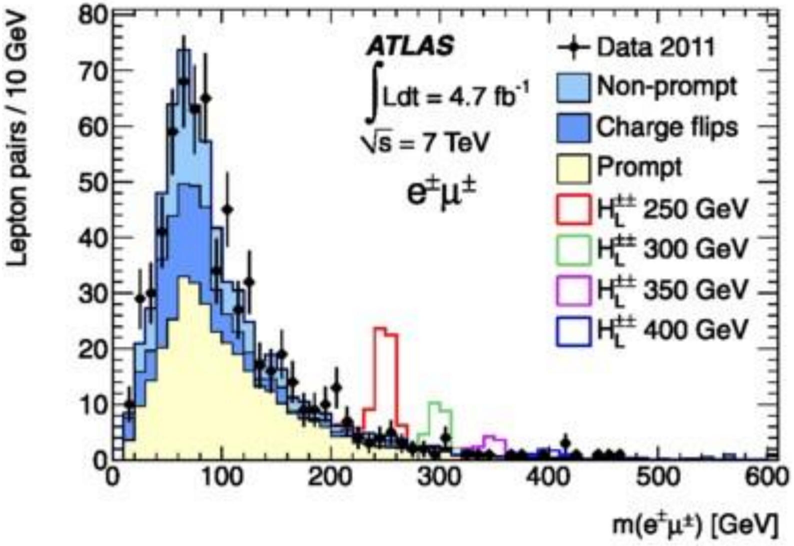
# FYST17 Lecture 7
# MC and Simulation

Thanks to M. Asai, T. Sjöstrand, J. Morris



**Suggested reading:** this is not well described in the book. Chap 9.4.3 in the statistics book has a few details. Better to follow a class with the world-famous Lund phenomenologists

# Today’s topics

- Simulation, Monte Carlo (MC) and why we use it
- MC generators
  - Examples
  - Different specialities
- Detector simulation
  - GEANT
- Performance, some examples

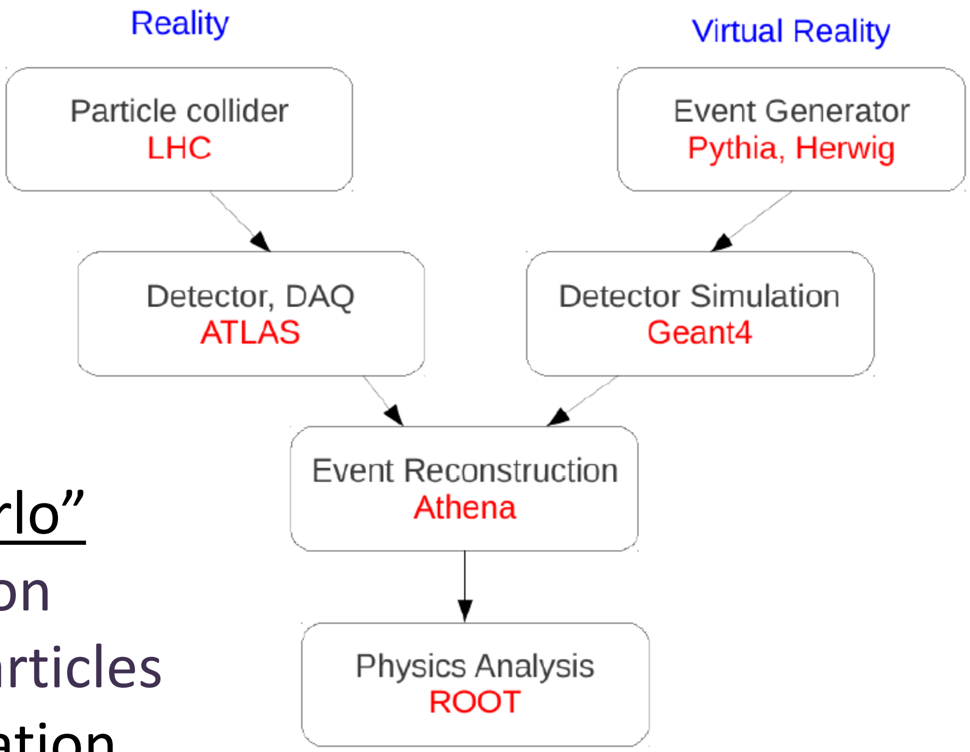
# Why simulation?

We want to be able to compare data to expectations

*”Virtual experiment”*

”Simulation” typically consists of two steps:

- <u>Event generation ”Monte Carlo”</u>
  - Calculations, hadronization
  - 4-vectors of final state particles
- <u>Detector simulation + digitization</u>
  - The particles’ paths through the detector material
  - Detector and electronic response



Reality

Virtual Reality

Particle collider
LHC

Event Generator
Pythia, Herwig

Detector, DAQ
ATLAS

Detector Simulation
Geant4

Event Reconstruction
Athena

Physics Analysis
ROOT

# Why simulation?

## 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

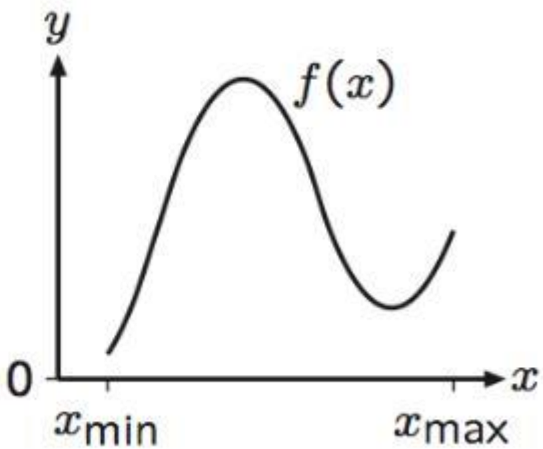
- Predict cross sections and topologies of various processes
  - ⇒ Feasibility study - Can we find the theoretical particle X?
- Simulate background processes to the signal of interest
  - ⇒ Can devise analysis strategies
- Study detector response
  - ⇒ Optimise trigger & detector selection cuts
- Study detector imperfections
  - ⇒ Can evaluate acceptance corrections

- Remove the effect of the apparatus from the measurement
  - ⇒ *Unfold* the data. Correcting the data for detector effects

Impossible to do analytically!

# The Monte Carlo method

”Monte Carlo” refers to any numerical method that uses random numbers in order to simulate probabilistic processes



Select x at random* according to f(x)

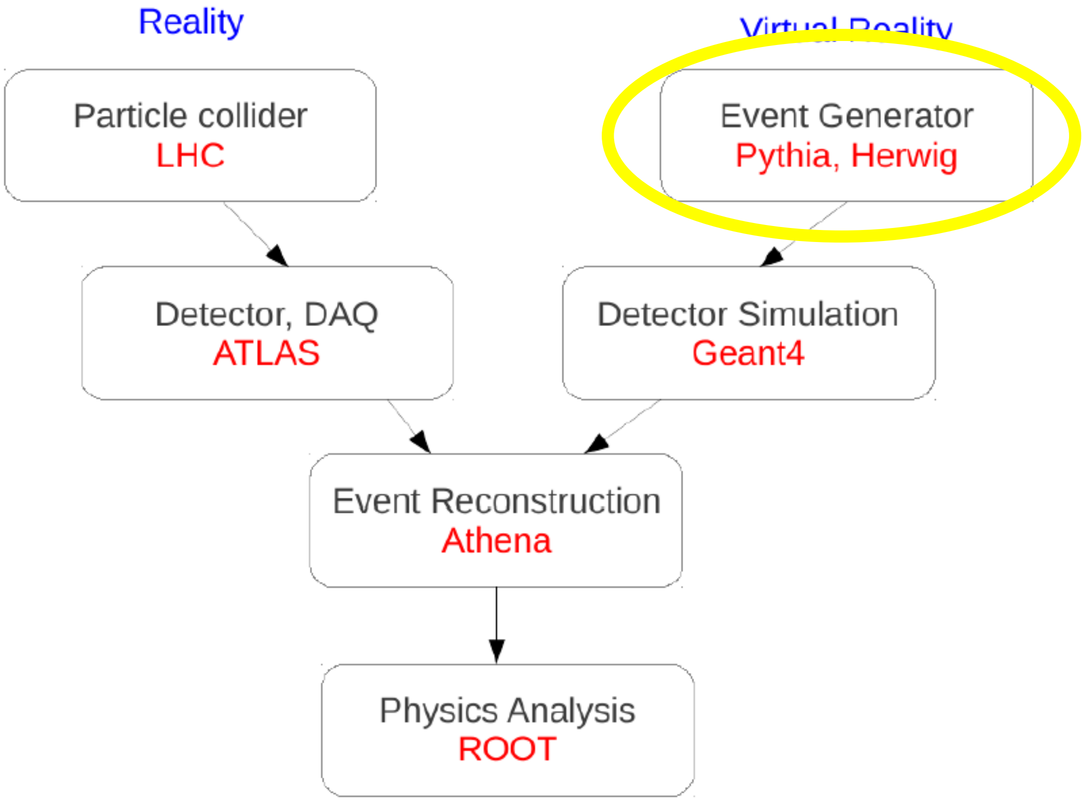
Integral $I = \int_{x1}^{x2} f(x)dx = (x_2 - x_1)\langle f(x)\rangle$

Draw N values from a uniform distribution:

$$I \approx I_N \equiv (x_2 - x_1)\frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

Cross section randomly sampled over phase space. Method governed by the Central Limit Theorem: errors $\propto \frac{1}{\sqrt{N}}$

*In particle physics applications: Random numbers represent QM choices



Reality

Virtual Reality

Particle collider
LHC

Event Generator
Pythia, Herwig

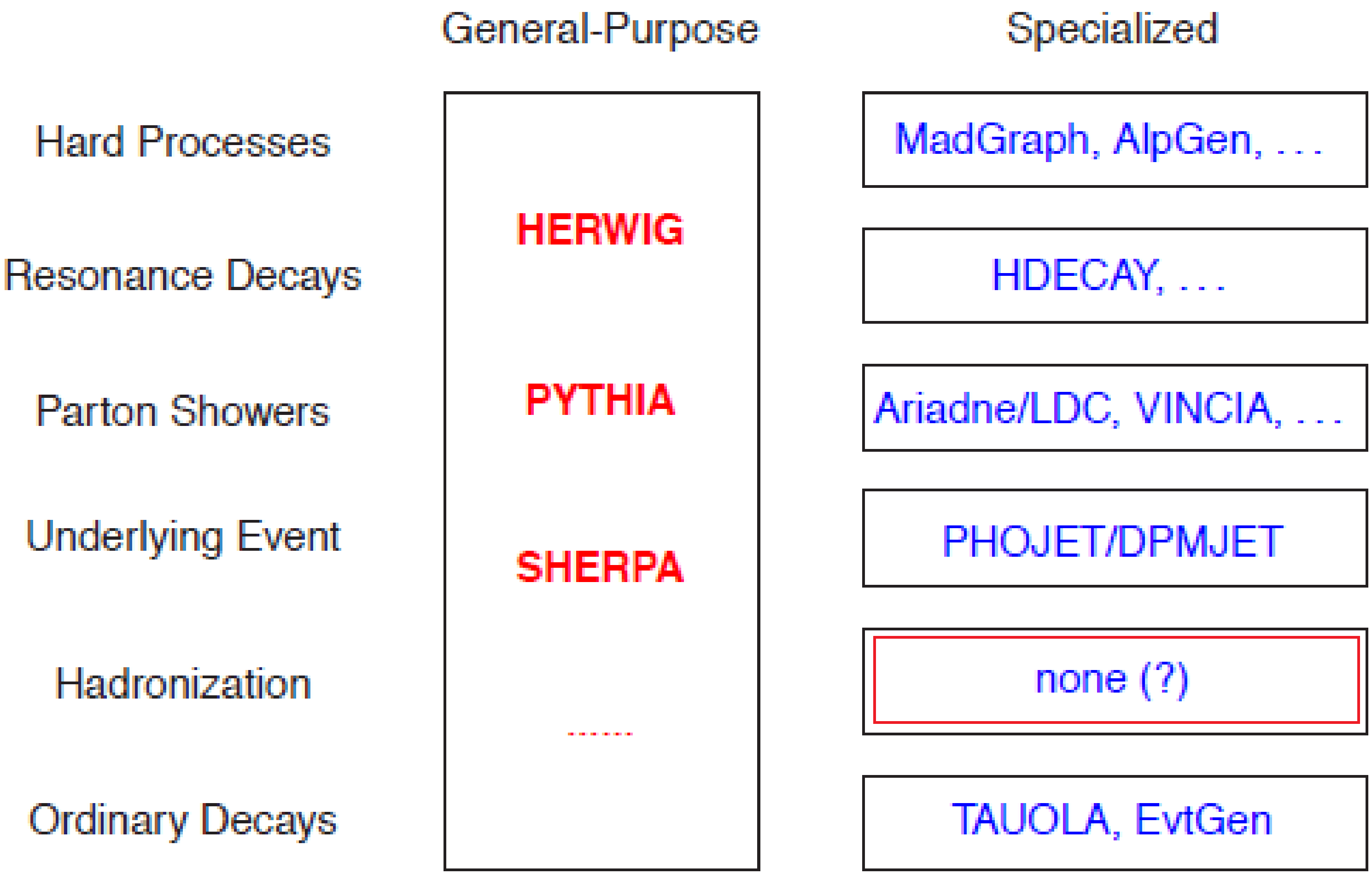
Detector, DAQ
ATLAS

Detector Simulation
Geant4

Event Reconstruction
Athena

Physics Analysis
ROOT

# Event generators



General-Purpose

Specialized

Hard Processes

Resonance Decays

Parton Showers

Underlying Event

Hadronization

Ordinary Decays

**HERWIG**

**PYTHIA**

**SHERPA**

……

MadGraph, AlpGen, . . .

HDECAY, . . .

Ariadne/LDC, VINCIA, . . .

PHOJET/DPMJET

none (?)

TAUOLA, EvtGen

Specialized often best at given task, but need General-Purpose core

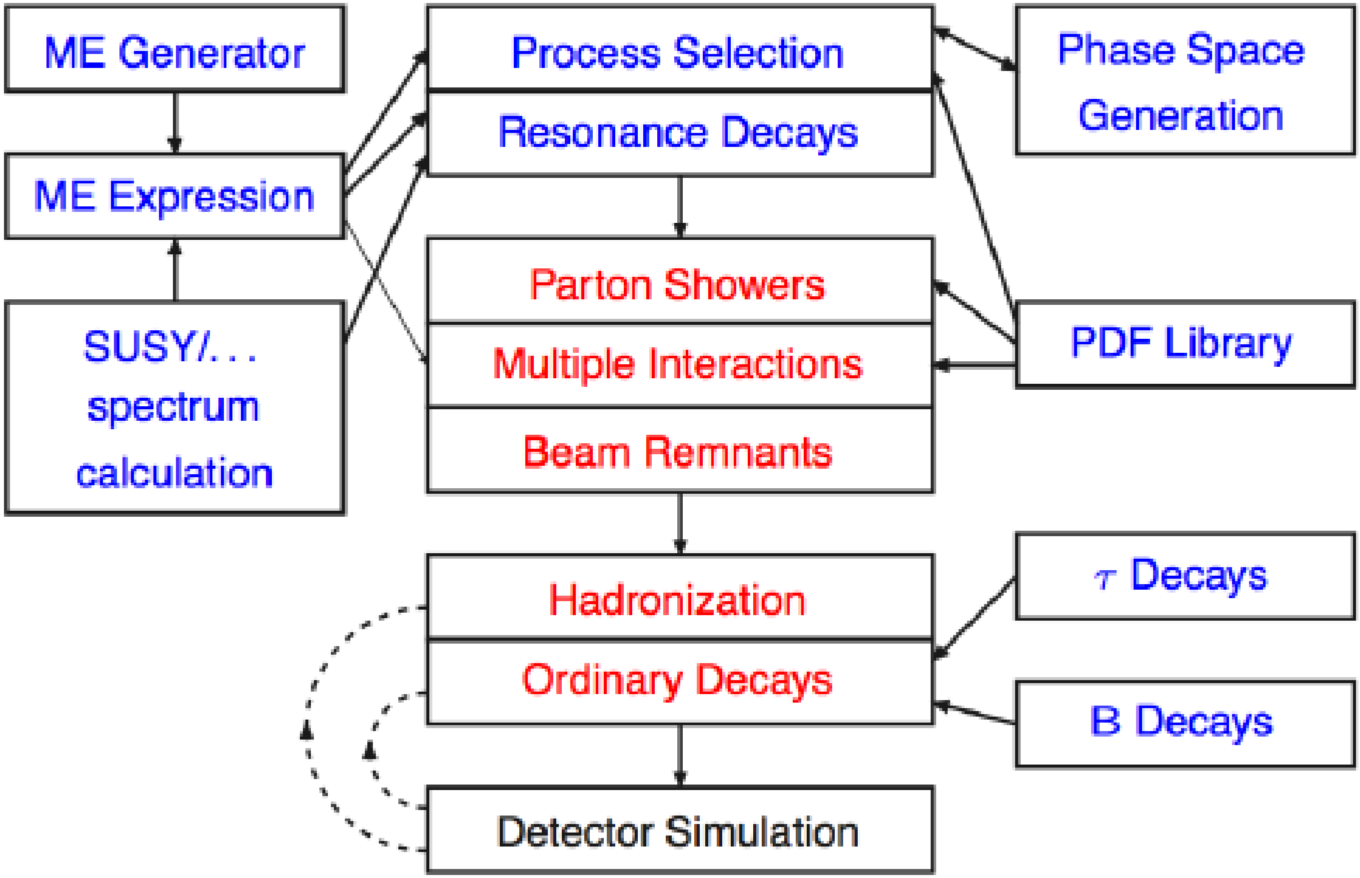
# Event generators

From Lund U phenomenology group!

| | General-Purpose | Specialized |
|---|---|---|
| Hard Processes | **HERWIG**<br>**PYTHIA**<br>**SHERPA**<br>....... | MadGraph, AlpGen, . . . |
| Resonance Decays | | HDECAY, . . . |
| Parton Showers | | Ariadne/LDC, VINCIA, . . . |
| Underlying Event | | PHOJET/DPMJET |
| Hadronization | | none (?) |
| Ordinary Decays | | TAUOLA, EvtGen |

Specialized often best at given task, but need General-Purpose core

# What they do

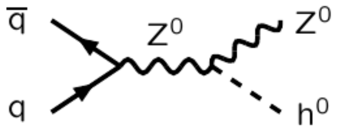


Several standardized interfaces!

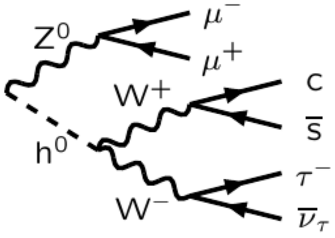
# Monte Carlo generation

**Matrix elements (ME):**

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

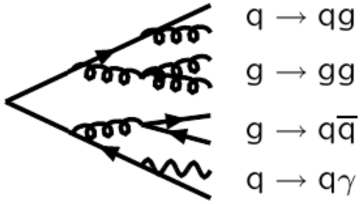


2) Resonance decays:
includes correlations.

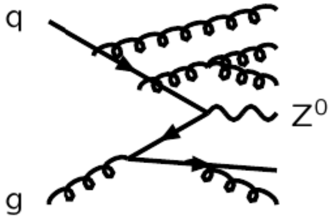


**Parton Showers (PS):**

3) Final-state parton showers.

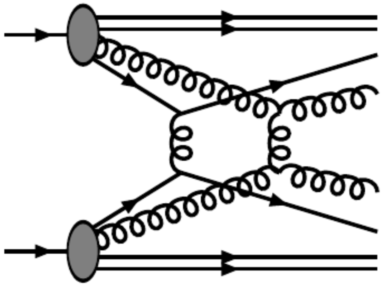


4) Initial-state parton showers.

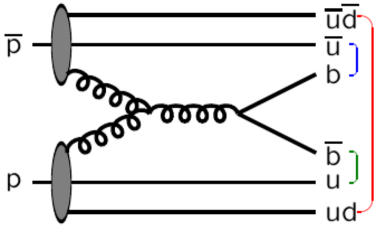


# Monte Carlo generation

5) Multiple parton–parton interactions.

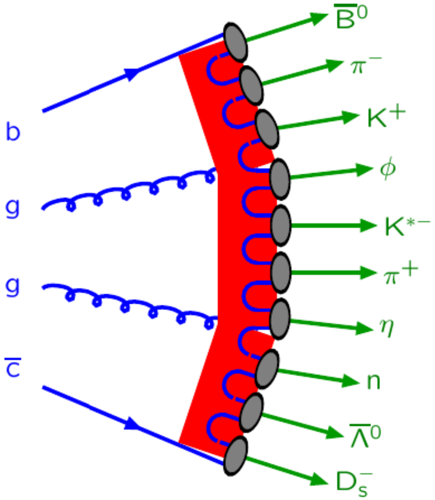


6) Beam remnants, with colour connections.



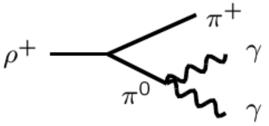
5) + 6) = Underlying Event

**7) Hadronization**

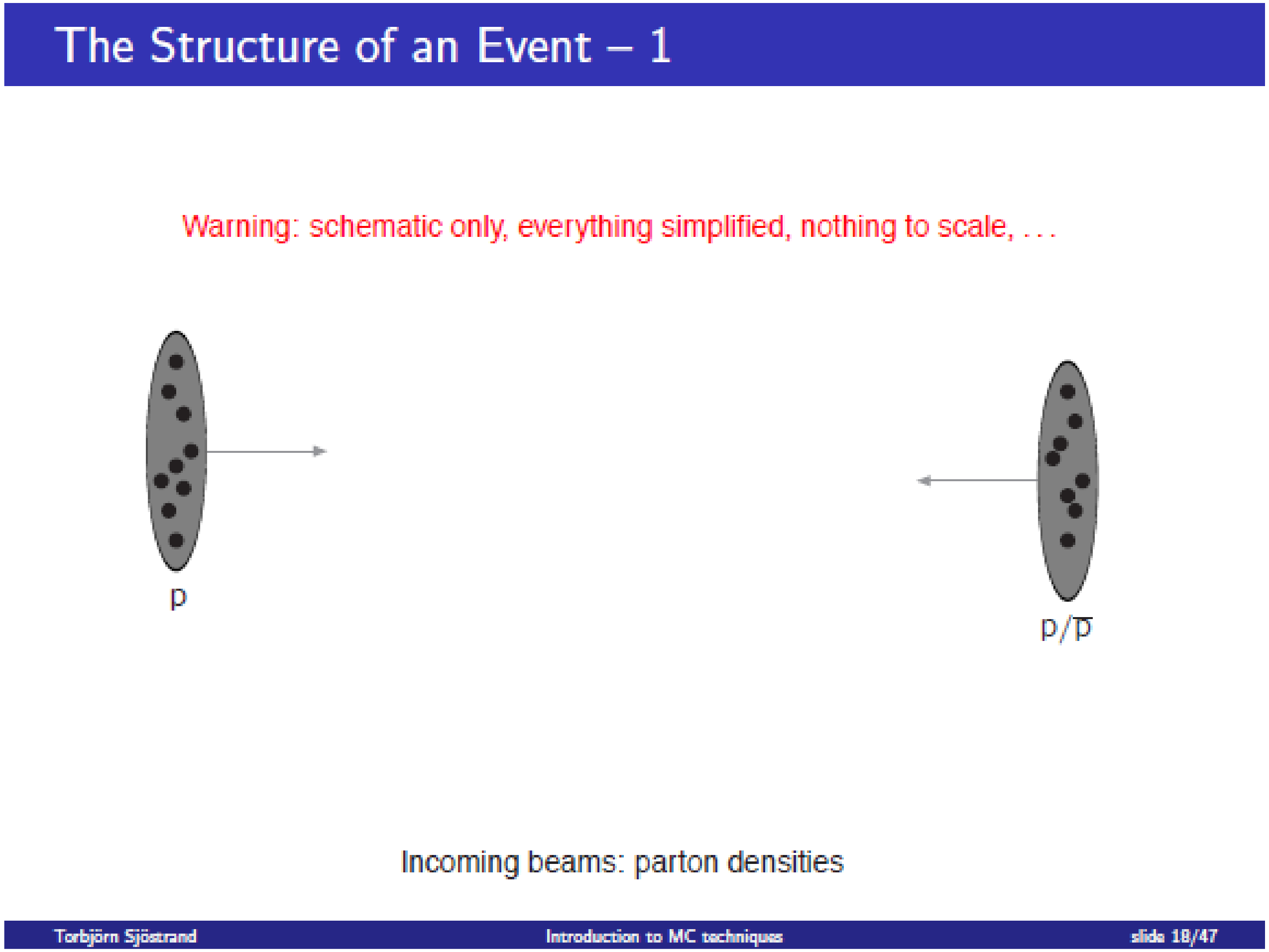


**8) Ordinary decays**:

hadronic, $\tau$, charm, . . .



# Slides from Torbjörn Sjöstrand



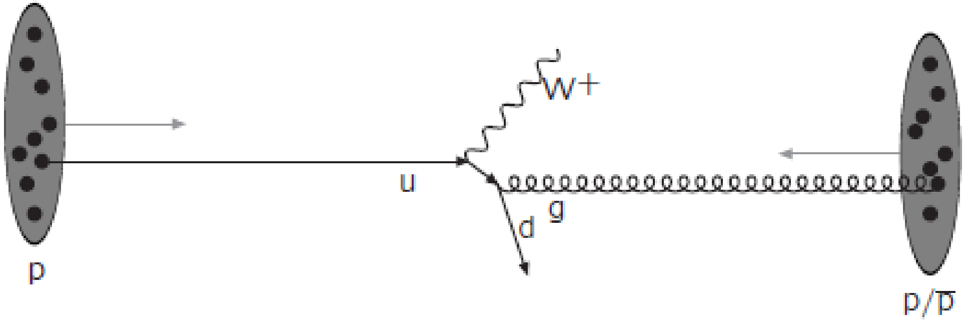
## The Structure of an Event – 1

Warning: schematic only, everything simplified, nothing to scale, . . .

Incoming beams: parton densities

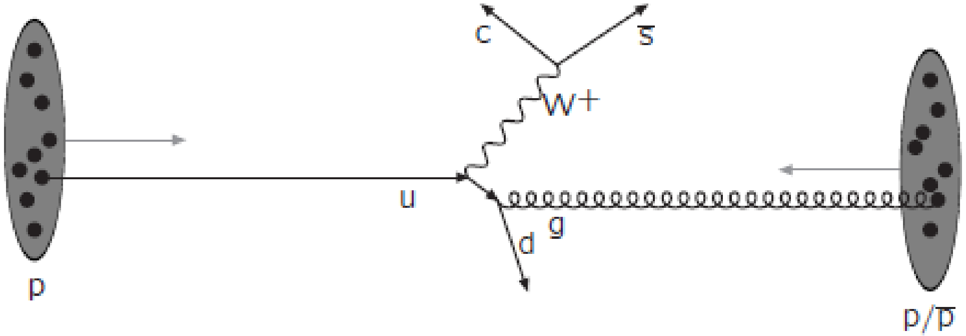
Torbjörn Sjöstrand Introduction to MC techniques slide 18/47

# The Structure of an Event – 2



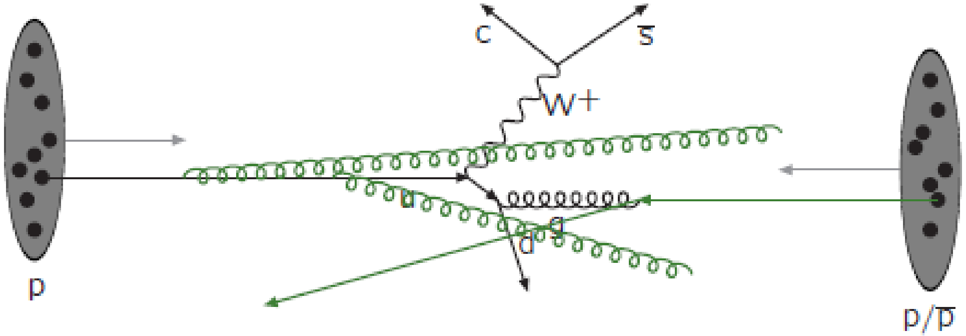
Hard subprocess: described by matrix elements

# The Structure of an Event – 3



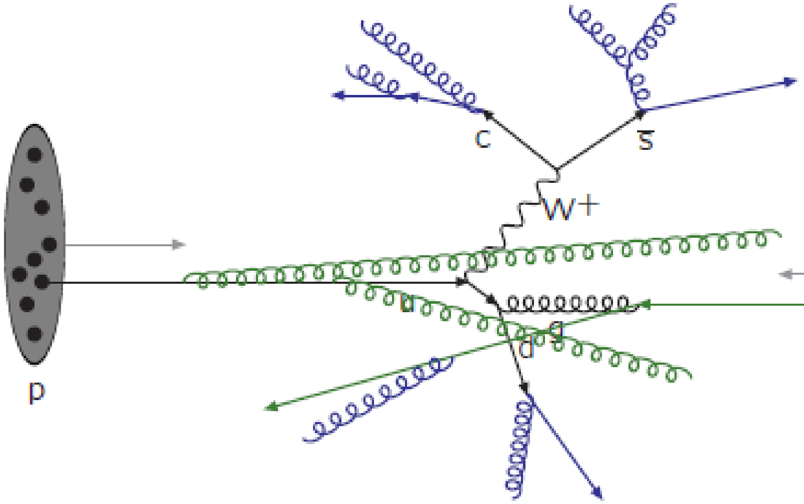
Resonance decays: correlated with hard subprocess

# The Structure of an Event – 4

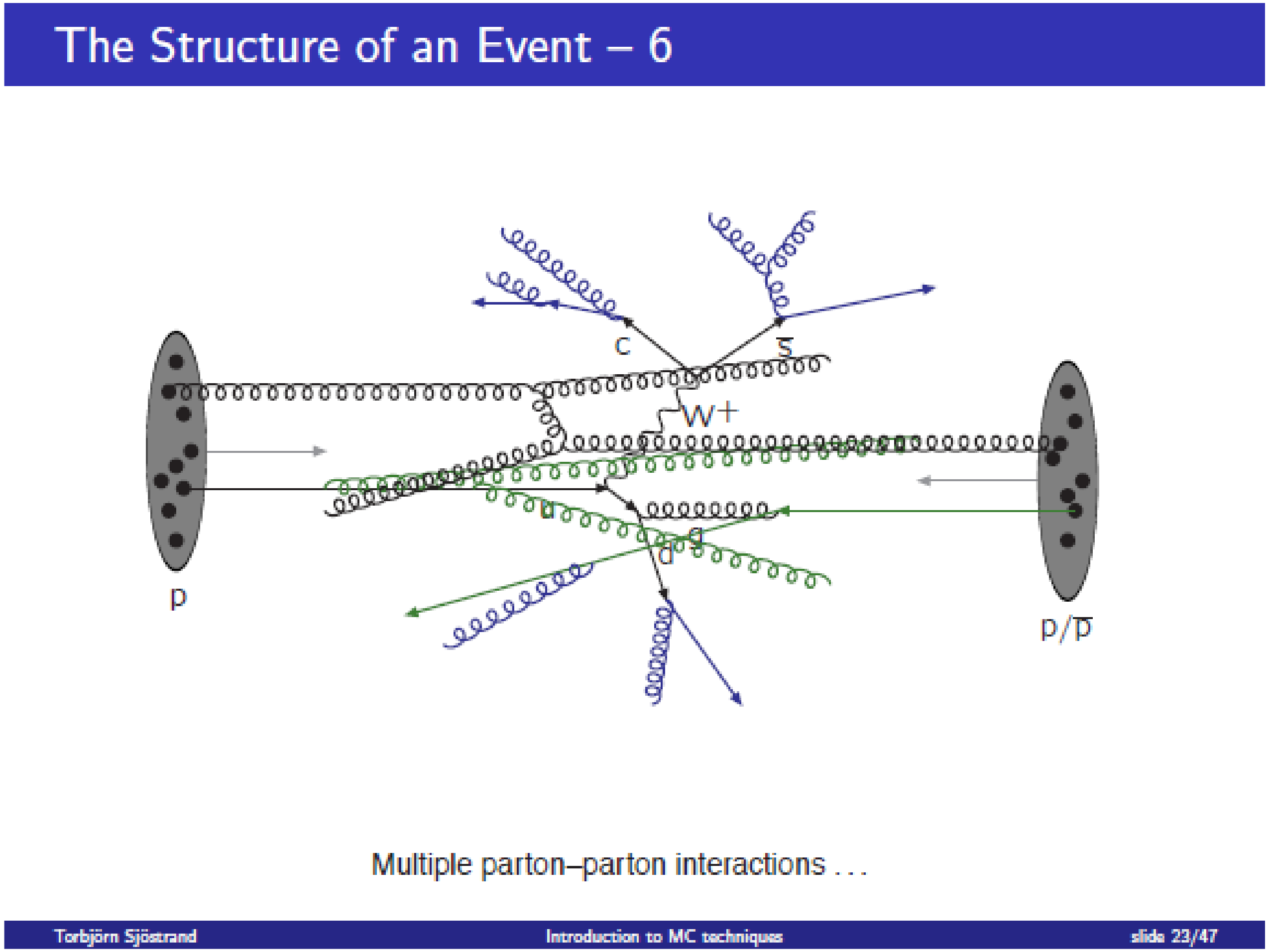


Initial-state radiation: spacelike parton showers

# The Structure of an Event – 5



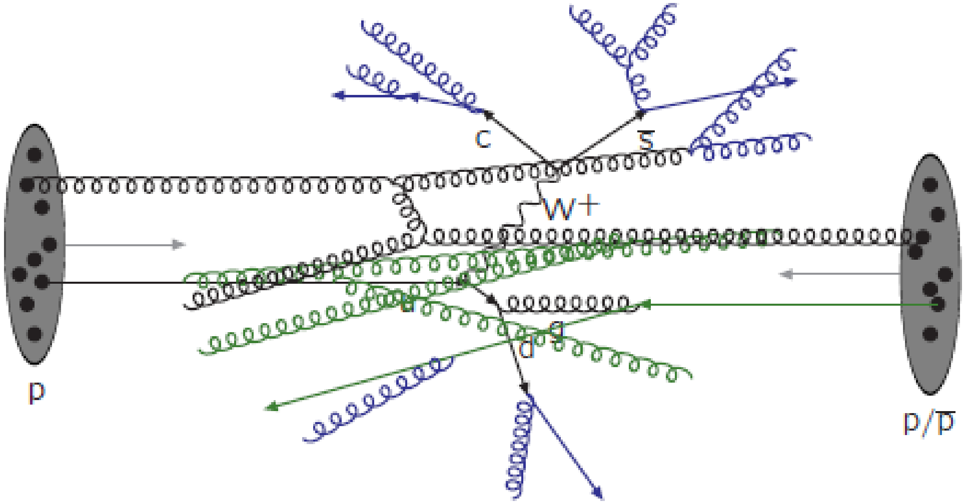
Final-state radiation: timelike parton showers



# The Structure of an Event – 6

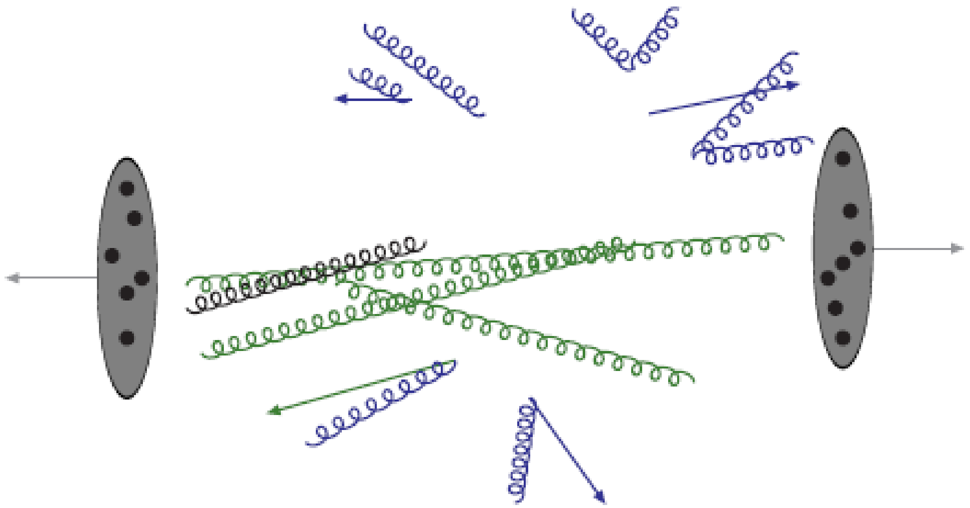
Multiple parton–parton interactions . . .

# The Structure of an Event – 7



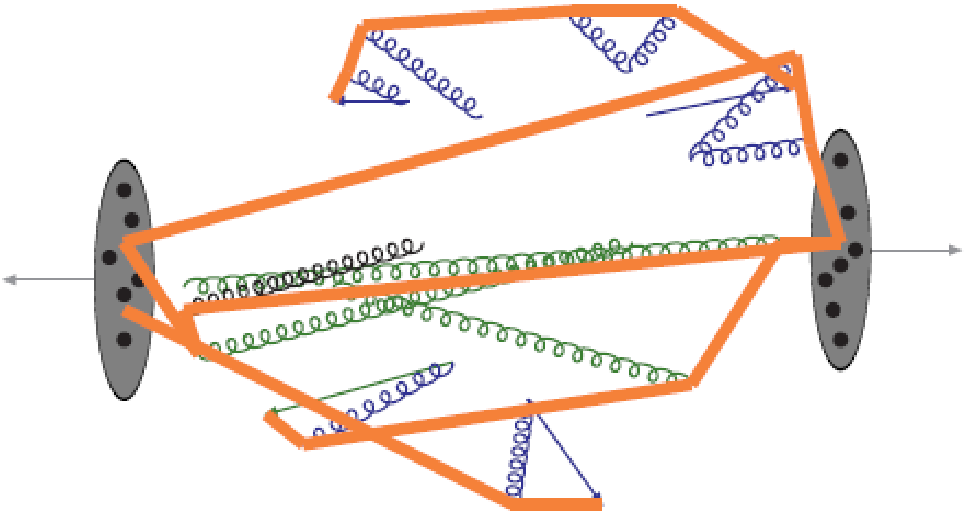
. . . with its initial- and final-state radiation

# The Structure of an Event – 8



Beam remnants and other outgoing partons

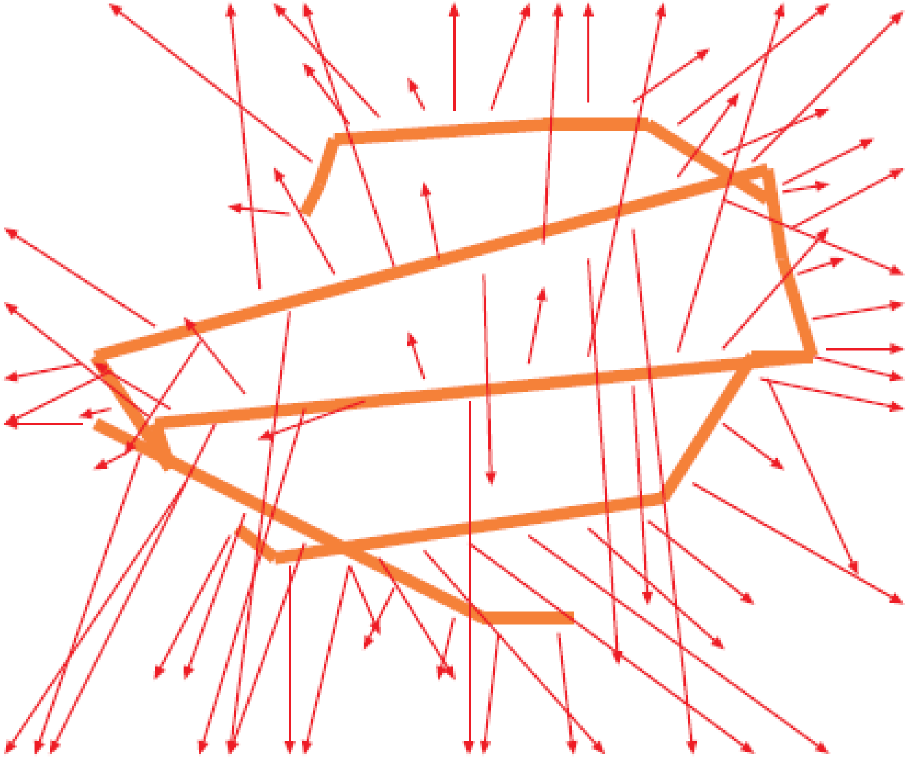
# The Structure of an Event – 9



Everything is connected by colour confinement strings

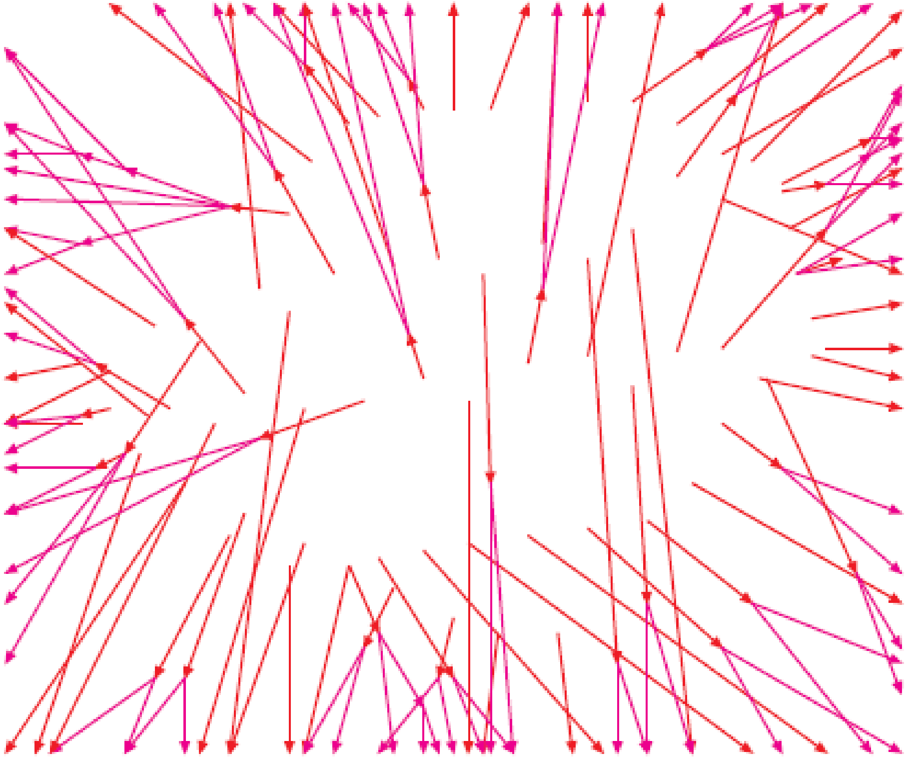
Recall! Not to scale: strings are of hadronic widths

# The Structure of an Event – 10



The strings fragment to produce primary hadrons

# The Structure of an Event – 11

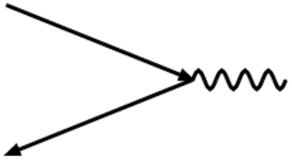


Many hadrons are unstable and decay further

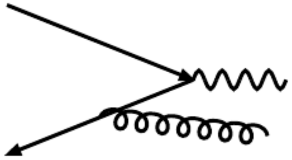
# Matrix Element calculation

Normally done to LO or NLO

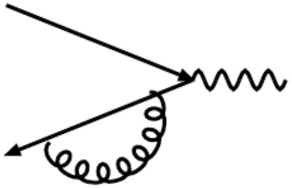
I. Lowest order,
$\mathcal{O}(\alpha_{\mathrm{em}})$:
$q\overline{q} \rightarrow Z^0$



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

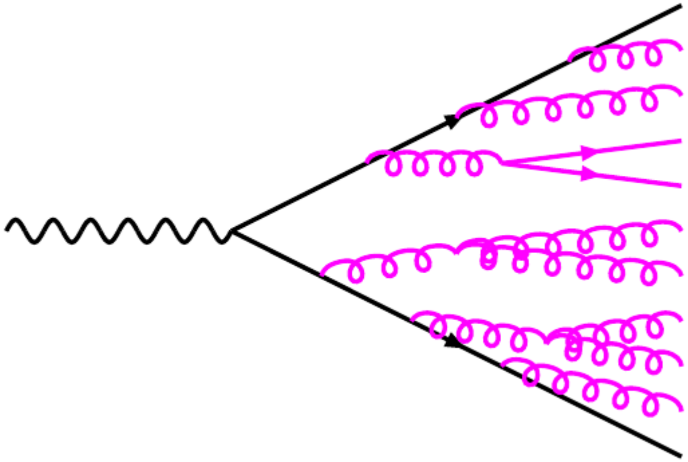


III. First-order virtual,
$\mathcal{O}(\alpha_{\mathrm{em}}\alpha_{\mathrm{s}})$:
$q\overline{q} \rightarrow Z^0$ with loops



- 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
  - $\Rightarrow$ 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 $p_{\mathrm{T}}$ is calculated first
- HERWIG : $Q^2 \approx E^2\left(1 - \cos\left(\theta\right)\right)$
  - 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

# Hadronization

## 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

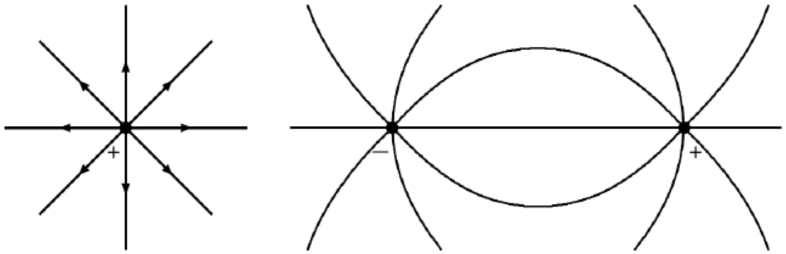
## This is another theoretical modelling uncertainty

- 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

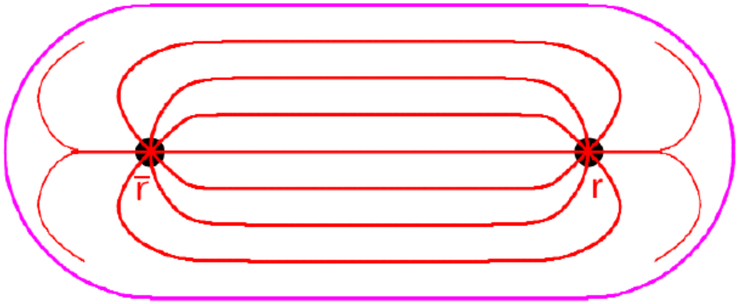
# Hadronization

The Lund String model:

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



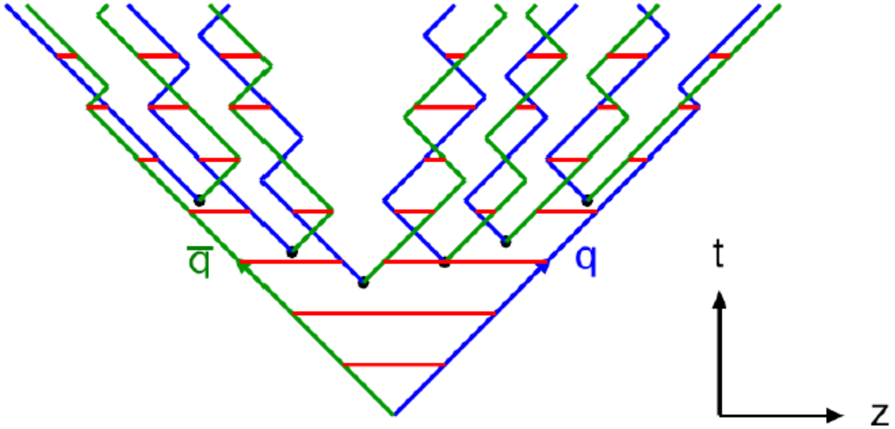
- In QCD, for large charge separation, field lines seem to be compressed into tube-like regions ⇒ string(s)
- Self-interaction among soft gluons in the vacuum



# Hadronization

The Lund String model:

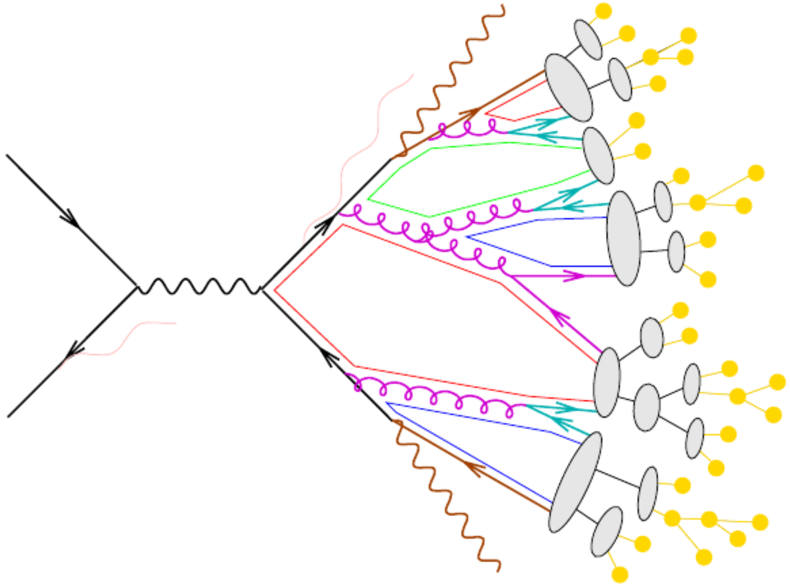
- The strings connecting the 2 partons breaks as they move apart
- Fragmentation starts in the middle and spreads out



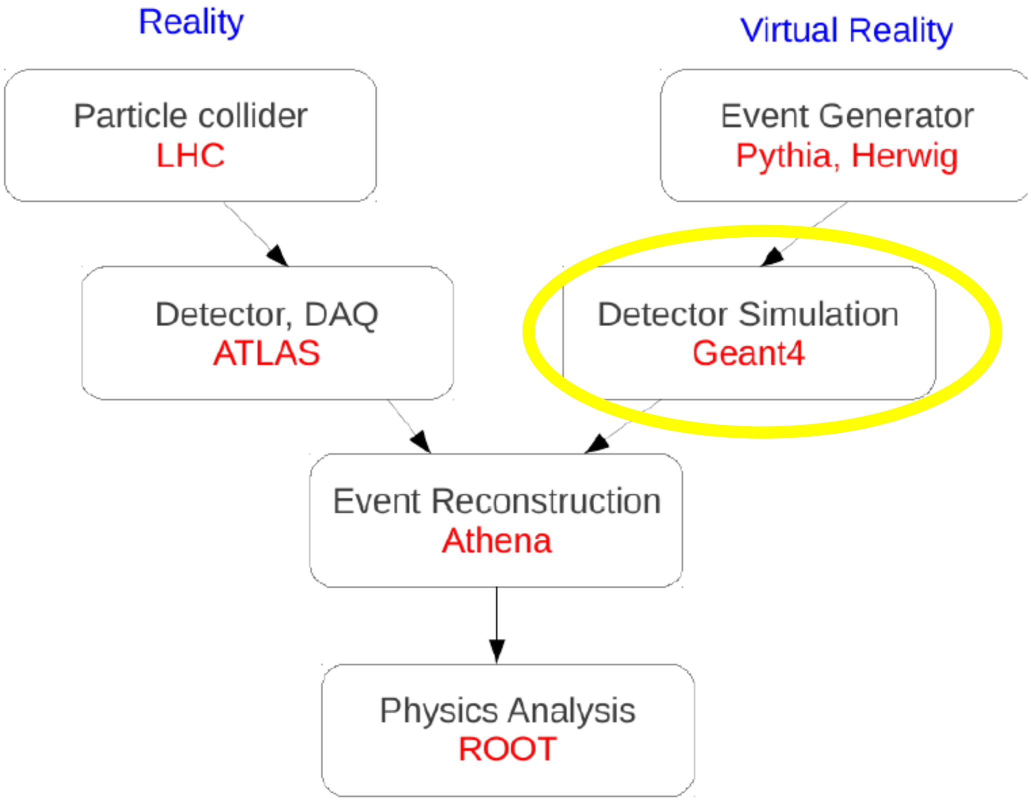
- The breakup vertices become causally disconnected
- This is governed by many internal parameters
- Implemented by the PYTHIA MC program

# Hadronization

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



Reality

Virtual Reality

Particle collider
LHC

Event Generator
Pythia, Herwig

Detector, DAQ
ATLAS

Detector Simulation
Geant4

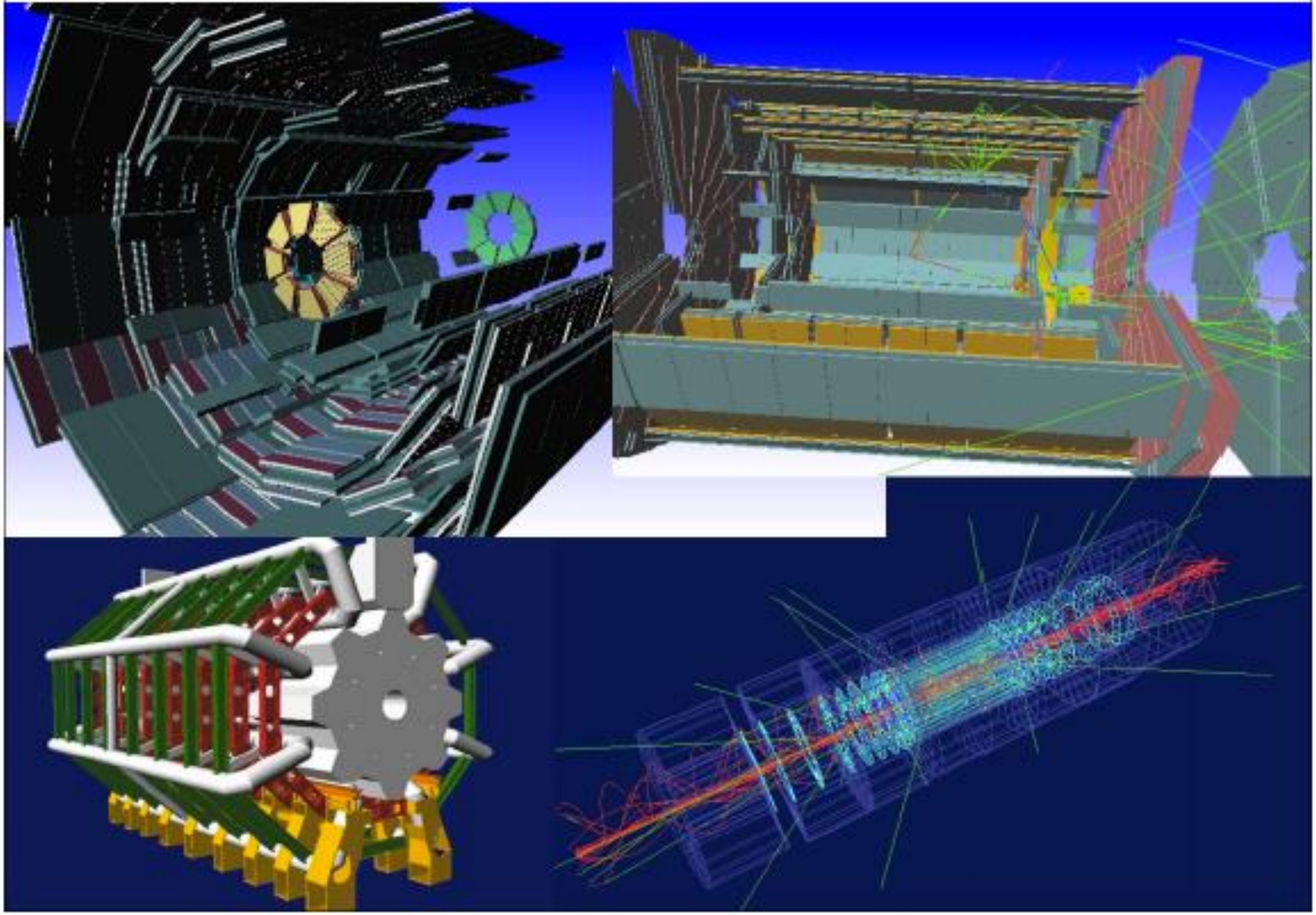
Event Reconstruction
Athena

Physics Analysis
ROOT

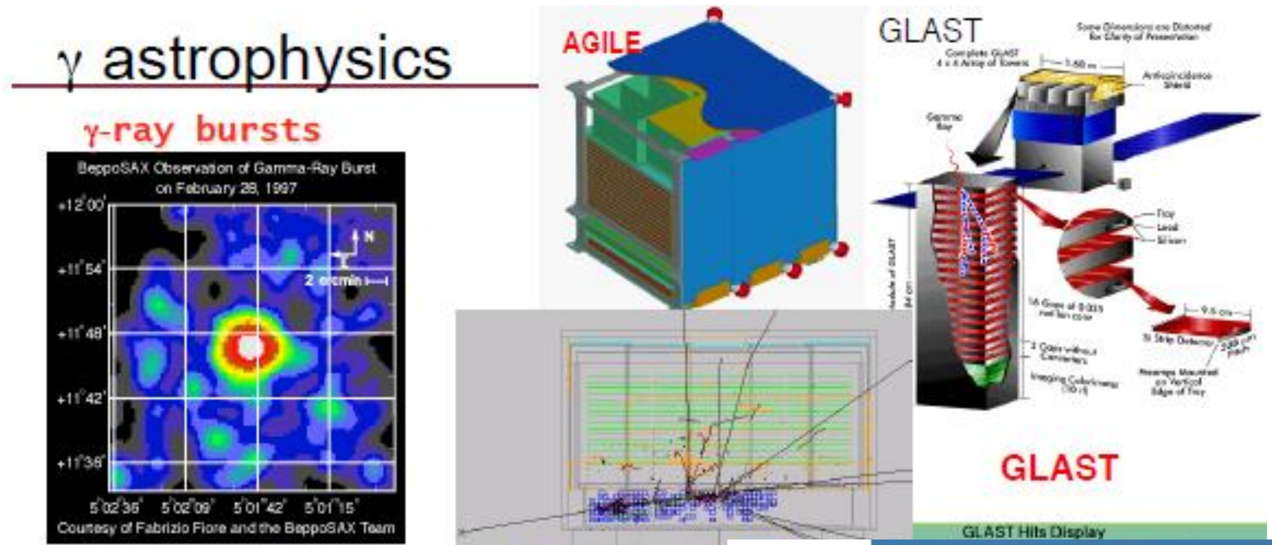
# Detector simulation

- Next step is simulating the particles paths through the detector:
  - Tracking chambers, calorimeters, muon system
  - but also cables, cooling pipes etc
  - and also faulty detector modules/electronics!
- Takes time: need to simulate all interactions, ionization, energy deposits, secondary interactions and decays, scattering …
- Mostly used: **GEANT4** a C++ program. Takes as input 4-vectors from event generators and outputs ”raw data”
- Takes up to 10 mins/event! Short-cut ***Fast simulation***: Smear the 4-vectors instead of calorimeter simulation

# Geant4 in High Energy Physics (ATLAS at LHC)



# Not just used in high energy physics

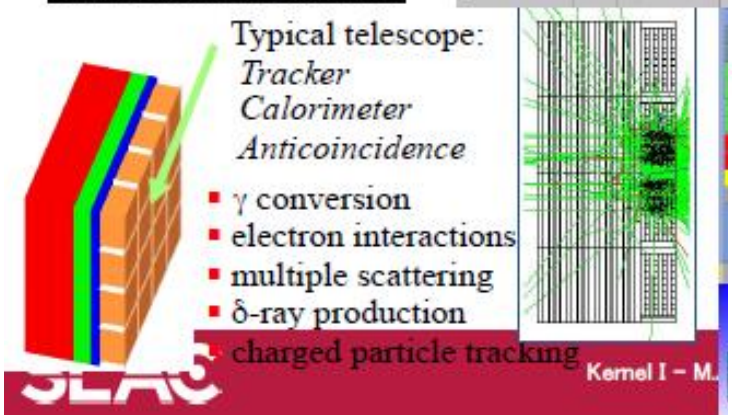


γ astrophysics

γ-ray bursts

BeppoSAX Observation of Gamma-Ray Burst on February 28, 1997

Courtesy of Fabrizio Fiore and the BeppoSAX Team



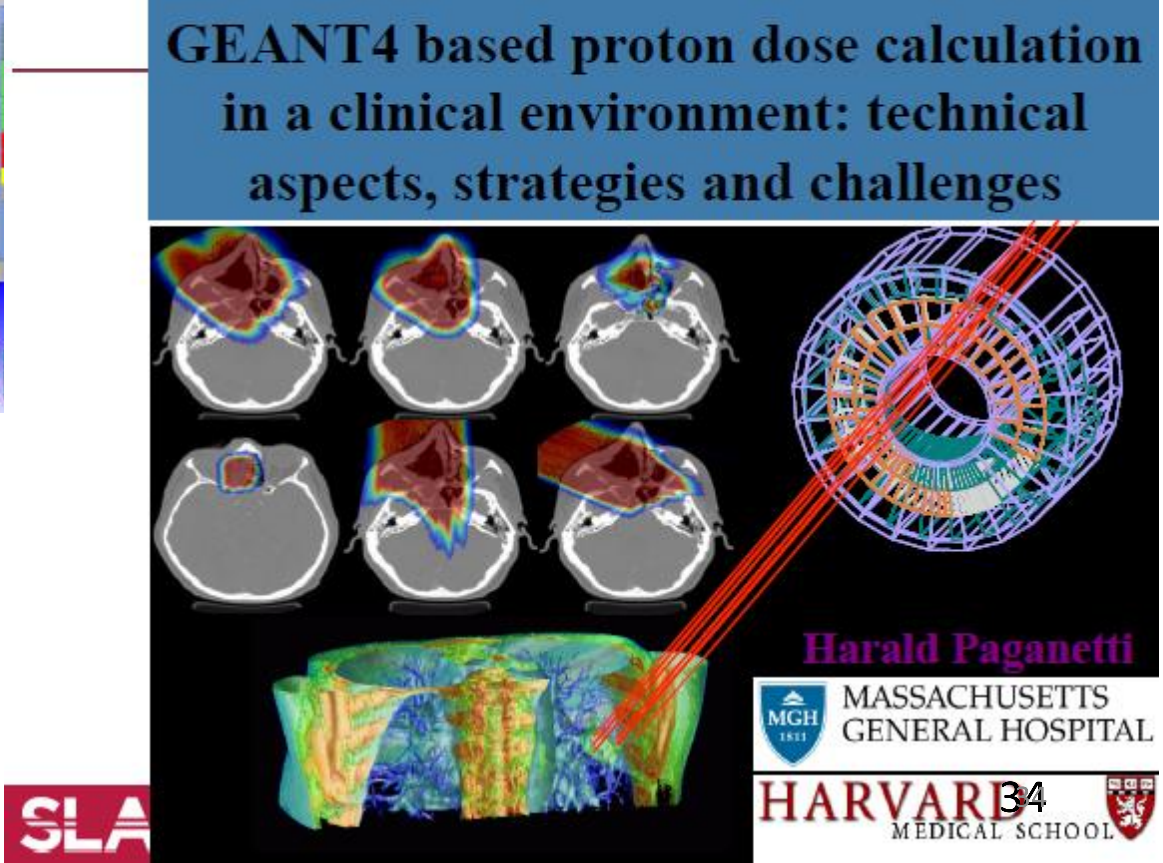
Typical telescope:

*Tracker*

*Calorimeter*

*Anticoincidence*

- γ conversion
- electron interactions
- multiple scattering
- δ-ray production
- charged particle tracking



**GEANT4 based proton dose calculation in a clinical environment: technical aspects, strategies and challenges**

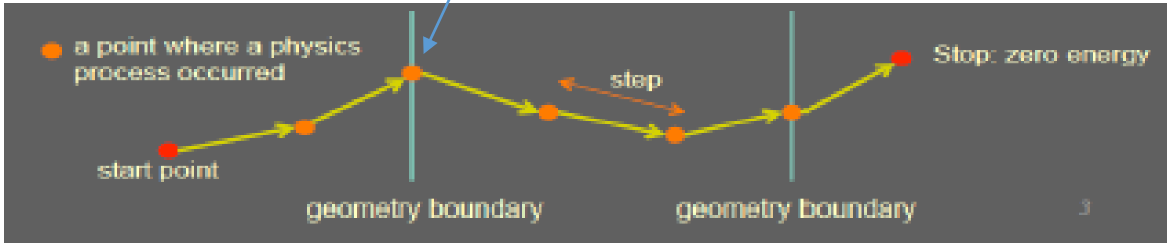
**Harald Paganetti**

MASSACHUSETTS GENERAL HOSPITAL

HARVARD MEDICAL SCHOOL

# How does it Work ?

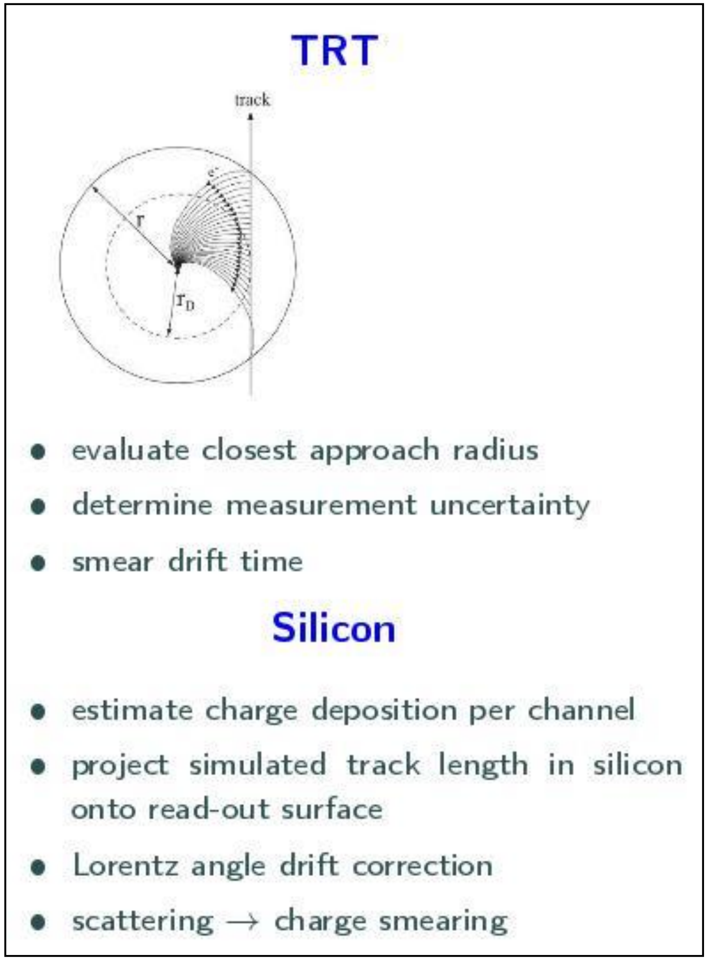
- Treat one particle at the time
- Treat a particle in **steps**



+ a photon?

- For each step
  - the step length is determined by the cross sections of the physics processes and the geometrical boundaries; if new particles are created, add them to the list of particles to be transported;
  - local energy deposit; effect of magnetic and electric fields;
  - if the particle is destroyed by the interaction, or it reaches the end of the apparatus, or its energy is below a (tracking) threshold, then the simulation of this particle is over;
    else continue with another step.
- Output - new particles created (indirect)
  - local **energy deposits** throughout the detector (direct)

# Digitization



**TRT**

- evaluate closest approach radius
- determine measurement uncertainty
- smear drift time

**Silicon**

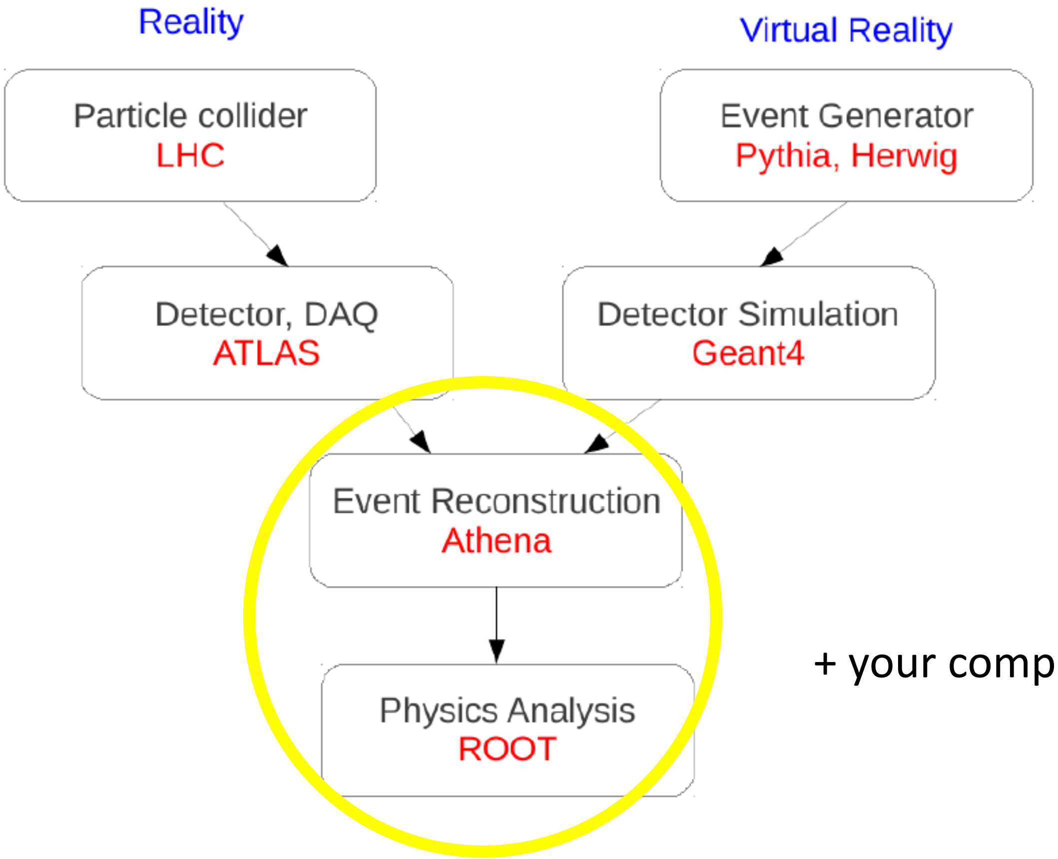
- estimate charge deposition per channel
- project simulated track length in silicon onto read-out surface
- Lorentz angle drift correction
- scattering → charge smearing

Before we are ready to run the same reconstruction algorithms as on data, the GEANT output needs to be ***digitized***

That is, converting the simulated hits in detectors into signals in read-out electronics

Also trigger simulation can be done at this level

Time consumption dominated by inner detector (most channels)


Reality

Virtual Reality

Particle collider
LHC

Event Generator
Pythia, Herwig

Detector, DAQ
ATLAS

Detector Simulation
Geant4

Event Reconstruction
Athena

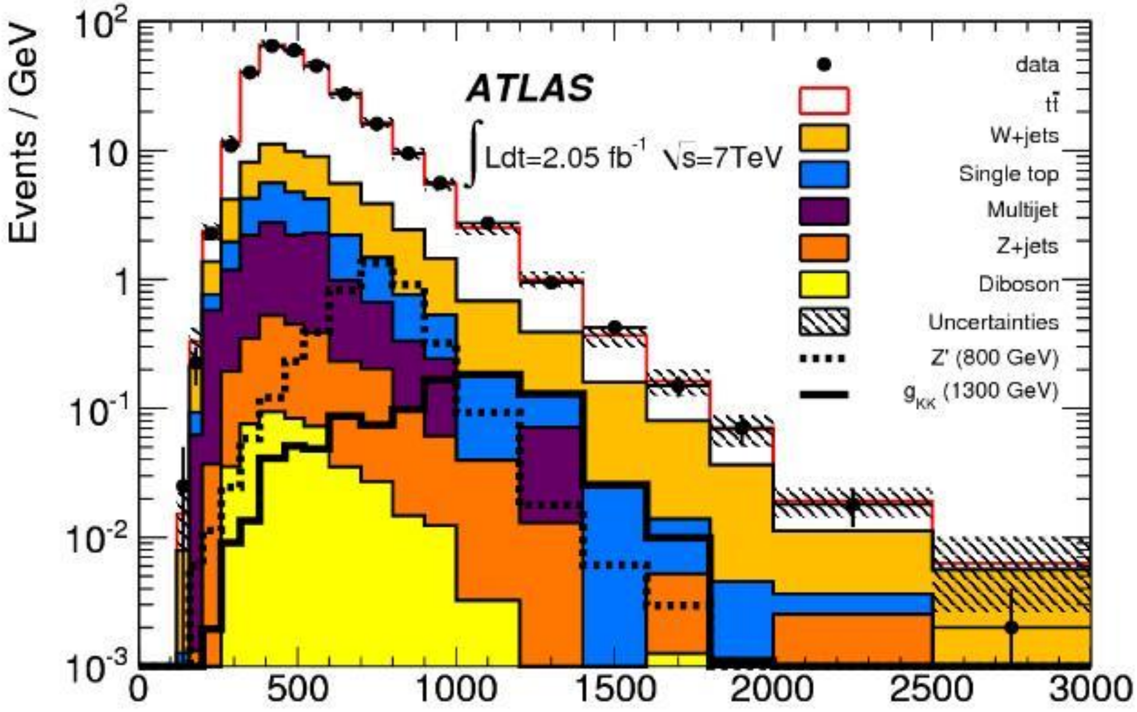
Physics Analysis
ROOT

+ your computer exercise

# Putting it all together

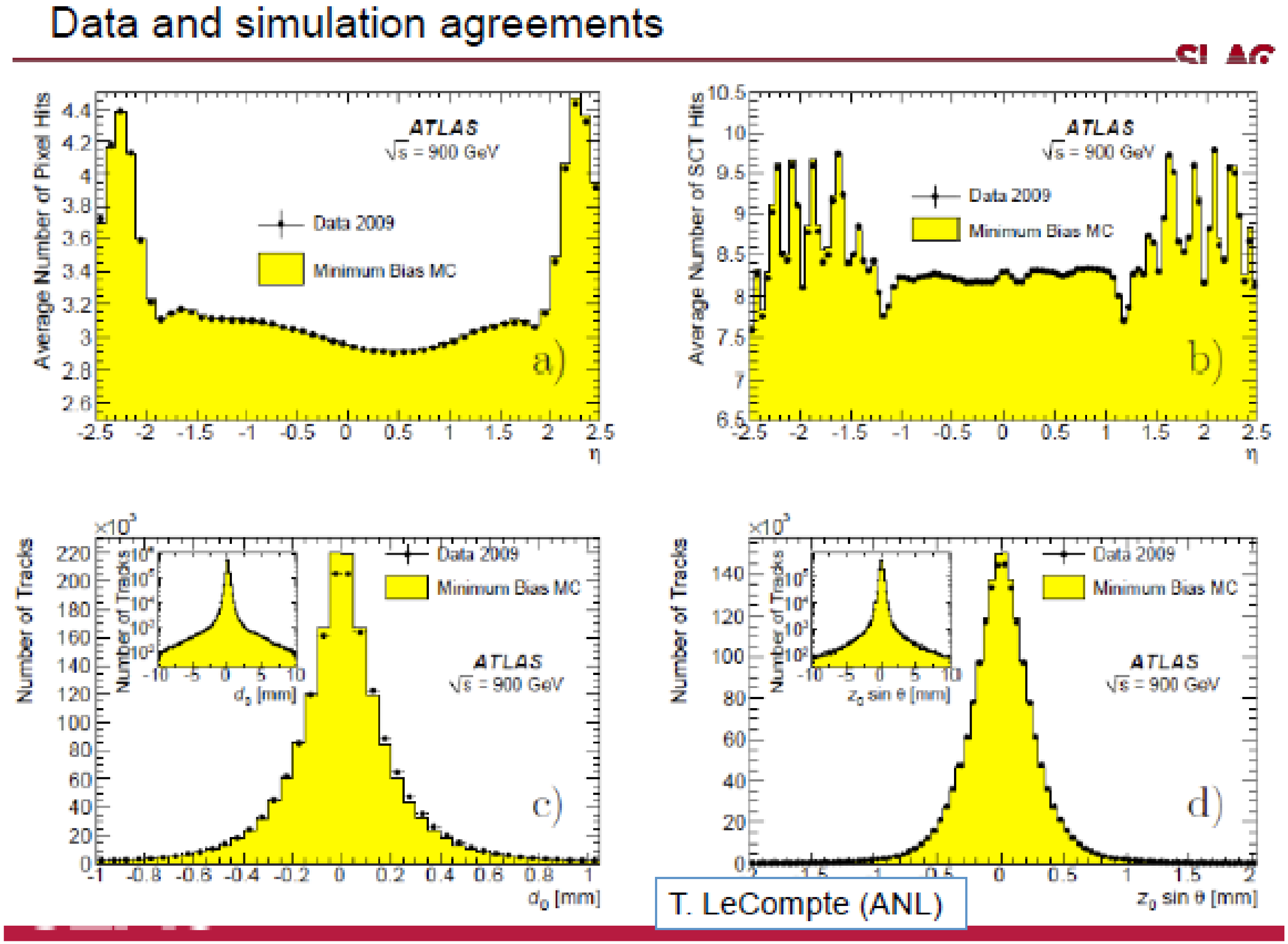
MC is not the truth! – tests/validation necessary

Some features are time-dependent ie amount of pile-up, technical problems with the detector, center of mass energy etc



Need to update (and test!) the simulation regularly

# Minimum bias events

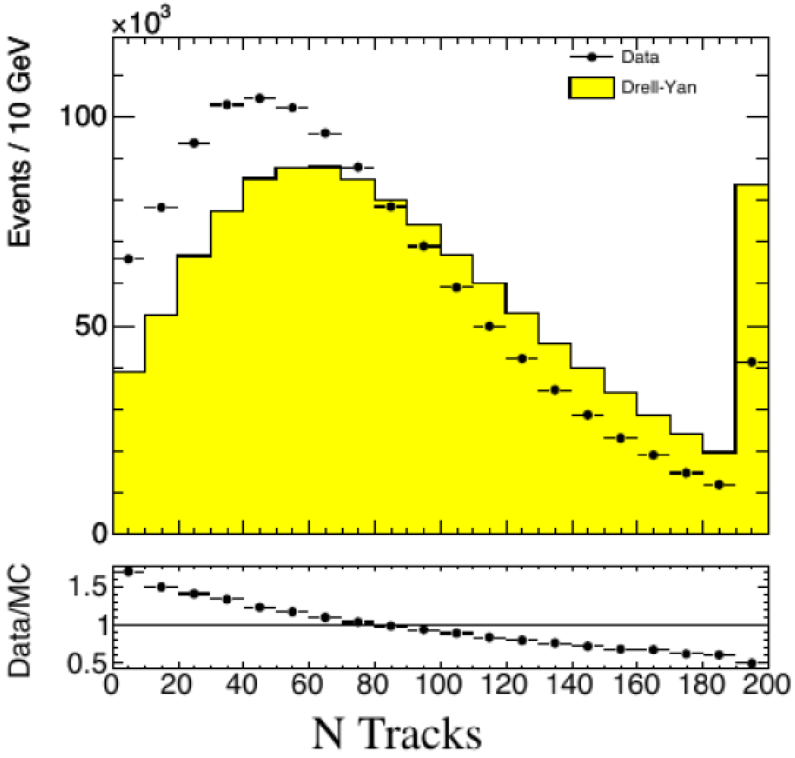


Data and simulation agreements

T. LeCompte (ANL)

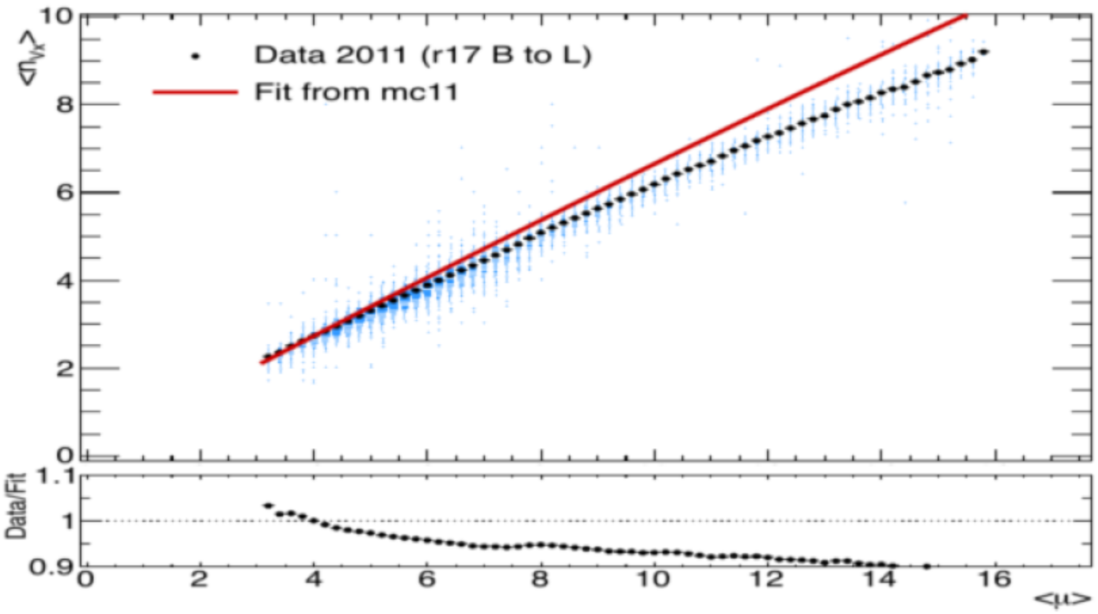
# It doesn’t always work

## Number of tracks (ATLAS)



# Re-weighting effect of pile-up

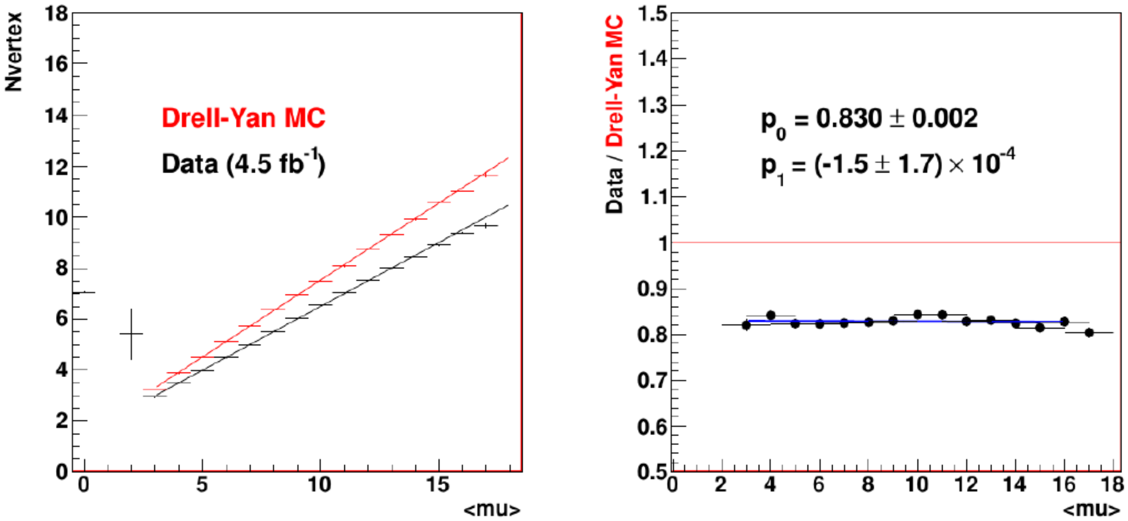
N vertices Vs average N interactions per bunch crossing



- 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

# Re-weighting the MC

## Need to determine re-weighting factors



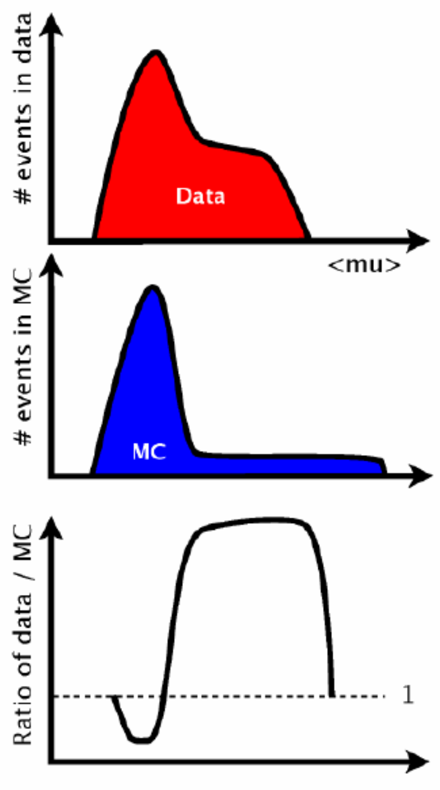
(a) Data-MC comparison

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

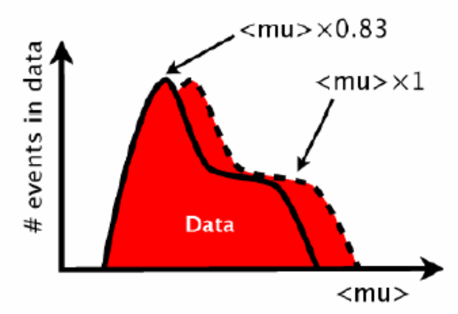
- Divide Data by MC to determine correction
- In this case, fit the ratio and determine a weight
- Use this weight for each MC event
  - histogram → Fill(x, weight);

# Re-weighting the MC

## Illustration of re-weighting procedure



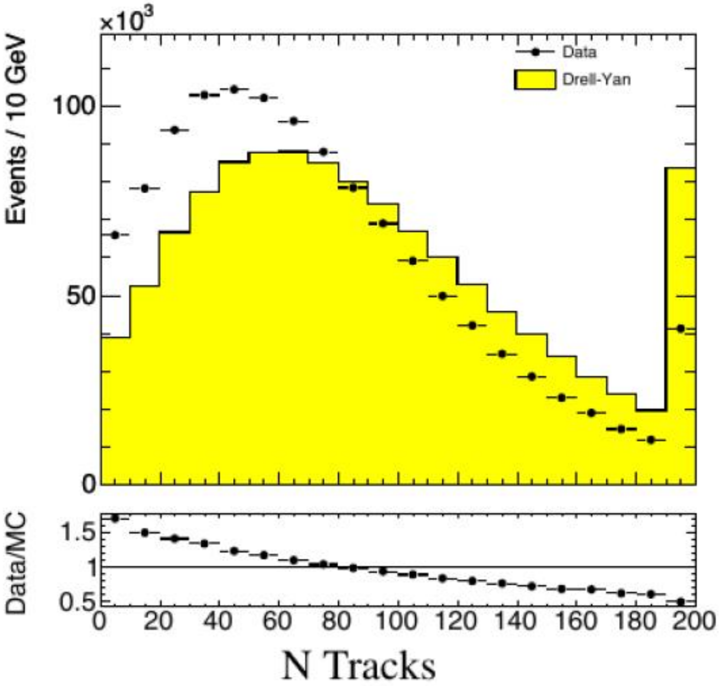
(a) Reweight procedure



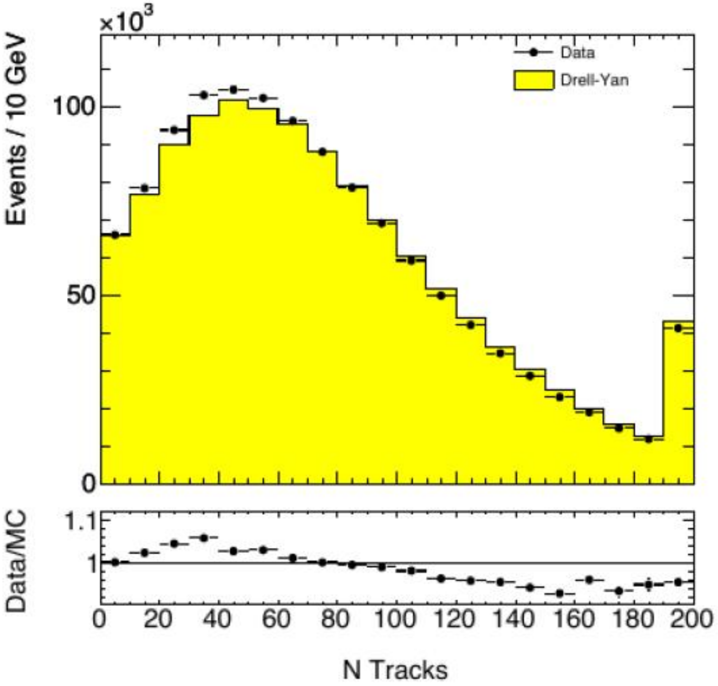
(b) Rescale procedure

Pile-up for instance , is hard to get right, we only know the exact conditions *after* data-taking is over

# After reweighting the agreement is much better
# Main problem is understanding the number of vertices

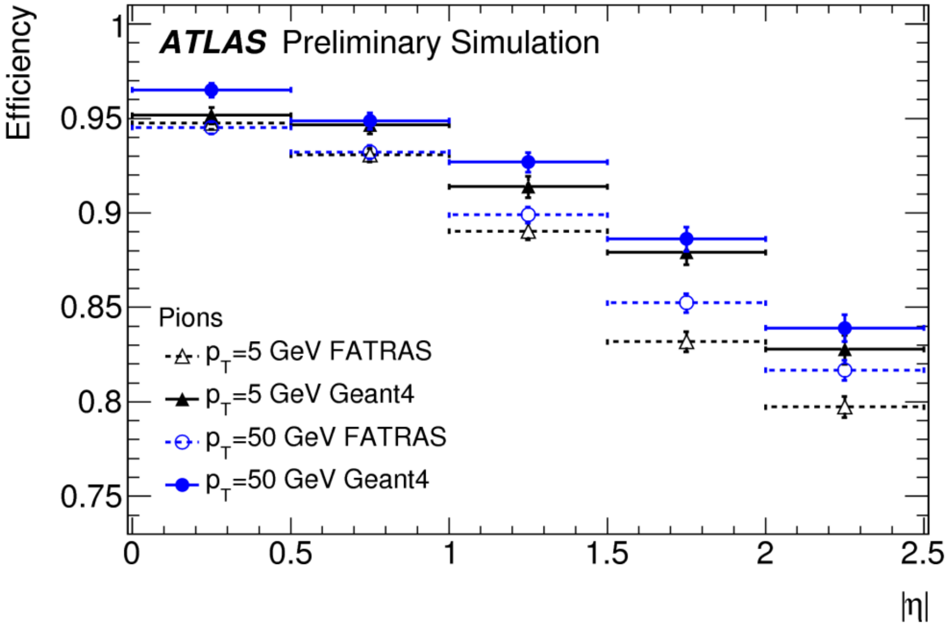


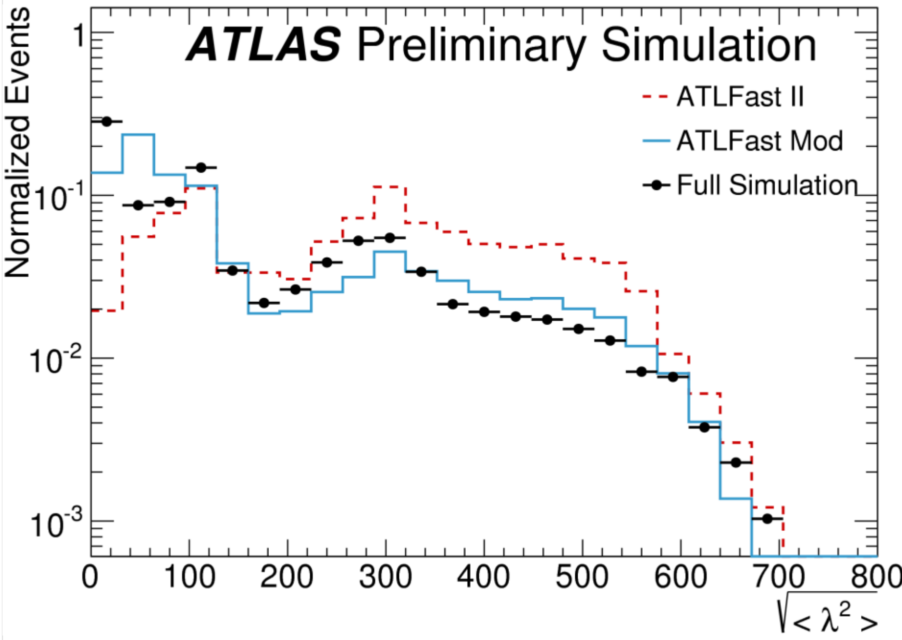
(a) ⟨$\mu$⟩ rescale factor= 1 (no rescale)



(b) ⟨$\mu$⟩ rescale factor= 0.83

# Fast vs full simulation





Although less meticulous, the fast simulation can be easily tuned to GEANT – or to data!

# Summary

- Most processes are impractical or impossible to calculate analytically
  - Therefore we use simulation to prepare for analysis

- Two steps: event generation (the physics process) and detector simulation (interaction with materials + electronics)
  - Several choices when it comes to event generators. Each have the pros and cons
  - Detector simulation = GEANT4 + digitization code
  - PYTHIA is a Lund product. You can try it yourself at: [http://home.thep.lu.se/~torbjorn/Pythia.html](http://home.thep.lu.se/~torbjorn/Pythia.html)

- It works! Many good comparisons between data and MC gives us confidence that we should notice the first non-SM physics!