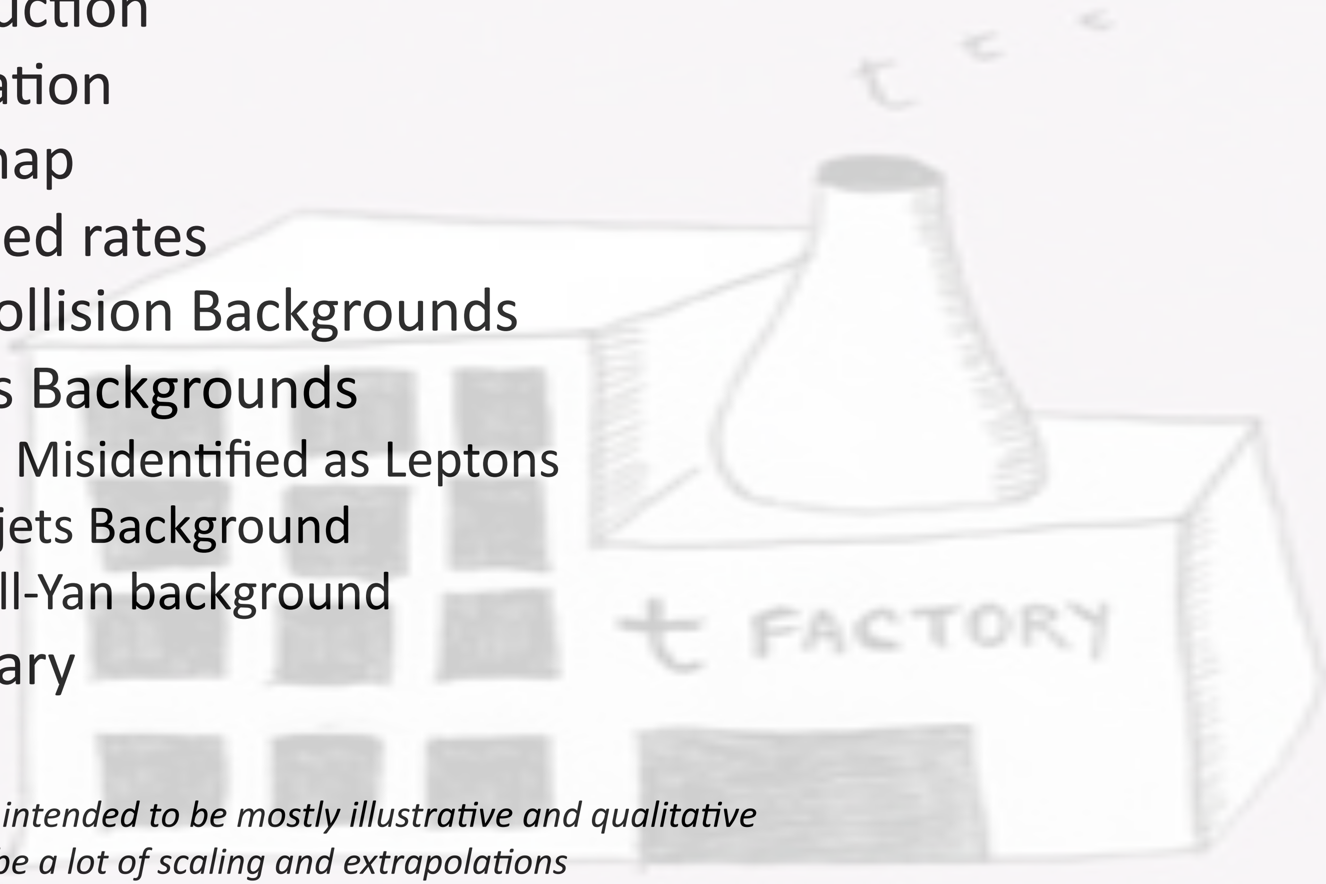
Backgrounds estimation strategies in ATLAS

Andrey Loginov
Yale University
For the ATLAS Collaboration

Outline

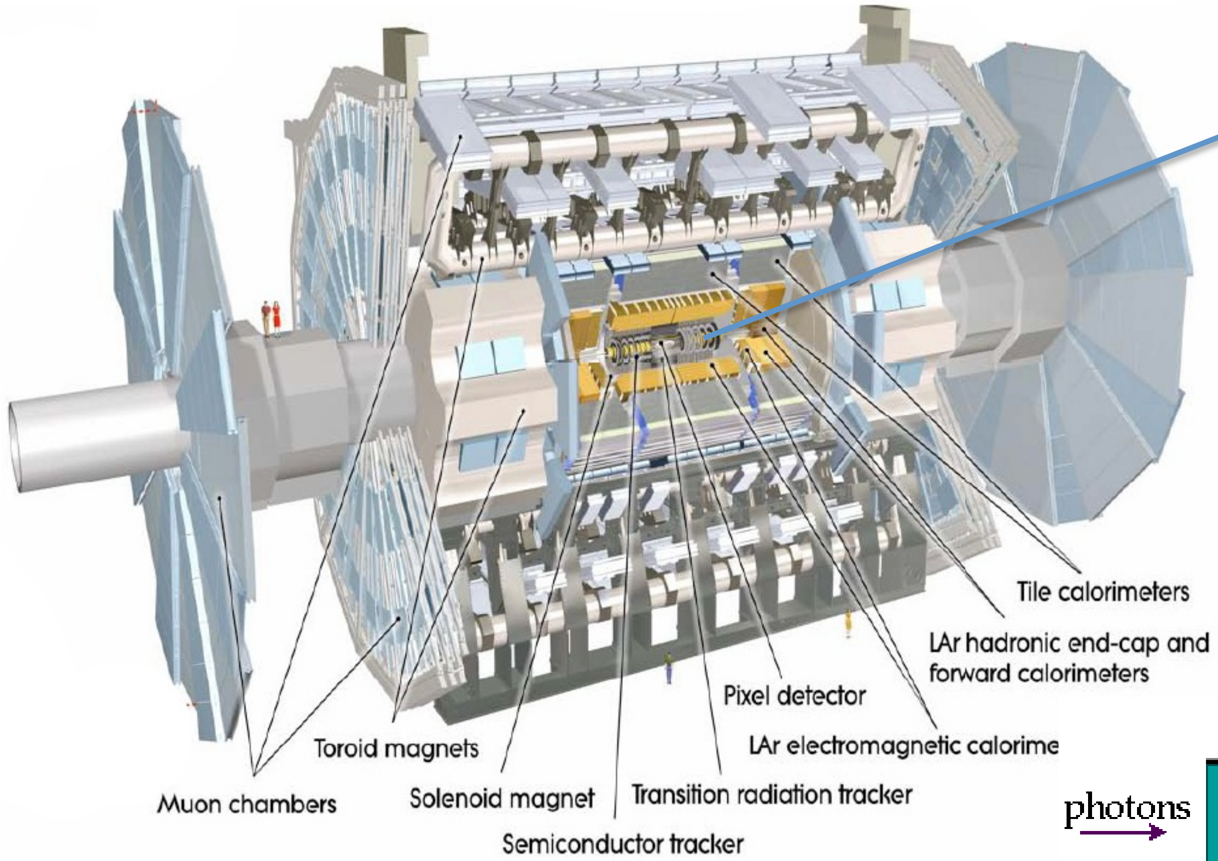
- ATLAS
- Introduction
- Motivation
- Roadmap
- Expected rates
- Non-Collision Backgrounds
- Physics Backgrounds
  - Jets Misidentified as Leptons
  - W+jets Background
  - Drell-Yan background
- Summary

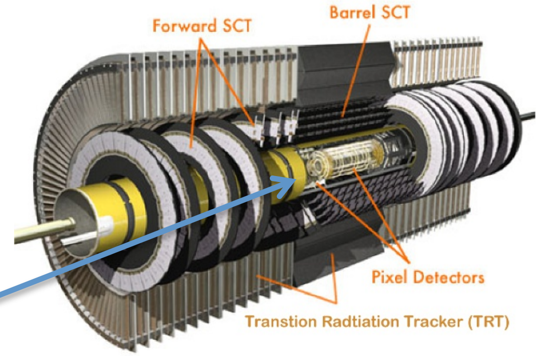


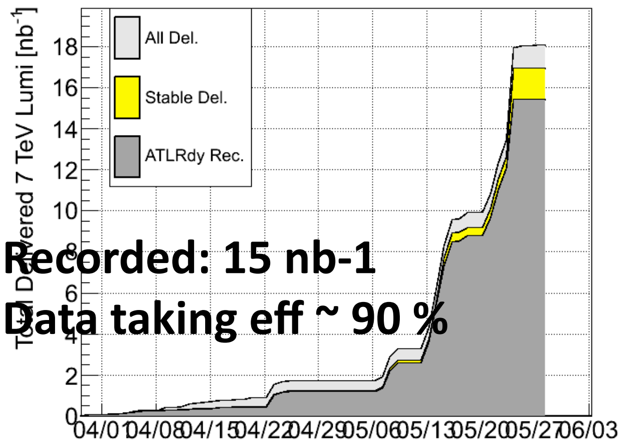
Disclaimer:

- *This talk is intended to be mostly illustrative and qualitative*
- *There will be a lot of scaling and extrapolations*
- *I will omit far more than I can include*
- *Will concentrate on data-driven techniques*

ATLAS



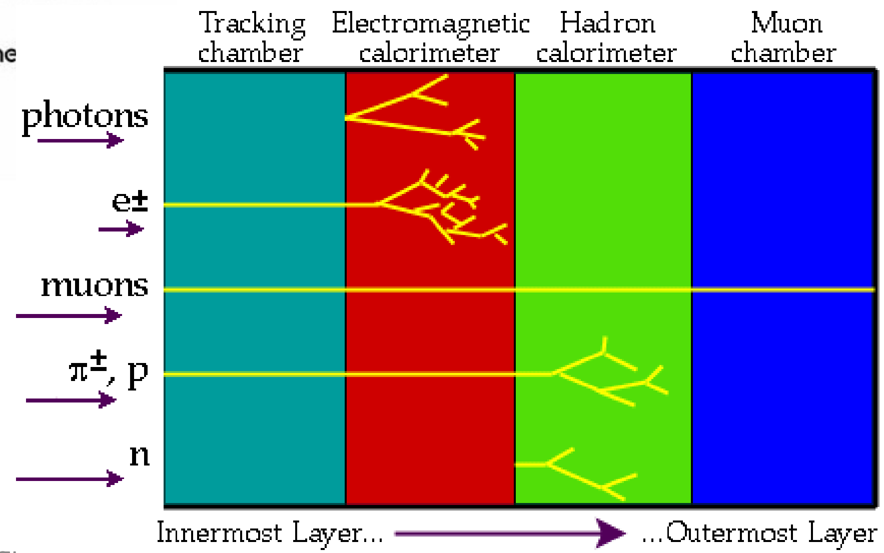




Recorded: 15 nb-1

Data taking eff ~ 90 %

Expect a few pb^{-1} by the end of summer 2010

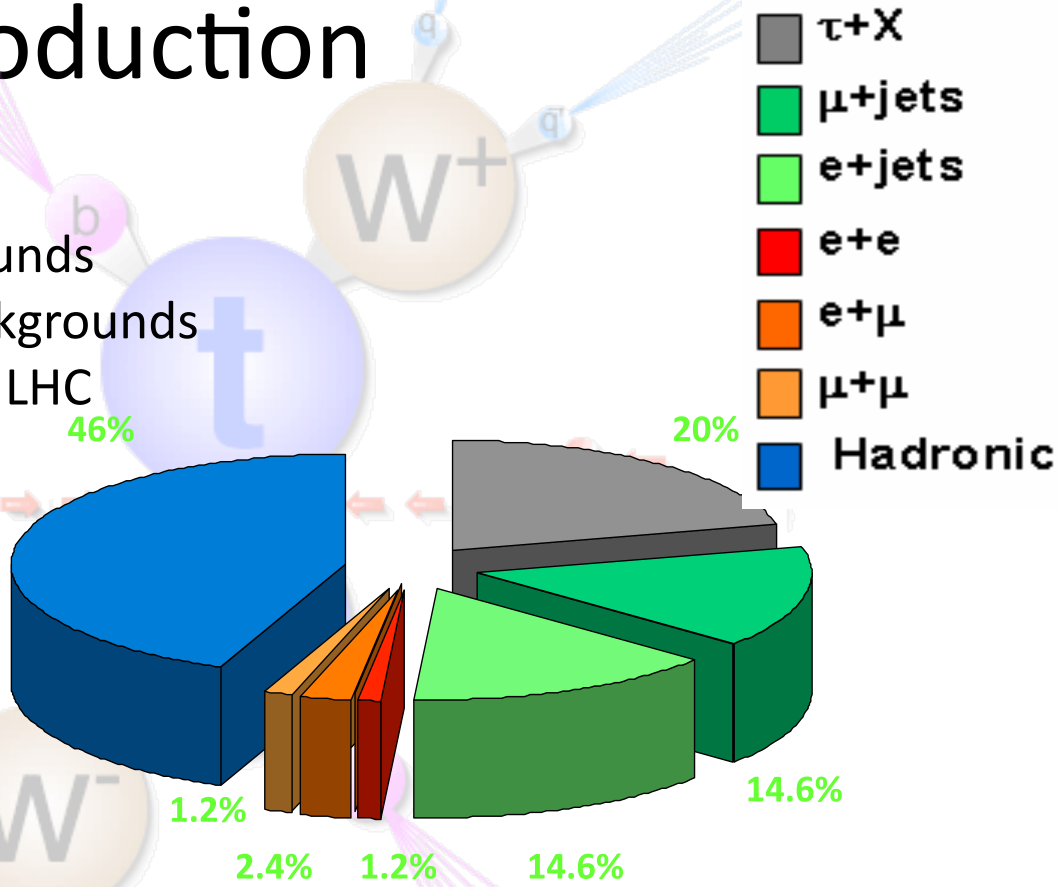


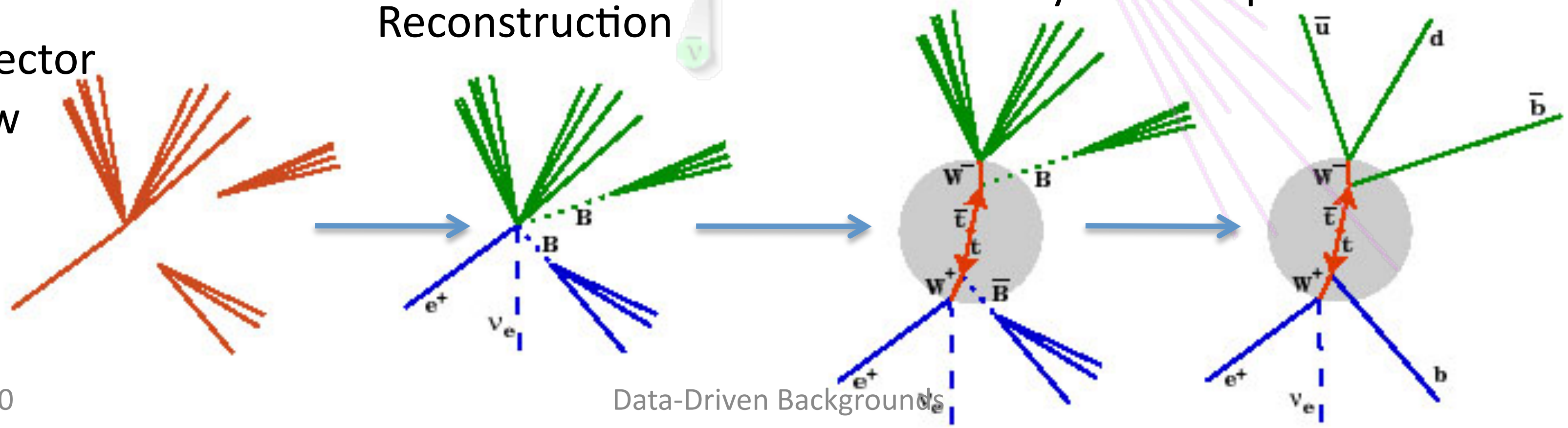
Detectors and identified objects:

- Trackers: e, mu, jets, photons
- Calorimeters: e, mu, jets, MET, photons
- Muon Detectors: mu

Introduction

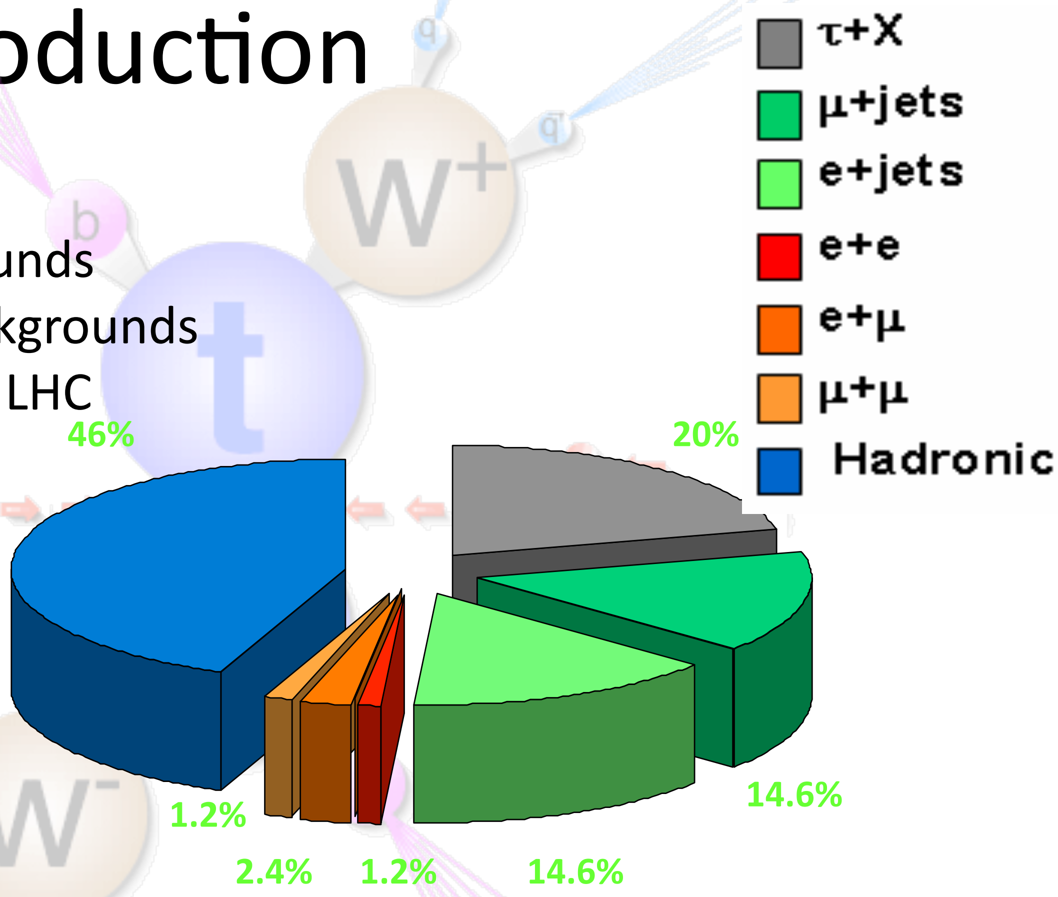
- Dilepton: small BR, smallest backgrounds
- Lepton+Jets: large BR, moderate backgrounds
- Leptons are the key to top physics at LHC (trigger and offline)
- Study the processes that give rise to the ttbar-like dilepton and lepton + jets signatures

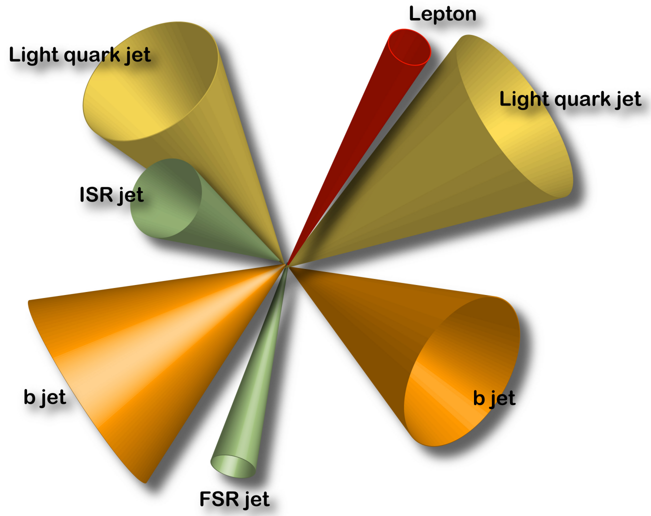


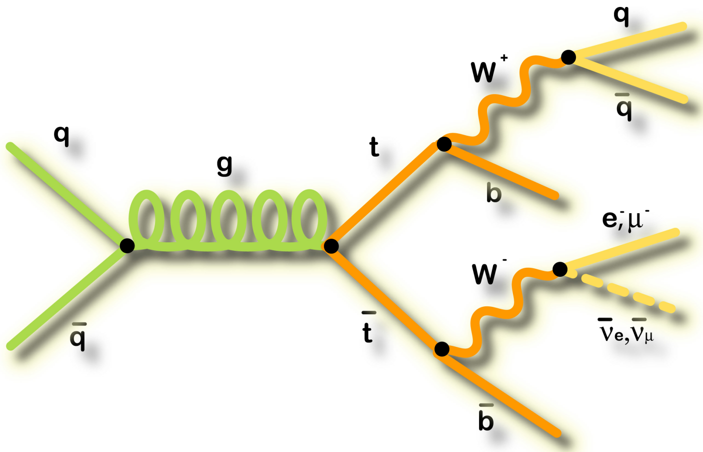


Introduction

- Dilepton: small BR, smallest backgrounds
- Lepton+Jets: large BR, moderate backgrounds
- Leptons are the key to top physics at LHC (trigger and offline)
- Study the processes that give rise to the ttbar-like dilepton and lepton + jets signatures





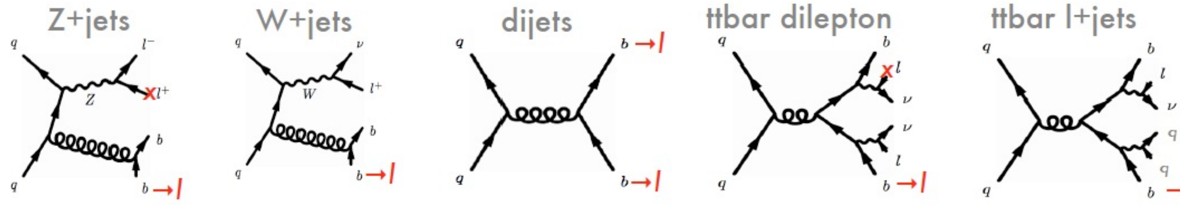


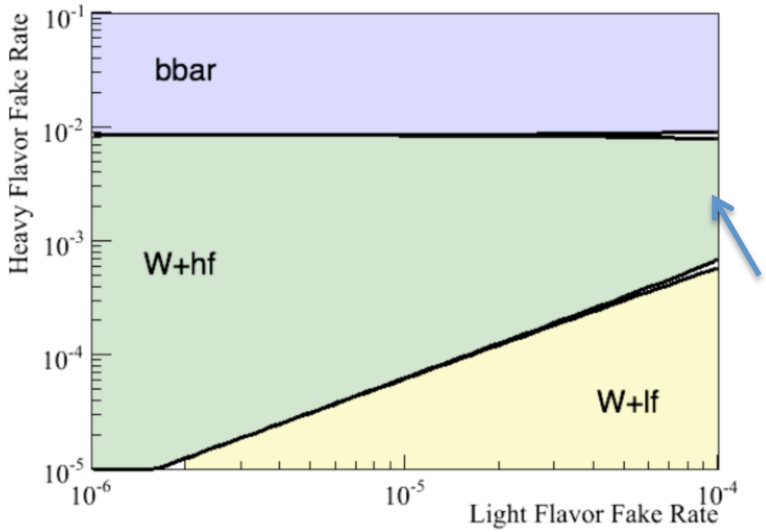
Motivation

- Why data-driven methods?
  - If the MC were perfect, no need, but at some level, detector simulation, Matrix elements, PDF's, will never be perfect, and proving they are good enough can be very hard
- Early days: establish confidence in background estimates without relying on
  - *MC description of material, detector performance and complicated backgrounds: W+jets, Z+jets, multi-jets (QCD).*
  - *Simulating many events (QCD)*
- *Estimate and/or reduce backgrounds*

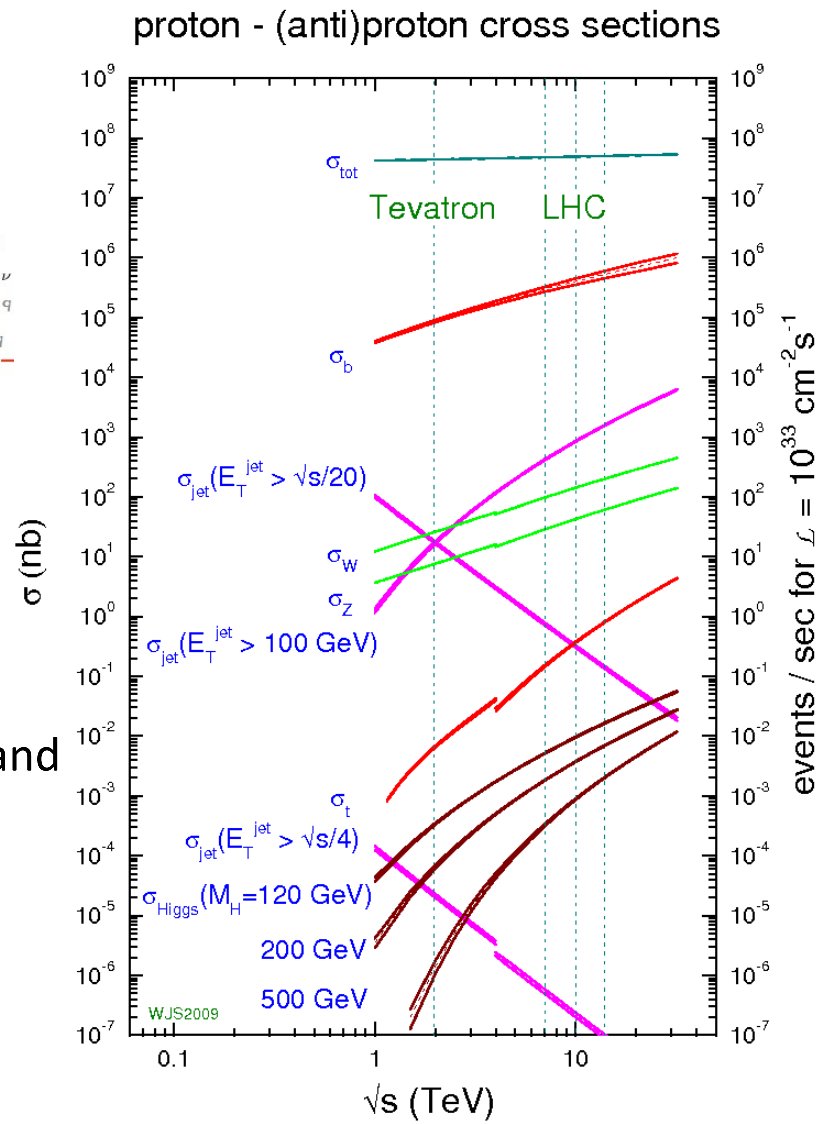
Motivation

- Jets misID as leptons
  - W+jets bkg for dileptons
  - QCD bkg for l+jets / dileptons





Different Heavy Flavor and Light Flavor fake rates

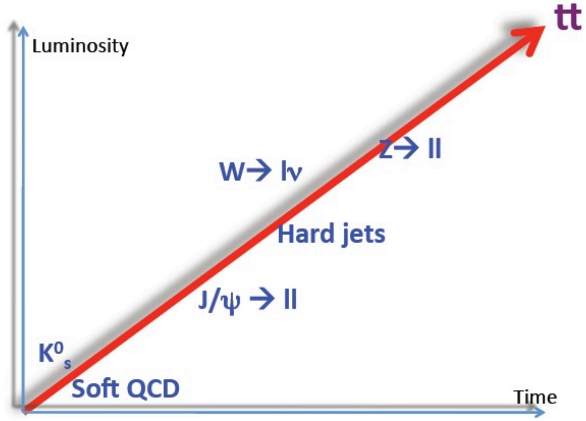


- Fake MET (mismeasured jets / muons)
  - Drell-Yan (Z/γ^*) bkg for dileptons
- W+jets for l+jets

Roadmap

- Begin by understanding the basic objects:
  leptons, jets, b-tagging, Missing transverse energy (MET)
- Continue by combining them to low-mass resonances
- Establish W signal, study vs N jets, estimate QCD background
- Measure ID and trigger lepton efficiencies, study backgrounds to top in W +jets and Z+jets samples
- Establish Z signal, study vs N jets, estimate W+jets / QCD backgrounds
- Validate MET in complex final states
- Look for the emergence of a top signal
- Study top final states (background for many searches for physics beyond the Standard Model)
- Higgs and / or beyond the Standard Model

Peter Maettig



Roadmap

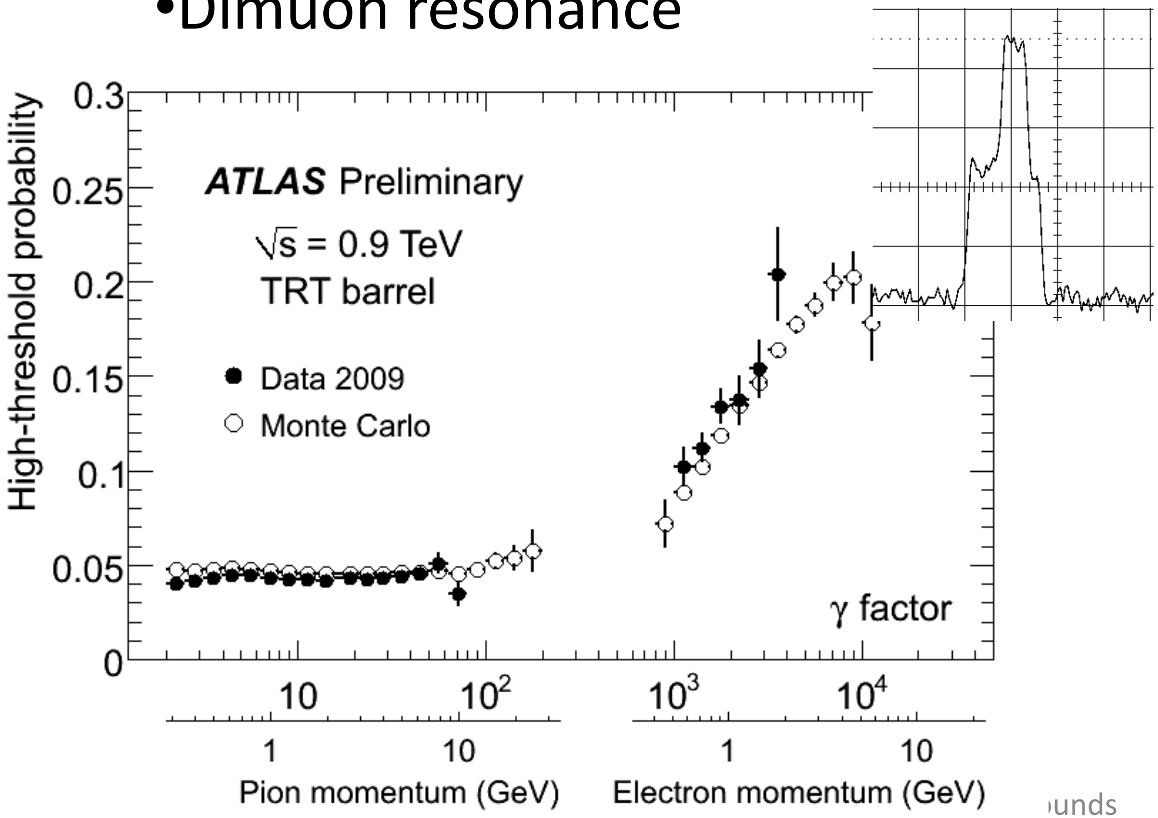
Ivo van Vulpen's talk

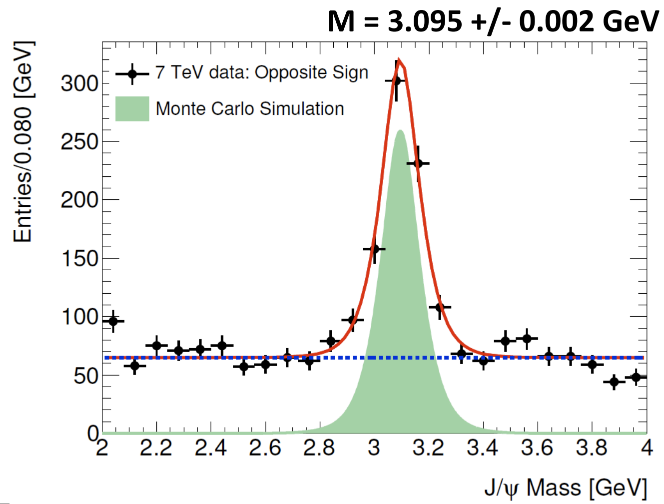
Nabil Ghodbane's talk

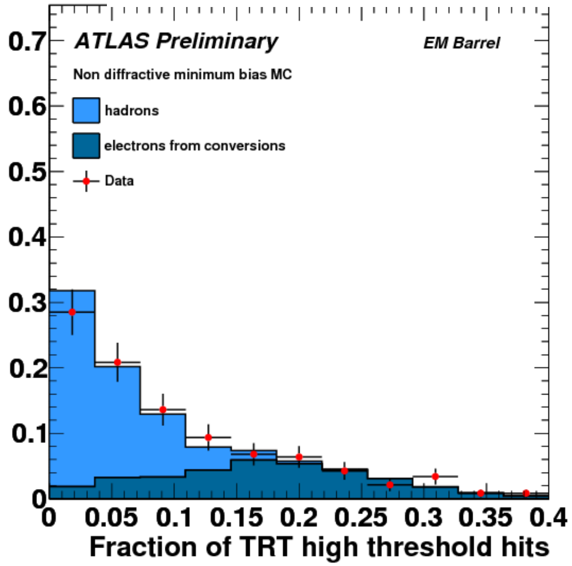
- **Begin by understanding the basic objects: jets, leptons, HF, MET**
- Many plots – see Ivo van Vulpen's talk. I will just show few favorites:

•High Threshold TRT hits (e ID)

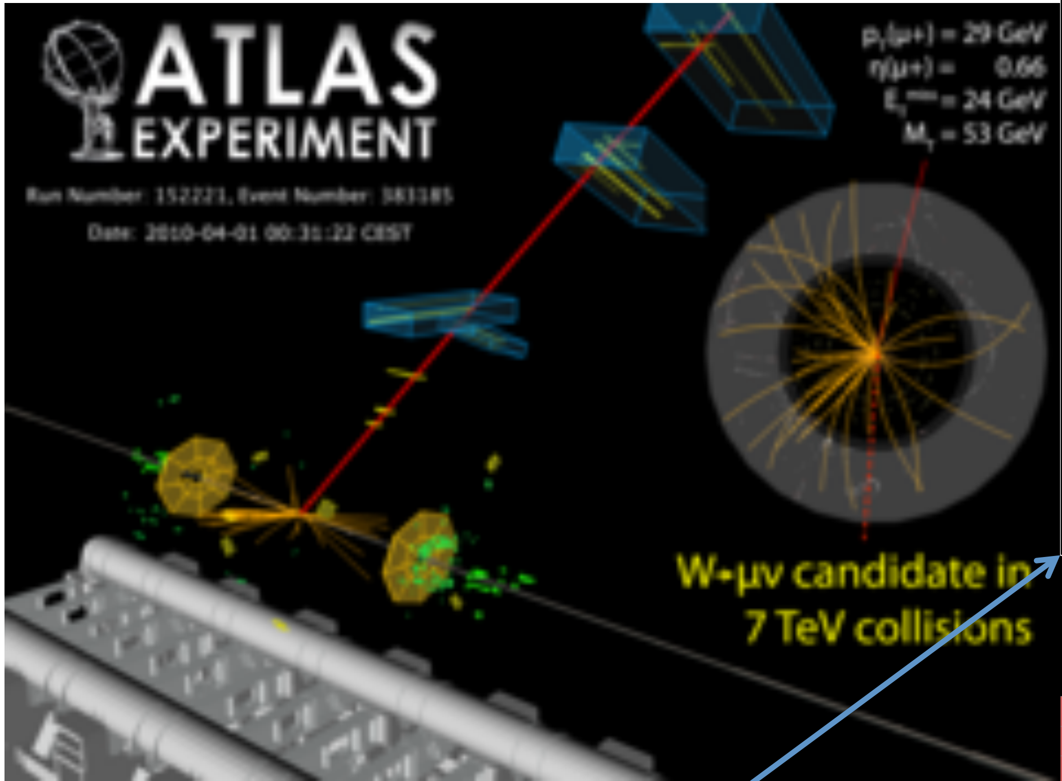
•Dimuon resonance

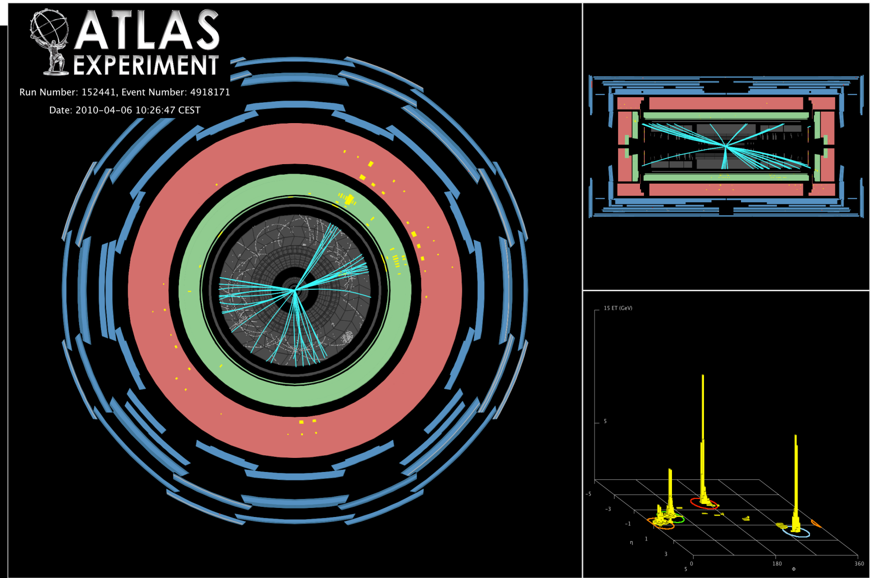


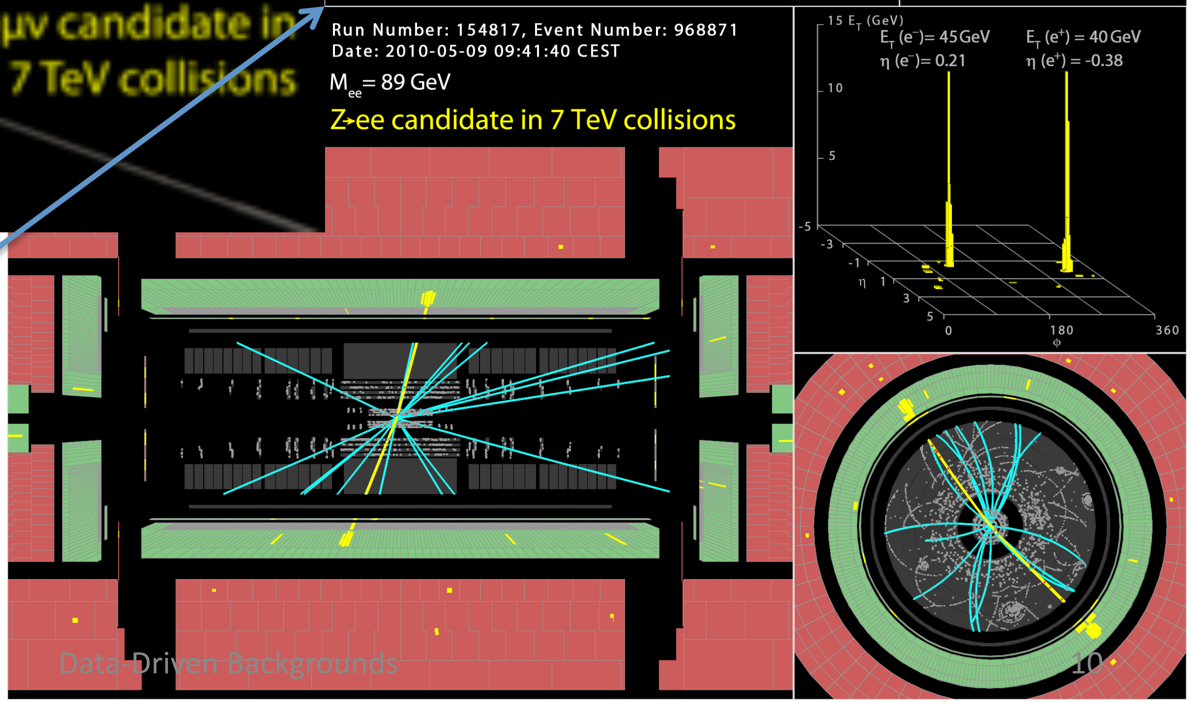




Roadmap







On the way to

- **W+jets**
- **QCD background estimates**
- **Z+jets**

Expected Rates

Will show results based on ATLAS top cross-section 10 TeV MC notes for

- Dilepton channel
- Single Lepton channel

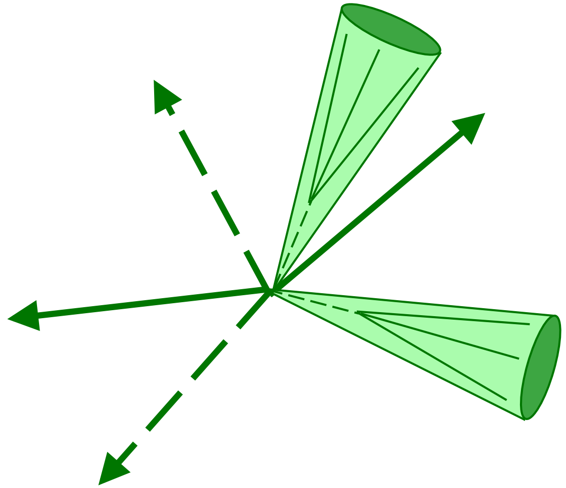
- Details are in
  - ATL-PHYS-PUB-2009-086
  - ATL-PHYS-PUB-2009-087
  - Also shown on the poster of M.Saleem

Dilepton: Expected Rates

$$t\bar{t} \rightarrow \ell^+ \nu \ell^- \nu b \bar{b}$$

$$\ell = e, \mu$$

- 2 opposite sign (OS) leptons, pt>20 GeV
- At least 2 jets, pt>20 GeV
- MET>35 for the ee and mumu, 20 for the emu
- Remove events dilepton inv.mass close to the Z



$$\ell^+ \ell^- jj + MET(t\bar{t}, diboson, Z \rightarrow \tau\tau ...)$$

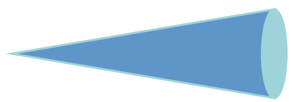
$$\ell^+ (j \rightarrow \ell^-) jj + MET$$

$$(j \rightarrow \ell^+) \ell^- jj + MET$$

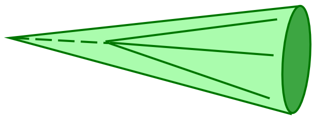
$$(j \rightarrow \ell^-)(j \rightarrow \ell^+) jj + MET$$

$$\ell^+ \ell^- jj + FakeMET$$


Electron or muon


Jet: shower of particles

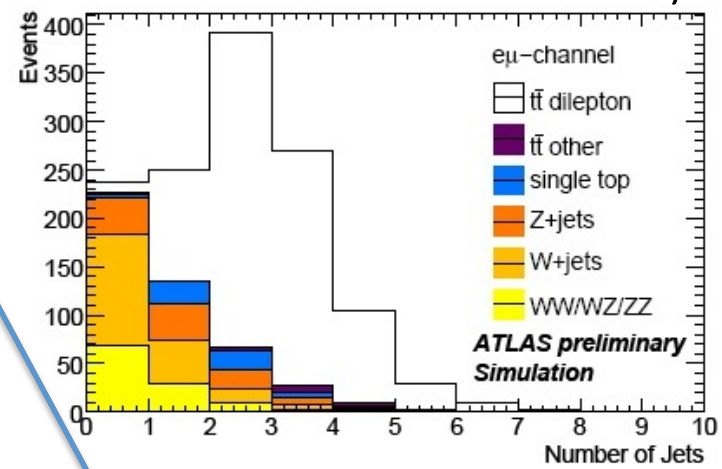

Neutrino: Missing E_T


b-jet: identified with secondary vertex tag

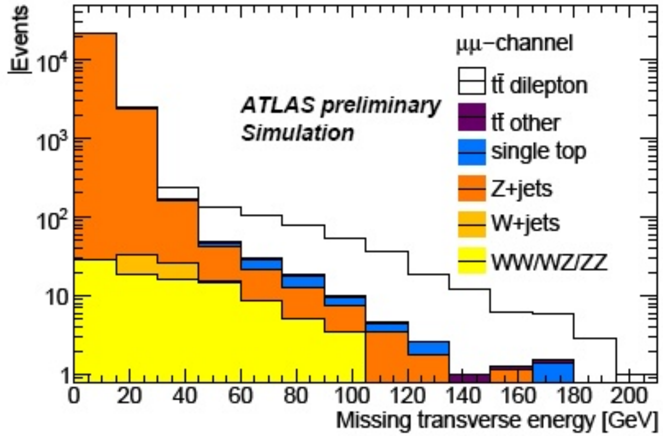
Dilepton: Expected Rates

- 2 opposite sign (OS) leptons, pt>20 GeV
- At least 2 jets, pt>20 GeV
- MET>35 for the ee and mumu, 20 for the emu
- Remove events dilepton inv.mass close to the Z
- **We show expected rates at 200pb-1 at 10 TeV**

ATL-PHYS-PUB-2009-086)



ATL-PHYS-PUB-2009-086)



$\Delta\sigma/\sigma$ (%)	ee channel	$\mu\mu$ channel	$e\mu$ channel	combined
Stat only	-7.5 / 7.8	-6.0 / 6.2	-4.0 / 4.1	-3.1 / 3.1
Luminosity	-17.3 / 26.3	-17.4 / 26.2	-17.4 / 26.2	-17.4 / 26.2
Electron Efficiency	-4.5 / 5.0	0.0 / 0.0	-2.2 / 2.4	-1.9 / 1.9
Muon Efficiency	0.0 / 0.0	-4.6 / 5.2	-2.1 / 2.2	-2.2 / 2.3
Lepton Energy Scale	-0.3 / 1.6	-2.4 / 2.0	-0.5 / 0.5	-0.8 / 0.8
Jet Energy Scale	-3.4 / 3.2	-3.0 / 4.5	-2.5 / 2.5	-2.8 / 3.0
PDF	-2.1 / 2.3	-1.4 / 1.6	-1.6 / 1.8	-1.7 / 1.8
ISR FSR	-4.0 / 4.2	-3.6 / 3.7	-3.5 / 3.5	-3.6 / 3.7
Signal Generator	-4.7 / 5.4	-4.6 / 5.4	-4.7 / 5.3	-4.7 / 5.3
Cross-Sections	-0.3 / 0.3	-0.3 / 0.3	-0.3 / 0.3	-0.3 / 0.3
Drell Yan	-1.4 / 1.3	-2.2 / 2.2	-0.5 / 0.5	-0.8 / 0.9
Fake Rate	-9.7 / 9.5	-1.1 / 1.1	-6.2 / 6.2	-4.0 / 4.0
All syst but Luminosity	-12.7 / 13.9	-8.9 / 10.2	-9.4 / 10.2	-8.7 / 9.6
All systematics	-21.0 / 30.3	-19.3 / 28.3	-19.5 / 28.5	-19.3 / 28.1
Stat + Syst	-22.3 / 31.3	-20.2 / 29.0	-19.9 / 28.8	-19.5 / 28.3

Jet->lepton systematics 50%
DY systematics 15% (sensitive to statistics)

What about 10 pb-1 at 7 TeV ?

Dilepton: Expected Rates

- 2 OS leptons, pt>20 GeV
- At least 2 jets, pt>20 GeV
- MET>35 for the ee and mumu, 20 for the emu
- Remove events dilepton inv.mass close to the Z
- **We scale rates to 10pb-1 at 7 TeV**
- (based on ATL-PHYS-PUB-2009-086)

Expected 10 pb^{-1} sensitivity (per experiment)

Channel	N(Signal)	N(background)
$e - \mu$	14	2.5
$e - e$	4.3	1.1
$\mu - \mu$	6.6	1.9
Total	**25**	**5.5**

ATL-PHYS-PUB-2009-086 + scaling to 10 pb^{-1} @ 7 TeV.

DY bkg:

0.5 for the ee,
2 for the mumu,
1 for emu

Fakes:

1.0 for the ee
0.5 for the mumu
1.5 for emu

Most of the backgrounds will be estimated in a data-driven way

When using data-driven background estimates for fake leptons, care must be taken to remove double counting with backgrounds measured from MC

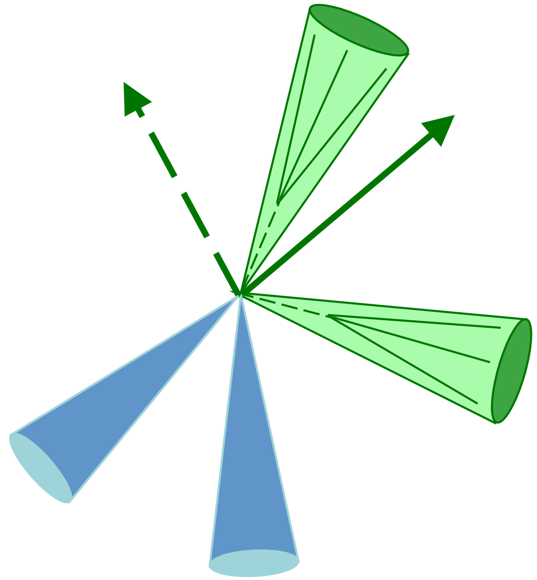
Can trade off more signal for lower S/B in order to optimize, so we might have different selection criteria than those I have shown.

Lepton+Jets: Expected Rates

- 1 lepton, pt>20 GeV
- At least 4 jets pt>20 GeV, 3 jets pt>40 GeV
- MET>20 GeV

$$t\bar{t} \to \ell \nu q q b \bar{b}$$

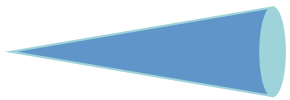
$$\ell = e, \mu$$



$$\ell jjjj + MET(t\bar{t}, W + jets, \ldots)$$

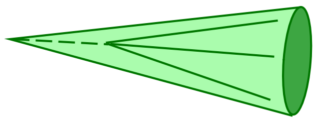
$$(j \to \ell)\, jjjj + MET(QCD)$$

Electron or muon



Jet: shower of particles

Neutrino: Missing E_T



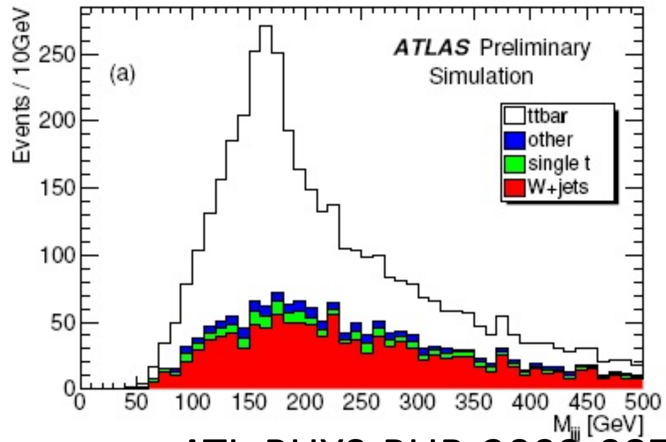
b-jet: identified with secondary vertex tag

Lepton+Jets: Expected Rates

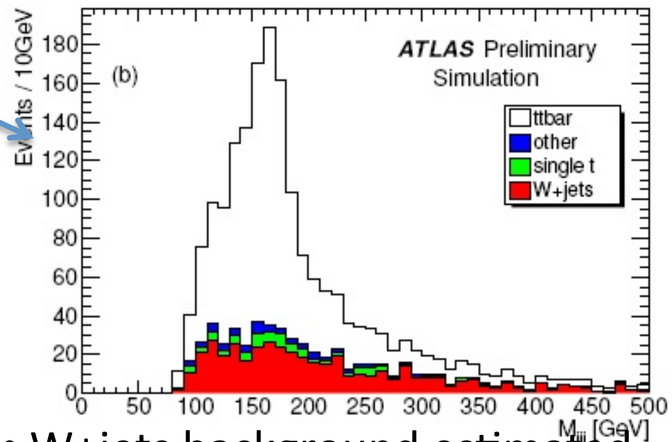
- 1 lepton, pt>20 GeV
- At least 4 jets pt>20 GeV, 3 jets pt>40 GeV
- MET>20 GeV
- **We show expected rates at 200pb-1 at 10 TeV**

	Cut and Count method			
Source	*e*-analysis		μ-analysis	
	default (%)	$+M_W$-cut (%)	default (%)	$+M_W$-cut (%)
Stat.	± 2.5	± 3.4	±2.3	±3.1
Lepton ID eff.	±1.0	**±1.0**	±1.0	**±1.0**
Lepton trig. eff.	±1.0	**±1.0**	±1.0	**±1.0**
50% W+jets	±25.1	±17.4	±28.1	±19.8
20% W+jets	±10.0	**±7.0**	±11.2	**±7.9**
JES (10%,-10%)	+24.8-23.4	+15.9-19.1	+20.5-22.3	+11.9-17.9
JES (5%,-5%)	+12.3-11.9	**+8.6-9.3**	+10.4-10.9	**+6.1-8.4**
PDFs	±1.6	**± 1.9**	±1.2	**± 1.4**
ISR/FSR	+9.1-9.1	**+7.6-8.2**	+8.2-8.2	**+5.2-8.3**
Signal MC	±3.3	**±4.4**	±0.3	**±2.8**
Back. Uncertainty	±0.6	**±0.4**	±0.5	**±0.4**
Fitting Model	-	-	-	-
10% Lumi.	±11.6	±11.2	±11.4	±11.1
20% Lumi.	±23.2	**±22.3**	±22.8	**±22.2**
Tot. without Lumi.	+18.8-18.5	**+14.4-15.2**	+17.5-17.7	**+11.9-14.7**

ATL-PHYS-PUB-2009-087



ATL-PHYS-PUB-2009-087)



- W+jets background estimated with W/Z ratio method
- QCD not shown: expect to be smaller than W+jets

Lepton+Jets: Expected Rates

- 1 lepton, pt>20 GeV
- At least 4 jets pt>20 GeV, 3 jets pt>40 GeV
- MET>20 GeV
- **We show expected rates at 10pb-1 at 7 TeV**

  (based on ATL-PHYS-PUB-2009-087)

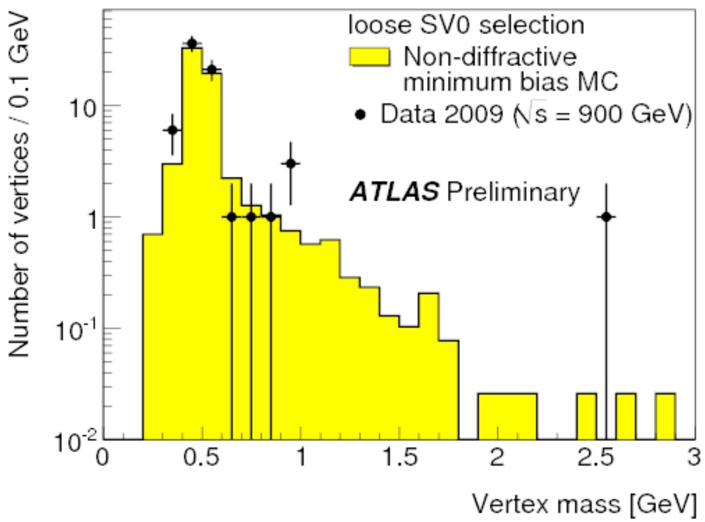
Numbers of Selected Events				
	Electron Analysis		Muon Analysis	
Sample	default	$+M_W$-cut	default	$+M_W$-cut
$t\bar{t}$	53	26	64	32
W+jets	29	10	40	14
single top	5	2	25	2
Other	6	2	6	2

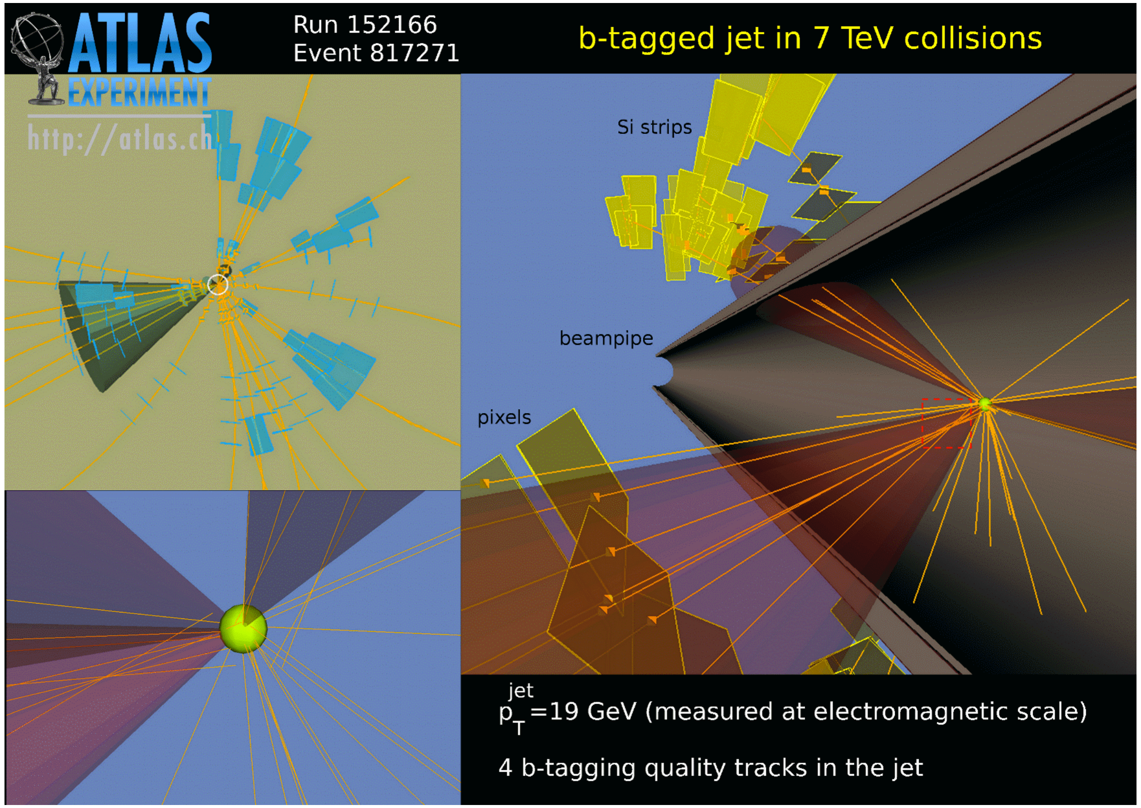
- no b-tagging
- QCD: Work focus on how to estimate it from data, and how to reduce it if rates higher than expected

Can trade off more signal for lower S/B in order to optimize, so we might have different selection criteria than those I have shown.

Do not depend on b-tagging, but…

ATLAS is studying b-tagging with the data in hand.
– Many tagged jets have been found, sometimes correlated with nearby leptons or second tags in the event

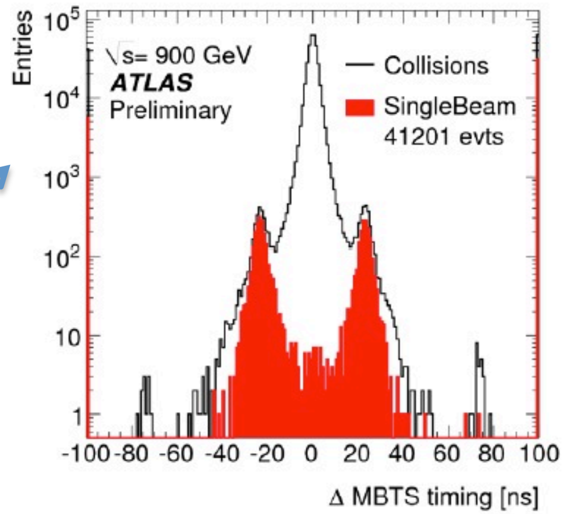


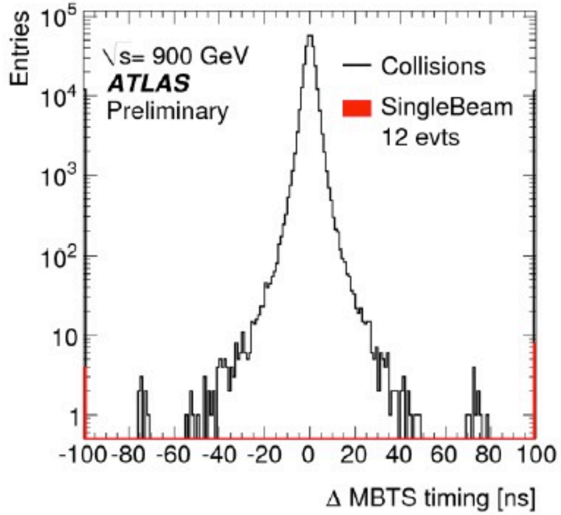


Will suppress W+jets, QCD, Drell-Yan backgrounds

Non-Collision Backgrounds

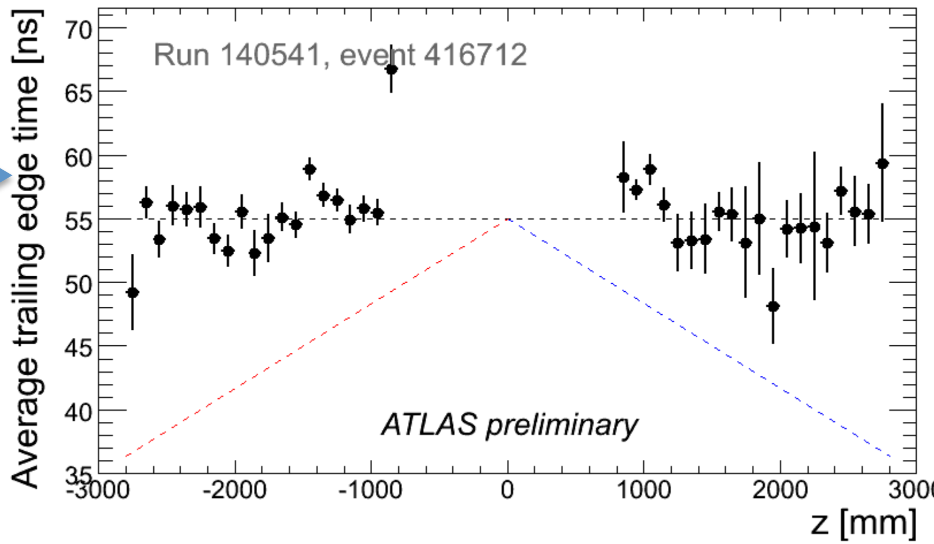
- **Beam Induced**
  - **Beam Halo mu's**
  - **Beam-Gas collisions**
- **Cosmic rays**





Handles:

- MBTS (MinBias Trigger Scintillators) timing
- Liquid Argon Calorimeter timing
- TRT EndCap timing



We have good handles to reject the backgrounds, and reasonably confident from our data analysis so far, that they will not be a big background for top analyses.

Crucial for MET and muons

Physics “Fakes” Backgrounds

Sources of fake lepton background

muons:

- Decays in flight of π or K meson in light jets
- Punch through muons
- Semileptonic decay in a heavy flavor jet

electron:

- Jet with high EM fraction
- Photon associated with a track, conversions
- Semileptonic decay in a heavy flavor jet

Semileptonic channel

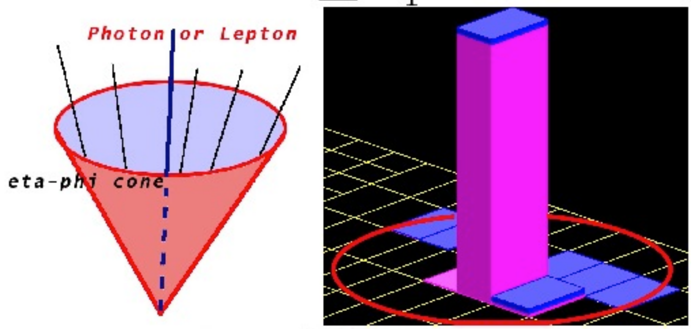
- QCD multijet
- gamma+jet

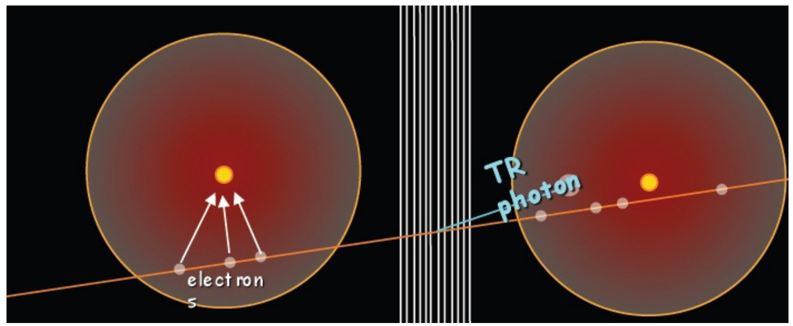
Dileptonic channel

- W+jets: jet fakes isolated
- QCD multijet: two jets fake isolated leptons

Tracking Iso: $\sum \mathrm{p}_\mathrm{T}^{track}$ in $\eta - \phi$

Calorimeter Iso: $\sum \mathrm{E}_\mathrm{T}^{tower}$ in $\eta - \phi$

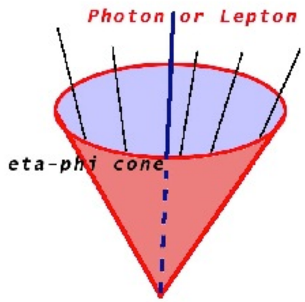


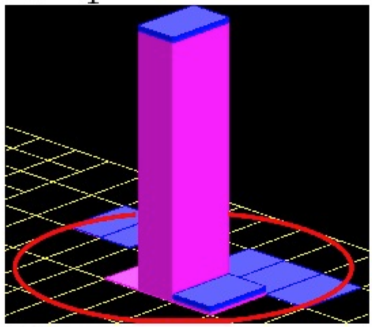


Physics "Fakes" Backgrounds

Tracking Iso: $\sum \mathrm{p}_\mathrm{T}^{track}$ in $\eta - \phi$

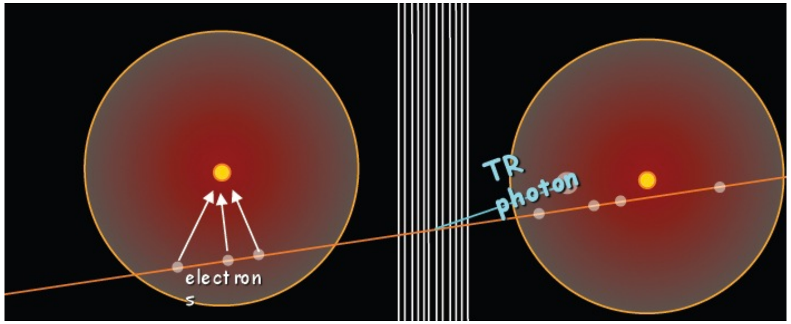
Calorimeter Iso: $\sum \mathrm{E}_\mathrm{T}^{tower}$ in $\eta - \phi$





Handles:

- calorimeter and track isolation (e and mu)
  - Track iso is less sensitive to the pile-up
- d0 for the mu's and for the e's
  - Small d0 for the prompt e's and mu's, large d0 for HF
- TRT High Threshold (HT) hits (e)
  - Large HT fraction for e's, low HT fraction for hadrons
- Number of hits in different parts of the ID (pixels, SCT, TRT)
  - Conversions will have no hits in the pixels
- DeltaPhi between MET and leptons, jets
  - Same or Opposite direction for fake MET



Control Samples / Checks

- Dileptons
  - OS / SS (*) + 0 / 1 jets
  - OS + 2-or-more jets in Z mass window
  - SS + 2-or-more jets
  - With and/or without MET
- Lepton+Jets
  - *l*+MET+0/1/2/3 jets
- “Stability Plots”
  - Rates of leptons+jets+X and dilepton+X per pb-1

    help check the stability of the detector calibration, data quality selections etc

____

(*) OS: Opposite Sign leptons, SS: Same Sign leptons

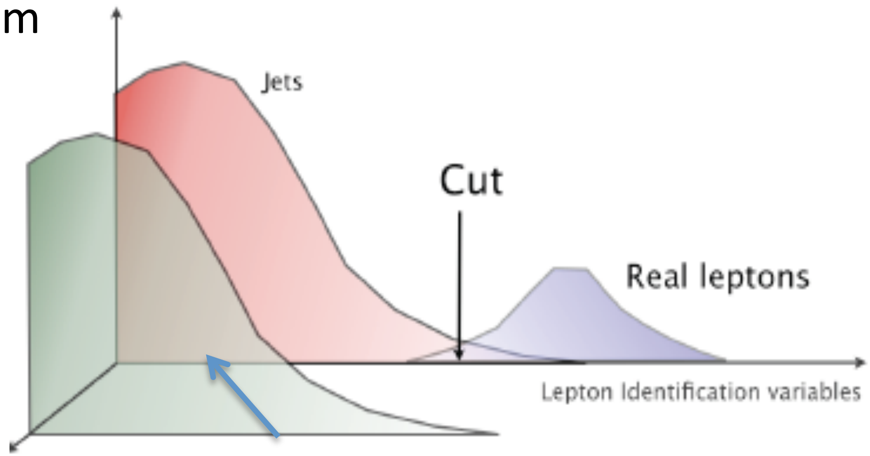
Data-Driven Estimates

- Jets misID as leptons
  - W+jets bkg for the ttbar dileptons
  - QCD bkg for the ttbar l+jets
- Fake MET (mismeasured jets / muons)
  - DY bkg for the ttbar dileptons
- W+jets for l+jets

Jets misID as leptons

"Fake": not a real lepton or a lepton not from a particle of interest

How much background passes the cut?



Orthogonal selection uncorrelated with first variable

Problem I: measure background fake rate

- Rates vary with region
- background region contains signal
- parametrization is limited

Problem 2: different components (LF, HF)

- with different rates
- different fake estimate calculations

- Many solutions exist from previous experience
- Some new facets for LHC

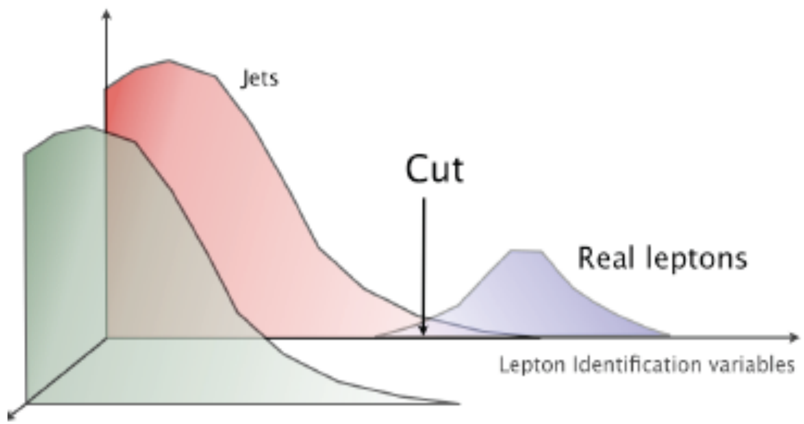
Matrix Method

$$N^{\text{loose}} = N^{\text{loose}}_{\text{real}} + N^{\text{loose}}_{\text{fake}},$$
$$N^{\text{tight}} = \varepsilon_{\text{real}} N^{\text{loose}}_{\text{real}} + \varepsilon_{\text{fake}} N^{\text{tight}}_{\text{fake}}$$

l+jets

two lepton selections:
"Real" (from Z sample)
"Fake" (from QCD sample)

Define a loose and tight isolated lepton sample
Determine the probability to find isolated lepton in each sample



Efficiencies must be:
- sufficiently different
- relatively independent of the event topology
-parameterized as a function of the relevant kinematic variables
- Independent of MET

The definition of the loose sample can be adjusted in order to meet most of the above conditions
Contamination from real leptons of control sample has to be taken into account

dileptons

$$\begin{bmatrix} N_{TT} \\ N_{TL} \\ N_{LT} \\ N_{LL} \end{bmatrix} = \begin{bmatrix} rr & rf & fr & ff \\ r(1-r) & r(1-f) & f(1-r) & f(1-f) \\ (1-r)r & (1-r)f & (1-f)r & (1-f)f \\ (1-r)(1-r) & (1-r)(1-f) & (1-f)(1-r) & (1-f)(1-f) \end{bmatrix} \begin{bmatrix} N_{RR} \\ N_{RF} \\ N_{FR} \\ N_{FF} \end{bmatrix}$$

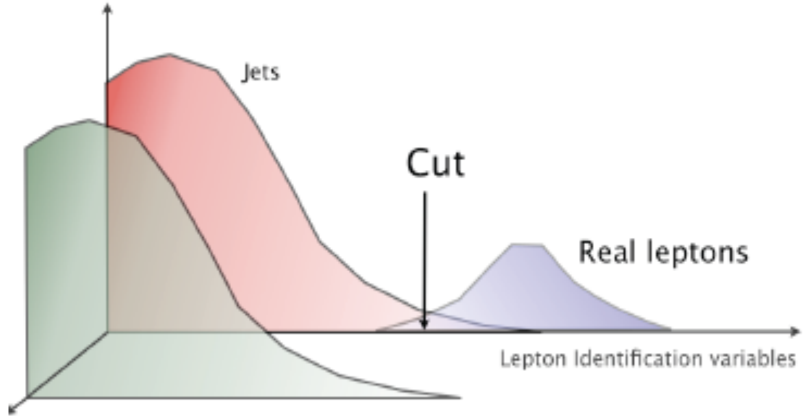
ABCD Method

- Define "loose" and "tight" lepton selection
- The ratio R^{QCD} between loose and tight background leptons is measured with an appropriate QCD control sample (B+D)

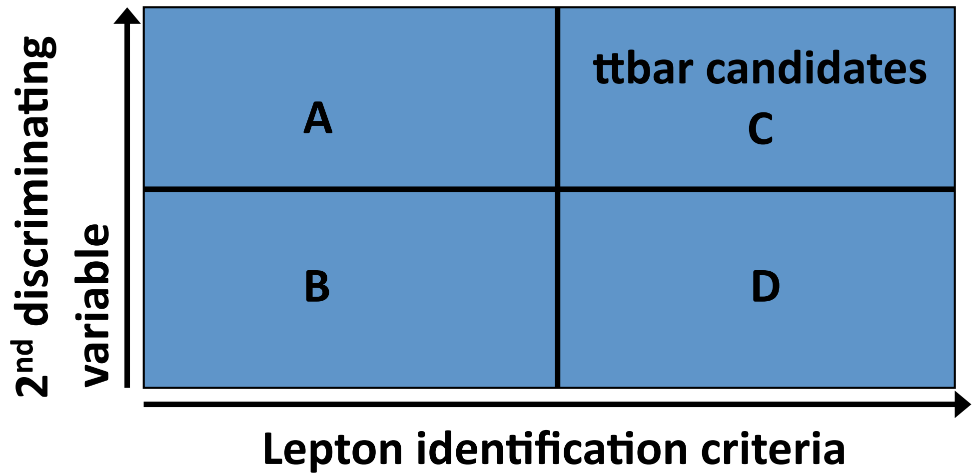
Example for B+D: invert etmiss cut

- **Neglecting signal leptons in loose sample A:**

$N_{QCD} = R^{QCD} * N_A = N_A * N_D / N_B$



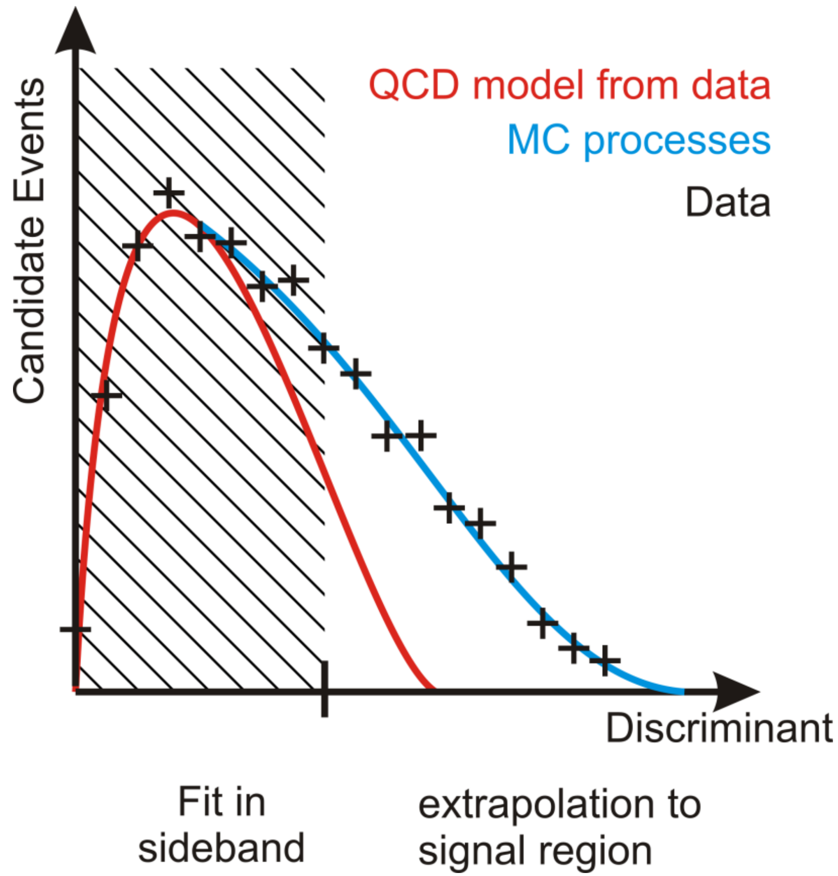
The signal contamination in A can be taken into account measuring the ratio R^W between loose and tight W leptons from the Z with tag and probe



Anti-electron method

The method used at the Tevatron

- **„Anti-electron“ sample:** orthogonal to the standard selection (inverted e ID cuts), triggered by jet trigger
- The sample will mostly consist of QCD
- Fit ttbar e+jets and mu+jets MET with the anti-electron MET shape (after subtracting other backgrounds in the low-MET region)
- If this leads to problems/bias for the mu +jets, will invent “anti-muons”

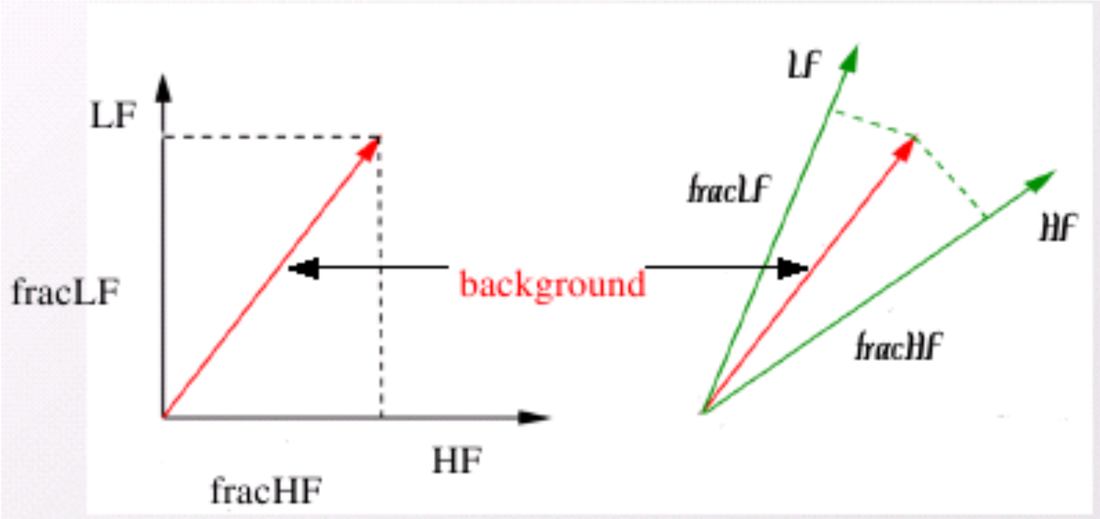


Data-Driven Templates

• Impossible to get a completely pure set of HF and LF templates

• If don't care where the fake come from, need to fit for the total:

- Obtain QCD templates that are HF enriched ("HF")
- Obtain QCD templates that are LF enriched ("LF")



Using for the electrons: TRT HT variable and isolation are uncorrelated

• Large HT for the electrons / Small for the HF and LF fakes

• Small Isolation for the electrons / Large for the HF and LF fakes

Data-Driven Estimates

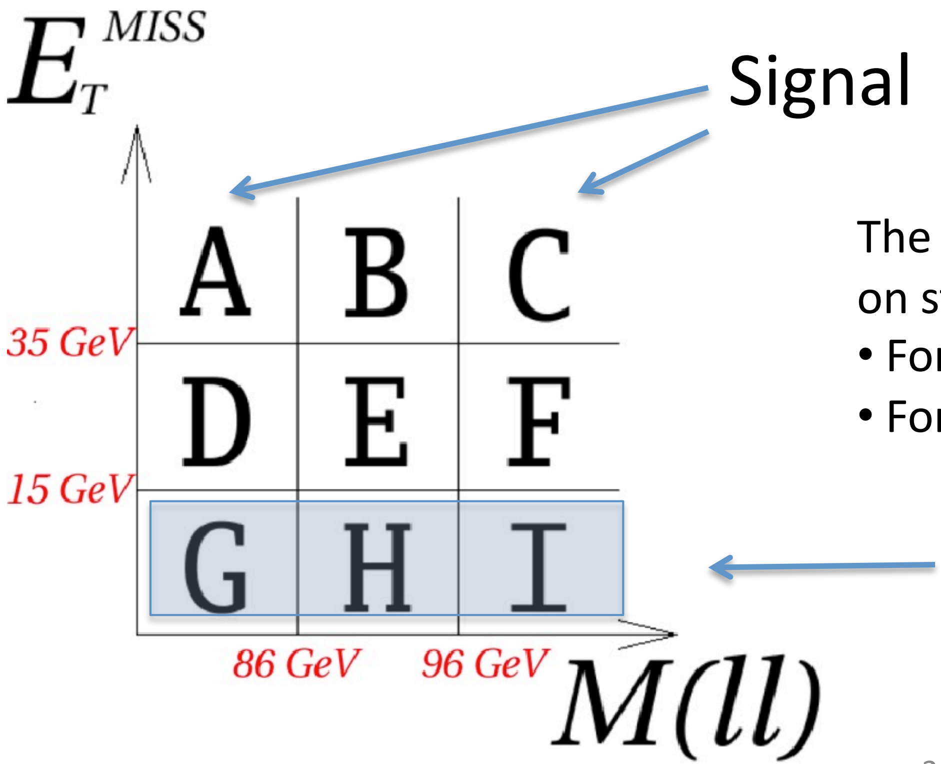
- Jets misID as leptons
  - W+jets bkg for the ttbar dileptons
  - QCD bkg for the ttbar l+jets
- Fake MET (mismeasured jets / muons)
  - DY bkg for the ttbar dileptons
- W+jets for l+jets

Z+jets

- G and I: normalization
- B/H: Extent of Z-region Drell-Yan MET tail
- A/G: Extent of low-mass Drell-Yan MET tail
- C/I: Extent of high-mass Drell-Yan MET tail

$$A_{Est} = G_{Data}\left(\frac{A_{MC}}{G_{MC}}\right)\left(\frac{B_{Data}}{H_{Data}}\right)\left(\frac{H_{MC}}{B_{MC}}\right)$$

$$C_{Est} = I_{Data}\left(\frac{C_{MC}}{I_{MC}}\right)\left(\frac{B_{Data}}{H_{Data}}\right)\left(\frac{H_{MC}}{B_{MC}}\right)$$



Background

The uncertainty of the method depends on statistics in Z+jets+MET sample

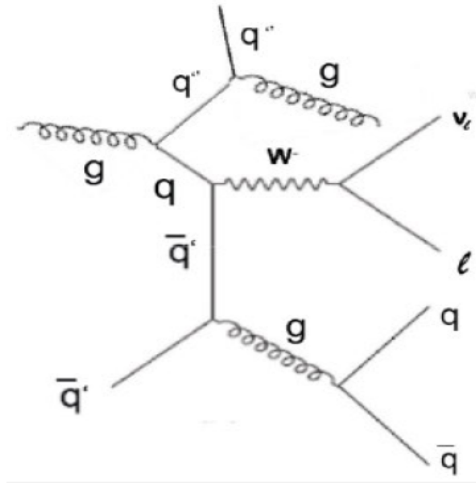
- For 200pb-1 at 10 TeV estimated to be ~15%
- For 10pb-1 at 7 TeV expected to be much higher

The method is not perfect, hence it is an active area of research in ATLAS – stay tuned

Data-Driven Estimates

- Jets misID as leptons
  - W+jets bkg for the ttbar dileptons
  - QCD bkg for the ttbar l+jets
- Fake MET (mismeasured jets / muons)
  - DY bkg for the ttbar dileptons
- W+jets for l+jets

W+jets



- Expected to be 3/4th of the total background
  (if no btagging is used)
- No NLO calculation of W+4j, MC estimates uncertain
- Few data-driven approaches well developed:

  W/Z ratio method

  $$(W^{SR}/W^{CR})_{data} = (Z^{SR}/Z^{CR})_{data} \cdot C_{MC}, \qquad C_{MC} = \frac{(W^{SR}/W^{CR})_{MC}}{(Z^{SR}/Z^{CR})_{MC}}$$

  Measure Z+4jets, rescale by
  W/Z ratio measured in lower Njets
  control sample

  $$r = \sigma(W^+)/\sigma(W^-) = \frac{u(x_1)\bar{d}(x_2) + c(x_1)\bar{s}(x_2)}{\bar{u}(x_1)d(x_2) + \bar{c}(x_1)s(x_2)}$$

  Charge asymmetry method

  $$N_{W^+} + N_{W^-} = \left(\frac{r_{MC}+1}{r_{MC}-1}\right)(N_{W^+} - N_{W^-})_{DATA}$$

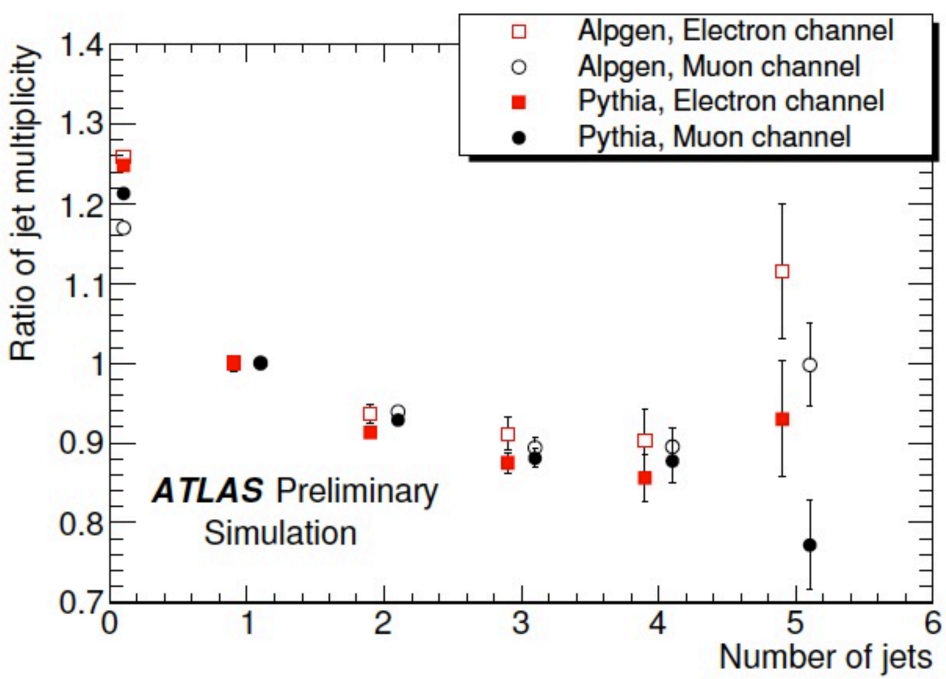
  Ttbar and most backgrounds are
  charge symmetric, while W+ and W- rates are different.

W+jets: W/Z ratio method

$$W^{SR} = C_{MC} * (Z^{SR} * (W^{CR}/Z^{CR}))$$ → measured from data

Estimated from MC: $C_{MC} = (W^{SR}/W^{CR})_{MC} * (Z^{CR}/Z^{SR})_{MC}$

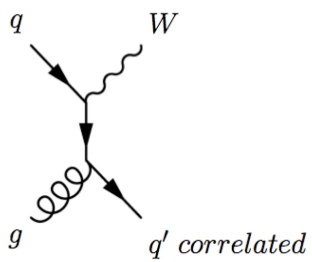
- W/Z ratio in CR (low jet multiplicity) can be measured very early
- Measurement of Z in signal region SR limited by statistics (few Zs in 10pb-1)
- The W/Z ratio is not constant with the number of jets, there is a small correction factor from MC.
- Many uncertainties cancel in the double ratio $(W/Z)^{CR}/(W/Z)^{SR}$

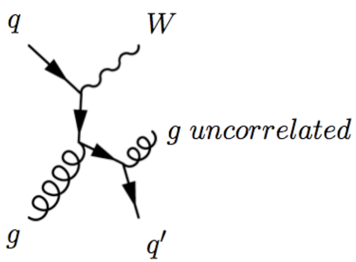


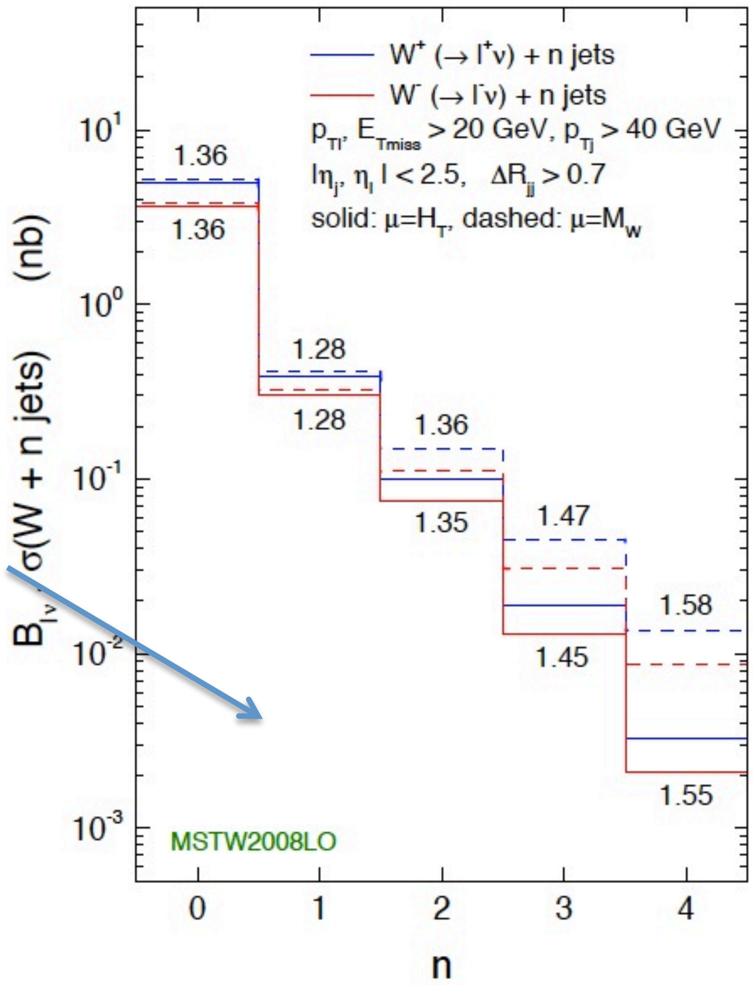
W+jets: Charge Asymmetry Method

ttbar is charge symmetric, W is not

- MC: $R = (NW^+ + NW^-) / (NW^+ - NW^-)$
- The W background is $R_{MC}(N+ - N-)$
- Need to subtract small t-channel single top contribution to asymmetry
- Main systematics is expected to come from PDF
- Statistically limited for early data
- **arXiv:1004.3404, [C. H. Kom](), [W. J. Stirling]()**
- Can be used as an additional handle for the dileptons







Summary

- We have developed a set of methods to estimate backgrounds from data, which have been thoroughly exercised on MC (started cross-checks with data). There are at least two methods for each major background:
  - Fake leptons for the ttbar l+jets and dileptons
  - W+Jets for ttbar l+jets
  - Z+Jets for the ttbar dileptons
- Non-Collision backgrounds are under control – have complimentary information from many subsystems
  - MBTS, LAr, TRT...
- Working on other methods – couldn't show all ideas
- We are in the midst of the physics program, moving step by step…

Roadmap

- Begin by understanding the basic objects: leptons, jets, HF, MET
- Continue by combining them to low-mass resonances
- Establish W signal, study vs N jets, estimate QCD background
- Measure ID and trigger lepton efficiencies, study backgrounds to top in W +jets and Z+jets samples
- Establish Z signal, study vs N jets, estimate W+jets / QCD backgrounds
- Validate MET in complex final states
- Look for the emergence of a top signal
- Study top final states (often a background for searches for physics beyond the Standard Model)
- Higgs and/or beyond the Standard Model

Stay tuned - we are moving forward!