A study of boosted $Z^0 \rightarrow e^+e^-$ signatures at CMS

James Jackson

Rutherford Appleton Laboratory

Boost 2010, Oxford

24-06-2010

CMS

- One of 2 general purpose LHC detectors

- Main features:
  - 3.8T solenoidal magnet
  - Tracking up to $|\eta| = 2.5$
  - Calorimetry up to $|\eta| = 3$ (Forward hadronic up to 5.2)
  - Muon systems up to $|\eta| = 2.4$
  - Level-1 trigger relies on coarse calorimetry + muon systems
  - High Level Trigger adds tracking information and fine grain calorimetry / muon information
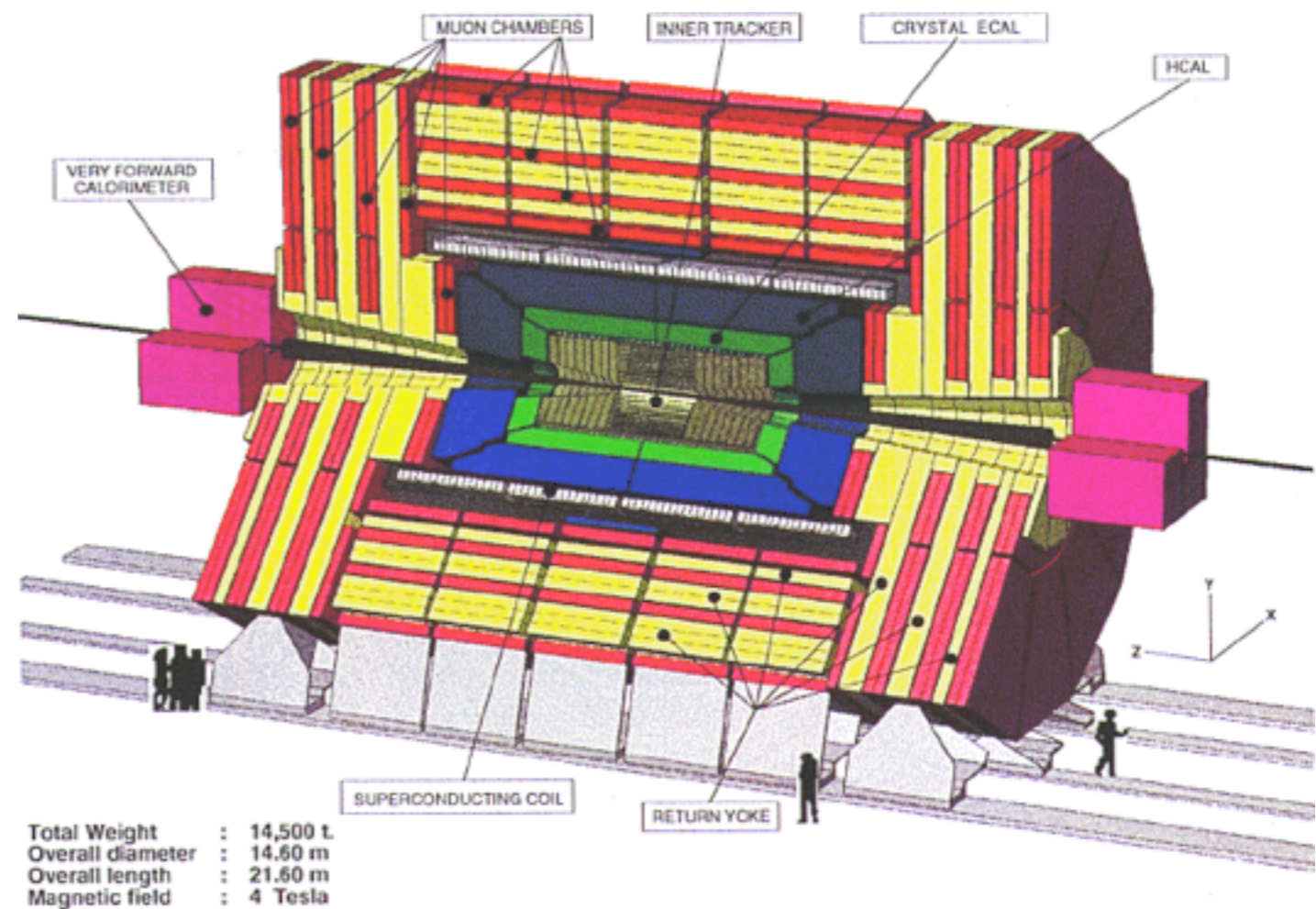
Analysis introduction

- Aim is to produce a model independent search for new heavy resonances decaying to $Z^0 + X$

- Use a reference excited quark model to benchmark the analysis

- Excited fermions are taken to be spin 1/2, isospin 1/2 partners, assumed to acquire a mass before EWK symmetry breaking. The matter content becomes:

$$l_L \equiv \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad l_R \equiv e_R; \quad l_L^* \equiv \begin{pmatrix} \nu_e^* \\ e^* \end{pmatrix}_L, \quad l_R^* \equiv \begin{pmatrix} \nu_e^* \\ e^* \end{pmatrix}_R$$
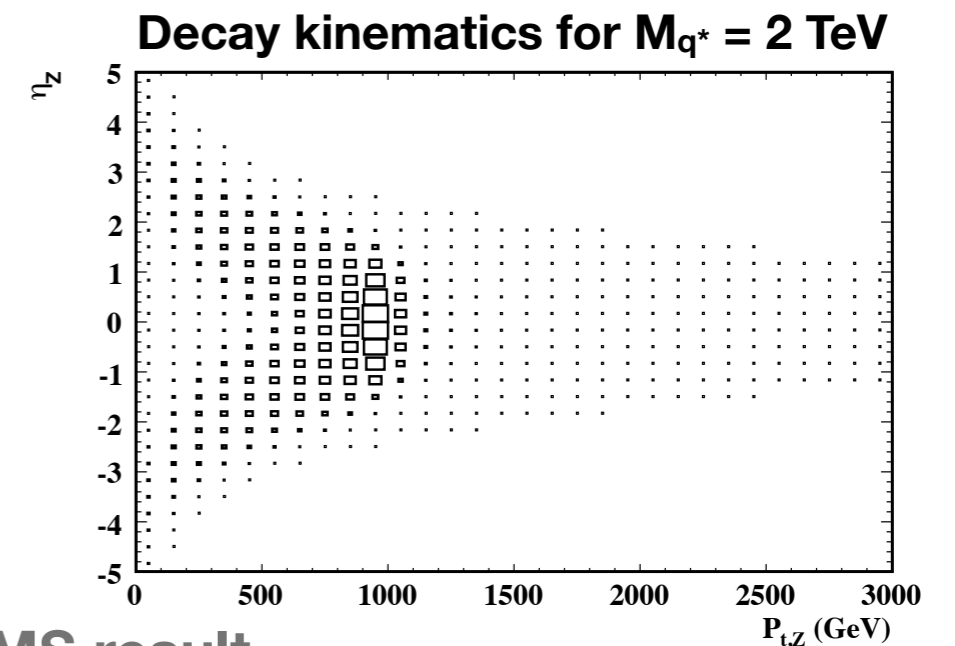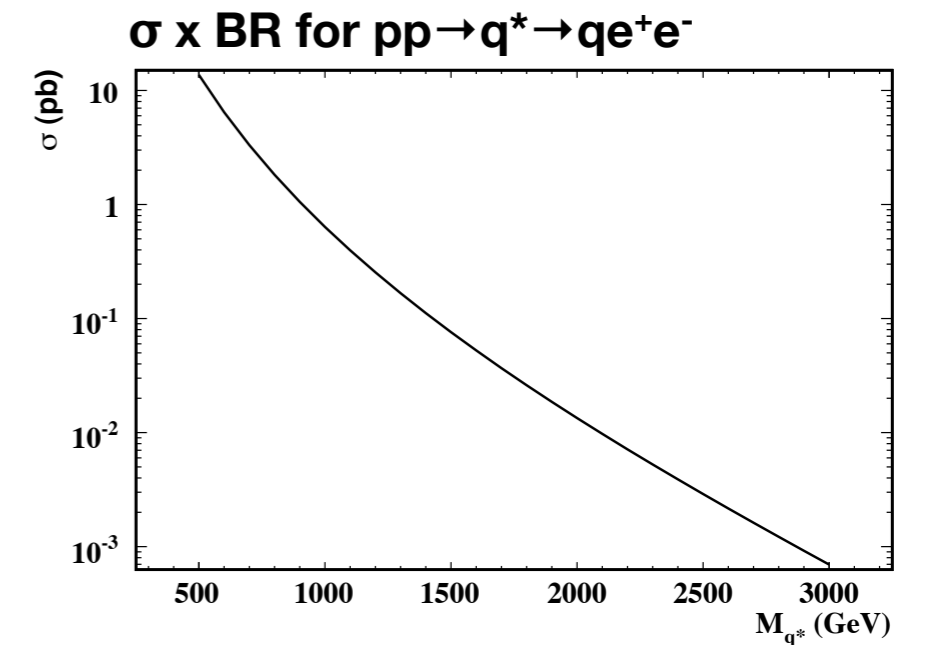
$$q_L \equiv \begin{pmatrix} u \\ d \end{pmatrix}_L, \quad q_R \equiv u_R, d_R; \quad q_L^* \equiv \begin{pmatrix} u^* \\ d^* \end{pmatrix}_L, \quad q_R^* \equiv \begin{pmatrix} u^* \\ d^* \end{pmatrix}_R$$

- Transitions between SM and excited fermion states are given by:

$$\mathcal{L}_{\text{eff}}^2 = \frac{1}{2\Lambda} \bar{f}_R^* \sigma^{\mu\nu} \left(f_s g_s \frac{\lambda^a}{2} G_{\mu\nu}^a + f g \frac{\tau^a}{2} W_{\mu\nu}^a + f' g' \frac{Y}{2} B_{\mu\nu} \right) f_L$$

- Use the choice $f_s = f = f' = 1$, and set $\Lambda = m_{q^*}$

Note that this is a thesis analysis, and not an official CMS result

σ x BR for pp→q*→qe⁺e⁻

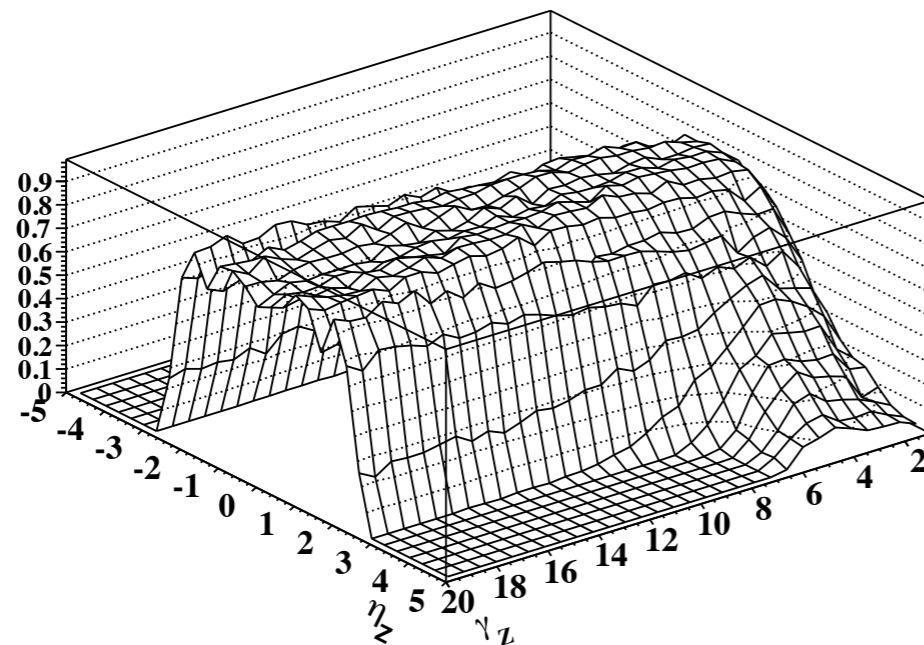**Decay kinematics for M_{q^*} = 2 TeV**

Close electron reconstruction

- Z decay electrons can become very close in the calorimeter - ~0.1 rad for $M_X > 2$ TeV

- This causes a problem due to Bremsstrahlung recovery algorithms at the SuperClustering stage

- Clusters from electrons which are close and aligned in phi are combined into one SuperCluster

- Modified algorithm:

  - Run `FixedMatrix5x5` clustering algorithm in EB (currently only run in EE) to avoid a BasicCluster merging the two clusters

  - Promote all BasicClusters to SuperClusters (with 15 GeT E_t cut)

 - Re-run GSF electron reconstruction with new SuperCluster collections
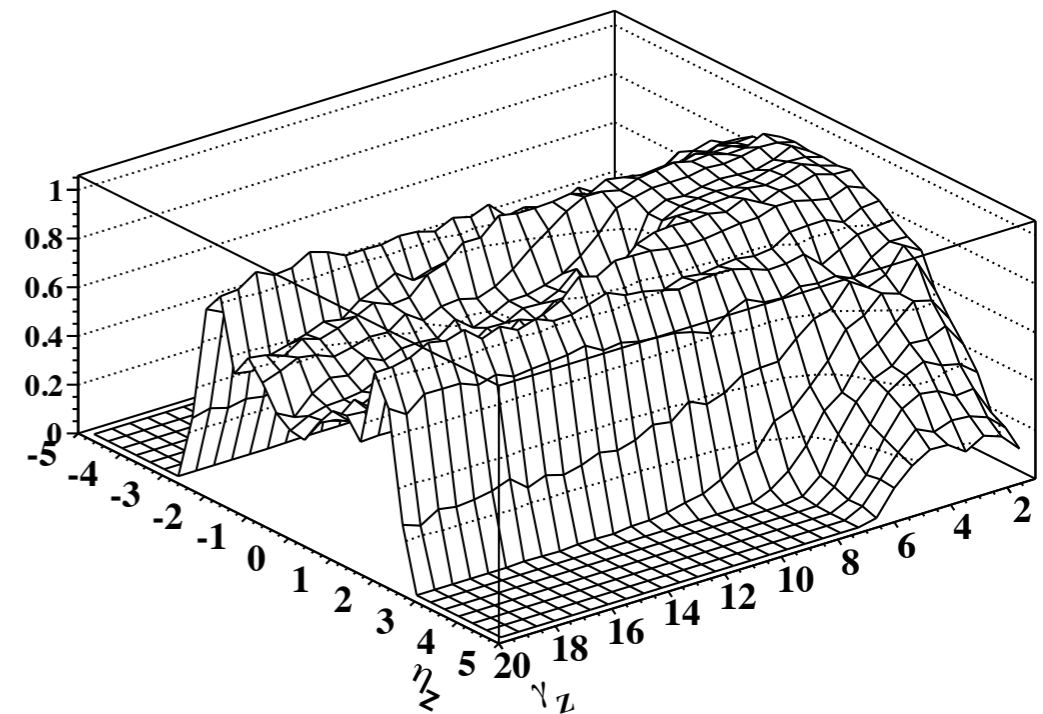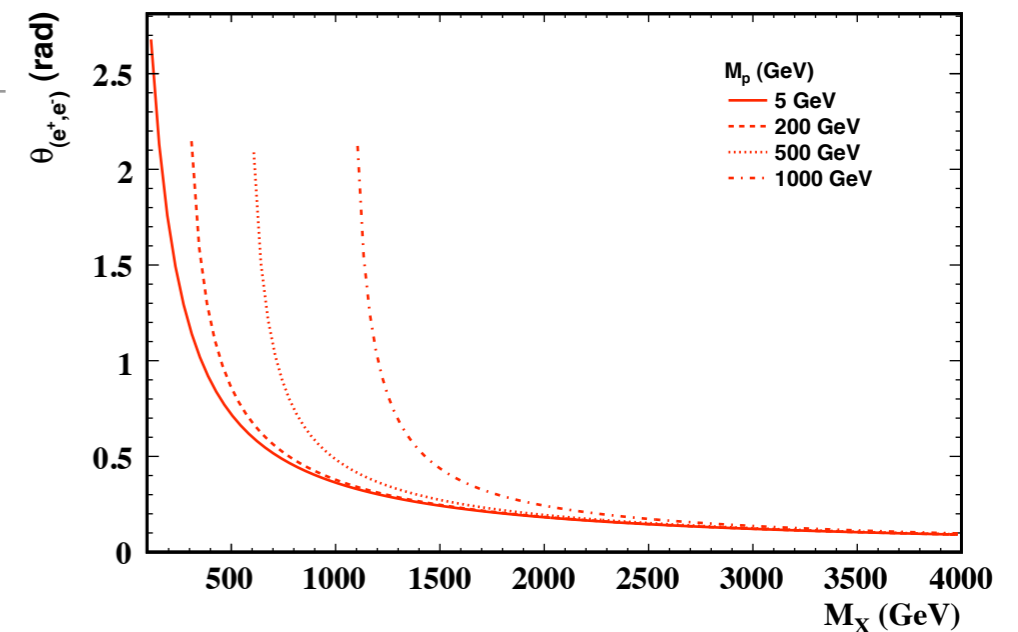
Close electron reconstruction

- Performance with new algorithm shows good improvement at high γ_Z

- Some loss in efficiency at low γ_Z due to lack of Bremsstrahlung recovery
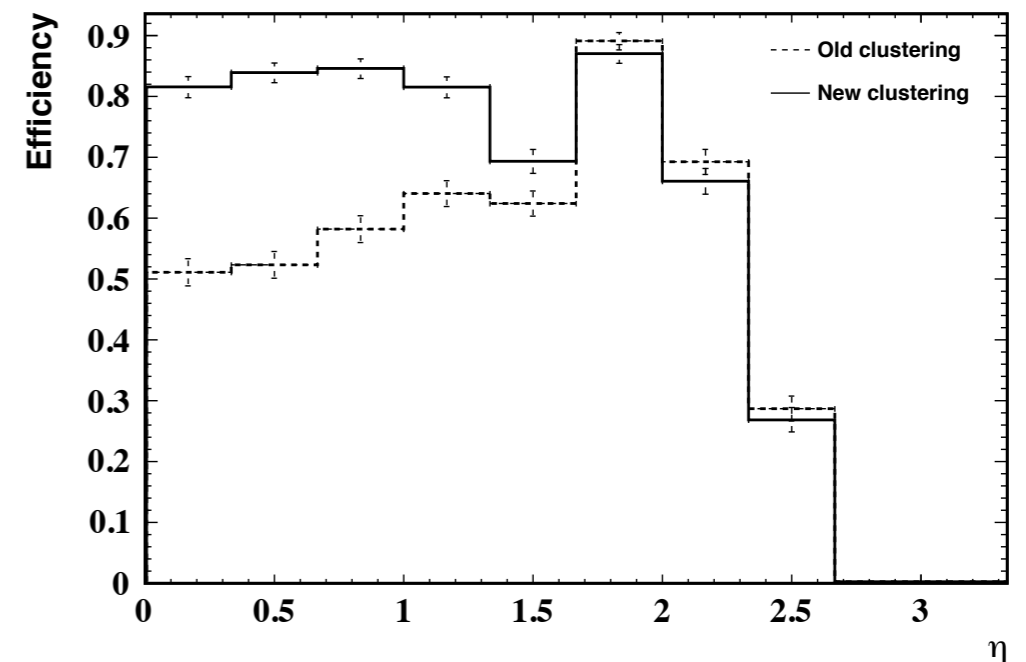
- Need to check this is limited to low P_t

Z reconstruction efficiency comparison ($\gamma_Z = 1$)
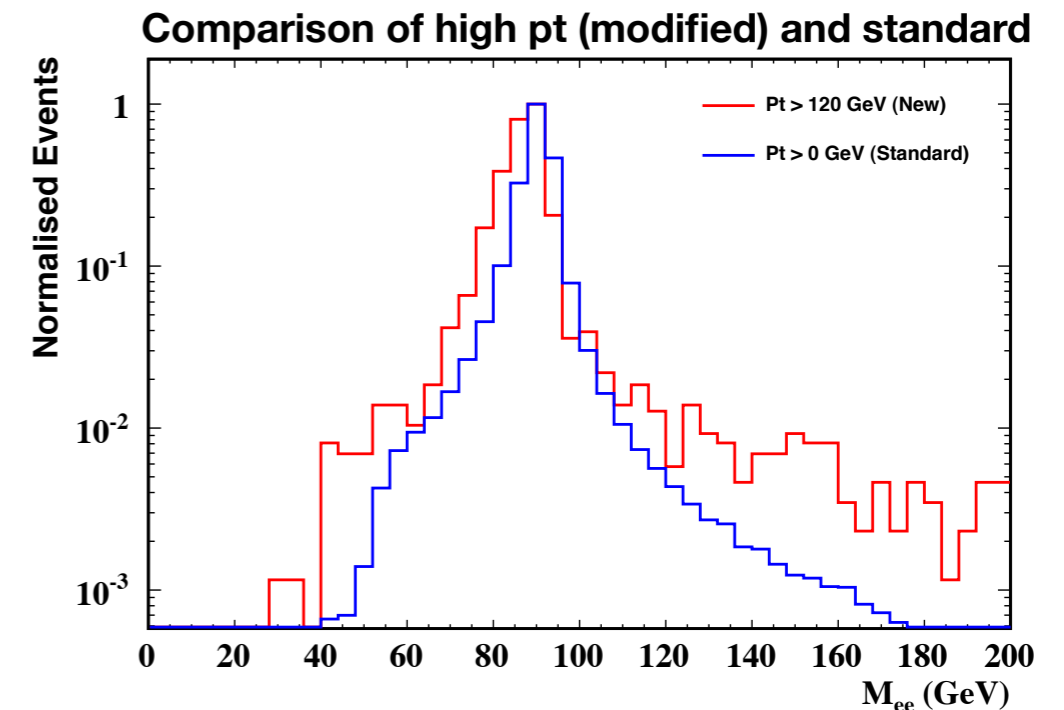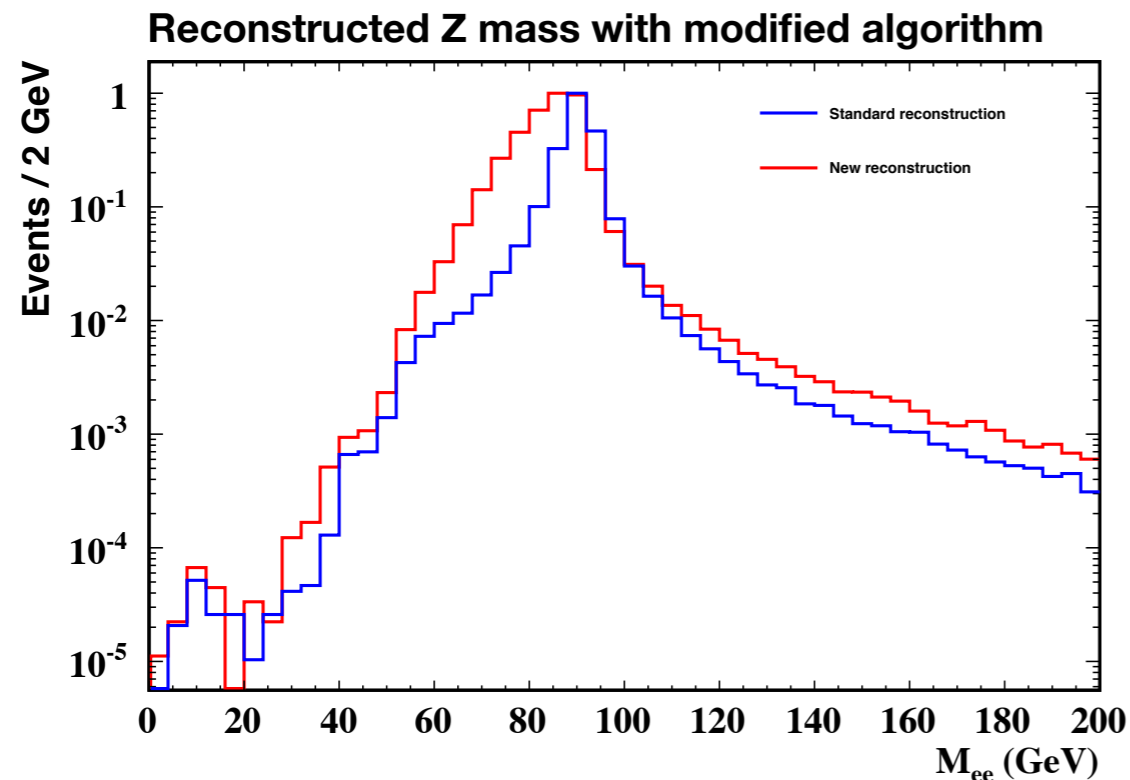
Z reconstruction efficiency (new algorithm)
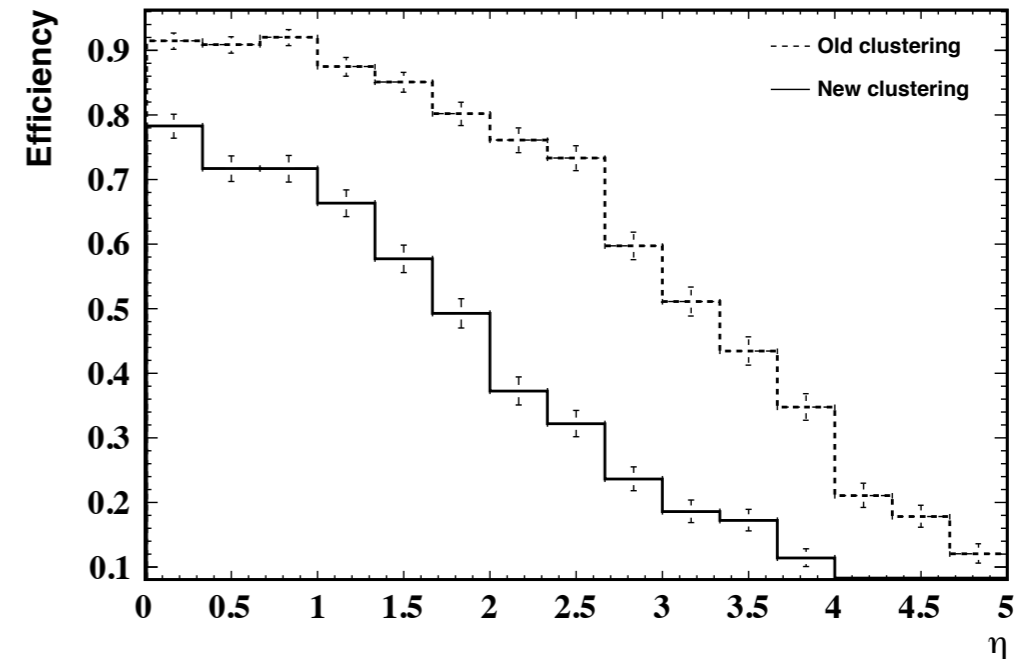
Z reconstruction efficiency comparison ($\gamma_Z = 20$)

Quantification of energy loss

Reconstructed Z mass with modified algorithm

Standard reconstruction
New reconstruction

Comparison of high pt (modified) and standard

Pt > 120 GeV (New)
Pt > 0 GeV (Standard)

Fractional energy loss

- Z peak shows low sideband behavior typical of missing energy (due to Bremsstrahlung photons not included)

- First check this is limited to low pt electrons by imposing pair pt cut. Sideband behaviour restored

- Cross-check by (from simulation) calculation fractional energy loss for electrons - dominates in the barrel to endcap transition region around E = 100 GeV
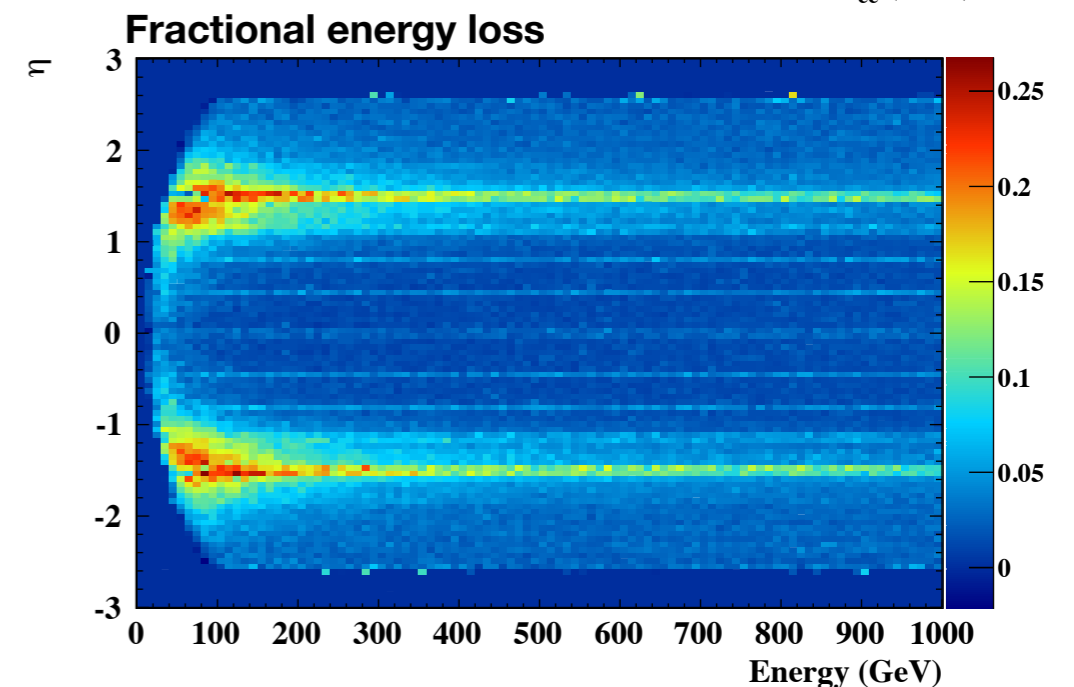
Electron selection

Barrel HEEP selection performance

Cut	1 TeV	1.25 TeV	1.5 TeV	1.75 TeV	2 TeV
E_t	0.96	0.97	0.98	0.98	0.98
$\vert\eta_{\mathrm{SC}}\vert$	0.98	0.99	0.99	0.99	0.99
$\vert\Delta\eta_{\mathrm{in}}\vert$	0.99	0.99	0.99	0.99	0.99
$\vert\Delta\phi_{\mathrm{in}}\vert$	1.00	1.00	1.00	0.99	1.00
H/E	0.99	0.99	0.99	0.99	0.99
$E^{2\times5}/E^{5\times5}$	0.96	0.96	0.96	0.96	0.97
EM + Had D1 Isolation	0.90	0.7	0.57	0.45	0.36
Track p_t Isolation	0.96	0.89	0.71	0.56	0.46

Endcap HEEP selection performance

Cut	1 TeV	1.25 TeV	1.5 TeV	1.75 TeV	2 TeV
E_t	0.96	0.96	0.96	0.97	0.97
$\vert\eta_{\mathrm{SC}}\vert$	0.93	0.92	0.92	0.92	0.92
$\vert\Delta\eta_{\mathrm{in}}\vert$	0.98	0.99	0.99	0.99	0.99
$\vert\Delta\phi_{\mathrm{in}}\vert$	1.00	1.00	1.00	1.00	1.00
H/E	0.99	0.99	0.99	0.99	0.99
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	1.00	0.99	0.99	0.99	0.99
EM + Had D1 Isolation	0.96	0.93	0.87	0.83	0.79
Had D2 Isolation	0.97	0.97	0.97	0.97	0.96
Track p_t Isolation	0.99	0.98	0.94	0.90	0.88

Original HEEP selection performance

m_{u^*} (TeV)	$Efficiency$
1	0.676 ± 0.004
1.25	0.566 ± 0.004
1.5	0.434 ± 0.004
1.75	0.332 ± 0.003
2.0	0.264 ± 0.002

Modified HEEP selection performance

m_{u^*} (TeV)	$Efficiency$
1	0.797 ± 0.005
1.25	0.823 ± 0.005
1.5	0.842 ± 0.006
1.75	0.853 ± 0.006
2.0	0.864 ± 0.005

- EM + Had Depth 1 and Track p_t isolation cuts perform badly - these are removed to give the modified HEEP selection
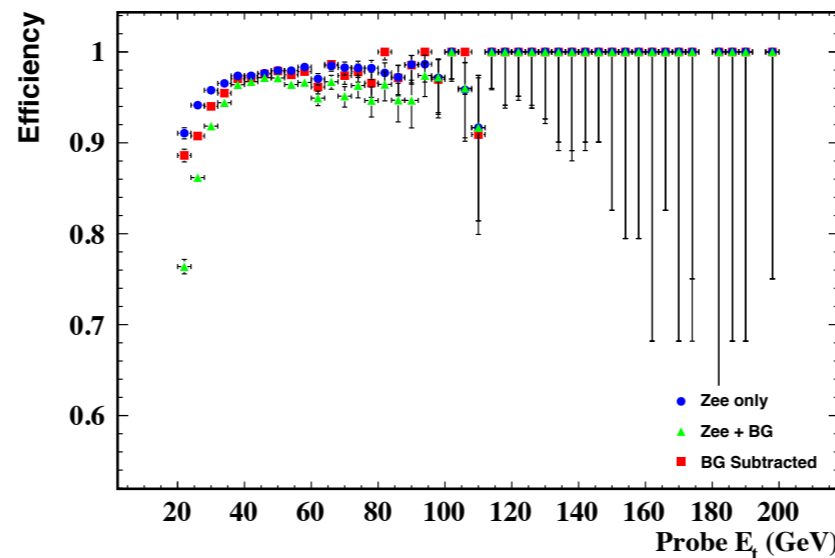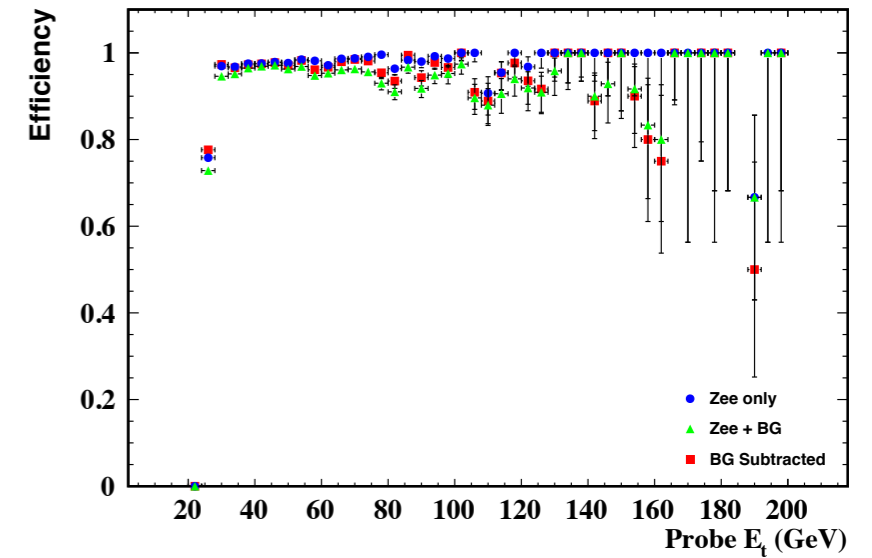
All efficiencies measured from data with Tag+Probe

- Backgrounds estimated with simple sideband counting technique
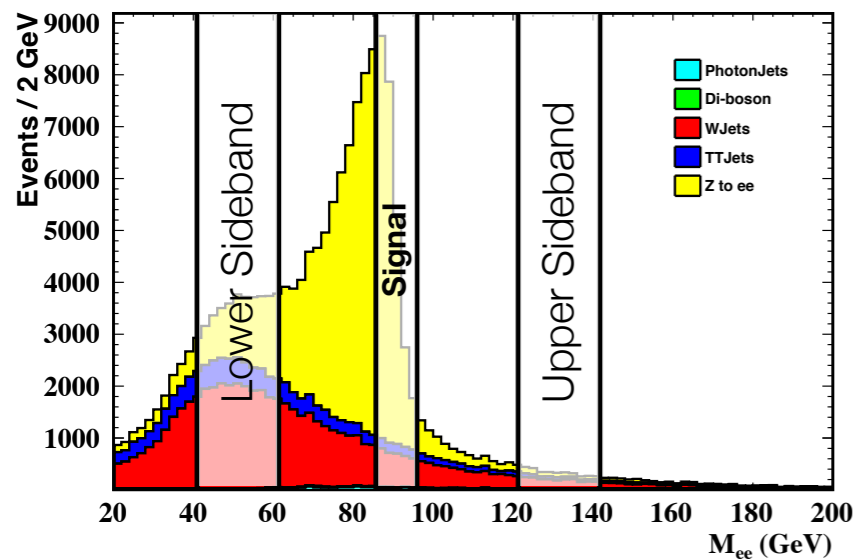
- All efficiencies > 95% after turn on regions

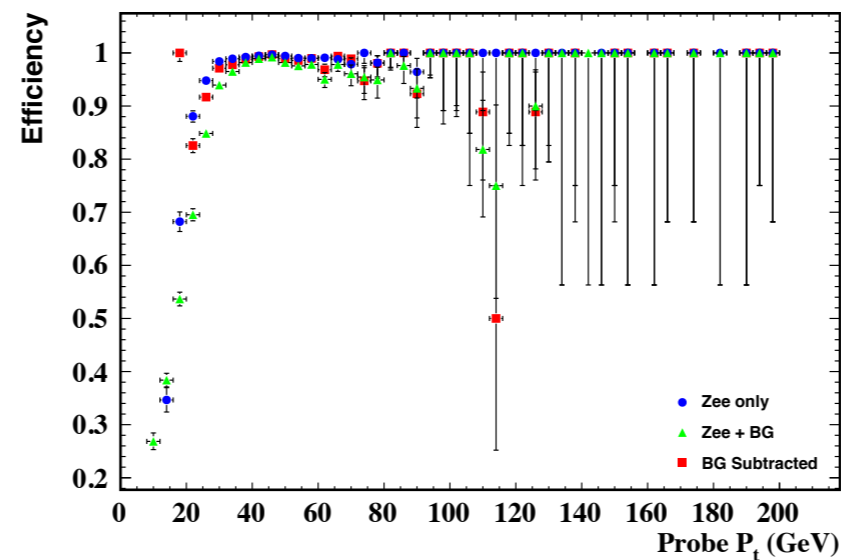GSF reconstruction efficiency

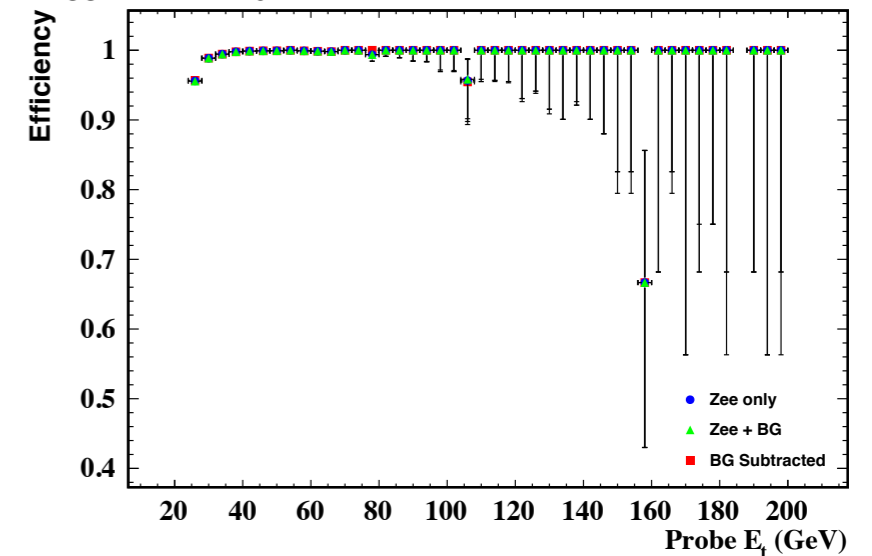

HEEP ID efficiency

Cluster All: Di-electron Mass

Cluster efficiency

Trigger efficiency

Background estimation

- After all event selection cuts are applied, there are three backgrounds to be individually estimated

  - X + Jets
    - Estimate with the fake-rate method
  - tt
    - Estimate with the b-tagging method
  - SM Z→e$^+$e$^-$
    - Estimate from MC or W from data

- Estimations are not used by statistical tools, but to check that the sample is understood and under control

Final selected events
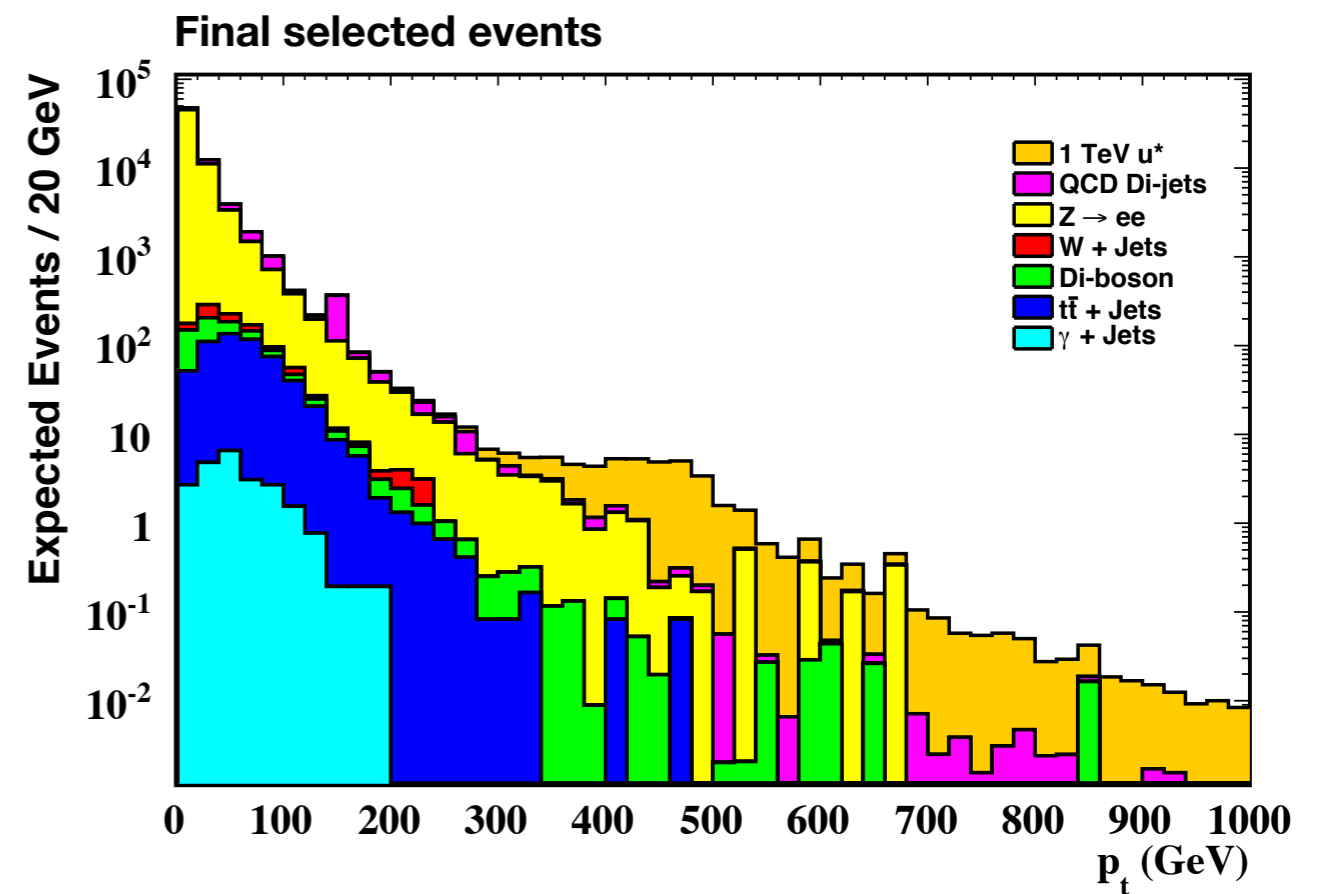
Combination of background estimations


Estimates without signal / Estimates with 1TeV u*
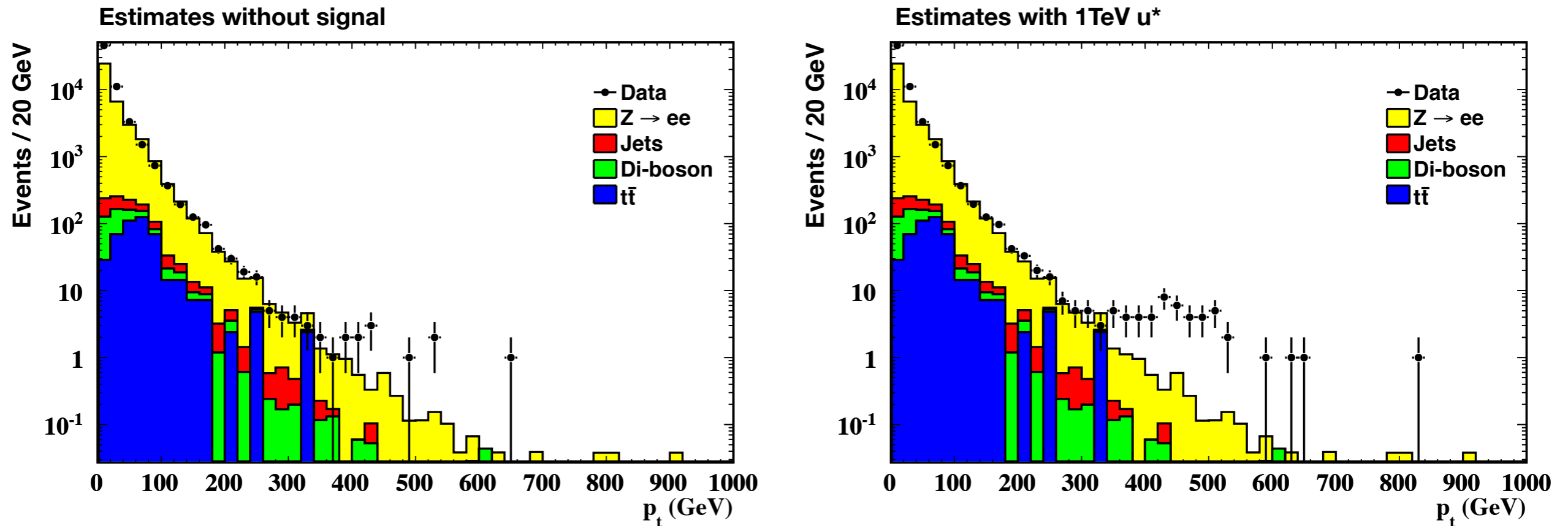
- The combination is shown for 200 pb^{-1} psuedo-experiments with and without a 1 TeV u*

Determining signal significance

- Given some set of data, is there an excess, and at what significance?
- Take the hypothesis that background follows the functional form $e^{-\alpha p_t} p_t^{-\beta}$
- Run a fit (RooFit) to this PDF in the range 100-1000 GeV
- Use result of the fit to construct background hypothesis histogram with same binning as data
- Compute the probability that the contents of each bin±1 (sliding window of 60 GeV) are due to a Poisson fluctuation (p(N >= obs)) around the background (use $-\log_{10}(p)$ for convenience)

~10$^{-8.1}$ chance this is due to BG fluctuations

200pb^{-1}, 1 TeV u*
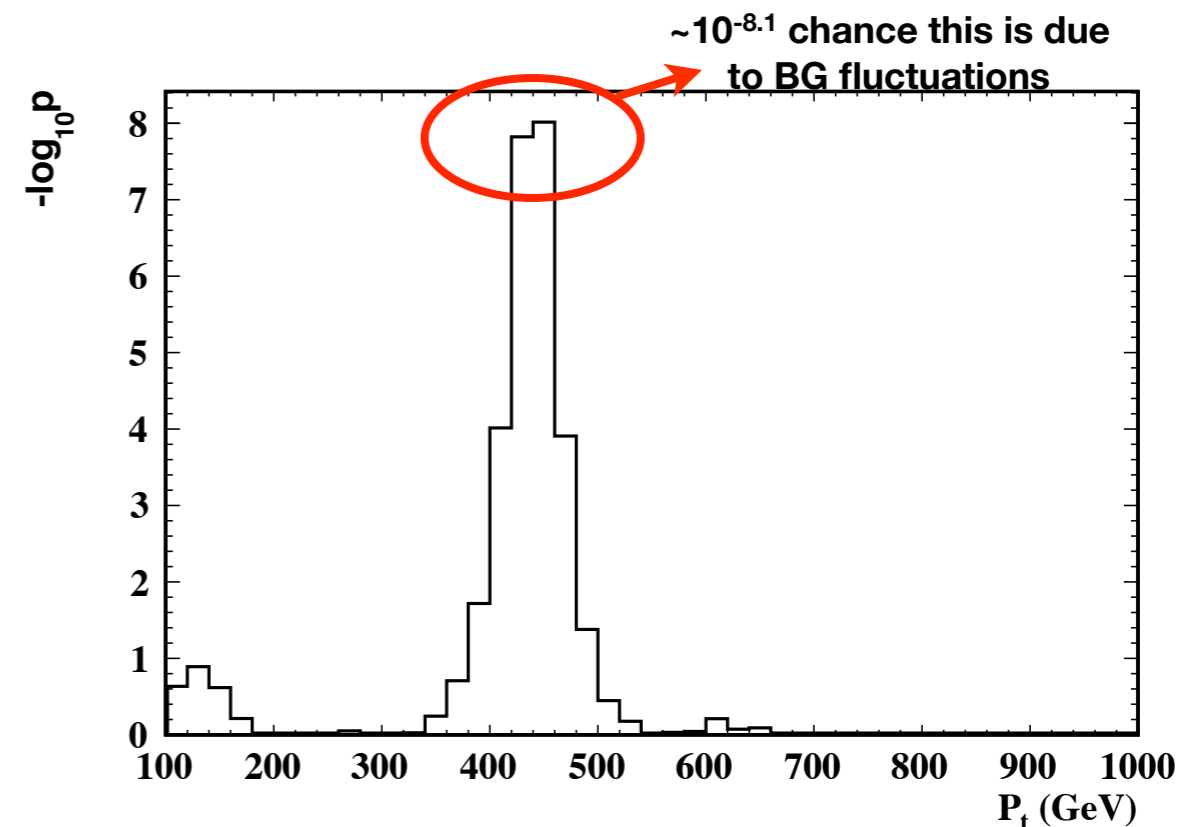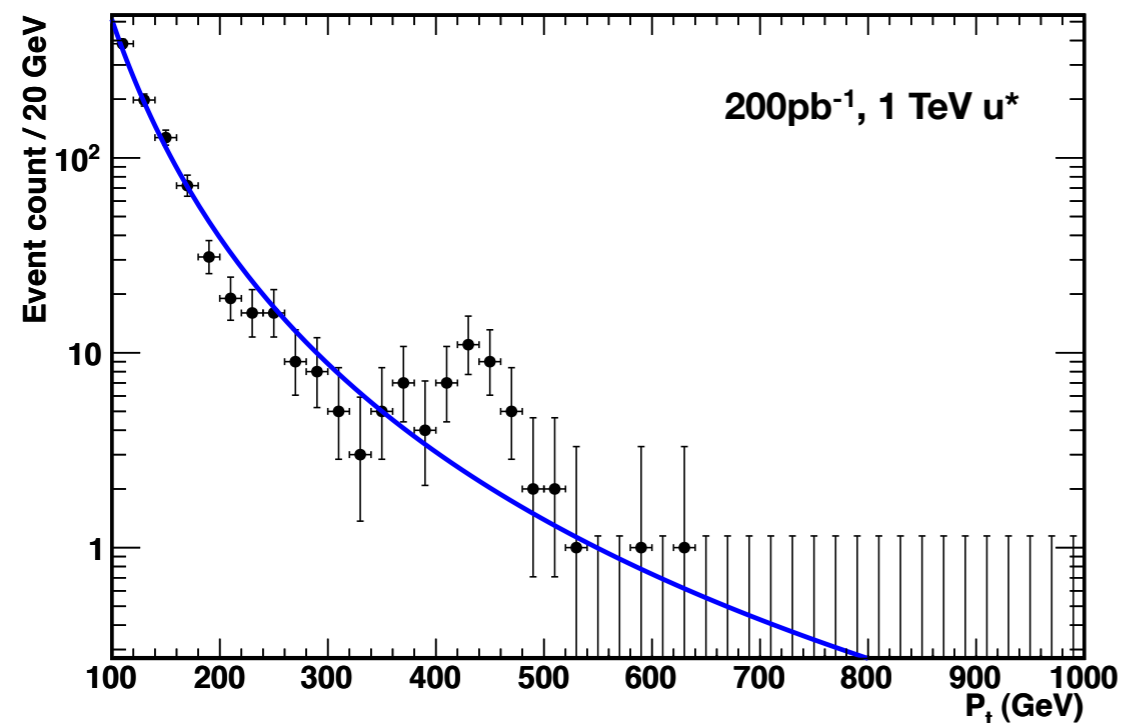
What does $p_{BG} = 10^{-8.1}$ mean?

- With an experiment p-value at hand, need to determine if it is significant
- *Look-elsewhere effect*, or likelihood of same significance occurring due to BG fluctuations anywhere in the search region, to take into account the many correlated p-value measurements
- For a given integrated luminosity, run 1B background only pseudo-experiments, where bin contents are allowed to vary following Poisson statistics
- For each experiment, run a scan to determine the minimum p-value
- Histogram as a function of $-\log_{10}(p)$, and construct a weighted mean to determine the most likely p-value from background fluctuations

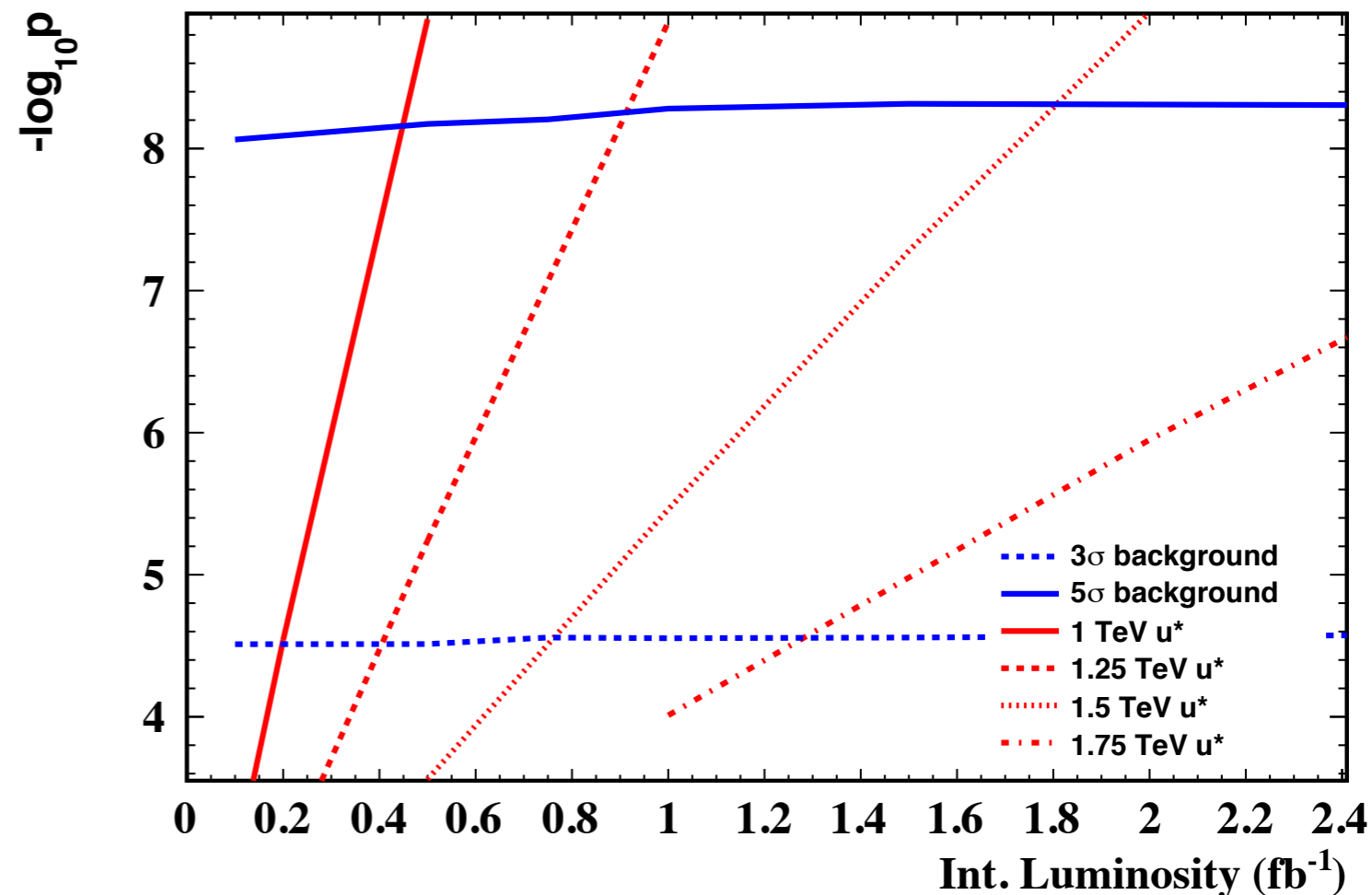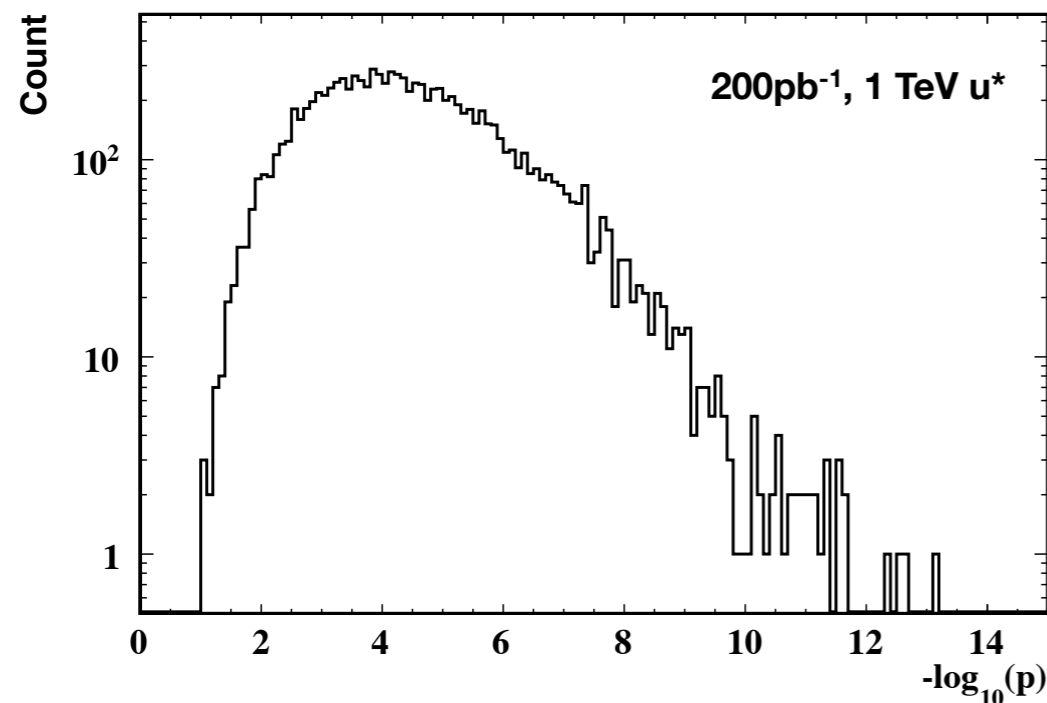$$p_{\text{likely}} = \frac{1}{\sum_{m=0}^{N} B(m)} \sum_{n=0}^{N} B(n)M(n)$$

- 3σ (evidence) and 5σ (discovery) limits are then determined by finding the p-value for which 0.14% and 2.87×10^{-5}% of the experiments have the expected p-value or less (use 100k bins)

200pb^{-1} example with 1B experiments, and coarse binning
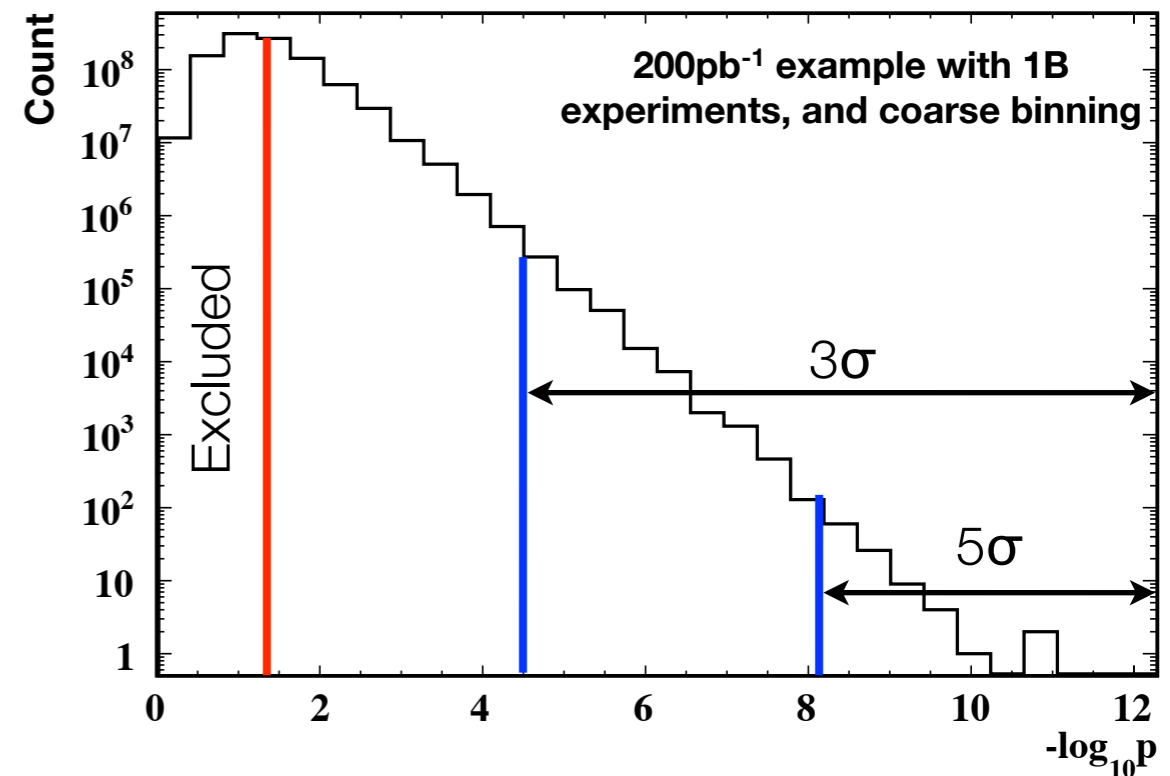
Search reach determination

- With the ability to compute 3σ and 5σ p-value limits for BG exclusion, a method is required to calculate the most likely p-value a signal + BG experiment will have for a given luminosity
- Throw 10k pseudo-experiments from the S+BG PDF
  - Run the BG fit, determine minimum p-value found, and histogram them
- Construct the weighted mean, as before, to determine the most-likely p-value for the luminosity and mass
- Plot as a function of luminosity; intersection with BG curves show evidence / discovery potential

200pb^{-1}, 1 TeV u*

Legend:
- 3σ background
- 5σ background
- 1 TeV u*
- 1.25 TeV u*
- 1.5 TeV u*
- 1.75 TeV u*

Measured systematic uncertainties

Table 32: Combination of systematic uncertainties to maximise search reach

| Channel	p.d.f. uncert.	p.d.f. choice	μ scale	Ele. ID	Cal. & Align.	Combination
Di-boson	−5%	−4.5%	−2%	−5%	−2.5%	−11%
$t\bar{t}$ + Jets	−5%	−5.5%	−13%	−5%	−2.5%	−18%
W + Jets	−5%	−5%	−2%	−5%	−2.5%	−12%
Z + Jets	−5%	−3.5%	−6%	−5%	−2.5%	−12%
γ + Jets	−5%	−10%	−6%	−5%	−2.5%	−17%
u* (1 TeV)	+3%	+3%	+10%	0%	0%	+12%
u* (1.25 TeV)	+4.8%	+4%	+10%	0%	0%	+13%
u* (1.5 TeV)	+6.5%	+4.5%	+10%	0%	0%	+15%
u* (1.75 TeV)	+8.3%	+4.5%	+10%	0%	0%	+16%
u* (2 TeV)	+10%	+4.5%	+10%	0%	0%	+18%

Table 33: Combination of systematic uncertainties to minimise search reach

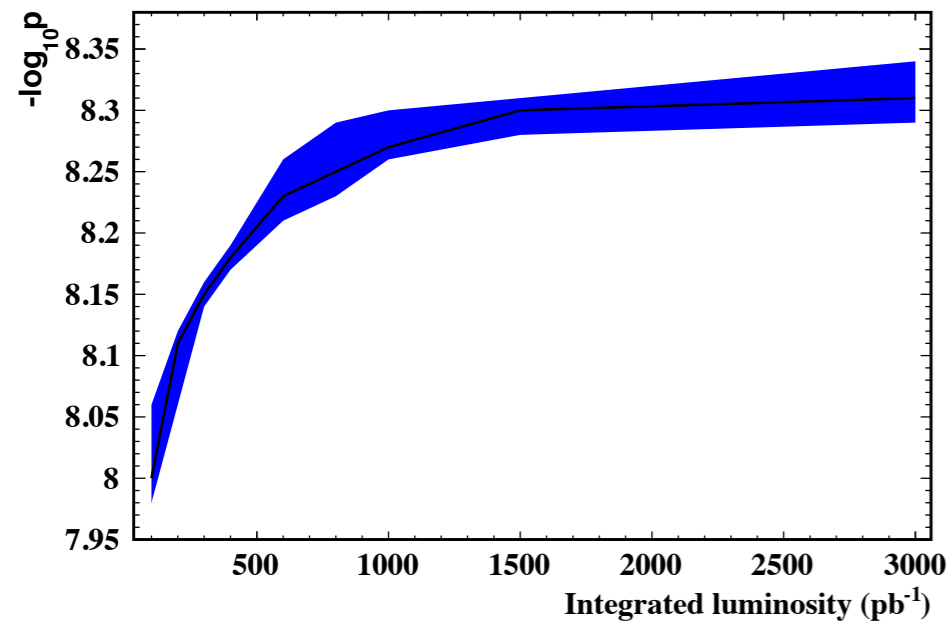| Channel	p.d.f. uncert.	p.d.f. choice	μ scale	Ele. ID	Cal. & Align.	Combination
Di-boson	+5%	+4.5%	+2%	0%	0%	+9%
$t\bar{t}$ + Jets	+5%	+5.5%	+14%	0%	0%	+18%
W + Jets	+5%	+5%	+2%	0%	0%	+10%
Z + Jets	+5%	+3.5%	+4%	0%	0%	+9%
γ + Jets	+5%	+10%	+1%	0%	0%	+15%
u* (1 TeV)	−3%	−3%	−8%	−5%	−2.5%	−11%
u* (1.25 TeV)	−4.8%	−4%	−7%	−5%	−2.5%	−13%
u* (1.5 TeV)	−6.5%	−4.5%	−8%	−5%	−2.5%	−15%
u* (1.75 TeV)	−8.3%	−4.5%	−8%	−5%	−2.5%	−16%
u* (2 TeV)	−10%	−4.5%	−8%	−5%	−2.5%	−17%

Final search reach with systematics
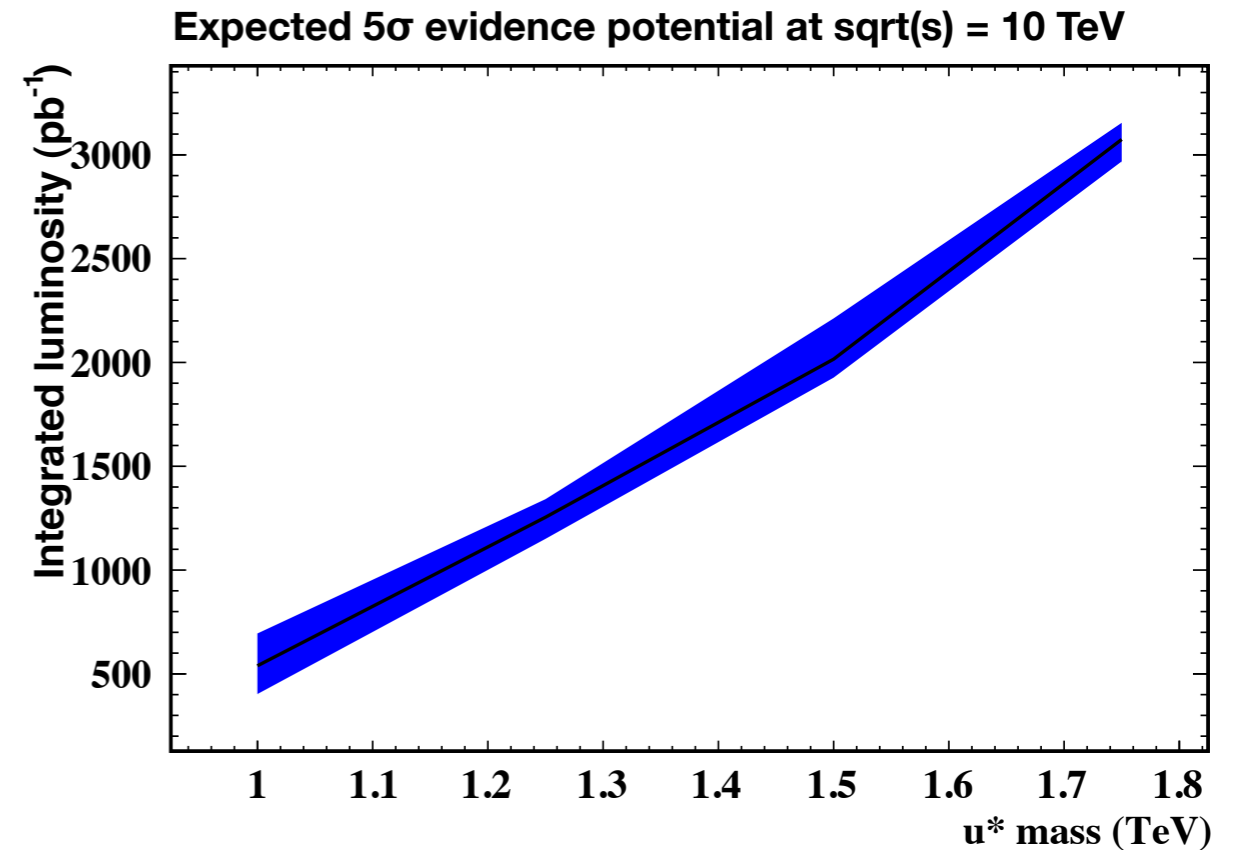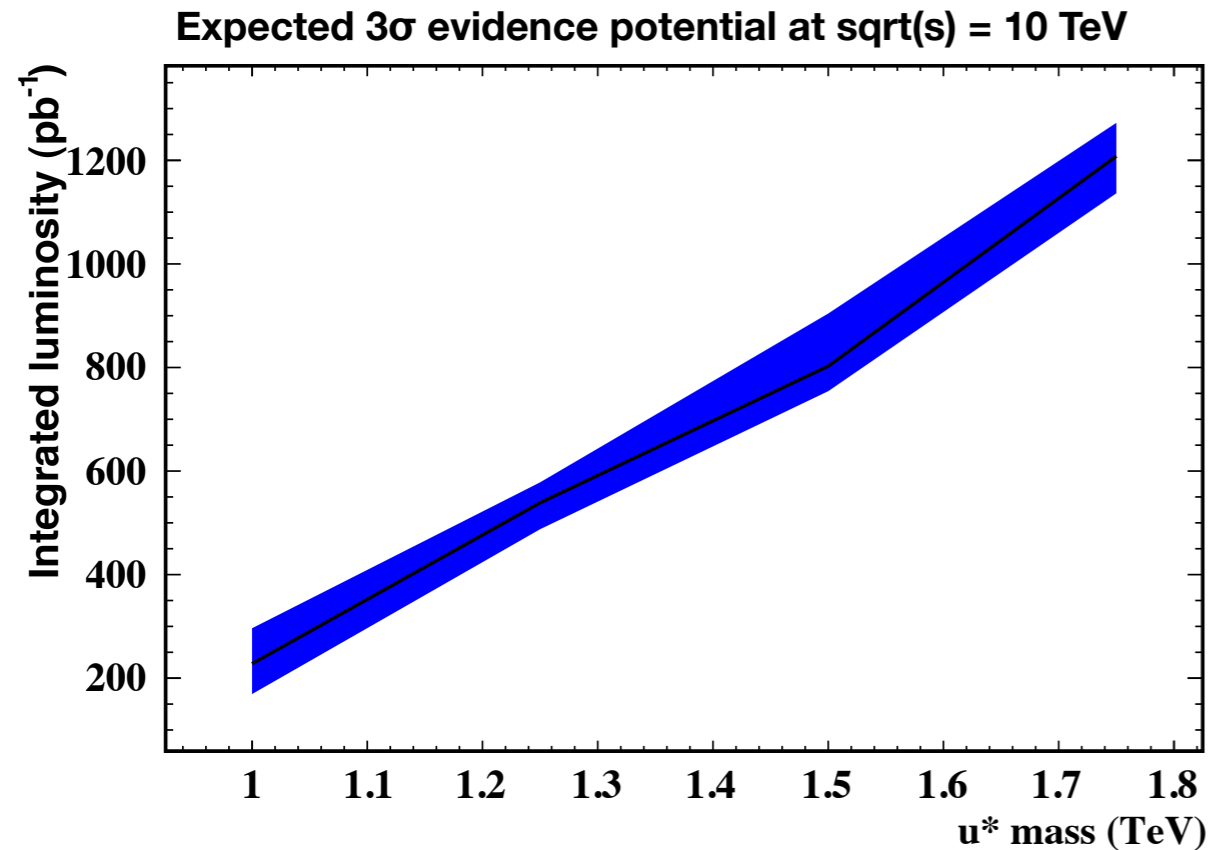

3σ background rejection p-value criteria
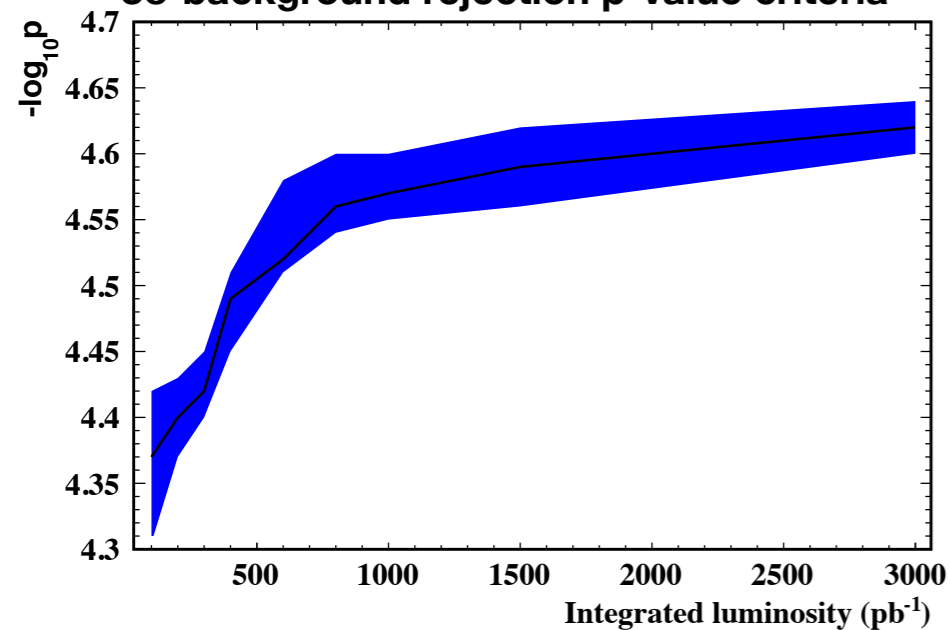
5σ background rejection p-value criteria

Final search reach determination plot

5σ background
3σ background
1 TeV u*
1.25 TeV u*
1.5 TeV u*
1.75 TeV u*

Final search reach with systematics

Expected 3σ evidence potential at sqrt(s) = 10 TeV

Expected 5σ evidence potential at sqrt(s) = 10 TeV
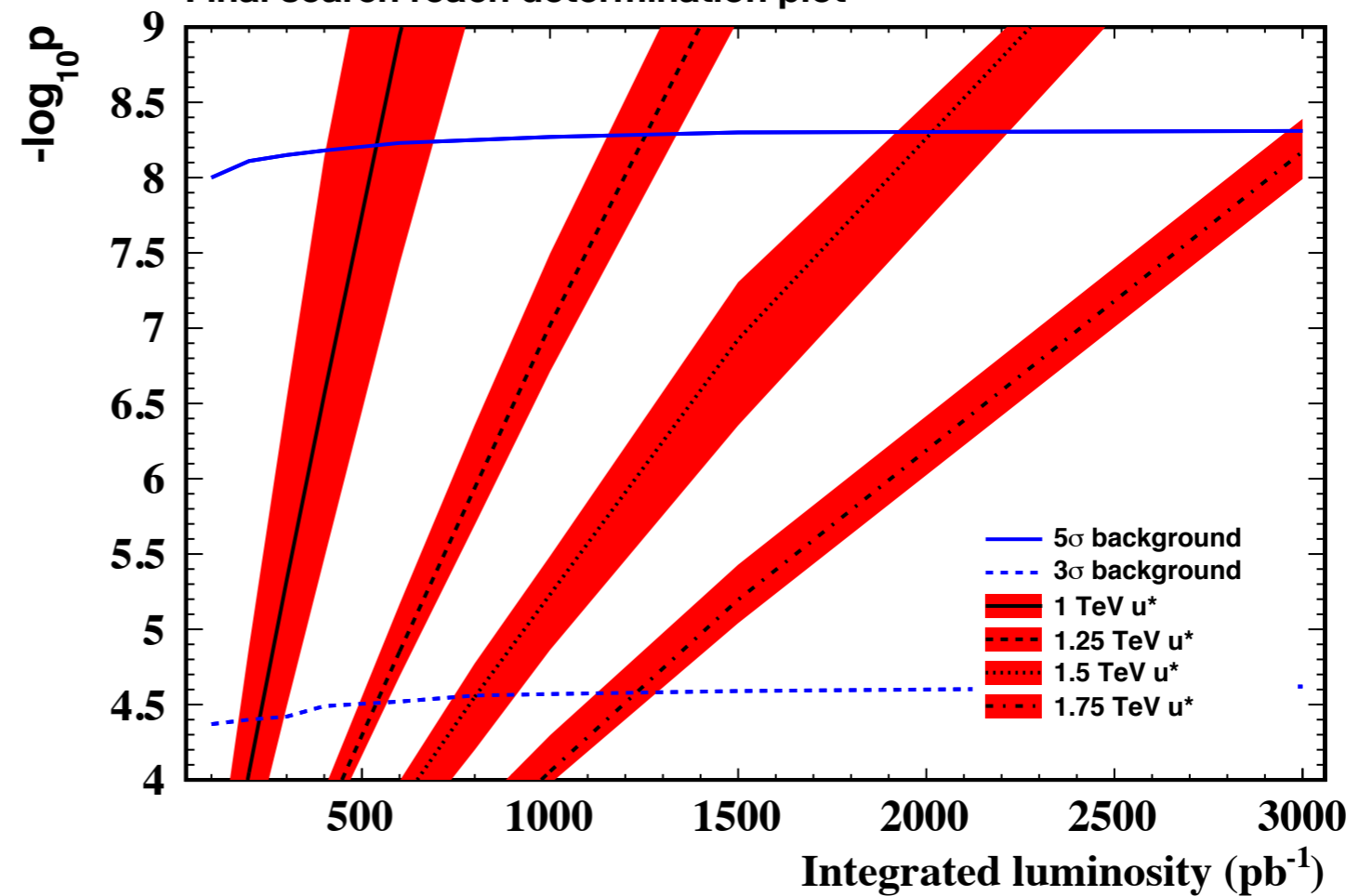
- Plots determined from intersection of signal and background curves on previous slide

- **For 1 TeV u* with input model assumptions, 3σ evidence could be found with 200 pb^{-1} of integrated luminosity, and 5σ with 500 pb^{-1} at sqrt(s) = 10 TeV**
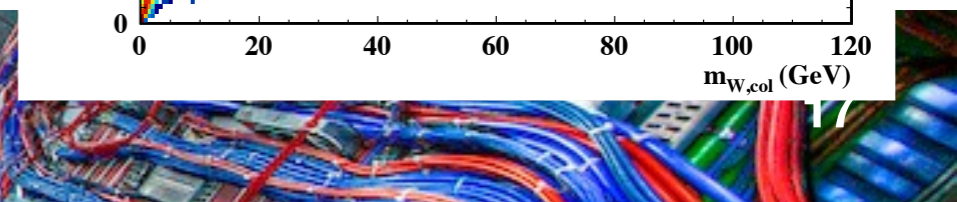
Bonus slide - Boosted $W^{\pm} \to e^{\pm}\nu$

- Missing E_t will be strongly correlated with the boost direction. Reconstruct the neutrino three-vector in the collinear approximation

- The electron-neutrino invariant mass is then plotted against the opening angle in phi between the electron and missing E_t

$$\vec{p}_{\nu_e} = \left(\not{E}_x, \not{E}_y, \frac{\sqrt{\not{E}_x^2 + \not{E}_y^2}}{\sqrt{p_{x,e}^2 + p_{y,e}^2}} p_{z,e} \right)$$
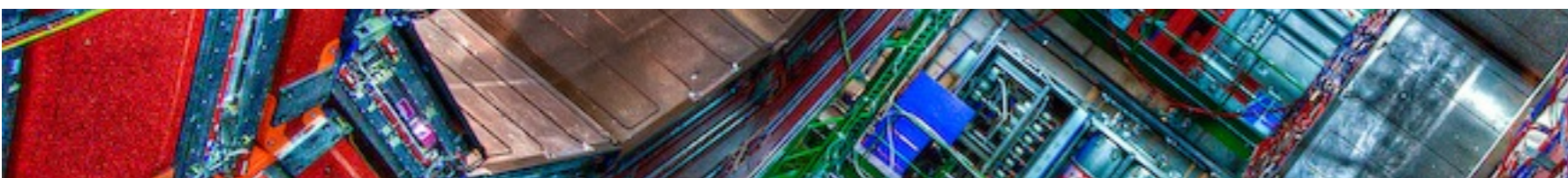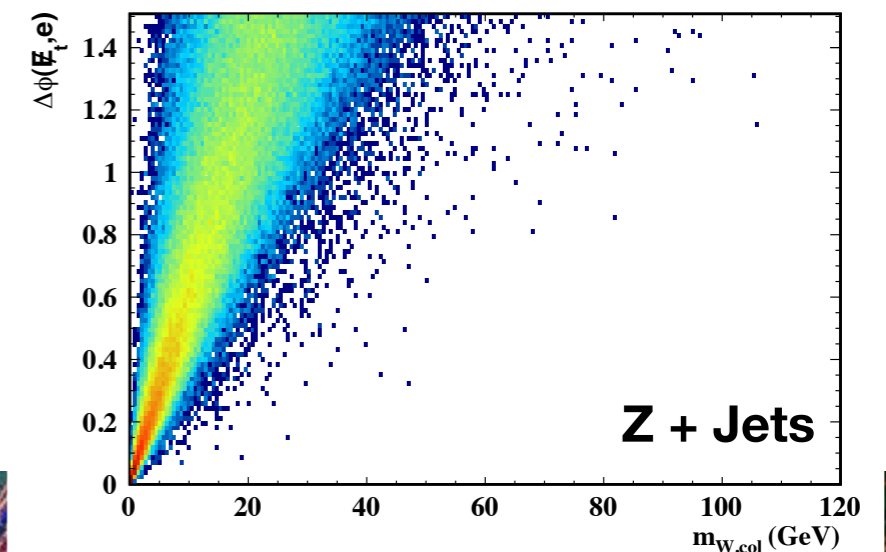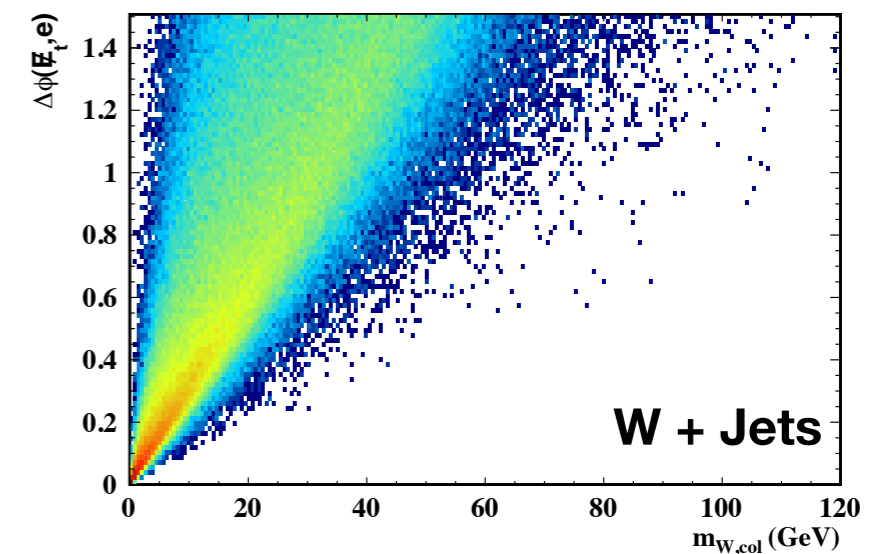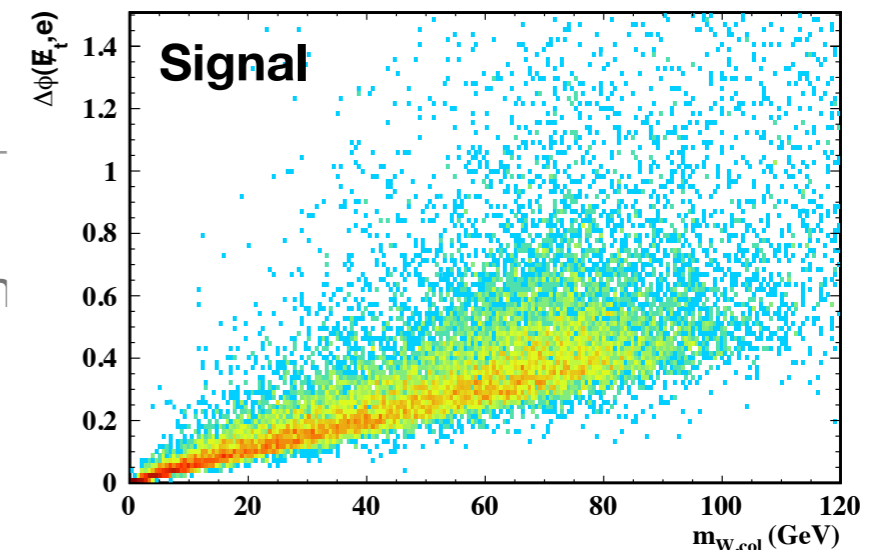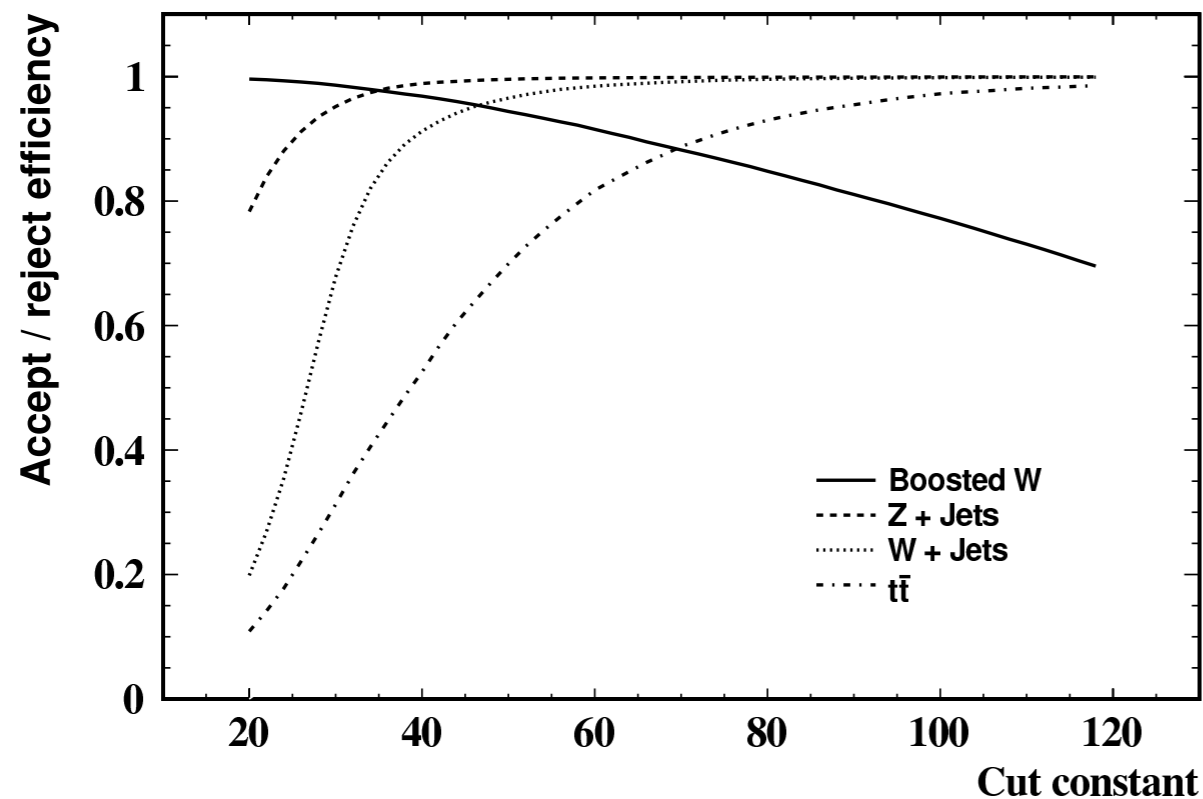
BACKUP

Electron selection

- Aim to follow the HEEP high energy electron selection as closely as possible
  - Shared code, efficiencies, commissioning etc

Variable	Barrel	Endcap
E_t	$> 25\,\mathrm{GeV}$	$> 25\,\mathrm{GeV}$
$|\eta_{SC}|$	< 1.422	$1.560 < |\eta_{SC}| < 2.5$
$|\Delta\eta_{\mathrm{in}}|$	< 0.005	< 0.007
$|\Delta\phi_{\mathrm{in}}|$	$< 0.09\,\mathrm{rad}$	$< 0.09\,\mathrm{rad}$
H/E	< 0.05	< 0.05
$\sigma_{i\eta i\eta}$	n/a	< 0.0275
$E^{2\times5}/E^{5\times5}$	> 0.94 OR $E^{1\times5}/E^{5\times5} > 0.83$	n/a
EM + Had Depth 1 Isolation	$< 3 + 0.002 E_t\,\mathrm{GeV}$	$< 5.5\,\mathrm{GeV}$ for $E_t < 50\,\mathrm{GeV}$ else $< 5.5 + 0.05(E_t - 50)\,\mathrm{GeV}$
Had Depth 2 Isolation	n/a	$< 0.5\,\mathrm{GeV}$
Track p_t Isolation	$< 7.5\,\mathrm{GeV}$	$< 15\,\mathrm{GeV}$

- The performance of the cuts was measured by defining the event selection efficiency as

$$\epsilon_E = \frac{\text{Events with} \geq 2 \text{ fiducial electrons passing HEEP cuts}}{\text{Total number of events with} \geq 2 \text{ fiducial electrons}},$$

- Each individual efficiency was measured by matching reconstructed electrons to MC truth (ΔR < 0.1), and measuring each cut individually

Jet backgrounds with the fake rate method

- Make use of the fact that events with one selected electron are more likely than those with two
- Two stage method:
    - Use a sample unbiased with respect to the signal selection to measure the probability that a jet fakes a signal electron
    - Apply this probability to all the jets in an event with only one reconstructed signal electron to estimate the background
- The unbiased sample is selected with jet triggers, taken from 1E31 v0.6 menu. To make use of the available QCD samples, a pseudo-HLT reweighting scheme was used. Each event was scaled by the inverse of the trigger prescale, to allow all events to be used
- 'Jets' are defined as loosely selected GSF electrons. This removes the requirement of a jet scale correction step. The ΔR cut further removes any trigger bias

Loose electron selection

Cut	Value
$\Delta R(\text{Trig.}, \text{Cand.})$	> 0.2
$|\eta|$	< 2.5
E_t	$> 20\,\text{GeV}$
Had / EM	< 0.2

Tight electron selection

Cut	Value
$\Delta R(\text{Trig.}, \text{Cand.})$	> 0.2
$|\eta|$	< 2.5
E_t	$> 20\,\text{GeV}$
Modified HEEP selection cuts	Must pass

Jet triggers used (1E31 v0.6 menu)

Trigger	L1 Prescale	HLT Prescale	Total Prescale
Jet30	1000	5	5000
Jet50	100	2	200
Jet80	10	2	20
Jet110	1	1	1

Jet backgrounds with the fake rate method

$$F(E_t) = \frac{\sum \text{Jet objects passing tight electron selection}}{\sum \text{Jet objects passing loose electron selection}}.$$
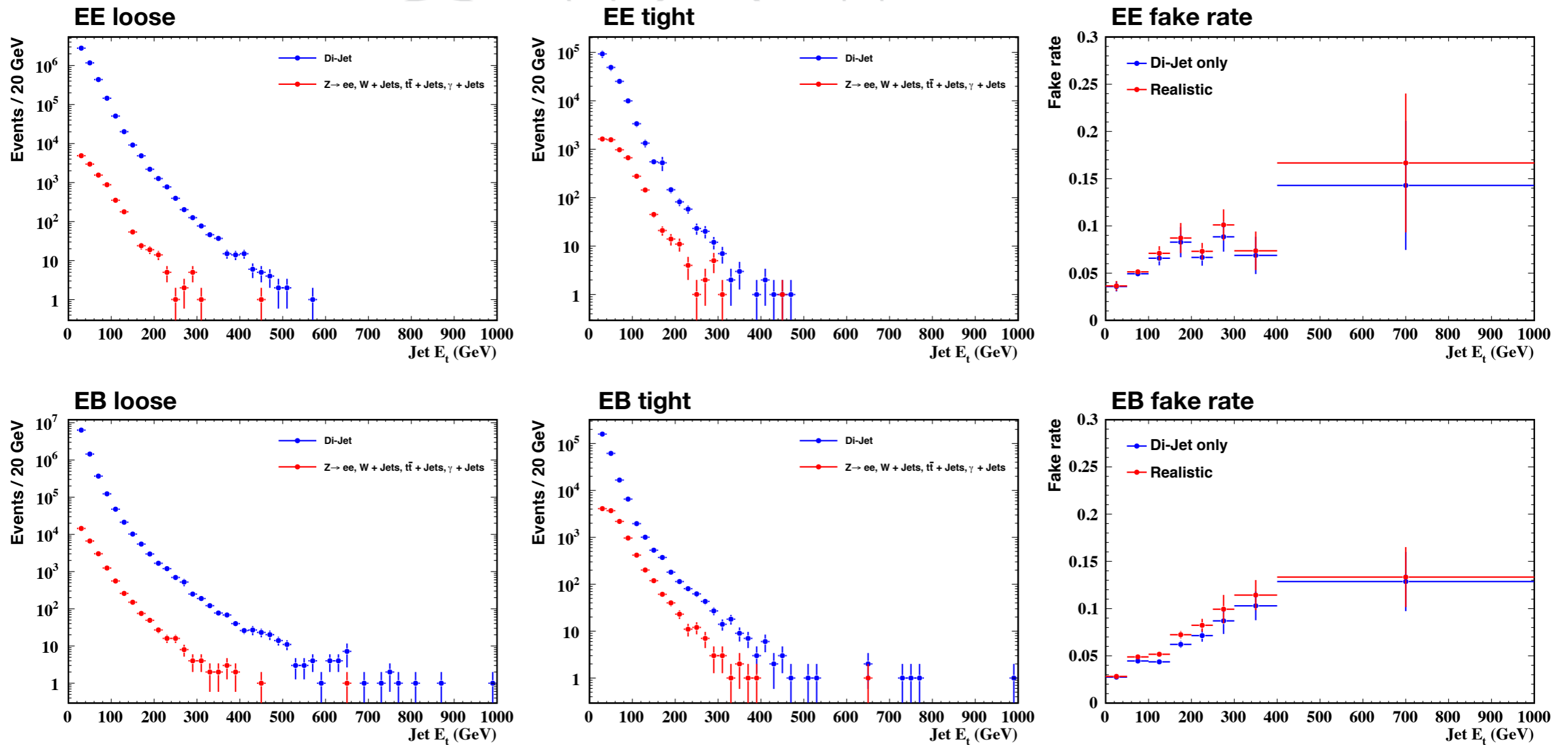
Jet backgrounds with the fake rate method

- For all events in the signal trigger set, the highest Et electron is taken as the triggered lepton, and must pass the tight selection criteria
- All other objects passing the loose electron selection, but not the tight selection (to remove signal bias), are histogrammed by trigger-fake pair p_t, weighted according to the fake rate given by the loose object
- Pairs are excluded if they lie in the range 70 < M < 110 GeV to further remove signal contamination

$$N(p_t) = \sum_{\text{loose}} w_e \frac{F(E_t)}{1 - F(E_t)},$$

Jet fake rate with QCD samples included
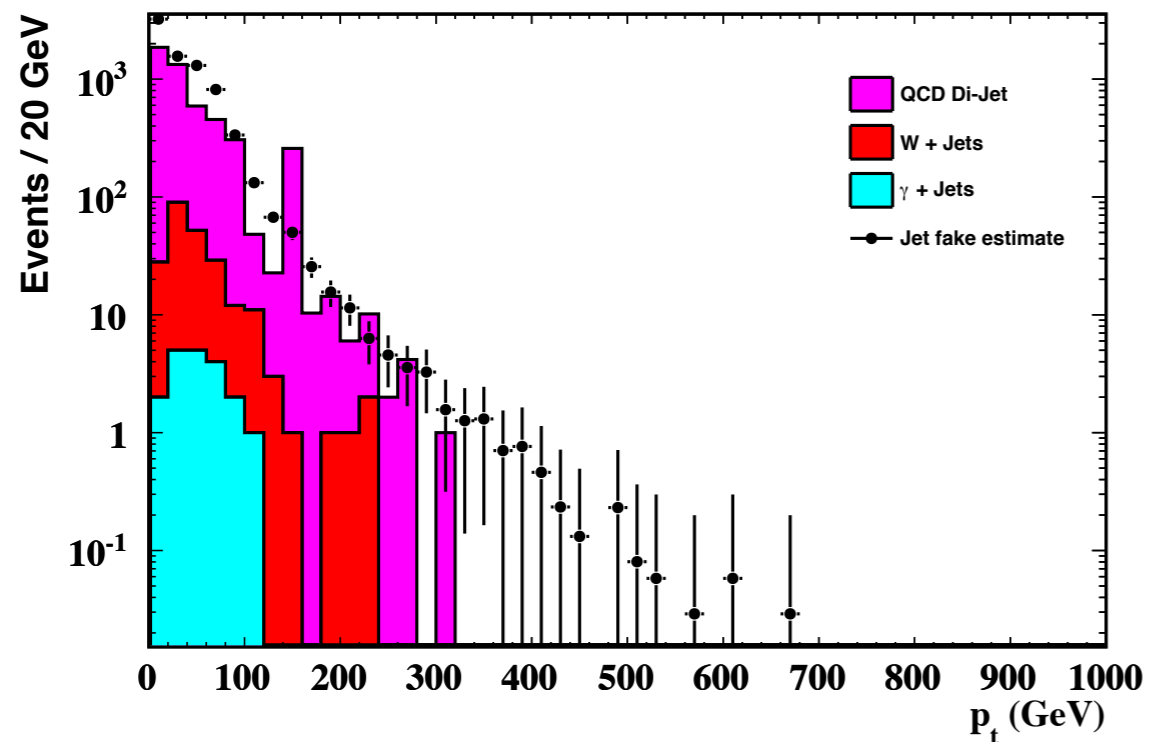
Jet fake rate with realistic (non-QCD) dataset
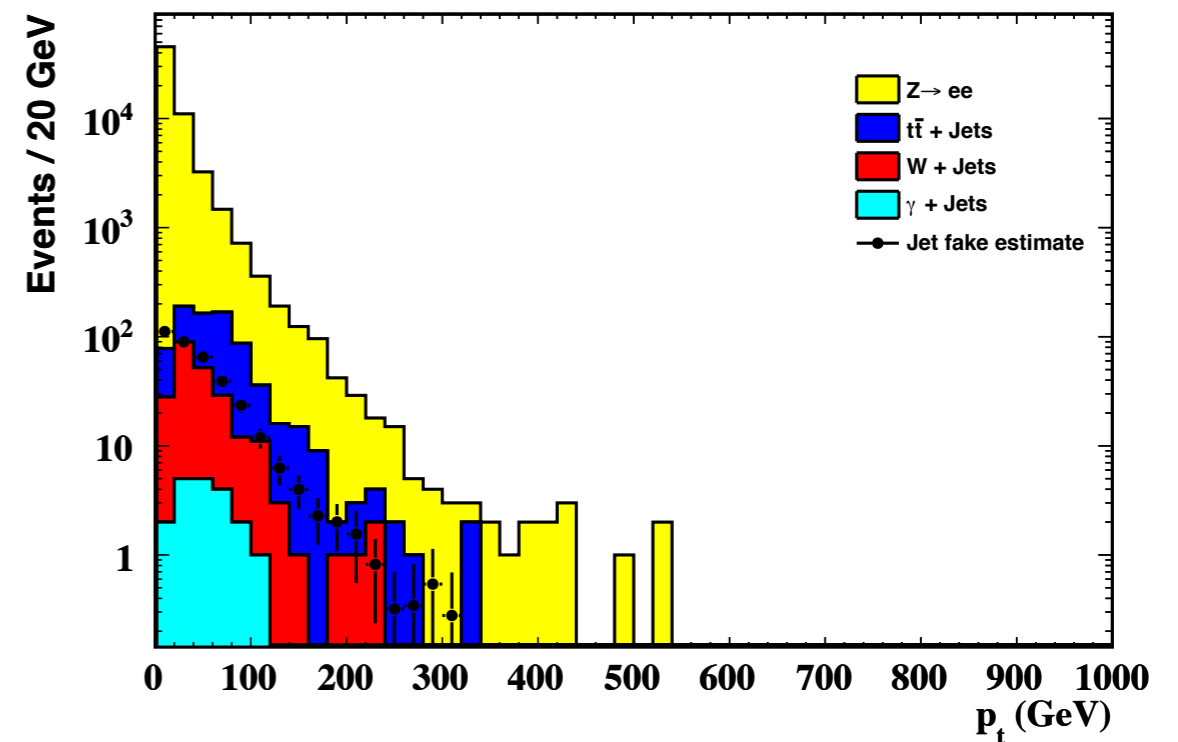
tt with the b-tagging method

- The b-tagging method is robust against b-tagging commissioning, and can be applied on top of the existing event selection

- The observed number of events with exactly one and two b tags are given by n_1 and n_2. These are related to the actual number of events (N_1, N_2) within detector acceptance by

$$
\begin{aligned}
n_1 &= N_1 \epsilon_b + 2 N_2 \epsilon_b \left(1 - \epsilon_b\right) \\
n_2 &= N_2 \epsilon_b^2,
\end{aligned}
$$

- where ϵ_b is the b-tagging efficiency. N_1 and N_2 are related to the true number of tt events by

$$
\begin{aligned}
N_1 &= N A_1 \\
N_2 &= N A_2
\end{aligned}
$$

- Where A_1 (A_2) is the geometric acceptance for events containing exactly 1 (2) b jets from a tt event. From these expressions, and A_1 (A_2) measured from MC, ϵ_b can be determined as

$$
\epsilon_b = \frac{A_1 / A_2 + 2}{n_1 / n_2 + 2}
$$

- With this measurements performed, N can be calculated from either the n_1 or n_2 samples as:

$$
\begin{aligned}
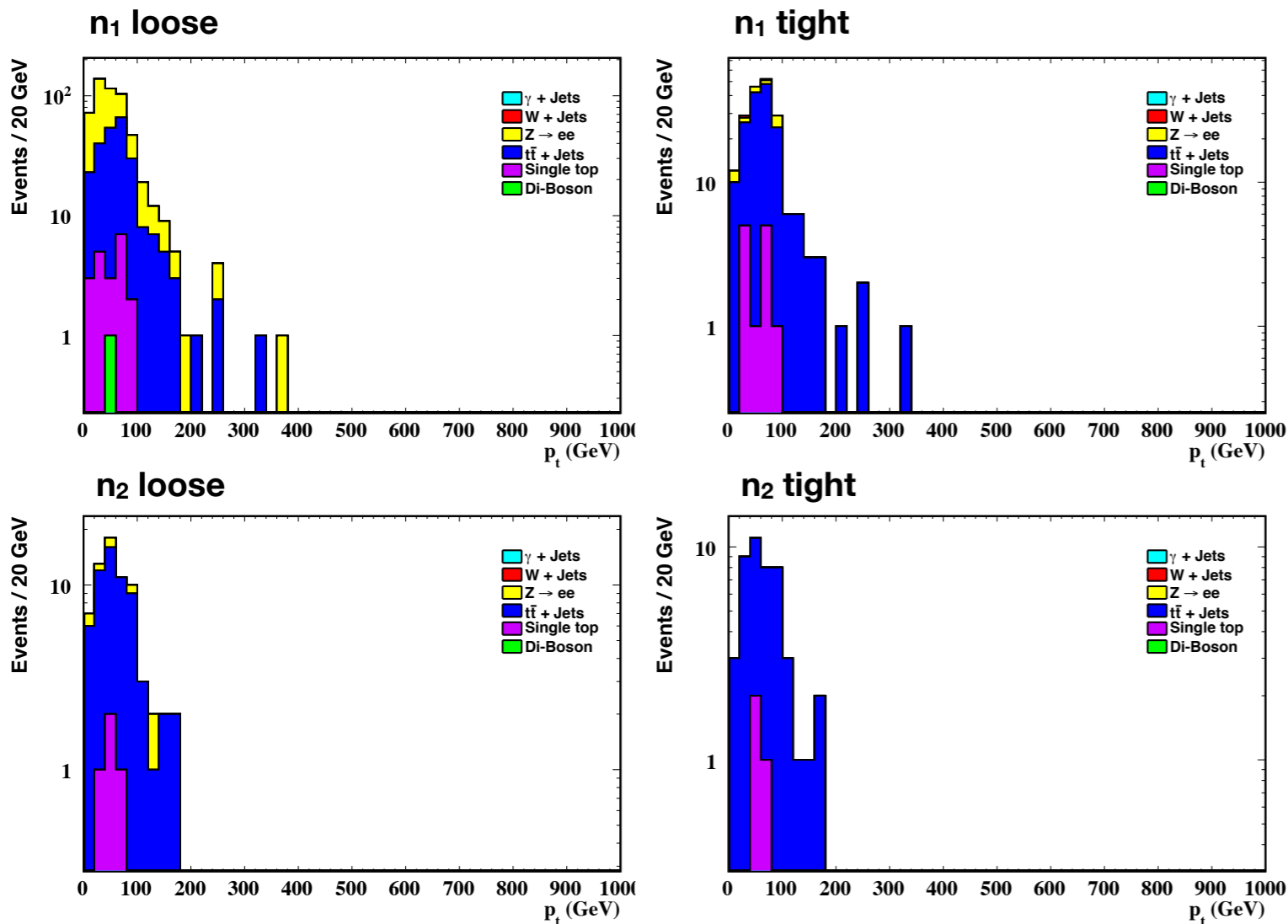N &= \frac{n_1}{\epsilon_b \left(A_1 + 2 A_2 (1 - \epsilon_b)\right)}, \\
N &= \frac{n_2}{A_2 \epsilon_b^2}.
\end{aligned}
$$

- To ensure the samples are of equivalent purity, a tight selection is defined which vetoes events with $70 < M < 110$ GeV

tt with the b-tagging method
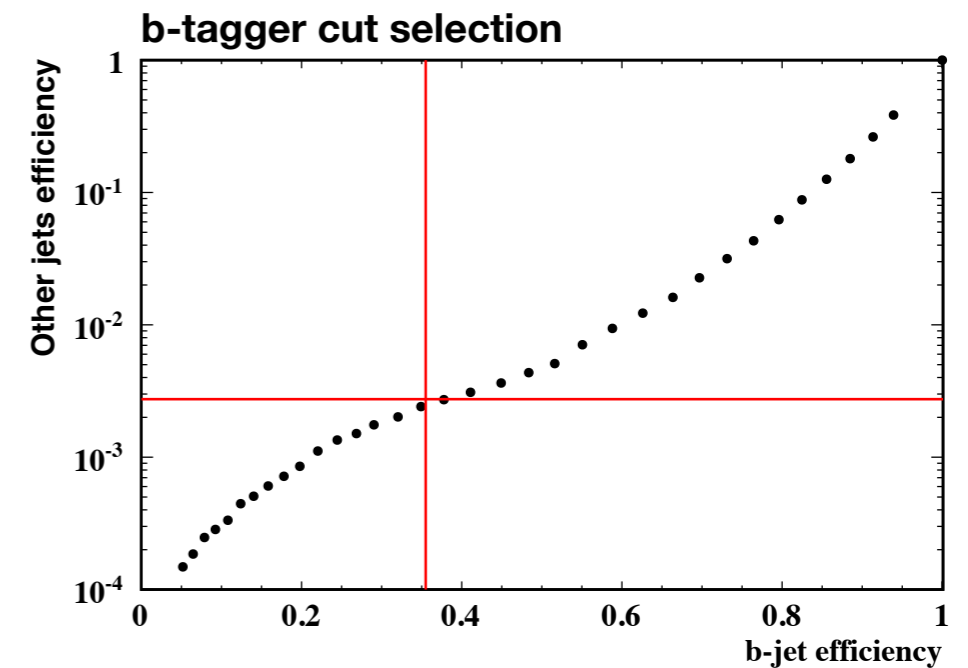

n₁ loose

n₁ tight

n₂ loose

n₂ tight

Legend for each plot:
- γ + Jets
- W + Jets
- Z → ee
- $t\bar{t}$ + Jets
- Single top
- Di-Boson

Jet selection criteria

Jet Algorithm	iterativeCone5CaloJets
Jet E_t	> 20 GeV
Jet $|\eta|$	< 2.4
B Discriminant	jetBProbabilityJetTags
Discriminant cut	> 4.0

b-tagger cut selection

- From MC, $A_1 = 0.146 \pm 0.005$, $A_2 = 0.79 \pm 0.01$

tt with the b-tagging method

Complete n₁, n₂ estimation

Legend:
- $Z \rightarrow ee$
- W + Jets
- γ + Jets
- Di-Boson
- Single top
- $t\bar{t}$ + Jets
- N1 Estimate
- N2 Cross-Check
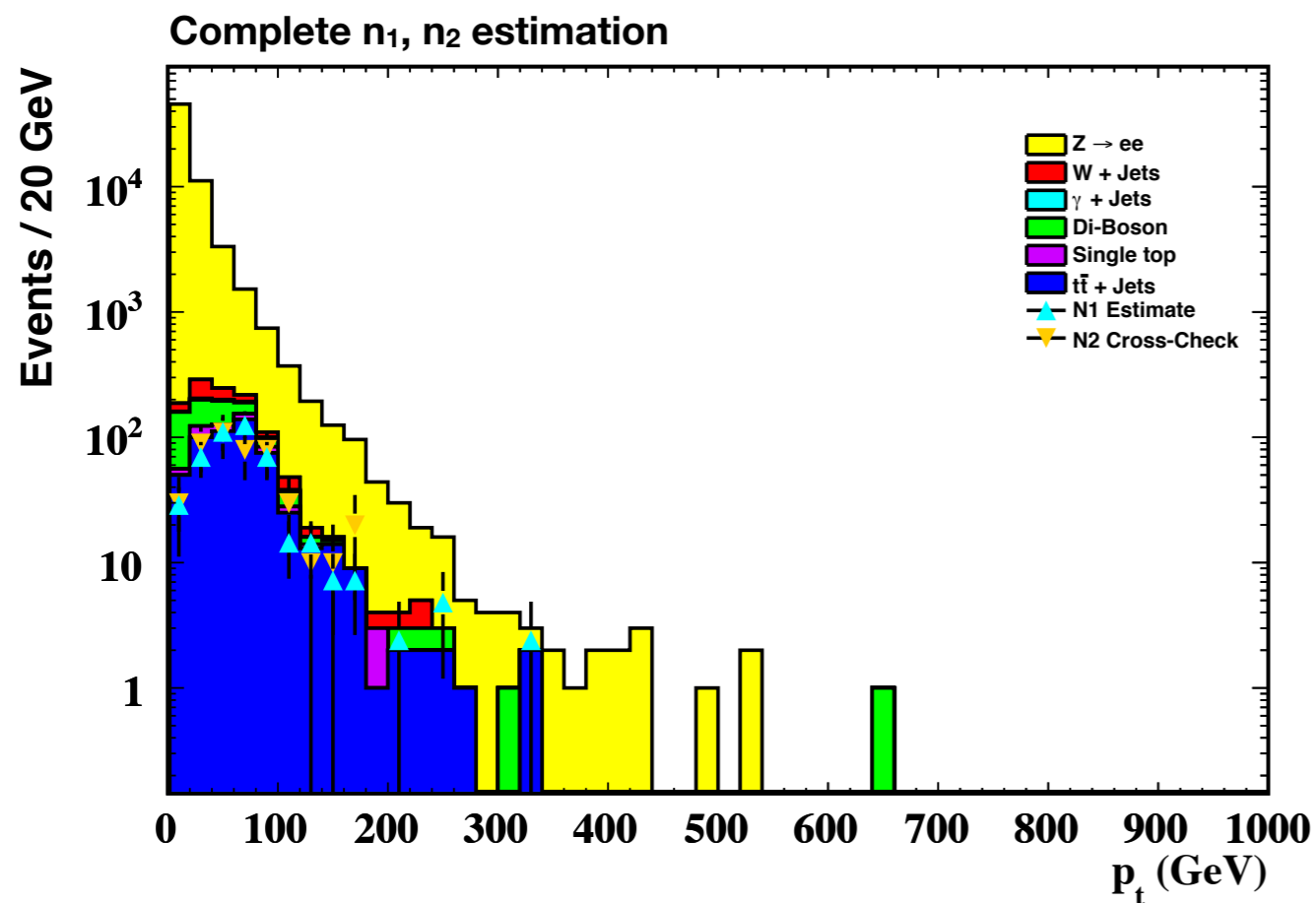
Measured b-tagging efficiency

Data sample	Efficiency
$t\bar{t}$ only	0.377 ± 0.058
$t\bar{t}$ only (tight)	0.382 ± 0.067
Realistic (tight)	0.356 ± 0.062

Total number of estimated tt events

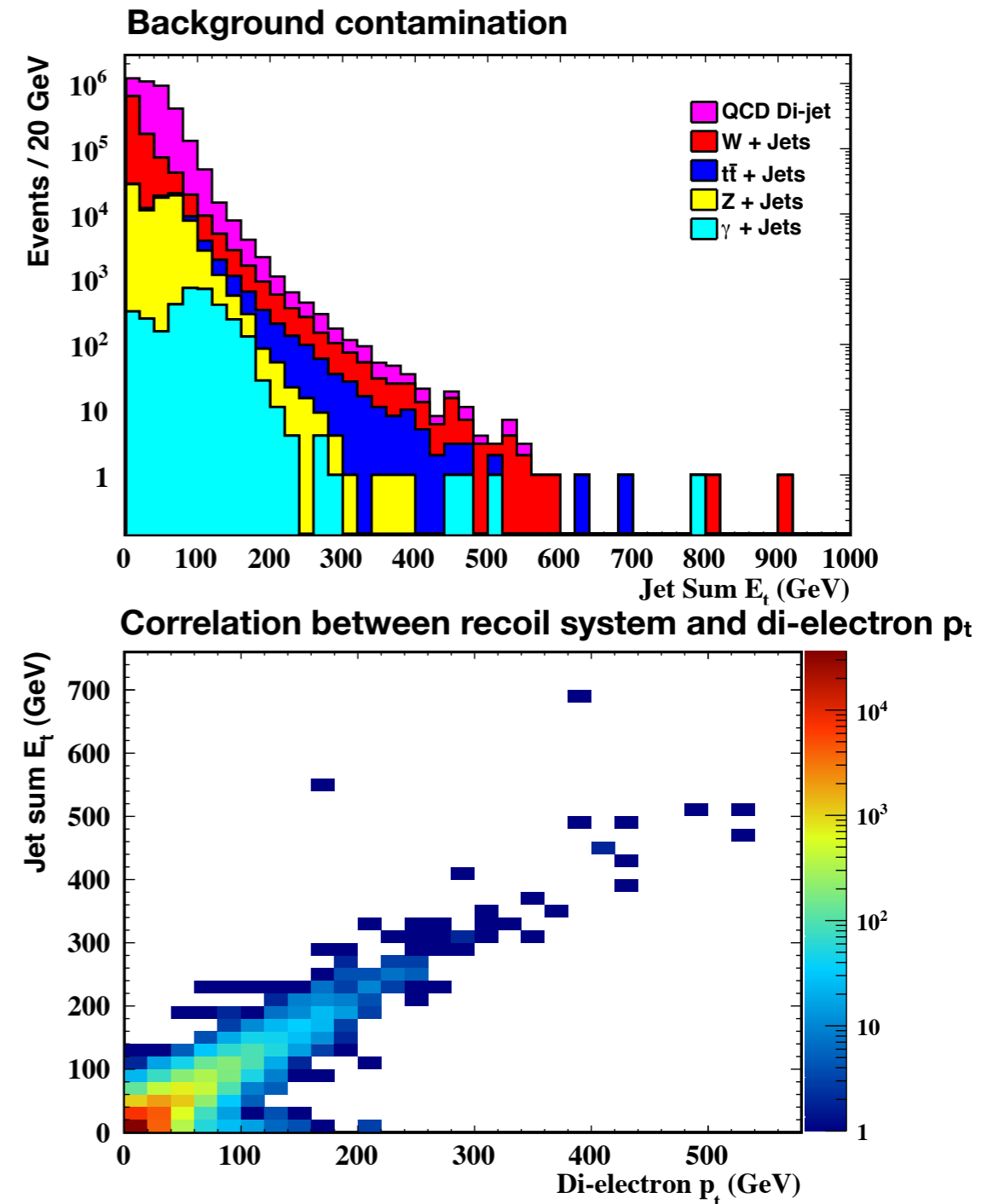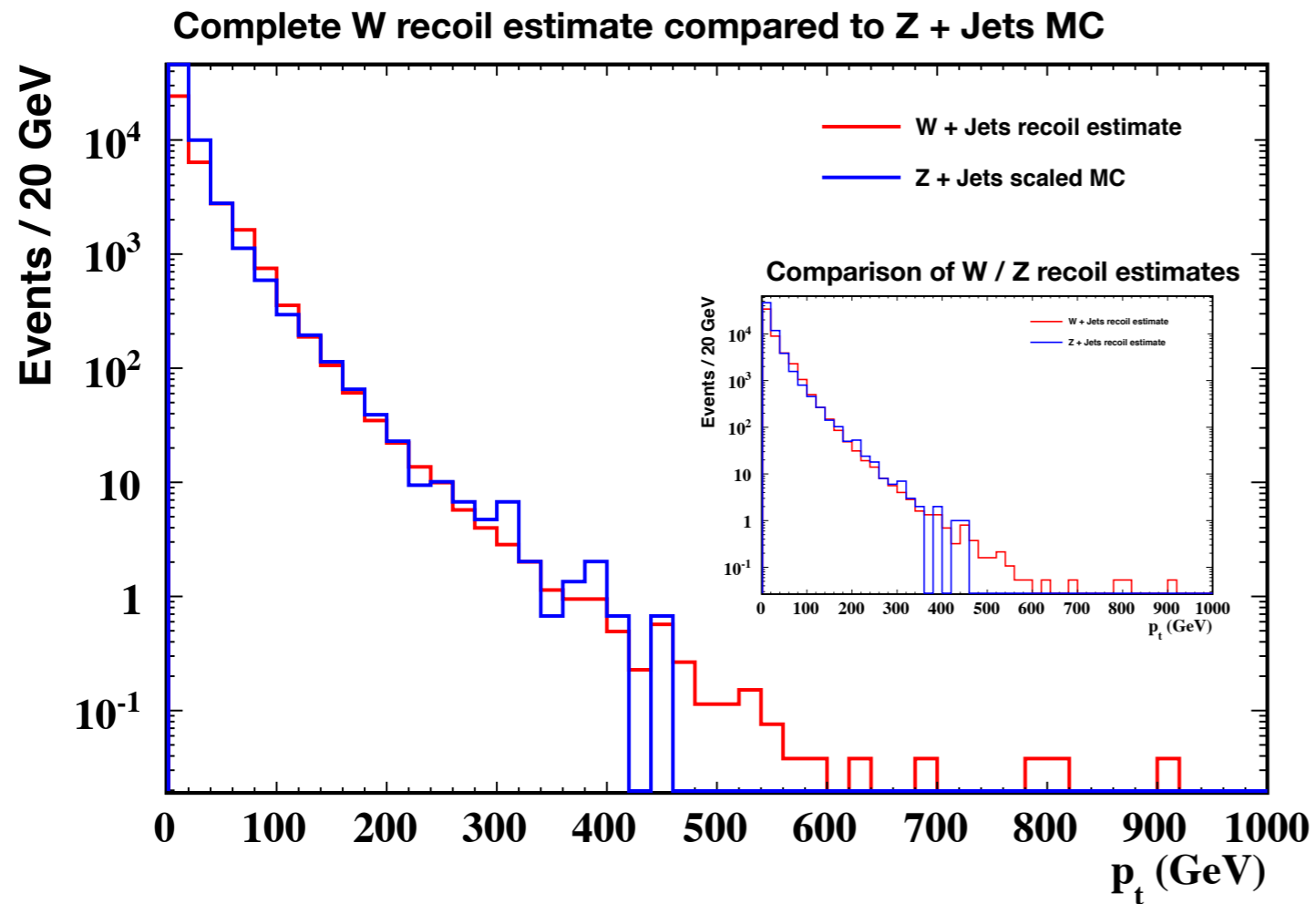| Data sample	Events with $70 < M_{ee} < 110 (GeV)$
$t\bar{t}$ true	378
$t\bar{t}$ only (n_1)	374 ± 95
$t\bar{t}$ only (n_2)	381 ± 101
Realistic (n_1)	428 ± 119
Realistic (n_2)	428 ± 124

Z→e⁺e⁻ with W hadronic recoil

- The irreducible Z→ee background can be estimated from MC, but this requires complete understanding of the simulation in the region where new physics is expected
- In the kinematic region above the W and Z masses, the W and Z can be considered to have identical production kinematics. The W cross section is ~3 times that of the Z, and the branching ratio W→ev is ~3 times that of Z→ee, a factor of 10 more W than Z events are expected
- By computing the p_t of the hadronic recoil system, the p_t of the W can be determined, and therefore an estimate of the Z p_t spectrum can be computed, given a suitable normalisation (taken to be the region 150 - 250 GeV to minimise the QCD di-jet influence)
- Event are selected with one well isolated (passing the full HEEP selection) electron
- The four-vectors of all jets which are separated ($\Delta R > 0.4$) from the electron, and with loose selection cuts ($E_t > 20$ GeV, $|\eta| < 2.5$) are summed, and the p_t of the resulting four-vector determined
- As a cross-check, the hadronic recoil p_t of events containing two selected electrons can be computed. This also allows a check of the jet energy scale, and an appropriate correction to be derived if required
- New physics coupling to Z would also be expected to couple to W - may force use of MC. However, some discriminating power is available using W mass and electron / neutrino phi separation
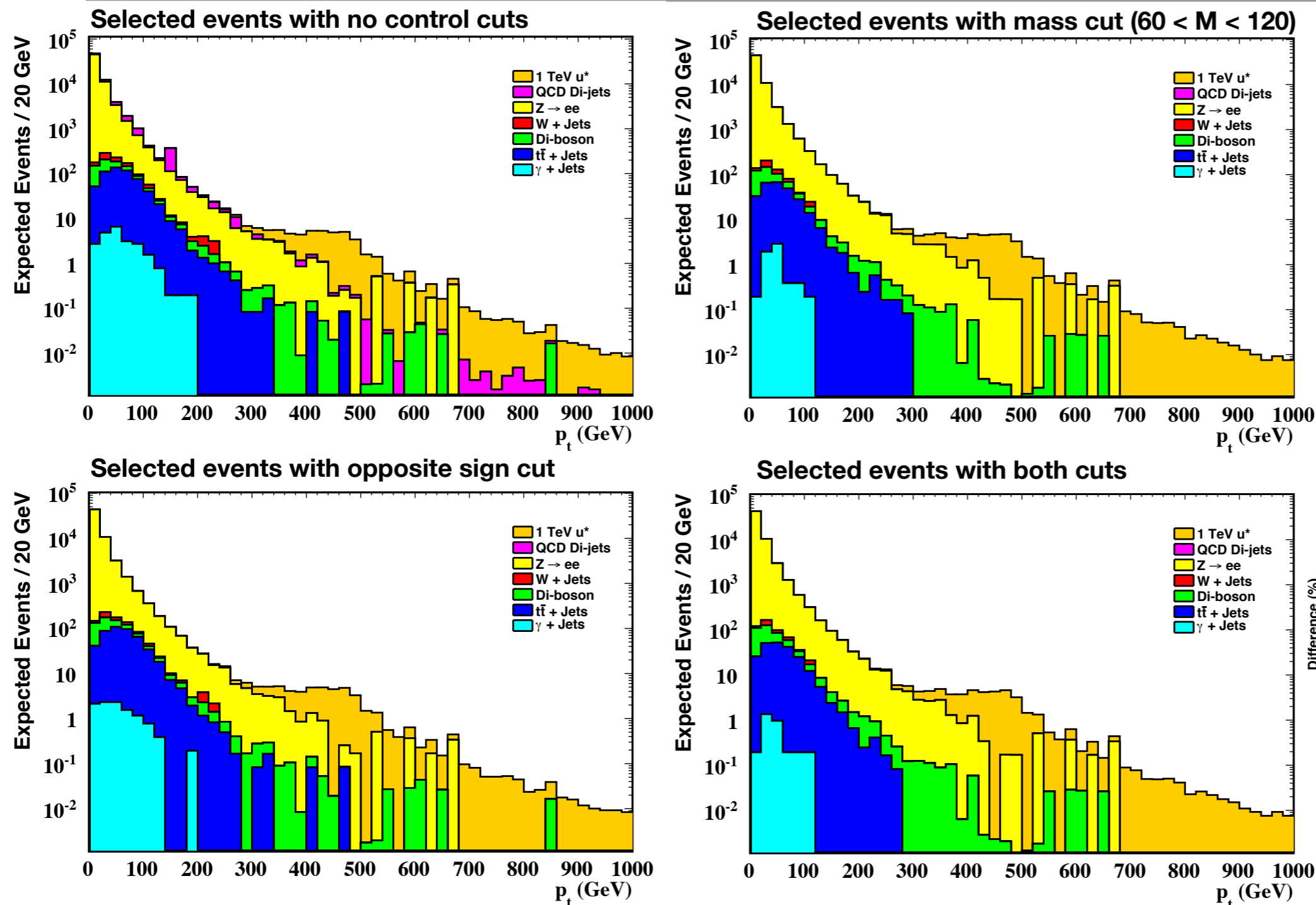
Z→e+e− with W hadronic recoil

Complete W recoil estimate compared to Z + Jets MC

- W + Jets recoil estimate
- Z + Jets scaled MC

Comparison of W / Z recoil estimates

- W + Jets recoil estimate
- Z + Jets recoil estimate

Background contamination

- QCD Di-jet
- W + Jets
- $t\bar{t}$ + Jets
- Z + Jets
- γ + Jets

Correlation between recoil system and di-electron p_t

Control of backgrounds

Selected events with no control cuts

Selected events with mass cut (60 < M < 120)

Selected events with opposite sign cut

Selected events with both cuts
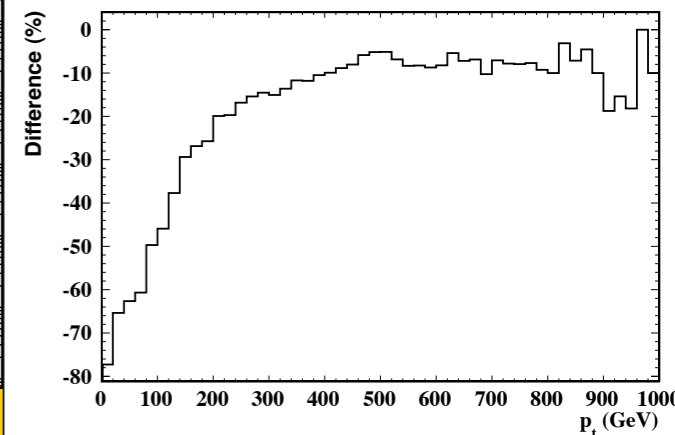

- Propose two cuts in case backgrounds are larger than expected
- Opposite sign leptons
- Pair invariant mass

- Not used in the analysis

### Signal reduction with both cuts

Discriminating signal / background W$^\pm$

- For a boosted W, the missing E$_t$ will be strongly correlated with the boost direction. Use this to reconstruct the neutrino three-vector in the collinear approximation

- The electron-neutrino invariant mass is then plotted against the opening angle in phi between the electron and missing E$_t$

$$\vec{p}_{\nu_e} = \left(\not{E}_x, \not{E}_y, \frac{\sqrt{\not{E}_x^2 + \not{E}_y^2}}{\sqrt{p_{x,e}^2 + p_{y,e}^2}} p_{z,e} \right)$$
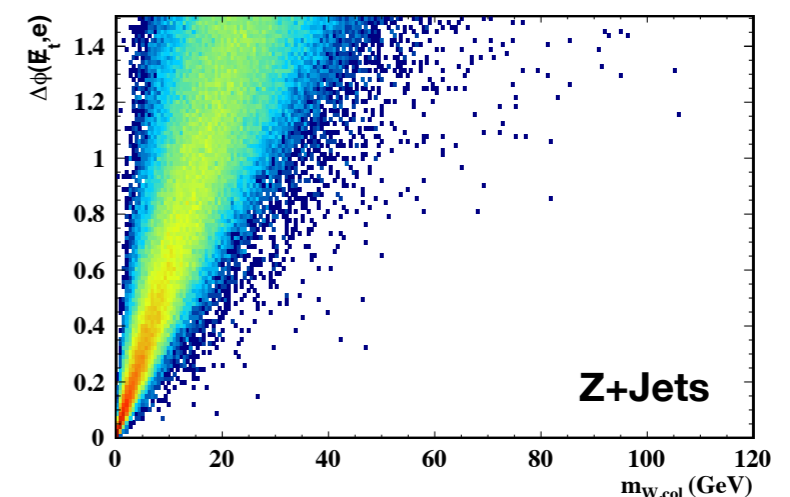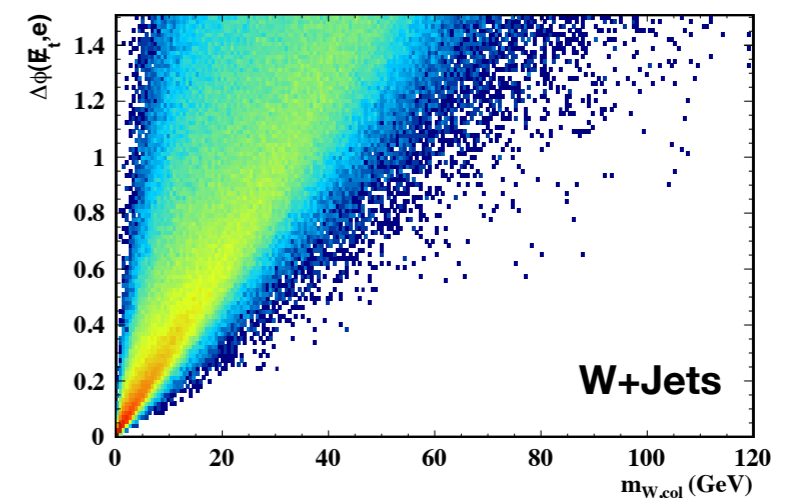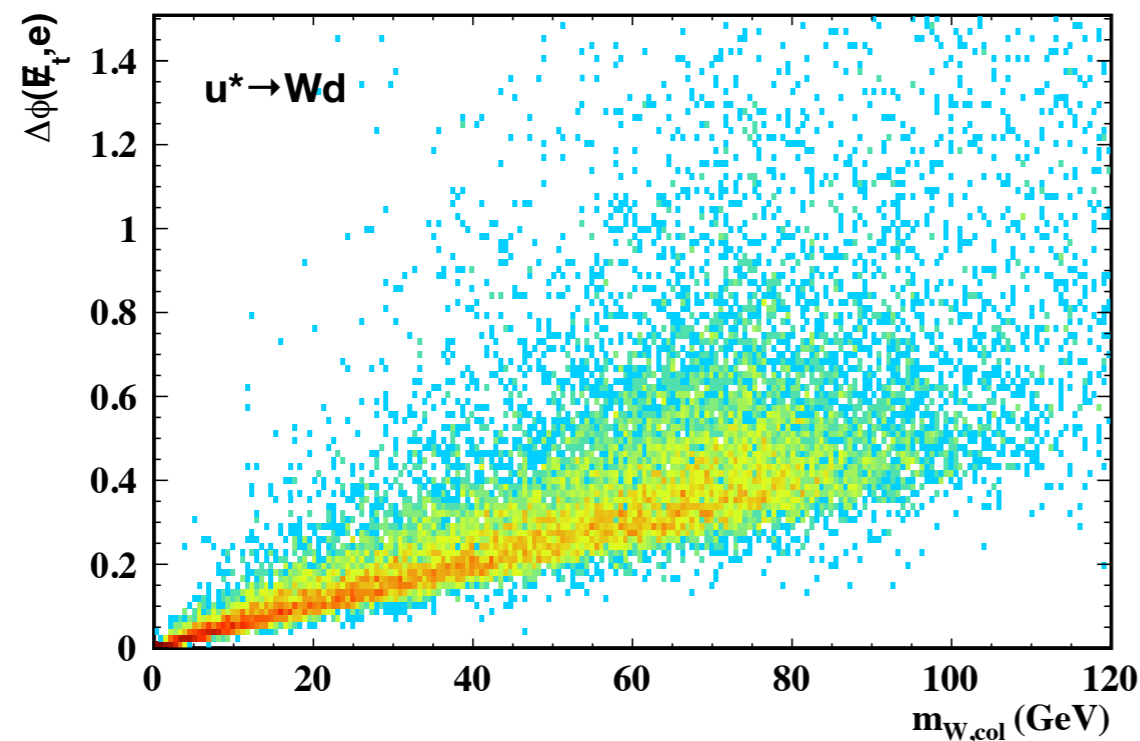
14 TeV analysis potential

- Signal + BG yields scaled for parton luminosities at 14 TeV
- Statistical tool re-run on scaled datasets
- Systematics not considered
- Shows well-behaved scaling with change in parton lumis

Relative 14 to 10 TeV parton luminosities

Search reach determination plot

Expected 3σ evidence potential at sqrt(s) = 14 TeV

Expected 5σ evidence potential at sqrt(s) = 14 TeV