From Raw Data to Physics: Reconstruction and Analysis

Reconstruction: Particle ID

How we try to tell particles apart

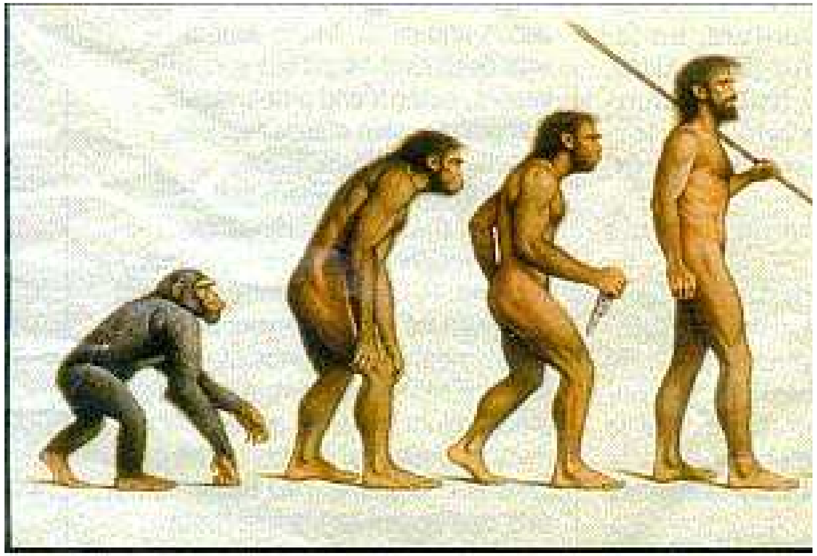
Analysis: Measuring α_S in QCD

What to do when theory doesn't make clear predictions

Alignment

We know what we designed; is it what we built?

Summary



From Raw Data to Physics: Reconstruction and Analysis

Reconstruction: Particle ID

How we try to tell particles apart

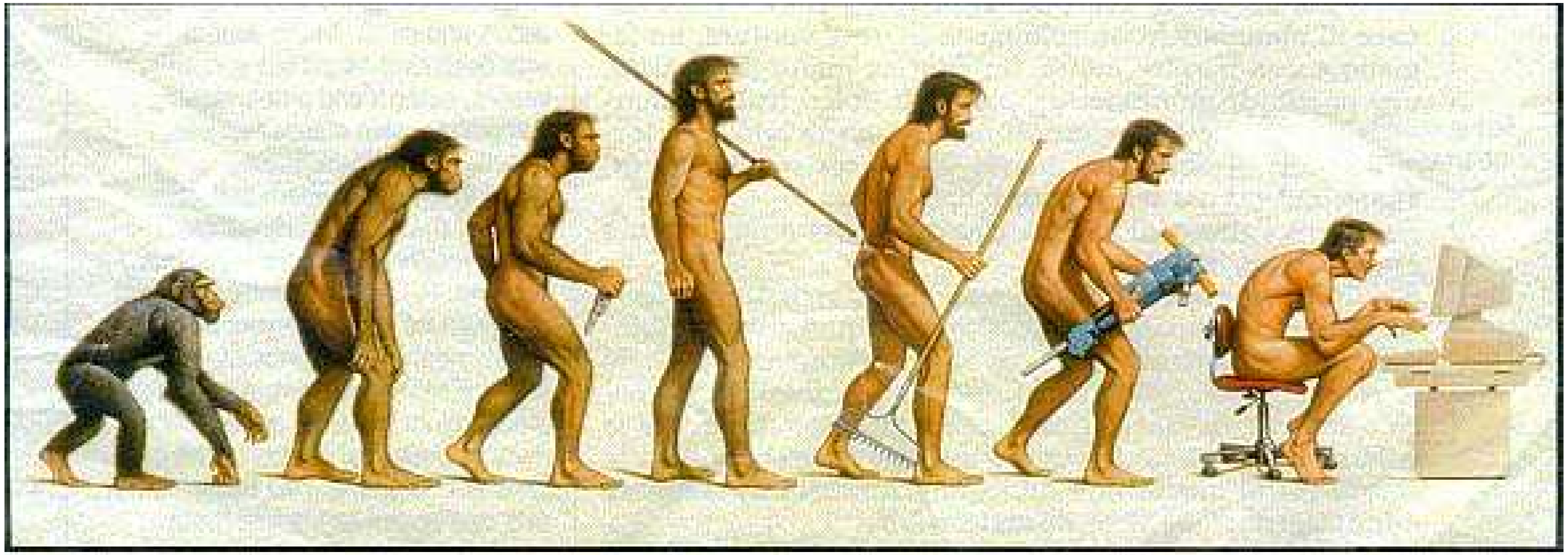
Analysis: Measuring α_S in QCD

What to do when theory doesn't make clear predictions

Alignment

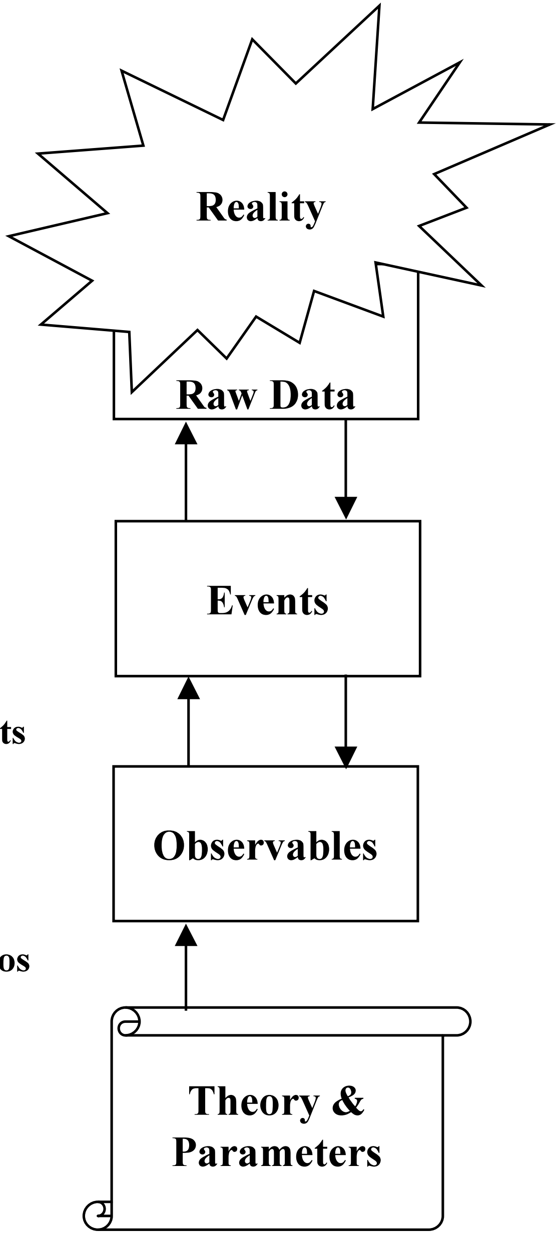
We know what we designed; is it what we built?

Summary



Somewhere, something went terribly wrong

Monte Carlo simulation's role



Calculate what imperfect detector would have seen for those events

Treat that as real data and reconstruct it

Randomly pick decay paths, lifetimes, etc for a number of events

Compare to original to understand efficiency

Calculate expected branching ratios

What about the computing behind this?

BaBar records about 70k B events per day

- Hidden in 7 million events recorded/day
- Take data about 300 days per year

'Prompt processing'

- Want data available in several days
- Reconstruction takes about 3 CPU seconds/evt
- Processed multiple times
  - E.g. new algorithms, constants, etc

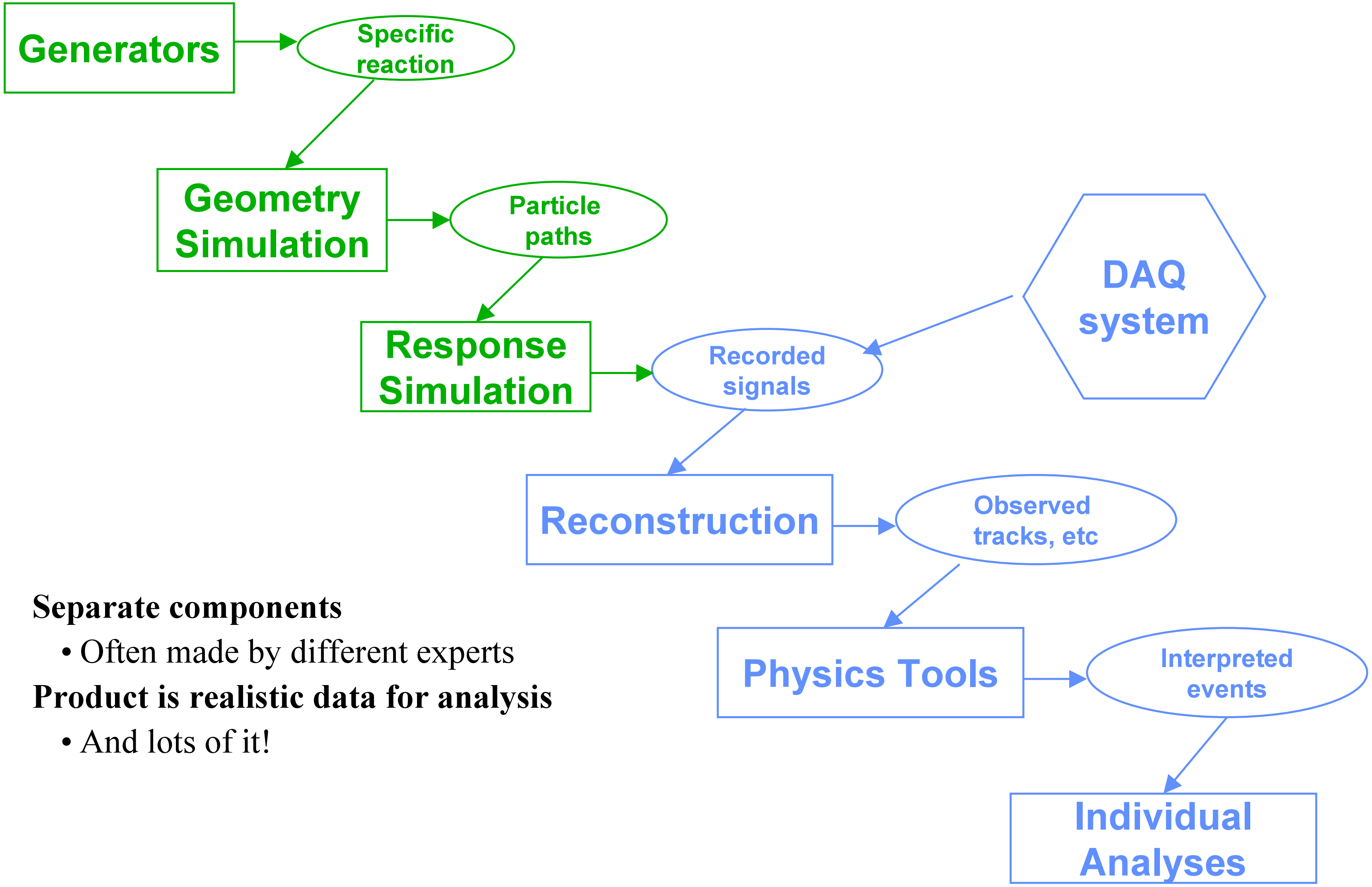
We have about 600 million simulated events to study

- About half in specific decay modes
- Half 'generic' decays to all modes

About 4 million lines of code in simulation and reconstruction programs

- Plus the individual analyses

Traditional flow of data - real and simulated



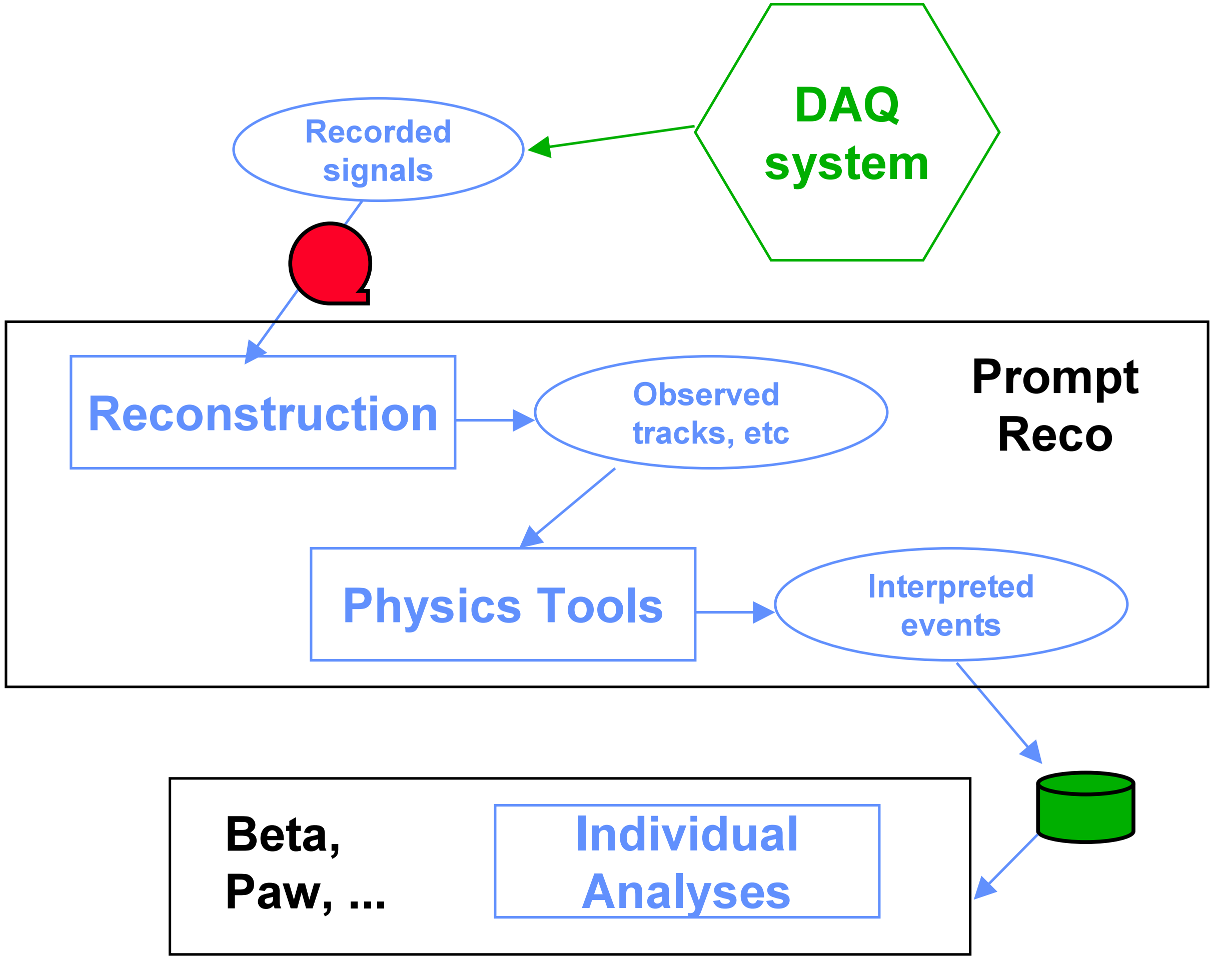
Separate components

- Often made by different experts

Product is realistic data for analysis

- And lots of it!

Processing real data



DAQ system

Recorded signals

Reconstruction

Observed tracks, etc

Prompt Reco

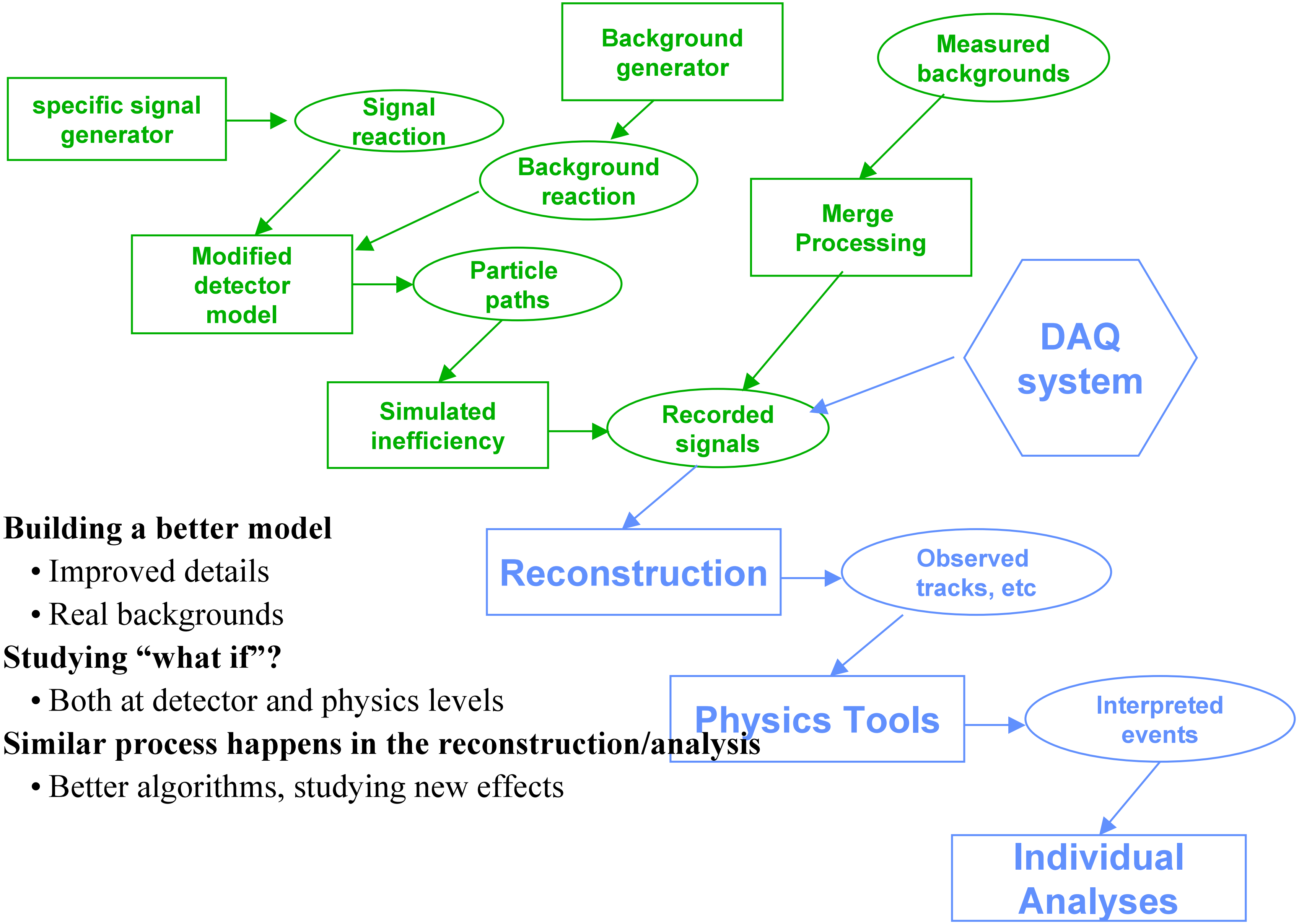
Physics Tools

Interpreted events

Beta, Paw, ...

Individual Analyses

More detailed studies via more detailed simulation



Building a better model

- Improved details
- Real backgrounds

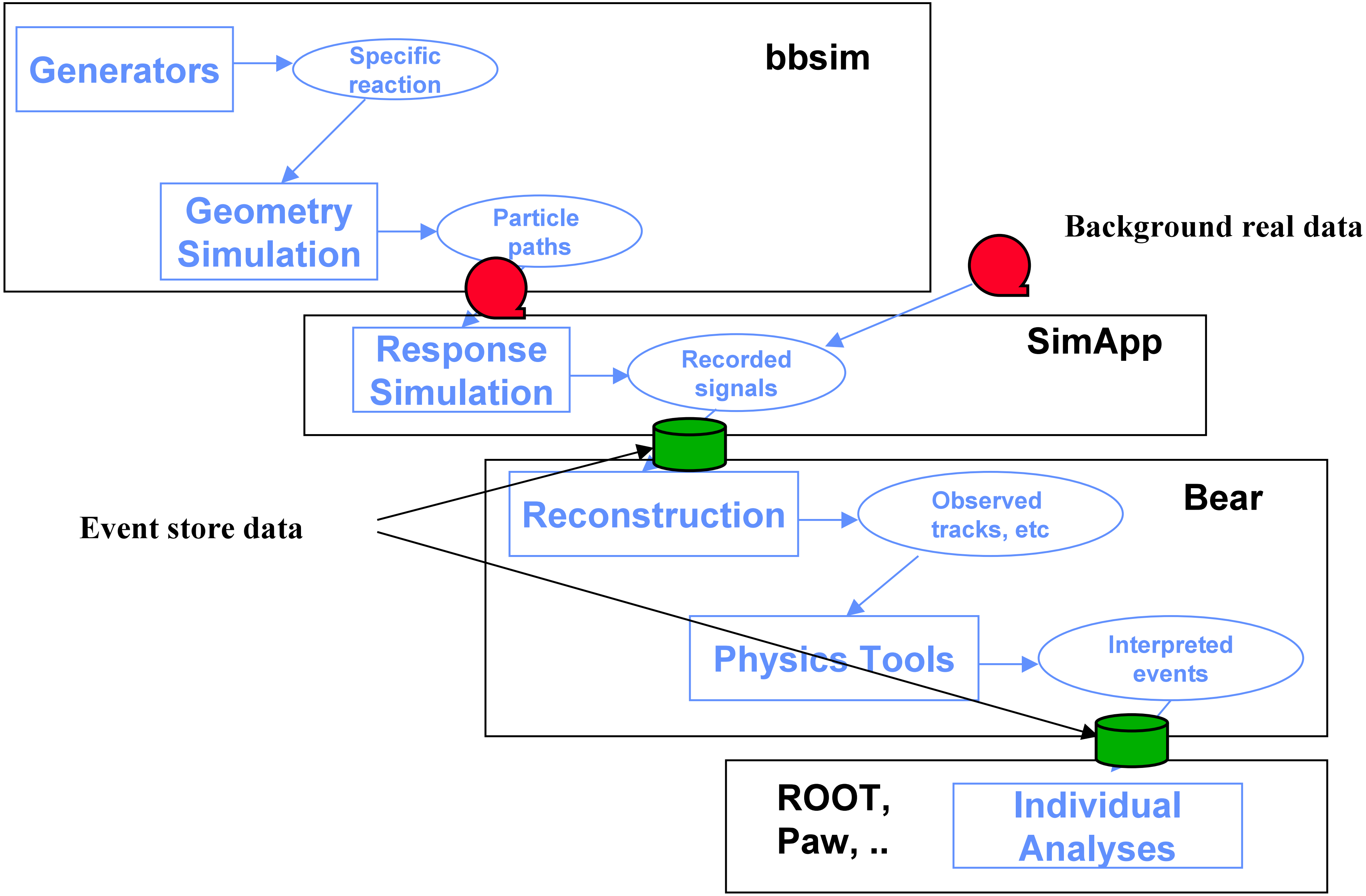
Studying "what if"?

- Both at detector and physics levels

Similar process happens in the reconstruction/analysis

- Better algorithms, studying new effects

Partitioning production system into programs



bbsim

Generators

Specific reaction

Geometry Simulation

Particle paths

Background real data

SimApp

Response Simulation

Recorded signals

Bear

Event store data

Reconstruction

Observed tracks, etc

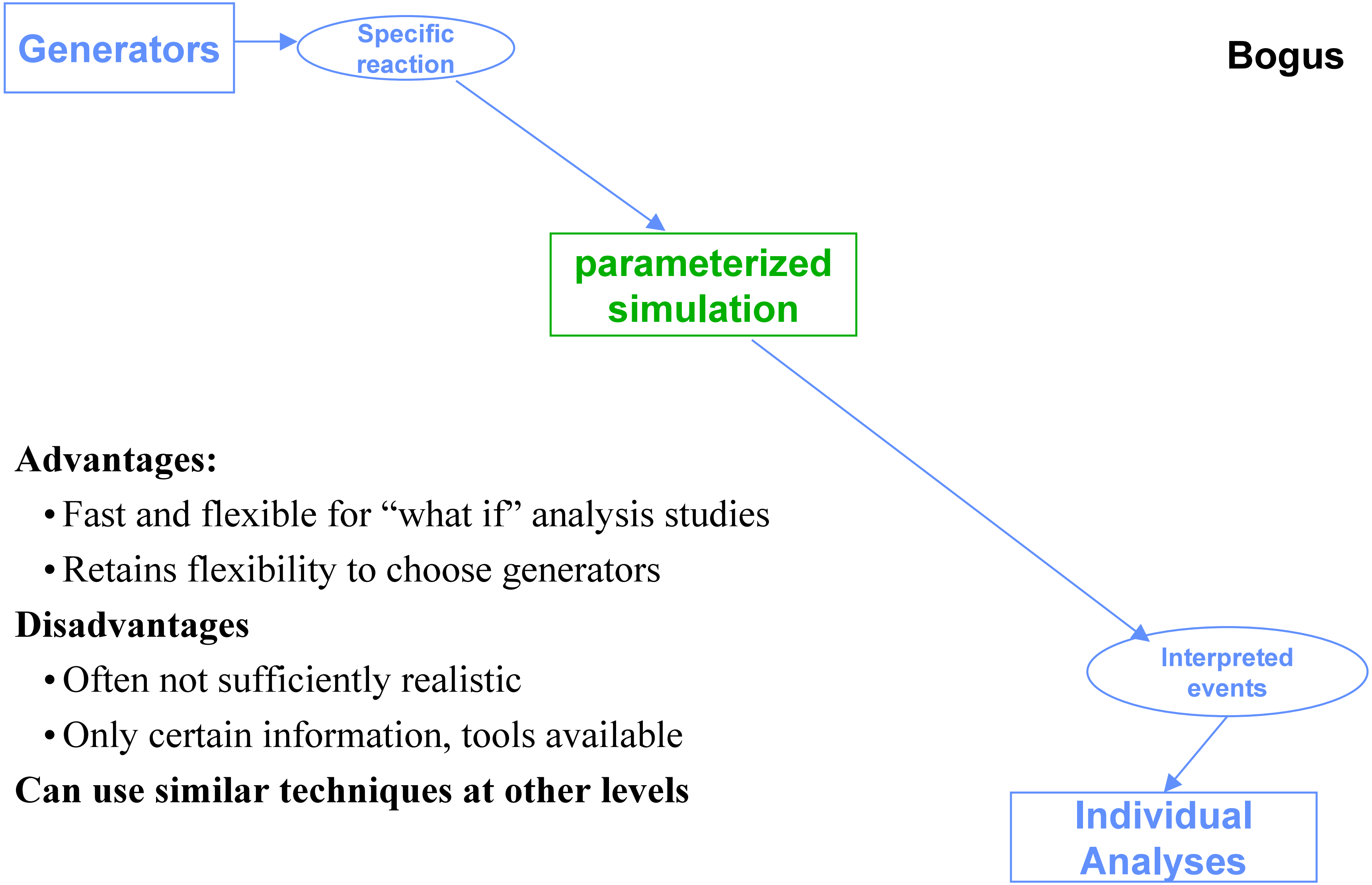
Physics Tools

Interpreted events

ROOT, Paw, ..

Individual Analyses

Speed, simplify simulation by crossing levels



Generators

Specific reaction

Bogus

parameterized simulation

Interpreted events

Individual Analyses

Advantages:

- Fast and flexible for “what if” analysis studies
- Retains flexibility to choose generators

Disadvantages

- Often not sufficiently realistic
- Only certain information, tools available

Can use similar techniques at other levels

Particle ID (PID)

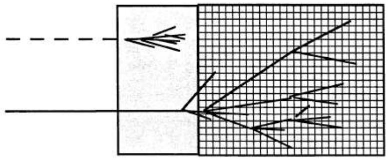
Track could be e, μ, π, K, or p; knowing which improves analysis

- Vital for measuring B->Kπ vs B->ππ rates
- Mistaking a π for e, μ, K or p increases combinatoric background

Leptons have unique interactions with material

- e deposits energy quickly, so expect E=p in calorimeter
- μ deposits energy slowly, so expect penetrating trajectory

But hadronic showers from π, K, p all look alike



Can't you measure mass from $m^2=E^2-p^2$?

For p=2GeV/c, pion energy = 2.005 GeV, kaon energy = 2.060 GeV

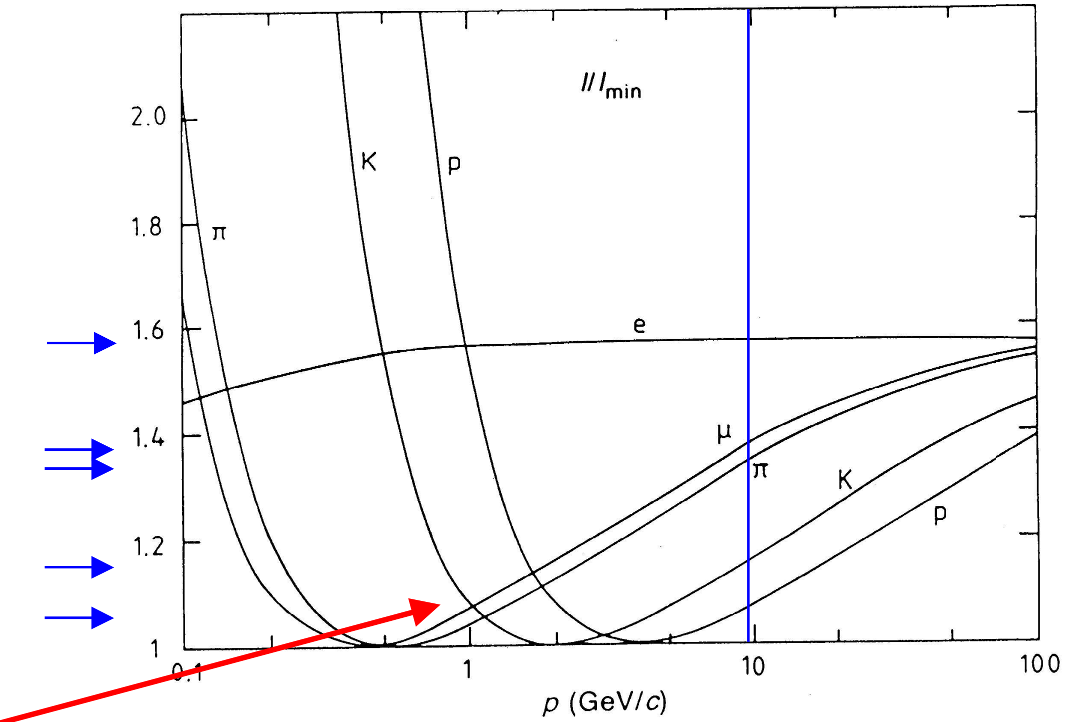
Calorimeters are not that accurate

(We usually cheat and calculate E from p and m)

dE/dx

Charged particles moving through matter lose energy to ionization

Loss is a function of the speed, $\beta \equiv \frac{v}{c}$ so a function of mass and momentum



With certain ambiguities!

Alternately, measuring $m = \frac{p}{\gamma\beta}$ lets us identify the particle type

Its hard to make this precise

Minimize material -> small loses

- Hard to measure dE well

Geometry of tracking is complex

- Hard to measure dx well

Typical accuracy is 5-10%

- "2 sigma separation"

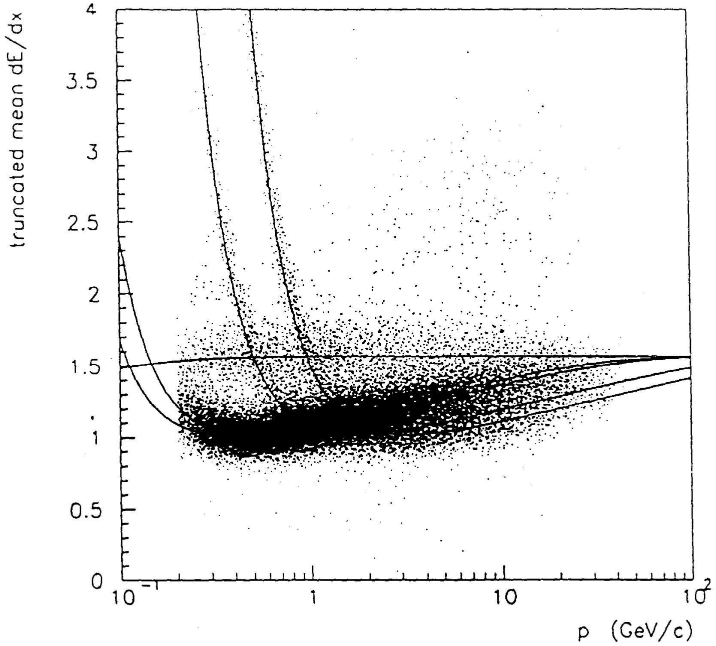


Fig. 8: Scatter plot of the ionisation measurement for a large set of hadronic Z_0 decays

During analysis, can choose

- **efficiency**
- **purity**

But can't have both!

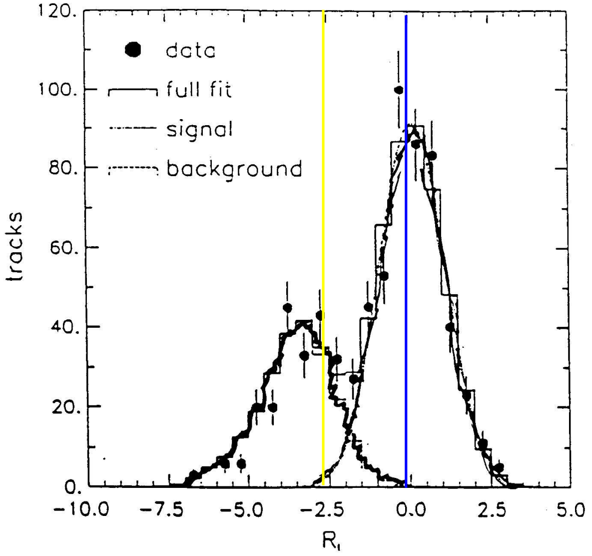
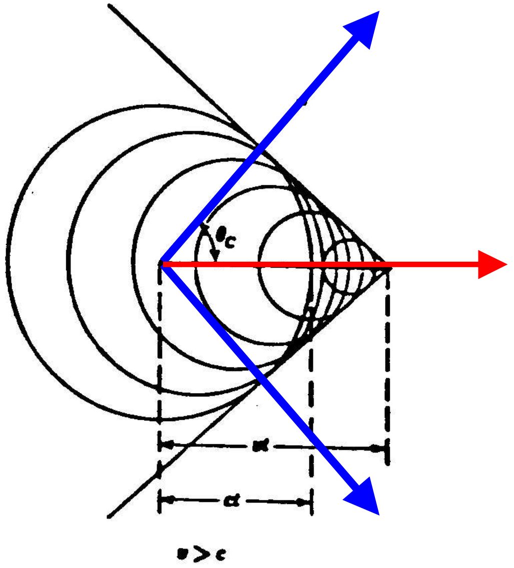


Fig. 10: Histogram of electron candidates using the dE/dx information of the TPC

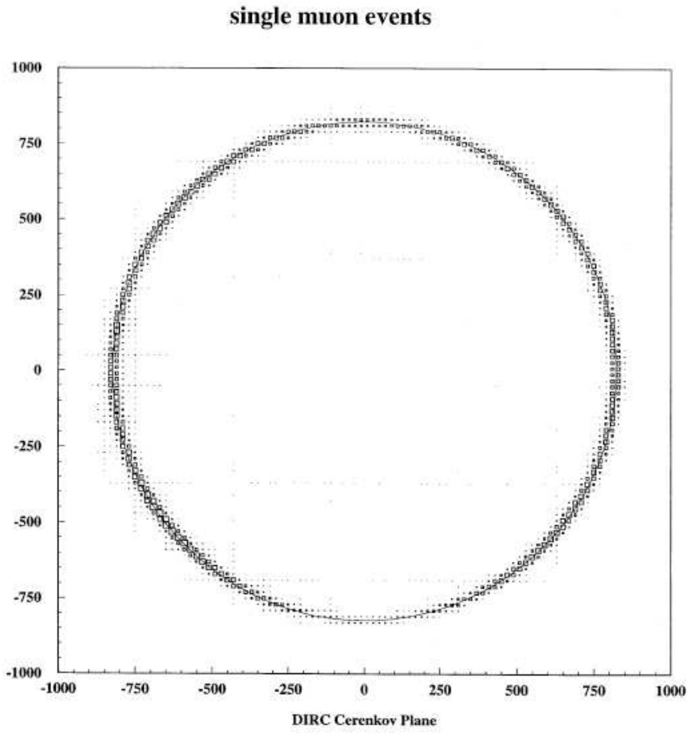
Another velocity-dependent process: Cherenkov light

Particles moving faster than light in a medium (glass, water) emit light

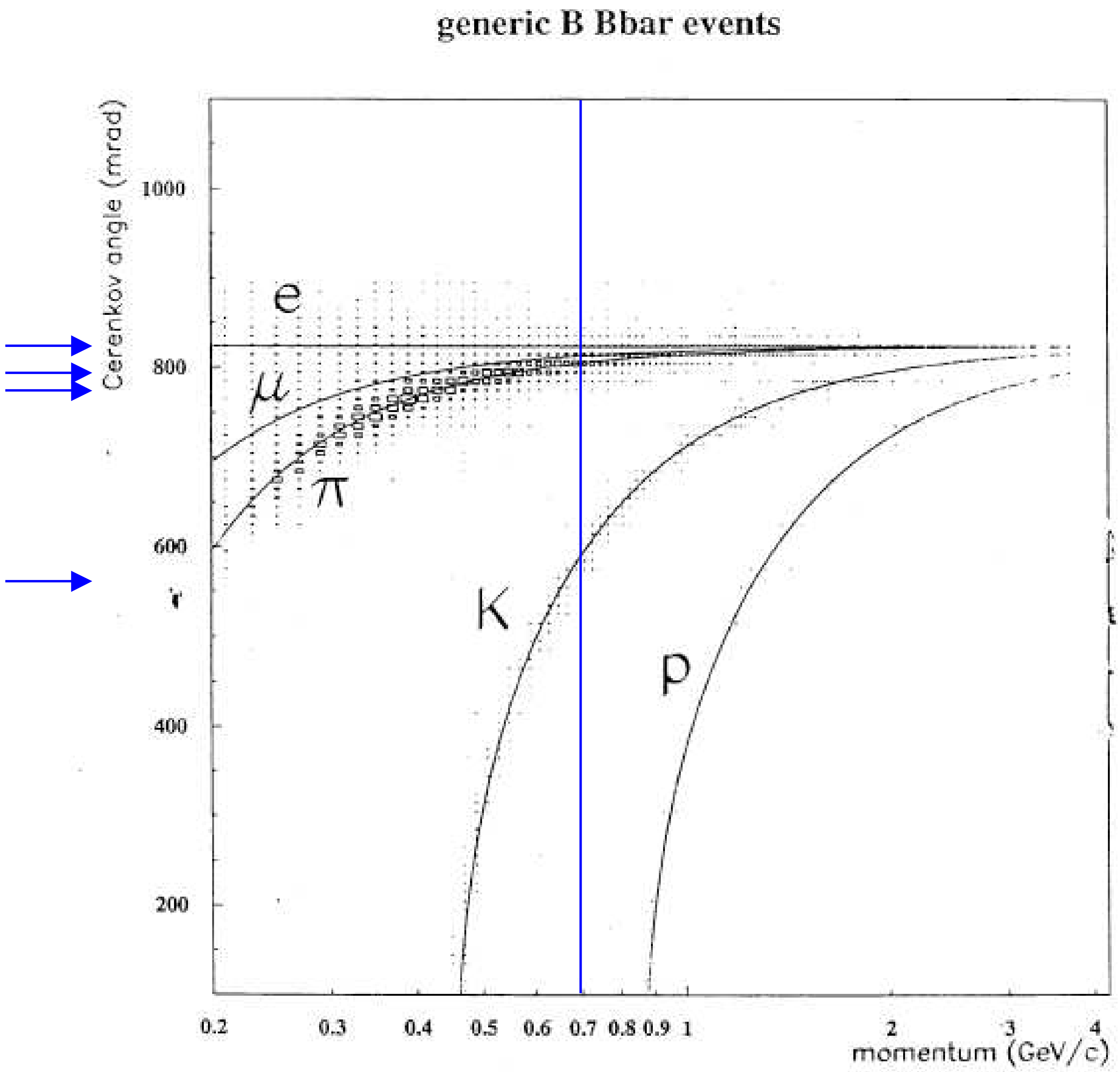
- Angle is related to velocity
- Light forms a cone



Focus it onto a plane, and you get a circle:



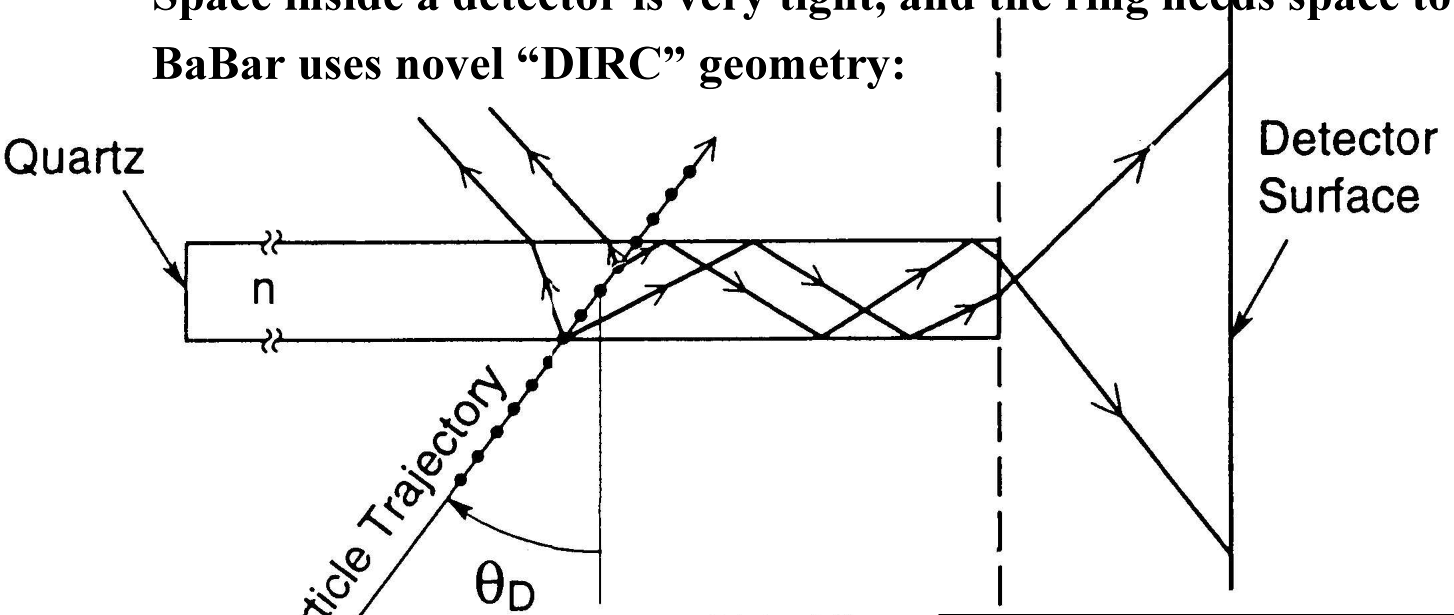
Radius of the reconstructed circle give particle type:

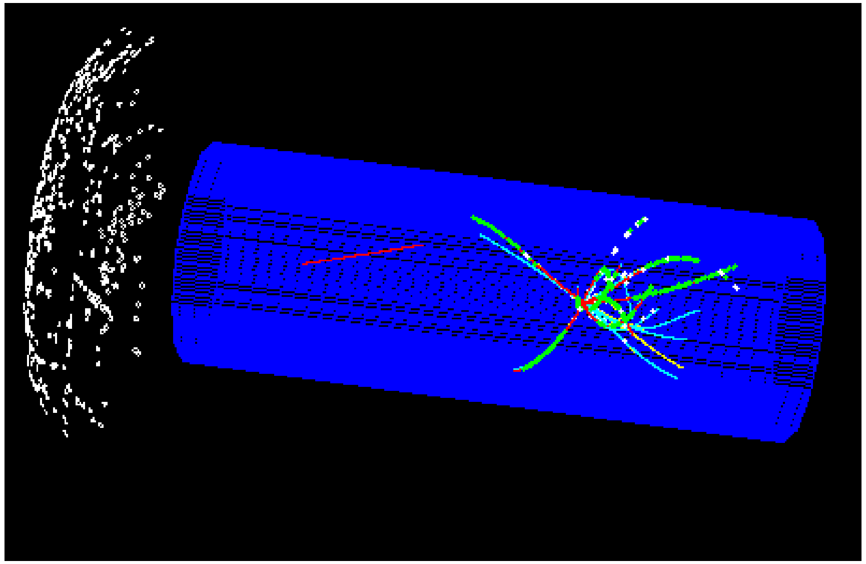


How to make this fit?

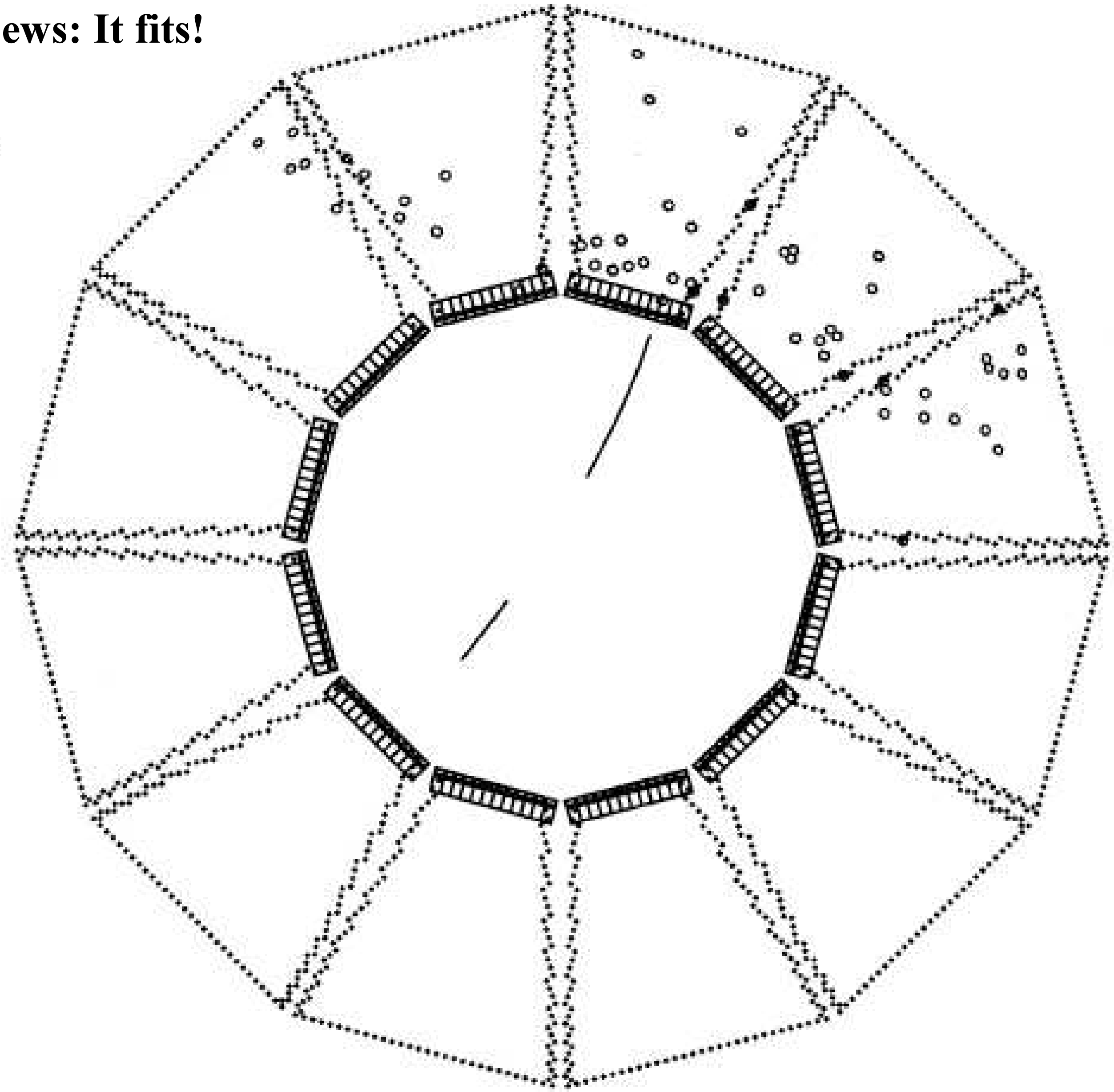
Space inside a detector is very tight, and the ring needs space to form

BaBar uses novel “DIRC” geometry:



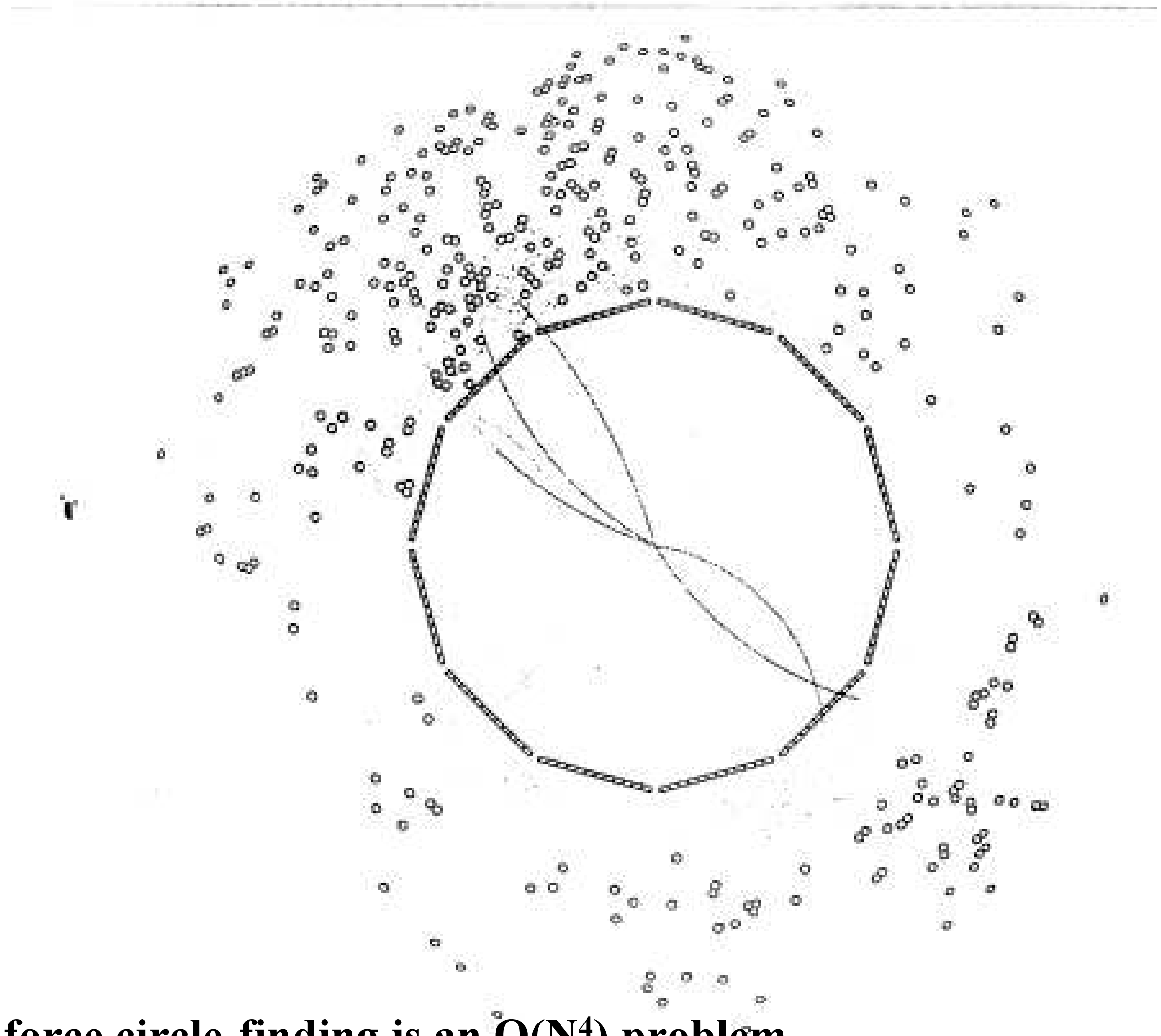


Good news: It fits!



Bad news: Rings get messy due to ambiguities in bouncing

Simple event with five charged particles:



Brute-force circle-finding is an $O(N^4)$ problem

Realistic solution?

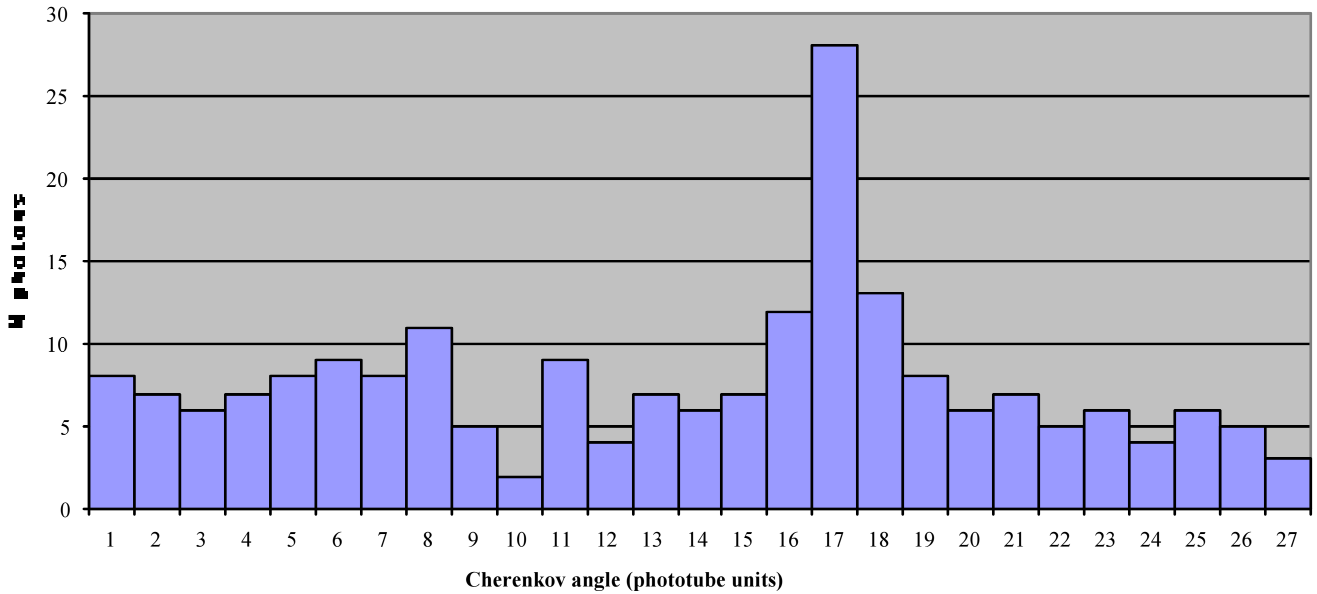
Use what you know:

- Have track trajectories, know position and angle in DIRC bars
- All photons from a single track will have the same angle w.r.t. track

  No reason to expect that for photons from other tracks

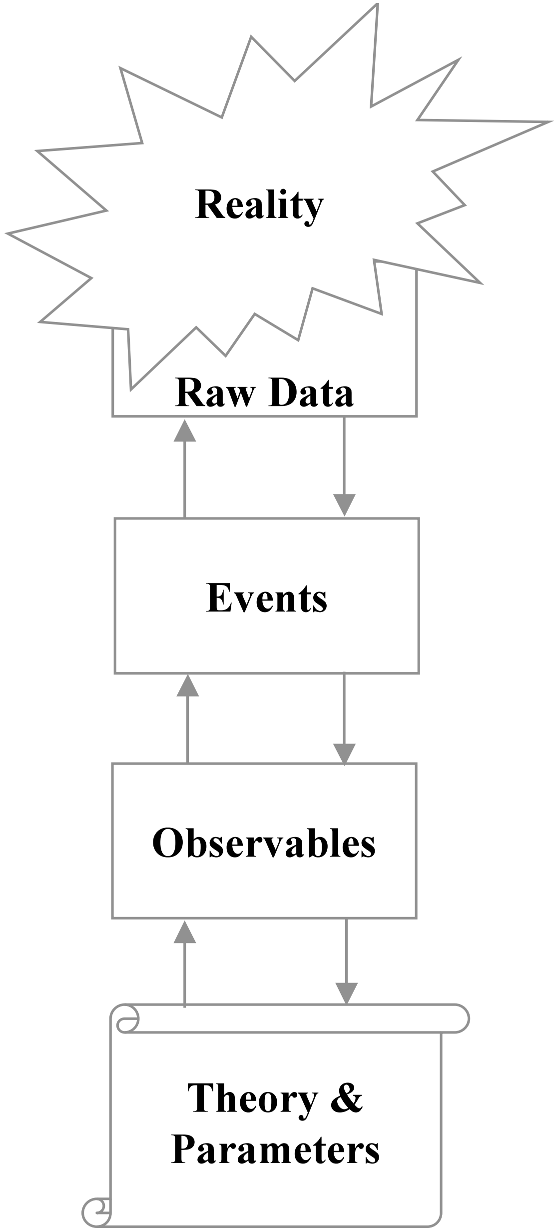
For each track, plot angle between track and <u>every</u> photon

- Don't do pattern recognition with individual photons
- Instead, look for overall pattern



Not perfect, but optimal?

Will do better as we understand more



Reality

Raw Data

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

Events

A unique happening:
Run 21007, event 3916 which
contains a J/psi -> ee decay

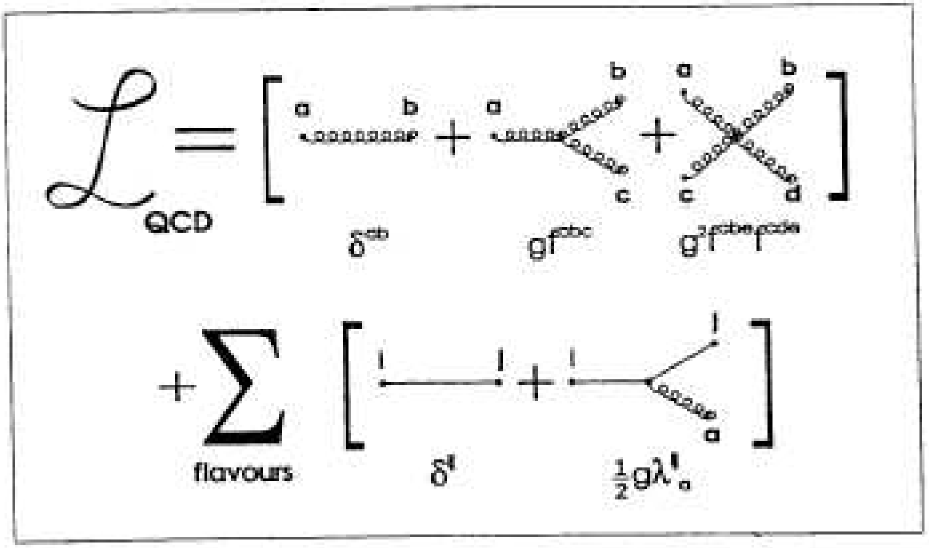
Observables

Specific lifetimes, probabilities, masses,
branching ratios, interactions, etc

**Theory &
Parameters**

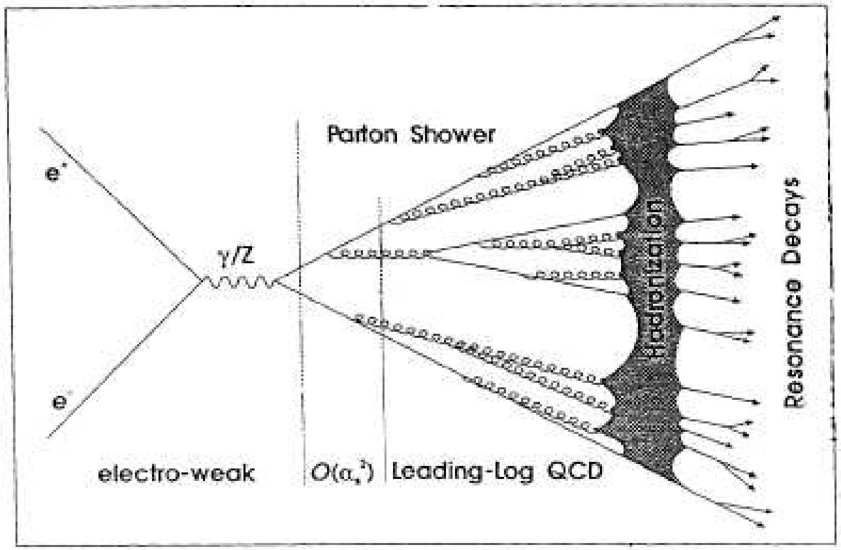
A small number of general equations, with specific
input parameters (perhaps poorly known)

Analysis: Measuring α_S in QCD



QCD predicts a set of basic interactions:

- You can measure the strong coupling constant by the relative rates



Unfortunately, QCD only makes exact predictions at high energy

- Low energy QCD, e.g. making hadrons, must be "modeled"

Compare models to observations in lots of different variables

Over time, new models get created and old ones improve

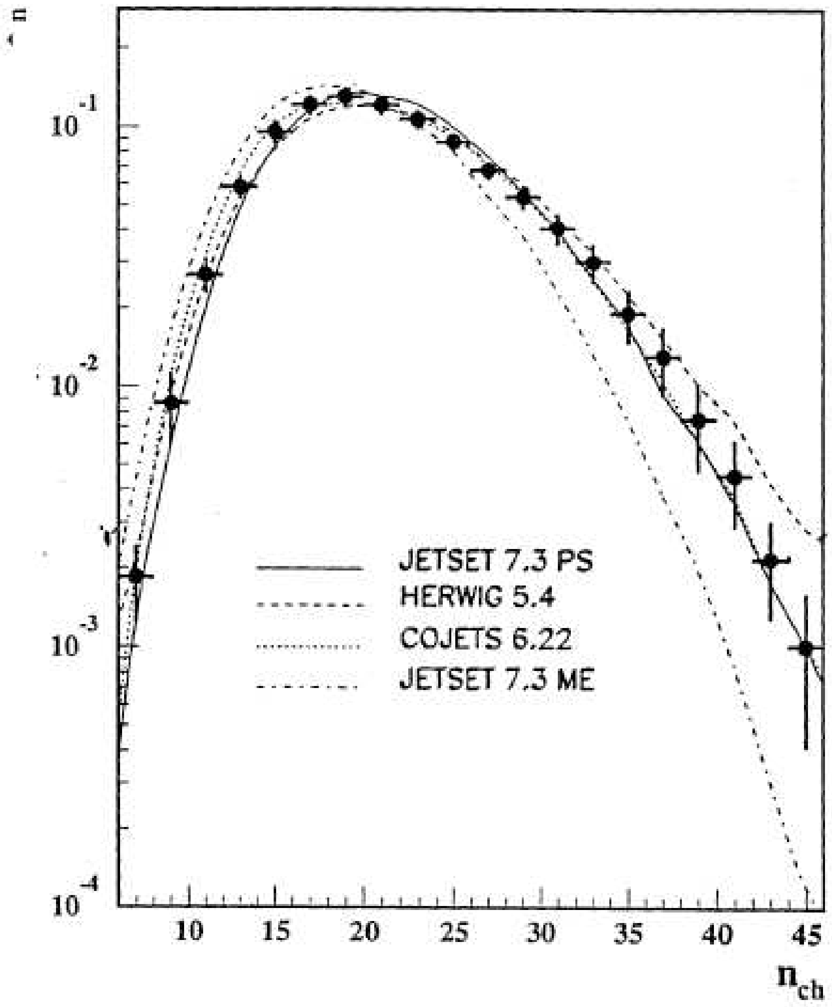
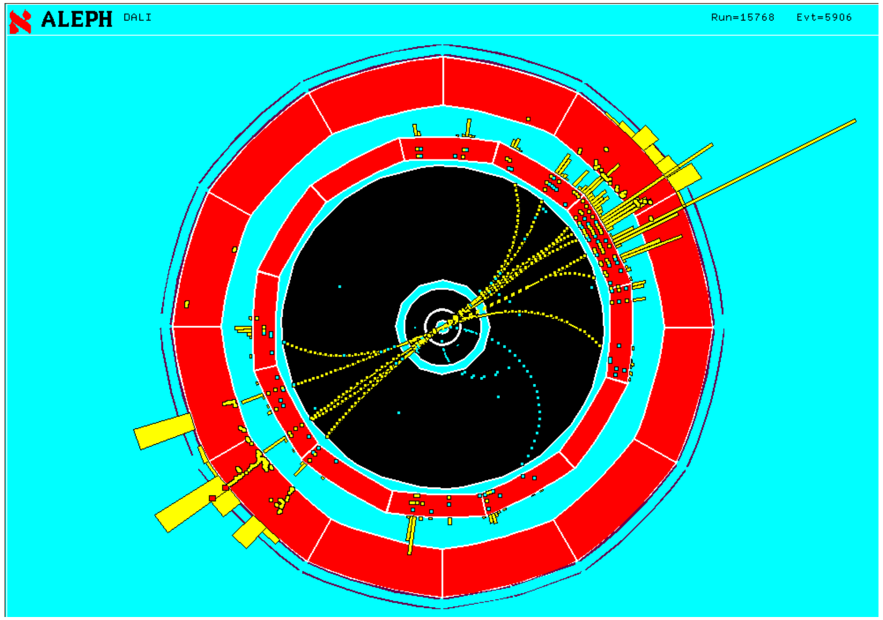
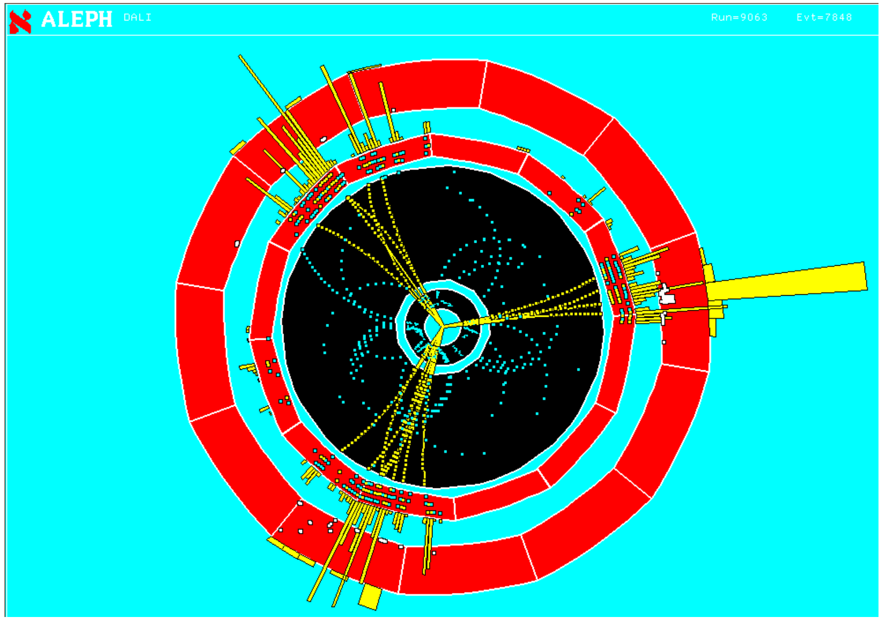


Figure 5: Charged multiplicity distribution measured by the L3 collaboration [28]. The points with error bars are the experimental data, the curves are model predictions.

“Jets”

Groups of particles probably come from the underlying quarks and gluons





But how to make this more quantitative?

- Don’t want people “guessing” at whether there are two or three jets
- Need a jet-finding algorithm

Simple one:

- Take two particles with most similar momentum and combine into one
- Repeat, until you reach a stopping value “y_{cut}”

What about that arbitrary cut?

Nature doesn’t know about it

- If your model is right, your simulation should reproduce the data at any value of the cut
- Pick one (e.g. 0.04), and use the number of 2,3,4, 5 jet events to determine α_S.
- Then check consistency at other values, with other models

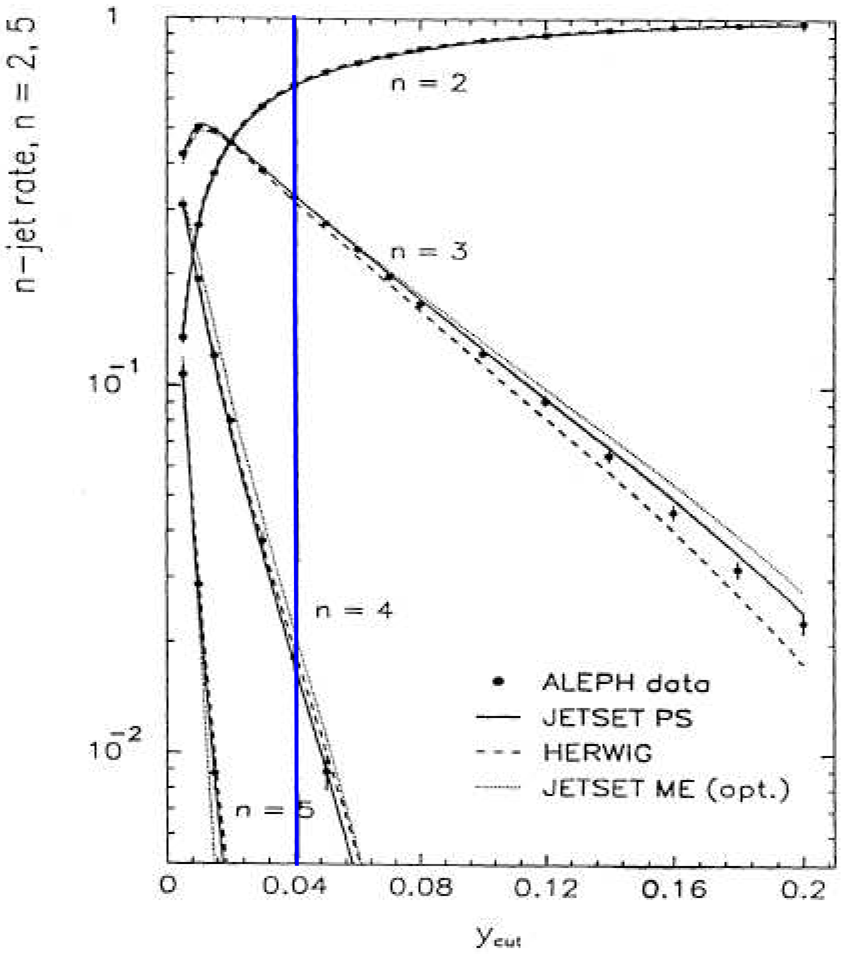


Figure 8: Jet rates determined by the ALEPH-collaboration [29] as function of the jet resolution parameter y_{cut}. The experimental results are compared to model calculations. Note that neighbouring points are highly correlated.

Many ways to measure α_S

If the theory's right, all get same value because all are measuring same thing

If the values are inconsistent, perhaps a more complicated theory is needed

Or maybe we just made a mistake...

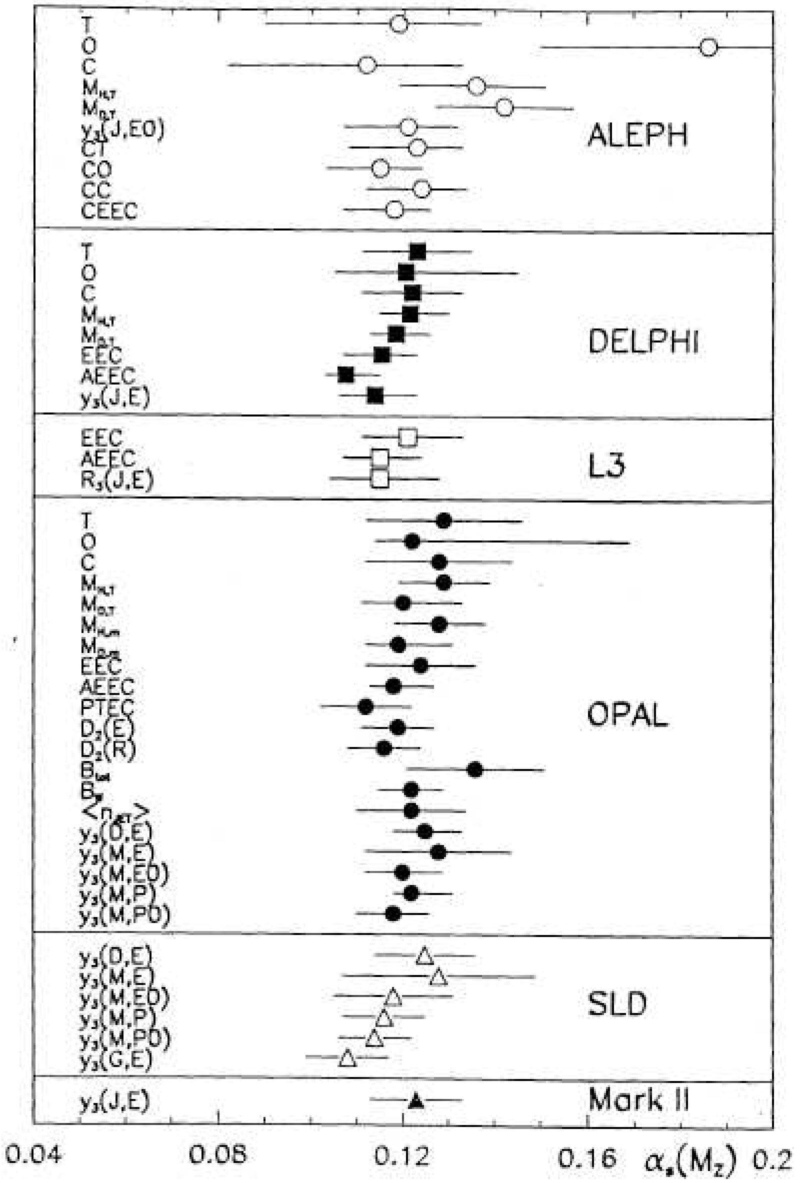


Figure 12: Measurements of the strong coupling constant from event shape variables based on second order QCD predictions.

Alignment & Calibration

How do you know the gain of each calorimeter cell?

- What's the relationship between ADC counts and energy?
- You designed it to have a specific value; does it?

How do you know where the tracking hits are in space?

- Need to know Si plane positions to about 5 microns

Start with

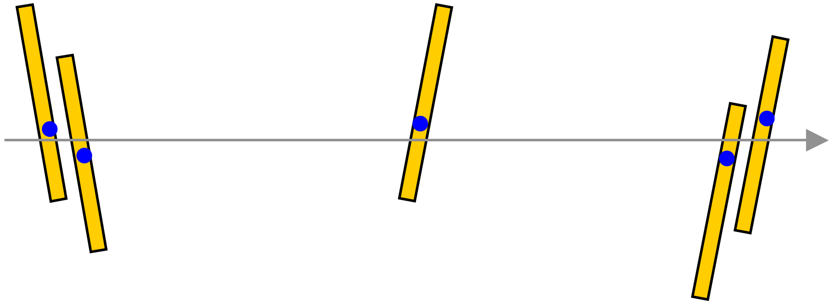
- Test beam information
- Surveys during construction
- Simulations and tests

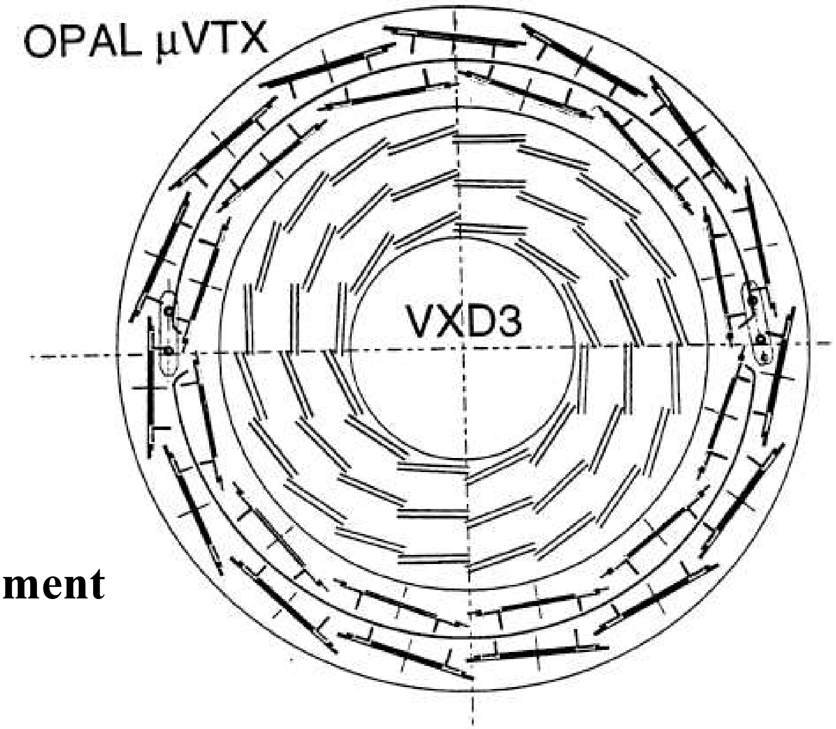
But it always comes down to calibrating/aligning with real data

Example: BaBar vertex detector alignment

About 700 Si wafers

- Each with 6 degrees of freedom
- => 4200 alignment constants to find





Small motions => small changes in alignment

=> change χ^2 of track

Approach 1: Take 10^5 tracks

Calculate sum of track χ^2s

For each of 4200 constants, generate equation from $$\frac{\partial \chi^2}{\partial c_i} = 0$$

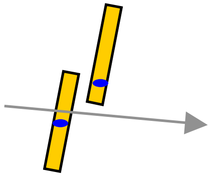
Solve 4200 equations in 4200 unknowns

Computationally infeasible

- Even worse, non-linear fit won't converge

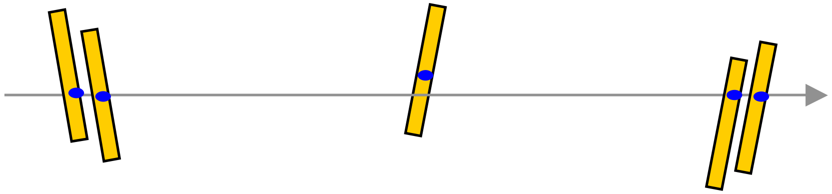
Instead, break problem into pieces:

- Two mechanical halves => 2x6 "global alignment constants"
- "local" constants within the halves



Do local alignment iteratively

- Look at pairs of adjacent wafers, and try to position them
- Then use tracks to position entire layers



- And iterate as needed

Iterative, sensitive process

- Manually guided from initial knowledge to final approximation
- Requires judgement on when to stop, how often to redo

Summary

Reconstruction and analysis is how we get from raw data to physics papers

Throughout, you deal with:

- Too little information
- Too much detail
- Little prior knowledge

You have to count on

- Lots of cross checks
- Prior art
- Tuning and evolutionary improvement

But you can generate wonderful results from these instruments!