Particle Physics and the Grid

Randall Sobie
Institute of Particle Physics
University of Victoria

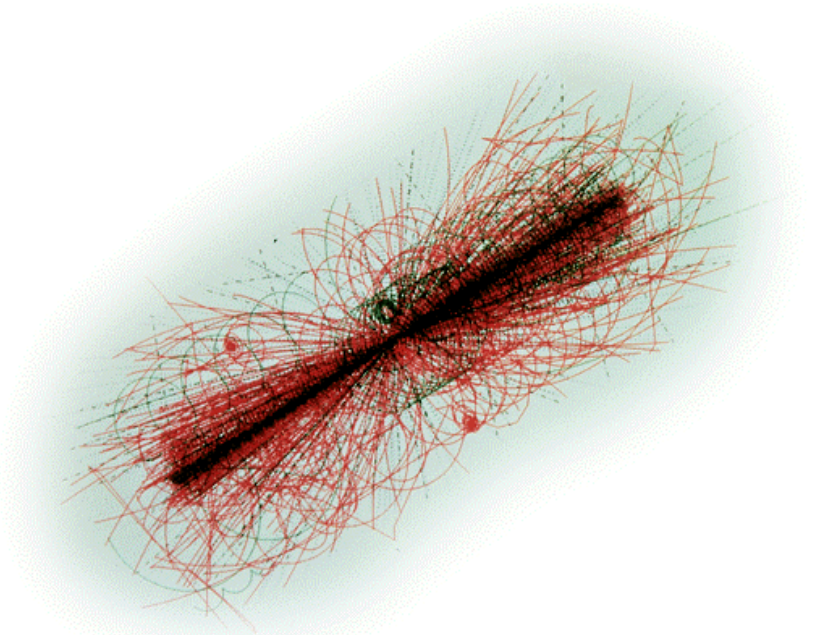
- Motivation
- Computing challenge
- LHC Grid
- Canadian requirements
- Local initiatives

Particle Physics

Goal is to understand the nature of matter and forces

We require the highest possible energy
- LHC will probe deep into matter (10^{-20} cm)
- glimpse at 1st picoseconds of the big bang

Requires large experimental facilities, dispersed around the world, and large international collaborations.



Proton-proton collision in ATLAS at LHC

Particle physics computing

Traditionally at the forefront of computing

- information technology (WWW)
- work together in an Open Software environment
- culture of remote coordination and collaboration

Current requirements (ATLAS at CERN)

- petabytes of distributed data
- tens of thousands of computing resources
- thousands of users

Grid seems to be the answer to our distributed computing requirements.

Particle Physics is an ideal field to experiment and test the Grid technologies.

Particle physics laboratories

Canadians are involved in experiments at:

SNO, TRIUMF in Canada

Fermilab, SLAC, Brookhaven in the US

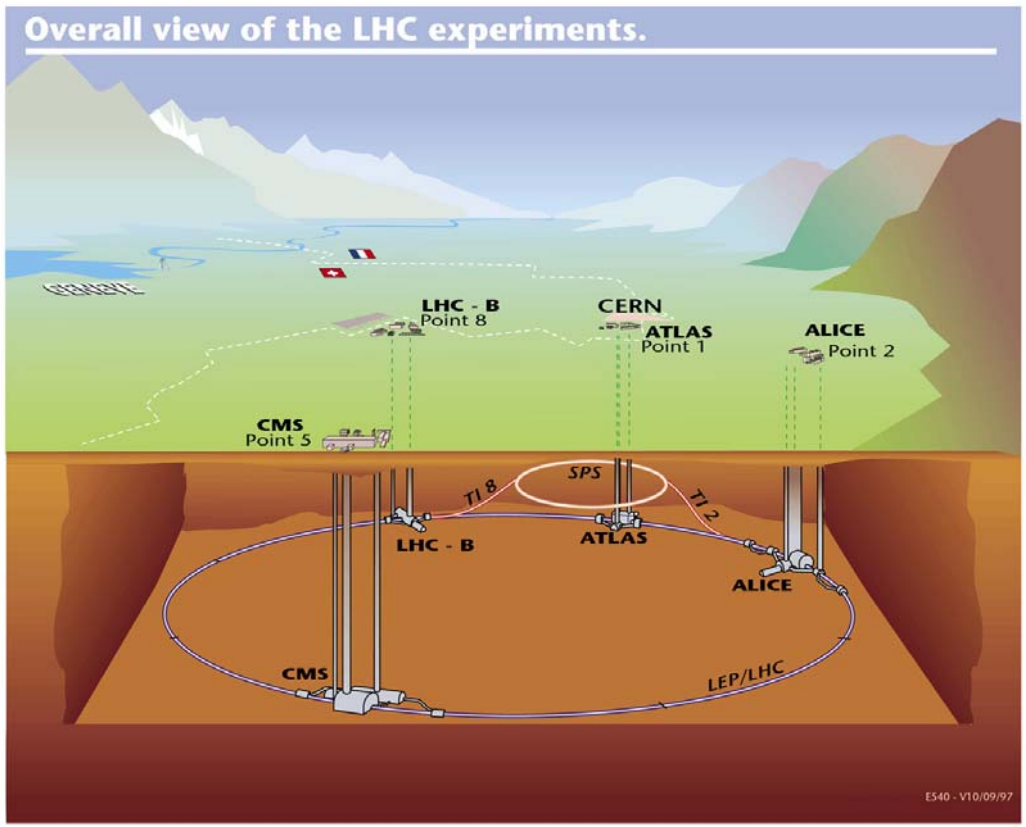
DESY in Germany

CERN

Birthplace of the worldwide web

Large Hadron Collider (LHC)

The LHC at CERN in Geneva will accelerate protons to the highest man-made energies.

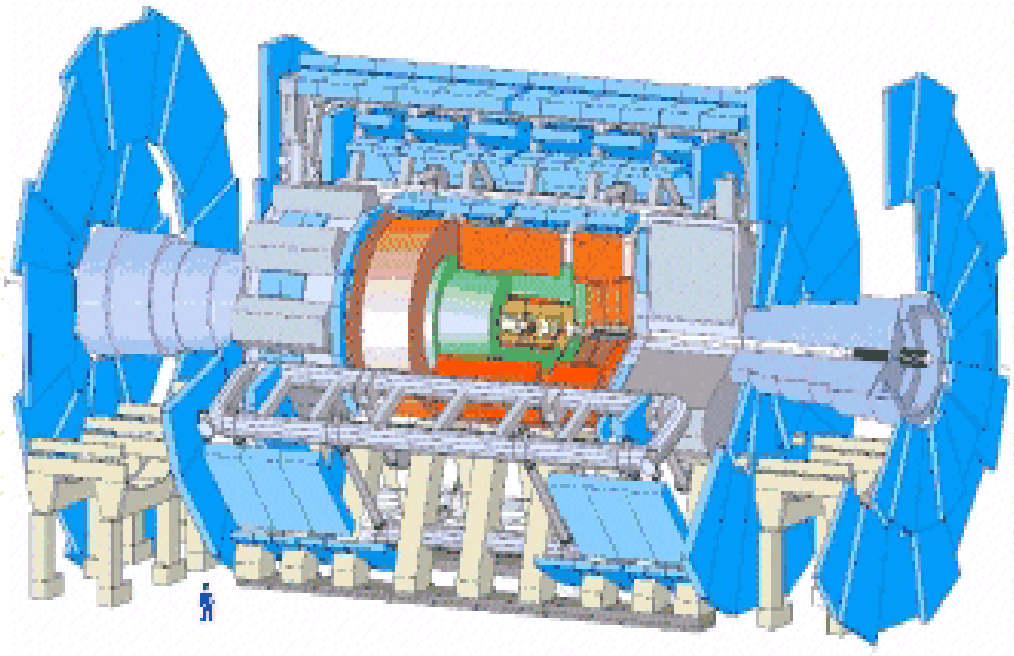


- 27 km circumference with superconducting magnets
- 7 TeV protons (energy of freight train)
- LEP beam energy was sensitive to tidal effects, the level of the lake and the passing of the TGV

ATLAS Collaboration

1800 Researchers from 35 countries

ATLAS Detector



General purpose detector:
high precision tracking
energy measurement
muon detection

$500 M detector
$100 M computing

There is one other general purpose detector at LHC called CMS and two detectors dedicated for specific physics (LHCb and Alice)

Canadian Participation

Montreal
Toronto
York
Carleton

Victoria
UBC
TRIUMF
Alberta

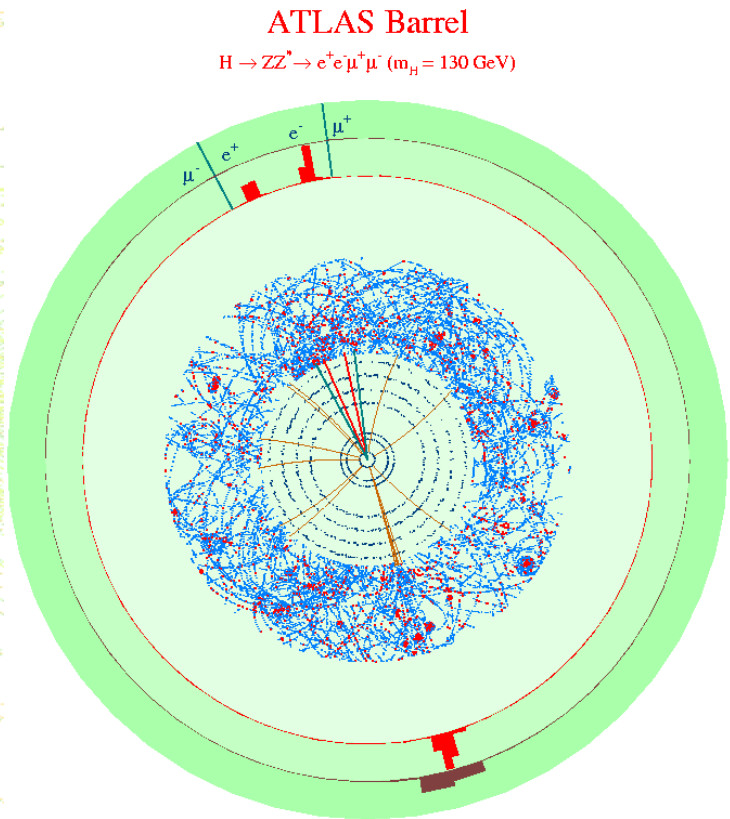
30 faculty at 8 institutions
100 Canadians

NSERC MIG $15 million for detector components that are currently under construction

TRIUMF has contributed $30 million to the LHC accelerator through in-kind contributions of items such as magnets

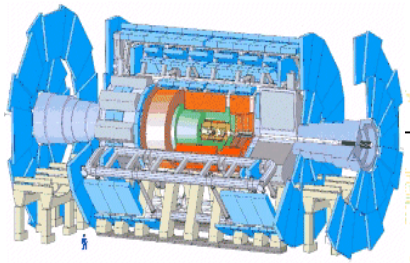
Data rates and sizes

- ***Event*** = result of p-p collision
- 30 million events per second
- Online selection:
  - reduces rate to 100 events/s
  - size to 1-2 MB per event
- Data rate:
  - 100 MB/s or 10 TB/day
  - 100 days/year = 1 PB/year



1 interesting event is buried in a background of 1E13 events

Data flow



Filter selection

Analog and digital information

1 PB/year

Raw Data

Simulated data

1 PB/year

Reconstruction

Recon Data

1 PB/year

Particle tracks and energies

User analysis

LHC Computing Model

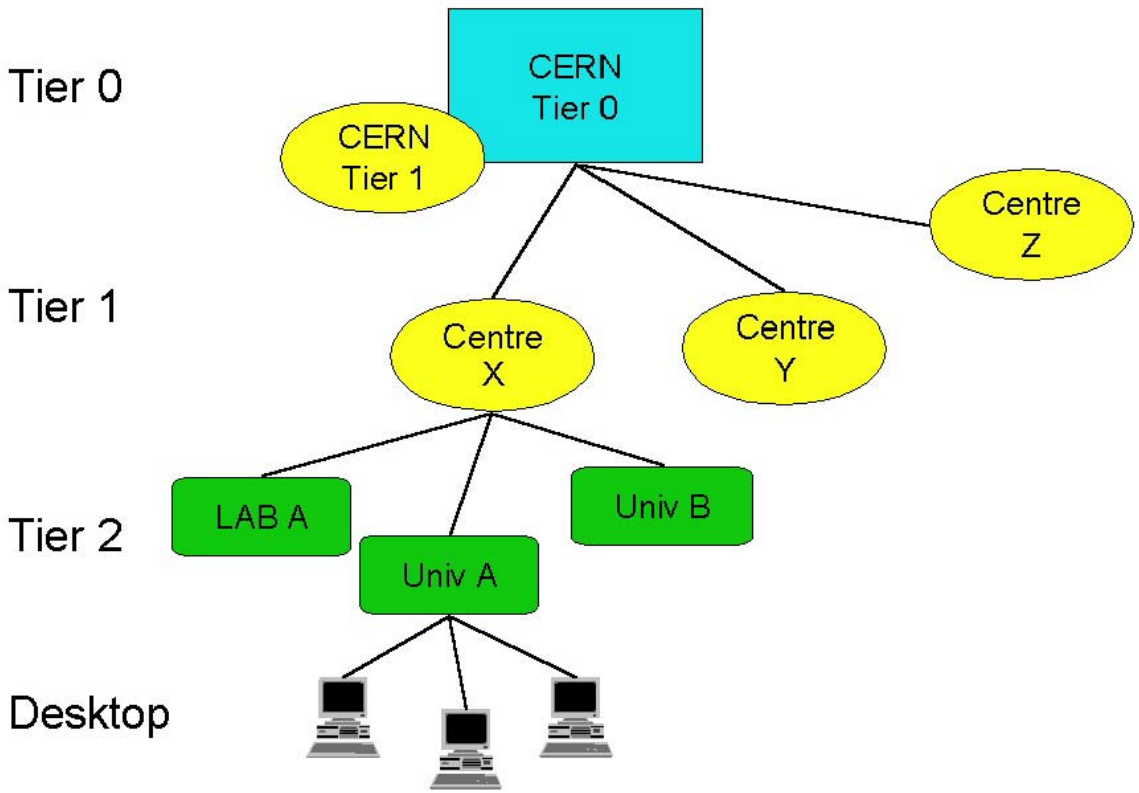
The computing requirements cannot be satisfied at a single site

- Raw data will be stored at the experimental site in CERN
- need to exploit geographically distributed resources
- utilize established computing expertise and infrastructure
- tap into funding sources not usually available to particle physics

LHC Computing Model:

- hierarchical structure with the CERN at the top tier
- Large national or intra-national centres
- National or institutional centres
- Desktop

LHC Computing Model



Tier 0+1 site at CERN

400 TB disk, 9 PB tape,
700,000 SpecInt95

Multiple Tier 1 sites for reconstruction and simulation

400 TB disk, 2 PB tape
200,000 SpecInt95

Tier 2 sites for user analysis

ATLAS Computing in Canada

Tier1:
400 TB disk, 2 PB tape
200,000 SpecInt95

10% in 2004
40% in 2005
100% in 2006

Particle physics computing

Our needs are different than other research groups:

- calculations are mainly integer-based
- our jobs are not parallelizable - we run multiple instances
- Linux (and Solaris) operating systems
- Objectivity database (OO DBMS)
  - sequential events with random access within an event
- simulation: *high CPU and low I/O*
- reconstruction: *low CPU and high I/O*

Contrast with other groups in Canada interested in the Grid

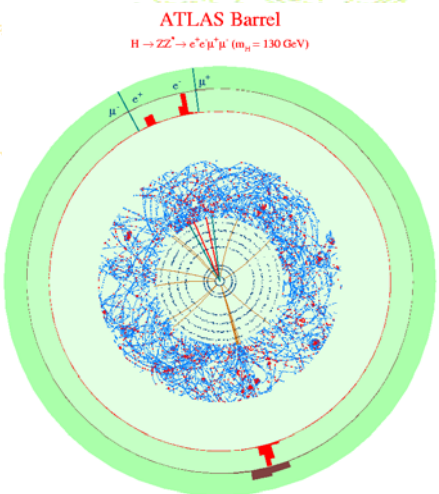
- goal is to create a Grid to run parallel applications

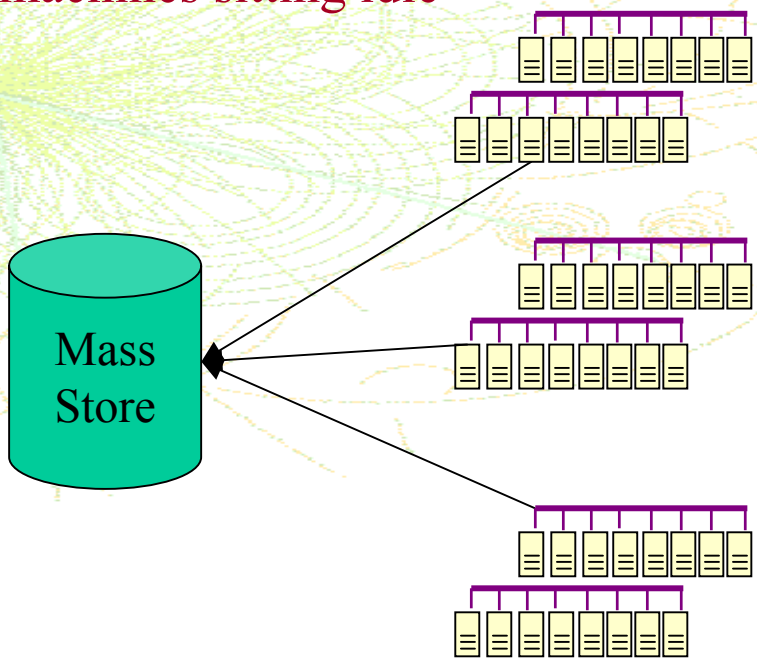
GridUseCase I: Simulation

Particle physics simulations: high CPU and low I/O demands

Compute Grid:

- enable us to use as many machines as possible
- eg. UVic has many classroom machines sitting idle 16 hours/day





GridUseCase II: Reconstruction

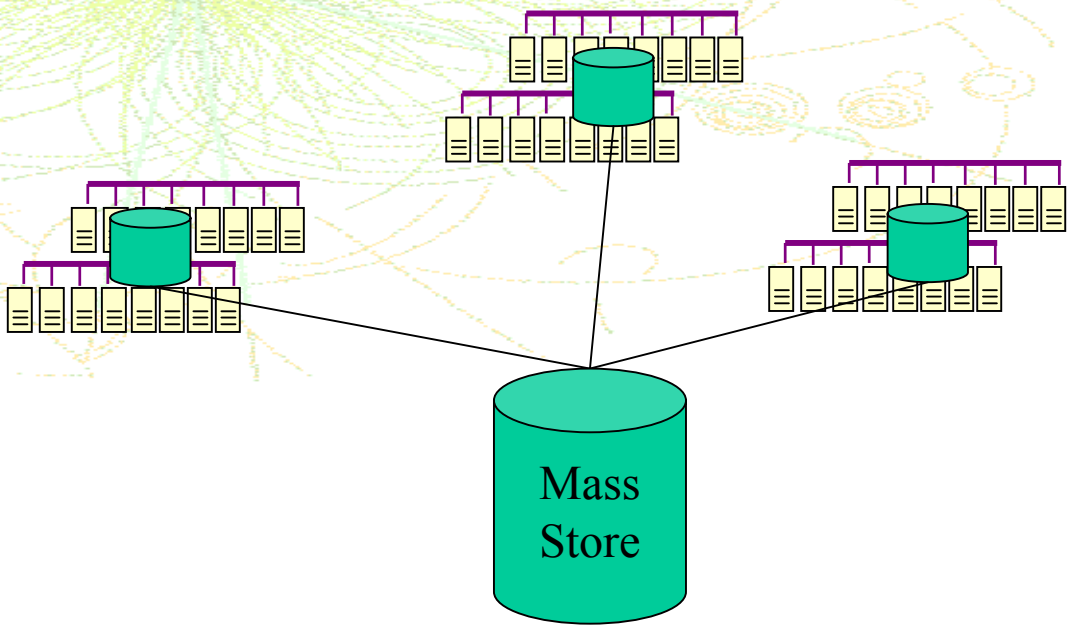
Reconstruction and analysis: low CPU and high I/O demands

Data Grid:

Do we run the jobs where the data is?

Do we move the data to the available CPU's?

Do we put frequently used data at CPU sites?



Grid Activities in CERN

The goal is to make many activities Grid-aware

- data transfers between the regional centres
- production of simulated data
- reconstruction of real/simulated data
- data analysis

Tools (middleware) for the Grid are being developed

- management tools for the jobs, data, computing farms and storage
- monitoring tools for the network and computing facilities
- security and authentication

CERN is leading the Data-Grid Project which will address these issues for particle physics, genome project and space science

Deployment at CERN

- Prototype stage: 2001-2004
  - goal is to have a facility with 50% of the scale required by a single LHC experiment
  - test Grid tools with Regional Centres
- Pilot stage: 2005
  - services set up for each experiment
  - establish LHC Grid
- Production: 2006-2007
  - expansion and conversion of pilot service into production mode

Grid Activities in ATLAS

How can we use the Grid tools for our applications?

- Just beginning to study this area
- How do we use the Grid tools at CERN, our local facilities and couple them into the LHC Grid?

Mock data challenges in ATLAS:

- prototyping and tests are an essential part of learning how to create a world-wide Grid for particle physics
- MDC every year from 2003 starting with 10 TB data sets

Activities in Canada

We are looking at 3 key areas:

- mass storage and hierarchical storage managers (HSM)
  - CFI proposal for 400 TB facility for many applications including particle physics
  - HSM tests with our existing HPC facility
- Linux commercial and commodity clusters
  - 40 node Linux cluster shared between Cosmology and BaBar
  - learn how to link the mass storage and the cluster (high rates)
- Grid
  - Globus testbed for particle physics simulation

CFI International Funds

Project Outlines to the 2 CFI International Funds due July 3

Joint Venture - $100M for 4 projects in Canada

- *Canadian-Tier 1 and multiple Tier 2 facilities starting in 2004*

Access Venture - $100M for projects outside Canada

- *Establish a team of 5 HQ Canadians at CERN to contribute to the LHC Grid for 2002-2007*

Phoenix Grid Project at Victoria

A.Agarwal, R.Kowalewski,
R.Sobie, J.van Uytven

A.De Silva

Funded in part by

Learn how we can create a Grid of our particle physics computing resources in Canada

Project milestones Oct 2000-Oct 2001:

- •establish a TRIUMF-Victoria Testbed using Globus toolkit
- •I/O in the Grid - how to synchronize I/O
- •Use the "Grid-enabled" CONDOR batch scheduler
- •Expand the Grid to a more realistic size - add the Victoria Linux farm and other remote farms

Project Evolution

Long term:

- BaBar experiment is an active experiment at SLAC and uses similar software as ATLAS
- Create a Grid for running BaBar simulation jobs (creating simulated data)

We would like to expand our Grid activities

- increased involvement on particle physics specific applications
- get involved in the Data-Grid project at CERN
  - former graduate student and RA are working on Data-Grid
  - links with Oxford particle physics group in the Data-Grid project
- work together with groups interested in the Grid
  - HPC group at NRC with C3.ca

C3 Grid

We should emulate the DataGrid project for the C3 Grid

Four main areas each with separate work packages:

- middleware
  - work scheduling, data management, monitoring services, fabric management, mass storage management
- infrastructure
  - testbed and demonstrators, network services
- applications
  - Particle Physics, …?
- management
  - dissemination, project management

Strategy

Create a Grid Project rather than fund through MFA
(follow the European and US models)

- give access to funds from other NSERC GSC's
  - subatomic physics GSC would likely provide funding
  - other Grid applications
- look for provincial and industrial matching

Ideas? Go for a small request to NSERC this fall ($250K).

Summary

Exciting period in particle physics

- physics, accelerator, detector and computing

Grand computing challenge

- petabtyes of data, thousands of computers and users

The Grid may be the answer to our computing requirements.

- A large world-wide effort has begun to develop the Grid
- Particle Physics is an ideal field to experiment and test the Grid technologies.

Canada needs to significantly increase its effort in the development of the infrastructure and applications