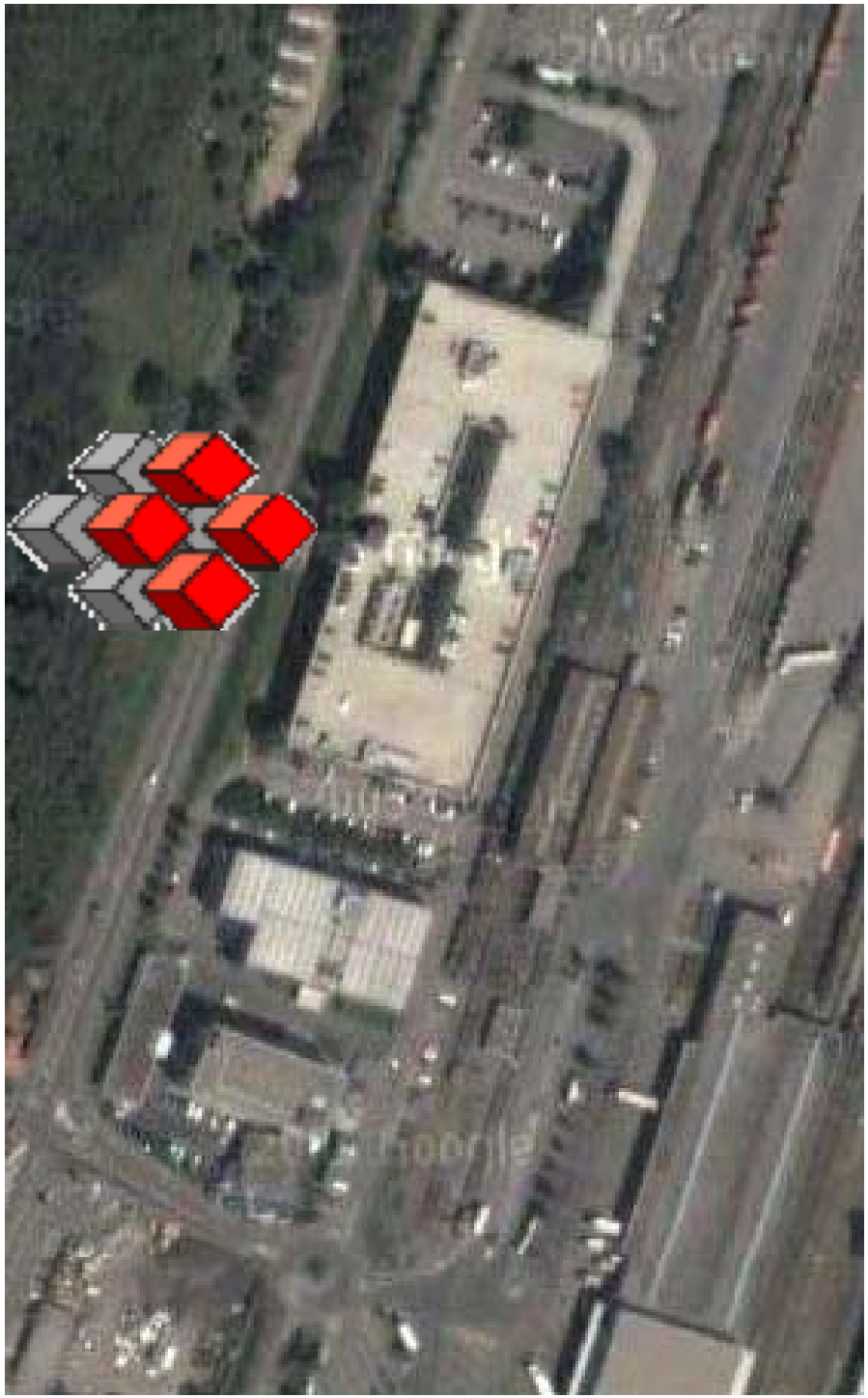


Swiss Institute of Particle Physics

LHC Computing and Analysis Workshop

Physics Cases - ATLAS

25-26 August 2005

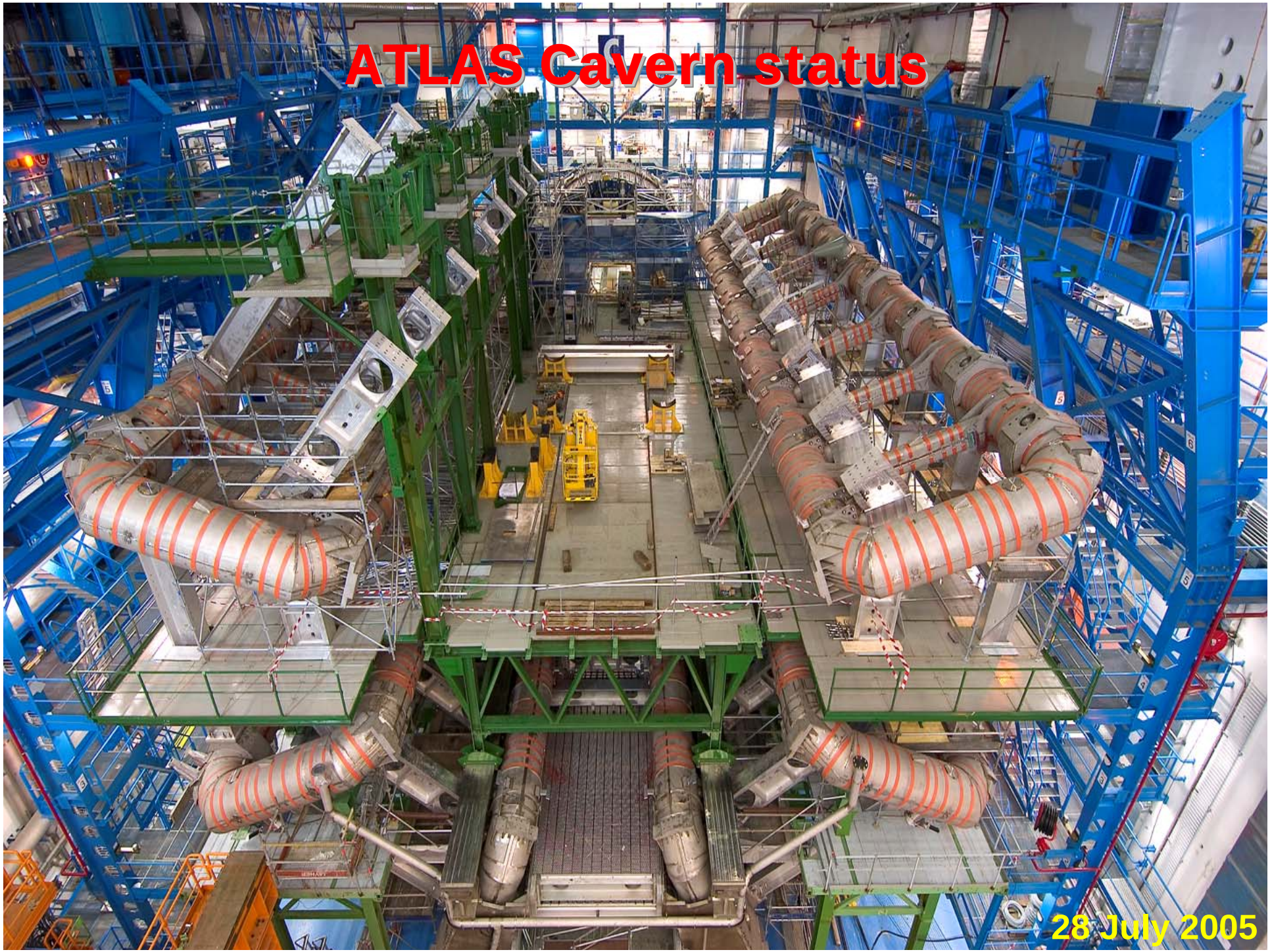
CSCS

Manno

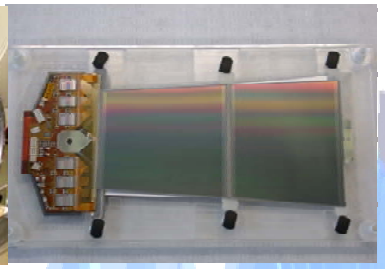
HansPeter Beck

LHEP – Uni Bern

ATLAS Cavern status



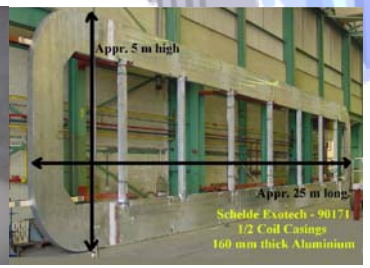
Swiss Contribution to ATLAS



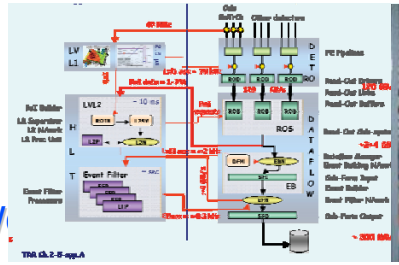
Semiconductor Tracker
Barrel Support Structure
Forward Modules
Geneva

Calorimeter
Readout Electronics
Geneva

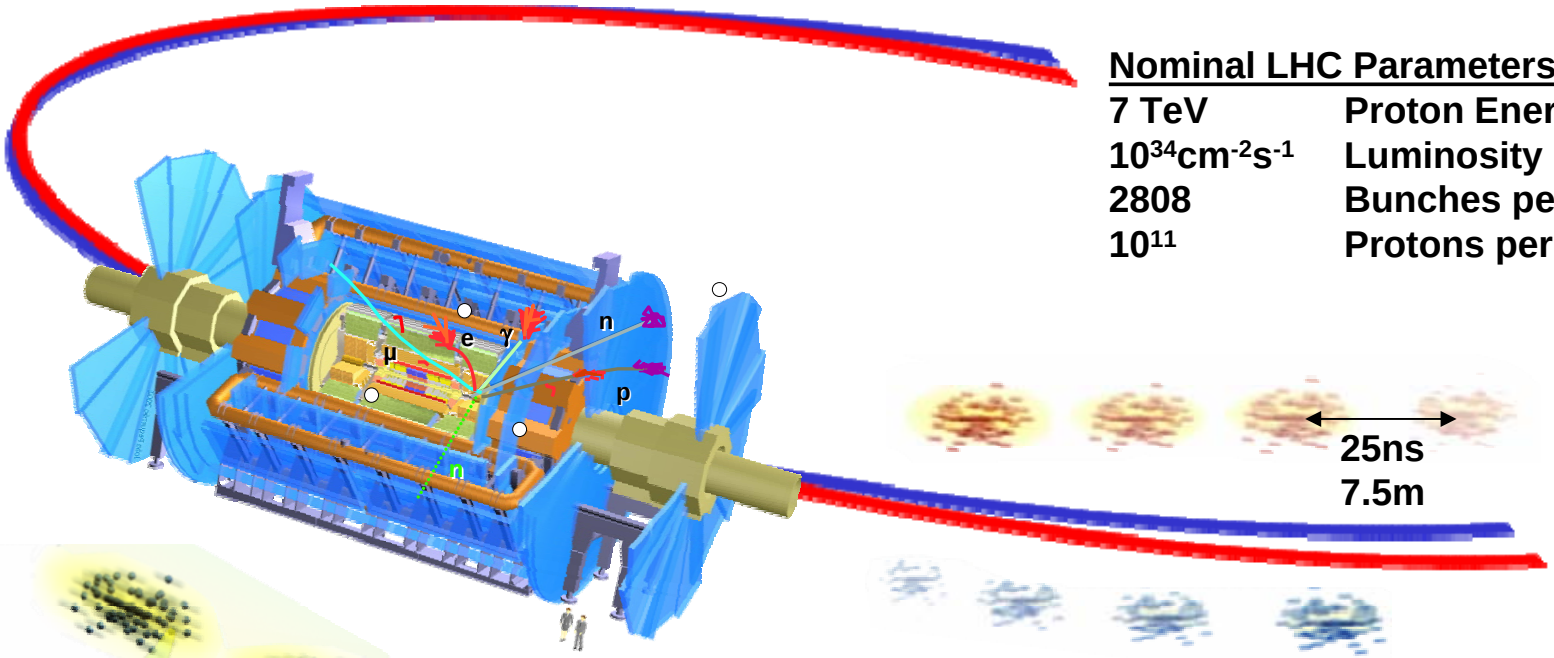
Trigger & DAQ
Dataflow
High Level Triggers
Bern



Barrel Toroid
Superconductor
Coil Casings
Bern and Geneva



p-p Collisions @ LHC

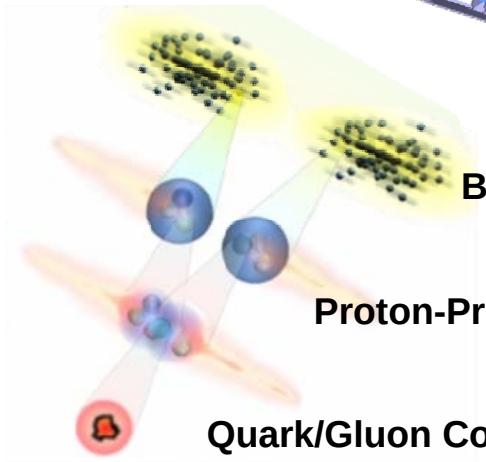


Nominal LHC Parameters:

7 TeV	Proton Energy
$10^{34}cm^{-2}s^{-1}$	Luminosity
2808	Bunches per Beam
10^{11}	Protons per Bunch

25ns
7.5m

0.6 A

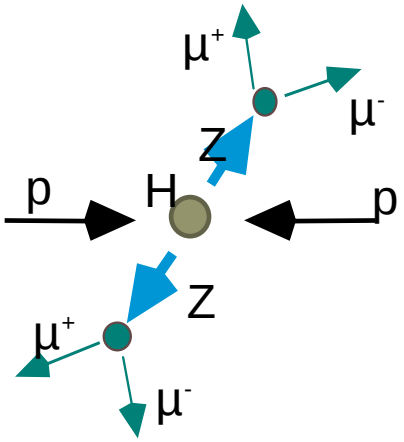


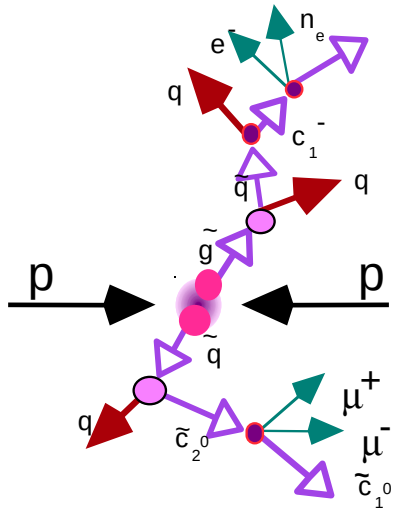
Bunch Crossings $4x10^7$ Hz

Proton-Proton Collisions 10^9 Hz

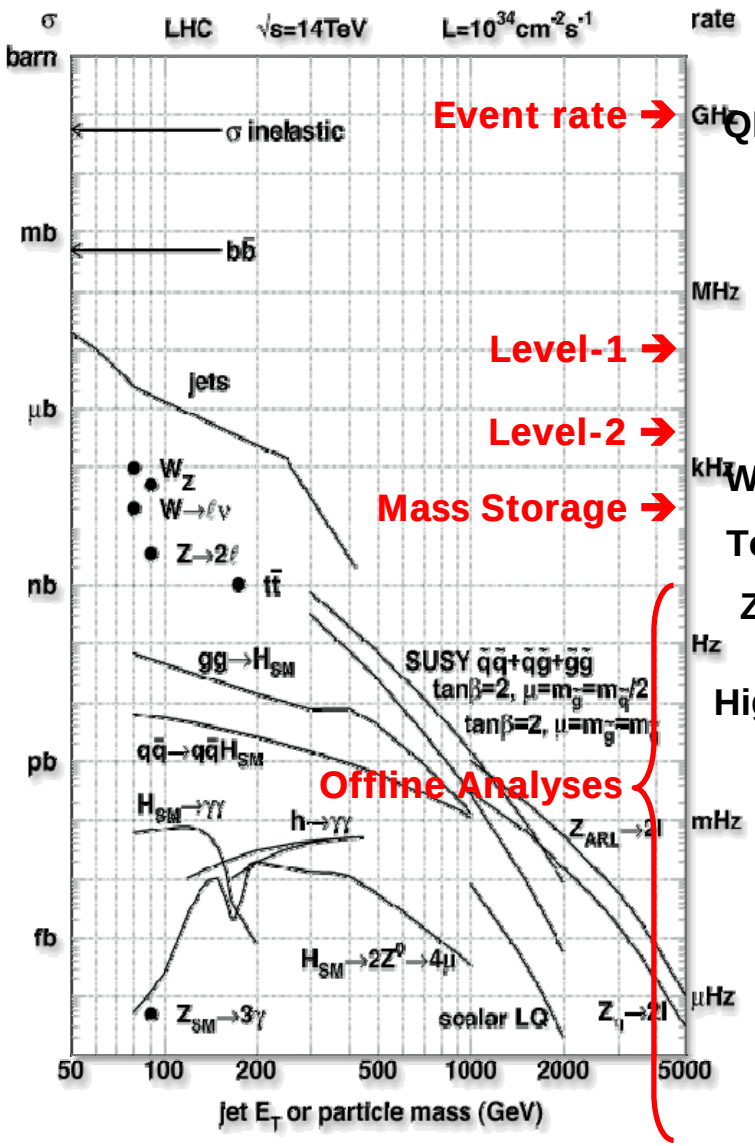
Quark/Gluon Collisions

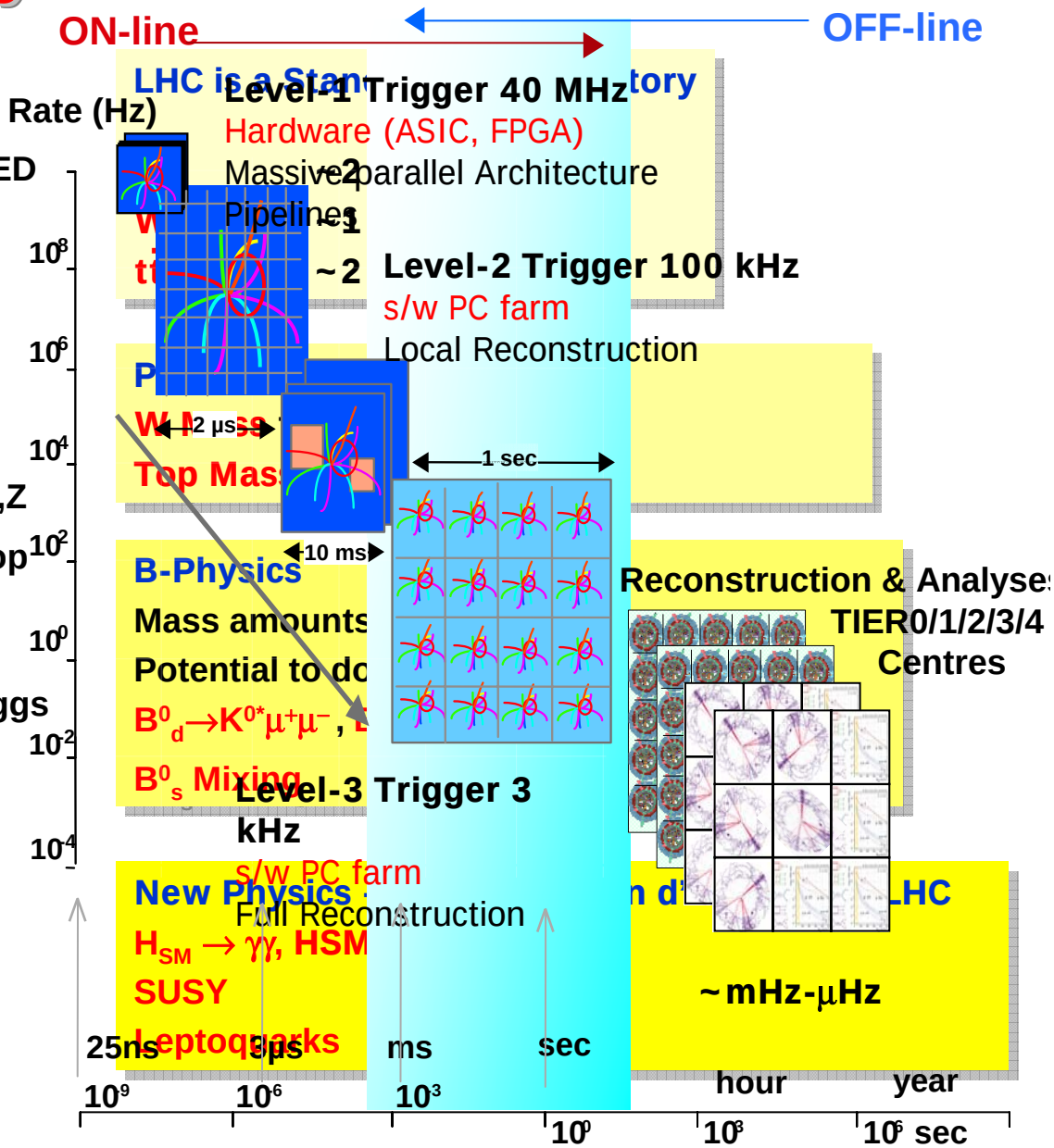
Heavy particle production $10^{+3...-7}$ Hz
(W, Z, t, Higgs, SUSY,…)





Physics Rates



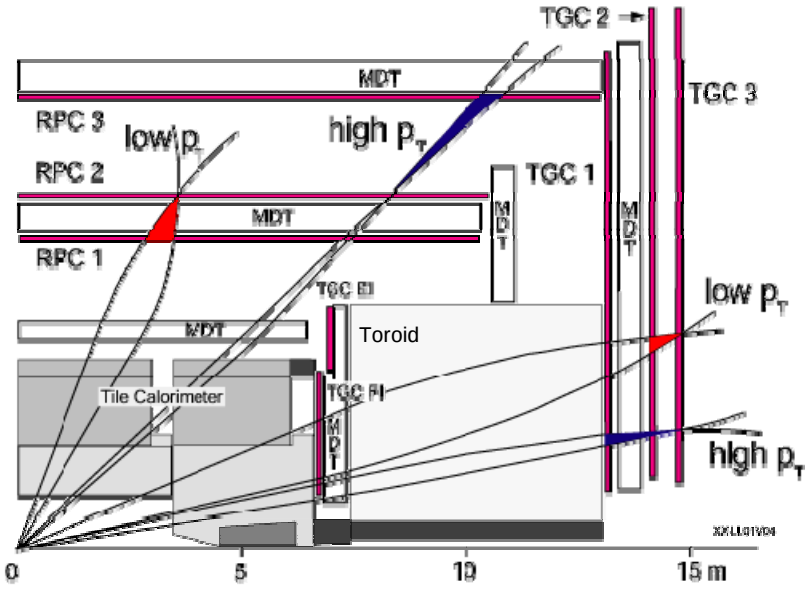


Physics Selection Strategy

- **ATLAS has an inclusive trigger strategy**
  - **LVL1 Triggers on individual signatures / objects**
    - EM / Had Cluster
    - Total Energy
    - Missing Energy
    - Muon track
  - **LVL2 confirms & refines LVL1 signature**
    - seeding of LVL2 with LVL1 result – i.e. Region of Interest [RoI]
  - **EventFilter confirms & refines LVL2 signature**
    - seedig of EventFilter with LVL2 result
    - tags accepted events according to physics selection

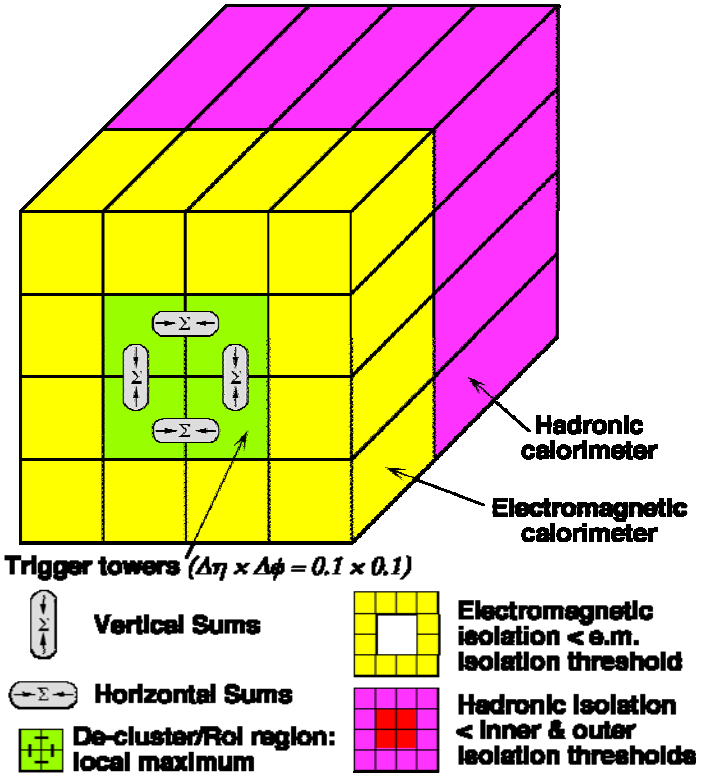
- **Offline Analysis is based on trigger samples**
  - **an individual analysis will always run over a (tag) of events**
  - **need to understand trigger object selection efficiencies**

LVL1 - Muons & Calorimetry



Muon Trigger looking for coincidences in muon trigger chambers
2 out of 3 (low-p_T; >6 GeV) and 3 out of 3 (high-p_T; > 20 GeV)

Triger efficiency 99% (low-p_T) and 98% (high-p_T)



Calorimetry Trigger looking for $e/\gamma/\tau +$ jets

- **Various combinations of cluster sums and isolation criteria**
- **$\Sigma E_T^{em,had}$, E_T^{miss}**

LVL1 Trigger Rates

Selection		$2*10^{33}$ cm^{-2}s^{-1}	10^{34} cm^{-2}s^{-1}
MU20	(20)	0.8	4.0
2MU6		0.2	1.0
EM25I	(30)	12.0	22.0
2EM15I	(20)	4.0	5.0
J200	(290)	0.2	0.2
3J90	(130)	0.2	0.2
4J65	(90)	0.2	0.2
J60 + xE60	(100+100)	0.4	0.5
TAU25 + xE30	(60+60)	2.0	1.0
MU10 + EM15I		0.1	0.4
Others (pre-scales, calibration, ...)		5.0	5.0
Total		~ 25	~ 40

- Rates given in kHz — No safety factor included!
  - → E_T thresholds imply 95% efficiency values

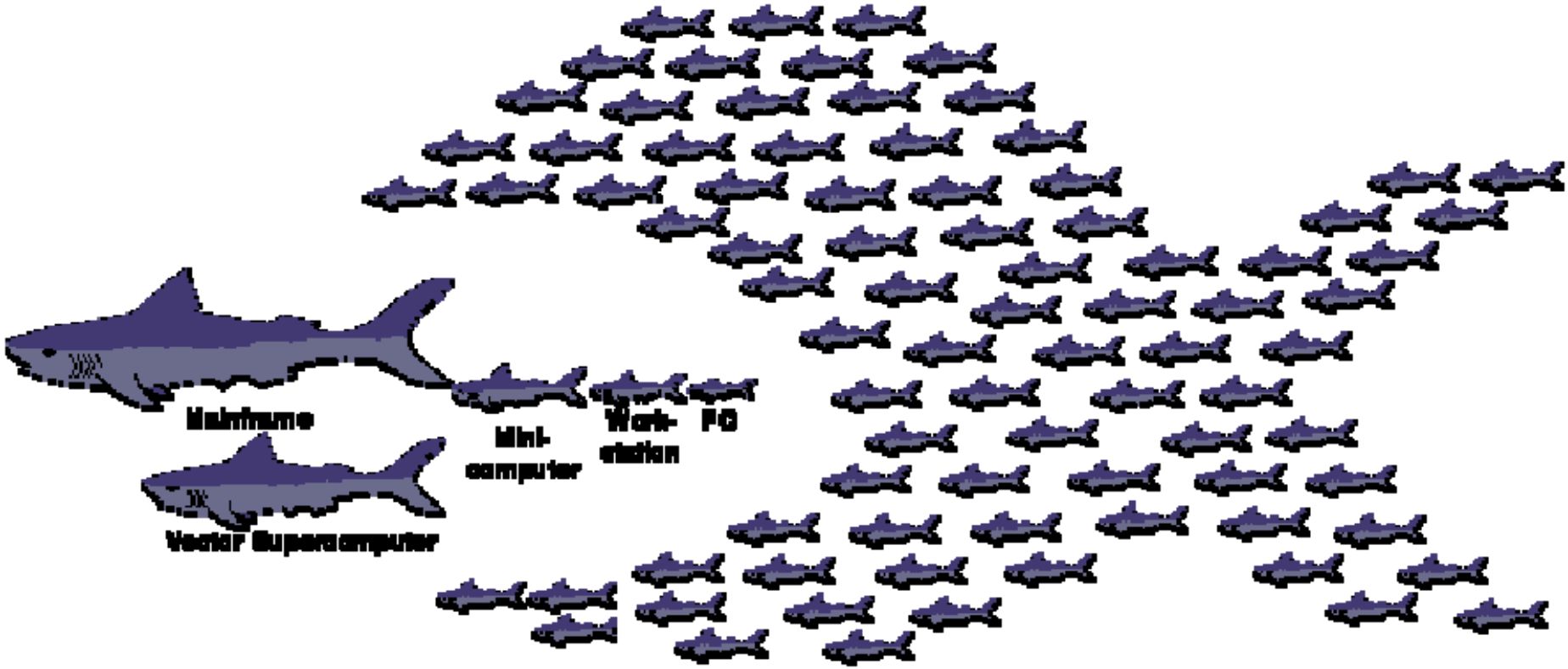
LVL1 rate is dominated by electromagnetic clusters: 78% of physics triggers

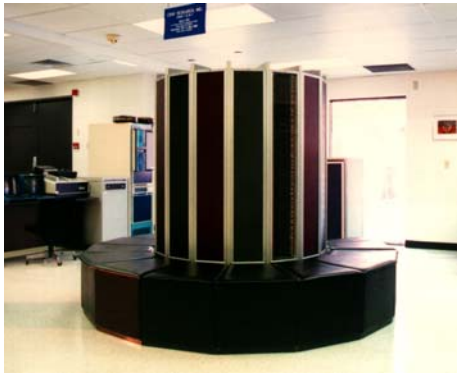
Inclusive Higher Level Trigger Event Selection

Selection	$2x10^{33}$ $cm^{-2}s^{-1}$	Rates (Hz)
Electron	e25i, 2e15i	~40
Photon	γ60i, 2γ20i	~40
Muon	μ20i, 2μ10	~40
Jets	j400, 3j165, 4j110	~25
Jet & E_T^{miss}	j70 + xE70	~20
tau & E_T^{miss}	τ35 + xE45	~5
B-physics	2μ6 with m_B / $m_{J/\psi}$	~10
Others	pre-scales, calibration, ...	~20
Total		~200

HLT rate reduces e/γ a lot: 45 % of physics triggers

DAQ/HLT/TIER0-1-2-3 are all based on PC Farms





ATLAS Event Size

Inner Detector	Channels	Fragment size - kB
Pixels	1.4×10^8	60
SCT	6.2×10^6	110
TRT	3.7×10^5	307

Muon Spectrometer	Channels	Fragment size - kB
MDT	3.7×10^5	154
CSC	6.7×10^4	256
RPC	3.5×10^5	12
TGC	4.4×10^5	6

Calorimetry	Channels	Fragment size - kB
LAr	1.8×10^5	576
Tile	10^4	48

Trigger	Channels	Fragment size - kB
LVL1		28

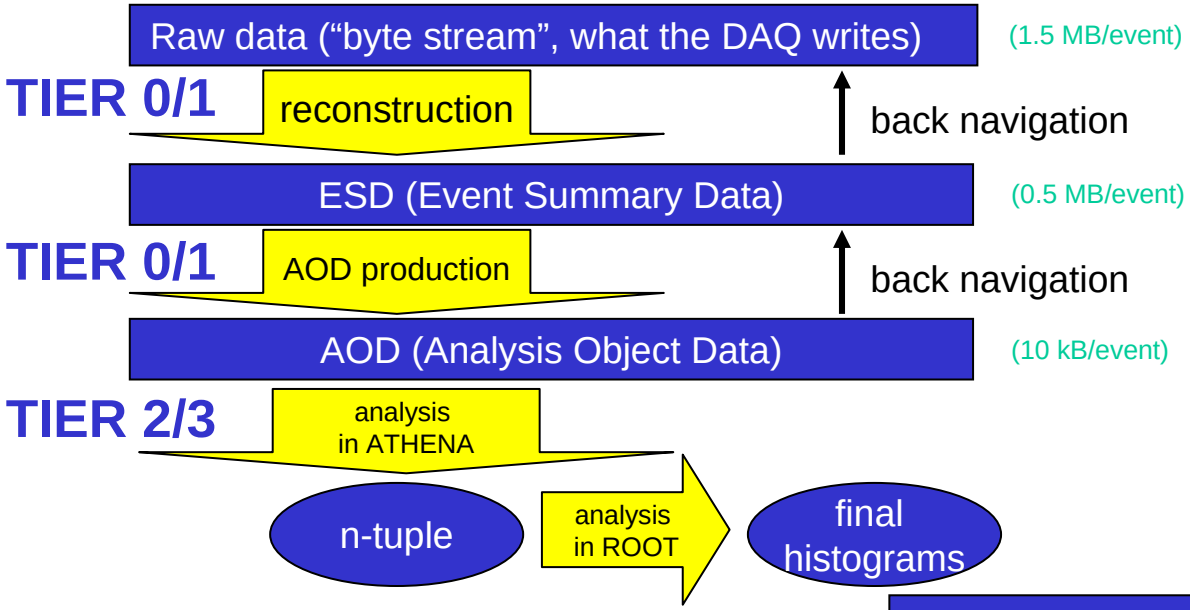
Atlas event size: 1.5 MBytes
140 million channels
organized into ~1600 Readout Links

Affordable mass storage b/w:

300 MB/sec A 200 Hz

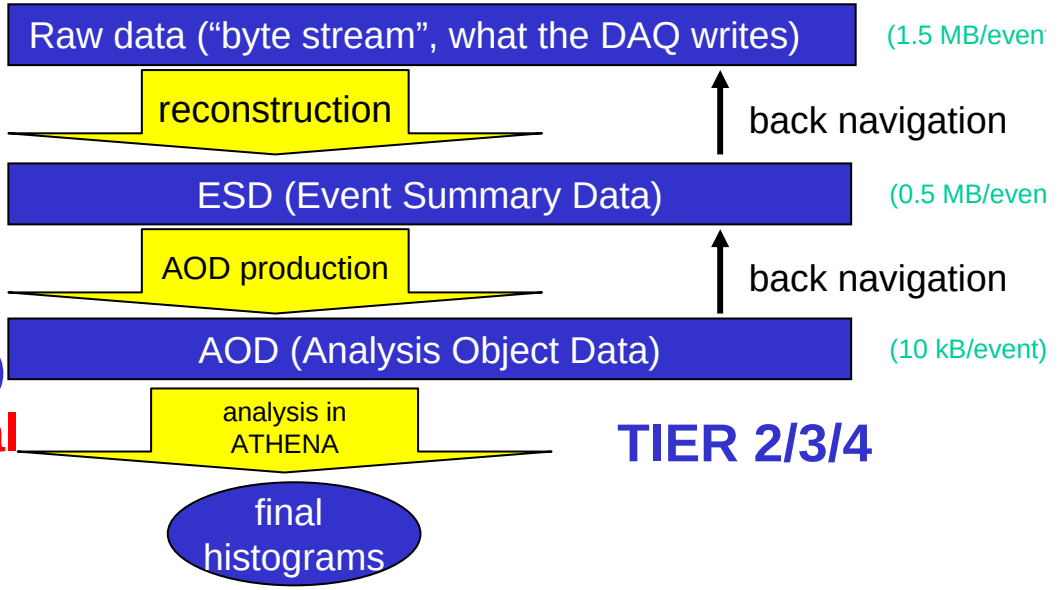
☛ **3 PB/year**
of carefully triggered data
for offline analysis

How to Analyze 3 PB/year while still getting the Physics out?



Today:
Joe Physicist analyzing ROOT n-tuples

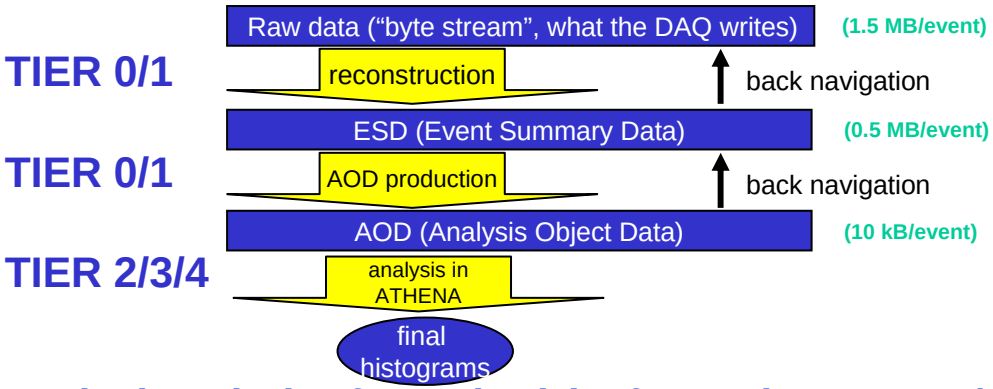
Tomorrow:
Joe Physicist analyzing AODs directly
(using ATHENA, Python, ROOT,…)
today still too slow to be practical



Analysis of ATLAS data (cont)

- **What Joe Physicist (aka Heiri / Henri Füsikus, Vreni / Giselle Füsika,) will do**
  - **Joe is organized in a physics working group**
    - e.g. Higgs
  - **Joe is looking at a specific / inclusive channel**
    - e.g. Higgs → four electrons
  - **Joe needs to understand how electrons are**
    - **triggered**
      - LVL1, HLT
    - **reconstructed**
      - efficiencies
      - fakes
    - **receive there measured properties**
      - 4vector; i.e. energy and 3 momenta
      - error matrix (correlations)
  - **Joe runs over many different data samples**
    - real data on tight trigger selection (full statistics)
    - Background samples (i.e. complementarory trigger samples; e.g. jet-samples)
    - real data on loosened trigger samples (full statistics not needed – need to keep the systematical errors under control)
    - Monte Carlo data (needs to be generated first, simulated, reconstructed)
      - MC produced ATLAS wide via Higgs group
      - Special dedicated MC samples produced by Joe
      - amount defined by precision needed. balance statistical and systematica error
  - **With increasing understanding of his task, Joe will do this on**
    - ntuples full statistics
    - AODs full statistics
    - ESD maybe full statistics –but more likely only a moderate subset of ESD needed
    - RAW only a small subset a RAW needed

Analysis of ATLAS data (cont)



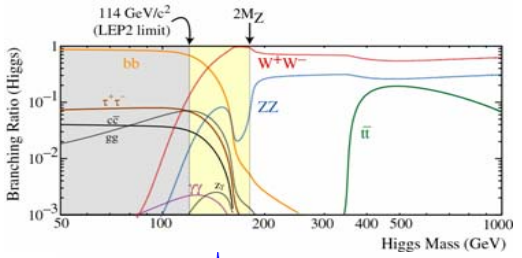
- **Final Analysis of Joe Physicist focused on TIER 2/3/4 centres**
  - **TIER4 (Laptop): data presentation, job preparation, coding of user specific code**
  - **TIER3/4: running over small samples**
    - **skimmed data** providing full statistic (specialized AODs or ntuples)
    - **test samples** (AOD, ESD, RAW) small fraction for debugging and testing
    - producing some dedicated (Joe's) Monte Carlo data (gen, sim, rec, ESD, AOD)
  - **TIER2: running over big samples**
    - full statistics of Joe's pre-selection (AOD and sometimes even ESD) needs to be available
      - real data and Monte Carlo data
    - fast data movement from TIER1, enough storage at TIER2
    - Joe producing a lot of dedicated Monte Carlo data (gen, sim, rec, ESD, AOD)
- **TIER 1/2**
  - **ATLAS physics groups decide on global production needs**
    - i.e. production of a very big Higgs sample can be proposed by Joe – but priorities will be decided ATLAS wide
- **Joe Physicist is only competitive if these TIERs give him high priorities for his jobs. I.e. do not try to do any of above on lxplus at Cern**
- **There is not only Joe there, also Vreni and Heiri have their own and independent analysis...**

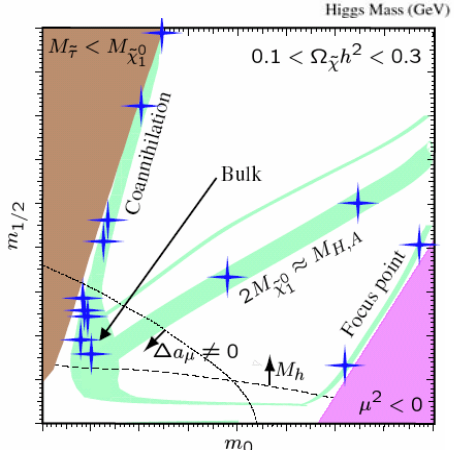
Swiss ATLAS Groups will do....

- **Exploiting ATLAS Physics potential** — **TIER2/3**
  - **based on tagged AOD (and ESD) data**
  - **specific trigger samples, but full statistics**
  - **every physicist has his own preferred trigger samples**
    - Higgs, SUSY, Beyond Standard Model, Standard Model
    - Heavy Ion ??
- **Detector studies** — **TIER2/3**
  - **based on AOD, ESD and RAW data**
  - **calibration, alignment, efficiencies**
  - **events from calibration stream**
  - **often, full statistic sample not needed (a sub-sample is often enough=**
    - **SCT & LAr**
    - more systems will be needed in the course of analyzing data
  - **RAW data will just be accessed on a small sub-sample of the events available – but needed for debugging/improving of reconstruction code**
- **Trigger studies** — **TIER2/3**
  - **based on AOD and ESD data**
  - **efficiencies**
    - specific trigger samples and control samples
    - needed for LVL1, LVL2 and EF
- **Based on Real Data and Monte Carlo Data** — **TIER 1/2/3**
  - **Real Data needs to be processed and re-processed (calibration, conditions and alignment)**
  - **Monte Carlo Data needs to be produced and processed and re-processed (calibration, conditions, alignment)**

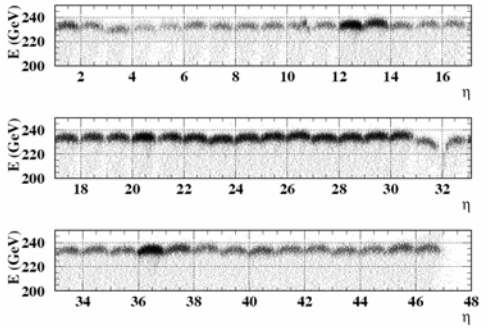
Swiss ATLAS Analysis
Actual Examples and Experiences…

- Monte Carlo based studies of
  - **Offline and Trigger analyses**
    - H→ZZ* → 4e
    - H → ZZ* → 2e2μ
    - H → WW* → 2e2ν via VBF
  - **SUSY**
    - Participation in DataChallenges for MC production
    - sparticle masses (edges) at various mSUGRA points. Including stop and stau coannihilation; bulk region; etc.





- Real Data studied from Combined Test Beam
  - **LAr energy calibration**
  - **electron / pion separation**
  - **track reconstruction**



… and Problems seen on Phoenix, Ubelix,…

- **Releases installation**
  - **Many flavors of ATLAS s/w on the market, caused confusion**
  - **libraries (system + ATLAS specific) were missing**
  - **differences seen between OS versions**
    - Reco (ESD) 9.0.4 went to infinite loop with SL3 build, Not with RH73 build (under SL3!) (Phoenix)
    - pythom boost lib libboost_python-gcc.so changed by admin triggered crashes on 32 bit machines
    - /usr/bin/time directory missing in some Ubelix nodes
- **Grid**
  - **NFS problems with nordugrid/Ubelix front end. (NFS failed also on some Phoenix nodes)**
  - **Problems in handling too many jobs submission, even if below the (500) limit(lheppc10).**
  - **#time limit of 500' from some hidden default settings in SGE scheduler, fixed (Ubelix).**
  - **gridftp server problems (various, wrong permission) (Ubelix)**
  - **Oracle DB at cern get stuck (too many requests)**
- **Atlas Software**
  - **Memory leak in reco, memory exceeds requirements (1GB). Problem in 9.0.4. Still not solved in**
  - **10.0.1. Workaround: require 1.5Gb to reco 50 evt.**
  - **Tauola bug (prodution restarted from scratch)**
- **Hardware**
  - **PCs that did not supported the load, over-heated, etc.**

What Joe Physicsist Expects

applies also to Heiri/Henri Phüsikus and Vreni/Giselle Phüsika

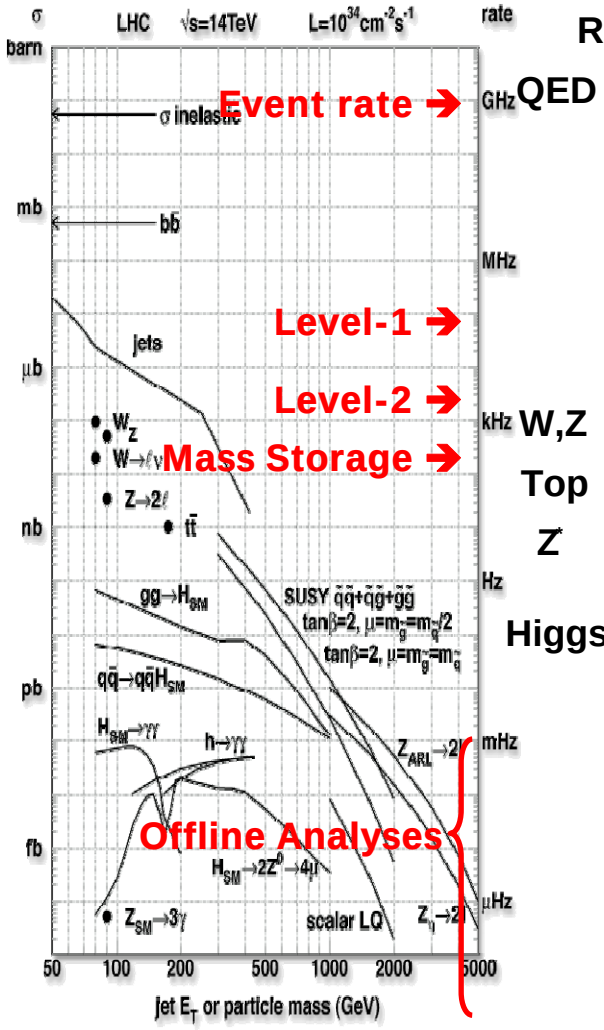
- **Analysis of ATLAS data is a complex procedure**
  - **need to go not only over ntuples or AOD data**
    - but some access needed to ESD and RAW data
  - **easy use of middleware**
    - i.e. Joe Physicist doesn't even want to know about
  - **available ATLAS s/w releases**
    - more than one in parallel
      - » Joe may need version x while Vreni wants y and Henri z
  - **local Database access**
    - condition, calibration, alignment data
  - **infrastructure robust against s/w still under development**
    - memory leaks, infinite loops, crashes
  - **flexible environment**
    - transparent use of TIER 4, TIER 3 and TIER 2 usage
      - Joe does not want to learn every time a new environment
  - **availability of data**
    - at various TIERs
- **Independence from Cern lxplus/lxshare**
  - **provided there are enough resources**
    - CPU, storage, bandwidth

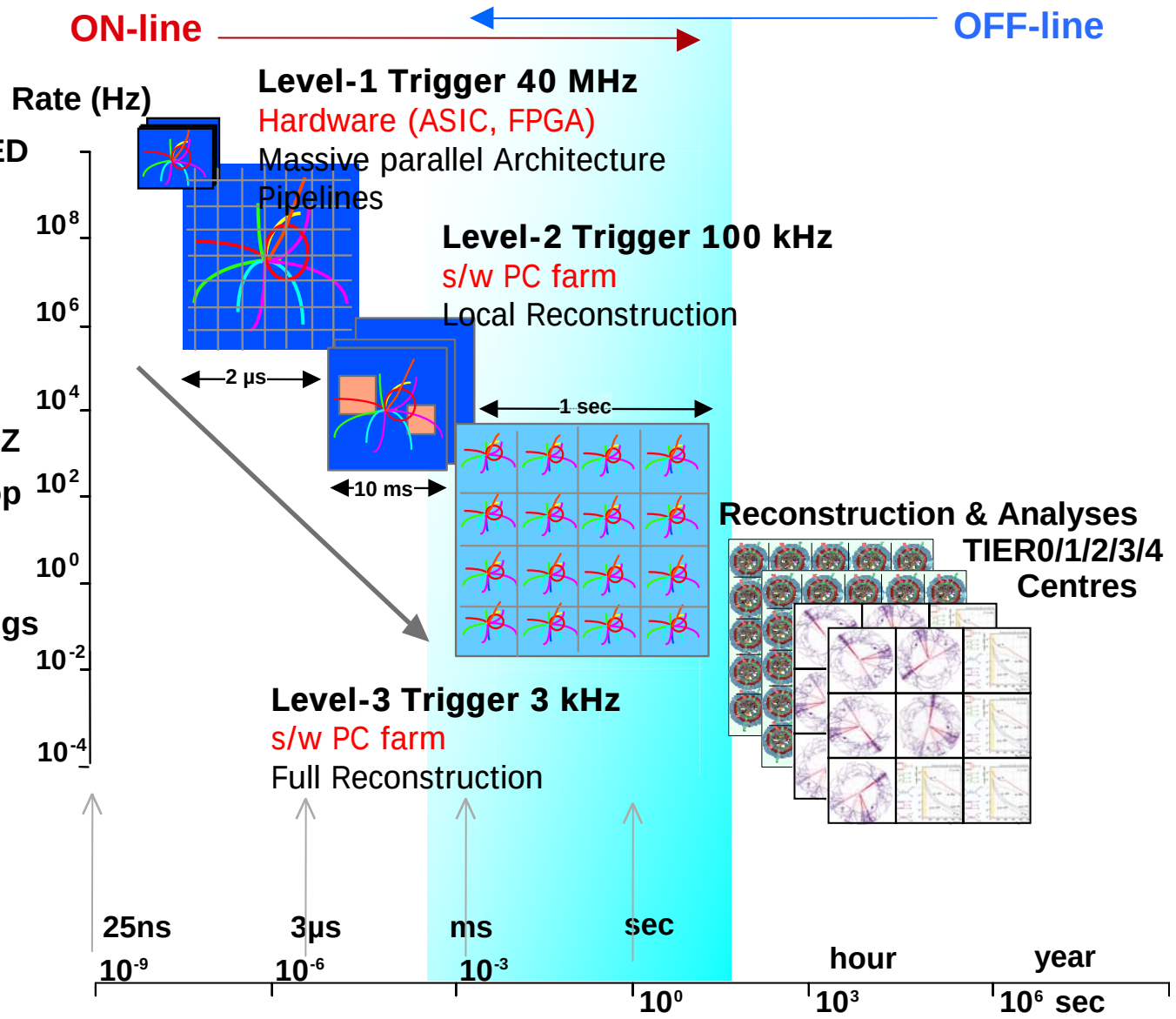
Conclusions

- **ATLAS is well on track**
- **First data can be expected by mid 2007**
  - **However**
    - Combined Testbeam data already now
    - Cosmics data starting up now (individual detector components)
    - Monte Carlo data existing now
  - **Can start implementing and applying policies alread now**
- **A flexible and robust TIER 2/3/4 system needed for individual analysis**
  - **otherwise no impact of Swiss physicists to LHC exploitation possible**
  - **usable resources (CPU, storage, bandwidth) for Swiss physicists concentrated in TIER2 (i.e. Manno) fits well with analysis procedure presented here**
  - **need for TIER3 (i.e. institutes cluster)**
    - for fast turn around in development and debug cycles
    - final analysis over skimmed data sets
- **Need to establish policies on how to share and use resources like the Manno cluster**
  - **this workshop**

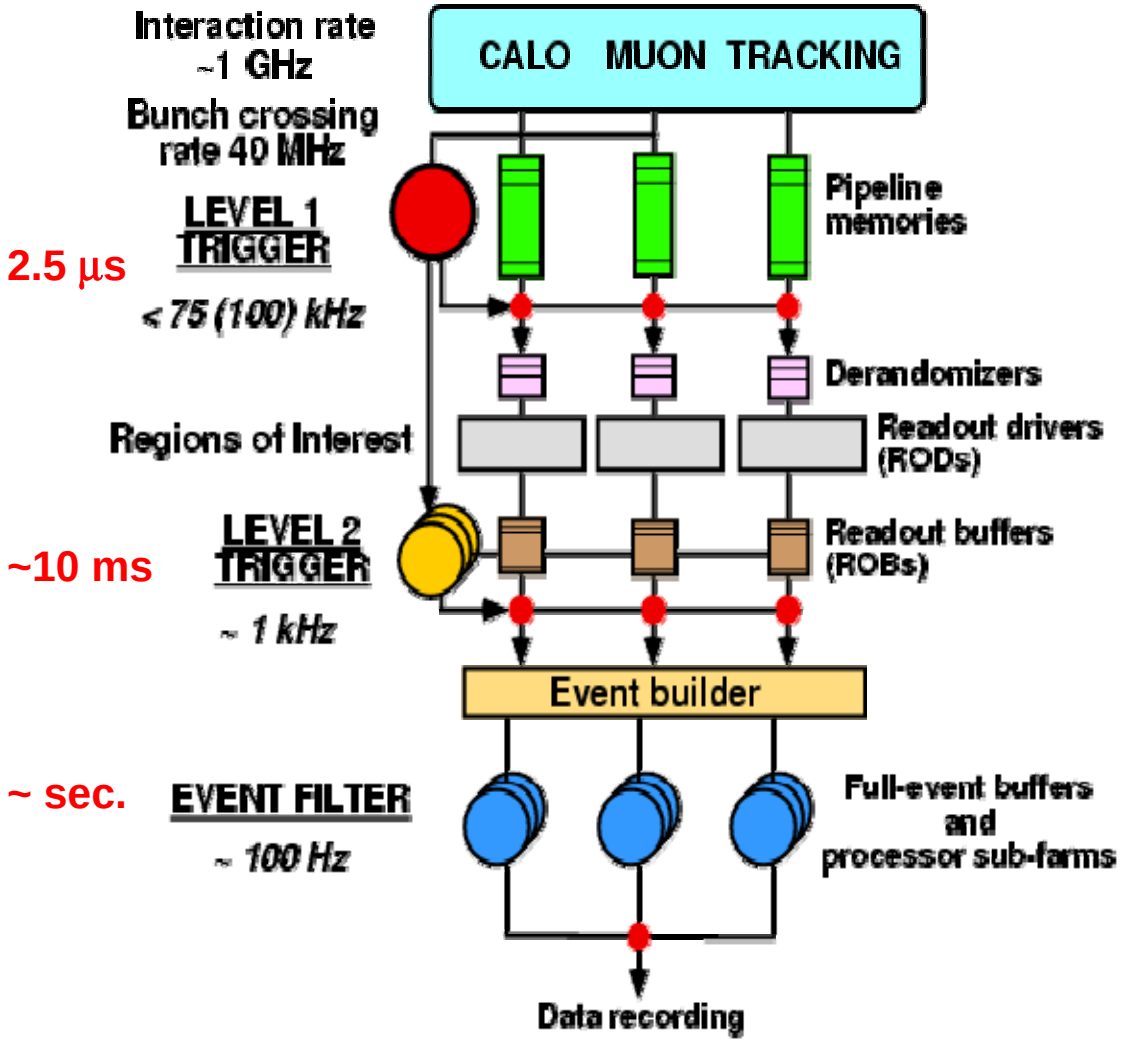
BackUp Slides

From Bunch Crossings to Physics Analyses



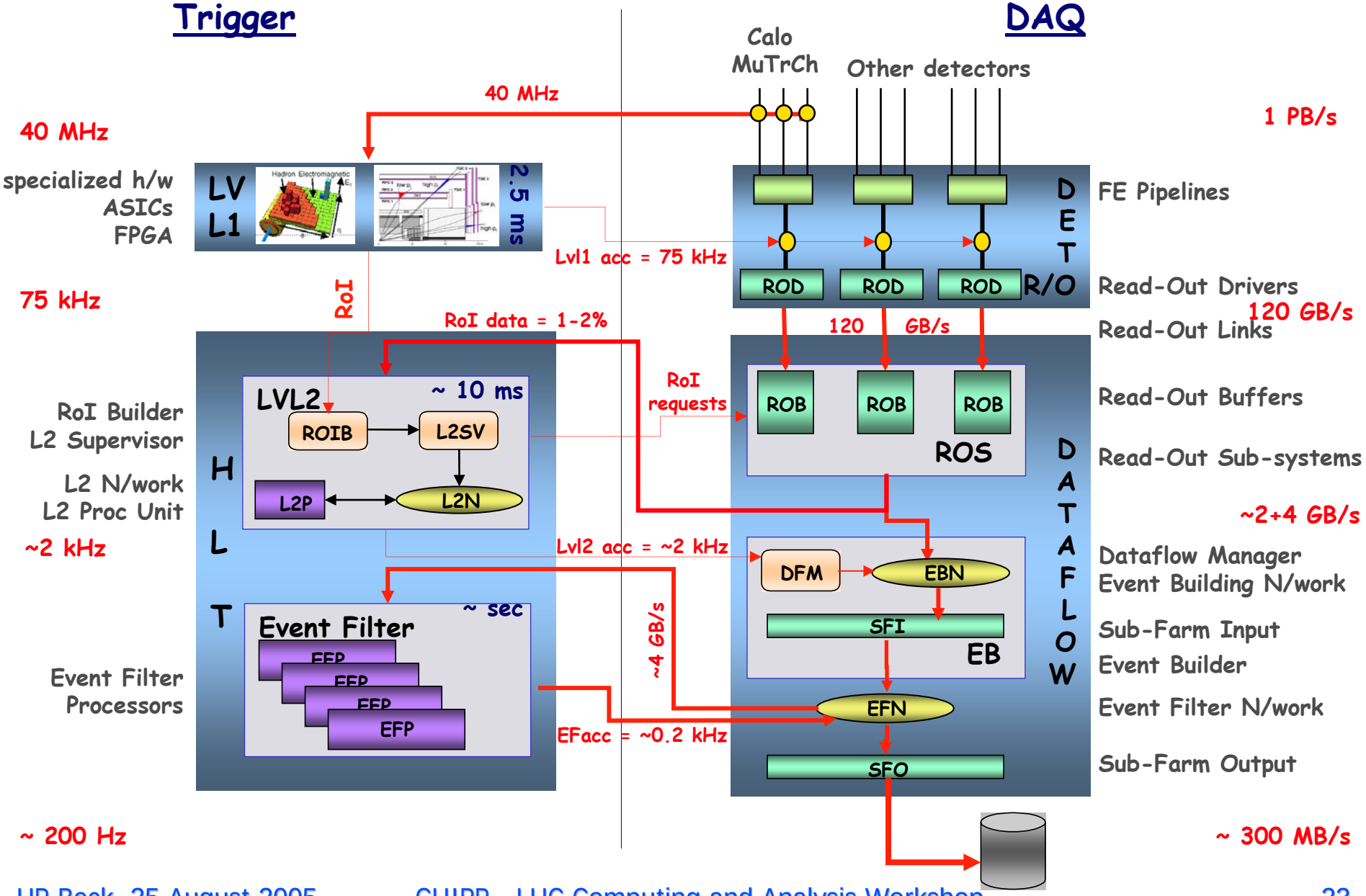


ATLAS Three Level Trigger Architecture



- LVL1 decision made with calorimeter data with coarse granularity and muon trigger chambers data.
  - Buffering on detector
- LVL2 uses Region of Interest data (ca. 2%) with full granularity and combines information from all detectors; performs fast rejection.
  - Buffering in ROBs
- EventFilter refines the selection, can perform event reconstruction at full granularity using latest alignment and calibration data.
  - Buffering in EB & EF

ATLAS TDAQ Architecture



Resources needed for Simulating 100k SUSY events

- **Generation**
  - **100k events**: 20 files of 5k evgen events
  - SIM+DIGI+REC: File size: 50 events/files (**limited by memory leak**; i.e. s/w bugs)
    20 x 100 jobs, about 10h each

- **CPU**
  - Time to process 1 event: **650 sec on a 2.8 GHz processor**
    - GEN 0.5 kSI2-sec
    - SIM+DIGI 100 kSI2-sec
    - REC 15 kSI2-sec
    - Analyze 0.5 kSI2-sec
  - 100k hours GHz –or- 4.5k day GHz – or- **15 days on a 100x3GHz cluster**

- **DISK SPACE**
  - Event size
  - Generated : **0.07 MB/evt**
  - Simulated and digitalized **(raw data): 2.1 MB/evt**
  - Reconstructed **(ESD) 0.9MB/evt AOD's 0.025 MB/evt**
  - **Total 3MB/event**

  - **Need about 300GB to store 100k events**