ATLAS TDAQ system: current status and performance

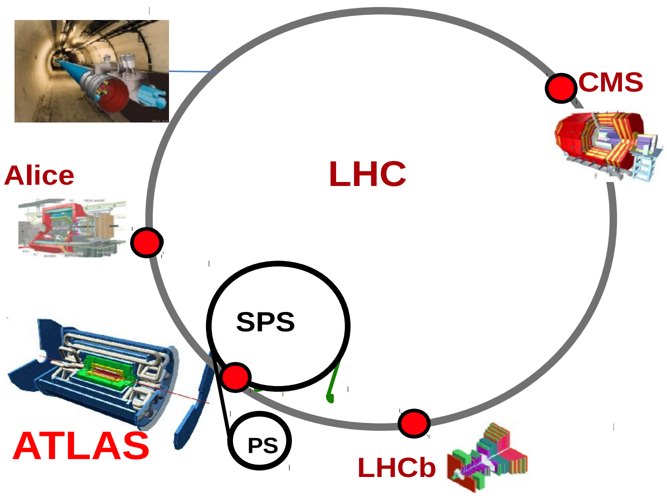
Sergio Ballestrero
University of Johannesburg
& CERN PH/ADT

on behalf of the
ATLAS Collaboration

Outline

- Introduction
  - LHC&ATLAS
  - TDAQ Design, scheme
  - 2010 status
- Evolution from 2010 to 2011
  - Luminosity
  - Trigger and data rates
  - Event Filter expansion
  - ROS optimisation and updates
- Status in 2011
  - Efficiency
  - TDAQ & Controls Software
- Conclusion
- Outlook

LHC & ATLAS



Nominal working conditions:

p-p beams: $\sqrt{s}$=14 TeV; $\mathcal{L}=10^{34}$ cm^{-2}s^{-1}; Bunch Cross every 25 ns

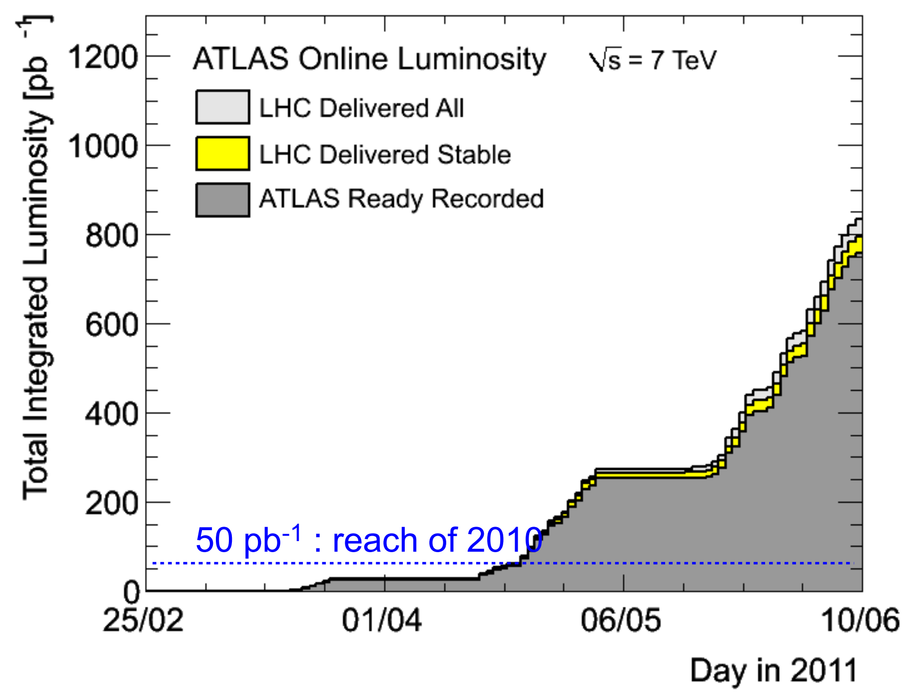
Pb-Pb beams: $\sqrt{s}$=5.5 TeV; $\mathcal{L}=10^{27}$ cm^{-2}s^{-1}

2010

- $\sqrt{s}$=7 TeV
- up to 233 colliding bunches in ATLAS
- Peak $\mathcal{L}\sim10^{32}$ cm^{-2}s^{-1} (October)
- Pb ion run (November)

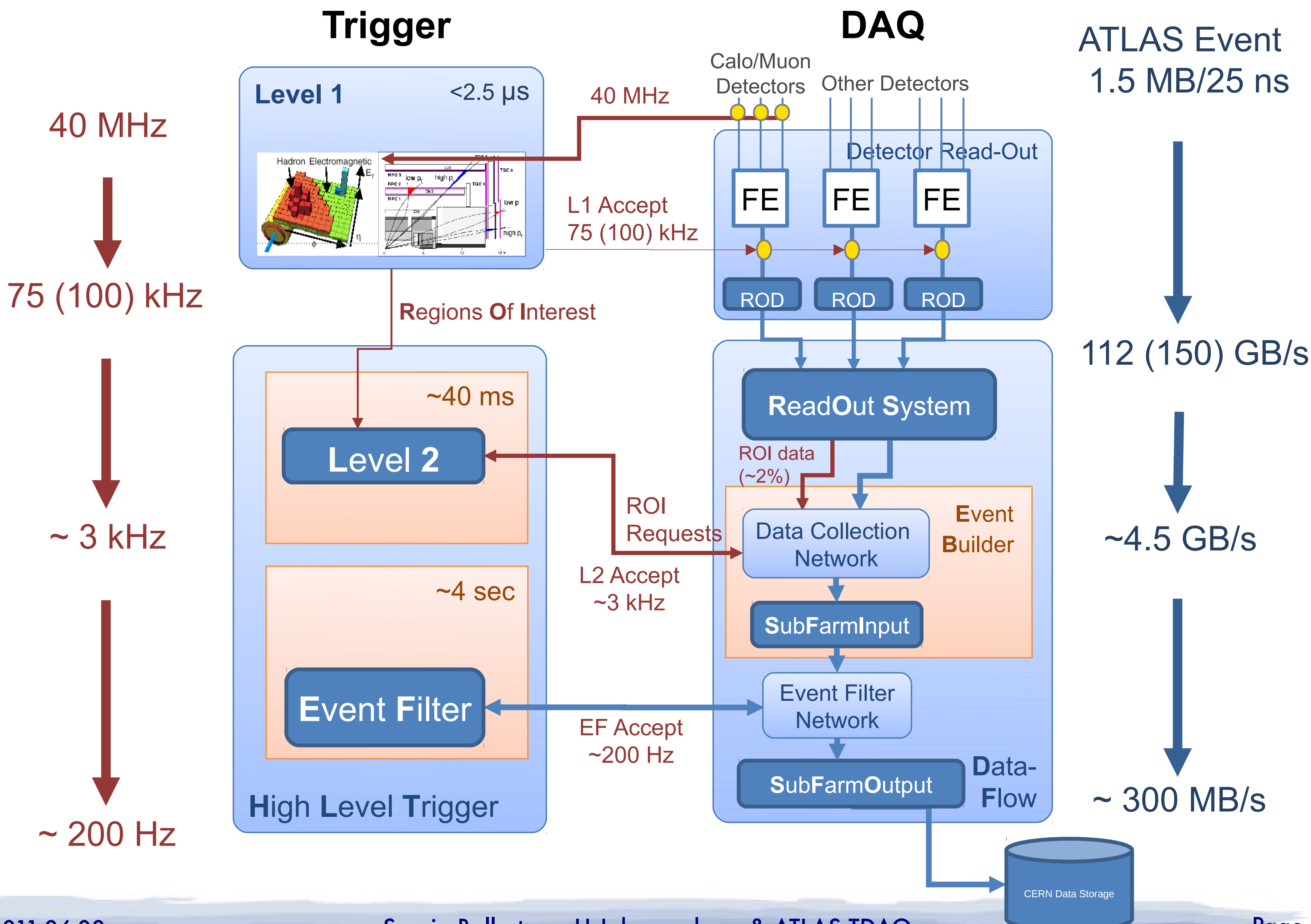
2011

- $\sqrt{s}$=7 TeV
- May: 1092 bunches, $\mathcal{L}_{peak}\sim1.26\cdot10^{33}$ cm^{-2}s^{-1}
- goal: 1400 bunches, $\mathcal{L}\sim1.6\cdot10^{33}$ cm^{-2}s^{-1}
- possible higher $\mathcal{L}$ with beam optimisations



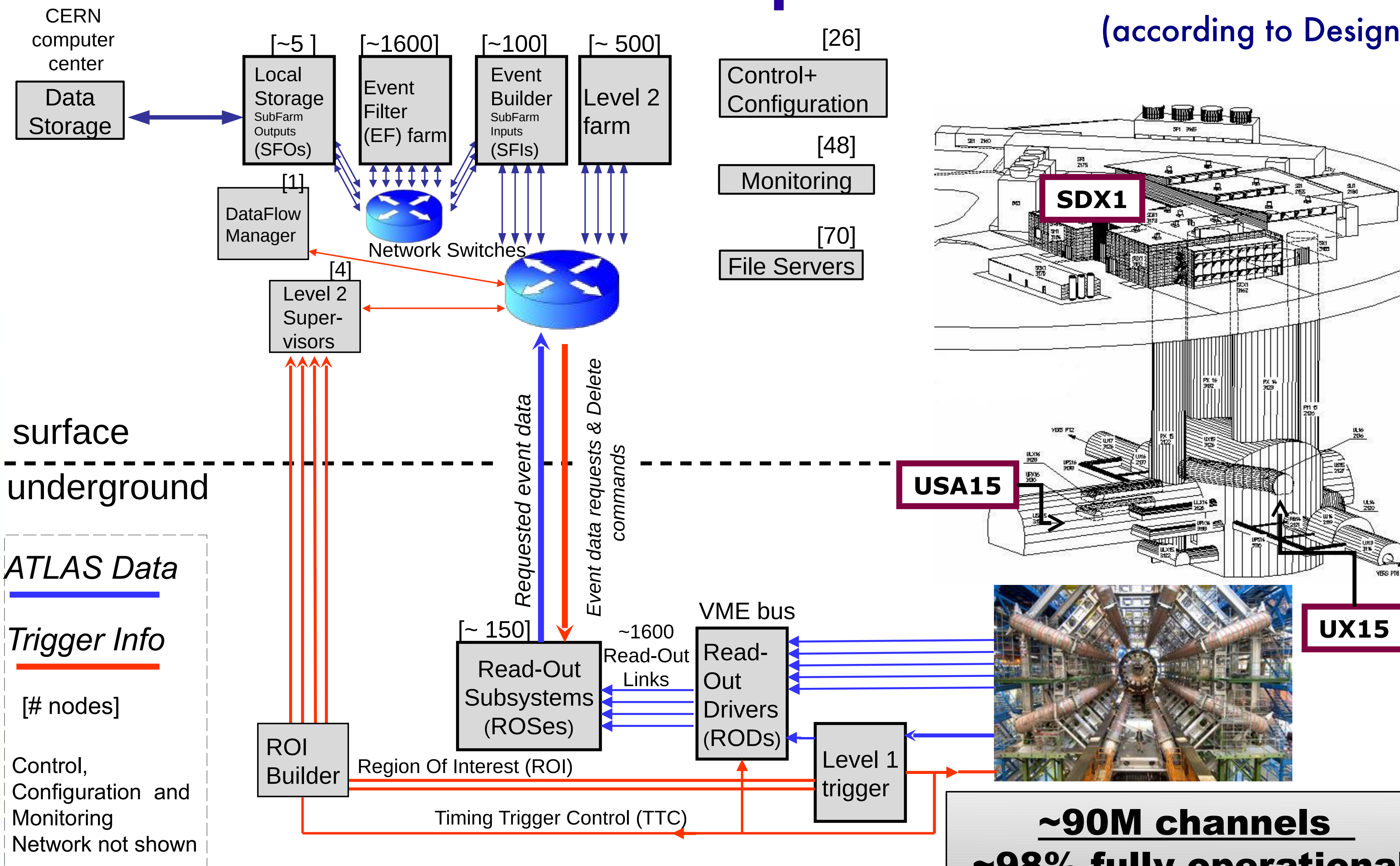
ATLAS TDAQ Design

ATLAS Data
Trigger Info



ATLAS TDAQ Implementation

(according to Design)



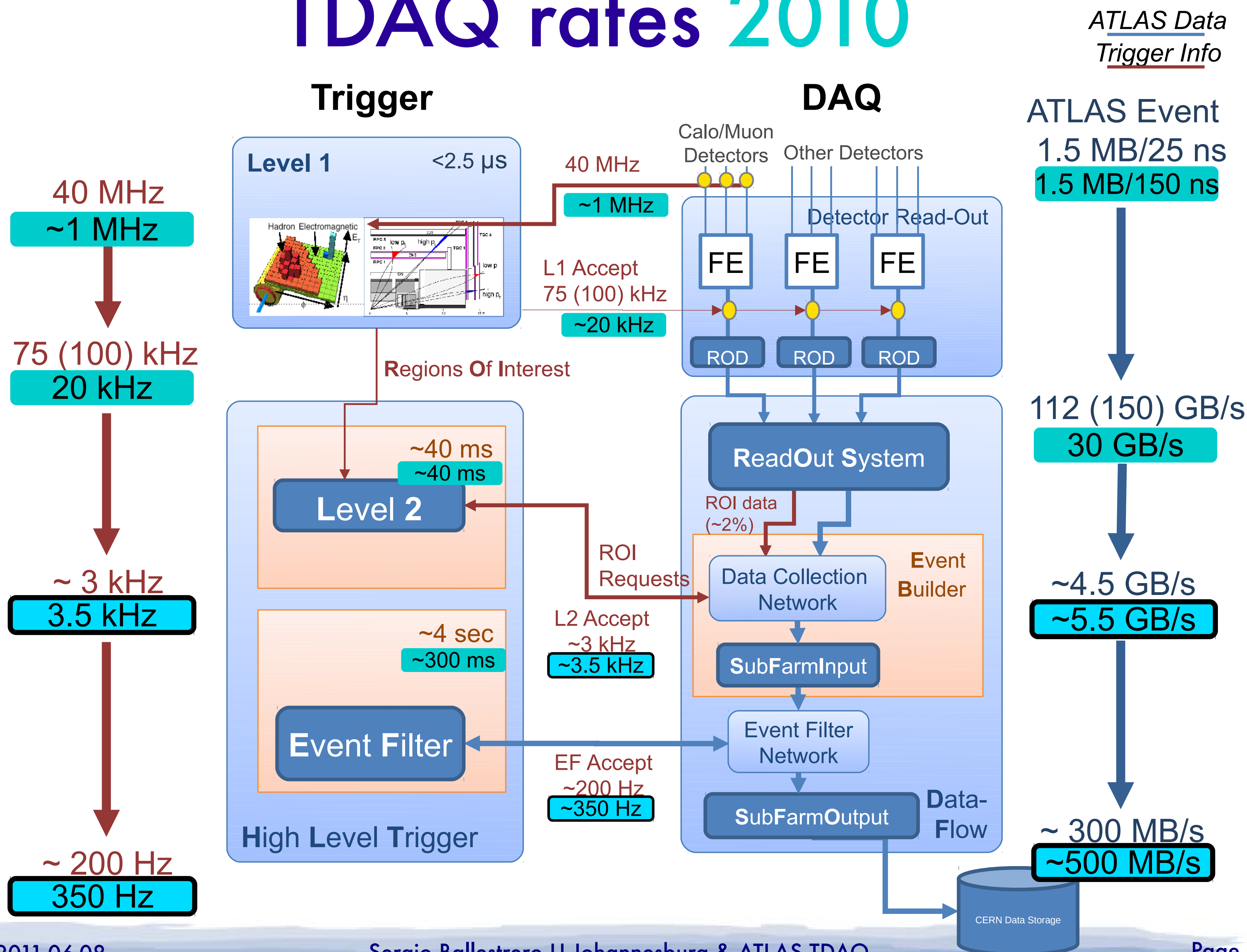
~90M channels
~98% fully operational
in 2010

TDAQ Farm status: 2010 & early 2011

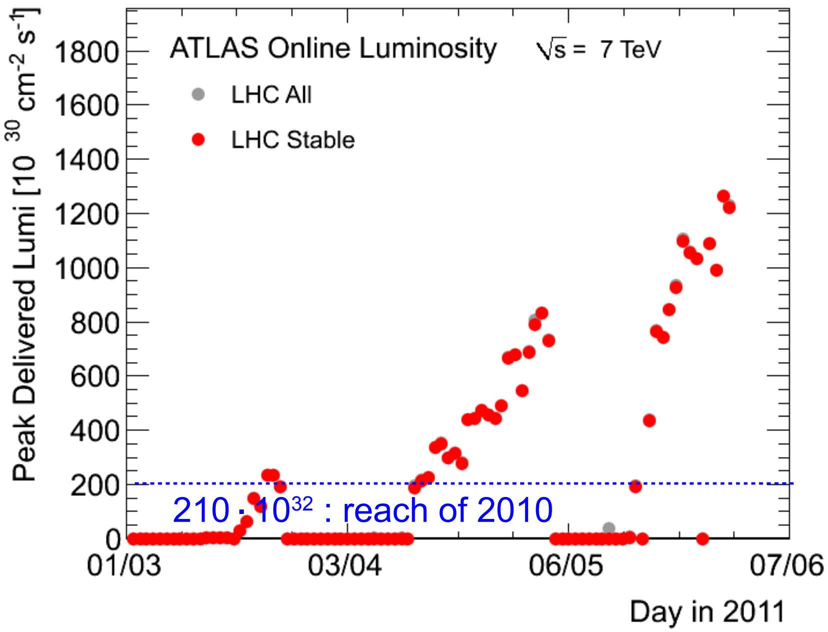
Component	Installed	Comments
Online & Monitoring	100%	~60 nodes
ROSes	100%	~150 nodes
ROIB & L2SVs	100%	
HLT (L2+EF)	~50%	~800 xpu nodes; ~300 EF nodes (since October 2010)
Event Builder	100%	~60 nodes (exploiting multi-core)
SFO	100%	Headroom for high instantaneous throughput
Networking	100%	Redundancy deployed in critical areas

- 27 XPU racks ~800 XPU nodes
  - XPU = L2 or EF Processing Unit (connected to both EF & L2 networks) on a "run by run" basis can be configured to run either as L2 or EF, allows high flexibility to meet the trigger needs
  - Functional role assignment not automated
- 10 EF racks ~300 EF nodes
  - EF nodes are dedicated (connected to EF network only)

TDAQ rates 2010



LHC from 2010 to 2011

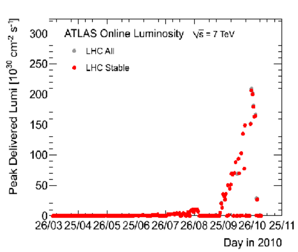


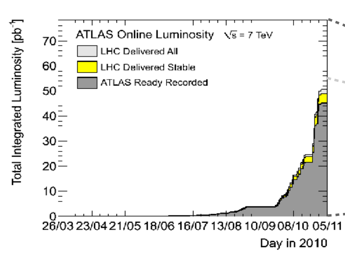
LHC has delivered excellent luminosity since the beginning of 2011

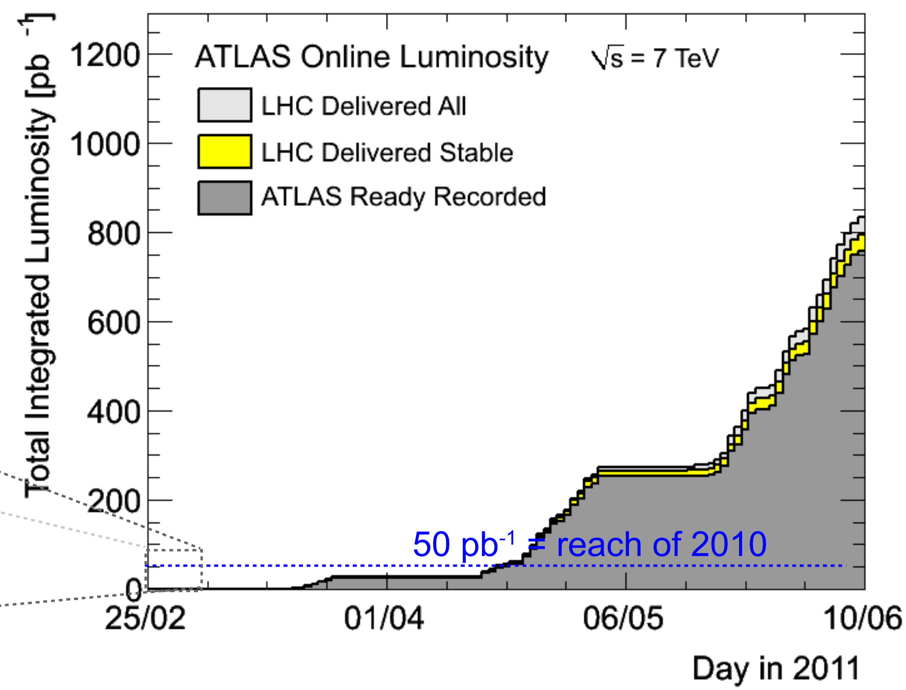
$\mathcal{L}_{peak}$=1.26·10^{33} cm^{-2}s^{-1}

$\mathcal{L}_{int}$= ~~539~~ 768 pb^{-1}

so good that we can't put the 2010 plots to scale !

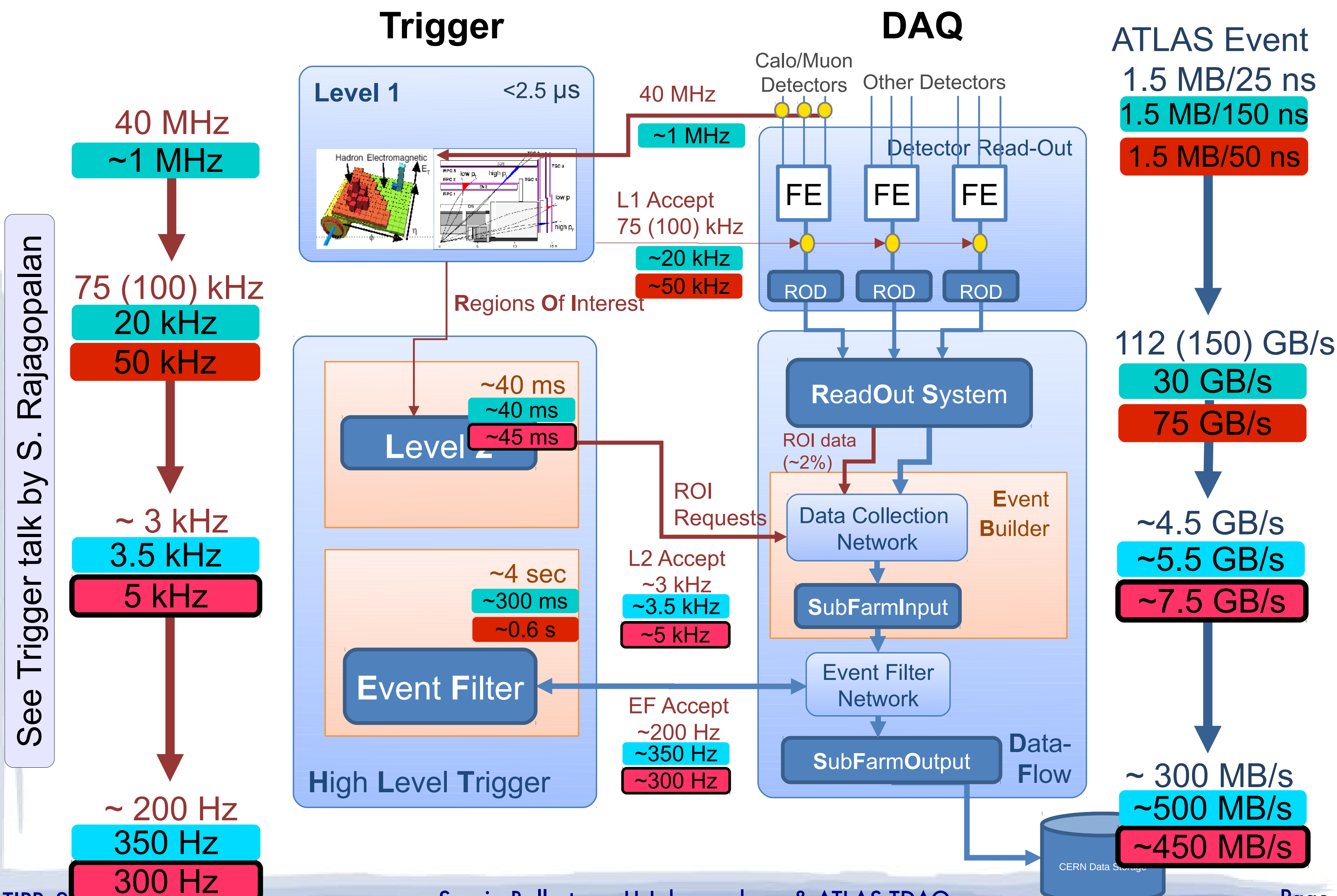




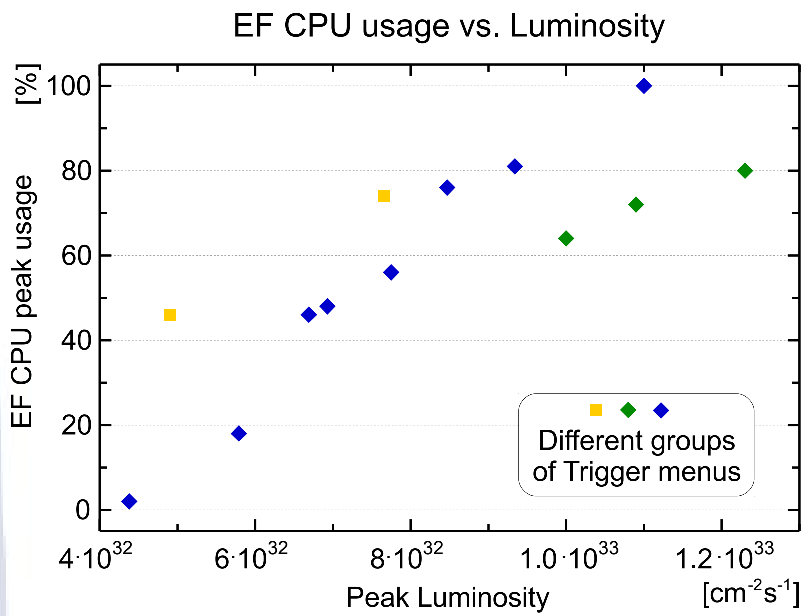


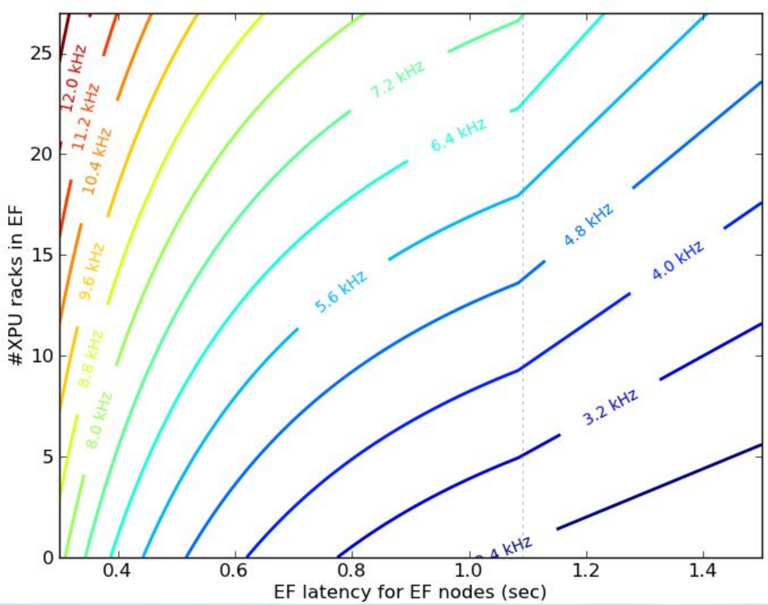
TDAQ rates 2010 → 2011

ATLAS Data Trigger Info



Luminosity vs EF CPU load





- EF CPU usage
  ~ linear with $\mathcal{L}$
  > linear with μ
  (beam optim. will increase pileup)

Complementary info in talk by R. Ospanov

- Balance XPU assignment between L2 and EF
  given other constraints as
  - EF processing time
  - ROS load
  - XPU network B/W
  - Tier0 storage B/W

Details in bonus slide (pg.20)

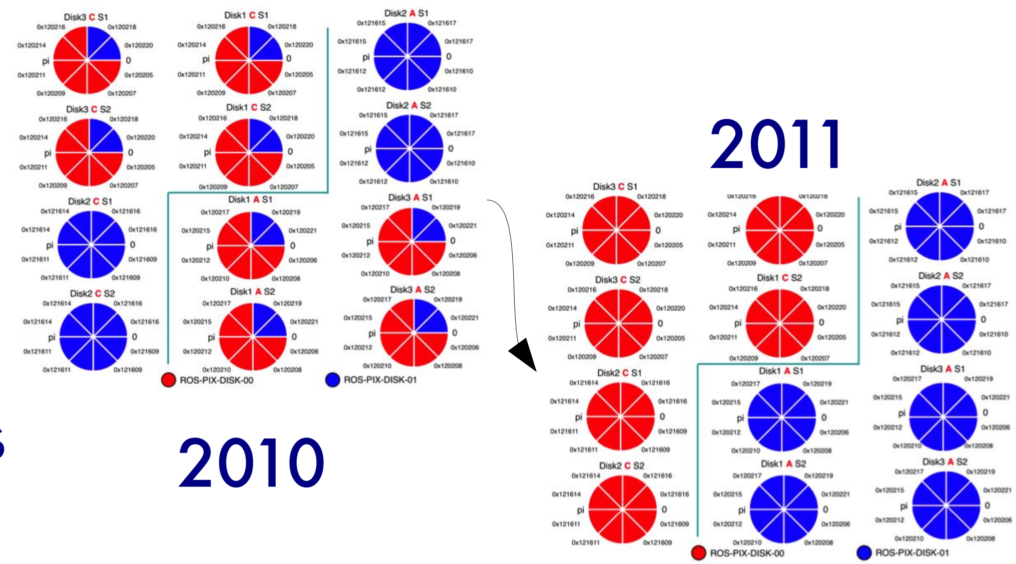
TDAQ Farm 2011: HLT expansion

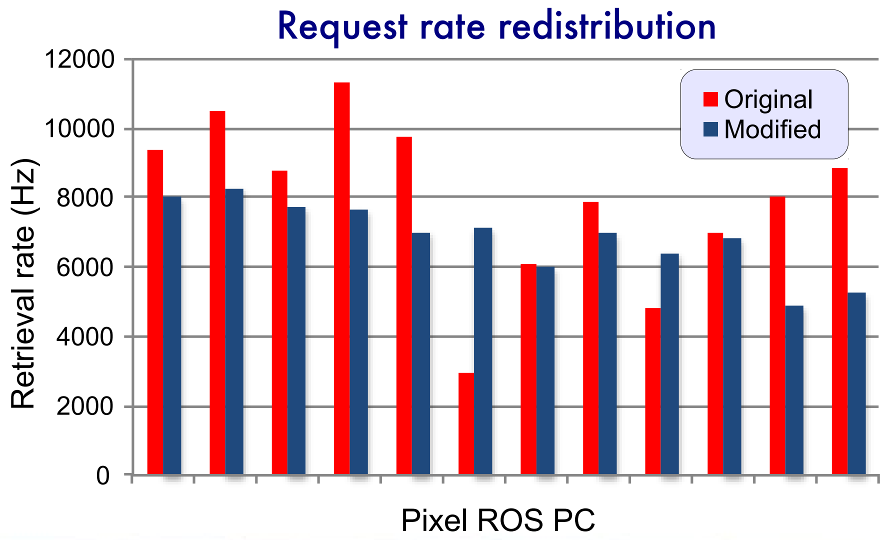
Component	Installed	Comments	Mid 2011
Online & Monitoring	100%	~60 nodes	100%
ROSes	100%	~150 nodes	100%
ROIB & L2SVs	100%		100%
HLT (L2+EF)	~50%	~300 EF nodes -> 750 ~800 xpu nodes -> assign more to L2	~70% of design #
Event Builder	100%	~60 nodes (exploiting multi-core)	100%
SFO	100%	Headroom for high instantaneous throughput	100%
Networking	100%	Redundancy deployed in critical areas	100%

- Using full CPU capacity on EF racks
- Network limit for EF on XPU racks (beyond design rates)
  2 Gbit links/rack → maximum possible EF bandwidth with XPU racks ~6 GB/s
- Conclusion:
  - it's time to install more EF racks, according to design plans, and reassign XPUs to L2
  - also rolling update of XPU racks to current CPUs

Pixel ROS mapping optimisation

- RODs for adjacent detector sections connected to different ROSes
- for a single ROI requires requests to >1 ROS, increases total requests
- optimal ROD-ROS links allow rates further above design
- implemented January 2011

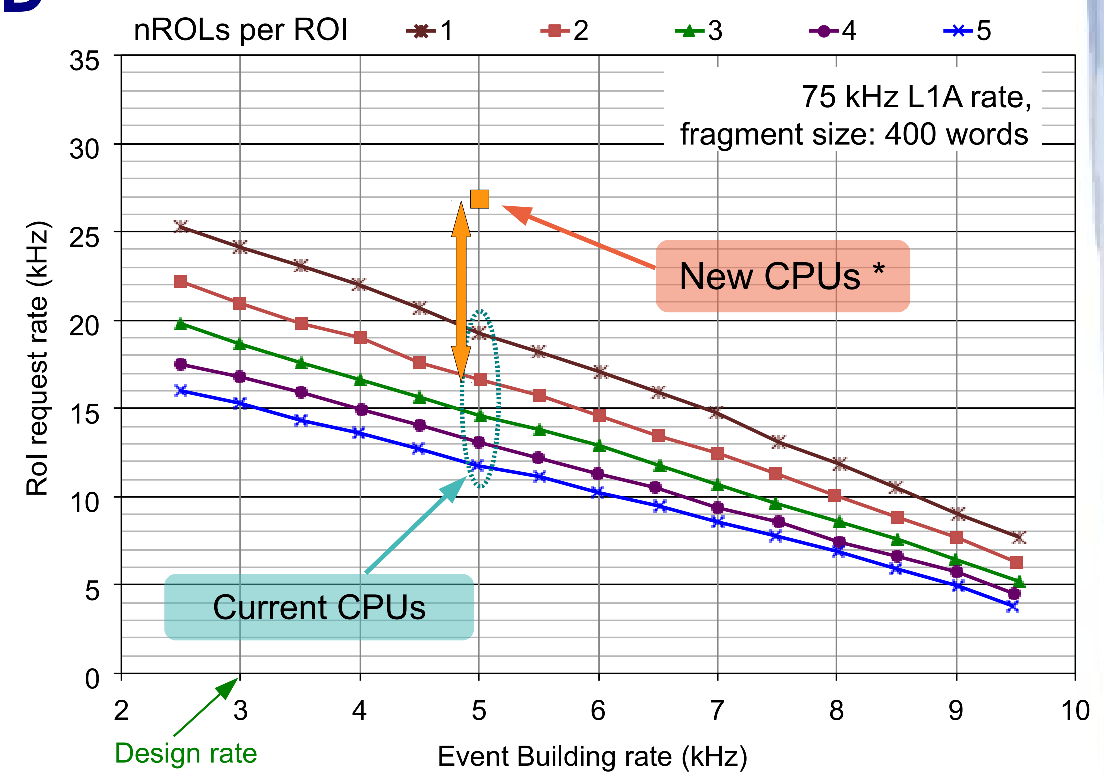




More details in talk by R. Ospanov

ROS L2 vs EB load

- ROS provide L2 & EB
  - L2: ROI fragments
  - EB: all fragments
- Total nr. of requests limited by CPU load
- 5KHz @EB
  above design (3kHz)



* not optimised, does reach network B/W limits

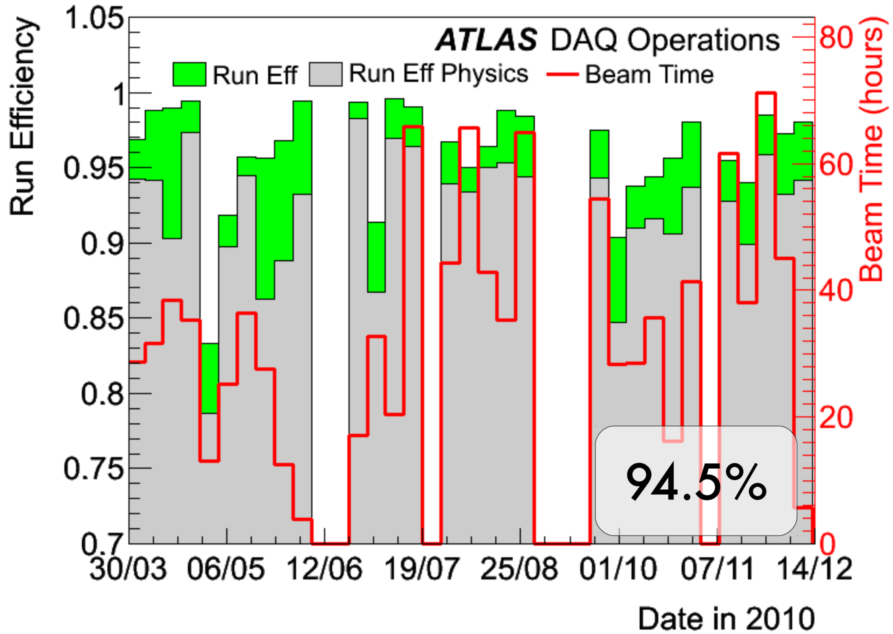
→ too little headroom for L2 requests

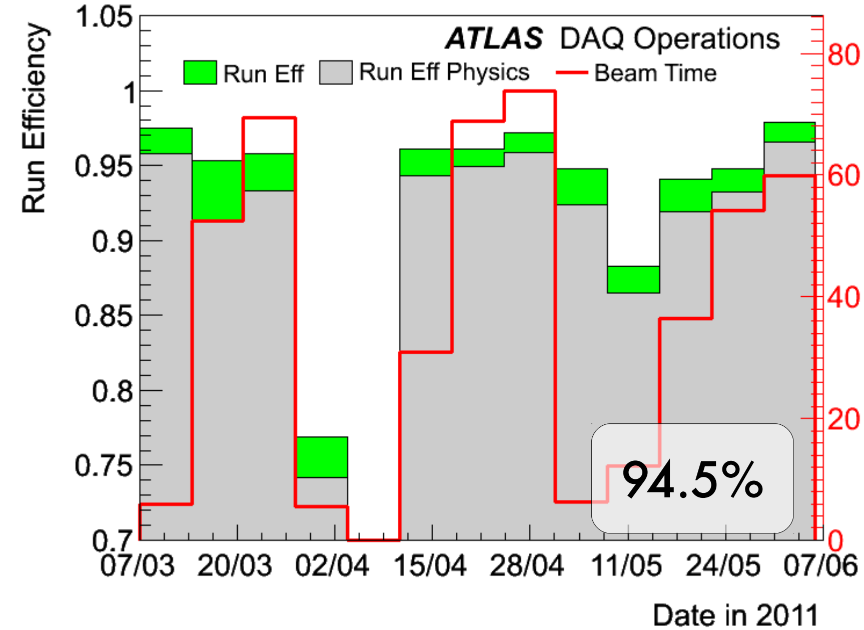
TDAQ Farm 2011: update ROS & SFO

Component	Installed	Comments	Mid 2011
Online & Monitoring	100%	~60 nodes	100%
ROSes	100%	~150 nodes -> update each	100%
ROIB & L2SVs	100%		100%
HLT (L2+EF)	~50%	~300 EF nodes -> 750 ~800 xpu nodes -> assign more to L2	~70%
Event Builder	100%	~60 nodes (exploiting multi-core)	100%
SFO	100%	Increase disks for 400Hz sustained	100%
Networking	100%	Redundancy deployed in critical areas	100%

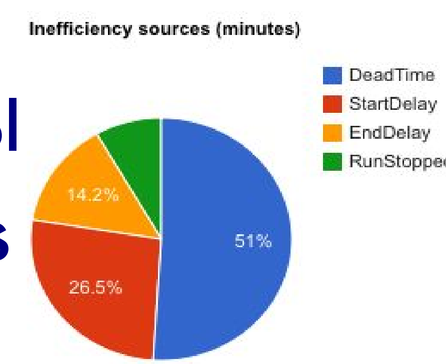
- ROS request rates are CPU limited
  a limited update provides headroom beyond design
  - update ROS PCs with more performant CPUs
  - sufficient for 7 TeV LHC run (until *Long Shutdown* of 2013)
- SFO disk capacity increase for 2-day buffer at 400Hz sustained
  - rate capability proven with (burst) up to 1.3kHz in 2010

TDAQ Efficiency



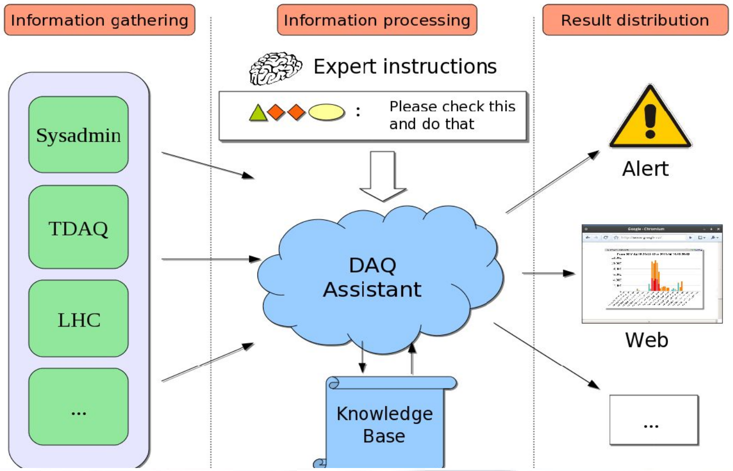


- Run efficiency tools matured in 2010
  - Identify & keep dead-time sources under control
- Stop-less & automatic recovery procedures
  - minimise dead time, avoid full stop-start cycles
  - supported by an extensible Expert System



TDAQ & Controls Software

- New TDAQ SW Release
  - consolidate platform
  - additional scalability and fault tolerance
- Support Missing E_T trigger at Level2
  - now precise E_T^{miss} trigger requires a full Event Build
  - L2 will use instead ROI + E_T information from front-ends
  - more rate headroom at ROSes
  - ready for deployment
- Data compression at Tier0
  - soon in production
- DAQ Shifter Assistant
  - collect information streams from multiple sources
  - Knowledge Base rules provide alerts and suggestions to shifter
  - ease shift crew reduction



Conclusions

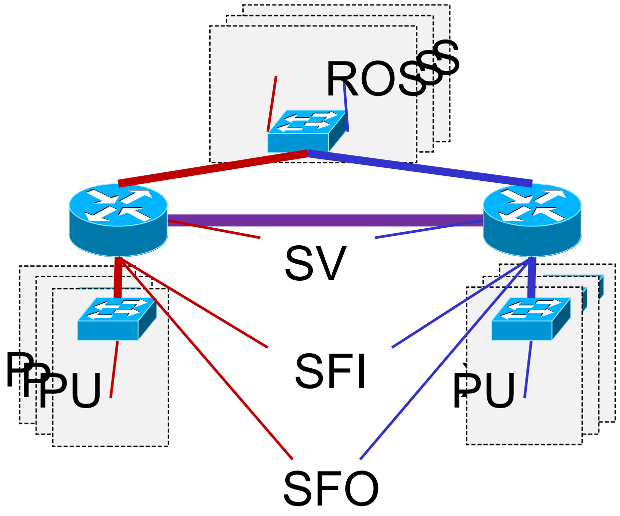
- ATLAS DAQ coped well with the 2010 operation, from low rates to beyond design
  - detector commissioning tested dataflow bandwidth & peak storage rates far beyond design specifications
- 2011 is higher luminosity and fully physics-oriented
  - now L2 and EF capacity will be increased as planned, to keep up with fast LHC progress
- The high efficiency (~95%) of 2010 has been maintained in these first months of 2011
  along a luminosity growth of 6 orders of magnitude

Bonus slides

Outlook on Upgrades

- ## Network

Upgrade of network cores proposed for 2013. Basic option maintains current architecture, improving scalability and fault tolerance. Investigating a new network topology that would unify L2 and EF networks. More flexible in balancing L2 and EF, and possibly better fitting one proposed evolution of the DataFlow software, but could be more difficult to scale.



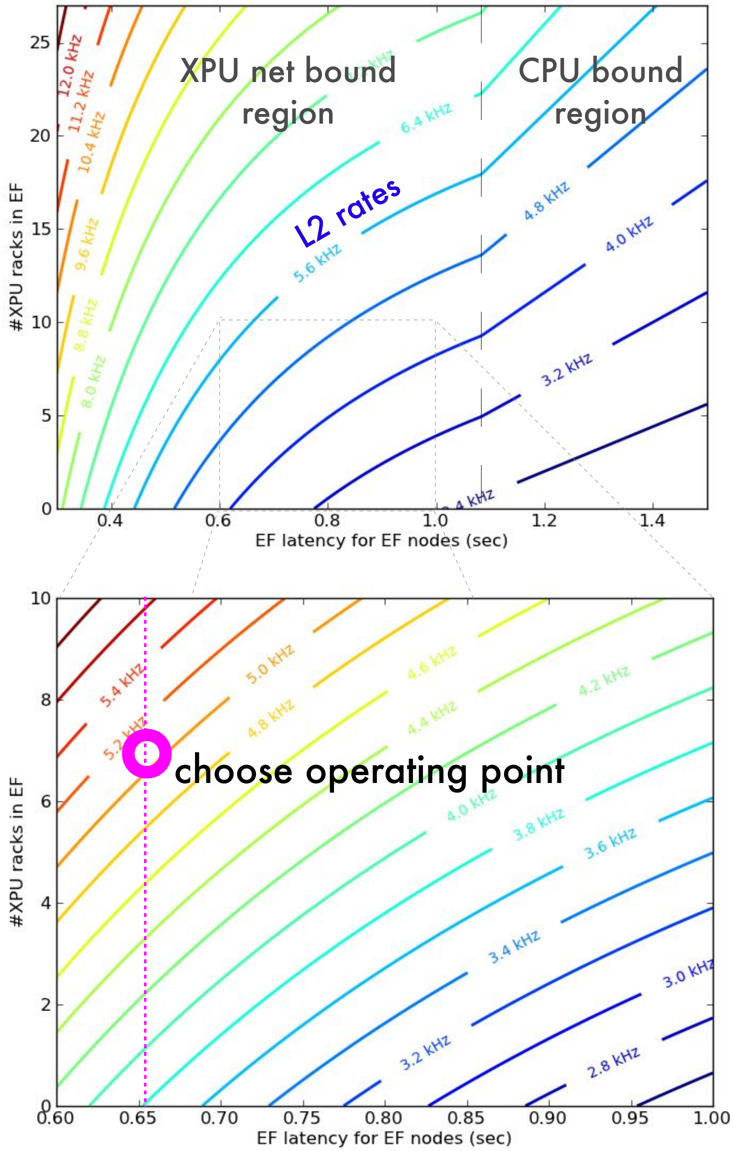
- ## MultiCore CPU

CPUs did not follow the clock speed increase path expected at the time of design, multi-core prevailed instead. Redesigns of certain SW components will allow better efficiency and new features, e.g. data compression at the SFOs.

HLT Farm Systems / CPUs

- 1st generation XPUs
  - 2 socket Xeon E5320 4 cores 1.8GHz "Clovertown"
  - 16GB RAM
- 2nd generation XPUs
  - 2 socket Xeon E5420 4 cores 2.5GHz "Harpertown"
  - 16 GB RAM
- 3rd generation EFs
  - 2 socket Xeon E5540 4 cores+HT 2.5GHz "Gainestown"
  - 24 GB RAM
- Coming XPU updates and EF expansion:
  - 2 socket Xeon X56xx 2.6GHz 4 cores+HT "Gulftown"
  - 24 GB RAM

TDAQ Phase diagrams



- TDAQ model includes constraints on EF/XPU allocation, network bandwidth, rates, rejection etc
- chosen one set of "given" parameters, e.g. EF latency, explore impact of resource allocation options
- provide input to Trigger group