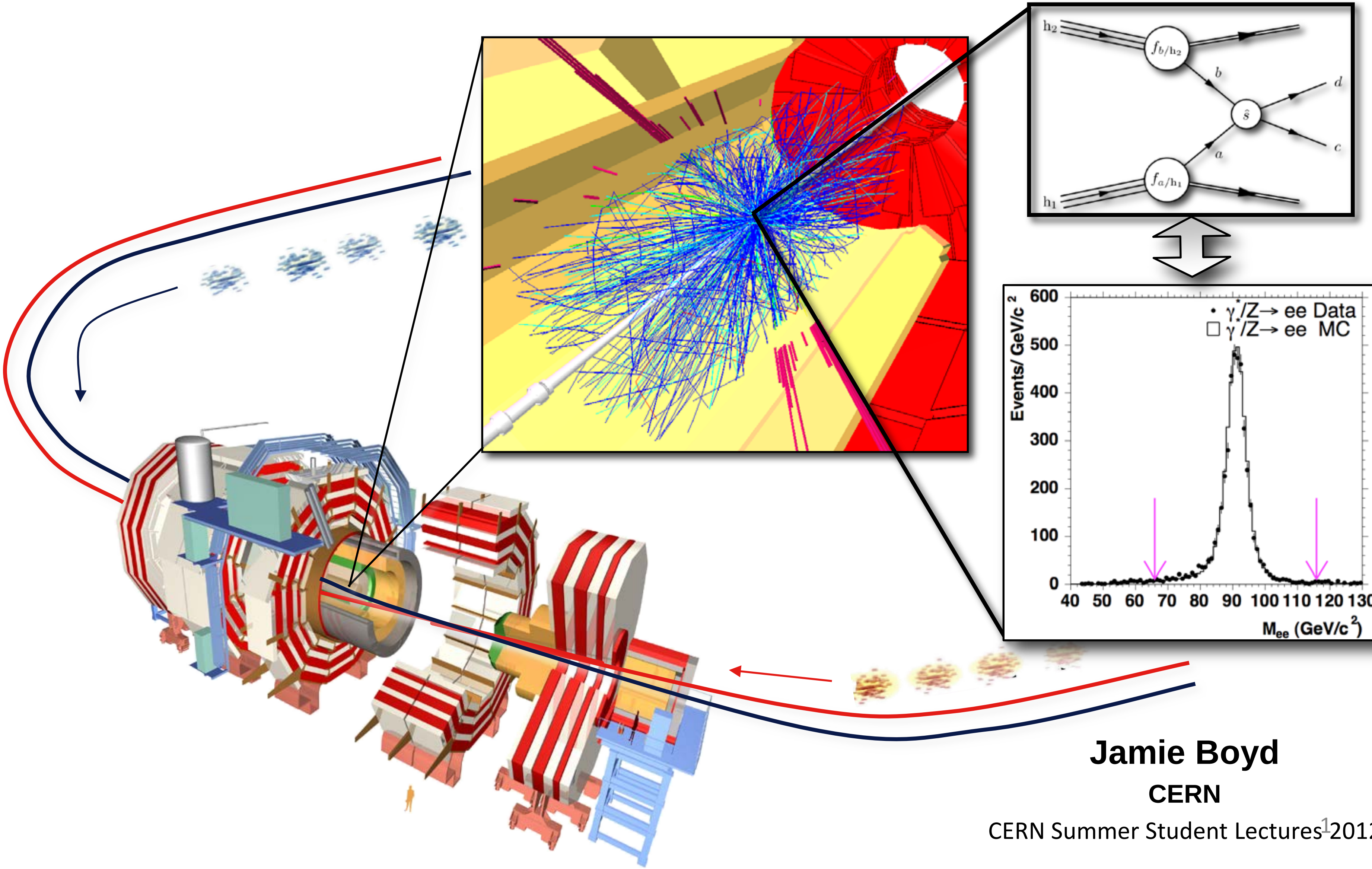
From Raw Data to Physics Results



Jamie Boyd

CERN

CERN Summer Student Lectures 2012

Outline

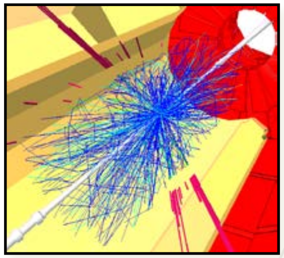
- Summary
  - Brief overview of the full lecture course
- A simple example
  - Measuring the Z^0 cross-section
- Reconstruction & Simulation
  - Track reconstruction
  - Calorimeter recostruction
  - Physics object reconstruction
  - Simulation
- Physics Analysis
  - Data Quality
  - Z'->ll
  - H->γγ
  - H->ZZ->4l
- Computing infrastructure
- The End!

} Todays Lecture

Disclaimer : Much of the content based on previous years lectures
Thanks to G. Dissertori

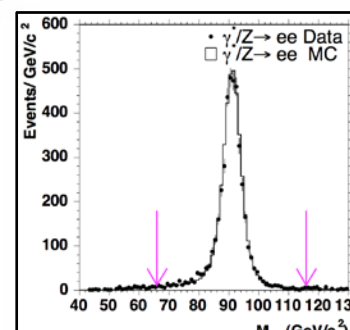
Section 1: Summary

Data Analysis Chain

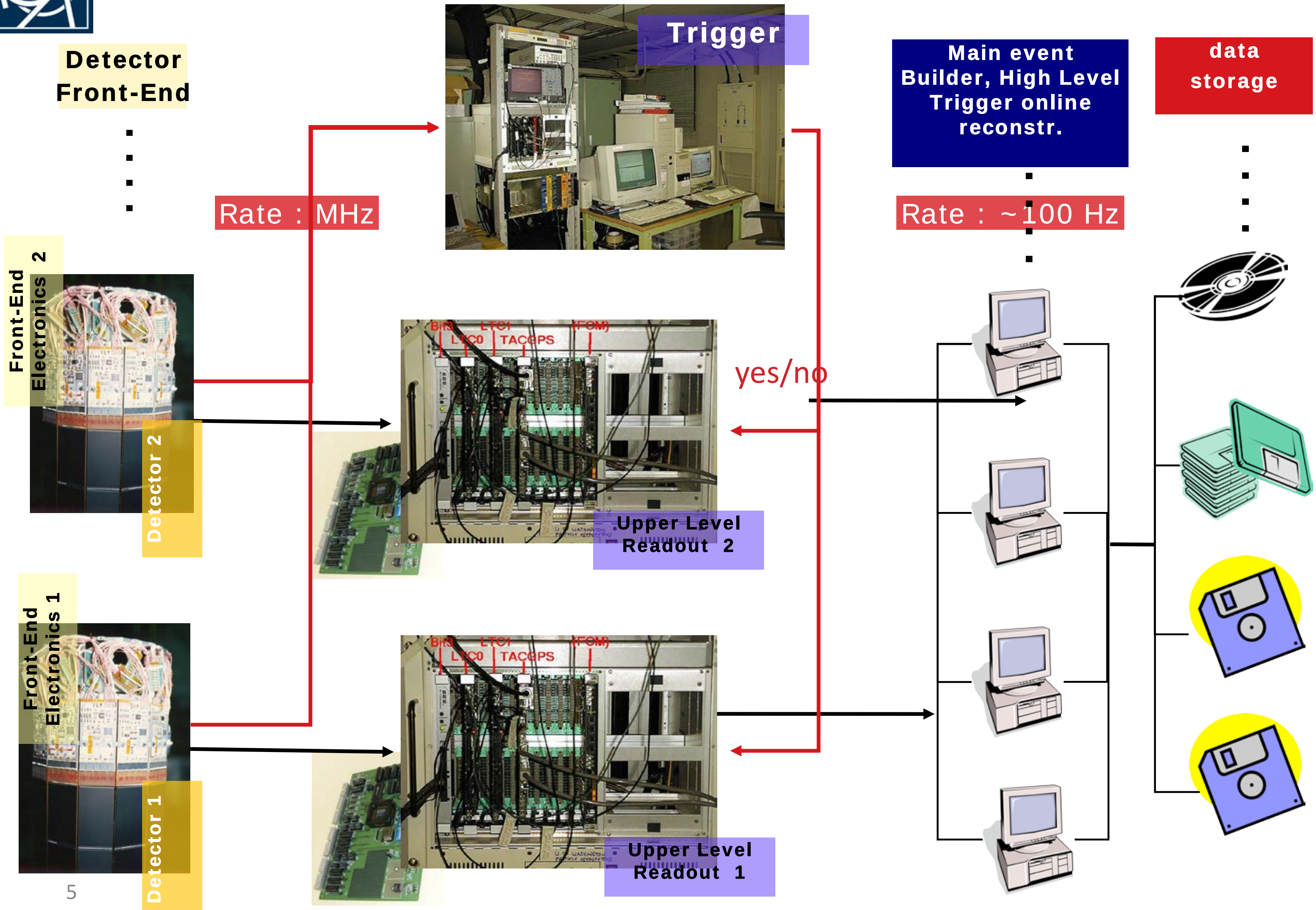


- Have to collect data from many channels on many sub-detectors (millions)
- Decide to read out everything or throw event away (Trigger)
- Build the event (put info together)
- Store the data
- Analyze them
  - reconstruction, user analysis algorithms, data volume reduction
- do the same with a simulation
  - correct data for detector effects
- Compare data and theory

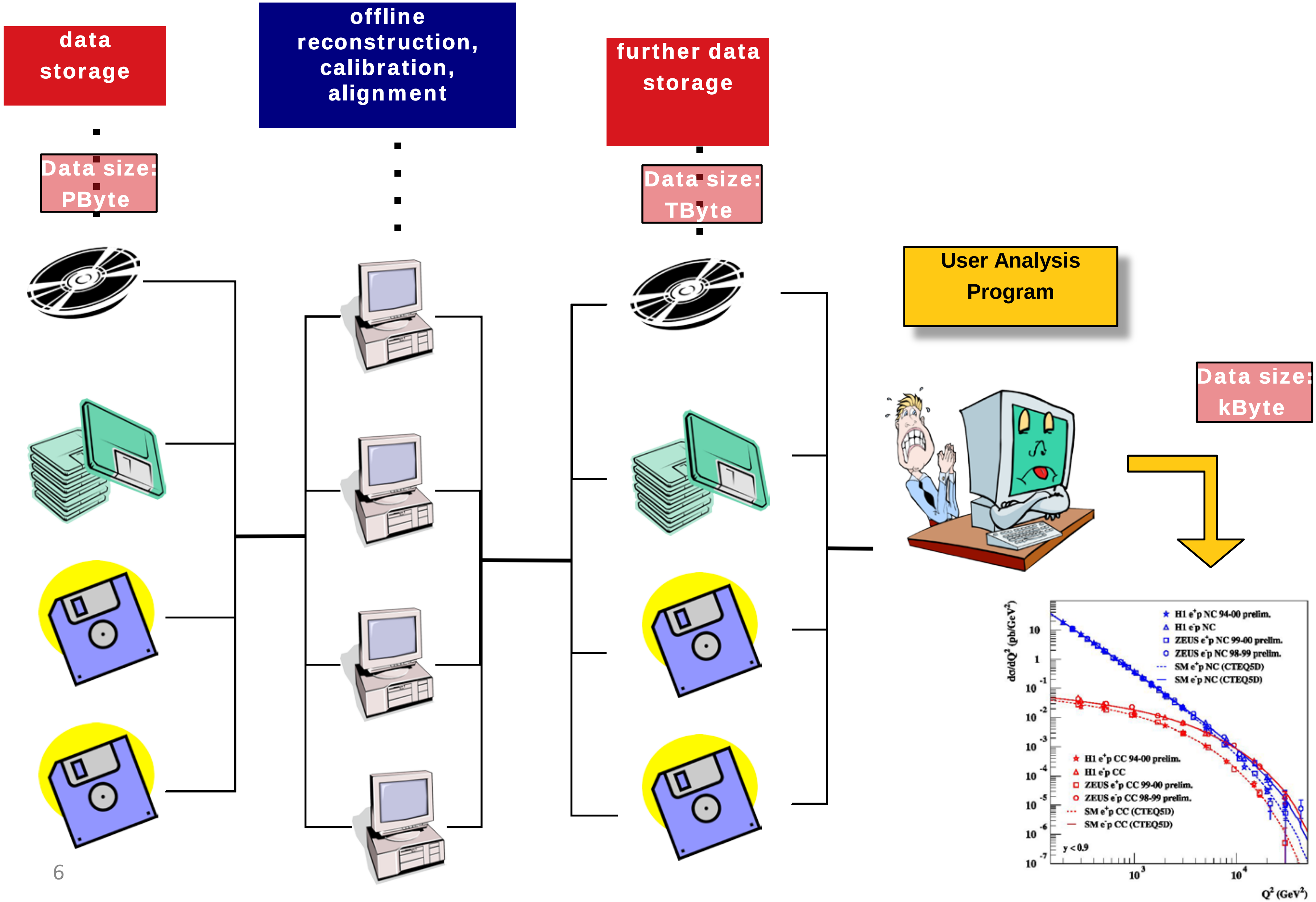
This lecture course!!



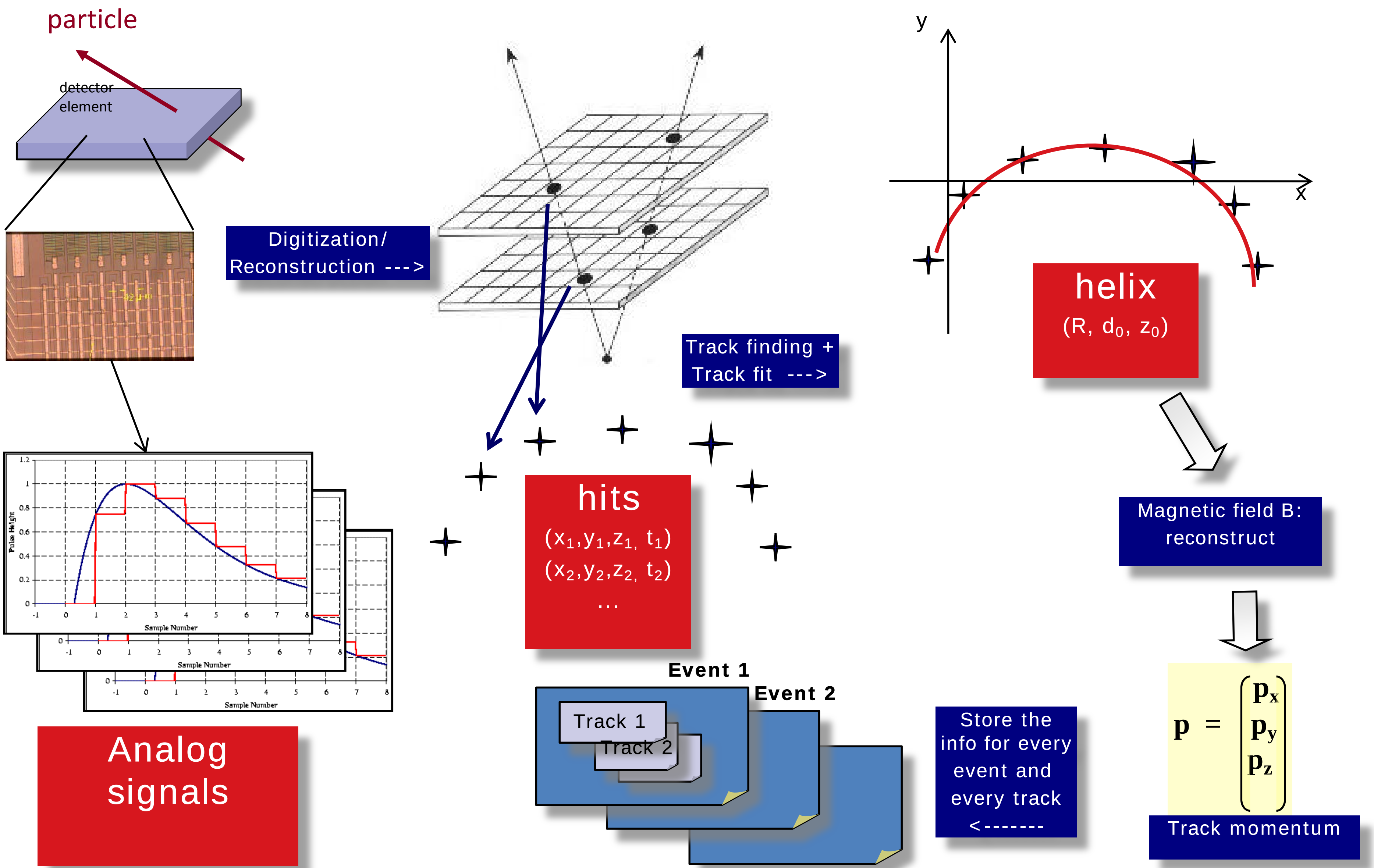
DAQ chain (see lectures by W. Vandelli & B. Dahmes)



Offline Analysis Chain



Data reduction/abstraction



High Level Data Storage

- Data are stored sequentially in files...

Event 1

Nch (charged tracks) :

2

Pcha

(Momentum of each track):

{{"-7.65698","42.9725","14.3404"},
{" 7.54101","-42.1729","-14.0108"}}

px py pz

Qcha

(Charge of each track):

{-1,1}

Event 2

Nch (charged tracks) :

3

Pcha

(Momentum of each track):

{{"-12.9305","12.2713","40.5615"},
{" 12.2469","-11.606","-38.7182"},
{"0.143435","-0.143435","-0.497444"}}

px py pz

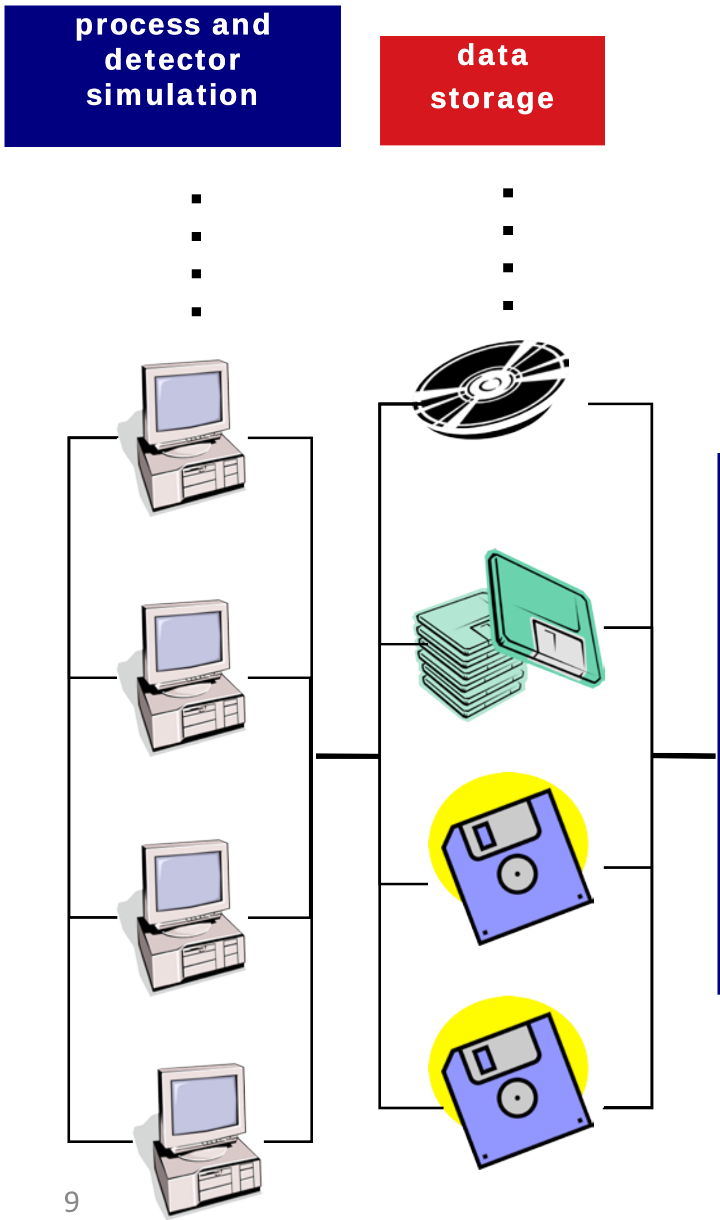
Qcha

(Charge of each track):

{-1,1,-1}

File A

Simulation



Exactly the same steps as for the data

Simulation of many (millions) of events

- simulate physics process
e.g. p p –> Z
or p p –> H

- plus the detector response to the produced particles

- understand detector response and analysis parameters (lost particles, resolution, efficiencies, backgrounds)

- and compare to real data

- Note : simulations present from beginning to end of experiment, needed to make design choices

Our Task

Reality

We use experiments to inquire about what "reality" (nature) does

We intend to fill this gap

Theory

$$S = \mathrm{i} \int \mathrm{d}^4x \, \mathcal{L}(x)$$

The goal is to understand in the most general; that's usually also the simplest.

- A. Eddington

Theory...

CERN

$$\mathcal{L} = -\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}$$

$W^{\pm}, Z, \gamma$ kinetic energies and self-interactions

$$+\overline{L}\gamma^{\mu}\left(i\partial_{\mu} - g\frac{1}{2}\tau\cdot\mathbf{W}_{\mu} - g'\frac{Y}{2}B_{\mu}\right)L$$
$$+\overline{R}\gamma^{\mu}\left(i\partial_{\mu} - g'\frac{Y}{2}B_{\mu}\right)R$$

lepton and quark kinetic energies and their interactions with $W^{\pm}, Z, \gamma$

$$+\left|\left(i\partial_{\mu} - g\frac{1}{2}\tau\cdot\mathbf{W}_{\mu} - g'\frac{Y}{2}B_{\mu}\right)\phi\right|^{2} - V(\phi)$$

$W^{\pm}, Z, \gamma$ and Higgs masses and couplings

$$-(G_1\overline{L}\phi R + G_2\overline{L}\phi_c R + h.c.)$$

lepton and quark masses and coupling to Higgs

L ... left–handed fermion (l or q) doublet
R ... right–handed fermion singlet

$\mathcal{L}$ from QCD:

$$\mathcal{L} = \bar{q}\underbrace{(i\gamma^{\mu}\partial_{\mu} - m)q}_{E_{kin}(q)} - g\underbrace{(\bar{q}\gamma^{\mu}T_a q)G^a_{\mu}}_{\text{Interaction } q,\, g} - \underbrace{\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a}_{E_{kin}(g)}$$

$E_{kin}(q)$ — Interaction q, g — $E_{kin}(g)$ includes self-interaction between gluons

eg.
the Standard Model

has parameters

coupling constants

masses

predicts:
cross sections,
branching ratios, lifetimes, ...

Experiment...

```
0x01e84c10:   0x01e8 0x8848 0x01e8 0x83d8 0x6c73 0x6f72 0x7400 0x0000
0x01e84c20:   0x0000 0x0019 0x0000 0x0000 0x01e8 0x4d08 0x01e8 0x5b7c
0x01e84c30:   0x01e8 0x87e8 0x01e8 0x8458 0x7061 0x636b 0x6167 0x6500
0x01e84c40:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c50:   0x01e8 0x8788 0x01e8 0x8498 0x7072 0x6f63 0x0000 0x0000
0x01e84c60:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c70:   0x01e8 0x8824 0x01e8 0x84d8 0x7265 0x6765 0x7870 0x0000
0x01e84c80:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c90:   0x01e8 0x8838 0x01e8 0x8518 0x7265 0x6773 0x7562 0x0000
0x01e84ca0:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cb0:   0x01e8 0x8818 0x01e8 0x8558 0x7265 0x6e61 0x6d65 0x0000
0x01e84cc0:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cd0:   0x01e8 0x8798 0x01e8 0x8598 0x7265 0x7475 0x726e 0x0000
0x01e84ce0:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cf0:   0x01e8 0x87ec 0x01e8 0x85d8 0x7363 0x616e 0x0000 0x0000
0x01e84d00:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d10:   0x01e8 0x87e8 0x01e8 0x8618 0x7365 0x7400 0x0000 0x0000
0x01e84d20:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d30:   0x01e8 0x87a8 0x01e8 0x8658 0x7370 0x6c69 0x7400 0x0000
0x01e84d40:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d50:   0x01e8 0x8854 0x01e8 0x8698 0x7374 0x7269 0x6e67 0x0000
0x01e84d60:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d70:   0x01e8 0x875c 0x01e8 0x86d8 0x7375 0x6273 0x7400 0x0000
0x01e84d80:   0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d90:   0x01e8 0x87c0 0x01e8 0x8718 0x7377 0x6974 0x6368 0x0000
```

eg.

$1/30^{th}$ of an event in the BaBar detector

- get about 100 evts/sec

"Address" :

- which detector element took the reading

"Value(s)" :

- what the electronics wrote out

Making the connection

Reality

Raw Data — The imperfect measurement of a (set of) interactions in the detector

↓

Reconstructed Events — A unique happening:
eg. Run 23458, event 1345
which contains a Z → $\mu^+\mu^-$ decay

↓

Analysis : We "confront theory with experiment" by comparing the measured quantity (observable) with the prediction.

Observables — cross sections (probabilities for interactions), branching ratios (BR), ratios of BRs, specific lifetimes, ...

↑

Theory

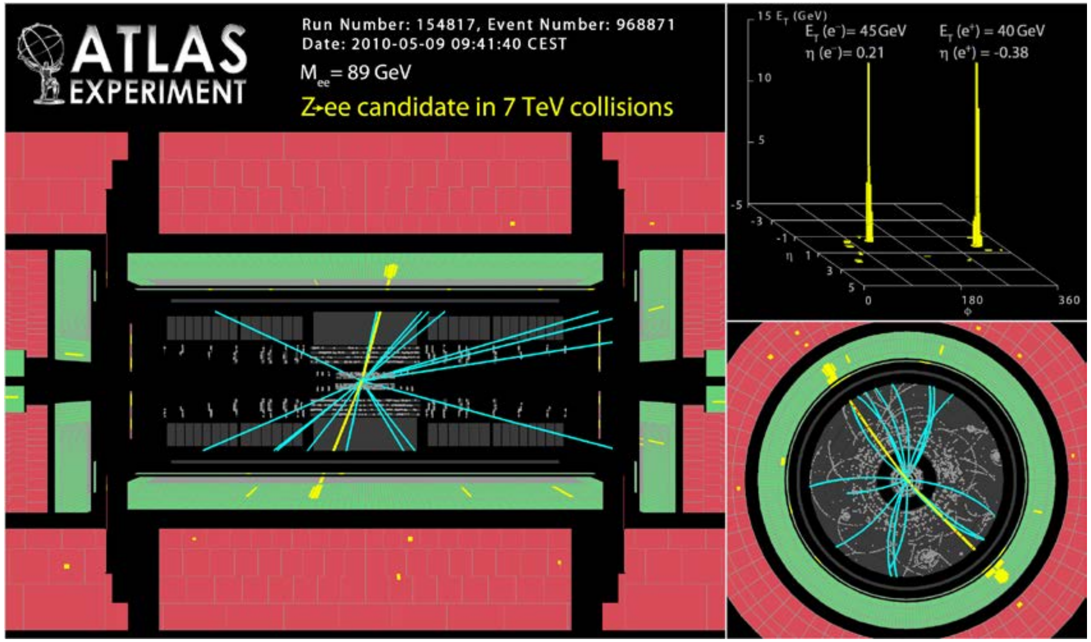
$$S = \mathrm{i}\int \mathrm{d}^4x\,\mathcal{L}(x)$$

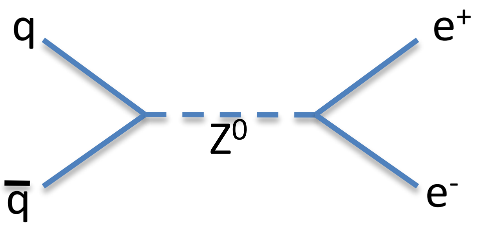
A small number of general equations, with some parameters (poorly or not known at all)

Section 2: A simple example

Measuring Z^0 cross-section at LHC

- Z^0 boson decays to lepton or quark pairs
  - We can reconstruct it in the e^+e^- or $\mu^+\mu^-$ decay modes
- Discovery and study of the Z^0 boson was a critical part of understanding the electroweak force
- Measuring the Z^0 cross-section at the LHC important test of theory
  - Does the measurement agree with the theoretical prediction at LHC collision energy?
- Now we use the Z^0 as a tool for studying electron and muon reconstruction and deriving calibrations (have now recorded millions of Z decays)



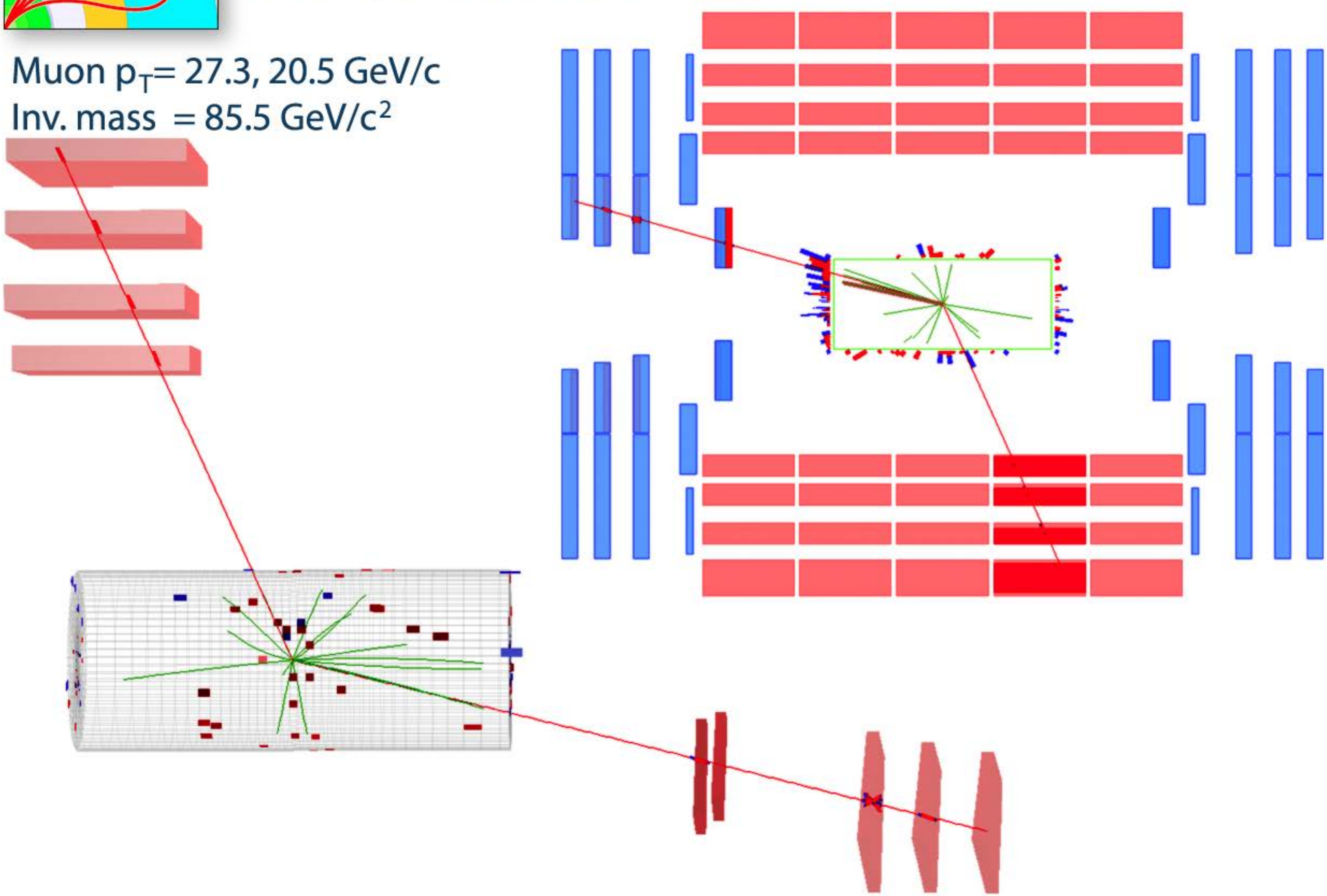


Z^0 cross-section is related to the probability that we will produce a Z^0 at the LHC

CMS Experiment at LHC, CERN
Run 136087 Event 39967482
Lumi section: 314
Mon May 24 2010, 15:31:58 CEST

Z->μμ event in CMS

Muon p_T= 27.3, 20.5 GeV/c
Inv. mass = 85.5 GeV/c^2



Reconstructing Z^0's

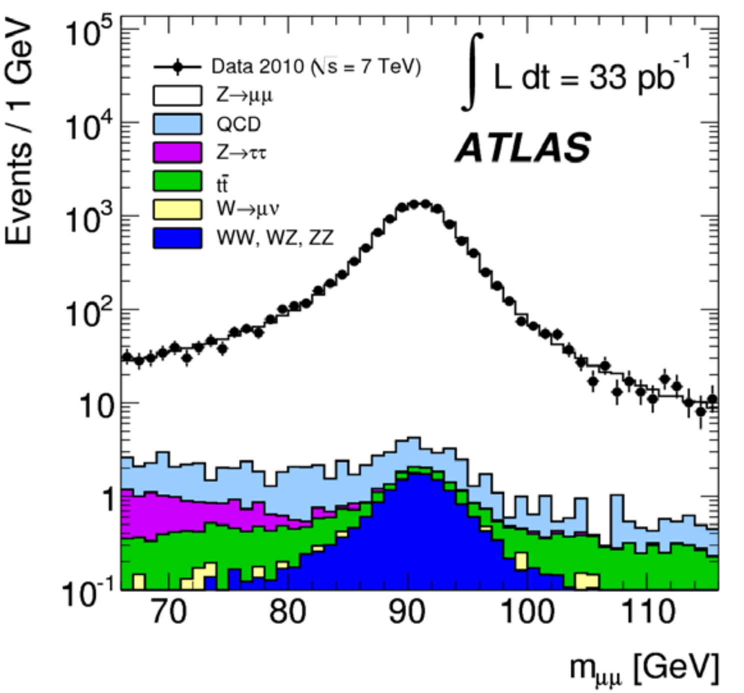
How do we know if it's a Z^0:

Identify Z decays using the invariant mass of the 2 leptons
$M^2 = (L_1+L_2)^2$ where $L_i = (E_i,\underline{\mathbf{p}}_i)$ = 4-vector for lepton i

Under assumption that lepton is massless compared to mass of Z^0
=> $M^2 = 2 E_1 E_2 (1-\cos\vartheta_{12})$ where ϑ_{12}= angle between the leptons

So need to reconstruct the electron and muon energy and direction.
Then can calculate the mass.

Select Z^0 events with 'analysis cuts':
-Events with 2 high momentum electrons or muons
-Require the electrons or muons are of opposite charge
- With di-lepton mass close to the Z^0 mass (e.g. $70<m_{l+l-}<110$ GeV)

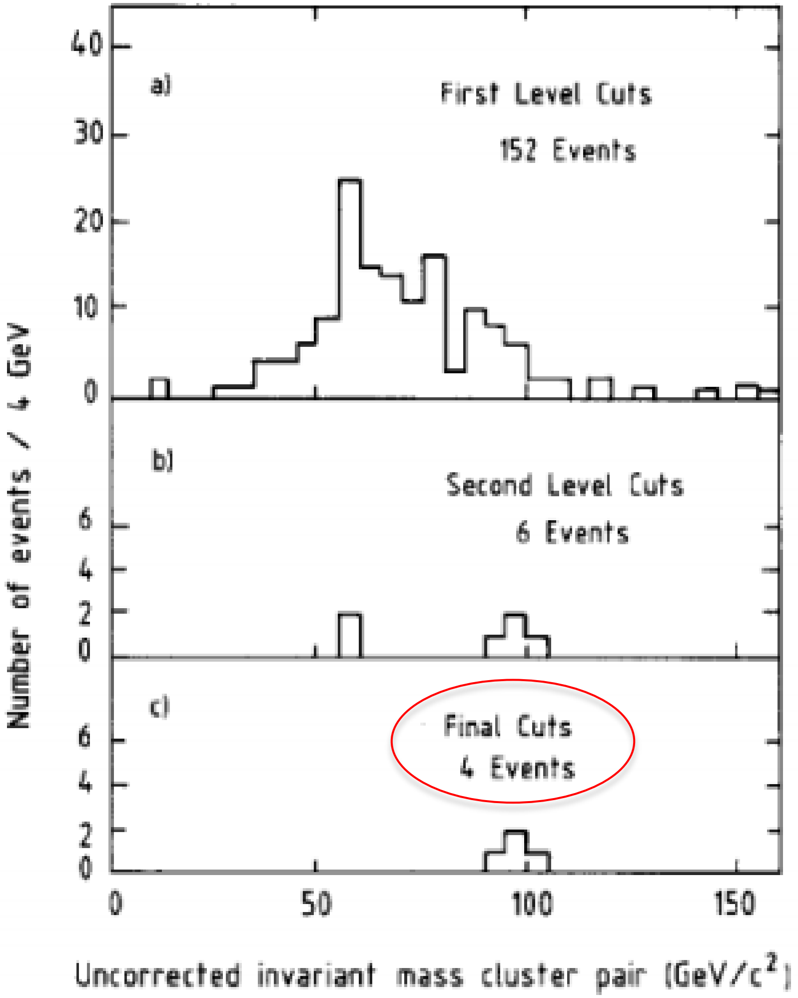


Very little background in the Z^0 mass region

UA1: observation of $Z \rightarrow e^+ e^-$

(May 1983)

The Nobel Prize in Physics 1984
Carlo Rubbia, Simon van der Meer



Two energy clusters (p_T > 25 GeV) in electromagnetic calorimeters; energy leakage in hadronic calorimeters consistent with electrons

Isolated track with p_T > 7 GeV pointing to at least one cluster

Isolated track with p_T > 7 GeV pointing to <u>both</u> clusters

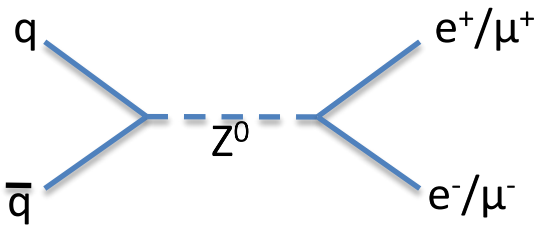
Slide taken from: "The achievements of the CERN proton – antiproton collider", Luigi DiLella, 2004

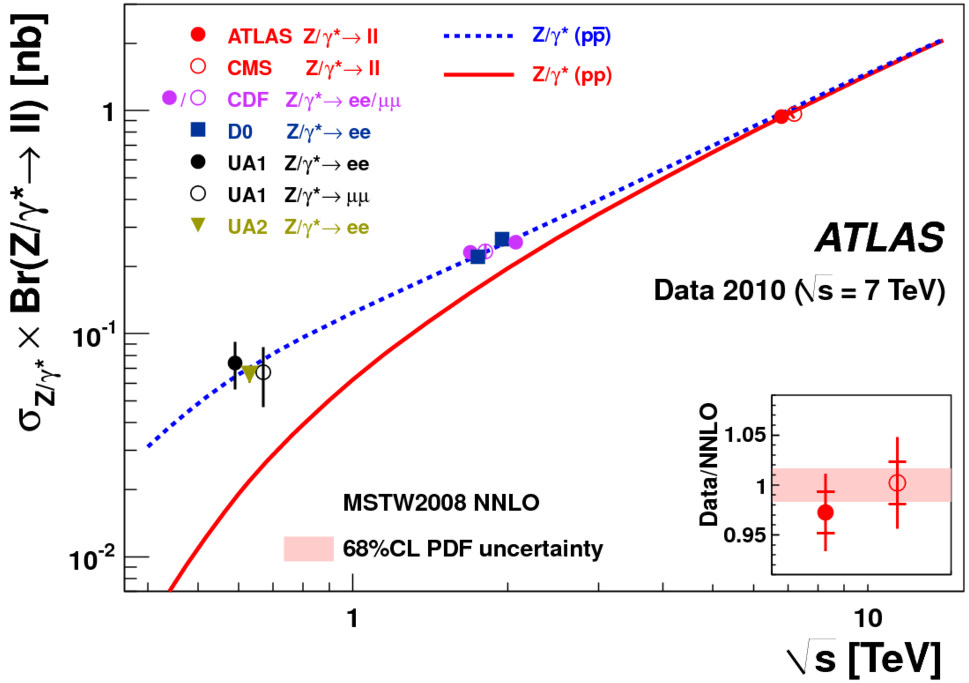
Measuring the Z^0 cross-section

Theoretically:

Cross-section calculated for:

- Specific production mechanism (pp, p$\bar{p}$, e^+e^-)
- Centre-of-Mass of the collisions (7TeV at LHC)





Experimentally:

$\sigma(pp \to Z) = (N_{OBS} - N_{BKG}) / L\, \varepsilon$

Where:

N_{OBS} = Number of observed events passing the selection

N_{BKG} = Estimate of the number of background events

L = Luminosity of the data samples (amount of data)

ε = Efficiency of the selection on Z^0 events

(how often would we select a true Z^0 event with our selection?)

Can use simulated data to evaluate ε and N_{BKG}

Measuring the Z^0 cross-section

$\sigma(pp \rightarrow Z) = (N_{OBS} - N_{BKG}) / L\, \varepsilon$

Looks like simple counting experiment.

But need to also calculate **uncertainty** on the cross-section – measurement without an uncertainty is useless.

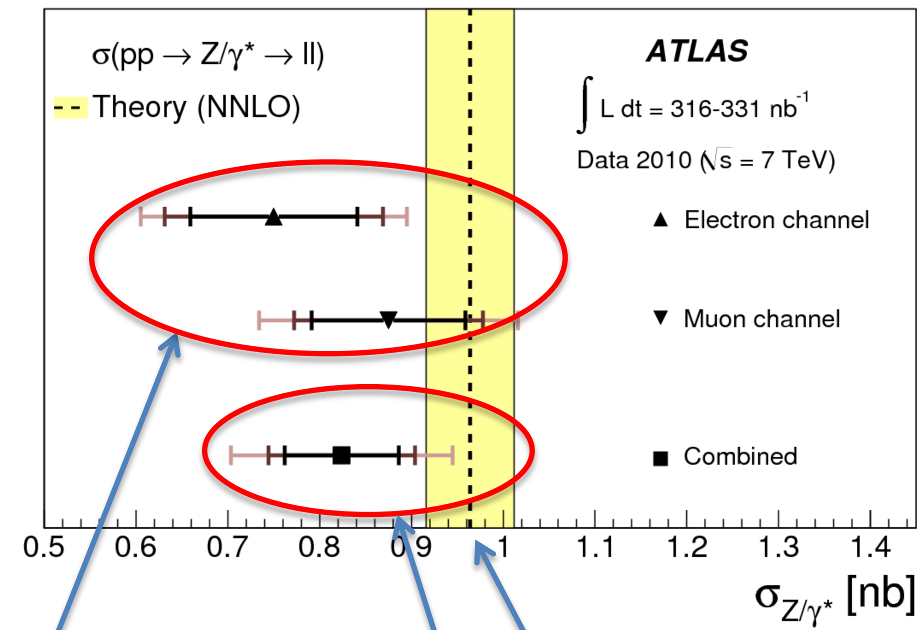
Two components to the uncertainty:

Statistical: $\sim\sqrt{N_{OBS}}$

Systematic:

- How well do we know the background?
- How well do we know the efficiency?
- How well do we know the luminosity?

Most of the work in the physics analysis is trying to understand the systematic uncertainties related to the above questions.



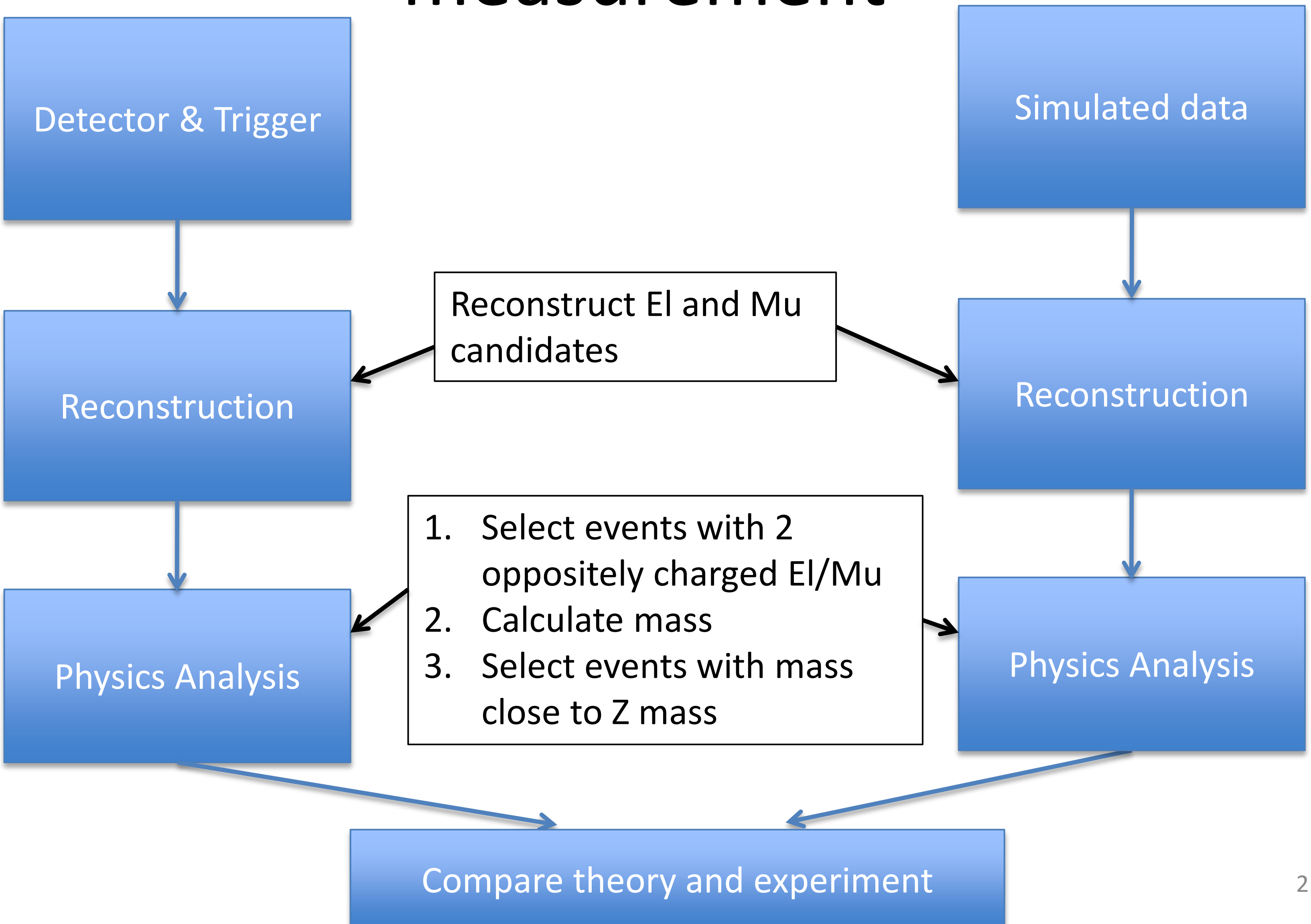
Electron and Muon channel agree within uncertainties

Theory prediction

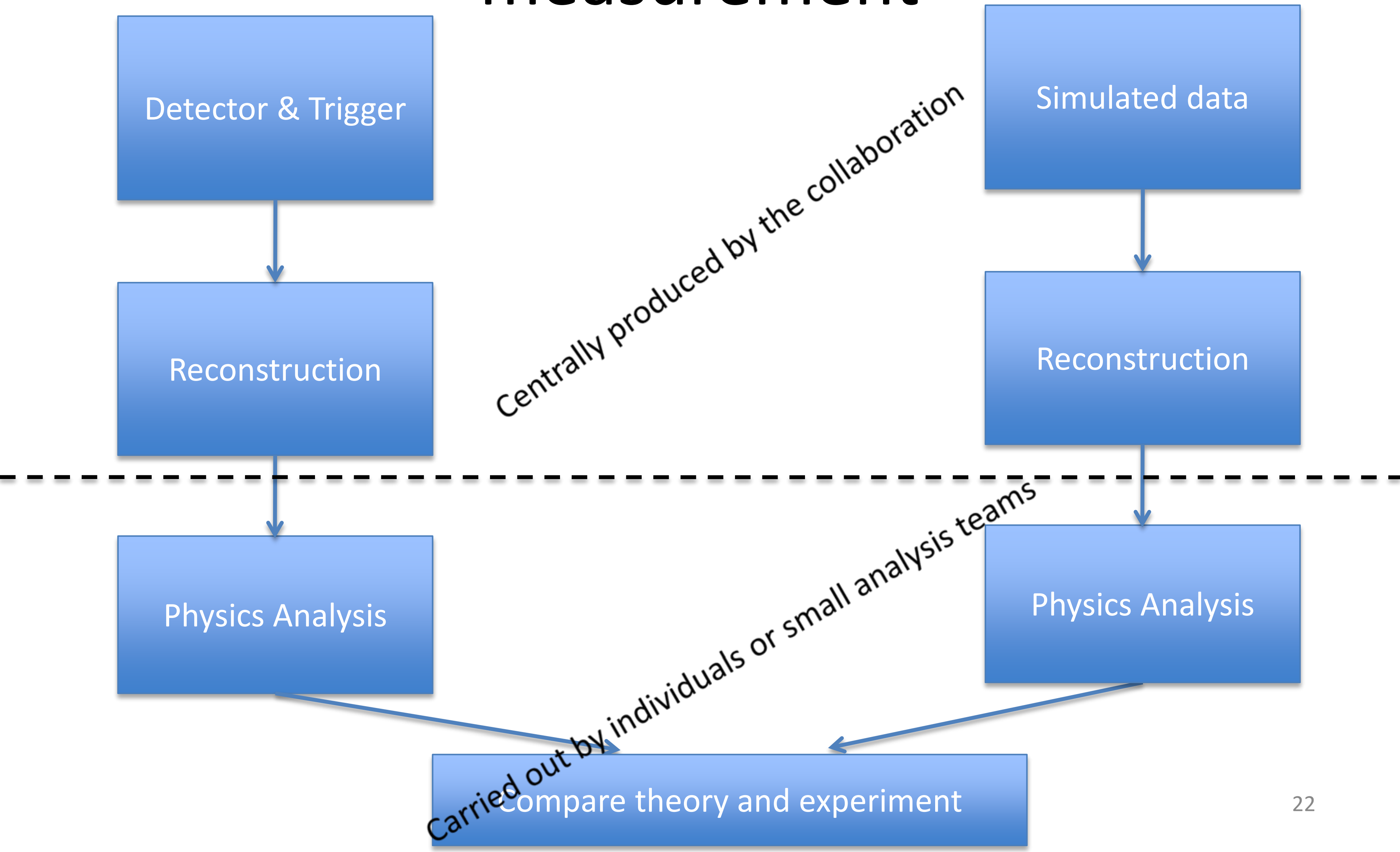
Measurement consistent with prediction within uncertainties

Measurements of the Z cross-section were one of the first physics measurements from ATLAS and CMS.

Analysis flow in Z cross-section measurement



Analysis flow in Z cross-section measurement



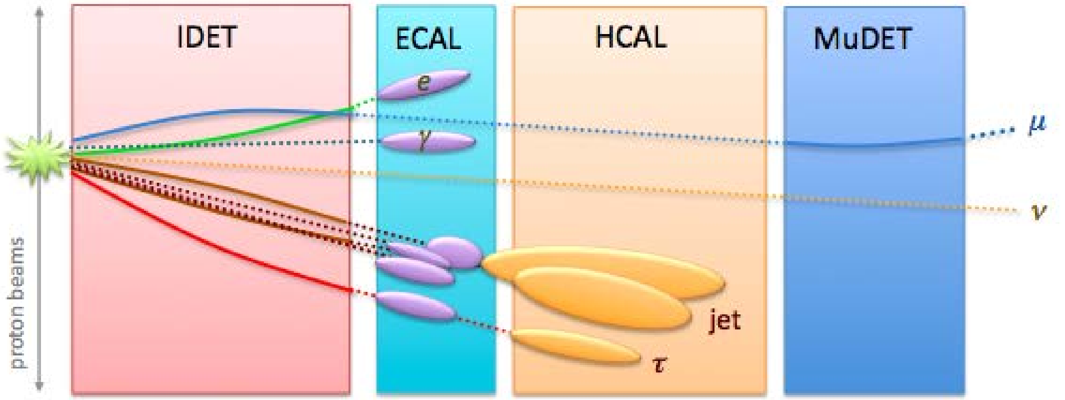
Summary

- Path from Raw data to physics results contains many steps
    - Online path (Trigger and DAQ)
    - Offline path
        - Reconstruction
        - Physics Analysis
    - Use simulation in order to compare data with theoretical predictions
- Above points illustrated with the example of the Z^0 cross-section measurement at the LHC
- More details tomorrow

Section 3: Reconstruction

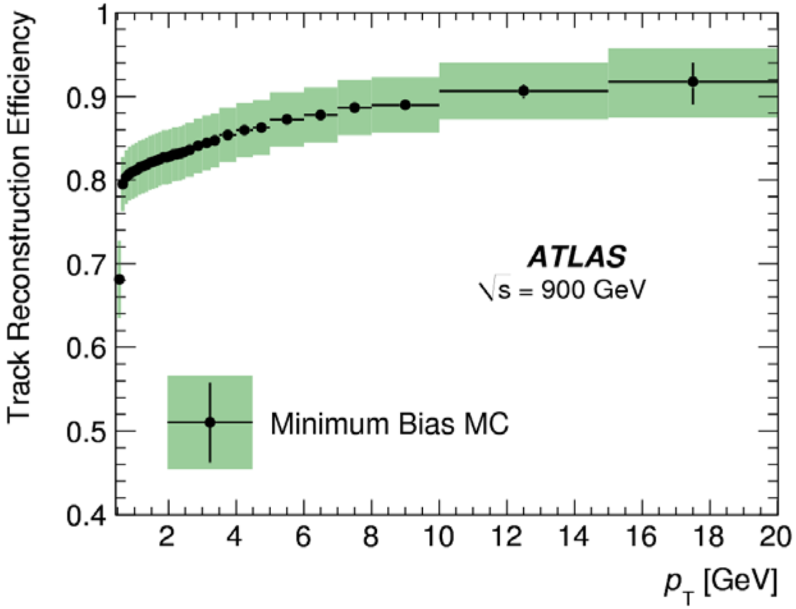
Reconstruction

- Detector reconstruction
  - Tracking
    - finding path of charged particles through the detector
  - Calorimeter reconstruction
    - finding energy deposits in calorimeters from charged and neutral particles
- Combined reconstruction
  - Electron/Photon identification
  - Muon identification
  - Jet finding
- Calibrations and alignments applied at nearly every step



Important figures of merit for reconstructed objects

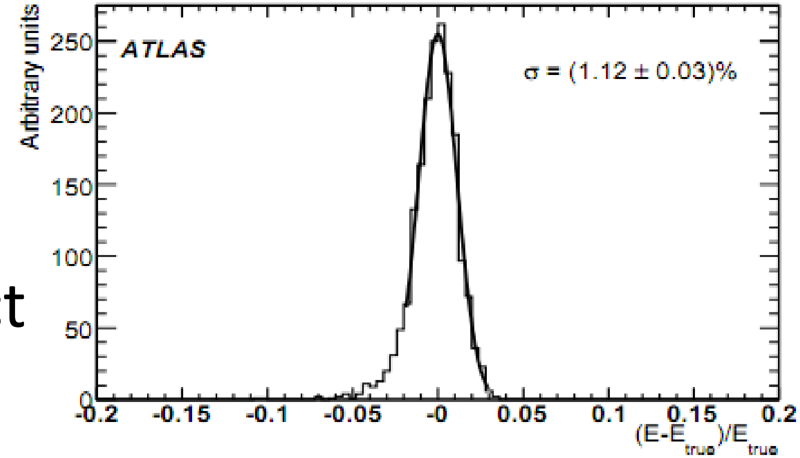
- **Efficiency**
  - how often do we reconstruct the object – e.g. tracking efficiency



Efficiency = (Number of Reconstructed Tracks) / (Number of True Tracks)

Important figures of merit for reconstructed objects

- **Efficiency**
  - how often do we reconstruct the object – e.g. tracking efficiency
- **Resolution**
  - how accurately do we reconstruct it – e.g. energy resolution

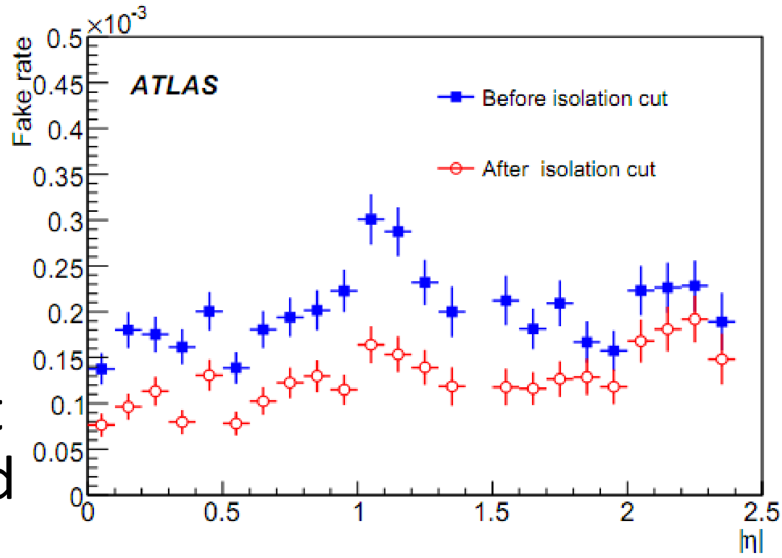


Electron energy resolution from simulation

Energy resolution = (Measured_Energy – True_Energy)/ True_Energy

Important figures of merit for reconstructed objects

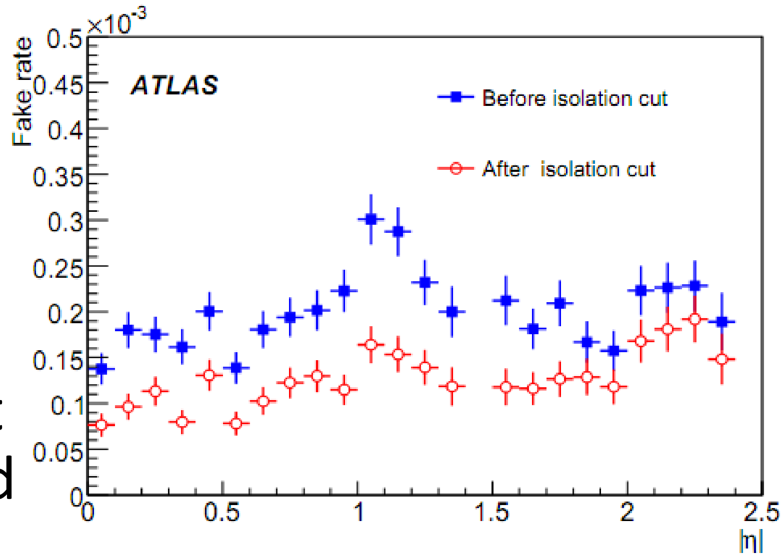
- **Efficiency**
  - how often do we reconstruct the object – e.g. tracking efficiency
- **Resolution**
  - how accurately do we reconstruct a quantity – e.g. energy resolution
- **Fake rate**
  - how often we reconstruct a different object as the object we are interested in – e.g. a jet faking a electron



Fake rate = (Number of jets reconstructed as an electron) / (Number of jets)

Important figures of merit for reconstructed objects

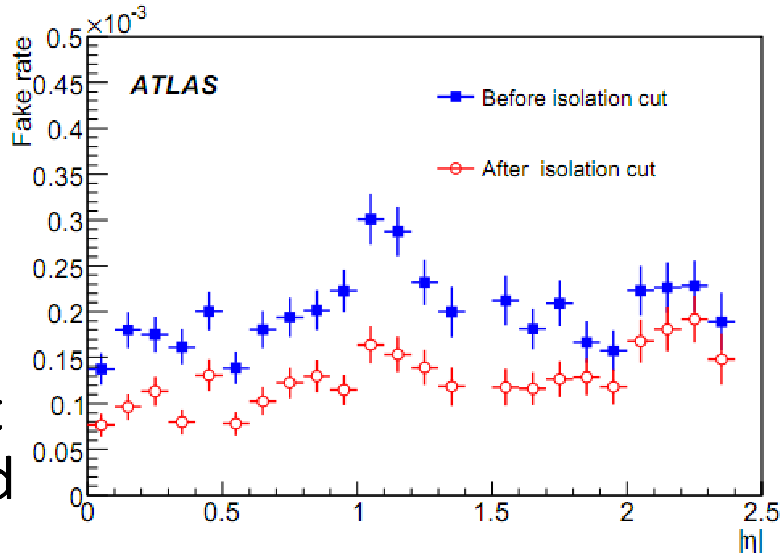
- **Efficiency**
  - how often do we reconstruct the object – e.g. tracking efficiency
- **Resolution**
  - how accurately do we reconstruct a quantity – e.g. energy resolution
- **Fake rate**
  - how often we reconstruct a different object as the object we are interested in – e.g. a jet faking a electron



These quantities depend on the detector, but also on the reconstruction and calibrations and alignment!

Important figures of merit for reconstructed objects

- **Efficiency**
  - how often do we reconstruct the object – e.g. tracking efficiency
- **Resolution**
  - how accurately do we reconstruct a quantity – e.g. energy resolution
- **Fake rate**
  - how often we reconstruct a different object as the object we are interested in – e.g. a jet faking a electron



For physics analysis it is important

i) to have high efficiency, good resolution, and low fake rates

ii) to be able to measure the efficiencies, resolutions and fake rates and their uncertainties (not easy)

Reconstruction Goals

- High efficiency
- Good resolution
- Low fake rate
- Robust against detector problems
  - Noise
  - Dead regions of the detector
- Be able to run within the computing resources limitations
  - CPU time per event
  - Memory use