χ^2 method to select the hadronic top combination in the ttbar event

Abideh jafari

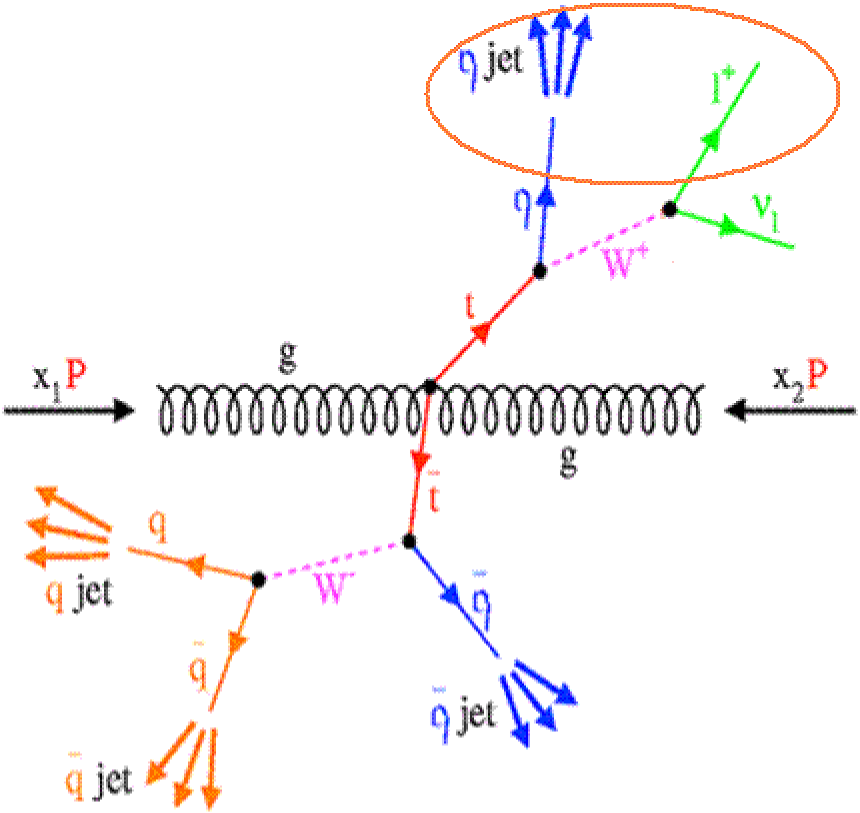
Introduction

- The aim of the analysis is to find the b-tagging efficiency using top events. Top events are a very rich source of b-jets.

- The semiElectronic decay channel of ttabr will be used in this analysis

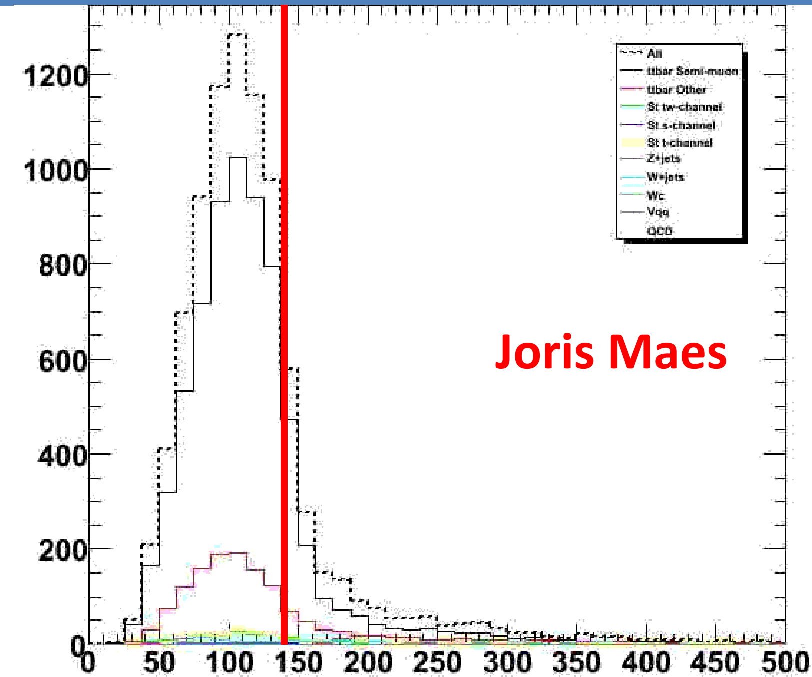
- The key point to take the b-jet among all four jets without biasing the efficiency, is the invariant mass of the lepton and the b-jet from the leptonic side of the ttbar event.

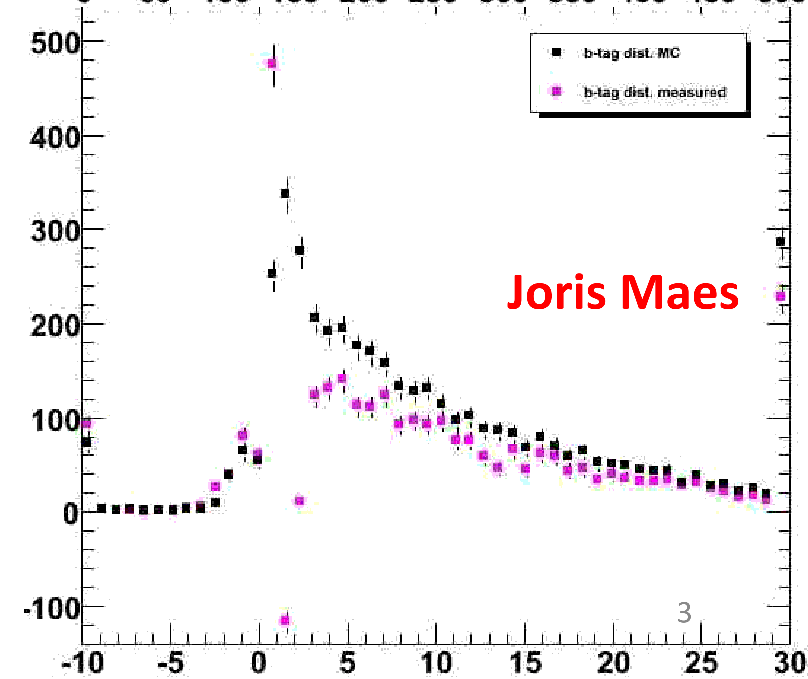
- So, one of the challenges is to find the b-jet coming from the leptonic decay of one of the top-quarks.
- We try to find it using a χ^2method. (coming later)



Introduction

- Another challenge is to subtract the backgrounds, (light jets from W+Jet events, QCD ,etc.) from the signal region
- It can be done either by using the MC information or by looking at some variables in bkg dominated region.
- A variable like b-tag discriminator (BD) is to have the same shape for light jets in signal and background dominated region.
- One can extract the shape of BD from the right side, multiply it by a factor and subtract it from the distribution of BD in the left side. Then, taking the b-jet from left side, the b-tag efficiency can be found
- The analysis in the muon channel is done by Joris Maes and the details can be found in sources mentioned in the backup.





How to recognize the hadronic top?

- The first step is to reconstruct the hadronic Top with 3 jets and let the other jet (say leptonic b-jet) play the role in the rest of the analysis, i. e. mlb calculation.
- Our tool to find the best jet combination for hadronic top, is a chiSquaed variable. It is calculated under some constraints like the mass of the W/Top and their mass resolution.
- These constraints are coming from the result of fitting a function on the mass distribution of the reconstructed Top/W.

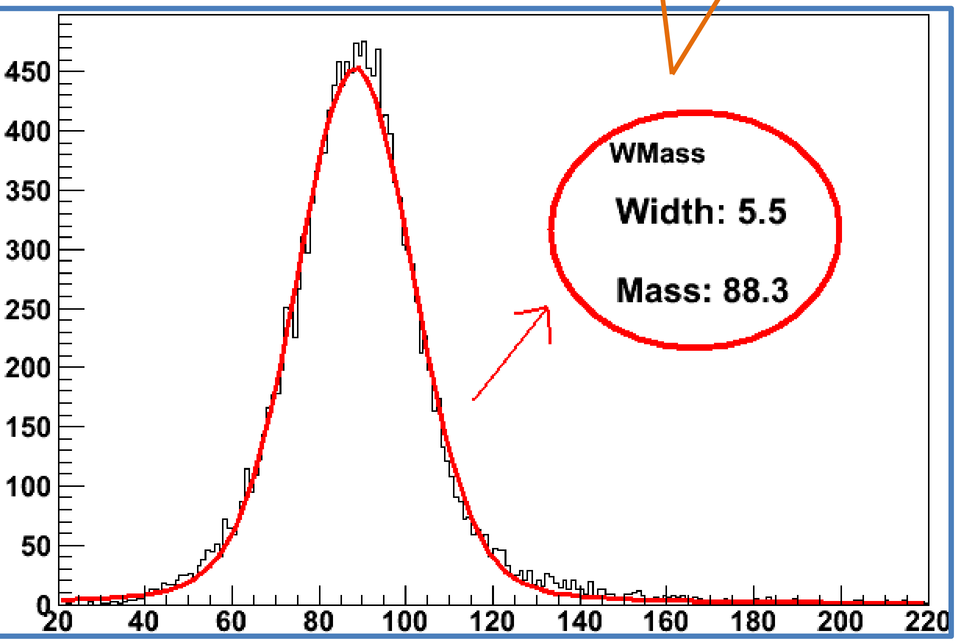
- In each event, there are different jet combinations of which, only one is corresponding to the decay products of the Hadronic TopQuark.
- This combination is expected to give us the minimum chiSquared value.
- So, in each event, this minimum value is to be found and it should be less than some χ^2_0 value to be accepted as coming from the hadronic top.
- This χ^2_0 value is to be determined by studying the purity and efficiency of the cut on χ^2.

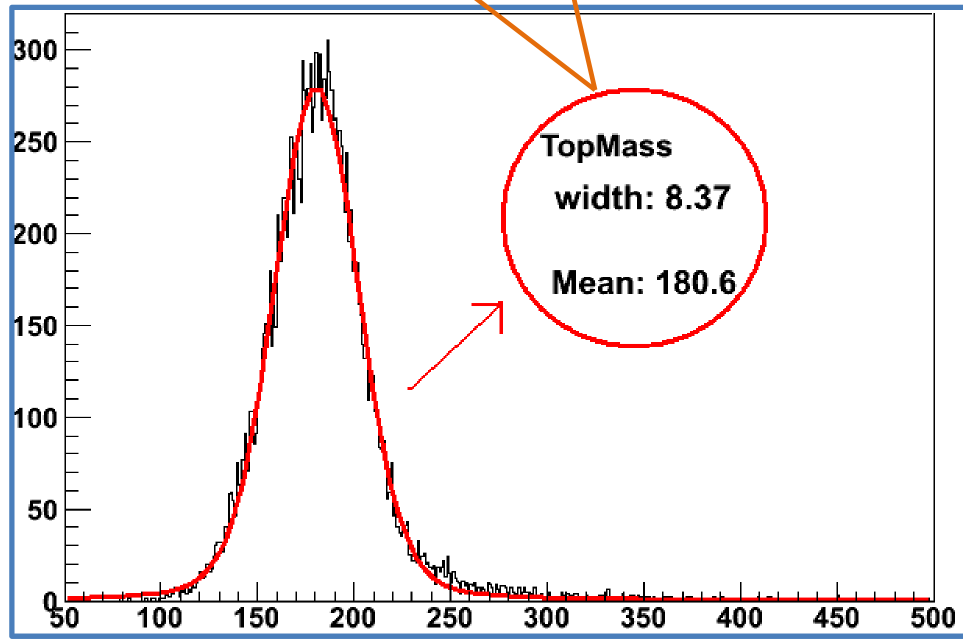
*** The data-set is TtBar Pythia sample summer09. Skimmed to semiElectronic decay channel*

*** The sofware is CMSSW314*

How to calculate the χ2

$$\chi^2 = \left(\frac{recW.Mass() - WMass}{\sigma_W}\right)^2 + \left(\frac{recTop.Mass() - TopMass}{\sigma_{Top}}\right)^2$$





- All jets meet the criteria of |η|<2.4, Pt>25 GeV, nCaloTowers>5
- Events only asked to have at least 4 of these jets
- Jets are matched to the genQuarkes and then are used in W/Top reconstruction.
- The closest jet to a genQuark is matched to it only if ΔR(jet,Q) < 0.3.
- The fit function is a Breit-Wigner convoluted with a Gaussian.
- The constraints, encircled values, are resulted from fitting.

Event Selection

- Events are selected according to what we had discussed in the meeting except here there is no request for "*at least on b-jet*".

  - Passing the HLT of having a non-Isolated electron.(68019)

  - At least one electron, $|\eta|<2.4$ (gap is excluded), Pt > 20 GeV, isolated, identified, with d0<200 μm.(38061)
  - Exactly one electron with the criteria above.(38041)
  - No second electron with $|\eta|<2.4$ (gap is excluded), Pt > 20 GeV, and identified as robustLoose.(36568)

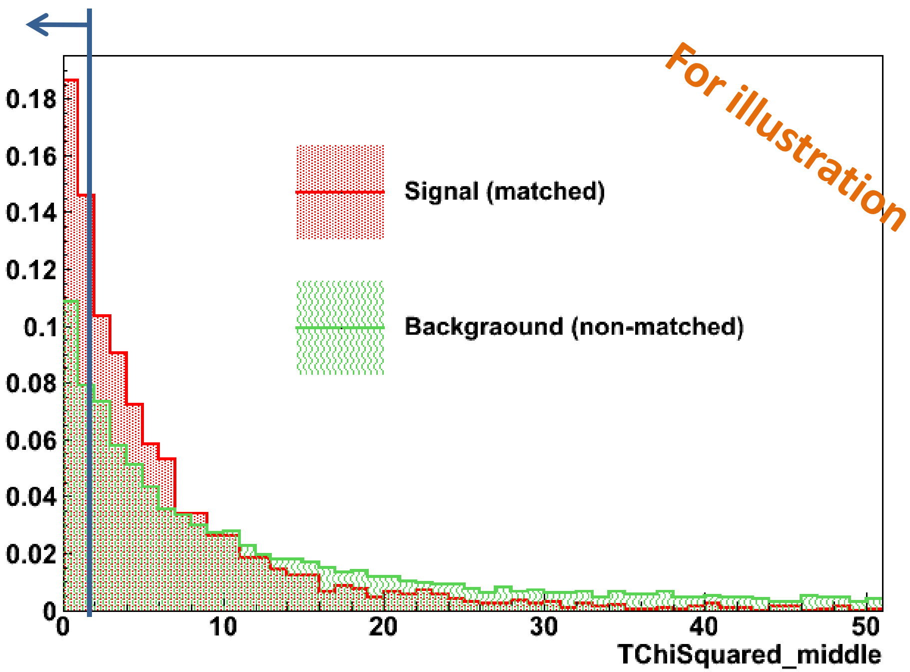
  - No isolated muon with with $|\eta|<2.1$, Pt > 20 GeV, d0<200 μm, $\chi^2<10$, nValidHits > 11. (36567)

  - Jets are cleaned from electrons (Maryam's Method)
  - At least 4 jets with $|\eta|<2.4$, Pt>25 GeV, nCaloTowers>5 (17998)

❖ *initial number of events: 70925*

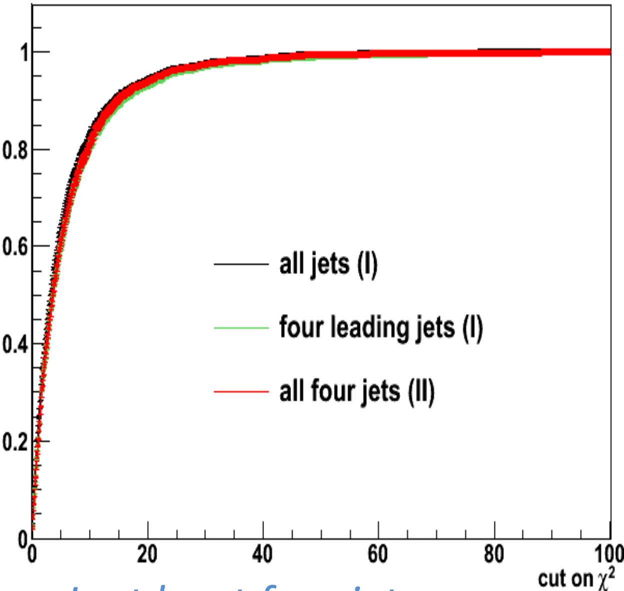
How does the χ^2 work?

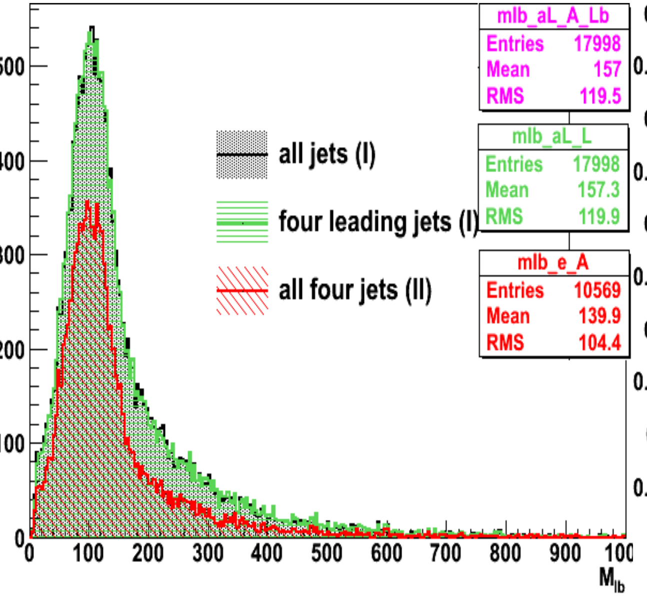
- In order to see how does this method works, if it is stable enough, the analysis is split as follows:
  - We may select the events with
    - At least 4 jets
    - Exactly 4 jets
  - In each event we can take all jets or only the first 4 leading jets in to account to make the combinations
- On the other hand, applying a cut on the χ^2 gives us an efficiency and purity. The behaviors of them are investigated vs. the cut on the χ^2.

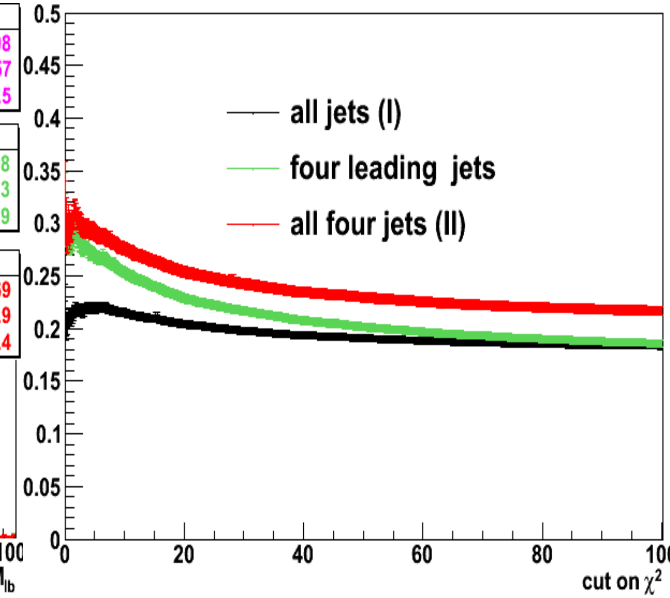


- Signal efficiency: number of signals (events with matched combination) that passed the cut.(red left/red right)
- Purity: the signal fraction in events pass the cut. (left side: red/red+green)

First results







- *I: at least four jets*
- *II: exactly four jets*

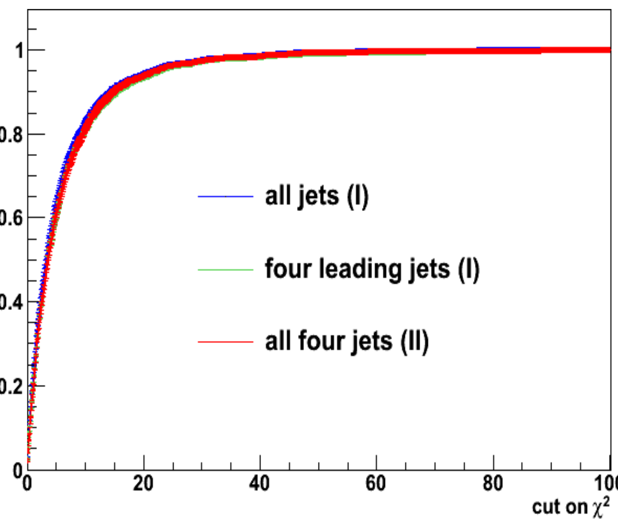
- **Efficiencies are the same.**
- **When we asked for at least 4 jets, working with all jets or the leadings has no difference in mlb.**
- **Working with all jets, results in a very low purity.**
- **Asking for exactly 4 jets gives, makes the purity slightly better (we are not interested in high values of χ^2).**
- **In case of exactly 4 jets, we loose many events from the peak range.**
- **Asking for at least 4 jet and working with leading jets seems fine.**

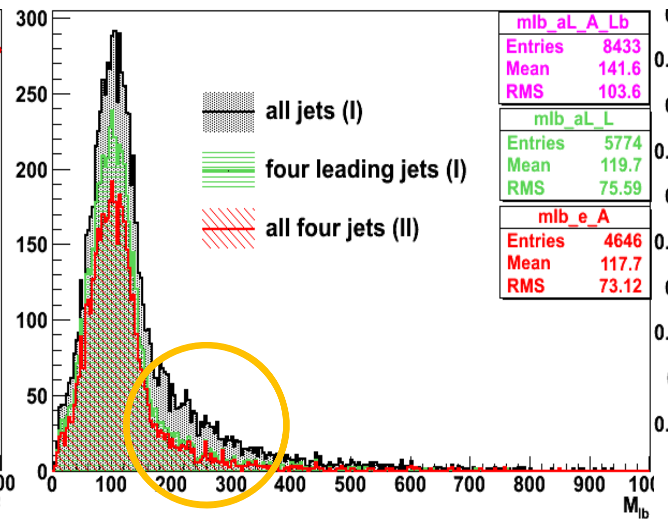
Accepting events with matched combination

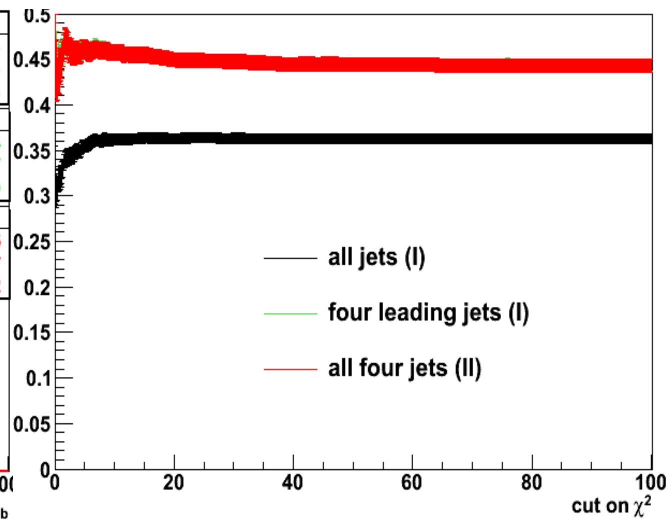
Comments from Jorgen/Stephanie:

Events enter the game, should have the potential of being matched. We have events in which none of the combinations are matched. These events give us a minimum χ^2 value but they can never give a matched combination.

As we are interested in seeing the power of our χ^2 method, it makes sense to give to the method events that have the possibility of being correctly selected, i.e. events with one matched combination.

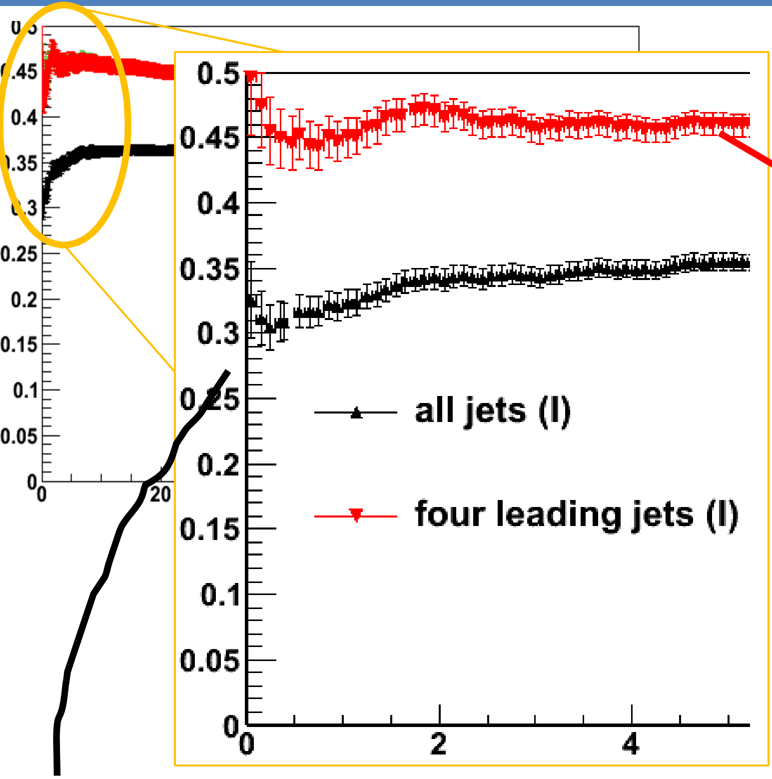




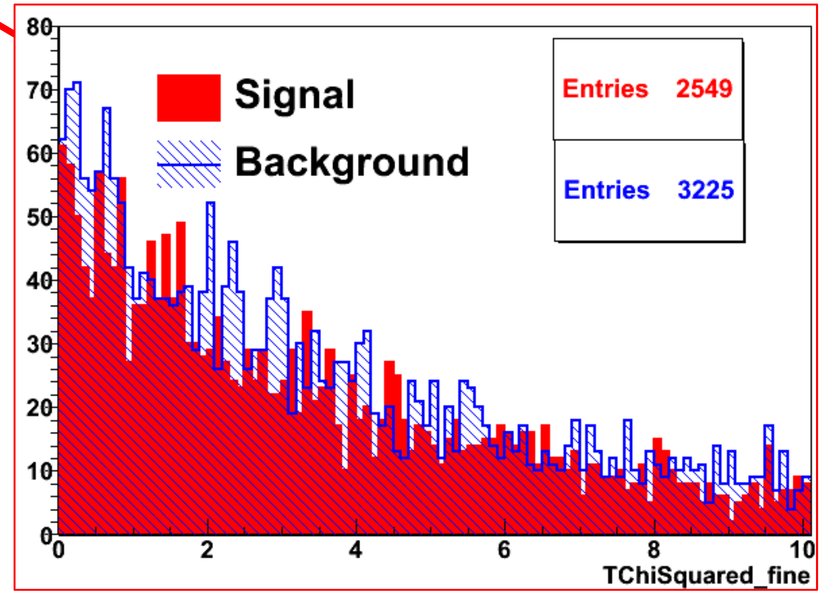


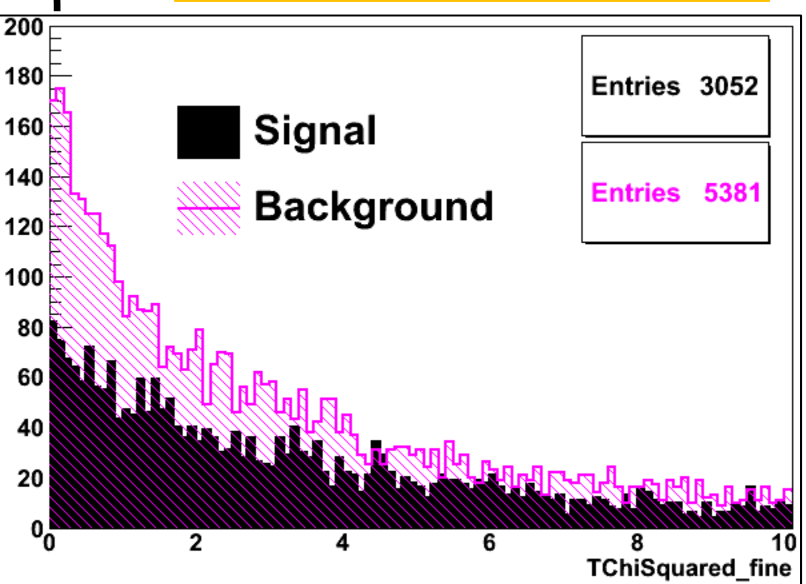
- **Working with all combinations, adds entries to the non-peak range of M_lb distribution. Also the purity is again very low.**
- **The purity of events with exactly for jets is the same as those events with at least four jets when only the leading jets are considered.**
- **Again asking for exactly 4 jets reduces the statistics in the peak range while keep the efficiancy and the purity similar to case of working with leading jets.**

A question



- Why the purity is not always falling? Specially in the black line, it is always raising!

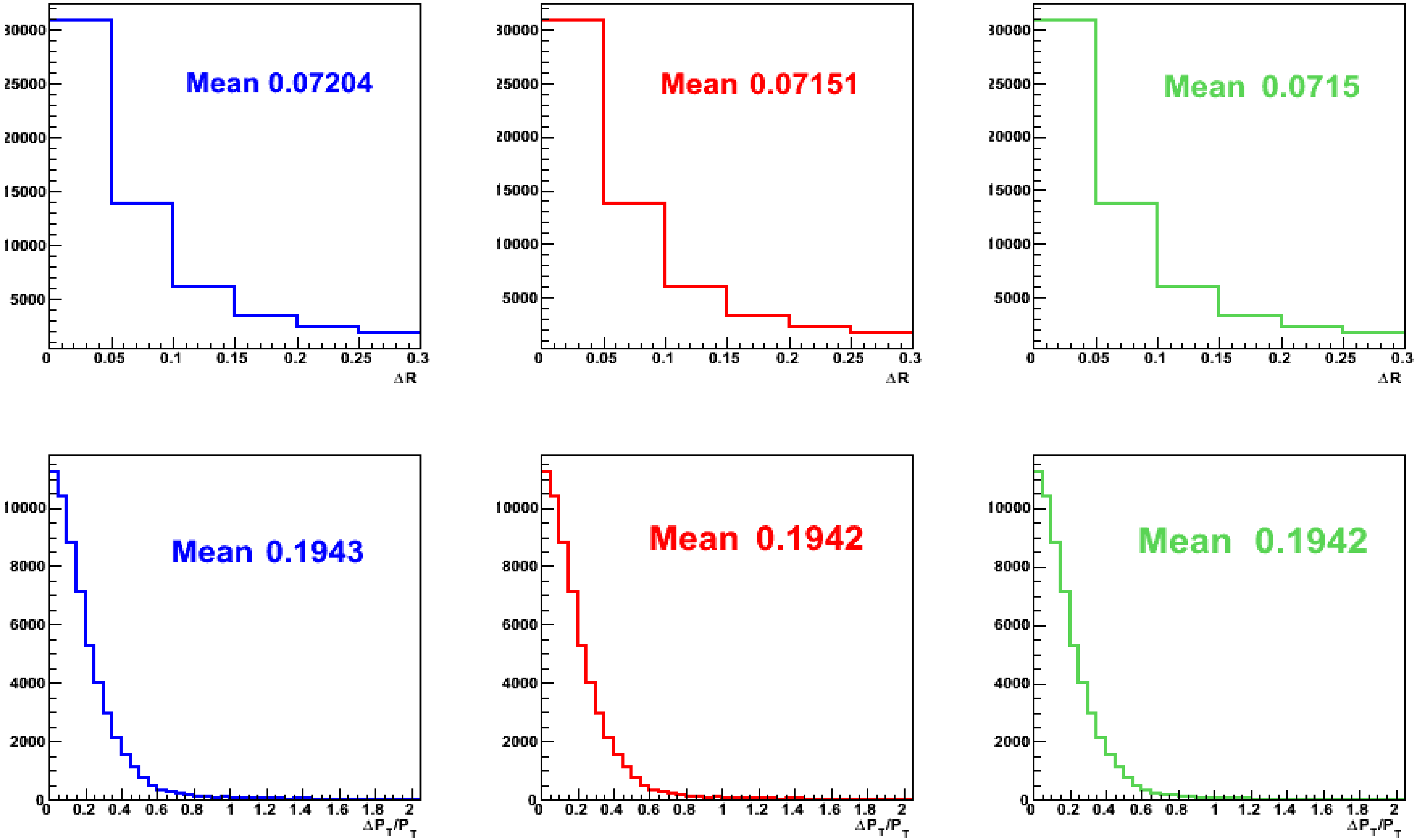




- **In the red plot which is to be our way of making combinations, the shape of the signal and background are not different that much. While we expect a distribution peaky near zero for signal and a more flat distribution for background.**
- For the black plot the situation is even worse. Instead of signal, bkg has a peak near zero

Is the matching algorithm problematic?

- **The first guess is that matching algorithm can be problematic:**



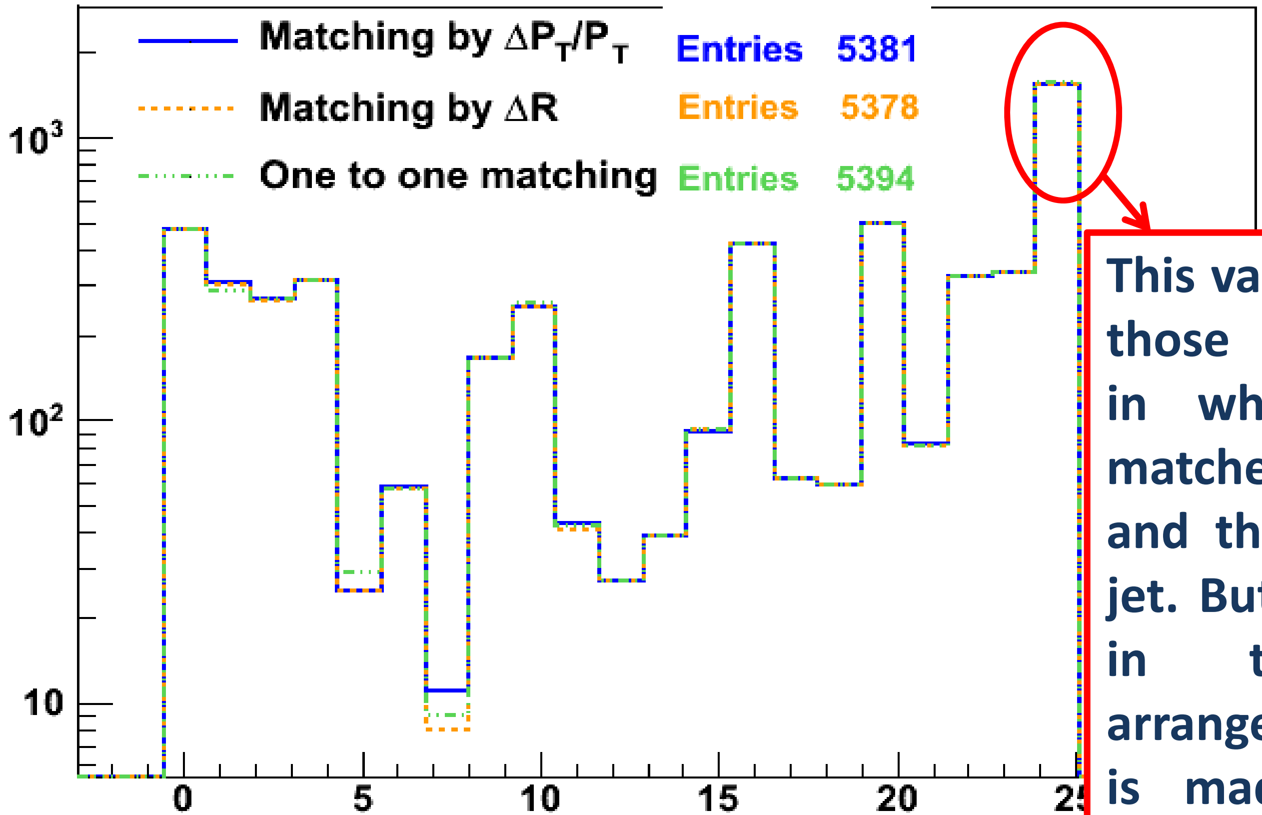
By ΔR

By ΔPt/Pt

By ΔR, check to be one to one

Go deeper …

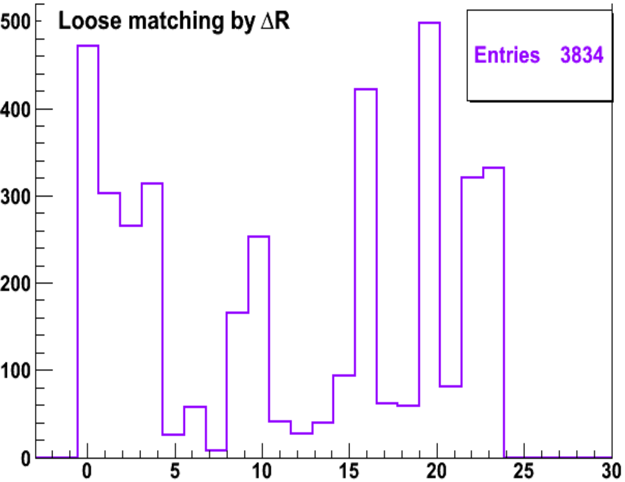
- The non-matched combinations can have either non-matched jets or matched jets in a bad arrangement. I have categorized them in 24 categories. This plot shows the type of non-matched combinations for different matching algorithms. The combinations here are made using all jet s.

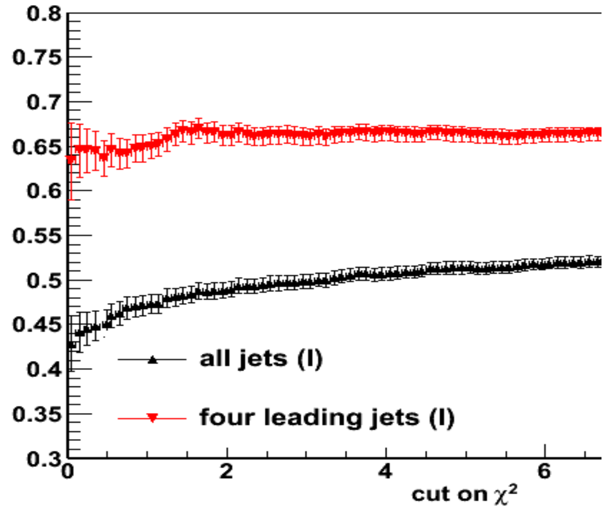


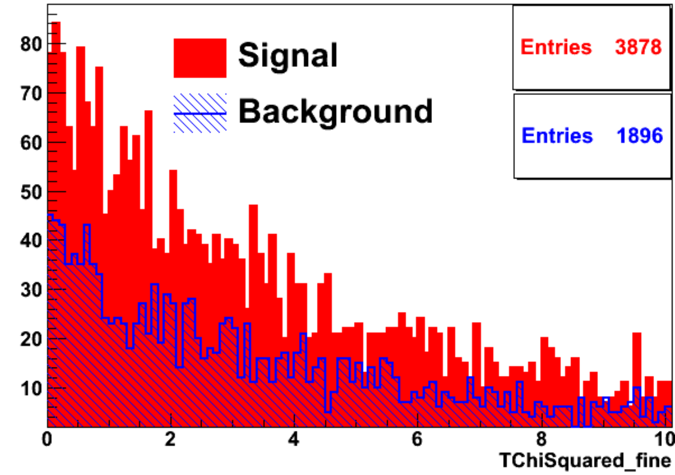
This value belongs to those combinations in which jets are matched to q, qbar and the hadronic b-jet. But they are not in the correct arrangement i.e. W is made with one light jet and the hadronic b-jet

Loosening the matching criterion of the combinations

- Here, I accept the combination as matched if it contains true jets even in bad arrangement. By this kind of matching, The remaining b-jet is still leptonic b.



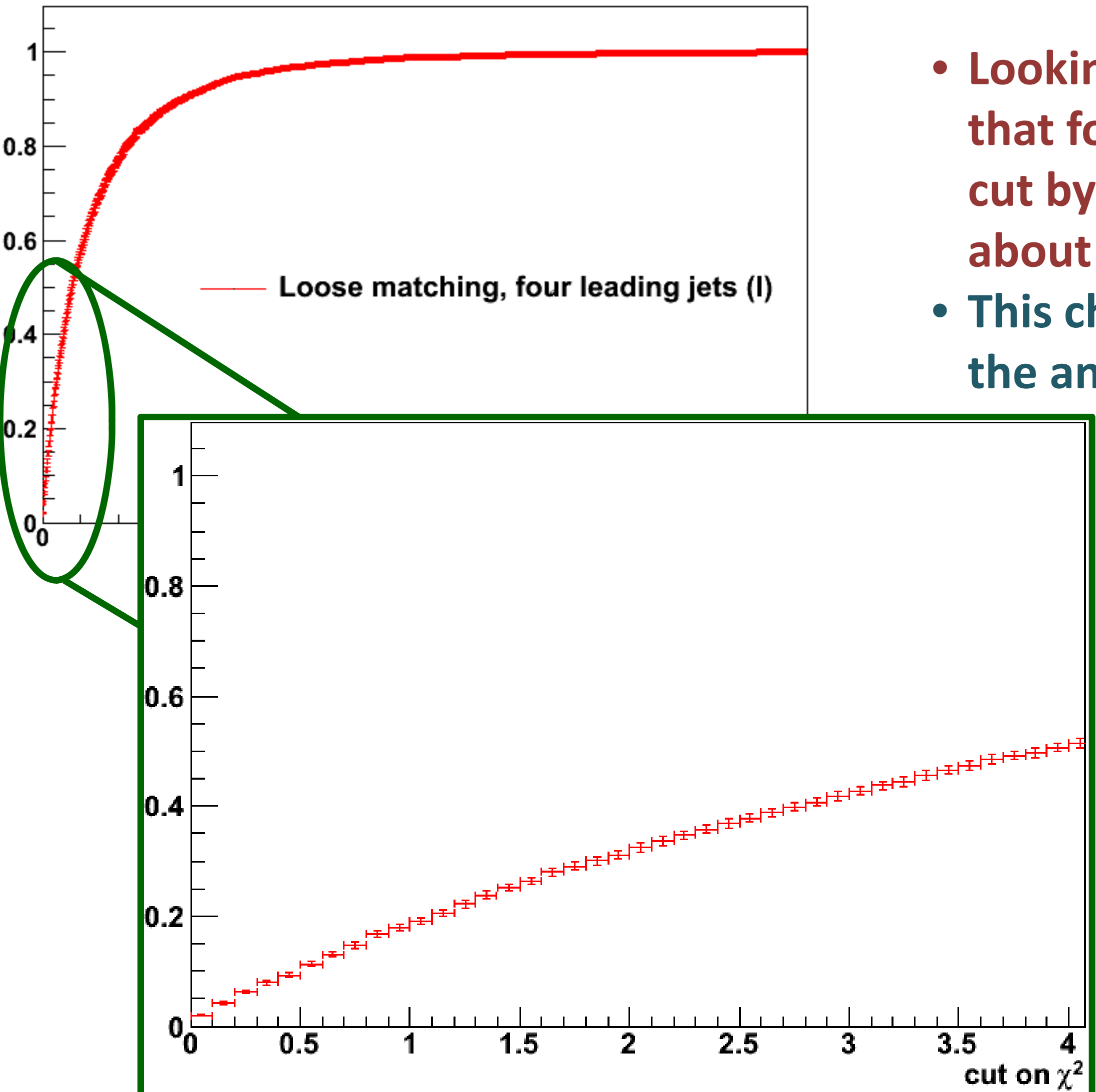




- The overall purity has increased as is expected.
- In the red plot, the fluctuations have decreased but the whole behavior has not changed that much.
- Although th e amount of signal has enhanced, but the shape signal and background are still very similar.
- For the back plot, the situation becomes even worse! It is raiding more sharply!

Loosening the matching criterion of the combinations

- If we work in low χ^2 range, where the purity is not decreasing but at least is kind of flat, then we may suffer from the systematic uncertainty. Because the efficiency is increasing sharply there.



- Looking at plot in green box, one can see that for example at χ^2 =0.5, if we change the cut by 0.1, the efficiency will change by about 2%.
- This change can have no effect at the end of the analysis or can be harmful.

- This can be investigated only at end of the analysis

Summary and conclusion

- Different ways of combining jets are investigated for different event selections.
  - The best choice is asking for at least 4 jets and using only 4 leading jets

- The purity of the selected events, is raising vs. the cut on χ^2. Different matching algorithms are studied to see if the strange behavior of purity vs. the cut, is coming from the way of matching.

- Ignoring the jet arrangement criterion in selecting the matched combination is not helpful.

- The χ-squared method is not so discriminating between the good and bad combinations in signal, i.e. semiElectronic decay channel of ttbar. The reason is still not clear for me.
- Perhaps adding backgrounds like W+Jets, can change the game.

- It is still a question that can we ignore this raising behavior as we are to work in low χ^2 range.

Back up

- CMS CR-2008/059
- http://indico.cern.ch/getFile.py/access?contribId=0&resId=2&materialId=slides&confId=63265