A Search for Resonant and Non-Resonant Di-Higgs Production in the $\gamma\gamma$bb Channel Using the ATLAS Detector

Dissertation Defense
Tyler James Burch
March 26, 2020

Northern Illinois University

Committee:
Jahred Adelman (Chair)
Dhiman Chakraborty
Michael Eads

Welcome!

Thanks for staying flexible!

How this defense will proceed:

- Will start with this presentation (~1 hour) in this room
  - Open Zoom meeting, plus cross-streaming to YouTube
- After, committee and I will transition to a separate room to conclude
- Once we've finished I'll return to this room, and the committee will deliberate and return to this room afterwards

Some general guidelines:

- Please mute if not speaking
- Given Brianna Stamas room host privileges to manage low-level Zoom concerns. Thanks Bri!

Thank you for attending!

Northern Illinois University

Outline

Introduction

- The Standard Model
- The LHC and ATLAS Detector
- Di-Higgs Production

HH→$\gamma\gamma$bb search using 2015+2016 data

VBF HH→$\gamma\gamma$bb

- Motivation
- Integrating into $\gamma\gamma$bb analysis

Photon Identification Optimization

- Background and current methods
- Topological Cluster inputs
- MVA Approaches

Conclusion

Introduction

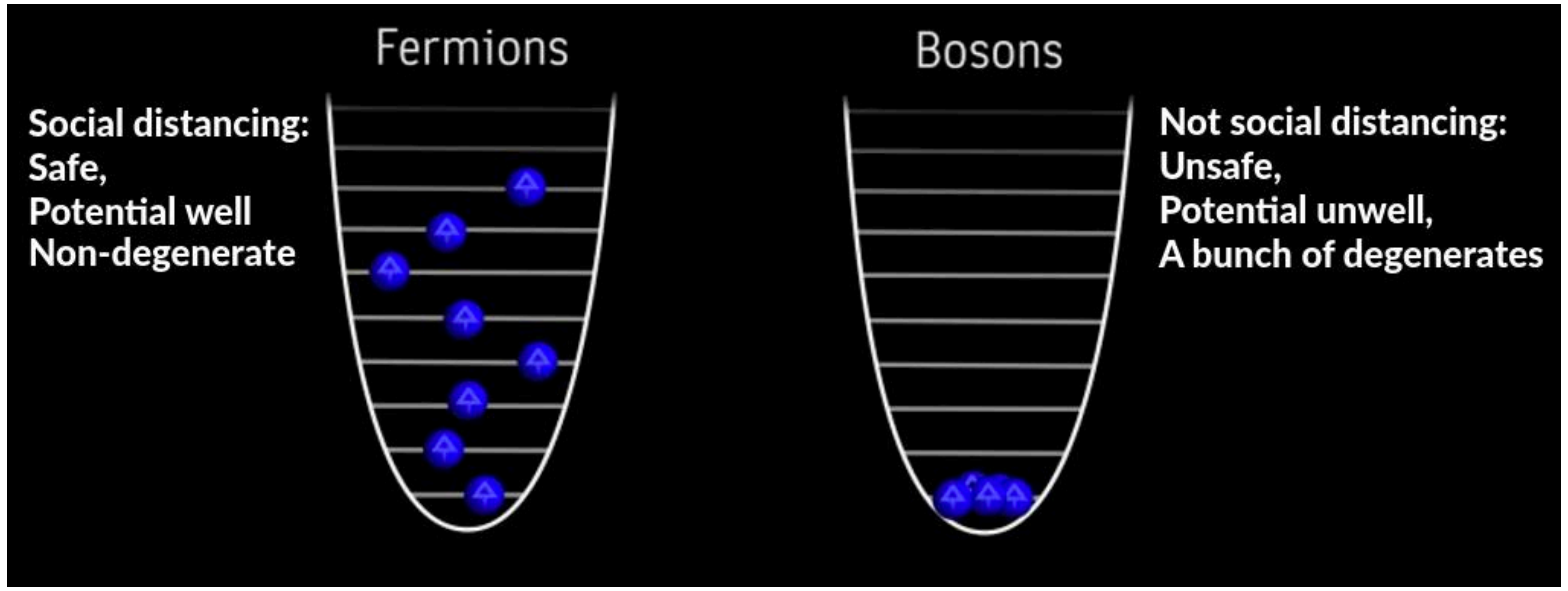
The Standard Model (SM)

wikipedia.org

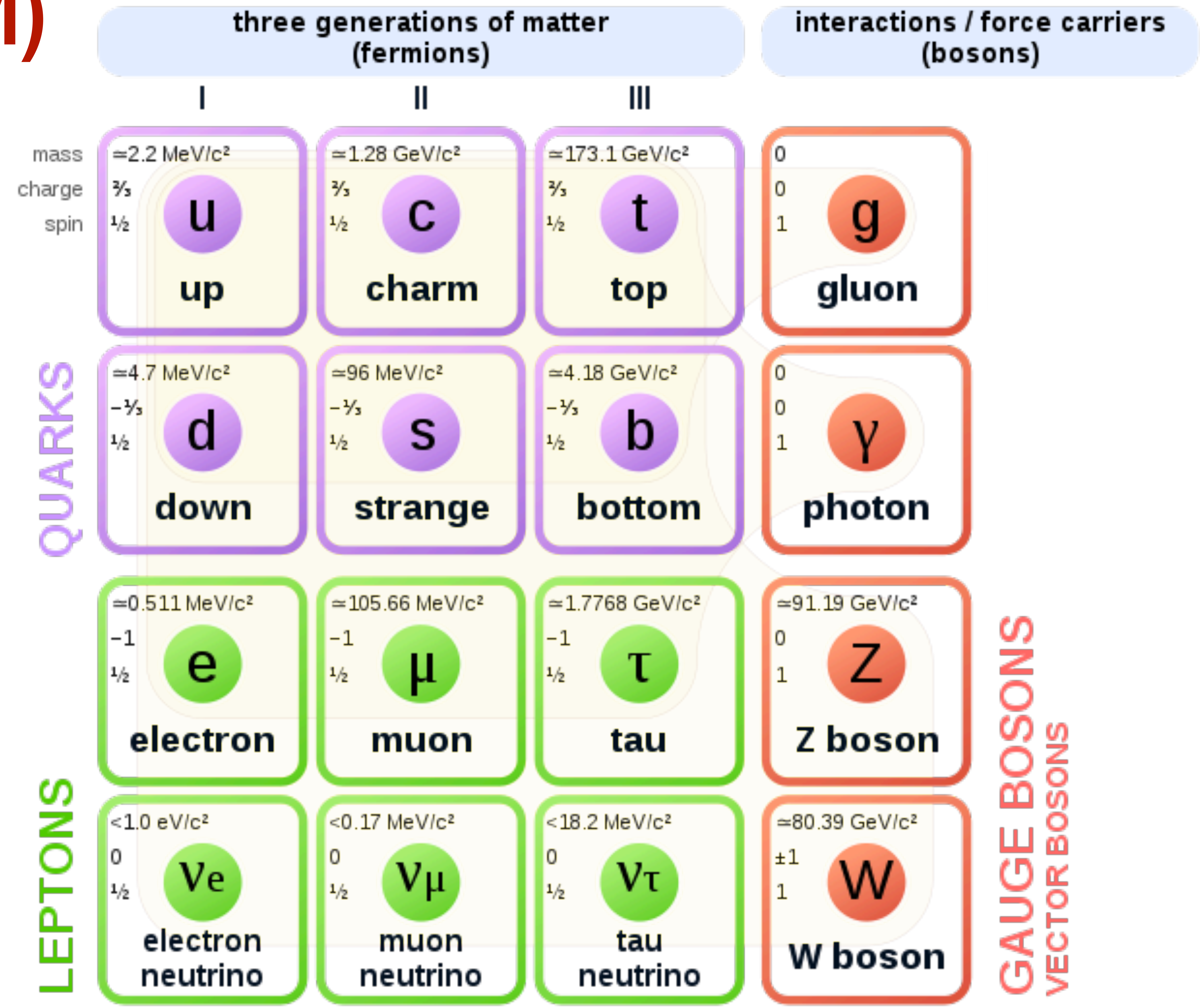
Best working description of the fundamental particles and their interactions

Fermions - particles that “take up space.”
(Precisely, they obey the Pauli exclusion principle)

Bosons - force carriers, explain interactions
- Photon → Electromagnetic force
- $W^{\pm}$ and Z → Weak nuclear force
- Gluon → Strong nuclear force

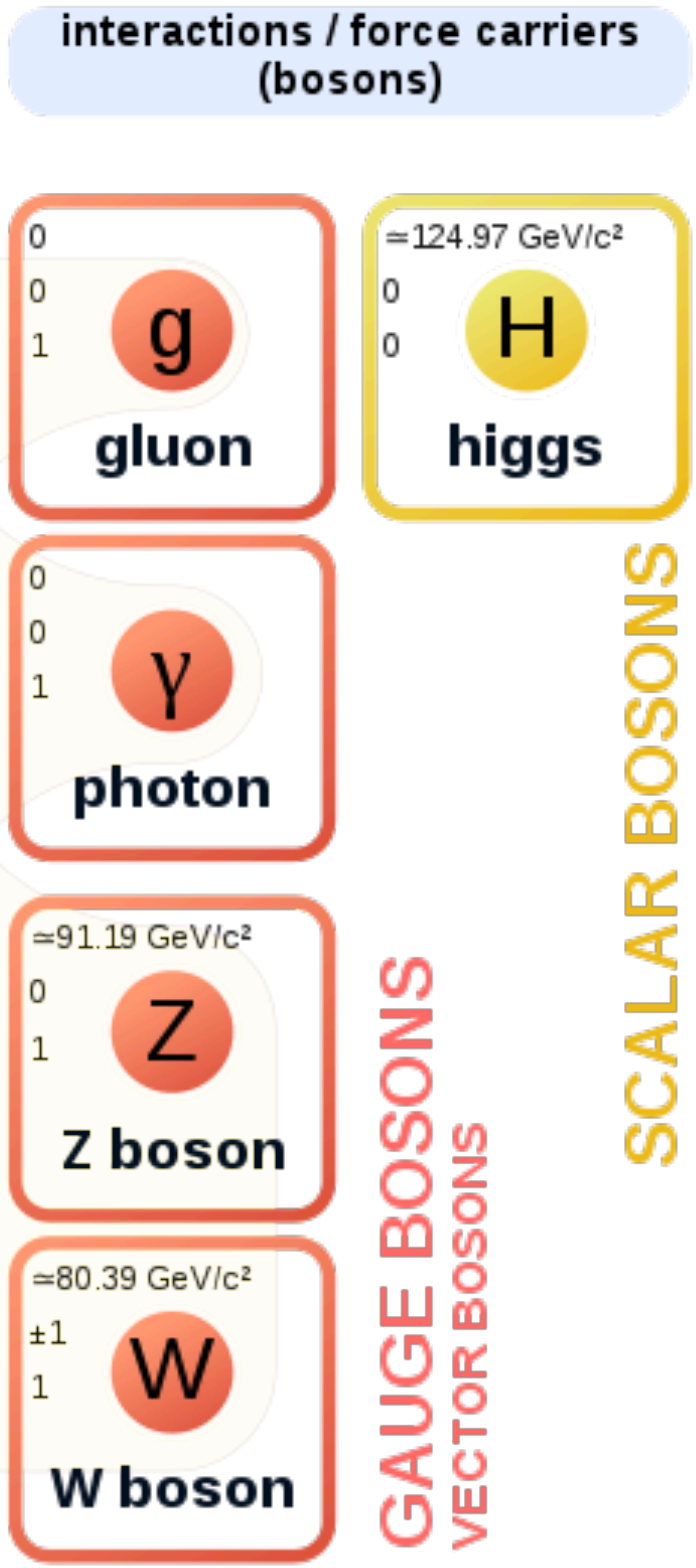


Westerdale



Northern Illinois University

The missing piece: the Higgs Boson



Electroweak Theory

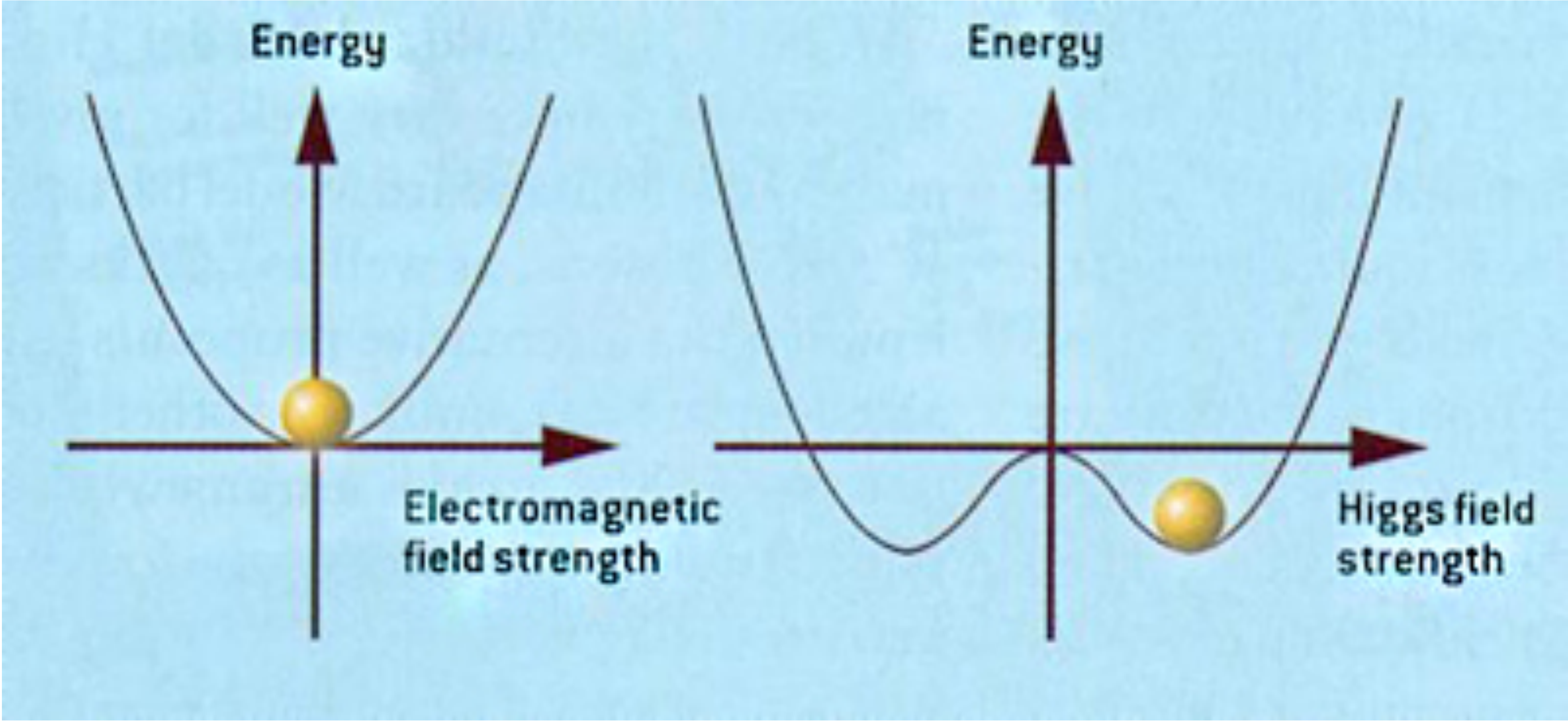
Weak and electromagnetic force are manifestations of the same “electroweak" force, which split as universe cooled (under ~160 GeV)

- W and Z gained mass while photon remained massless… but how?

Brout-Englert-Higgs

Interactions with a field present everywhere in space leads to mass

- Non-zero vacuum expectation value!



enigmass

Northern Illinois University

Small particles require large solutions…

The Large Hadron Collider

27 kilometers in circumference, situated on the Franco-Swiss border

Primarily *pp* collisions, accelerated to 99.9999990% the speed of light, collide at 4 primary points where detectors are situated

Center-of-mass energy of 13 TeV (6.5 TeV per beam)

CERN

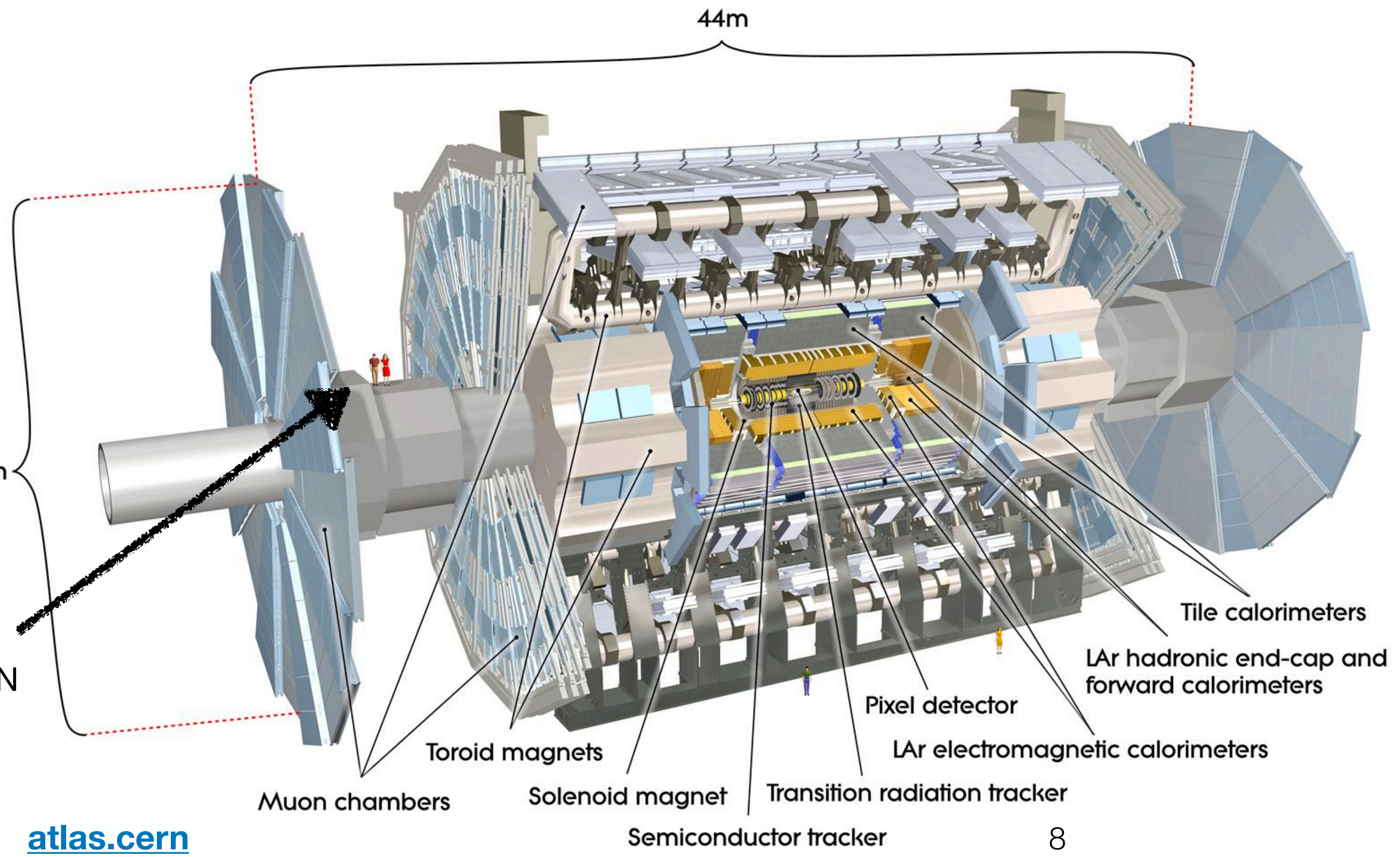
Small particles require large solutions...

The ATLAS Detector

General purpose detector, aiming to measure signals resulting from *pp* collisions to cover a vast range of analyses
- Beyond just the Higgs program, searches for supersymmetry, dark matter, precision measurements, etc.

Collaboration of over 3,000 scientists from all over the world

Composed of several subsystems, each aiming to collect specific information about collisions



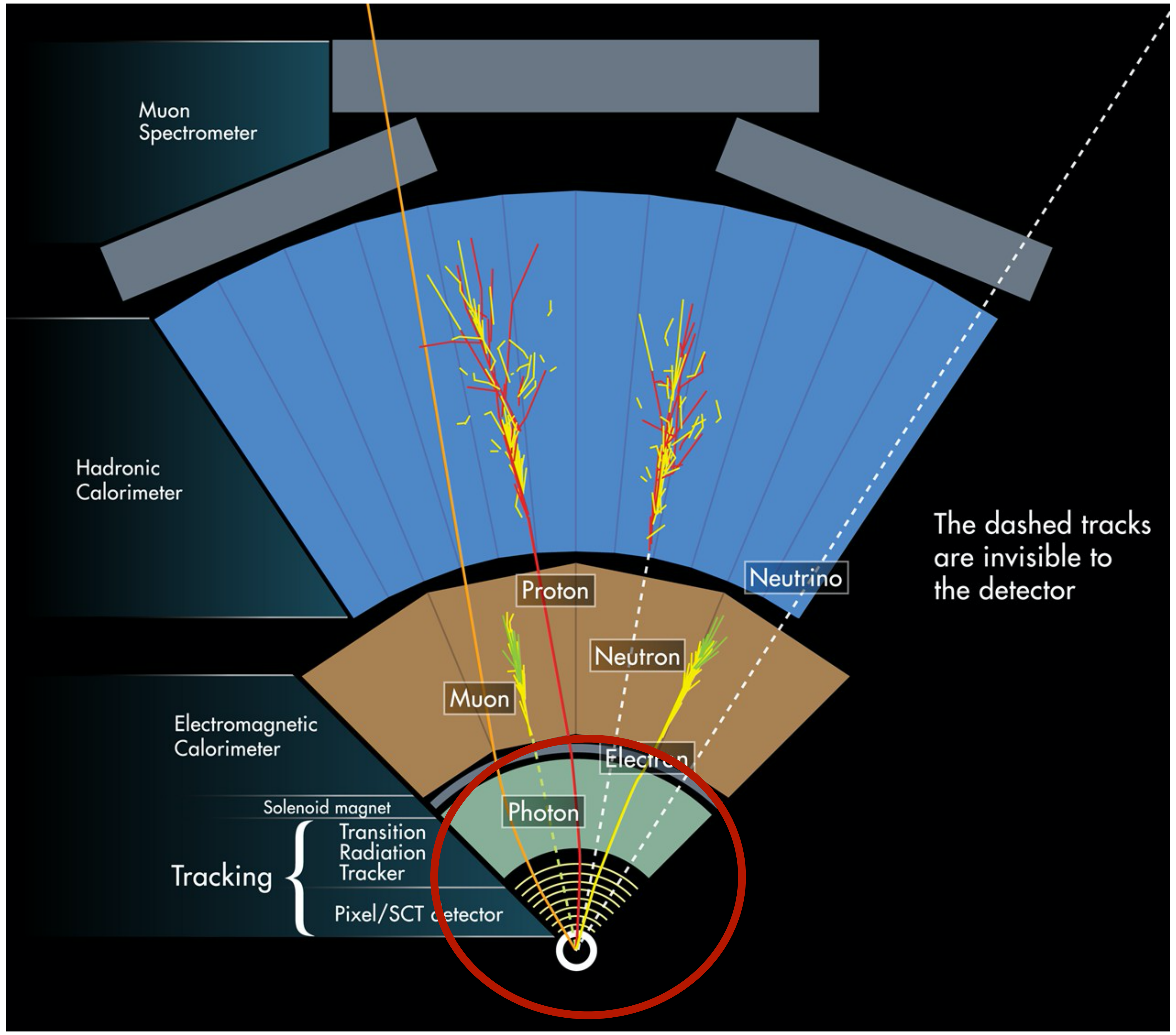
Average human height (for scale only, CERN safety would not approve of this)

atlas.cern

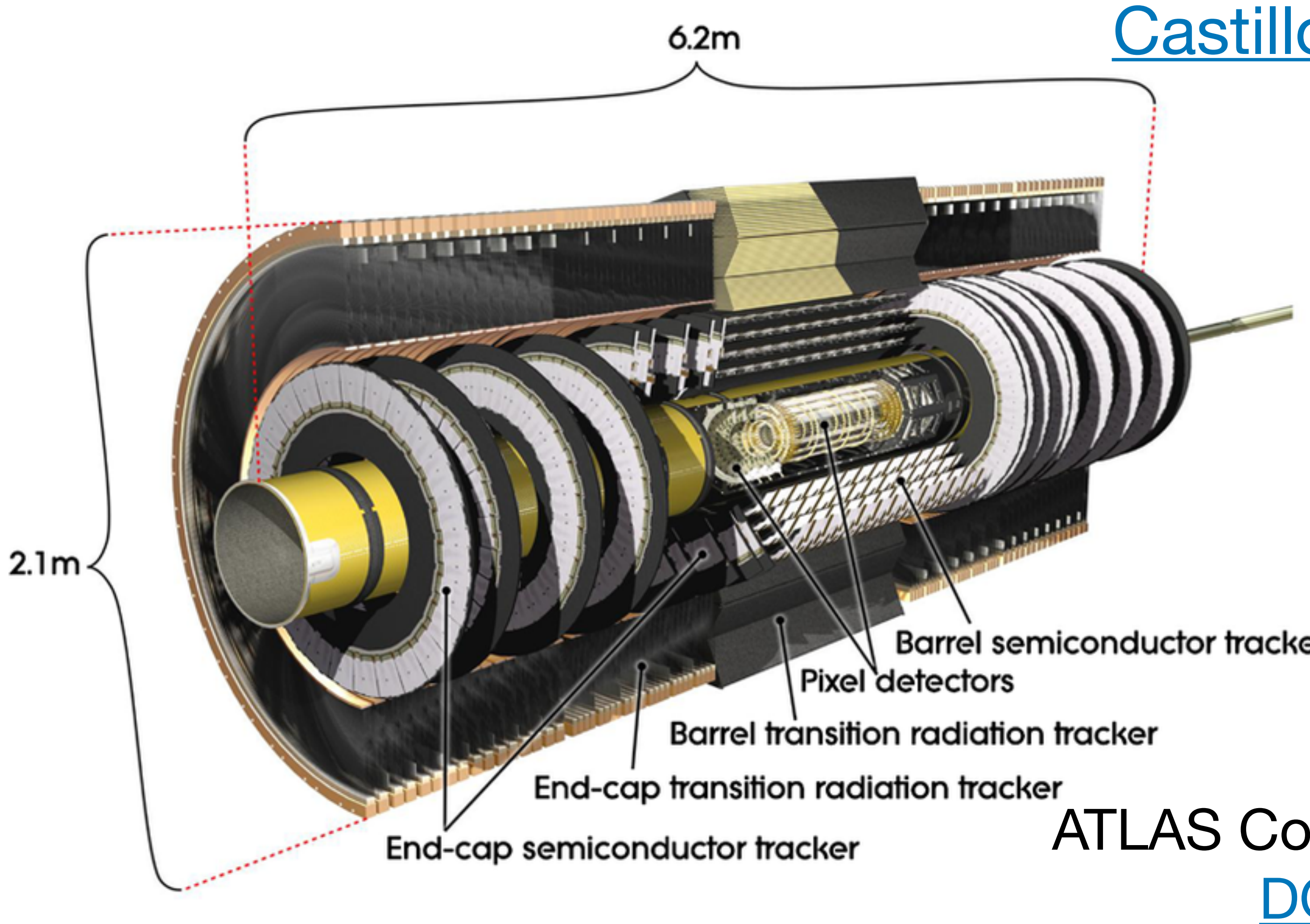
Particles in…

ATLAS is a many-layered detector
Each layer targets specific information from collision decay products

<u>Inner detector</u> - Tracking information
- Describes charged particle trajectory through the detector and magnetic field



Castillo, 2015

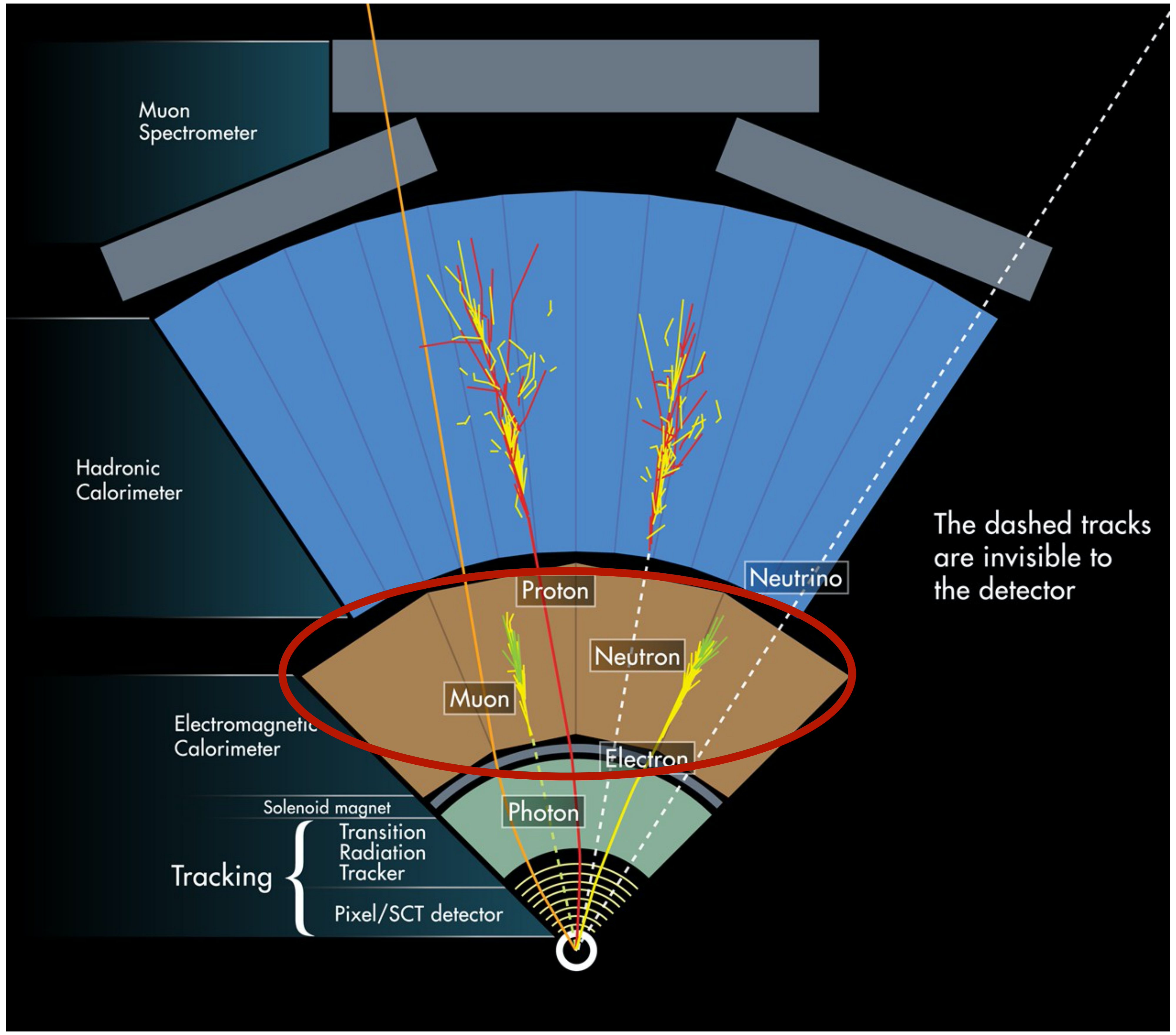


ATLAS Collaboration
DOI 10.1088

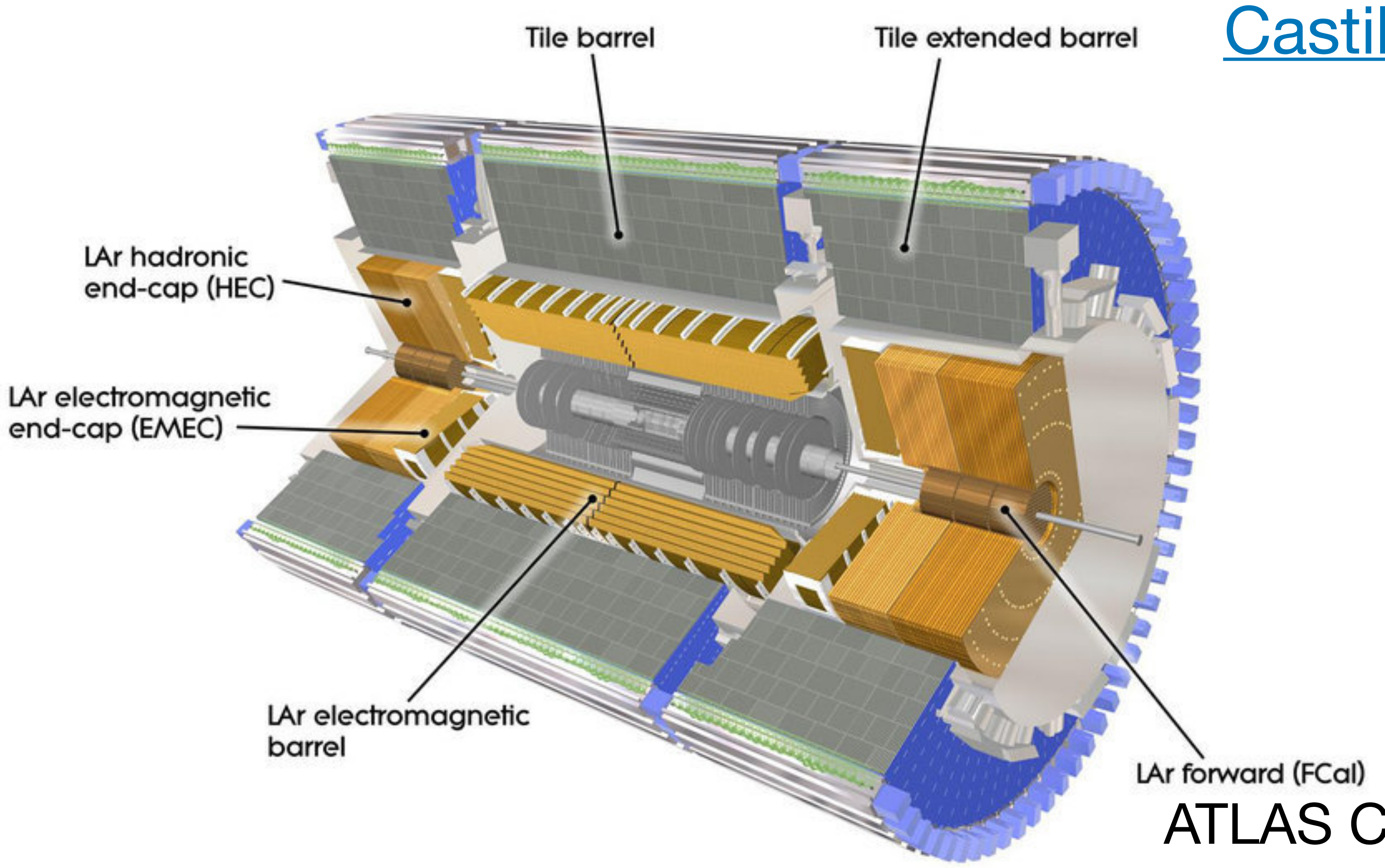
Particles in...

ATLAS is a many-layered detector
Each layer targets specific information from collision decay products

Inner detector - Tracking information
Electromagnetic Calorimeter - Electromagnetic (EM) signatures (γ,e)



Castillo, 2015

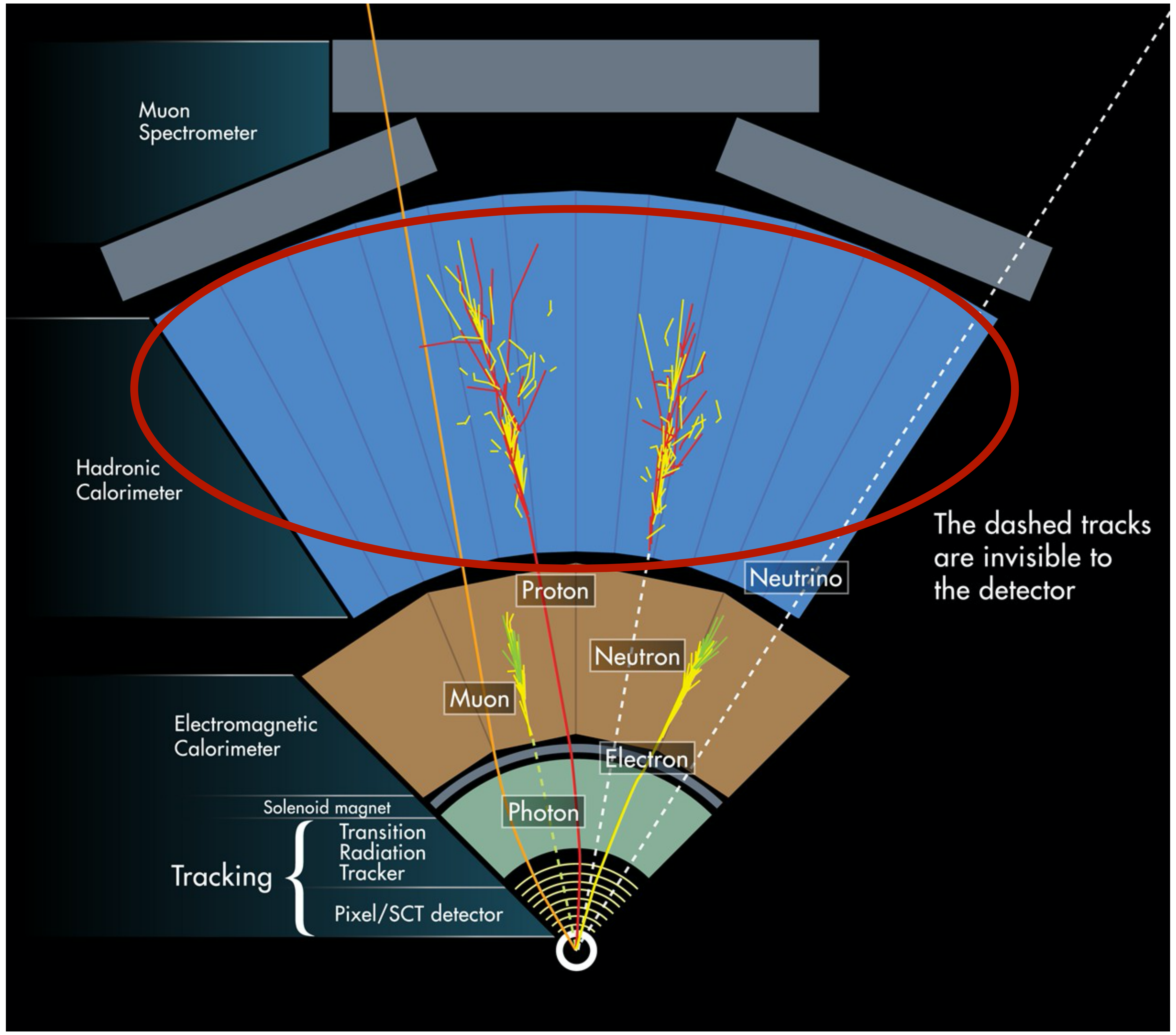


ATLAS Collaboration
DOI 10.1088

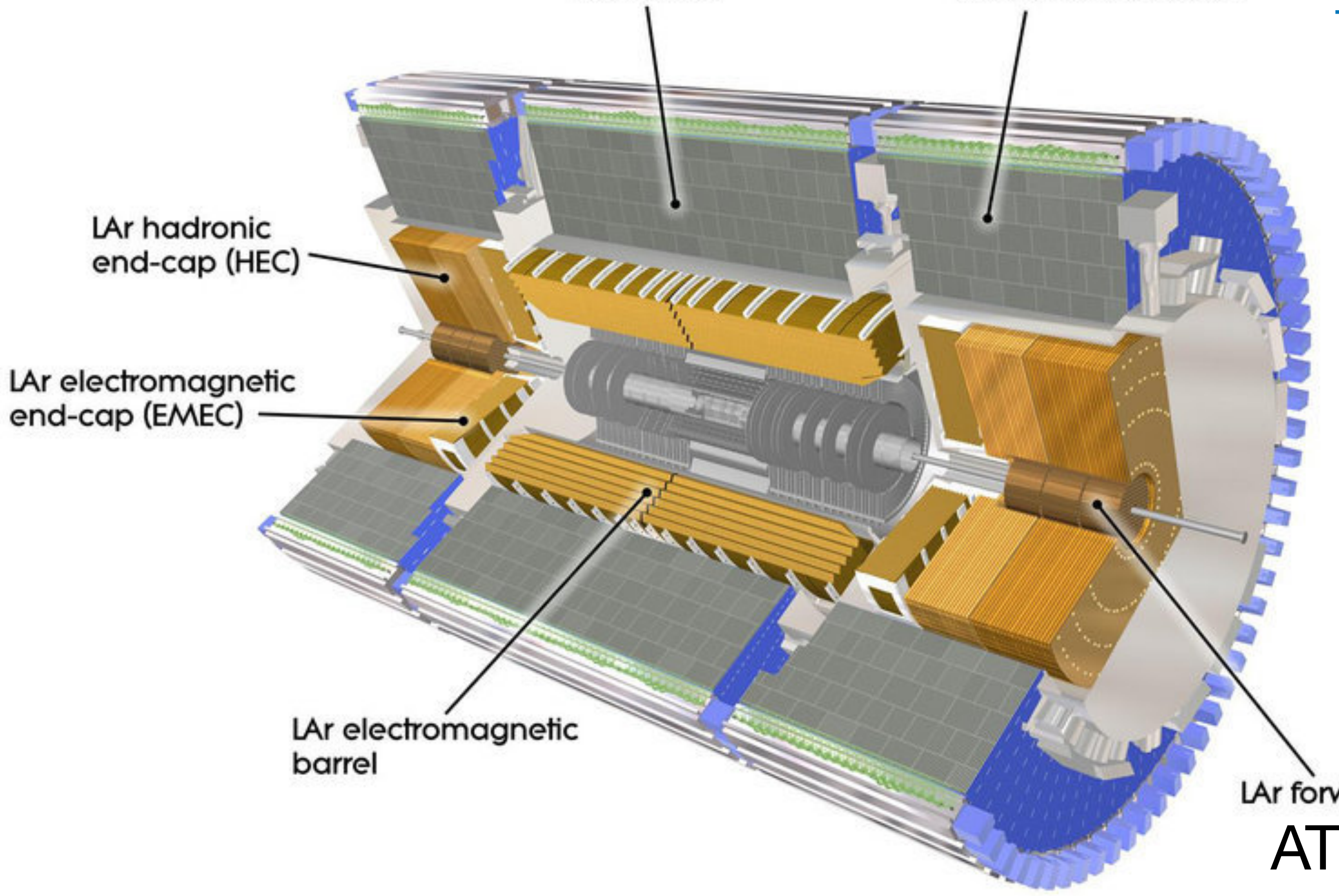
Particles in…

ATLAS is a many-layered detector
Each layer targets specific information from collision decay products

Inner detector - Tracking information
Electromagnetic Calorimeter - Electromagnetic (EM) signatures (γ,e)
Hadronic Calorimeter - Particles that interact via the strong force (quarks, gluons)
- Initiating strongly interacting particles may radiate additional quarks and gluons
- Must all be in colorless bound states due to *confinement*, so eventually turn into hadrons (π, K, p, n, etc.)



Castillo, 2015

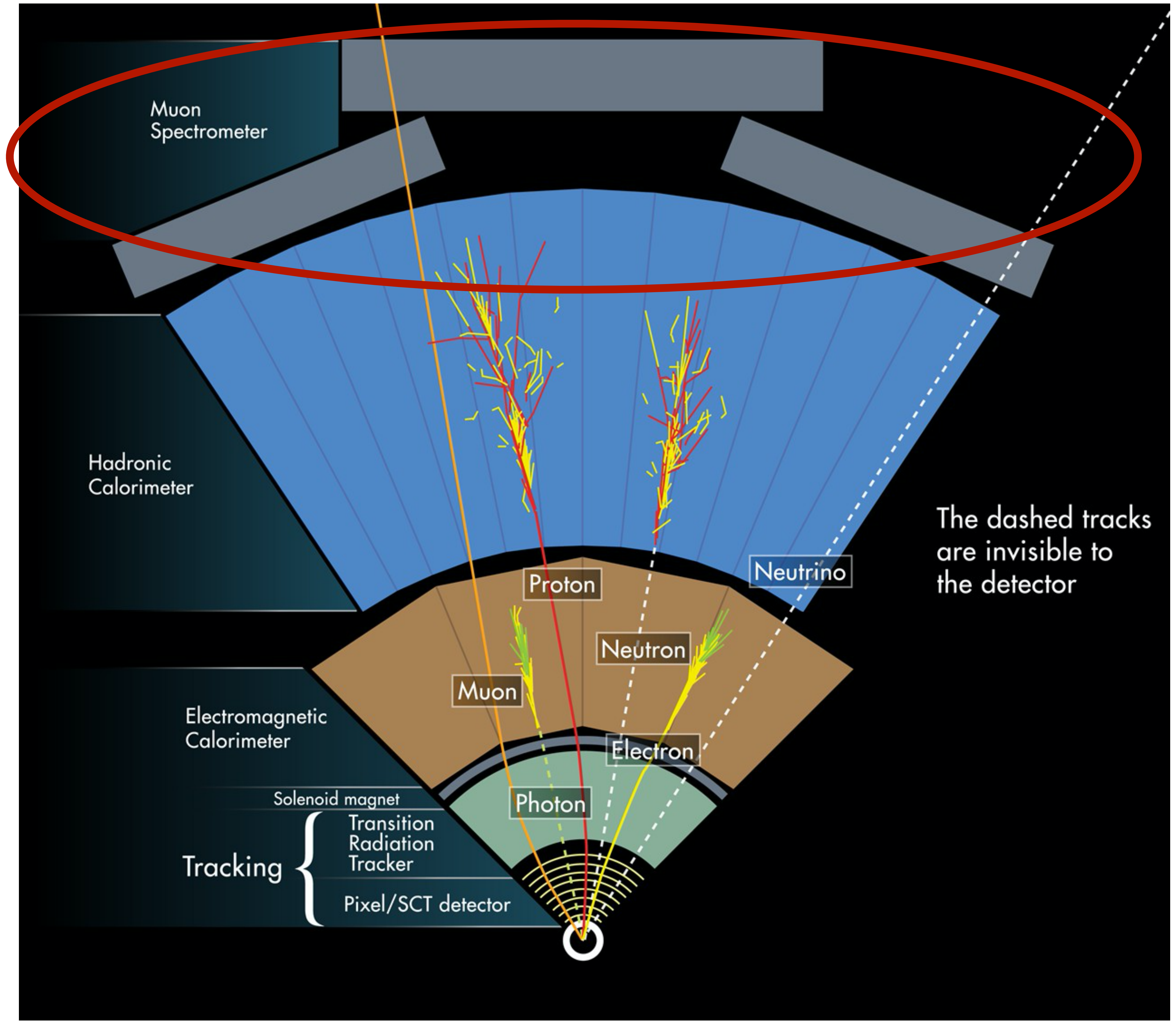


ATLAS Collaboration
DOI 10.1088

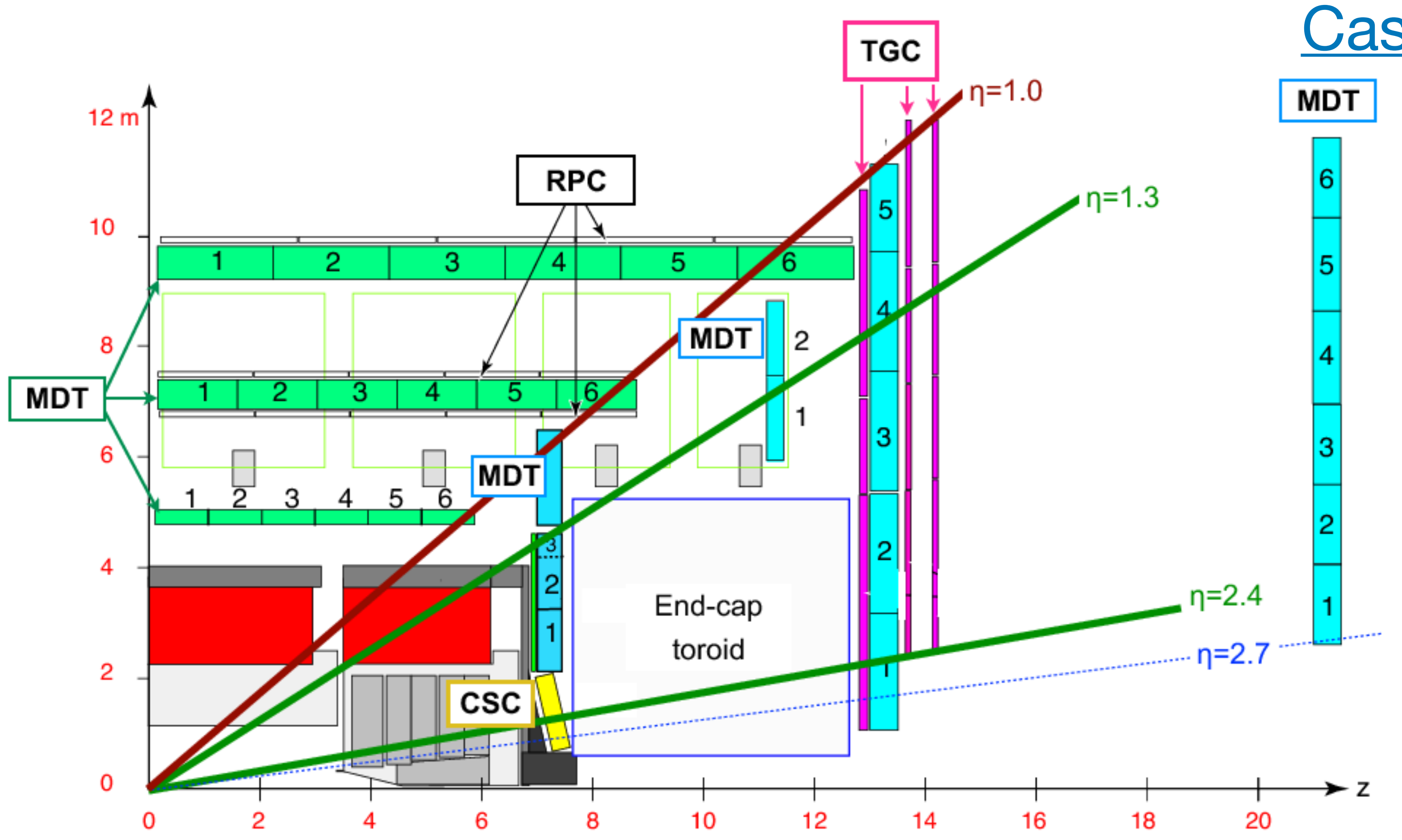
Particles in…

ATLAS is a many-layered detector
Each layer targets specific information from collision decay products

Inner detector - Tracking information
Electromagnetic Calorimeter - Electromagnetic (EM) signatures (γ,e)
Hadronic Calorimeter - Particles that interact via the strong force (quarks, gluons)
Muon Spectrometer - Muons
- Don't interact in upstream material, so dedicated subsystem for detecting



Castillo, 2015

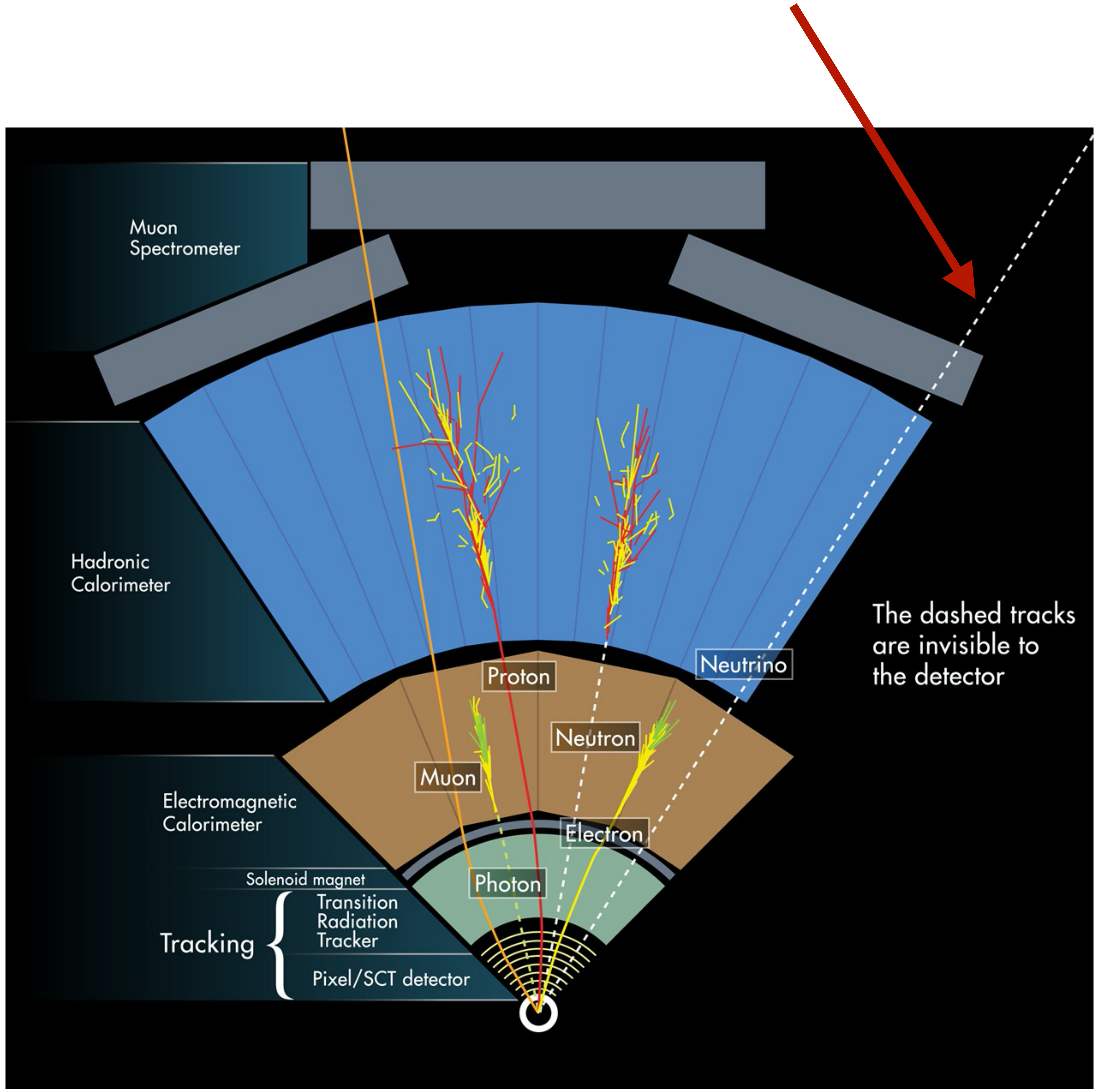


ATLAS Collaboration DOI 10.1140

Particles in...

ATLAS is a many-layered detector
Each layer targets specific information from collision decay products

Inner detector - Tracking information
Electromagnetic Calorimeter - Electromagnetic (EM) signatures (γ,e)
Hadronic Calorimeter - Particles that interact via the strong force (quarks, gluons)
Muon Spectrometer - Muons
Missing Transverse Energy (MET) - Neutrinos
- Measurement for particles that don't interact with any portion of the detector, using conservation of momentum



Castillo, 2015

...objects defined...

Use signatures in various subsystems to define various physics objects

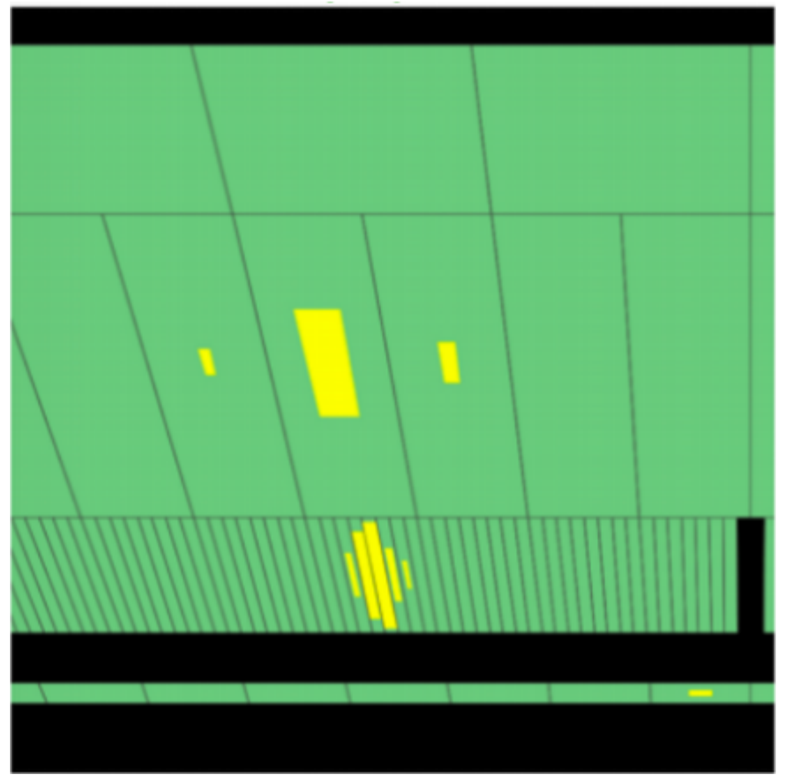
Most relevant to this analysis are photons and jets

Photons

Signatures in the electromagnetic calorimeter used

- Measures electromagnetic showers based on 2 processes
  - Bremsstrahlung photon emission
  - Electron-positron conversion

Topological clusters of dynamic size constructed moving out from high signal to noise cells, then grouped into superclusters, ultimately defined as photons



ATLAS

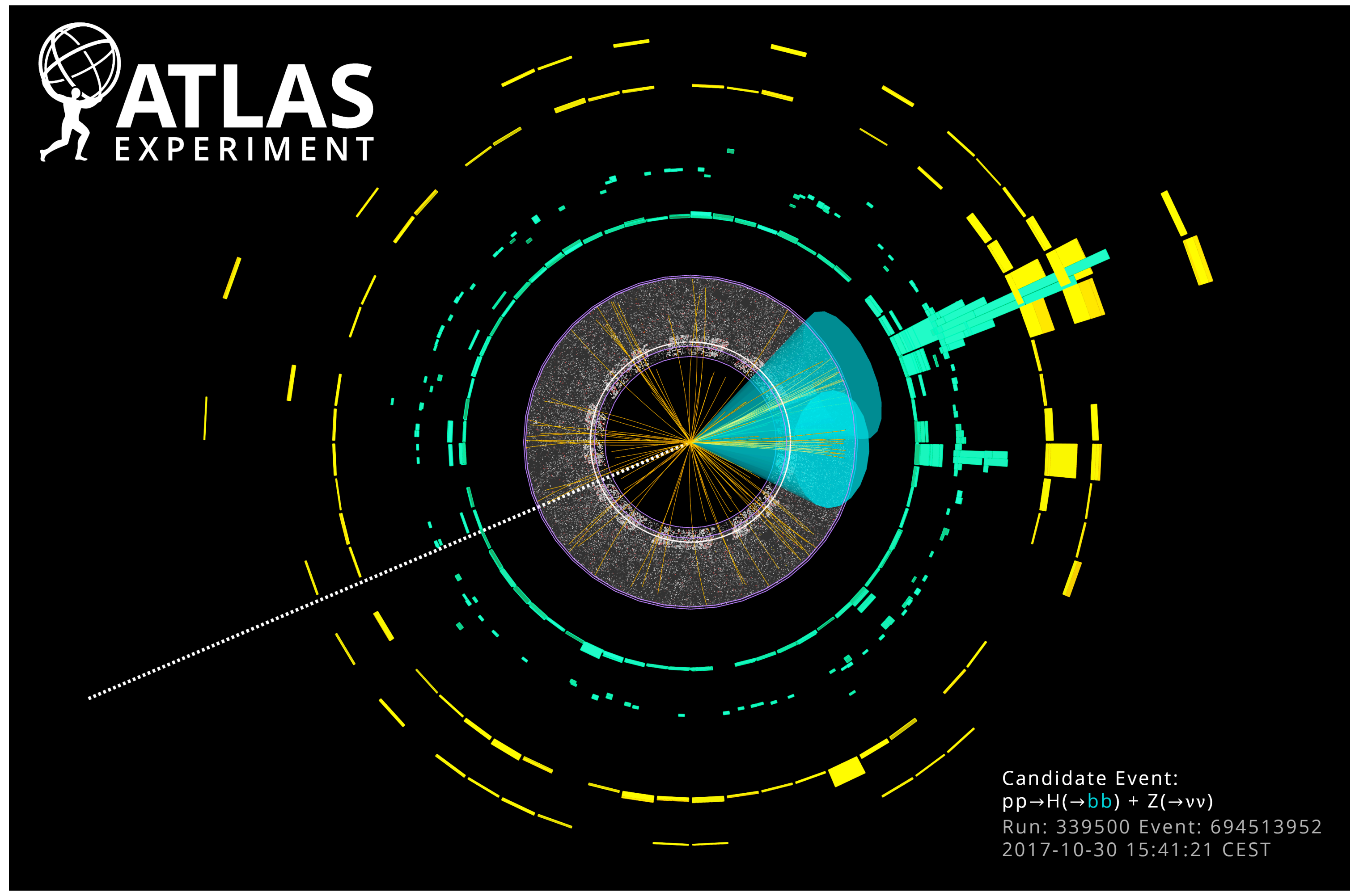
Jets

Final signature of strongly interacting particles is a collimated spray of hadrons called "jets." Information in both calorimeters used

- As hadrons interact with matter they scatter and produce additional hadrons, cascade known as hadron shower
- Cone-based algorithm (Anti-k_t) used

Algorithms used to discern the source of the jet, specifically, b-jets

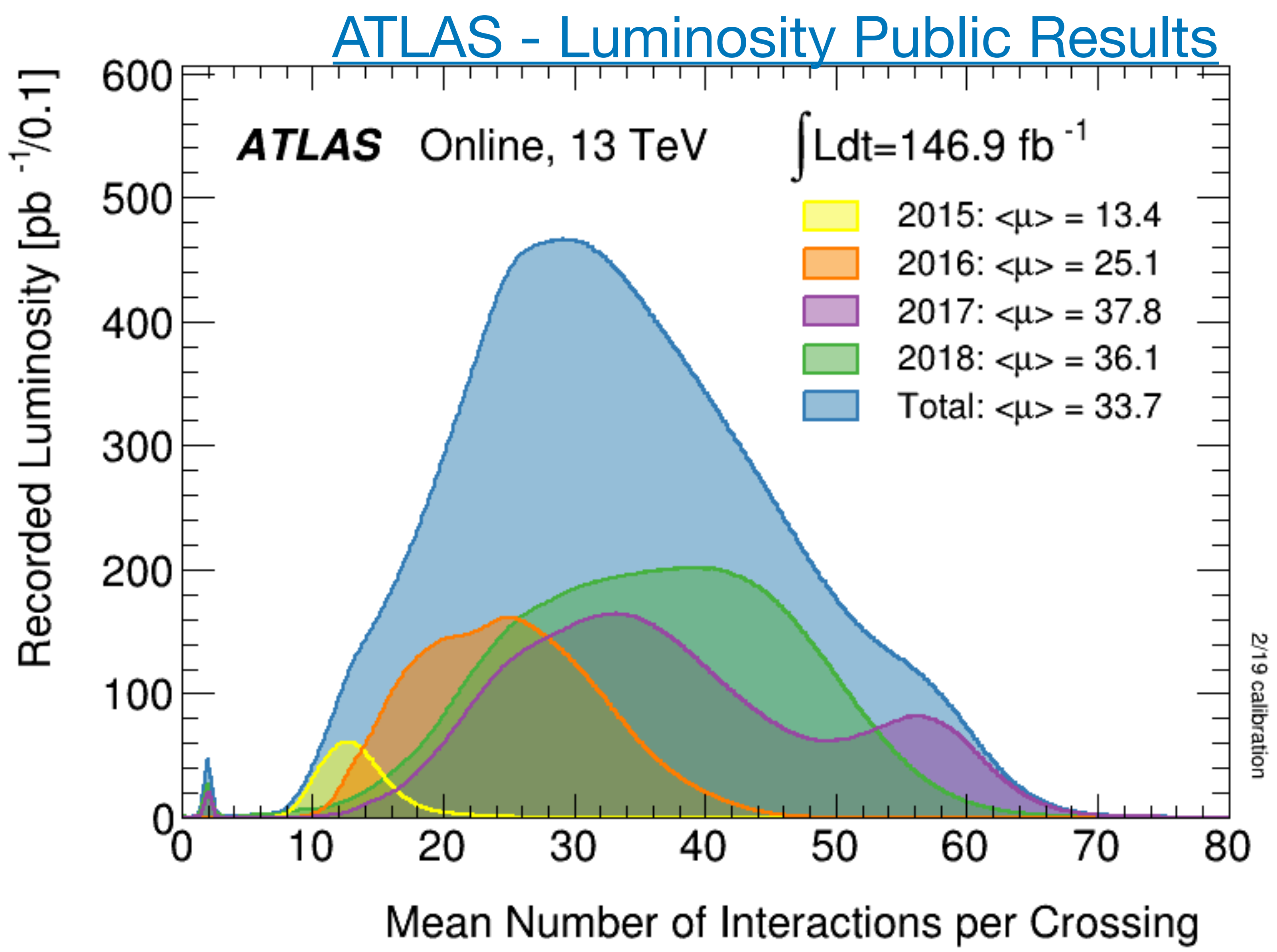
- For this work a BDT-based algorithm
- Several calibrated working points defined based on signal purity



ATLAS

...data out

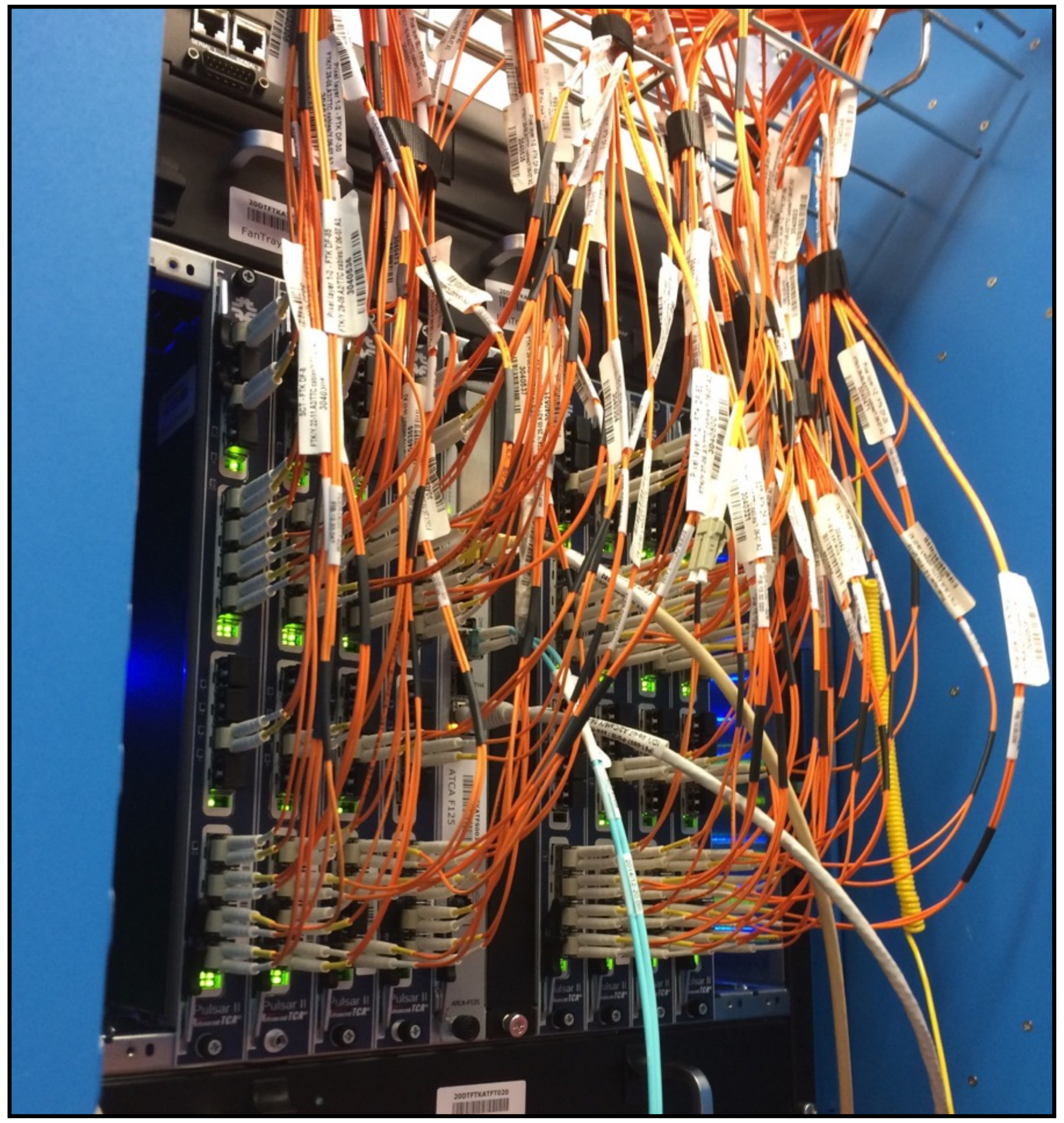
Bunches of protons collide at a rate of 40 MHz (every 25 ns)
- On average, 33.7 collisions per crossing in Run 2
  - 23.7 on average through 2015-2016 (presented analysis)



Each raw event ~1.6 MB of information, so around 64 TB/s data created

No possible way to save all of this data!

<u>ATLAS employs a 2 stage trigger system</u>
- Level 1: hardware based, performing coarse object reconstruction. Brings rate to under 100 kHz
- High-Level Trigger (HLT): Software based, performing object reconstruction as close to offline as possible. Brings rate to 1 kHz



Still, incredible amounts of data

200 petabytes milestone passed in June 2017

and with this data....

Discovery of the Higgs Boson

July 4, 2012 - Standard Model-like Higgs boson discovered by ATLAS and CMS!

- Mass reconstructed at ~125 GeV
- Observed in multiple channels with rates consistent with the Standard Model (SM) prediction
- Only fundamental scalar boson, which motivates further study beyond just the discovery

NYTimes

Plenty of outstanding questions! Among these…

- Does the discovered Higgs boson couple as predicted by the SM?
- Are there any additional Higgs bosons?
- Do particles beyond those predicted by the SM exist? If so, how do they couple to the Higgs boson?

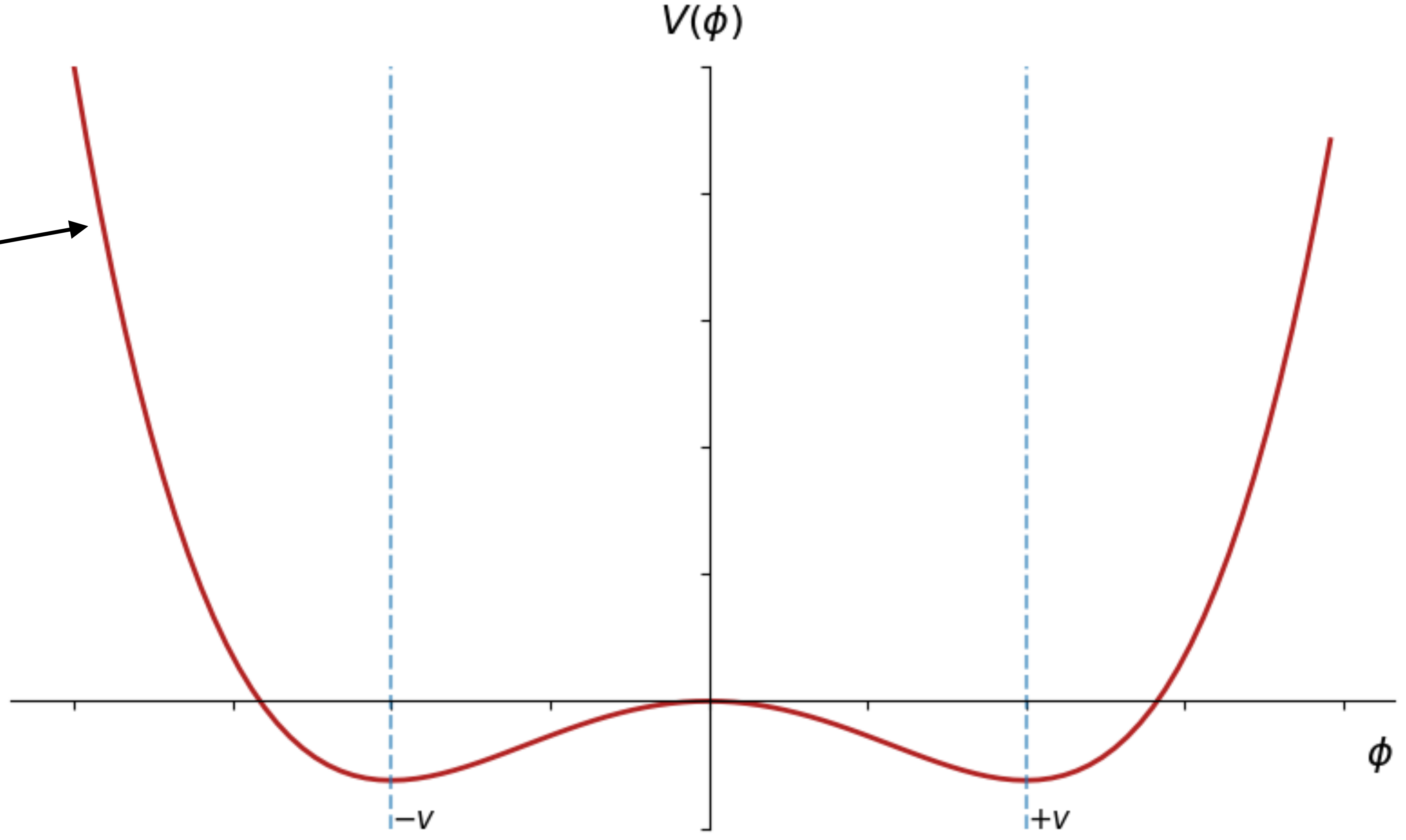
Studying di-Higgs production gives insight to these questions

Northern Illinois University

Di-Higgs Production

Higgs potential function is given by

$$V(\phi) = \mu^2 |\phi|^2 + \lambda(|\phi|^2)^2$$



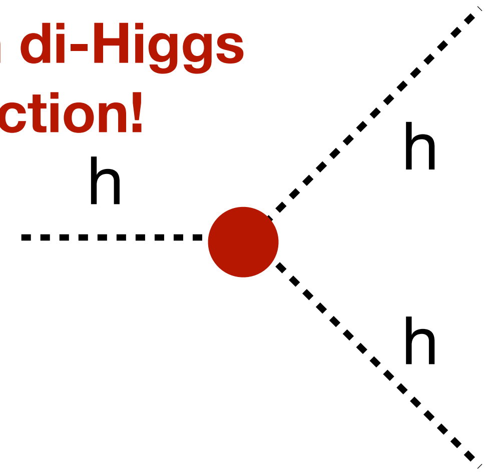
for scalar field ϕ with mass μ, and Higgs self coupling strength λ. For $\mu^2 < 0$, potential shown, with minima of found at

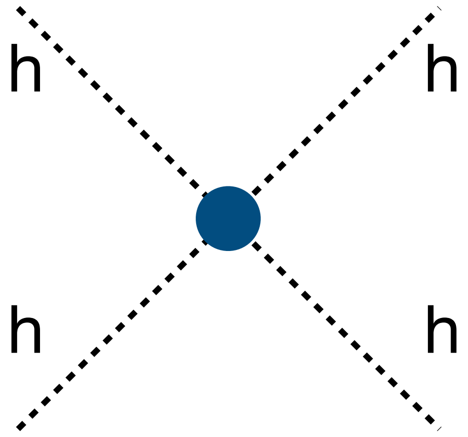
$$v = \sqrt{-\mu^2/\lambda} = 246$$ GeV (from experiment),

which meets earlier criteria of **non-zero values in empty space**. Pick a minima and expand:

$$V(h) = \lambda v^2 h^2 + \underbrace{\lambda v}_{\lambda_{HHH}^{SM}} h^3 + \frac{\lambda}{\underbrace{4}_{\lambda_{HHHH}^{SM}}} h^4$$

Results in di-Higgs production!





Quartic coupling - extraordinarily rare, out of reach of LHC

Picking this minima is known as **electroweak symmetry breaking,** and implies the Higgs **self-couples**. Measurement of the Higgs self-coupling provides a direct probe on the Higgs potential.

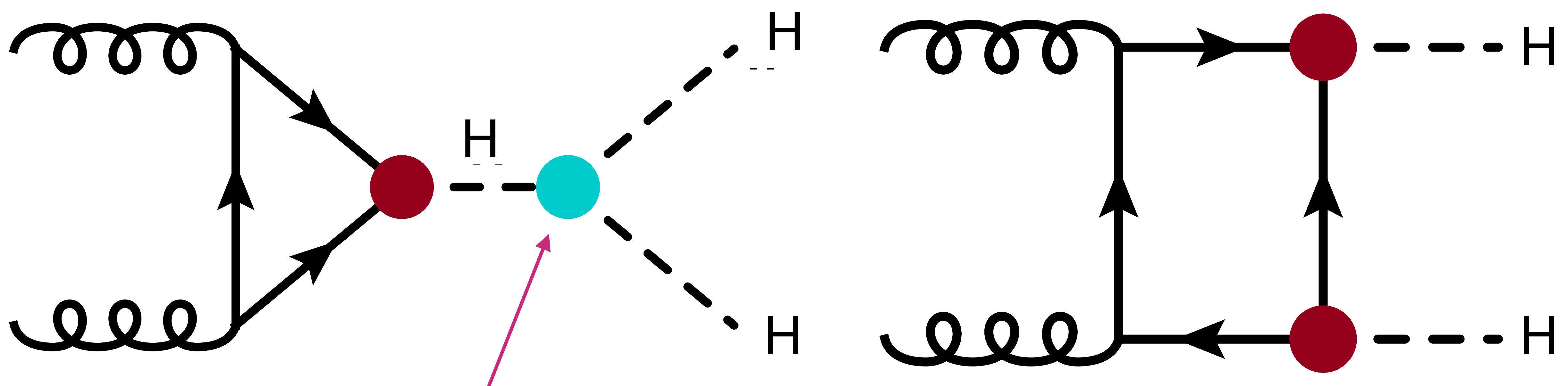
Northern Illinois University

Di-Higgs Production in the SM

Gluon-gluon fusion (ggF) is the dominant di-Higgs production mode, accounting for ~87% of di-Higgs events at 13 TeV, most published analyses target this production mode

SM non-resonant HH Production via gluon-gluon fusion

JHEP 11 (2018) 040



Higgs Self-Coupling, κ_λ

Not sensitive to κ_λ

Occurs through two diagrams which interfere destructively

Very small cross section!

$\sigma(gg \to HH)_{SM} \approx 33$ fb @ 13 TeV

di-Higgs production is a rare process, which will test the limits of the LHC

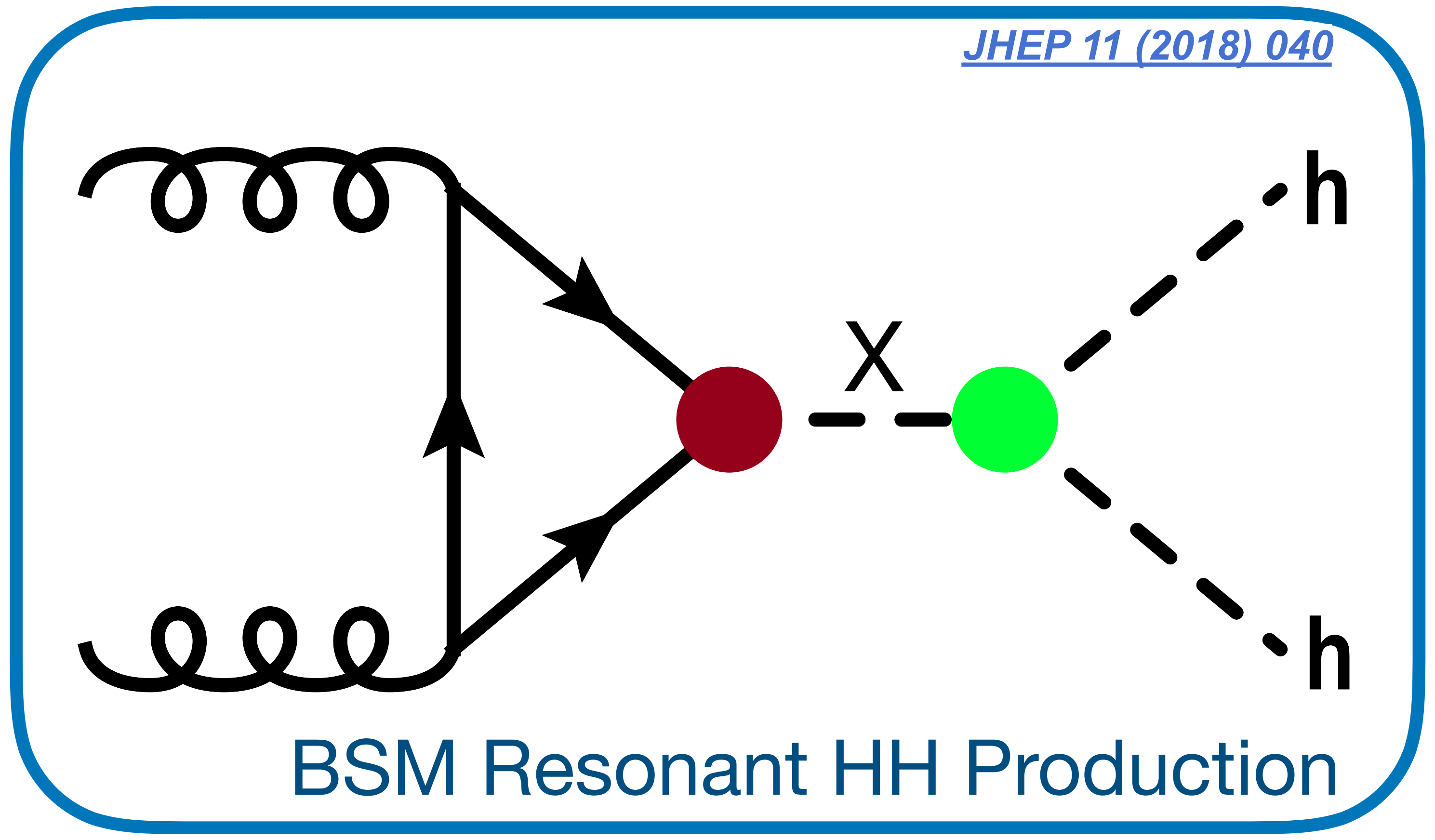
Northern Illinois University

BSM Di-Higgs Production

Resonant enhancements:

Enhancements can occur through a theoretical particle beyond the Standard Model (BSM) that decays to two Higgs bosons

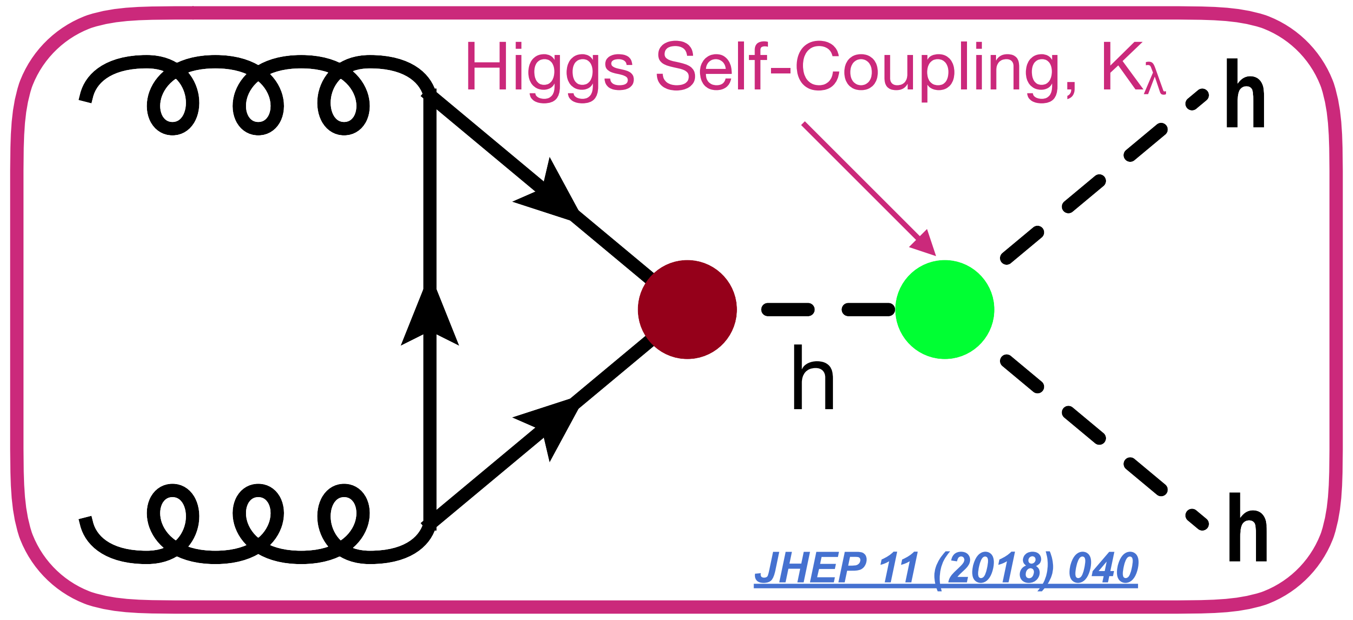
- Two-Higgs-doublet model (THDM) - Resonance is a heavy scalar

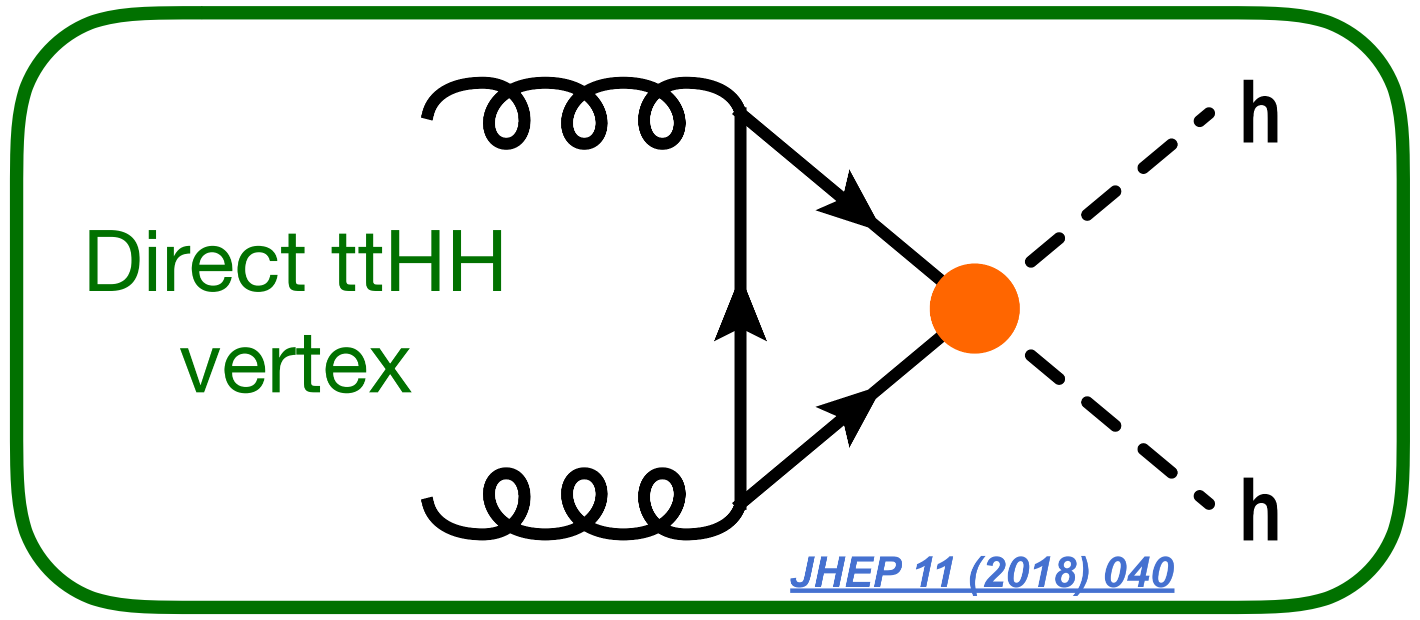


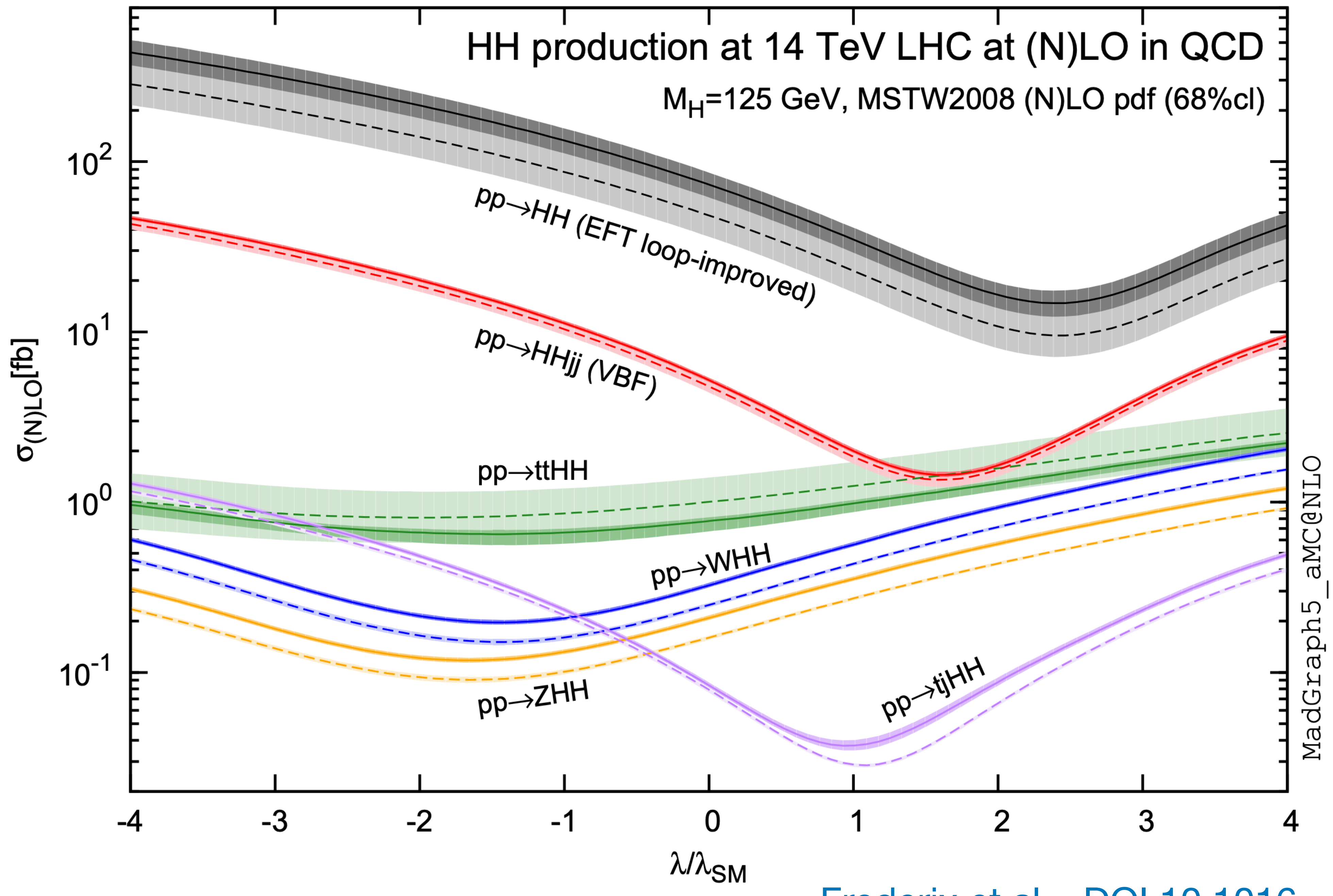
Non-Resonant enhancements:

Can also look toward enhanced production caused by couplings that differ from the SM

- Higgs Self-Coupling (K_λ) - strength of coupling deviating from SM value can lead to enhanced production
- Additional couplings not predicted by the SM, e.g. direct ttHH vertex







Frederix et al., DOI 10.1016

Models with enhanced production make di-Higgs already interesting to study with Run-2 data

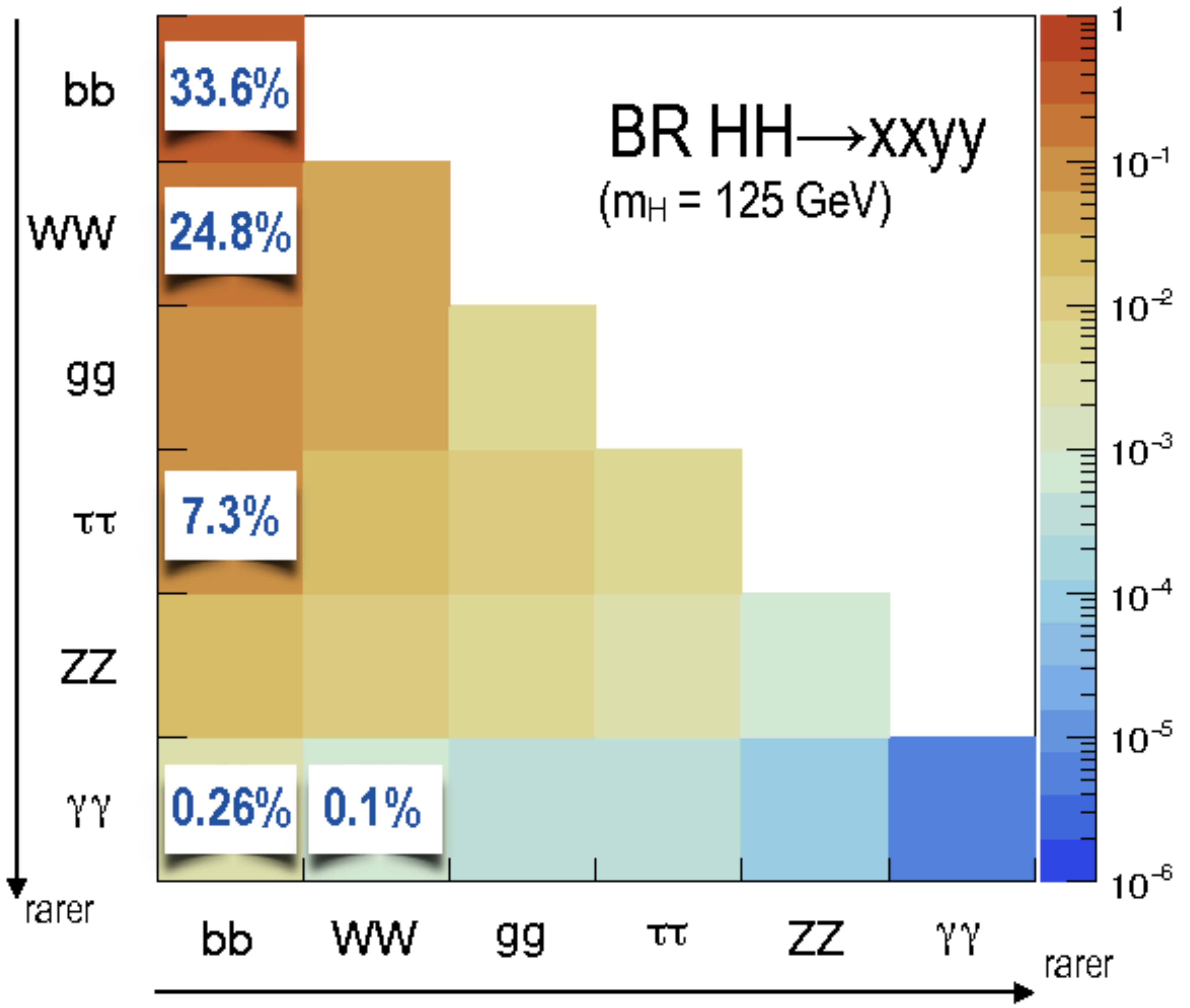
Decay Modes

Strong motivation to look at certain channels based on properties of decay products

Strongest contribution to current limits:

- $b\bar{b}b\bar{b}$: Fully takes advantage of high bb branching ratio, but suffers from large multijet background
- $\gamma\gamma b\bar{b}$: Excellent trigger and mass resolution for photons, high bb branching ratio
- $b\bar{b}\tau\tau$: Taus are relatively clean while still having a large branching ratio, high bb branching ratio

Morse



Northern Illinois University

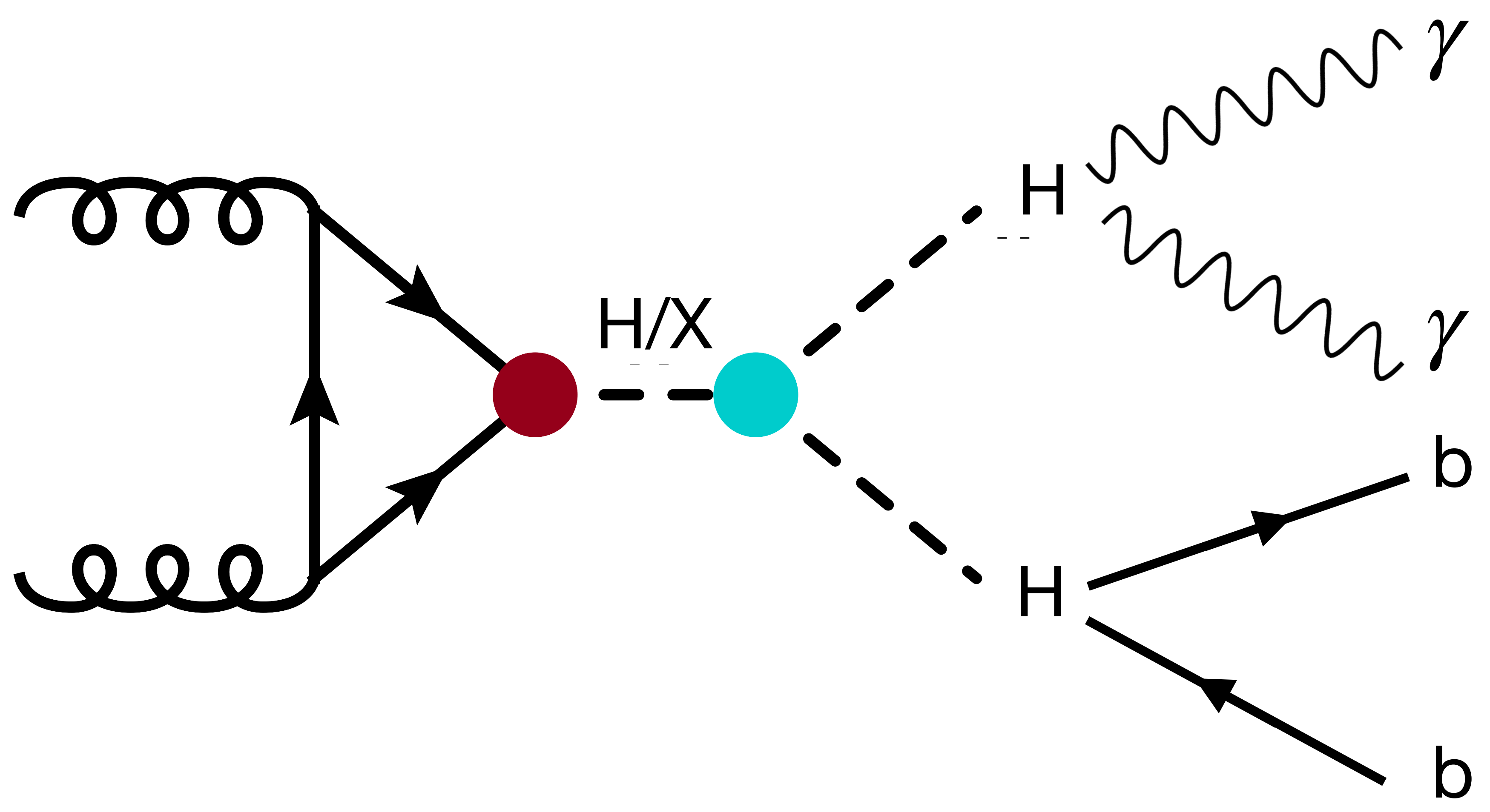
HH→$\gamma\gamma$bb search using 2015+2016 data

Analysis Overview

Searching for resonant and non-resonant di-Higgs production

Final state involves one Higgs decaying to photons, the other to a b quark pair

- Search for this final state in the detector - 2 photons and 2 b-jets



Full decay shown (similar for box diagram)

Analysis provides sensitivity to 3 values:

- The SM di-Higgs cross-section
- The Higgs trilinear coupling strength
- The existence of a generic scalar particle (resonance) which decays to Higgs boson pairs
  - Generic scalar chosen to remain model independent

Northern Illinois University

Analysis Selection

Preselection - target decay products

Matches other $H \to \gamma\gamma$ analyses - 2 photons passing tight criteria, within the mass window of [105, 160] GeV

- Diphoton trigger used - one photon with $p_{T,\gamma 1} > 35$ GeV, $p_{T,\gamma 2} > 25$ GeV

2 jets within b-tagging region for $H \to bb$

Split into 3 categories using b-tagging information

2-tag	**1-tag**	**0-tag**
2 jets passing medium b-tagging (70% signal efficiency)	1 jet passing tight b-tagging (60% signal efficiency)	Passes preselection, but fails other categories
Used as a signal region	Used as a signal region, second jet chosen using a NN	Used for shape information

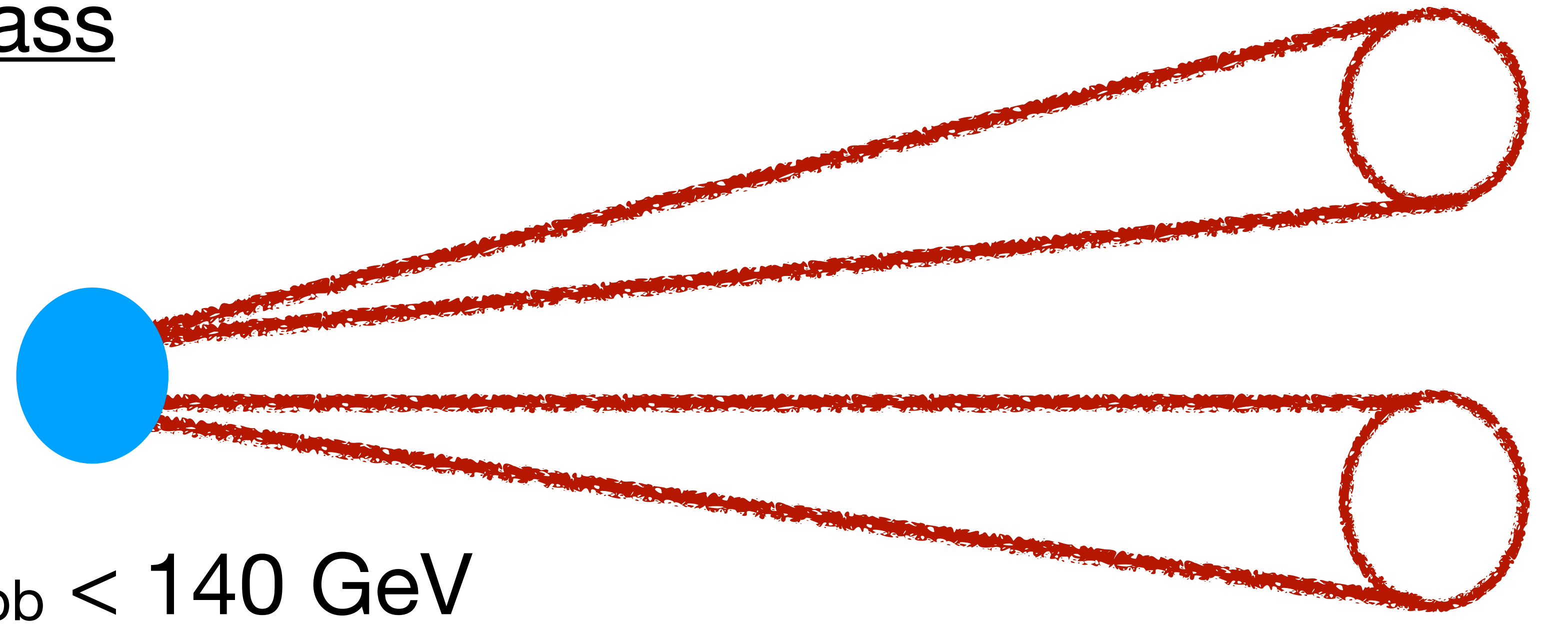
Northern Illinois University

Analysis Selection

2 Overlapping Analysis Categories

Built to target different portions of the analysis, applying constraints to jets and bb system

<u>Low mass</u>



$p_{T,j1} > 40$ GeV

$p_{T,j2} > 25$ GeV

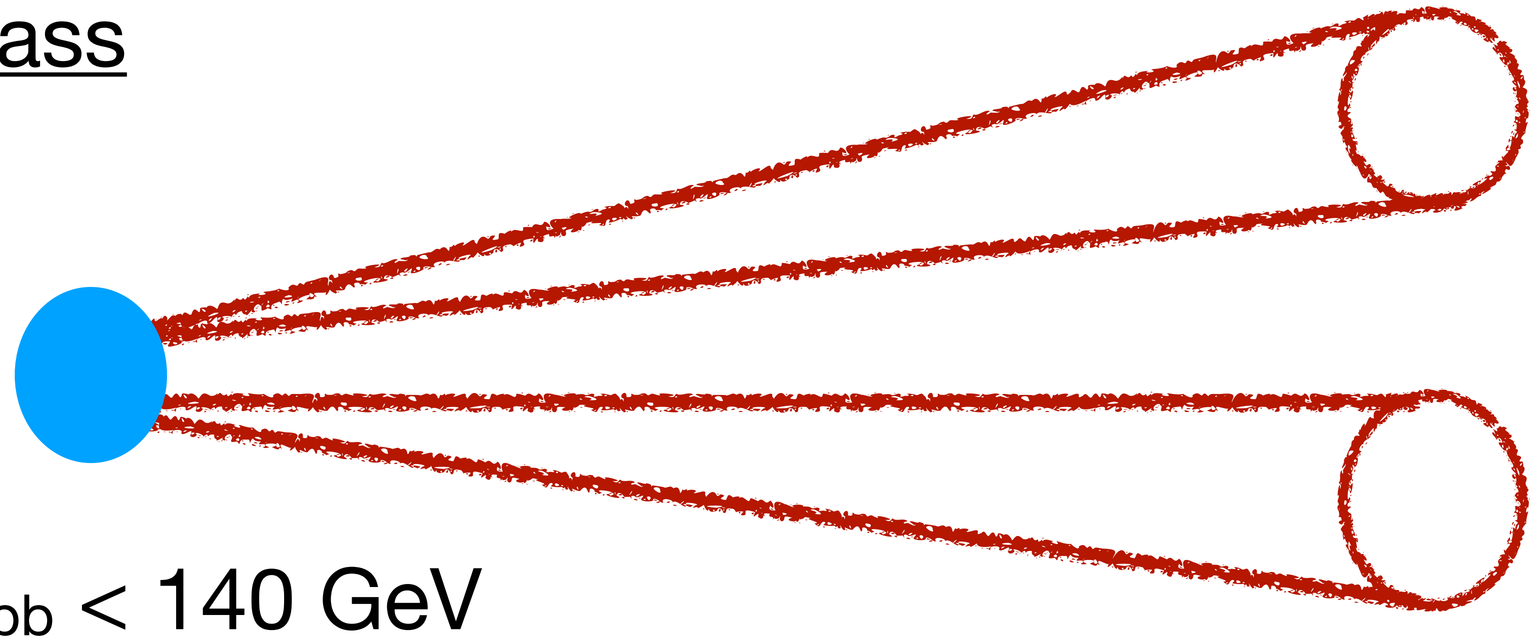
$80 < m_{bb} < 140$ GeV

Used for:

Resonant analysis, mass hypotheses below 500 GeV

K_λ scan

<u>High Mass</u>



$p_{T,j1} > 100$ GeV

$p_{T,j2} > 30$ GeV

$90 < m_{bb} < 140$ GeV

Used for:

Resonant analysis, hypotheses above 500 GeV

Non-resonant Analysis

Approach - Perform unbinned likelihood fits to signal and background

Non-resonant analysis - fit $m_{\gamma\gamma}$ spectrum

Resonant analysis - cut in $m_{\gamma\gamma}$ and fit $m_{\gamma\gamma bb}$ spectrum

Northern Illinois University

Background Estimation

Largest background in analysis is $\gamma\gamma$-continuum processes, subdominant is single Higgs boson production

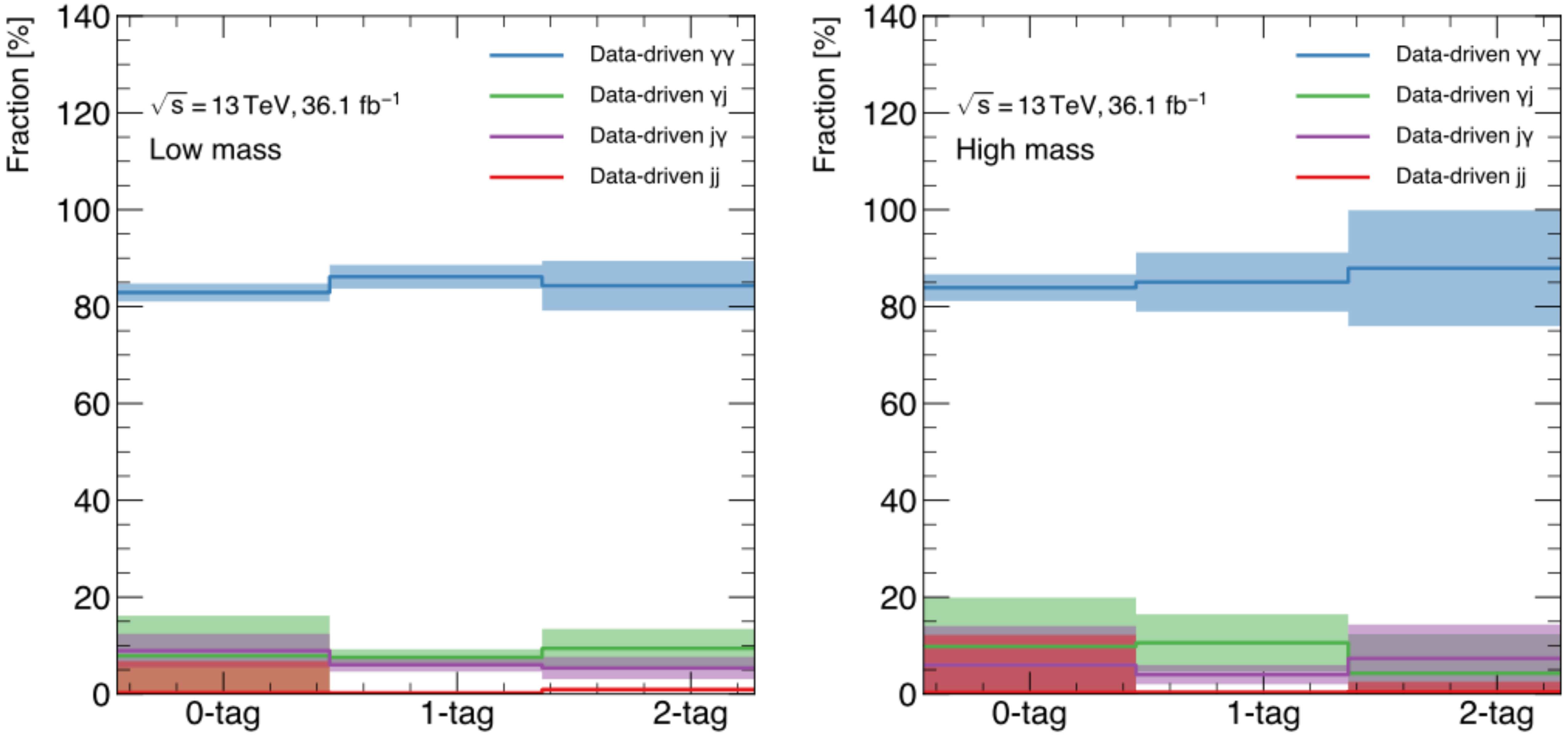
Use data-driven method at estimating $\gamma\gamma$-continuum contribution from real photons and jet fakes ($\gamma\gamma$, γj, jγ, jj)

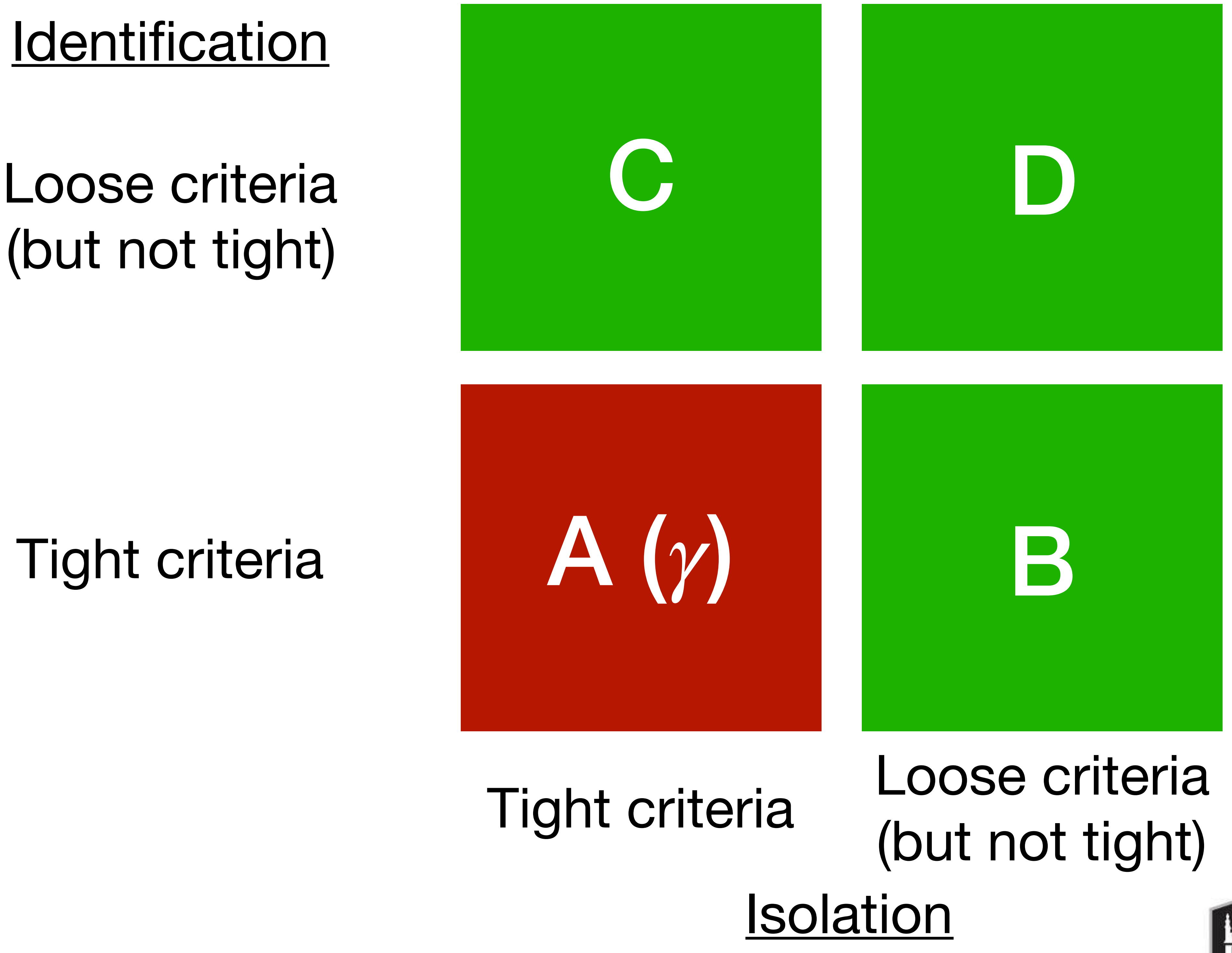
- Different kinematics not well modeled in MC

Use 2x2D sideband method

Method constructs 4 regions for each photon, using orthogonal cuts on photon identification and isolation criteria.

- 4 regions ** 2 photons = 16 total regions
- Use to establish a system of 16 equations on yield with 19 unknowns, constrain 4 from MC, 2 are assumed to be 1.0
- Perform bin-wise in $m_{\gamma\gamma}$ 0-tag (statistics), to establish a smooth reweighting function, apply across categories



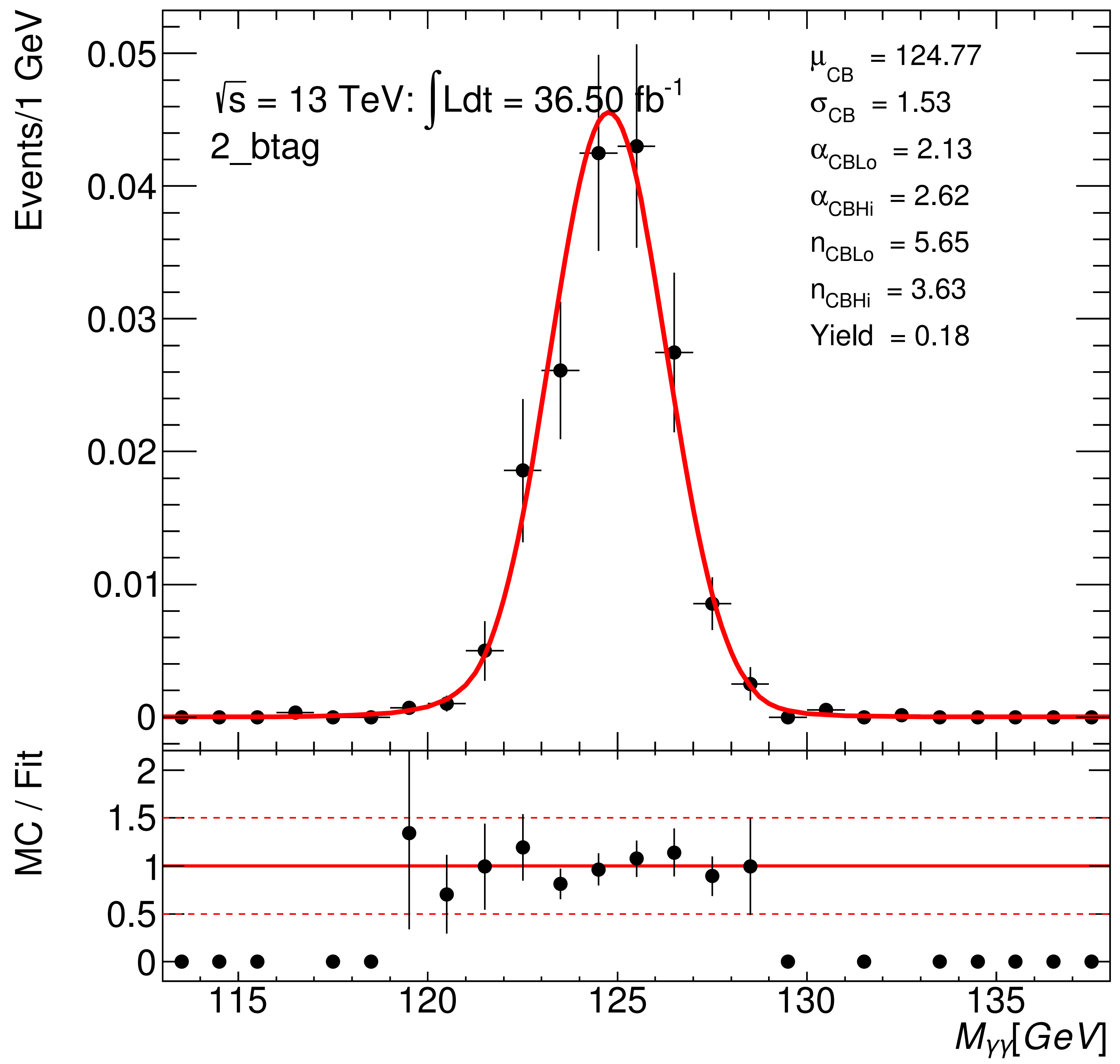


Northern Illinois University

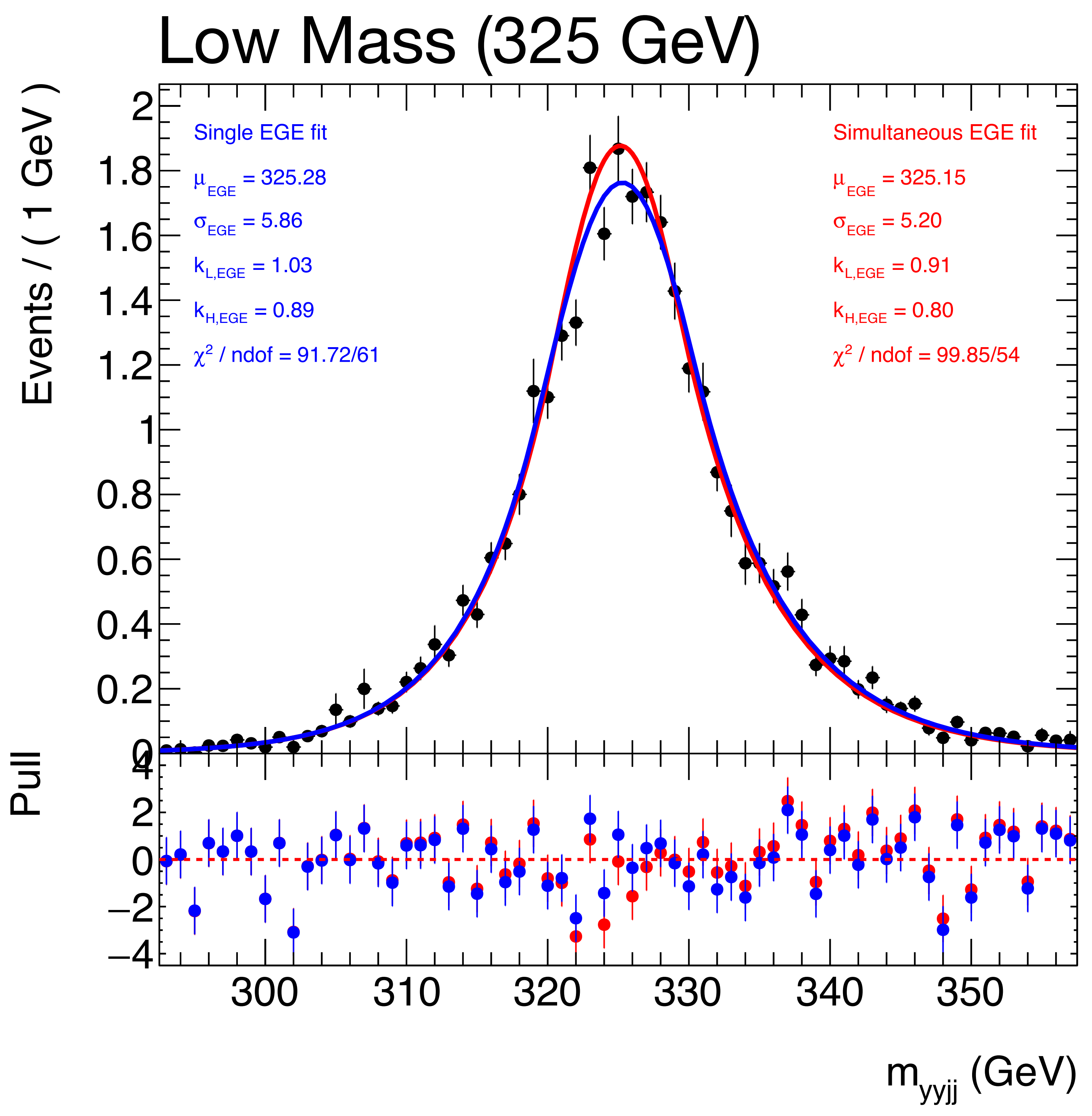
Analysis Approach - Signal Modeling

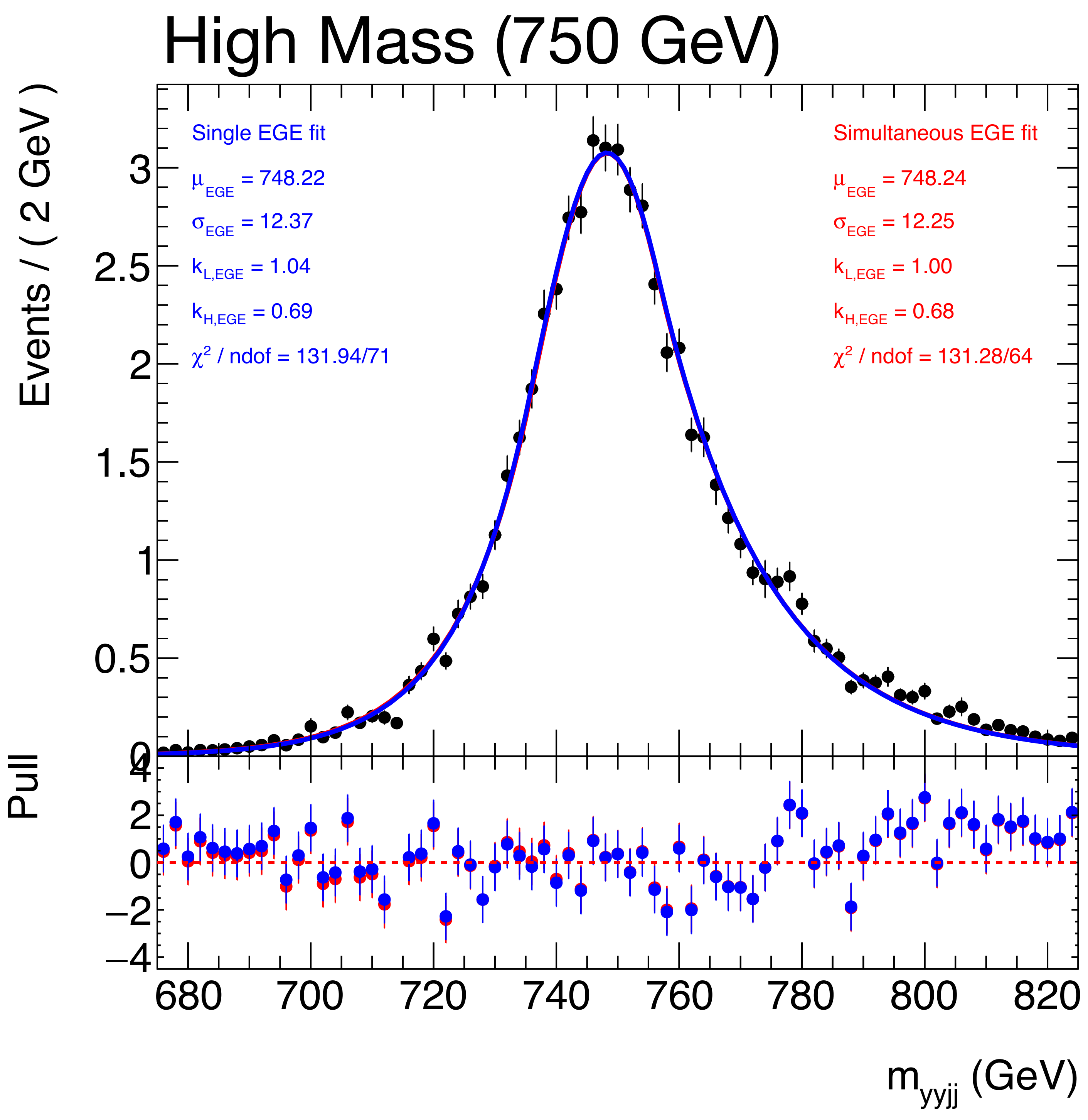
Signal modeling - parametrize Higgs peak (non-resonant) or peak for generic scalar particle decaying to Higgs pairs (resonant)

Non-resonant Analysis:	**Resonant Analysis:**
Fit $m_{\gamma\gamma}$ spectrum	Cut $120 < m_{\gamma\gamma}/\text{GeV} < 130$ Fit $m_{\gamma\gamma bb}$ spectrum, low and high mass categories separately
Double-sided crystal ball - Gaussian with power law tails	Gaussian with exponential tails - Selected converging fit with minimal number of parameters



$m_H = 125$ GeV





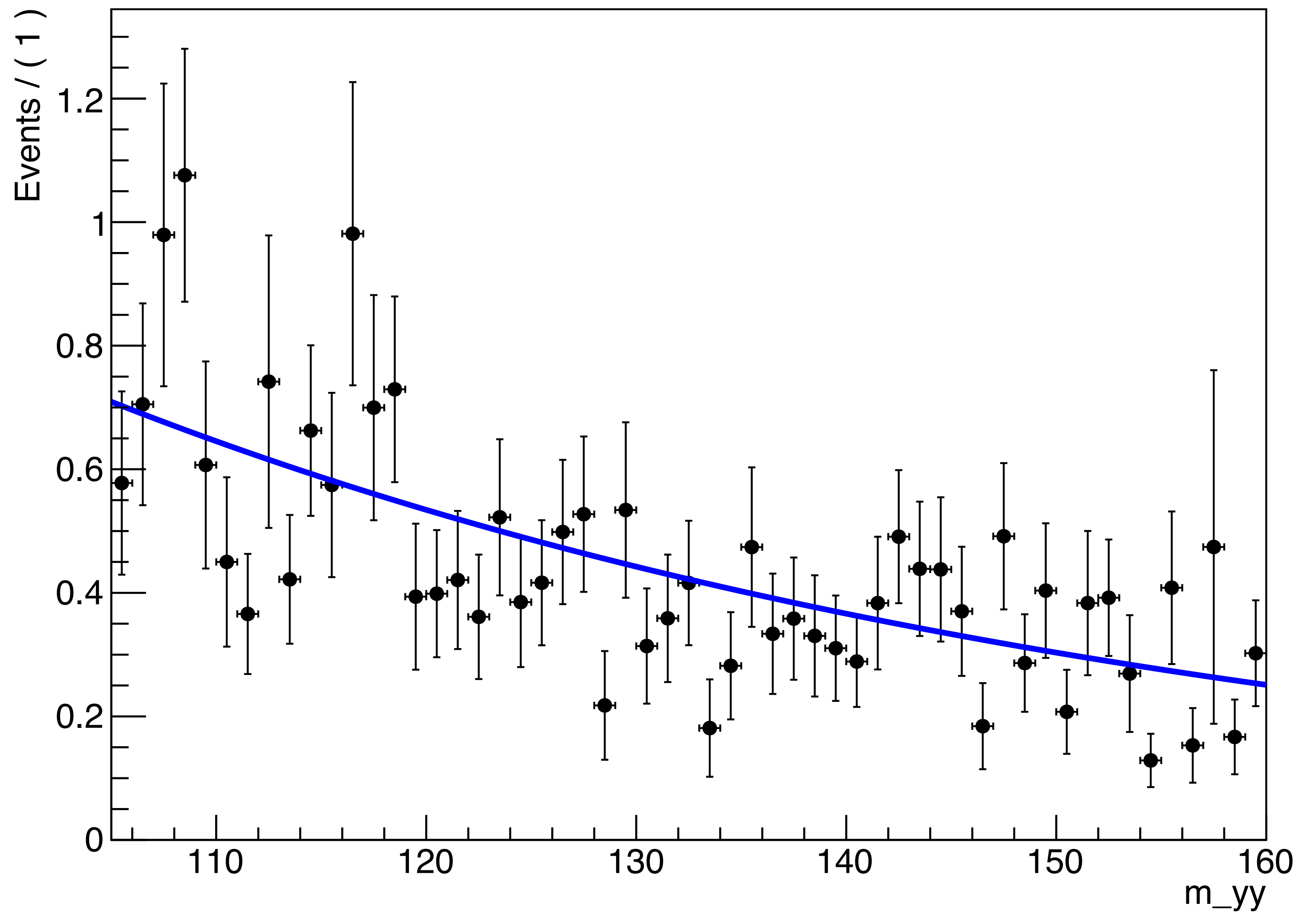
Northern Illinois University

Analysis Approach - Background Modeling

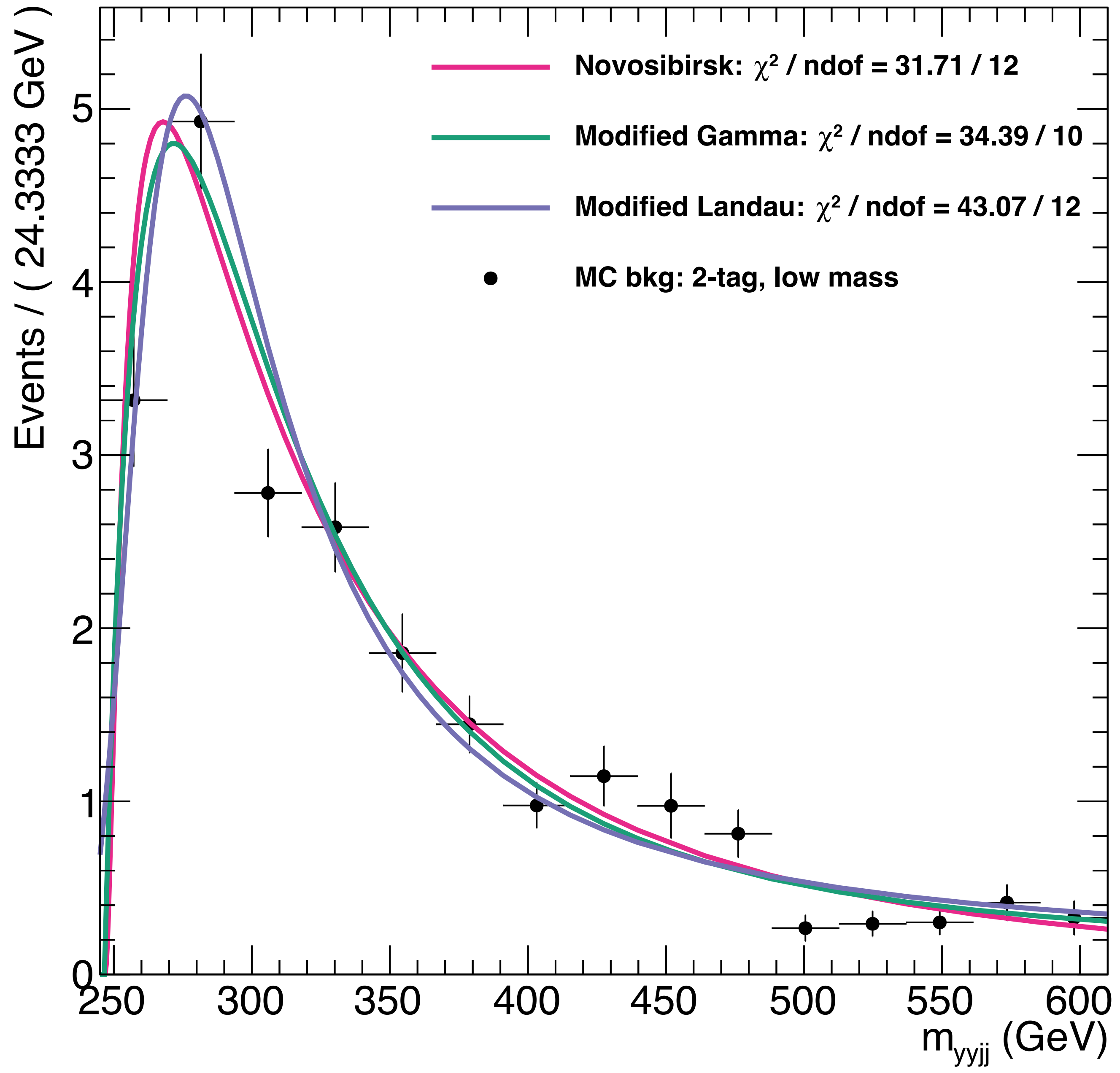
Perform unbinned profile likelihood fit

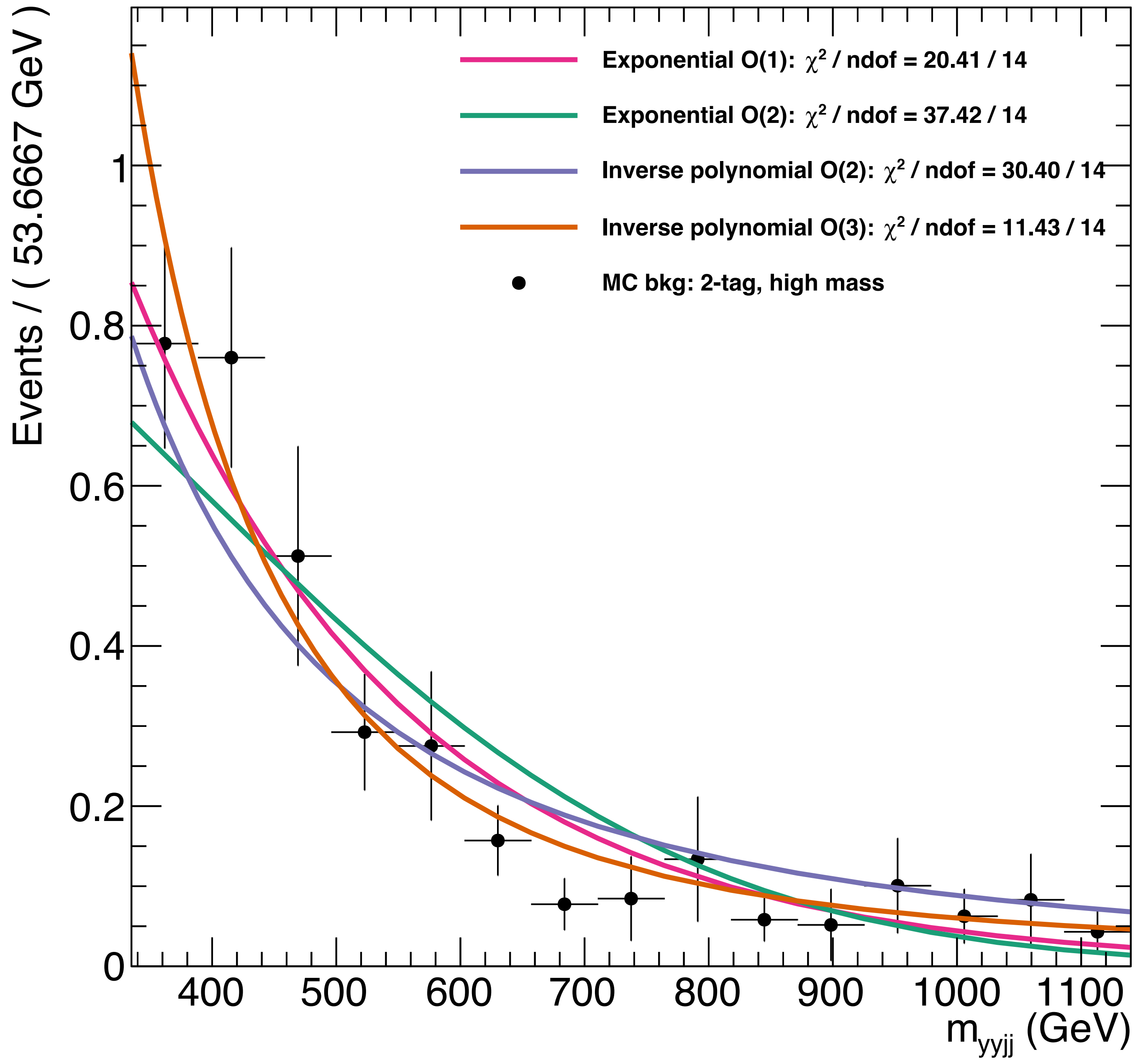
Models selected through “spurious signal” test - perform signal+background fits to just the background, take as bias. Select function that minimizes this bias

Non-resonant Analysis:	**Resonant Analysis:**	
Fit $m_{\gamma\gamma}$ spectrum	Cut $120 < m_{\gamma\gamma}$/GeV< 130 Fit $m_{\gamma\gamma bb}$ spectrum, low and high mass categories separately	
First order exponential	Low mass - Novosibirsk (Used for $m_X < 500$)	High mass - Exponential (Used for $m_X \geq 500$)



m_H = 125 GeV





* Additionally single Higgs background contributions modeled with double sided crystal ball functions

Perform signal plus background fits using functions - No excesses seen, so set limits

Uncertainty

Statistically dominated search!

- **Non-resonant Signal and single Higgs:** Shape and yields extracted from fit to MC, uncertainties propagated to yield, peak, and width
- **Resonant Signal:** Shape and yield evaluated at each mass hypothesis, maximum taken for all
- **Background Parameters:** Unconstrained nuisance parameters

Source of systematic uncertainty		% effect relative to nominal in the 2-tag (1-tag) category			
		Non-resonant analysis		Resonant analysis: BSM HH	
		SM HH signal	Single-H bkg	Loose selection	Tight selection
Luminosity		±2.1 (±2.1)	±2.1 (± 2.1)	±2.1 (±2.1)	±2.1 (±2.1)
Trigger		±0.4 (±0.4)	±0.4 (± 0.4)	±0.4 (±0.4)	±0.4 (±0.4)
Pile-up modelling		±3.2 (±1.3)	±2.0 (± 0.8)	±4.0 (±4.2)	±4.0 (±3.8)
Photon	identification	±2.5 (±2.4)	±1.7 (± 1.8)	±2.6 (±2.6)	±2.5 (±2.5)
	isolation	±0.8 (±0.8)	±0.8 (± 0.8)	±0.8 (±0.8)	±0.9 (±0.9)
	energy resolution	-	-	±1.0 (±1.3)	±1.8 (±1.2)
	energy scale	-	-	±0.9 (±3.0)	±0.9 (±2.4)
Jet	energy resolution	±1.5 (±2.2)	±2.9 (± 6.4)	±7.5 (±8.5)	±6.4 (±6.4)
	energy scale	±2.9 (±2.7)	±7.8 (± 5.6)	±3.0 (±3.3)	±2.3 (±3.4)
Flavor tagging	b-jets	±2.4 (±2.5)	±2.3 (± 1.4)	±3.4 (±2.6)	±2.5 (±2.6)
	c-jets	±0.1 (±1.0)	±1.8 (±11.6)	-	-
	light-jets	<0.1 (±5.0)	±1.6 (± 2.2)	-	-
Theory	PDF+α_S	±2.3 (±2.3)	±3.1 (± 3.3)	n/a	n/a
	Scale	+4.3 (+4.3)	+4.9 (+ 5.3)	n/a	n/a
		−6.0 (−6.0)	+7.0 (+ 8.0)	n/a	n/a
	EFT	±5.0 (±5.0)	n/a	n/a	n/a

Dominant:
Yield: Flavor tagging, JES, Photon ID, Pile-up
Scale: e/gamma Scale
Width: e/gamma Resolution, e/gamma Scale, JES, Pile-up, flavor tagging

Northern Illinois University

Results

No significant excesses seen, so exclusion limits set

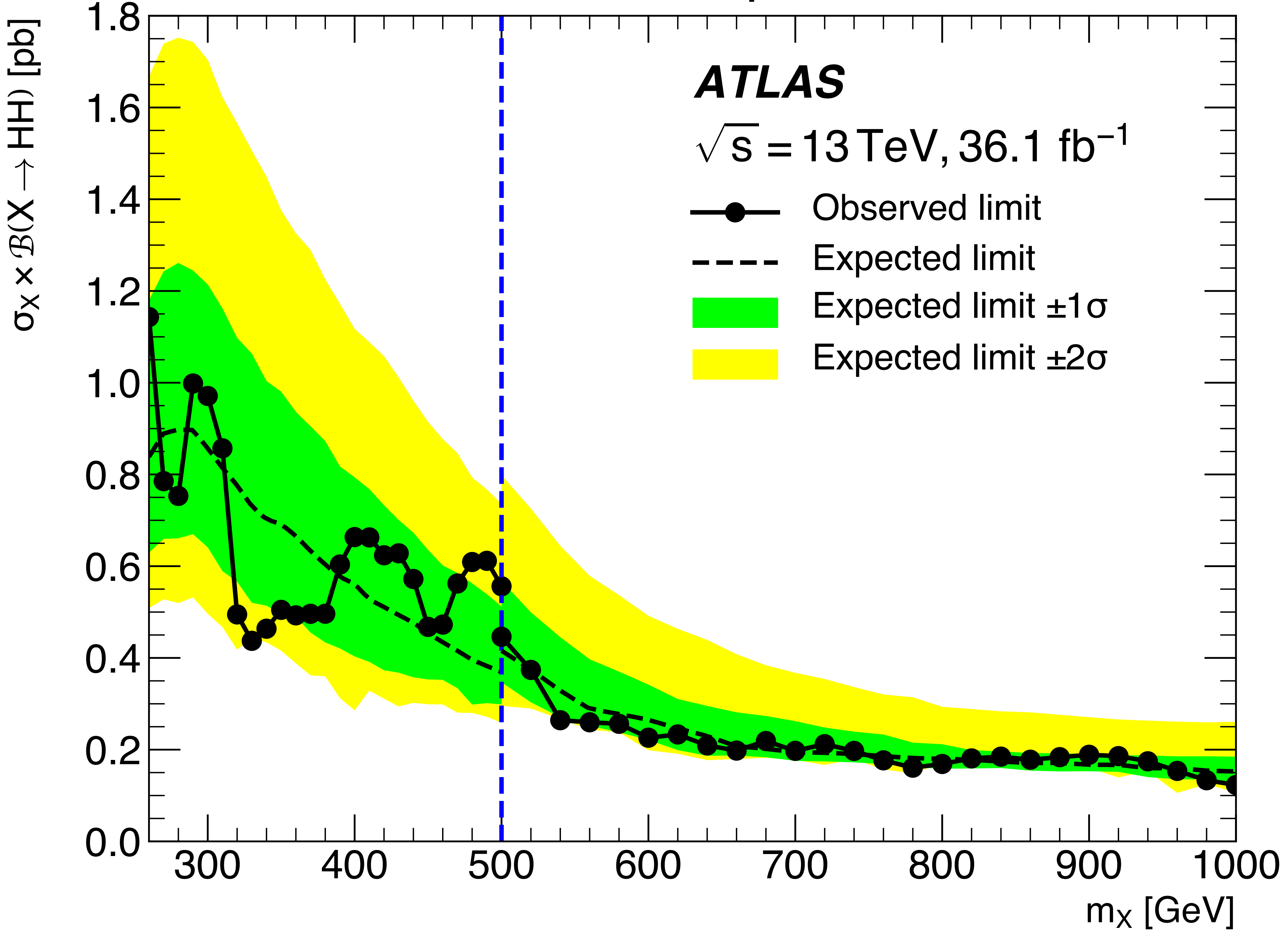
Limits on non-resonant HH production

	Observed	Expected	-1σ	$+1\sigma$
$\sigma_{gg\to HH}$ [pb]	0.73	0.93	0.66	1.3
Multiple of σ_{SM}	22	28	20	40

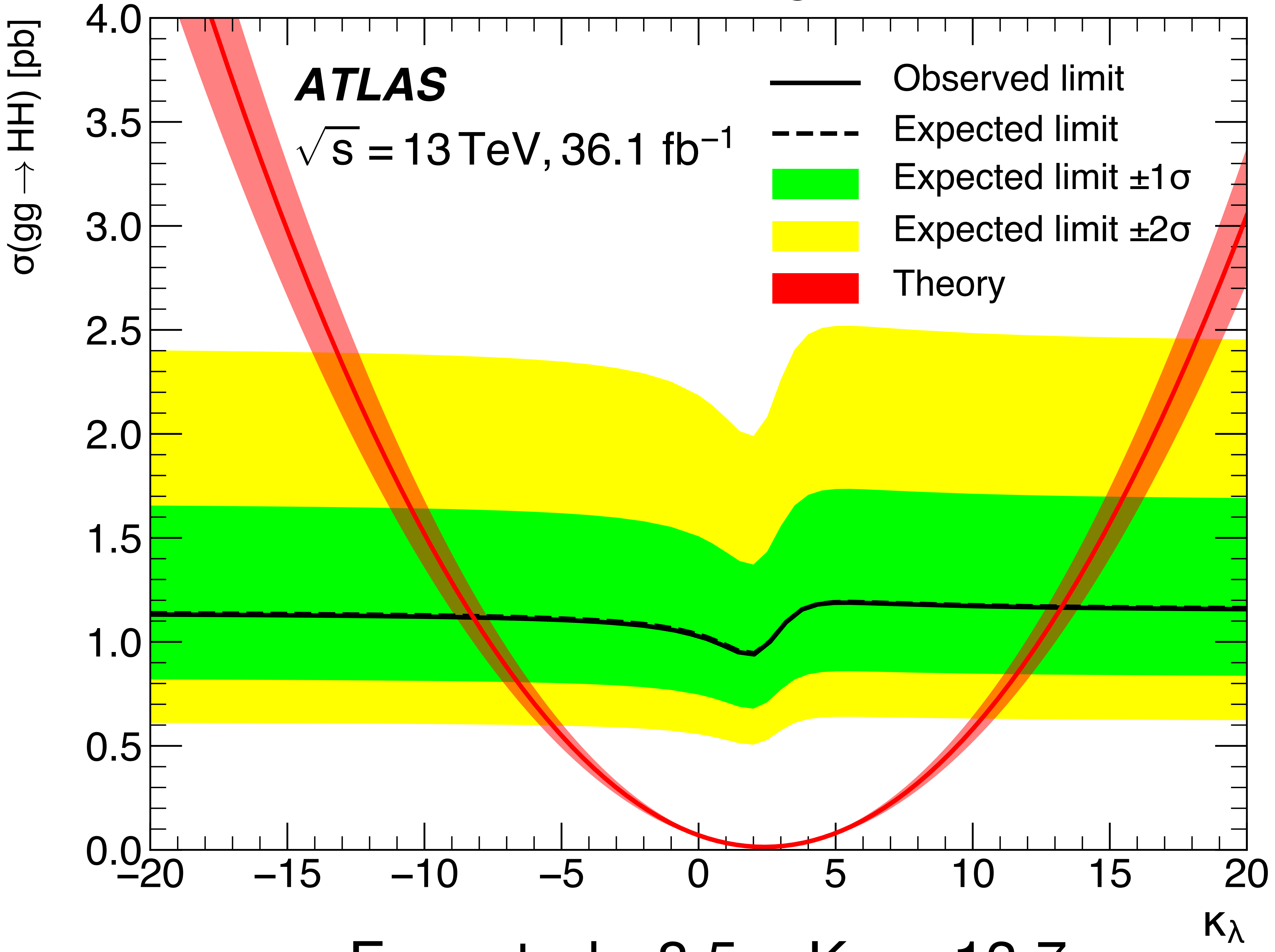
Use profile likelihood ratio test statistic on $m_{\gamma\gamma bb}$ ($m_{\gamma\gamma}$) for the (non-)resonant analysis to derive signal yields

CLs prescription used on POI to set 95% CL
Limits are generated by using "toy" pseudo-experiments

Limits on resonant HH production



Limits on trilinear coupling



Expected: $-8.5 < K_\lambda < 13.7$
Observed: $-8.2 < K_\lambda < 13.3$

But also looking forward to future di-Higgs analyses!

- ~4x more data when using full Run 2 dataset
- **New analysis techniques and approaches**

Northern Illinois University

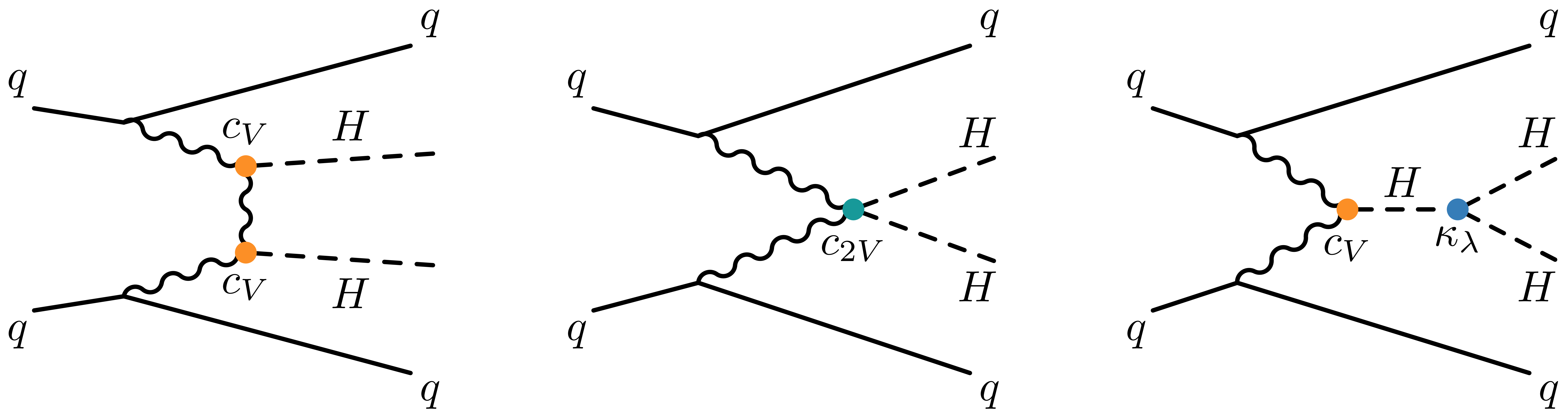
Vector Boson Fusion HH Production

SM Vector Boson Fusion Production

ggF is the dominant HH production mode, but what if we consider subdominant modes?

Vector Boson Fusion

- Subdominant production mode, cross-section factor of 20 smaller than ggF



$$\sigma_{HH}^{ggF} = 33.4 \text{ fb}^{-1}$$

$$\sigma_{HH}^{vbf} = 1.64 \text{ fb}^{-1}$$

Small cross-section motivates use of machine learning (ML) to get the most pure possible signal selection

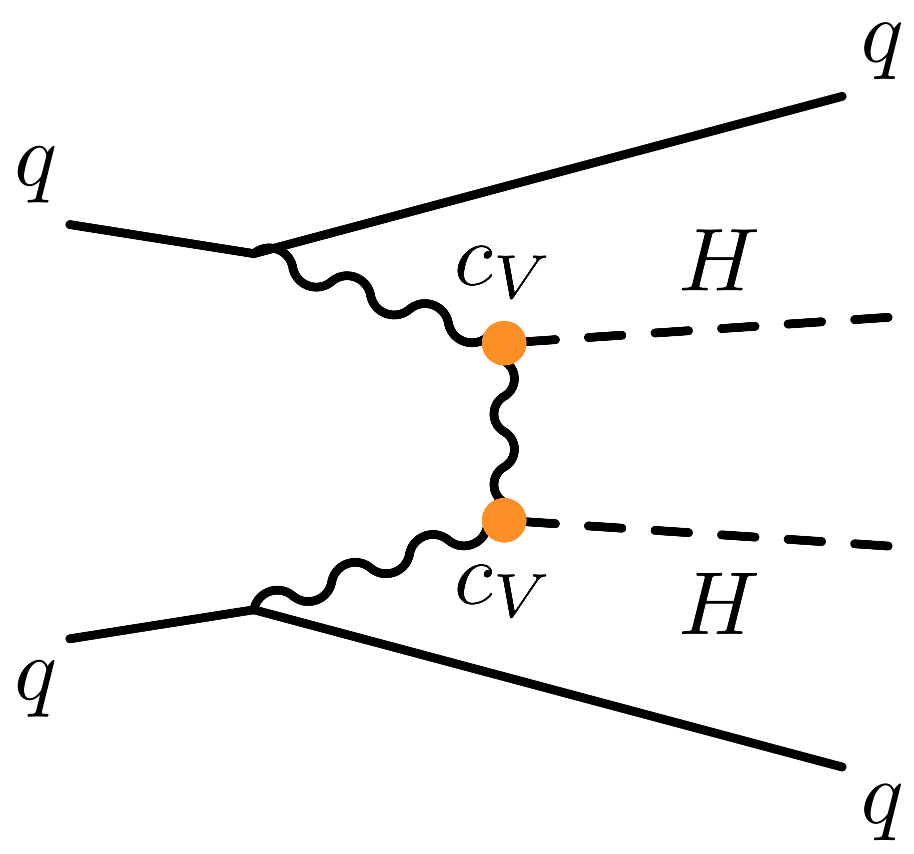
However - distinct topology that makes for easy signal discrimination

- Production yields 2 extra jets from quarks radiating vector bosons
- Extra jets have unique properties
  - Very forward in the detector
  - High m_{jj}, compared to H→bb jets

Northern Illinois University

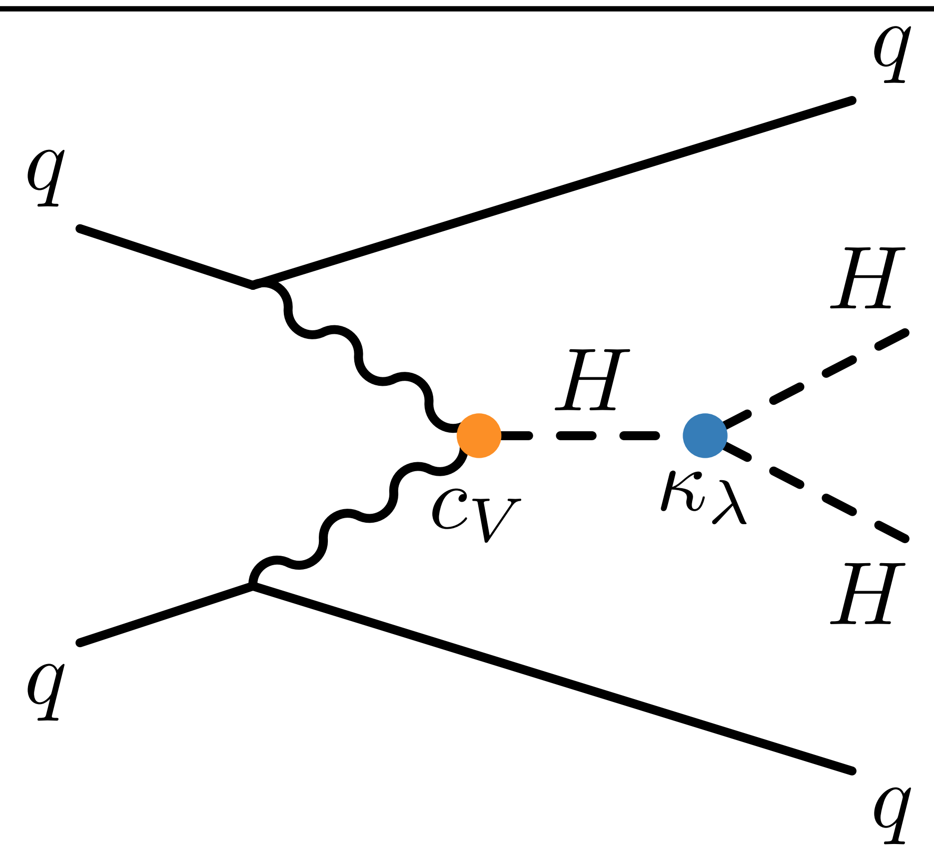
BSM Vector Boson Fusion Production

What if coupling strengths deviate from SM predictions?

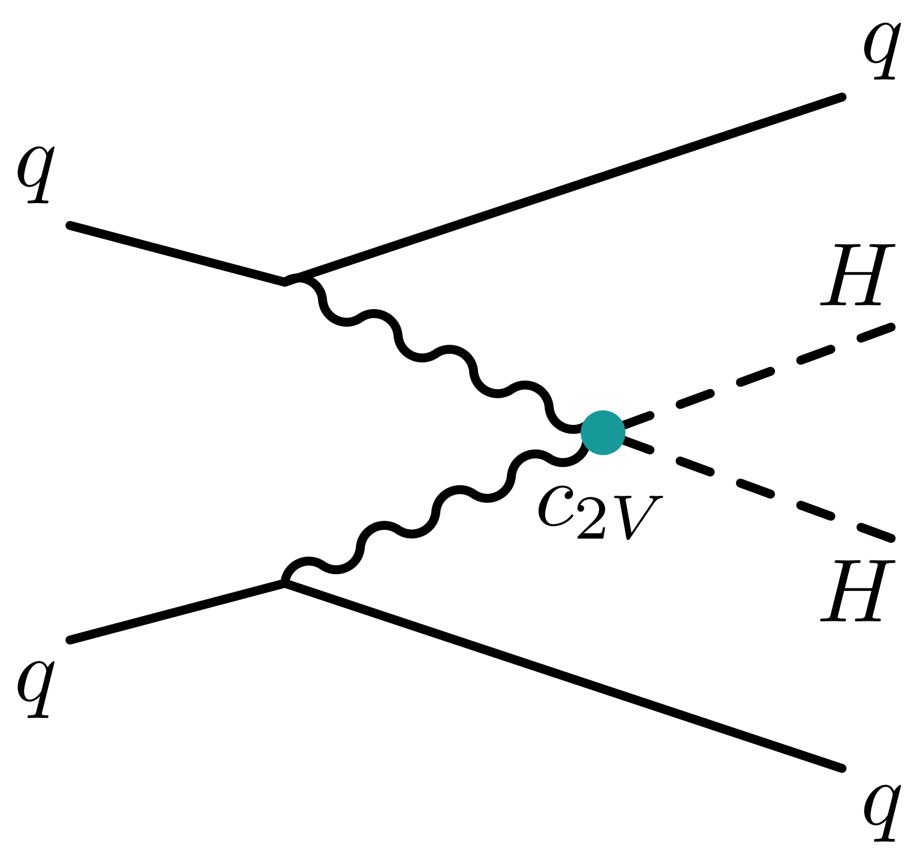


C_V (VVH) - Constrained by mono-Higgs analyses

ATLAS-CONF-2018-031 $\sigma_{C_V}/\sigma_{C_V,SM} = 1.21\pm^{0.22}_{0.21}$



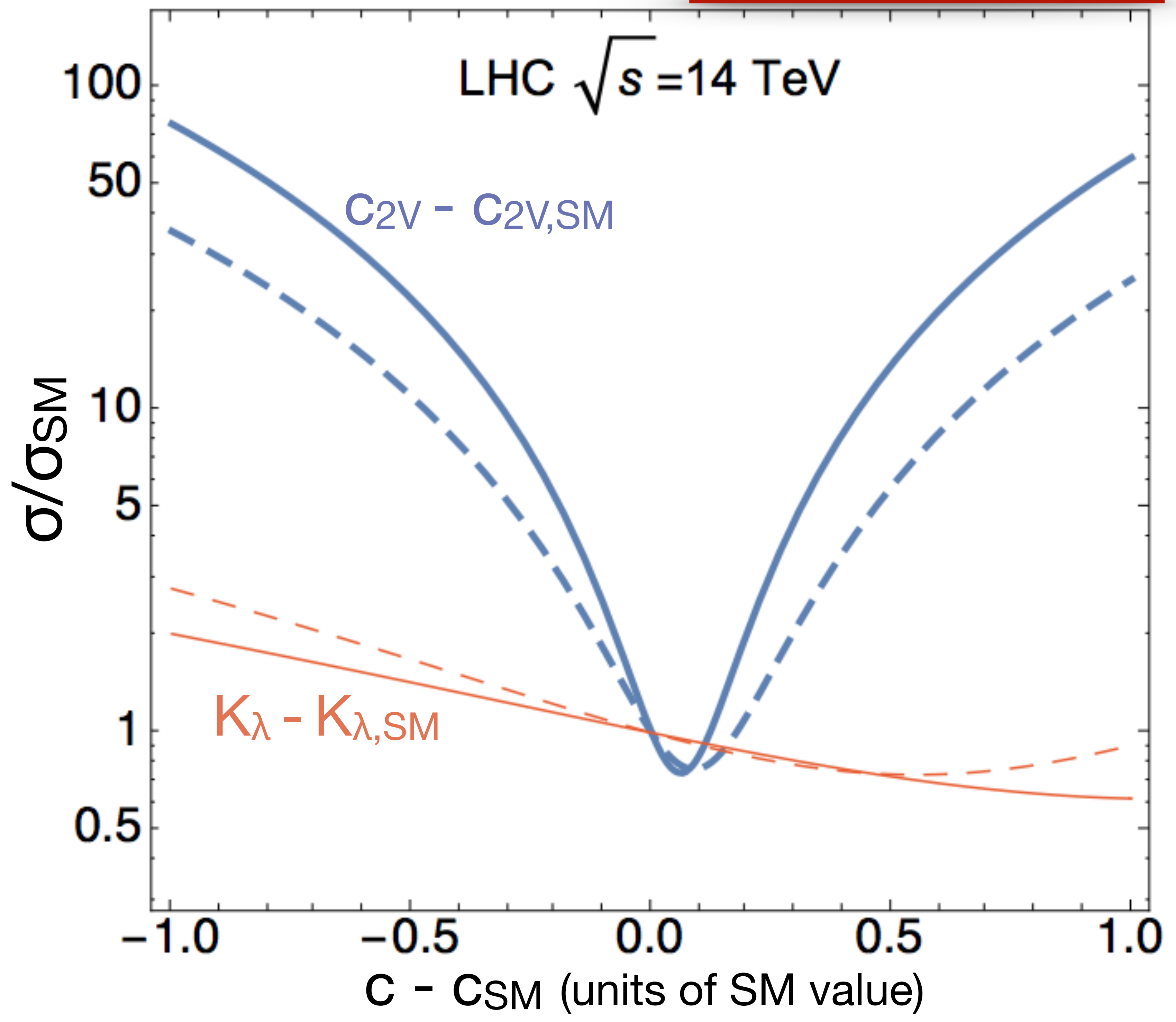
κ_λ (HHH) - VBF HH production is also sensitive to deviations in κ_λ, and can contribute to limits



C_{2V} (VVHH) - VBF HH provides unique handle at tree level to this coupling!

Greatly enhances HH cross-section, even at small deviations from predictions

Bishara, Contino, Rojo
Eur. Phys. J. C 77 (2017) 481

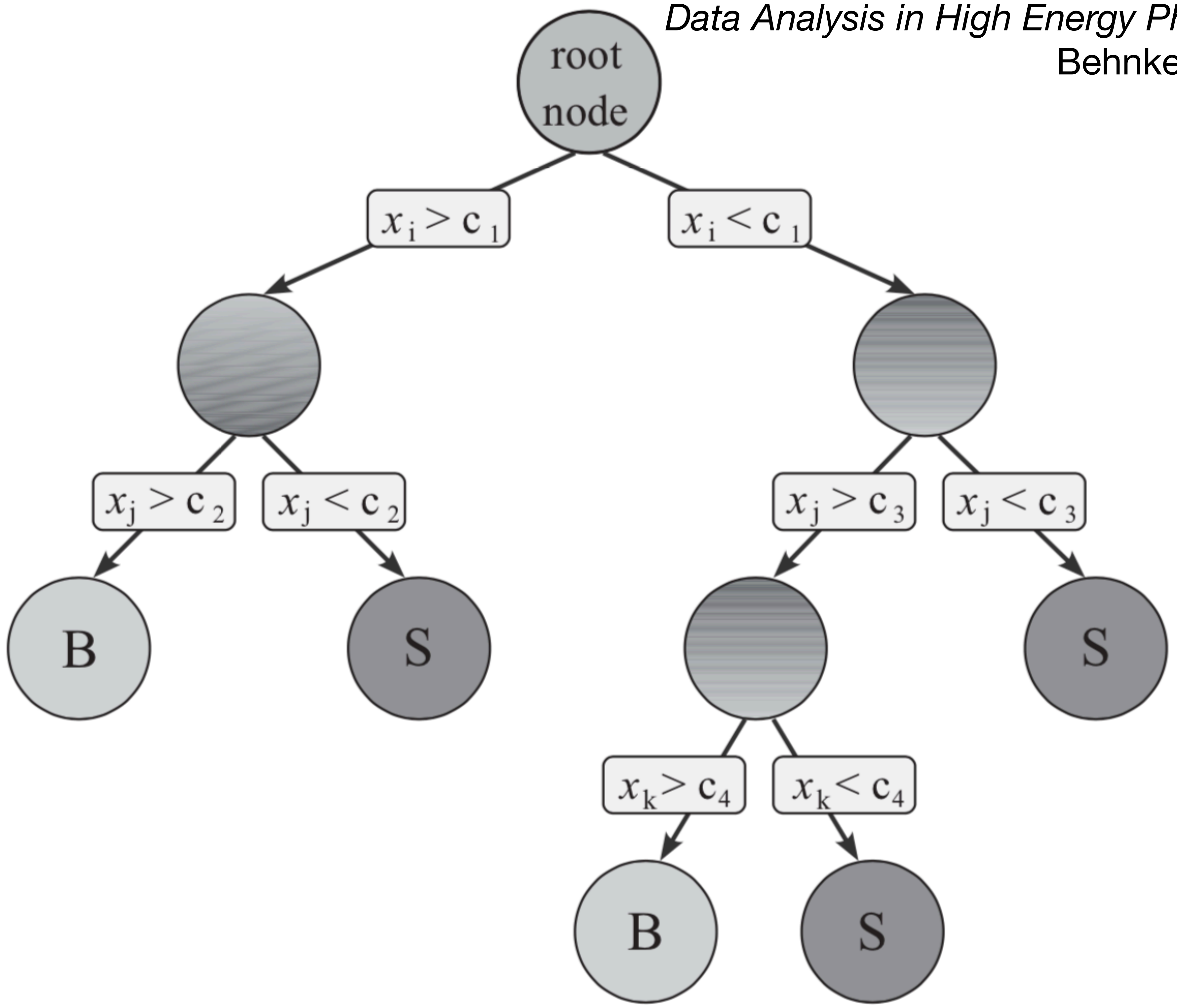


Lines correspond to cuts in pheno study:
Solid - broad acceptance cuts
Dashed - tight analysis cuts (targeting VBF HH)

Northern Illinois University

Machine Learning Interlude #1
Boosted Decision Trees

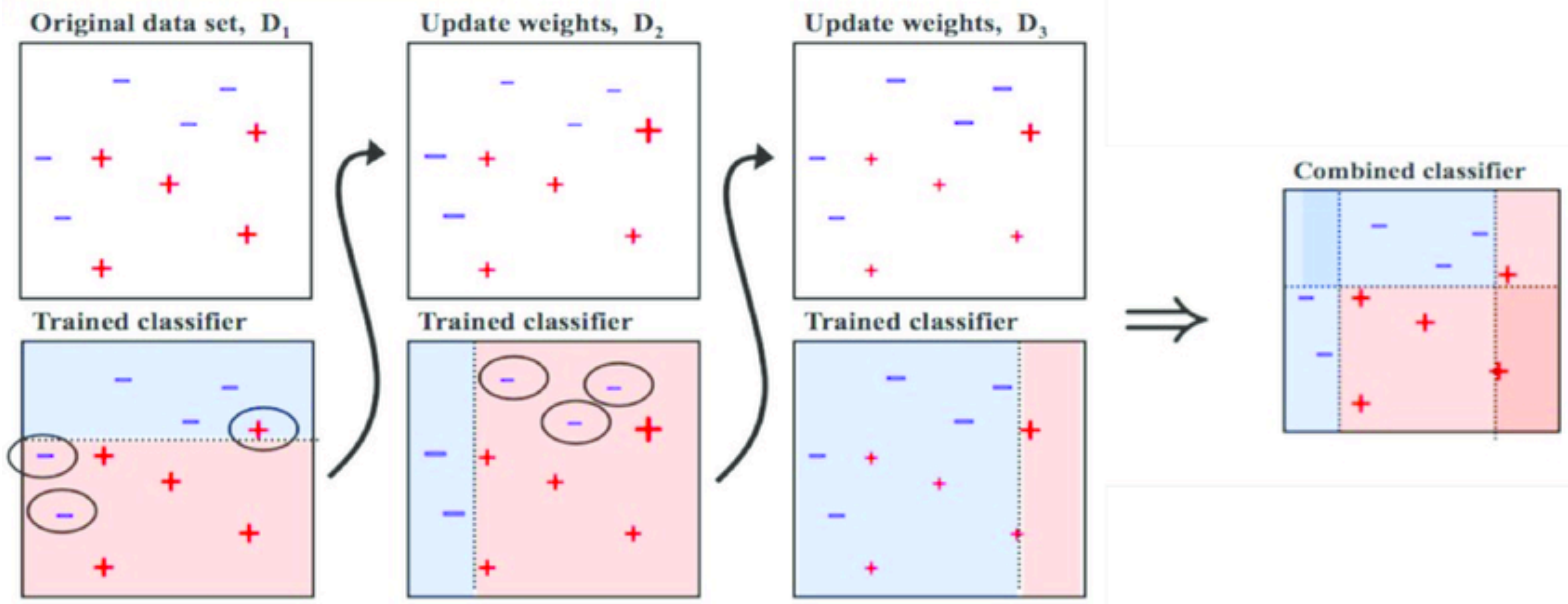
Data Analysis in High Energy Physics
Behnke et. al



Decision Tree:
- Familiar "tree" based learning - split data using input variables, maximize a splitting criteria (e.g. signal/background separation)
- Continue splitting until a stopping criteria is met (e.g. depth, samples in a node)

Boosting:
- After building tree, find misclassifications, upweight, and derive a new tree, do this for a defined number of trees
- Yields complex output space!
- "Ensemble method" - combining many weak learners to make one strong learner

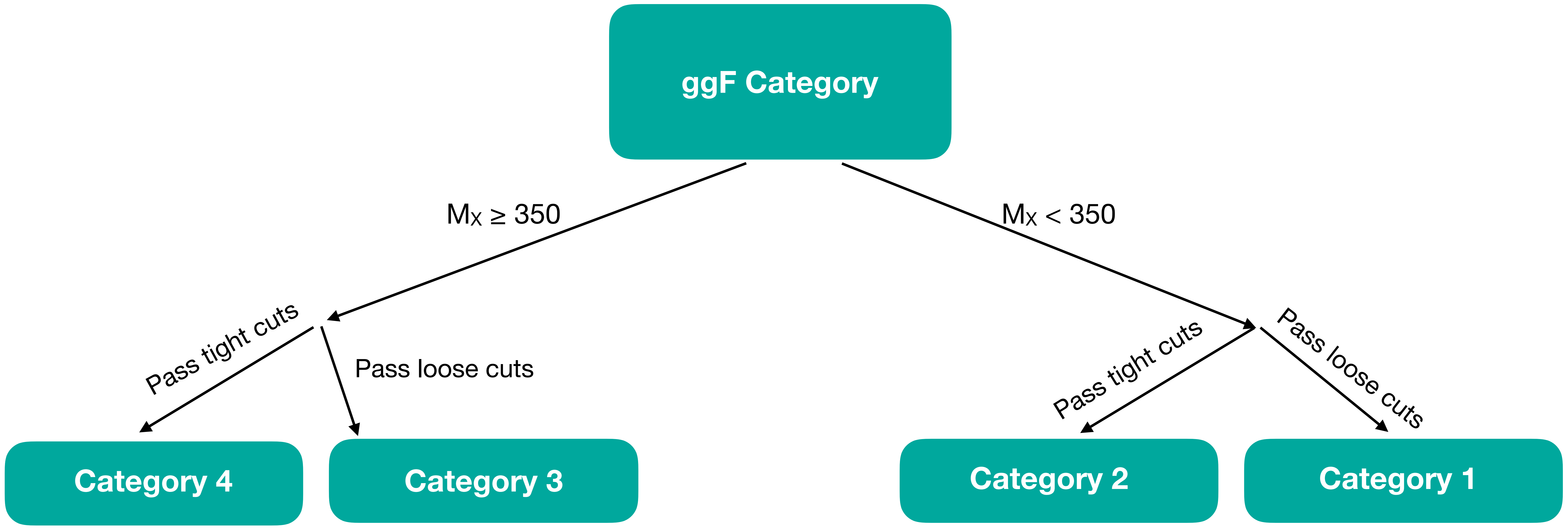
Parameters in ML adapt by learning from observations (data and/or MC) - "learning" stage is called *training*



Notes
- Commonly used in physics, easy to integrate into ATLAS software
- Good interpretability

Full Run 2 Analysis Selection

Next iteration of analysis will construct 4 signal regions using a BDT trained on ggF signal
- Targets differences in di-Higgs kinematics from backgrounds



Tight and loose cuts defined using BDT scores

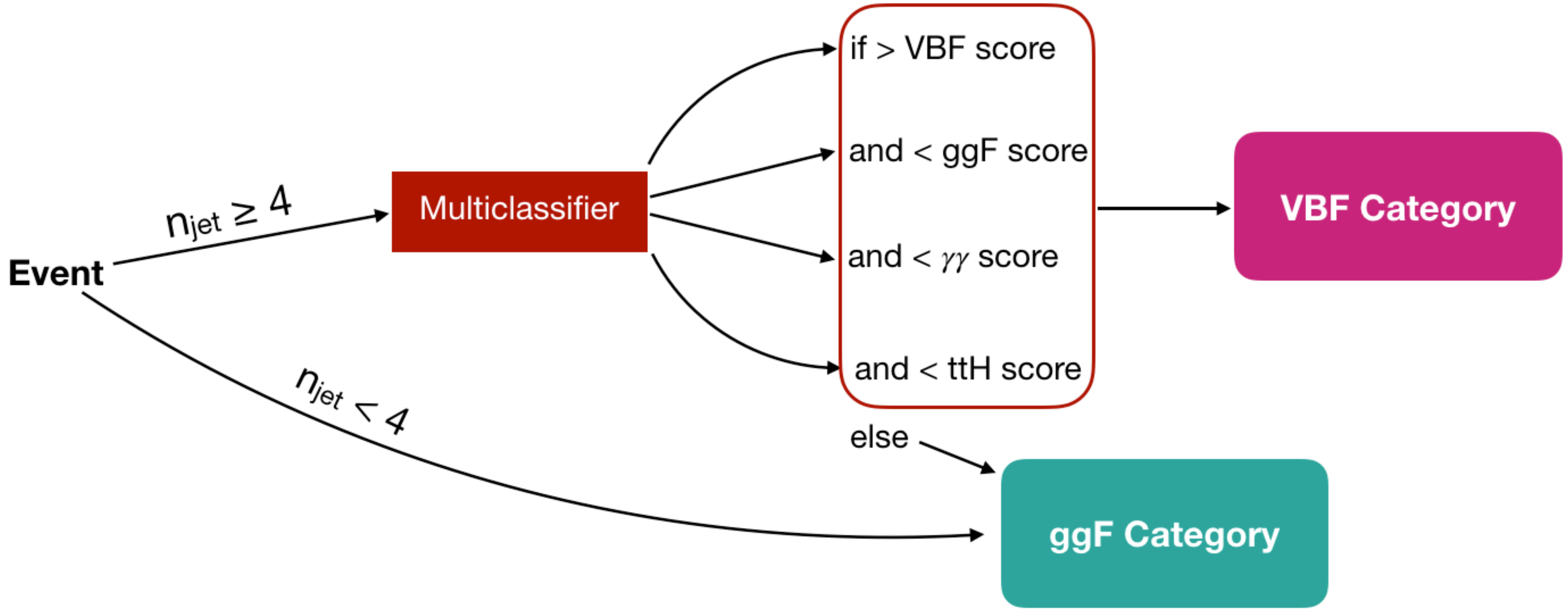
How to fit VBF production into this?

Northern Illinois University

Implementing into Analysis

Create a VBF-enriched signal region - use unique VBF topology to isolate this production from background and ggF HH production

- First considered just $\gamma\gamma$-continuum background, but found large contamination from ttH so developed a **multiclassifier** BDT, independent probabilities for VBF HH, ggF HH, $\gamma\gamma$, and ttH



(actually 4 categories shown in prior slide)

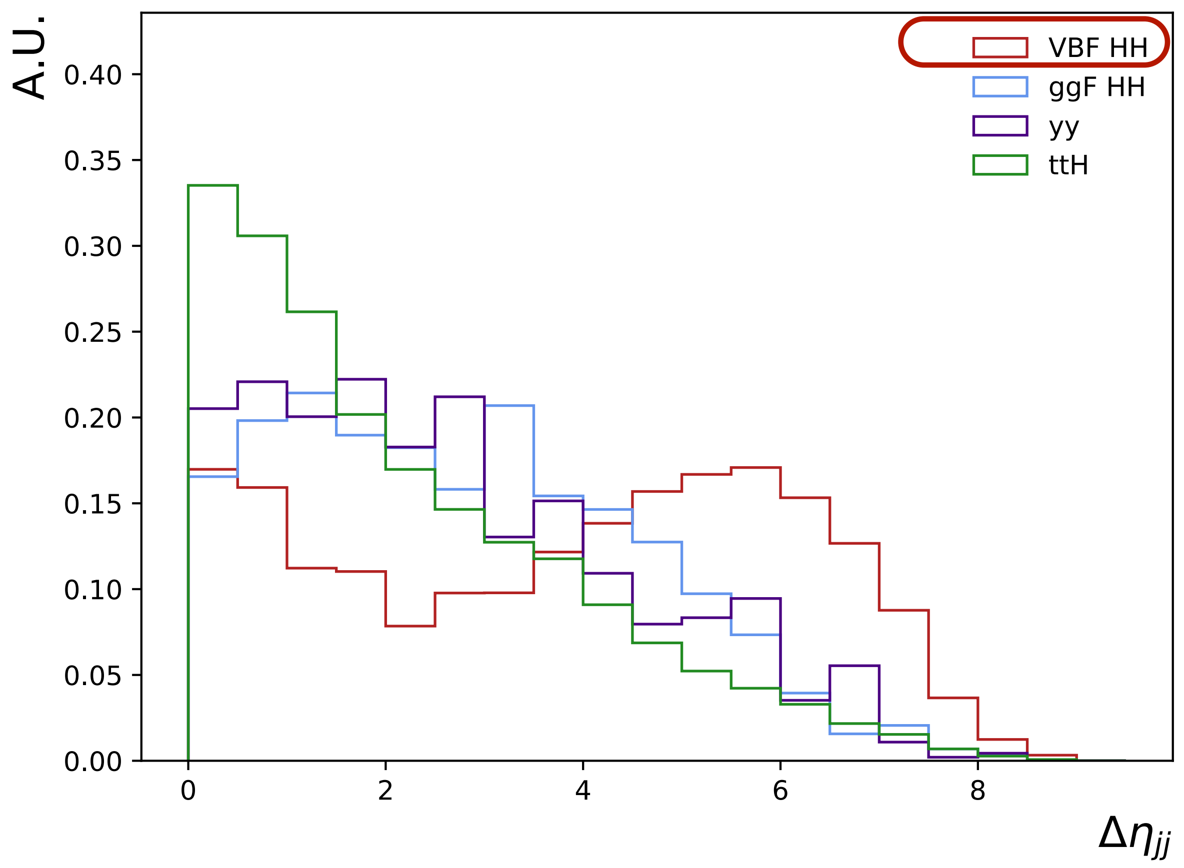
In total **26 variables** investigated as BDT inputs, through a variable pruning, **reduced to 10** with equivalent performance
- Pruning motivated by cross correlations, correlation to final output, and final exhaustive search of all possible subsets

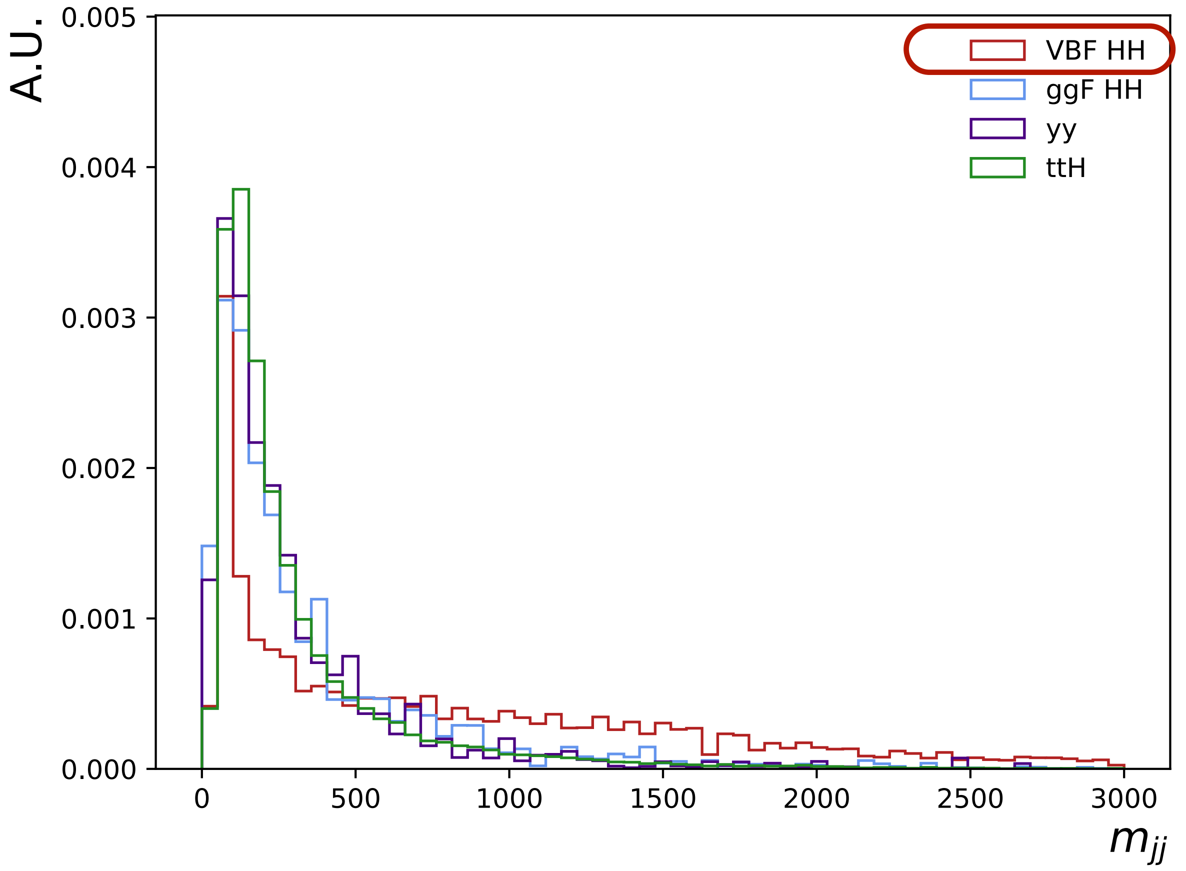
BDT Inputs **(not all shown here)**

BDT developed using XGBoost

VBF-targeting

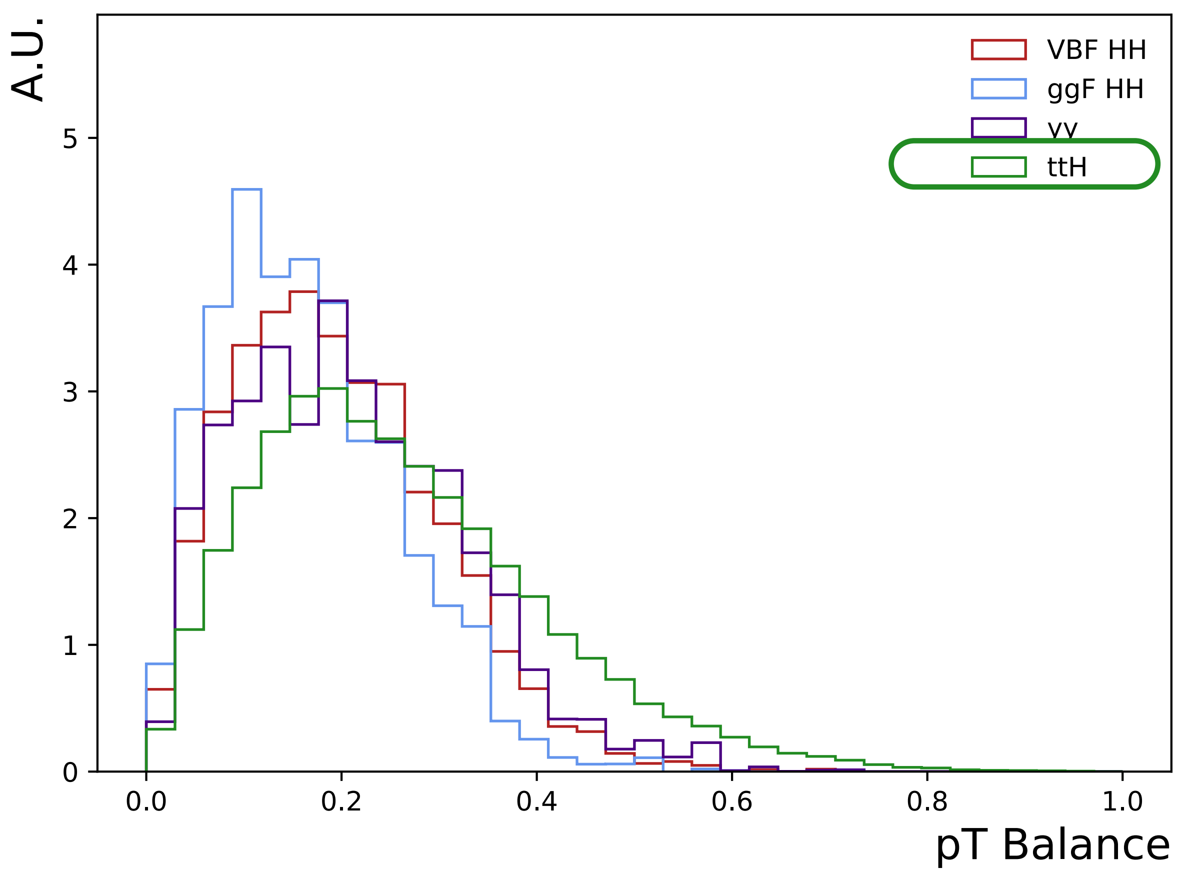
Take advantage of unique VBF topology to separate out production mode

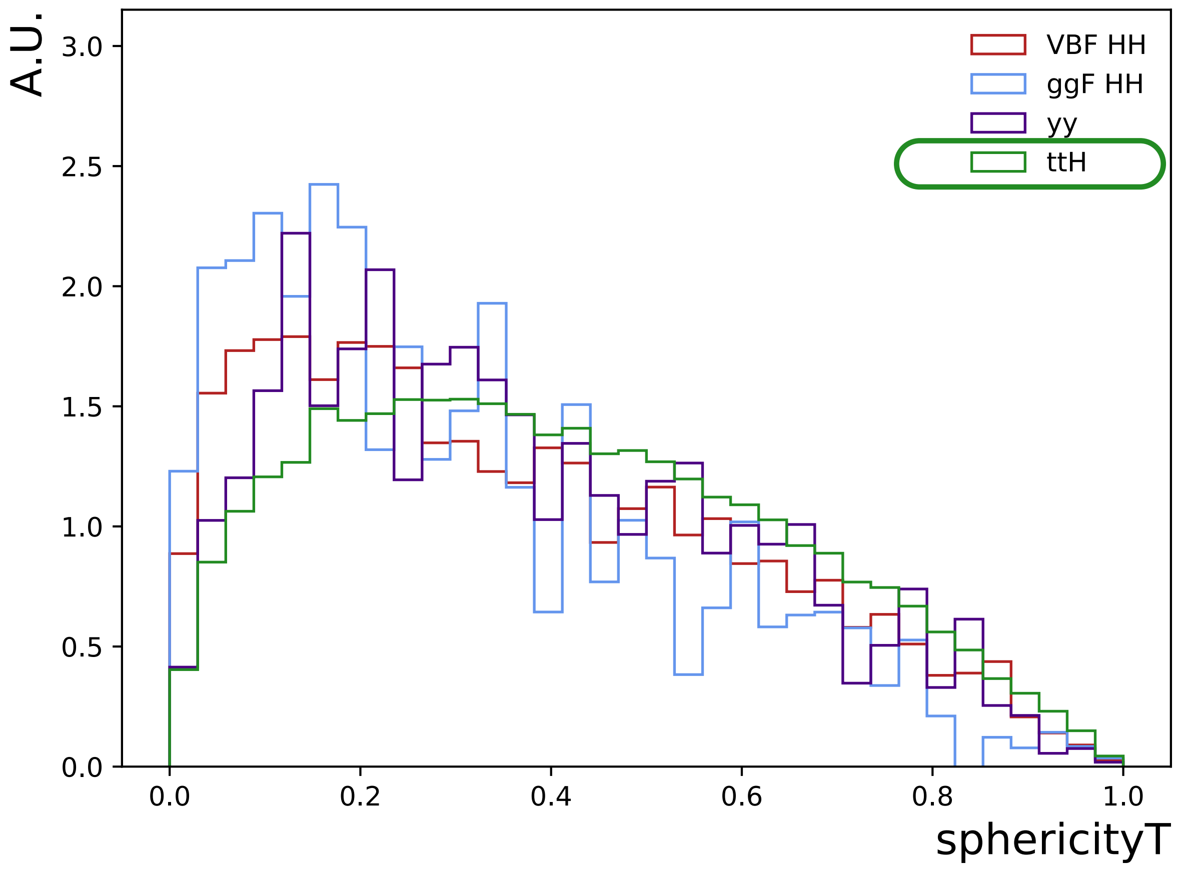




ttH-Rejecting

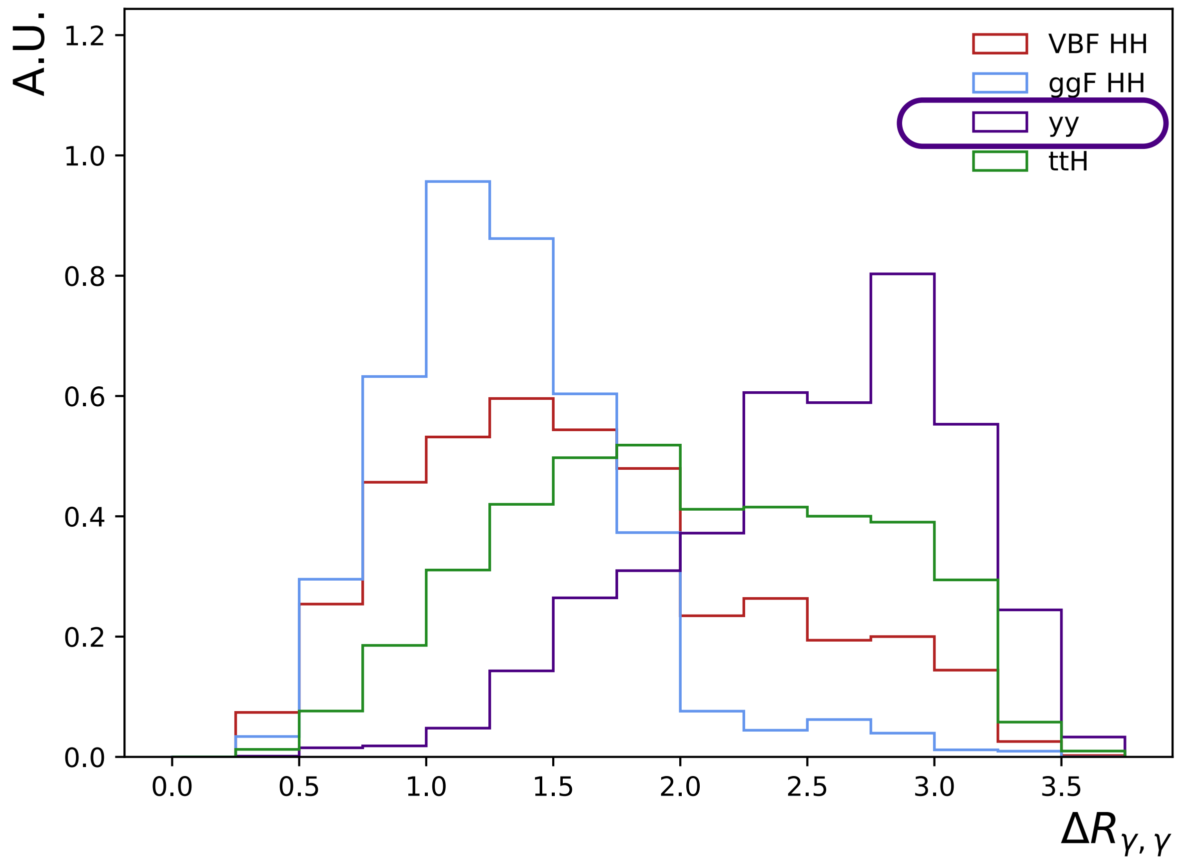
"Event shape" variables - describe the flow and distribution of energy in an event

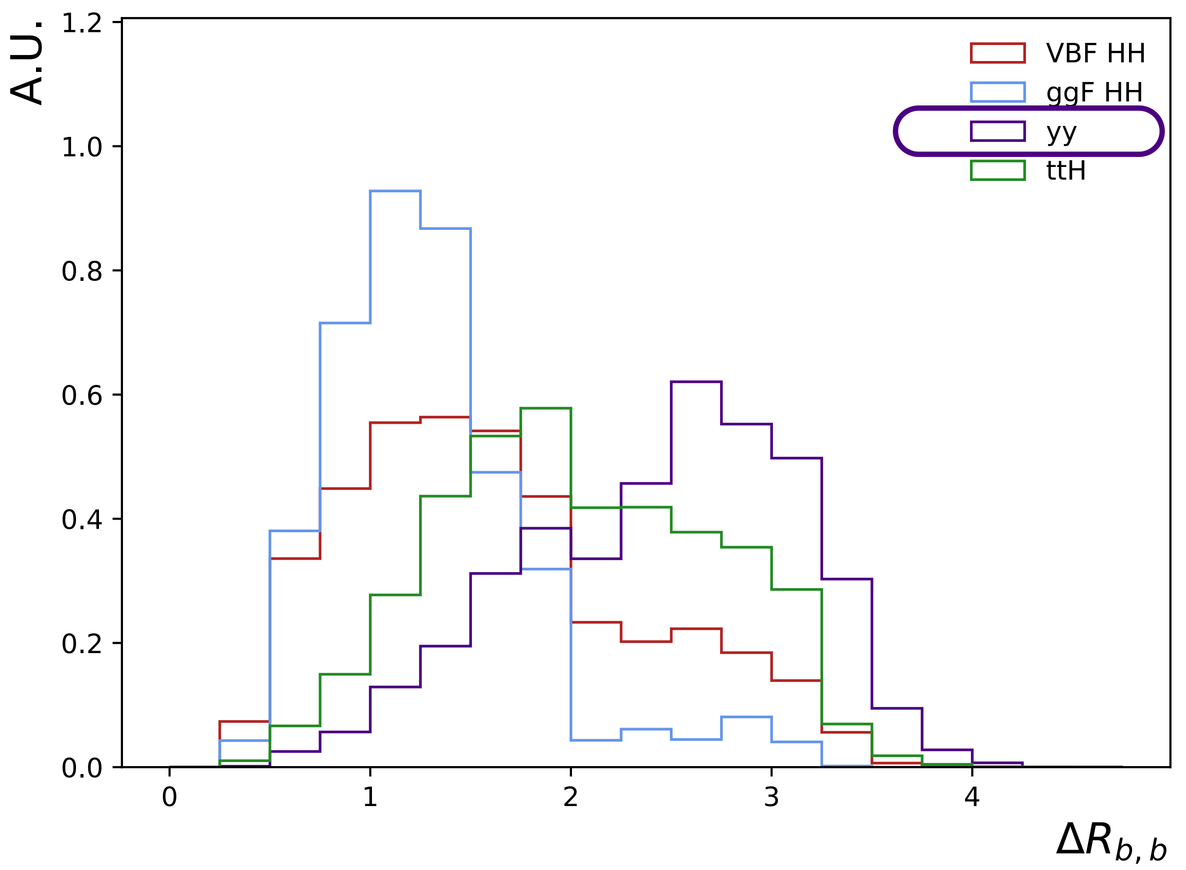


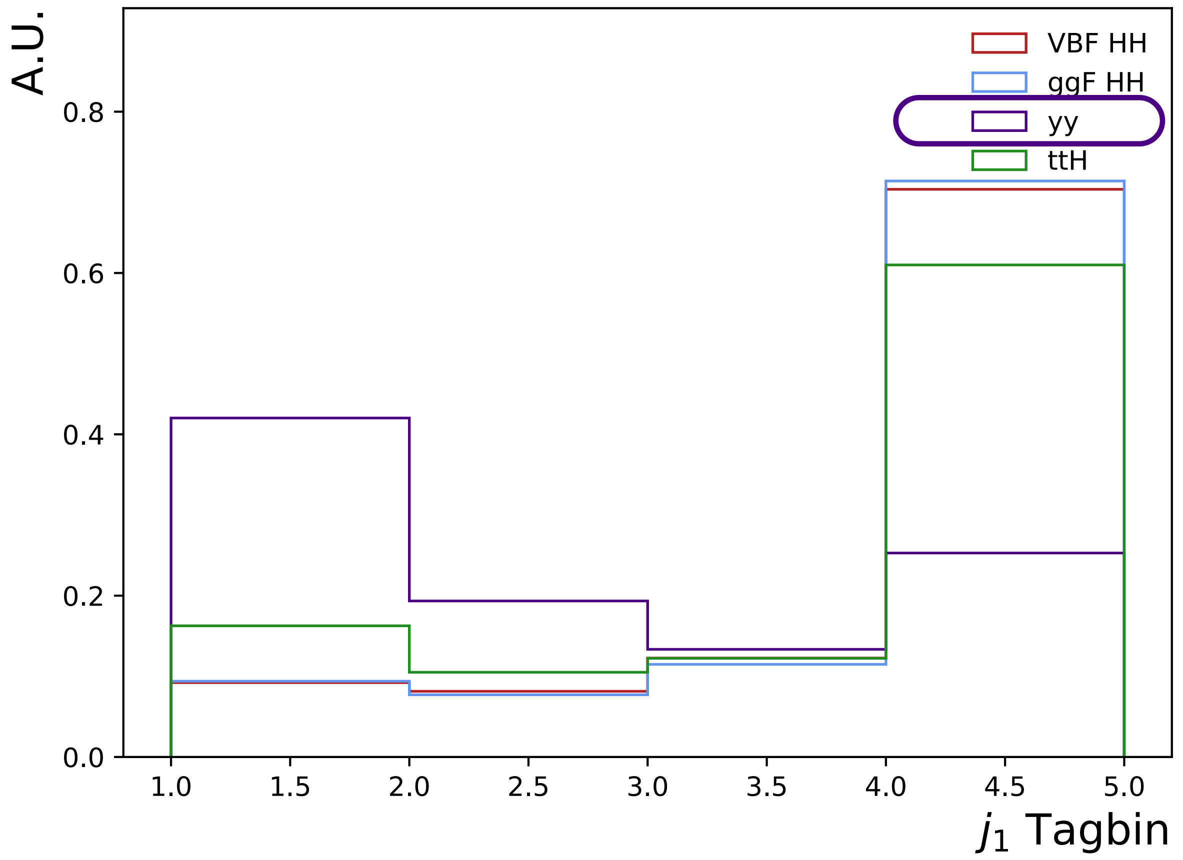


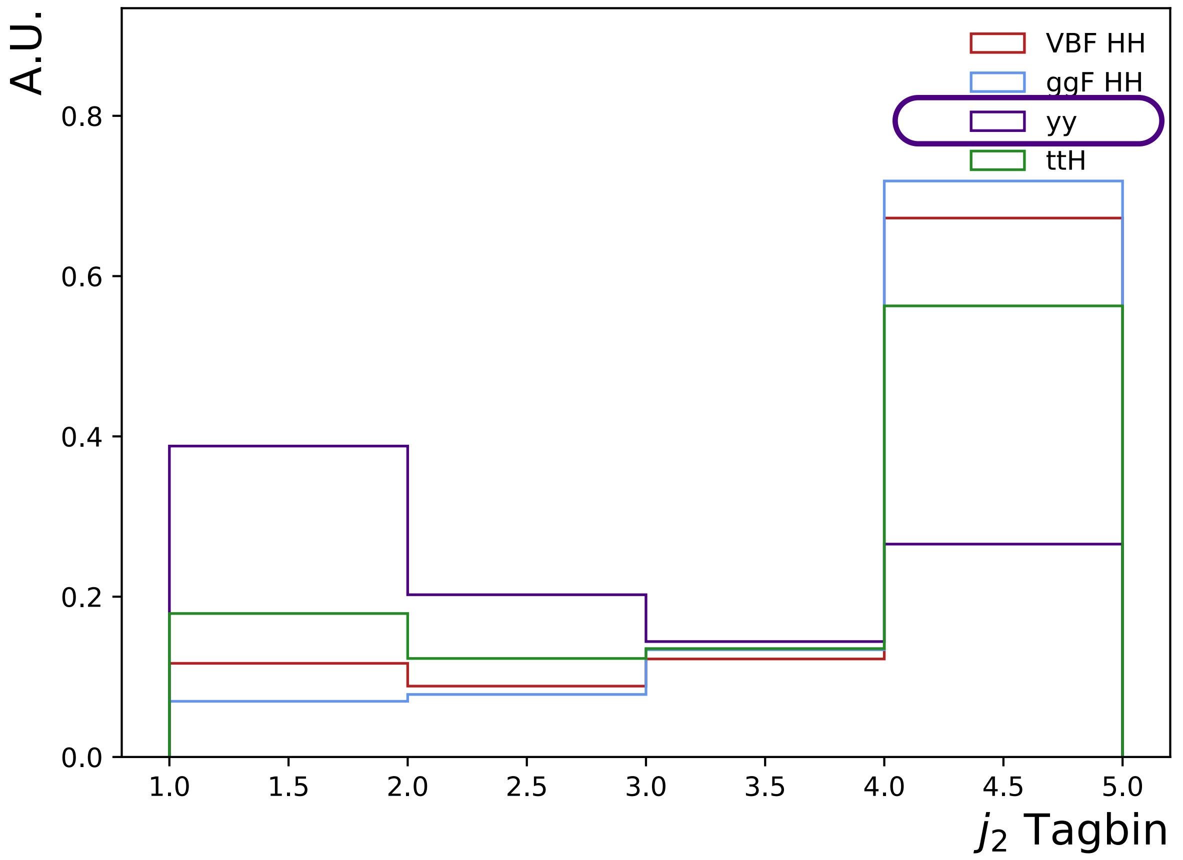
Higgs Kinematics (reject $\gamma\gamma$)

Use differences in Higgs decay kinematics and b-tagging to differentiate from dominant background





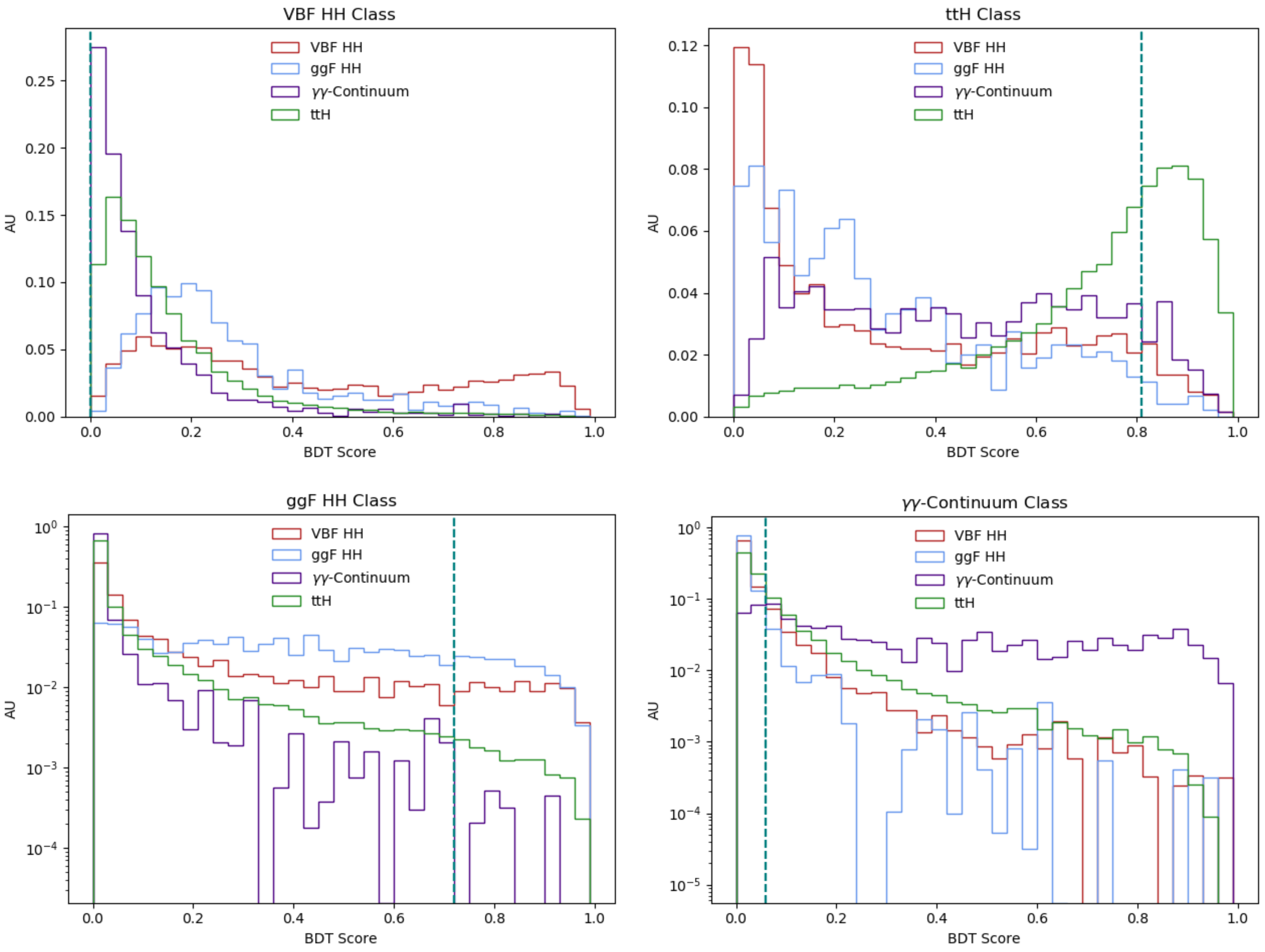




BDT Optimization

Probability cuts to define category selected via a 4-d scan over the 4 BDT scores.

- On each step, calculate the combined Asimov significance, and select highest value



Asimov significance given by

$$Z = \sqrt{2[(s+b)\ln(1+s/b) - s]}$$

At each step, calculate value for each category, add in quadrature (sensitivity to full HH cross section)

$$Z_{\textbf{total}} = \sqrt{Z^2_{ggF_{cat1}} + Z^2_{ggF_{cat2}} + Z^2_{ggF_{cat3}} + Z^2_{ggF_{cat4}} + Z^2_{VBF}}$$

selecting most optimum point gives

$$Z = 0.459$$

which is a 9.7% improvement over not defining such a category (Z=0.415)

Northern Illinois University

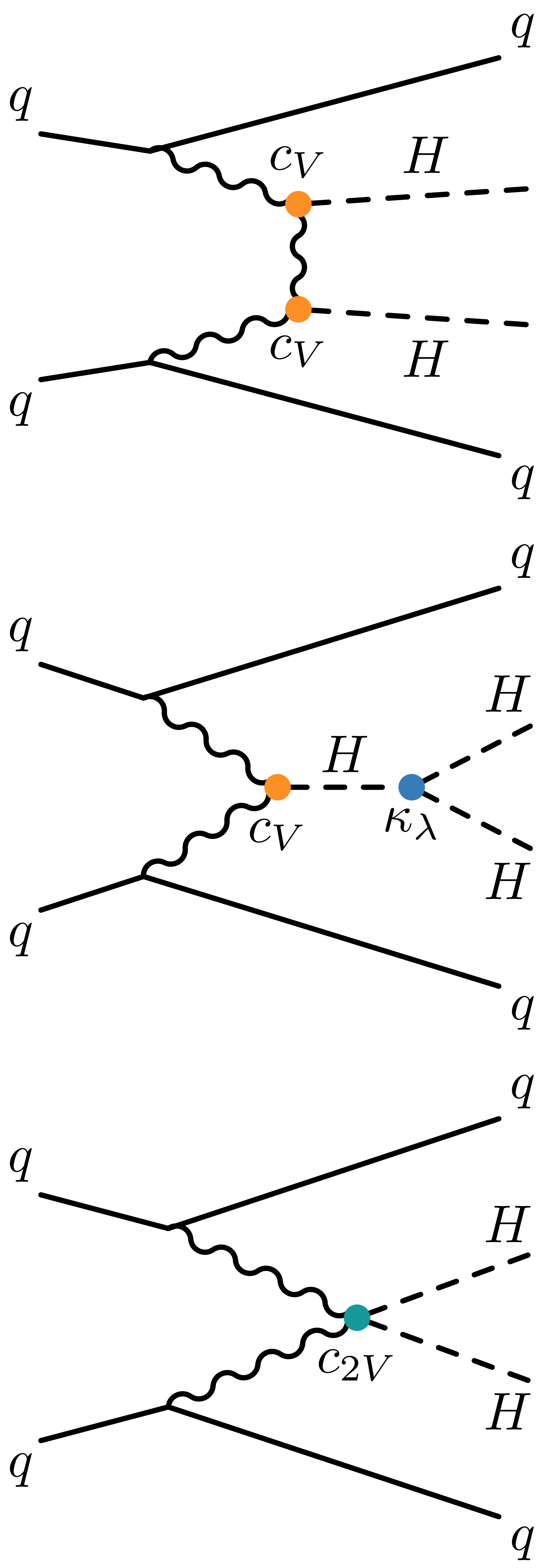
VBF HH Conclusion

Studying VBF HH production provides interesting insights into di-Higgs production

- VBF HH production is characterized by a unique topology, good for signal/background discrimination
- Studying VBF HH production provides sensitivity to new couplings, specifically c_{2V} (VVHH vertex), and deviations from SM prediction can cause large enhancements

The upcoming HH→$\gamma\gamma$bb analysis will incorporate a VBF-enriched signal region

- A multiclassifier BDT is used to separate VBF HH production from ggF HH as well as $\gamma\gamma$-continuum and ttH backgrounds
- After optimizing cuts on 4 probability scores, an improvement of 9.7% in Asimov significance can be found by adding this category



Northern Illinois University

Photon Identification

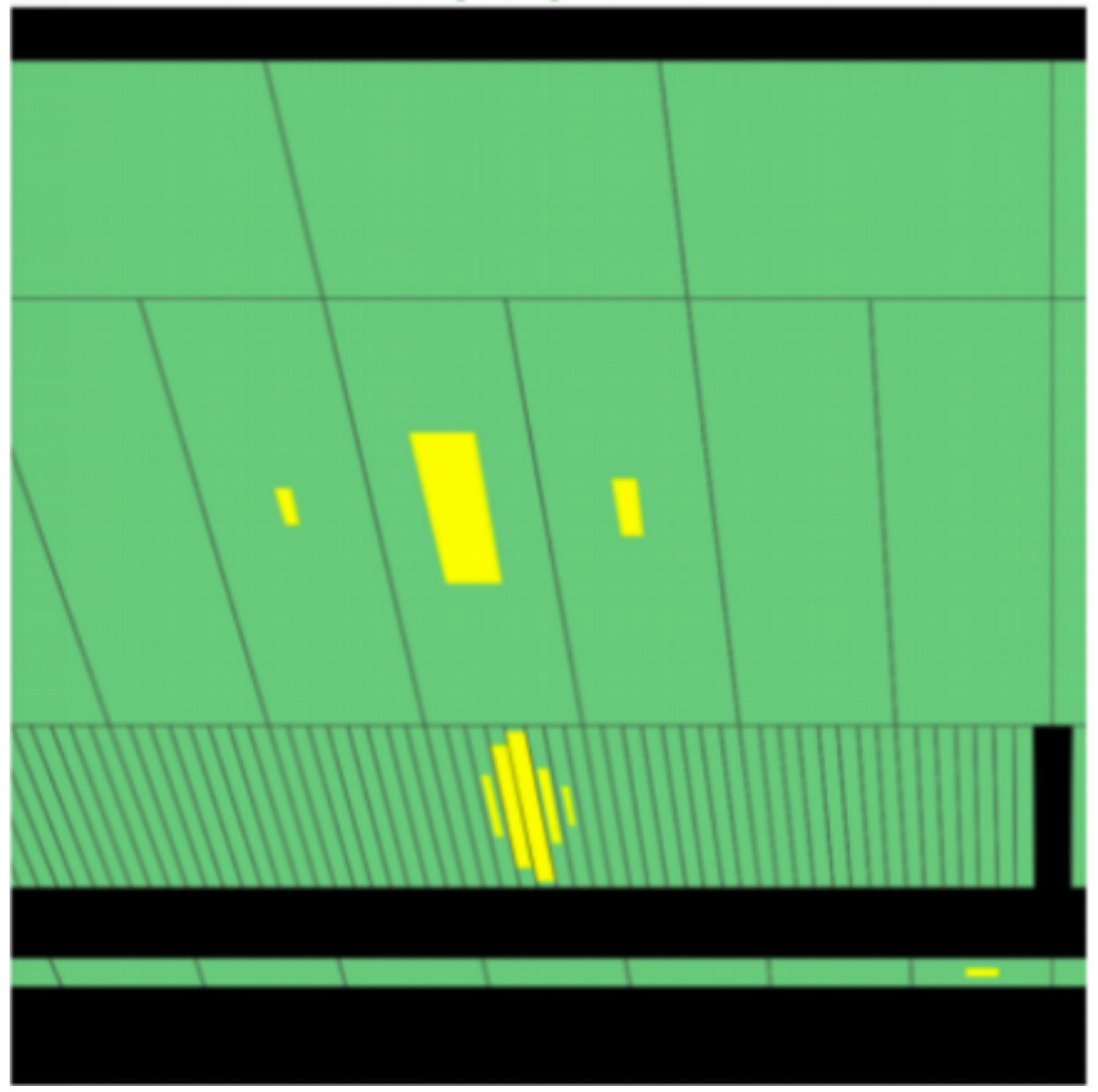
$\gamma\gamma bb$ is sensitive to the square of the photon identification efficiency - motivates searching for improvements in photon ID

Photon Identification in ATLAS

Goal: Classify energy depositions in the EM calorimeter
- Identify "prompt" photons - come from colliding protons, process of interest
- Reject photons resulting from background processes (e.g. hadronic decays in jets)

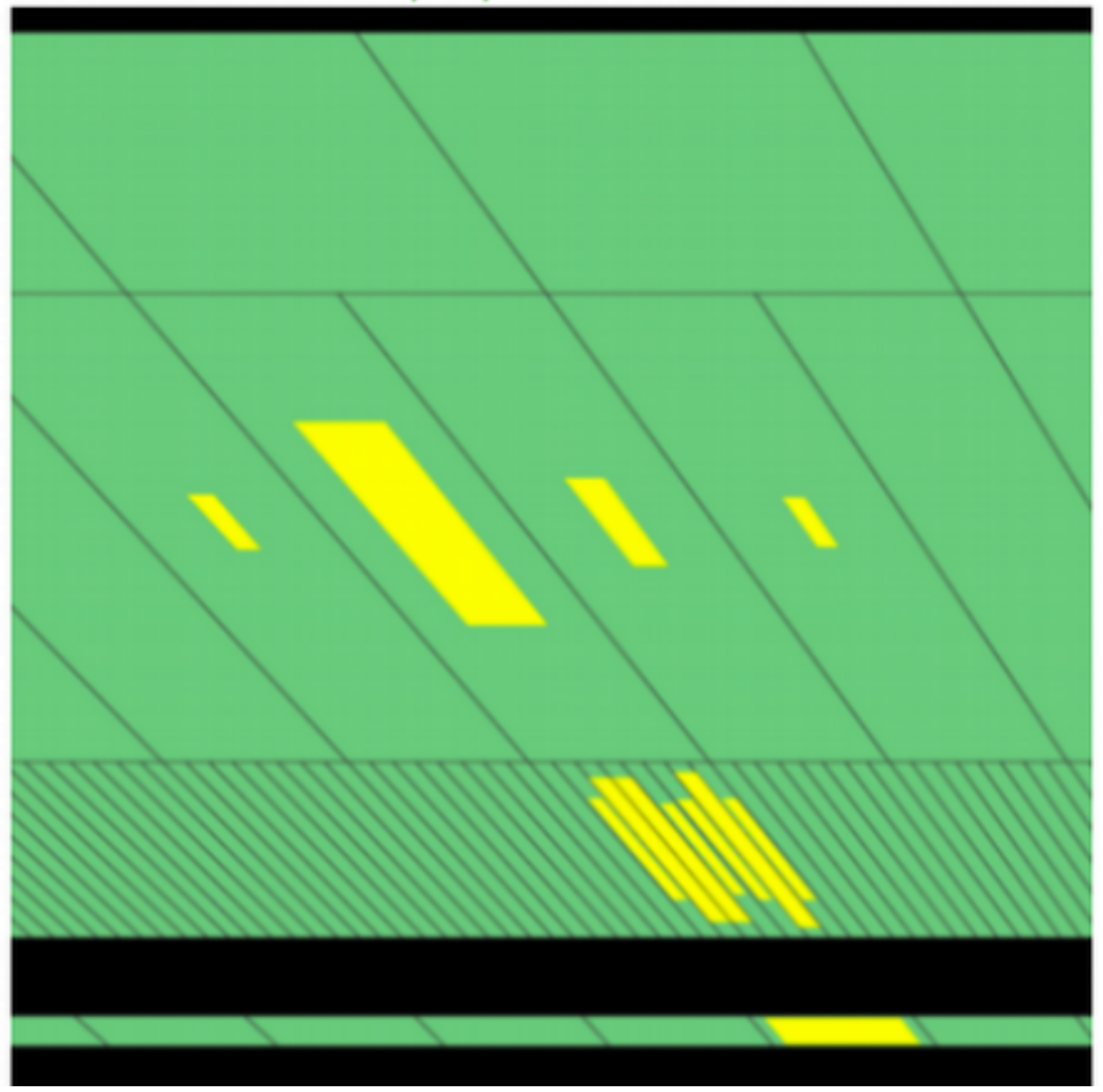
Want to discriminate

Prompt photon (γ)



from

$\pi^0 \rightarrow \gamma\gamma$



Northern Illinois University

Current Methodology

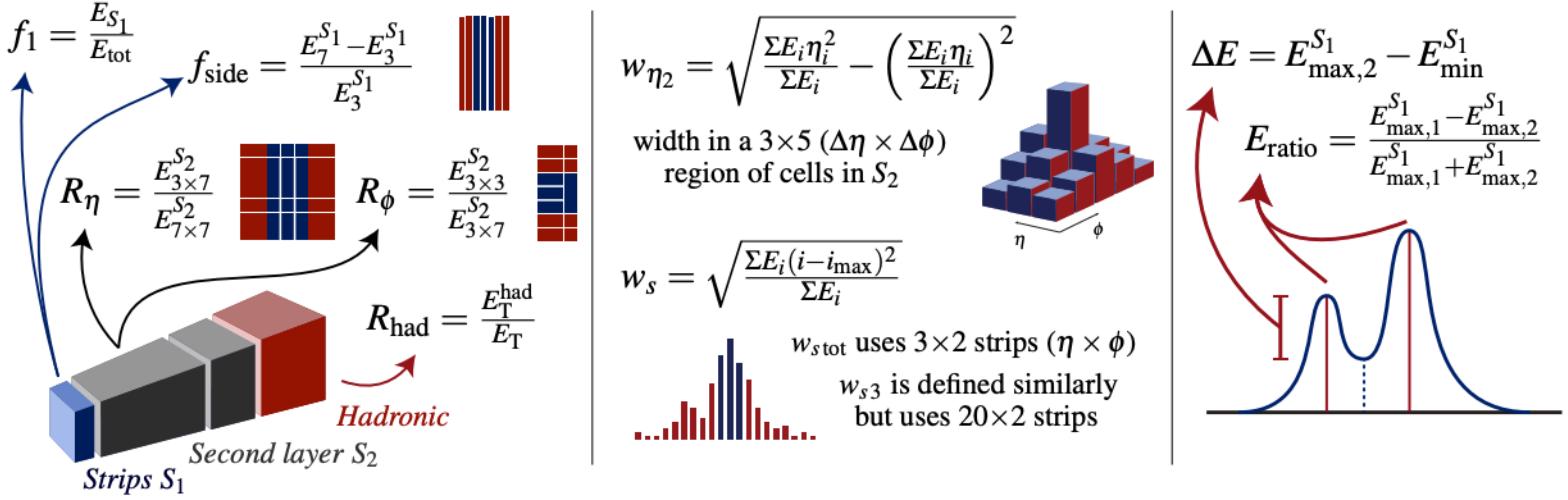
Showering - Description of particle interaction with detector material

- For photons, this is successive conversion to electron-positron pairs and bremsstrahlung photon emission

Cutting on shower shape variables

- 9 variables that describe longitudinal and lateral shower development in the calorimeter
- Showering dependent on detector segmentation - **bin in η, p_T, and conversion status**

Shower Shapes | Widths | Energy Ratios



How can we improve this?

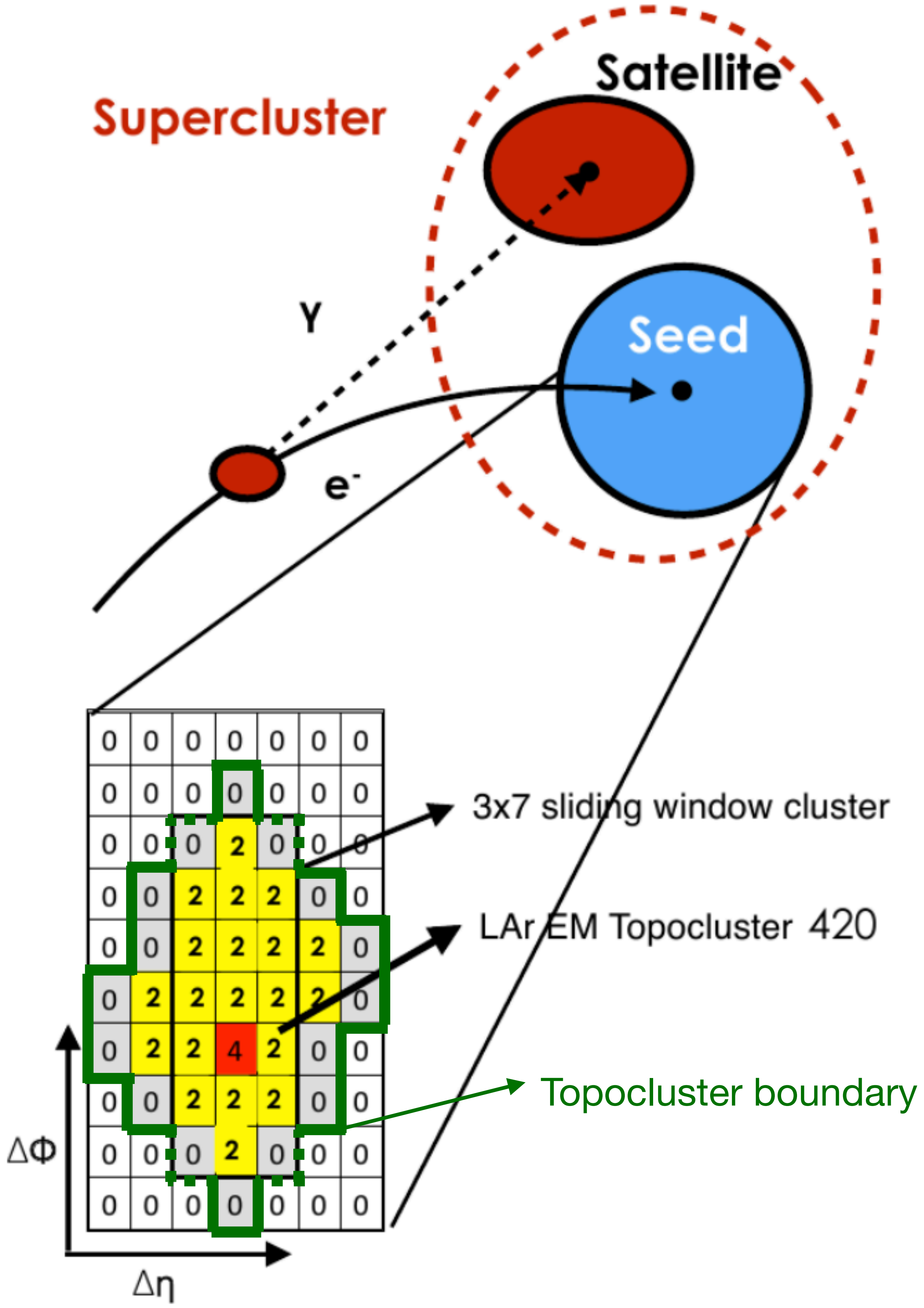
New Variables

Better algorithm to separate signal/background

ATL-COM-PHYS-2017-726

Northern Illinois University

Topological Clusters for Photon ID



Topological Clusters (topo-clusters)

- Photons in ATLAS are defined from objects known as "superclusters"
  - These are groupings of topological clusters, made through the logic shown
  - Studied various topo-cluster observables (moments) as inputs to photon ID algorithm to see if they can bring additional discriminating power

Samples Used

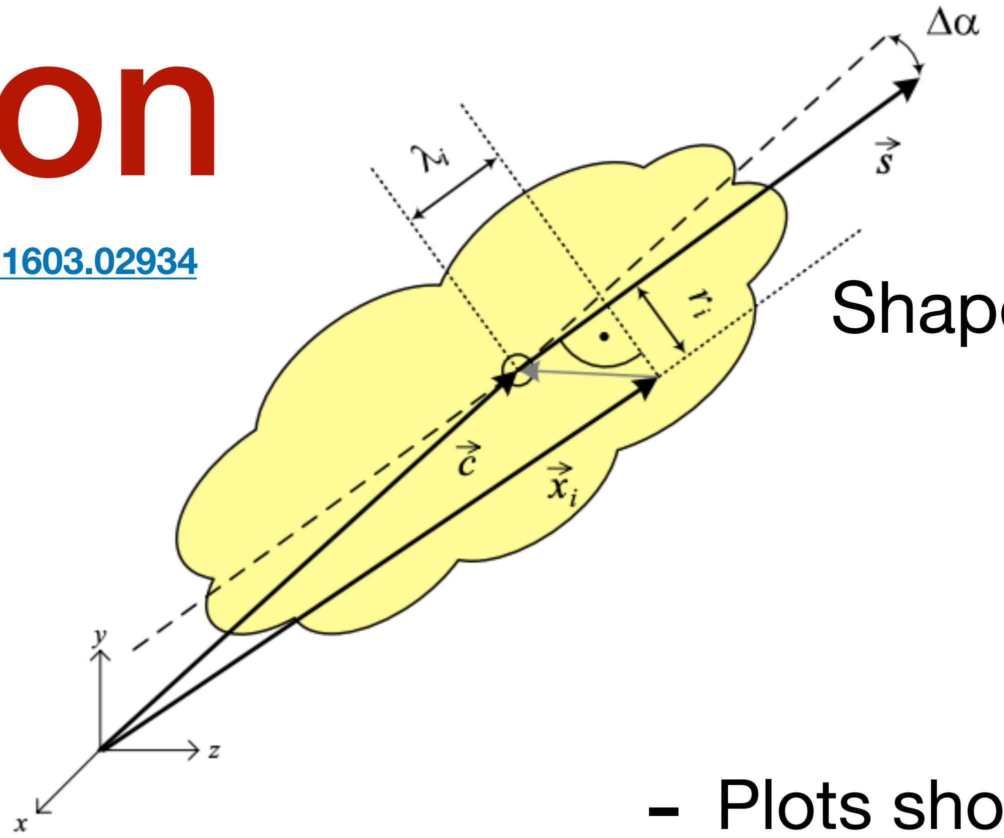
- γ+jets signal, jet-fakes background
- Monte Carlo truth information used to require signal sample be a prompt photon, vetoed for background sample

<u>Formed by:</u>
Find "seed" - cell significance > 4σ
Scan neighbors. If > 2σ, add cell and neighbors to cluster

Northern Illinois University

Topo-Cluster Shape Information

arxiv:1603.02934



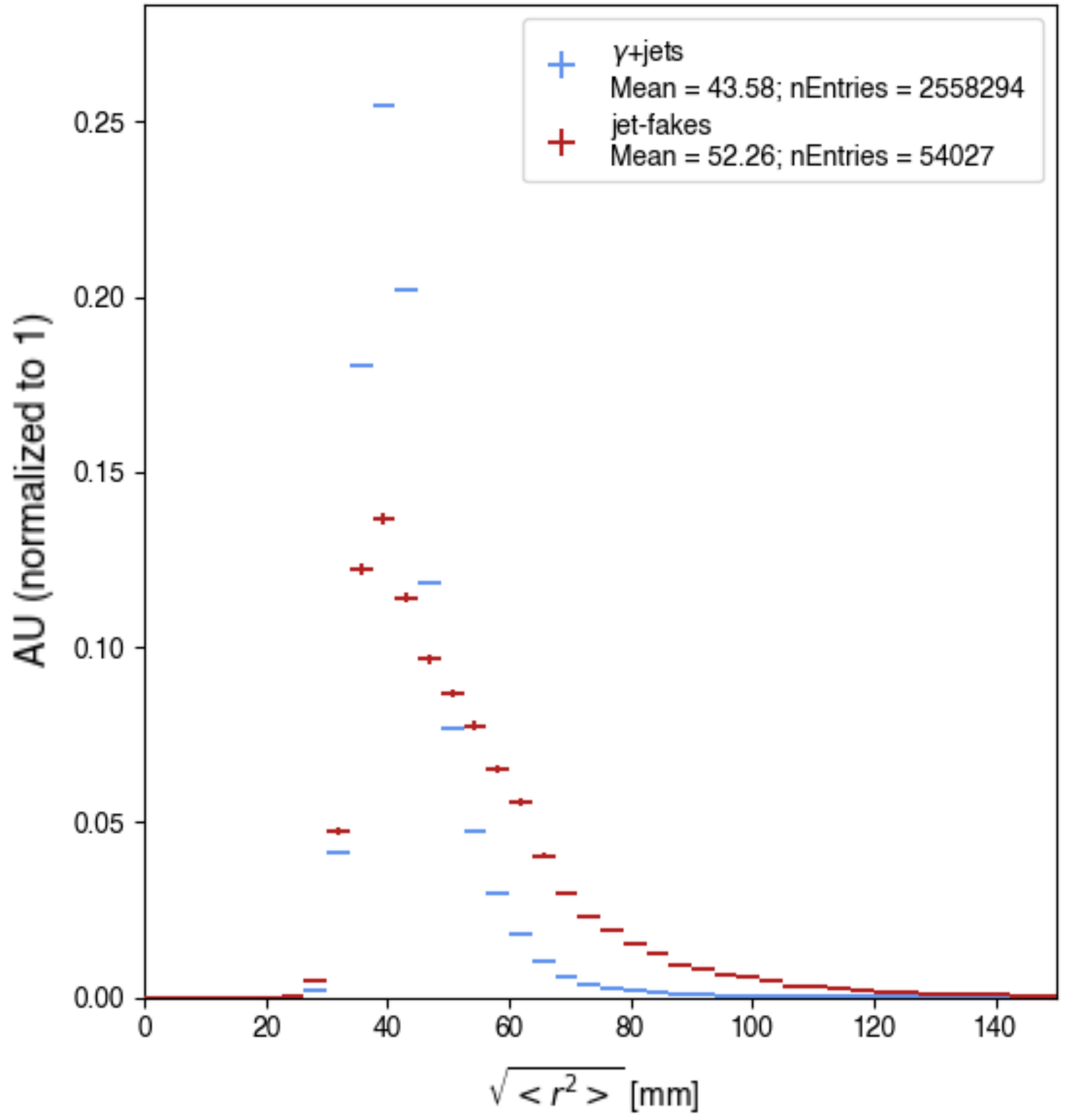
$$r_i = |(\vec{x} - \vec{c}) \times \vec{s}|$$
$$\lambda_i = (\vec{x_i} - \vec{c}) \cdot \vec{s}$$

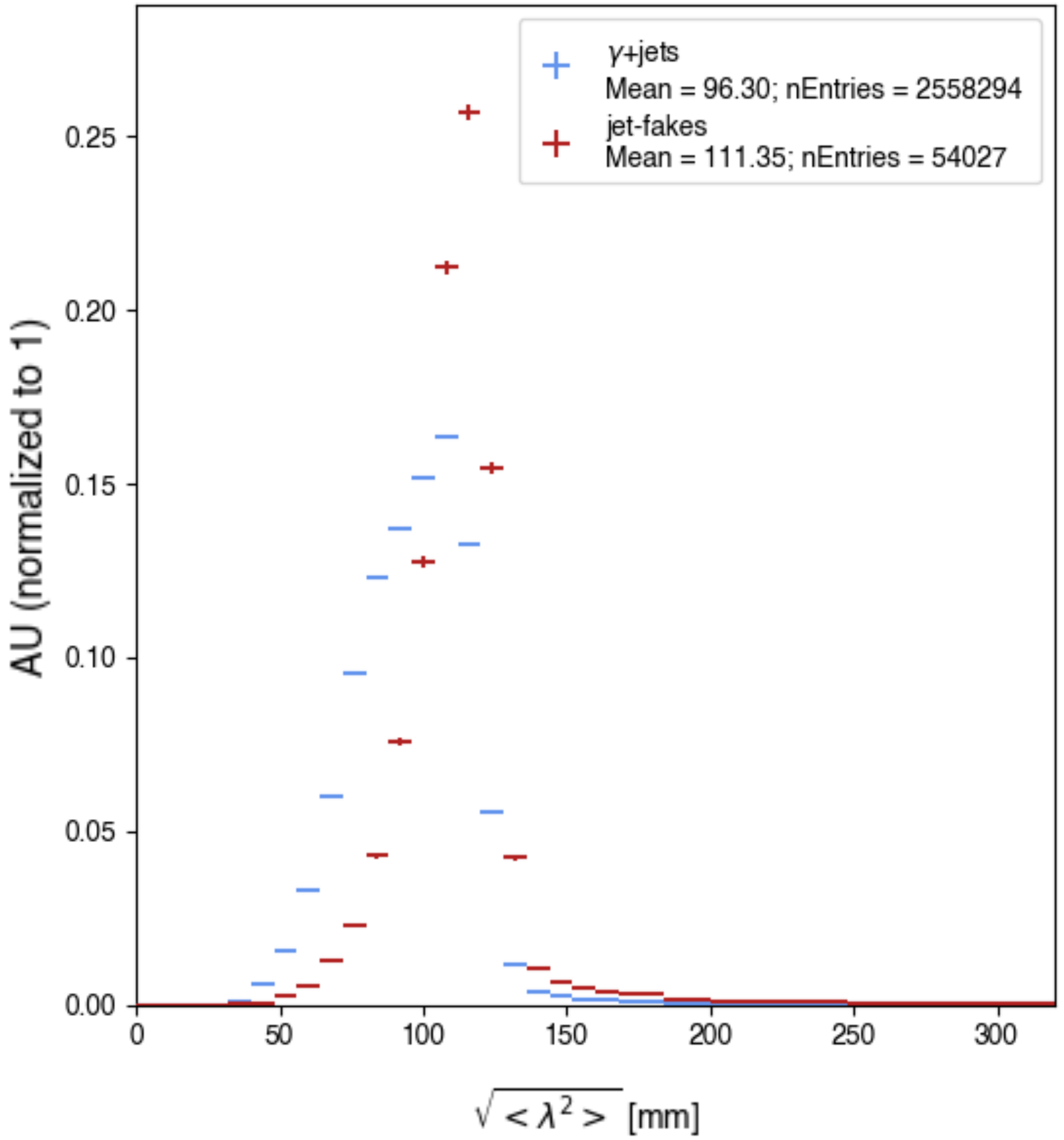
Shape can be defined using these quantities:

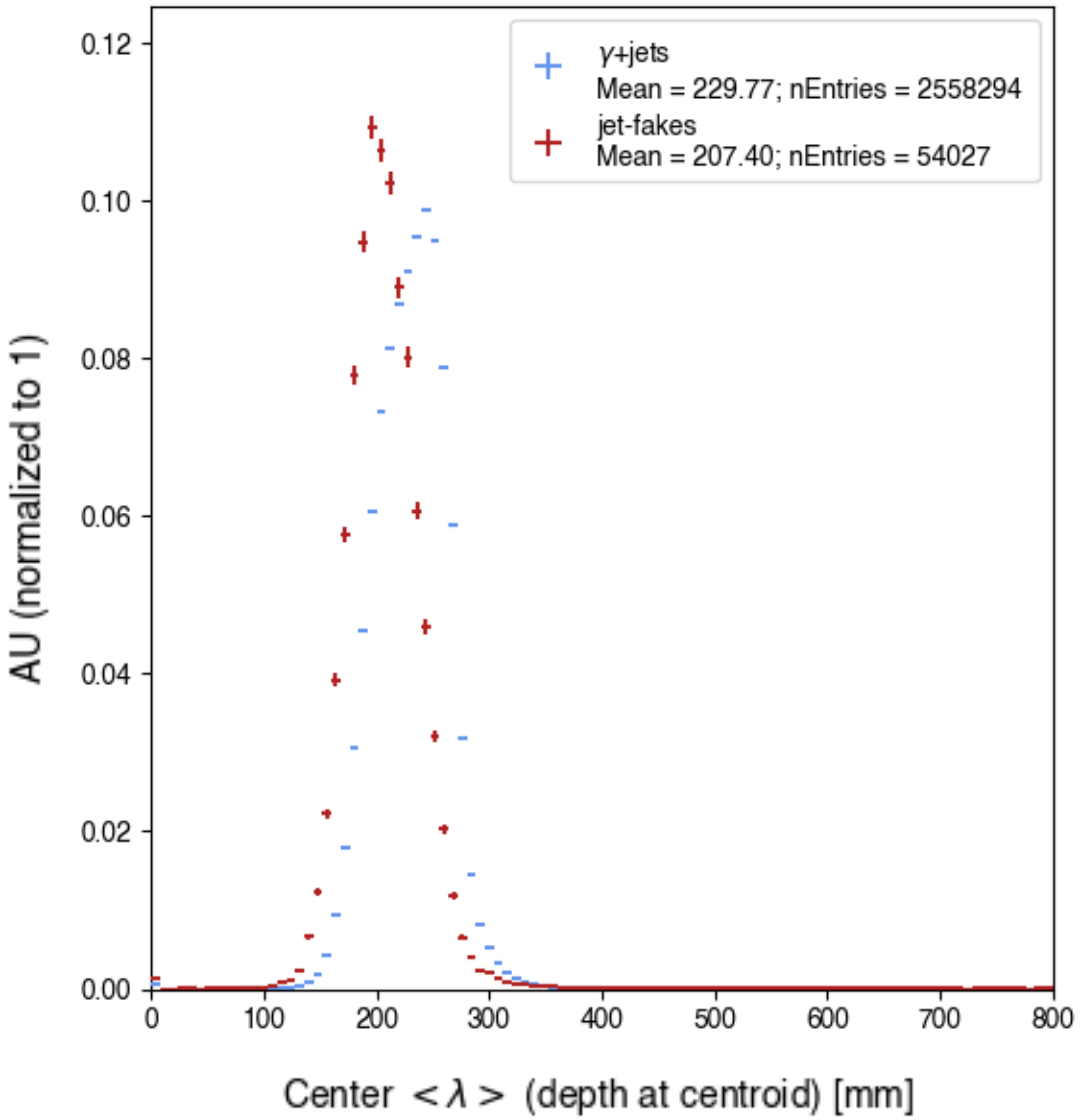
$\sqrt{\langle r^2 \rangle}$ = **Semi-major axis in width**

$\sqrt{\langle \lambda^2 \rangle}$ = **Semi-major axis in depth**

- Plots shown are in |η| < 0.6 region, pT inclusive





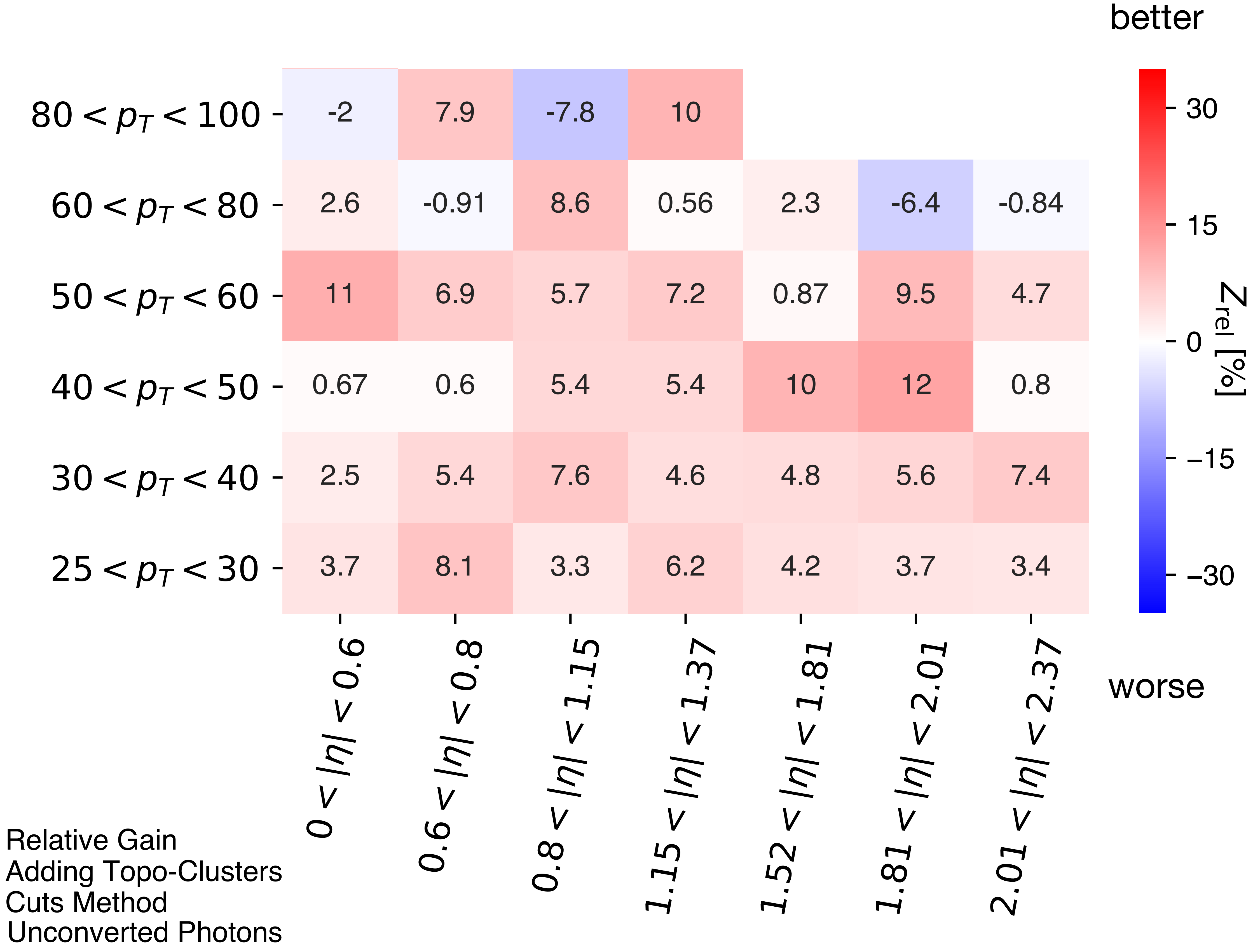


Differences exist in topo-cluster shape after current tight ID - additional information that can be used to further separate signal and background

*current Tight ID applied

Improvement by Adding Topo-clusters

Run existing cut-based optimization, adding the topo-cluster variables shown in prior slide



85% Signal Efficiency corresponds to current Tight ID

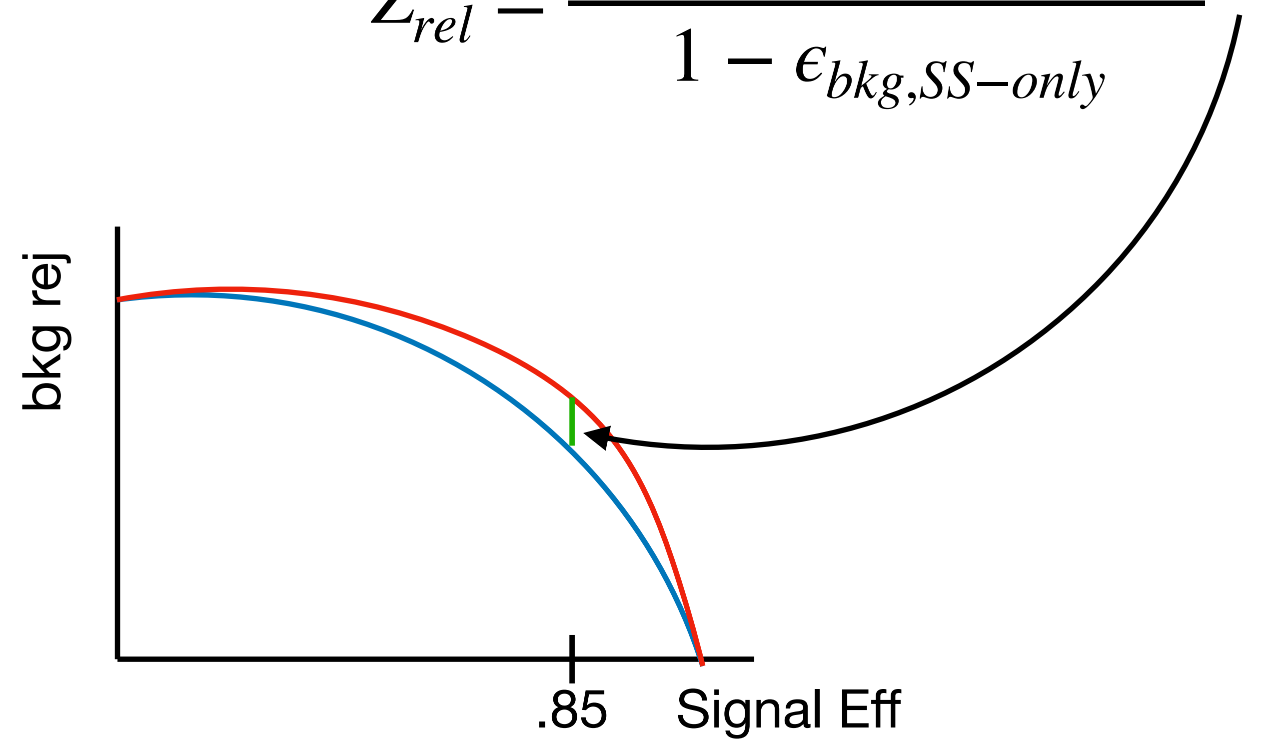
Figure of merit - improvement in background rejection at same signal efficiency as the current "tight" identification (85%)

Values shown:

$$Z_{rel} = \frac{\%\, Rej_{bkg,topo} - \%Rej_{bkg,SS-only}}{\%\, Rej_{bkg,SS-only}}$$

Equivalently,

$$Z_{rel} = \frac{\epsilon_{bkg,SS-only} - \epsilon_{bkg,topo}}{1 - \epsilon_{bkg,SS-only}}$$



Adding topo-cluster information can add **up to an additional ~10% background rejection** for the same signal efficiency using the current cut based method

Northern Illinois University

Machine Learning Approaches

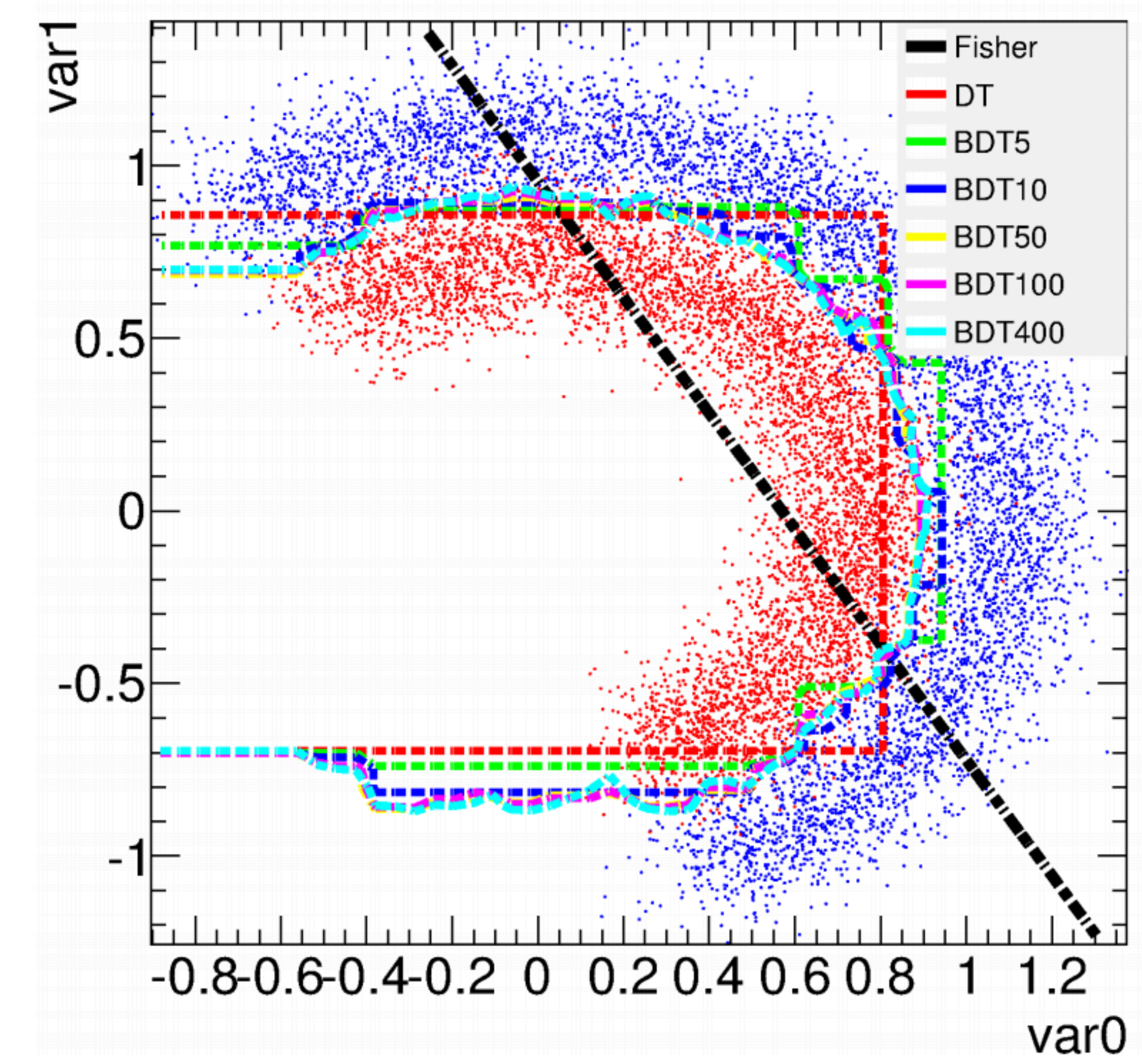
MVA Advantages	MVA Disadvantages
- More flexible solution space compared to cuts - Can exploit non-linearities	- Loss of clearly defined cuts, transparency

Current algorithm is a 9-dimensional cut optimization - could we consider a more sophisticated algorithm?

<u>Machine learning</u> - Two models considered:

Boosted Decision Tree
- Explained earlier! Ensemble method following tree logic
- Developed using TMVA, library for MVA in ROOT

Neural Network
- More flexible, can model more arbitrary functions
- Developed in Keras using Tensorflow backend



Of course, this is a toy example - cuts could do just as well with a variable transformation!

Northern Illinois University

Boosted Decision Tree

Evaluated improvement replacing cut-based method used presently with a BDT

While training, found statistics to be problematic in high p_T, high η bins

- Navigated by training *inclusively* in p_T. Variables uncorrelated to p_T, and provided better performance in all bins

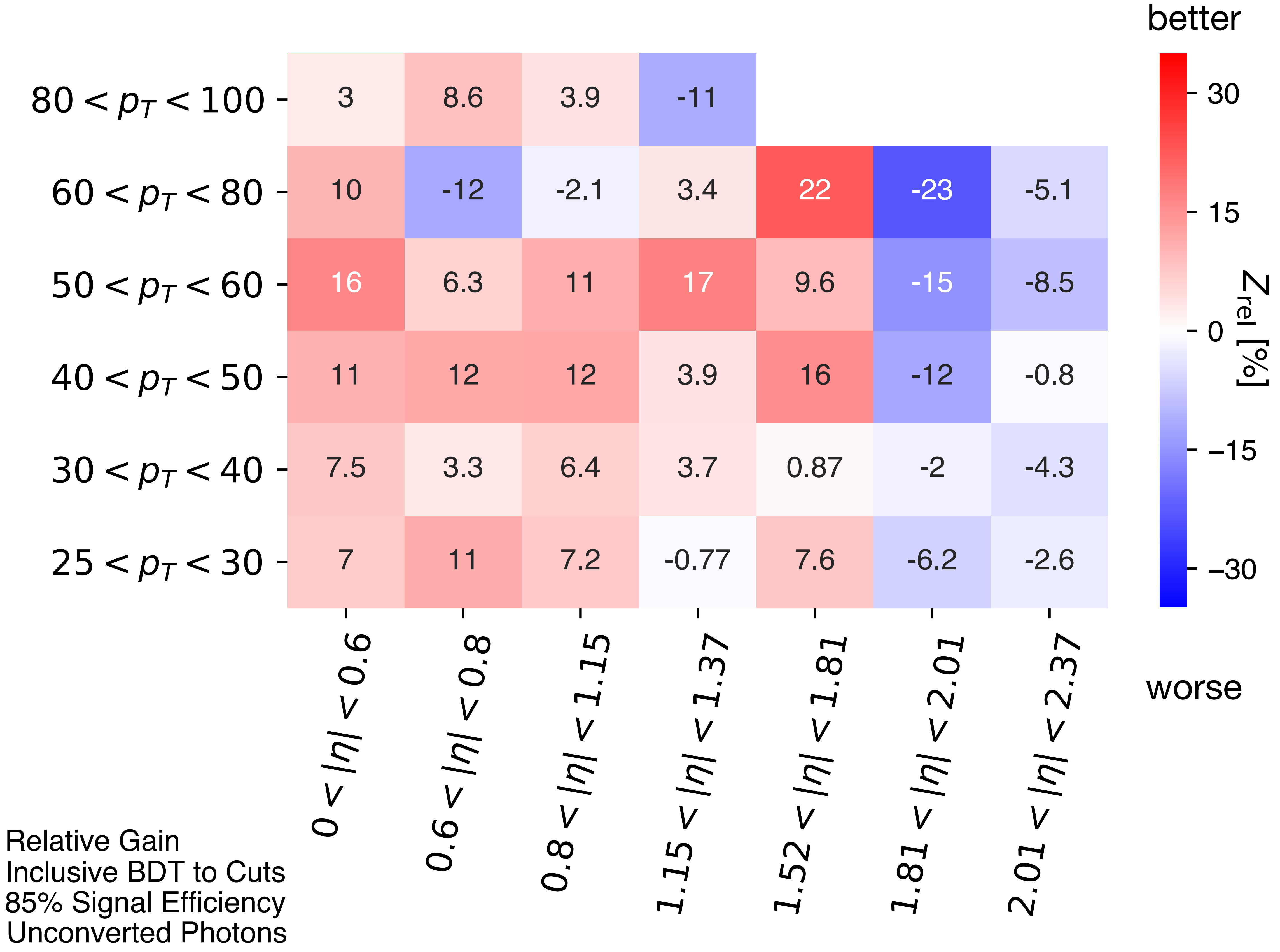
Same FOM used:

$$Z_{rel} = \frac{\%\, Rej_{BDT} - \%Rej_{cuts}}{\%\, Rej_{cuts}}$$

Only current shower-shape variables as input

Bins in which Z is worse with the BDT can be interpreted as the cut-method being better

- Limited statistics



Northern Illinois University

Machine Learning Interlude #2

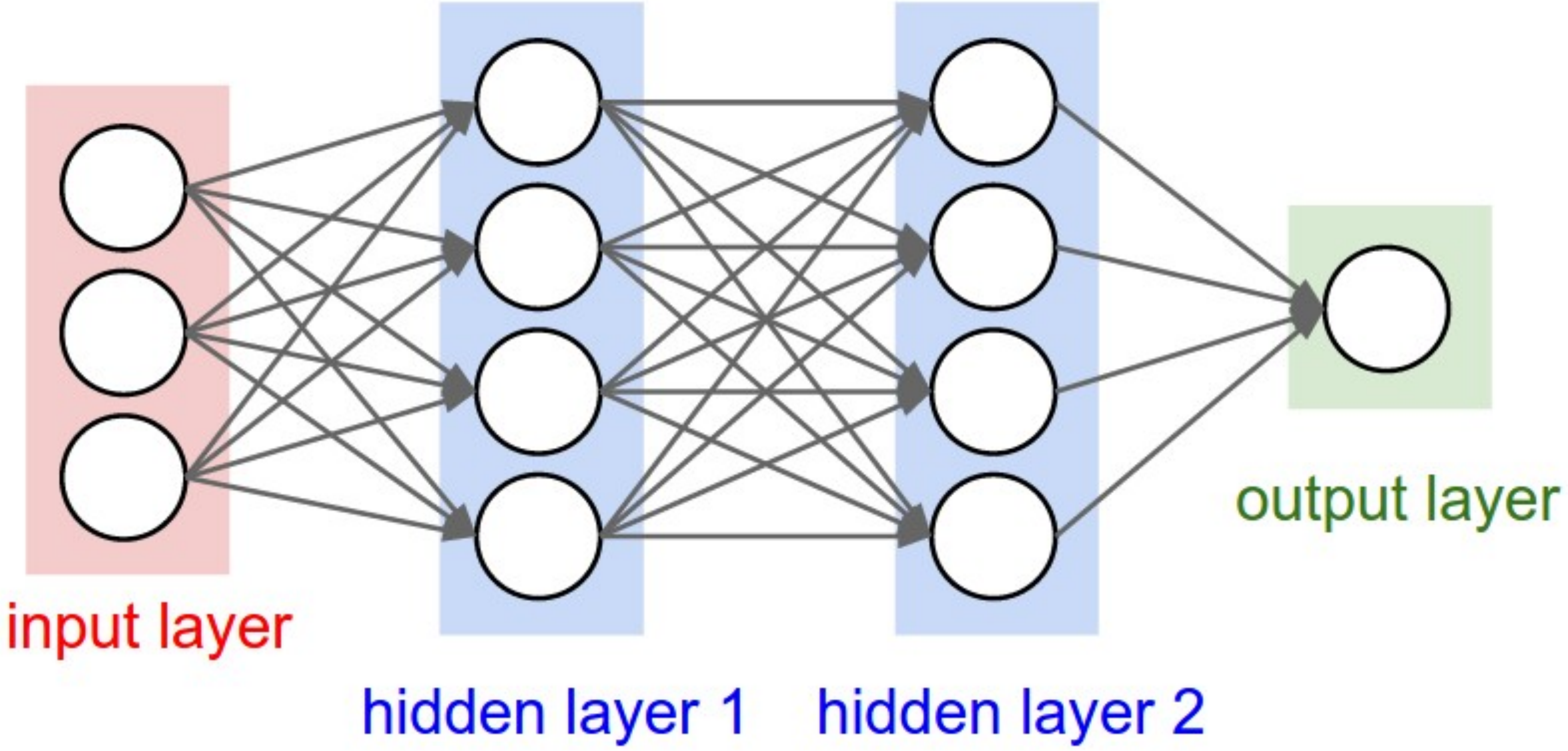
Neural Networks

Successive layers of nonlinear basis functions (*activation* functions), connected by weights

- Training stage learns the values of weights and the parameters
- In standard NNs (used in this work), data proceeds strictly from input to output (*feed forward*)
- Name comes from neuroscience analogy, where activation functions are analogous to neurons

- Over a sufficient number of activation functions, any possible function can be approximated

Andrej Karpathy



Nodes represent activation functions
Connections represent weights

Notes

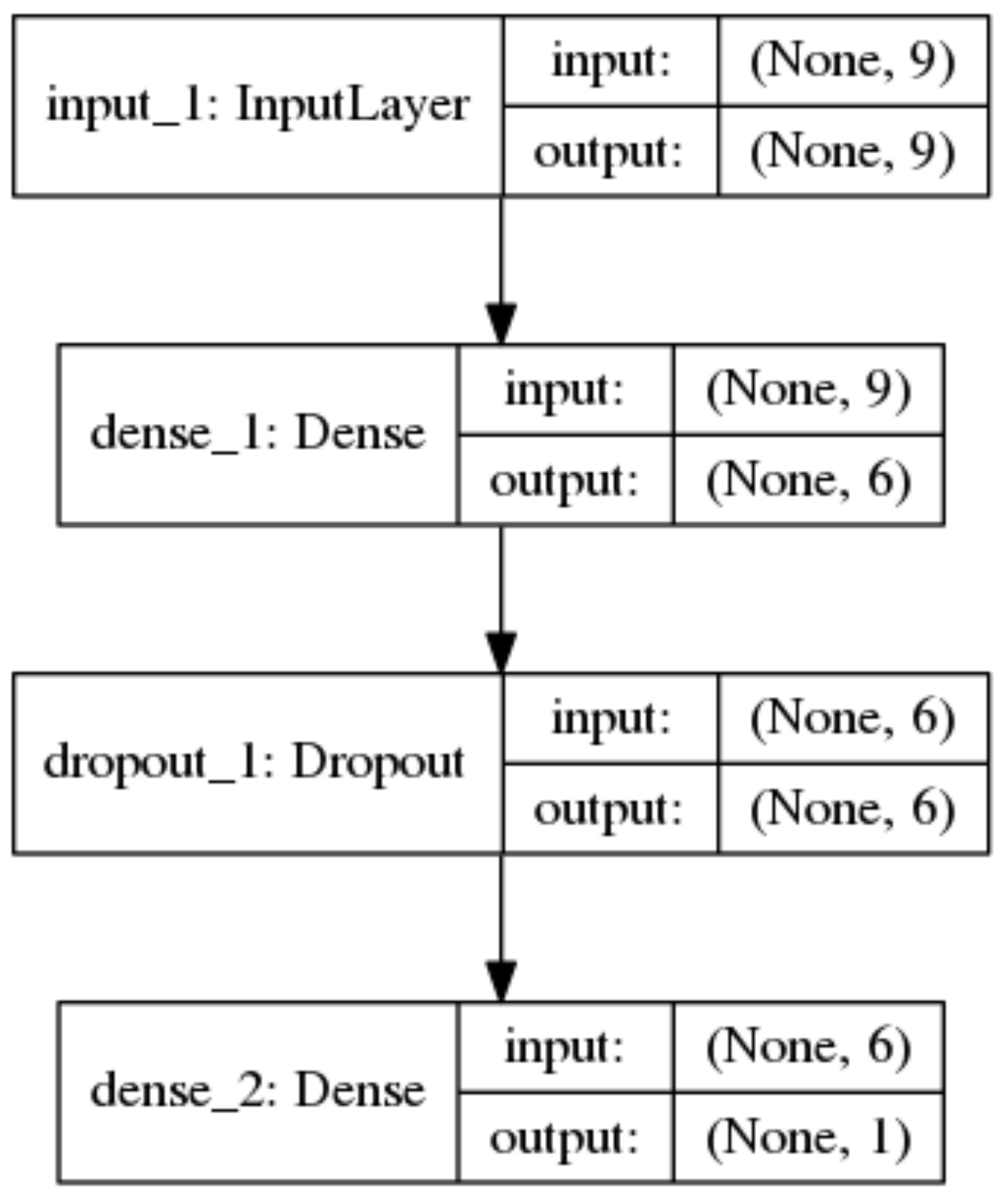
- Incredibly flexible algorithm!
- Requires *significant* training data

Northern Illinois University

NN Training Notes

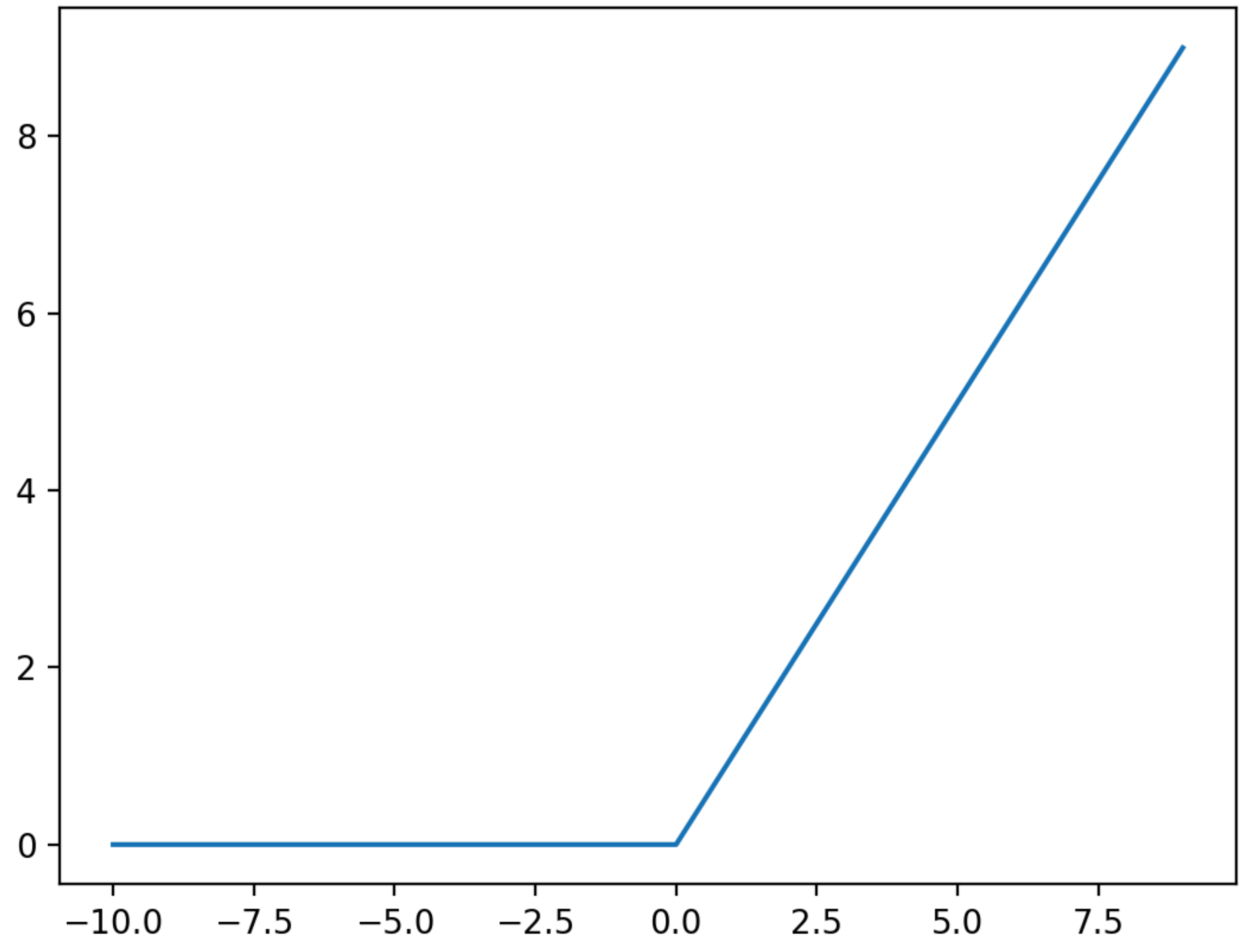
<u>Topology and notes</u>

- Constructed in Keras, topology shown, selected by a course optimization by hand
  - Very simple topology, few parameters compared to modern networks
  - Deeper/wider networks tested but found no additional improvement
- ReLU activation used for internal nodes, sigmoid on output
- Binary Cross-Entropy loss
- Epochs dictated by early stopping - stops after no improvement for 20 epochs

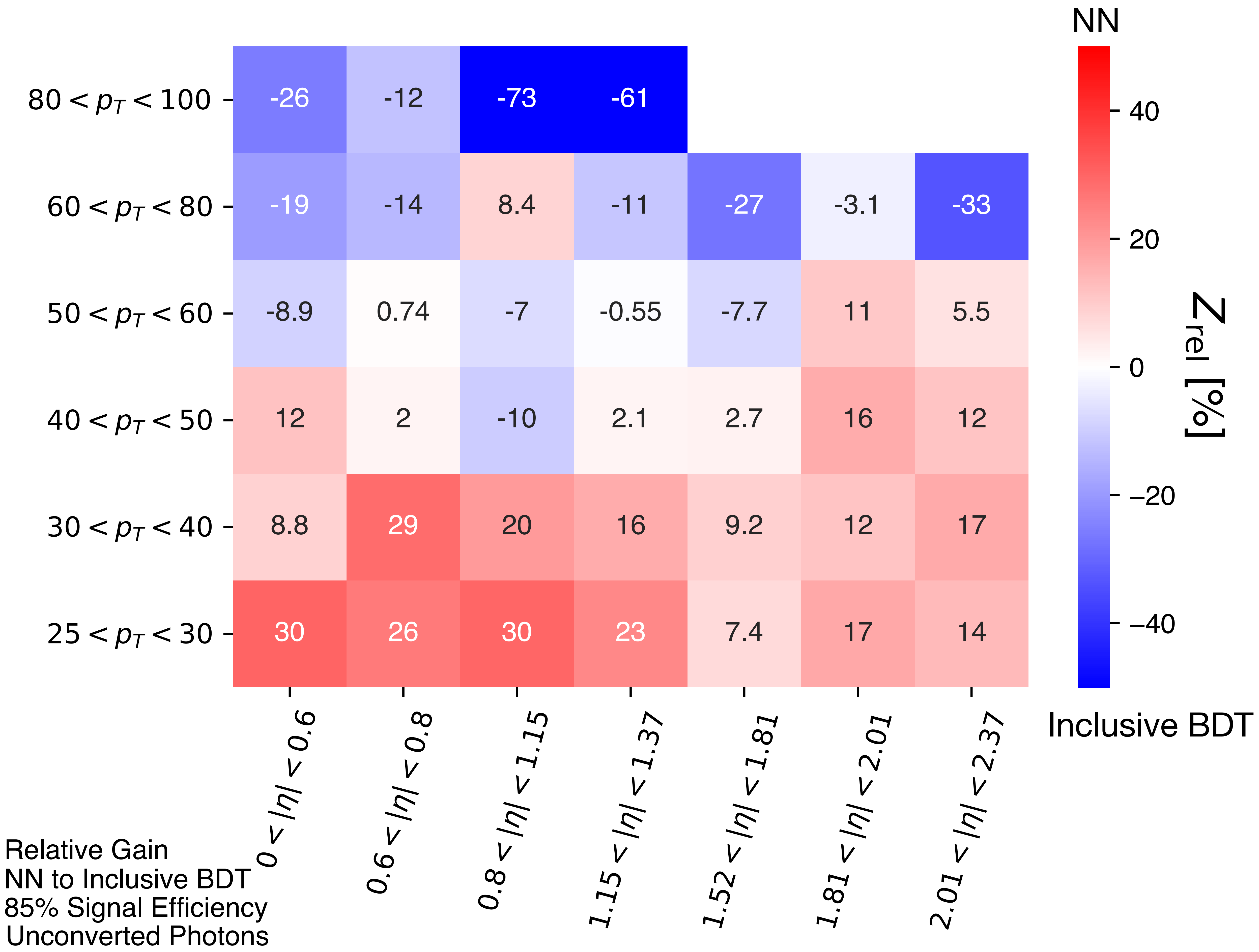


67 trainable parameters

ReLU Activation



Neural Network Results



$$\frac{Rej_{bkg,NN} - Rej_{bkg,BDT}}{Rej_{bkg,BDT}}$$

Again, can think of negative values as BDT performing better, positive as NN performing better

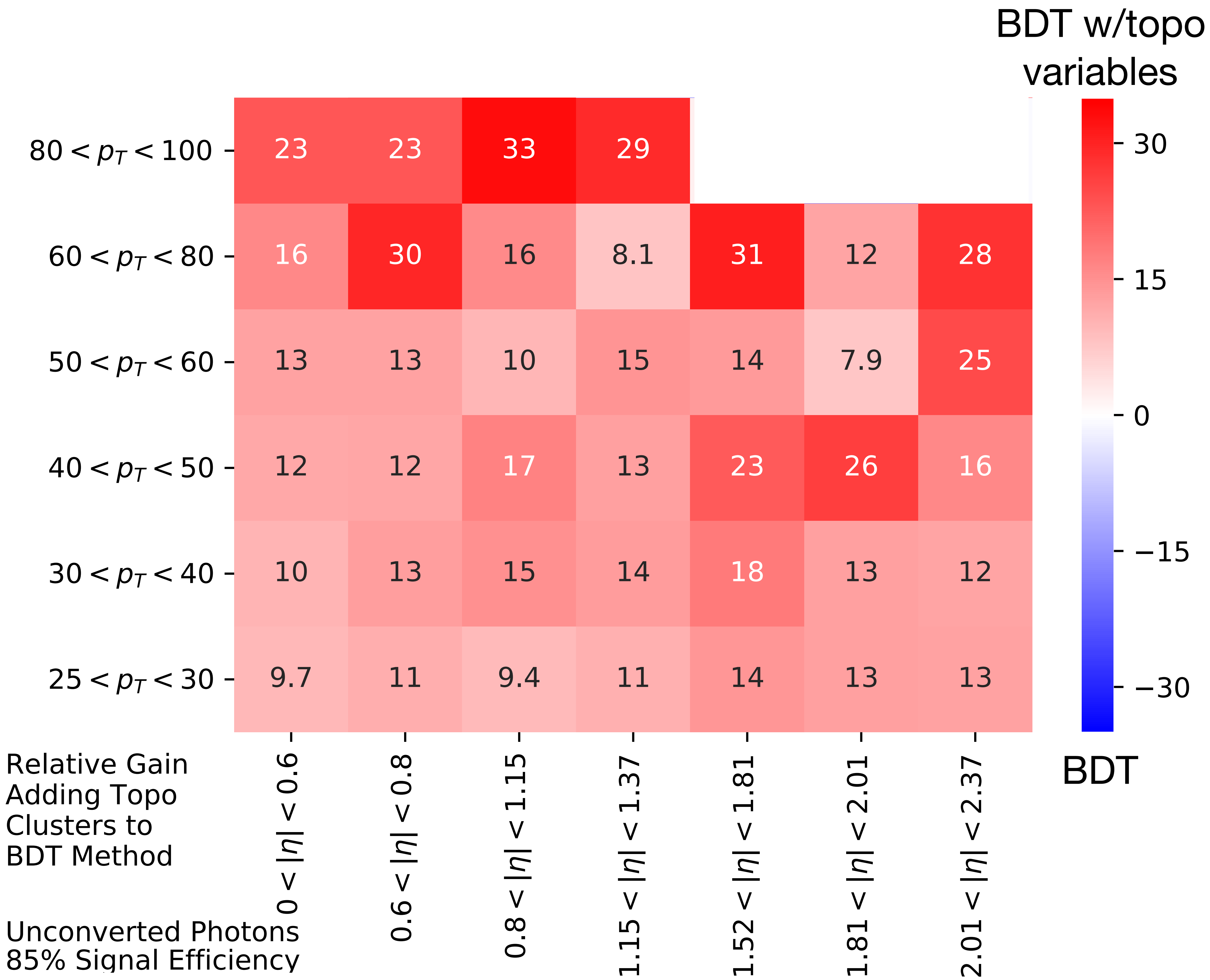
NN trained binwise does better than BDT in the lowest pT regions

Inclusively trained NN was evaluated, but significant gains only seen at pT > 80 GeV, and much worse performance seen below

Northern Illinois University

Putting Improvements Together

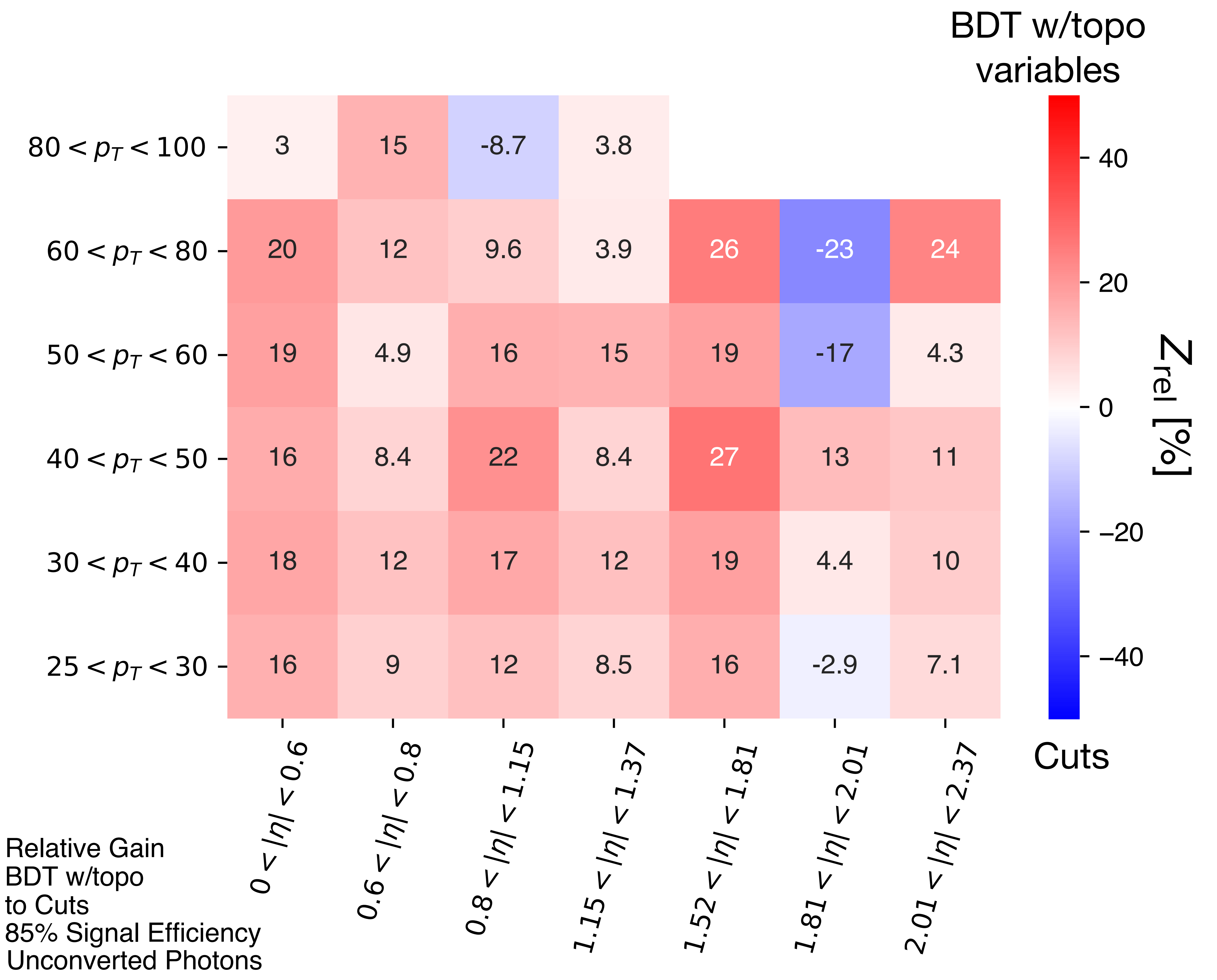
Want to also consider how the algorithms can improve by including the aforementioned topo-cluster variables
- Shown for test case using a BDT



As much as a 30% improvement in the BDT by including the topo-cluster variables

Putting Improvements Together

Last, as a final metric of improvement, compare the BDT including the topo-cluster variables to the present method

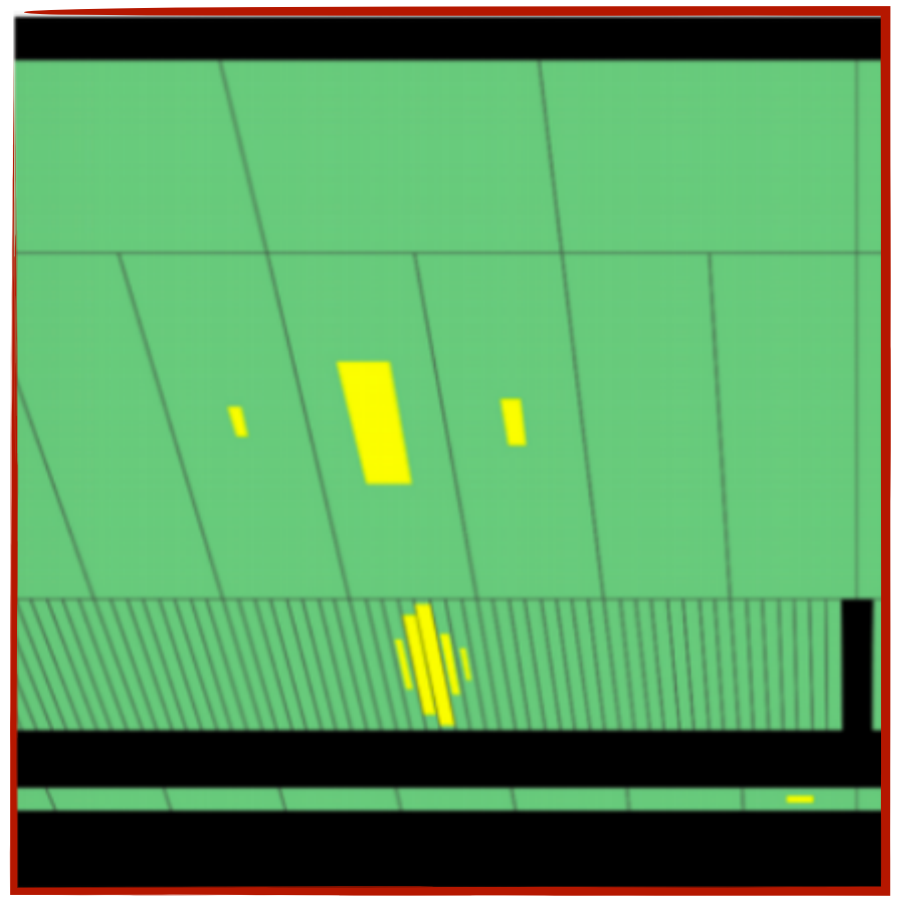


As much as a 25% gain through all improvements compared to current methods

Most bins show an improvement of > 10%

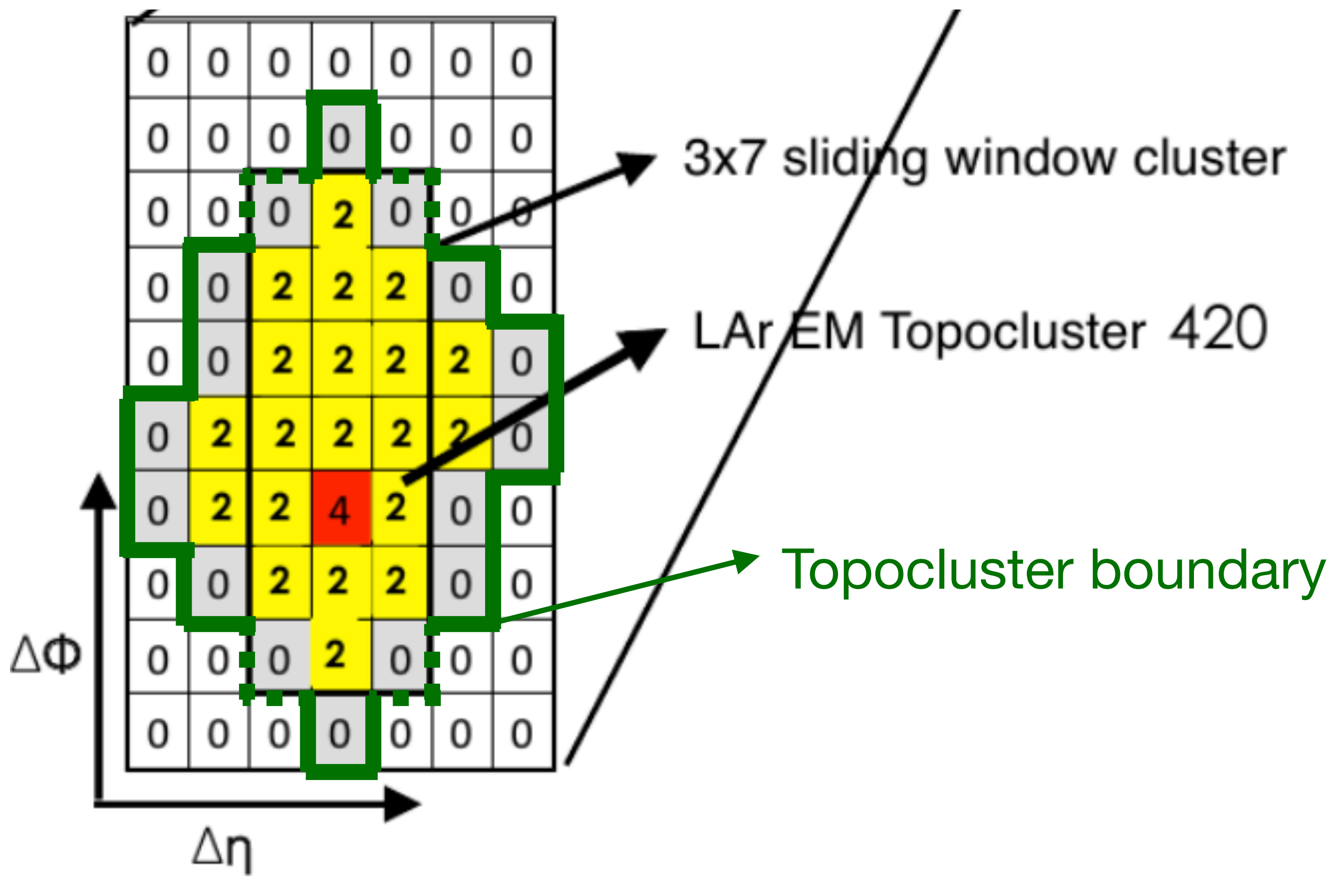
Northern Illinois University

Photon Identification Conclusion



The HH→$\gamma\gamma$bb analysis is sensitive to the square of the photon identification efficiency, motivates searching for improvements
Methods to optimize photon identification studied, using improvement in background at current tight WP as a FOM

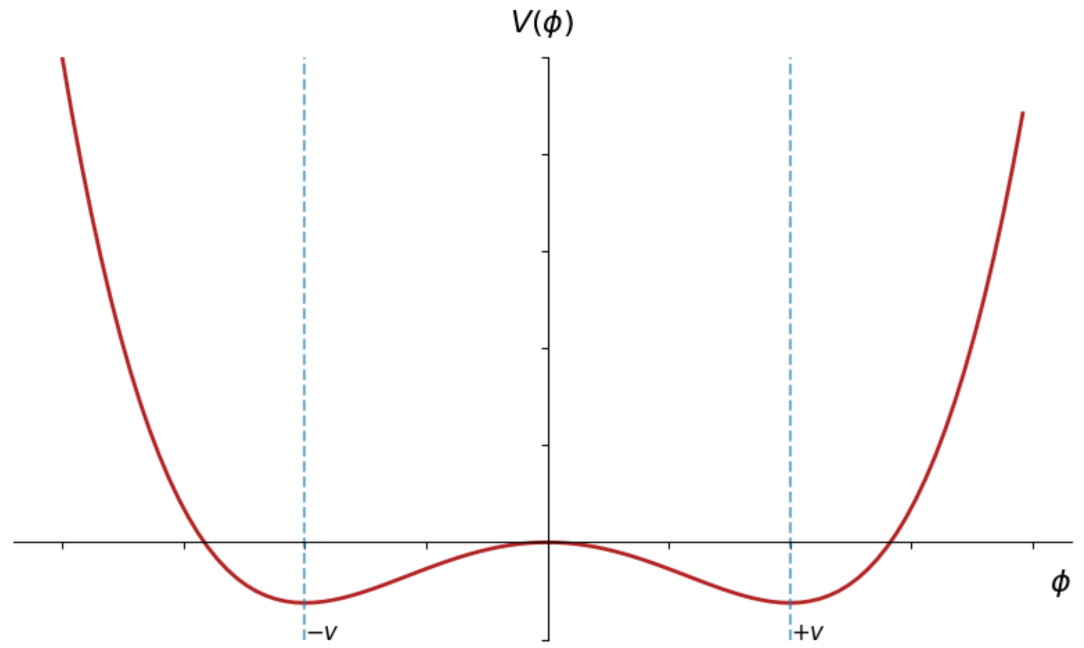
1. Including new variables as inputs based on the topo-cluster variables, notably the shape observables
2. Switching photon identification from a univariate cuts method to a multivariate algorithm, such as a BDT or NN



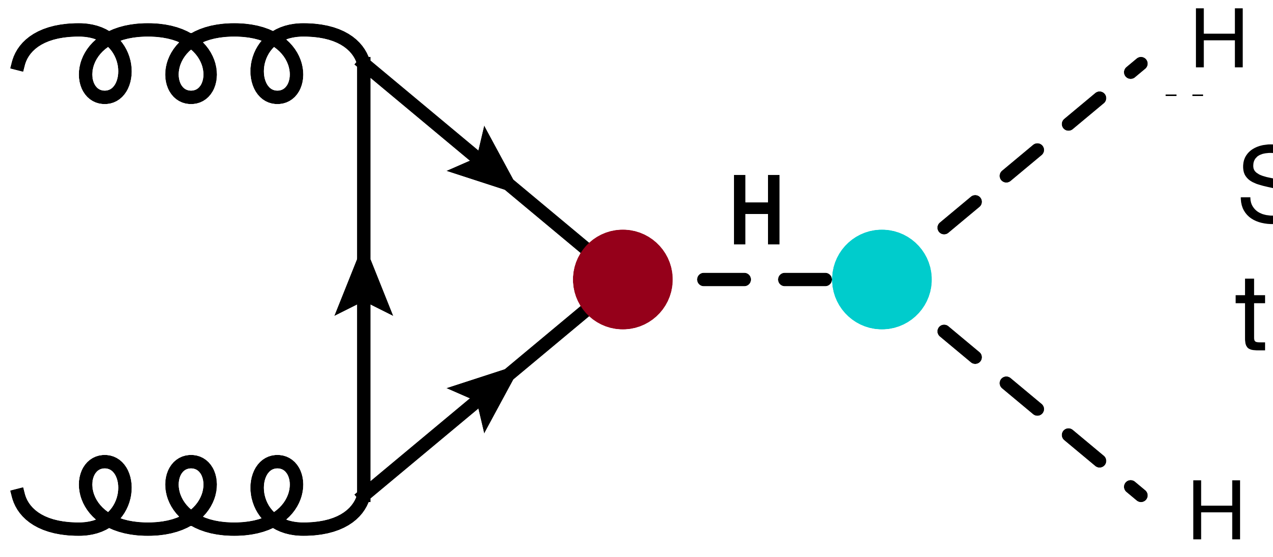
Through implementing these methods, an improvement of as much as 25% background rejection for the same signal efficiency can be found, greater than 10% improvement in most η-p_T bins

Northern Illinois University

Conclusion

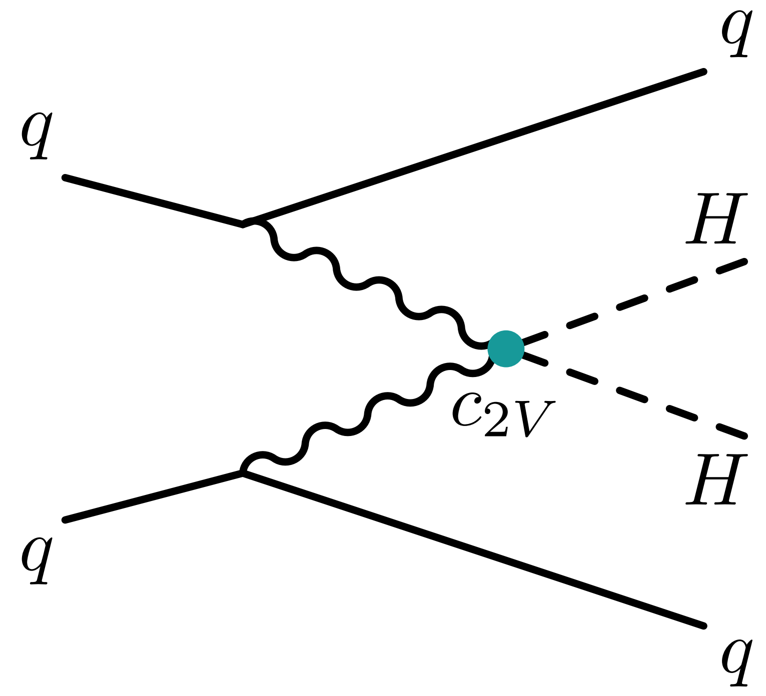


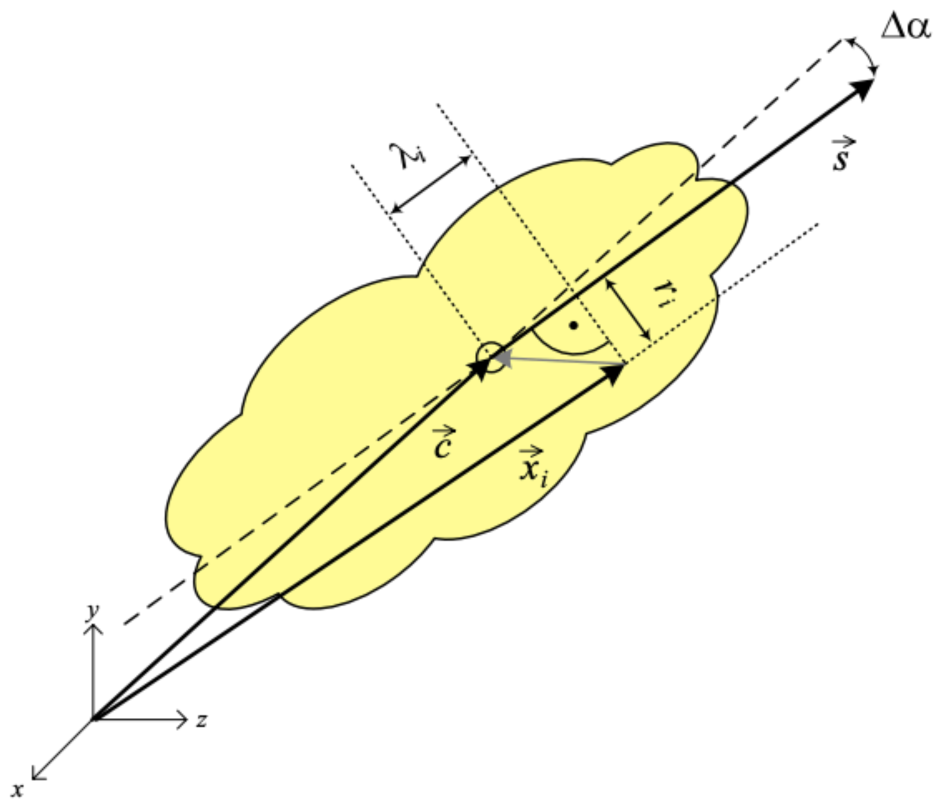
The discovery of the Higgs boson has been a marker of success for the LHC physics program, and further understanding Electroweak Symmetry breaking is a leading priority of the future



Studying di-Higgs production gives a handle on the Higgs self-coupling and the structure of the Higgs potential, ATLAS searches have provided the best current limits on the self-coupling

HH→$\gamma\gamma$bb is a key channel contributing to this limit, and is working toward results using the Full Run 2 data. Showed a projected 9.8% improvement by including a VBF-enriched category defined via a multiclass BDT



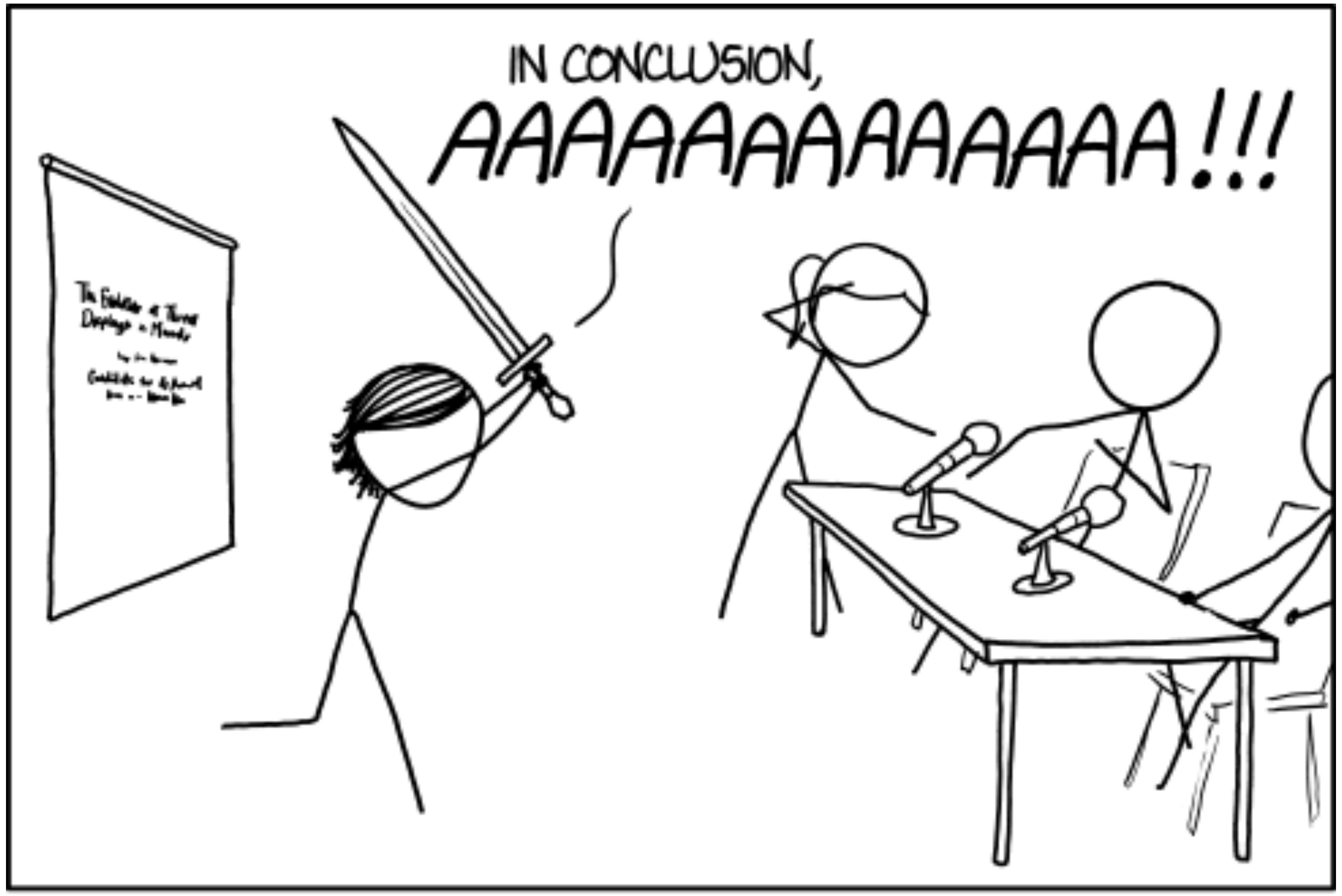


Improvement to future photon identification in ATLAS presented, through inclusion of new discriminating variables, as well as through implementing machine learning methods

Northern Illinois University

Questions?

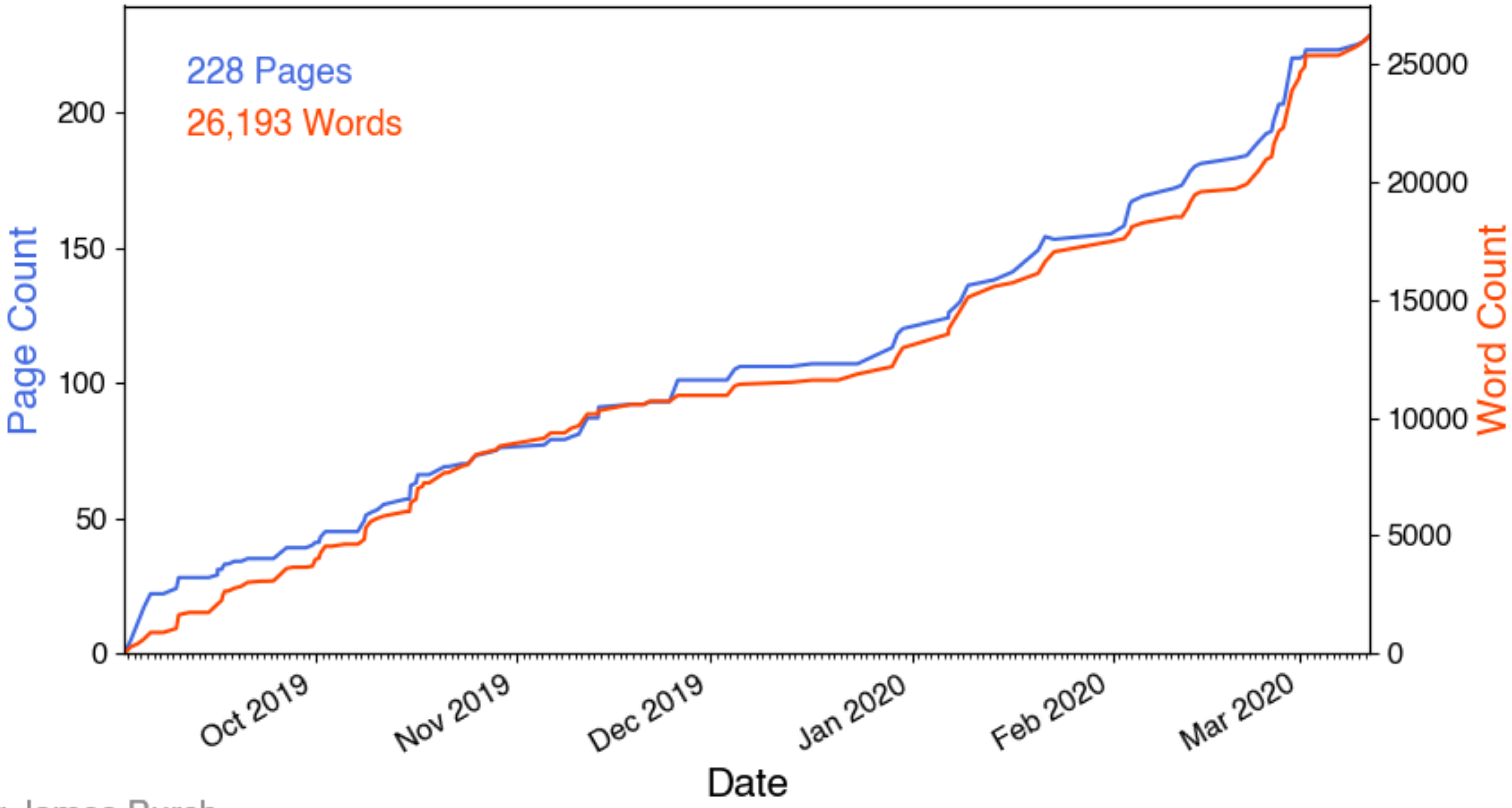
xkcd



THE BEST THESIS DEFENSE IS A GOOD THESIS OFFENSE.

Northern Illinois University

Backup



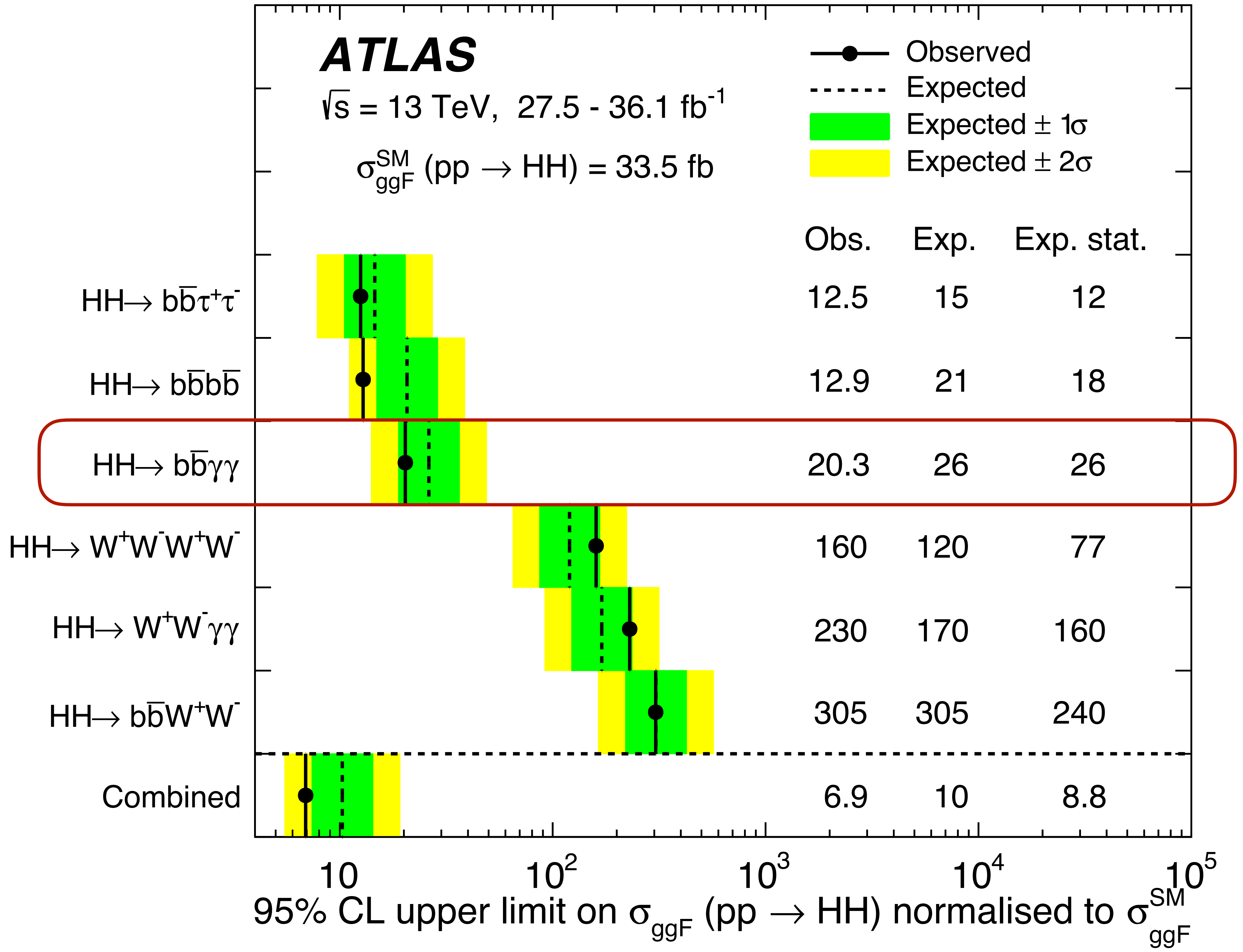
Tyler James Burch

Northern Illinois University

Combined Limit

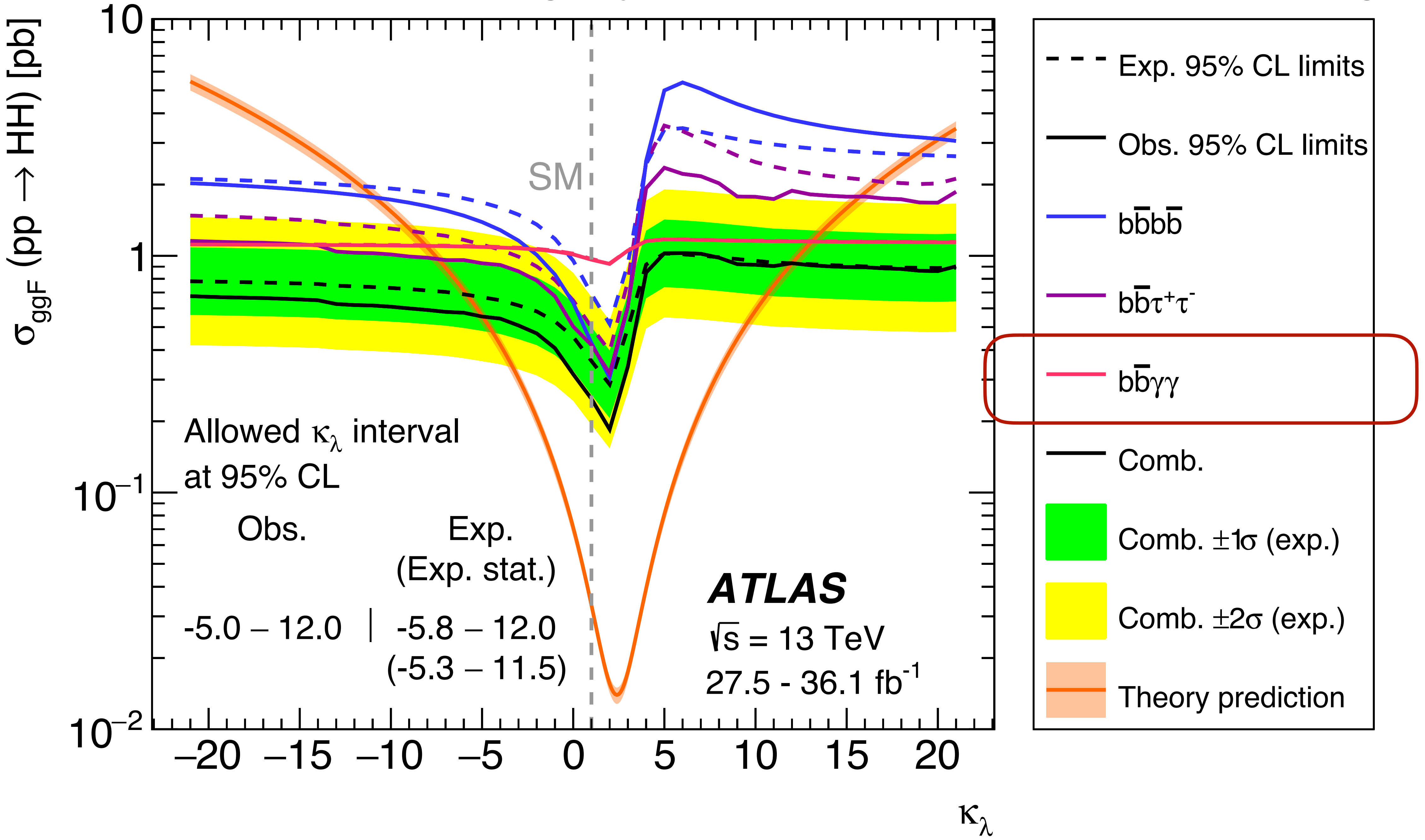
$\gamma\gamma$bb combined with other leading channels to set best ever di-Higgs limits

[1906.02025]



Upper limit = 6.9*SM expectation

Using asymptotic approximations for limit-setting



Trilinear Coupling

Expected: $-5.8 < K_\lambda < 12.0$

Observed: $-5.0 < K_\lambda < 12.0$

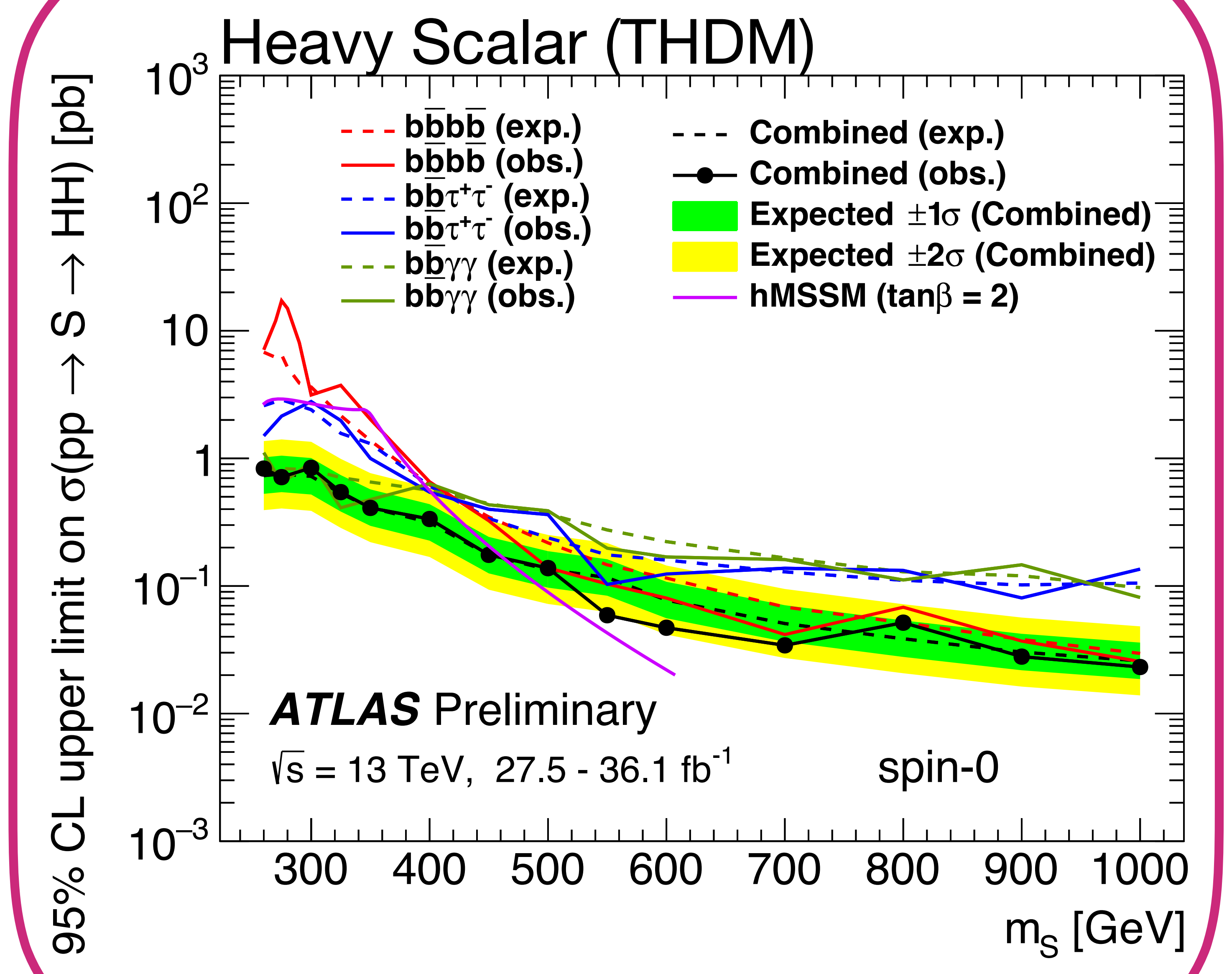
Northern Illinois University

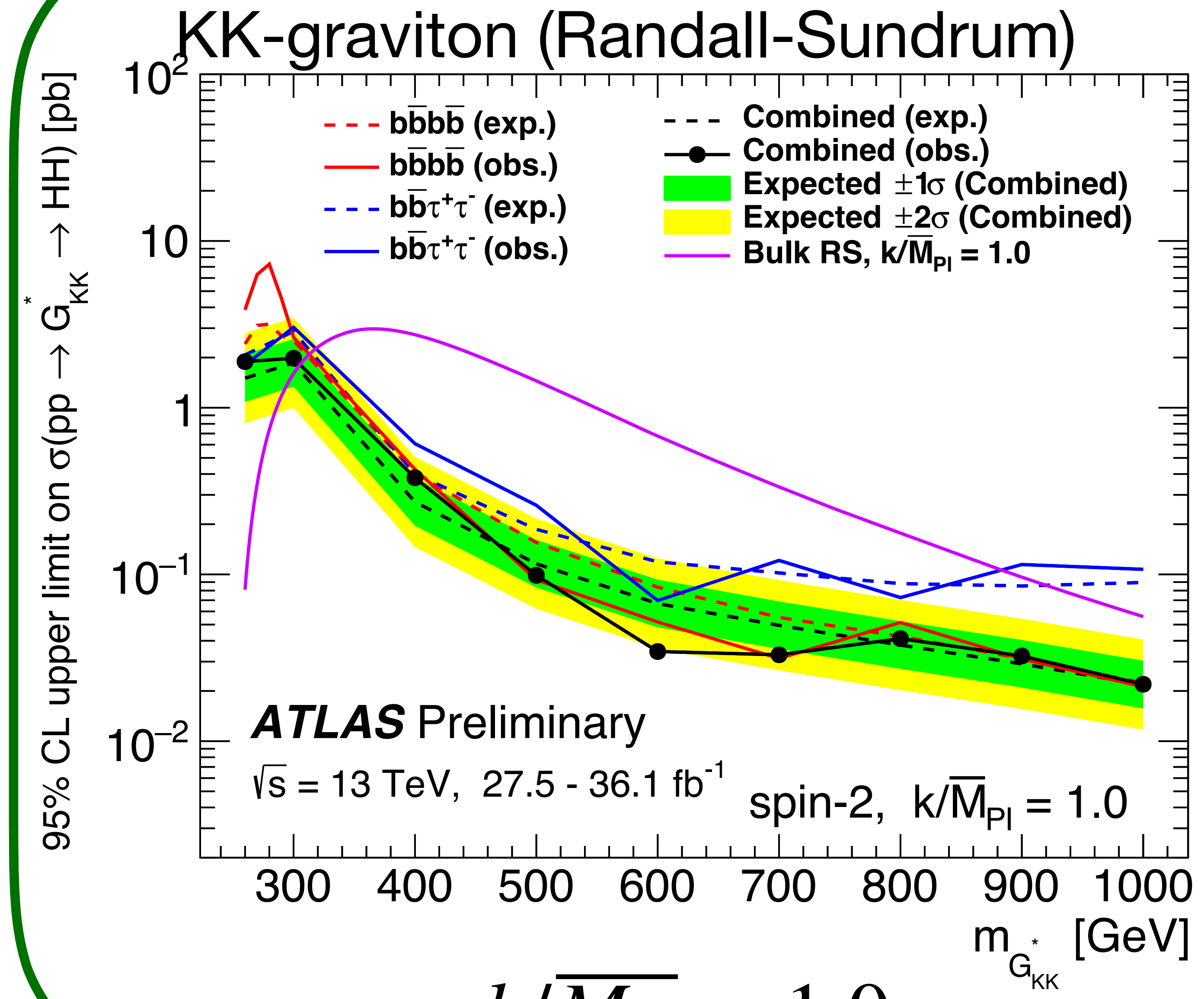
Resonant HH Combination Limits

Two-Higgs Doublet Model
mX < 462 GeV @ 95% CL in hMSSM

Randall-Sundrum Model
k/M_{Pl} =1.0 constraints: 307 < m_G < 1362 GeV
k/M_{Pl} =2.0 constraints: m_G < 1744 GeV

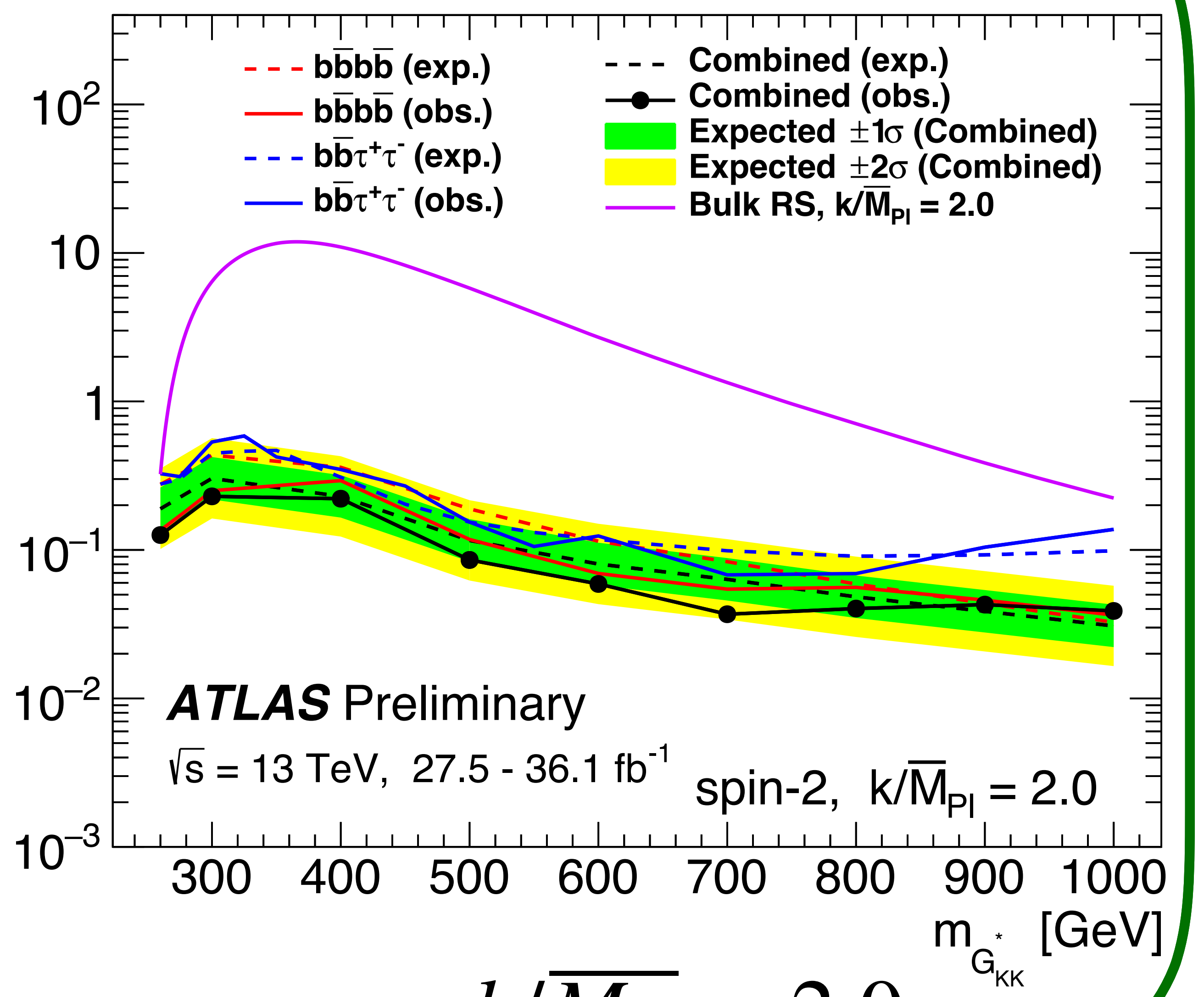
tanβ = 2: ratio of the vacuum expectation values of the two Higgs doublets





$k/\overline{M}_{Pl} = 1.0$

ATLAS-CONF-2018-043



$k/\overline{M}_{Pl} = 2.0$

k: curvature of the warped extra dimension
M⁻ Pl: the effective four-dimensional Planck scale

Northern Illinois University

VBF Current Constraints

Current constraints set by HH->bbbb publication

$-1.02 < c_{2V} < 2.71$

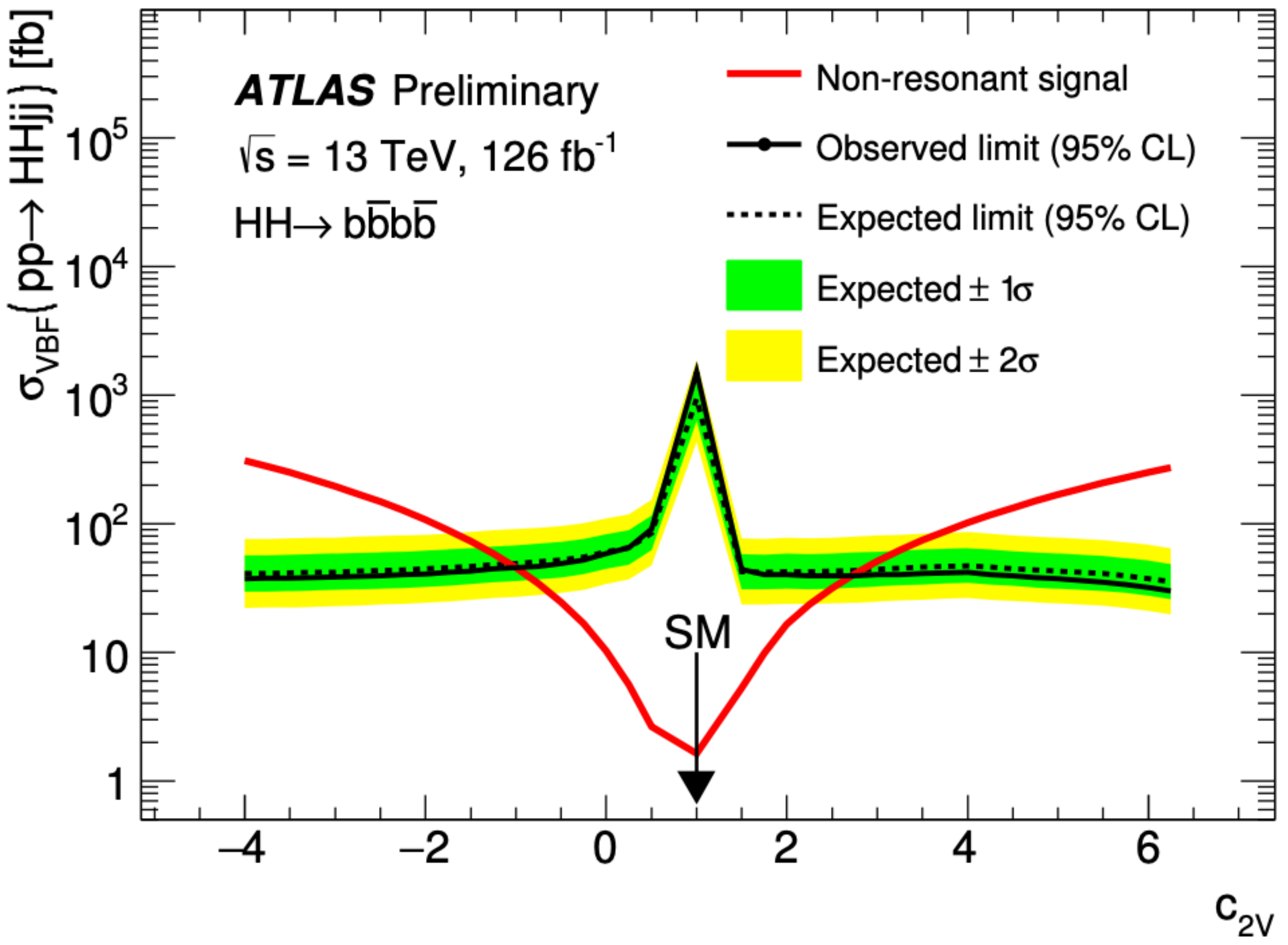


Figure 6: Observed and expected 95% CL upper limits on the production cross-section for non-resonant HH production via VBF as a function of the di-vector-boson di-Higgs-boson coupling c_{2V}. The red line corresponds to variations of the leading-order cross-section with c_{2V}.

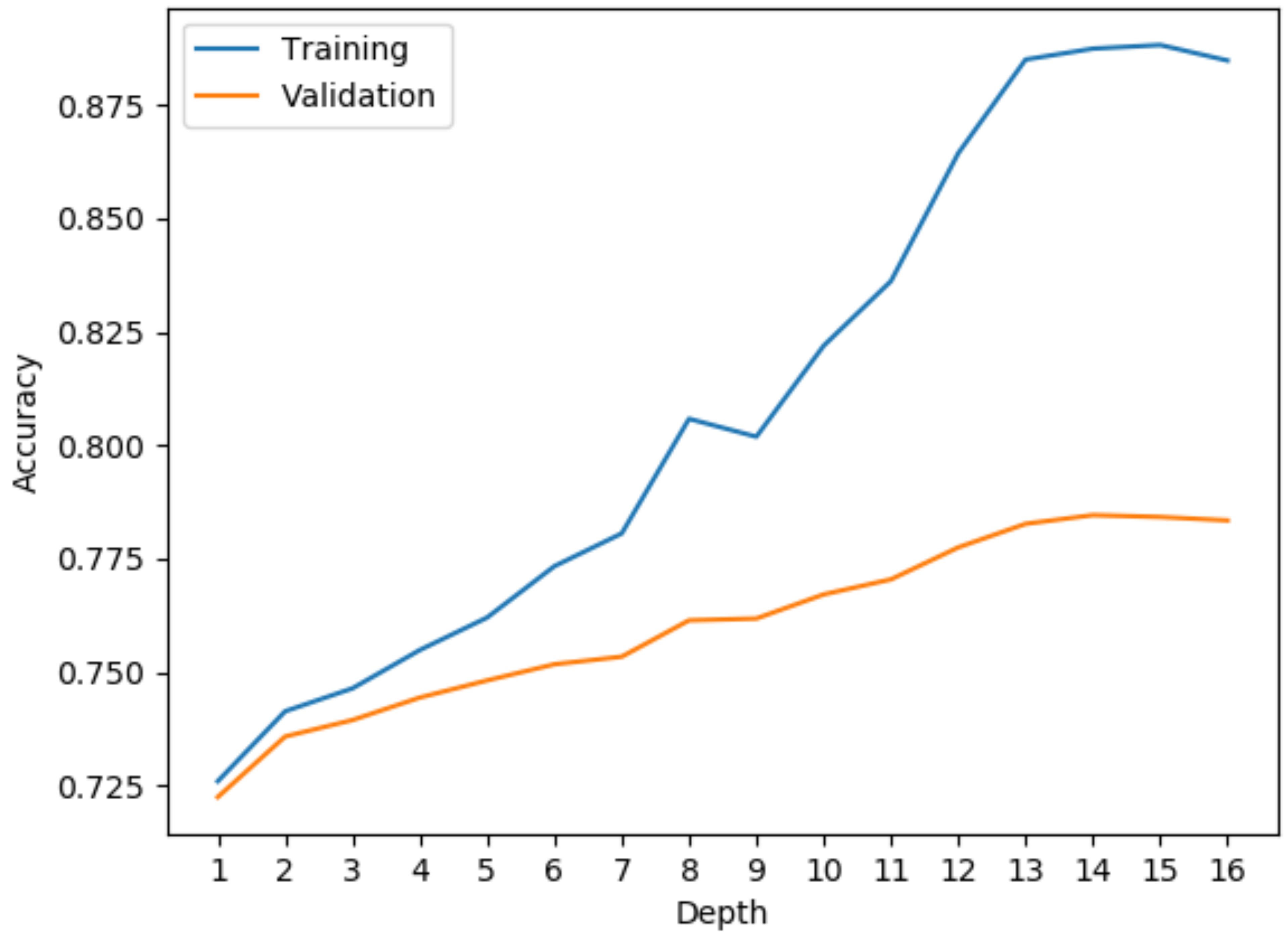
ATLAS-CONF-2019-030

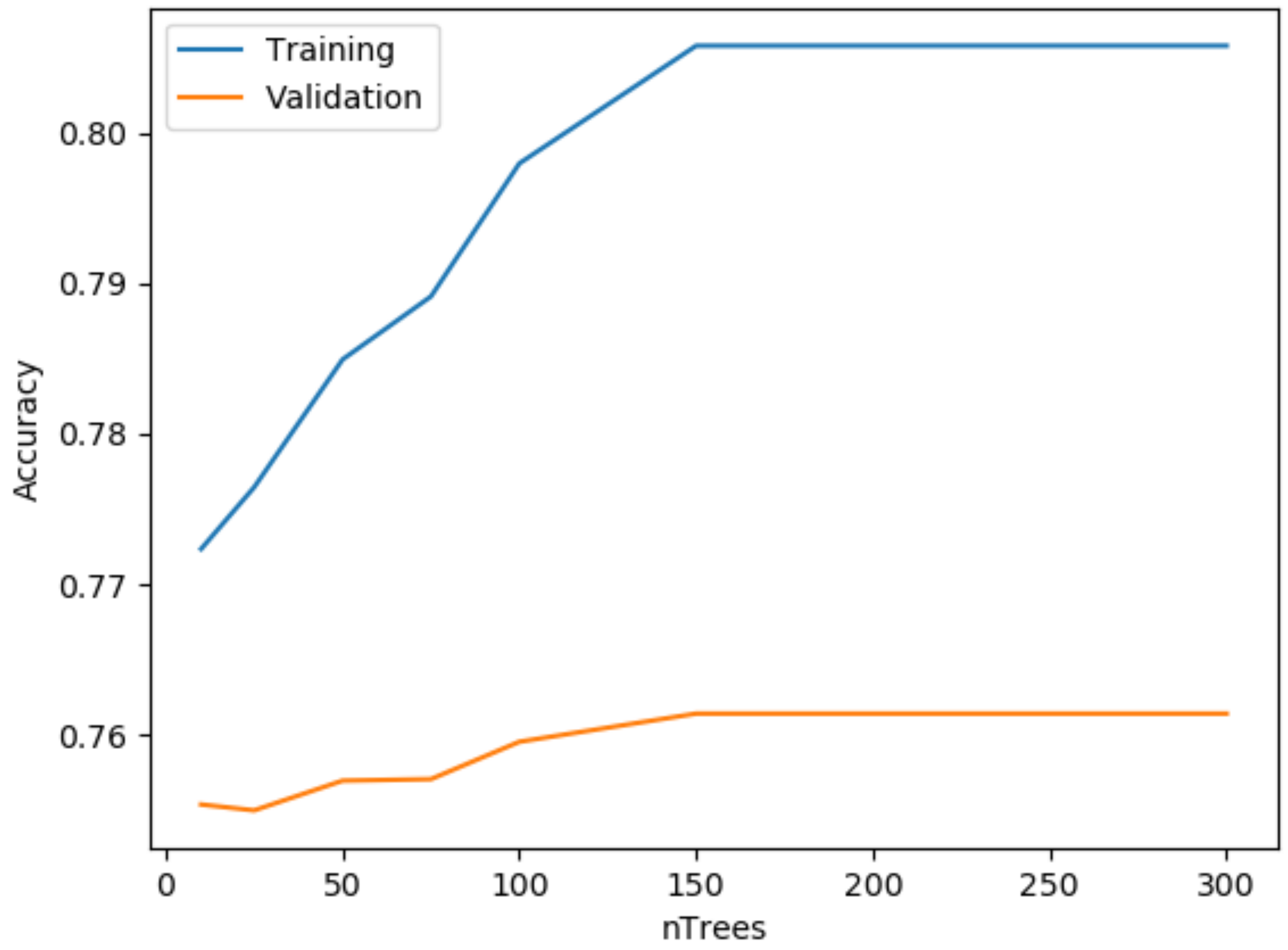
Northern Illinois University

Input List:

- $\Delta\eta_{jj}$, m_{jj}, m_{yybb}, $dR_{\gamma\gamma}$, dR_{bb}, j_1 tagbin, j_2 tagbin, sphericityT, planar flow, pT balance

Hyperparameters for VBF BDT

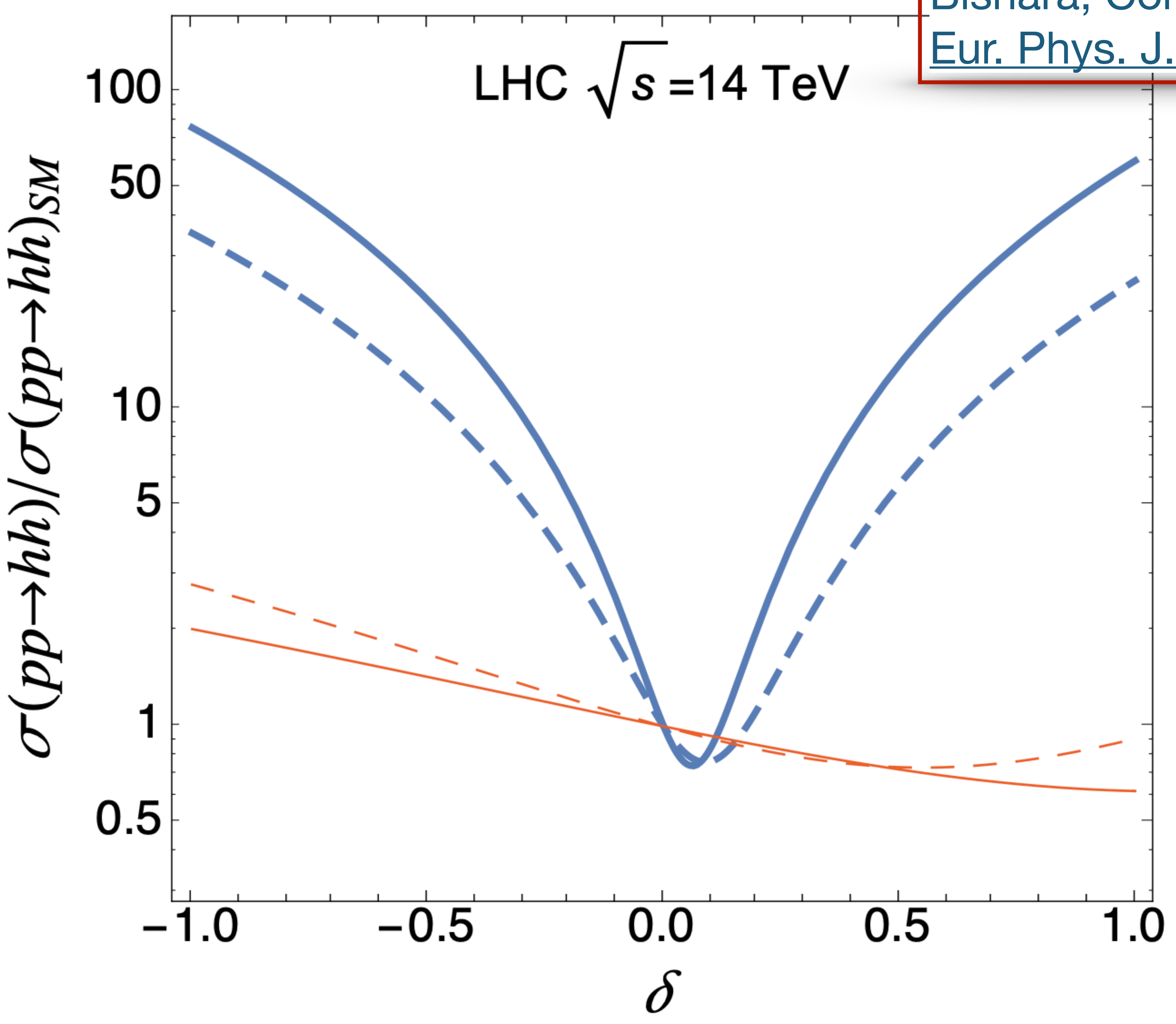




Northern Illinois University

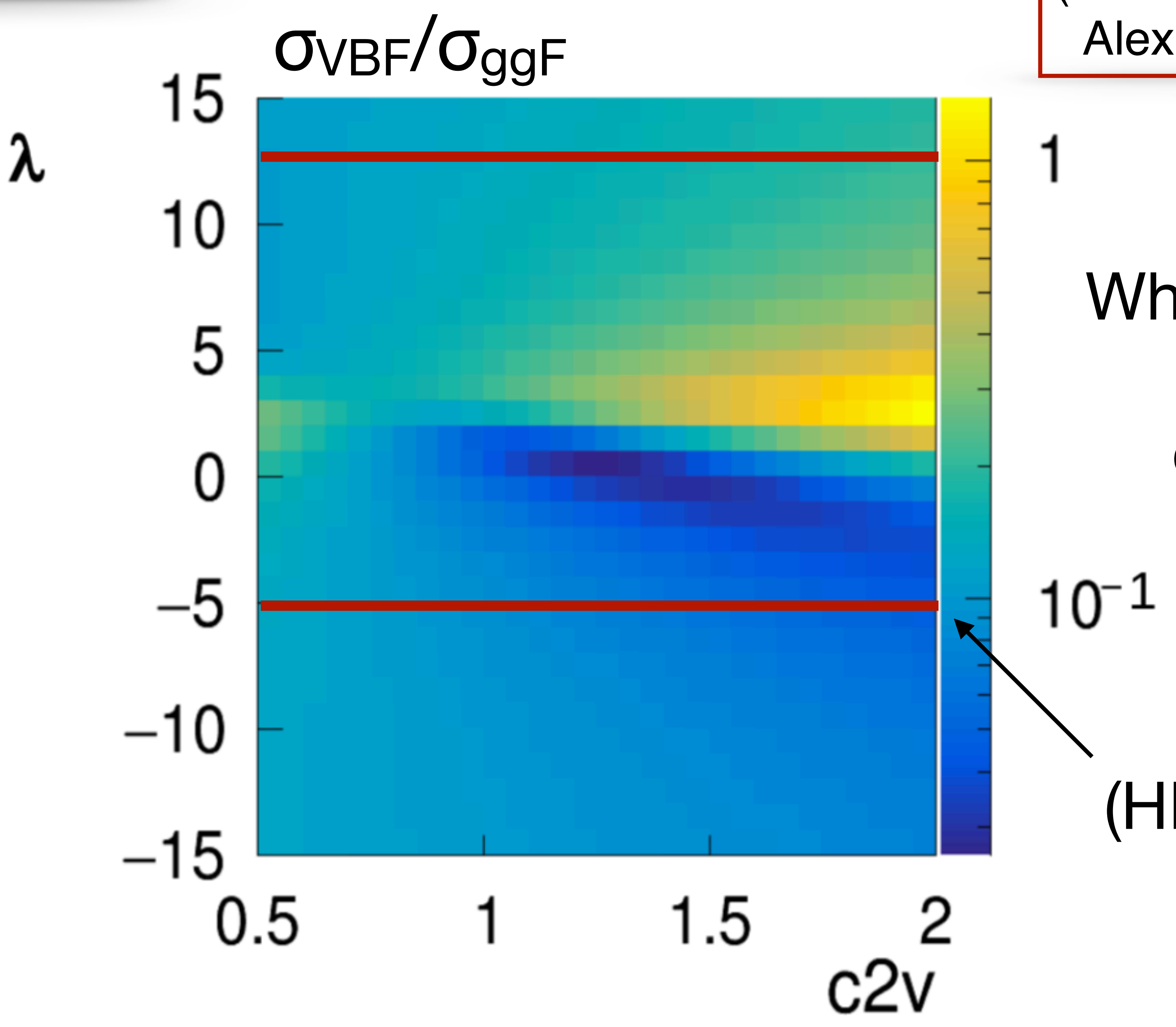
A Note on Enhancements

Coupling variations, specifically to c_{2V}, could cause scenario in which VBF HH production *only* is enhanced



Bishara, Contino, Rojo
Eur. Phys. J. C 77 (2017) 481

Anamika Aggarwal (Masters Thesis) and Alexandra Carvalho



When >1 along Z axis (yellow), VBF is dominant mode

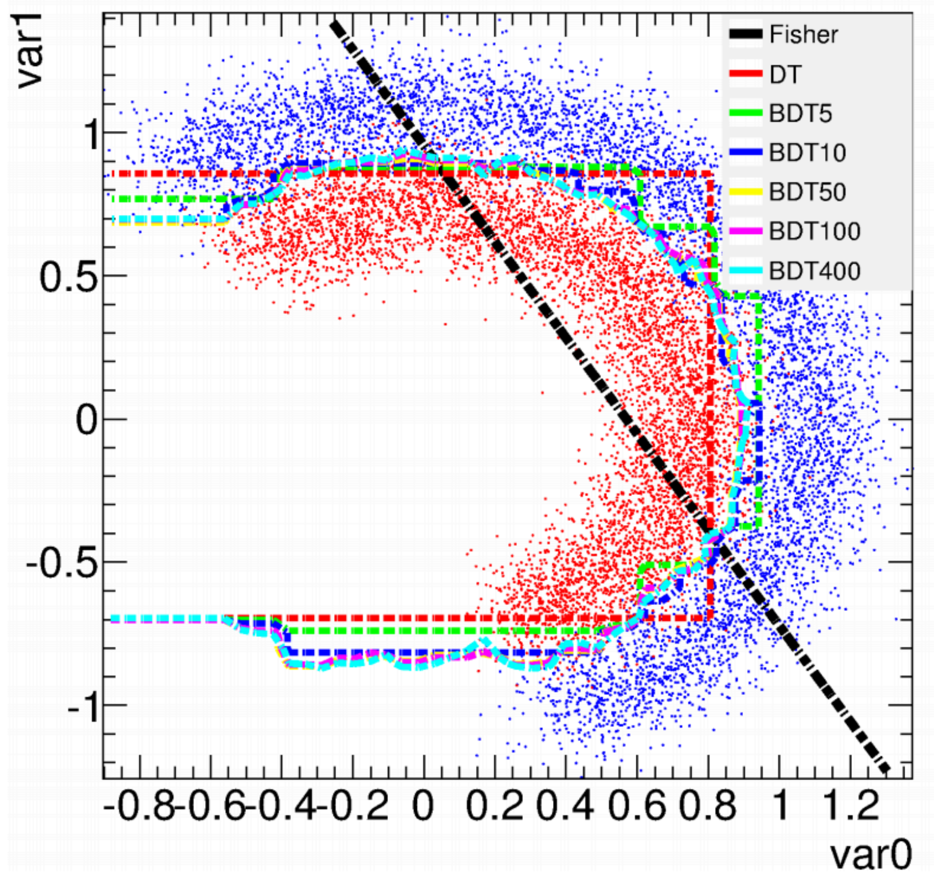
(HH combination constraints)

4b analysis constrained to $-1.0 < c_{2V} < 2.82$, so still regimes in which large enhancements to VBF production are allowed, including some in which VBF would theoretically be dominant

Northern Illinois University

MVA Techniques for Photon ID

from TMVA Documentation



Current methods of photon ID use rectangular cuts optimization over 9 shower shape variables

- Limited solution space in classification

Consider other methods for more sophisticated solution space - MVA techniques

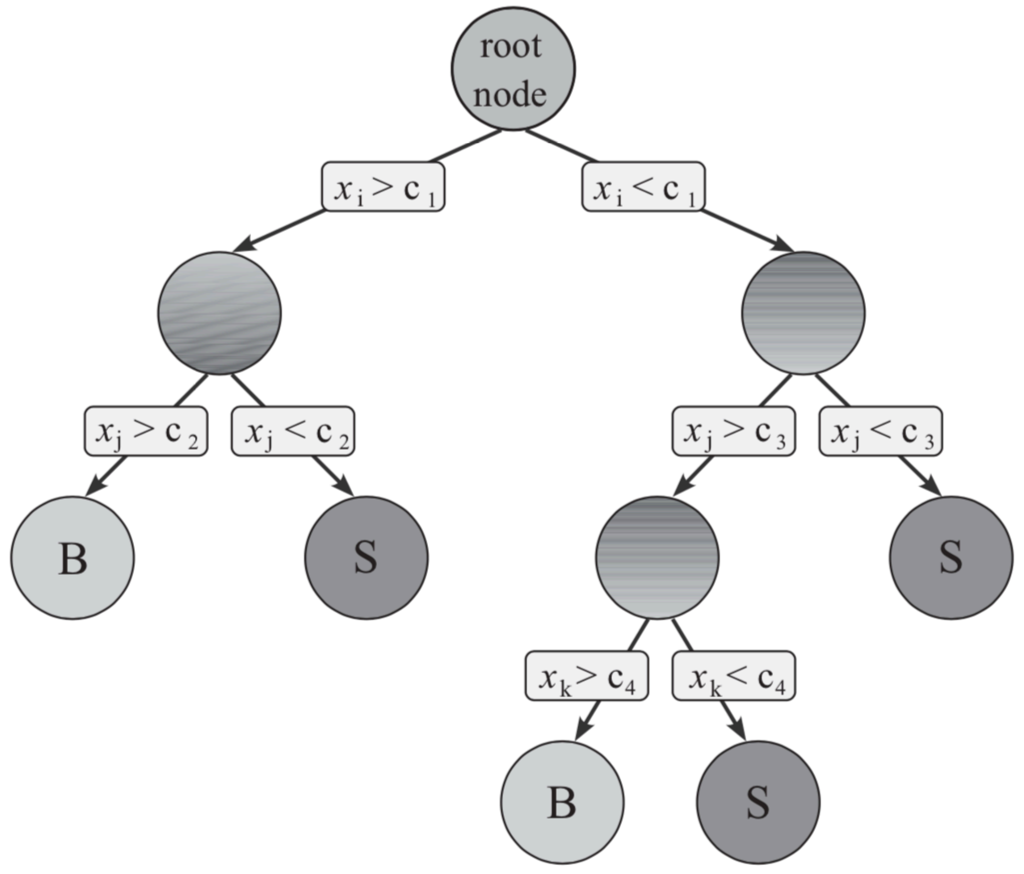
- Two considered, boosted decision tree and neural network

Boosted decision Tree

Single decision tree: familiar “flow-chart” style method of classification

Boosting - combine many weak learners to make one strong learner

Use **gradient boosting** - subsequent trees upweight misclassified events based on residuals of previous misclassifications

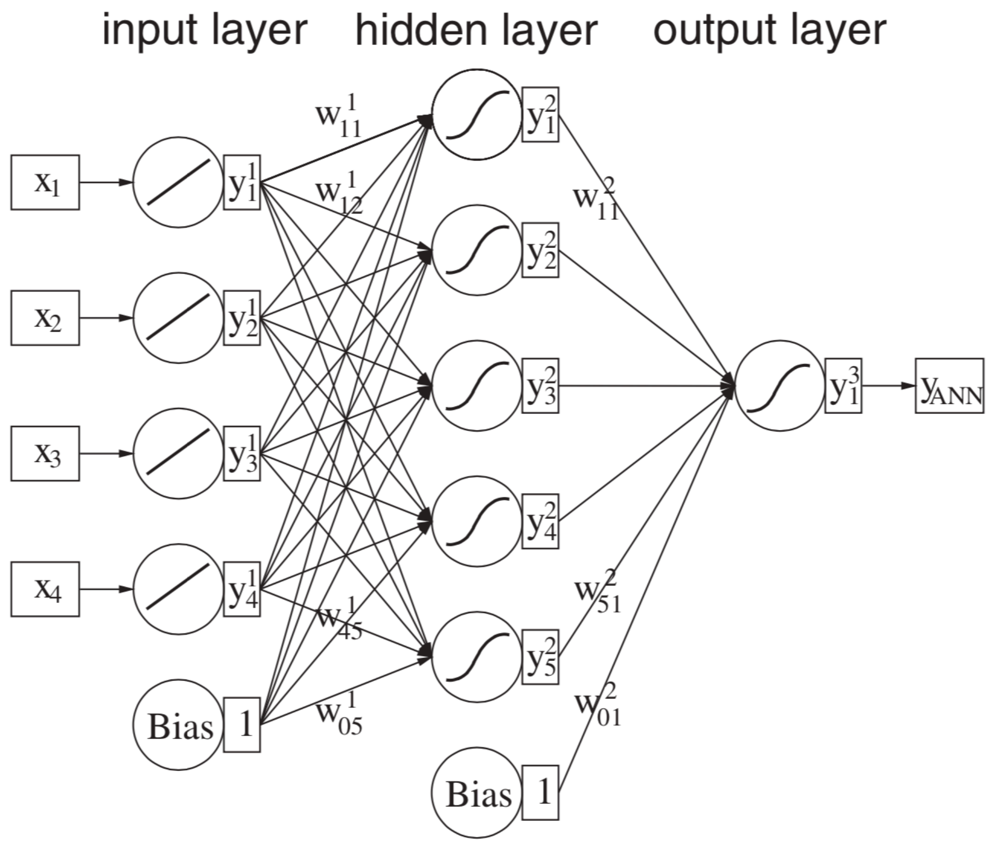


Neural Network

Inputs evaluated over connected nonlinear basis (“activation”) functions, and ultimately issue a classification probability

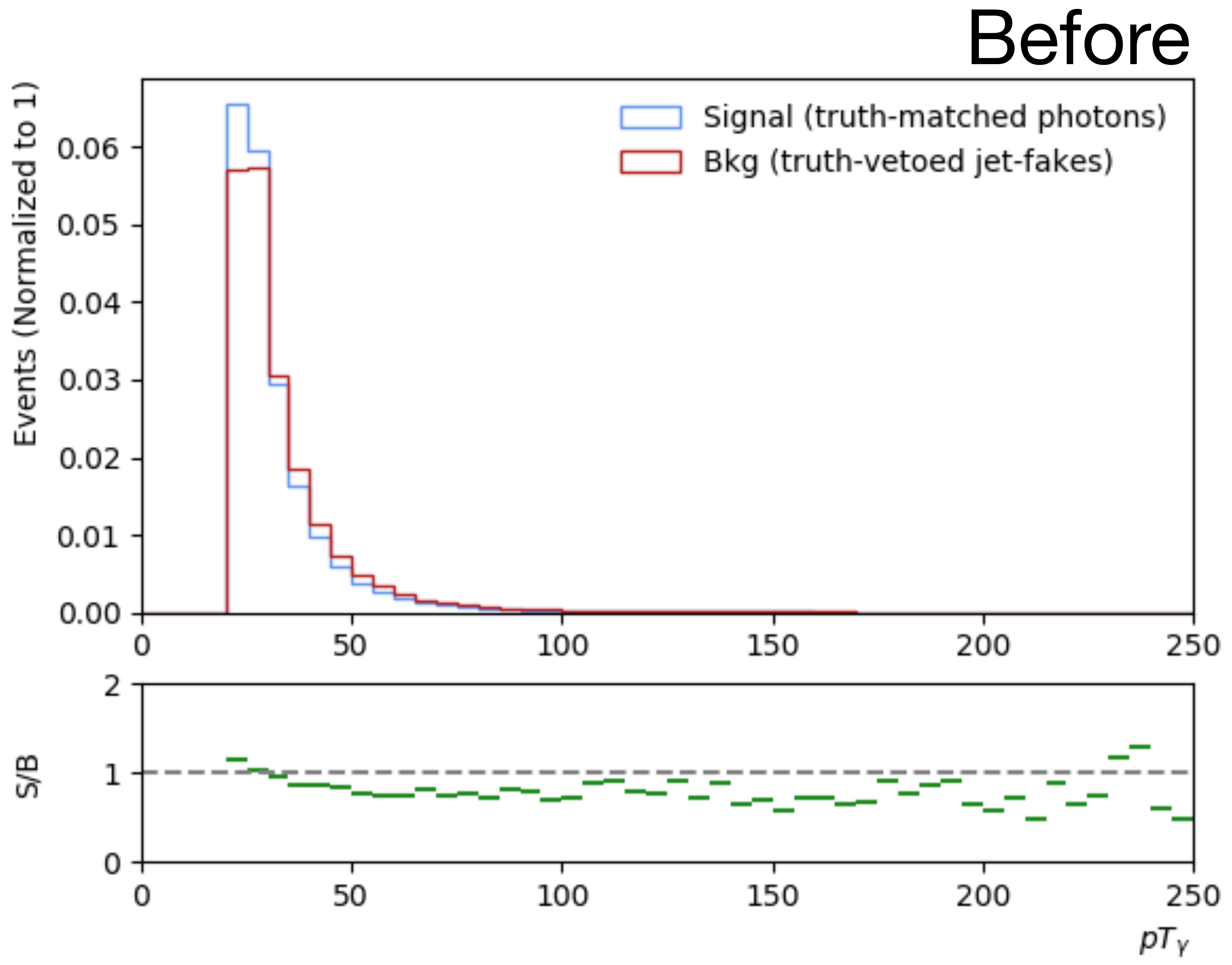
Any possible function can be approximated over sufficient number of activation functions[1]

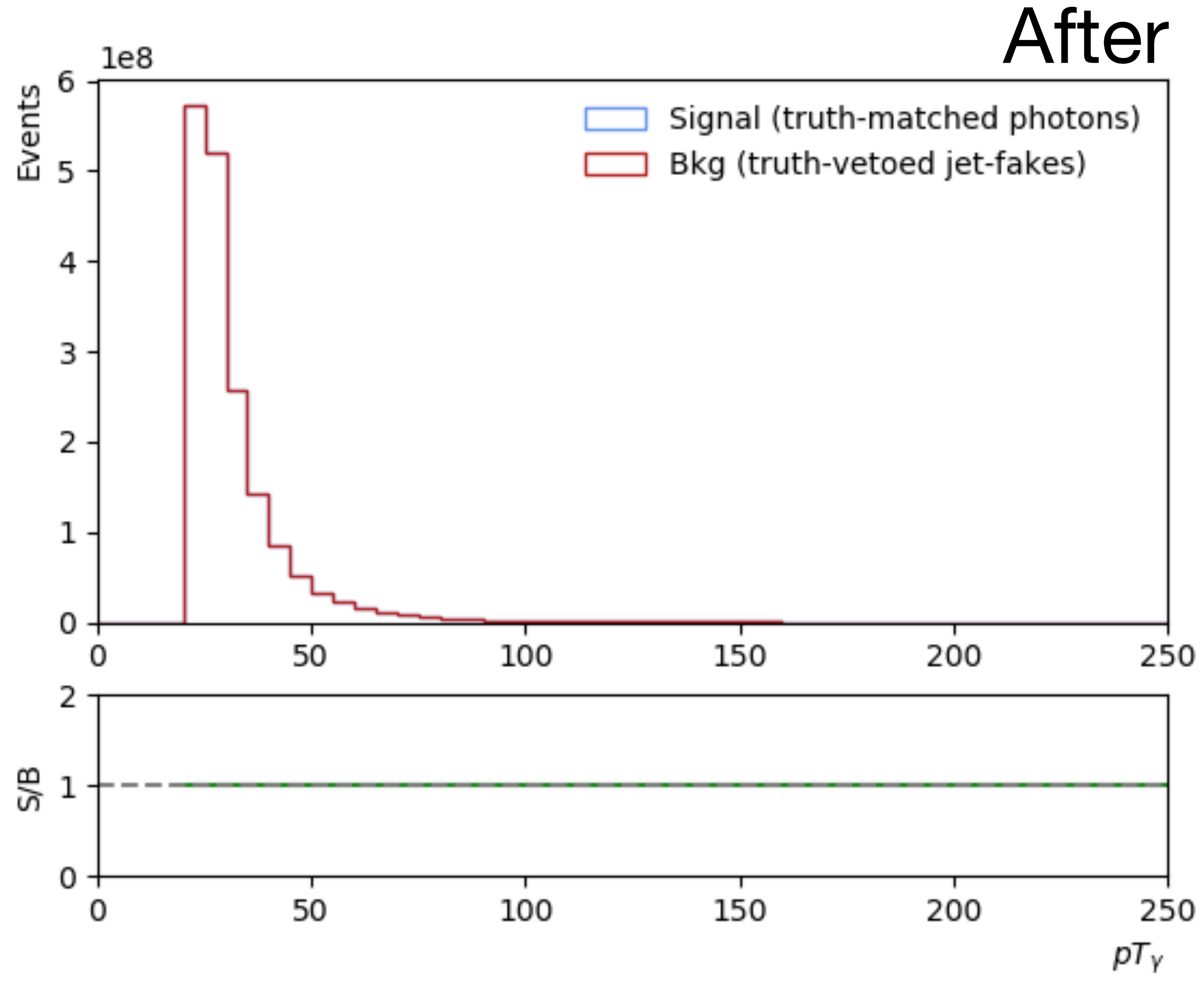
Weights and biases learned during the training procedure

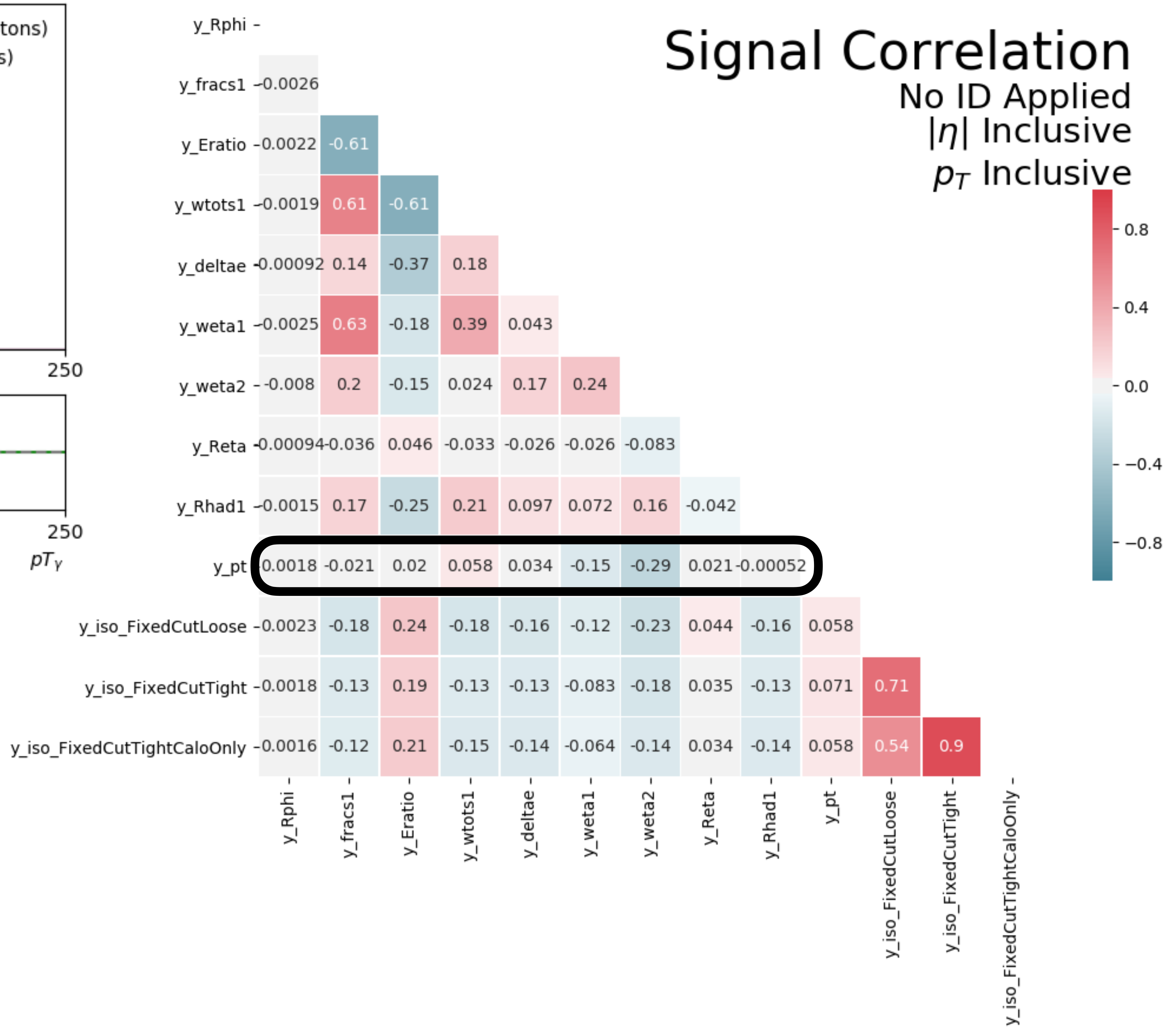


Training Notes

- Training statistics
  - Specifically at high pT, training statistics become a major problem - approach taken was to train inclusively in pT and apply bin-wise
  - Concern about pT dependence, want to make sure it's not just learning pT differences between signal/ background sample - rescale background pT spectrum to look like signal

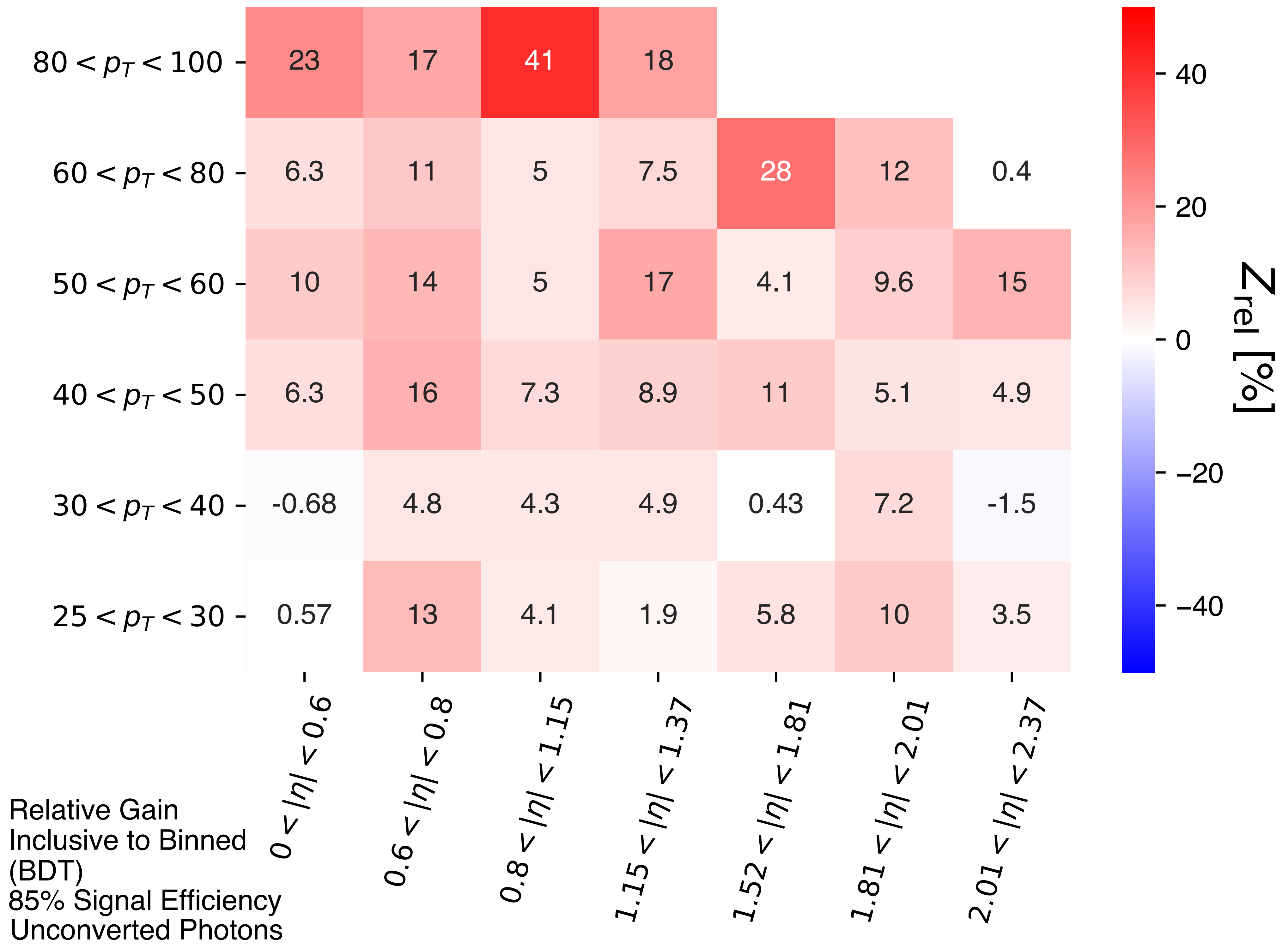






- Found rescaling didn't change training at all, didn't understand why for some time
- Ultimately discovered that shower shape has generally low correlation to pT, so correction not needed

BDT Inclusively Trained vs binwise



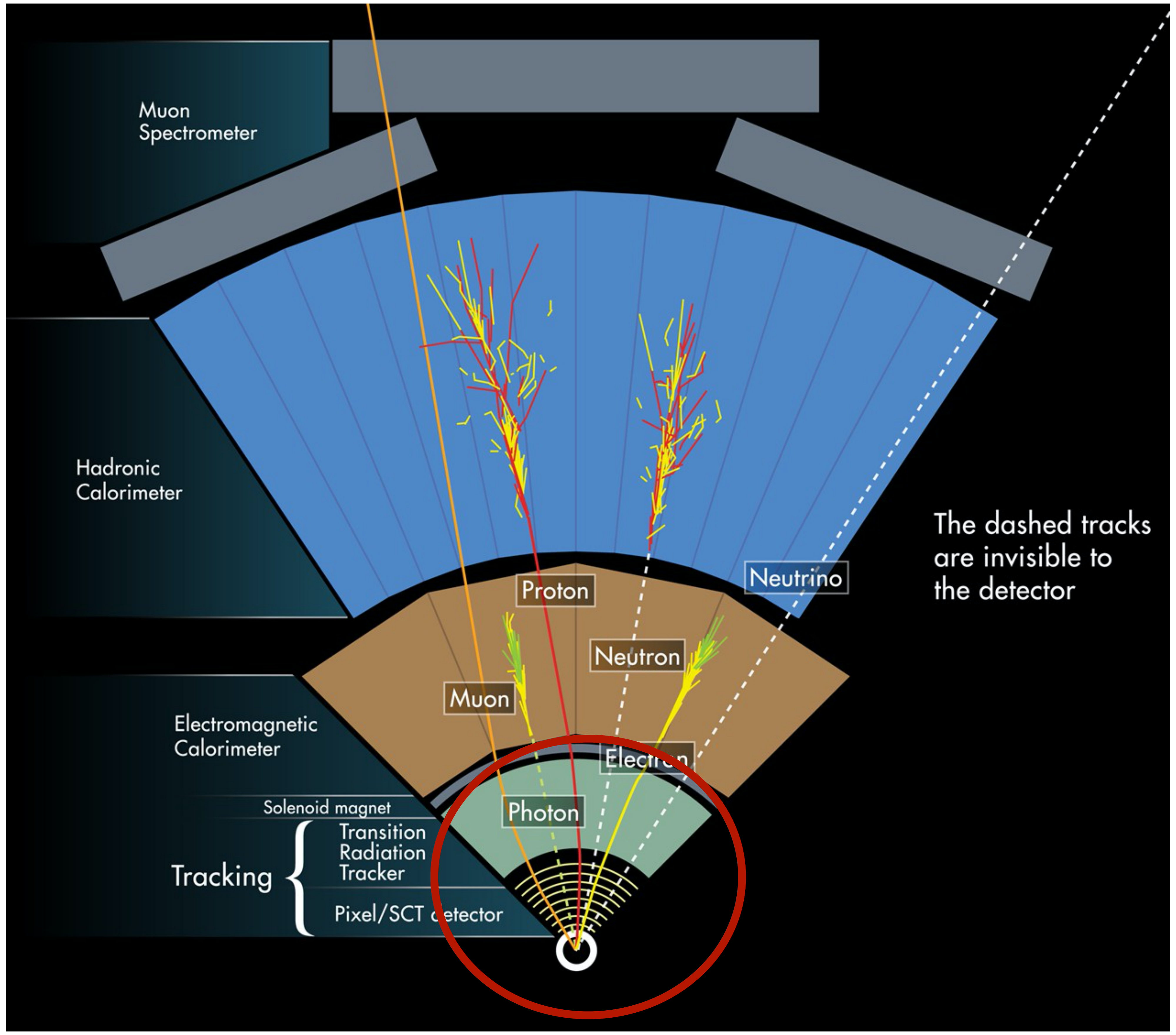
Particles in…

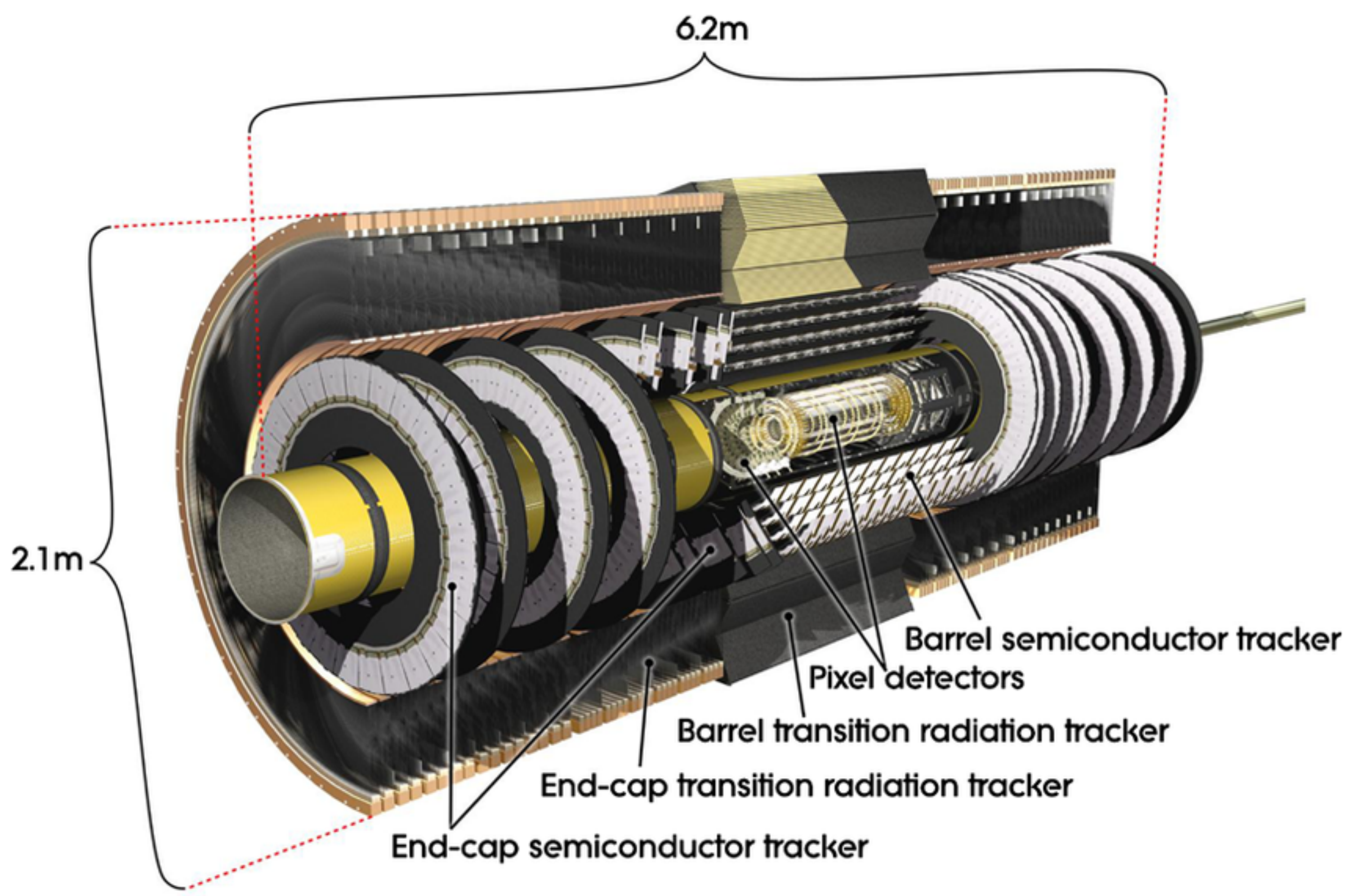
ATLAS is a many-layered detector

Each layer targets specific information from collision decay products

Inner detector - Tracking information

- Describing the particle trajectory through the detector and magnetic fields
- Provides tracking information, vertex measurement, momentum and charge information
- Consists of 3 subsystems:
  - Pixel Detector
  - SemiConductor Tracker
  - Transition Radiation Tracker



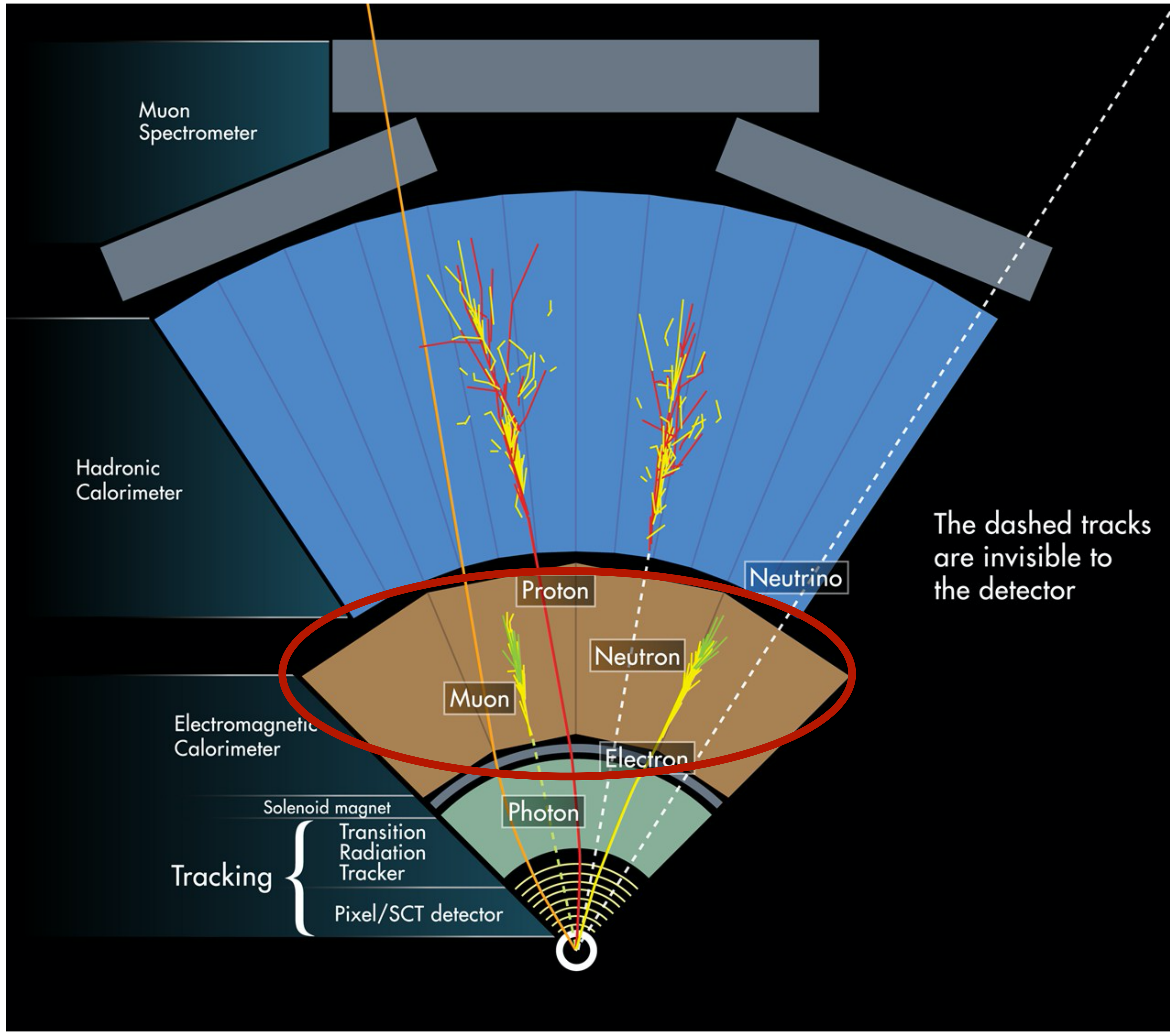


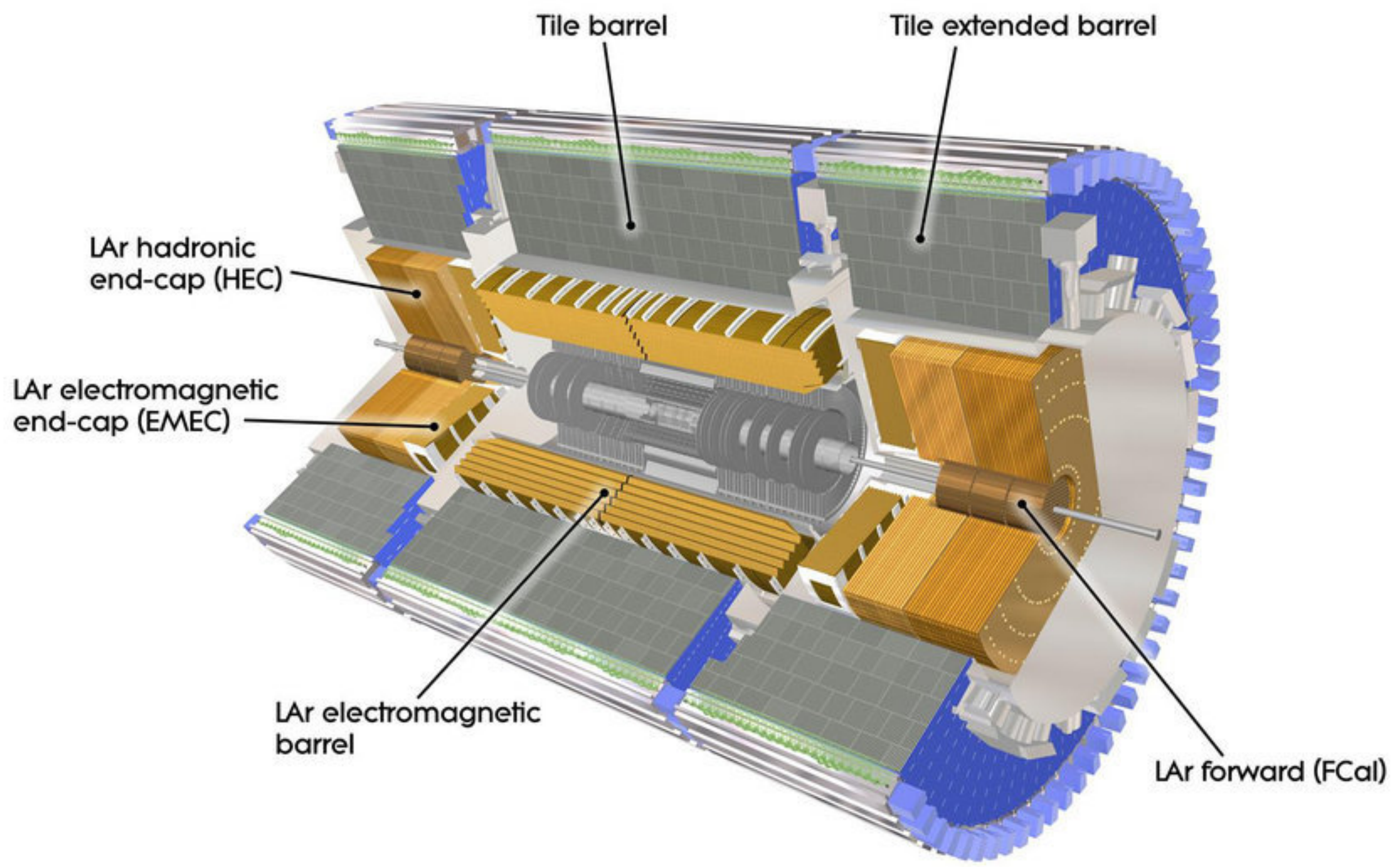
Particles in…

ATLAS is a many-layered detector
Each layer targets specific information from collision decay products

Inner detector - Tracking information
Electromagnetic Calorimeter - Electromagnetic (EM) signatures (γ,e)

- Liquid Argon calorimeter
- Accordion-style geometry - high granularity
- In material, electromagnetic showers occur, which are based on 2 processes
  - Bremsstrahlung photon emission
  - Electron-positron conversion





Particles in…

ATLAS is a many-layered detector

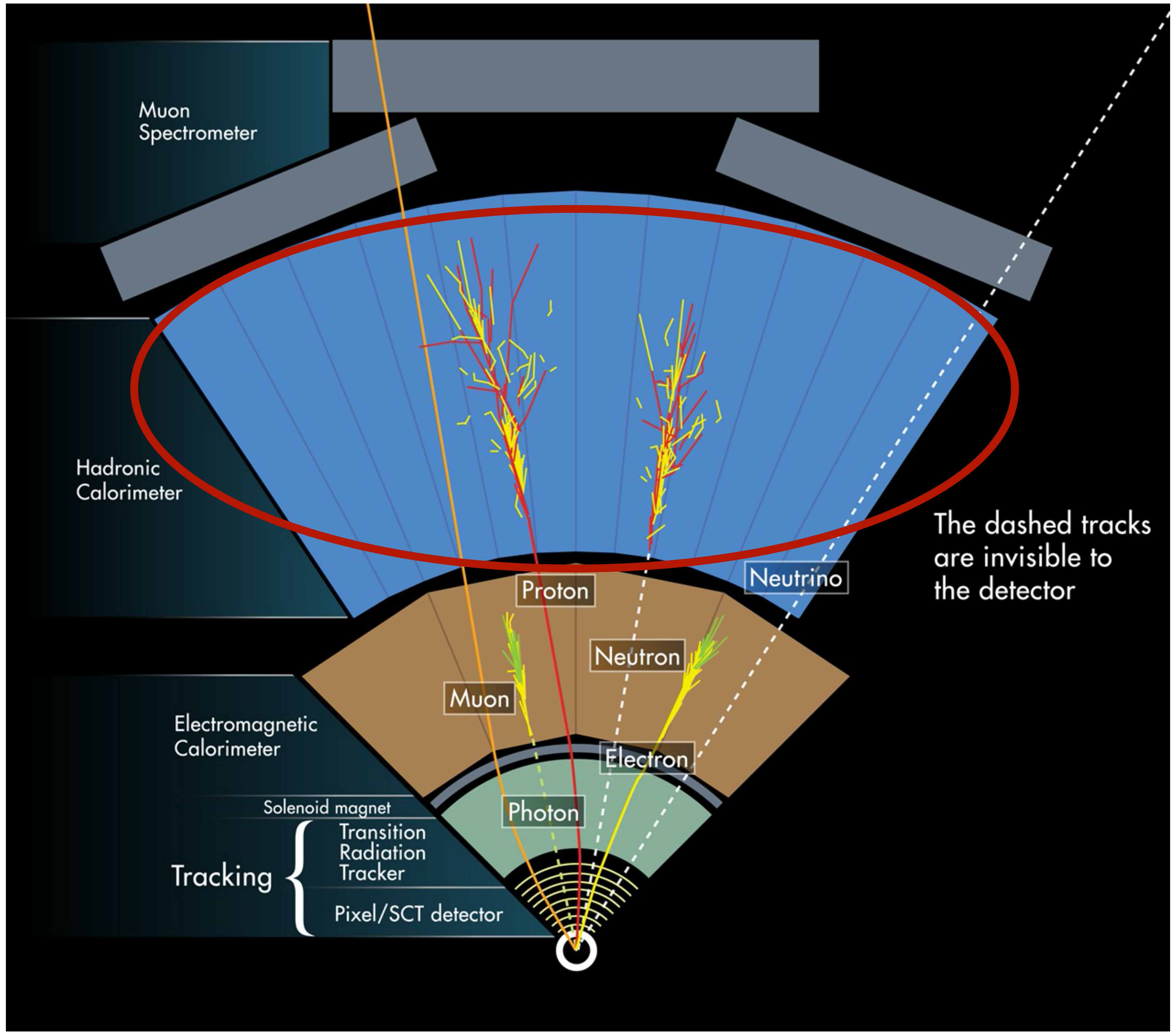
Each layer targets specific information from collision decay products

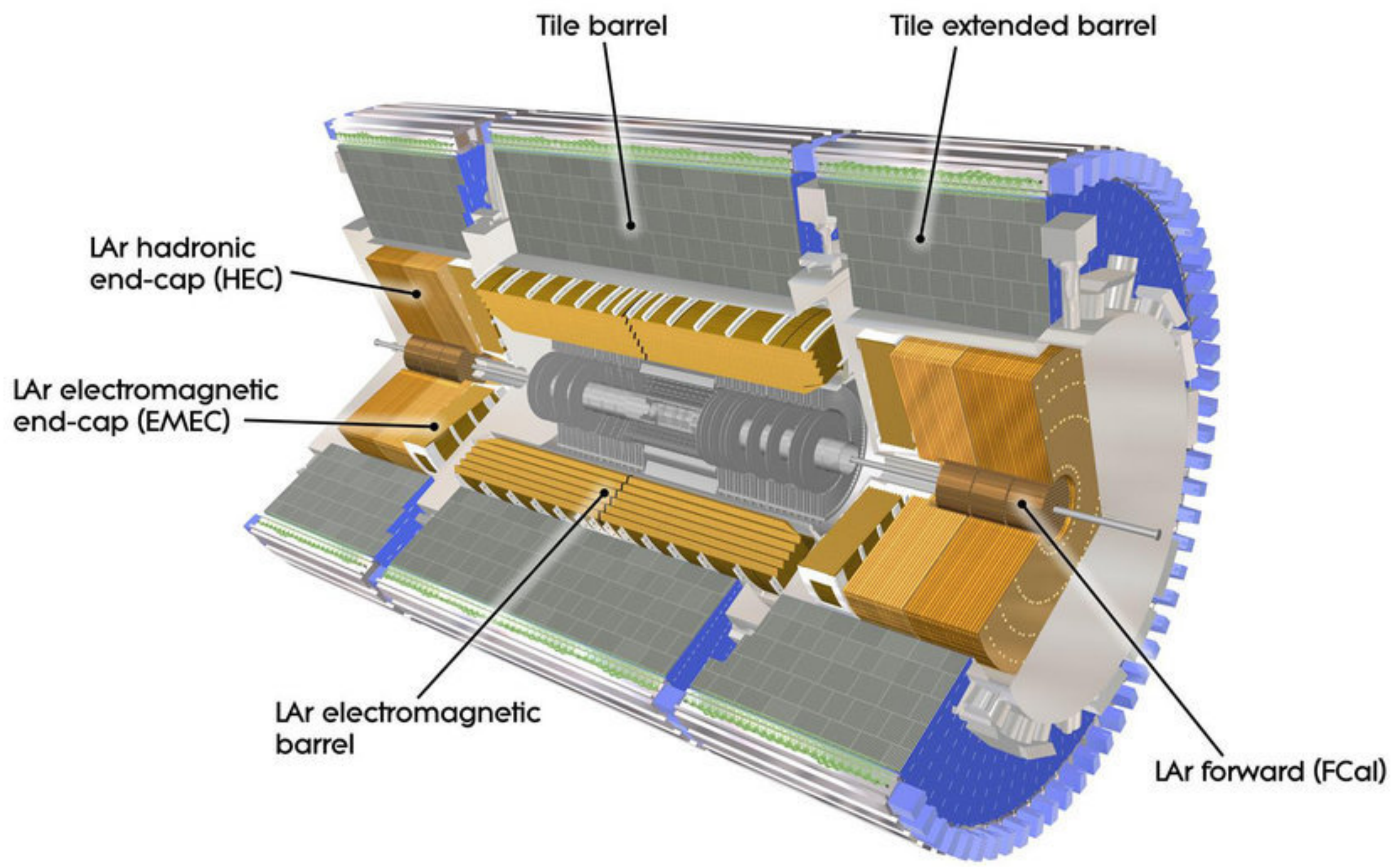
<u>Inner detector</u> - Tracking information

<u>Electromagnetic Calorimeter</u> - Electromagnetic (EM) signatures (γ,e)

<u>Hadronic Calorimeter</u> - Particles that interact via the strong force (quarks, gluons)

- Initiating strongly interacting particles may radiate additional quarks and gluons
- Must all be in colorless bound states due to *confinement*, so eventually turn into hadrons (π, K, p, n, etc.)
- Final signature is a collimated spray of hadrons "jets"
- As hadrons interact with matter they scatter and produce additional hadrons, cascade known as hadron shower
- Calorimeter to measure showers - steel absorber, scintillating tile active material





Particles in…

ATLAS is a many-layered detector

Each layer targets specific information from collision decay products

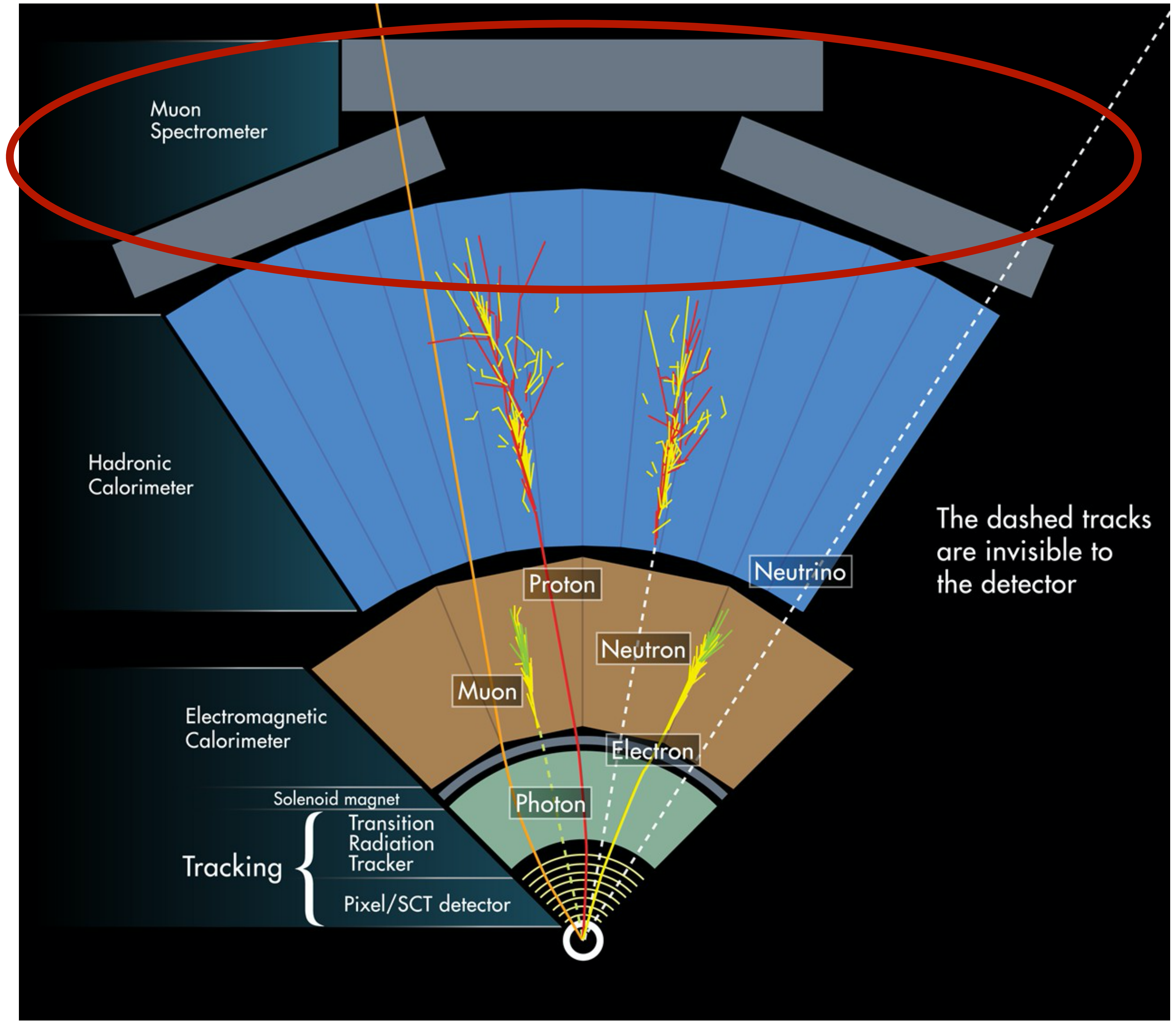
Inner detector - Tracking information

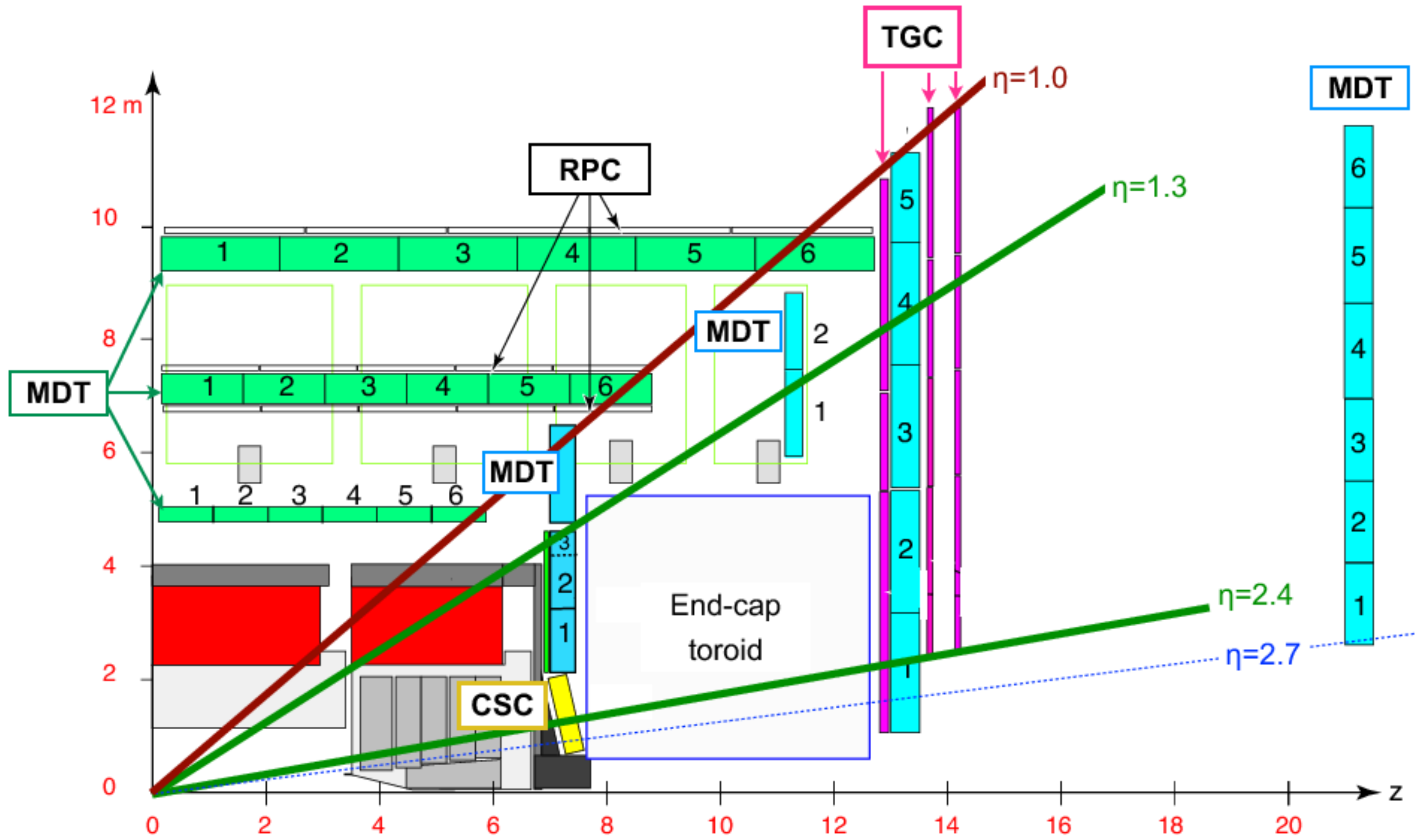
Electromagnetic Calorimeter - Electromagnetic (EM) signatures (γ,e)

Hadronic Calorimeter - Particles that interact via the strong force (quarks, gluons)

Muon Spectrometer - Muons

- Don't interact in upstream material, so dedicated subsystem for detecting
- Due to distance from IP, independent triggering, combine with ID information for final decision





Particles in…

ATLAS is a many-layered detector

Each layer targets specific information from collision decay products

<u>Inner detector</u> - Tracking information

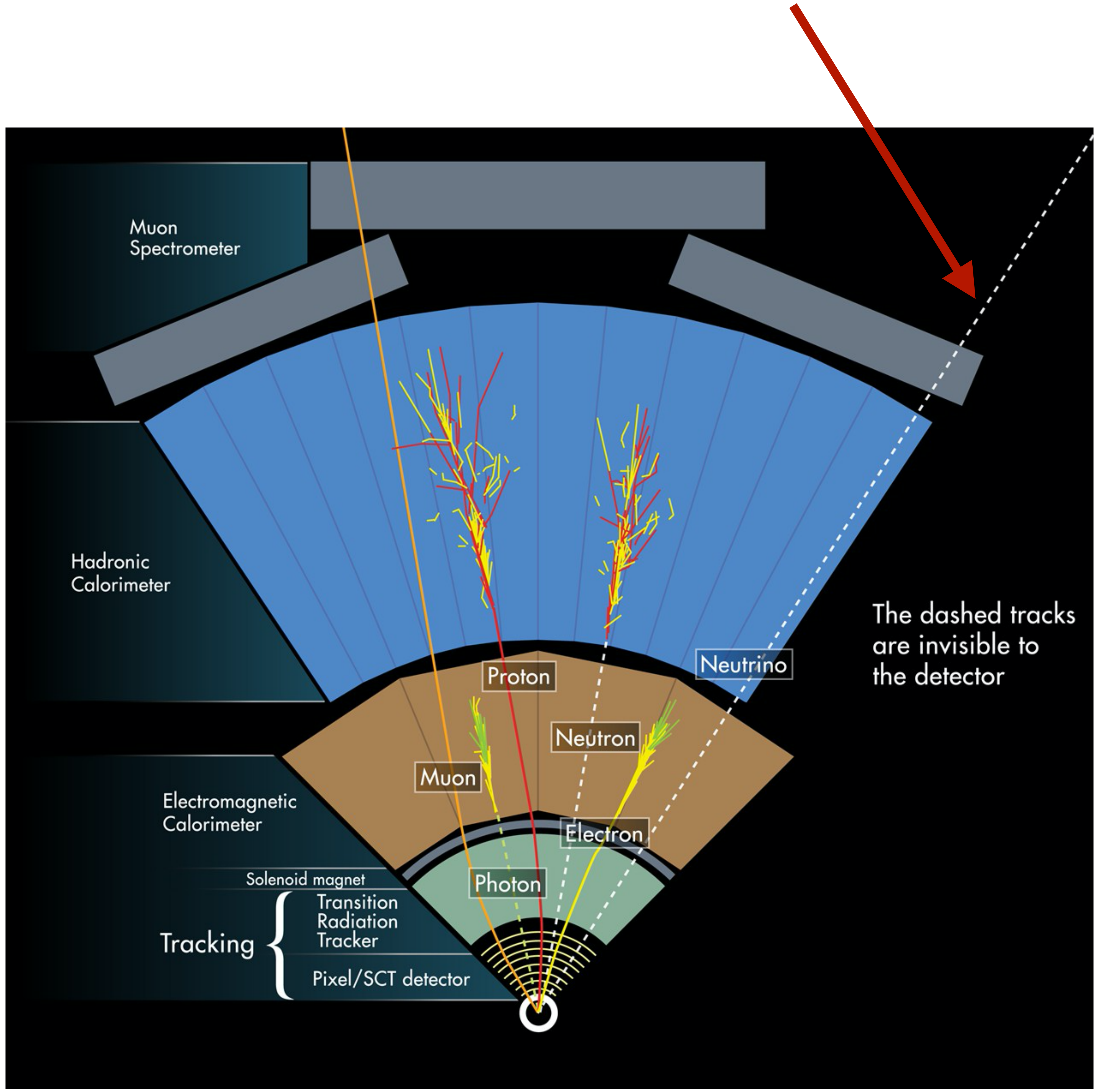
<u>Electromagnetic Calorimeter</u> - Electromagnetic (EM) signatures (γ,e)

<u>Hadronic Calorimeter</u> - Particles that interact via the strong force (quarks, gluons)

<u>Muon Spectrometer</u> - Muons

<u>Missing Transverse energy (MET)</u> - Neutrinos, BSM Signatures

- Particles that don't interact with any portion of the detector
- Conservation of momentum - all particles reconstructed should sum to 0, imbalances in this are taken as MET



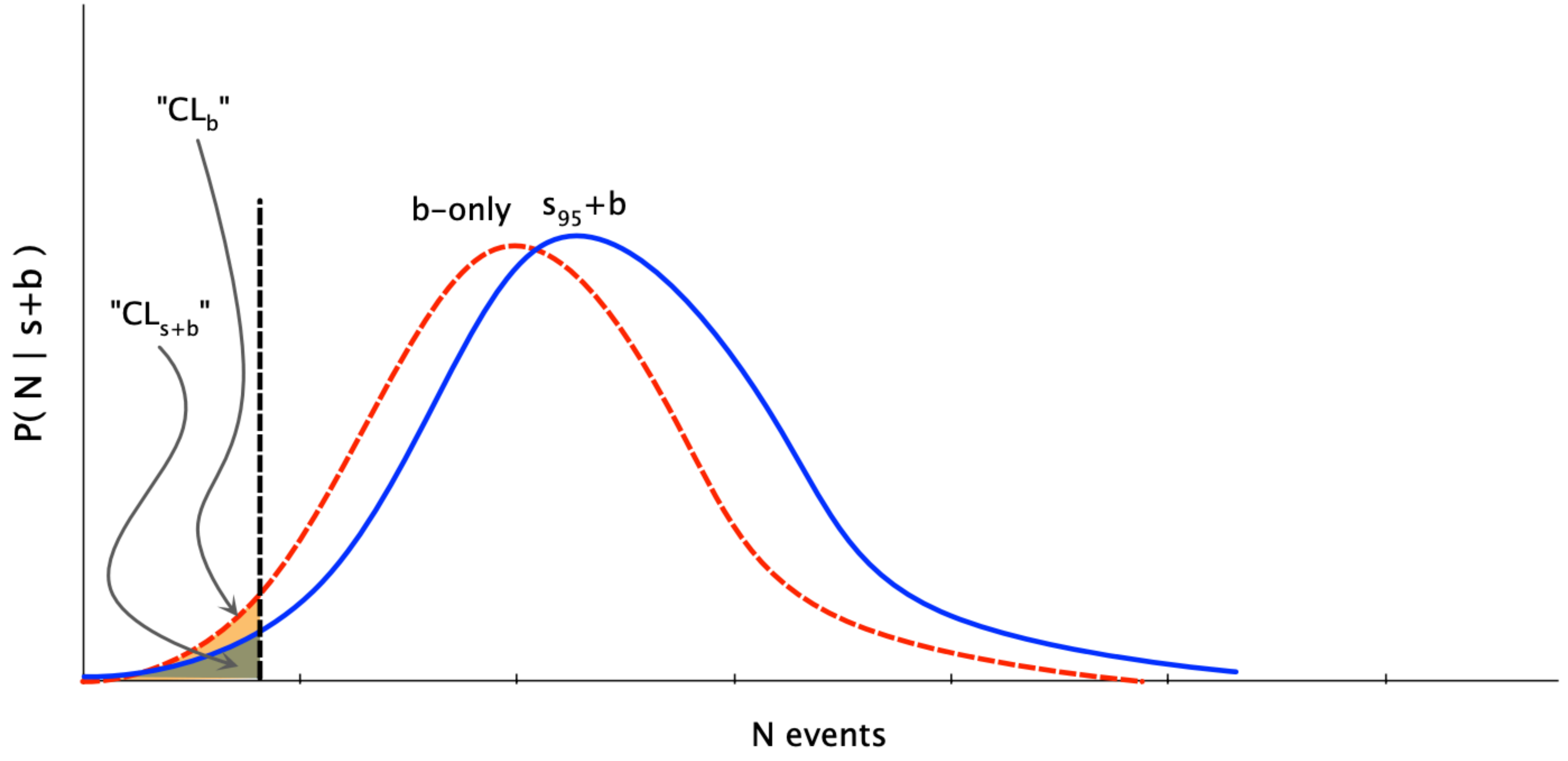
CLs

To address the sensitivity problem, CLs was introduced http://inspirehep.net/record/599622

- common (misused) nomenclature: $CL_s = CL_{s+b}/CL_b$
- idea: only exclude if CL_s<5% (if CL_b is small, CL_s gets bigger)

CL_s is known to be "conservative" (over-cover): expected limit covers with 97.5%

- Note: CL_s is NOT a probability



Cranmer

Northern Illinois University