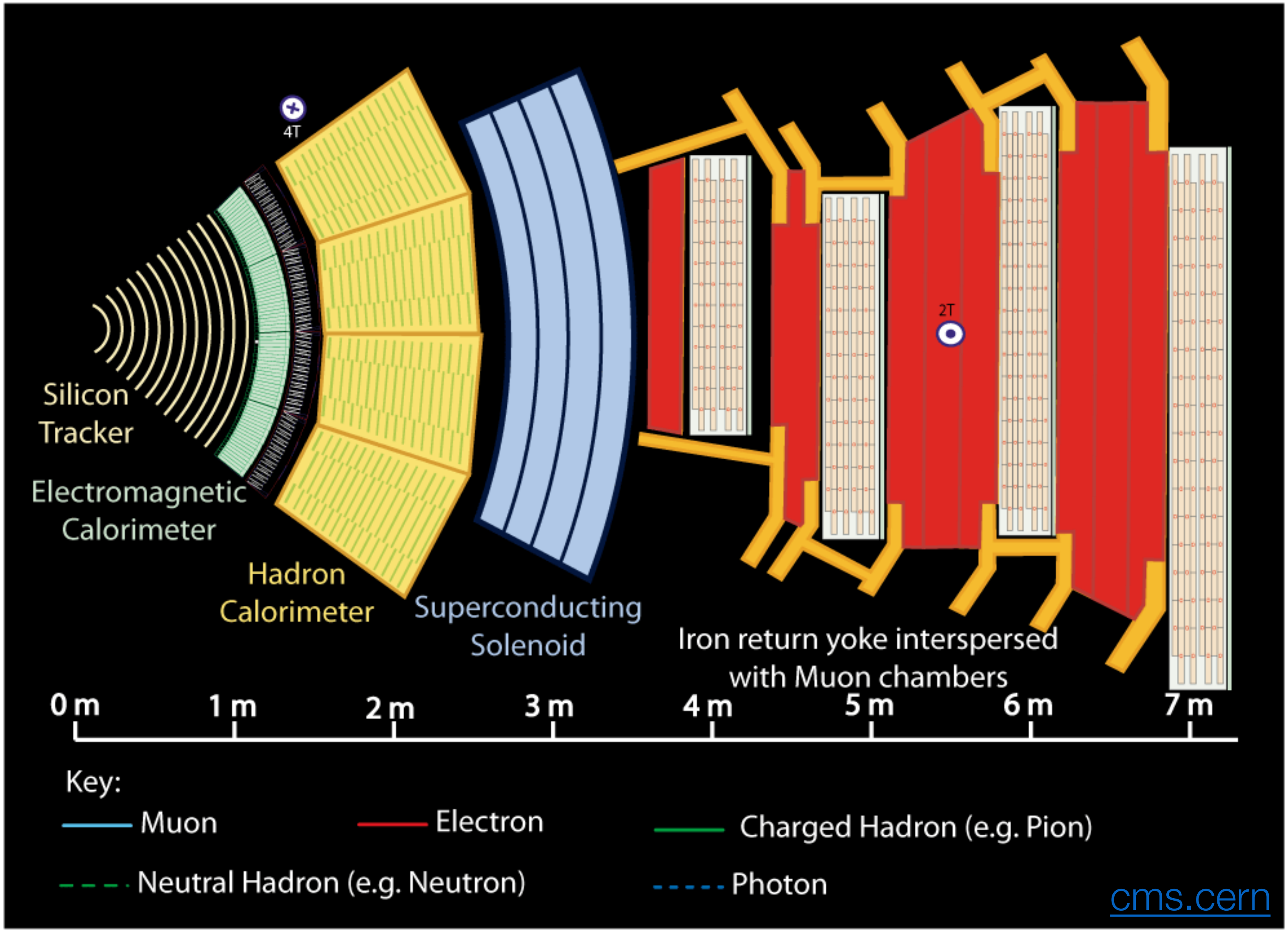
# Jet Calibration and Search for Vector-Like Quark decaying into top+Higgs in hadronic final state using Run 2 CMS data with Neural Network

Oct 25, 2022

Jieun Choi

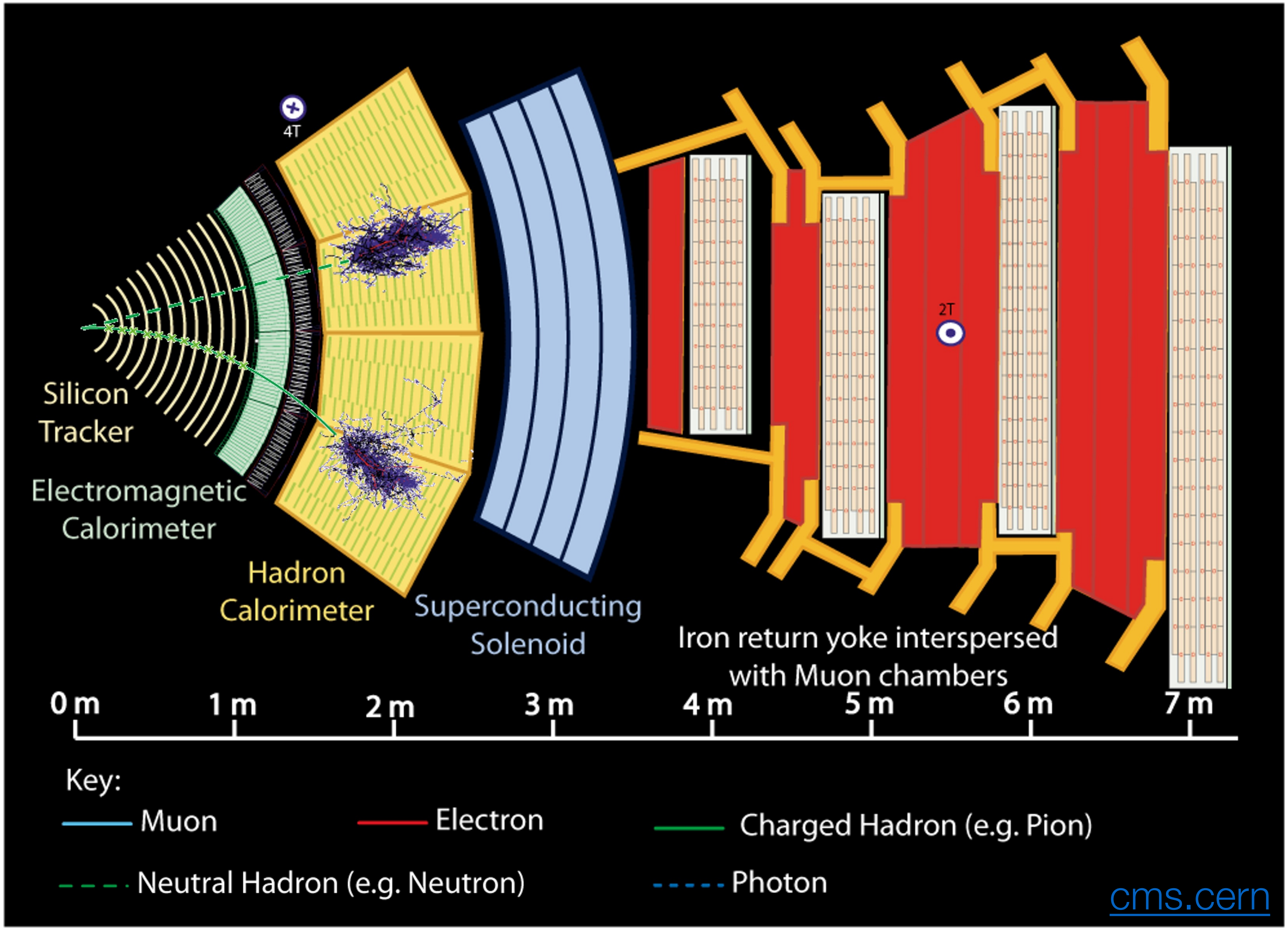
IP2I Lyon

# Introduction to CMS detector




CMS experiment has many layers to detect different kinds of objects

# Introduction to CMS detector

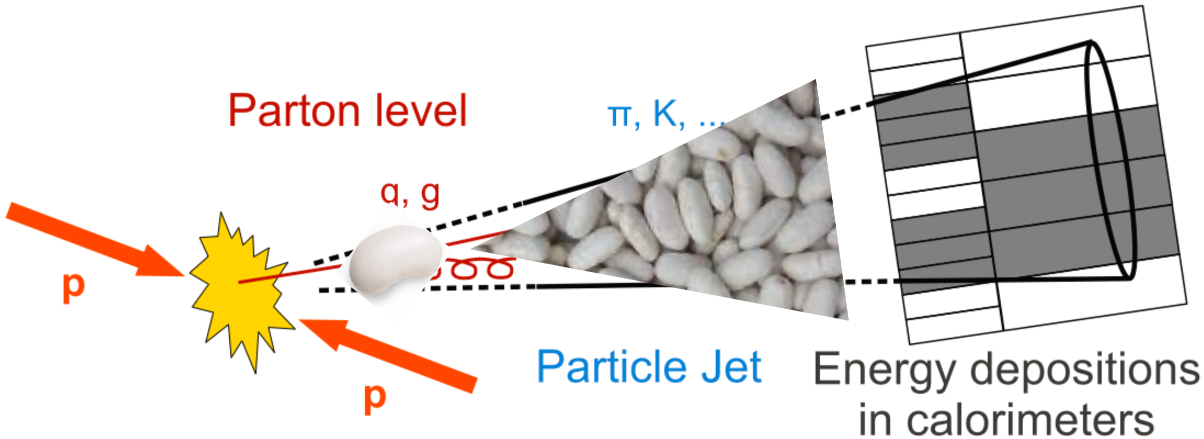



CMS experiment has many layers to detect different kinds of objects

# Jet in CMS experiment

## What is a Jet ?

- Particle Object from Hadronization




- Quarks and Gluons cannot exist freely due to color confinement
- Form color-neutral hadrons → a *shower* of hadrons

- In theoretical calculation or Monte-Carlo simulation:
  - The final state stable particles
- In the real experiment:
  - Energy block having finite position / energy resolution

https://cms.cern/news/jets-cms-and-determination-their-energy-scale

# Jet in CMS experiment

## How we build up a Jet in CMS

**Detector Clustering**

- Grouping energy deposit in Calorimeter

**Particle Flow**

- Building particles by linking all sub-detectors
- Subtract additional energy from low energy proton-proton collisions

**Jets / MET Clustering**

- Grouping PF particles with given cone size

**Calibration & Tagging**

- *Determine Jet Energy Scale and Resolution*
- Flavor + Heavy object tagging algorithm

**Analysis**

- Comparison of data and simulation

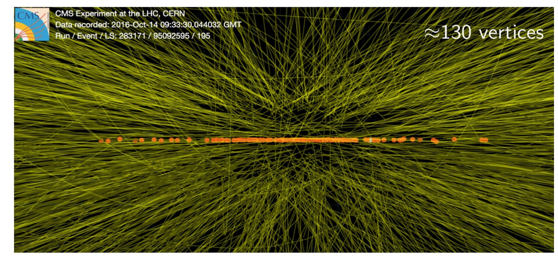
# Jet in CMS experiment

## Jet Energy Correction in CMS

Applied to data →

DP-2021/033


Applied to simulation →



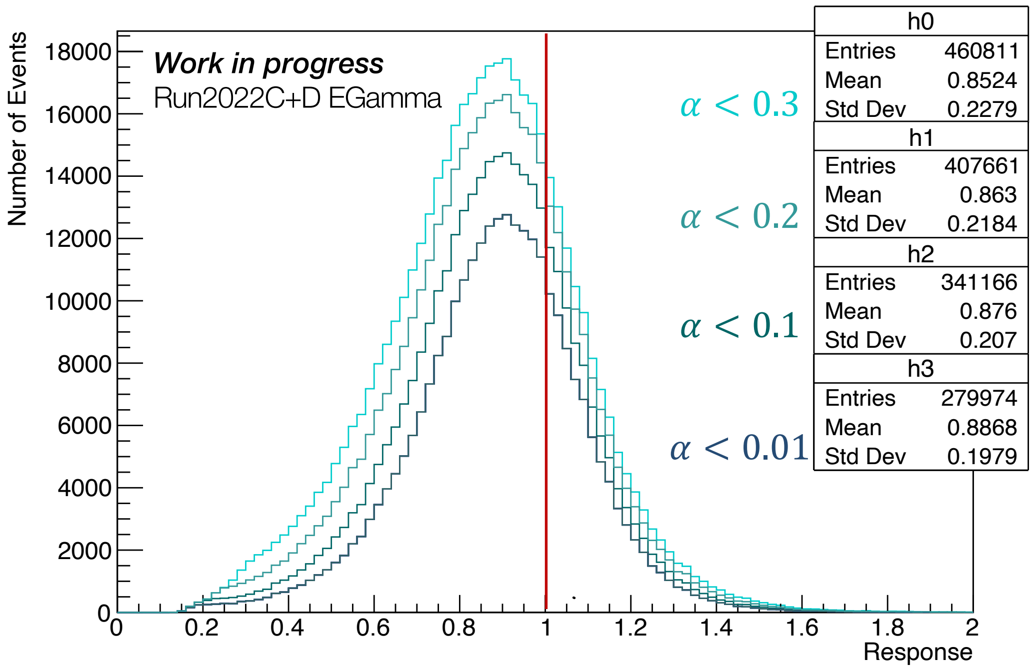

- Factorized approach
  - Subtract additional energy from collisions happening simultaneously (Pileup ↑)
  - Compensate non-linear response of calorimeter (+ Angular differences)
  - Residual corrections from Data and MC Comparison

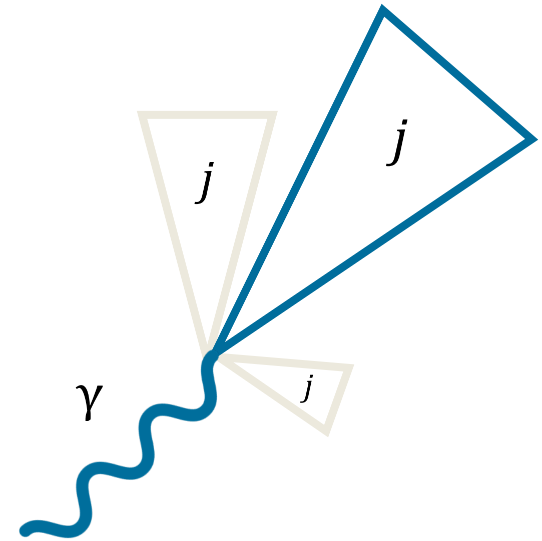
A better understanding of the scale uncertainties
→ more precise measurements possible!

# Jet in CMS experiment

## Response with Run 3 data

- Response distribution with respect to alpha variation
  - Response: Leading jet pT / photon pT
  - Alpha: sub-leading jet pT / photon pT



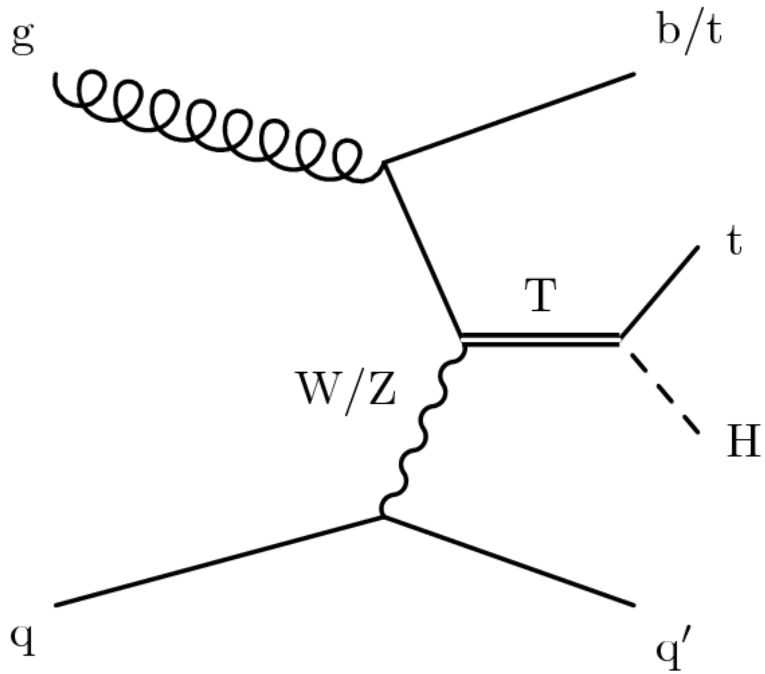




- In Ideal case (MC): 1 photon – 1 jet back-to-back → Response ~ 1, Alpha ~ 0
- HCAL issue observed: miscalibration arise lower energy reconstruction ~ 70 %
- By doing such study, we can validate the data in early stage

We need to understand “Jet”!

# Motivation

## Search for Vector Like Quark in hadronic final states




### What is a Vector-Like Quark?

- Vector-like: Spin = 1
- Evaluate many underlying models:
  - Stabilize the Higgs boson mass
  - Offers a potential solution to the hierarchy problem

    …

### T’ decay in full hadronic final state

- T’ decaying into top and Higgs
  - t → Wb → qqb
  - H → bb
- Main background:
  - ttbar in hadronic decay (tt → bbqqqq)
  - multi-jet event (QCD)

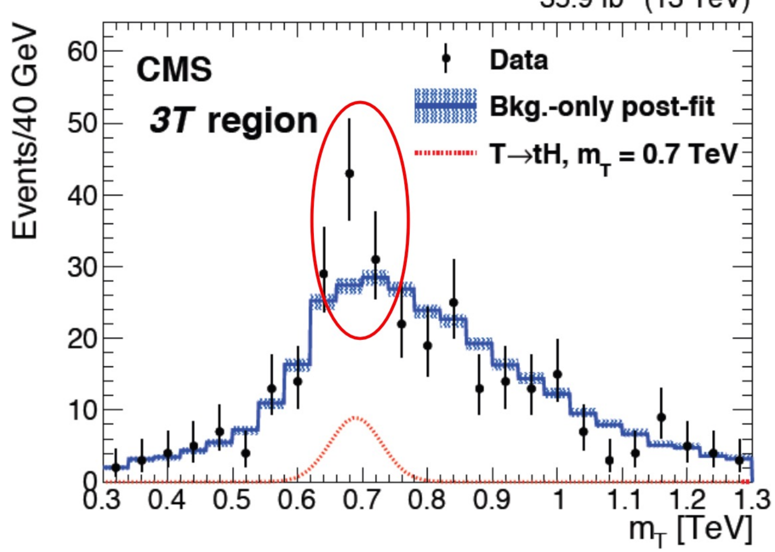
# Strategy

## Analysis using 2016 data in CMS

- Excess in T' mass @ 700 GeV is observed!

→ Might be able to improve significance with NN using Run2 data (even Run 3) 🤔

[JHEP 01 (2020) 036]



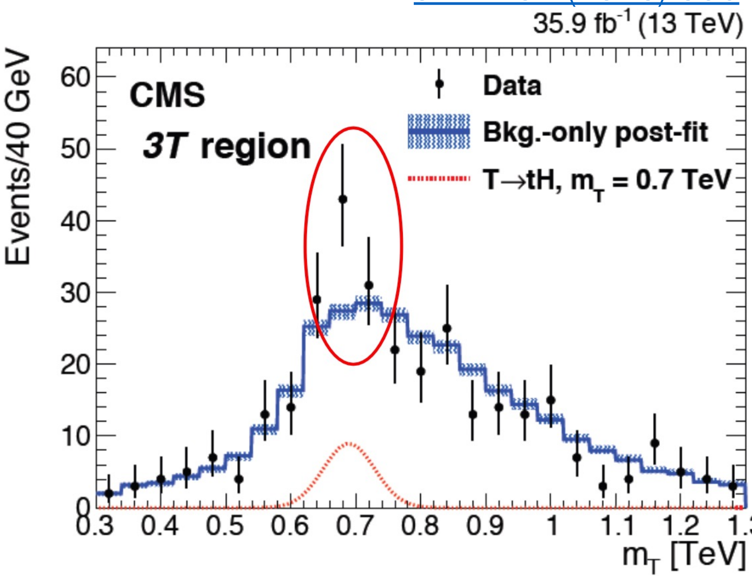

## Cut-based method → Neural Network

- Cut-based method: Categorizing events with a certain "selection" criterion on a data
- Selections are already optimized based on kinematic information for maximizing significance


Feed the network!

# Strategy

JHEP 01 (2020) 036



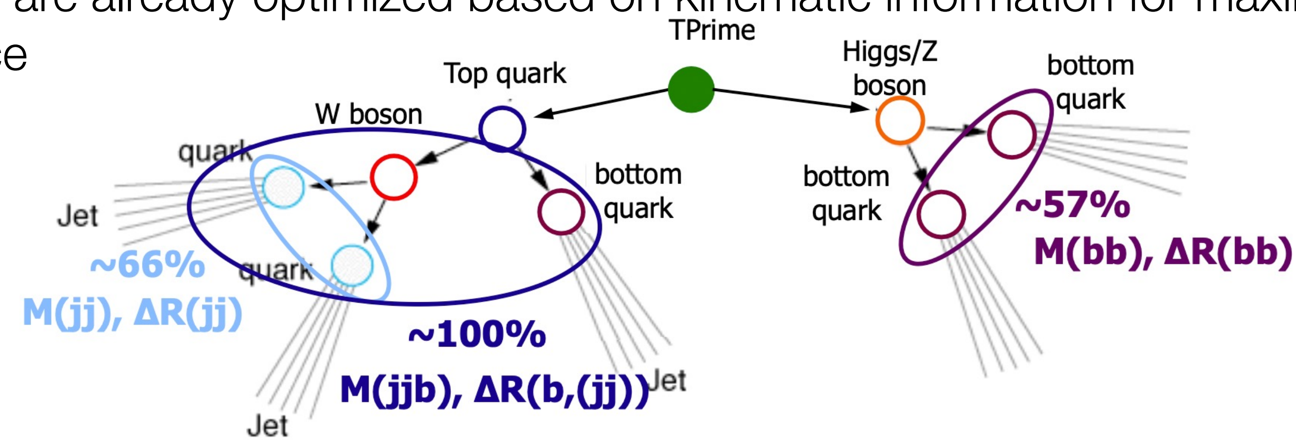

## Analysis using 2016 data in CMS

- Excess in T' mass @ 700 GeV is observed!

→ Might be able to improve significance with NN using Run2 data (even Run 3)

## Cut-based method → Neural Network

- Cut-based method: Categorizing events with a certain "selection" criterion on a data
- Selections are already optimized based on kinematic information for maximizing significance

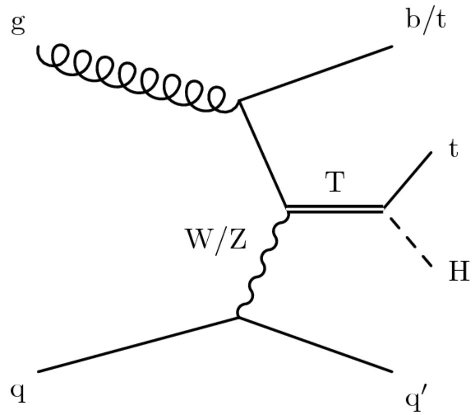



Feed these information to neural network!

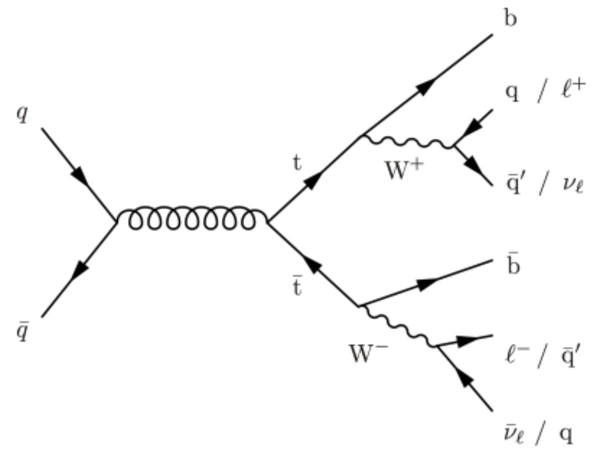
# Strategy

## Target Process

- Signal and Background Classification




Signal: T’→tH hadronic decay



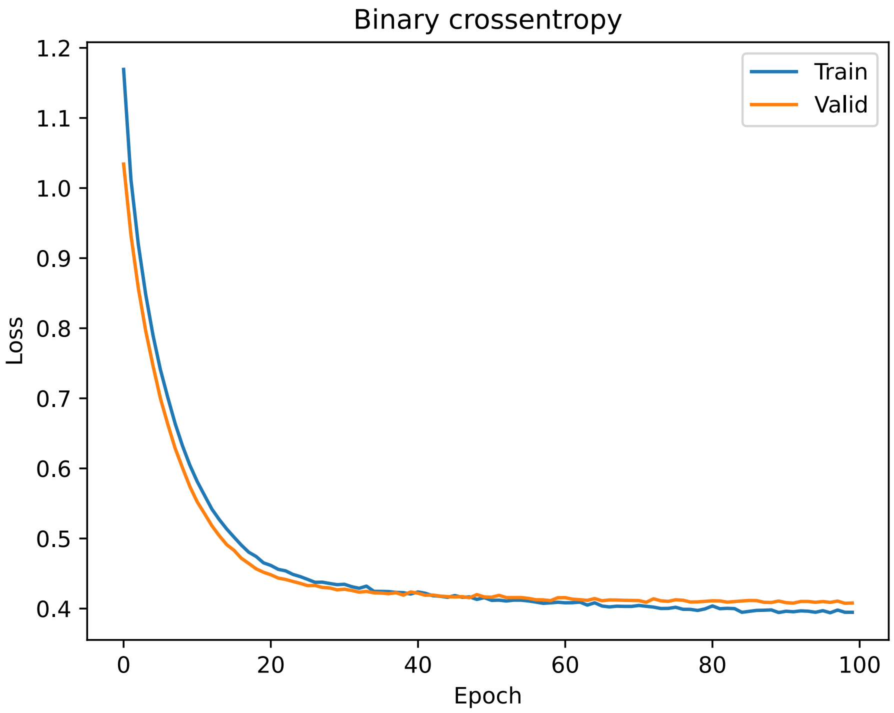

Background: ttbar hadronic decay

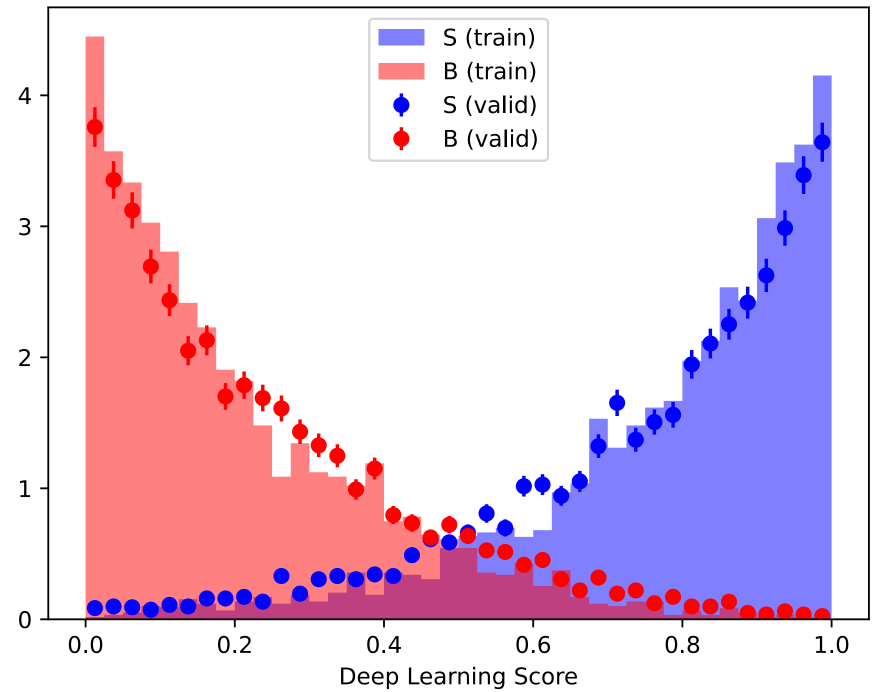
DNN Structure:
3 layers with 100 nodes
Dropout: 0.2
Activation: relu+sigmoid
Optimizer: Adam
Loss: binary_crossentropy
Batch size: 2048

## Neural Network details

- Structure: Simple DNN (3 layers with 100 nodes)
- Input set:
  - T’ signal MC sample (M=700 GeV) : ttbar hadronic decay = 1:1
  - 80 % for training, 20 % for validation
- Input features: 33
  - low-level features (angular position in the detector, energy of jets, …)
  - high-level features (features used in cut-based, angle between jets, ...)

# Current Status



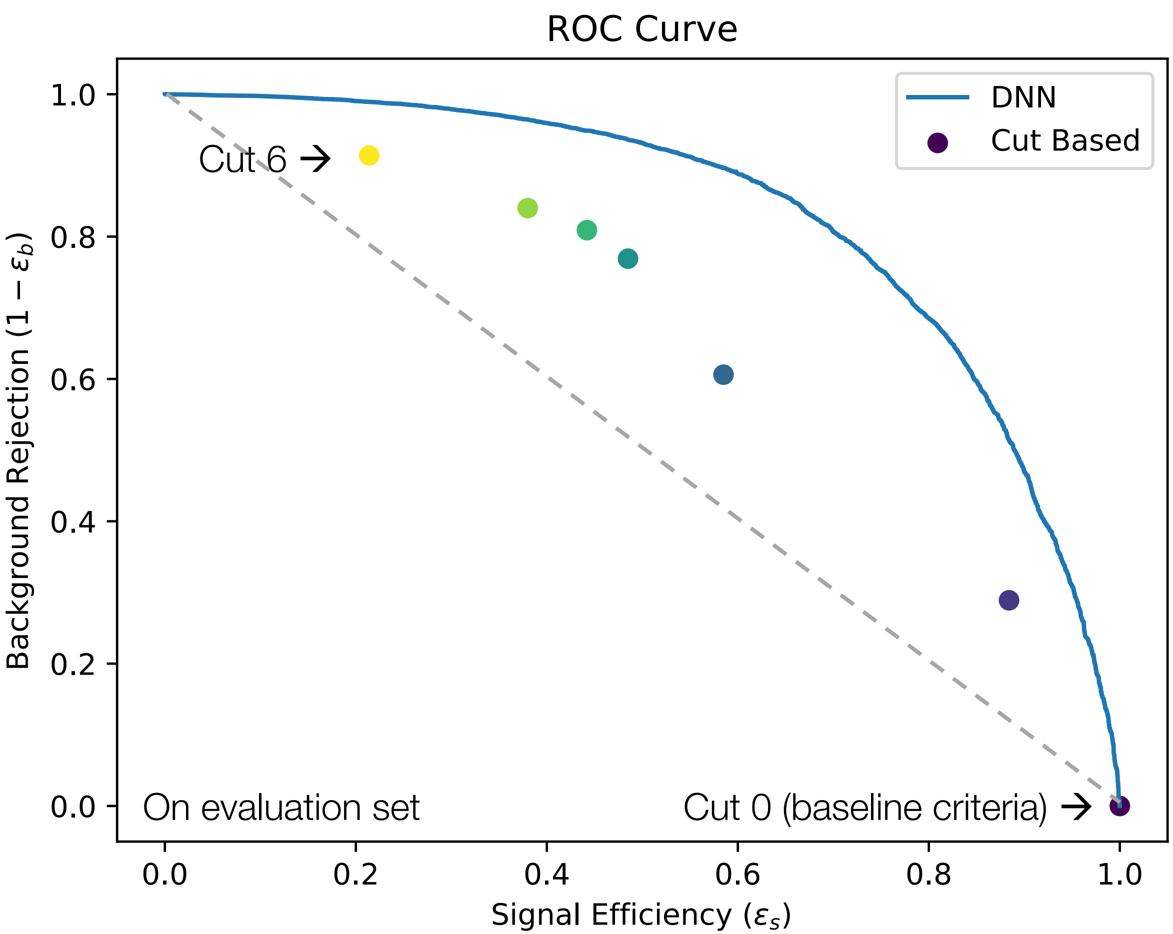




- Overtraining check:
  - Make sure if model is working not only on training set, but also on the real data
  - Check Loss curve + output distribution from Training and Validation set
  - Overtraining has not occurred!

Will perform on data too!

# Current Status

## Receiver Operating Characteristic curve

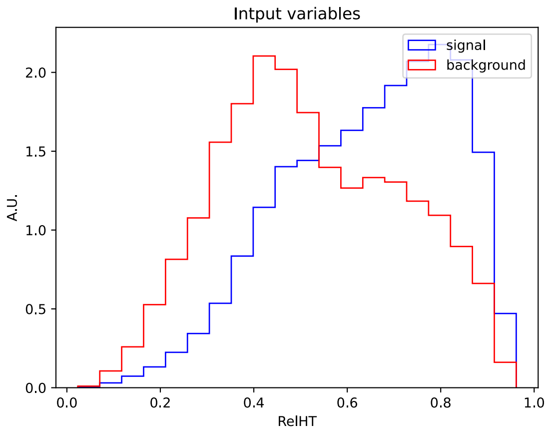


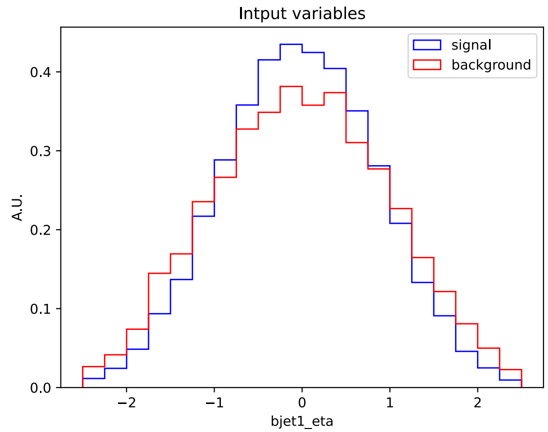

NN works better than cut-based method even without any optimization!

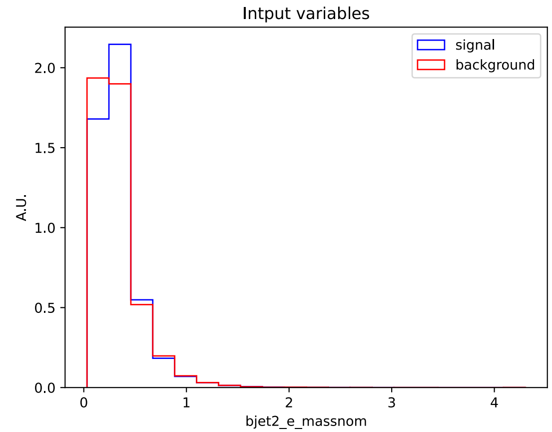
# Current Status

## Study on input features

- Initial question:
  - How should we determine which physics observables are “more important”?
  - Which input feature has the largest impact on the NN output node?




Which are “good” and “bad” observables?

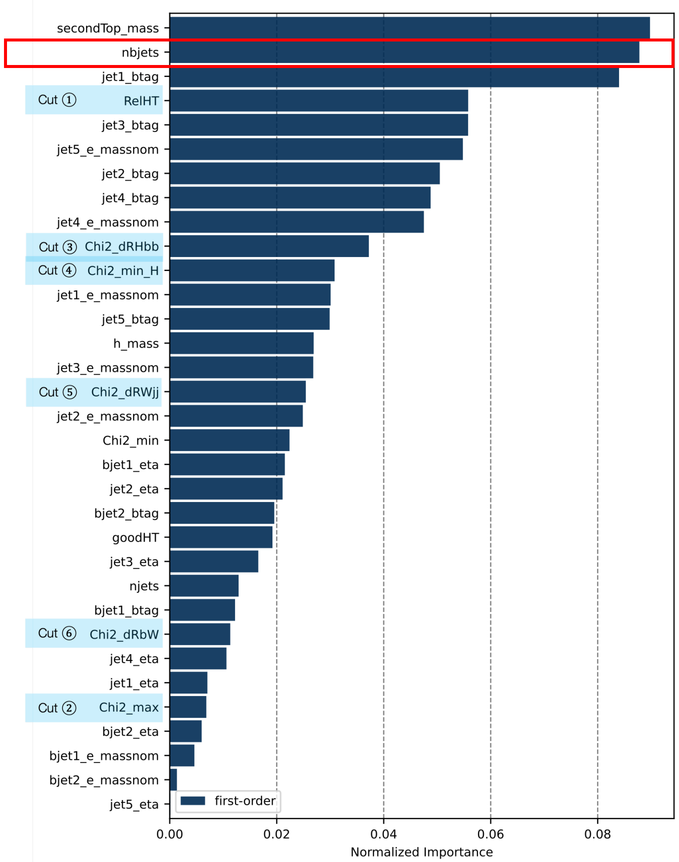
## Methodology

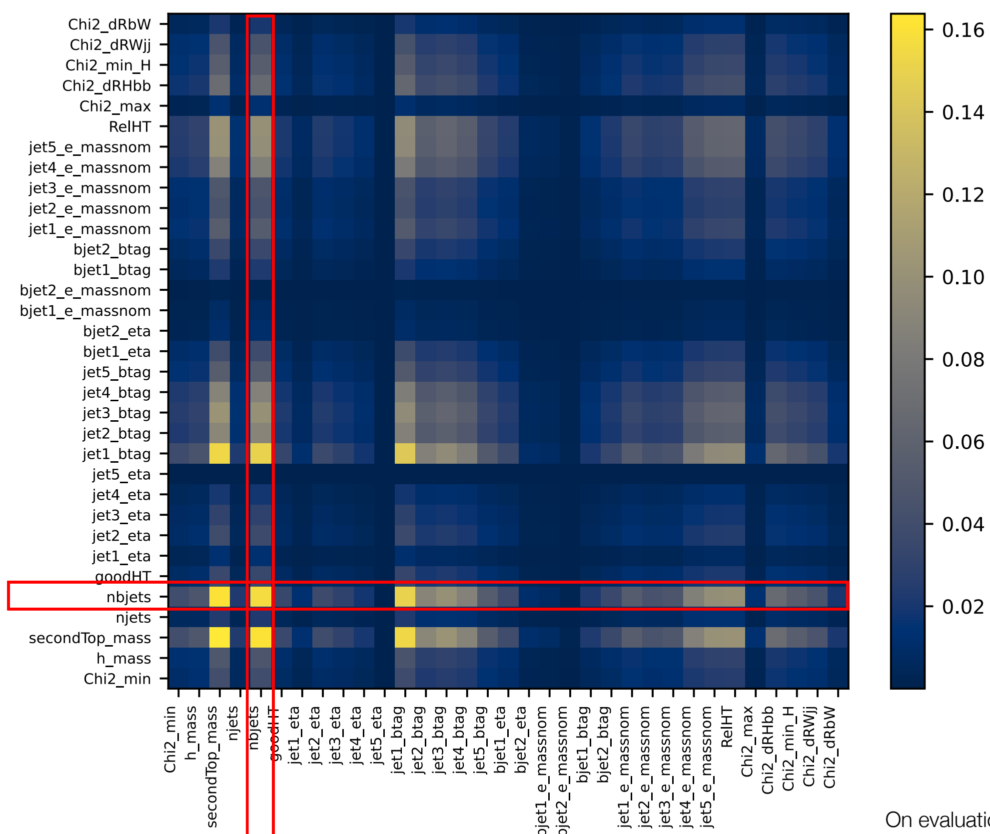
- Talyor expansion of the output function at the minima (model)
  - [arXiv:1803.08782](arXiv:1803.08782)
  - Calculate gradients of output(node) w.r.t. inputs(event)
  - Extract average gradient for each input features
  - Will be able to ”see” how much each variable “effects” on training model

# Current Status

## Feature importance from Talyor expansion

- $1^{st}$ order gradient:
  - Physical location of feature/marginal distributions: weight w_i for x_i
- $2^{nd}$ order gradient:
  - Gradient of each element of the source w.r.t target: weight w_ij for x_i * x_j




On evaluation set
Normalized by 1

# Conclusion

- Jet Calibration is important for data analysis.

- Analysis of Search for Tprime in Hadronic Final state is ongoing, while excess to 2016 CMS data was observed.

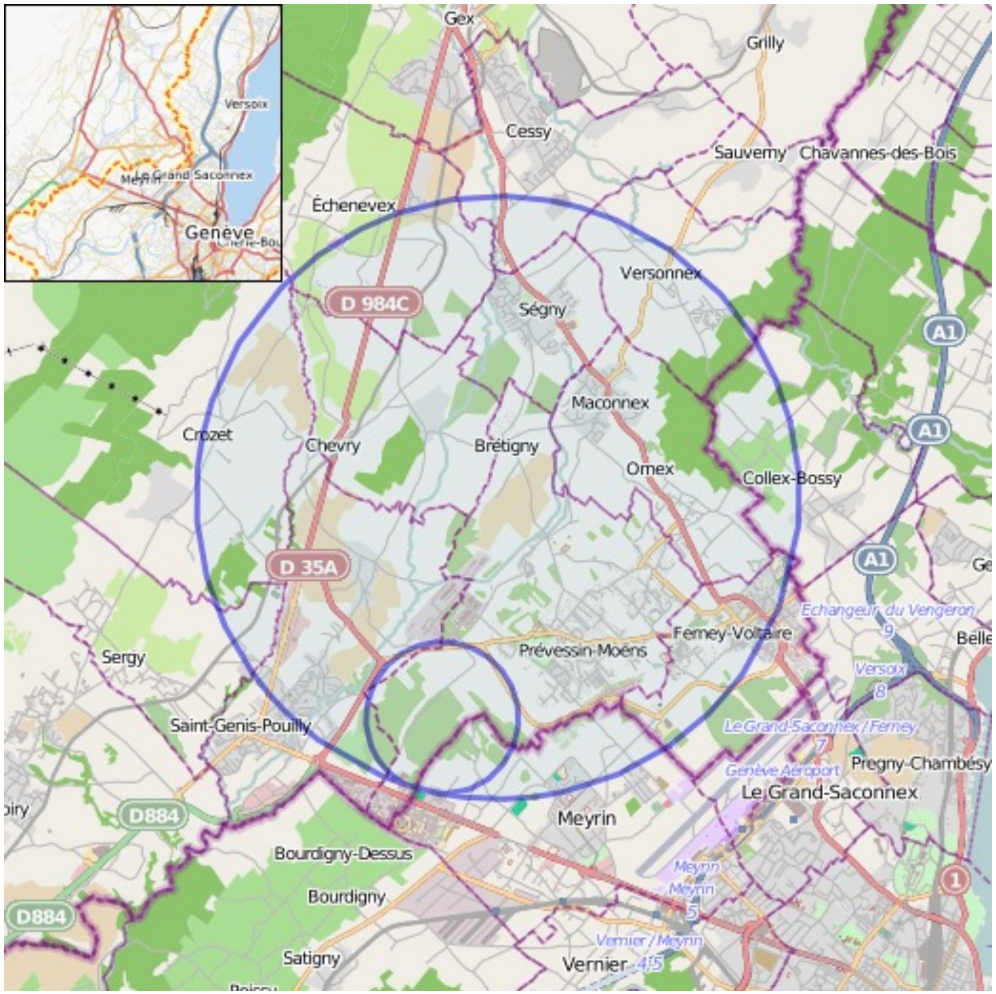
- Improving the significance of Tprime analysis using Neural Network is under study.
  - Has more performance than the cut-based method

- To do list:
  - Add more feature candidates and check the importance
  - Add other mass variation T' M=600 ~ 1200 GeV for training
  - Hyperparameter optimization
  - Adapt Graph Neural Network / parameterized Neural Network
  - Continue working on Run 3 data

# BACKUP

# Introduction

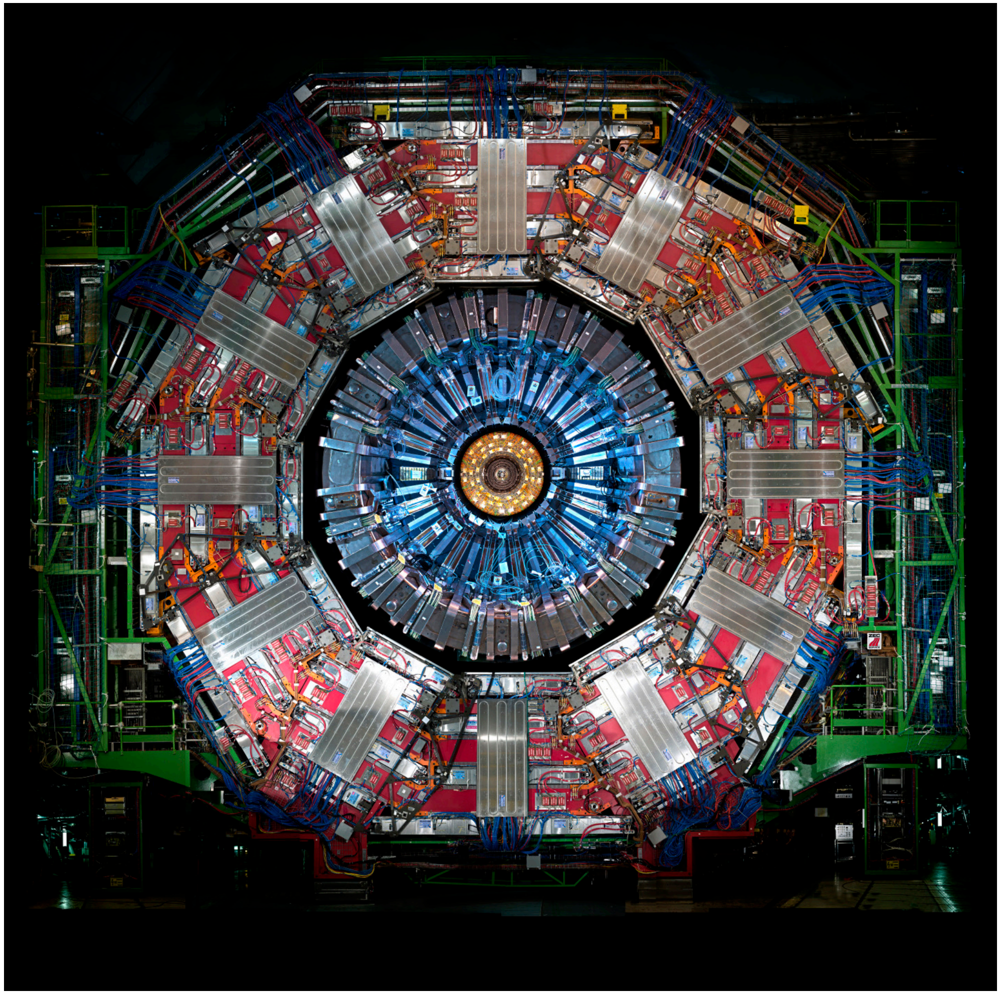
## Large Hadron Collider




- The LHC is a particle accelerator that pushes protons to near the speed of light
- It consists of a 27 km ring of superconducting magnets with accelerating structures that boost the energy of the particles along the way
- It produces lots of particle physics phenomena from proton-proton collisions at the center of mass energy = 13 TeV

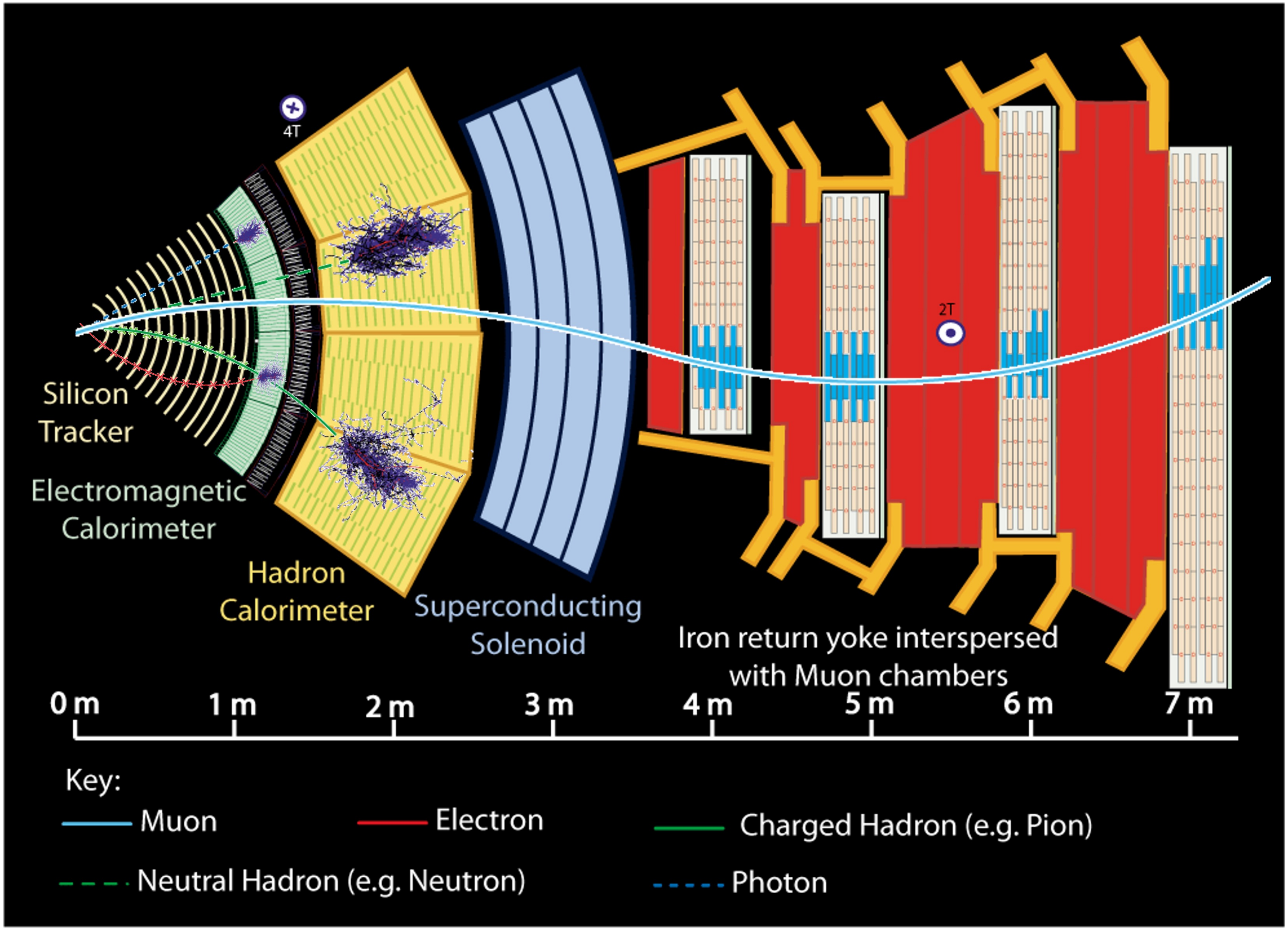
# Introduction

## Compact Muon Solenoid



- The CMS detector is located at one of the four collision points in LHC
- With 15 meters high and 21 meters long, CMS is “compact” for all detectors it contains
- It has the most powerful solenoid magnet ever made
- The discovery of Higgs boson at CMS and ATLAS detector in 2012 completed standard model
- However, some phenomena still exist that are not described by standard models

# Slice of CMS detector

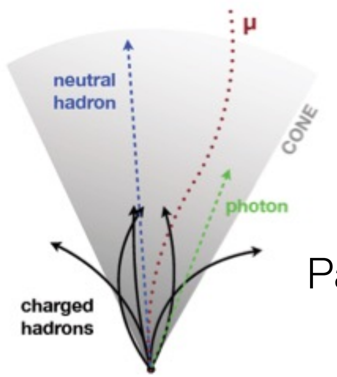


# Particle Flow

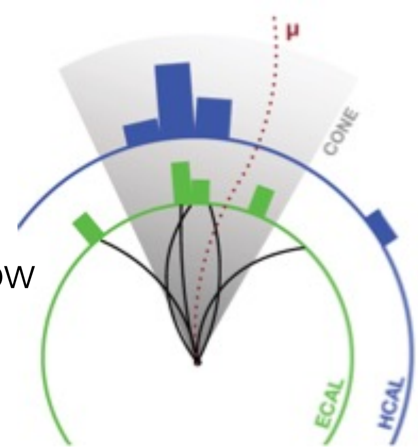
## How we reconstruct jet in CMS

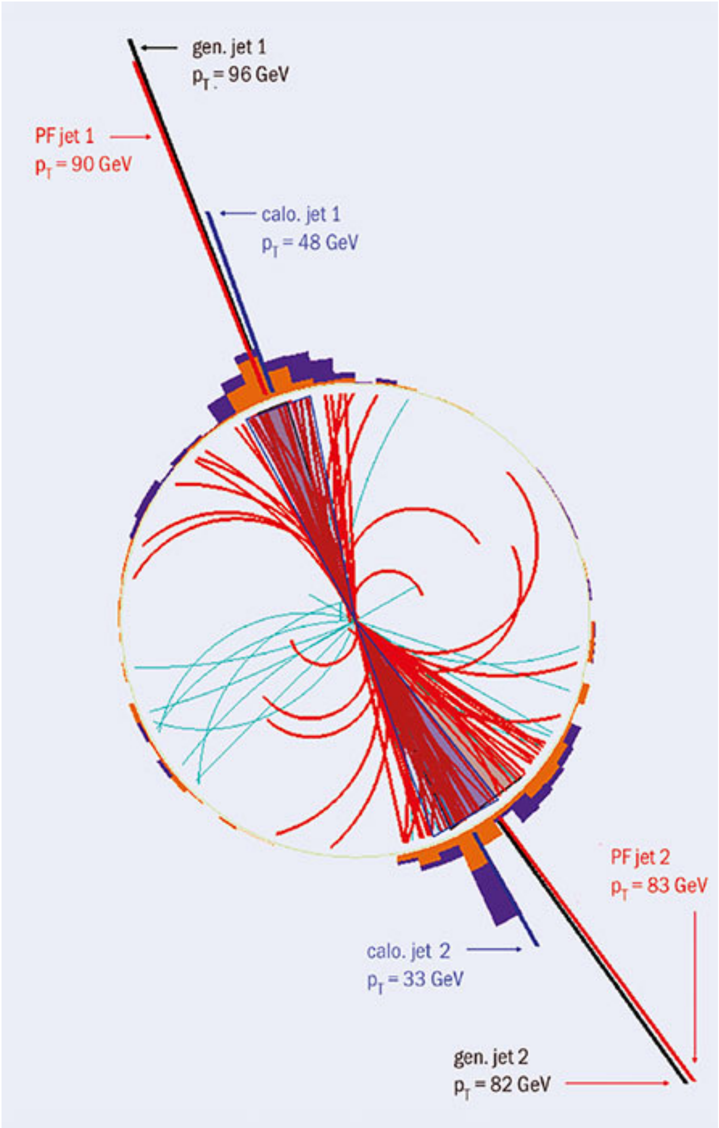
- Calorimeter based approach
- Jet-Plus-Track approach: Calorimeter jet + tracks

- Particle Flow approach
  - Reconstruct each particle individually in the event based on information from all sub-detectors
  - Jet composition:
    ~ 65% charged hadrons
    ~ 25% photons
    ~ 10% neutral hadrons




Detector
⇌
Particle Flow



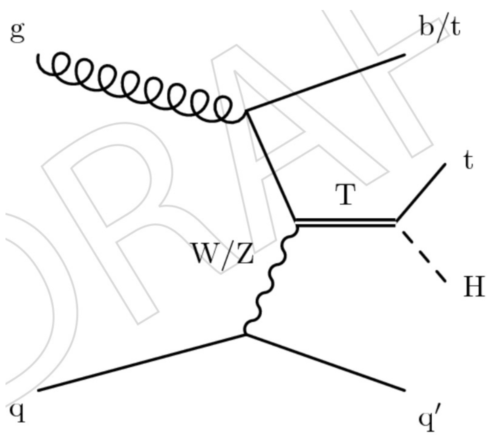




https://indico.in2p3.fr/event/14415/#2-pflow-in-cms by Florian Beaudette in Particle Flow Day @ LLR

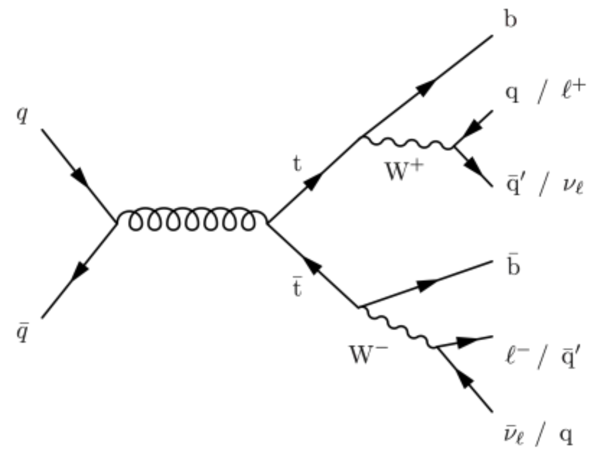
# How to avoid bias in NN

## Target Process

- Signal and Background Classification



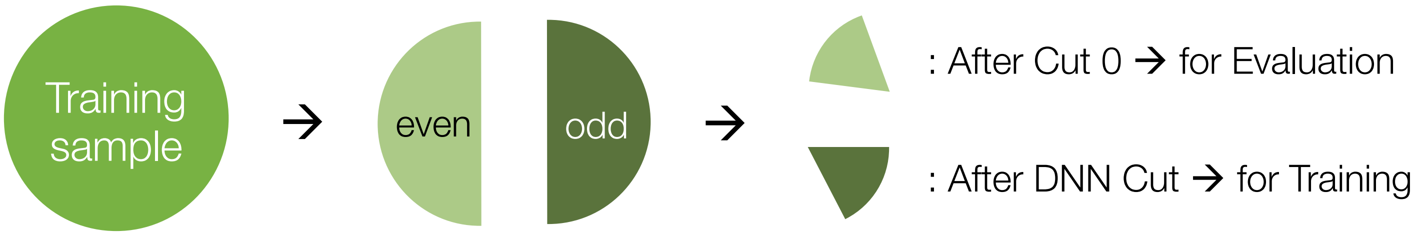

Tprime



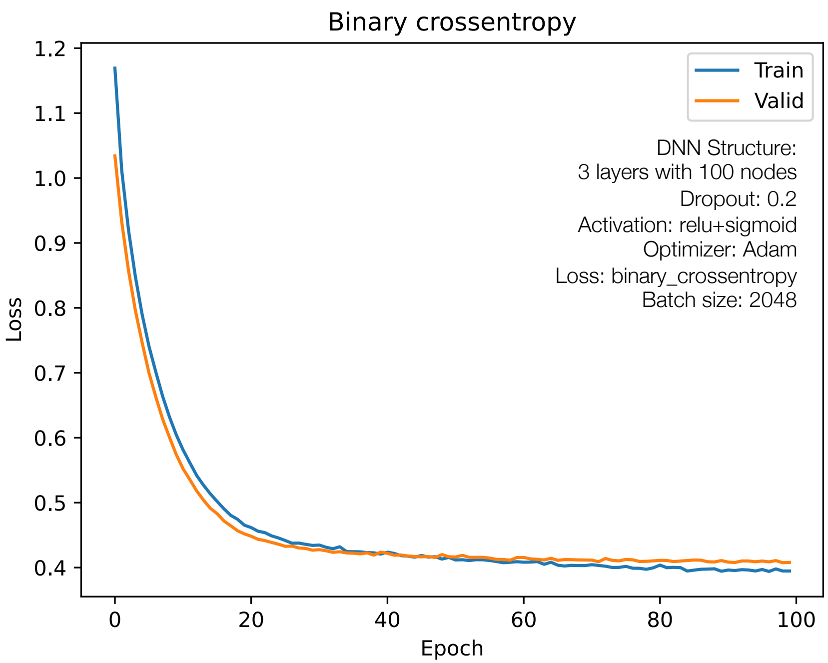

ttbar hadronic decay

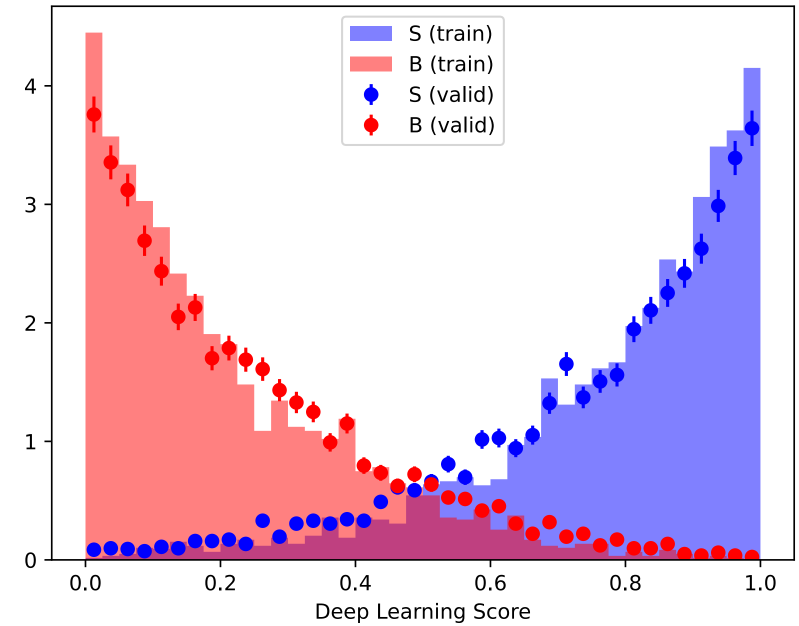
## Strategy

- Train on Tprime700 Hadronic + TTToHadronic (1:1 training)
- Selection for DNN: HLT + njets >= 6 + nbjets (DeepJet Medium) >= 3
- Compare ROC curves with cutBased (signal efficiency vs background rejection)
    - Evaluate NN at the level of Cut 0 for the pair comparison



# Overtraining check

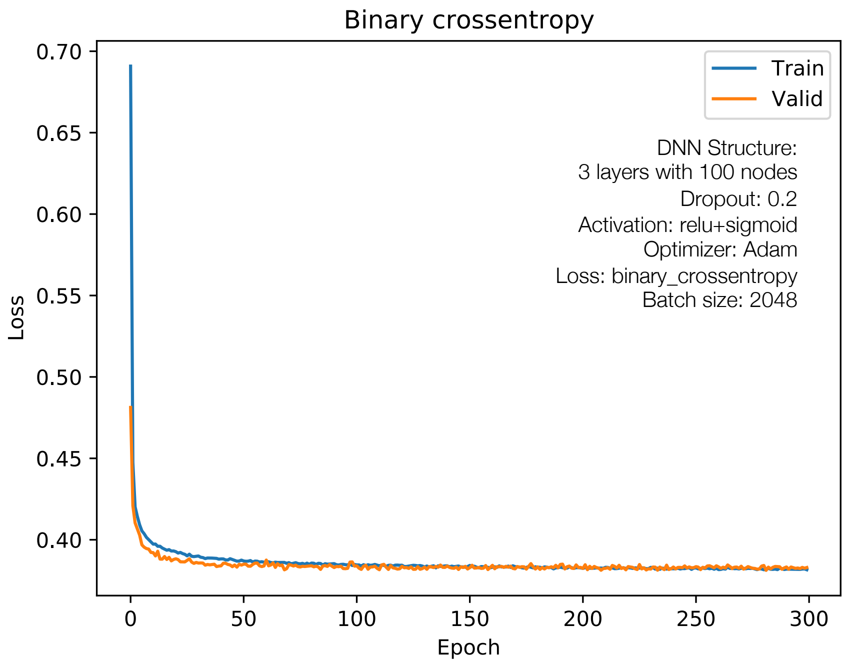


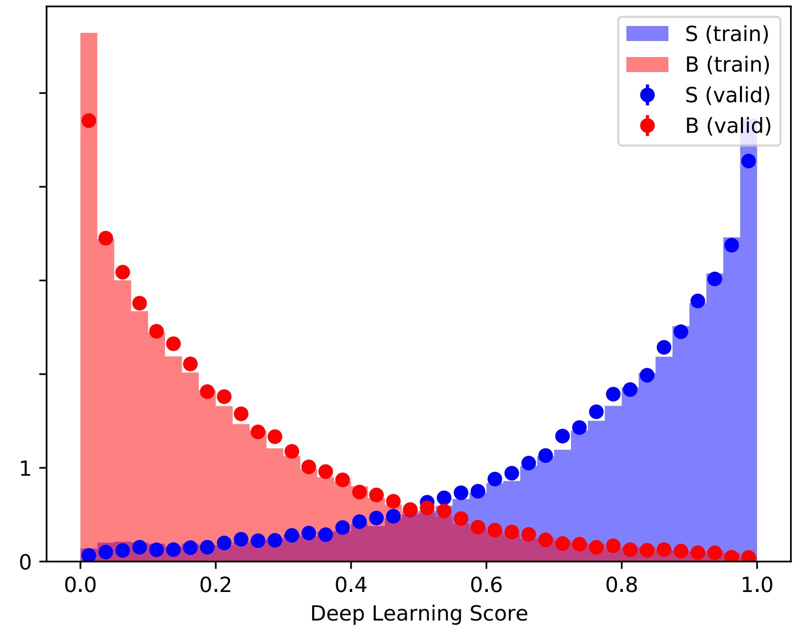




## Detail

- Trained in CC server (CPU without slurm - Training time: 10ms/epoch)
- Input set : Half of TprimeBToTH_M-700 after selection (odd numbered event, 23210 entries)
  - 80 % for training, 20 % for validation
  - Keep even numbered event for evaluation: to avoid bias (using the same event) for performance estimation
  - Epoch: 100 → Validation Loss / Acc are stable, does not diverge yet

# Training with more statistics



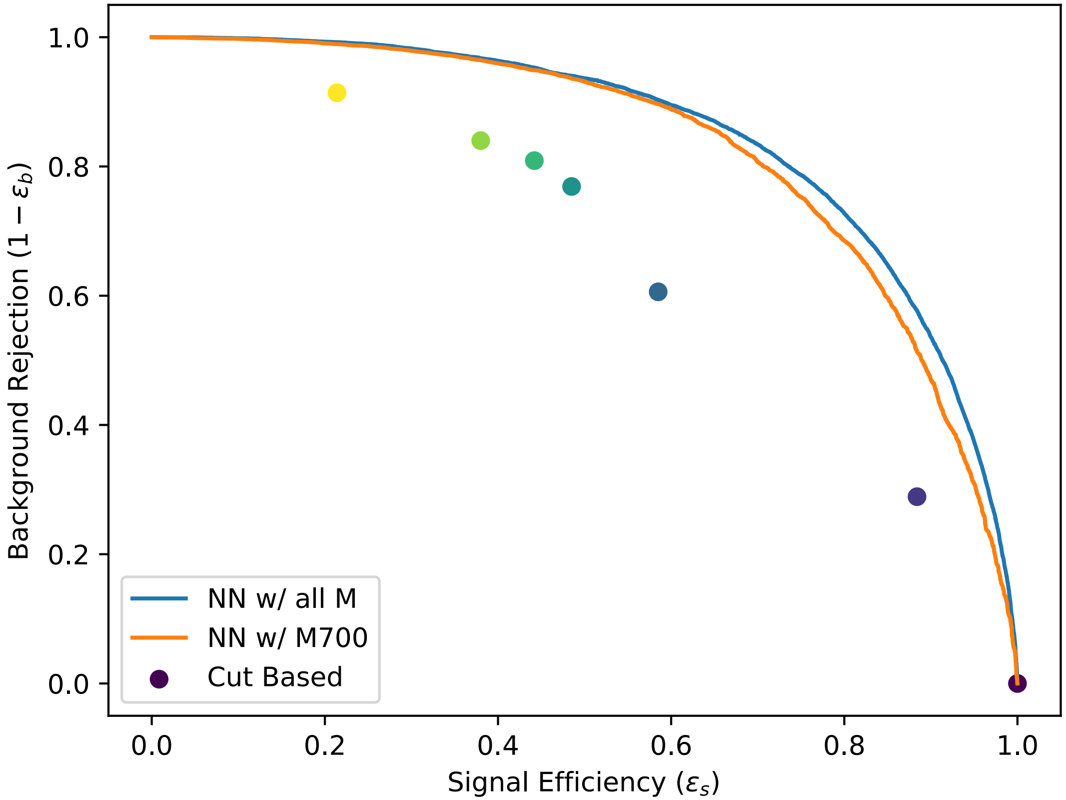

## Strategy

- Trained in CC server (Training time: 1s 8ms/epoch)
- Do the same with more statistics from different mass range
- Train on signal samples M=600~1200 GeV (181724 entries (M700 entries * 7)) + TTToHadronic
- With the same input features, same architecture (but more epoches)

# ROC cureves

## Evaluation

- Evaluation is performed in Cut 0 with odd numbered events




- Performance is slightly increased
- M-{600,700,....,1200} GeV samples were used

# Feature importance

## Motivation

- Initial question: What are the input features with the largest impact on the NN output nodes?
- Extract average gradient for each input features
- Will be able to ”see” how much each variable “effects” on training model
- arXiv:1803.08782

## Methodology

- Talyor expansion of the output function at the minima (model)
- Tensorflow.GradientTape()
  - Allow us to record the history of operations applied to target input features
  - Calculate gradients of output(node) w.r.t. inputs(event)
- $1^{st}$ order: Physical location of feature/marginal distributions – weight w_i for x_i
- $2^{nd}$ order: Curvature of NN output function - correlations across two features: Gradient of each element of the source w.r.t target – weight w_ij for x_i * x_j

$$\langle t_\alpha \rangle = \frac{1}{N} \sum_{k=1}^{N} \left| t_\alpha \left( \left\{ x_j^{(k)} \right\} \right) \right|$$

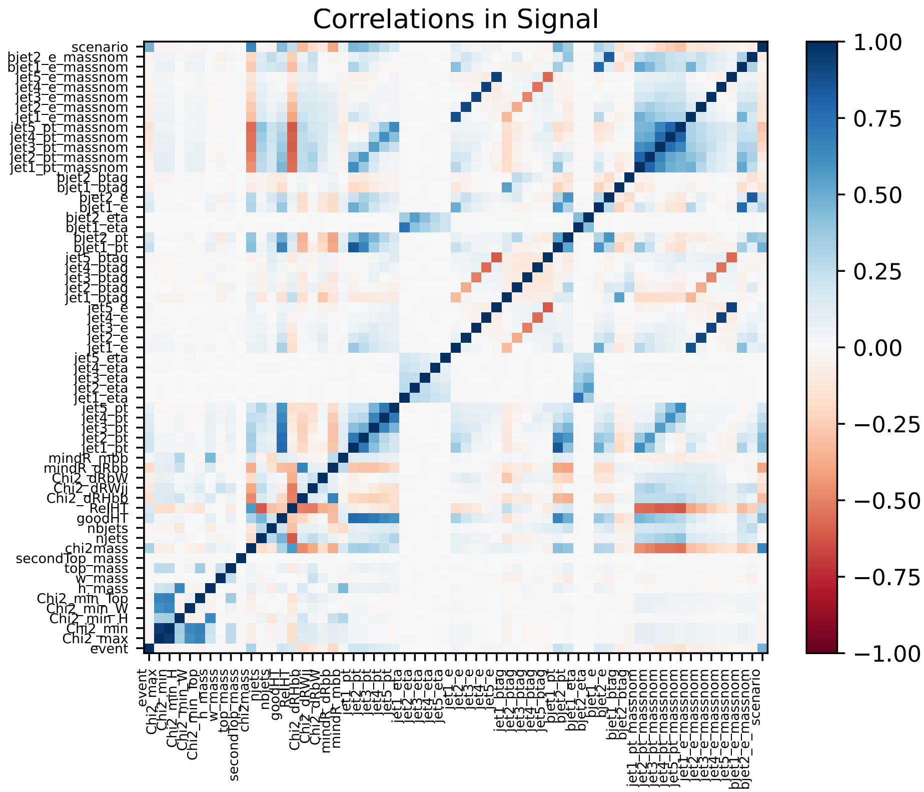
$N$ : Sample size

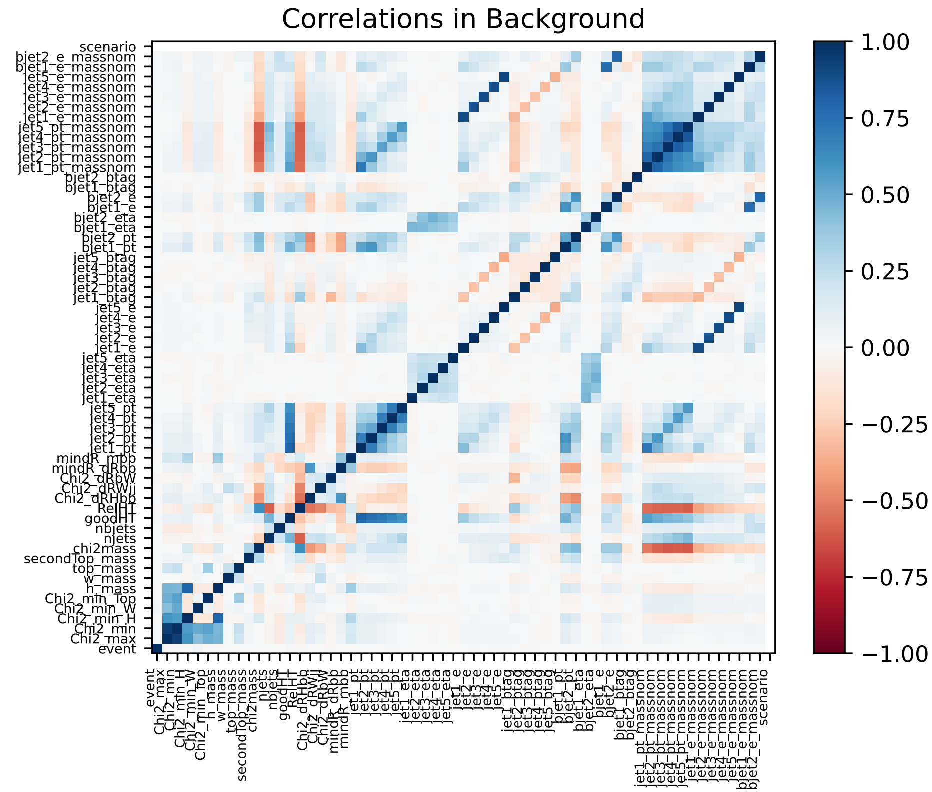
$t_\alpha$ : Taylor coefficient labeled by $\alpha$

- $\langle t_\alpha \rangle$ is the arithmetic mean of $|t_\alpha|$, evaluated on the whole input space that is sampled by the test data set.
- Introduce nomenclature of *generalized features* of the input feature space:

| | |
|---|---|
| $\alpha = x_1, x_2, \ldots$ | 1. order feature of input space (~ 1. order derivative) |
| $\alpha = x_1 x_1, x_1 x_2, \ldots$ | 2. order feature of input space (~ 2. order derivative) |
| $\alpha = x_1 x_1 x_1, x_1 x_1 x_2, \ldots$ | 3. order feature of input space (~ 3. order derivative) |
| $\vdots$ | $\vdots$ |

# Correlation Matrices



# Physics observables in cut-based method

| Basic Selection Criteria |
|---|
| Trigger and $p_T$,$\eta$ and $n_b^{DeepCSVT} \geq 3$ |
| $j^1_{p_T} > 170$ GeV/$c$, $j^2_{p_T} > 130$ GeV/$c$, $j^3_{p_T} > 80$ GeV/$c$ and $H_T > 500$ GeV/$c$ |
| $\chi^2 < 15$ |
| 2nd Top Mass$> 250$ GeV/$c^2$ |
| Higgs Mass$> 100$ GeV/$c^2$ |

| Label | Cuts |
|---|---|
| Cut 0 | Basic selection |
| Cut 1 | Relative $H_T > 0.4$ |
| Cut 2 | Max($\chi^2$) $< 3$ |
| Cut 3 | $\Delta R(b_{Higgs}, b_{Higgs}) < 1.1$ |
| Cut 4 | $\chi^2_{Higgs} < 1.5$ |
| Cut 5 | $\Delta R(j_W, j_W) < 1.75$ |
| Cut 6 | $\Delta R(b_{Top}, W) < 1.2$ |

| Criteria | Quick description |
|---|---|
| $P_T$ of each jets | Signal should have harder $P_T$ than QCD |
| $p_T$ (T′) | |
| Nb Good Jets | QCD could have larger jet multiplicity |
| $\chi^2$ | Signal peaks at 0 |
| $\chi^2_{Higgs}$ | Signal peaks at 0, background is larger |
| $\chi^2_{Top}$ | Signal peaks at 0, QCD is larger |
| $\chi^2_W$ | Signal peaks at 0, QCD is larger |
| Max($\chi^2$) | Maximum ($\chi^2_{Higgs}$,$\chi^2_{Top}$,$\chi^2_W$) |
| M($Higgs_{cand}$) | Invariante mass of Higgs candidate |
| M($top_{cand}$) | Invariante mass of Top candidate |
| M($W_{cand}$) | Invariante mass of W candidate |
| M($W_{cand}$+$Higgs_{cand}$) | Invariante mass of sum of Higgs and W candidate [4 jet mass] |
| M(6 Jets) | Invariante mass of the 6 selected jets |
| 2nd Top Mass | Invariante mass of Higgs candidate and $6^{th}$ $jet$ |
| $\frac{M_{top}-M_{Higgs}}{M_{top}+M_{Higgs}}$ | Ratio of invariante masses |
| $\frac{M^{2nd}_{top}+M^{2nd}_W}{M_{Higgs}}$ | Ratio of invariante masses |
| $\frac{M(W+H)}{M(Top+H+6^{th}Jet)}$ | Ratio of invariante masses |
| $H_T$ | |
| Relative $H_T$ | $\frac{p_T(H_{cand})+p_T(top_{cand})}{H_T}$ |
| New Relative $H_T$ | $\frac{p_T(H_{cand})+p_T(top_{cand})+p_T(6^{th}Jet)}{H_T}$ |
| $\Delta R(T', 6^{th}Jet)$ | LO signal tends to give back to back results |
| $\Delta R(b_{Higgs}, b_{Higgs})$ | Separation between the two jets making the Higgs candidate |
| $\Delta R(j_W, j_W)$ | Separation between the two jets making the W candidate |
| $\Delta R(Higgs, Top)$ | Separation between the Higgs and Top candidates |
| $\Delta\eta(W, H)$ | Eta separation between the Higgs and W candidates |
| $\Delta R(b_{Top}, W)$ | Separation between b-jet making the top candidate and the W candidate |
| $\Delta R \times \Delta R$ | Product of all $\Delta R$ in the event. All $\Delta R$ are computed to be peaking at 0 |
| Max($\Delta R$) | Maximum of all $\Delta R$ in the event. All $\Delta R$ are computed to be peaking at 0 |
| $\Delta\phi(Higgs, Top)$ | Phi separation between the Higgs and Top candidates |
| $\frac{P^{2ndtop}_T-P^{top}_T}{P^{2ndtop}_T}$ | Ratio of $P_T$ candidates |
| $\frac{p_T(H_{cand})-p_T(top_{cand}}{T'}$ | Ratio of $P_T$ candidates |
| $\frac{T'}{p_T(H_{cand})} - \frac{T'}{p_T(top_{cand})}$ | Ratio of $P_T$ candidates |