15th International Workshop on
advanced computing and analysis techniques in physics

May 16-21, 2013, Beijing, China
http://acat2013.ihep.ac.cn

The Evolution of the Trigger and Data Acquisition System in the ATLAS Experiment

Attila Krasznahorkay
on behalf of the ATLAS Collaboration

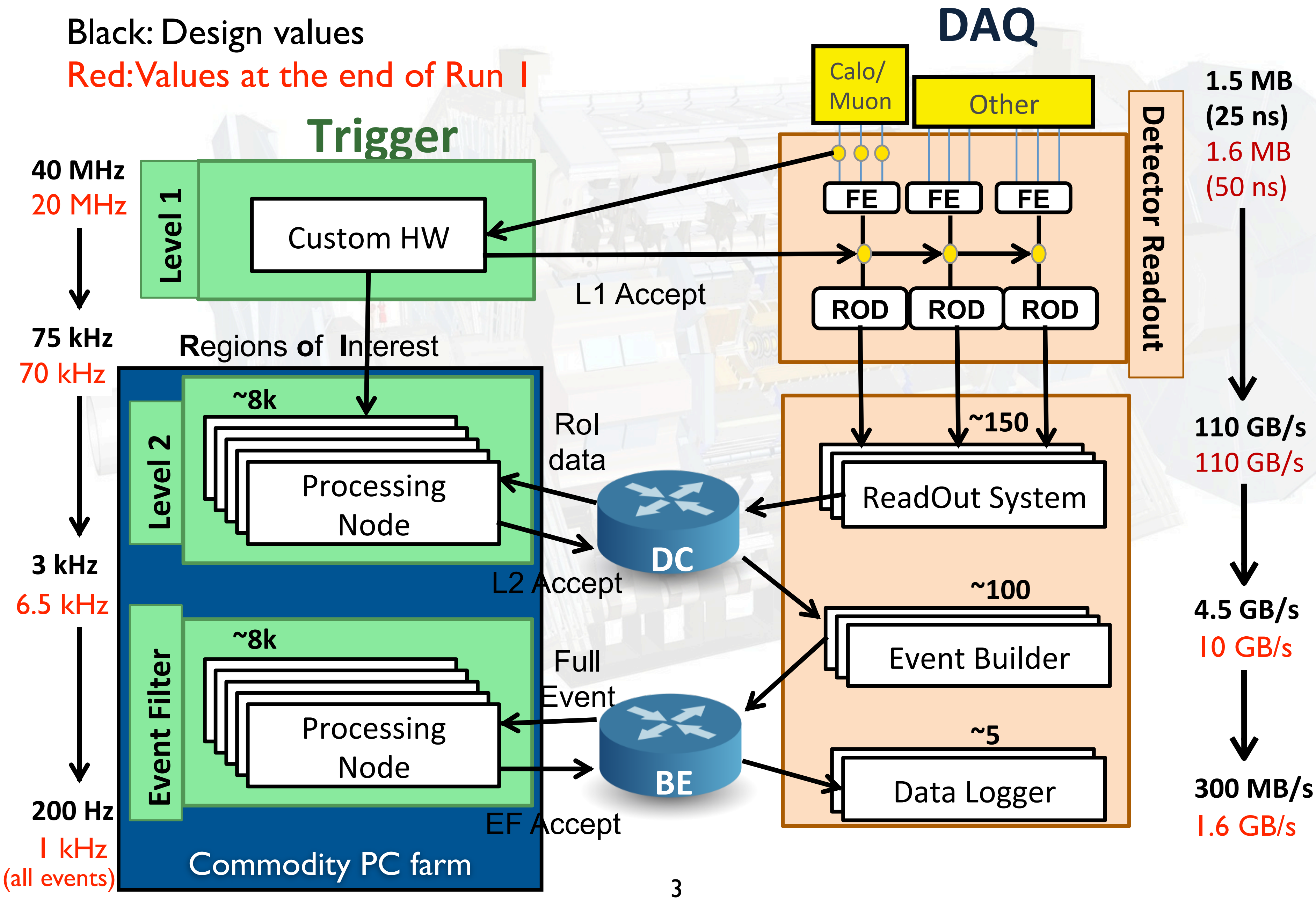
ATLAS EXPERIMENT
http://atlas.ch

NEW YORK UNIVERSITY

Overview

- ATLAS T/DAQ design used in Run 1
    - Achievements and limitations of the existing system
- Changes/additions to the hardware
    - Description of planned network changes
    - Introduction of the new hardware
        - Level-1 Topological Trigger
        - Fast Tracking hardware
        - Upgraded **R**ead **O**ut **S**ystem nodes
- Planned changes to the High Level Trigger
- Preliminary results for the new design with demonstrator code

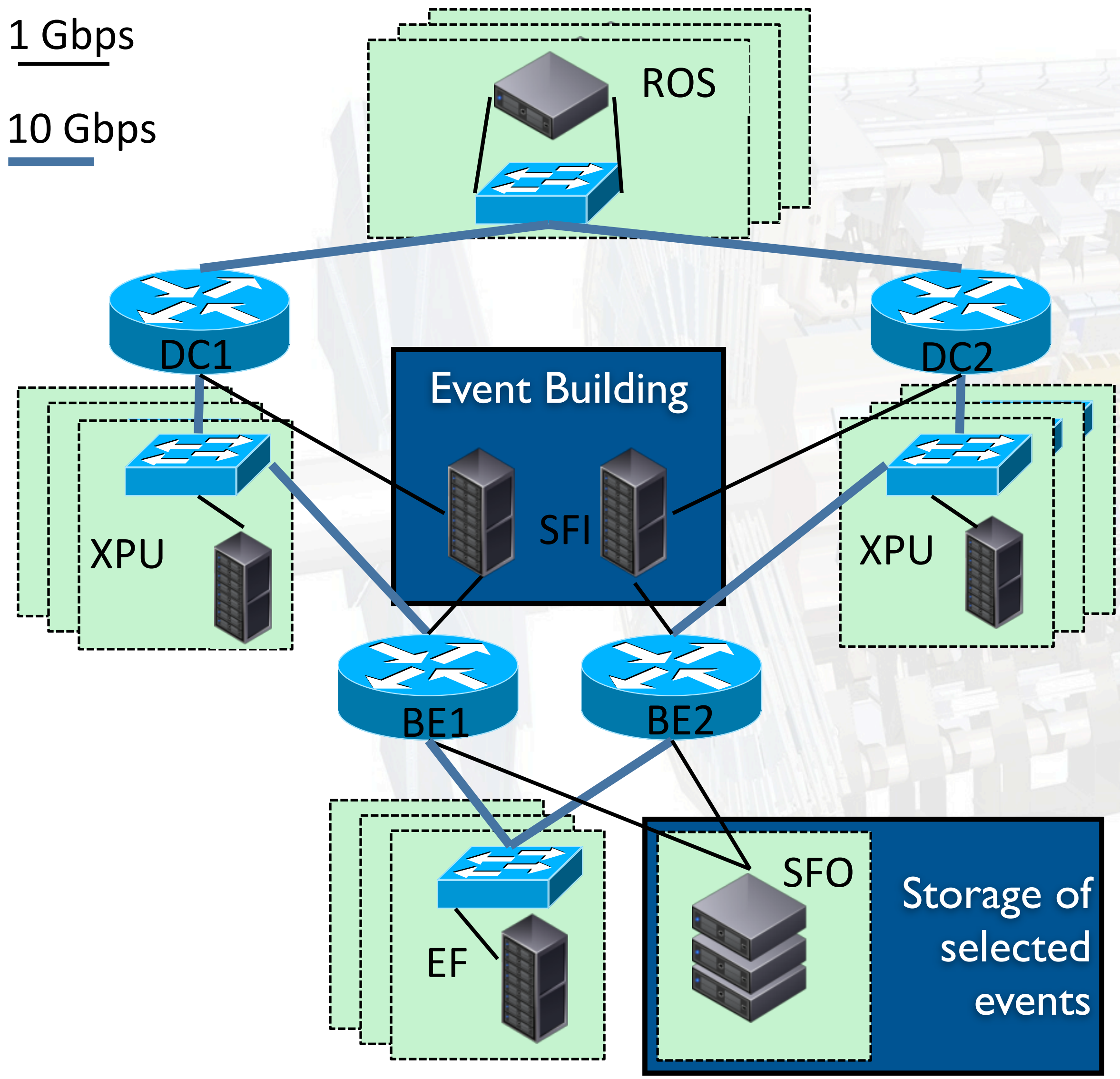
ATLAS T/DAQ in Run I



ATLAS T/DAQ in Run 1

- Trigger system is separated into 3 levels
  - ***Level-1:*** Custom built hardware selecting the interesting events, and providing information about the locations of the objects identified (**R**egions **o**f **I**nterest)
  - ***Level-2:*** Mostly dedicated trigger algorithms running on commodity PC hardware, making decisions about the events by only accessing parts of the full event
  - ***Event Filter:*** A mixture of trigger specific and offline reconstruction algorithms selecting interesting, fully-built events. Receives input both from Level-1 and Level-2.
- Managed to use the architecture well beyond the design specifications
  - Changed physics requirements necessitated higher selected event rates
  - Does not scale well to the specifications of Run 2

DAQ Network in Run I



- Two separate data networks coupled by event building to separate different data traffic
  - **D**ata **C**ollection (**R**ead **O**ut **S**ystem fragments)
  - **B**ack-**E**nd (fully built events)
- Level-2 (XPU nodes) and EF nodes strictly separated on the network
- Request rates to ROS farm posed the most stringent limits on triggering rate by the end of Run I
  - Despite of the system performing beyond the original specifications

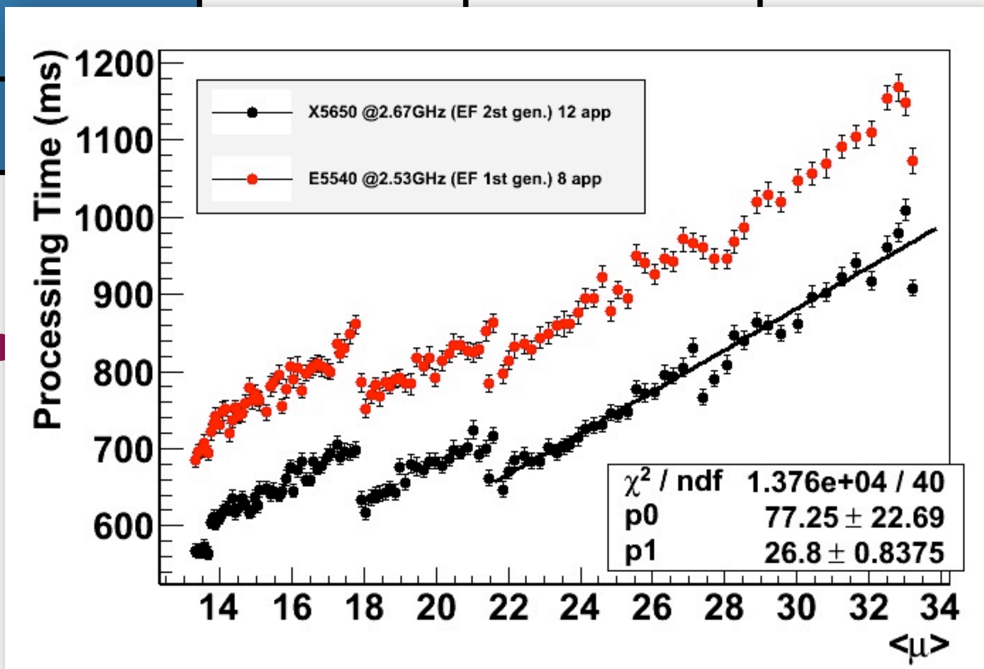
Run 2 vs. Run 1

Parameter	E_{cm} [TeV]	# of bunches	Bunch separation [ns]	Peak lumi. [$cm^{-2}s^{-1}$]	Pile-up	Event size
Run 1	8	1380	50	~$7x10^{33}$	~35	~1.5 MB
Run 2	~13	1380 - 2700	25 - 50	$1.5\text{-}2x10^{34}$	up to 80	1.7 - 2.1 MB

- (Some) Physics goals:
  - Electroweak searches
    - High efficiency lepton triggers
  - Precision Higgs measurements
    - Dedicated, high efficiency triggers for each production and decay channel
- To achieve this:
  - Speed up the code
  - Make the trigger selection more similar to the offline one (simplify analysis)
  - Introduce more specialized, more selective triggers

Run 2 vs. Run 1

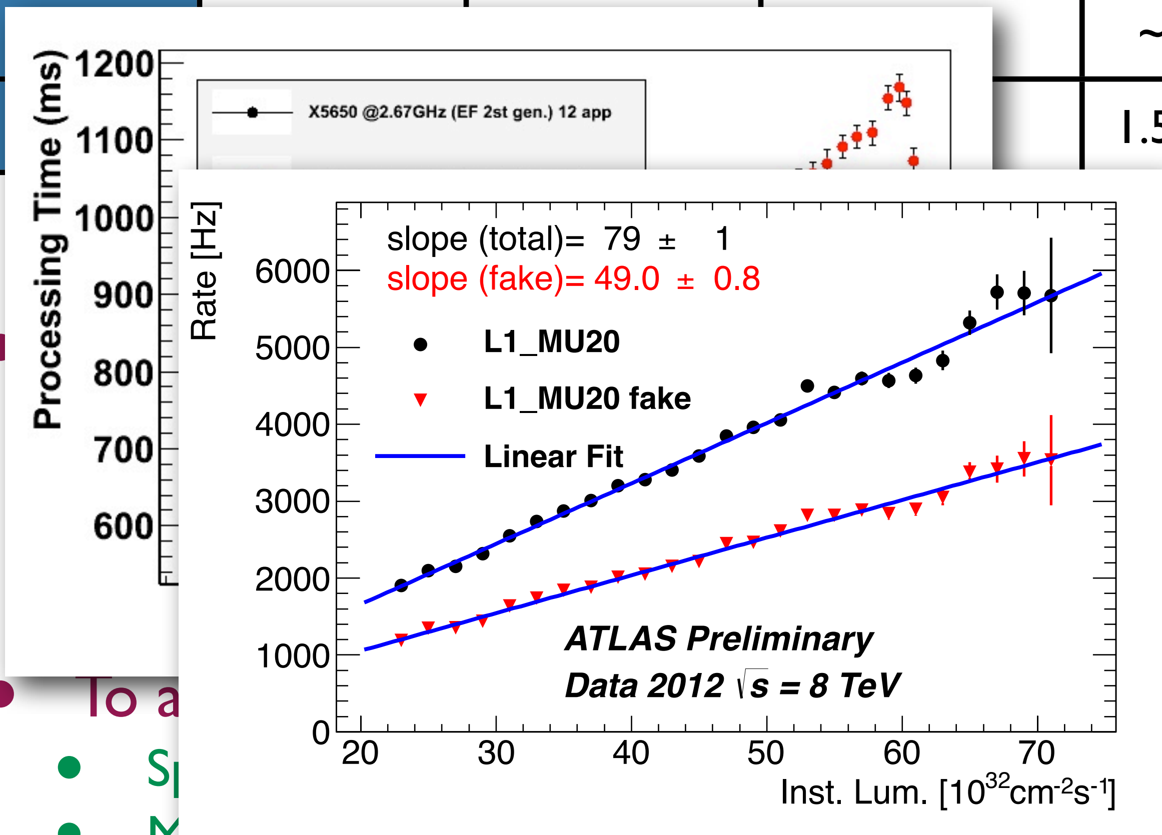
Parameter	E_{cm} [TeV]	# of bunches	Bunch separation [ns]	Peak lumi. [$cm^{-2}s^{-1}$]	Pile-up	Event size
				~$7x10^{33}$	~35	~1.5 MB
				1.5-$2x10^{34}$	up to 80	1.7 - 2.1 MB



- r each production and decay channel
- To achieve this:
  - Speed up the code
  - Make the trigger selection more similar to the offline one (simplify analysis)
  - Introduce more specialized, more selective triggers

Run 2 vs. Run 1

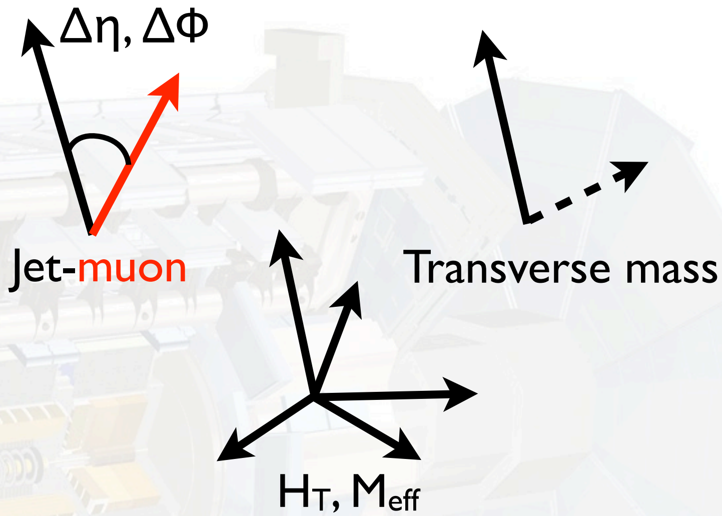
Parameter	E_{cm} [TeV]	# of bunches	Bunch separation [ns]	Peak lumi. [$cm^{-2}s^{-1}$]	Pile-up	Event size
				$\sim 7\times10^{33}$	~35	~1.5 MB
				$1.5\text{-}2\times10^{34}$	up to 80	1.7 - 2.1 MB



- ...roduction and decay channel
- To a...
    - Sp...
    - Make the trigger selection more similar to the offline one (simplify analysis)
    - Introduce more specialized, more selective triggers

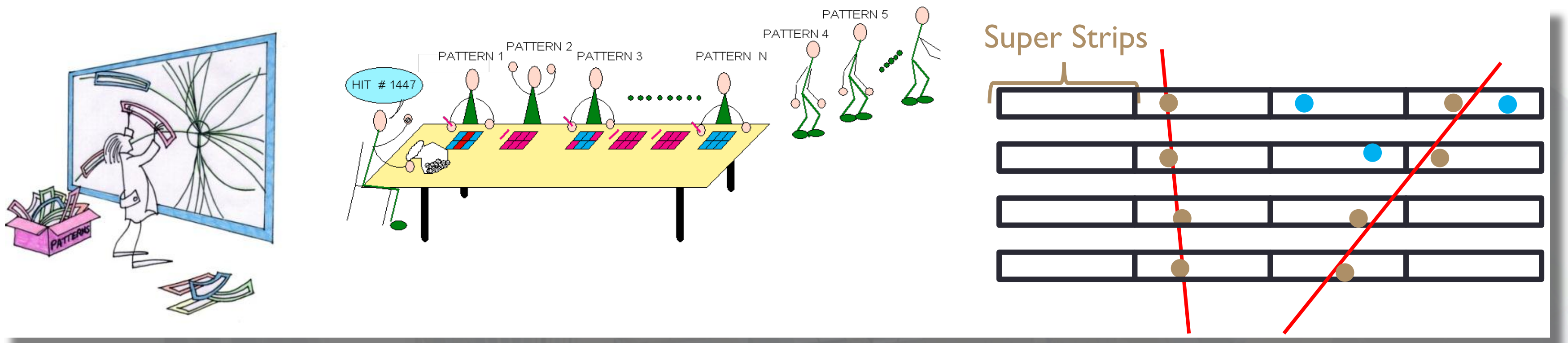
LVL1 Topological Trigger

- Introducing a new element in the Level-1 trigger hardware
  - Will receive detailed information about the object candidates found by the Level-1 calorimeter and muon trigger
  - Algorithms implemented in its FPGA will make a decision based on “complicated” kinematic requirements



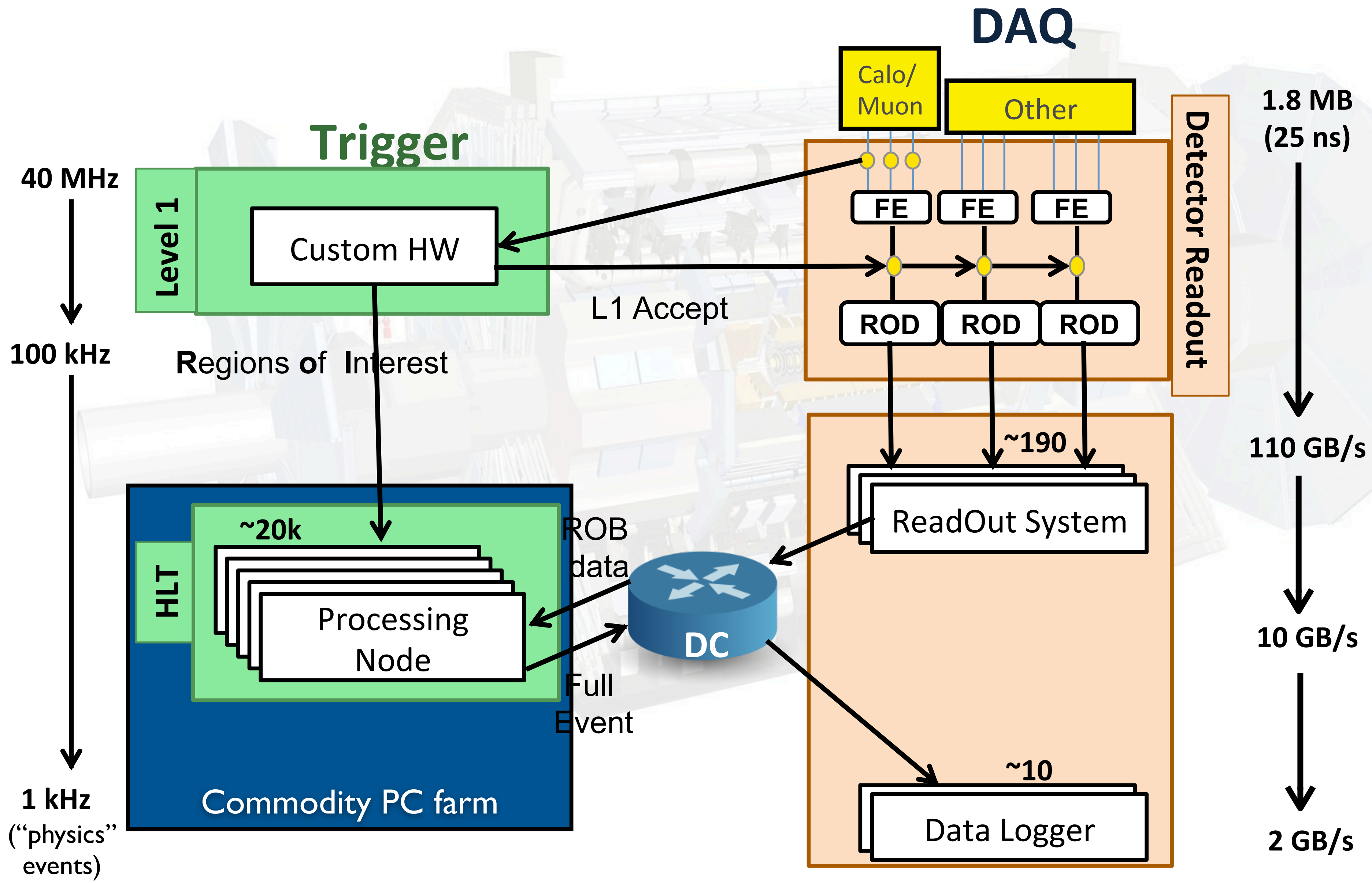
- Hardware will only be able to output information about which algorithms passed, not on which input object combinations were found to pass them
  - Information would be too heavy for the 100 kHz readout
- *HLT* will need detailed information on which *RoIs* to seed the topological trigger chains with
  - *L1Topo* simulation will be run in the *HLT* to reconstruct this info
- More detailed information about the hardware is available in ATL-DAQ-PROC-2012-041

Fast Tracking (FTK)



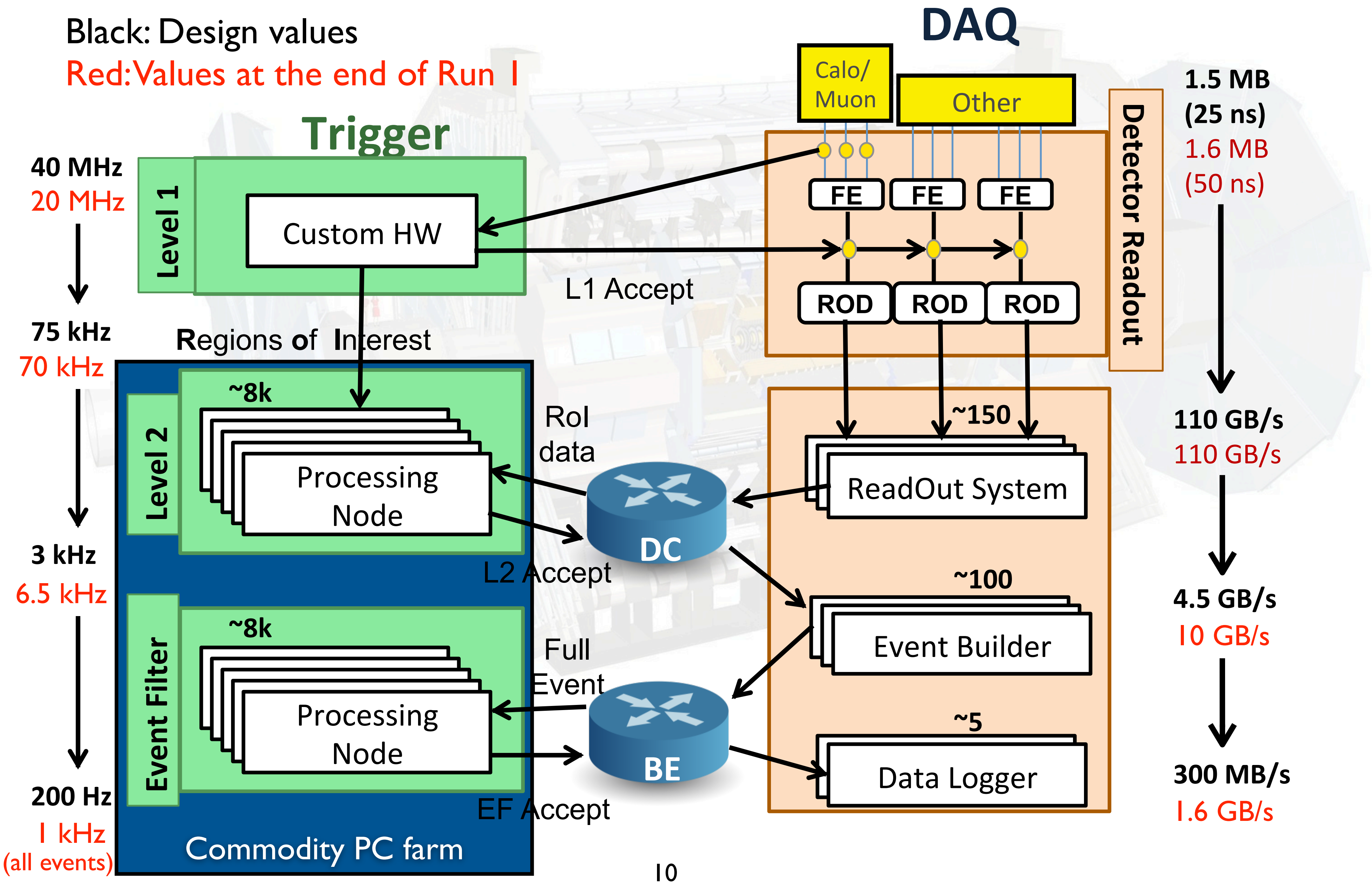
- Special hardware will receive a copy of all silicon data at each Level-1 accept (up to 100 kHz)
  - Can reconstruct all tracks in the event using pattern matching on specially pre-processed data in an associative memory
- Useful for most trigger signatures, by providing tracking information for the whole detector very quickly
  - Seed track finding in the HLT for more precise measurements in selected regions
  - Can for instance improve b-tagging and tau reconstruction
- More details about the hardware are available in ATL-DAQ-PROC-2012-020

T/DAQ Design: Merged L2&EF

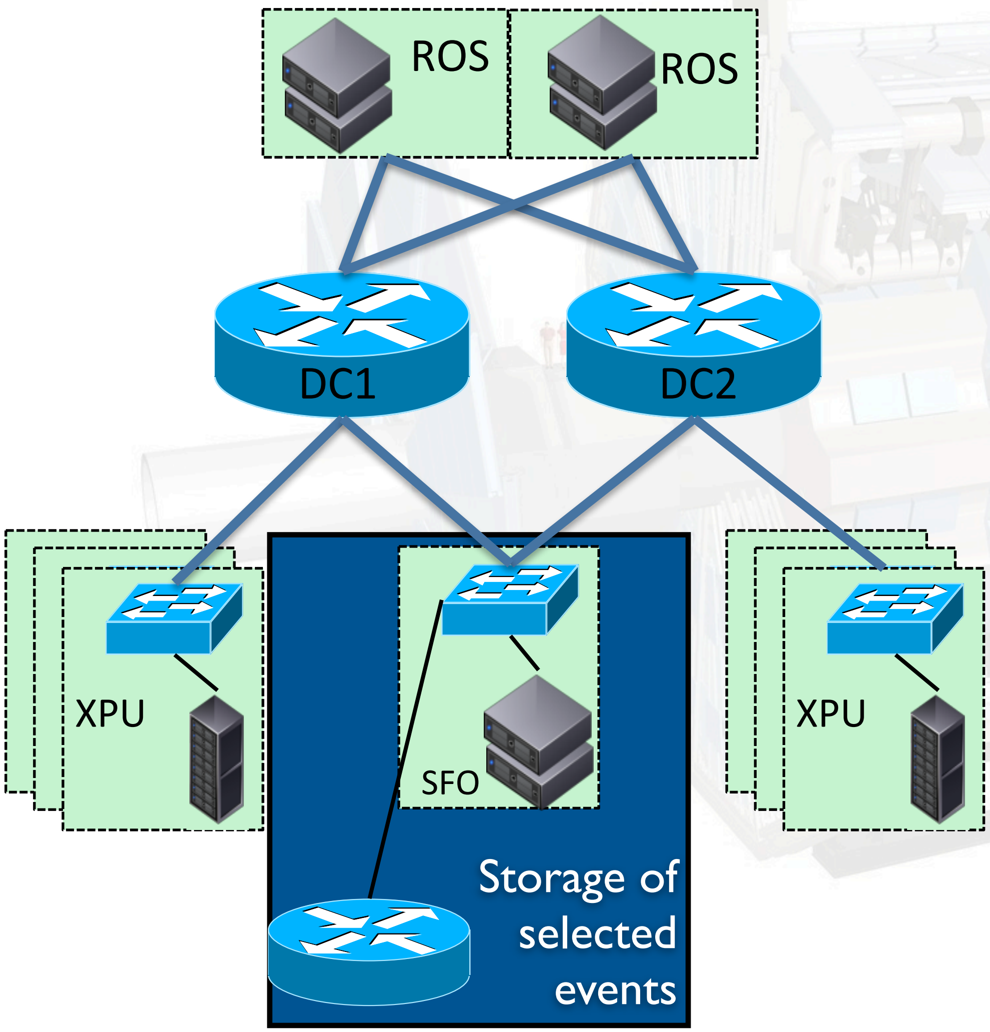


Current Design Reminder

Black: Design values
Red: Values at the end of Run 1

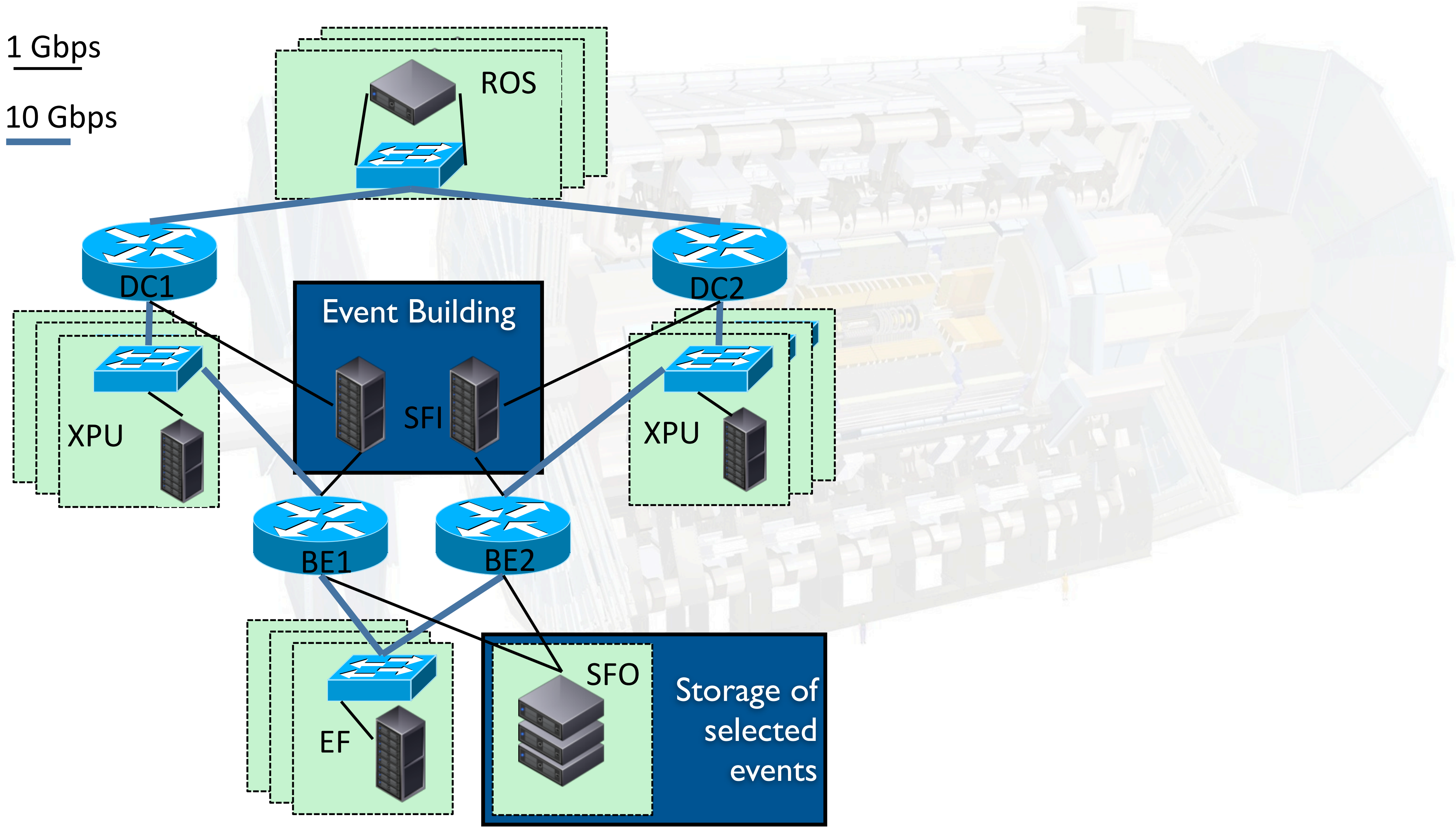


DAQ Network at Run 2 Start

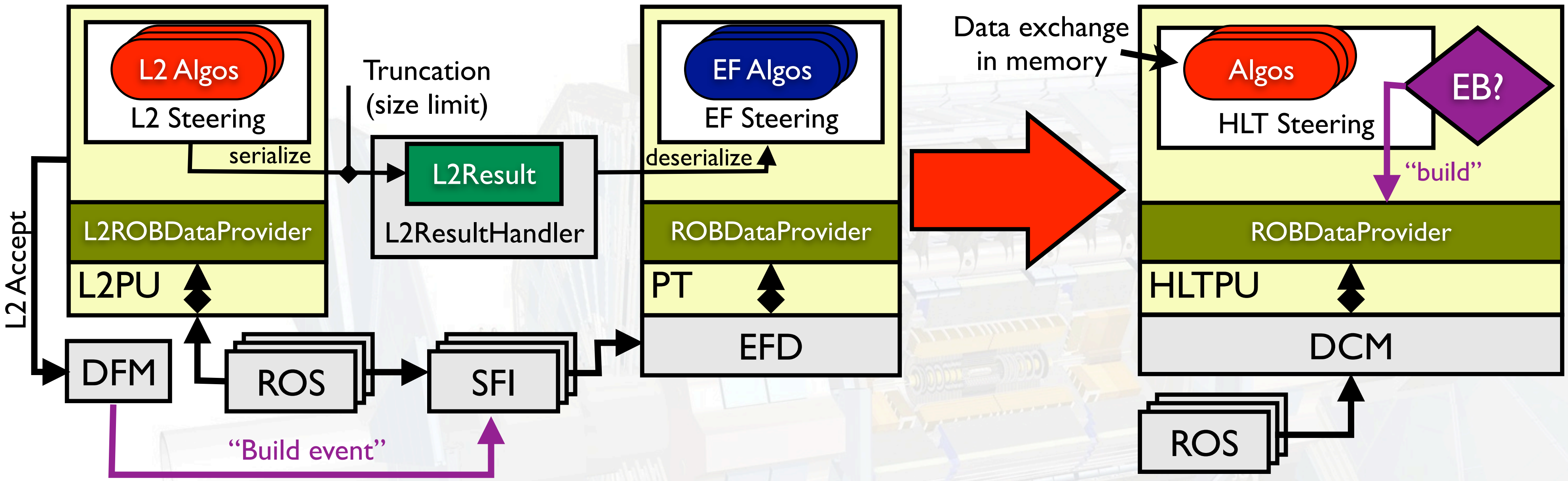


- Single data network
- Separate event building farm no longer used
- Final setup will contain up to 4 **D**ata **C**ollection networks distributing the *ROS* –> *HLT* processing node (*XPU*) traffic better
  - Purchases will be made according to the LHC ramp-up in luminosity
  - Design much more scalable to meet observed requirements
- *XPU* and *SFO* farms much more easily scalable than in current system
  - The plan is to extend the network during Run 2 as the need arises

Current Network Reminder



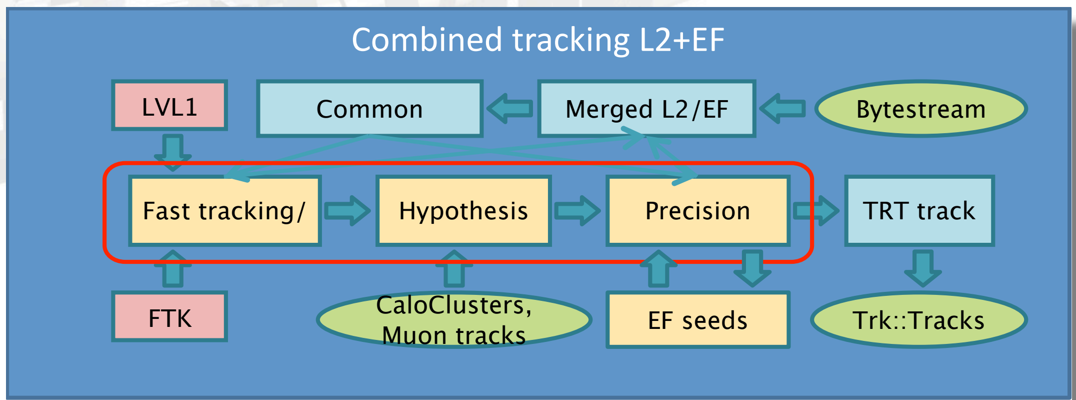
HLT Re-Design: Merged L2&EF



- Execute former Level-2 and EventFilter algorithms in one process
  - ROS data requested by L2 algorithms need not be requested again for EF: reduces ROS requests by few %
  - Limited network transfer of data from L2 to EF algorithms replaced by unlimited in-memory transfer: allows tighter coupling and reduces CPU & network use
  - Opportunity for re-optimization of algorithms (see next slide)
- Flexible event building point depending on trigger chains
  - Optimum should be found for final trigger menu
  - New reconstruction strategies being explored

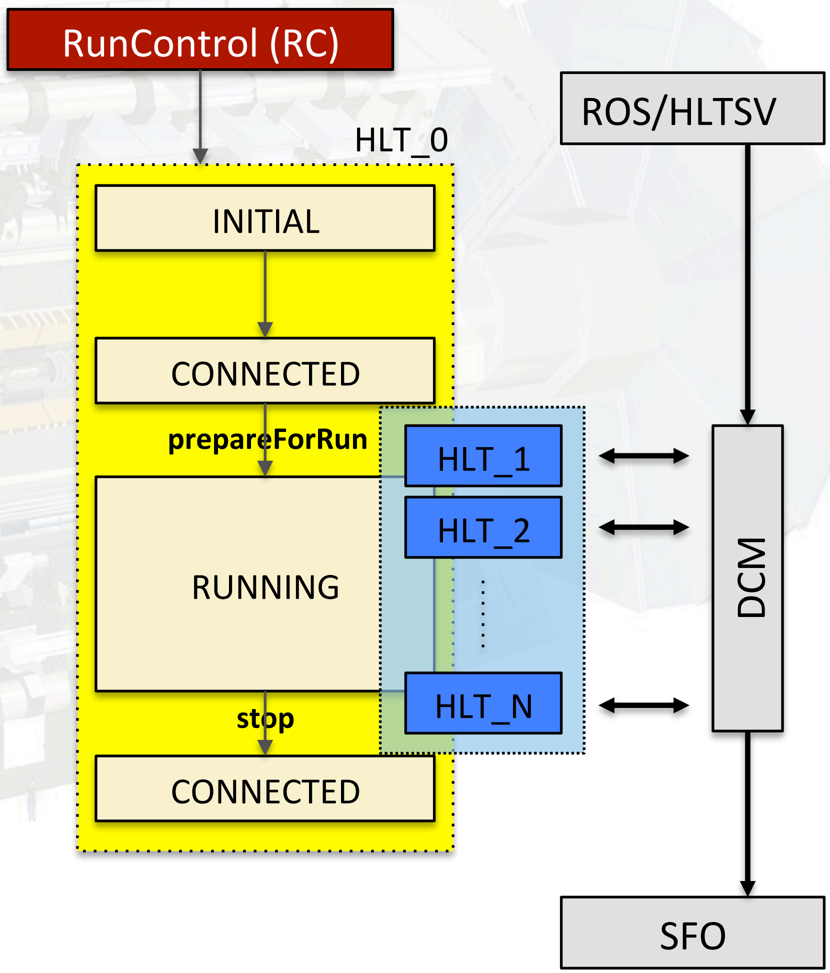
Reconstruction in the Merged HLT

- No hardware limit on algorithm seeding
  - Objects reconstructed by a previous step can be fully re-used in later steps
- No need to do the same sort of calculations twice in the two levels
  - e.g. spacepoint formation for the tracking
- Much room for development
  - For example, much work going into re-organizing the HLT track reconstruction
  - In current system LVL2 and EF has to re-do a lot of things
  - In new system much more possible to be re-used

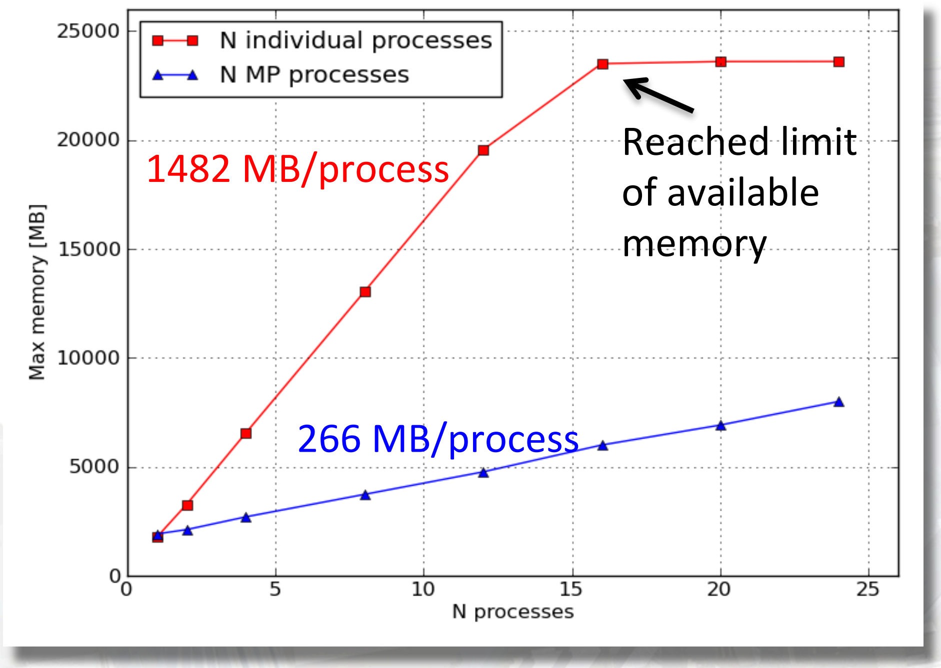


Multi-Process HLT

- Servers get more and more CPU cores, but memory can't be increased linearly
  - Industry moving towards massively parallel processes
- *In the long term:* Will have to re-design the HLT code together with the offline code to make it much more parallel
- *In the short term:* Run many separate processes as before, but share as much memory between them as possible
  - Only event-wise data is different between processes. Geometry, conditions, magnetic field, etc. are the same.
  - Prototype simply uses the Linux kernel's Copy-On-Write feature. Forking the worker processes after main process got configured.



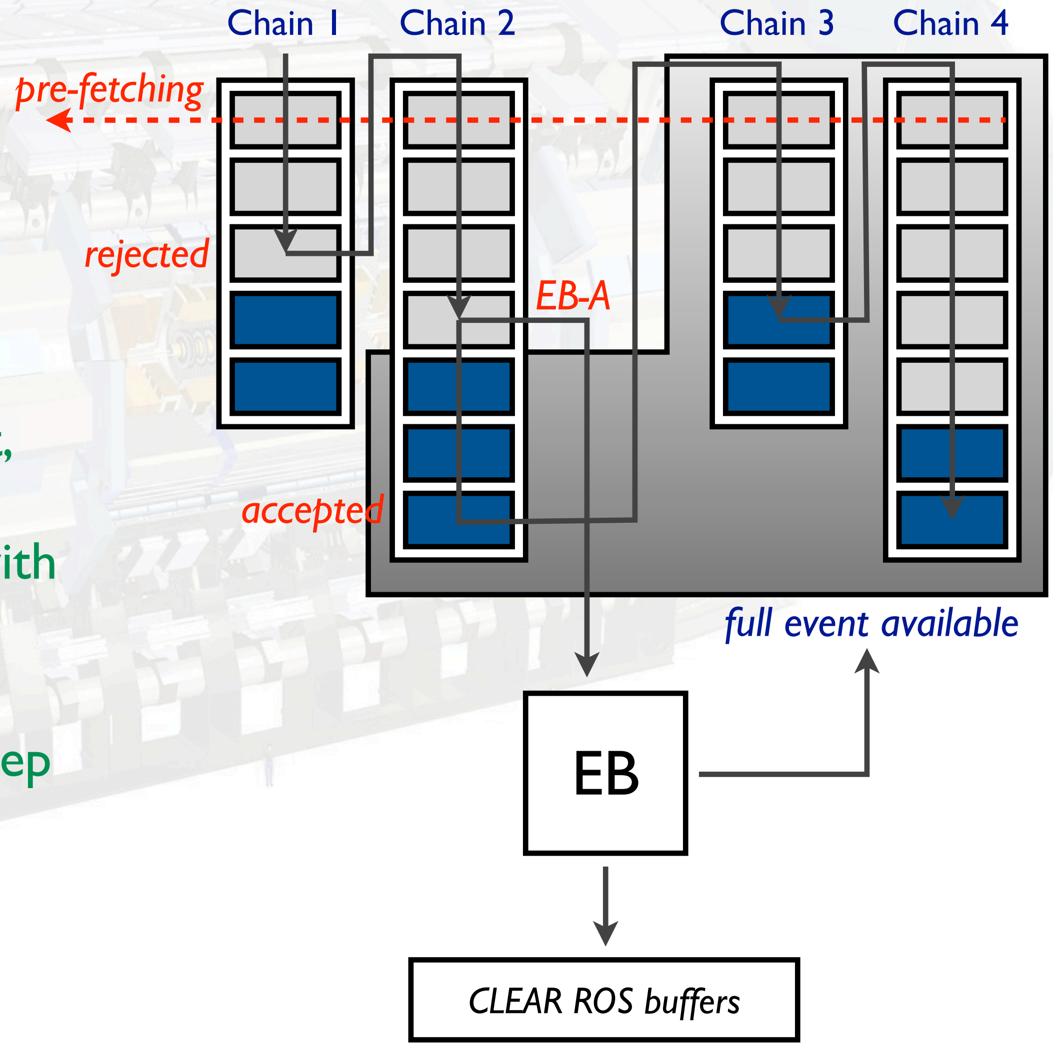
Multi-Process HLT



- Prototype providing much reduction in the per-process memory requirements
  - ~1500 MB/process -> ~300 MB/process
- System can easily coexist with future multi-threaded developments

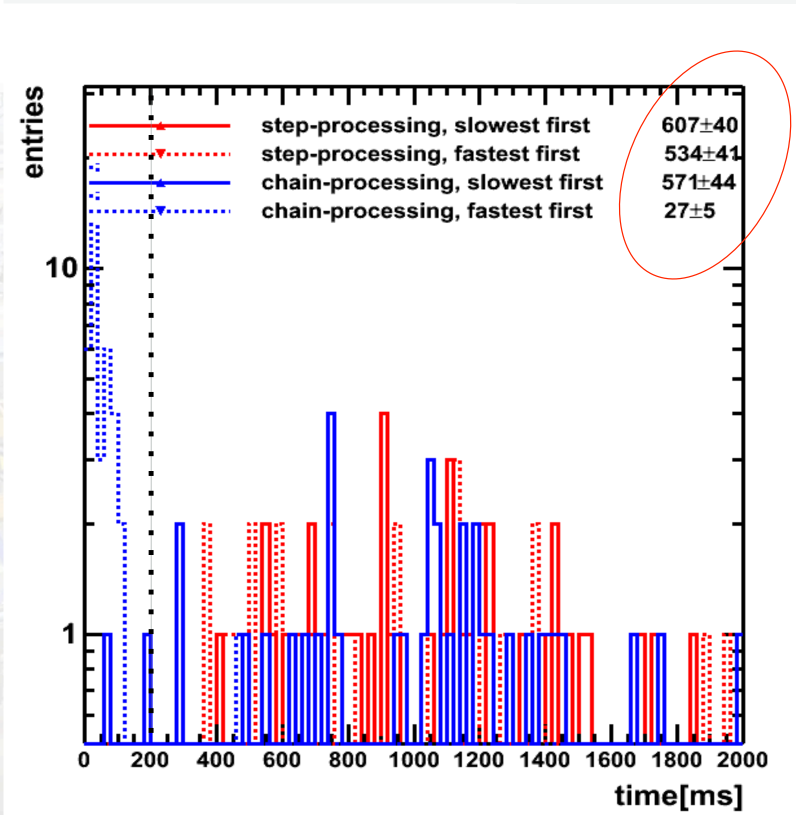
Alternative EB Strategies

- Technique used until now: Step-wise processing
  - The first step of each chain is executed one after the other, then the second step of the still active chains is executed, etc.
- Possible alternative approach: Process chains one by one
  - When first chain gets to *EB* Accept, build the event
  - Rest of the chains are processed with the full event in memory
- Needs changes to pre-fetching:
  - Now: Pre-fetch data before each step
  - New system: Fetch all data for the chain before the first step
- Non-trivial to optimize (CPU time, time to ROS clear, ROS access pattern, etc.)



Chain-wise Processing

- Average time needed to issue event building signal depends strongly on the chain ordering
- General rule is to first execute the chains that are
  - most likely to accept the event
  - the fastest
- Would need automatic, reproducible ordering
  - Complicated by overlaps
  - Could base on some new chain attributes like slow/fast set by hand
  - Or automatic optimization based on a recent run
- Just an example of the studies happening, not necessarily what we will use

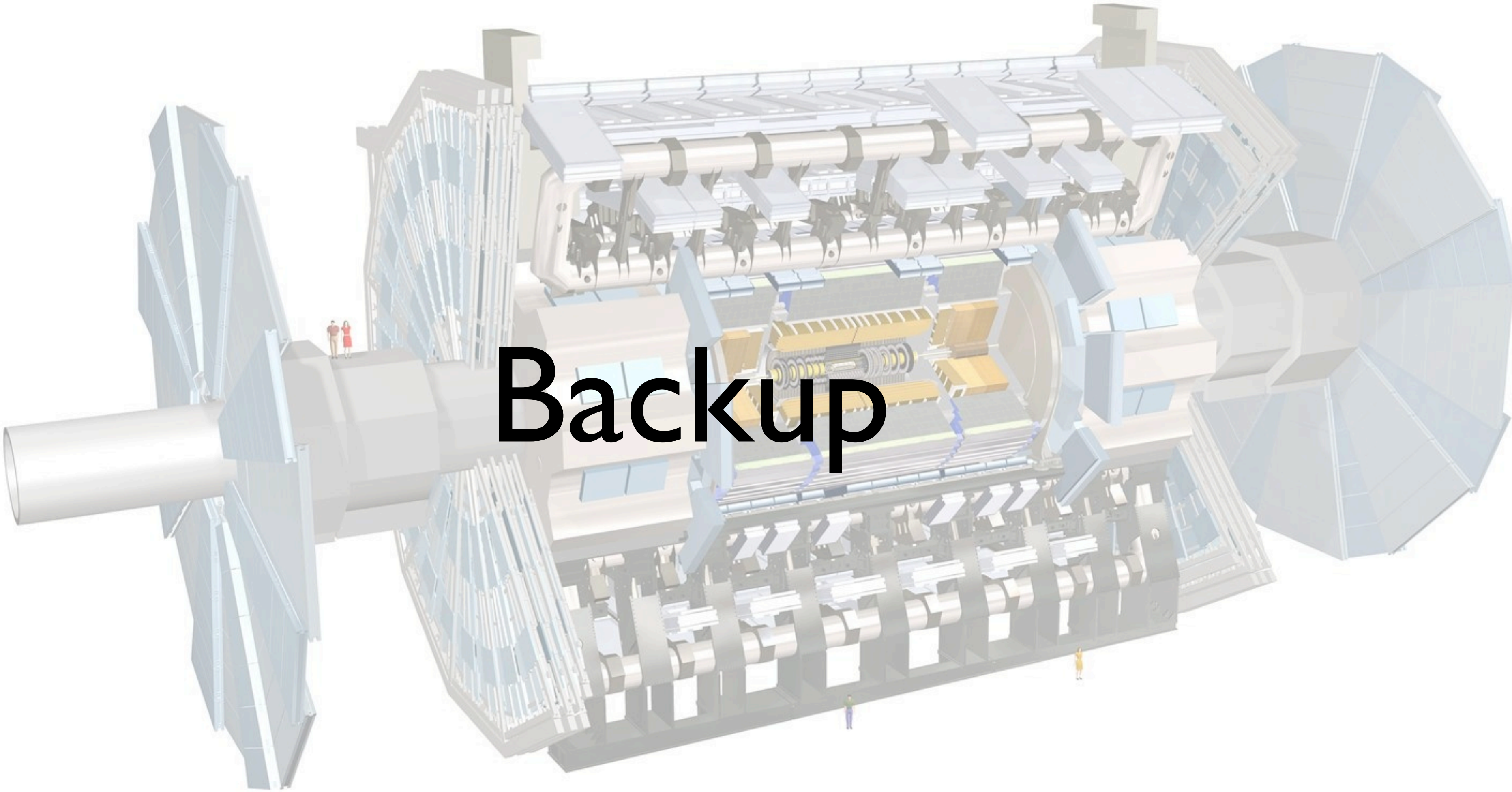


Algorithm	Rejection	Timing [ms]
Single μ, p_T > 10 GeV	0.78	5
Single μ + J/Ψ->μμ	0.98	75
Single μ + ϒ->μμ, full-scan	0.87	600

Summary

- Many changes planned for the ATLAS T/DAQ hardware infrastructure
  - New Level-1 components
  - FTK
  - Streamlined network configuration
- Multiple upgrade projects going on in parallel to be ready with the High Level Trigger for LHC's Run 2
  - Integrating the usage of the new hardware components
  - Simplifying the infrastructure by creating a homogenous HLT system
  - Incorporating the changes made in the offline reconstruction
- Already preparing the software for the requirements posed after Long Shutdown 2...

Backup



Vocabulary

- ***FE***: Front End
- ***ROD***: Read Out Device
- ***ROS***: Read Out System
- ***ROB***: Read Out Buffer
- ***SFI***: Sub-Farm Input
- ***SFO***: Sub-Farm Output

ATLAS Trigger Menu

- Trigger algorithms organized into “trigger chains”
    - Quality of reconstructed objects checked at multiple steps
    - If reconstructed object(s) passes all requirements, the chain passes the selection
- The trigger menu is a collection of such trigger chains
    - Menu used in 2012 held O(1k) chains