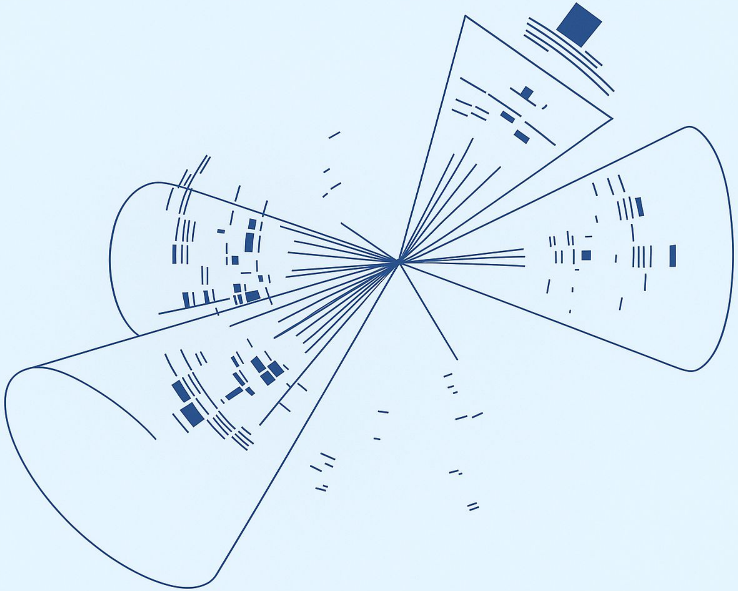
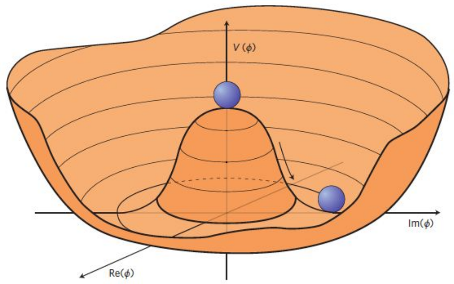
$X \to SH \to b\bar{b}b\bar{b}$ search in ATLAS



Marta Lanzac Berrocal (IFIC, CSIC-UV)

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS
VNIVERSITAT DE VALÈNCIA
IFIC INSTITUT DE FÍSICA CORPUSCULAR
EXCELENCIA SEVERO OCHOA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
UNIÓN EUROPEA
FONDO EUROPEO DE DESARROLLO REGIONAL
"Una manera de hacer Europa"
AGENCIA ESTATAL DE INVESTIGACIÓN
ATLAS EXPERIMENT

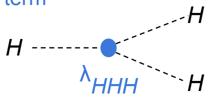
Higgs boson potential



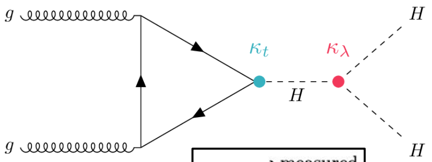
- The Higgs boson potential remains partially unconstrained:

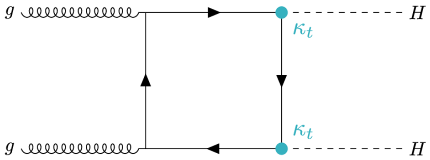
Higgs boson mass term · Trilinear self interaction term

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2 \longrightarrow V_H = \frac{1}{2}m_H H^2 + \lambda v H^3 + \ldots$$

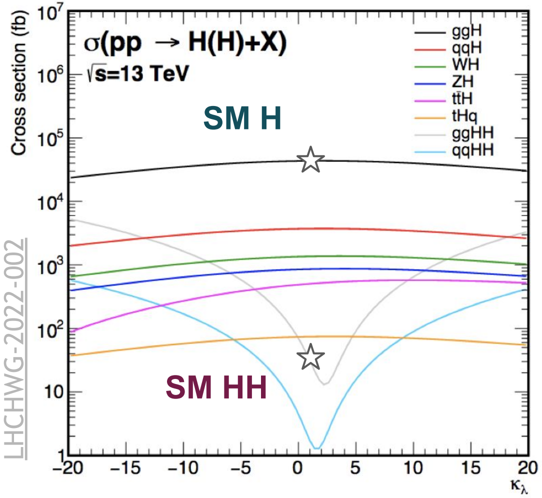


- Measuring **Higgs boson pair production (HH)** is the only direct probe of the Higgs boson self-coupling (λ_{HHH})





$$\kappa_\lambda = \frac{\lambda_{HHH}^{\text{measured}}}{\lambda_{HHH}^{\text{SM}}}$$

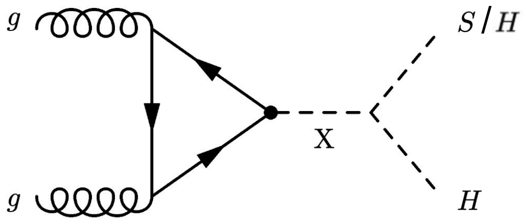


- In the Standard Model, the HH production is rare ($\sigma_{SM} \approx 33$ fb at $\sqrt{s} = 13$ TeV)
- **No observation yet** - expected to be observed at HL-LHC

Beyond the Standard Model Theories in Higgs Physics

- **Non-resonant production (pp → HH)** enhancements can arise from:
  - **Loop corrections** involving new particles
  - **Anomalous coupling** between the Higgs boson and SM particles
  - **Modified Higgs potential,** affecting the trilinear self-coupling (λ_{HHH})

- **Resonant production (pp → X → HH / SH)** motivated by different BSM theories:
  - **TRSM** (Two Real Singlet Model)**:**
    → adds to two CP even neutral bosons alongside the SM Higgs
  - **2HDM** (Two Higgs-Doublet Model):
    → two new Higgs doublets; e.g. CP-conserving: 1 lighter and 1 heavier (wrt. to the SM-Higgs) CP-even scalar
  - **NMSSM** (Next-to-Minimal Supersymmetric Standard Model): 2HDM+S model
    → multiple charged and neutral scalars
    → less constrained than the pMSSM



H: SM Higgs (m = 125 GeV)
S: additional scalar

Signal topologies and phase space in $X \to SH \to 4b$

The resonant HH searches in the 4b final state have similar characteristics as in the non-resonant case:

- Main advantage: **largest branching ratio**
- Challenges: complex **QCD multijet background estimation**

Mass coverage and regimes:

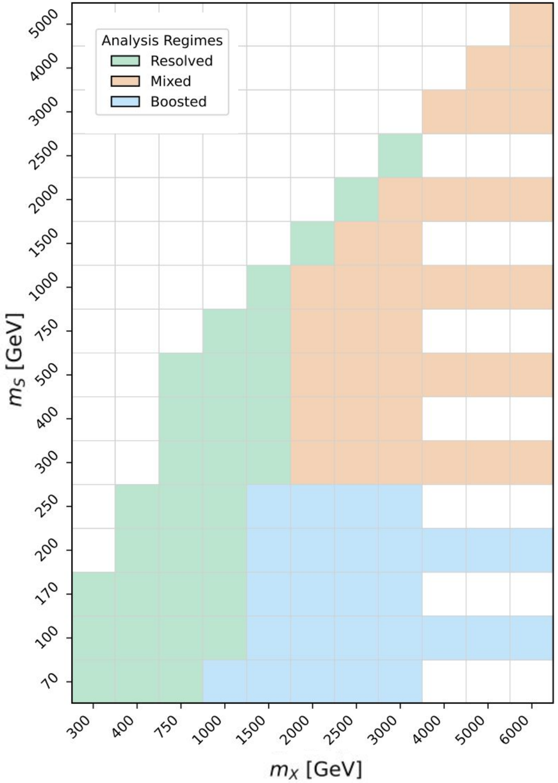
- Full analysis scan: $\mathbf{m_X \in [300, 6000]}$ **GeV**, $\mathbf{m_S \in [70, 5000]}$ **GeV**
- Phase space divided by topology:

Resolved

Mixed

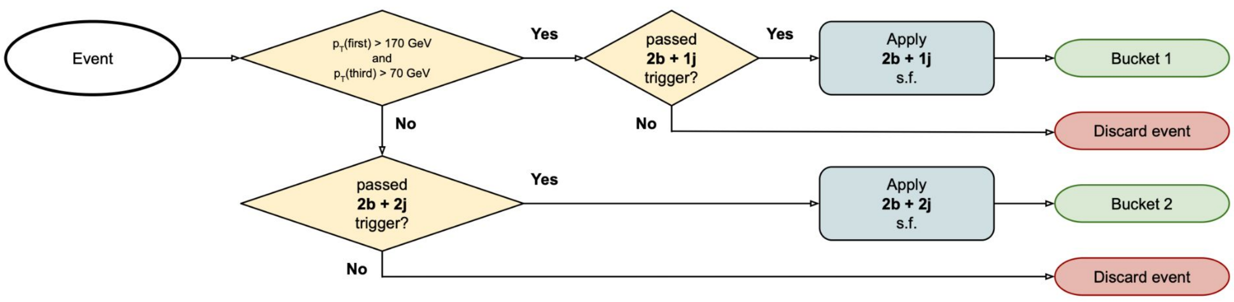
Boosted

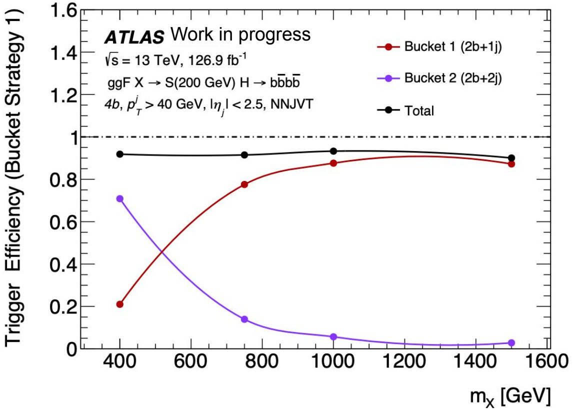
In this talk we will focus on the **resolved regime**



Trigger strategy

- **Multi-b-jet triggers (2b+2j and 2b+1j)** used in the analysis
- Trigger efficiencies measured at the jet-level for each trigger leg as a function of the reconstructed jet kinematics
- Scale Factors (SFs) applied to MC events to correct the simulated trigger efficiency to the performance measured in data
- Event-level SFs are calculated from jet-level inputs, taking into account the combination of different jets and the correlation between different triggers
- To simplify the calculation of event-level SFs, a **trigger bucket strategy** is used (events split in different categories, in each category only a single trigger is used)
- Triggers and trigger bucket scheme chosen to **maximize the yield of X → SH → 4b signal events** using MC samples





Event selection

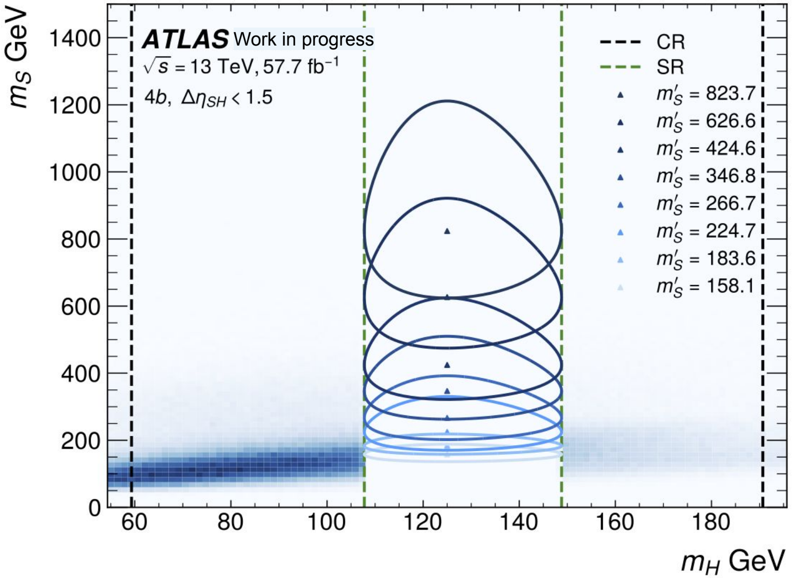
Offline selection:

- ≥4 jets reconstructed using the anti-k_t algorithm with R=0.4:
  - p_T > 40 GeV
  - $|\eta|$ < 2.5
  - Jet Vertex Tagger using NN
  - b-tagged with GN2 @ 85% efficiency
- BDT for S and H reconstruction
- $|\Delta\eta_{SH}|$ < 1.5 for QCD background rejection

Kinematic regions:

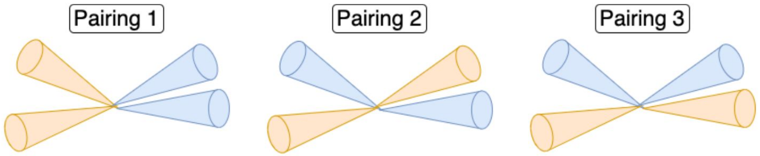
- ***Signal Region (SR) strip:*** m_H window
- ***Control Region (CR):*** m_H sidebands
  - $m_H \in$ [59.3, 107.8] GeV
  - $m_H \in$ [148.8, 190.7] GeV
- ***SRs:*** Ellipses in m_S - m_H mass plane defined using X_{SH}

$$m_H \in [107.8,\ 148.8]\ \text{GeV and } X_{SH} = \sqrt{\left(\frac{m_H - m'_H}{r_{f,H} m_H}\right)^2 + \left(\frac{m_S - m'_S}{r_{f,S} m_S}\right)^2} < 1.6$$

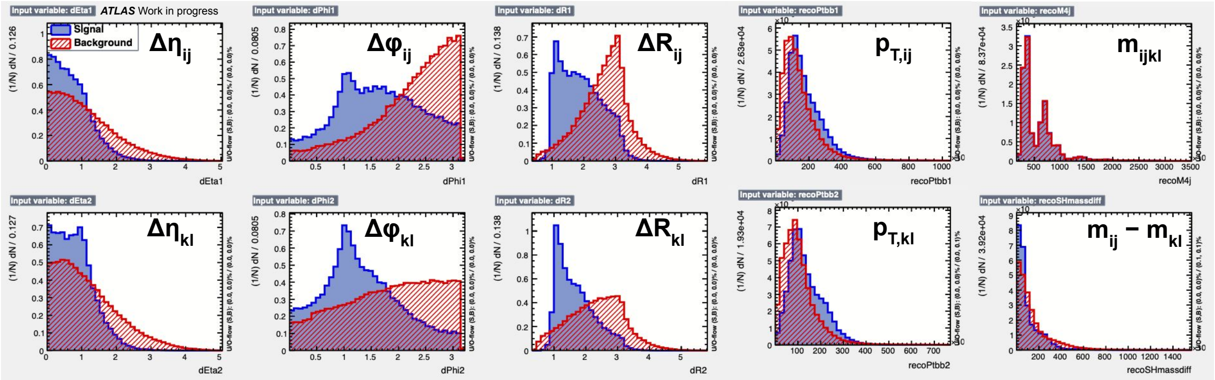


Jet pairing strategy: BDT

- Three possible combinations to pair four b-tagged jets, having two dijet systems to reconstruct the H and S candidates:



- **BDT** developed for b-jet pairing:
  - → BDT trained to assign scores to each combination
  - → Pairing with highest score = correct pair
  - → Training variables: $\Delta\eta_{ij}$, $\Delta\varphi_{ij}$, ΔR_{ij}, $p_{T,ij}$, $\Delta\eta_{kl}$, $\Delta\varphi_{kl}$, ΔR_{kl}, $p_{T,kl}$, m_{ijkl}, $m_{ij} - m_{kl}$
  - → Training performed using MC X → SH → 4b (using correct pairs as signal and incorrect pairs as background)

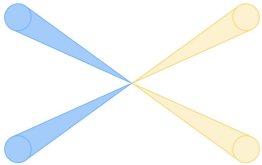


Jet pairing strategy: BDT

- **Training groups**:

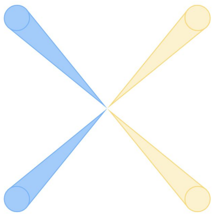
  → BDT training split in two regions based on the X and S masses and final state topology:

Non-compressed regime: $m_X > m_S + m_H$

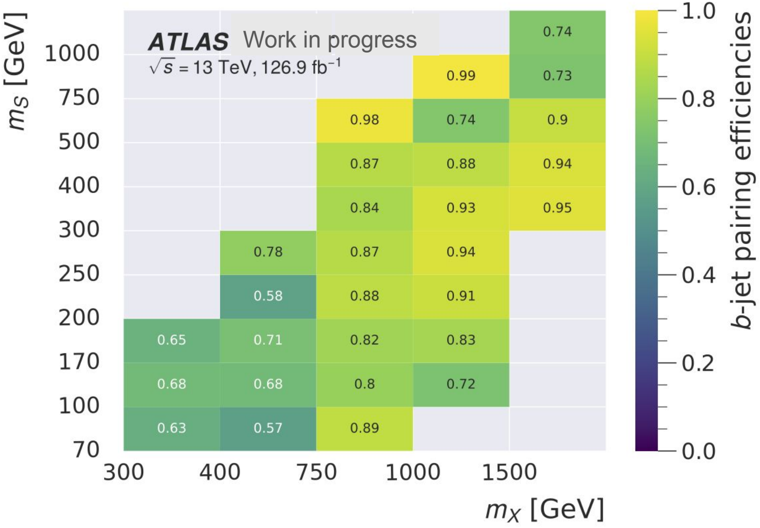


- *S* and *H* decay with a slight boost
- Resolved but collimated topology

Compressed regime: $m_X \approx m_S + m_H$

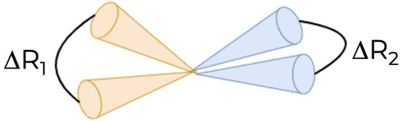


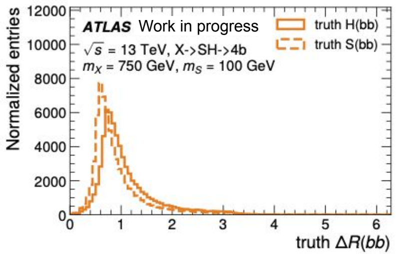
- *S* and *H* decay almost at rest
- Back-to-back topology

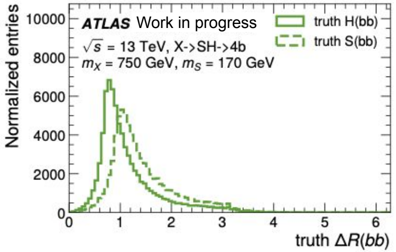


Scalar assignment strategy:

- Assign b-jet pair with the largest ΔR_{bb} as the more massive reconstructed scalar (S or H)



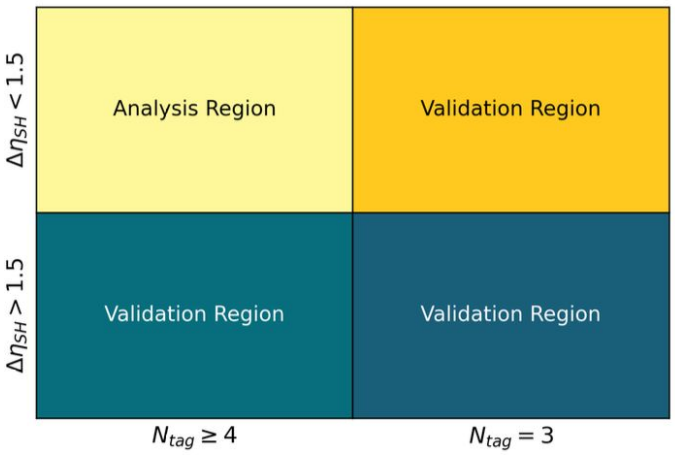




Kinematic regions

Define additional kinematic regions (Validation Regions, VRs) orthogonal to SR where predictions in m_H window can be validated using collision data:

- ***4b nominal $\Delta\eta_{SH}$:*** Analysis Region (signal expected here, full analysis selection applied, m_H window blinded)
- ***3b nominal $\Delta\eta_{SH}$ VR:*** kinematics similar to 4*b* ⇒ used for background validation
- ***4b and 3b reversed $\Delta\eta_{SH}$ VRs:*** almost purely QCD ⇒ conservative validation of analysis workflow



Nominal $\Delta\eta_{SH}$: $\Delta\eta_{SH} < 1.5$

Reversed $\Delta\eta_{SH}$: $\Delta\eta_{SH} > 1.5$

4*b*: ≥4 jets *b*-tagged with GN2v01@85%

3*b*: 3*b* + 1 jet failing GN2v01@85%

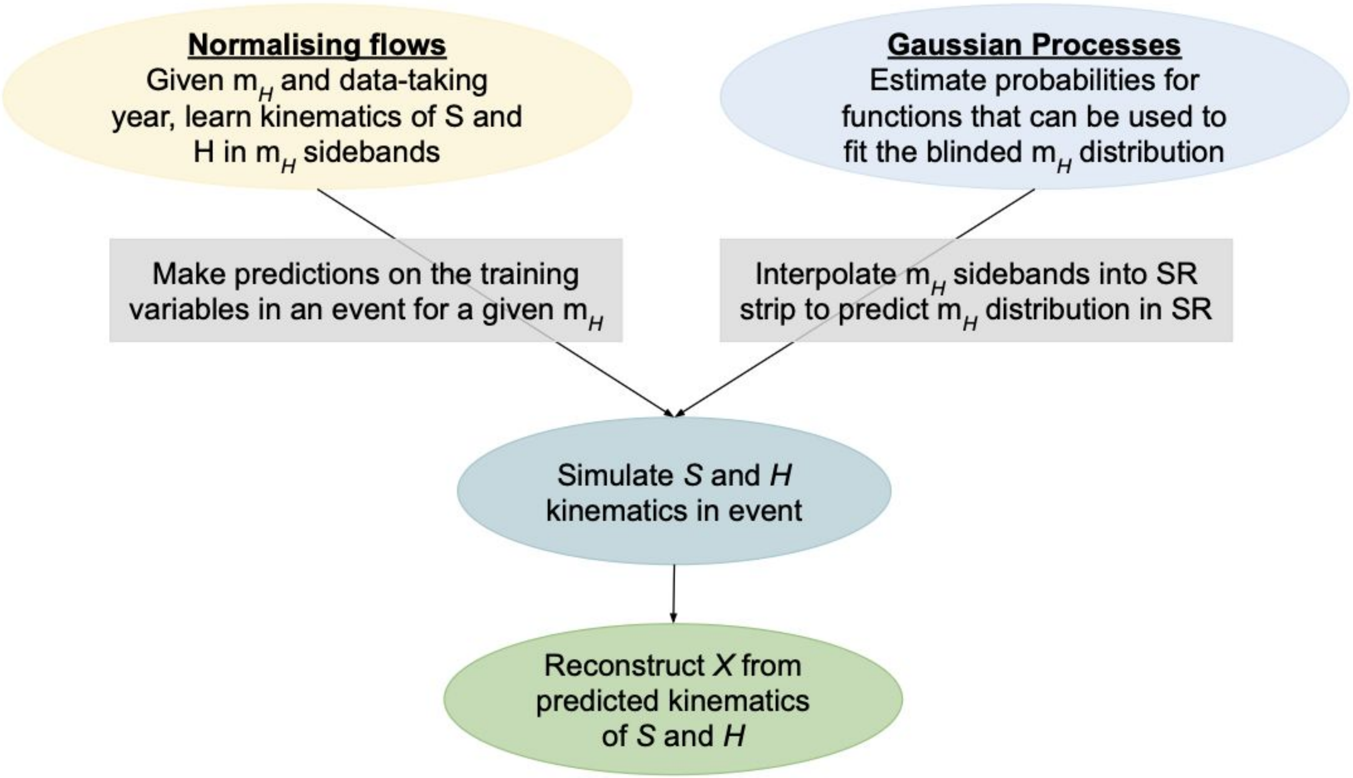
Discriminant variable: m_{SH}

Analysis using **6 categories**:

- $|\Delta\eta_{SH}| \in [0, 0.5, 1.0, 1.5]$
- $X_{SH} \in [0, 0.95, 1.96]$

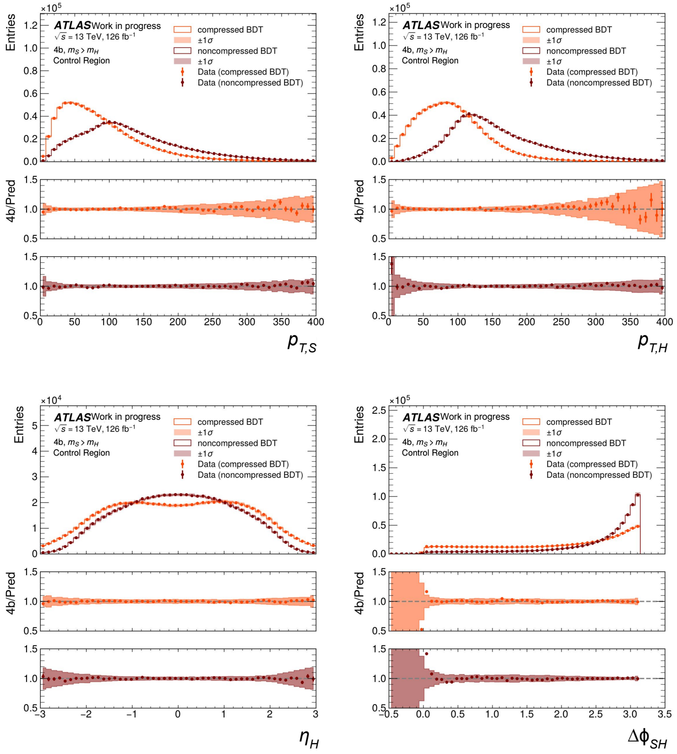
Background estimation

Dominant backgrounds: **QCD multijet** and ***tt̄* production** → entirely **data-driven background estimation**

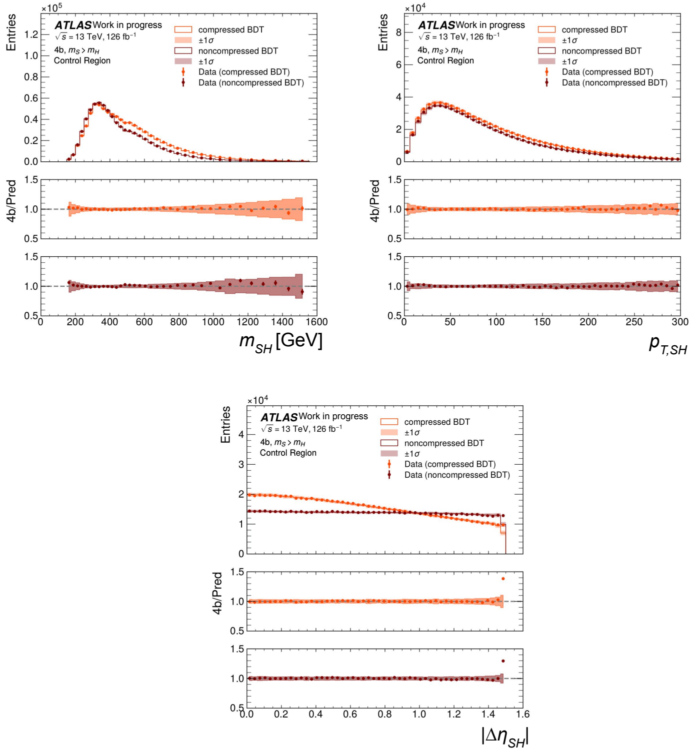


Background validation

Training Variables



Reconstructed Variables

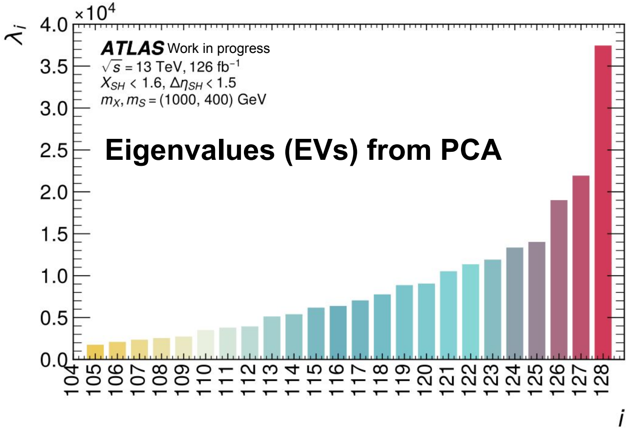


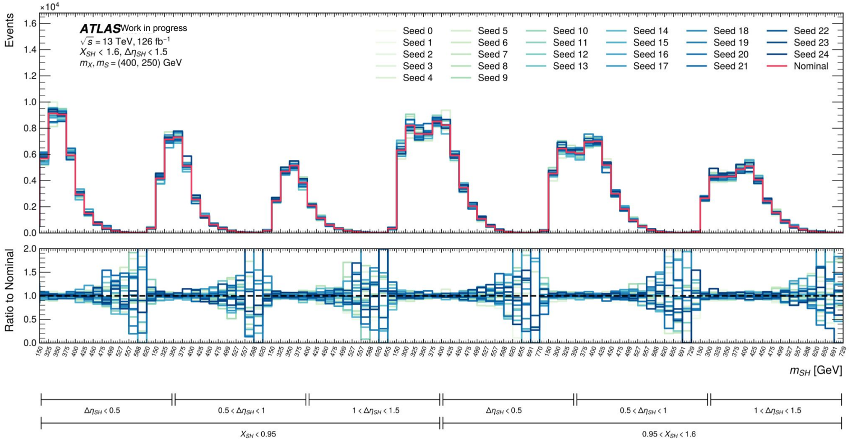
Systematic uncertainties (I)

Dominant uncertainties arise from the **data-driven background estimation** technique

Background systematics:

- **Principal Component Analysis (PCA):**
  - Used to reduce dimensionality and capture the dominant shape variations in the m_{SH} spectrum
  - Eigendecomposition of 1σ uncertainty band on 25 NF+GP predictions
  - Large EVs → correlated (shape) uncertainties, smaller EVs → uncorrelated uncertainties (applied to account for residual discrepancies in the validation region)

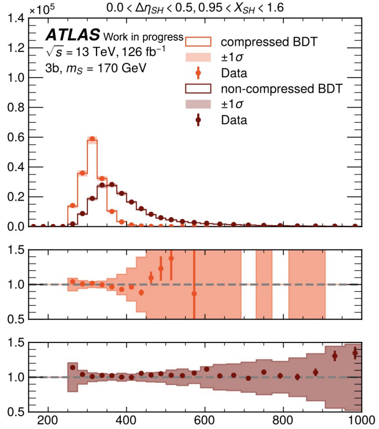


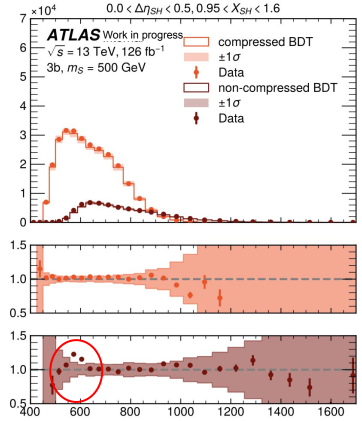


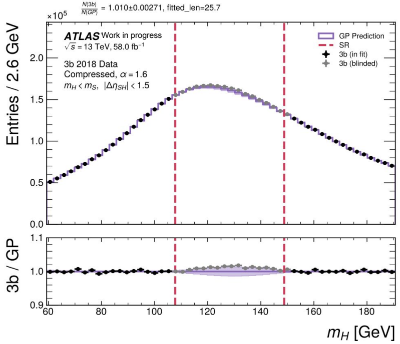
Systematic uncertainties (II)

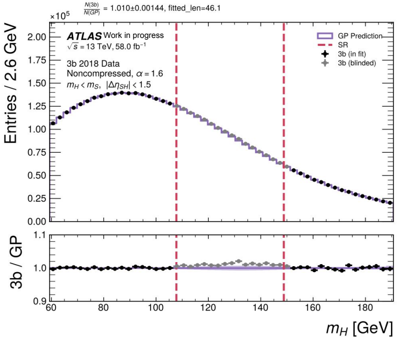
Background systematics:

- **NF+GP 3b non-closure uncertainty** (shape uncertainty):
  - Quantifies disagreement between prediction and data in the 3b validation region
  - Captures potential mismodeling of QCD multijet background
- **GP non-closure** (normalisation uncertainty):
  - Apply normalisation discrepancy in 3*b* as additional normalisation uncertainty on background



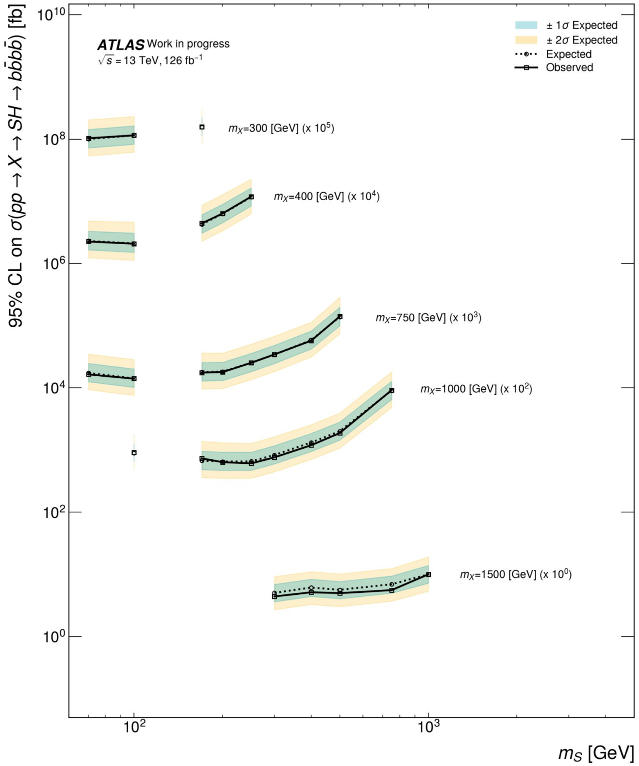


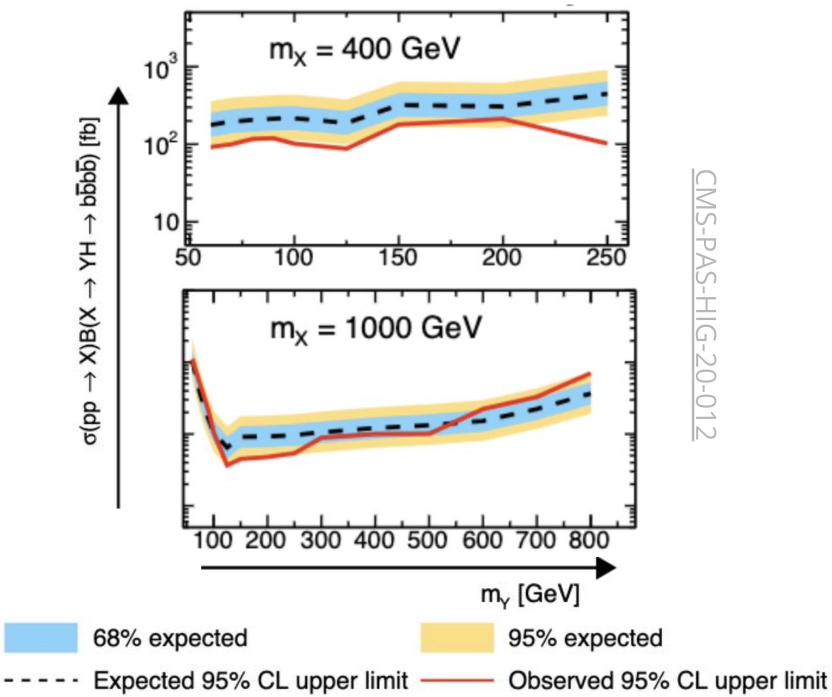




Results

- Upper limits set on $\sigma(pp \rightarrow X \rightarrow SH \rightarrow b\bar{b}b\bar{b})$, no excess observed in any category
- Mass range explored: $m_X \in [200, 3000]$ GeV, $m_S \in [70, 2500]$ GeV

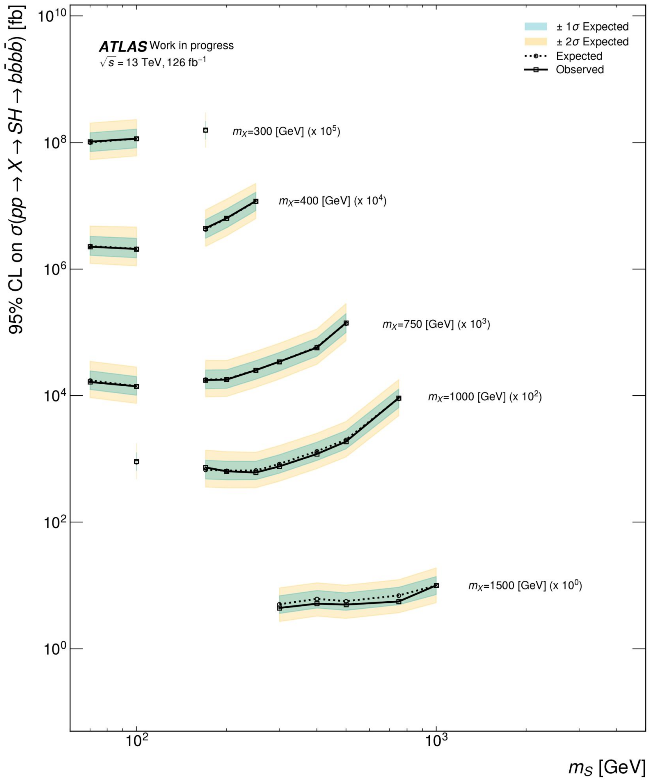


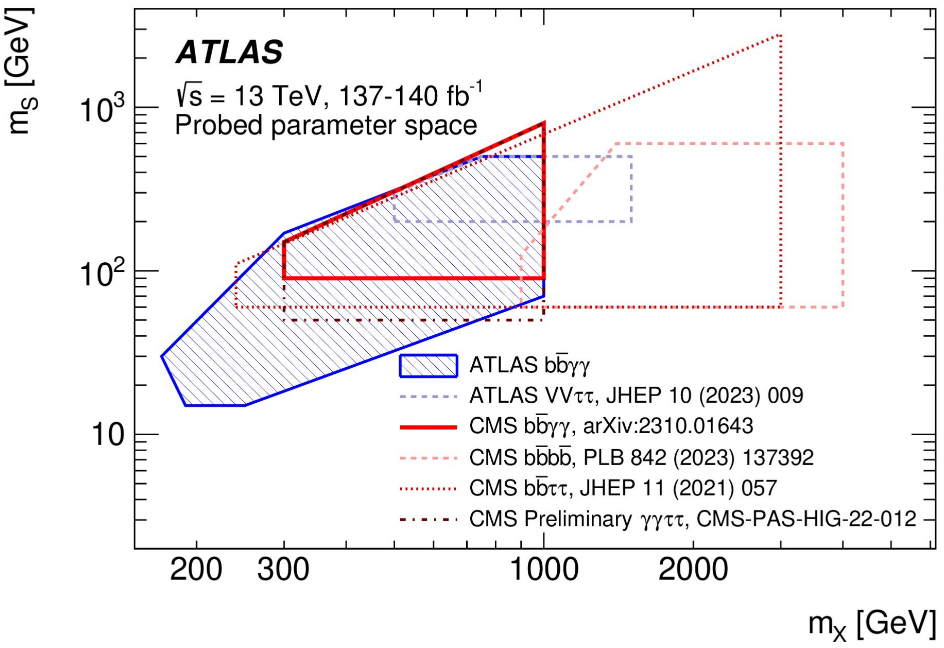


CMS-PAS-HIG-20-012

Results

- ATLAS search for $X \to SH \to b\bar{b}b\bar{b}$ including resolved, boosted and mixed topologies covers the widest mass grid yet (**$m_X \in$ [300, 6000] GeV, $m_S \in$ [70, 5000] GeV**)





arXiv:2404.12915

Summary

$X \to SH \to b\bar{b}b\bar{b}$ search with Run 2 ATLAS data:

- Asymmetric scalar cascade decays to 4b final states performed for the **first time in ATLAS**
- Analysis focuses on the resolved topology, covering $m_X \in [300, 1500]$ GeV, $m_S \in [70, 1000]$ GeV
- Targets a broad mass plane using a dedicated strategy per signal region
- Categorisation of the signal regions to harmonised sensitivity across the full grid

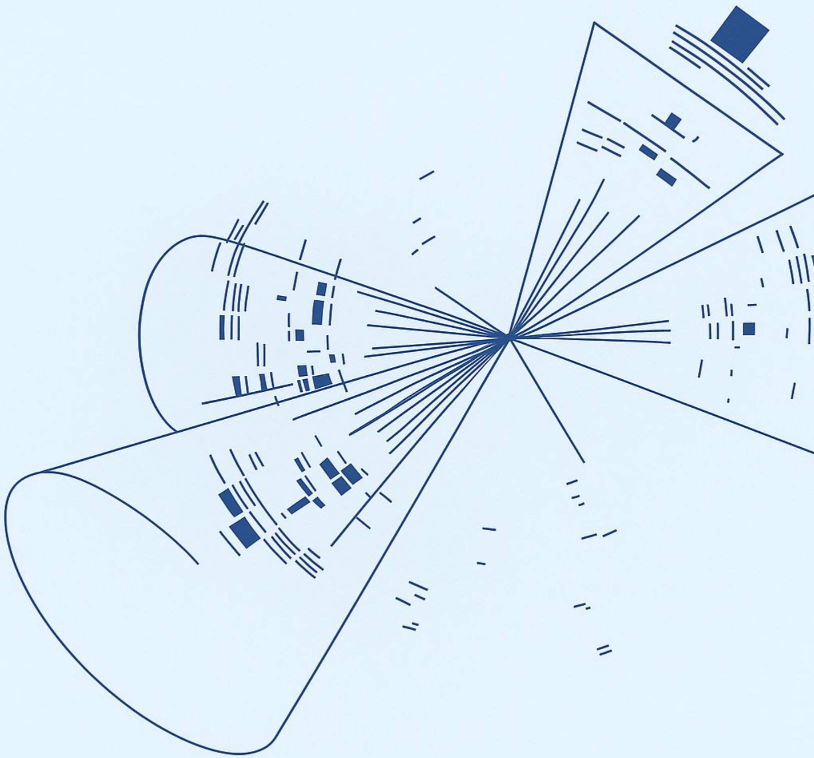
Novel techniques:

- **Transformer** (**GN2)** for jet flavour tagging and **BDT** for jet pairing
- **Normalising Flows + Gaussian Process** for fully data-driven background estimation

Results:

- No excess observed
- 95% CL upper limits on $\sigma(pp \to X \to SH \to b\bar{b}b\bar{b})$ set for full resolved grid
- Limits competitive with CMS

THANKS!



Trigger strategy

- List of all triggers considered:

Trigger Type	Trigger Name
	2016
2b + 2j	HLT_2j35_bmv2c2060_split_2j35_L14J15p0ETA25
2b + 1j	HLT_j100_2j55_bmv2c2060_split
1b	HLT_j225_bmv2c2060_split
	2017
2b + 2j	HLT_2j15_gsc35_bmv2c1040_split_2j15_gsc35_boffperf_split_L14J15p0ETA25
2b + 1j	HLT_j110_gsc150_boffperf_split_2j35_gsc55_bmv2c1070_split_L1J85_3J30
2b + HT	HLT_2j35_gsc55_bmv2c1050_split_ht300_L1HT190_J15s5pETA21
1b	HLT_j225_gsc300_bmv2c1070_split
	2018
2b + 2j	HLT_2j35_bmv2c1060_split_2j35_L14J15p0ETA25
2b + 1j	HLT_j110_gsc150_boffperf_split_2j45_gsc55_bmv2c1070_split_L1J85_3J30
2b + HT	HLT_2j45_gsc55_bmv2c1050_split_ht300_L1HT190_J15s5pETA21
1b	HLT_j225_gsc300_bmv2c1070_split

Event Selection: signal regions

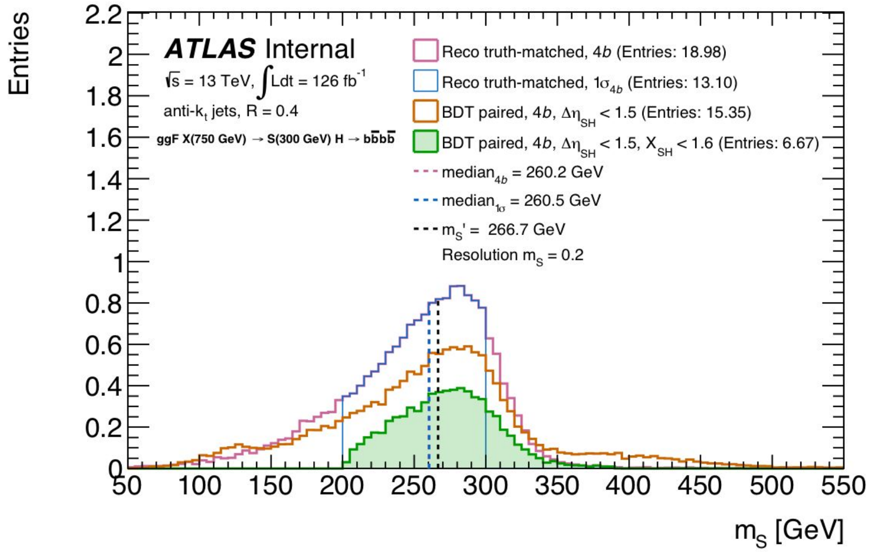
125 GeV (true m_H)

Median of inner 68% quantile of reconstructed m_S

$$X_{SH} = \sqrt{\left(\frac{m_H - m'_H}{r_{f,H} \cdot m_H}\right)^2 + \left(\frac{m_S - m'_S}{r_{f,S} \cdot m_S}\right)^2} < 1.6$$

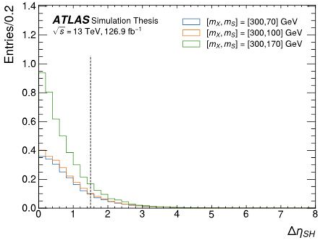
0.1 (from legacy *HH→4b* analysis)

$$\frac{\Delta m \text{ (inner 68\% quantile)}}{2 \cdot m_S'}$$

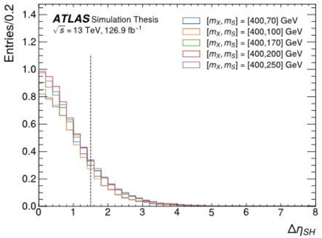


Event Selection: signal regions

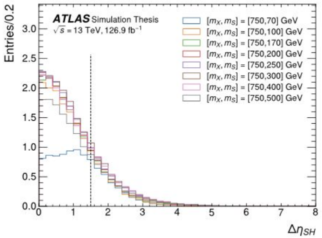
- Cut on $|\Delta\eta_{SH}| < 1.5$ for QCD background rejection



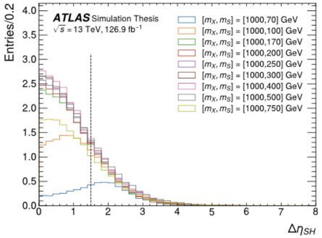
(a) $m_X = 300$ GeV



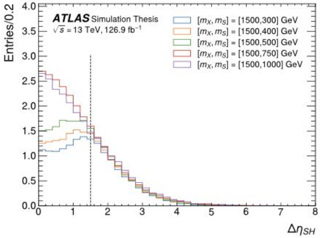
(b) $m_X = 400$ GeV



(c) $m_X = 750$ GeV



(d) $m_X = 1000$ GeV



(e) $m_X = 1500$ GeV

Background estimation

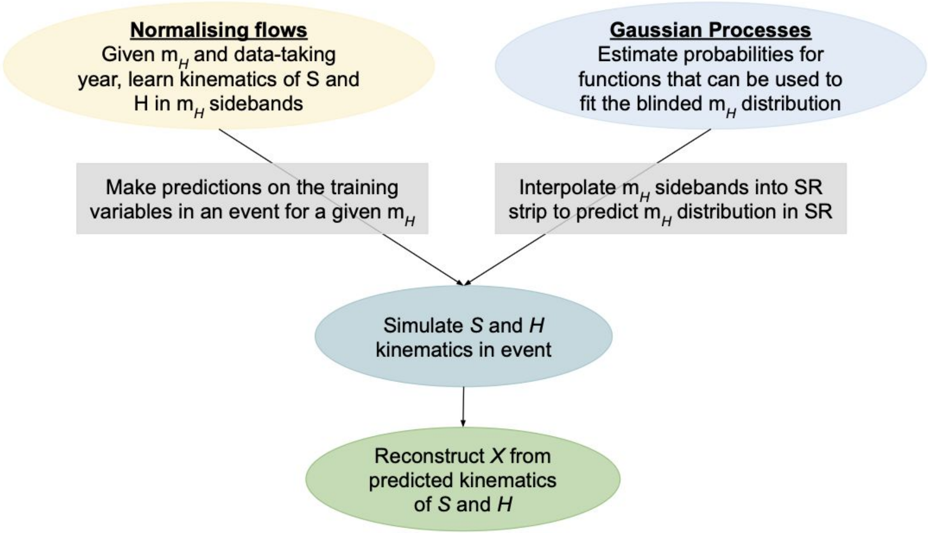
Dominant backgrounds: **QCD multijet** and **tt production** → entirely **data-driven background estimation**

Normalising Flows (NF):

- Used to model the $\mathbf{m_{SH}}$ **distribution**
- Uses a series of invertible transformation to learn known distributions into random data
- Capture nonlinear correlations and tails
- Trained using unbinned CR data for each (m_X, m_S) point
- Provide sampling, interpolation and uncertainty quantification

Gaussian Process (GP):

- Model the SR/CR as a function of m_{SH}
- Bayesian regression technique: learns a function with a prior over function space
- Provides **smooth interpolation** and **pointwise uncertainties**
- Trained on sideband data to extract SR/CR correction
- Applied as a multiplicative weight to the NF output

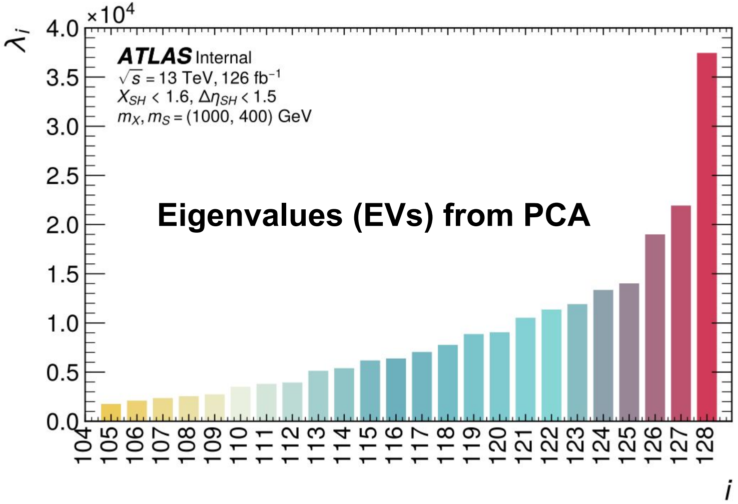


Systematic uncertainties (I)

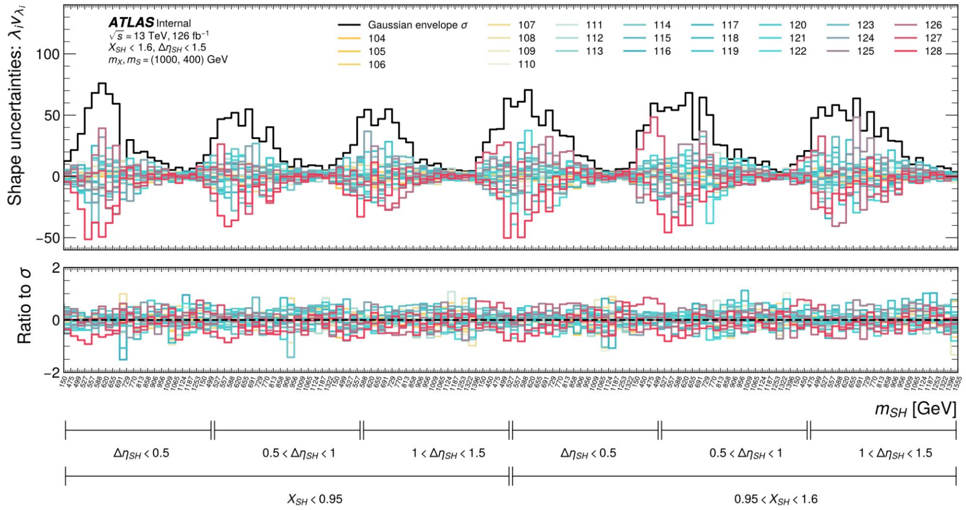
Dominant uncertainties arise from the **data-driven background estimation** technique

Background uncertainties contribution:

- *Principal Component Analysis (PCA)*:
  - Used to reduce dimensionality and capture the dominant shape variations in the m_{SH} spectrum
  - Perform 25 retrainings and reduce to their principal components
  - Largest components included as shape uncertainties on the discriminant



Shape variations from eigenvalues



Fit validation

Statistical uncertainties:

- Discriminant variable (mSH) rebinning:
  - initial binned with 25 GeV in [150, 500] GeV → many empty bins
  - rebinned requiring at least 16 predicted background events per bin

Fit validation strategy:

- Full fit performed in validation regions
- NF + GP background models are fixed; only nuisance parameters float
- Compare post-fit prediction with data for mSH spectrum
- Good closure observed in all six categories
- Systematics are profiled; pulls and constraints checked