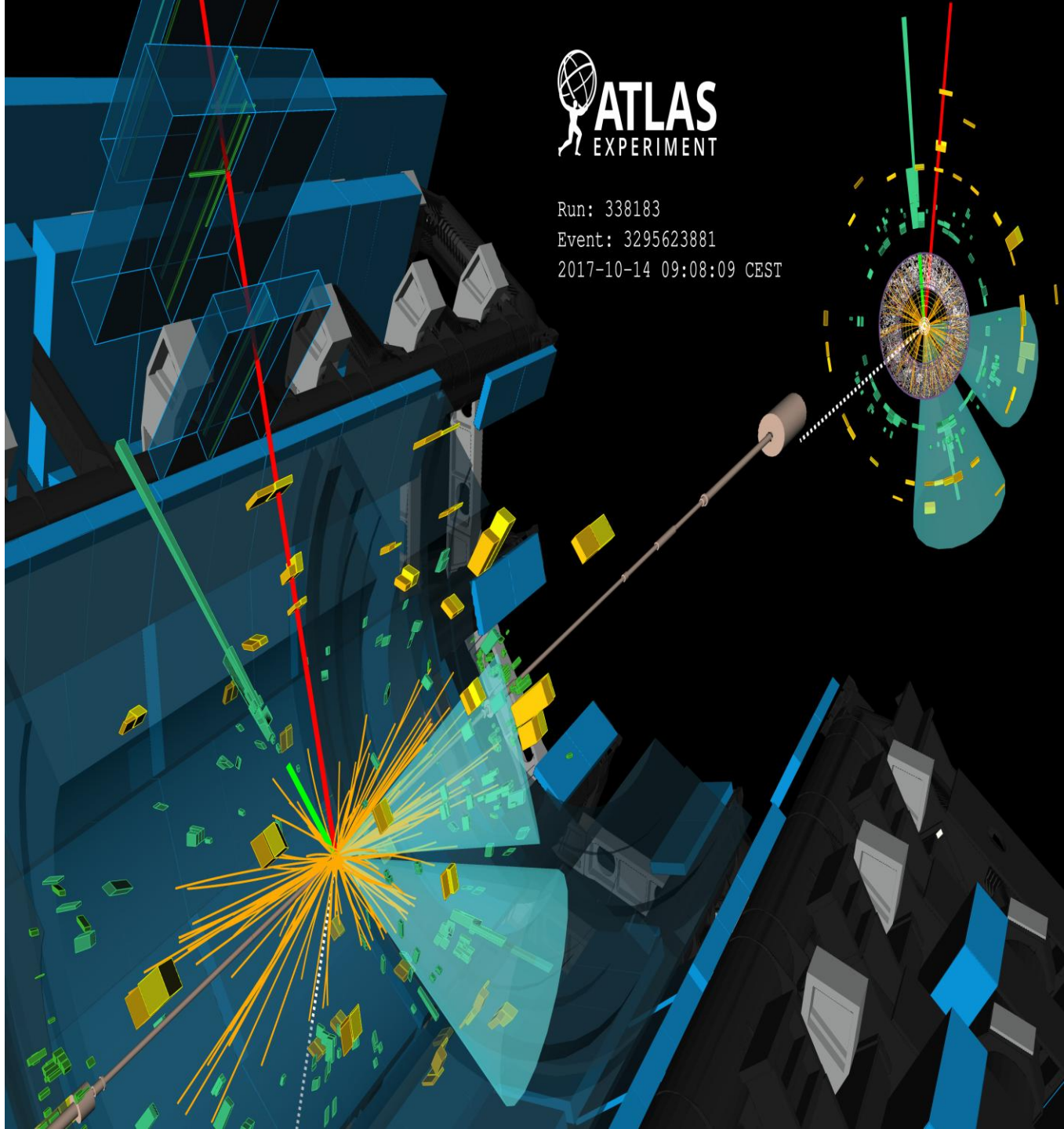
# Unravelling the origins of top pairs: production mode analysis in ATLAS

## DESY Summer School final presentation

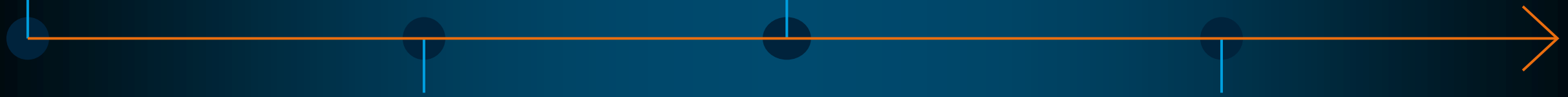
Valerio Tinari
ATLAS group

Supervised by: *K. Behr, E. Jones, M. Gonzalez , F. Jolly*

11/09/2025



# Table of contents




**Top quark**
What is it? What do we know about it?

**Dataset**
Producing the bricks for our building

**DNN implementation**
Implementing a MultiVariate Analysis

**Conclusions and outlook**
Could this help with toponium?

# Top quark

## What is it?

m = 172.52 GeV/c²

q = +2/3

**t**

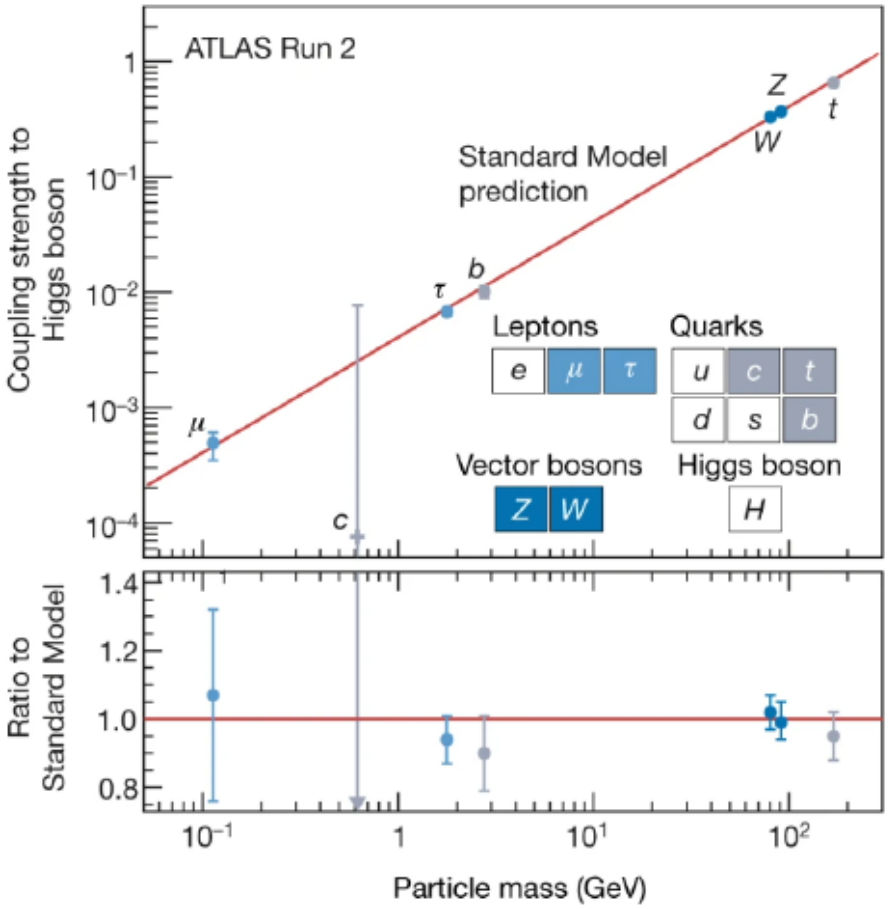
**top**

The most massive quark, 40 times more than the previous one (b quark); strongly coupled to Higgs boson

$$\tau_t \sim 10^{-25} s < \text{hadronisation time}$$

Phys. Rev. D 110, 030001 (2024)

Nature 6075 52-59 (2022)

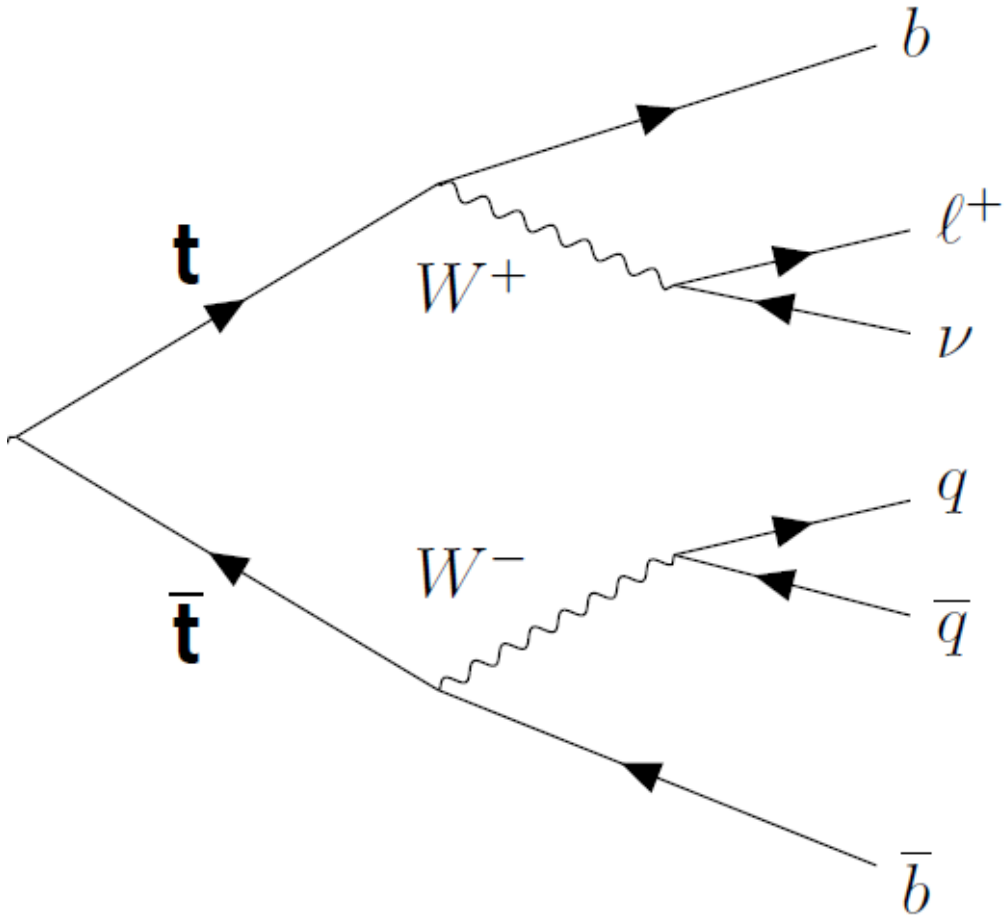


# Top quark

## How does the top decay?

$\tau_t \sim 10^{-25} s <$ hadronisation time

Phys. Rev. D 110, 030001 (2024)



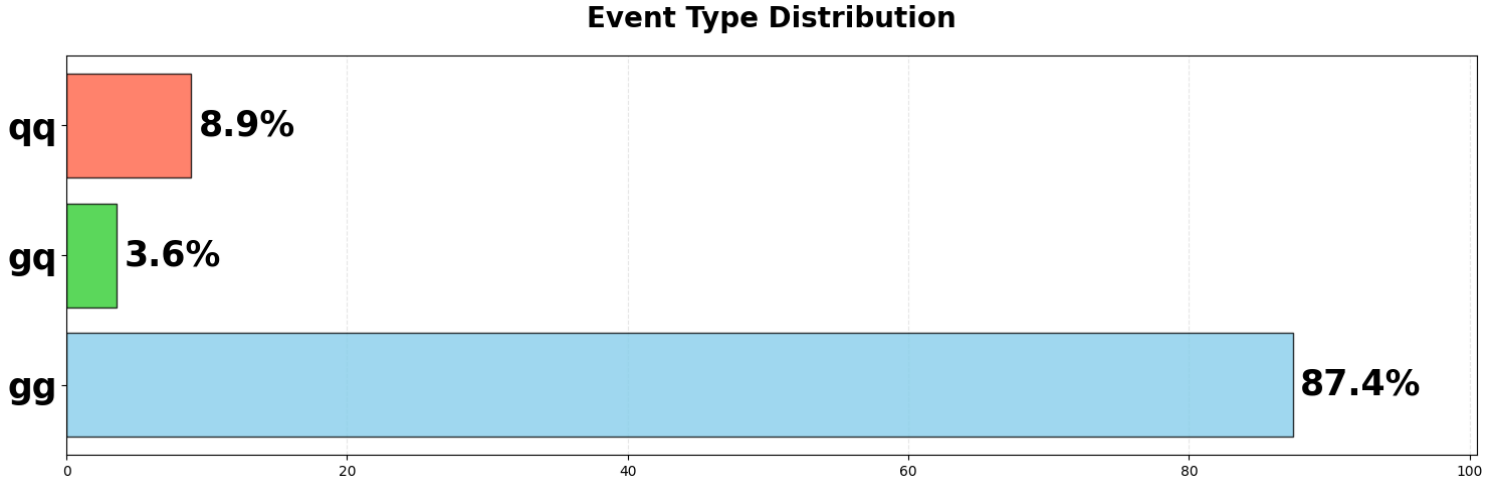

- Semi-Leptonic (1L) decay: one W decays hadronically, the other leptonically
- Detector signature: ≥ 4 jets + MET + 1 lepton
- Jet Tagging using SPANet model → Leptonic t and Hadronic t

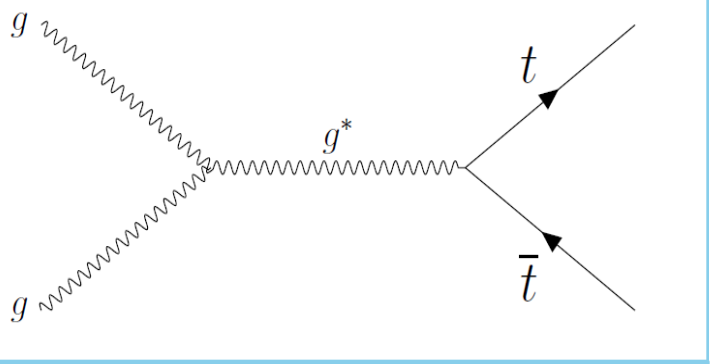
SciPost Phys. 12, 178 (2022)

# Top quark

## How is top produced?

At LO in QCD:



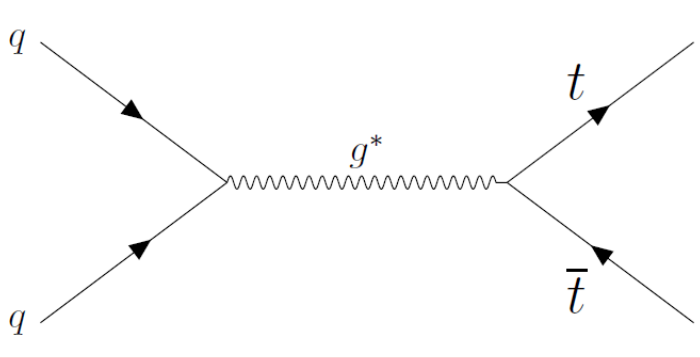




**gg channel**

Most abundant one

10% attractive colour singlet $^1S_0^{[1]}$

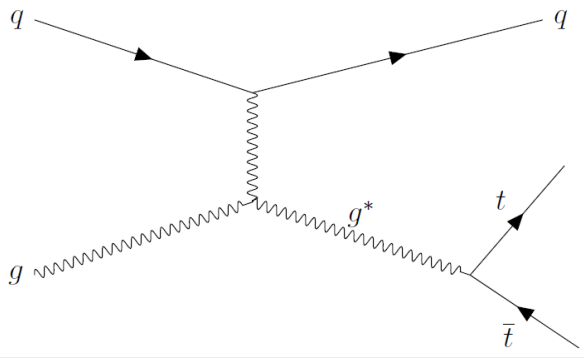
90% repulsive colour octet $^1S_0^{[8]}$




**qq channel**

Final state similar to gg

100% repulsive colour octet $^3S_0^{[8]}$




**gq channel**

Different final state, more jets

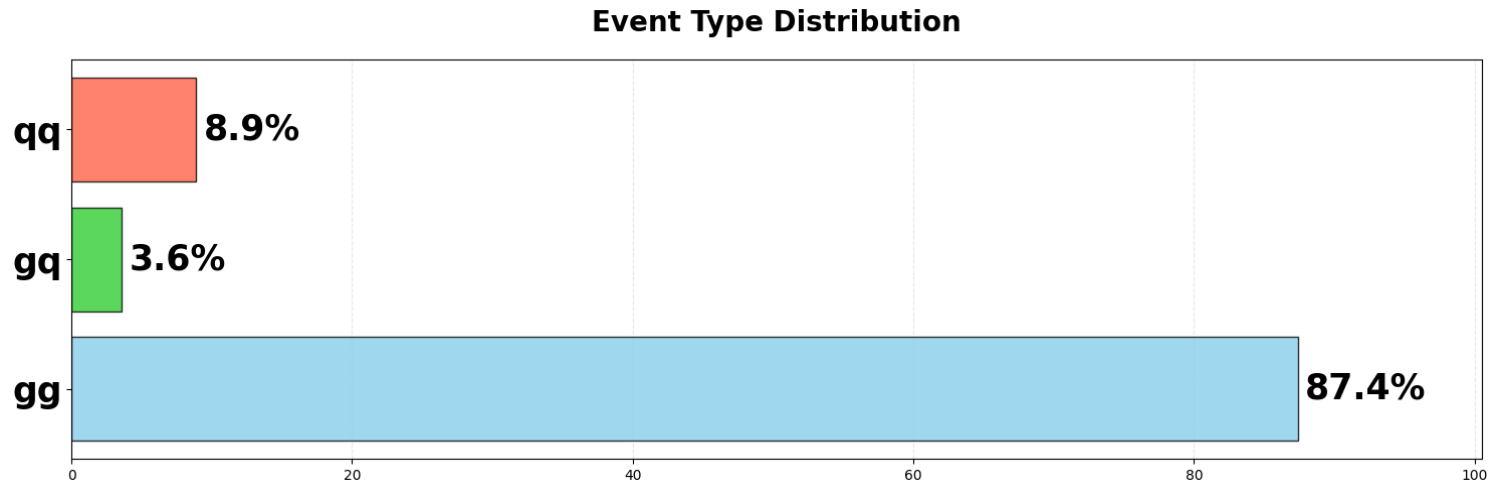
Partially colour singlet → mixed

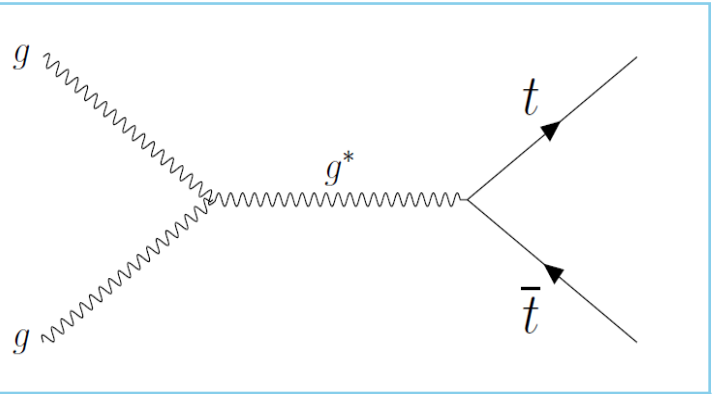
K. Behr- Hamburg International Summer School 2025

# Top quark

## How is top produced?

At LO in QCD:



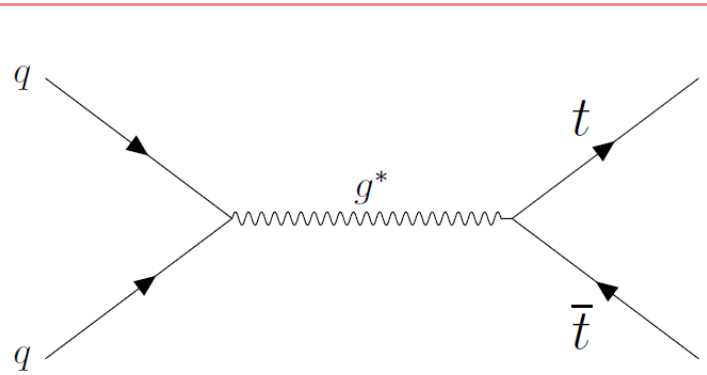




**gg channel**

Most abundant one

10% attractive colour singlet $^1S_0^{[1]}$

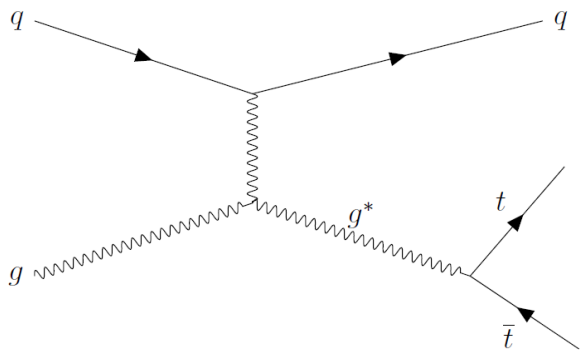
90% repulsive colour octet $^1S_0^{[8]}$




**qq channel**

Final state similar to gg

100% repulsive colour octet $^3S_0^{[8]}$




**gq channel**

Different final state, more jets

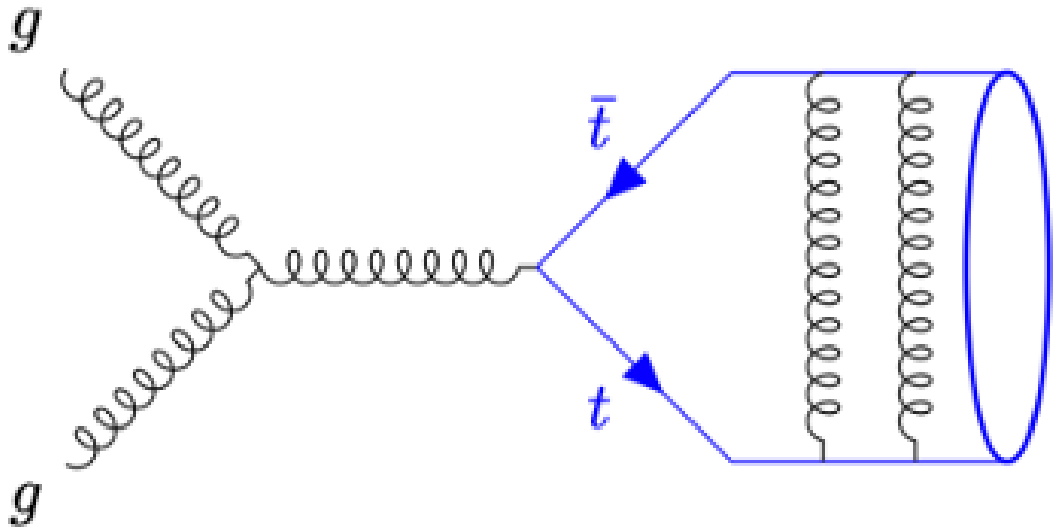
Partially colour singlet → mixed

K. Behr- Hamburg International Summer School 2025

# Top quark

## Quasi-bound state

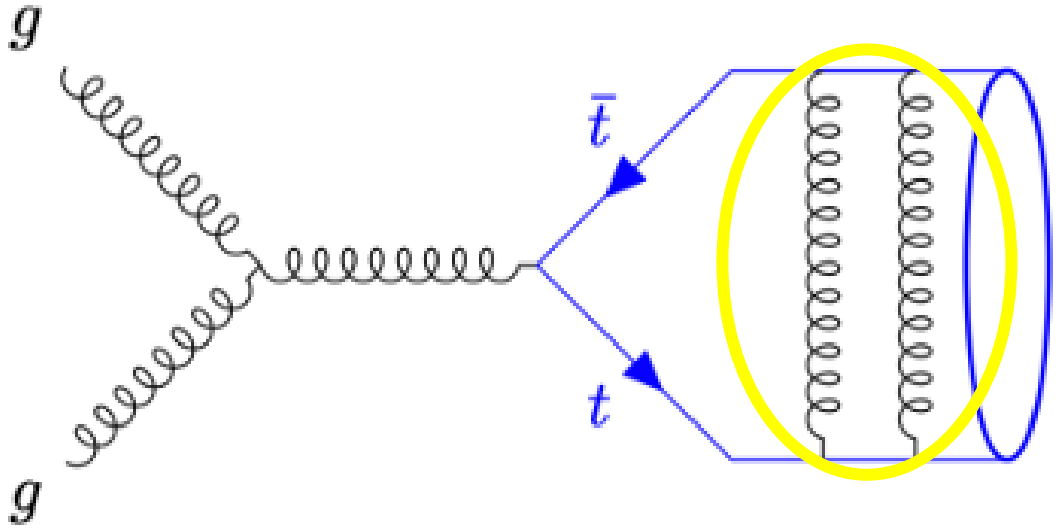
- Multiple gluon exchanges between t and tbar
- Creation of a quasi-bound state, called “toponium”
- Mass just below the production threshold
- Good opportunity to probe early hadron formation



# Top quark

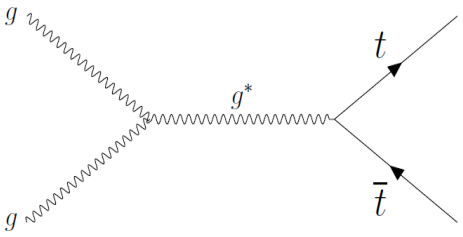
## Quasi-bound state

- Multiple gluon exchanges between t and tbar
- Creation of a quasi-bound state, called “toponium”
- Mass just below the production threshold
- Good opportunity to probe early hadron formation



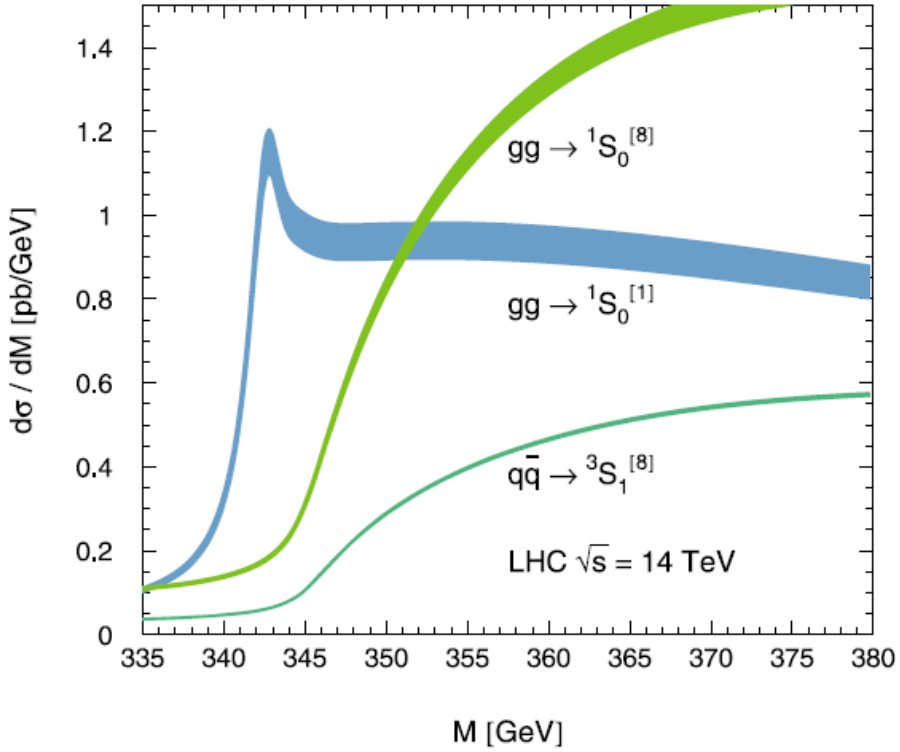
# Top quark

## Attractive colour singlet




- Attractive colour singlet $^1S_0^{[1]}$ is the good candidate for a quasi-bound state in the threshold region
- First predicted in [1987] (before top observation)
- Threshold region: $M = 2 \cdot m_t$

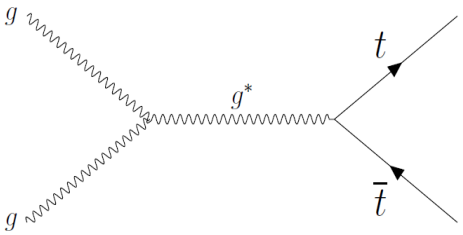
Kiyo et al: EPJC 60(2009) 375-386




Peak smaller than detector resolution, low cross-section: NEVER THOUGHT OF OBSERVING IT AT THE LHC

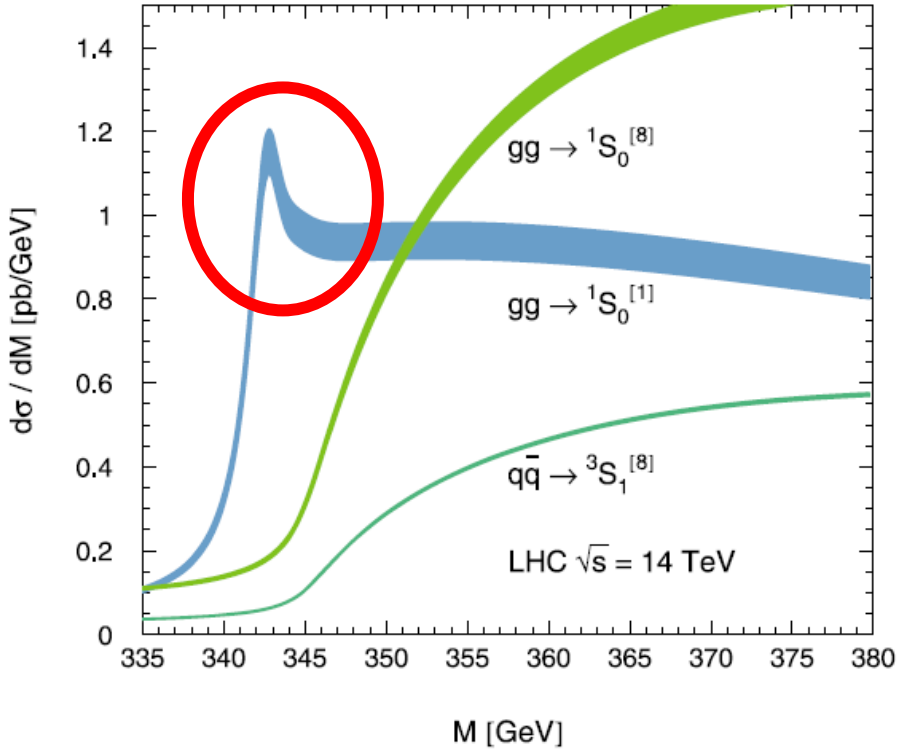
# Top quark

## Attractive colour singlet




- Attractive colour singlet $^1S_0^{[1]}$ is the good candidate for a quasi-bound state in the threshold region
- First predicted in [1987] (before top observation)
- Threshold region: $M = 2 \cdot m_t$

Kiyo et al: EPJC 60(2009) 375-386

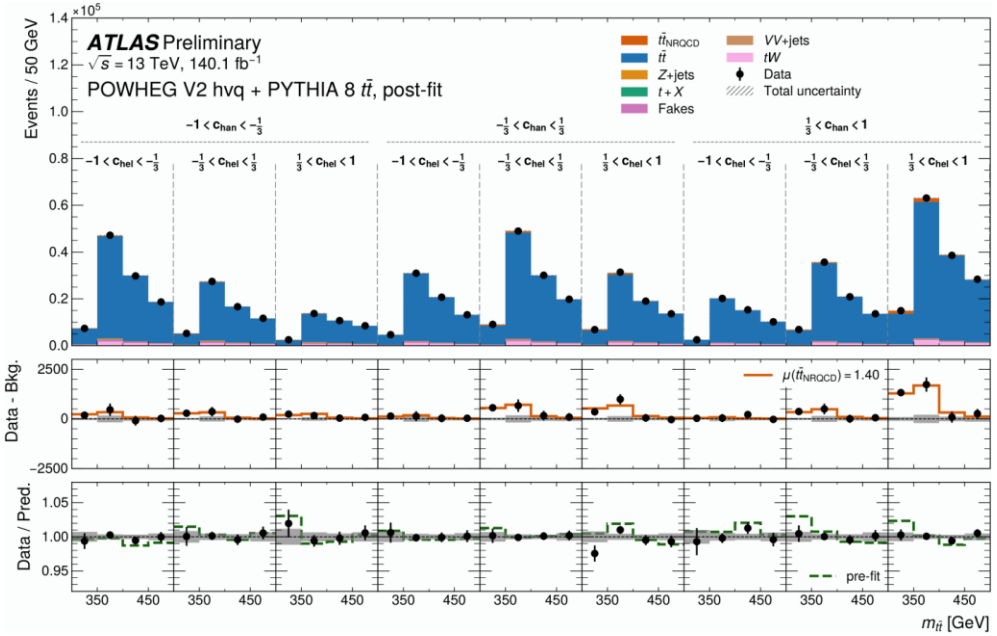



Peak smaller than detector resolution, low cross-section: NEVER THOUGHT OF OBSERVING IT AT THE LHC

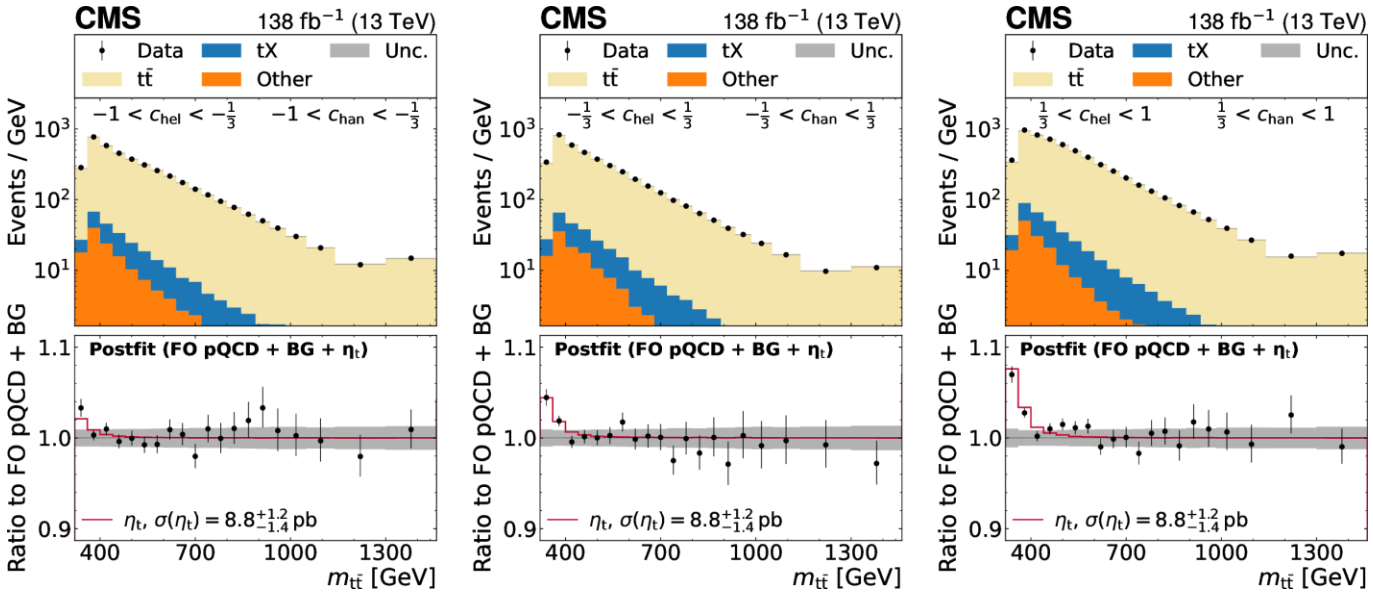
# Top quark

## Quasi-bounded state - observation

But, in 2025, it was observed by ATLAS and CMS collaborations,
where they did not attempt to distinguish between *gg* and background:



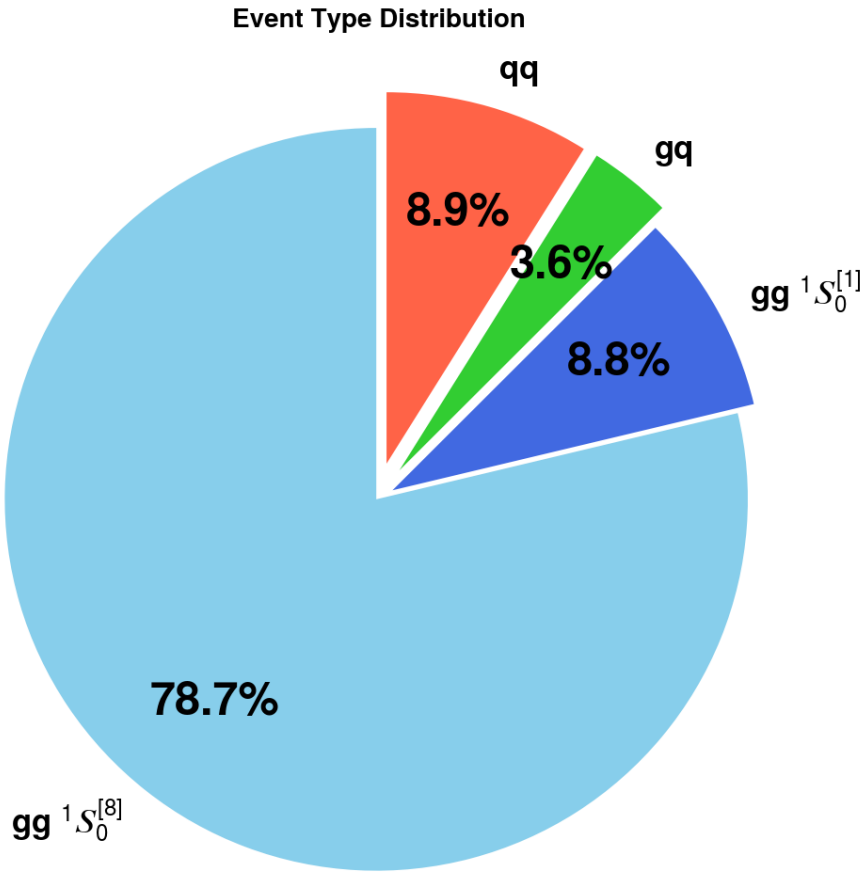

ATLAS-CONF-2025-008




CMS-TOP-24-007

# Goal of the project

## What is our destination?



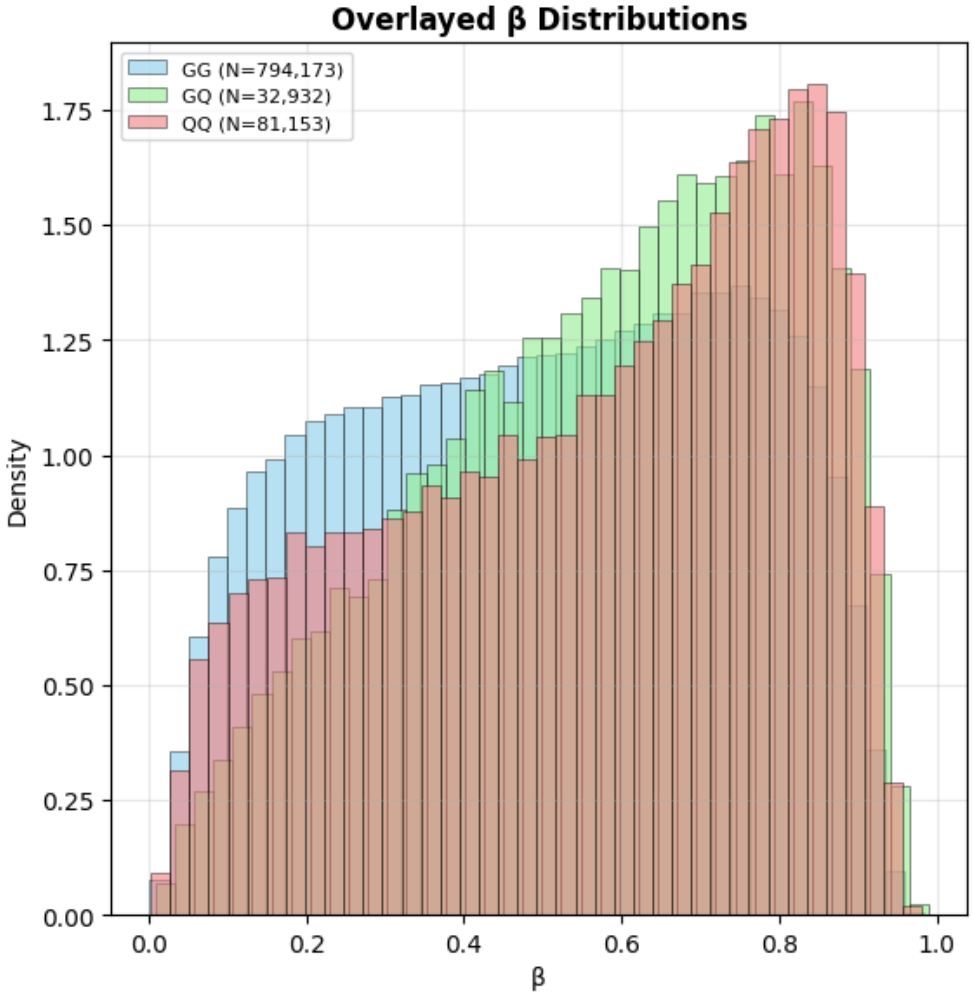

GOAL OF THIS PROJECT:

Classify the different colour states using properties of the production mode, since the signal in *gg* has the same percentage of *background*

# Dataset

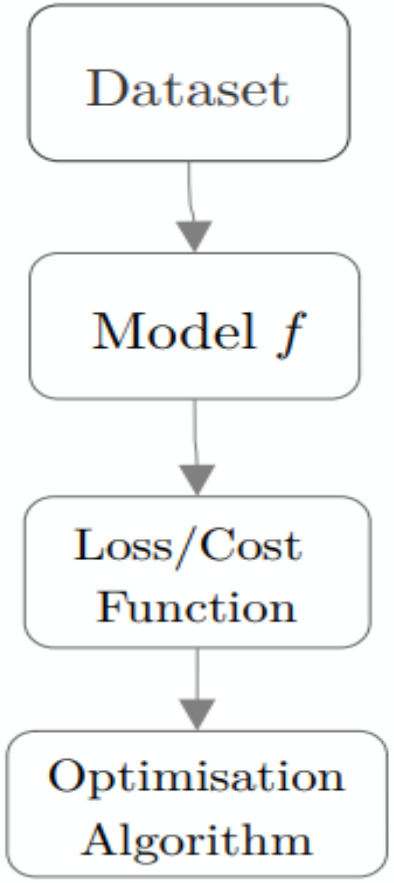
## Building needed bricks

- MC simulated ttbar data
- The production channel of each event is known
- Computing kinetic variables:
  - $p_T$ for $t\bar{t}$
  - $\Delta\eta$ , $\Delta$R for $t\bar{t}$ and $b\bar{b}$
  - $\beta$, $|\beta_z|$
- Also computed angular variables:
  - $\cos\varphi_{ij} = chel$
  - $chan$
  - $\cos\theta^*$
- Good discriminating power



# Deep Neural Network

## DNN in a nutshell



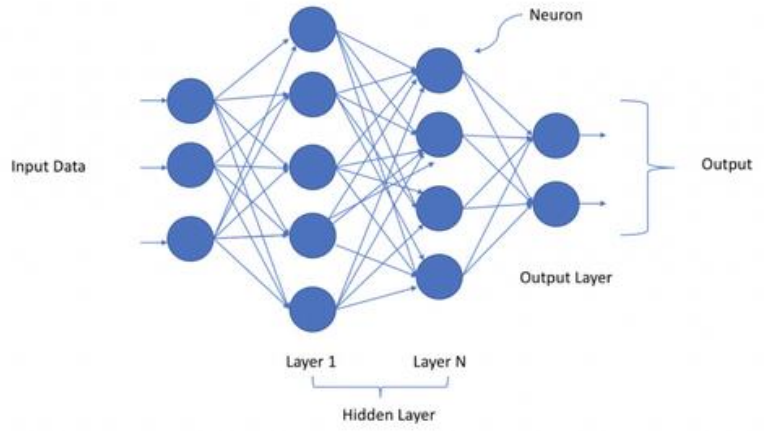

S. Jiggins - DESY Summer Lecture 2025

In this study:

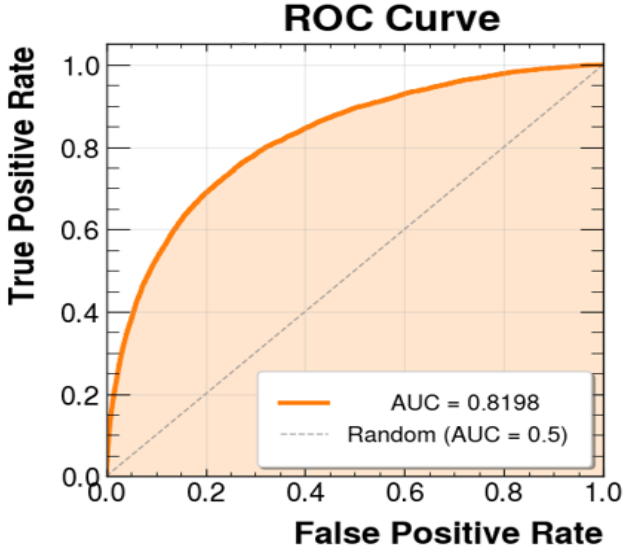
- **LOSS**: Binary and categorical crossentropy
- **METRIC**: weighted accuracy

- **13 FEATURES**: $\beta$, $|\beta_z|$, $m_{t\bar{t}}$, $p_{T\,t\bar{t}}$, *jets_per_event*, $\Delta\eta_{\,b\bar{b}}$, $\Delta R_{\,b\bar{b}}$, $\Delta\eta_{\,t\bar{t}}$, $\Delta R_{\,t\bar{t}}$, $\cos\varphi$, c_han, $\cos\theta^{*}_{had}$, $\cos\theta^{*}_{lep}$

- **DATASET SPLIT**:
(training, validation, testing) = (60:20:20)

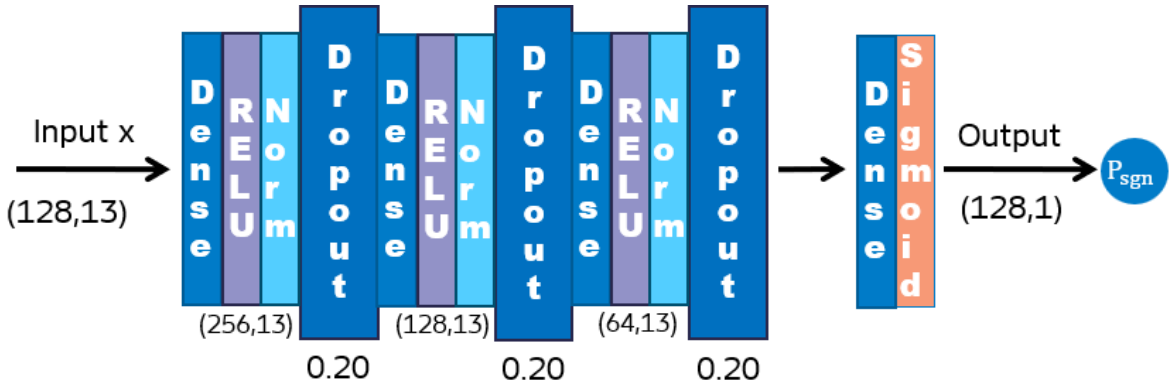



T. Lesort. Continual Learning: Tackling Catastrophic Forgetting in DNNs

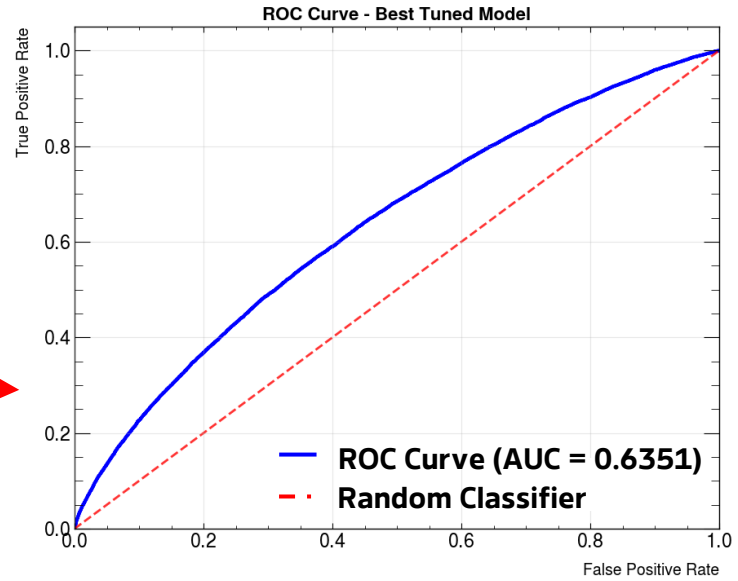


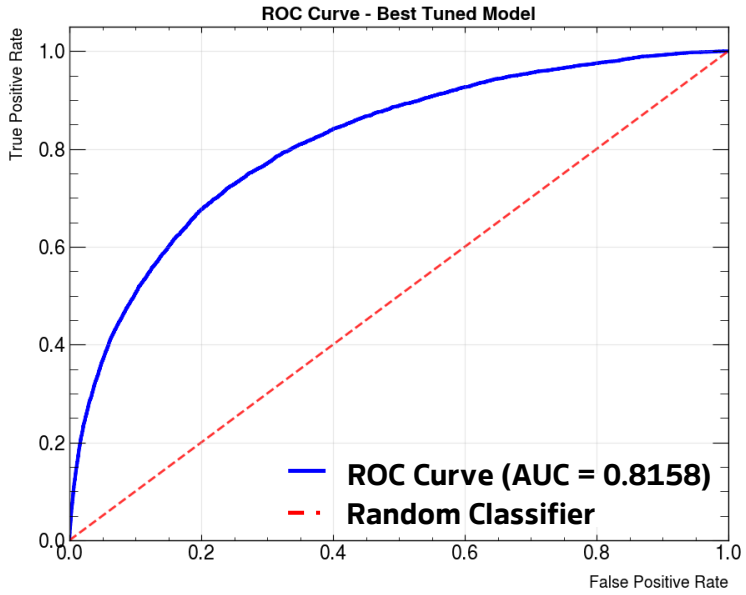
# DNN implementation

## Binary classification – separate analysis



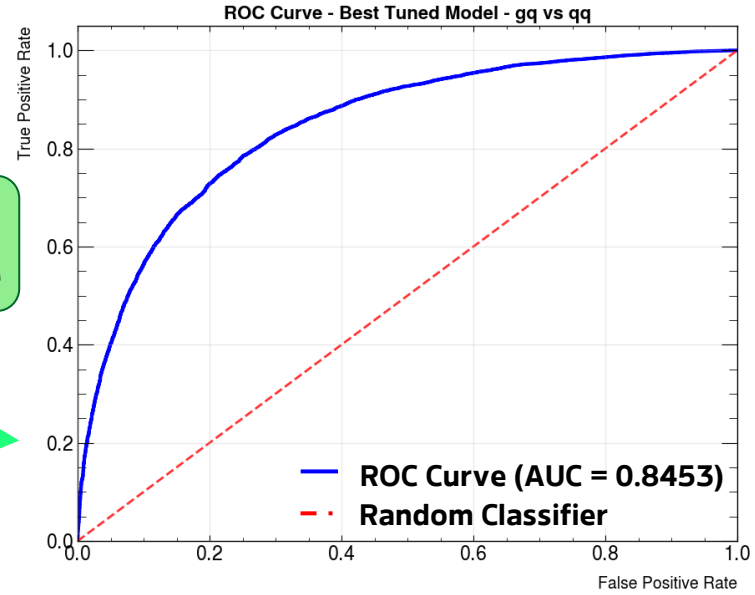

gg vs qq classifier
Not as good as other channels!



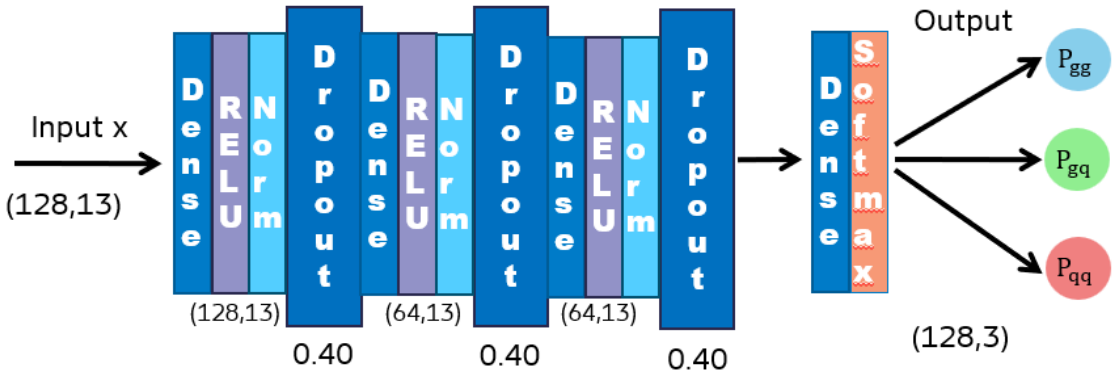

gg vs gq classifier
Good performance

gq vs qq classifier
Good performance



# DNN implementation

## 3 classes classification – [gg,gq,qq]

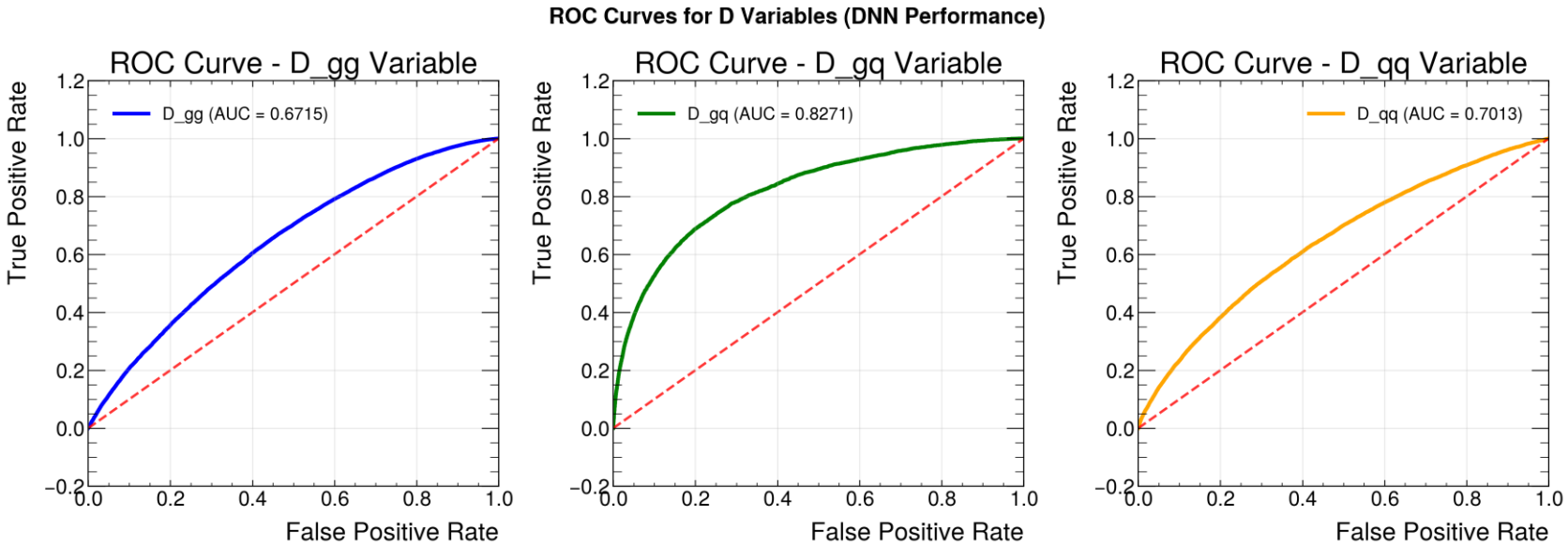



Discriminator is used to evaluate model's performance:

$$D_{gg} = \frac{P_{gg}}{f_{gq}P_{gq} + f_{qq}P_{qq}}$$

$$D_{gq} = \frac{P_{gq}}{f_{gg}P_{gg} + f_{qq}P_{qq}}$$

$$D_{qq} = \frac{P_{qq}}{f_{gg}P_{gg} + f_{gq}P_{gq}}$$



# DNN implementation

## Feature ranking and optimisations

- **FEATURE ANALYSIS:**

Gradient-Based Feature Importance by Tensorflow,
→ ttbar_pT most discriminant variable

- **OPTIMISATIONS**:

Testing different hyperparameters and choosing the configuration with the maximum AUC

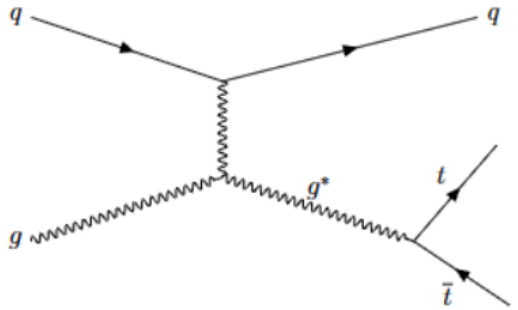
ROC Curve computed over single variables distributions:

DNN got higher AUC in all cases

| Variable | gg vs gq | gg vs qq | gq vs qq |
|---|---|---|---|
| Beta | 59.7 % | 57.4 % | 58.1 % |
| Beta_z | 51.3 % | 57.9 % | 56.1 % |
| ttbar pT | 77.0 % | 56.5 % | 53.2 % |
| DNN 1 node | 81.6 % | 63.5 % | 84.5 % |

# Conclusions

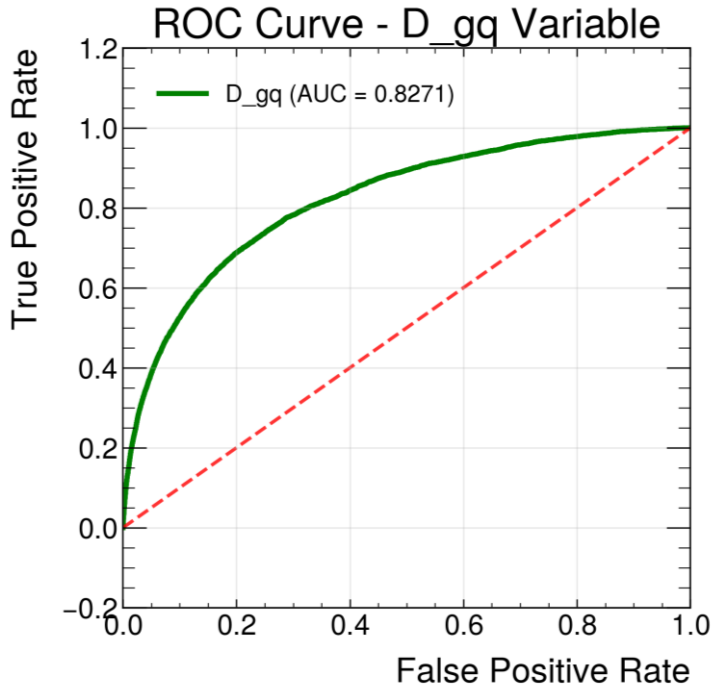
## Could this help with toponium?




- Distinguishing between *gg* and other channels is crucial for top measurements (e.g.: *toponium, spin correlation* etc) → it provides a cleaner data sample

- gq channel is the one we can identify better, in all cases
- DNN is more performant than classical cut-based analysis

- *Future outlook:*
  - *Increasing statistics*: from 900K to 10M events
  - Implementing *more complex architectures* (e.g. Transformers), running on low-level variables
  - Prepare the field to use Run3 data



# Thank you

Thanks to:
*K. Behr, E. Jones, M. Gonzalez , F. Jolly* for the guidance
*ATLAS group* for the hospitality
*DESY for the opportunity*

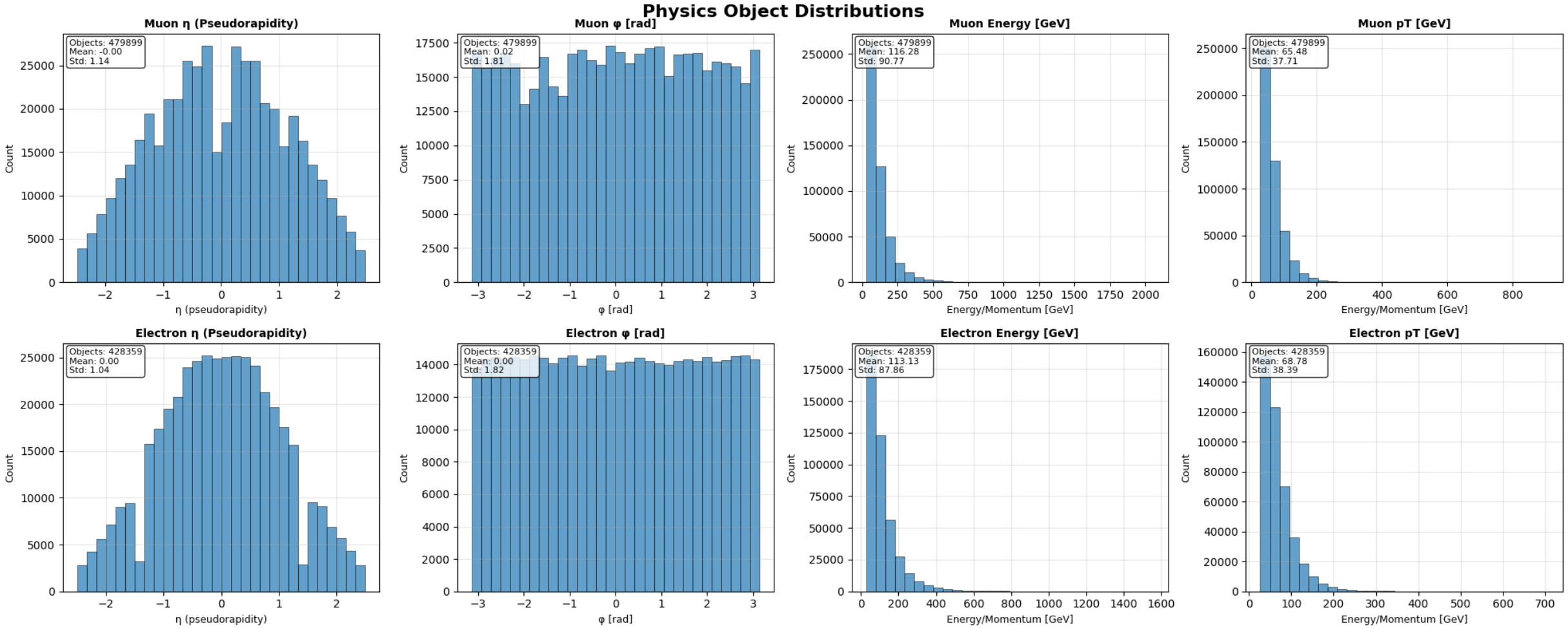
**Contacts**
Valerio Tinari

E-Mail:
valerio.tinari@desy.de
tinari.1998628@studenti.uniroma1.it

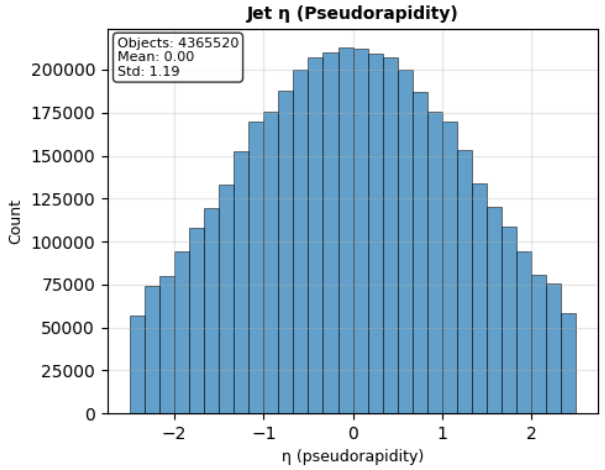
**Deutsches Elektronen-Synchrotron DESY**
ATLAS Group
Bldg. 01c | O1.301
Notkestraße 85
22607 Hamburg
www.desy.de

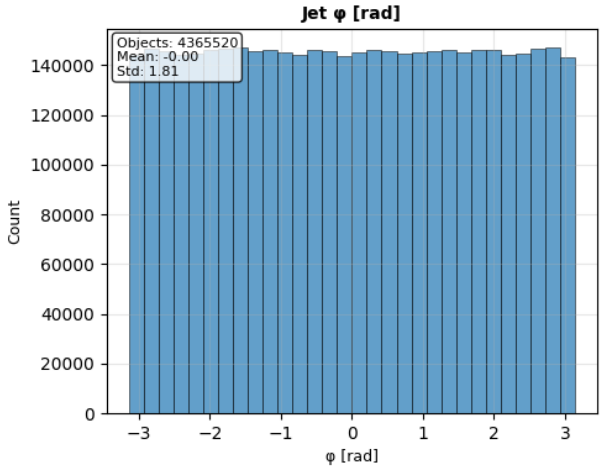
# BACKUP

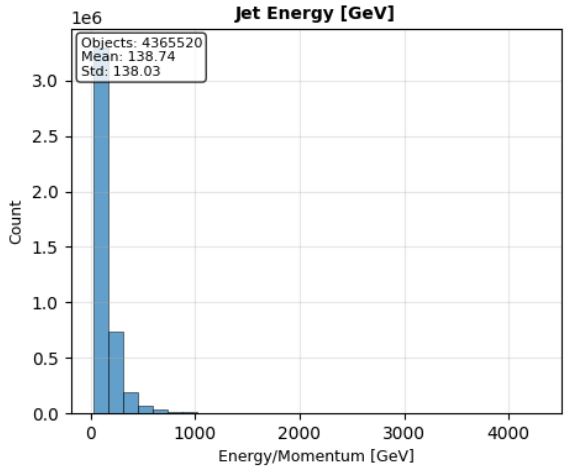
# Dataset

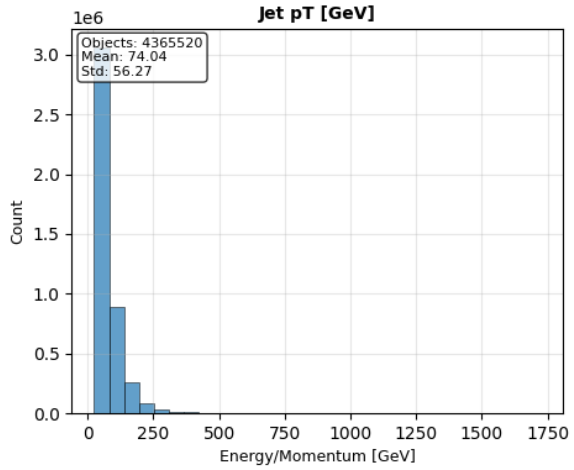

Physics Object Distributions
Muon η (Pseudorapidity)
Objects: 479899
Mean: -0.00
Std: 1.14
Count
η (pseudorapidity)
Muon φ [rad]
Objects: 479899
Mean: 0.02
Std: 1.81
Count
φ [rad]
Muon Energy [GeV]
Objects: 479899
Mean: 116.28
Std: 90.77
Count
Energy/Momentum [GeV]
Muon pT [GeV]
Objects: 479899
Mean: 65.48
Std: 37.71
Count
Energy/Momentum [GeV]
Electron η (Pseudorapidity)
Objects: 428359
Mean: 0.00
Std: 1.04
Count
η (pseudorapidity)
Electron φ [rad]
Objects: 428359
Mean: 0.00
Std: 1.82
Count
φ [rad]
Electron Energy [GeV]
Objects: 428359
Mean: 113.13
Std: 87.86
Count
Energy/Momentum [GeV]
Electron pT [GeV]
Objects: 428359
Mean: 68.78
Std: 38.39
Count
Energy/Momentum [GeV]

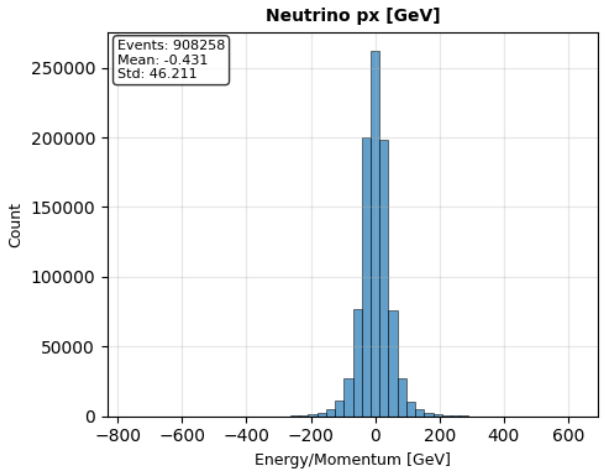
# Dataset

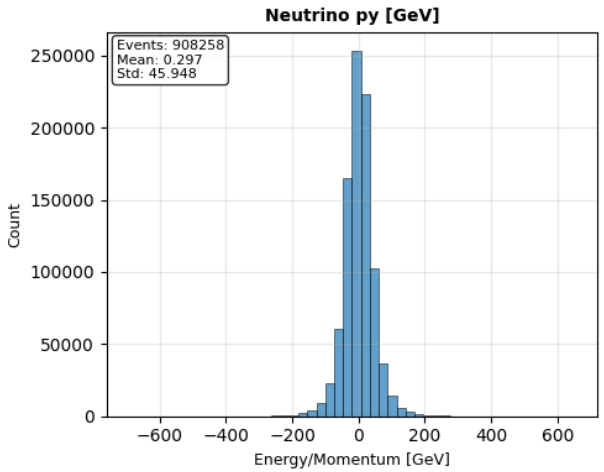

Jet η (Pseudorapidity)
Objects: 4365520
Mean: 0.00
Std: 1.19
Count
η (pseudorapidity)

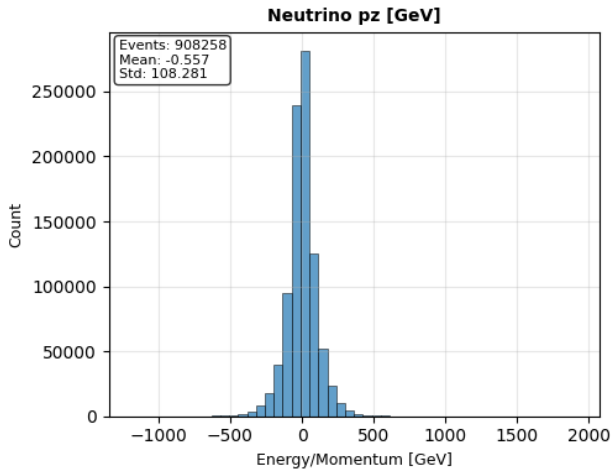

Jet φ [rad]
Objects: 4365520
Mean: -0.00
Std: 1.81
Count
φ [rad]

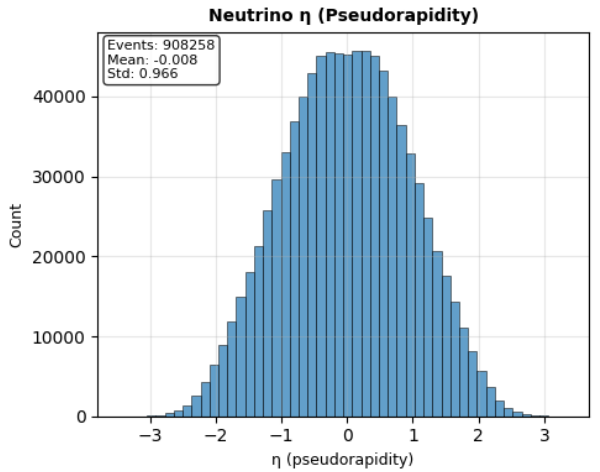

1e6
Jet Energy [GeV]
Objects: 4365520
Mean: 138.74
Std: 138.03
Count
Energy/Momentum [GeV]


1e6
Jet pT [GeV]
Objects: 4365520
Mean: 74.04
Std: 56.27
Count
Energy/Momentum [GeV]


Neutrino px [GeV]
Events: 908258
Mean: -0.431
Std: 46.211
Count
Energy/Momentum [GeV]


Neutrino py [GeV]
Events: 908258
Mean: 0.297
Std: 45.948
Count
Energy/Momentum [GeV]


Neutrino pz [GeV]
Events: 908258
Mean: -0.557
Std: 108.281
Count
Energy/Momentum [GeV]


Neutrino η (Pseudorapidity)
Events: 908258
Mean: -0.008
Std: 0.966
Count
η (pseudorapidity)

# Dataset

## Kinematic variables computation

- Beta for ttbar system: $\beta_{t\bar{t}} = \frac{p_{t\bar{t}}}{E}$ → *boost of the system*
- Beta_z for ttbar system: $|\beta_{z\,t\bar{t}}| = \frac{p_{z\,t\bar{t}}}{E}$ → *boost of the system along z-axis*
- *Delta eta*: $\Delta\eta = \eta_t - \eta_{\bar{t}}$
- *Delta R*: $\Delta R = R_t - R_{\bar{t}}$

# Dataset

## Angular variables computation

- $\cos\varphi_{ij}$ = chel : angle between lepton and down-type quark , after boosting them to top parent ref frame

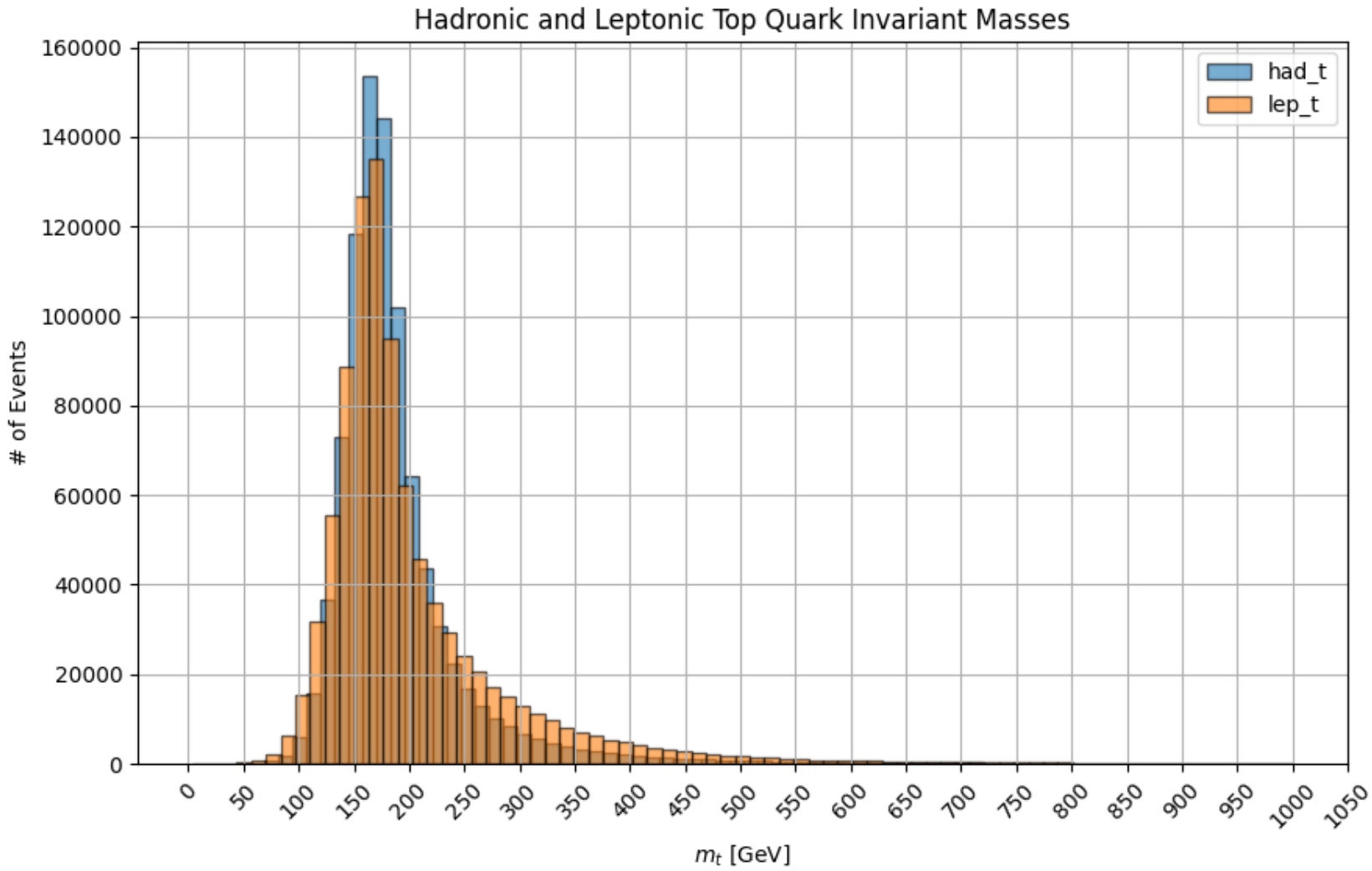
$$< \cos\varphi > = < \hat{\ell} \cdot \hat{d} >$$

- Chan: the same of $\cos\varphi$, but with flip of the z-direction of lepton before the product
- $\cos\theta^*$ : angle between one top and flight direction of ttbar

# Dataset

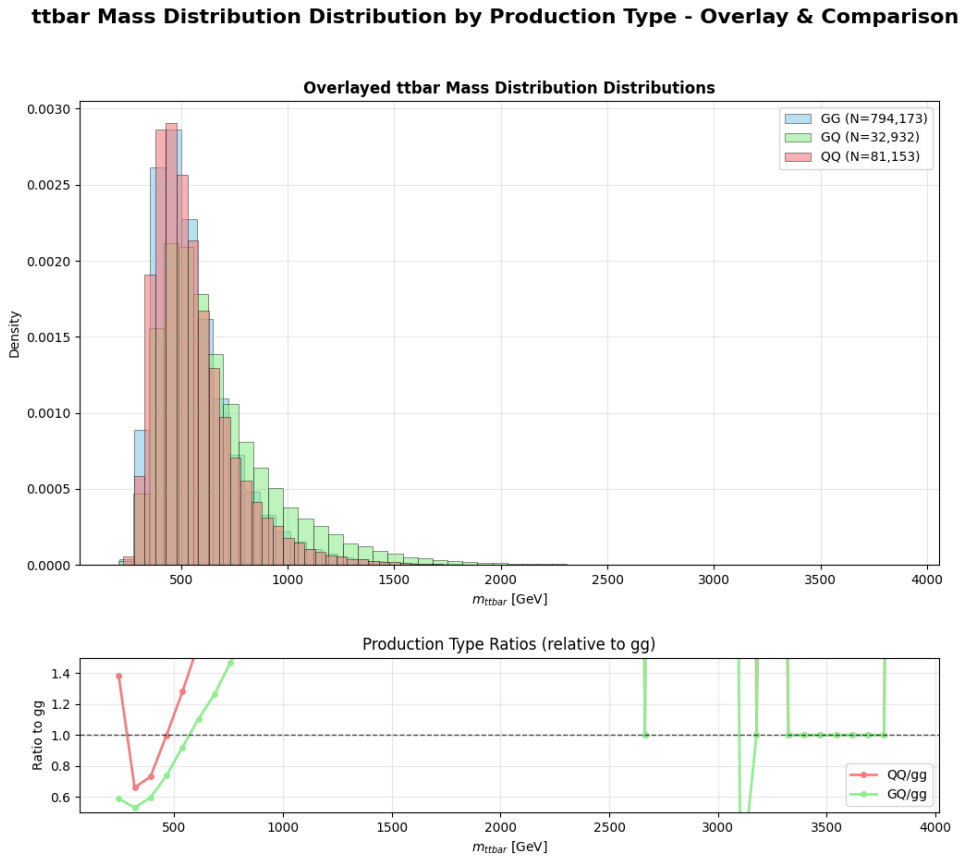
## Invariant mass of both hadronic and leptonic top

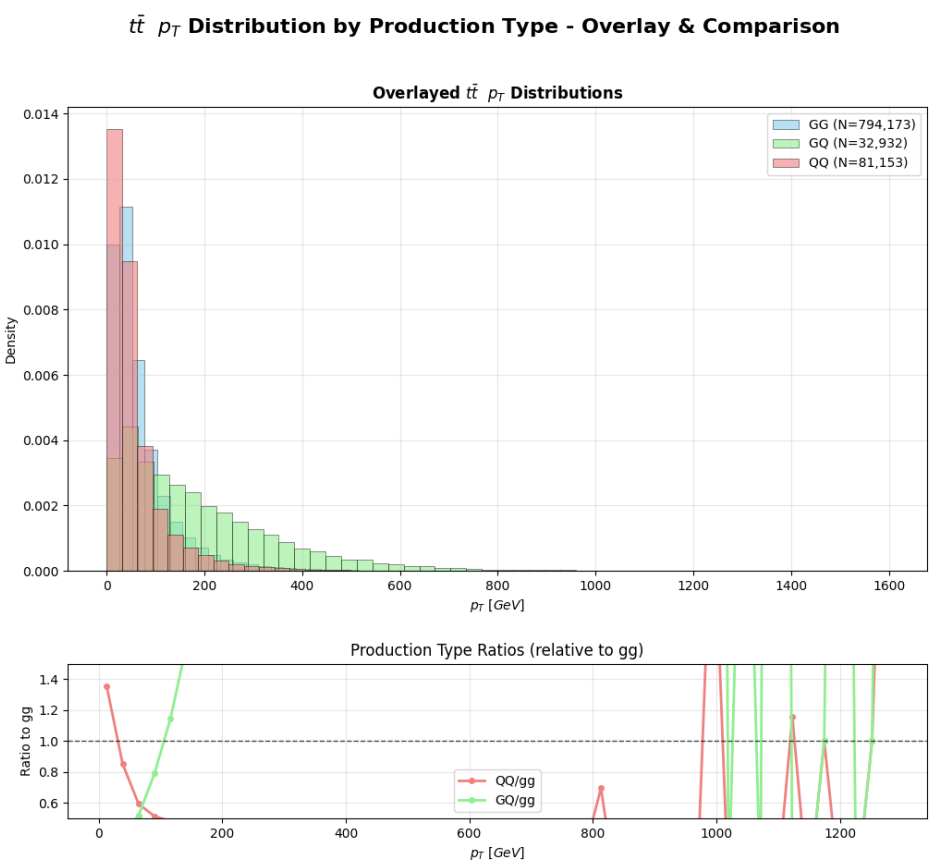
*Excess in the leptonic top due to the presence of a neutrino to reconstruct*



# Dataset

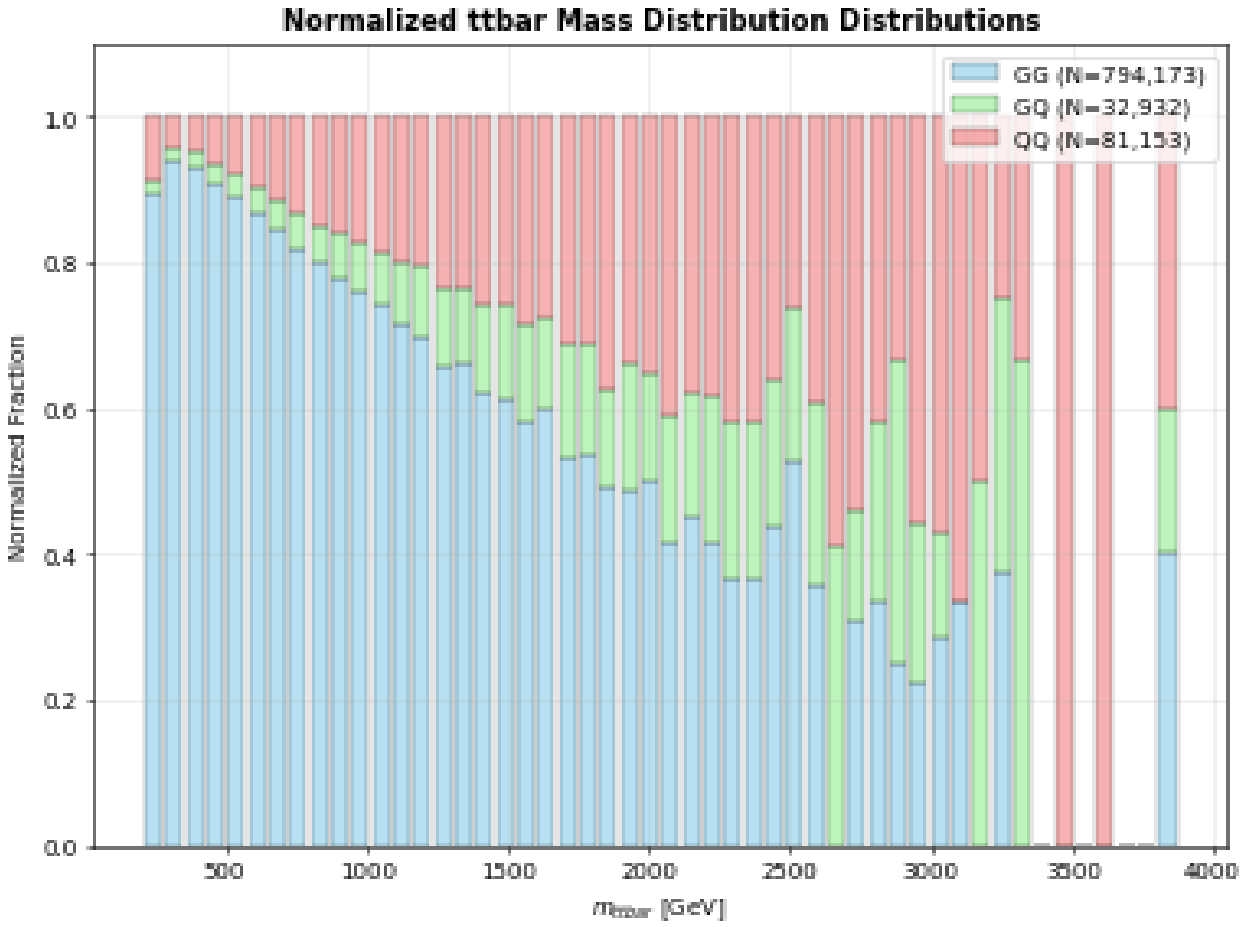
## ttbar invariant mass and pT of ttbar



# Dataset

## Normalized ttbar invariant mass per production mode

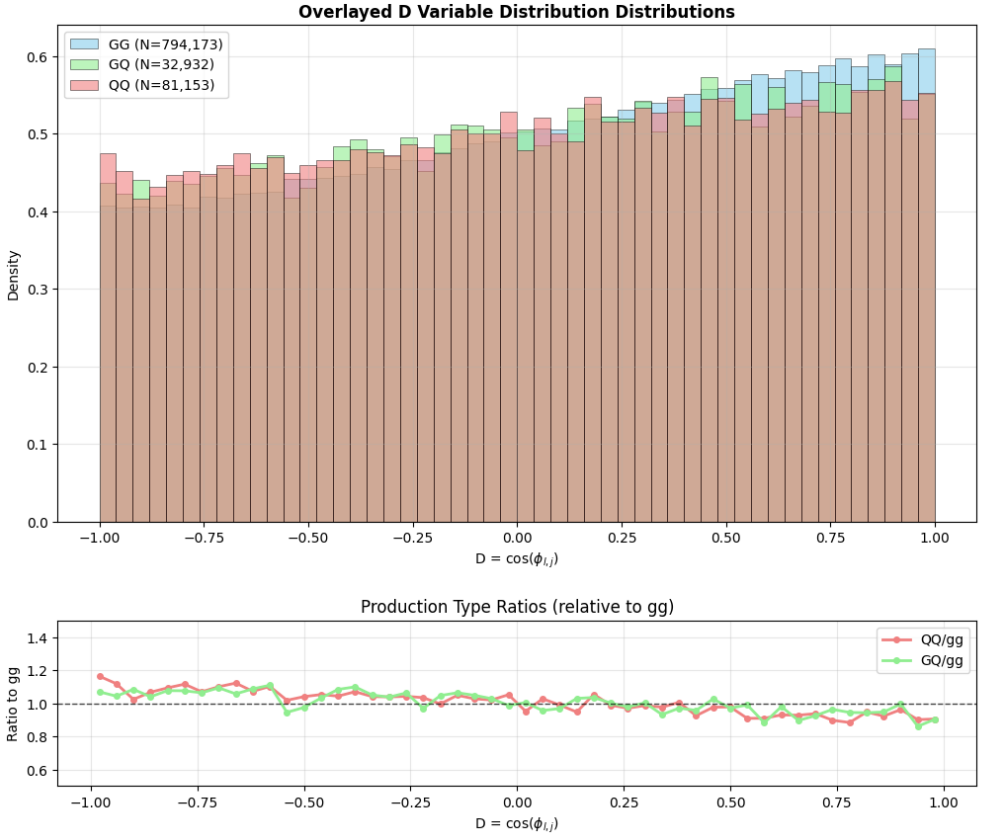



*gg* is dominating in the threshold region;
*qq* is dominating in the high energy region

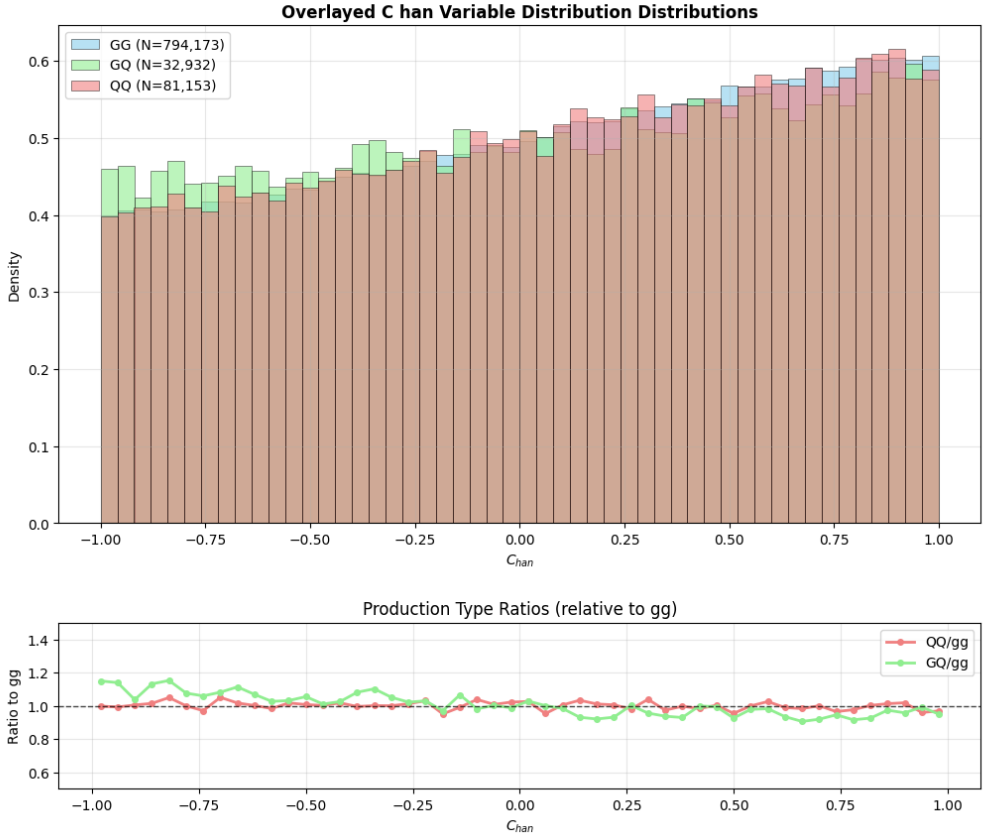
# Dataset

## Angular variables per production mode

**D Variable Distribution Distribution by Production Type - Overlay & Comparison**

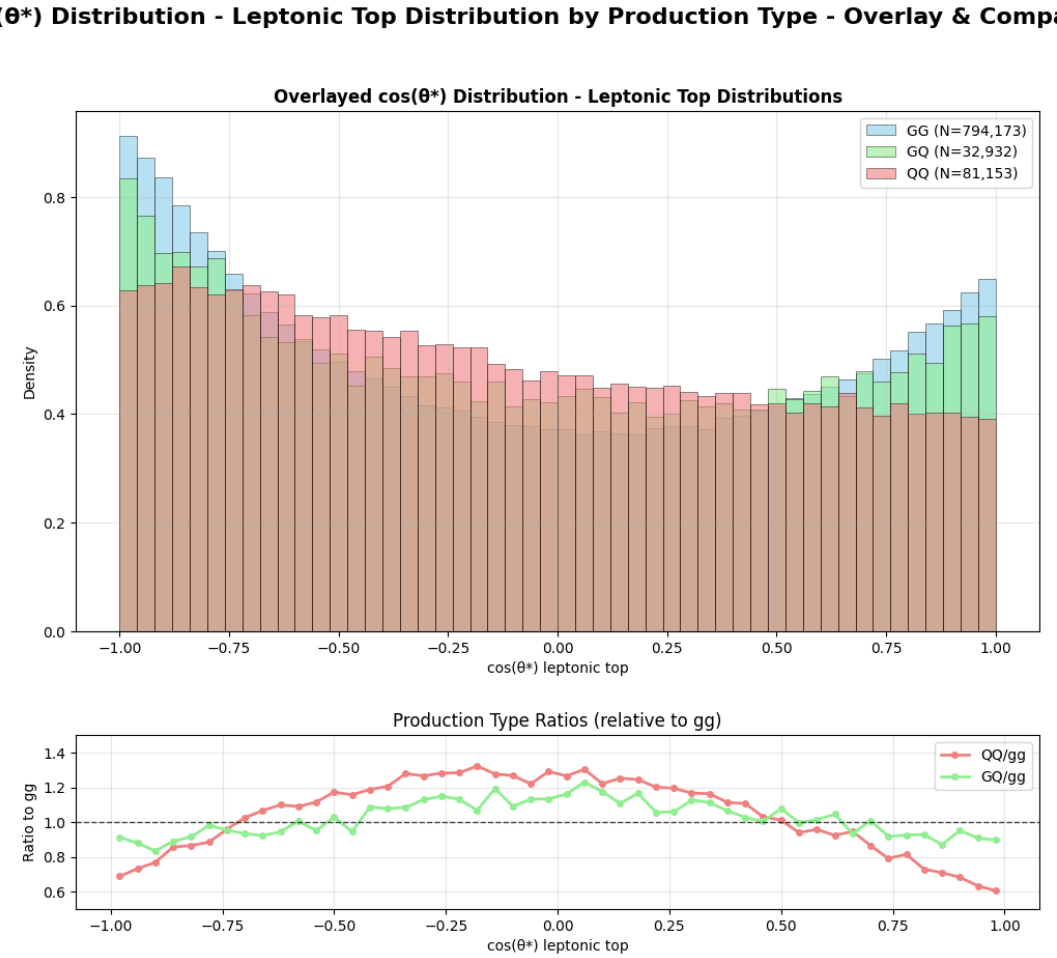



**C han Variable Distribution Distribution by Production Type - Overlay & Comparison**



# Dataset

## Angular variables per production mode



# SPANet architecture

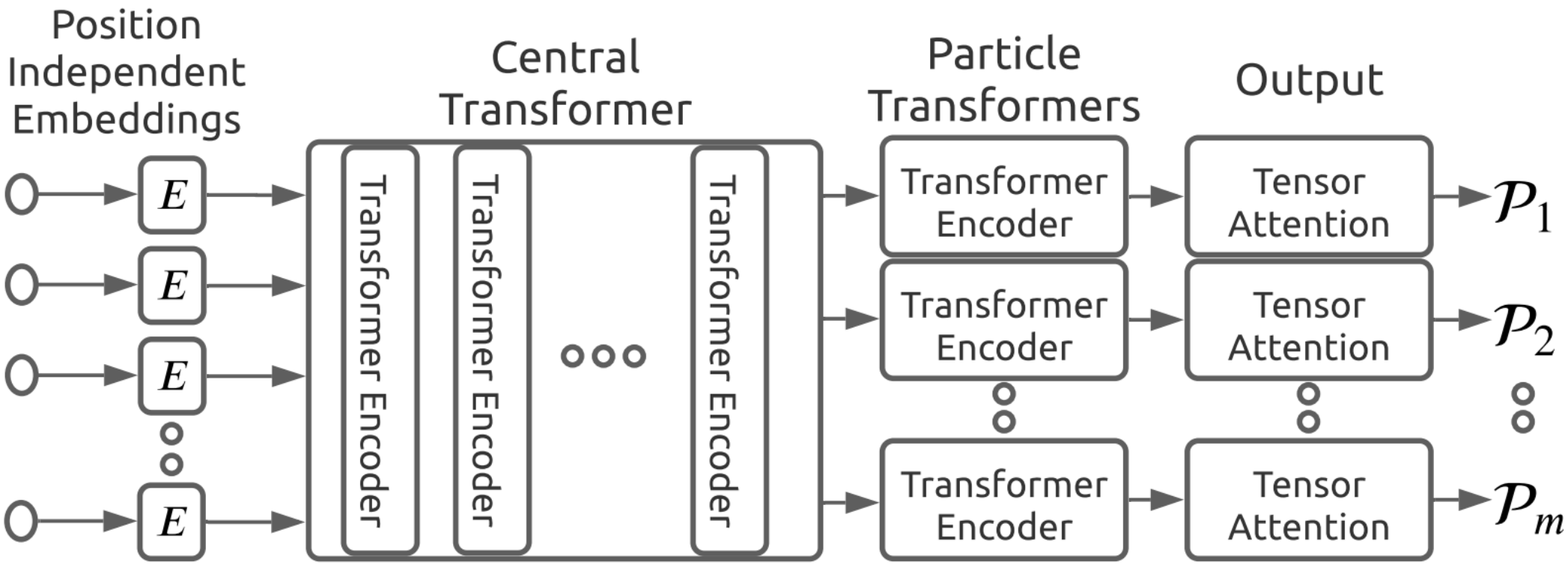
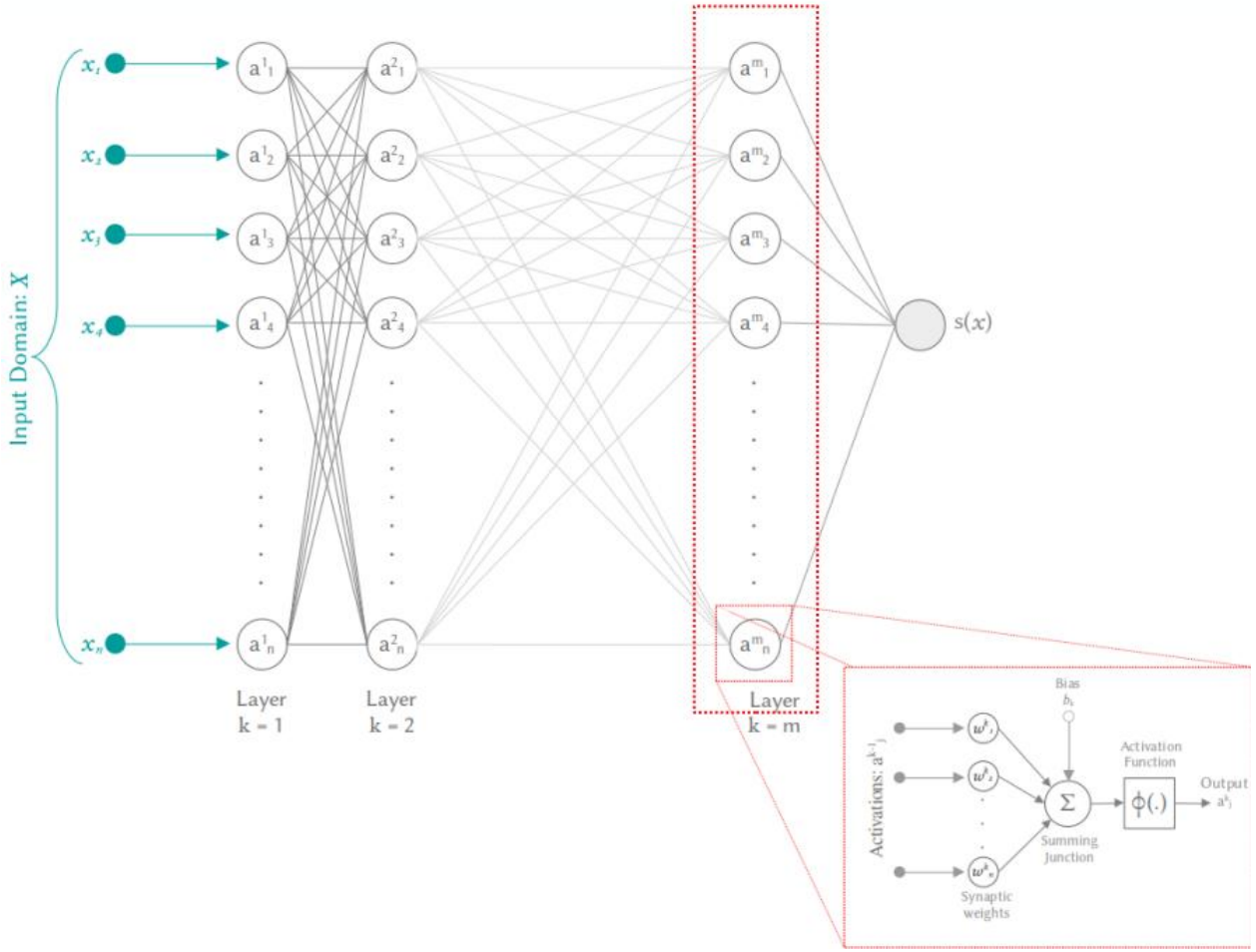



Figure 2: A visualization of the high level structure of SPA-NET.

SciPost Phys. 12, 178 (2022)

# Deep Neural Network

## What is it?




S. Jiggins - DESY Summer Lecture 2025

- Neuron structure:

$$a_i^{(k+1)} = \phi\left(\sum_j^n w_{i,j}^{(k)} a_j^{(k)} + b^{(k)}\right)$$

- Neural Network:

$$\begin{bmatrix} a_0^{(k+1)} \\ \dots \\ a_n^{(k+1)} \end{bmatrix} = \phi\left(\begin{bmatrix} w_0^{(0)} & \dots & w_n^{(0)} \\ \dots & \dots & \dots \\ w_0^{(k)} & & w_n^{(k)} \end{bmatrix} \begin{bmatrix} a_0^{(k)} \\ \dots \\ a_n^{(k)} \end{bmatrix} + \begin{bmatrix} b_0^{(k)} \\ \dots \\ b_n^{(k)} \end{bmatrix}\right)$$

# Binary and categorical crossentropy loss

- Binary crossentropy (BCE):

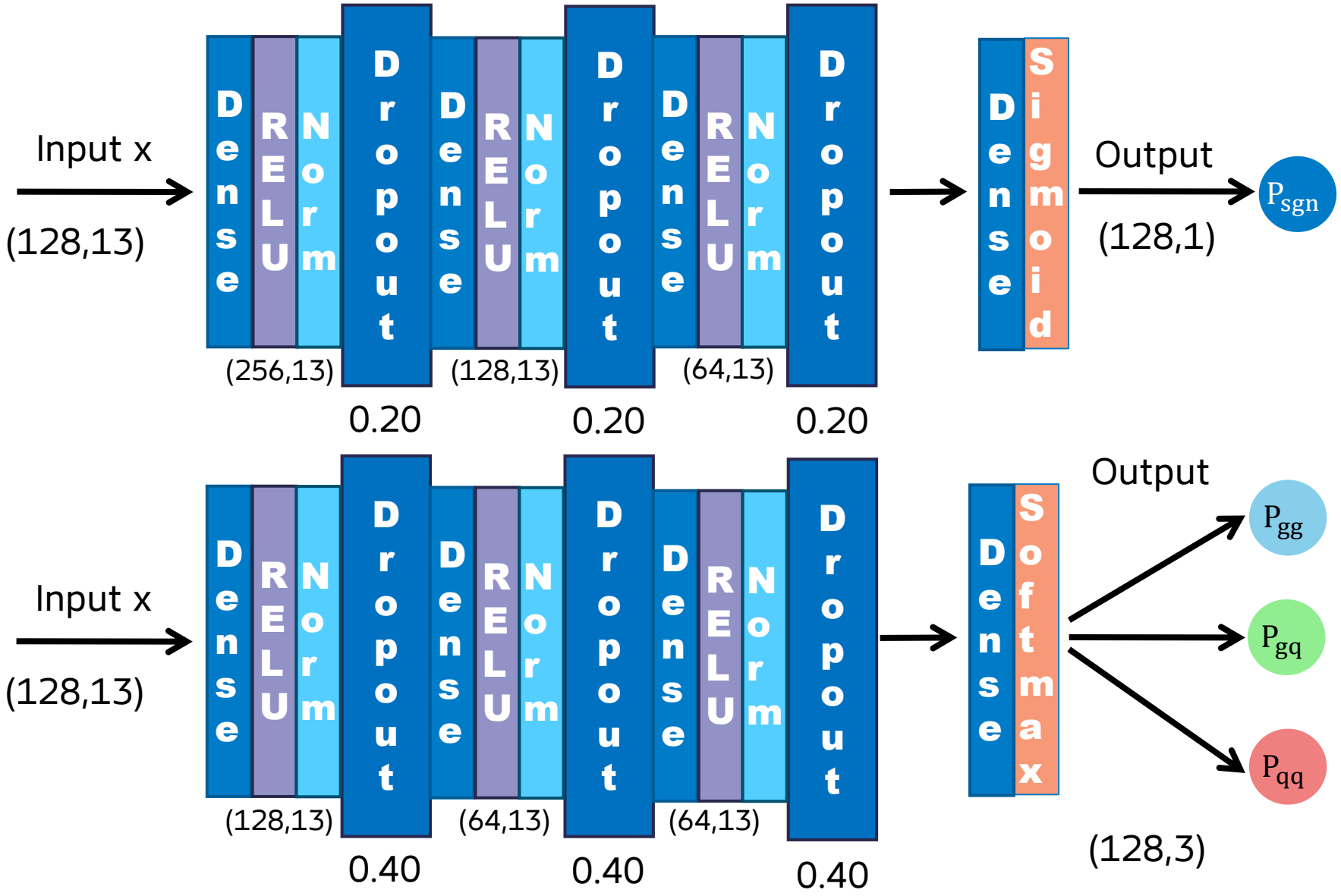
$$\mathrm{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$$

Binary Cross-Entropy measures the distance between the true labels and the predicted probabilities. When the predicted probability pipi is close to the actual label yiyi, the BCE value is low, indicating a good prediction.

- Categorical crossentropy

$$L(y,\hat{y}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

# Model architecture



# DNN implementation

## Consistency checks

BDT confirms results from DNN

- BDT training to cross-check results
- ROC curve and AUC score over most important variables' distribution (Beta, Beta_z, ttbar pT)
- Run without angular variables → no major changes

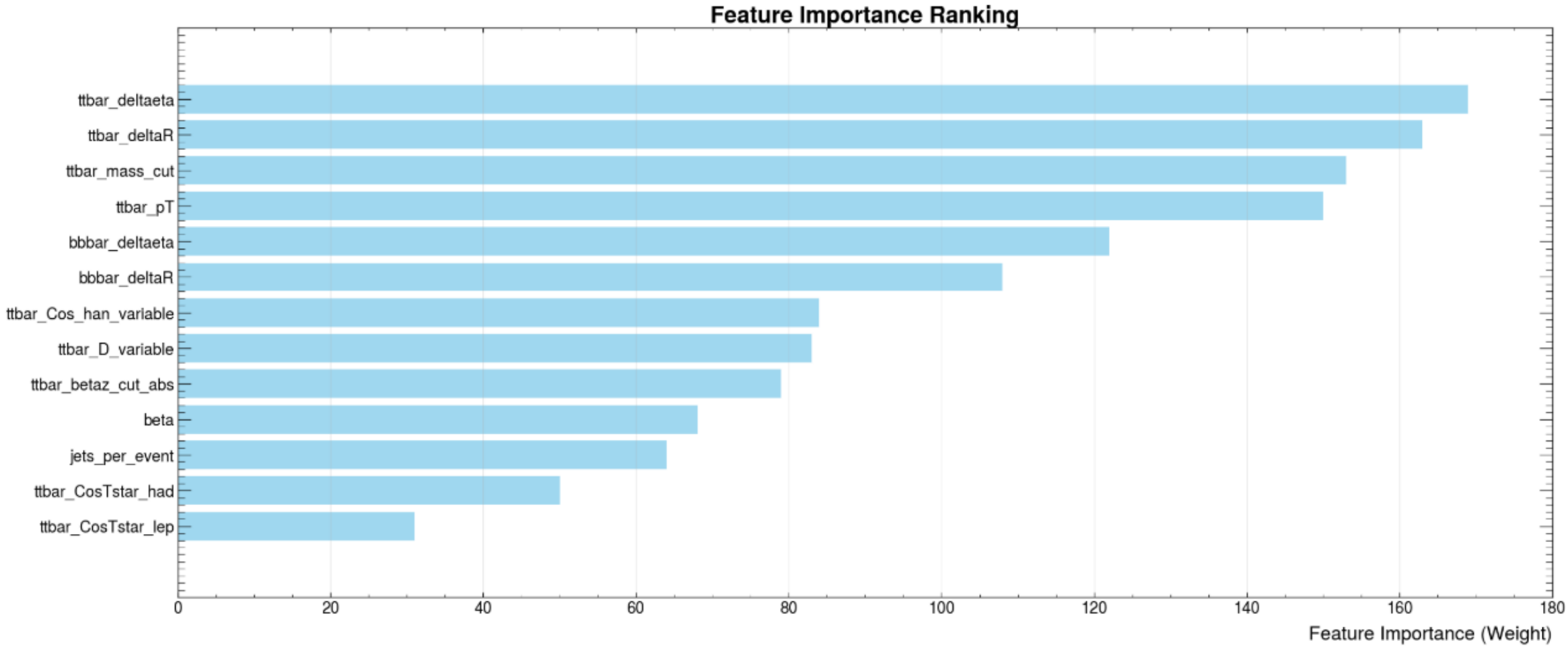
| Classical AUC | gg vs gq | gg vs qq | gq vs qq |
|---|---|---|---|
| BDT | 80.9 % | 83.8 % | 63.7 % |
| Beta | 59.7 % | 57.4 % | 58.1 % |
| Beta_z | 51.3 % | 57.9 % | 56.1 % |
| ttbar pT | 77.0 % | 56.5 % | 53.2 % |

DNN is performing better

| Discriminator AUC | gg signal | gq signal | qq signal |
|---|---|---|---|
| BDT 3 classes | 64.3 % | 81.9 % | 64.5 % |

BDT documentation

# Features ranking – BDT (gg vs qq channel)



# ATLAS detector