Study of the heavy flavor content of jets produced in association with *W* bosons at $\sqrt{s} = 1.8$ TeV

F. Ptohos/INFN-Frascati

CTEQ –ME/MC Tuning Working Group
Fermilab
April 29-30 2003

Outline of the talk

- Review of the tools used for the selection of tt events and the measurement of $\sigma_{t\bar{t}}$
  - Tagging Algorithms
  - Cross-checks on independent data samples

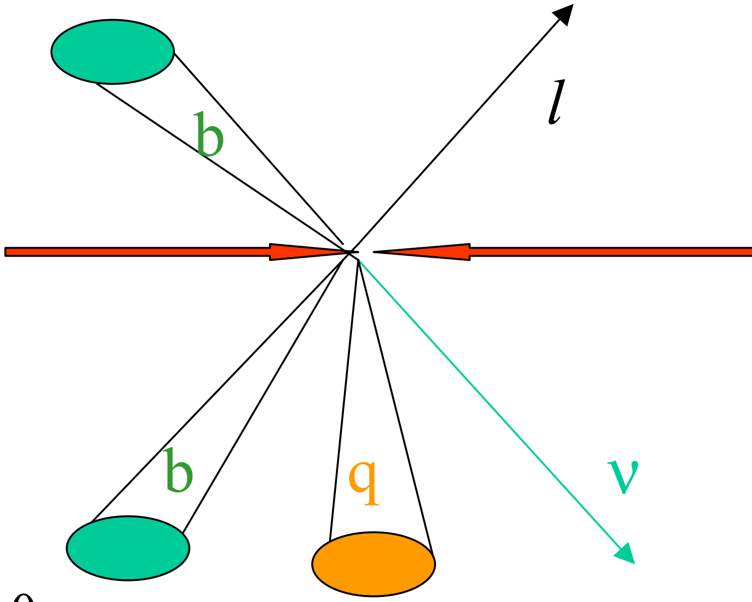
Phys. Rev. D64:032002, 2001

- Assuming the Standard Model (i.e. using $\sigma_{t\bar{t}} = 5.1 \pm 1.54\, pb$), compare rates of various combinations of tagged W + jets events to Standard Model expectations

- Describe the observation and characteristics of anomalous events found in the previous study

Phys. Rev. D65, 052007 (2002)

Data sample : lepton + jet events



- Kinematic cuts:
  - One high E_T lepton (e,μ)
    - E_T >20 GeV
    - Central (| η | <1)
    - Isolated
  - $\not{E}_T > 20$ GeV
  - at least 1 jet with E_T >15 GeV and |η| <2.0
- b-identification : Efficiency
  - SECondary VerTeX (SECVTX) 50%
  - Jet-ProBability (JPB) 50%
  - semileptonic decay (**S**oft**L**epton**T**agging) 20%

Composition of W+≥1 jet sample

- The following processes have been considered:
  - non-W background (estimated from data)
  - mistags in the W+jet sample (estimated from data)
  - $Z \rightarrow \tau^+\tau^-$, $t\bar{t}$, single top production, diboson production (estimated from simulation)
  - Wc production (estimated from simulation)
  - W +bb and W +cc production (estimated from simulation which is calibrated on data)
  - Z + heavy flavor production (estimated from simulation and data)
- **Mistags**

  Tags in jets without h.f.content. Estimated using parametrized probability functions derived in jet data. Error on mistag estimate 10%
- **$t\bar{t}$ production**

  Use Pythia and the theoretical estimate of $\sigma_{t\bar{t}} = 5.1\text{pb}$ with 15% error
- **Single Top**

  Use Herwig for W-g fusion and NLO $\sigma = 1.5 \pm 0.4$ pb

  Use Pythia for $qq \rightarrow W^* \rightarrow tb$ and NLO $\sigma = 0.74 \pm 0.05$ pb
- **Wc**

  $gs \rightarrow Wc$ and $gd \rightarrow Wc$ contributions are evaluated in Herwig. Uncertainty dominated by the uncertainty in the strange sea content of the proton

Composition of W+≥1 jet sample

- W/Z + g (g→bb or g→cc)

  The major source of heavy flavor production in the W+jets tagged sample

  - ✖ LO matrix element calculations have large uncertainty (~ 40%)
  - ➢ Use number of W events in the data before tagging in each jet bin and multiply by the fraction of W+bb and W+cc events measured in the simulation.
  - ✖ No Monte Carlo describes the full phase space of the events
  - ➢ VECBOS:

    Matrix Element MC

    Requires cut-off values for parton P_T and ΔR separation to regulate singularities

    Heavy flavor partons are well separated resulting to separate jets

    Interface with parton shower MC for initial final state evolution
  - ➢ HERWIG:

    Parton shower MC

    g → heavy flavor pair most often inside one jet

    - ✖ Gluon splitting uncertainty between 25% and 40% ⟹ **Requires calibrations**
  - ➢ **Use combination of the two MC to obtain the W+h.f. event fractions but calibrate these fractions using data**

Calibration of the heavy flavor content in the simulation

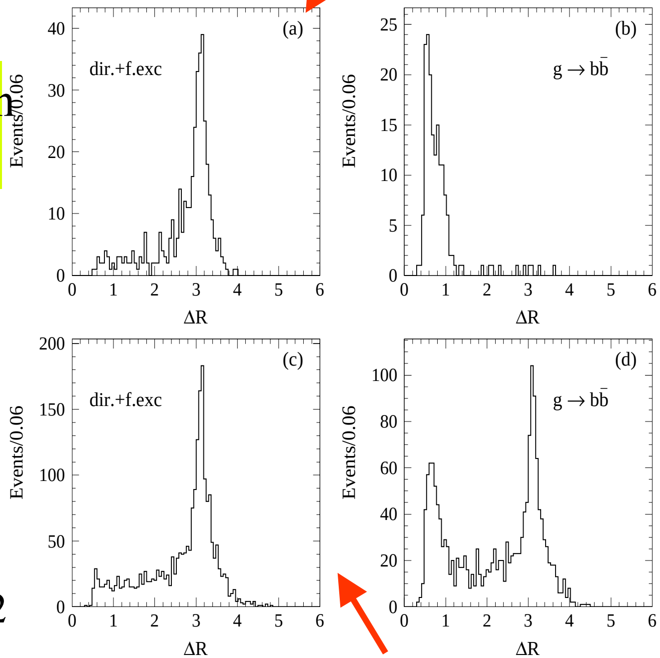
- JET50 and JET100 data sets are compared to Herwig simulations (option 1500, generic 2→2 hard scattering, use MRS(G) PDFs)
- $b\bar{b}$ and $c\bar{c}$ production are generated through processes of order α_s^2 such as $q\bar{q} \rightarrow b\bar{b}$
- Processes of order α_s^3 are implemented through flavor excitation diagrams, such as gb→gb, or gluon splitting, in which the process gg→gg is followed by $g \rightarrow b\bar{b}$
  - Gluon splitting comparable to the other production mechanisms

	direct production + flavor excitation		gluon splitting		
Sample	b-jets	c-jets	$g \rightarrow b\bar{b}$	$g \rightarrow c\bar{c}$	Total
JET 50	2.14×10^{-2}	3.04×10^{-2}	1.67×10^{-2}	3.79×10^{-2}	10.64×10^{-2}
JET 100	2.15×10^{-2}	2.89×10^{-2}	2.58×10^{-2}	5.73×10^{-2}	13.35×10^{-2}

Calibration of the heavy flavor content in the simulation

- Use kinematical differences to distinguish gluon splitting from other h.f. production mechanisms.
- Use ΔR (in $\eta-\phi$ space) to separate direct production and flavor excitation from gluon splitting
- Use large jet multiplicity since relative rate of gluon splitting increases with jet multiplicity
- Fit simulation to data for
  - tagged jets in events with one taggable jet
  - tagged jets in events with one taggable jet and three or more jets (rich in gluon splitting)
  - events with two tagged jets
  - tagged events with a companion jet within $\Delta R<1.2$
  - double tagged events with a companion jet with $\Delta R < 1.2$
- Use SECVTX and JPB tags in order to discriminate the flavor type ($\varepsilon_c^{JPB} \sim 2\times\varepsilon_c^{SECVTX}$)

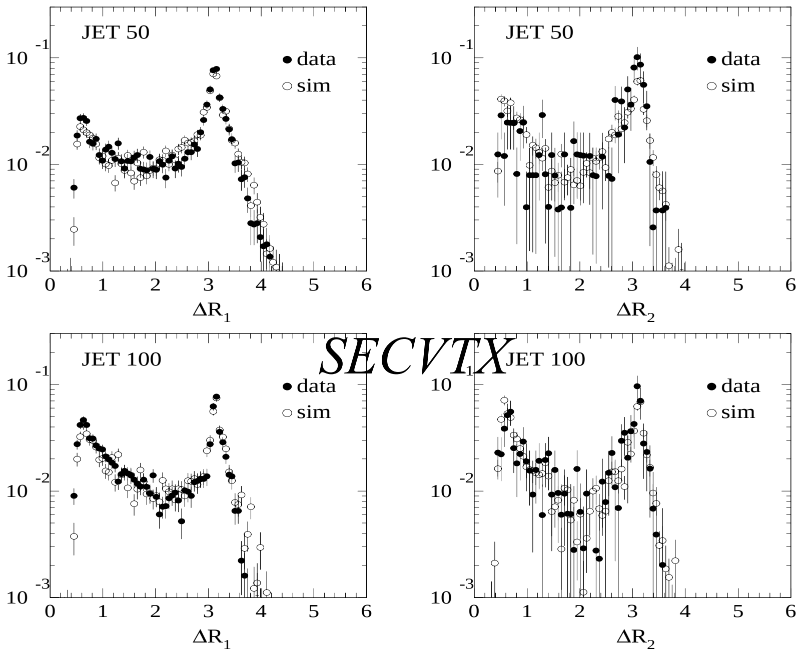
Contributions and distances between 2 tagged SECVTX jets in JET50 MC



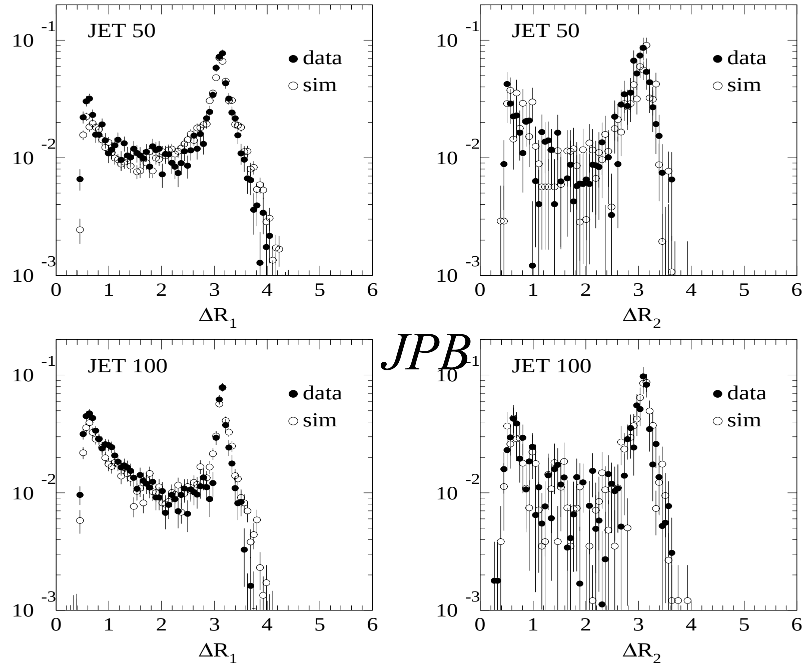
Contributions and distances between 1 tagged SECVTX jet and the closest jet with E_T> 10 GeV in JET50 MC

Calibration of the heavy flavor content in the simulation

- Fit results:
  - b,c direct production+flavor excitation σ weight 1.11 ± 0.16
  - gluon splitting σ weight 1.36 ± 0.22



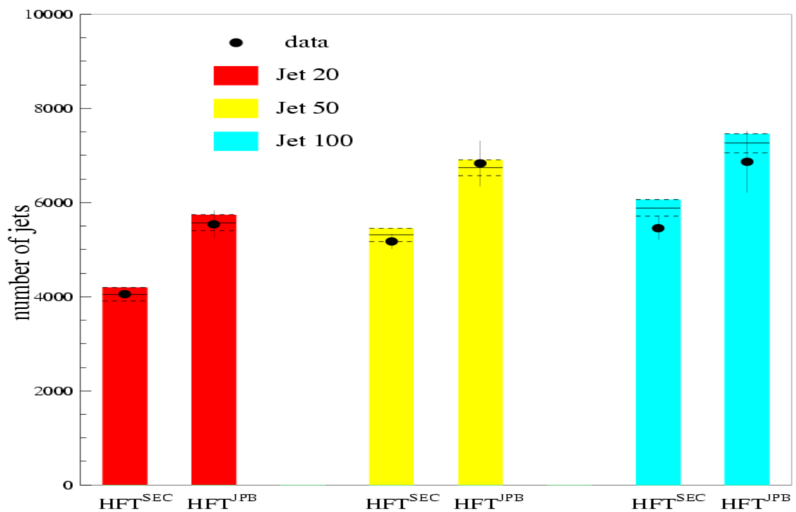
SECVTX

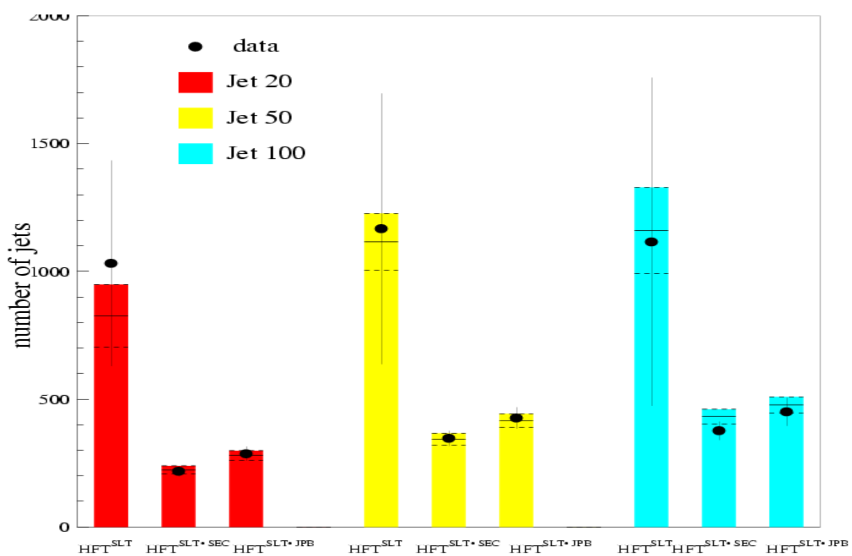


JPB

The fraction of $g \to b\bar{b}$ and $g \to c\bar{c}$ in the Herwig simulation needs to be increased by (39±19)% and (35±36)% respectively

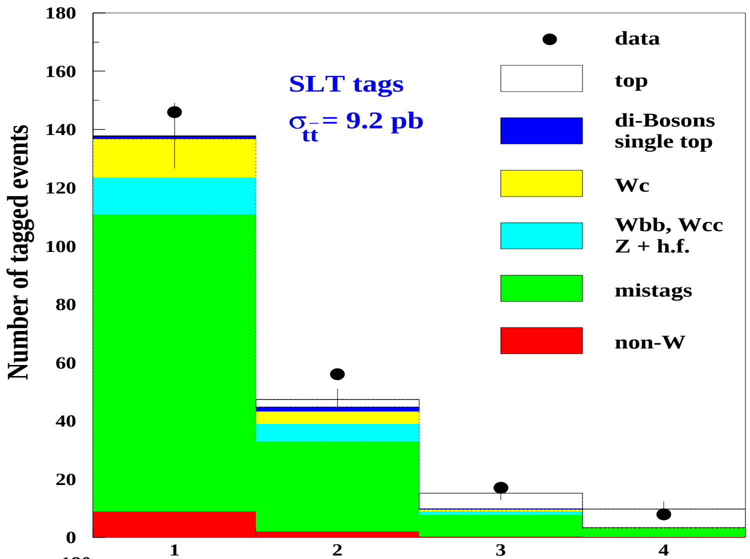
Cross check of the calibration

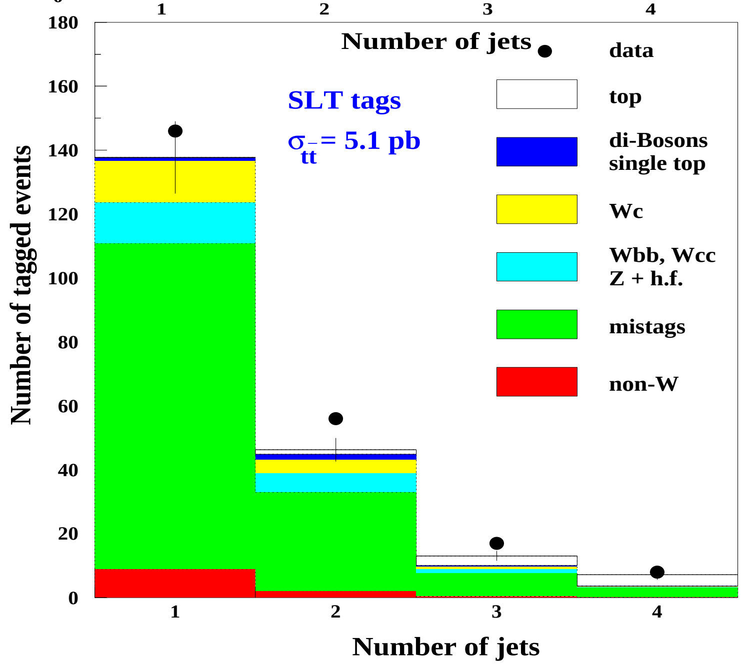


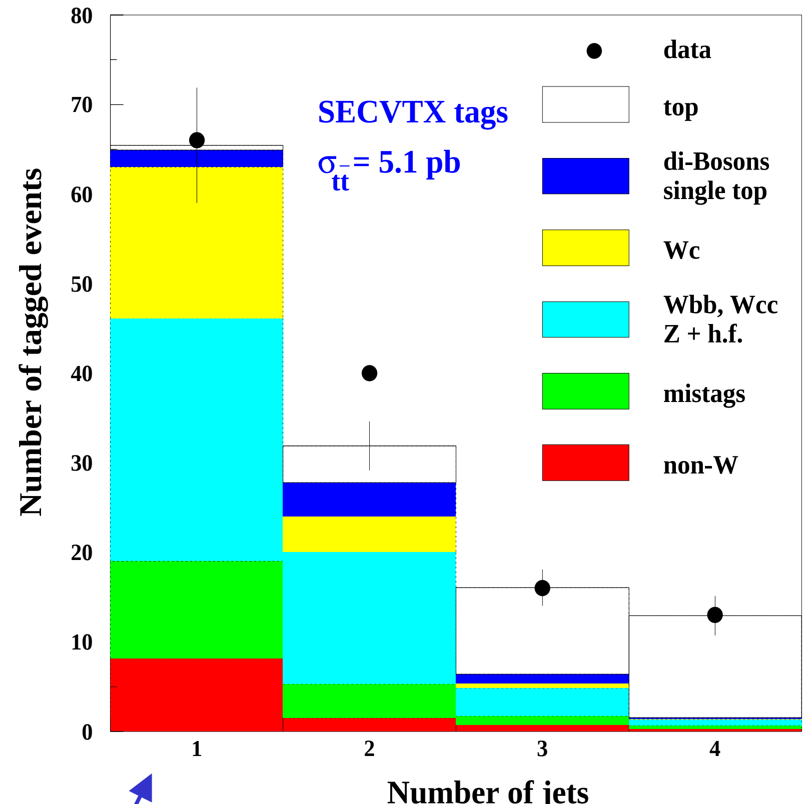


- HERWIG tuning:
  (Flavor excitation + Direct production) × (1.1±0.16)
  $g \to \bar{b}b \quad \times \quad (1.4 \pm 0.19)$
  $g \to \bar{c}c \quad \times \quad (1.35 \pm 0.36)$
- Corrections are of the same size as those measured by SLC and LEP
- Good description of JET 20 which was not used in the tuning
- SLT tags (not used in the fitting) are well described
- Supertag finding efficiencies are lower in the data than in simulation by (85±5)% independently of energy and flavor type

W + jets after tagging



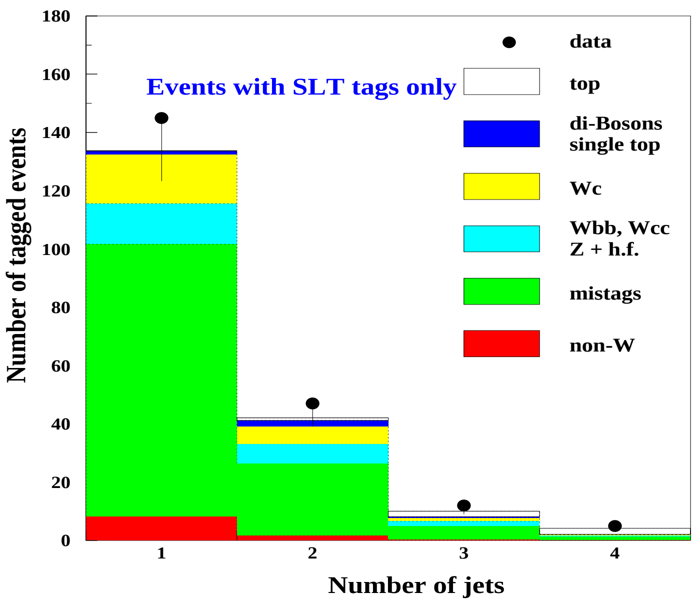




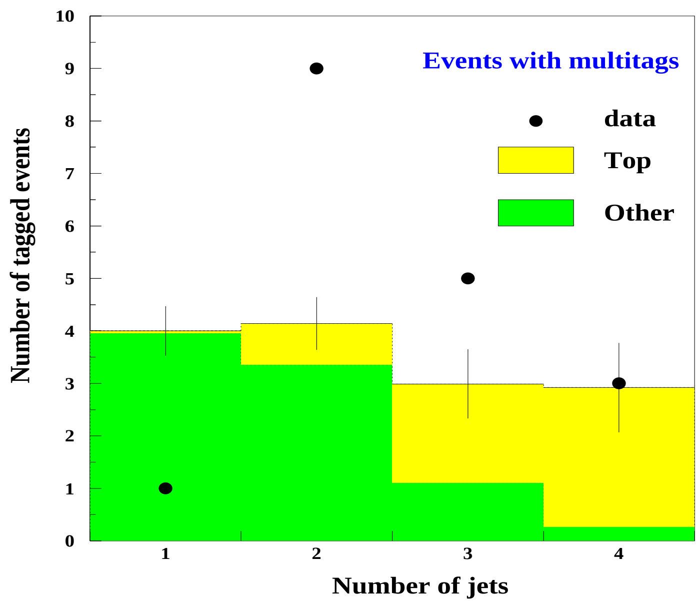
P~50%

SECVTX and SLT correlations

P~50%



P~3%

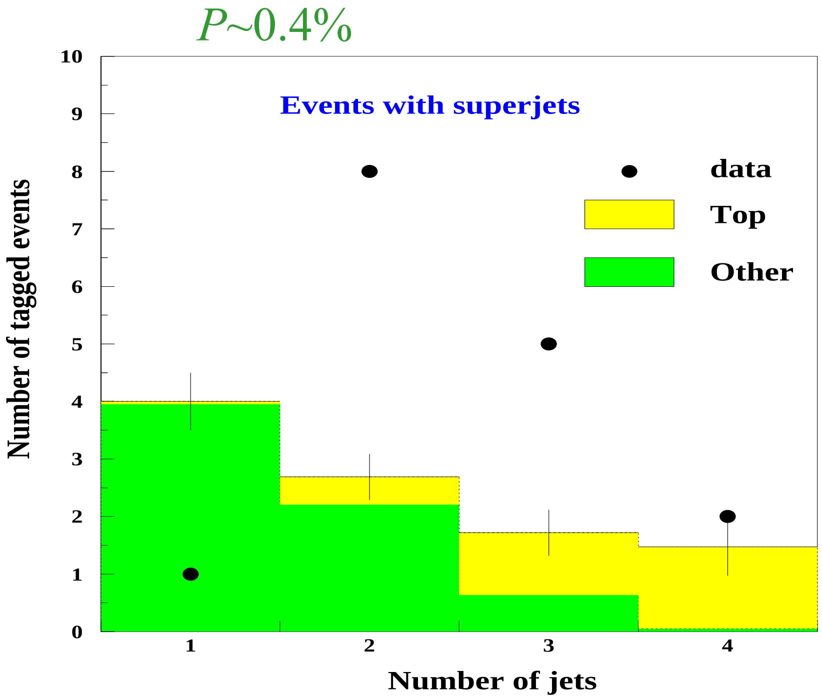


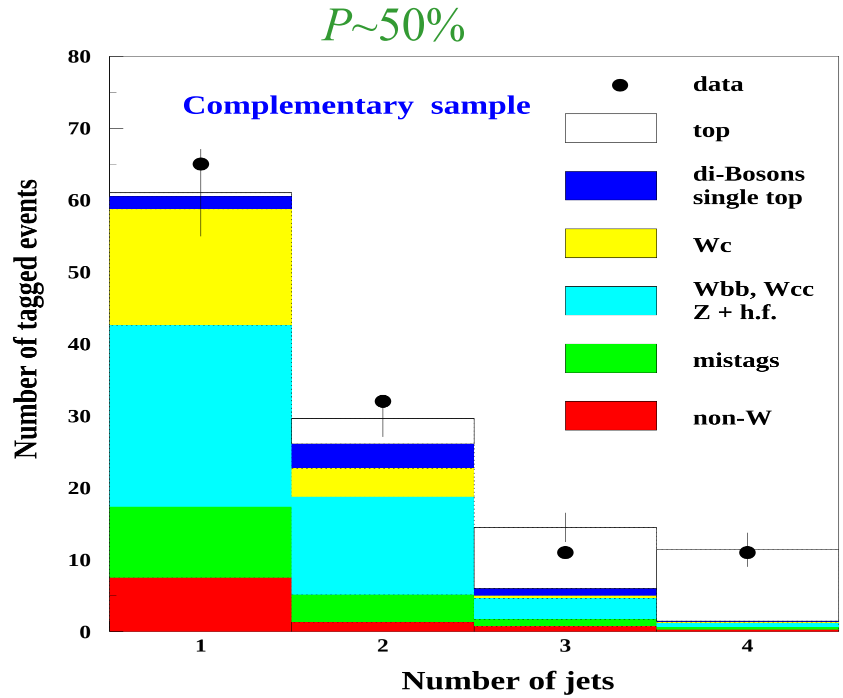
- The numbers of events with SLT and no SECVTX tags are consistent with prediction. Most of these tags (~75%) are fakes.
- Clean sample: The composition of SECVTX tags is 70% b-jets and 20% c-jets
- The numbers of observed and predicted events with both SLT and SECVTX tags are not very consistent

Check the semileptonic BR of heavy quark jets by studying the fraction of SECVTX tagged jets which contain also a SLT tag (supertag)

  - Expect 7% of SECVTX tagged jets to contain an additional SLT tag.

Events tagged by SECVTX w/o supertags





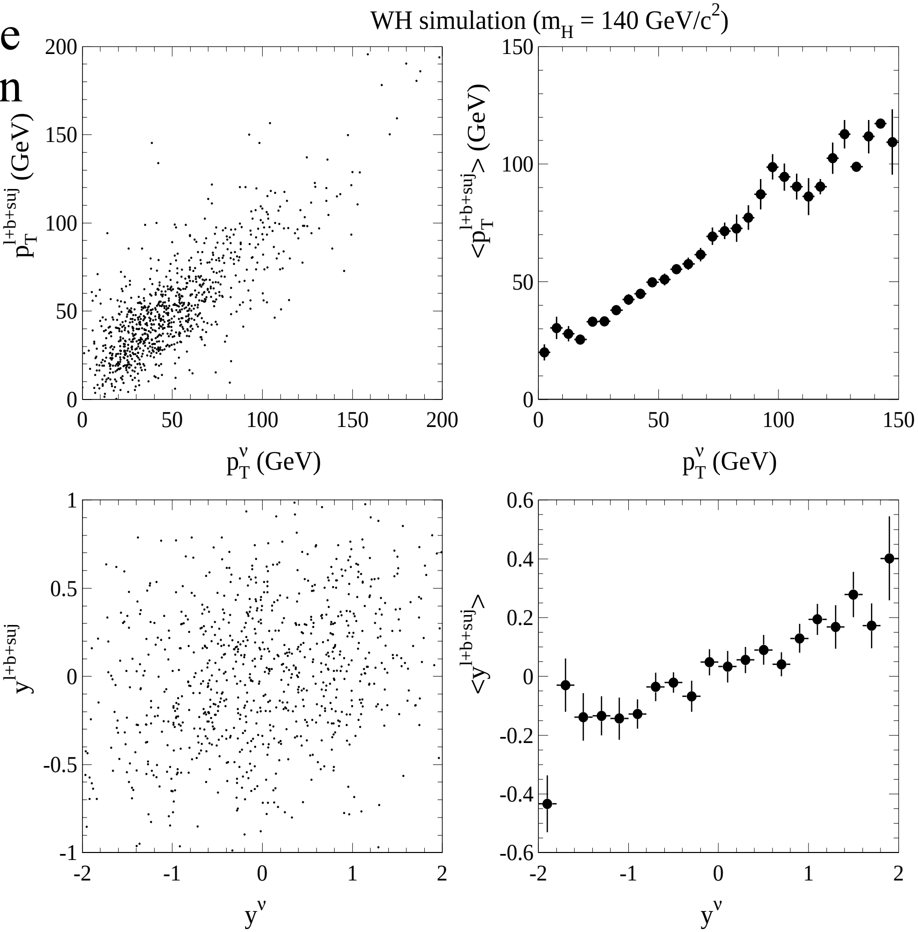
- In the superjet sample the probability of consistency with SM in all jet bins is 0.4%
- In the n_{jet} = 2,3 bins the supertag sample has 13 events when 4.4 ± 0.6 are expected. The *a posteriori* probability of observing 13 or more events is $P=10^{-3}$
- We define a complementary/control sample as the SECVTX tagged events which pass all cuts defining superjets + SLT candidate track inside the tagged jet.
- The complementary sample has 43 events and 43.6 ± 3.3 are expected from S.M.

Study of the kinematics

- If the 13 events are a statistical fluctuation, the kinematics of this sample will be consistent with the S.M. simulation and the complementary sample
- We chose two sets of 9 variables to look for differences

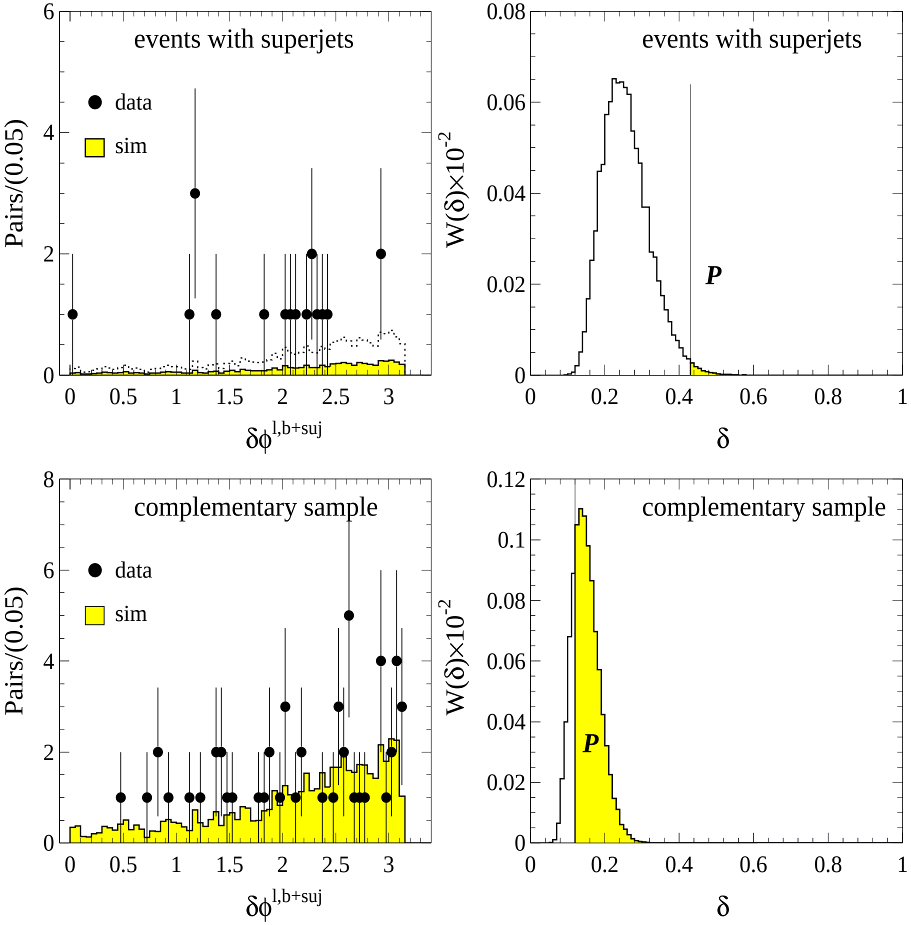
First set

- Studies $d^2\sigma / (dp_T\, d\eta)$ for every different object in the final state (8 var.)
- Replace $\not{E}$ with the system l+suj+b
- Add the angle between the lepton and *W* (check if events are consistent with the production and decay of *W* bosons)
- This set of 9 variables fully describes the kinematics of the final state with modest correlations
- Compare distributions in the data and the SM simulation using a K-S test. Use the Kuiper's definition.



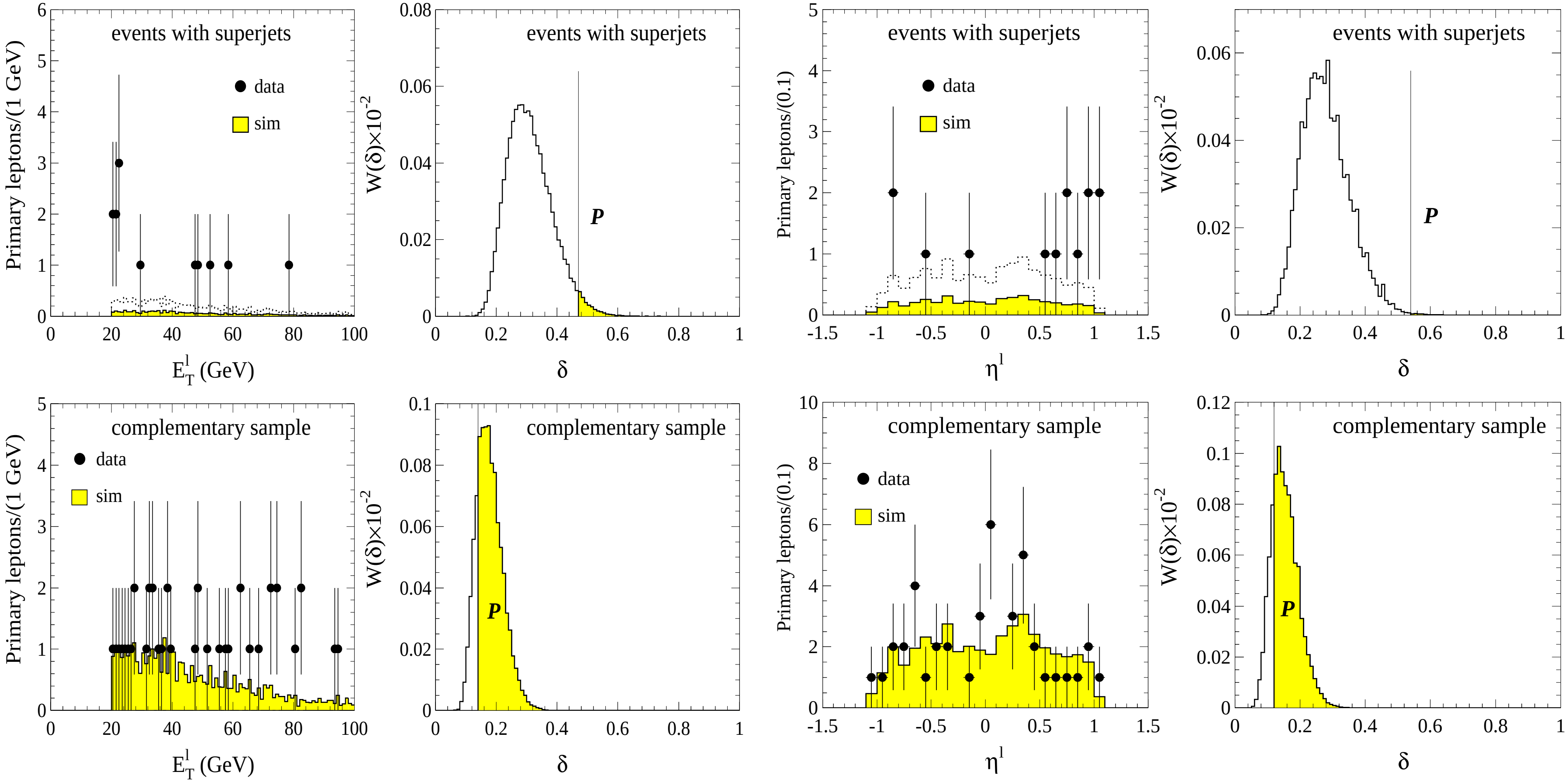
Azimuthal angle, $\delta\phi^{l,b+suj}$

- The probability distribution of the K-S distance, δ, is determined with pseudo-experiments
- In each pseudo-experiment, we construct SM “running” templates which account for Poisson fluctuations and Gaussian uncertainties of each SM process
- From each “running” template we randomly generate a distribution with the same number of entries of the data
- We then compare this distribution to the nominal SM template and derive δ



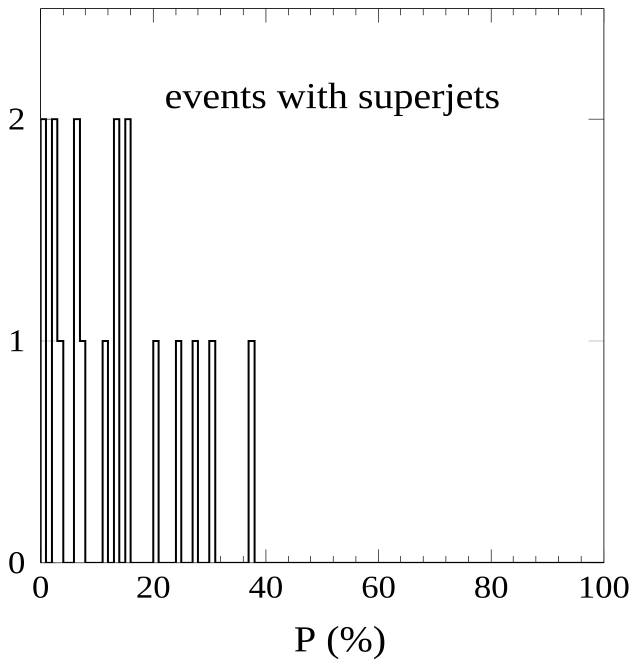
Primary lepton

E_T η

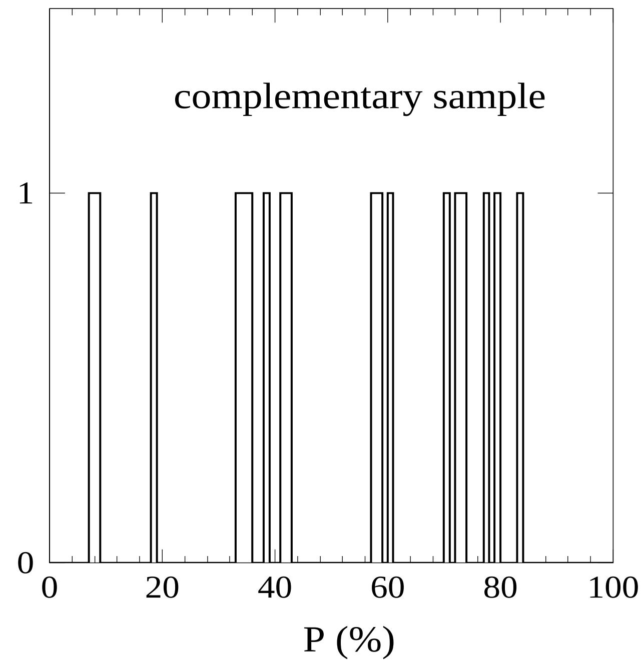


18 Kinematical variables

< P>=0.13 RMS=0.11



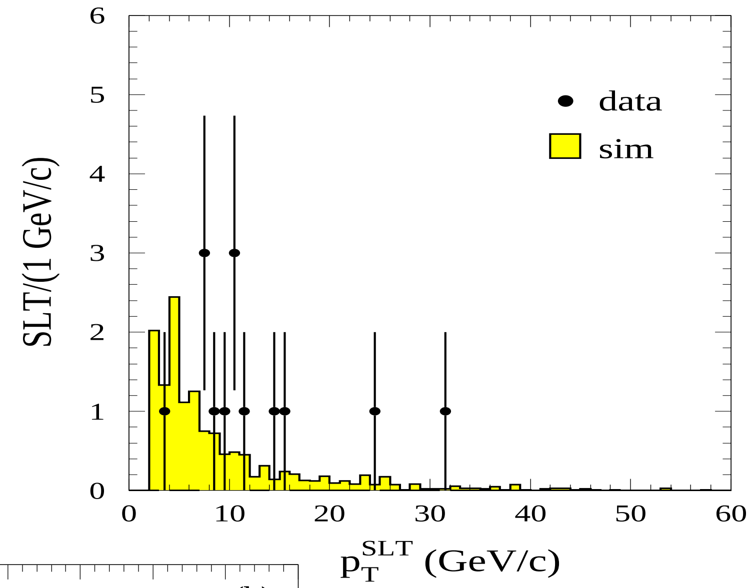
< P>=0.50 RMS=0.24

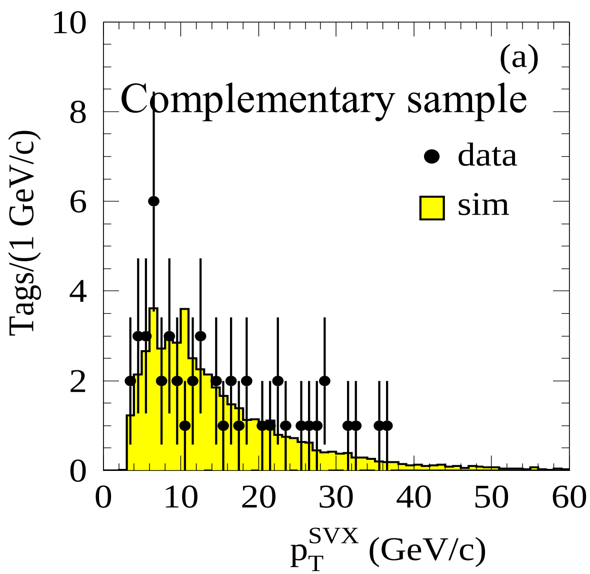


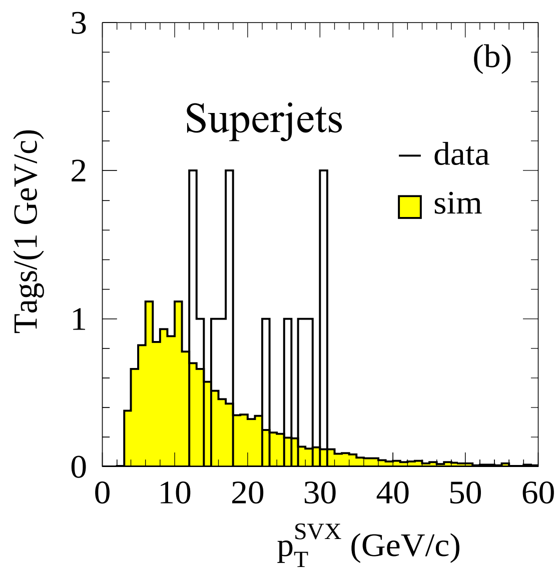
- The complementary sample probabilities are flatly distributed as expected we expected from simulation.
- All the probabilities of the superjet events have low values

Additional curiosities

- Compare to a SM simulation, in which the superjet transverse momentum distribution in each SM process has been sculpted to reproduce the data
- the usual K-S test yields P= 0.1%
- A check for inadequate modeling of the hadronization process is provided by the p_T^{SVX} tag



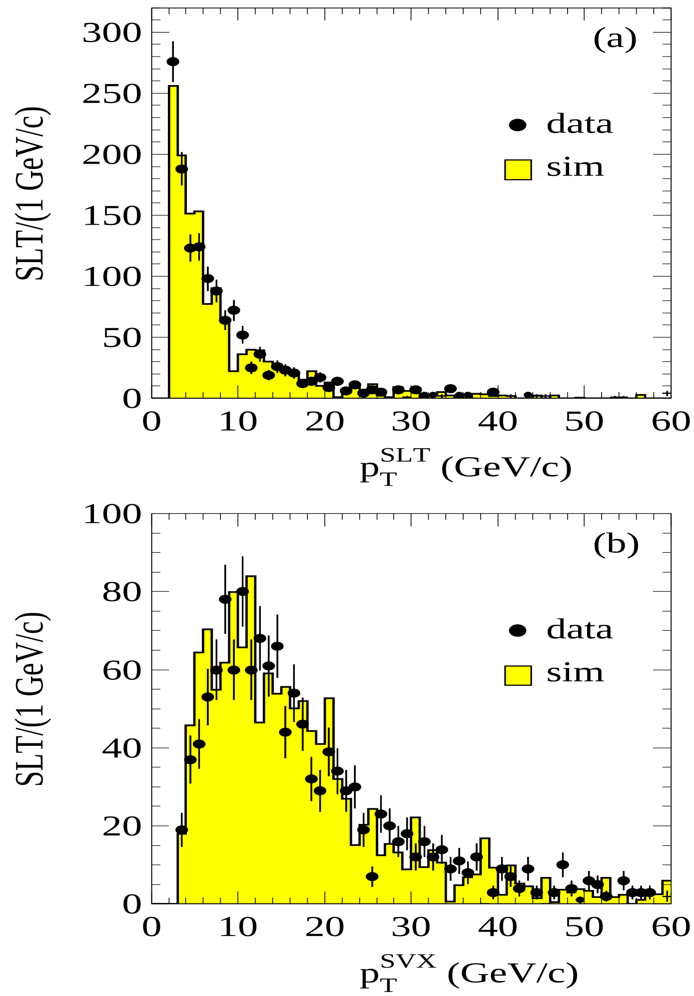




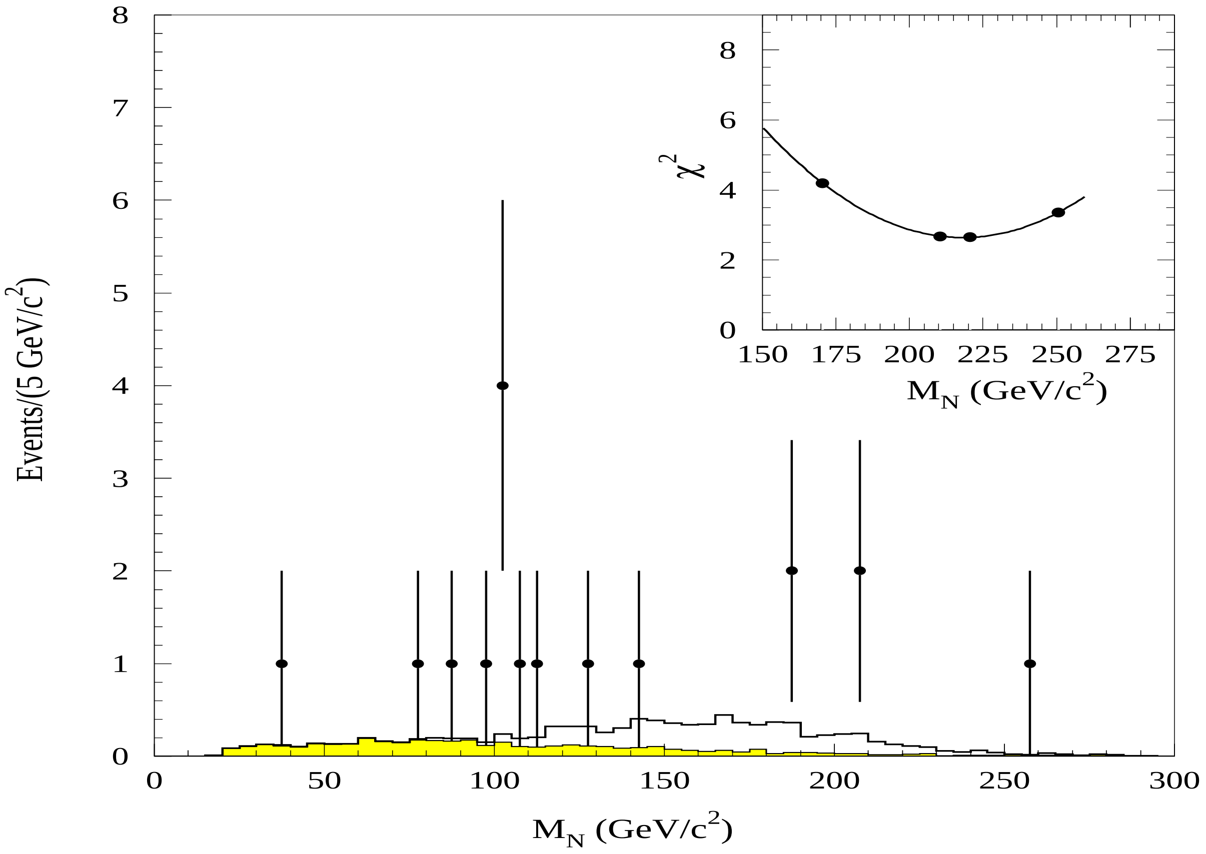
Check of the fragmentation with generic-jet data

- 550,000 generic-jet events in the data and in the Herwig simulation (JET20, JET50, and JET100).
  - 1324 supertags in the data
  - 1342 simulated supertags

- b-jets and c-jets are correctly modeled by the simulation

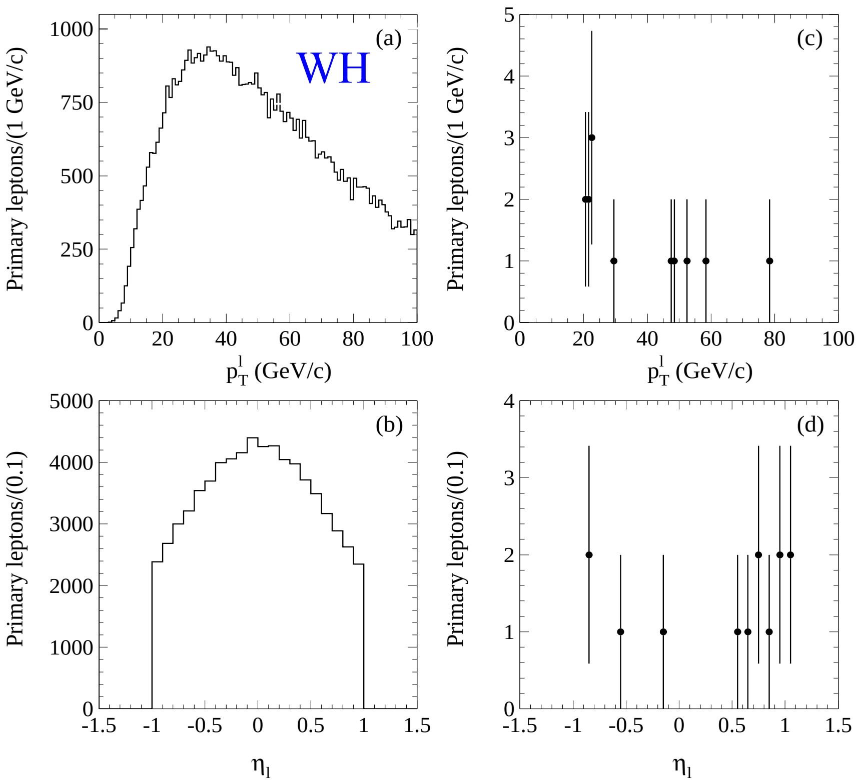


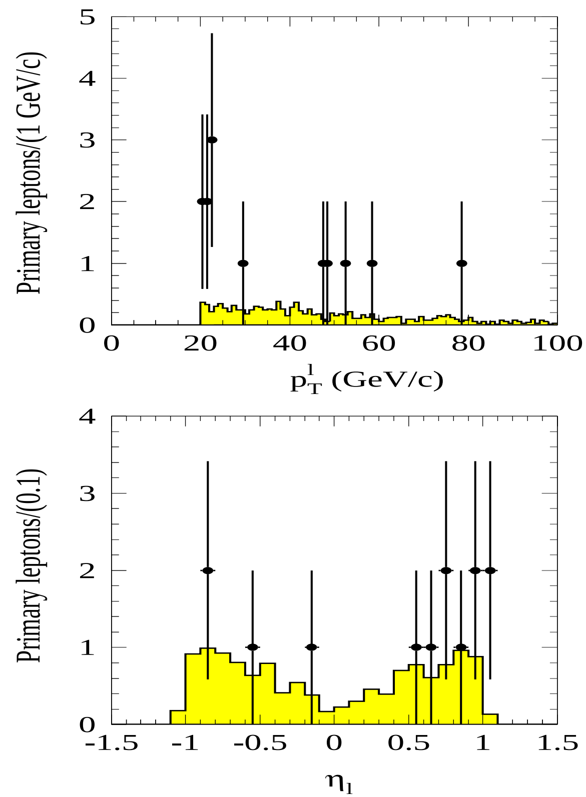
Invariant mass of the two highest E_T jets



l,$\not{E}_T$ from W decay

W→lν decay



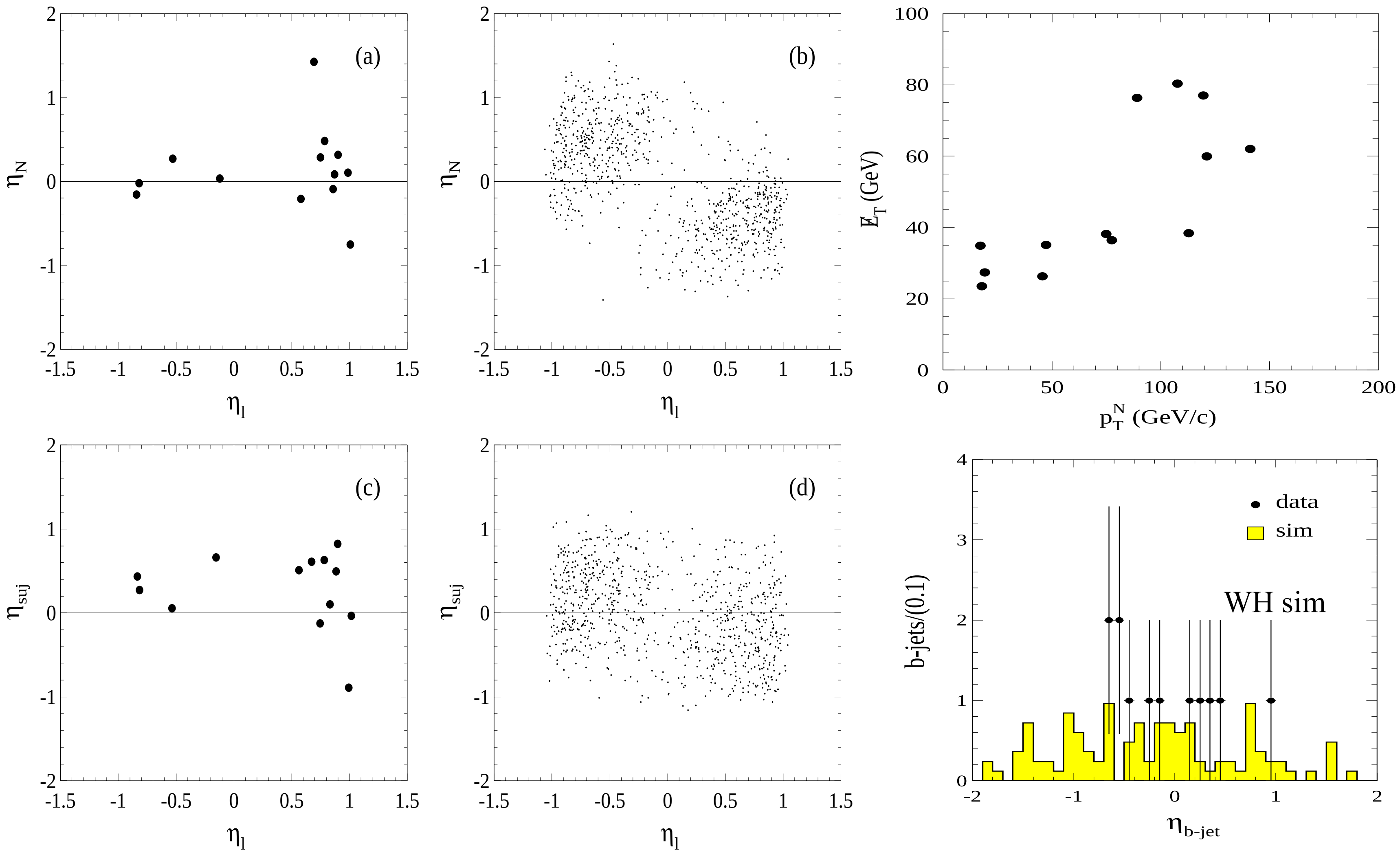


- weight the lepton polar angle distribution as $z^4 = \cos^4 q$ in the rest frame of the initial partons

Other distributions

Superjet data

WH simulation



η– φ correlations of tracks