

Méthode des éléments de matrice : activités HPC au LPNHE et au LLR

Tristan Beau, Aurélien Demilly, Frédéric Derue,
Didier Lacour, Sylveste Pires

Laboratoire de Physique Nucléaire et de Hautes Energies, Paris



Florian Beaudette, David Chamont, Gilles Grasseau
Laboratoire Leprince Ringuet, Palaiseau



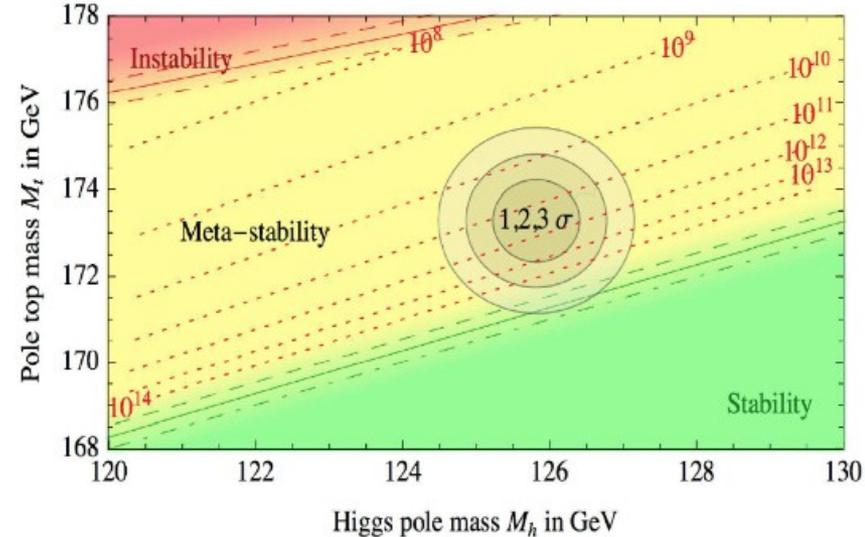
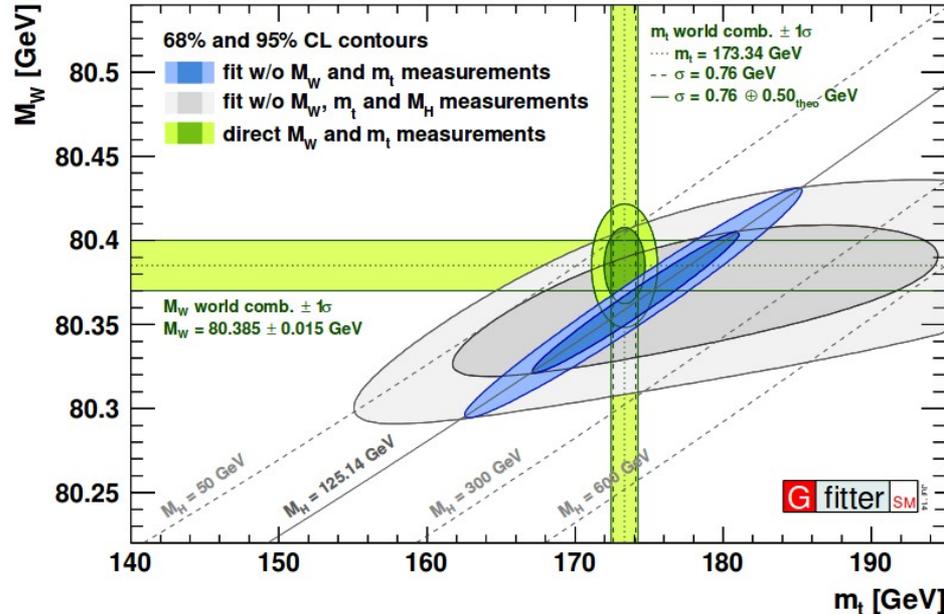
Rencontres LCG France,
IRFU Saclay, 1-2 december 2014

Mesure de la masse du quark top

test de cohérence interne
du modèle standard

Lié au potentiel de Higgs
et à la stabilité de l'Univers

arXiv:1407.3792



Du point de vue expérimental la masse peut être mesurée dans différents états finaux et avec de nombreuses techniques, le tout avec une bonne précision

Au LPNHE, nous travaillons sur une mesure avec des événements $t\bar{t}b\bar{b}$ avec deux leptons dans l'état final (i.e on ne peut pas reconstruire la masse directement car il y a des neutrinos/énergie manquante).

La méthode choisie est celle dite des éléments de matrice

La méthode des éléments de matrice (MEM)

For each event a probability P_{evt} is defined as a function of the probabilities to be signal like (P_{sig}) or background like (P_{bkg}) :

$$P_{evt}(x | m_{top}) = f_{top}P_{sig}(x | m_{top}) + (1 - f_{top})P_{bkg}(x)$$

where x : measured quantities and f_{top} : signal fraction

The probability to be signal like is related to the differential cross section $\frac{d\sigma}{dx}(m_{top})$ through :

$$P_{sig}(x | m_{top}) = \frac{1}{\sigma_{obs}(m_{top})} \frac{d\sigma}{dx}(m_{top})$$

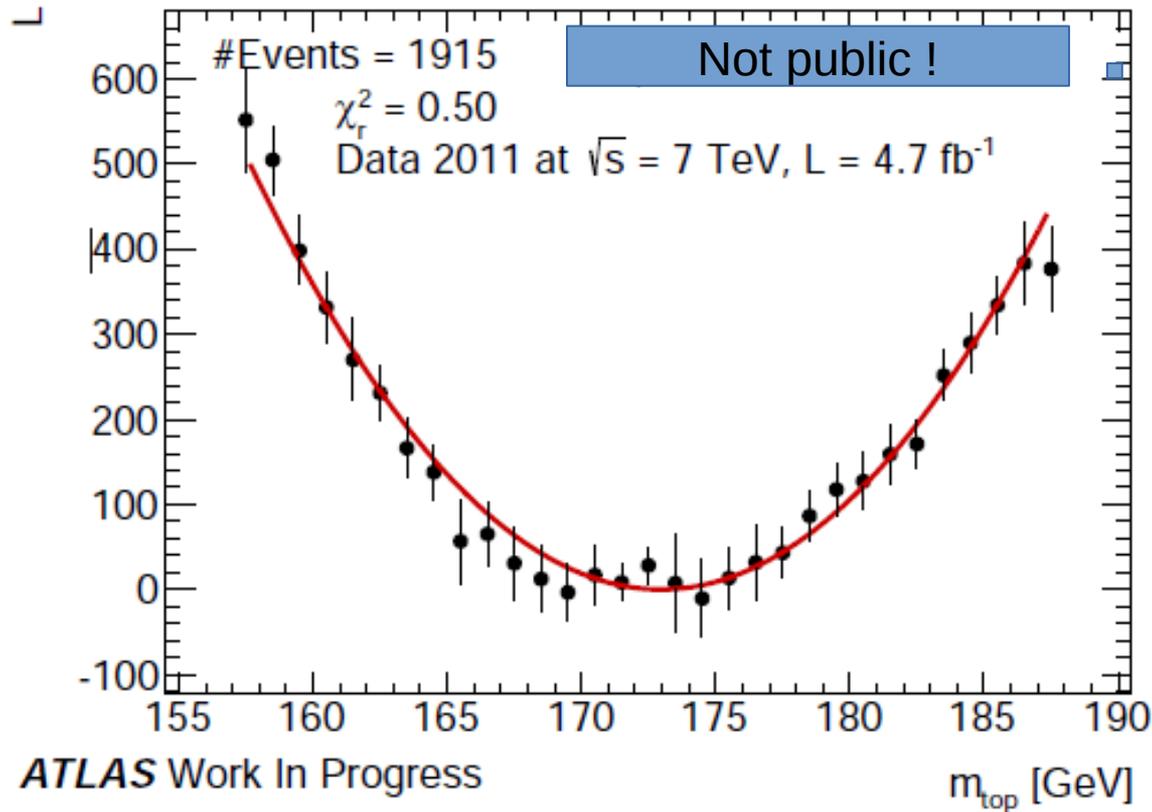
Computing the differential cross section is the heart of the method :

$$\frac{d\sigma}{dx}(m_{top} | k) \propto \int dg_1 dg_2 d\Phi |M_{\bar{t}\bar{t}}(g_1, g_2, x, m_{top})|^2 f_{pdf}(g_1) f_{pdf}(g_2) W(x | k)$$

- $|M_{\bar{t}\bar{t}}(g_1, g_2, x, m_{top})|$: matrix element for $g_1 g_2 \rightarrow \bar{t}\bar{t} \rightarrow b\bar{b}e\mu\nu_e\nu_\mu$
- g_1 and g_2 : incoming partons
- x : observables of the final state at the parton level
- $d\Phi$: phase space for the $pp \rightarrow \bar{t}\bar{t} \rightarrow b\bar{b}e\mu\nu_e\nu_\mu$ process
- f_{pdf} : parton density function
- k : observables of the final state at the reconstructed level
- $W(x | k)$: transfer function from k to x

Résultat obtenu à 7 TeV

Aurélien Demilly : thèse soutenue en Sept. 2014 avec les données à 7 TeV
Sylvestre Pires : thèse à défendre à l'été 2015 avec les données à 8 TeV



NB: résultat n'est pas publique (pas dans une ATLAS CONF note ou un papier) et seul l'étudiant a pu le montrer lors de sa soutenance. Mais le résultat est dispo ici: Aurélien Demilly, Mesure de la masse du quark top dans le canal en dileptons au LHC avec le détecteur ATLAS auprès du LHC, 19 septembre 2014, Beau T., Lacour D. (dir), CERN-THESIS-2014-140

Limitations de la méthode

Pour le résultat à 7 TeV (2011) utilisation du programme MadWeight basé sur le générateur Madgraph

- MadGraph 4 : seulement Leading Order only, cette version n'est plus maintenue, pas spécifique à la physique du top, pas facile à paralléliser
- intégrateur basé sur VEGAS

- 1 événement, pour une masse hypothèse de masse du top, avec ~ 10000 pas d'intégration = ~ 30 s
on a besoin de ~ 30 hypothèses de masse $\implies 15$ mn par événement
- sélection d'environ 2000 événements dileptons à 7 TeV
+ Monte Carlo (~ 20 fois la stat des données) + toutes les systématiques (>100), les hypothèses signal vs signal+bruit de fond etc

 \implies environ 10^7 heures de calcul
en pratique ~ 4 mois de calcul sur la grille

Pour la thèse de Sylvestre, avec les données à 8 TeV :

~ 10000 evts de données et $\sim 150k$ evts MC (ttbar)

\implies nous allons commencer la prod

Ressources HPC à l'UPMC

Création en 2012 de l'Institut de Calcul et de Simulation à l'UPMC.
Acquisition d'une machine HPC
Dans le cadre de l' Equipex
Equip@MESO
<http://www.ics.upmc.fr/>

La machine est localisée dans
l'une des salles serveurs
du LPNHE

HPC :

SGI UV2000

Performances : 19 Tflop/s

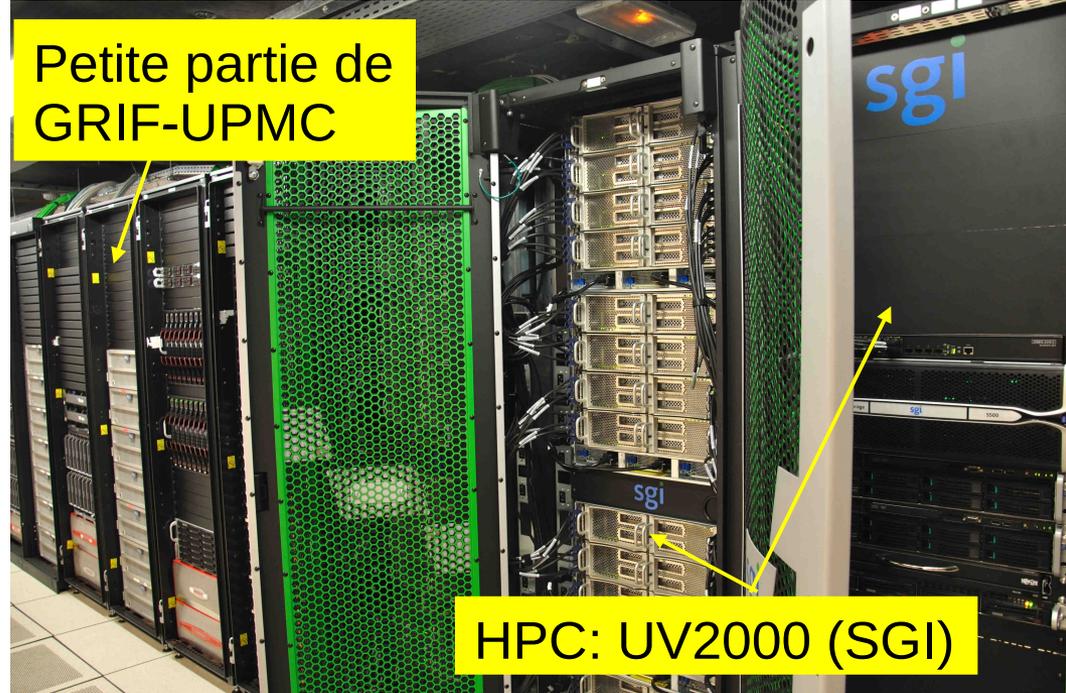
Processeurs : 1024 cpu, 64 nodes each with two Intel Xeon E5-4650L
(Sandy Bridge) 2.6 GHz, 8 cores

Mémoire : 16 TB DDR3 1600 MHz

Interconnection : Numalink 6 12.5 Gb/s

Stockage : 90 TB

Job scheduling : Altair PBSPro 12.1.0



Utilisation de nouveaux outils et parallélisation

Nouveaux outils sont disponibles

- MadGraph 5 (<https://cp3.irmp.ucl.ac.be/projects/madgraph/wiki/MadWeight>)
Next to Leading Order, nouvelles possibilités (ISR) +
bien plus rapide ~facteur 60 !
- MeMTk : développé dans ATLAS par le groupe de Berlin, NLO,
intégrateur basé sur CUBA/VEGAS

MeMTk : 2000 evts en 120 mn sur 1 coeur (gain ~100 wrt MadGraph4)
14 mn sur 8 coeurs
7 mn sur 16 coeurs
5 mn sur 32 coeurs
4 mn sur 128 coeurs (extra gain de ~30)

Le sample de données 2011 à 7 TeV en ~5mn au lieu de ~1 jour sur la grille

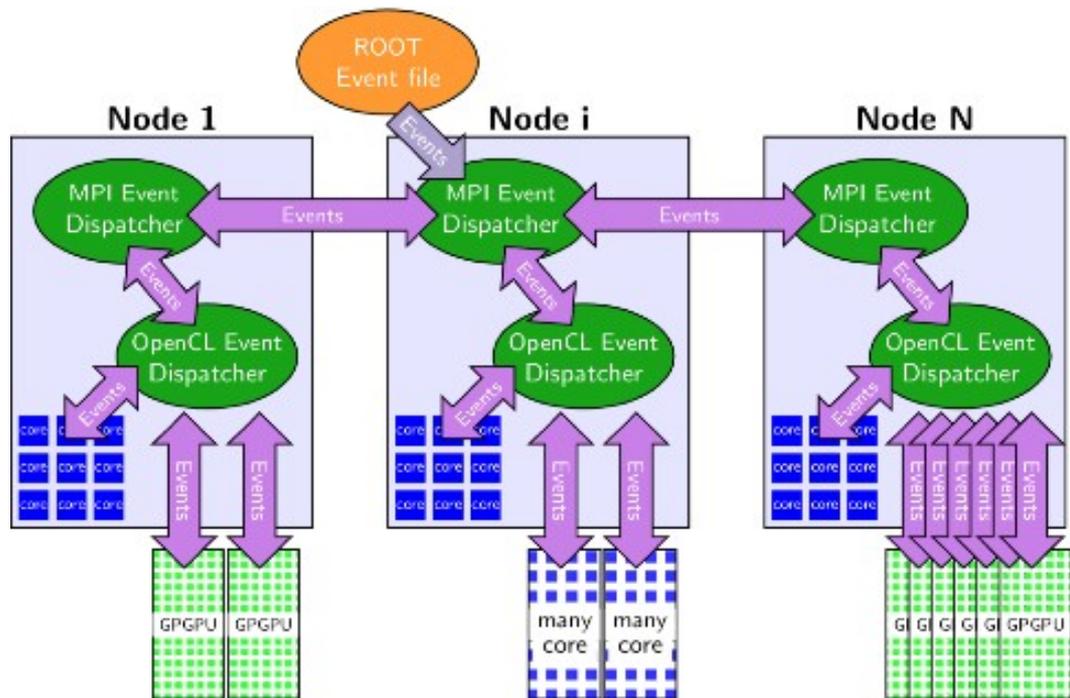
Mais nous avons encore beaucoup de travail de comparaison
et de validation de cet outil

Ce travail s'inscrit aussi comme une activité dans le cadre de LPaSo

- ❑ Candidature ANR « défi de tous les savoirs », préselectionnée 2014, relancée en 2015
- ❑ Objectif : Gagner en performance en exploitant les cœurs multiples, les instructions vectorielles, les accélérateurs, ..., afin de pouvoir absorber la montée en luminosité du LHC.
- ❑ Partenaires:
 - LAL (ATLAS, LHCb)
 - ▮ LLR (CMS)
 - ▮ LRI
 - ▮ LPNHE (ATLAS,LHCb)
- ❑ Thématiques
 - Parallélisme de tâches avec GaudiHive (ATLAS, LHCb).
 - ▮ Reconstruction (offline et trigger)
 - ▮ Traitement de données avec accélérateurs
 - ▮ Parallélisation/vectorisation d'outil d'analyse (Matrix Element Method)



MEM Integration with VEGAS



HPC MEM:

- Extensible deployment
- Thanks to MPI+OpenCL
- Hybrid application
CPUs, Xeon Phi, NVidia

VEGAS Benchmarking:

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

- Talk in conference:
GPU in HEP'14 (G. Grasseau),
Pisa, Sept. 2014
- Speed up 300 factor reached

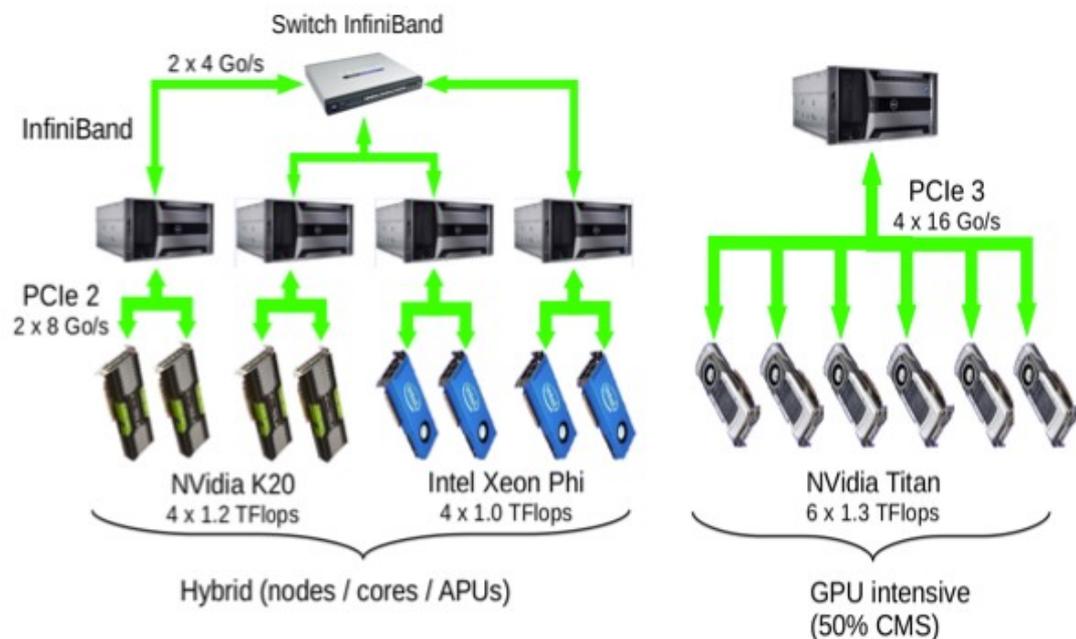
Towards MEM for $H \rightarrow \tau\tau$ channel:

- VEGAS integration ; very promising first results
- LHAPDF: Fortran \rightarrow C99 done
- ROOT Lorentz vector : C++ \rightarrow C99 done
- MadGraph plug-in: C++ \rightarrow C99 to be done

CHEP'15 talk: MEM ($H \rightarrow \tau\tau$ channel) and HPC hybrid application (CPUs, GPUs, ...)



GridCL platform (P2IO-CMS)



Development platform:

- New // programming paradigm for computing accelerators
- Based on a computer abstract model
- Very efficient: parallelize and vectorize

Laboratories:

- LLR, IAS, IRFU, LAL, IMNC, IPNO, CSNSM
- IN2P3 school planned for Spring 2016

“hybrid application: parallelization, vectorization”

Pros:

- High computing power, peak performance: GridCL ~ 68 nodes
- Best costs investment and operation:
cost (\$/GFlop) divided by ~4
power (W/GFlop) divided by ~3.5

Cons:

- Low abstraction : C-like (code refactoring)
- Find/change to vectorize algorithm
- Low memory capacity (6 Gb)

Conclusion

La méthode des éléments de matrice a été utilisée avec succès au Tevatron et à donné les mesures individuelles de mesure du quark top les plus précises. Mais c'est une méthode complexe à mettre en oeuvre, en particulier du fait de sa lenteur

De nouveaux outils sont disponibles et semblent prometteurs (MemTk, MadGraph5) : ils sont d'eux memes plus rapides (~ facteur 100) et sont parallélisables.

Ces activités se sont développées de manière indépendante au LPNHE et au LLR
mais s'inscrivent "naturellement" dans le cadre de l'un des WorkPackage (MEM) de la demande ANR LPaSo