



# Flavor tagging performance in ATLAS and CMS

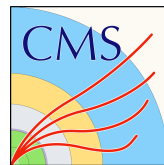
with an emphasis on HH and multi-Higgs searches



**Nicole Hartman**

nicole.hartman@tum.de

obo ATLAS and CMS collaborations



**CO**mprehensive **M**ulti-boson  
**E**xperiment-**T**heory **A**ction workshop

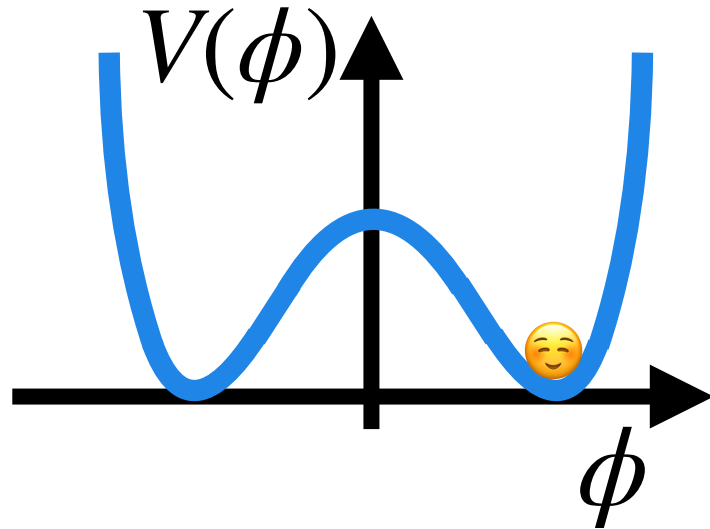
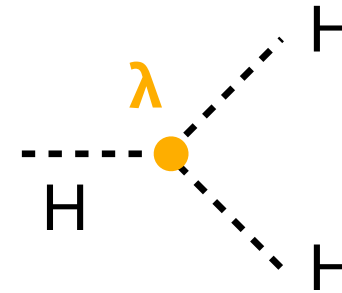
Izmir, Turkey

29 Feb 2024

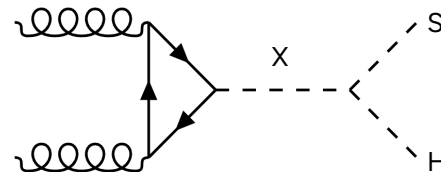


# Motivation – why do we need b-tagging?

$$V(\phi) = \mu^2 h(x)^2 + \lambda v h(x)^3 + \frac{1}{4} \lambda h(x)^4$$



... or directly search for resonances that could modify this potential



Focus for today!  
Main HH channels

Higgs 1 decay

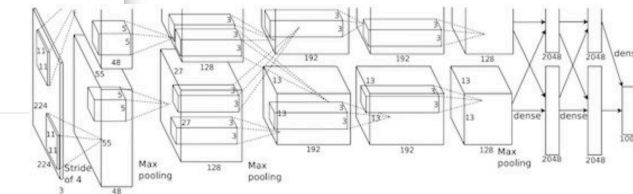
	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
$\gamma\gamma$	0.26%	0.10%	0.028%	0.012%	0.0005%

# History of b-tagging

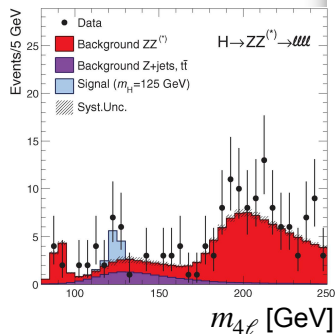
**ATLAS TDR 1999** [T]he  $H \rightarrow bb$  decay mode is dominant ... [but] the extraction of a signal from  $H \rightarrow bb$  decays in the  $WH$  channel will be very difficult at the LHC, even under the most optimistic assumptions for the  $b$ -tagging performance and calibration of the shape and magnitude of the various background sources from the data itself.

## "Alex Net"

Krizhevsky, Sutskever, Hinton, 2012



Today  $H(bb)$  key for  
 $HH$  and Higgs-like  
resonance searches!



[Higgs discovery](#)

The New York Times

### Physicists Find Elusive Particle Seen as Key to Universe

Share full article | 122 comments

Scientists in Geneva on Wednesday applauded the discovery of a subatomic particle that looks like the Higgs boson. Pool photo by Denis Balibouse

By Dennis Overbye  
July 4, 2012

[article](#)

The New York Times

### Scientists See Promise in Deep-Learning Programs

Share full article

A voice recognition program translated a speech given by Richard F. Rashid, Microsoft's top scientist, into Mandarin Chinese. Hao Zhang/The New York Times

By John Markoff  
Nov. 23, 2012

[article](#)

# FTAG history



**ATLAS**



**CMS**

early Run 2

Shallow  
NNs / BDTs

MV2c10

DeepCSV

full Run 2

Deep NNs  
(RNNs / Sets)

RNNIP /  
DL1r

DeepJet

late Run 2 /  
early Run 3

Graph NNs

GN1

ParticleNet

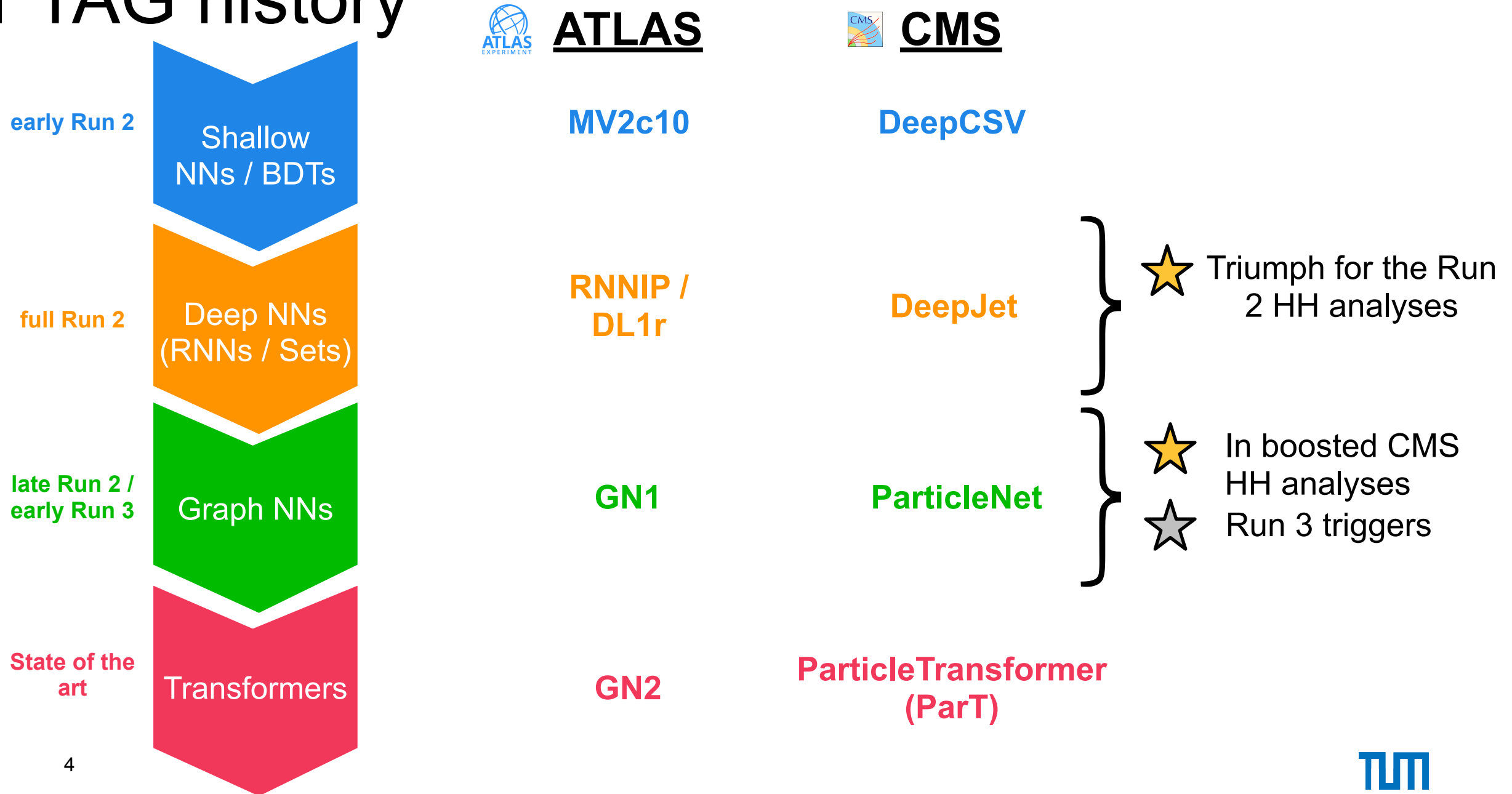
State of the  
art

Transformers

GN2

ParticleTransformer  
(ParT)

# FTAG history



# What does a “b”-jet look like?

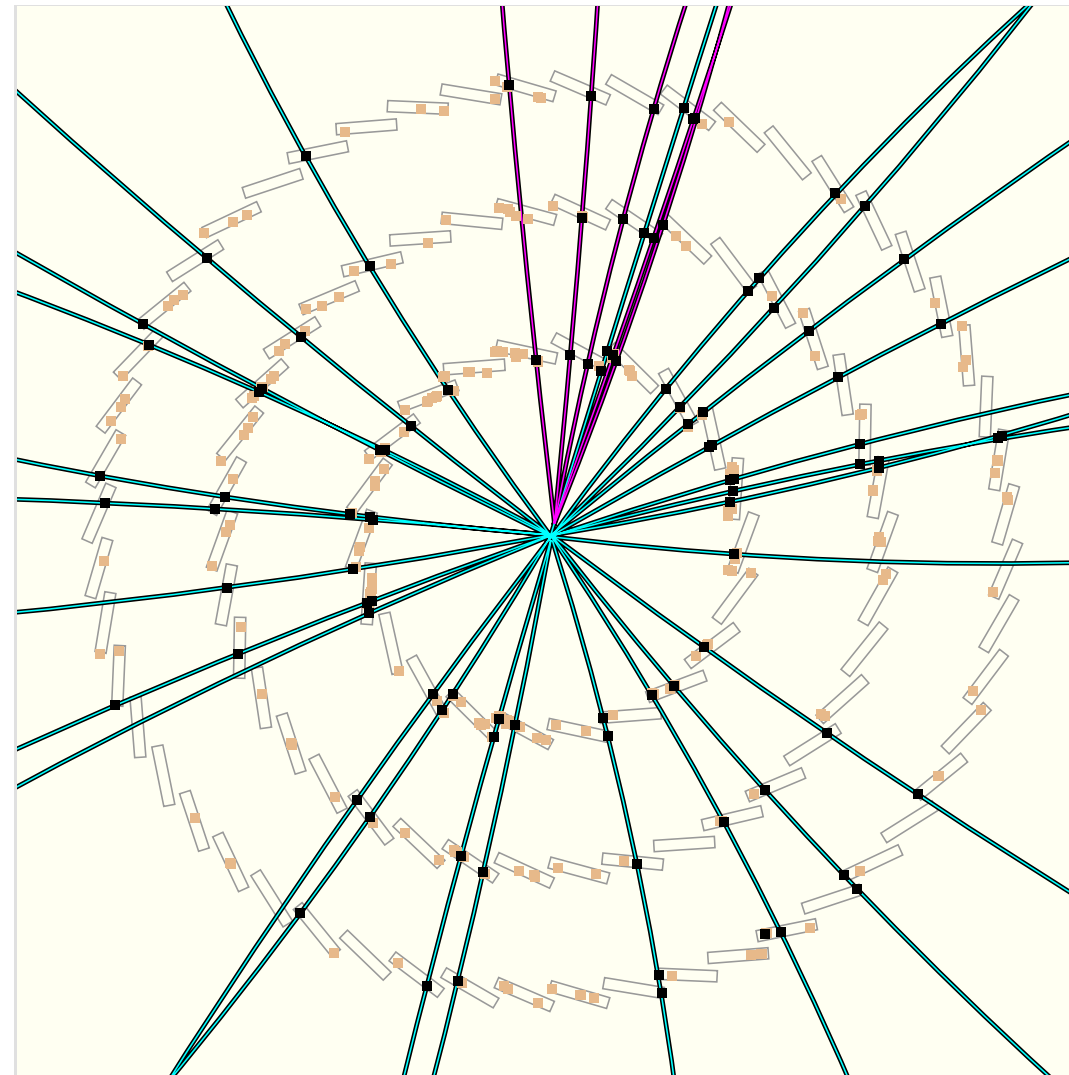
baseline

## b-jet properties:

- “Long” lifetime  $\tau = 1.2$  ps
- Massive B-hadron ( $\approx 5$  GeV)
- $\approx 5$  displaced tracks
- Harder fragmentation

RNNs  
+Sets

Graphs

Trans-  
formers

Run Number: 142195, Event Number: 284154

Date: 2009-12-12 12:24:43 PST

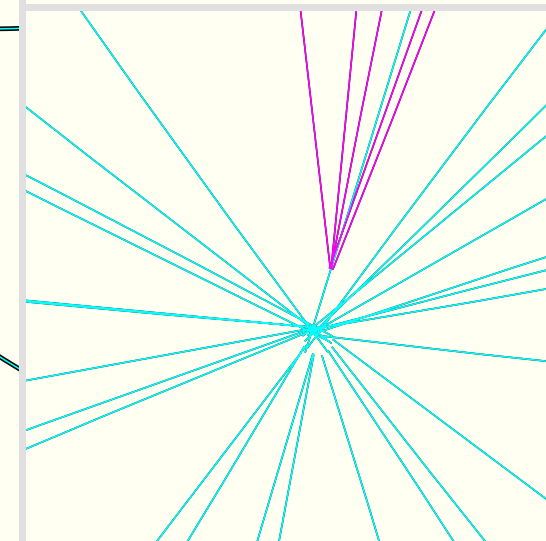
Decay length = 3.7 mm

Decay length significance = 22

Lifetime = 3.1 ps

Vertex mass = 2.5 GeV

Number of tracks = 5



# Early Run 2 taggers

Key idea: Hand engineered algorithms aggregated via a high-level classifier

baseline

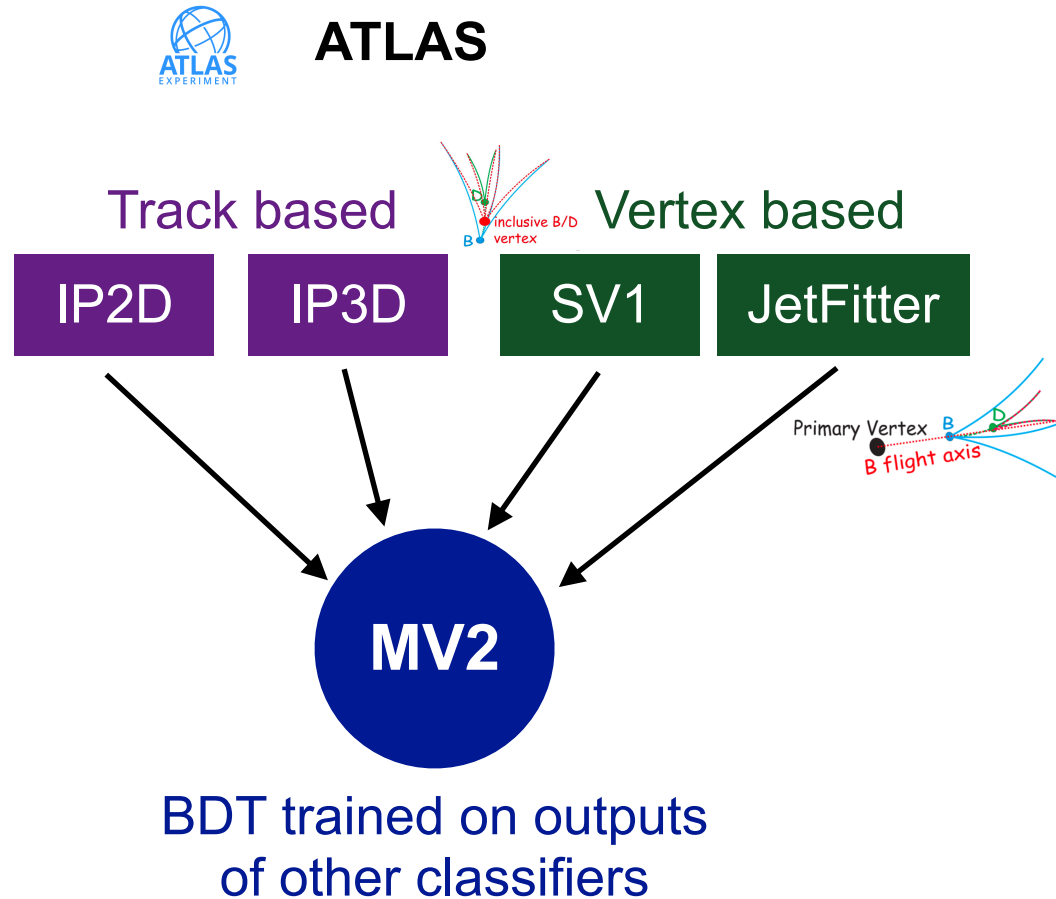
RNNs + CNNs

Sets + Graphs

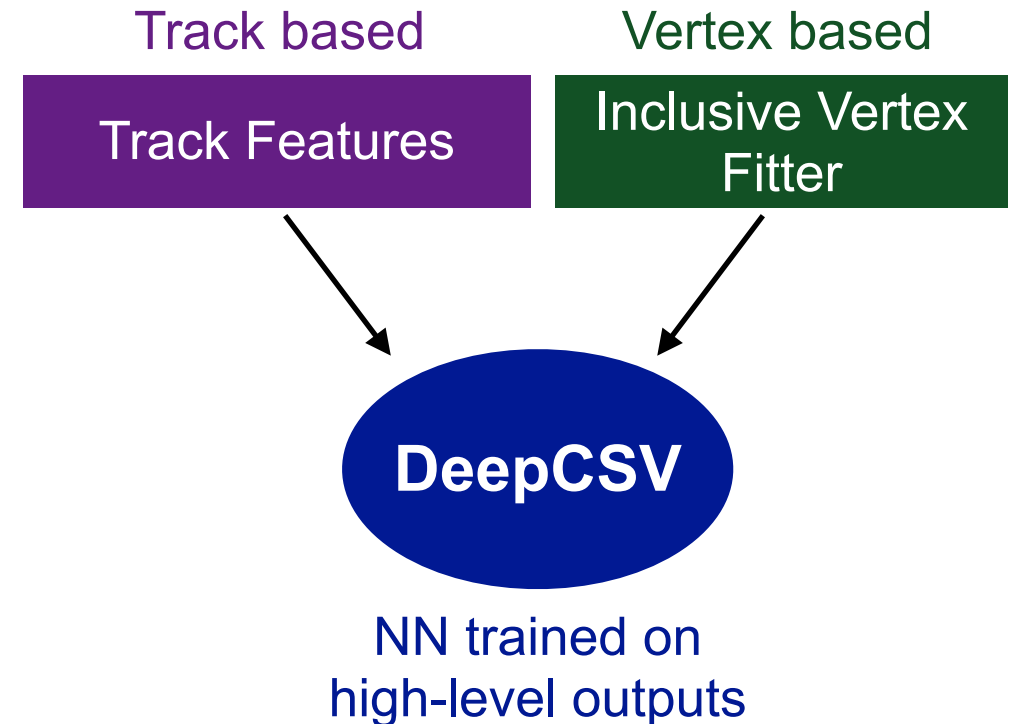
Transformers

Low level

High level





## CMS: Deep CSV



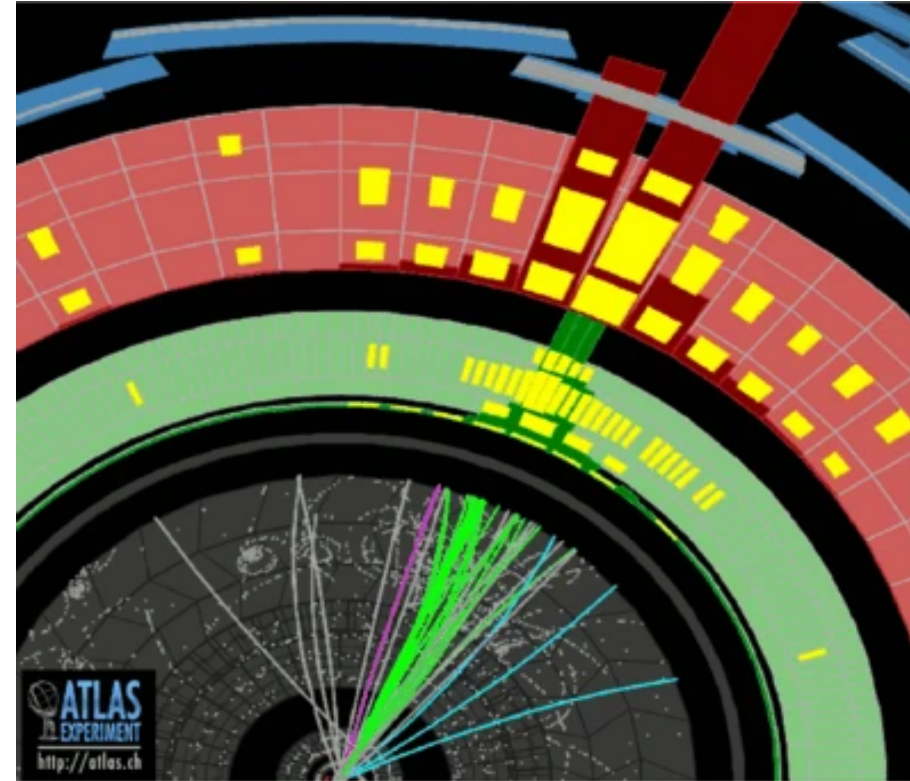
\* [slide](#) with input details

# How do we aggregate this information?

## What we have:

- Collection of tracks
  - $X_i : i = \{1, \dots, \underline{n}\}$   
 variable # of tracks
- Each track has features
  - $X_i \in \mathbb{R}^m$   E.g, impact parameters  
momenta, quality
- Jet has labels
  - $Y: \{\mathbf{b}, \mathbf{c}, \mathbf{light}\}$  – or –
  - $Y: \{\mathbf{bb}, \mathbf{cc}, \mathbf{top}, \mathbf{QCD}\}$

What we want:  $p(Y | X_1, \dots, X_n)$



High dimensional problem

$$n \cdot m \sim \mathcal{O}(10^3)$$



# Key development: Recurrent Neural Network

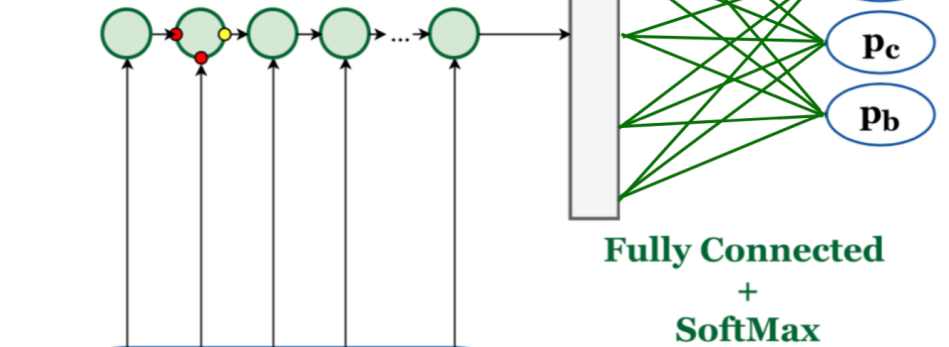
5m jets



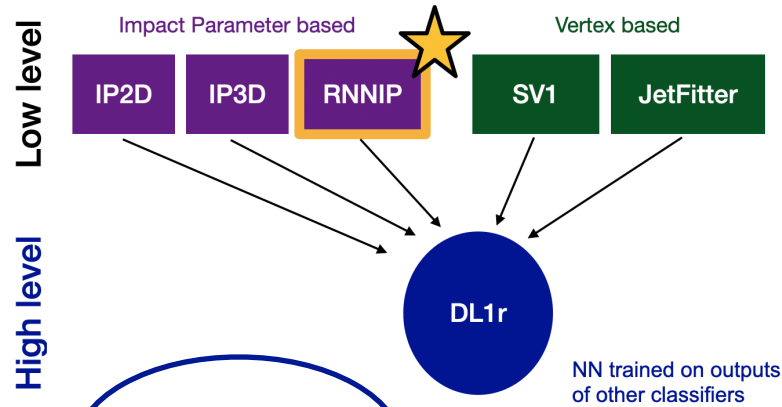
## Model the jet as a sequence

- ☑ Account for correlations between tracks
- ☑ Allow for variable # of tracks in the jet
- ☑ Avoids curse of dimensionality - add more features

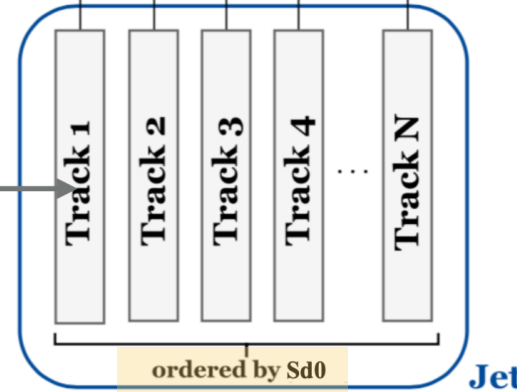
Unrolled RNN



Fully Connected + SoftMax



$S_{d0}$
$S_{z0}$
$P_{T,frac}$
$\Delta R$
hits



20m jets

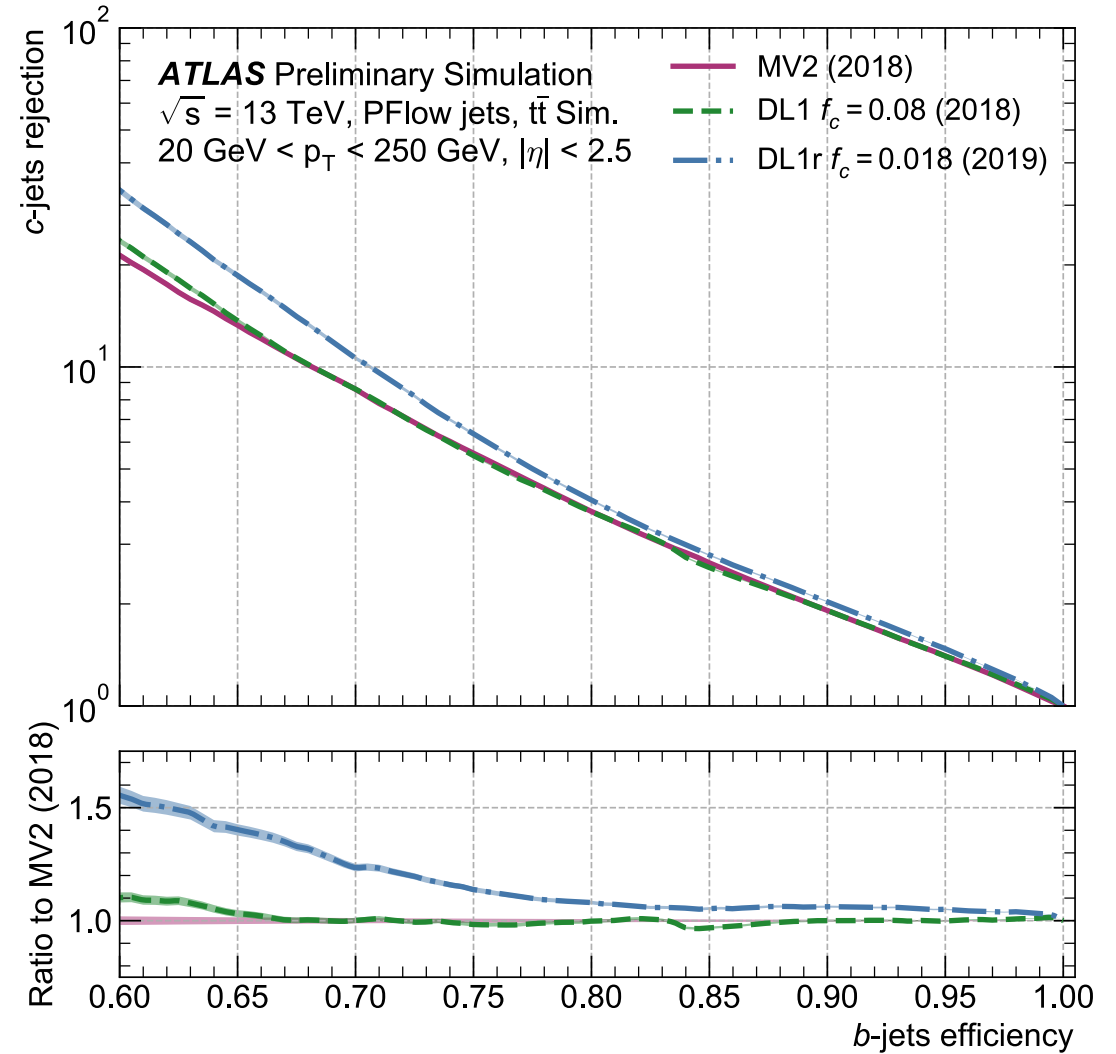
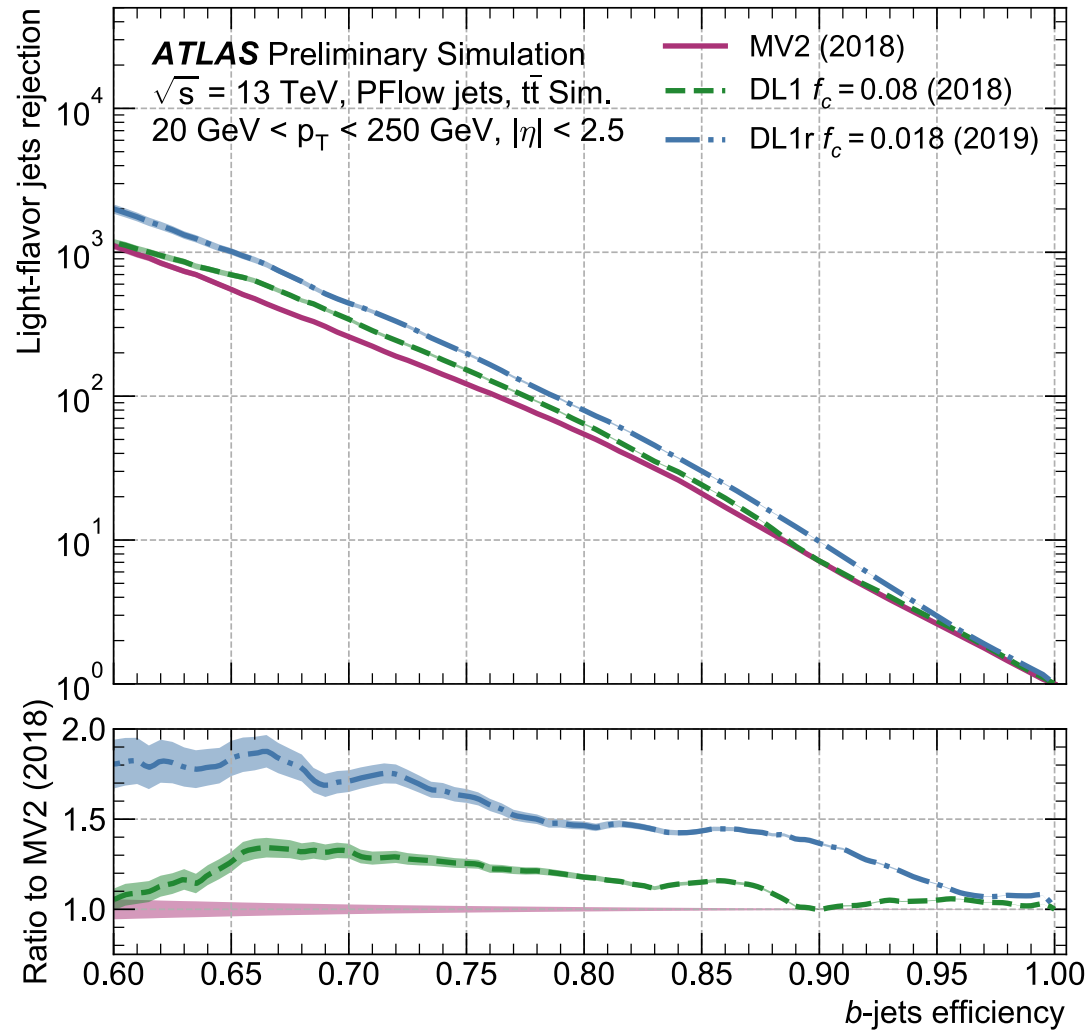
baseline

RNNs + Sets

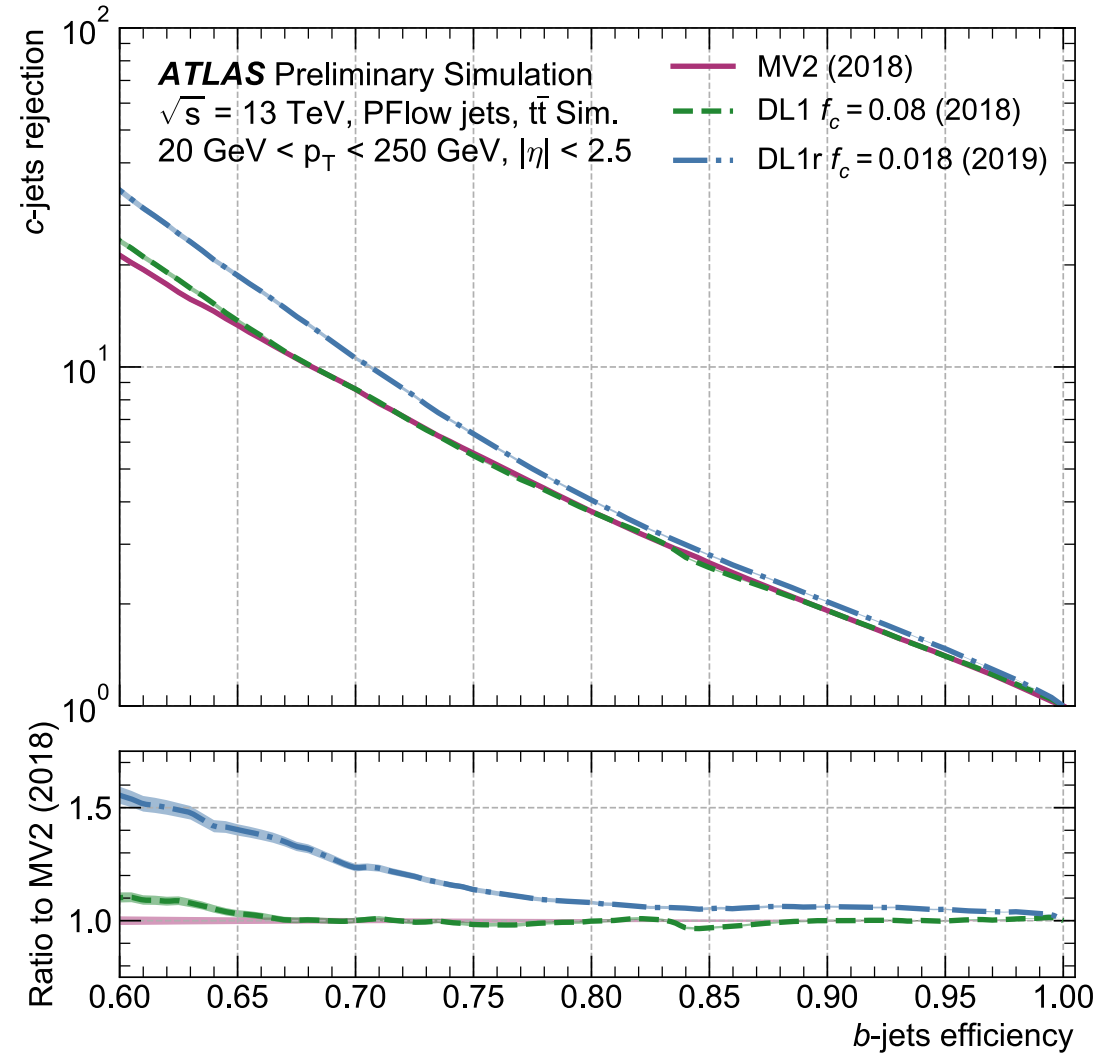
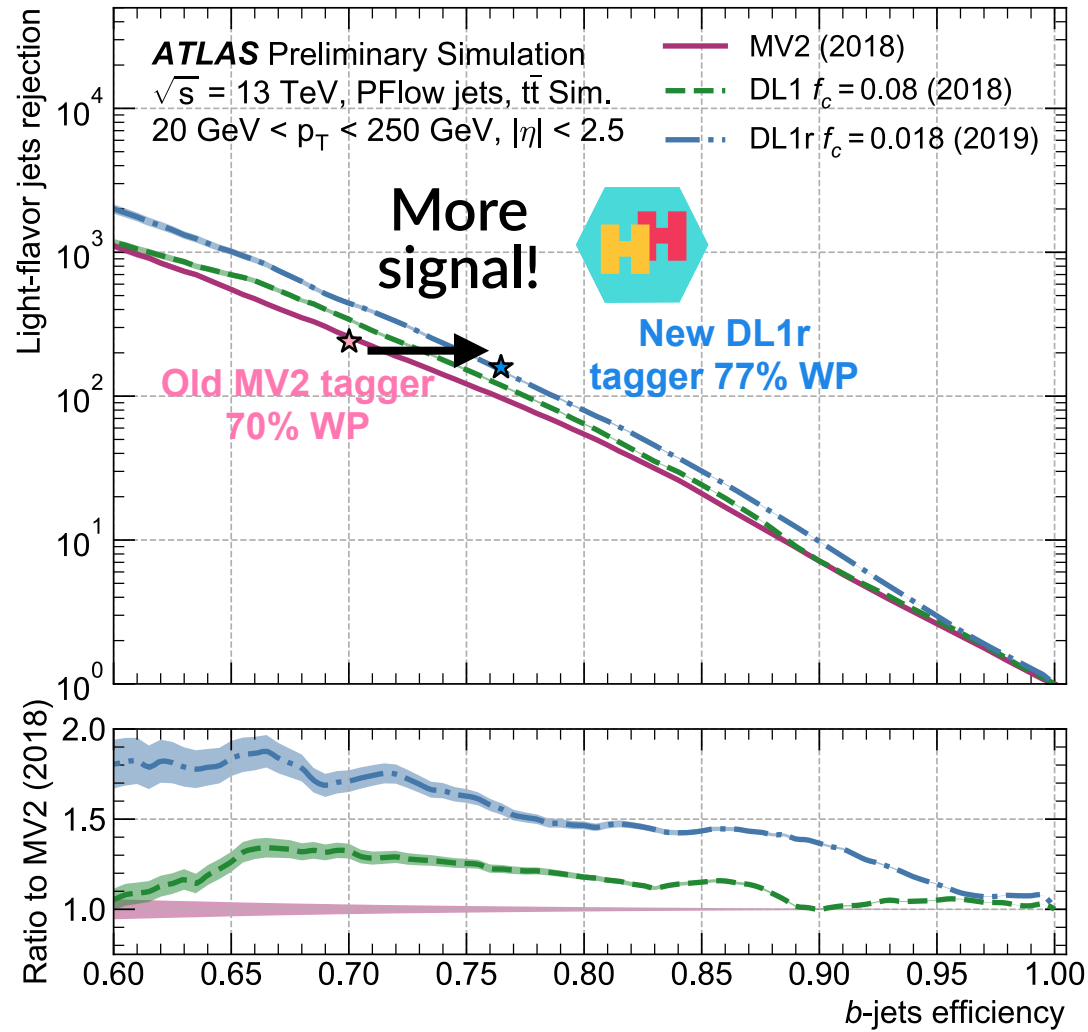
Graphs

Transformers

# b-tagging



# b-tagging



# DIPS

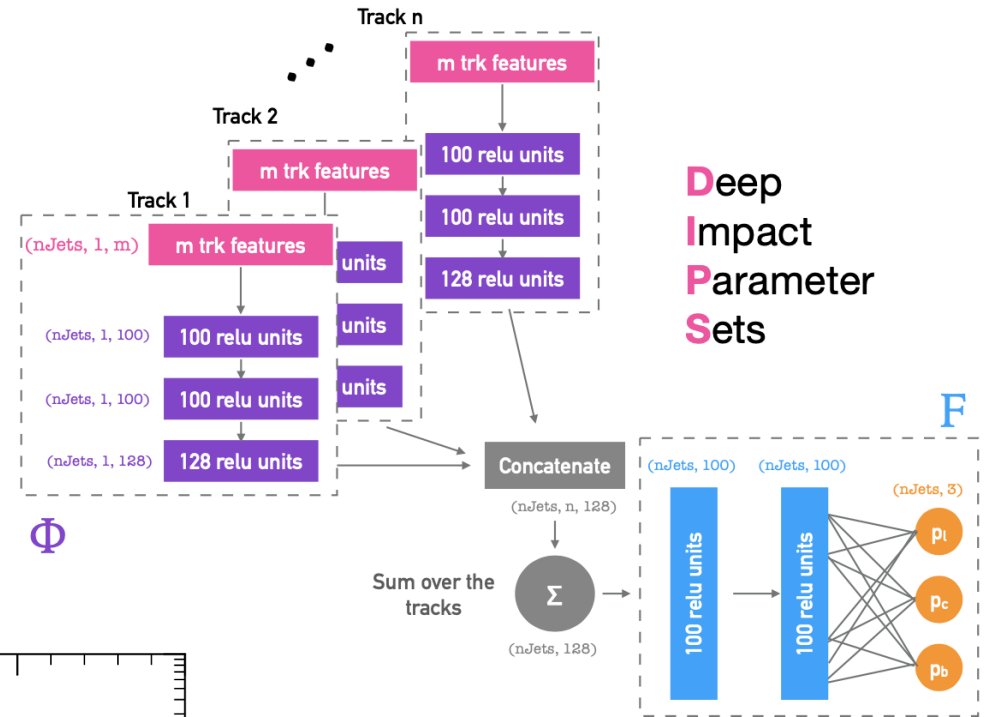
baseline



Model the jet as a ~~sequence~~ **set**

$$O(\{X_1, \dots, X_n\}) = F\left(\sum_{i=1}^n \Phi(X_i)\right)$$

Same inputs as b-tagging RNN



Deep  
Impact  
Parameter  
Sets

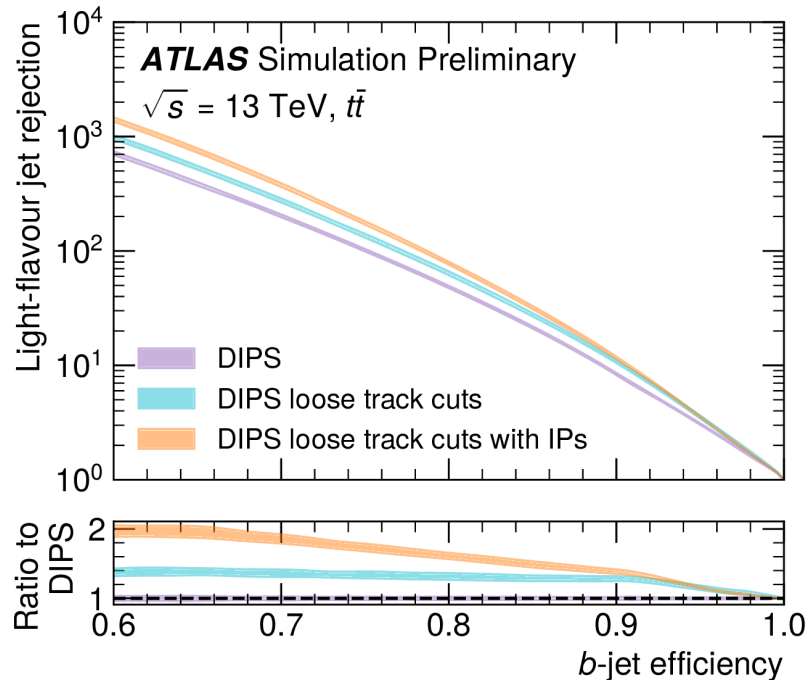
F

RNNs  
+Sets

Graphs



4x speed up in training time c.f. RNNIP



DIPS → DL1d baseline for the Run 3 tagger optimizations

# DeepJet

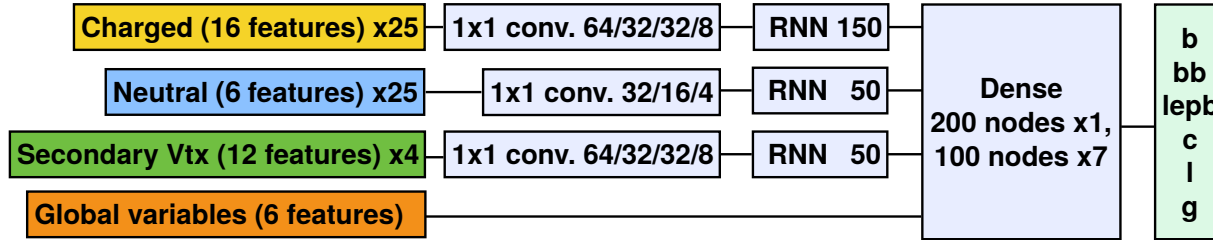
baseline

RNNs  
+Sets

Graphs

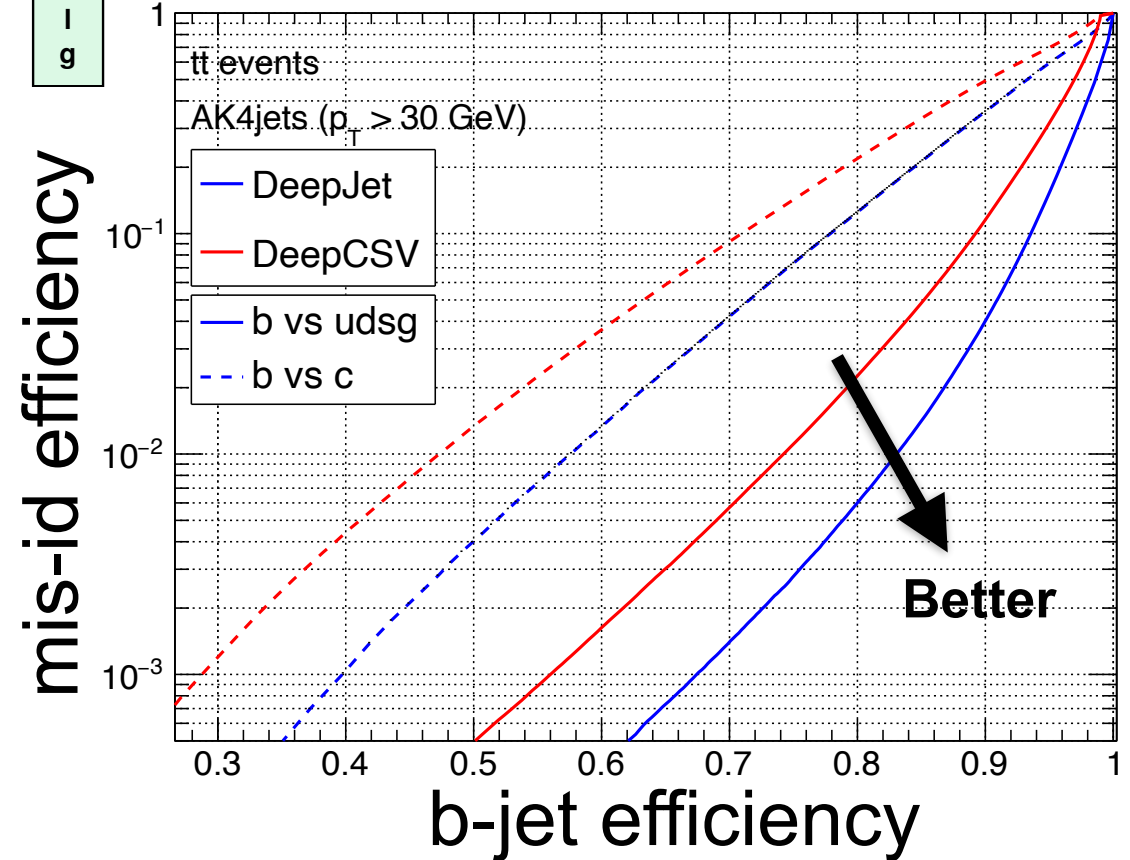
Trans-  
formers

$\Phi$ : per-particle network



Dramatically increased number of inputs  
→ Over 600 in total

100m jets



# DeepJet

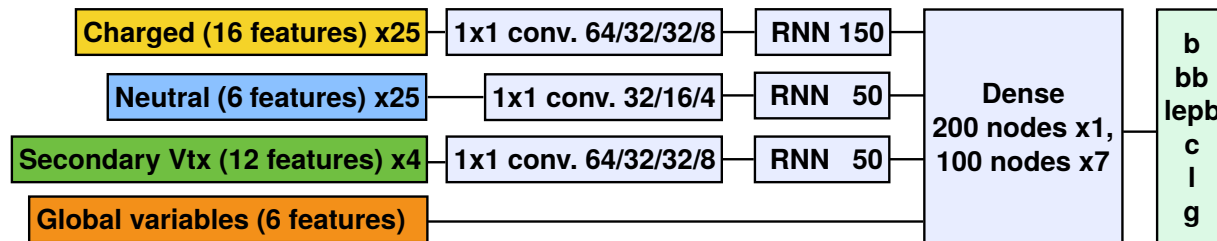
baseline

RNNs  
+Sets

Graphs

Trans-  
formers

$\Phi$ : per-particle network

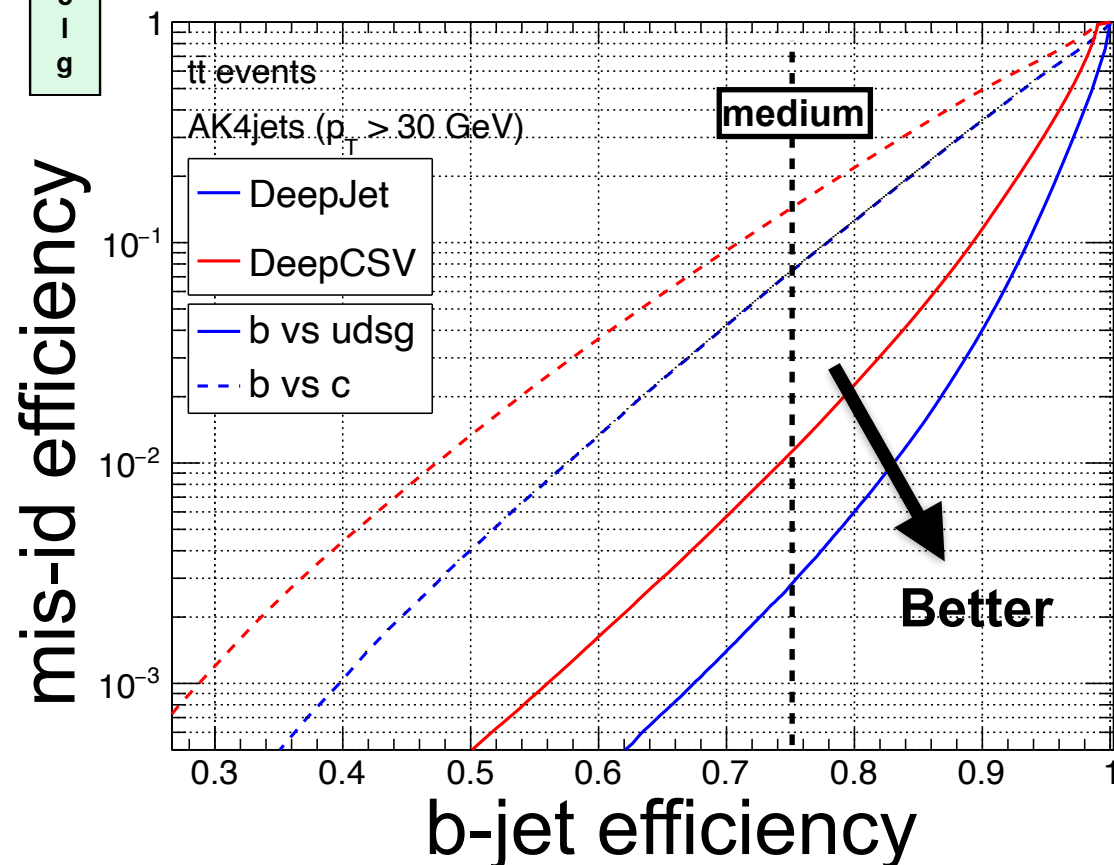


Dramatically increased number of inputs  
 → Over 600 in total

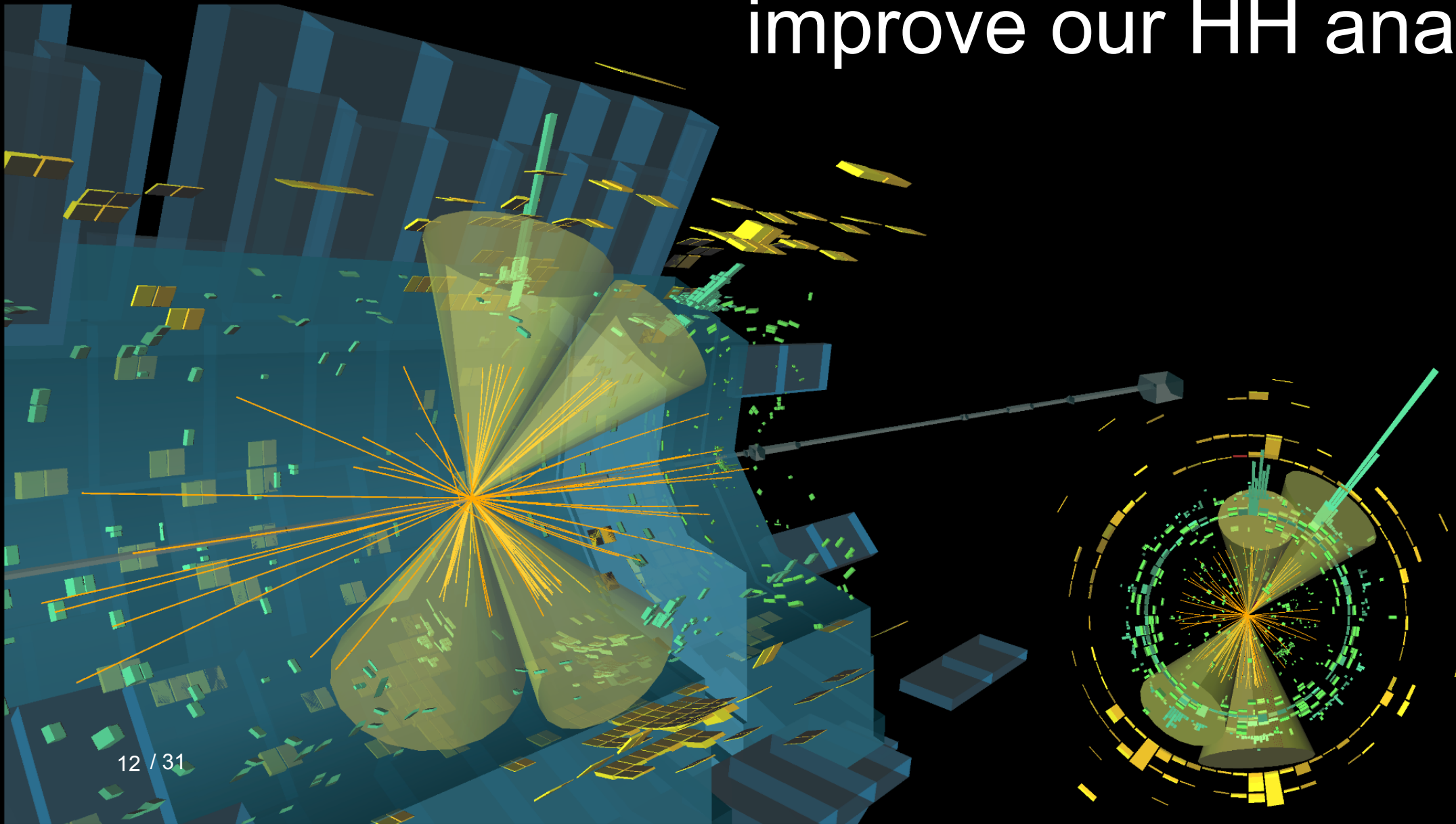
**medium working point (58% eff)**

- ▶ 1.7x improvement in charm rejection
- ▶ 5x improvement in light rejection

100m jets



# How did these **b-tagging** improvements improve our HH analyses?





**ATLAS**  
**CMS**



4b



bbγγ



bbττ



Combination

baseline

RNNs  
+Sets

Graphs

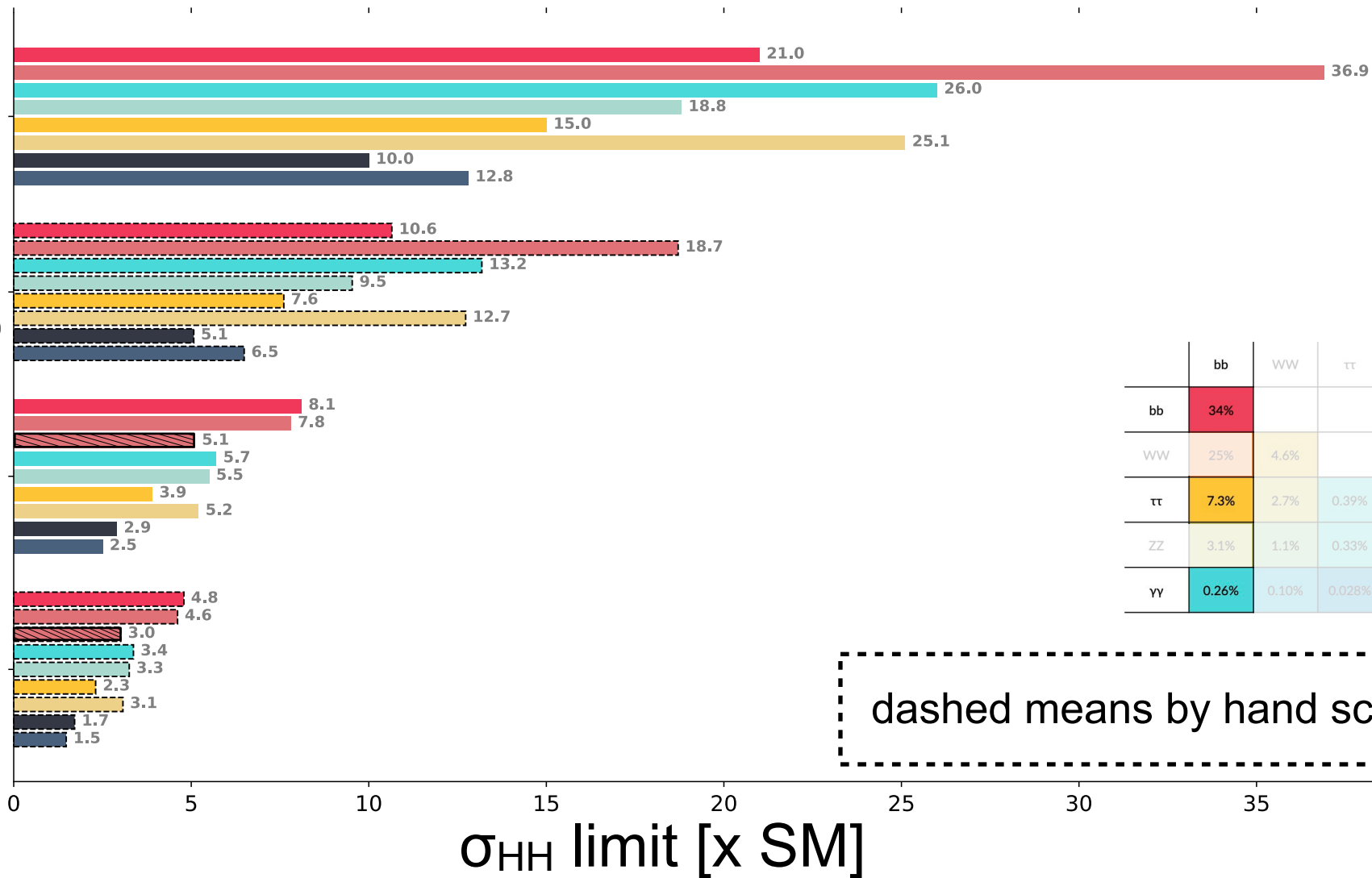
Trans-  
formers

Early Run 2

Full Run 2  
(lumi scaling)

Full Run 2

Run 2+3  
(lumi scaling)



	bb	WW	ττ	ZZ	γγ
bb	34%				
WW	25%	4.6%			
ττ	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
γγ	0.26%	0.10%	0.028%	0.012%	0.0005%

dashed means by hand scaling

\* Note: Some places omitting sub-dominant channels to emphasize ones driving the sensitivity, full results in backup







**ATLAS**  
**CMS**



4b



bbγγ



bbττ



Combination

baseline

RNNs  
+Sets

Graphs

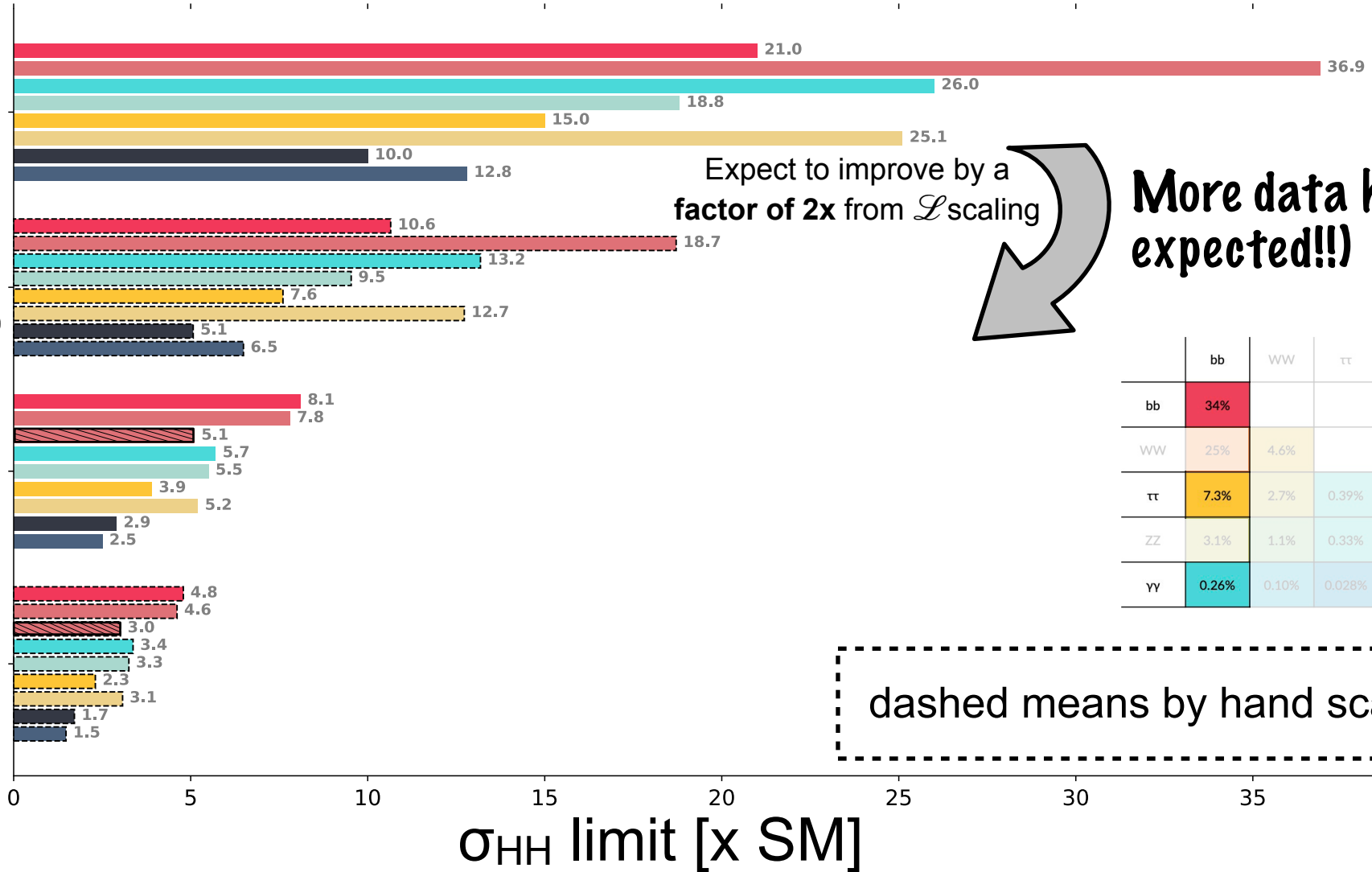
Trans-  
formers

Early Run 2

Full Run 2  
(lumi scaling)

Full Run 2

Run 2+3  
(lumi scaling)



Expect to improve by a factor of 2x from  $\mathcal{L}$  scaling

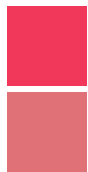
**More data helps (as expected!!)**

	bb	WW	ττ	ZZ	γγ
bb	34%				
WW	25%	4.6%			
ττ	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
γγ	0.26%	0.10%	0.028%	0.012%	0.0005%

dashed means by hand scaling



**ATLAS**  
**CMS**



4b



bbγγ



bbττ



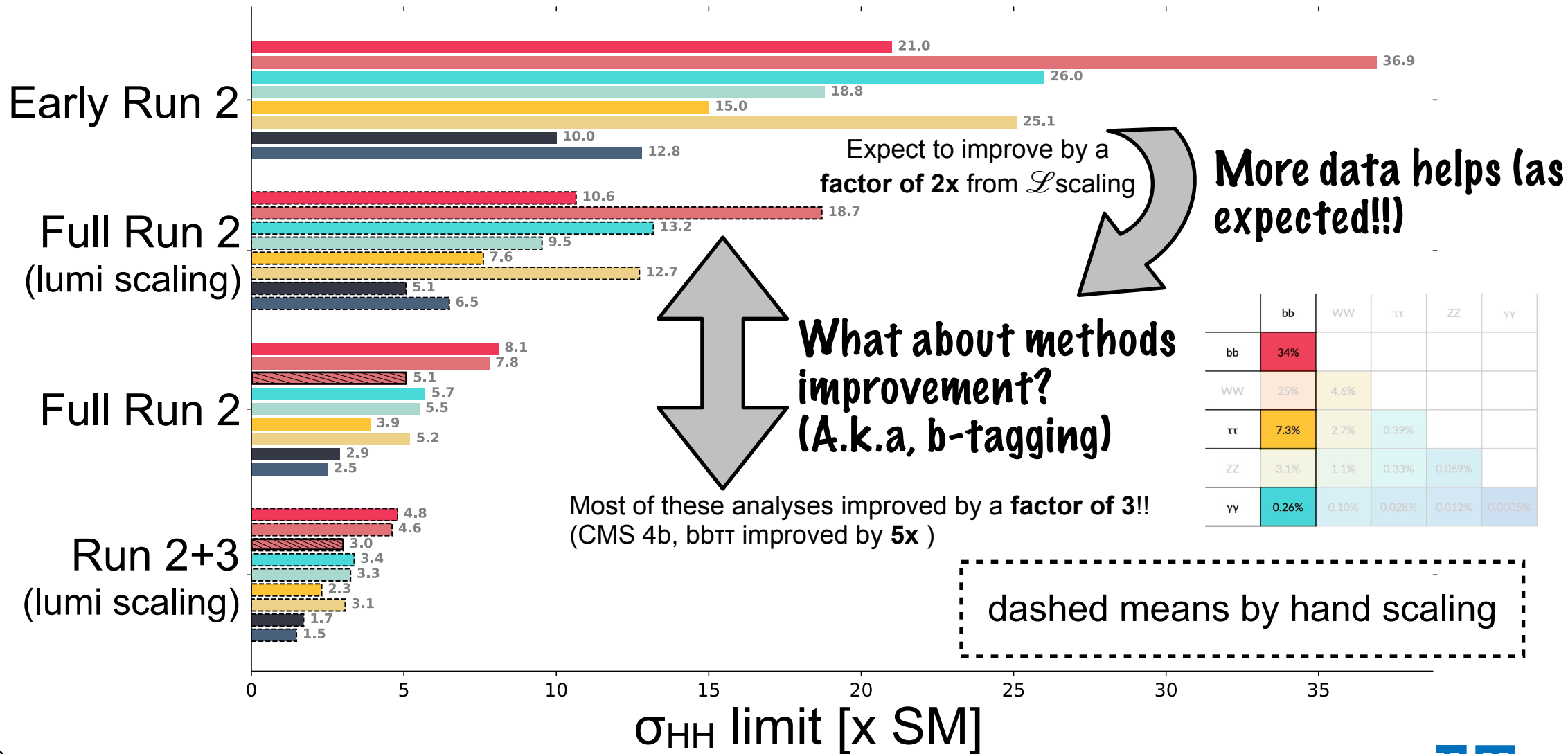
Combination

baseline

RNNs  
+Sets

Graphs

Trans-  
formers





**ATLAS**  
**CMS**



4b



bbγγ



bbττ



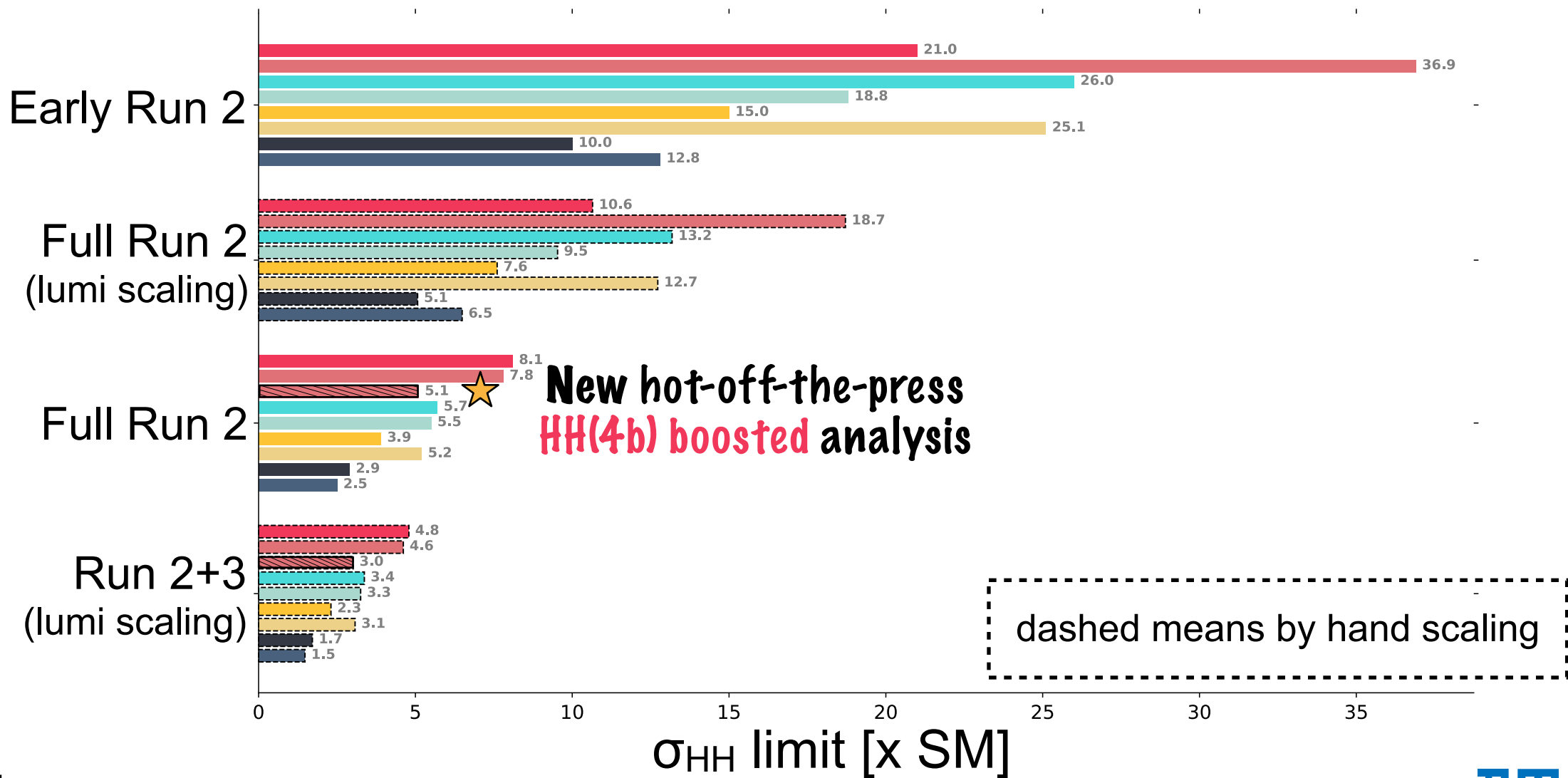
Combination

baseline

RNNs  
+Sets

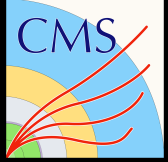
Graphs

Trans-  
formers



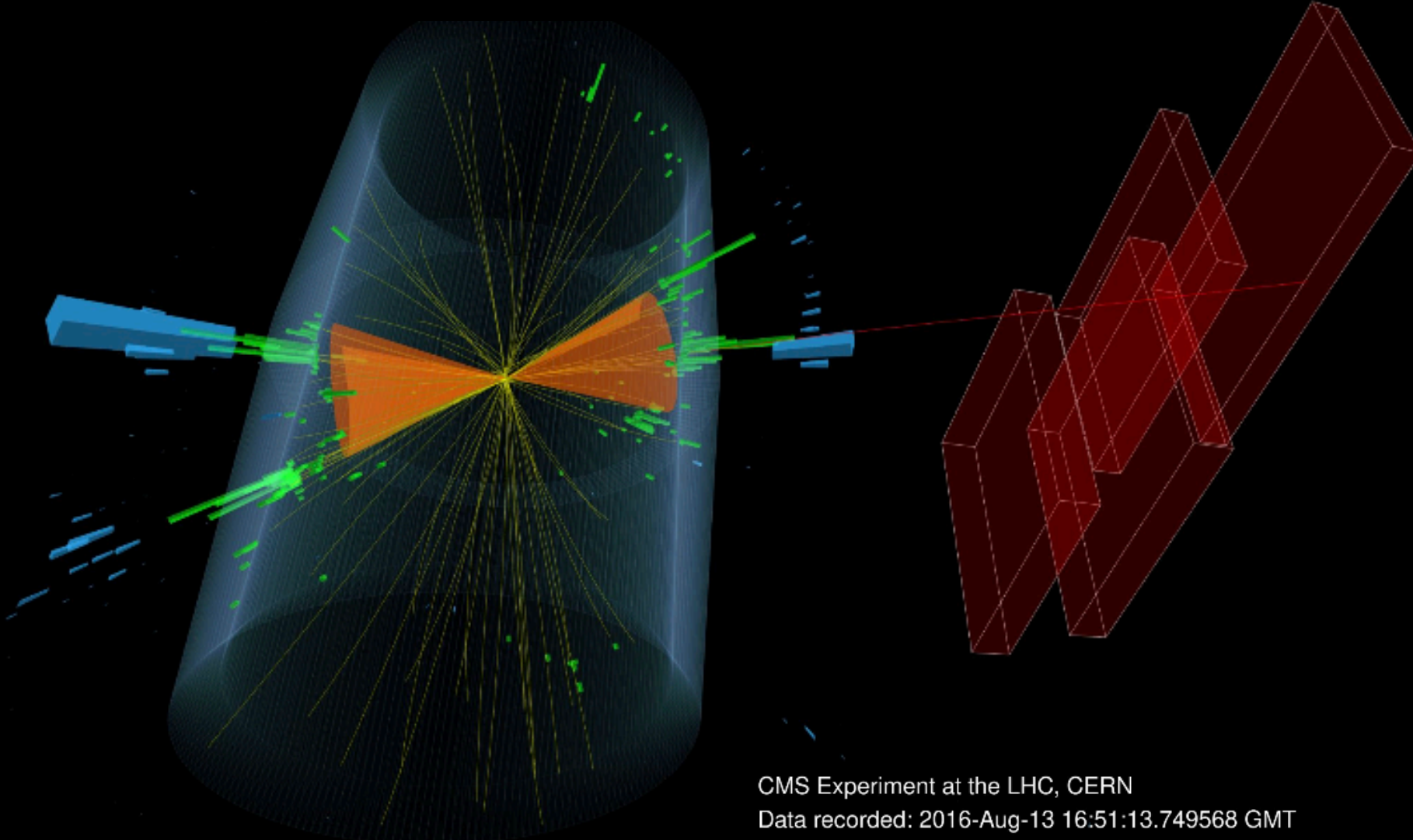
\* Note: Some places omitting sub-dominant channels to emphasize ones driving the sensitivity, full results in backup





# Boosted analyses

Highlighting CMS boosted  
H(bb)-tagger, ParticleNet



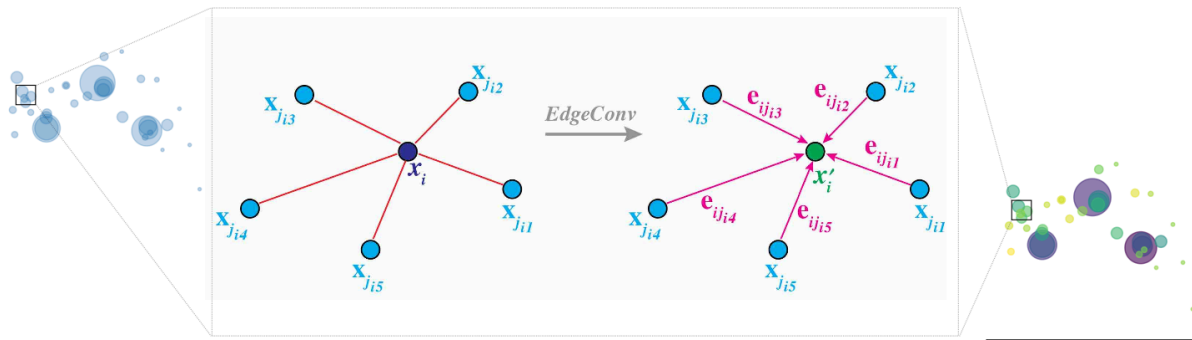
CMS Experiment at the LHC, CERN  
Data recorded: 2016-Aug-13 16:51:13.749568 GMT  
Run / Event / LS: 278803 / 465417690 / 259



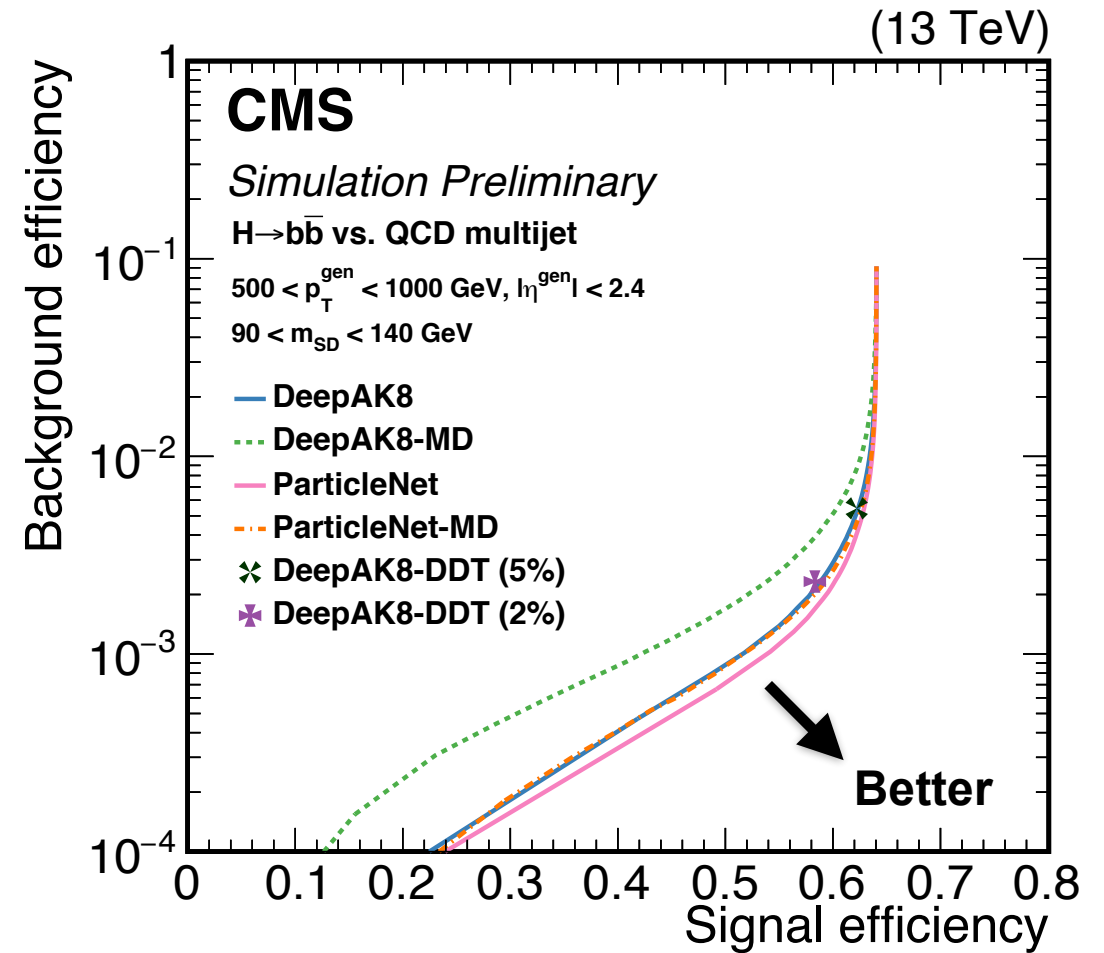


# ParticleNet

## Dynamic Graph Convolutional NN k nearest neighbors



The graph connectivity can *update* in through successive EdgeConv layers



baseline

RNNs  
 +Sets

Graphs

Trans-  
 formers



# ParticleNet Analyses

baseline

RNNs  
+Sets

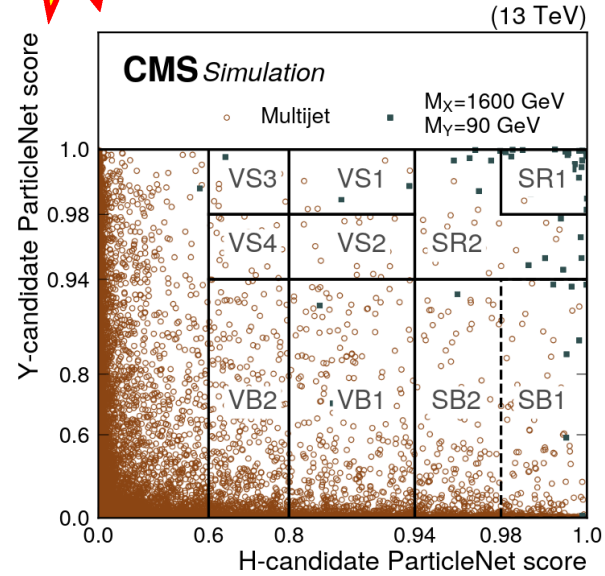
Graphs

Trans-  
formers

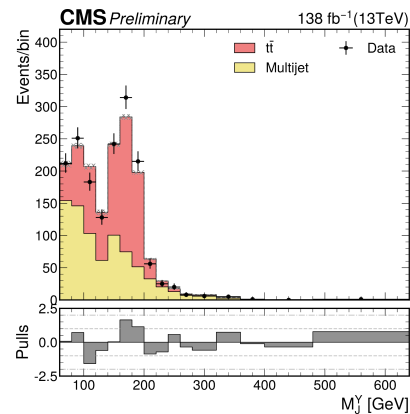


**X→YH→4b**

Tagger is so powerful at background rejection that the bkg estimate becomes simpler: a (few) normalization factors



Epic QCD rejection





# ParticleNet Analyses

baseline

RNNs  
+Sets

Graphs

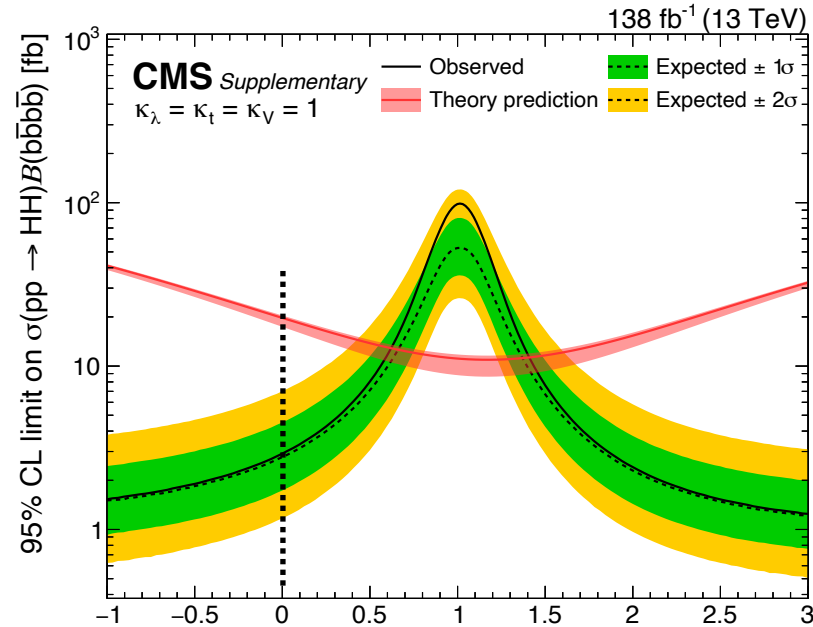
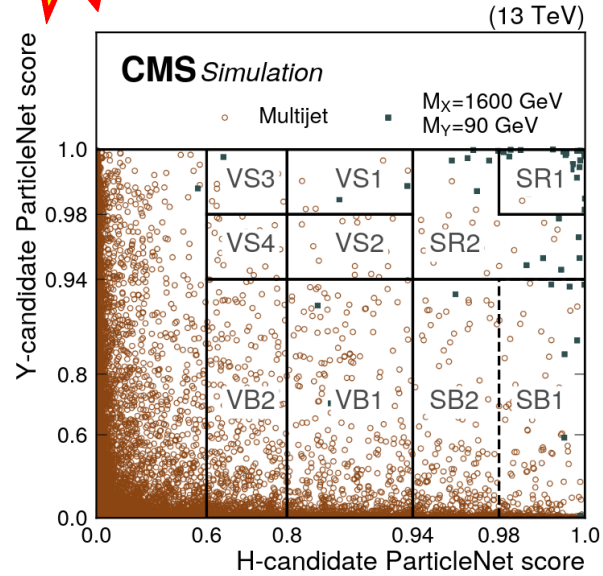
Trans-  
formers



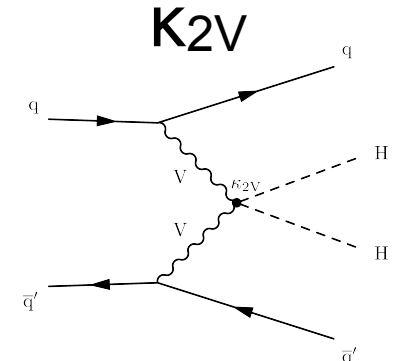
**X→YH→4b**

Tagger is so powerful at background rejection that the bkg estimate becomes simpler: a (few) normalization factors

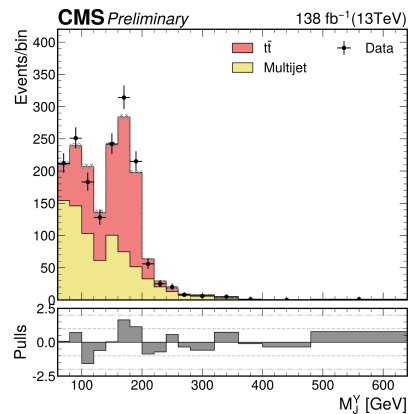
## VBF HH→4b



First analysis  
excluding  $\kappa_{2V} = 0$



Epic QCD rejection





# ParticleNet Analyses

baseline

RNNs  
+Sets

Graphs

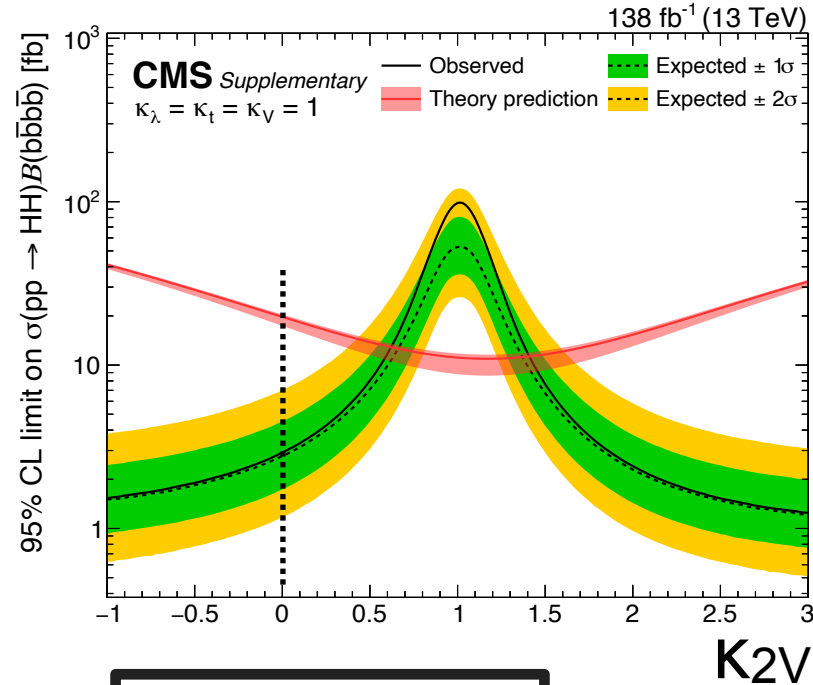
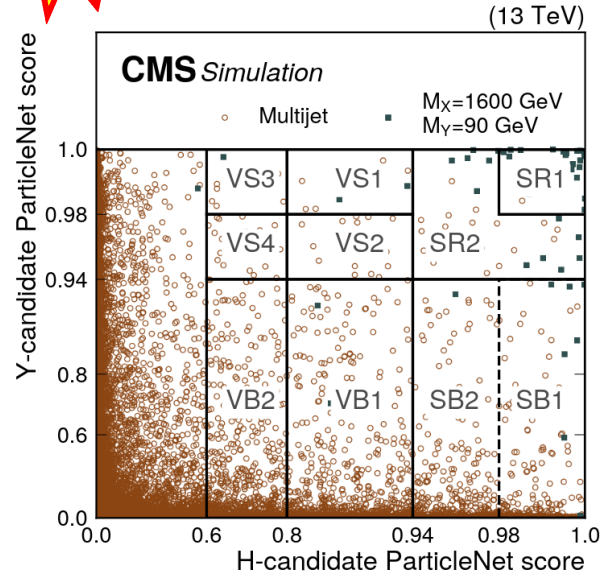
Transformers



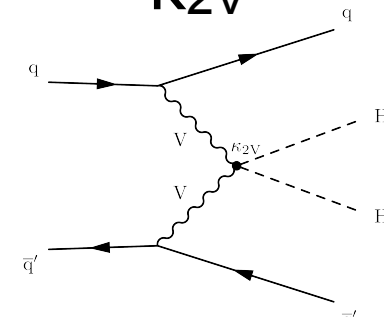
**X → YH → 4b**

Tagger is so powerful at background rejection that the bkg estimate becomes simpler: a (few) normalization factors

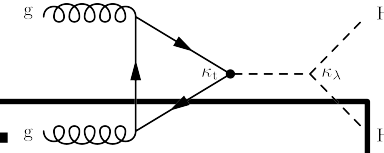
**VBF HH → 4b**



First analysis excluding  $k_{2V} = 0$



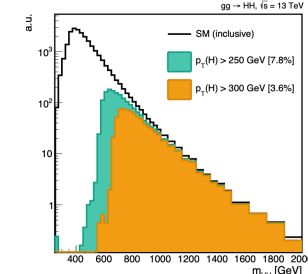
**ggF HH → 4b**



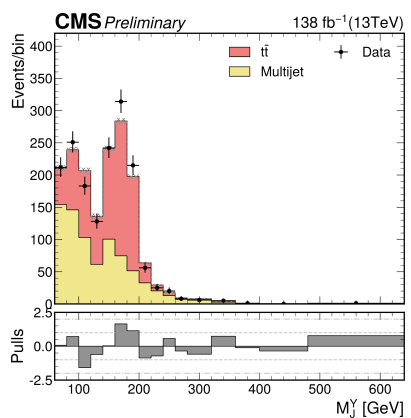
**Boosted:**  
 Obs (exp): 9.9 (5.1)

**Resolved:**  
 Obs (exp): 5.4 (8.1)  
 Obs (exp): 3.8 (7.8)

Note: only **3.6%** of the SM HH signal is boosted (Luca's slides)



Epic QCD rejection





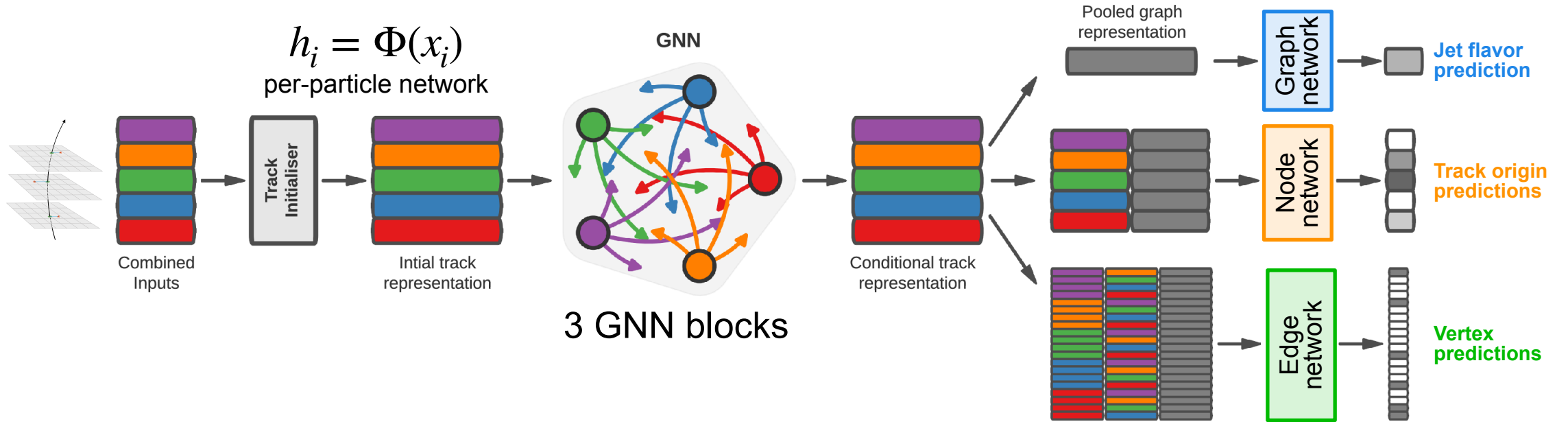
# GN1: architecture

baseline

RNNs  
+Sets

Graphs

Trans-  
formers



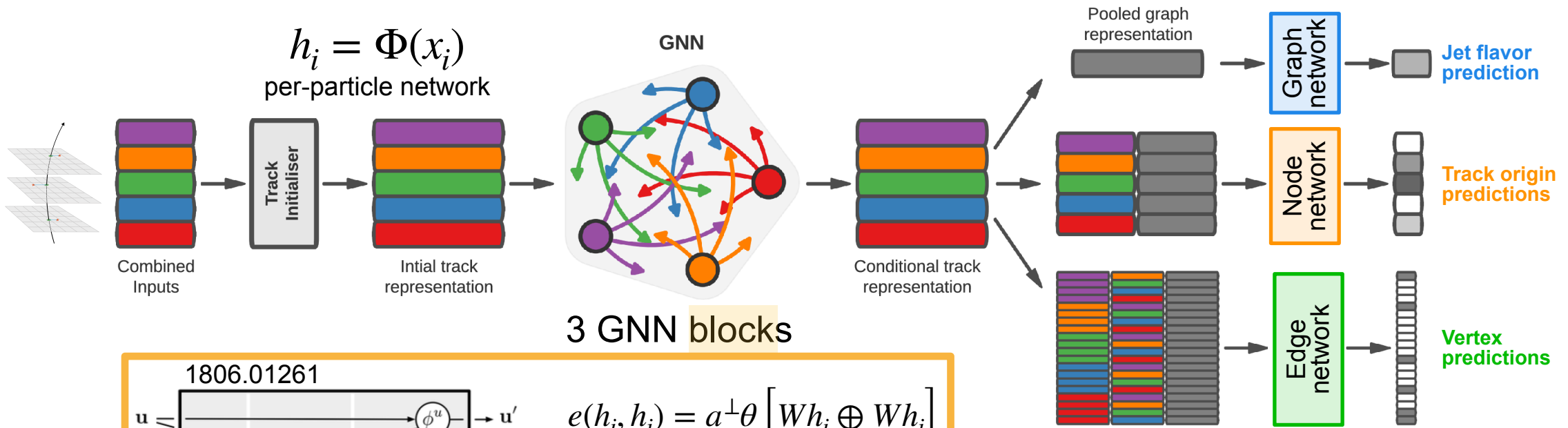
# GN1: architecture

baseline

RNNs  
+Sets

Graphs

Trans-  
formers



1806.01261

The detailed diagram shows the internal structure of a GNN block. It is divided into three sections: 'Edge block', 'Node block', and 'Global block'. The 'Edge block' takes inputs  $u$  and  $v$  and produces  $E$ . The 'Node block' takes  $u$  and  $v$  and produces  $u'$  and  $v'$ . The 'Global block' takes  $E$  and produces  $E'$ . The diagram includes functions  $\phi^u$ ,  $\phi^v$ , and  $\phi^e$ , and variables  $\rho^{v \rightarrow u}$ ,  $\rho^{e \rightarrow v}$ , and  $\rho^{e \rightarrow u}$ .

$$e(h_i, h_j) = a^\perp \theta [Wh_i \oplus Wh_j]$$

$$a_{ij} = \text{Softmax}_j [e(h_i, h_j)]$$

$$h'_i = \sigma \left[ \sum_{j \in \mathcal{N}_i} a_{ij} \cdot Wh_j \right]$$

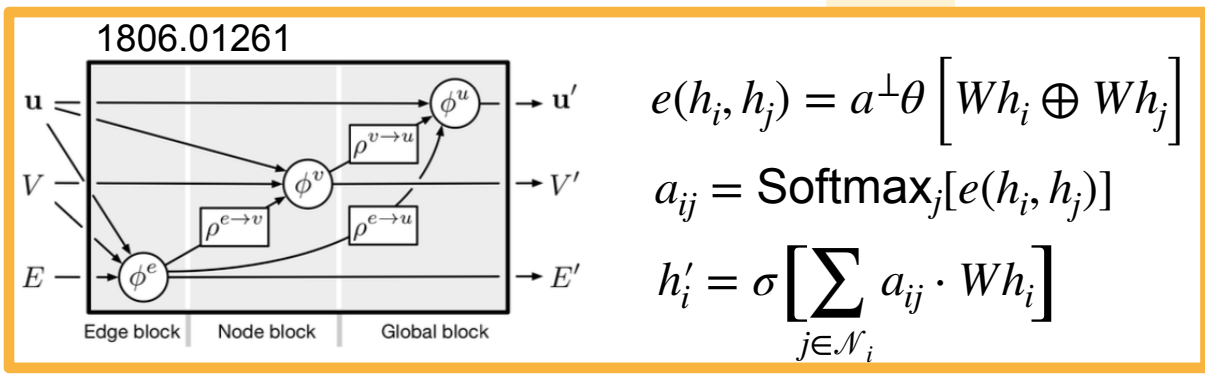
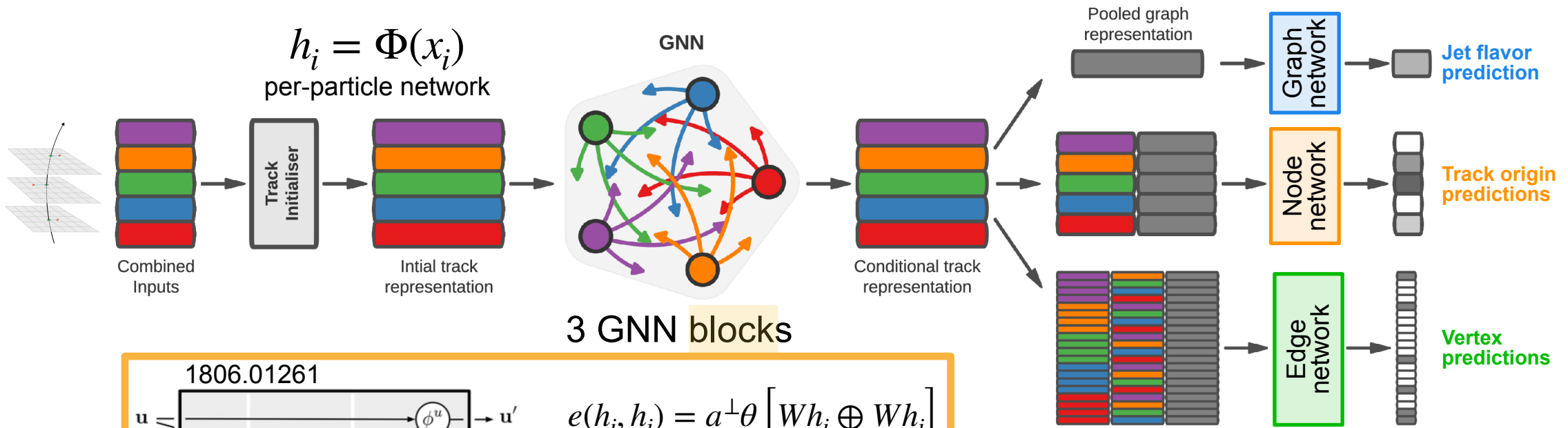
# GN1: architecture

baseline

RNNs  
+Sets

Graphs

Trans-  
formers

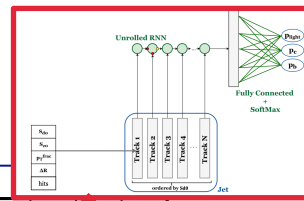


**Multi-modal learning!**

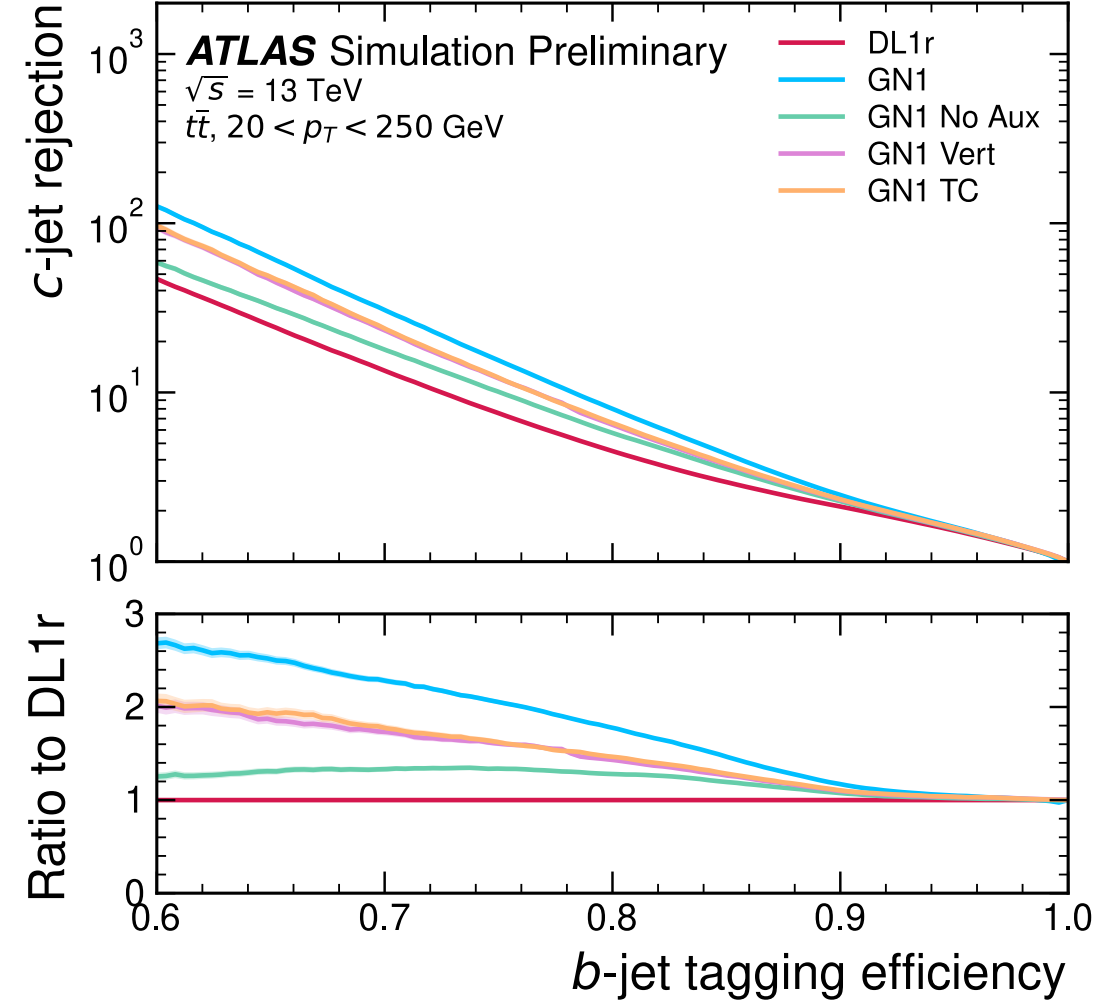
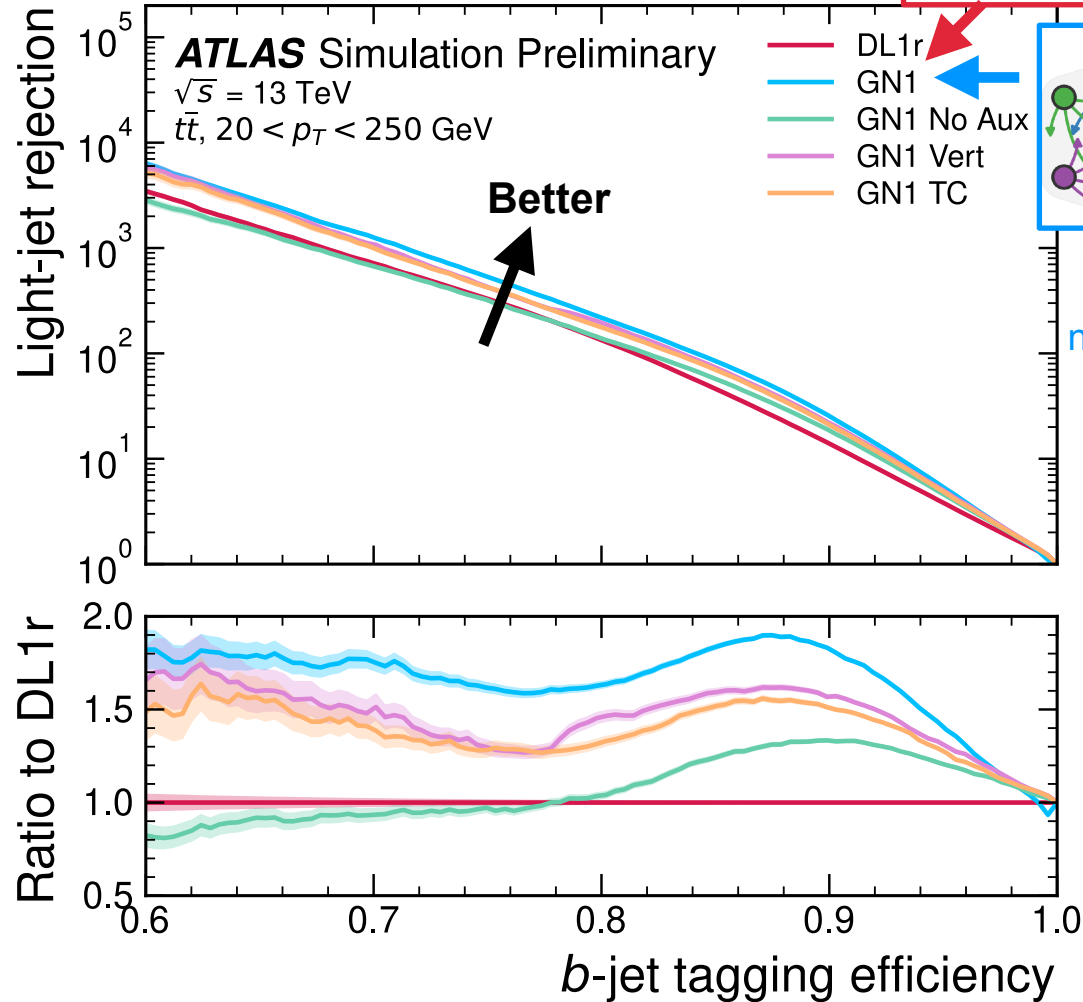
$$\mathcal{L}_{tot} = \mathcal{L}_{jet} + \alpha \mathcal{L}_{trk} + \beta \mathcal{L}_{vtx}$$

$\alpha = 0.5$                        $\beta = 1.5$

# GN1: auxiliary tasks

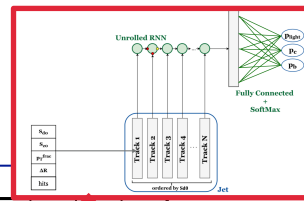


Recurrent  
Neural  
Network

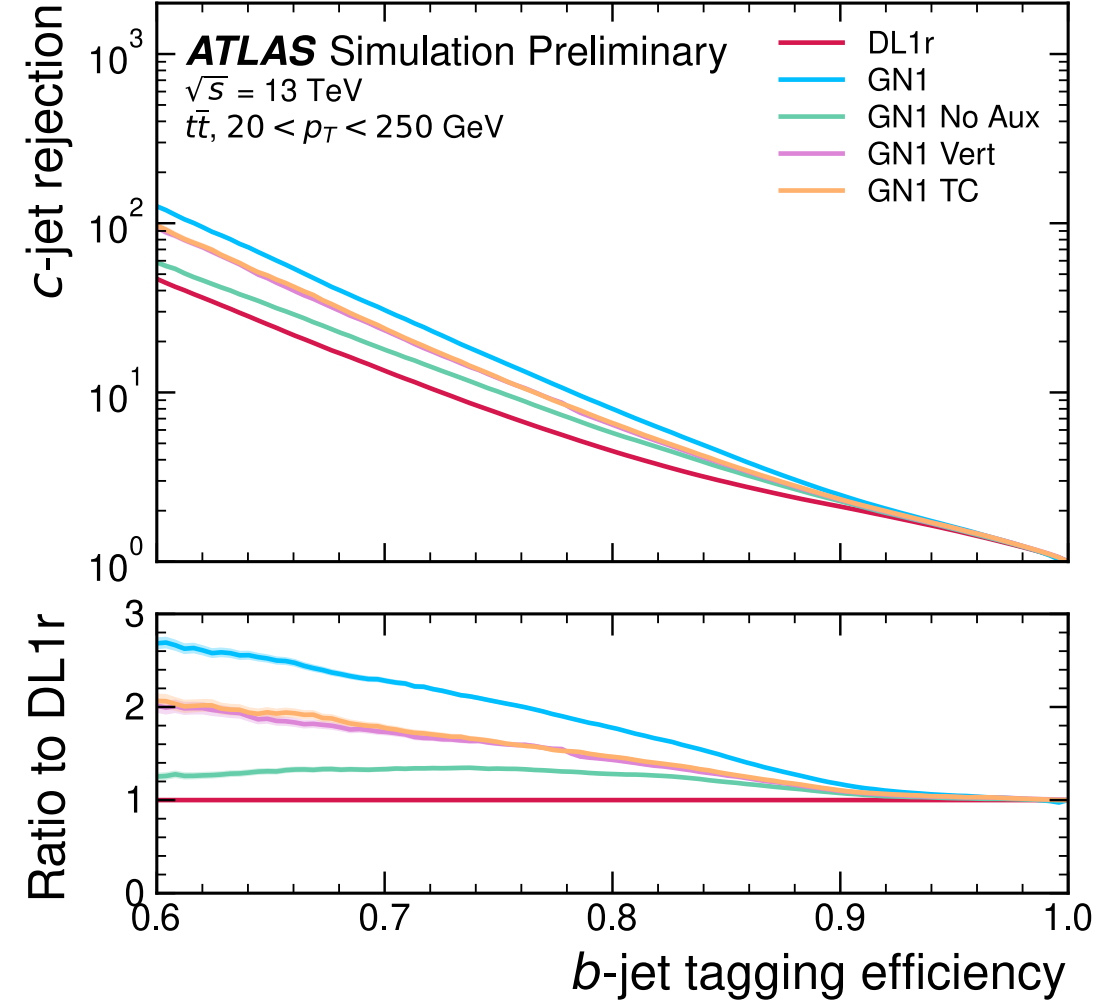
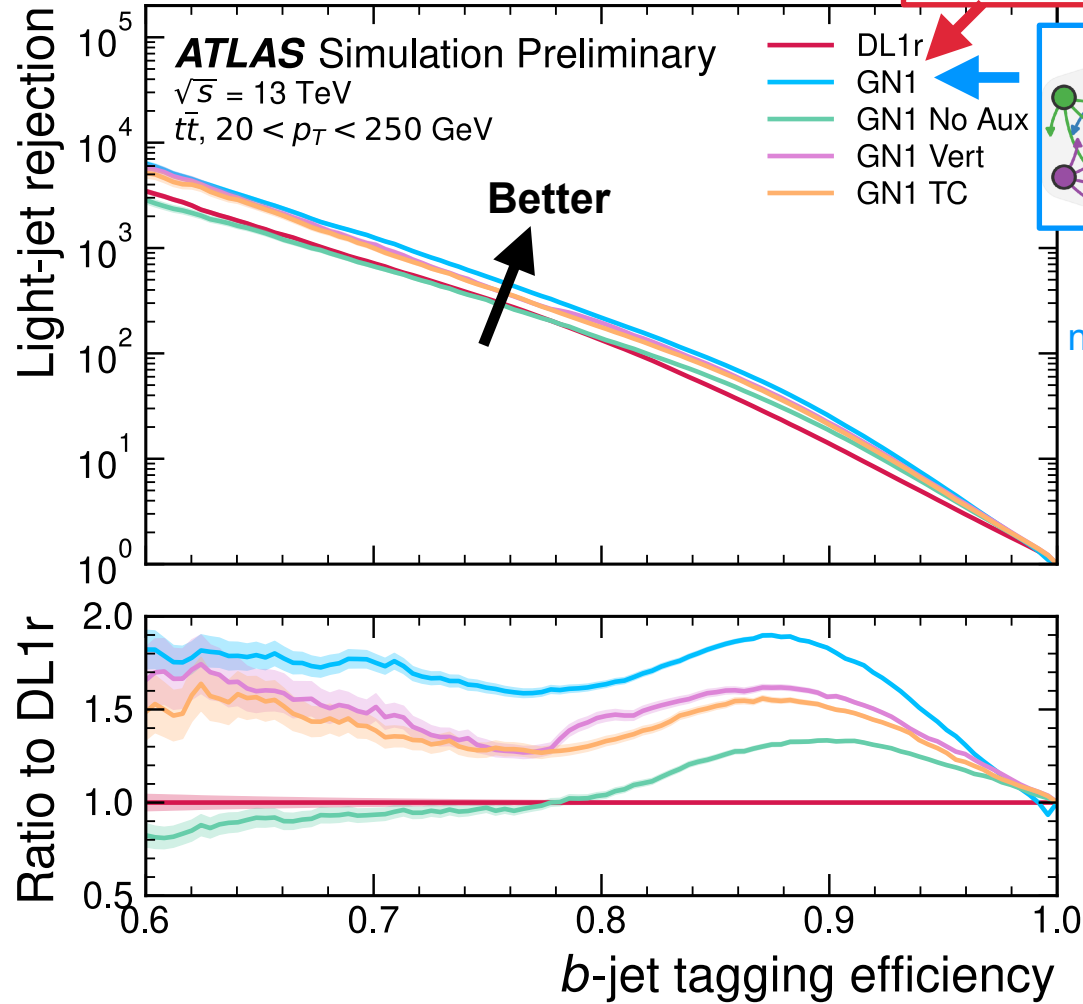


**GN1: 2x increase in performance for light and charm rejection!**

# GN1: auxiliary tasks



Recurrent  
Neural  
Network

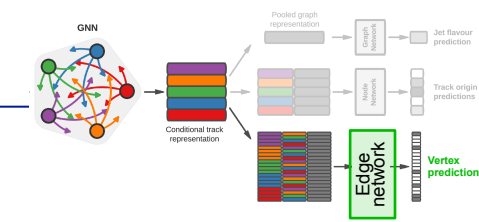


**GN1: 2x increase in performance** for light and charm rejection!

Gain coming from the **vertex finding** and **track classification** aux tasks.

# GN1: vertex finding + track classification

Group tracks together if they have a **pair-wise compatibility** > 0.5

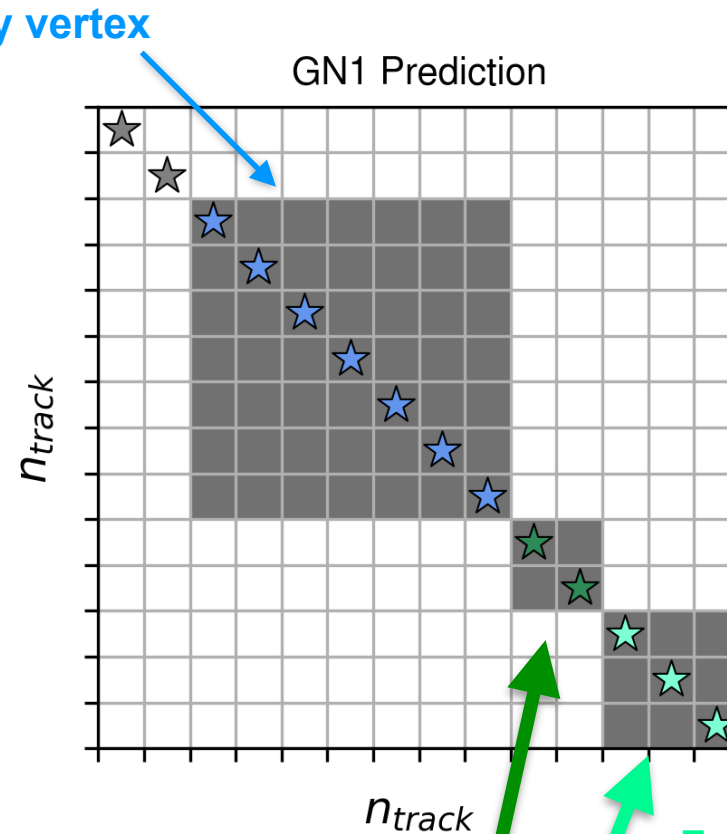
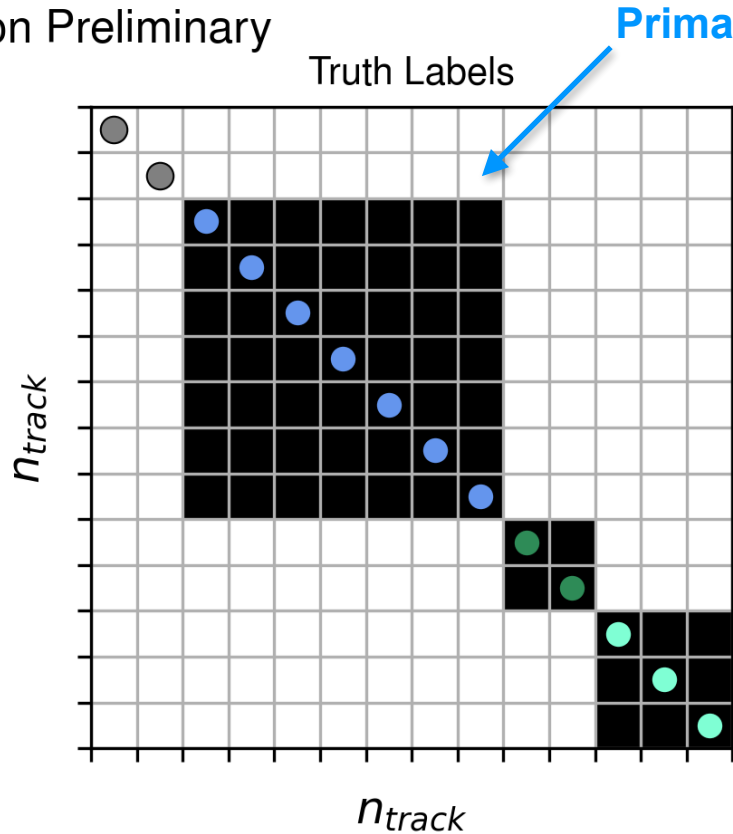


**ATLAS** Simulation Preliminary

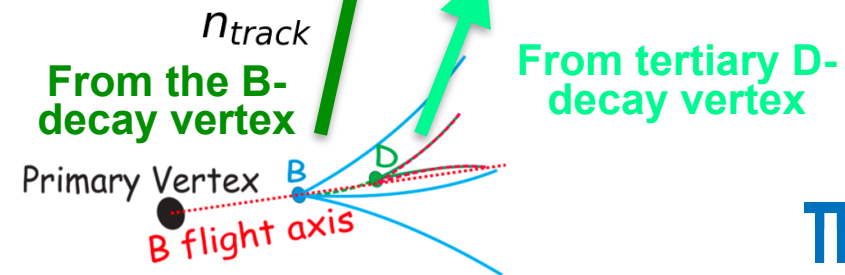
$\sqrt{s} = 13$  TeV  
 $t\bar{t}$  jets

**Truth  $b$ -jet**  
 $p_T = 134.1$  GeV

$p_b = 0.995$   
 $p_c = 0.005$   
 $p_u = 0.000$



- Truth
- ★ Predicted
- Pileup
- Fake
- Primary
- FromB
- FromBC
- FromC
- FromTau
- OtherSecondary



# For small-R jets.. in the trigger

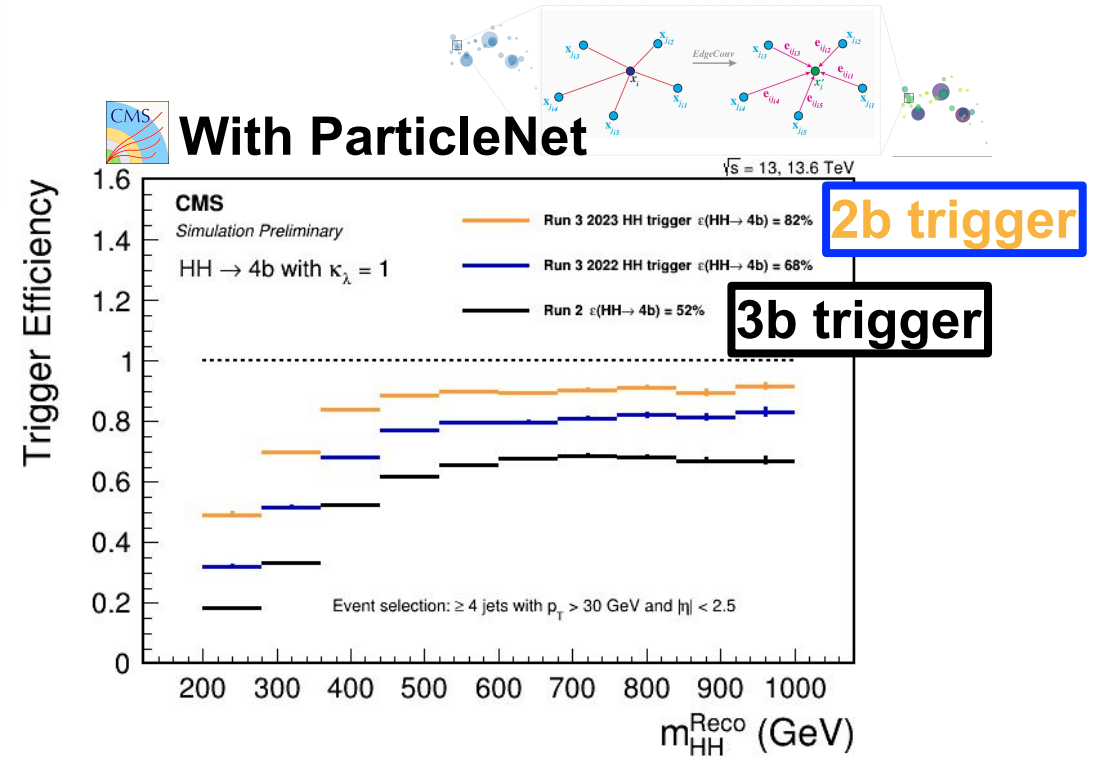
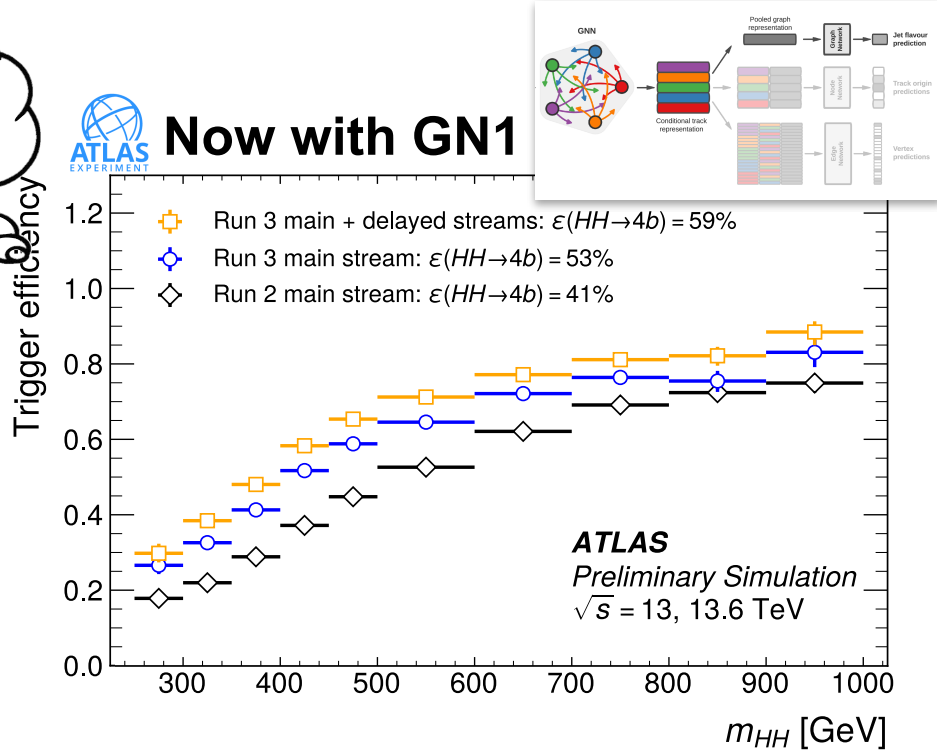
baseline

RNNs  
+Sets

Graphs

Trans-  
formers

Recall from Petar's [talk](#)



# For small-R jets.. in the trigger

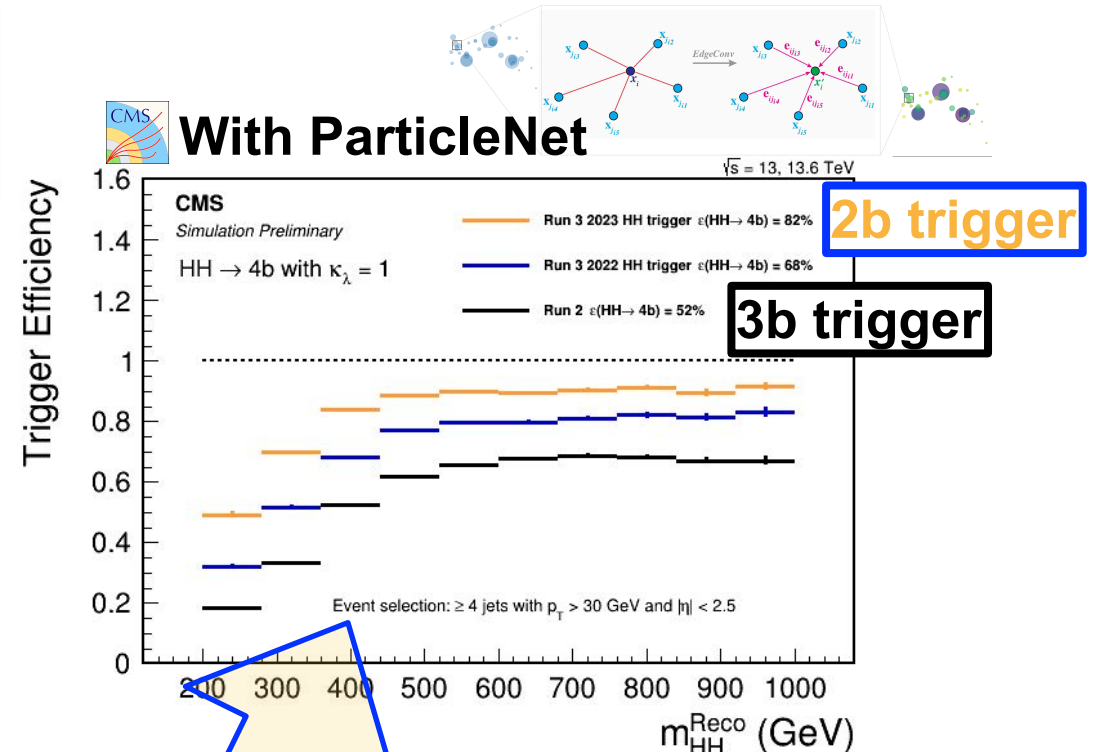
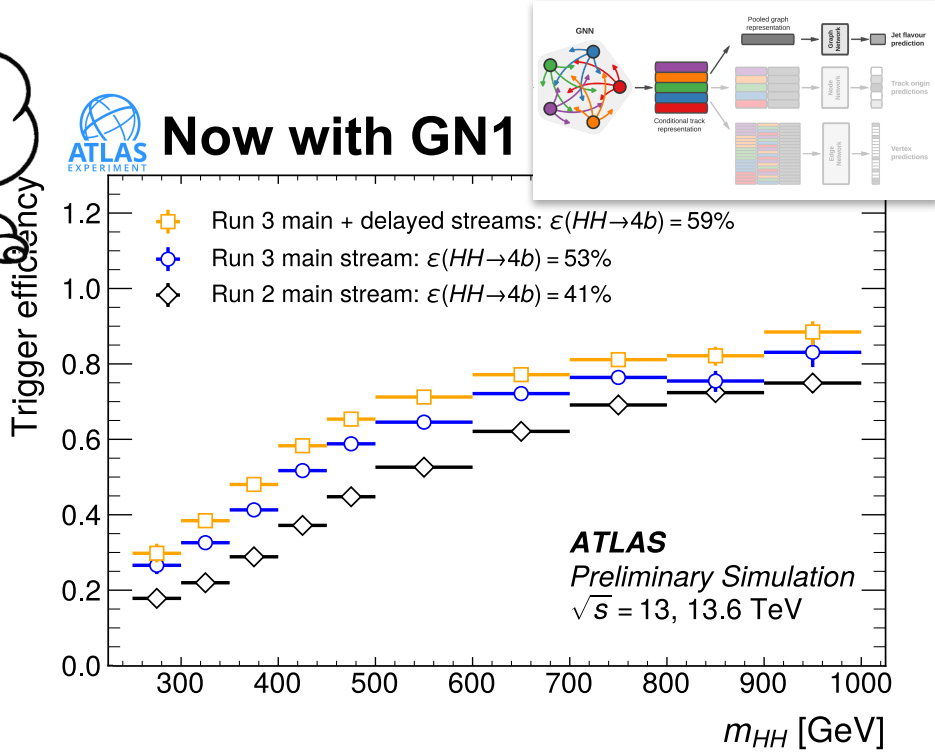
baseline

RNNs  
+Sets

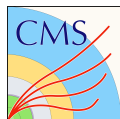
Graphs

Trans-  
formers

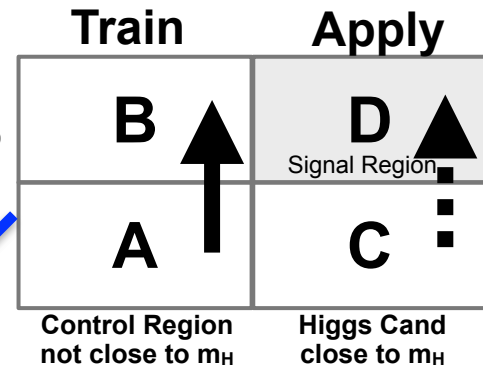
Recall from Petar's [talk](#)



Background estimation with BDT



4b  
2b 3b



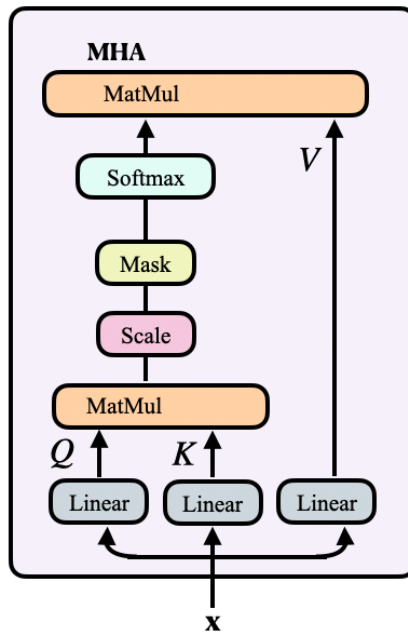
New trigger will reduce key background systematic from the limited **3b Control Region** statistics



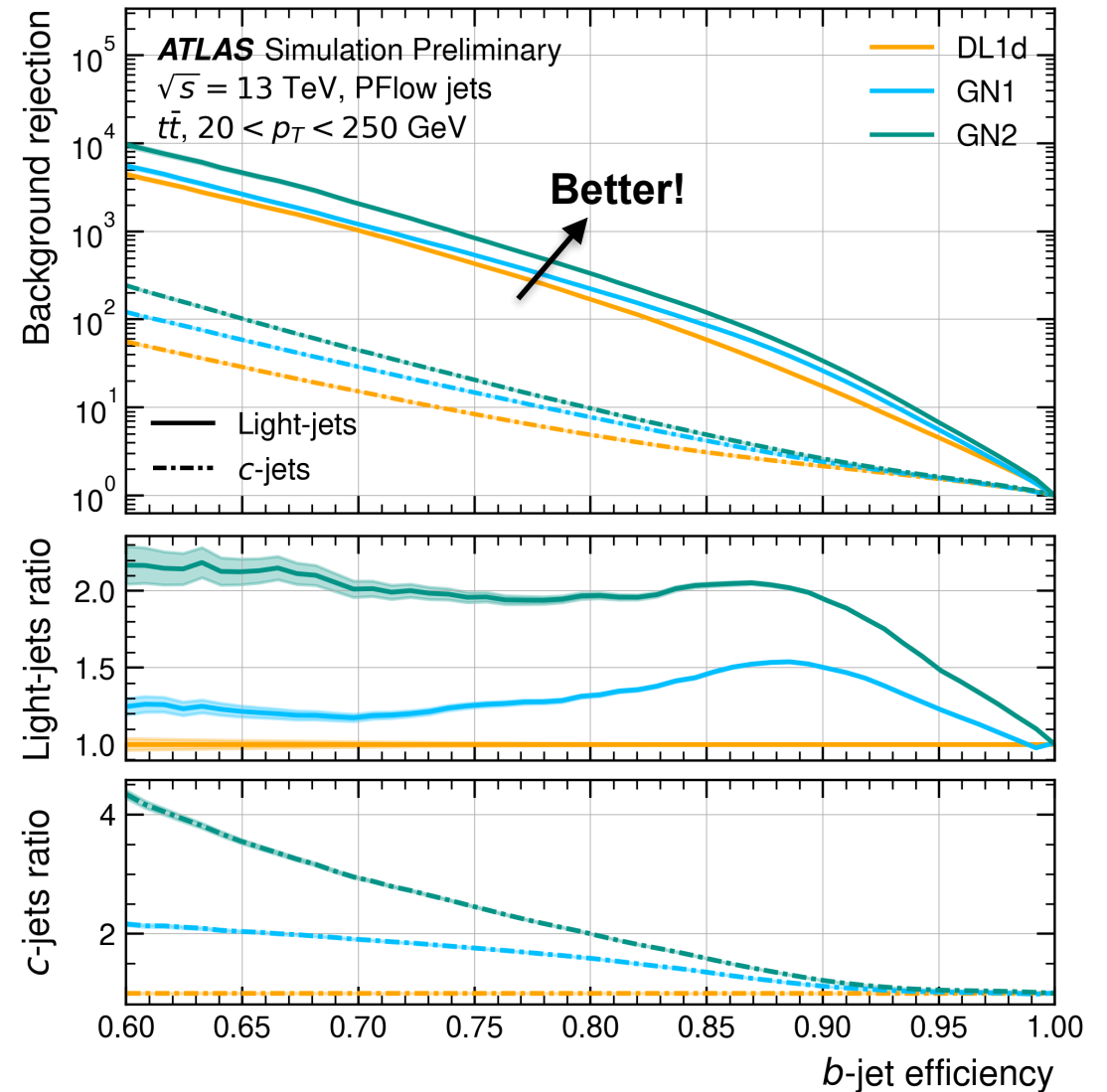
# GN2: $t\bar{t}$ eval

Transformer special type of GNN... but much easier to train!

Attention is All You Need: 1706.03762



Ex: chatGPT



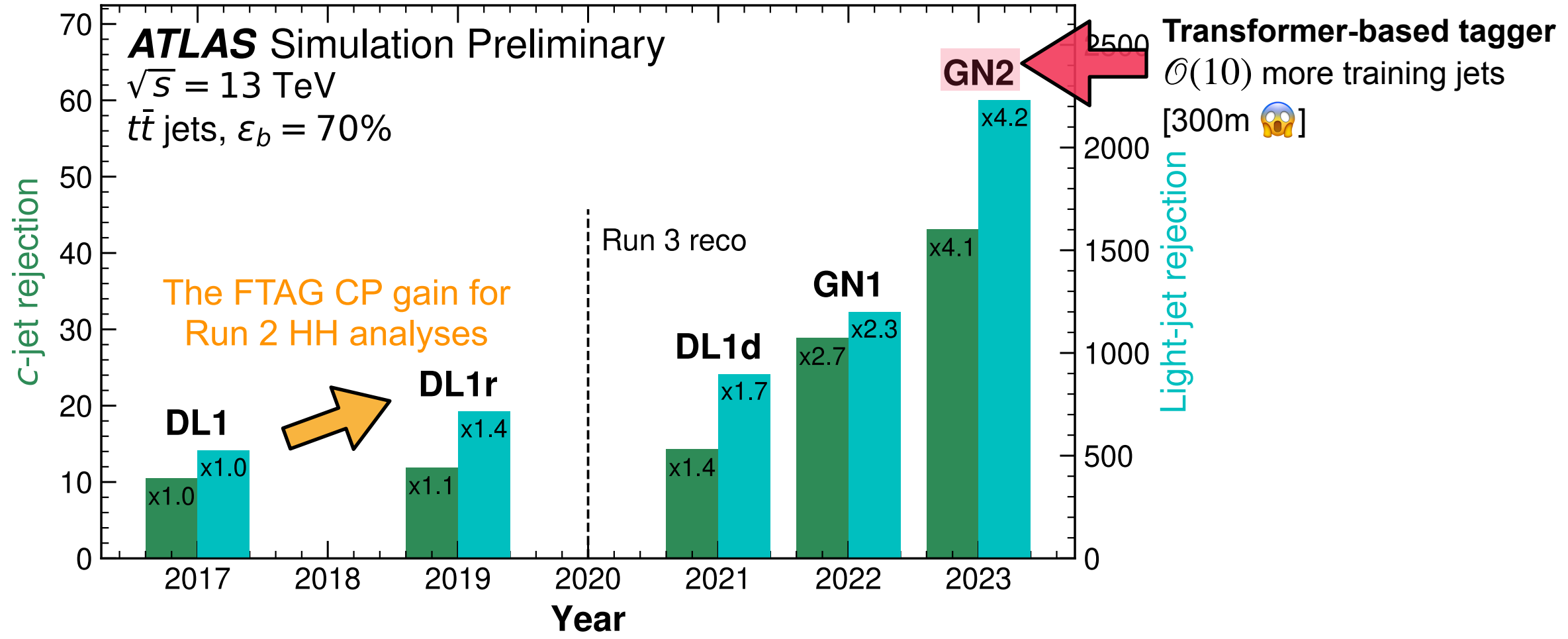
# GN2

baseline

RNNs  
+Sets

Graphs

Trans-  
formers





# ParticleTransformerAK4

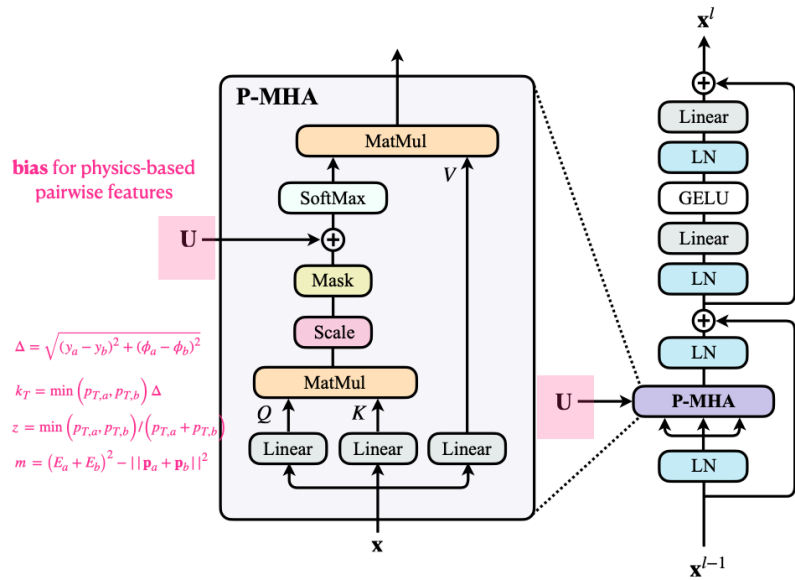
Particle Transformer (ParT) also has access to info about the **two track pairs**.

baseline

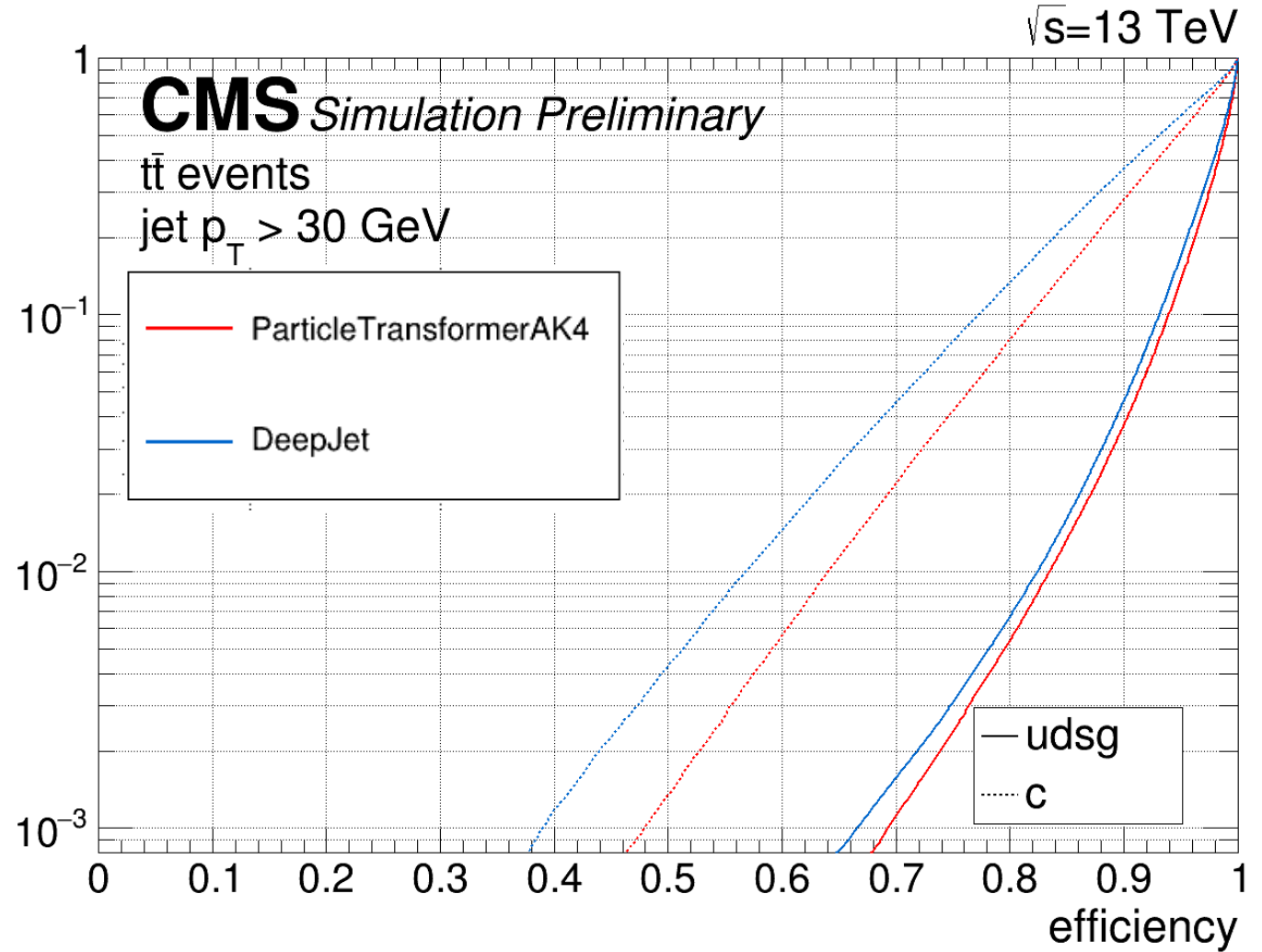
RNNs + Sets

Graphs

Transformers



misid. probability



# Conclusion

early Run 2

Shallow  
NNs / BDTs

full Run 2

Deep NNs  
(RNNs / Sets)

late Run 2 /  
early Run 3

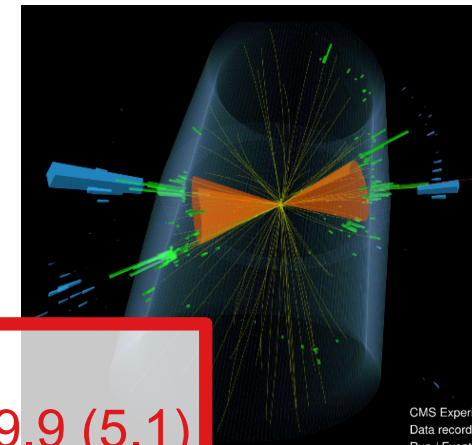
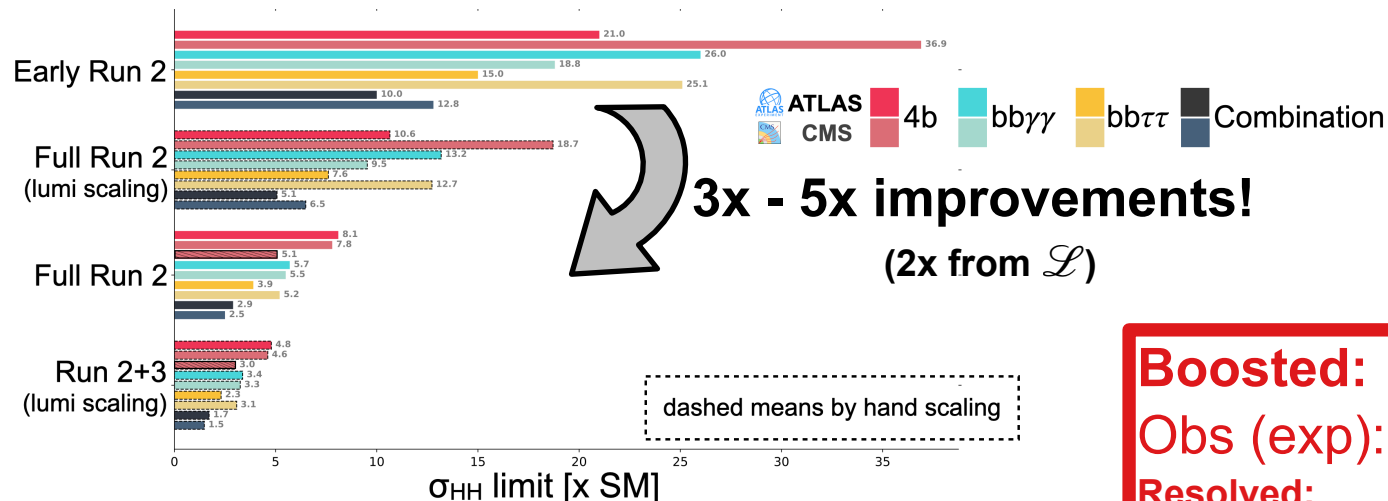
Graph NNs

State of the  
art

Transformers

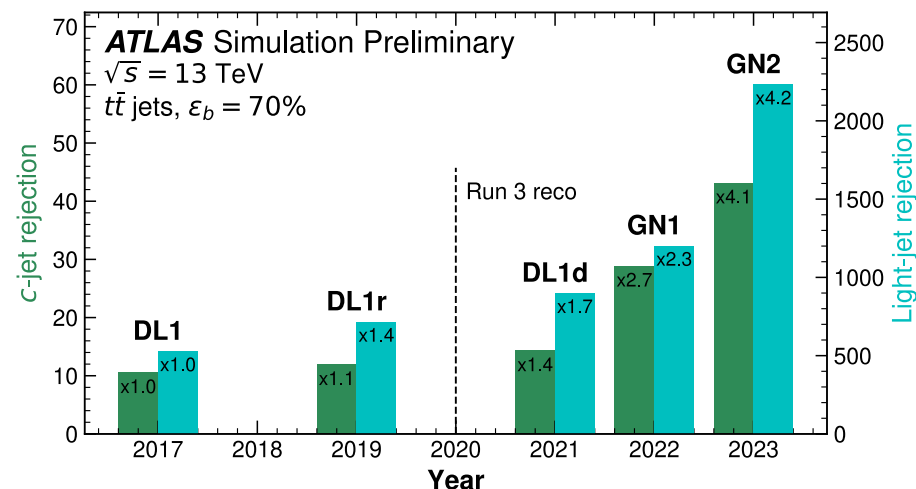
25

ML is **transforming** the way we do physics and the H(bb) results that are now key driver for our HH / YH search programs

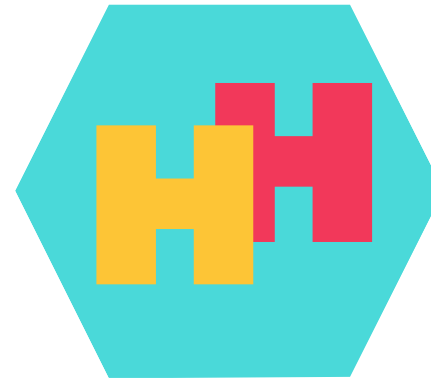
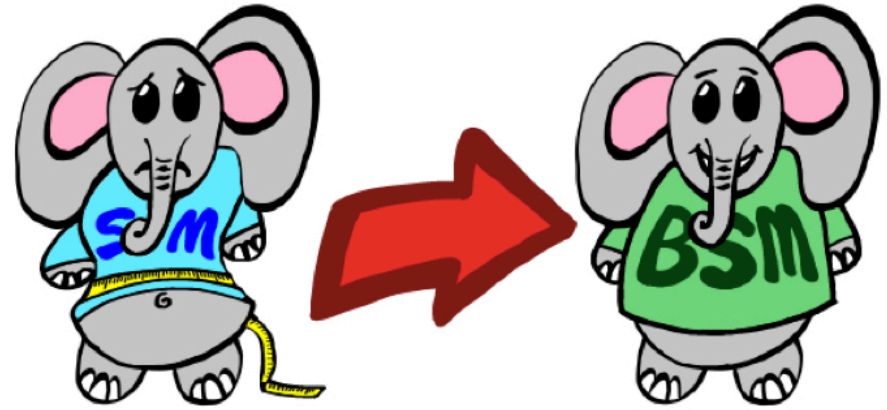


**Boosted:**  
Obs (exp): 9.9 (5.1)

**Resolved:**  
 Obs (exp): 5.4 (8.1)  
 Obs (exp): 3.8 (7.8)

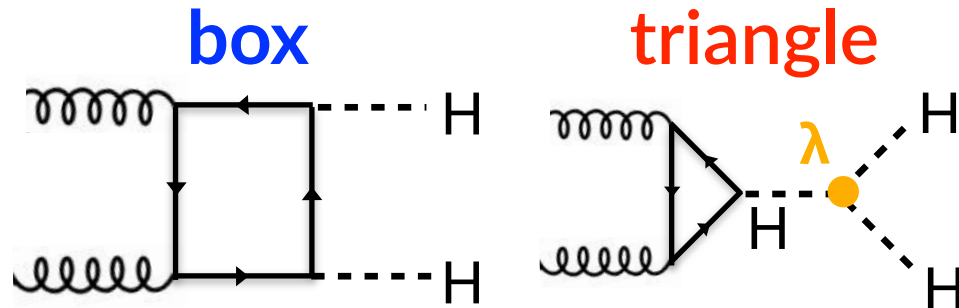


# Backup



# Why so small?

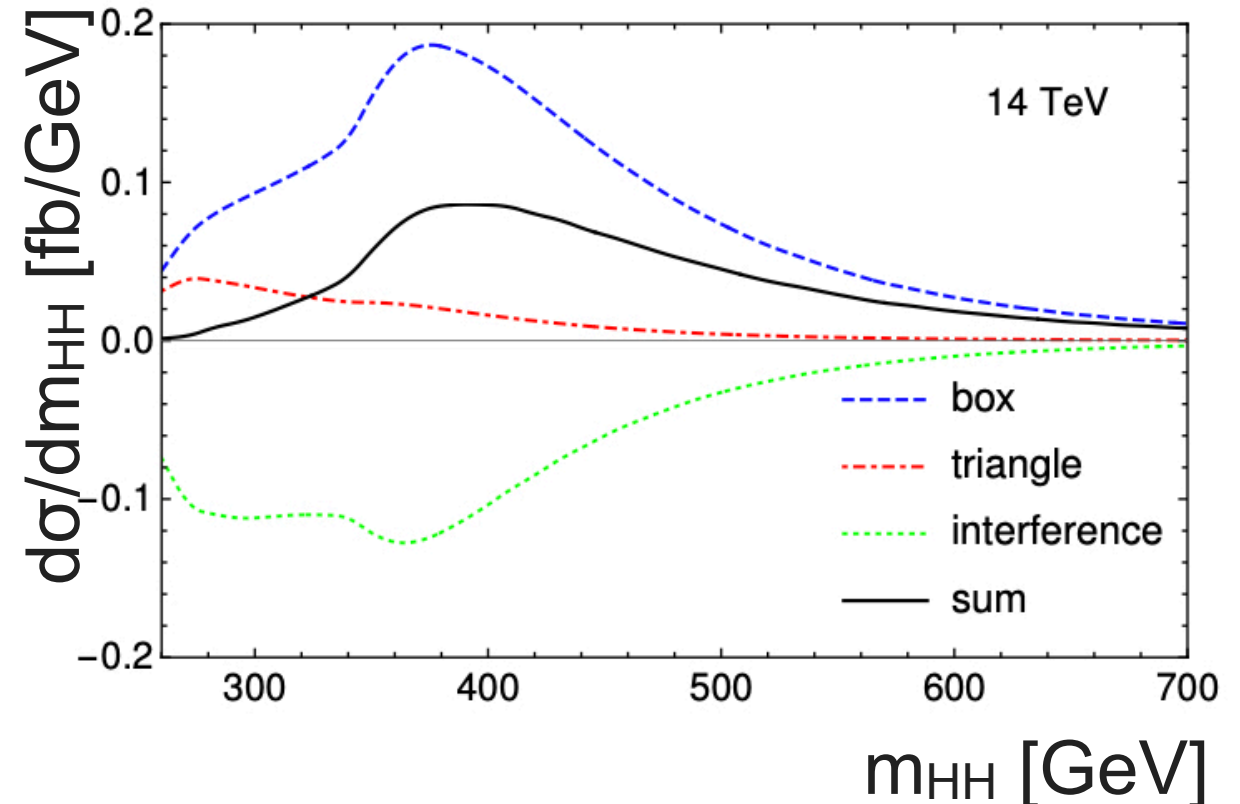
$$V(\phi) = \mu^2 h(x)^2 + \lambda v h(x)^3 + \frac{1}{4} \lambda h(x)^4$$



$$\sigma_{ggF,HH} = 31.05 \text{ fb}$$

!!! (tiny)

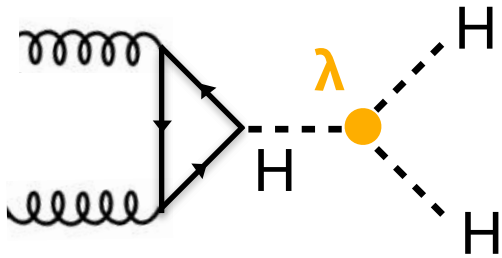
1000x lower than single Higgs



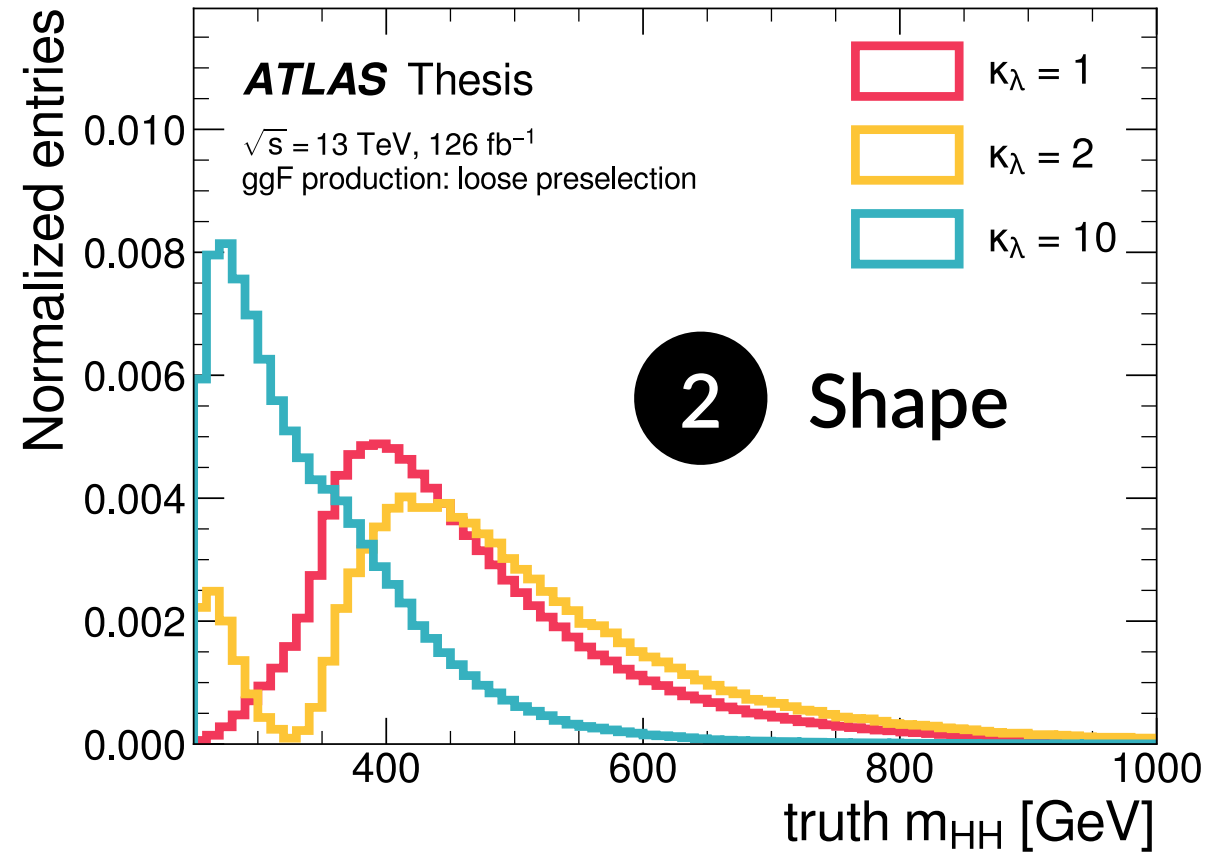
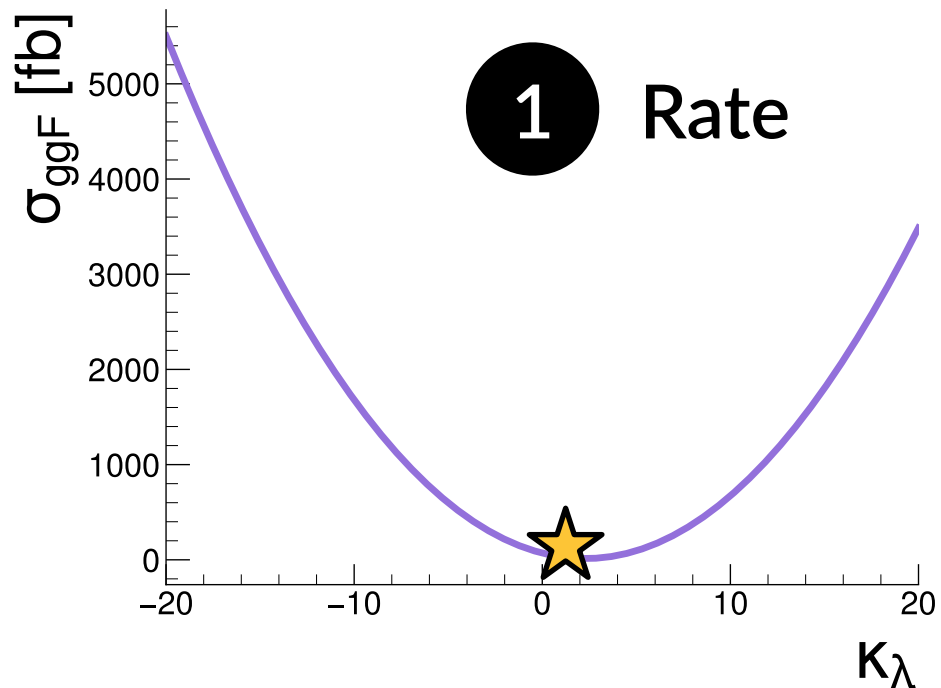
$m_{HH}$  [GeV]

TUM

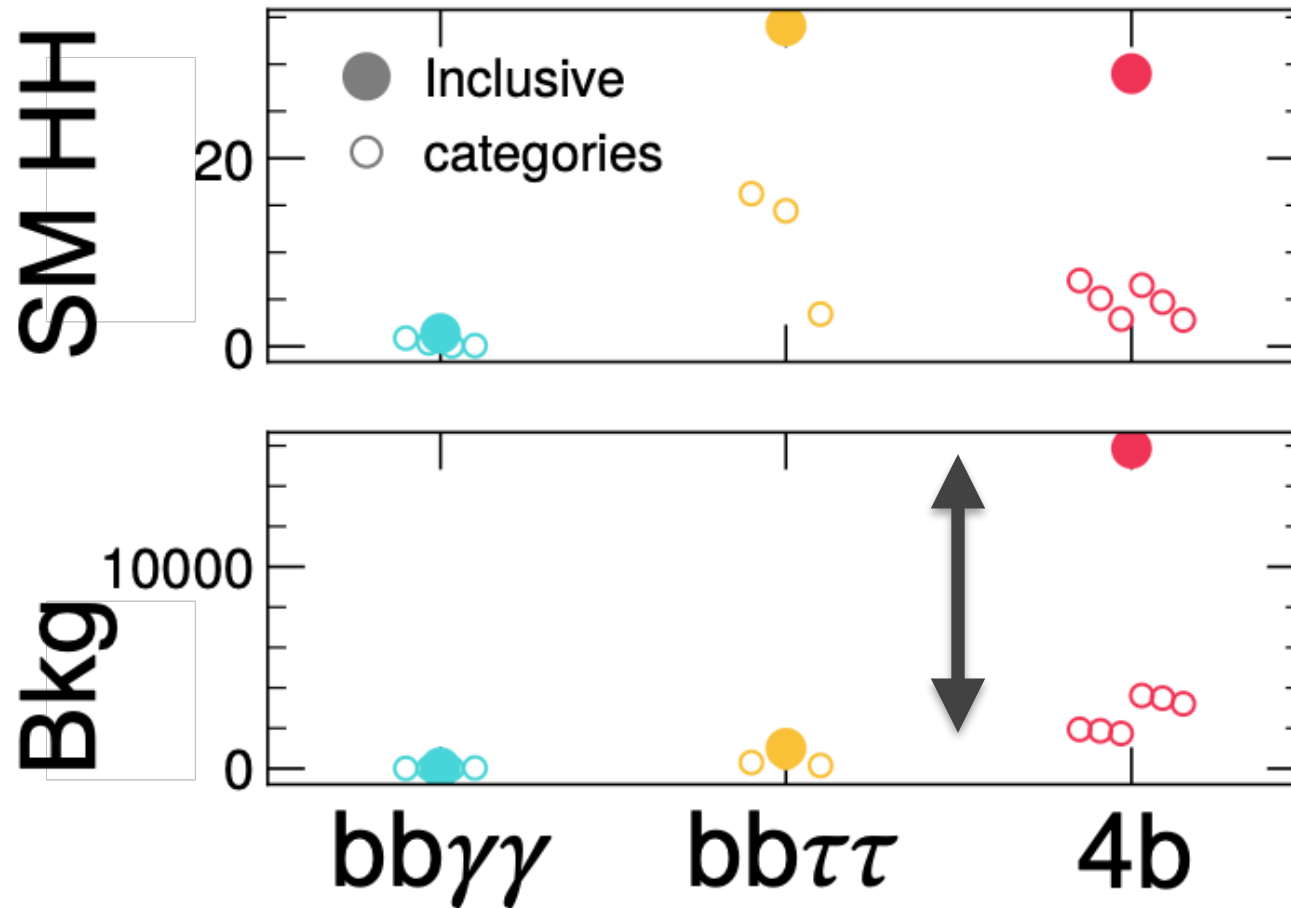
# What does it look like?



Parametrize deviations as:  $\kappa_\lambda = \frac{\lambda}{\lambda_{SM}}$



# ATLAS golden channels



From ATLAS  
 $bb\gamma\gamma$  [HDBS-2018-34](#),  
 $bb\tau\tau$  [HDBS-2018-40](#), and  
4b [HDBS-2019-29](#) analyses.

**Orders of  
magnitude higher  
background for 4b.**



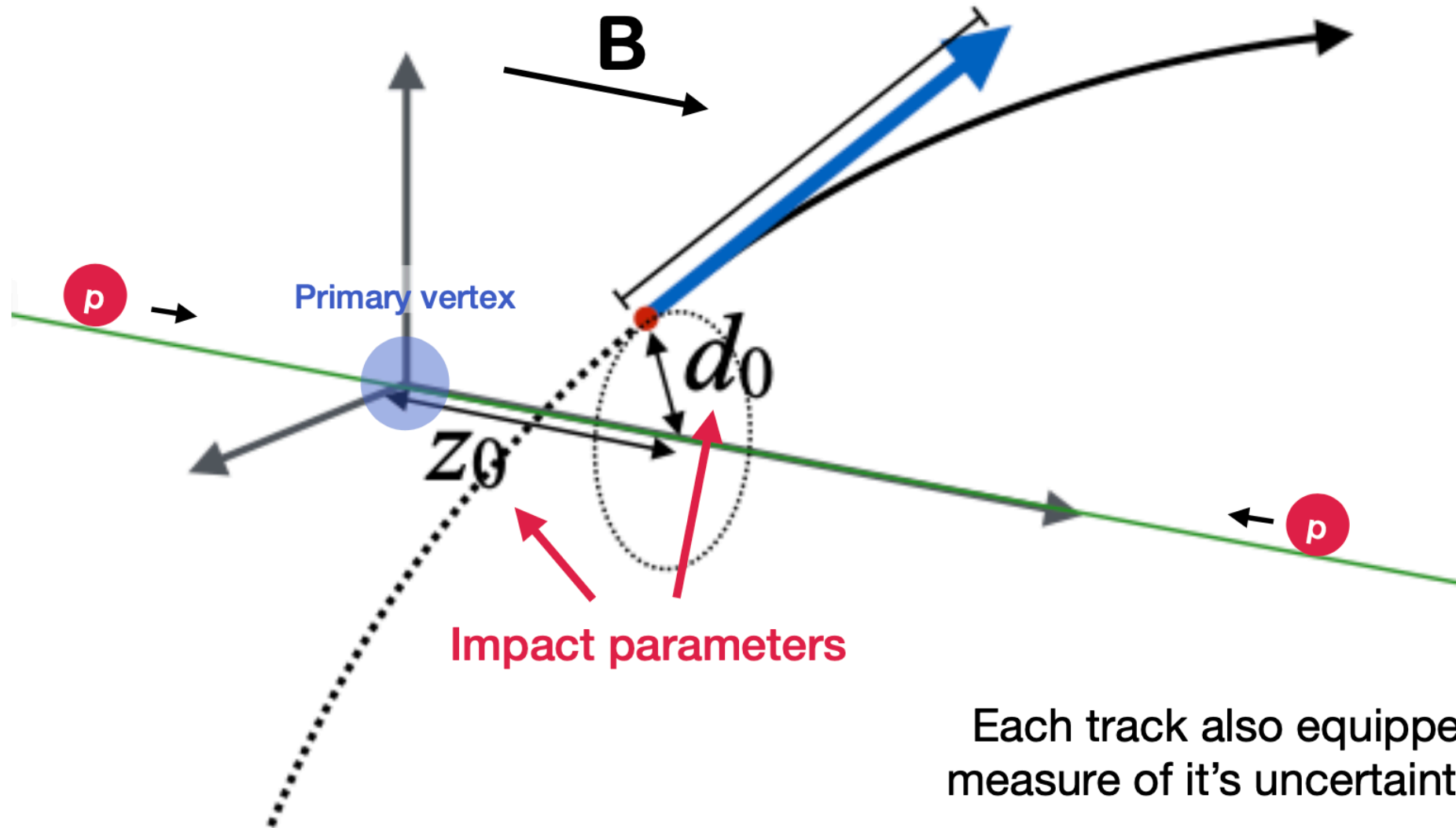
# The baseline algorithms

---

## Small R:

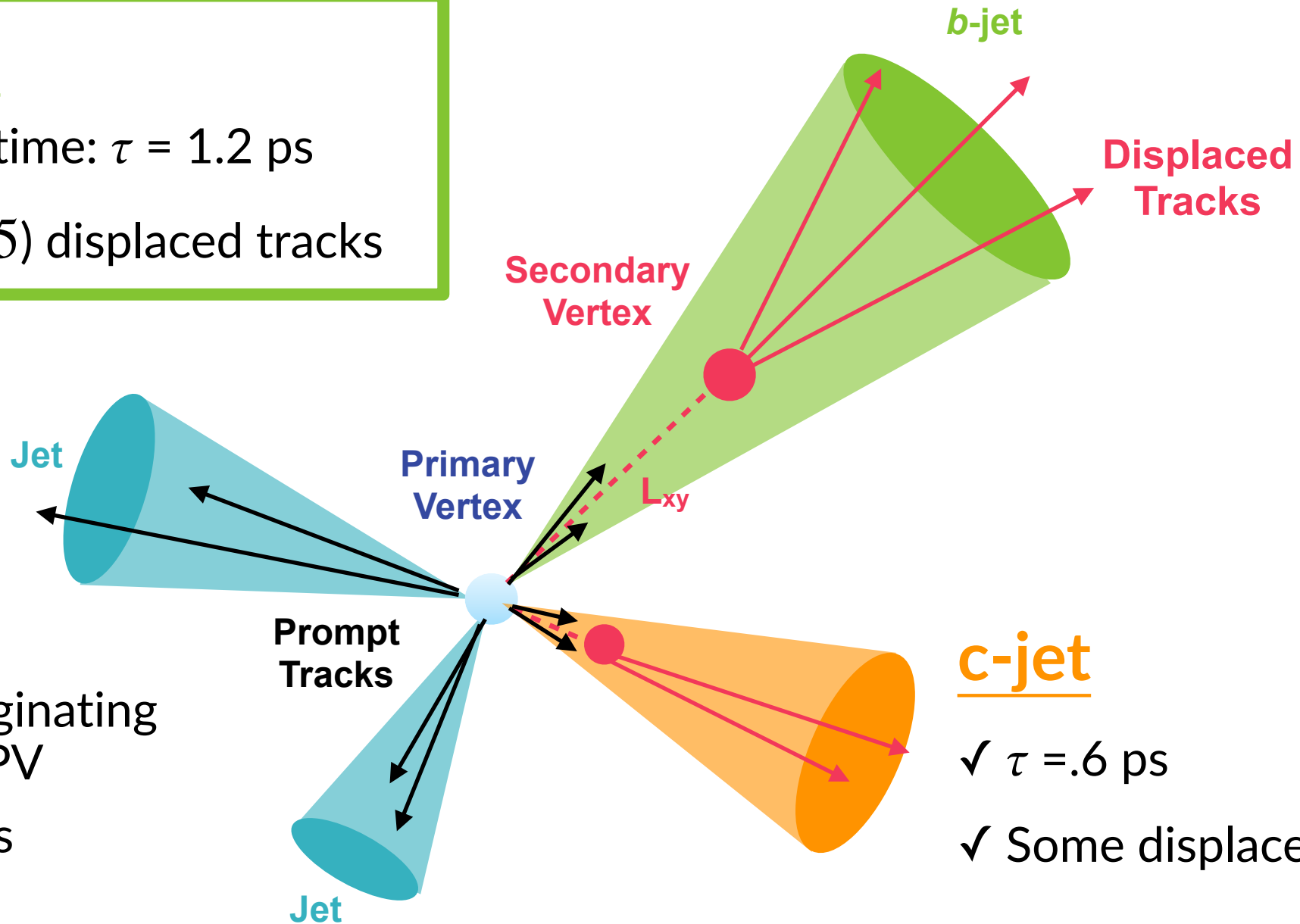
- ATLAS: MV2
- CMS: DeepCSV (?)

# Impact parameter



💖 b-jet

- ✓ “Long” lifetime:  $\tau = 1.2$  ps
- ✓ Many ( $\approx 5$ ) displaced tracks



light jet

- ✓ Tracks originating from the PV
- ✓ Few tracks

c-jet

- ✓  $\tau = .6$  ps
- ✓ Some displaced tracks

# 💖 b-jet

✓ “Long” lifetime:  $\tau = 1.2$  ps

✓ Many ( $\approx 5$ ) displaced tracks

Variable # of tracks

b-jet

Displaced Tracks

Secondary Vertex

Jet

# light jet

✓ Tracks originating from the PV

✓ Few tracks

Prompt Tracks

Vertex

c-jet

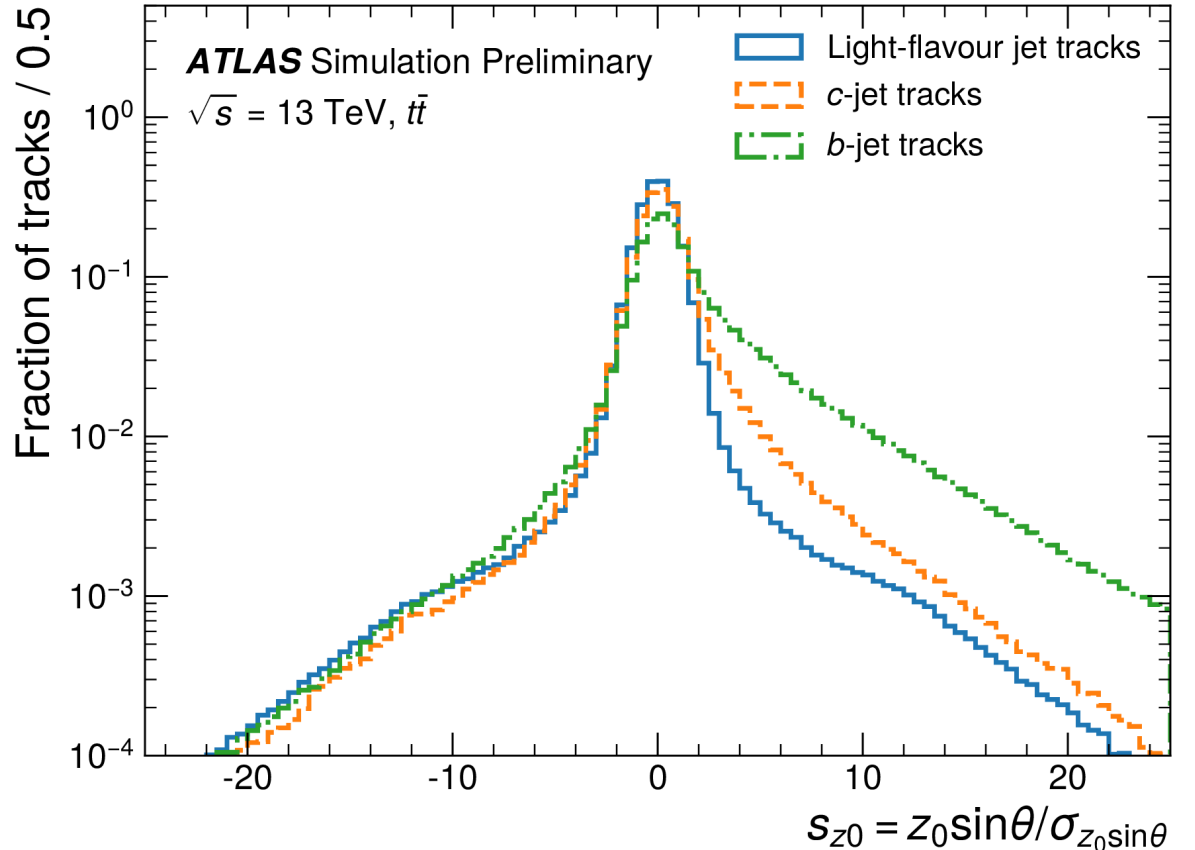
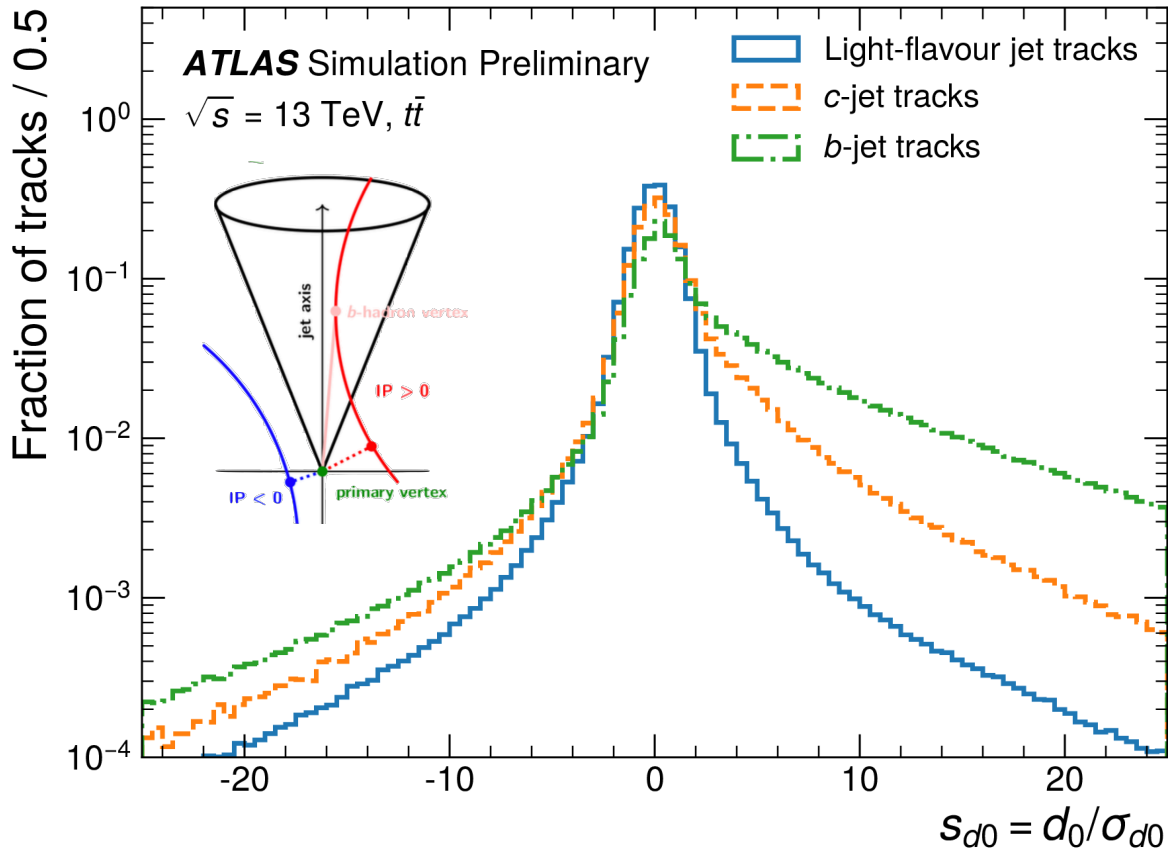
$\tau = .6$  ps

✓ Some displaced tracks

Jet

# b-tagging

Key variable: impact parameter



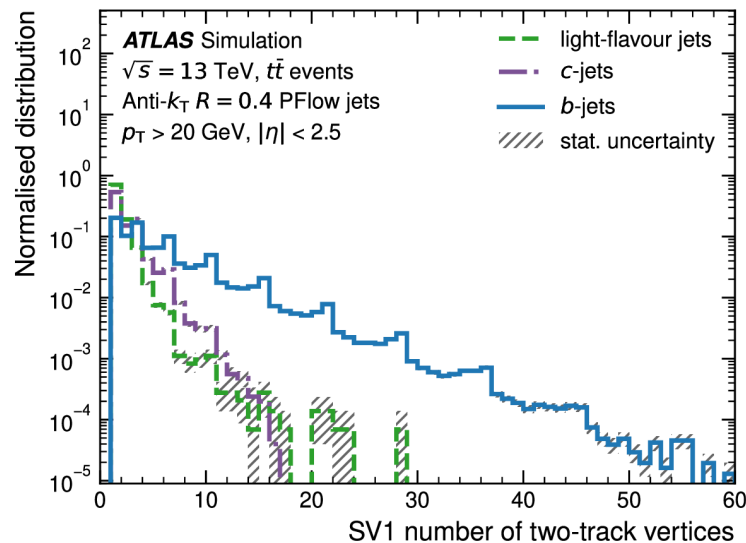
# Combined Secondary Vertex (CSVv2)

	Input variable	Run 1 CSV	CSVv2
<b>Vertex based</b>	SV 2D flight distance significance	x	x
	Number of SV	—	x
	Track $\eta_{\text{rel}}$	x	x
	Corrected SV mass	x	x
	Number of tracks from SV	x	x
	SV energy ratio	x	x
<b>Track based</b>	$\Delta R(\text{SV, jet})$	—	x
	3D IP significance of the first four tracks	x	x
	Track $p_{\text{T,rel}}$	—	x
	$\Delta R(\text{track, jet})$	—	x
	Track $p_{\text{T,rel}}$ ratio	—	x
	Track distance	—	x
	Track decay length	—	x
	Summed tracks $E_{\text{T}}$ ratio	—	x
	$\Delta R(\text{summed tracks, jet})$	—	x
	First track 2D IP significance above c threshold	—	x
	Number of selected tracks	—	x
	Jet $p_{\text{T}}$	—	x
	Jet $\eta$	—	x

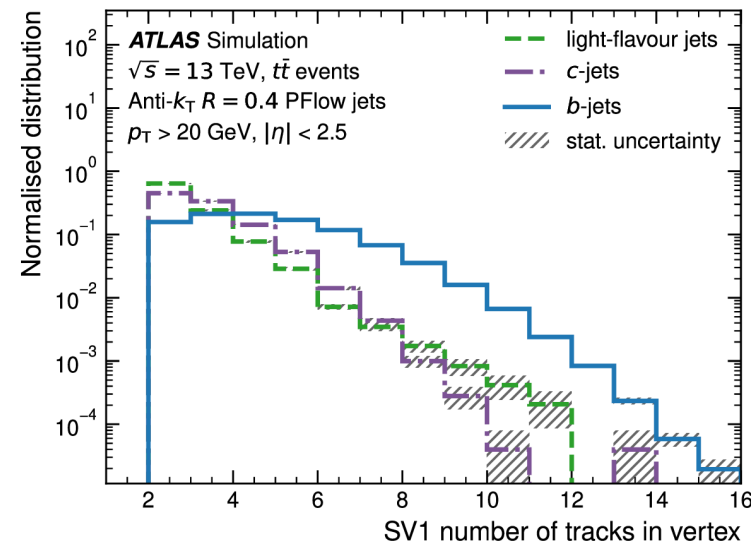
# SV1

## Secondary vertices are characteristic for b-jets.

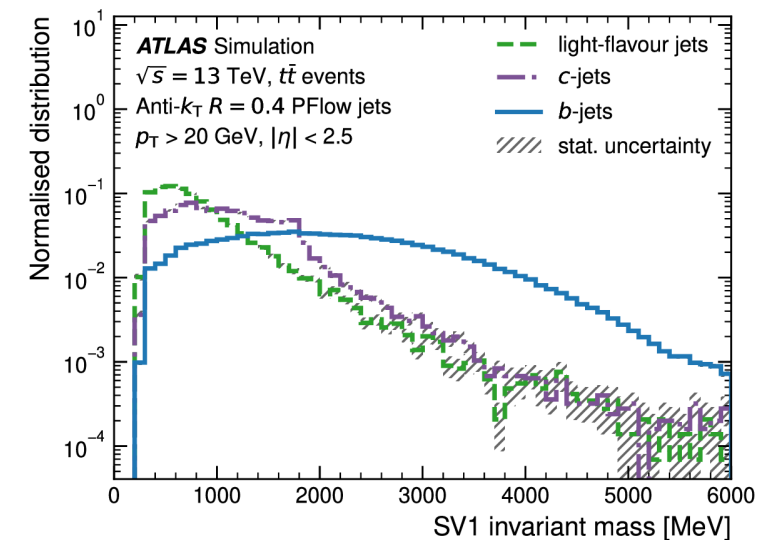
SV1 is based on single-secondary-vertex-finding algorithms searching for combinations of tracks originating from a common secondary vertex.



**Two-track vertices** are built from all tracks associated with jet and are then **iteratively merged** until one secondary vertex (SV) remains.



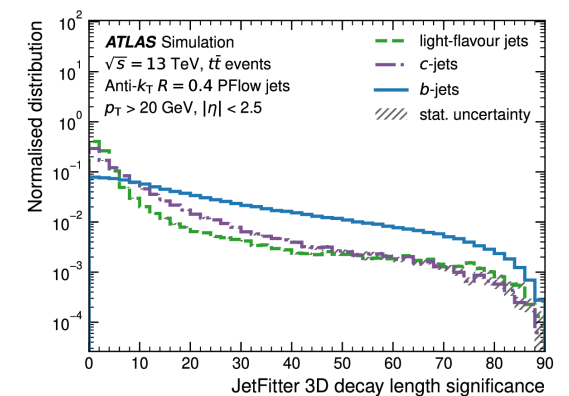
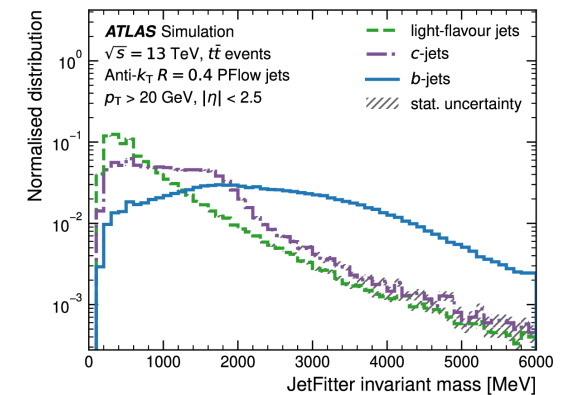
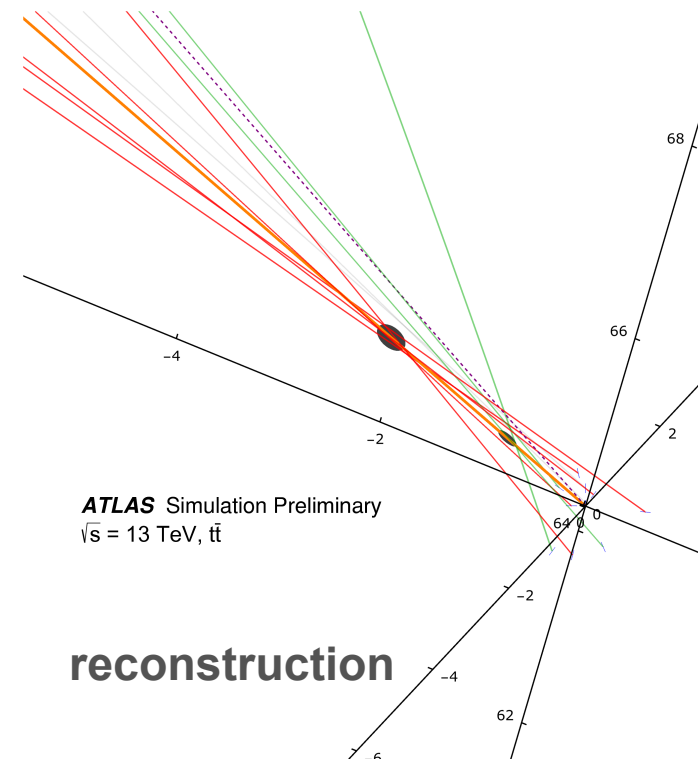
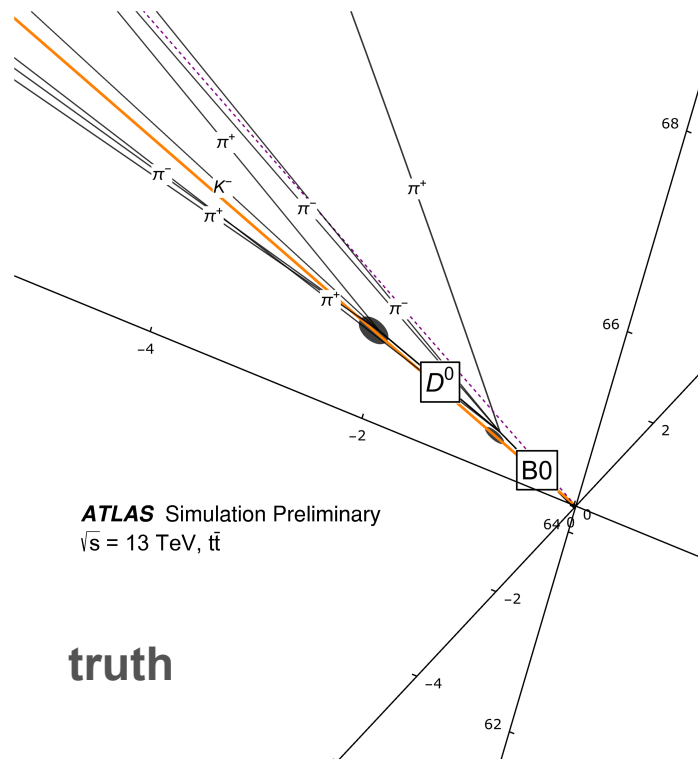
**Ks,  $\Lambda^0$ , and photon conversions** are **removed** by imposing requirements on  $\chi^2$  and mass.



Reconstructed secondary vertex provides discriminating observables like **SV mass, SV energy fraction, decay length significance, ...**

# JetFitter

**JetFitter follows cascade decays of B hadrons to charmed hadrons (D mesons) and tries to reconstruct both secondary and tertiary vertices secondary using an extended Kalman Filter for the search for a common axis connecting the three vertices.**





# Deep Neural Networks

---

Recurrent Neural Networks  
Convolutional Neural Networks

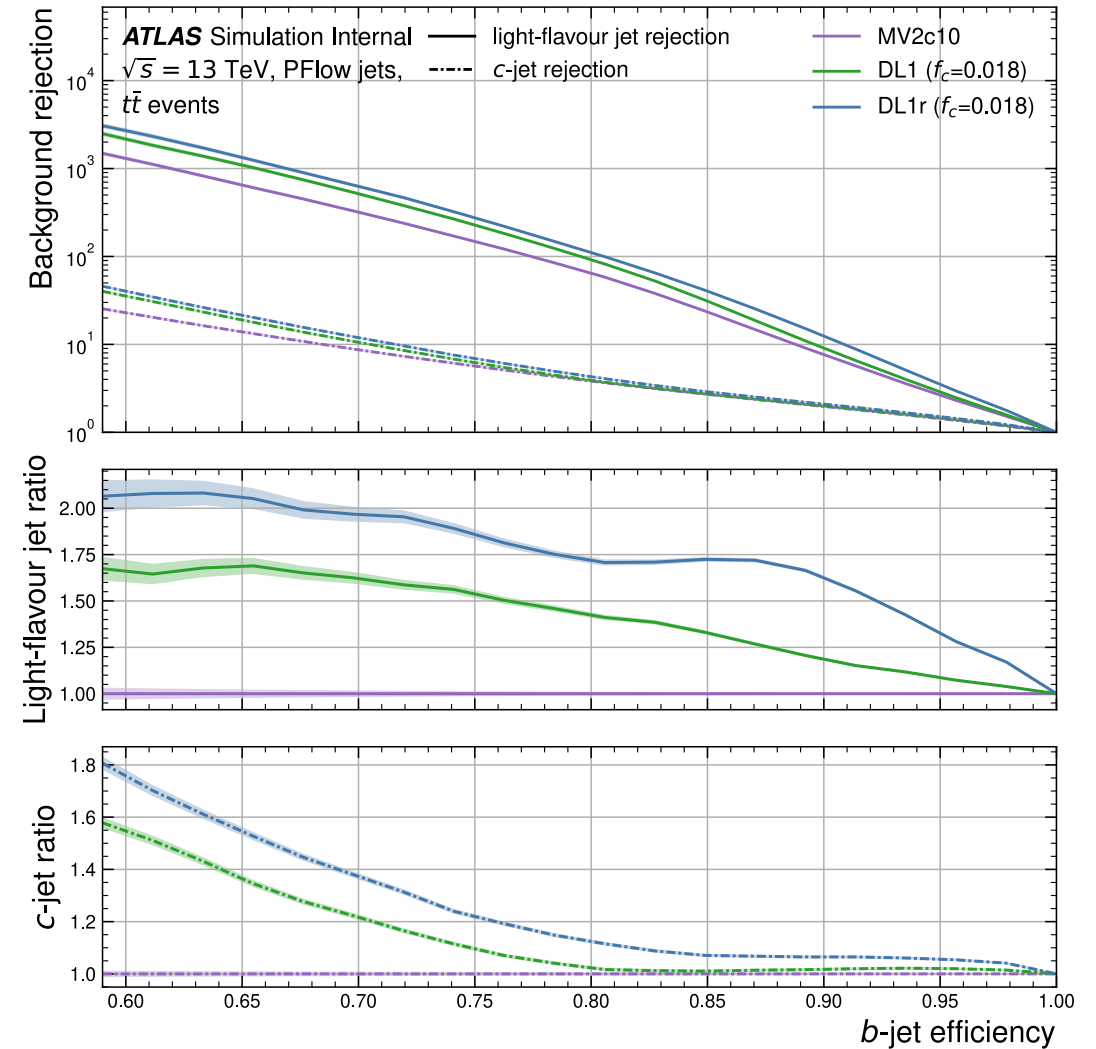
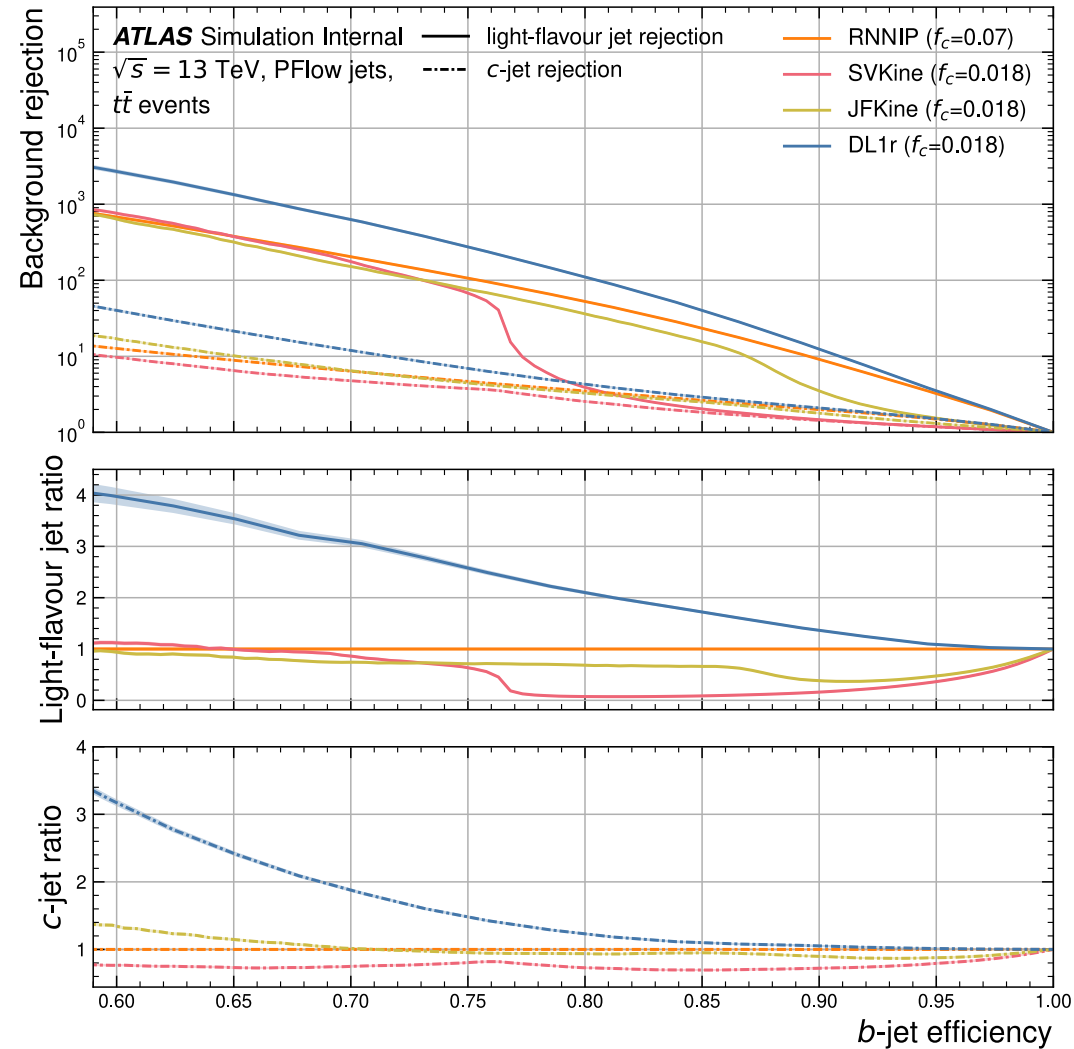
## Small R:

- ATLAS: DL1r (PFlow)
- CMS: DeepJet

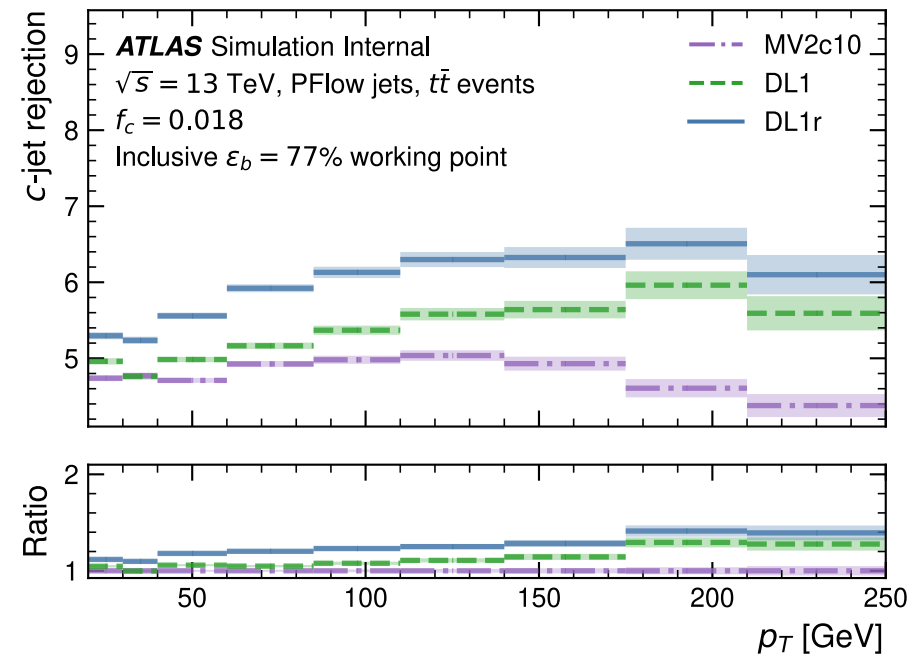
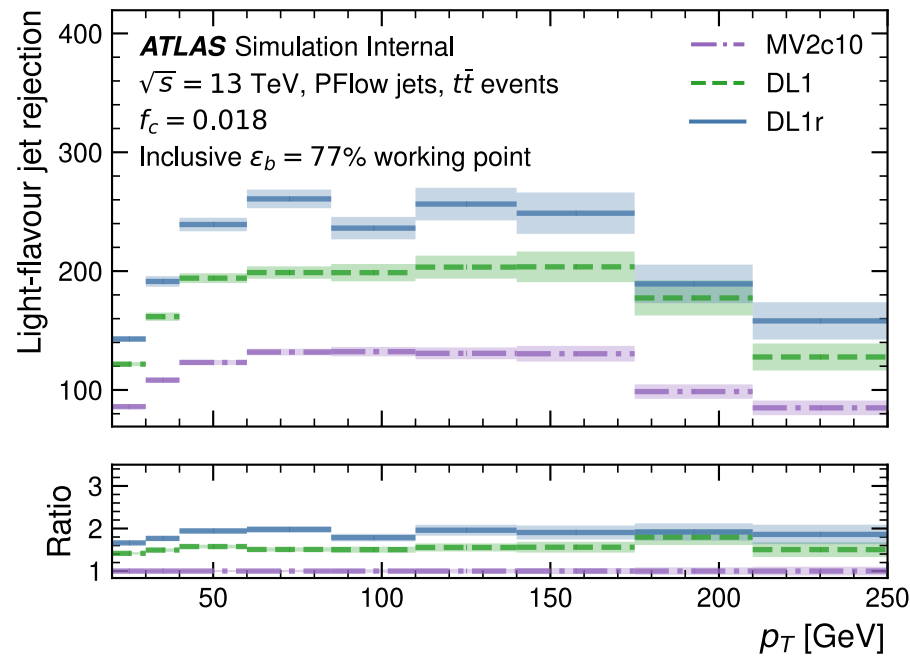
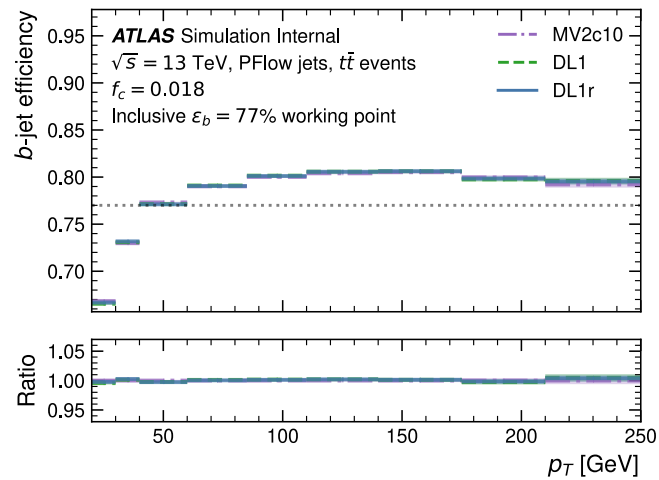
## Large R:

- ATLAS: Xbb
  - DNN trained on leading 3 associated VR track jets w/ the dedicated DL1r training
- CMS: DeepAK8

# DL1r improvement

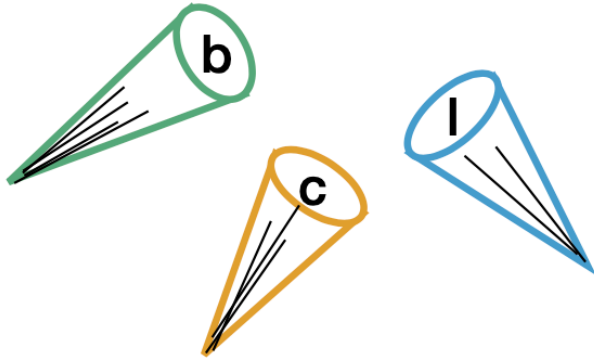


# DL1r $p_T$ dependence



# Calibration

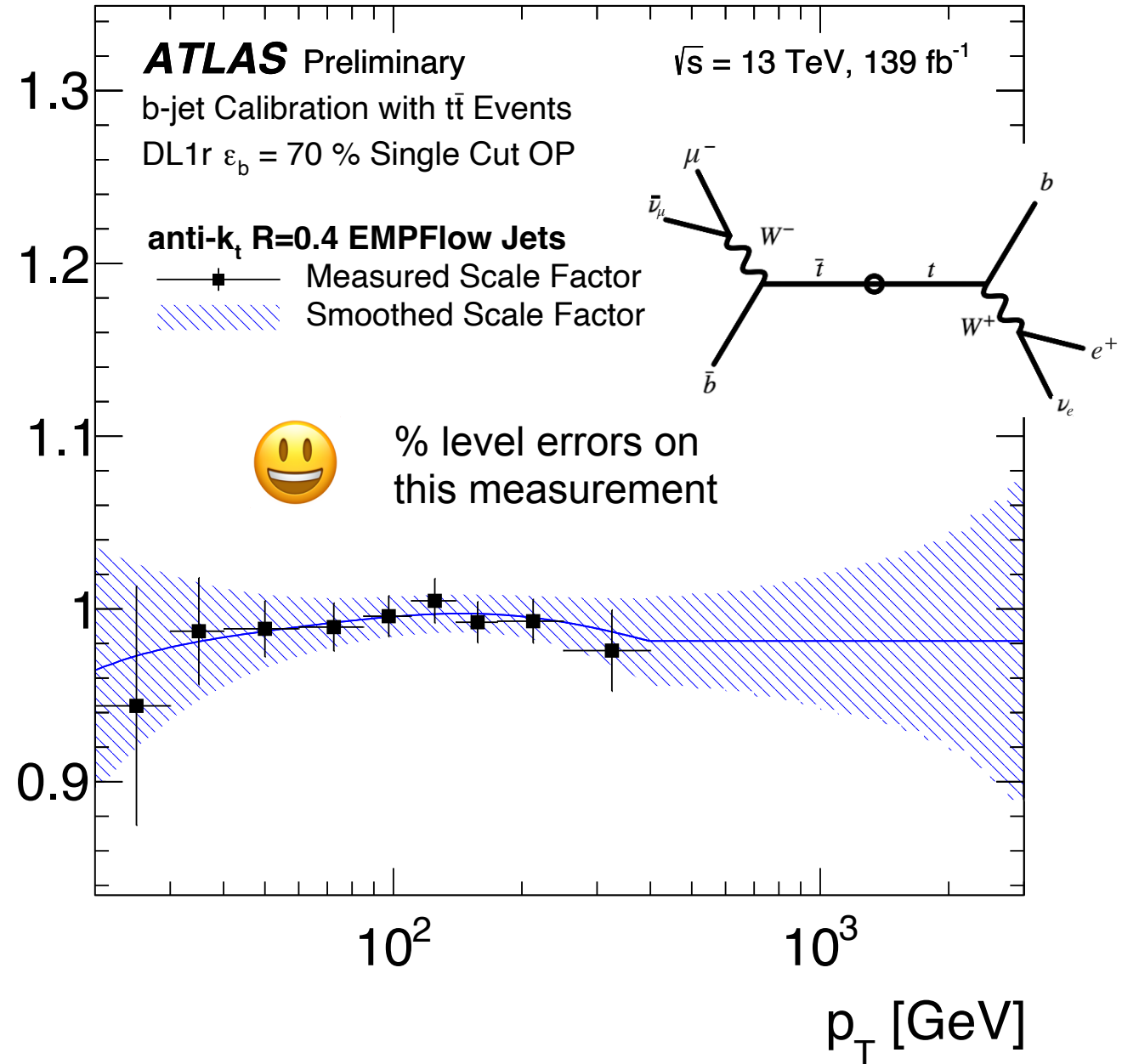
Domain adaptation



First time RNN-based b-jet tagger was calibrated....

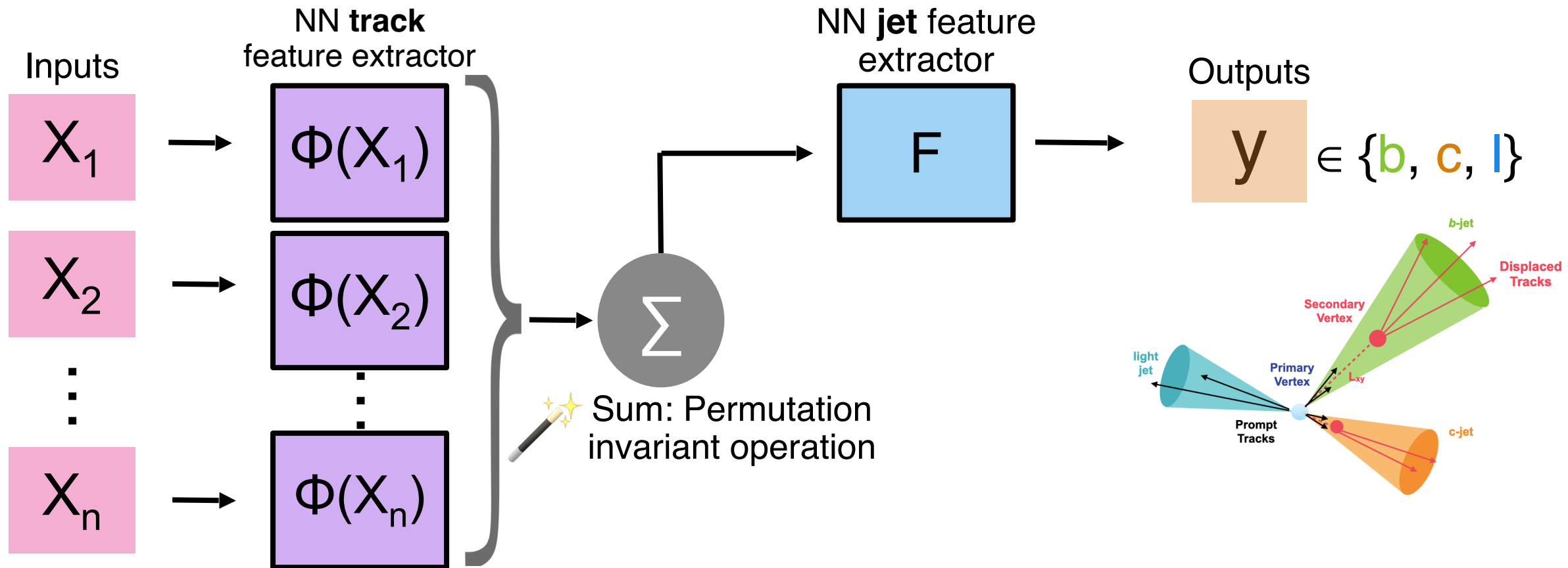
Key aspect of using in a physics analysis

b-jet Efficiency SFs



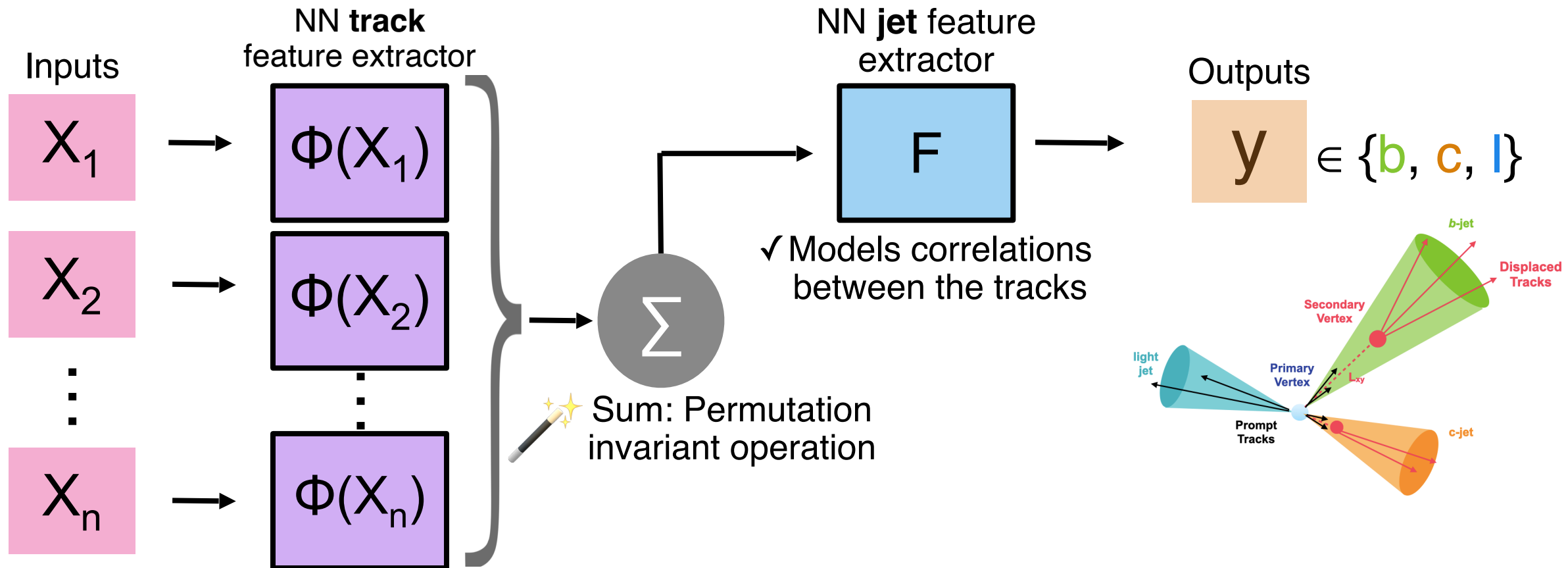
# Deep Sets

## Network of networks!!



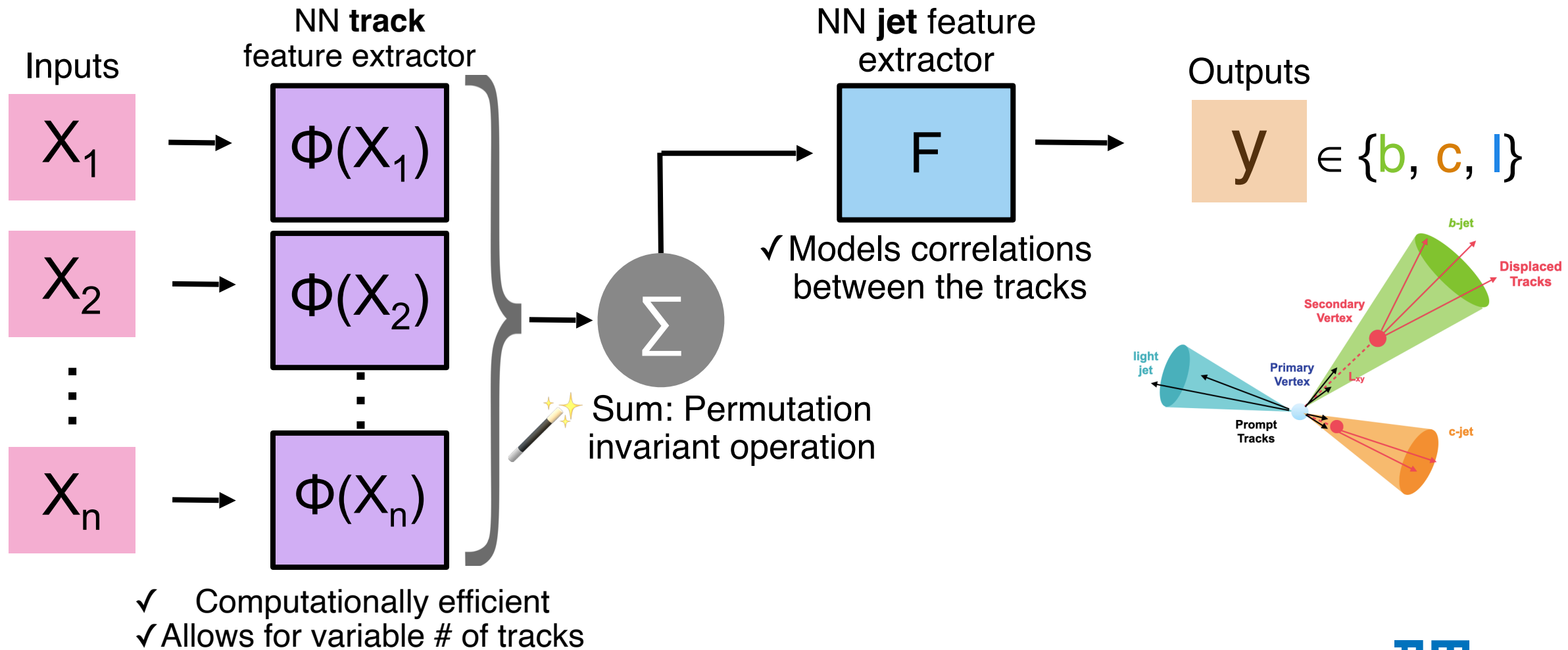
# Deep Sets

## Network of networks!!

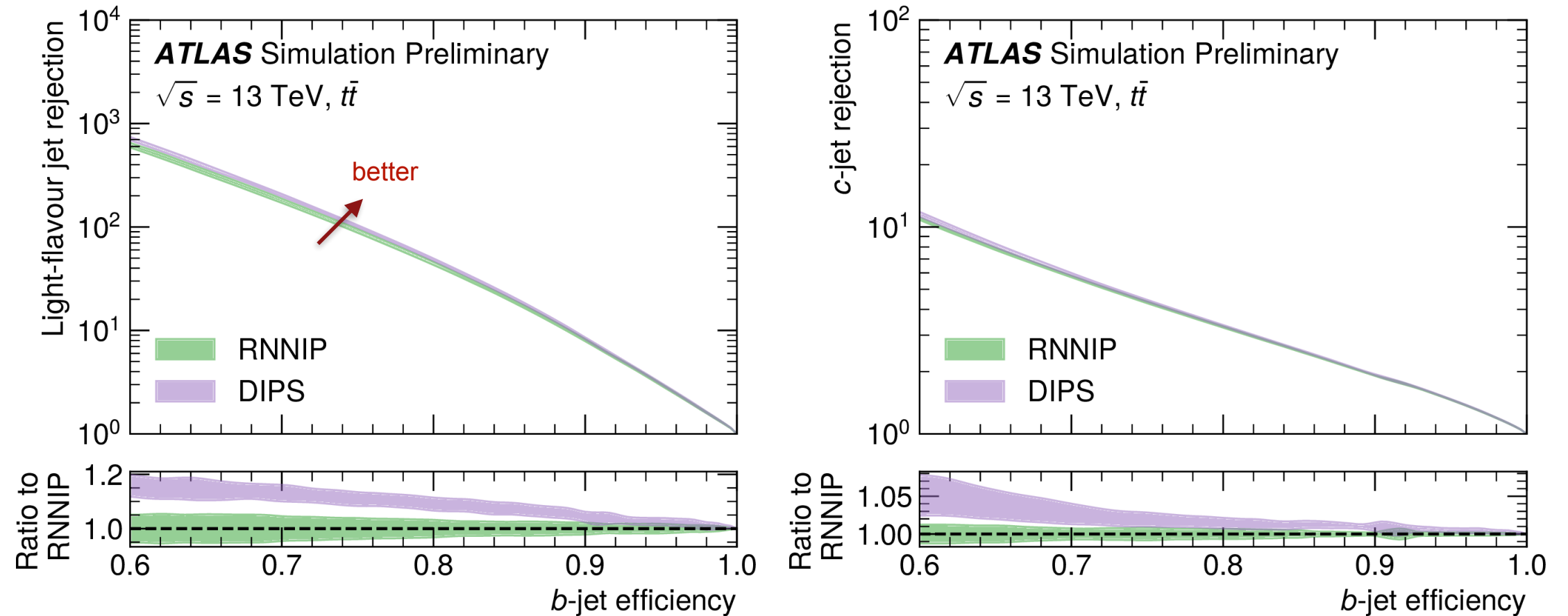


# Deep Sets

## Network of networks!!



# DIPS



Similar performance with the same inputs

**4x speed-up in the training time!!!**



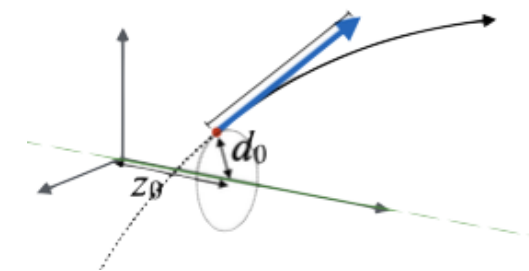


# Faster turn around time for physics optimizations

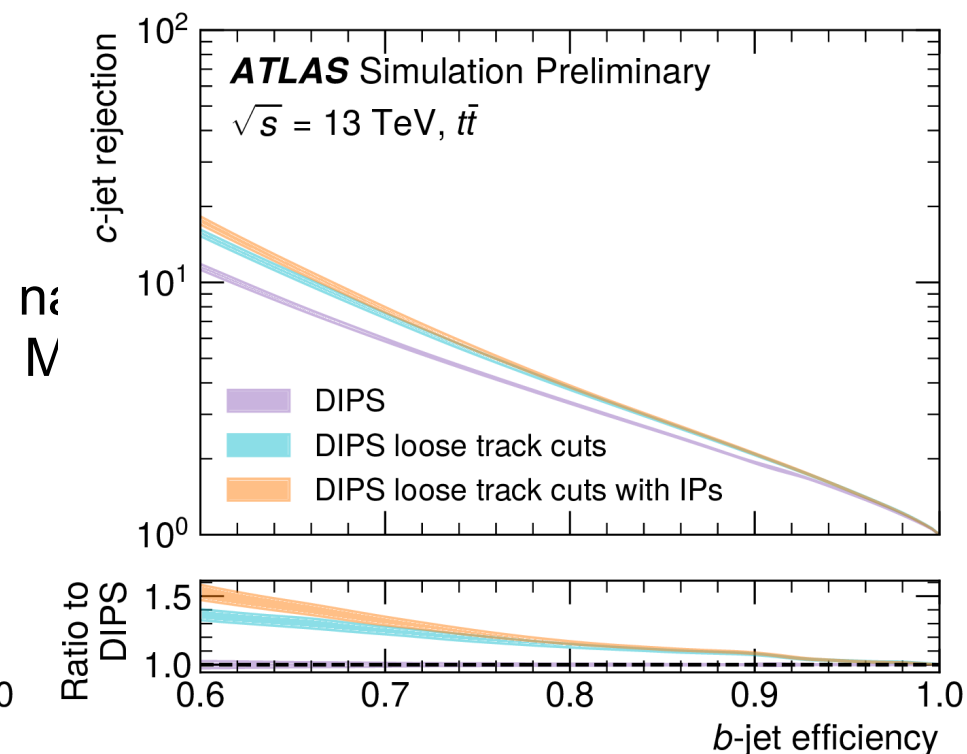
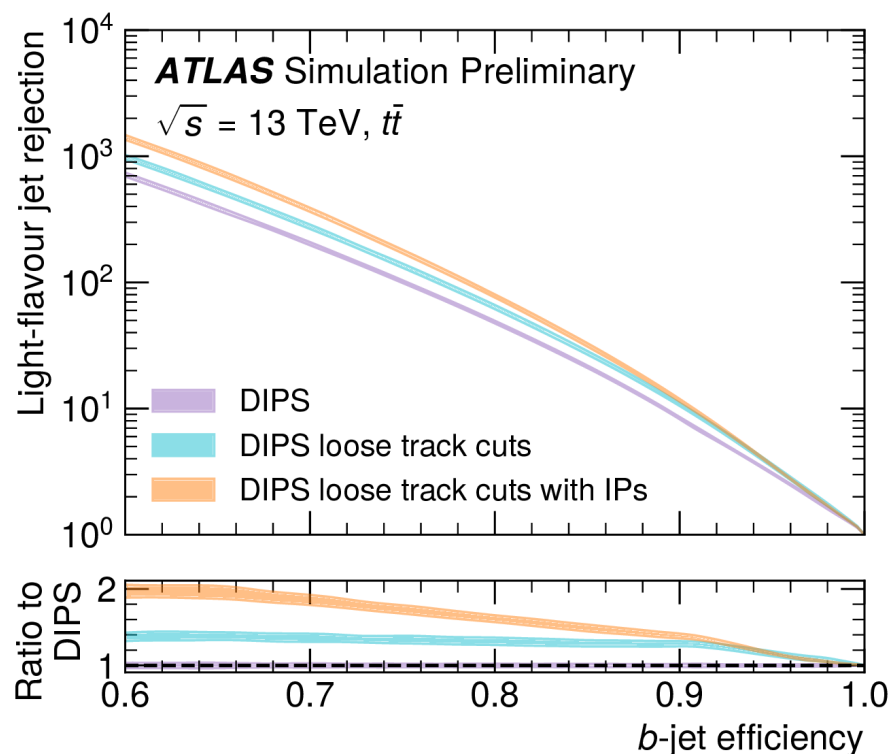
**nominal:** from before ( $p_T > 1$  GeV,  $|d_0| < 1$  mm,  $|z_0 \sin \theta| < 1.5$  mm)

**loose:** poorer quality ( $p_T > 1$  GeV,  $|d_0| < 1$  mm,  $|z_0 \sin \theta| < 1.5$  mm)

**Optimized DIPS:** loose selection + new inputs ( $d_0$ ,  $z_0 \sin \theta$ )

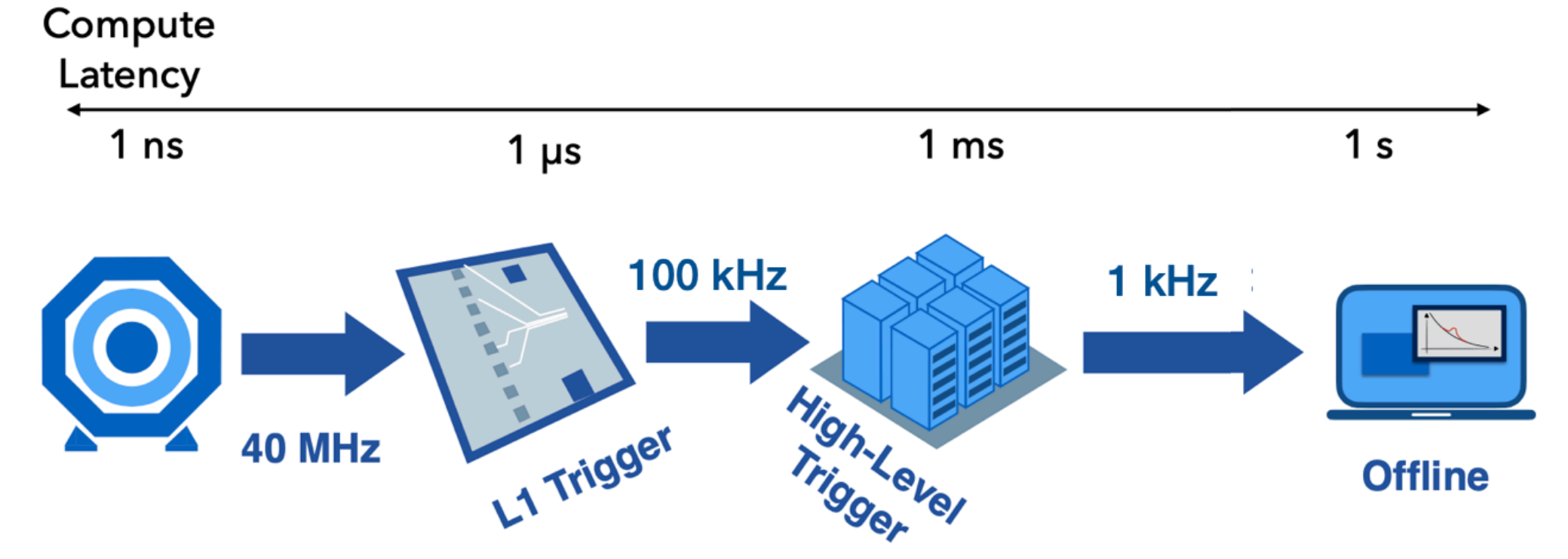


DIPS more performant for a high efficiency track selection



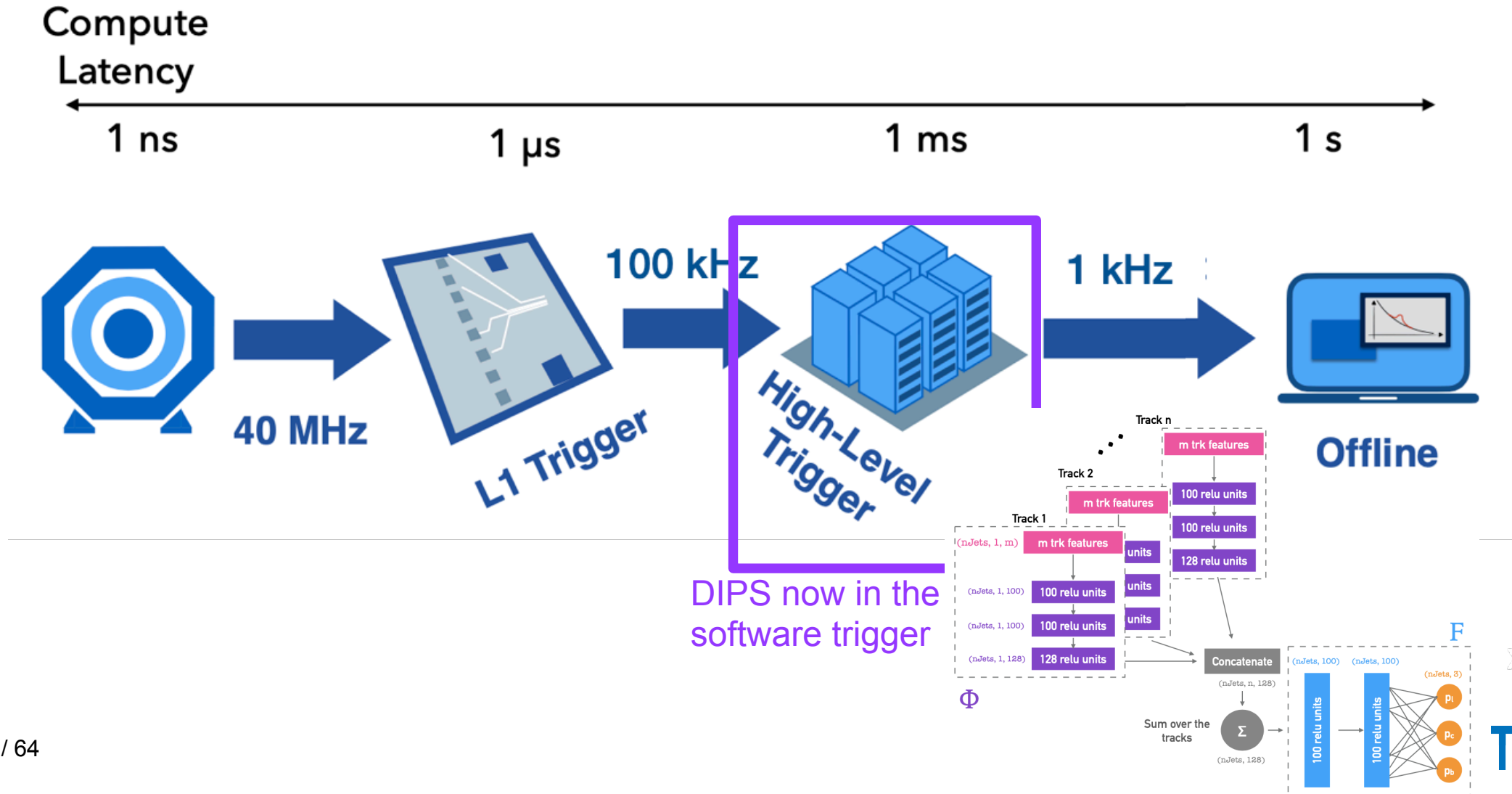
# DIPS in Data Acquisition

How do we decide which events to save?



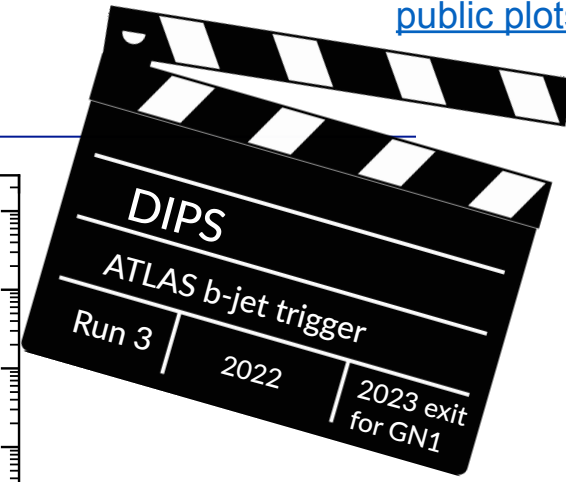
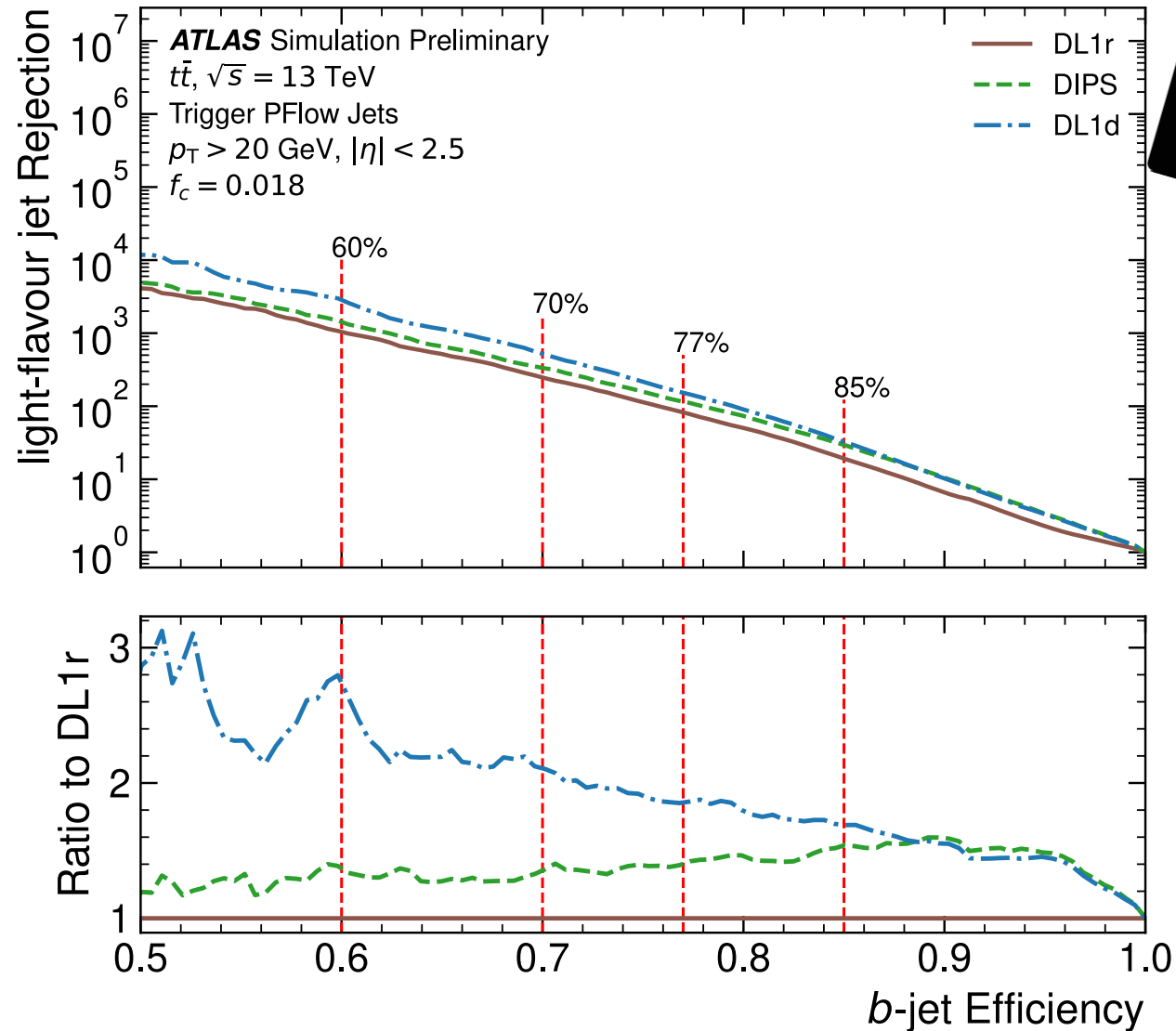
# DIPS in Data Acquisition

How do we decide which events to save?



# In real time

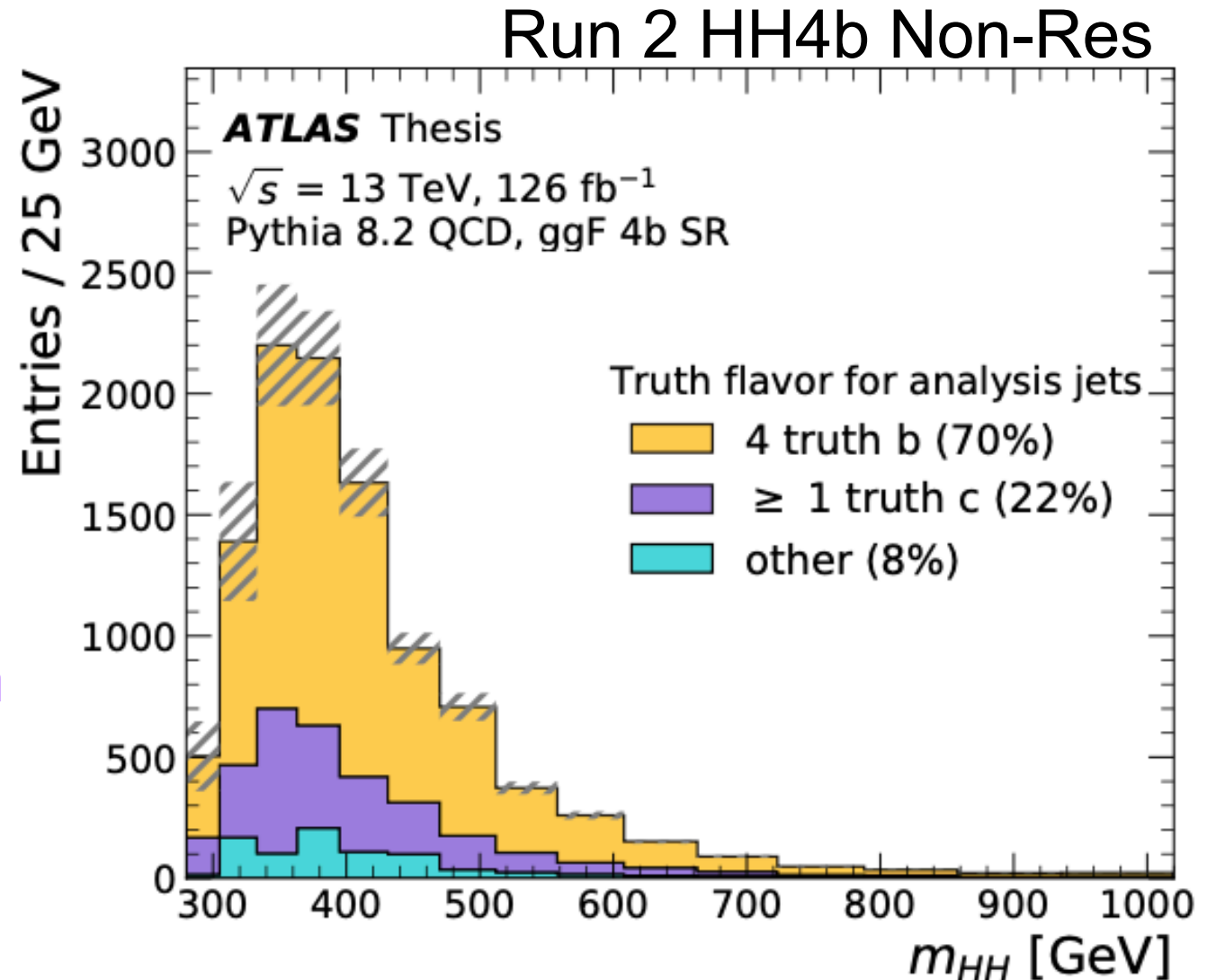
DIPS is deciding which events to keep during data-taking *happening now!!*



# Background composition: Case study

$$D_b = \log \frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_l}$$

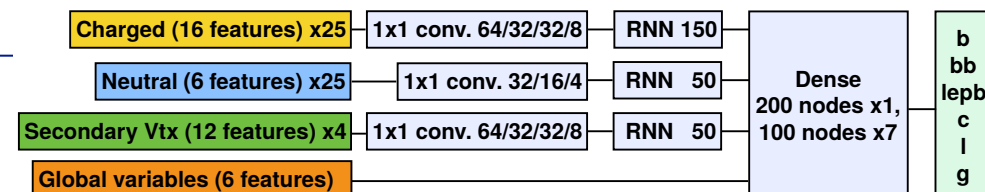
How shall we balance **charm** versus **light** rejection?



# Deep Jet inputs

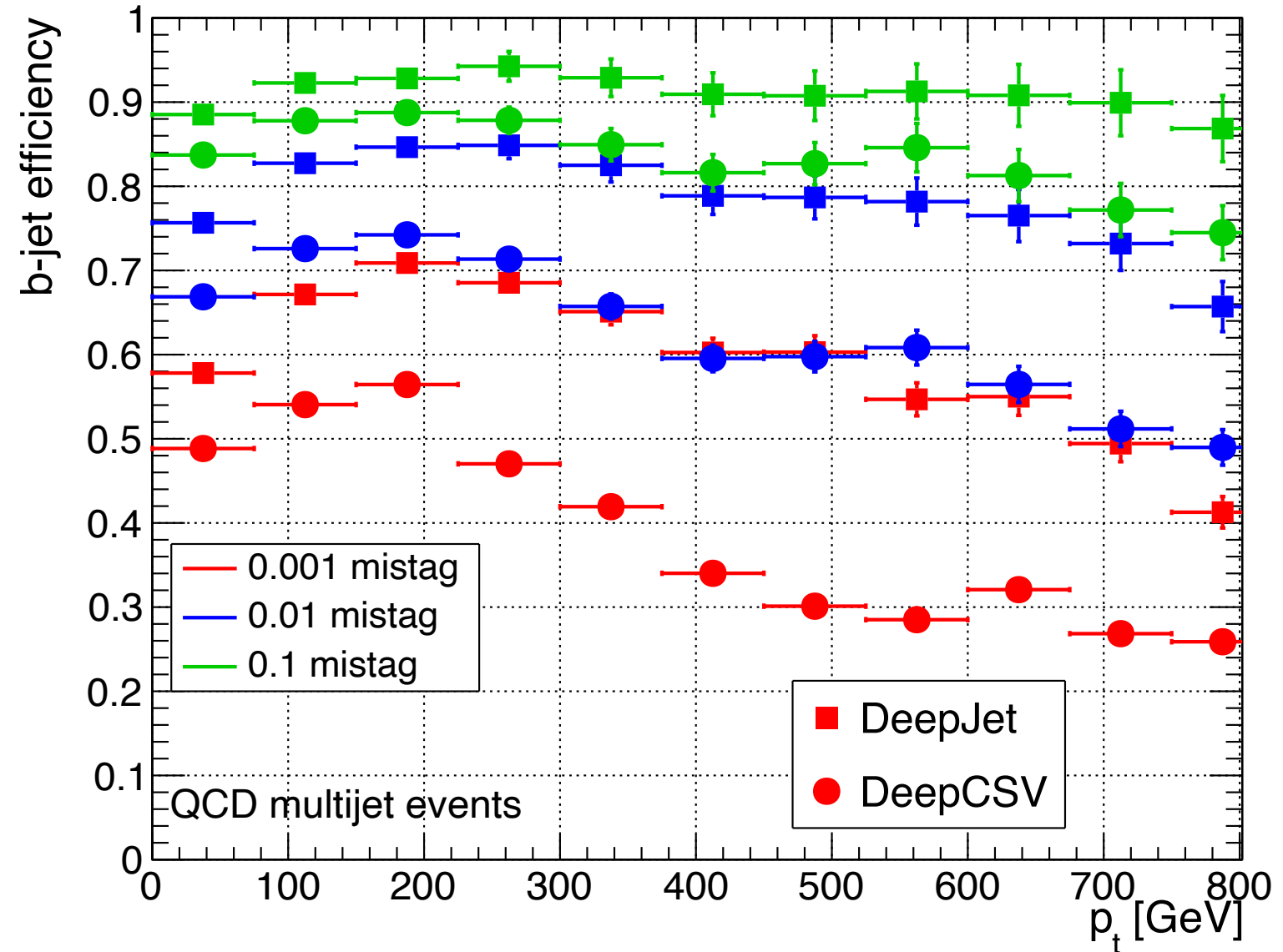
## 650 inputs!

<b>Global variables</b>	Jet $p_T$
	Jet $\eta$
	# charged pflow cand
	# neutral pflow cand
	# SV in jet
<b>Charged pflow variables [up to 25]</b>	# PV in event
	$\Delta\eta(\text{trk, jet})$
	$p_{T^{\text{rel}}}$
	$\vec{p}_{jet} \cdot \vec{p}_{trk}$
	$ \vec{p}_{jet} \cdot \vec{p}_{trk}  /  \vec{p}_{jet} $
	$\Delta R(\text{trk, jet})$
	Track 2d IP
	Track 2d IP significance
	Track 3d IP
	Track 3d IP significance
	Track distance to jet axis
	Fraction of jet momentum carried by the track
	$\Delta R(\text{trk, closest SV})$
	is in PV fit (bool)
	PUPPI weight
	$\chi^2$ of the trk fit
	track quality (category, int)



<b>Neural pflow variables [up to 25]</b>	Fraction of jet momentum carried by neutral cand
	$\Delta R(\text{neutral, jet})$
	is photon (bool)
	fraction E in the HCAL
	$\Delta R(\text{neutral, closest SV})$
<b>Secondary vertex variables [up to 4]</b>	PUPPI weight
	SV $p_T$
	$\Delta R(\text{SV, jet})$
	SV mass
	# tracks in the SV
	$\chi^2$ of the SV fit
	Reduced $\chi^2$ of the SV fit
	SV 2d IP
	SV 2d IP significance
	SV 3d IP
	SV 3d IP significance
	cosine of angle b/w SV flight direction and SV momentum
	Ratio of SV E to jet E

# Deep Jet $p_T$ dependence



# Deep Jet – where do the performance gains come from?

baseline

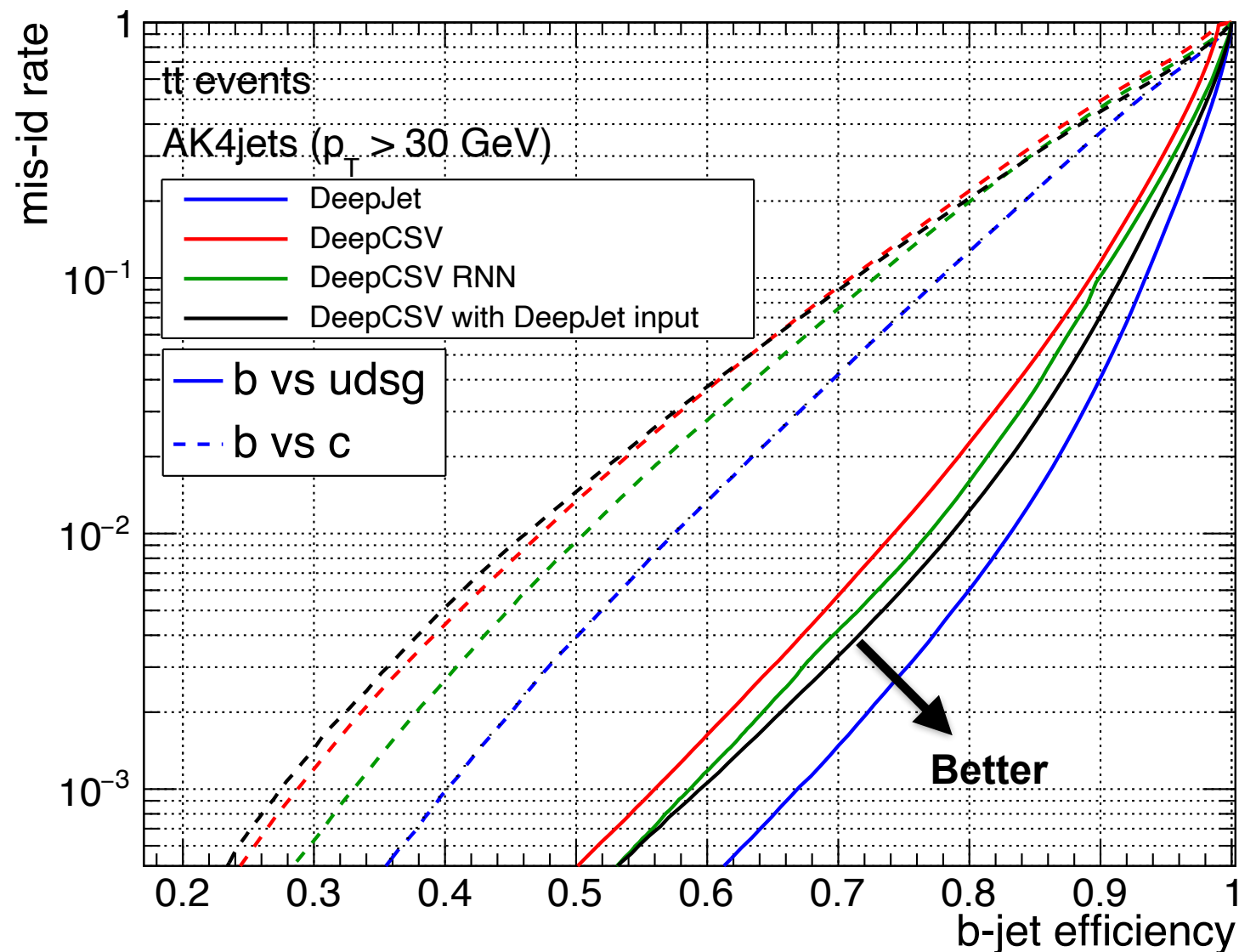
**DeepJet:** 1d-CNN+RNN “SOTA”  
Run 2 tagger

**DeepCSV:** Early Run 2 tagger  
(shallow DNN)

**Deep CSV:** Same inputs as  
CSV, but w/ the the DeepJet  
RNN architecture

**DeepCSV with DeepJet input:**  
Pass the extended inputs with  
the older architecture

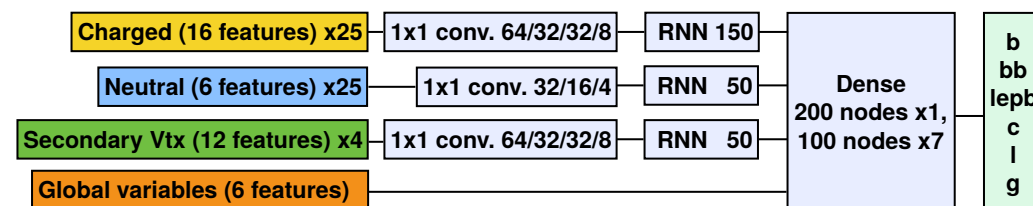
∴ Need the fancier architecture to  
take advantage of the low-  
dimensional inputs

RNNs  
+CNNsSets +  
GraphsTrans-  
formers

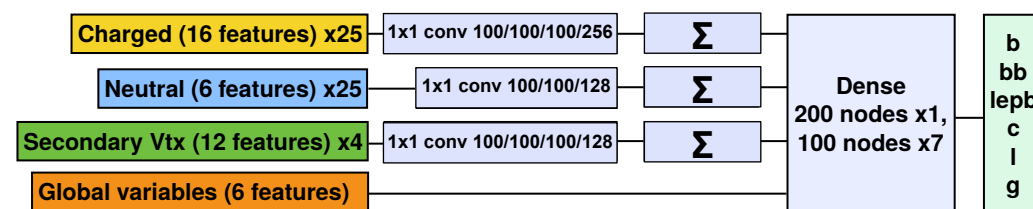


# Deep Jet compared to Deep Set

## DeepJet: Default (R2 SOTA)

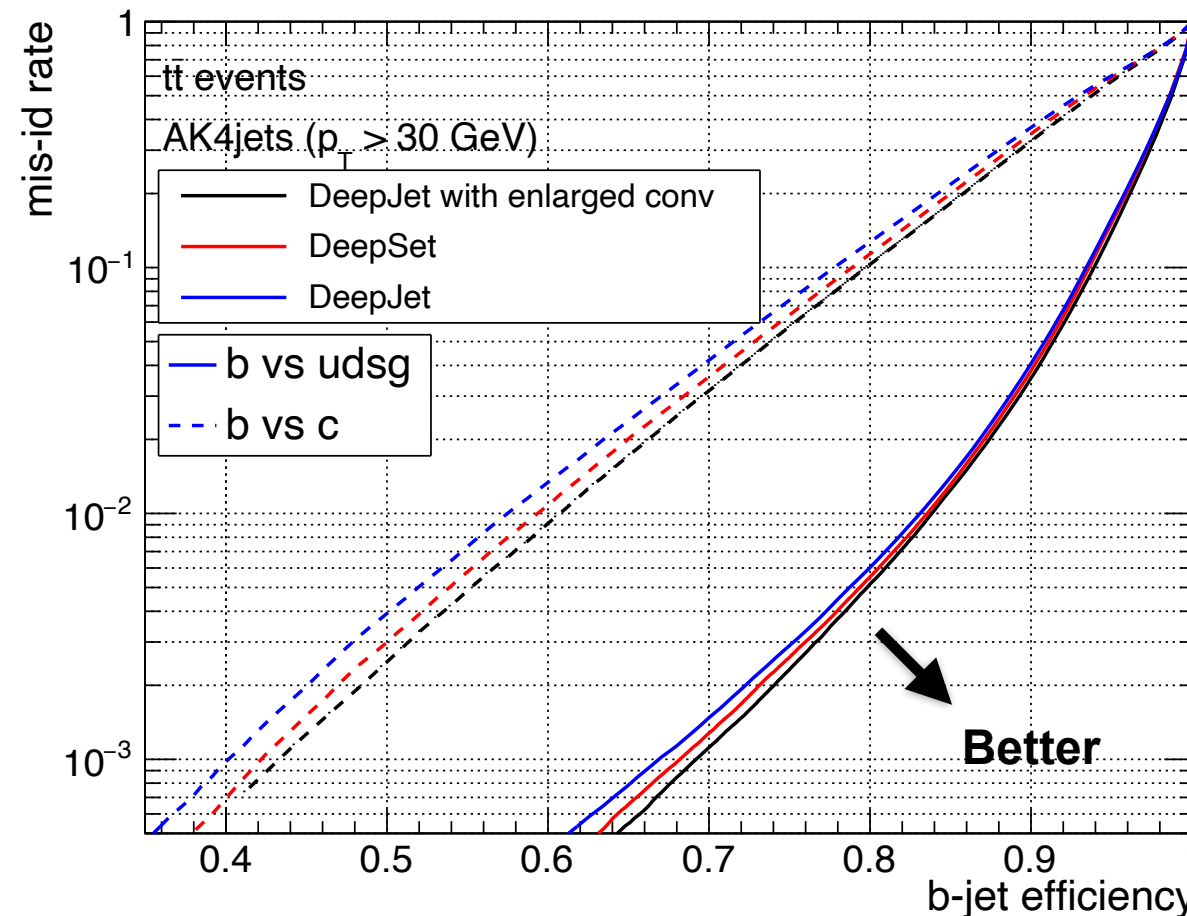


**DeepSet:** Mimic the filter sizes of  $\Phi$  per-particle network of the Deep Sets architecture

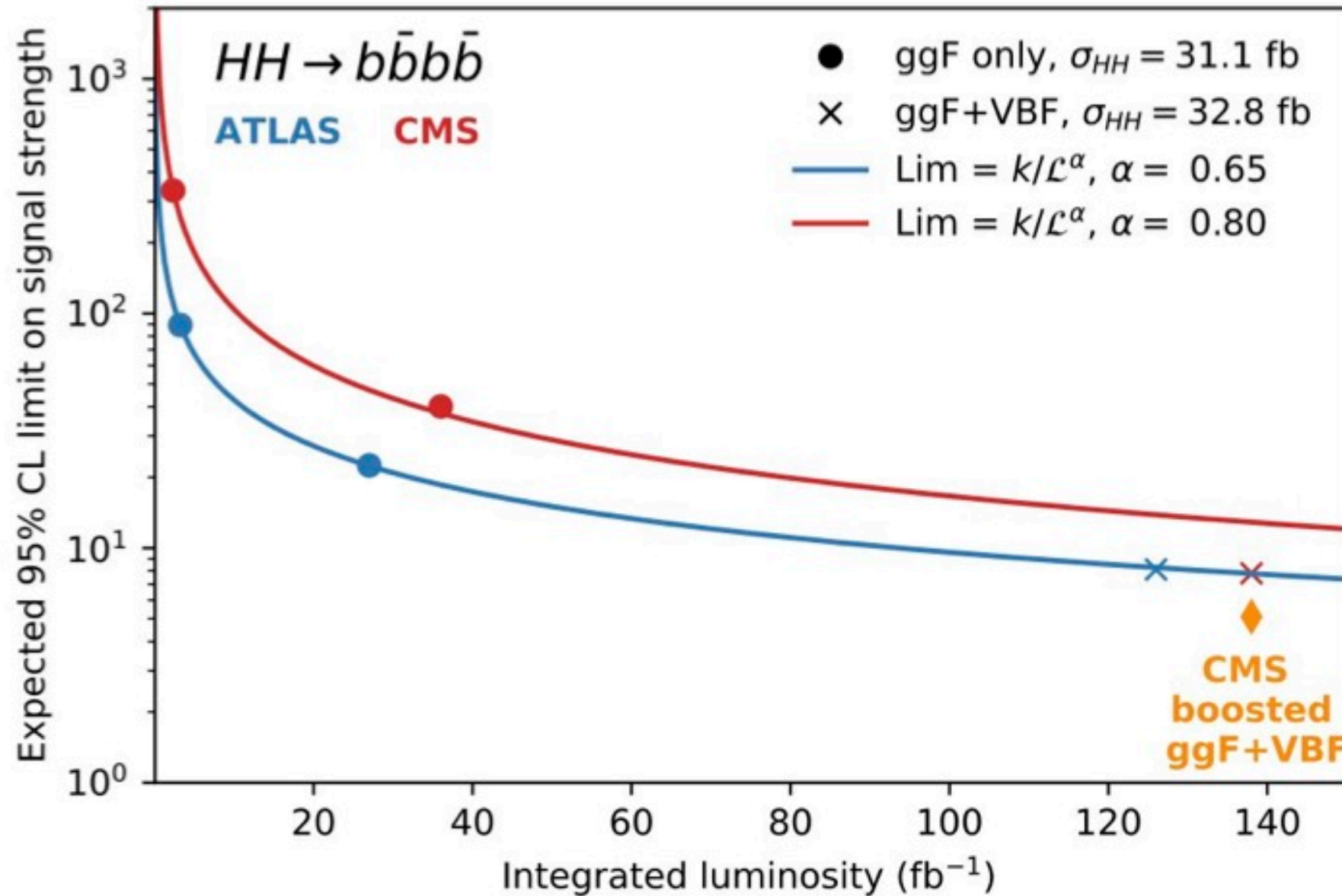


**DeepJet w/ enlarged conv:** Use the larger filters of DeepSet, but keep the RNN aggregation

∴ Similar performance, but dataset size of 130m jets, still gaining w/ the larger architectures



# Impact in the 4b analysis



baseline

RNNs  
+CNNs

Sets +  
Graphs

Trans-  
formers

# Early Run 2 status

[j.physletb.2019.135103](https://arxiv.org/abs/1905.05820)

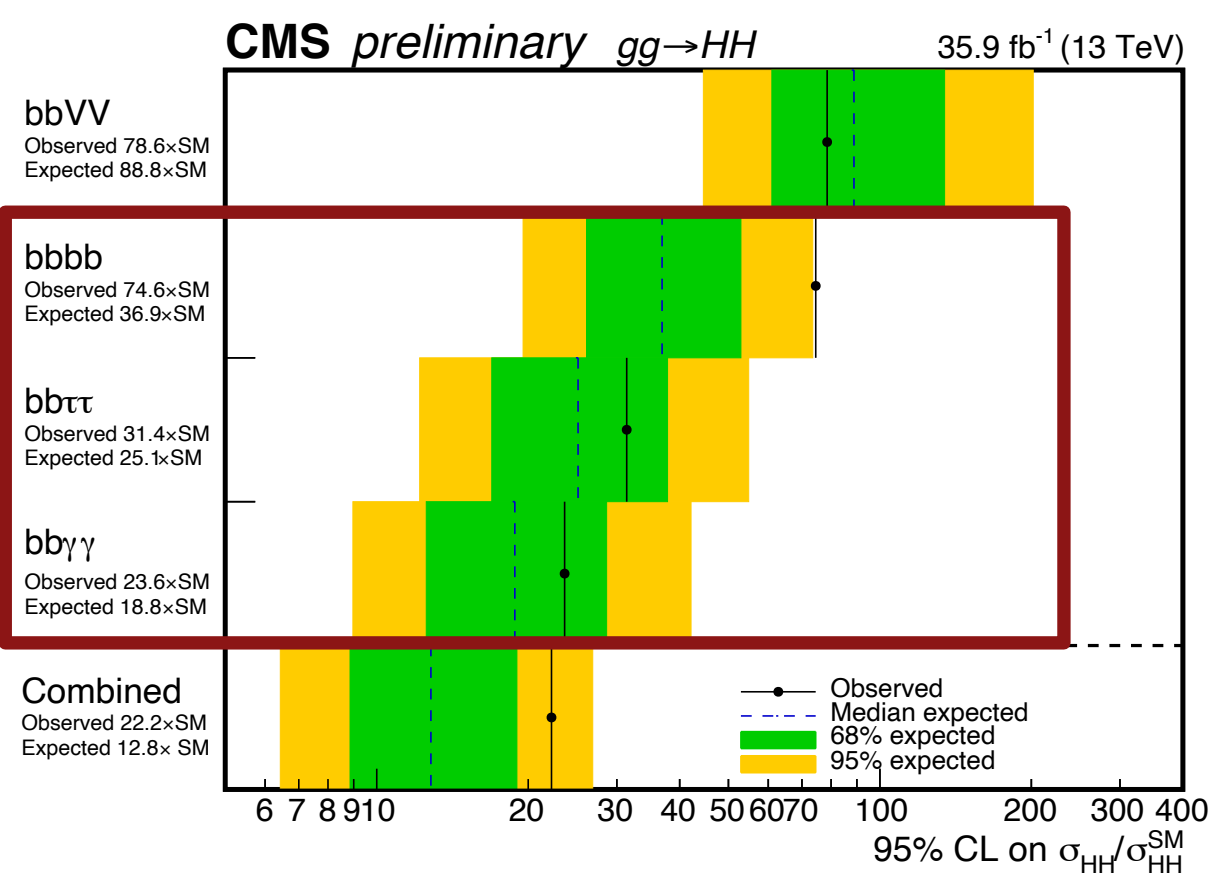
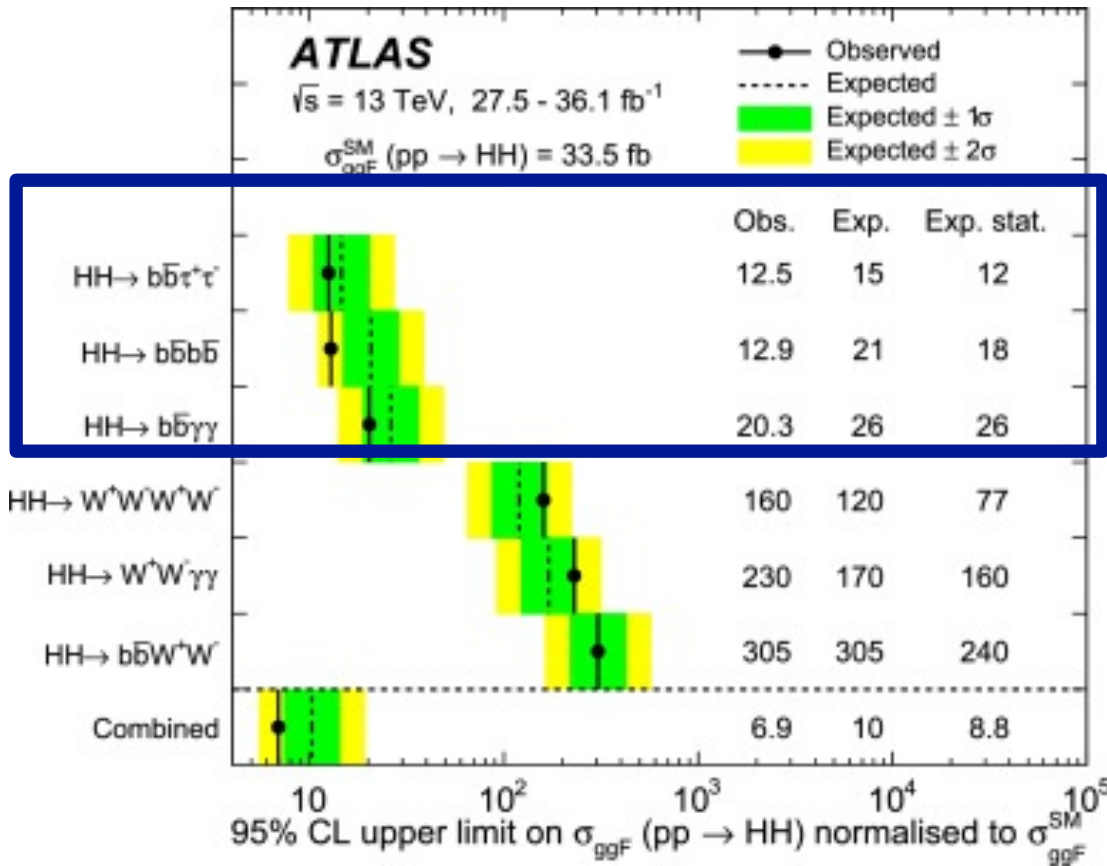
[HIG-17-030](#)

baseline

RNNs + CNNs

Sets + Graphs

Transformers



Emphasizing our main HH channels

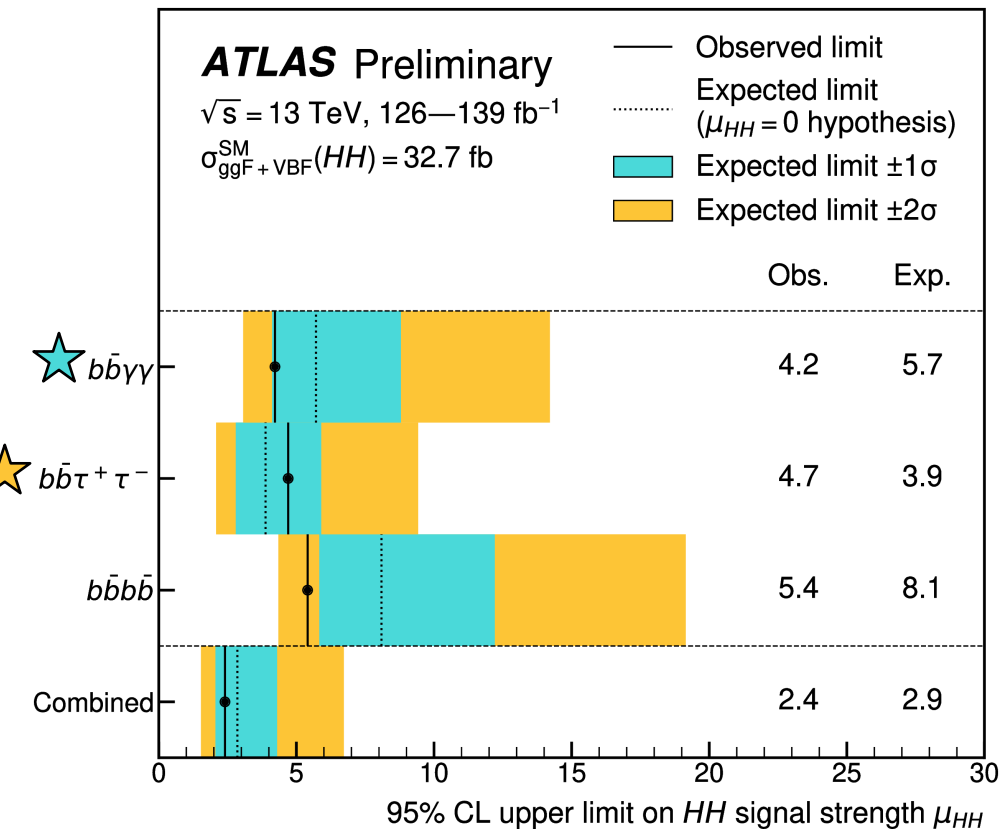
# Full Run 2 status

baseline

RNNs + CNNs

Sets + Graphs

Transformers



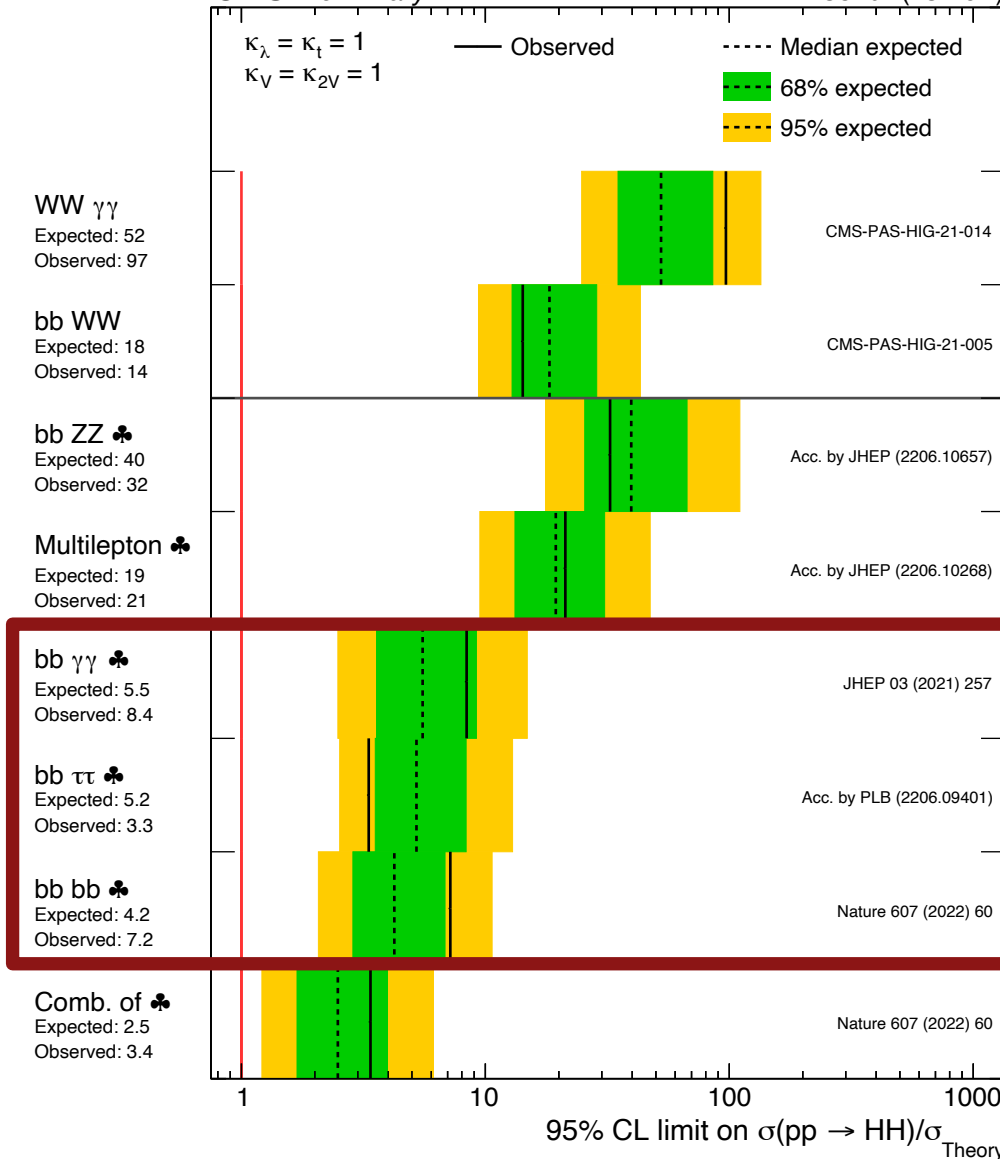
## Updated legacy Run 2 results

★  $b\bar{b}\gamma\gamma$ : 4.0 (5.0) obs (exp) [JHEP 01 \(2024\) 066](#)

★  $b\bar{b}\tau\tau$ : 5.9 (3.1) obs (exp) [ATLAS-CONF-2023-071](#)

53

## CMS Preliminary CMS HIG [twiki](#) 138 fb<sup>-1</sup> (13 TeV)



Emphasizing main HH channels

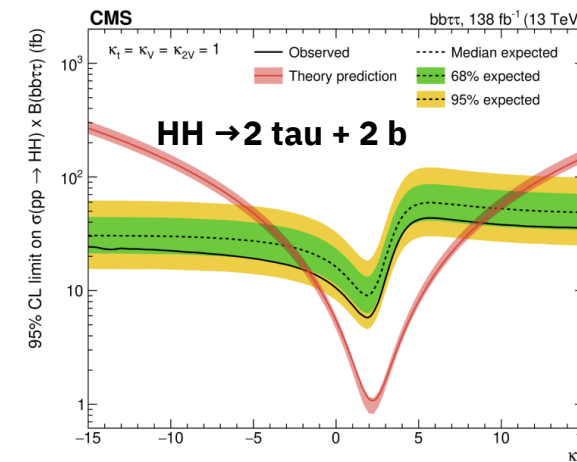
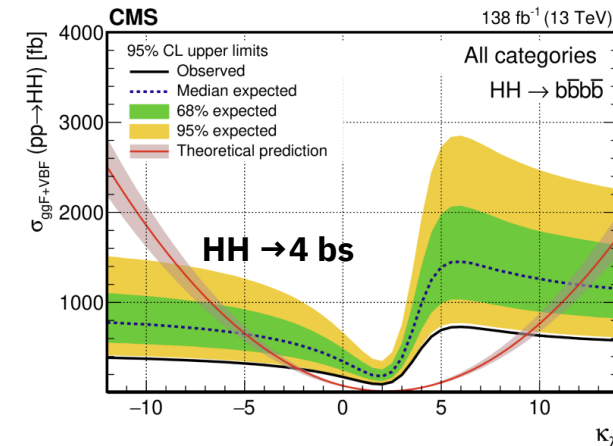
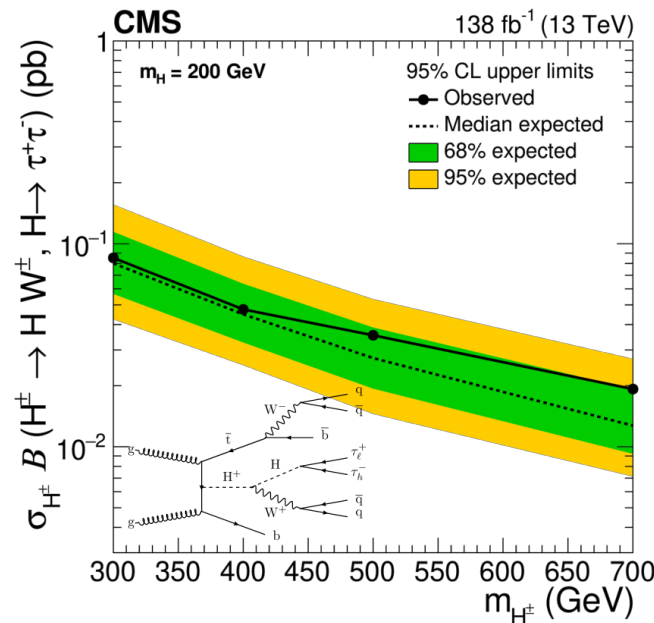
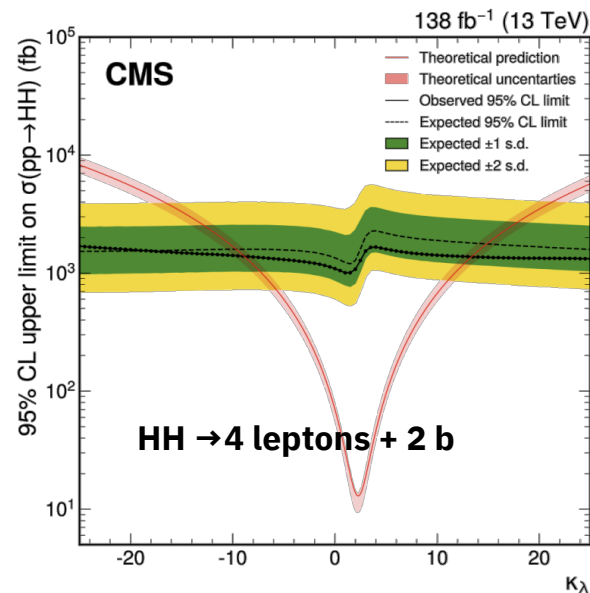


# Also enabling new physics searches!!

[HIG-20-004](#) [HIG-20-010](#)  
[HIG-21-010](#) [HIG-20-005](#)

## b-Tagging with DeepJet

- Number of HH results with resolved b-jets
- Strongest observed HH cross-section constraints
- First  $H^\pm \rightarrow HW^\pm$  results at LHC



# Graph Neural Networks

---

## Impact on analysis strategy

---

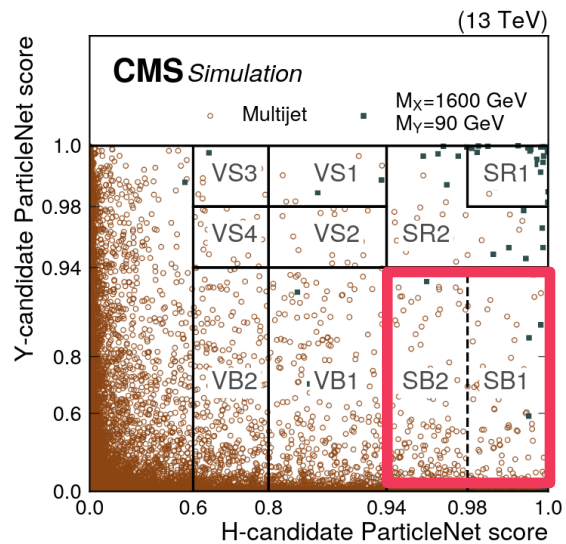
When you put *so much* ML into the Higgs-jet identification the analysis strategy simplifies dramatically!

- Stat uncertainties now dominant, no need for a detailed background description, a simple QCD normalization is (often) sufficient.
- As the QCD bkg plummets, more important to model the  $t\bar{t}$  contribution separately and constrain in dedicated CRs

# Background estimation.. boosted ParticleNet analyses

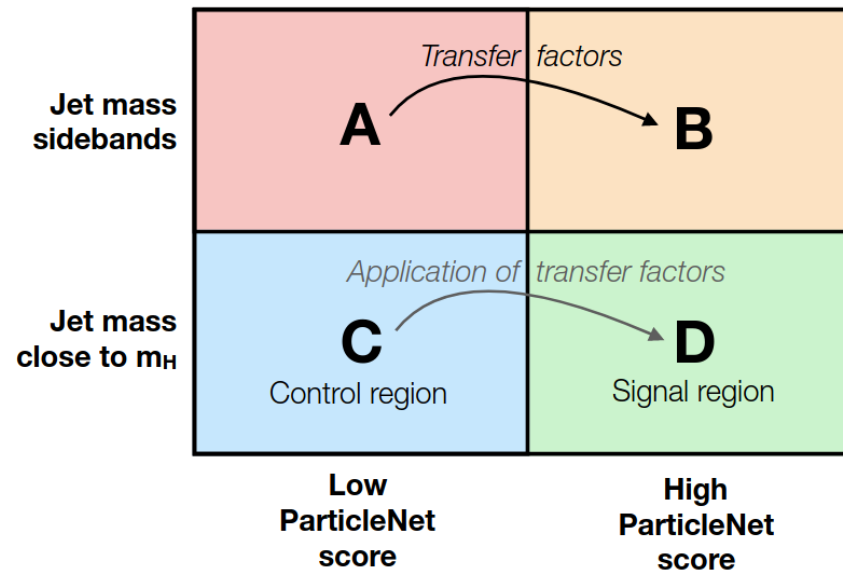
## HY resonant

[PhysLetB.2022.137392](https://arxiv.org/abs/2202.13739)

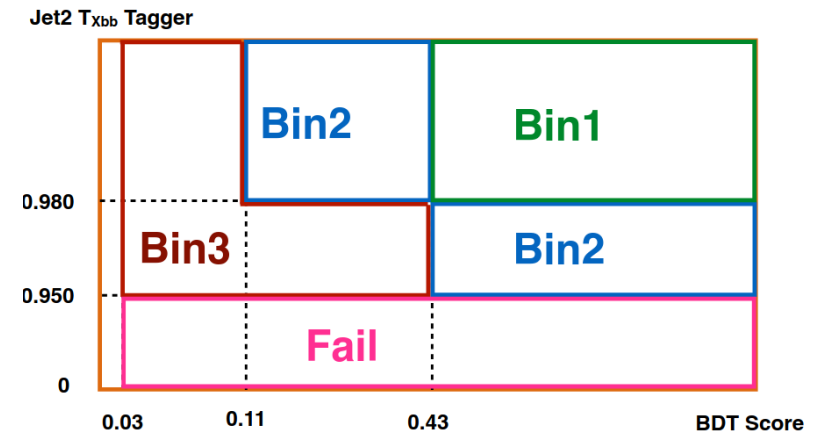


## HH NR VBF

[PhysRevLett.131.041803](https://arxiv.org/abs/131.041803)



## HH NR ggF



AK8 $m_{\text{reg}}$ region	$50 < m_{\text{reg}}^{\text{lead}} < 110 \text{ GeV}$	$110 < m_{\text{reg}}^{\text{lead}} < 150 \text{ GeV}$	$150 < m_{\text{reg}}^{\text{lead}} < 200 \text{ GeV}$
$50 < m_{\text{reg}}^{\text{subl}} < 90 \text{ GeV}$	Transfer factor regions (A & B)		
$90 < m_{\text{reg}}^{\text{subl}} < 145 \text{ GeV}$	Validation region (D)	search region (D)	Validation region (D)
$145 < m_{\text{reg}}^{\text{subl}} < 200 \text{ GeV}$	Transfer factor regions (A & B)		

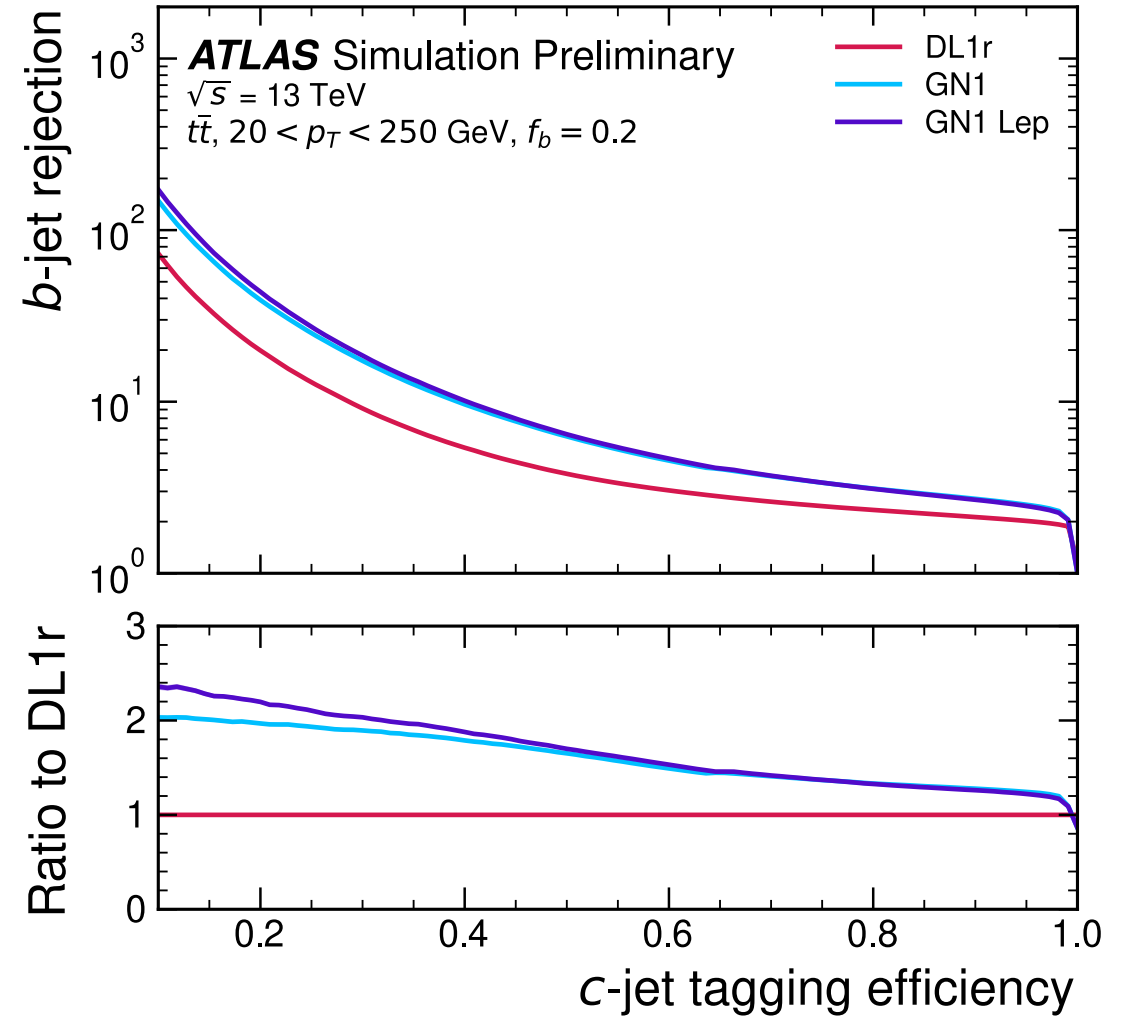
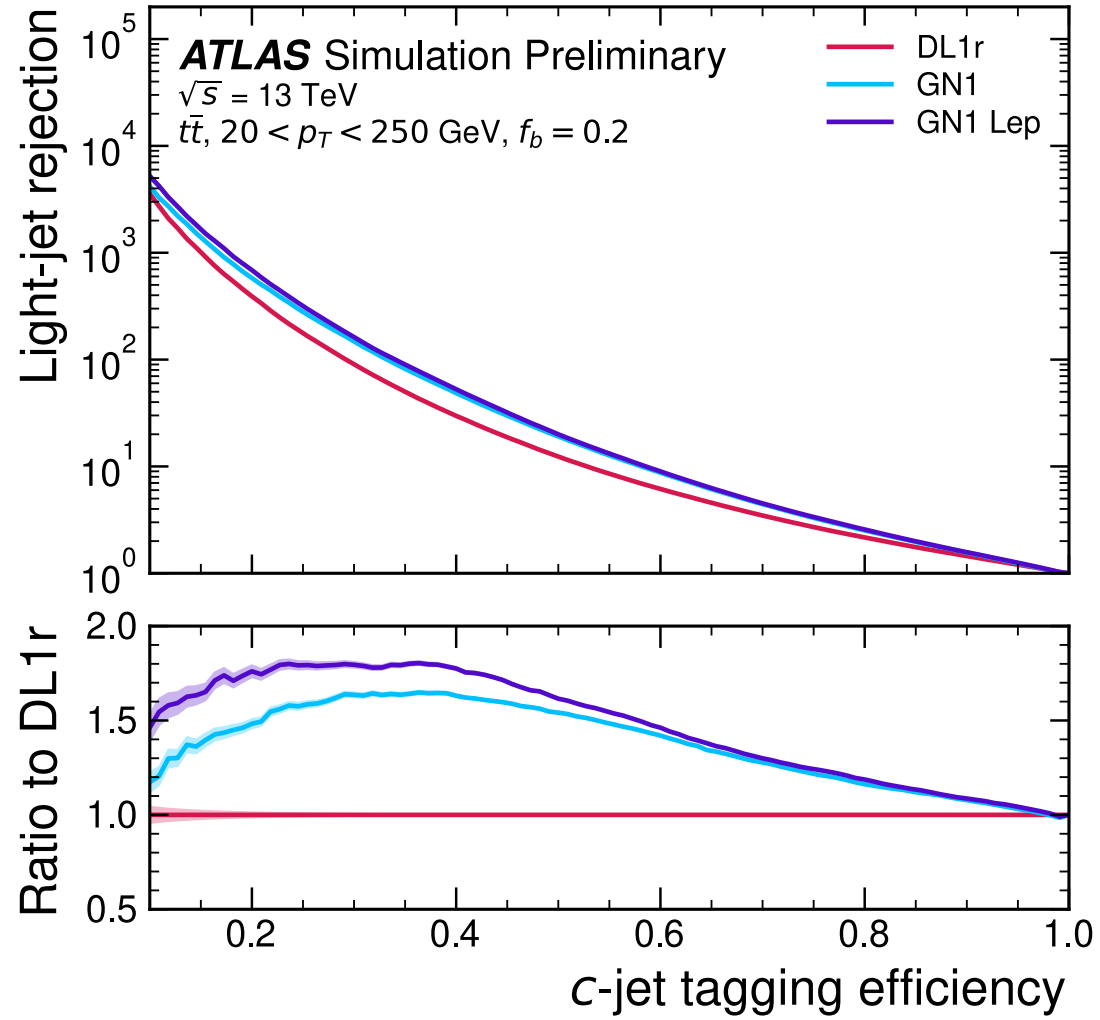


# Errors

Uncertainty source	$\Delta\mu$	
Statistical	+2.55	-2.30
Signal extraction	+2.32	-2.06
QCD multijet modeling	+1.12	-1.01
$t\bar{t}$ modeling	+0.28	-0.19
Systematic	+2.09	-0.89
Simulated sample size	+0.55	-0.55
$D_{b\bar{b}}$ selection	+0.72	-0.32
Jet energy and mass scale and resolution	+0.54	-0.39
Trigger selection	+0.26	-0.03
Luminosity measurement	+0.13	-0.04
Pileup modeling	+0.05	-0.06
Other experimental uncertainties	+0.05	-0.03
Theoretical	+0.63	-0.63
Total	+3.30	-2.47

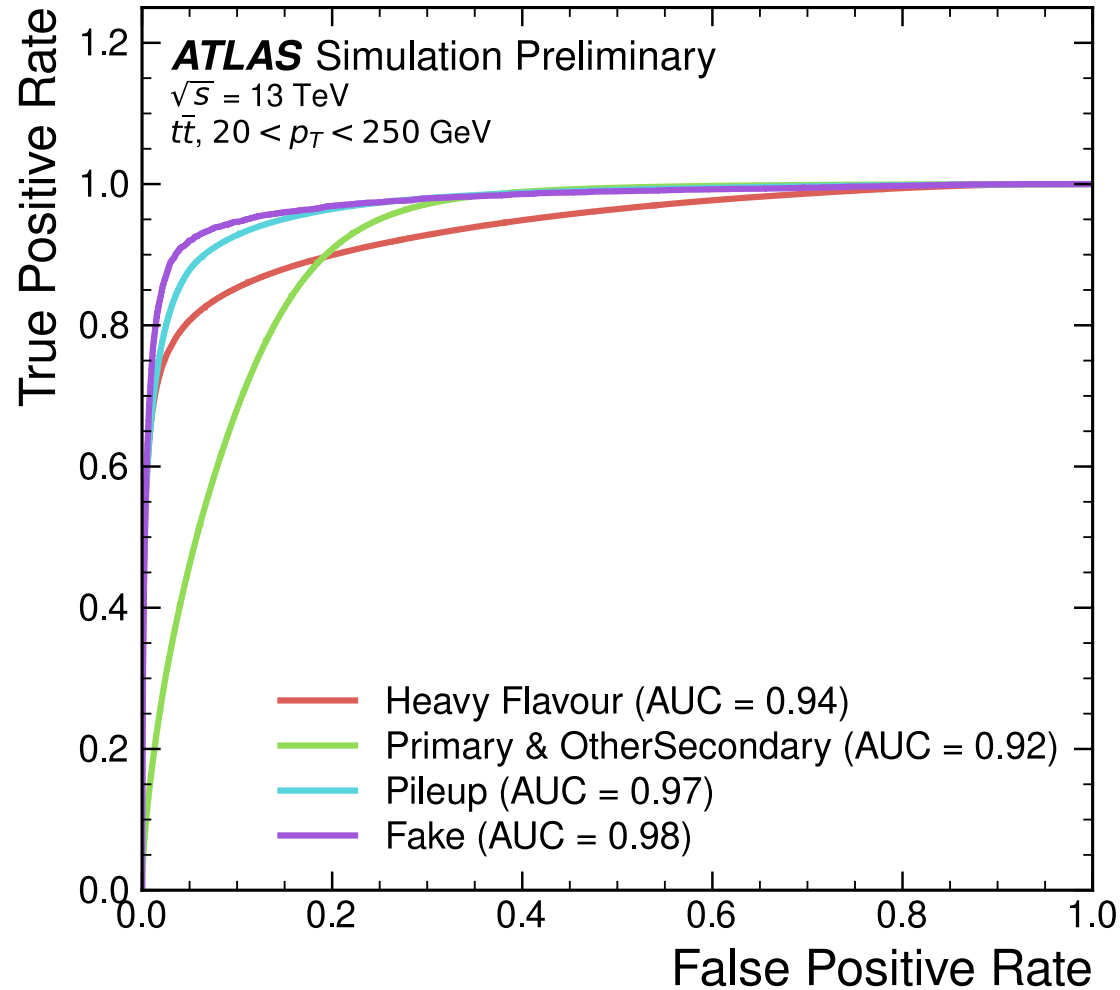
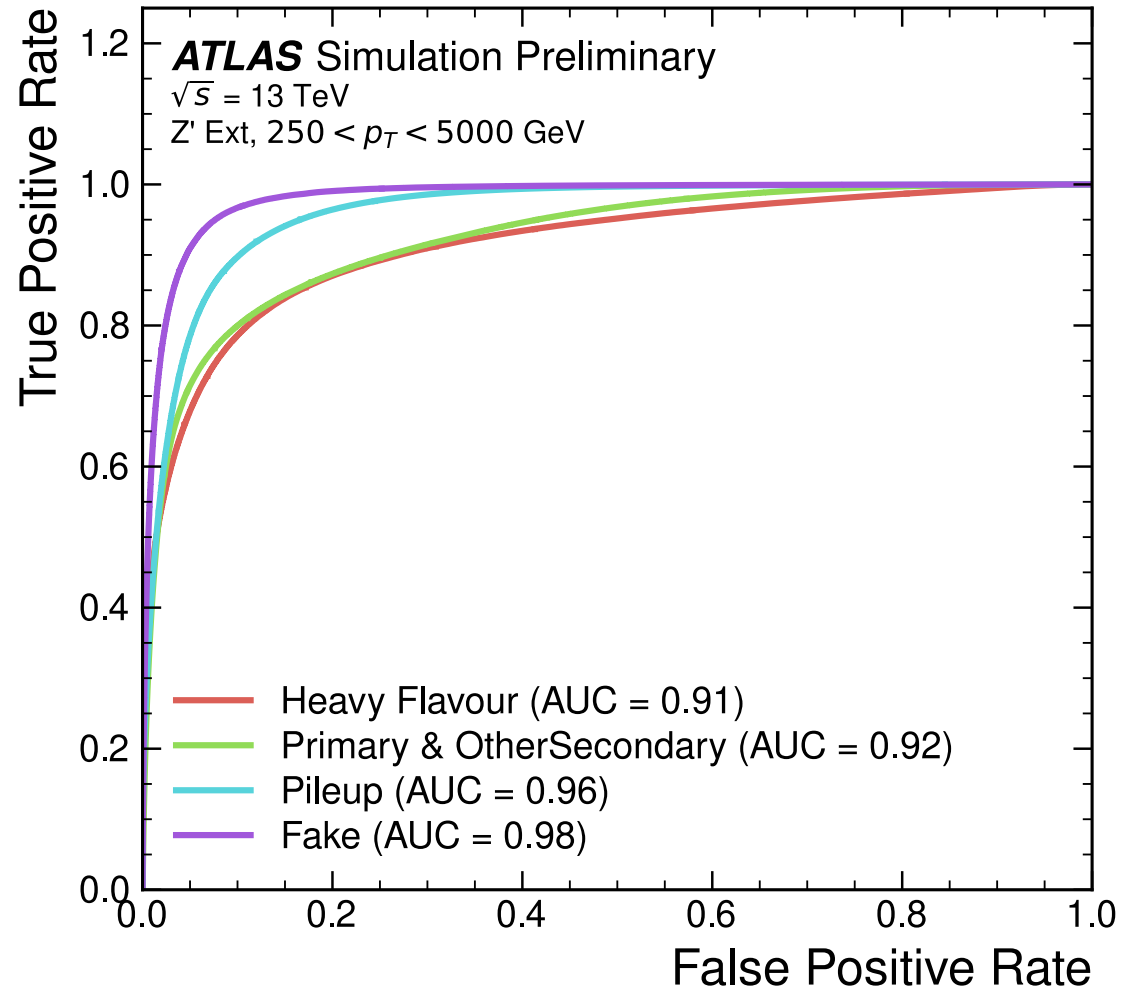
# GN1 performance

baseline

RNNs  
+CNNsSets +  
GraphsTrans-  
formers

# GN1: Track classification

baseline

RNNs  
+CNNsSets +  
GraphsTrans-  
formers
 $t\bar{t}$ 

 $Z' \rightarrow q\bar{q}$ 


# GN1: track classification

baseline

RNNs + CNNs

Sets + Graphs

Transformers

**ATLAS Simulation Preliminary**

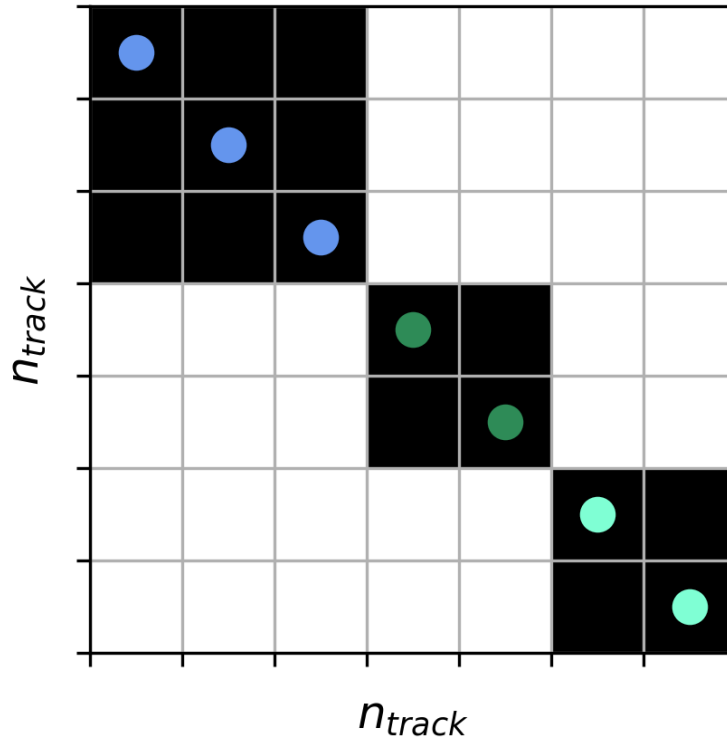
$\sqrt{s} = 13 \text{ TeV}$

$t\bar{t}$  jets

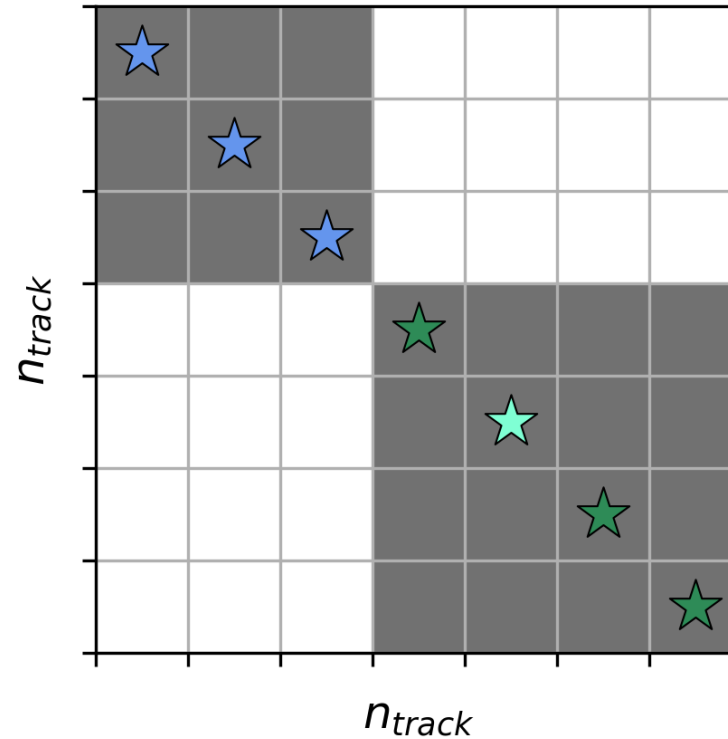
**Truth  $b$ -jet**  
 $p_T = 32.2 \text{ GeV}$

$p_b = 0.997$   
 $p_c = 0.003$   
 $p_u = 0.000$

Truth Labels



GN1 Prediction



- Truth
- ★ Predicted
- Pileup
- Fake
- Primary
- FromB
- FromBC
- FromC
- FromTau
- OtherSecondary

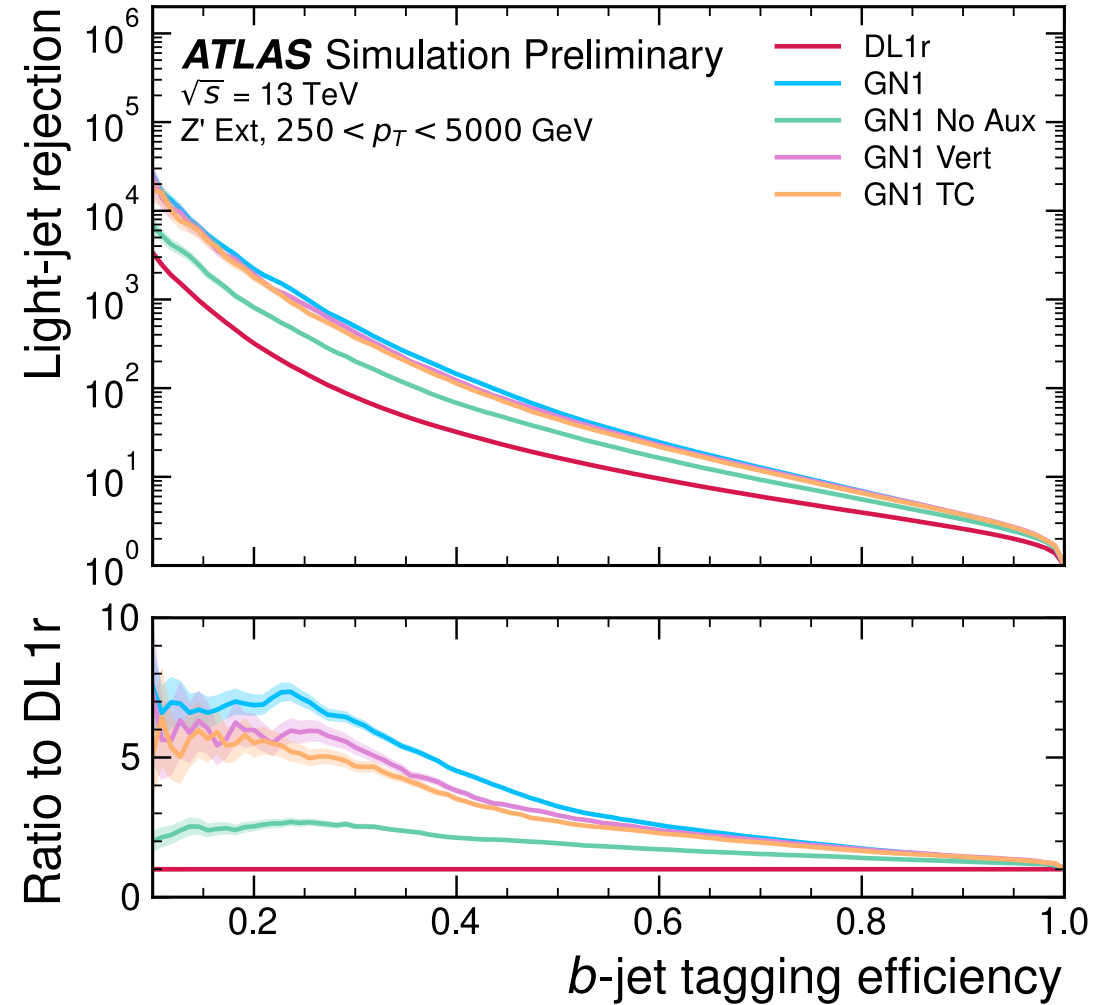
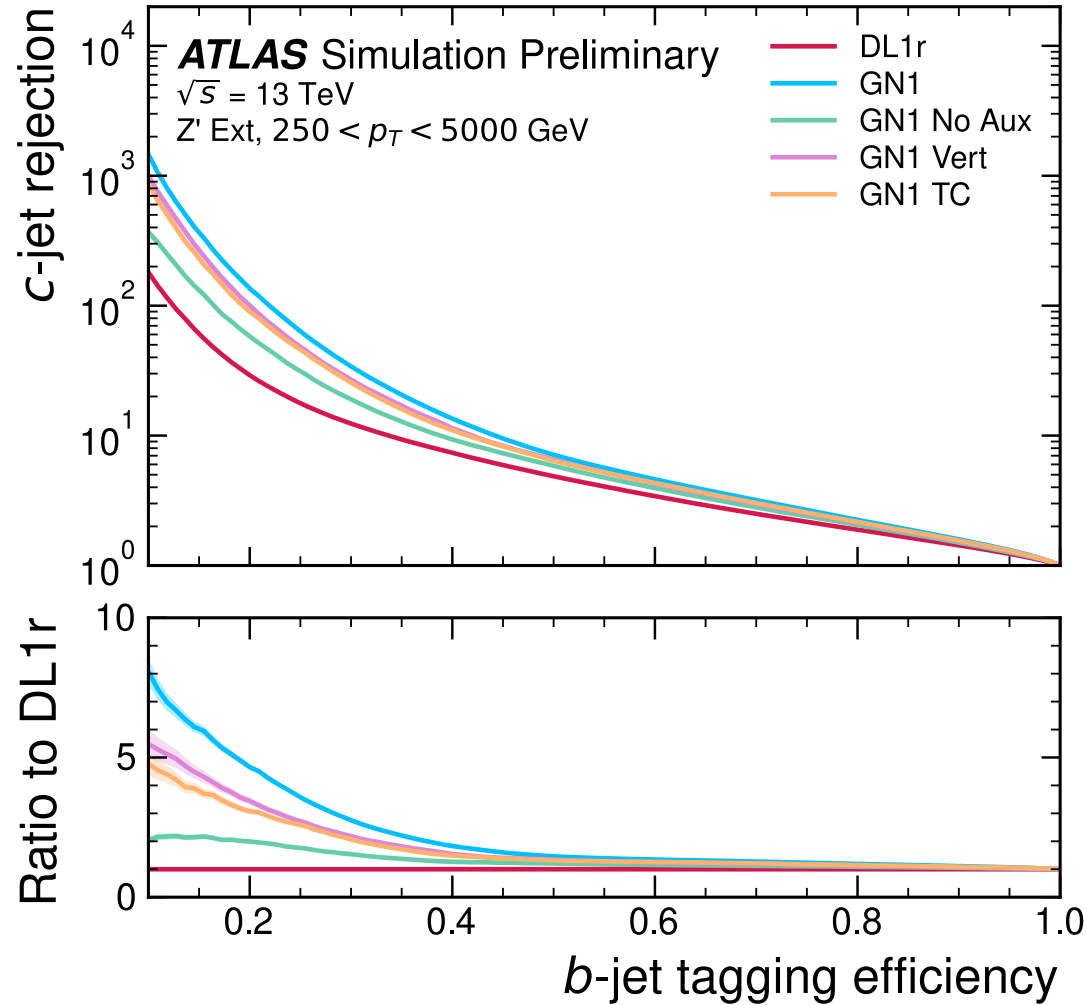
# GN1: inputs

baseline

RNNs  
+CNNsSets +  
GraphsTrans-  
formers

Jet Input	Description
$p_T$	Jet transverse momentum
$\eta$	Signed jet pseudorapidity
Track Input	Description
$q/p$	Track charge divided by momentum (measure of curvature)
$d\eta$	Pseudorapidity of the track, relative to the jet $\eta$
$d\phi$	Azimuthal angle of the track, relative to the jet $\phi$
$d_0$	Closest distance from the track to the PV in the longitudinal plane
$z_0 \sin \theta$	Closest distance from the track to the PV in the transverse plane
$\sigma(q/p)$	Uncertainty on $q/p$
$\sigma(\theta)$	Uncertainty on track polar angle $\theta$
$\sigma(\phi)$	Uncertainty on track azimuthal angle $\phi$
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0)$	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
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes

# GN1: auxiliary tasks

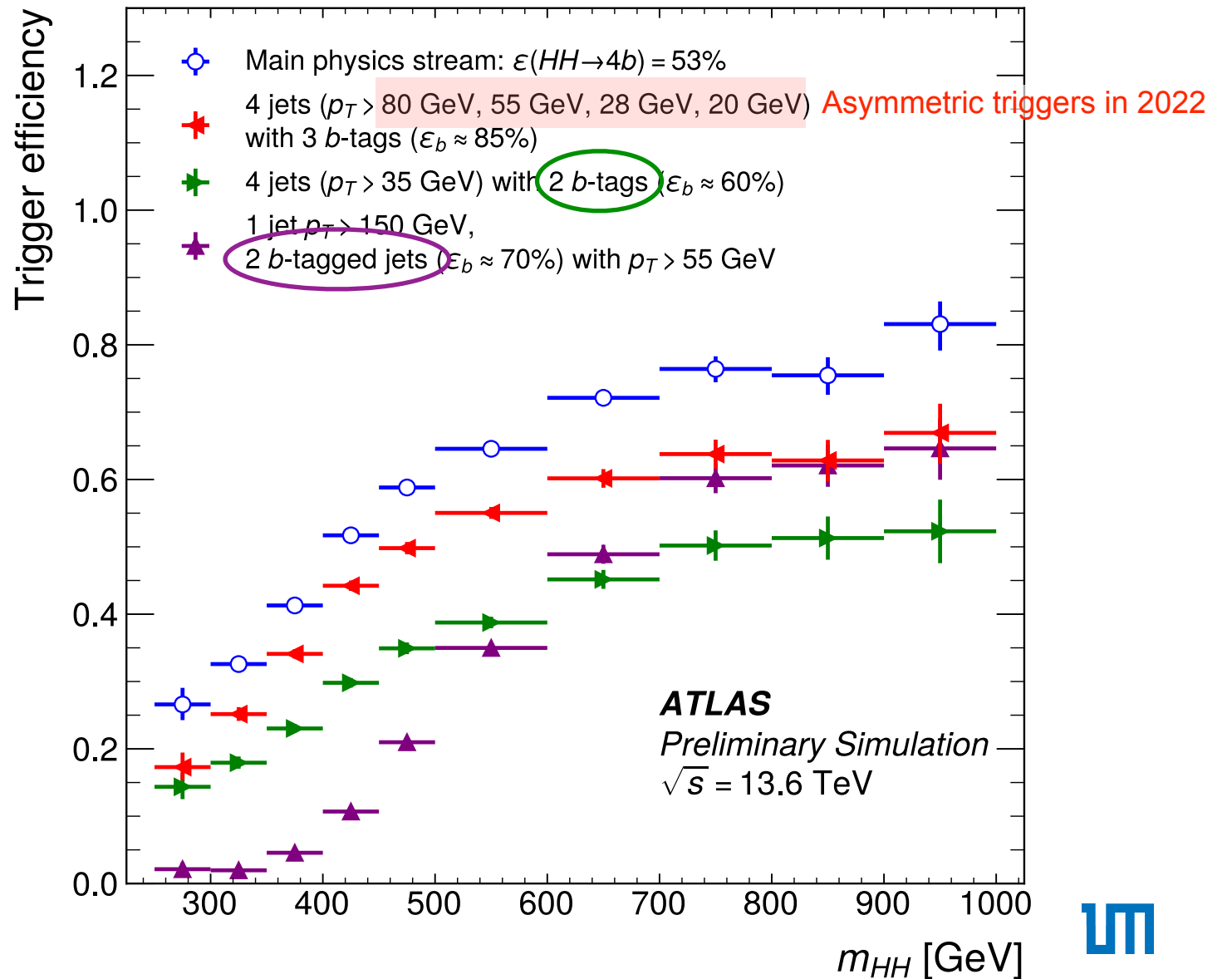


baseline

RNNs  
+CNNsSets +  
GraphsTrans-  
formers

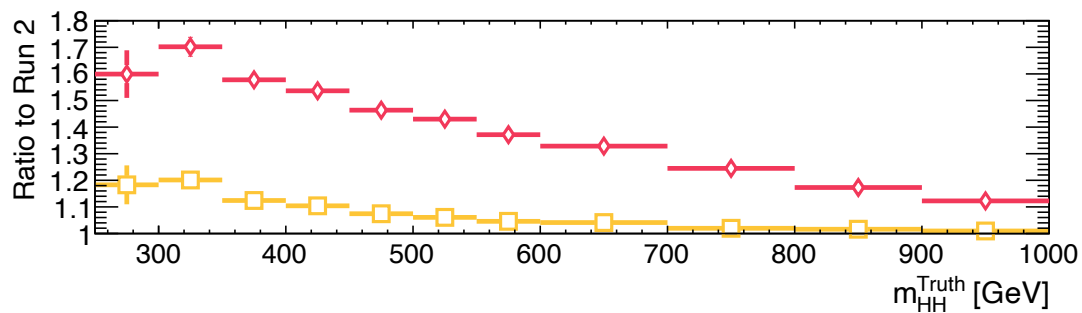
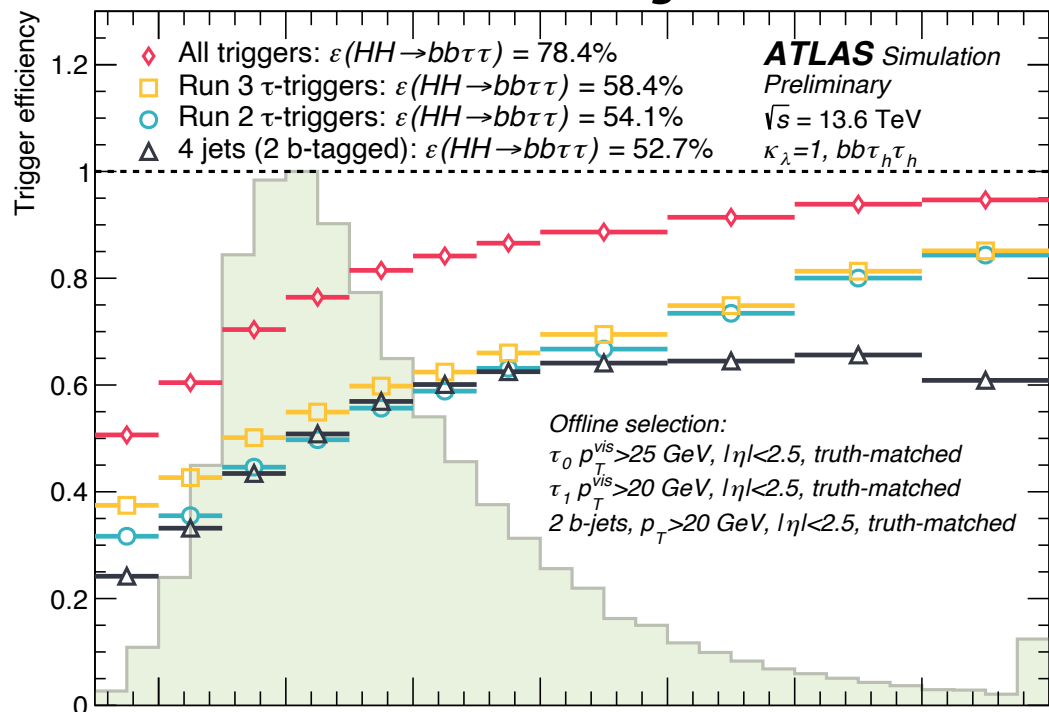
# Triggers

The composition of triggers making up the main physics stream

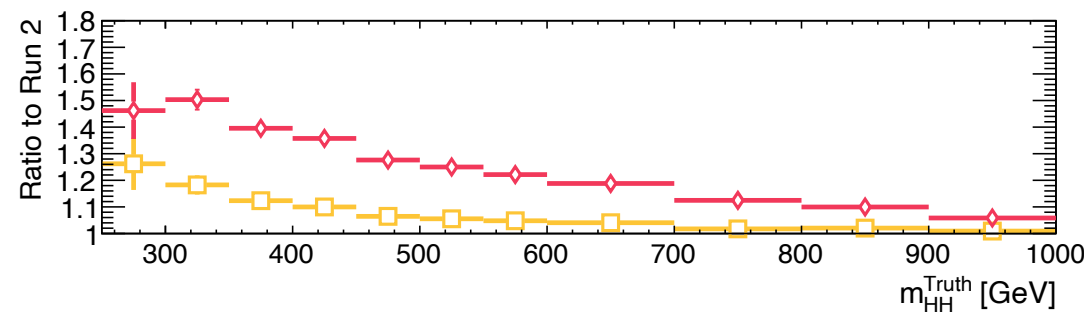
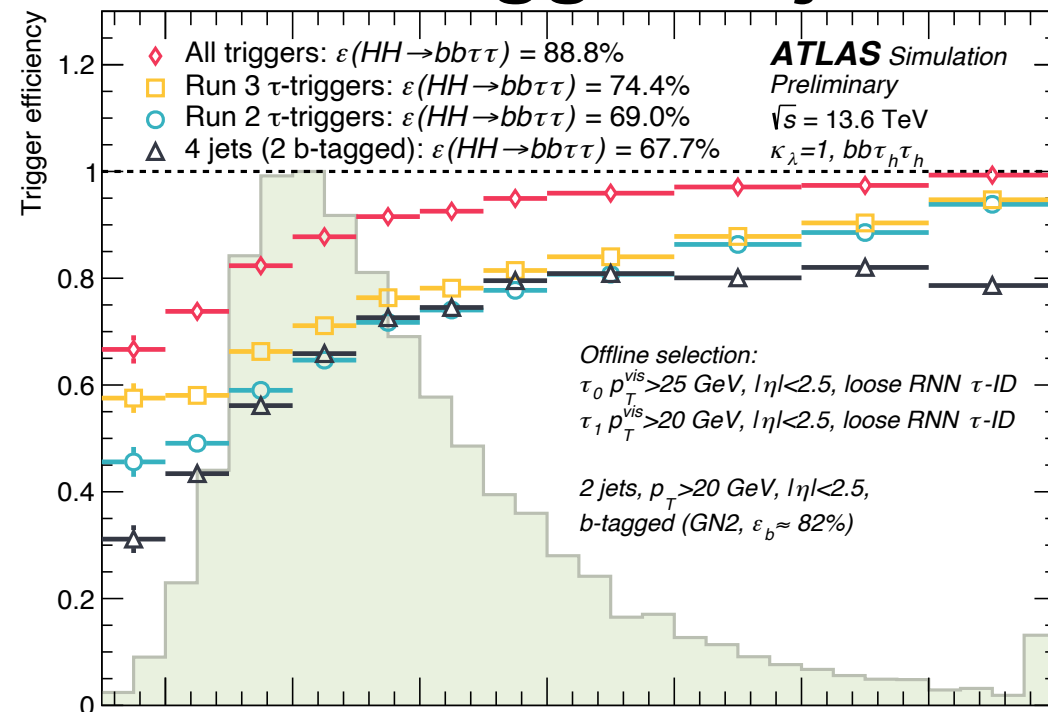


# HH $\rightarrow$ bb $\tau_{\text{had}}\tau_{\text{had}}$

## Truth b-jets

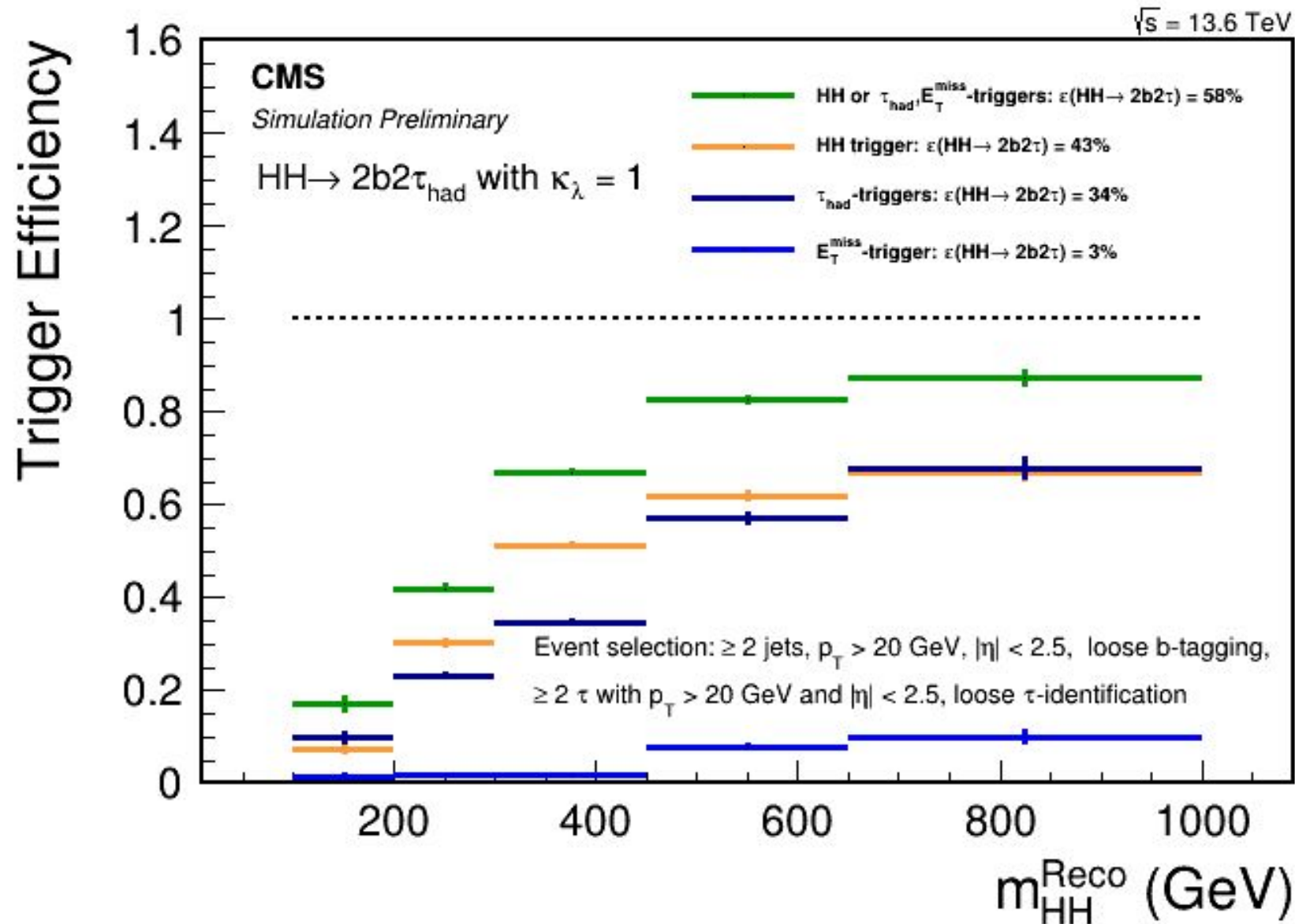


## GN2-tagged b-jets





# CMS: bbtatau trigger




# Transformers

ATLAS: GN2

CMS: ParticleTransformer (ParT)

 **You**

What would be useful for experimentalists and theorists to discuss at a workshop for "Combined Experiment Theory Action"

 **ChatGPT**

A workshop on "Combined Experiment Theory Action" would likely focus on fostering collaboration and synergy between experimentalists and theorists in the field of particle physics. Here are some topics that could be useful for discussion:

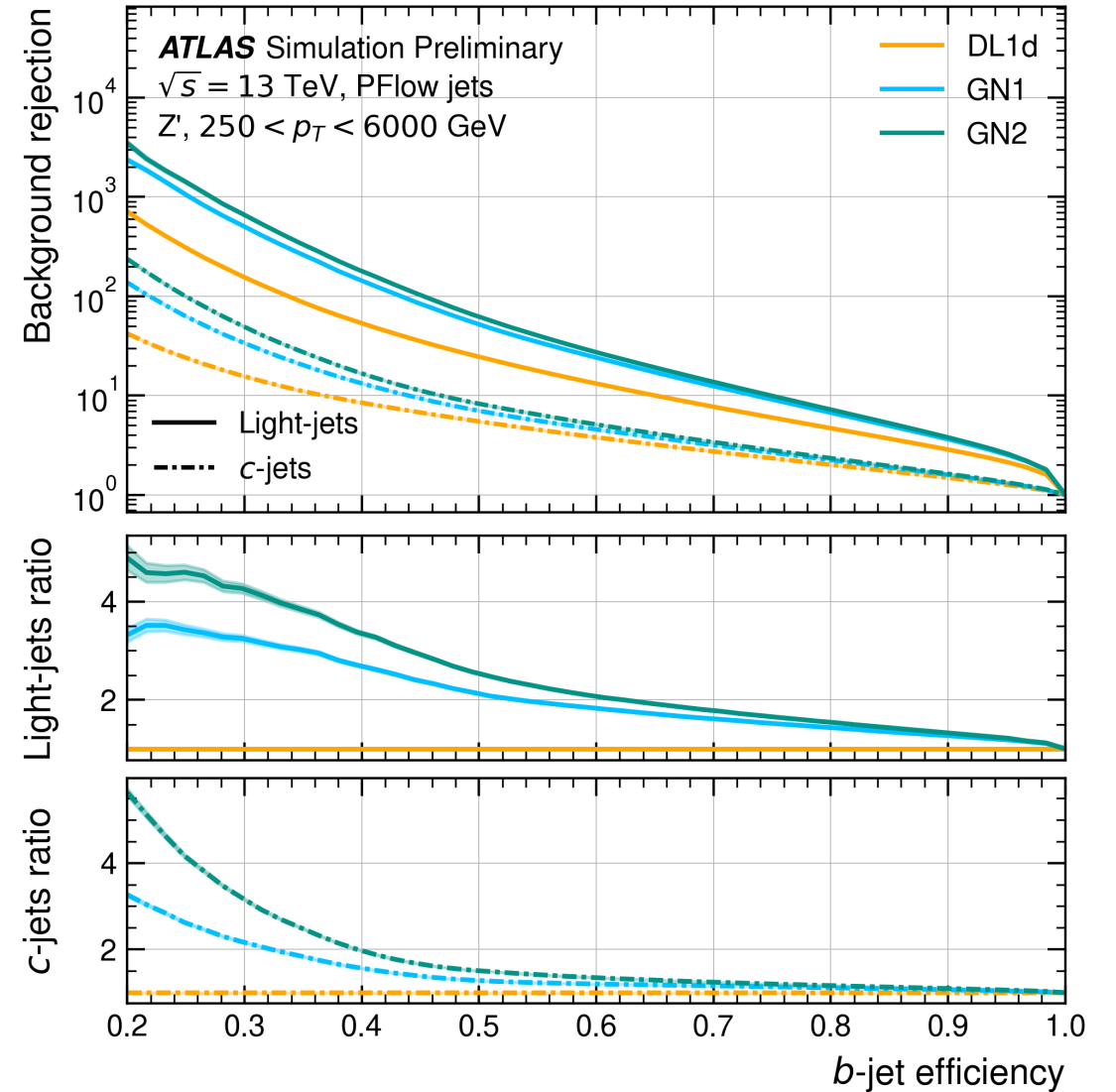
1. **Current Challenges and Open Questions:** Discussing the current challenges and open questions in particle physics that require a combined effort from experimentalists and theorists to address. This could include topics such as the nature of dark matter, the properties of neutrinos, the search for new particles beyond the Standard Model, and the origin of matter-antimatter asymmetry.
2. **Experimental Constraints and Theoretical Frameworks:** Exploring how experimental data can constrain theoretical models and vice versa. This involves understanding the limitations of experimental measurements and how theoretical predictions can guide experimental searches.
3. **Data Analysis Techniques:** Sharing best practices in data analysis techniques used by both experimentalists and theorists. This could include statistical methods, machine learning approaches, and simulation techniques.
4. **Model Building and Phenomenology:** Discussing how theoretical models are constructed and how they can be tested against experimental data. This involves considering both established frameworks like the Standard Model and speculative theories beyond it.
5. **Collider and Non-collider Experiments:** Highlighting the different types of experiments in particle physics, including collider experiments like those at the Large Hadron Collider (LHC) and non-collider experiments such as neutrino experiments and dark matter searches.
6. **Computational Tools and Resources:** Sharing computational tools and resources that are useful for both experimentalists and theorists. This could include software packages for Monte Carlo simulations, theoretical calculations, and data analysis.
7. **Interdisciplinary Collaborations:** Exploring opportunities for interdisciplinary collaborations with researchers from other fields, such as astrophysics, cosmology, and condensed matter physics. These collaborations can provide new insights and perspectives on particle physics phenomena.
8. **Education and Training:** Discussing strategies for educating and training the next generation of particle physicists, with an emphasis on interdisciplinary skills and collaboration.

By discussing these topics and fostering collaboration between experimentalists and theorists, a workshop on "Combined Experiment Theory Action" can help advance our understanding of fundamental physics and address some of the most pressing questions in the field.

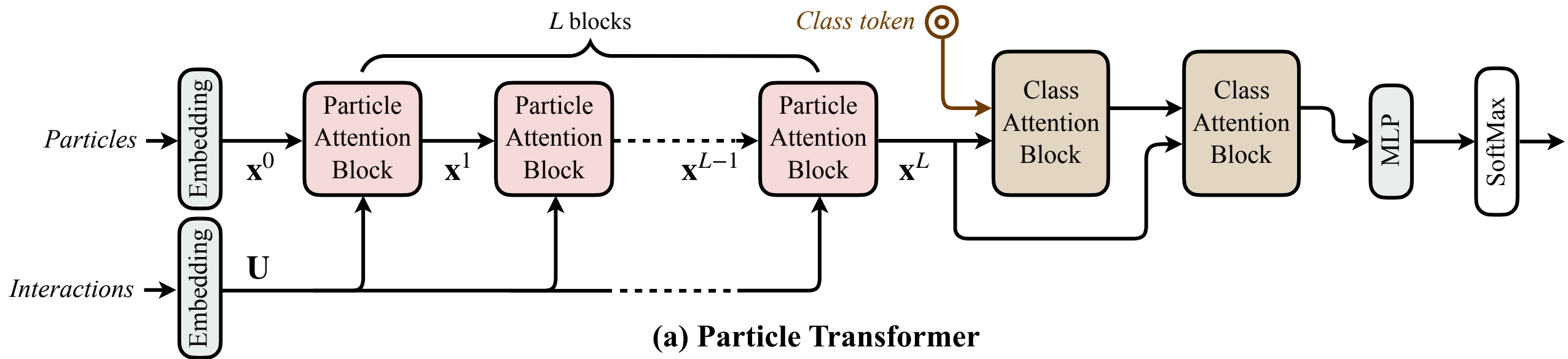


# GN2: Z' eval

baseline

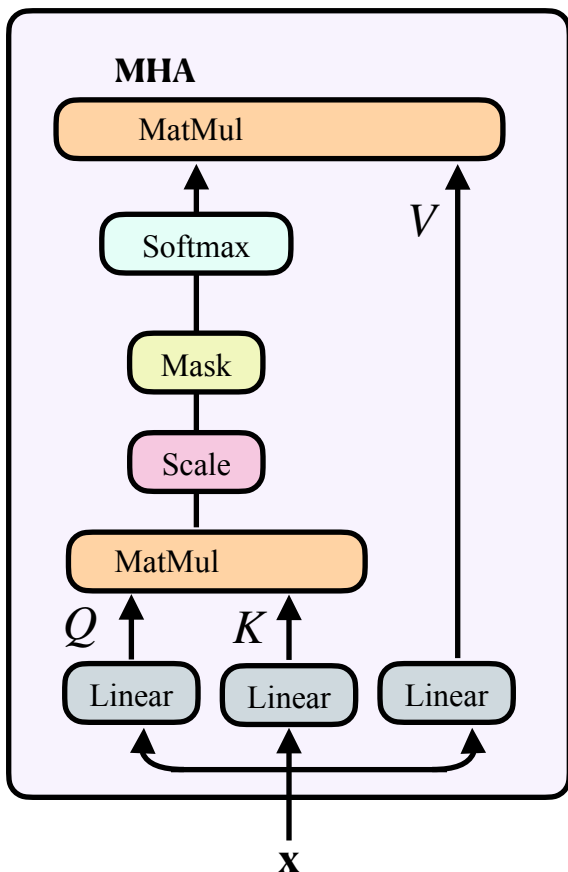
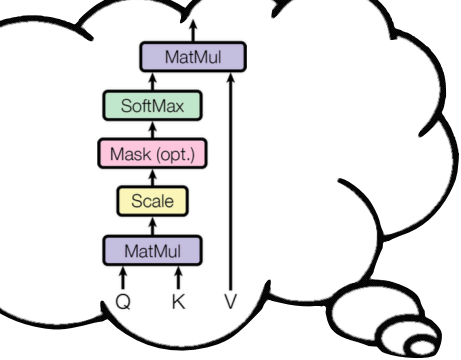
RNNs  
+CNNsSets +  
GraphsTrans-  
formers

# Particle Transformer



# Particle Attention Attention Head

## Attention Head



bias for physics-based pairwise features

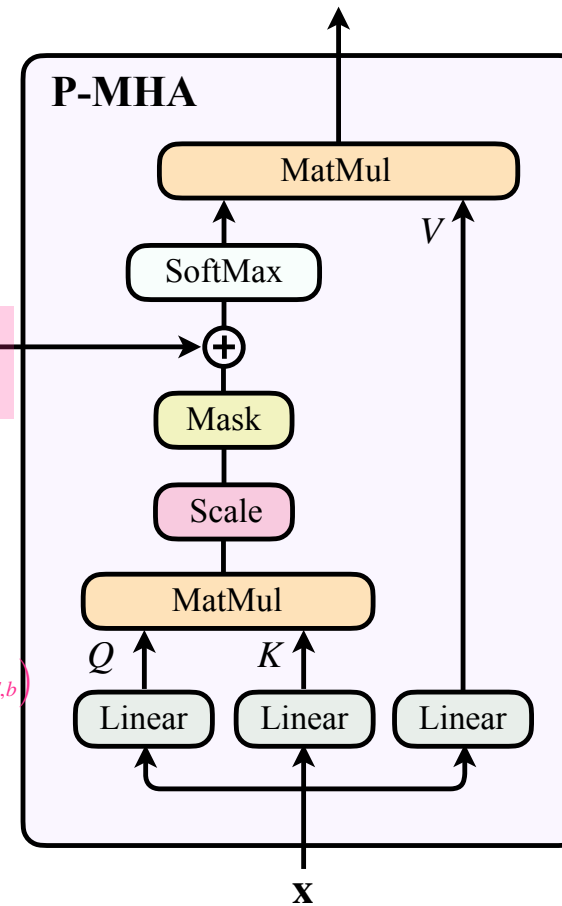
U

$$\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2}$$

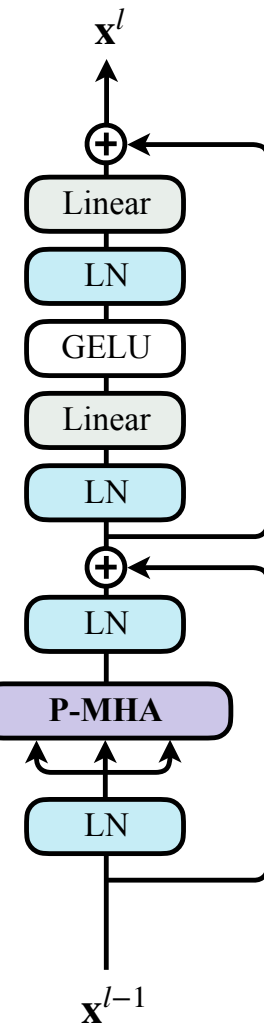
$$k_T = \min(p_{T,a}, p_{T,b}) \Delta$$

$$z = \min(p_{T,a}, p_{T,b}) / (p_{T,a} + p_{T,b})$$

$$m = (E_a + E_b)^2 - \|\mathbf{p}_a + \mathbf{p}_b\|^2$$

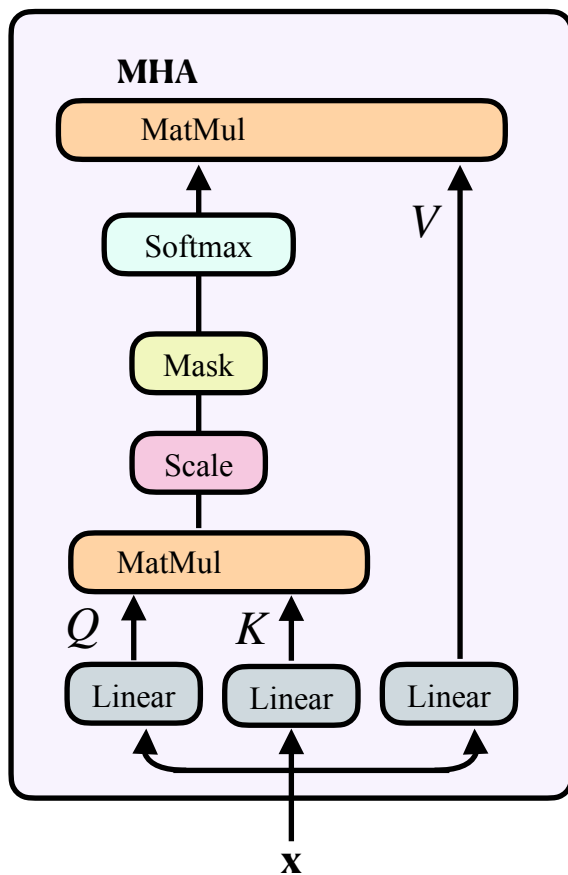


U



(b) Particle Attention Block

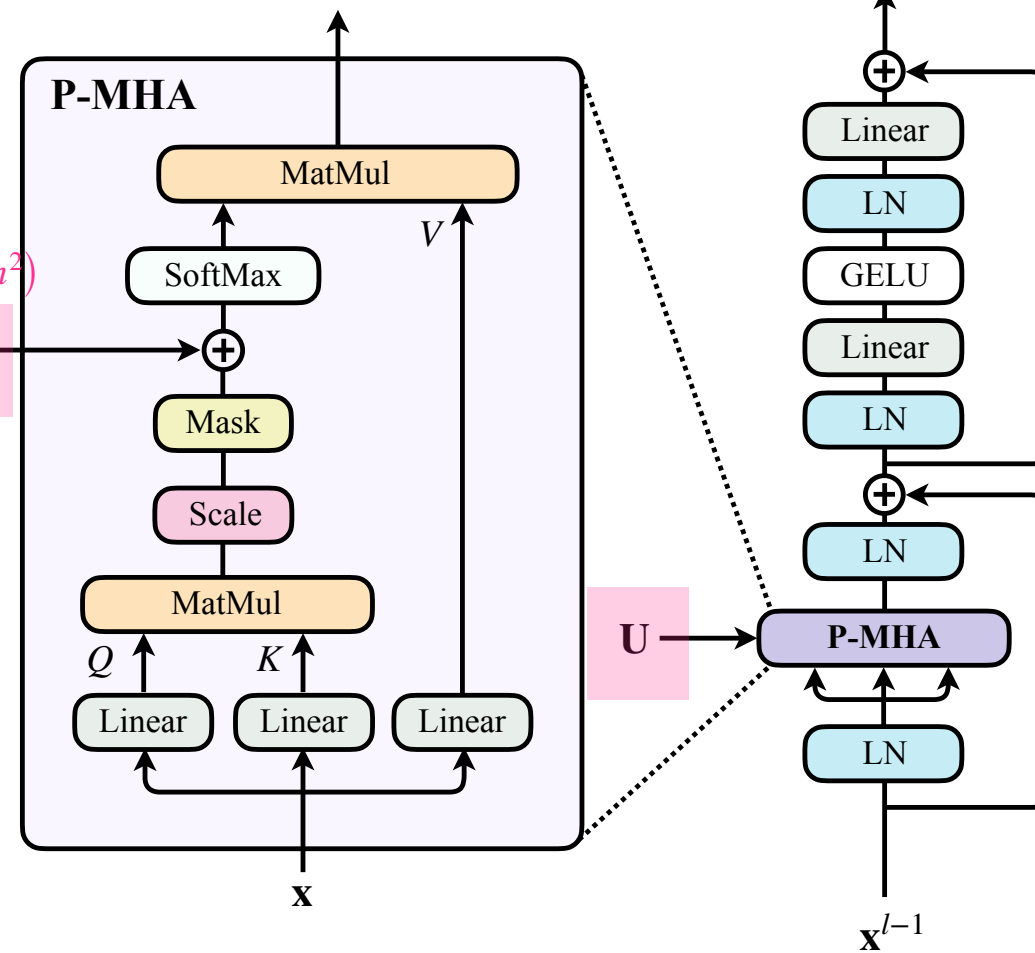
## Attention Head



## 🥰 ParT Attention Head

**bias**  
for physics-based  
pairwise features:  
( $\log \Delta$ ,  $\log k_T$ ,  $\log z$ ,  $\log m^2$ )

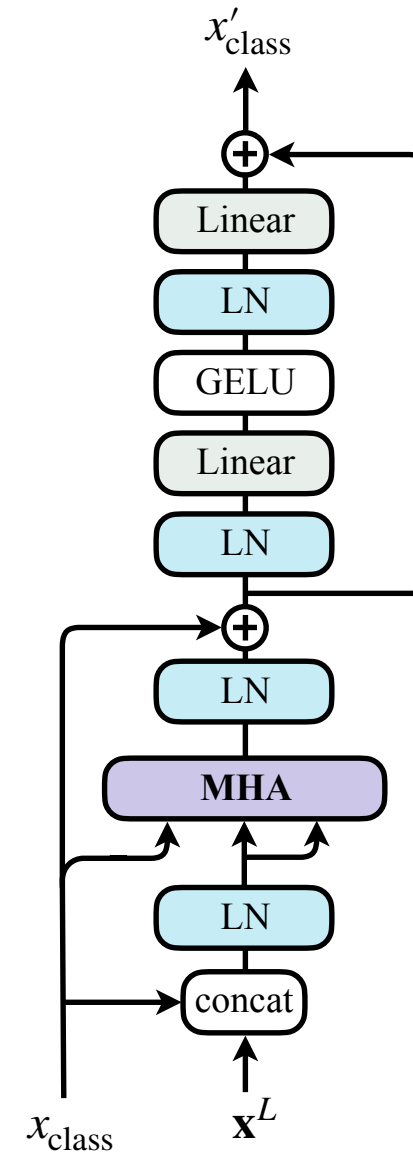
**U**



(b) Particle Attention Block

# Class Attention Block

- Attention between the target classes and the jet latent space
  - Conceptually similar to **slot attention**
    - 2006.15055
    - Used for global PFlow studies, 2212.01328
- For Xbb tagging, expected not as big of a difference as the particle attention block



(c) Class Attention Block