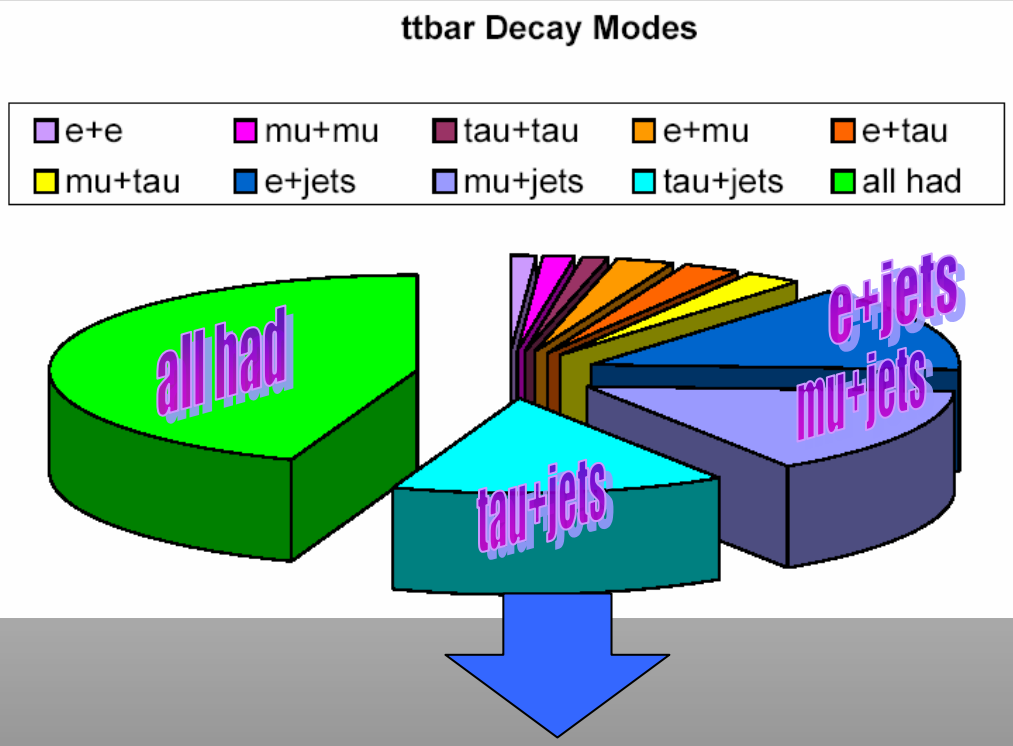
$t\bar{t} \rightarrow \tau$+jets Analysis Overview

Giorgio Cortiana,
University of Padova and INFN

Outline:

- *Introduction and motivation*
- *Datasets and method summary*
- *Method-I background prediction*
- *Kinematical selection optimization*
- *Preliminary results*

Introduction



tau+jets

☺ Not observed yet, can add a new piece of knowledge in the top physics sector

☹ Large background: QCD, EWK+HF

❑ taus are reconstructed as jet-like objects: more challenging signature compared to e/μ.

❑ Will use MET rather than lepton ID: Extra Acceptance from "dirty" e/μ+jets events

- Need optimized kinematical + topology selection.
- Need b-jet identification to increase S/N ratio.

- b-jet identification rates are different on ttbar and background processes: can distinguish the two components:

→ B-tag rate parametrizations → Tagging matrix

Datasets & Method

Datasets and trigger:

TOP_MULTI_JET *dataset* up to Aug 2004: 311 pb^{-1}.

- **L1: ≥1 cal. tower with E_T≥10 GeV;**
- **L2: ≥4 cal. clusters with E_T ≥ 15 GeV, ΣE_T ≥ 125 GeV;**
- **L3: ≥4 jets, R=0.4, E_T ≥10 GeV**

MC : (167 fb^{-1}), Pythia ttbar, M_{top} = 178 GeV

Good Run List: v7.0 (1,1,4,1)

Offline version: 5.3.3_nt5

Jet Corrs: jetCorr04b

High missing Et Selection?

yes

τ+jets analysis
ttbar→blν bbarjj

no

All Had analysis
ttbar→bjj bbarjj
See Andrea's talk

Method-I approach
+
ad hoc Kinematical selection

- **Method 1: positive tagging matrix approach to predict the absolute amount of background**
- **Kinematical Selection + ≥ 1 SECVTX positive tag**

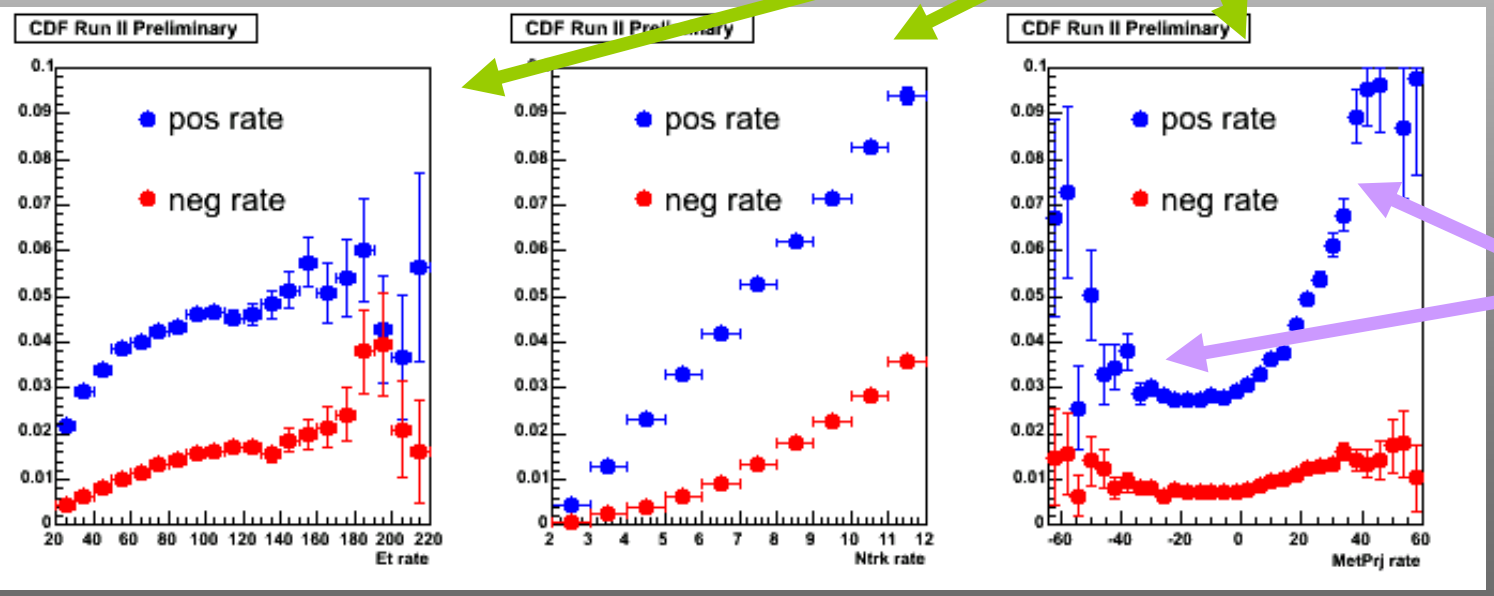
ttbar cross section measurements in multi-jet final states

Method-I Background Prediction

- b-jet identification rates are different on ttbar and background processes: can distinguish the two components:

- Look at the B-tag rates directly from TOP_MULTI_JET data
- Use 3 (E_T^{L5} > 15 GeV, $|\eta|$<2.0) jet events: $F_{top} = 2\times10^{-5}$
- Take the vars by which the tag-rate mainly depends
- Construct the positive tagging matrix

- jet E_T
- jet N_{TRK}
- Met *cos $\Delta\phi$(met,jet)



Asymmetric distribution: due to different concurring processes, HF+jets EW+jets

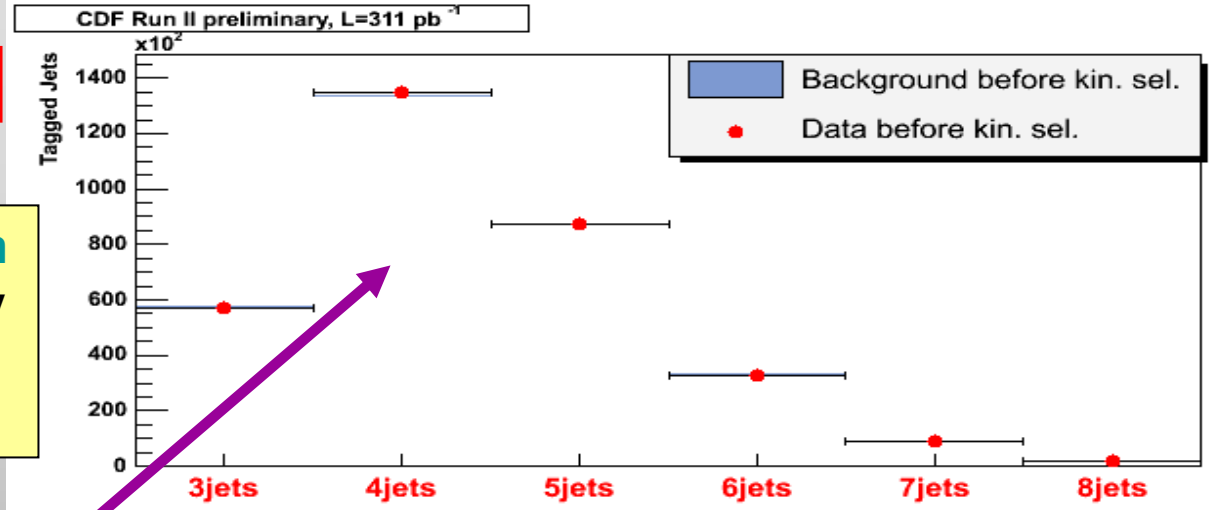
Allow to track sample composition changes with MET

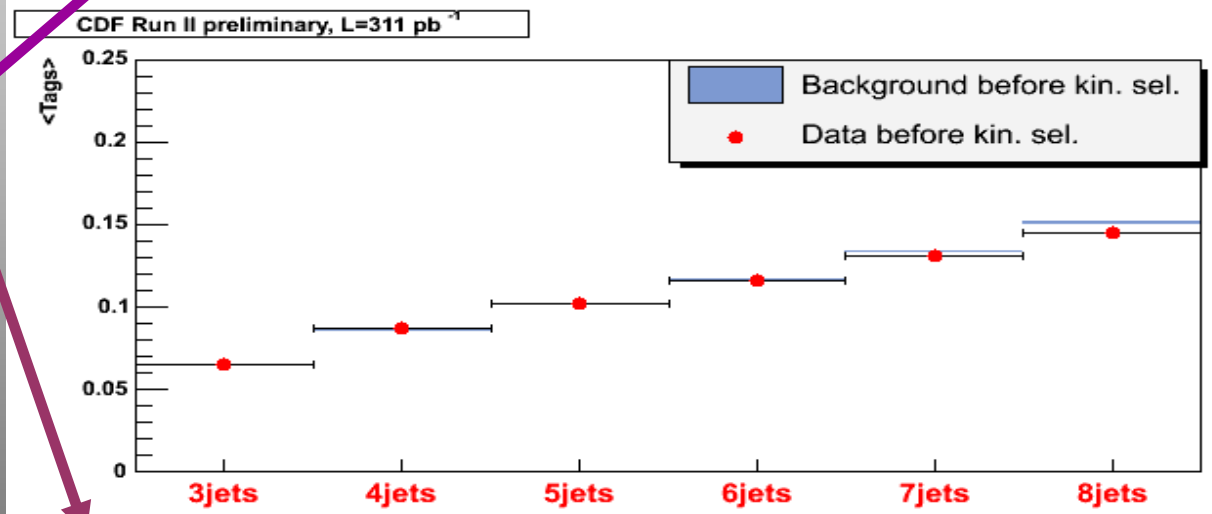
3-d (E_T, N_{TRK}, Met_{PRJ}) Positive Tagging Matrix constructed on 3 (E_T^{L5}>15GeV, $|\eta|$<2.0) jet data events. Use it to extrapolate the tag rate to higher jet multiplicities

Mtx Check #1

Extrapolate the tag rate from 3 jet to higher jet multiplicity events, before kinematical selection.

The agreement between observed and matrix-predicted positive tagged jet is good for all jet multiplicities





N jet	3	4	5	6	7	8
Nevts	879,187	1,553,525	859,543	284,062	68,628	13,237
N tgbl jets	1,781,645	4,197,888	2,877,148	1,134,089	318,410	69,966
Obs + tags	57,314	135,056	87,332	32,914	8,992	1,914
Exp + tags	$57,314 \pm 233$	$13,3275 \pm 546$	$87,156 \pm 370$	$33,184 \pm 149$	$9,147 \pm 43$	$2,000 \pm 10$
					Total observed + tags:	324,008
					Total expected + tags:	$322,575 \pm 1337$

Kinematical Selection

Clean up selection:

- **Tight leptons (e/μ) veto** (no overlap w/ other L+J top analyses)
- **Trigger requirement simulation** (for MC events)
- **Vertex requirements**:
  - $|Z_{vert}| < 60$ cm
  - $|Z_{vert}^{jet} - Z_{pvert}| < 5$ cm
  - $N_{vertices}(Q>12) \geq 1$

Optimize the kinematical selection in order to minimize the relative statistical error on xsec using both the expected amount of tags for inclusive ttbar and background (from matrix)

Optimization procedure:

- **Start by selecting ≥4 jets** (matrix is computed with =3 jet events)
- **Scan different sets of requirements (metsig, A, min$\Delta\phi$)**
- **Calculate the amount of expected bkg tags for a given cut set**
- **Instead of N_{obs}^{tag} use $N^{tag}_{mc} + N^{tag}_{exp}$**

$$\sigma_{ttbar} = \frac{N_{obs}^{tag} - N_{\exp}^{tag}}{\varepsilon_{kin} \cdot \varepsilon_{tag}^{ave} \cdot L}$$

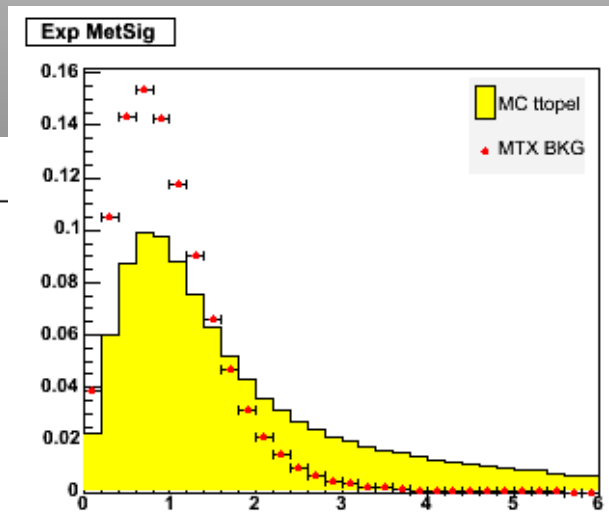
Choose the set of cuts that minimizes the expected (stat. only) relative error on xsec

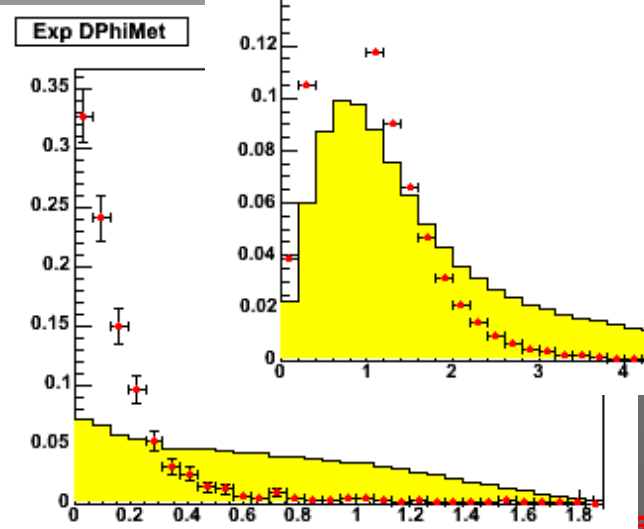
Kinematical Selection – cont'd

- N jets(E_T≥15 GeV; |η|<2.0) ≥ 4
- $\not{E}_T / \sqrt{\sum E_T} \geq 4.00$
- min Δφ(met,jet) ≥ 0.4 rad

cut set				MC^{evt}_{inc}	MJ^{evt}	MC^{tag}_{inc}	BKG^{tag}	$S/\sqrt{N}$	$\frac{\sigma_{xsec}}{xsec}$
1	4.00	0.00	0.40	92	597	73	68 ± 4	8.83	17.59%
2	4.00	0.00	0.50	81	461	64	46 ± 3	9.45	17.61%
3	4.00	0.01	0.40	89	549	70	63 ± 4	8.84	17.73%
4	4.00	0.01	0.50	79	426	62	43 ± 3	9.49	17.74%
5	4.50	0.00	0.30	84	446	67	59 ± 3	8.72	17.95%
...	...	...	...	...	...	...	...	...	...

Extra acceptance come from e/μ + jets ttbar events failing the tight lepton identification requirements





before tagging

N evt	$MC_{(\tau+jets)}$	$MC_{Incl.}$	MJ
Total	149,234	1,021,924	4,249,644
Prereq	74,786	556,575	3,897,755
NJet	72,597	547,309	2,781,788
MetSig	29,768	77,545	3,996
DPhiMetJet	19,035	49,421	597
in 311 pb^{-1}:	35.36	91.81	597
S/N $\tau + jets$: $N^{\tau}_{mc}/(N^{data}_{obs} - N^{incl}_{mc})$			0.07
S/N Inclusive: $N^{incl}_{mc}/(N^{data}_{obs} - N^{incl}_{mc})$			0.18

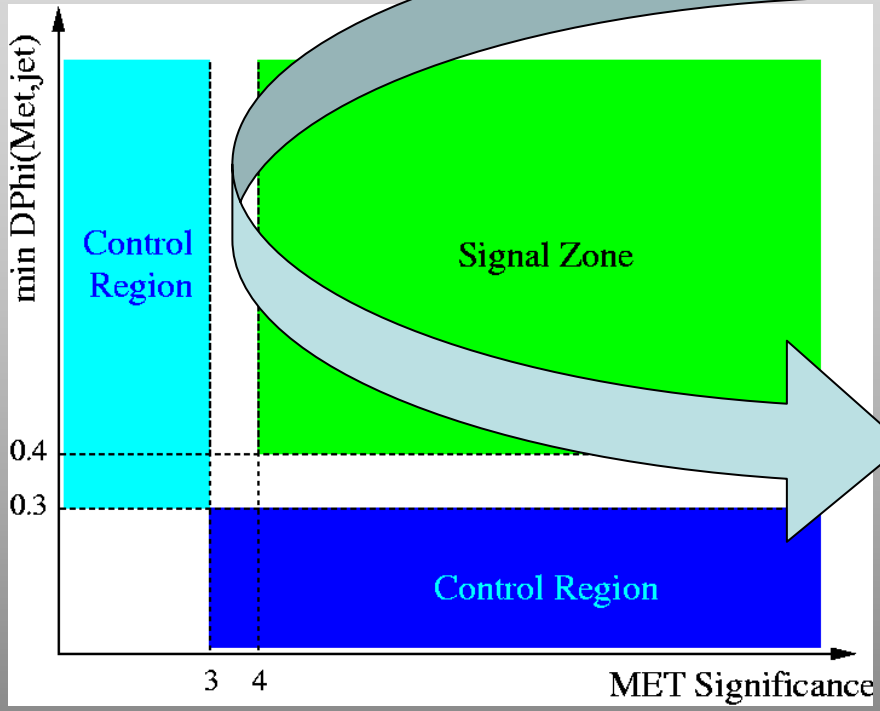
after tagging

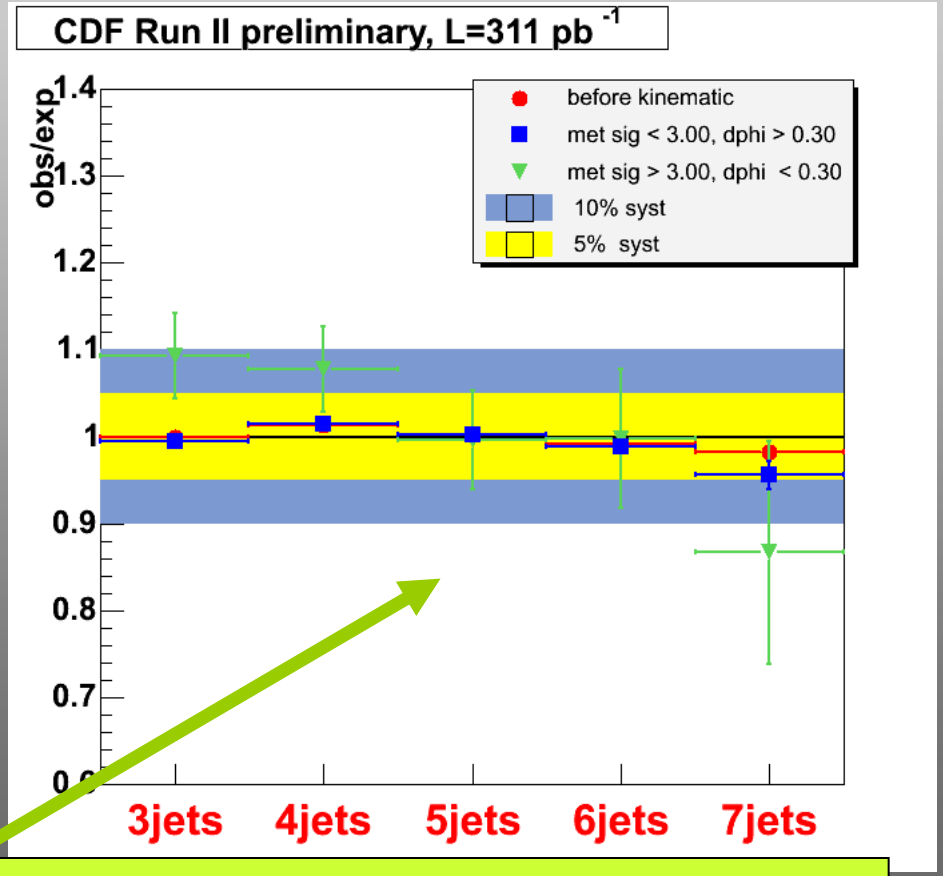
DPhiMetJet	19,035	49,421	597
≥ 1 tag	11,609	30,182	106
in 311 pb^{-1}:	21.6	56.1	106

More on Matrix checks:

Once we have an optimized kin selection:

The tagging matrix background predictions can be checked in control samples obtained from multi-jet data itself:





- data before kinematical selection
- data: met sig < 3 and min$\Delta\phi$ > 0.3
- data: met sig > 3 and min$\Delta\phi$ < 0.3

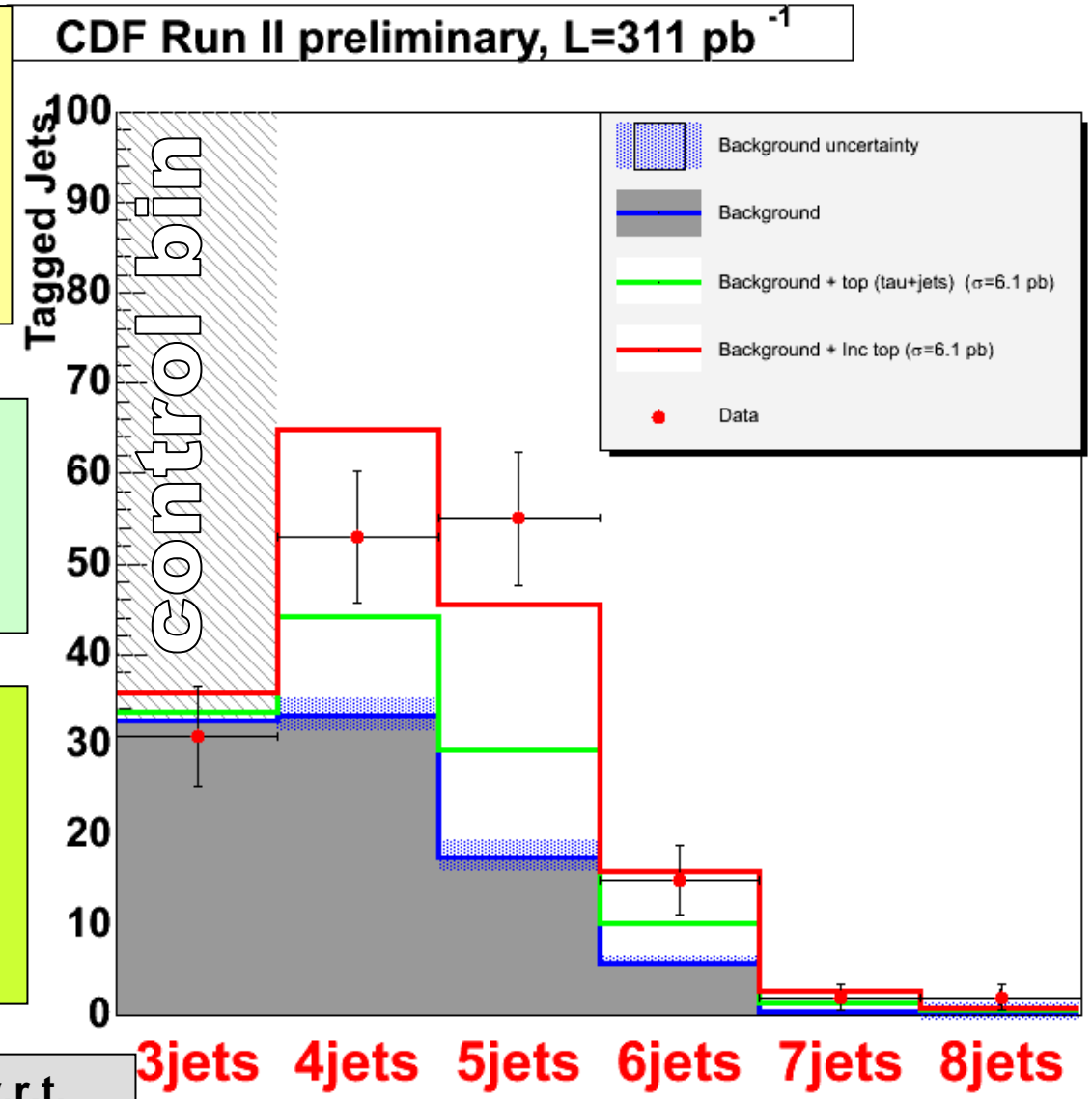
The matrix performs well in the control samples, the discrepancies in terms of the ratio obs/exp tags being limited at 10 %.

Kin sel + ≥1 tag Sample:

Once we feel confident about our matrix parametrization we can look at its prediction in the data sample after kinematical selection and compare it with SecVtX tagged data.

- N jets(E_T^{L5} ≥15; |η|<2.0) ≥ 3
- $\not{E}_T / \sqrt{\sum E_T} \geq 4.00$
- min Δφ(met,jet) ≥ 0.4 rad

- matrix-based background prediction is corrected with an iterative procedure to account for the ttbar presence in the pre-tag sample.

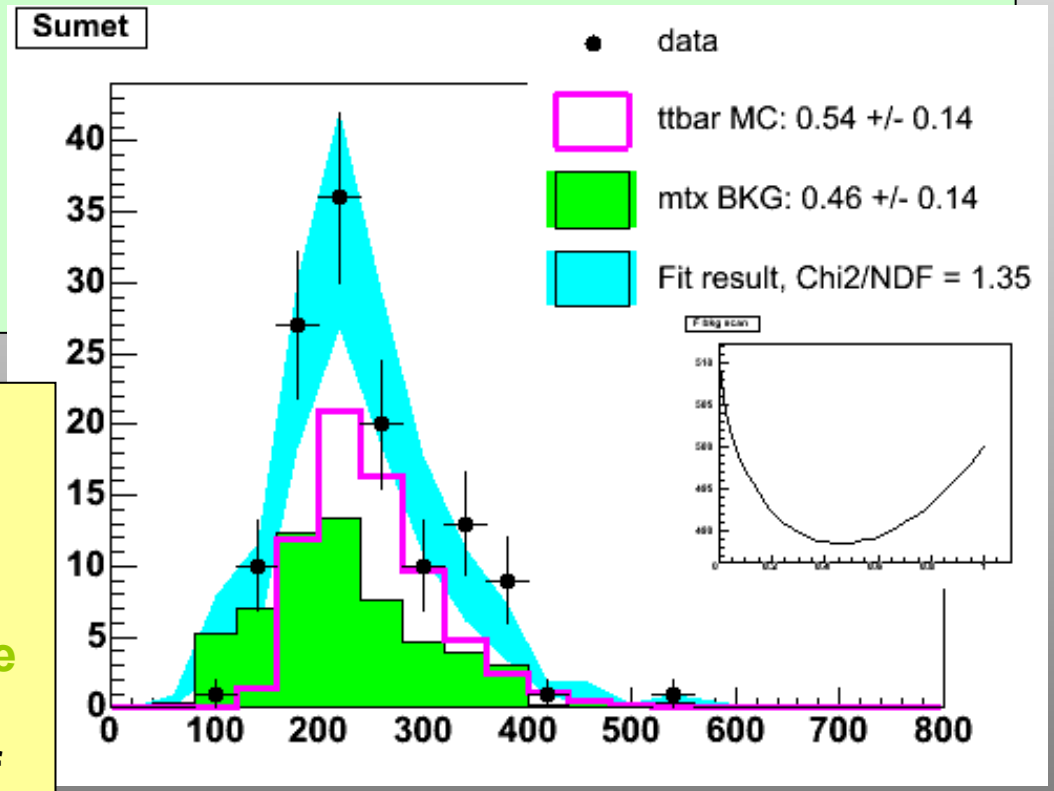


The excess is well consistent w.r.t. MC+BKG expectations in all jet bins!

Checking F_{top} using 2-c fits

We can cross-check the excess we attribute to ttbar production by looking to kinematical variables.

In particular we can use the positive tagging matrix to extract background shapes.



Then fit data distribution after kin sel + ≥1 tag to the sum of:

- Inclusive ttbar template
- Matrix extracted bkg template

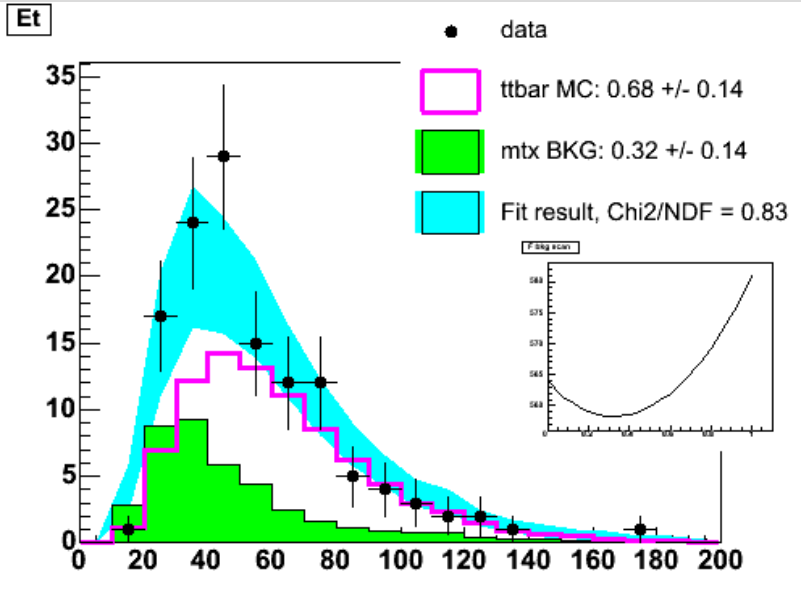
And extract the relative fractions of signal and background in the data

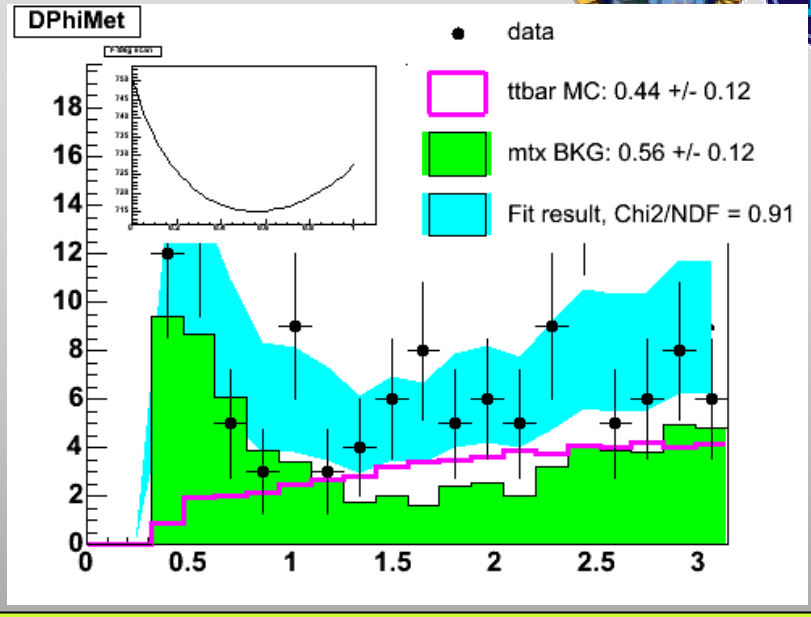
Used binned likelihood fits:

$$\begin{cases} L = -2 \cdot \sum_{i=1}^{Nbins} N_i \cdot \log(F) - F \\ where: F = f_b B_i + (1 - f_b) S_i \end{cases}$$

and checked the fitting procedure by pseudoexps.

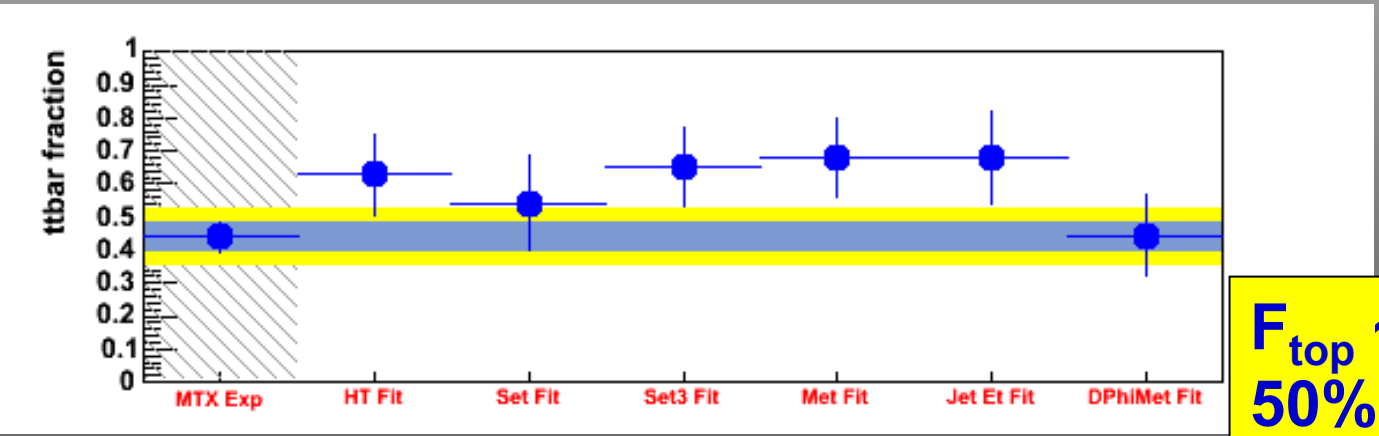
We also fitted, Set3, HT, jet ET and jet $\Delta\phi$ with MET distributions:





From data fits:

we found a ttbar fraction consistent with that calculated by the counting method.



Note that some of the determinations are correlated with each other (HT, Sumet, Sumet3, met)

F_{top} ~ 50%

After kin sel + ≥1 tag data: 50% top + 50% background

Systematic Uncertainties:

- **Almost done with systematic uncertainties determination, only ISR / FSR to be computed (waiting for prescriptions).**

Systematics uncertainty sources		
Source	Method	Uncertainty
ϵ_{kin} systematics		
Trigger simulation	turn-on curves	17.8 %
Generator dependence	$\frac{\lvert\epsilon_{PYTHIA}-\epsilon_{HERWIG}\rvert}{\epsilon_{PYTHIA}}$	8.2 %
PDFs	MC reweighting	1.6 %
ISR/FSR	samples comparison	N.A.
Jet Corrections	$\frac{\lvert\epsilon_{jetcorr,+1\sigma}-\epsilon_{jetcorr,-1\sigma}\rvert}{2\epsilon_{kin}}$	1.5 %
M_{top}	$\frac{\lvert\epsilon_{kin,M_{top}=170}-\epsilon_{kin,M_{top}=180}\rvert}{2*\epsilon_{kin,M_{top}=178}}$	2.8 %
ϵ_{tag} systematics		
SecVtX scale factor	$\frac{\lvert\epsilon_{tag,+1\sigma}-\epsilon_{tag,-1\sigma}\rvert}{2\epsilon_{tag}}$	5.5 %
Tagging matrix systematics		
Data control samples	N_{obs}/N_{exp}	10.0 %
Luminosity systematics		
Luminosity measurement	–	6.0 %

- **Total Systematic uncertainty: 24 %**

Cross Section:

$$\mathcal{L} = e^{-\frac{(L-\bar{L})^2}{2\sigma_L^2}} \cdot e^{-\frac{(\epsilon_{kin}-\bar{\epsilon}_{kin})^2}{2\sigma_{\epsilon_{kin}}^2}} \cdot e^{-\frac{(\epsilon_{tag}-\bar{\epsilon}_{tag})^2}{2\sigma_{\epsilon_{tag}}^2}} \cdot e^{-\frac{(N'_{exp}-\bar{N}'_{exp})^2}{2\sigma_{N'_{exp}}^2}} \cdot e^{-\frac{(N'_{exp}-\bar{N}'_{exp})^2}{2\sigma_{N'_{exp}}^2}} \cdot \frac{(\sigma_{t\bar{t}} \cdot \epsilon_{kin} \cdot \epsilon_{tag} \cdot L + N'_{exp})^n}{N_{obs}!} \cdot e^{-(\sigma_{t\bar{t}} \cdot \epsilon_{kin} \cdot \epsilon_{tag} \cdot L + N'_{exp})}$$

- **The cross section is measured by means of a likelihood function maximization:**

Variable	Simbol	Input Value	Output Value
Integrated Luminosity (pb^{-1}	$\mathcal{L}$	311 ± 18	311 ± 18
Observed Tags	N_{obs}	127	–
Expected Tags	N'_{exp}	57.43 ± 8.1	57.43 ± 8.06
Kin efficiency (%)	ϵ_{kin}	4.836 ± 0.962	4.837 ± 1.00
Ave Tag efficiency (%)	ϵ_{tag}	79.02 ± 4.34	79.03 ± 4.35

- $\sigma_{ttbar} = 5.8 \pm 1.1\ (stat)\ ^{+1.7}_{-1.0}\ (syst)$ pb

  $= 5.8\ ^{+2.0}_{-1.5}$ pb.

Conclusions:

- We wanted to isolate the ttbar tau+jets signal from multijet triggered data in the sample after kin sel + ≥ 1 SecVtX tag.
- We set up a Method-I analysis constructing a positive tagging matrix able to predict the amount of background tags in a given data sample with an uncertainty of 10%
- We optimized the kinematical selection using the matrix predicted background information by minimizing the expected statistical uncertainty in a xsec measurement.
- By using tag counting and kinematical distribution fits, 50% of the final sample was attributed to inclusive ttbar production
- The total systematic uncertainty was estimate to be 24%, and was found to be mainly driven by trigger simulation related systematic.
- With all these ingredients we measured a cross section of:

- $\sigma_{ttbar} = 5.8 \pm 1.1\ (stat)\ ^{+1.7}_{-1.0}\ (syst)\ pb$

  $= 5.8\ ^{+2.0}_{-1.5}\ pb.$

- We believe to have this results blessed by this summer
- A Full Status report is already scheduled for June 17th at L+J meeting.

Backup Slides

Kin sel + ≥ 1 tag Sample:

Let us see in deeper details which ttbar decay channel mainly contributes to the signal we expect:

We computed the Monte Carlo positive tags expectations for each decay channel as a function of the event jet multiplicity

- **N jets(E_T^{L5} ≥15; |η|<2.0) ≥ 3**
- $\not{E}_T / \sqrt{\sum E_T} \geq 4.00$
- **min Δφ(met,jet) ≥ 0.4 rad**

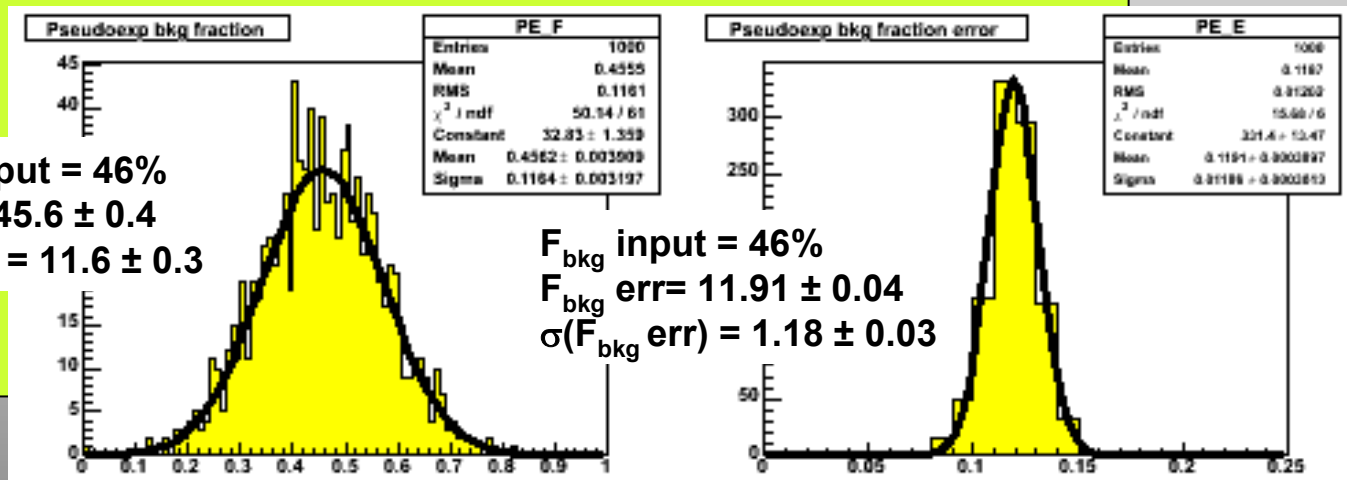
N jets	3	4	5	6	7	8
Monte Carlo contributions						
ee	0.08	0.40	0.19	0.04	0.01	–
$e\mu$	0.06	0.31	0.11	0.04	0.01	–
$\mu\mu$	0.02	0.05	0.01	0.01	–	–
$e\tau$	0.11	0.91	0.41	0.11	0.03	–
$\mu\tau$	0.05	0.28	0.15	0.05	0.01	–
$\tau\tau$	0.07	0.57	0.25	0.04	0.01	–
$e + jets$	0.65	6.56	8.70	3.32	0.74	0.20
$\mu + jets$	1.03	11.63	6.41	2.05	0.37	0.06
$\tau + jets$	1.01	11.06	11.73	4.31	0.98	0.18
$all - had$	0.01	0.09	0.14	0.15	0.04	0.03
$inclusive$	3.11	31.80	27.99	10.09	2.20	0.47

Checking the fit procedure..

Pseudo experiments:
Use the fitted bkg fraction to generate pseudoexp w/ same stats as data w/ the original shapes for signal and bkg.

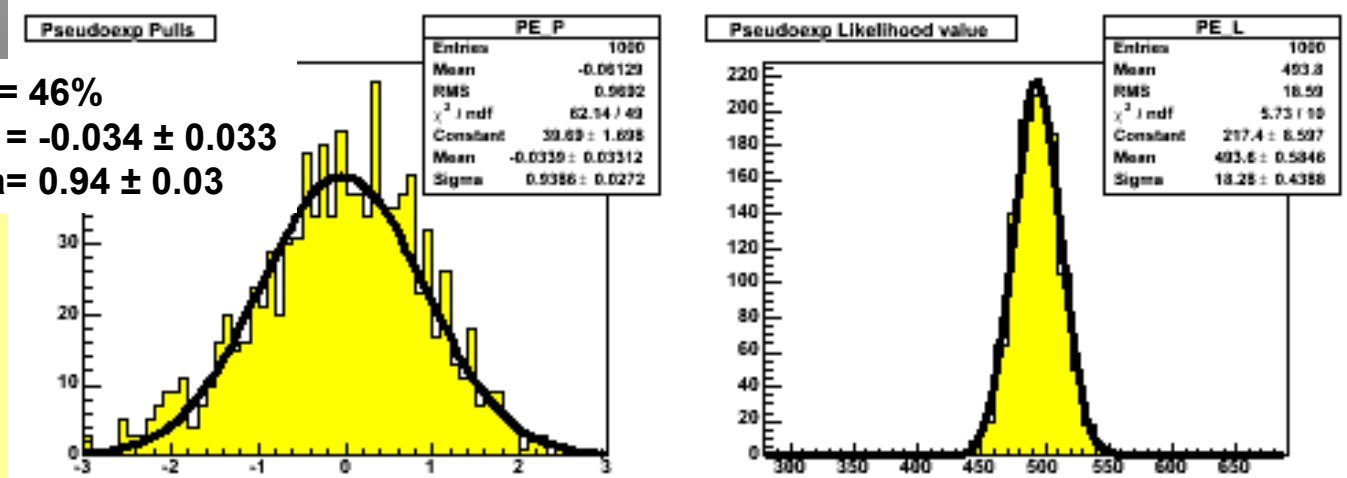
Fit them with the same fitting procedure…

…and iterate 1K times



F_{bkg} input = 46%
F_{bkg} = 45.6 ± 0.4
$\sigma(F_{bkg})$ = 11.6 ± 0.3

F_{bkg} input = 46%
F_{bkg} err= 11.91 ± 0.04
$\sigma(F_{bkg}$ err) = 1.18 ± 0.03



F_{bkg} input = 46%
Pull mean = -0.034 ± 0.033
Pull sigma= 0.94 ± 0.03

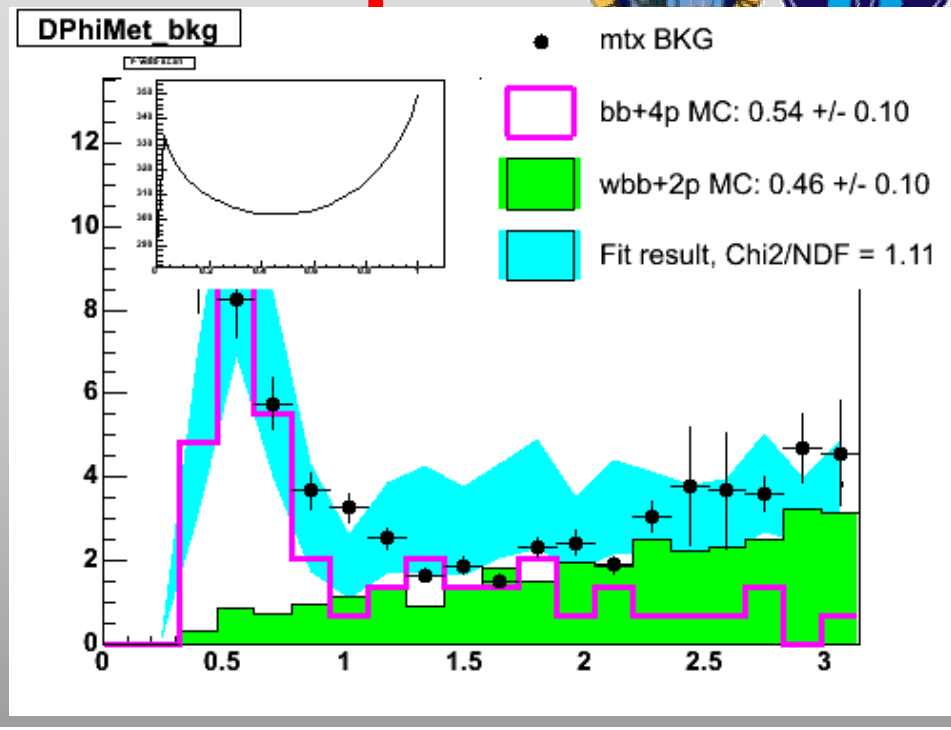
Fig refers to the Sumet fits P.E.

Fit technique does not show particular bias

Hints on bkg sample composition

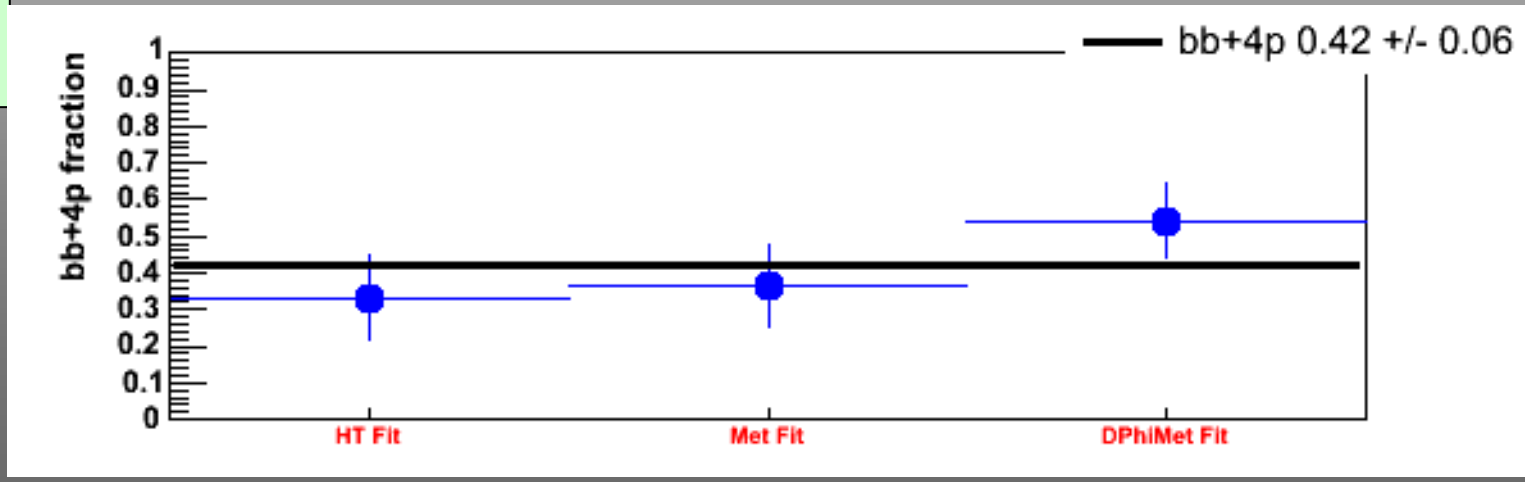
We can do more: we have the background shapes extracted from the tagging matrix information, we can fit them to the sum of two Alpgen Monte Carlo templates for the processes we expect to populate our signal region.

- Wbb+2P
- bb+4P



From bkg fits:

we found a bb fraction:



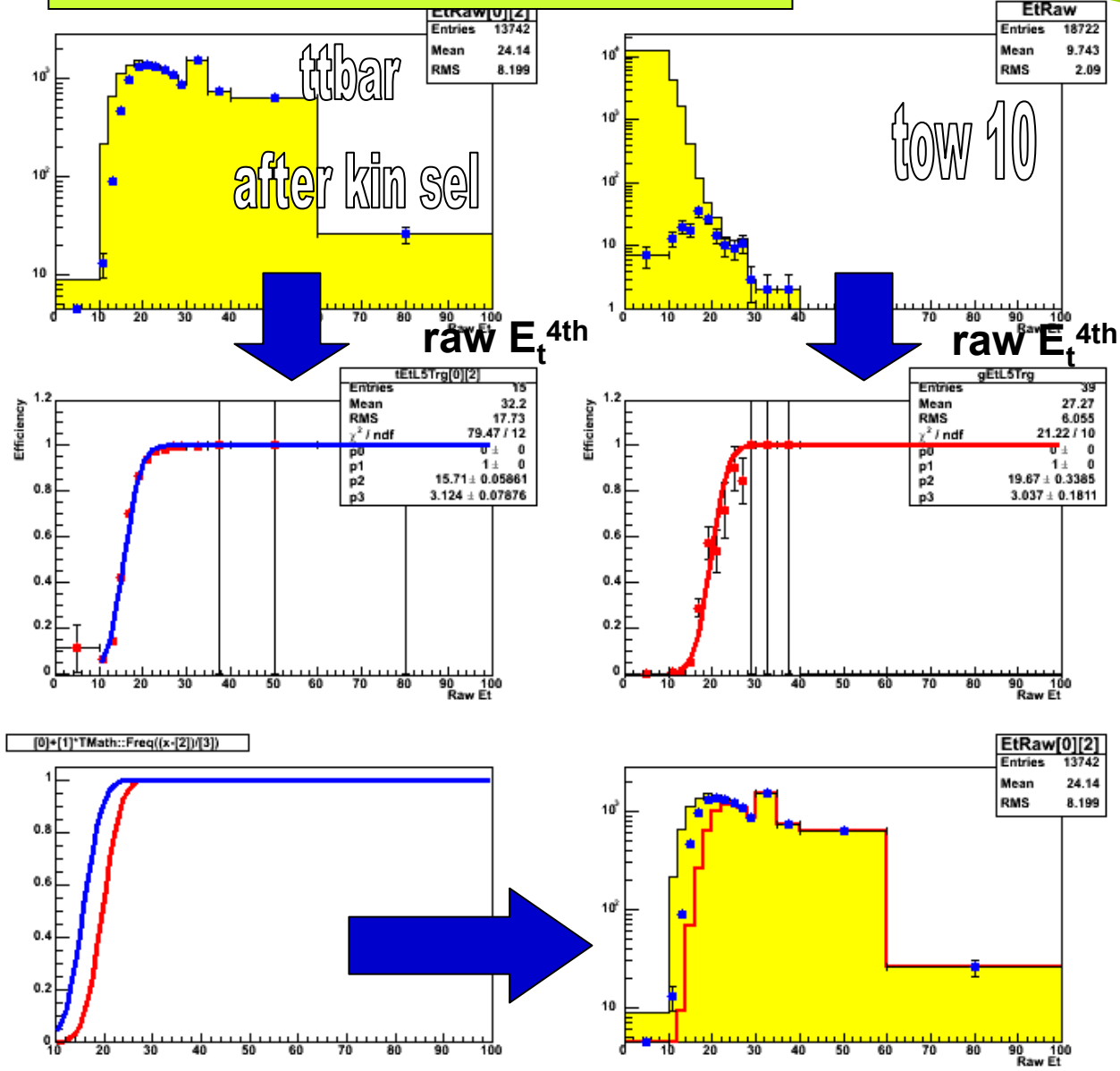
F_{bkg} ~ 42%

After kin sel + ≥1 tag data: 50% top + ~21% bb + ~ 29% Wbb

A bit more on systematics: Trigger

Trigger efficiency on signal events is determined using TRIGSIM++.

CDF-7473

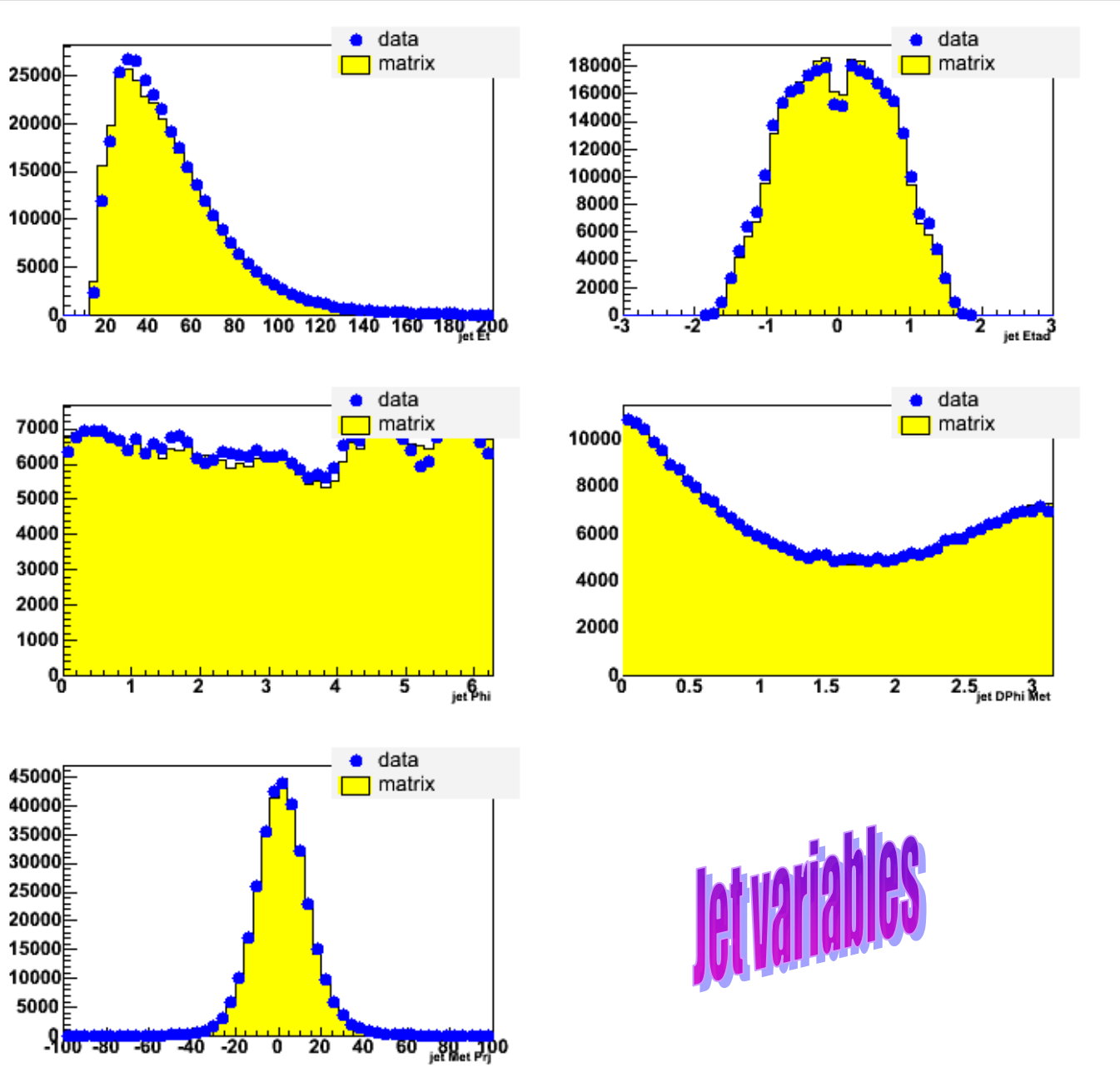


Need to evaluate related systematic:

comparing trigger turn-on curve (as a function of some offline variable) as returned by the simulation and as measured from Tower 10 data (same L1 as TOP_MULTI_JET).

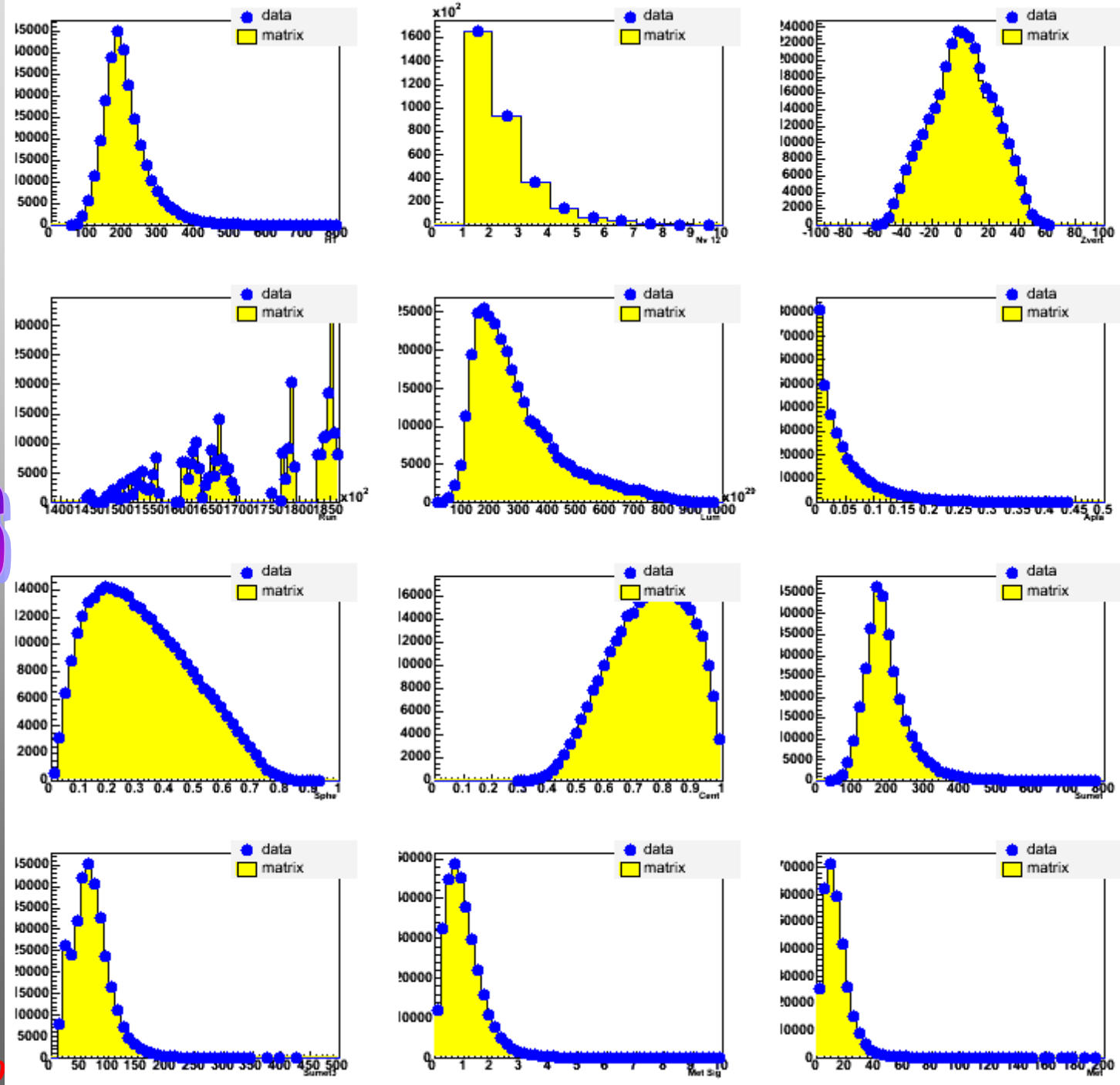
The mismatch between turn-on curve allows to quantify the systematic effect.

Note: the 4th offline jet is matched with the 4th L2 cluster within R=0.4 in order to preserve energy hierarchy



Tagging matrix based kinematical distribution compared to data ones before kinematical selection

Tagging matrix based kinematical distribution compared to data ones before kinematical selection



Evt variables