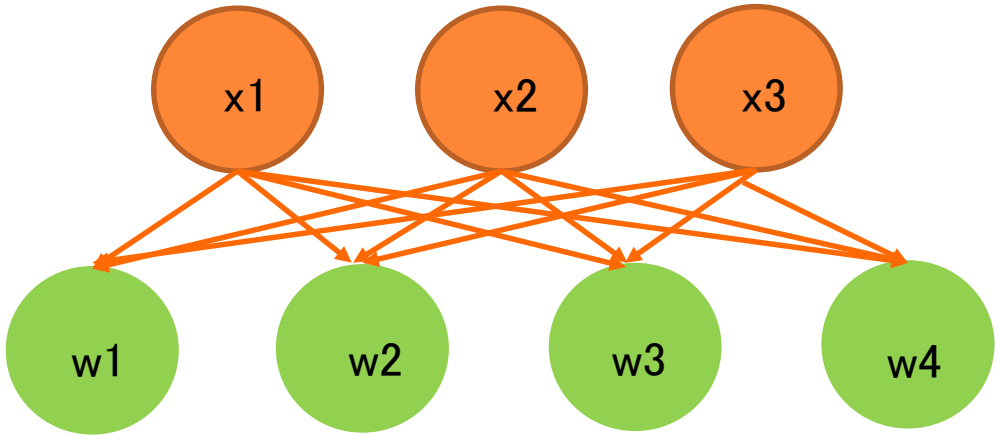
More explanation of Self-Organized Mapping for Jet Clustering and Trying some Jet Clustering Patterns

Masakazu Kurata

12/18/2015

Introduction

- SOM is one of an "unsupervised" neural network
  - Original idea can be got everywhere – Please check them!
- Looking for "winner" nodes
  - choose output node whose weight vector is similar with input data
  - This node is called as best matching unit(BMU)
- Adjusting the weight vector towards the input vector
  - Both BMU and the nodes which are close to BMU
  - Update formula: $w_j(t+1) = w_j(t) + \alpha_t(x - w_j(t)),$
  - α_t: neighborhood function(distance between BMU and other nodes)
- Overall procedure

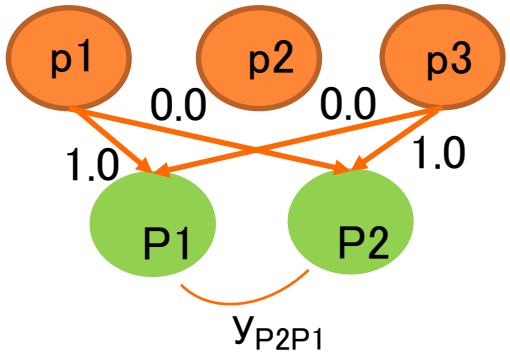


Input Layer

Output Layer

Procedure

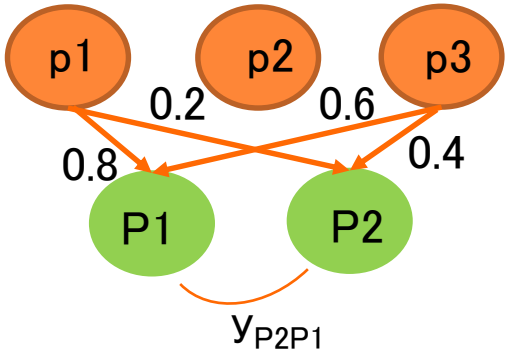
- Default: Durham result(e.g. p1 and p2 belong to P1, p3 belong to P2)
  - w11=1.0, w21=0.0
  - w12=1.0, w22=0.0
  - w13=0.0, w23=1.0



Input Layer

Output Layer

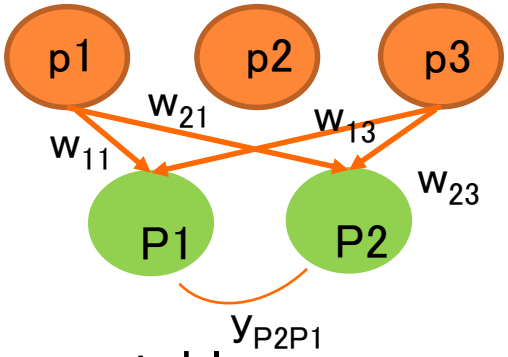
- Epoch 1:
  - For each input node:
    - Look for winner
    - Update the weights



Input Layer

Output Layer

- Epoch 2000(train end):
  - For each input node:
    - Look for winner
    - Update the weights



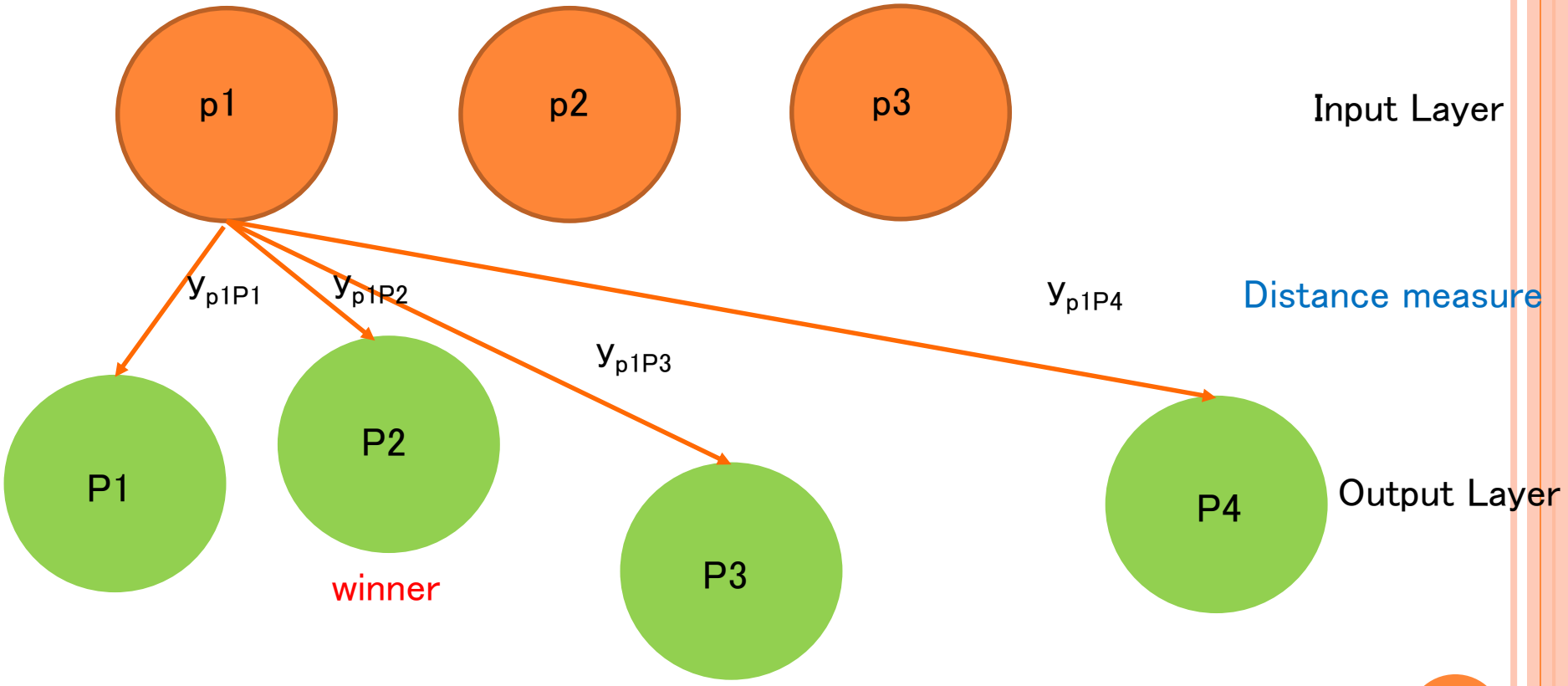
Input Layer

Output Layer

- Repeat until status becomes stable
- Reference jet 4-momentum: $P_i = \sum w_{ij} \cdot p_j , 0 \leq w_{ij} \leq 1$

Looking for winner node

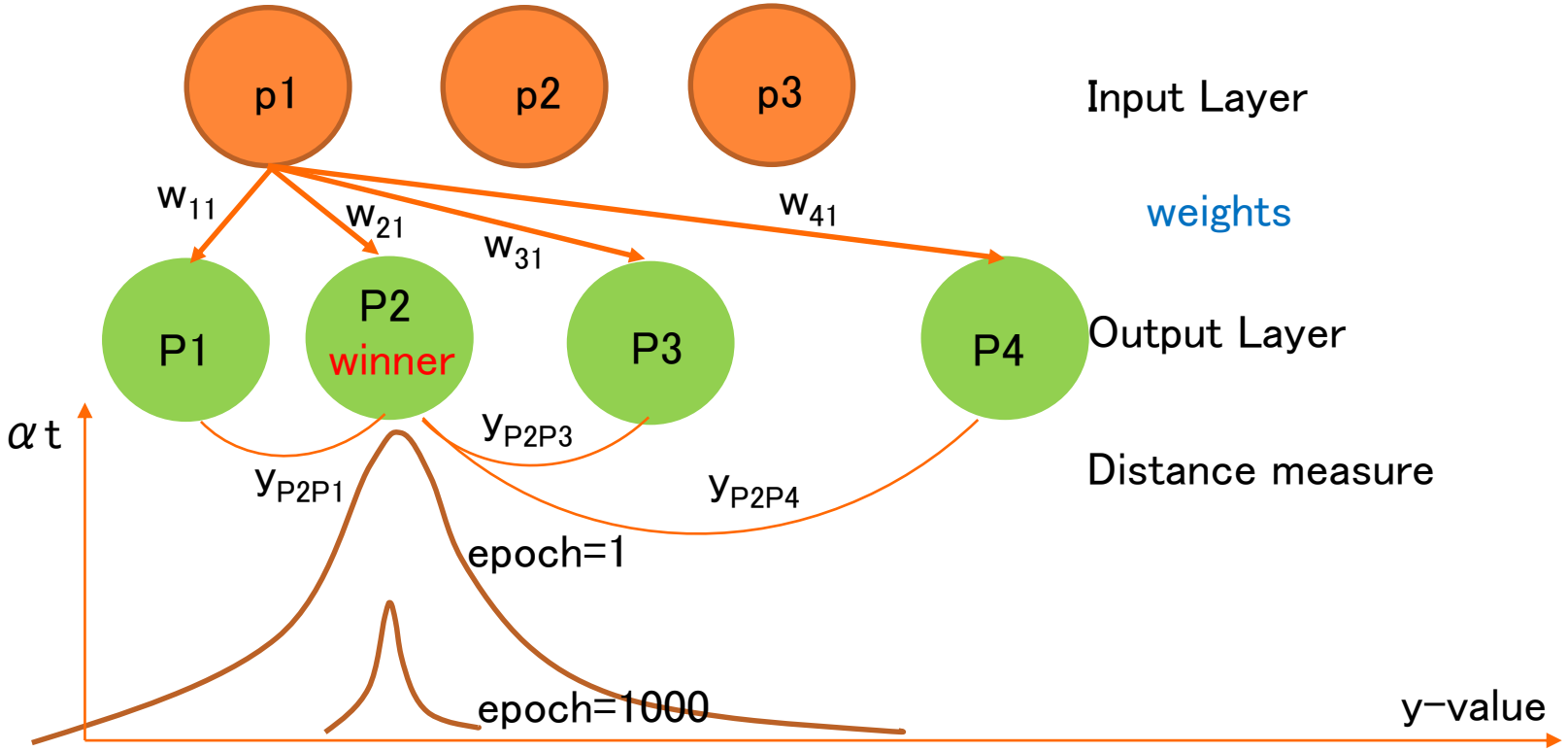
- "winner" node is decided using traditional distance measure
  - e.g.) Durham y-value $y_{ij} = \frac{2\min(e_i^2, E_j^2)(1.0 - cos\theta_{ij})}{E_{vis} \cdot E_{vis}}$
  - Example: Winner node for p1



  - If y_{p1P2} is smallest, winner node is P2

Updating strategy

- Distance between output node is defined using y-value
  - e.g.) Durham y-value $y_{ij} = \frac{2\min(E_i^2, E_j^2)(1.0 - cos\theta_{ij})}{E_{vis} \cdot E_{vis}}$



- wij is changed at each epoch:
  - Updating formula: $w_{ij}(t+1) = w_{ij}(t) + \alpha_t(1.0 - w_{ij}(t))$
  - αt shrinks gradually with epoch(repeated time)
  - Weight sum constraint: $\sum_i w_{ij} = 1$

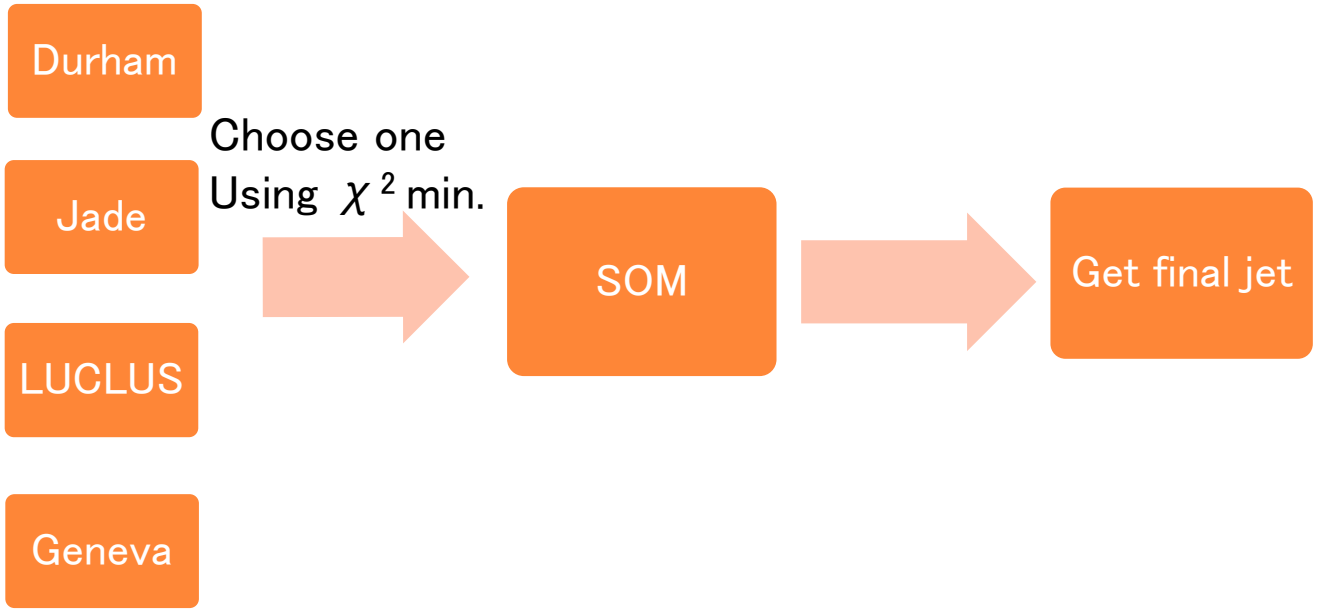
Procedure of SOM for jet clustering

1. Track is clustered until certain y-value using Durham
   - Each mini-jet is a input data for SOM
2. Default output vectors are result of Durham jet clustering
3. Choose BMU for a mini-jet p_j
   - Using distance measure of Jade with output vectors
4. Update weights of BMU and neighborhood output nodes
   - $w_{ij}(t+1) = w_{ij}(t) + \alpha_t(1.0 - w_{ij}(t))$
   - Using distance measure of Jade between BMU and other output nodes
   - αt shrinks gradually with epoch
5. 3. and 4. are performed for all the input data(mini-jets)
6. 5. is performed many times(called as "epoch")
7. After the training, each mini-jet is assigned to the output node which has smallest Jade distance measure
   - Node is regarded as a 'jet'

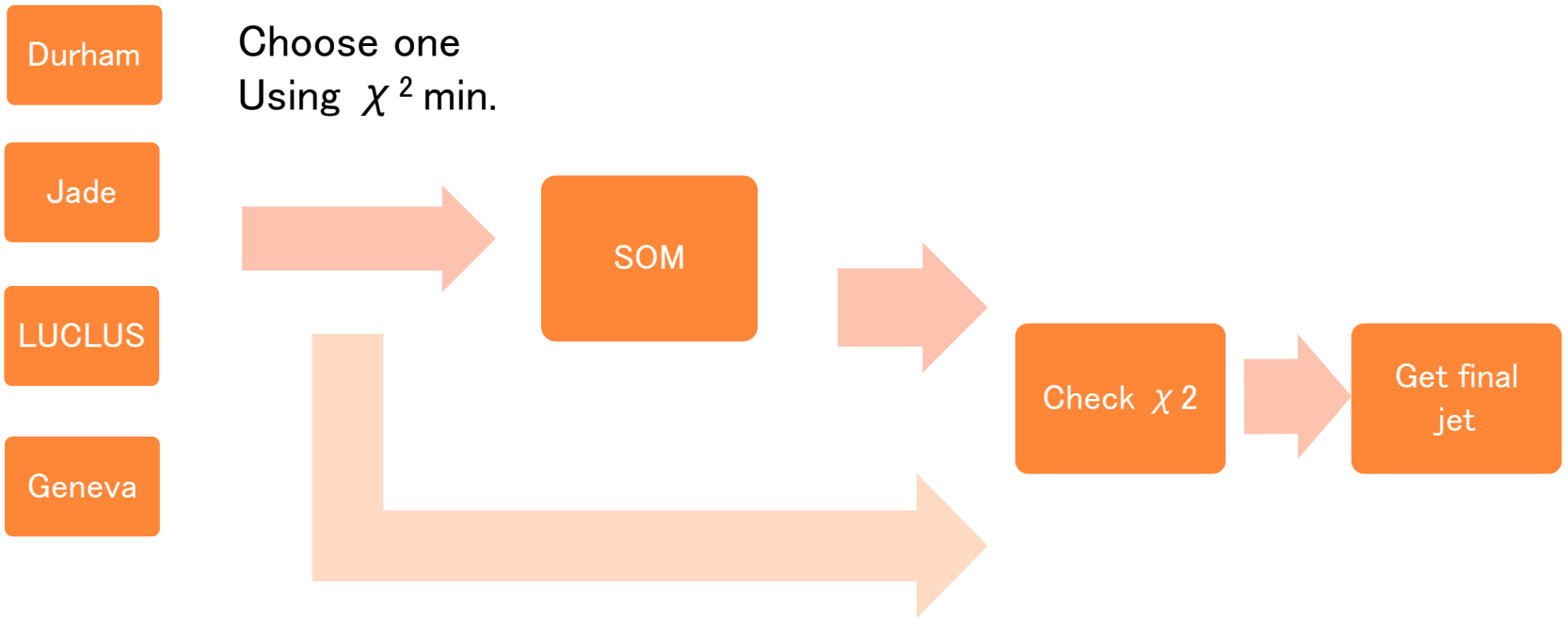
Trying some patterns

SOM Baseline

SOM USING 4 JET CLUSTERING

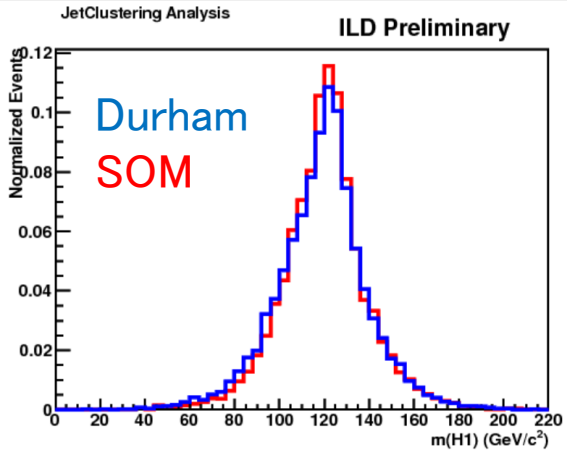


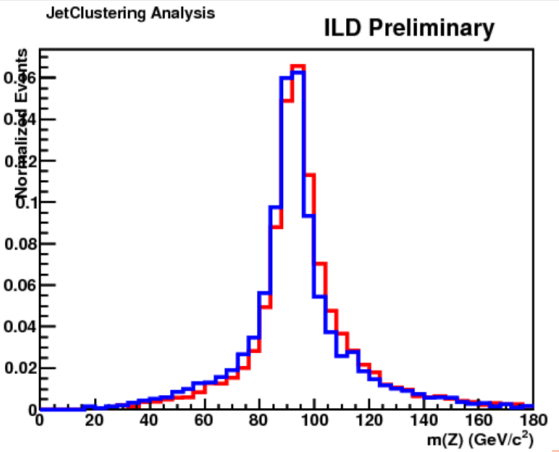
SOM USING CHI2



SOM RESULTS

- Using qqhh→qq(bb)(bb): 6 jet clustering(6 output nodes)
  - Use same event as original Durham clustering
  - Jet matching with MC truth is performed
  - Mass distribution:

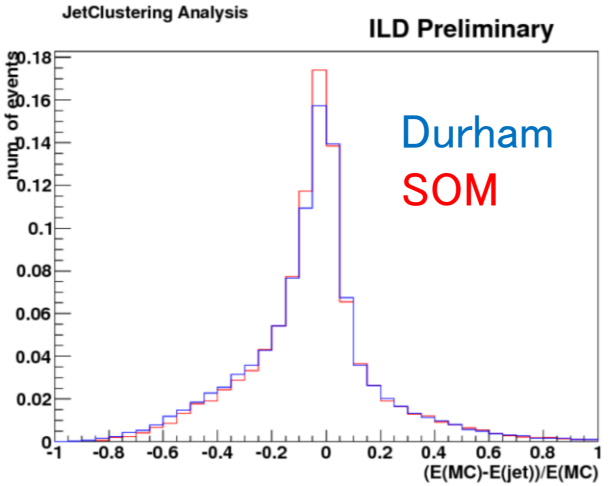




  - Num. of MC matched events:

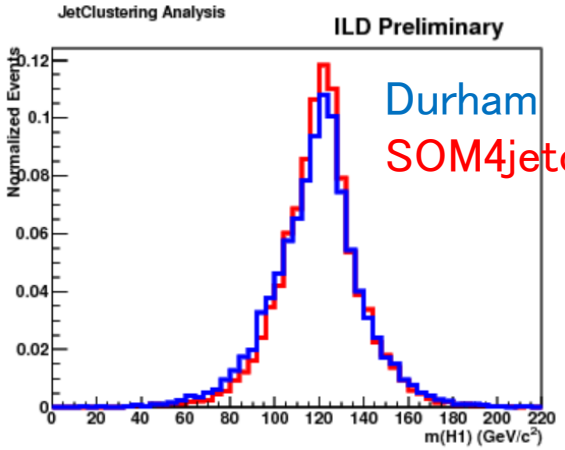
Method	Durham	SOM
Num. of events	6178	6278

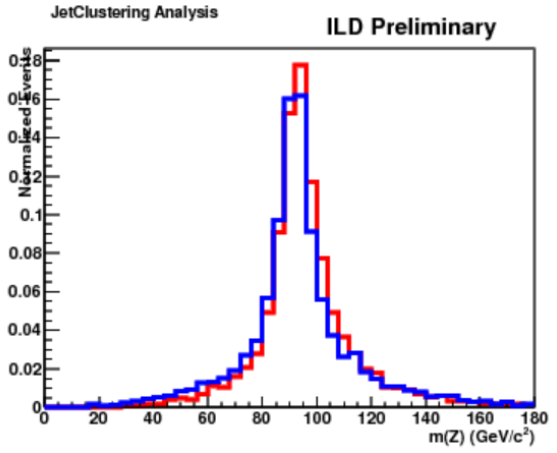
  - Jet Energy resolution of bjets



SOM USING 4JETCLU RESULTS

- Using qqhh→qq(bb)(bb): 6 jet clustering(6 output nodes)
  - Use same event as original Durham clustering
  - Jet matching with MC truth is performed
  - Mass distribution:

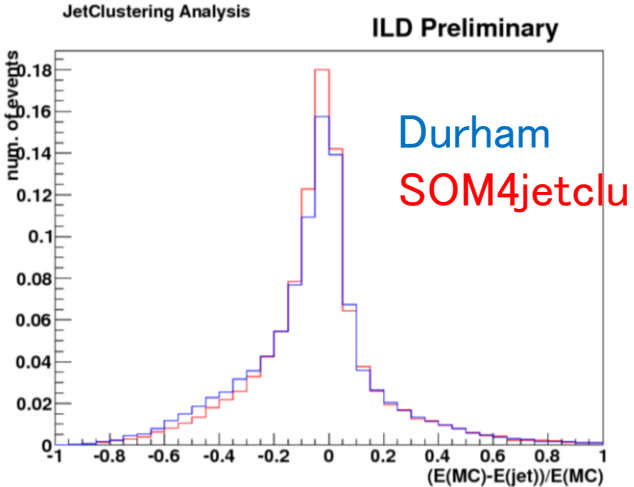




  - Num. of MC matched events:

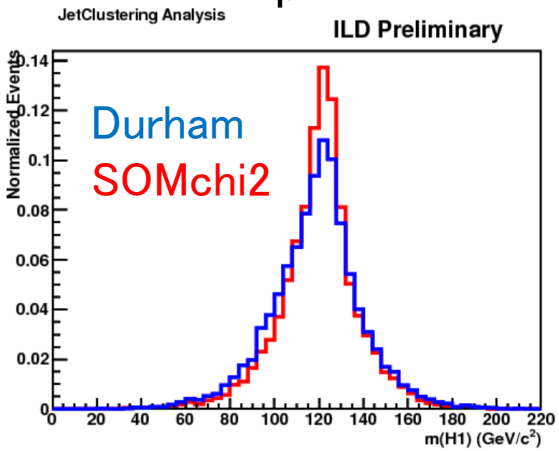
Method	Durham	SOM
Num. of events	6178	6254

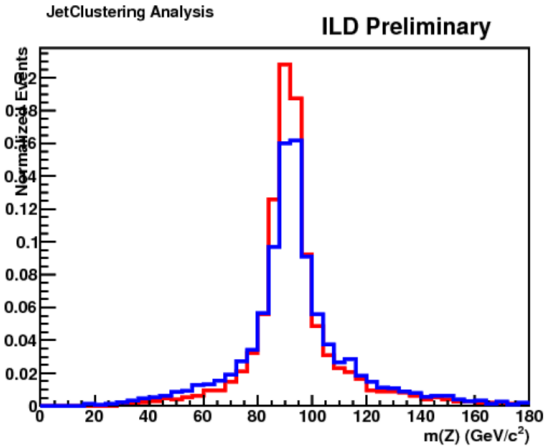
  - Jet Energy resolution of bjets
    - Slightly better



SOM+chi2 results

- Using qqhh→qq(bb)(bb): 6 jet clustering(6 output nodes)
  - Use same event as original Durham clustering
  - Jet matching with MC truth is performed
  - Mass distribution:

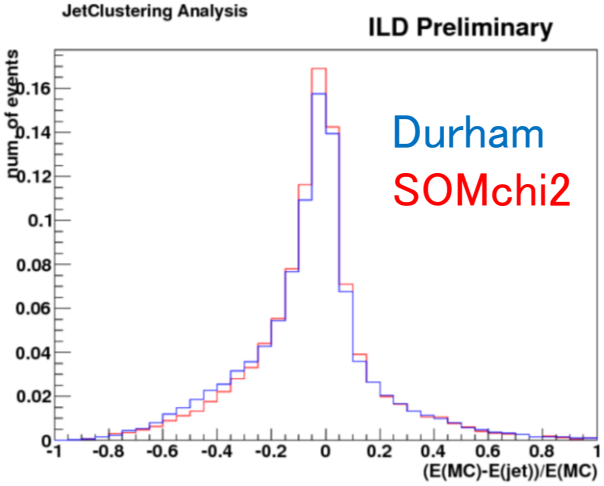




  - Num. of MC matched events:

Method	Durham	SOM
Num. of events	6178	6211

  - Jet Energy resolution of bjets
    - Better mass resolution
      does not go to better JER…
      →why?

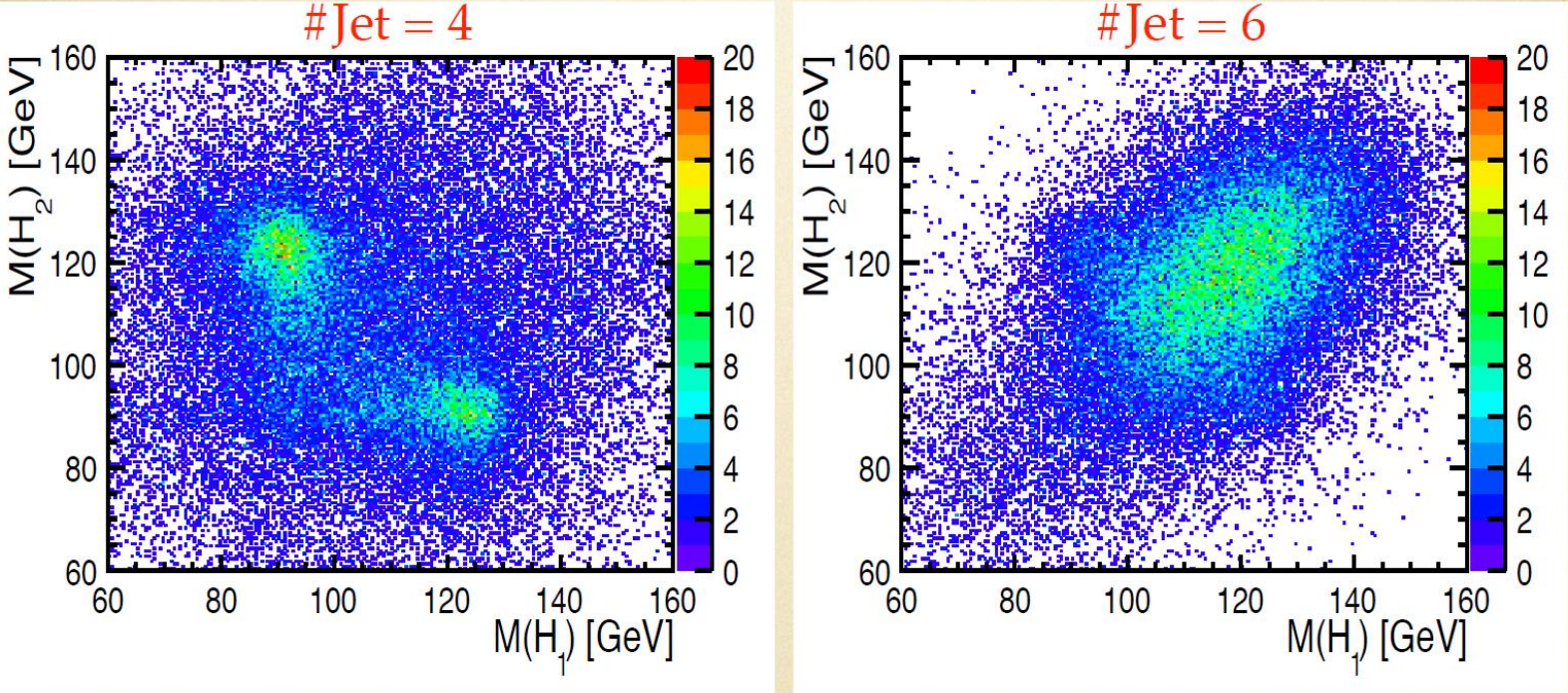


We are afraid if

- Mass resolution will become better, but··· if

No… Improvement should be addressed in terms background separation as well (more importantly in λ_{HHH} analysis)

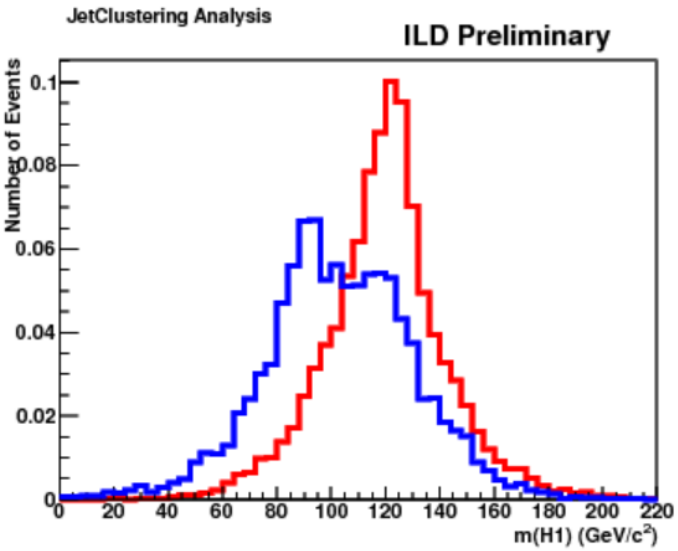
take a look at background $\nu\nu ZH$ @ 500 GeV

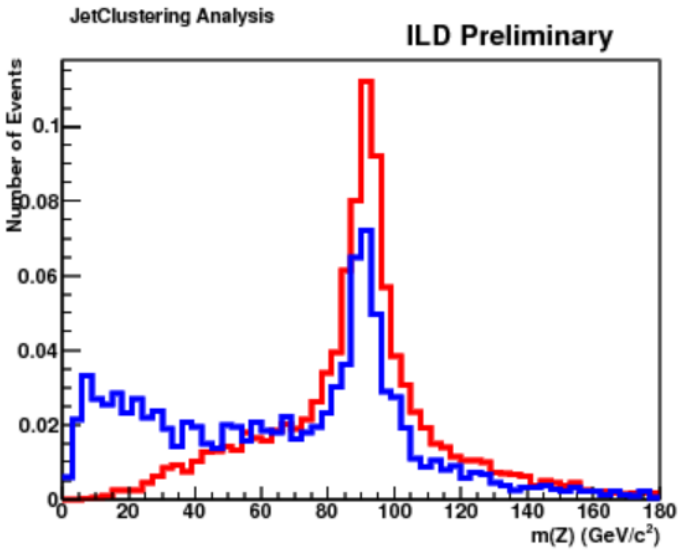


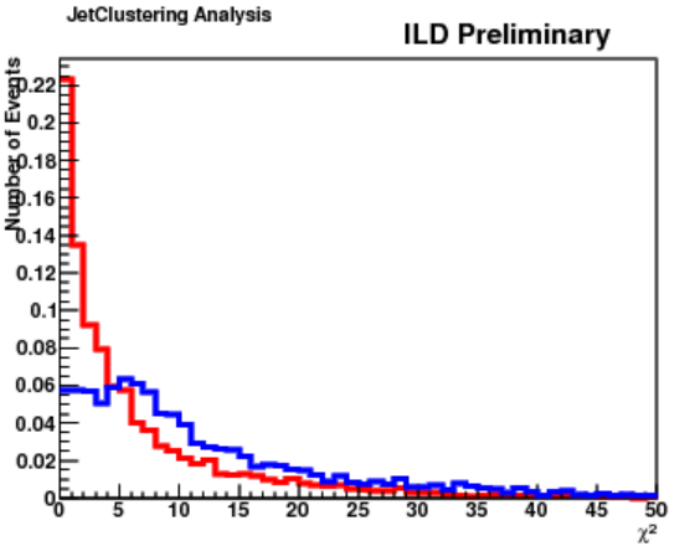
just using mass to pair jets is completely a disaster…

DURHAM – SIGNAL V.S. ZZH

- Using qqHH and ZZH
- Use events ～12000(signal) v.s. ～40000（ZZH)







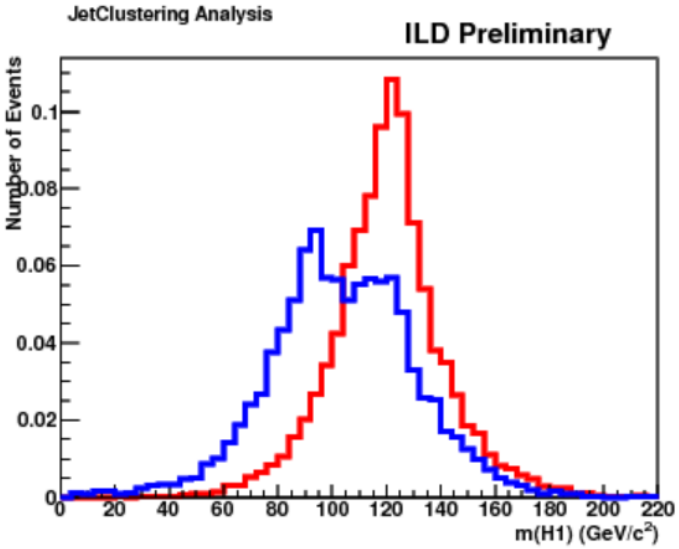
signal

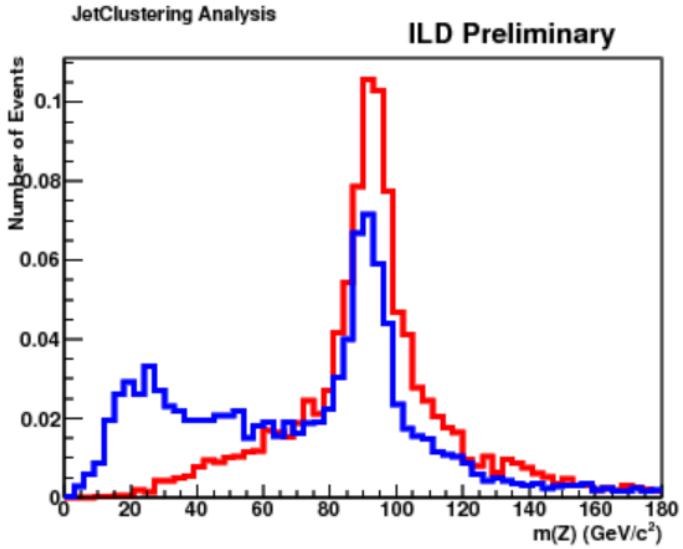
ZZH

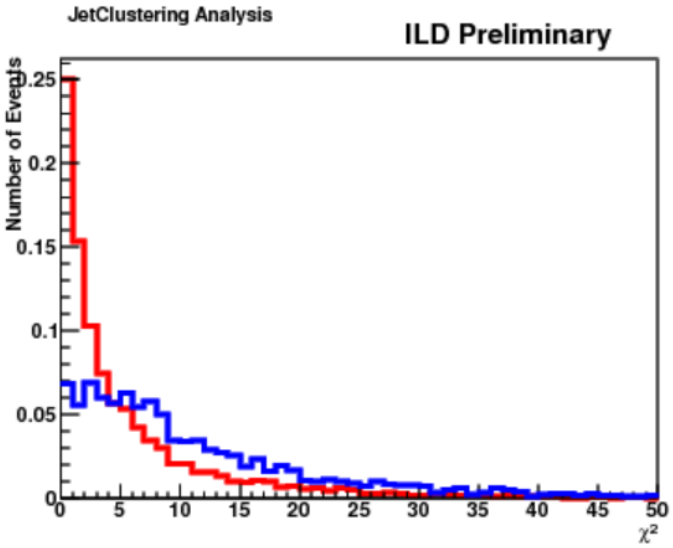
Btag>0.30 for 4 jets

SOM – SIGNAL V.S. ZZH

- Using qqHH and ZZH
- Use events ～12000(signal) v.s. ～40000（ZZH)







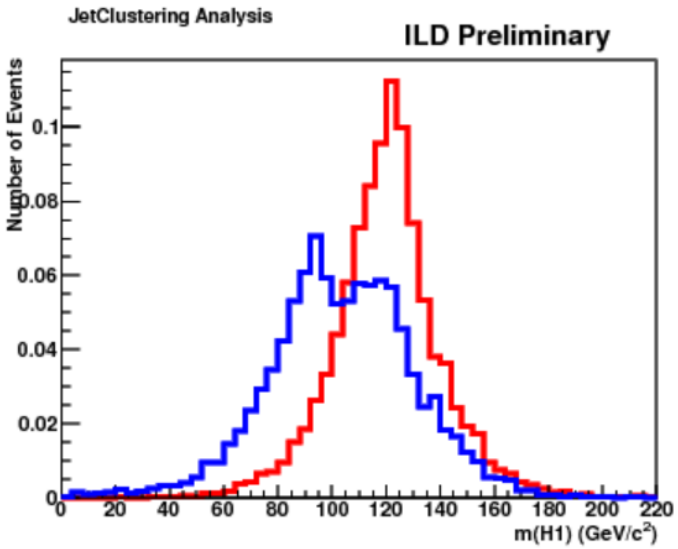
signal

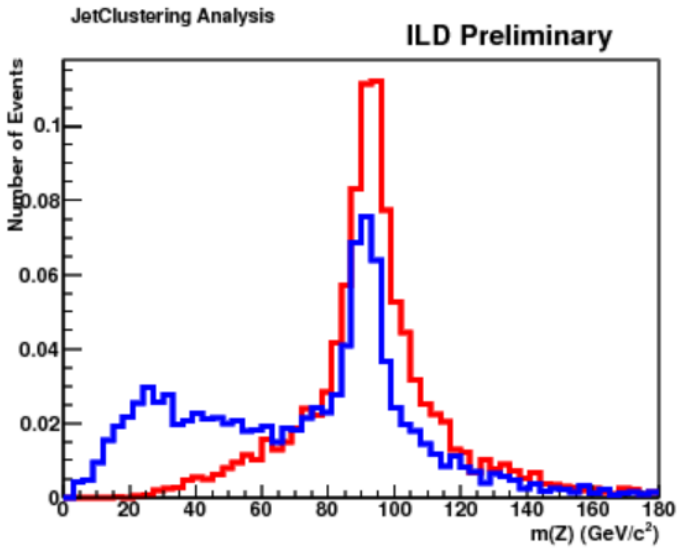
ZZH

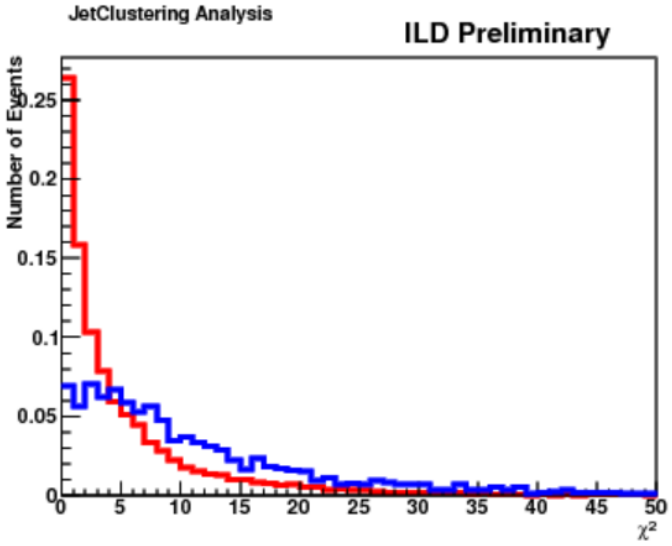
Btag>0.30 for 4 jets

SOM USING 4 JETCLU – SIGNAL V.S. ZZH

- Using qqHH and ZZH
- Use events ～12000(signal) v.s. ～40000（ZZH)







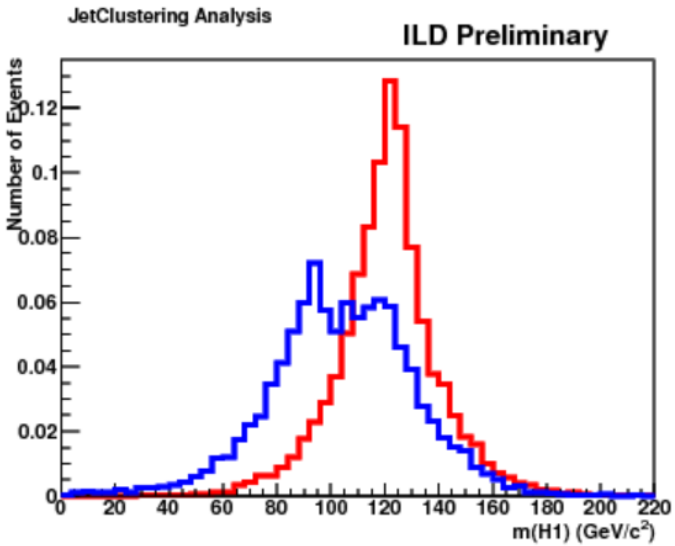
signal

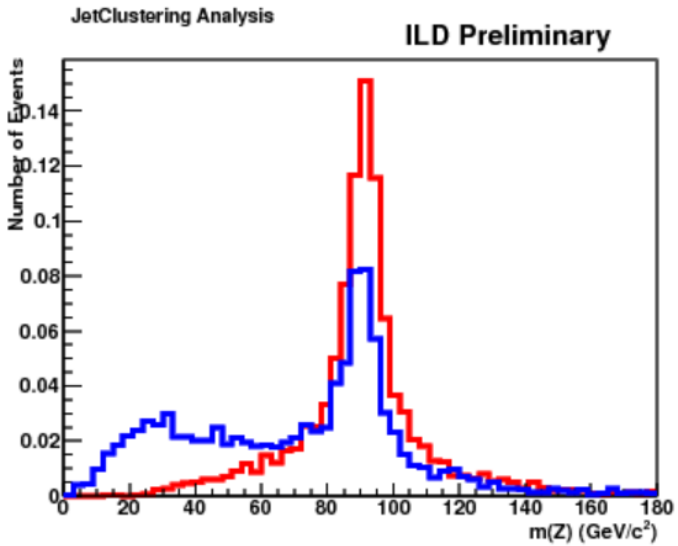
ZZH

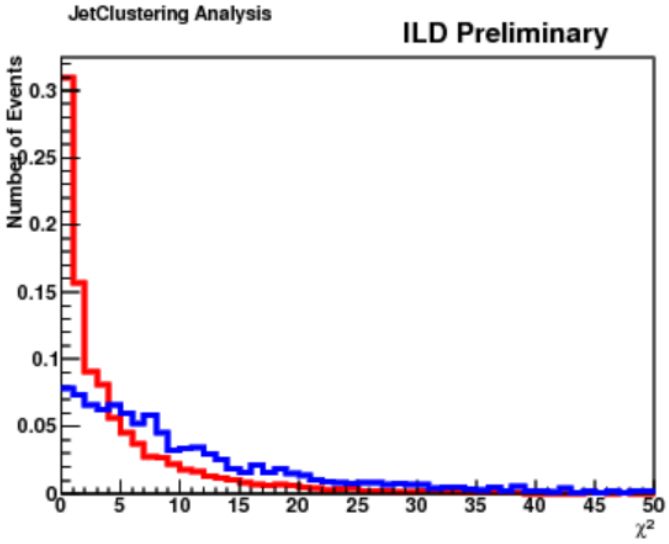
Btag>0.30 for 4 jets

SOM+CHI2 – SIGNAL V.S. ZZH

- Using qqHH and ZZH
- Use events ～12000(signal) v.s. ～40000（ZZH)







signal

ZZH

Btag>0.30 for 4 jets

Check num. of events using some SOM patterns

- Trying some patterns using SOM and mass constraint
- Check MC matching(all the direction of quarks are matched with reco. jets) and btag(>0.30) cut
- MC matching events is slightly increased using SOM
- All the patterns have almost same num. of events with btag cut
  - 8→6 clustering makes unnatural clustering possible, so num. of events is slightly decreased

Method	MCMatch	Btag>0.30	Note
SOM	6278	6165	Baseline, bias free
SOM using 4jetclu	6254	6181	best JER, but will have small bias
SOM + chi2	6211	6184	Best mass reso., but will have large bias
8→6 + SOM	6175	6078	Unnatural clustering is possible, and bias
Org. Durham	6178	6189	Nominal so far, bias free

CHECK NUM. OF EVENTS USING CHI2 CUT

- How many events are remained using chi2 cut?
  - Compare between some SOM patterns and orig. Durham
  - Check S/N simply – which one is best?
- Mass reso. and bias are trade-off
  - As long as using mass constraint only
- Mass reso. advantage? or small bias?
  - Full analysis necessary?? What is a good estimator?

qqHH	Btag>0.30	Chi2<5.0	Chi2<10.0	Chi2<15.0
SOM use4jetclu	6181	4110	5217	5634
SOM + chi2	6184	4293	5268	5692
SOM	6165	3933	5049	5516
Org. Durham	6189	3645	4800	5322

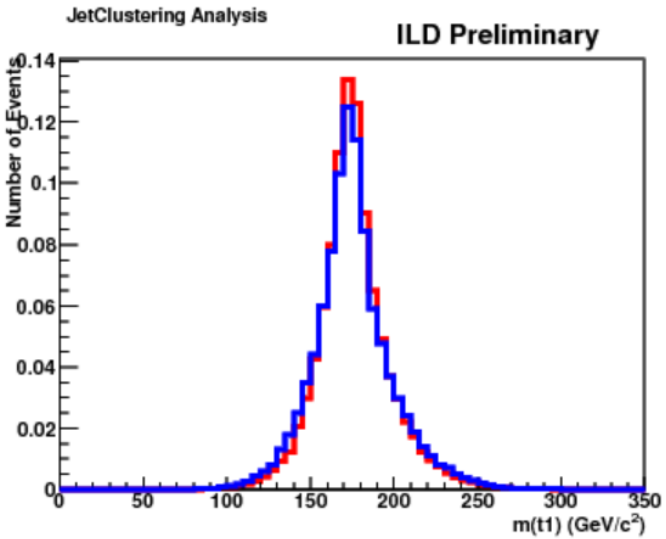
ZZH	Btag>0.30	Chi2<5.0	Chi2<10.0	Chi2<15.0
SOM use4jetclu	3301	1047	1906	2412
SOM + chi2	3309	1147	1966	2432
SOM	3300	1016	1872	2358
Org. Durham	3326	936	1836	2328

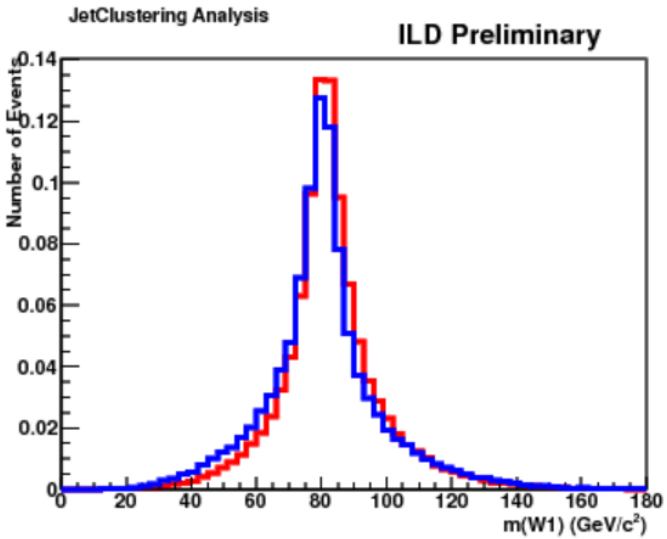
Problems and prospects

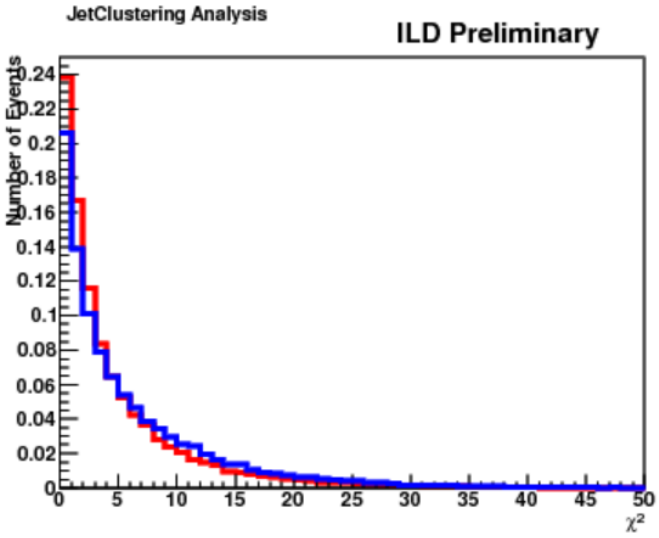
- Problems
  - SOM method doesn't reflect physics perfectly…
    - I don't know why only Jade can obtain such results…
  - I don't know the arrangement for jet clustering is good…
  - In the case of backgrounds, especially ZZH
  - In the case of other processes
    - Can obtain similar result?
  - Physics process specific
- Prospects if OK:
  - Improvement of the choice of default output vectors
    - Using Matrix element? Only mass is not enough?
    - 12→6 for example → seems bad…
  - Optimization of several parameters
    - Neighborhood function
    - Learning rate
    - Mass resolution(σ) for mass constraint
    - Etc.
- Hope for jet clustering improvement??

SOM USING 4 JETCLU – TTBAR

- Using Durham or SOM4jetsclu
- Use events ～40000(signal)







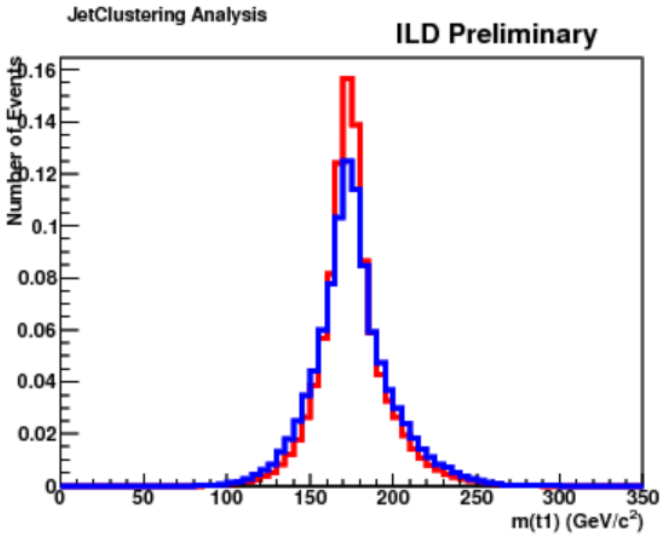
SOM4jetclu

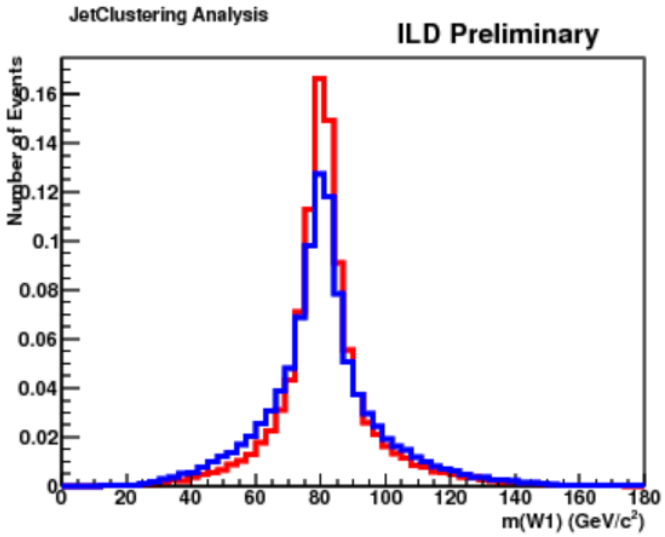
Durham

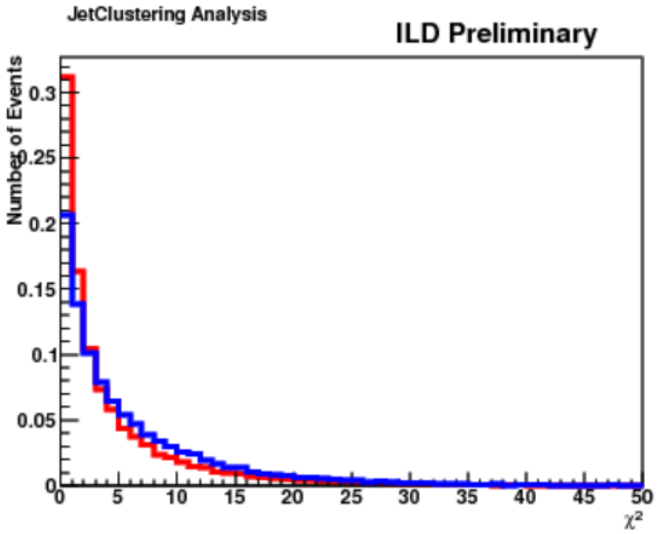
Btag>0.30 for 2 jets

SOM + CHI2 – TTBAR

- Using Durham or SOMchi2
- Use events ～40000(signal)







SOMchi2

Durham

Btag>0.30 for 2 jets

BACKUPS

Check using chi2 cut

- In SOM, basically Jade will obtain the best result!!
  - But worth trying in some patterns
- 6jet clustering
  - Using Durham, Jade, LUCLUS, and Geneva
  - Using mass constraint, min $\chi 2$ jet clustering result is used
  1. Use Jade only for SOM distance measure
  2. Use corresponding jet clustering for SOM distance measure

  e.g.) Durham jets have min. $\chi 2$ → use Durham distance measure in SOM
- 8jet clustering

	MCMatch	Btag>0.30	
SOM	6278	6165	Baseline, bias free
6+Jade SOM	6254	6181	best JER, but will have small bias
SOM +chi2	6211	6184	Best mass reso, but Will have large bias
6+correspond SOM	6228	6155	Jade is best?, small bias
8+Jade SOM	6175	6078	Unnatural clustering is possible, and bias
Org. Durham	6178	6189	Nominal so far, bias free