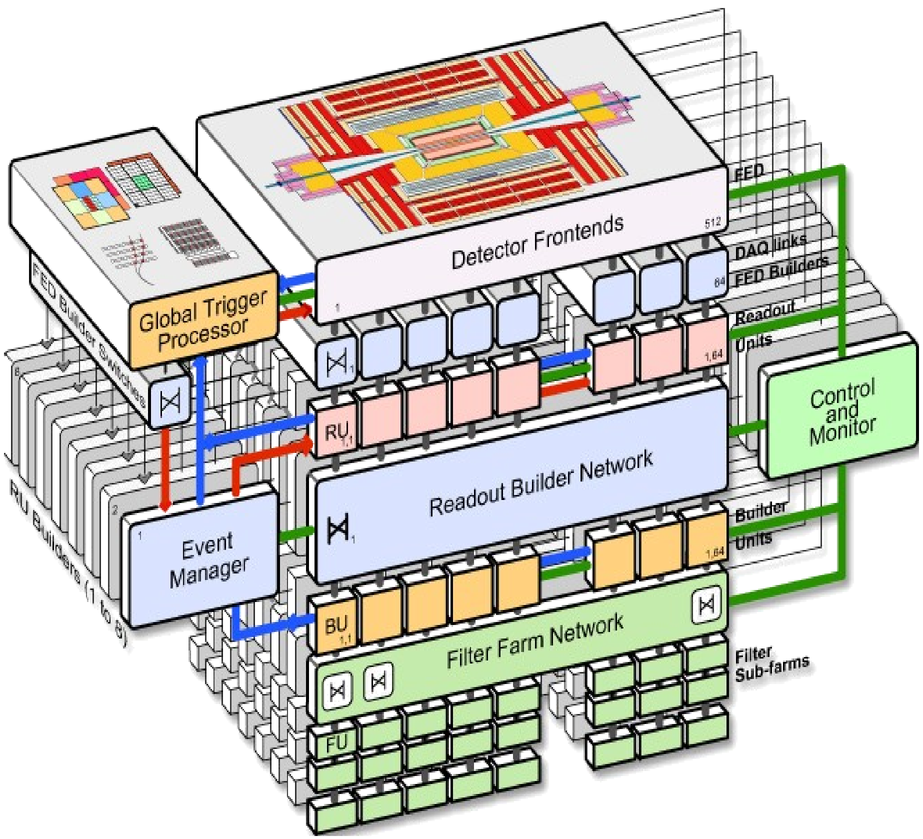


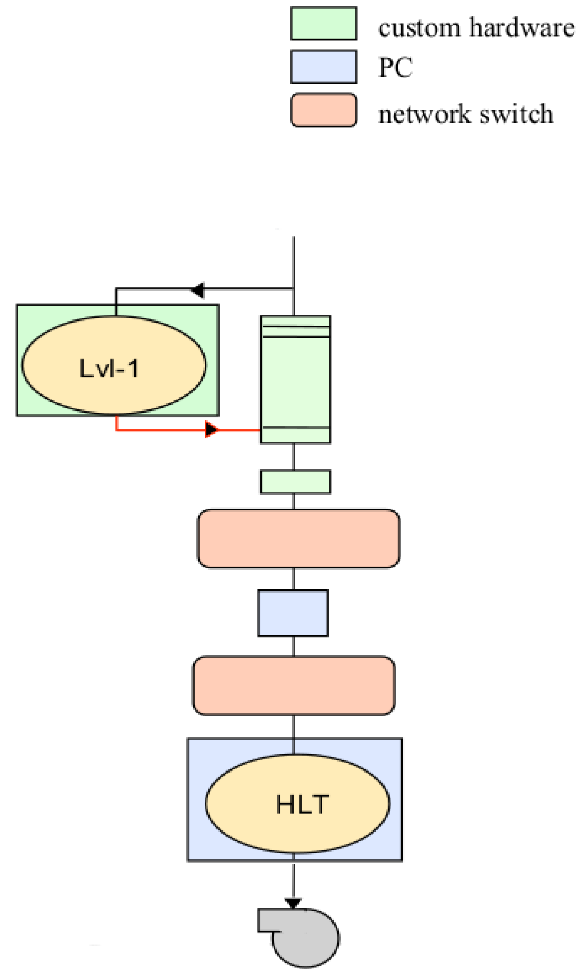
Level 1
<2.5 μs
40 MHz
Calo/Muon Detectors
Other Detectors
Detector Read-Out
FE
FE
FE
L1 Accept 75 (100) kHz
ROD
ROD
ROD
Regions Of Interest
~40 ms
Level 2
ReadOut System
ROI data (~2%)
ROI Requests
Event Builder
Data Collection Network
L2 Accept ~3 kHz
SubFarmInput
~4 sec
Event Filter
Event Filter Network
EF Accept ~200 Hz
SubFarmOutput
Data-Flow
High Level Trigger



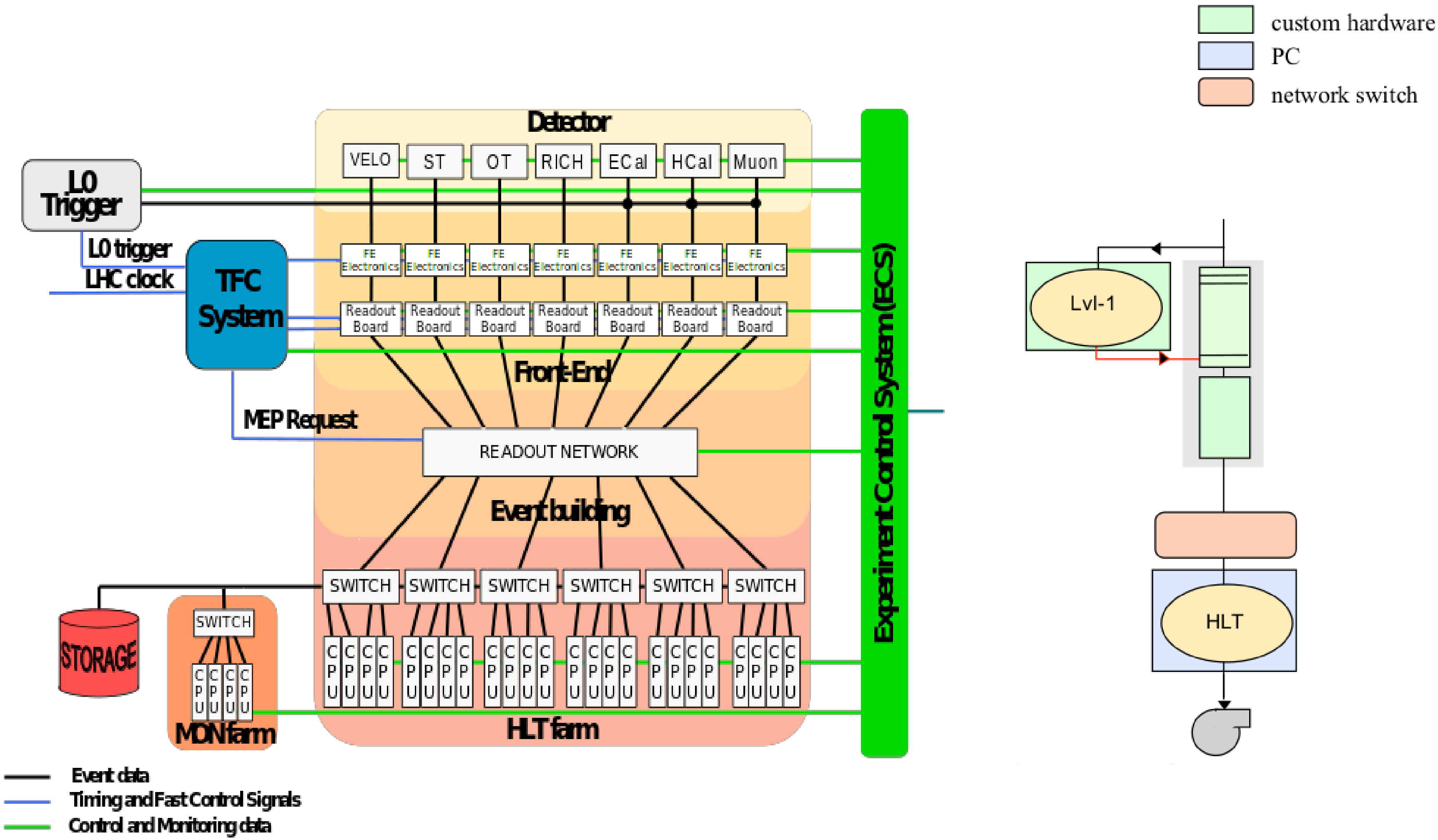
custom hardware
PC
network switch
Lvl-1
3μs lat
Regions Of Interest
Lvl-2
HLT

CMS

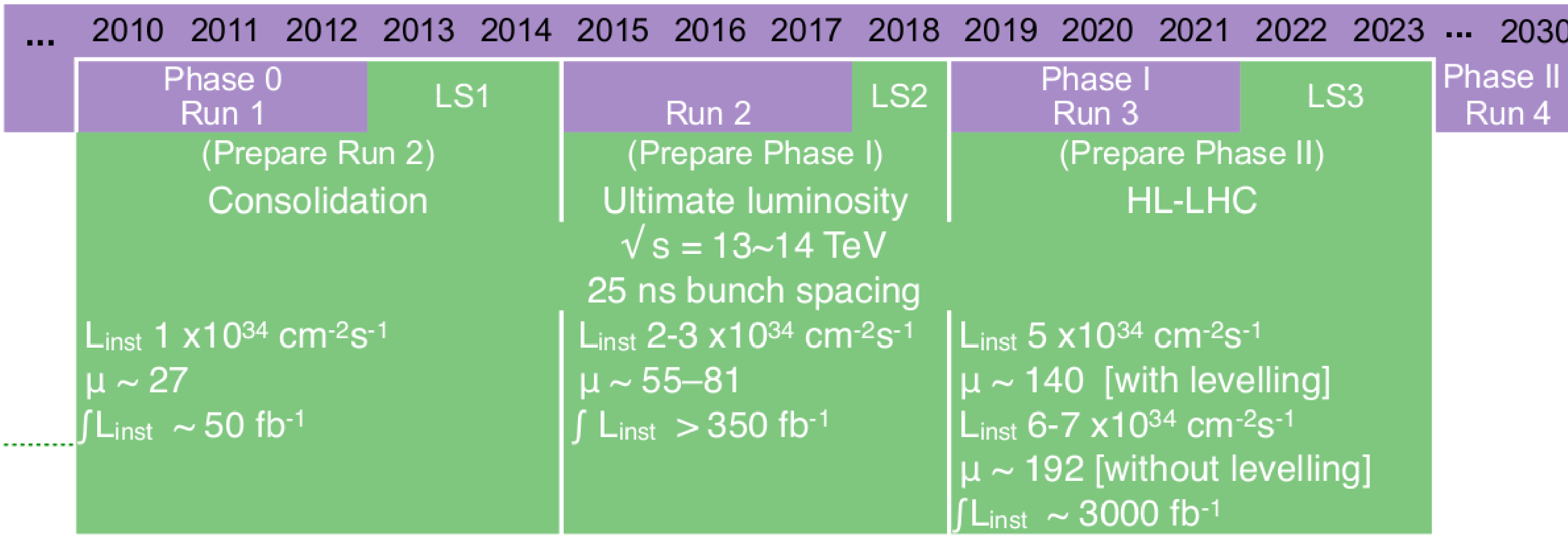

Detector Frontends
Global Trigger Processor
FED Builder Switches
RU Builders (1 to 8)
FED
DAQ links
FED Builders
Readout Units
RU
Readout Builder Network
Control and Monitor
Builder Units
Event Manager
BU
Filter Farm Network
Filter Sub-farms
FU


custom hardware
PC
network switch
Lvl-1
HLT

LHCb

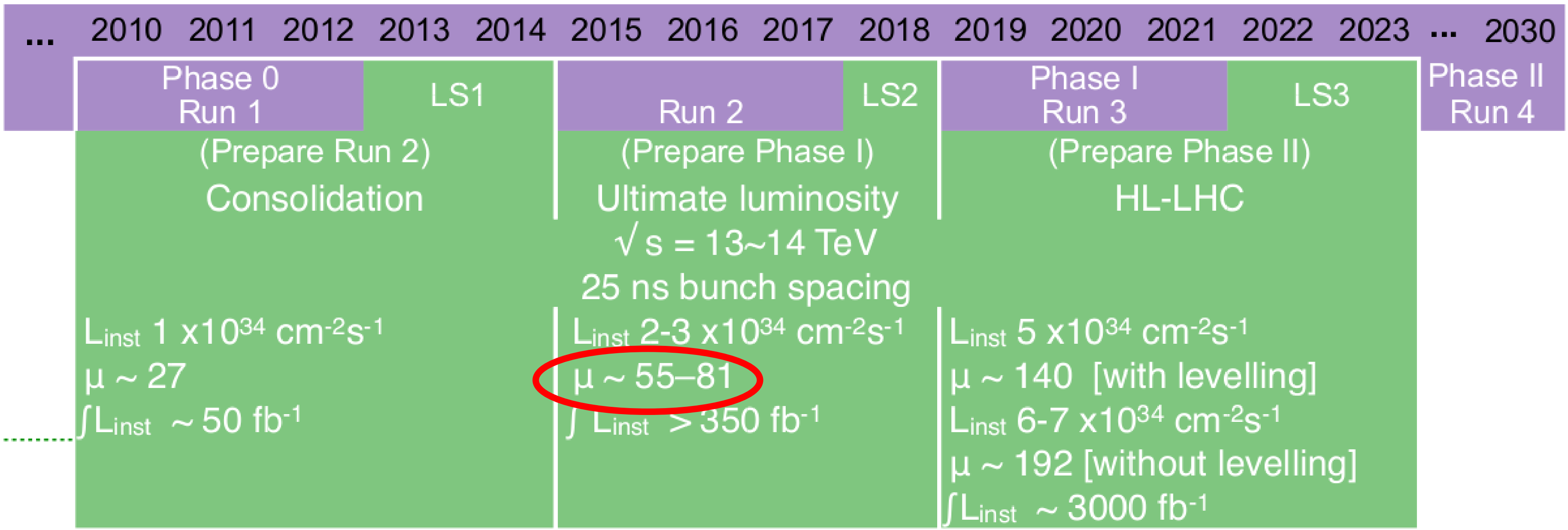

custom hardware
PC
network switch
Detector
VELO
ST
OT
RICH
ECal
HCal
Muon
L0 Trigger
L0 trigger
LHC clock
TFC System
FE Electronics
Readout Board
Front-End
MEP Request
READOUT NETWORK
Event building
SWITCH
CPU
STORAGE
MON farm
HLT farm
Experiment Control System (ECS)
Event data
Timing and Fast Control Signals
Control and Monitoring data
Lvl-1
HLT

Long Shutdown 1 – TDAQ Perspective



- LHC data-acquisition system backbones installed ~2007
  - during Run 1 → stability, efficiency, performance reach and optimization
- LS1 occasion to
  - upgrade core systems and review architectures
  - introduce new technologies, retire obsolete ones
  - follow changes of detector side
  - prepare for Run2 (and Run3) challenges

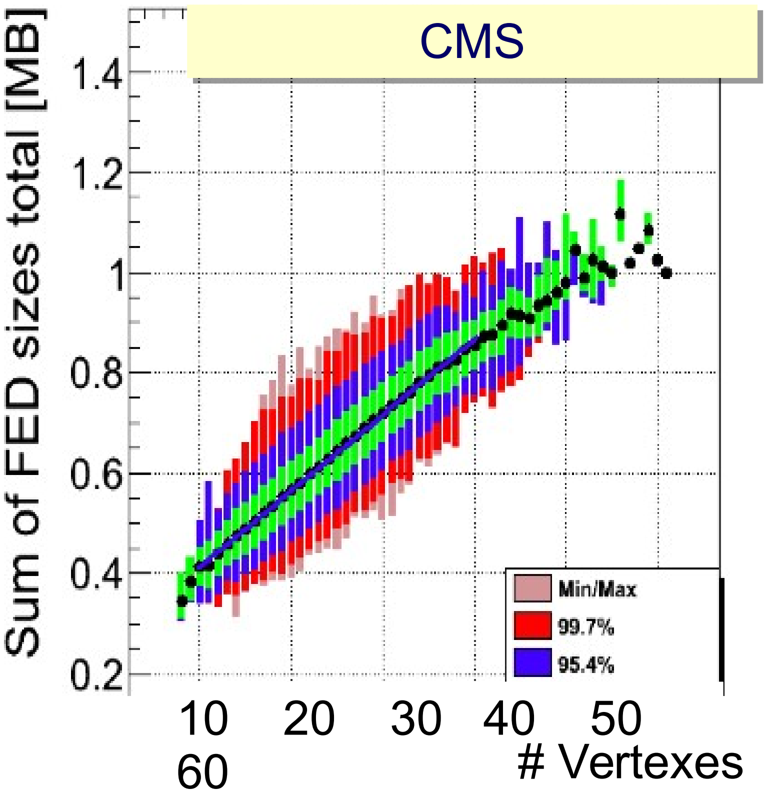
Long Shutdown 1 – TDAQ Perspective

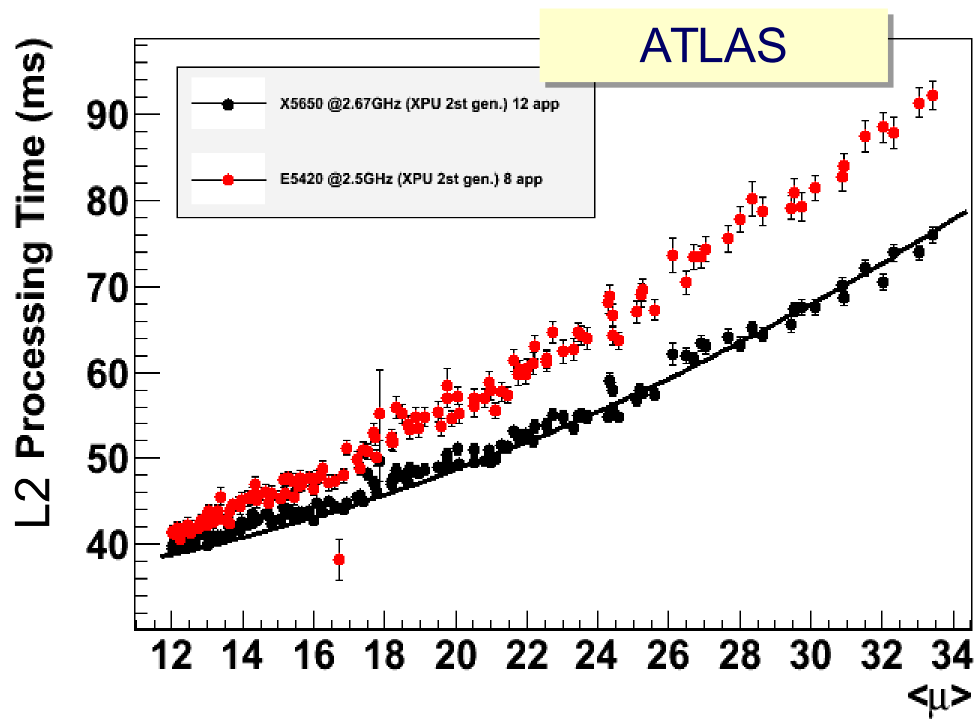


- LHC data-acquisition system backbones installed ~2007
  - during Run 1 → stability, efficiency, performance reach and optimization
- LS1 occasion to
  - upgrade core systems and review architectures
  - introduce new technologies, retire obsolete ones
  - follow changes of detector side
  - prepare for Run2 (and Run3) challenges

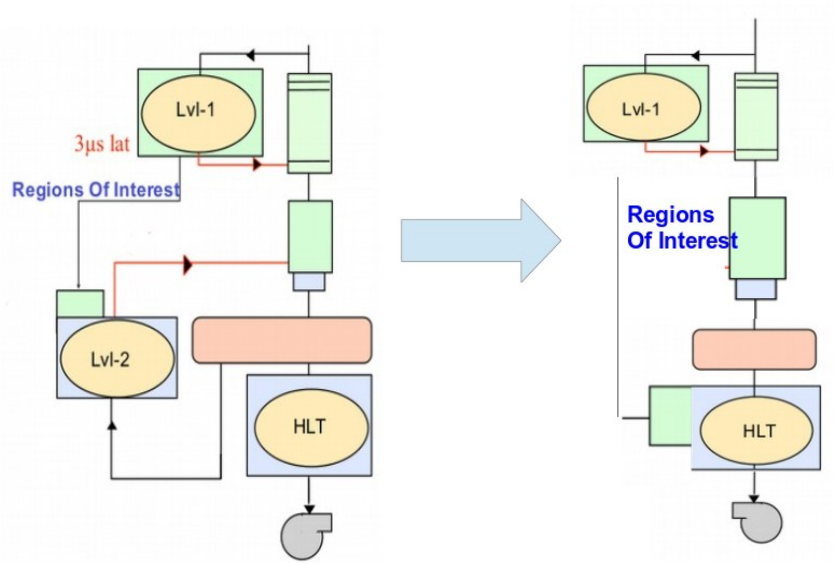
Run 2: Challenges

- Increased pile-up
  - larger data size → bandwidth and storage
  - more complex events → increased computing needs, trigger efficiency and rejection power

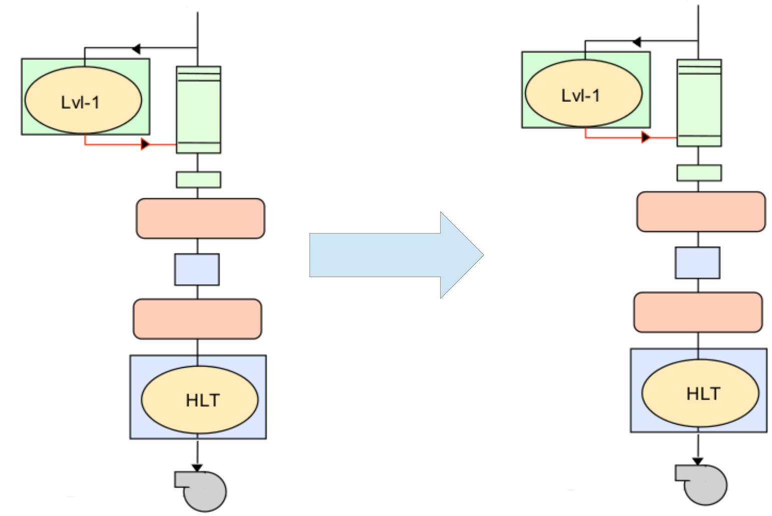




(Some) Run-2 Updates



- Merge L2 and L3 into a single HLT farm
  - preserve Region of Interest, but diluted the farm separation and fragmentation
  - increased flexibly, computing power efficiency



- No architectural changes, but
  - all network technologies replaced
    - Myrinet → Ethernet
    - Ethernet → Infiniband
  - file-based event distribution in the farm
    - achieve full decoupling between DAQ and HLT

Evolution for Run 2

ATLAS:

more like CMS

… still using "L2" ROI, but as first step of a unified L2/EB/HLT process

CMS:

more like ATLAS

… still doing full EB, but analyse ROI first

DAQ@LHC Joint Workshop 2013 :
http://indico.cern.ch/conferenceOtherViews.py?view=standard&confId=217480

Looking forward to LS2 and beyond

DAQ@LHC Workshop

On some long term, all experiments looking forward to significant increase in L1 trigger rate and bandwidth. ALICE and LHCb will pioneer this path during LS2

- First level trigger for Pb-Pb interactions **500 Hz → 50 kHz**
- 22 MB/event
  - **1 TB/s readout → <u>500 PB/month</u>**
- Data volume reduction
  - on-line full reconstruction
  - discard raw-data
- Combined DAQ/HLT/offline farm
  - COTS, FPGA and GPGPU

- **1 MHz → 40 MHz** readout and event building → trigger-less
  - trigger support for staged computing power deployment
- 100 kB/event
  - on-detector zero suppression → rad-hard FPGA
  - 4 TB/s event-building

Almost the End ...

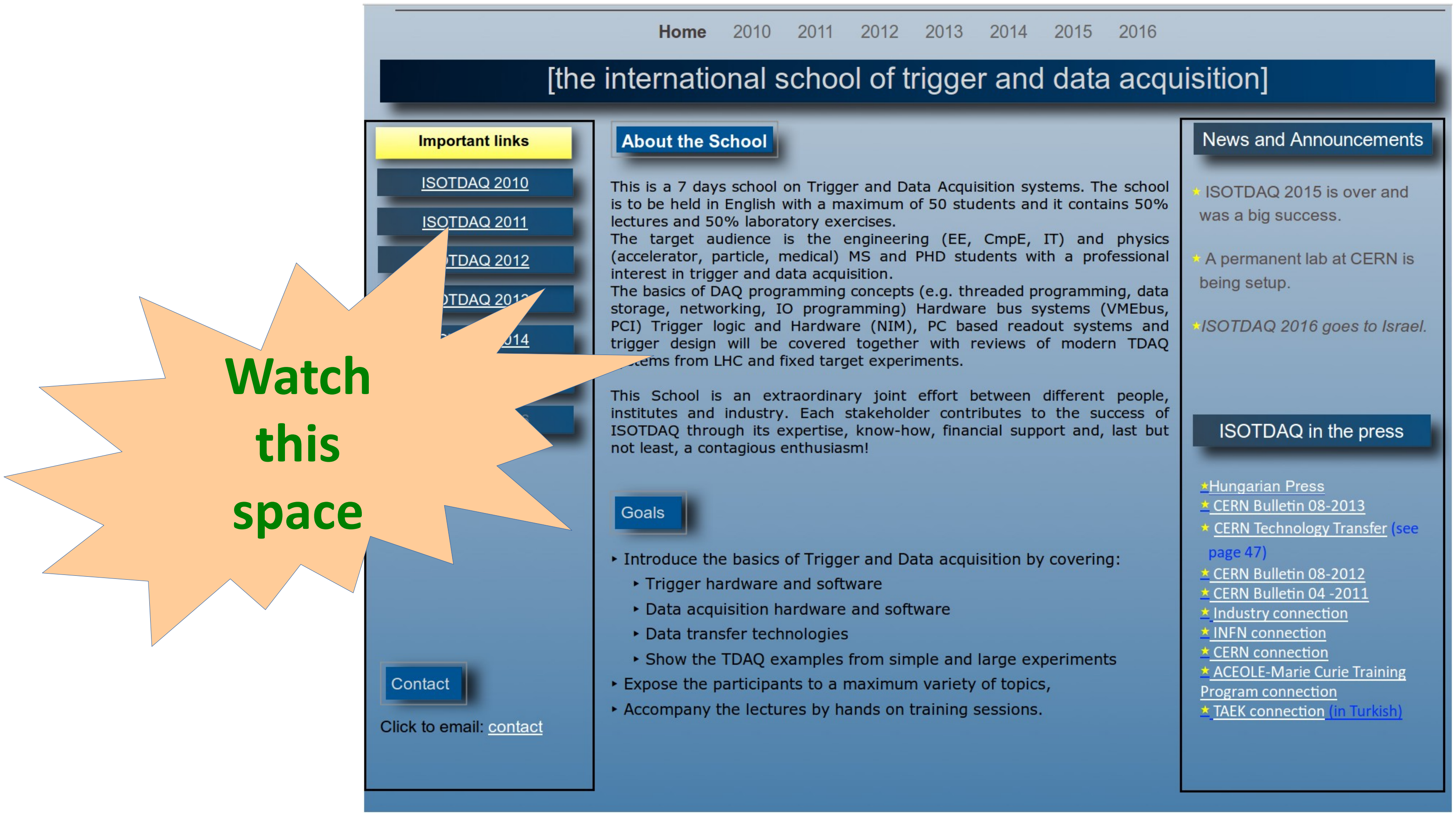
What I did not talk about …

- Many many topics, not exhaustive list:
  - Run Control → DAQ steering, Finite State Machine
  - Configuration → Storing, distributing and archiving SW, HW and trigger configuration
  - Monitoring → data quality, detector state, DAQ functionality

- Your chance of hearing about these and much more and learn through practice at …

ISOTDAQ 2017

- Seventh edition of the **International School of Trigger and Data Acquisition** will be held in Jan-Feb 2017 (location not yet known)



http://isotdaq.web.cern.ch/isotdaq/isotdaq/Home.html

References

- Lectures and papers from H. Spieler
  - http://www-physics.lbl.gov/~spieler/
- Lecture at ISOTDAQ schools
  - http://isotdaq.web.cern.ch/isotdaq/isotdaq/Home.html
- Of course, previous Summer Student courses
  - https://indico.cern.ch/category/345/
- DAQ@LHC Workshops
  - 1st in 2013: https://indico.cern.ch/event/217480/
  - 2nd in 2016: https://indico.cern.ch/event/471309/

Thanks for your patience …