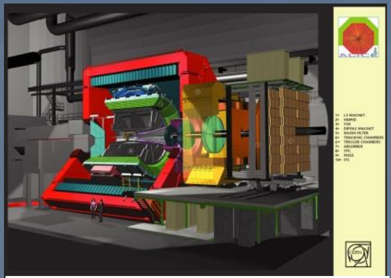
Online Triggers and DAQ

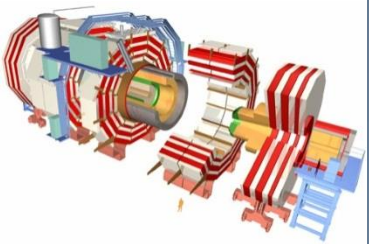
Niko Neufeld, CERN/PH

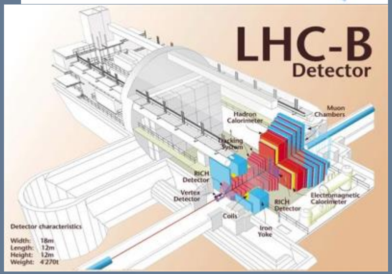
CERN openlab IT Challenges Workshop

The LHC Experiments today

- ALICE – “A Large Ion Collider Experiment”
  - Size: 26 m long, 16 m wide, 16m high; weight: 10000 t
  - 35 countries, 118 Institutes
  - Material costs: 110 MCHF
- ATLAS – “A Toroidal LHC ApparatuS”
  - Size: 4 6m long, 25 m wide, 25 m high; weight: 7000 t
  - 38 countries, 174 institutes
  - Material costs: 540 MCHF
- CMS – “Compact Muon Solenoid”
  - Size: 22 m long, 15 m wide, 15 m high; weight: 12500 t
  - 40 countries, 172 institutes
  - Material costs: 500 MCHF
- LHCb – “LHC beauty” (b-quark is called “beauty” or “bottom” quark)
  - Size: 21 m long, 13 m wide, 10 m high; weight: 5600 t
  - 15 countries, 52 Institutes
  - Material costs: 75 MCHF
- Regular upgrades … first 2013/14 (Long Shutdown 1)

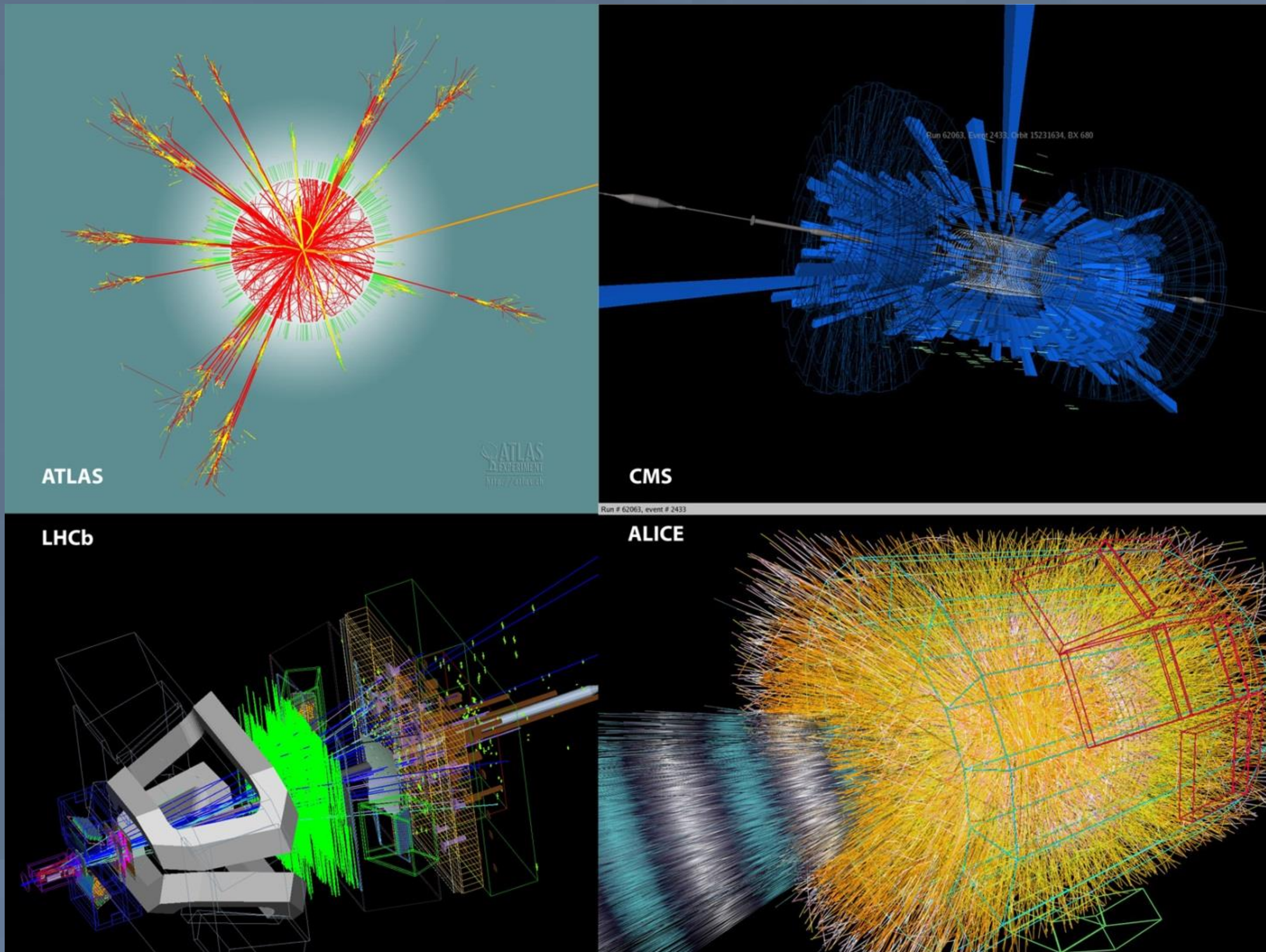






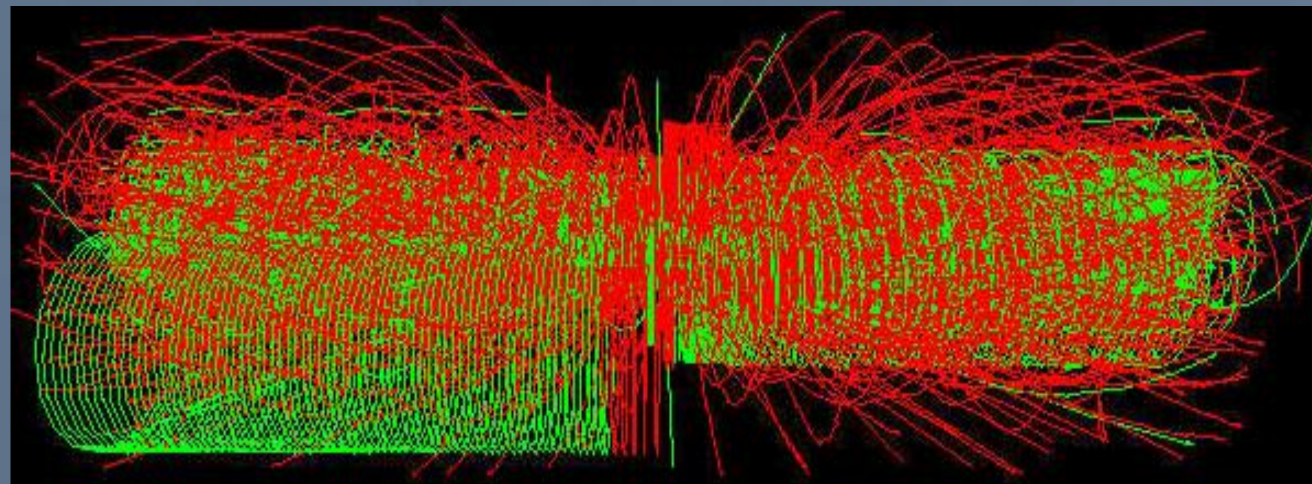
1 CHF ~ 1 USD

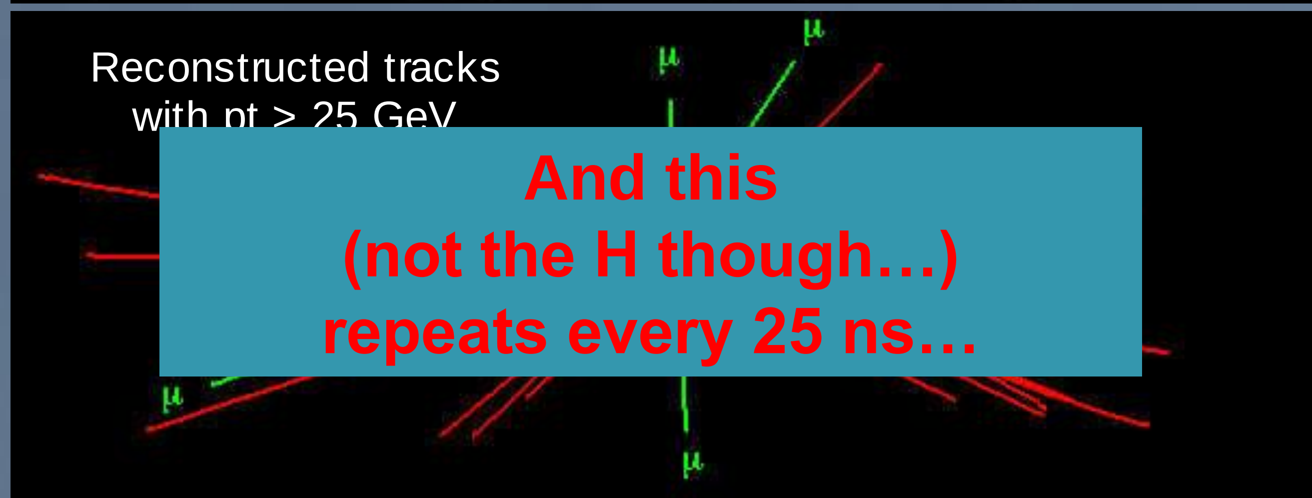
What do Events Look Like?



The needle in the hay-stack: pp Collisions at 14 TeV at 10^{34} $cm^{-2}s^{-1}$

- $\sigma(pp)$ = 70 mb --> >7 x 10^8 /s (!)
- In ATLAS and CMS* 20 – 30 min bias events overlap
- $H \to ZZ$
  $Z \to \mu\mu$
  $H \to$ 4 muons: the cleanest ("golden") signature





And this
(not the H though…)
repeats every 25 ns…

Data Rates @ LHC today

- Particle beams cross every 25 ns (40 MHz)
  - Up to 30 particle collisions per beam crossing (until 2023)
  - Up to 10^9 p-p collisions per second
- Two filter / trigger levels
  - Level 1 (custom hardware) sees everything
  - High Level Trigger (software) sees a fraction
  - Together they reduce 10^9 p-p collisions per second to O(1000)
- Raw data to be stored permanently: >15 PB/year

Physics Process	Events/s
Inelastic p-p scattering	10^8
b	10^6
$W \rightarrow e\upsilon$; $W \rightarrow \mu\upsilon$; $W \rightarrow \tau\upsilon$	20
$Z \rightarrow ee$; $Z \rightarrow \mu\mu$; $Z \rightarrow \tau\tau$	2
t	1
Higgs boson (all; m_H = 120GeV)	0.04
Higgs boson (simple signatures)	0.0003
Black Hole (certain properties)	0.0001

Stage	Incoming data rate	Outgoing data rate	Reduction factor
1) Level1 Trigger (custom hardware)	40000000 s^{-1}	10^5 – 10^6 s^{-1}	400-10,000
2) High Level Trigger (software on server farms)	2000-1000000 s^{-1}	1000 -10000 s^{-1}	10-2000

Future data rates @ LHC (trigger stage 2)

	Event-size [kB]	Rate [kHz]	Bandwidth [Gb/s]	Year [CE]
ALICE	20000	50	8000	2019
ATLAS	4000	200	6400	2023
CMS	4000	1000	32000	2023
LHCb	100	40000	32000	2019

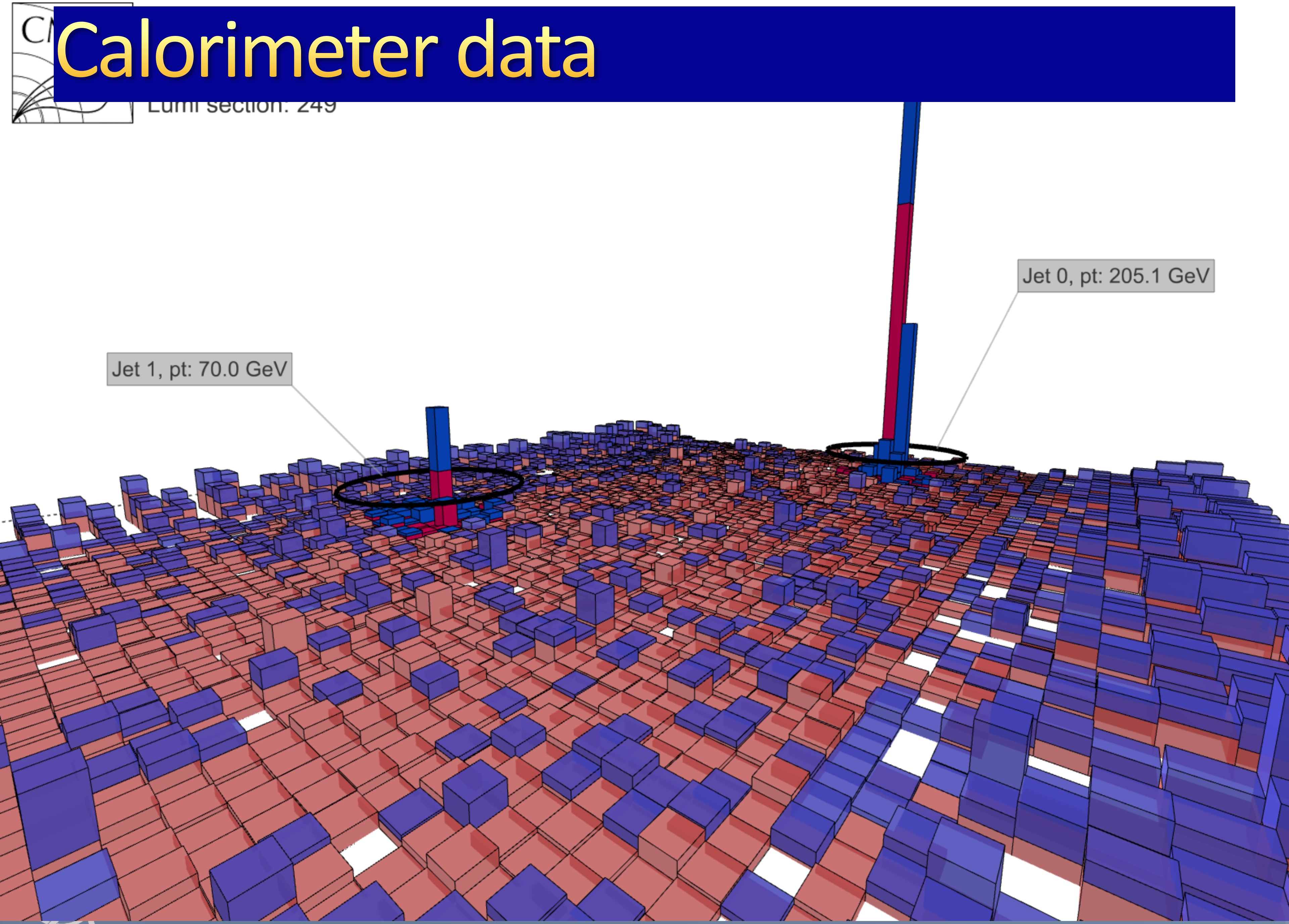
40000 kHz == collision rate
→ *LHCb abandons Level 1 for an all-software trigger*
O(100) Tbit/s networks required

“Use-cases” in Online

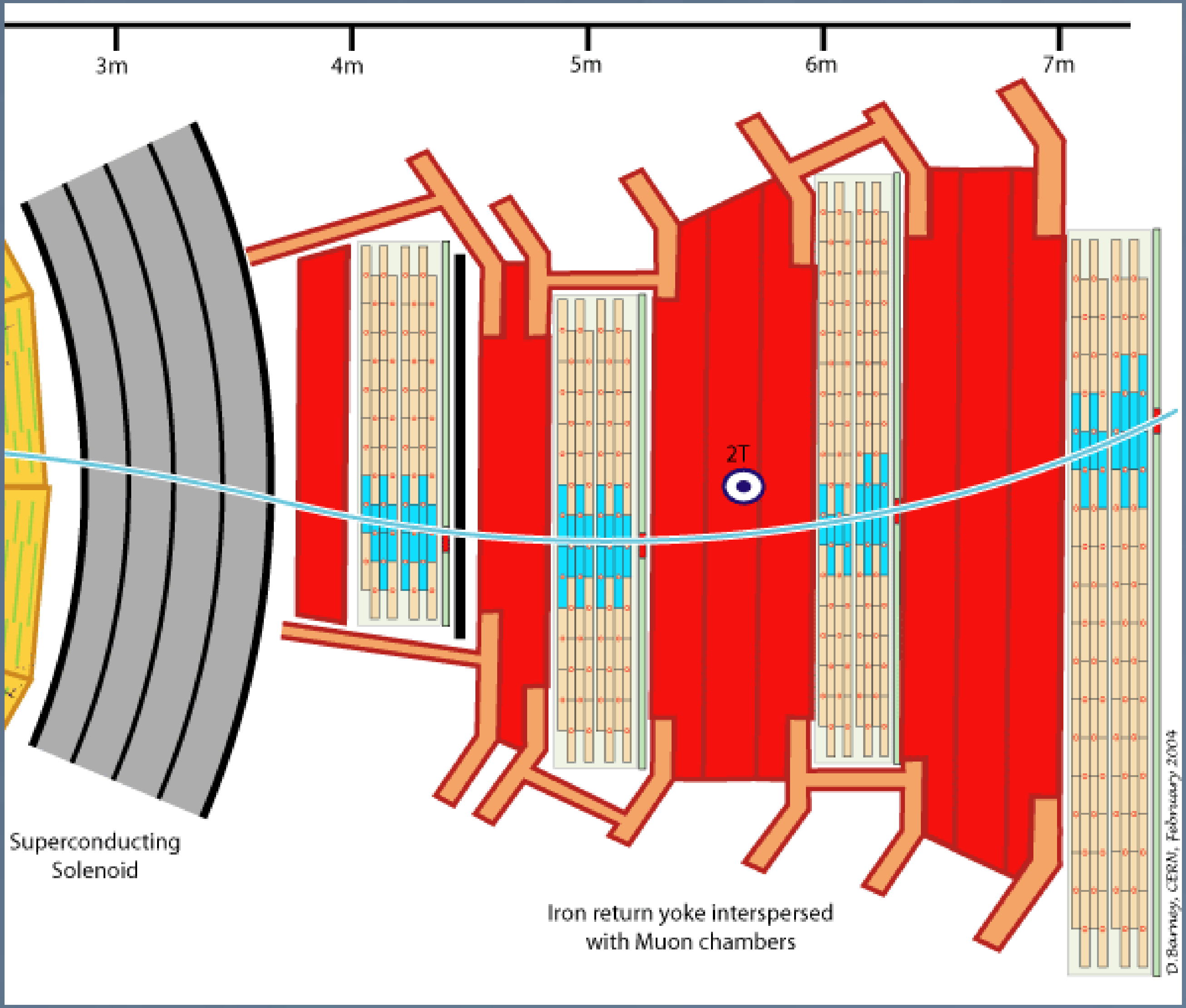
1) “Level-1” using (more/all) COTS hardware
2) “Data Acquisition”
3) “High Level Trigger”
4) “Controls”
5) “Online data processing”
6) “Exporting data”
7) “Online data storage”

Use-case #1
The first level trigger

Calorimeter data



Finding Muons (2d view)



Level 1 Trigger today

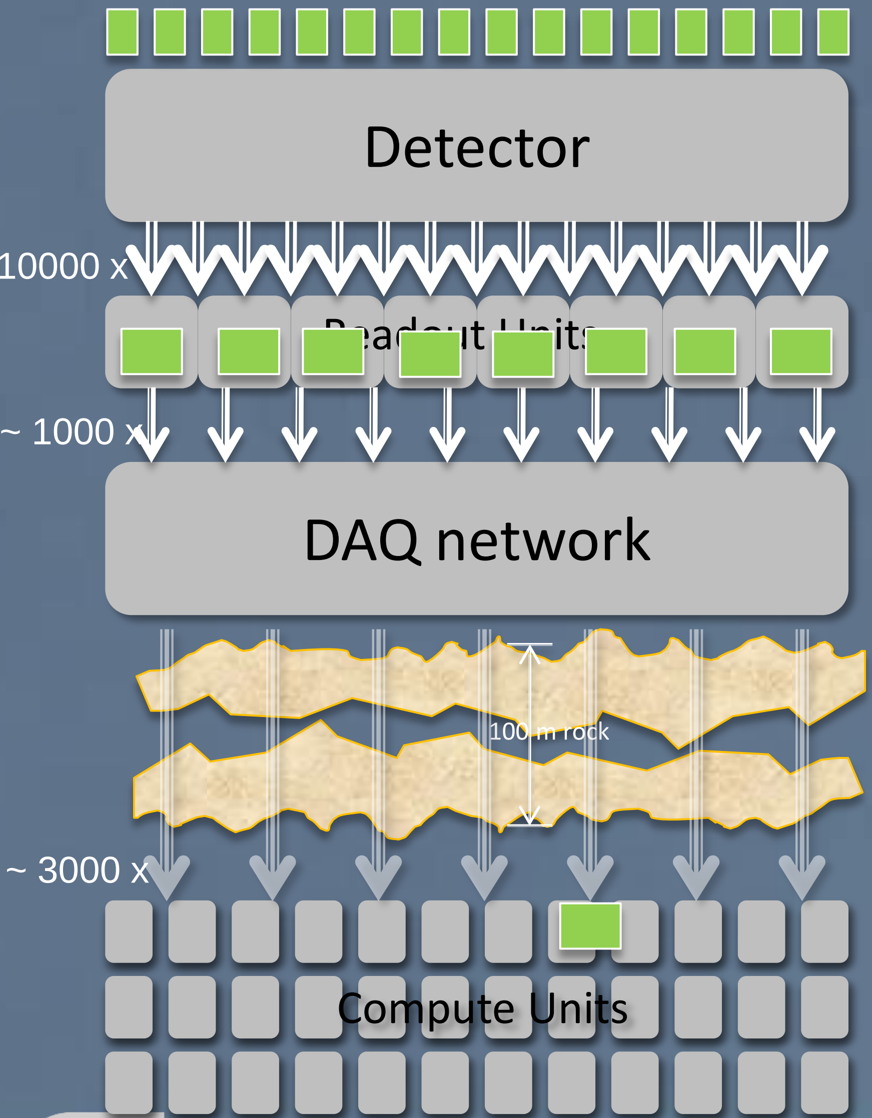
- The Level 1 Trigger is implemented in hardware: FPGAs and ASICs → difficult / expensive to upgrade or change, maintenance by experts only
- Decision time: ~ a small number of microseconds
- It uses simple, hardware-friendly signatures → looses interesting collisions
- Each sub-detector has its own solution, only the uplink is standardized →

Level 1 challenge

- Can we do this in software? Using GPGPUs / XeonPhis?
- We need low and near– deterministic latency
- Need an efficient interface to detector-hardware: CPU/FPGA hybrid?

Use-case #2 and #6
Data Acquisition and Export
→ fast data transport

Data Acquisition (generic example)



Every Readout Unit has a piece of the collision data
All pieces must be brought together into a single compute unit
The Compute Unit runs the software filtering (High Level Trigger – HLT)

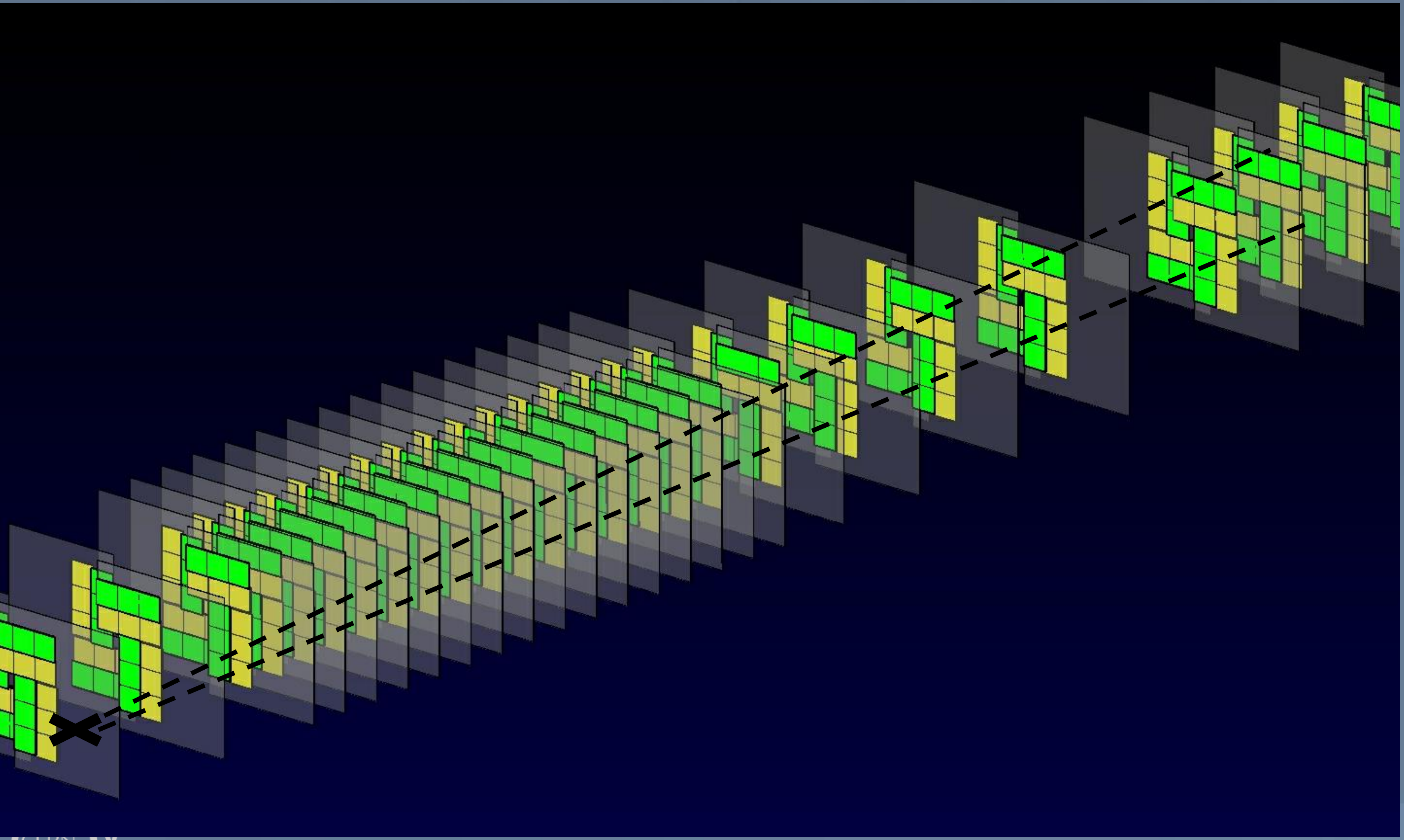
- custom radiation- hard link from the detector
- DAQ ("event-building") links – some LAN (10/40/100 GbE / InfiniBand)
- Links into compute-units: typically 1/10 Gbit/s

Data acquisition challenge

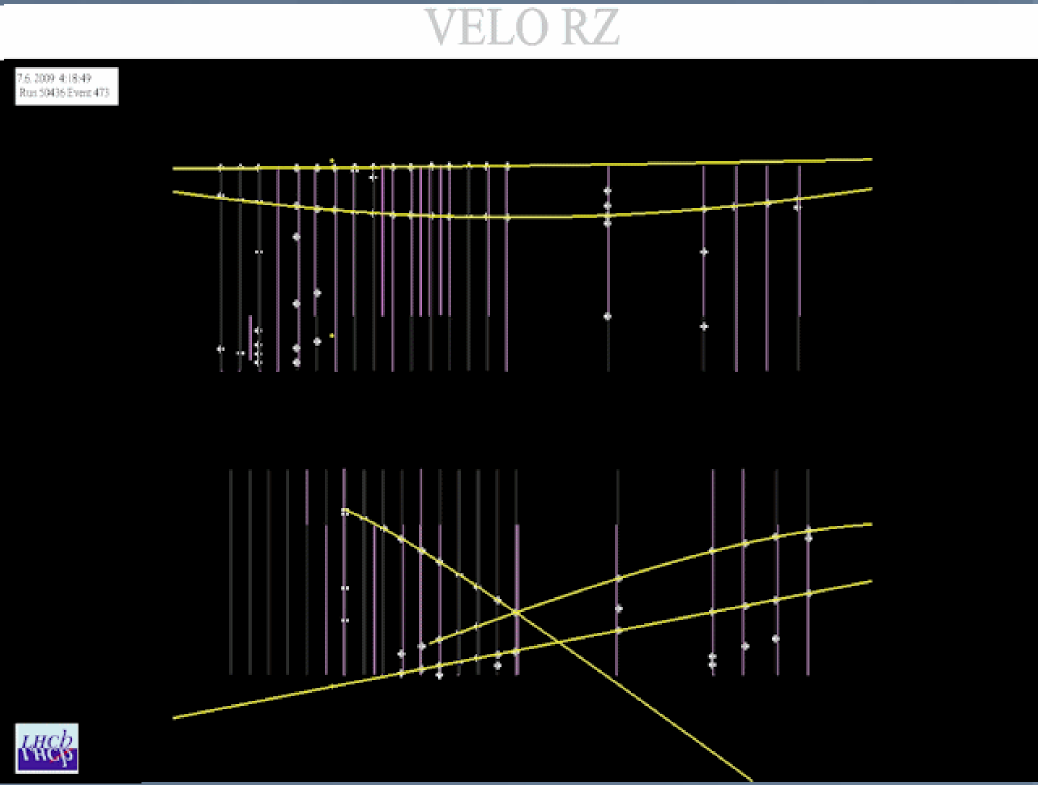
- Transport large amount of data (multiple Terabit/s @ LHC) reliably and cost-effectively
- Integrate the network closely and efficiently with compute resources (be they classical CPU or "many-core")
- Multiple network technologies should seamlessly co-exist in the same integrated fabric ("the right link for the right task"), end-to-end solution from online processing to scientist laptop (e.g. light-sources)

Use-case #3 and #5
Online Data processing

Pattern finding - tracks



Same in 2 dimensions



- Can be much more complicated: lots of tracks / rings, curved / spiral trajectories, spurious measurements and various other imperfections

LHC High Level Trigger: Key Figures

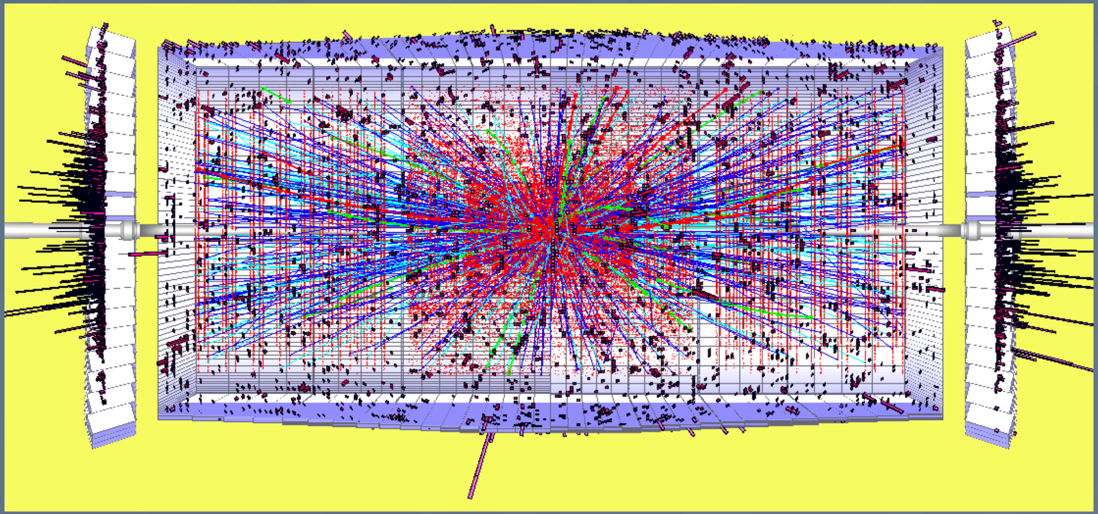
- Existing code base: 5 MLOC of mostly C++
- Almost all algorithms are single-threaded (only few exceptions)
- Currently processing time on a X5650 per event: several 10 ms / process (hyper-thread)
- Currently between 100k and 1 million events per second are filtered online in each of the 4 LHC experiments

Online Data processing challenge

- Make the code-base ready for multi/many-core (this is not Online specific!)
- Optimize the online processing compute in terms of cost, power, cooling
- Find the best architecture integrating “standard servers”, many-core systems and a high-bandwidth network

More material

A Track-Trigger at 40 MHz 2020++



- Goals:
  - Resolve up to 200÷250 collisions per bunch crossing
  - Keep latency within ~ 6 µs
- L1 tracking trigger data combined with calorimeter & muon trigger data
  - With finer granularity than presently employed.
- Physics objects made from tracking, calorimeter & muon trigger data transmitted to Global Trigger.

Design principles for data networks

- Minimize number of expensive “core” network ports
- Use the most efficient technology for a given connection
  - different technologies should be able to co-exist (e.g. fast for building, slow for end-node)
  - keep distances short
- Exploit the economy of scale → try to do what everybody does (but smarter ☺)

LHC Data Acquisition Key Figures – Example LHCb

- Minimum required bandwidth: > 32 TBit/s
- Number of 100 Gbit/s links: > 320
- Number of compute units: > 4000
- Event size: 100 kB
- Number of events per seconds: 10 – 40 Millions per second
- Number of events retained for permanent storage: 20k – 30k per second

High Level Trigger compared to HPC

- Like HPC:
  - full ownership of the entire installation → can choose architecture and hardware components
  - single "client" / "customer"
  - have a high-bandwidth interconnect
- Unlike HPC:
  - many independent small tasks which execute quickly → no need for check-pointing (fast storage)
    → no need for low latency
  - data driven, i.e. when the LHC is **not** running (70% of the time) the farm is idle → interesting ways around this (deferral, "offline usage)
  - facility is very long-lived, growing incrementally