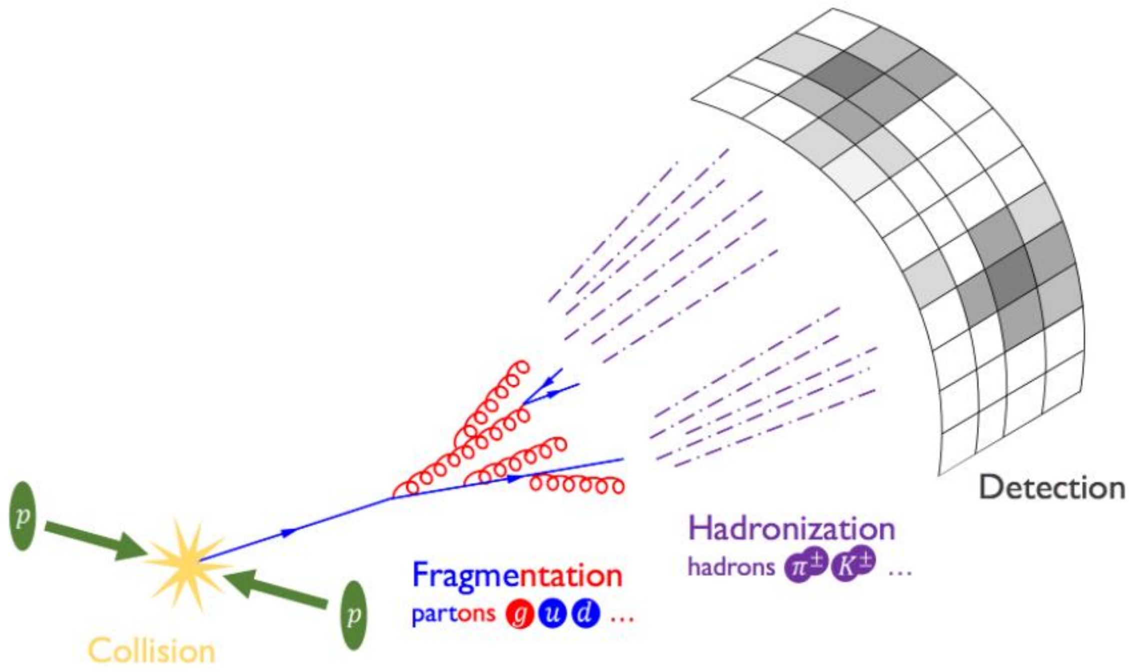
Matching VBF-quarks with jets

Santeri Salomaa
CMS Summer Student 2021
Supervisor: Santeri Laurila

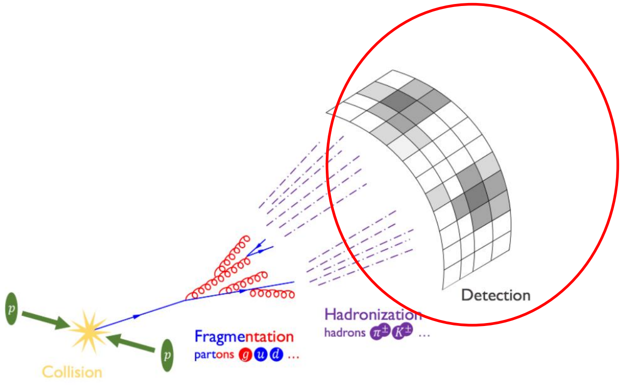


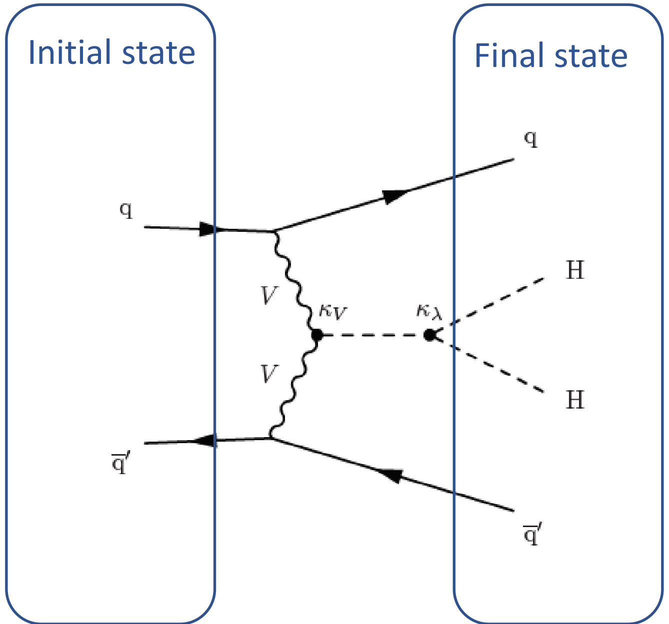
Introduction

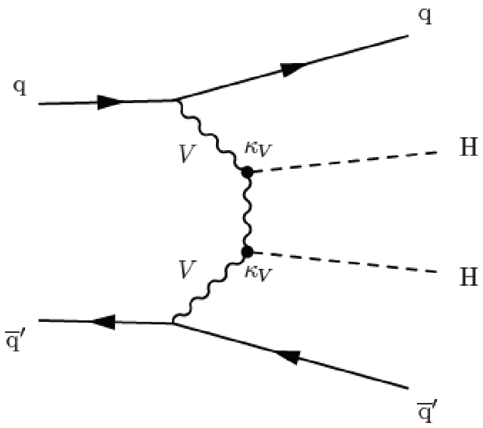
Higgs boson pair production in vector boson fusion

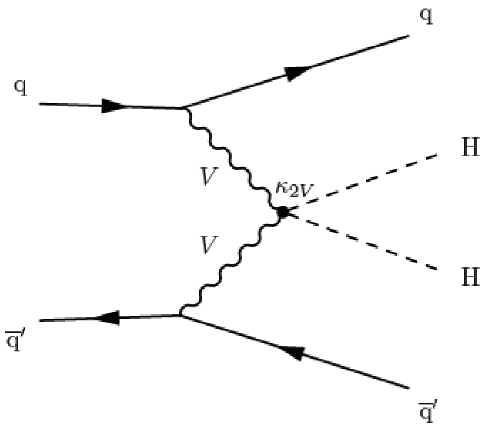
- Two quarks interact with vector bosons (weak interaction) and produce a Higgs boson pair (so called VBF-HH-event).
- Feynman diagrams below represent the interactions of the lowest complexity (and therefore highest probability).
- I have studied these interactions with computer simulations in “generator level” (in real experiments only information is from the detectors).

Detector data



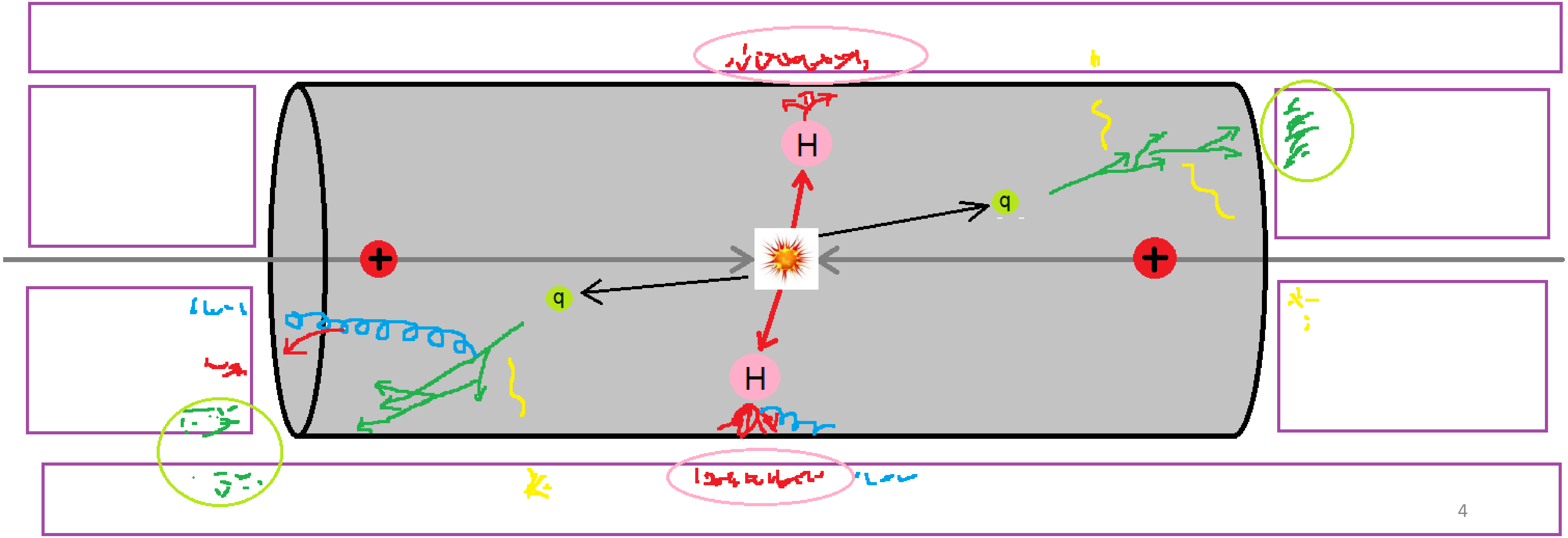






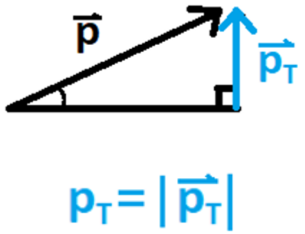
The interactions in the detector

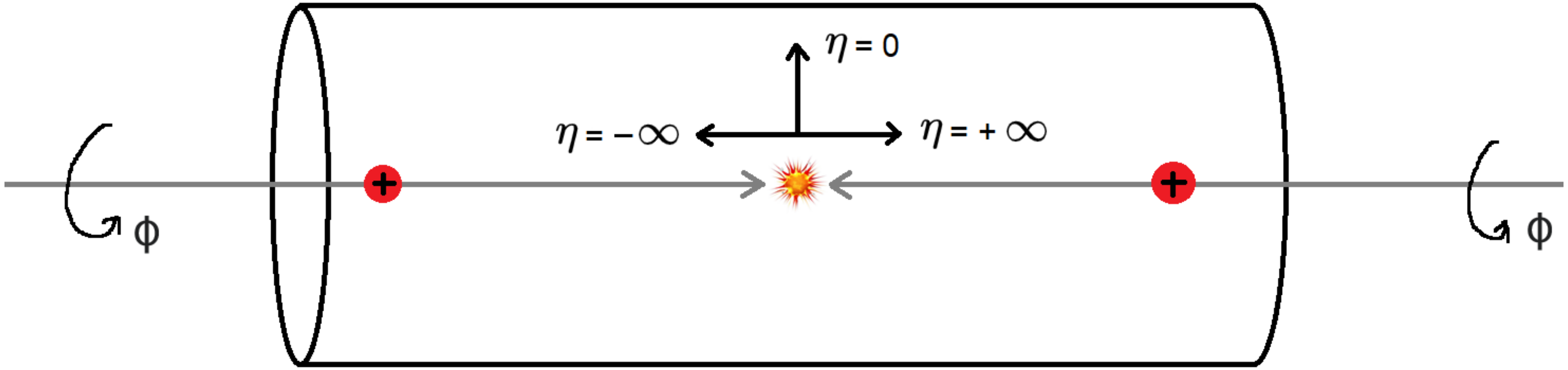
- In the simulation, always 2 higgs bosons and 2 quarks are produced with VBF-interaction.
  - Higgs bosons decay quickly into other particles and produce large jets near 90° angle.
  - The quarks also produce jets (decay and hadronisation) but in smaller degree angles.



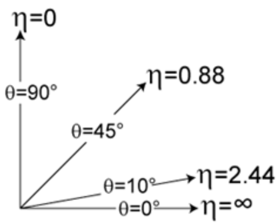
What is the coordinate system?

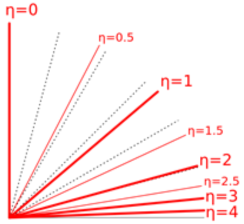
- (η , Φ)
  - Angle η is defined below.
  - The distance (radius) from the origin is uninteresting.
- Other interesting variable is the transverse momentum pT.
  - Higher | η | => lower pT .





$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$$



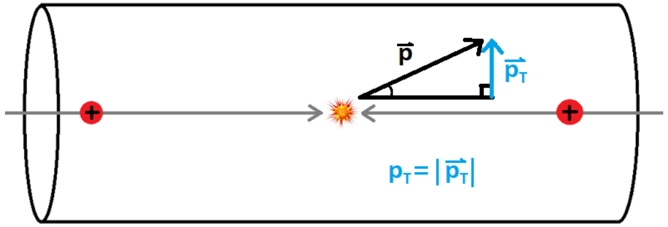


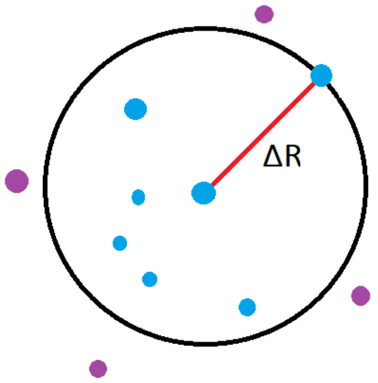
η = 0.5: θ = 62°
η = 1: θ = 40°
η = 2: θ = 15°
η = 3: θ = 5.7°
η = 4: θ = 2.1°
η = 5: θ = 0.8°

VBF-quarks

- Let’s focus on what happends to the VBF-quarks after the VBF interaction.
  - How is the matching from quarks to jets done optimally?

- Jets are clustered from the particles with ***anti-kT algorithm*** which produce circular clusters near **big-pT-particles**.
- The VBF quark jets are usually smaller (**AK4**) than the Higgs boson jets (AK8).



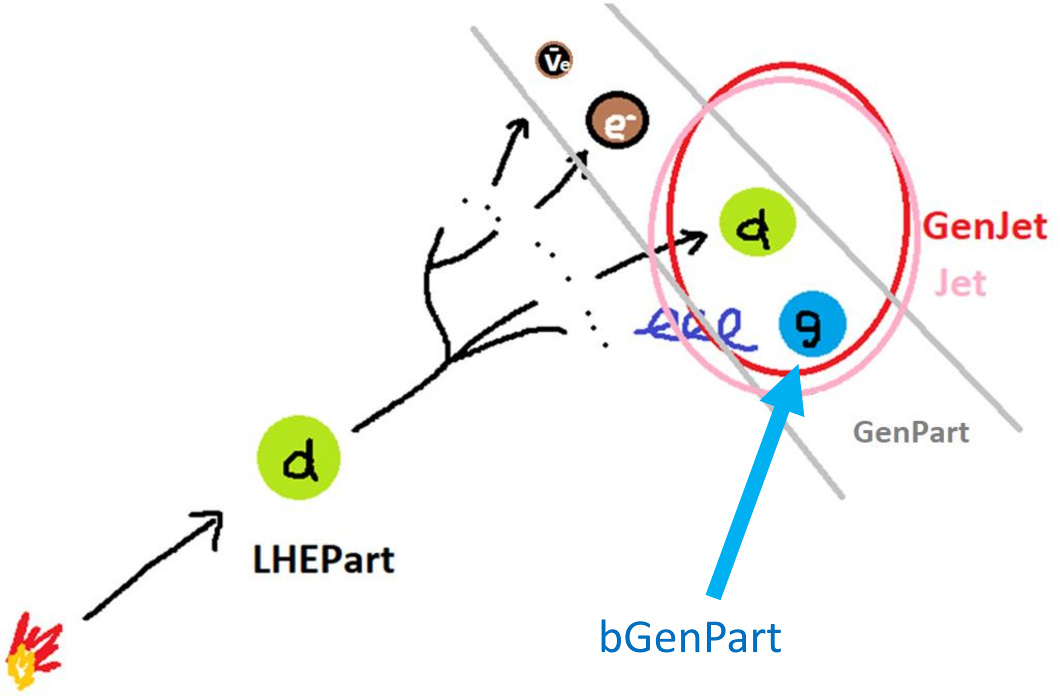


$$\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

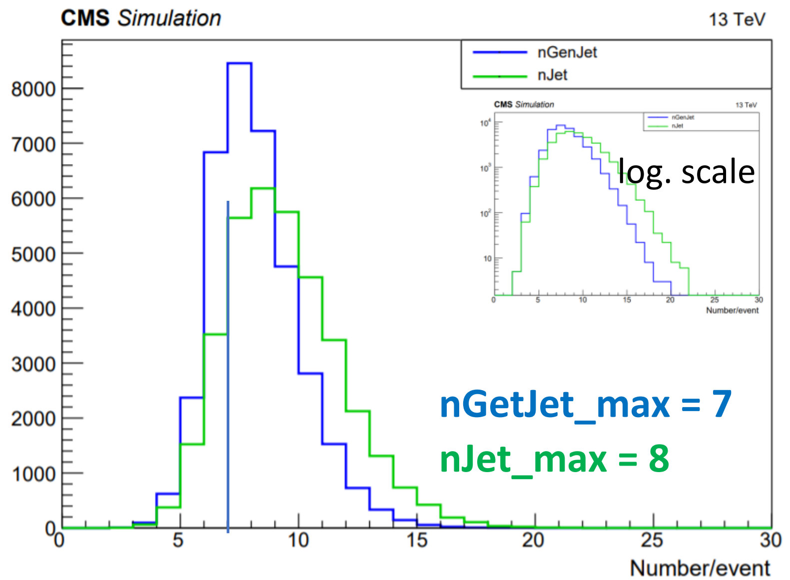
AK4 jet
= jet with radius 0.4

GenJets and Jets

- In the simulation, GenJets are jets that are clustered from the particles just before hitting the detectors.
- Jets are clustered from the detector data. In ideal case detector detects all the detectable particles, and the jets are very close to GenJets.

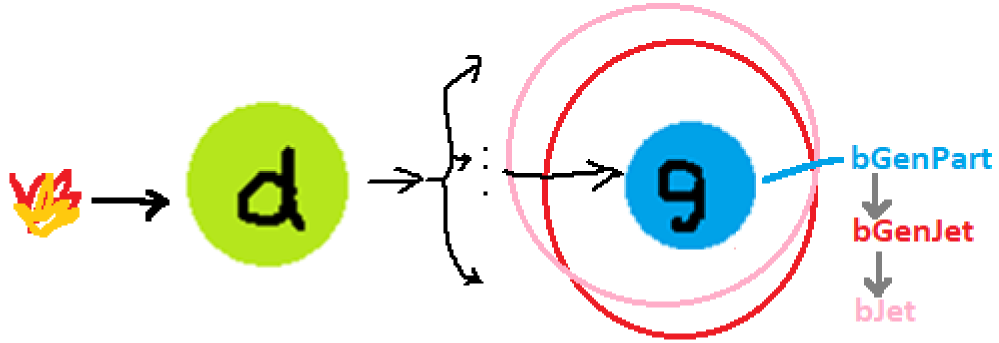


- What is a good way to choose a GenJet corresponding to the VBF-quark?
  1. Choose the biggest-pT (detectable) GenPart child (= **bGenPart**) and find it's nearest GenJet.
  2. If the GenJet is close enough (ΔR < 0.4), declare the matching successful (called **weak match**).
  3. If it's not successful. Choose the next biggest-pT child, until the matching is successful or the GenPart child -list ends.
- If there is also a Jet in ΔR < 0.4 radius from the GenJet, the match is called **strong match.**



Explanation of the variables

- bGenPart = best GenPart child
- bGenJet = closest GenJet for the bGenPart
- bJet = closest Jet for the bGenJet



- Weak match = ($\Delta R < 0.4$ match for bGenPart → bGenJet)
- Strong match = ($\Delta R < 0.4$ match for bGenJet → bJet)

```
************************************************************************************
Event 45 with 2 weak quark(s):
------------------------------------------
Final state quarks:
['u', 'd']
------------------------------------------
Decay conclusion:
u ---> ['u', 'anti_d', 'd'] ---> d
d ---> ['e-', 'anti_e-', 'g'] ---> g
- - - - - - - - - - - - - - - - - - - - -
with coordinates (eta, phi, pT):
(3.95, -2.74, 88) ---> [(3.77, -2.55, 10), (4.09, -2.68, 17), (4.03, -2.63, 35)] ---> (4.03, -2.63, 35) ---> (4.04, -2.72, 82)
(-4.05, 0.2, 30) ---> [(-0.46, -1.28, 0), (-0.94, -1.76, 0), (-3.77, 0.44, 6)] ---> (-3.77, 0.44, 6) ---> (-3.87, 0.4, 35)
- - - - - - - - - - - - - - - - - - - - -
All GenJet coordinates:
[(-4.99, -0.9, 17), (-3.87, 0.4, 35), (-3.61, 1.82, 15), (-1.26, -2.83, 631), (-0.26, 2.94, 19), (1.07, 0.35, 749), (4.04, -2.72, 82)]
------------------------------------------
************************************************************************************
```

bGenPart:

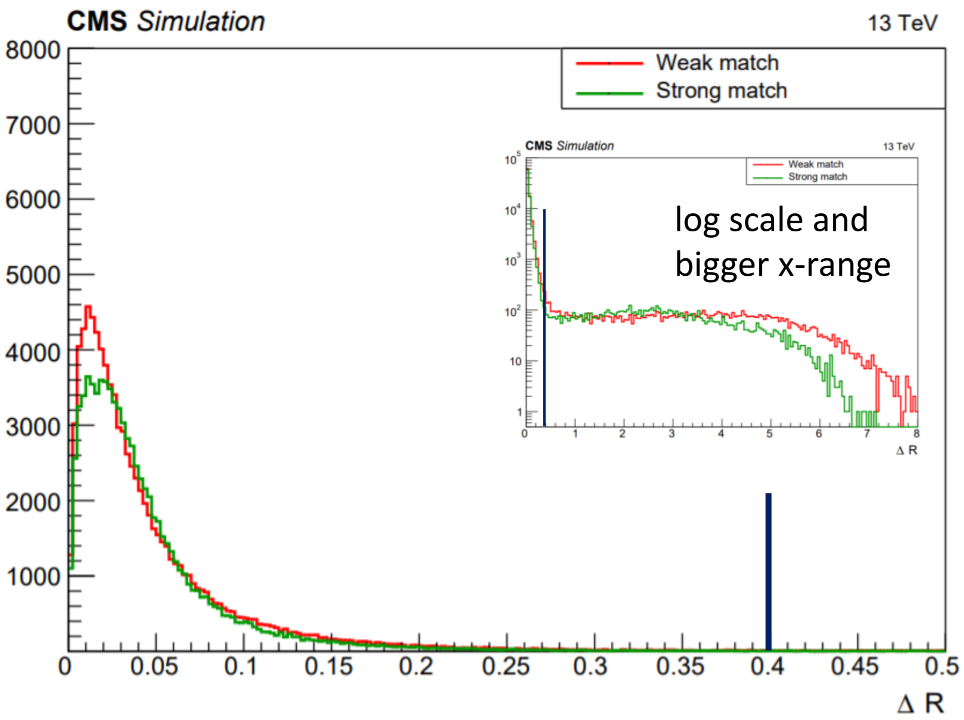
bGenJet:

Results of the matchings

Success rates for the weak and strong match

- Weak match success rate: 0.911
- Strong match success rate: 0.836

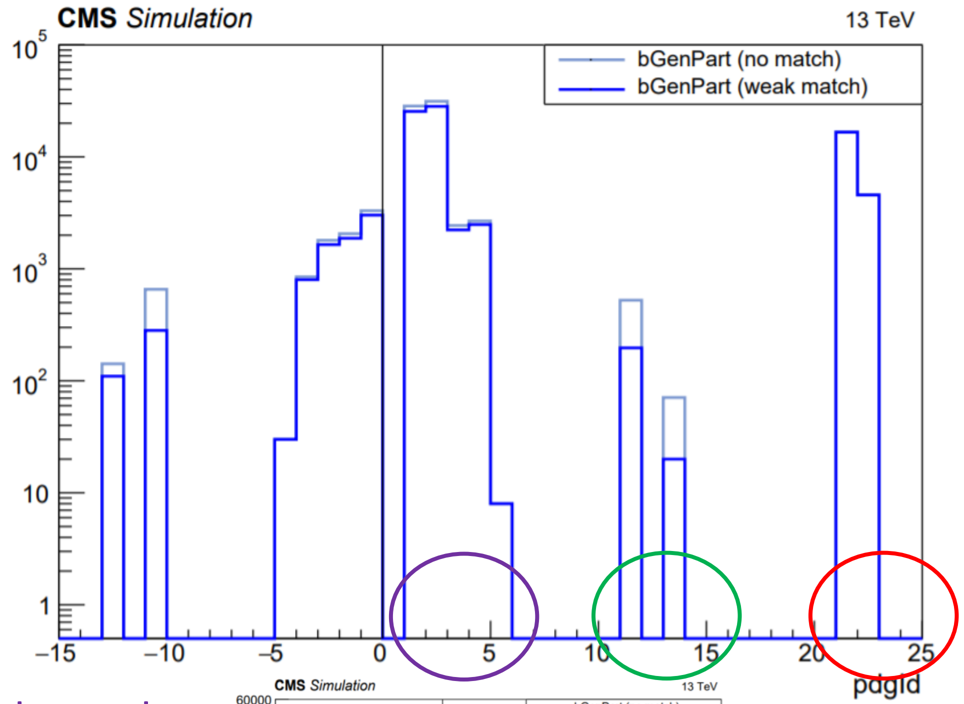
- Let's find out what are the problems in matchings.



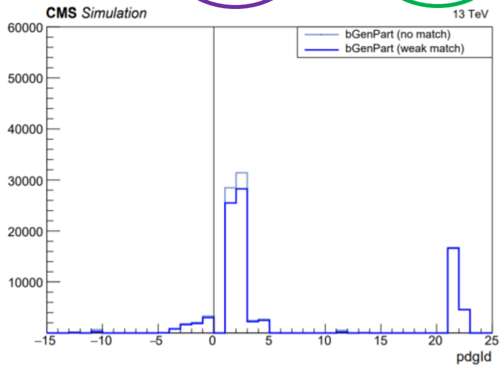

```
--------------------------------------------------------------
Statistics of 49031 events with deltaRmax = 0.4:
--------------------------------------------------------------
# particles/event:

= 2:  1.0
  1:  0.0
  0:  0.0

# all particles
= 98062
--------------------------------------------------------------
in weak match
= 2:  0.83
  1:  0.161
  0:  0.009

# weak matched particles / # all particles
= 89299/98062 = 0.911
--------------------------------------------------------------
and strong match
= 2:  0.701
  1:  0.27
  0:  0.029

# strong matched particles / # weak matched particles
= 82002/89299 = 0.918
# strong matched particles / # all particles
= 82002/98062 = 0.836
--------------------------------------------------------------
```

What are the bGenPart particles?

- Just after the weak interaction, the quarks are either *down, up, strange* or *charm* quarks.
- After decays and hadronisation, the distribution of particle types is larger.
- bGenPart pdgId distributions are shown in the histograms on the right side.
    - There are also mesons and baryons that are not showing in the histograms.



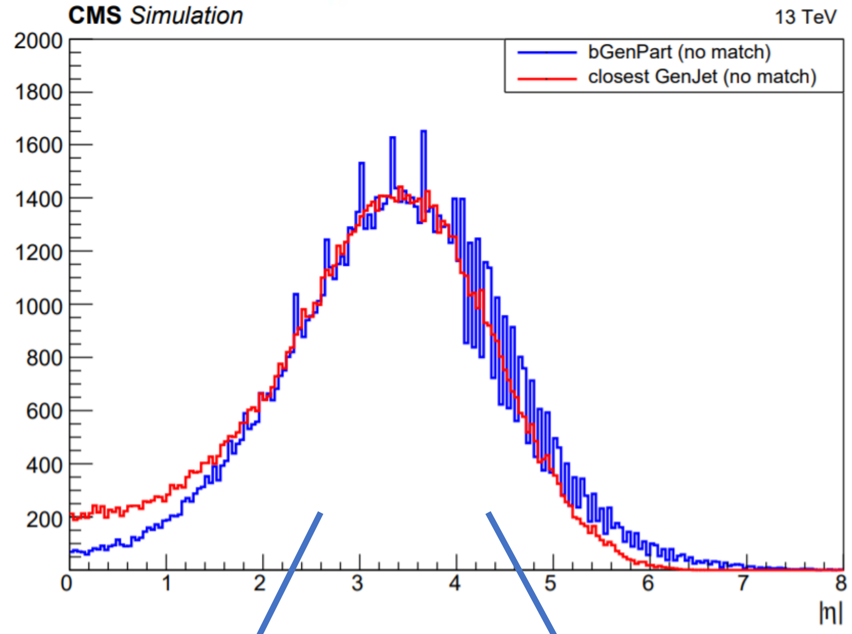
d, u, s, c, b quarks,
electrons, muons
gluons, photons,
mesons or baryons

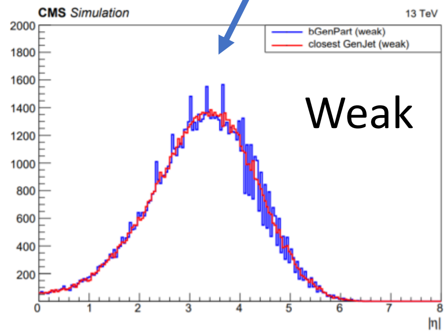


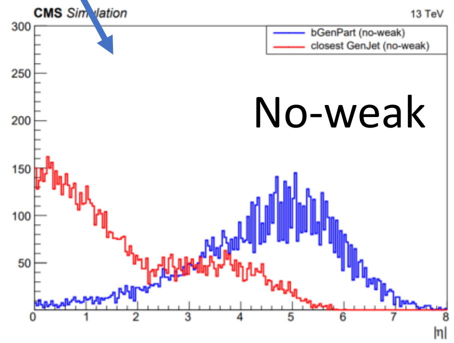
η = 0.5: ϴ = 62°
η = 1: ϴ = 40°
η = 2: ϴ = 15°
η = 3: ϴ = 5.7°
η = 4: ϴ = 2.1°
η = 5: ϴ = 0.8°

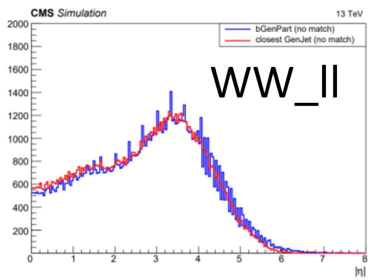
|η| in the weak match

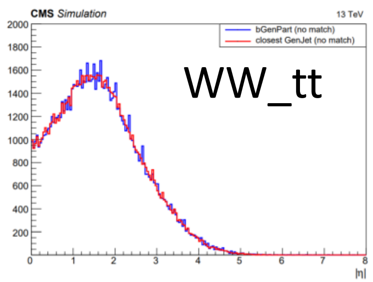
- The distribution of **η is very high valued** (small ϴ degrees from the collision pipe).
- Compare these with WW data at the bottom, where two W bosons are produced instead of higgs.
  - The difference between these is the polarisation (longitude and transverse polarisations).
- **The no-weak match cases are mostly with high η bGenParts.**
  - High η → low pT → no GenJet
- The wrong GenJet is usually a jet from Higgs decay and therefore near zero η.







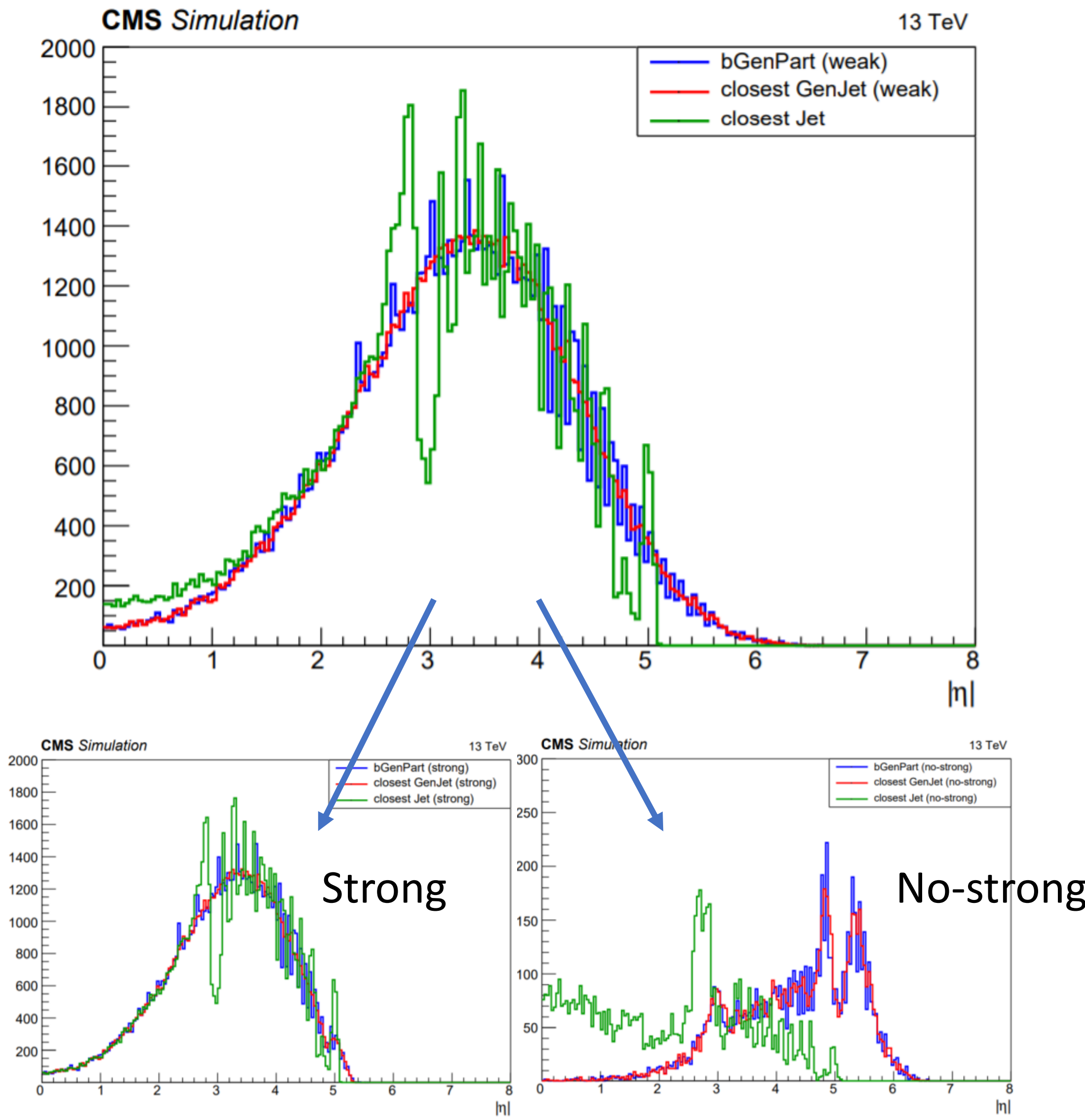




|η| in the strong match

- There is a blind spot at η = +/- 3 (Θ = 5.7°).
  - Therefore, there are also higher bars around it.

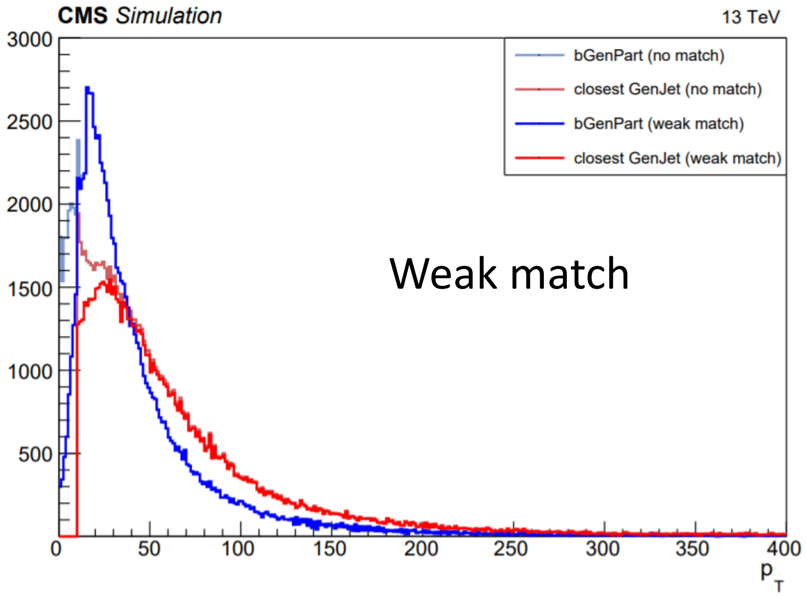
- The detector ends around η = +/- 5.1 (Θ = 0.7°).
  - The bGenParts with bigger η are mostly no-strong cases.

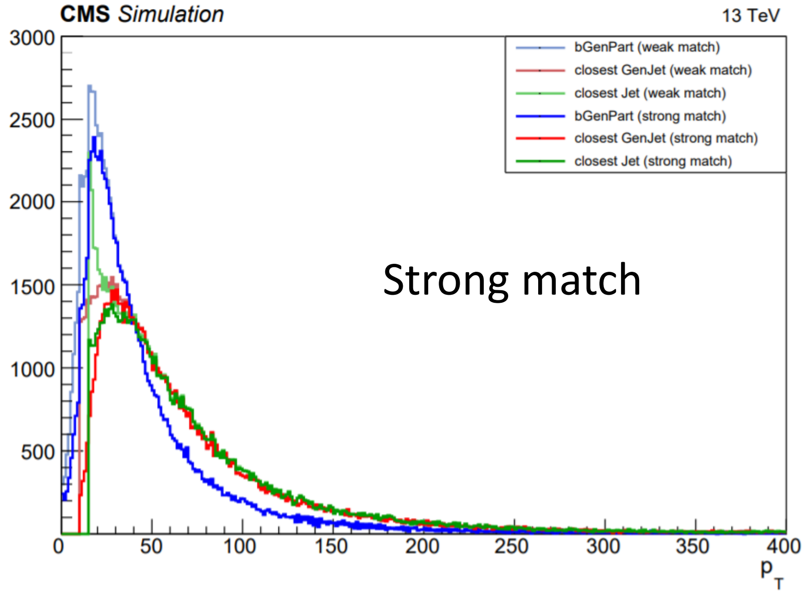


pT in small (linear) scale

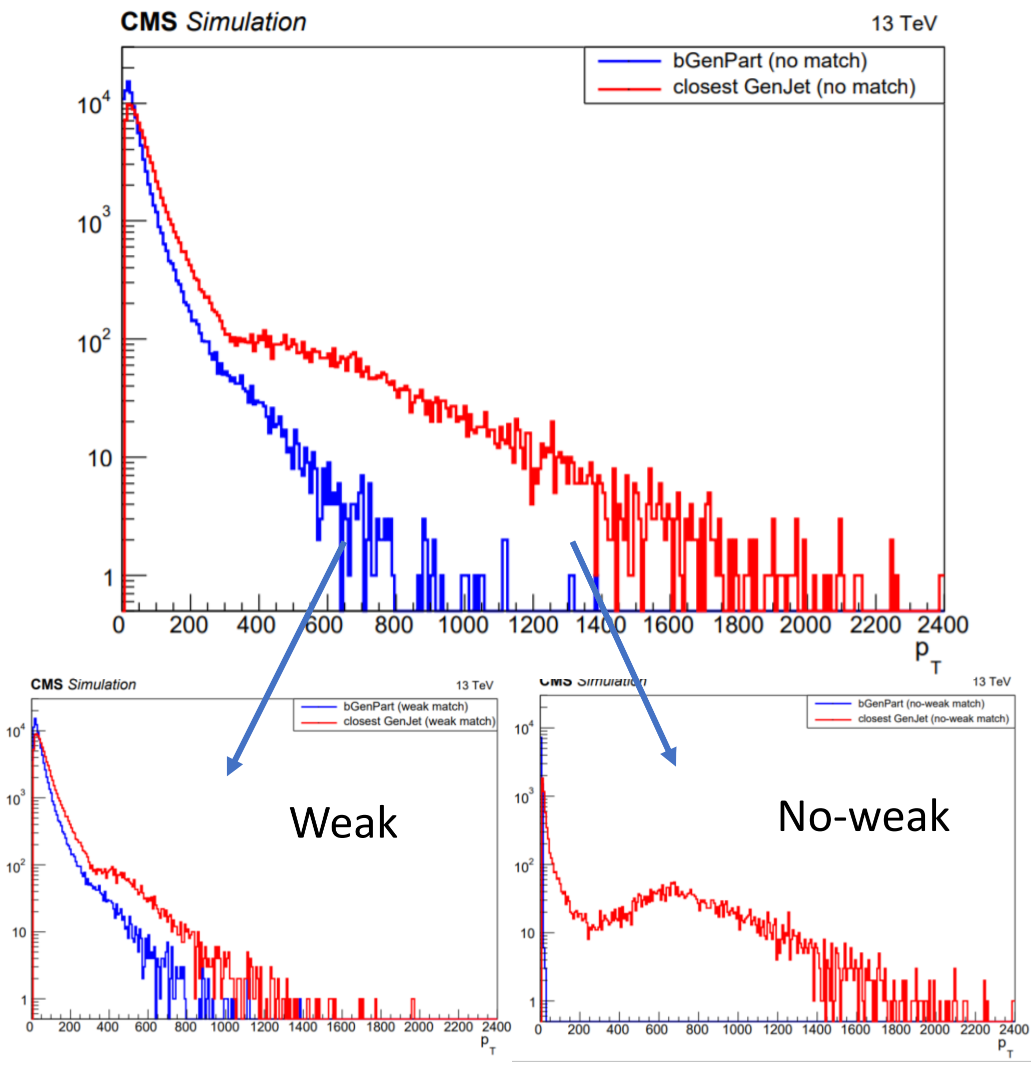
- pT distribution is focused on small pT values (0-100 GeV)
  - Notice that bGenPart gets a lot of no-match cases between 0 - 10 GeV.
  - Minimum pT for GenJet is 10 GeV.
  - Minimum pT for Jet is 15 GeV.

- Notice that GenJet and Jet pTs get higher values since it consists of many particles

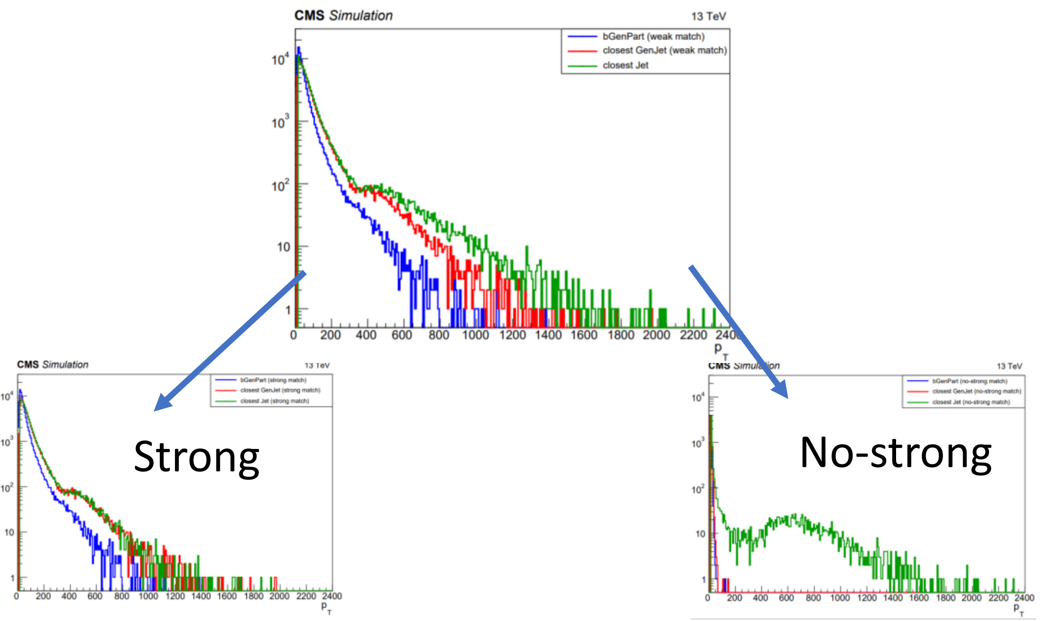




Big pT in logarithmic scale



- There are also higher pT cases.
- **In no-weak cases, the incorrect small-pT-bGenPart to higgs-GenJet match is shown**.
  - Similarly no-strong matches are incorrect matches between small-pT-GenJets and higgs-Jets



Successful matching rates using filters

- The matching success rates improve a lot with these filters
  - Without filters weak rate = 0.911 and strong rate = 0.836.

Detectable η

```
----------------------------------------------------------------
Statistics of 49031 events with deltaRmax = 0.4:
----------------------------------------------------------------
# particles/event with filter: detectable eta
(|bGenPart_eta| < 5.1 && |bGenJet_eta| < 5.1)

= 2:  0.885
  1:  0.111
  0:  0.004

# all filtered particles / # all particles
= 92222/98062 = 0.94
----------------------------------------------------------------
in weak match
= 2:  0.781
  1:  0.204
  0:  0.015

# weak matched particles / # all (filtered) particles
= 86620/92222 = 0.939
----------------------------------------------------------------
and strong match
= 2:  0.691
  1:  0.278
  0:  0.031

# strong matched particles / # weak matched particles
= 81386/86620 = 0.94
# strong matched particles / # all (filtered) particles
= 81386/92222 = 0.883
----------------------------------------------------------------
```

bGenPart_pT > 10 GeV

```
----------------------------------------------------------------
Statistics of 49031 events with deltaRmax = 0.4:
----------------------------------------------------------------
# particles/event with filter: detectable GenJets
(bGenPart_pT > 10)

= 2:  0.721
  1:  0.255
  0:  0.024

# all filtered particles / # all particles
= 83196/98062 = 0.848
----------------------------------------------------------------
in weak match
= 2:  0.715
  1:  0.259
  0:  0.025

# weak matched particles / # all (filtered) particles
= 82864/83196 = 0.996
----------------------------------------------------------------
and strong match
= 2:  0.641
  1:  0.317
  0:  0.043

# strong matched particles / # weak matched particles
= 78371/82864 = 0.946
# strong matched particles / # all (filtered) particles
= 78371/83196 = 0.942
----------------------------------------------------------------
```

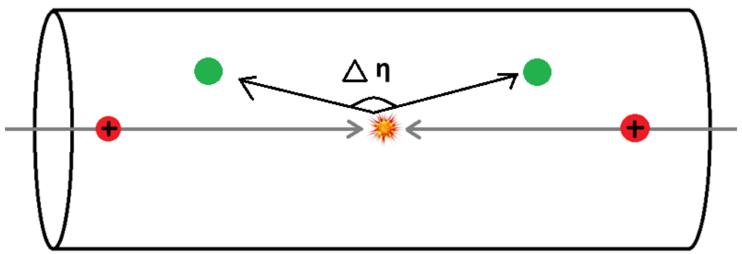
Both + bGenJet_pT > 15 GeV

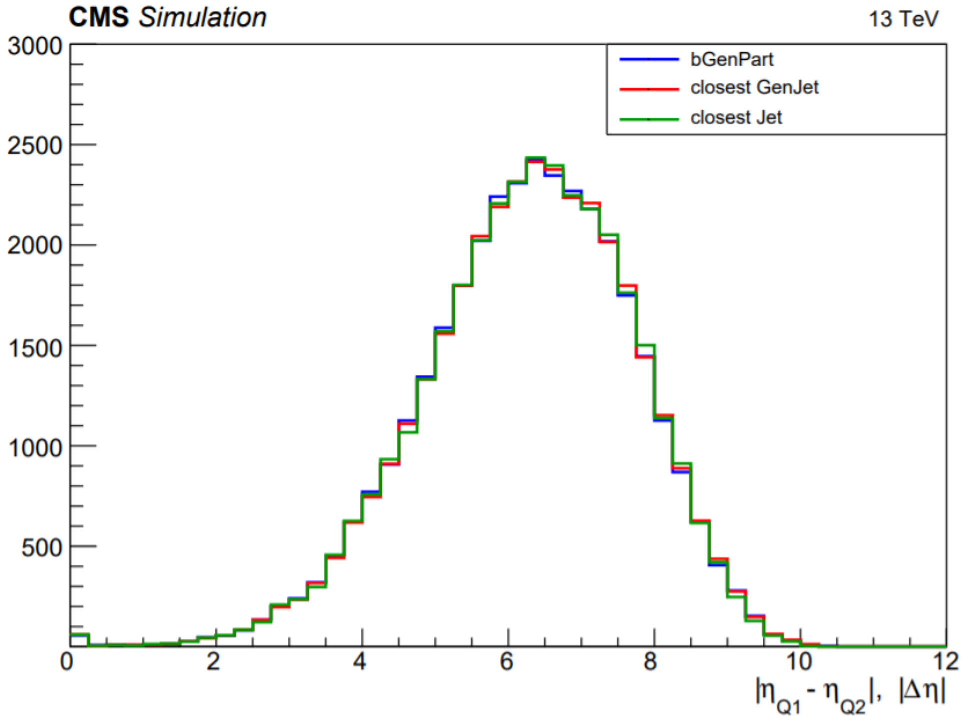
```
----------------------------------------------------------------
Statistics of 49031 events with deltaRmax = 0.4:
----------------------------------------------------------------
# particles/event with filter: detectable GenJets and Jets
(|bGenPart_eta| < 5.1 && |bGenJet_eta| < 5.1
&& bGenPart_pT > 10 && bGenJet_pT > 15)
= 2:  0.653
  1:  0.307
  0:  0.04

# all filtered particles / # all particles
= 79069/98062 = 0.806
----------------------------------------------------------------
in weak match
= 2:  0.65
  1:  0.309
  0:  0.041

# weak matched particles / # all (filtered) particles
= 78881/79069 = 0.998
----------------------------------------------------------------
and strong match
= 2:  0.62
  1:  0.332
  0:  0.048

# strong matched particles / # weak matched particles
= 77059/78881 = 0.977
# strong matched particles / # all (filtered) particles
= 77059/79069 = 0.975
----------------------------------------------------------------
```

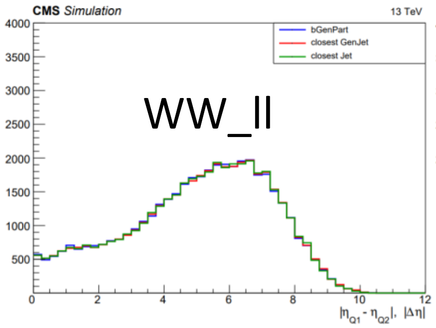
VBF quark vs. quark comparison inside the events

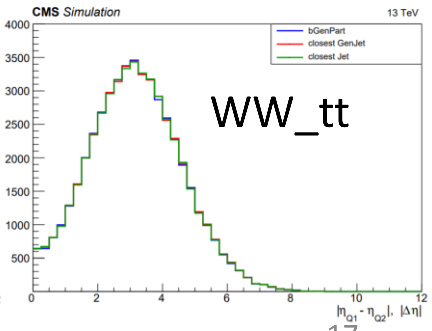
- I represent only one histogram here since there have been too much stuff already.

- **The difference between η angle is high**. This is due to the eta distribution (high and opposite sign etas).
- This pattern is important because it can be used to differenciate the di-higgs cases from other events.
  - Compare the distributions with WW events.







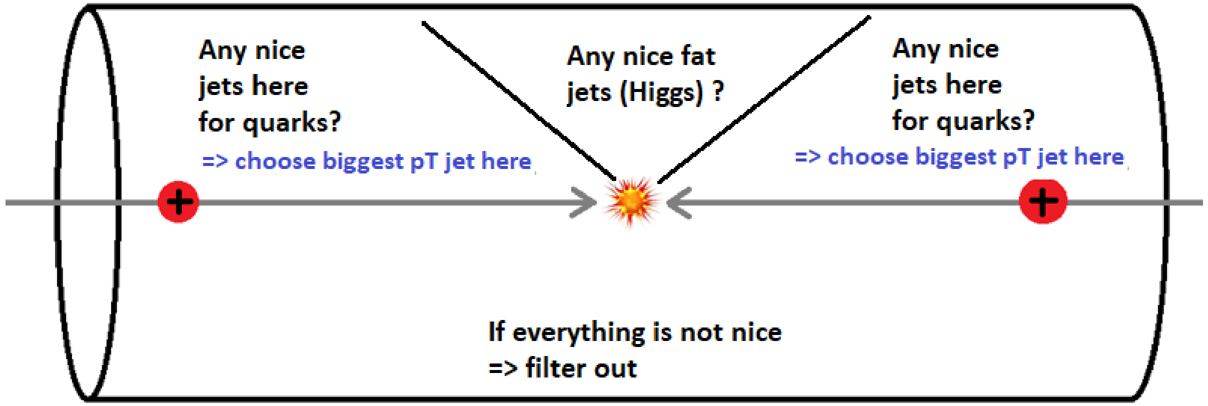


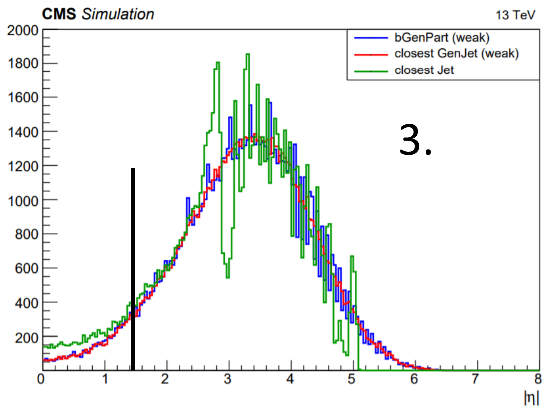
Filtering the signal events and finding the VBF quark jets

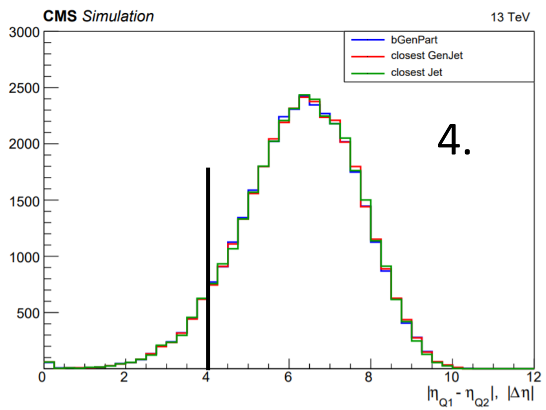
In the detector level

How filtering signal and finding VBF quark jets are done?

1. **Filter the collision events to find VBF-HH events** (this leaves only 1.22 % of the signal events). Some VBF quark jet filters that leaves only nice VBFCandidates:
   1. Only events without leptons (electrons and muons) near VBF jets.
   2. Difference in η is over 1.2 with Higgs candidate jet (1.2 = 0.8+0.4, and this guarantees **separate VBF quark jets from the higgs boson fat jets**).
   3. VBF jet **|η| > 1.5** (VBF jet that deviates max 25°).
   4. Difference in η is over 4 between the two VBF jets (**VBF jets are far away from one another**).
   5. VBF jet invariant mass > 500 GeV
2. After this, the two biggest pT jets are chosen from the VBFCandidate list.







How successful is this method?

- The VBFJetCandidate is defined correct if it is a strong matched jet.

- Actually, the results are surprisingly good:

1. VBFJetCandidate[0] (biggest pT jet) is one of the strong matched jets = 98.0 %
2. VBFJetCandidate[1] (2. biggest pT jet) is one of the strong matched jets = 89.5 %
3. #(correct VBFJetCandidates)/#(all VBFJetCandidates) = 93.7 %
4. Both VBFJetCandidates are the strong matched jets = 87.7 %

- Downsides of this method: very strong filter for the signal, and only very specified events.
- Also, notice that in 12,3 % of the events at least one VBFCandidate is wrong.

Conclusion

- Optimal way to match VBF-quarks with jets (with anti-kT algorithm) in the generator/simulation level is found.
  - Strong matched case = near the quark final childs there is a GenJet and a Jet.
  - Better matching success with larger pT and smaller η.

- The previous method of finding the quark jets in the data is pretty successful, but very simple and aggressive with the filters.
- What's next?
  - **Comparisons** with the filtered data and **improvements** to the matching methods.
  - Moving from the **"generator level" to "detector level"** to filter the VBF-HH events and find the VBF-quark-jets (machine learning algorithms applied?).