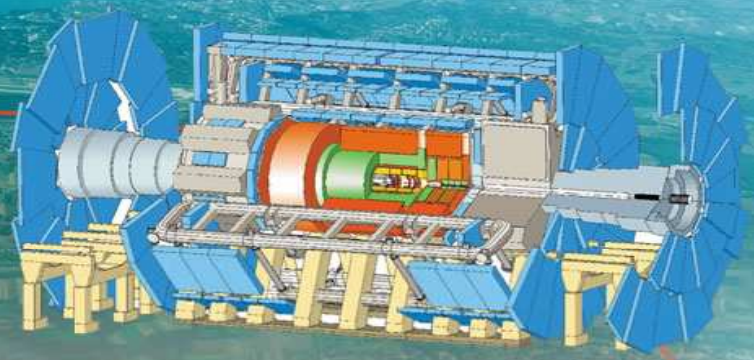
Distributed analysis for the ATLAS Experiment in the S.Co.P.E Project

Gianpaolo Carlino
INFN Napoli

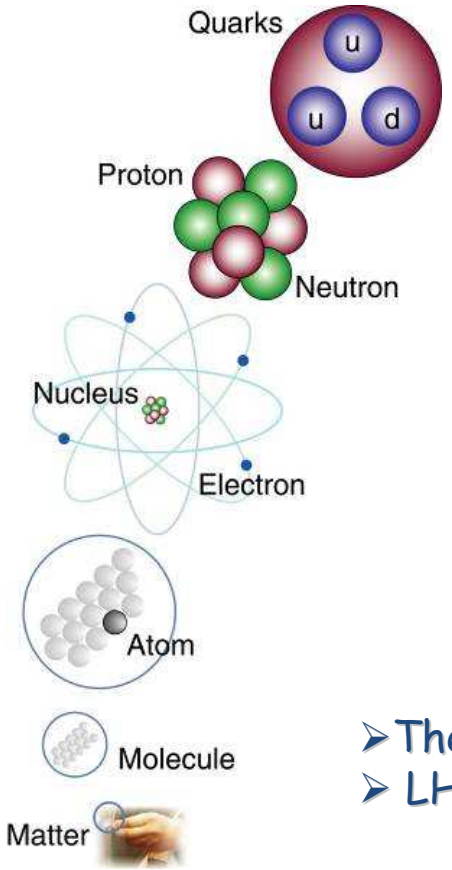
Workshop finale dei Progetti Grid
del PON Ricerca 2000-2006 - Avviso 1575
Catania, 10-12 February 2009



- The LHC Physics Motivations and the ATLAS Experiment
- The ATLAS Computing Model
- The Napoli ATLAS Tier2 & SCoPE

The LHC Physics Motivations

The study of elementary particles and fields and their interactions



Elements of the Standard Model

matter particles | gauge particles

	1st gen.	2nd gen.	3rd gen.
QUARK	u up	c charm	t top
	d down	s strange	b bottom
LEPTON	ν_e e neutrino	ν_μ μ neutrino	ν_τ τ neutrino
	e electron	μ muon	τ tau

Strong Force: g Gluon x8

Electro-Magnetic Force: γ photon

Weak Force: W^+ W^- W bosons; Z Z boson

scalar particle(s): H Higgs ? ? • • •

Main goal of LHC is the Higgs discovery

Direct search at LEP: $M_H > 114,4$ GeV

- The Standard Model is incomplete
- LHC addresses several leading questions in particle physics
  - ✓ Source of Mass ($M_\gamma = 0$ vs. $M_W = 89.4$ GeV)
    - Higgs particle?
    - Supersimmetry?
  - ✓ Are quarks fundamental? Are there more quarks?
  - ✓ Source of Dark Matter

The LHC Physics Motivations

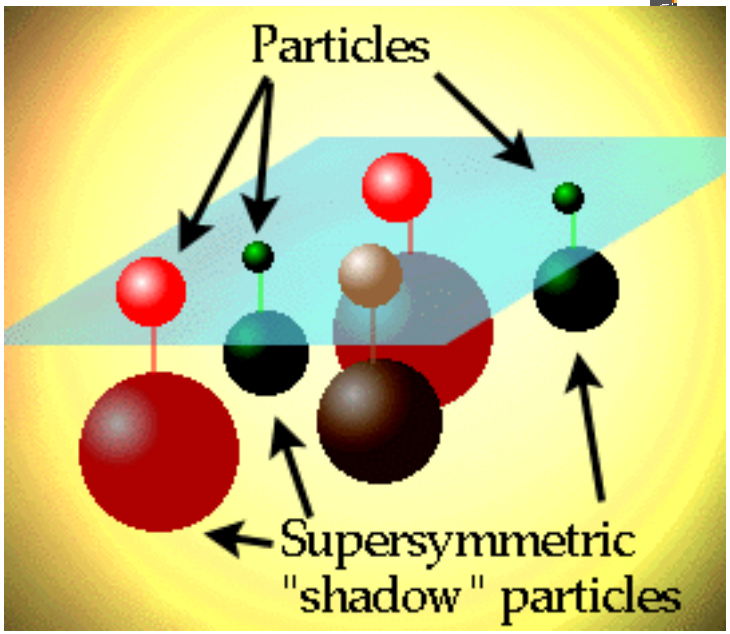
Supersymmetry (SUSY)

Establishes a symmetry between fermions (matter) and bosons (forces):

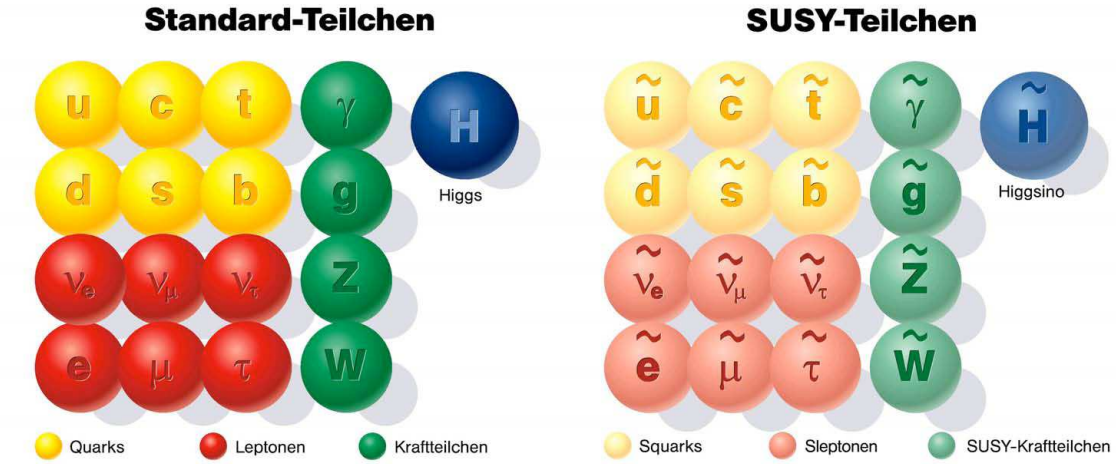
- Each particle p with spin s has a SUSY partner p with spin s -1/2

- Examples q (s=1/2) → $\tilde{q}$ (s=0) squark

  g (s=1) → $\tilde{g}$ (s=1/2) gluino



Our known world Maybe a new world?



Why SUSY is so interesting?

- Unification
- Solves some deep problems of the Standard Model
- If the Lightest SUSY particle is stable we have a "natural" explanation of *Dark Matter*

Physics at LHC: the challenge

Small x-sections

need highest luminosity

$L = 10^{34}$ cm^{-2}s^{-1}

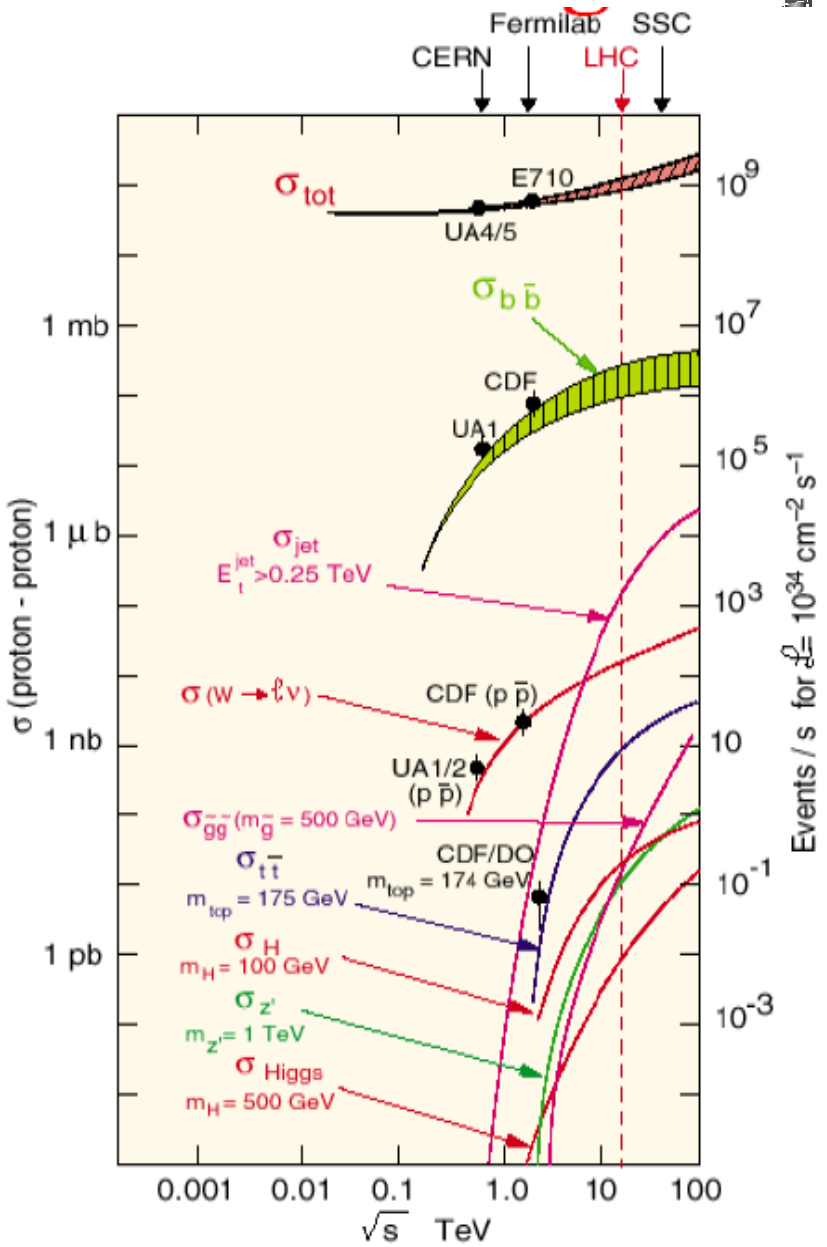
Orders of magnitude of event rates for various physics channels:

- **Inelastic :** 10^{10} Hz
- **W -> lν :** 10^{3} Hz
- **tt production :** 10^{2} Hz
- **Higgs (m=100 GeV) :** 1 Hz
- **Higgs (m=600 GeV) :** 10^{-1} Hz

(and include branching ratios: ~ 10^{-2})

Selection power for Higgs discovery ≈ 10^{14}

Rate (events/sec) = σ x L



The LHC machine

Lake of Geneva

CMS

The Large Hadron Collider is a 27 km long collider ring housed in a tunnel about 100 m underground near Geneva.

LHCb

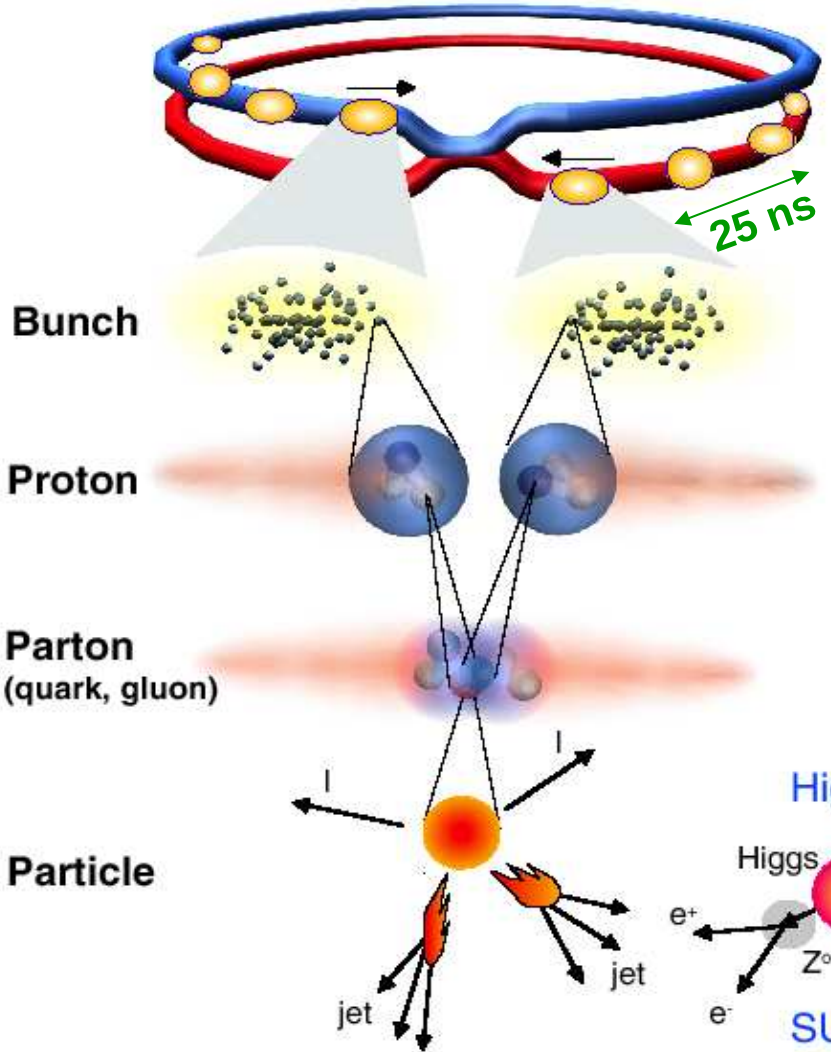
Airport

ALICE

ATLAS

The LHC machine is fully installed and was ready to start operation with single beams on 10th September 2008, but it is now delayed for several months until next spring after an incident that happened on 19th September

Collisions at LHC



Proton-Proton	2835 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	10^{34} cm^{-2} s^{-1}
Crossing rate	40 MHz

Selection of 1 in
10,000,000,000,000

Event rate:

N = L x σ (pp) ≈ 10^9 interactions/s

→ very powerful detectors needed

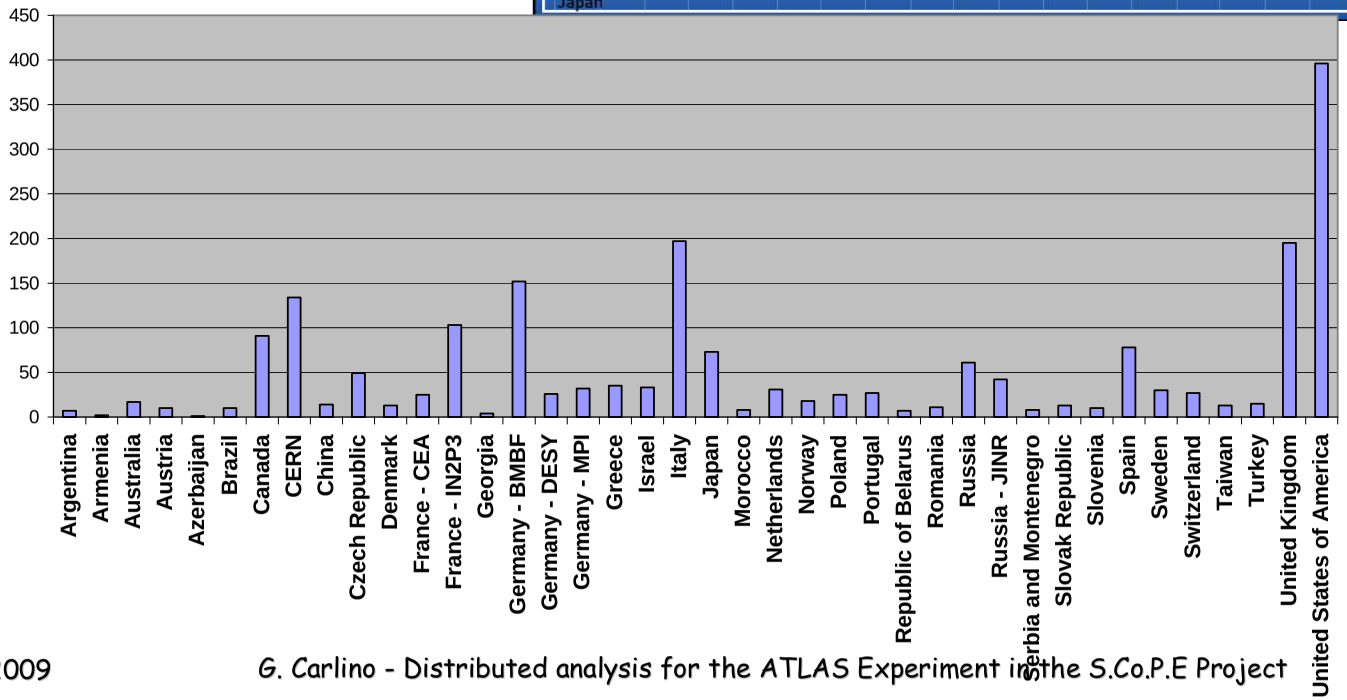
ATLAS Collaboration

A Toroidal LHC ApparatuS

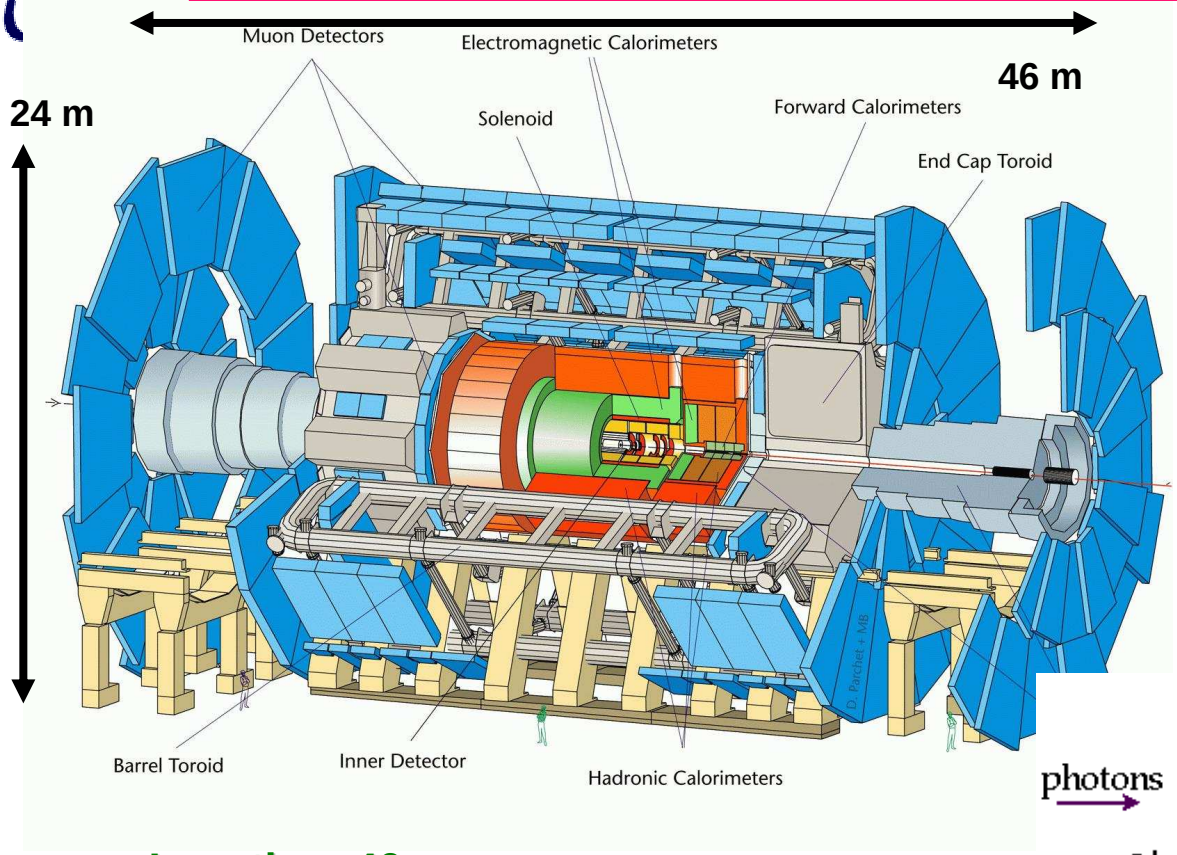
37 Countries
167 Institutions
2000 Scientific Authors total

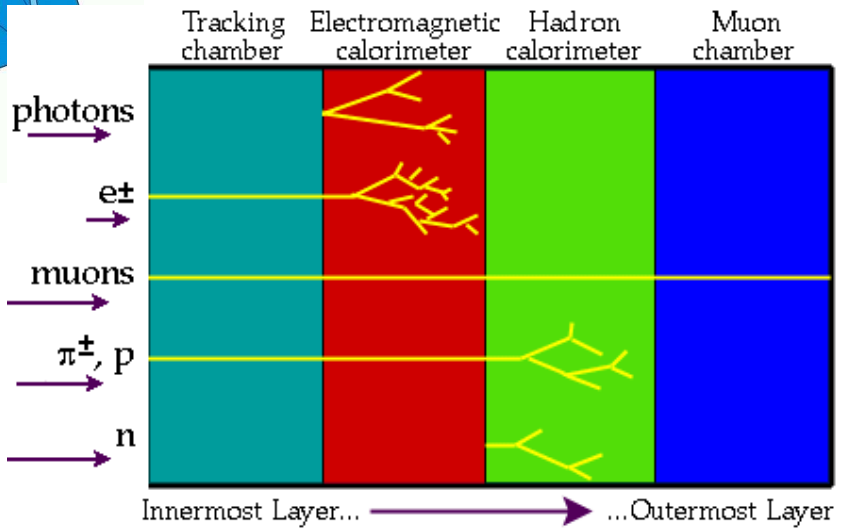
ITALY

13 Institutions
~ 200 Members



The ATLAS Experiment





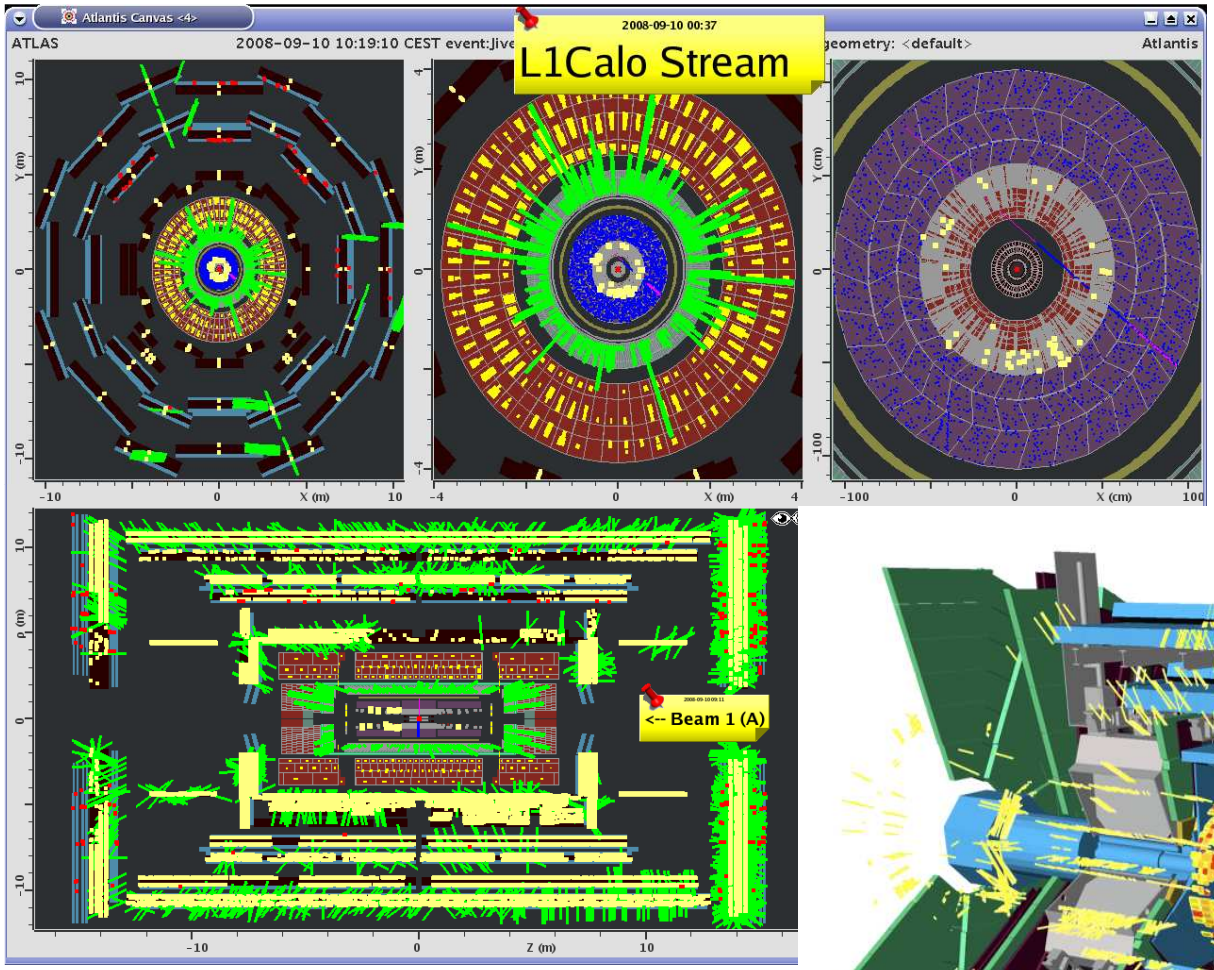
Length: ~ 46 m

Radius: ~ 24 m

Weight: 7000 Tons

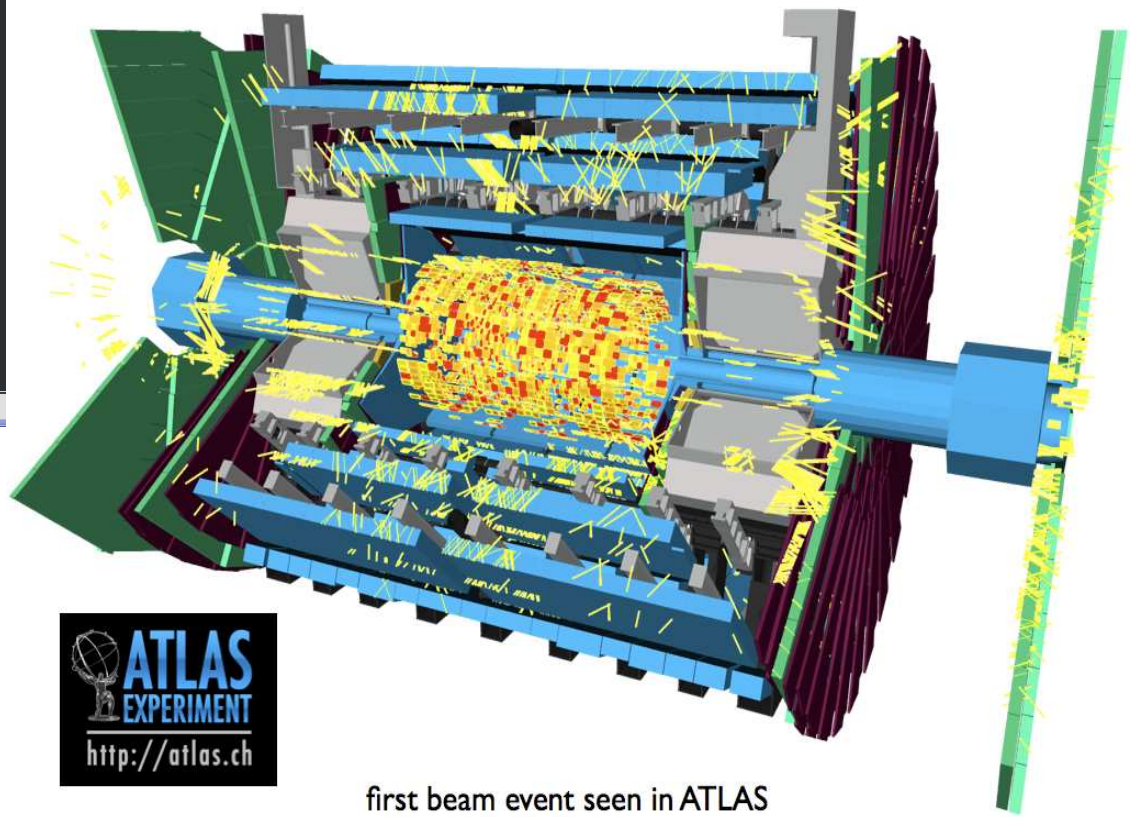
~ 10^8 electronic channels

~ 3000 km of cables



First LHC Events

The very first beam-splash event from the LHC in ATLAS on 10:19, 10th September 2008

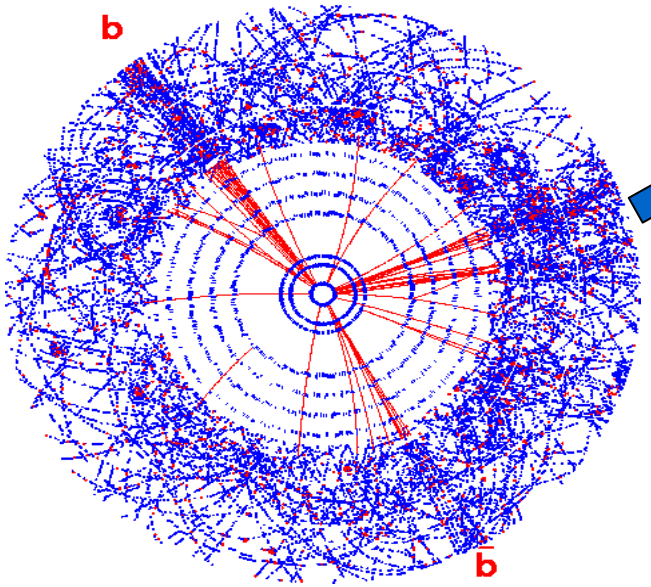


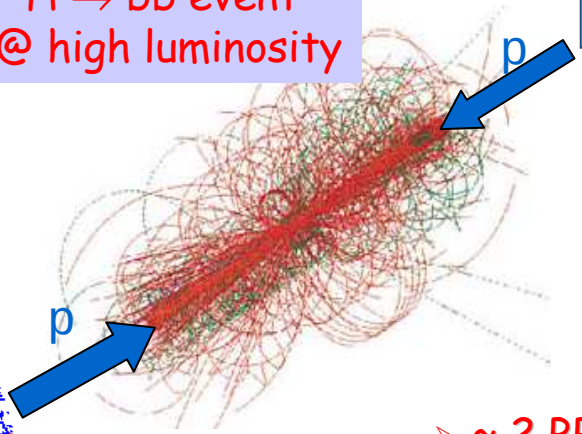
first beam event seen in ATLAS

The ATLAS events

H → bb event
@ high luminosity

… are heavy …





… and in large quantity

Event rate = 1 GHz

Trigger rate = 200 Hz

Event data flow from on-line to off-line = 320 MB/s

⇓

- ~ 2 PB/year of RAW data
- Processing, Reprocessing and user analysis need $\mathcal{O}(10^5)$ CPUs full time and produce $\mathcal{O}$(PB/year) of derived data

High granularity detectors needed (10^8 electronic channels)

⇓

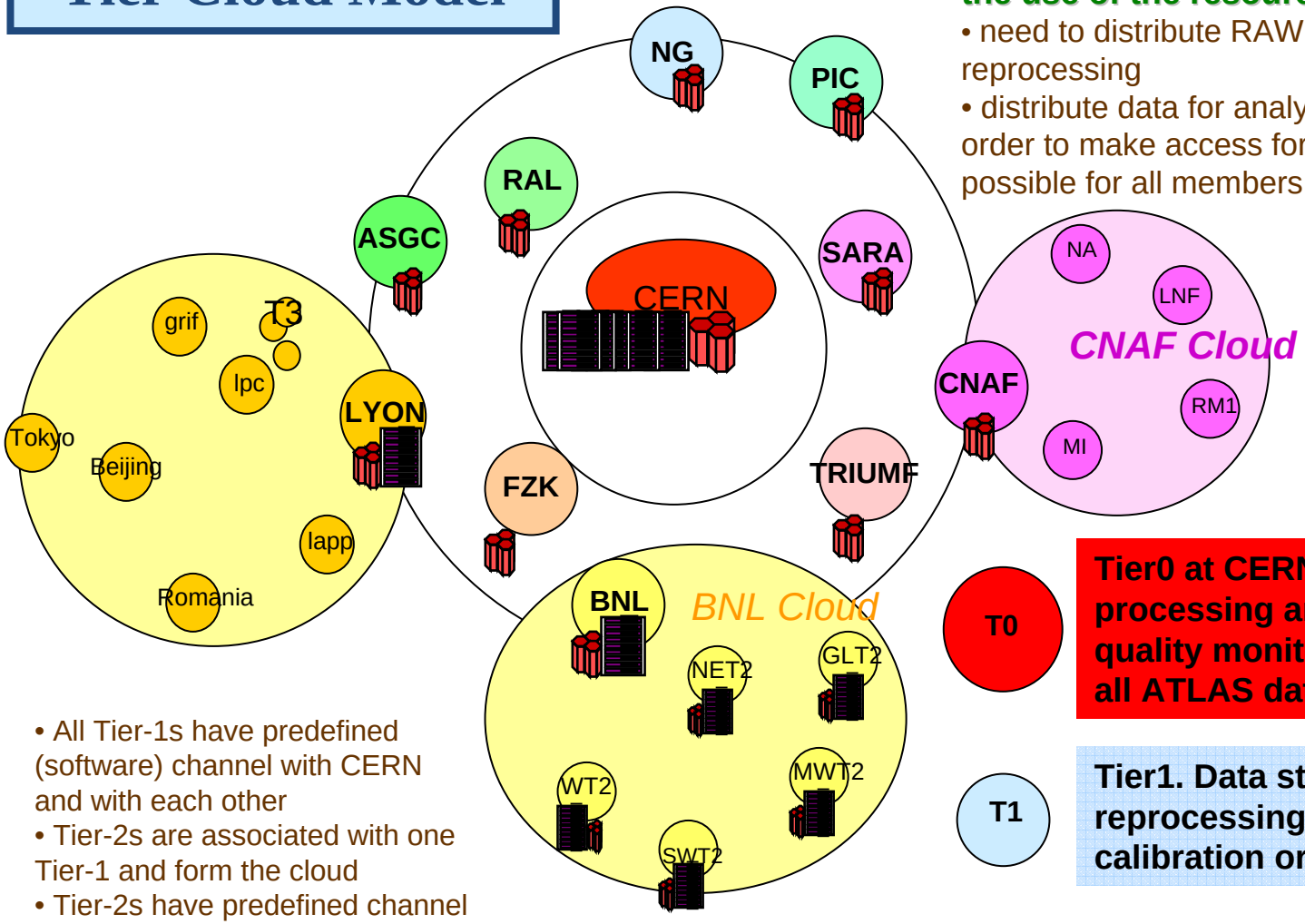
Very large event size = 1.6 MB

There is no way to concentrate all needed computing power and storage capacity at CERN

Necessity of Distributed Computing and GRID. The LHC Computing Grid (LCG) provides the world-wide infrastructure used by ATLAS to distribute, process and analyse the data.

The ATLAS Computing Model

"Tier Cloud Model"



Hierarchical Computing Model optimising the use of the resources

- need to distribute RAW data for storage and reprocessing
- distribute data for analysis in various formats in order to make access for analysis as easy as possible for all members of the collaboration

- All Tier-1s have predefined (software) channel with CERN and with each other
- Tier-2s are associated with one Tier-1 and form the cloud
- Tier-2s have predefined channel with the parent Tier-1 only.

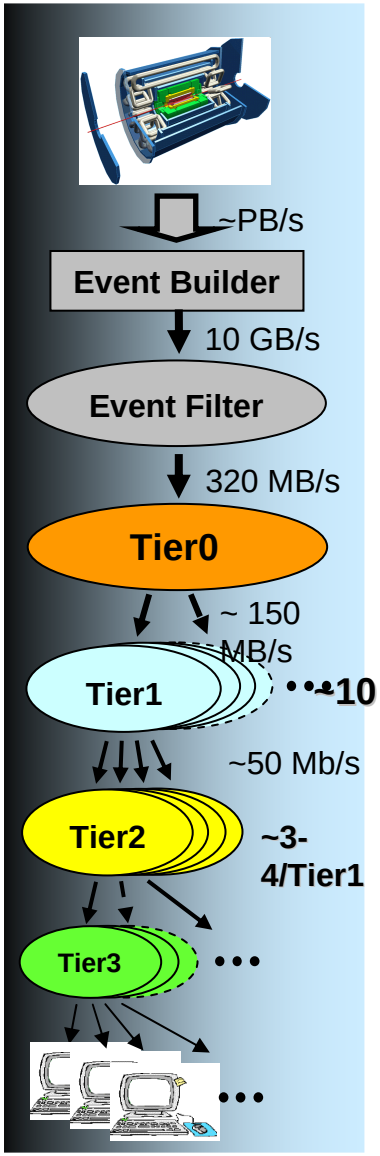
Tier0 at CERN. Immediate data processing and detector data quality monitoring. Stores on tape all ATLAS data

Tier1. Data storage and reprocessing of data with better calibration or alignment constants

Tier-2. Simulation and user Analysis

The ATLAS Computing Model

The ATLAS Computing Model



Three Grid middleware infrastructures are used by the ATLAS distributed computing project:

- EGEE (Enabling Grids for E-sciencE) in most of Europe and the rest of the world
- NorduGrid in the Scandinavian countries
- OSG (Open Science Grid) in the USA

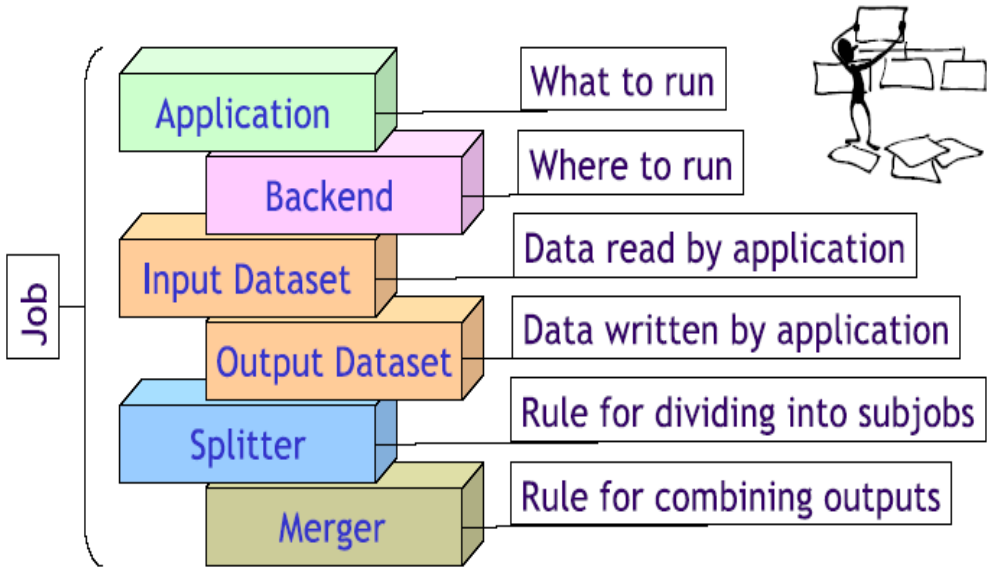
The Atlas Grid tools interface to all middleware types and provide uniform access to the Grid environment

- The VOMS database contains the computing information and the privileges of all ATLAS members; it is used to allow ATLAS jobs to run on ATLAS resources and store their output files on ATLAS disks
- the DDM (Distributed Data Management) system catalogues all ATLAS data and manages the data transfers
- The ProdSys/Panda production system schedules all organised data processing and simulation activities
- The Ganga and Pathena interfaces allow the analysis job submission: jobs go the site(s) holding input data and output data can be stored locally or sent back to the submitting site.

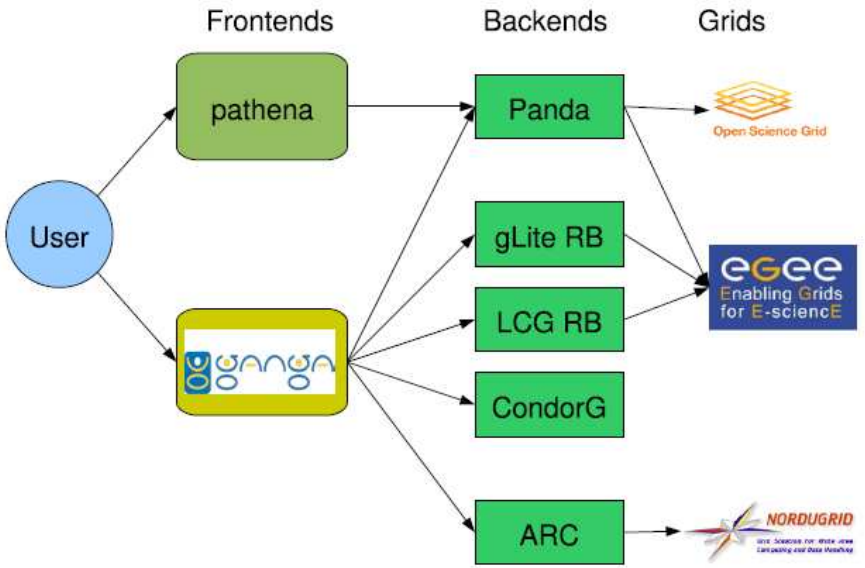
The Distributed Analysis

GANGA: a single tool to submit/manage/monitor the user jobs on GRID

Functional Blocks: a job is configured by a set of functional blocks, each one with a well defined purpose



Grid and local execution allowed. Grid backends include all the three middleware flavours



- Interface via a command-line or a GUI
- Users don't need to know where data are located and where jobs run.

Napoli ATLAS Tier2

Napoli is one of the four ATLAS Italian Tier2s with Milano, Roma1 and Frascati (sj).
Provides pledges resources to the whole ATLAS collaboration for Monte Carlo simulation and analysis and in particular to italian users.

- Computing Resources:
  - 50 WN, 260 cores, 580 kSI2k
- Storage Resources:
  - SAN architecture
  - 260 TB
  - 12 disk servers
- Local Services:
  - CE, SE, DPM, HLR, Monitoring

Two-site structure:
- INFN site in the Physics Dept.
- SCoPE "Federico II" site

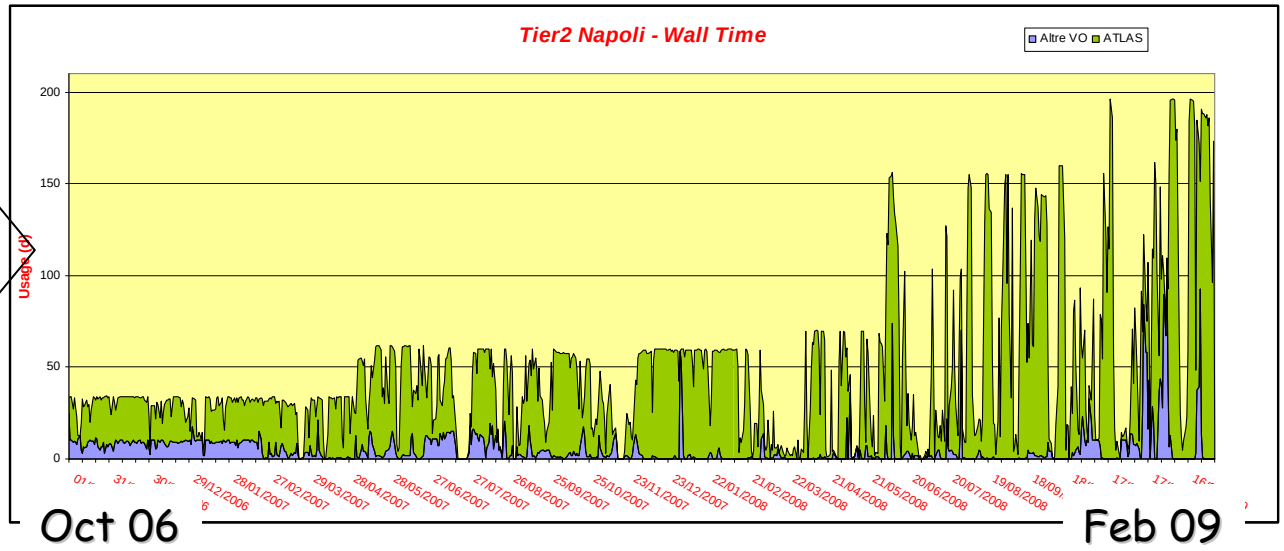
linked by 10 Gbps fibre connection

INFN site
- 3 ATLAS racks
- 1 SCoPE prototype rack

Napoli ATLAS Tier2

The Napoli ATLAS farm started its operation as official Tier2 in the first half of 2006. Since then we have been running continuously.

Wall time usage since Oct 06. It's evident the increasing in the available computing resources



Main activities in the Tier2 by the Napoli and other ATLAS groups:

- Muon Trigger (LVL1 and EF)
- Muon Reconstruction
- RPC calibration and analysis
- Supersimmetry Searches

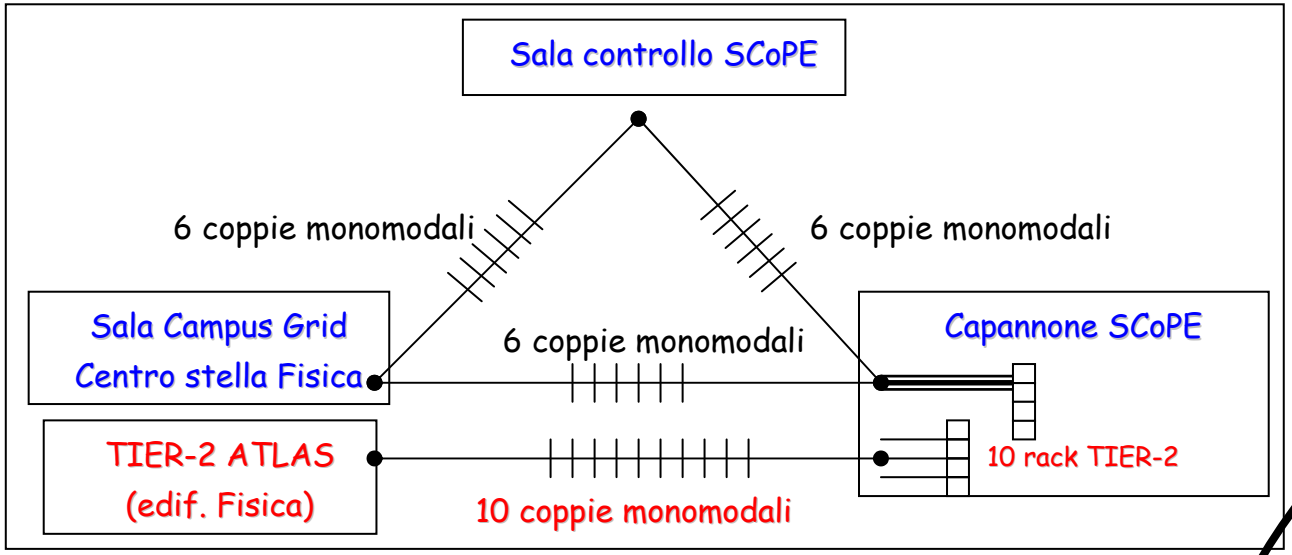
Waiting for data to get physics results

In the meantime we participated to every kind of test organised by ATLAS or WLCG:

- Functional Tests
- Throughput Tests
- Distributed Analysis Tests (Hammer Tests)
- CCRC08 (Combined Computing Readiness Challenge 2009) with all the LHC experiments

Napoli ATLAS Tier2

Network connection between SCoPE, INFN and POP GARR



10 fibres connecting at 10 Gbps the 10 ATLAS racks in the SCoPE site with the INFN site.

Napoli ATLAS Tier2

SCoPE site
- 10 racks for ATLAS

- the new ATLAS resources (WN and storage) have been installed in the SCoPE site
- the SCoPE resources will be used by ATLAS accordingly to the fair share policies among the supported VOs

So far
- the ATLAS job submission on the SCoPE prototype has been successfully tested
- Ganga configuration has been modified in order to use the SCoPE Grid services (CE and Resource Broker)
- The ATLAS central sw installation procedure includes the SCoPE CE.

Conclusions

- The LHC Physics Program has breakthrough discovery potential, will make an enormous number of precision measurements, and therefore is the leading component of the world High Energy Physics program from for the next ~20 years.
  - We are really excited with the physics starting soon! The new LHC schedule has been defined last week. We will be taking data starting from next autumn over the whole next year.
- The ATLAS computing model had been defined in order to exploit the advantages of the GRID and the large amount of resources spread over the world
- We have tested the functionality of the ATLAS sw on SCoPE and will be able to use, thanks to the interoperability, these and the other PON facilities