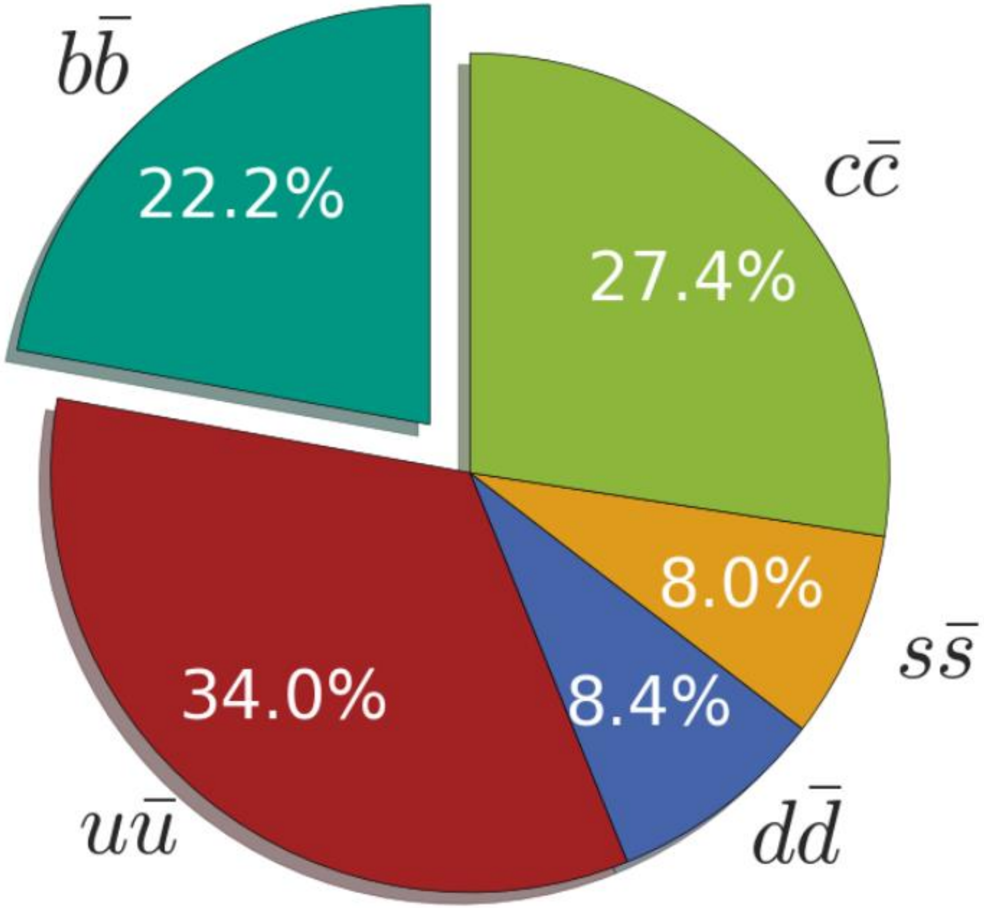
Deep Continuum Suppression

Oskar Tittel – Young Scientists Workshop Ringberg 2022

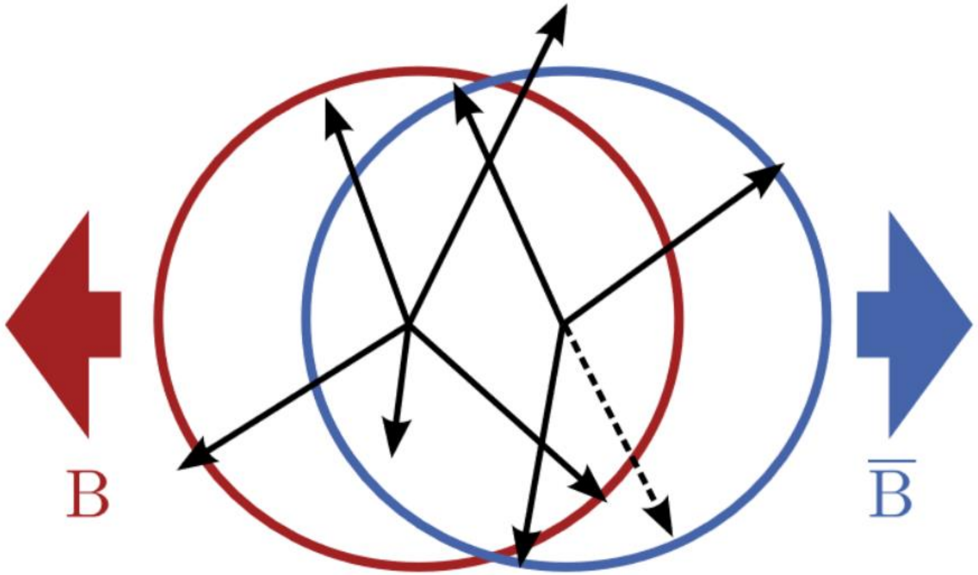
Continuum Background

- Belle II runs at $\sqrt{s} = 10.58\ GeV$
  → $\Upsilon\ (4S)$ resonance
- The dominating source of hadronic background comes from $q\bar{q}$ bar events below the center of mass energy ($u\bar{u}, d\bar{d}, c\bar{c}, s\bar{s}$)
- These events are called "continuum background" and are dominating in many analyses

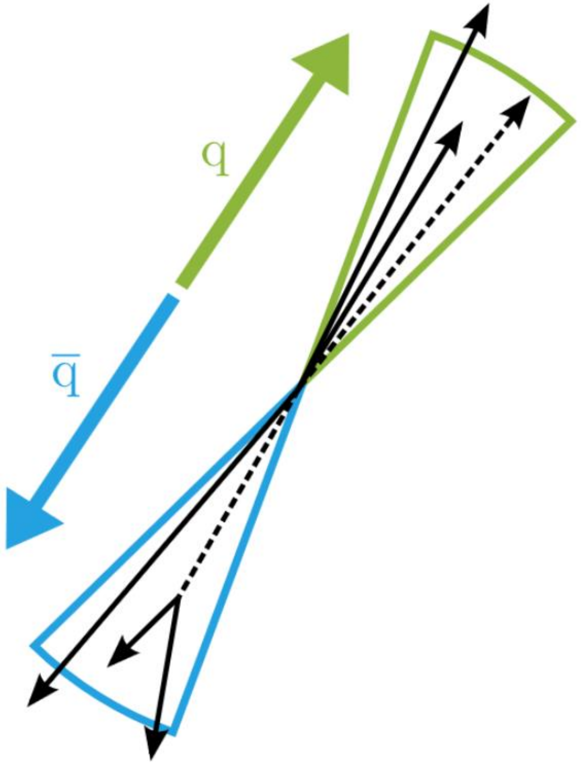


Event Shapes

- $B\bar{B}$-events:
  decay spherically in the CMS, since the B masses add up to $10.58\ GeV$

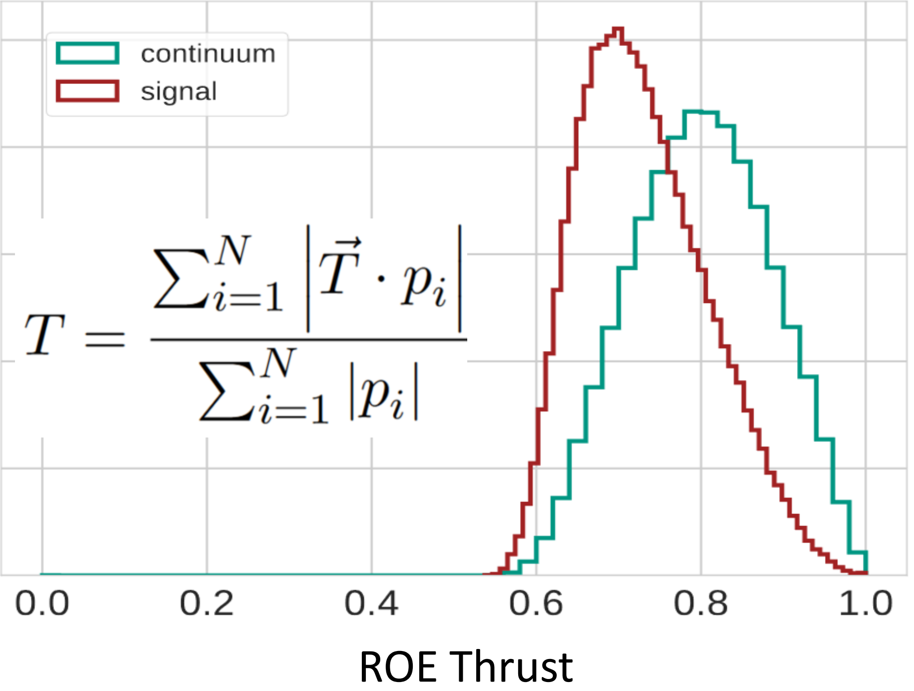


- $q\bar{q}$-events:
  are produced jet-like back-to-back in the CMS, since they are lighter, and the residual detector energy is transformed into momentum

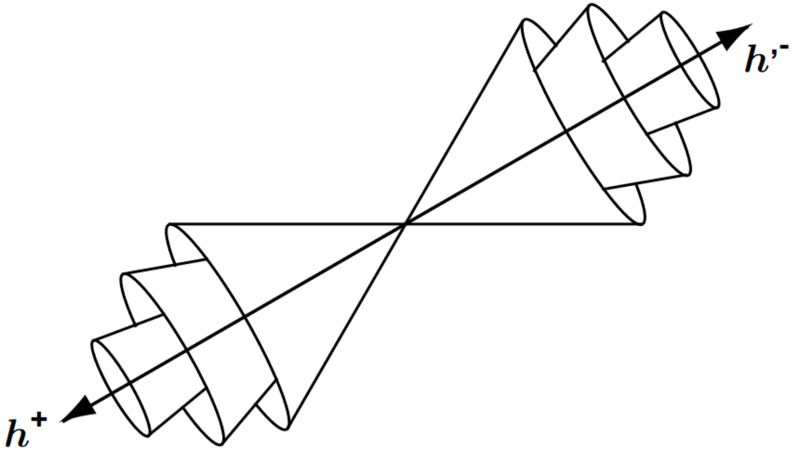


Classical Continuum Suppression

- Exploit the differences of the events shapes in the CMS by building special event shape variables: Thrust, CLEO Cones,…

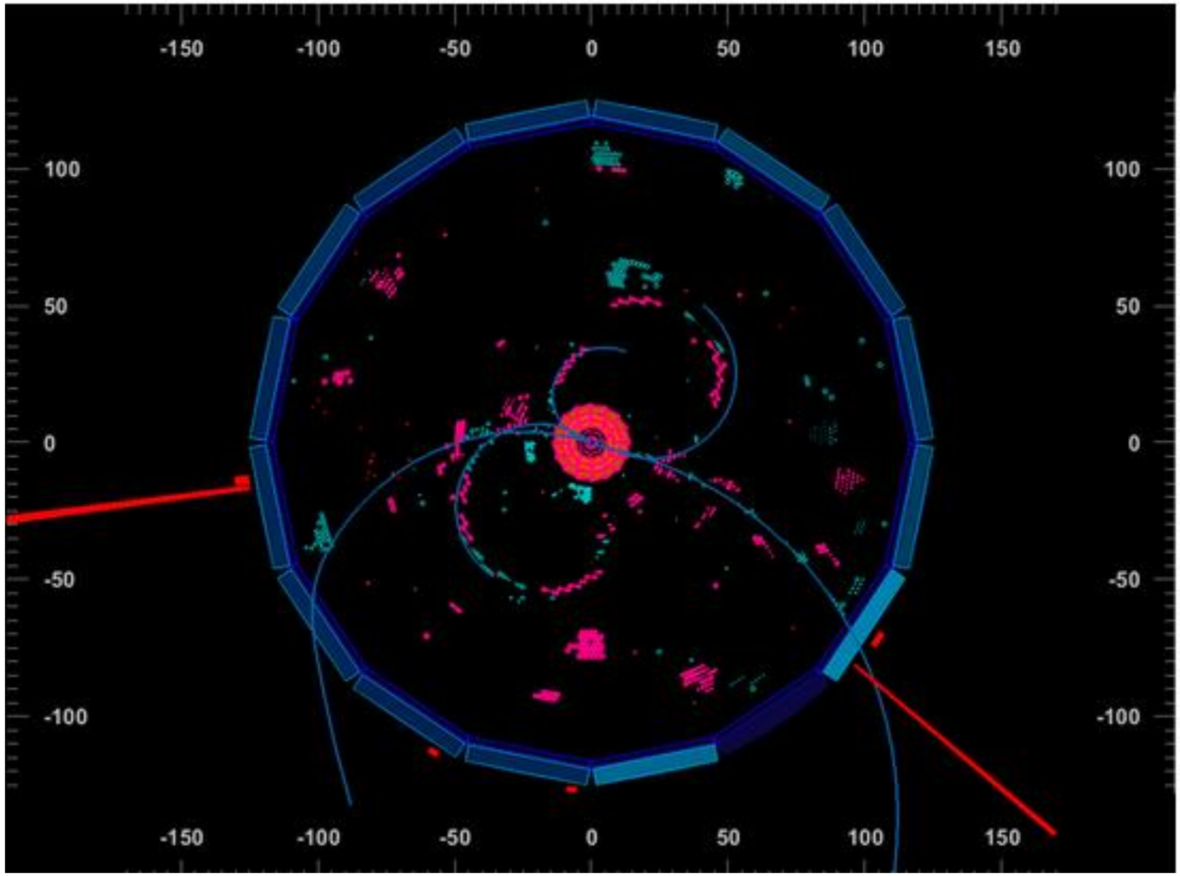


- In total, around 30 event shape variables are used and fed into a BDT, which is trained on MC signal and continuum events

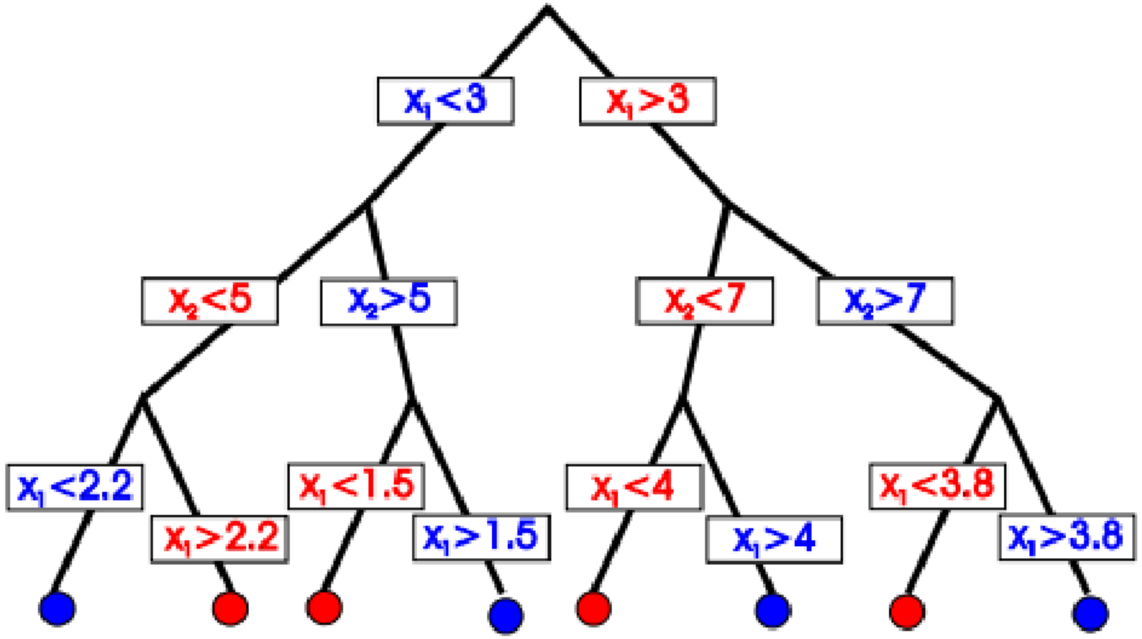


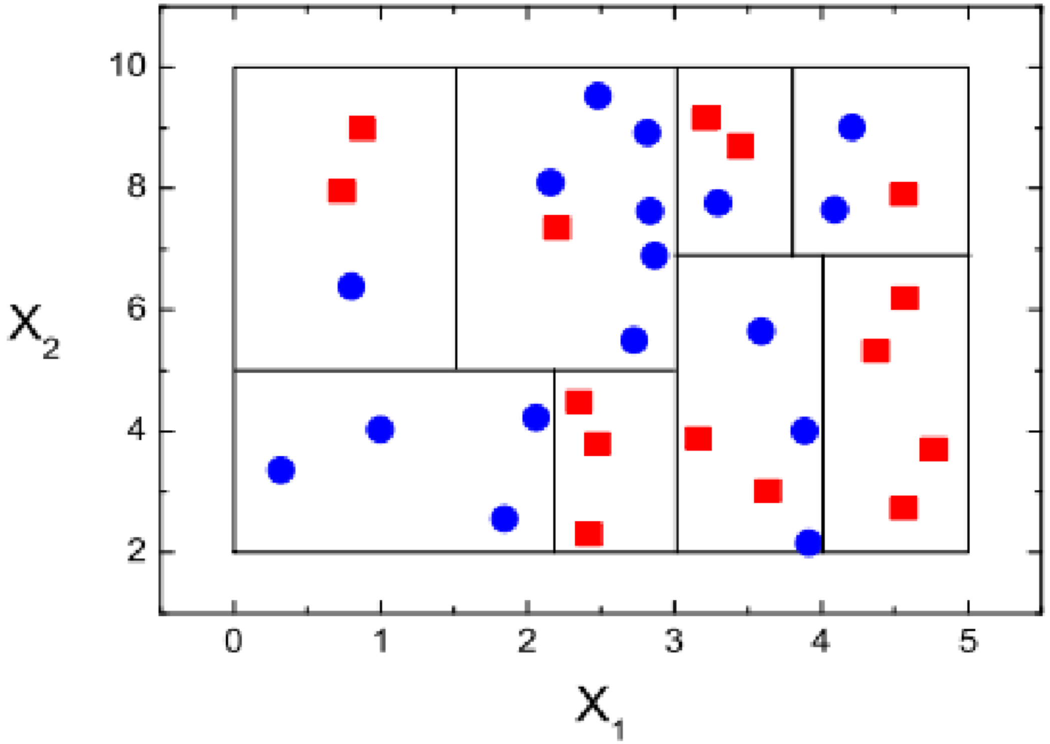
Low Level Variables

- Instead of constructing high-level event shape variables, directly access the measurements coming from the detector
- This includes momenta of tracks and cluster, specific cluster information, track variables like number of hits in the CDC or distance from the decay vertex to the IP →~360 variables



Decision Tree

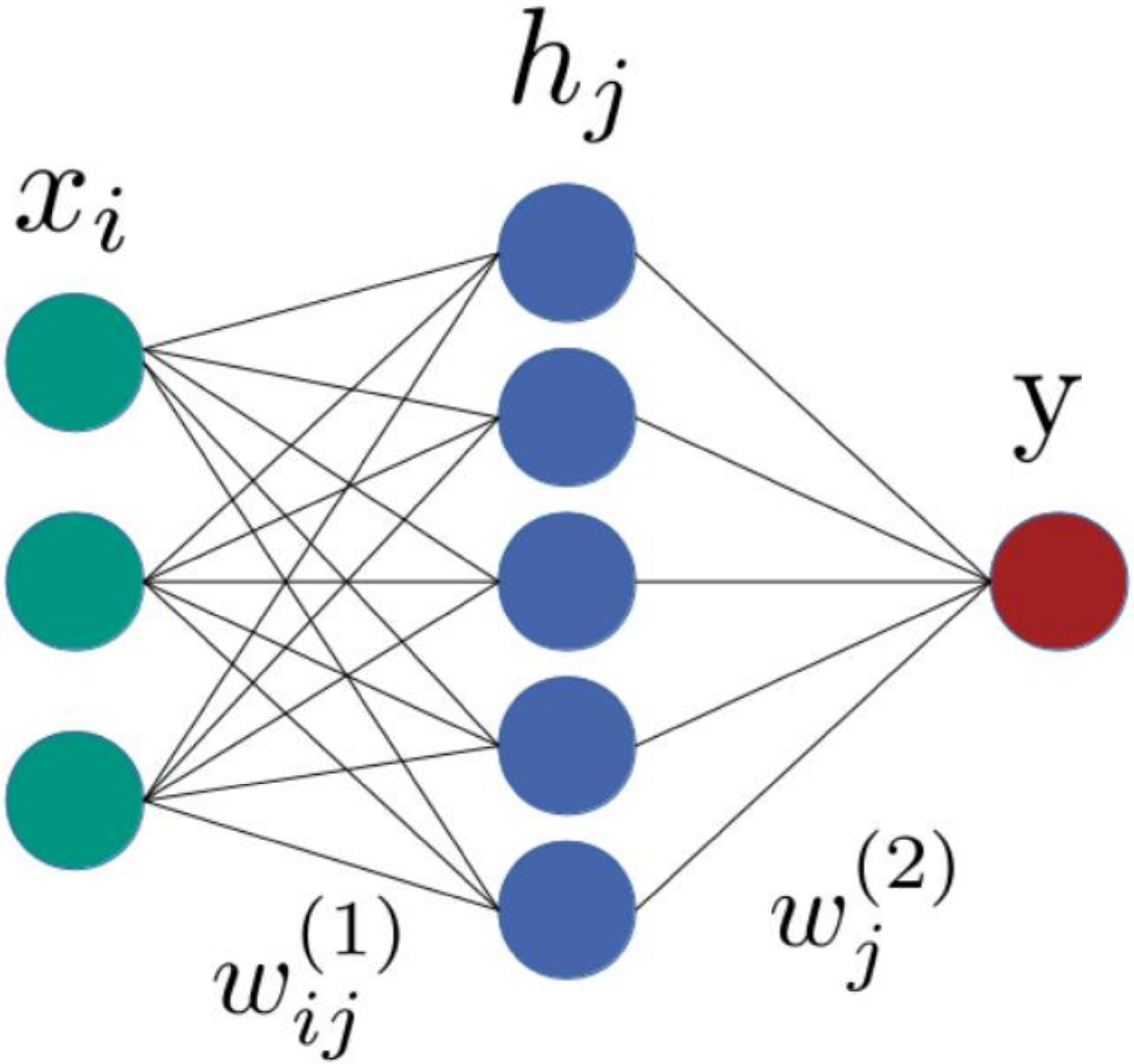




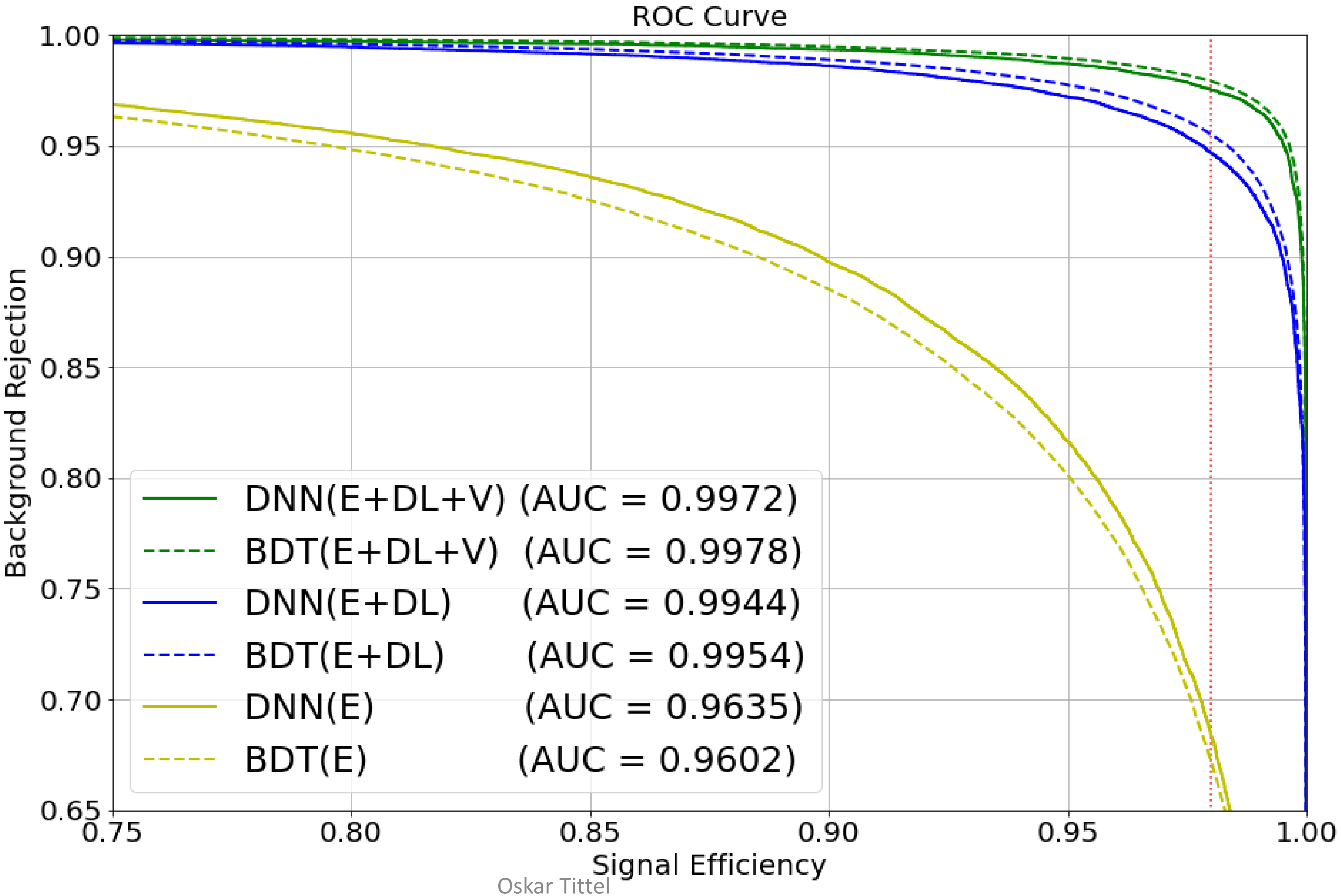
- Tree adds distinction rules until no gain in precision can be reached any more
- Might be slow with high amounts of data

Deep Neural Network

- Input data is fed into first layer of DNN
- Data is weighted and transformed by activation function in single or several hidden layers and then turned into output with final weighting
- Weights are then adjusted such that performance is increased

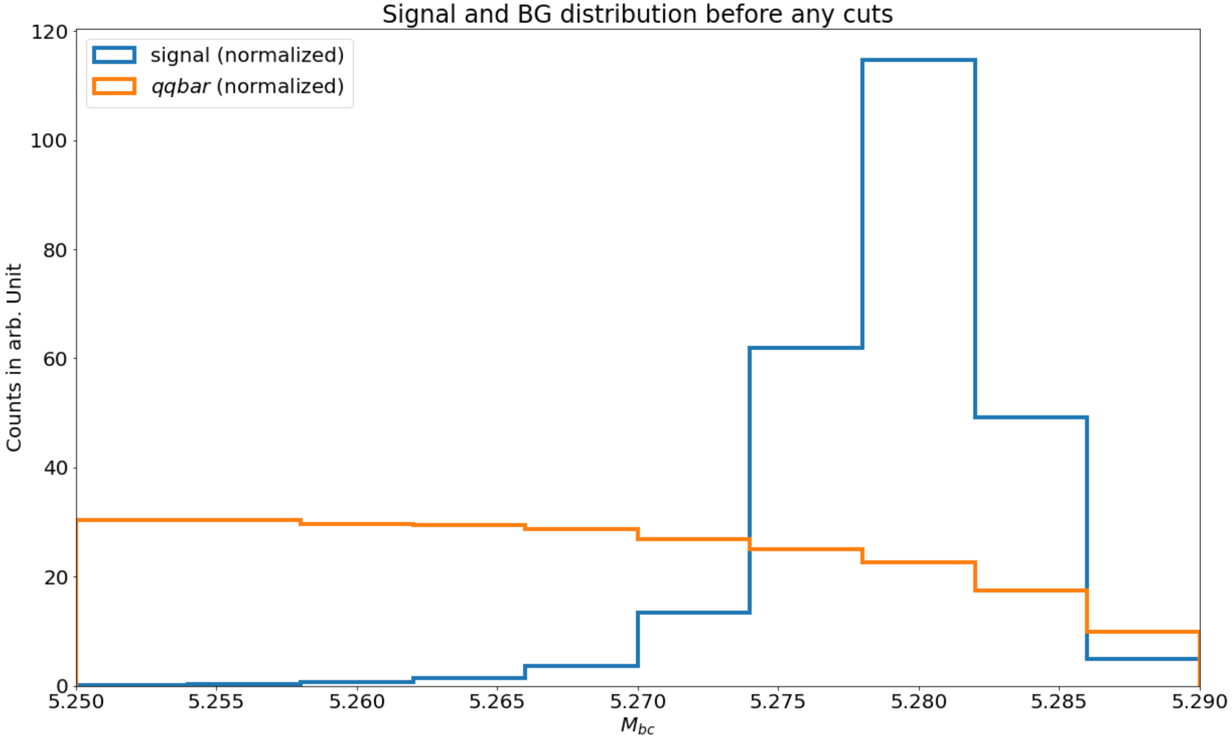


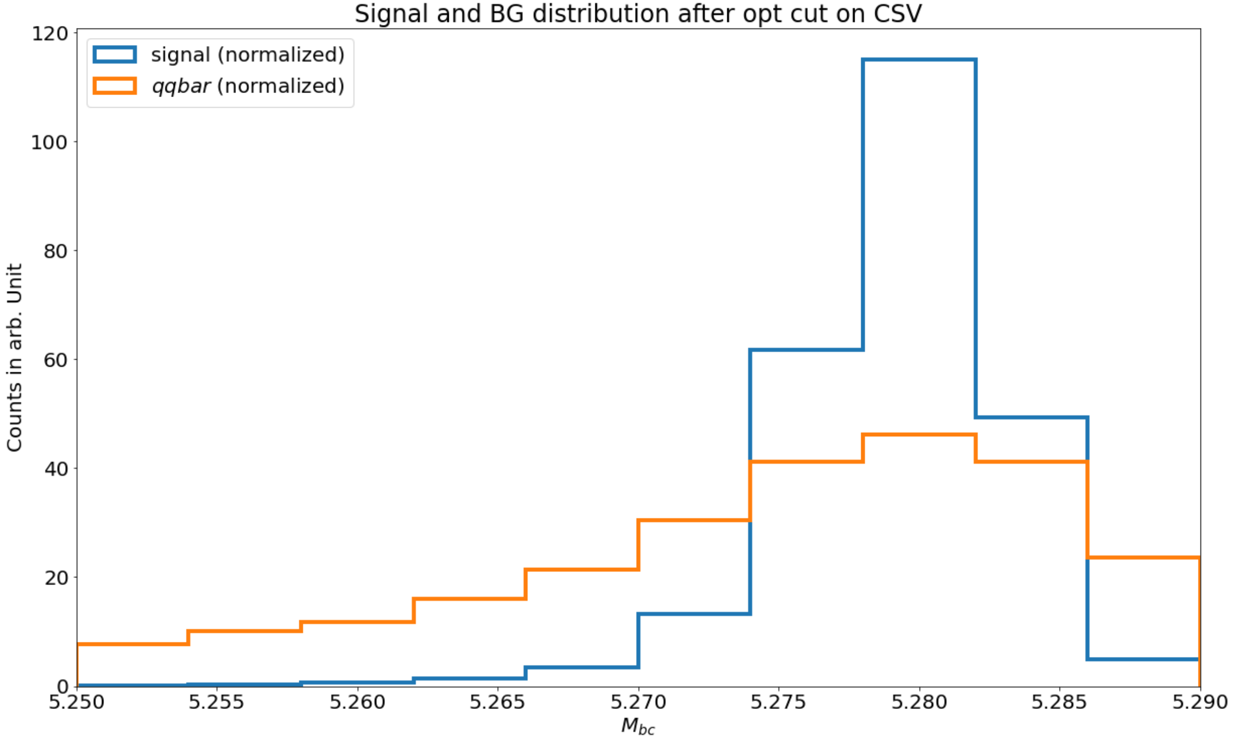
Performance



Challenges

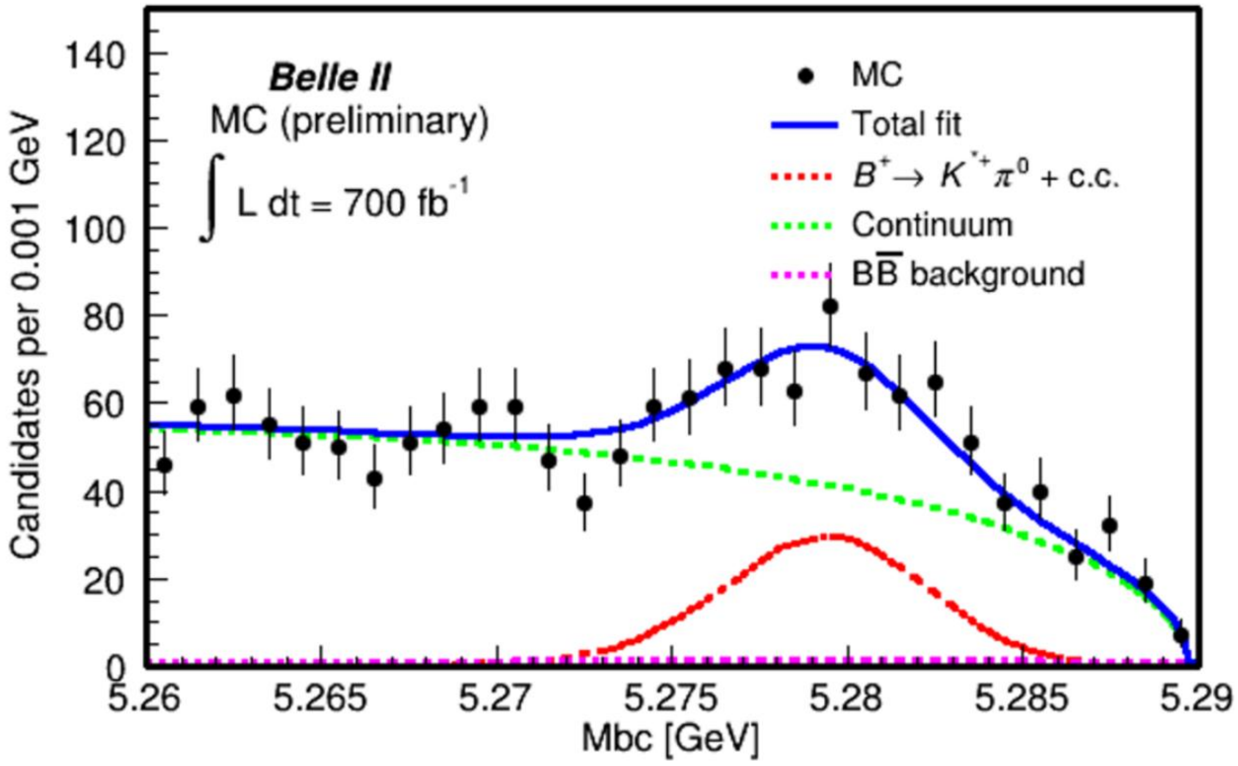
$$M_{bc} = \sqrt{E_{beam}^2/c^4 - p_B^2/c^2}$$



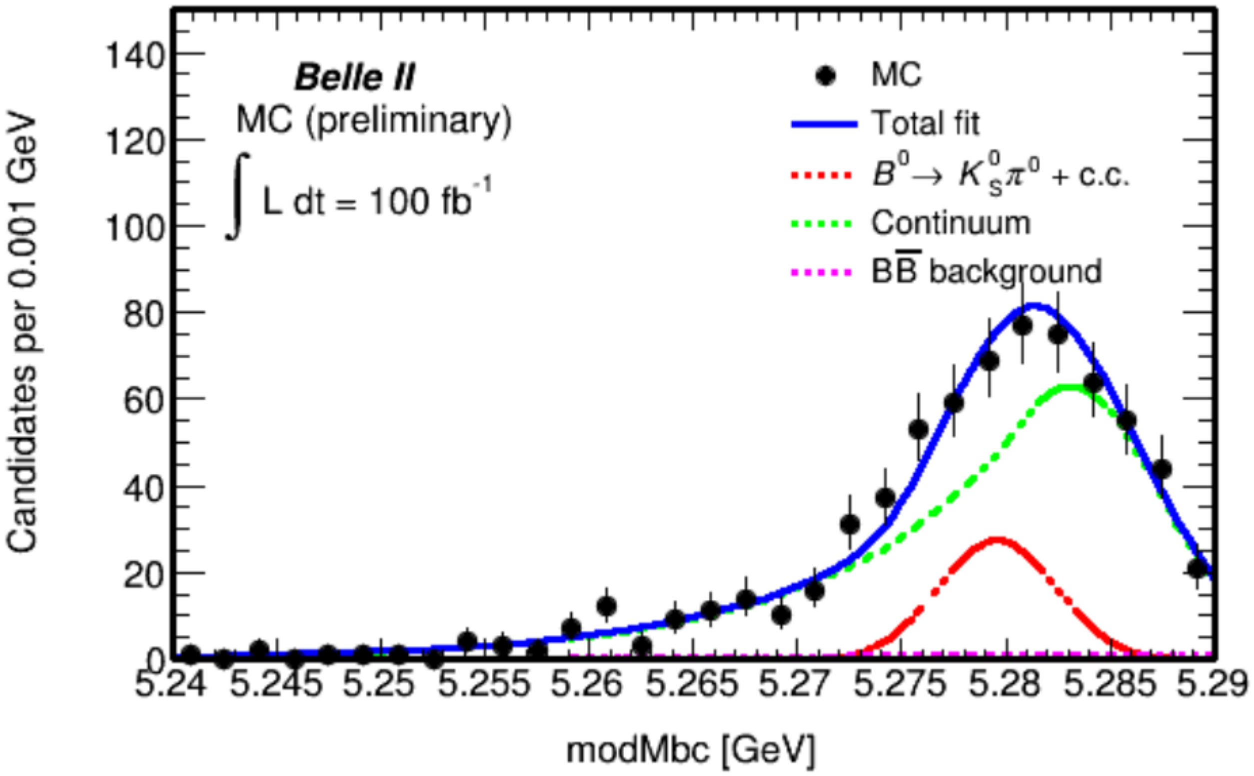


- As expected, the CSV cut shapes the $q\bar{q}$ events in M_{bc} in a signal like fashion
- Taking out individual variables showed no effect

Example



Belle II
MC (preliminary)
∫ L dt = 700 fb-1
MC
Total fit
B+→ K*+π0 + c.c.
Continuum
BB̄ background
Candidates per 0.001 GeV
Mbc [GeV]

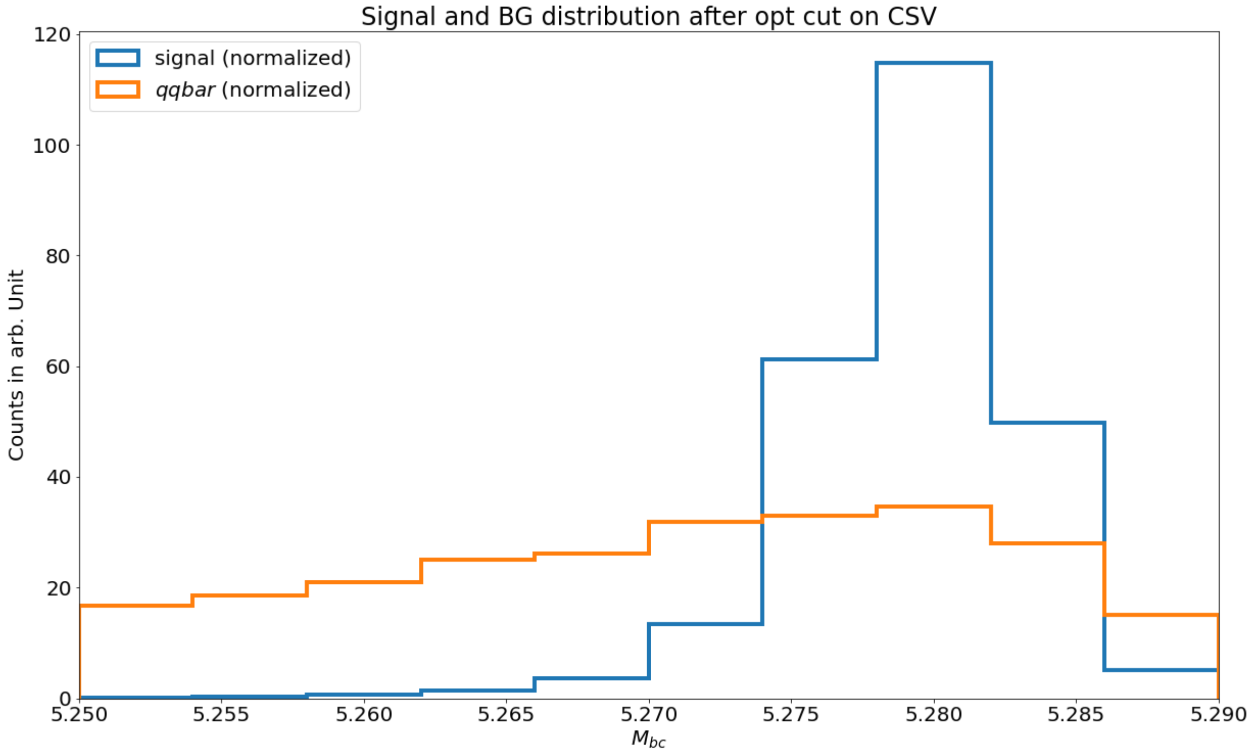


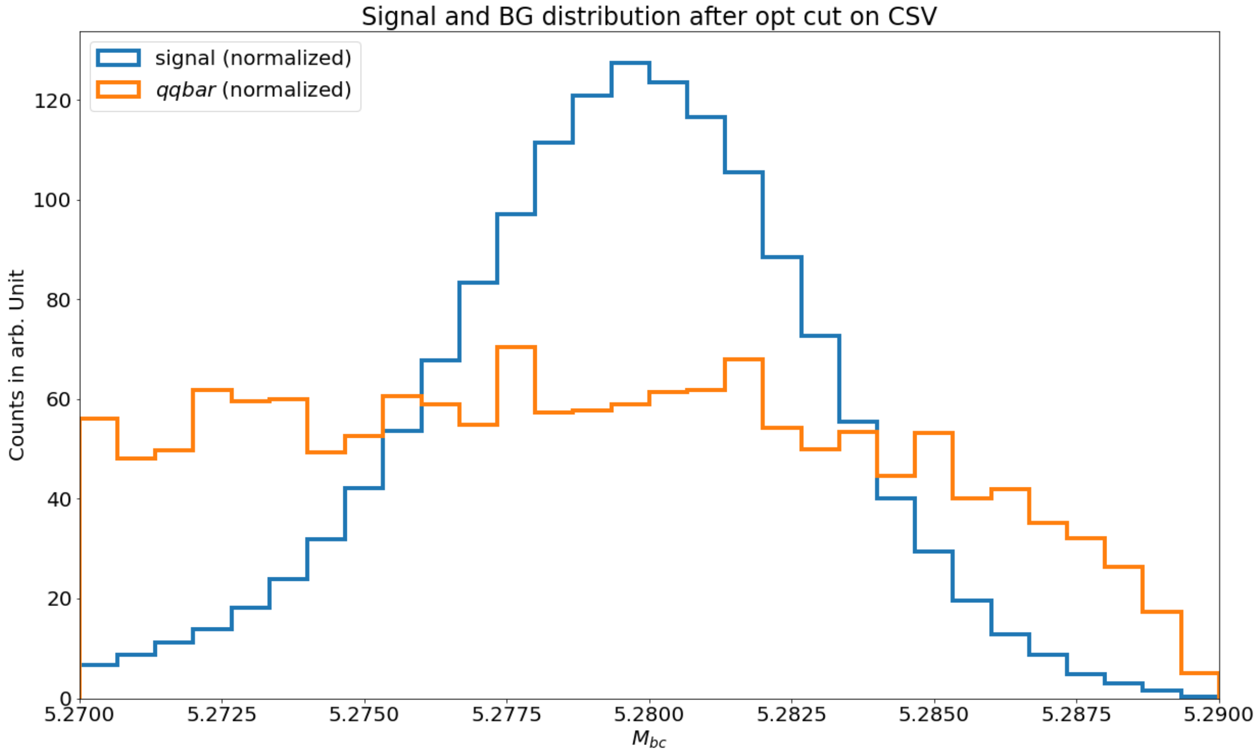
Belle II
MC (preliminary)
∫ L dt = 100 fb-1
MC
Total fit
B0→ K0Sπ0 + c.c.
Continuum
BB̄ background
Candidates per 0.001 GeV
modMbc [GeV]

Decorrelation

- Change loss function like $\mathcal{L}' = \mathcal{L} + \lambda \cdot corr(M_{bc})$
- Optimal phase space pockets are chosen automatically, since decorrelation and classification performance are simultaneously trained
- λ can be used to scale the correlation contribution to the loss function. Increasing the decorrelation will lead to a worse suppression of continuum
- Decorrelate CS classifier until the continuum background does not peak around the signal peak any more or has a much broader peaking structure

First Example





AUC Score: 0.9904

$$corr_{q\bar{q}}(M_{bc}, CSV) = 20\%$$
$$corr_{sig}(M_{bc}, CSV) = 5.4\%$$

MAX-PLANCK-INSTITUT
FÜR PHYSIK

Summary

- Continuum suppression classifiers on the base of machine learning techniques are very successful
- Expanding the input variable space to low level variables has increased their performance substantially
- However, with higher classification power comes higher correlation which makes the fitting more difficult
- The bigger variable space calls for new methods for the decorrelation

→The challenge is to find the sweet spot between high performance and low correlation for the best statistical significance after fitting