The Matrix Element Method and GPGPUs for the Higgs boson studies

F. Beaudette, D. Chamont, O. Davignon, G. Grasseau,
L. Mastrolorenzo, P. Paganini, Thomas Strebler

Laboratoire Leprince-Ringuet
CNRS/IN2P3, École Polytechnique

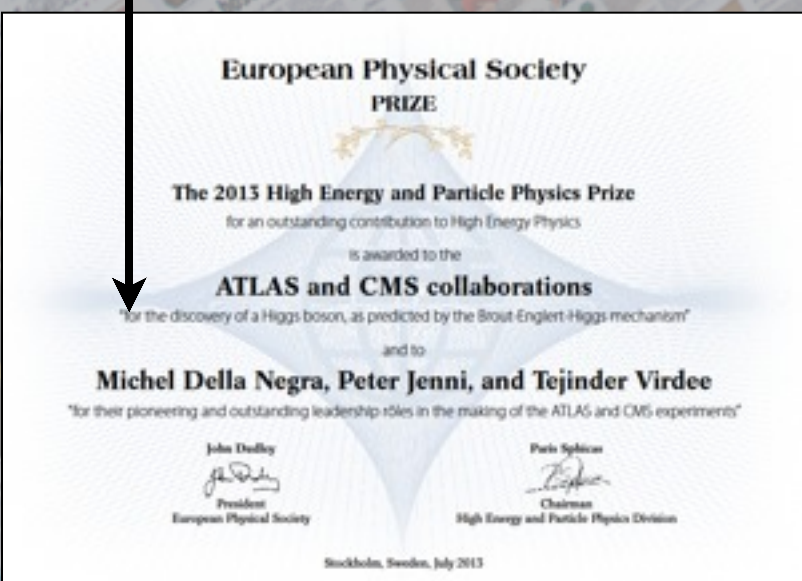
NVIDIA/CDS/UPSaclay meeting

The Higgs boson discovery

The discovery of the Higgs boson in 2012 had a worldwide coverage

Particle associated to the field that gives their mass to the fundamental particles through the electroweak symmetry breaking

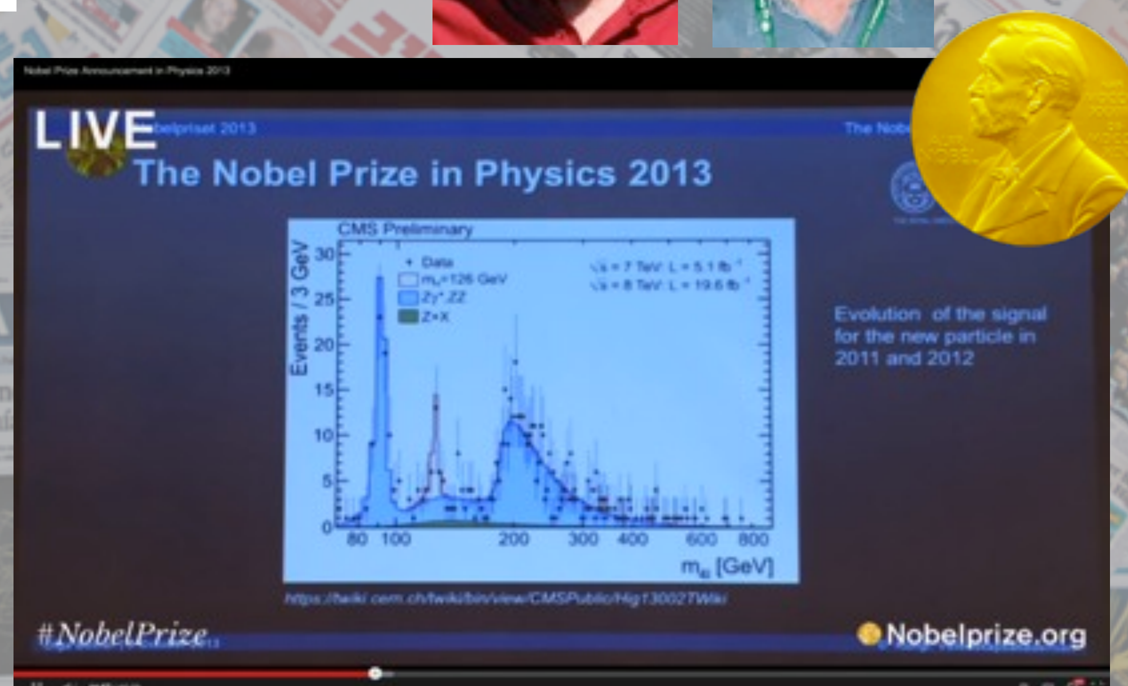
"for the discovery of a Higgs boson, as predicted by the Brout-Englert-Higgs" mechanism


European Physical Society
PRIZE

The 2013 High Energy and Particle Physics Prize
for an outstanding contribution to High Energy Physics
is awarded to the
ATLAS and CMS collaborations
"for the discovery of a Higgs boson, as predicted by the Brout-Englert-Higgs mechanism"
and to
Michel Della Negra, Peter Jenni, and Tejinder Virdee
"for their pioneering and outstanding leadership rôles in the making of the ATLAS and CMS experiments"

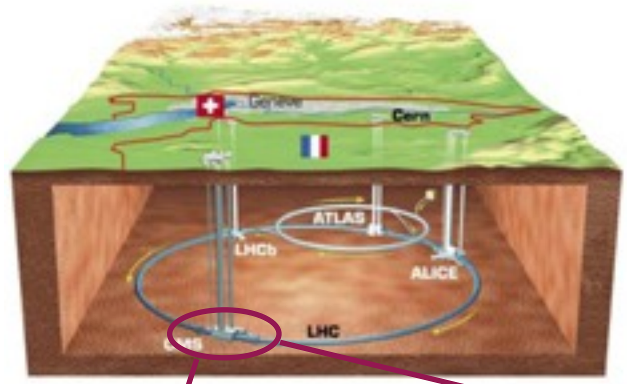
John Dudley, President, European Physical Society
Paris Sphicas, Chairman, High Energy and Particle Physics Division

Stockholm, Sweden, July 2013



The CMS experiment at CERN

Multi-purpose detector, designed for proton-proton collisions delivered by the Large Hadron Collider



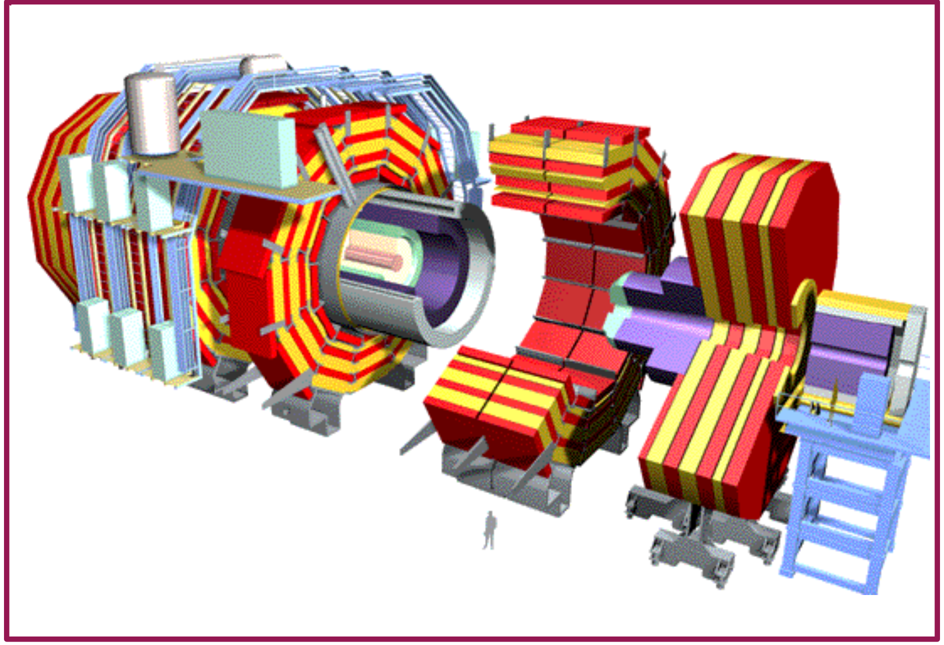
LHC

Run 1: 2010-2012
7 & 8 TeV

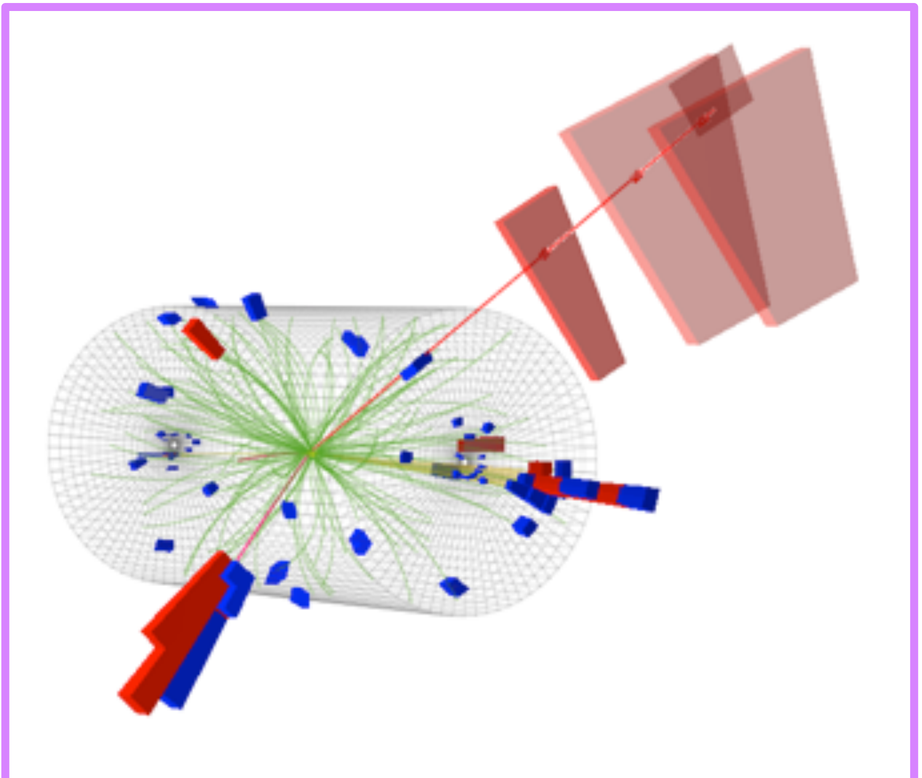
Run 2: 2015-2018
13 TeV

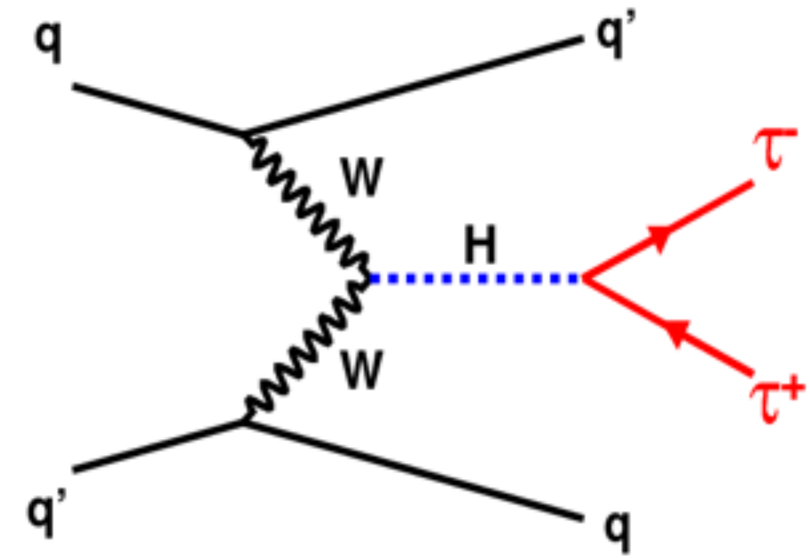
The Compact Muon Solenoid detector

- built around around a 3.8 T solenoid ∅ 6m - 13m long. Energy stored 2.6 GJ
- Largest Silicon detector in the world with 205m^2 of active area
- Total detector: 80M electronic channels
- Data taking rate > 100 Hz
- “Raw” event size : 1-2 Mb
- 20TB of data per day. This is *big data !*



Foreword

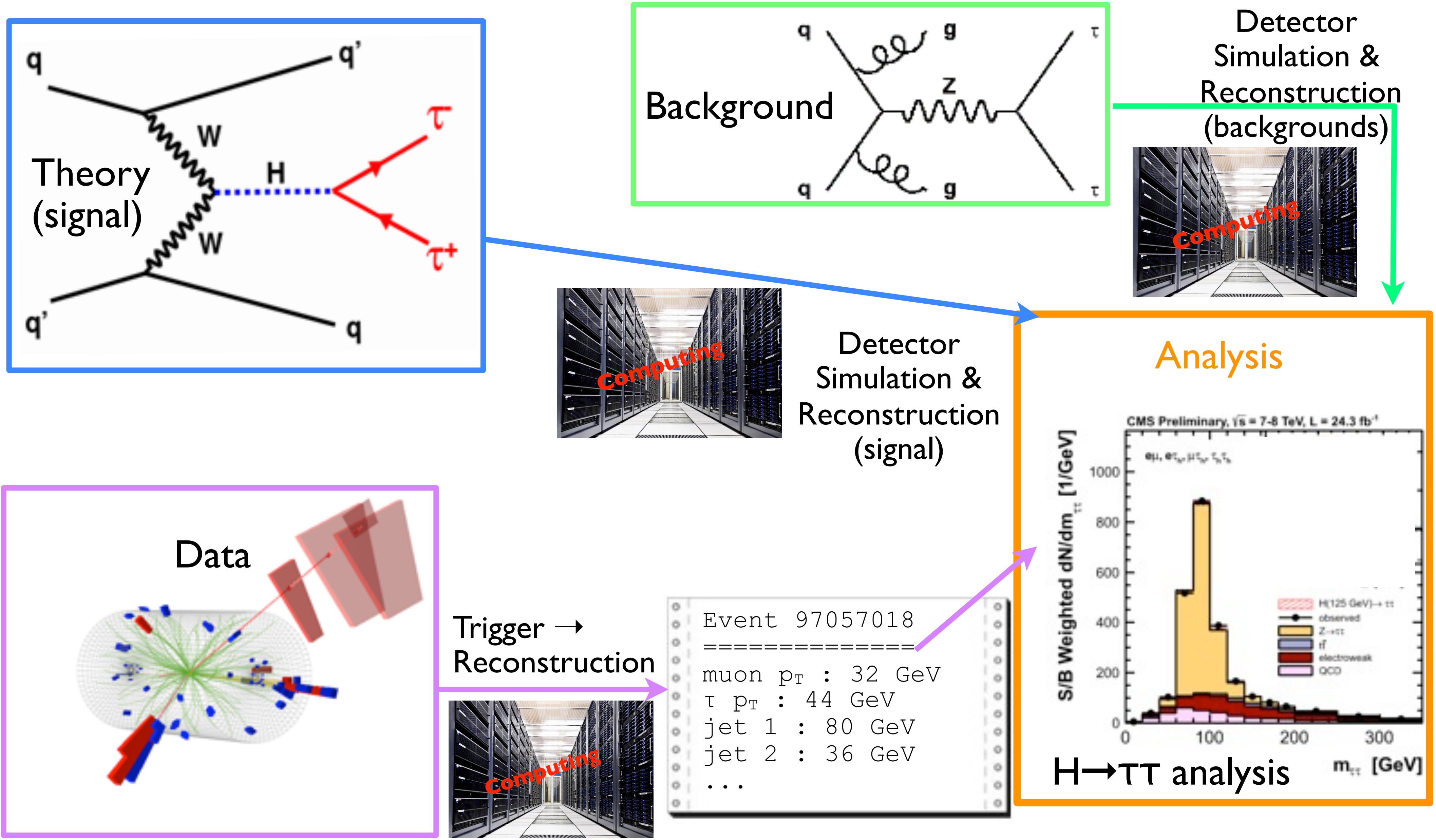




H→ττ with production in association with quarks (Vector Boson Fusion)

The goal is to compare :
the experiment
to
the theory

Classic approach



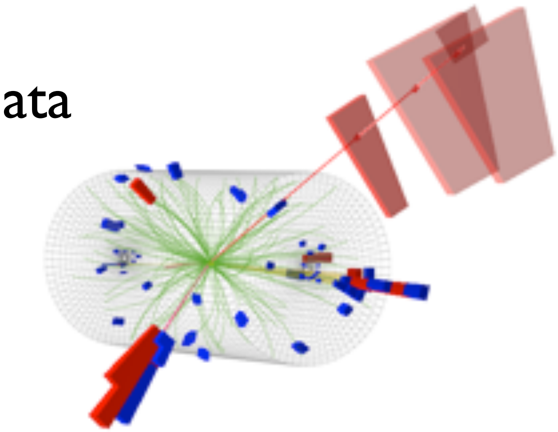
Matrix Element approach

The Matrix Elements allow a direct comparison between data and theory

It is an unsupervised method based on theory-based calculations of the probability

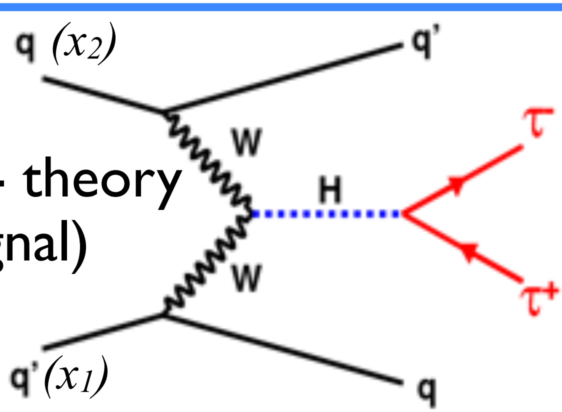
Trigger → Reconstruction

Data



```
Event 97057018
==============
muon pT : 32 GeV
τ pT : 44 GeV
jet 1 : 80 GeV
jet 2 : 36 GeV
...
```

Ω - theory (signal)



Matrix Element

simplified formula

$$P(\mathbf{x}|\Omega) = \frac{1}{\sigma_\Omega}\int\!\!\int\!\!\int dx_1 dx_2 d\mathbf{y}\, P_s(x_1, x_2)|M_\Omega(x_1, x_2, \mathbf{y})|^2 W(\mathbf{x}, \mathbf{y})$$

incoming ($dx_1 dx_2$), outgoing ($d\mathbf{y}$)

"Transfer function": response of the detector

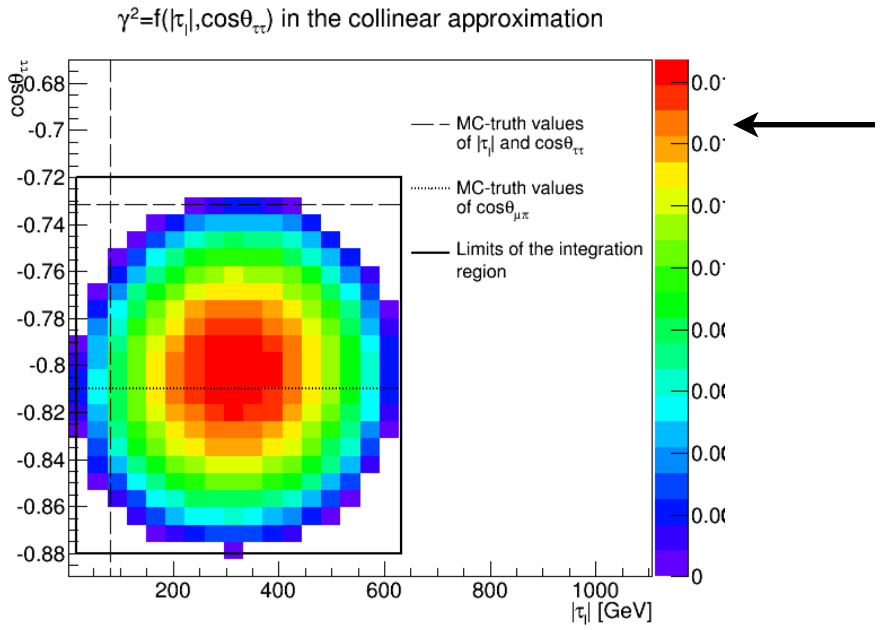
Novelty: the ME method has never been applied with τ before

- For each event a weight can be computed quantifying the probability that it arises from a given theory model
  - Use it as kinematic discriminant (to increase S/B)
  - Repeating the procedure for all events, and maximizing the result, find the theoretical model matching the observation

Computing aspects of the MEM in the H→ττ analysis (1/2)

- Multi-dimensional integrals (typically 5)
- To be competitive, it is of utmost importance to be able to run the entire analysis in a reasonable time-scale, of the order of two weeks (final detector calibrations usually arrive at the last minute)

- Fortunately enough, Matrix Element approaches are very well suited for parallel computing as the key element is the Monte-Carlo integration and different cores can work independently on different hyper volumes



2D-slice of the integration domain

Quite a lot of computations and iterations (by-hand) have been needed to write the integral so that the numerical integration is efficient.

Computing aspects of the MEM in the H→ττ analysis (2/2)

Tools involved and status of their parallelization:

$$P(\mathbf{x}|\Omega) = \frac{1}{\sigma_\Omega} \int\int\int dx_1 dx_2 d\mathbf{y} \, P_s(x_1, x_2) \, |\mathcal{M}_\Omega(x_1, x_2, \mathbf{y})|^2 \, W(\mathbf{x}, \mathbf{y})$$

well advanced — numerical integration (VEGAS)

~ 2000 lines of code

Tricky point: Random generators and reproducibility

done — parton density functions (LHAPDF)

~ 1600 lines of code

in progress — Matrix Element generated by Madgraph

~ 3000 lines of code +
~ 6000 lines (signal)
~10000 lines (background)

Can be reduced by a large factor doing some (reasonable) assumptions

done — Transfer functions home-made code, experiment specific

< 1000 lines
Some functions are semi-analytical.
Some of them are "look-up tables"

+ < 1000 lines (core software: event reading, kinematic transformations, etc....)

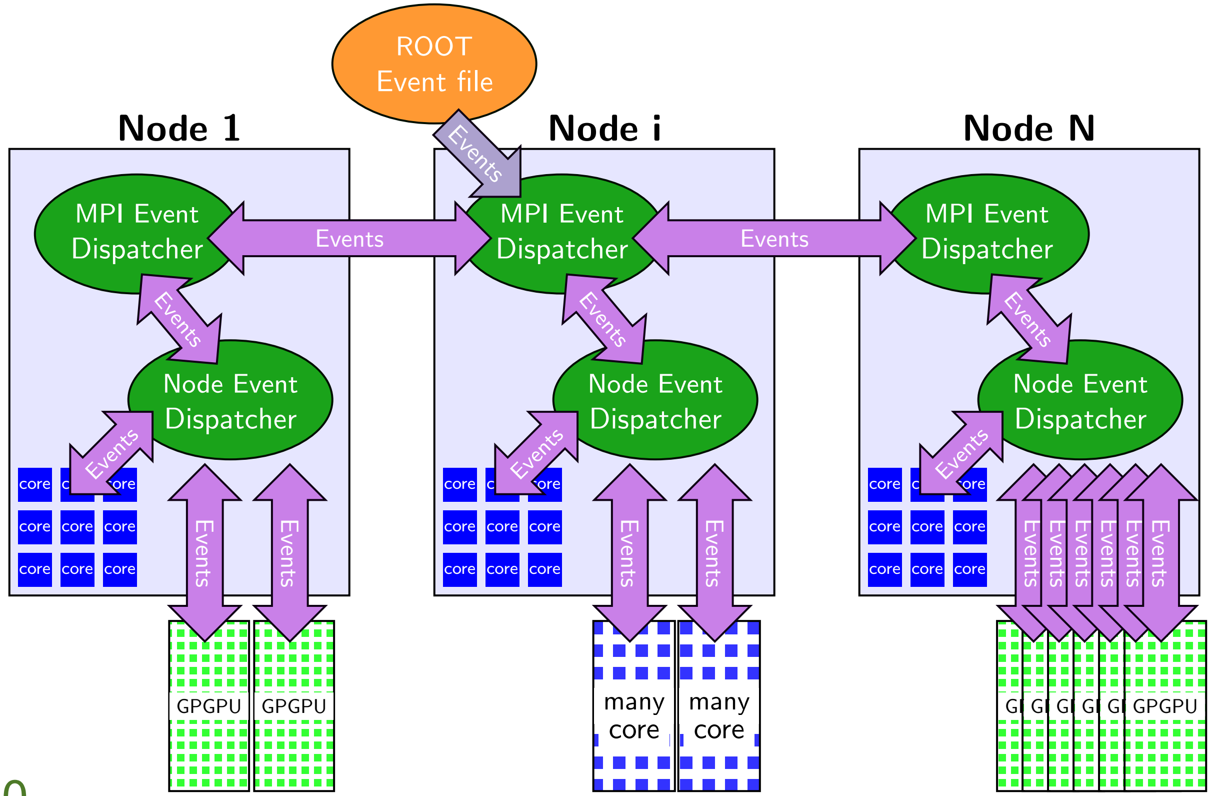
MEM integration with MC VEGAS

HPC MEM:
- Extensible deployment
- Hybrid application
  CPUs, Xeon Phi, NVidia
- All the available computing
  power has to be used

VEGAS Benchmarking:

$$\int_0^{2\pi} d^5\vec{x}\,\frac{\sin(\vec{x})}{\vec{x}}$$

- Speed up 300 factor reached
  wrt. single CPU (6X NVIDIA Titans)
- 5-6 factor reached with 2 NVIDIA K20
- Additional optimizations
  are envisioned



Talks (+proceedings) in conference:
GPU in HEP'14, Pisa, Sept. 2014
CHEP'15 talk: MEM (H→ττ channel) and HPC hybrid application (CPUs, GPUs, ...)

Conclusion

The team at LLR working on Matrix Element Method counts 7 MEMbers (2 research engineers & 5 physicists) and a large fraction of the work to put the method in place in the VBF $H\rightarrow\tau\tau$ channel has been done.
A significant effort to make the code able to run in parallel has been done since the very beginning of the project.

Indeed, the method is computing intensive and, in the competitive HEP environment we think it is crucial that our team is able to run the analysis in a 1-2 weeks time-scale.

As a result, parallel computing, in particular on NVIDIA-based platforms is in our opinion the way to go, it implies
- increasing our GPU-based resources connected with the grid
- sharing expertise