Study of Large Radius Jet Mass Reconstruction for Higgs Pairs

Talia Saarinen,
Supervisor: Maximilian Swiatlowski

Image Credit

Outline

- The ATLAS Experiment
- Why Study HH?
  - Why Boosted ggF HH→4b?
- Reconstruction Strategies and Missing Muons
- Quantifying Mass Sensitivity

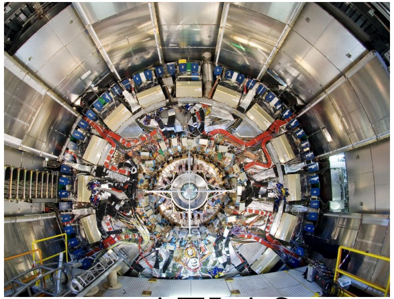
Outline

- **The ATLAS Experiment**
- Why Study HH?
  - Why Boosted ggF HH→4b?
- Reconstruction Strategies and Missing Muons
- Quantifying Mass Sensitivity

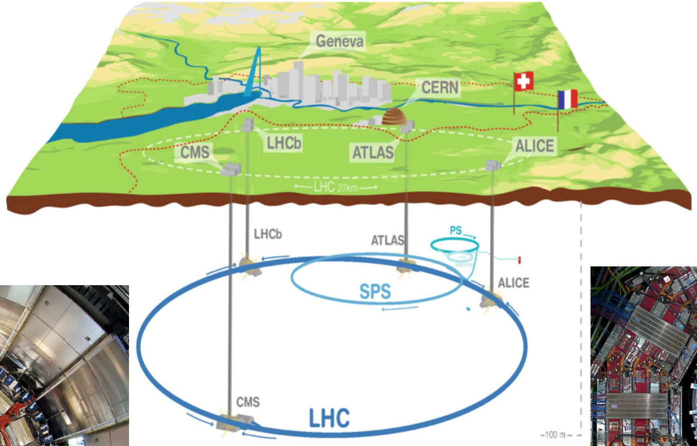
The LHC

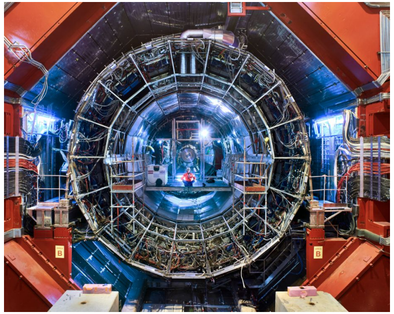
- 27 km-circumference circular synchrotron collider on Franco-Swiss border. Primarily proton-proton collisions
- Two beams of protons are brought together to collide in 4 points along the ring: ATLAS, ALICE, CMS, and LHCb

LHCb



ATLAS



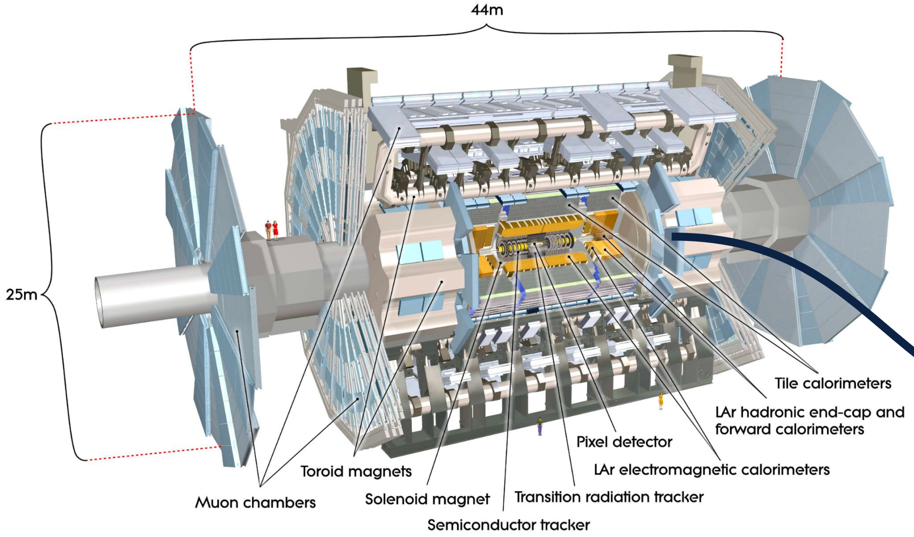


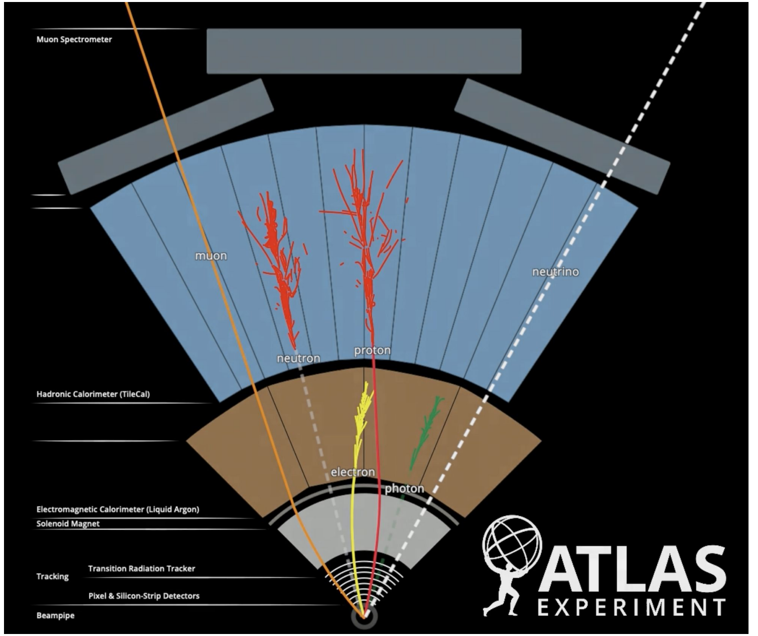
ALICE

CMS

Source: atlas.cern, home.cern, cms.cern, home.cern

The ATLAS Experiment

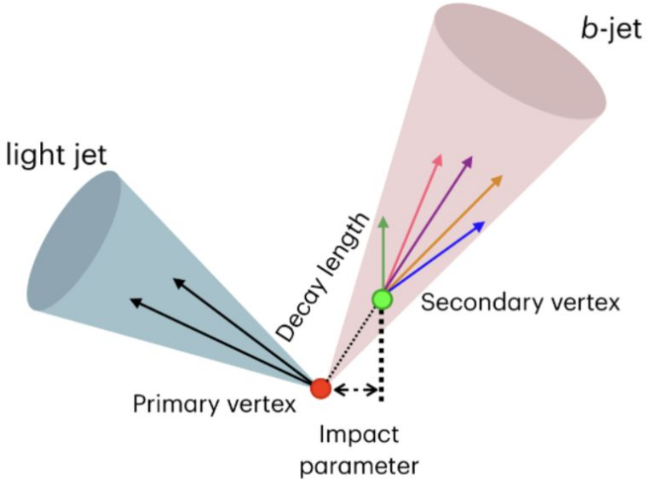




- General-purpose experiment
- 44m x 25 m toroid surrounding a collision point
- Components (innermost to outermost):
  - Inner tracker → EM Calorimeter → Hadronic Calorimeter → Muon Spectrometer

Source: ATLAS Twiki, CERN

The ATLAS Experiment: Tracks and Jets



- **Tracks** are paths individual charged particles take away from the collision point. Tracks make **hits** on the detector layers as they leave

- **Jets** are collections of particles that come from an original particle – like a b-quark

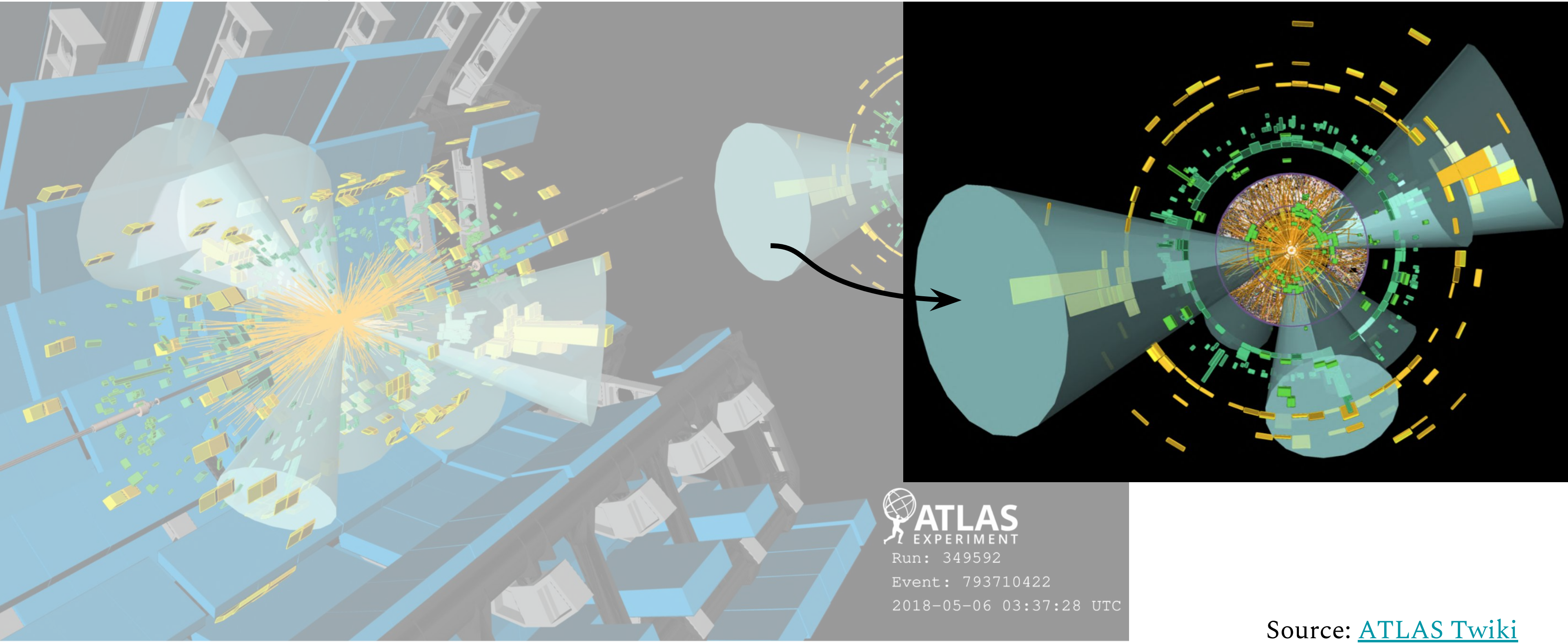
- **B-jets** start with a b-quark, forms B-hadrons → characterized by long lifetimes + secondary vertices

- Not this idealized in the actual detector – lots of pileup, overlapping tracks scramble reconstruction efforts…

Source: ATLAS Twiki

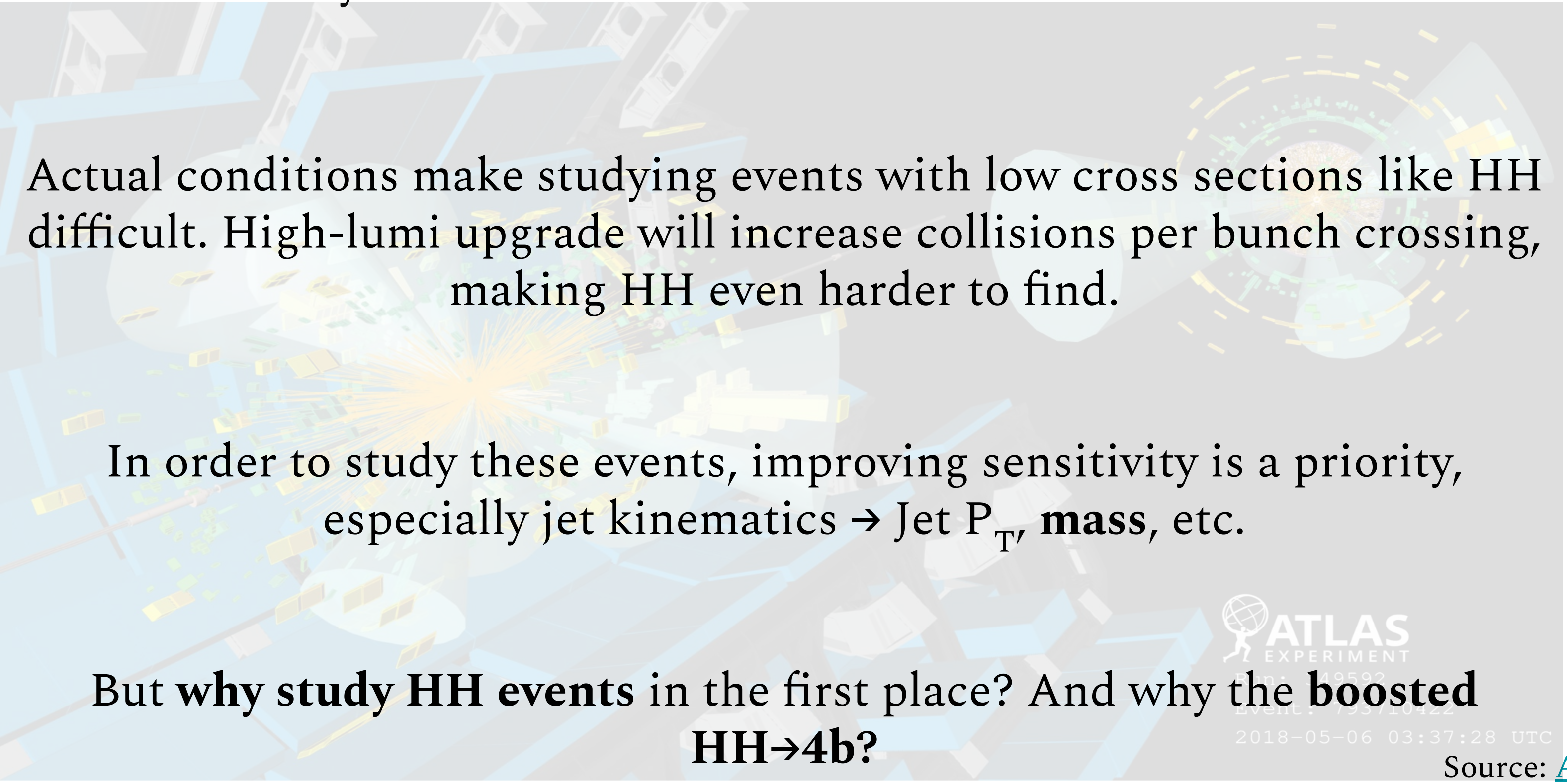
The ATLAS Experiment: Tracks and Jets

What this actually looks like: (actual HH4b candidate event)



Source: ATLAS Twiki

What this actually looks like:



Actual conditions make studying events with low cross sections like HH difficult. High-lumi upgrade will increase collisions per bunch crossing, making HH even harder to find.

In order to study these events, improving sensitivity is a priority, especially jet kinematics → Jet P_T, **mass**, etc.

But **why study HH events** in the first place? And why the **boosted HH→4b?**

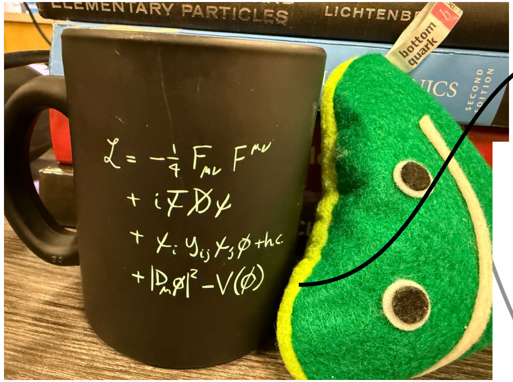
Source: ATLAS Twiki

Outline

- The ATLAS Experiment
- **Why Study HH?**
    - Why Boosted ggF HH→4b?
- Reconstruction Strategies and Missing Muons
- Quantifying Mass Sensitivity

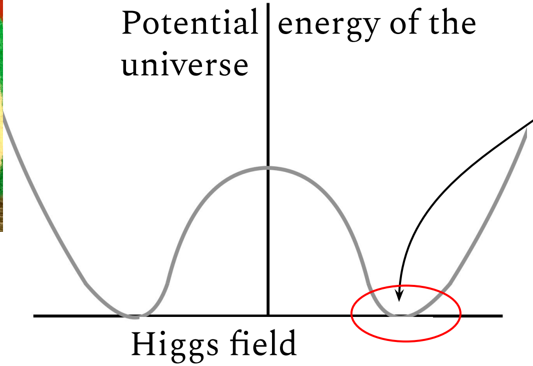
Why HH→4b?

SM Lagrangian



Expand Higgs Potential About Minimum

$$V = V_0 + \frac{1}{2} m_H^2 h^2 + \frac{m_h^2}{2v^2} v h^3 + \ldots$$

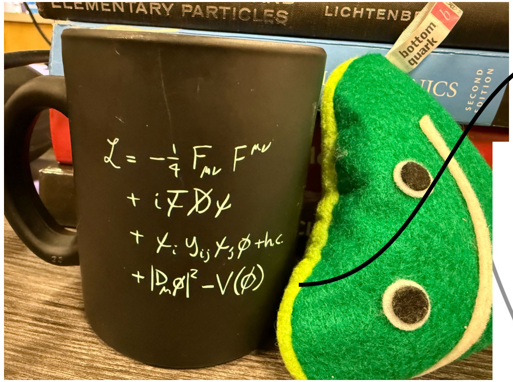


Coefficients of cubic (and higher) terms describe the **Higgs Potential**

Source: G. Aad-PhysRevD.108.052003

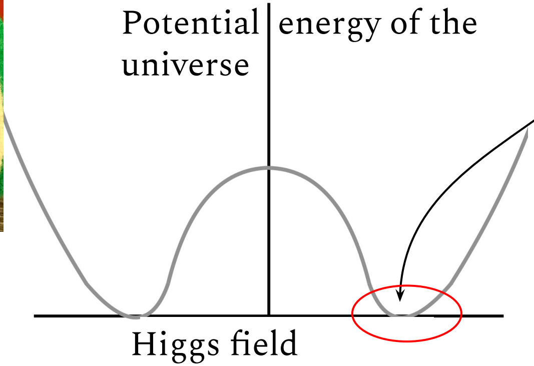
Why HH→4b?

SM Lagrangian



Expand Higgs Potential About Minimum

$$V = V_0 + \frac{1}{2}m_H^2 h^2 + \frac{m_h^2}{2v^2} v h^3 + \ldots$$



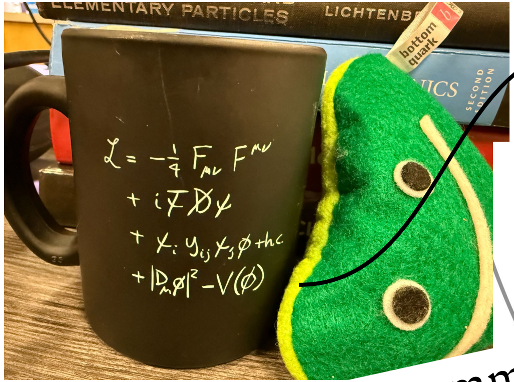
Coefficients of cubic (and higher) terms describe the **Higgs Potential**

Understanding the **Higgs Potential** tells us how particles gain mass → key to answer big-ticket questions

Source: G. Aad-PhysRevD.108.052003

Why HH→4b?

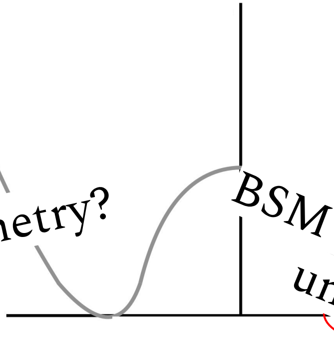
SM Lagrangian



Expand Higgs Potential About Minimum

$$V = V_0 + \frac{1}{2}m_H^2 h^2 + \frac{m_h^2}{2v^2} v h^3 + \ldots$$

Coefficients of cubic (and higher) terms describe the **Higgs Potential**



Matter-antimatter asymmetry?

BSM scenarios in how the universe evolved?

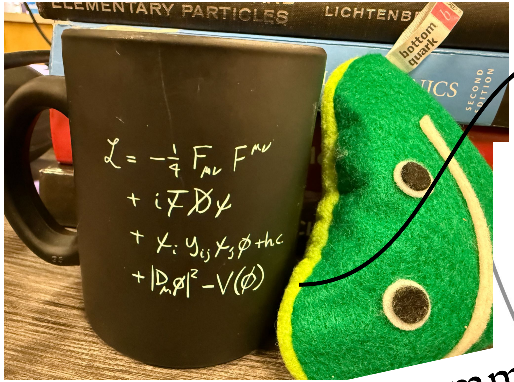
Understanding the **Higgs Potential** tells us how particles gain mass → key to answer big-ticket questions

Will our laws of physics change in the future?

Source: G. Aad-PhysRevD.108.052003

Why HH→4b?

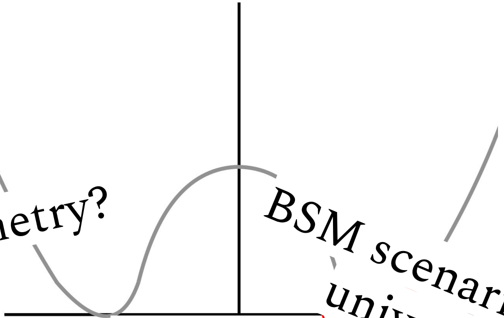
SM Lagrangian



Expand Higgs Potential About Minimum

$$V = V_0 + \frac{1}{2}m_H^2 h^2 + \frac{m_h^2}{2v^2} v h^3 + \ldots$$

Coefficients of cubic (and higher) terms describe the **Higgs Potential**



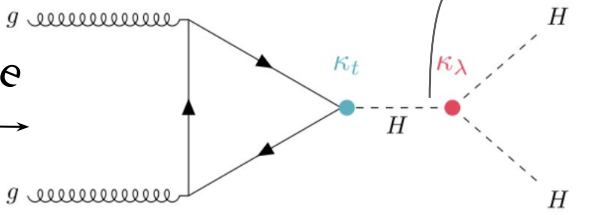
Matter-antimatter asymmetry?

BSM scenarios in how the universe evolved?

Will our laws of physics change?

Understanding the **Higgs Potential** tells us how particles gain mass → key to answer big-ticket questions

This is the goal: $\kappa_\lambda = \frac{\lambda_{HHH}}{\lambda_{HHH}^{SM}}$

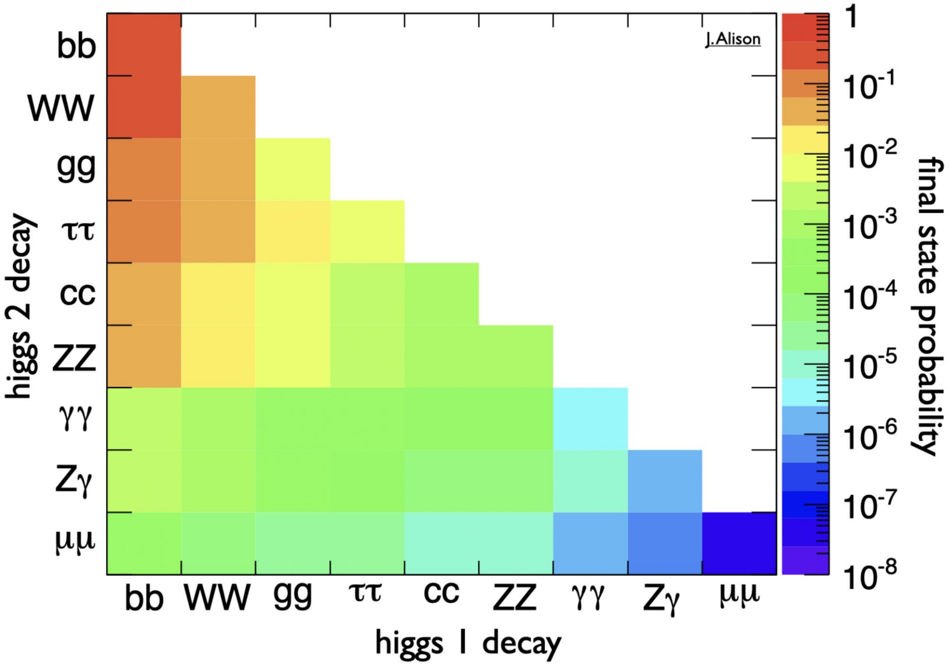


Source: G. Aad-PhysRevD.108.052003

Outline

- The ATLAS Experiment
- Why Study HH?
  - **Why Boosted ggF HH→4b?**
- Reconstruction Strategies and Missing Muons
- Quantifying Mass Sensitivity

Why HH→4b?



Why 4b?

- H→2b is the highest cross section decay channel for higgs with 58.2% branching ratio → HH4b most populated channel for HH

But

- Noisy channel as many other processes have bb end states:
  - QCD Jets
    - Top quark pair production
    - Jets from strong interaction
  - Light-quark-initiated jets misidentified as originating from heavy quarks

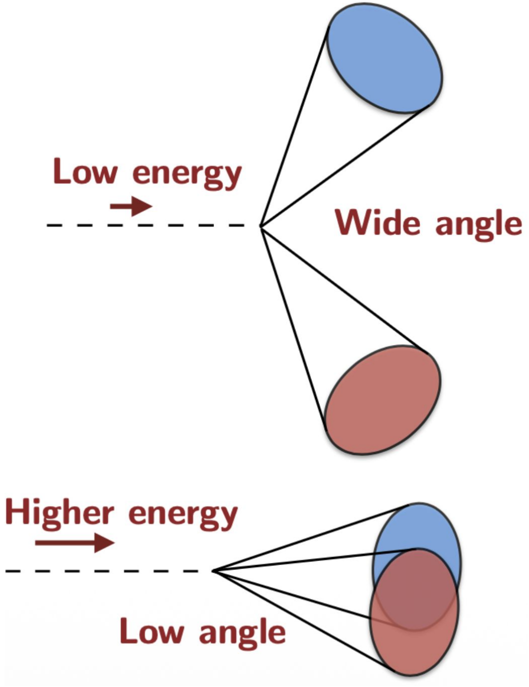
Source: G. Aad-PhysRevD.108.052003, John Alison (CMS)

Why Boosted HH→4b?

Sidestep sensitivity limitations by considering boosted regime

- At higher P_T ranges, jets are more attenuated

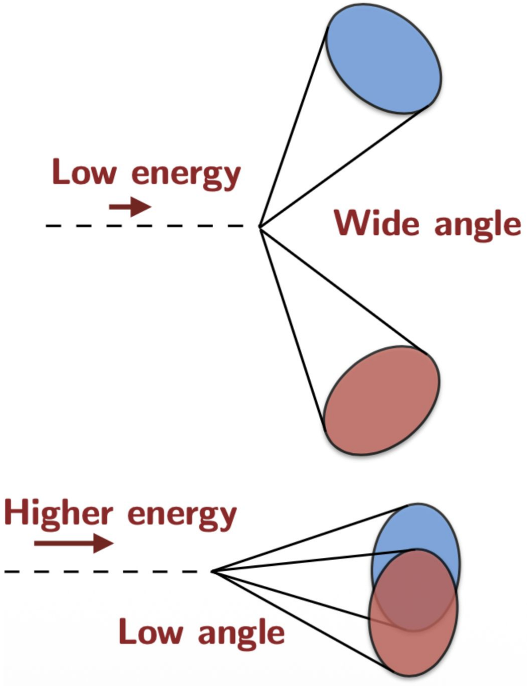
- Resolution becomes difficult, so grouping into two large R jets yields better results



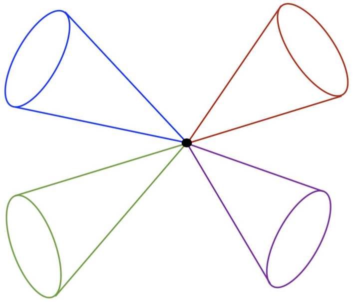
Source: G. Aad-PhysRevD.108.052003

Why Boosted HH→4b?

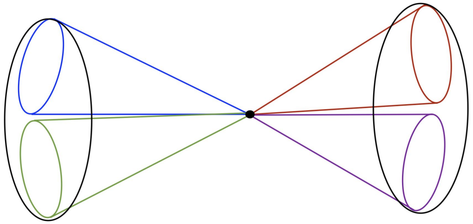
- At higher P_T ranges, jets are more attenuated
- Resolution becomes difficult, so grouping into two large R jets yields better results



Previously:
Separately resolve 4 bottom jets



Boosted Regime:
Tag 2 boosted large radius jets



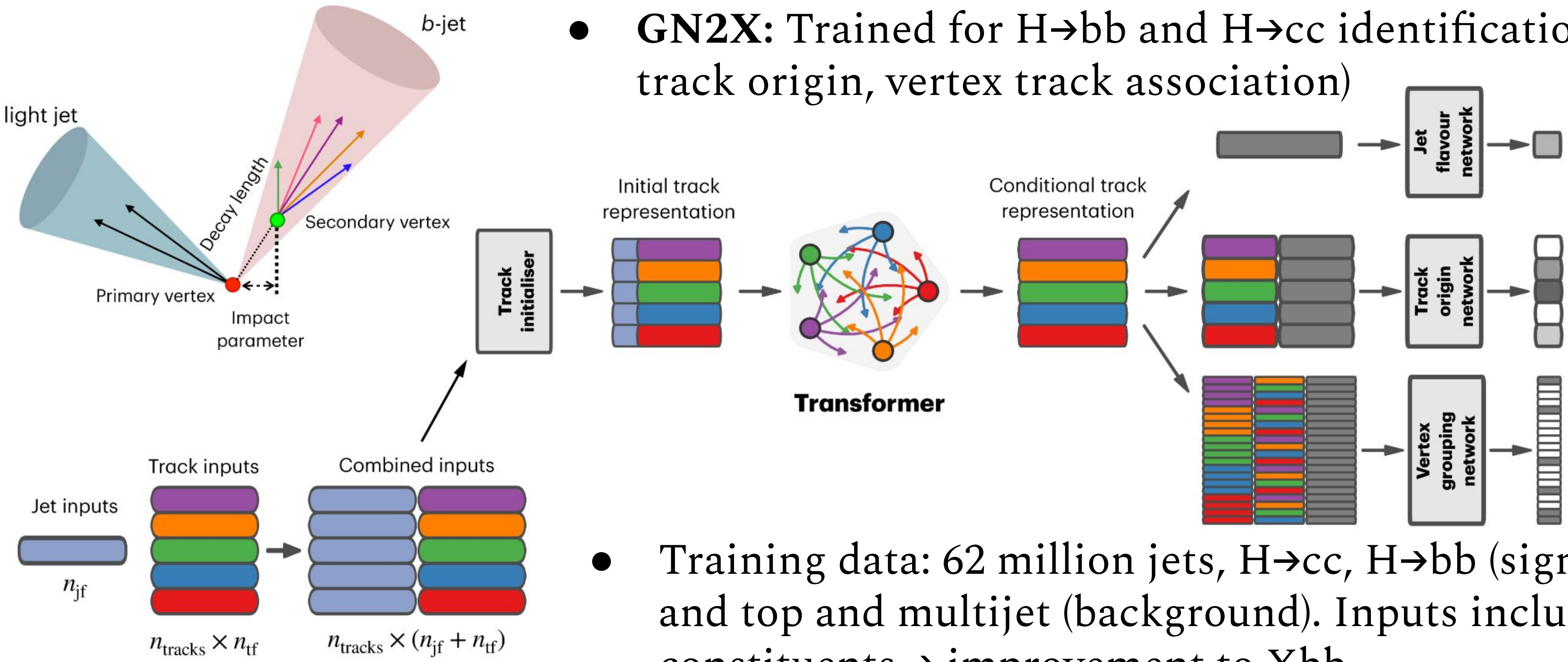
Source: G. Aad-PhysRevD.108.052003

Outline

- The ATLAS Experiment
- Why Study HH?
    - Why Boosted ggF $HH \rightarrow 4b$?
- **Reconstruction Strategies and Missing Muons**
- Quantifying Mass Sensitivity

Reconstruction Strategies and Missing Muons

- **GN2X:** Trained for H→bb and H→cc identification (and track origin, vertex track association)

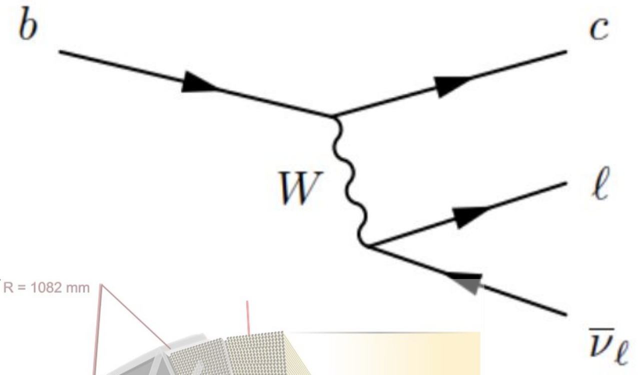


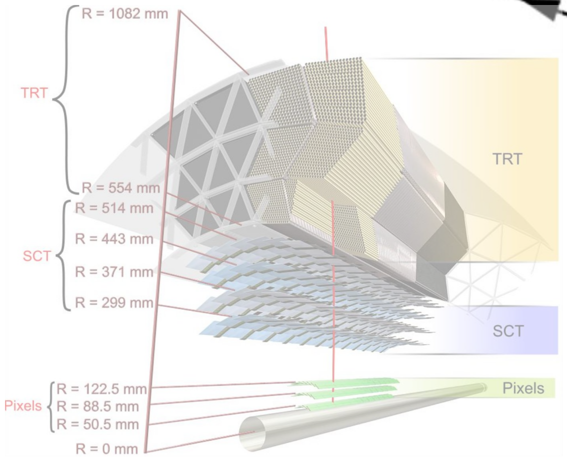
- Training data: 62 million jets, H→cc, H→bb (signal) and top and multijet (background). Inputs include jet constituents → improvement to Xbb

Source: FTAG-2023-05

Reconstruction Strategies and Missing Muons

GN2X is a good improvement from Xbb, but...





Missing Muons in Jet Reconstruction

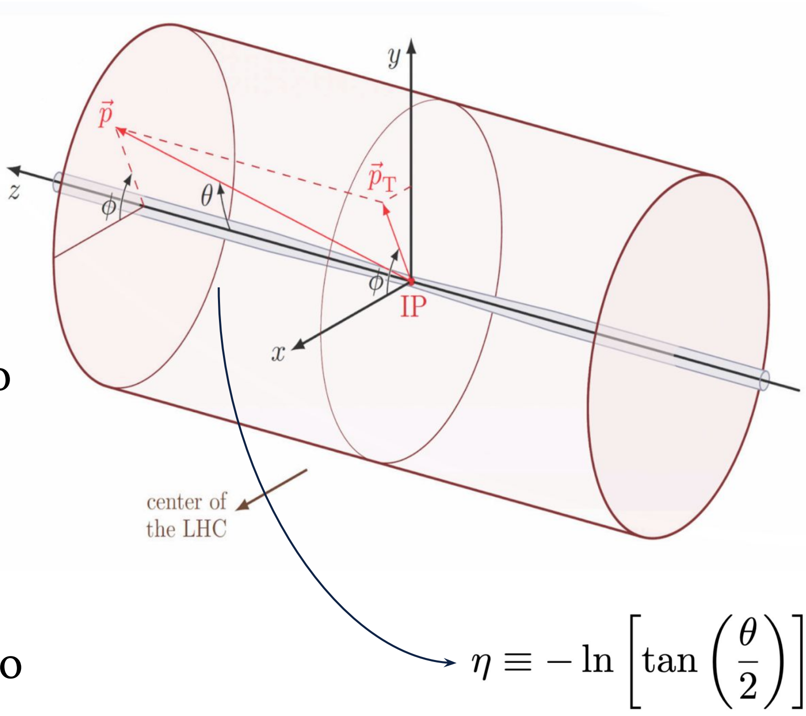
- Muons are radiated via semileptonic decays within b jets ~ 10%
- Muons are "lost" → not associated with the jet they originated in, happens mainly in data readout
- Jet energy consistently underestimated as energy associated with muons are not included in reconstruction calculations

Source: Marco Battaglia, Evripidis Koutsioumpas, Andrea Sciandra, CERN HFLAV

Reconstruction Strategies and Missing Muons

Solution: Add Back Missing Energy from Muons in Jets (MIJ)

- "Simple" fix - match muons with jets internal to large-R jet with $\Delta R = \sqrt{\phi^2 + \eta^2}$
  - Then add this missing energy back to that jet
  - Note only muon energy is accounted for, not the energy for ν_μ
- More complicated fix: Use machine learning to find regressed values for jet kinematics

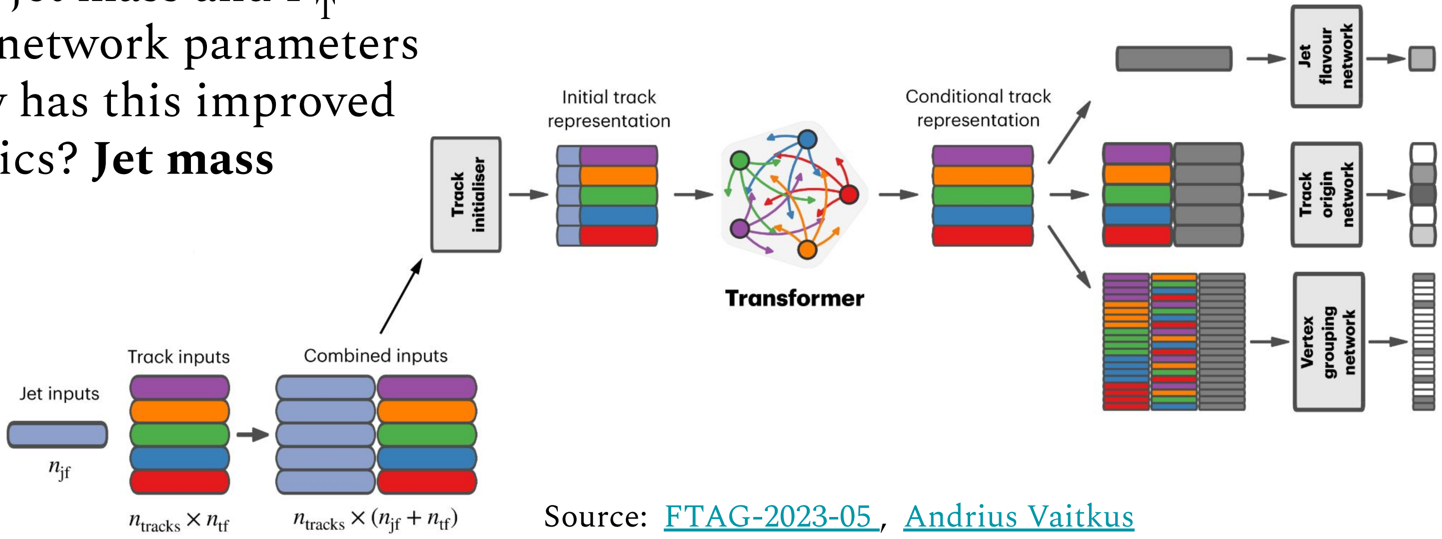


Source: Marco Battaglia, Evripidis Koutsioumpas, Andrea Sciandra, ATLAS-OUTREACH-2025-016

Reconstruction Strategies and Missing Muons

ML Regressions for Jet Mass and P_T

- Fixes MIJ problem + others
- Trained on some of the data used for GN2X
  - H->bb, H->cc, QCD (but no top) – 180 million jets, Trained for jet mass and P_T
  - 5.6 million network parameters
- My project: how has this improved jet kinematics? **Jet mass**



Source: FTAG-2023-05 , Andrius Vaitkus

Outline

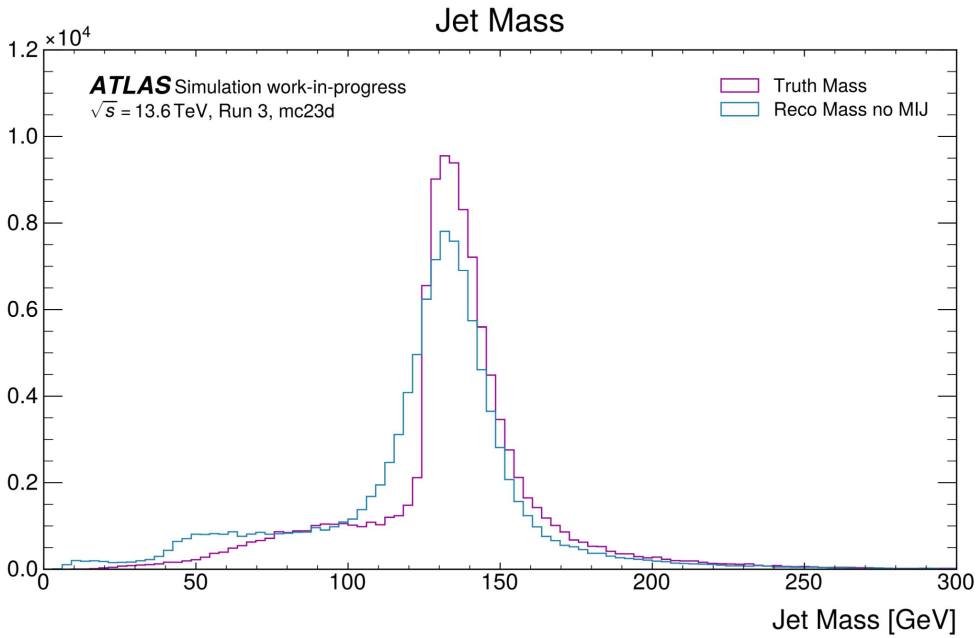
- The ATLAS Experiment
- Why Study HH?
  - Why Boosted ggF HH→4b?
- Reconstruction Strategies and Missing Muons
- **Quantifying Mass Sensitivity**

Comparing Mass Sensitivity

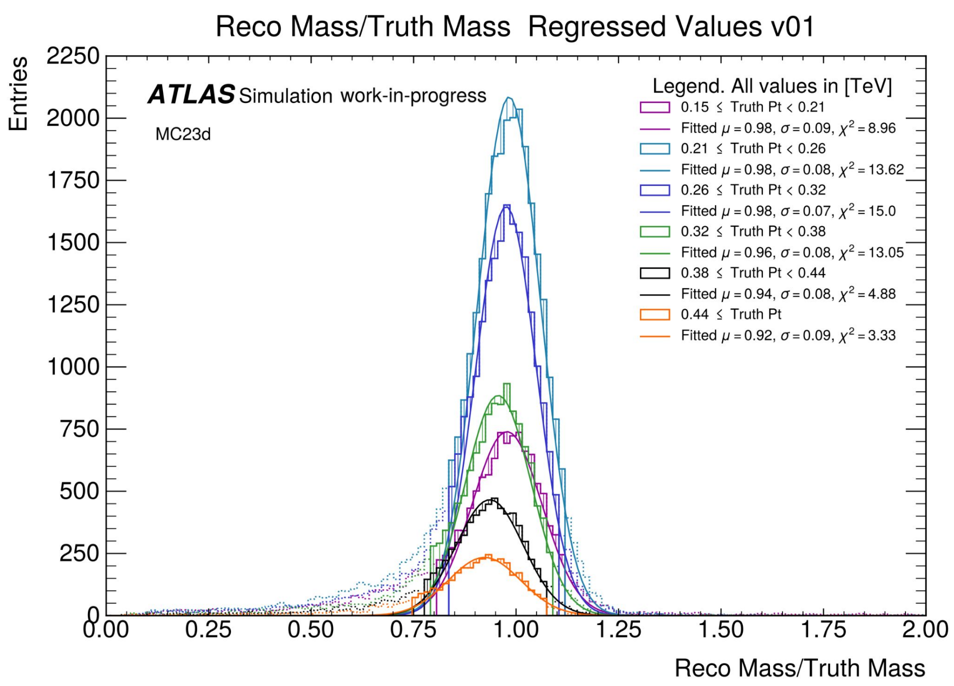
Strategy:

- Used simulated Run 3 signal HH→4b samples [50,000 jets] to compare:
  - Baseline reconstruction
  - Muon-in-jet correction added
  - Regressed values V01: Trained on data **<u>with</u>** muons in jets

Take a ratio of these two (between matched jets)



Comparing Mass Sensitivity

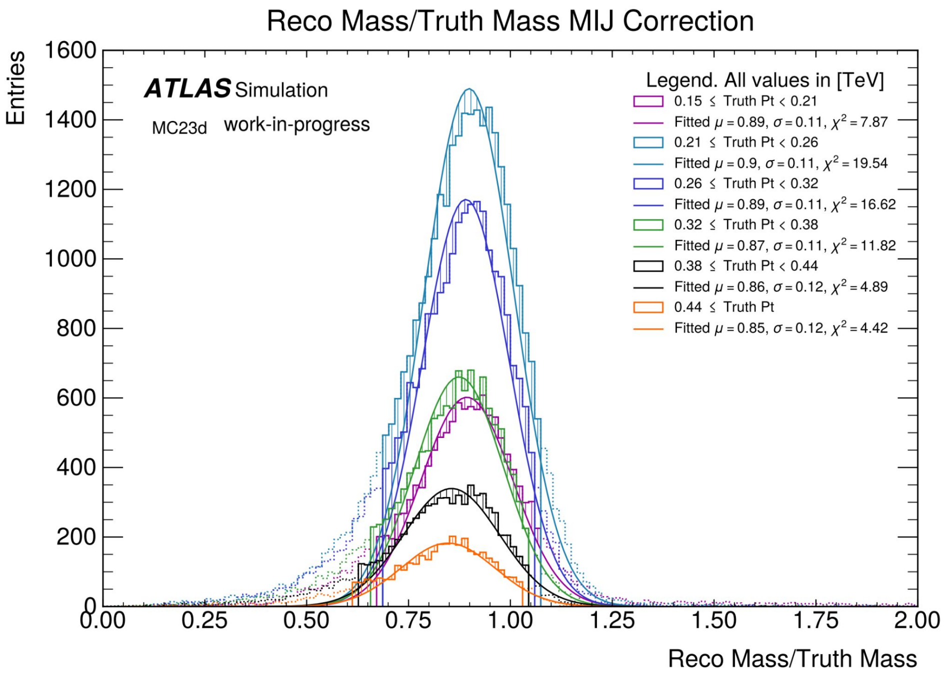


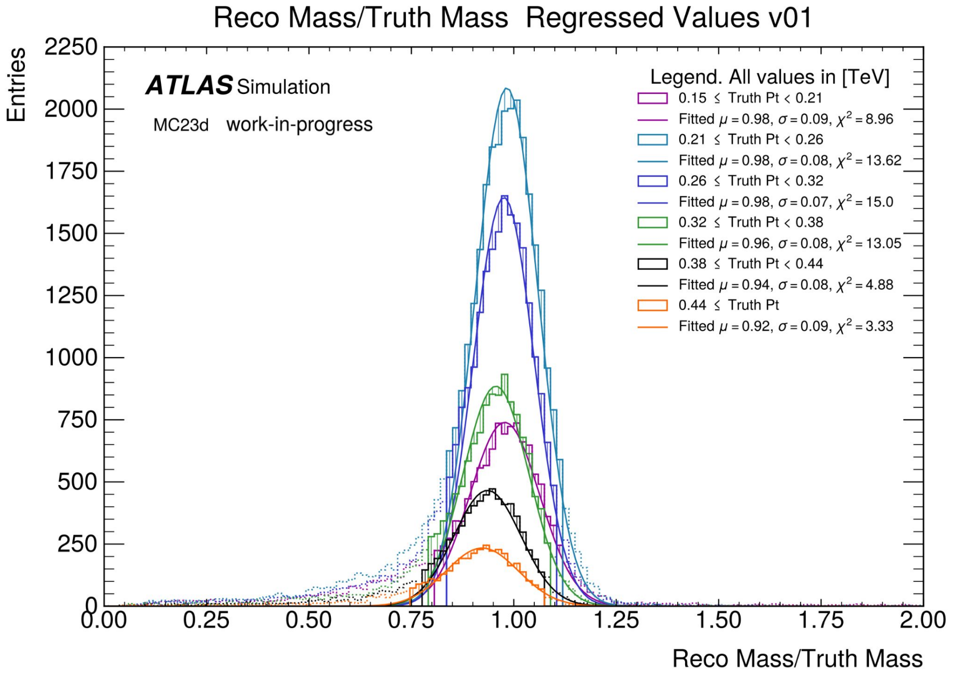
Procedure:

- Require at least two reconstructed jets in each event
- Required true combined HH mass of 0.5 TeV
- Ensured that truth reconstructed jets had minimal distance in ϕ - η space (truth-reco consistency) by minimizing $\Delta R = \sqrt{\phi^2 + \eta^2}$ (built-in for Newer Data)
- Plotted $\frac{Mass_{reco}}{Mass_{true}}$ partitioned by true P_T

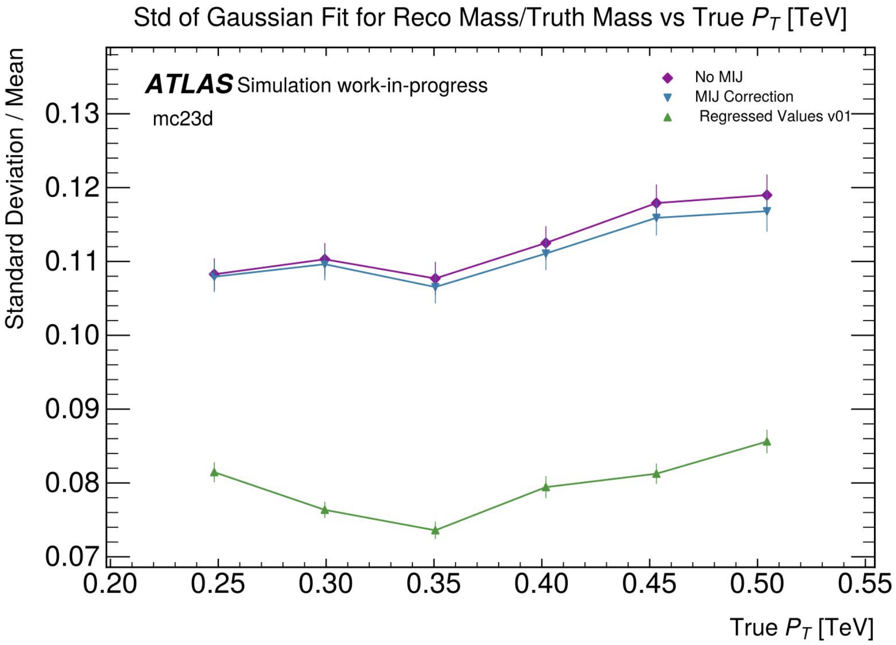
Comparing Mass Sensitivity

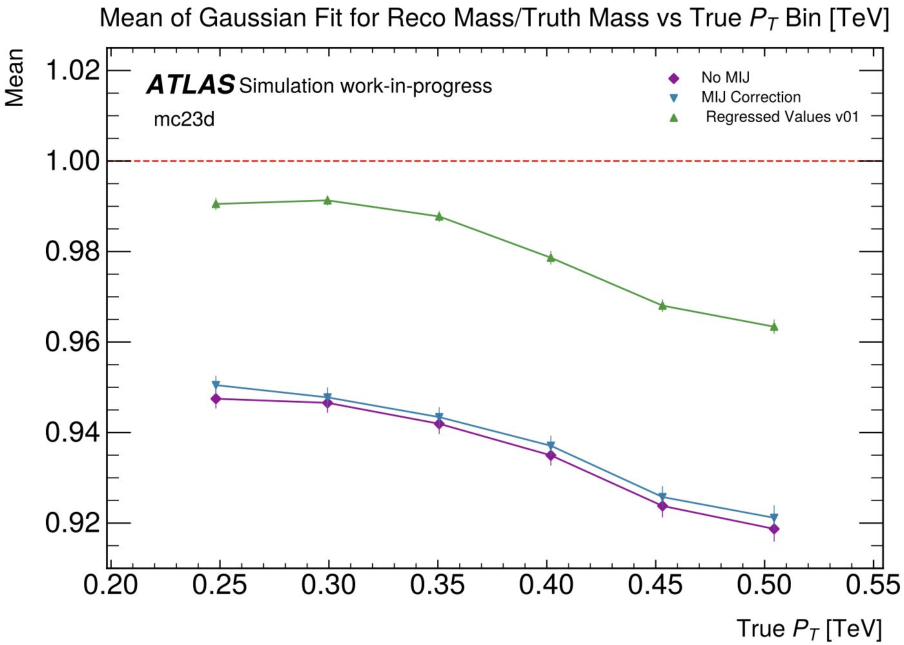
- Tails are long and asymmetrical, so truncate the central curve and fit a gaussian. Discard bins with counts smaller than $RMS = \left[\frac{1}{N_{bins}} \sum (counts)^2\right]^{1/2}$ (sensitive to binning)





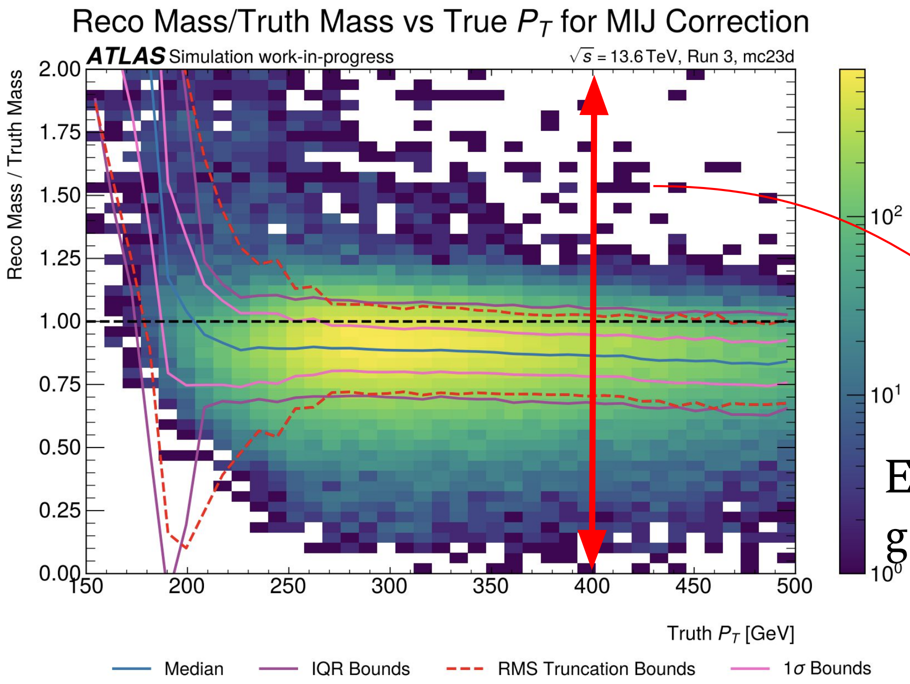
Comparing Mass Sensitivity

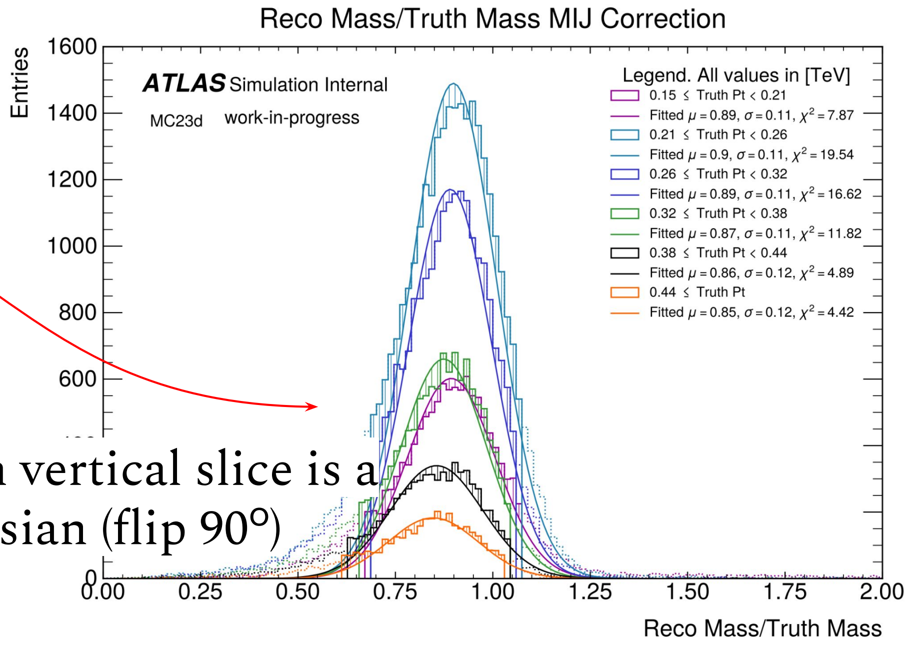




Comparing Mass Sensitivity

- Using finer P_T binning, median and interquartile range (IQR) to capture contribution on tails

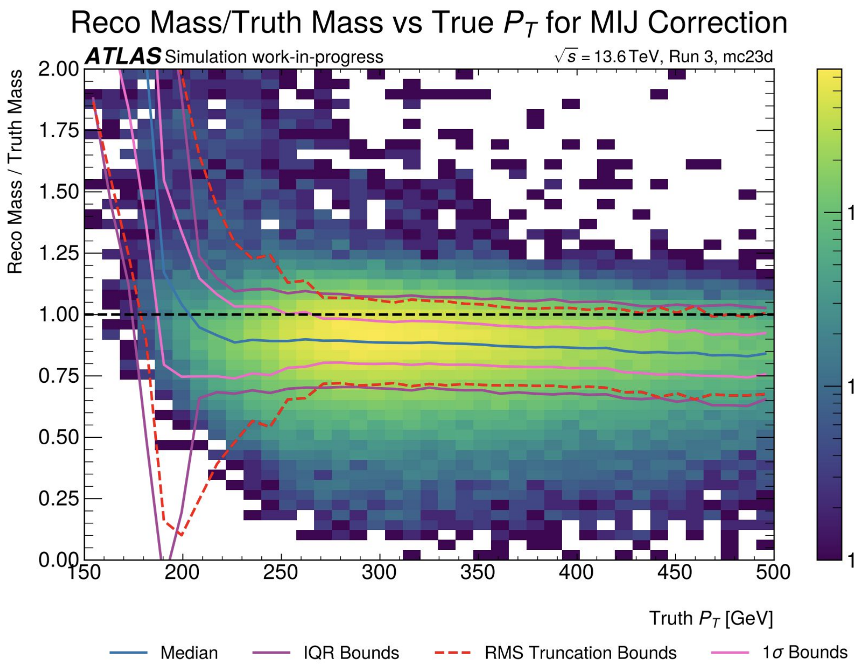


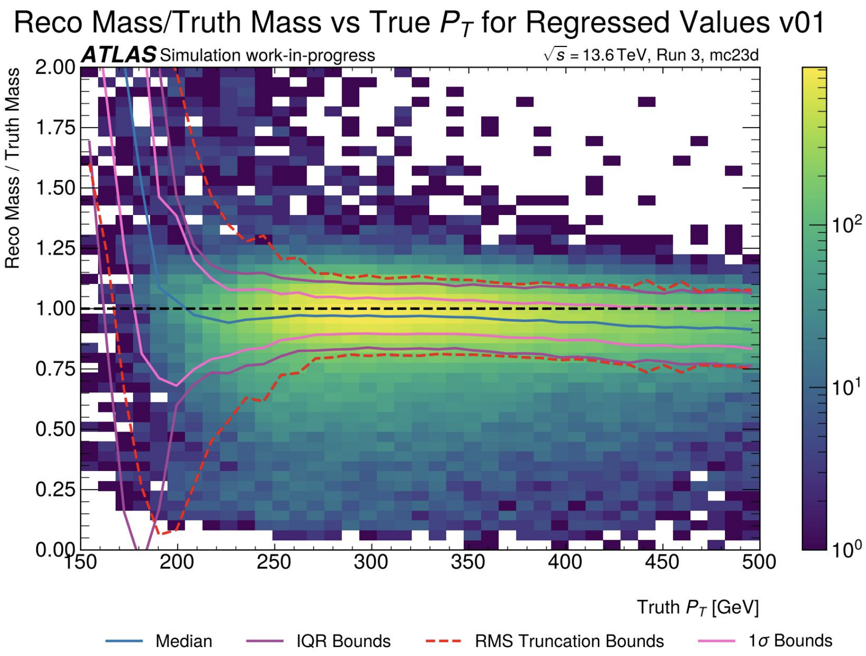


Each vertical slice is a gaussian (flip 90°)

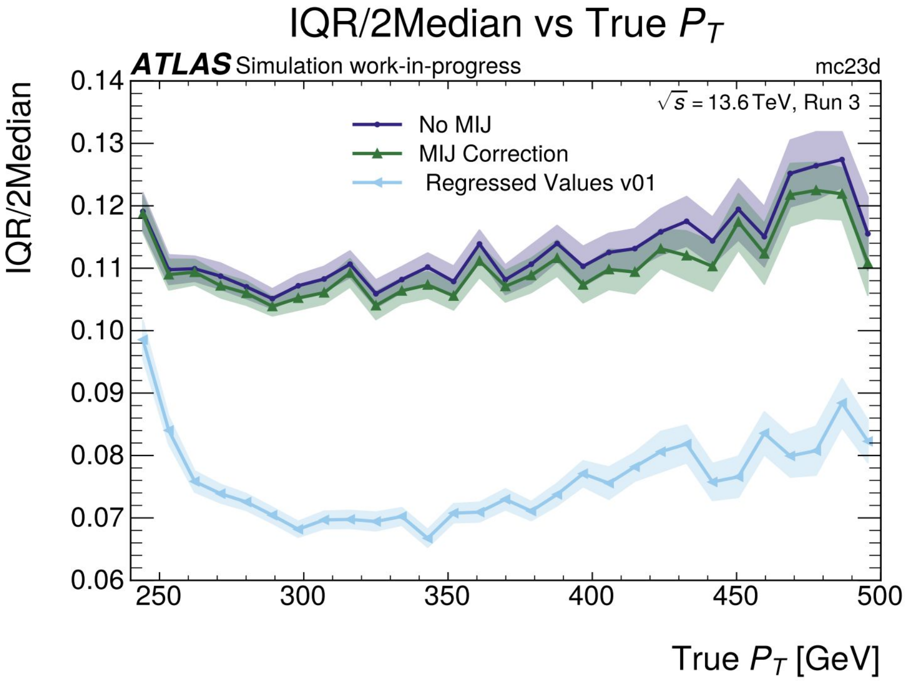
Comparing Mass Sensitivity

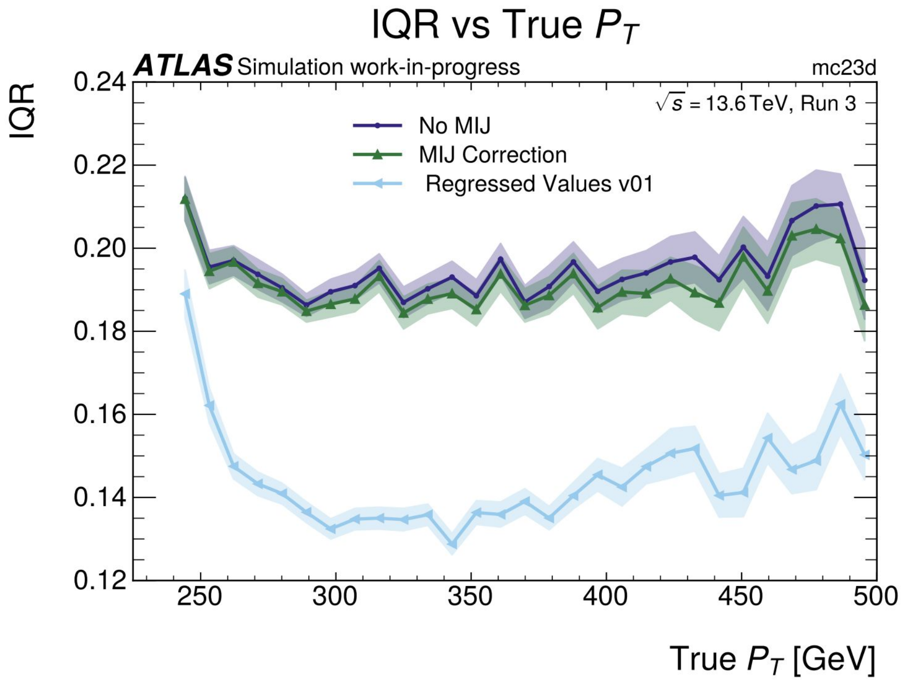
- Using finer P_T binning, median and IQR to capture contribution on tails





Comparing Mass Sensitivity



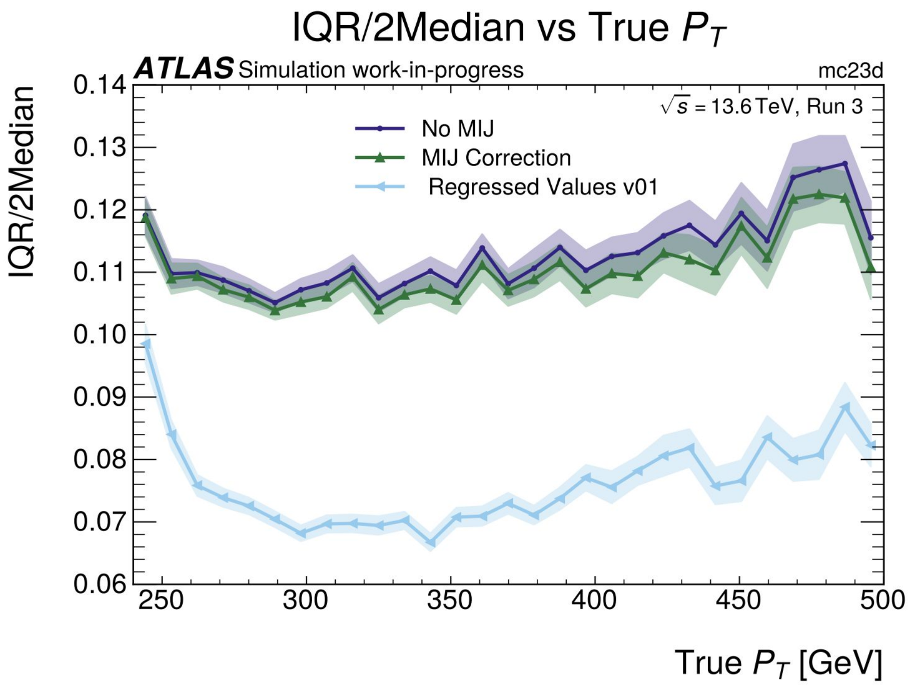


Conclusion

- We study HH4b to characterize the Higgs potential by measuring coupling constants: $\kappa_\lambda = \frac{\lambda_{HHH}}{\lambda_{HHH}^{SM}}$ → answers to even larger questions, BSM scenarios, etc.

- Accurate reconstruction of jet kinematics, including jet mass, is needed to increase sensitivity to detect events with small cross sections → HH4b

- Demonstrated improvement to mass sensitivity by adding missing MIJ energy to jets. Using ML regression for jet mass works even better



References

Figures

- ATLAS, ALICE, CMS, LHCb images: atlas.cern, home.cern, cms.cern, home.cern
- General Diagrams: ATLAS Twiki, CERN
- Event Displays: ATLAS Twiki
- ATLAS Coordinate System: ATLAS-OUTREACH-2025-016
- GN2X Schematic: FTAG-2023-05

Papers and Presentations:

- Muon-In-Jet Correction for the Inclusive/ggH boosted Hbb Analysis - Marco Battaglia, Evripidis Koutsioumpas, Andrea Sciandra

- Performance of Large- R Jet Regression: bJ10v00Ext, bJR10v01 - Andrius Vaitkus

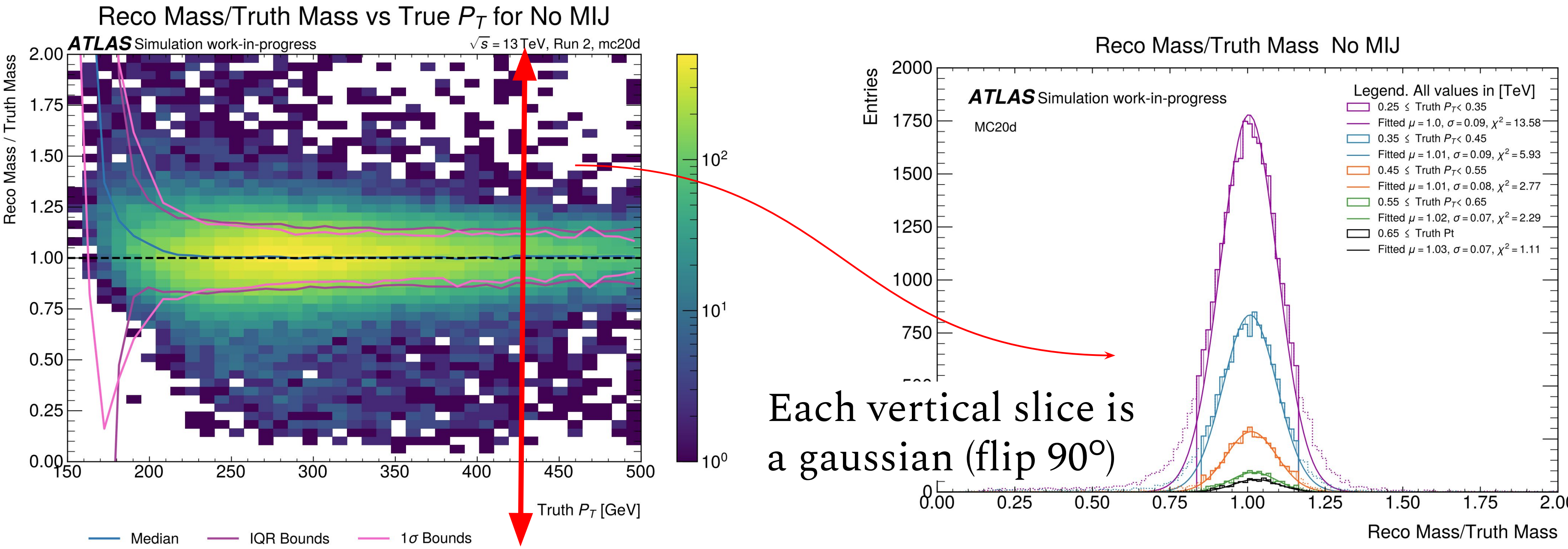
- Search for nonresonant pair production of Higgs bosons in the bbbb final state in pp collisions at sqrt{s}=13 TeV with the ATLAS detector - G. Aad-PhysRevD.108.052003

- How Pairs of Higgs Bosons Help Us Understand the Standard Model and Beyond - M.Swiatlowski, 2024

Backup

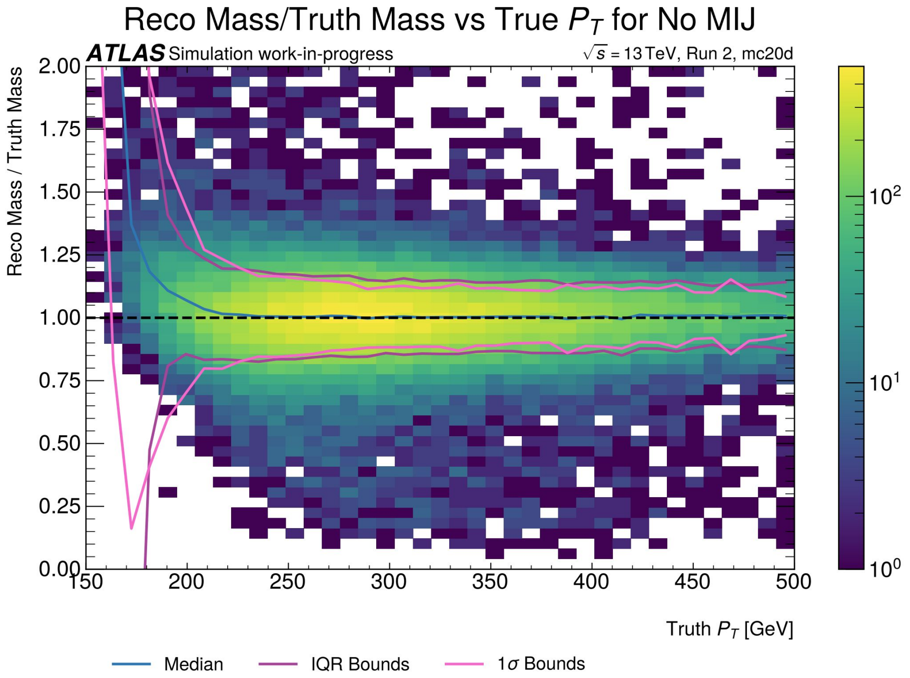
Comparing Mass Sensitivity

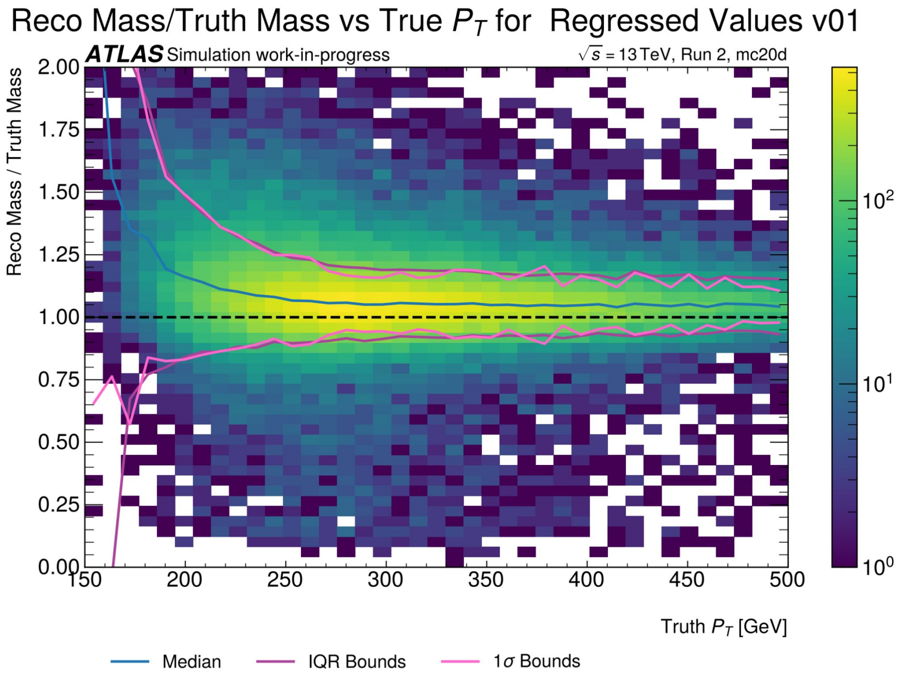
- Using finer PT binning, median and IQR to capture contribution on tails
- Truth in the denominator here does **not** include muons



Comparing Mass Sensitivity

- Using finer PT binning, median and IQR to capture contribution on tails
- Truth in the denominator here does **not** include muons

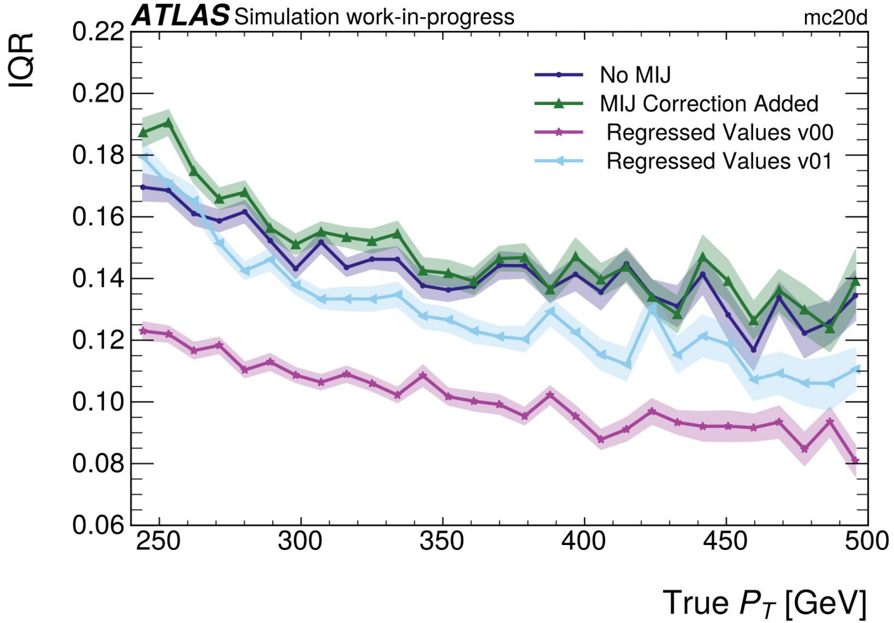




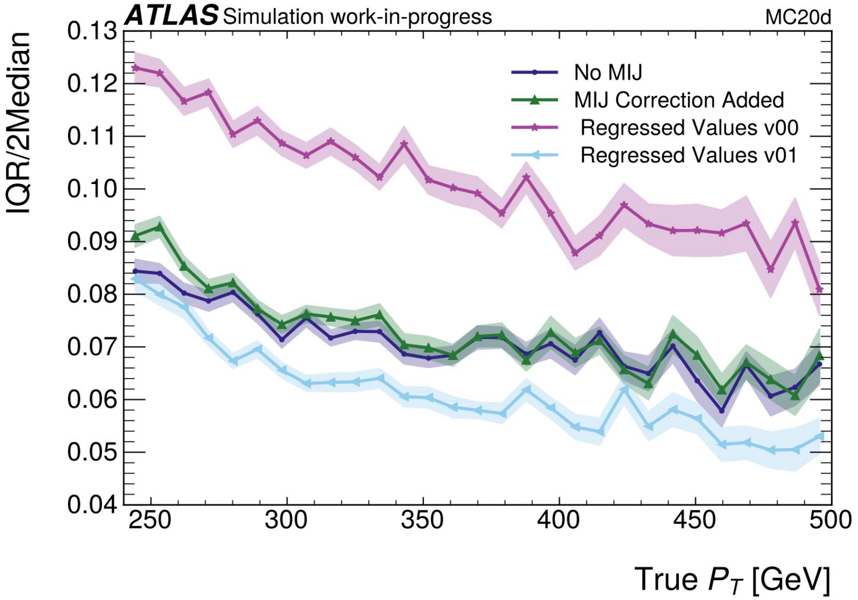
Comparing Mass Sensitivity (Older Data)

- Using finer PT binning, median and IQR to capture contribution on tails
- Truth in the denominator here does **not** include muons

IQR vs True Pt

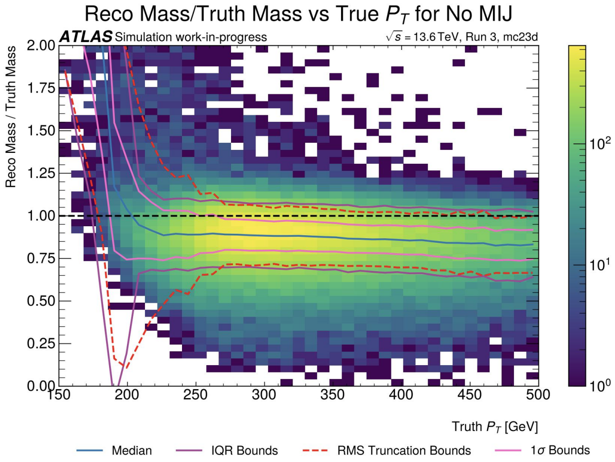


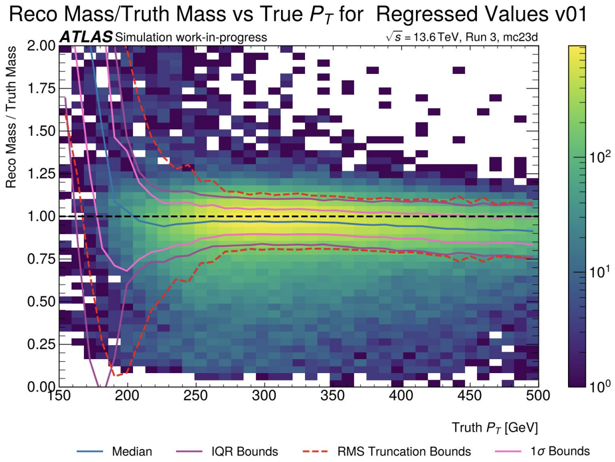
IQR/2Median vs True Pt

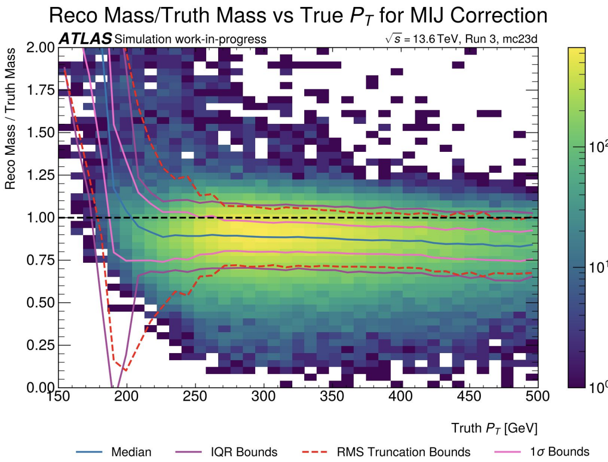


Comparing Regressed Mass

- Truth in the denominator here **does** include muons

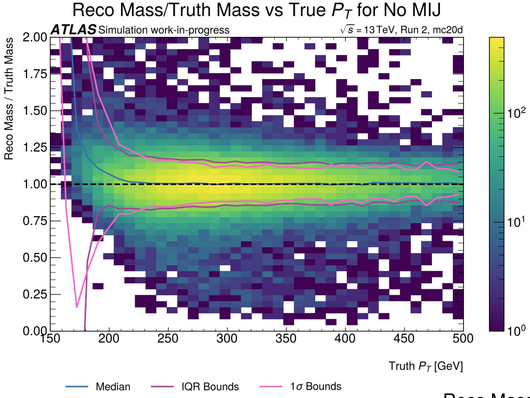


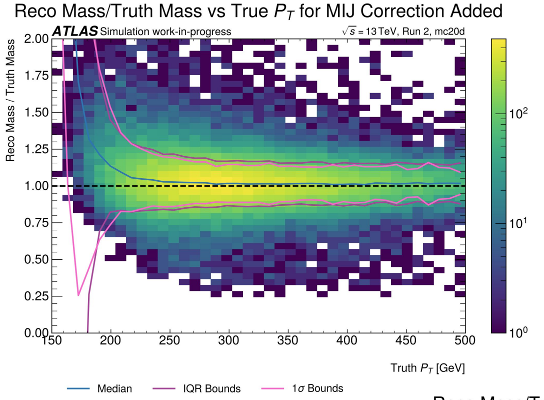


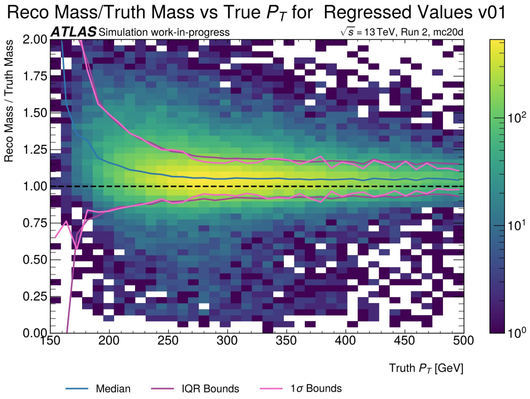


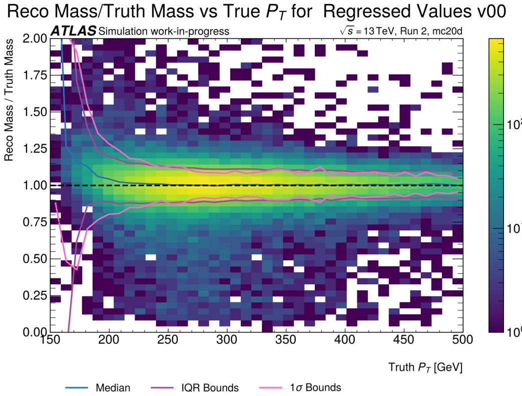
Comparing Regressed Mass (Older Data)

- Truth in the denominator here does **not** include muons

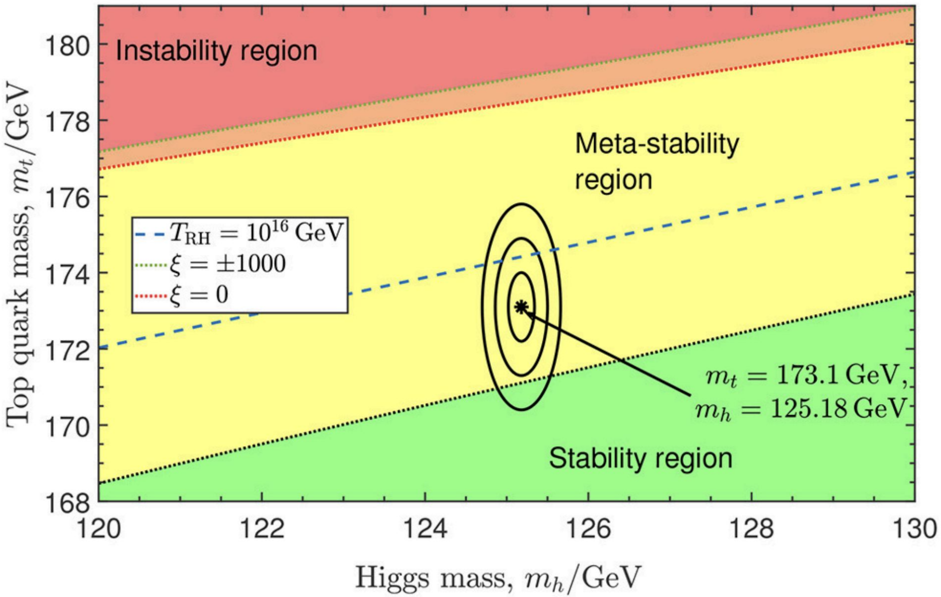








(In)stability of the universe



Source: Markkanen, Rajantie, Stopyra

GN2X Model Full Inputs

Jet Input	Description
p_T	Large-R jet transverse momentum
η	Signed large-R jet pseudorapidity
mass	Large-R jet mass

Track Input	Description
q/p	Track charge divided by momentum (measure of curvature)
$d\eta$	Pseudorapidity of track relative to the large-R jet η
$d\phi$	Azimuthal angle of the track, relative to the large-R jet ϕ
d_0	Closest distance from track to primary vertex (PV) in the transverse plane
$z_0 \sin\theta$	Closest distance from track to PV in the longitudinal plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0 \sin\theta)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
subjetIndex	Integer label of which subjet track is associated to (GN2X + Subjets only)

Subjet Input	Description (Used only in GN2X + Subjets)
p_T	Subjet transverse momentum
η	Subjet signed pseudorapidity
mass	Subjet mass
energy	Subjet energy
$d\eta$	Pseudorapidity of subjet relative to the large-R jet η
$d\phi$	Azimuthal angle of subjet relative to the large-R jet ϕ
GN2 p_b	b-jet probability of subjet tagged using GN2
GN2 p_c	c-jet probability of subjet tagged using GN2
GN2 p_u	light flavour jet probability of subjet tagged using GN2

Flow Input	Description (Used only in GN2X + Flow)
p_T	Transverse momentum of flow constituent
energy	Energy of flow constituent
$d\eta$	Pseudorapidity of flow constituent relative to the large-R jet η
$d\phi$	Azimuthal angle of flow constituent relative to the large-R jet ϕ

Source: ATL-PHYS-PUB-2023-021