Lepton+Jet+MET Topology Group Overview

Pratima Jindal

Purdue University Calumet

(On behalf of the LPC LJMET Topology Group)

JTERM IV

3rd-5th Aug 2009

Outline

- A number of analysis approved recently
  - ttbar cross section muon+jets channel - mass reconstruction
  - ttbar cross section in muon+jets channel using multivariate technique
  - Search for Exotic Top Partners: $T_{5/3}$ Top Partner Model
- Some very interesting work in progress
  - Data Driven Method for QCD and W+jets normalization
  - Simultaneous Heavy Flavor and Top Cross Sections Measurement
  - ttbar cross section in muon+jets channel using a soft muon tagger

ttbar cross section in muon+jets - mass reconstruction

Dan Green, Gail Hanson,
Geng Jeng, Francisco Yumiceva

- M3 Method
  - Choose combination of three jets with the highest vectorial sum p_T
- M3'(X^2-sorting) Method
  - Based in a mass hypotheses. For each permutation of 4-jets calculate a X^2:

$$\chi^2 = \frac{(M_{j_1 j_2} - M_W)^2}{\sigma^2_{jj}} + \frac{(M_{j_1 j_2 j_3} - M_{t-had})^2}{\sigma^2_{jjj}} + \frac{(M_{W_l j_4} - M_{t-lep})^2}{\sigma^2_{\mu\nu j}}$$

hadronic W | hadronic top | leptonic top

- Motivation: Full reconstruction of ttbar event
- Masses/resolutions are taken from MC
- For each combination of 4 jets : calculate the X^2 and put in ascending order
- No fit is performed. Select permutation of lowest X^2
- Method leads to the assignment of jets to the hadronic and the leptonic legs
- Reconstruct the leptonic W from the μ and MET using a W—mass constraint:
  - Two real solutions: use pz () closest to the muon pz (μ)
  - Complex solution(35%): use the real part of solution for pz ()
  - Second case leads to a high tail, thus apply M(μ) <150 GeV/c²

Event Selection

Trigger: HLT_Mu9 (at least one L3 muon with p_T >9 GeV/c)

Muon: Exactly one isolated muon

- GlobalMuon: p_T >20 GeV/c, $|\eta|$ <2.1
- Quality criteria: IP(PV)<200 μm, $N_{\mathrm{Hits}}^{\mathrm{tracker}} \leq 11$, χ^2_{NDF} <10, Veto-Cone deposits E_T^{ECAL} <4 GeV / E_T^{HCAL} <6 GeV
- Isolation:

$$\mathrm{RelIso}(\mu) = \frac{E_{iso}^{calo} + p_{T,iso}^{tracker}}{p_T(\mu)} < 0.05$$

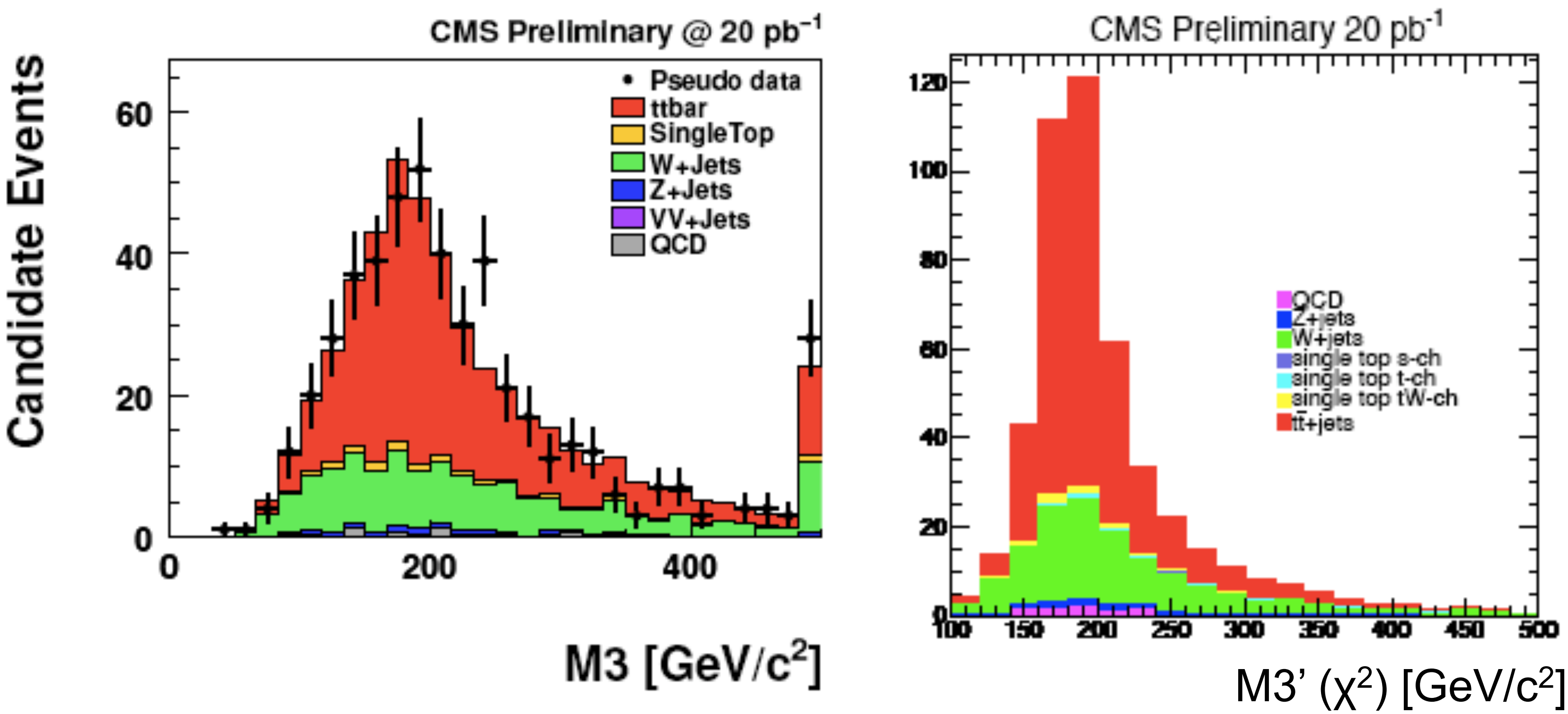
Jets: ≥4 jets with p_T >30 GeV/c, $|\eta|$ <2.1 (SisCone; R=0.5)

"Loose leptons": Veto on a second "loose" muon or a "loose" electron, i.e. Loose Muon: GlobalMuon, $|\eta|$ <2.5, p_T >10 GeV/c, RelIso<0.2 and Loose Electron: GsfElectron, $|\eta|$ <2.5, p_T >15 GeV/c, RelIso<0.2

	$t\bar{t}$+jets	$t\bar{t}$+jets	Single top			W+jets	Z+jets	VV+jets	QCD
	s.l. μ	other	s-Ch.	t-Ch.	tW				
AllEvents	1,220	7,060	32	832	580	912,000	76,240	236	2,546,279
Trigger	978	1,418	10	260	147	168,633	20,952	100	2,032,021
≥ 1 tight μ	620	345	5	140	69	110,509	15,296	73	7,200
< 2 tight μ	620	309	5	140	66	110,509	9,300	62	7,200
no tight e	620	264	5	140	62	110,508	9,292	53	7,200
veto on loose μ	618	228	5	140	60	110,503	5,492	44	7,192
veto no loose e	616	183	5	140	56	110,469	5,415	34	7,188
≥ 1 jet	614	180	4	125	55	16,998	1,325	18	2,701
≥ 2 jets	593	158	3	63	47	3,076	256	5	387
≥ 3 jets	489	99	1	18	27	651	51	1	60
≥ 4 jets	277	43	0	5	9	140	10	0	7

Signal : 320 events
Background :171 events

M3 and M3’ Distribution

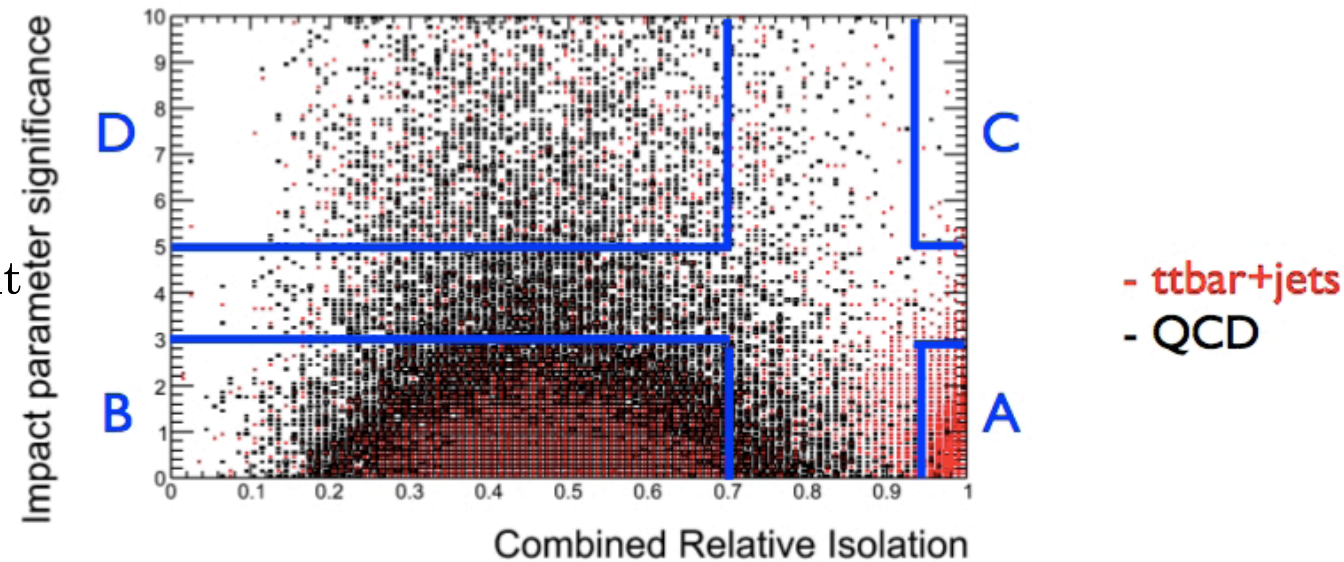


- Left: Combination of those 3 jets in the event, with the highest vectorially summed pT
- Right: Distribution of M3’ variable defined using the best χ^2 jet combination

Data Driven QCD Estimation

- Use two uncorrelated variables; divide in 4 phase-space regions, among which 3 are dominated by QCD
- Variables: Impact parameter significance with respect to the beam spot and combined relative isolation:

Use NA/NB = NC /ND
=> NA = NB · NC / ND
where NA is the amount of QCD in the signal region



Data-driven technique to estimate the contribution from QCD: the ABCD method -> yields an uncertainty of about 50%

Systematics

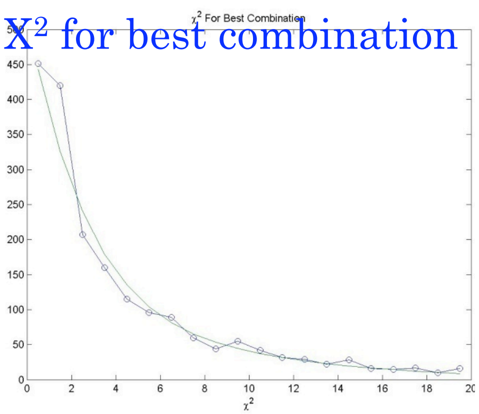
Summary of statistical and systematic uncertainties in the template fits to the $\eta(\mu)$, M3 and M3' variables

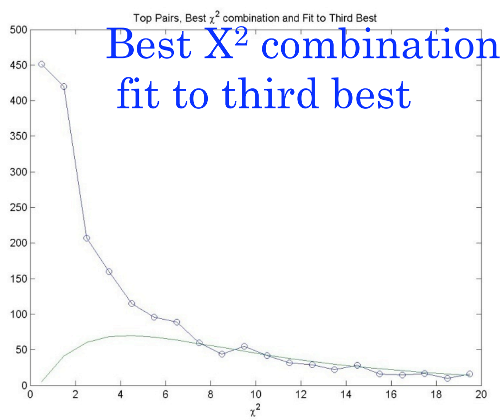
Source	Uncertainty [%]		
	Fit to $\eta(\mu)$	Fit to M3	Fit to M3'
Statistical Uncertainty (20 pb^{-1})	17.7	16.3	11.5
Jet Energy Scale	16.7	15.1	19
$t\bar{t}$ MC Generator	1.9	14.9	14
$t\bar{t}$ ISR/FSR	3.3	7.7	2
W+jets Factorization scale	4.4	4.7	4
W+jets Matching threshold	5.5	2.8	4
Single Top Shape	0.1	0.8	1
PDF Uncertainty	5.0	5.0	5.0
Total Systematic Error	19.2	23.8	25.0
Luminosity Error	10.0	10.0	10.0

- Cross section is extracted by template fits to discriminating distributions : M3, M3', $\eta(\mu)$
- ttbar cross section can be measured with about 12-18% stat. uncertainty for 20 pb^{-1} and about 20-25% systematic uncertainty dominated by the uncertainty on the JES

Next Steps

- Cuts on X^2 to reduce/study backgrounds
  - M_{jjj}: 2291 top pair events with $n_j > 3$
  - Cut on $M_{uv} < 150$ * $M_{jj} < 110$ * $(150 < M_{uvj} < 220)$
  - Have 1538 events (67 %) for best chisq combination, 800 for third best
  - Therefore, expect ~ 738 top pair events where the parton level 4 jets are in the sample
- Templates for X^2 – reconstructable top, other top and backgrounds





use third best chisq as a shape and then fit to large chisq tail
Find 1134"good" top pairs – compared to 738

- Multiplicity extrapolation as cross-check on template fits to X^2
- Flavor discrimination – b tags
- Extended X^2 using likelihood for flavor

ttbar cross section in muon+jets: multivariate technique

Pratima Jindal, Gennadiy Kukarzev,
Meenakshi Narain, Neeti Parashar

- The goal is a measurement of the ttbar cross section with early LHC data
- Using only kinematic and topological information in the event
- Not using b tagging
- Technique:
  - Event selection that limits the background rate to a fraction of the signal rate
  - Use several variables to improve separation between signal and background
  - Boosted Decision Trees is the method of choice for multivariate analysis
  - Template fit and the cross section evaluation
  - Estimate uncertainties

Event Selection

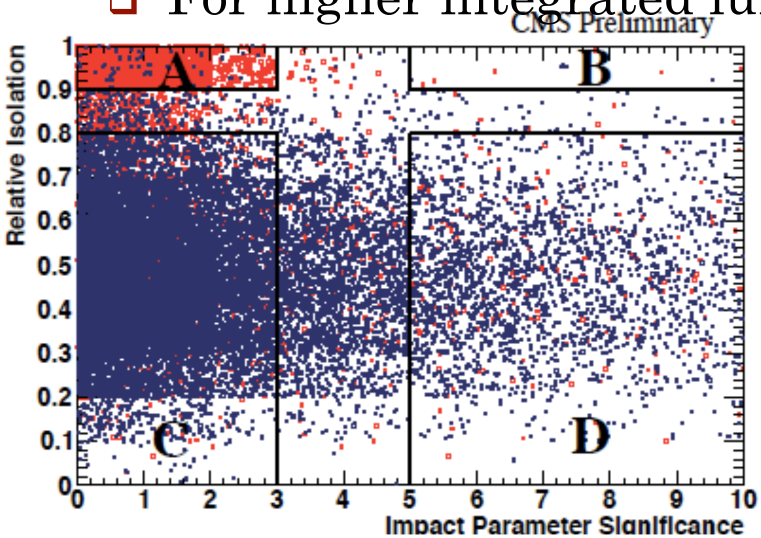
- Require that the event triggers the detector
- At least 4 jets and one muon with the following parameters
- Seedless Infrared Safe cone algorithm jets
  - p_T>30 GeV, $|\eta|$<2.4
- Muon
  - Relative isolation (ΔR<0.3) $\frac{1}{1+\frac{\text{track isolation}+\text{calorimeter isolation}}{p_T(\text{lepton})}} > 0.9$
  - p_T>20 GeV, $|\eta|$<2.1
  - Global fit of the silicon tracker and the muon system information
  - Quality requirements that are >90% efficient for an isolated muon
  - Small energy deposition in the EM and hadronic calorimeters that are consistent with a MIP
  - Impact parameter significance (relative to beam spot) < 3 standard deviations

Selection	$t\bar{t}+\text{jets}$	$W+\text{jets}$	$Z^0+\text{jets}$	$(tb)_t$	$(tb)_{tW}$	$(tb)_s$	QCD
preselection	952	236	45	12	31	0.5	24717
Trigger: HLT_Mu9	780	189	35	10	24	0.4	22035
muon quality	747	181	34	10	23	0.4	20605
RelIso > 0.9	421	148	29	6	13	0.2	62
E_T veto cone	395	134	25	5	11	0.2	28
$d_0/\sigma_{d_0}(\text{lepton}) < 3$	392	133	25	5	11	0.2	20

Signal : 392
Background : 194

Multijet Background

- We obtain QCD rate and template shape from data
- The template is derived from an enriched QCD data sample (isolation is reversed, no electromagnetic and hadronic vetos)
- The rate
  - For early data – estimated from the quadrant method (below)
  - For higher integrated luminosities – free parameter in the fit



Region	RelIso	d_0/σ_{d_0}(lepton)
A	> 0.9	< 3
B	> 0.9	> 5
C	< 0.8	< 3
D	< 0.8	> 5

$$A(\text{background}) = \frac{C \times B}{D}$$

Estimated: 9.5 ± 7.1 events
Expected: 19 events

Variables

η – pseudorapidity

E – energy

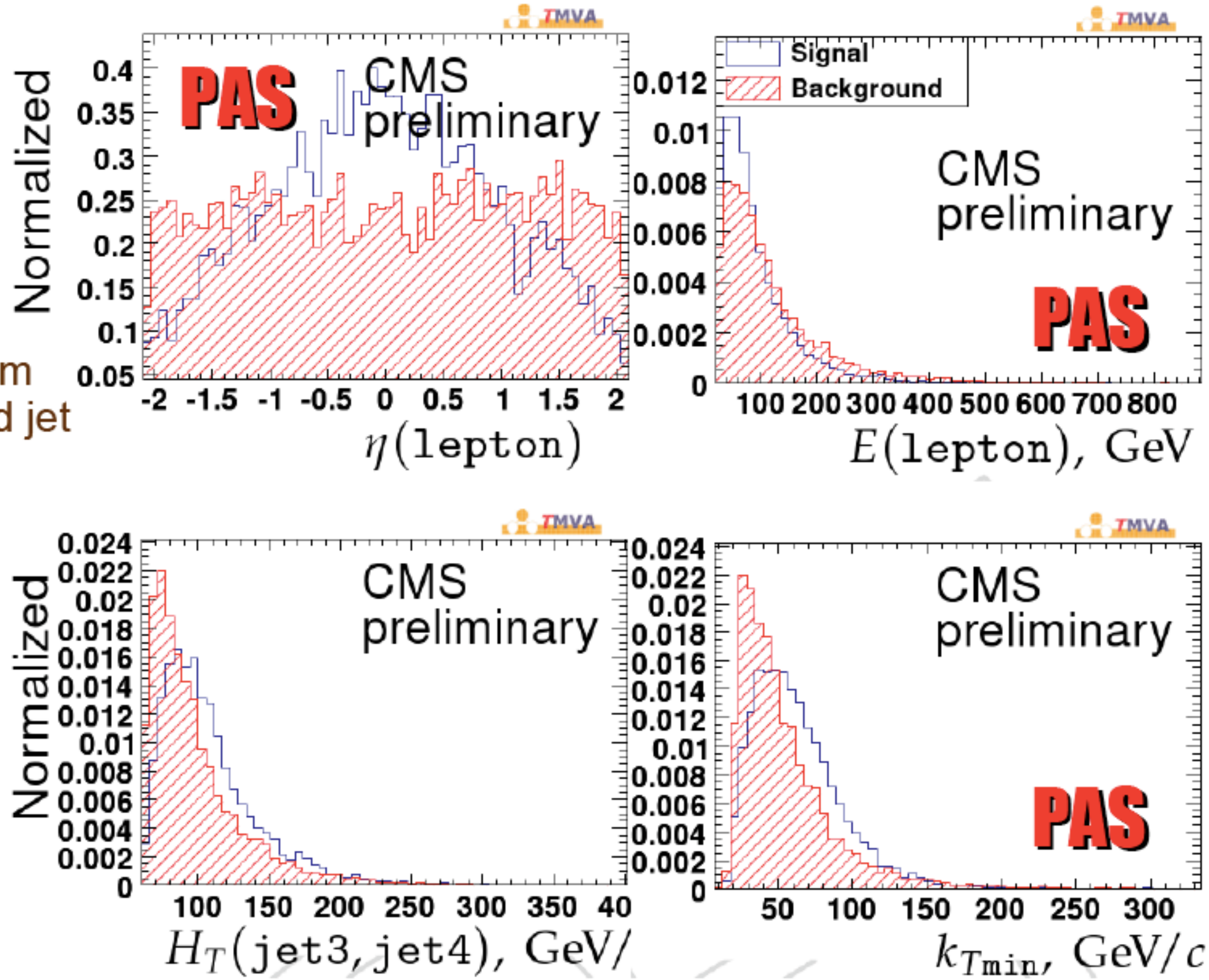
p_T – transverse momentum

k_{Tmin} – in a pair of jets, closest in R, smaller transverse momentum relative to the second jet axis

H_T – scalar sum of transverse momenta

$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$

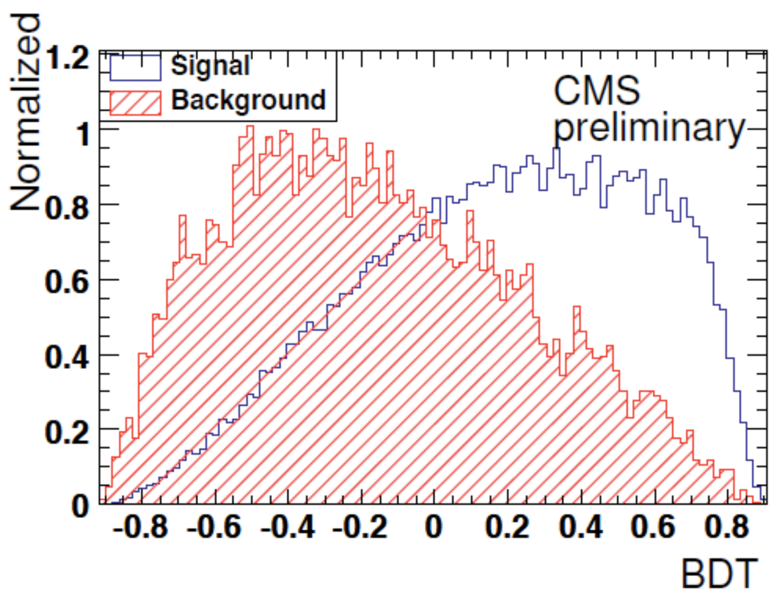
Aplanarity – 3/2 of the smallest eigenvalue of the normalized momentum tensor



Boosted Decision Trees

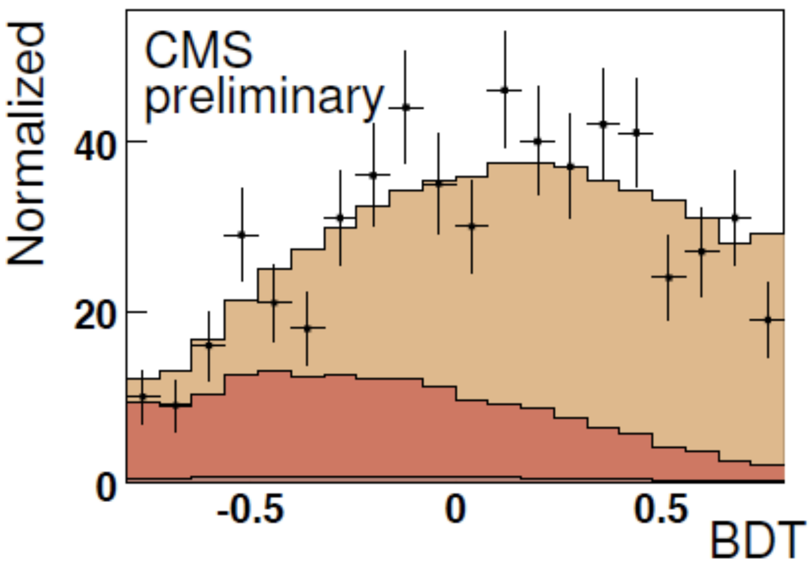
- A "forest" of binary decision trees
- Series of classification decisions, each based on a subrange of values of a single variable
- A misclassified event from the training sample gets larger weight in constructing the next tree (that's boosting)
- We use the TMVA implementation
- The variables used and the classifier shapes are shown

	variable	fraction of BDT nodes
1	$H_T(\mathtt{jet3,jet4})$	1.288e-01
2	$\Delta R(\mathtt{jet1,jet2})$	1.175e-01
3	aplanarity(jets only)	1.149e-01
4	$k_{T\mathtt{min}}$	1.114e-01
5	$\eta(\mathtt{lepton})$	1.113e-01
6	$H_T(\mathtt{jet3,jet4})/H_Z$	9.658e-02
7	$E(\mathtt{lepton})$	9.128e-02
8	$\Delta R_{\mathtt{min}}(\mathtt{lepton, jet1\ or\ jet2})$	8.125e-02
9	$p_T(\mathtt{jet4})$	7.821e-02
10	$\not{E}_T$	6.876e-02

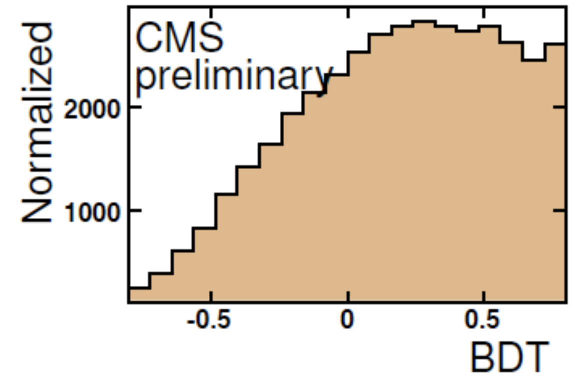


Templates and the Fit

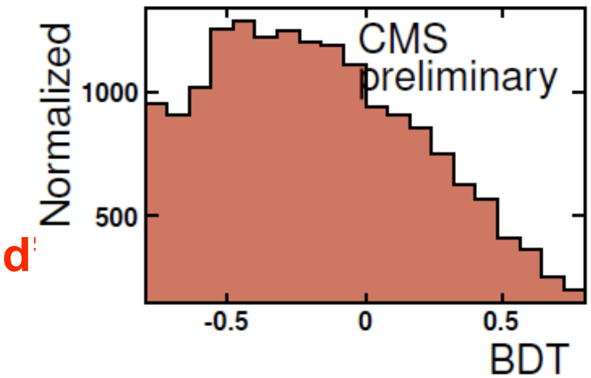
- We use TFractionFitter, the statistical uncertainty of the templates is accounted for
- "Physics background" includes W+jets, Z+jets and single top events
- QCD template is evaluated from data by using an enriched QCD sample (isolation requirement reversed, EM and hadron vetos removed)



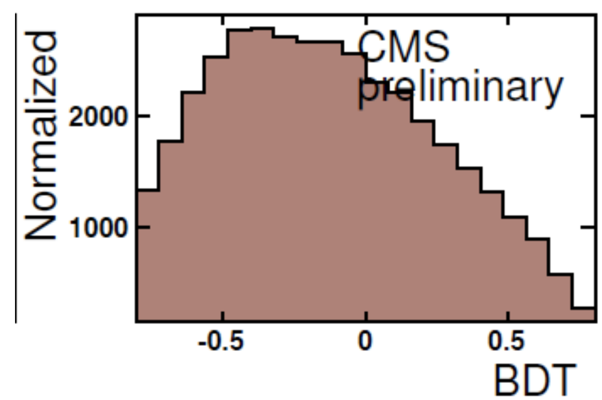
ttbar



"Physics Background"



QCD



Closure Test

- Mix the amount of events predicted for 20 pb^{-1}
  - Ttbar, W+jets, Z+jets, Single top and QCD
- Fit and extract fractions of signal and background
  - The yields agree with expected values
  - Estimated statistical error agrees with the senstivity measured using pseudoexperiments
- We derive the cross section as

$$\sigma = \frac{N_{signal}}{L \cdot \epsilon_{t\bar{t}}}$$

Signal yield (N_{signal})

Integrated luminosity (L)

Selection efficiency ($\epsilon_{t\bar{t}}$)

- Template fit results using an equivalent of 20 pb^{-1} of mock data set

parameter	true values	fit result (fraction)	extracted yield
$t\bar{t}$ yield	392	0.667 ± 0.054	391 ± 32
physics background yield	174	0.317 ± 0.049	186 ± 29
instrumental background yield	20	0.016(fixed)	–

Systematic Uncertainties

source	uncertainty
Muon ID and trigger	1.0%
Modeling of $t\bar{t}$ MC (MADGRAPH vs. PYTHIA)	6.2%
$W+$ jets MC matching threshold	1.2%
$W+$ jets MC factorization scale	1.7%
Multijet background normalization	0.2%
PDF	5.0%
Jet energy scale	17.1%
Initial and final state radiation (ISR/FSR)	
large Λ_{QCD} ISR/FSR	6.1%
"no power shower" ISR/FSR	8.5%
Total	21.7%

Summary of various sources of systematic uncertainties

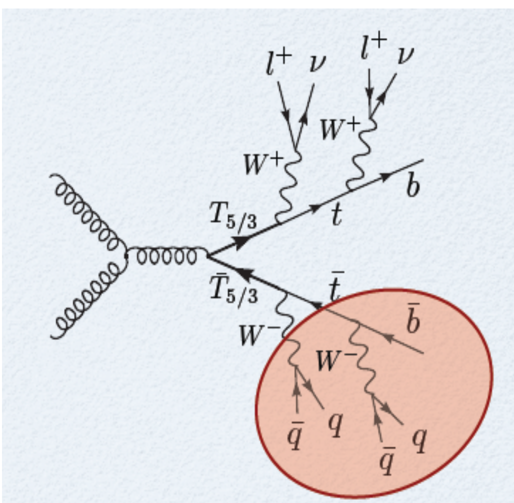
- The methods sensitivity is studied using Monte Carlo simulated events, and obtain an expected statistical uncertainty of 8.6% and an expected systematic uncertainty of 21.7% (and 10% expected uncertainty related to the integrated luminosity estimate)
- The largest contributions to the systematic uncertainty are jet energy scale, integrated luminosity and ISR/FSR

Search for Exotic Top Partners: $T_{5/3}$ Top Partner Model

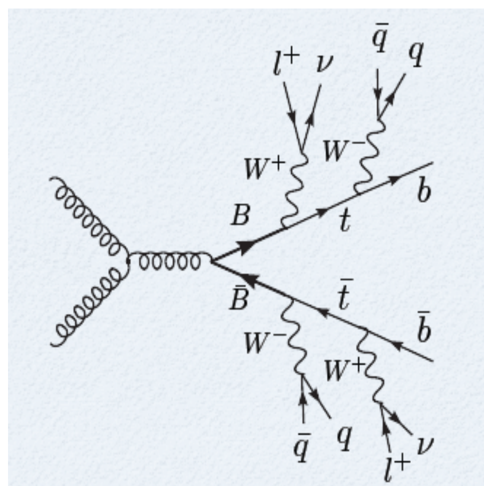
Aram Avetisyan, Tulika Bose
Meenakshi Narain

- Searches for two exotic particles- a heavy top quark partner with a fractional charge of 5/3, $T_{5/3}$, and its partner, the heavy B quark with charge -1/3
- These particles decay to a top quark and a W boson, leading to very busy events with multi-leptons and multi-jets
- Consider processes where same-sign dileptons are produced
- Backgrounds are predominantly from *ttbar*, QCD multi- jets, Z+jets, *ttbar*WW, *ttbar*W and multiple-W+jets

Heavy Top Partner



Heavy Bottom Partner



Signature:

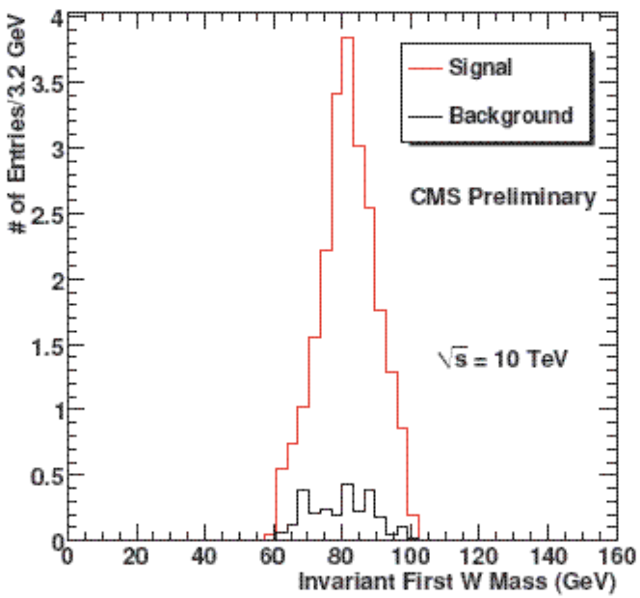
$$l^{\pm} l^{\pm} + n \text{ jets} + \text{MET} \ (n \geq 5)$$

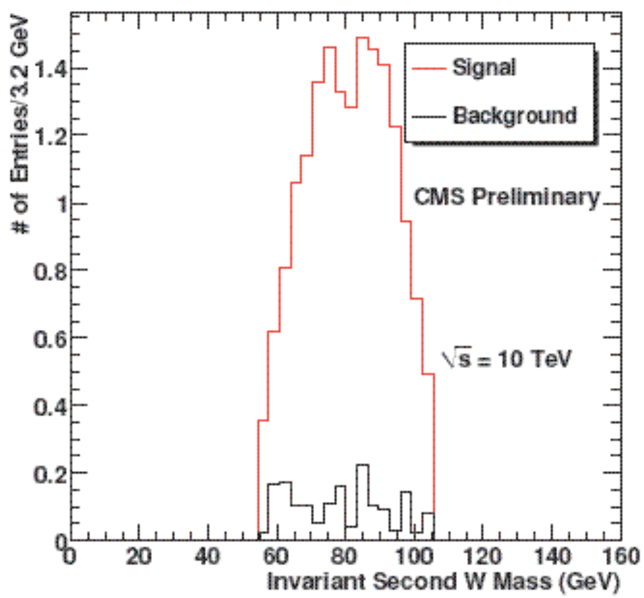
Event Selection

- At least 5jets with pT>30 GeV and | η | <3.5
  - p_T>100 GeV for the leading jet
  - p_T>80 GeV for the second-leading jet
- Two same sign isolated leptons, consider all possible pairs of electrons and muons: 2e, 2μ and e+μ
  - Muons are required to pass the tight quality cut. For the 2e sample, require the leading electron to be tight and the second electron to be medium or tight. For the e+μ sample, require a medium electron
- Electrons reconstructed in either the barrel or the endcap parts of the ECAL defined in the pseudorapidity regions of | η | <1.479 and 1.55<| η | <2.5
- Muons are required to be within | η | <2.4 pT>50 GeV for the leading lepton
- pT > 25 GeV for the second-leading lepton
- A 10 GeV veto around the Z mass for the 2e channel

Reconstruction of the Exotic Top Quark

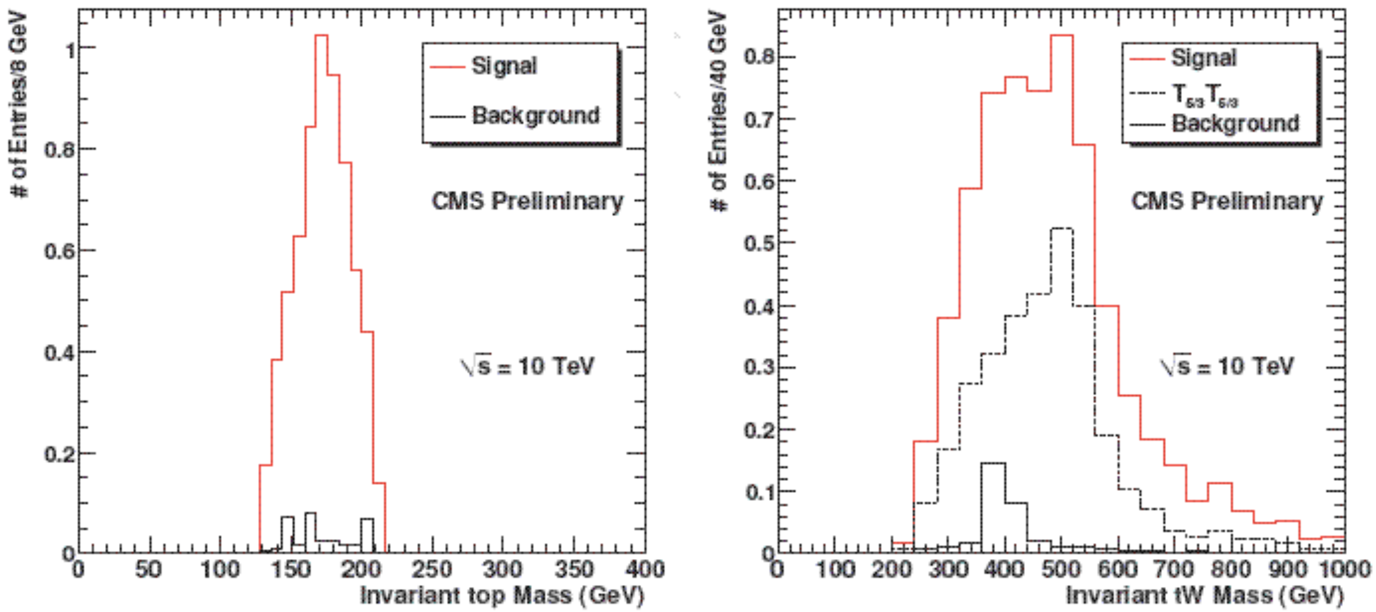
- The heavy top quark, $T_{5/3}$, with fully hadronic decay signature in the event can be fully reconstructed by reconstructing the two W bosons and the top quark in its decay chain
  - The two W bosons and the top quark are reconstructed
  - Reconstruct the other W boson
  - Reconstruct the top quark from $T_{5/3}$ decay
  - Finally, the top quark candidate is combined with the other W boson to yield the mass of the heavy top $T_{5/3}$ in the event





The invariant mass distributions of the first(left) and second (right) W bosons for signal at M=500 GeV and physics background samples at 10 TeV

Reconstruction of top and $T_{5/3}$



The invariant mass distributions of the reconstructed top quark (left) and tW (right) for signal at M=500 GeV and physics background samples at 10TeV

- A means of searching for exotic partners of the top quark, the $T_{5/3}$ and B
- At 10 TeV, ~1.6 fb^{-1} of integrated luminosity needed for a 5 σ observation of sum of $T_{5/3}$ and B

Data Driven Method for QCD and W+jets normalization

Ioana Anghel, Cecilia Gerber

- Iterative Kolmogrov Smirnov Method
- 2 W bosons present in the decay of a ttbar pair
  - Decay channel classified based on the decay of the W's

Event Selection

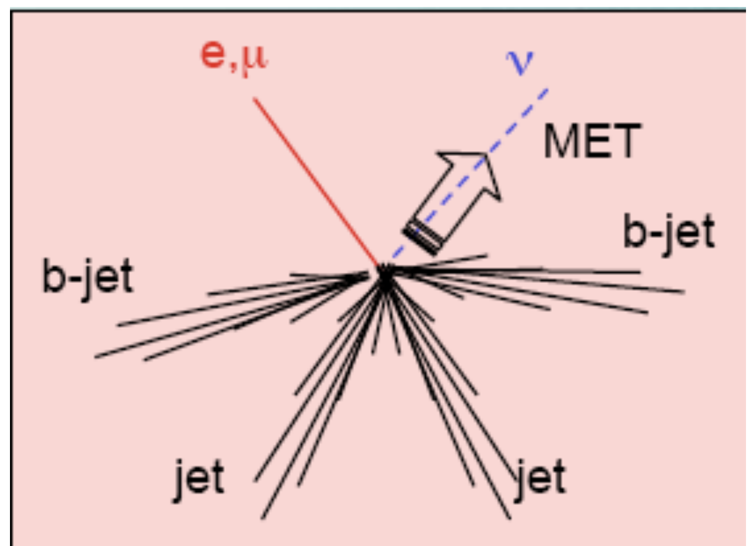
"good jets"
- Pt > 20 GeV
- -2.1 < Eta < 2.1

"good muons"
- Pt > 20 GeV
- -2.1 < Eta < 2.1
- normalized Chi2 < 10
- nHits >= 11
- d0 < 2 mm

muon is isolated
- RelIso : $RelIso2 = \frac{TkIso + ECalIso + HCalIso}{p_T}$

ΔR(good muon, good jet)
- ΔR > 0.3

MET > 15 GeV



Semi-leptonic Decay
(BR~30%, moderate bckg)

Method

- Method determines the absolute normalization of the multijet and the total W +jets sample, same idea as the DØ Matrix Method
- At Tevatron (DØ), it was found that the variables that are the most sensitive to QCD and W+jets samples are lepton p_T, E , M_T(W)

$$N^{\text{data}}_{\text{pretag}} - N^{\text{MC}}_{\text{bkgd}} = S_{\text{W+jets}} N^{\text{MC}}_{\text{W+jets}} + S_{\text{multijet}} N^{\text{data}}_{\text{multijet}}$$

where S_{W+jets} and $S_{multijets}$ are the normalization scale factors

- QCD template shape can be obtained from data (anti-isolation or similar)
- W+jets template shape obtained from MC, small background from MC

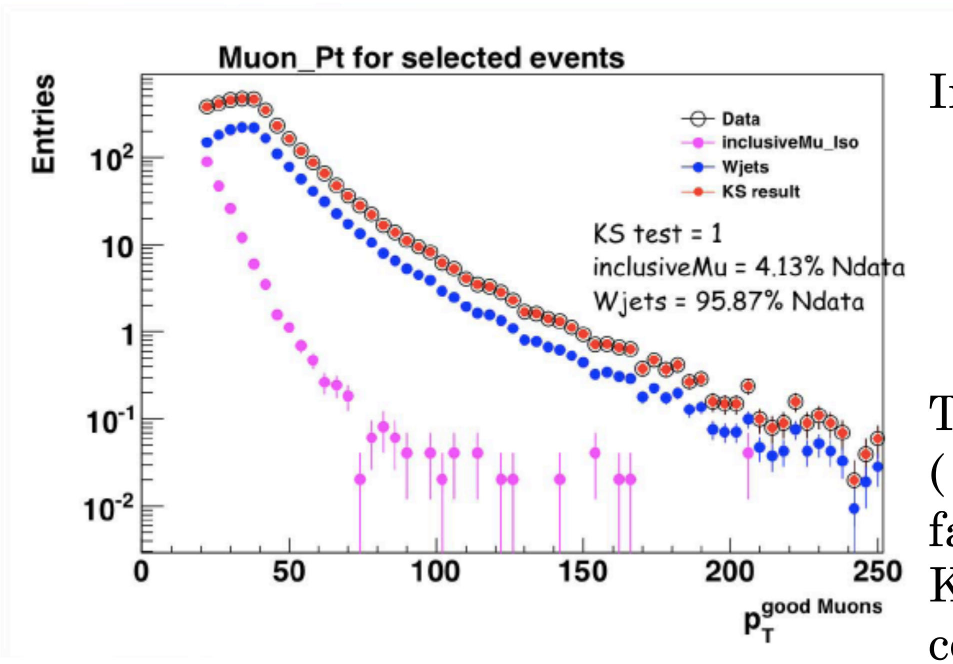
1. Select a kinematic variable sensitive to the normalization
2. Since S_{W+jets} and $S_{multijets}$ are anticorrelated, we start by setting S_{W+jets} = 1 and calculate the corresponding $S_{multijets}$ using equation from previous slide.
3. The differences between the mixed sample and the data can be seen from the KS test: KS takes values from 0 (totally different) to 1 (exactly the same)
4. Start increasing S_{W+jets} by 0.01 and repeat the steps until $S_{multijets}$ becomes negative.
5. Select the values of S_{W+jets} and $S_{multijets}$ which gives the maximum KS-test value.

The process is repeated for each of the chosen sensitive variables.

Closure Test

- Start with two templates of W+Jets and multijets
- In order to construct the “data” sample, make a linear combination of the 2 templates:

  $Param_{Data} = a^* \, Param_{Multijets} + b^* Param_{Wjets}$
- Perform a “closure” test



In the “Data” sample, there are:

~4% of N_{events} from Data represents inclusiveMu sample

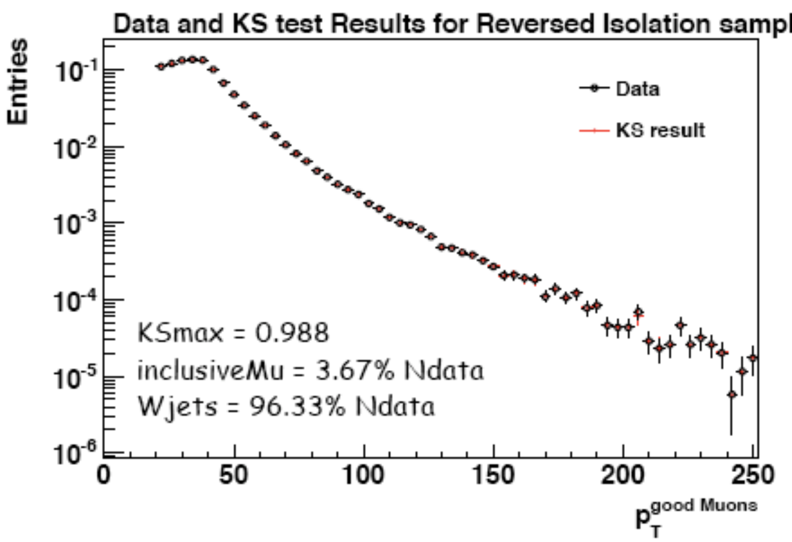
~96% of N_{events} from Data represents Wjets sample

The percentage of the two samples (corresponding to a,b scale factors) obtained for the maximum KS test are the same as used to construct the data

Test with a Different Multijet Sample

- Use different multijets samples (obtained from data) and apply the method, and compare the results with the percentages used to construct the data
- The multijets sample intended to be used in the method is obtained from the data by:
  - Reversing the muon Isolation (Iso > 0.1), keeping the rest of the preselection cuts (good muons, good jets, DR(muon,jet) > 0.3 and MET > 15 GeV)
  - Reversing the d0 cut (d0 > 0.04) keeping the rest of the preselection
  - Eliminate the closest jet to the muon and recalculate MET: newMET = oldMET + jet

~ 4% of Data is multijets and ~ 96% of Data is Wjets



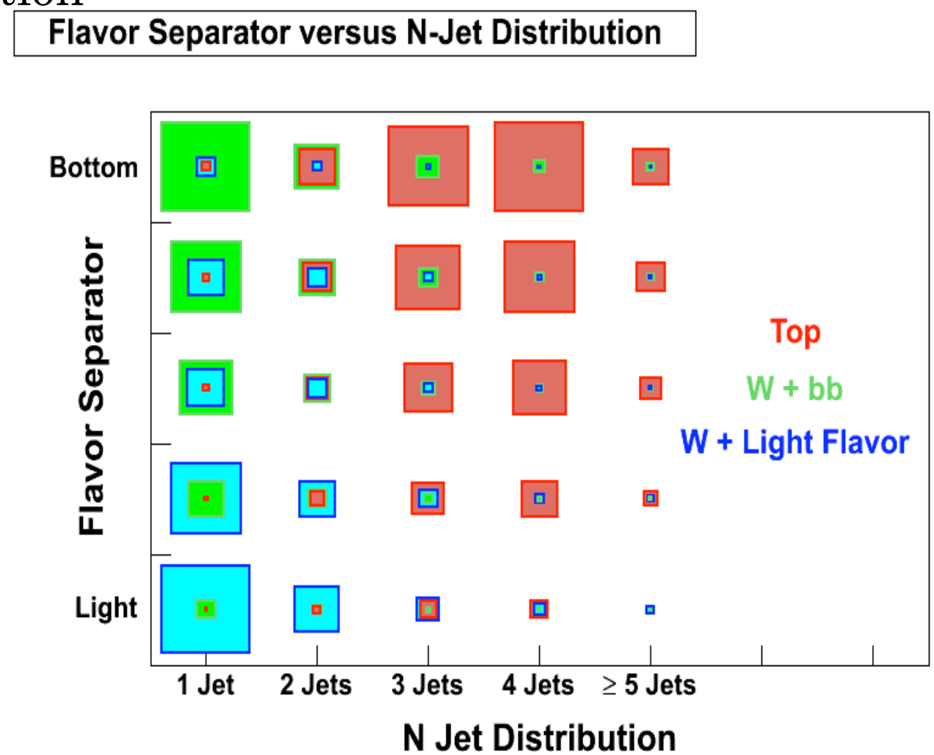
The scale factors obtained by the IKS method are very close to the ones present in the data

Simultaneous Heavy Flavor and Top Cross Sections Measurement

- Using a sample of Leptons + Jets:
  - High p_T muon
  - At least one jet and at least one b-tag
- Measure Physics observables:
  - Top pair and single top cross section
  - Wbb and W+charm cross section
- Measure Calibration observables
  - b-tag scale factor

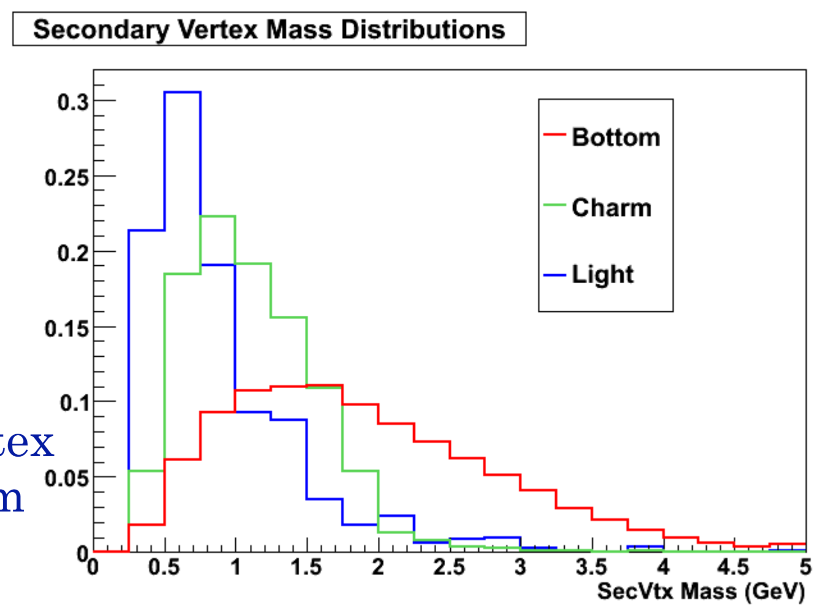
C. Plager, B. Hegner, E. Kennedy,
D. Fehling, G. Giurgiu, G. Hu,
P. Maksimovic, S. Rappoccio, A. Ivanov,
S. Malik, K. Lannon, J. Slaunwhite,
S. Boutle, C. Neu

- What different discriminating variables are available:
  - N-Jet distribution
    - Distinguish Top from Wbb
  - "Flavor separator"
    - Distinguish Wbb from Wc/Wcc and W Light Flavor



SHyFT Introduction

- Split events into single and double tags
- W + jets not treated as a background, but as another signal
- "Everything" is fit simultaneously
  - Top pair, single top, Wbb, Wcc, Wc, Wqq
  - EW (Z + jets, diboson, etc) and QCD are constrained to calculated/fit cross sections
- Systematic uncertainties are fit:
  - Tag uncertainties
  - Jet energy scale
  - Q^2 for W + jets
  - ISR/FSR for top, etc

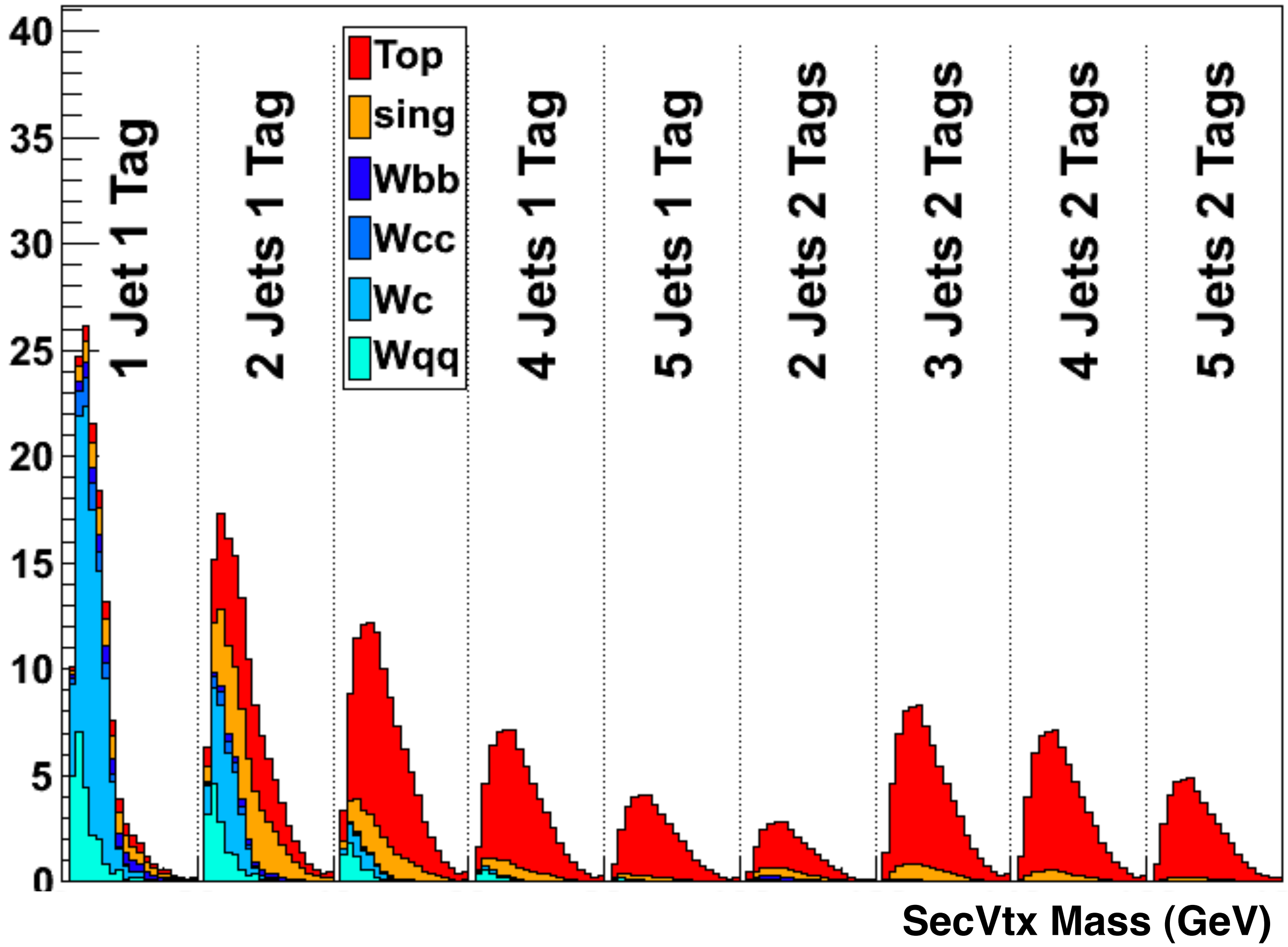


- Use invariant mass of tracks originating at identified secondary vertex
- Good separation of bottom from charm and light flavor jets

SHyFT Fit

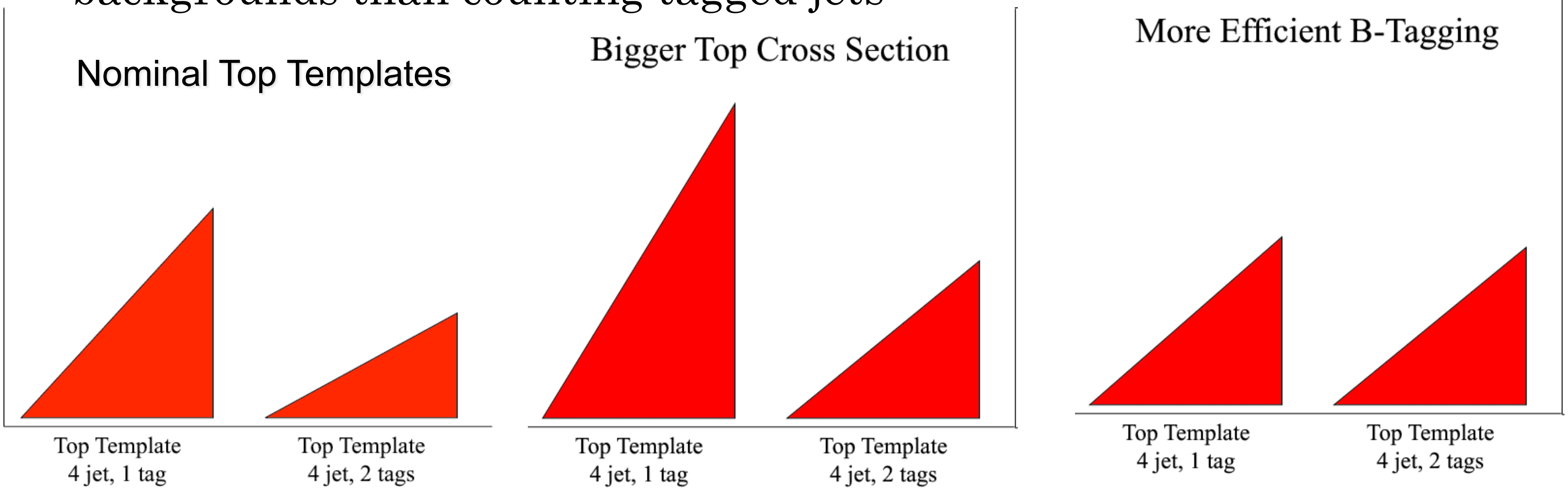
Method III Fit

$$\int \mathscr{L} = 10\ \mathrm{pb}^{-1}$$



Why Separate Double Tag Events?

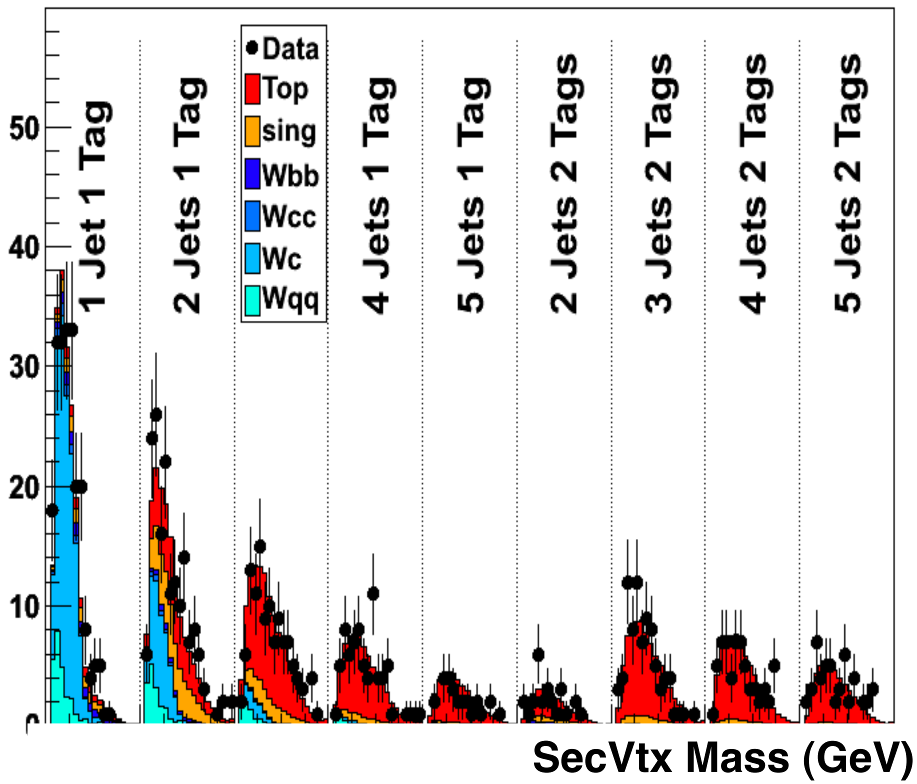
- Double tag events are much higher purity (for top and Wbb)
- Single tag events have much higher acceptance
- How can we have the best of both worlds?
  => fit single and double tag events separately
- Data can be used to help minimize our sensitivity to uncertainties in the b-tagging scale factor
- Using flavor separator makes us less sensitive to non-bottom backgrounds than counting tagged jets



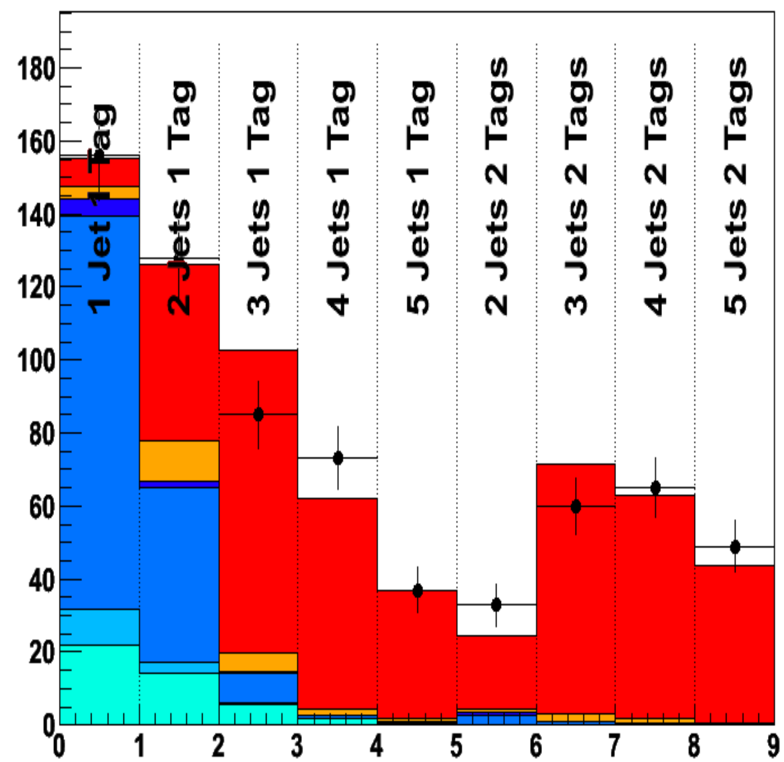
Example Fit to Pseudo-Experiment

$$\int \mathscr{L} = 10\ \text{pb}^{-1}$$

Method III Fit



Method III N Jet Distributions



SHyFT

- *SHyFT* works very well for measuring the top and W + heavy flavor cross sections
- Can also be used to estimate backgrounds for searches
- *SHyFT* will also help us calibrate our detector using the same data events we use for physics
- Currently only using muons and work on electrons coming soon
- Already good sensitivity to top with 10 pb^{-1} of integrated luminosity
- *SHyFT* will be orthogonal to many other methods of measuring cross sections
  - Allow cross checks with other methods
  - Better sensitivity for combinations

ttbar cross section in muon + jets using a soft muon tagger

Ken Bloom, Helena Malbouisson

- W+jets backgrounds are large for untagged analysis; they are probably still the largest backgrounds for a tagged analysis
- Tags in W+jets events can arise from
  - Actual heavy flavor produced in these events through gluon splitting; difficult to model in MC, will need to be measured in data
  - Fakes of some sort, e.g. decay in flight, hadron punch through, track mistakenly matched to something in muon system-> to be measure in data
- Main idea: use complementary event samples with related physics processes to model the SMT tag rate, apply the measured tag rate to l+jets events selected for ttbar analysis to estimate tagged W+jets contamination
- This assumes that the jets in the other samples are similar to those in W +jets. Works best if heavy-flavor is relatively small and fake tags dominate

Taggable Tracks

- Tags are due to tracks that are identified as muons - define a "taggable track" as one that can potentially cause a jet to be tagged.
- Identify taggable tracks in γ+jets events, and measure the probability that a track of some pT, η, ϕ causes a jet to be tagged.
  - γ+jets should be similar in structure to W+jets, and plentiful
- Take those tag rates and apply them to taggable tracks in Z+jets events. See if the predicted number of tags matches up with observed to some reasonable precision

Background estimation:Tag matrix (Fake + HF)

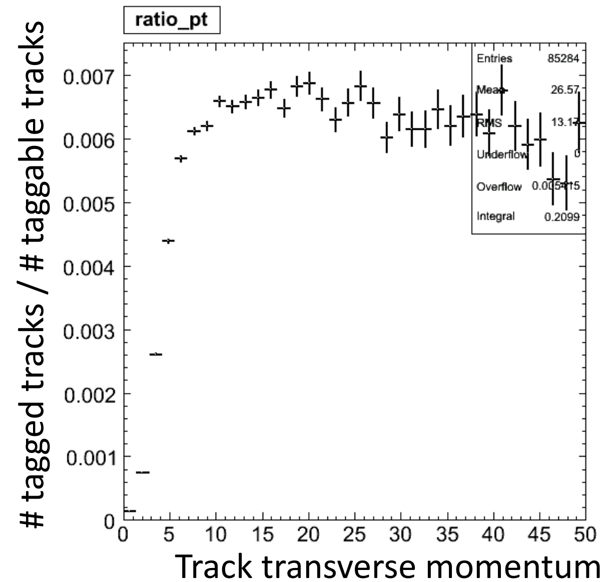
- Compute the probability of tagging a jet as a function of track p_T and detector η ➔ **tag matrix**;

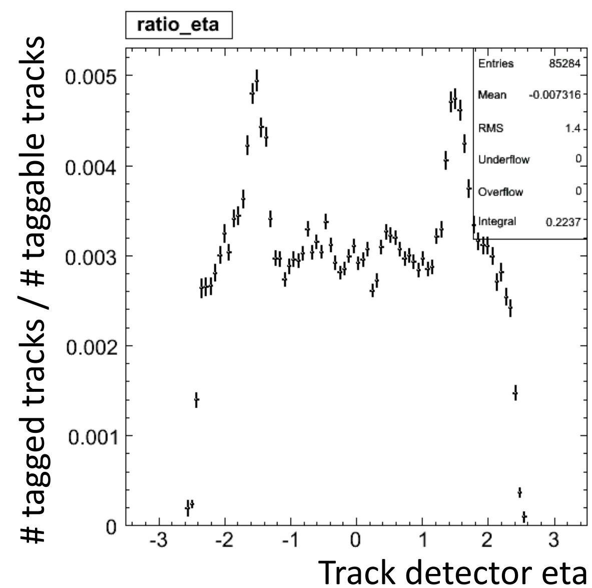
- The tag matrix is defined as: $\dfrac{\#\ tagged\ tracks}{\#\ taggable\ tracks}$

- Where tagged tracks are defined as the track matched to the (soft) muon that b-tagged a jet, and taggable tracks are tracks that passed track quality criteria and are close to a jet;

- **method:** obtain the tag matrix (probability of a given track in a jet to yield a SMT tag) from a γ + jets sample and use it as the estimate of tagged events of W+jets (including Heavy Flavor) and fakes;

- **Plots are still very preliminary**





Summary

- Lots of interesting work being done in the group
- Lots of constructive discussions and friendly feedback
- Many common issues, tools
  - Estimating multijet/fake lepton backgrounds from data
  - Understanding W+jets with data
  - ttbar both as a source of physics and an important background
  - Generator tuning
- There's a lot of work to be done so come join us and get involved!