Methods and Techniques in Experimental Particle Physics

Eram Rizvi

Queen Mary University of London

Aim to introduce methods useful for your analysis

Not found in textbooks - HEP folklore...

Details depend crucially on analysis

Only generalities given

- Monte Carlo Models
- Reweighting Techniques
- Event Signatures
- How to Measure a Cross Section
- Purity and Acceptance
- Systematic Uncertainties

Monte Carlo

- What is it?

A program to calc. cross-section of a specific physics process in a defined kinematic region.

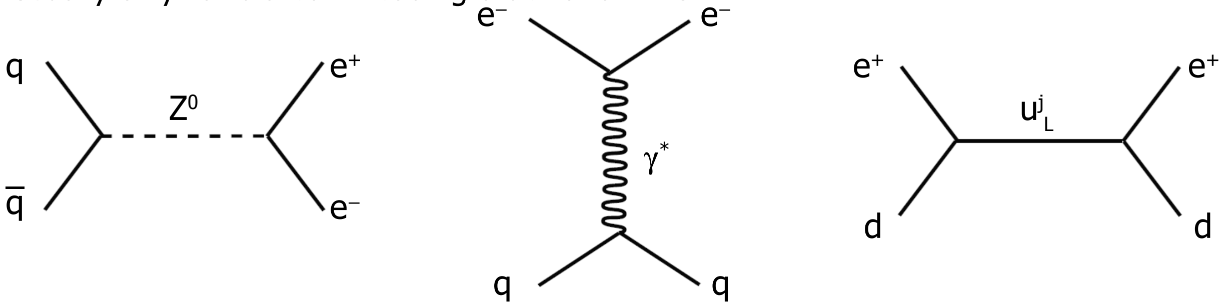
Code uses random number generator to integrate - hence name Monte Carlo

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

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

Cross-section randomly sampled over phase space. Events generated resulting in hard fermion scattering.

Usually only calculated in leading order or at NLO



Consider Drell-yan production of di-electron pair at LHC

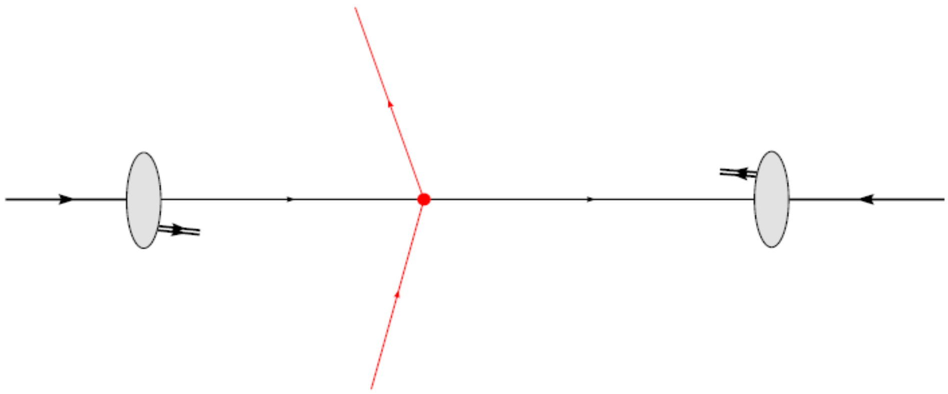
pp Event Generator

Start with two protons

Stefan Gieseke · DESY MC school 09 *11/42*

Generate hard sub-process

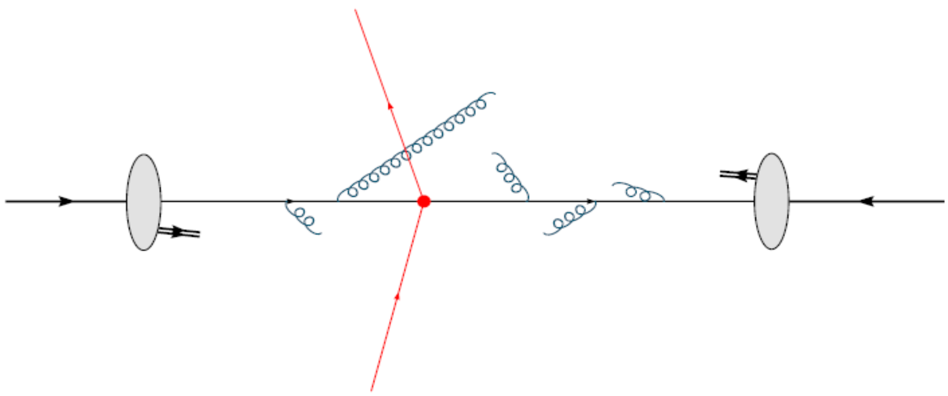
pp Event Generator



Stefan Gieseke · DESY MC school 09 11/42

Add initial state & final state bremstrahlung

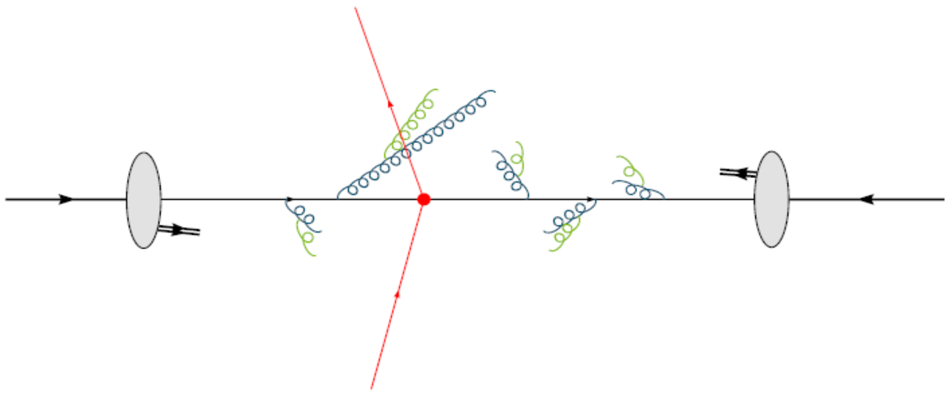
pp Event Generator



Stefan Gieseke · DESY MC school 09 11/42

Add parton showers

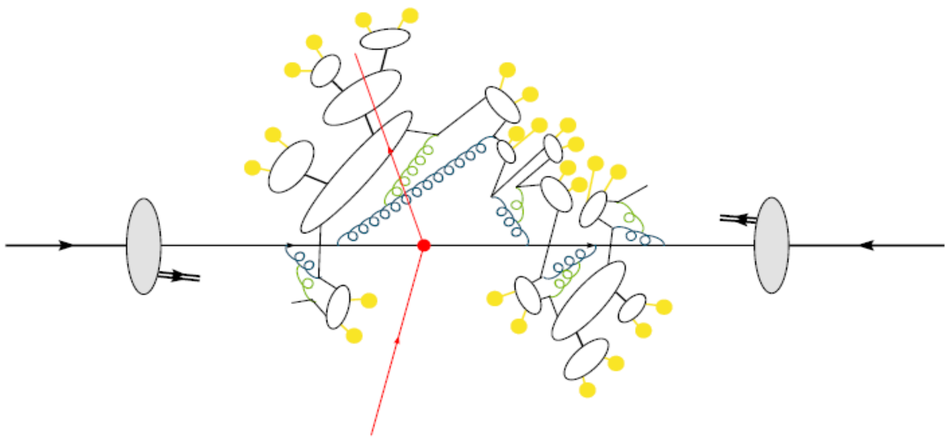
pp Event Generator



Stefan Gieseke · DESY MC school 09 11/42

Hadronise the event: partons → colourless hadrons

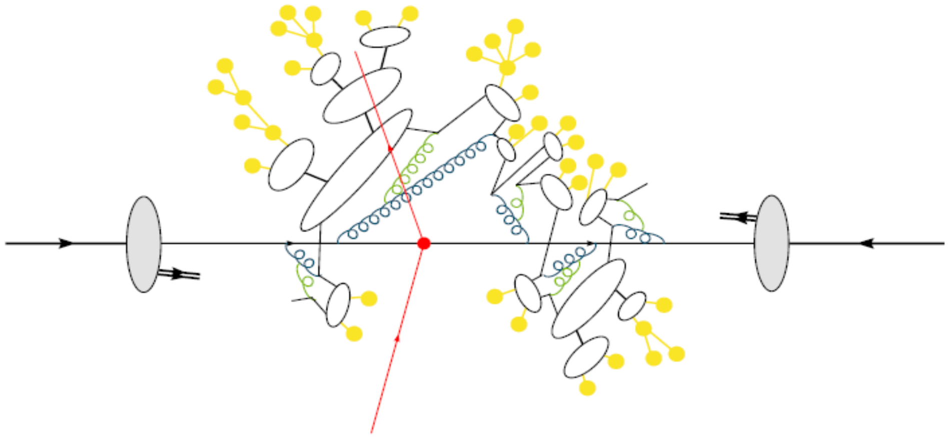
pp Event Generator



Stefan Gieseke · DESY MC school 09 11/42

Add hadronic decays

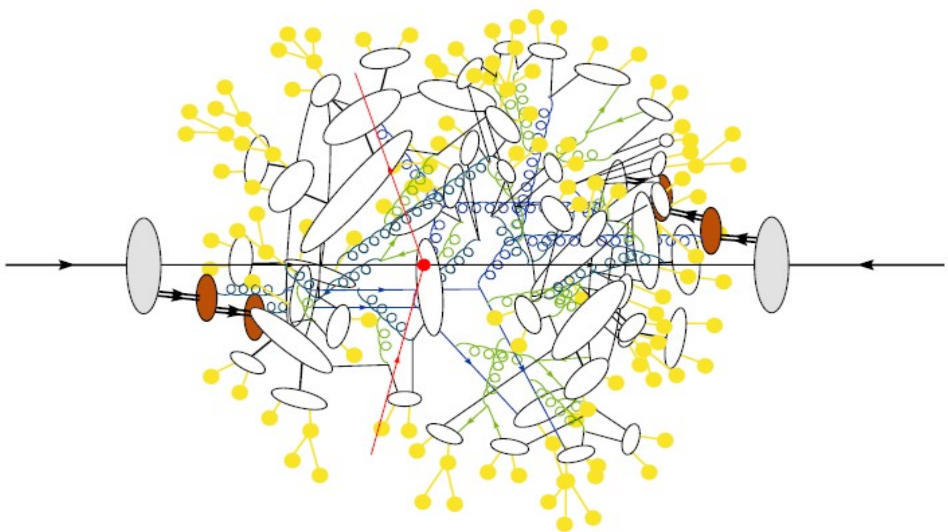
pp Event Generator



Stefan Gieseke · DESY MC school 09 11/42

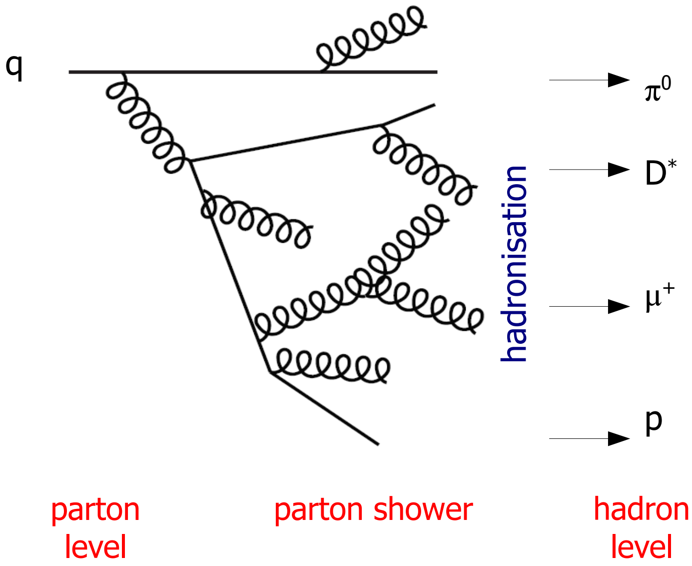
Deal with proton remnants → the underlying event!!!

pp Event Generator



Stefan Gieseke · DESY MC school 09 11/42

- Higher order processes are statistically simulated at parton level known as parton showering
- High energy particles radiate until all particles have low energy, then forced into hadrons → JETSET (and the LUND string model)



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

Monte Carlo 22/11/10

- Each experiment creates simulation of detector: description/location of all major materials in detector construction
- Program GEANT uses generator output (particle 4-vecs.). It simulates interaction of particles within detector volume:
  - particle ionisation in trackers
  - energy deposition in calorimeters
  - intermediate particle decays/radiation
- GEANT code merged with (experiment specific) detector simulation.
- Final output is raw data:
  - charges measured on each tracker wire
  - Elec. pulses in each calo. anode / photomultiplier
- MC events now in same format as real raw data

- Final stage to pass "raw data" through same reconstruction as real data. Output now in form useable for analysis:
  - particle trajectories
  - energy depositions in calorimeters
  - time of flight info
- Now have totally simulated "data" with complete info about underlying physics
- Detector simulation can be very detailed and CPU intensive (>10min/event!)
  - time dependent simulation of noise in detector
  - tracking all particles throughout complete detector volume
  - "sick channel maps" reflecting degraded parts of sub-detectors
- Simulations may be simple with only parametrised efficiencies and resolution functions (e.g. Atlfast)

Beware!

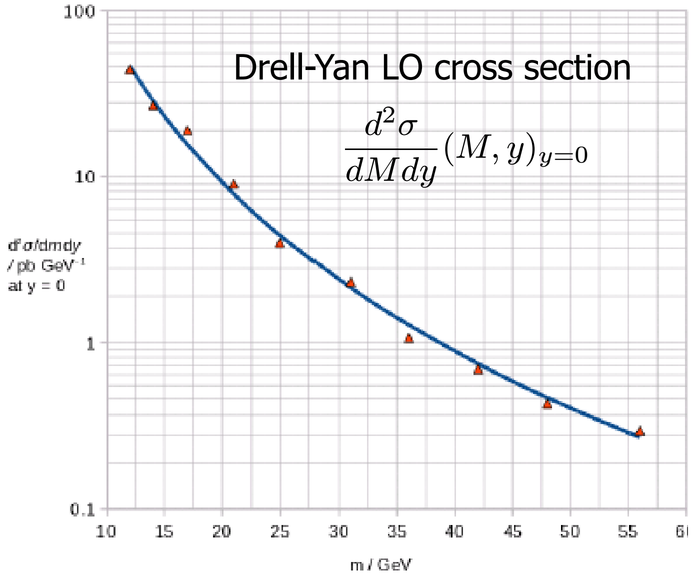
- MC is NOT the truth!
- Only simulates specific physics processes
- The cross-section calculation may be:
  - wrong / bugs
  - incomplete
  - inaccurate
  - have false kinematic dependence
  - etc.....
- Some processes calc'd in LO, the NLO corrections may be large
- Some MCs may include Phenomenological Models of higher order processes e.g.
  - Soft QED bremstrahlung
  - Hadronisation (combination of quarks/gluons into observable hadrons)

Higher order corrections are often included as a “K-factor”

$$K = \frac{\sigma^{NLO}(NLO\ PDFs)}{\sigma^{LO}(LO\ PDFs)}$$

or sometimes use NLO partons in LO matrix element

$$K = \frac{\sigma^{NLO}(NLO\ PDFs)}{\sigma^{LO}(NLO\ PDFs)}$$



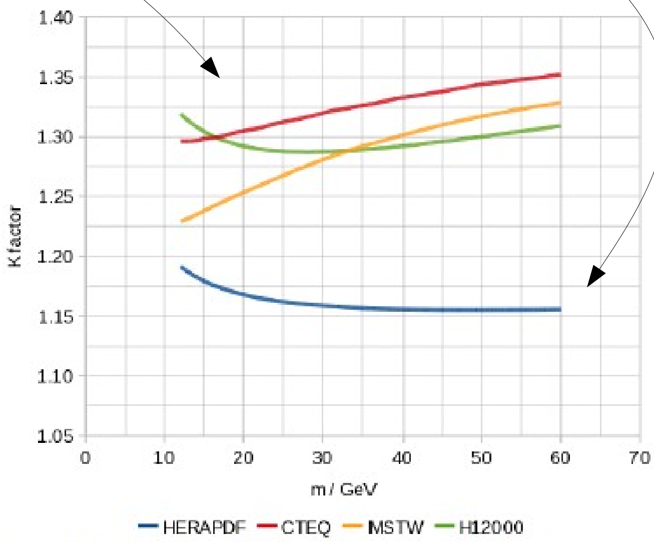


Figure 15: the K factors calculated using different PDF sets.

- MCs have many parameters describing details of models.
  - reduced prob. factor to produce s quarks compared to u or d quarks
  - initial beam energies
  - input value for α_s
- In general most generators work well, parameters are tuned
- BUT, bugs have been found (still) ! ...and new MCs are being written for LHC
- Also, simulation is most often is wrong - cannot take into account all detector problems
  - many probs not seen till long after data have been taken
  - simulated detector noise may not be accurate

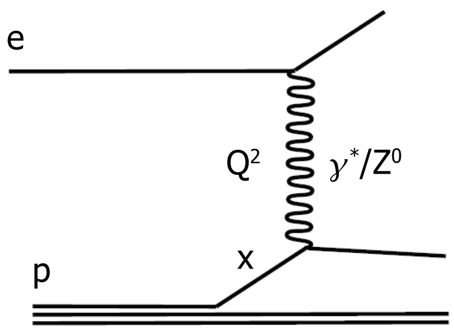
Example of MC jobOptions file

text list of all tunable parameters and values several pages long!

```
Feb 13, 01 15:17                 mcsteer.txt                     Page 1/2
*-----------------------------------------------------------------------
C0JG 0 / ! DJANGO 6.2  PARAMETERS
*-----------------------------------------------------------------------
*  DJANGO 6.2 parameters and steering variables available to the user
*  are stored in the following commons and can be changed as indicated
*  below. They are explained in the DJANGO and HERACLES manuals.
*
*
*---LEPTO parameter commons
*      COMMON /LEPTOU/ CUT(14),LST(40),PARL(30),X,Y,W2,Q2,U
*
*      The parameters and their defaults are descibed in the LEPTO
*      manual. Please note that LST(2) to LST(19) and PARL(1) to PARL(19)
*      are input parameters, whereas LST(20) to LST(40) and PARL(20) to
*      PARL(30) are output parameters.
*
*  Changing parameters:
*  The order and the number of parameters to be changed are arbitrary.
*  In case of a one-dimensional array only the first index is used, the
*  second index may have any value.
*  The basic rules (for details see the BOS manual) for writing card
*  images are:
*      1) columns 1-4 are free
*      2) items (names, indices, values) without inside blanks
*      3) blank(s) between items
*      4) the array names in PAR-NAM have to be shortened to four
*         characters if necessay
*      5) comment cards start with '*'
*      6) space behind "!" in each line reserved for comment
*
*   PAR-NAM  1.INDX  2.INDX NEW-VALUE ! COMMENT IF NEEDED
*   -------  ------  ------  -------   ---------------------------------
*
**************** HERACLES steering cards ******************************
*
*   'EELE'     0       0      26.7     ! (D=30.) Energy of electron
    'EELE'     0       0      27.6     ! (D=30.) Energy of electron
*                                        beam in GeV
*   'POLA'     0       0       1.      ! (D=0.) polarization of lepton
*                                        beam ( -1. ... 1. )
    'LLEP'     0       0      -1       ! (D=-1) Lepton charge
*                                        1=positron, -1=electron
    'EPRO'     0       0      920.     ! (D=820.) Energy of proton beam
*                                        in GeV
    'Q2WG'     0       0      $q2wg     ! (D=0.) Q^2-weighting of events
*                                        the value specifies below which
*                                        the events will be weighted
*   'ICUT'     0       0       1       ! (D=3) definition of cuts applied
*                                        1=cuts in x and lower Q2
*                                        2=cuts in x, lower Q2, W
*                                        3=cuts in x, y, Q2, W
*
*                                        must be 3 for the use of THMI!
    'XMIN'     0       0     $xmin    ! (D=1E-4) lower cut in x
    'XMAX'     0       0     0.999      ! (D=1.0 ) upper cut in x
*   'YMIN'     0       0     0.001     ! (D=0.0 ) lower cut in y
*   'YMAX'     0       0     0.99       ! (D=1.0 ) upper cut in y
    'Q2MI'     0       0     $q2min      ! (D=4.0 ) lower cut in Q**2
    'WMIN'     0       0     5.0       ! (D=5.0 ) lower cut in W (W2 90)
*ER generate electrons only up to 133 Deg
    'THMI'     0       0     $thmin      ! (D=0.0 ) lower cut in Theta_el.
*                                        in rad. wrt. the incoming electron,
Tuesday February 13, 2001                                    mcsteer.txt
```

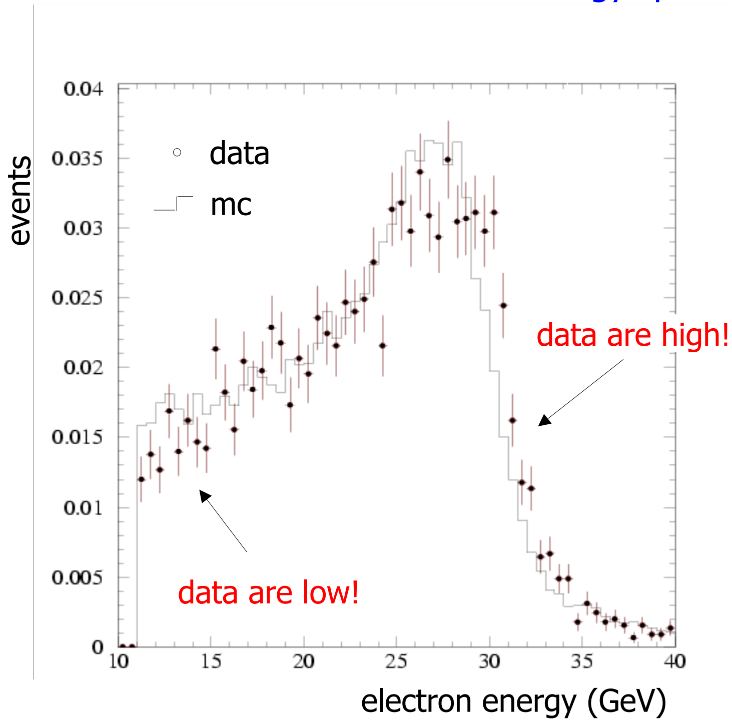
```
Feb 13, 01 15:17                 mcsteer.txt                     Page 2/2
*                                        ICUT must be 3, if THMI is used!
*                                        extremely small THMI-cuts (mrad) must
*                                        be avoided!
*   'PTMI'     0       0       0.      ! (D=0.0 ) lower cut on the transverse
*                                        momentum of the scattered electron
*   'EGMI'     0       0       0.5     ! (D=0.0 ) cutoff energy for
*                                        bremsstrahlung photons
*                                        0 =radiative and non-rad. ev.
*                                        >0=only hard photon bremsstr.
    'INT2'     1       0       1       ! (D=0) integration of the non-
*                                        radiative contribution (NC)
*                                        0=off, 1=on
*   'INT2'     2       0       1       ! (D=0) integration of the non-
*                                        radiative contribution (CC)
*                                        0=off, 1=on
    'INT3'     1       0       5       ! (D=0) integration of initial   (radcor)
*                                        state bremsstr. contr. NC
*                                        number of iterations for VEGAS
*                                        <100:VEGAS, >100:VEGAS1/2
    'INT3'     2       0       5       ! (D=0) as INT3(1) but for final (radcor)
*                                        state bremsstr. NC
    'INT3'     3       0       5       ! (D=0) as INT3(1) but for Compton (radcor
)
*
*                                        contribution NC
*   'INT3'     7       0       5       ! (D=0) integration of initial
*                                        state bremsstr. contr. CC
*                                        number of iterations for VEGAS
*                                        <100:VEGAS, >100:VEGAS1/2
    'NPVE'     0       0   10000       ! (D=1000) number of integration
*                                        points used for VEGAS
*   'LPIN'     1       0       1       ! (D=2) EW-parameters
*                                        1=from Mw,Mz; 2=from Mz, G
*   'LPIN'     2       0       0       ! (D=1) only Born x-section term   (radcor
)
*                                        or 1=including corrections
*   'LPIN'     4       0       0       ! (D=1) leptonic QED corrections
*                                        1=included, 0=excluded
*   'LPIN'     5       0       1       ! (D=0) quarkonic QED corrections
*                                        1=included, 0=excluded
*   'LPIN'     6       0       1       ! (D=0) lepton-quark interference
*                                        1=included, 0=excluded
*   'LPIN'     7       0       0       ! (D=2) fermionic contr. to
*                                        the photon self energy
*                                        0=no 1=with q-mass 2=param.
*   'LPIN'     8       0       1       ! (D=0) fermionic contribution
*                                        to the gamma-Z mixing
*   'LPIN'     9       0       1       ! (D=0) fermionic contr. to the
*                                        self energy of Z-boson
*   'LPIN'    10       0       1       ! (D=0) fermionic contr. to the
*                                        self energy of W-boson
*   'LPIN'    11       0       1       ! (D=0) purely weak contr. to self
*                                        energies, boxes, vertex corr.
*   'M111'     0       0       2       ! (D=1) =0 fixed alpha_s (P111)
*                                        =1 running alpha_s to 1st order
*                                        =2 running alpha_s to 2nd order
*   'M115'     0       0       2       ! (D=0) =0 diverge at q2-->0
*                                        =1 softened diverge at q2-->0
*                                        =2 fixed below q2=4
*   'P111'     0       0       0.2     ! (D=0.2) fixed alpha_s value
    'P112'     0       0       0.235   ! (D=0.25) Lambda used for alpha_s
                                                                 1/1
```


- Some example follow but are HERA specific ⇒ quick guide to HERA physics...
- At fixed beam energy Deep Inelastic Scattering (DIS) characterised by 2 variables: x, Q^2:
  - x = fraction of P^+ momentum carried by struck quark
  - Q^2 = virtuality of exchanged boson



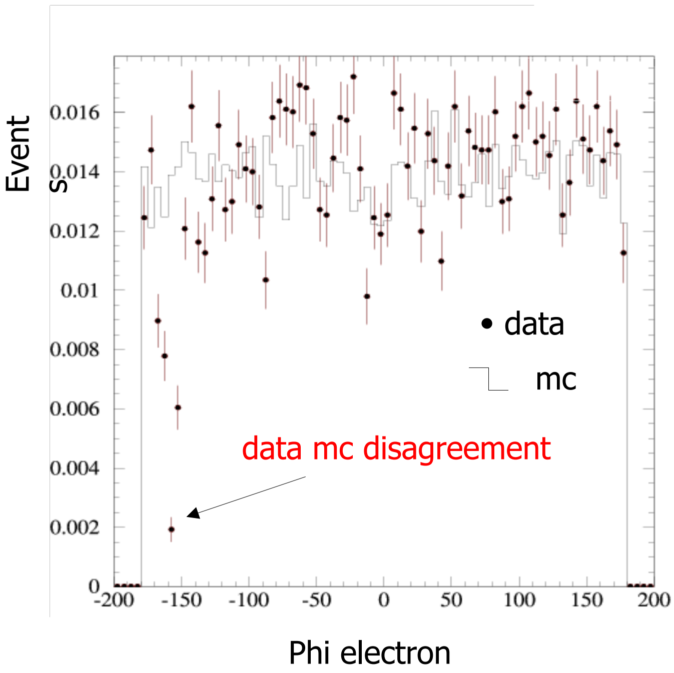
- electron : source of gauge bosons (W,gamma,Z) to probe p^+ ⇒ QCD!
- measurement of clean isolated e gives good measurement of x, Q^2 via θ_e and E_e
- scattered e^- energy spectrum sensitive to quark densities in p^+

- Back to the point....
- This is the electron energy spectrum for DIS data and MC



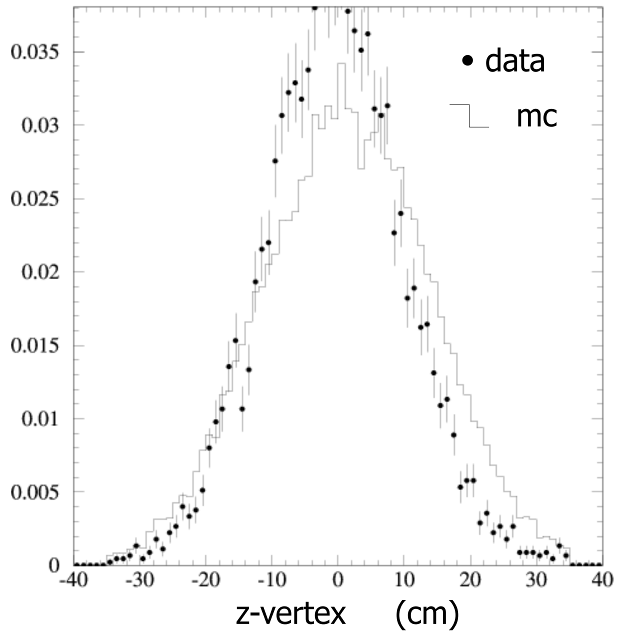
- Both spectra disagree!
- At low energy disagreement due to outdated input proton structure in MC
- At high energy disagreement due to better energy resolution of detector in MC
- Can be several reasons why data/mc disagree. Your data may be right!

- normalised phi spectrum of scattered electron
- physics is (generally) phi symmetric - oblems due to detector effects

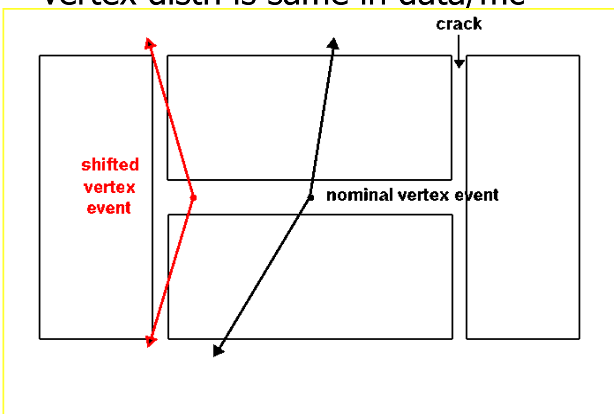


- data taken in 2000 during period when central tracking chamber broke one wire
- efficiency very much reduced in this region
- not (yet) simulated in mc!

- dist. of z position of event vertex can be important
- detector is designed with nominal mean vertex position
- if different then efficiency can be affected



- beam optics of accelerator changes mean vertex position for every fill
- detectors have cracks - should be simulated. Modelling only correct if vertex distn is same in data/mc



Summary

- MC is a useful tool - should contain best knowledge of physics + detector understanding
- This is often not the case
- What to do about it?
  - treat it with care
  - talk to people! you may have a problem that was understood and fixed months ago
  - your measurement should be independent on MC assumptions
  - try to isolate the problem e.g. generator or simulation?
  - reweight your mc to describe your data distributions

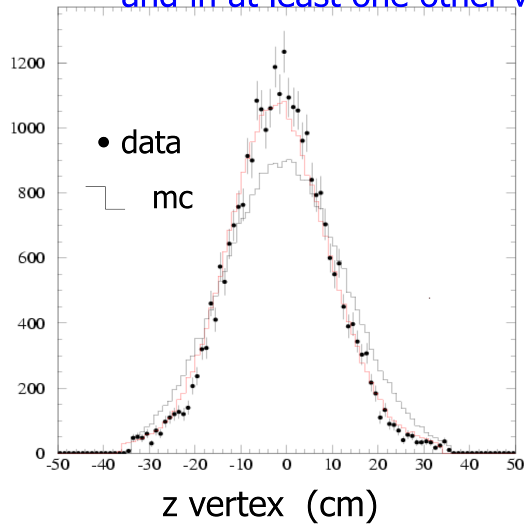
Reweighting MC

- Aim is to alter any given distribution of events to look like another, e.g. make mc spectra look like data distribution.
- Each event contributes to any histogram with a "weight" of 1
- The weight is altered to be a real number such that a distn. looks like another function
- See section on MC techniques in Particle Data Group Book
- Do it carefully - its subject to some constraints
- Ensure errors are calculated correctly!
- Simplest case is to set weight for all events to same value e.g. weight=2
  - this affects only normalisation of the distn.
  - this effectively doubles the number of events

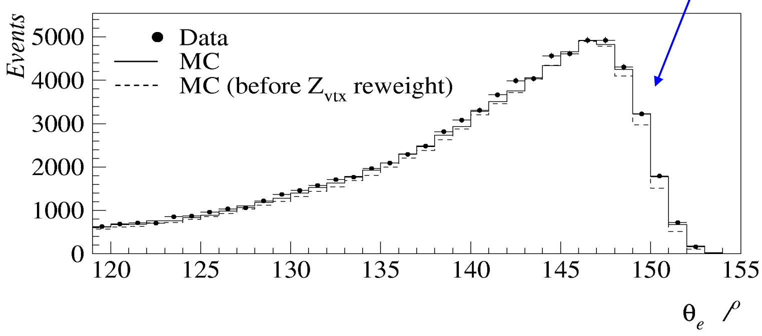
Look at example of reweighting z-vertex distn.

- This has a simple theoretical model
- Constrained to keep total number of events the same: mc luminosity unchanged by reweighting
- Theory: spread in z-vertex distn. due to finite length of each bunch of beam particles - expect Gaussian distribution
- At HERA σ_{vertex} ~ 10 cm. The <z> position changes with each machine fill dependent on beam optics etc...
- Generate MC with Gaussian vertex width of 13cm. MC now described by Gaussian function: $G_{mc}(z)$
- Data distn. is also Gaussian: $G_{data}(z)$
- Each mc event is now given a weight = G_{data} / G_{mc}

- Beware! MC generated with specific number of events + specific cross-section: reweighting SHOULD NOT affect normalisation of MC
- Add requirement: $\sum_i w_i = N\,(number\ of\ events\ generated)$
- Check the influence of the reweight
  - in variable used for reweight (z)
  - and in at least one other variable (sensitive to z-cracks in detector)



reweighting improves theta spectrum!

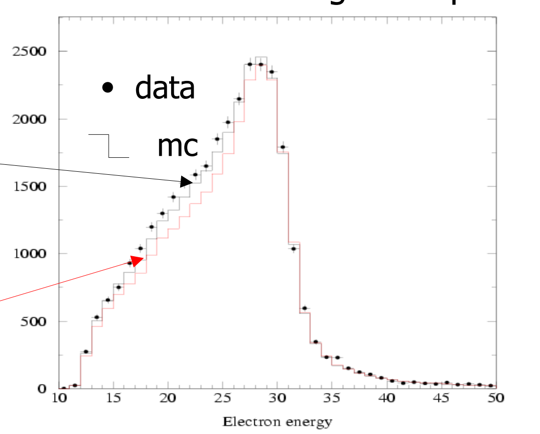


Example 2

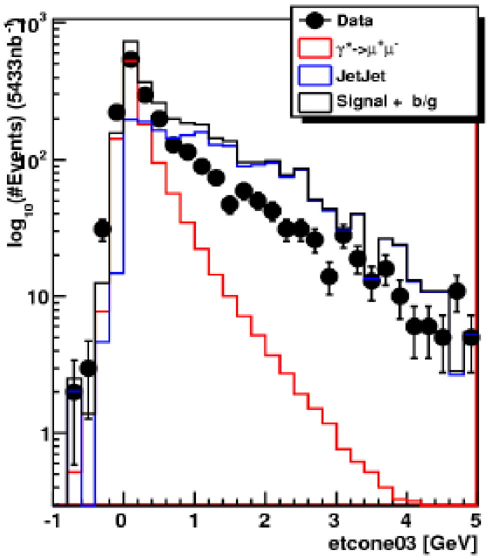
- Input physics of mc can be wrong - improve by reweighting. This time the cross-section calc. is 'wrong'
- Need fast analytical parametrisation of old and new cross-section.
- $w_i = \sigma_{new}(x_i, Q^2_i) \,/\, \sigma_{old}(x_i, Q^2_i)$ for every event i
- σ_{old} must be identical to mc input
- No constraint on normalisation since old cross-section can be wrong in shape and normalisation

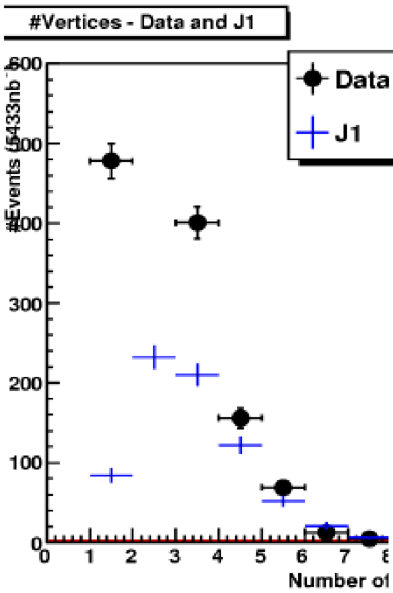
reweighted MC - better description of data (not perfect!)

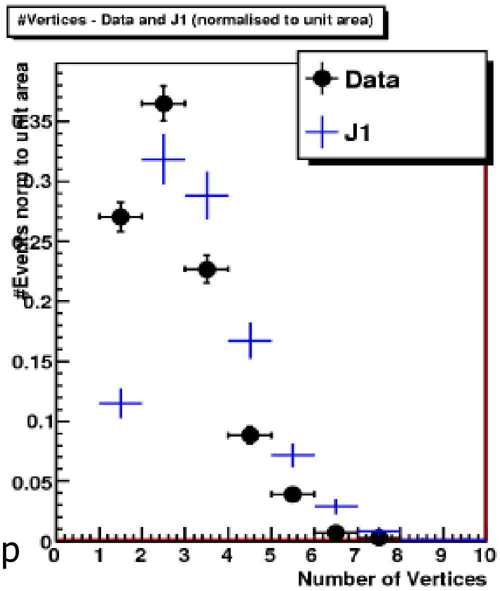
old MC has different shape and normalisation!



Example 3



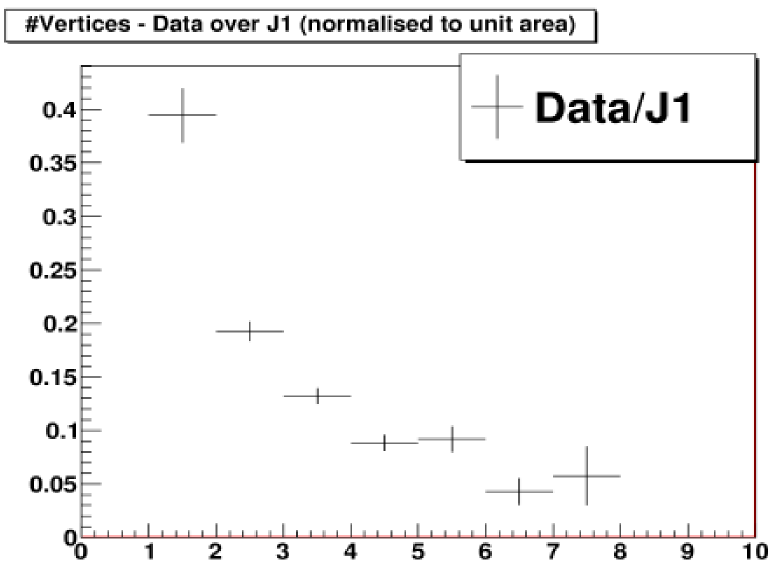


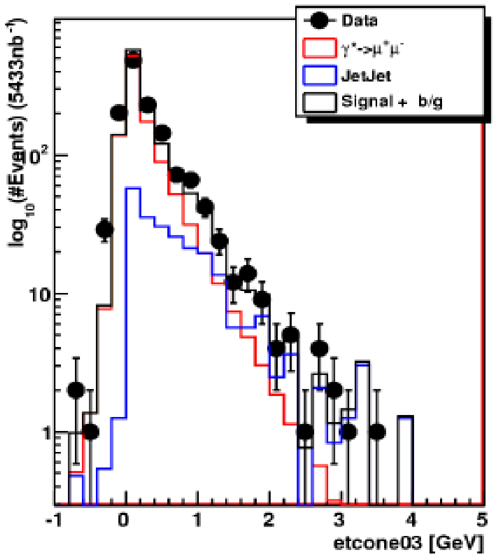


Try to describe isolation energy
Energy in cone around a muon

Depends on number of pile-up events in data and MC!

Example 3





reweight MC: ratio of number of vertices data/MC
Constraint: ensure average weight = 1

better description of isolation energy!

Summary

- Reweighting is very versatile technique
- Can also be used in simulation e.g. to model detector holes
- Two words of caution:
  - Calculate errors correctly: $\sqrt{\sum_i w_i^2}$
  - Watch out for events with large weights. If one corner of phase space has 2 events with w=100: distn is too spiky - weighting scheme too harsh!
- Two weighted events cannot describe the statistical properies of 100 unweighted events

When booking histos in RooT always call the `SumW2` function
RooT will then calc errors correctly whether your data are weighted or unweighted
Need to only call this function once per histo

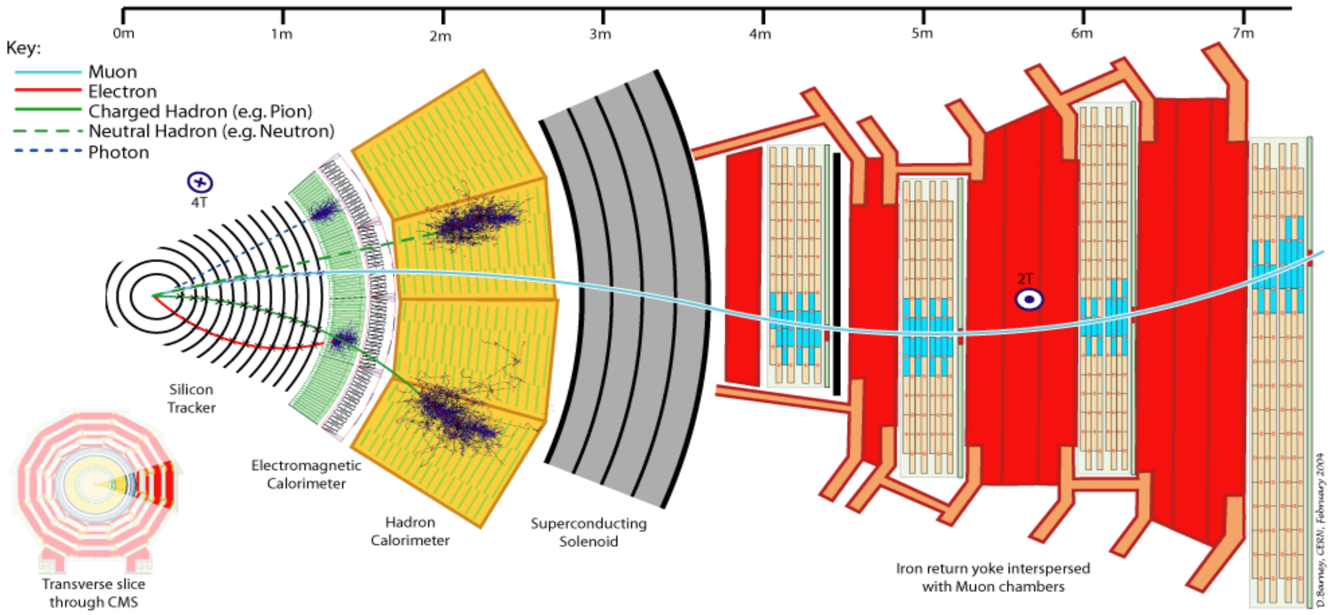
```
TH1F *hist = new TH1F(...);
hist->SumW2();
```


Event Properties

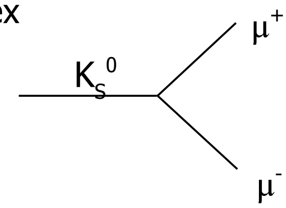
- Measuring cross-section of a process requires recognising event properties:
  - E.M. energy with a charged track — e+ or e-
  - E.M. energy without track — photon
  - hadronic energy — gluon/quark induced jet
  - penetrating charged track — μ+ or μ-
  - missing transverse energy — ν
  - missing longitudinal energy — beam remnants
  - displaced secondary vertex — in-flight decay of 'long lived' particle
- Look at the event topology

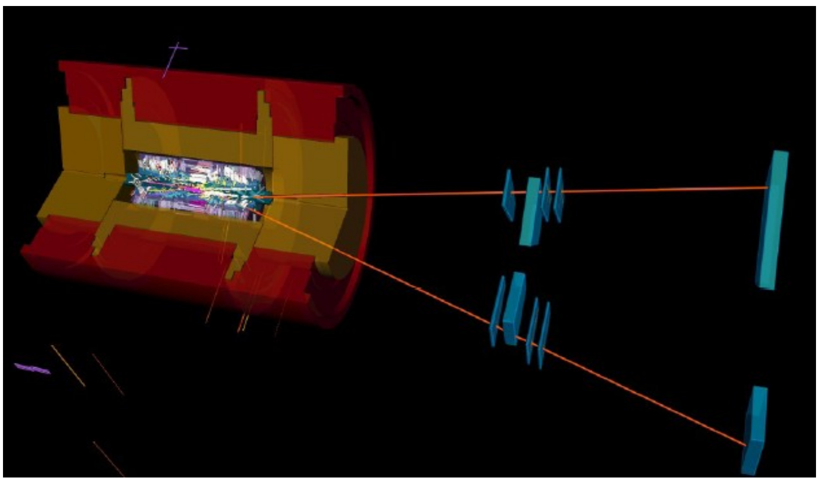
At low momenta some particles can be identified by their characteristic energy loss: dE/dx

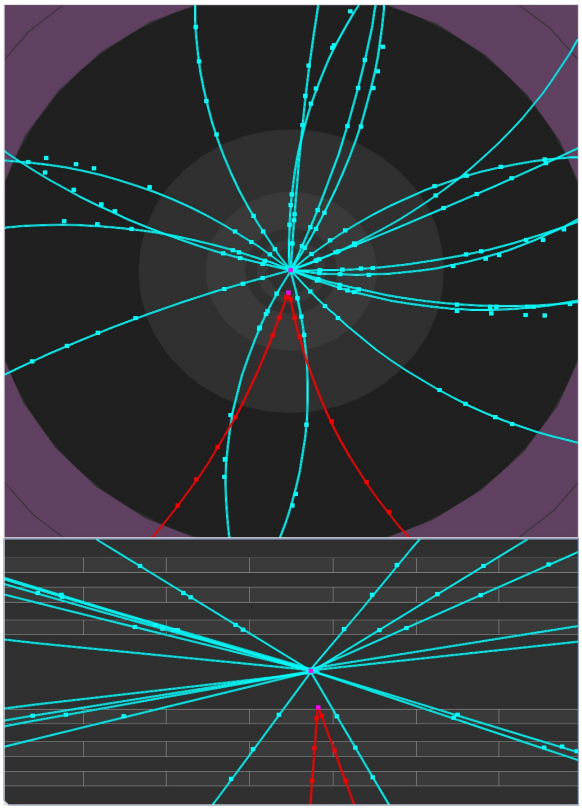
Take a slice through the CMS detector and see how each particle behaves

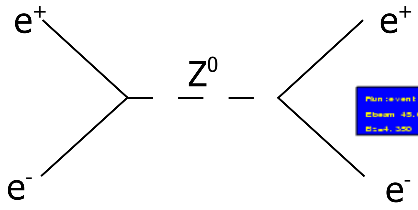


- Two oppositely curved tracks
- Penetrating tracks
- Displaced secondary vertex

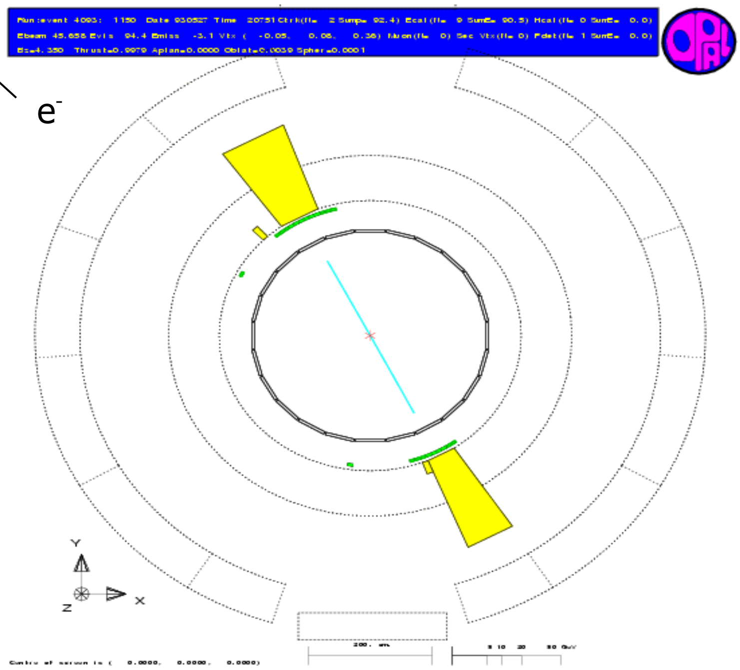




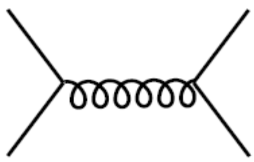




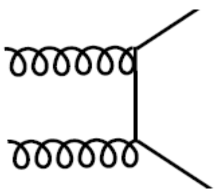
Two oppositely charged tracks back-to-back
Two localised EM energy deposits near tracks
No missing energy



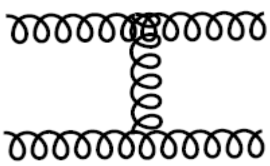
could be:



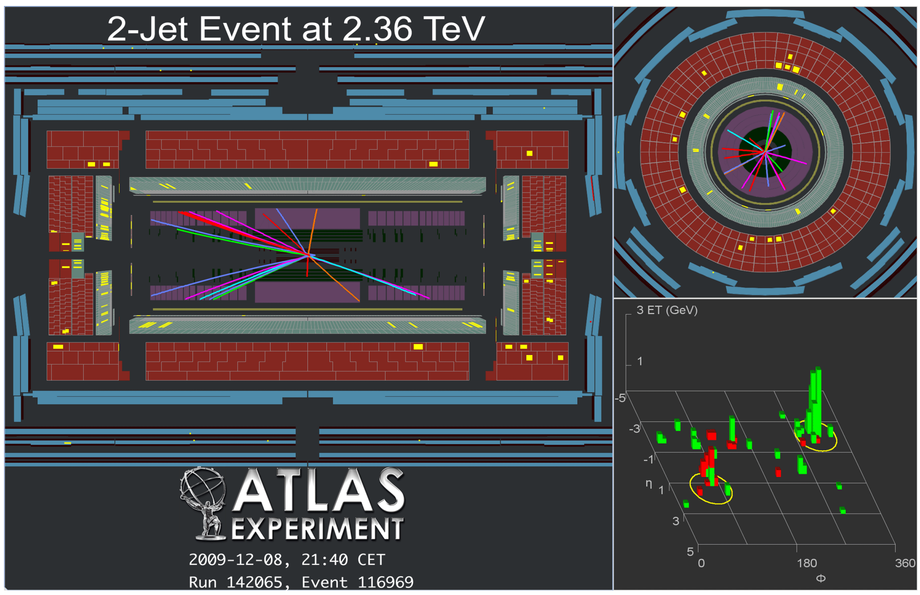
or



or



or...



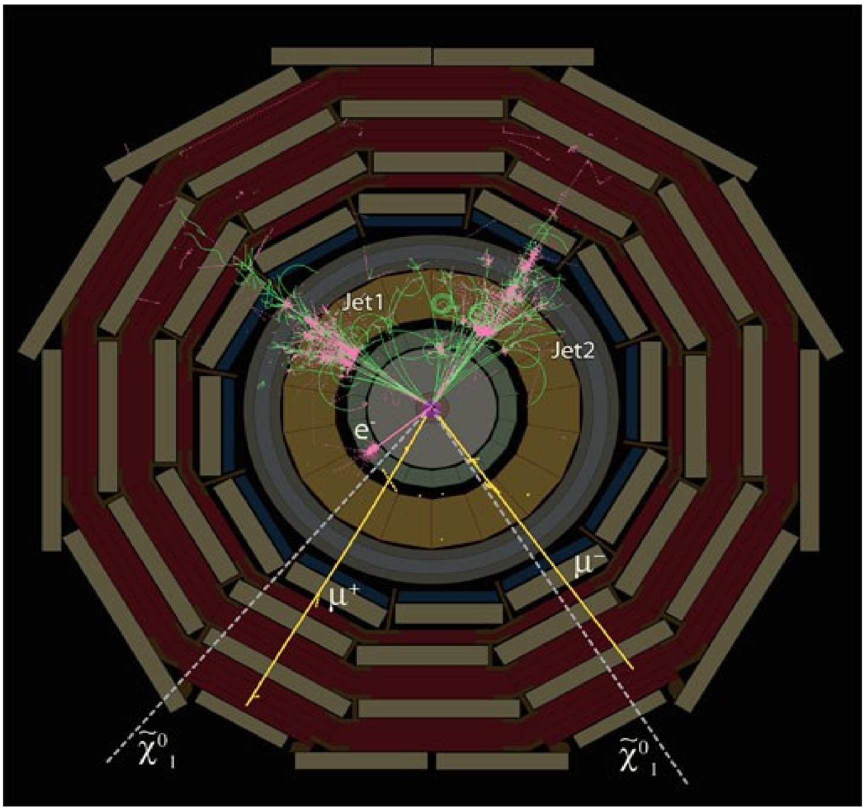
2 jet final state: 2 jets of particles

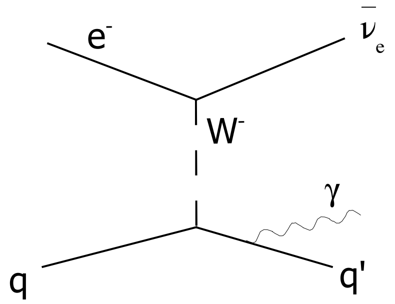
EM and hadronic energy

Event balanced in transverse momentum

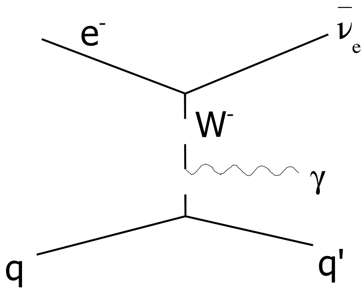
SuSy interaction at CMS

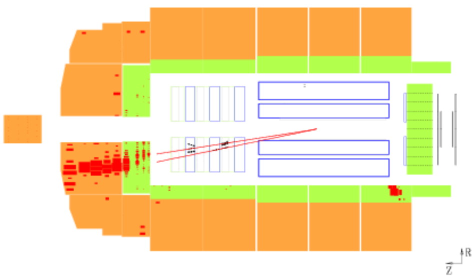
- 2 jets
- $\mu^+\mu^-$
- e^-
- 2 neutralinos

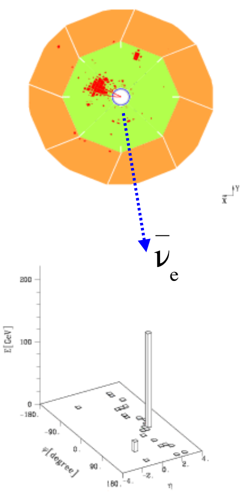




OR





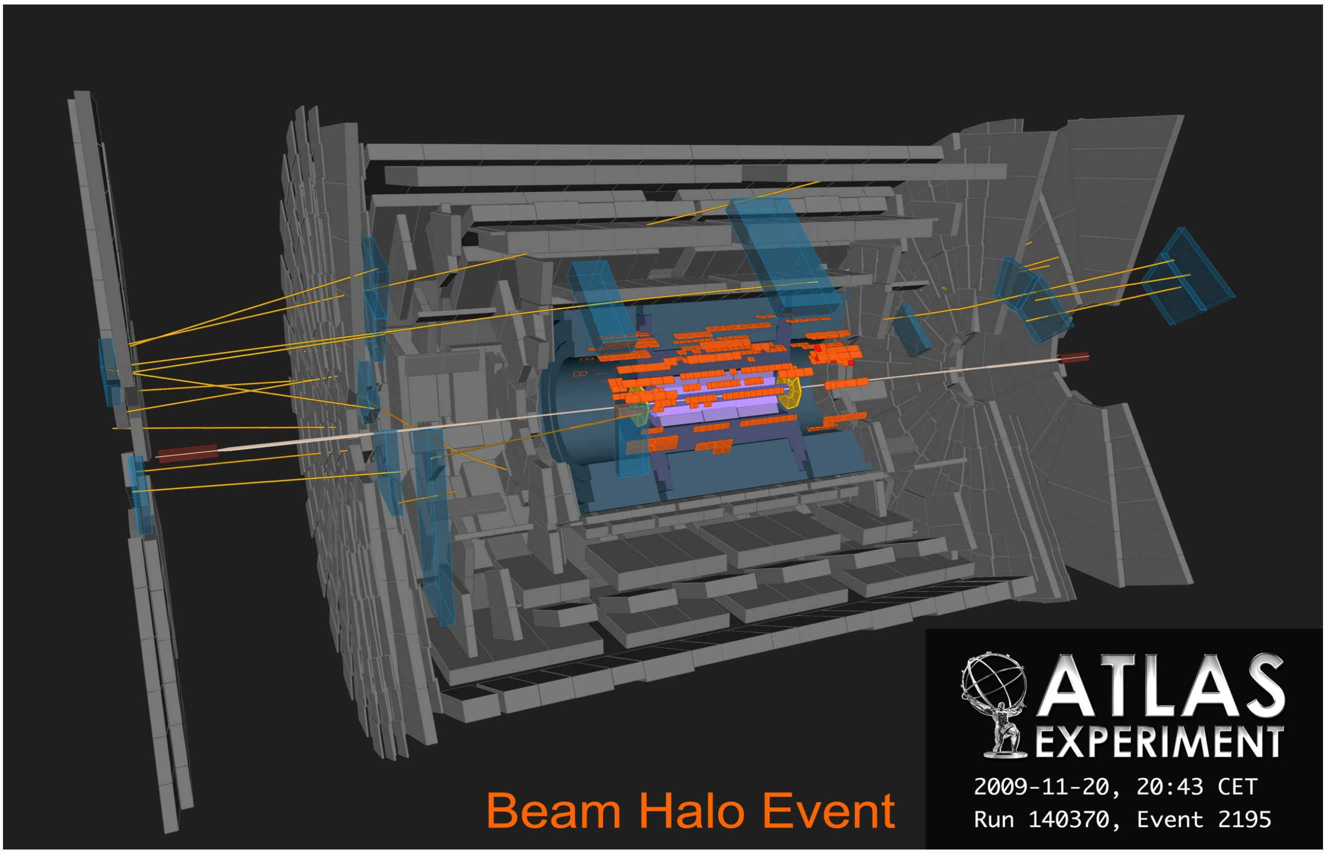


Radiative CC event:

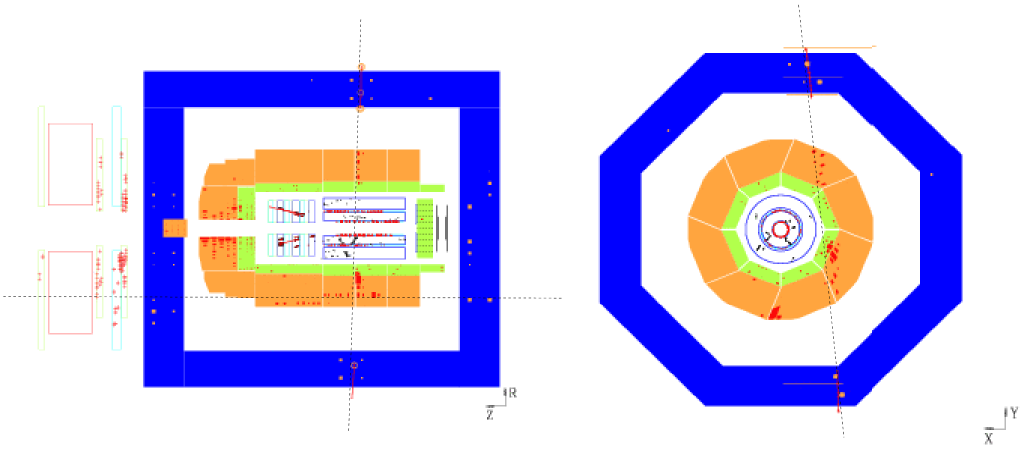
hadronic activity

isolated photon missing

transverse energy ($\bar{\nu}_e$)



A "mean" event - simultaneous overlay of 3 background processes

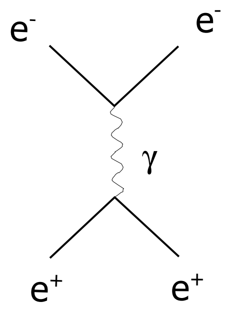


- cosmic muon - vertical penetrating particle
- halo muon - horizontal muon track parallel to p^+ beam
- beam/gas interaction - high activity in forward p^+ direction symmetric in phi

  $\rightarrow$ collision of p^+ with stray gas molecule

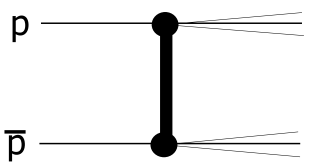
How To Measure a Cross Section

- A cross-section, σ, is defined by: $\sigma = N/\mathcal{L}$
  - N = number of events
  - $\mathcal{L}$ = luminosity collected
- Luminosity is universal measure of amount of data collected defined by above equation, BUT using reference cross-section
- $\mathcal{L}$ is function of beam currents and beam size: high current/small beam = large lumi $\Rightarrow$ higher beam particle density - more chance of collision*
- σ_{ref} is well known or calculable with little uncertainty
  - LEP used Bhabha scattering
  - large cross-section
  - very well known - QED calculable
  - all LEP cross-sections normalised to this process
  - uncertainty of ~0.1%

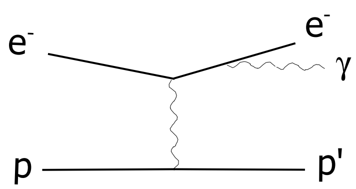


* luminosity is derived in Perkins

- At LHC use total pp cross-section
  - not calculable in QCD!
  - compare with previous measurements from e.g. CERN
  - uncertainty of ~7% (same at Tevatron...)



- At HERA use Bethe-Heitler process:
  - QED calculable process
  - large cross-section - measureable with high efficiency
  - uncertainty typically ~1%



- Measuring a cross-section is easy - simply count events
- NO! need to correct number of observed events to number of expected events given a perfect detector:
  - no holes / cracks
  - perfect efficiency for measuring particles
  - perfect resolution
  - no background events sneaking in
  - careful - must not depend on input MC model

Modified cross-section formula:

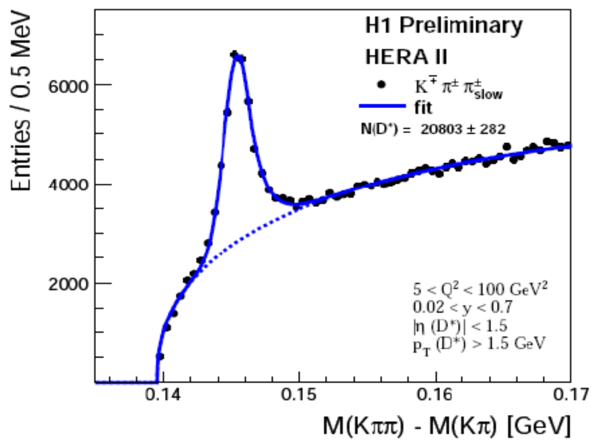
$$\sigma = \frac{N_{obs} - N_{b/g}}{\mathcal{L} \, . \, \varepsilon \, . \, A_{cc} \, . \, Br}$$

- N_{obs} = number of observed events
- $N_{b/g}$ = number of background events
- A_{cc} = Acceptance
- ε = efficiency
- Br = branching ratio

Background

Need to estimate background well
Need to compromise signal efficiency
with b/g rejection
At LHC signal:b/g ratio can be ~10^6 or more!

- try to rely on data driven methods
- use data in side-bands of a mass peak
- can also use MC to estimate

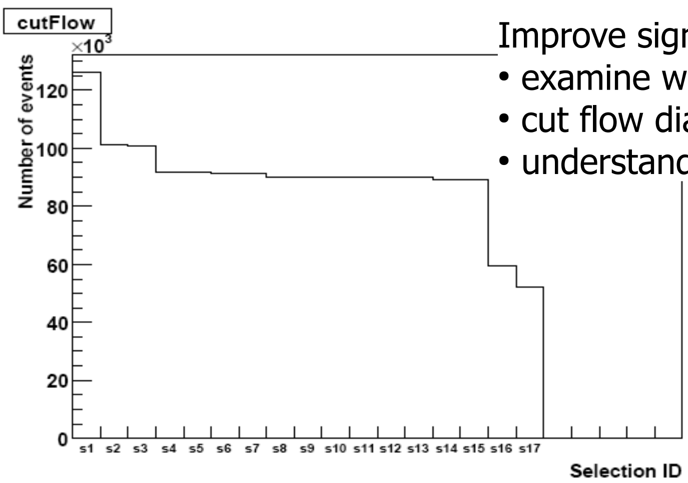


H1prelim-08-072

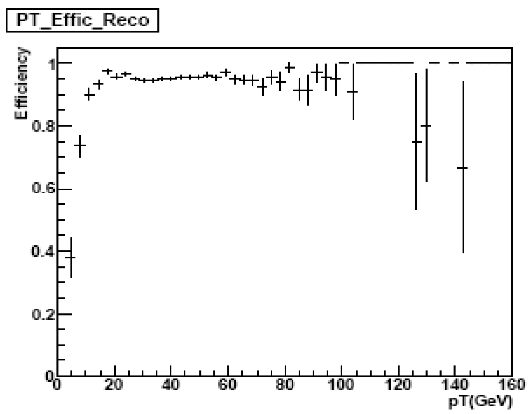
MC estimation:
- tune selection cuts to maximise signal efficiency & minimise b/g
- use this when the b/g is "well known"

Data estimation:
- look in nearby region of phase space dominated by b/g and no signal
- extrapolate into signal region
- be very careful in low statistics analyses!

easy to optimise selections based on handful of events recorded



Improve signal efficiency
- examine which cuts have large effects
- cut flow diagrams help
- understand why there are large jumps



Avoid selections which introduce complex efficiency behaviour
Best is a flat efficiency!
Remember efficiency can be improved with loose cuts
but this will introduce much more b/g

Acceptance

Acceptance corrects for:

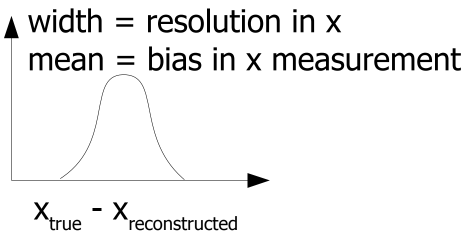
1. Finite resolution: events end up in the wrong bin.
2. Finite selection efficiency: some events end up in no bin at all.

$$A_{cc}=N_{rec}/N_{true}$$

N_{rec} = # Events reconstructed in a bin

N_{true} = # Events generated in a bin (truth)

A_{cc} calculated from Monte Carlo



- A_{CC} depends on shape of cross section
- Assign a systematic error to the acceptance correction
  (vary input cross section: determine change on measured cross section)
- Careful as some of acceptance may be purely geometrical !
  e.g. detector which only measures in one hemisphere has A_{CC}=0.5

Acceptance correction method reliable if:

- Bin widths greater than resolution
  Bin width of 2×resolution ⇒ 5% smearing into neighbour bins (Gaussian resolution)
- Monte Carlo reproduces data distributions well (i.e. resolutions & cross sections)

We are using a diagonal acceptance matrix:

estimate number of corrected events → observed number of events

$$\begin{pmatrix} N_1 \\ N_2 \end{pmatrix}_{est} = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}_{obs}$$

i.e.

$N_1^{est} = A_1 . N_1^{obs}$

$N_2^{est} = A_2 . N_2^{obs}$

we are not using all of the information !

Can use the full correlation matrix using various unfolding & matrix inversion techniques in the case of

1. non-Gaussian resolution,
2. very many dimensions,
3. poor resolutions,
4. pedantry

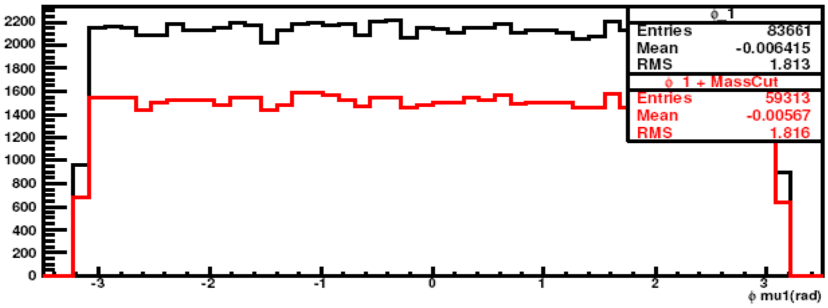
These techniques bring their own problems, and must NEVER be used as a "black box" or universal panacea !

Rule of thumb:

try to keep acceptance corrections ~ 10% i.e. A_{CC}: 0.90 – 1.10

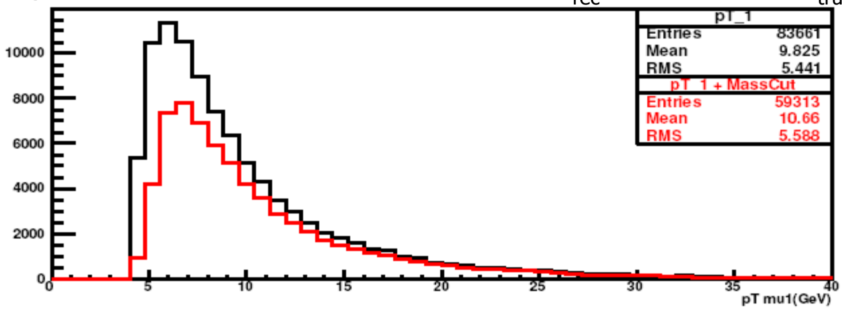
if uncertainty on acceptance is ±10% ⇒ cross section error= 0.1 × 0.1 = 1%

Cannot be achieved in all analyses!



Imagine measuring a 'flat' cross section
mismeasurement causes migrations from one bin to another (x_{rec} differs from x_{true})

Example of a steeply falling distribution: P_T



For steep falling distributions migrations to high P_T larger than migrations to low P_T
Bins at high P_t are impure - contaminated by mismeasured low P_T events
- high $P_{T,rec}$ bins contain many events with low $P_{T,true}$

Purity

Acceptance is not the whole story! Imagine a bin in which all events migrated out and were replaced by an equal number which had migrated in from elsewhere in the kinematic plane. A_{cc} = 1 but our binning is clearly stupid !

$$\text{Purity} = \frac{\text{Nr. Events generated AND reconstructed in bin}}{\text{Nr. events reconstructed in bin}}$$

Of all the (rec) events in a bin, what fraction belong there?

Flat cross section with perfect efficiency in 1D binning with 1σ width gives purity=68% (Gaussian errors).

Rule of thumb: ask for Purity > $(0.6)^n$ for an n dimensional binning

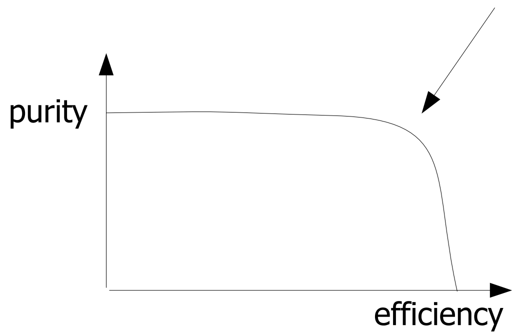
Purity is not used directly in evaluating the cross section. It is an example of a “control distribution” or “control plot”.

Need to find compromise between efficiency and purity
(and background contamination as well!)

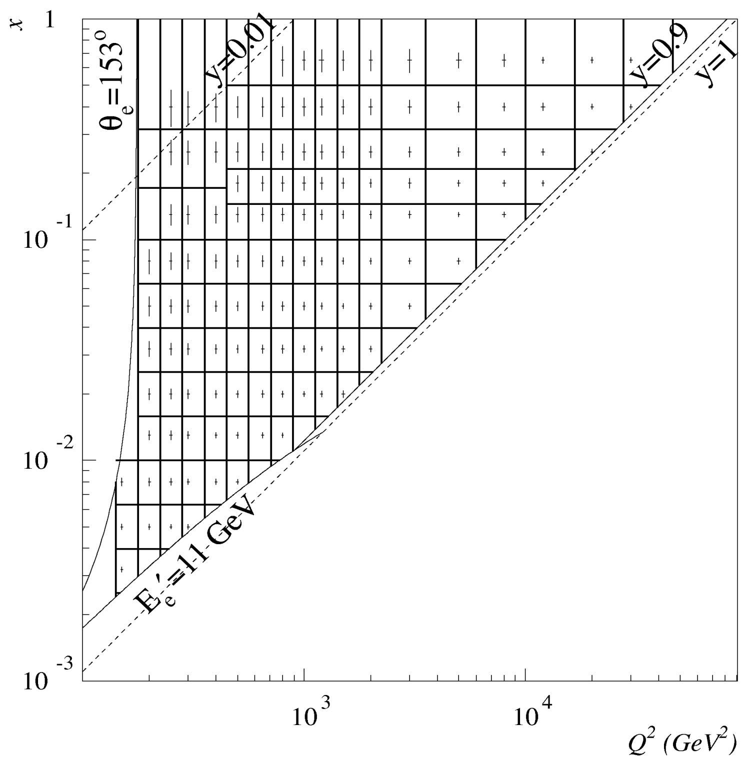
Easy to achieve high purity: make cuts very tight
loses signal due to resolution i.e. low efficiency

Easy to achieve high efficiency: make cuts very loose
get larger resolution migration i.e. purity falls
gains you background contamination

Find optimum working point maximising purity & efficiency



Example Binning



- Mean values reconstructed values centred in correct bins.
- Bin widths significantly bigger than s for most bins
- We can see where extra care is needed !

Bin Centre Corrections

We now have a cross section integrated over the phase space of our bin. To convert this to a differential measurement at (x_C, Q_C^2) we apply the correction:

$$\boxed{B_r = \frac{\dfrac{d^2\sigma(x_c, Q_c^2)}{dx\,dQ^2}}{\displaystyle\int_{bin} \frac{d^2\sigma}{dx\,dQ^2}\,dx\,dQ^2}}$$

which reduces to (bin volume)$^{-1}$ for a cross section which is uniform across the bin. This correction may be calculated analytically or with Monte Carlo. It depends on the shape of the cross section, so there is a systematic error associated with this.

Define Measured Cross-Section

- Specify your measurement in terms of configurations of observable particles
- A definition in terms of Lorentz Invariant quantities is preferable
- Specify any corrections you made to your data

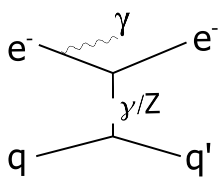
Why is this so important ?

Your measurement should not depend on the theory which you choose explain the data!

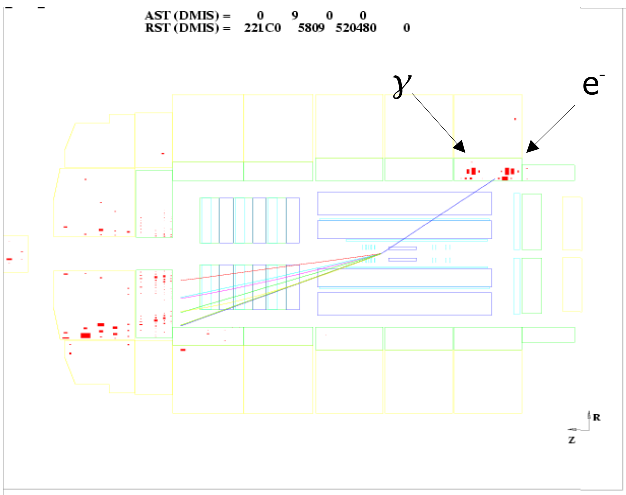
Can a theoretician who had an alternative to the Standard Model calculate your cross section? If not, something is wrong…

For example: Have you applied radiative corrections to your measurement i.e. is the running of α_{em} taken into account?

Remember some jet algorithms are "infra-red unsafe" i.e. cannot be defined on the hadron level such that theorists cannot compare models to your data!



• diagram modified by higher order QED corrections
• radiatied initial state photon = reduction in CMS energy
• photon emmitted at low angle and be undetected



radiated photon very close to electron - both energies should be summed in this case to get true energy of final state electron

• included in MC where possible - effects very large in some phase space areas!

• For some measurements QED corrections are "boring" - corrections well known - QED is well tested theory

• Corrections difficult to calculate depend on the exact analysis cuts etc.

• Some measurements should then be corrected to remove the effects of QED radiation.

• Experimentalists use MC to calc. rad-corrs for their cuts and detector. Then theorists fit the x-sections do not need to make the calculations for this well understood part of the cross-section.

Acceptance and Reweighting

For our acceptance corrections to work, our Monte Carlo must describe our data ! General procedure is:

1. Make measurement of cross section.
2. Compare measured cross section to that used in Monte Carlo.
3. Reweight or remake Monte Carlo with measured cross section as input
4. Go to 1 !

Note change in your measured cross section with each iteration

If analysis is ROBUST then

change in the measured cross section < change in Monte Carlo

convergence: depends crucially on sensible binning

Systematic Errors

Write down all the mistakes which can be made ! Make best guess of how much error is possible.

1. Central limit theorem helps to make most of our errors Gaussian.
2. Law of error propagation tells us how to combine both independent and correlated error sources together.

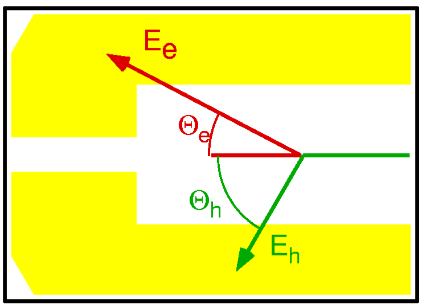
General Principle

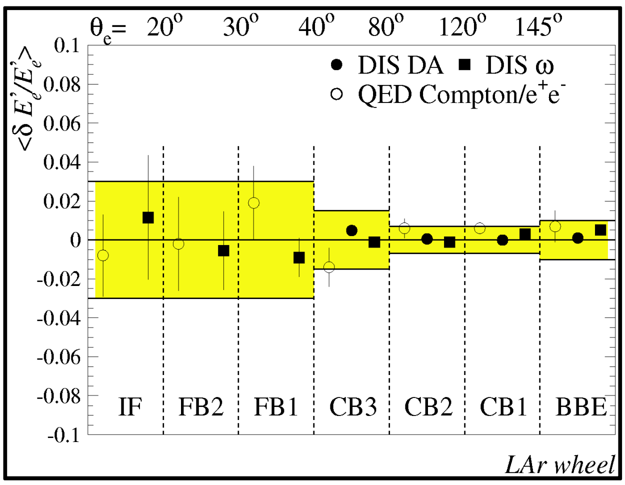
1. Determine 1σ variation on source of error (e.g. electron calibration).
2. Shift or reweight Monte Carlo by this amount.
3. Determine change in observed cross section in Monte Carlo.
4. Repeat for all independent sources of error.
5. Add resulting deviations in quadrature.

Example: Electromagnetic Calorimeter Calibration

- Double angle calibration of LAr electrons to 0.7%

4 measurements
2 degrees of freedom





We must use judgement to estimate 1σ ! Always try to check systematic effects in as many different ways as possible !

Other Sources of Systematic Error

Examples from My Analysis

1. Uncertainty in electron identification efficiency
2. Uncertainty in Trigger Efficiency
3. Uncertainty in theoretical model (e.g. radiative corrections).
4. Uncertainty in noise subtraction in calorimeter.
5. Uncertainty in background subtraction.
6. Uncertainty in luminosity measurement…

Correlated and Uncorrelated Systematic Errors

A systematic error is called "correlated" if it tends to shift some or all of the measured data points in a correlated way. E.g. the luminosity uncertainty shifts all data up or down together (100% correlated). In the case of correlated systematics, all the estimated shifts to the cross section need to be kept: they must be reapplied in any fit to the data to get the errors correct !

How to reduce systematic errors

Cross section ratios often allow large systematic uncertainties to cancel!
±7% luminosity uncertainty ought to apply to all cross sections at LHC
e.g. $\sigma(qq \to Z \to \mu\mu)$

But measuring ratio $R = \sigma(qq \to Z \to \mu\mu) / \sigma(qq \to W \to \mu\nu)$
will have smaller uncertainty - luminosity error cancels entirely!
other uncertainties may also cancel partially or completely

We do sacrifice information though:
Our measurement of R will be consistent with
$\sigma(qq \to Z \to \mu\mu) = 0.01$ pb AND
$\sigma(qq \to Z \to \mu\mu) = 1000$ pb

Ratios are good if you are mostly interested in the shape of a cross section not absolute value

Phillips' Law of Luck Cancellation

If you are lucky and get a lot of events, you are also unlucky because your cross section is wrong !

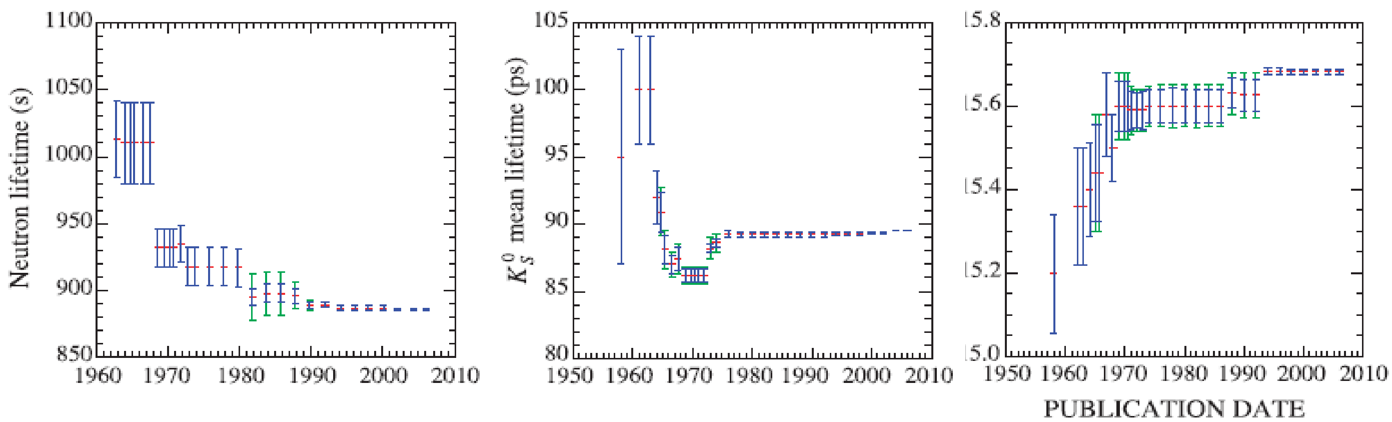
Particular care must be taken in low statistics analyses, especially in first measurements. Imagine you demand a minimum of 5 events to define a good measurement. If the true cross section predicts 2, you will only measure bins with upward fluctuations and thus get wrong results.

- Resist the temptation to fiddle too much with bin boundaries and cuts to "improve" your results (presupposes you know what they should be…)
- Adjustments of cuts and bin boundaries should always be aimed towards improving acceptance, efficiency or purity, and not getting "smoother" points or "smoother" errors

For the example above, choose to measure in the region where *theory* predicts 5 or more events. At most you have biased your phase space. Revise this if the data suggest that the theory is badly wrong.

History Plots of Measurements

- First measurements are often susceptible to subjective bias



Killer Questions for Physics Meetings

We can summarise some of what we have learned by listing questions which are always worth asking about your own or other people's analyses:

1. How have you defined the measured cross section ? If the answer relies on theoretical models, there may well be errors not properly treated.
2. What is the minimum purity in your measurement ? It's amazing how many people don't even know ! Be very suspicious if an evasive answer is given.
3. What is the largest source of systematic uncertainty ? Someone who has done a careful job will always know the answer, including:
   - what the next largest effect is,
   - what the largest effect is in different regions of phase space,
   - what they could do to reduce the error and why they didn't do it.
4. How did you check the size of your systematic errors ? For a good analysis, several different methods will have been used to check.
5. What is the physics message of your analysis ? Ask this question at the end for some fun discussion ! (You will know what I mean in a year…)

Final Words

- Analysis in HEP is complicated ! Common sense is very important.
- HEP Collaborations are chaotic: much useful knowledge is "folklore". Getting to know people and their work is vital for you:
  - Attend working group and collaboration meetings
  - Try to understand other analyses !
- Make a useful contribution ! A Ph.D. which does not contribute to a journal publication is a waste of time scientifically !
  - Do not reinvent the wheel: use all that exists and build on it to invent something genuinely new !
  - Ask yourself where your analysis is heading: if it isn't contributing in some way to a publication, ask yourself why not. Ask your supervisor or RAs as well.