# ATLAS $H \rightarrow \gamma\gamma$ Differential Cross Section Measurements

-- with a special focus on the background modelling --

Nils Gillwald

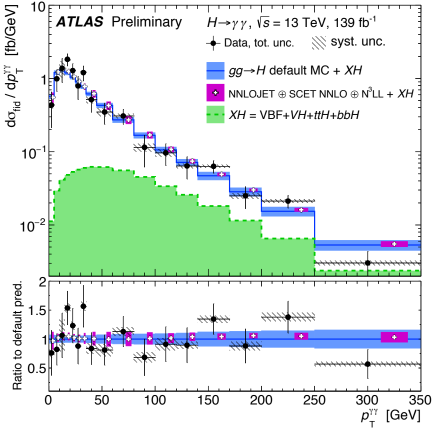
November 23, 2021

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

European Research Council
Established by the European Commission

# Why Measure Differential Cross Sections?

Source: ATLAS-CONF-2019-029




- Selected differential measurements sensitive to many Higgs boson properties:
  - $p_T^{\gamma\gamma} \rightarrow$ ggF & perturbative QCD, Yukawa & BSM couplings
  - $|y_{\gamma\gamma}| \rightarrow$ ggF & perturbative QCD, proton PDFs
  - $\Delta\phi_{jj} \rightarrow$ Higgs boson spin & CP properties
  - …
- Better measurements constrain unknown physics!
- Measurements are almost model independent $\rightarrow$ allow direct comparison to predictions

# Analysis Overview

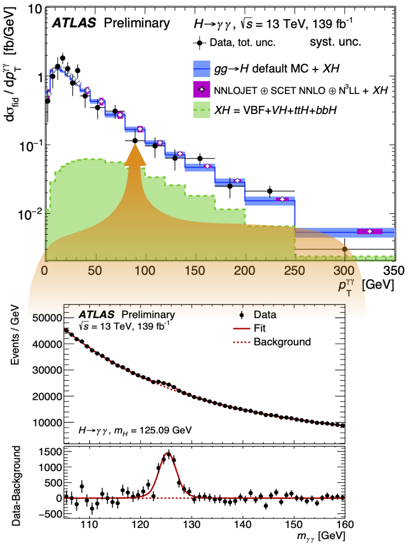
## Fiducial and differential cross sections in the $H \to \gamma\gamma$ channel

- $\gamma\gamma$ excellent channel to work with
  - High photon selection efficiency
  - Excellent $m_{\gamma\gamma}$ resolution $\to$ robust background subtraction
  - Balances low BR of 0.23%!
- Conf using full Run-2 dataset, 139.0 fb$^{-1}$ at $\sqrt{s} = 13$ TeV
- Total cross section measured in inclusive fiducial region
- Differential measurement of six variables:
  - $p_T^{\gamma\gamma}$, $|y_{\gamma\gamma}|$, $N_{\text{jets}}$, $p_T^{j1}$, $m_{jj}$, $\Delta\phi_{jj}$
- Two interpretations:
  - $b$ and $c$ Yukawa couplings
  - SMEFT

# Analysis Strategy

Source: ATLAS-CONF-2019-029

- Measure differential cross section by simultaneously fitting the $m_{\gamma\gamma}$ distributions in all bins of a variable
  - $\rightarrow$ Get signal $\sigma \times Br$ and background yield for every bin
- Fit uses analytical $S$ and $B$ shape parametrizations
  - Need to choose a background function for each bin $\rightarrow$ Background modelling
- Systematic uncertainties shared between all bins of a variable
- Unfold detector effects for easier cross-experiment & theory comparison

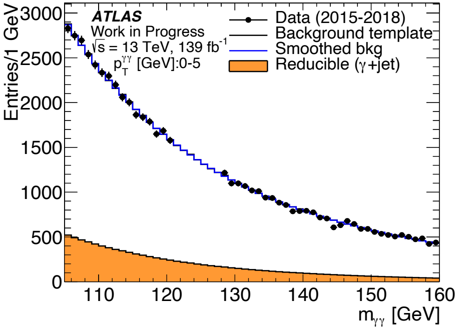


# Background Model

## Strategy

- Parametrize total background component with simple analytical function
  - But which functional form should be used?

→ Build background template closely matching data
  - Used to determine the background functional form for the final $S + B$ fit
  - Not used to determine the background function parameters



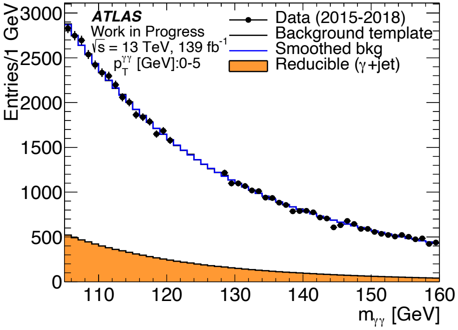

→ Need shapes & fractions of all background sources to build complete template

# Background Model

## Strategy

- Parametrize total background component with simple analytical function
  - But which functional form should be used?

→ Build background template closely matching data
  - Used to determine the background functional form for the final $S + B$ fit
  - Not used to determine the background function parameters




→ Need shapes & fractions of all background sources to build complete template

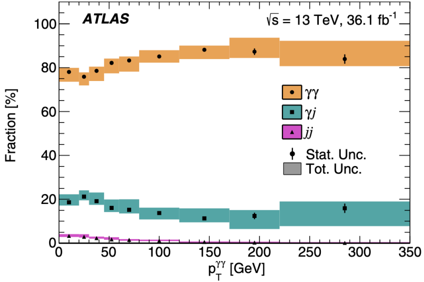
- Smooth background templates using Gaussian Process Regression (GPR)
- Background model bias studies
  - Estimate potential impact of background model choice on signal yield: spurious signal (SS)
  - Background function selection

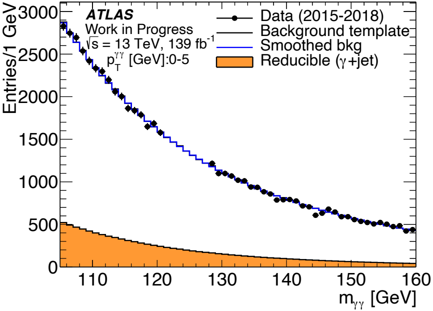
# Background Model

## Template Building

- Backgrounds are continuous non-resonant $\gamma\gamma$ (irreducible) and $\gamma j$, $jj$ (reducible)
- Build background template by adding different background distributions:
  - $\gamma\gamma$ shape from $\gamma\gamma$ MC simulation
  - $\gamma j$ shape from data control region
  - $jj$ neglected (no impact on template shape)
- Relative background fractions from data
  - Control regions defined by inverting the selection criteria for each $\gamma$ separately
- Inclusive fiducial composition typically $\gamma\gamma$ ~76%, $\gamma j$ ~21%, $jj$ ~3%

Source: 1802.04146v2

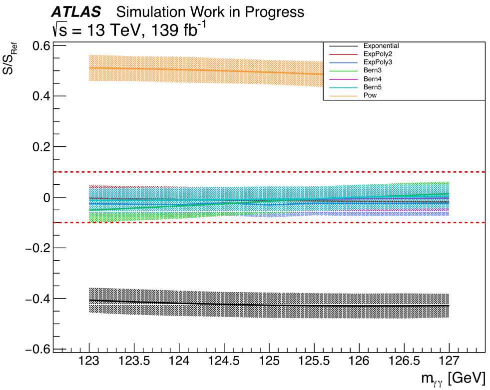


# Background Model

## Spurious Signal: Potential bias on $\sigma \times Br$ by background model choice

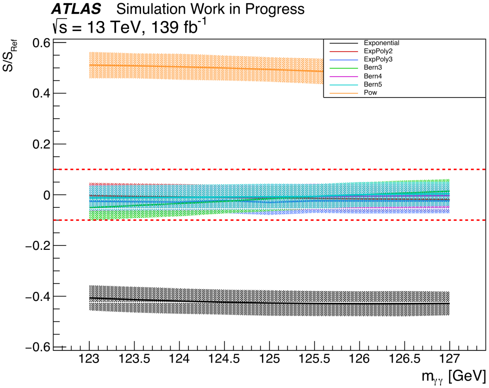
- A good background model describes the data well, with only small bias on the extracted $\sigma \times Br$ value
  - Evaluate goodness-of-fit ($\chi^2$) and SS of each considered function
  - Considered functions: Exp(polynomial $\mathcal{O}$(1,2,3)), Bernstein polynomials $\mathcal{O}$(3,4,5), Power Law



# Background Model

## Spurious Signal: Potential bias on $\sigma \times Br$ by background model choice

- A good background model describes the data well, with only small bias on the extracted $\sigma \times Br$ value
  - Evaluate goodness-of-fit ($\chi^2$) and SS of each considered function
  - Considered functions: Exp(polynomial $\mathcal{O}$(1,2,3)), Bernstein polynomials $\mathcal{O}$(3,4,5), Power Law
- SS is evaluated by $S + B$ fits to the background-only template: 17 fits scanning $m_{\gamma\gamma} \in [121, 129]$ GeV, max($S$) chosen as SS

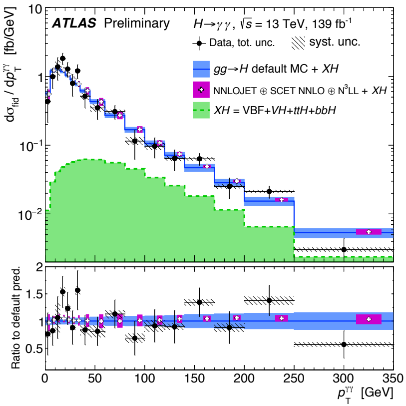


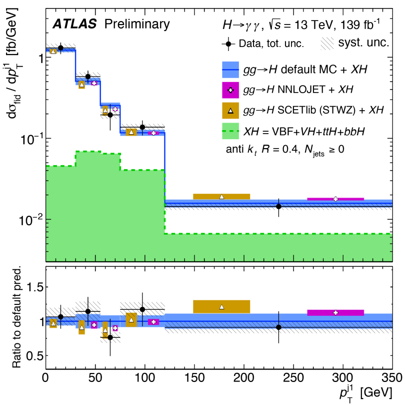

- Choose function with lowest #d.o.f. passing SS and $\chi^2$ criteria
- Spurious signal used as systematic uncertainty in final data fit

# Results

## Selected Highlights

- Spectrum probed to high $p_T^{\gamma\gamma}$ values
- Provide differential variables for precise theory comparisons
- Provide total cross section measurement
- Overall good agreement with SM predictions



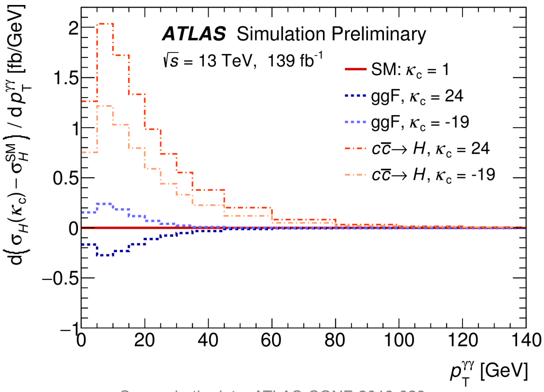

$$\sigma_{\text{fid}} = 65.2 \pm 4.5\ \text{(stat.)} \pm 5.6\ \text{(syst.)} \pm 0.3\ \text{(theo.) fb}$$

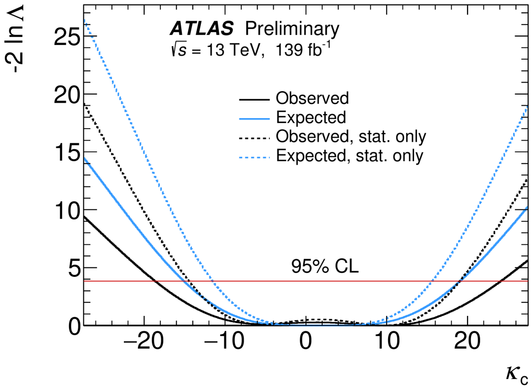
$$\sigma_{\text{SM}}^{\text{predicted}} = 63.6 \pm 3.3\ \text{fb}$$

# $\kappa_b$ and $\kappa_c$ Interpretations

## How to use differential measurements to constrain physics beyond the SM

- Consider overall rescaling of the $b/c$-quark Yukawa couplings $y$ with a factor $\kappa_{b/c} \sim y_{b/c}^{\text{modified}} / y_{b/c}^{\text{SM}}$
- Constraints on $\kappa_b$ and $\kappa_c$ from measured $p_T^{\gamma\gamma}$ distribution (profile likelihood)



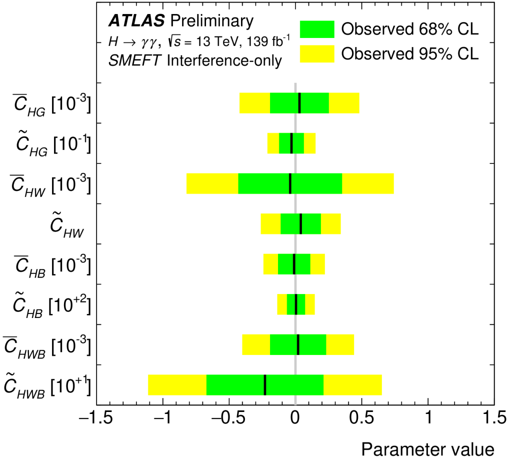




Source both plots: ATLAS-CONF-2019-029

# EFT Interpretation

- SM EFT adding eight dimension-6 operators to SM Lagrangian
  - Basically (re-)combining SM operators to higher orders
- Warsaw basis, new physics scale at $\Lambda = 1$ TeV
- Limits obtained by likelihood fit to $p_T^{\gamma\gamma}$, $N_{\text{jets}}$, $m_{jj}$, $\Delta\phi_{jj}$ and $p_T^{j1}$
- Limits are set on the dim-6 operators

Source: ATLAS-CONF-2019-029



# Summary

- Differential cross section measurements provide a precision view into Higgs boson properties, possible BSM physics, and physics modelling
- $H \to \gamma\gamma$ is an excellent channel for such measurements
  - Good $m_{\gamma\gamma}$ resolution and selection efficiency
- A reliable background model is essential for the $H \to \gamma\gamma$ analysis
- Measurements show good agreement with SM predictions, with $\sim 11\%$ precision
- Looking forward to include more variables and to combine different channels!

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 678215

**Contact**

DESY. Deutsches Elektronen-Synchrotron

Nils Gillwald
nils.gillwald@desy.de

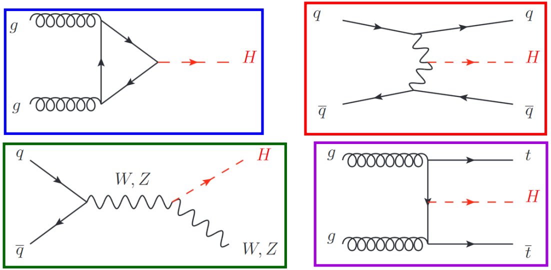
www.desy.de

# Backup

European Research Council
Established by the European Commission

# How to produce Higgs bosons at the LHC...

- Four major production modes:
  - Gluon-gluon fusion (ggF), vector boson fusion (VBF), $W/Z$ Strahlung, top-quark associated production



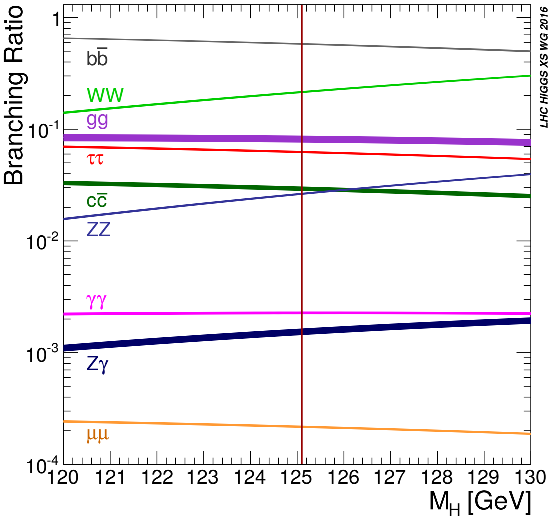

- ggF dominating, then VBF, VH($WH/ZH$), $ttH$

| Mode | Cross section [pb] |
|---|---|
| ggF | 48.61 |
| VBF | 3.766 |
| $WH$ | 1.358 |
| $ZH$ | 0.880 |
| $ttH$ | 0.507 |

Source: LHCHXSWG

# ...and how they decay

- $H$ dominantly decays into $b\bar{b}$
- Next $WW$, $gg$, $\tau\tau$
- Decay into $gg$ and $\gamma\gamma$ only possible via loops (massless particles!)
- Decay into photons ($\gamma\gamma$) relatively suppressed, BR $\sim$ 0.227%
- $\gamma\gamma$ still excellent channel to work with
  - High photon selection efficiency
  - Excellent $m_{\gamma\gamma}$ resolution
    $\rightarrow$ robust background subtraction
  - Balances low BR!



# Event Selection & Fiducial Region

- Harmonize truth and reconstruction cuts as much as possible
  - Avoid acceptance effects: reduces model dependencies

**Photon and Event Selection**

- Two energetic, isolated, central $\gamma$
- $m_{\gamma\gamma} \in [105, 160]$ GeV – most signal in $[120, 130]$ GeV
- Neural network selects $\gamma\gamma$ vertex: better four-momentum and $m_{\gamma\gamma}$ resolution, jet association

**Jet Selection**

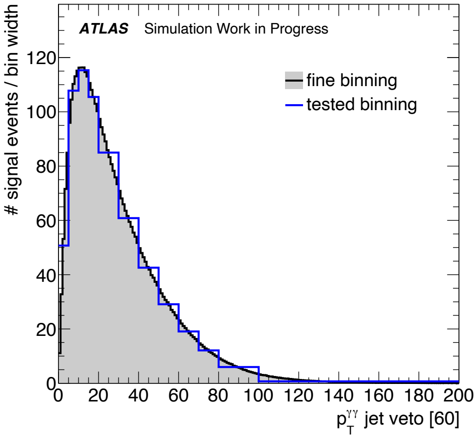
- Central $R = 0.4$ anti-$k_t$ jets
- Jet vertex tagger to suppress pileup
- Overlap removal: prioritize $e$, $\gamma$ over jets

# Electron and Photon ID

- Shapes of the showers in the electromagentic (EM) calorimeter used for $e$ and $\gamma$ ID
- Shapes can be sorted into four categories:
  - Track properties ($e$), lateral development, longitudinal development, track-cluster spatial compatibility
- $e$ ID: discriminate $e$ from $e$ from hadronic jets, converted $\gamma$, and $e$ from heavy-flavour hadrons
- $\gamma$ ID: discriminate $\gamma$ from hadronic jet backgrounds (bkd), $\pi^0$ decays
- $e$ ID constructs a likelihood discriminant from these, using data $Z \to ee$ ($E_T > 15$ GeV) and $J/\Psi \to ee$ ($E_T < 15$ GeV) tag and probes
- $\gamma$ uses cut-based approach: MC $Z \to \ell\ell\gamma$ signal with data $Z$+jets bkd (10 $< E_T <$ 25) GeV and inclusive-photon production MC with MC dijet bkd ($E_T > 25$ GeV)

# Choice of Binnings

- Strategy: Start with fine binning, merge until optimisation criteria satisfied
- Optimise variable binnings with respect to:
  - Expected significance $\gtrsim 2\sigma$
  - Harmonization with categories in Higgs couplings analysis (STXS), $H \to 4\ell$ and CMS



# Signal Extraction

## Including Systematic Uncertainties in Likelihood Fits -- Profile Likelihood

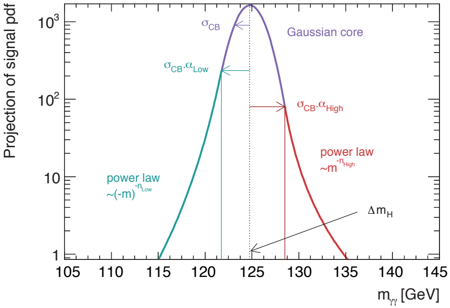
- Systematics are due to imperfect knowledge of auxiliary parameters
- But we have some knowledge: $\theta = \theta_0 \pm \Delta\theta$
- Modify likelihood function accordingly:

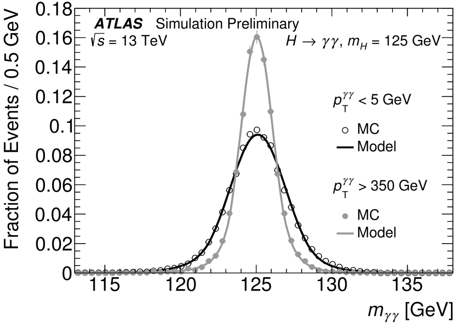
$$\mathcal{L}(\underbrace{n, \theta^0}_{\text{Observables}}; \underbrace{\mu, \theta}_{\text{Parameters}}) = \prod_{i \in \text{bins}} \mathcal{P}(n_i | \mu, \theta) \times \prod_{j \in \text{syst}} \mathcal{G}(\theta_j^0 | \theta_j, \Delta\theta_j)$$

- $\mathcal{P}$ is the probability of the observables given a model – i.e., the signal and background model enter here!
- A priori knowledge interpreted as ”auxiliary measurement”
  - Implemented as constraint / penalty term, i.e. probability density function
  - Often Gaussian, interpreting $\pm\Delta\theta$ as $\sigma$

# Signal Model

Source: ATLAS-CONF-2019-029



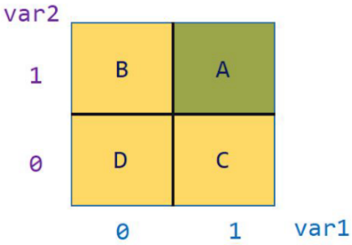

- Signal modelled by a double sided Crystal Ball function (CB)
- Good description of the $H$ signal peak

- Signal shape fixed from fit to $m_H = 125$ GeV MC
- mean + width of CB fixed, but can be effectively modified by constrained energy scale and resolution and Higgs mass nuisance parameters

# ABCD Sideband Method & use for Background Templates

- Invert $\gamma$ isolation and ID in order to create signal and background enriched regions

Source: Particle Wiki




- E.g., var1 = ID, var = iso
- A is signal regions, B, C, D background enriched
- IF var1 and var2 are *uncorrelated*, then

$$\frac{N_A}{N_B} = \frac{N_C}{N_D} \Leftrightarrow N_A = \frac{N_B N_C}{N_D}$$

$\Rightarrow$ Get an estimate on signal region from backround regions!

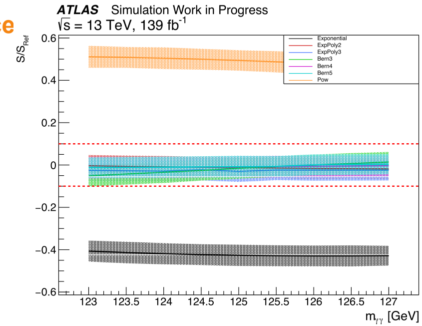
- Apply a similar method to $H \to \gamma\gamma$ background decomposition, treating each photon separately $\to$ 2x2D sideband method

# Background Model

## Spurious Signal: Potential bias by background model choice

Aim: Estimate potential bias to $S$ by background model choice

- Run $S + B$ fits on background template for all considered models
  - 9 steps for $m_{\gamma\gamma} \in [123, 127]$ GeV
- Max($S$) chosen as this model's potential bias
- Accept model if:
  - $S < 20\%$ of expected background uncertainty, or
  - $S < 10\%$ of expected # signal events
  - If no model passes, use $S \pm x\Delta_S$: $\Delta_S$ statistical uncertainty of $S$, $x \in \{1, 2\}$ can be further relaxed
  - $p(\chi^2)$ for model on template sidebands > 1%
- Choose lowest d.o.f. model passing criteria



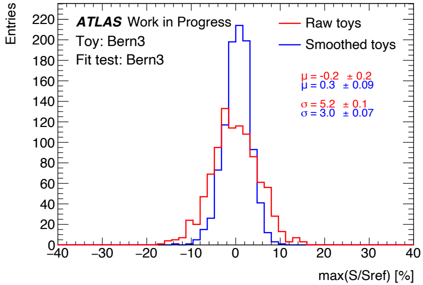

- Spurious signal scan, $p_T^{\gamma\gamma}$ bin 7
- Functions: Exp(polynomial $\mathcal{O}$(1,2,3)), Bernstein polynomials $\mathcal{O}$(3,4,5), Power Law

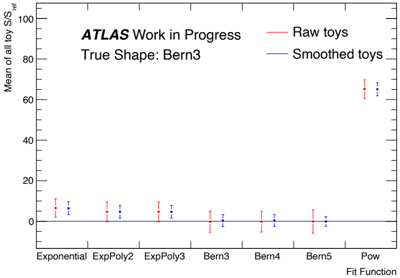
# Background Model

## GPR Bias Study

- If we smooth our background templates, do we spoil our spurious signal bias estimate?
- Strategy:
  - Assume a true background shape and throw background template toys from it
  - Compare the spurious signal results on raw and smoothed templates
  - If unbiased, mean value of distributions should agree within uncertainties
- Examples from $p_T^{\gamma\gamma}$, bin 45 - 60 GeV

Based on this study, smoothing is unbiased!



# Unfolding in a nutshell

## Removing Efficiency and Reconstruction Effects

**The problem.**

- Experiments measure detector-level quantities
- Would need to run full reconstruction on MC theory prediction
  - Cannot be done outside of ATLAS!
  - Difficult to compare results to updated theoretical predictions
- Impossible to compare to analytical theory predictions

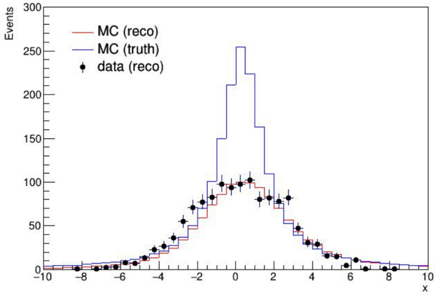
**The solution.**

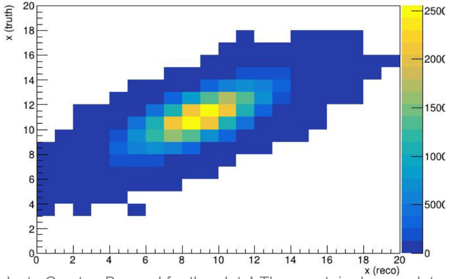
- Convert experimental data to ”truth quantities”!
- More useful for theorists
- Easier comparison to MC generator output
- Easier cross-experiment comparisons
- Don’t depend on knowledge on how to run the detector simulation

# Unfolding in a nutshell, II

## Removing Efficiency and Reconstruction Effects

- Different approaches to unfolding:
  - Bin-by-bin correction factor (top): $c =$ truth / reco
  - Matrix approach (bottom): $\vec{n}_{\text{truth}} = R^{-1} \times \vec{n}_{\text{reco}}$
- Each method has different strengths and weaknesses
- Many more points, such as biases, regularization (fixing problems with matrix inversion), interpolations, model dependence due to use of simulation, ...
- Our analysis uses the **matrix approach**
  - No regularization, directly in likelihood fit




Thanks to Carsten Burgard for the plots! They contain dummy data only.

# The Final Likelihood Function

## in this Analysis

$$\mathcal{L}\left(m_{\gamma\gamma}; \nu^{\text{sig}}, \nu^{\text{bkg}}\right) =$$

# The Final Likelihood Function

## in this Analysis

$$\mathcal{L}\left(m_{\gamma\gamma}; \nu^{\text{sig}}, \nu^{\text{bkg}}\right) = \qquad \nu^{\text{sig}} \underbrace{\mathcal{S}\left(m^{j}_{\gamma\gamma}; \theta_k\right)}_{\text{signal model PDF}}$$

# The Final Likelihood Function

## in this Analysis

$$\mathcal{L}\left(m_{\gamma\gamma};\nu^{\text{sig}},\nu^{\text{bkg}}\right) = \nu^{\text{sig}}\underbrace{\mathcal{S}\left(m^{j}_{\gamma\gamma};\theta_k\right)}_{\text{signal model PDF}} + \nu^{\text{bkg}}\underbrace{\mathcal{B}\left(m^{j}_{\gamma\gamma}\right)}_{\text{bkg model PDF}}$$

# The Final Likelihood Function

## in this Analysis

$$\mathcal{L}\left(m_{\gamma\gamma}; \nu^{\text{sig}}, \nu^{\text{bkg}}\right) = \underbrace{\prod_{j}^{n} \left[ \nu^{\text{sig}} \underbrace{\mathcal{S}\left(m^{j}_{\gamma\gamma}; \theta_k\right)}_{\text{signal model PDF}} + \nu^{\text{bkg}} \underbrace{\mathcal{B}\left(m^{j}_{\gamma\gamma}\right)}_{\text{bkg model PDF}} \right]}_{\text{product over all events in a } m_{\gamma\gamma} \text{ bin}}$$

# The Final Likelihood Function

## in this Analysis

$$\mathcal{L}\left(m_{\gamma\gamma};\nu^{\text{sig}},\nu^{\text{bkg}}\right) = \underbrace{\frac{e^{-\nu}}{n!}\prod_{j}^{n}\left[\nu^{\text{sig}}\underbrace{\mathcal{S}\left(m_{\gamma\gamma}^{j};\theta_k\right)}_{\text{signal model PDF}} + \nu^{\text{bkg}}\underbrace{\mathcal{B}\left(m_{\gamma\gamma}^{j}\right)}_{\text{bkg model PDF}}\right]}_{\text{product over all events in a } m_{\gamma\gamma} \text{ bin}}$$

# The Final Likelihood Function

## in this Analysis

$$\mathcal{L}\left(m_{\gamma\gamma};\nu^{\text{sig}},\nu^{\text{bkg}}\right) = \underbrace{\prod_i^m \frac{e^{-\nu}}{n!} \underbrace{\prod_j^n \left[\nu^{\text{sig}} \underbrace{\mathcal{S}\left(m^j_{\gamma\gamma};\theta_k\right)}_{\text{signal model PDF}} + \nu^{\text{bkg}} \underbrace{\mathcal{B}\left(m^j_{\gamma\gamma}\right)}_{\text{bkg model PDF}}\right]}_{\text{product over all events in a } m_{\gamma\gamma} \text{ bin}}}_{\text{product over all bins of a variable}}$$

# The Final Likelihood Function

## in this Analysis

$$\mathcal{L}\left(m_{\gamma\gamma};\nu^{\text{sig}},\nu^{\text{bkg}}\right) = \underbrace{\prod_{i}^{m} \frac{e^{-\nu}}{n!} \underbrace{\prod_{j}^{n} \left[\nu^{\text{sig}} \underbrace{\mathcal{S}\left(m_{\gamma\gamma}^{j};\theta_k\right)}_{\text{signal model PDF}} + \nu^{\text{bkg}} \underbrace{\mathcal{B}\left(m_{\gamma\gamma}^{j}\right)}_{\text{bkg model PDF}}\right]}_{\text{product over all events in a } m_{\gamma\gamma} \text{ bin}}}_{\text{product over all bins of a variable}} \times \underbrace{\prod_{k} C_k\left(\theta_k;0,1\right)}_{\text{nuisance parameter constraints}}$$

# The Final Likelihood Function

## in this Analysis

$$\mathcal{L}\left(m_{\gamma\gamma};\nu^{\text{sig}},\nu^{\text{bkg}}\right) = \underbrace{\prod_i^m \frac{e^{-\nu}}{n!} \underbrace{\prod_j^n \left[\nu^{\text{sig}} \underbrace{\mathcal{S}\left(m_{\gamma\gamma}^j;\theta_k\right)}_{\text{signal model PDF}} + \nu^{\text{bkg}} \underbrace{\mathcal{B}\left(m_{\gamma\gamma}^j\right)}_{\text{bkg model PDF}}\right]}_{\text{product over all events in a } m_{\gamma\gamma} \text{ bin}}}_{\text{product over all bins of a variable}} \times \underbrace{\prod_k C_k\left(\theta_k;0,1\right)}_{\text{nuisance parameter constraints}},$$

where $\quad \nu_i^{\text{sig}} = \sum_l R_{il}\sigma_l B_{\gamma\gamma} \cdot \mathcal{L}_{\text{int}}$

# The Final Likelihood Function

## in this Analysis

$$\mathcal{L}\left(m_{\gamma\gamma};\nu^{\text{sig}},\nu^{\text{bkg}}\right)=\underbrace{\prod_{i}^{m}\frac{e^{-\nu}}{n!}\underbrace{\prod_{j}^{n}\left[\nu^{\text{sig}}\underbrace{\mathcal{S}\left(m_{\gamma\gamma}^{j};\theta_k\right)}_{\text{signal model PDF}}+\nu^{\text{bkg}}\underbrace{\mathcal{B}\left(m_{\gamma\gamma}^{j}\right)}_{\text{bkg model PDF}}\right]}_{\text{product over all events in a } m_{\gamma\gamma} \text{ bin}}}_{\text{product over all bins of a variable}}\times\underbrace{\prod_{k}C_k\left(\theta_k;0,1\right)}_{\text{nuisance parameter constraints}},$$

where $$\nu_i^{\text{sig}}=\sum_{l}R_{il}\sigma_l B_{\gamma\gamma}\cdot\mathcal{L}_{\text{int}}$$

> $R_{il}$ is the response matrix used for the unfolding, $n_i^{\text{detector}}=R_{il}n_j^{\text{particle}}$

# The Final Likelihood Function

## in this Analysis

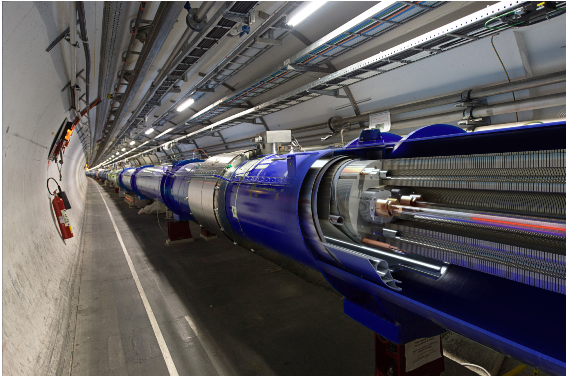
$$\mathcal{L}\left(m_{\gamma\gamma};\nu^{\text{sig}},\nu^{\text{bkg}}\right)=\underbrace{\prod_i^m \frac{e^{-\nu}}{n!}\underbrace{\prod_j^n\left[\nu^{\text{sig}}\underbrace{\mathcal{S}\left(m^j_{\gamma\gamma};\theta_k\right)}_{\text{signal model PDF}}+\nu^{\text{bkg}}\underbrace{\mathcal{B}\left(m^j_{\gamma\gamma}\right)}_{\text{bkg model PDF}}\right]}_{\text{product over all events in a } m_{\gamma\gamma} \text{ bin}}}_{\text{product over all bins of a variable}}\times\underbrace{\prod_k C_k\left(\theta_k;0,1\right)}_{\text{nuisance parameter constraints}},$$

where $\nu_i^{\text{sig}}=\sum_l R_{il}\sigma_l B_{\gamma\gamma}\cdot\mathcal{L}_{\text{int}}$

- $R_{il}$ is the response matrix used for the unfolding, $n_i^{\text{detector}}=R_{il}n_j^{\text{particle}}$
- $\sigma_l B_{\gamma\gamma}$ is the cross section in bin $l$ times the $H\to\gamma\gamma$ BR

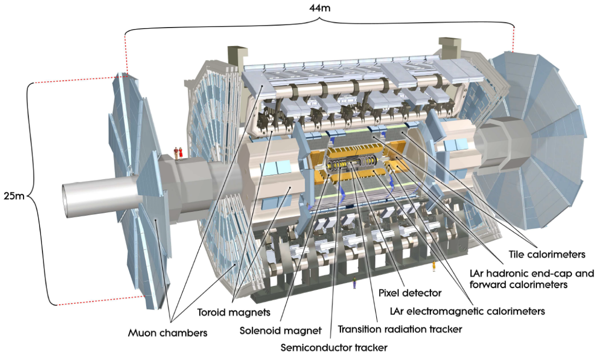
# The LHC

- Operating since 2008
- 27 km long, $pp$ collisions at $\sqrt{s} = 13$ TeV (since 2015)
- ATLAS peak luminosity: $2.14 \times 10^{34}$ cm$^{-2}$s$^{-1}$
- 2808 bunches, $\sim 10^{11}$ protons each; temporal spacing 25 ns
- Main experiments: ATLAS/CMS (multi-purpose), LHCb ($b$ physics / CP), ALICE (heavy ions / $qg$-plasma)



# The ATLAS experiment

- 44 $\times$ 25 m, 7000 t
- Divided into barrel and end-caps, nearly $4\pi$ solid angle coverage
- Onion structure: Tracking layers, EM and hadronic calorimeters, muon chambers
- Magnets between layers bend charged particles for easier reconstruction and measurements
- Trigger selects interesting events ($\sim$1/40,000 or 0.0025%)



# Some thoughts on template statistics: GPR Smoothing

European Research Council
Established by the European Commission

ATLAS
EXPERIMENT

DESY.

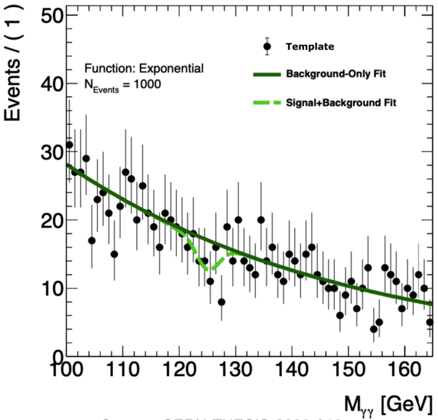
# Background Model

## GPR Smoothing - Why it's needed

- $H \to \gamma\gamma$ analysis has very small $S/B$ ratio
- Rely on smooth background template to estimate spurious signal (SS)
  - Model choice and associated SS very sensitive to even small fluctuations in template
  - SS might just be large due to random fluctuation in the generated distribution!

  $\Rightarrow$ Suboptimal function choice, overestimated systematic uncertainties
- Can't just generate more background events - already generated 1.025 billion!

Note: Black points are MC template, not data!!




Source: CERN-THESIS-2020-013

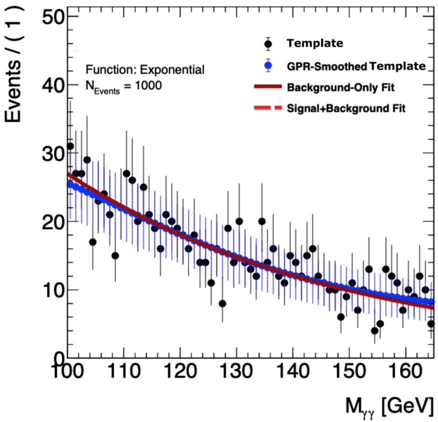
# Background Model

## GPR Smoothing - Why it's needed

- $H \to \gamma\gamma$ analysis has very small $S/B$ ratio
- Rely on smooth background template to estimate spurious signal (SS)
  - Model choice and associated SS very sensitive to even small fluctuations in template
  - SS might just be large due to random fluctuation in the generated distribution!

  $\Rightarrow$ Suboptimal function choice, overestimated systematic uncertainties
- Can't just generate more background events - already generated 1.025 billion!
- Use Gaussian Process Regression (GPR): smooth out fluctuations for better background model choice

Note: Black points are MC template, not data!!




Source: CERN-THESIS-2020-013

# Background Model

## GPR Smoothing - What it is & how it works

- A Gaussian process (GP) is a set of random variables $x_i$ where all subsets have a multivariate normal distribution
  - All linear combinations of subsets are Gaussian distributed
  - Correlations encoded in covariance matrix $\Sigma$

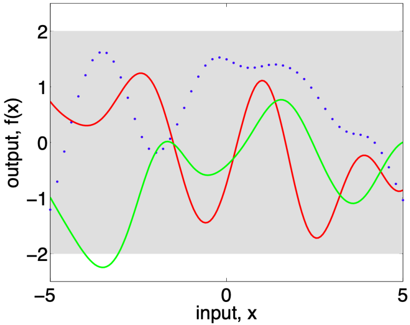
$$\Sigma = \begin{pmatrix} \sigma_0 & c_{01} & c_{02} & \dots & c_{0n} \\ c_{10} & \sigma_1 & c_{12} & \dots & c_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ c_{n0} & c_{n1} & c_{n2} & \dots & \sigma_n \end{pmatrix}$$

# Background Model

## GPR Smoothing - What it is & how it works

- A Gaussian process (GP) is a set of random variables $x_i$ where all subsets have a multivariate normal distribution
  - All linear combinations of subsets are Gaussian distributed
  - Correlations encoded in covariance matrix $\Sigma$

$$\Sigma = \begin{pmatrix} \sigma_0 & c_{01} & c_{02} & \dots & c_{0n} \\ c_{10} & \sigma_1 & c_{12} & \dots & c_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ c_{n0} & c_{n1} & c_{n2} & \dots & \sigma_n \end{pmatrix}$$



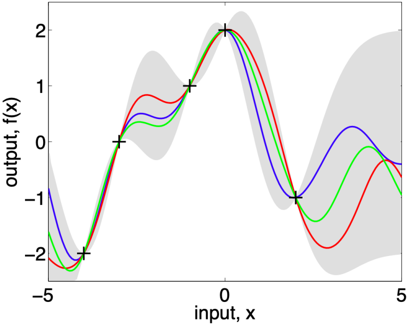

Source: Rasmussen, Williams: Gaussian Processes for Machine Learning

# Background Model

## GPR Smoothing - What it is & how it works

- A Gaussian process (GP) is a set of random variables $x_i$ where all subsets have a multivariate normal distribution
  - All linear combinations of subsets are Gaussian distributed
  - Correlations encoded in covariance matrix $\Sigma$
- Fit GP by fitting $\mu$, $\Sigma$ to $n$ supporting points
  - Important to choose a good prior $\mu$ and $\Sigma$!
- For a histogram, supporting points would be the single bins
- Can use fitted $\mu$, $\Sigma$ as smoothed version of a distribution!

$$\Sigma = \begin{pmatrix} \sigma_0 & c_{01} & c_{02} & \cdots & c_{0n} \\ c_{10} & \sigma_1 & c_{12} & \cdots & c_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{n0} & c_{n1} & c_{n2} & \cdots & \sigma_n \end{pmatrix}$$




Source: Rasmussen, Williams: Gaussian Processes for Machine Learning

# GPR: Choice of Prior - Kernels

- Choice of mean function depends on the distribution to model:
  - Constant, Linear, Exponential, ...

- Correlation matrix can be simplified by using a Kernel
  - Analytically describes level of correlation between two distinct points
- Different kinds of Kernels possible:
  - Radial Basis Function (RBF) Kernel:
    $k(x_i, x_j) = \exp\left(-\frac{1}{2}d(x_i/l, x_j/l)^2\right)$ with $d$ Euclidian distance, $l$ length scale
  - Gibbs Kernel (useful for smoothly falling distributions):
    Like RBF, but $l \rightarrow l(x)$
- Smoothing gets more agressive with increasing length scale
- Can also add a Kernel sensitive to the uncertainties of the supporting points
  - White Kernel: $constant \times Id(N)$
  - Linear Error Kernel: $Id(N)$ matrix with linearly decreasing values as a function of $x$