Deployment of a Matrix Element Method code for the *ttH* channel analysis on GPU's platform

G. Grasseau[1], F. Beaudette[1], A. Zabi[1], C. Martin Perez[1], A.Chiron[1], T. Strebler[2], G. Hautreux[3]

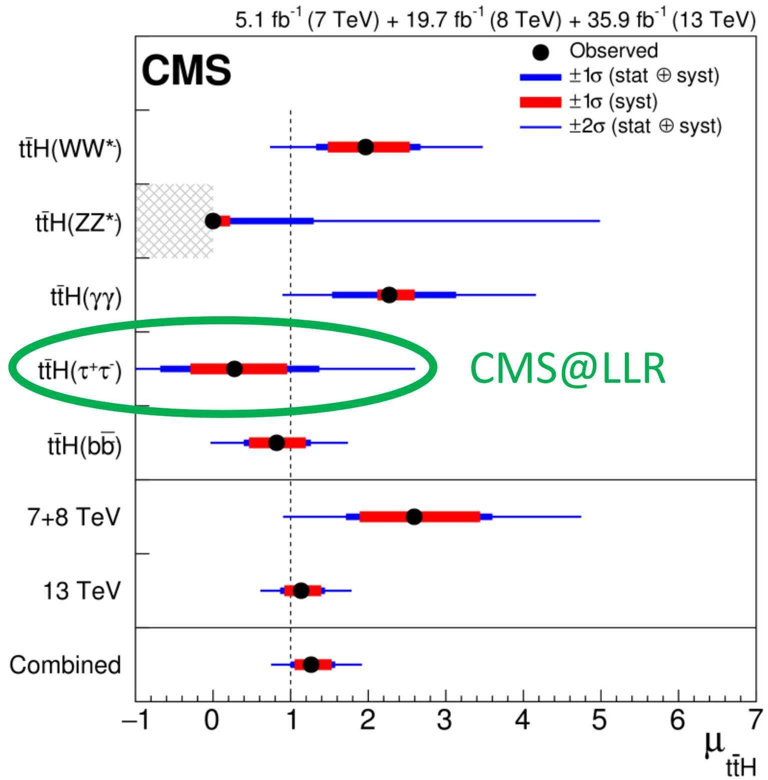
CHEP 2018 Conference, 9-13 July, Sofia, Bulgaria

[1] Leprince-Ringuet Laboratory (LLR), Ecole Polytechnique, Palaiseau

[2] Imperial College, London

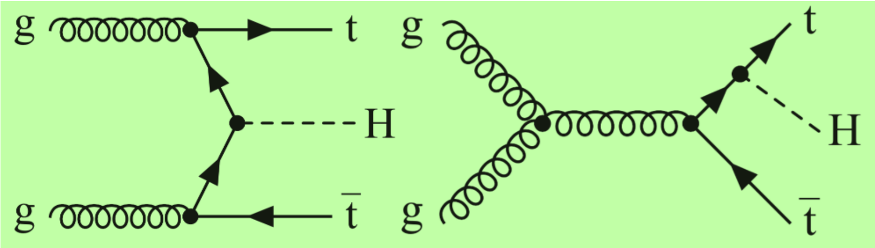
[3] GENCI, Grand Equipement National pour le Calcul Intensif, Paris

Recent discovery of H boson in *ttH* channel



- We (CMS@LLR) contributed to the $t\bar{t}H \rightarrow \tau\tau$ sub-channel

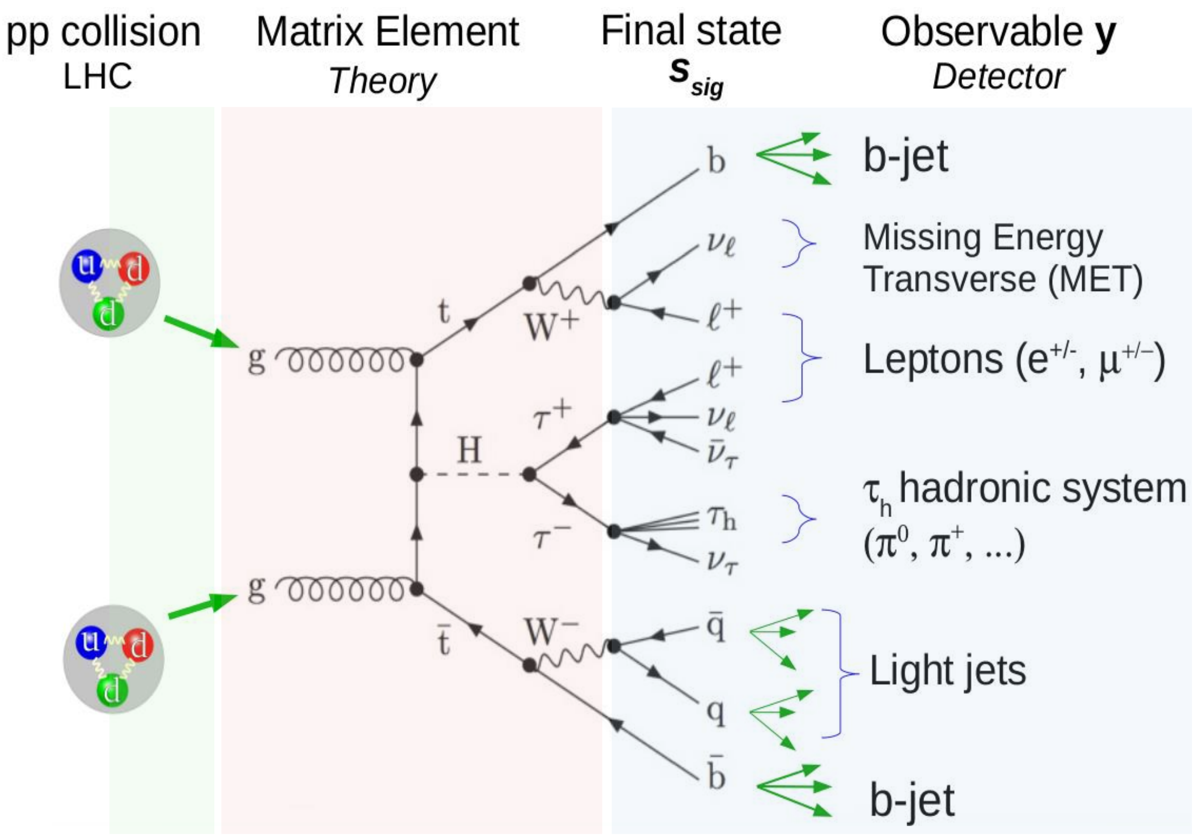
- Higgs decays into $\gamma\gamma, ZZ, WW$ and $\tau\tau$ final states have been observed (discovery 2012) and there is evidence for the direct decay to the bb final state
- In the Standard Model, the Higgs boson couples to fermions with a strength proportional to the fermion mass (Yukawa coupling)
- The decay to the $t\bar{t}$ final state is not kinematically possible
- Probing the coupling of the Higgs boson to the t quark, the *heaviest* known fermion, is a high priority
- The Higgs boson in association with $t\bar{t}$ final state can result from the fusion of a $t\bar{t}$ pair or through a radiation of t quark



- First observation* of the simultaneous production of a Higgs boson with a $t\bar{t}$ pair (channel) April 2018

*A. M. Sirunyan et al. (CMS Collaboration), “Observation of tt̄ H Production”, Phys. Rev. Lett. 120, 231801 (2018)

Matrix Element Method (MEM)



MEM is an unsupervised method (theory-driven) which is important to have among the supervised ones (Machine Learning, ...)

Principle:

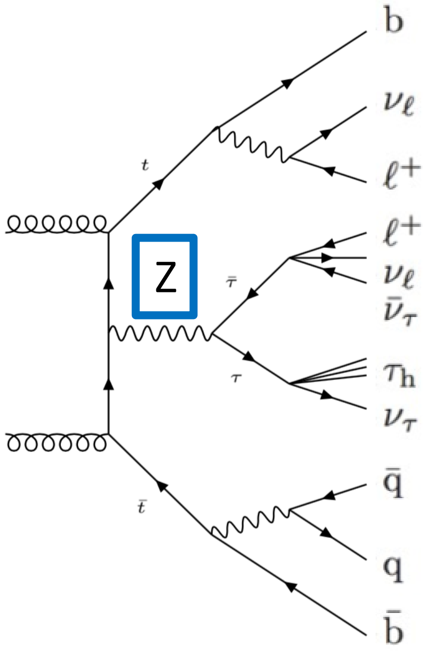
- select a Signal final state S_{sig} :
  $b\bar{b}, q\bar{q},\ \tau_{had}$, *2 leptons same sign*
- compute a weight quantifying the probability that an observed event matches a theoretical model
- vary the theoretical model (Signal, background(s))
- deduce a likelihood ratio

Weight of an event **y** — *p* processes — Parton Density Function (PDF) — Kinematics constrains — Matrix Element — Transfer Function *Response of the detector*

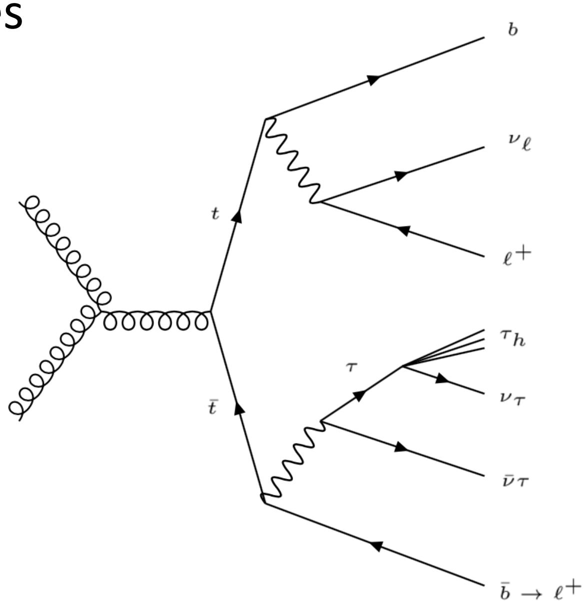
$$w_i(\mathbf{y}) = \frac{1}{\sigma_i} \sum_p \int d\mathbf{x} dx_a dx_b \frac{f(x_a, Q) f(x_b, Q)}{x_a x_b s} \delta^2(x_a P_a + x_b P_b - \sum p_k) |\mathcal{M}_i(\mathbf{x})|^2 W(\mathbf{y}||\mathbf{x})$$

MEM: time-consuming computations

- Multiple scenarios to consider (compute one integral for each) : the signal process and the background processes

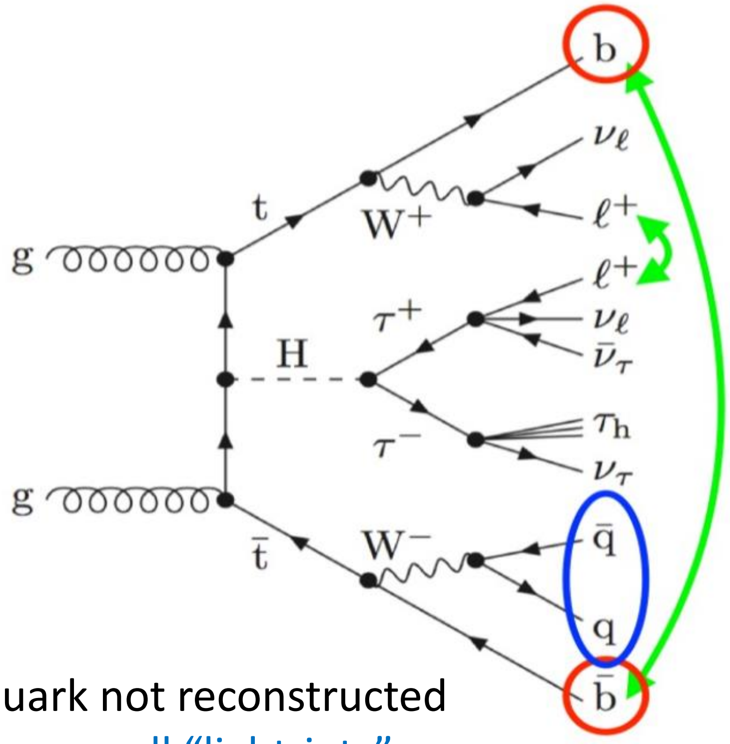


Irreducible background



One background: one non-prompt lepton produced in a b decay

- For each scenario : 4 permutations (green arrows)

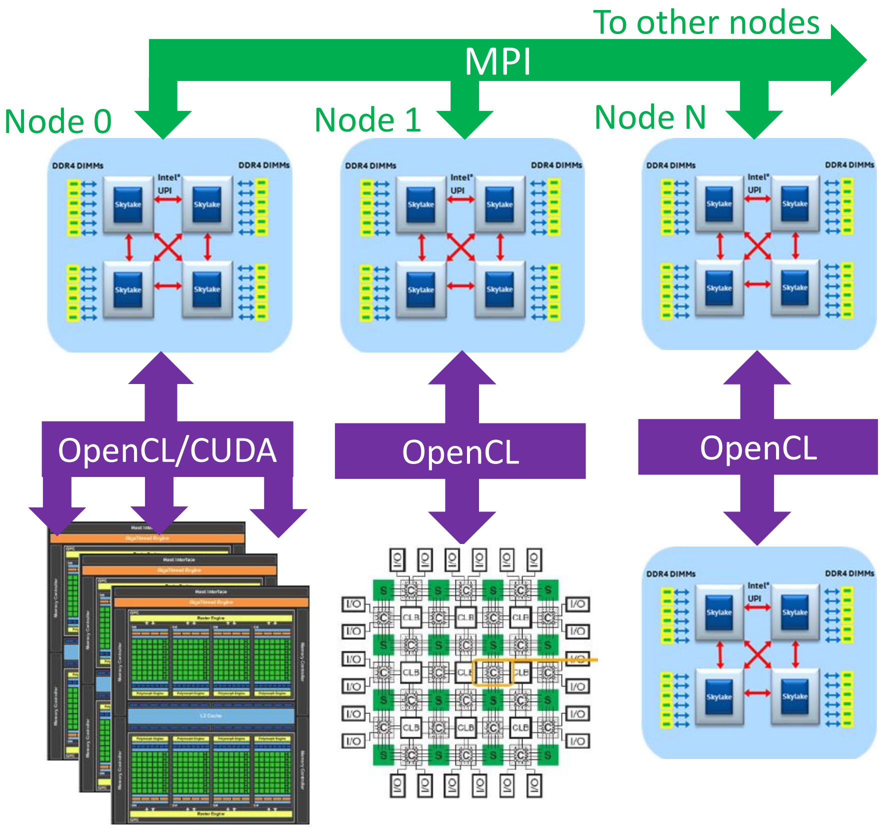


Only one quark not reconstructed (blue) → loop on all “light-jets”

*(1+3) * 4 [* #Ligth-jets]* Integrals with a dimension from 3 to 7.

They are computed if they are kinematically possible

The MEM Code



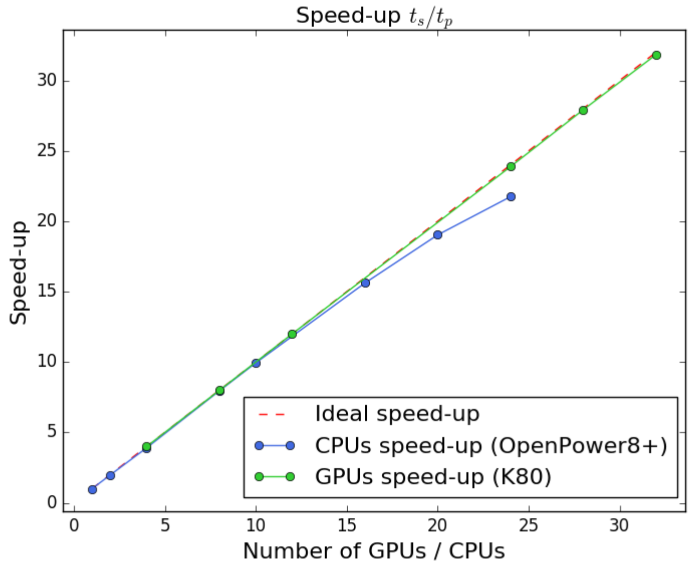
Multi-GPUs
on one node

Other hardware
OpenCL compliant

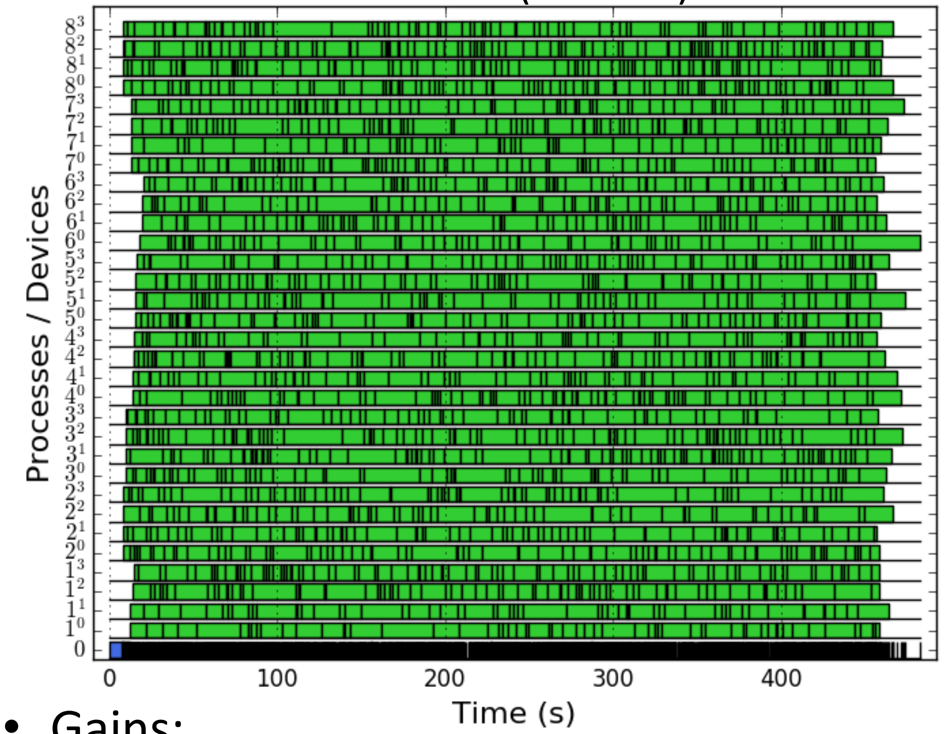
- The processing time for a typical data set (2395 evts) 55 days (14 hours / 96 cores)
- MEM code features: MPI/OpenCL/Cuda to aggregate numerous computing resources (HPC)
- Main kernel (one Vegas iteration)
  - developed a MadGraph extension to generate the OCL/Cuda kernel codes
  - LHAPDF lib.: Fortran to C-kernel translation
  - ROOT tools: Lorentz/geometric arithmetic's
  - →big kernels (10-20 x 10^3 lines)
- OpenCL / Cuda bridge (IBM+NVidia)

MEM code performance

- MPI C++ version versus MPI / OpenCL / CUDA - compilation -O 3, nvcc
- 1 node @CC-IN2P3:
  - Intel Xeon 2 x E5-2640, 2 x 8 cores@2.6 GHz
  - 2 NVidia K80 cards -> 4 Kepler GPUs per node
- Good scalability (MPI & kernels asynchronous mechanisms ok)



- Computing time of a data set with 2395 evts :
  - 55 days on 1 core (or 3. 5 days on a node)
  - 450 sec. on 32 GPUs (8 nodes)

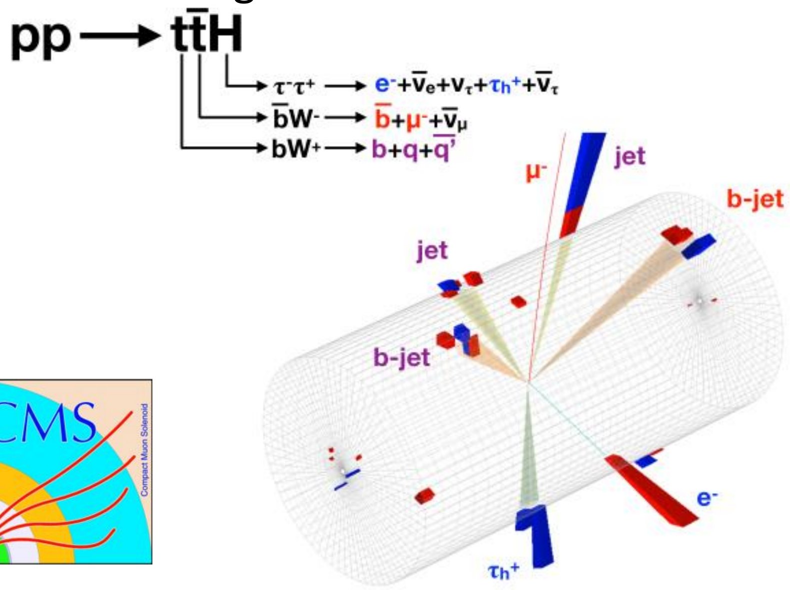


- Gains:
  - C++ → C kernel (careful) rewriting
  - CPU → GPU, the use of GPUs

Conclusion / perspective

- The MEM has proven to be an efficient method for signal extraction and our (CMS@LLR) results were combined to achieve the *ttH* production mode observation in 2018 *Phys. Rev. Lett. 120, 231801 (2018)*

2 lepton same sign and 1 tau channel



- Gain
  - Restitution time: several days against ~10 mn
  - Computing efficiency (cost, power supply, cooling, …)
    1 K80-GPUs is equivalent - for C++ MEM case - to ~20 nodes (2x8 cores)
  - In HL-LCG computing challenge, save the computing resources for other jobs.
- Physic program
  - For 2017 and 2018 data, new computations only with GPUs for *ttH(ττ)* analysis
- New developments
  - if we get the funding, project to have one code for CPU and GPUs, with the principles used by the MadGraph code generator
  - Optimizations: improve the computing load on GPUs

Acknowledgments

- Funding project P2IO
  Accelerated Computing for Physics
- Tiers 1 CC-IN2P3 benchmark platform
- Computing Center GENCI/IDRIS
- IN2P3 project: DECALOG/Reprises
- Google Summer of Code 2018
  HAhRD project : DL & HGCAL
- CHEP 2018 organizers