Boosted Top Quarks in CMS

Francisco Yumiceva
(Fermilab)

Oct 8, 2008 UMD/JHU HEP Seminars

CMS $m_{t\bar{t}}$/Boosted Top Group

Aachen	M. Erdmann, A. Hinzmann, J. Steggemann
CNRS/IPN Lyon	R. Chierici*, S. Perries, E. Chabert, F. Fassi, T. Le Grand
Fermilab	D. Green, F. Yumiceva*
Johns Hopkins	D. Fehling, G. Giurgiu, P. Maksimovic, S. Rappoccio, M. Swartz
Karlsruhe	T. Peiffer, J. Wagner-Kuhr, J. Ott, T. Muller
Maryland	N. Hadley, M. Kirn, J. Temple
UC Davis	J. Conway, J. Dolen, M. Searle, M. Squires, R. Vasquez
UC Riverside	J. Babb, S. Kao, S. Wimpenny

* Contact persons

Outline

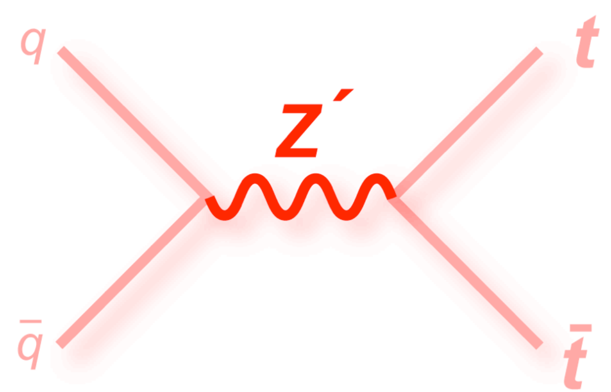
- Motivation
  - Kinematic distributions
- Reconstruction of Boosted top quarks
  - jet algorithm
  - b- tagging
- Overview of analyses and roadmap to discovery
  - Early data analysis
  - Searches at low-medium and high $m_{t\bar{t}}$ regions.
- Summary

Motivation: New physics

In many models, there is a high potential to discover new physics in the top sector by searching for heavy resonances:

$$pp \to X \to t\bar{t}$$

- We expect new physics to have strong couplings to the top sector because of the large top Yukawa coupling.



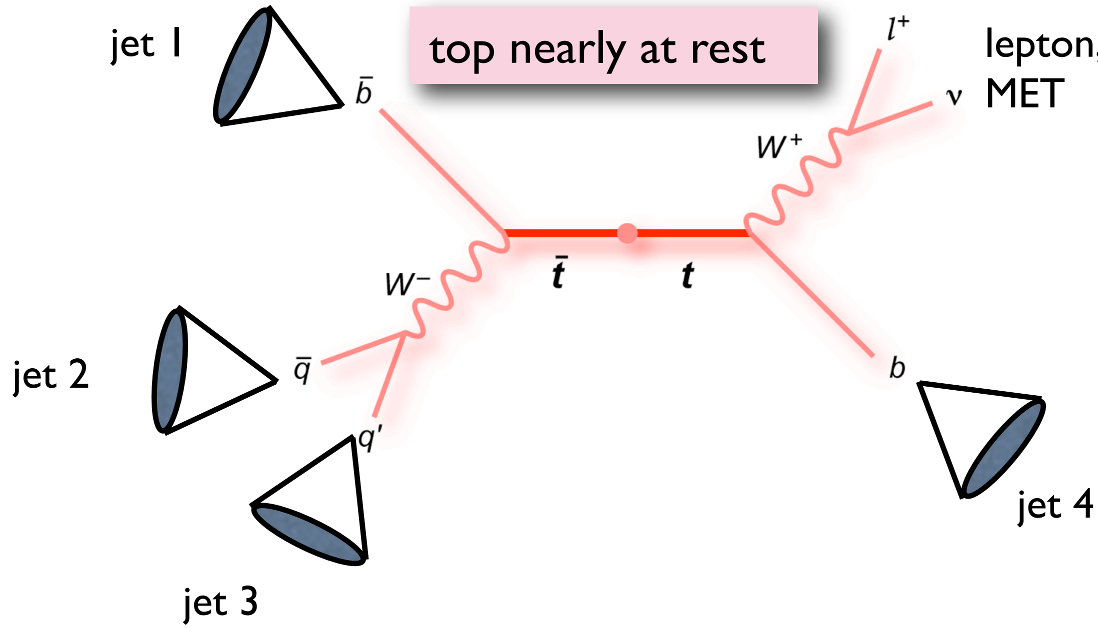
Detection of these high-mass exotic resonances requires the selection of top quark pair which are highly boosted:

This is experiementally very challenging!

The Challenge

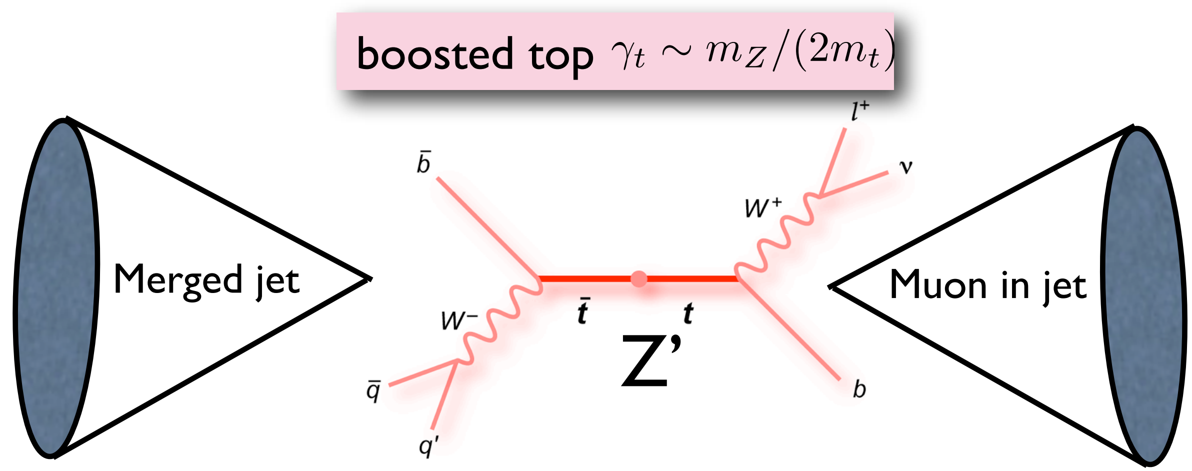
Standard model

e.g. top semileptonic decay channel: four jets, isolated lepton, missing Et



New physics: a heavy resonance

e.g. top semileptonic decay channel: di-jet topology, merged jets, lepton in jet.



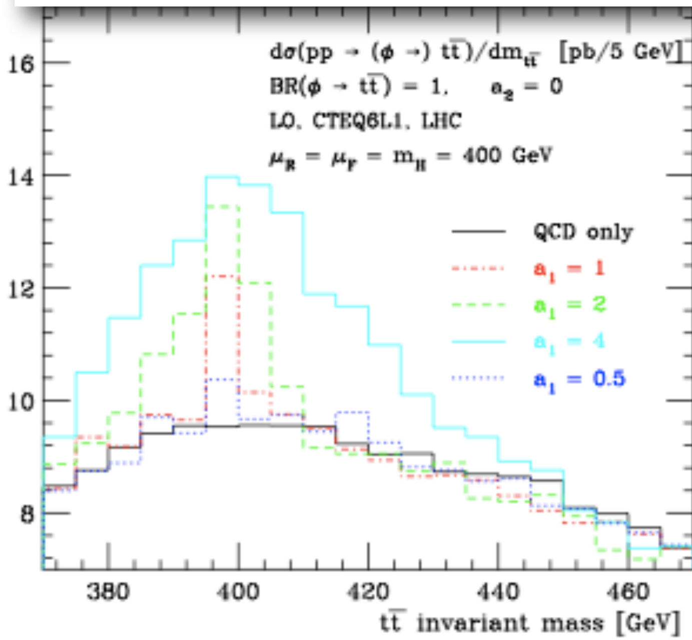
Can we reconstruct efficiently these jets?
Can we get a good Signal/Background ratio?

Observable for searches

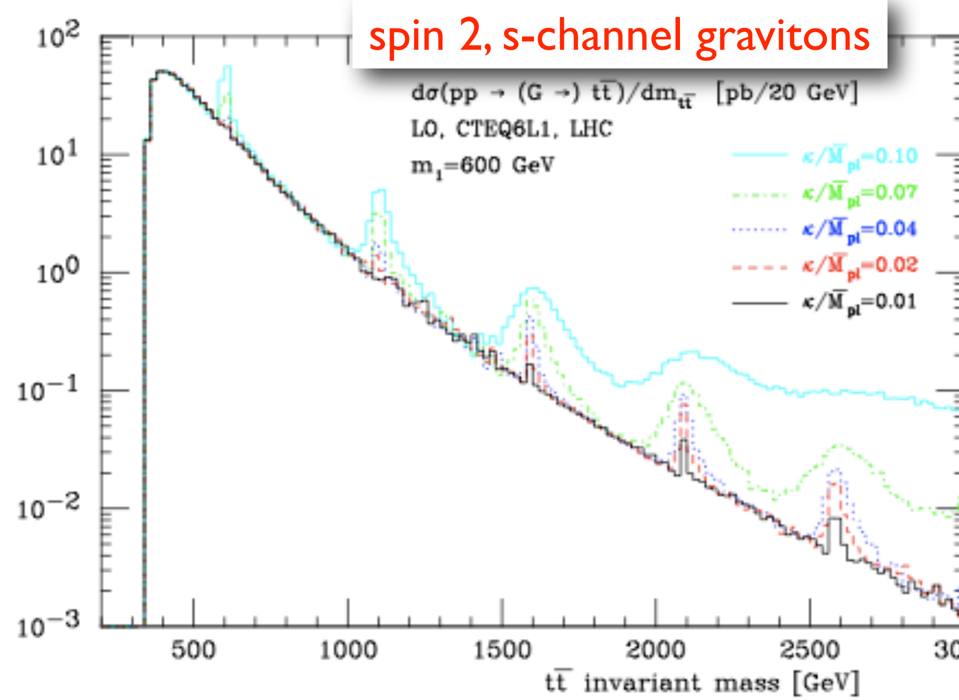
New phenomena can be observed in the $m_{t\bar{t}}$ distribution as shape distortions or peaks.

- Distortions can be deviations from the theoretical prediction due to enhancements or interferences of new physics.

400 GeV color singlet spin 0, scalar



spin 2, s-channel gravitons



R. Frederix and F. Maltoni [arXiv:0712.2355]

The $m_{t\bar{t}}$ distribution is a good observable and we will use it to carry on a model independent search.

Why are these studies important at this time?

We need to be ready for discovery, (good) surprises may appear any time.
e.g. observation of first high-pT jets, those could be boosted top jets!

Strain our current tools in special scenarios.
e.g. know beforehand the limits of our current techniques and algorithms.

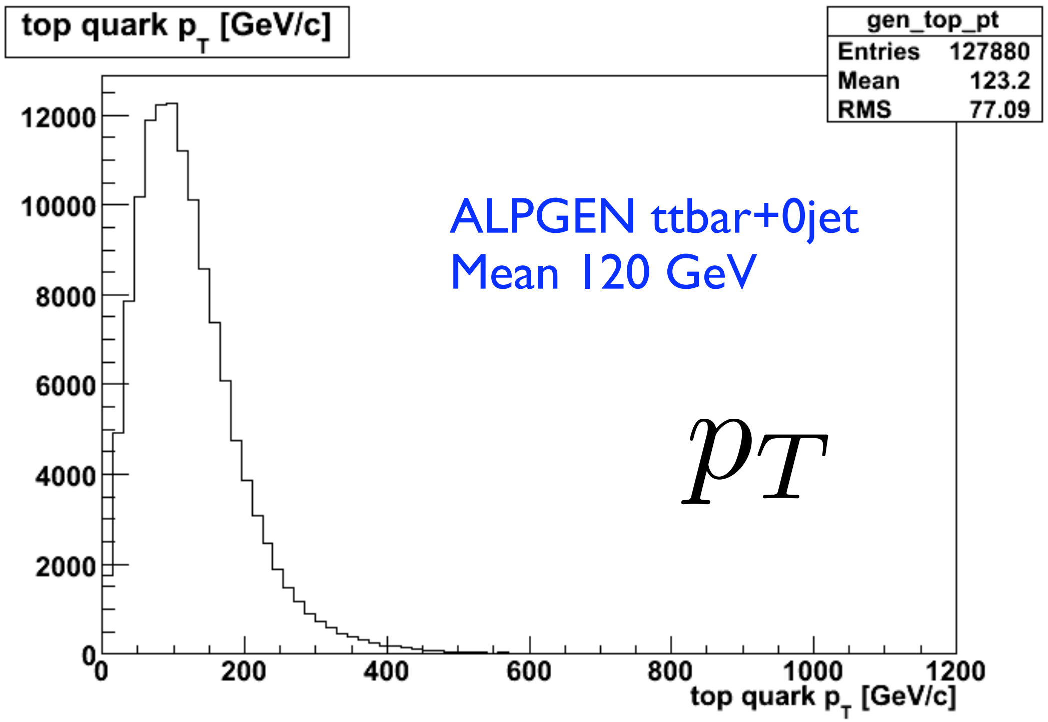
Implementation of new algorithms that could be useful for other analysis
e.g. new discriminant variables, new jet algorithms.

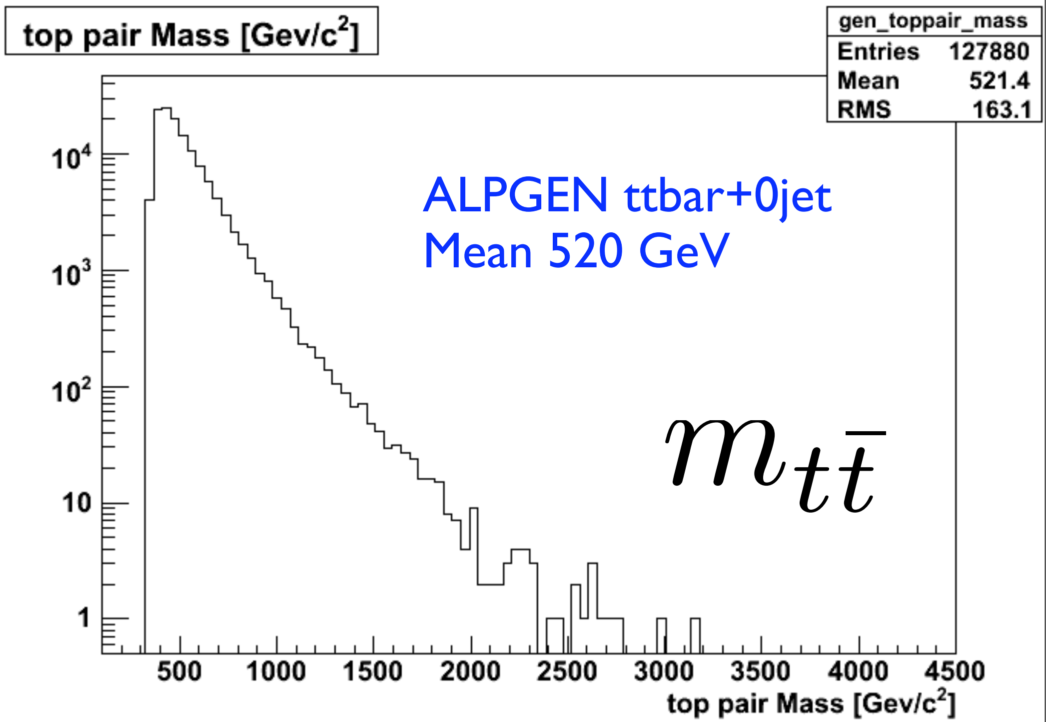
Complementary studies for standard analyses:
e.g. study of non-isolated muons, reconstruction and ID of muons in jets.

Bread & butter analyses (which can be done with first data samples)
e.g. mttbar distribution, top mass using mttbar distribution.

Kinematic Regions

Let's first get familiar with the spectrum of top decays from continuum QCD and from exotic resonances in the LHC.

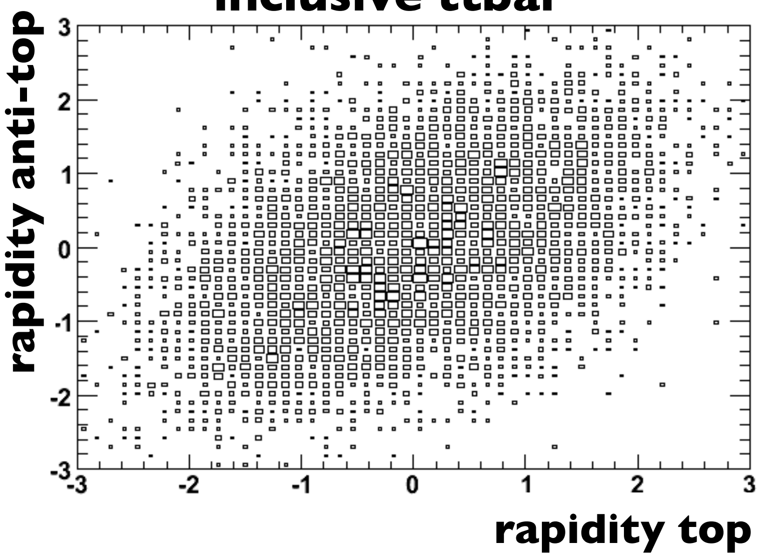


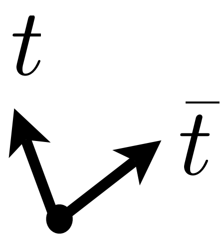


Top pairs from continuum have a mean pT ~ 120 GeV with most of top quarks below 400 GeV, and the top pair invariant mass has mean of ~ 520 GeV with a high tail extending below 2.5 TeV.

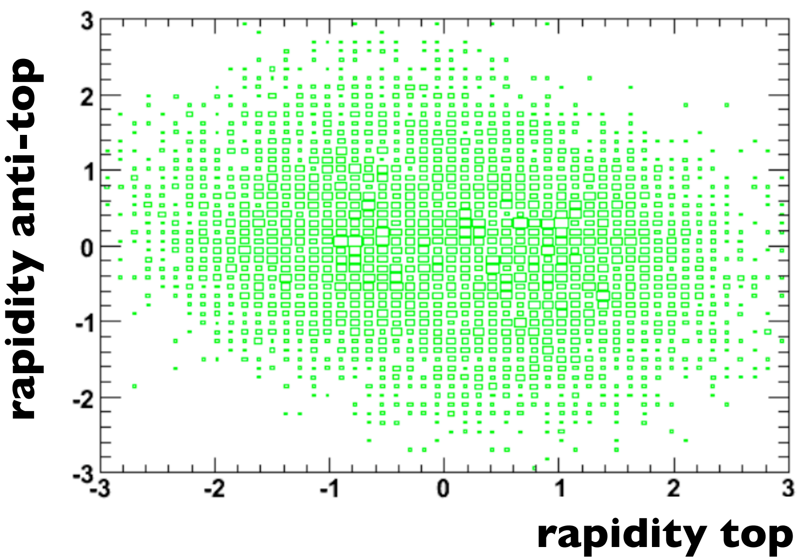
Kinematics (2): rapidity of top pairs

inclusive ttbar

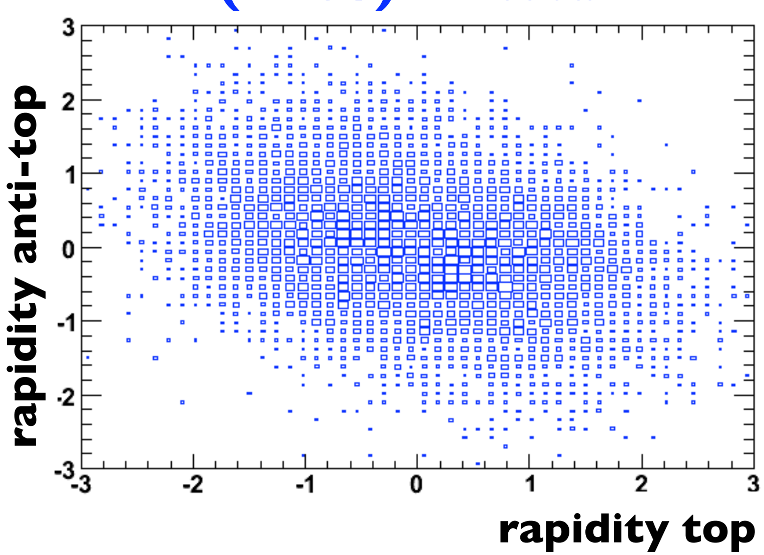


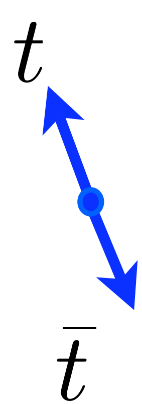


Z'(3 TeV) -> ttbar



Z'(4 TeV) -> ttbar

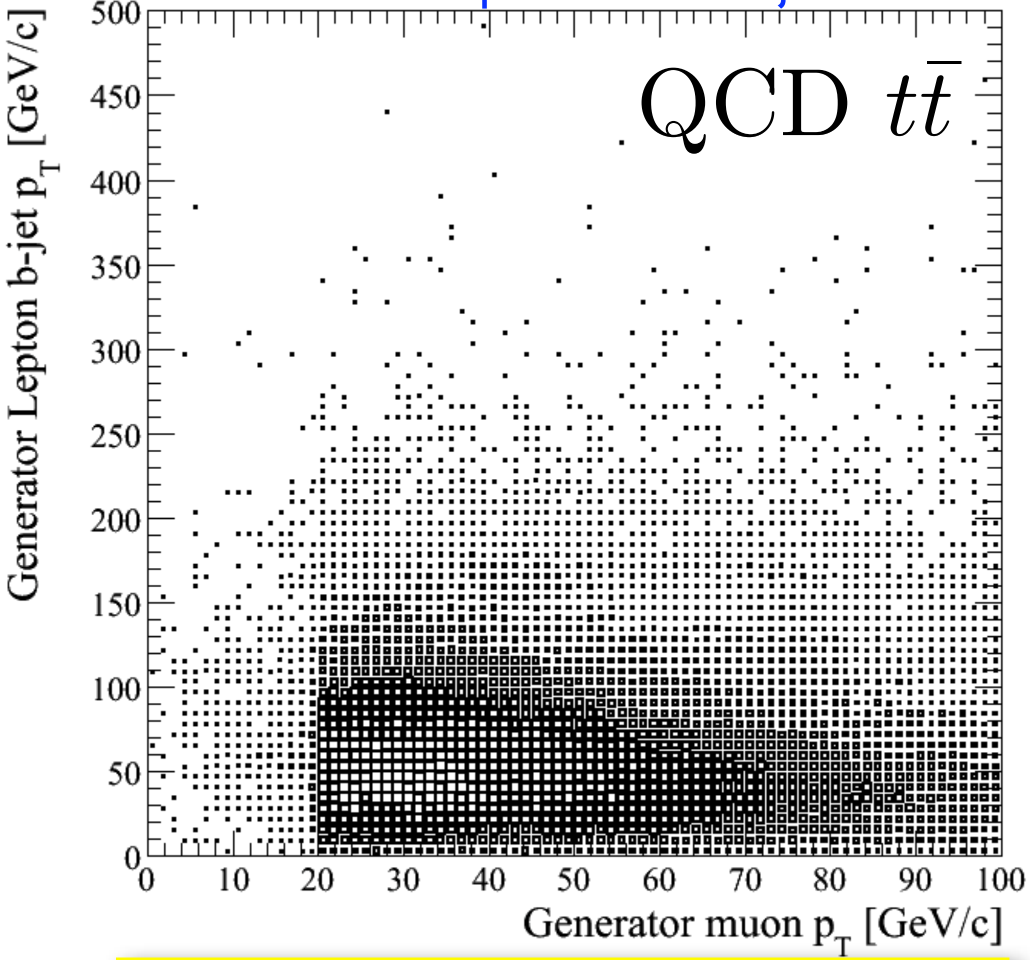




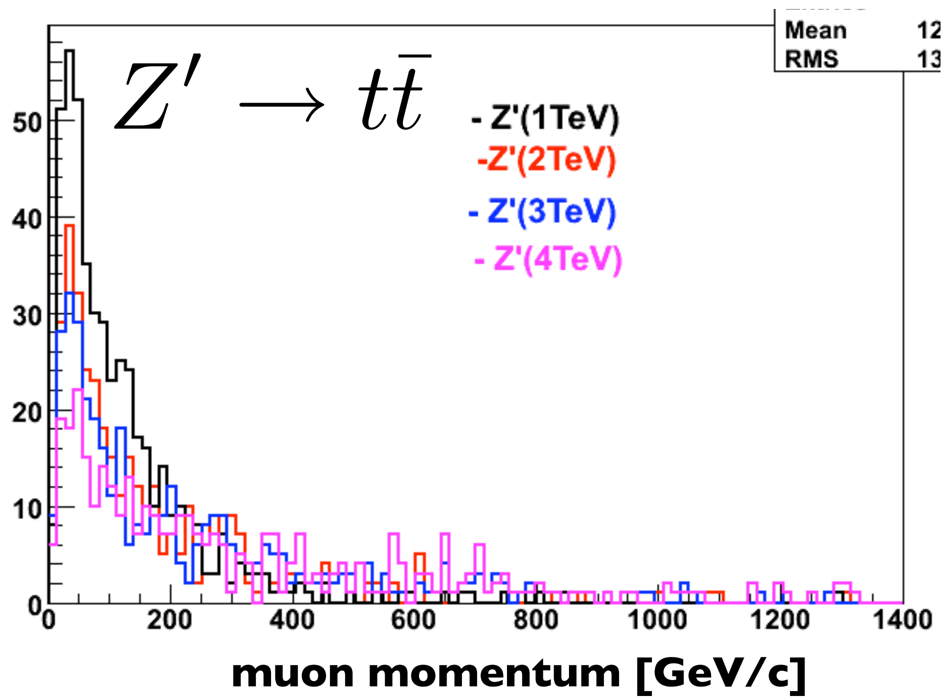
Top pairs are produced centrally and tend to be back-to-back in rapidity as the mass of a resonance increases.

Kinematics - Semileptonic Decays: Muon momentum

Muon semileptonic ttbar+0jet



Muon momentum from semileptonic Z' decays



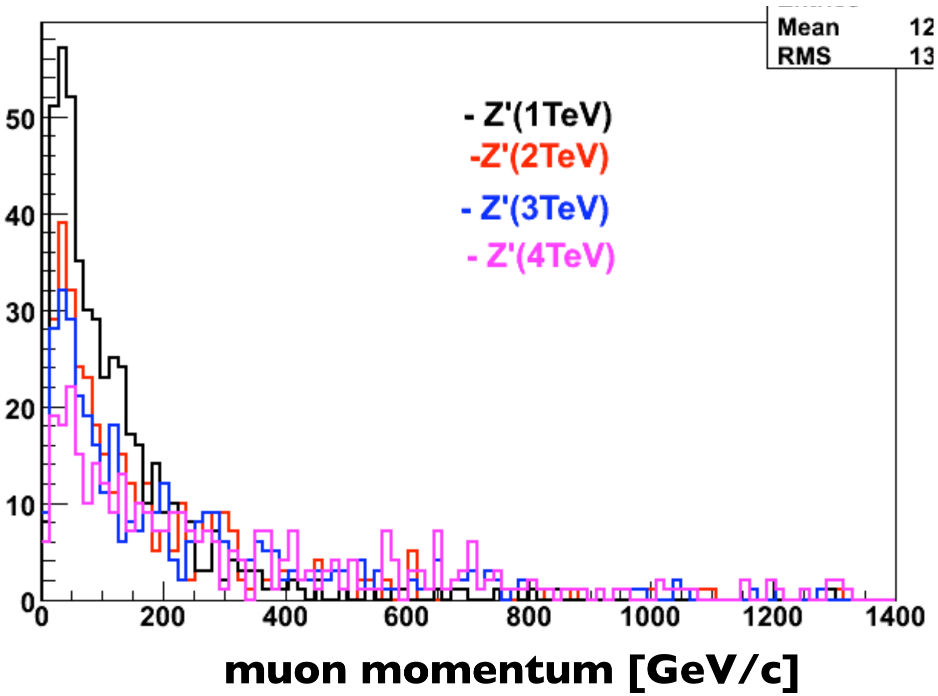
Isolated muon from W decays has mostly a pT > 20 GeV and below 90 GeV.

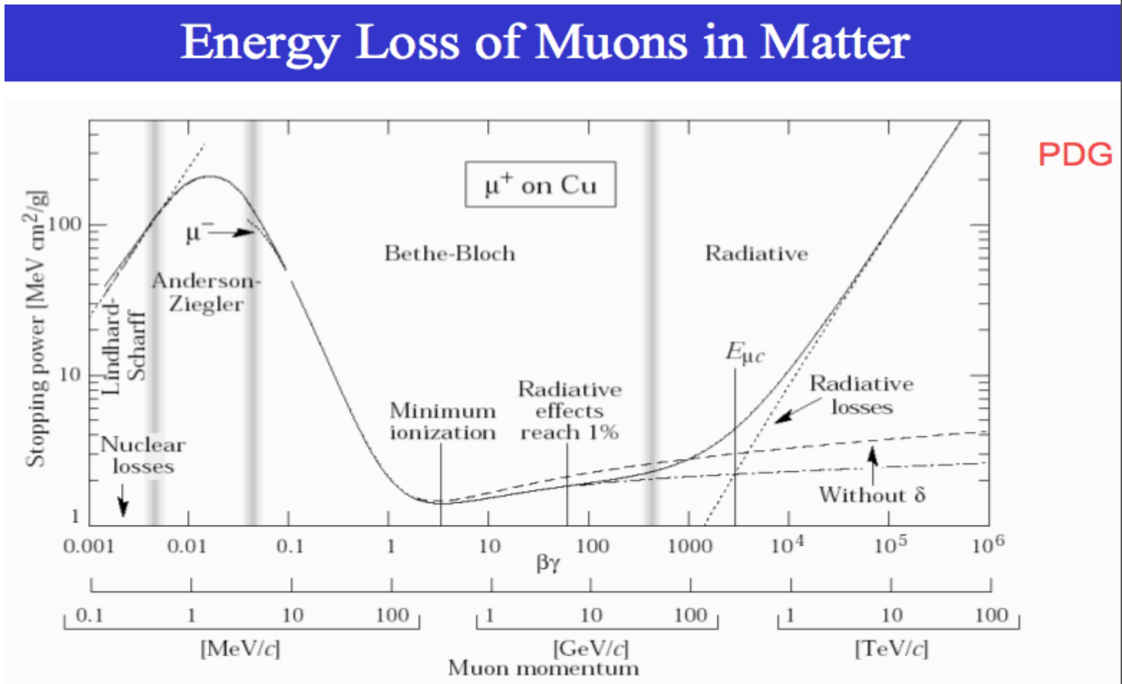
The b-jet from the leptonic top decay has a mean pT of 50 GeV.

Muons from high-pT top decays. The spectrum is harder than muons from QCD top pairs but still below the extreme region < 300 GeV.

Kinematics (4): Muon Spectrum

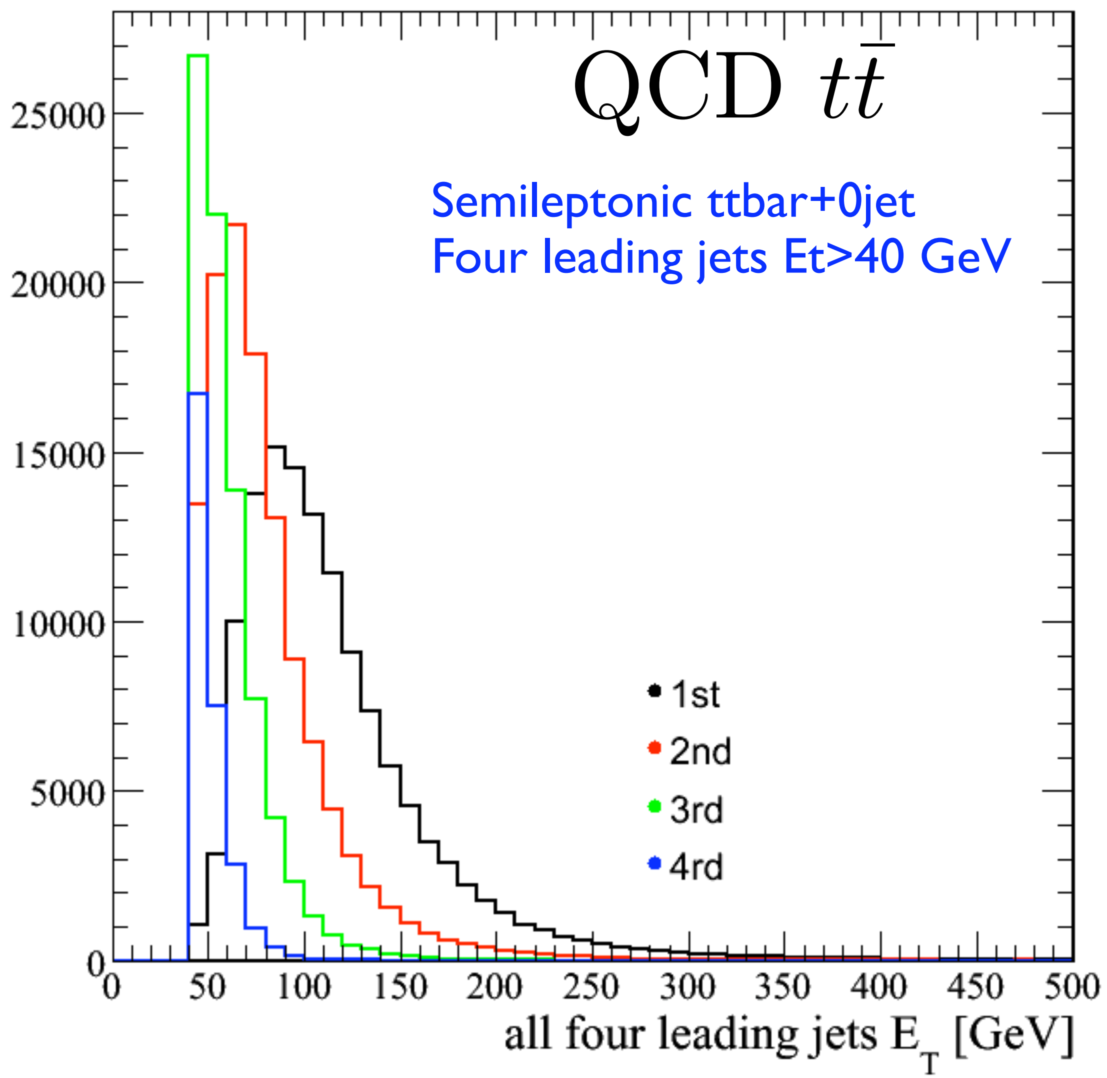
Muon momentum from semileptonic Z' decays





In the case of the muon semileptonic channel, the momentum of most of the muons from several heavy Z' resonances is below 400 GeV or away from the showering region. Life is a lot easier since we would not have to worry about the momentum resolution, radiation of these muons.

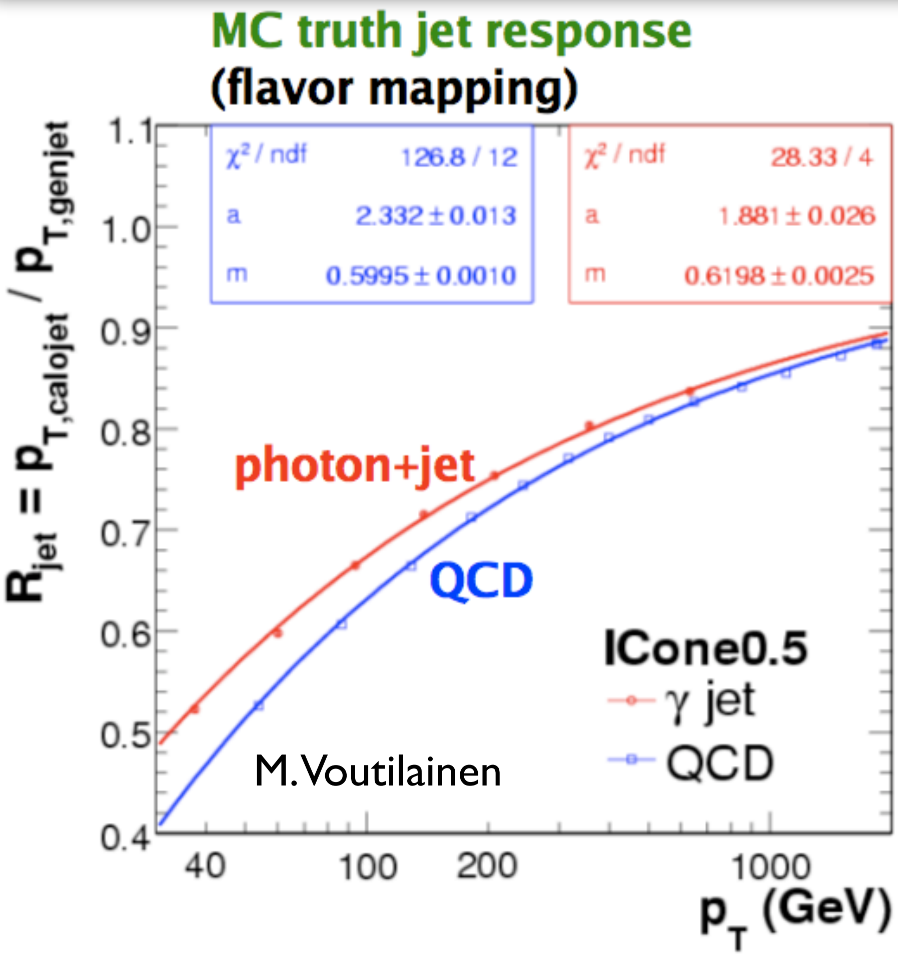
Kinematics (5): Leading Jets

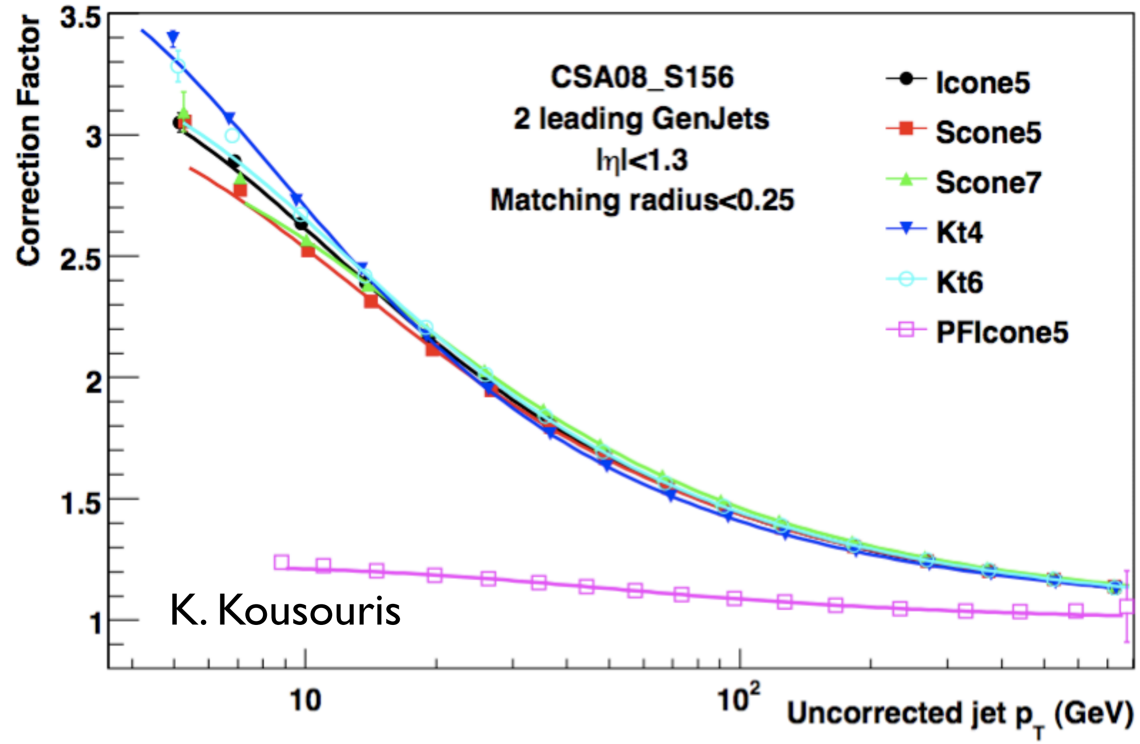


Mean Et of the four reconstructed leading jets with Et>40 GeV:

- 116 GeV
- 80 GeV
- 60 GeV
- 50 GeV

How large are the jet energy corrections at high-pT?

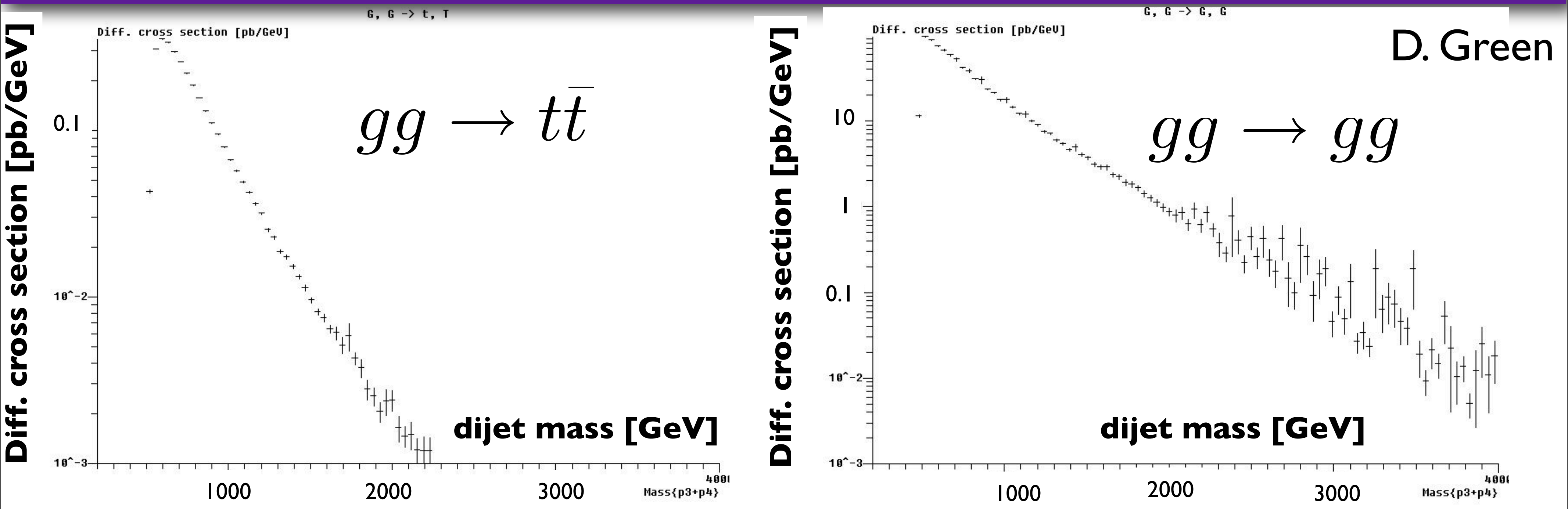




Absolute jet corrections approach unity for high-pT jets. The corrections are independent of the jet algorithm and sample in the high-pT region.

Experimental Challenges

QCD Background in the high-mass region

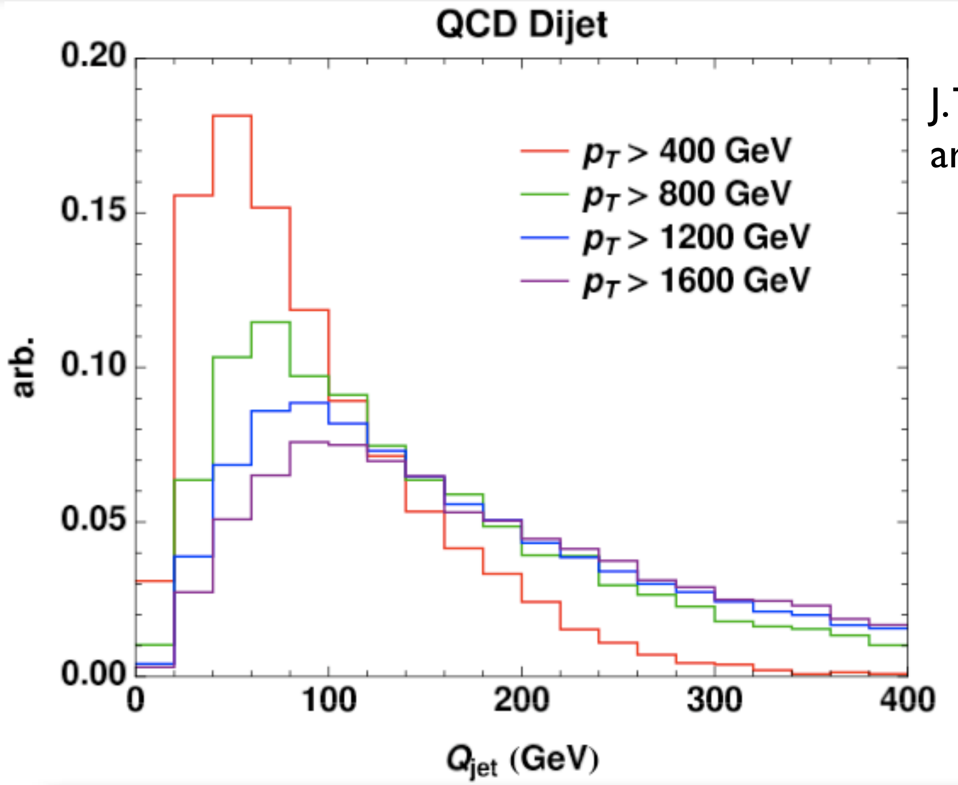


COMPHEP generated cross section for ttbar and QCD process as a function of dijet mass.

The b pair and the top pair cross sections are similar for masses well above the top mass. No need to reject bbar too hard.

On the other hand the top pair cross section above 1 TeV mass with jet pT>200 gives a ratio of gluon jets to top pairs of 320. We need a good rejection.

QCD Background



J. Thaler, L.T. Wang
arXiv:0806.0023

Jet invariant mass of the hardest jet in QCD dijets

Average jet mass increases as ~ 10% x pT

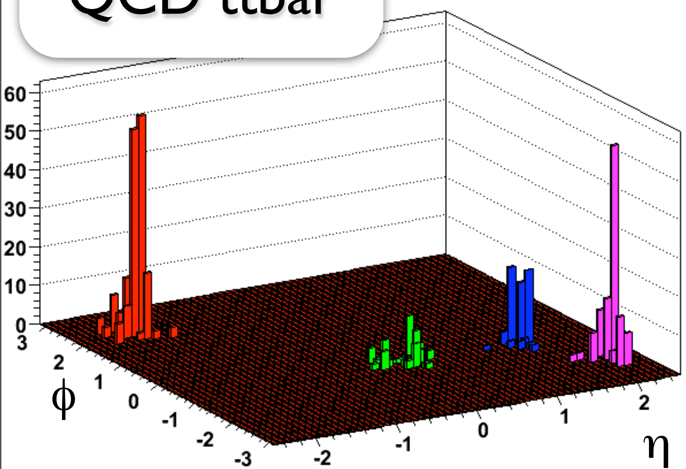
So once QCD jets pT~1TeV, the average jet invariant mass ~ the top mass. Kinematics of top decays ~ kinematics to QCD radiation.

We need tools to have a good rejection of QCD jets

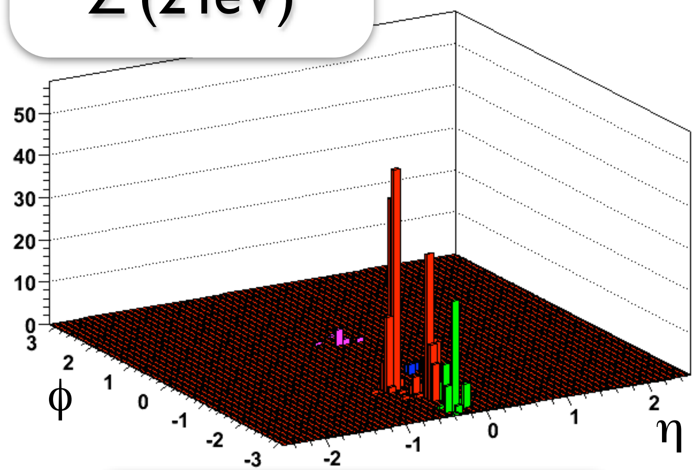
Experimental Challenges of boosted top quarks

- Top decay products are merged at high-pT. Need different reconstruction technique other than the standard top analyses.
- Lego plots of the calorimeter tower energy of the four leading jets:

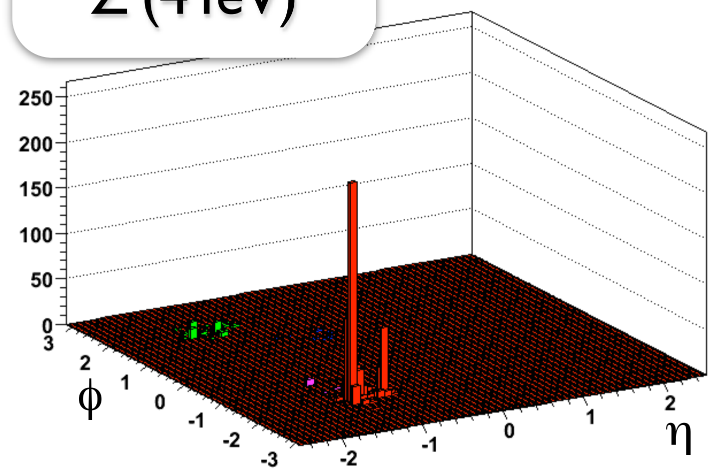
QCD ttbar



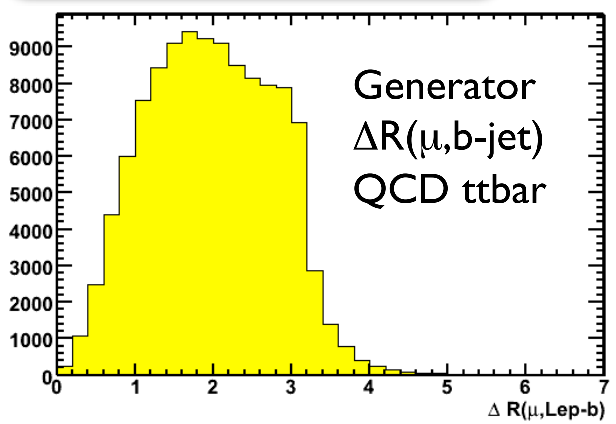
Z'(2TeV)

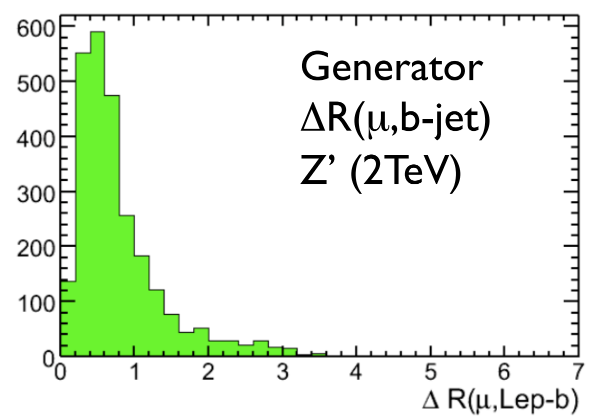


Z'(4TeV)



distinguishable 4 jets → jets closer to leading jet → jets merged into leading jet. Di-jet topology



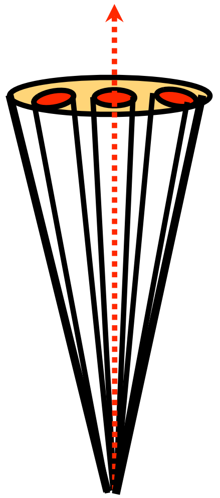


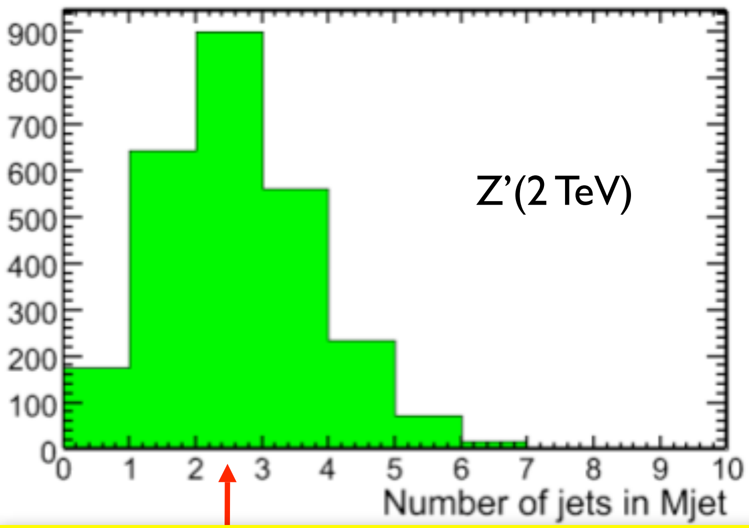
Jet Association for (fat) high-pT tops

- A simple approach to associate jets is to use ΔR(jet,leading-jet) to reconstruct a merged jet:
  - Start with the leading jet.
  - The rest of jets around the leading jet are added vectorially to the leading one.

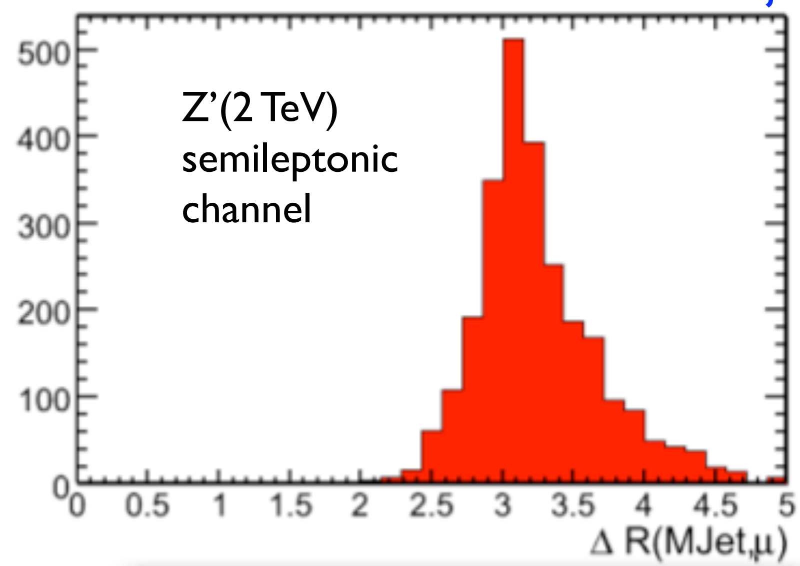
$$p^{\beta}_{MJet} = \sum_{i}^{jets} p^{\beta}_{i} \text{ if } \Delta R(\text{leading jet}, ith\text{-jet}) < 1.6$$

  - The resulting jet is called a merged jet and the invariant mass is estimated.





Number of merged jets into fat jet. In average a couple of jets around leading jet are pick up into a merged jet.

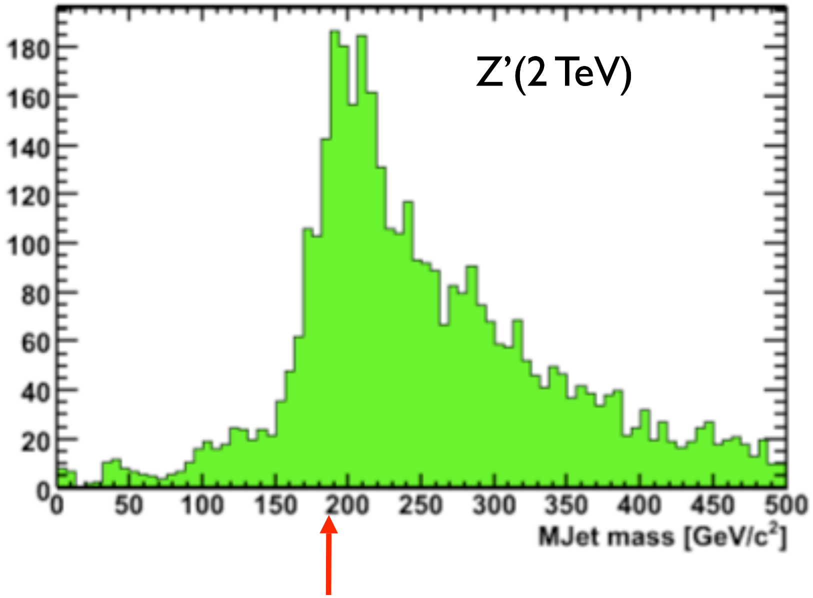


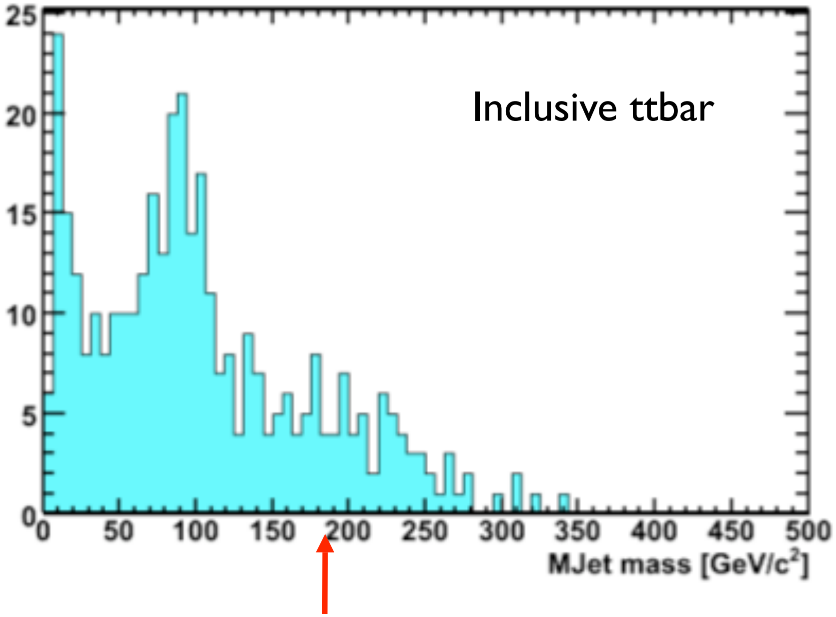
Merged jet is opposite to muon, ~ back-to-back topology

Using the Simple jet-merging to Reconstruct Z'

The mass of the merged jet picks around the resonance mass. Using this simple jet association increases the reconstruction efficiency of Z' even without optimization of cuts.

The same procedure is applied to a sample of inclusive ttbar events. This selection reduces the reconstruction efficiency of continuum ttbar events.

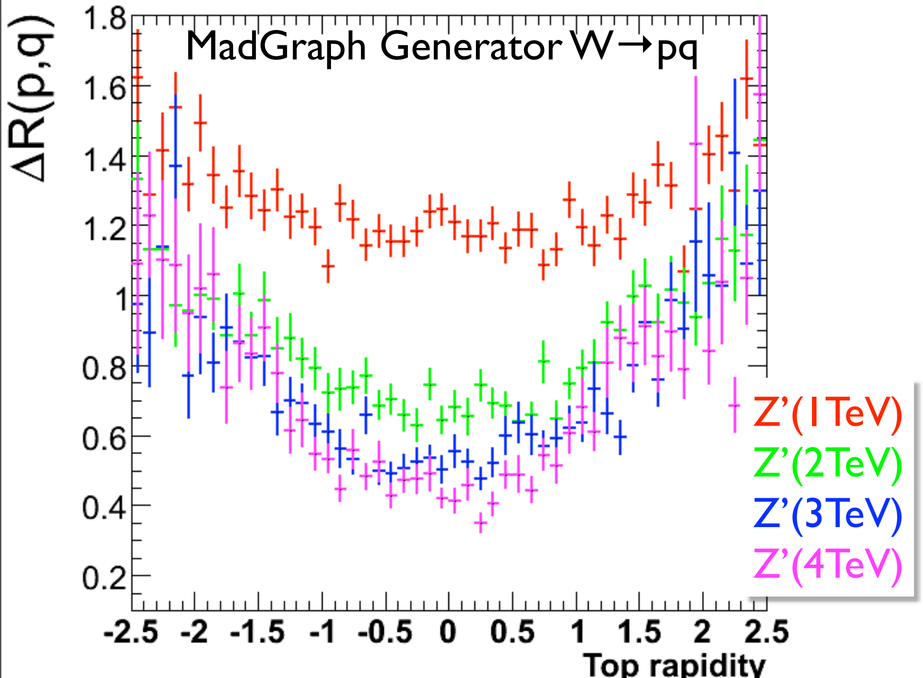




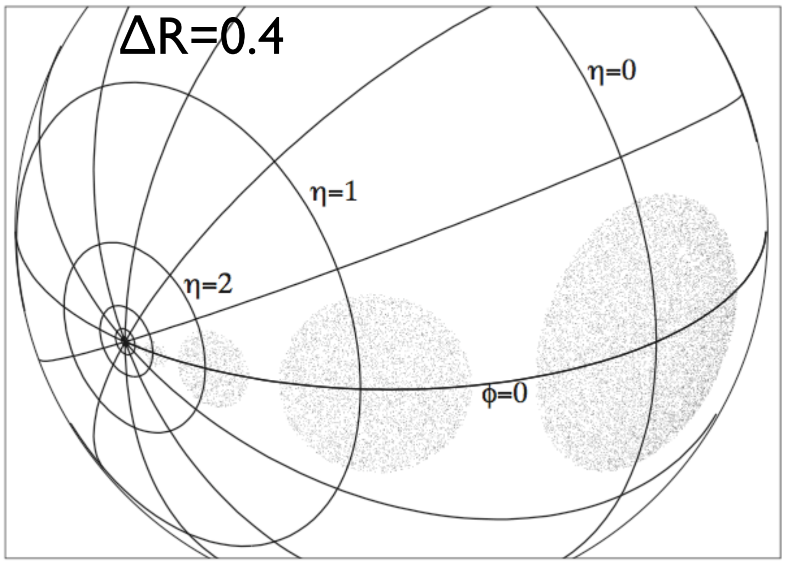
Jet Association for high-pT tops (II)

However, ΔR is not a good variable to merge jets from boosted top quarks because of its strong dependency on c.m. decay angle

- Distortion of angular distribution and hence spin determination



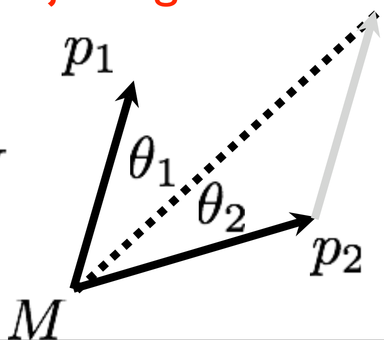
ΔR cones shrink as η increases



- A new variable ψ has been studied to replace ΔR. Inspired by the Kt jet-algorithm:

$$\psi = (p_1 + p_2)\sin((\theta_1 + \theta_2)/2)[\min(p_1/p_2)]^{1/\alpha}/M$$

optimal value α=4 using a toy MC



Recap about Jet Algorithms

Jet clustering algorithms in the market: Kt, Cambridge/Aache, cone jet algorithms (SISCone, iterative cone).

Basis:
A distance *dij* is introduced between entities (particles, pseudojets) and between entity *i* and the beam. The clustering proceeds by identifying the smallest of the distances:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} ,$$

$$d_{iB} = k_{ti}^{2p} ,$$

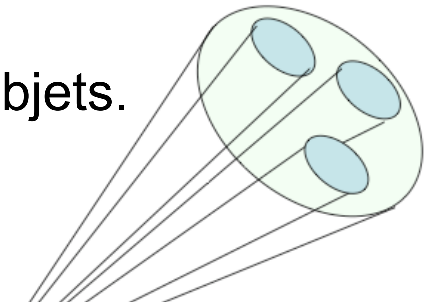
where *kt* is the transverse momentum of particle *i*. R is the radius parameter.

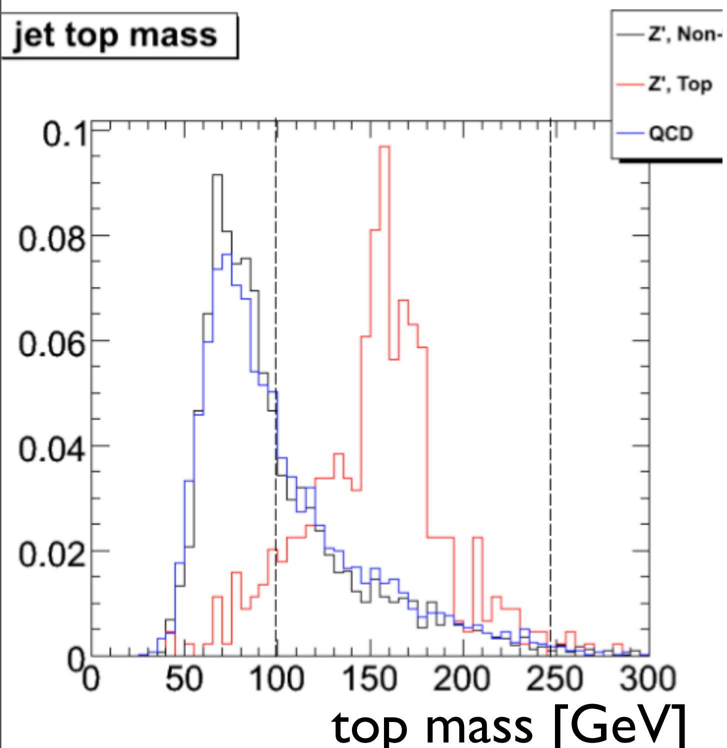
For p=1, inclusive kt algorithm,
For p =0, inclusive Cambridge/Aagen algorithm,
For p= -1, anti-kt jet clustering.

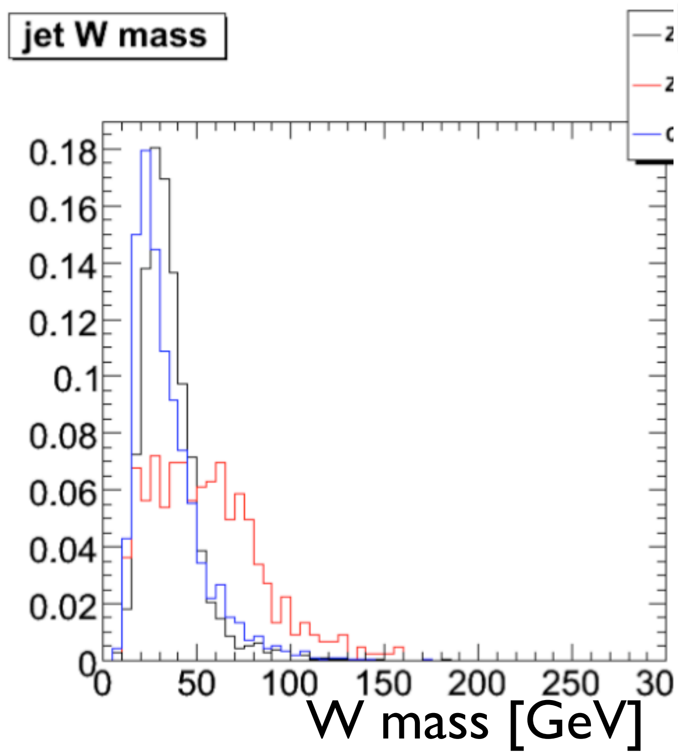
Implementation of new Jet Algorithms

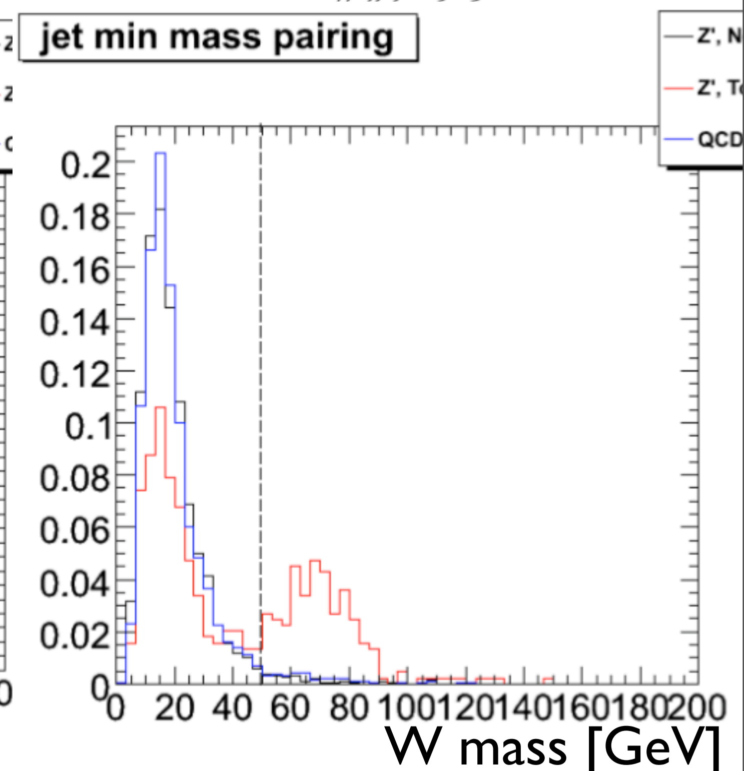
- New jet algorithms and event shapes have been proposed by theorists to separate high-pT top jets from QCD jets.
- For example, top-tagging by Kaplan et. al.(Johns Hopkins) hep-ph:0806.0848
- Based in sub-jet structure. Run the Cambridge-Aachen algo, then
- Iterative declustering. Impose angular and kinematic constraints to subjets.
- Claim ~99% QCD rejection
- Preliminary implementation in CMS:

  Samples: 5k Z' inclusive decays, 10k QCD (pt 500-1000 GeV)









S. Rappiccio et al.

Potential discrimination power against non-top jets

Jet Studies

There are many jet issues that still need to be done, just to mention a few:

- Data/Monte Carlo comparisons (when data available).
  - Sideband studies, can we select a QCD sample without top contamination

- How sensitive are the jet algorithms to the parton showering models?
  - Can we have an experimental observable which provides more information about the parton showering history?

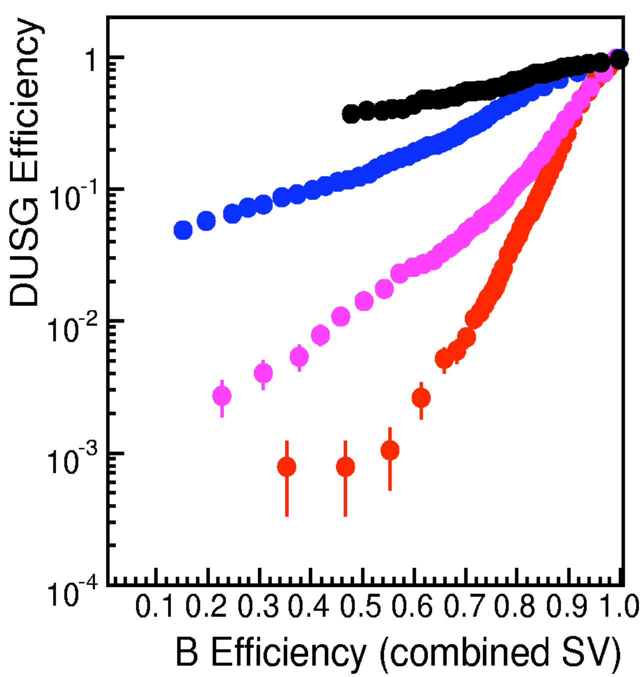
- For jet finding algorithms, is there an optimal ΔR-cone?
  - It is better to use $\Delta R<0.5$ or $\Delta R>0.5$ cone sizes?

- Single top tagging versus versus correlations between top pairs.

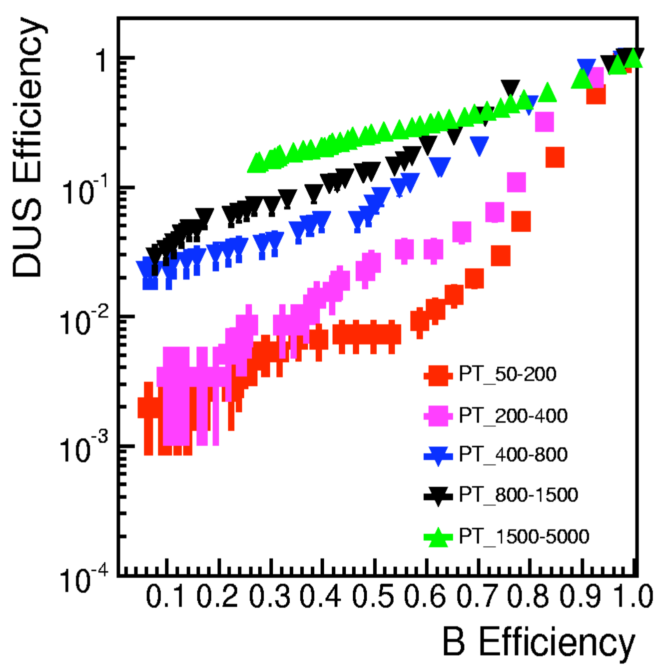
b-tagging at very high-pT jets

- b-tagging has been studied in different high-pT jet samples:
- top pairs at 100 GeV, 500 GeV, 1 TeV, 3 TeV (only hard interaction)
- Z' to ttbar 1,2,3,4 TeV
- QCD with pt>3000 GeV
- Very large degradation of b-tag performance observed at high-pT.
- High mis-tagging rate is present in all taggers: lifetime, Sec.Vtx., Combined

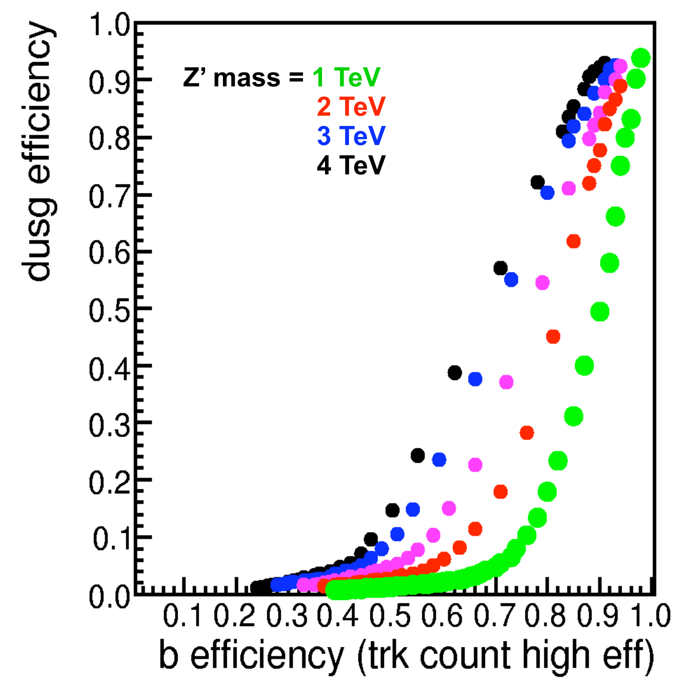
top gun samples



QCD 3 TeV sample



Z' samples

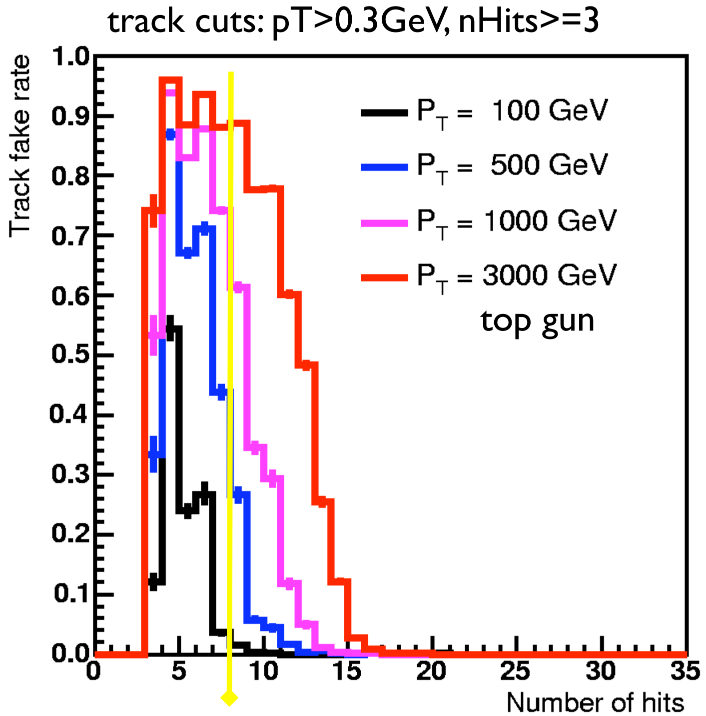


G. Giurgiu et al.

b-tagging of high-pT jets (II)

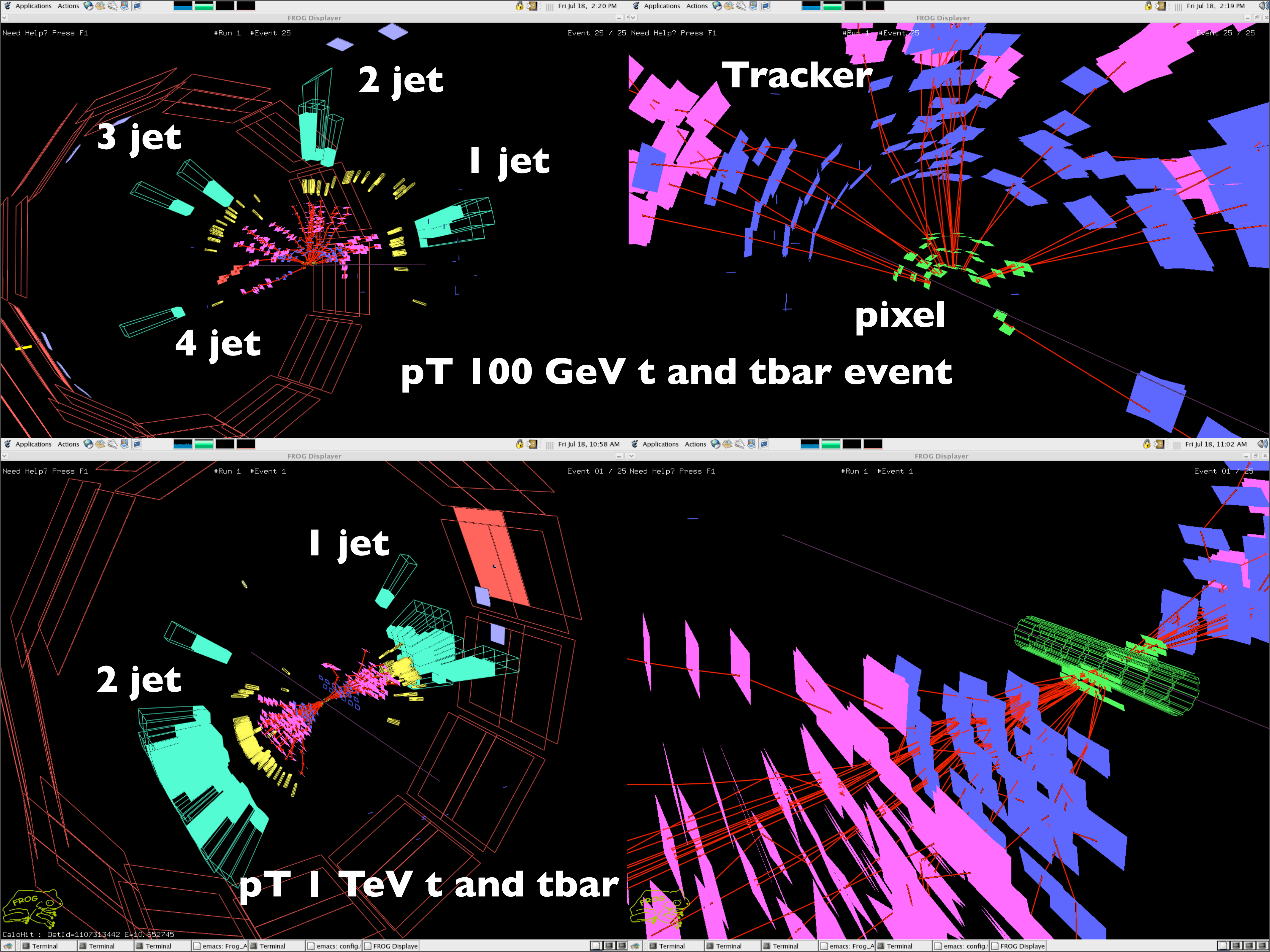
• This is a combined effort between Tracking and b-tagging groups to understand problem and provide solutions.

• Source of the problem seem to be a high fake track rate and badly reconstructed tracks.

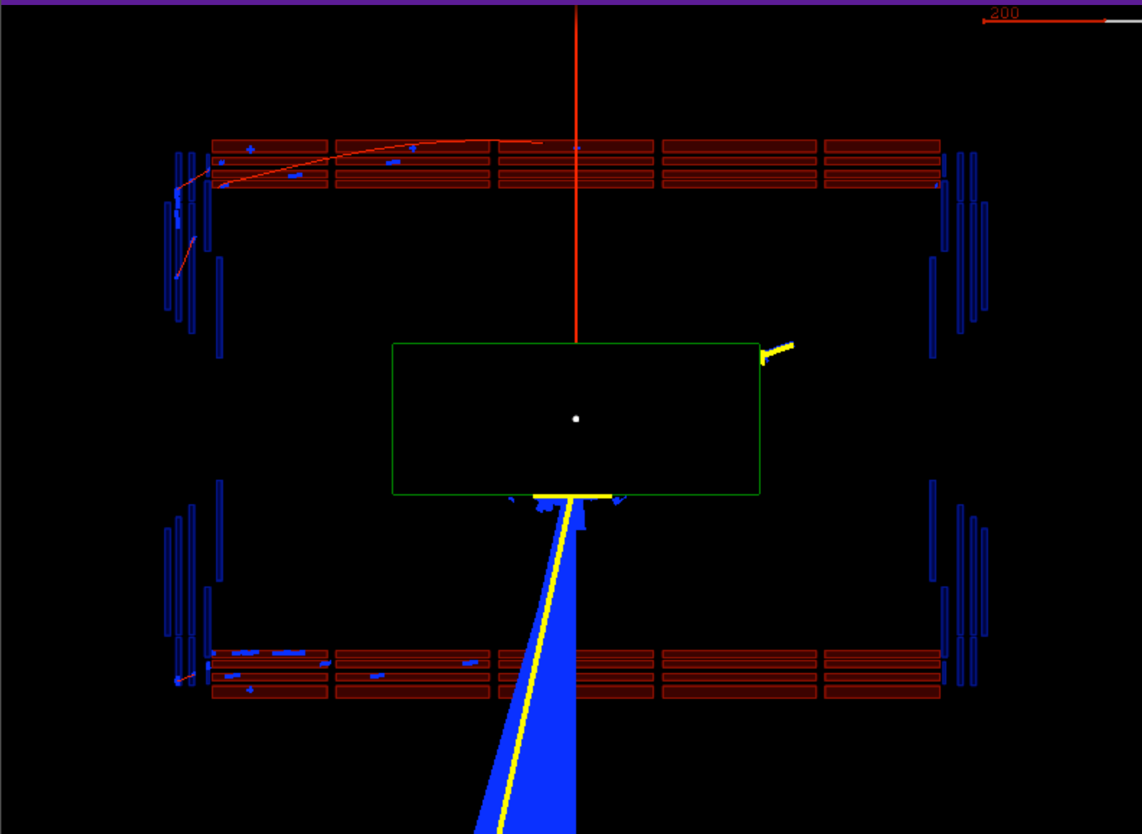


• b-tag algorithms already have track quality cuts, tighten cuts reduce track fake to ~10%
• Looking at the Tracking Particles, large rate of shared hits (SimHits) are observed in the Pixel:
• 32% for barrel Layer 1, 36% barrel L123
• Exploring other options:
    • Deterministic Annealing Filter (DAF)
    • Splitting merged hits (cluster splitting)
• No clear solution for the moment.

G. Giurgiu et al.

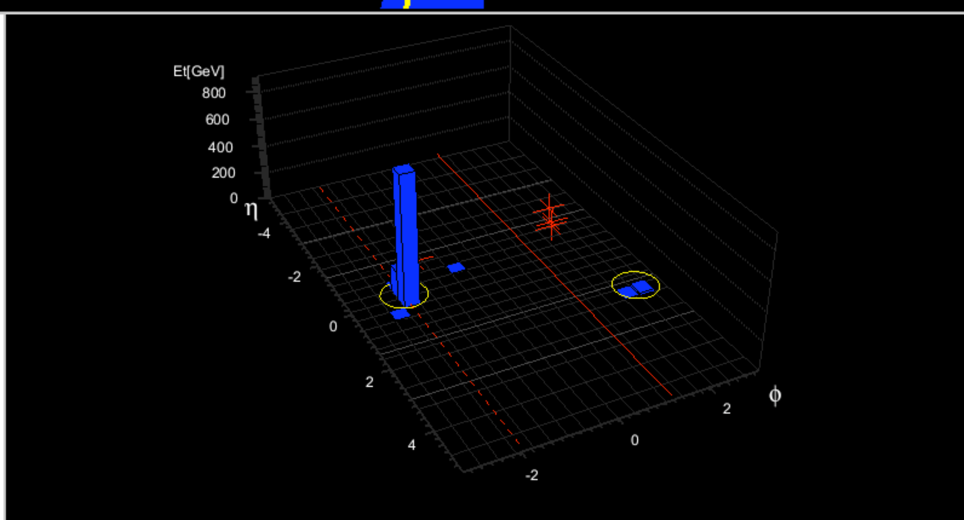

FROG Displayer
Need Help? Press F1
#Run 1 #Event 25
Event 25 / 25
2 jet
3 jet
1 jet
4 jet
Tracker
pixel
pT 100 GeV t and tbar event
#Run 1 #Event 1
Event 01 / 25
1 jet
2 jet
pT 1 TeV t and tbar
CaloHit : DetId=1107313442 E=10.652745

and a “real” TeV jet in beam halo data



A few events during the beam injection tests have some high-pT jets like this one of 1.2 TeV jet in the central region. Could be a beam gas interaction?

These kind of events are being studied now so they can be removed later on.



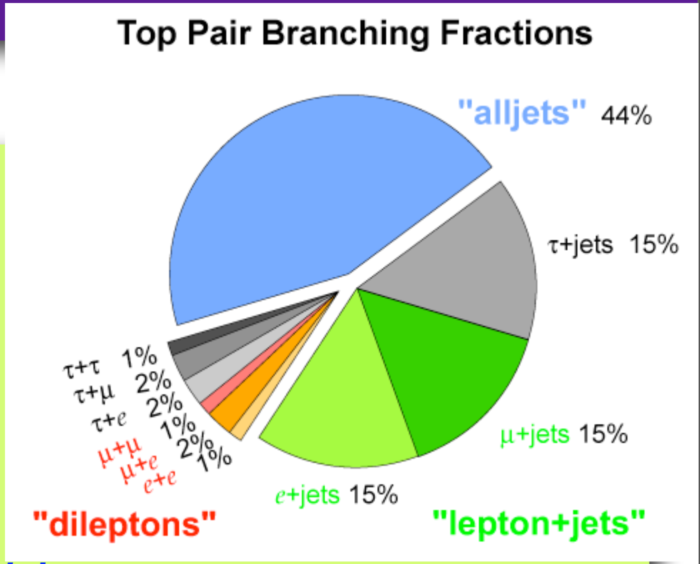
Fri Sep 12 13:05:49 2008 GMT

Run 62285 Event 1290
MET 1151.9 GeV MET phi 1.159 Sum ET 1195.7 GeV

Jets

Et	eta	phi	ECAL	HCAL	emf
1152.4	-0.061	-1.964	0.0	1154.5	0.000
27.4	1.798	2.493	0.0	85.1	0.000
7.2	0.495	-2.138	0.0	8.1	0.000
6.5	-0.784	-0.691	0.0	8.7	0.000

Overview of Analysis



Analyses with an early data sample ~10/pb

- Observation of top quarks.
- Reconstruction of mttbar distribution.
- Top mass from the mttbar distribution.

Decay channels:

- *lepton + jets overall best choice: good purity and yield.*
- dilepton: yield is ~5 times smaller, high purity.
- all hadronic: better yield, lower purity: large QCD background.

Search for new phenomena

Once top has been observed in the LHC and we have a confidence on the background content in the mttbar distribution, we can start searching for new physics using mttbar observable:

Searches on two regions: low-mass and high-mass regions of mttbar.

Early Observation of top: Roadmap to Discovery

Currently, we are studying procedures for the first observation of top quarks at LHC:

- Simulation analysis assumes 10/pb of data.
- Focus on muon semileptonic top decays.
- Fist step, a very simple cut-based analysis:
    - Hard cuts on jet and muon pT.
    - Simple selection of jet combinations.
    - Assume that b-tagging and MET are not ready to be used in analyses.
- Second step, a more optimal analysis (also useful for samples with c.m.<14TeV)
    - Efficient selection of jets and muons.
    - Selection of the best jet combination using a χ^2 distribution.
    - Use MET as input to reconstruct leptonic top decays.
- Once the detector has been aligned and calibrated, we can use other tools like:
    - Impact parameter significance and b-tagging to reduce background.

Simple Cut-based Analysis

CMS Analysis Note 2008/014:

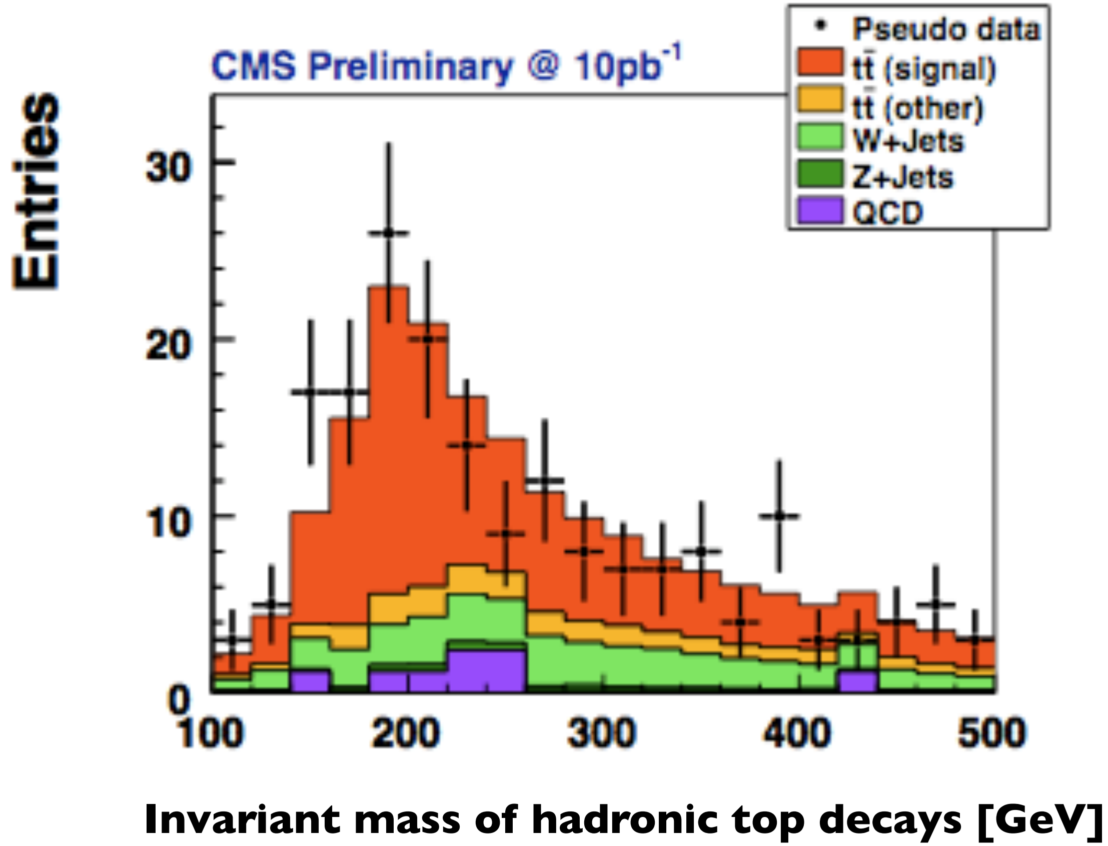
- Trigger: single non-isolated muon, pT>16 GeV.
- Loose preselection: muon pT>20 GeV, at lest one raw jet ET>30 GeV.
- Selection:
  - Muon: one isolated muon pT>30 GeV, |eta|<2.1. Tracking and Calorimeter isolation is applied.
  - Jets: at least 4 jets |eta|<2.4 jet ET 65/40/40/40 GeV.
  - Minimum ΔR(muon,jets) > 0.3

Data samples reconstructed with 100/pb conditions:
ALPGEN ttbar+jets, W+jets, Z+jets.
PYTHIA muon enriched QCD.

Simple cut-based analysis (2)

Selected Events for 10/pb:

	$t\bar{t}$ (signal)	$t\bar{t}$ (other)	W+jets	Z+jets	QCD	S/B(QCD)	S/B
Preselection	749	527	7474	1430	–	–	–
$dR_{min} > 0.3$ & $E_{iso}^{calo} < 1$ GeV	128	25	45	7	11	11.62	1.47



Invariant mass of hadronic top decays.

Jet combinations: The combination of three jets with the highest vectorially summed Et.

The pseudo data correspond to a poissonian smeared selection which is diced from the summed MC distribution.

Optimize Jet Selection

A simple way to optimize the jet selection is to estimate a χ^2 for each combination of jets:

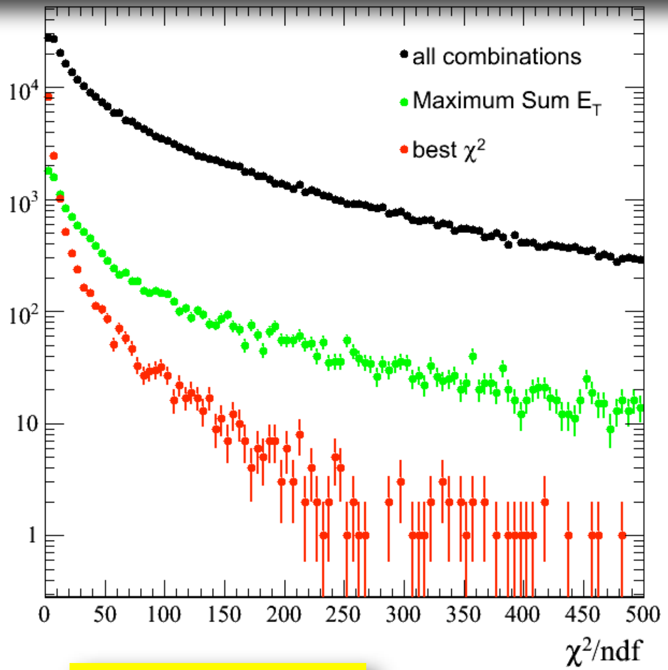
$$\chi^2 = \frac{(M_{j_1 j_2} - M_W)^2}{\sigma_{jj}^2} + \frac{(M_{j_1 j_2 j_3} - M_t)^2}{\sigma_{jjj}^2} + \frac{(M_{W j_4})^2 - M_t}{\sigma_{\mu\nu j}^2}$$

Use MC to estimate sigma. For dijet W mass is 7.6 GeV, for three jet top mass is 12.5 GeV, and for the jet + leptonic W is 15.6 GeV.

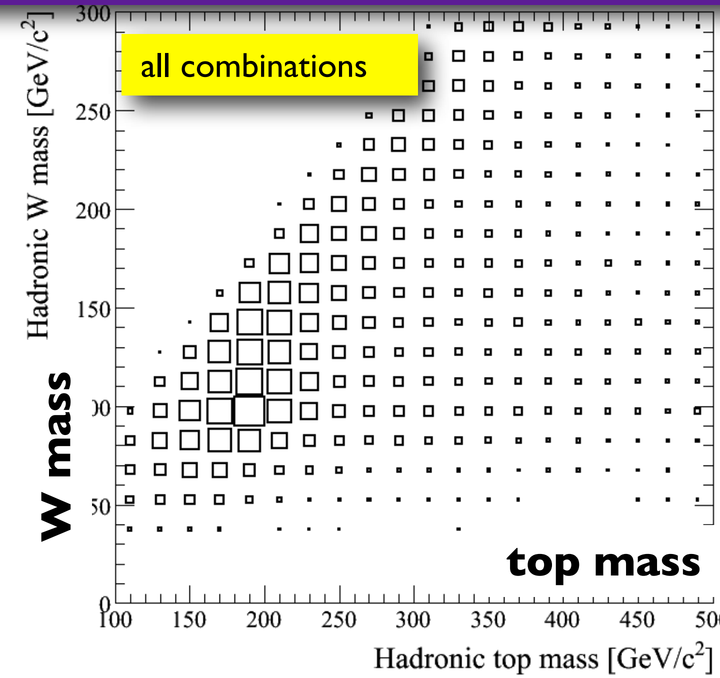
Select maximum 6 jets and pick the best combination using χ^2
If only 4 good jets then we have 12 combinations.
If 5 jets, there are 5 x 12 combinations,
If 6 jets, there are 15x12 combinations.

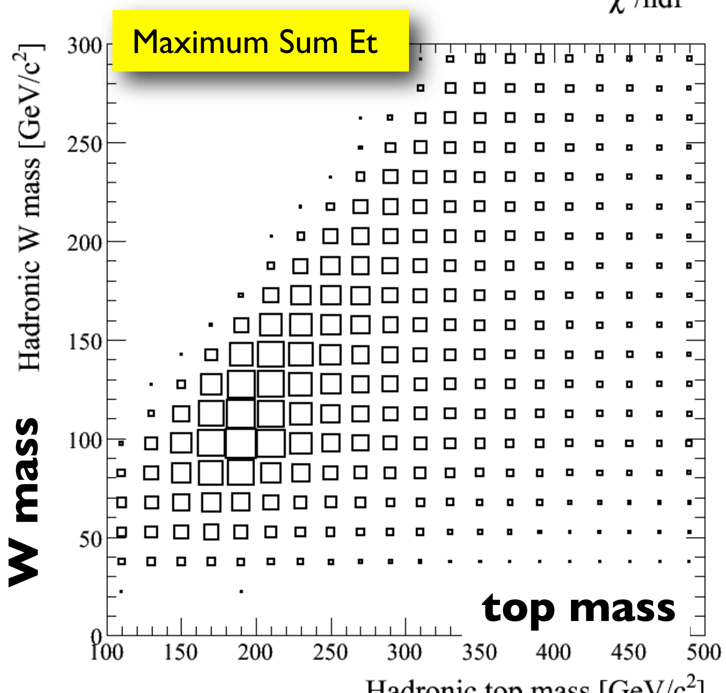
We can also apply cuts on the invariant hadronic and leptonic W masses.

Optimization of Jet Combinations(2)

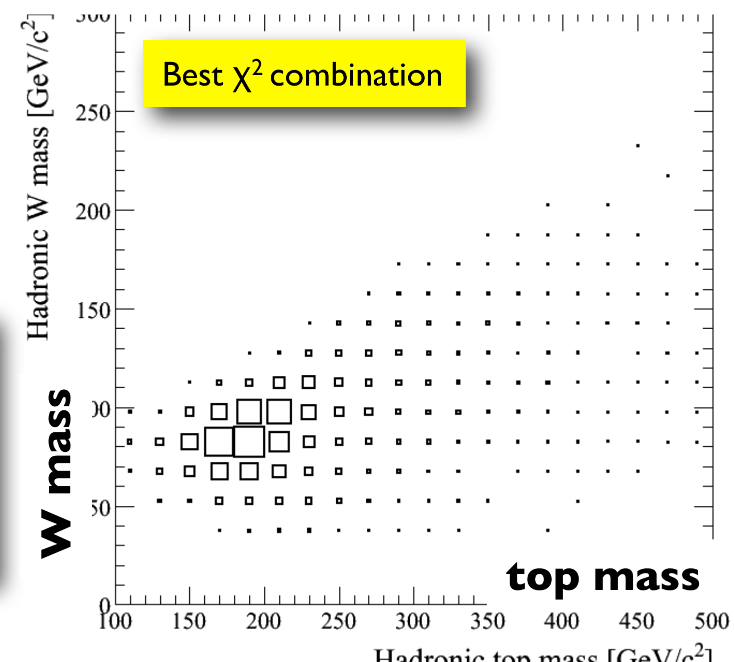


Jet Combinations for signal ttbar +0jet sample

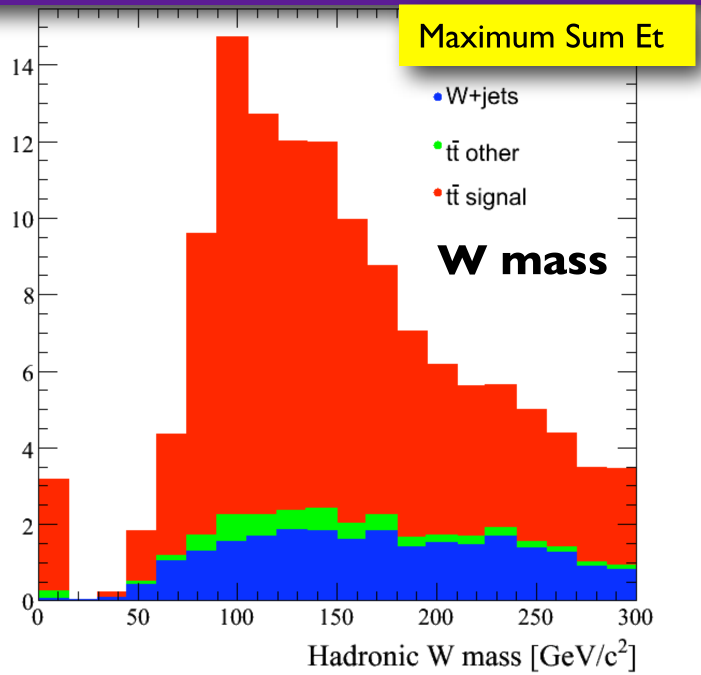


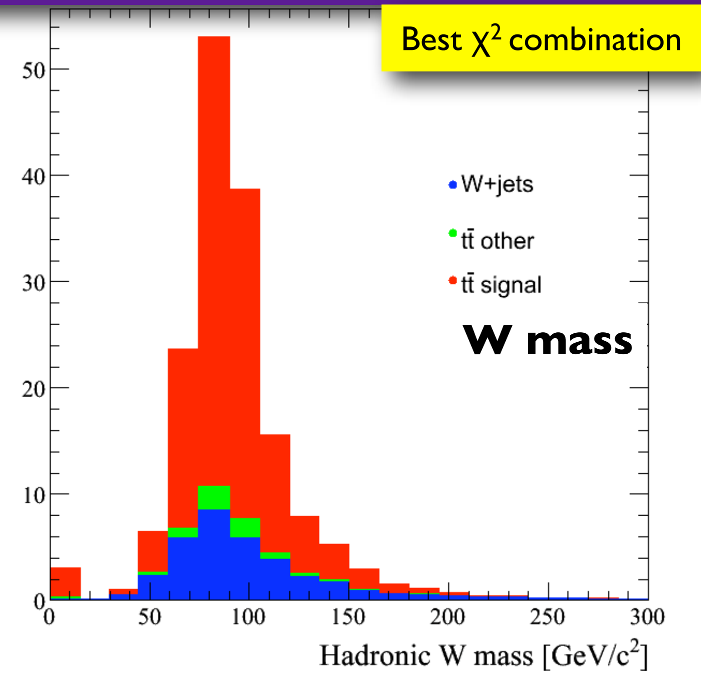


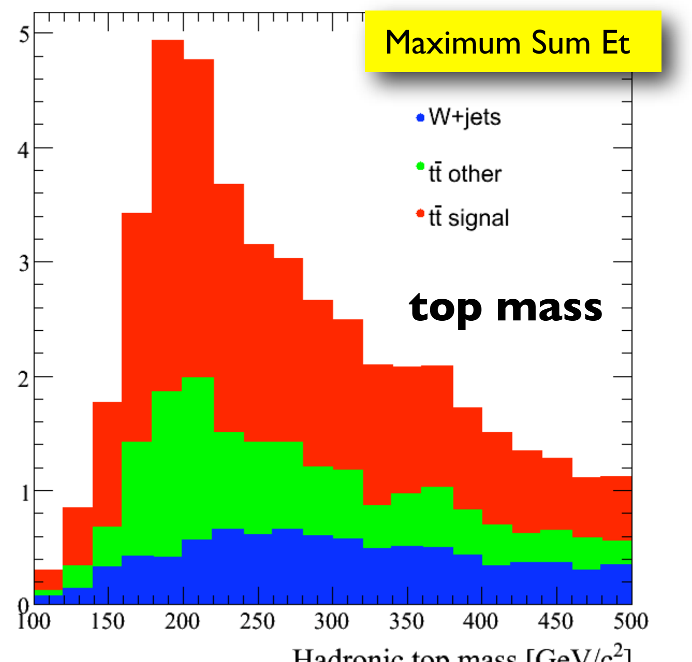
Using the best χ^2 the hadronic W and top mass resolution is improved. The correct set of jets are better chosen with this method.

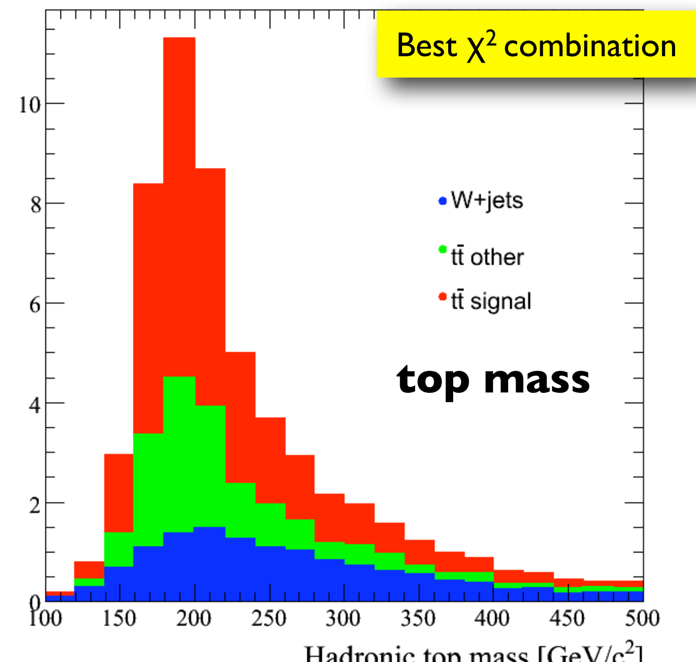


Optimization of Jet Combinations(3)



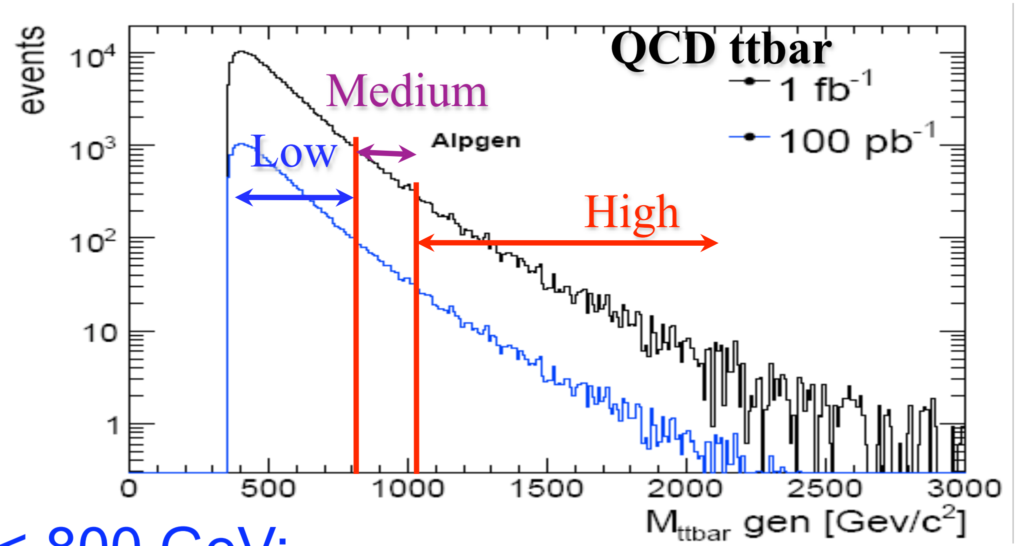






$m_{t\bar{t}}$ Analyses and Searches

At the moment, analyses are divided in regions of Mttbar:



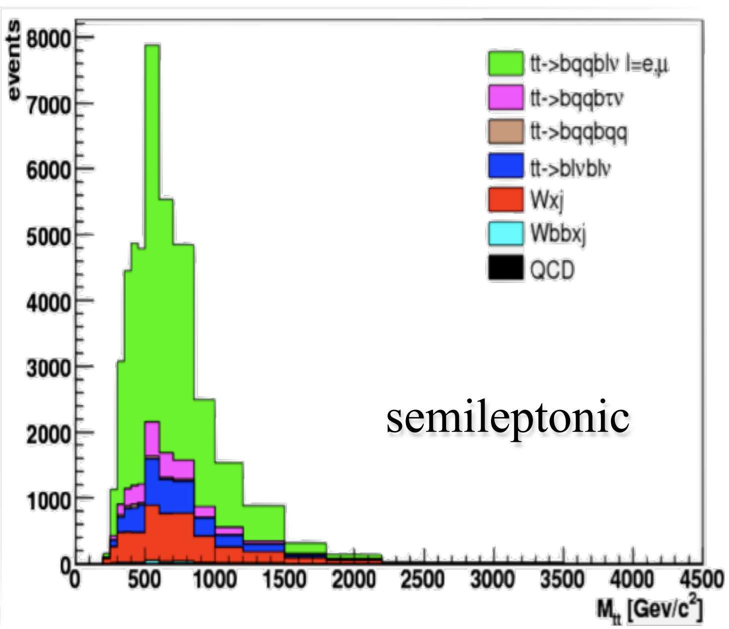
➡Low mass region < 800 GeV:
Most of the SM processes peaks.
Standard tools for top can be applied e.g. identifying the W, find three jets whose mass ~ top-mass, b-tagging.

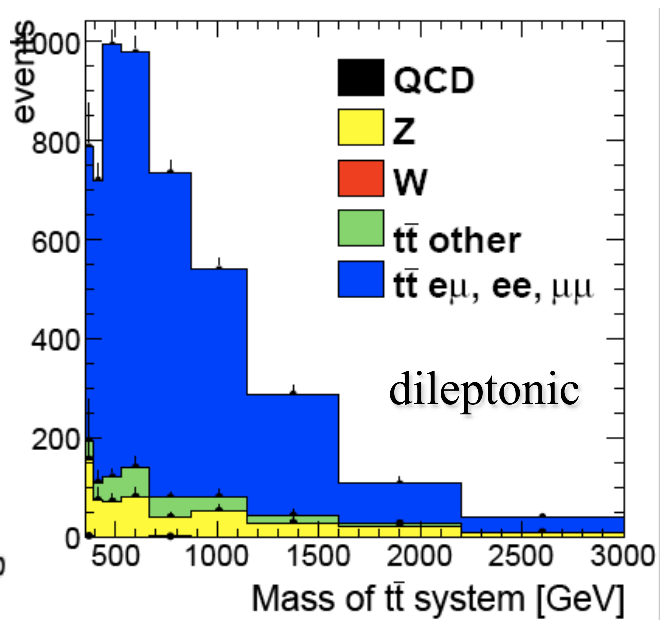
➡Medium mass region 800-1000 GeV:
Standard tools begin to have low efficiency because of the high boost.

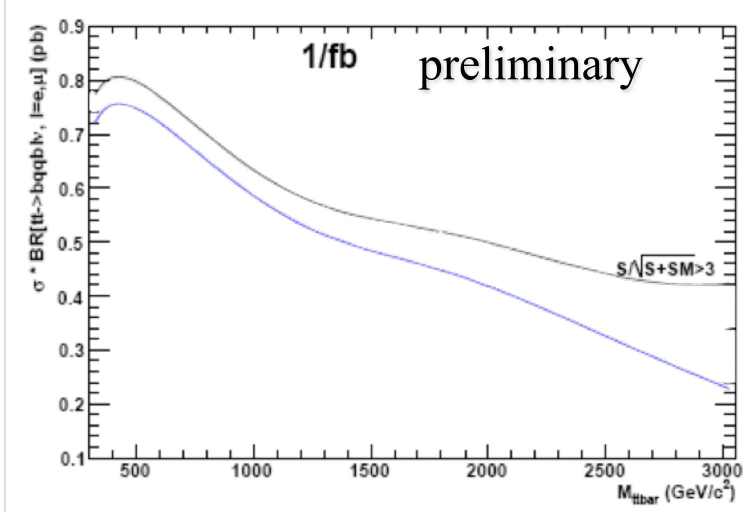
➡High mass region > 1000 GeV: boost factor $E/m \gtrsim 5$
A different approach is needed to reconstruct the highly boosted products.

$m_{t\bar{t}}$ Analysis: low & medium mass region

- Use most of the default selection criteria from standard top analysis.
  - Use a kinematic fit to improve mass resolution in the semileptonic channels.
- Study both semileptonic (e, muon) and dilepton channels.
- Details in CMS AN-2007-027

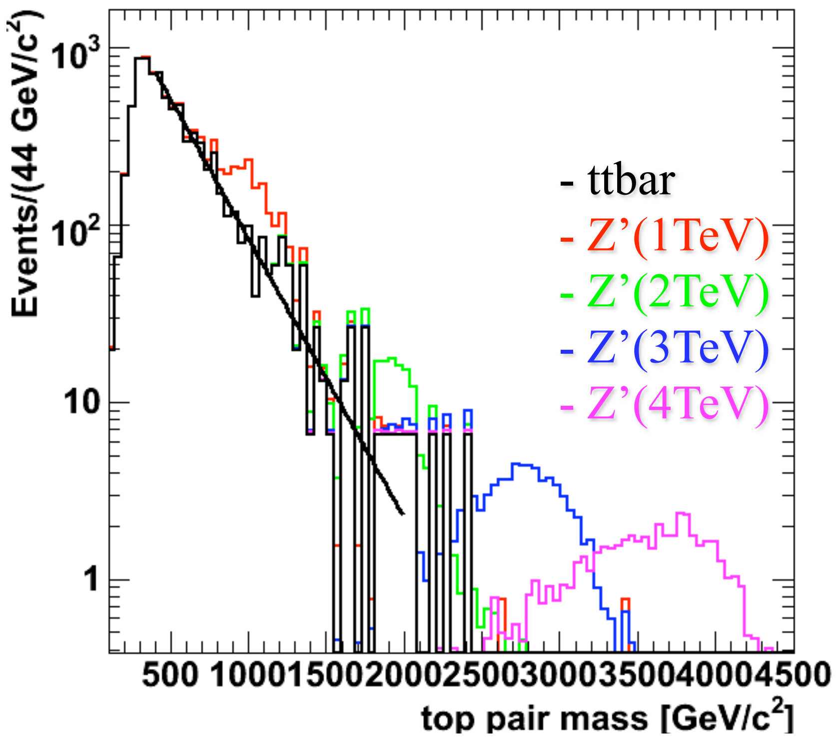


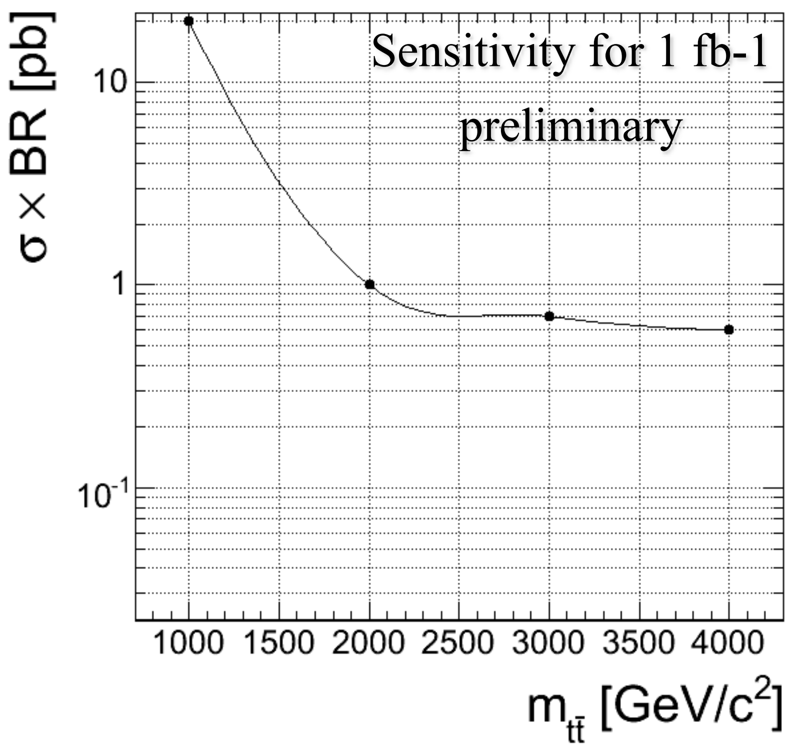




$m_{t\bar{t}}$ Analysis: high mass region

- Search for narrow resonances at masses above 1 TeV.
- Only muonic channel has been explored so far.
- For merged jets, a simple approach is used to associate jets based in ΔR(jet, leading-jet).
    - Efficiency to reconstruct Z' is improved.
- Details in CMS AN-2008-011



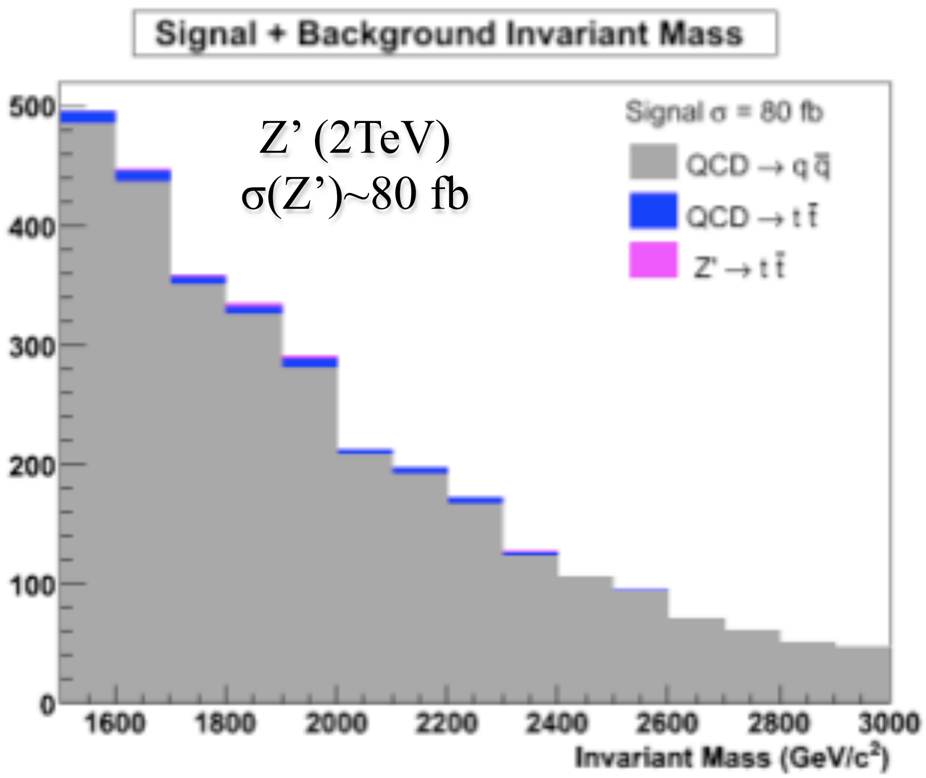


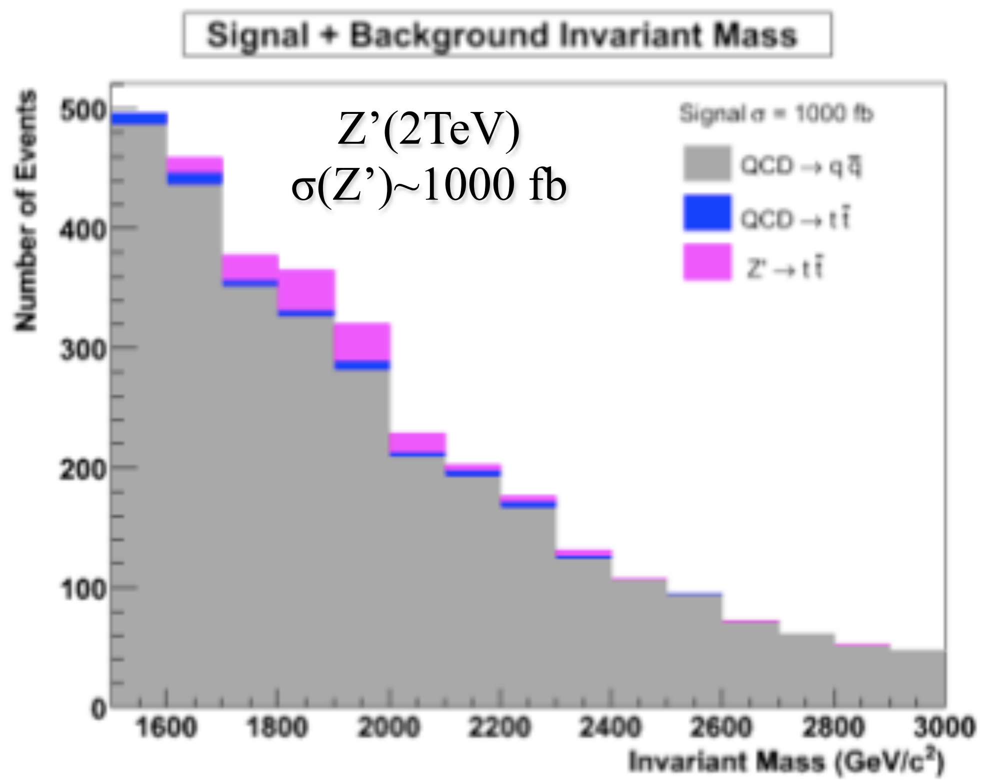
Fully hadronic channel in the high $m_{t\bar{t}}$ region

- This channel has been explored using the fast simulation.

- Reconstruction: constrained MET, find leading jet and associate jets with a 3D angle in two groups, then apply cuts on the mass of both jet groups.

- Preliminary studies shown that *QCD background is huge*. Needs optimization and new jet algorithms or event shapes to separate QCD.





Other Ideas to be explored

J.~Thaler and L.~T.~Wang,
“Strategies to Identify Boosted Tops”. JHEP 0807, 092 (2008)
arXiv:0806.0023

Variables to identify boosted tops. Semileptonic and hadronic channels.

L.~G.~Almeida, S.~J.~Lee, G.~Perez, G.~Sterman, I.~Sung and J.~Virzi,
“Substructure of high-p_T Jets at the LHC”
arXiv:0807.0234

Propose observables to reduce QCD background

P.~Ferrario and G.~Rodrigo,
“Massive color-octet bosons and the charge asymmetries of top quarks at hadron colliders”
arXiv:0809.3354

Charge asymmetry as an observable for new physics

... and just another idea came out

Top Jets at the LHC

Leandro G. Almeida[a], Seung J. Lee[a,b], Gilad Perez[a,b], Ilmo Sung[a] and Joseph Virzi[c]

[a] *C. N.Yang Institute for Theoretical Physics, Stony Brook University, Stony Brook, NY 11794-3840, USA*

[b] *Department of Particle Physics, Weizmann Institute of Science, Rehovot 76100, Israel*

[c] *University of California; LBNL, Physics Division, 1 Cyclotron Rd., Berkeley, CA 94720, USA*

arXiv:0810.0934v1 [hep-ph] 6 Oct 2008

Abstract

We investigate the reconstruction of high p_T hadronically-decaying top quarks at the Large Hadron Collider. One of the main challenges in identifying energetic top quarks is that the decay products become increasingly collimated. This reduces the efficacy of conventional reconstruction methods that exploit the topology of the top quark decay chain. We focus on the cases where the decay products of the top quark are reconstructed as a single jet, a "top-jet". The most basic "top-tag" method based on jet mass measurement is considered in detail. To analyze the feasibility of the top-tagging method, both theoretical and experimental aspects of the large QCD jet background contribution are examined. Based on a factorization approach, we derive a simple analytic approximation for the shape of the QCD jet mass spectrum. We observe very good agreement with the Monte Carlo simulation. We consider high-p_T $t\bar{t}$ production in the Standard Model as an example, and show that our theoretical QCD jet mass distributions can efficiently characterize the background via sideband

Summary

- We are straining our current tools and algorithms in extreme conditions (high-pT jets)
  - We are already finding limitations and problems. However, we are working to solve these issues before data comes.

- Current efforts are ongoing to improve understanding of:
  - Fast vs full Monte Carlo simulation of high-pT objects,
  - trigger efficiencies at low and high Mttbar regions,
  - muon isolation and identification in jetty environment,
  - new jet algorithms,
  - tracking and b-tagging at high-pT jets.

- New ideas for jet algorithms and event&jet shape parameters are being explored to improve reconstruction of merged jets and reject non-top jets.

- A strategy for a roadmap to discovery using top quarks is being prepared in CMS:
  1. Demonstrate detector is operational: Observation of top quarks.
  2. Demonstrate we understand control samples (bkg): Optimal selection of top quarks.
  3. Search for new physics: analyses of Mttbar distribution.

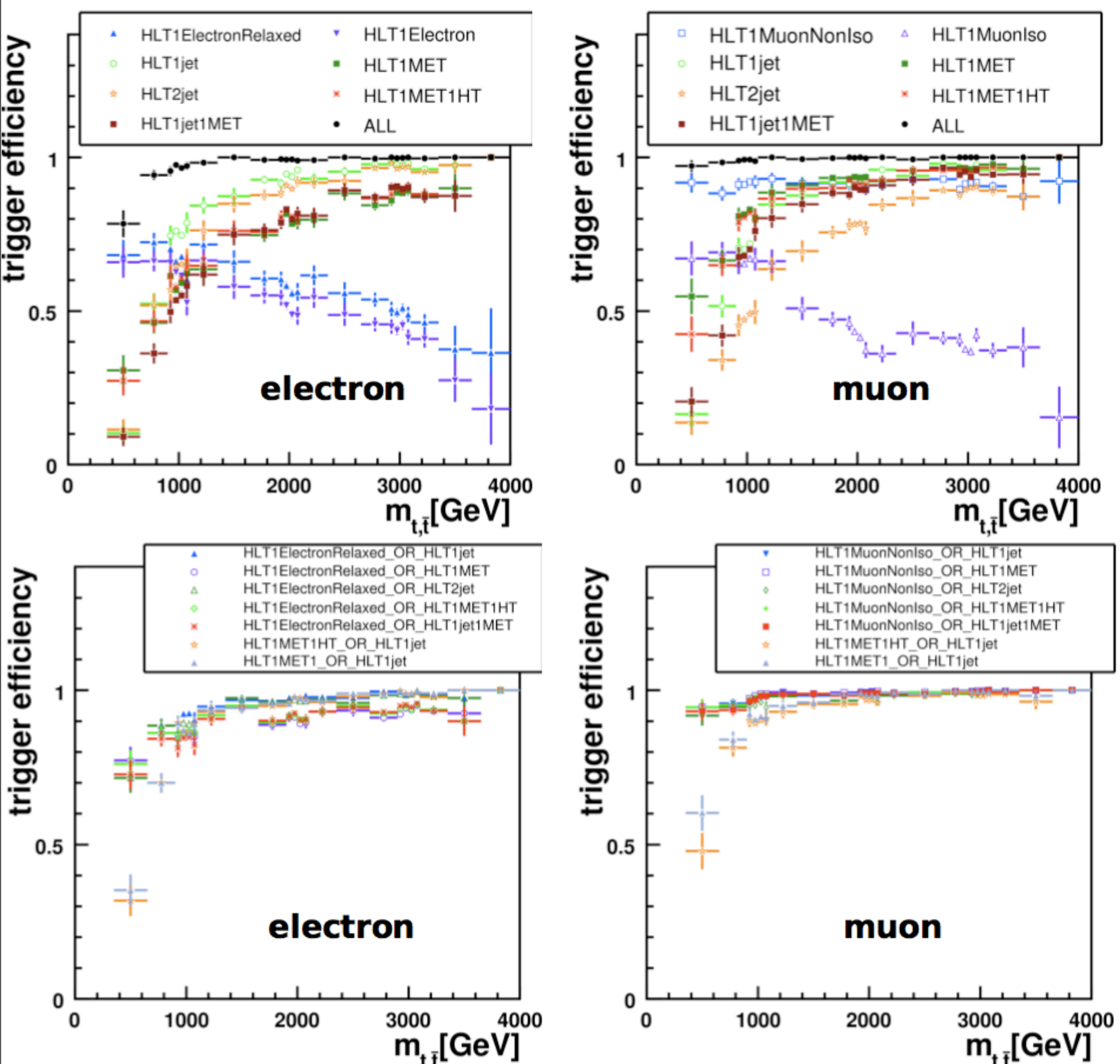
- Plan to expand data-driven techniques to our scenarios to estimate background
  - e.g. tag&probe as function of isolation.

- This is a very rich analysis that touches every part of the detector and software tools.

Backup slides

Trigger Studies as a function of $m_{t\bar{t}}$

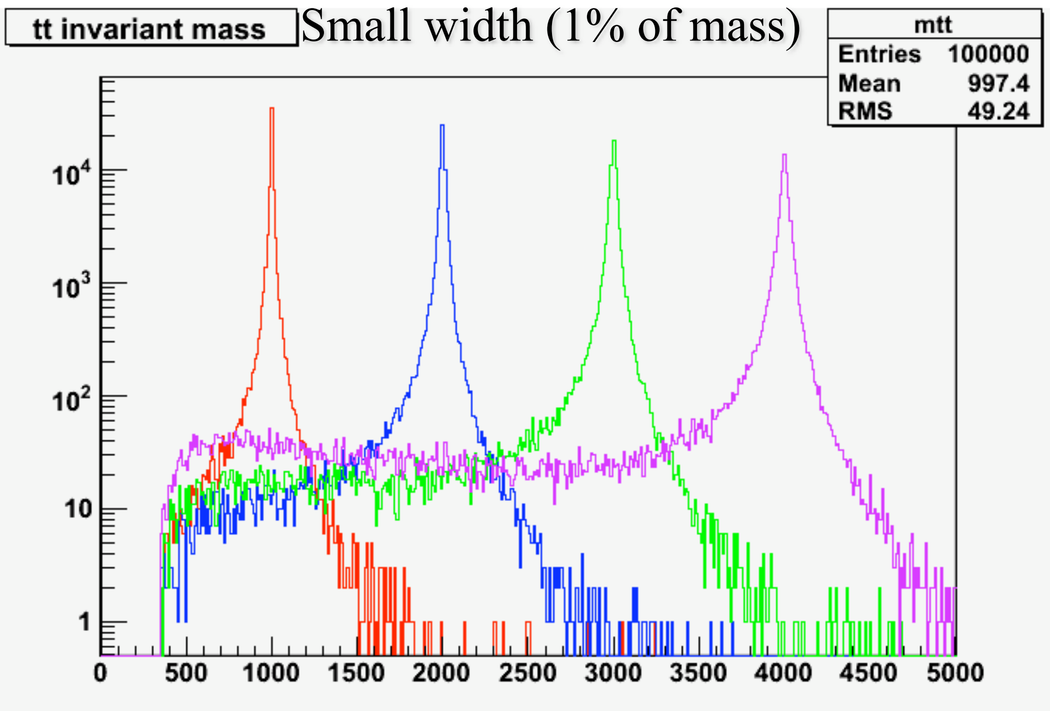
Semi-leptonic channels, all Z' samples combined

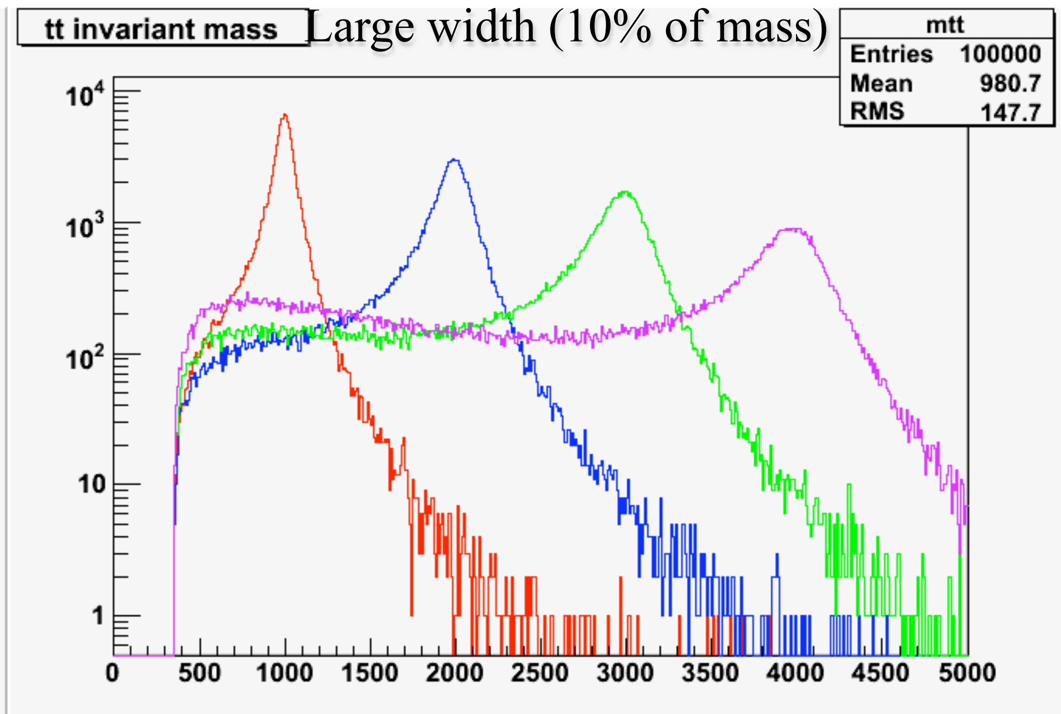


- Using PAT and CMSSW 1.6.11.
- Studying lepton triggers, jet triggers, and triggers using MET.
- Lepton triggers become inefficient for highly boosted top.
- Combination of non-iso lepton triggers and jet triggers (HLT1jet) provides efficiencies of ~1.

Group MC Production

- For more detail studies, we are going to produce a new set of signal MC samples:
- Use CMSSW 2.1.X which has several fixes like Geant4 long-live bug fix.
- Plan a model independent parameter scan (resonance mass, width, spin)
    - For several mass and width (1%, 10% of mass) points.
    - For several spin points: 0 (sizable interference effect with SM), 1, 2
- Use MadGraph generator





Competition between the Breit-Wigner of the resonance and the decreasing pdf of the partons creating the resonance → sizable low mass tail for the large width

Plans: data-driven studies

For many quantities we should use data to calibrate

- efficiencies (lepton id)
- background (fake rates)

Task in common with the top group, with the extra complication that we need to do it as a function of the boost of the top or lepton isolation.

We will focus on the (semi)leptonic channel with a muon and learn from there.

Lepton ID: ideas

- traditional tag&probe not enough, we want from data $\varepsilon(\eta, p_T, \text{isolation})$.
  - Z+jets and tag&probe, do we have enough statistics to map everything out?
- J/Ψ in jets. Too soft a spectrum? Too dense an environment?
- Energy extrapolation: trust MC as a first step.
- Mixing events: hybrid(MC+data).

Fake lepton rate: ideas

- use high p_T di-jet QCD selections (light jets, b-tag jets):
  - remove the ttbar component by differentiating also in number of tracks,calotowers or jet broadness.

Group Organization

- A recent effort: ~3 months ago.
- Contact persons: Roberto Chierici, Francisco Yumiceva
- Meetings:
  - Working meetings on-demand, at least one per month.
  - Reports are presented in the Top and Exotica PAGs.
- Hypernews: TopPairBSM
- More details in group twiki page:
  - https://twiki.cern.ch/twiki/bin/view/CMS/BoostedTop
- Group tools:
  - Common analysis package based in PAT
  - Common analysis skim