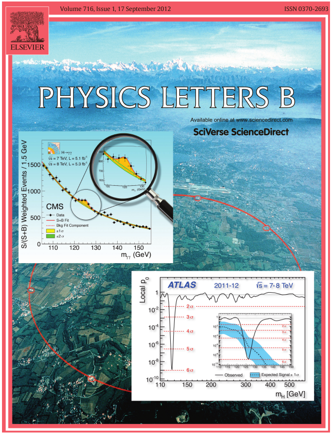
The discovery of the Higgs boson, and precision measurements at the energy frontier

Josh Bendavid (CERN)

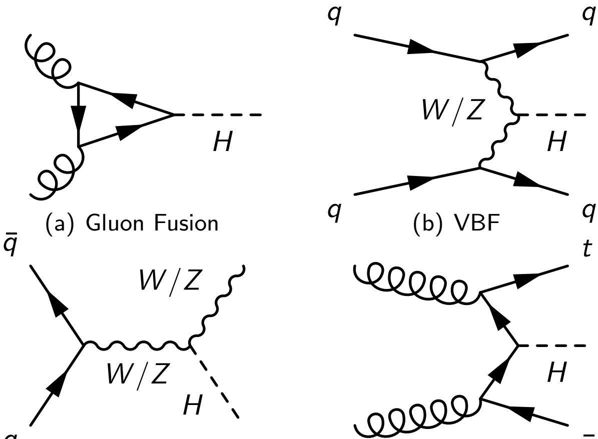
Apr. 8, 2019
Guido Altarelli Award 2019
XXVII International Workshop on Deep Inelastic Scattering and Related Subjects (DIS2019)
Torino, Italy

Introduction

- Higgs Boson discovered by ATLAS and CMS collaborations in summer 2012, substantial additional data incorporated at 8 and 13 TeV since then
- Higgs$\rightarrow \gamma\gamma$ in CMS played a critical role in the Higgs discovery announced in 2012
- Will discuss some details of the analysis, included associated use of machine learning techniques to increase the sensitivity
- Will discuss some other interesting technical developments related to reconstruction algorithms, data analysis, machine learning, Monte Carlo generators



Higgs Production at the LHC



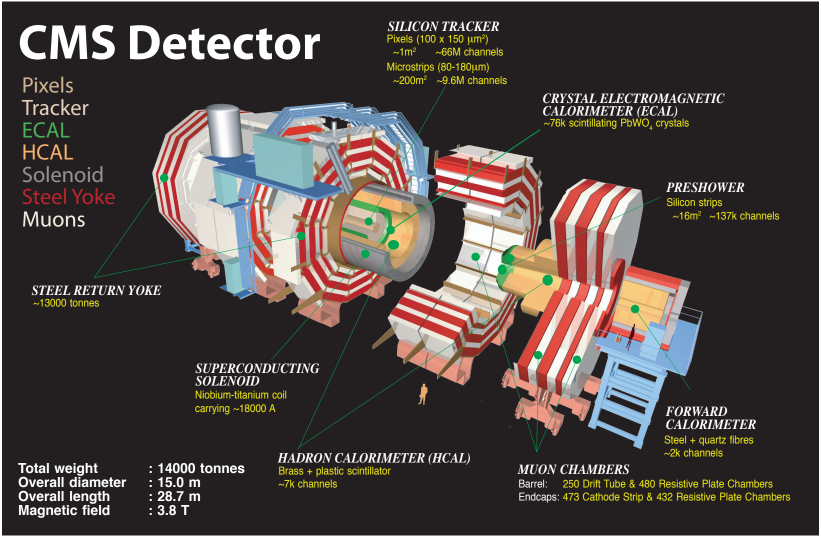
(a) Gluon Fusion

(b) VBF

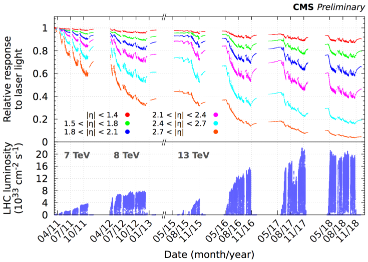
(c) W/ZH

(d) $t\bar{t}H$ ($b\bar{b}H$)

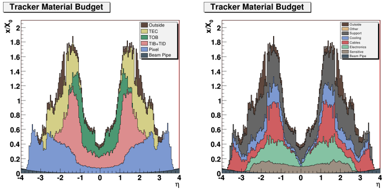
The CMS Detector



The CMS Detector: Some Challenges



(a) ECal Transparency Loss

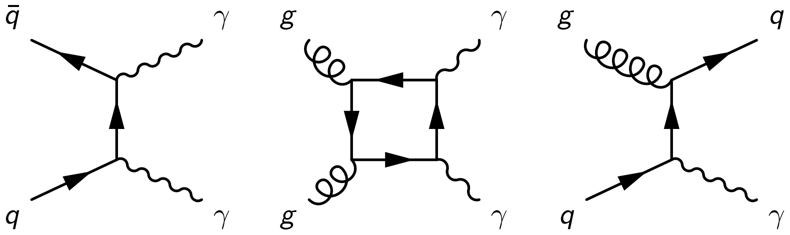


(b) Tracker Material Budget (Phase 0)

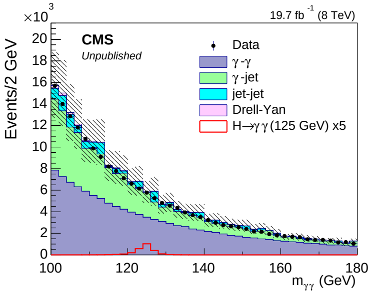
- ECal crystals lose and recover transparency under exposure to radiation
- Monitored in situ with LED/laser monitoring system, but still a major challenge for calibration
- Lots of material in front of the ECal

Higgs$\rightarrow \gamma\gamma$ Analysis Overview

- Higgs$\rightarrow$diphoton search at CMS simple in principle: Search for a small but narrow mass peak on a large, smoothly falling background
- Irreducible background from QCD di-photon production, reducible background from QCD γ+jets and multi-jet production with one or more jets faking a photon



Higgs→ $\gamma\gamma$ Analysis Overview



Inclusive selection with coarse binning

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1-\cos\theta_{12})}$$

- Standard Model search is carried out in inclusive, vector-boson-fusion tagged, W/Z, and $t\bar{t}$ associated production tagged channels
- Analysis makes extensive use of multivariate techniques to optimize the sensitivity, but basic principle of "bump hunt" is preserved

Higgs$\rightarrow \gamma\gamma$ Analysis Overview

1. Primary Vertex Selection (Vertex Selection MVA)
2. Photon Selection (Preselection + Photon-jet MVA discriminator)
3. **Multivariate Regression for EM Cluster corrections with per-photon resolution estimate**
4. Energy Scale and Resolution corrections from $Z \rightarrow ee$
5. **Event Categorization (MVA Discriminator)**
6. Signal modeling from Monte Carlo with smearing and scale factors applied
7. Background modeling from fit to data
8. Statistical Interpretation: Limits/Significance using maximum likelihood fit to $m_{\gamma\gamma}$ distribution in event categories categories

Multivariate Analysis Techniques

- Higgs search in this channel was statistically limited
- LHC data is valuable and finite
- Need to maximally exploit the large amount of information in each collision event
- Optimal discrimination between signal and background from full multidimensional log-likelihood ratio $L_R = \frac{\mathcal{L}_s(\bar{x})}{\mathcal{L}_s(\bar{x})+\mathcal{L}_b(\bar{x})}$
- Not known analytically in general, need to estimate from finite Data or Monte Carlo "training" samples
- Preferred tool for this analysis at the time of the Higgs discovery: Boosted Decision Trees (for both classification and regression)
- Intervening years have seen a significant move towards Deep Neural Networks given ongoing developments in industry, etc

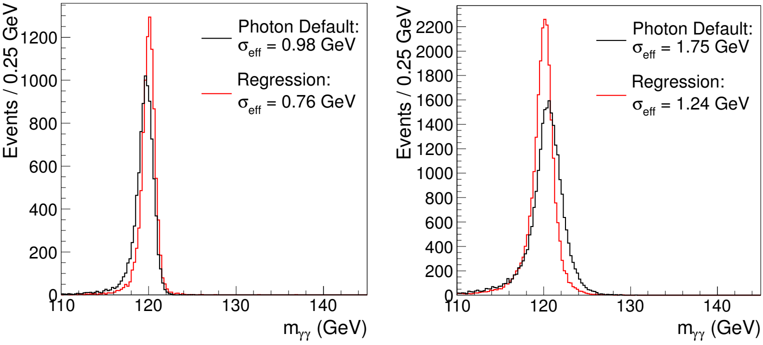
Regression Energy Corrections

- Photon energy reconstruction in CMS:

$$E_{e/\gamma} = F_{e,\gamma}(\bar{x}) \times \sum_{i}^{N_{crystals}} G(GeV/ADC) \times S_i(t) \times c_i \times A_i$$

- Two main components to photon energy resolution which at least partly factorize:
  1. Crystal level calibration (ADCtoGEV, Intercalibration, transparency corrections)
  2. Higher level reconstruction (Shower containment, crack/gap corrections, PU effects)
- Shower containment is complex and not clear if/how different contributions factorize
- Best performance is obtained with multivariate regression using BDT with cluster η, ϕ, shower shape variables, local coordinates, and number of primary vertices/median energy density as input
- Regression is trained on real photons in Monte Carlo, using the ratio of the generator level energy to the raw cluster energy, also provides a per photon estimate of the energy resolution

Regression Performance: Simulation

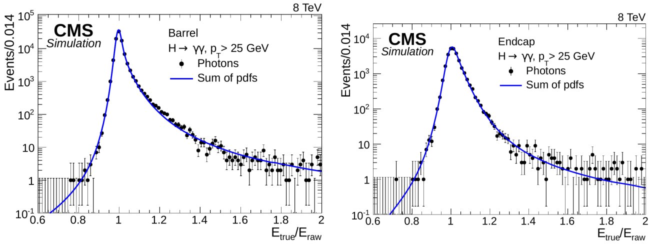


(a) Central Unconverted

(b) Central 1/2 Converted

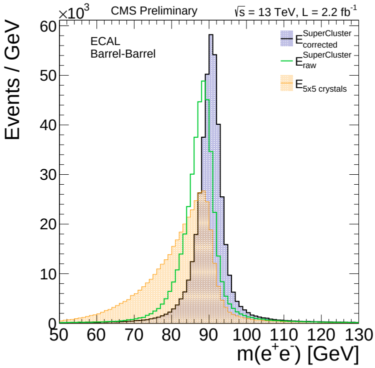
- Substantial improvement in diphoton mass resolution in simulation compared to simpler parameterized corrections (representative plots here)

Energy Regression: Predicted Response Distribution



- Semi-parametric regression provides a prediction for the full lineshape of the energy response **event-by-event**, given assumption of crystal-ball-like behaviour

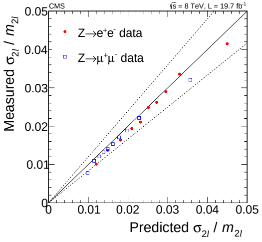
Energy Reconstruction: Data



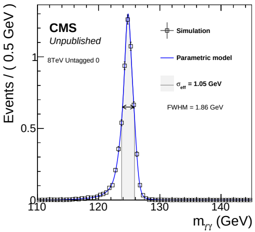
(a) Central (b) Forward

- Reconstructed Z mass in data with different levels of energy reconstruction and corrections
- Progression clearly visible even with 2.5 GeV natural Z width

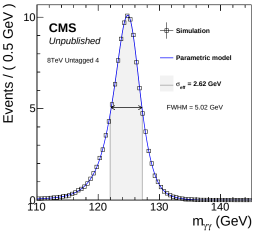
Per-photon Resolution Estimate



(c) Observed vs predicted σ_m



(d) $H \to \gamma\gamma$ Best Category



(e) $H \to \gamma\gamma$ Worst Category

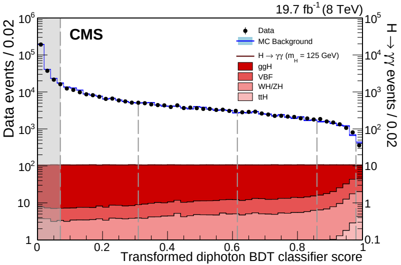
- In a resonance search, per-photon resolution estimate can be used to construct a per-event mass resolution estimate $\frac{\sigma_m}{m_{\gamma\gamma}} = \frac{1}{2}\sqrt{\frac{\sigma_{E1}^2}{E_1^2} + \frac{\sigma_{E2}^2}{E_2^2}}$
- Can be used to select or categorize events to make optimal use of highest resolution events (two unconverted photons in the center of the detector, incident on the center of the crystal, far from module boundaries)

Di-Photon MVA

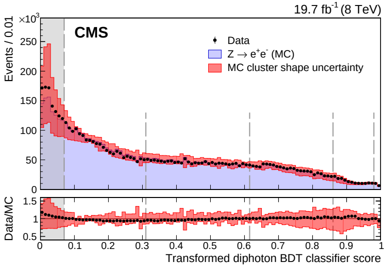
- Basic Strategy: Train di-photon MVA on Signal and Background MC with input variables which are to 1st order independent of $m_{\gamma\gamma}$
- Goal is to encode all relevant information on signal vs background discrimination (aside from $m_{\gamma\gamma}$ itself) into a single variable
- Can then simply categorize on Diphoton MVA output (5 categories, with cut values optimized against expected limit/significance using MC background, plus additional VBF/VH/ttH tagged categories with loose cut on di-photon MVA)
- Input variables cover kinematics (sans mass), per-event mass resolution and vertex probability, and photon ID

Di-Photon MVA Output

- Lowest score region not included in the analysis
- Diphoton MVA output for signal-like events can be validated with $Z \to ee$ events by inverting electron veto in the pre-selection
- Analysis does not rely on MVA shape of Monte Carlo background

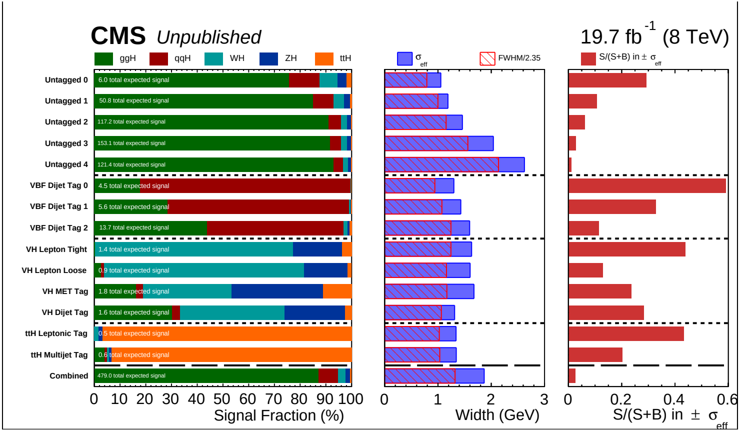


(a) Full Selection



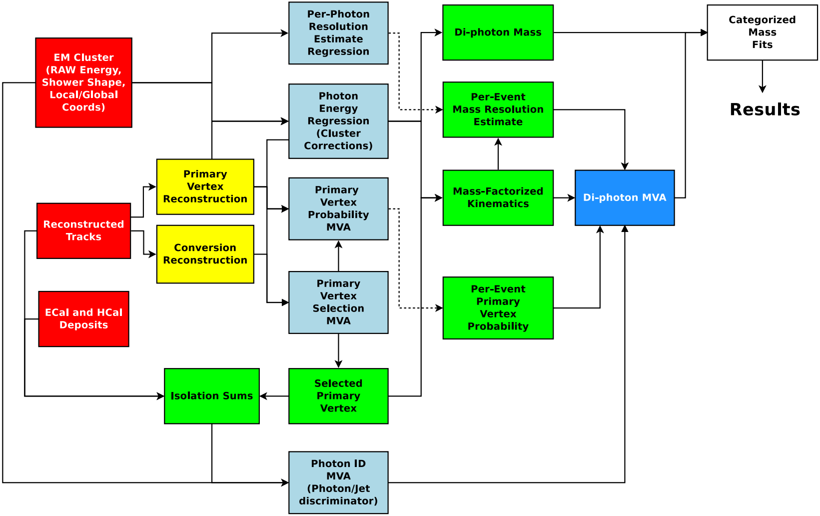
(b) Inverted e-Veto

Event Classification



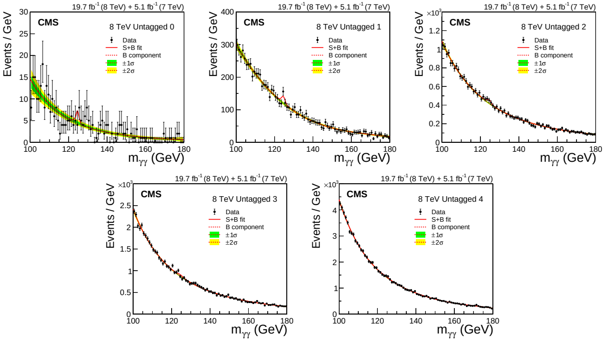
- Events classified according to di-photon MVA output plus tagging of additional objects
- Large variation in resolution and S/B across categories

Higgs$\rightarrow \gamma\gamma$: All Together



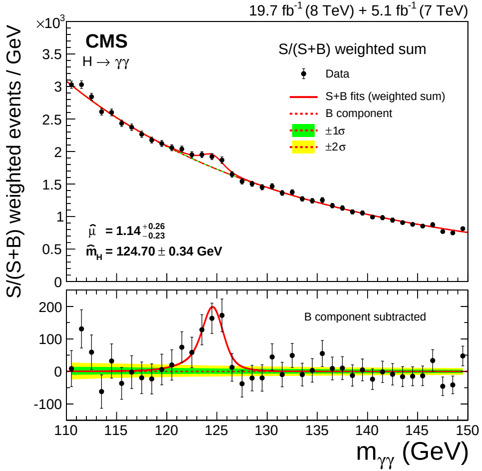
- Strategy: Process available information into quantities with straightforward physical interpretations in order to combine per-event knowledge of expected mass resolution and S/B into a single "Diphoton MVA" variable

S+B Fits - 8 TeV



- Plus 20 more distributions for exclusive-tagged modes and 7 TeV

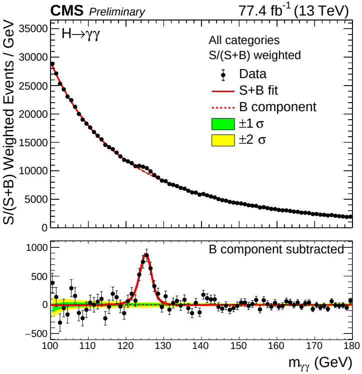
S+B Fit - Weighted Combination

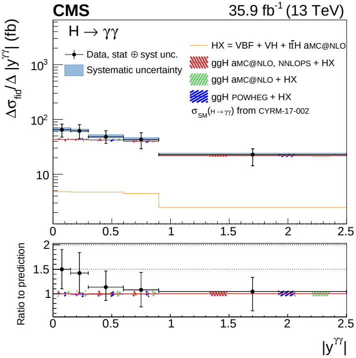


5.7σ observed significance (5.2σ expected)

- Results extracted from simultaneous fit to 25 event classes, but combined mass spectrum useful for visualisation
- Combination of all 25 event classes, weighted by $S/(S+B)$ for a $\pm\sigma_{eff}$ window in each event class
- Weights are normalised to preserve the fitted number of signal events

Current Run 2 $H \to \gamma\gamma$ Results

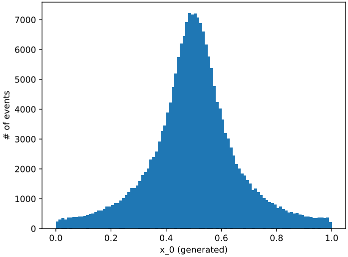




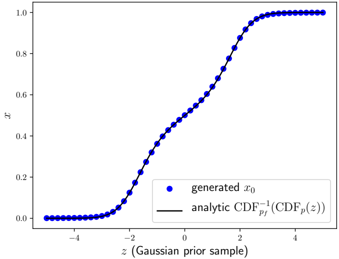
- Coupling analysis following similar methodology as in Run 1, with evolution to simplified template cross sections
- Differential xsec measurements become increasingly relevant with more data

Machine Learning Monte Carlo Integration

- Use machine learning to improve Monte Carlo integration efficiency in generators beyond what is achievable with VEGAS
- Generative Deep Neural Networks used to transform random noise (unit Gaussian) to target distribution, trained e.g. on KL divergence
- Can be thought of as learning empirically a multidimensional generalization of the inverse CDF



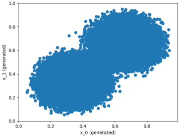
(a) Cauchy toy distribution



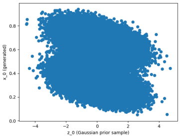
(b) DNN vs analytic inv. CDF

JB, arXiv:1707.00028

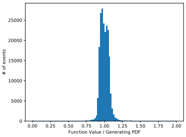
DNN 4D Camel Function Example



(a) Generated (2D Slice)



(b) Generated vs Prior (1D pair)



(c) Integration Weight

- 3x smaller weight variance to foam with 10x less function evaluations

Algorithm	# of Func. Evals	$\sigma_w / < w >$	σ_I / I (2e6 add. evts)
VEGAS	300,000	2.820	$\pm 2.0 \times 10^{-3}$
Foam	3,855,289	0.319	$\pm 2.3 \times 10^{-4}$
Generative DNN	300,000	0.082	$\pm 5.8 \times 10^{-5}$
Generative DNN	294,912	0.083	$\pm 5.9 \times 10^{-5}$
Generative DNN (staged)	294,912	0.030	$\pm 2.1 \times 10^{-5}$

Some results - 9D Camel Function Integration

- Comparing Vegas, GBRIntegrator, Generative DNN for 9-dimensional camel function

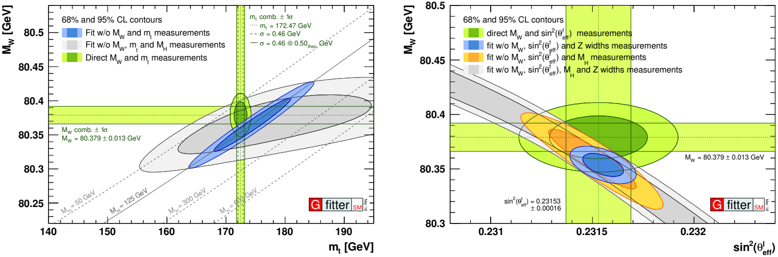
Algorithm	# of Func. Evals	$\sigma_w / < w >$	σ_I / I (2e6 add. evts)
VEGAS	1,500,000	19	$\pm 1.3 \times 10^{-2}$
GBRIntegrator	3,200,000	0.63	$\pm 4.5 \times 10^{-4}$
GBRIntegrator (staged)	3,200,000	0.31	$\pm 2.2 \times 10^{-4}$
Generative DNN	294,912	0.15	$\pm 1.1 \times 10^{-4}$
Generative DNN (staged)	294,912	0.081	$\pm 5.7 \times 10^{-5}$

- 50x smaller weight variance to Vegas with 2x function evaluations (BDT)
- DNN approach scales much better with dimensionality ($>$ 100x smaller weight variance than Vegas with 5x **fewer** function evaluations

Outlook: Machine Learning Monte Carlo Integration

- Large improvements with novel algorithms already demonstrated on toy cases
- Exploring alternative DNN architectures including auto-regressive and invertible models, along with further optimizations to training procedure
- Integration into Madgraph_aMC@NLO and tests with QCD matrix elements in progress
- Ultimate goal: significant increases in efficiency for phase space integration and event generation for complex/high multiplicity processes

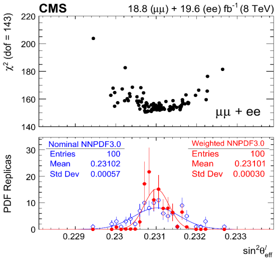
Electroweak Parameters

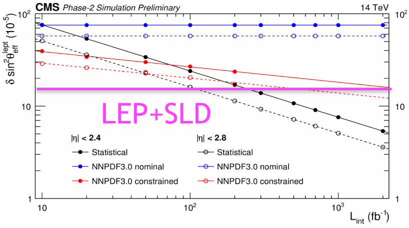


Eur. Phys. J. C78, 675 (2018)

- Precise measurements of the Higgs mass enable more precise consistency tests of the Standard Model using m_W and $\sin^2\theta_W$

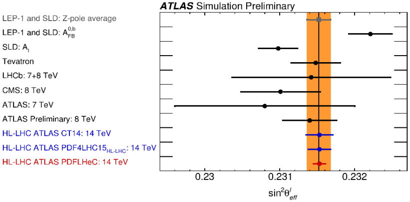
Weak Mixing Angle Prospects





Eur. Phys. J. C 78 (2018) 701, CMS-PAS-FTR-17-001,

ATL-PHYS-PUB-2018-037



- Existing measurements already reduce PDF uncertainties with in-situ constraint
- Measurements with full HL-LHC data can reach or surpass LEP+SLD precision, depending also on improved knowledge of PDFs from external sources

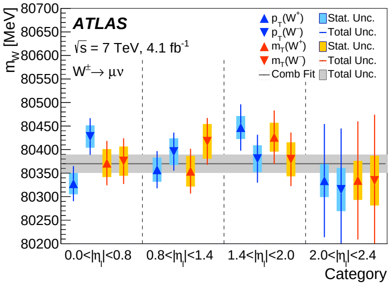
W mass: PDF Uncertainties

Eur. Phys. J. C 78 (2018) 110

$$m_W = 80370 \pm 7(\text{stat.}) \pm 11(\text{exp. syst}) \pm 14(\text{mod. syst.})\ \text{MeV}$$
$$m_W = 80370 \pm 7(\text{stat.}) \pm 11(\text{exp.}) \pm 8.3(\text{QCD})\ \pm 5.5(\text{EWK}) \pm 9.2(\text{PDF})\ \text{MeV}$$

	PDF Uncertainty (MeV)
per $|\eta|$-charge cat.	20-34
per-charge	14-15
full combination	9.2

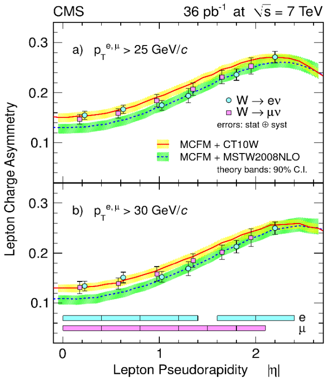
- PDFs determine the W rapidity spectrum and lepton decay angles through W polarization
- Well-defined correlations between phase space regions and processes which are already partly exploited in present measurement to reduce uncertainty
- Can be further exploited in the future



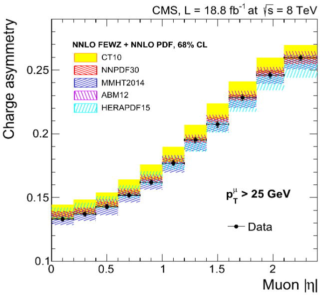
Ultimate Precision on Electroweak Parameters at LHC

- Ultimate precision on precision electroweak parameters at LHC may depend on in-situ constraints of PDFs to reduce associated uncertainty
- Must be handled with care with respect to experimental and other theoretical uncertainties
- Can result in e.g. technically challenging maximum likelihood fit
- WIP: Adapting modern tools (e.g. Tensorflow) and GPUs to binned maximum likelihood fits for e.g. W mass measurements $\rightarrow$ 100x speedup in likelihood+gradient evaluation and major improvement in numerical/minimization stability $\rightarrow$ **flexibility for more complex/accurate models of uncertainties together with PDF profiling**

W production at CMS



(a) 2010 Data



(b) 2012 Data

- W mass measurement with in-situ PDF constraints closely related to differential cross section and charge asymmetry data

Conclusions

- Rich program of measurements of electroweak gauge bosons and Higgs at the LHC through the HL-LHC era
- Precision tests of the Standard Model have the potential to reveal hints of new physics
- Differential cross section measurements also act as searches in kinematic tails
- Continued development of detectors and analysis/reconstruction techniques essential to fully exploiting LHC/HL-LHC and future machines