第三章：实验统计分析工具ROOT

授课人：董燎原
中国科学院高能物理研究所

Introduction to ROOT

基于面向对象设计的数据分析软件框架！

Nature 522 (2015) 68-72

What is CERN ?

- An international research organization „*Conseil Européen pour la Recherche Nucléaire*" 欧洲核子研究中心

- Founded in 1954 by 12 members

- These days there are 22 member states

- One of the world's largest research labs at the Franco-Swiss border near Geneva 瑞士日内瓦
    – more than 10,000 (visiting) scientists work here
- Finding out what the Universe is made of and how it works.
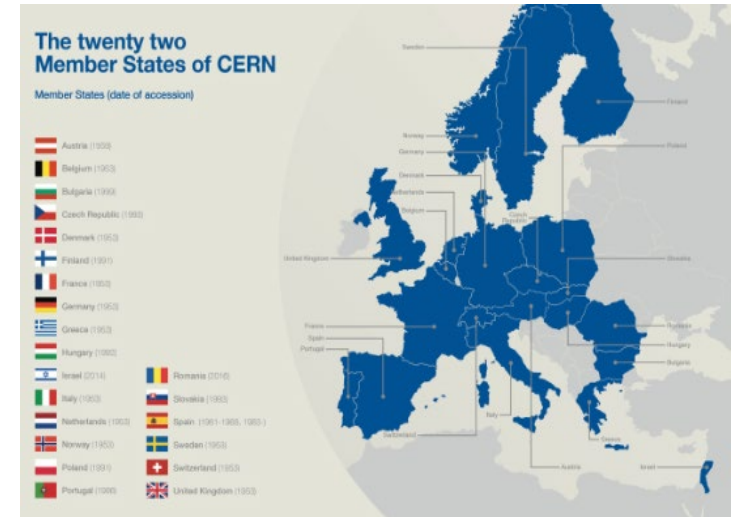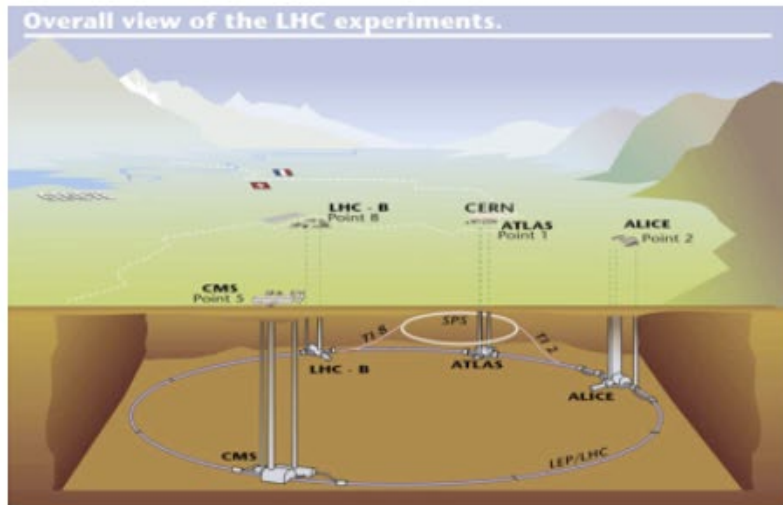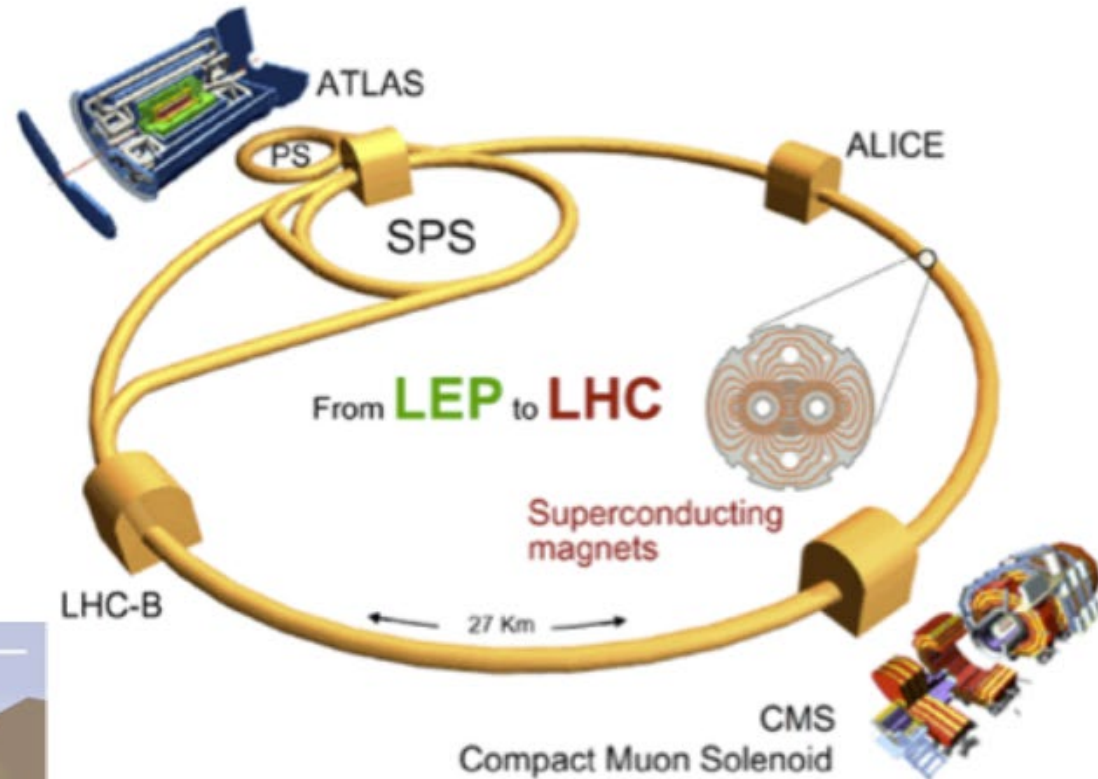
The LHC at CERN

LHC — the Large Hadron Collider — is a 27-kilometre collider ring.

Four large experiments use the LHC as a giant microscope

Started in 2009 and will run until approx. 2030

Overall view of the LHC experiments.

100 m below surface

ATLAS和CMS是交叉确认任何新发现的关键！如 2013年3月两个实验宣布发现**Higgs玻色子**。

Computer Science at CERN

- To fulfill its goals，CERN has always been pushing the limits of computing technology.

✓ While searching for a solution for remote information exchange between physicists Tim Berners-Lee invented the web.

世界第一个网站于1990年12月20日上线，当属的网页属于CERN。直到1991年8月才正式公开，成为互联网的基础！

✓ The immense data collected at CERN required the LHC Computing Grid for distributed processing and storage, anticipating many cloud concepts.

✓ In general, strong do-it-yourself attitude.

What's ROOT ?
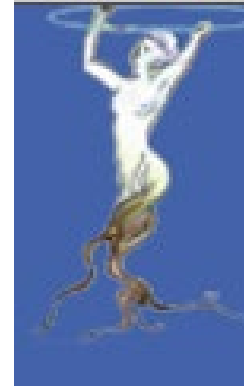
Brief overview of ROOT and its functionality before starting to use it

ROOT in a nutshell

- Framework for large scale data handling
- Provides, among others,
  - an efficient data storage, access and query system (PetaBytes)
  - advanced statistical analysis: histogramming, fitting, minimization and multi-variate analysis algorithms
  - scientific visualization: 2D and 3D graphics, Postscript, PDF, LateX
  - geometrical modeler
  - PROOF parallel query engine

- An Open Source Project
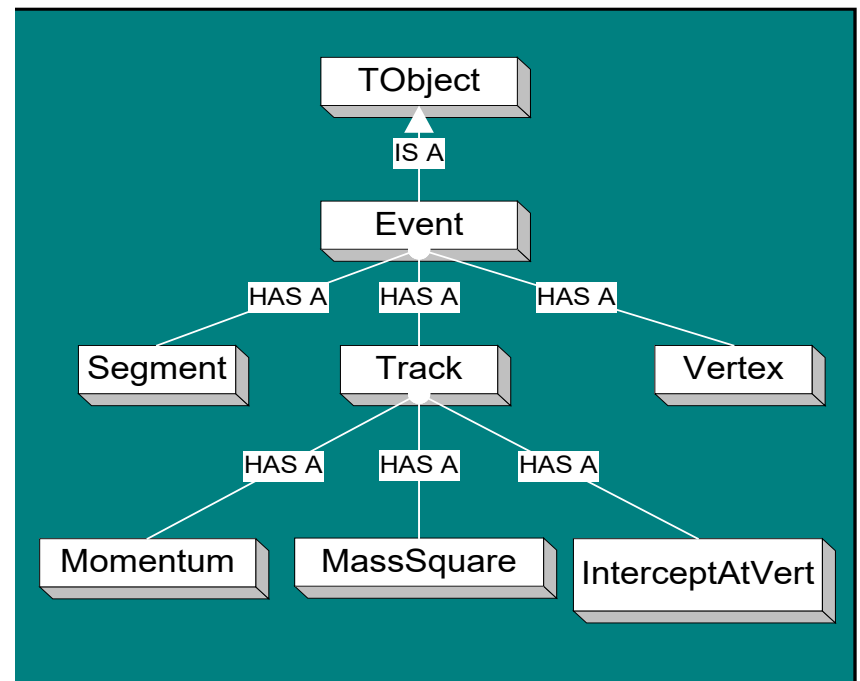
ROOT: An Open Source Project

- Started in 1995 by *R. Brun* and *F. Rademakers.*

- Available (including the source) under GNU LGPL.

- 7 full time developers at CERN, plus 2 at Fermilab (Chicago).

- Many contributors from high energy physics experiments which uses ROOT as base for their software frameworks.

- Large number of part-time developers.

- Several thousands of users giving feedback and a very long list of small contributions.
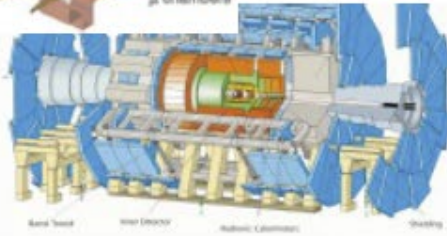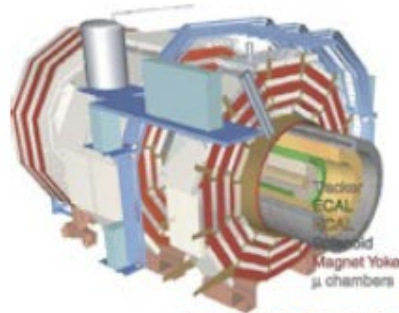
Why ROOT ?

- The analysis of data coming from LHC experiments and not only requires a powerful and general toolkit
  - Visualisation
  - Statistical studies
  - Data reduction
  - Multivariate techniques

- A scalable and reliable persistency method is needed to write the data on disks and tapes.
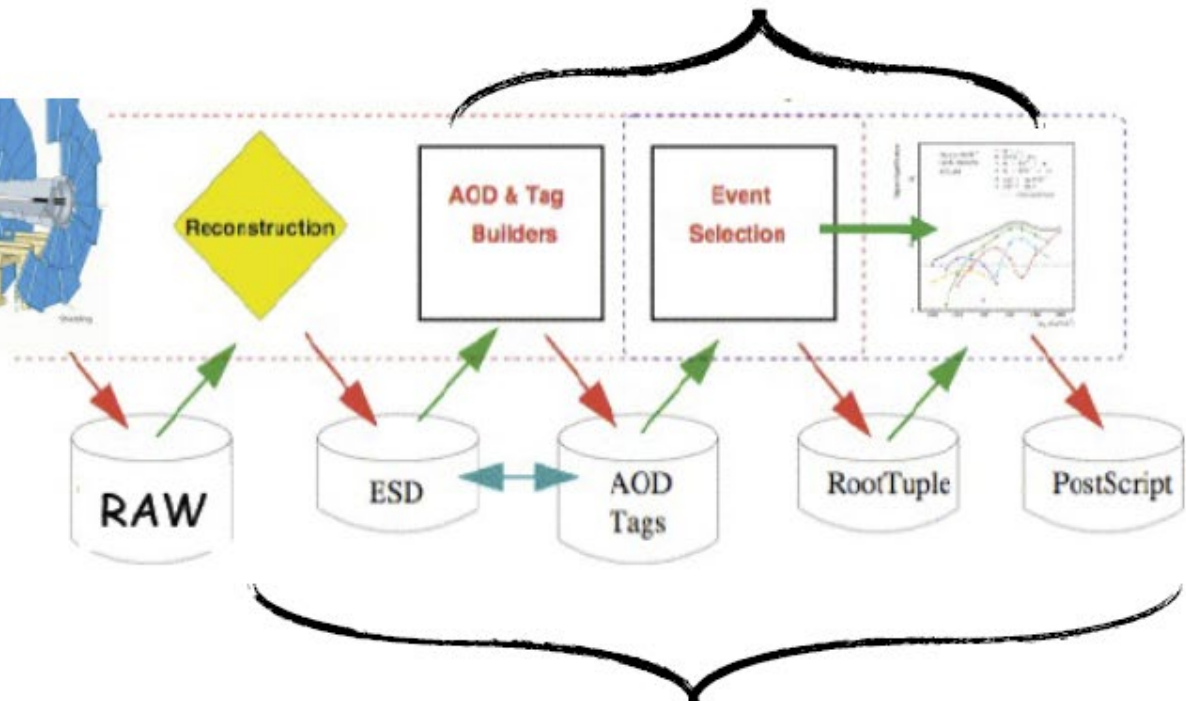
Object Oriented Concepts

- Class: the description of a "thing" in the system
- Object: instance of a class
- Methods: functions for a class

- Members: a "has a" relationship to the class.
- Inheritance: an "is a" relationship to the class.

ROOT Application Domains

Data Analysis & Visualization

General Framework

Reconstruction

AOD & Tag Builders

Event Selection

RAW

ESD

AOD Tags

RootTuple

PostScript

Data Storage: Local, Network

ROOT Services/Utilities

- Histogramming, Fitting, Statistics

- Graphics (2D, 3D)

- I/O to file or socket: specialized for histograms, Ntuples (Trees)

- Collection Classes and Run Time Type Identification

- User Interface
  - GUI: Browsers, Panels, Tree Viewer
  - Command Line interface: C++ interpreter CINT
  - Script Processor (C++ compiled ⇔ C++ interpreted)

ROOT Libraries

- Overview of ROOT libraries and their dependencies

  - 1,700,000 lines of code.
  - More than 100 shared libraries
  - Fully cross-platform: Unix/ Linux, MacOS and Windows.
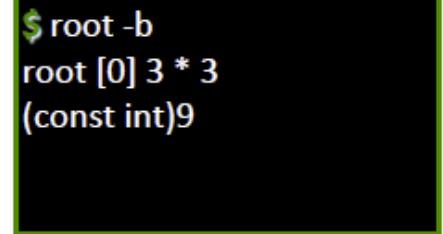 - More than 500000 downloads (since 1997)

ROOT Libraries Dependencies
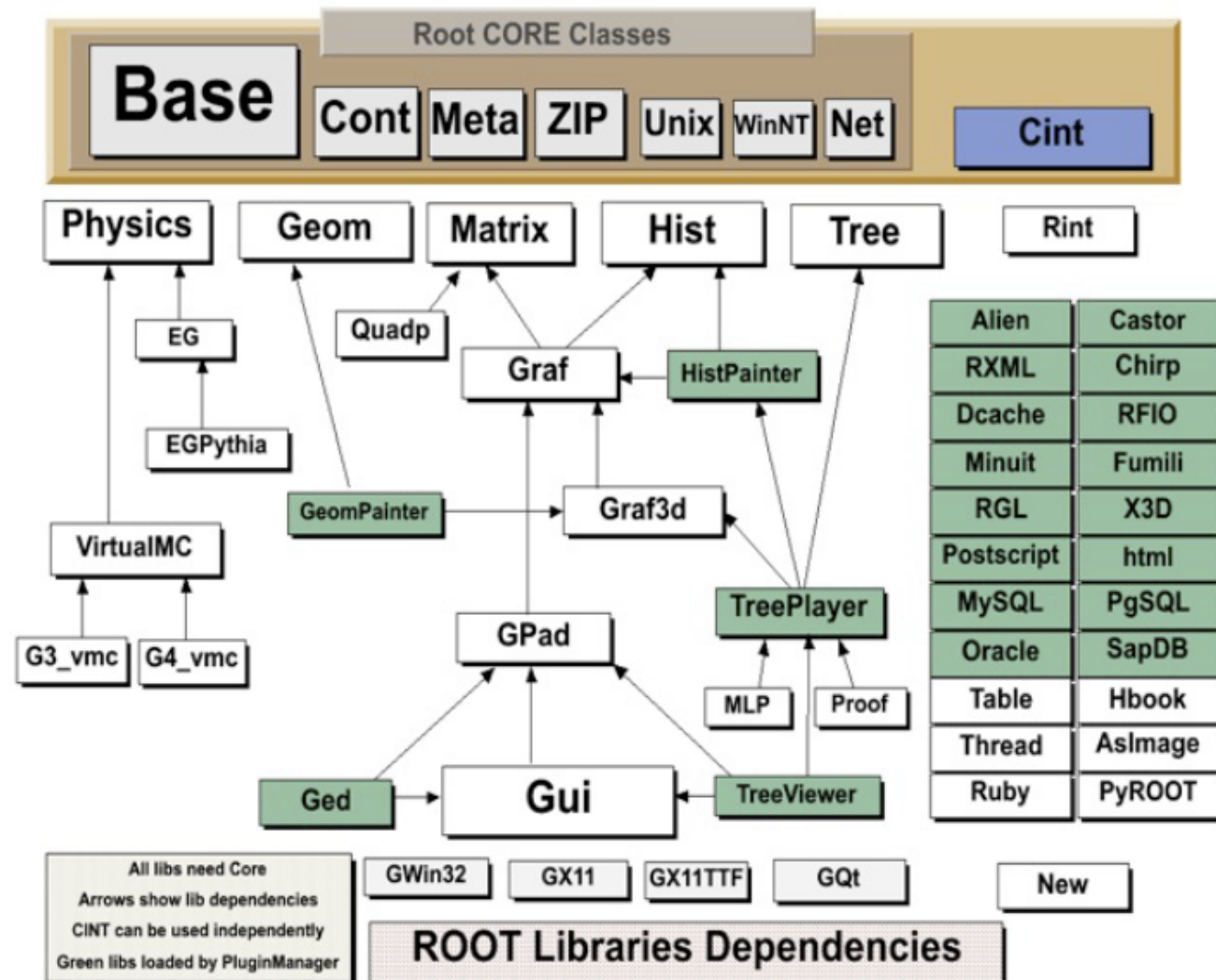
More ROOT Packages

The packages provided by ROOT include those for

- *histogramming* and *graphing* to visualize and analyze distributions and functions,
- *curve fitting* (regression analysis) and minimization of functionals,
- *statistics* tools used for data analysis,
- *matrix* algebra,
- *four-vector* computations, as used in high energy physics,
- standard *mathematical functions*,
- *multivariate data analysis*, e.g. using neural networks,
- *image manipulation*, used e.g. to analyze astronomical pictures,
- access to *distributed data* (in the context of the Grid),
- *distributed computing*, to parallelize data analyses,
- *persistence* and *serialization* of objects, which can cope with changes in class definitions of persistent data,
- *access to databases*,
- 3D visualizations (geometry)
- creating files in various *graphics formats*, like PostScript, JPEG, SVG,
- interfacing Python and Ruby code in both directions,
- interfacing Monte Carlo event generators.

The Interpreter

- ROOT is shipped with an interpreter, CINT*
  - C++ not trivial to interpret and not foreseen in the language standard!
- Provides interactive shell.
- Can interpret "macros" (not compiled programs)
  - Rapid prototyping possible.
- ROOT provides also Python bindings:
  - Can use Python interpreter directly after a simple import ROOT !!

```
$ root -b
root [0] 3 * 3
(const int)9
```

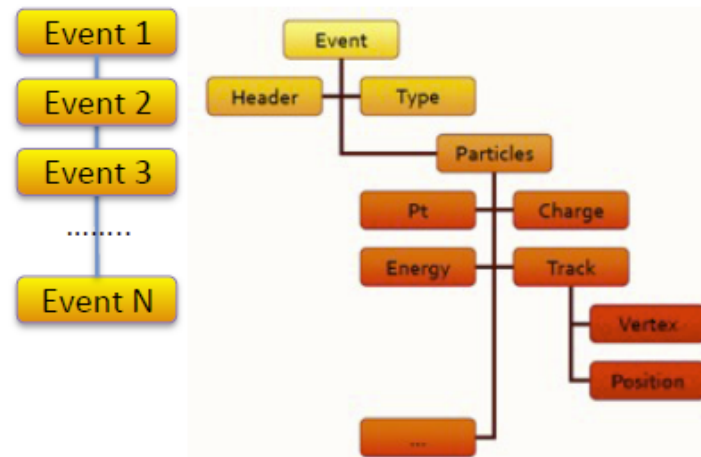* Starting from ROOT 6, the interpreter will be Cling (based on Clang)

Persistency (I/O)

- ROOT offers the possibility to write C++ objects to file
  - Extraordinary: impossible in native C++!
  - Used for petabytes/year rates of LHC detectors.

  > Cornerstone of the storage of experimental data

  - Achieved with serialization of the object using the reflection provided by the CINT dictionary.
- Single objects, collections, object trees
  - Basically one method for all ROOT objects: `TObject::Write`

LEP style flat n-tuples evolved in more efficient trees (fast access, read ahead)

ROOT Math/Stat Libraries

- ROOT provides a reach set of mathematical libraries and tools needed for event reconstruction, simulation and statistical data analysis
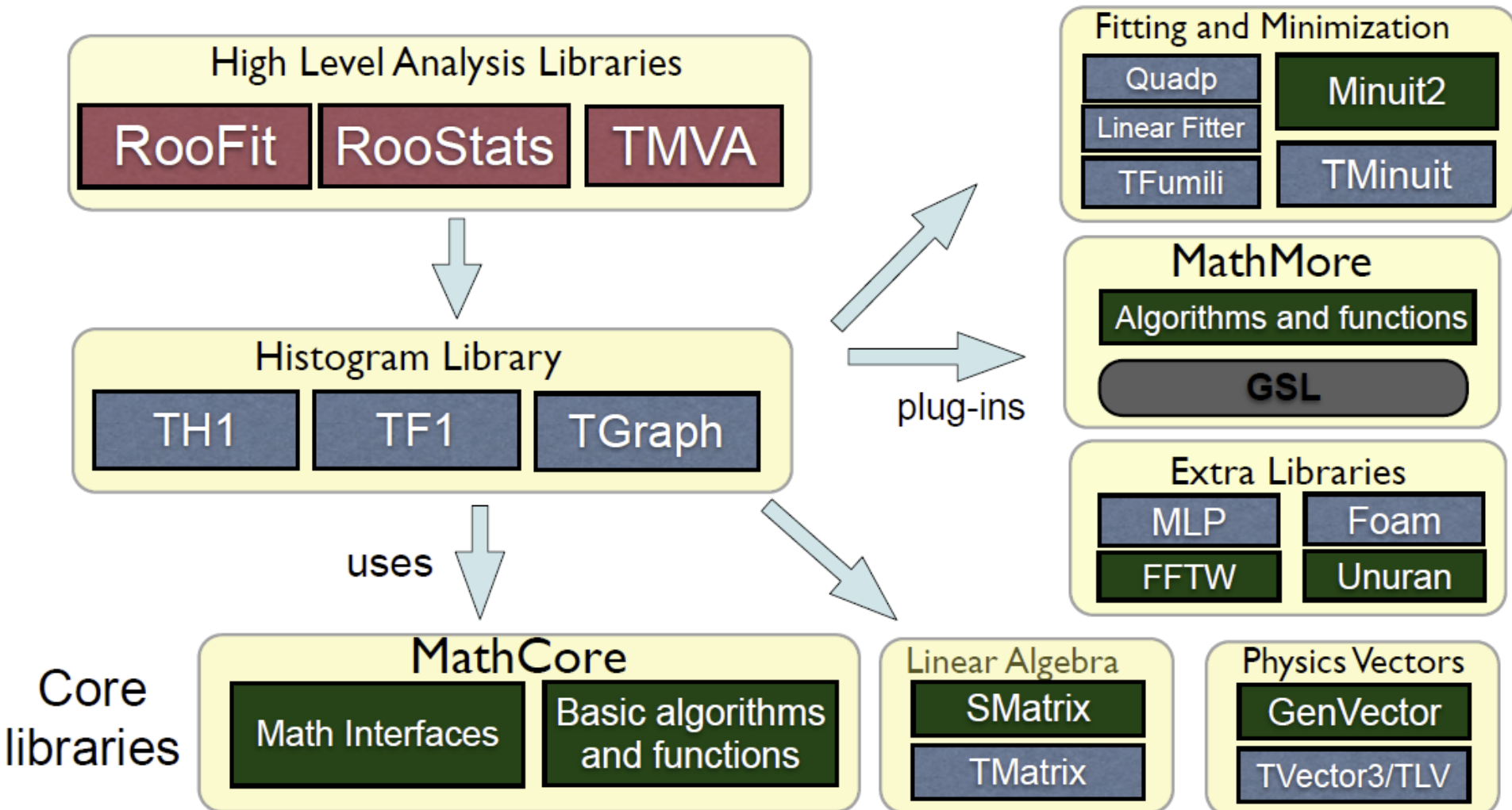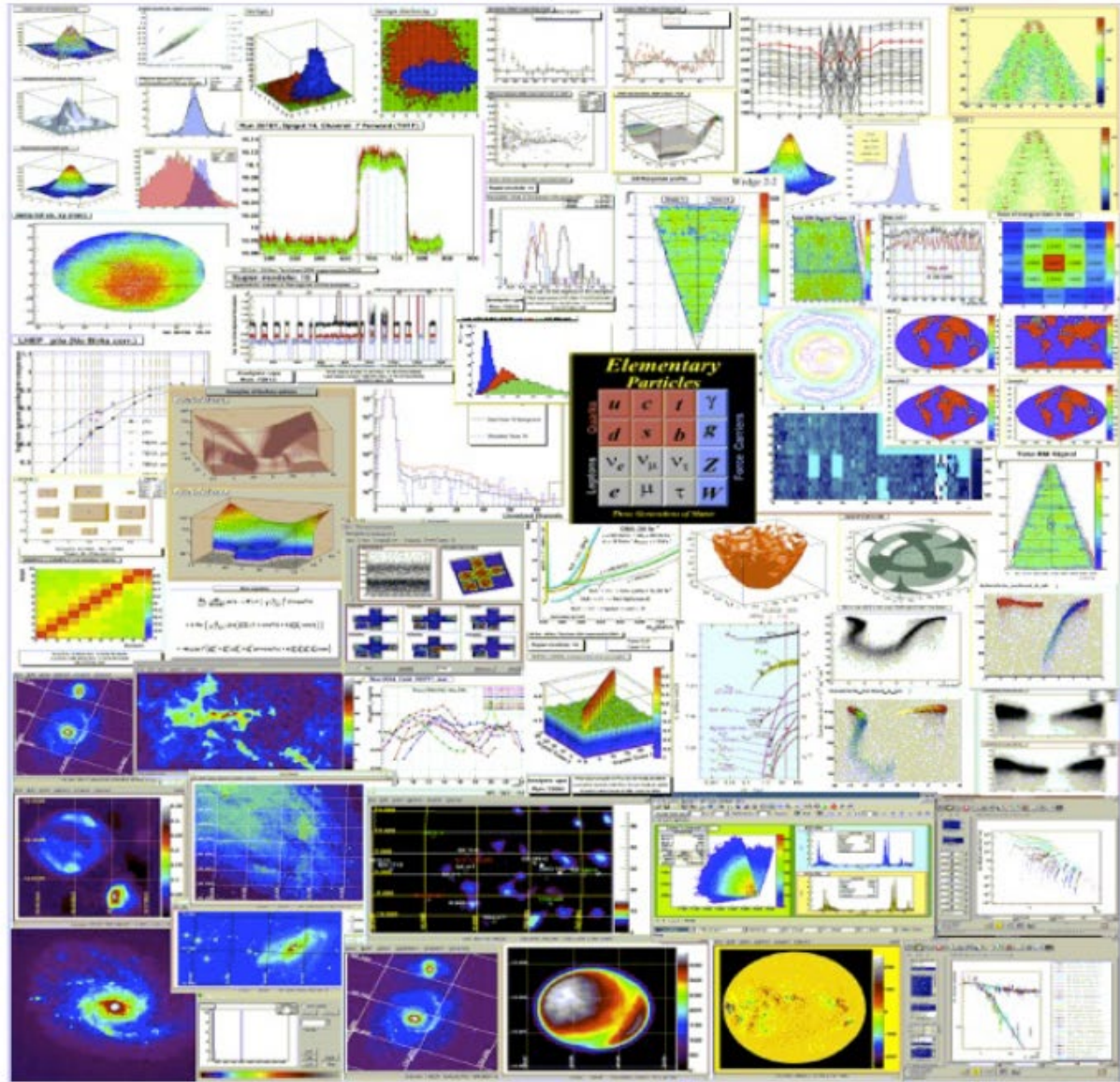
Graphics in ROOT
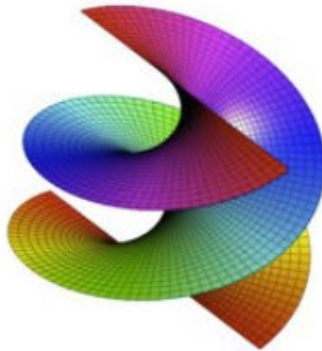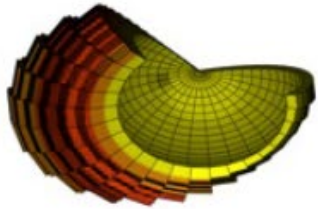
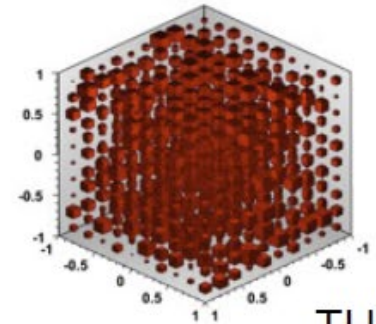- ## Many plots available for data analysis:
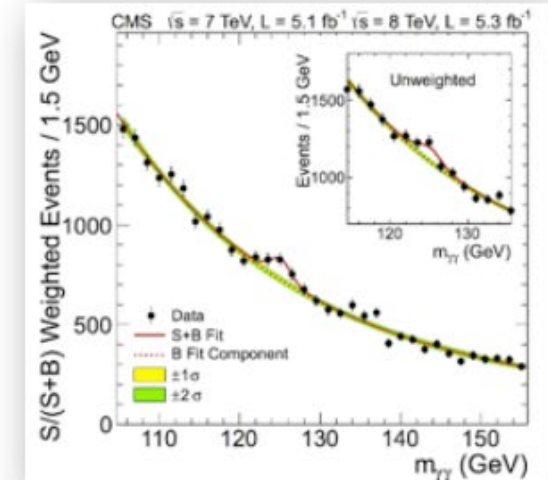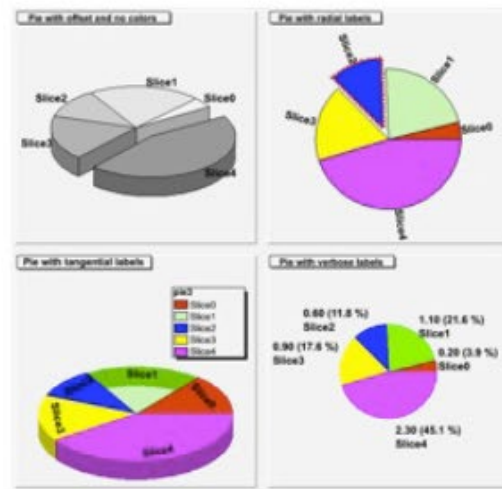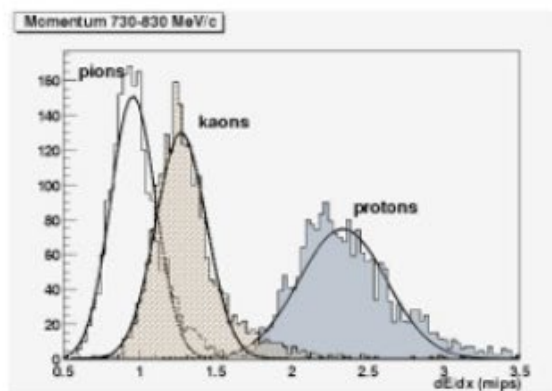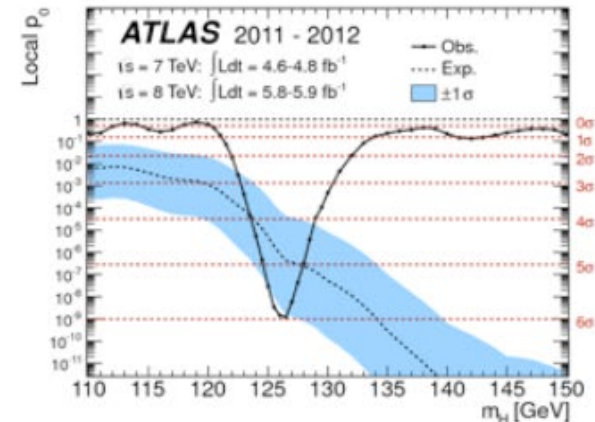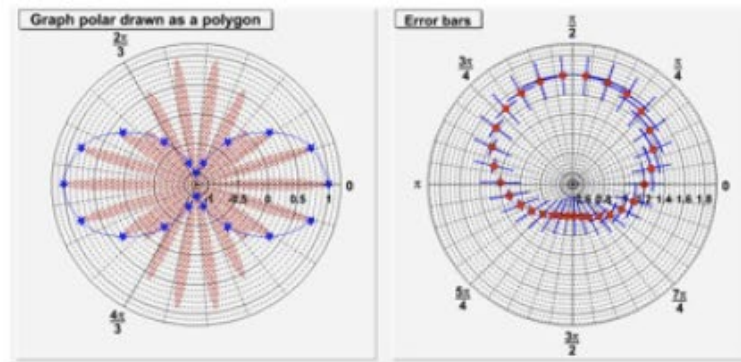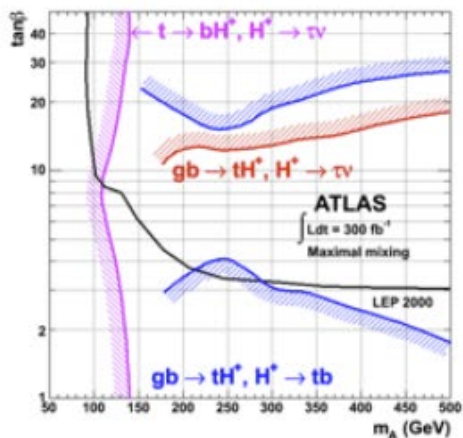
2D Graphics

- New functionality added at every new release

- Always new requests for new style of plots

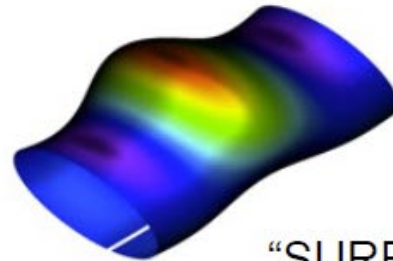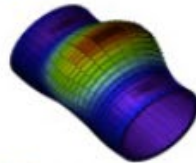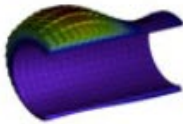- Can save graphics in many formats: `ps,pdf,svg,jpeg, png,c,root`

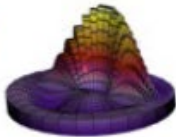# 3D Graphics

TGLParametric

TH3

"LEGO"

"SURF"

TF3

cftweb.cern.ch
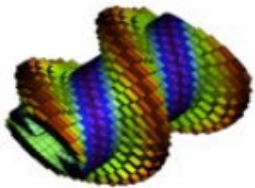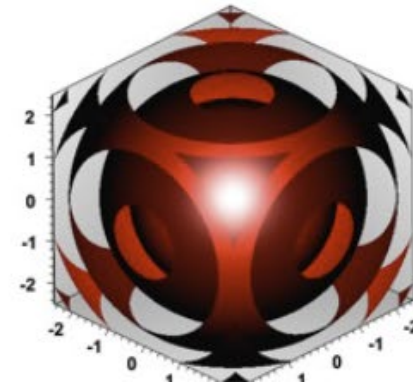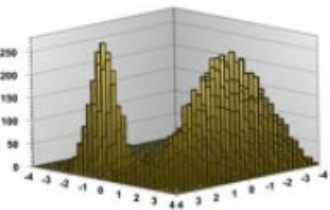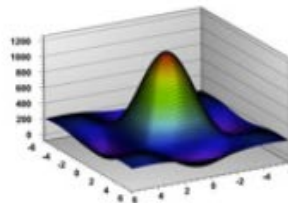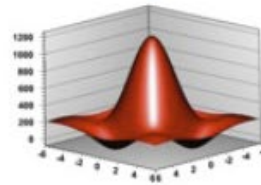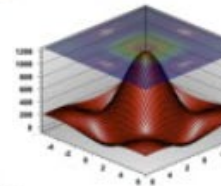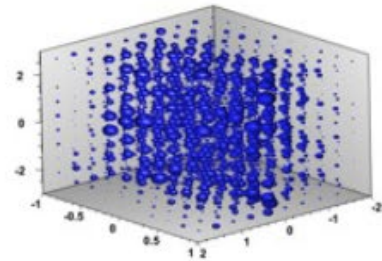
Other ROOT Features

- Geometry Toolkit
  - Represent geometries as complex as LHC detectors.
- Event Display (EVE)
  - Visualise particles collisions within detectors.
- PROOF: Parallel ROOT Facility
  - Multi-process approach to parallelism
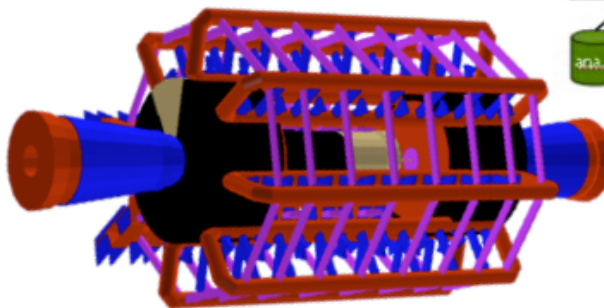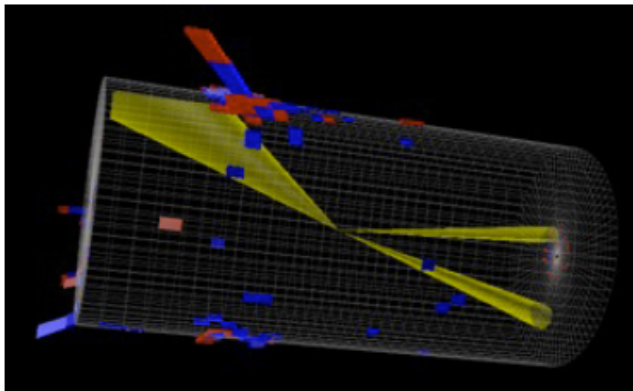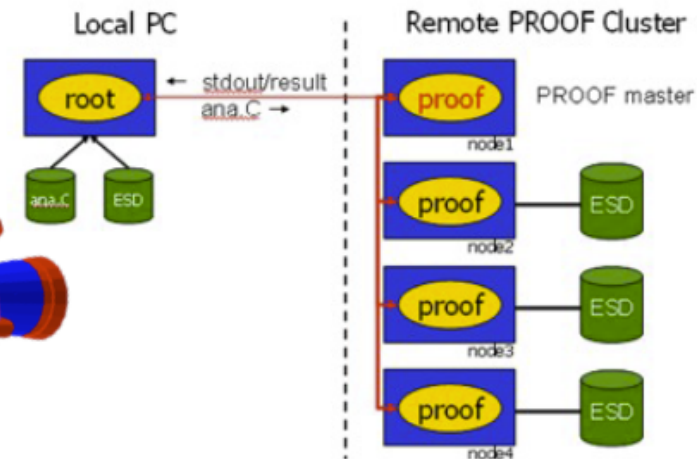 - A system to run ROOT queries in parallel on a large number of distributed computers
  - Proof-lite: does not need a farm, uses all the cores on a desktop machine

PROOF Schema

Getting Information

- **ROOT website**: http://root.cern.ch

  You can find ROOT user references and guides there.

- **the ROOT mailing list**: roottalk@root.cern.ch

- **subscribe to the mailing list:**

  http://root.cern.ch/root/Registration.phtml

- **forums and archives:**

  http://root.cern.ch/root/roottalk/AboutRootTalk.html

  http://root.cern.ch/phpBB2/

Summary

- CERN represents excellence in the field of Particle Physics research. To cope with the needs, it always pushed innovation in IT technologies.
- ROOT: complete analysis framework offering a C++ interpreter, a powerful persistency mechanism, advanced mathematical tools and even a parallel facility
  - it is designed for data analysis of very large shared data sets
  - it is the result of 18 years of cooperation between the development team and thousands of heterogeneous users
  - it is used extensively and profitably by physicists
    - all Higgs discovery plots made with ROOT

- We will see now how to start using ROOT !