Detector Traveling from Now to the Future: Study of the Calorimeter Performance for Future Colliders with the High Granularity Calorimeter in CMS and the Full Simulation based on FCC

Chih-Hsiang Yeh

Department of Physics and Center for High Energy and High Field Physics, National Central University, Chung-Li, Taiwan

Abstract

The discovery of the standard-model-like Higgs boson at Large Hadron Collider(LHC) in 2012 opened a new door for particle physics. Since then, people have been hard-working on upgrading the LHC to the higher luminosity and center-of-mass (C.M.) energy, so to study the properties of Higgs boson and to discover new particles predicted by the physics Beyond Standard Model(BSM). Some people even proposed the idea to build new colliders, such as Future Circular Collider(FCC), etc. near the working LHC space now. For my study, I focus on three topics which are related to the detector performance. First, because the radiation tolerance problem and increasing pileup at High-Luminosity LHC(HL-LHC), the Compact Muon Solenoid(CMS) experiment at LHC will replace the current endcap calorimeter with a silicon-pad High Granularity Calorimeter(HGCAL). I did some studies on the performance for them using data from test beam in CMS and the Hexaboard of its electronics in National Taiwan University(NTU). Second, I used the GEANT4 full simulation to simulate the FCC detector under the very high energy condition and study the jet performance. We expect to help the future facilities to design a high resolution detector, and could have a large potential to find the new physics with the new detector.

Contents

1 Introduction

1.1 Motivation and Expectation

After identifying the Higgs boson at LHC in 2012, people are eager to explore the new particle and the physics BSM , such as dark matter candidates, Z' bosons, heavy Higgs, and so on. CMS and ATLAS collaboration are also upgrading the detectors to improve the functions to cope with the new challenging conditions in the future colliders. For the operation in the future, one of the most important issues of detector performance is: How can we boost the efficiency of the distinguishability of signal from the background?

In the upcoming HL-LHC era, the pileup will be raised significantly compared with the condition the LHC is operating at currently. Therefore, first of all, we need to identify the particle very well. Otherwise, we are not able to reject the unwanted contaminant particles. My first topic with HGCAL in CMS focuses on the particle identification with electrons and pions. Because we will keep searching on more new physics with the electron final states , but some pions will be misidentified as the electrons, our requirements are removing these mistagged "fake" electrons. It is expected that we can find out the optimized cuts for them and apply in the cut-based analysis, aiding the signal and background studies that are very sensitive to pions. Furthermore, I studied some cross-talk phenomena on the PCB boards which are attached with the sensor used in the HGCAL, and try to find out the correlation between the injection pulse strength and cross-talk. We expect to understand the noise created in the electronics and reduce them in the future.

Some have proposed colliders with the higher C.M. energy for the next generation, such as Future Circular Collider (FCC), Circular Electron-Positron Collider (CEPC), Super proton-proton Collider (SppC) etc. When those colliders operate, other crucial problems with higher C.M. energy will arise. Under these circumstances, the jets from the segmentation will be very boosted. In this case, we need to explore other ways to look into this boosted jets structure. Otherwise, we can't separate the signal from the background very well. My second topic with FCC detector focuses on the boosted conditions, using different jet substructures with various configurations of the FCC detector to see whether the smallest detector configuration will give the best separation power to distinguish signal from the background.

1.2 Summary for research topics

The research activities of Chih-Hsiang Yeh from July 1, 2018 to Febraury 28, 2019 includes the following:

- Research with electronics of HGCAL : Cross-talk studies with the no sensor based Hexaboard in NTU
- Research with HGCAL in CMS : Pion-rejection studies with a CMS HGCAL test-beam prototype Electromagnetic calorimeter(ECAL)
- Research with simulation FCC detector : Studies of granularity of a hadronic calorimeter for tens-of-TeV jets at a 100 TeV pp collider

I will describe in more detail for all of the items above.

2 Cross-talk studies with the no sensor based Hexaboard in NTU

Cross-talk is the one basic phenomenon exist in the electronics, which is a signal transmitted on one circuit or channel of a transmission system creates an unwanted effect in another circuit or channel. For example, it has two types cross-talk could happen:

- Correlated cross-talk
  The capacitors in the circuit are too closed wth each other. One capacitor is charging, and it could influence the other capacitors by the EM effect.
- Anti-correlated cross-talk
  It could happen in the circuits which have the problem in power supplication. When the process in one channel doesn't have enough energy to process, it could "borrow" the energy from other channels, and lead to the cross-talk also.

Then, both of them will lead to giving out the wrong signal from the electronics. We need to quantify this value and know more about this kind of noise, otherwise, it will be a big problem when we use this electronics to reconstruct the events from the sensor in the High Energy Physics(HEP) experiments.

I did this study with Prof.Stathes Paganis, Prof.Rong-Shayang Lu, master student Chia-Hong Chein from NTU, and Prof.Shin-Shan Eiko Yu from NCU. The contributions of Chih-Hsiang Yeh to this study includes the following:

- Study the correlation between the stability of pedestals(SCA0) and other SCA sampling.
- Study the correlation between the injection pulse strength and the cross-talk.

I will describe the detail as following.

2.1 Experiment apparatus and Event reconstruction

For the experiment, we used the electronics which are composed of three parts, including the module (version 2), RPI Hexa and Raspberry Pi 3(RPI). In the Figure1, it presents the electronics and data flow we applied. For the module, it also be called "Hexaboard", because of its shape is hexagonal-like. The main power of it is to collect analog signal and convert them into the digital format. It can be bonded to silicon sensor, such as the HGCAL, to collect real charge from particle go through the sensor and do physical analysis. It also can be given the electrical pulse manually to simulate the sensor-attached condition. For the RPI Hexa, it is the bridge between FPGAs and Raspberry Pi(RPI). When the signal come out from the four CHIPs of hexaboard, first of all, we will collect them to the "slave" FPGA on the HexaBoard. Then after finishing collecting the data, the "slave" FPGA will transfer the data to MAX10 FPGA("master" FPGA) on the RPI Hexa. For Raspberry Pi 3(RPI), acting like the small computer, can do many things communicate with the hexaboard. The main power for it is to receive

the MAX10 data and can encode to read back the data, and finish reconstructing the events.
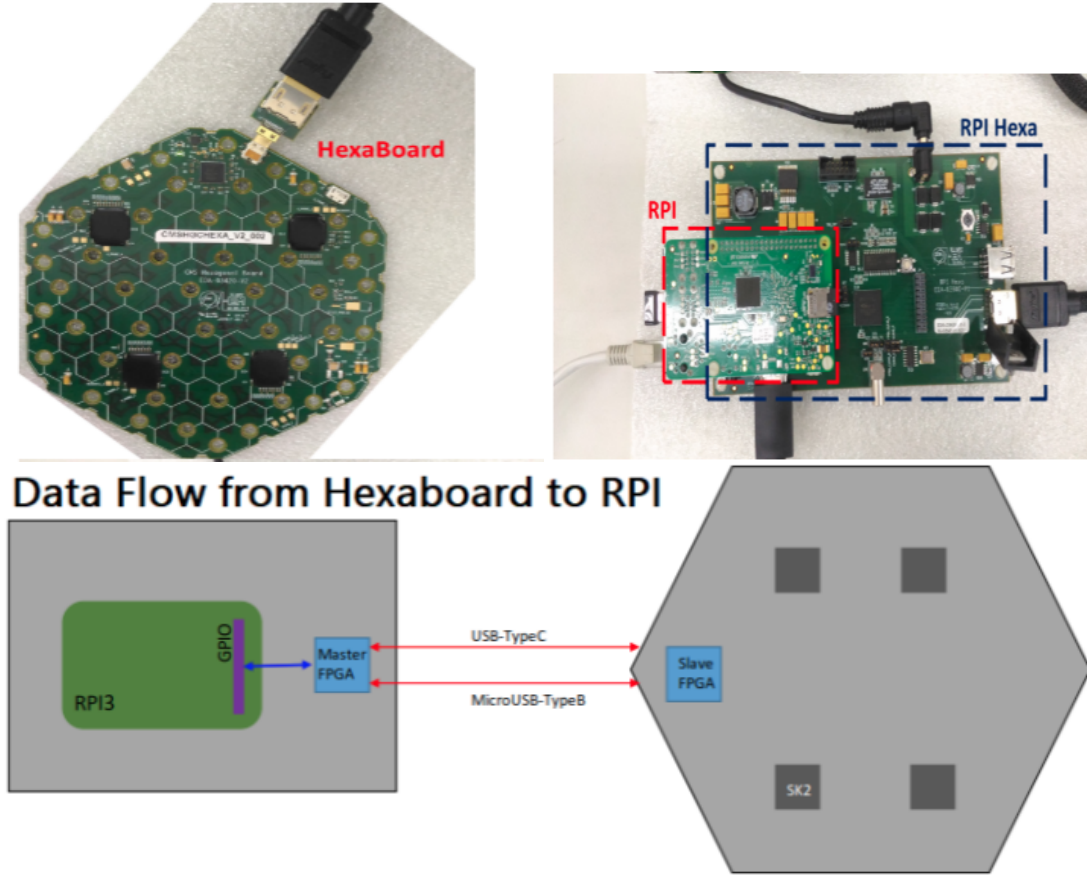

Figure 1: (Top left) The Hexaboard which was applied in the study. On the board, there are four SKIROC2cms CHIPs (black hexagonal) used to receive the electrical pulse which is given by the cable, and give out the signal to the Hexa RPI. (Top right) The Hexa RPI and RPI which is applied to transform the data and communicate with Hexaboard. (Down) The data flow pictures for introducing the data transfom from Hexaboard to Hexa RPI and RPI.

Every SKIROC2cms CHIP on the HexaBoard corresponds to 64 channels, while each channel has its own readout circuit(Pre-amplifier, shapers,etc.). In each event(run), it recorded 30 numbers to reconstruct the event for every channel. Totally 30 numbers are given by 13 ADC counts in both highgain(HG) and lowgain(LG) plus 2 TOTgain(Time Over Threshold) and 2 TOAs(Time Of Arrival). The 13 ADC counts of HG and LG come from 13 SCA(Switched Capacitor Array) units in the circuit, which sample the input signal every 25ns[1]. They can be used to define the hardware noise and differences between each capacitor. At the same time, we recorded the time stamp with the pulse trigger[2] for each event. Time stamp label the whole pulse in order, so we can use

---

[1]There exist thirteen different capacitors with the number SCA0, SCA1....SCA12, and they will record one ADC value in order every 25ns for the each channel of CHIPs. When going to SCA12 and finishing recording, it will return to SCA0. The order of SCA numbers are not related to the order of the electrical pulse we give, but the time stamps are.

[2]In real case, it recorded the starting SCA and the end of SCA with roll position array when the pulse

the time stamp to mark the pulse location, specially for the peak of pulse.

2.2 Events and Methods for analysis

For simplifying the study, we used the Hexaboard without the sensor on it. In this studies, we used two types of runs to do the researches through the whole process:

- Pedestal run: Run without injecting the electrical pulse, and record the non-pulse run to be our reference.
- Charged run: Run with injecting the electrical pulse to the certain channel, and record the with-pulse run to do the study.

Be attention with, because we simulated the sensor-attached condition, and sensor only apply the 32 channels on the Hexaboard in real life, we used 32 channels of the Hexaboard to study. In the Fig.2 is the channel map that we used in the study.
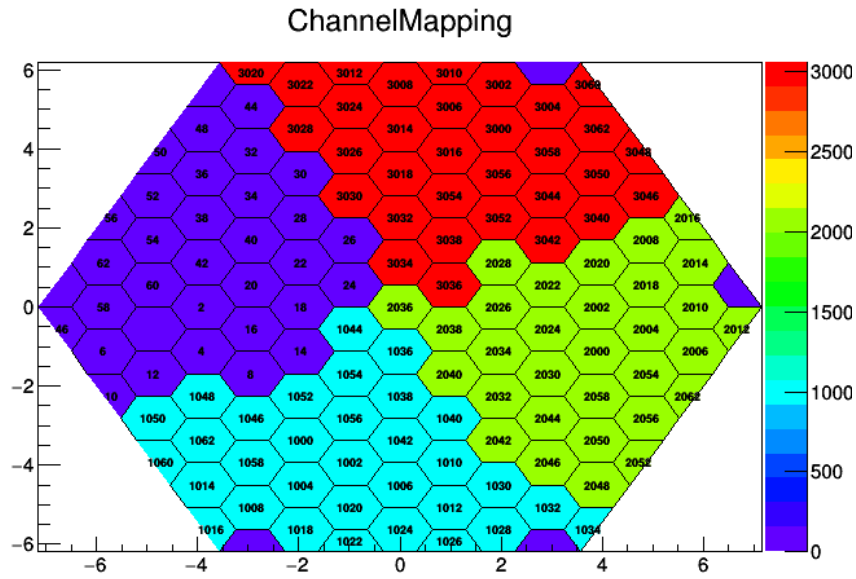

Figure 2: The channel map which was applied in the research, there are four regions with the different color represent the different SKIROC2cms CHIPs. A digit in thousands is the order of the CHIPs, and a digit in tens and ones represent the channels on each CHIPs.

The following three cases which were used in our studies are shown in the fig.3. (1)Pedestal run were used to evaluate the pedestals and noise for each channel. The value of mean of ADC-counts will be defined as the pedestal value from each channel, and it is the hardware-dependent(SCA-dependent) value. (2)Charged run were used

---

trigger is on and off, and rearrange the SCAs with this array as the time stamp, so it can be specified as the pulse information.

to record the responses with ADC-counts values from every channel with the electrical pulse injecting, and simulated the real physics. (3)In the end, the "Pedestal-subtracted" value will be used in our study for subtracting the "reference". The mean value of the ADC-counts for it is calculated after Charged run ADC-counts subtract the mean of ADC-counts from Pedestal run with same SCA number. And the error(width of gaussian) means the noise in the channels including the electronics noise, sensor noise(if it is installed), etc. Note, in the study, we always used SCAs of HG to do. For simplifying the case, we fixed the injected-channel to number 20 in CHIP 0 and saw all cases.
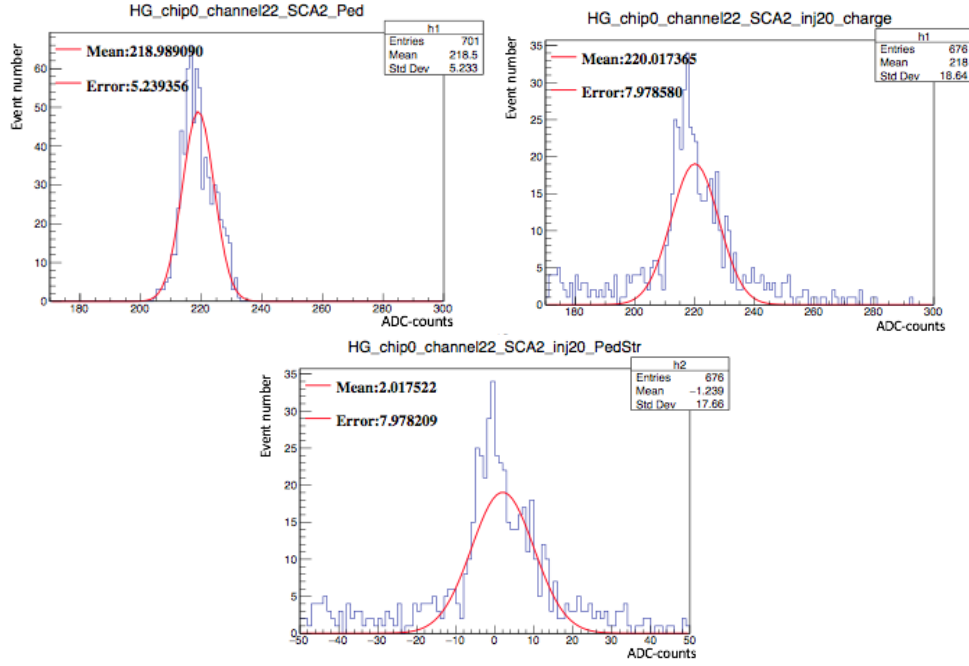

Figure 3: These figures perform the histograms that using in the study. All this cases are shown with fixing at SCA=2 and channel=22. (Top left) The pedestal run histograms which is used to calculate the pedestal and noise. Fitting the line is to get the mean of ADC-counts. (Top right) The charged run for example. (Down) The pedestal-subtracted case is shown. Also fit the line to get the mean and error of ADC-counts, and use in the study directly

2.3 Results and Conclusion

First of all, because the first SCA (the 0th) is useful for making on-line pedestal subtraction (because there may be some low frequency noise), and usually our signal comes after SCA0, we need to explore the correlation between the stability of pedestals(SCA0) and other SCA sampling. In the Fig.4, we can see that the nearest channel of the injection channel (channel 20) with channel 18 and 22 as examples, the stability of them are very good. Because the fitting lines are pretty flat in both of them, and that means in the different SCA number other than SCA0, they are slightly different. And we can see the same condition in other channels also.

Second, we wanted to explore the correlation between injection pulse strength and the cross-talk. In real physics, because particles will have different energy individually in the collider, they will see the different pulse strength in the real case. At there, we
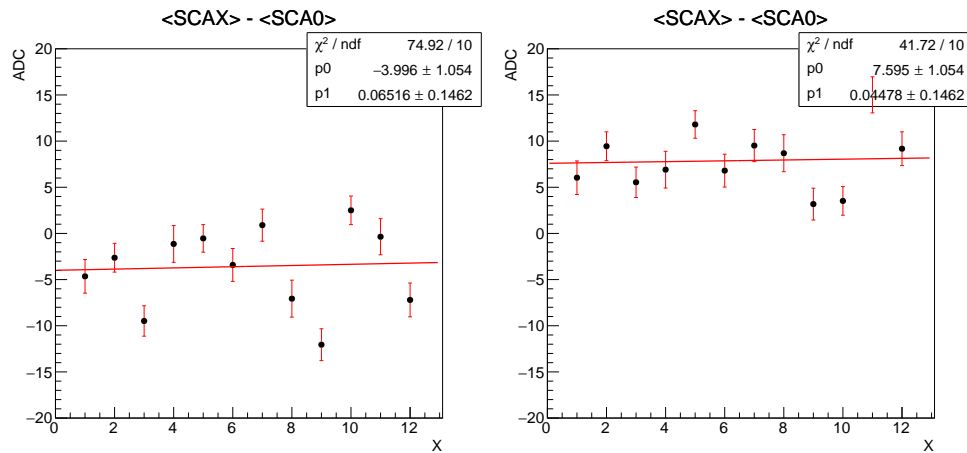
Figure 4: The figures show the difference mean of ADC-counts between the first SCA and other SCAs. (Left) The channel 18 response and (Right) The channel 22 response.

wanted to see whether the different pulse strength will give out more or less the cross-talk. In the fig.5, they show that the mean and error values from Pedestal-subtracted value for different DAC of injection pulse. The results for them are quiet obviously showing that the cross-talk in two cases:

- Anti-correlated cross-talk
  DAC from 0 to 1000: No matter the closer or farther the channels, the slopes are negative for all of them. That means the channels, except the injection channels, are borrowed the energy by injection channels.

- Correlated cross-talk
  DAC after 1000: When the channels are closed to the injection channels, the slopes are bigger compared with the channels which are farther from the injection channels. That means, the injection channels give more energy to the closed channels compared with the farther channels.

For the conclusion, we can see from our study that the cross-talk exist in the Hex-aboard, and we used the ADC-counts from the Pedestal-subtracted values to quantify the cross-talk. We expected that this study can help to quantify the cross-talk and can decrease it in the future.

3   Pion-rejection studies with a CMS HGCAL test-beam prototype ECAL

In most cases, the electrons leave most of the energy in ECAL by the processes of pair production and bremsstrahlung. Different from them, Chagred-pions deposit most of the energy in Hadronic Calorimeter(HCAL) through the process of segmentation. Unfortunately, in some cases, those pions could give out the fully Electromagnetic(EM) showering in ECAL, and put most of the energy in it. If we can't explore the way to tag those pions and reject them, it will be very terrible. Because they will be misidentified as the electrons, and those "fake electrons" could contaminate with the real electrons. In addition, it could influence the analysis which is sensitive to the electron/pion dis-
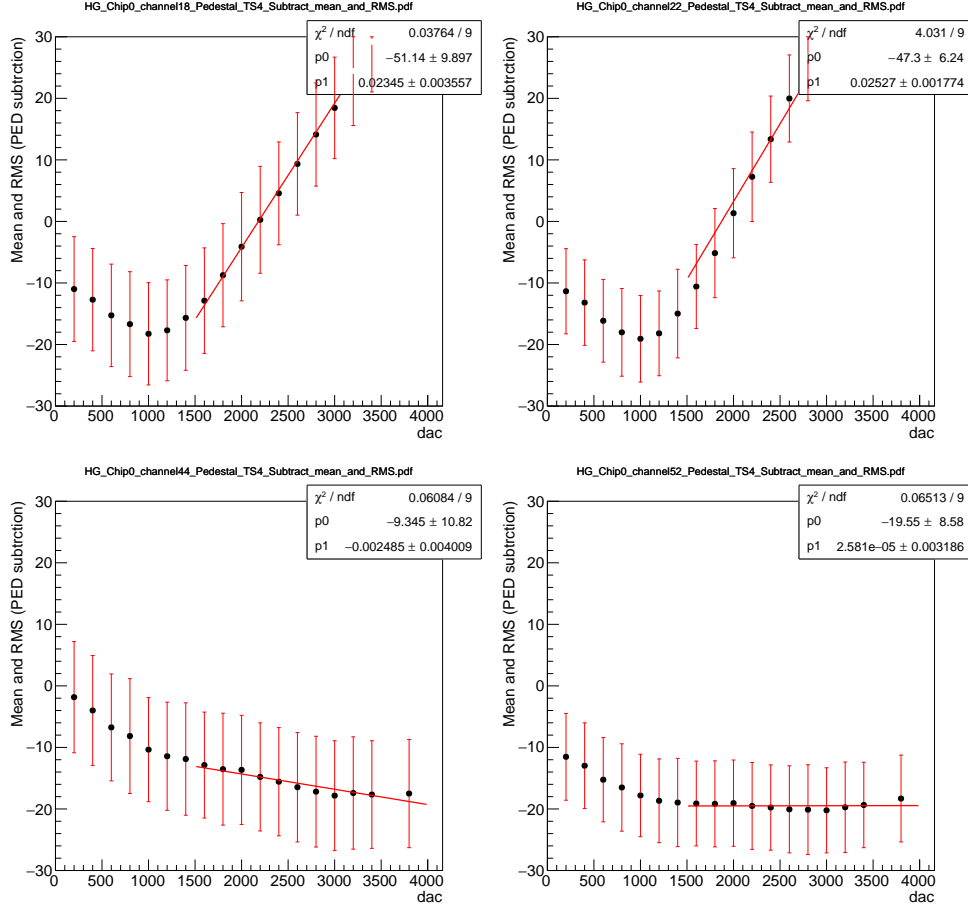
Figure 5: These figures present the preliminary results for the correlation between the different DAQ input and Pedestal-subtracted mean and error. The Top left and right pictures perform the channel of 18 and 22, which are closed to the injection channel. The down left and right pictures show channel of 44 and 52, which are farther from the injection channel.

tinguishability. We expected to solve this issue and help the physics analysis with this type of background in the future.

Because of the significant challenges for radiation tolerance and event pileup on the detectors of the CMS experiment at LHC, they will replace the current Endcap calorimeter with a Si-pad HGCAL, a new generation state-of-the-art calorimeter, which can perform 3D imaging of the shower as well as provide 30ps timing. To test the performance of HGCAL, including calibration, particle identification, etc. before it will be installed in CMS Phase II upgrade , we did the "test-beam"(Fig.6), which was held to incident the test particles, such as electron, pion and muon into HGCAL prototype, and analyze its performance from the information of the hits come out with the detector. At the same time, The Monto Carlo(MC) samples are provided with the same prototype of the detectors to let the people study as well. We used the test-beam MC samples which are done in CMS on June and October 2019 to do this study. The group of test-beam divided the studies into many different topics, and my topic was in the section of "electron/pion identification".
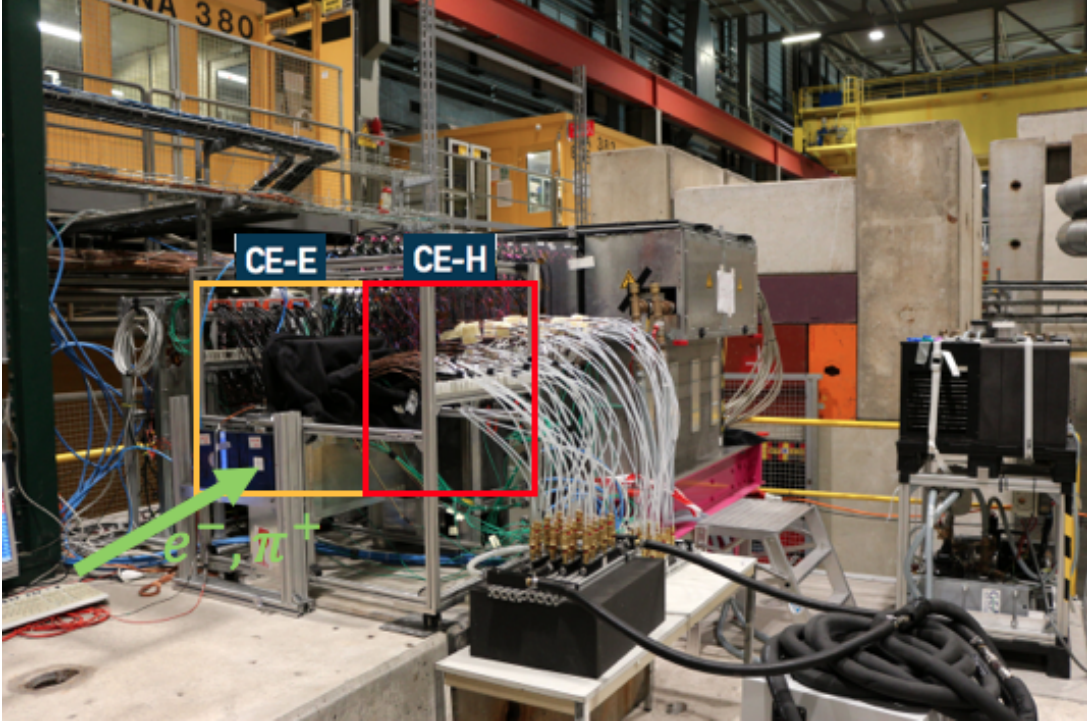
Figure 6: The figure show the test beam experiment setup in October. The orange and red lines box mark the CE-E and CE-H part of the detector including the module(in the metal box) and readout electronics.

Charged-pion tagging and rejection are one of the strong points of HGCAL. Traditionally the most powerful discriminants for pion rejection are lateral shower containment and longitudinal energy leakage in the back of the EM section of HGCAL. In this study we explore the capability of HGCAL to tag pions that are fully EM-showering in the ECAL and they pass the transverse containment and leakage from the back requirements. For our expectation, we wanted to find the best cuts and invited the new variables to distinguish "fake electrons" from "real electrons".

I did this study with Prof. Stathes Paganis, Prof. Rong-Shayang Lu, master student Chia-Hong Chein from NTU, Dr. Shilpi Jain from University of Minnesota and Prof. Shin-Shan Eiko Yu from NCU. The contributions of Chih-Hsiang Yeh to this study includes the following:

- Study the optimized cuts for finding the "fake electron" from pions, and introduce the new variable with the longitudinal segmentation using the test-beam MC samples from June.
- Apply the found optimized cuts and new variable in test-beam MC samples from October to see the electron efficiencies and pion rejections.

I will describe the detail as following.

3.1 Introduction for HGCAL

During the Run 1 (2010-2012) in LHC operated at $\sqrt{s} = 7\,\text{TeV}$, delivering $\approx 6\,\text{fb}^{-1}$, and at $\sqrt{s} = 8\,\text{TeV}$ in 2012, delivering $\approx 23\,\text{fb}^{-1}$. The most significant physics results

from this period was the discovery of the Higgs boson, and along with the Nobel prize for it in 2013. After Run 1, Run 2 started in 2015 at a C.M of $\sqrt{s} = 13\,\text{TeV}$ and the instantaneous luminosity, reaching to the value for $1.7 \times 10^{34} cm^{-2} s^{-1}$. In the Figure.7, it shows the summary plot for the luminosity from Run1 to Run2. Surprisingly, it operated the exceed value compare with the original design. More studies with this unbelievable operation, such as Higgs boson, SM processes, BSM, will be carried out.
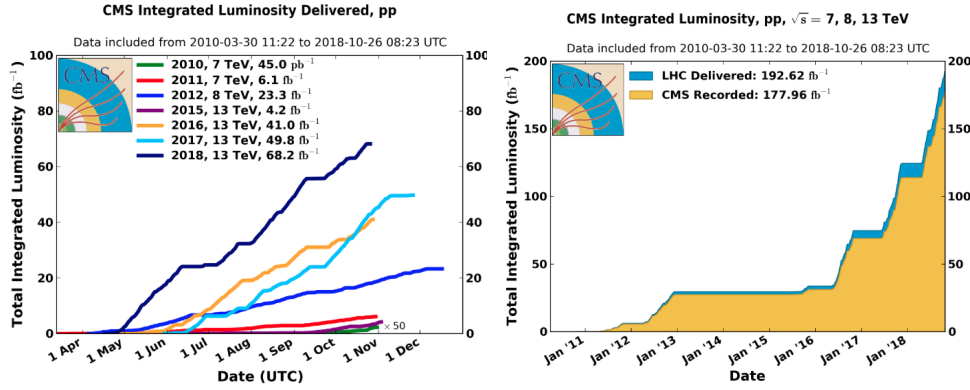

Figure 7: The figures show the summary plot of luminosity from the Run1(2010-2012) to Run2(2012 to 2018) in LHC. The left is the luminosity for the individual years, and the right is the totally integrated luminosity from 2010 to 2018.

For the future in Run3(2023), LHC intended to accumulate around $300\,\text{fb}^{-1}$. After the third long shutdown (LS3), they plan to design the value to the instantaneous luminosity at $5 \times 10^{34} cm^{-2} s^{-1}$ with the target of integrating around $3000\,\text{fb}^{-1}$ by the time of mid-2030s, and it will go into the HL-LHC era. The corresponding mean number of collisions (pileup) per bunch crossing will be 140. In addition, the LHC has the ability to deliver 50% higher values for both the instantaneous and integrated luminosities.

In this case, It will bring out the two significant issues in the HL-LHC era: (1)radiation damage and (2)event pileup on CMS detectors, especially for calorimetry in the forward region. As part of its HL-LHC upgrade program, the CMS Collaboration is proposing to build a HGCAL to replace the existing endcap calorimeters. There are many requirements for the HGCAL upgrade to let it preserve the good performance. Write some of them as following:

- This detector must be "radiation tolerance", otherwise, it will happen the unrecoverable condition, same as the circumstance of scintillator now in CMS. To solve this issue, active layer with the silicon sensor will be applied in the HGCAL.

- This detector need to be designed with the find granularity for lateral and longitudinal segmentation:
    - For the lateral part, it can help with separating two showers and can observe the narrow jets.
    - For the longitudinal part, it can help us to probe the longitudinal development of showers, providing good energy resolution, and also can know more about the patterns in the physics processes.

    Both of them can remove the "pileup event" if they are great enough to

distinguish processes of different particles.

The HGCAL prototype consists of an electromagnetic compartment (CE-E) followed by a hadronic compartment (CE-H) with Forward region(FH) and Backward region(BH) in the Fig.8. For the CE-E, it consists of 28 sampling layers with a depth of approximately 26 X_0 and 1.7 λ. In the Fig.9 top left shows one sampling layer of the CE-E. The element of active layer for it is a hexagonal silicon sensor, which is sandwiched between layer of WCu (75%, 25%) baseplate and a printed circuit board(PCB), which is used to study my cross-talk studies in NTU, that carries the front-end electronics to form a silicon module in the Fig.9 top right. Modules are tiled on a Cu cooling plate, which together with the two WCu baseplates form one absorber layer. The alternate absorber layer is formed by lead planes clad with stainless steel (SS) sheets that are placed on module-cooling plate sandwich.

Active Elements:
- Hexagonal modules based on Si sensors in CE-E and high-radiation regions of CE-H
- "Cassettes": multiple modules mounted on cooling plates with electronics and absorbers
- Scintillating tiles with SiPM readout in low-radiation regions of CE-H

Key Parameters:
- CE covers $1.5 < \eta < 3.0$
- ~215 tonnes per endcap
- Full system maintained at -30°C
- ~600m^2 of silicon sensors
- ~500m^2 of scintillators
- 6M si channels, 0.5 or 1 cm^2 cell size
- ~27000 si modules
- Power at end of HL-LHC: ~110 kW per endcap

Electromagnetic calorimeter (**CE-E**): Si, Cu & CuW & Pb absorbers, 28 layers, 25 X_0 & ~1.3λ
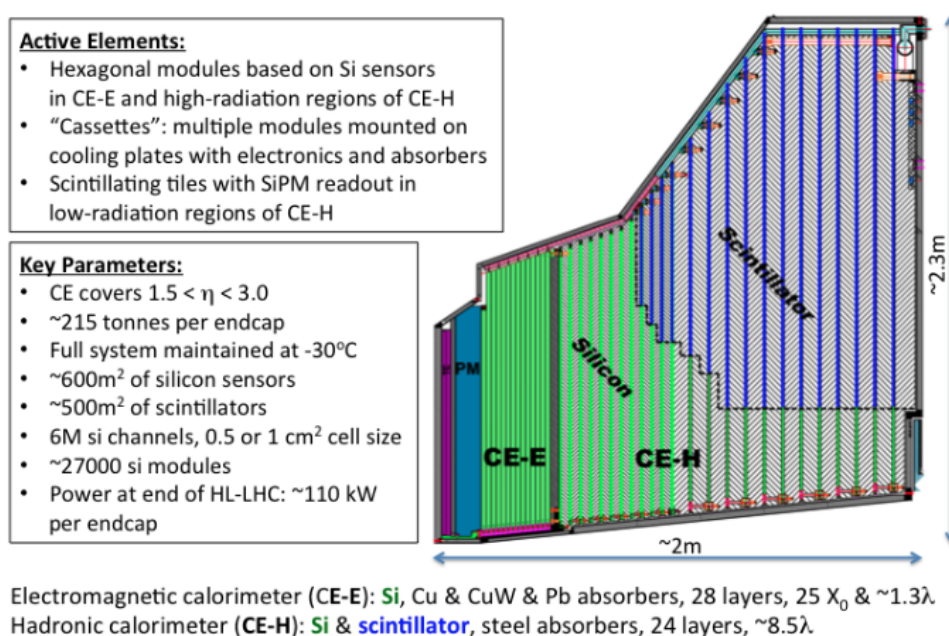Hadronic calorimeter (**CE-H**): Si & scintillator, steel absorbers, 24 layers, ~8.5λ

Figure 8: Summary for the basic design and coverage of the CMS high-granularity Endcap Calorimeter (CE) including CE-E and CE-H

For the hadronic compartment of HGCAL, they consist of 12 planes of thick SS plates followed by another 12 SS planes with different thickness. Between these absorber plates sit silicon modules(show one sampling layer in Fig.9 down left) in most regions and silicon modules mixed with scintillator tileboards(show one sampling layer in Fig.9 down left) in part of low-radiation regions mounted on a copper cooling plates to form the wide cassettes. These cassettes are similar to those in the EM compartment, but include sensors on only one side of the cooling plate, which are formed as a separate mechanical structure. This leads to a total interaction length with number 10.7λ, including the CE-E and the neutron moderator layer in front of the calorimeter. All layers are read out for use in energy measurement, but only alternate layers in CE-E, and all in CE-H, are used for producing L1 trigger primitives.

In the test-beam, we used the material based on prototype of HGCAL written below, but for the detail of its, each components will use the CMS phase 2 design. I will write
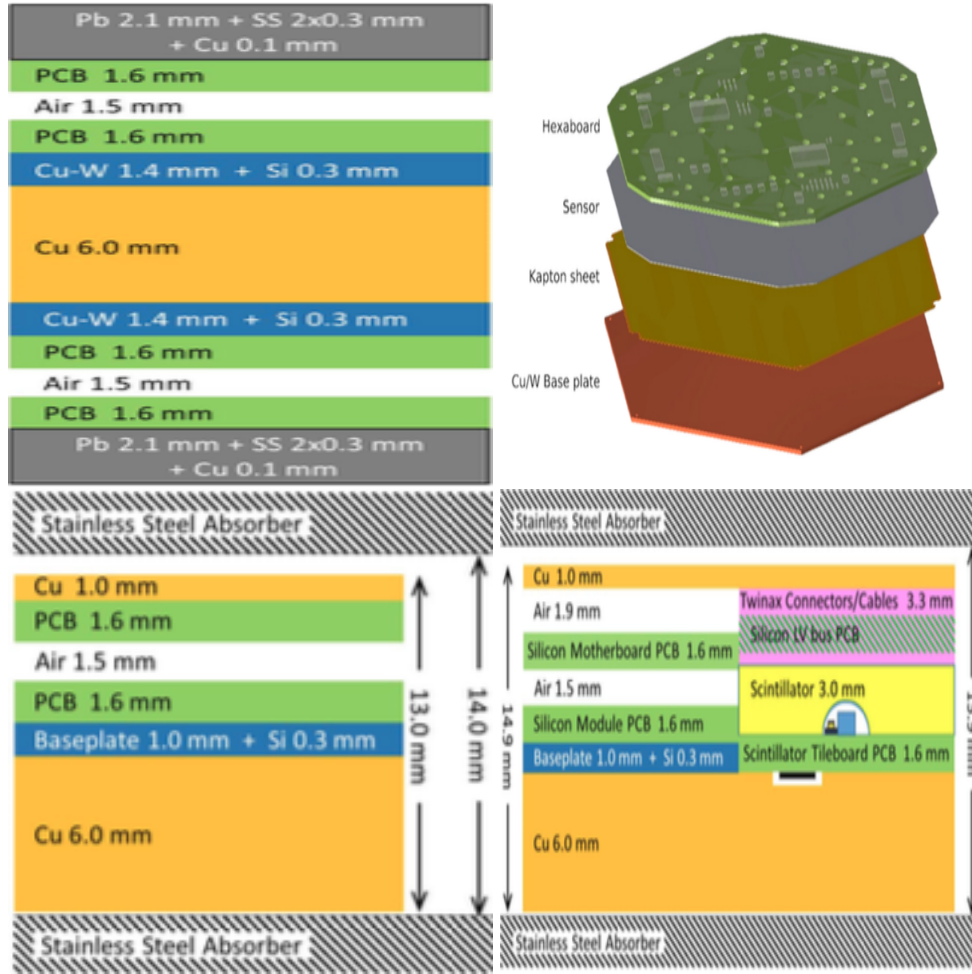
Figure 9: The figures show the different material arrangement in the different part of the detector. The top left is the one sampling layer of the CE-E. The top right show the silicon module used in the CE-E and CE-H. The down left is the one sampling layer of the CE-H which is sit on the high-radiation region. Oppositely, the down right is the one sampling layer of the CE-H which is based on the place with low-radiation.

the results of the test-beam in 2018 on June and October.

3.2   Tag-and-probe the optimized cuts and new variable with June test-beam MC

Because we wanted to tag the "electron-like"(e-like) pion, we need to explore the properties for the electron first, including shower shape, energy deposition, etc. in detector. And then, apply all of them in pion runs to tag those bad pion. I will describe the detail for the cuts and selections as following. In this studies, we used the sample with 50GeV electron and pion MC.

The first selection, because electron and pion have very different shower shape(width) on the transverse plane in ECAL, we studied on them at starting. From the theory, it tells us that electron has the narrower shower shape than pion in ECAL, we explored the cut with the different number of rings on the transverse plane of detector. In the
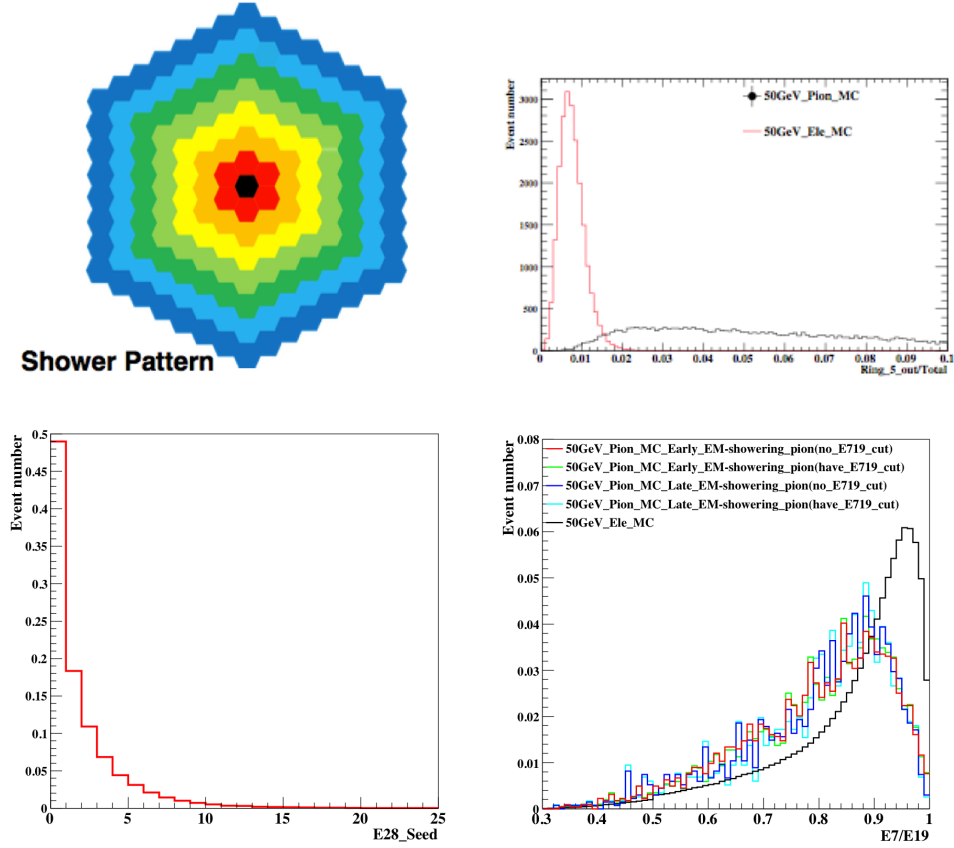
Figure 10: These figures present the cuts we used in the studies. Top left is the shower pattern on the transverse plane in the detector. Different rings mean the different sizes of the clustering on it. For example, the 5-Rings means the fifth circle calculated from the black pad to the shadow green. The top right plot is the energy distribution of $\frac{\text{energy out of the 5 Rings}}{\text{total energy}}$. The down left is the last layer energy with electrons. The down right is the cut with $\frac{\text{energy in 2 Rings}}{\text{energy in 3 Rings}}$, we plot every layer $\frac{\text{energy in 2 Rings}}{\text{energy in 3 Rings}}$ value with and without the $\frac{E_7}{E_{19}}$ cut

Fig.10 top left shows the distribution plot of ring. In the end, we chose the energy of 5 Rings for being the cuts. Fig.10 top right is presented to show the cut we used with "$\frac{\text{energy out of the 5 Rings}}{\text{total energy}}$". Because we can see in the plot, the energy of electrons showering are collected out of the 5 Rings with at most 1% compared with total energy in ECAL. We chose this standard to tag the e-like pions first. Although, it is the tight cut because it also discarded some of electrons as well, but we wanted to make sure that we cut-off the bad pion really. We called this cut as "5-Rings cut".

The second selection, because we wanted to select the "fully EM-showering", and we didn't have the information of HCAL energy in this month test-beam, we only could use the "poor man's solution"-finding the last layer energy of electrons and using it to be our cut. In the Fig.10 down left is the plot which was used to find the cut of last layer energy with the events which passed 5-Rings cut . Eventually, we used the last layer energy smaller than 20MIPs to tag the fully showering events, we called this cut

as "The last layer energy cut".

The third selection, it was the cut always used to find the EM-showering in pion traditionally, and related to the shower shape also, similar with the first selection. In every run, we applied the cut with "$\frac{\text{energy in two Rings}}{\text{energy in three Rings}} = \frac{E_7}{E_{19}}$" need to be at least 0.75 at the shower depth maximum layer. Shower depth(unit:X_0) is the energy weighted in the detector, it means the center-of-energy in the shower. In the Fig.10 down right shows the events with the values of $\frac{E_7}{E_{19}}$ of all layers. Compare between with and without the cut, most of the events with cut have bigger value than without the cut, this means we cut-off some of the boarder shower shape events in pion runs. Both of Early EM-showering and Late EM-showering in the plot are two kinds of e-like pion, will describe the detail later.

The fourth selection, because we applied those cuts from first to third, and we looked the shower depth for those passing events in the Fig.11, we found there exists two types of electron-like pion with the cut described in the next paragraph:

- Early EM-showering: The events have the similar shower depth with the electrons in ECAL, which means the energy distributions for these events are similar with the electrons. We can't distinguish them from electrons very well.
- Late EM-showering: The events have the late showering compared with the electrons in ECAL, and they put most of the energy in the backward ECAL, we can use some variables to distinguish this kind of pion from electrons.
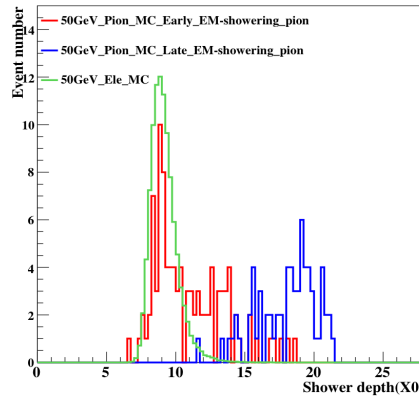

Figure 11: This figure shows the shower depth for both Early EM-showering and Late EM-showering.

We show some event display for normal electron and pion showering, Early EM-showering and Late EM-showering in Fig.12. And we present the event display in the We used the cut that, we require a consecutive number of hits in CE-E starting from the first layer and checking until the 14 Layer. We record the first layer found with more than 3MIPs. And at least 2 consecutive layers with more than two seeds with Energy>3MIPs. If the layer of start showering is from the first to fifth layer, we called it "Early EM-showering", on the other hand, if it starts at sixth to later layers, we called it "Late EM-showering". For note, because this is the convenient way to let us
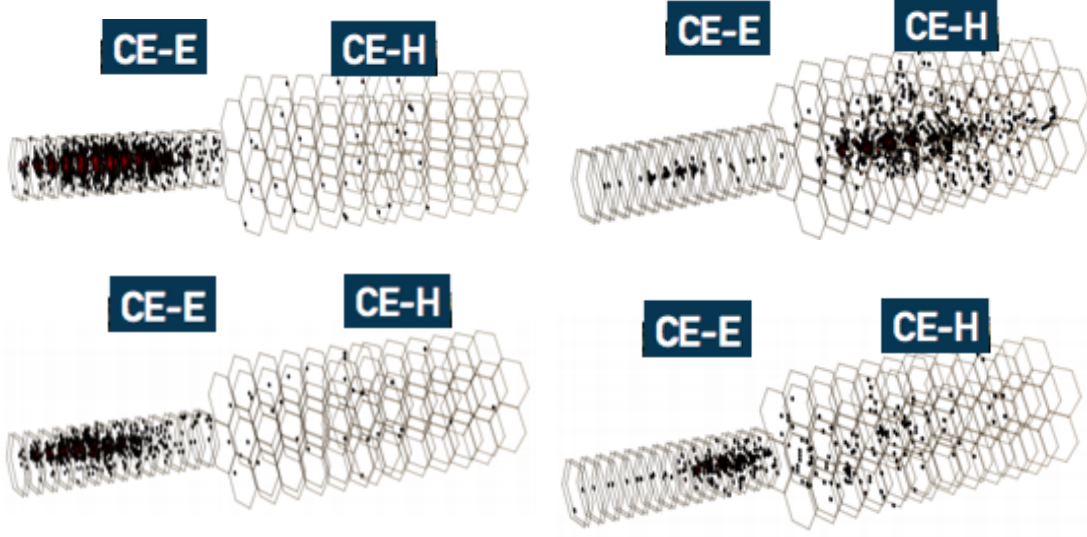
Figure 12: These figures show the event display with the cases which are the used to probe in the studies. (Top left) is the normal electron, most of the energy put in the CE-E, oppositely, (Top right) is the normal pion, which put most of the energy in CE-H. The both cases below the normal particles are the "e-like pion" cases, which we target to distinguish them from the normal electron. The left plot is the case for "Early EM-showering" and the right plot is the case for "Late EM-showering".

see the dynamic of the shower in ECAL approximately, we used this cut to distinguish them. In the real case, we don't do this cut, just used one variable to distinguish "both" of them from electrons. I will show the results later.

Tight cut types	value
5-Rings cut	> 0.99
The last layer energy	$< 20\text{MIPs}$
$\frac{E_7}{E_{19}}$	> 0.75

Table 1: E10/Etotal Tight cuts selection used in June test-beam

Until now, we can see the one strong point of the difference between electron and e-like pion - the shower energy deposit in the ECAL. Because some of them put energy in the different section of succeeding layers in ECAL, specially for comparing the electrons with Late EM-showering. In this case, we thought about the variables with "longitudinal segmentation" with ECAL. We developed the discrimination variable as following:

$$E_{10} = \frac{E_{\text{first 10 layers energy}}}{E_{\text{total energy}}} \tag{1}$$

Intuitively, because the electrons put most of the energy in the front of the detector, they will have the bigger value for E_{10}. Oppositely, for the e-like pions, because they put all of the energy in the different sections, they will have many values for E_{10} from all events. In the Fig.13, we can see that we can use E_{10} to reject some of e-like pions out of the electron region.

In my study, we need to find the critical point of E_{10} to be the threshold, because we wanted to reserve the electrons as many as possible, and cut-off the bad pions for
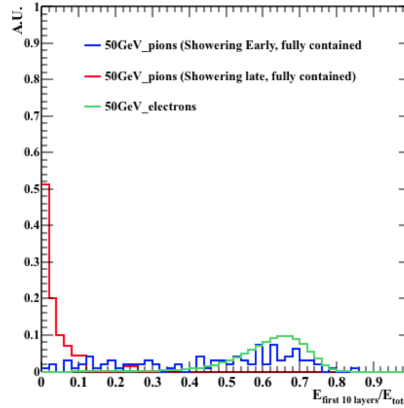
Figure 13: The E10 value for electrons, Early EM-showering and Late EM-showering.

our best. We used the quantities about the "Electron Efficiency" and "Background Rejection" as our standard in the following formula, to see whether this cut is good:

$$\text{Electron Efficiency} = \frac{\text{Pass electrons}}{\text{Total electrons}} \tag{2}$$

$$\text{Background Rejection} = \frac{\text{Total e−like pions}}{\text{Pass e−like pions}} \tag{3}$$

Our goal is to cut-off the most e-like pion with the highest background rejection (pass the least e-like pions) in (2), but reserve the most electrons with the highest electron efficiency. (pass the most electrons) in (3).

First, we tried to use the critical point at 0.2, 0.3 and 0.4, and see whether it can give us the best results. In the Table.2, it tells us that the background rejection is good (more than half of e-like pions are rejected), but we hoped that our electron efficiency will go to 99.8% or more. On the other hand, we still wanted to keep the electrons alive after the higher critical point to reserve most important researches on them.

E10/Etotal cut	Background Rejection	Electrons Efficiency
>0.2	2.15+/-0.18(stat)	99.85%+/-0.01%(stat)
>0.3	2.53+/-0.24(stat)	99.30%+/-0.01%(stat)
>0.4	2.74+/-0.28(stat)	97.32%+/-0.01%(stat)

Table 2: E10 cuts of Electrons Efficiency and Background Rejection comparison with point 0.2,0.3 and 0.4

In this case, we tried to use another points near 0.2 and see whether we can get the great results without losing more electrons. In the Table.3, it said that near the point 0.2, it can reserve most of the electrons near 99.9% and can get the background rejection with the expectation value near 2 (Half of rejection).

For the conclusion of this month test-beam results, we found that the Background Rejection can be near 2 without killing many electrons with 99.9% Electron Efficiency. For our expectation, we wanted to apply this cut in the October test-beam, to see whether we can see the similar results.

E10/Etotal cut	Background Rejection	Electrons Efficiency
>0.16	2.02+/-0.15(stat)	99.93%+/-0.01%(stat)
>0.18	2.07+/-0.16(stat)	99.89%+/-0.01%(stat)
>0.20	2.15+/-0.17(stat)	99.85%+/-0.01%(stat)

Table 3: E10 cuts of Electrons Efficiency and Background Rejection comparison with point 0.16,0.18 and 0.20

3.3   The results of application in October test-beam MC

In this month test-beam, because we had the information of HCAL, we used the cut no more than 0.4% energy fraction in the HCAL energy to replace the poor man's solution. And also, we changed some parameters of the cuts to get the best results. In the Table.4, the summary table for the cuts is shown.

Tight cut types	value
5-Rings cut	> 0.99
$\frac{\text{HCAL total energy}}{\text{ECAL total energy}}$	< 0.004
$\frac{E_7}{E_{19}}$	> 0.85

Table 4: Tight cuts selection used in October test-beam

In the June test-beam, we only did the rejection for E_{10} after the tight-cut selections. In this month, we wanted to studied the power of the cuts from tight-cut selections to E_{10}, it means we wanted to study totally how many e-like pions will be rejected after all cuts. We have two steps rejection, and we used the same definitions with the equation (2) and (3).

- Tight cuts rejection power:
  This is the first step rejection study. Because we need to find out the e-like pions with the cut in the Table.4 first, and then we can use our segmentation variable to reject them for the next step. For this step, we found that the number of the background rejection is 3600. Most of the pions are rejected by those cuts.

- E_{10} cuts rejection power:
  This is the second step rejection study. After we found out the e-like pions, we used the discrimination variable which we have invited in June test-beam,E_{10}, to distinguish them from the real electrons. In this step, we used the critical point at 0.18,0.20 and 0.22, to see whether at the critical point near 0.2 will give us the good results. In the Table.5 summarize the results for them. Although, this time didn't have the better results compared with the June test-beam, we still got the number about 1.5.

E10/Etotal cut	Electron Efficiency	Background Rejection
> 0.18	99.56%+/-0.01%(syst)+/-0.20%(stat)	1.52+/-0.27(stat)
> 0.20	99.05%+/-0.01%(syst)+/-0.20%(stat)	1.53+/-0.26(stat)
> 0.22	98.91%+/-0.01%(syst)+/-0.20%(stat)	1.56+/-0.24(stat)

Table 5: E10 cuts electrons efficiency and Background rejection after the tight cuts

In the end, totally we can get the total rejection power of cuts which is calculated by the number 3600 with the Tight cuts rejection power and the number 1.5 with the E_{10} cuts rejection power. After times them together, we can get the number $3600 \times 1.5 = 5400$. In A Toroidal LHC ApparatuS(ATLAS), the expected value for it is between 5000 to 10000, and in our case, we arrived the value for the expectation.

3.4 Conclusion

In this section, we obtained the cuts and new variable E_{10} to distinguish the e-like pions from real electrons with the June test-beam MC samples, and we applied in the October test-beam MC samples, happily, we got the expected results compared with the previous studies in ATLAS. This cuts and variable have never been applied in CMS analysis before, and we expect that we can help to reduce the background which is sensitive to electron/pion identification with HGCAL in the future.

4 Studies of granularity of a hadronic calorimeter for tens-of-TeV jets at a 100 TeV pp collider

After the discovery of the standard-model-like Higgs boson at 2012[1], we have established a milestone in HEP field. We used the LHC at CERN to solve an important puzzle in the SM of particle physics, Higgs boson, and it gave us the answer why the elementary particles retain their mass[2] from the early universe to now. We expect more particles predicted by BSM can be found in the future. A lot of these new particles may be heavier than 13 TeV[3][4], which is operating now in LHC. Therefore, particle physicists plan to build a very high energy collider with $\sqrt{s} = 100\,\text{TeV}$ in the future, such as high-energy LHC (HE-LHC), future circular pp colliders of the European initiative, FCC-hh [5] and the Chinese initiative, SppC [6].

In the very high energy era, the C.M. energy will be increased compared with now in LHC, and the instantaneous luminosity is also expected to increase. In this case, more interesting events will occur within a shorter amount of time, but incidentally, it will accompany with the unexpected background. We need to figure out some methods to pick out the signals which we are interested in, and filter the unwanted huge background events extremely. The silicon future circular collider (SiFCC) detector, which is based on the silicon detector (SiD)[7] designed for international linear collider (ILC)[8][9], is our simulated detector prepared for the high energy collider for the future era in Ref.[10].We used this detector to simulate the condition under very high C.M., and applied the different configurations of detector with some methods such as jet substructure variables, to see the detector performance of distinguishing the signal from background. I will describe the detail later.

Traditionally, we used scintillators as the active layers of the ECAL and HCAL to measure the energy of the showering of particles which are induced by the absorbers (with material such as lead, brass, or tungsten, etc.). But when the radiation dosage is much higher than before, scintillators can deteriorate easily. This is the same circumstance

which CMS is bumped into now. In addition, because of time consuming for making
the special shape for fitting the detector, it is not easy to sub-divide scintillators into
smaller areas. And also, we need to have a readout with high position resolution to
match them. In this case, we need to use the new design- "high granularity" detector to
resolve the resolution problems, i.e., we used the different configurations of detector to
record the precise position and energies. For the radiation problem, because the silicon
detector is known to be radiation tolerant.[11], we could use it in the future collider,
same as HGCAL which was presented in my previous studies.

I did this study with Prof. Shin-Shan Eiko Yu from NCU, Prof. Ashutosh Kotwal
and student Sourvan Sen from Duke University, Dr. Sergei Chekanov and Dr. Proud-
foot James from Argonne National Laboratory(ANL), and Nhan Viet Tran from Fermi
National Accelerator Laboratory. The contributions of Chih-Hsiang Yeh to this study
includes the following:

- Study of detector performance with soft drop mass.
- Studies of signal and background separation using the different jet substruc-
  ture variables.

I will describe the detail as following.

4.1   Simulation of detector response and event reconstruction

The description of the detector and software used for our study is discussed in [10].
We used the geometry of SiFCC detector with the software package that represents
a various environment for simulations of detector performance. It is used to prepare
for new technology options, event reconstruction techniques for future 100 TeV collid-
ers. Figure 14 is the SiFCC detector simulated by the GEANT4 package[12] and it was
the detector studied in Ref.[10]. The baseline detector discussed in [10] uses a steel-
scintillator hadronic calorimeter with a transverse cell size of 5×5 cm^2, which corre-
sponds to $\Delta\eta \times \Delta\phi = 0.022 \times 0.022$. In addition, because we wanted to test the detector
performance of granularity effect with hadronic processes of signal and background,
several geometry variations were considered, such as 20×20 cm^2 and 1×1 cm^2,
which correspond to $\Delta\eta \times \Delta\phi= 0.087 \times 0.087$ and 0.0043×0.0043, respectively.

SiFCC detector which was simulated in our study contains five main parts: tracker,
ECAL, HCAL, solenoid, and muon chambers. When particles are produced in the
collider, they will go through the detector and leave signals to be recorded. First, the
tracker can record the trajectory of a charged particle because it can leave the energy in
and bended by the solenoid, we can obtain the transverse momentum of this charged
particle. Electrons and photons will leave energy in ECAL , and hadronic particles,
such as neutrons, protons, etc., will deposit energy in HCAL mainly. Muons will go
through the whole detector and could be detected by muon chamber. Figure **??** lists the
material and the configuration for the sub-detectors in SiFCC. Note, since for the time
being we have more experimental results coming from the scintillator-based HCAL
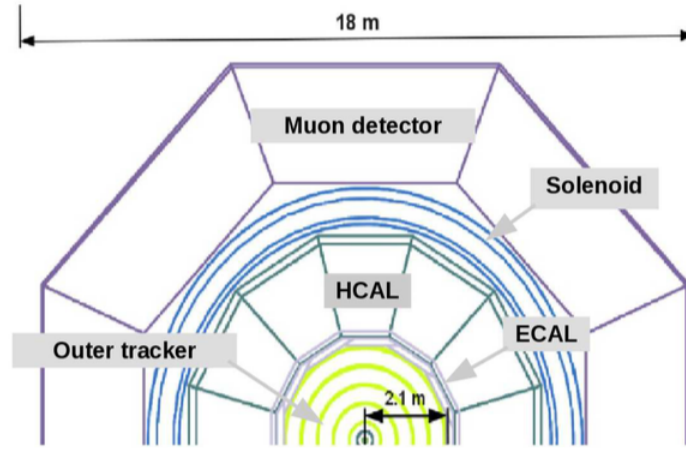in CMS and ATLAS, we have been using this type of HCAL as a first study. After

Figure 14: The azimuthal cross section of the SiFCC detector.

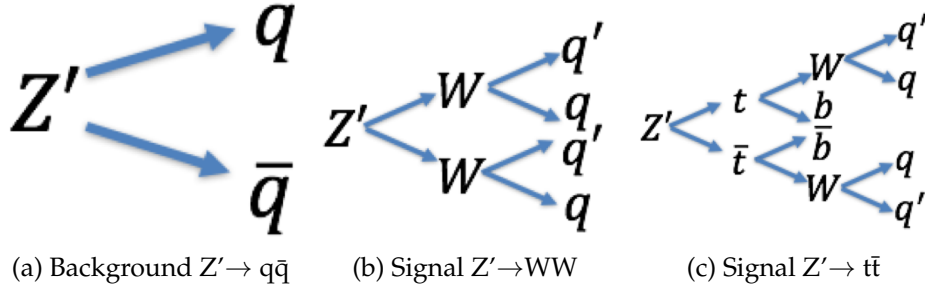(a) Background Z′→ qq̄   (b) Signal Z′→WW   (c) Signal Z′→ tt̄

Figure 15: The signals and background we study.

finishing the first study, we will move to the study of HCAL with silicon sensors as active layers later, same as HGCAL material.

In our studies, we used the simulations of a heavy Z' boson, a hypothetical gauge boson that arises from extensions of the electroweak symmetry of the SM. The Z' bosons were simulated with the masses, $M = 5$, 10, 20 and 40 TeV. The lowest value 5 TeV represents a mass that is the achieved C.M energy of the LHC experiments. The value 40 TeV represents the physics for 100 TeV colliders. The Z' particles are forced to decay to two light-flavor jets (qq̄) with one-prong jet as background, W^+W^- or tt̄ as signal, where $W(\rightarrow q\bar{q})$ with two-prong jets, and $t(\rightarrow W^+ b \rightarrow q\bar{q}b)$ with three-prong jets decay hadronically. The processes for signal and background are shown in Fig.15. In all such scenarios, two highly boosted jets are produced, which are typically back-to-back in the laboratory frame. The main difference between considered decay types lays in different jet substructure. The events of signals were generated using the generator with the default settings, ignoring interference with SM processes. The event samples which were used in this paper are available from the HepSim database [13].

4.2   Study of detector performance with soft drop mass

In this section, we use the jet mass computed with a specific algorithm, soft drop declustering, to study the performance of detector with various detector cell sizes and

center-of-mass (c.m.) energies.

4.2.1 The technique of soft drop declustering

The soft drop declustering [14] is a grooming method that removes soft wide-angle radiation from a jet. The constituents of a jet j_0 are first reclustered using the Cambridge-Aachen (C/A) algorithm [15, 16]. Then, the jet j_0 is broken into two subjets j_1 and j_2 by undoing the last stage of C/A clustering. If the subjets pass the following soft drop condition, jet j_0 is the final soft-drop jet. Otherwise, the algorithm redefines j_0 to be the subjet with larger p_T (among j_1 and j_2) and iterates the procedure.

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}}(\frac{\Delta R_{12}}{R_0})^\beta, \tag{4}$$

where p_{T1} and p_{T2} are the transverse momenta of the two subjets, z_{cut} is soft drop threshold, ΔR_{12} is the distance between the two subjets in the η-ϕ plane, R_0 is the characteristic radius of the original jet, and β is the angular exponent.

In our study, we compare the performance of future detector when setting $\beta = 0$ versus when setting $\beta = 2$. For $\beta = 0$, the soft drop condition depends only on the z_{cut}. For $\beta = 2$, the condition depends on the angular distance between the two subjets and z_{cut} and the algorithm becomes infrared and collinear safe.

4.2.2 Analysis method

We employ the following method to quantify the detector performance and find out the cell size that gives the best separation power to distinguish signal from background. For each configuration of detector and c.m. energy, we draw the receiver operating characteristic (ROC) curves in which the x-axis is the signal efficiency (ϵ_{sig}) and y-axis is the inverse of background efficiency ($1/\epsilon_{\text{bkg}}$). In order to scan the efficiencies of soft drop mass cuts, we vary the mass window as follows. We first look for the median bin i_{med}[3] of the soft drop mass histogram from simulated signal events. Taking the right boundary of bin i_{med} as the center of mass window x_{center}, we start increasing the width of mass window symmetrically on the left and on the right of x_{center}, in steps of 5 GeV, i.e. the narrowest mass window is $[x_{\text{center}} - 5, x_{\text{center}} + 5]$. If one side reaches the boundary of the mass histogram, we only increase the width on the other side, also in steps of 5 GeV. For each mass window, there will be corresponding ϵ_{sig} and ϵ_{bkg}, which gives a point in the ROC curves.

4.2.3 Results and conclusion

Figures 16, 18, 20, and 22 present the distributions of soft drop mass for $\beta = 0$ and $\beta = 2$ with different c.m. energies and detector cell sizes; the signals considered are Z′→WW and Z′→tt̄. In Figs. 17, 19, 21, and 23, ROC curves from different detector cell sizes are compared for each c.m. energy, respectively.

Figures 17 and 19 show that for $\beta = 0$ the smallest detector cell size, 1 cm × 1 cm, has the best separation power at $\sqrt{s}$ =5, 10, and 20 TeV when the signal is Z′→WW

---

[3]The integral from bin 0 to bin i_{med} ($i_{\text{med}} - 1$) should be greater (less) than half of the total number of events. Note, the bin width is 5 GeV.
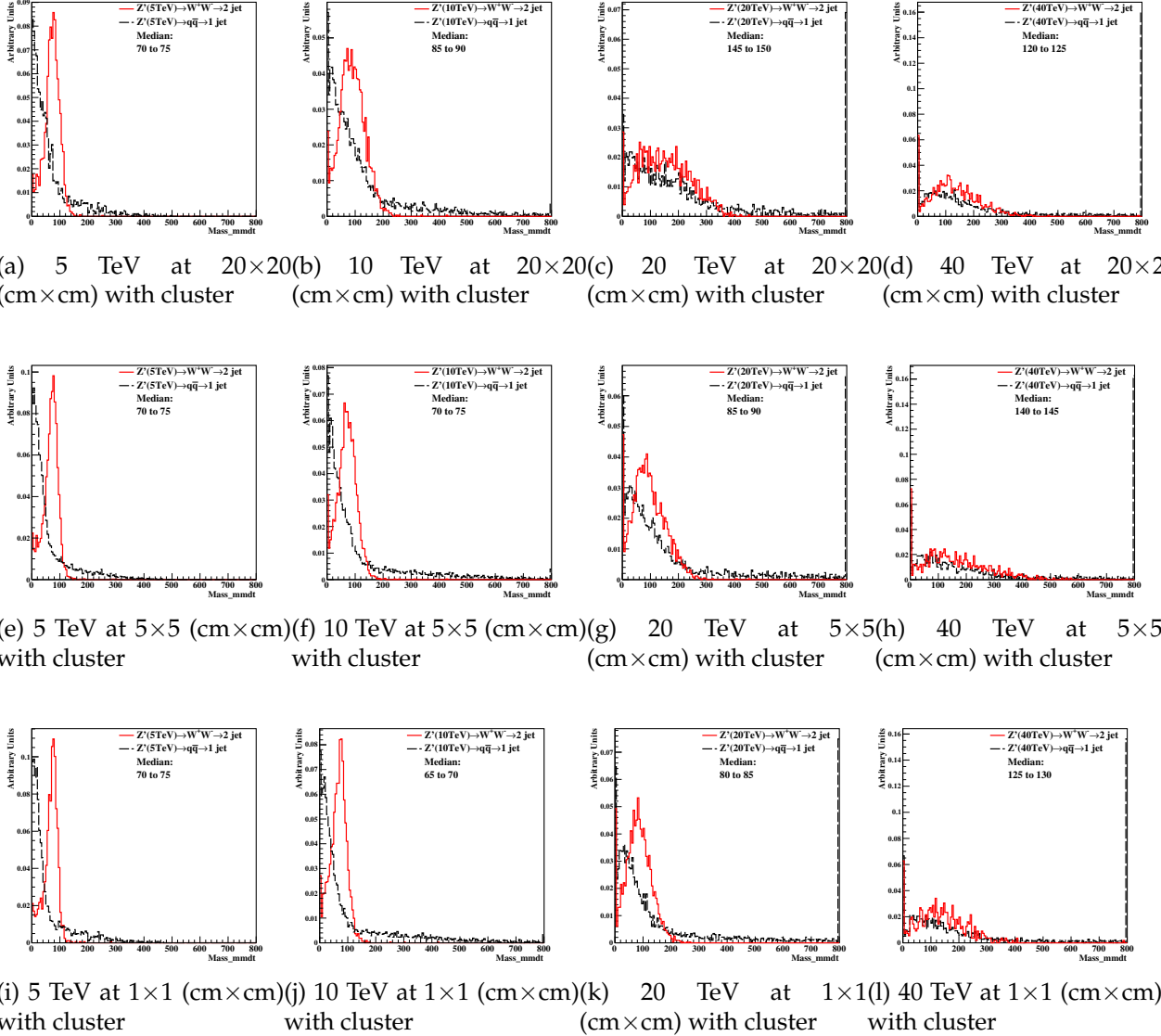
(a) 5 TeV at 20×20 (cm×cm) with cluster

(b) 10 TeV at 20×20 (cm×cm) with cluster

(c) 20 TeV at 20×20 (cm×cm) with cluster

(d) 40 TeV at 20×20 (cm×cm) with cluster

(e) 5 TeV at 5×5 (cm×cm) with cluster

(f) 10 TeV at 5×5 (cm×cm) with cluster

(g) 20 TeV at 5×5 (cm×cm) with cluster

(h) 40 TeV at 5×5 (cm×cm) with cluster

(i) 5 TeV at 1×1 (cm×cm) with cluster

(j) 10 TeV at 1×1 (cm×cm) with cluster

(k) 20 TeV at 1×1 (cm×cm) with cluster

(l) 40 TeV at 1×1 (cm×cm) with cluster

Figure 16: Distributions of soft drop mass for β=0, with 5, 10, 20, and 40 TeV c.m. energies and three different detector cell sizes: 20×20, 5×5, and 1×1 (cm×cm). The signal (background) process is Z′→WW (Z′→q$\bar{q}$).

and at $\sqrt{s}$ =10 and 20 TeV when the signal is Z′→t$\bar{t}$. On the contrary, Figs. 21 and 23 show that for $\beta = 2$ the smallest detector cell size does not have improvements in the separation power with respect to those with larger cell sizes. In fact, the performances of the three cell sizes are similar. In addition, sometimes bigger detector cell sizes, 5 cm × 5 cm or 20 cm × 20 cm have the best separation power.

We also find compared to $\beta = 2$, soft drop mass with $\beta = 0$ has better performance for distinguishing signal from background. Therefore, we will apply requirements on this variable when studying the other jet substructure variables.
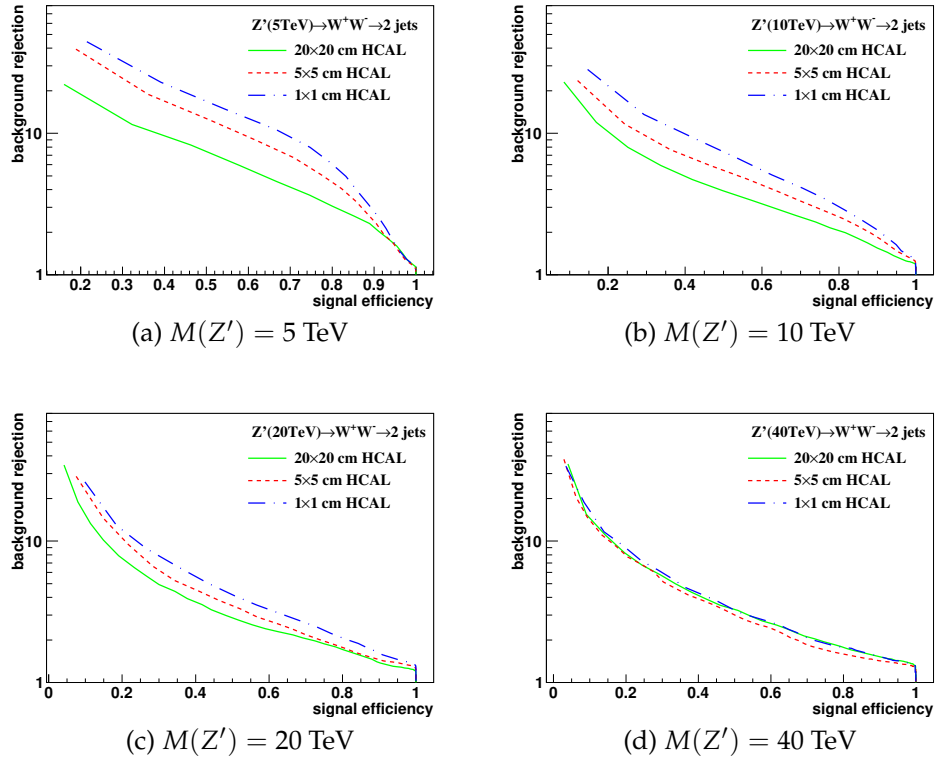
(a) $M(Z') = 5$ TeV

(b) $M(Z') = 10$ TeV

(c) $M(Z') = 20$ TeV

(d) $M(Z') = 40$ TeV

Figure 17: The ROC curves of soft drop mass selection for β=0 with 5, 10, 20, and 40 TeV c.m. energies. Three different detector cell sizes are compared: 20×20, 5×5, and 1×1 (cm×cm). The signal (background) process is Z'→WW (Z'→qq̄).

(a) 5 TeV at 20×20 (cm×cm) with cluster

(b) 10 TeV at 20×20 (cm×cm) with cluster

(c) 20 TeV at 5×5 (cm×cm) with cluster

(d) 40 TeV at 5×5 (cm×cm) with cluster

(e) 5 TeV at 1×1 (cm×cm) with cluster

(f) 10 TeV at 1×1 (cm×cm) with cluster

(g) 20 TeV at 20×20 (cm×cm) with cluster

(h) 40 TeV at 20×20 (cm×cm) with cluster

(i) 5 TeV at 5×5 (cm×cm) with cluster

(j) 10 TeV at 5×5 (cm×cm) with cluster

(k) 20 TeV at 1×1 (cm×cm) with cluster

(l) 40 TeV at 1×1 (cm×cm) with cluster
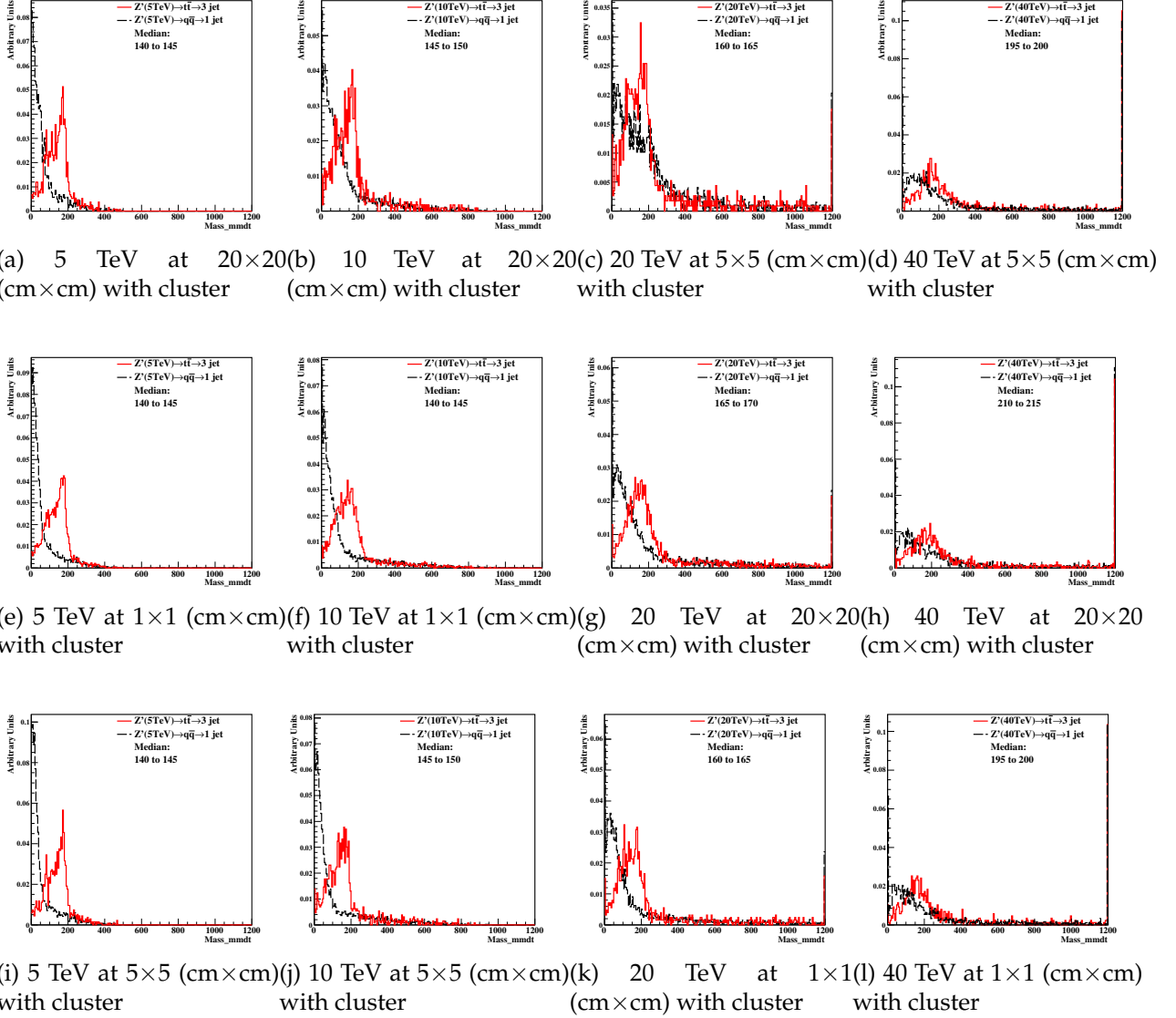
Figure 18: Distributions of soft drop mass for β=0, with 5, 10, 20, and 40 TeV c.m. energies and three different detector cell sizes: 20×20, 5×5, and 1×1 (cm×cm). The signal (background) process is Z'→t$\bar{\text{t}}$ (Z'→q$\bar{\text{q}}$).

(a) $M(Z') = 5$ TeV

(b) $M(Z') = 10$ TeV
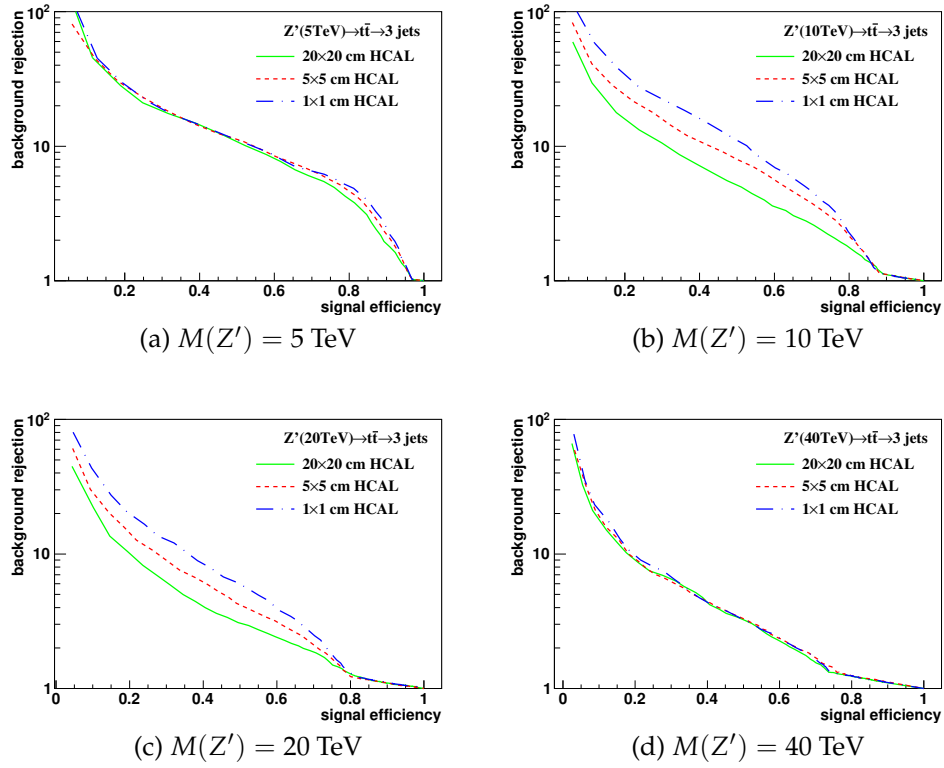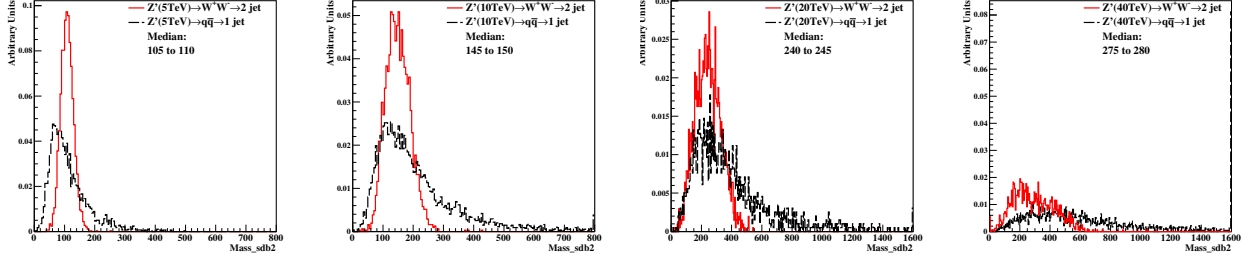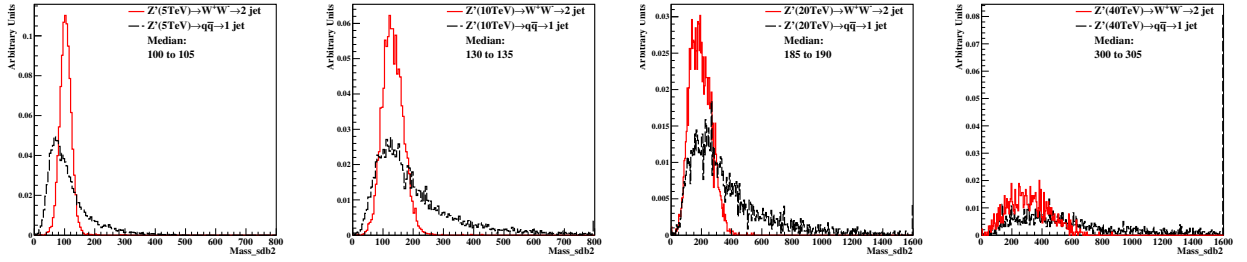
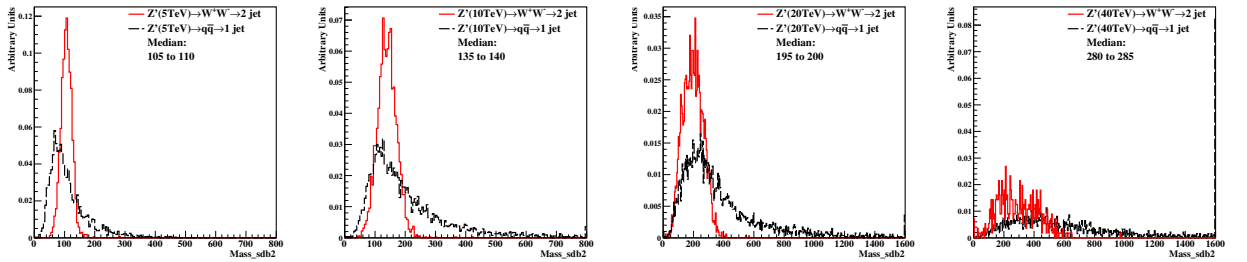(c) $M(Z') = 20$ TeV

(d) $M(Z') = 40$ TeV

Figure 19: The ROC curves of soft drop mass selection for β=0 with 5, 10, 20, and 40 TeV c.m. energies. Three different detector cell sizes are compared: 20×20, 5×5, and 1×1 (cm×cm). The signal (background) process is Z'→tt̄ (Z'→qq̄).

(a) 5 TeV at 20×20 (cm×cm) with cluster (b) 10 TeV at 20×20 (cm×cm) with cluster (c) 20 TeV at 20×20 (cm×cm) with cluster (d) 40 TeV at 20×20 (cm×cm) with cluster

(e) 5 TeV at 5×5 (cm×cm) with cluster (f) 10 TeV at 5×5 (cm×cm) with cluster (g) 20 TeV at 5×5 (cm×cm) with cluster (h) 40 TeV at 5×5 (cm×cm) with cluster

(i) 5 TeV at 1×1 (cm×cm) with cluster (j) 10 TeV at 1×1 (cm×cm) with cluster (k) 20 TeV at 1×1 (cm×cm) with cluster (l) 40 TeV at 1×1 (cm×cm) with cluster

Figure 20: Distributions of soft drop mass for β=2, with 5, 10, 20, and 40 TeV c.m. energies and three different detector cell sizes: 20×20, 5×5, and 1×1 (cm×cm). The signal (background) process is Z′→WW (Z′→qq̄).
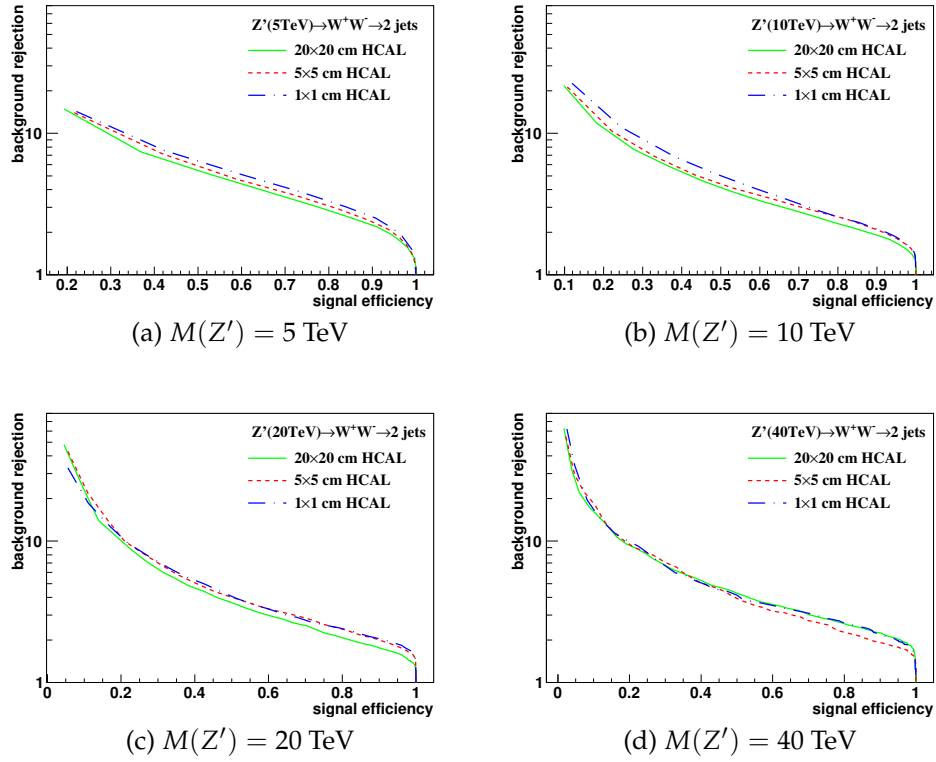
Figure 21: The ROC curves of soft drop mass selection for β=2 with 5, 10, 20, and 40 TeV c.m. energies. Three different detector cell sizes are compared: 20×20, 5×5, and 1×1 (cm×cm). The signal (background) process is Z'→WW (Z'→qq̄).

(a) 5 TeV at 20×20 (cm×cm) with cluster (b) 10 TeV at 20×20 (cm×cm) with cluster (c) 20 TeV at 20×20 (cm×cm) with cluster (d) 40 TeV at 20×20 (cm×cm) with cluster
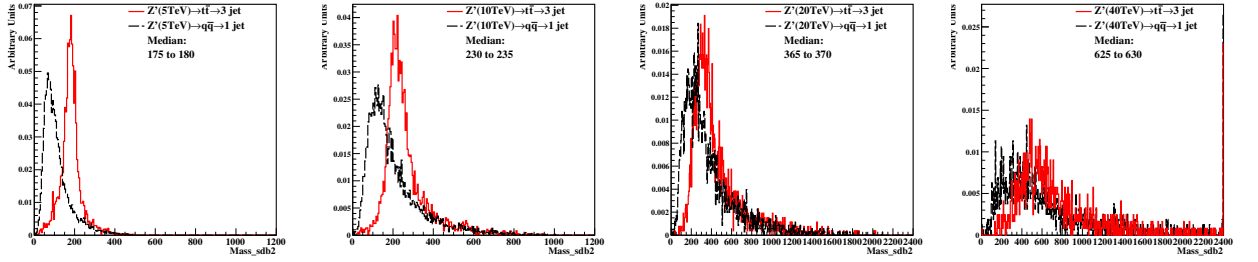
(e) 5 TeV at 5×5 (cm×cm) with cluster (f) 10 TeV at 5×5 (cm×cm) with cluster (g) 20 TeV at 5×5 (cm×cm) with cluster (h) 40 TeV at 5×5 (cm×cm) with cluster

(i) 5 TeV at 1×1 (cm×cm) with cluster (j) 10 TeV at 1×1 (cm×cm) with cluster (k) 20 TeV at 1×1 (cm×cm) with cluster (l) 40 TeV at 1×1 (cm×cm) with cluster

Figure 22: Distributions of soft drop mass for β=2, with 5, 10, 20, and 40 TeV c.m. energies and three different detector cell sizes: 20×20, 5×5, and 1×1 (cm×cm). The signal (background) process is Z'→t$\bar{\text{t}}$ (Z'→q$\bar{\text{q}}$).

(a) $M(Z') = 5$ TeV

(b) $M(Z') = 10$ TeV

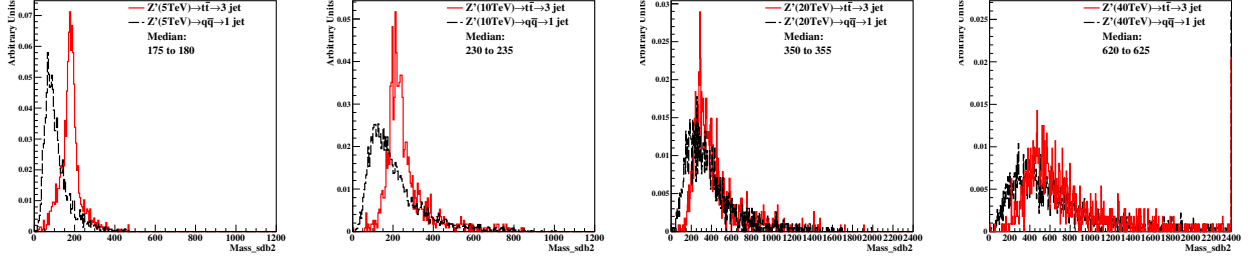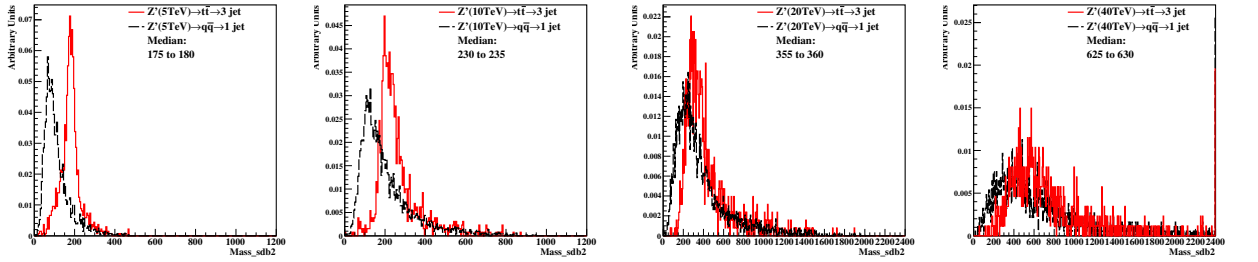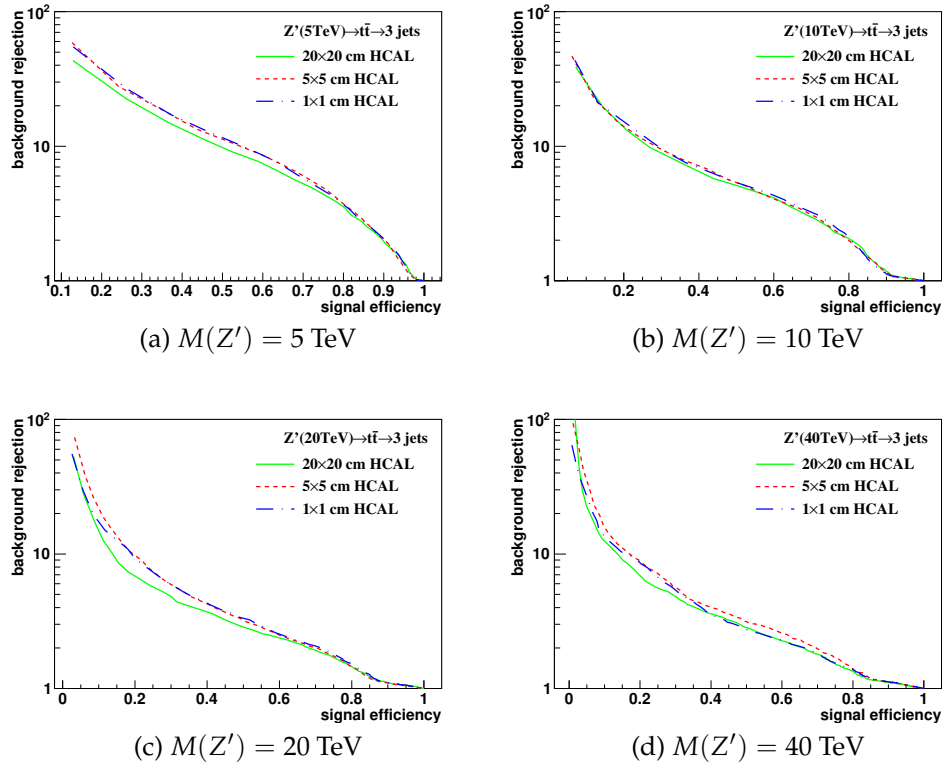(c) $M(Z') = 20$ TeV

(d) $M(Z') = 40$ TeV

Figure 23: The ROC curves of soft drop mass selection for β=2 with 5, 10, 20, and 40 TeV c.m. energies. Three different detector cell sizes are compared: 20×20, 5×5, and 1×1 (cm×cm). The signal (background) process is Z'→t$\bar{\text{t}}$ (Z'→q$\bar{\text{q}}$).

4.3   Studies of signal and background separation using jet sub-structure variables

In this section, we study different jet substructure variables and compare their ability to separate signal from background with different detector sizes and c.m. energy using ROC curves.

4.4   N-subjettiness

N-subjettiness[**??**] is the detection technique of jet substructure that is used to identify boosted hadronically-decaying objects under the high c.m. conditions. We use τ variable to distinguish the number of subjet in a fatjet to separate signal and background with different detector sizes and c.m. energy.

4.4.1   The technic of N-subjettiness

The formula and the technique are as following:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} min\{\Delta R_{1,k}, \Delta R_{2,k}, ...\Delta R_{N,k}\} \tag{5}$$

$$d_0 = \sum_k p_{T,k} R_0 \tag{6}$$

k runs over all constituent particles in the given jets (fatjet), $p_{T,k}$ are their transverse momentum, $\Delta R_{J,k} = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$ is the distance between the constituent particles k and the candidate subjet J on the $\eta - \phi$ plane. R_0 is the characteristic jet radius used in Anti-kt(AK) jet algorithm at starting. d_0 is the normalization factor.

1. First, Anti-kt(AK) algorithm is used to reconstruct jets

2. Second, after reconstructing the AK4 jets, exclusive k_T algorithm[**??**] is used in finding the jet axis in a fatjet.

3. Third, start running formula [5] and loop all constituent particles in a fatjet.

4. Finally, when finishing running all particles, it will give out τ_N, where N is positive integer.

If a fatjet has N subjet(s)[**??**], its τ_N is smaller than the τ_N of the fatjet with different number of subjets. For example, if we compare the τ_2 of one-subjet fatjet and two-subjets one, two-subjets fatjet has smaller τ_2 than one-subjet one. On the other hand, one-subjet fatjet has smaller τ_1 than two-subjets one. In the end, we can use the ratio of τ_2 and τ_1 (τ_{21}) to distinguish fatjet with one-subjet case and-two subjets case.τ_{21} is used to discriminate the fatjet shape, and it can be modified with different number of subjet.

In our study, we use τ_{21} and τ_{32} in distinguishing two-subjets fatjet and three-subjets fatjet from one-subjet fatjet individually. We want to use this two ratio values to distinguish signal from background.

4.4.2 Analysis method

First, we select the events in mass window by using SD with $\beta = 0$ and 75% signal efficiency. Then, we find the highest ratio bin to be our seed bin. Next, we compare the left and right of ratio bin, and add the higher bin to be our width. Finally, We can use this width to draw the ROC curves.
2

4.4.3 The results and conclusion

In the figure [26][28], they show the histograms of τ_{21} and τ_{32} after selecting the events. In all figures, they also include the Mann-Whitney U value in ir the legend.

As a result of figure [27][29], they perform the ROC curves of τ_{21} and τ_{32} with different detector cell sizes and c.m. energy. The smallest detector cell (1×1) doesn't have the best separation power to distinguish signal from background. Some of them have the best separation power with the bigger cell size (5×5 and 20×20).

In Figure [??](a)(b), they show the summary plots of τ_{21} and τ_{32} with the rawhit cut with 0.5GeV using Mann Whitney U test. In τ_{21}, 5TeV has better separation power when detector sizes get smaller. When energy increases, there is no improvement in the smallest detector cell size (1×1). In τ_{32}, the case is similar to τ_{21}. Even worse, with some c.m. energies, the bigger detector sizes (5×5 and 20×20) have better separation power than the smallest detector sizes (1×1).

4.5 Energy correlation function

Energy correlation function (ECF) [??] is another kind of detection technique of jet sub-sturcture that is used to distinguish the number of subjets in a fatjet under high c.m. energy conditions. This method only uses the momenta of particles and the angles between them without additional algorithm.

4.5.1 The technic of energy correlation function

The basic ECF formula is as following:

$$ECF(N,\beta) = \sum_{i_1 < i_2 < .... < i_N \in J} \left(\prod_{a=1}^{N} E_{ia}\right)\left(\prod_{b=1}^{N-1} \prod_{c=b+1}^{N} \theta_{i_b i_c}\right)^{\beta} \tag{7}$$

In the formula 7, the sum loop all particles in the jet J, E are the energy of particles, and θ are the angles between the particles.

We apply two approximation. First, because under the high energy limitation $p >> m$, $E \approx p$. Second, we use Radius R between particles naturally, so our ECF formula (7) can be modified to the formula:

$$ECF(N,\beta) = \sum_{i_1<i_2<....<i_N\in J} (\prod_{a=1}^{N} P_{ia})(\prod_{b=1}^{N-1} \prod_{c=b+1}^{N} R_{i_b i_c})^{\beta} \tag{8}$$

From the modified ECF formula (8), in order to use the dimensionless observation to determine whether the number of subjets in system, parameter τ_N is defined as:

$$\tau_N^{(\beta)} \equiv \frac{ECF(N+1,\beta)}{ECF(N,\beta)} \tag{9}$$

The idea of formula (9) is from N-subjetness, because the behavior of it is very similar to N-subjetness as reference [**??**]. In general, if the system has N subjets, $ECF(N+1,\beta)$ should be significantly smaller than $ECF(N,\beta)$, so we can use this advantage to distinguish different number of subjets. FInally, because it is suggested to used τ_{21}, τ_{32} [**??**] to distinguish two-subjets fatjet and three-subjets fatjet from one-subjet fatjet, in the ECF, it also defines the ratio of τ there, and define the energy correlation double ratio that is used in our study:

$$C_N^{(\beta)} \equiv \frac{\tau_N^{(\beta)}}{\tau_{N-1}^{(\beta)}} = \frac{ECF(N-1,\beta)ECF(N+1,\beta)}{ECF(N,\beta)^2} \tag{10}$$

We set N=2 and $\beta = 1$ (C_2^1) to distinguish two-subjets fatjet from one-subjet fatjet.

4.5.2 Analysis method

First, we select the events in mass window by using SD with $\beta = 0$ and 75% signal efficiency. Then, we find the highest ratio bin to be our seed bin. Next, we compare the left and right of ratio bin, and add the higher bin to be our width. Finally, We can use this width to draw the ROC curves.

4.5.3 The results and conclusion

In the figure [24], they show the histograms of τ_{21} and τ_{32} after selecting the events. In all figures, they also include the Mann-Whitney U value in their legend

As a result of figure [25], they perform the ROC curves of C_2^1 with different detector cell sizes and c.m. energy. The smallest detector cell (1×1) doesn't have the best separation power to distinguish signal from background. In addition, in some cases such like (a), the biggest one (20×20) has the best distinguish power under the same c.m. energy.

In Figure [**??**](c), it shows the summary plots with the rawhit cut with 0.5GeV using Mann Whitney U test. For conclusion, all separation power aren't improved by small-

(a) 5TeV at 20×20(cm×cm) 0.5GeV

(b) 10TeV at in20×20(cm×cm) in 0.5GeV

(c) 20TeV at 20×20(cm×cm) in 0.5GeV

(d) 40TeV at 20×20(cm×cm) in 0.5GeV

(e) 5TeV at 5×5(cm×cm) in 0.5GeV

(f) 10TeV at 5×5(cm×cm) in 0.5GeV

(g) 20TeV at 5×5(cm×cm) in 0.5GeV

(h) 40TeV at 5×5(cm×cm) in 0.5GeV

(i) 5TeV at 1×1(cm×cm) in 0.5GeV

(j) 10TeV at 1×1(cm×cm) in 0.5GeV

(k) 20TeV at 1×1(cm×cm) in 0.5GeV

(l) 40TeV at 1×1(cm×cm) in 0.5GeV

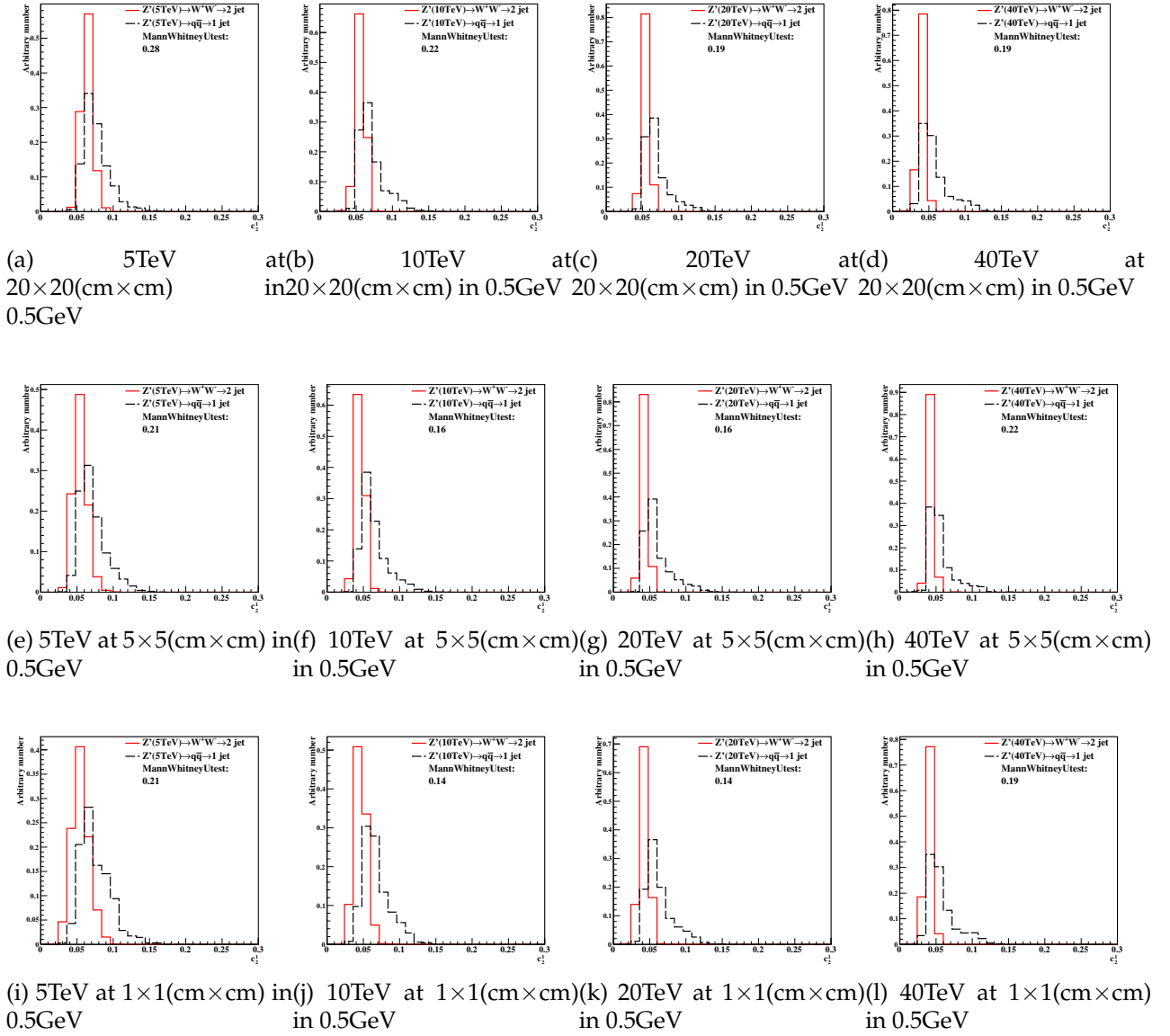Figure 24: Distributions of Mann-Whitney value U in 5, 10, 20, 40TeV energy collision for c2b1 in different detector sizes. Cell Size in 20×20, 5×5, and 1×1(cm×cm) are shown here.

est cell size (1×1).

References

[1] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

(a) 5 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

(b) 10 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

(c) 20 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

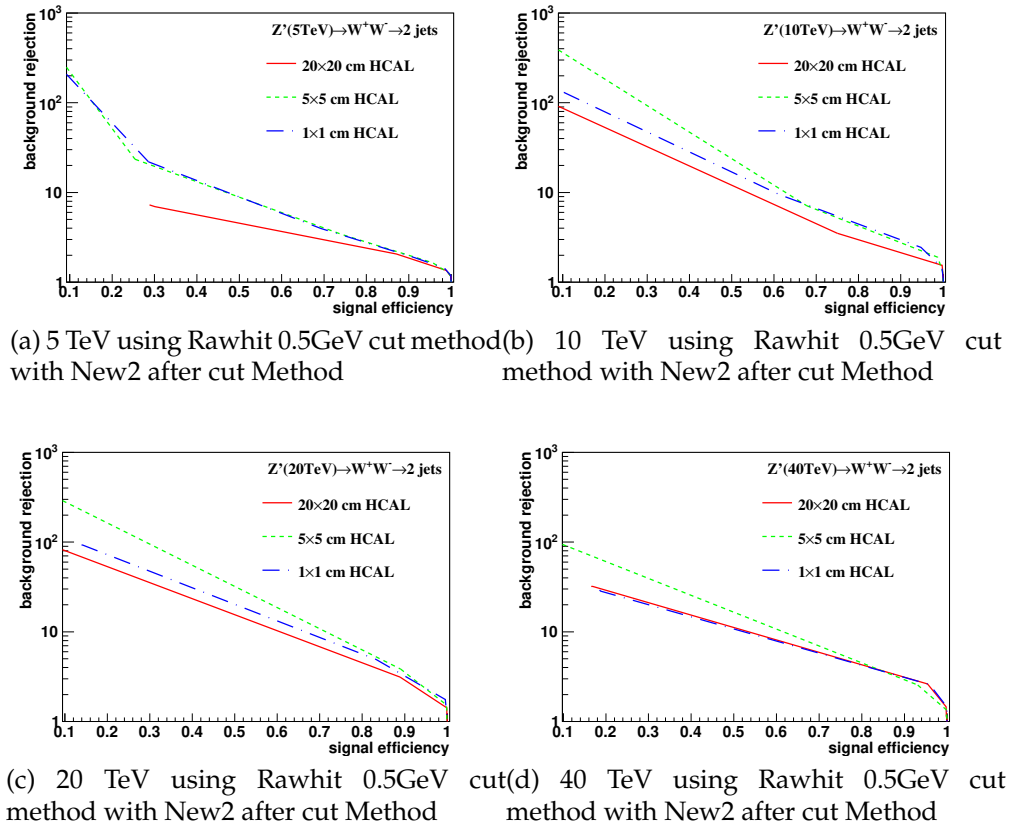(d) 40 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

Figure 25: Signal efficiency versus background rejection rate using c2b1.The energies of collision at (a)5, (b)10, (c)20, (d)40TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

(a) 5TeV at 20×20(cm×cm) 0.5GeV

(b) 10TeV at in20×20(cm×cm) in 0.5GeV

(c) 20TeV at 20×20(cm×cm) in 0.5GeV

(d) 40TeV at 20×20(cm×cm) in 0.5GeV

(e) 5TeV at 5×5(cm×cm) in 0.5GeV

(f) 10TeV at 5×5(cm×cm) in 0.5GeV

(g) 20TeV at 5×5(cm×cm) in 0.5GeV

(h) 40TeV at 5×5(cm×cm) in 0.5GeV

(i) 5TeV at 1×1(cm×cm) in 0.5GeV

(j) 10TeV at 1×1(cm×cm) in 0.5GeV

(k) 20TeV at 1×1(cm×cm) in 0.5GeV

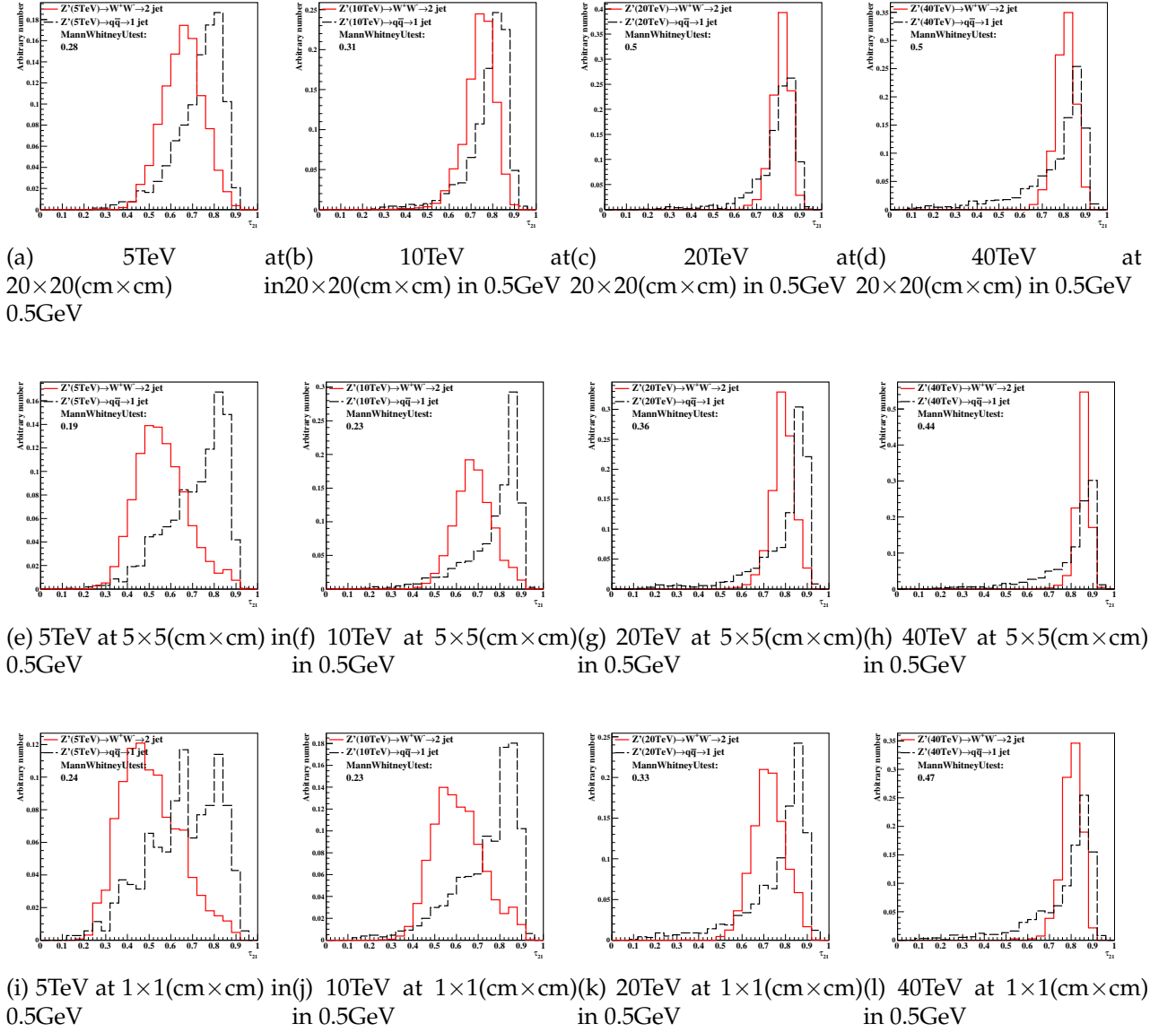(l) 40TeV at 1×1(cm×cm) in 0.5GeV

Figure 26: Distributions of Mann-Whitney value U in 5, 10, 20, 40TeV energy collision for τ_{21} in different detector sizes. Cell Size in 20×20, 5×5, and 1×1(cm×cm) are shown here.

(a) 5 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

(b) 10 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

(c) 20 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

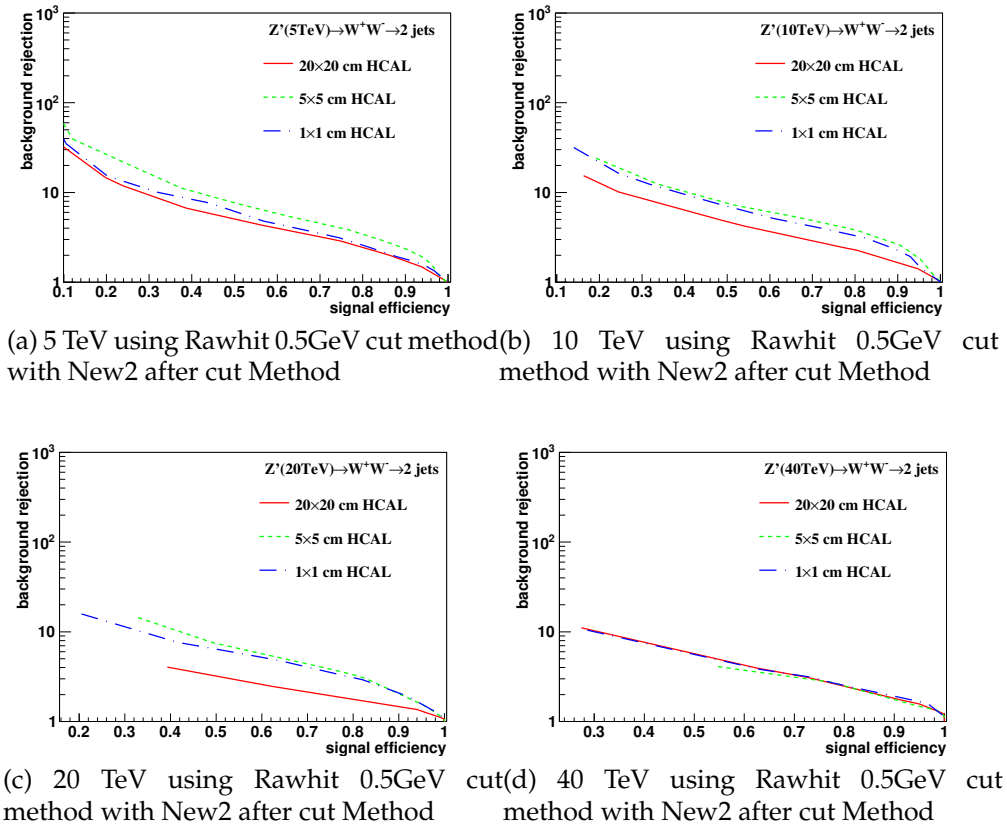(d) 40 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

Figure 27: Signal efficiency versus background rejection rate using τ_{21}. The energies of collision at (a)5, (b)10, (c)20, (d)40TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

(a) 5TeV at 20×20(cm×cm) 0.5GeV

(b) 10TeV at in20×20(cm×cm) in 0.5GeV

(c) 20TeV at 20×20(cm×cm) in 0.5GeV

(d) 40TeV at 20×20(cm×cm) in 0.5GeV

(e) 5TeV at 5×5(cm×cm) in 0.5GeV

(f) 10TeV at 5×5(cm×cm) in 0.5GeV

(g) 20TeV at 5×5(cm×cm) in 0.5GeV

(h) 40TeV at 5×5(cm×cm) in 0.5GeV

(i) 5TeV at 1×1(cm×cm) in 0.5GeV

(j) 10TeV at 1×1(cm×cm) in 0.5GeV

(k) 20TeV at 1×1(cm×cm) in 0.5GeV
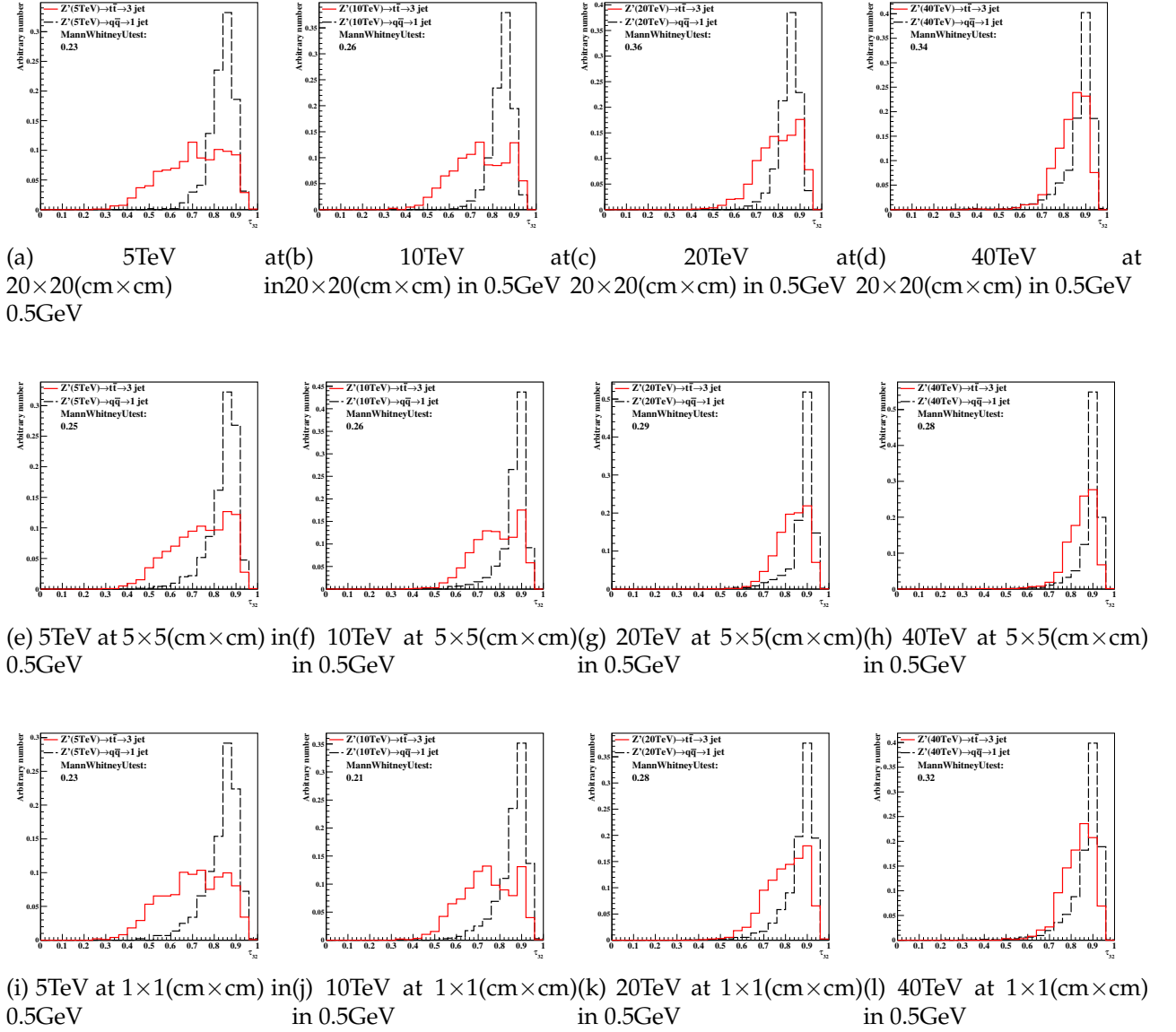
(l) 40TeV at 1×1(cm×cm) in 0.5GeV

Figure 28: Distributions of Mann-Whitney value U in 5, 10, 20, 40TeV energy collision for τ_{32} in different detector sizes. Cell Size in 20×20, 5×5, and 1×1(cm×cm) are shown here.
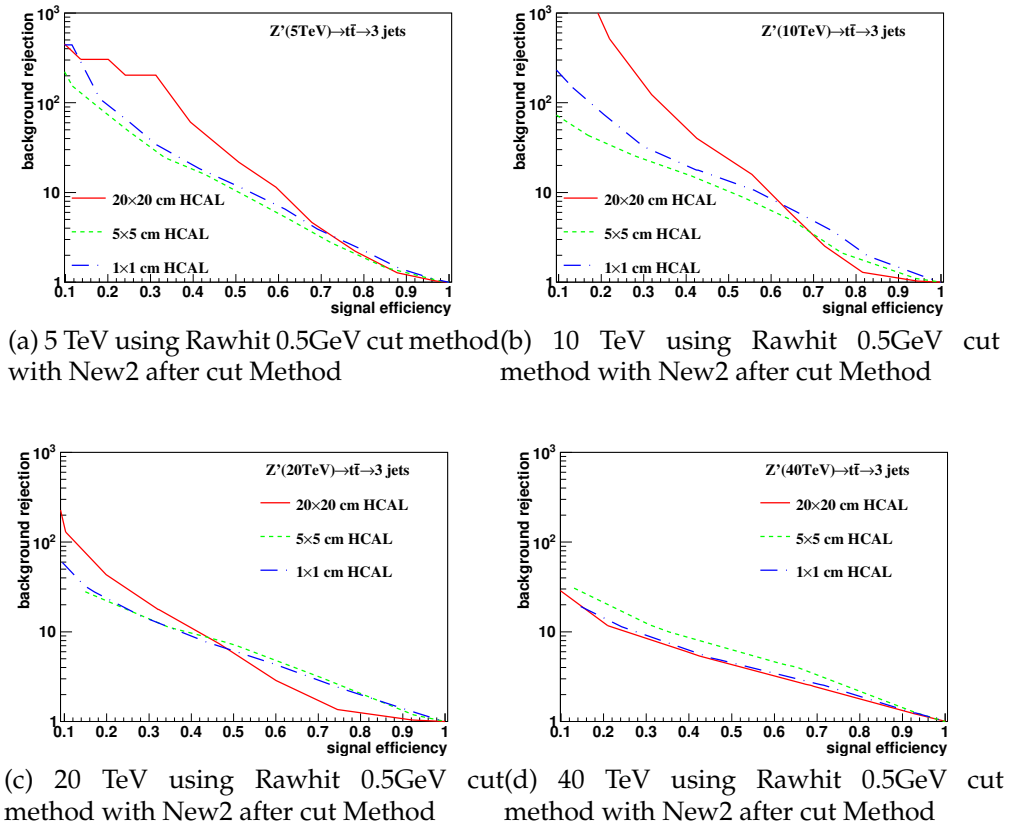
(a) 5 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

(b) 10 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

(c) 20 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

(d) 40 TeV using Rawhit 0.5GeV cut method with New2 after cut Method

Figure 29: Signal efficiency versus background rejection rate using τ_{32}. The energies of collision at (a)5, (b)10, (c)20, (d)40TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

[2] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, `doi:10.1103/PhysRevLett.13.508`.

[3] R. Contino et al., "Physics at a 100 TeV pp collider: Higgs and EW symmetry breaking studies ", `arXiv:1606.09408`. CERN-TH-2016-113.

[4] M. L. Mangano et al., "Physics at a 100 TeV pp collider: Standard Model processes ", `arXiv:1607.01831`. CERN-TH-2016-112, FERMILAB-FN-1021-T.

[5] M. Benedikt, "The Global Future Circular Colliders Effort",. CERN-ACC-SLIDES-2016-0016. Presented at P5 Workshop on the Future of High Energy Physics, BNL, USA, Dec. 15-18, 2013.

[6] J. Tang et al., "Concept for a Future Super Proton-Proton Collider", 2015.

[7] H. Aihara et al., "SiD Letter of Intent", 2009. `arXiv:0911.0006`. Presented to ILC IDAG.

[8] C. Adolphsen et al., "The International Linear Collider Technical Design Report - Volume 3. II: Accelerator Baseline Design", 2013. `arXiv:1306.6328`.

[9] H. Abramowicz et al., "The International Linear Collider Technical Design Report - Volume 4: Detectors", 2013. `arXiv:1306.6329`.

[10] S. V. Chekanov et al., "Initial performance studies of a general-purpose detector for multi-TeV physics at a 100 TeV pp collider", *JINST* **12** (2017), no. 06, P06009, `doi:10.1088/1748-0221/12/06/P06009`, `arXiv:1612.07291`.

[11] Z. Li, "Radiation hardness / tolerance of Si sensors / detectors for nuclear and high energy physics experiments", in *Semiconductor pixel detectors for particles and X-rays. Proceedings, International Workshop, PIXEL2002, Carmel, USA, September 9-12, 2002*. 2002.

[12] J. Allison et al., "Recent developments in Geant4", *Nuclear Instruments and Methods in Physics Research A* **835** (2016) 186.

[13] S. Chekanov, "HepSim: a repository with predictions for high-energy physics experiments", *Advances in High Energy Physics* **2015** (2015) 136093. Available as `http://atlaswww.hep.anl.gov/hepsim/`.

[14] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop", *JHEP* **05** (2014) 146, `doi:10.1007/JHEP05(2014)146`, `arXiv:1402.2657`.

[15] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", *JHEP* **08** (1997) 001, `doi:10.1088/1126-6708/1997/08/001`, `arXiv:hep-ph/9707323`.

[16] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering", in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*. 1998. `arXiv:hep-ph/9907280`.