Matrix Element Method

an application on $t\bar{t}H$ and $t\bar{t}Z$ analysis

Nicolas Chanon, Institut de Physique Nucléaire de Lyon, CNRS/IN2P3

Jing Li, Qiang Li, *Peking University, Beijing, China*

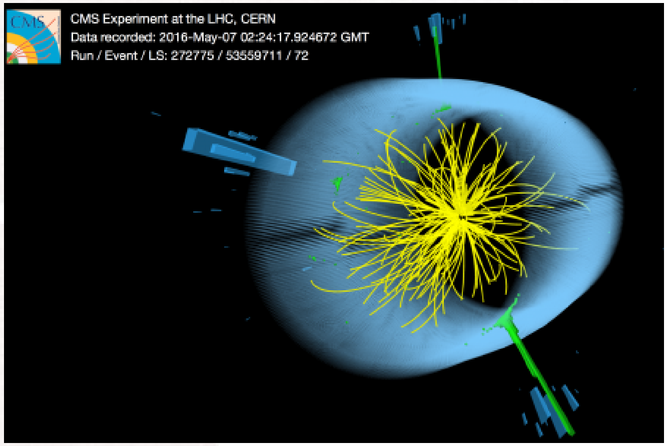
On behalf of the CMS Collaboration

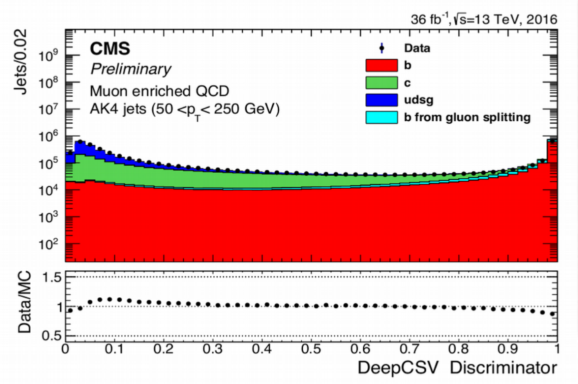
22nd December 2017

The Third China LHC Physics Workshop
22-24 December 2017 Nanjing University

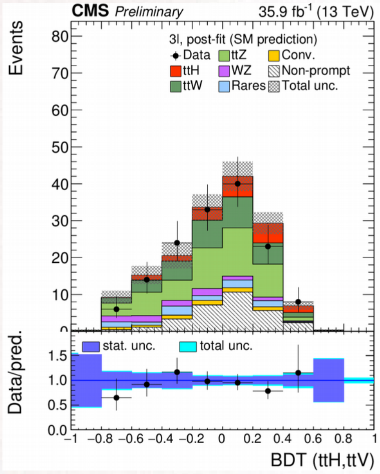
Motivation

- One of the biggest challenges in Run II LHC: extract rare and novel signal events from a large number of background events





Deep Neural Network for b tagging



BDT in CMS-SMP-17-004

Model-dependent approach

- theoretical assumption as a starting point
- probability distribution as discriminator among different theory model (hypothesis)
- exploit and provide more information
- MultiVariate Analysis (MVA) techniques, e.g. neural networks or boosted decision trees
- Matrix Element Method (MEM)

Introduction

- Matrix Element Method (MEM) is a powerful experimental technique widely employed to maximize the amount of information that can be extracted from a collider dataset
  - the matrix element contains the maximal amount of theoretical information available
  - use the measured particles' momenta as direct input to the evaluation of the matrix element
- For each event, the MEM compute a weight quantifying the probability that it arises from a given theory model
- Then the most probable hypothesis can be obtained through a likelihood maximization method
- Especially useful when the expected experimental signatures involve a complex final state, or has a topology not fully reconstructed at detector level

Description of the method

- The goal of the MEM is to perform a measurement using the matrix element to create a probability function
- MEM weight: conditional probability P(**x**|**α**)
  - Experimentally quantities: **x**
  - Theoretical information: **α**
  - Parton-level configuration: **y**
  - The evolution of **y** into **x** (transfer function): W(**x**,**y**)

$$P(\boldsymbol{x}|\boldsymbol{\alpha}) = \int d\boldsymbol{y} P_\alpha(\boldsymbol{y}) W(\boldsymbol{x}, \boldsymbol{y}).$$

Description of the method

- The parton-level probability $P_\alpha(\mathbf{y})$ can be expressed as a product of
  - the squared matrix element $|M_\alpha|^2(\mathbf{y})$,
  - the parton distribution functions $f_1(q_1)$ and $f_2(q_2)$
  - the phase-space measure $d\Phi(\mathbf{y})$

$$P(\boldsymbol{x}|\boldsymbol{\alpha}) = \frac{1}{\sigma_\alpha} \int d\Phi(\boldsymbol{y}) dq_1 dq_2 f_1(q_1) f_2(q_2) |M_\alpha|^2(\boldsymbol{y}) W(\boldsymbol{x}, \boldsymbol{y}) \, .$$

- The normalization by the total cross section σ_α ensures that $P(\mathbf{x}|\boldsymbol{\alpha})$ is a probability density

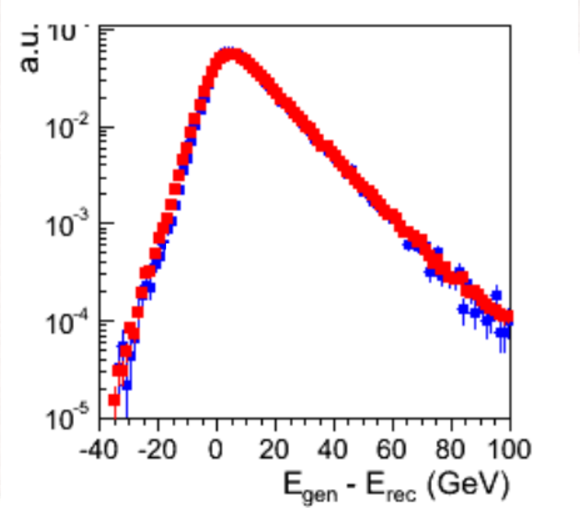
$$\int P(\boldsymbol{x}|\boldsymbol{\alpha}) d\boldsymbol{x} = 1.$$

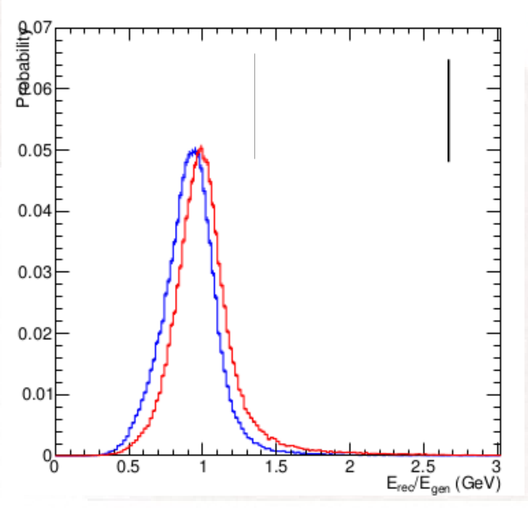
Transfer functions

- One assumptions: the transfer functions are "factorisable"
  - can be written as the product of single-particle resolution functions

$$W(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{n} W_i(x^i, y^i),$$

- Transfer functions could be parameterized from MC simulation
- Match reconstructed objects with partons stored by the generator history information, e.g. within a ΔR range





MEM Setup

MEM weight

Element of phase space corresponding to unmeasured quantities

Squared matrix element
MadGraph C++ standalone

$$w_{i,\alpha}(\Phi') = \frac{1}{\sigma_\alpha}\int d\Phi_\alpha \cdot \delta^4\left(p_1^\mu + p_2^\mu - \sum_{k\geq 2} p_k^\mu\right)\cdot \frac{f(x_1,\mu_F)f(x_2,\mu_F)}{x_1x_2s}\cdot \left|\mathcal{M}_\alpha(p_k^\mu)\right|^2 \cdot W(\Phi'|\Phi_\alpha),$$

Integration:
VEGAS in ROOT

Enforcing 4-momentum conservation

Parton distribution function
LHAPDF interface
NNPDF2.3 LO

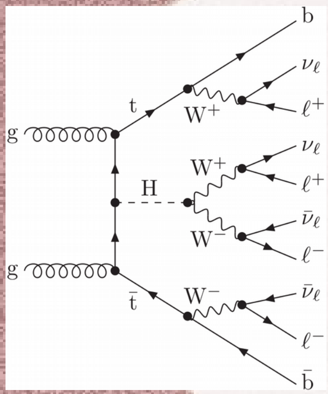
Transfer functions
Evaluated in MC

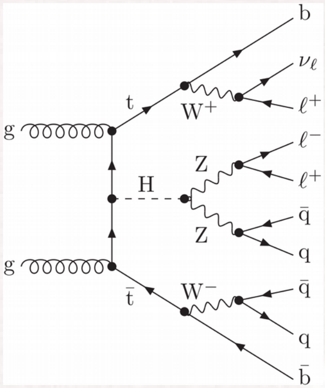
Advantages
- Makes maximal use of both experimental information and the theoretical model on an event-by-event basis
- Good discrimination vs irreducible background, especially for complex final state

Limits
- CPU intensive
- Subject to numerical inaccuracies
- Matrix element LO only

Application of the MEM in $t\bar{t}H$ multilepton analysis

- Standard model Higgs boson in association with a top quark pair





CMS-PAS-HIG-15-008, CMS-PAS-HIG-16-022, CMS-PAS-HIG-17-004

Higgs bosons decays to WW^*, $\tau\tau$, ZZ^* (→ multileptons)

Leptonic decay of at least one of the top quarks

Hadronic τ vetoed (measured in CMS-PAS-HIG-17-003)

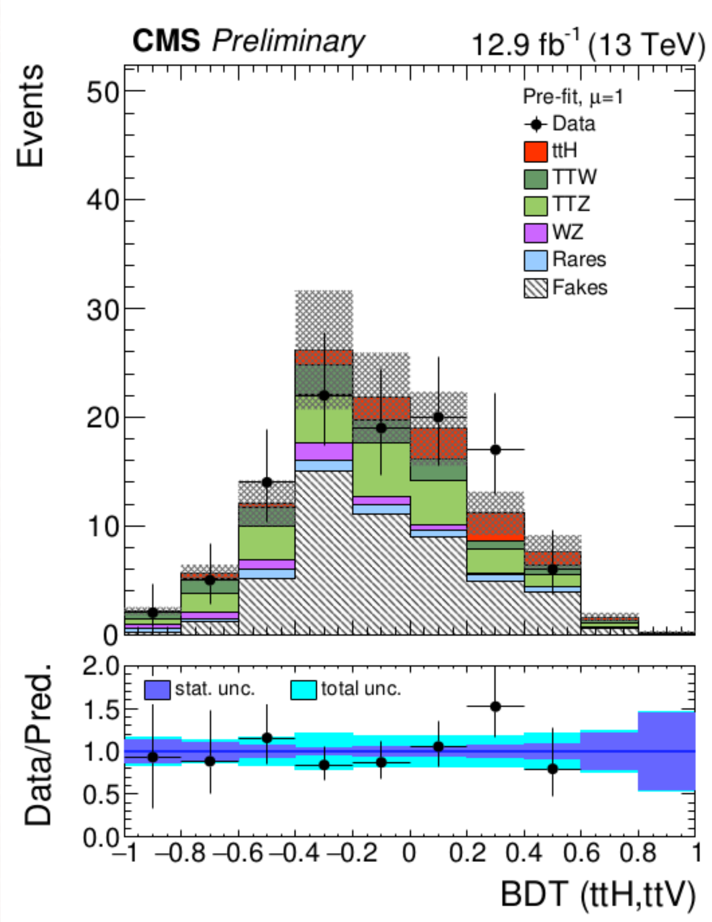
- Main irreducible background: $t\bar{t}V$ ($t\bar{t}W$, $t\bar{t}Z/\gamma^*$)
- A BDT classifiers MVA($t\bar{t}H$ vs $t\bar{t}V$) is trained to improve the separation between the signal and background
- Include the likelihood ratio of $t\bar{t}H$ and $t\bar{t}V$ from MEM as input of the BDT

Application of the MEM in $t\bar{t}H$ multilepton analysis

- Evaluate MEM weights under $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z/\gamma^*$ hypotheses
- MEM weights is the average weight of all possible lepton, jets, b-jets permutations
- The likelihood ratio of $t\bar{t}H$ and $t\bar{t}V$ from MEM as input of the BDT

$$\mathcal{L}_{t\bar{t}H\,vs\,t\bar{t}V} = -log\left(\frac{\sigma_{t\bar{t}V}w_{t\bar{t}V}}{\sigma_{t\bar{t}H}w_{t\bar{t}H} + \sigma_{t\bar{t}V}w_{t\bar{t}V}}\right)$$

- **Improved discrimination by 10% in 3l category in CMS-PAS-HIG-16-022**
- The significant of observation over 3 σ in CMS-PAS-HIG-17-004
  - **First evidence!** Hot topic at Moriond 2017



See talk from Na PENG about this analysis: http://indico.ihep.ac.cn/event/7102/session/7/contribution/30

MEM as a kinematic fit

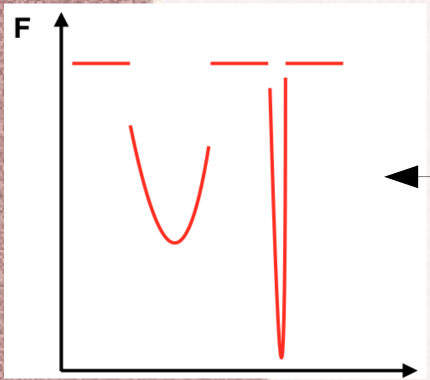
- In order to evaluate the weights, a non trivial multi-dimensional integration of complicated functions over the phase space has to be undertaken
- The numerical efficiency (and therefore the speed) of such integration is currently a serious limitation
- Instead of integration, look for the kinematic configuration having maximum probability

$$w_{i,\alpha}(\Phi') = \frac{1}{\sigma_\alpha}\int \underbrace{d\Phi_\alpha \cdot \delta^4\left(p_1^\mu + p_2^\mu - \sum_{k\geq 2} p_k^\mu\right)\cdot \frac{f(x_1,\mu_F)f(x_2,\mu_F)}{x_1 x_2 s}\cdot \left|\mathcal{M}_\alpha(p_k^\mu)\right|^2 \cdot W(\Phi'|\Phi_\alpha)}_{\textbf{MEM function } f(\Phi')}$$

- **Caveat : work still ongoing !**

Different minimization algorithms

- Easy way to find the maximum:
  - Obtained with the highest integrand value tried by VEGAS among all iterations of the integration
  - Repeat for all permutations and select the permutation with highest value



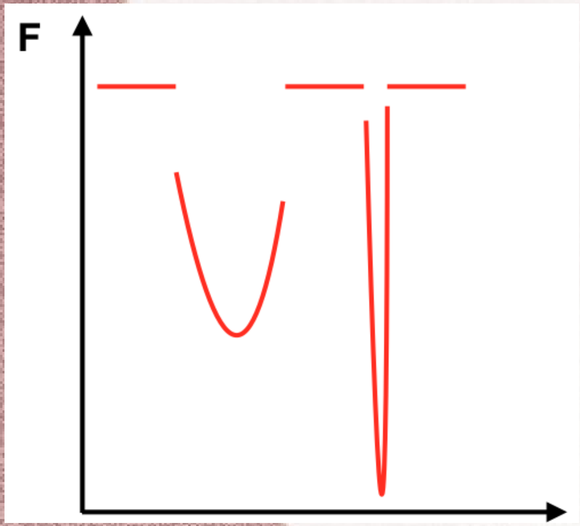
- Minimize the MEM function:

  -log(f(Φ’)), if f(Φ’) ≠ 0

  1000, if f(Φ’) = 0 (this happens when the phase space is forbidden by kinematics)
- Difficulties:
  - The function is convex by parts: jumps when the phase space is forbidden
  - Global minimum to be found among many local minima

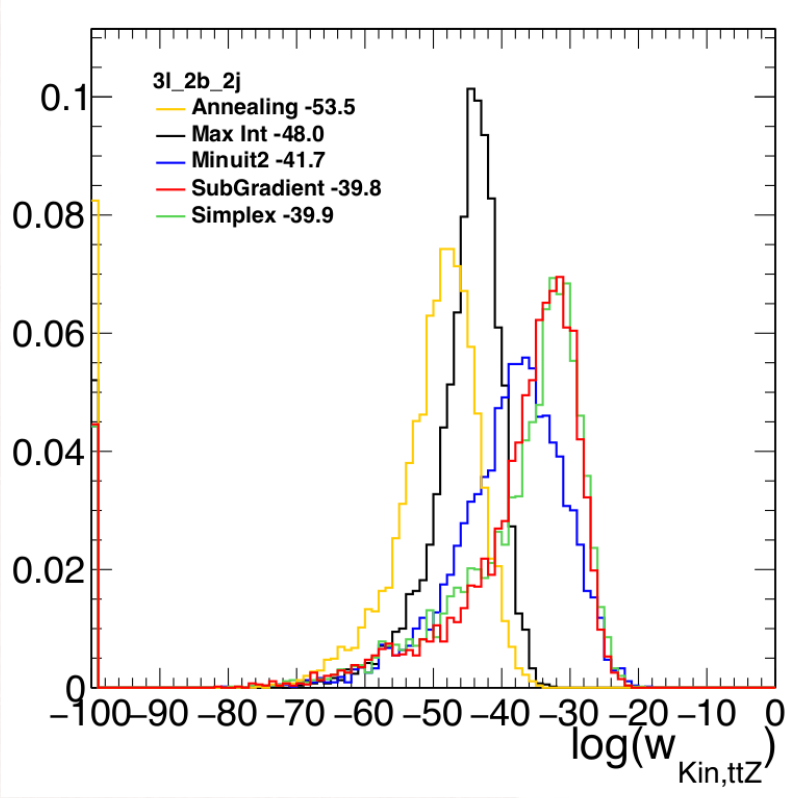
Comparison of Minimizers

- Try several minimization algorithms
  - "Max int": minimum is found during VEGAS integration
  - "SubGradient": custom minimization with steepest descent, choosing lowest gradient between left/right side, to avoid gaps where the function is null
  - Minuit2: usual Migrad minimization, variable metric method heavily using first derivatives
  - Simplex: simplex algorithm is adaptative, based on barycenters of previous steps, but does not use any derivative
  - Annealing: simulated annealing from GSL algorithm
- Initialization: repeat random initialization of variables for each event, until a set is found to get a non null MEM value



Minimization results

- Simplex algorithm is the best minimizer: most of the problem is solved when tackling the “jumps”
- “Subgradient" almost as good as the simplex algorithm: a custom algorithm, promising
- Minuit2 reaches more difficulties computing derivatives
- “MaxInt” not performing very well: can be improved with increasing number of VEGAS calls, but would increase CPU time
- Annealing: to be tuned



Conclusion

- The matrix element method is a powerful discriminator that makes maximal use of both experimental information (**x,** W(**x,y**)) and the theoretical model ($|M_\alpha|^2$(**y**)) on an event-by-event basis
- In the measurement of the production of standard model Higgs boson in association with a top quark pair, including the likelihood ratio of MEM in the classifier could improve the discrimination power by 10%
- Instead of integration, looking for the kinematic configuration having maximum probability could be one possible solution
  - work still ongoing

Thank you

Thank you

Backup

Backup

$$d\Phi_{top,had} \propto dE_b d\theta_b d\phi_b \cdot d\theta_{j1} d\phi_{j1} \cdot d\theta_{j2} d\phi_{j2} \cdot dm_W$$

$$d\Phi_{top,lep} \propto dE_b d\theta_b d\phi_b \cdot dE_l d\theta_l d\phi_l \cdot d\phi_\nu dm_W$$

$$d\Phi_{H\to 2l2\nu} \propto dE_{l1} d\theta_{l1} d\phi_{l1} \cdot dE_{l2} d\theta_{l2} d\phi_{l2} \cdot dE_{\nu 1} d\theta_{\nu 1} d\phi_{\nu 1} \cdot d\phi_{\nu 2} dm_{W2}$$

$$d\Phi_{H\to l\nu jj} \propto dE_{j1} d\theta_{j1} d\phi_{j1} \cdot dE_{j2} d\theta_{j2} d\phi_{j2} \cdot dE_{l1} d\theta_{l1} d\phi_{l1} \cdot d\phi_{\nu 1} dm_{W1}$$

$$d\Phi_Z \propto dE_{l1} d\theta_{l1} d\phi_{l1} \cdot dE_{l2} d\theta_{l2} d\phi_{l2}$$

$$d\Phi_W \propto dE_l d\theta_l d\phi_l \cdot d\theta_\nu d\phi_\nu dm_W$$

- Assumptions:
  - Assume narrow-width for Top quark and Higgs boson
  - Treat final-state b from top as massive
  - Keep full W and Z propagators in the top ME: follows a Breit-Wigner
  - Dileptons: Z and γ^* contributions included
- Transfer functions
  - The lepton energy and its direction is assumed to be perfectly measured
  - The direction of quarks is assumed to be perfectly measured by the direction of the reconstructed jet
  - Jets and b-jets energy transfer functions are evaluated in MC simulation – histograms parameterized as a function of E(rec) /E(gen)

Sample

- ttZ 3 leptons sample generated with MG5_aMC@NLO
  - 13 TeV, semi-leptonic top decay
  - LO
  - Pythia8 for showering
- CMS Detector simulation with Delphes3
  - No PU included yet
- Event selection ttZ control region