Matrix Element Methods in Higgs Experimental Physics

Eric Feng
Argonne National Laboratory

Zurich Phenomenology Workshop
University of Zurich
Jan 8-10, 2014

Introduction

- Matrix element method (MEM) important in many Higgs analyses and other areas of HEP
  - Separate signal vs. background in search for Higgs boson and measurements of its rates/couplings
  - Distinguish different spin/CP hypotheses, particularly in H->ZZ->4l channel
  - Analytical fits to extract underlying parameters of model
- In this talk, discuss technical aspects of MEM from experimental perspective and compare/contrast with machine learning algorithms
  - Use public examples from LHC and Tevatron experiments to illustrate points
  - Disclaimer: Not official talk on behalf of ATLAS – own views
- Non-exhaustive list of papers/analyses cited, so please excuse if your favorite isn't listed

A Brief (Non-Exhaustive) History of MEM

- Note complementary CDF & D0 analysis for each topic below

- Matrix element method first proposed:
  K. Kondo. "Dynamical Likelihood Method for Reconstruction of Events with Missing Momentum." J. Phys. Soc. Japan 57, 4126 (1988)
- Top quark mass:
  D0 Collaboration. Nature 429, 638 (2004).
- Helicity of W boson in top decays:
  D0 Collaboration. Phys. Lett. B617, 1–10 (2005).
- Electroweak single top:
  CDF Collaboration. Phys. Rev. Lett. 103, 092002 (2009).
- Search for Higgs boson in W/ZH(bb):
  CDF Collaboration. Phys. Rev. D. 85, 072001 (2012)
- Discovery of Higgs boson:
  ATLAS Collaboration. Phys. Lett. B716, 1 (2012), 1207.7214.
  CMS Collaboration. Phys. Lett. B716, 30 (2012), 1207.7235.

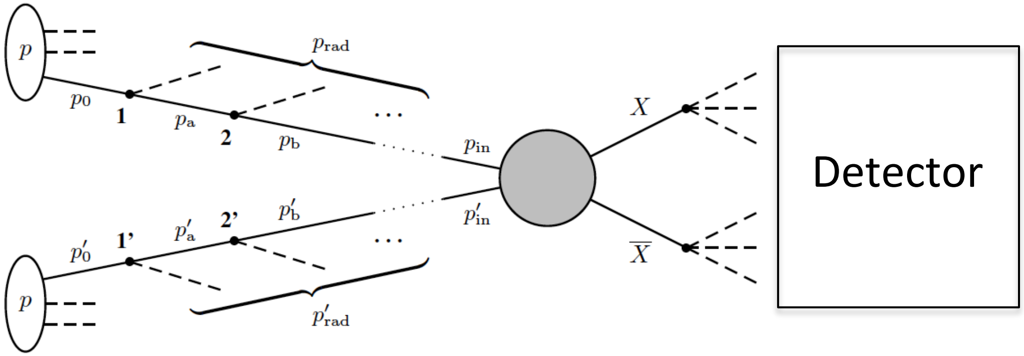
Nomenclature

“A rose by any other name would smell as sweet.”
- William Shakespeare, Romeo and Juliet

- Many different names for discriminant in MEM:
    - Event probability discriminant (EPD)
    - Likelihood discriminant (LD)
    - Kinematic discriminant (KD)
    - Matrix element likelihood analysis (MELA)
    - Matrix element kinematic discriminant (MEKD)
    - Etc.

Matrix element method

Procedure (I)



$$\underbrace{P(x|\alpha)}_{\text{Event weight}} = \underbrace{\frac{1}{\sigma}\int d\phi(y)|M|^2}_{\text{ME-squared (norm. to unity)}} \underbrace{dw_1 dw_2 f_1(w_1) f_2(w_2)}_{\text{PDF}} \underbrace{W(x,y)}_{\text{Transfer function}}$$

- Event weight *P(x|* α*)* = probability to observe kinematic variables *x* at detector-level from parton-level configuration *y*
  - Integrate over unmeasured observables (momenta of initial and non-reconstructed particles, e.g. neutrinos)
  - *dφ(y)* is phase space element – kinematic acceptance of detector/analysis

Procedure (II)

- Weight each configuration (event) by probability *P(x|α)* to produce observed measurement
  - Analytical/numerical integration with RooStats, JHU, MadWeight, etc
  - Integration challenging for arbitrary process, typically using importance sampling (adaptive MC) like VEGAS
- Transfer function W(x,y) = probability to reconstruct parton state *y* as measurement *x*
  - Most important experimental component (account for ISR/FSR, soft physics, and detector effects) in many analyses
  - Reflects efficiencies and energy/momentum resolutions (scales assumed calibrated)
- Combine weights into likelihood for parameter of interest α (top quark mass, Higgs signal strength, etc):

$$L(\alpha) = e^{-N\int \bar{P}(x,\alpha)dx} \prod_{i=1}^{N} \bar{P}(x_i, \alpha)$$

- Maximize likelihood (minimize NLL) to measure that parameter

Initial/final state radiation (I)

- One of main limitations is that ME generally only considered at LO
  - Initial/final radiation generally requires explicit treatment through transfer function W(parton, detector)
    - Thus W: parton -> hadron -> jet
    - Underlying event and hadronization also absorbed here
  - By contrast Monte Carlos at LO+PS (Pythia, Herwig), multi-leg LO+PS (Alpgen, Sherpa), NLO+PS (MC@NLO, POWHEG), etc
    - Different tunes of given parton shower, UE, hadronization
- Activity on-going to extend to NLO ME:

  J. Campbell, W. Giele, C. Williams. JHEP 1211, 043 (2012)

  J. Campbell, K. Ellis, W. Giele, C. Williams. 1301.7086 (2013)
  - Some indications that LO->NLO gives improvements in S/B of the order of 10% since NLO covers larger (real) phase space
  - Again additional radiation (and UE, hadronization) still must be included separately in transfer function

Initial/final state radiation (II)

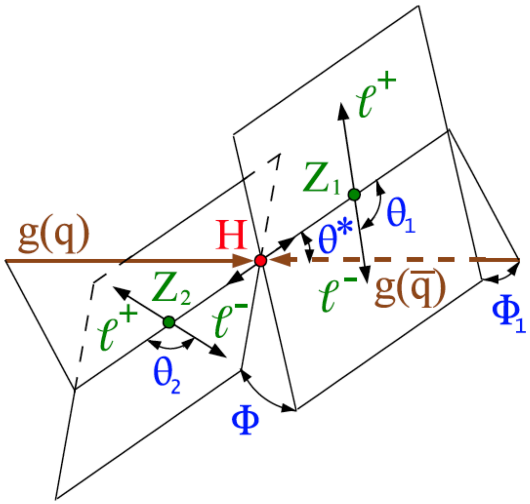
- QCD ISR from interacting partons can be important
  - For example, change in momentum boosts Higgs system
- ISR often (but not always) soft and collinear with emitting partons, so transverse pT not typically large
  - Primarily important for separation of signal vs. background, and different Higgs production modes (ggF vs. VBF for coupling measurements)
  - Higgs pT does not have large impact on CP separation, although plays role in qq vs. gg production of spin-2 RS graviton-like particle
- QCD FSR important for final states with jets/MET, not leptons/photons
- QED FSR treated explicitly by detector reconstruction/calibration
  - For electrons, photons already included during reconstruction from energy deposits in EM calorimeter shower
  - For muons, photons are reincorporated during reconstruction
- Side note: At lepton collider e.g. LEP, no PDF and QED ISR instead, but analogous issues

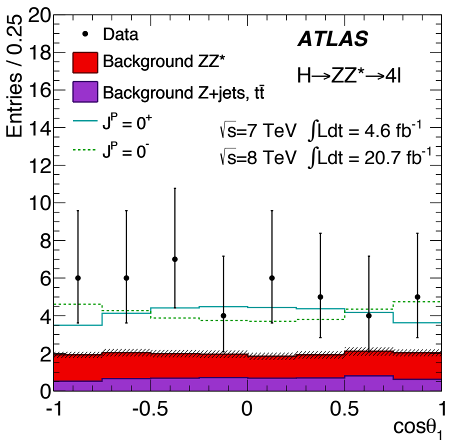
Matrix element
vs.
Machine learning algorithms

Spin and CP hypothesis testing with MVAs

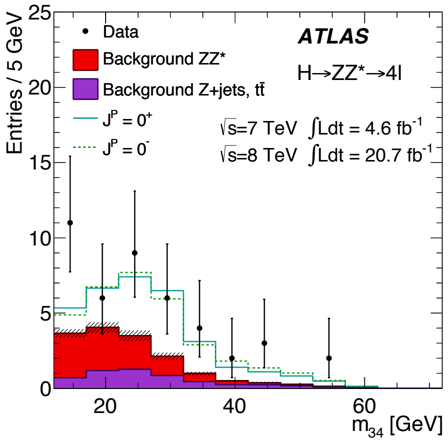
- H->ZZ->4l ideal: 3 masses (m_{4l}, m_{12}, m_{34}) and 5 angles ($\cos(\theta^*)$, φ_1, $\cos(\theta_1)$, $\cos(\theta_2)$, φ)
    - To test CP and spin hypotheses, separation from $\cos(\theta_1)$ and m_{34} respectively
- Separation power of individual kinematic observables very marginal, particularly given low statistics
    - Much higher separation power from multivariate discriminant such as BDT
- Phys. Lett. B 726, 120 (2013):

JHU - arXiv:1208.4018

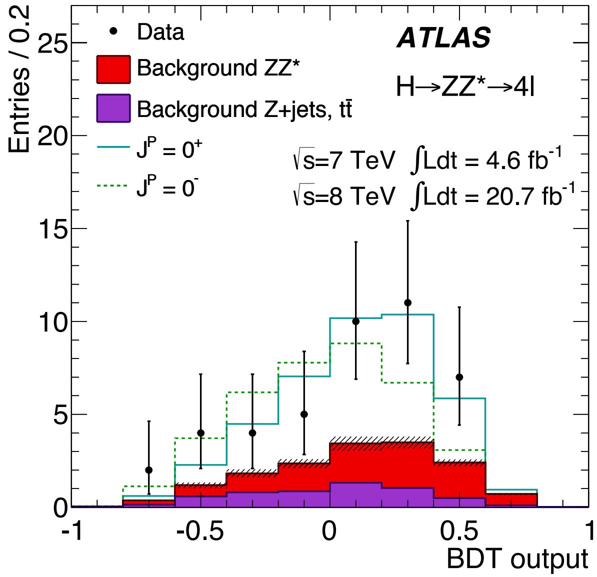




x

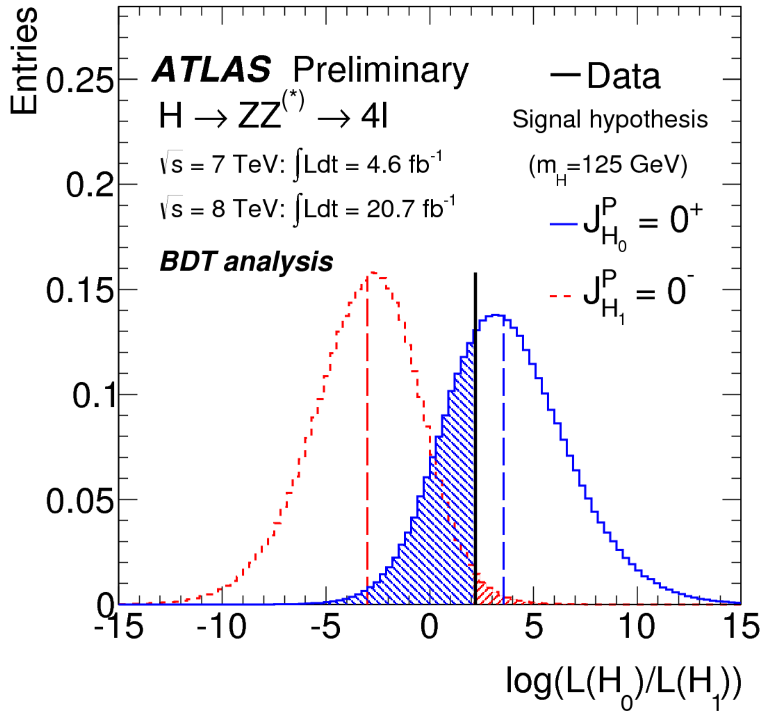


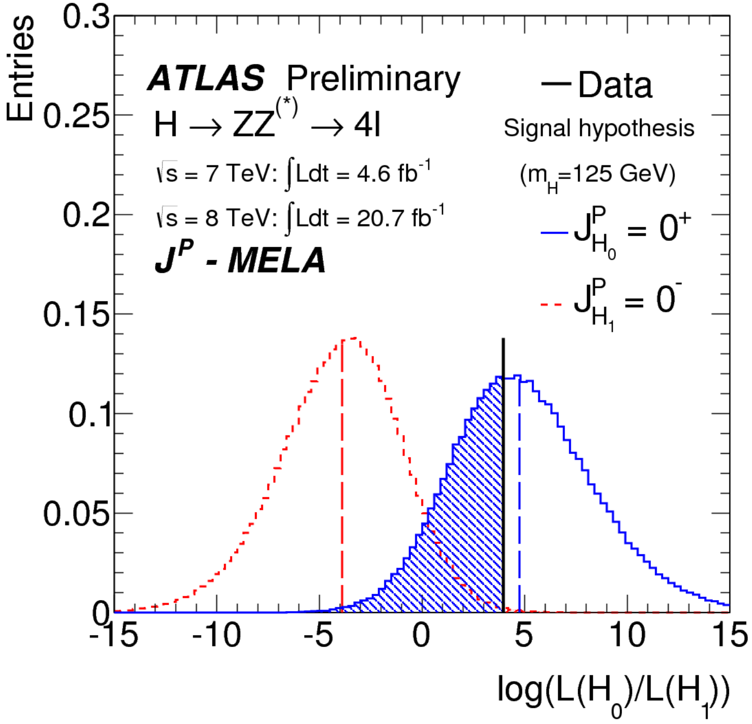
x … =



BDT vs. Matrix Element

- ATLAS-CONF-2013-013
- Better expected (and observed) separation for ME approach compared to BDT
  - $CL_s = p_0$ (alternative J^P) / $(1 - p_0(0^+))$
  - 0^- hypothesis excluded at 97.8% (BDT) and 99.6% (MELA) confidence in favor of 0^+ hypothesis





ME vs. Machine learning

- Both used to combine several variables into more sensitive one
- **Sensitivity:** All physics information *at given order* is in ME by construction
  - Lacks higher-order radiative corrections and correlations
  - Machine learning may better correlations using full simulation, compared to analytical transfer functions
- **Robustness:** Since all physics information already encoded in matrix element:
  - No need to re-learn using BDT, NN, etc -- possibly learn suboptimal/wrong dependence
  - MC mismodeling of data affects both ME and machine learning discriminants
- **Transparency:** Conceptually simple method – easy to spot potential problems, e.g. with transfer function mismodeling
  - Machine learning algorithms are black box
- **Power:** Fit fundamental parameters of model in large, continuous phase space
  - Compare to pair-wise testing of finite discrete SM-like hypotheses
  - Difficult to generalize machine-learning algorithms to large phase space of continuous parameters because computationally intensive
- Often (not always) gain best of both worlds with ME as input to BDT / NN

Technical issues in MEM analyses

Tevatron Higgs searches using ME

- Search for Higgs boson in Wh(bb) using MEM
    - CDF - Phys. Rev. D. 85, 072001 (2012)
- Similar idea in other searches e.g. for Zh(bb) – see talk later
    - CDF - Phys. Rev. D 80, 71101 (2009)
- Event probability discriminant for 1-btag and 2-btag samples
    - Relative probability normalized wrt all tagged/non-tagged processes:

$$EPD \equiv \frac{b\ \hat{P}_{WH}}{b\ (\hat{P}_{WH} + \hat{P}_{Wb\bar{b}} + \hat{P}_{t\bar{t}} + \hat{P}_s + \hat{P}_t) + (1-b)(\hat{P}_{Wc\bar{c}} + \hat{P}_{Wcj} + \hat{P}_{W+l} + \hat{P}_{Wgg} + \hat{P}_{dib})}$$

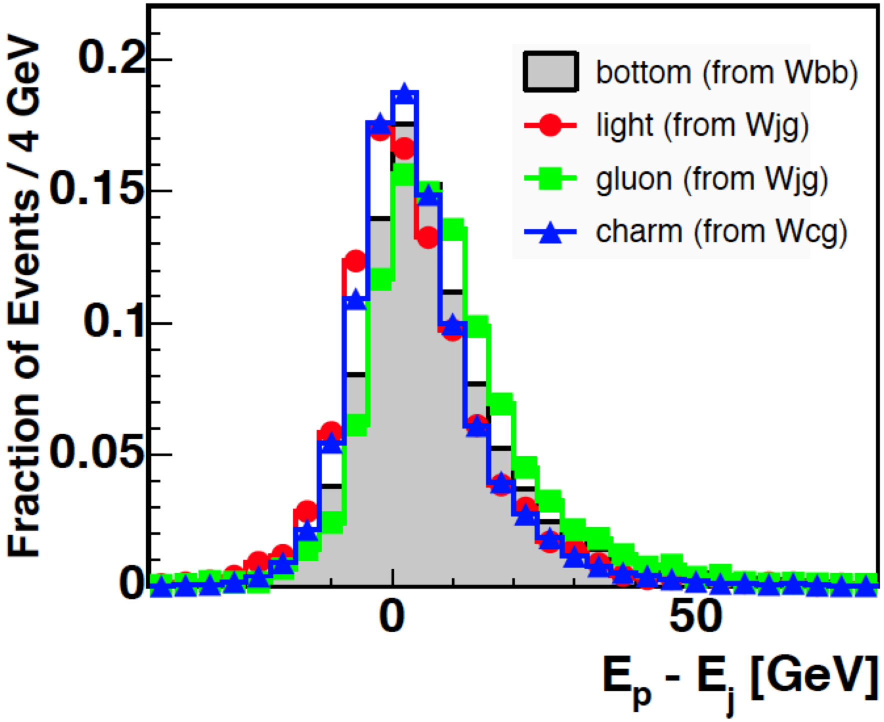
$$EPD \equiv \frac{b_1\ b_2\ \hat{P}_{WH}}{b_1 b_2(\hat{P}_{WH} + \hat{P}_{Wb\bar{b}} + \hat{P}_{t\bar{t}} + \hat{P}_s) + b_1(1-b_2)\hat{P}_t + (1-b_1)(1-b_2)(\hat{P}_{Wc\bar{c}} + \hat{P}_{Wcj} + \hat{P}_{W+l} + \hat{P}_{Wgg} + \hat{P}_{dib})}$$

Transfer function for jet energy resolution

$$W(y,x) = \delta^3(\vec{p}_l^{\,y} - \vec{p}_l^{\,x}) \prod_{i=1}^{2} \delta^2(\Omega_i^y - \Omega_i^x) \prod_{k=1}^{2} W_j(E_{p_k}, E_{j_k})$$

$$W_{\rm jet}(E_{\rm parton}, E_{\rm jet}) = \frac{1}{\sqrt{2\pi}(p_2 + p_3 p_5)} \left(\exp \frac{-(\delta_E - p_1)^2}{2p_2^2} + p_3 \ \exp \frac{-(\delta_E - p_4)^2}{2p_5^2} \right)$$

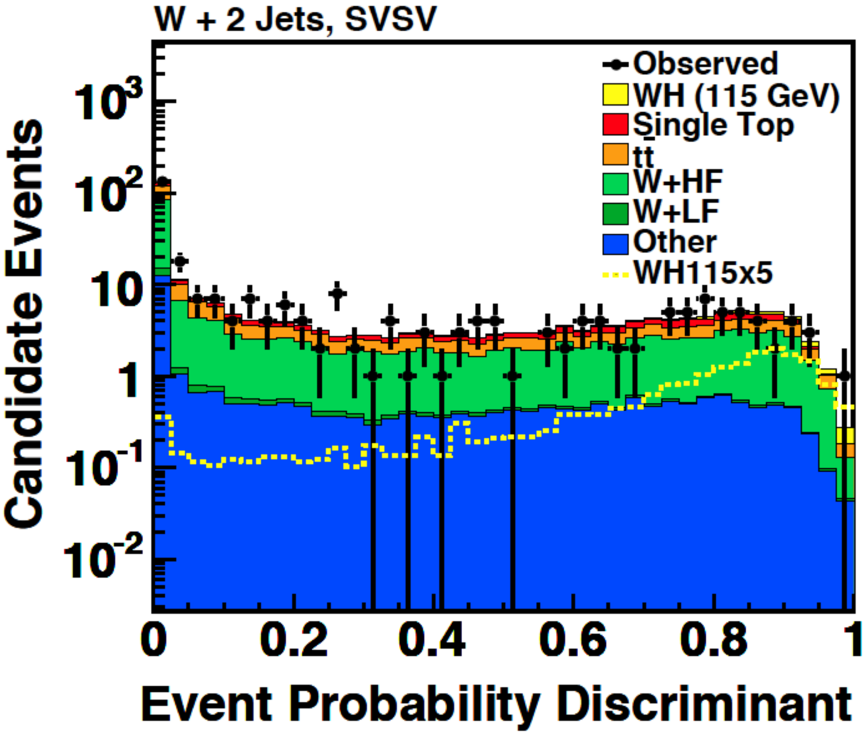
- Double Gaussian transfer function
- Energy resolution terms p_i (noise, stochastic, constant): $p_i/E = p_{i,0}/E + p_{i,1}/\mathrm{sqrt}(E) + p_{i,2}$ as function of pseudorapidity using matched partons and jets in MC
- Gaussians possible for resolutions in lepton momenta and object angles
  - Here δ functions good approx. and faster computationally
- Biases assumed calibrated away
- Efficiencies for trigger and reconstruction explicitly corrected



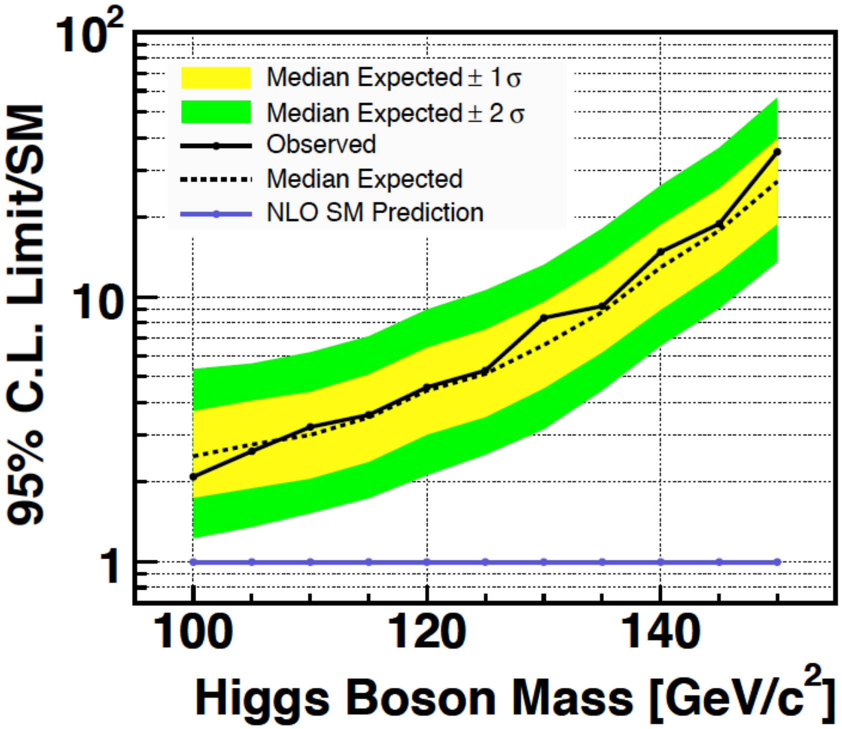
Finite width and combinatorics

- Important drawback is implicit 1:1 correspondence between partons and reconstructed objects (jets/leptons/etc), but theoretically invalid
- For quarks/gluons (color triplets/octets), cannot map 1:1 to jets (color neutral)
  - Initial/final state radiation and hadronization result in parton->jet map that is many -> many
- For indistinguishable particles e.g. leptons from Z decay, cannot know what assignment is right one
  - In principle, require weighted sum over all possible permutations
  - In practice, just pair same-flavor, opposite-sign leptons closest to Z mass since relatively similar result
- Efficiency, purity, and resolution depend on ΔR matching criterion
  - Tighter ΔR cut: Lower efficiency, but better purity and resolution
  - Looser ΔR cut: Higher efficiency, but worse purity and resolution along with larger non-Gaussian tails
- Optimization and systematic uncertainty of ΔR cut important
- All above issues identical to those in jet energy calibration

Tevatron Higgs ME search results



- Higgs signal would peak at higher EPD than backgrounds (ttbar, W+jets)
- Unfortunately statistics relatively low with ~5.6 fb^{-1} at 1.96 TeV



- Observed (and expected) limit at ~$5\sigma_{SM}$ for m_H = 125 GeV
  - Best limit until LHC surpassed

Correlations and background in H->ZZ->4l

- Predictions must accurately describe all 1D distributions of the observables for both signal and background
  - 2D correlations must also be described well if they are significant
- 10 pairs of production angles {θ^*, φ_1} and decay angles {cos(θ_1), cos(θ_2), φ} in H->ZZ->4l
  - Final states: 4e, 2e2μ, 4μ
- Production/decay angles φ_1 and φ are strongly anti-correlated for signal and continuum ZZ
- Weak correlation between cos(θ_1) and cos(θ_2)
- For given m_{4l}, m_1 and m_2 also anti-correlated since they sum to m_{4l}
- Analytical calculation for ZZ background by R. Vega-Morales et al (arXiv:1211.1959)
  - 75,000 lines for 64 diagrams contributing to SM ZZ (!)
  - 15,000 lines for 16 diagrams for 0^+ signal (!)
  - Authors say these formulas are already "simplified" compared to their initial results

Systematic uncertainties

- Experimental uncertainties:
  - Object performance
    - Energy/momentum scale & resolution of leptons, photons, jets
    - Efficiency of trigger and reconstruction
  - Modeling of ISR/FSR and detector effects by transfer functions
    - Closure test by injecting given signal (e.g. SM 0^+) from reco MC, and verifying that analysis correctly recovers input signal
      - Similarly for injection of non-standard signals of 0^-, 2^+, etc.
    - Checks that transfer functions can reproduce MC
  - Tests of theory predictions (in framework) against data in control sample
    - Verify corrected analytical predictions model actual data well
  - Luminosity
- Theoretical uncertainties:
  - Renormalization/factorization scales
  - PDF and α_S
  - Tune (PS, UE, hadronization) of MCs for transfer function
  - Higgs pT

Background and spin/CP discriminants in H->ZZ->4l

Discriminants

- CMS – arXiv:1312.5353
- Discriminants against background and other spin/CP hypotheses
  - Correlations between discriminants taken into account

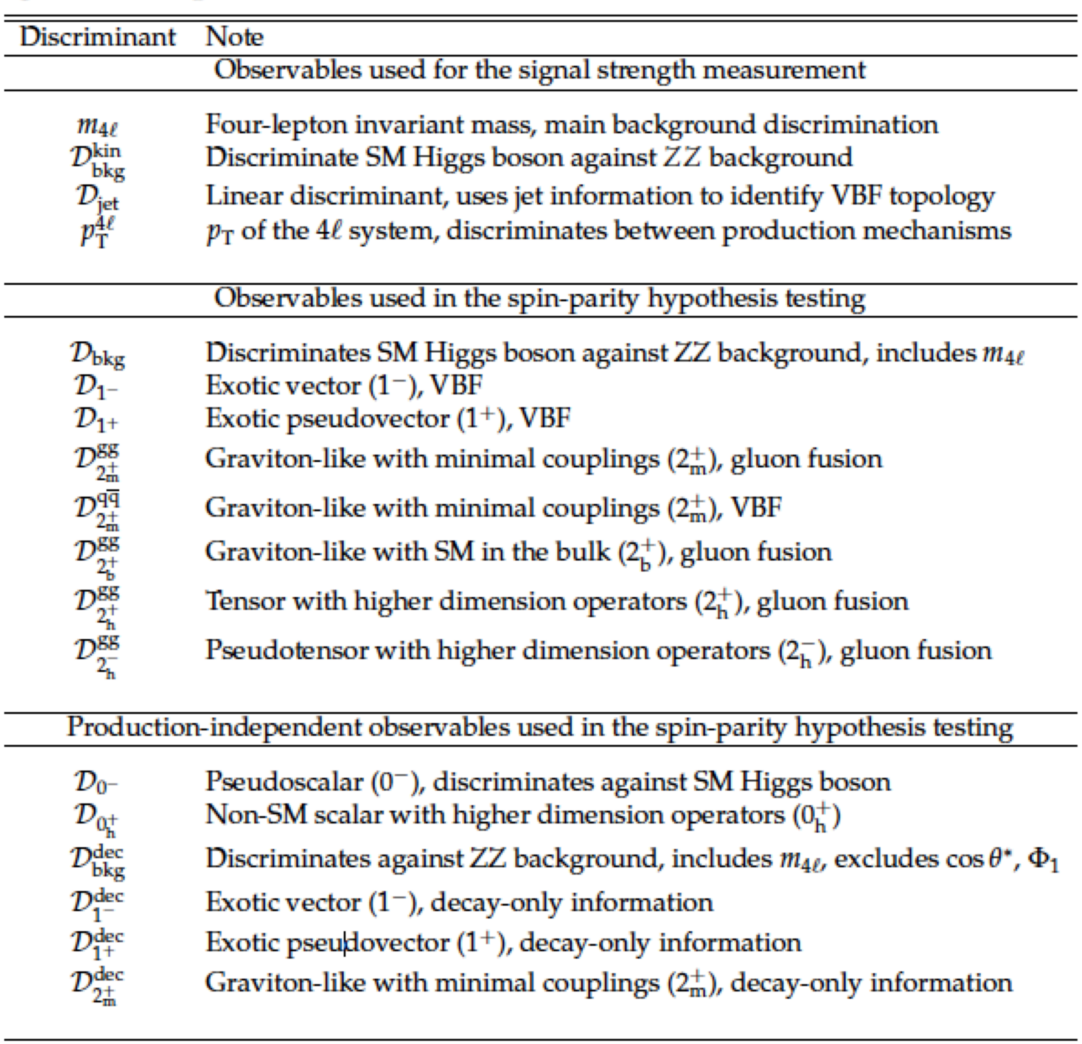
$$\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}} = \frac{\mathcal{P}_{0^+}^{\mathrm{kin}}}{\mathcal{P}_{0^+}^{\mathrm{kin}} + \mathcal{P}_{\mathrm{bkg}}^{\mathrm{kin}}} = \left[1 + \frac{\mathcal{P}_{\mathrm{bkg}}^{\mathrm{kin}}(m_{Z_1}, m_{Z_2}, \vec{\Omega} | m_{4\ell})}{\mathcal{P}_{0^+}^{\mathrm{kin}}(m_{Z_1}, m_{Z_2}, \vec{\Omega} | m_{4\ell})}\right]^{-1}$$

$$\mathcal{D}_{\mathrm{bkg}} = \left[1 + \frac{\mathcal{P}_{\mathrm{bkg}}^{\mathrm{kin}}(m_{Z_1}, m_{Z_2}, \vec{\Omega} | m_{4\ell}) \times \mathcal{P}_{\mathrm{bkg}}^{\mathrm{mass}}(m_{4\ell})}{\mathcal{P}_{0^+}^{\mathrm{kin}}(m_{Z_1}, m_{Z_2}, \vec{\Omega} | m_{4\ell}) \times \mathcal{P}_{\mathrm{sig}}^{\mathrm{mass}}(m_{4\ell} | m_{0^+})}\right]^{-1}$$

$$\mathcal{D}_{J^P} = \left[1 + \frac{\mathcal{P}_{J^P}^{\mathrm{kin}}(m_{Z_1}, m_{Z_2}, \vec{\Omega} | m_{4\ell})}{\mathcal{P}_{0^+}^{\mathrm{kin}}(m_{Z_1}, m_{Z_2}, \vec{\Omega} | m_{4\ell})}\right]^{-1}$$

Discriminants

- All discriminants against backgrounds and alternative spin/CP hypotheses:

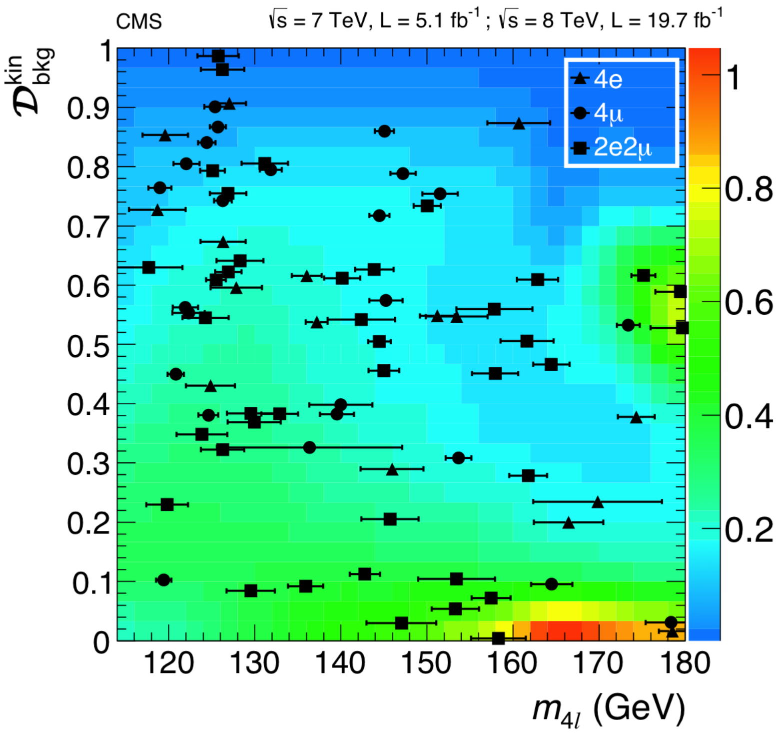


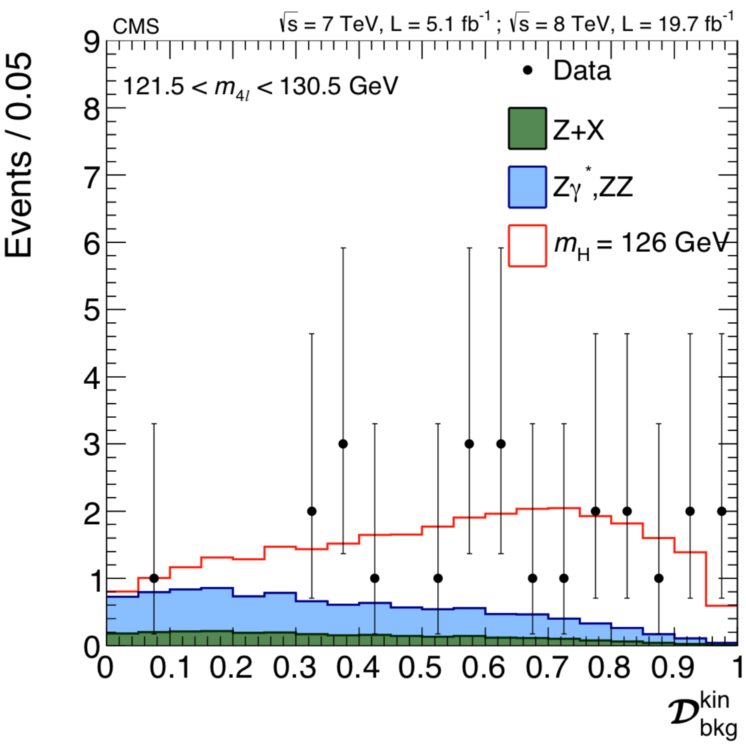
Discriminant	Note
Observables used for the signal strength measurement	
$m_{4\ell}$	Four-lepton invariant mass, main background discrimination
$\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$	Discriminate SM Higgs boson against ZZ background
$\mathcal{D}_{\mathrm{jet}}$	Linear discriminant, uses jet information to identify VBF topology
$p_{\mathrm{T}}^{4\ell}$	p_{T} of the 4ℓ system, discriminates between production mechanisms
Observables used in the spin-parity hypothesis testing	
$\mathcal{D}_{\mathrm{bkg}}$	Discriminates SM Higgs boson against ZZ background, includes $m_{4\ell}$
$\mathcal{D}_{1^-}$	Exotic vector (1^-), VBF
$\mathcal{D}_{1^+}$	Exotic pseudovector (1^+), VBF
$\mathcal{D}_{2_{\mathrm{m}}^+}^{\mathrm{gg}}$	Graviton-like with minimal couplings (2_{m}^+), gluon fusion
$\mathcal{D}_{2_{\mathrm{m}}^+}^{\mathrm{q\bar{q}}}$	Graviton-like with minimal couplings (2_{m}^+), VBF
$\mathcal{D}_{2_{\mathrm{b}}^+}^{\mathrm{gg}}$	Graviton-like with SM in the bulk (2_{b}^+), gluon fusion
$\mathcal{D}_{2_{\mathrm{h}}^+}^{\mathrm{gg}}$	Tensor with higher dimension operators (2_{h}^+), gluon fusion
$\mathcal{D}_{2_{\mathrm{h}}^-}^{\mathrm{gg}}$	Pseudotensor with higher dimension operators (2_{h}^-), gluon fusion
Production-independent observables used in the spin-parity hypothesis testing	
$\mathcal{D}_{0^-}$	Pseudoscalar (0^-), discriminates against SM Higgs boson
$\mathcal{D}_{0_{\mathrm{h}}^+}$	Non-SM scalar with higher dimension operators (0_{h}^+)
$\mathcal{D}_{\mathrm{bkg}}^{\mathrm{dec}}$	Discriminates against ZZ background, includes $m_{4\ell}$, excludes $\cos\theta^*$, Φ_1
$\mathcal{D}_{1^-}^{\mathrm{dec}}$	Exotic vector (1^-), decay-only information
$\mathcal{D}_{1^+}^{\mathrm{dec}}$	Exotic pseudovector (1^+), decay-only information
$\mathcal{D}_{2_{\mathrm{m}}^+}^{\mathrm{dec}}$	Graviton-like with minimal couplings (2_{m}^+), decay-only information

CMS
arXiv:1312.5353

Discriminant against background

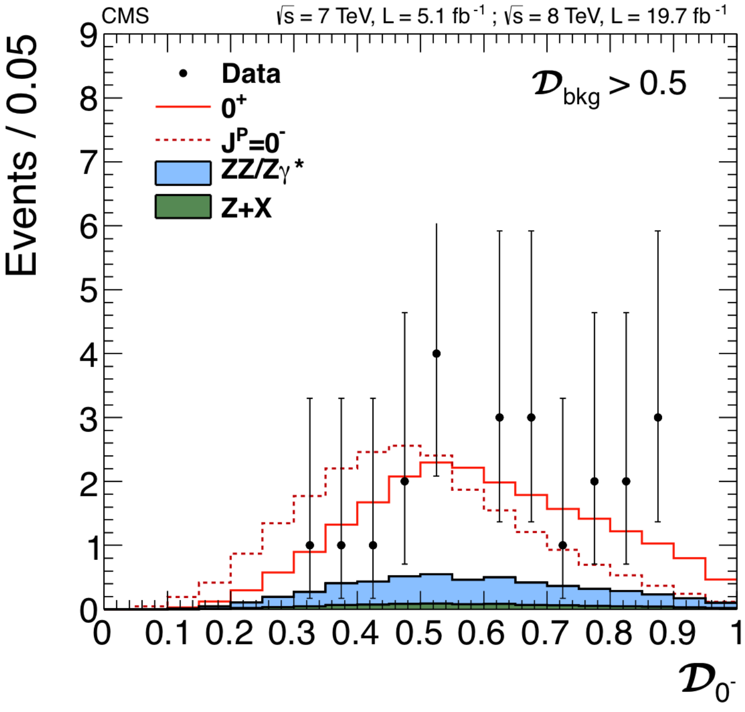
- CMS - arXiv:1312.5353
- Background discriminant values of 4-lepton events in mass window around 126.5 GeV favor Higgs signal
  - Even clearer from 1D projection of discriminant (right)
- Events at other masses consistent with background

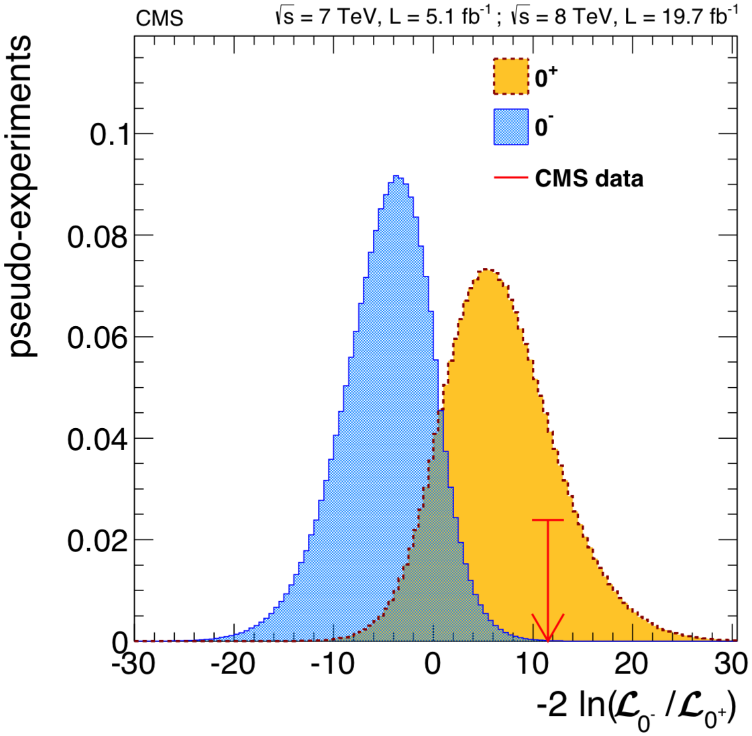




Discriminant against alternate hypotheses

- CMS - arXiv:1312.5353
- Spin/CP discriminant for signal-like events ($D_{bkg} > 0.5$)
- 0- hypothesis excluded at 99.9% CL_s (3.6σ)





Analytical fits for CP fraction and phase (operators) in H->ZZ->4l

CP fraction via fit to analytical ME

- ATLAS: ATL-PHYS-PUB-2013-013
- Extend ME hypothesis tests to determine CP fraction in H->ZZ->4l final state (including interference between operators) via fit to analytical prediction
  - Unbinned maximum likelihood fit to simultaneously measure spin, CP, and helicity amplitudes
  - Recovers existing results from discrete hypothesis tests as limiting cases
- Analytical ME from JHU calculations: arXiv:1208.4018
- Observables for ML fit:
  - Masses: m_{4l}, m_{12}, m_{34}
  - Angles: production $\{\cos(\theta^*), \varphi_1\}$ and decay $\{\cos(\theta_1), \cos(\theta_2), \varphi\}$
- Transfer functions model detector effects using fits to Fourier series of polynomials (numerical approximation)

CP fraction and operator phase

- Amplitude in terms of CP-related operators: JHU - arXiv:1208.4018

$$A(H \to VV) \sim (a_1 M_H^2 g_{\mu\nu} + a_2(q_1 + q_2)_\mu (q_1 + q_2)_\nu + a_3 \epsilon_{\mu\nu\alpha\beta} q_1^\alpha q_2^\beta)\epsilon_1^{*\mu} \epsilon_2^{*\nu}$$

$$a_1 = g_1 \frac{m_V^2}{m_H^2} + \frac{s}{m_H^2}\left(2g_2 + g_3 \frac{s}{\Lambda^2}\right); \quad a_2 = -\left(2g_2 + g_3 \frac{s}{\Lambda^2}\right); \quad a_3 = -2g_4$$

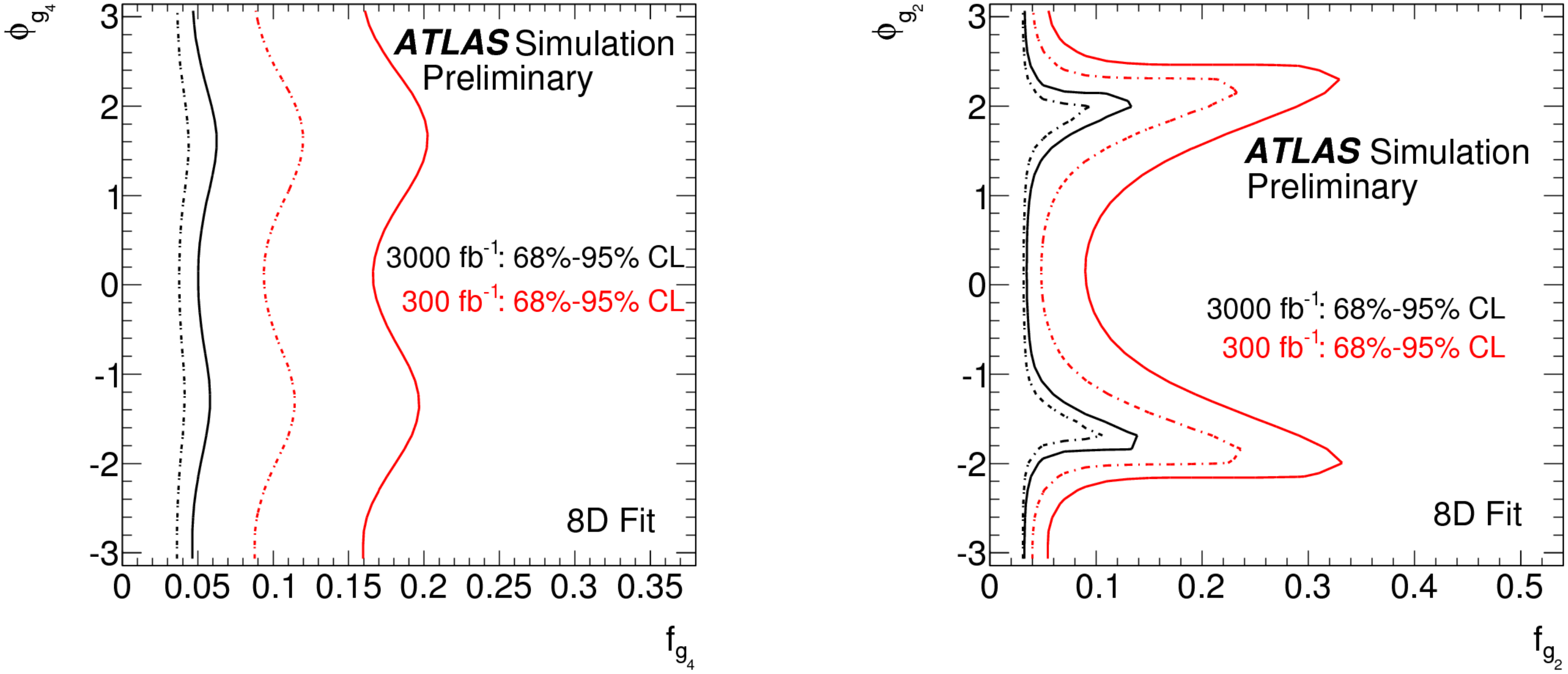
- 0^+: $g_1=1$, $g_4=0$ (SM)
- 0^-: $g_1=0$, $g_4=1$

$$f_{g_i} = \frac{|g_i|^2 \sigma_i}{|g_1|^2 \sigma_1 + |g_2|^2 \sigma_2 + |g_4|^2 \sigma_4}; \quad \phi_{g_i} = \arg\left(\frac{g_i}{g_1}\right)$$

- f_{g4} is observed CP-odd fractional yield
- φ_{g4} is phase between g_4 and g_1 (real/imaginary components)
  - Discussion within LHXSWG whether φ is useful to measure
- Analogous for g_2, which is higher-dimension CP-even operator

Expected limits on CP fraction

- With 300 (3000) fb^{-1} of data at 14 TeV, expected limits at 95% CL:
  - $f_{g4} < 0.15\ (0.04)$
  - $f_{g2} < 0.32\ (0.13)$
- Expected 2D exclusion limits for CP fractional yields and phase (real/imaginary components translated to polar coordinates)



Conclusions

- Wide-spread development of ME method in last 15 years
  - Common use in searches for the Higgs boson and measurements of its spin/CP, particularly H->ZZ->4l channel
- Multivariate approach is similar in many ways to other machine learning techniques like BDT, NN, etc
  - Strengths over those include maximal sensitivity, robustness, transparency, and direct fit of model parameters
  - Large drawback is that LO ME requires transfer function for ISR/FSR and soft physics, and difficult to model all correlations
    - NLO ME will be significant improvement
- MEM is ideal in final states not sensitive to QCD radiation (no jets/MET) and/or transverse boosts wrt lab frame
  - Best compromise in many cases may be some hybrid where ME is fed as input to machine learning algorithm used on MC
  - MEM standalone is still most practical solution for fits to analytical prediction to extract many fundamental parameters of theory