Experimental particle physics at the LHC (3)

Kerstin Tackmann (DESY)

GRK1504/2: Autumn Block Course 2016

Photon pointing and primary vertex selection (I)

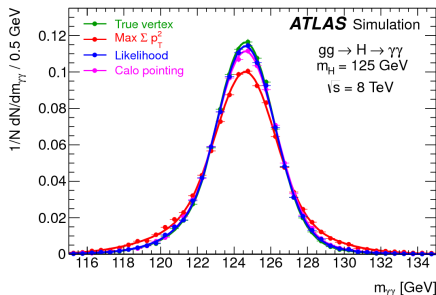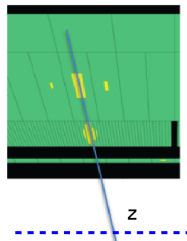$$m_{\gamma\gamma}^2 = 2E_1 E_2 (1 - \cos\alpha)$$

Improve photon angle measurement using neural network (8 TeV) based on

- Photon pointing
  - ★ Photon direction measured from calorimeter using longitudinal segmentation
  - ★ Position of conversion vertex for converted photons (with Si hits)
- $\sum p_T^2$, $\sum p_T$ (over tracks) and angular balance in ϕ between tracks and diphoton system
- → Contribution of angle measurement to mass resolution negligible already without primary vertex information
- → Good primary vertex selection needed for selection of signal jets

Photon pointing and primary vertex selection (I)
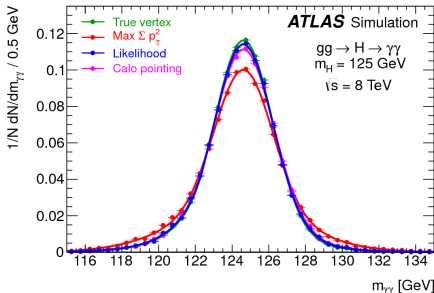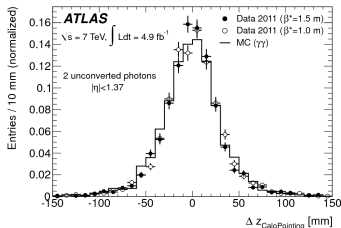
$$m_{\gamma\gamma}^2 = 2E_1 E_2 (1 - \cos \alpha)$$

Improve photon angle measurement using neural network (8 TeV) based on

- Photon pointing
  - ★ Photon direction measured from calorimeter using longitudinal segmentation
  - ★ Position of conversion vertex for converted photons (with Si hits)
- $\sum p_T^2$, $\sum p_T$ (over tracks) and angular balance in ϕ between tracks and diphoton system
- → Contribution of angle measurement to mass resolution negligible already without primary vertex information
- → Good primary vertex selection needed for selection of signal jets

$$m_{\gamma\gamma}^2 = 2E_1 E_2 (1 - \cos\alpha)$$

Improve photon angle measurement using neural network ($8\,\mathrm{TeV}$) based on

- Photon pointing
  - ★ Photon direction measured from calorimeter using longitudinal segmentation
  - ★ Position of conversion vertex for converted photons (with Si hits)
- $\sum p_T^2$, $\sum p_T$ (over tracks) and angular balance in ϕ between tracks and diphoton system
- → Contribution of angle measurement to mass resolution negligible already without primary vertex information
- → Good primary vertex selection needed for selection of signal jets
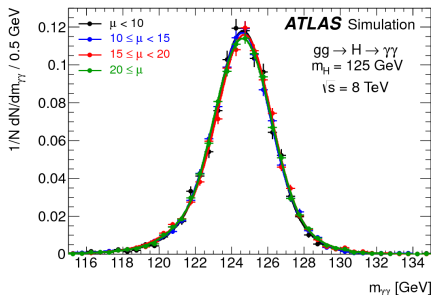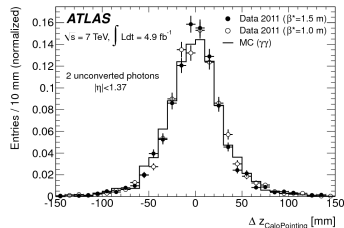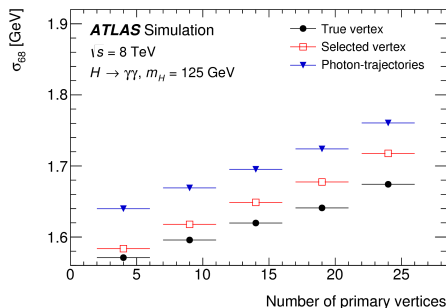
Photon pointing and primary vertex selection (II)

- Efficiency of primary vertex selection can be measured with $Z \to ee$, disregarding the electron tracks

- Primary vertex selection more efficient in lower pileup
  - ★ Fewer primary vertices with larger average spacing

- Primary vertex selection more efficiency for events with higher Higgs p_T
  - ★ More recoil track p_T

- Efficiency of primary vertex selection can be measured with $Z \to ee$, disregarding the electron tracks

- Primary vertex selection more efficient in lower pileup
  - ★ Fewer primary vertices with larger average spacing

- Primary vertex selection more efficiency for events with higher Higgs p_T
  - ★ More recoil track p_T
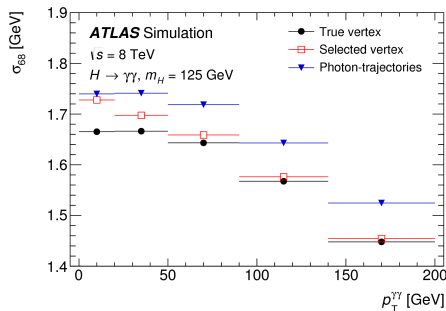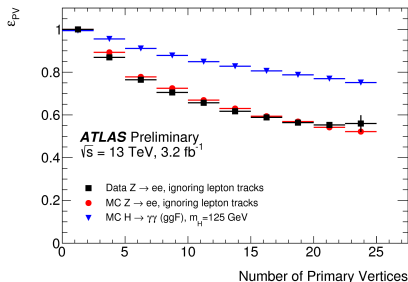
Invariant mass resolution – CMS vs ATLAS

Calorimeter resolution

- CMS crystal calorimeter with excellent intrinsic resolution

  $$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{0.12}{E} \oplus 0.3\%$$
  vs ATLAS

  $$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$$

$\Rightarrow$ Narrower core of resolution function in CMS compared to ATLAS, e.g. best resolution event category
  - ★ CMS 1.18 GeV
  - ★ ATLAS 1.39 GeV

Primary vertex selection

- ATLAS longitudinally segmented calorimeter allows for pileup-independent input to primary vertex selection
  - ★ CMS primary vertex selection relies entirely on tracker
- ATLAS resolution function less affected by long non-Gaussian tails arising from wrong primary vertex choice
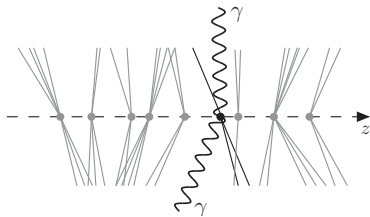
Invariant mass spectrum
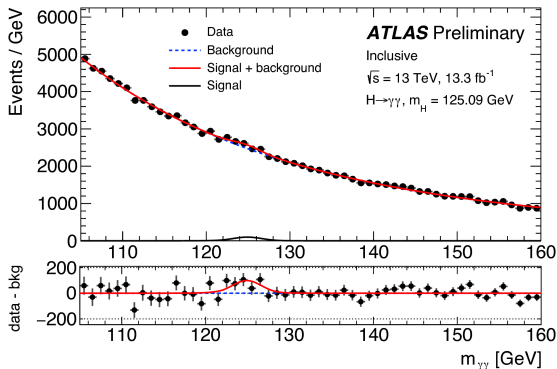
Diphoton selection

Identified and isolated photons
$p_T^{\gamma 1} > 0.35 m_{\gamma\gamma}$
$p_T^{\gamma 2} > 0.25 m_{\gamma\gamma}$

Background+signal fit, signal constrained to
125.09 GeV

CMS took some data with $B = 0\,\mathrm{T}$ in 2015.

How are the various steps affected?

Note: this data was not used in any $H \to \gamma\gamma$ analysis, but was used in the search for high-mass diphoton resonances with 2015 data

Specifics for CMS 0 T dataset

No energy spread from bremsstrahlung

- Dedicated calibration and photon identification
  - ⋆ 5-15% lower id efficiency
- Better intrinsic energy resolution
- Better representation of photons by electrons

No measurement of track momenta

- Track isolation relying on track counting instead of track momenta
- Primary vertex selection based on track counting (largest multiplicity)
  - ⋆ Reduced efficiency for identifying correct PV of 60% (instead of 90%)
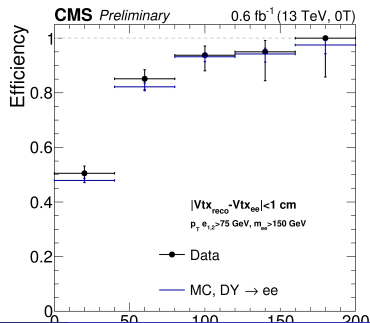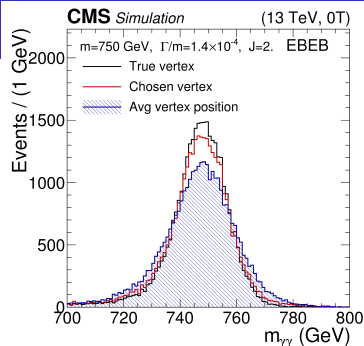
I will discuss two measurements in the $H \to \gamma\gamma$ decay channel in detail:

- Fiducial and differential cross section measurements
- "Coupling measurements": studies of the Higgs production processed

- I am showing plots from several different analyses, using different data sets, depending on which plot is available for which analysis/data set
  - ⋆ General principle is the same (details and results are of course different)

Cross section measurements

Principle of a cross section measurement

Recall

$$\sigma = \frac{N}{\int \mathcal{L} \mathrm{d}t} = \frac{N_{\mathrm{meas}} - N_{\mathrm{bkgd}}}{\epsilon \cdot A \cdot \mathcal{B} \cdot \int \mathcal{L} \mathrm{d}t}$$

Experimental steps

- Estimate and subtract the background(s)
- Correct for detector acceptance, and for efficiencies
- If needed/wanted, correct for branching ratio(s)
- Determine the luminosity (see earlier lecture)

Differential cross section in variable x: $\frac{\mathrm{d}\sigma}{\mathrm{d}x}$

- In practice: bin-averaged cross section $\frac{\Delta\sigma}{\Delta x}$
- Background estimation and subtraction, efficiency and acceptance corrections performed for every bin
- Requires correction of resolution effects in x: unfolding

Why cross section measurements for Higgs?

- Almost model-independent measurements of production and decay kinematics

- Measure kinematic distributions of Higgs, of associated jets, ...

- Sensitivity to Higgs production processes, QCD effects, CP, ...

- Measure inclusive cross section, and cross section in phase space enriched with VBF, and with a lepton

- Differentially in $p_T^{\gamma\gamma}$, N_{jet}, p_T^{jet}, ...

- $H \to \gamma\gamma$ and $H \to 4\ell$ decays well suited thanks to good signal invariant mass resolution $\to$ comparably "simple" analyses
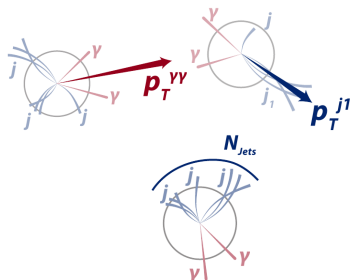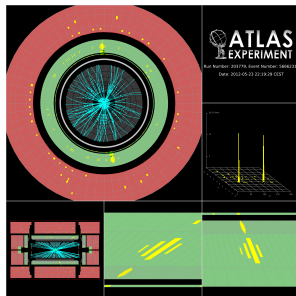
Backgrounds

- Irreducible backgrounds: events with two photons, e.g.
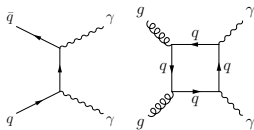
- Reducible backgrounds: events where at least one photon candidate is a misidentified jet, e.g.
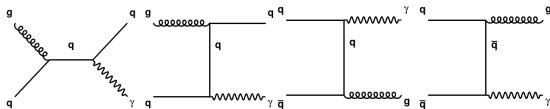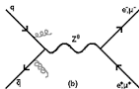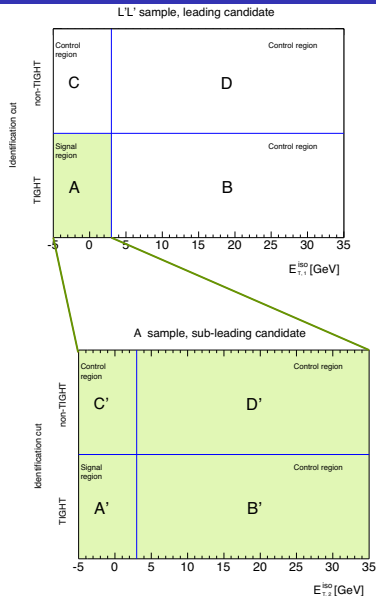

- $Z \to ee$ with the electrons misreconstructed as photons (mass tail reaches beyond $m_Z = 90\,\mathrm{GeV}$)

Understanding the backgrounds (I)

- Define control regions enriched in background
  - ★ Photon candidates that fail a given set of the shower shape cuts and/or
  - ★ Photon candidated that are less isolated

- Fit determines (given numbers of events in signal and control regions and photon identification and isolation efficiency)
  - ★ Efficiencies for jet to pass photon identification and isolation for γjet and jetjet events, separately for higher and lower p_T candidate
    - ▶ Correlation for both jets to pass isolation in jetjet events
  - ★ Number of $\gamma\gamma$, γjet, jetγ and jetjet events
    - ▶ $Z \to ee$ included in $\gamma\gamma$ as e look most like γ in id and isolation

L'L' sample, leading candidate

Control region — C

Control region — D

Signal region — A

B

$E_{T,1}^{iso}$ [GeV]

A sample, sub-leading candidate

Control region — C'

Control region — D'

Signal region — A'

B'

$E_{T,2}^{iso}$ [GeV]

Understanding the backgrounds (II)
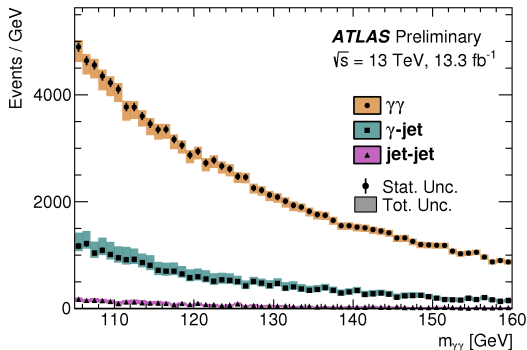
$\gamma\gamma$ $\quad (78.9 \pm 0.2^{+1.9}_{-4.0})\%$
γjet $\quad (18.6 \pm 0.2^{+3.5}_{-1.7})\%$
jetjet $\quad (2.5 \pm 0.1^{+0.5}_{-0.4})\%$

Largest uncertainty: definition of control regions

- Study performed in every bin and every measured region of phase space
- Understanding of background composition not important directly to derive results, but for studies of background parametrization and photon identification

Understanding the backgrounds (III)

- Misreconstruction of electrons usually as single-track converted photons

- Suppressed by requiring no hit in innermost silicon layer for single-track conversions

- Probability for an electron to be reconstructed as photon estimated using $Z \to ee$ events
  - Select "$Z \to e\gamma$" and $Z \to ee$

- For $H \to \gamma\gamma$, only $Z \to ee$ tail relevant

- Not considered separately

Understanding the backgrounds (III)

- Misreconstruction of electrons usually as single-track converted photons

- Suppressed by requiring no hit in innermost silicon layer for single-track conversions

- Probability for an electron to be reconstructed as photon estimated using $Z \to ee$ events
  - ⋆ Select "$Z \to e\gamma$" and $Z \to ee$

- For $H \to \gamma\gamma$, only $Z \to ee$ tail relevant
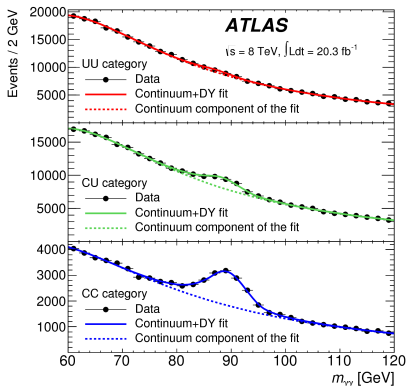
- Not considered separately
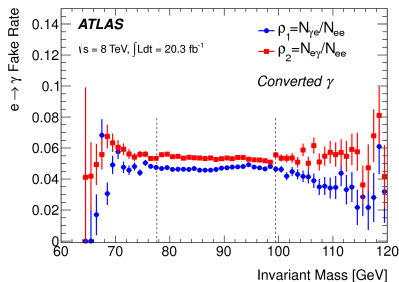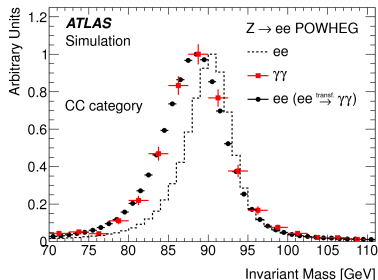
Parametrizing the signal

- Signal is extracted by a signal+background fit to $m_{\gamma\gamma}$ spectrum

- Signal is parametrized by a double-sided Crystal Ball function
  - ⋆ Gaussian with exponential tail
- SM Higgs width $4\,\mathrm{MeV}$ $(m_H = 125\,\mathrm{GeV})$
- Parameters that determine the shape are determined on simulation
- Peak position (= Higgs mass) and Gaussian width (= detector resolution) constrained within uncertainties
  - ⋆ Energy scale and resolution, and m_H
  - ⋆ Peak position unconstrained for measurement of the Higgs mass
    - ▶ To be done with Run2 data once precision energy calibration achieved
  - ⋆ Run1 Higgs mass measurement $m_H = (125.09 \pm 0.21(\mathrm{stat}) \pm 0.11(\mathrm{syst}))\,\mathrm{GeV}$
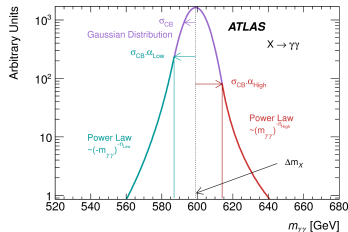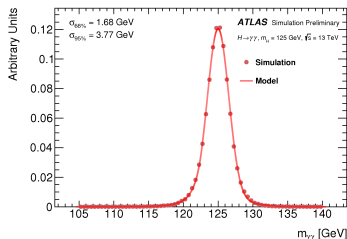
Parametrizing the backgrounds (I)
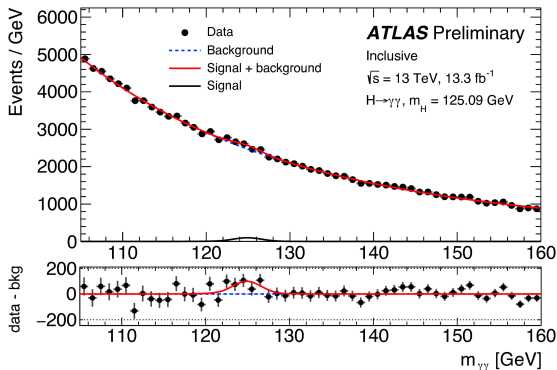
Background+signal fit, signal constrained to 125.09 GeV

Background modelled by smooth, monotonously falling function

- Polynomials (typically 3rd or 4th order)
- Exponentials of polynomials (typically 1st or 2nd order)

shape and normalization determined by the fit

Studied on high-statistics MC and chosen to give good statistical power while keeping potential biases acceptable

Potential bias accounted for as systematic uncertainty

Parametrizing the backgrounds (II)

- S+B fits to high statistics background samples ($\rightarrow$ "spurious signal")
- Choose the most performant parametrization that has small enough systematics:
  1. Accept model if the spurious signal is $< 10\%$ of expected signal or $< 20\%$ of fitted signal uncertainty
  2. Among the models left, choose the one with the smallest number of free parameters
- Run1: $\gamma\gamma$, γjet, jetjet simulation, detector effects included through weighting and smearing
- Run2: γjet, jetjet shapes from data control regions
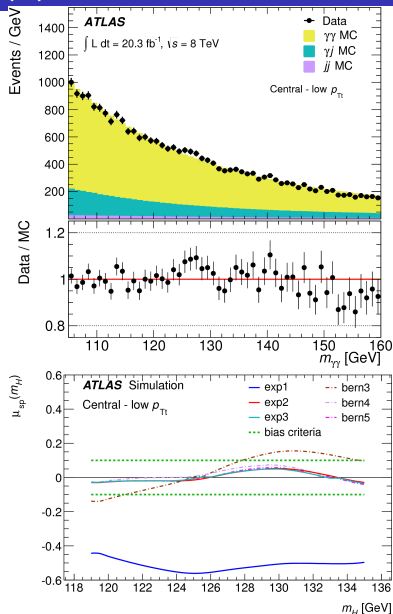
Signal+background fit

...carried out for...

- ...all selected events → fiducial cross section
- ..after specific selections → fiducial cross section for that selection
- ...in bins of a given variable → differential spectrum
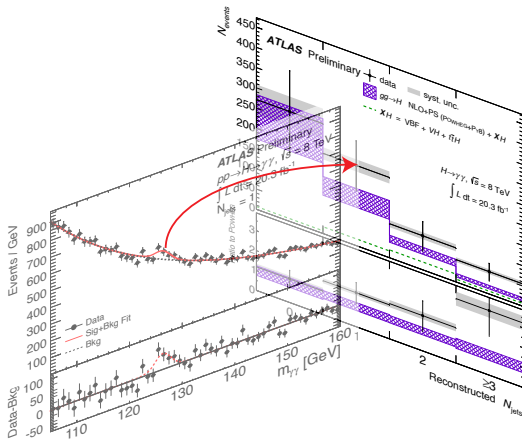
Signal+background fit

Likelihood function to be maximized

$$\mathcal{L} = \prod_i \left\{ \frac{e^{-\nu_i}}{n_i!} \prod_j^{n_i} \left[\nu_i^{\mathrm{sig}} \, \mathcal{F}_i^{\mathrm{sig}}(m_{\gamma\gamma}^j, \theta; m_H) + \nu_i^{\mathrm{bkg}} \, \mathcal{F}_i^{\mathrm{bkg}}(m_{\gamma\gamma}^j) \right] \right\} \times \prod_l G_l(\theta)$$

- for bin i and event j
- n_i number of events in bin i
- $\nu_i^{(\mathrm{sig, bkg})}$ expected number of total/signal/background events
- $\mathcal{F}_i^{(\mathrm{sig, bkg})}$ signal/background shape
- θ nuisance parameters associated with systematic uncertainties, constraint via $G_l(\theta)$
- Energy scale and resolution uncertainties, and uncertainty on m_H correlated between all bins
  - $\rightarrow$ Nuisance parameters common between all bins

But wait a moment... is there a signal?

...back to summer 2012

★ Signal or statistical fluctuation of the background?

Compare compatibility of data with B-only and with S+B hypothesis with a signal scaling factor μ

Profile likelihood ratio

$$\tilde{q}_\mu = -2 \ln \frac{L(\text{data}|\mu, \hat{\hat{\theta}}_\mu)}{L(\text{data}|\hat{\mu}, \hat{\theta})}$$

- Numerator and denominator are maximized independently
- $\hat{\hat{\theta}}_\mu$ conditional maximum given μ; $\hat{\mu}, \hat{\theta}$ corresponding to global maximum of the likelihood
- Large $\tilde{q}_\mu$ correspond to disagreement between data and hypothesis μ
- $\tilde{q}_\mu$ behaves as χ^2 for large data samples and Gaussian θ
- Denominator is only normalization term, independent of μ
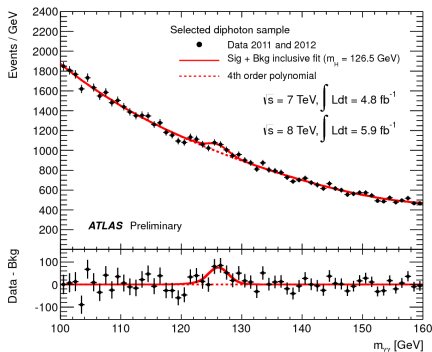
Frequentist limit setting procedure

- Construct likelihood function $L(\mu, \theta)$

- Construct test statistics $\tilde{q}_\mu$

- Perform fit to data and determine observed $\tilde{q}_{\mu,\mathrm{obs}}$ for hypothesis μ

- Generate pseudo MC to construct PDF $p_\mu(\tilde{q}_\mu | \mu, \hat{\theta}_{\mu,\mathrm{obs}})$ of $\tilde{q}_\mu$
  - ⋆ MC generation done with $\hat{\theta}_{\mu,\mathrm{obs}}$, but $\hat{\theta}_\mu$ allowed to float in the fits

- Determine the observed p-value for hypothesis μ:
  $P(\mu) = \int_{\tilde{q}_{\mu,\mathrm{obs}}}^{\infty} p_\mu(\tilde{q}_\mu | \mu, \hat{\theta}_{\mu,\mathrm{obs}}) \mathrm{d}\tilde{q}_\mu$

- Perform "discovery" test by computing $P(\mu = 0)$

- Find the 95% upper bound $\mu = \mu_{95,\mathrm{obs}}$ for which $P(\mu) = 0.05$
  - ⋆ To be conservative and to avoid that upward fluctuations of the background contribute to the p-value, LHC experiments compute upper limit from
    $P_{\mathrm{CL}_s}(\mu) = P(\mu)/P(0) = 0.05$
    - ▶ CL_s usually over-covers, so less than 5% of repeated experiments would lie outside the given bound
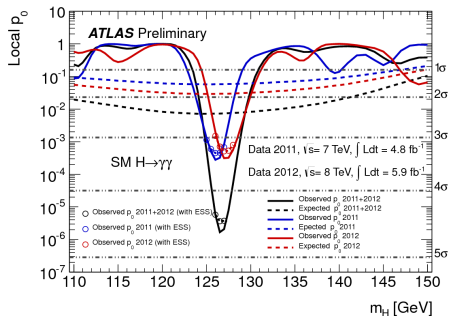
For complex fits pseudo-MC procedure can be very CPU intensive. Asymptotic formulae exist for cases with enough events.

Testing background-only for $H \to \gamma\gamma$ ICHEP 2012

Maximum deviation from background-only expectation at $m_H = 126.5 \, \text{GeV}$

- Local significance $4.5 \, \sigma$ (expected $2.4 \, \sigma$)
- Global significance $3.6 \, \sigma$
- Need to take into account "look-elsewhere effect": probability for a fluctuation somewhere in the studied mass range larger than for a given mass

- Require $5 \, \sigma$ for discovery ($p = 2.9 \cdot 10^{-7}$)
  - ⋆ Reached at ICHEP in combination with $H \to ZZ^* \to 4\ell$
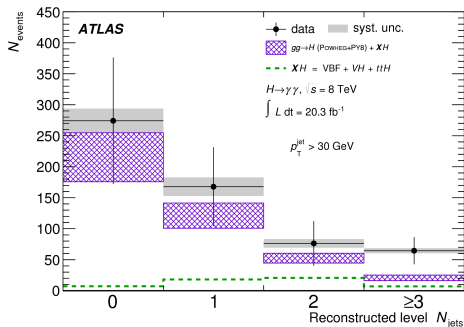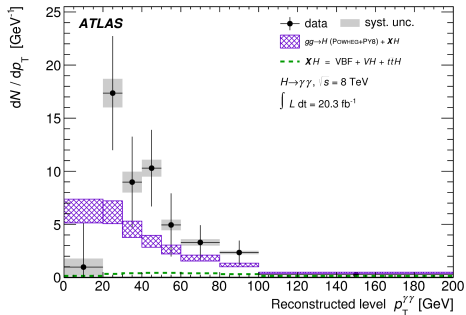
- First part achieved:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x} = \frac{N_{\mathrm{meas}} - N_{\mathrm{bkgd}}}{\epsilon \cdot A \cdot \mathcal{B} \cdot \mathrm{d}x \cdot \int \mathcal{L}\mathrm{d}t}$$

- ..although not quite...

Not mentioned so far: resolution corrections/unfolding

- $A_{ij}x_j = b_i$ (b – measured, x – true)
- Detector response matrix A encodes resolution (can also include efficiency and acceptance)
  - $\star$ A_{ij} = Probability for event in true bin j to be reconstructed in reco bin i
  - $\star$ A_{ij} is largely model independent, although there could be caveats in some cases

- "Naive" matrix inversion: $x = A^{-1}b$
  - $\star$ Unfolded spectrum x usually dominated by statistical fluctuations
    - ▶ Statistical fluctuations in measured spectrum get amplified
    - ▶ Nice explanation of this effect here
  - $\star$ Unbiased estimator with smallest possible variance (typically see large negative correlations between adjacent bins)

Nucl.Instrum.Meth. A372 (1996)
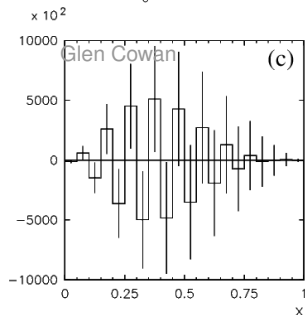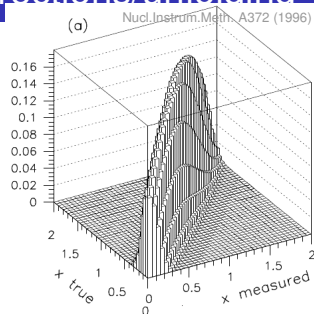
Glen Cowan

Unfolding

Unfolding is a complicated business and one is well advised to ask in each problem if it can be avoided. (Glen Cowan)

Link to a (very readable) survey of unfolding methods from Glen Cowan here

Bin-by-bin correction (the simplest method)

- Used for $H \to \gamma\gamma$ differential cross section measurements
- Each bin is corrected with multiplicative factor, $x_i = b_i/c_i$, with:

$$c_i = \frac{N_i^{\text{reco}}}{N_i^{\text{true}}} \quad \epsilon_i = \frac{N_i^{\text{true+reco}}}{N_i^{\text{true}}} \quad p_i = \frac{N_i^{\text{true+reco}}}{N_i^{\text{reco}}}$$

with ϵ_i efficiency and p_i purity

- Pro: simple and robust
- Con: prone to produce biased results
  - ⋆ Data biased towards the MC used to compute the c_i
  - → Carefully evaluate uncertainties!
- My personal take: ok to be used as long as biases are estimated properly and small compared to e.g. statistical uncertainties

MC2 corrected with MC1

- BBB corrected MC2
- MC2 truth
- MC1 truth

Stefan Schmitt

r_{BBB}
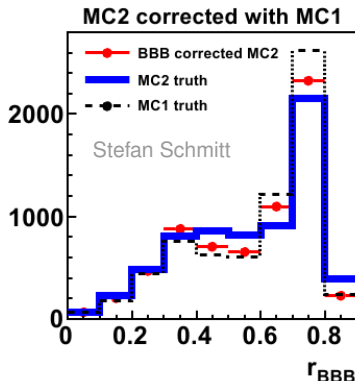
Regularization (I)

- Regularization dampens the oscillations by suppressing insignificant bins in the data distribution and detector response matrix

- In simplified form, can write unfolding problem as minimization of

  $$\chi^2(x^{\mathrm{data}}) = (A^{\mathrm{MC}}x^{\mathrm{data}} - b^{\mathrm{data}})^T(A^{\mathrm{MC}}x^{\mathrm{data}} - b^{\mathrm{data}}) + \tau(Cx^{\mathrm{data}})^T(Cx^{\mathrm{data}})$$

  where x^{data} the unfolded data distribution, C is a matrix and Cx^{data} e.g. the sum of squared of the 2nd derivative of x^{data} (other choices possible)

- $\tau = 0$ corresponds to exact matrix inversion, $\tau \neq 0$ regularizes inversion
  - $\star$ τ too small $\rightarrow$ oscillations
  - $\star$ τ too large $\rightarrow$ unfolded spectrum biased towards MC used for unfolding
  - $\star$ Good size of τ depends on statistics, binning, resolution, how well the MC describes the data, ... and needs to be carefully determined for every case

Regularization (II)

Toy example
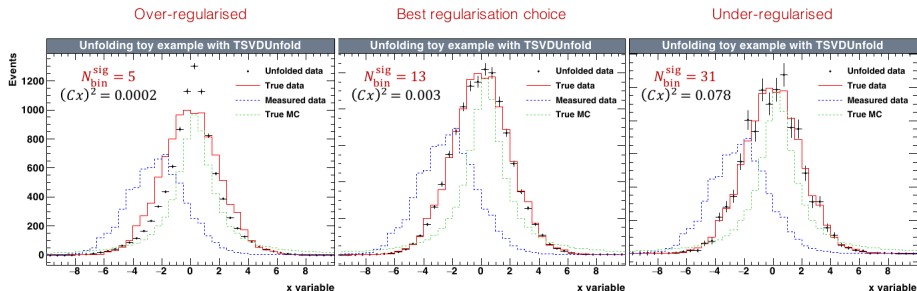
- $\tau = 0$ corresponds to exact matrix inversion, $\tau \neq 0$ regularizes inversion
  - ⋆ τ too small $\rightarrow$ oscillations
  - ⋆ τ too large $\rightarrow$ unfolded spectrum biased towards MC used for unfolding
  - ⋆ Good size of τ depends on statistics, binning, resolution, how well the MC describes the data, ... and needs to be carefully determined for every case

Fit-based and SVD unfolding

- Solve unfolding problem by minimizing χ^2
- SVD: Based on singular value decomposition of detector response matrix (here)
  - ⋆ Allows association of "physics" with large singular value contributions and "statistical fluctuations" with small singular value contributions
  - ⋆ Regularization through suppression of small singular value constributions ("beyond kth singular value")

- Pro: in principle always possible to find decent working point (might require adaptation of discrete k choice)
- Con: curvature regularization iffy for spectra with sharp features (e.g. narrow peaks)

Iterative unfolding

Iterative Bayesian unfolding (D'Agostini's method, here) is widely used

- Employs Bayes theorem to infer the unfolding matrix

$$P(\text{true bin } j|\text{reco bin } i) \propto P(\text{reco bin } i|\text{true bin } j)P(\text{true bin } j)$$

and iteratively determines

$$x_j = \frac{1}{\varepsilon_j} \sum_{i=1}^{N} P(\text{true bin } j|\text{reco bin } i)b_i$$

starting with an initial estimate for $P(\text{true bin } j)$ (e.g. from MC)

- Effectively regularized by using a finite number of iterations
  - ⋆ Small number of iterations → biased towards input MC
  - ⋆ Large number of iterations → large variance

- Pros: able to handle high-dimensional problems (e.g. picture deblurring), used in SM for 3d unfolding, some adaptation of "MC" to data (if enough iterations possible)

- Cons: iteration might not be possible with low statistics (seen in $H \to \gamma\gamma$), can be problematic if there is a badly modelled quantity that is not measured but averaged over (no help from iteration)

Back to the analysis: efficiency corrections

- Efficiency of the reconstruction and selection

$$\epsilon = \frac{\text{Number of events reconstructed and selected}}{\text{Number of signal events in the kinematic range}}$$

Main contributions to inefficiencies in $H \to \gamma\gamma$

- photon identification
- photon isolation
- diphoton trigger

Efficiencies are measured in control samples

- Sometimes, efficiencies are determined from simulations
  - ⋆ Requires good simulation of detector and/or physics process

Photon id efficiency measurements

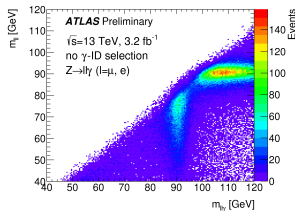Id efficiency for isolated photons: $E_T^{iso} <$ 4 GeV

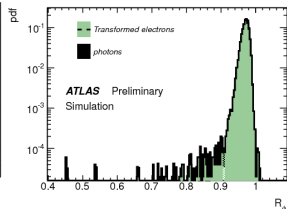Radiative Z decays:
$Z \to \ell\ell\gamma$

E_T^γ of 10-80 GeV
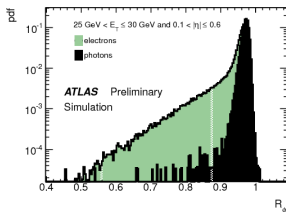Photon purity

- $\sim 90\%$ (10-15 GeV)
- $\geq 98\%$ (> 15 GeV)
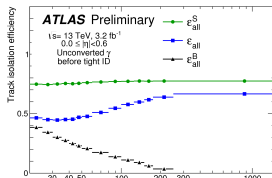

$Z \to ee$ tag-and-probe

+ transformation of electron showers to resemble photon showers

"Matrix method"

Purity determination from track isolation before and after id → id efficiency
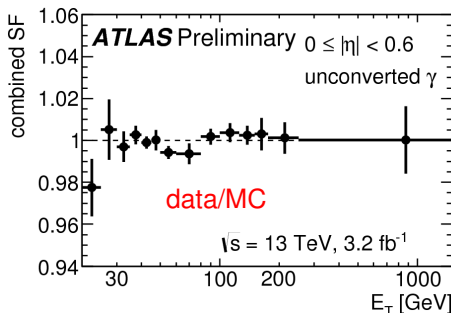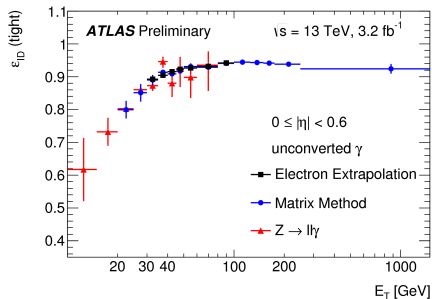
Photon id efficiency measurements

- Partial overlap in E_T regions covered by the different methods
→ Combination of measurements in overlap regions
- 1-2% uncertainties for $E_T <$ 40 GeV, 0.5-1% above 40 GeV

Uncertainty on $H \to \gamma\gamma$ signal yield

ICHEP 2012	10.8%
Dec 2012	5.3%
Moriond 2013	2.4%
ICHEP 2014	1%

Second-largest experimental uncertainty on $H \to \gamma\gamma$ signal strength (final Run1 paper)

Acceptance corrections (I)

- Acceptance of the kinematic selection

$$A = \frac{\text{Number of signal events in the kinematic range}}{\text{Number of all signal events}}$$

- Experimentally accessible kinematic region is limited
  - ⋆ Small E_T photons not used due to large backgrounds
  - ⋆ Detector acceptance limited in η
- Need to use theoretical predictions to extrapolate
  - ⋆ Usually in the form of simulations
  - ⋆ Introduced dependence on theoretical predictions and their uncertainties
- Unfold to a fiducial region defined by photons (and jets) to minimize acceptance corrections

  - ⋆ $p_T^{\gamma 1(\gamma 2)} > 0.35\ (0.25)\ m_{\gamma\gamma}, \quad |\eta^{\gamma 1,2}| < 2.37$
  - ⋆ $p_T^{\mathrm{iso}} < 0.05\ p_T^{\gamma}$ with $p_T^{\mathrm{iso}} \sum p_T$ of all charged particles with $p_T > 1$ GeV within $\Delta R = 0.2$ around photon
  - ⋆ $p_T^j > 30$ GeV, $\qquad\qquad |y^j| < 4.4$

Acceptance corrections (II)

Correcting from fiducial region to the full phase space would be a sizeable correction
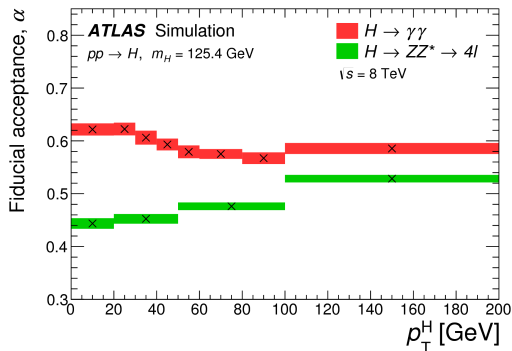
- ...of course this means that theoretical predictions will have to be done for the same fiducial region
- where not available (yet), correction factors are derived from simulation