HEP Analysis

John Morris
Queen Mary University of London
2012

HEP Analysis

Outline

- Monte Carlo (MC)
  - What is MC and why do we use it?
  - MC event generation
  - MC detector simulation
- What happens when MC does not agree with data?
  - Event re-weighting
  - 4-vector smearing
- Cross Section measurement
- Systematic uncertainties

- These notes will focus on the *ATLAS* experiment, but also apply to other experiments
- Borrowing slides, plots and ideas from E. Rizvi, T. Sjostrand and *ATLAS* public results

HEP Analysis

This is not an in-depth MC course

- The subject of MC is vast
- There are many different models, generators and techniques
    - Could spend hours, weeks, months on each
- These slides are just a brief outline
    - Focus on the generalities
    - Explain the steps involved
    - Not going to get bogged down in details

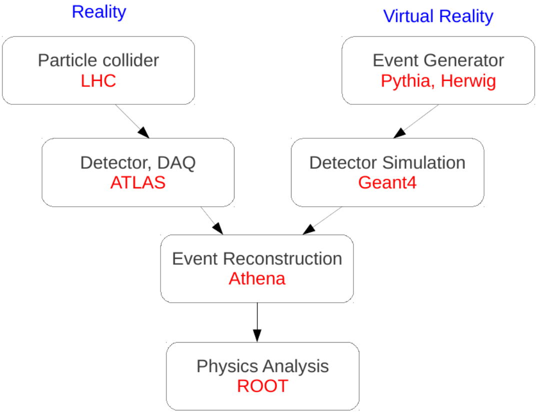
Aim of this course

- Aim is to provide a working knowledge of MC basics
    - Steps involved in generation
    - Introduction to various systematics
- Techniques for correcting MC to describe data
- Overview of a measurement
- Systematics

Monte Carlo (MC) - What is MC?

- MC is a simulation of collisions between particles
  - For these notes, *pp* collisions in *ATLAS* at the *LHC*
  - MC is used in all particle physics experiments
- MC generation is broadly split into 2 parts:
  1. Generation of the physics process
     - Matrix element calculation
     - ...(lots of steps)...
     - 4-vectors of final state hadrons
  2. Simulation of the detector
     - Trigger, tracking and calorimeter simulation
     - ...(lots of steps)...
     - Analysis objects like electrons and jets
- End result:
  - MC Samples in the same format as the actual data
  - With additional information from Step 1, the truth record

Monte Carlo (MC) - What is MC?



Reality

Virtual Reality

Particle collider
LHC

Event Generator
Pythia, Herwig

Detector, DAQ
ATLAS

Detector Simulation
Geant4

Event Reconstruction
Athena

Physics Analysis
ROOT

- MC simulates what happens at the *LHC* and *ATLAS*
- Many different programmes can be used at each stage

Monte Carlo (MC) - Why use MC?

Why use generators?

- Allows studies of complex multi-particle physics
- Allows studies of theoretical models
    - ⇒ What does a SUSY signal look like?

Can be used to

- Predict cross sections and topologies of various processes
    - ⇒ Feasibility study - Can we find the theoretical particle X?
- Simulate background processes to the signal of interest
    - ⇒ Can devise analysis strategies
- Study detector response
    - ⇒ Optimise trigger & detector selection cuts
- Study detector imperfections
    - ⇒ Can evaluate acceptance corrections
    - See next week for a discussion of acceptance
- Remove the effect of the apparatus from the measurement
    - ⇒ *Unfold* the data. Correcting the data for detector effects

MC Generation

Use random numbers for integration

- Code uses random number generator to integrate

$$\int_{x1}^{x2} f(x)\,.dx = (x_2 - x_1)\langle f(x)\rangle$$

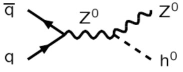
$$\langle f(x)\rangle = \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

- Cross section randomly sampled over phase space
- Want to generate events that simulate nature
  - Get average and fluctuations right
  - Need to make random choices, as in nature
  - An event with n particles involves $\mathcal{O}(10n)$ random choices
  - *LHC* events involve 1000's of random choices

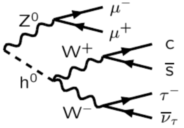
Monte Carlo Generation

Matrix elements (ME):

1) Hard subprocess:
$|\mathcal{M}|^2$, Breit-Wigners,
parton densities.

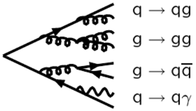


2) Resonance decays:
includes correlations.

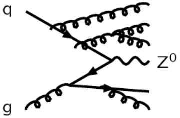


Parton Showers (PS):

3) Final-state parton showers.

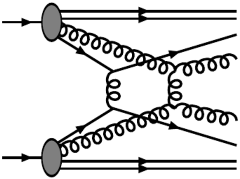


4) Initial-state parton showers.

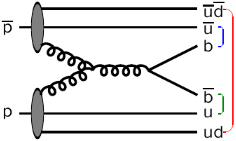


Monte Carlo Generation

5) Multiple parton–parton interactions.

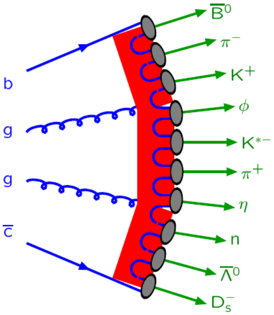


6) Beam remnants, with colour connections.



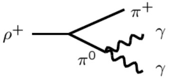
5) + 6) = Underlying Event

7) Hadronization



8) Ordinary decays:

hadronic, τ, charm, . . .



Monte Carlo Generation

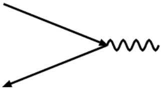
Identifying theoretical modelling uncertainties

- Let's look at some of these steps in more detail
- There are several ways to perform each step
- This choice leads to theoretical modelling uncertainties
  - These uncertainties enter into most *LHC* results
  - Want to understand where they come from
- Theoretical modelling uncertainties include:
  - Generator uncertainties
  - PDF uncertainties
  - Parton shower uncertainties
  - Hadronisation uncertainties

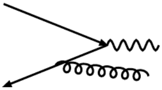
Matrix element calculation

Normally calculated at LO or NLO

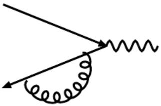
I. Lowest order,
$\mathcal{O}(\alpha_{\mathrm{em}})$:
$q\overline{q} \rightarrow Z^0$



II. First-order real,
$\mathcal{O}(\alpha_{\mathrm{em}}\alpha_s)$:
$q\overline{q} \rightarrow Z^0 g$ etc.



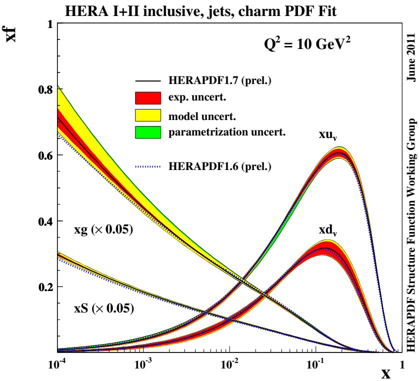
III. First-order virtual,
$\mathcal{O}(\alpha_{\mathrm{em}}\alpha_s)$:
$q\overline{q} \rightarrow Z^0$ with loops



- Higher order corrections are important:
  - Normalisation and shape of kinematic distributions
  - Multiplicity of objects like jets
- Higher order corrections are hard to calculate and CPU intensive
- Several programs that will do the calculation
  - Different calculation techniques
  - Different assumptions
  - Different results
  - ⇒ Theoretical modelling uncertainty

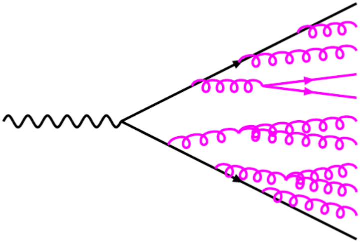
The Proton Parton Density Function (PDF)

- Proton is not a point-like particle, it's full of partons
- Need to calculate:
  - Probability of propagator interacting with quarks/gluon
  - Needed as a function of Q^2 and x



- Various groups provide PDFs
- Parametrised differently
- *LHC* uses:
  - CTEQ
  - MSTW
  - NNPDF
- Get a different result using different PDFs
- ⇒ Theoretical modelling uncertainty

Parton Showering



- Need to go from 2→2 scattering to 100's of particles
    - A particle can decay into more particles
    - A particle can emit another particle
    - All controlled by random numbers
- Parton shower evolution is a probabilistic process
    - Occurs with unit total probability

Parton Showering

2 Common approaches to parton showering

- Need to avoid divergences and infinities in calculations
  - See your QCD course for why these occur
  - Solution requires the final state partons to be ordered
- There are 2 common approaches to do this
- PYTHIA : $Q^2 = m^2$
  - The parton with the highest p_{T} is calculated first
- HERWIG : $Q^2 \approx E^2 (1 - \cos(\theta))$
  - The parton with the largest angle is calculated first

This represents a theoretical modelling uncertainty

- Both provide a good description of data but which is correct?
  - Neither is correct, but nature is unknown, we only have models
- All physics measurements need to take this into account
  - Expect to see a parton shower systematic for every result
  - Use both methods for calculation of physics result
  - Difference between results is a theoretical modelling systematic

Hadronisation

Going from partons to hadrons

- Partons are not observed directly in nature, only hadrons
- Hadronisation occurs at low energy scales
    - Perturbation theory is not valid
    - Cannot calculate this process from first principals
- Require models to simulate what happens
- 2 common approaches are used
    - PYTHIA : Lund string model
    - HERWIG : Cluster model

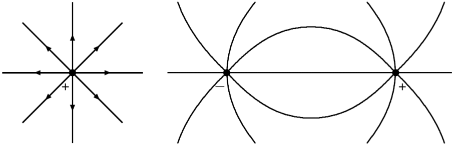
This is another theoretical modelling uncertainty

- Similar type of uncertainty as for parton showering
    - We don't know exactly how nature works
    - We have 2 reasonable models
    - Calculate physics result using each method
    - Difference is a theoretical modelling systematic

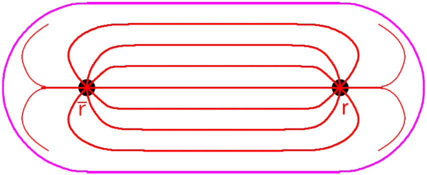
Hadronisation

The Lund string model

- In WED, field lines go all the way to infinity
- Photons do not interact with each other



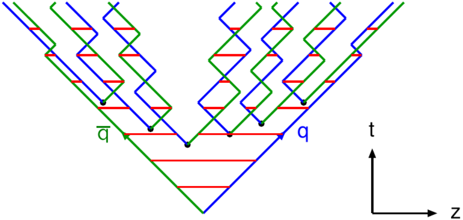
- In QCD, for large charge separation, field lines seem to be compressed into tube-like regions ⇒ string(s)
- Self-interaction among soft gluons in the vacuum



Hadronisation

The Lund string model

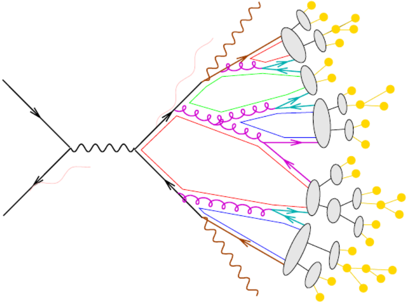
- The strings connecting the 2 partons breaks as they move apart
- Fragmentation starts in the middle and spreads out



- The breakup vertices become causally disconnected
- This is governed by many internal parameters
- Implemented by the PYTHIA MC program

Hadronisation

The Cluster model



- Pre-confinement colour flow is local
- Forced $g \to q\bar{q}$ branchings
- Colour singlet clusters are formed
- Clusters decay isotropically to hadrons
- Relatively few internal parameters
- Implemented by the HERWIG MC program

MC Generation

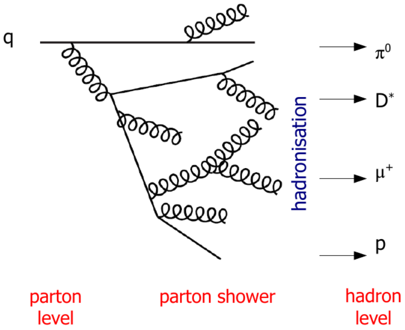
Summary

- There are many steps involved in generating MC
- There are competing models for each step
    - ⇒ Leads to theoretical modelling uncertainties
- General practice for a physics measurement
    - Decide on a "Nominal" choice for all options
    - Vary the generator, PDF, parton shower and hadronisation
    - Use the difference in your result as a systematic

Why only a few models?

- Don't really know what is happening, only have models
- Choice a few different models and take the difference
    - If model is very poor, doesn't this inflate the uncertainty?
- Why not use my "back-of-an-envelope" model?
    - Feel free to come up with a new model!
    - Need to describe the data accurately
    - Need to convince physics community

MC Generation

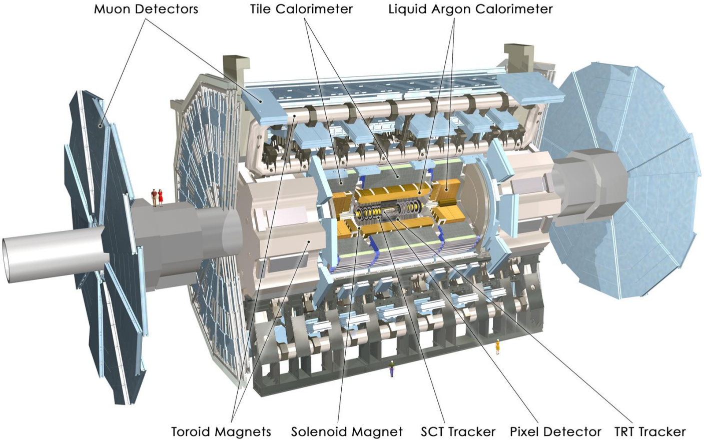


- MC Generator stops with set of "stable" final state particles
- Complete 4-vector info is known about every particle
- All parent-daughter relations are known and stored
- High energy parton state known as parton level
- Stable particle state known as hadron level

- This level of information is often called the truth record
- This is the pure event before it interacts with any apparatus

Detector Simulation

Now need to simulate the detector



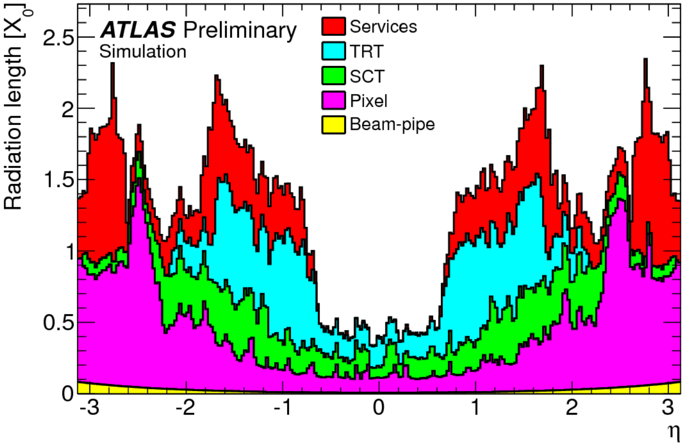
Detector Simulation

Detector simulation is complex and detailed

- Simulation of all major components and materials
  - Tracking, calorimetry, magnets, muon chambers
  - Support structure, cooling pipes, cables
  - Faulty components leading to missing readout
- GEANT4 program uses generator output 4-vectors
- Simulates interaction of particles within the detector volume
  - Particle ionisation in trackers
  - Energy deposition in calorimeters
  - Intermediate particle decays, radiation and scattering
- Full Simulation can take 10 minutes per event!
  - Possible to do a fast simulation - AtlFastII
  - Smears 4-vectors instead of doing calorimeter simulation
- Final output is raw data
  - Charges measured on each tracker wire
  - Electronic pulses in each calorimeter photomultiplier
  - Same format as raw data

Example of *ATLAS* material

Inner detector material in *ATLAS*



- Radiation length of various material in *ATLAS* ID Vs η
- This is what is in the current simulation
- Does it match reality? Probably not 100%

MC Warning

MC is not the truth

- What happened in the MC generator didn't happen at the *LHC*
  - The MC is our best guess
- The MC only simulates specific physics processes
- The cross section calculation may be:
  - Wrong, Incomplete, Inaccurate
  - Have a false kinematic dependence
  - The MC code may have unknown bugs in it
- Some processes require higher order corrections which have not been calculated
- Simulation cannot account for all detector effects

Reconstruction

Going from electronic pulses to analysis objects

- Data and MC pass through the same reconstruction algorithms
- Raw electronic pulses reconstructed into:
  - Tracks
  - Calorimeter deposits
- Which are then reconstructed into:
  - Jets, electron, muons, taus,
  - Photons, tracks, missing E_{T}

Real life issues need to be reflected in the MC

- Some parts of the detector become faulty over time
- e.g. - A section of the calorimeter readout dies and cannot be repaired until the detector is opened up in a shutdown
- Lets say that this affects x% of the data luminosity
- Need to generate MC with this problem in x% of the MC
  - Cannot know x until end of year
  - $\Rightarrow$ Need to reprocess the MC at the end of the year
- Some MC bugs do not become apparent for some time

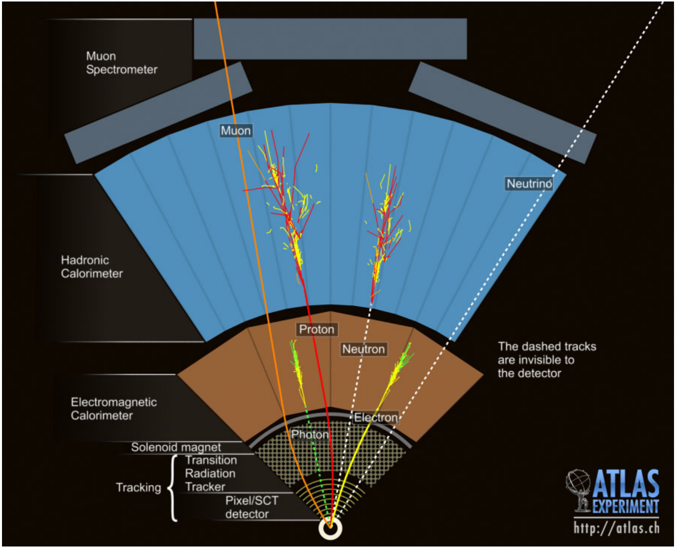
Event Properties

- Final state particles from the Standard Model:
    - Electrons, Muons, Taus, Neutrinos, Photons, (un)charged Hadrons
- Each interacts with detector in different ways
    - e, μ and τ are very similar in theory, however:
    - e leaves a track and doesn't penetrate further than EM Calo
    - μ leaves a track and passes through Calo into Muon chambers
    - τ looks very much like a jet
        - Decays within the inner detector
        - ⇒ Lots of tracks, EM and Hadronic Calo deposits

EM energy without a track	Photon
EM energy with a track	Electron
Hadronic energy without a track	Neutral Hadron (eg Neutron)
Hadronic energy with a track	Charged Hadron (eg Proton)
Hadronic energy with many tracks	Jet, Tau
ID and Muon chamber track	Muon
Missing transverse energy	Neutrino
Missing longitudinal energy	Beam remnants
Displaced secondary vertex	in-flight decay

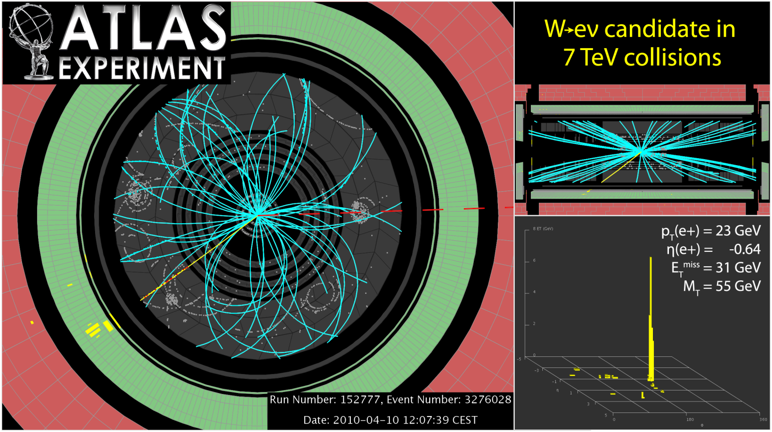
Event Properties

- Reconstructing an event requires recognising event properties



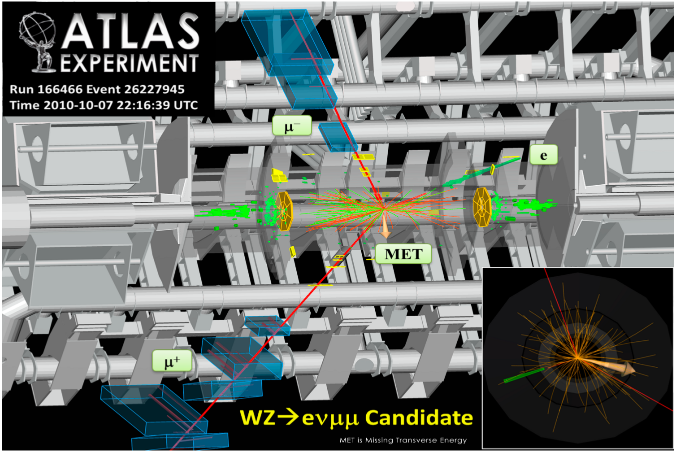
The dashed tracks are invisible to the detector

Event Display : $W \to e\nu$



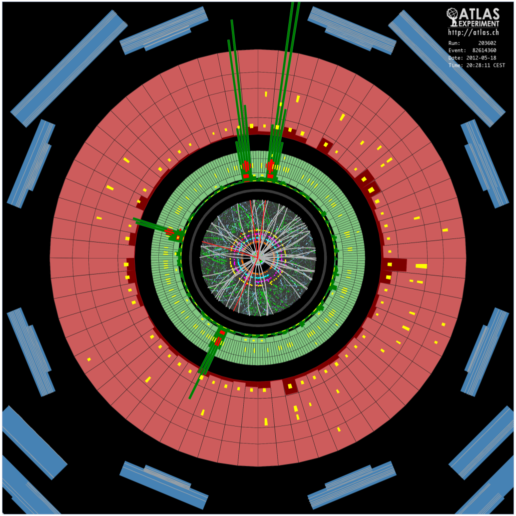
- Electron (yellow) leaves track and EM Calo deposit
- Missing E_{T} (red) identified with a neutrino
- $W \to e\nu$ Candidate event

Event Display : $WZ \rightarrow e\nu\mu\mu$



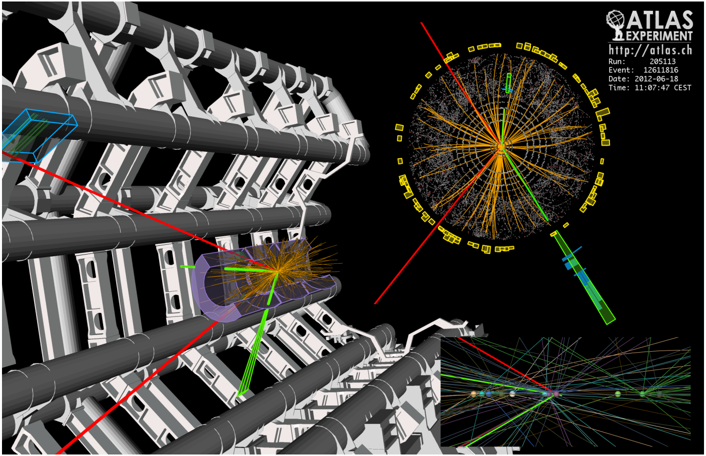
- Electron (green) and Missing E_{T} (orange) are reconstructed as $W \rightarrow e\nu$, like in the previous slide
- 2 additional Muons (red) reconstructed as $Z \rightarrow \mu\mu$

Event Display : $H \to 4e$



- 4 electrons reconstructed to candidate Higgs Boson

Event Display : $H \rightarrow ee\mu\mu$



- 2 electrons (green) and 2 muons (red) reconstructed to candidate Higgs Boson

When MC gets it wrong

Simulating the *LHC* and *ATLAS* is difficult

- As you can see, this is a very complex task
- Far more involved that my brief overview
- MC description of data often not good enough
    - ⇒ Re-generate MC ⇒ time consuming and difficult to get right
    - ⇒ Apply corrections to the MC ⇒ pragmatic and usable

How to deal with inaccurate descriptions

- This is broadly split into 3 different methodologies
- Re-weight the MC:
    - A data event contributes 1 entry to a histogram
    - A MC event may contribute 0.9 or 1.13 entries
    - Commonly used for trigger and reconstruction efficiencies
- Smear the MC:
    - Use random numbers (yet more!) to alter 4-vectors
    - Commonly used for jet, electron and muon p_T
- Tune the MC - Regenerate the MC with tuned input parameters

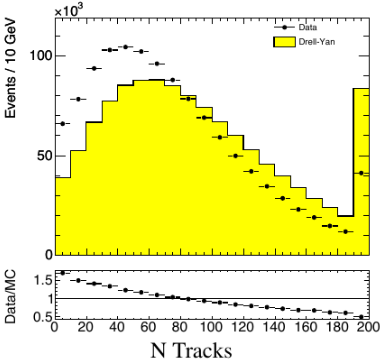
Re-Weighting basics

Histograms in ROOT

- You use histograms to plot data and MC
- An entry in a histogram contains a weight
  - histogram→ Fill(x, *weight*) ;
- For MC we will often want to give weight to an event
- The default in ROOT is to do simple errors $\sigma = \sqrt{N}$
- This will get the errors wrong in a weighted distribution
- Need to use the sum of squares of weights $\sigma = \sqrt{\sum w^2}$
- In your code, please ensure than when you book a histogram
  - TH1D* histo = new TH1D(“name”,“title”,nBins,MinX,MaxX);
  - histo→sumw2();

Re-Weighting MC

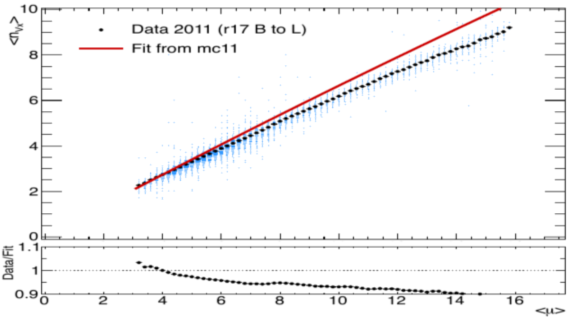
Number of tracks in *ATLAS* events



- The MC clearly does not describe the data
- What has gone so wrong?
- Is it parton showering, hadronisation or something else?
    - Need to understand the root cause of this disagreement

Pileup Re-Weighting

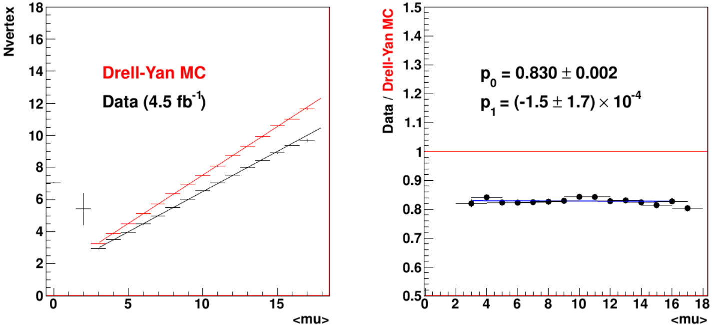
N vertices Vs average N interactions per bunch crossing



- Classic *ATLAS* example of MC not describing data accurately
- This shows that the MC gets the number of vertices wrong
  - Problem simulating proton bunches with 10^{11} protons
  - Understandably a very difficult task!
- Unfortunately this has big effects for many distributions

Pileup Re-Weighting

Need to determine re-weighting factors



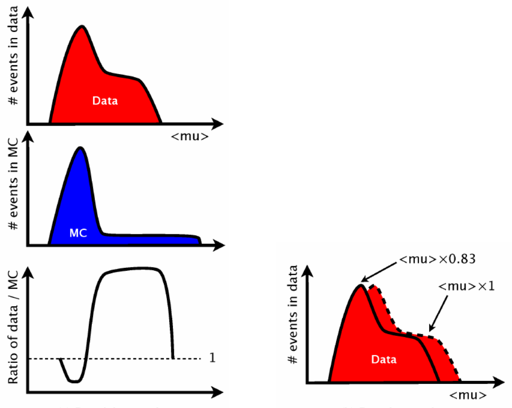
(a) Data-MC comparison

(b) Fit of the ratio of the distributions in (a)

- Divide Data by MC to determine correction
- In this case, fit the ratio and determine a weight
- Use this weight for each MC event
  - histogram → Fill(x, weight);

Pileup Re-Weighting

Cartoon illustrating re-weighting procedure

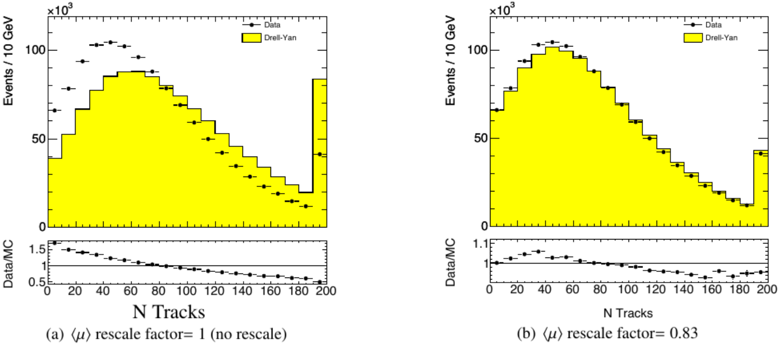


(a) Reweight procedure

(b) Rescale procedure

Pileup Re-Weighting

Effect of re-weighting on N Tracks



(a) $\langle\mu\rangle$ rescale factor= 1 (no rescale)

(b) $\langle\mu\rangle$ rescale factor= 0.83

- (a) Before re-weighting MC describes the data poorly
- (b) After re-weighting the MC description is much better
- We have identified the underlying problem - N vertices
  - We have re-weighted the N vertices distribution
  - Can see effect in the number of tracks distribution

Re-Weighting

We require many different weights

- Pileup is just 1 example where we need weights
- Electrons, muons, taus, photons all require weights for:
    - Trigger, reconstruction, identification
- Jets require weights for:
    - Reconstruction, resolution, jet vertex fraction
- All of these weights provide a correction to specific aspects of data/MC disagreement
- Combined all (via multiplication) to provide an overall weight

$$W_{\text{Event}} = W_e \; x \; W_\mu \; x \; W_\tau \; x \; W_{\text{jets}} \; x \; W_{\text{Pileup}} \; x \; W_{\text{other}}$$

$$\text{where } \; W_e = W_{\text{Trigger}} \; x \; W_{\text{Reco}} \; x \; W_{\text{ID}}$$

- The vast majority of weights are determined via dedicated studies using the Tag & probe methodology
    - ⇒Ideal service task for *ATLAS* authorship

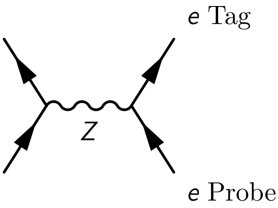
Tag & Probe Methods

A very common methodology for *ATLAS*

- Tag & Probe is a method to study analysis objects (jets, μ, etc)
- Study trigger, reconstruction efficiencies
- Data and MC often disagree in many places
    - Determine many different weights, often binned in p_T and η
- Use Standard Model candles like $Z \to ee$, $W \to \mu\nu$
    - What to use decay products from well known particles
    - Know that $Z \to ee$ has 2 electrons
    - Know the invariant mass of a Z very well
    - “Tag” one electron and study the other, “Probe” electron
- Use data-driven methods to determine the detector response
    - No need for MC in determining trigger, reco efficiencies
    - MC generally only used to determine weights
- In *ATLAS* weights are often called Scale Factors

Tag & Probe Methods

Tag & Probe : $Z \rightarrow ee$ example



- Tag electron:
    - Matched to Trigger
    - Tight reconstruction
    - Satisfies all conditions for being an excellent electron
- Probe electron:
    - Minimal selection

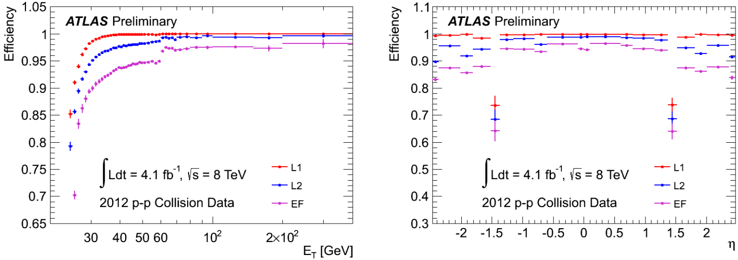
Use probe to determine efficiencies

$$\text{Some Efficiency} = \frac{\text{Probe Passes Selection}}{\text{All probes}}$$

$$\text{For Example} \;::\; \text{Trigger Efficiency} = \frac{\text{Probe Fires Trigger}}{\text{All probes}}$$

Tag & Probe Methods

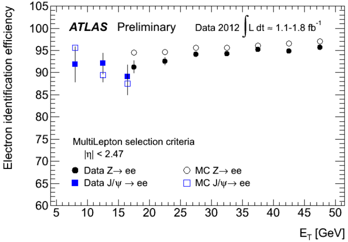
Tag & Probe : Electron trigger efficiency

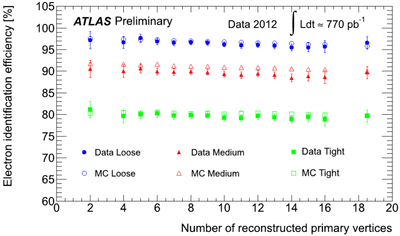


- $Z \to ee$ event selection applied
- Tag electron used to select events
    - No requirement on probe to pass trigger
- Probe electron asked "Did you fire the trigger?"
- Trigger efficiency determined for L1, L2 and EF
- Shown as a function of electron E_T and η

Tag & Probe Methods

Tag & Probe : Electron Identification efficiency

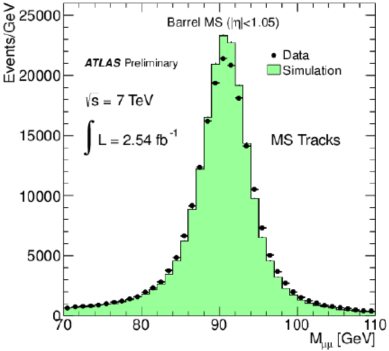




- $J/\Psi \rightarrow ee$ and $Z \rightarrow ee$ event selection applied
- Tag electron used to select events
  - No requirement on probe to pass identification cuts
- Probe electron asked "Did you pass ID selection?"
- Shown as a function of electron E_{T} and number of vertices
- Note how MC does not match data ⇒ weights are required

$Z \to \mu\mu$ Mass

What's wrong with this plot?



- Reconstructed Z mass using invariant mass of 2 muons
- Is this a scale or resolution effect?

Smearing the MC

Muons in *ATLAS*

- Muons are reconstructed using 2 detectors in *ATLAS*
    - Inner detector : Pixel + SCT + TRT
    - Muon Spectrometer
- Each detector has it's own momentum resolution

$$\frac{\sigma(p)}{p} = \frac{p_0^{\mathrm{MS}}}{p_{\mathrm{T}}} \oplus p_1^{\mathrm{MS}} \oplus p_2^{\mathrm{MS}} \, . \, p_{\mathrm{T}}$$

$$\frac{\sigma(p)}{p} = p_1^{\mathrm{ID}} \oplus p_2^{\mathrm{ID}} \, . \, p_{\mathrm{T}}$$

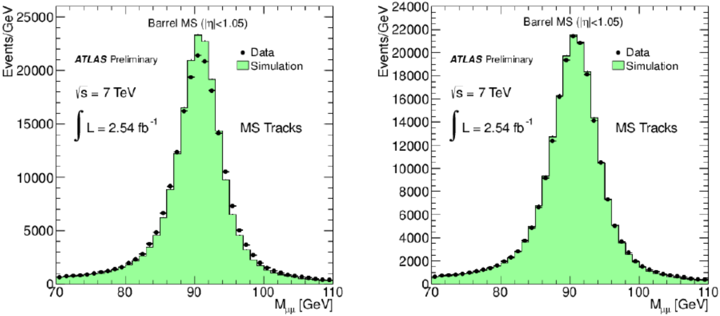
- These equations quantify the how well we can measure the momentum of any given muon
- A MC muon should be subject to the same uncertainties
    - ⇒ Smear the MC muon momentum using random numbers

Smearing the MC

Muons in *ATLAS*

$$p'_\mathrm{T} = p_\mathrm{T}\left(1 + g\Delta p_1^\mathrm{ID,MS} + g\Delta p_2^\mathrm{ID,MS} p_\mathrm{T}\right)$$

- p'_T : MC muon p_T after applying the corrections $\Delta p_i^\mathrm{ID,MS}$
- g : normally distributed random number, mean 0 and width 1



- Z mass before(left) and after(right) MC muon smearing

Smearing the MC

Smear depending on detector region

- Detector resolution varies with region
- Resolution in Barrel is different from that in End-Cap

η region	p_0^{MS} (TeV)	p_1^{MS} (%)	p_2^{MS} (TeV^{-1})
barrel	0.25 ± 0.01	3.27 ± 0.05	0.168 ± 0.016
transition	0	6.49 ± 0.26	0.336 ± 0.072
end-caps	0	3.79 ± 0.11	0.196 ± 0.069
CSC/No-TRT	0.15 ± 0.01	2.82 ± 0.58	0.469 ± 0.028
η region	p_0^{ID} (TeV)	p_1^{ID} (%)	p_2^{ID} (TeV^{-1})
barrel	n.a	1.55 ± 0.01	0.417 ± 0.011
transition	n.a	2.55 ± 0.01	0.801 ± 0.567
end-caps	n.a	3.32 ± 0.02	0.985 ± 0.019
CSC/No-TRT	n.a	4.86 ± 0.22	0.069 ± 0.003

- This is just for muons
- All sub-detectors have similar resolution tables

Summary so far

Monte Carlo (MC)

- MC generation:
    - Random numbers used extensively
    - Simulate matrix element
    - Simulate decays and hadronisation
- Detector simulation:
    - Output 4-vectors put through a complex detector description
    - MC and data are reconstructed using the same algorithms

MC often doesn't describe the data

- Scale Factors used to provide MC event weights
- Physics objects have their momentum smeared
- "Tag" and "Probe" methodology is widely used in *ATLAS*

Hopefully the MC now describes the data

- After simulation, extensive study and many corrections
- ⇒ Can now use the MC in physics measurements

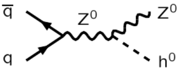
Day 2

- Review of Day 1
  - MC Generation
  - Detector simulation
  - Correcting MC with re-weighting and smearing
- Making a measurement
  - Cross section
  - Acceptance & Purity
  - Luminosity
  - Systematics
- Searches and exclusions

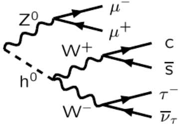
MC Generation

Matrix elements (ME):

1) Hard subprocess: $|\mathcal{M}|^2$, Breit-Wigners, parton densities.

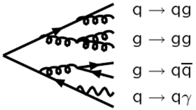


2) Resonance decays: includes correlations.

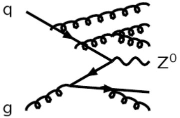


Parton Showers (PS):

3) Final-state parton showers.

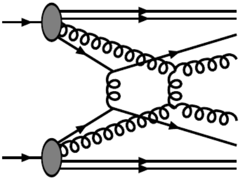


4) Initial-state parton showers.

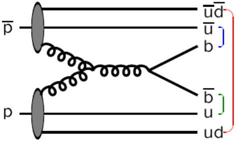


MC Generation

5) Multiple parton–parton interactions.

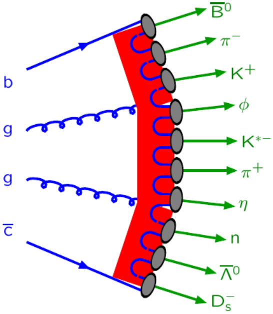


6) Beam remnants, with colour connections.



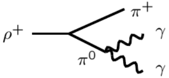
5) + 6) = Underlying Event

7) Hadronization



8) Ordinary decays:

hadronic, τ, charm, . . .



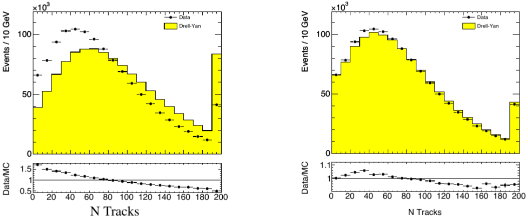
Detector Simulation

Detector simulation is complex and detailed

- Simulation of all major components and materials
  - Tracking, calorimetry, magnets, muon chambers
  - Support structure, cooling pipes, cables
  - Faulty components leading to missing readout
- GEANT4 program uses generator output 4-vectors
- Simulates interaction of particles within the detector volume
  - Particle ionisation in trackers
  - Energy deposition in calorimeters
  - Intermediate particle decays, radiation and scattering
- Final output is raw data
  - Charges measured on each tracker wire
  - Electronic pulses in each calorimeter photomultiplier
  - Same format as raw data
- This represents out best guess of the detector
  - Detector simulation is very complex
  - Difficult to model detector imperfections
  - Difficult to model bits of detector that break during running

Correcting MC

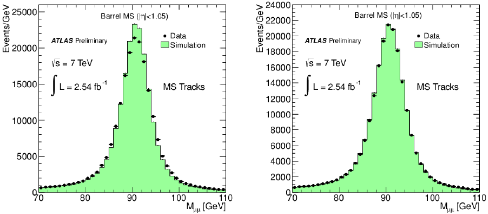
Use event weighting and 4-vector smearing



(a) $\langle\mu\rangle$ rescale factor= 1 (no rescale)

(b) $\langle\mu\rangle$ rescale factor= 0.83

- Re-weight MC events to achieve same efficiency in MC as in data
- Smear MC 4-vectors to achieve same resolution in MC as in data



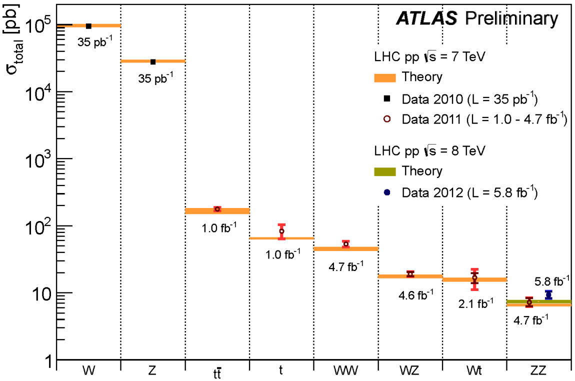
Cross section

- A cross section is defined by

$$\sigma = \frac{N}{\mathcal{L}}$$

- Where:
  - N is the number of events
  - $\mathcal{L}$ is the total integrated luminosity
- Represents a probability that an event will occur
- Given a fixed luminosity, you expect
  - X number of Z-boson events
  - Y number of $t\bar{t}$ events
  - Z number of Higgs events
- All have different cross sections

Cross section



- Shows cross section of many different processes at $\sqrt{s} = 7\,\text{TeV}$
- Represents a probability that a process will occur
  - For a fixed luminosity, you can expect X events

Purity & efficiency

What happens at the LHC

- 2 protons interact and a physics process occurs
- All processes that can occur, do occur
- The process you are interesting in is in there somewhere

Purity - reduced by background

- You will not only get the events you want
- You will get other events as well, ones you don't want
- Purity is percentage of your events that are signal

$$\text{Purity} = \frac{\text{Signal events in sample}}{\text{All events in sample}}$$

- Need MC to properly estimate purity

Purity & efficiency

Efficiency - reduced by detector

- Efficiency is percentage of all signal events that you reconstruct

$$\text{Efficiency} = \frac{\text{All events in sample}}{\text{All generated signal events}}$$

- Your process will have final states objects at all p_{T}
  - The p_{T} of pretty much all objects obeys an exponential
  - You will not be able to trigger objetcs at low p_{T}
  - You will not be able to reconstruct objects at low p_{T}
  - $\Rightarrow$ Reduction in efficiency
- Your process will have final states objects at all η
  - Your detector will only cover a fixed range in η
  - $\Rightarrow$ Reduction in efficiency

Cross section

Not so simple

- Naively a cross section is simply counting events
- It's not that easy!
- Assumes that we only measure the process of interest
    - All processes occur ⇒ will will get some background
- Assumes that we have a perfect description of our detector
    - No holes / cracks
    - Perfect efficiency for measuring particles
    - Perfect resolution
    - No background events
    - No dependence on MC models
- We already know that we need to correct for these issues
- ⇒ Cross section calculation needs modification

Cross section

Modified Cross section formula

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR}$$

- Where:
  - N_{obs} is the number of observed events
  - N_{bkg} is the number of expected background events
  - $\mathcal{L}$ is the total integrated luminosity
  - ϵ is the acceptance efficiency
  - BR is the branching ratio

Luminosity

Need to know the *LHC* luminosity

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L} \,.\, \epsilon \,.\, BR}$$

$$\mathcal{L} = \frac{\mu n_b f_r}{\sigma_{inel}} = \frac{\mu^{meas} n_b f_r}{\epsilon \sigma_{inel}} = \frac{\mu^{meas} n_b f_r}{\sigma_{vis}}$$

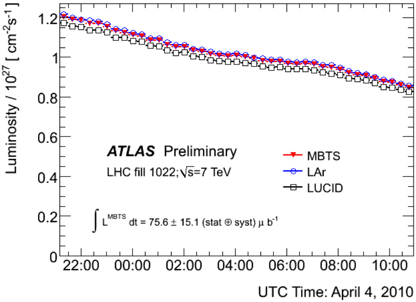
- μ is the average number of interactions per bunch crossing
- n_b is the number of bunches colliding at the interaction point
- f_r is the machine revolution frequency
- σ_{inel} is the total inelastic cross section
- $\mu^{meas} = \epsilon\mu$ is the measured μ
- ϵ is the detector luminosity reconstruction efficiency
- σ_{vis} is the visible cross section

Luminosity

Luminosity determination

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR}$$

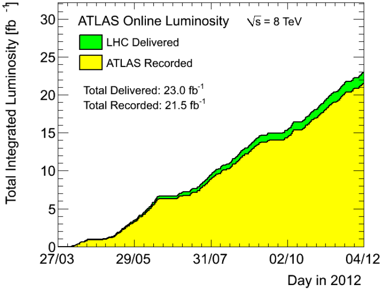
- Several methods used to determine the luminosity
- LUCID used for primary result, others are cross-checks
- Instantaneous luminosity decreases over time



Cross section

ATLAS Luminosity

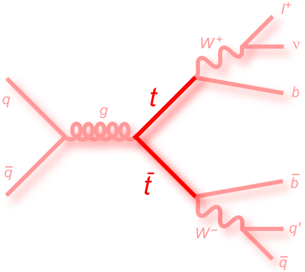
$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L} \,.\, \epsilon \,.\, BR}$$



- *ATLAS* only records a percentage of what the *LHC* delivers
  - Sub-detectors not all working
  - Transition time from “Stable Beams” to recording
- Absolute luminosity known to $\approx 4\%$

Cross section Example

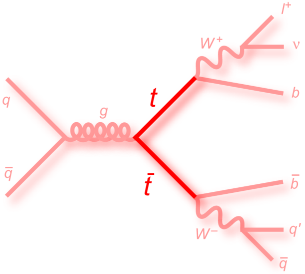
Let's work through a $t\bar{t}$ cross section measurement



- Top quark decays $t \rightarrow bW$
- W decays either $W \rightarrow q\bar{q}$ or $W \rightarrow l\nu$
- 3 classifications of event:
  - $W \rightarrow q\bar{q}$ and $W \rightarrow q\bar{q}$ - Fully hadronic
  - $W \rightarrow q\bar{q}$ and $W \rightarrow l\nu$ - Semileptonic
  - $W \rightarrow l\nu$ and $W \rightarrow l\nu$ - Dilepton

Cross section Example

Focus on the semileptonic classification



- Final state:
  - 1 lepton - e or μ. Don't use τ as it looks like a jet
  - Missing E_{T} - from the neutrino
  - 4 jets
    - 2 of these jets are from b-quarks
    - Require that we have 1 jet tagged with a b-tagging algorithm

Event selection

Need to apply an event selection

- *LHC* experiments record millions of events
  - We are looking for a specific process
  - We want to select only the events of our process

- Use only good data
  - Only use data where all components have given a green light
  - Red light if a sub-detector isn't working properly
- Trigger : Require that event passes a electron or muon trigger
- Place p_{T} and η requirements on the lepton and jets
  - Only look at areas of the detector we can measure well
- Place quality cuts on the lepton and jets
  - Reconstruction algorithms usually class object quality
  - Only use "good" objects - reject "bad" objects
- Require missing E_{T} for the neutrino
- Require at least 1 jet is *b*-tagged

Observed events

Number of observed events

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR}$$

- N_{obs} is relatively simple
- Apply event selection to the data
- Count how many events we observe

Background events

Number of background events

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR}$$

- N_{bkg} is not so simple
- How many background events have passed our selection?
- For the $t\bar{t}$ process, backgrounds are:
  - Single Top - Contains a t quark, W-boson and b-jet
  - W+Jets - Contains a W-boson and sometimes a b-jet
  - Z+Jets - Sometimes a lepton is reconstructed as a jet
  - DiBoson events WW, WZ, ZZ - can pass the selection
  - QCD Multijets
- b-tagging algorithms have "fake-rates" and often tag light jets
- Backgrounds are estimated using MC and data-driven methods

Background events

Number of background events

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L} \; . \; \epsilon \; . \; BR}$$

MC estimation of the background

- MC is used to estimate the number of background events
- Have to trust MC cross section calculation
- Have to trust MC generation process and detector simulation
- Simply count number of MC events expected:
  - Normalised MC events to data Luminosity
  - Put MC samples through event selection
- Done for single top, Z+Jets and DiBoson processes

Background events

Number of background events

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR}$$

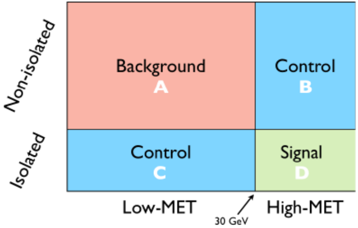
Data-driven estimation of the background

- There are processes where we don't trust the MC
- W+Jets process:
  - Very difficult to calculate
  - Large theoretical uncertainties in normalisation
  - Large theoretical uncertainties in heavy flavour composition
    - What proportion of jets are light, c, b ?
- QCD Multijet processes:
  - Standard Model processes involving light quarks and gluons
  - Dominates all events at the *LHC*
  - We do not trust the MC

Background events

Data-driven background estimation example

- Split phase space into 4 regions
- Use 2 variables
- Muon isolation variable
- Missing E_{T} of event



- Assume that QCD background is the same in all 4 regions
    - Big assumption ⇒ large uncertainties
- Count events in regions A,B & C
- Extrapolate to number of events expected in signal region D
- This is just 1 method used
    - Typically use several methods, all with different assumptions
    - Use difference of methods as a systematic

Acceptance Efficiency

Acceptance Efficiency

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR}$$

- If X number of signal events were generated
- How many make it into our final selection?
- Events fail to make the selection cuts because:
  - Events fails to fire a trigger
  - Inefficient reconstruction algorithms
  - Final state objects:
    - Have too low p_{T} to pass cuts
    - Are outside of detector volume - η
    - Go into cracks or broken bits of detector
- The acceptance efficiency is calculated on the signal $t\bar{t}$ MC
  - We rely on the MC for our measurement
  - We must have high level of trust in MC

Branching Ratio

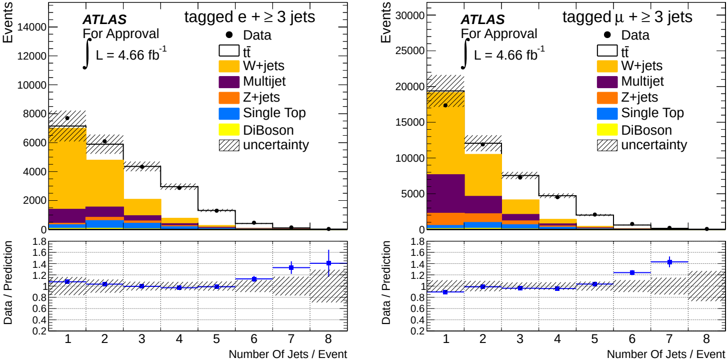
Branching Ratio

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR}$$

- Branching ratio is process dependant
- Determined by theoretical calculation
- In the case of $t\bar{t}$, there are 3 decay types:
  - $W \rightarrow q\bar{q}$ and $W \rightarrow q\bar{q}$ - Fully hadronic
  - $W \rightarrow q\bar{q}$ and $W \rightarrow l\nu$ - Semileptonic
  - $W \rightarrow l\nu$ and $W \rightarrow l\nu$ - Dilepton
- For the semileptonic case $BR = 0.543$

Control plots

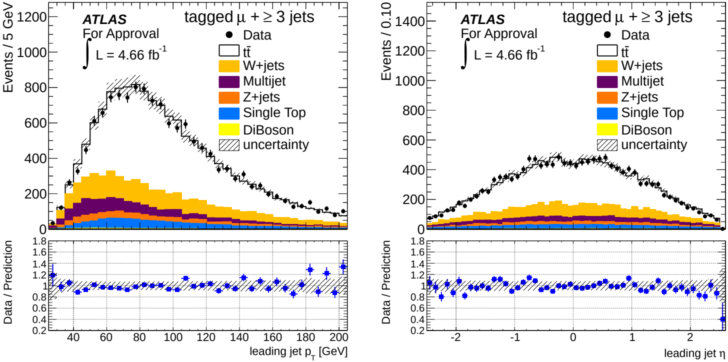
Does the MC describe the data?



- Shows the number of jets after all event selection
- Simulation histograms are stacked on top of each other
- Uncertainty is shown with hatched area
- MC does a good job of describing data

Control plots

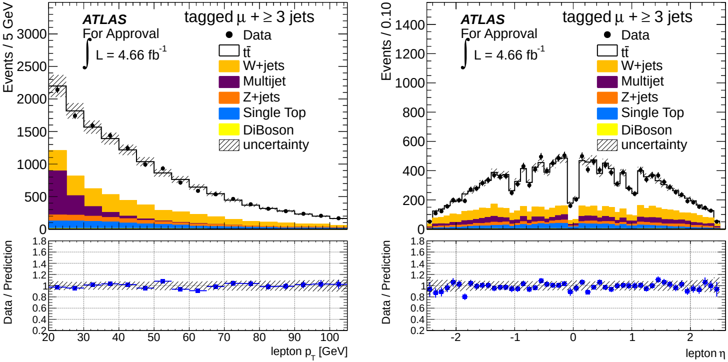
Does the MC describe the data?



- Shows the leading (highest p_{T}) jet p_{T} and η
- Data / Prediction plot shown underneath
- Ideally this would be at unity, not always the case
- Keep in mind the uncertainty band and statistical uncertainties

Control plots

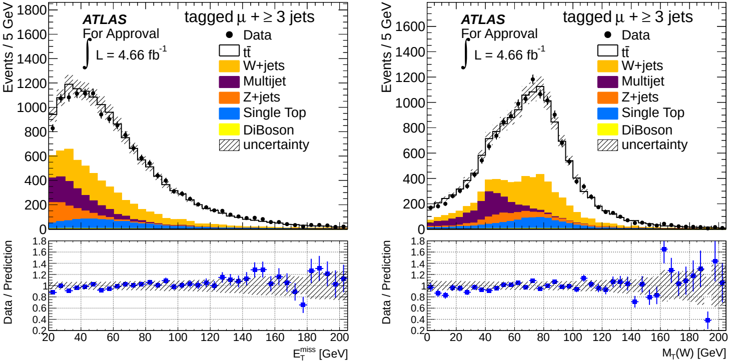
Does the MC describe the data?



- Shows the muon from the W-boson p_{T} and η
- Data / Prediction plot shown underneath
- Statistical fluctuations mean that a few points will be away
- from unity and outside the uncertainty band

Control plots

Does the MC describe the data?



- Shows the Missing E_T and W-boson transverse mass
- MC needs to describe data across a broad range of distributions
- A good analysis should have many control plots
- A good rule is to plot every variable you cut on

Cross section

Pulling it all together

- Count events for each process and make a yield table

	Muon Channel
Data	14940
$t\bar{t}$ (MC)	9084
Single top (MC)	980
Z+Jets (MC)	780
Diboson (MC)	59
Multijet (DD)	1310
W+Jets (DD)	2880
Backgrounds	6009
$t\bar{t}$ + Backgrounds	15093
$t\bar{t}$ acceptance ϵ	0.0215

- $N_{obs} = 14940, \; N_{bkg} = 6009$
- $\epsilon = 0.0215, \; BR = 0.543, \; \mathcal{L} = 4.66 fb^{-1}$

$$\sigma_{t\bar{t}} = 164.2 pb$$

We have a result!

Can we publish yet?

$$\sigma_{t\bar{t}} = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR} = 164.2pb$$

- This result is currently unusable - not even close to publication
- What is the uncertainty?
  - $\sigma_{t\bar{t}} = 164.2 \pm 1.0pb$ - fantastic result! World leading
  - $\sigma_{t\bar{t}} = 164.2 \pm 150.0pb$ - Never mind, will struggle to publish

Now the real work begins

- The uncertainty on your measurement is critical
  - More important than central value
- Crudely speaking
  - Small uncertainty = Good measurement
  - Large uncertainty = Bad measurement

Evaluating uncertainties

Determining your measurements' uncertainty

- There are, broadly speaking, 4 main categories of uncertainty
- Let's go through them individually
- Statistical
    - Determined entirely by how many signal events you have
- Luminosity
    - We only know it to $\approx 4\%$
- Experimental
    - Background estimation
    - Uncertainties from re-weighting MC
    - Uncertainties from smearing MC
- Theoretical modelling uncertainties
    - Generator, PDF, parton shower, ISR/FSR
    - Introduced in 1^{st} lecture

Offset Method

Adding uncertainties in quadrature

- Imagine we have 3 uncertainties, α_A, α_B and α_C
- Add uncertainties in quadrature
- Systematics tend to be asymmetrical
  - $\Rightarrow$ Calculate $\uparrow$ and $\downarrow$ contributions separately

$$\alpha_{tot} = \sqrt{\alpha_A^2 + \alpha_B^2 + \alpha_C^2} = \sqrt{\sum_i^N \alpha_i^2}$$

Offset method of combining uncertainties

- We are going to want to combine many different uncertainties
- We normally use the offset method to do this
- Change something and recalculate the final result
- Difference in results is the uncertainty

$$\alpha_A = \sigma_{t\bar{t}}^{Syst\ A} - \sigma_{t\bar{t}}^{Nominal}$$

Statistical uncertainty

Statistical uncertainty

- Determined by number of data events you have
- $\alpha = \sqrt{N_{events}}$
- In our $t\bar{t}$ cross section
  - $N_{obs} = 14940 \Rightarrow \alpha_{N_{obs}} = \sqrt{14940} = 122.2$
  - Recalculate cross section using $N_{obs} \pm \alpha_{N_{obs}}$
  - $\sigma_{t\bar{t}}^{Stat\uparrow} = 166.4pb \qquad \sigma_{t\bar{t}}^{Stat\downarrow} = 162.0pb$
- The statistical uncertainty is
  - $\alpha_{Stat\uparrow} = 166.4 - 164.2 = 2.2pb$
  - $\alpha_{Stat\downarrow} = 162.0 - 164.2 = -2.2pb$
- Our updated measurement is now:

$$\sigma_{t\bar{t}} = 164.2 \pm 2.2\,(\mathrm{Stat.})$$

Luminosity uncertainty

Luminosity uncertainty

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR}$$

- The Luminosity at *ATLAS* is only known to ≈ 4%
- Has a direct impact on the cross section
- Vary the luminosity and recalculate the cross section
  - $\sigma_{t\bar{t}}^{Lumi\uparrow} = 152.7pb \quad \sigma_{t\bar{t}}^{Lumi\downarrow} = 175.7pb$
- Our updated measurement is now:

$$\sigma_{t\bar{t}} = 164.2 \pm 2.2\,(\mathrm{Stat.}) \pm 7.0\,(\mathrm{Lumi.})$$

- As $\sqrt{2.2^2 + 7.0^2} = 7.3$, adding in quadrature gives

$$\sigma_{t\bar{t}} = 164.2 \pm 7.3\,(\mathrm{Stat.} \oplus \mathrm{Lumi.})$$

Experiment uncertainties

Mis-modelling of the detector

- Background estimations can have significant uncertainties
- We know that data and MC do not agree out-of-the-box
- Event re-weighting:
    - Correct MC efficiencies to match data efficiencies
    - Applied to most (if not all) reconstructed objects
    - Applied to Pile-up and missing E_{T}
- 4-vector smearing:
    - Correct MC resolutions to match data resolutions
    - Applied to most (if not all) reconstructed objects
- Each and every re-weighting/smearing has an uncertainty
- Each uncertainty needs to be accounted for in final result
- Varies between measurements according to event selection
- Clearly, different experiments will apply different re-weightings

Background uncertainties

Background estimation uncertainties

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L}\ .\ \epsilon\ .\ BR}$$

- Data-driven methods tend to have large uncertainties
    - Don't trust the MC for these processes
    - Use several different data-driven techniques
    - Uncertainty often taken as difference of methods
- In the case of the $t\bar{t}$ cross section
    - QCD Multijet estimation $= 1310 \pm 130$ events
    - W+Jets estimation $= 2880 \pm 350$ events
- Recalculate cross section for ↑ and ↓ cases
    - That's 4 more calculations from this page
- Calculate difference from the nominal result
- Add differences together in quadrature

Experiment uncertainties

- All further uncertainties effect ϵ

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{L} \,.\, \epsilon \,.\, BR}$$

Re-weighting and smearing

- All re-weighting routines should also provide uncertainties
  - If not, bug the author
  - If you are the author - make sure you provide uncertainties
- Simply a case of re-running over your MC N times
  - Where N is the total number of possible variations ↑ and ↓
  - Use different weights each time to reflect different uncertainties
  - Recalculate cross section
  - Calculate difference from the nominal result
  - Add differences together in quadrature

Theoretical uncertainties

Theoretical modelling uncertainties

- Theoretical understanding (or lack of it) of what is happening
- See 1^{st} lecture for introduction to
    - Calculation of matrix element
    - Parton Density Function (PDF) of the proton
    - Parton showing and hadronisation
    - Initial and final state radiation (ISR/FSR)
- Typically evaluated by using different signal MC samples
    - $\alpha_{Theo.}$ is often not simply the difference from the nominal
    - For some prescriptions you take half the difference
    - Let's look at the PDF for a complicated case
- Should be common across experiments
    - Reality is that *ATLAS* and *CMS* do things differently
    - Working group set up
    - Expect convergance in 20XY

PDF uncertainties

The PDF4LHC prescription

- Test 3 different PDF sets
    - CTEQ (Nominal), MSTW and NNPDF
- Each PDF set gives an event weight, based on EventNumber
- The first weight is the PDF nominal value, followed by n different tests varying a PDF eigenvector $\pm 1\sigma$
- Uncertainties are calculated differently for each PDF set
- Total uncertainty is $\frac{1}{2}$ the envelope
- Nominal cross section does not enter systematic

	PDF Set		
	CTEQ	MSTW	NNPDF
N Weights	53	41	101
N PDF eigenvectors varied $\pm 1\sigma$	26	20	50
Uncertainty method	Symmetric Hessian	Asymmetric Hessian	Standard Deviation

PDF uncertainties

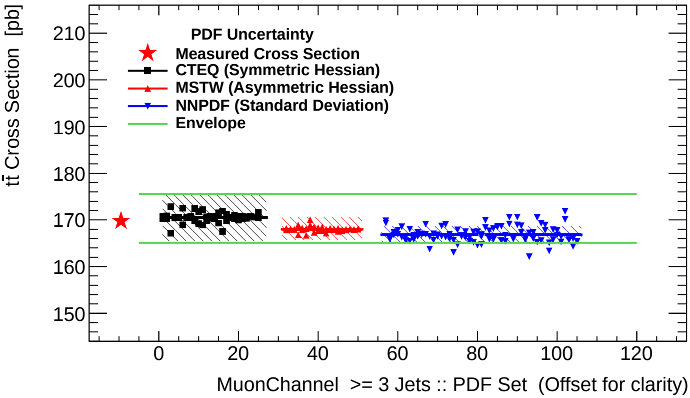
The PDF4LHC prescription

- If:
  - X_0 = Nominal PDF cross section (first weight, not our $\sigma_{t\bar{t}}$)
  - $X_i^{\pm}$ = PDF eigenvector test i, varying the eigenvector by $\pm 1\sigma$
  - $\bar{X}$ = Mean of all PDF eigenvectors (Used for NNPDF)
  - N = All PDF eigenvectors. This does not include the first value
- Then :
  - CT10 $\alpha = \frac{1}{2}\sqrt{\sum_{i=1}^{N/2}\left(X_i^{+} - X_i^{-}\right)^2}$
  - MSTW $\alpha^{\uparrow} = \sqrt{\sum_{i=1}^{N}\left(X_i - X_0\right)^2} \quad : \text{if } (X_i - X_0) > 0$
  - MSTW $\alpha^{\downarrow} = \sqrt{\sum_{i=1}^{N}\left(X_i - X_0\right)^2} \quad : \text{if } (X_i - X_0) < 0$
  - NNPDF $\alpha = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X} - X_i\right)^2}$

PDF uncertainties

- CTEQ $= 170.5 \pm 5.0$ pb
- MSTW $= 168.0^{+2.6}_{-2.2}$ pb
- NNPDF $= 166.8 \pm 1.7$ pb

- Envelope Up $= 175.5$ pb
- Envelope Down $= 165.1$ pb
- $\frac{1}{2}$Envelope $= 5.2$ pb (3.07%)



Combining channels

Combining electron and muon channels

- Do not want a electron results and a muon result
- Want to combine the channels together
- Relatively simple to do:
    - Add event yields from electron and muon channels
- Must do this for each systematic
    - Takes into account correlations

Full systematic table

	Relative cross section uncertainty [%]		
Source	e+jets	μ+jets	Combined
Statistical Uncertainty	±1.5	±1.3	±1.0
Object selection			
Lepton energy resolution	+0.4 /-0.3	+0.2 /-0.1	+0.2 /-0.1
Lepton reco, ID, trigger	+2.4 /-2.5	+1.5 /-1.5	+1.7 /-1.8
Jet energy scale	+3.8 /-4.3	+3.2 /-3.6	+3.5 /-3.8
Jet energy resolution	±0.2	±0.5	±0.2
Jet reconstruction efficiency	±0.06	±0.06	±0.06
Jet vertex fraction	+1.2 /-1.4	+1.2 /-1.4	+1.2 /-1.4
$E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainty	±0.06	±0.08	±0.07
SMT muon reco, ID	±1.3	± 1.3	±1.3
SMT muon χ^2_{match} efficiency	±0.6	±0.6	±0.6
Background estimates			
Multijet normalisation	± 5.2	± 3.9	± 4.4
W+jet normalisation	± 5.2	± 5.7	± 5.5
Other bkg normalisation	± 0.2	± 0.2	± 0.1
Other bkg systematics	+1.6 /-1.5	+2.5 /-2.0	+2.2 /-1.8
Signal simulation			
$b \rightarrow \mu X$ Branching ratio	+2.9 /-3.0	+2.9 /-3.1	+2.9 /-3.1
ISR/FSR	± 2.4	± 0.9	± 1.5
PDF	± 3.2	± 3.0	± 3.1
NLO generator	± 3.2	± 3.2	± 3.2
Parton shower	± 2.2	± 2.2	± 2.2
Total systematics	±11.2	±10.2	±10.5
Integrated luminosity	± 3.8	± 3.8	± 3.8

Final result

The final result

- Combining the electron and muon channels
- The final cross section for $t\bar{t}$ at $\sqrt{s} = 7\,\text{TeV}$ is

$$\sigma_{t\bar{t}} = 165 \pm 2(\mathrm{Stat.}) \pm 17(\mathrm{Syst.}) \pm 7(\mathrm{Lumi.})\ \mathrm{pb}$$

- The theoretical calculation is

$$\sigma_{t\bar{t}}^{Theo.} = 167^{+17}_{-18}$$

- Majority of measurement is evaluating systematics
- Uncertainty tells you the precision of a measurement

Searches & Exclusions

Searches & Exclusions

- Cross sections are a basic measurement
  - They are still very involved and complex
  - Must be measured
  - We have to understand the Standard Model before we search for physics beyond the Standard Model
- The *LHC* is a discovery machine
- Designed to search for the Higgs and BSM physics
  - Exciting PhD topics with plenty of work that needs doing
- We search for new physics and there are 2 options:
  1. Discovery!!
  2. No discovery - exclude a theory within certain limits
- Option 2 is by far the most common

Searches & Exclusions

Methods for searching for new physics

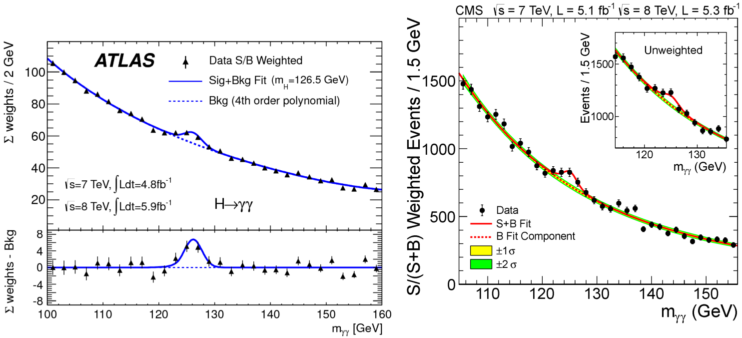
- Bump hunting
    - Plot invariant mass distributions
    - Understand the backgrounds
    - Look for a localised excess of events - a bump
    - Statistical analysis
    - $\Rightarrow$ New particle $\Rightarrow$ Higgs candidate
- Gradient searches
    - Plot mass or p_{T} distributions
    - Understand the backgrounds
    - Study gradient of distribution
    - Deviation from SM prediction could indicate
        - Quark compositness - quark sub-structure
        - Extra dimensions
        - TeV gravity

Higgs

Search for the Higgs boson

- This is not a definitive Higgs analysis
  - Only a quick overview
- *ATLAS* Higgs search combines results from 5 channels
  - $H \rightarrow ZZ \rightarrow 4\ell$
  - $H \rightarrow \gamma\gamma$
  - $H \rightarrow WW \rightarrow e\nu\mu\nu$
  - $H \rightarrow bb$
  - $H \rightarrow \tau\tau$
- Combined with advanced statistical techniques
  - More complex than cross section $e + \mu$ combination
- Combined result provides greater significance to the result
  - No individual channel provides a $\sigma = 5.0$ discovery
  - But the combination.....

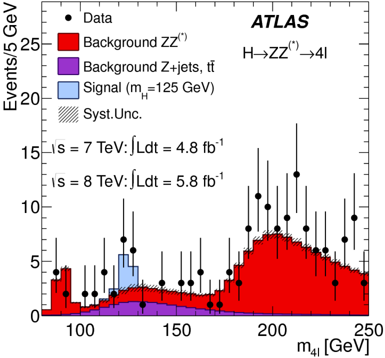
Search for the Higgs boson $H \to \gamma\gamma$

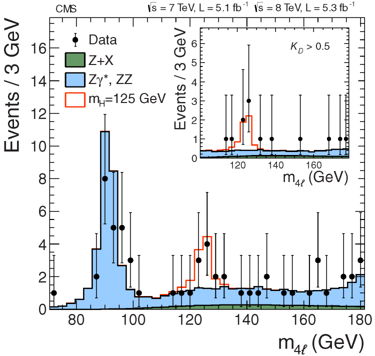


- Weighted data from 2011 and 2012 shown with background
- Invariant mass of 2 photons
- Clear bump observed around 125 GeV
    - Is is statistically significant? Not by itself

Higgs

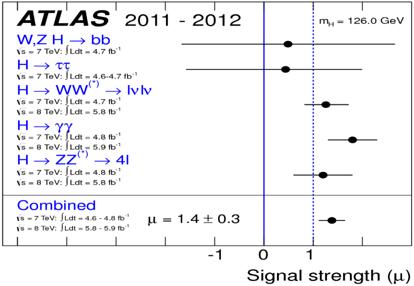
Search for the Higgs boson $H \to 4\ell$





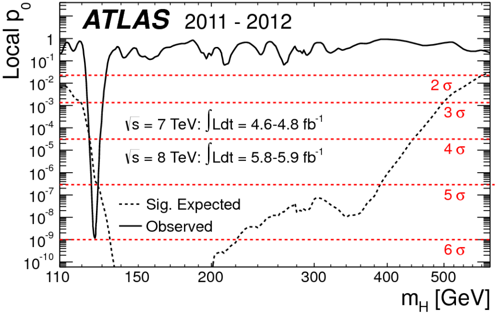
- Data from 2011 and 2012 shown with background
- Invariant mass of 4 leptons
- Bump observed around 125 GeV
    - Only a few events, but are they significant? Not by themselves

Higgs boson - Combining results



- Global signal strength
  - $\mu = 0$ corresponds to background-only hypothesis
  - $\mu = 1$ corresponds to SM Higgs + background hypothesis
- Combined result appears to be in excess of SM Higgs
  - Uncertainties are still to large
  - Result is consistent with SM Higgs hypothesis

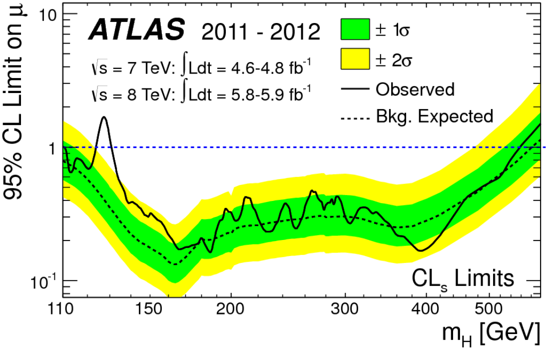
Higgs boson - Combined p_0 result



- p_0 is the probability that the background can produce a fluctuation greater or equal to the excess observed in data
- Dotted line shows SM Higgs + background hypothesis
- In range 122-131 GeV, $p_0 = 1.7x10^{-9}$ corresponding to $\sigma = 5.9$
  - $\sigma = 5.0$ is the de-facto standard for discovery in physics

Higgs

Excluding the Higgs boson via CL_s



- CL_s is a method used to exclude theories
- *"It is our duty to be sceptical i.e. to try to falsify or exclude and this is not the same as, but rather complementary to, the determination of confidence intervals"* - A.L. Read (2000)
- All Higgs masses below the blue dotted line are excluded at 95%

Recent Higgs results

ATLAS Higgs results

- Mass of new particle $M_H = 126 \pm 0.4\,(\mathrm{Stat.}) \pm 0.4\,(\mathrm{Syst.})$ GeV
- Signal strength parameter $\mu = 1.4 \pm 0.3$ is consistent with SM Higgs hypothesis
- $p_0 = 1.7x10^{-9}$ corresponding to $\sigma = 5.9$
- New particle decays to pairs of vector bosons whose net electric charge is zero $\Rightarrow$ It's a neutral boson
- Observation in diphoton ($H \rightarrow \gamma\gamma$) disfavours spin-1 hypothesis
- More data needed

CMS Higgs results

- Mass of new particle $M_H = 125.3 \pm 0.4\,(\mathrm{Stat.}) \pm 0.5\,(\mathrm{Syst.})$ GeV
- Local excess observed with $\sigma = 5.0$
- Observation in diphoton ($H \rightarrow \gamma\gamma$) disfavours spin-1 hypothesis
- More data needed

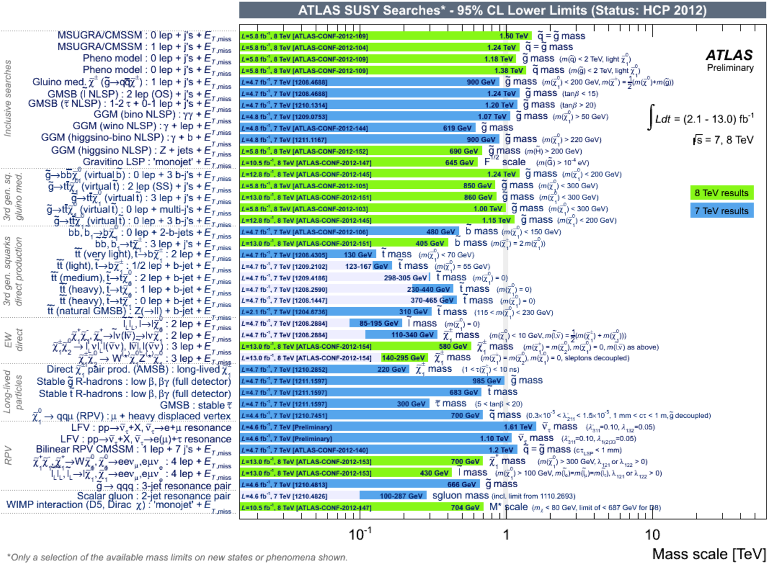
Beyond the standard model

Higgs isn't the only game in town

- *ATLAS* and *CMS* are involved in many searches
- Super Symmetry (SUSY)
  - Many many different models
  - $\approx$ 1.3 models per SUSY theorist (some have more than 1)
- Extra dimensions
  - How many extra dimensions?
  - What are their scale?
  - How strong is gravity in these dimensions?
- Additional Bosons - Z' and W'
- Additional Quarks - b' and t'
- Quark compositness

- All analysis are producing exclusion plots

SUSY Exclusions



*Only a selection of the available mass limits on new states or phenomena shown.
All limits quoted are observed minus 1σ theoretical signal cross section uncertainty.

Other Exclusions

*Only a selection of the available mass limits on new states or phenomena shown

Conclusion

Summary of what I hope you have learnt

- Simulating physics collisions is complex, difficult but necessary
- Simulating detectors is complex, difficult but necessary
  - We do a very good job
  - We can't get the simulation 100% accurate
- We can correct MC performance to match data performance
  - Re-weight MC to match MC efficiencies to data
  - Smear MC 4-vectors to match MC resolutions to data
- Cross section analysis is actually quite involved
- Systematic uncertainties are most of a measurement
- Higgs (probably) exists and SUSY (probably) doesn't