

1. Studies of signal and background separation using Mann-Whitney U test

In this section, we study different jet substructure variables and compare their ability to separate the signal and background for different detector cell sizes using the Mann-Whitney U test.

By the definition of the Mann-Whitney U test, if the value of U is close to 0.5, it means that the signal and background distributions have almost identical shapes, i.e. the separation power of the variable is bad. On the other hand, if the U value is close to 0, it means that the distributions of the signal and the background are very different from each other and the separation power of the variable is great.

Figure 6 shows the representative sample of the distributions for τ_{21}, τ_{32} , in different detector sizes with their corresponding U value. In τ_{21} , the separation power is better when the detector size is smaller. However, the separation power of tau32 does not improve when the detector size gets smaller.

Figure 7 shows the summary plots of the clustering in Mann-Whitney U test for the three different variables. In τ_{21} , 5 TeV has the better separation power when the detector size is smaller. However, there is not much improvement in higher energy collisions. In τ_{32} , 5 TeV has also better separation power on smaller detector size but higher energy collisions seem to have better separation power when the detector size is bigger. The $c_2^{(1)}$ does not seem to have any significant improvement in its separation power as the detector size gets smaller for all energy collision. Nevertheless, the U values of the $c_2^{(1)}$ are better than the τ_{21} and τ_{32} . In conclusion, the $c_2^{(1)}$ variable is the best parameter as its separation power is better than the other variable and does not have a significant improvement in higher energy collision.

Figure 8 shows the summary plots of the rawhit cut at 0.5 GeV in Mann Whitney U test for the three different variables. In τ_{21} , 5 and 10 TeV have better separation power on smaller detector sizes. However, there is no significant improvement in the separation power of higher energy collisions. The τ_{32} does not have any significant improvement in its separation power as the detector size gets smaller for all energy collision. Lastly, there is a slight improvement in separation power for 5, 10, 20, 40, and 20 TeV energy collision in $c_2^{(1)}$. At 40 TeV, no significant improvement is observed. In conclusion, the $c_2^{(1)}$ variable is the best parameter as it has the best separation power at 40 TeV energy collision than the other variables.

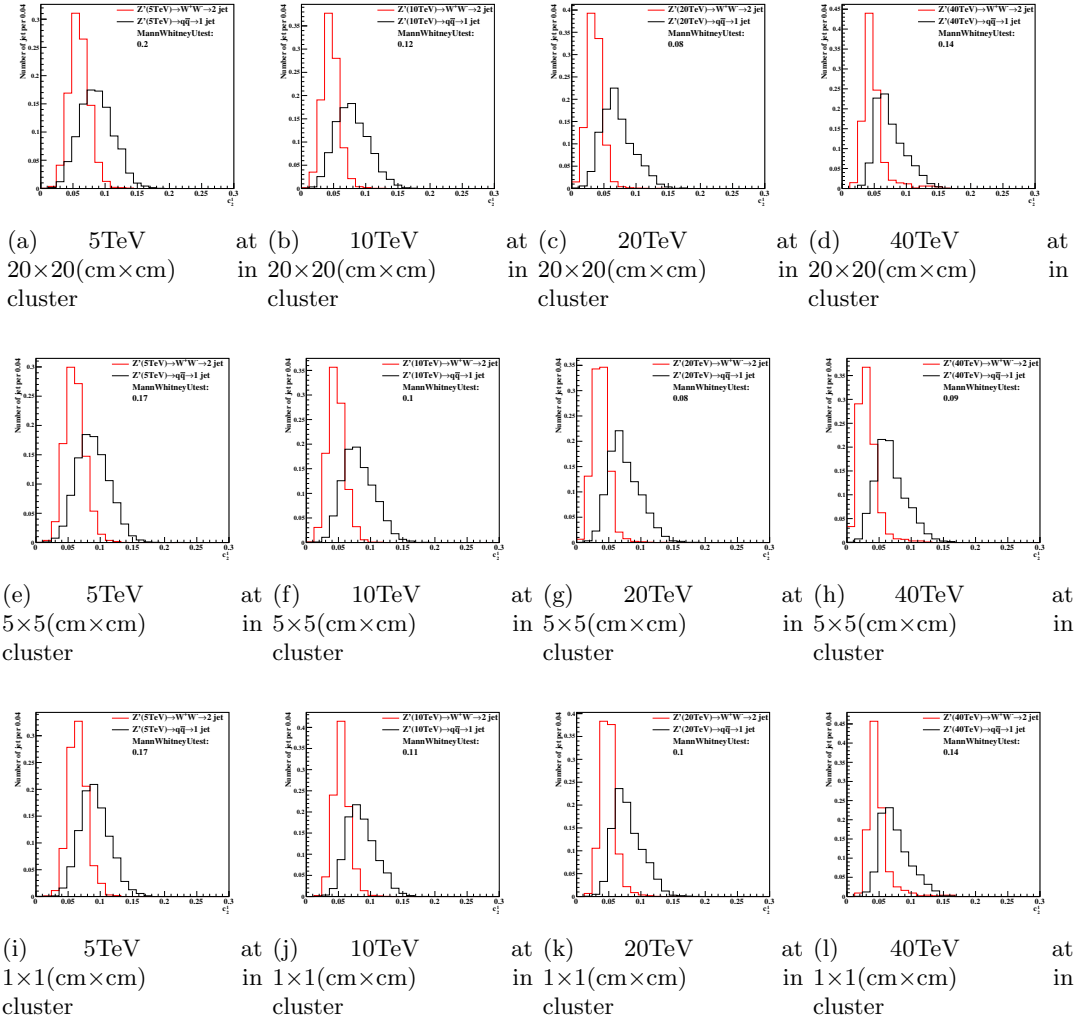
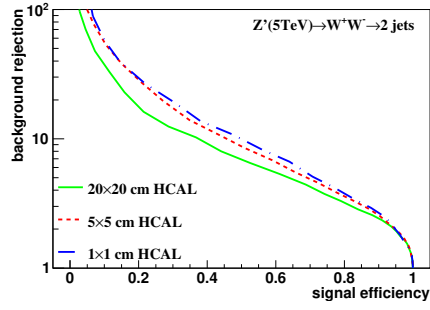
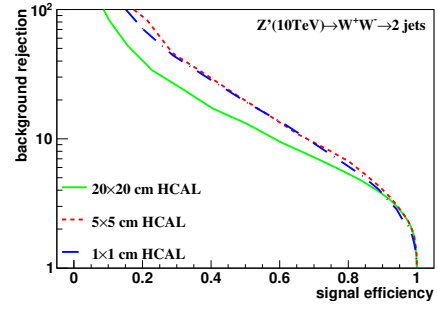


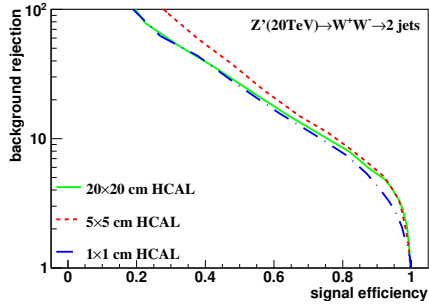
Figure 1: Distributions of Mann-Whitney value U in 5, 10, 20, 40 TeV energy collision for $c_2^{(1)}$ in different detector sizes. Cell Size in 20×20 , 5×5 , and $1 \times 1(\text{cm} \times \text{cm})$ are shown here.



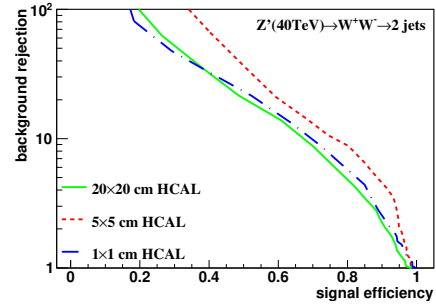
(a) 5 TeV in cluster



(b) 10 TeV in cluster



(c) 20 TeV in cluster



(d) 40 TeV in cluster

Figure 2: Signal efficiency versus background rejection rate using $c_2^{(1)}$. 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.

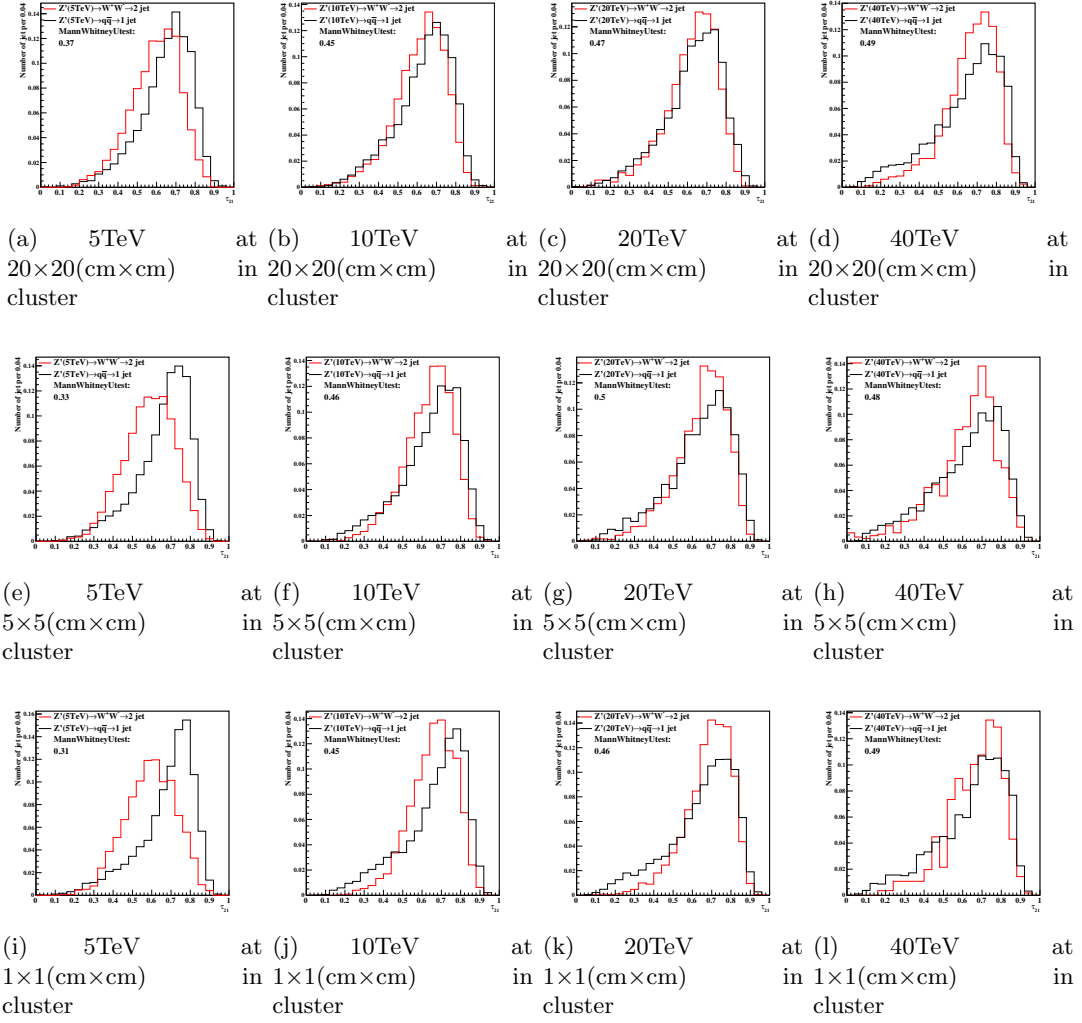


Figure 3: Distributions of Mann-Whitney value U in 5, 10, 20, 40TeV energy collision for τ_{21} in different detector sizes. Cell Size in 20x20, 5x5, and 1x1(cm x cm) are shown here.

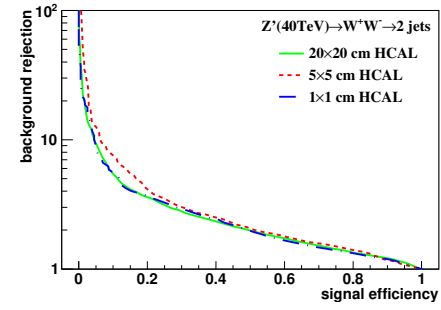
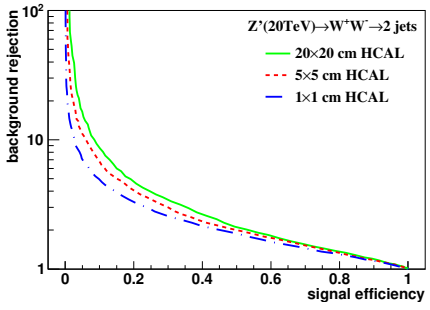
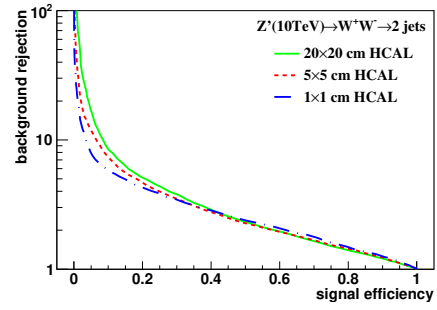
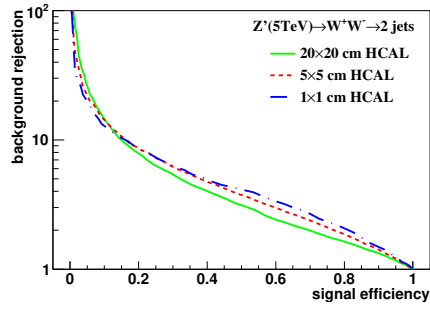


Figure 4: Signal efficiency versus background rejection rate using τ_{21} . The energies of collision at (a) 5, (b) 10, (c) 20, (d) 40 TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

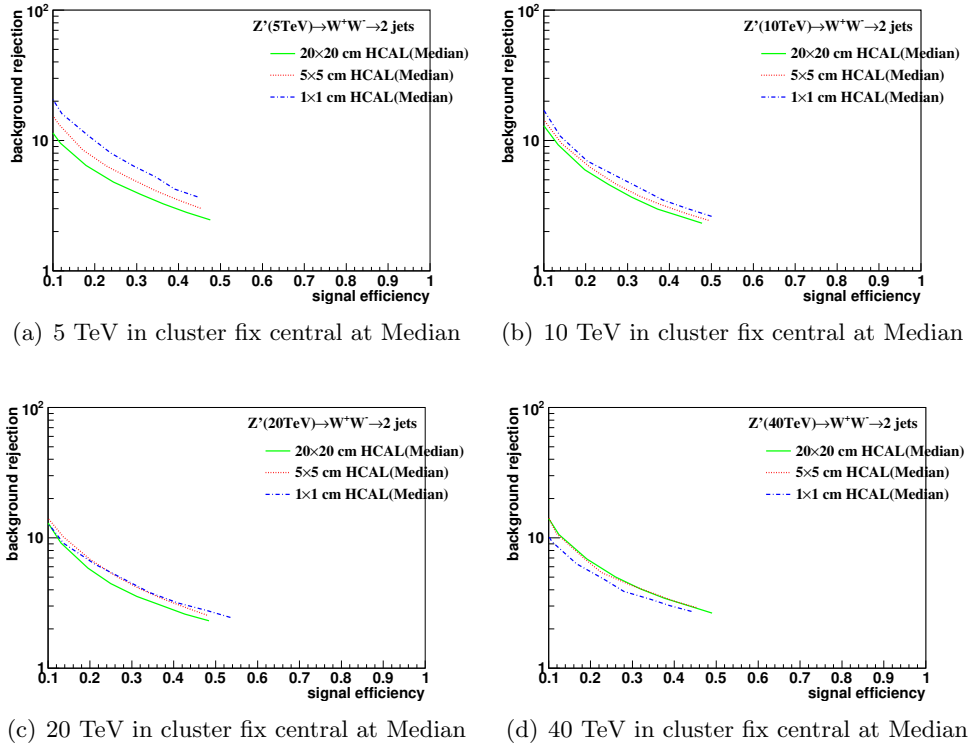


Figure 5: Signal efficiency versus background rejection rate using τ_{21} . The energies of collision at (a) 5, (b) 10, (c) 20, (d) 40 TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

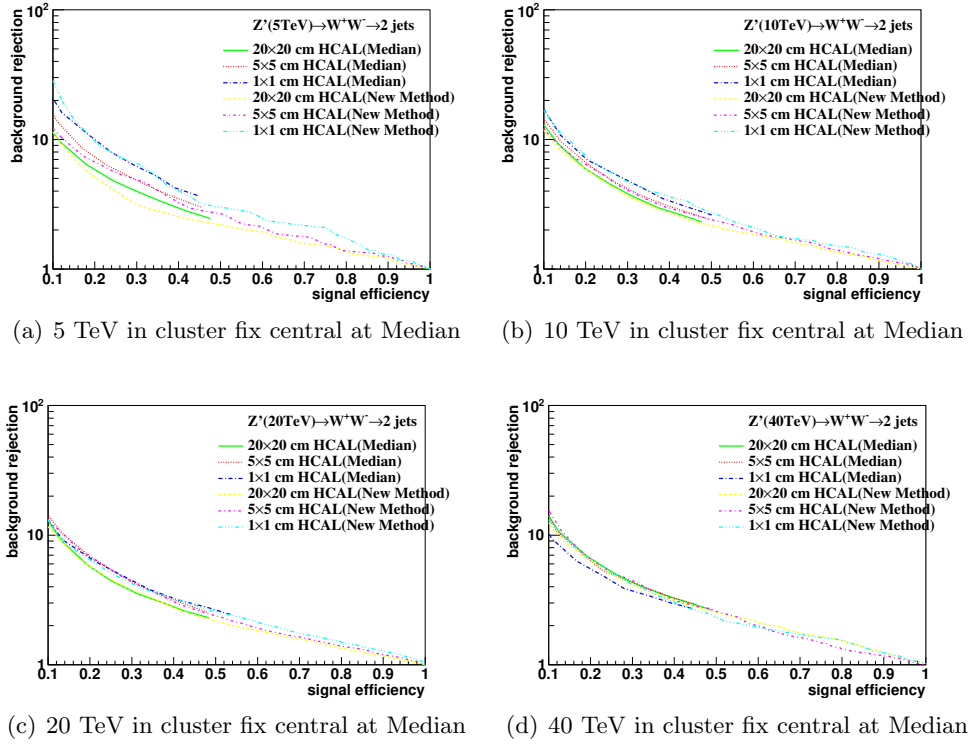


Figure 6: Signal efficiency versus background rejection rate using τ_{21} . The energies of collision at (a)5, (b)10, (c)20, (d)40TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

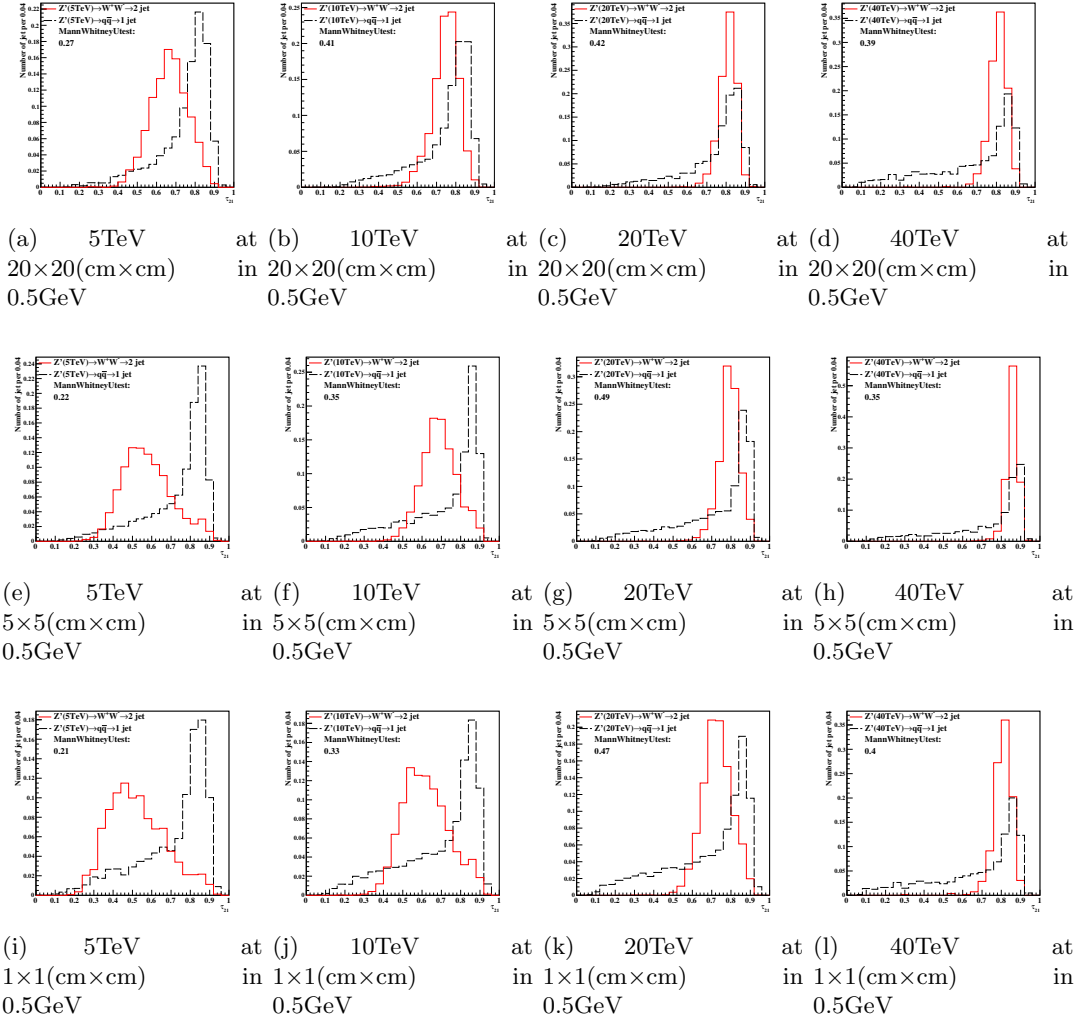
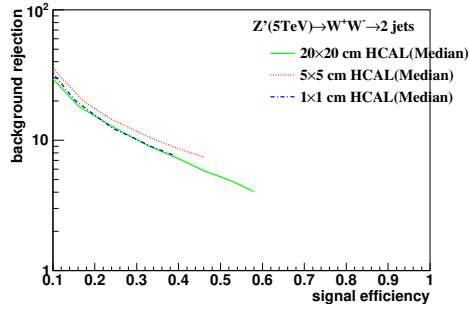
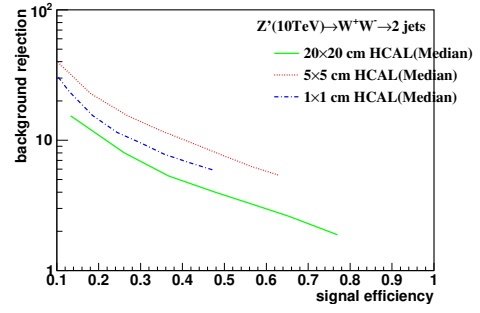


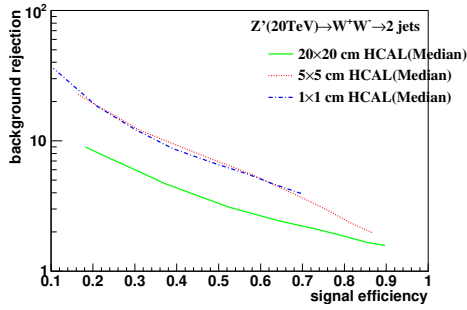
Figure 7: Distributions of Mann-Whitney value U in 5, 10, 20, 40TeV energy collision for τ_{21} in different detector sizes. Cell Size in 20x20, 5x5, and 1x1(cm x cm) are shown here.



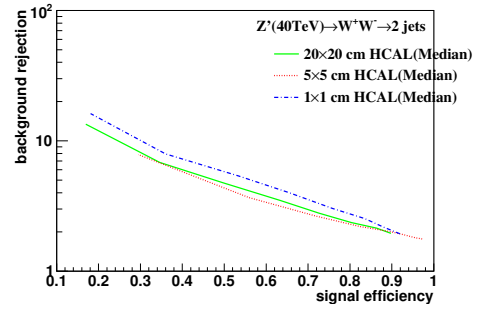
(a) 5 TeV rawhit cut at 0.5GeV fix central at Median



(b) 10 TeV rawhit cut at 0.5GeV fix central at Median

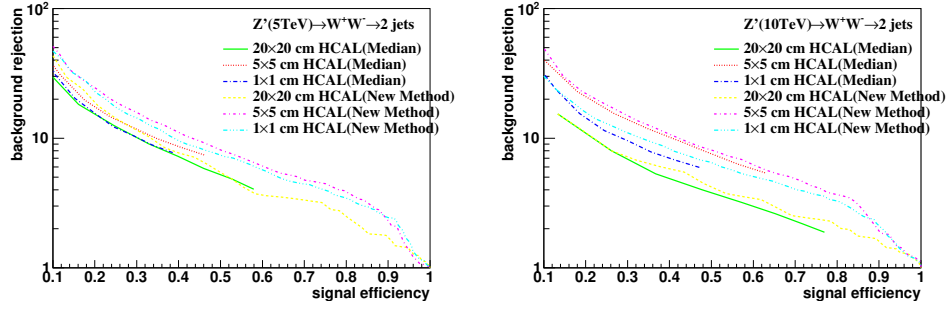


(c) 20 TeV rawhit cut at 0.5GeV fix central at Median

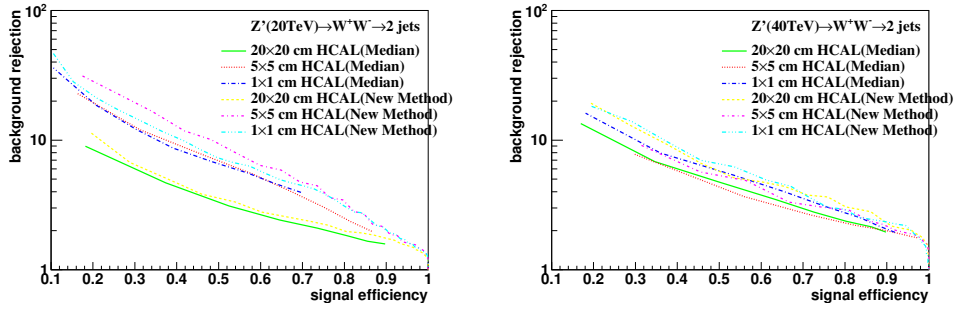


(d) 40 TeV rawhit cut at 0.5GeV fix central at Median

Figure 8: 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) 5 TeV rawhit cut at 0.5GeV fix central at Median (b) 10 TeV rawhit cut at 0.5GeV fix central at Median



(c) 20 TeV rawhit cut at 0.5GeV fix central at Median (d) 40 TeV rawhit cut at 0.5GeV fix central at Median

Figure 9: Signal efficiency versus background rejection rate using τ_{21} . The energies of collision at (a) 5, (b) 10, (c) 20, (d) 40 TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

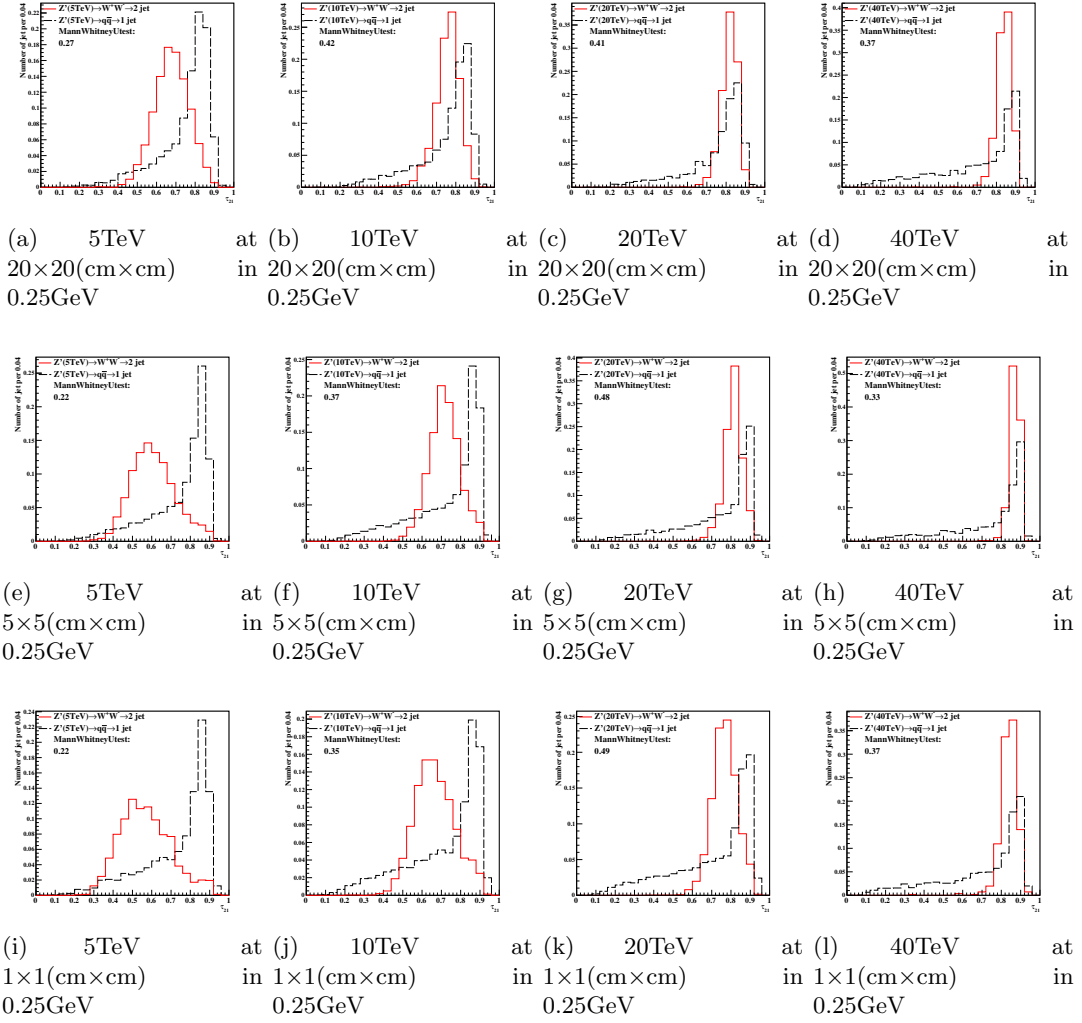
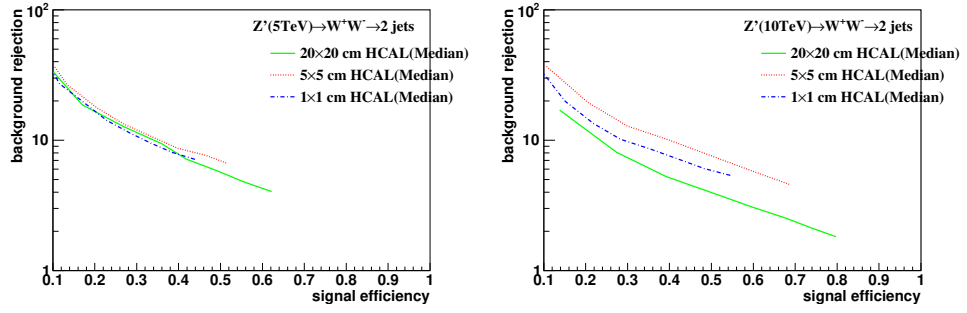
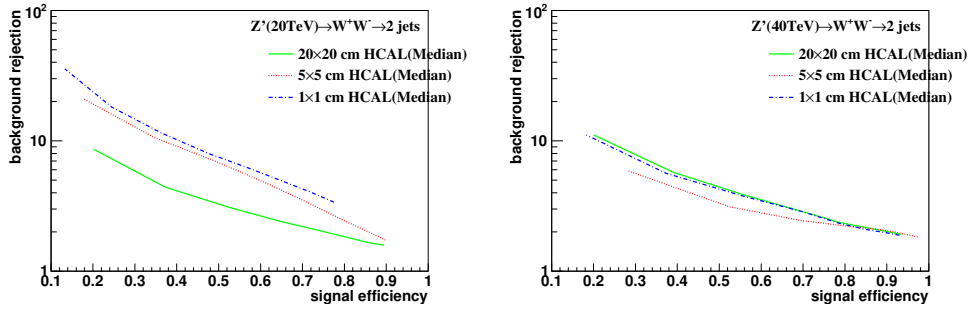


Figure 10: Distributions of Mann-Whitney value U in 5, 10, 20, 40TeV energy collision for τ_{21} in different detector sizes. Cell Size in 20x20, 5x5, and 1x1(cm \times cm) are shown here.

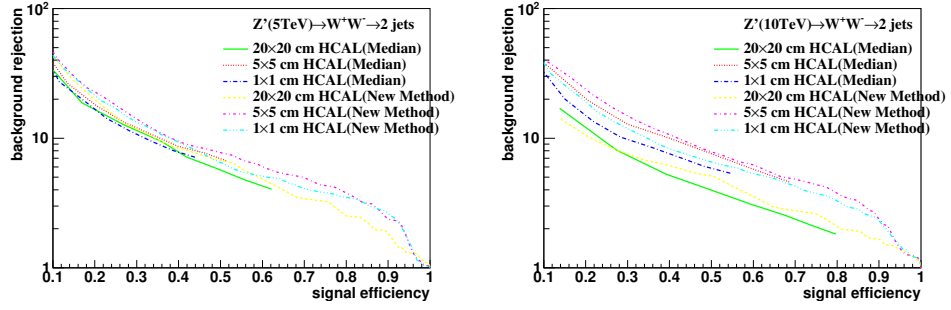


(a) 5 TeV rawhit cut at 0.25GeV fix central at Median (b) 10 TeV rawhit cut at 0.25GeV fix central at Median

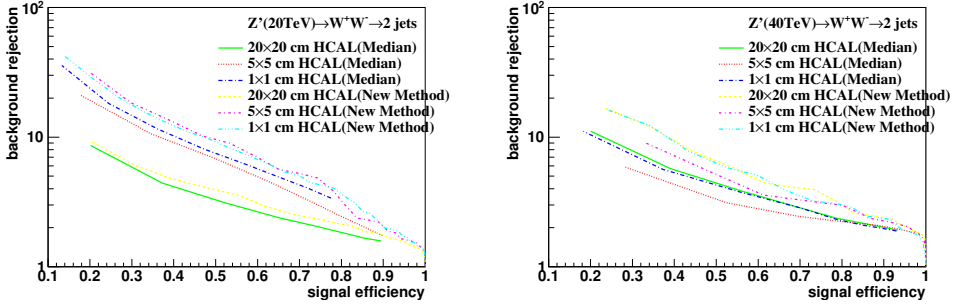


(c) 20 TeV rawhit cut at 0.25GeV fix central at Median (d) 40 TeV rawhit cut at 0.25GeV fix central at Median

Figure 11: 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) 5 TeV rawhit cut at 0.25GeV fix central at Median (b) 10 TeV rawhit cut at 0.25GeV fix central at Median



(c) 20 TeV rawhit cut at 0.25GeV fix central at Median (d) 40 TeV rawhit cut at 0.25GeV fix central at Median

Figure 12: Signal efficiency versus background rejection rate using τ_{21} . The energies of collision at (a) 5, (b) 10, (c) 20, (d) 40 TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

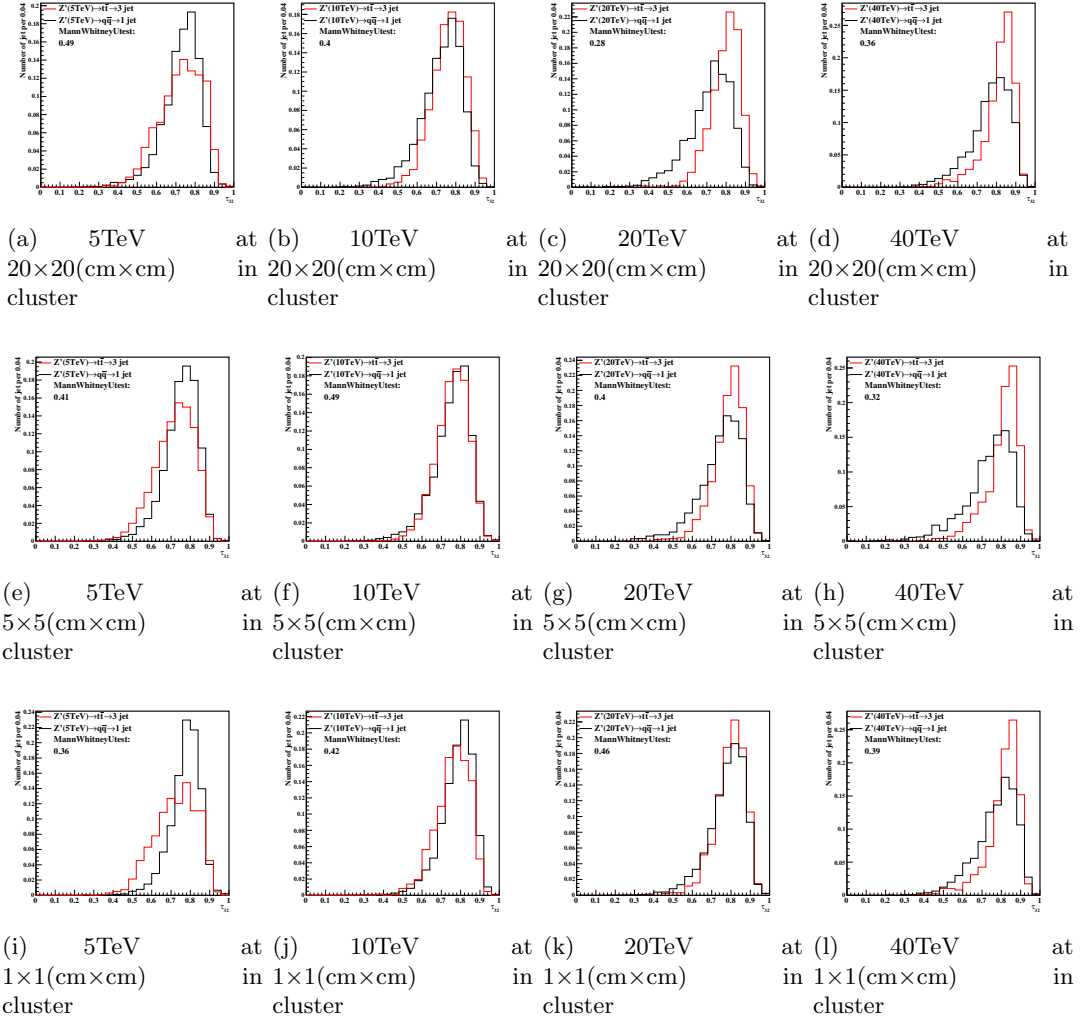
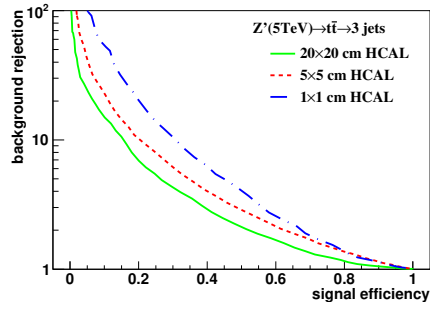
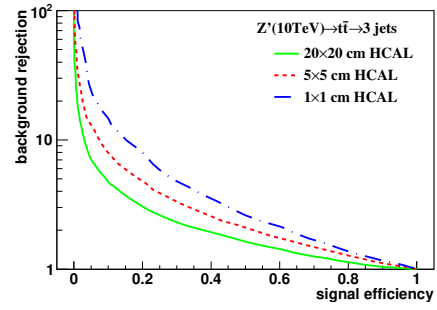


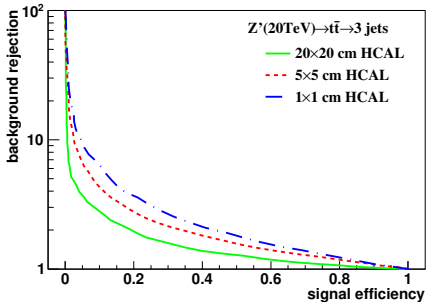
Figure 13: Distributions of Mann-Whitney value U in 5, 10, 20, 40 TeV energy collision for τ_{32} in different detector sizes. Cell Size in 20x20, 5x5, and 1x1(cm x cm) are shown here.



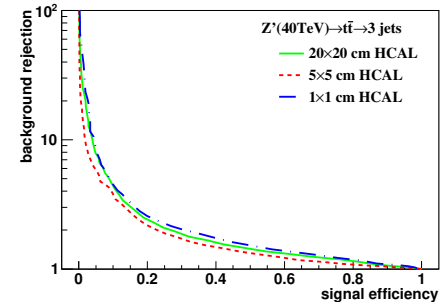
(a) 5 TeV in cluster



(b) 10 TeV in cluster



(c) 20 TeV in cluster



(d) 40 TeV in cluster

Figure 14: Signal efficiency versus background rejection rate using τ_{32} . The energies of collision at (a) 5, (b) 10, (c) 20, (d) 40 TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

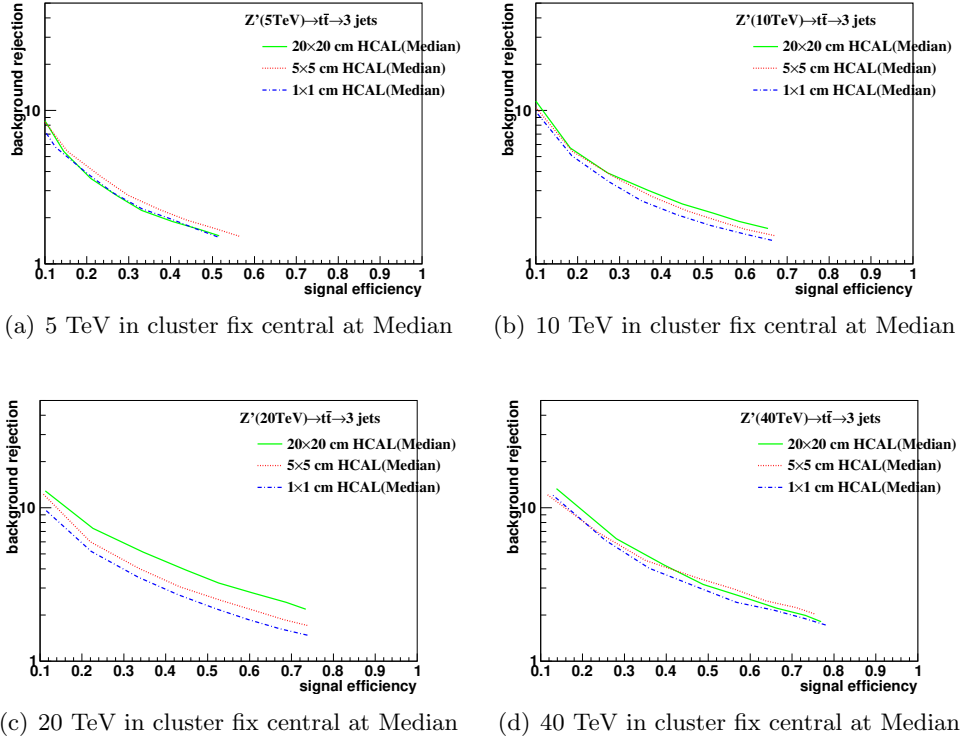


Figure 15: Signal efficiency versus background rejection rate using τ_{32} . The energies of collision at (a) 5, (b) 10, (c) 20, (d) 40 TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

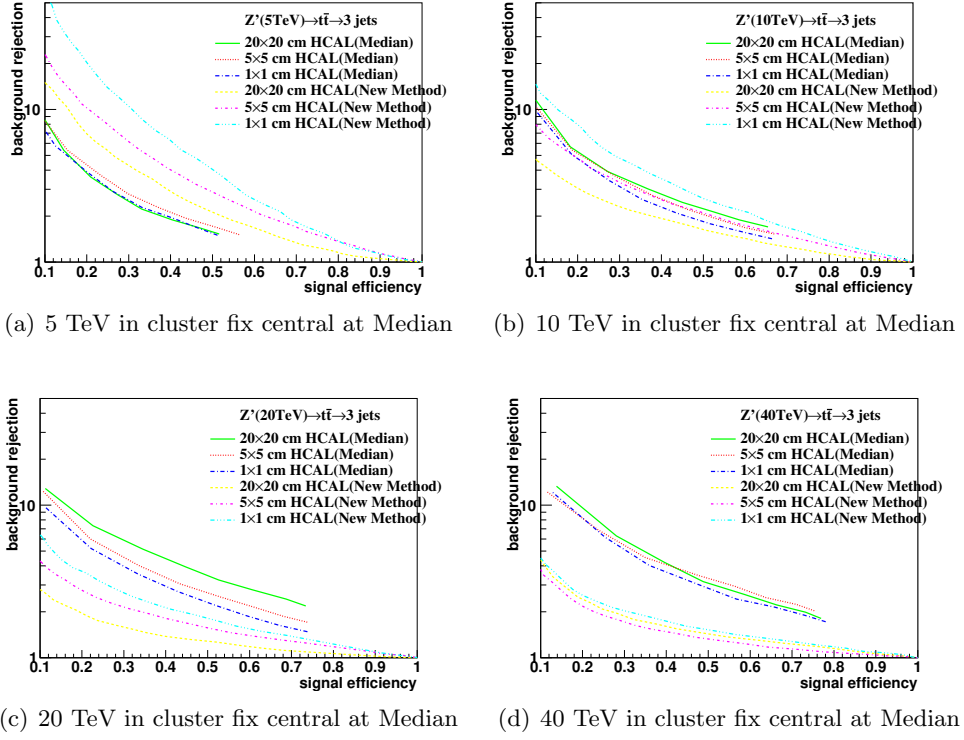


Figure 16: Signal efficiency versus background rejection rate using τ_{32} . The energies of collision at (a) 5, (b) 10, (c) 20, (d) 40 TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

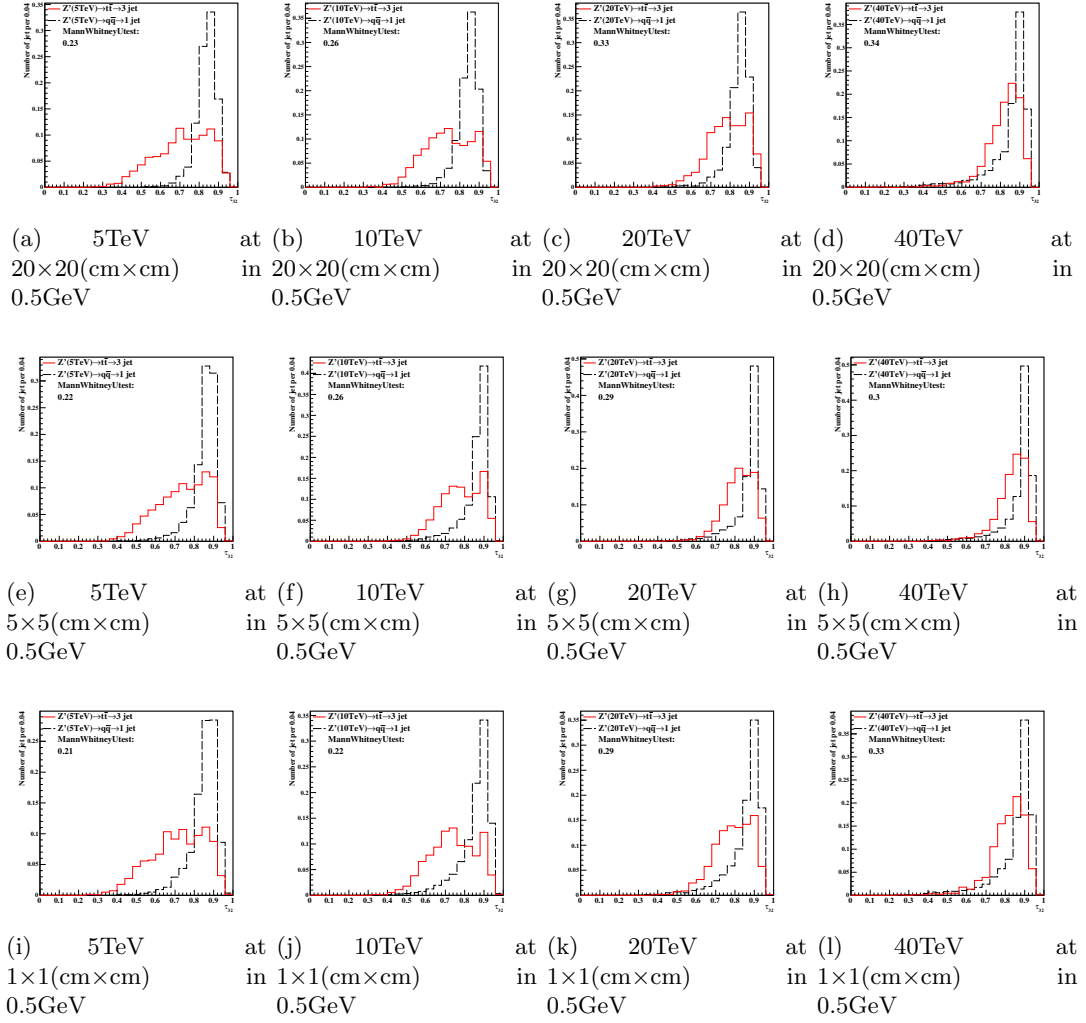
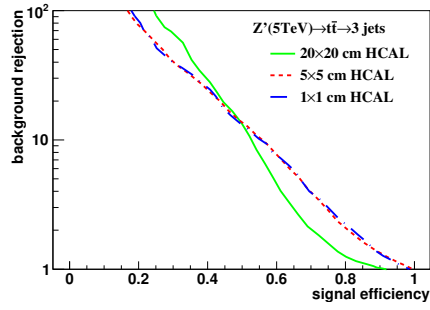
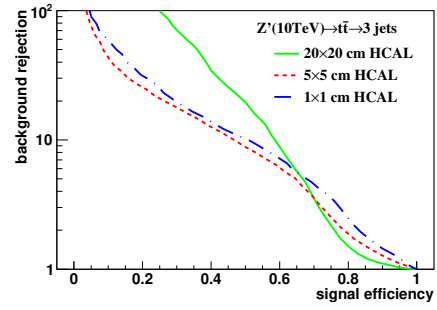


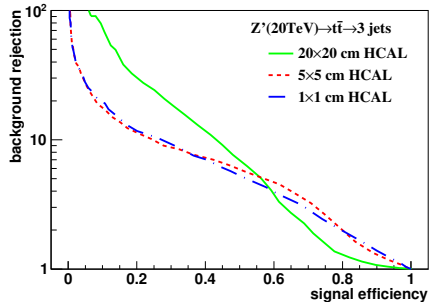
Figure 17: Distributions of Mann-Whitney value U in 5, 10, 20, 40 TeV energy collision for τ_{32} in different detector sizes. Cell Size in 20x20, 5x5, and 1x1(cm x cm) are shown here.



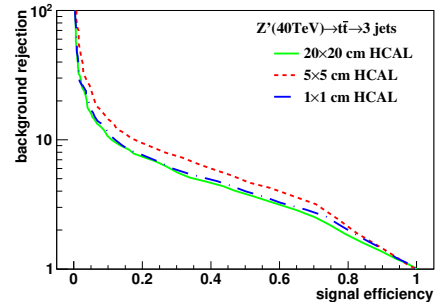
(a) 5 TeV rawhit cut at 0.5GeV



(b) 10 TeV rawhit cut at 0.5GeV



(c) 20 TeV rawhit cut at 0.5GeV



(d) 40 TeV rawhit cut at 0.5GeV

Figure 18: 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.

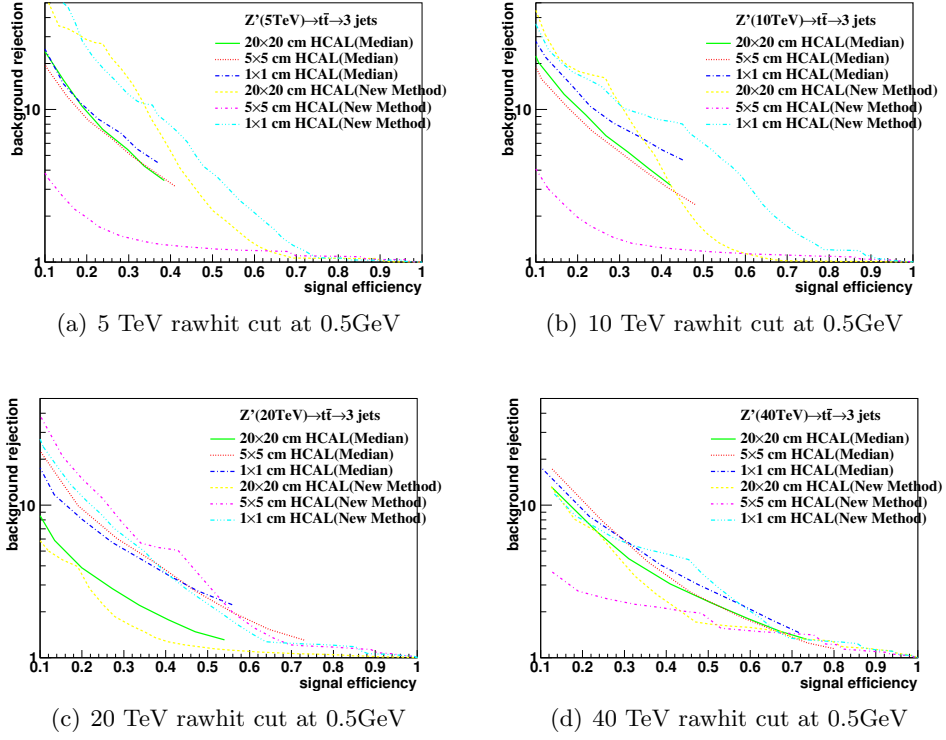


Figure 19: Signal efficiency versus background rejection rate using τ_{32} . The energies of collision at (a) 5, (b) 10, (c) 20, (d) 40 TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.

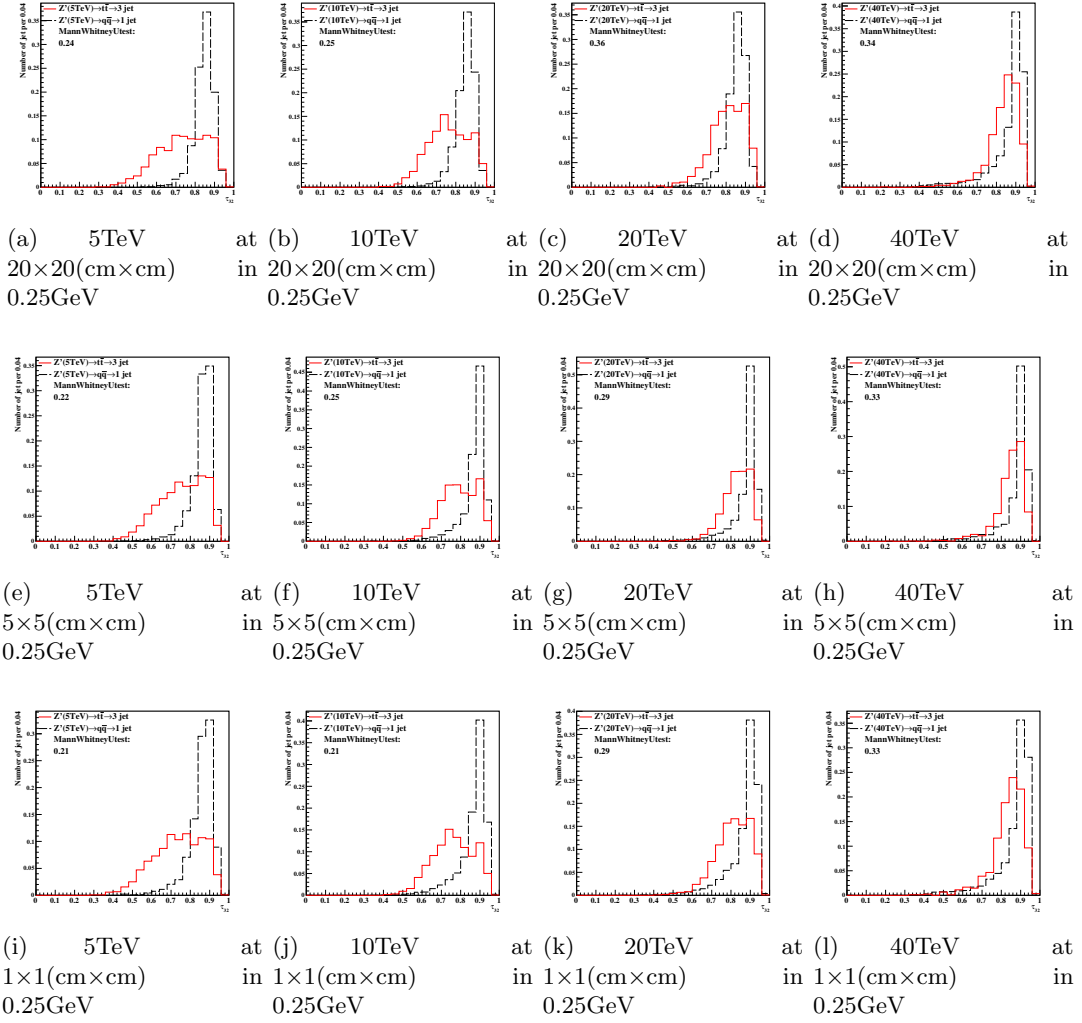
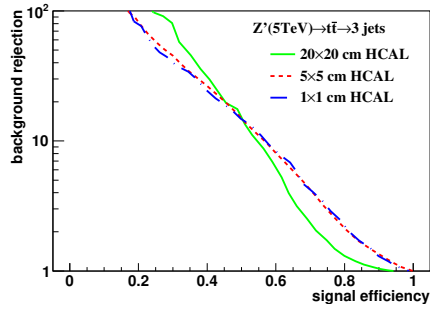
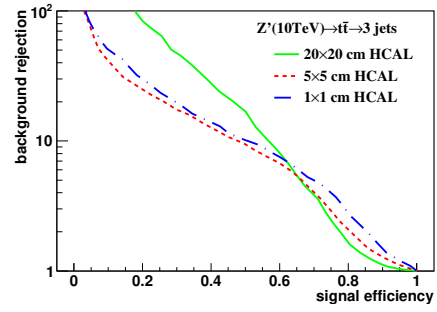


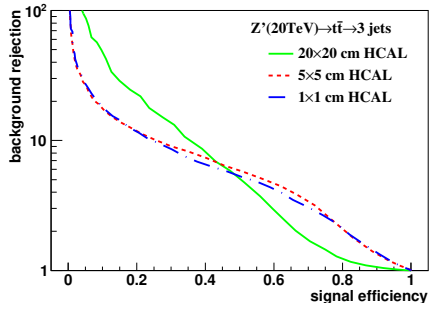
Figure 20: Distributions of Mann-Whitney value U in 5, 10, 20, 40 TeV energy collision for τ_{32} in different detector sizes. Cell Size in 20×20 , 5×5 , and 1×1 (cm x cm) are shown here.



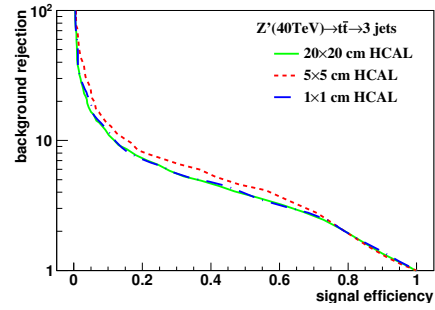
(a) 5 TeV rawhit cut at 0.25GeV



(b) 10 TeV rawhit cut at 0.25GeV

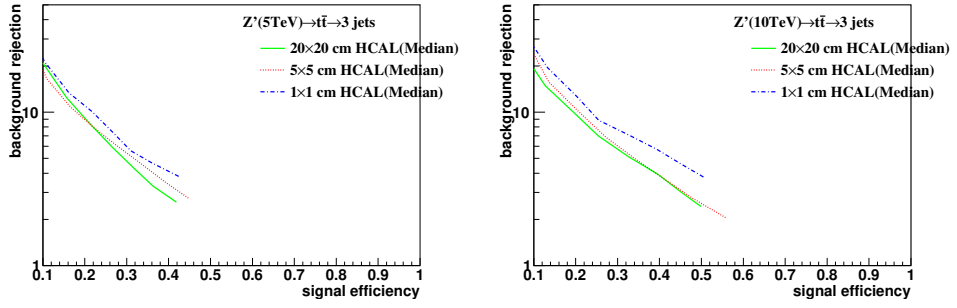


(c) 20 TeV rawhit cut at 0.25GeV

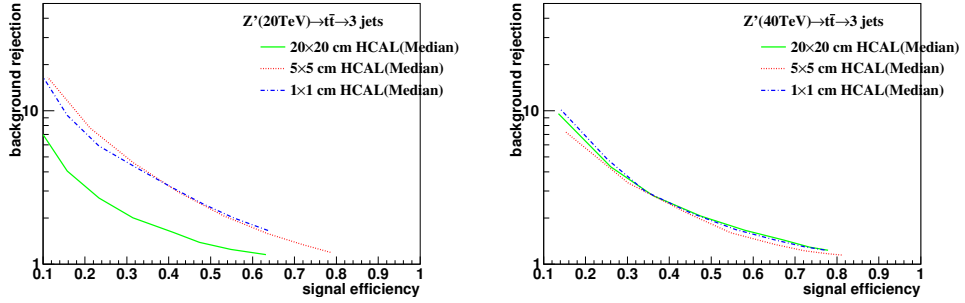


(d) 40 TeV rawhit cut at 0.25GeV

Figure 21: 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.

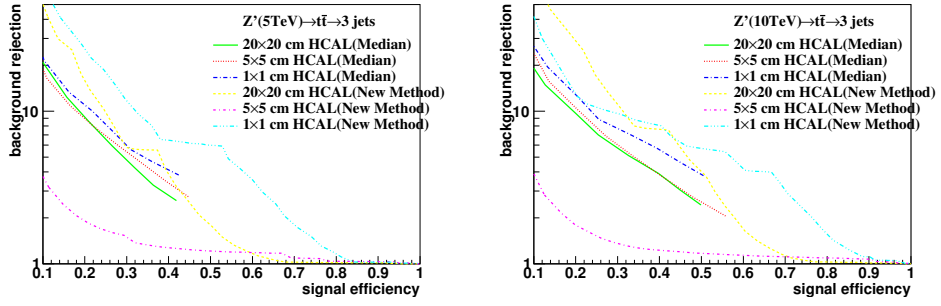


(a) 5 TeV rawhit cut at 0.25GeV fix central at Median (b) 10 TeV rawhit cut at 0.25GeV fix central at Median

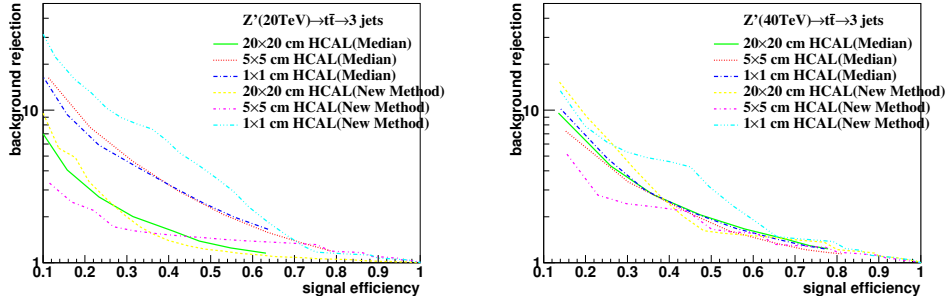


(c) 20 TeV rawhit cut at 0.25GeV fix central at Median (d) 40 TeV rawhit cut at 0.25GeV fix central at Median

Figure 22: 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.

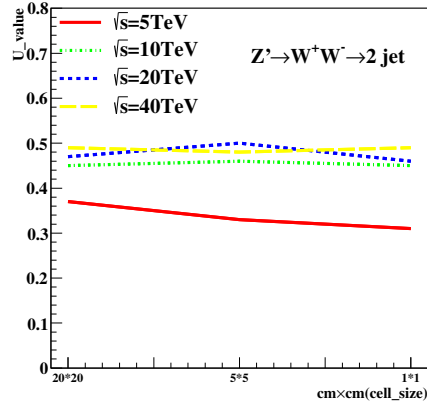


(a) 5 TeV rawhit cut at 0.25 GeV fix central at Median (b) 10 TeV rawhit cut at 0.25 GeV fix central at Median

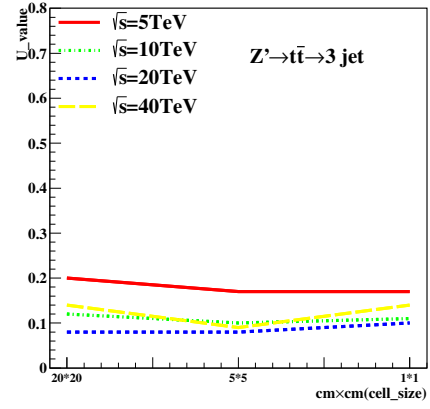


(c) 20 TeV rawhit cut at 0.25 GeV fix central at Median (d) 40 TeV rawhit cut at 0.25 GeV fix central at Median

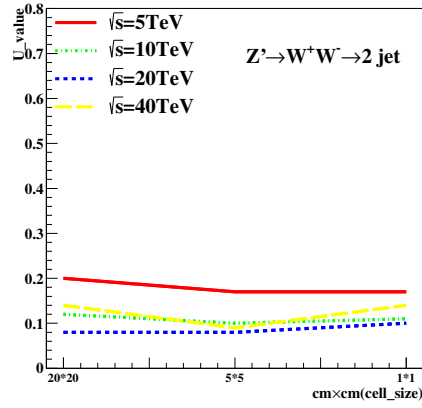
Figure 23: Signal efficiency versus background rejection rate using τ_{32} . The energies of collision at (a) 5, (b) 10, (c) 20, (d) 40 TeV are shown here. In each picture, the three ROC curves correspond to different detector sizes.



(a) τ_{21} in cluster

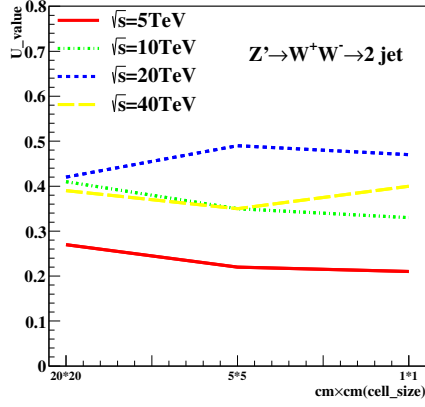


(b) τ_{32} in cluster

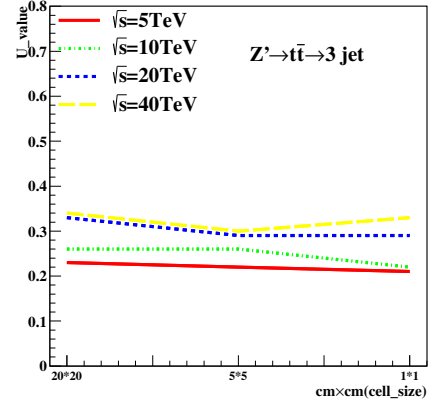


(c) $c_2^{(1)}$ in cluster

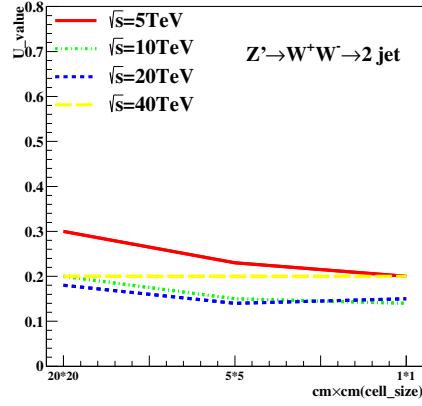
Figure 24: The Mann-Whitney U values for τ_{21}, τ_{32} and $c_2^{(1)}$ reconstructed from calorimeter clusters at different collision energies correspond to different detector sizes in cluster. The energies of collision at 5, 10, 20, 40, 20, 40TeV are shown in each figure.



(a) τ_{21} rawhit cut at 0.5 GeV

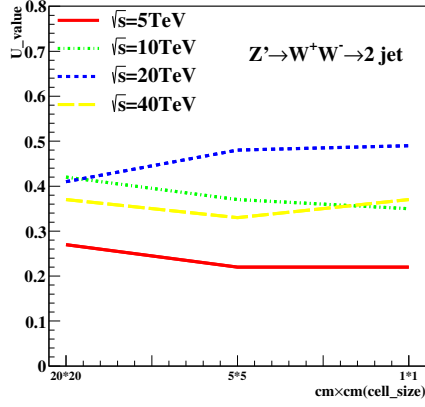


(b) τ_{32} rawhit cut at 0.5 GeV

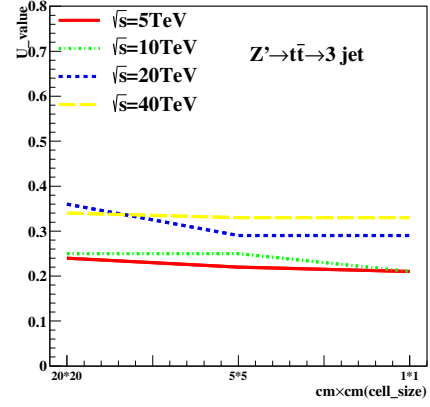


(c) $c_2^{(1)}$ rawhit cut at 0.5 GeV

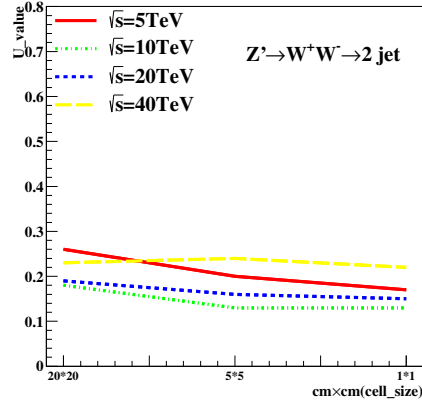
Figure 25: The Mann-Whitney U values for τ_{21}, τ_{32} and $c_2^{(1)}$ reconstructed from calorimeter hit at 0.5 GeV cut at different collision energies correspond to different detector sizes in rawhit cut at 0.5 GeV. The energies of collision at 5, 10, 20, 40, 20, 40 TeV are shown in each figure.



(a) τ_{21} rawhit cut at 0.25GeV



(b) τ_{32} rawhit cut at 0.25GeV



(c) $c_2^{(1)}$ rawhit cut at 0.25GeV

Figure 26: The Mann-Whitney U values for τ_{21}, τ_{32} and $c_2^{(1)}$ reconstructed from calorimeter hit at 0.25GeV cut at different collision energies correspond to different detector sizes in cluster. The energies of collision at 5, 10, 20, 40, 20, 40TeV are shown in each figure.