ROOT

Data Analysis Framework

Andrew Savchenko

NRNU MEPhI, Moscow, Russia

24 August 2014

LVEE

Linux Vacation / Eastern Europe

Outline

1. Introduction
2. ROOT Features
3. Applications
4. Summary

Disclaimer: most images are taken from official ROOT sources

Introduction

HEP (high energy physics) aside from scientific research pushes leading edge technology as its byproduct.

Main demands for data processing:

- for petabytes of data:
  - effective storage
  - fast analysis
- extensible framework

And so ROOT (ROOT Object Oriented Toolkit) was born in 1995 for NA49 experiment. PAW is its ancestor, Rene Brun is founder of both projects.

Introduction

HEP (high energy physics) aside from scientific research pushes leading edge technology as its byproduct.

Main demands for data processing:

- for petabytes of data:
  - effective storage
  - fast analysis
- extensible framework

And so ROOT (ROOT Object Oriented Toolkit) was born in 1995 for NA49 experiment. PAW is its ancestor, Rene Brun is founder of both projects.

Introduction

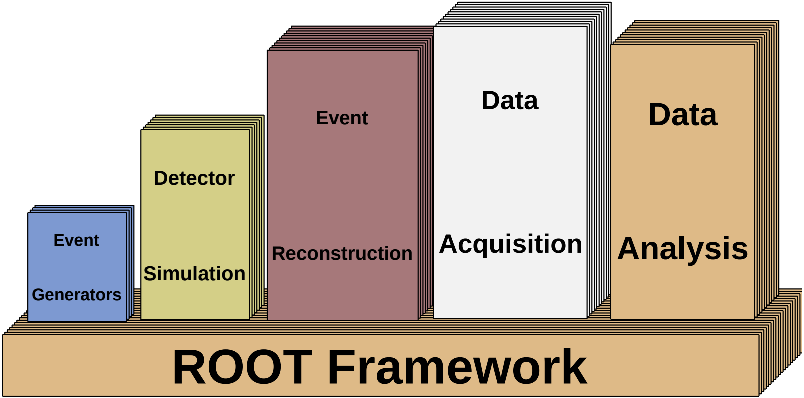
HEP (high energy physics) aside from scientific research pushes leading edge technology as its byproduct.

Main demands for data processing:

- for petabytes of data:
  - effective storage
  - fast analysis
- extensible framework

And so ROOT (ROOT Object Oriented Toolkit) was born in 1995 for NA49 experiment. PAW is its ancestor, Rene Brun is founder of both projects.

ROOT framework



C++11 framework

Typically each experiment creates own classes and applications based on ROOT framework.

Key features

- Cross-platform: Linux, MacOS, Windows
- All major compilers: gcc, clang, icc
- LGPL-2.1 (+ other free licenses for aux components)

- C++11 framework for building applications
- Analysis tools

- Regular compilation/linking
- C++ interpreter: Cling (LLVM/Clang based)
- Automatic interface to compiler (ACLiC)

Key features

- Cross-platform: Linux, MacOS, Windows
- All major compilers: gcc, clang, icc
- LGPL-2.1 (+ other free licenses for aux components)

- C++11 framework for building applications
- Analysis tools

- Regular compilation/linking
- C++ interpreter: Cling (LLVM/Clang based)
- Automatic interface to compiler (ACLiC)

Key features

- Cross-platform: Linux, MacOS, Windows
- All major compilers: gcc, clang, icc
- LGPL-2.1 (+ other free licenses for aux components)

- C++11 framework for building applications
- Analysis tools

- Regular compilation/linking
- C++ interpreter: Cling (LLVM/Clang based)
- Automatic interface to compiler (ACLiC)

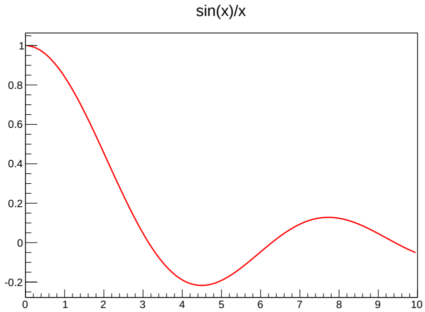
Advanced features

- ~ 2700 C++ classes
- Statistical analysis tools (fitting, minimizing)
- Multivariate analysis (MVA)
- Neural networks
- Visualisation tools (including OpenGL)
- Effective data queries in large data sets
- Client/server networking
- Parallel computing facilities (PROOF)
- Grid, AFS
- ...

Advanced features

- ~ 2700 C++ classes
- Statistical analysis tools (fitting, minimizing)
- Multivariate analysis (MVA)
- Neural networks
- Visualisation tools (including OpenGL)
- Effective data queries in large data sets
- Client/server networking
- Parallel computing facilities (PROOF)
- Grid, AFS
- ...

C++ Interpreter




```
$ root -l
root [0] TF1 f1(”func”,”sin(x)/x”,0,10);
root [1] f1.Draw();
Info in <TCanvas::MakeDefCanvas>:  created default TCanvas with
name c1
root [2] c1->Print(”sin.pdf”);
Info in <TCanvas::Print>: pdf file sin.pdf has been created
root [3] .q
```

C++ Interpreter

- <=ROOT-5.x CINT (C Ineterpreter)
- >=ROOT-6.x Cling
- ACLiC support for fast compiling and linking
- Can be separated from ROOT!

Cling features:

- based on Clang and LLVM
- JIT
- stricter C++11 support

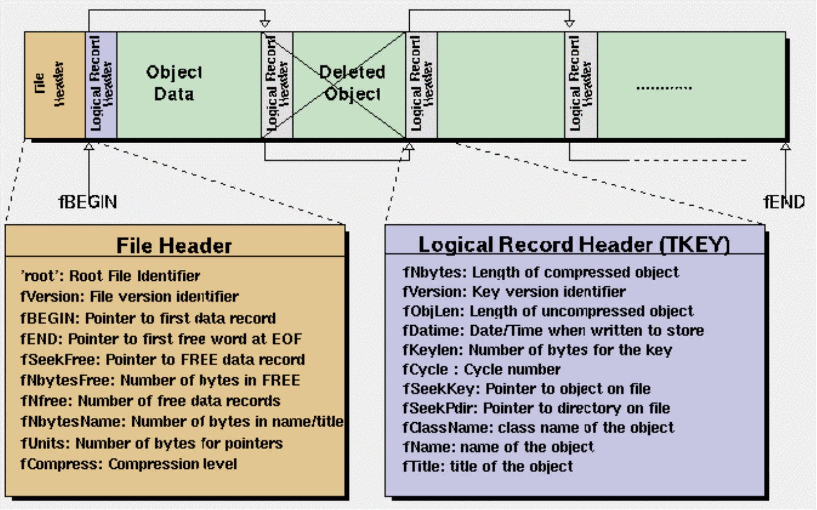
C++ Interpreter

- $<=$ROOT-5.x CINT (C Ineterpreter)
- $>=$ROOT-6.x Cling
- ACLiC support for fast compiling and linking
- Can be separated from ROOT!

Cling features:

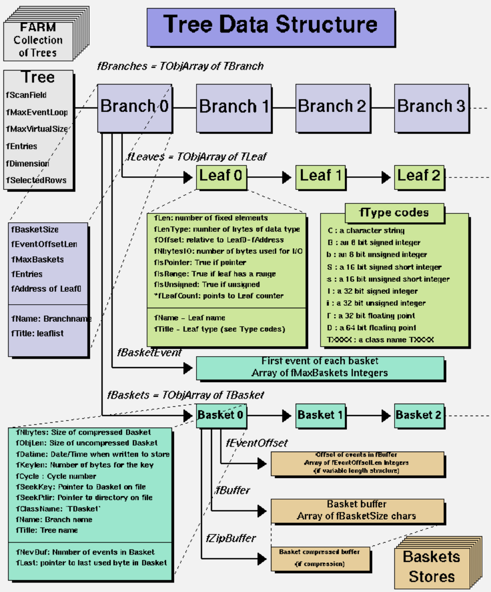
- based on Clang and LLVM
- JIT
- stricter C++11 support

ROOT File



File Header
Logical Record Header
Object Data
Logical Record Header
Deleted Object
Logical Record Header
Logical Record Header
....
fBEGIN
fEND
File Header
'root': Root File Identifier
fVersion: File version identifier
fBEGIN: Pointer to first data record
fEND: Pointer to first free word at EOF
fSeekFree: Pointer to FREE data record
fNbytesFree: Number of bytes in FREE
fNfree: Number of free data records
fNbytesName: Number of bytes in name/title
fUnits: Number of bytes for pointers
fCompress: Compression level
Logical Record Header (TKEY)
fNbytes: Length of compressed object
fVersion: Key version identifier
fObjLen: Length of uncompressed object
fDatime: Date/Time when written to store
fKeylen: Number of bytes for the key
fCycle : Cycle number
fSeekKey: Pointer to object on file
fSeekPdir: Pointer to directory on file
fClassName: class name of the object
fName: name of the object
fTitle: title of the object

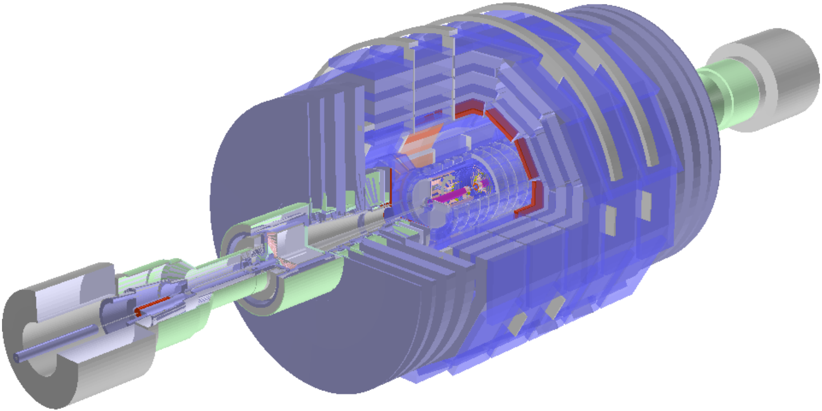
Tree structure



Math

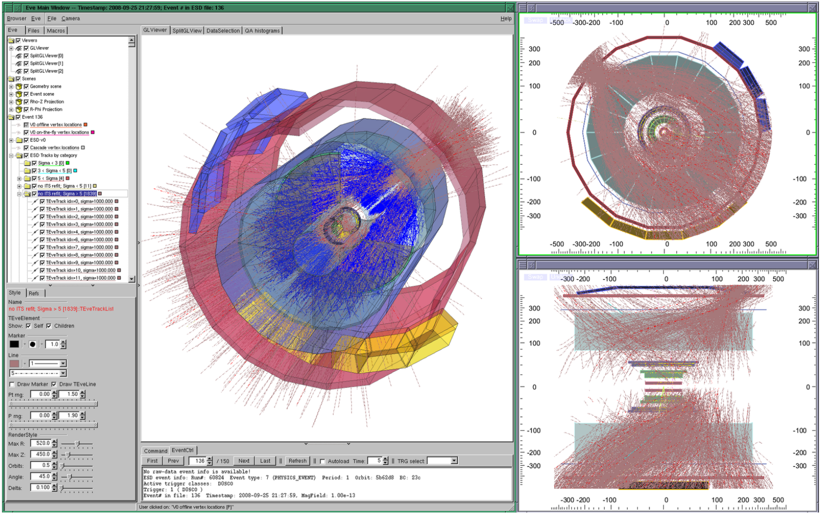
- 3-Vectors and transform
- 4-Vectors and Lorentz Transformations
- Matrix computations
- Numerical algos: derivation, integration, etc
- Minimization functions
- PDF-based analysis
- All standard C, C++ functions: GSL, STL, Boost, etc

Visualisation

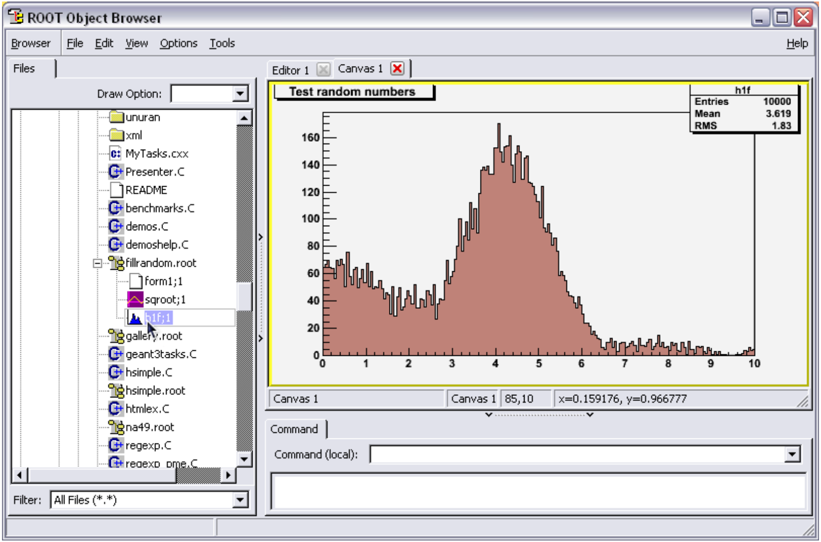


LVEE
Linux Vacation / Eastern Europe

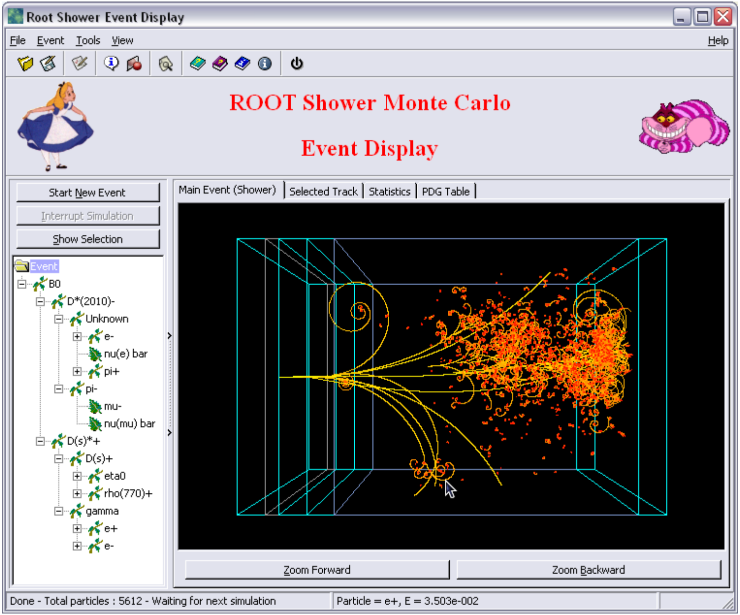
Event Display



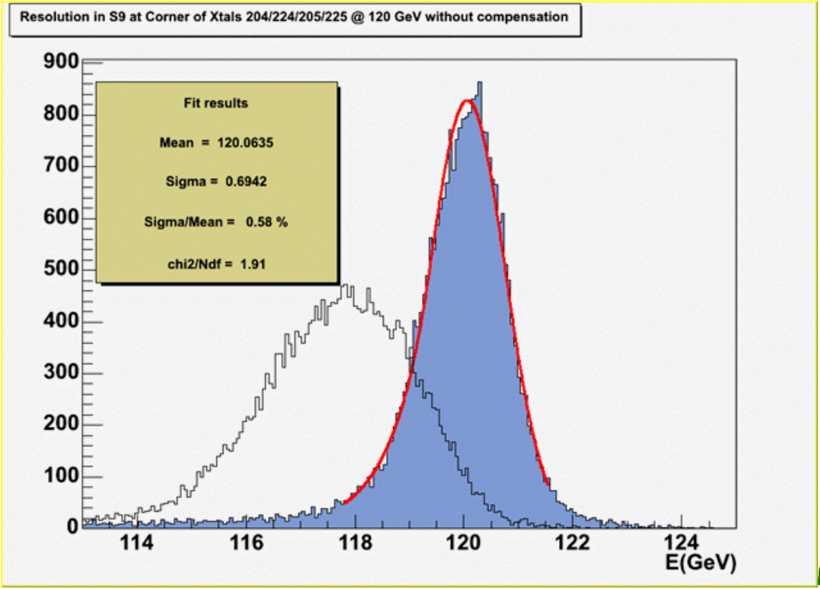
GUI


ROOT Object Browser
Browser File Edit View Options Tools Help
Files
Draw Option:
unuran
xml
MyTasks.cxx
Presenter.C
README
benchmarks.C
demos.C
demoshelp.C
fillrandom.root
form1;1
sqroot;1
h1f;1
gallery.root
geant3tasks.C
hsimple.C
hsimple.root
htmlex.C
na49.root
regexp.C
Filter: All Files (*.*)
Editor 1
Canvas 1
Test random numbers
h1f
Entries 10000
Mean 3.619
RMS 1.83
Canvas 1
Canvas 1 85,10
x=0.159176, y=0.966777
Command
Command (local):

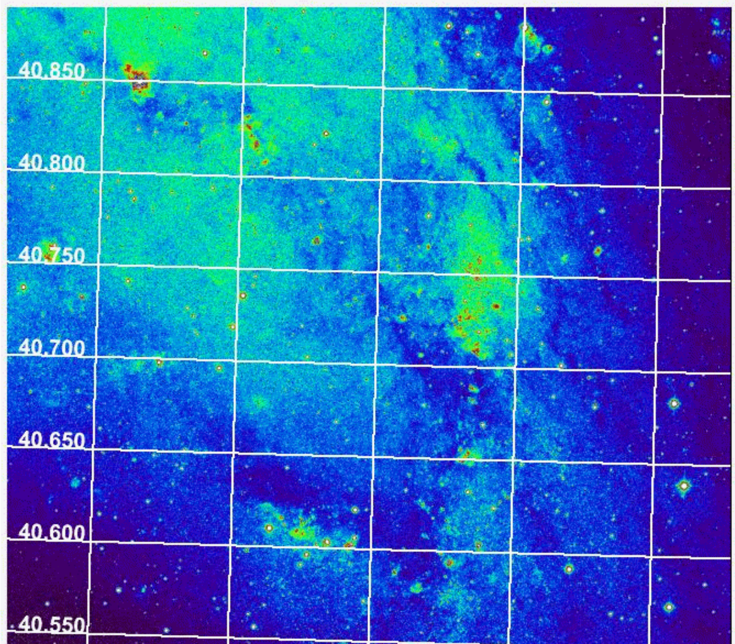
GUI


Root Shower Event Display
File Event Tools View Help
ROOT Shower Monte Carlo
Event Display
Start New Event
Interrupt Simulation
Show Selection
Event
B0
D*(2010)-
Unknown
e-
nu(e) bar
pi+
pi-
mu-
nu(mu) bar
D(s)*+
D(s)+
eta0
rho(770)+
gamma
e+
e-
Main Event (Shower)
Selected Track
Statistics
PDG Table
Zoom Forward
Zoom Backward
Done - Total particles : 5612 - Waiting for next simulation
Particle = e+, E = 3.503e-002

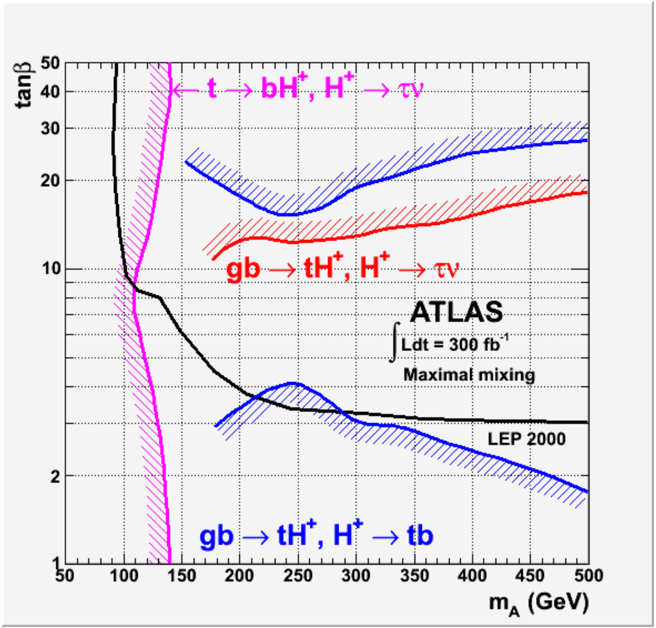
Data analysis



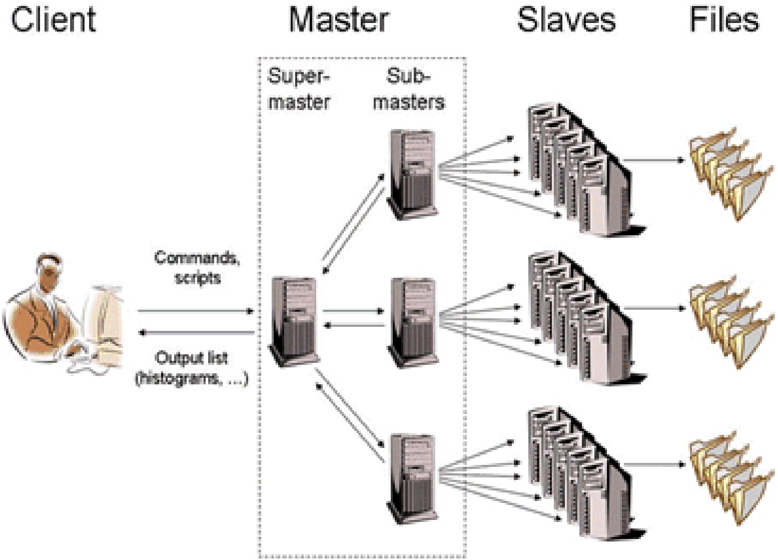
Data analysis



Data analysis



PROOF


Client
Master
Slaves
Files
Super-
master
Sub-
masters
Commands,
scripts
Output list
(histograms, ...)

Integration with other tools

Interpreters:

- C++ (Cling)
- Python
- Ruby (not in ROOT6 yet)

Simulation software:

- Geant-3, Geant-4
- Pythia-6, Pythia-8

- DB: MySQL, Postgres, Oracle, SQLite, ODBC
- CAD: OpenCascade
- OpenFOAM, R
- Grid, AFS, ...

Integration with other tools

Interpreters:

- C++ (Cling)
- Python
- Ruby (not in ROOT6 yet)

Simulation software:

- Geant-3, Geant-4
- Pythia-6, Pythia-8

- DB: MySQL, Postgres, Oracle, SQLite, ODBC
- CAD: OpenCascade
- OpenFOAM, R
- Grid, AFS, ...

Integration with other tools

Interpreters:

- C++ (Cling)
- Python
- Ruby (not in ROOT6 yet)

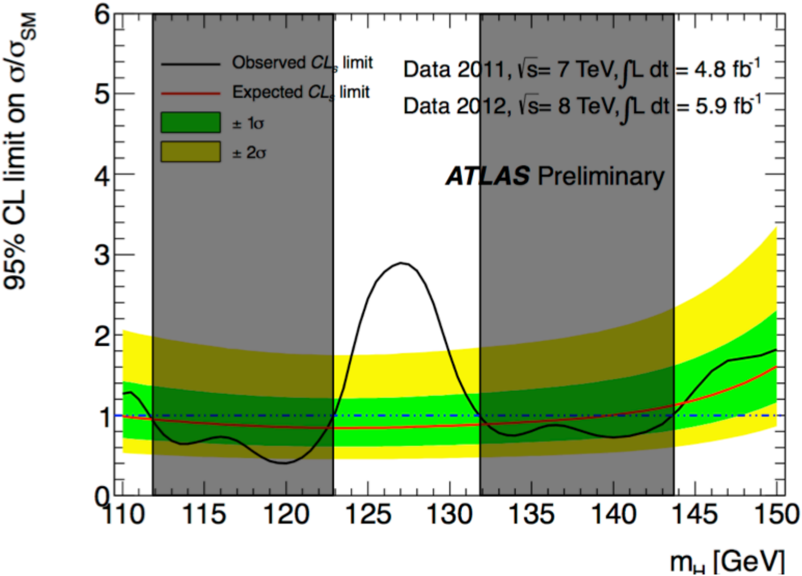
Simulation software:

- Geant-3, Geant-4
- Pythia-6, Pythia-8

- DB: MySQL, Postgres, Oracle, SQLite, ODBC
- CAD: OpenCascade
- OpenFOAM, R
- Grid, AFS, ...

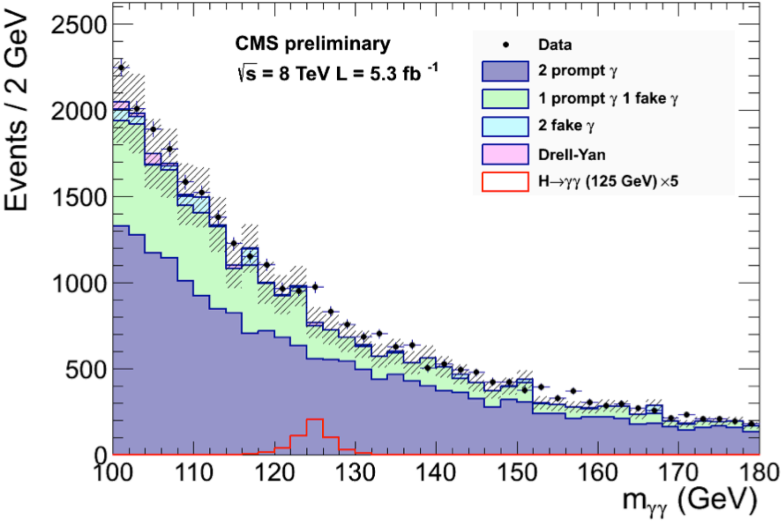
LVEE
Linux Vacation / Eastern Europe

Higgs discovery

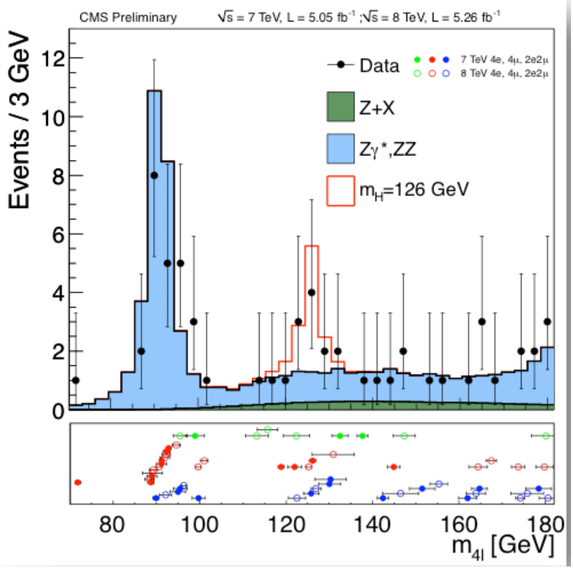


Higgs discovery

2012 8 TeV



Higgs discovery



Scope of applications

Outside of HEP and Nuclear Physics ROOT is used in:

- astronomy
- biology, bioengineering, bioinformatics
- computational neuroscience
- finance
- machine learning
- medicine
- natural language processing

Community

Visit root.cern.ch for details and downloads!

Mail list, forums and jira are available for discussion and reports.

Most user issues from novice till expert level are discussed.

Upstream is quite effective in patch review.

Thank you for attention!