Extracting backgrounds to SUSY searches from LHC data

Vadym Zhuravlov

(Max-Planck-Institut für Physik)

on behalf of the ATLAS and CMS collaborations

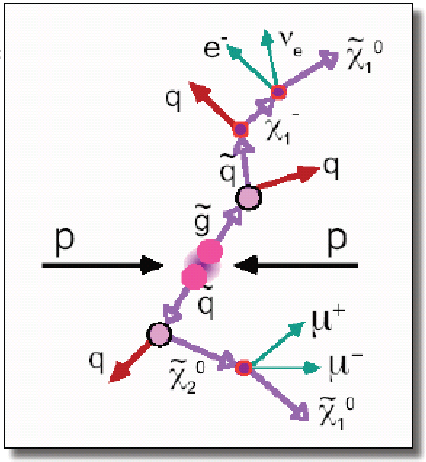
Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

Outline

- Multi-jet + lepton veto — no lepton mode
  - $Z \to \nu\nu$
    - from $Z \to ee$ and $Z \to \mu\mu$
    - from $W \to \mu\nu$
    - from γ+jets
  - $tt \to bb\ qq\ l\nu$ with not-identified lepton
    - replacement of lepton
- Multi-jet + one lepton — 1 lepton mode
  - top-anti-top and W+ jets

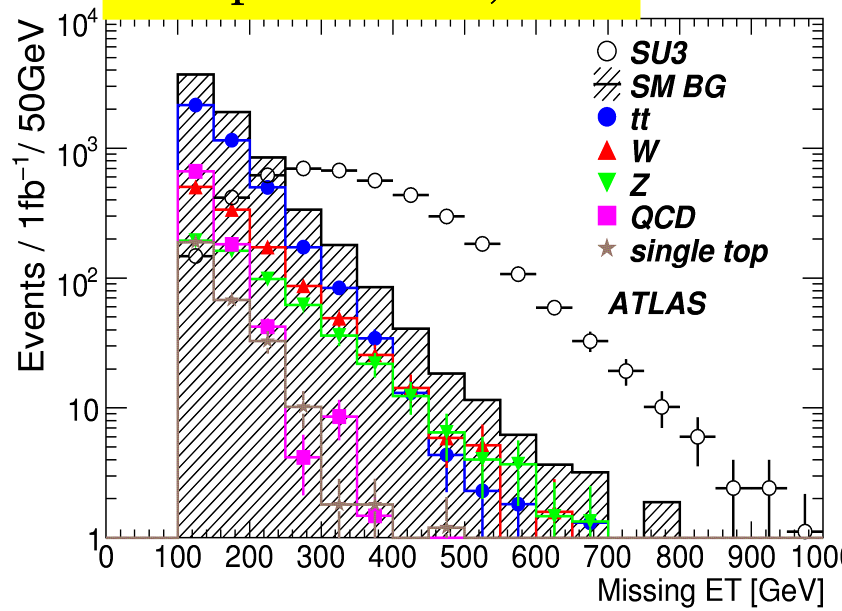
Introduction: SUSY searches

- First LHC data: search for events with multi-jets and large E_T^{miss}



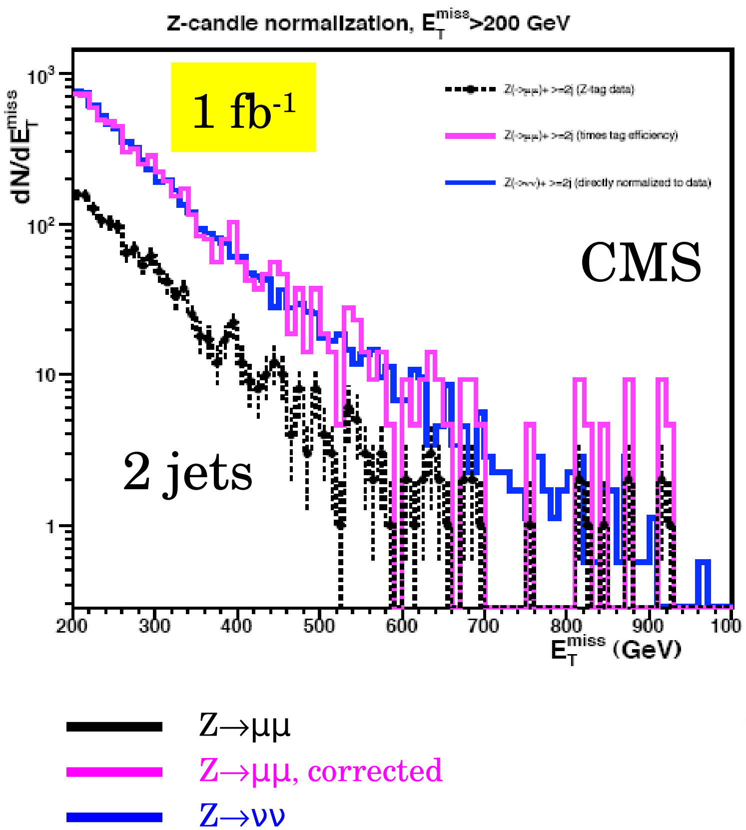
- Example: R-parity conserving mSUGRA
  - cascade decays, 0,1,2, ... leptons
  - pair of LSP: no mass peak
- Counting of number of events in "signal region" and compare to SM prediction
- Monte Carlo SM prediction has large uncertainty:
  - p.d.f. for protons @ 14 TeV
  - cross-sections
  - detector response

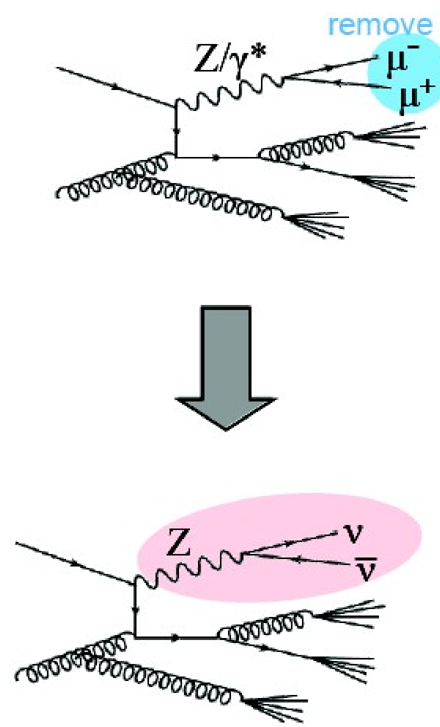
no-lepton search, 1 fb^{-1}

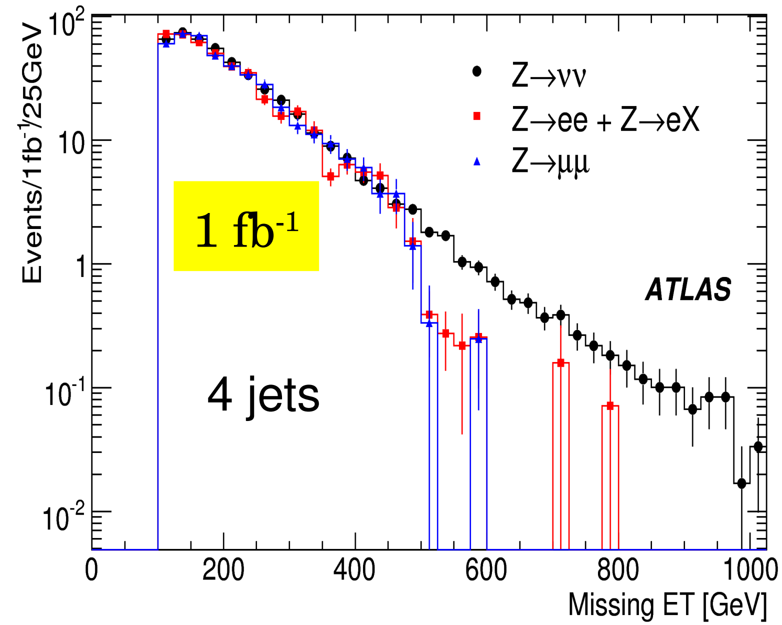


SU3: ATLAS benchmark point with SUSY mass scale ~600 GeV

$Z \rightarrow \nu\nu$ from $Z \rightarrow ll$







- Low statistics at high E_T^{miss} N(μμ)/N(νν)~1/12
  - Use MC shape
  - Lower jet multiplicity
  - $Z \rightarrow e + X$

for 1 fb^{-1}:

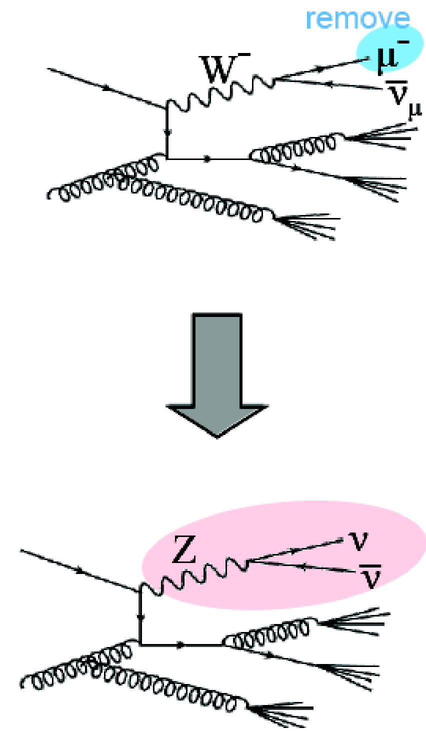
stat. uncert. ~ 13%

syst. uncert. ~ 8%

$Z \to \nu\nu$ from $W \to \mu\nu$

no lepton mode

- Control sample selection: no-lepton search + 1 muon $N(\mu\nu)/N(\nu\nu) \sim 1.5$
  - backgrounds:
    - top-anti-top: data-driven technique (ratio of b-tagged and untagged events)
    - QCD: anti-isolated muons

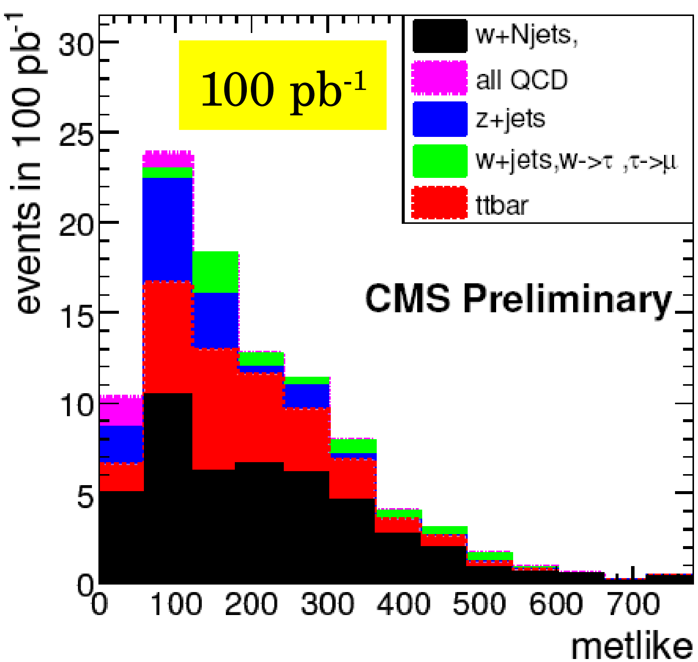


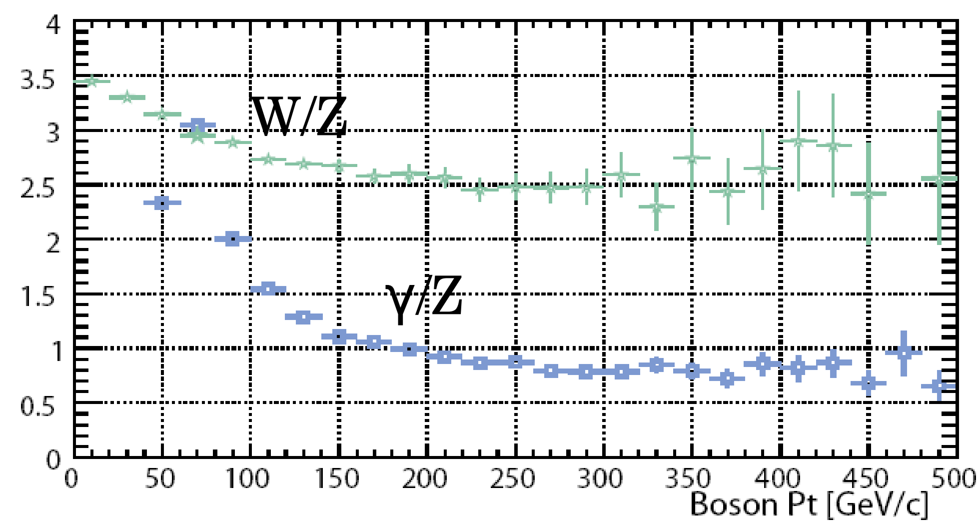
for 100 pb^{-1}:

stat. uncert. ~ 29%

syst. uncert. ~ 23%

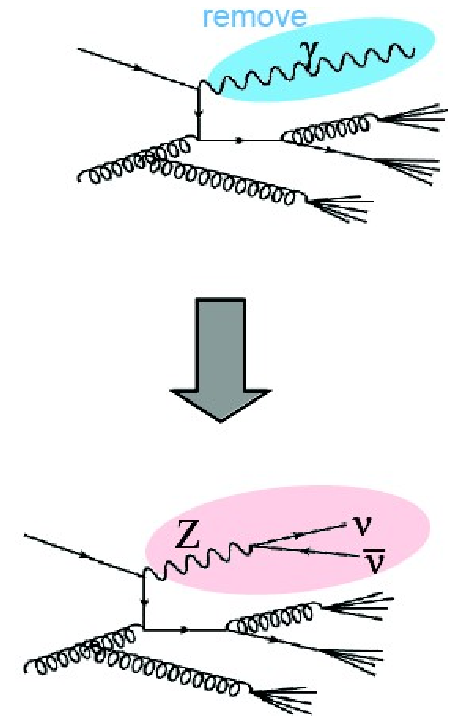
theory ~ 8%



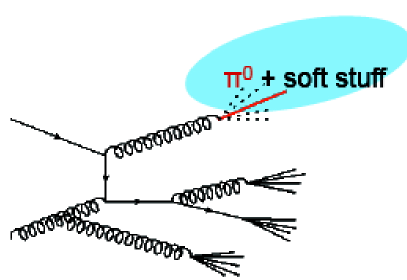


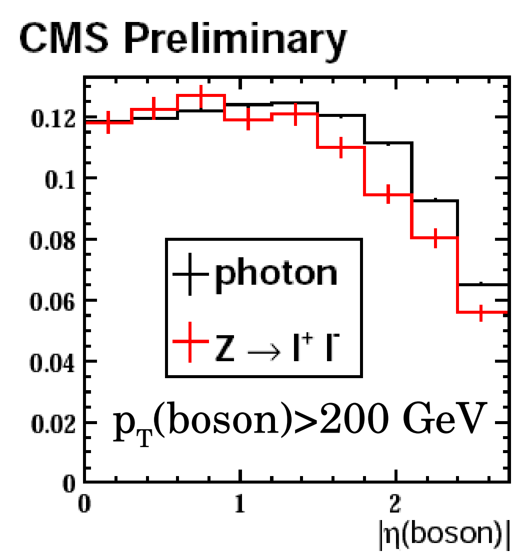
$Z \to \nu\nu$ from γ+jets

no lepton mode



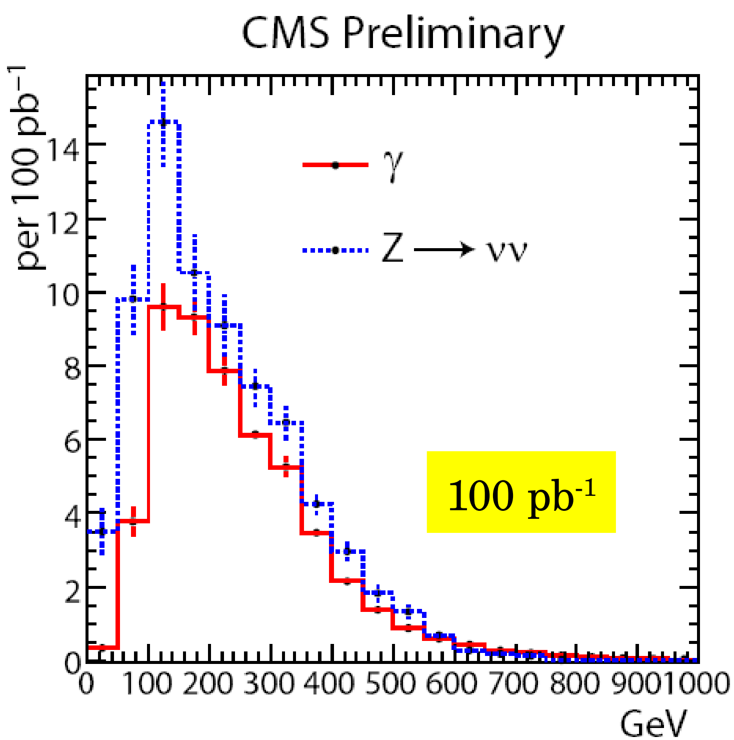
- Photon selection:
  - $E_T(\gamma) > 100$ GeV, electron veto
  - isolation: tracker isolation+relative calorimetric isolation:
    - factor 100 in rejection of QCD
    - 90% efficiency to prompt photons
    - data-driven estimation of efficiency
- Backgrounds (<3% each):
  - QCD secondary photons
  - electron faking photon
  - data-driven estimation (backup slide)
- N(γ+jets)/N(νν)~4

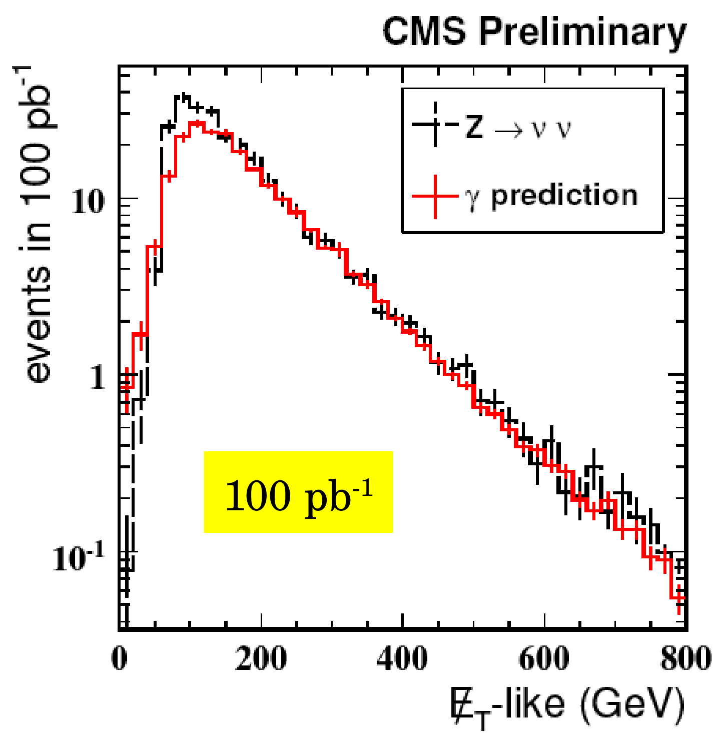




$Z \to \nu\nu$ from γ+jets (III)

no lepton mode





- corrected for:
  - $Z \to \nu\nu$ branching
  - selection efficiency

- theory correction is applied

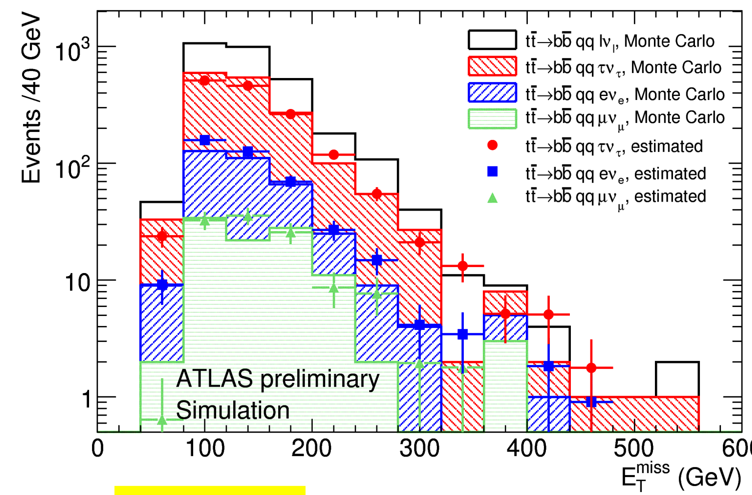
<u>for 100 pb^{-1}:</u>
stat. uncert. ~ 10%
syst. uncert. ~ 20%

ATL-PHYS-PUB-2009-083

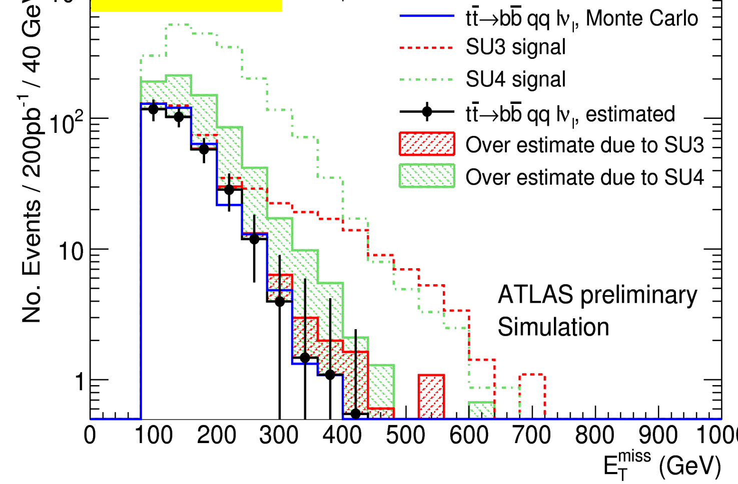
$t\bar{t}$ background

no lepton mode

- Dominant contribution:
  - τνqq
  - eνqq and μνqq with non-identified lepton
- Select top-anti-top sample with identified lepton and replace lepton by τ or by non-identified lepton
  - τ: take into account neutrino, tau-jet and lepton
  - electron: replace by “electronic jet”
  - μ: consider probability of inclusion of momentum to E_T^{miss}



200 pb^{-1}



<u>for 200 pb^{-1}:</u>

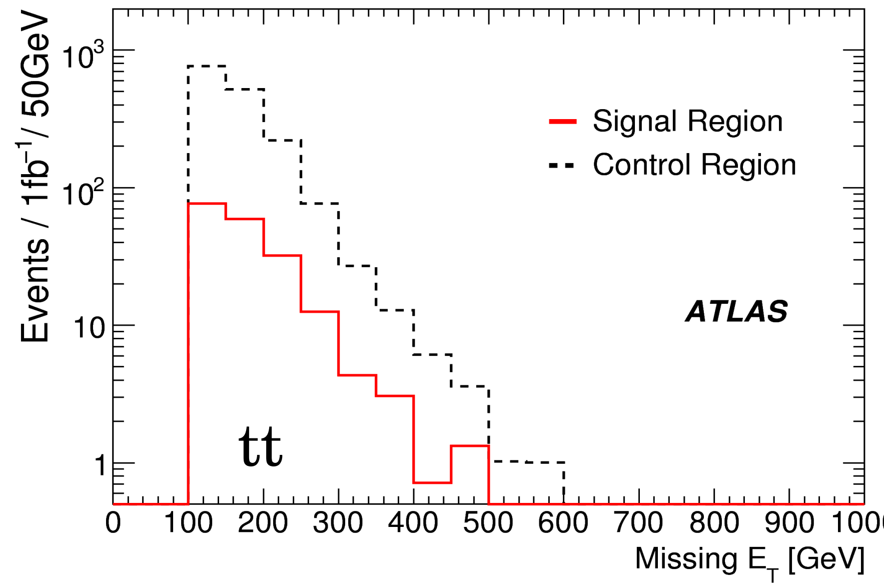
stat. uncert. ~ 8%

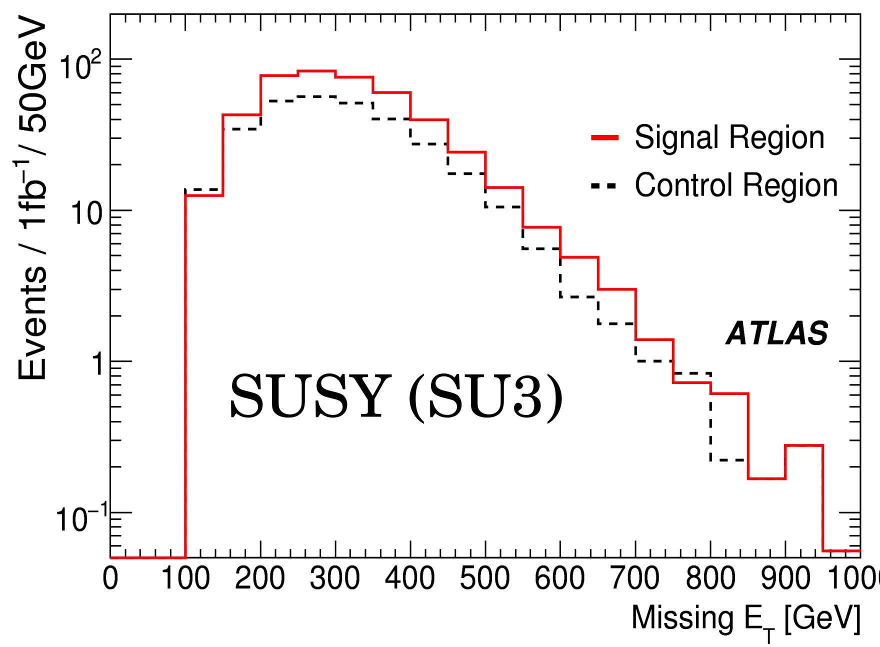
syst. uncert. ~ 36%

m_T-method: reversing m_T cut

1 lepton mode

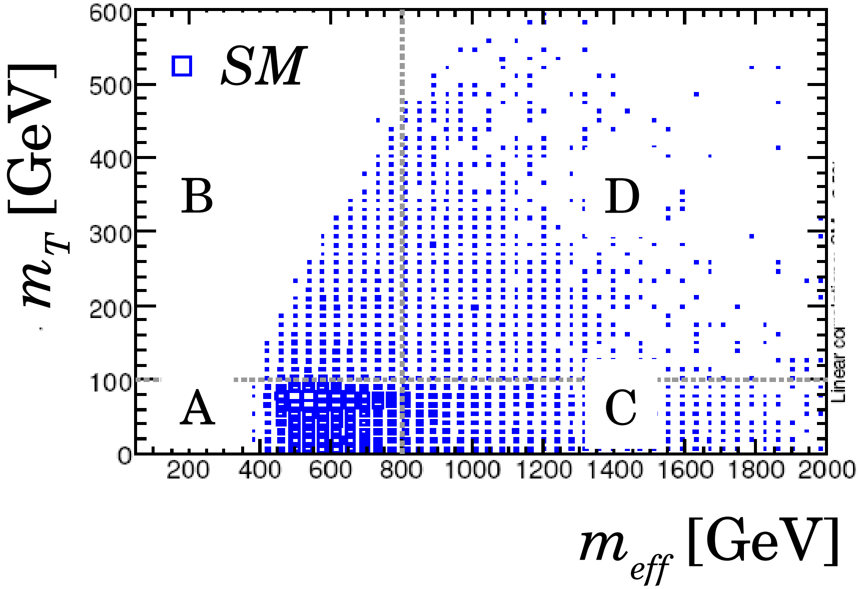
- m_T < 100 GeV: control sample
- m_T > 100 GeV: signal region
- take shape of background in the control sample
- normalize in 100 GeV < E_T^{miss} < 200 GeV
- works if
  - m_T and E_T^{miss} are uncorrelated
  - no SUSY in control sample and in signal sample for 100 GeV < E_T^{miss} < 200 GeV

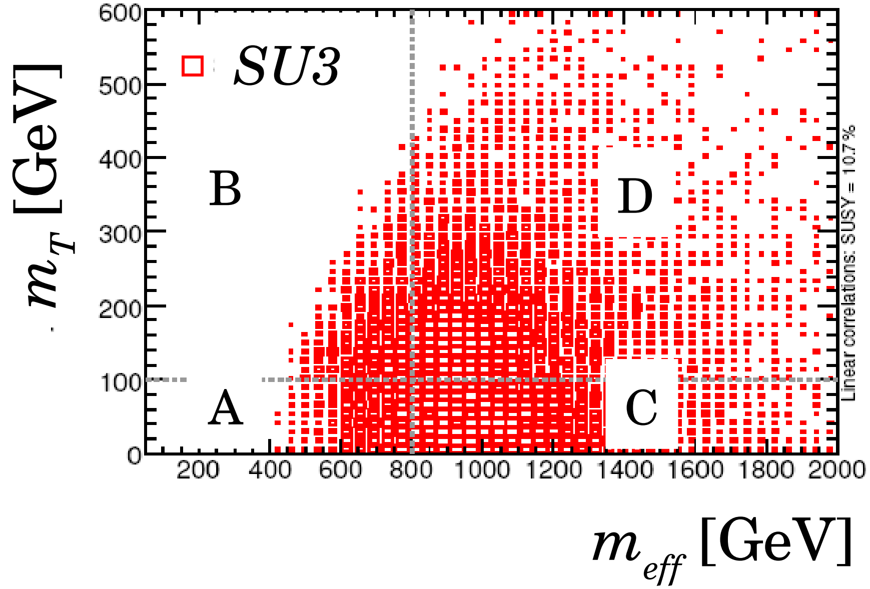




Tiles method, 2x2

1 lepton mode





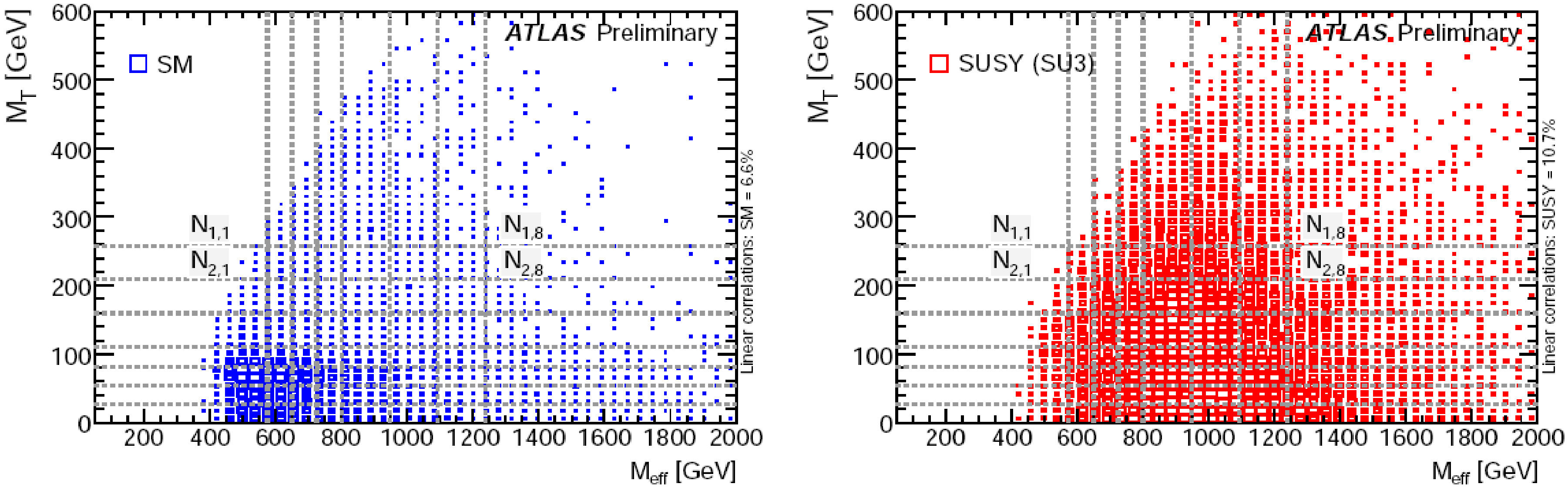
$$
\begin{cases}
N_A = f_A \cdot N_{SM} + SUSY_A \\
N_B = f_B \cdot N_{SM} + SUSY_B \\
N_C = f_C \cdot N_{SM} + SUSY_C \\
N_D = f_D \cdot N_{SM} + SUSY_D \\
SUSY_A \,/\, \mathrm{SUSY}_B = \mathrm{SUSY}_C \,/\, \mathrm{SUSY}_D
\end{cases}
$$

red quantities are unknown

- Divide m_{eff}-m_T plane into several tiles
- Each tile has contributions from signal and background
- Efficiencies to background are from the Monte Carlo (f_A, f_B, f_C, f_D)
- Assumption on signal: no correlation

Tiles method, generalisation NxN

1 lepton mode



- System of equations is over-constrained
- Minimize *extended* log-likelihood estimator

$$-\ln\mathscr{L} = \sum_{i,j=1}^{n} \left(\overline{N}_{ij} - N_{ij}\ln\overline{N}_{ij}\right)$$

Conclusion

- $Z \to \nu\nu$ background to SUSY search in the no-lepton mode is estimable from 3 control samples:
  - $Z \to ee$ and $Z \to \mu\mu$
  - $W \to \mu\nu$
  - γ+jets

impacted differently by SUSY

- Estimation of top-anti-top by replacement technique
- Several techniques of data-driven estimation of the combined background in the 1-lepton mode are developed
  - m_T-method
  - tiles method
- Plurality of (uncorrelated) techniques are necessary for understanding of SUSY background in early data

Backup slides

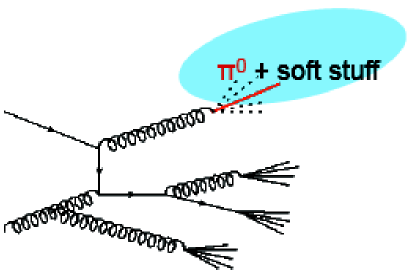
Definitions

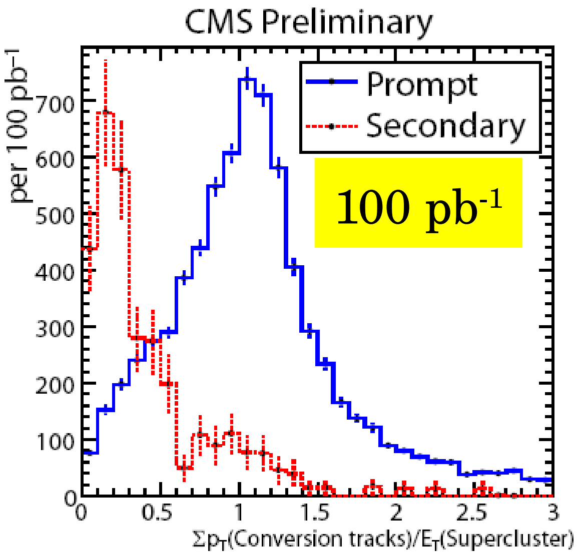
- m_T - "transverse mass", invariant mass, constructed from E_T^{miss} and lepton p_T
- m_{top} (combined fit) – invariant mass of 3 jets with largest vector-summed p_T
- $M_{\mathrm{eff}} \equiv \sum_{i=1}^{4} p_T^{\mathrm{jet},i} + \sum_{i=1} p_T^{\mathrm{lep},i} + E_{\mathrm{T}}^{\mathrm{miss}}$
- *mSUGRA* points
  - SU3: m_0=100 GeV, $m_{1/2}$=300 GeV, A_0=-300 GeV, $\tan\beta$=10, μ>0
  - SU4: m_0=200 GeV, $m_{1/2}$=160 GeV, A_0=-400 GeV, $\tan\beta$=10, μ>0
  - LM1: m_0=60 GeV, $m_{1/2}$=250 GeV, A_0=0 GeV, $\tan\beta$=10, μ>0

$Z \to \nu\nu$ from γ+jets (II)

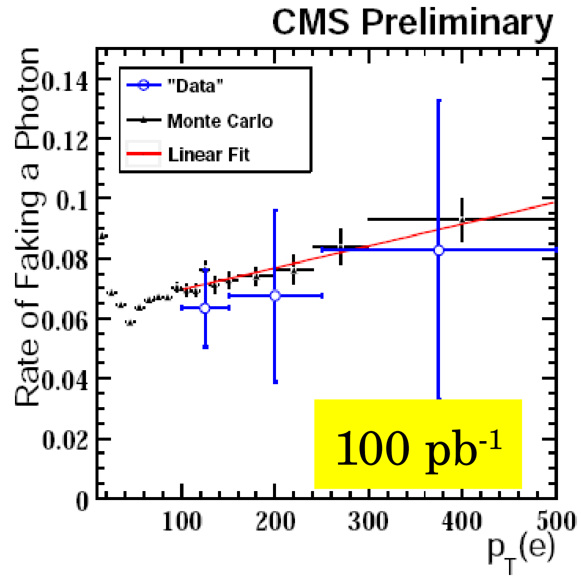
no lepton mode

- QCD background:
  - separated from the prompt photons by Σp_T(track)/E_T(ECAL) (converted electrons)
  - control sample is constructed by reversing the calorimetric isolation cut
  - normalize before SUSY selection cuts

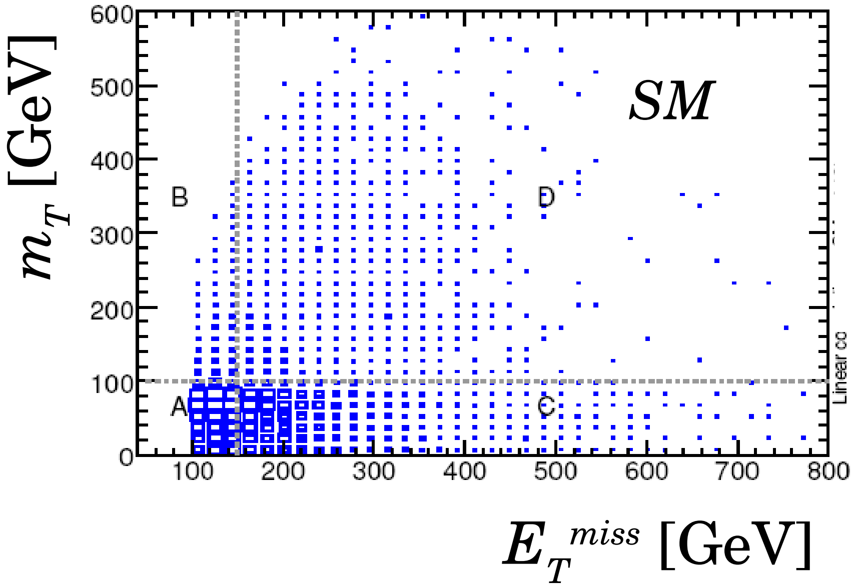


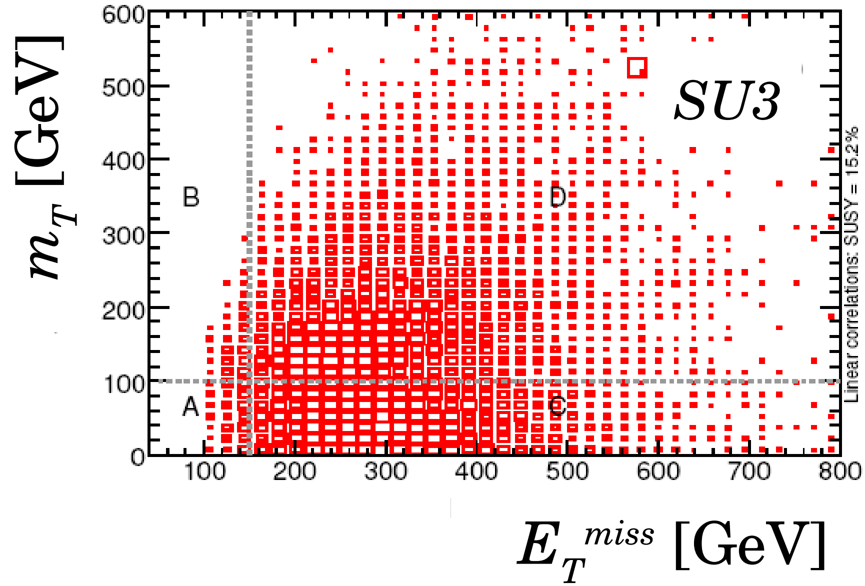


- Electron faking Photon
  - tag-and-probe method $Z \to ee$
  - 7 % in Barrel
  - 14 % in End Cap



m_T-method (II)

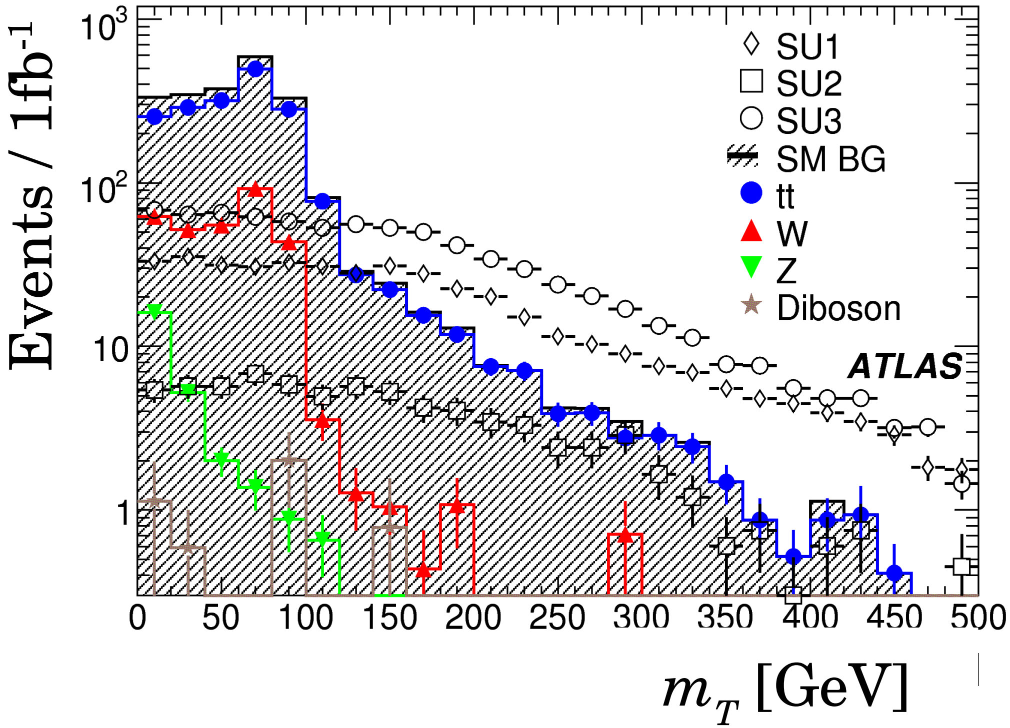




- signal contamination in A, B, and C
  - background in D is over-estimated
- Correlation of variables
  - E_T^{miss} is part of m_T
  - fractions of background contributions depend on m_T
  - over-estimated background for low-mass SUSY and under-estimated for high-mass SUSY

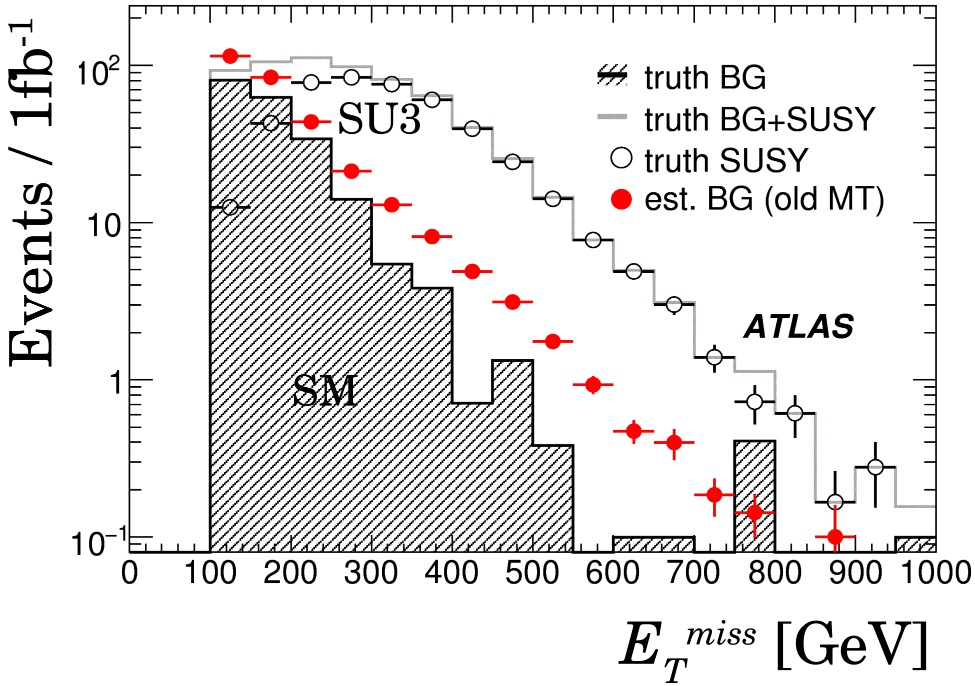
Correcting for signal contamination

- background: m_T falls off steeply
- signal: m_T falls only slowly
- correct for signal contamination:
  - estimate background in m_T>100 GeV
  - estimate signal in m_T>100 GeV
  - translate to low-m_T region
  - subtract from background

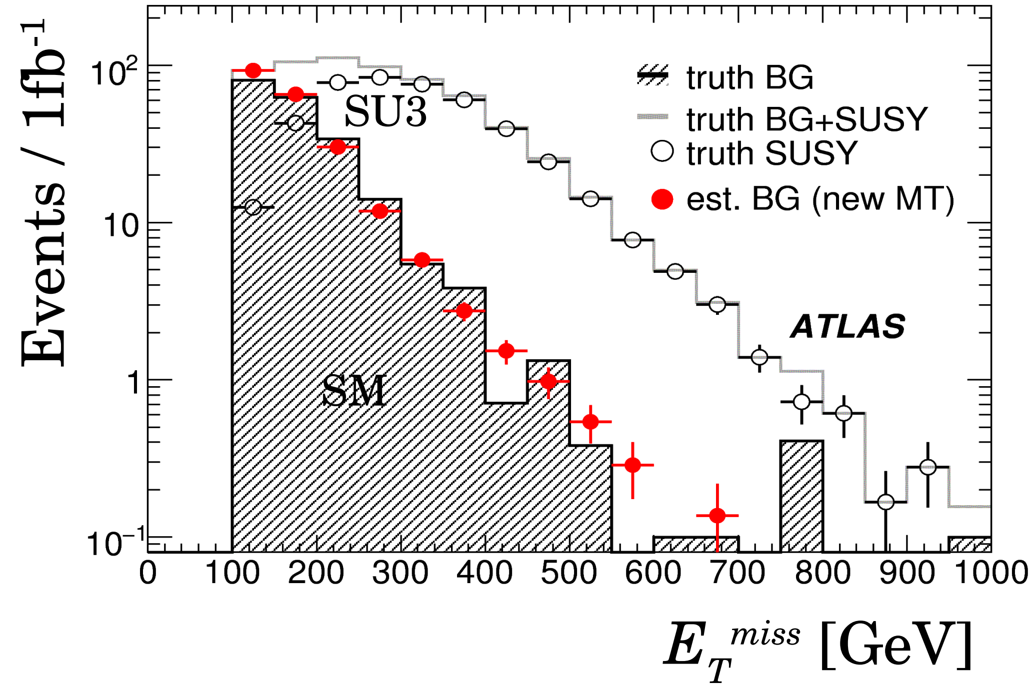


Correcting for signal contamination

m_T-method



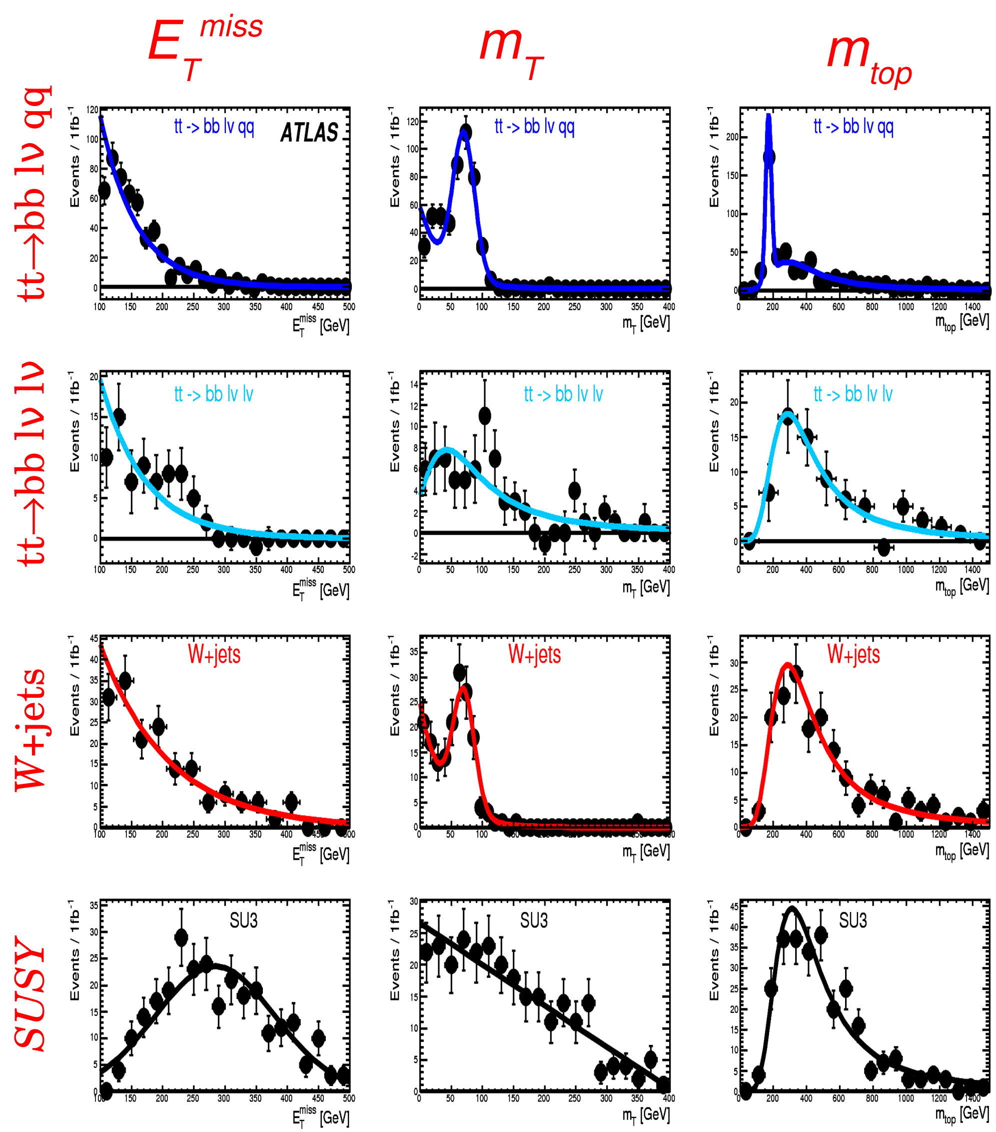
"New" m_T-method (corrected for SUSY presence in the control sample



- May not work for low-mass SUSY
- Assumption on signal is needed to transfer the signal from signal region to control sample

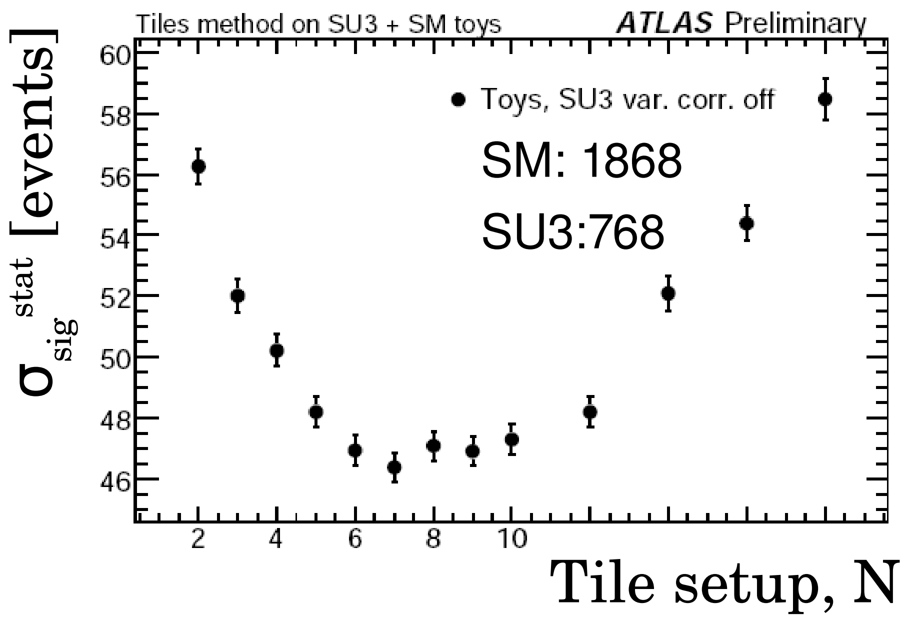
Further development: combined fit

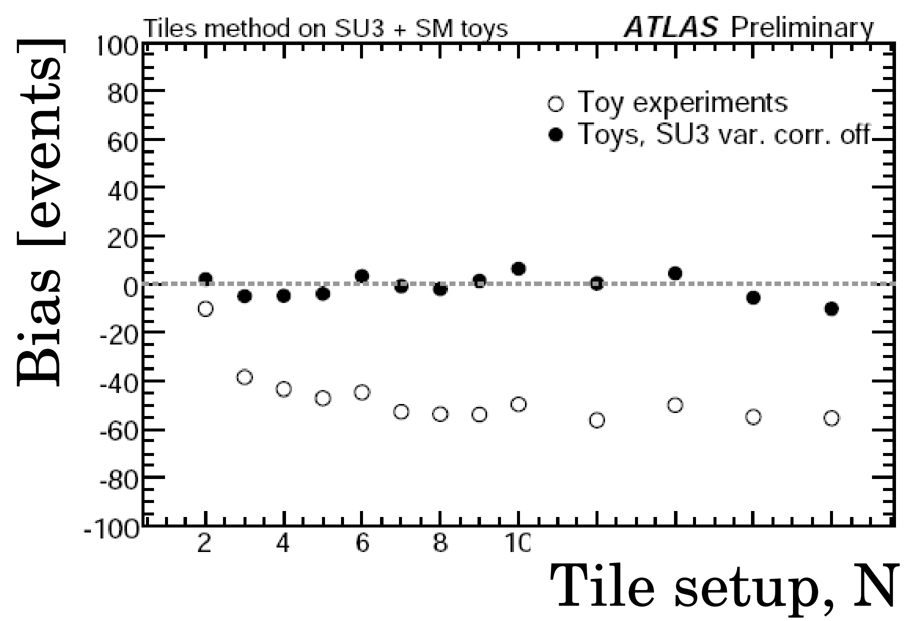
- Also take into account correlation between E_T^{miss} and m_T
- Build "empirical model" (p.d.f.) of each background component based on Monte Carlo and control samples
- "Generic shape" for signal contamination in the control sample
- Fit the sum of background "empirical models" and signal "generic shape" to the data in E_T^{miss}, m_T and m_{top} 3D space.
- Fit parameters:
  - normalization of each background
  - "empirical model" parameters



Tiles method: NxN optimization

- Few tiles:
  - ☹ few degrees of freedom
  - ☺ less sensitive to signal correlation
- many tiles:
  - ☹ large statistical fluctuation
  - ☺ goodness of fit
  - ☺ sensitivity to signal shape
  - ☺ new parameters:
    - linear correlation factor
    - separate background contributions





Tiles method vs Combined Fit

	Combined fit	**Tiles method**
Variables	E_T^{miss} m_T m_{Top}	E_T^{miss} m_{eff}
Background	Empirical model, constructed separately for each contribution	Total background shape is from MC
Signal assumption	"generic shape"	No correlation
"Hidden" SUSY	shape close to linear combination of background	shape close to total background

Replacement techniques

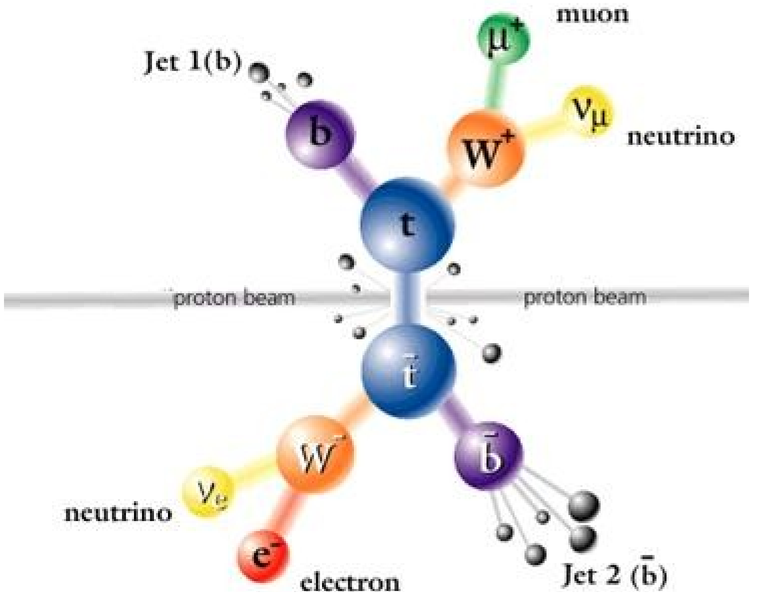
Basic idea: isolate events with “simple” topology and replace one reconstructed object by another

- top re-decay: select ttbar events with low E_t^{miss}, reconstruct both top's kinematics, remove their decay products and re-decay
- $Z \rightarrow \nu\nu$ from $Z \rightarrow ll$
- $tt \rightarrow bb\ l\nu\ \tau\nu$ from $tt \rightarrow bb\ l\nu\ l\nu$
- $tt \rightarrow bb\ qq\ \tau\nu$ from $tt \rightarrow bb\ qq\ l\nu$

Replacement technique: top re-decay

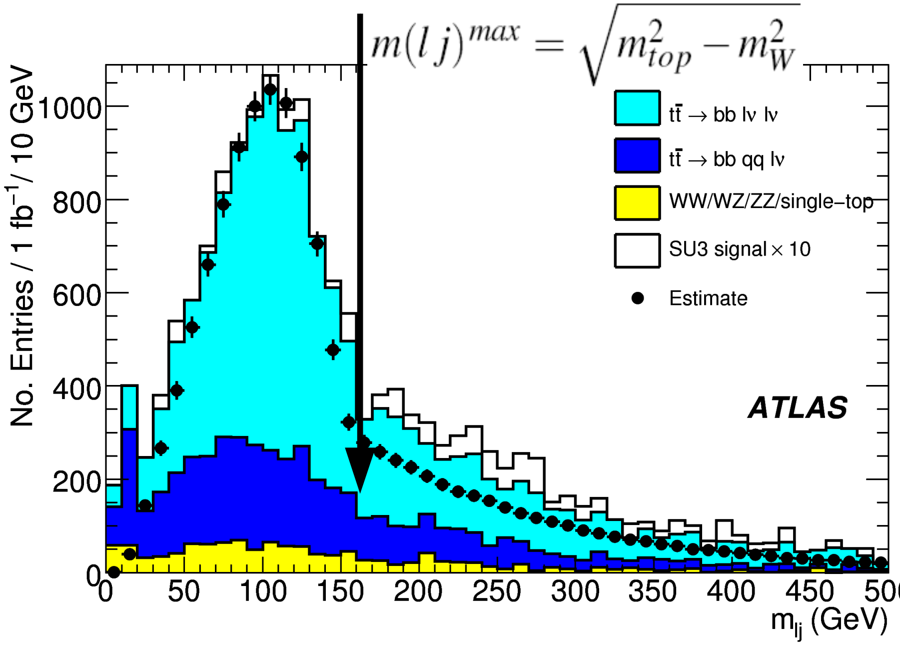
- seed events selection:
  - 2 opposite sign leptons, 2 jets
  - low E_T^{miss}
  - only one of two combinations jet-lepton

    $m(l, j) < \sqrt{m_{Top}^2 - m_W^2}$



- top kinematic reconstruction

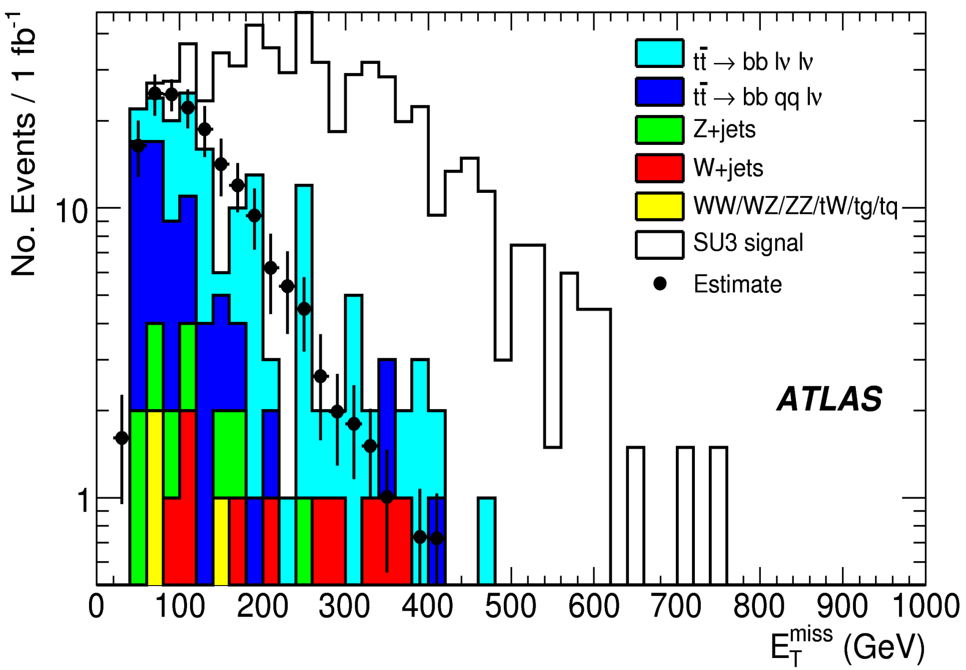
$$m_W^2 = (p_{l1} + p_{\nu 1})^2$$
$$m_W^2 = (p_{l2} + p_{\nu 2})^2$$
$$m_t^2 = (p_{l1} + p_{\nu 1} + p_{b1})^2$$
$$m_t^2 = (p_{l2} + p_{\nu 2} + p_{b2})^2$$
$$E_x^{miss} = p_{(\nu 1)x} + p_{(\nu 2)x}$$
$$E_y^{miss} = p_{(\nu 1)y} + p_{(\nu 2)y}$$



Replacement technique: top re-decay (2)

- Top decay simulation:
  - PYTHIA 6.4 + Atlfast, leptonic decays
  - 1000 decays for each seed event:
    - increase statistics but introduce bin-to-bin correlations

- Remove original top decay products and merge resimulated top's
- Apply SUSY search cuts
- Normalize to data in low-E_T^{miss} region
- robust against SUSY contamination



Other methods

- QCD background in no-lepton mode: jet smearing
  - measure jet response function from jet+γ and 3 jet events
- $HT_2 = \sum_{i=2}^{4} p_T^{\text{jet}\,i} + p_T^{\text{lepton}}$ versus E_T^{miss} significance
  - a pair of (almost) uncorrelated variables with high separation power
- τ reconstruction
  - tt→bb lν τν in 1-lepton mode and tt→bb τν qq in no-lepton mode
- Topbox: tt→bb lν qq in 1-lepton mode
  - selection of clean sample of tt→bb τν qq events reconstructing the invariant masses or top and W decay products
- QCD background in 1-lepton mode: lepton isolation
  - estimate E_t^{miss} shape from QCD events with non-isolated lepton and normalize to number of events with isolated lepton in low-E_t^{miss} region