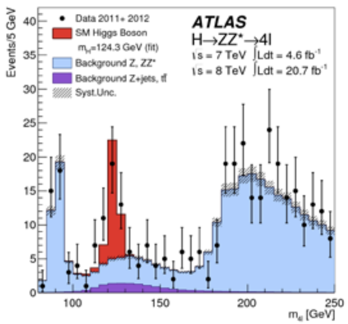
Software and computing evolution: the HL-LHC challenge

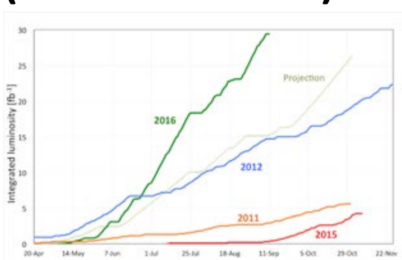
Simone Campana, CERN

The Large Hadron Collider at CERN

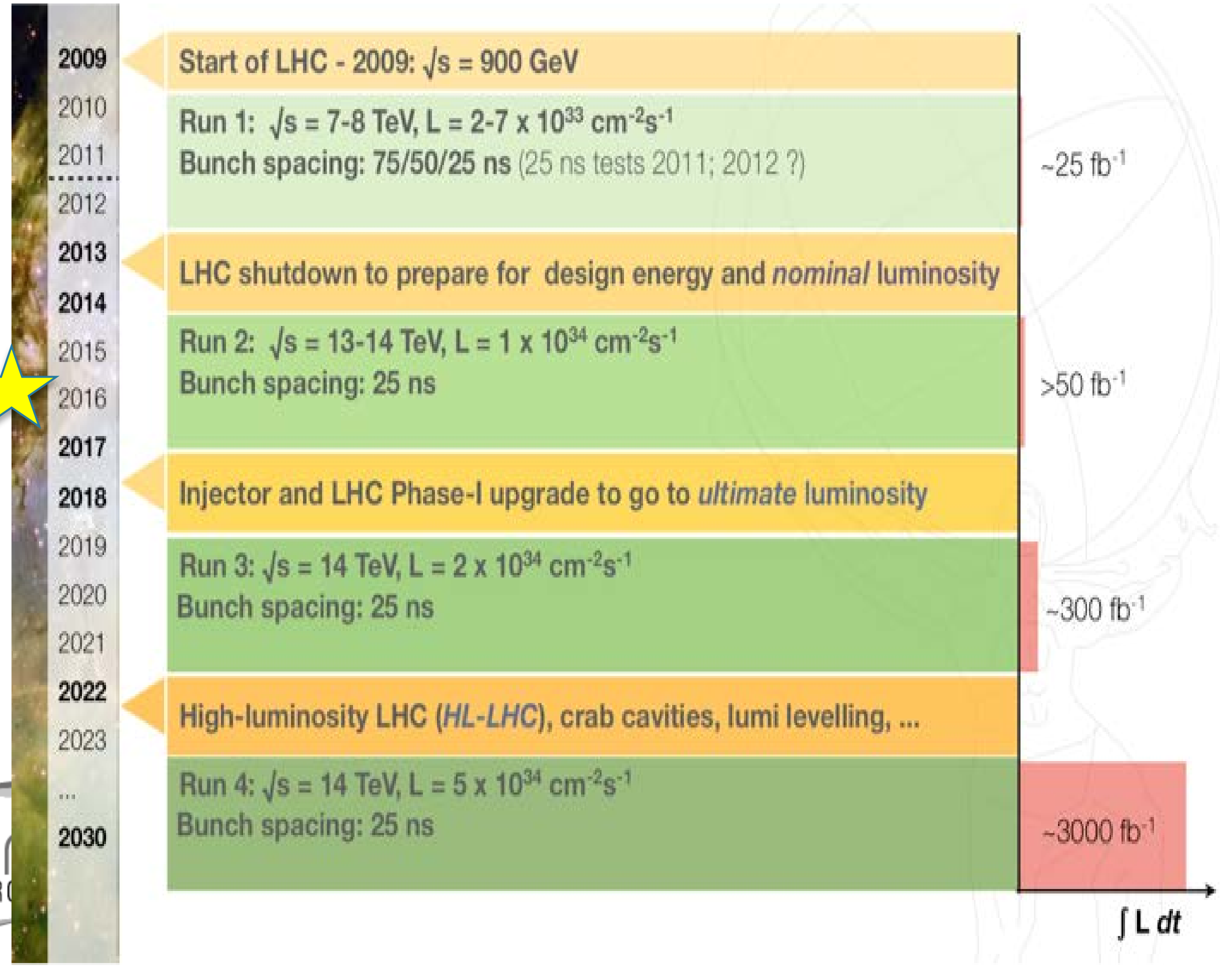
Higgs discovery in Run-1



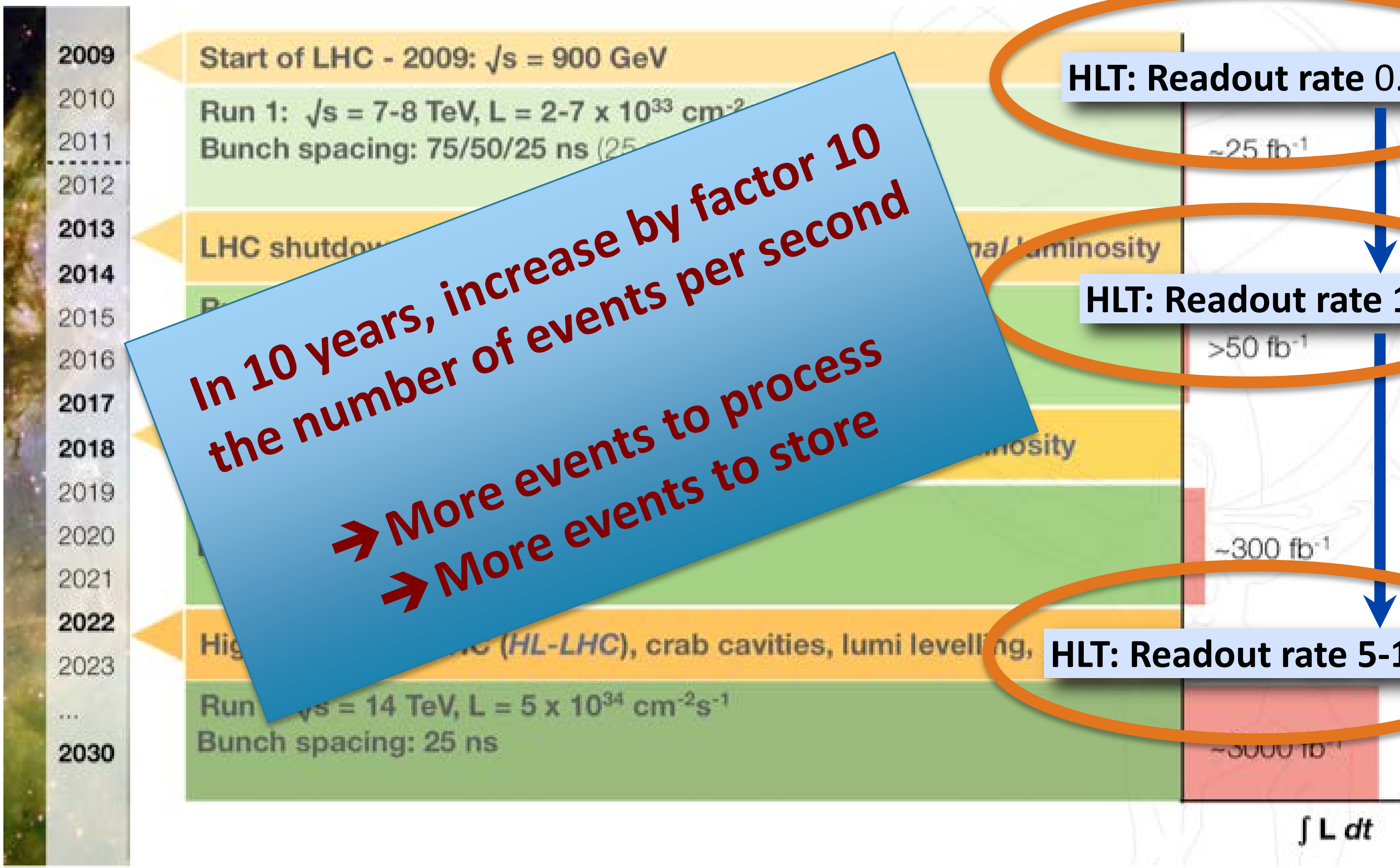
We are here: Run-2 (Fernando's talk)



High Luminosity: the HL-LHC challenge (Simone's talk)



The data rate and volume challenge



HLT: Readout rate 0.4 **kHz**

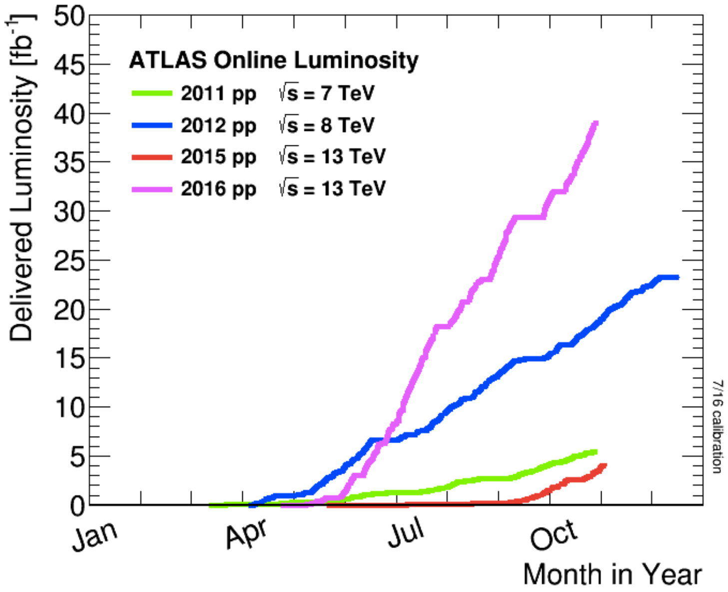
HLT: Readout rate 1 kHz

HLT: Readout rate 5-10 kHz

In 10 years, increase by factor 10 the number of events per second

➔ More events to process
➔ More events to store

Success story of 2016 data taking

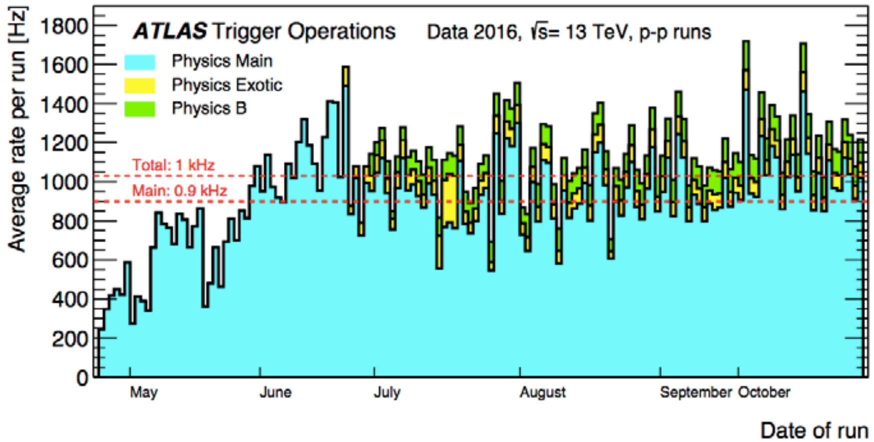


7.6 Billion events in pp collisions
1.4 Billion events in pPb collisions

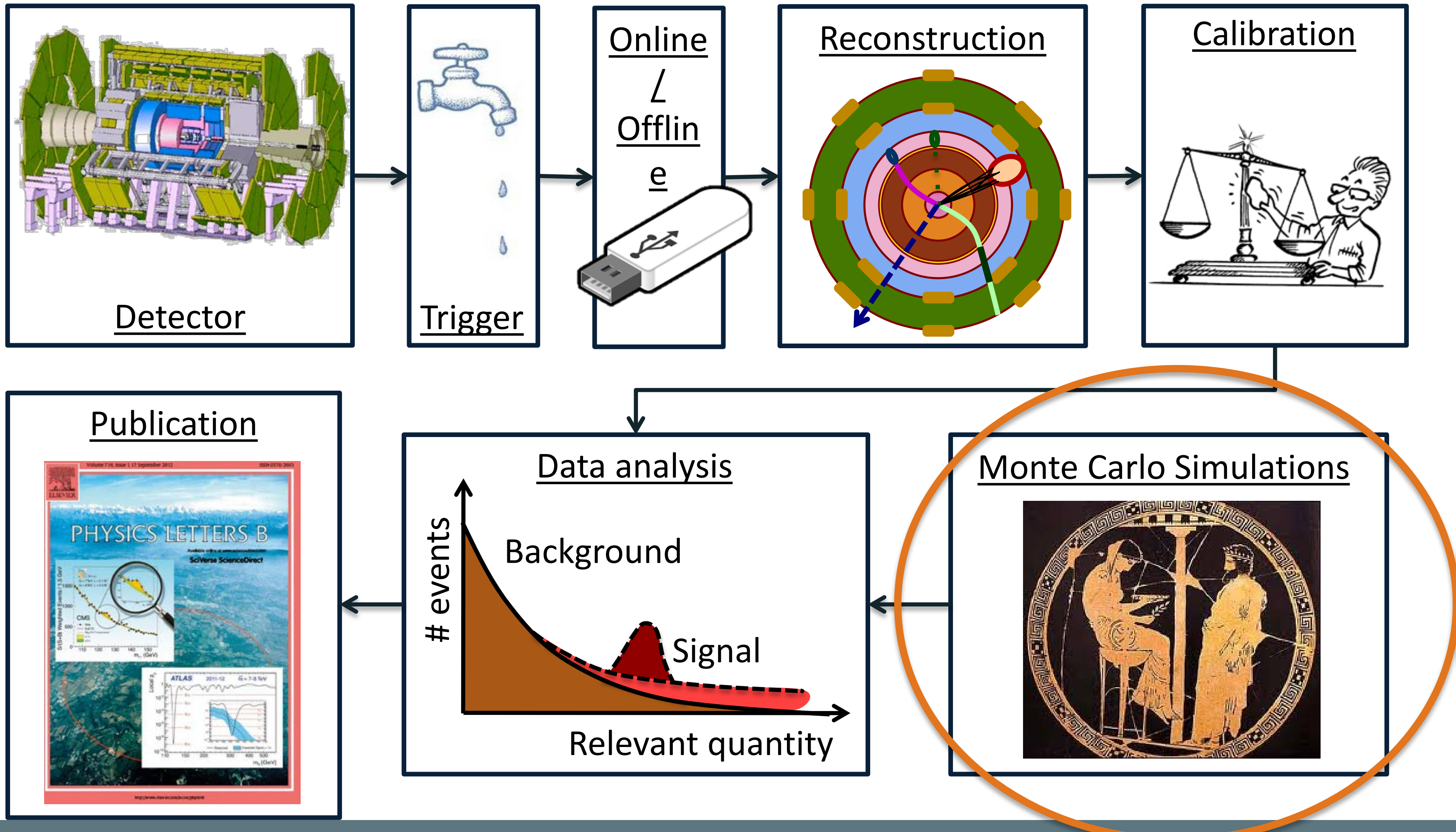
50% more data than expected

The LHC is a fantastically performing machine

ATLAS continuously adapts and improves, to take maximum benefit in terms of physics



Not only real data …



Monte Carlo production chain

How much processing time needed for each step?

Event Generation
simulate the physics process.

From < 1s to a few hours / event.

Detector Simulation
simulate the interaction of the particles with the detector material.

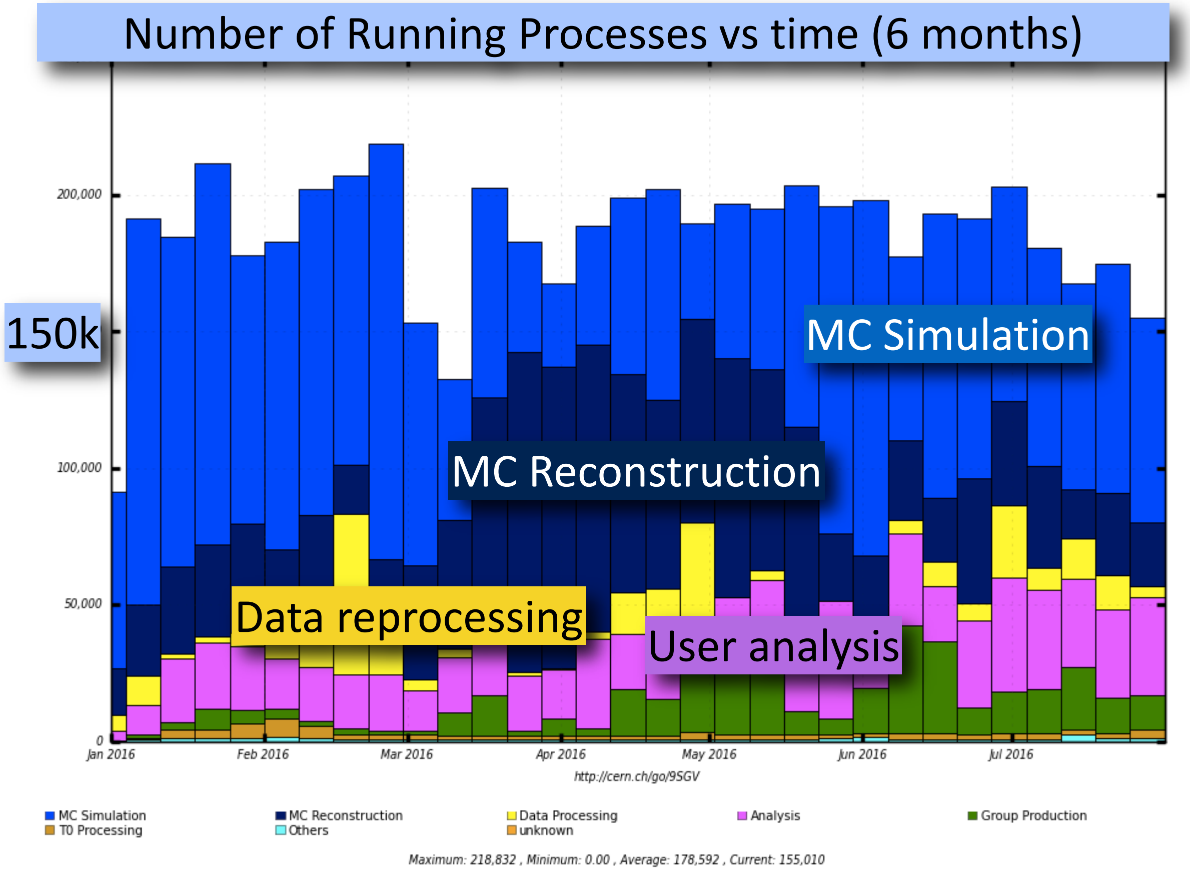
From 1 to 10min / event

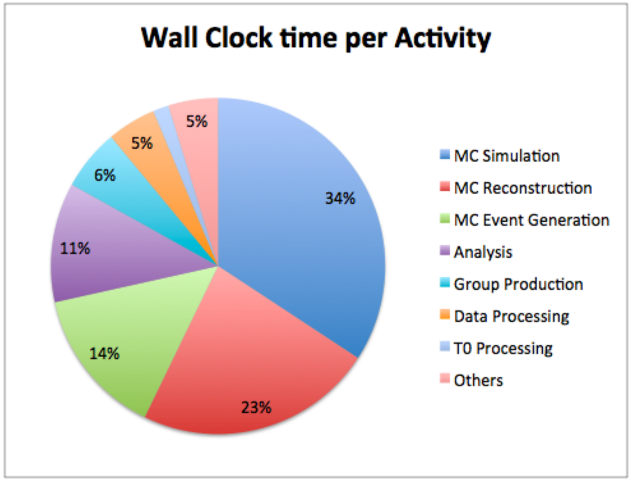
Digitization
Translate interactions with detector into realistic signals.

Reconstruction
Go from signals back to particles, as for real data.

From 5 to 60s / event

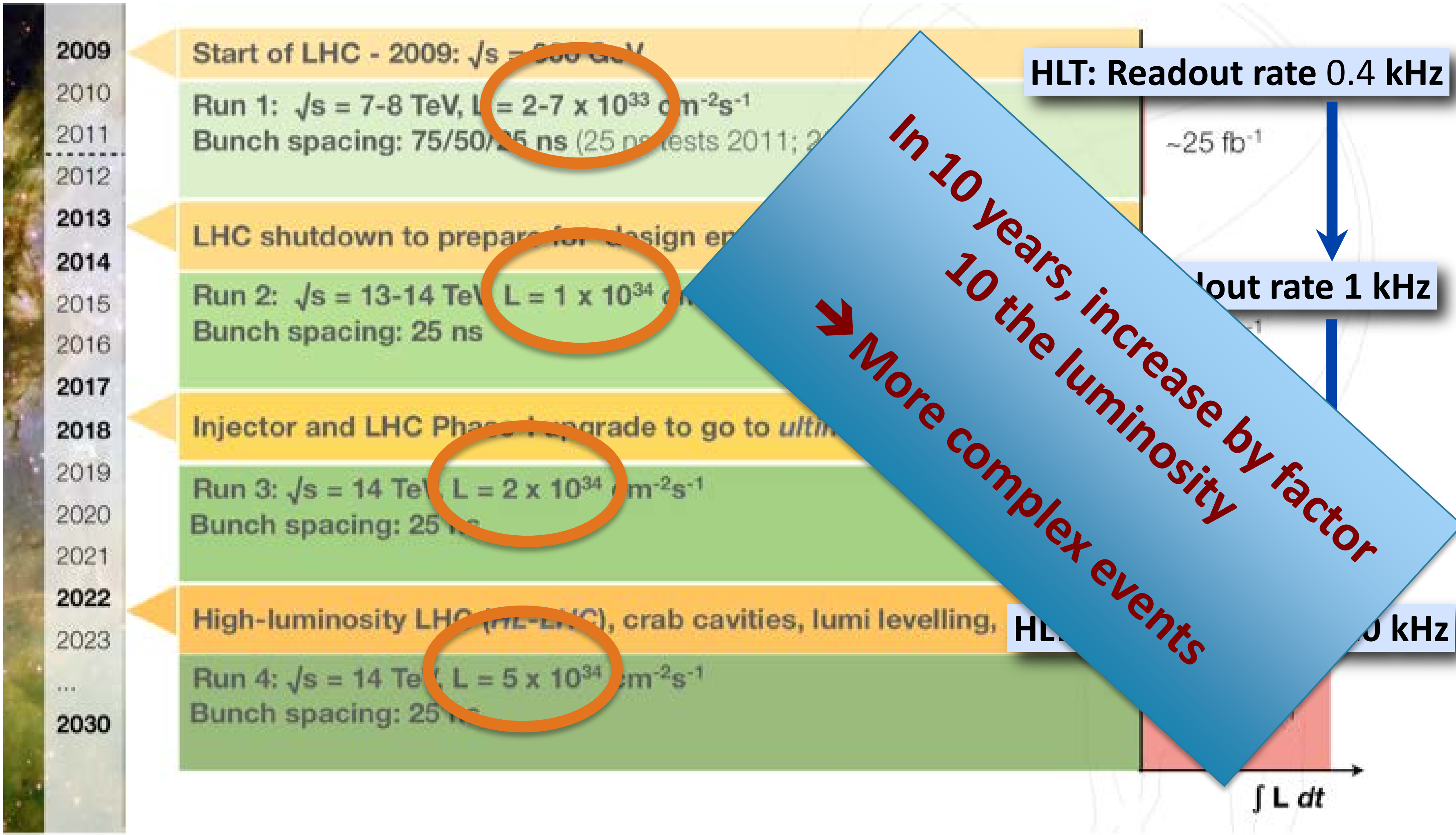
ATLAS computational load



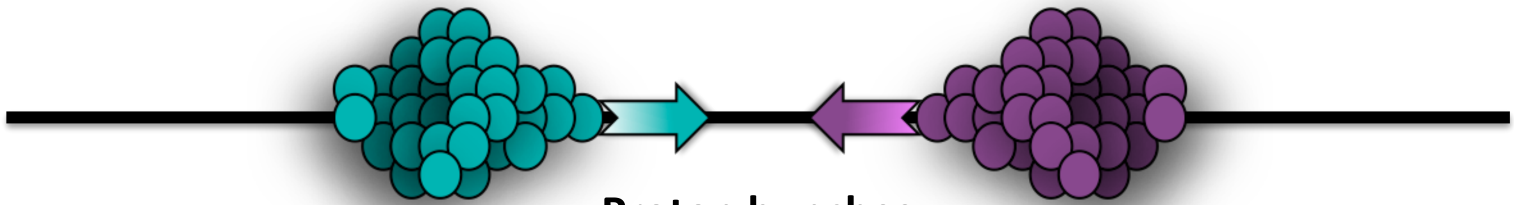


All together, 70% of ATLAS computing resources are utilized to produce simulated events samples

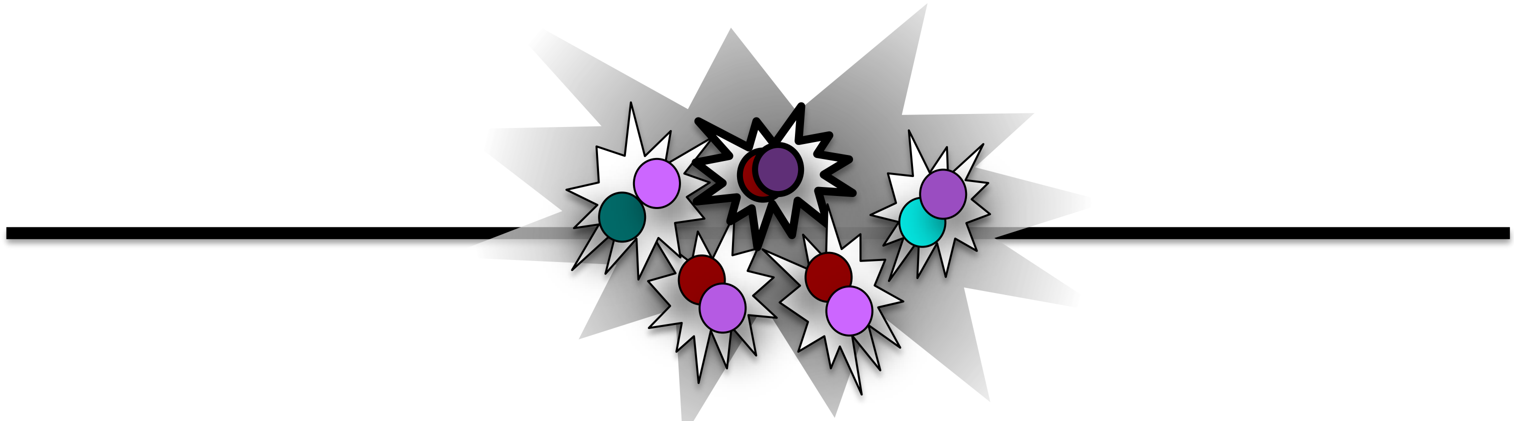
The data complexity challenge



Pile-up

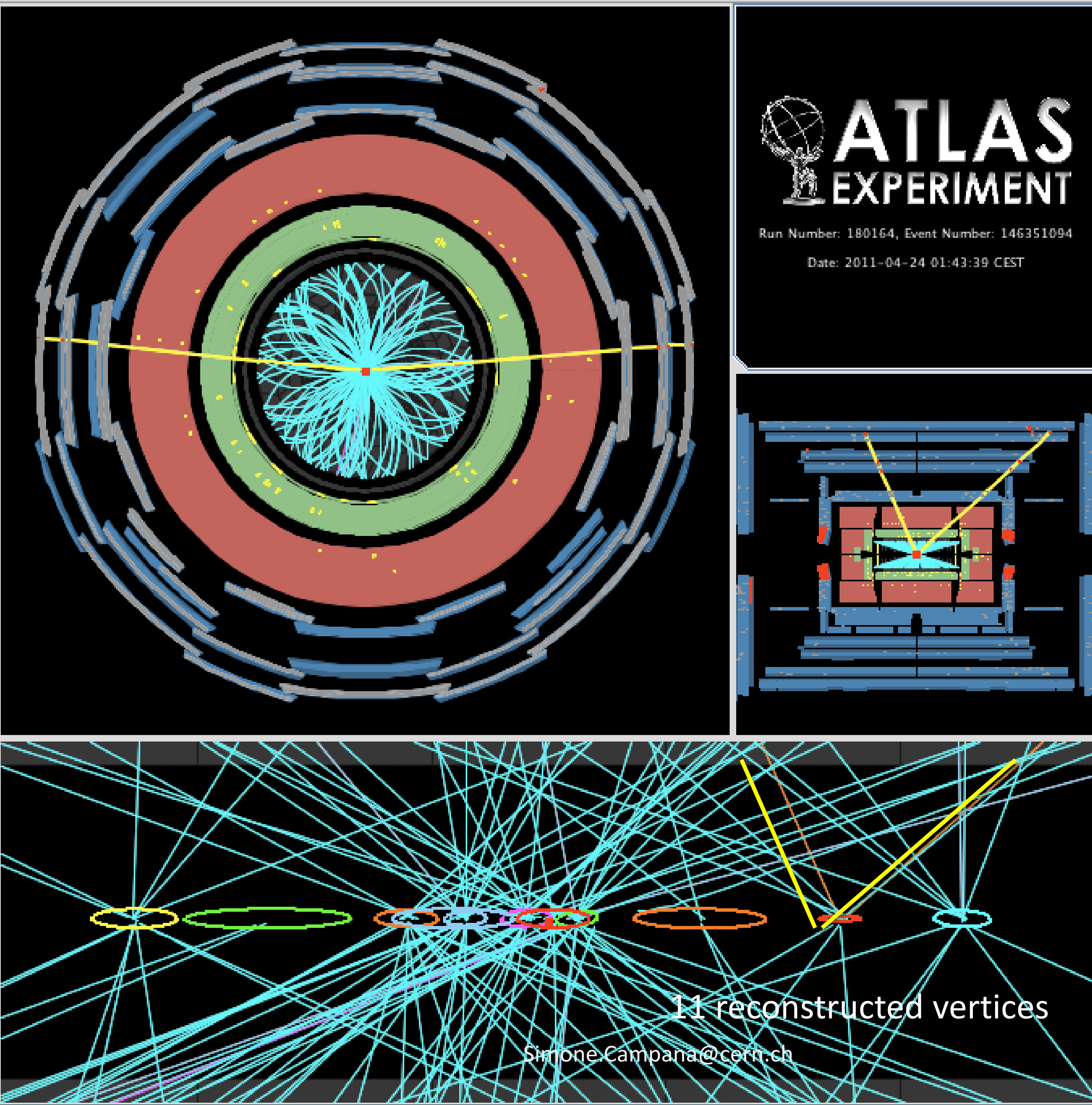


Proton bunches
>10^{11} protons/bunch
(colliding at ~**40MHz in run2)**



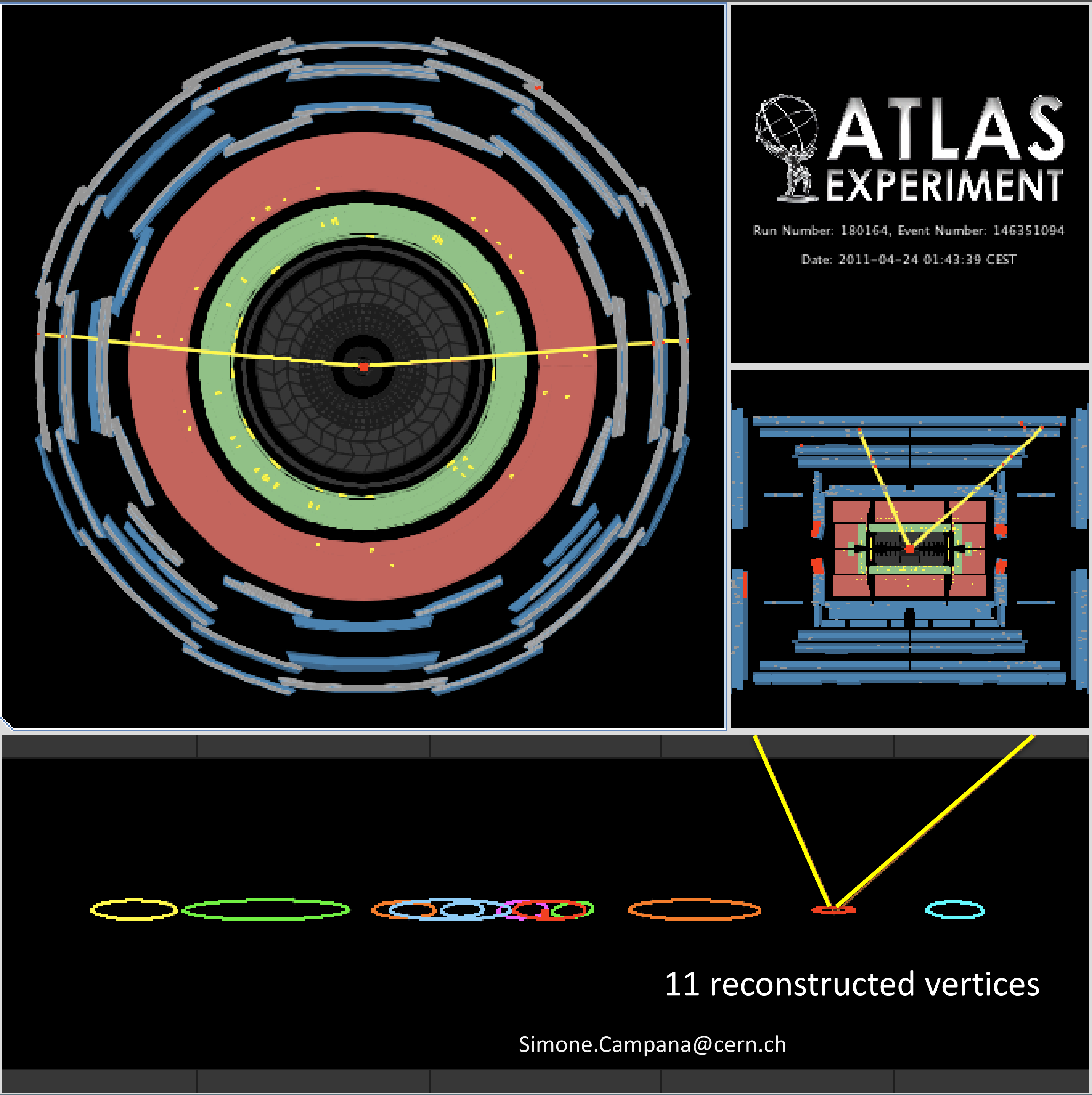
~30 p-p collisions / bunch crossing in 2016 data taking conditions

Z->μμ event;
2011 data.



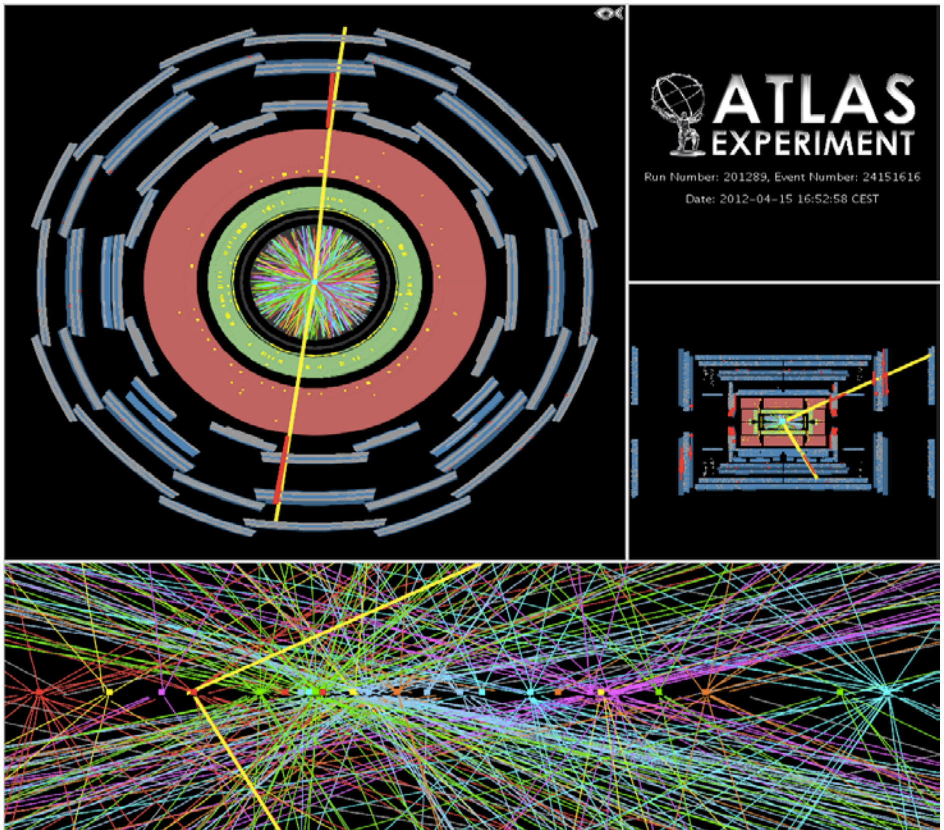
Track pT > 0.5 GeV

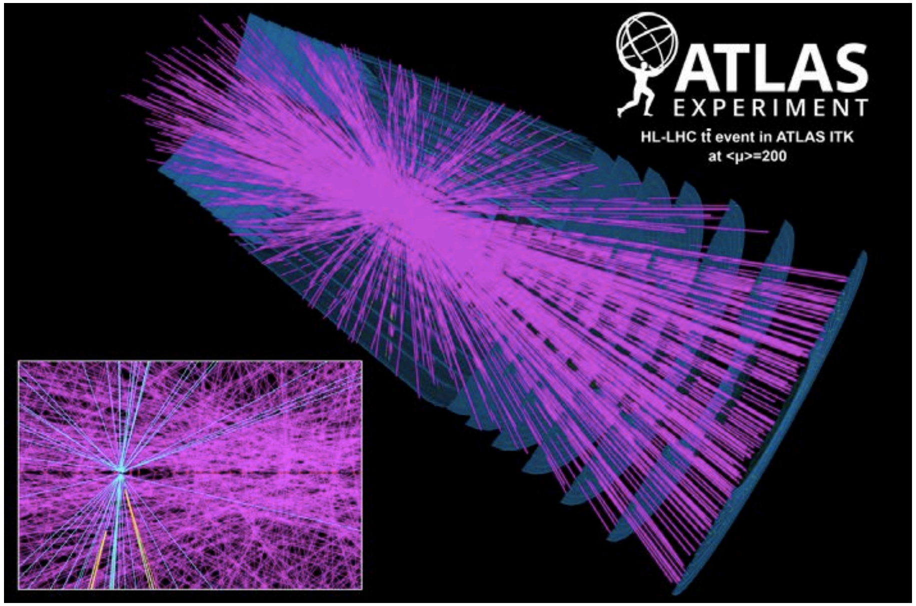
Z->μμ event;
2011 data.



Track pT > 10 GeV

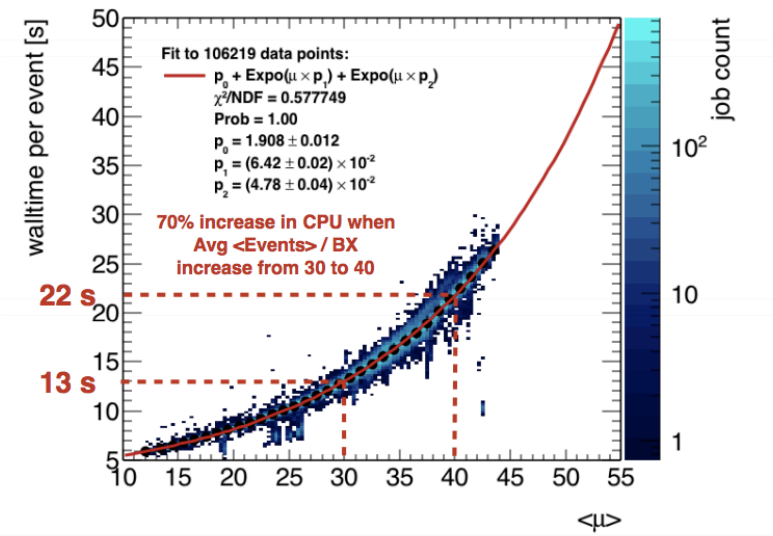
An event in 2016 …





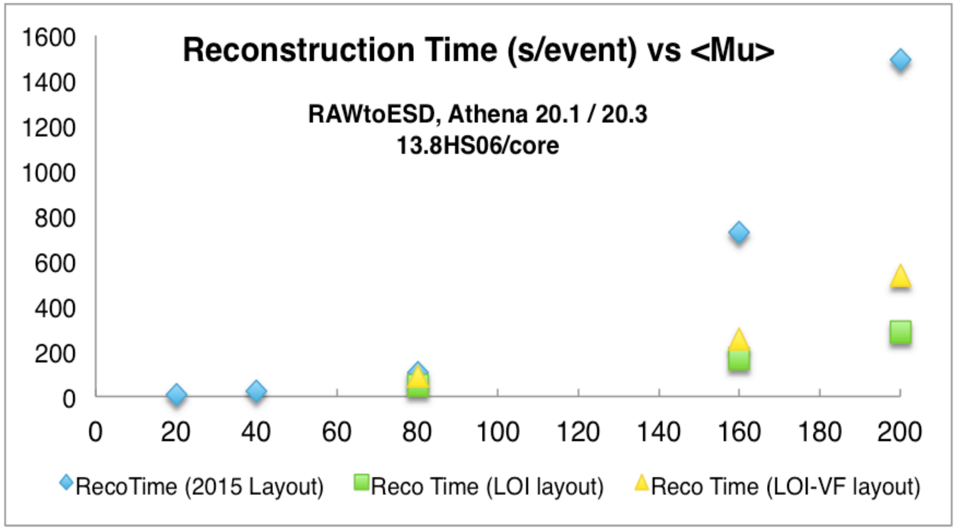
.. and a simulated event in 2025 with 200 vertexes

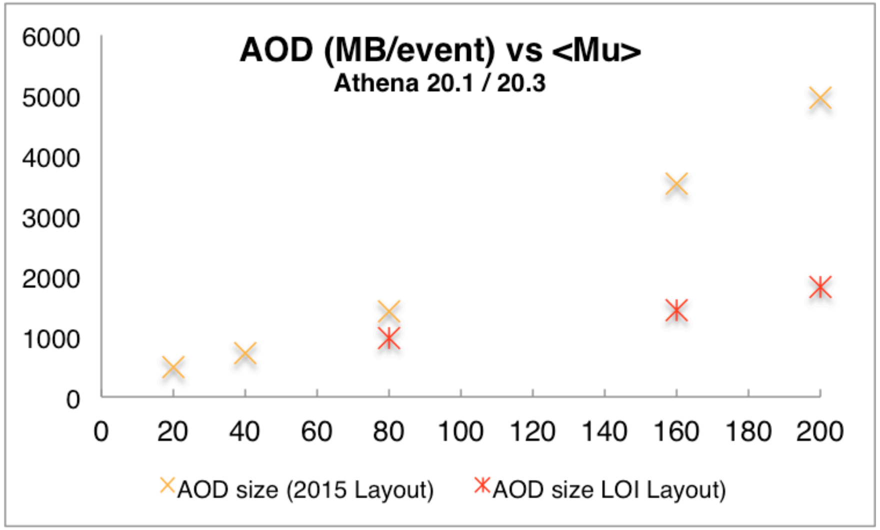
2016 1st pass reconstruction



Higher pileup means:

- Linear increase of digitization time
- Factorial increase of Reco time
- Larger events
- Much more memory

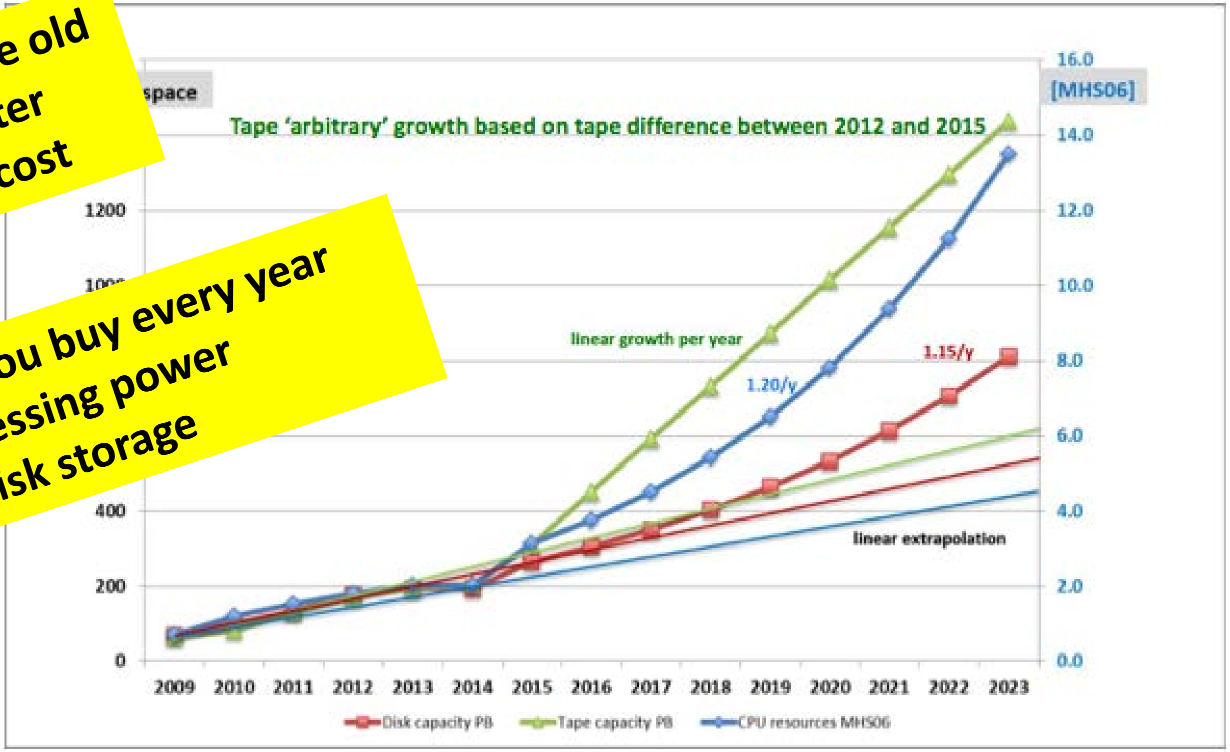




- The world's economy is not doing great and HEP can not overspend
- We consider a "**Flat budget**" scenario = same amount of funding for computing hardware every year
- Funding needs to cover the cost of hardware replacement

Fortunately, you replace old hardware with better one for the same cost

For the same money, you buy every year 20% more processing power 15% more disk storage



Input parameters, assumptions, disclaimers

Input Parameters at HL-LHC
(LOI = the ATLAS Letter of Intent for Upgrade Phase-2)

Output HLT rate: 10kHz (5 to 10 kHZ in LOI)
Reco time: 288s/event, Simul Time: 454 s/event at mu=200
Nr Events MC / Nr Events Data = 2
Fast Simulation: 50% of MC events
LHC live seconds /year: 5.5M

Simplified Computing Model with respect to 2016/2017 resource requests:

Data from previous years not taken into account
=> Little difference at the beginning of the Run-4 but huge difference for Run-2 and Run-3

Projection of available resources in HL-LHC:

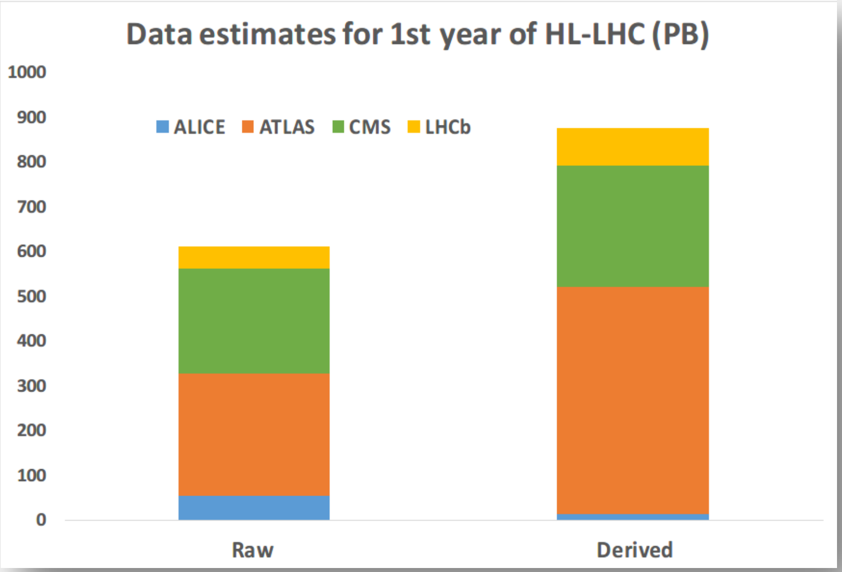
20% more CPU/year
15% more storage/year

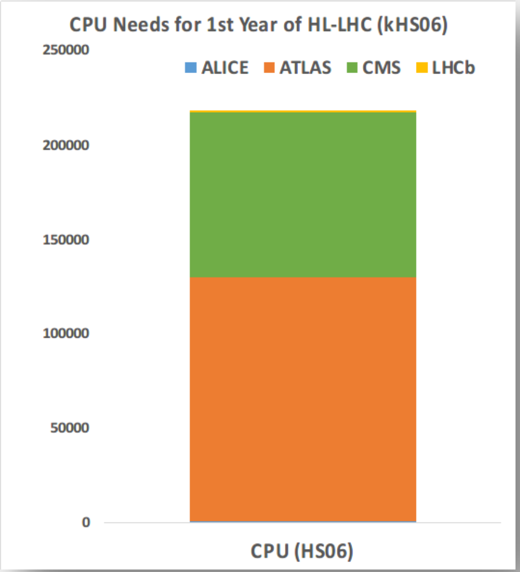
For the same cost

Projections evolve 2017 values **OF THIS SIMPLIFIED MODEL (not the 2017 WLCG pledges)**

Conclusion: looking at absolute numbers makes little sense.
Relative differences between needs and projections at HL-LHC are meaningful. With caveats.

Estimates of resource needs for HL-LHC





Storage

Raw 2016: 50 PB → 2027: 600 PB

Derived (1 copy): 2016: 80 PB → 2027: 900 PB

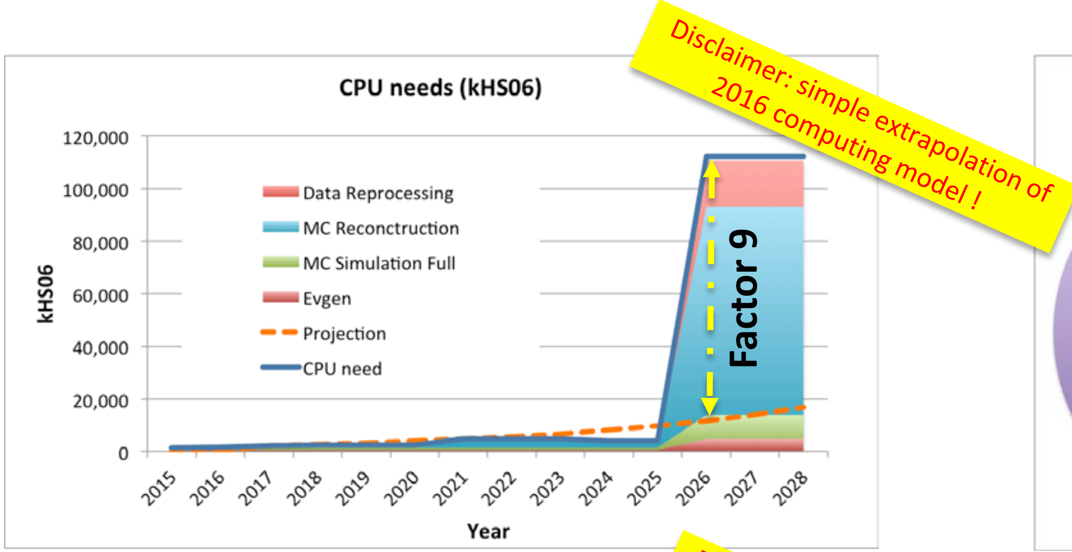
CPU

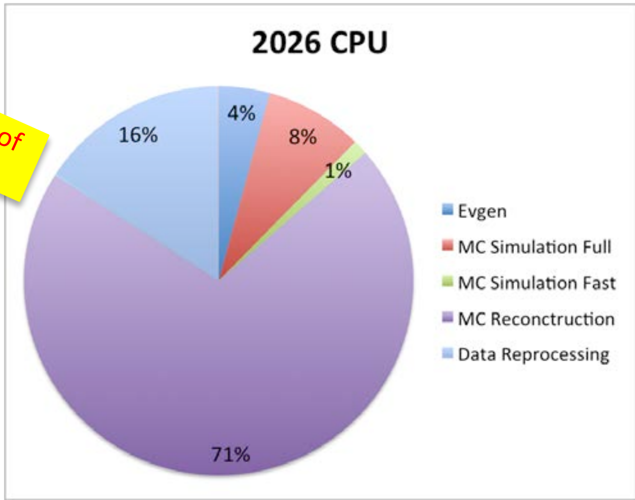
x60 from 2016

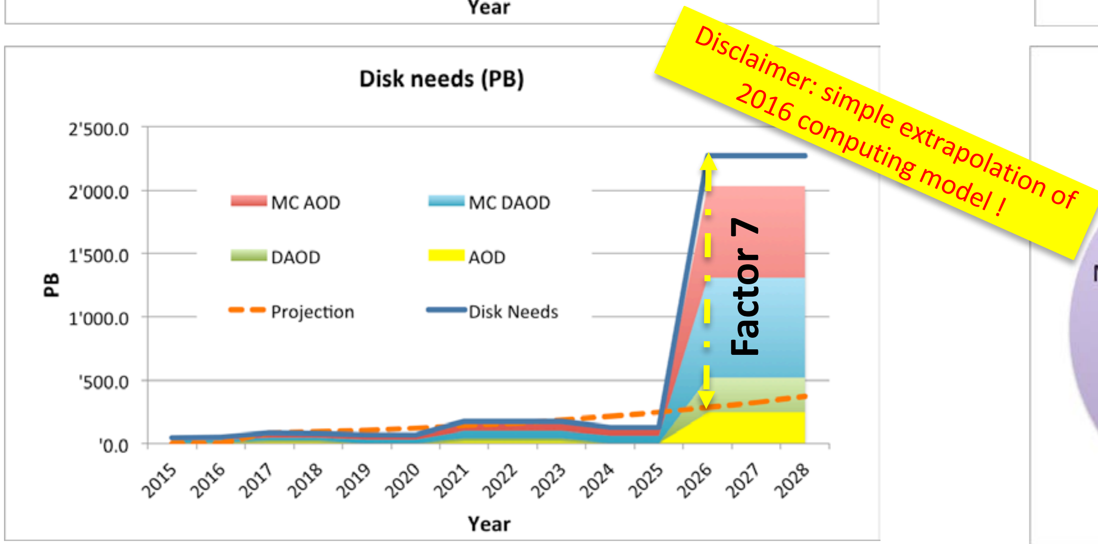
Technology at ~20%/year will bring **x6-10** in 10-11 years

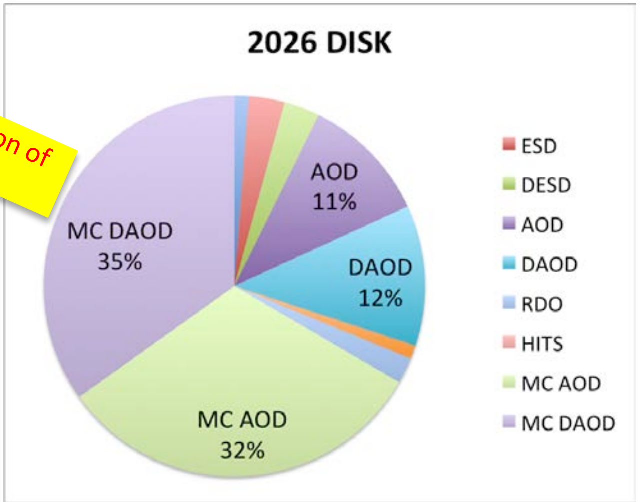
=> x10 above what is realistic to expect from technology with constant cost

HL-LHC baseline resource needs









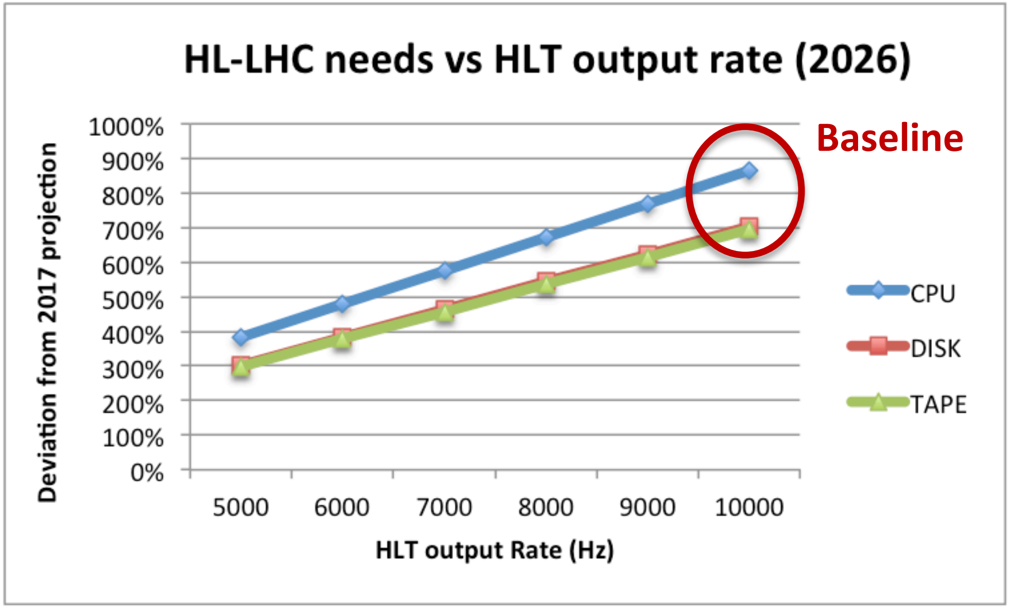
HLT output rate

The output trigger rate does not determine only the amount of data per year but also the amount of Monte Carlo to be produced.

The LOI foresees a value between 5 kHz and 10kHz. We use the latter as baseline in this study

The possibility to reduce the trigger rate to a lower value without impacting the ATLAS physics program will be analyzed in the years to come

If we consider the lower LOI limit (5kHz) the discrepancy with the projection of available resources reduces to x4 for CPU



Monte Carlo needs

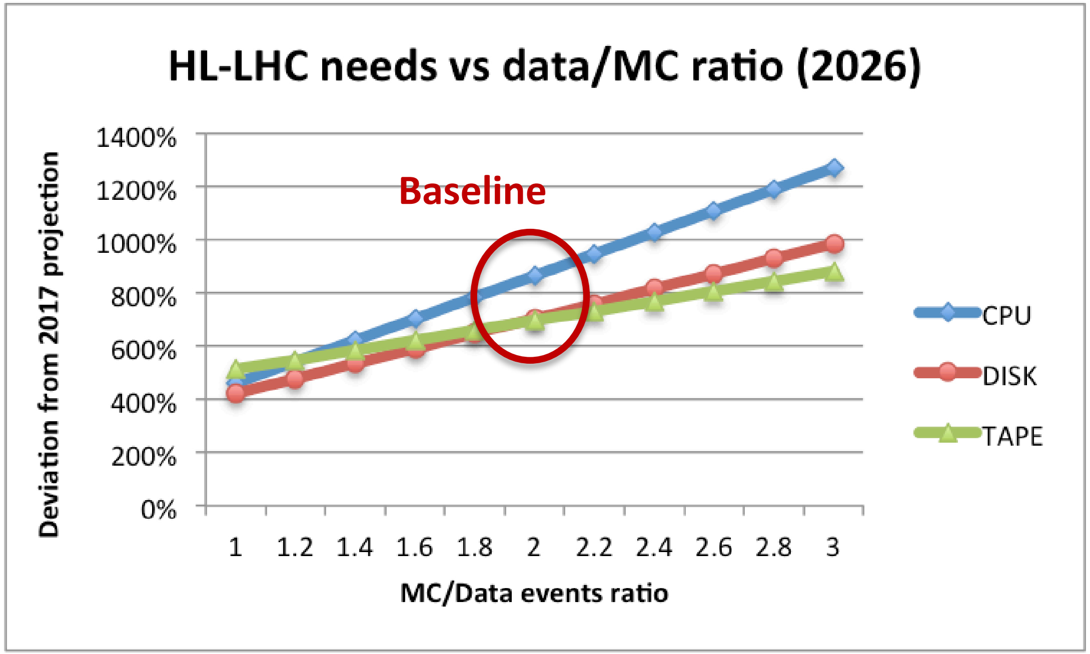
The physics case for HL-LHC will evolve in the next years. The high statistics of data collected in HL-LHC reduces the significance of statistical uncertainties. Therefore one might assume a lower need of MC with respect to data

<u>HOWEVER</u>

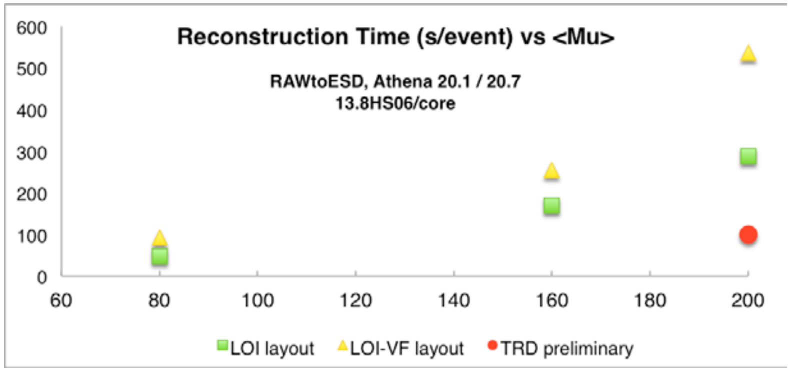
Things might change significantly once the physics case for HL-LHC evolves

Generators might become very expensive if we go to NNLO

In 2004 we expected a factor x0.3 MC with respect of data. We are at x2.0.



Layouts and Reconstruction



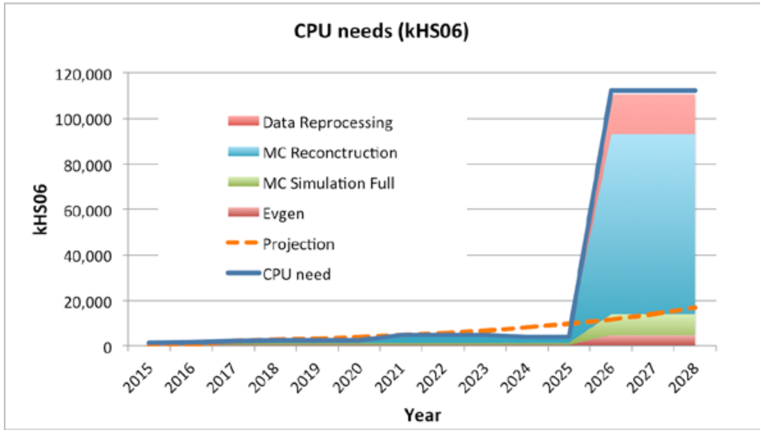
Reconstruction time dominates the CPU consumption in HL-LHC

Especially for MC, where trigger simulation utilizes the same offline algorithms (so it impacts twice as much)

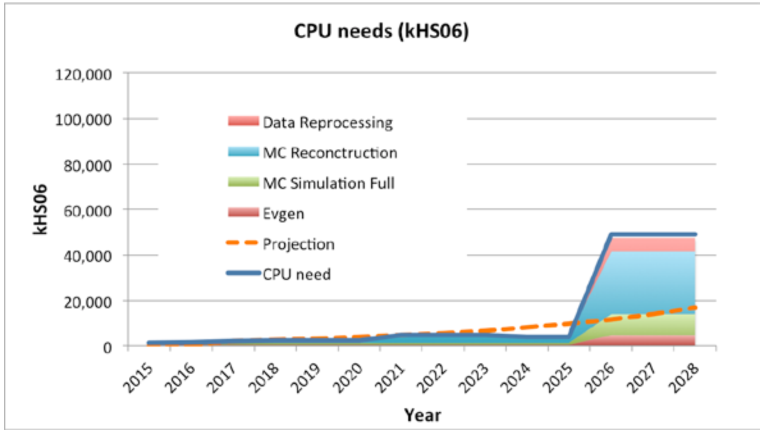
The detector layout will play an important role, together with the optimization/tuning of algorithms. Tracking will be the main consumer

Alternatives are also being investigated as R&D e.g. Machine Learning techniques

LOI Layout



Possible TDR Layout



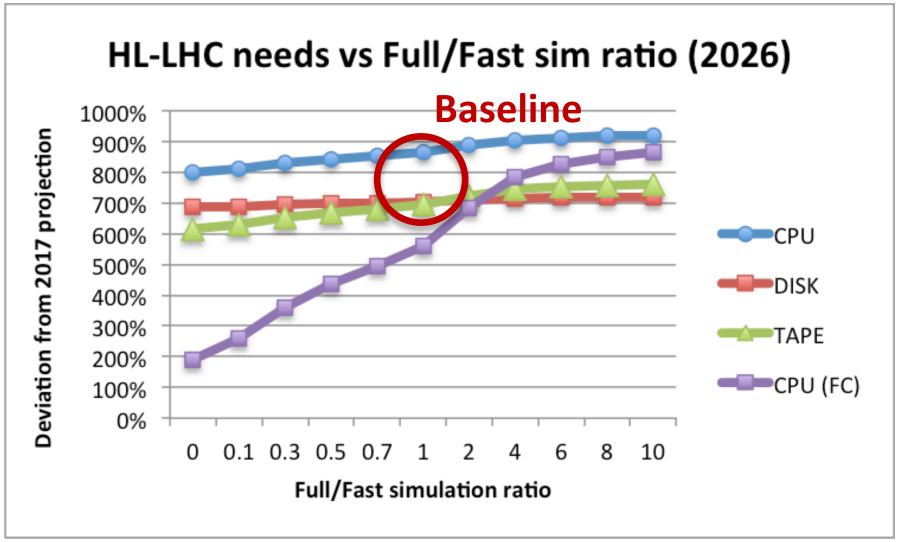
Fast Simulation and Fast Chain

Fast Simulation in Run-2 is x10 faster than Full Simulation (G4)

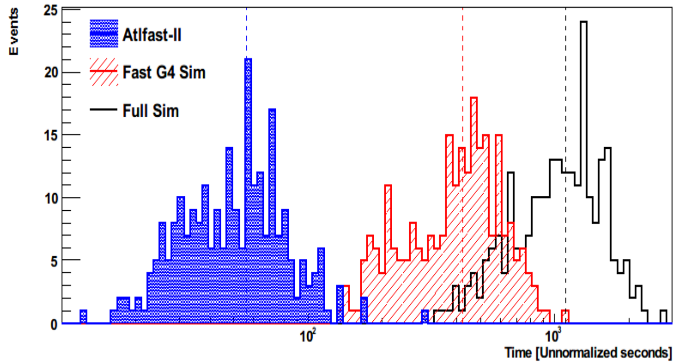
Fast Simulation can be used today only for a subset of analyses

Detector Simulation in general is not the driving cost in HL-LHC

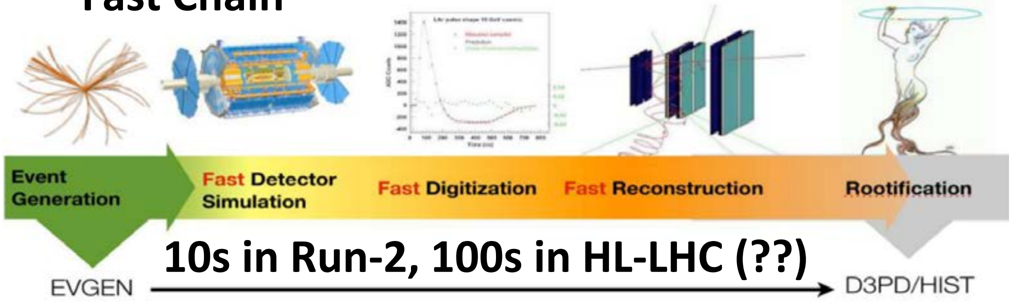
The gain will come with Fast Chain



Fast Simulation

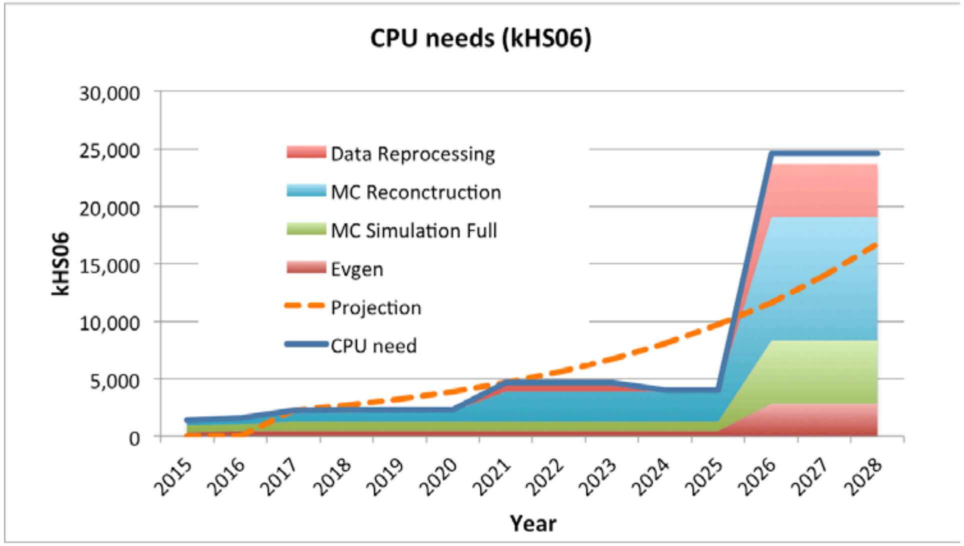


Fast Chain



10s in Run-2, 100s in HL-LHC (??)

If we want a very optimistic scenario …



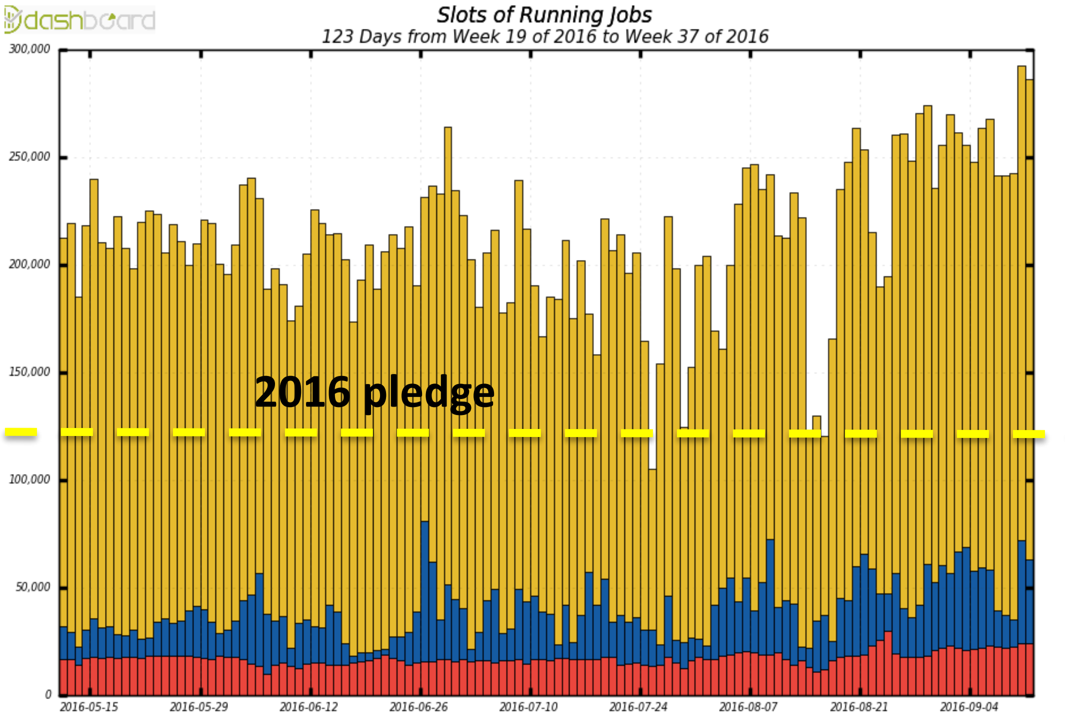
In a very optimistic scenario, the discrepancy for CPUs reduces to 200% (from almost 900%).

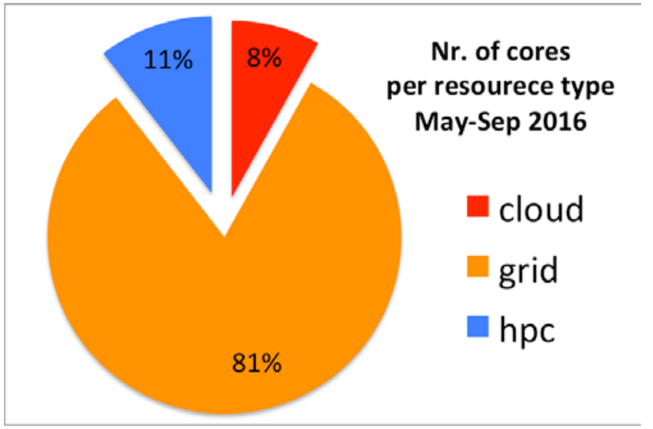
Which, given all the uncertainties, means problem solved

DO NOT GET TOO EXCITED AND LISTEN TO THE REST OF THE TALK

	Baseline Scenario	Optimistic Scenario
HLT output rate	10kHz	7.5kHz
Reco and Simul Time/Evt	from LOI	From preliminary TDR studies
Nr. Events MC / Nr. Events Data	2.0	1.5
Fast Simulation	50% of MC events	50% of MC events
Fast Chain	None	50% of MC events
LHC live seconds/year	5.5M	5.5M

Heterogeneous Resources





Integration of non Grid resources in ATLAS is a big investment with the potential of a big return

Challenges: resource provisioning, non standard architecture, GPU processing capacity, memory

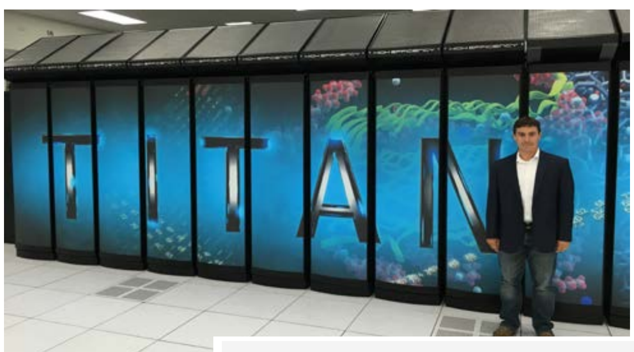
Challenges in HPCs utilization

High Performance Computers were designed for massively parallel applications (different from data intensive HEP use case) but we can parasitically benefit from empty cycles that others can not use (e.g. single core job slots)

The ATLAS production system has been extended to leverage HPC resources

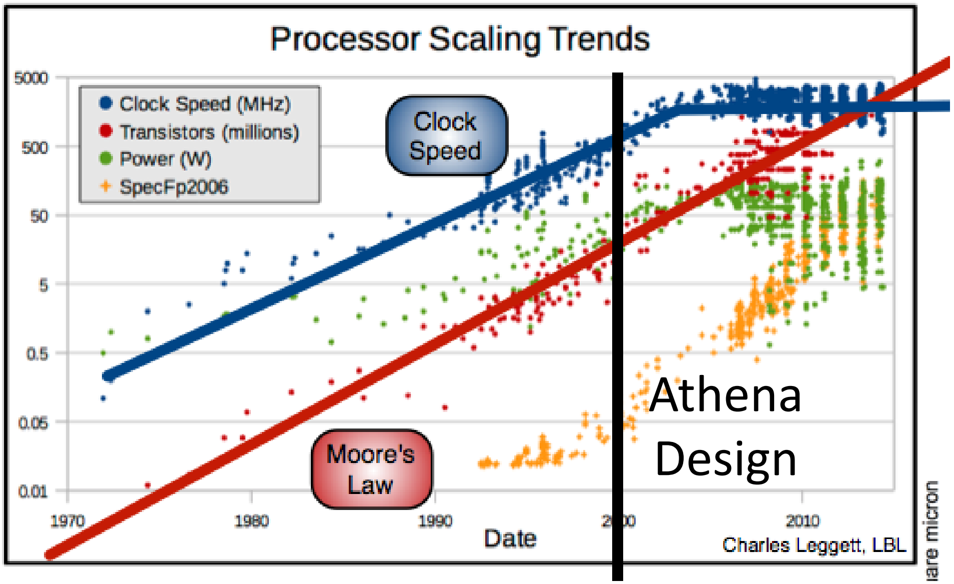
Heterogeneous site policies, inbound/outbound connectivity, #jobs/#threads, kind of grant/agreement

Blue Gene: PowerPC architecture



Theoretical Peak Floating-point Performance per node	166.4 Gigaflops (Intel® Xeon® E5-2670) 1311.0 Gigaflops (NVIDIA® Tesla® K20X)

Hardware trend and implications



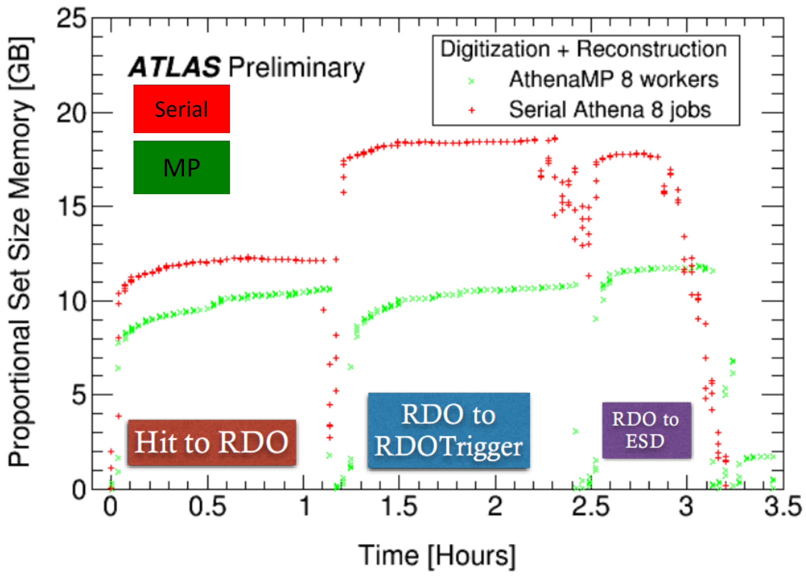
Clock Speed stalled but transistor density keeps increasing. Exploiting hardware becomes more complicated (vectors, memory…)

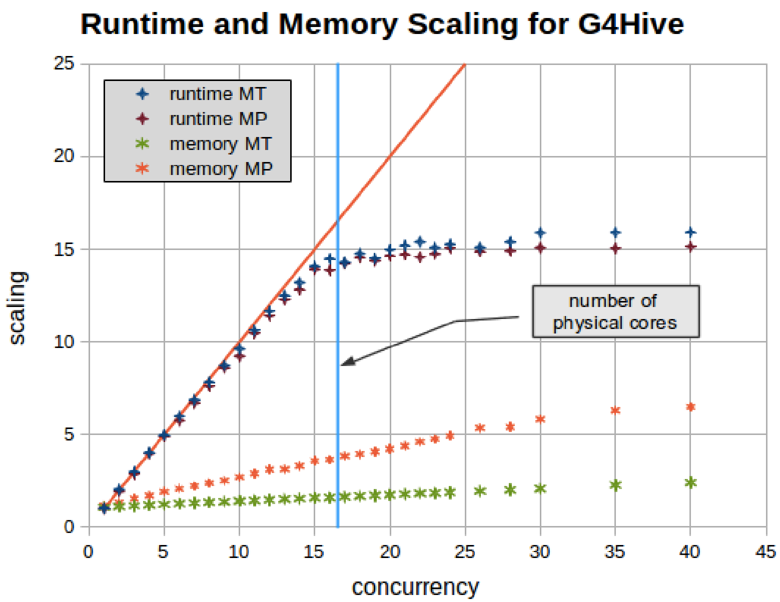
Example: Cori@NERSC (Intel Knights Landing)
1PB of Memory, 9304 nodes
68 cores/node, 4 HW threads/core
=> Approx 300 MB/thread

From Multi Processing to Multi Threading

AthenaMP (multiprocessing) will not be sufficient anymore. We will need (and we are developing) AthenaMT (multithreading). Will be in production for Run-3 (2020) already.

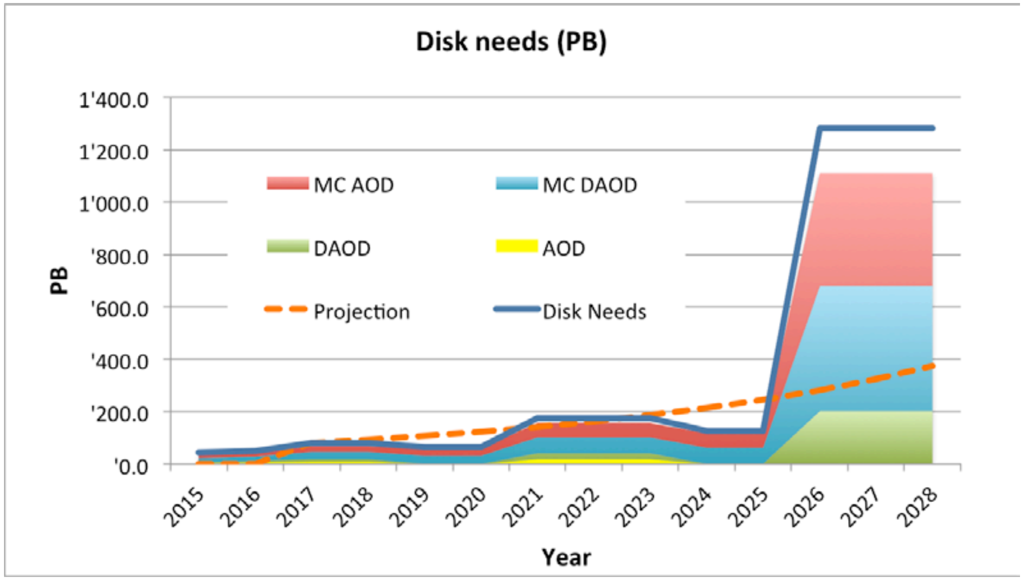
Parallel processing in a multithreaded environment will come with its challenges both for developers, operations and infrastructures





What about Storage ?

Optimistic Scenario + No AOD on disk



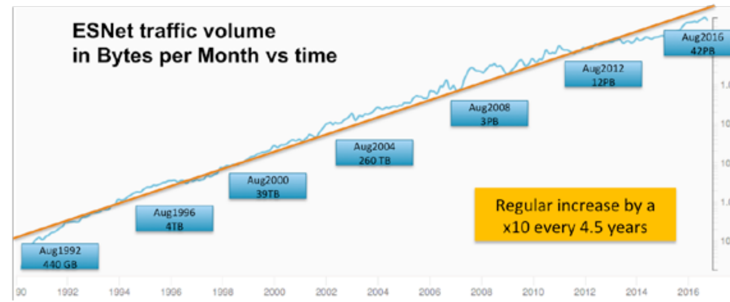
Even in the optimistic scenario, we are still far from solving the problem

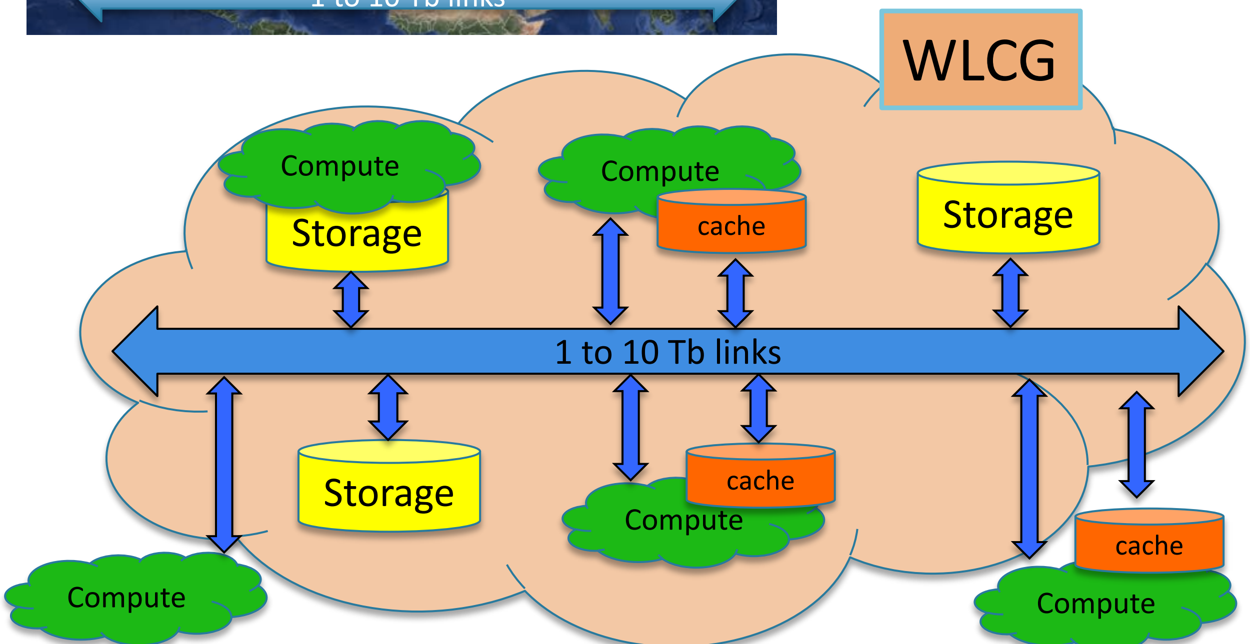
AODs and DAODs are the main consumers.

With no AOD on disk (run Train Analysis from AODs on TAPE) you get x4 above the resource projection

The remaining gain must come from re-thinking of distributed data management, distributed storage and data access. A network driven data model allows to reduce the amount of storage, particularly for disk. Tape today costs at least 4 times less than disk.

Computing infrastructure in HL-LHC





A data cloud for science

Storage and Compute loosely coupled but connected through a fast network

Heterogeneous Computing facilities (Grid/Cloud/HPC/ …) both in and outside the cloud

Different centers with different capabilities, fo different use cases

Conclusions

- HL-LHC will present unprecedented computing challenges
- To keep cost of computing under control in 2026 we need to invest effort from now
- The effort spans many areas: online, offline software, distributed computing, physics, infrastructure and facilities. The detector layout will play a crucial role
- It is important to consider cost of computing when choices are made
- HEP will need to adapt to market trends, therefore flexibility is the key