Lepton fakes

Attila Aranyi RMKI-KFKI, Budapest

CMS SUSY Group

Weekly Budapest-Debrecen-CERN meeting

(NKTH, OTKA: NK67974, 74153)

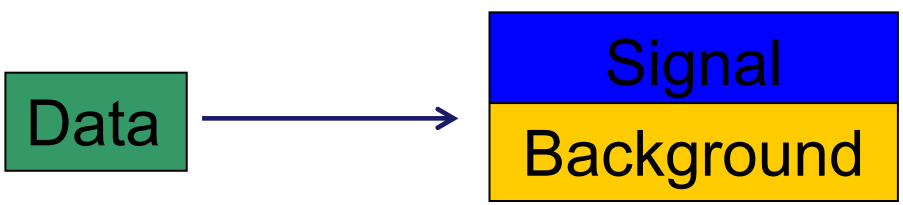
06/07/2009

Contents

- Motivation
- Muons
- Electrons
- Fake rate method
- Useful docs
- To Do

Motivation

- In a simple counting experiment useful strategy for the analysis is to separate the signal from background
  - define an event selection for the signal
  - count the number of candidates
  - compare this number of expected candidates from all non-signal sources (background)
  - ascribe the excess of events to signal (important to be able to validate the estimation of the background)

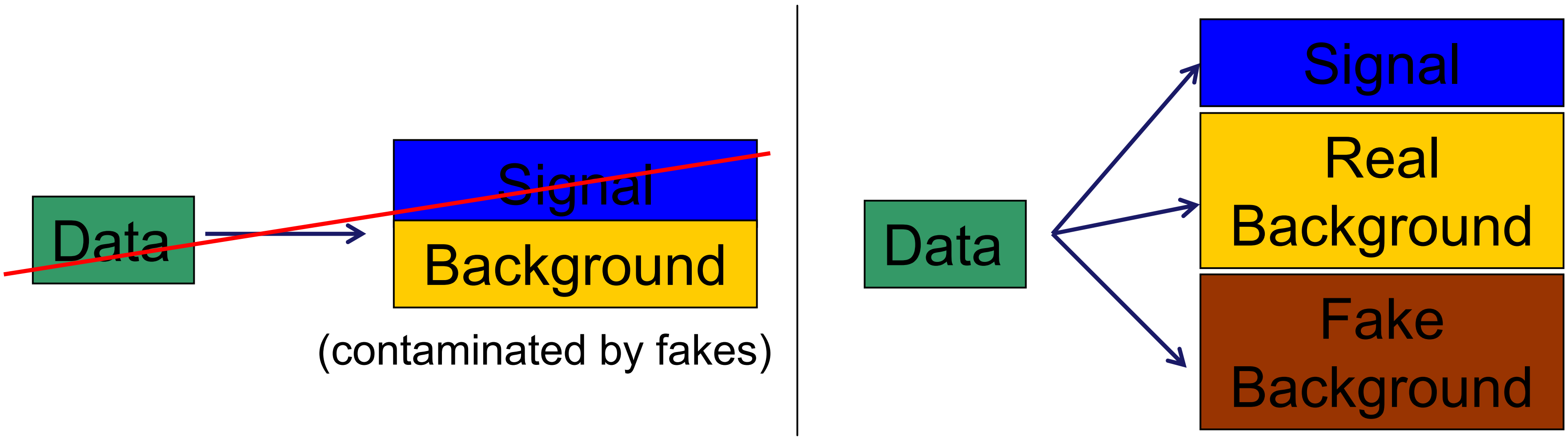


One of the many problems could be that the signal or background is contaminated by fakes (misidentified objects)

Insufficient signal/background ratio

Goal

•The goal is to estimate the fake lepton contribution to a signal lepton selection:



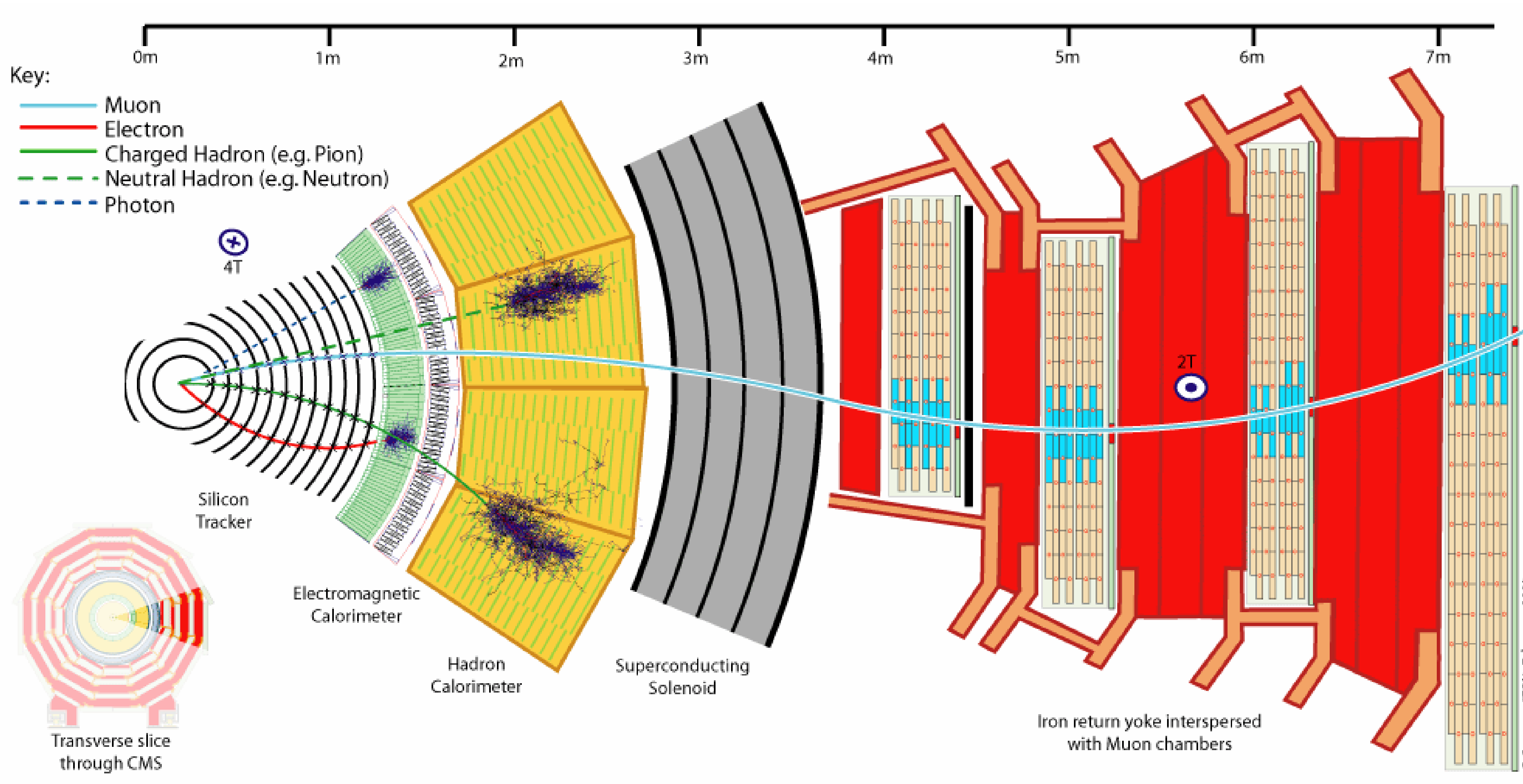
What we would like to measure and what we really measure?

Need to understand the physics and our hardware!

How to do:

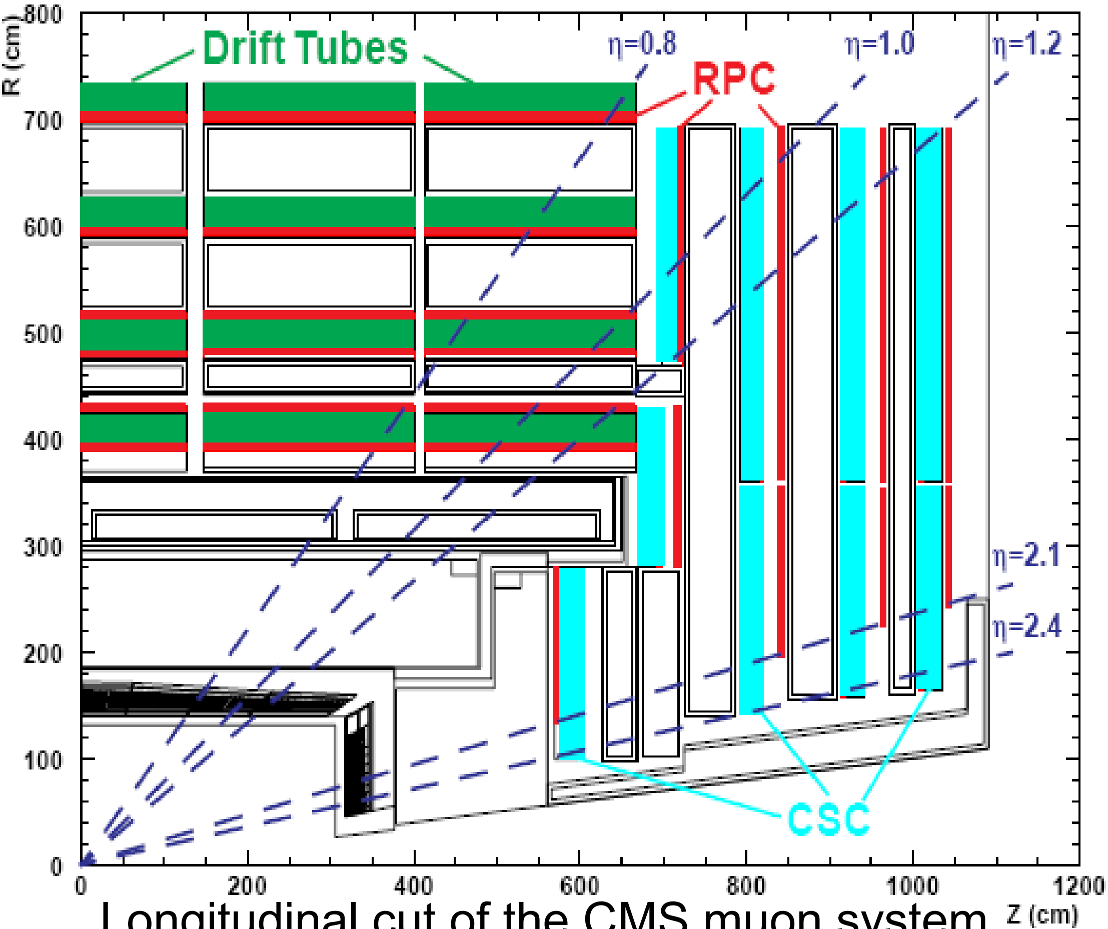
- signal and background
- used simulation
- reconstruction
- detector system
- …

Detector overview



Transverse slice of the CMS detector

Muon System



Longitudinal cut of the CMS muon system

- The CMS is specifically optimized for muon measurement
- a particle is measured by fitting a curve to hits among the four muon stations
- By tracking its position through the multiple layers of each station, combined with tracker measurements the detectors precisely trace a particle's path
- CMS magnet is very powerful so we can bend even the paths of very high-energy muons and calculate their momenta

250 drift tubes (DTs) and 540 cathode strip chambers (CSCs) track the particles' positions and provide a trigger, while 610 resistive plate chambers (RPCs) form a redundant trigger system, which quickly decides to keep the acquired muon data or not

Muon background

The basic tasks of the Muon Trigger are:
- muon identification,
- transverse momentum measurement,
- bunch crossing identification.

→ difficult because of the presence of severe background

3 main sources of background:
- proton-proton interactions themselves
- beam halo muons
- cosmic rays

Background classification:

detected objects	caused by	dominant source
tracks	low p_T muons	b- and c-quark decays, π and K decays
track segments	hadrons	punch-through and back-splashes
correlated hits	electrons	muon bremsstrahlung, δ-rays, e^+e^- production
uncorrelated hits	electrons	thermal neutrons $\rightarrow \gamma \rightarrow$ e

More detailed in: CMS Trigger TDR - Chapter 8 - Muon Trigger Introduction (15/12/2000)

Where the muons come from?

1. Proton-proton interactions
 (a) decays of heavy objects like W, Z, top, higgs, etc.,
 (b) b- and c- quark decays,
 (c) decays of hadrons composed with quarks
 u, d and s (mainly π and K),
 (d) punch-through of hadronic showers.

2. Beam losses because of the limited
 LHC aperture (sometimes called *beam halo muons*).

3. Cosmic rays.

Often called

Prompt muons

Fake muons

- In fact, the definitions depend on the analysis!
 (sometimes only 1.(a) are the Prompt muons)

- Need for muon identification due to the presence of fake muons!

- Goal: use cuts to decrease the rate of the fake muons ('fake rate')

Track quality cuts:

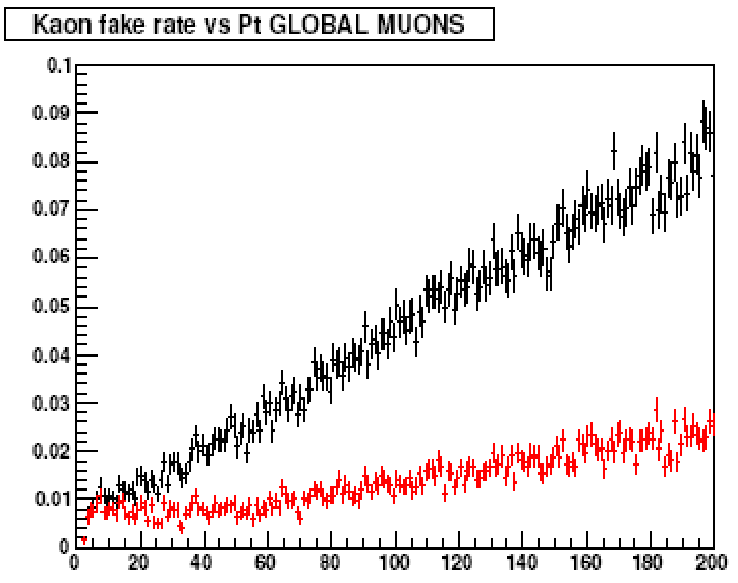
- Normalized-chi2 of the silicon fit.
- Impact parameter of the silicon or global fit.
- Number of hits of the silicon fit.

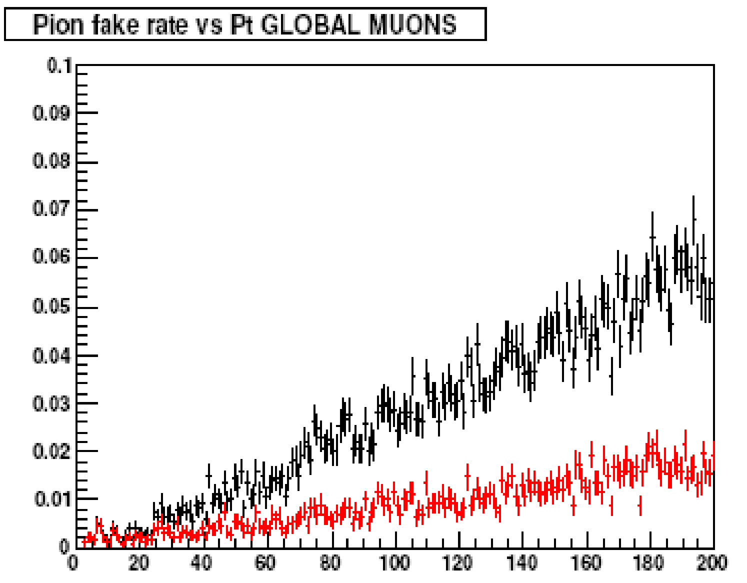
More detailed in: CMS Trigger TDR - Chapter 8 - Muon Trigger Introduction (15/12/2000)

Muon ID: π and K decays

1. (c) decays of hadrons composed with quarks u, d and s (mainly π and K),

- kaons and pions of pT < 200 GeV and |eta|< 2.4 generated flat in both pT and eta
- check the reply of the detector
- fake rate: ratio of fake muons to generated kaons and pions
- 'Decays in flight' (black) vs. 'Punch-Through' (red)

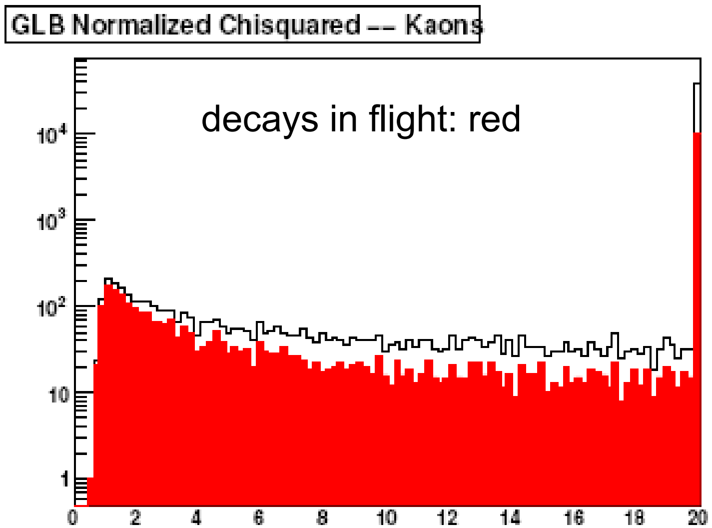


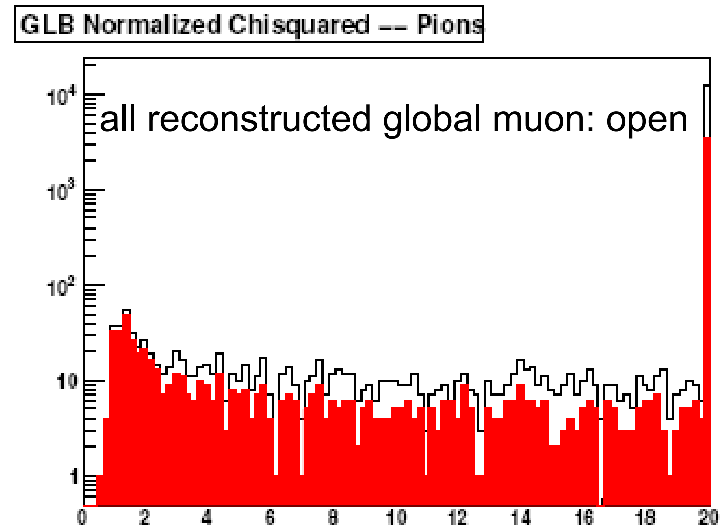


More detailed in: Muon Identification in CMS (03/11/2008)

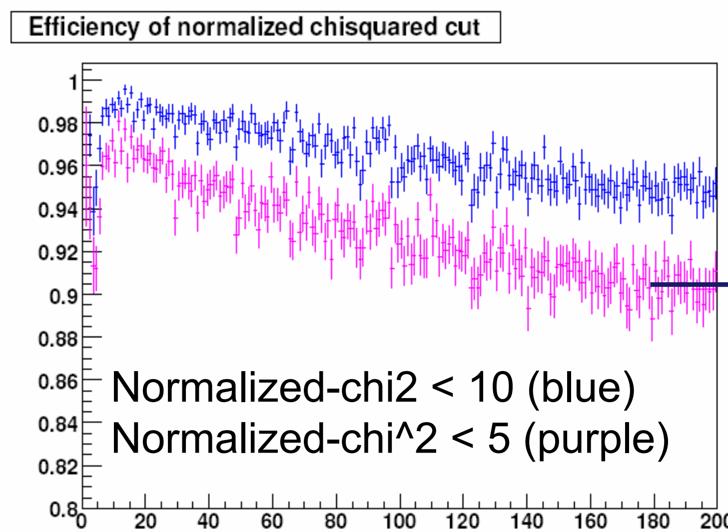
Muon ID: normalized-chi^2 cut

- Global muon normalized-chi^2 (of the track match) is a powerful tool to reject both decays in flight and punch through:





Use cuts



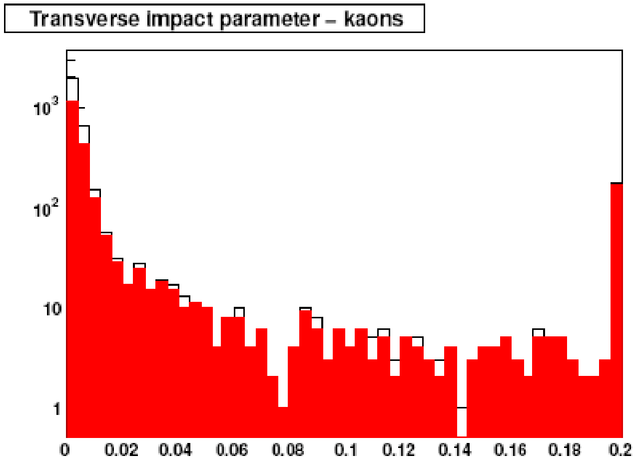
Use

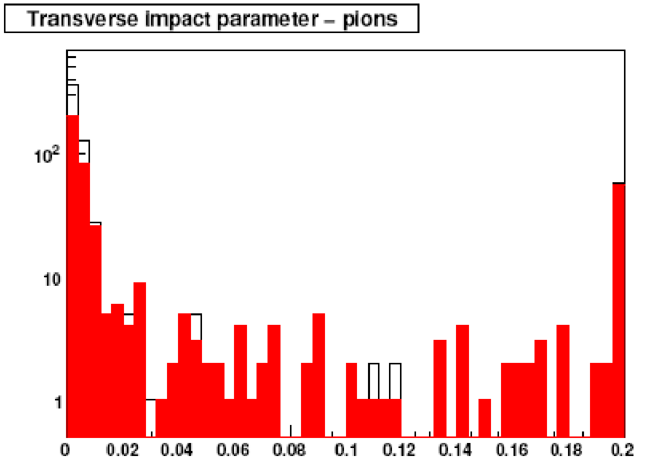
Normalized-chi2 < 10

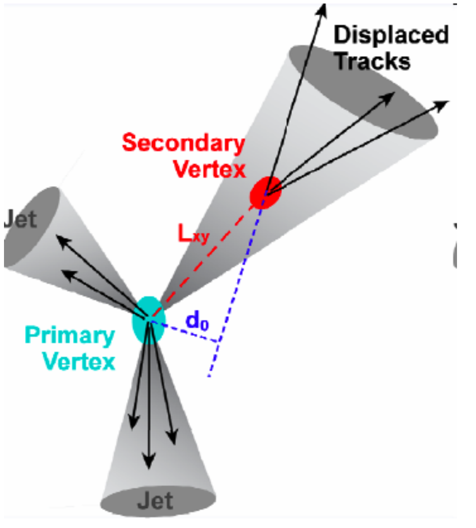
More detailed in: Muon Identification in CMS (03/11/2008), Muon Reconstruction in the CMS Detector (04/12/2008)

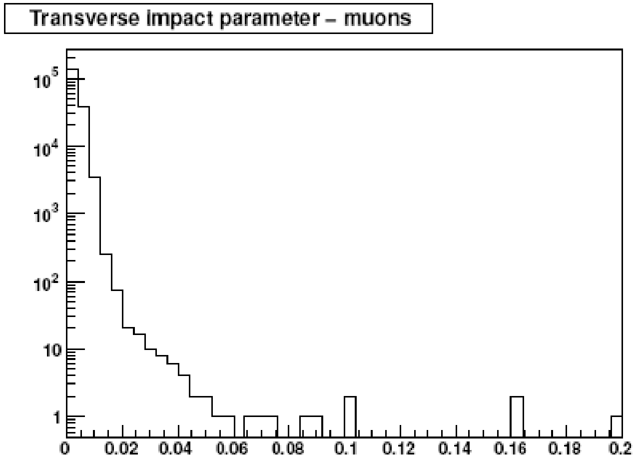
Muon ID: impact parameter

• d0, Impact parameter of the silicon fit (distance between the track and the primary vertex)









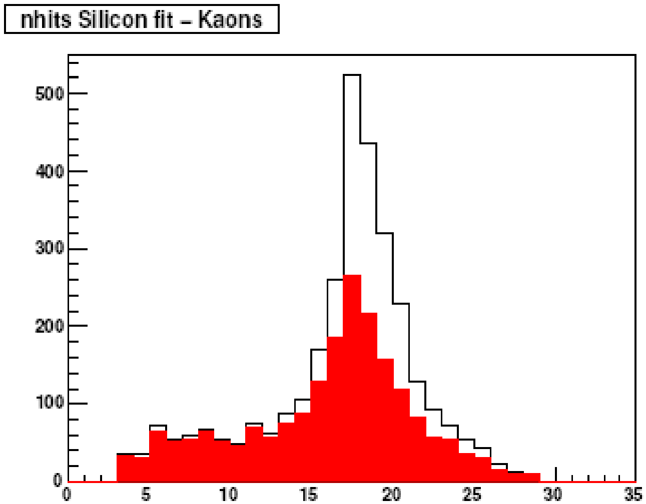
→ Useage of |d0| < 2mm

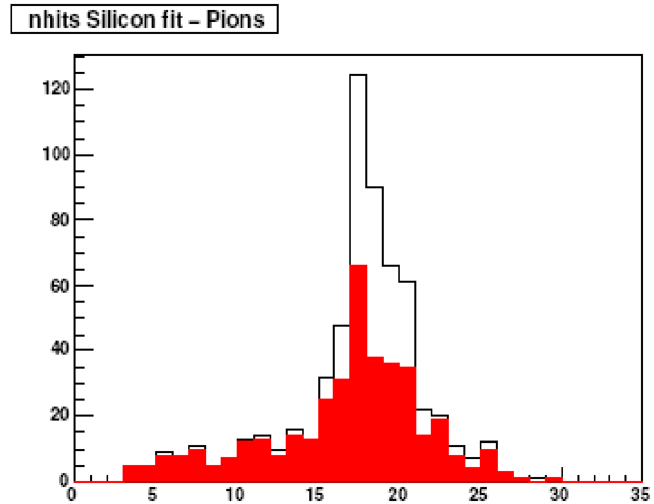
can be reject muons from b and c-quark decays also!

More detailed in: Muon Identification in CMS (03/11/2008), Muon Reconstruction in the CMS Detector (04/12/2008)

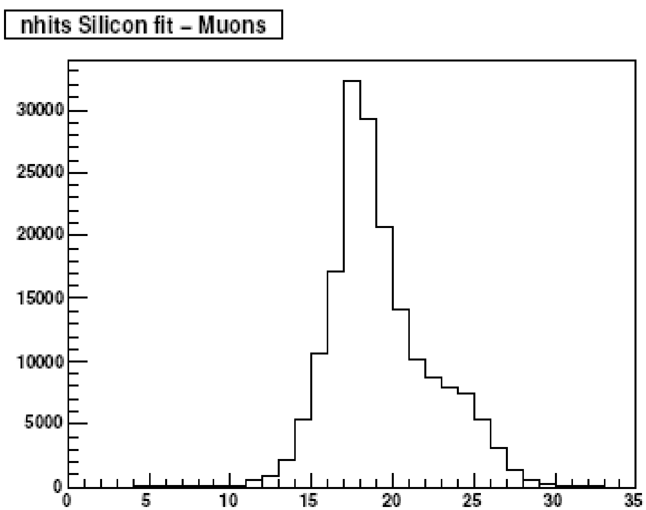
Muon ID: valid hits for the silicon fit

- The number of valid hits also has discriminating power





After apply previous cuts!



Use N_hits>11 for loose cut

More detailed in: Muon Identification in CMS (03/11/2008), Muon Reconstruction in the CMS Detector (04/12/2008)

Muon isolation

• Isolation criteria can be applied to the muon candidates to provide additional rejection
• suppress non-prompt muons from b, c, π, and K decays.

Three isolation techniques have been studied:

• *Calorimeter isolation*: based on the standard technique of summing the calorimeter energy in a cone around the muon. (in HCAL+ECAL)

• *Pixel isolation*: based on the partial reconstruction of tracks in the silicon pixel detector isolation is determined on the basis of the sum of the transverse momenta of the tracks in a cone around the muon.

• *Tracker isolation*: ΣPT of tracks reconstructed in a cone around the direction of the muon, neglecting the contribution from the muon itself.

Commonly used cut: let the relative isolation (or its reciprocal) above (below) a given threshold

(Cal_Iso+Pixel_Iso+Tracker_Iso) /PT

More detailed in: Technical Design Report, Volume 2: (15/12/2002)

Muon selection criteria

Examples:

- $p_T(\mu) > 30$ GeV
- $|\eta(\mu)| < 2.1$

→ Strongly depend on analysis

- $N_{\text{tracker hits}} > 8$ → Track quality cut
- $R_{\text{track isolation}} = \frac{\sum p_T \text{ of tracks in 0.3 cone}}{p_T(\mu)} < 0.1$ → Isolation cut
- $\frac{\chi^2/dof}{N_{\text{valid hits}}} < 1$ → Track quality cut

	loose cut	tight cut
transverse momentum	$p_T^\mu \geq 20$ GeV	$p_T^\mu \geq 20$ GeV
pseudo rapidity	$\|\eta_\mu\| \leq 2.5$	$\|\eta_\mu\| \leq 2.5$
track+calo based isolation, see B.2	$p_T^\mu/(p_T^\mu + sumIso) > 0.75$	$p_T^\mu/(p_T^\mu + sumIso) > 0.92$
number of tracker+μ-chamber hits		$n_{Hits} \geq 11$
global μ fit:	$\chi^2/NDoF \leq 20$	$\chi^2/NDoF \leq 10$
	more fakes	less fakes
	higher efficiency	lower efficiency

More detailed in:

Search Strategies for mSUGRA in the Muons + Jets + MET Final State (12/06/2008)

Data-driven methods to estimate the electron and muon fake rate contributions to lepton analyses (21/03/2009)

RA4 Muon selection (baseline cuts)

Quantity	PAT Object and Member Function	Cut	Comment
Mu type	pat::Muon => isGood("GlobalMuonPromptTight")	GlobalMuonPromptTight	Chi^2<10
p_T	pat::Muon => pt()	e 20 GeV	
abs(eta)	pat::Muon => eta()	d 2.1	
Rel. Isolation	pat::Muon => hcalIso(), ecalIso(), trackIso(), pt()	< 0.1	
chi^2/dof	pat::Muon => combinedMuon()–>chi2(), combinedMuon() –>ndof()	< 10	
abs(d_0)	pat::Muon => track()–>d0 *	< 0.2 cm	
N hits	pat::Muon => track()–>numvalhits()	e 11	
HCal Iso Deposit E	pat::Muon => hcalIsoDeposit–>candEnergy()	< 6	
ECal Iso Deposit E	pat::Muon => ecalIsoDeposit–>candEnergy()	< 4	

* this is the d0 from the track. The cut should be placed on the d0 w/ respect to the beam spot. The correspondence between the two is: $d0_{beamspot} = d0_{mutrack} - beamspot_x * sin(\phi_{mutrack}) + beamspot_y * cos(\phi_{mutrack})$

Note: relative isolation is calculated as

$$Isol = \left(\sum_{\Delta R<0.3} E_T(ECAL) + \sum_{\Delta R<0.3} E_T(HCAL) + \sum_{\Delta R<0.3} p_T(tracker)\right) / p_T(\mu)$$

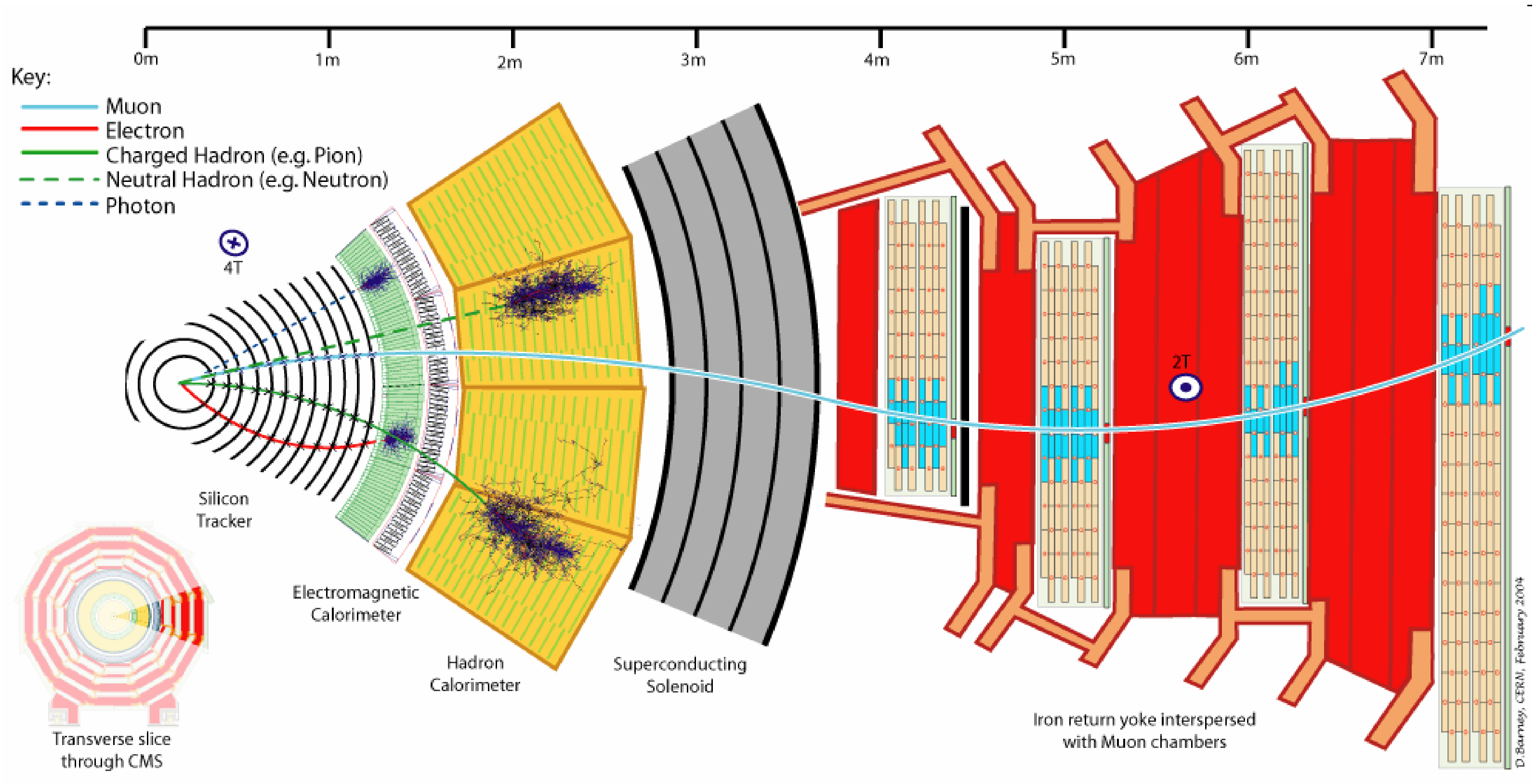
More detailed in: https://twiki.cern.ch/twiki/bin/view/CMS/SusyRA4SingleMuonProjectTable#Muon_Selection

RA4 Muon selection (baseline cuts)

Most recent talks about the Muon selection in RA4 from Carsten Magass:

Selection status report for SUSY Muon Id (12/02/2009)
Baseline Muon Id for SUSY selection (26/02/2009)
Study of Muon Id for RA4 (07/05/2009)
Muon Id for RA4 – Follow up (02/07/2009)

Electron system



• pixel detector	⟶	2-3 hits / track
• silicon strip tracker	⟶	14-15 hits / track
• ECAL	⟶	barrel, preshower device, endcap

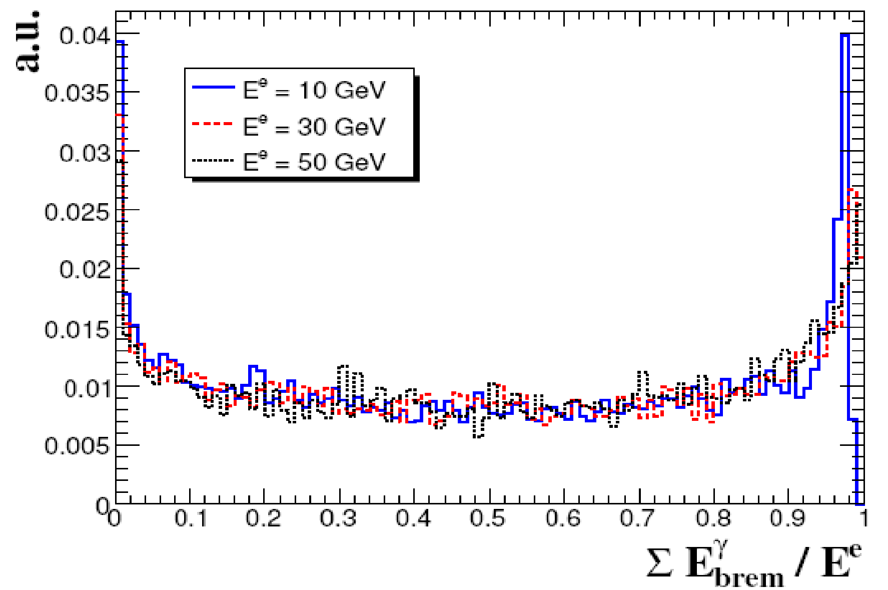
Problems

- there are unusal features:
  - high magnetic field
  - thick tracker, …

• Electrons traversing the silicon layers of the tracker radiate bremsstrahlung photons and the energy reaches the ECAL with a significant spread in the azimuthal direction.

generated electron energy E^{e}

radiated bremsstrahlung photons $\sum E^{\gamma}_{\mathrm{brem}}$



• About 35% of the electrons radiate more than 70% of their initial energy before reaching the ECAL. In about 10% of the cases, more than 95% of the initial energy is radiated.

More detailed in: Electron Reconstruction in CMS (02/2006)

Origin of fake electrons

- <u>Charge exchange interactions of charged pions</u>: As a charged pion turns into a neutral pion via charge exchange, the neutral pion decays to two photons. The entire energy of the charged pion is thus deposited in the calorimeter. The result is a track and an ECAL deposit with perfect E/p and impact parameter consistent with the primary vertex.

- <u>Fake electrons from photon conversion</u>: Any photon can convert early and asymmetrically and lead to a signal electron. Electrons from conversions tend to have impact parameters, inconsistent with the primary vertex.

- <u>Leptons from heavy flavor decays</u>: As only prompt leptons from WW/$t\bar{t}$ are considered signal, these leptons are considered fake.

- <u>matching ambiguity</u>: a photon cluster in the electromagnetic calorimeter is matched to a random track.

- <u>Final State Radiation (FSR) followed by conversions</u>: The tails of the FSR distribution in both energy and distance from the lepton may result in isolated electrons from conversion of the FSR photon.

Reconstruction and identification is much more difficult than in case of muons!

More detailed in: Data-driven methods to estimate the electron and muon fake rate contributions to lepton analyses (21/03/2009)

How to handle?

• Electron clustering: building of superclusters (clusters of clusters), is designed to take into account the Phi spread and collect the bremsstrahlung energy

• Electron classification: different treatment can be used in each category

→ Find physically meaningful variables that can differe from fakes!

• E/p (supercluster energy / track momentum at the vertex) is often well measured for electrons and it is not often measured to be less than 1. Many fakes from jets have $E/p < 1$ partly because of the low response of ECAL to charged pions.

• The electron signature in the detector greatly depends on how much energy it radiates. We define the fractional energy loss to bremsstrahlung as determined by the track momentum p_in estimated at the interaction point and the track momentum p_out when the track exits the tracker.

$$f_{\mathrm{brem}} = \frac{p_{\mathrm{in}} - p_{\mathrm{out}}}{p_{\mathrm{in}}}$$

fakes from jets usually have f_brem around 0 they are usually just charged pion tracks

More detailed in: [Electron Reconstruction in CMS (02/2006)](), [A cut based method for electron identification in CMS (24/09/2008)]()

Category based ID

Requirement: E/p > max[0.8, 0.9(1 − fbrem)]

- Category-0, bremming.: f_brem > 0.06 barrel (0.10 endcap) and 0.8 < E/P < 1.2.
- Category-1, low brem.: f_brem < 0.06 barrel (0.10 endcap).
- Category-2, bad track: otherwise.

more variables used to distinguish between real and fake electrons:

- H/E is the ratio of the E_T of the closest HCAL tower to the E_T of the electron reconstructed in the ECAL.
- $\sigma_{\eta\eta}$ is the covariance defined to be $\sum_i w_i(\eta_i - \bar{\eta}_{5\times5})^2/\sum_i w_i$ where $w_i = 4.7 + \ln(E_i/E_{5\times5})$.
- $\Delta\phi_{in}$ is the difference between the ϕ co-ordinate of the electron measured in the ECAL and the projection of the track as measured at the primary vertex and projected through the magnetic field to the ECAL.
- $\Delta\eta_{in}$ is the difference between the η co-ordinate of the electron measured in the ECAL and the projection of the track as measured at the primary vertex and projected to the ECAL.
- E_{seed}/P_{in} is the ratio of the energy of the seed cluster that initiated the bremsstrahlung recovery process to the momentum of the track at the vertex.

More detailed in: A cut based method for electron identification in CMS (24/09/2008),

Data-driven methods to estimate the electron and muon fake rate contributions to lepton analyses (21/03/2009)

Loose electron ID

• primarily designed to have enough fake rejection to clearly extract Z->ee signal from the QCD background.

		bremming	low brem	bad track	$E/p_{in} > 1.5$
H/E	(barrel)	0.115	0.10	0.055	
	(endcap)	0.145	0.12	0.15	
$\sigma_{\eta\eta}$	(barrel)	0.014	0.012	0.0115	
	(endcap)	0.0275	0.0265	0.0265	
$\Delta\eta_{in}$	(barrel)	0.009	0.0045	0.0085	
	(endcap)	0.0105	0.0068	0.010	
$\Delta\phi_{in}$	(barrel)	0.05	0.025	0.053	0.09
	(endcap)	0.07	0.03	0.092	0.092
E_{seed}/p_{in}	(barrel)	0.11	0.91	0.11	
	(endcap)	0.	0.85	0.	

More detailed in: A cut based method for electron identification in CMS (24/09/2008)

Tight electron ID

- a tighter level of identification is needed in W analysis.
  (sometimes called 'robust' tight ID)

		bremming	low brem	bad track	$E/p_{in} > 1.5$
H/E	(barrel)	0.022	0.02	0.015	
	(endcap)	0.04	0.02	0.02	
$\sigma_{\eta\eta}$	(barrel)	0.01	0.0095	0.009	
	(endcap)	0.0275	0.025	0.024	
$\Delta\eta_{in}$	(barrel)	0.00525	0.0025	0.004	
	(endcap)	0.005	0.005	0.0055	
$\Delta\phi_{in}$	(barrel)	0.015	0.01	0.015	0.09
	(endcap)	0.025	0.016	0.025	0.092
E_{seed}/p_{in}	(barrel)	0.3	0.93	0.60	
	(endcap)	0.3	0.85	0.65	

More detailed in: A cut based method for electron identification in CMS (24/09/2008)

Electron selection criteria

	loose cut	tight cut
transverse momentum	$p_T^e \geq 20$ GeV	$p_T^e \geq 20$ GeV
pseudo rapidity	$\|\eta_e\| \leq 2.5$	$\|\eta_e\| \leq 2.5$
closest muon veto	$\Delta R_{e-\mu} > 0.1$	$\Delta R_{e-\mu} > 0.1$
had/elmag energy fraction	$H/E \leq 0.2$	included in electron identification
track based isolation, see B.1	$p_T^e/(p_T^e + tkIso) > 0.92$	
track+calo based isolation, see B.2	$p_T^e/(p_T^e + sumIso) > 0.75$	$p_T^e/(p_T^e + sumIso) > 0.92$
impact parameter		$\|d_0^e\| \leq 0.025$
electron identification		Tight electron ID (see App. F and [1])
	more fakes	less fakes
	higher efficiency	lower efficiency

RA4 Electron selection (baseline cuts)

Quantity	PAT Object and Member Function	Cut	Comment
p_T	pat::Electron => pt()	e 20 GeV	
abs(eta)	pat::Electron => eta()	d 2.5	
Id	pat::Electron => electronID("eidRobustTight")	robustTight	
Rel. Isolation	pat::Electron => hcalIso(), ecalIso(), trackIso(), et()	< 0.1	
abs(d_0)	pat::Electron => gsfTrack()->d0	< 0.2 cm	

* this is the d0 from the track. The cut should be placed on the d0 w/ respect to the beam spot. The correspondence between the two is: $d0_{beamspot} = d0_{eltrack} - beamspot_x * sin(\phi_{eltrack}) + beamspot_y * cos(\phi_{eltrack})$

Note: relative isolation for the electron is calculated as

$$Isol = \left(\sum_{\Delta R<0.3} E_T(ECAL) + \sum_{\Delta R<0.3} E_T(HCAL) + \sum_{\Delta R<0.3} p_T(tracker)\right) / E_T()$$

More detailed in: https://twiki.cern.ch/twiki/bin/view/CMS/SusyRA4SingleMuonProjectTable#Electron_Selection

Definition of fake leptons

- ‘Prompt’: reconstructed leptons that are not from the primary decay of Z,W, hard interactions (DY), or new physics (SUSY)
- ‘Fake’: not Prompt

  Jet deposits: large amount of energy in the ECAL that is matched to a single high-P_T track

  Heavy quark decays: kick from quark decay may be enough to knock electron out of the jet a bit, making it appear isolated

  Photon conversion: any photon can convert early and lead to a signal electron

  etc…

- More detailed, sources of fakes in: AN-2009/041

  How many fake leptons are out there?

  How to handle them?

Need to predict lepton fake contribution to signal selection

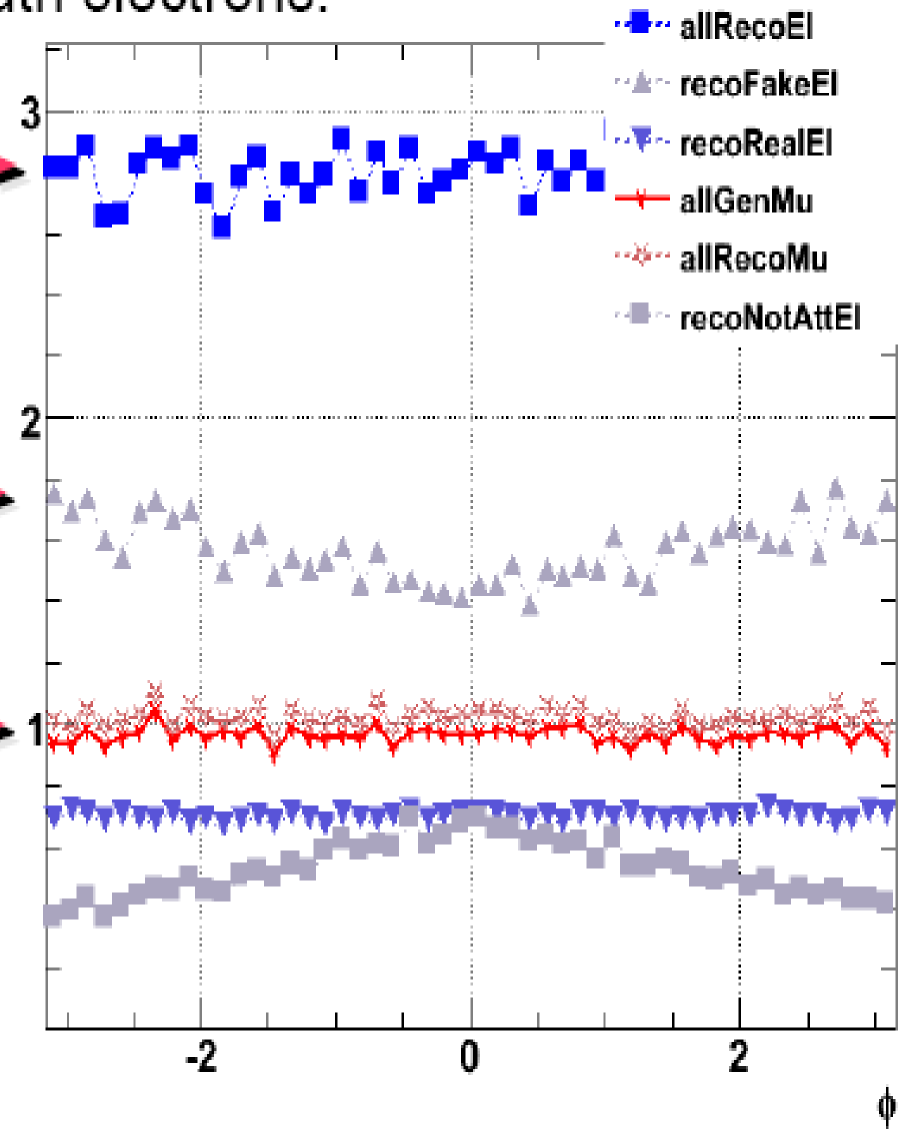
How many fake leptons are out there?

- Normalized to number of MC-truth electrons:

Total number or RECO electrons is almost 3 times greater than number of MC-truth electrons

...because fake rate is at least 1.6 times greater than number of MC-truth electrons

Check using lepton universality:
number of MC-truth electrons
= number of MC-truth muons
= number of RECO muons
The latter is true because muon fake rate is so small and because muon RECO eff. is so high. (Global muons)



For a given background data sample

(this is for ttbar)

The fake rate method

• The goal is to estimate the fake lepton contribution to a signal lepton selection.
• The method consists of a determination and an application part where both parts use statistically independent data samples.

• select a fake dominated background sample

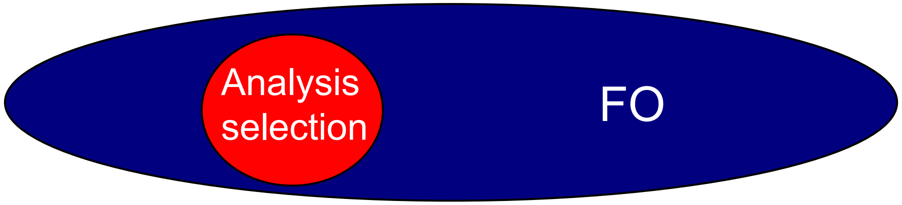
• used to measure the fraction of signal leptons (numerator) to looser-defined leptons (denominator) called the fake rate.

• Loose leptons: Fakeable Object

$$F(p_T,\eta,...) = \frac{\text{analysis object } (p_T,\eta,...)}{\text{denominator object } (p_T,\eta,...)}$$

• measure of how often a fakeable object (e.g. a jet or a loosely matched track) results in a lepton passing your analysis selection cuts (id, isolation, etc.)

• the fake contribution to the signal sample is estimated by weighting the denominator leptons contribution with the fake rate.



The numerator is determined by the analysis lepton selection,
• The choice of the denominator is un-constrained and is usually the key to accuracy and stability of the procedure.

Defining the fake rate

There is more than one way possible to define a fake rate

- Advantages and disadvantages to any method
- All methods will depend on electron selection

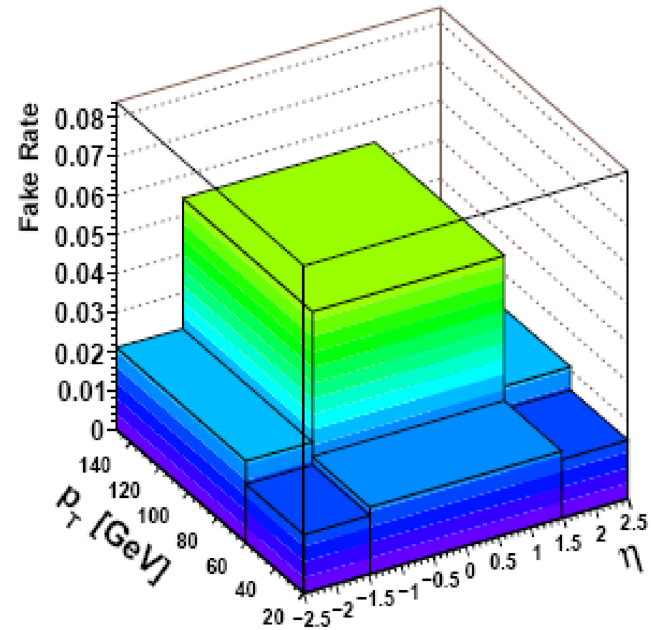
Two common fake rate definitions:

Method 1

- Numerator: jets matched to leptons with tight cuts
- Denominator: jets
- Weakness: numerator is a jet and not a lepton

Method 2

- Numerator: leptons with tight cuts
- Denominator: leptons with loose cuts
- Weakness: more susceptible to contamination from real leptons from W and Z bosons



Estimate QCD lepton fake contribution

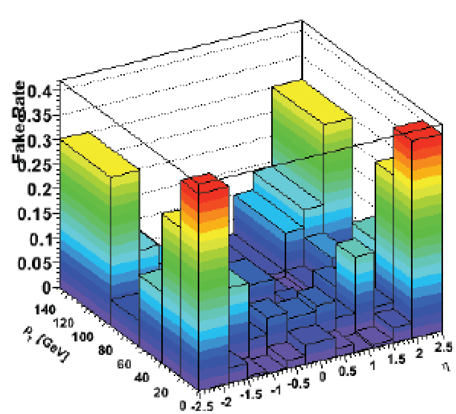
- WW with di-lepton final state
- For this selection the W+jets process is the dominant source of “fake background”
- one real lepton from W and one fake lepton from a jet

Defining the fake rate:

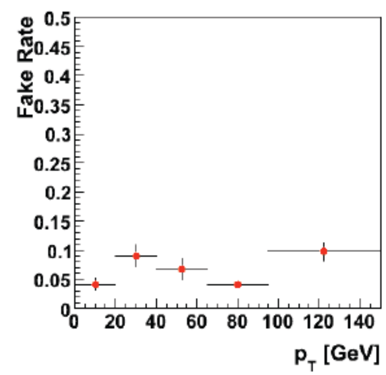
Cut	Denominator	Numerator
transverse momentum	$p_T^e \geq 20$ GeV	$p_T^e \geq 20$ GeV
pseudo rapidity	$\lvert\eta_e\rvert \leq 2.5$	$\lvert\eta_e\rvert \leq 2.5$
closest muon veto	$\Delta R_{e-\mu} > 0.1$	$\Delta R_{e-\mu} > 0.1$
had/elmag energy fraction	$H/E \leq 0.2$	included in electron identification
track based isolation, see B.1	$p_T^e/(p_T^e + tkIso) > 0.92$	
track+calo based isolation, see B.2	$p_T^e/(p_T^e + sumIso) > 0.75$	$p_T^e/(p_T^e + sumIso) > 0.92$
impact parameter		$\lvert d_0^e\rvert \leq 0.025$
electron identification		Tight electron ID (see App. F and [1])

Fake rate determination from the QCD background

2D fake rate

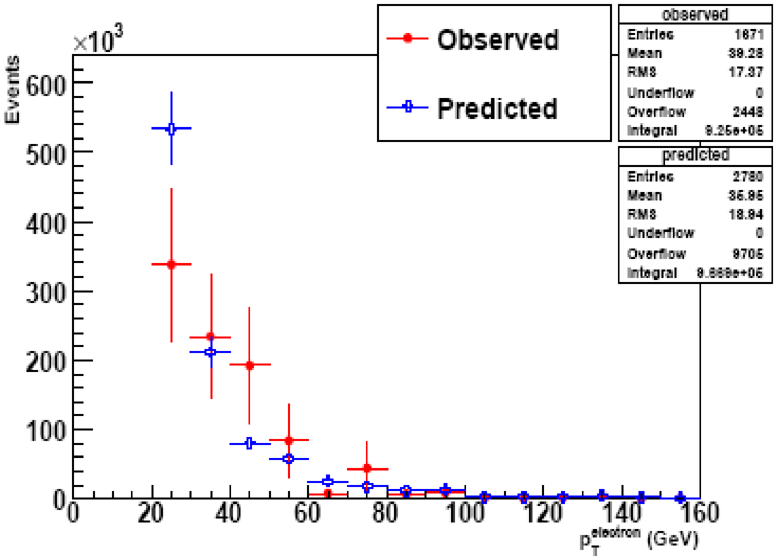


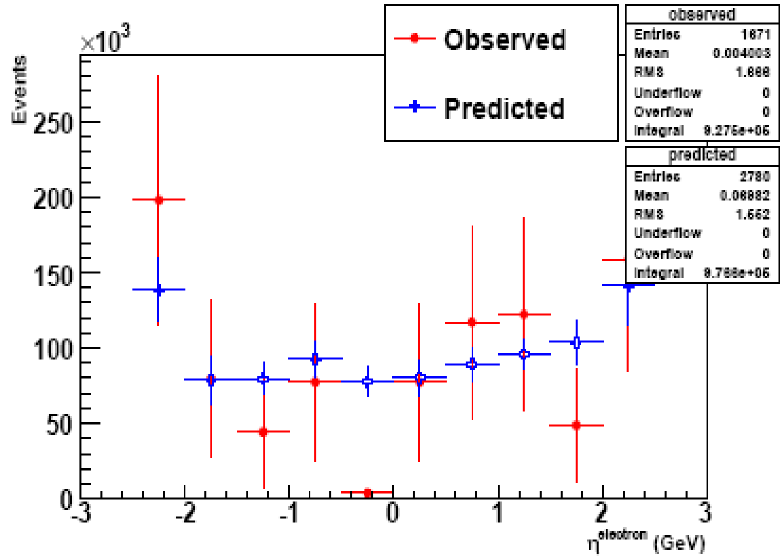
1D fake rate



Fake Rate Test

• As a consistency check, one can test lepton fake prediction on the W+jets background samples.





Observed: numerator selection

Predicted: denominator selection * fake rate

	fake electrons for $1\ fb^{-1}$
observed	927479 ± 176101
predicted	958151 ± 26822

	fake muons for $1\ fb^{-1}$
observed	527992 ± 137421
predicted	509499 ± 49321

Predict fake contribution

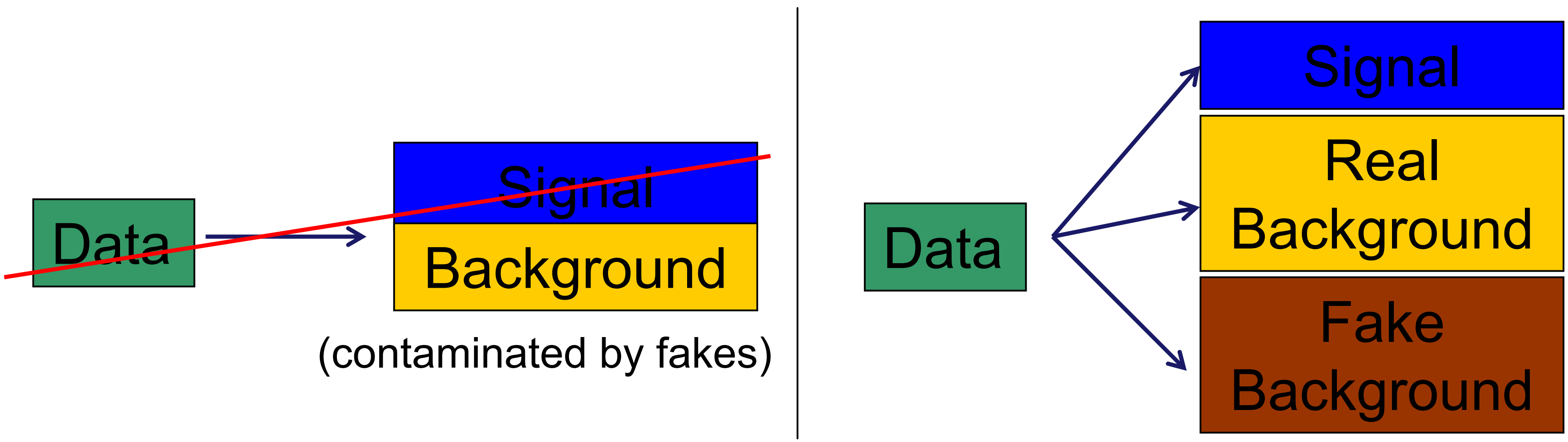
- compare prediction to observed reconstructed leptons in all of the backgrounds

Apply numerator selection yields:

	ww	wz	zz	wjets	dyee	dymm	dytt	ttbar	tw	total
em	344.5 ± 7.3	7.9 ± 1.0	0.3 ± 0.2	48.7 ± 10.8	0.0 ± 0.0	3.4 ± 1.9	22.1 ± 5.5	85.5 ± 6.1	31.1 ± 2.1	543.5 ± 15.7

Apply denominator selection && !numerator selection leptons and multiply by FR/(1-FR) to exclude numerator events yields:

	ww	wz	zz	wjets	dyee	dymm	dytt	ttbar	tw	total
em	8.4 ± 0.6(stat) ± 1.7 (fake)	0.2 ± 0.1(stat) ± 0.1 (fake)	0.1 ± 0.0(stat) ± 0.0 (fake)	50.6 ± 3.7(stat) ± 9.7 (fake)	0.0 ± 0.0(stat) ± 0.0 (fake)	1.1 ± 0.4(stat) ± 0.4 (fake)	2.2 ± 1.0(stat) ± 0.6 (fake)	1.4 ± 0.3(stat) ± 0.3 (fake)	0.3 ± 0.1(stat) ± 0.1 (fake)	64.3 ± 3.9



Useful docs:

Search Strategies for mSUGRA in the Muons + Jets + MET Final State (12/06/2008)

Data-driven methods to estimate the electron and muon fake rate contributions to lepton analyses (21/03/2009)

Prospects for measuring the WW production cross section in pp collisions at sqrt(s)=10 Tera eV (21/06/2009)

Expectations for observation of top quark pair production in the dilepton final state with the early CMS data (10/06/2009)

Search for Supersymmetry with Trimuons (30/05/2009)

The whole collection can be found in our twiki page:

http://www.grid.kfki.hu/twiki/bin/view/CMS/SusySearch

http://www.grid.kfki.hu/twiki/bin/view/CMS/HUSUSYFAKES

TO DO

- upload relevant talks to the twiki page (this afternoon)
- understand deeper the physics and methods
- explore the RA4 and top_stop signal
- electrons or muons?

- expecting advices…