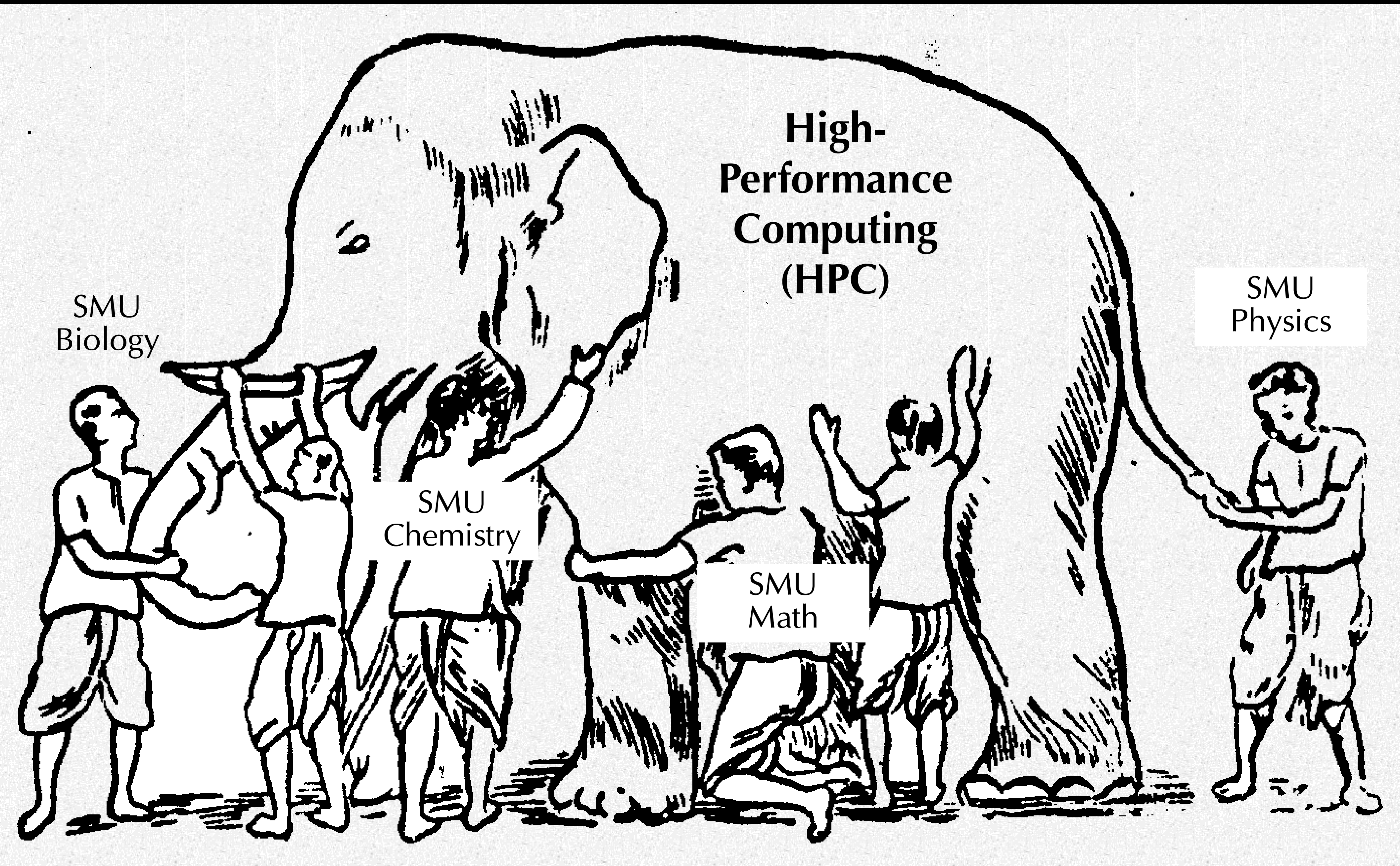
Computing the Collider

Stephen Sekula
SMU Department of Physics
Presented at the
SMU Interdisciplinary STEM Forum
April 30, 2012


High-
Performance
Computing
(HPC)
SMU
Biology
SMU
Chemistry
SMU
Math
SMU
Physics

What's the Problem We Are Solving?

- **We want to understand the laws of physics as they existed near the beginning of time**
- Problem: the universe, as far as we know, came into being once and we were not there to see it
- Solution:
  - recreate a moment in time near the beginning of time
  - do this many times to build up a solid, statistical picture of those moments in time
  - understand the laws of physics by studying these moments in aggregate

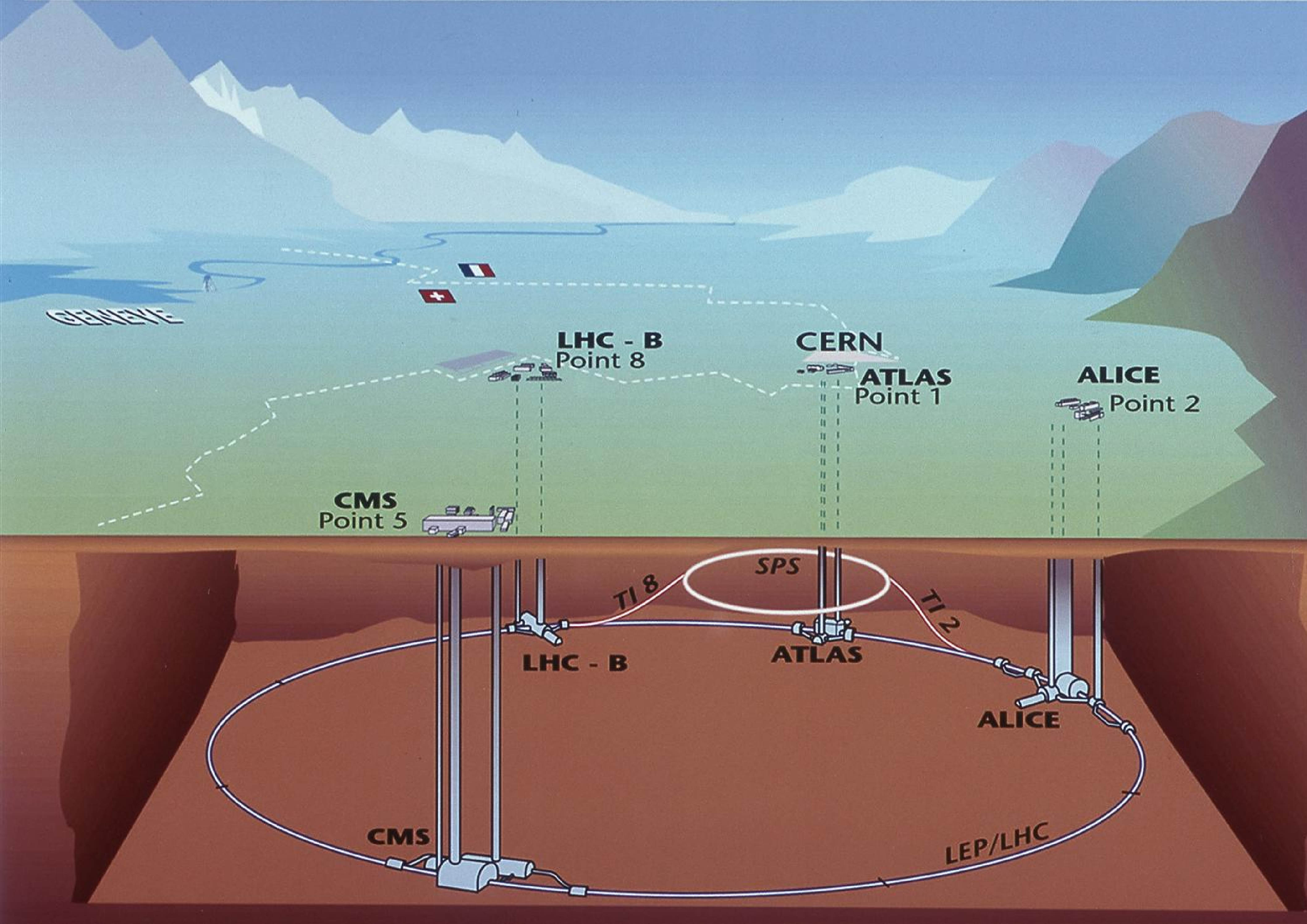
What is the Principle?

- Quantum physics and Relativity have together taught us that energy, time, and space are related to one another
- High energies probe short distances and early times in the universe
  - ANALOGY: short wavelengths of light allow you to probe small structures. Cells can be observed with visible light, but DNA can only be observed with x-rays, which have a much shorter wavelength
- The early universe was small, hot, and dense . . . this means we need high energies to understand it.

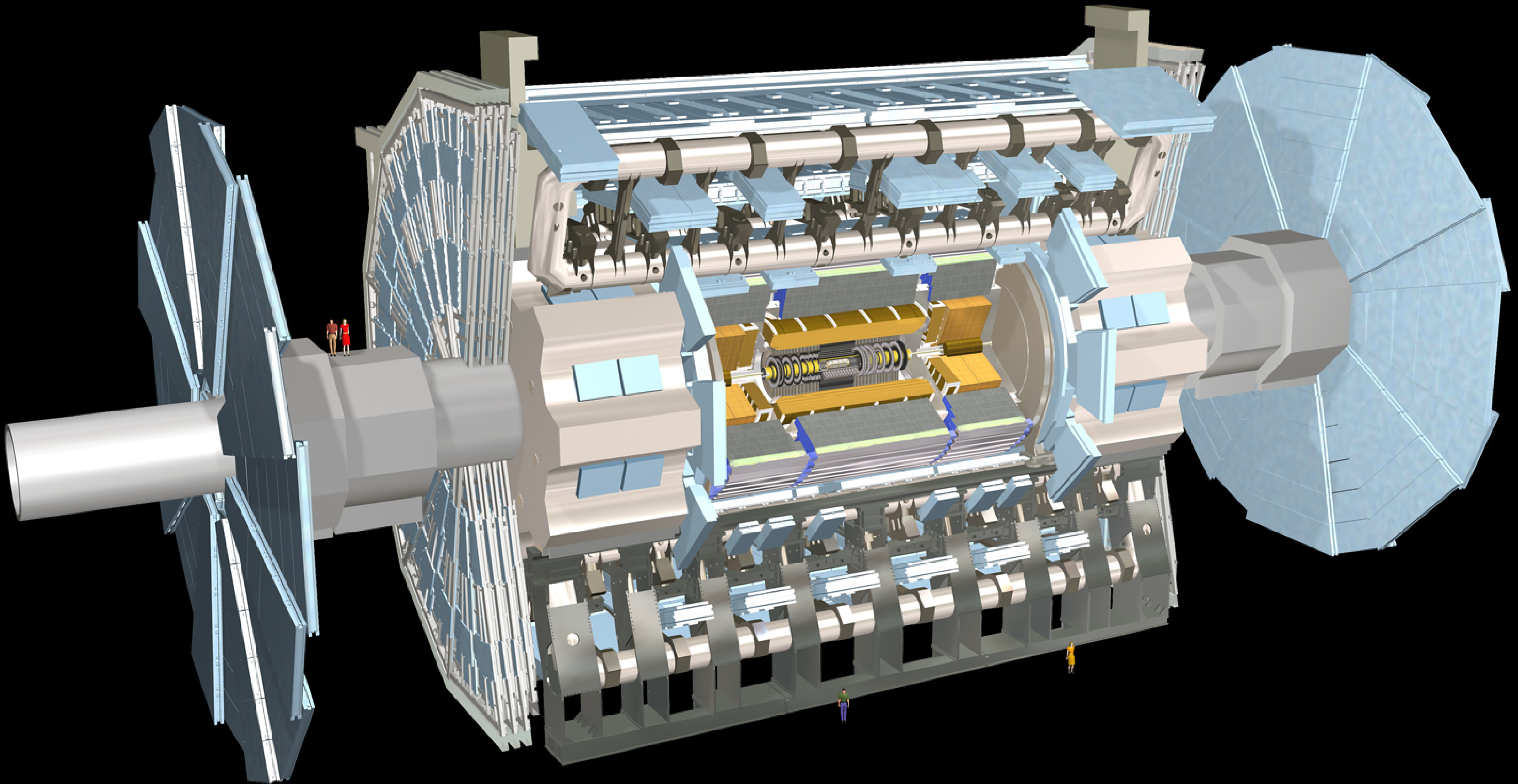
A Comment on Energy

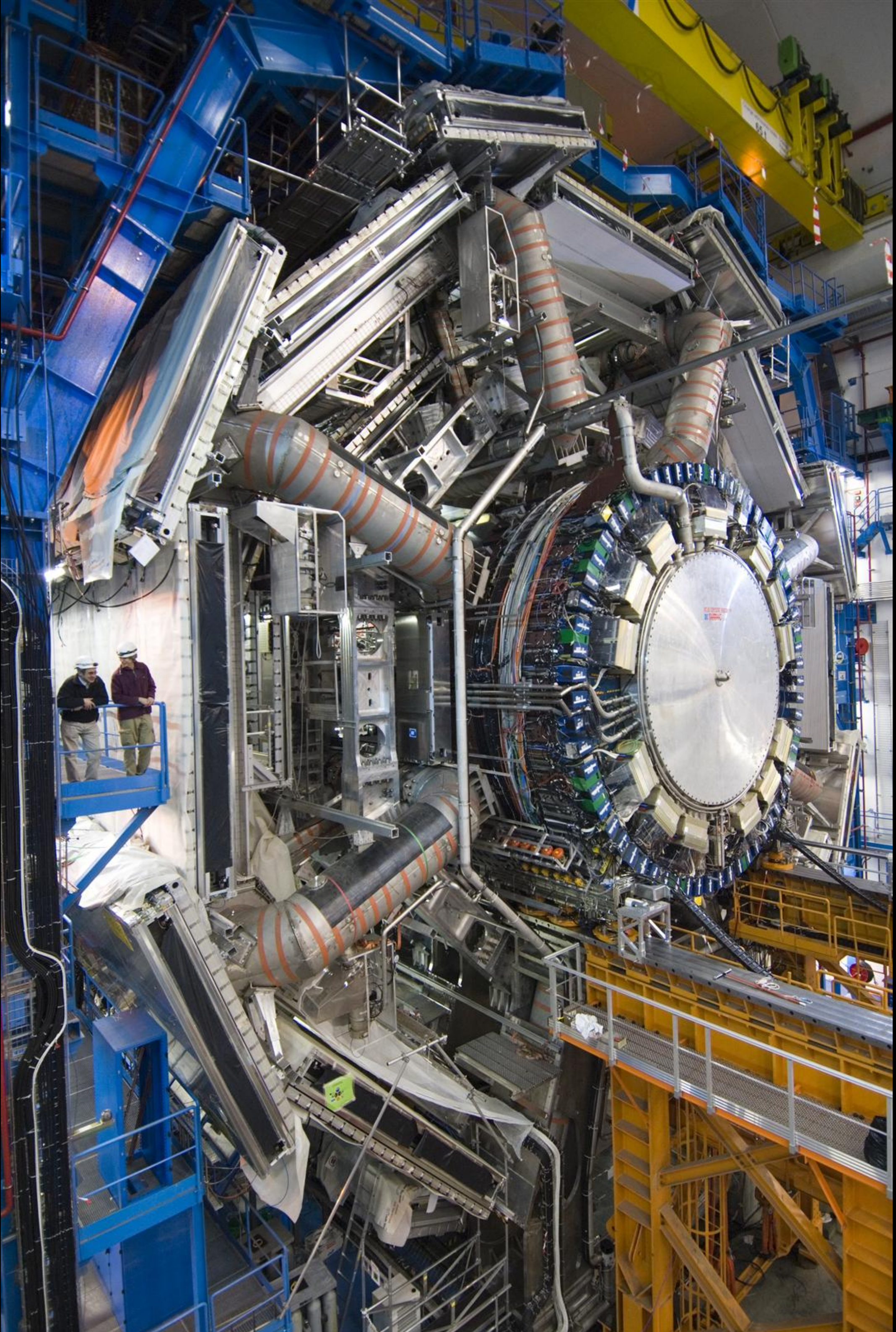
- In subatomic physics, we use the “electron-Volt” (eV) as the basis of energy – it’s conveniently sized
- We accelerate protons (volume of approximately 4.2×10^{-45} m^3) up to 8 Trillion eV (TeV)
  - how do you understand what this means? A 12 oz soda (Coke) contains 140 food calories, which is about 3.7×10^{15} TeV
  - our protons don’t seem impressive, until you consider the ENERGY DENSITY (energy per unit volume)
    - Coke Energy Density: 9.8×10^{16} TeV/m^3
    - LHC Proton Energy Density: 1.9×10^{45} TeV/m^3

Overall view of the LHC experiments.



The ATLAS Experiment

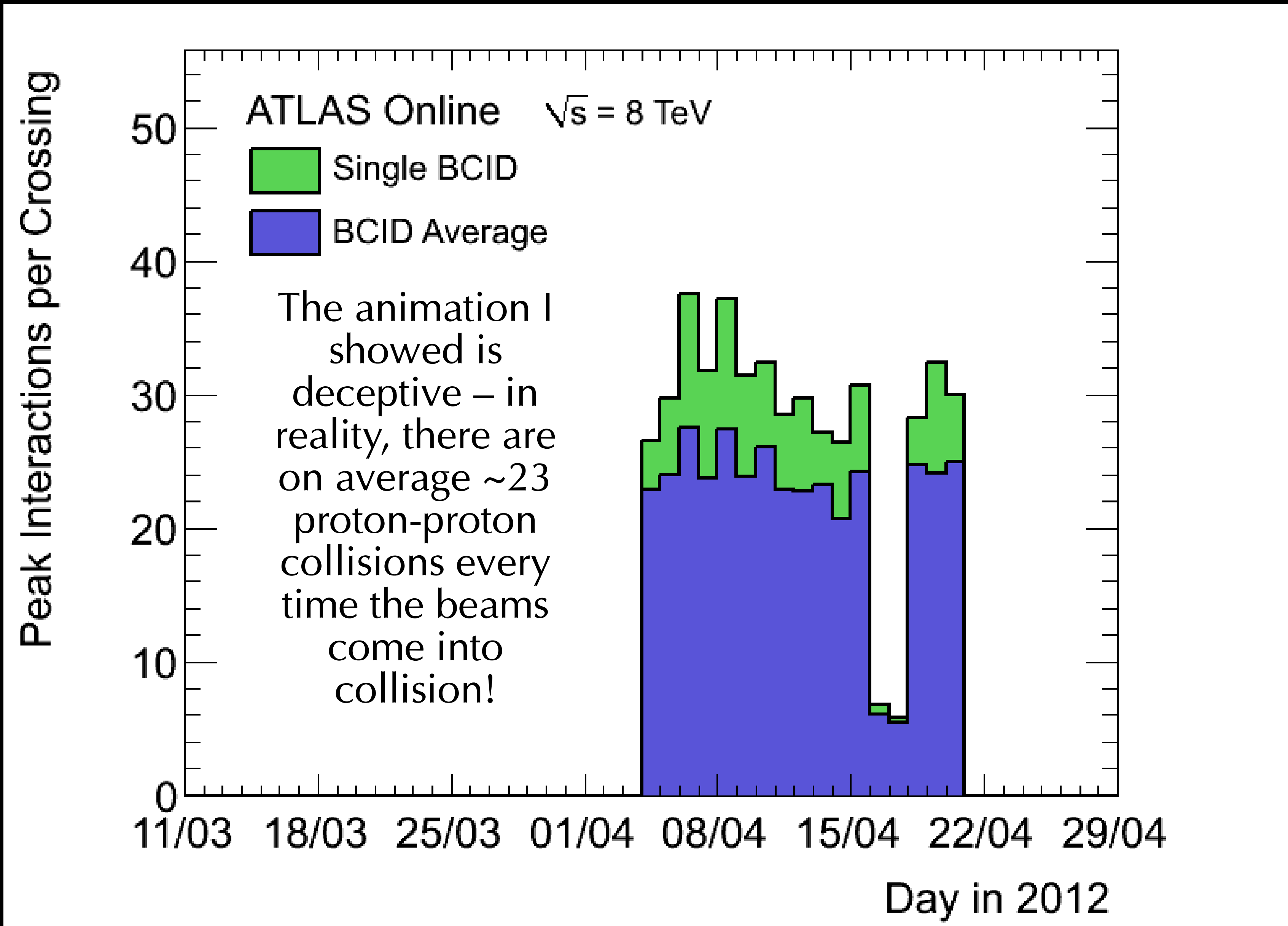




ilicon Strips	6.3 M
ransition Radiation Tracker	350 k
l Calorimeter	170 k
lorimeter	9800
nic endcap LAr calorimeter	5600
rd LAr calorimeter	3500
Calo trigger	7160
Muon RPC trigger	370 k
Muon TGC trigger	320 k
Muon Drift Tubes	350 k
athode Strip Chambers	31 k
arrel Muon Chambers	370 k

The ATLAS Detector, with its millions of channels of data, is a tremendous information challenge.

The Computing Challenge - I

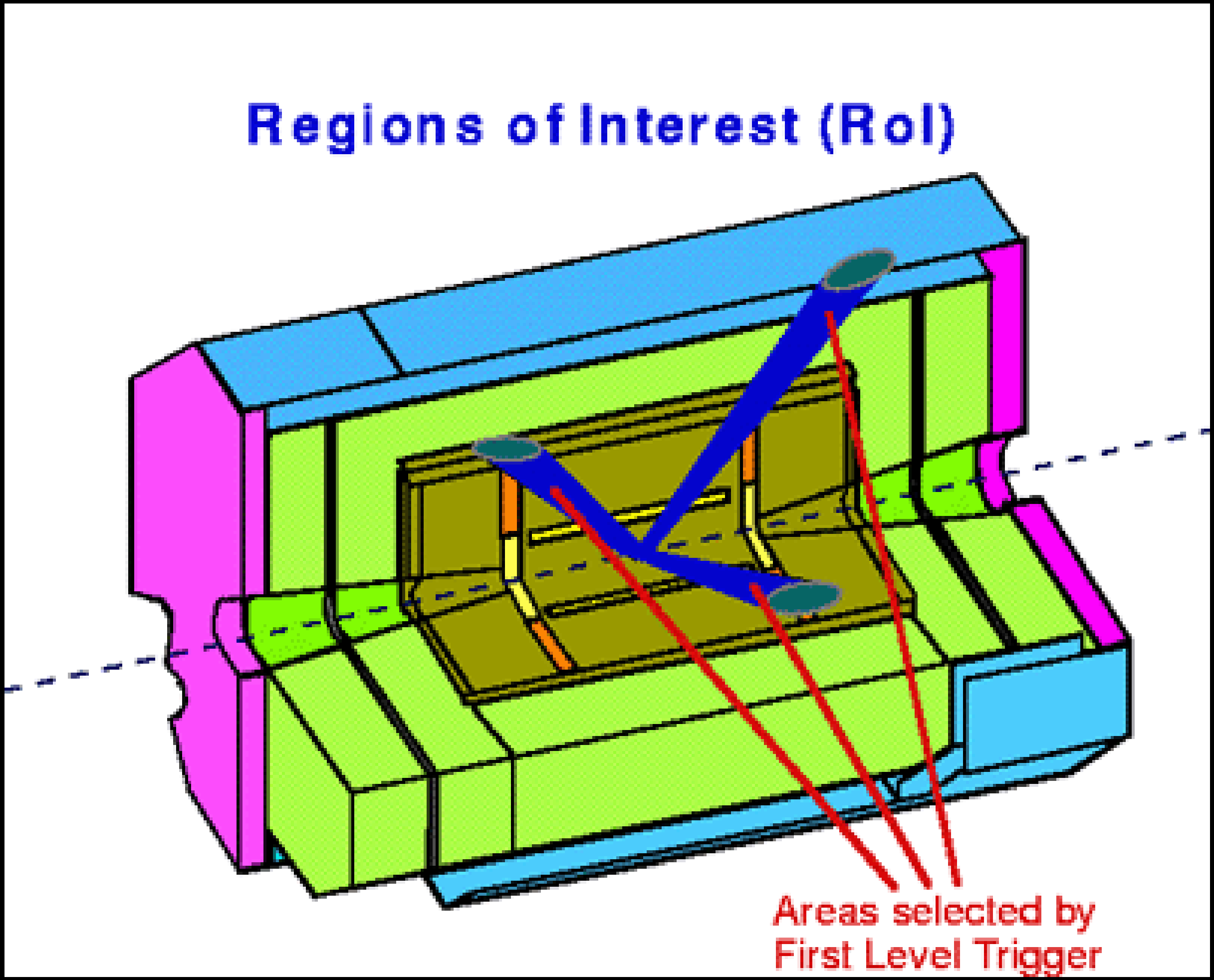


The Computing Challenge - II

- There are 40 million LHC proton bunch collisions each second that the LHC is operating.
- Each proton-proton collision yields about 25 MB of information (1.6 MB if zero suppression is applied)
- There are, on average, about 24 proton-proton collision events per bunch crossing
- This means, if we kept EVERY bunch crossing, we would have to store 23 PETABYTES each second.
- Do the math – what if we tried to keep EVERY "bunch collision?"
  - Assume the LHC ran 24 hours per day for 9 months in 2011.
  - That would yield 2.2×10^7 PETABYTES of data per year.

  The total amount of printed material in the world is estimated at 200 Petabytes; the total amount of estimated digitally stored information in the world was 5×10^5 Petabytes as of 2009.

The Computing Challenge - III



The ATLAS Trigger systems helps us to prune the uninteresting events and retain the interesting ones.

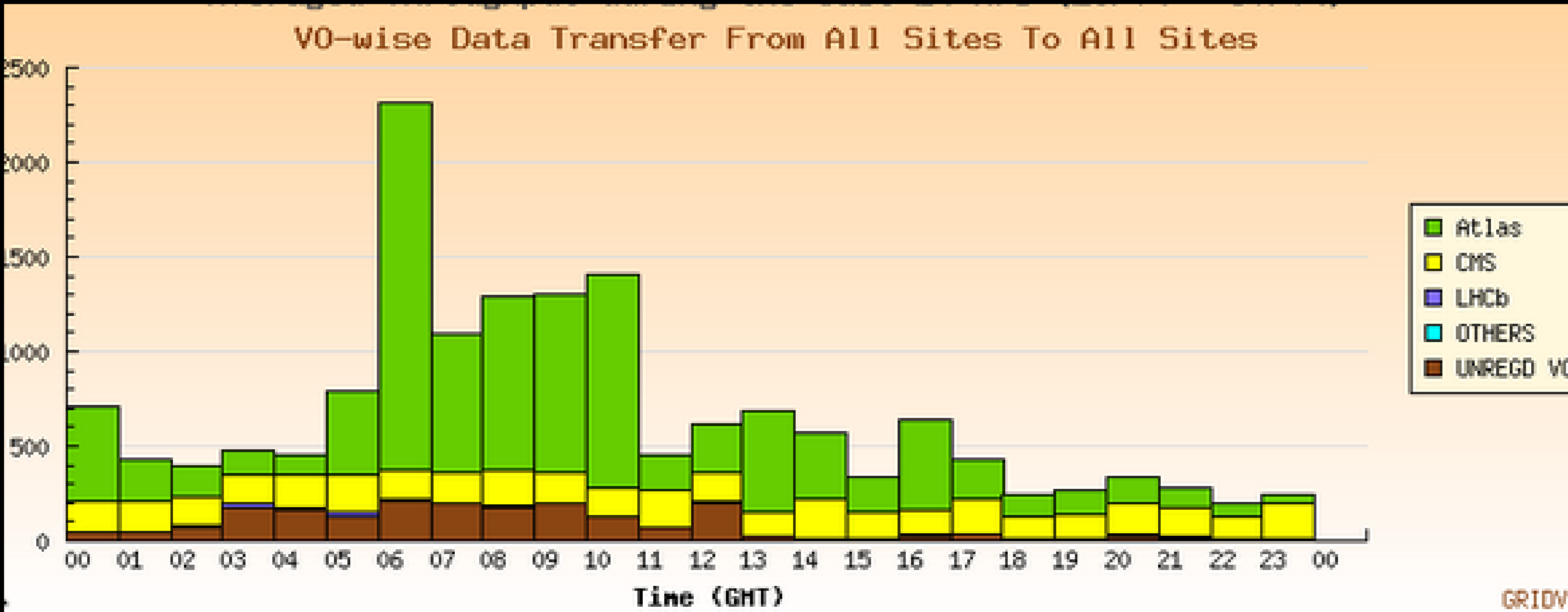
- LEVEL-1: identifies “regions of interest” in the detector and reduces the data rate from 40 MHz to 100 kHz
- LEVEL-2: further refines the assessment of LEVEL-1, reducing data stream to a rate of 1 kHz
- High-Level Trigger: a full event reconstruction is performed on events from LEVEL-2. Output data rate is a few hundred events/s

The Computing Challenge - IV

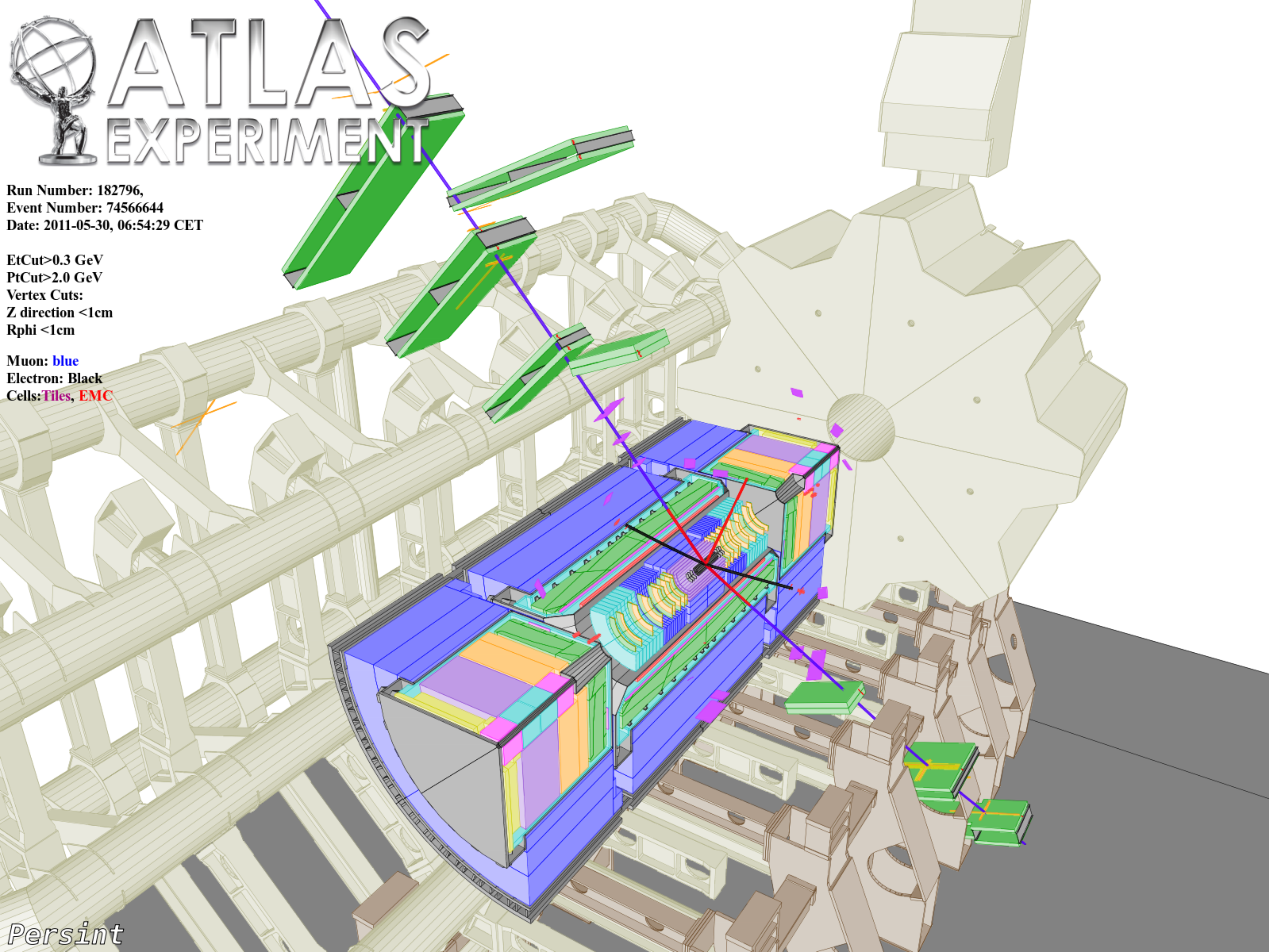
- RAW data events are stored at a rate of about 100 MB/s after the application of triggering
  - this means at least a PETABYTE of data per year of operation – still too much to do analysis at a single location in the world
- ATLAS and other LHC experiments have helped develop and are utilizing “The GRID” - a world-wide, connected system of computers and networks
  - Principle: you develop your analysis locally, submit it to the GRID. The analysis is “agnostic” about where data comes from or where it had to be run; the GRID handles the details and returns results to you

The GRID and SMUHPC

- The GRID sounds great, but . . .
  - with thousands of users, your priority for access to resources is never guaranteed
  - high-priority tasks from the experiments take precedent over individual user analysis
  - running your analysis could take hours, days, or weeks, and require constant vigilance to resubmit failed jobs
- SMUHPC is a critical complimentary resource
  - typical physicist-level data formats can be copied to SMU and take up ~10-20 TB of disk space
  - these can be processed in a matter of hours, with priority dictated by past usage levels and the number of other concurrent users (as well as special requests to briefly raise priority)



This map shows only GRID sites with "Tier 1" or "Tier 2" status


ATLAS EXPERIMENT
Run Number: 182796,
Event Number: 74566644
Date: 2011-05-30, 06:54:29 CET
EtCut>0.3 GeV
PtCut>2.0 GeV
Vertex Cuts:
Z direction <1cm
Rphi <1cm
Muon: blue
Electron: Black
Cells:Tiles, EMC
Persint

The Computing Challenge - V

- But wait . . . there is more!
  - we don't just collect data – we simulate what we expect from the collider using all previous accumulated physics knowledge
  - this allows us to make predictions:
    - what will the data look like in the absence of new discoveries (e.g. in the absence of a Higgs Boson)?
    - what will the data look like if there IS a new particle present?
- SMUHPC has provided a critical platform for generating simulated data for use in official ATLAS physics program, including the search for the Higgs Boson


Data, 2011 – a hint of the Higgs Boson
Will it be stuff we already knew about?
W- side
2-SIDED
W+
gluon
top quark
or the real Higgs Boson?