

Data Driven Background Determination



Isabell-A. Melzer-Pellmann





Dominant backgrounds



- ✦ $t\bar{t}$ (mainly for high N_{jet})
- ✦ W +Jets (mainly for low N_{jet})
- ✦ QCD multijet production



Method overview



- ✦ HCAL Noise suppression
- ✦ SM MET subtraction
- ✦ α_T jet balancing method
- ✦ $\chi^2(\text{ttbar})$
- ✦ Top Box Method
- ✦ Combinatorial reconstruction of ttbar / Topbox method
- ✦ Dilepton Background (in single lepton events)
- ✦ ttbar background estimation using b-tagging
- ✦ W polarisation method



HCAL Noise Suppression



Noise suppression based on precise hit timing

- ✦ Time resolution $O(1\text{ns})$ for high energies, worse for lower energies
- ✦ reject events which are out of time window around event time
- ✦ window needs to have broader shape for lower energies
- ✦ MET reduction by factor 5-10



SM MET subtraction



Largest backgrounds for lepton+jets+MET:

- ✦ $t\bar{t}$ +jets with $t \rightarrow bW(l\nu)$
- ✦ W +jets with $W \rightarrow l\nu$

Two components to MET: artificial and ν -induced

✦ Artificial MET:

- ✦ jets, detector, beam related backgrounds, non-collision effects
- ☛ Model MET with a pool of multi-jet QCD events based on:
 - ✦ N_{jet} above high p_T (e.g. 50 GeV or higher)
 - ✦ $J_T = \text{SUM } |p_{T\text{Jet}}|$ with $p_T(\text{jet}) > 20$ GeV

✦ ν -induced MET:

- ☛ Model using charged l p_T spectra (l and ν spectra should be the same if W is not polarised in transverse plane, need small corrections as W is polarised in $t\bar{t}$ decays)

Assume that the two components interfere at random angle ϕ ($0-\pi$)

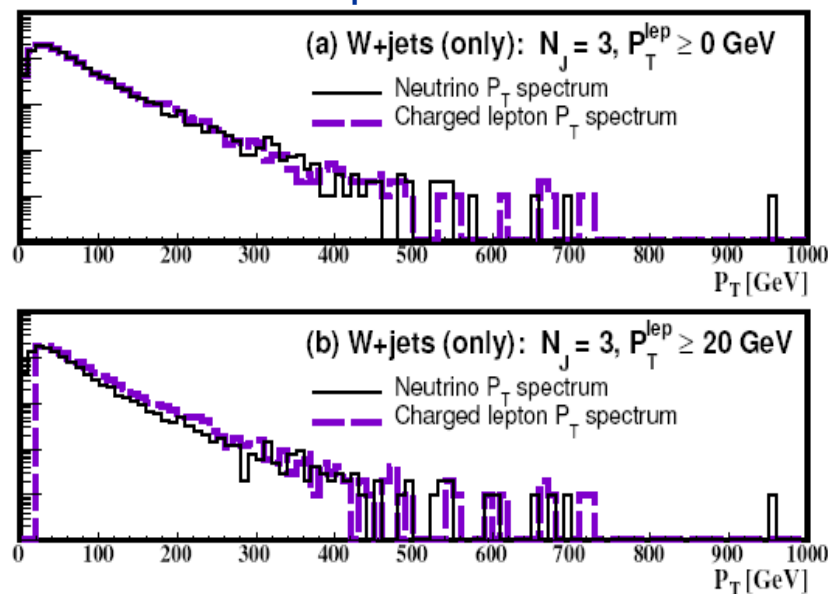
- ☛ Smear charged l p_T with artificial MET prediction at angle ϕ
- ☛ Works better for events with large N_{jet}



SM MET subtraction (2)

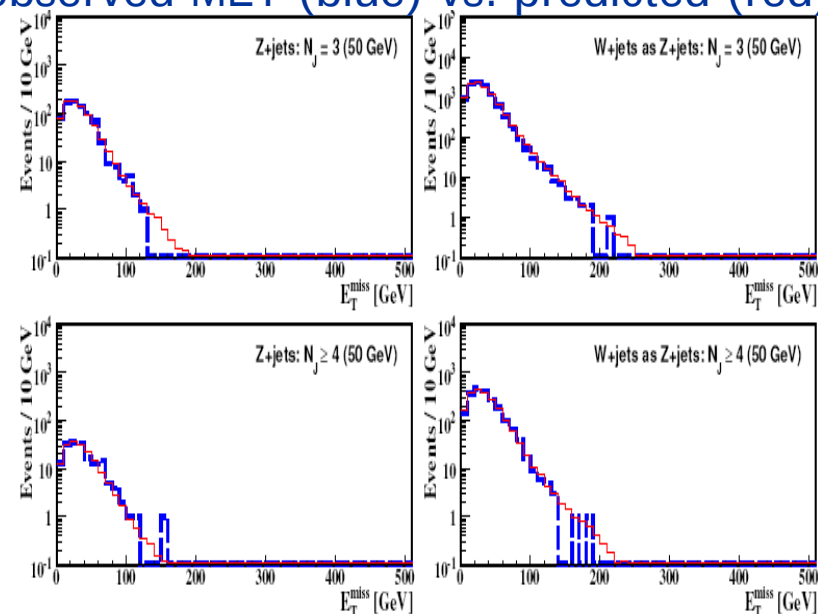


l – v comparison for W



Harder cut on charged lepton p_T
makes ν spectrum softer
➡ needs to be corrected

observed MET (blue) vs. predicted (red)



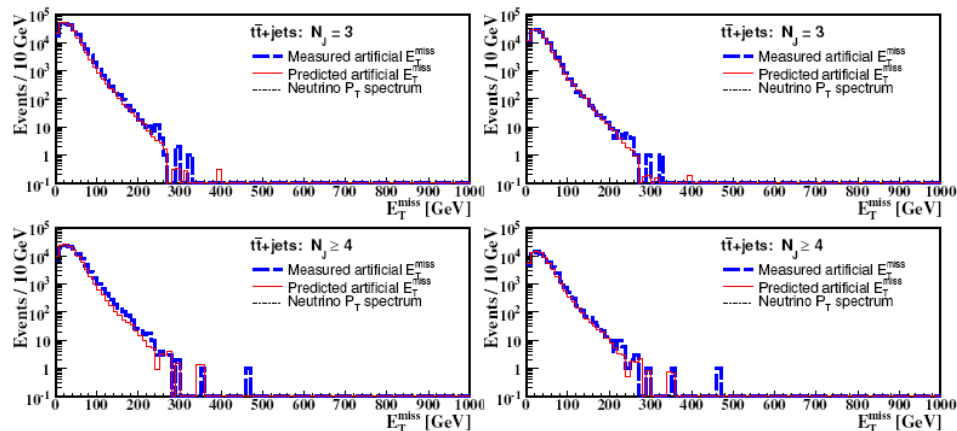
W+jets as Z+jets with ν treated
as 2nd mu using generator p_T



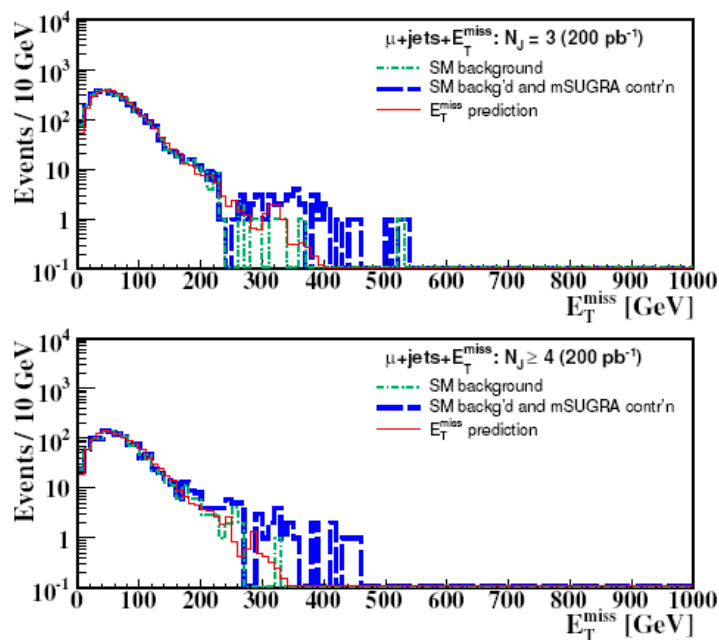
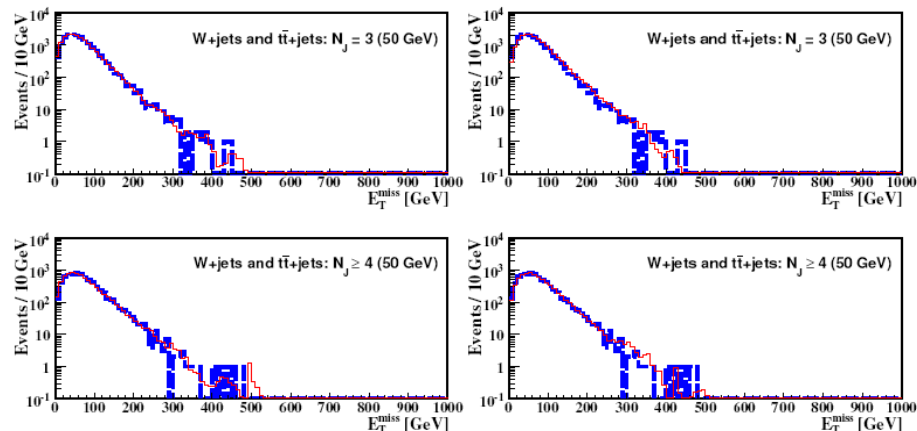
SM MET subtraction (3)



artificial MET



artificial + genuine MET



Prediction works best for models with soft lepton spectra



α_T jet-balancing method



Lepton isolation: key tool to reduce background from fake or heavy-flavor electrons and muons;

For $p_T > 30$ GeV: standard recommendations from V+jets Cross PAG

For $p_T < 30$ GeV: $\text{TrkIso}_{\text{abs}} = \sum_{\Delta R < 0.3} p_T^{\text{track}} < 3 \text{ GeV (e)} \text{ or } < 5 \text{ GeV (mu)}$

α_T : from all-hadronic analysis to reproduce kinematics in di-jet event

Idea: construct two pseudo-jets, which balance each other in H_T ,
where pseudo-jet H_T (scalar sum of p_T of all jets in pseudo-jet)

Jets are combined in pseudo-jets by minimizing $\Delta H_T = |H_{T1} - H_{T2}|$

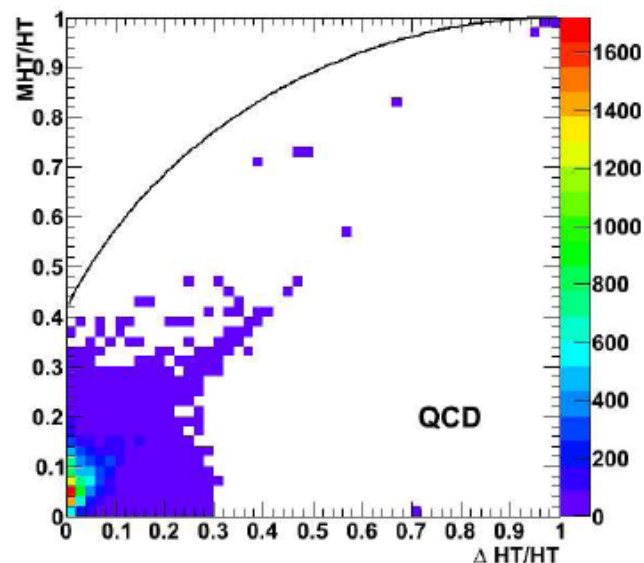
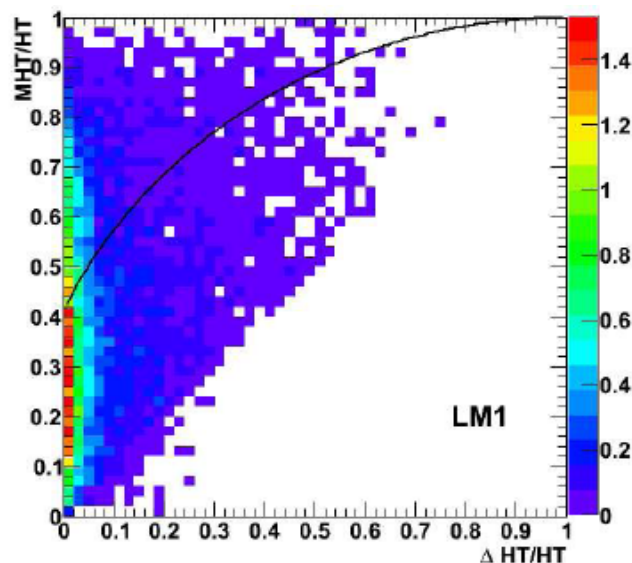
$$\alpha_T = \frac{1}{2} \frac{H_T - \Delta H_T}{M_T} = \frac{1}{2} \frac{H_T - \Delta H_T}{\sqrt{H_T^2 - M^2 H_T^2}}$$

Completely balanced system: $\Delta H_T = 0$

Leptonic α_T : similar to all-hadronic, just including lepton in variables



α_T jet-balancing method (2)

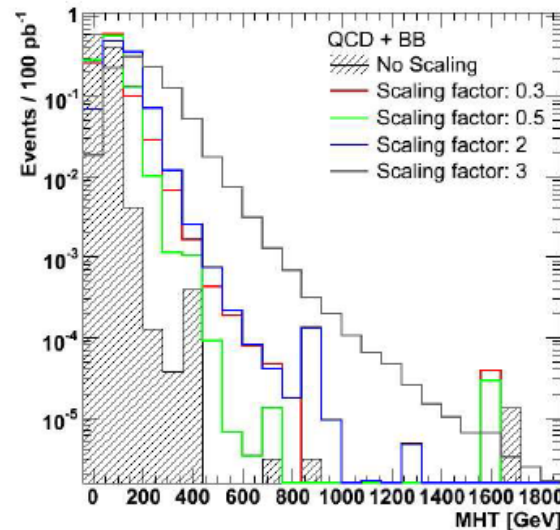
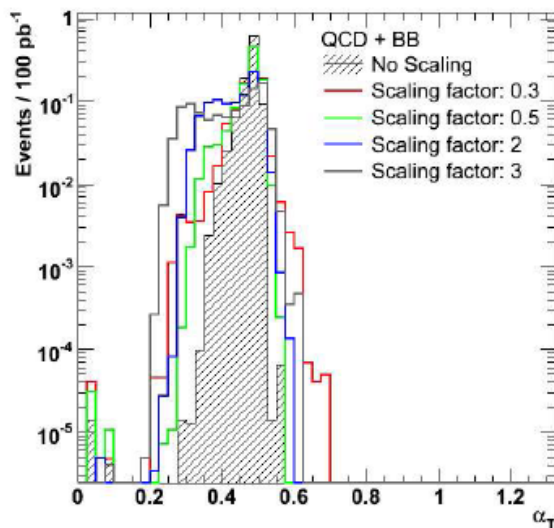
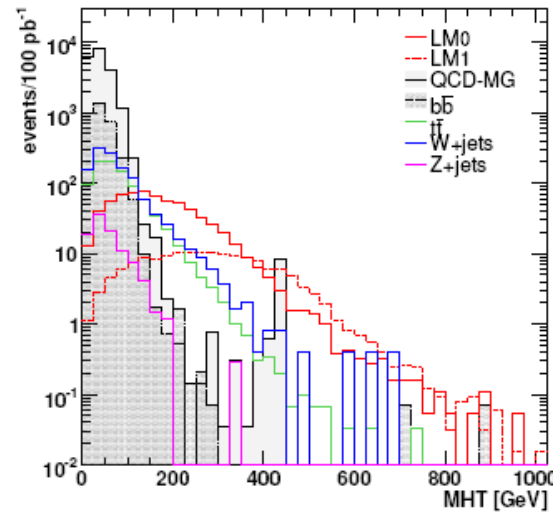
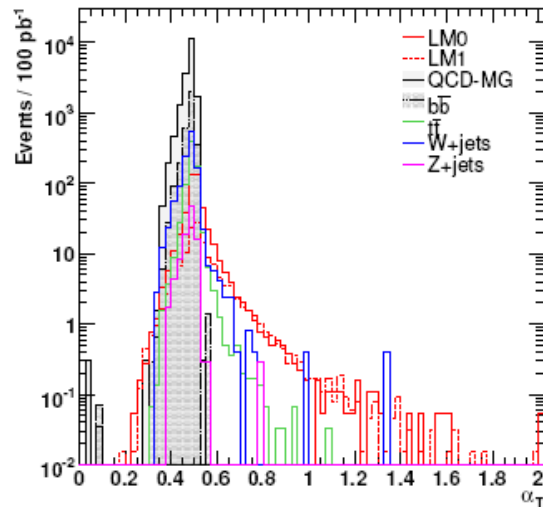
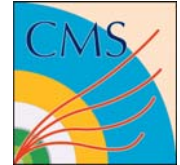


$H_T > 350$ GeV
curve for
 $\alpha_T = 0.55$

Good separation power



α_T jet-balancing method (3)



Robustness of α_T
for bad jet calibration
(MHT not so good)



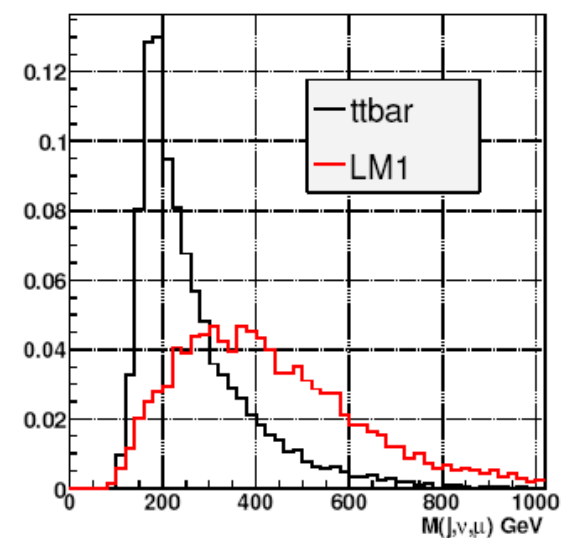
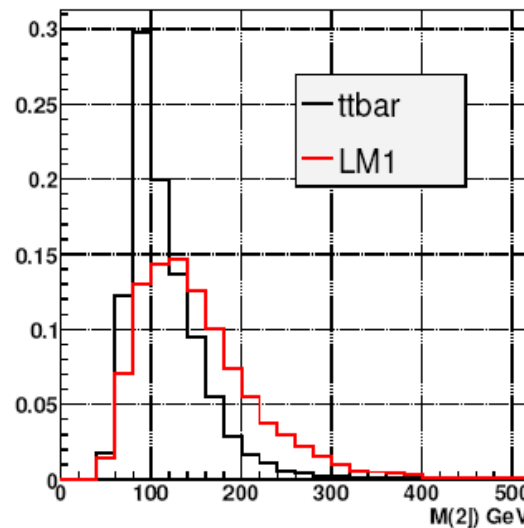
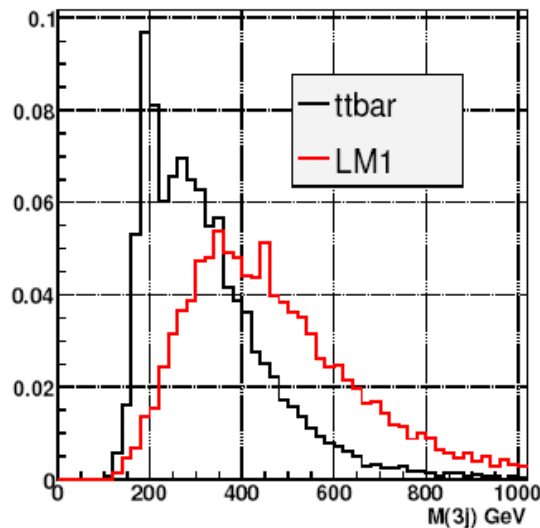
ttbar background using $\chi^2(\text{ttbar})$



Construct χ^2 by choosing the four highest pt jets and taking the lowest χ^2 permutation as the combination of jets assigned to hadronic W, b-jet of hadronic top, b-jet of leptonic top:

$$\chi^2(\text{t}\bar{\text{t}}) = \frac{(M_{j_1 j_2} - M_W)^2}{\sigma_{jj}^2} + \frac{(M_{j_1 j_2 j_3} - M_t)^2}{\sigma_{jjj}^2} + \frac{(M_{W_{\ell\nu} j_4} - M_t)^2}{\sigma_{\mu\nu j}^2}$$

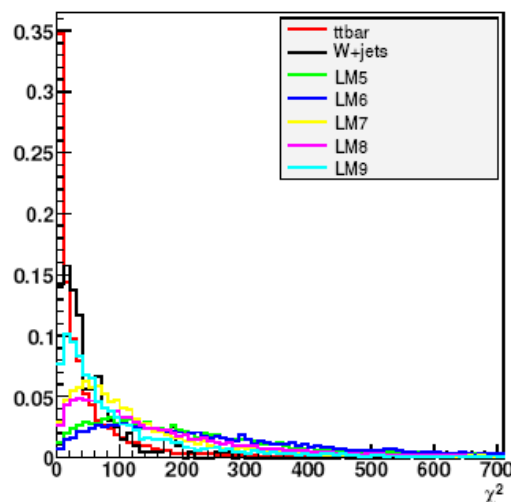
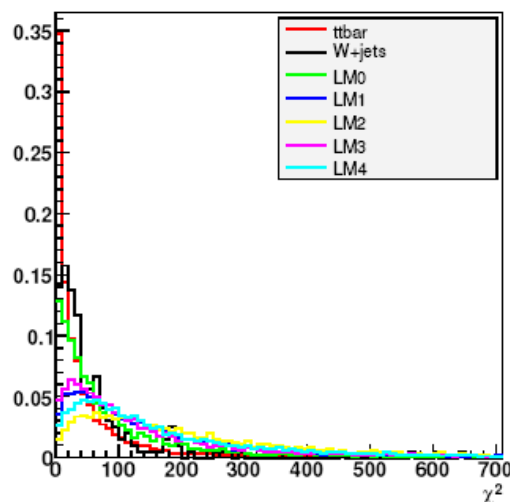
$$\sigma_{jj}^2 = 10.5, \sigma_{jjj}^2 = 19.3, \sigma_{\mu\nu j}^2 = 21.2$$



Masses for the lowest χ^2



ttbar background using $\chi^2(\text{ttbar})$ (2)



$\chi^2 < 20$:
50% ttbar
30% W+jets
2-24% SUSY

Try to mix this with 2nd variable exploiting the ABCD method

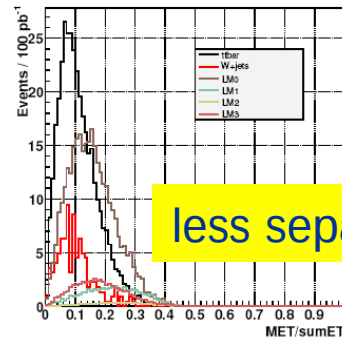
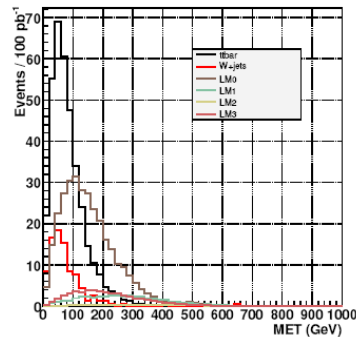


χ^2 vs MET/SUM(ET) ABCD method

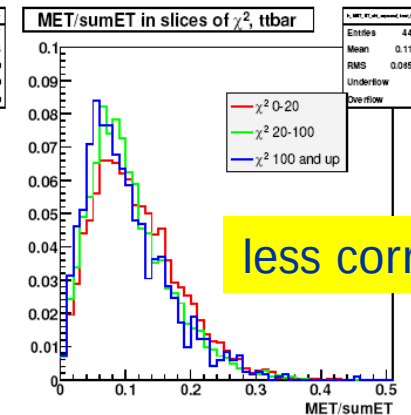
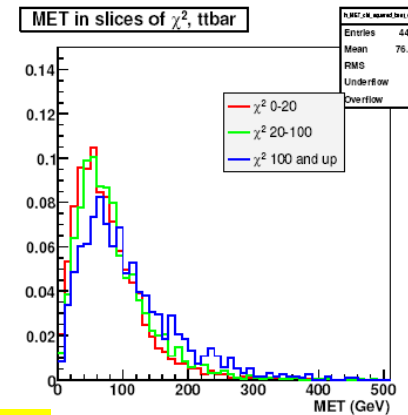


Strong correlation between χ^2 and MET, better (3 times smaller) for MET/SUM(ET)

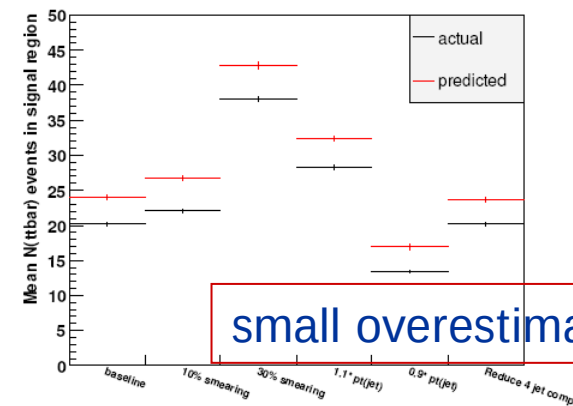
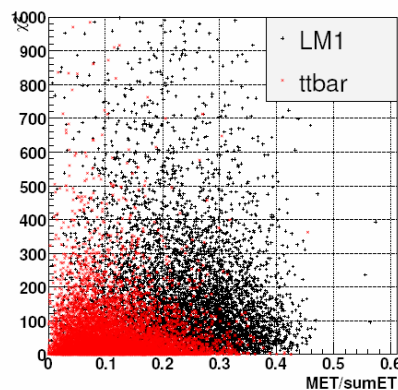
$$\sum E_T = E_T + E_T^{\text{jets}}$$



less separation



less correlation



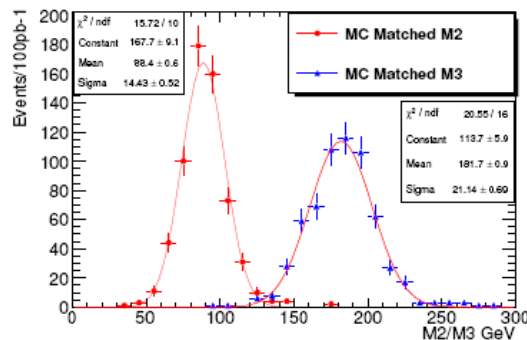
small overestimation of BG



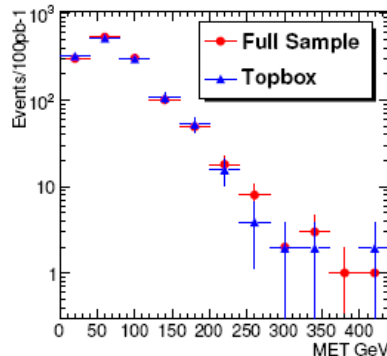
ttbar background from TopBox



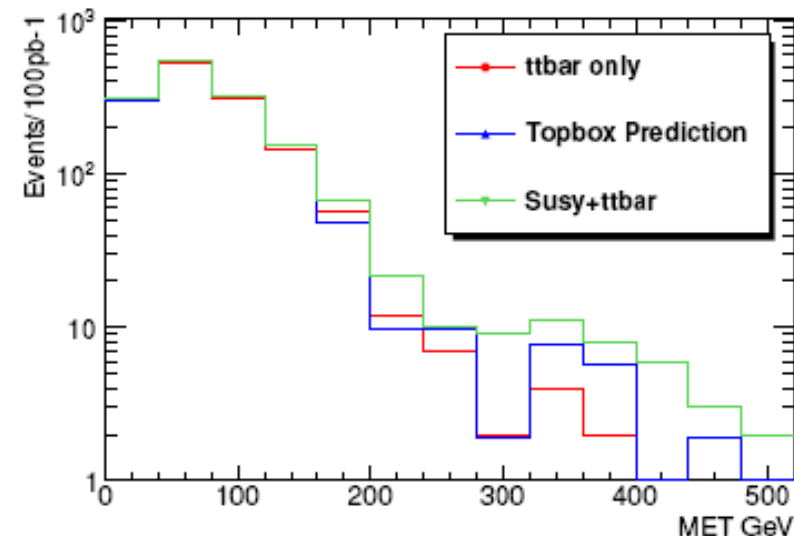
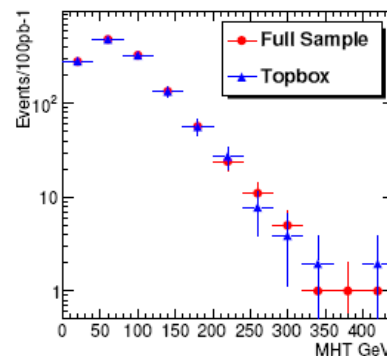
Calculate 2-jets (M2) and 3-jets (M3) invariant mass combinations
 $M_i = \text{SUM (4-Momentum(Jet) with Mass(Jet) set to 0)}$



Choose MC based cuts:
 $70 < M2 < 110 \text{ GeV}$
 $150 < M3 < 210 \text{ GeV}$



TopBox and full ttbar sample
look similar



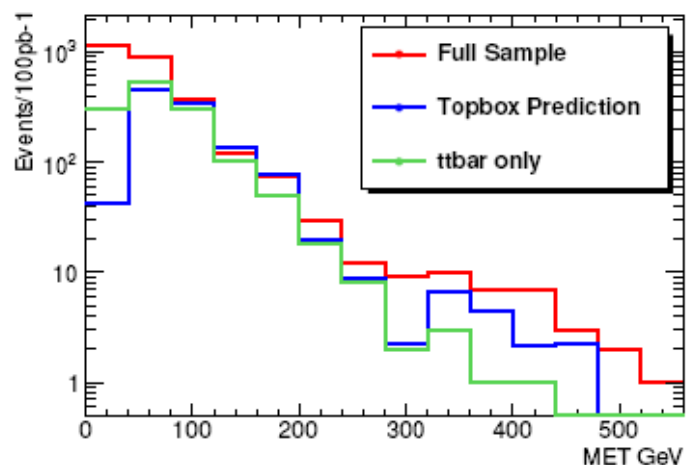
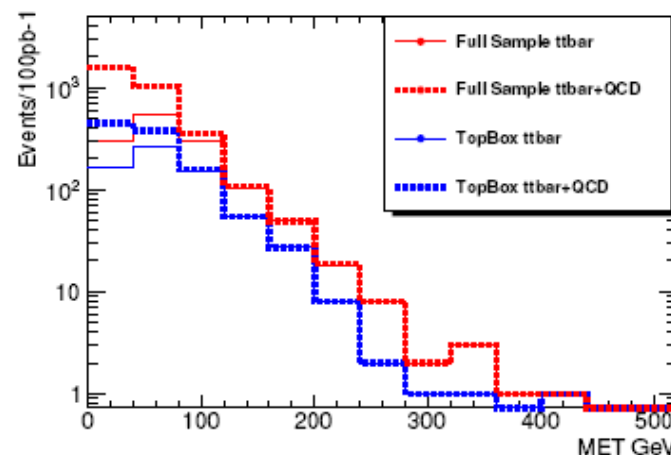
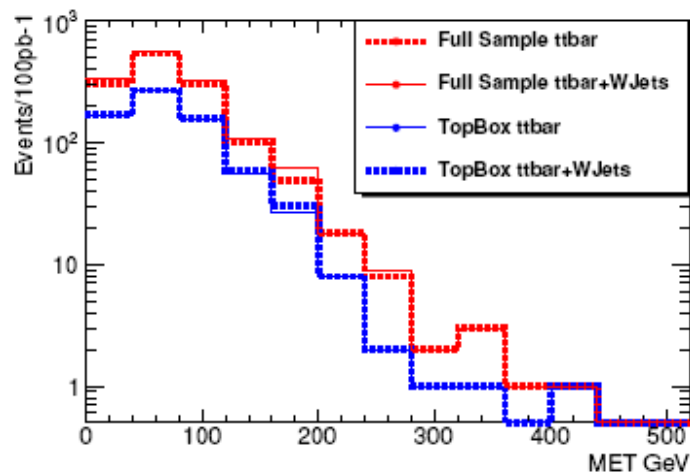
TopBox prediction with signal
contamination



ttbar background from TopBox (2)



W+jets and QCD jets contribution:



Slight overestimation
with Topbox prediction



ttbar background with χ^2 sorting and sideband subtraction



Again, use χ^2 :

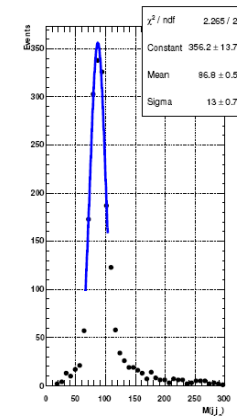
$$\chi^2(tt) = \frac{(M_{j_1 j_2} - M_W)^2}{\sigma_{jj}^2} + \frac{(M_{j_1 j_2 j_3} - M_t)^2}{\sigma_{jjj}^2} + \frac{(M_{W \ell \nu j_4} - M_t)^2}{\sigma_{\mu \nu j}^2}$$

Then use hadronic W mass for sideband subtraction (reduce SUSY signal in ttbar sample, as SUSY is relatively flat):

- ✦ signal region: $60 < M < 120$ GeV
- ✦ sideband: $120 < M < 180$ GeV

ttbar background tail also in sideband region

- ✦ subtraction will affect overall number of ttbar events
- ✦ only shape considered
- ✦ normalize sideband subtracted MET shape in low MET ($50 < \text{MET} < 100$ GeV) region where ttbar (and W+jets) backgrounds dominate

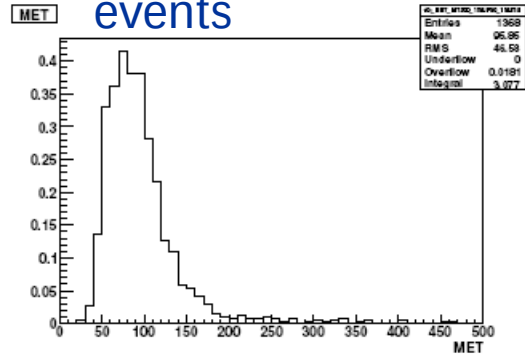




Dilepton background from ttbar

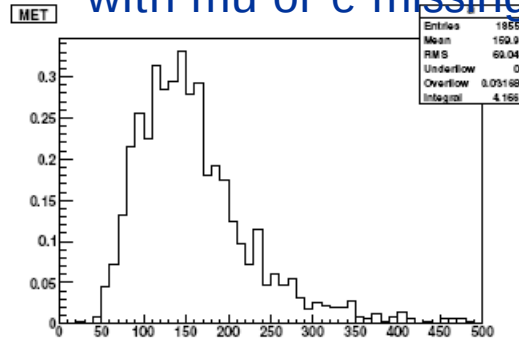


true single lepton events



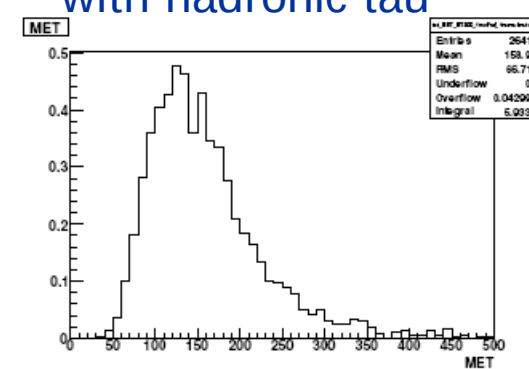
(a) MET from events with $1\nu_\mu$: $MET_{True}^{1\nu_\mu}$

fake from di-lepton with mu or e missing



(b) MET from events with $2\nu_\mu$ and $1\nu_\mu 1\nu_e$: $MET_{True}^{2\nu_\mu + 1\nu_\mu 1\nu_e}$

fake from di-lepton with hadronic tau

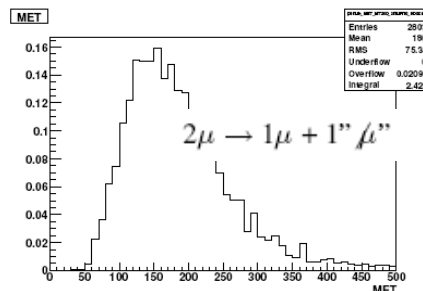


(c) MET from events with $1\nu_\mu 1\nu_\tau$: $MET_{True}^{1\nu_\mu 1\nu_\tau}$

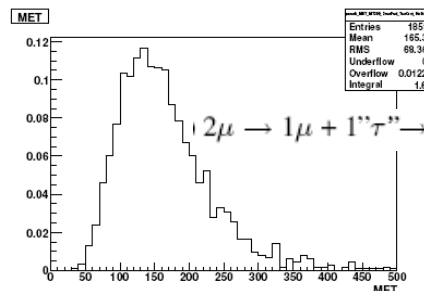
Recipe:

take di-muon sample (from data) and replace one muon:

- by adding muon pT to MET and recalculate MT (do this for each muon)
- simulate one mu as hadronic tau by adding half of the energy to MET and use the other half for a jet (usually < 50 GeV and below cuts)



$2\mu \rightarrow 1\mu + 1'\mu$

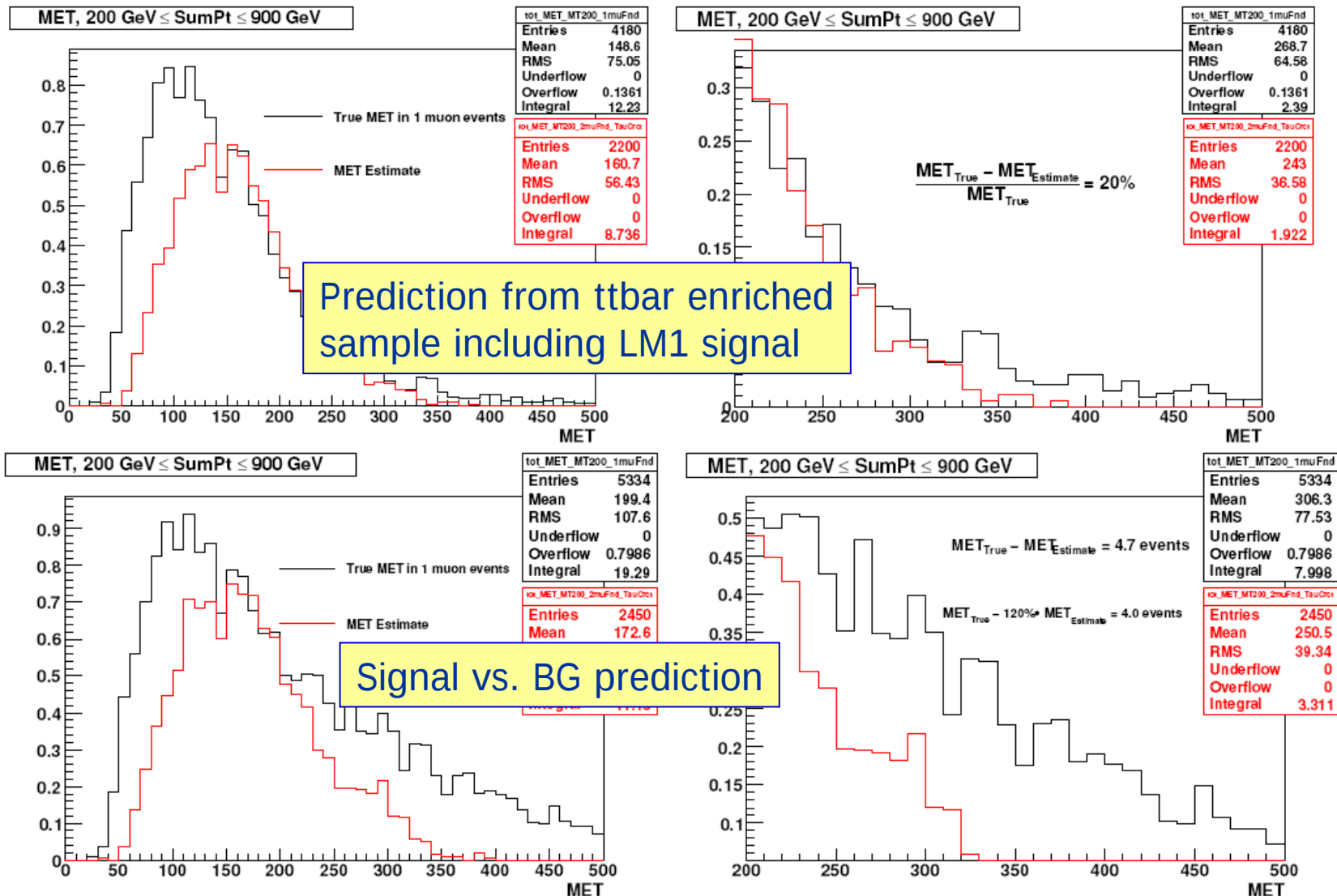


$2\mu \rightarrow 1\mu + 1'\tau \rightarrow \text{jets}$

add these two distributions as $\alpha * MET(\text{mumumiss}) + \beta * MET(\text{mutau})$
Ratio taken completely from MC, distributions extracted from data!



Dilepton background from ttbar (2)





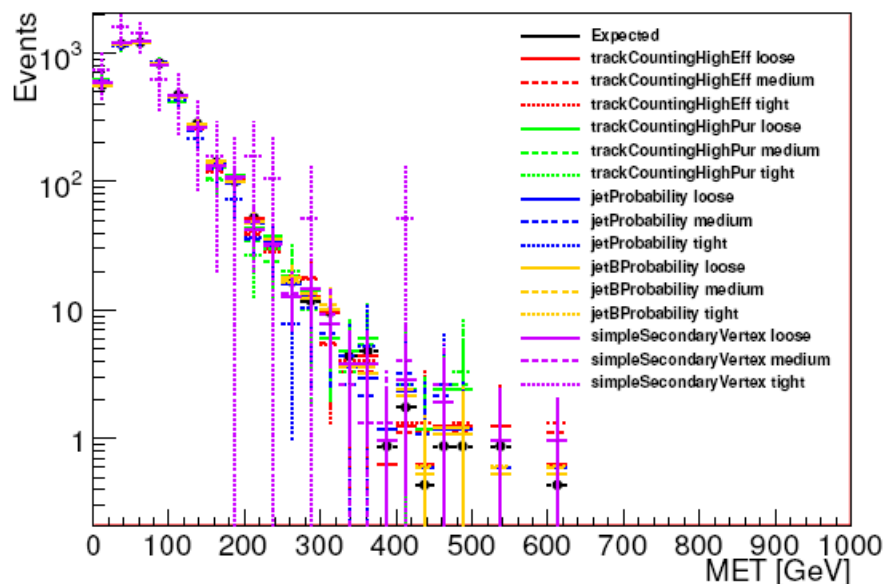
ttbar background using b-tagging



Determination of MET in Control Region [0:100 GeV]

3 Variables useful to determine background from semi- and di-leptonic ttbar background:

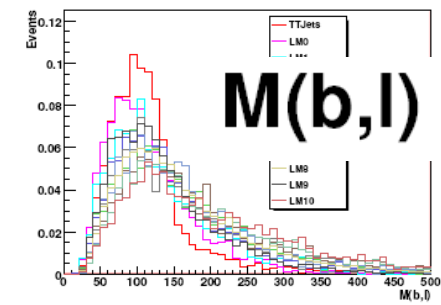
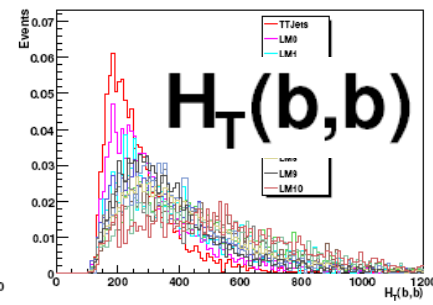
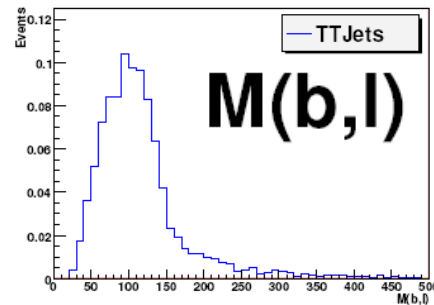
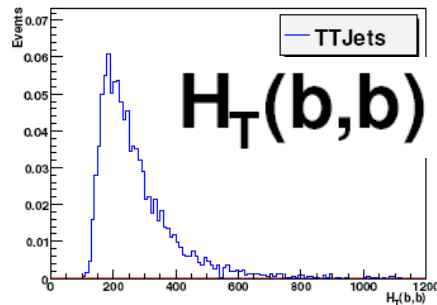
- $H_T(b, b)$: the scalar sum of the P_t of the 2 b-tagged jets
- $\Delta R(b, b)$: the angle between the 2 b-tagged jets
- $Mass(b, l)$: the invariant mass of the muon and the closest b-tagged jet



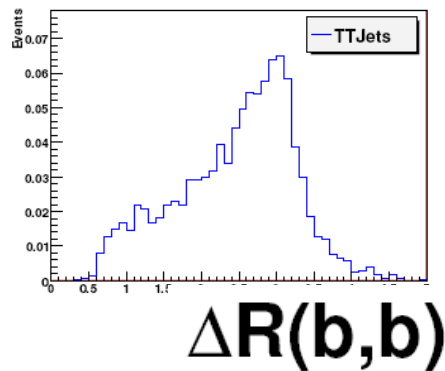
MET is independent from b-tagging!



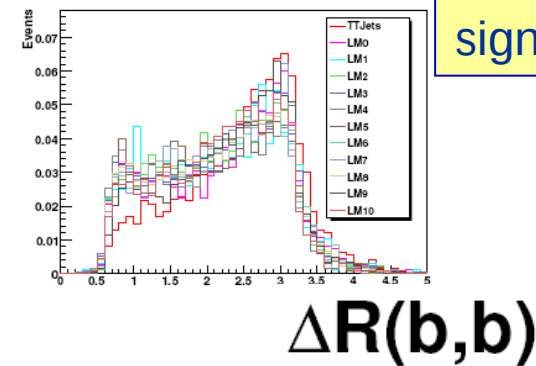
ttbar background using b-tagging (2)



ttbar



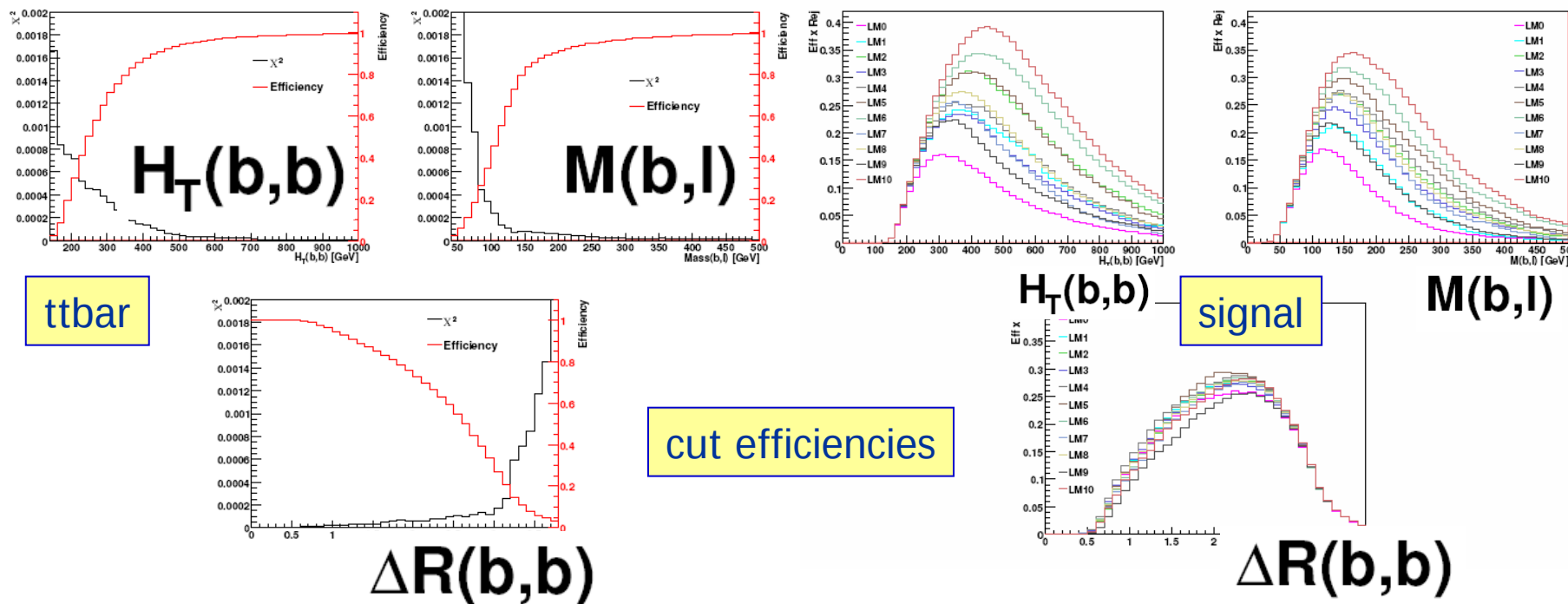
signal



Use these variables for a kind of ABCD region BG estimate

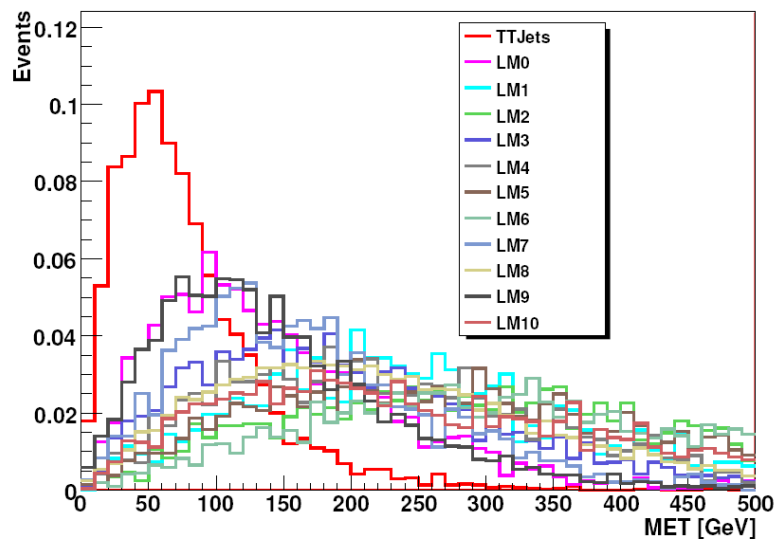


ttbar background using b-tagging (3)

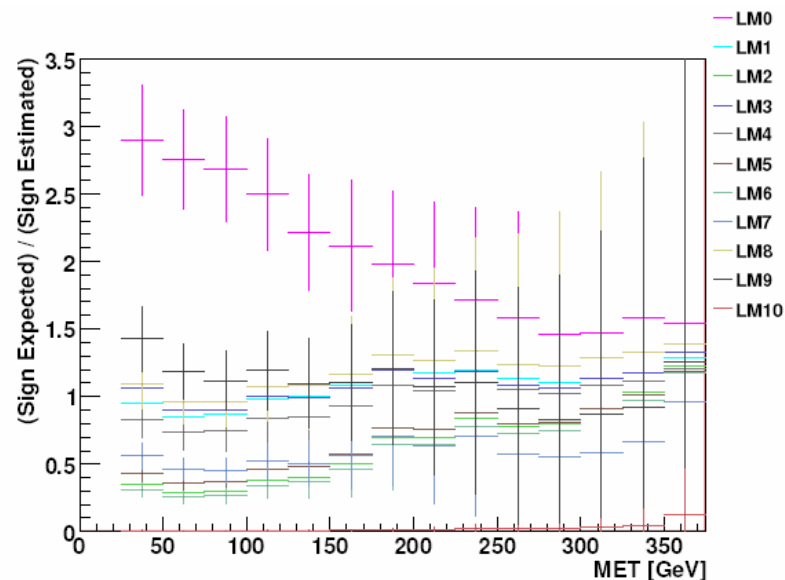
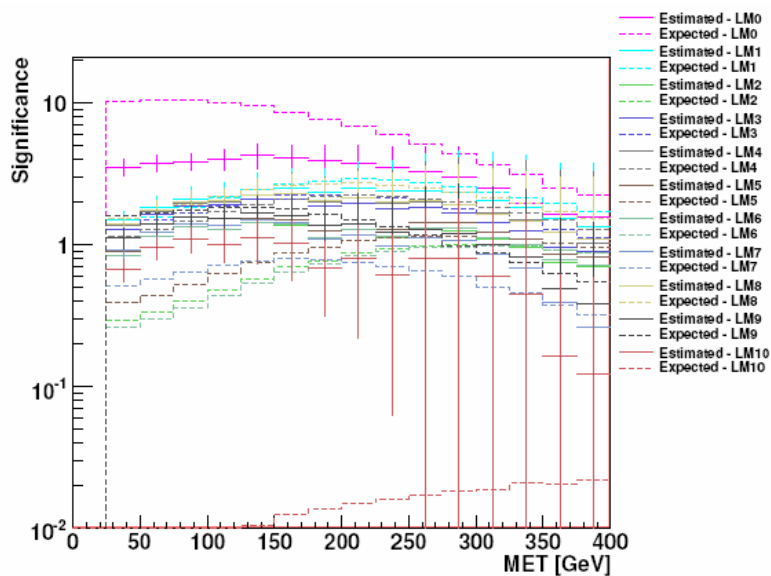




ttbar background using b-tagging (4)



MET is always signal contaminated
→ calculated significance based on
different MET cuts





W Polarisation



See talk from Markus Stoye:

<http://indico.cern.ch/getFile.py/access?contribId=5&resId=0&materialId=slides&confId=76024>



Formulae collection





Important formulas



MET for muons:

$$MET = |p_{miss}| = \left| - \left(\sum_{calotowers} p_T - \sum_{muons} p_T^{deposit} - \sum_{muons} p_T \right) \right|$$

estimated CAL energy deposits of muons are removed and replaced by the measured pT from the global track fit

$$H_T = \sum_{4 \text{ highest } p_T \text{ jets}} |p_T|$$

$$M_{eff} = \sum_{4 \text{ highest } p_T \text{ jets}} |p_T| + p_T^{\text{single lepton with } p_T > 10 \text{ GeV}}$$

$$\sum E_T = \cancel{E}_T + jets$$