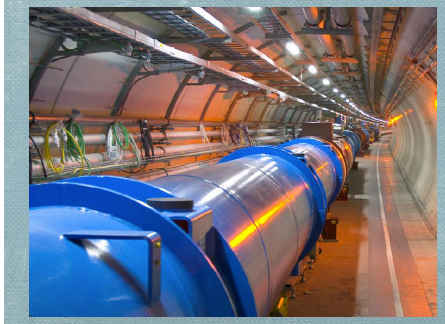
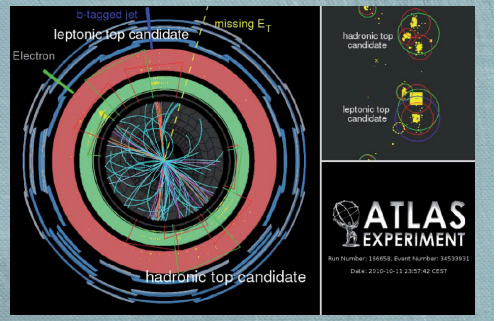




FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR Portugal

UNIÃO EUROPEIA
Fundo Social Europeu

POPH

Optimization Processes in High Energy Physics:

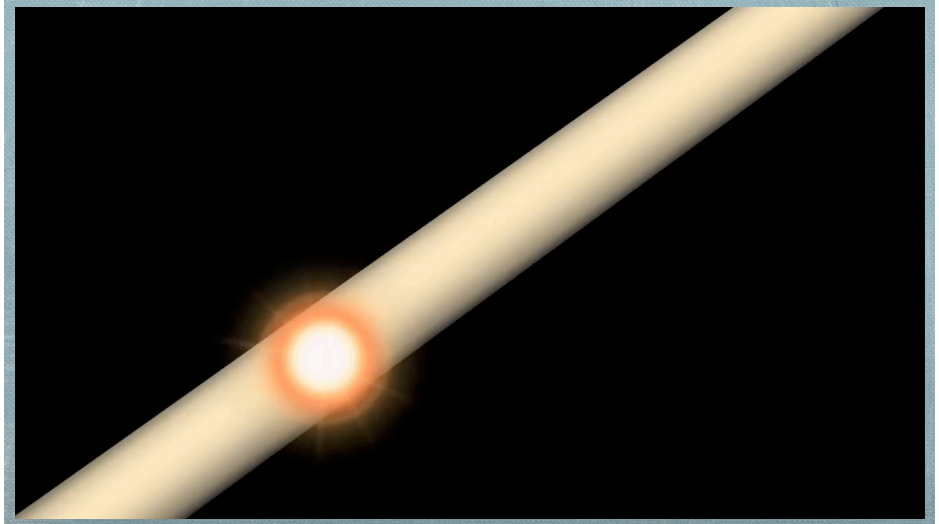
Challenges at the LHC

(A.Onofre, N.Castro, M.Oliveira,H.Carvalho,
E.Gouveia,R.Alves, J.Espinosa,P.Chennavajhula)
antonio.onofre@cern.ch

LIP

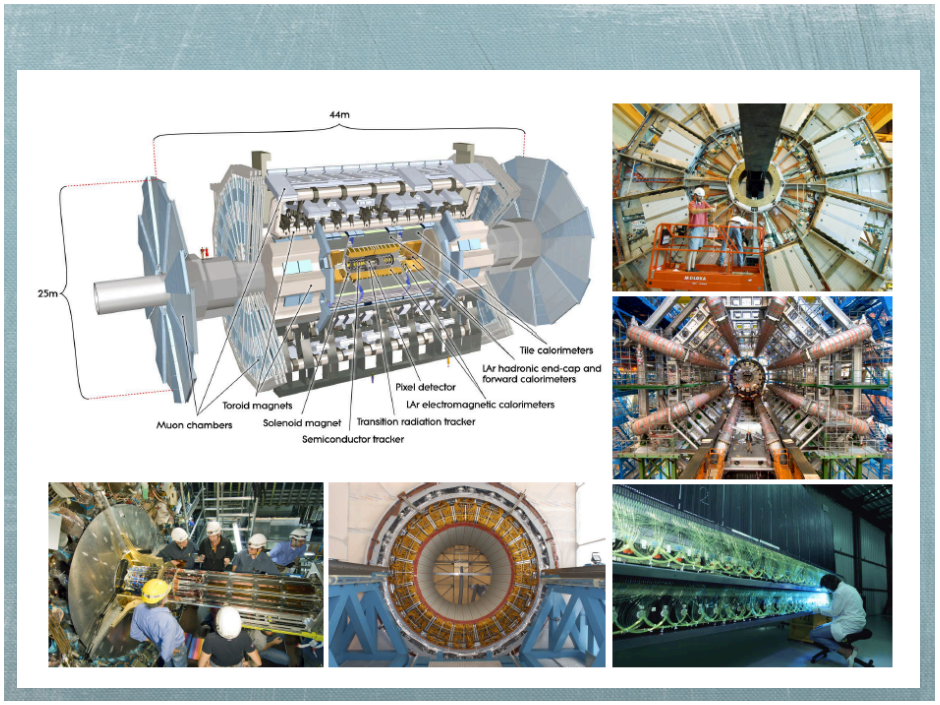
The LHC

The LHC

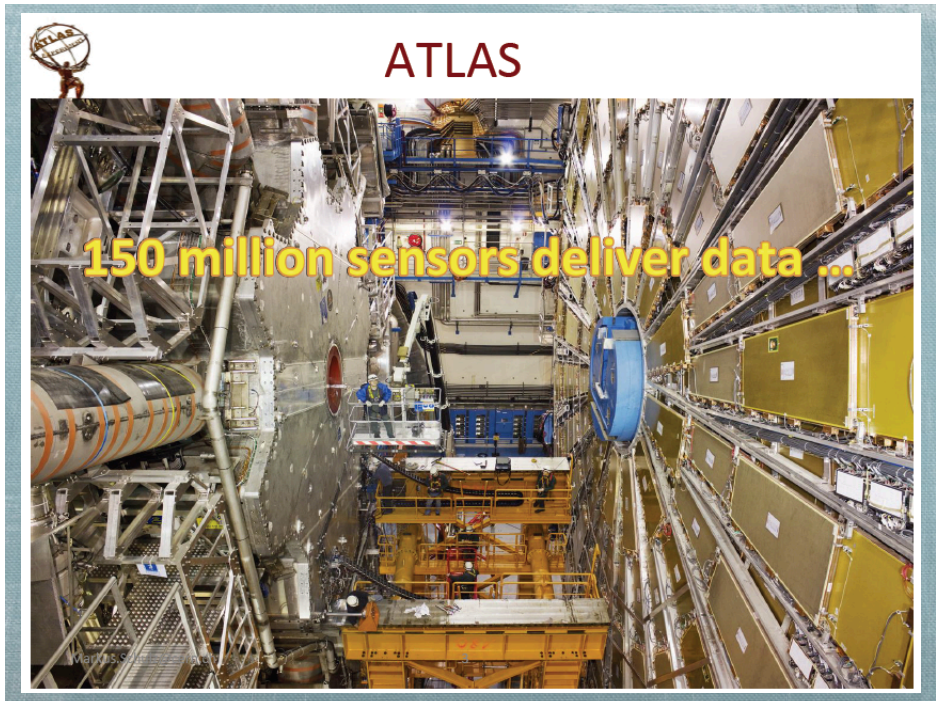


A Thoroidal LHC ApparatuS

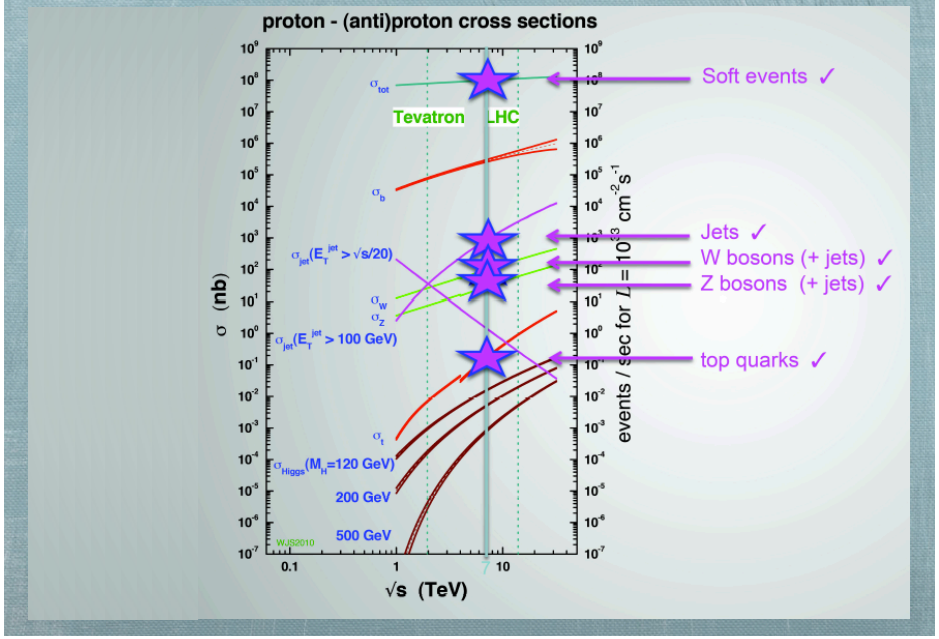
- ATLAS is one of the six particle detectors experiments at Large Hadron Collider (LHC) at CERN
- The project involves more than 3000 scientists and engineers in ~40 countries
- ATLAS has 44 meters long and 25 meters in diameter, weighs about 7,000 tons. It is about half as big as the Notre Dame Cathedral in Paris and weighs the same as the Eiffel Tower or a hundred 747 jets



ATLAS

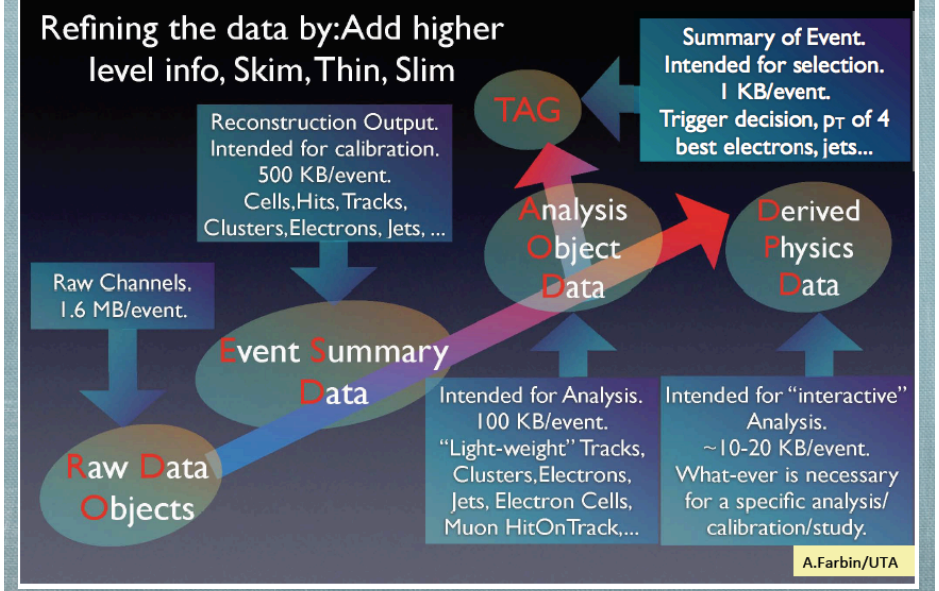


Summary of *pp* collisions results



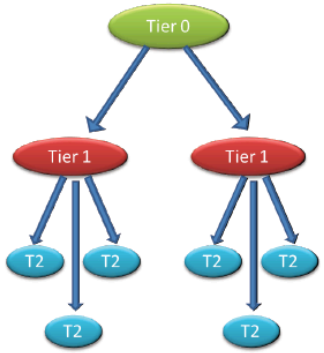
The Event Data Model

Refining the data by:Add higher level info, Skim, Thin, Slim



Grid Data Processing

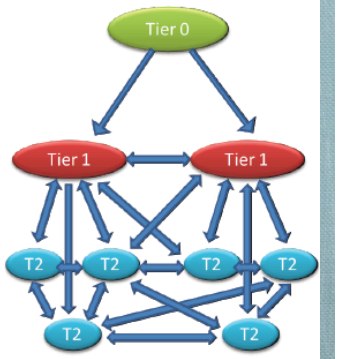
- ATLAS Grid Data Processing (GDP) uses Grids with three different interfaces split in ten "clouds" organized as large computing centres with tape data storage (Tier-1 sites) each associated with 5-6 other computing centres (Tier-2 sites)
  - ATLAS clouds evolve from the hierarchy (left) to the mesh (right)



15%: Tier-0 center at CERN
- Data recording
- First-pass data reconstruction
- Data distribution

40%: Tier-1 centers
- Permanent storage
- Reprocessing
- Group data processing

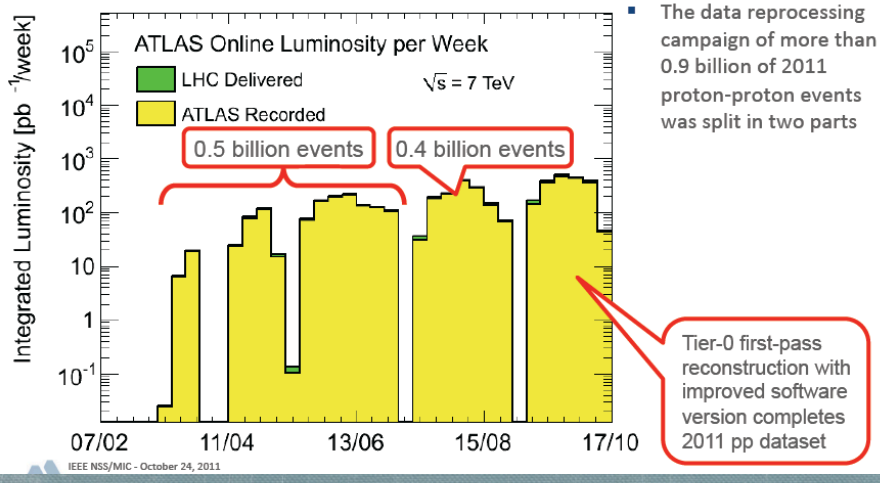
45%: Tier-2 centers
- Simulation
- End-user analysis



- GDP system empowered further data processing steps performed by dozens of ATLAS physics groups with coordinated access to computing resources worldwide

Coherent 2011 Dataset

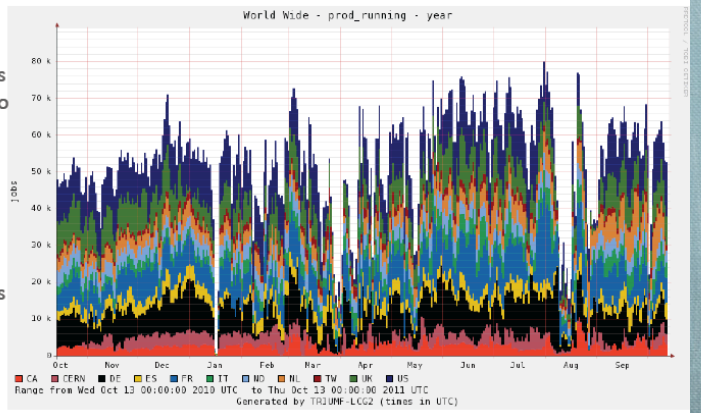
- Preparations for 2011 data reprocessing campaign were going on since January
  - Achieving numerous improvements in physics performance and data quality
    - Such as providing 8% more luminosity-weighted data to be used in physics analysis



- The data reprocessing campaign of more than 0.9 billion of 2011 proton-proton events was split in two parts

From Reprocessed Data to Physics Analysis

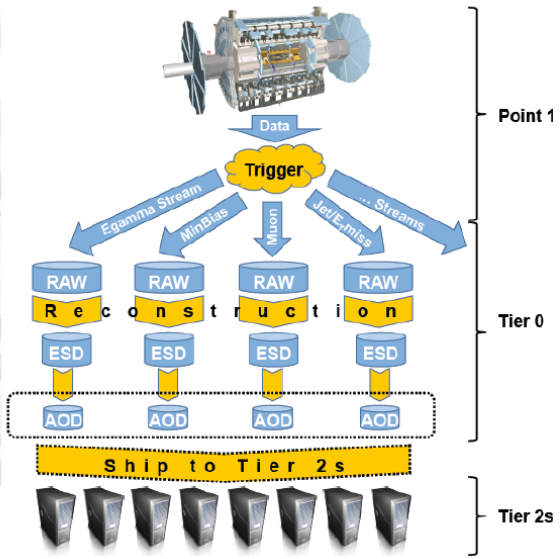
- In contrast with the major reprocessing campaign that are conducted only few times per year, the centrally managed production for physics groups process full available dataset once every few weeks, providing further improvements in the data used for ATLAS physics analysis shortly after the reprocessing or data taking
- Physics analysis of the detector data relies on the results of the "Monte Carlo experiments" produced by the CPU-intensive detector response simulations tasks mostly running on the Grid Tier-2 sites



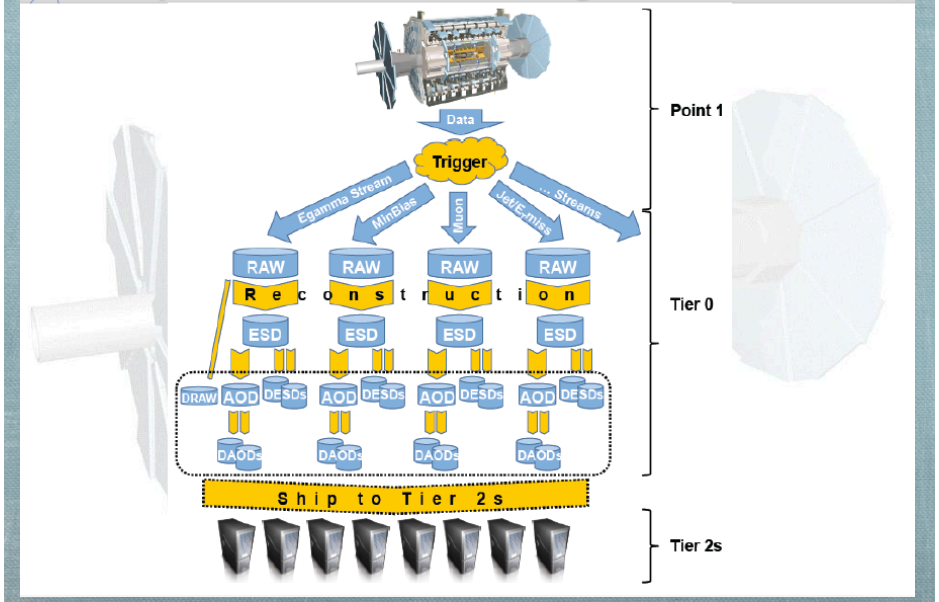
Data placement – original

Available to all users:

- Only data that are on disks at the tier-2s is readily available for user analysis.
- Only the AOD (and the TAG) is shipped to the tier-2s.
- Very detailed information that resides only in the ESD is NOT accessible to users.
- The AOD is still very large for frequent analyzes by users.

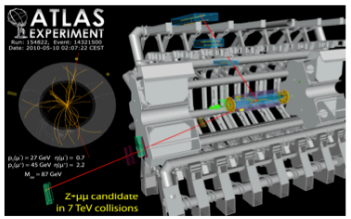


Data production chain – adding DAODs

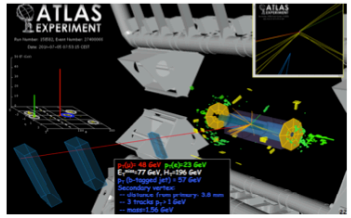


The first collisions at $\sqrt{s} = 7$ TeV

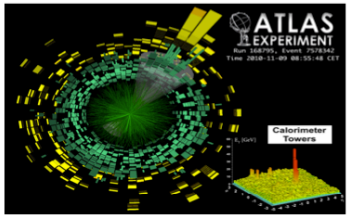
Marsch 2010: ATLAS Control Room



Z Candidate (10 May 2010)



Top Candidate (5 July 2010)

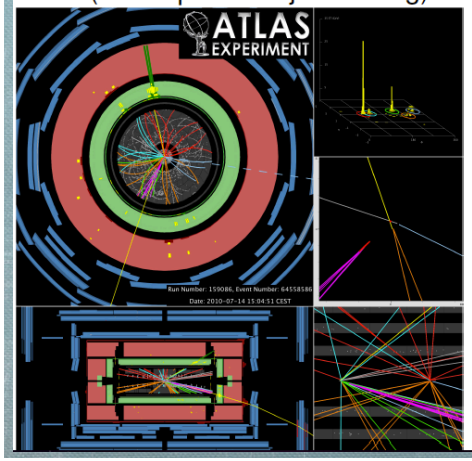


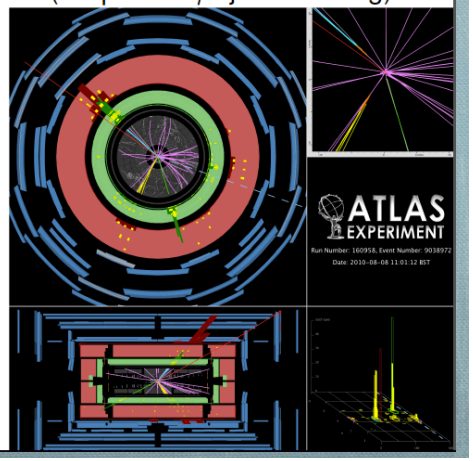
Pb-Pb collision (9 November 2010)

$t\bar{t}$ Semi and Dileptonic Events

☞ top candidates seen at ATLAS

(semileptonic e+jets w/btag) (dileptonic $e\mu$+jets w/2 btag)

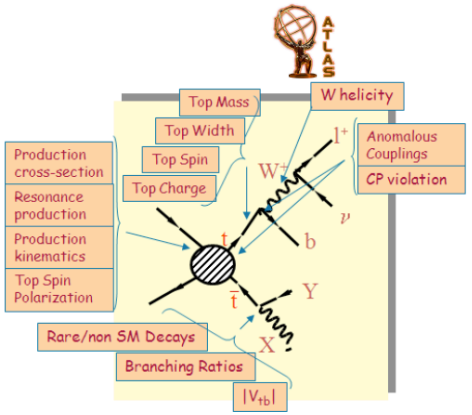




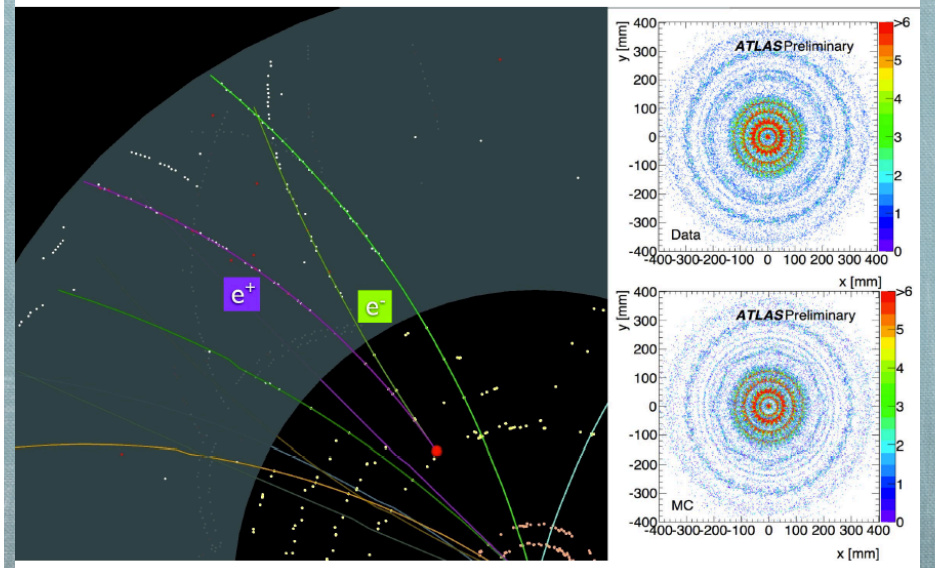
Outline

Top quark @ LHC

- $t\bar{t}$ production
  - $\sigma_{t\bar{t}}$
  - Mass
  - W polarization and the $t \to bW$ decay
  - Anomalous couplings
  - FCNC
  - Charge Asymmetry
- Single top production
  - cross section



Detector mapping with $\gamma \rightarrow e^+e^-$ conversions



Outstanding Challenges

- **Analysis Code Optimization**
  - Use current analysis code (LipCbrAnalysis), identify possible inefficiencies (way data is read, identify redundant operations, identify obsolete code, etc.) and correct them, without major changes in the underlying framework
  - We need to gain a significant factor in speed in order to remain competitive with increasing flow of data

Outstanding Challenges

- **Viability Studies: run analysis code using GPU's ?**
  - Example of previous tests in HEP (NA62 Experiment@CERN): "The usage of graphical card processing in the construction of a trigger system, both in hardware and software, allowed to make decisions in real-time and increased the computing capability in the trigger."
  - Tests on Matrix Multiplication were performed on Nvidia GeForce 8400 GS with a host processor Intel Core 2 Duo CPU E8400 (3.0 GHz) with 2GB DDR2 SDRAM. Results of tests:

Matrix Dimension	Time Taken in CPU (in seconds)	Time taken in GPU (in seconds)
1024X1024	4.0	0.52
2048X2048	130.62	1.51
3072X3072	490.98	5.25
4096X4096	1270.22	9.94

Going Even Further:

- **I/O Optimization**
  - Is it possible to have a more efficient I/O handling within our framework (we read on average 5TB of data/per analysis)?
  - We crucially depend on ROOT package (not trivial to optimize!)
- **Local Resources Optimization:**
  - test the use of Paralel ROOT Facility (PROOF)
- **GRID Resources Optimization:**
  - GPU's ???