W→eν cross-section measurement at CMS with the first $3pb^{-1}$ of pp collision data

Nikolaos Rompotis
Imperial College London

Introduction (I)

- In this presentation I will discuss only a part of my contribution to the recent CMS W→eν cross-section measurement

CMS ECAL:
data certification,
ECAL software

CMS EGAMMA:
Electron Id, Supercluster cleaning, electron efficiency, egamma skimming

CMS EWK GROUP:
Wev and Zee studies:
electron efficiency, signal extraction, event and electron selection, software and ntuple production

Other research work:
Stochastic thermostats: Phys. Rev. B**78**, 094305 (2008),
J. Phys.: Cond. Matt. **22,** 074205 (2010)
Scanning Tunneling Microscopy: Phys. Rev. B**78**, 165302 (2008)
Optics (Solitons in LHM): Phys. Rev. E**79**, 037601 (2009)

Introduction (II)

- Major CMS Responsibilities:
  - Contact for CMS Egamma Electron Selections (Oct 09 till present)
  - ECAL data certification expert (Aug 09 till present)
  - EWK Wenu/Zee software contact, ntuple production (May 09 till present)
  - Co-editor in CMS PAS EGM-10-001, Section for Electron Isolation (January 2010).
  - Main contributor in CMS-PAS-EWK-09-004
  - Contact for Egamma skimming operations and EWK electron operations (CMS October 09 Exercise)
  - Contact for the CMS electron efficiency (May 08-Oct08)

Overview

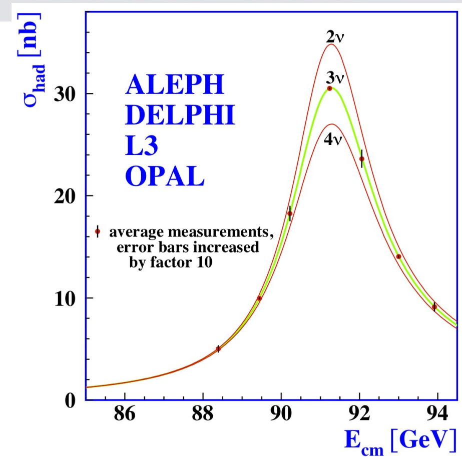
- Introduction: Weak Vector Bosons
  - Wenu channel
- CMS Experiment
- Overview of the Cross-Section Measurement
- $W \to e\nu$ Event Selection
  - Electron Selection
  - Tuning with a Genetic Algorithm
  - Tuning with the “Iterative Technique”
- Electron Selection Efficiency
- Signal Extraction
  - Data-driven jet template
  - Extrapolation-based signal extraction
- Cross-Section Results
- Outlook

Introduction: Weak Vector Bosons

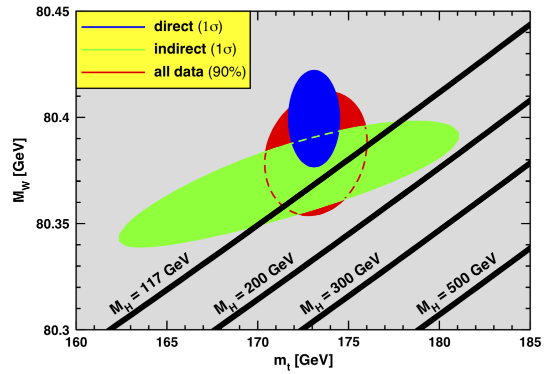
- The weak vector bosons (W,Z) have been discovered through their production in $p\bar{p}$ collisions and their leptonic decays (UA1 and UA2, 1983).
- Since then the measurement of their properties have contributed to the establishment of the Standard Model. Few examples follow:

Introduction: Weak Vector Bosons

- Number of Neutrino Species



- Higgs mass constraint from W and top-quark mass



- Electroweak precision tests (mostly from precision Z measurements)

Introduction: Weak Vector Bosons

- In pp collisions weak vector bosons play a very important role in physics and performance studies. Some examples:
  - W+jets, Z+jets, multiboson production is an important background to new or rare SM physics
  - Top-quark physics: $t \rightarrow Wb$; Higgs physics: $H \rightarrow ZZ$, $H \rightarrow WW$
  - Energy scale calibration using precisely known Z mass
  - W is a major source of prompt leptons and real missing transverse energy (MET)

Introduction: Weak Vector Bosons

- In pp collisions weak vector bosons play a very important role in physics and performance studies. Some examples:
  - W+jets, Z+jets, multiboson production is an important background to new or rare SM physics
  - Top-quark physics: $t \rightarrow Wb$; Higgs physics: $H \rightarrow ZZ$, $H \rightarrow WW$
  - Energy scale calibration using precisely known Z mass
  - W is a major source of prompt leptons and real missing transverse energy (MET)

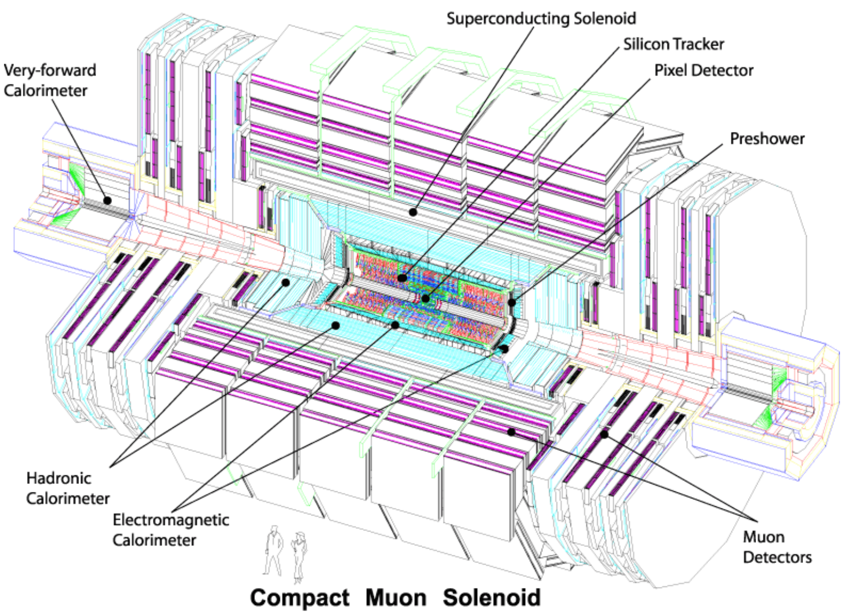
Very important tool for lepton and MET commissioning, which is the most important use of the measurement that is discussed in this talk

Introduction: W→eν process

- W→eν decay is the most abundant source of high-pT electrons and high MET events
  - Cross section of ~10nb or ~5000 good electrons (p_T>20GeV/c) per pb^{-1} in CMS
- Very useful in commissioning of
  - electron reconstruction and identification
  - MET
- Some physics measurements are also possible
  - pp → WX; W → eν cross-section measurement is a test of perturbative QCD
    - » Can also be used as a luminosity estimator
  - The ratio of W and Z production cross sections can give a precise (but indirect) measurement of Γ_W

The CMS Experiment

- General purpose detector designed to measure the output of LHC pp (and heavy ion) collisions

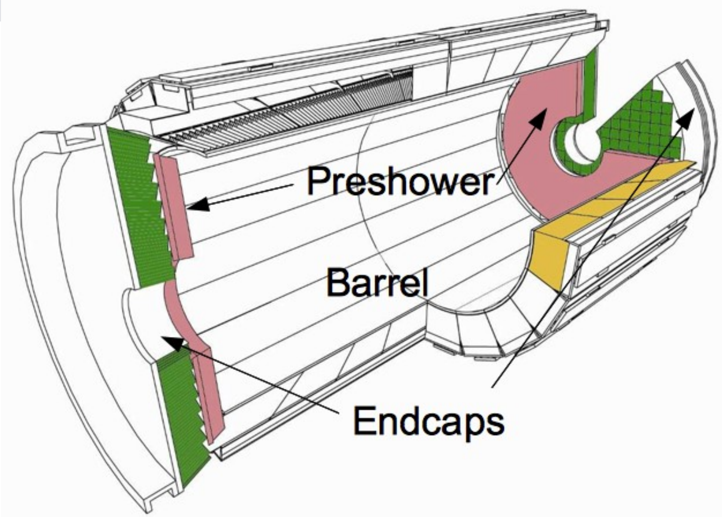


Basic features:

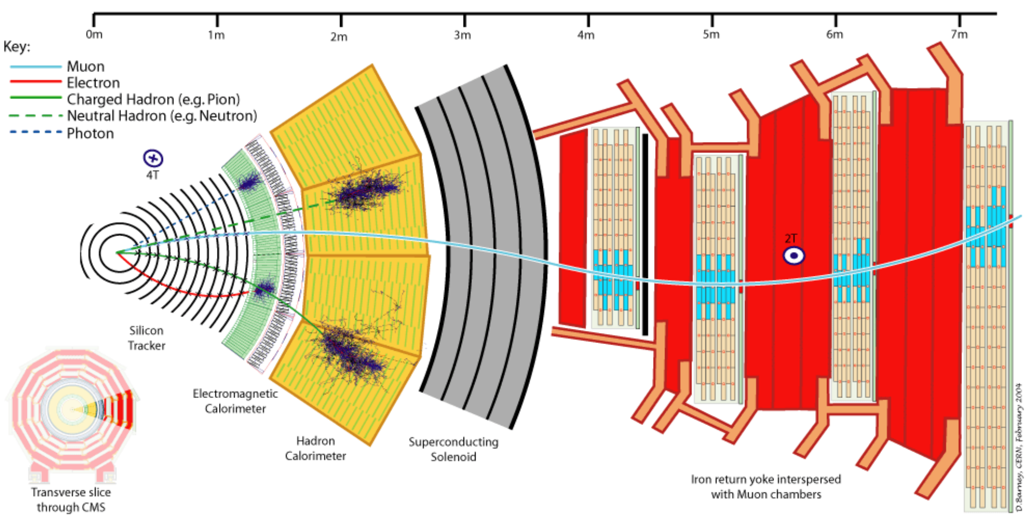
- Large solenoid magnet that encloses inner tracking and calorimetry systems
- All-silicon tracker
- Homogeneous electromagnetic calorimeter (ECAL)
- Hermetic calorimetric coverage (up to $|\eta|<6.5$ including the very forward calorimeters)

The CMS Experiment

- CMS ECAL is a homogeneous lead tungstate crystal calorimeter
  - Designed to fit in the very compact CMS design
  - Good energy resolution (stochastic term ~3% cf. ~10% for the ATLAS LiAr ECAL)



- Example of typical particle interactions in CMS



Overview of the Cross-section Measurement

- How to calculate a cross section

Signal extraction/ bkg removal

$$\sigma = \frac{N_{candidates} - N_{bkg}}{A \epsilon \int L\, dt}$$

Integrated luminosity:
Measured with Hadronic Forward calorimeter, normalized from beam parameters

Acceptance of kinematic cuts
Estimated from simulation

Efficiency of selection criteria

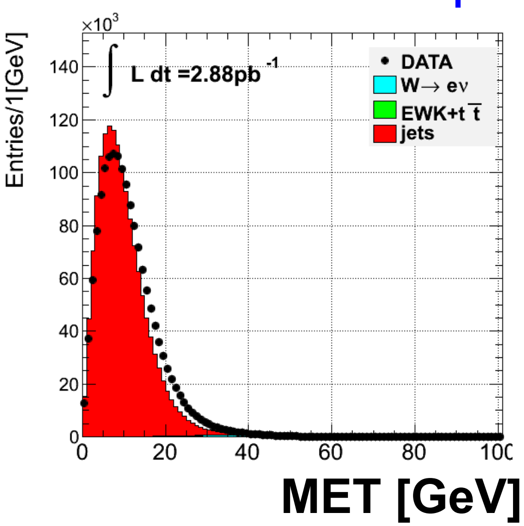
- In this talk I will focus on
  - W-candidate selection
  - Selection efficiency
  - Signal extraction

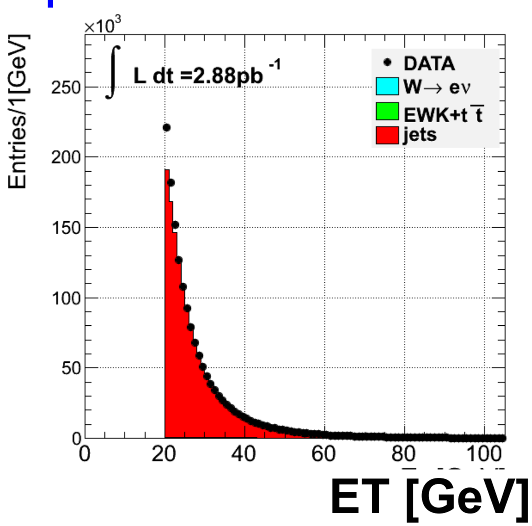
Dataset in use corresponds to $3pb^{-1}$

Measurement also described in CMS-PAPER-EWK-10-002 (accepted by JHEP)

W→eν Event Selection

- W→eν events are characterized by a high-pT electron (>20GeV/c) and high MET (>20GeV)
- However, these criteria are not sufficient to extract a pure W sample
  - A single reconstructed electron sample contains a very small number of prompt electrons





Single reconstructed electron sample with electron ET>20GeV

Sources of Electron Background

Charged hadron - π^0 overlap: matched in space with a photon shower from π^0

Charged Hadrons showering early in ECAL, Charge exchange ($\pi^+ n \to \gamma p$)

Electrons from **conversions** or from **heavy flavor quark decays** (real electrons)

$W \to e\nu$ Event Selection

- A method to enrich the single electron candidate sample with prompt electrons is to apply selection criteria on the electron candidates based on prompt electron properties like:
  - Isolation, Shower width and length,
    Track-ECAL cluster matching in η and φ directions (Δη, Δφ)
  - Conversion rejection:
    search for a conversion partner track
    search for missing hits in the inner tracker layers, before the first hit that belongs to the electron candidate track

Electron Selection

- Electron Selection: specific case of a classification problem
- Statistical theory tells us (Neyman-Pearson lemma) that the best classifier is the likelihood ratio
  - But difficult to calculate → approximations/use of other classifiers
  - Classification with cuts on variables will use this classifier

$$t(\vec{x};\vec{c}) = \prod_i H(c_i - x_i),\quad c_i = \textit{tunable parameter: cut value on electron property } x_i$$

$$H: \textit{step function}$$

And the tuning of the classifier parameters (c_i, or "cuts") is done by minimizing the function:

$$f(\vec{c};\{\vec{x}\},S) = \sum_{j\in Bkg} t(\vec{x}_j;\vec{c}) + \lambda\left(\sum_{i\in Sig} t(\vec{x}_i;\vec{c}) - S\right) \quad \textbf{(1)}$$

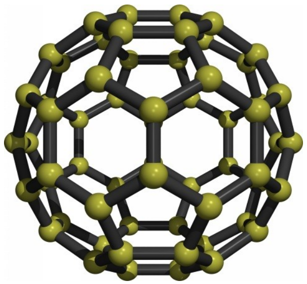
this effectively means that for a given signal, the cuts are chosen such that the background rejection is maximized – nothing new: exactly what the Neyman-Pearson lemma does, but in the case of a specific classifier

Electron Selection Tuning

- There are many ways in the market to minimize Eq (1) of the previous slide
  - All of them start from the definition of signal and background samples, which are used for training/testing the classifier
  - In the following I will focus on describing 2 techniques that I have worked in the past
    - » Tuning based on a **Genetic Algorithm**
    - » Tuning based on an **Iterative Technique**

Tuning with a Genetic Algorithm Implementation

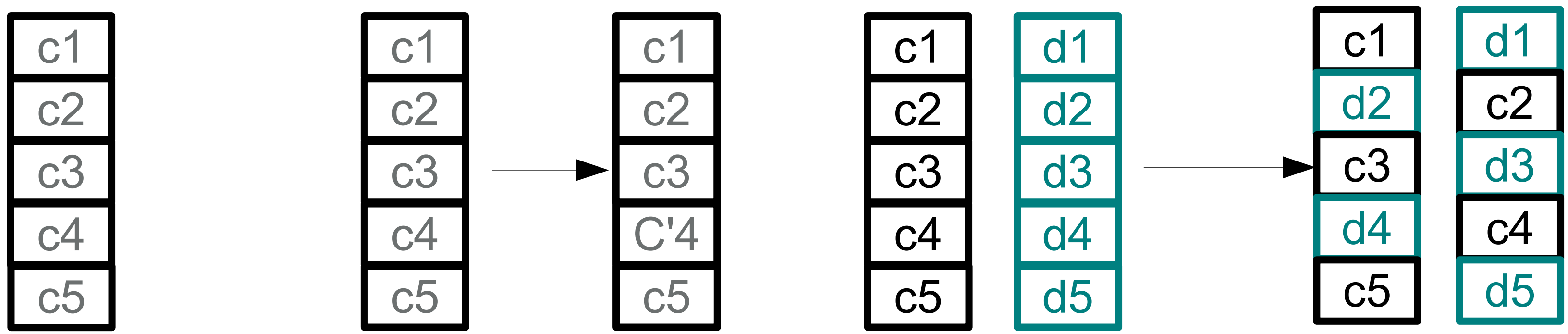
- Genetic algorithm is a well established technique, first used in the 1960's
  - First application on physics in the 1990's with simulations of the fullerene structure



- Elements of the method
  - A potential solution to the problem is codified in a "**chromosome**", C in $\overline{\overline{\mathbf{C}}}$
  - Definition of **operators**:
    - » Mutation $\mathbf{M} : \overline{\overline{\mathbf{C}}} \rightarrow \overline{\overline{\mathbf{C}}}$
    - » Crossing $\boldsymbol{\aleph} : \overline{\overline{\mathbf{C}}} \times \overline{\overline{\mathbf{C}}} \rightarrow \overline{\overline{\mathbf{C}}} \times \overline{\overline{\mathbf{C}}}$
  - **Ordering principle**: $C_1 > C_2$

Tuning with a Genetic Algorithm Implementation

Chromosome **Mutation operator** **Crossing operator**



Ordering through a "fitness function" : $r(\vec{x}) \times Gaus(\epsilon(\vec{x}) - \epsilon_0)$

Bkg rejection (r), Signal Efficiency ($\epsilon(\vec{x})$), Target Efficiency (ϵ_0)

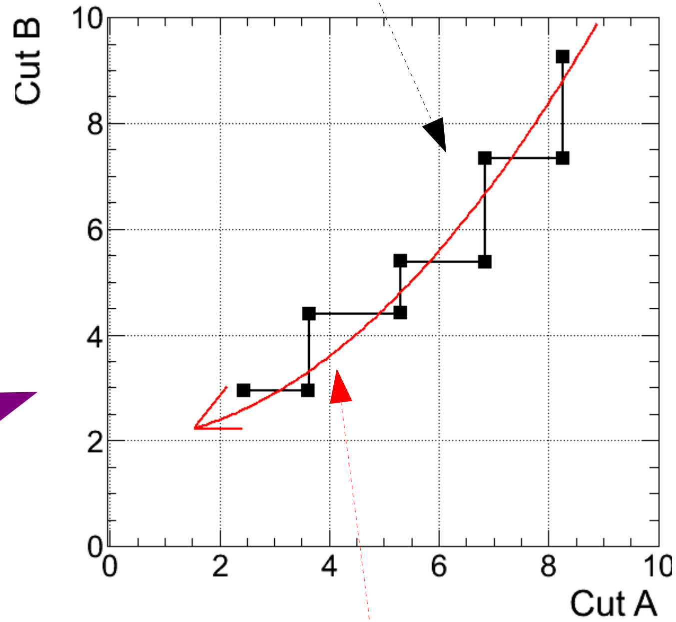
Steps : Create an initial chromosome population O(100)
Perform Mutations and Crossings to increase its size
Keep the best members of the population
Iterate

The "Iterative Technique" for Selection Tuning

- The "Iterative Technique" is an approximation of the gradient descent minimization of Equation (1)
  - It starts from a signal and background sample and a configuration with very loose cuts
  - Steps:

1. define a target in bkg rejection that is slightly higher than the current one
2. find which **single** cut can achieve this bkg rejection target with the highest signal efficiency
3. change this single cut only to obtain a new selection
4. iterate

path followed by the iterative technique



iterative algorithm concept illustration for a 2 cut case

optimal curve that the algorithm tries to approximate

The "Iterative Technique" for Selection Tuning

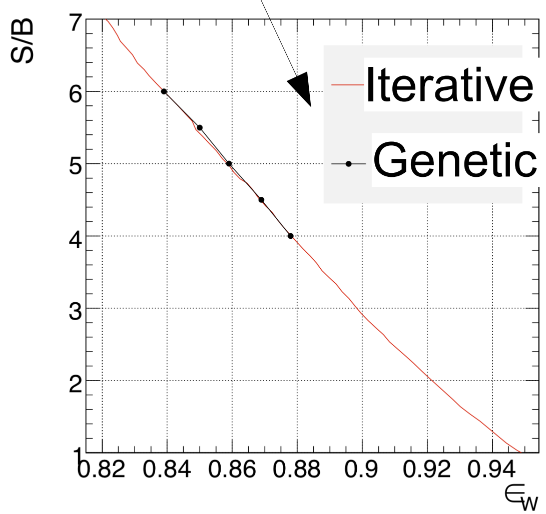
- Validation of the algorithm using simulation
  - Signal from W→eν; Bkg from jets+EWK bkgs to W→eν

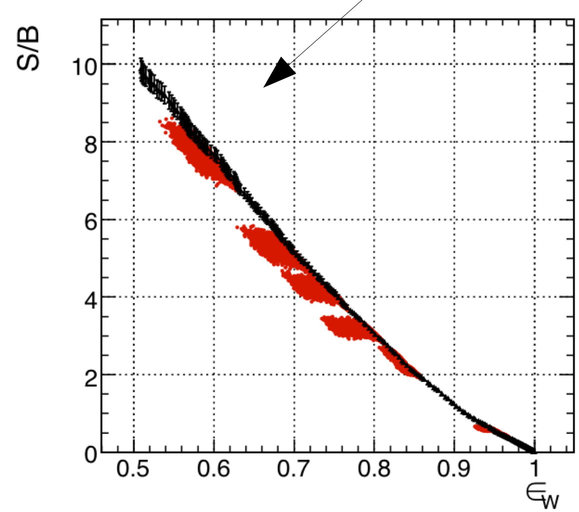
using electrons with ET>30GeV

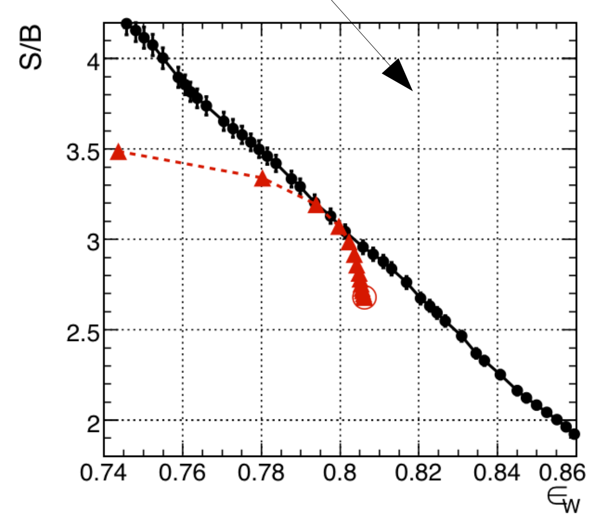
Moving the cut on a single variable (here ECAL isolation in EB)

Comparison with the Genetic Algorithm Tuning

Randomnly generated points, seeded by working points that the iterative produces



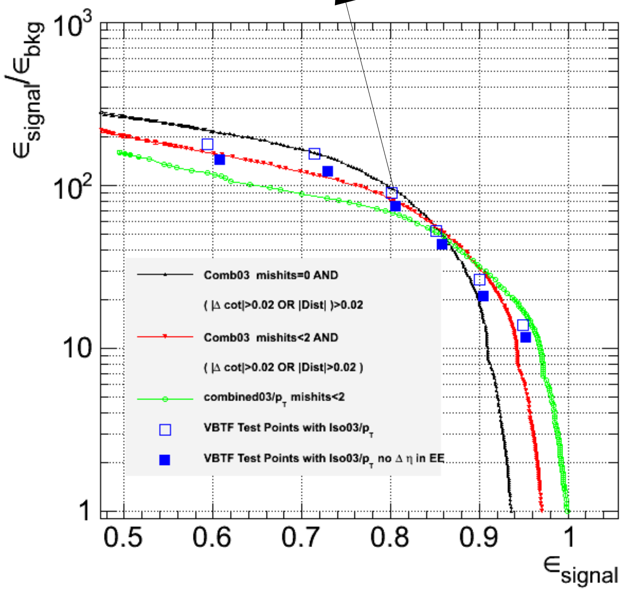




The "Iterative Technique" with MC Samples

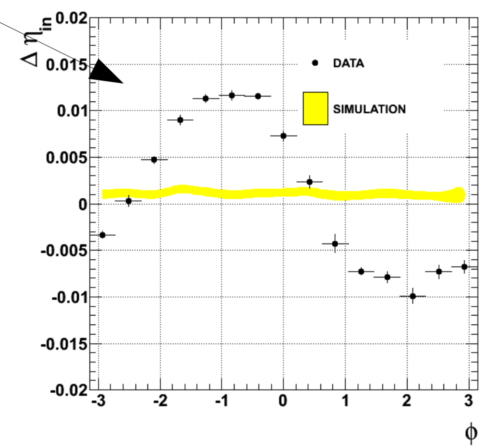
- The "Iterative Technique" was used with MC samples to tune electron selections for different conversion rejection tightnesses

"WP80" selection used for $W \rightarrow e\nu$ cross section



- When the first data became available the simulation was found to describe electrons pretty well
  - » Modulo the ECAL Endcaps–tracker misalignment

The MC-tuned electron selections have been used without the $\Delta\eta$ cut in ECAL EE for electron identification in data throughout the first year of data taking for all CMS analyses that use electrons



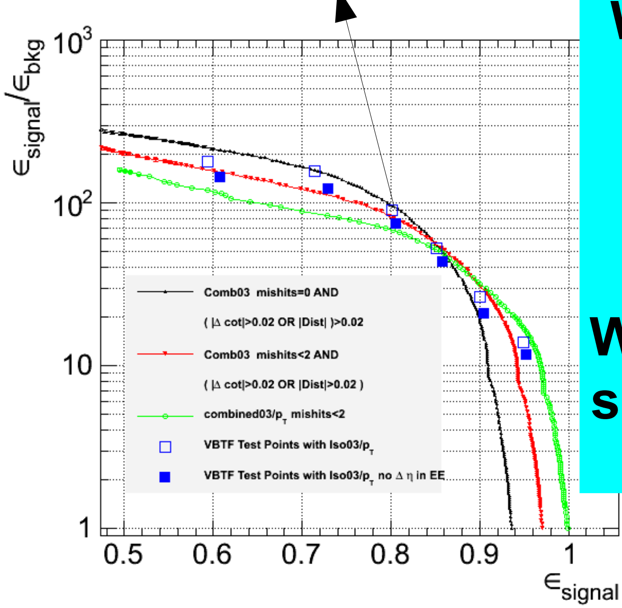
The “Iterative Technique” with MC Samples

- The “Iterative Technique” was used with MC samples to tune electron selections for different conversion rejection tightnesses

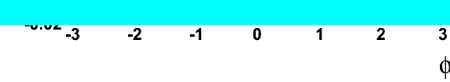
All CMS analyses in 2010 with electrons have used one of these electron selections, e.g.

W and Z cross sections CMS-PAPER-EWK-10-002
Top quark production CMS-PAPER-TOP-10-001
Search for b' CMS-PAPER-EXO-10-018
W charge asymmetry CMS-PAPER-EWK-10-006
W polarization CMS-PAPER-EWK-10-014
WW production observation CMS-PAS-EWK-10-009
single-lepton SUSY searches (e.g. CMS-CR-10-030)
...

“WP80” selection used for W→eν cross sec

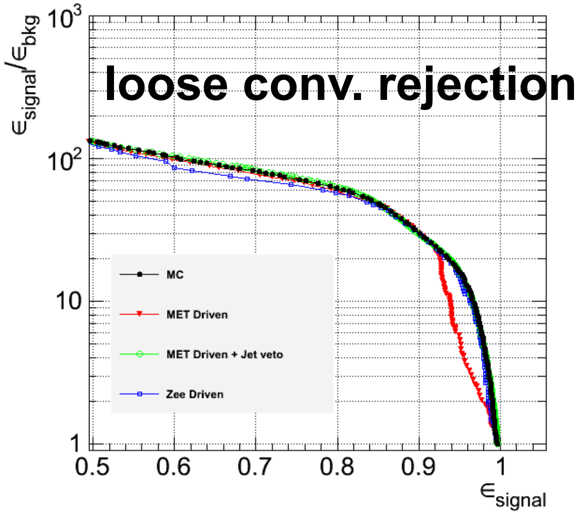


of data taking for all CMS analyses that use electrons



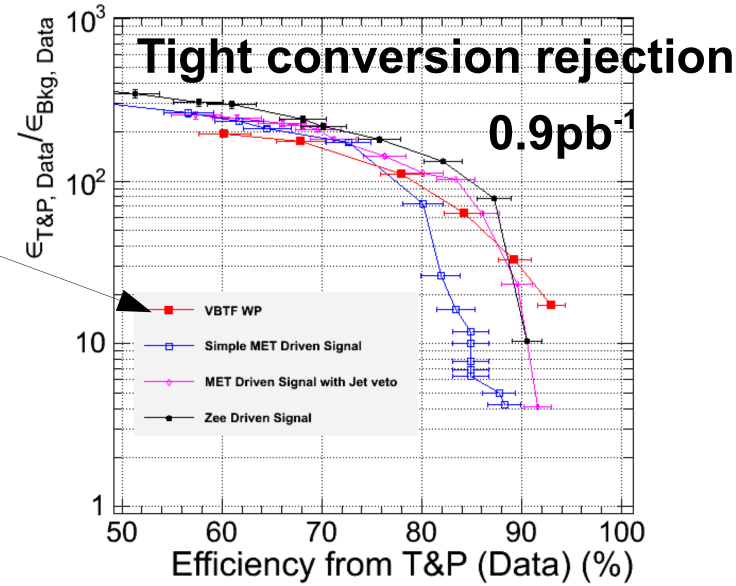
Data Driven Selection Tuning with the "Iterative Technique"

- Data-driven definitions of signal/bkg samples are also possible from a single electron (ET>20GeV) sample:
  - Bkg: MET < 20GeV
  - Signal: 3 different ways
    - » MET > 30GeV
    - » MET > 30GeV plus jet veto
    - » electrons from Zee



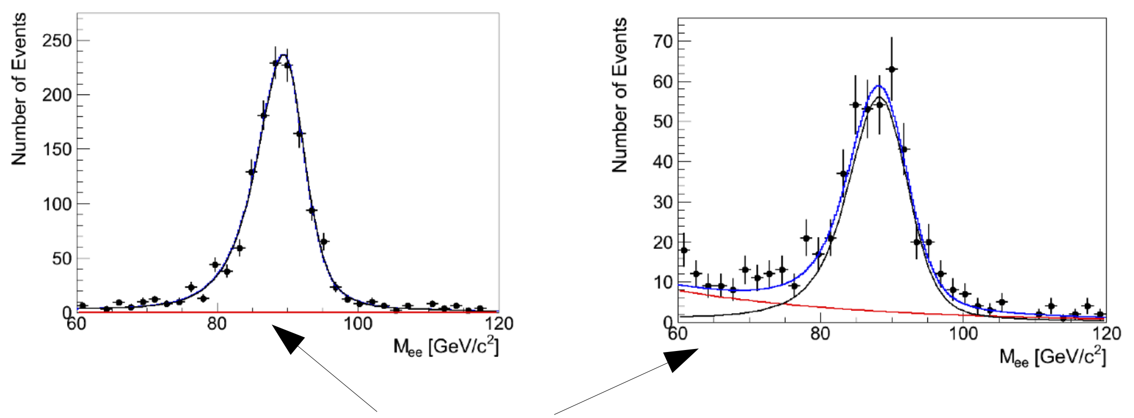
Example of MC test of the data-driven sample definitions

Tuning example with real data!



Electron Selection Efficiency

- Electron selection efficiency is measured from data using a pure electron sample from Z decays (Tag-and-Probe)
  - One well identified electron tags the event and a second electron (probe) is used to estimate the efficiency
  - Efficiency is estimated by a template fit of the tag-probe invariant mass spectrum for the tag+(probe passing selection) and the tag+(probe failing selection)



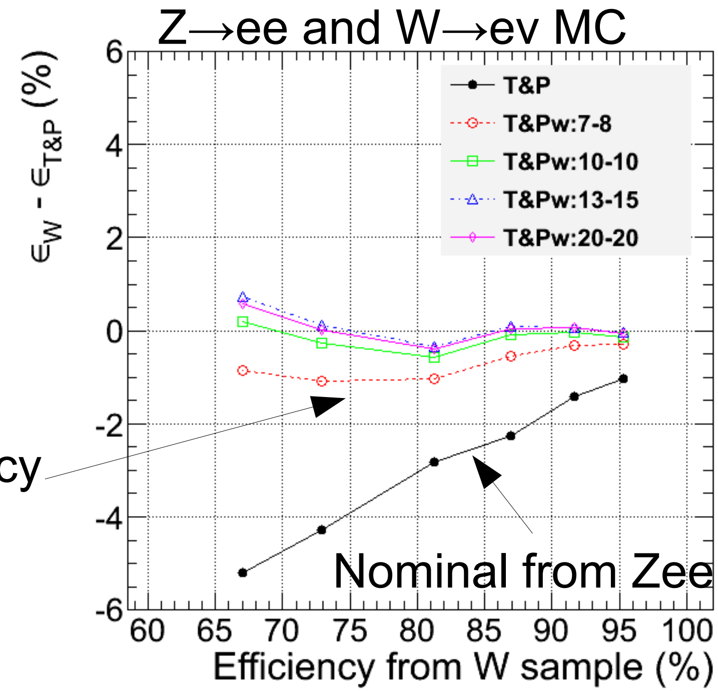
Example Fits:
probes are reconstructed electrons that **pass** or **fail** WP80 selection

Electron Selection Efficiency

- The W efficiency as it is measured using Z events is biased
  - Kinematic differences between the W and Z lead to differences of the efficiency of the same selection

To allow for these kinematic differences the measured efficiency is corrected using simulation:

$$\epsilon_{sele} = \frac{\epsilon_{W,MC}}{\epsilon_{TP,MC}} \epsilon_{TP,DATA}$$



Electron selection efficiency is estimated

$$\epsilon_{sele} = 72.0 \pm 2.8 \%$$

(including electron reconstruction + trigger efficiencies)

Signal Extraction

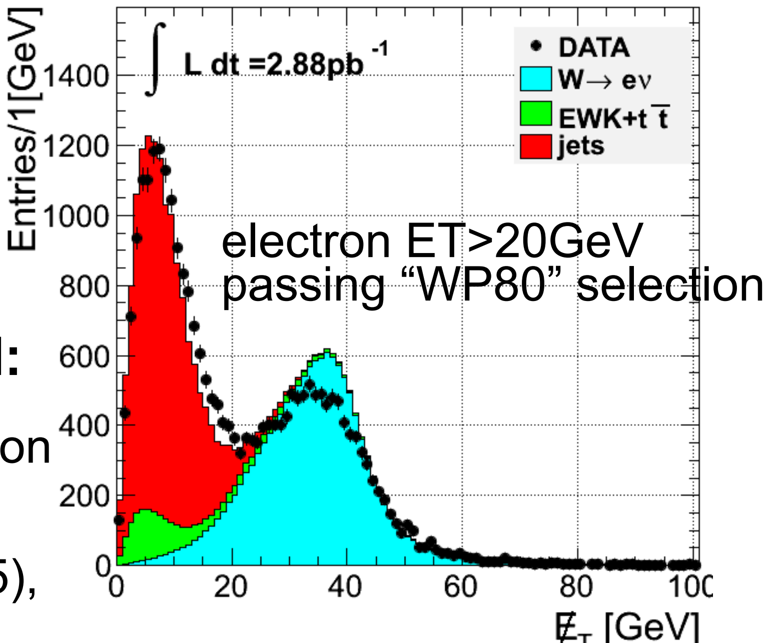
- Despite the electron selection the final collection of W candidates contains a considerable amount of background
- Different methods to extract the signal

“Template”-based:
Estimate signal and bkg shapes and extract the signal from a fit

e.g. this study

“Extrapolation”-based:
Extrapolate bkg to signal region from a bkg-rich region

e.g. D0 W width (2000), CDF W,Z production (2005), CMS-PAS-EWK-09-004, this study (as a cross-check)



MET distribution of W candidate events

"Template"-based Signal Extraction

- What it is all about:
  - Estimate somehow the MET shape of the components of the W candidate sample
  - Perform a fit to the data to extract the number of signal events

$$Nf_{DATA}(MET) = N_{jet} f_{jet}(MET) + N_W f_W(MET)$$

Many options on how to construct templates

Data-driven:
using a selection that rejects signal

Ansatz-based:
assuming a priori a functional form

Data-driven:
using Zee events

Simulation-based:
needs corrections for possible differences in MET resolution between data-MC

"Template"-based Signal Extraction

- What it is all about:
  - Estimate somehow the MET shape of the components of the W candidate sample
  - Perform a fit to the data to extract the number of signal events

$$Nf_{DATA}(MET) = N_{jet} f_{jet}(MET) + N_W f_W(MET)$$

Many options on how to construct templates

Data-driven:
using a selection that rejects signal

Ansatz-based:
assuming a priori a functional form

Data-driven:
using Zee events

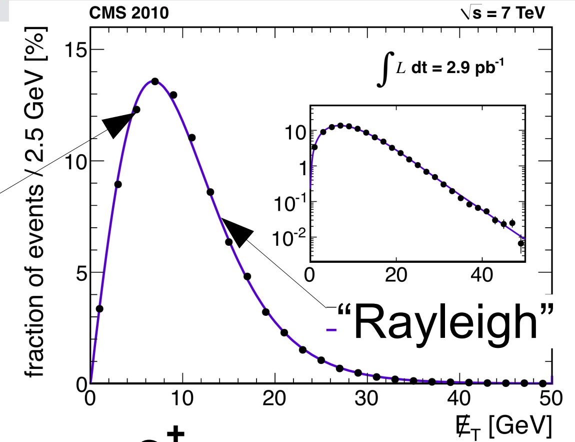
Simulation-based:
needs corrections for possible differences in MET resolution between data-MC

Selected templates for the final result
(the other methods used as a cross-check when lumi allows)

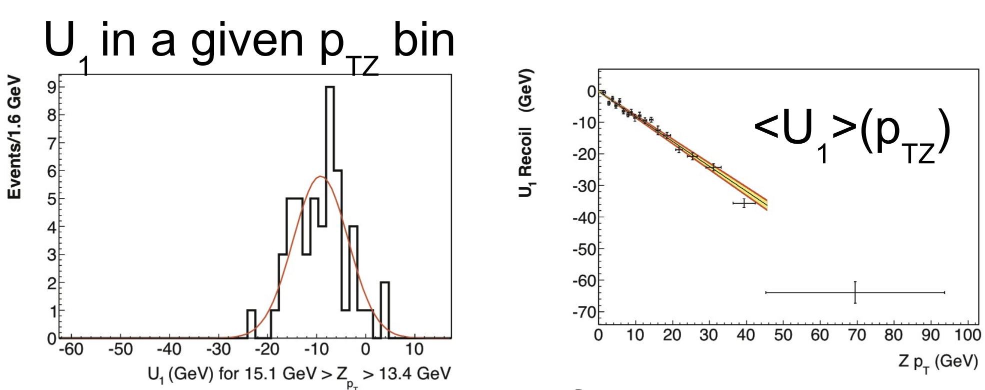
"Template"-based Signal Extraction

- Jet "template":
  physics-motivated Rayleigh function ansatz

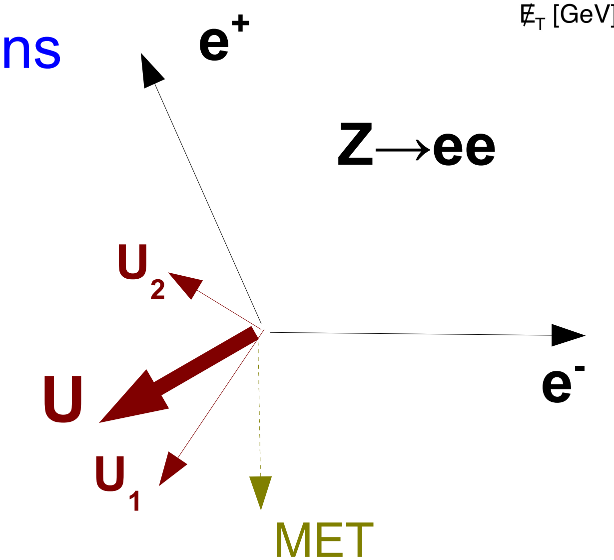
$$P_{jet}(x;\sigma_0,\sigma_1) = x\exp\left(-\frac{x^2}{2(\sigma_0+\sigma_1 x)^2}\right),$$



- W "template": from simulation
  - model from data the components of U in bins of boson pT assuming gaussian behavior
  - correct the Wev simulation



Example recoil fits on Z data

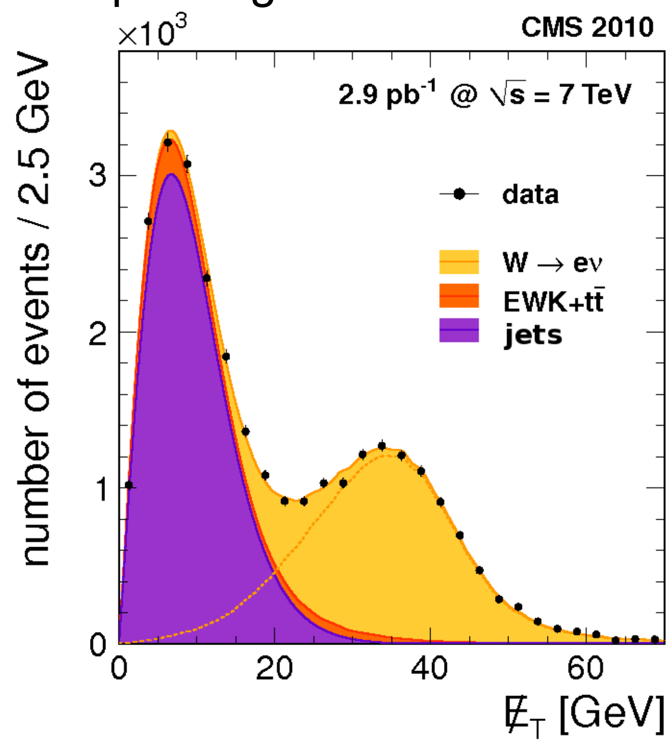


$$\vec{U} = -(\vec{E}_{T,e1} + \vec{E}_{T,e2}) - \vec{MET}$$

"Template"-based Signal Extraction

- Other bkgs apart from jets are modeled directly from simulation and added to the signal "template"
  - Most important processes: $W \to \tau\nu$, $Z \to \tau\tau$, top production
  - contribute ~13% of signal yield
- Excellent agreement with data

electron ET>20GeV passing "WP80" selection

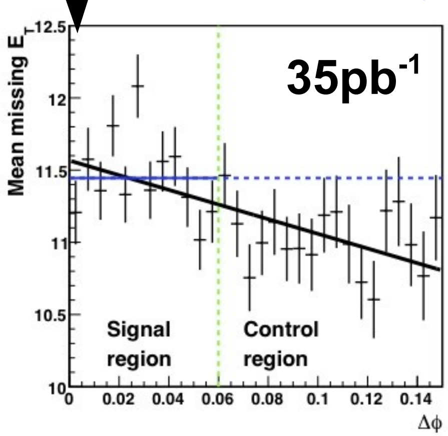


Number of signal events from the fit:

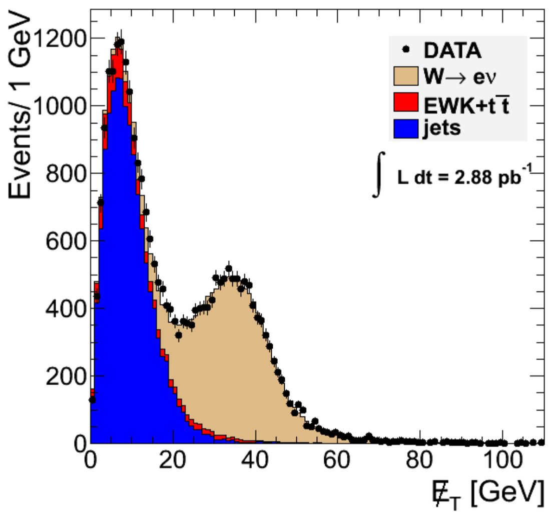
N = 11 895 ± 115
(statistical uncertainty only)

Data-driven Jet-"template"

- Data-driven jet "templates" have been used as a cross-check of the result
  - Defined by a selection that rejects signal: invert track-ECAL cluster matching cuts (Δη and Δφ)
  - Assumption: the inverted cuts are uncorrelated with MET
    - » And this is not quite true: possible to derive a correction with more data.
    - » This is the major source of systematic uncertainty of this method



With $3pb^{-1}$ the result using this jet template is in agreement with the Rayleigh "template" within 1.2%



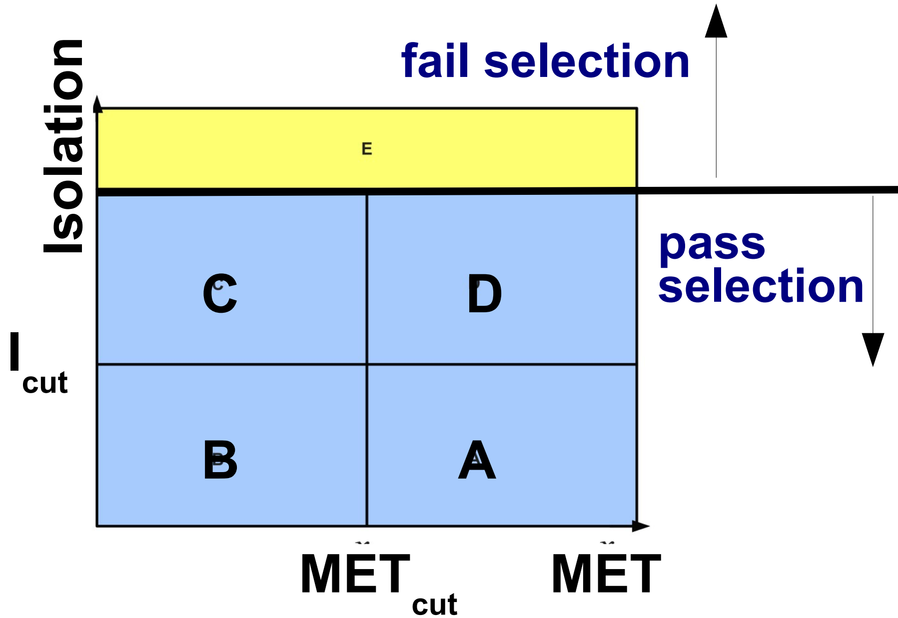
Imperial College London

"Extrapolation-based" signal extraction

- Benchmark method in the CMS before data taking
- Based on finding 2 uncorrelated variables as far as the jet distribution is concerned: here MET and Isolation

Inputs of the method:

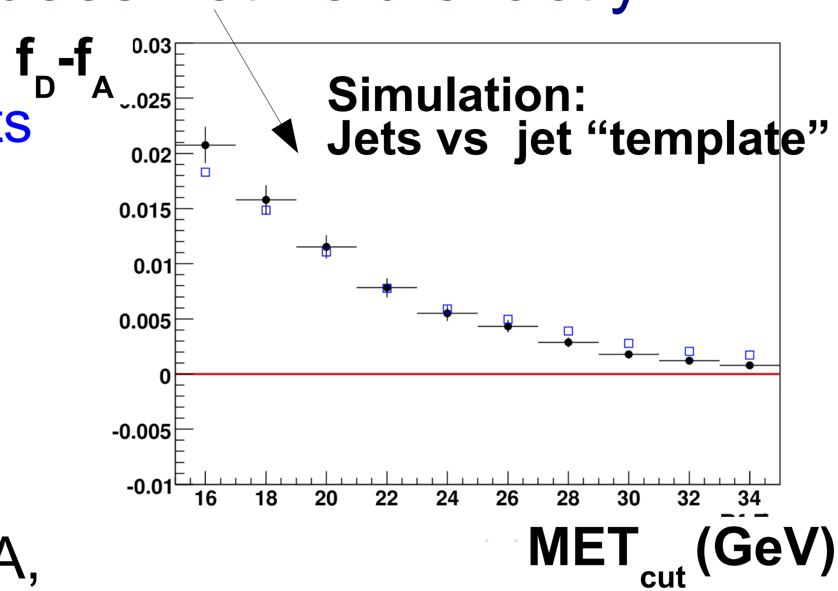
1. Events in A,B,C,D regions (from data)
2. MET efficiencies for signal events: $\varepsilon_A = S_A/S_{AB}$, $\varepsilon_D = S_D/S_{CD}$ (signal MET template needed)
3. Isolation efficiency for signal $\varepsilon_P = S_{AB}/S_{ABCD}$ (Tag-and-Probe from Z electrons)



By assuming that for jet events $f_A \equiv Q_A/Q_{AB} = f_D \equiv Q_D/Q_{CD}$
(i.e. jet shape extrapolation from CD to AB)
The total signal S_{ABCD} can be extracted

"Extrapolation-based" signal extraction

- Jet shape assumption $f_A = f_D$ does not hold exactly
  - This will result in a biased predicted number of signal events
  - A correction can be derived from data using the shapes from the data-driven jet "template"



But for all these you really need DATA, more data than $3pb^{-1}$

With $36pb^{-1}$ a data-driven correction of about 0.4% was derived
The overall systematic in the signal extraction is 1.2% (very preliminary) and dominated by efficiency uncertainties

Systematics

- Total systematic uncertainty (without luminosity error) ~5%
  - c.f. Statistical error: 1% and luminosity error: 11%

This figure will become smaller with more data
either because of their statistical nature or due to the fact that more data will permit the implementation of other methods

source	uncertainty value (%)
Efficiency	3.9
PDF uncertainty on acceptance	0.8
Theoretical uncertainties on acceptance	1.3
Electron energy scale/resolution	2.0
Jet $\not{E}_T$ shape modelling	1.3
Signal $\not{E}_T$ shape modelling	1.8
Total	5.1

Highest systematic in the efficiency measurement: most of it is of statistical nature

Electron energy scale determined from Zee

"template" related errors ~2.2% (more data will allow other potentially more accurate methods)

Cross-Section Measurement (3pb^{-1})

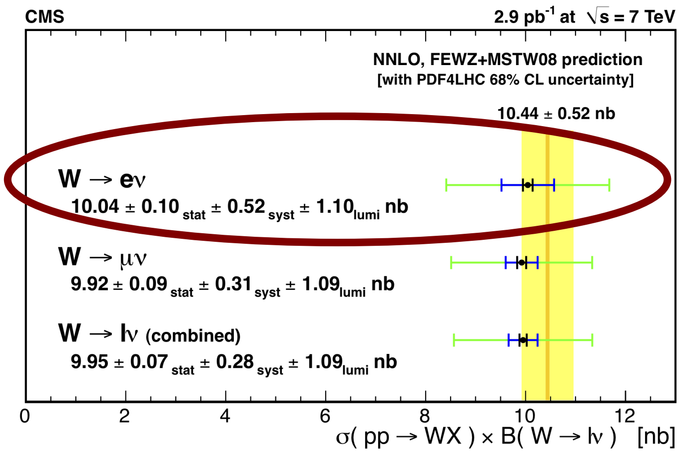
- **Results:** $N_{candidates} - N_{bkg} = 11895 \pm 115\,(stat)$

$$\sigma = \frac{N_{candidates} - N_{bkg}}{A \epsilon \int L\,dt}$$

$$A = 0.571 \pm 0.09\,(theory)$$

$$\epsilon = 0.720 \pm 0.028$$

$$\int L\,dt = 2.88 \pm 0.32\ pb^{-1}$$



- Final result in very good agreement with theory

$$\sigma(pp \to W + X) \times BR(W \to e\nu) = 10.04 \pm 0.10\,(\text{stat}) \pm 0.52\,(\text{syst}) \pm 1.10\,(\text{luminosity})\ \text{nb},$$

- Combined electron+muon result for the ratio

$$\frac{\sigma_W Br(W \to l\nu)}{\sigma_Z Br(Z \to l l)} = 10.64 \pm 0.40$$

Sensitive to Γ_W at about 4% level!

(c.f. **2%** all direct measurements combined)

Prospects for the Future

- Still room for improvement in the systematics
  - Electron selection efficiency and energy scale are determined from Z data, which are statistically limited
  - **Preliminary** results suggest that template systematics also reduced with more data, e.g.
    - » data-driven jet template is at 0.6% (c.f. currently: 1.3%)
    - » "extrapolation-based" signal extraction is about 1.2% (c.f. Currently: $2.2 \oplus 2\%$)
- New cross-section measurement with ~35pb^{-1} ongoing

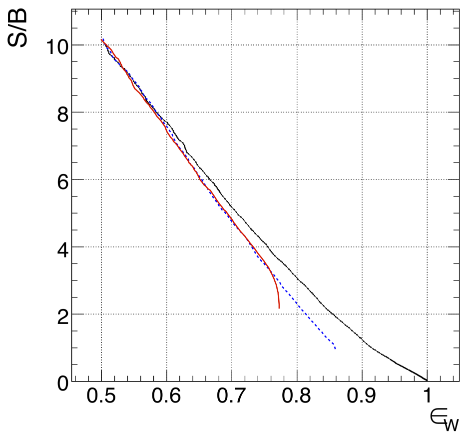
Acknowledgements

- Special thanks to C. Seez, G. Daskalakis, D. Wardrope, J. Hays, D. Futyan N. Wardle, J. Virdee,C. Timlin, J. Berryhill and many other CMS members for collaboration and useful comments
  - And to you for your attention!

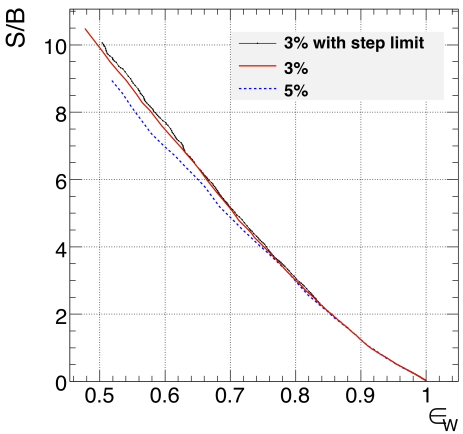
Additional Slides

"Iterative Technique" internal parameter tuning

- The algorithm internal parameters have to be chosen appropriately in order to achieve the optimal performance



Initial Conditions



Step convergence