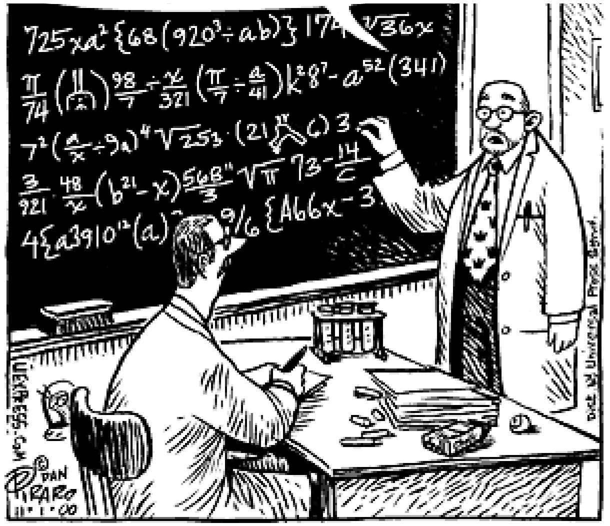
From Raw Data to Physics: Reconstruction and Analysis

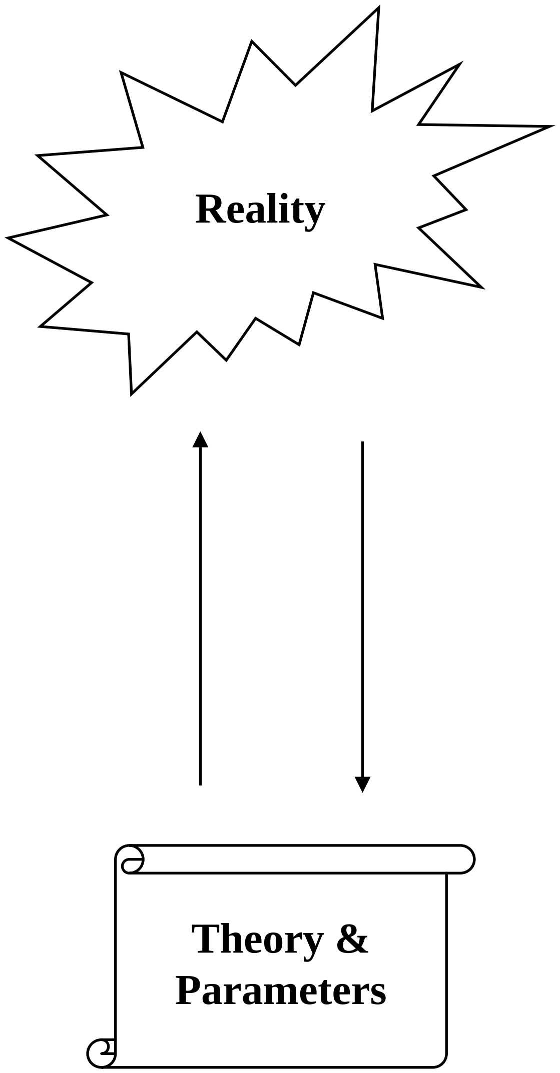
Introduction

Sample Cases

A Model

Basic Features





We use experiments to inquire about what "reality" does.

We intend to fill this gap

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

Theory

10. ELECTROWEAK MODEL AND CONSTRAINTS ON NEW PHYSICS

Revised August 1999 by J. Erler and P. Langacker (Univ. of Pennsylvania).

10.1. Introduction

The standard electroweak model is based on the gauge group [1] SU(2) × U(1), with gauge bosons W_μ^i, $i = 1, 2, 3$, and B_μ for the SU(2) and U(1) factors, respectively, and the corresponding gauge coupling constants g and g'. The left-handed fermion fields $\psi_i = \begin{pmatrix} \nu_i \\ \ell_i^- \end{pmatrix}$ and $\begin{pmatrix} u_i \\ d_i' \end{pmatrix}$ of the i^{th} fermion family transform as doublets under SU(2), where $d_i' \equiv \sum_j V_{ij}\, d_j$, and V is the Cabibbo-Kobayashi-Maskawa mixing matrix. (Constraints on V are discussed in the section on the Cabibbo-Kobayashi-Maskawa mixing matrix.) The right-handed fields are SU(2) singlets. In the minimal model there are three fermion families and a single complex Higgs doublet $\phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$

After spontaneous symmetry breaking the Lagrangian for the fermion fields is

$$\begin{aligned}
\mathscr{L}_F = & \sum_i \overline{\psi}_i \left(i\, \partial\!\!\!/ - m_i - \frac{g m_i H}{2 M_W} \right) \psi_i \\
& - \frac{g}{2\sqrt{2}} \sum_i \overline{\psi}_i\, \gamma^\mu\, (1 - \gamma^5)(T^+\, W_\mu^+ + T^-\, W_\mu^-)\, \psi_i \\
& - e \sum_i q_i\, \overline{\psi}_i\, \gamma^\mu\, \psi_i\, A_\mu \\
& - \frac{g}{2\cos\theta_W} \sum_i \overline{\psi}_i\, \gamma^\mu (g_V^i - g_A^i \gamma^5)\, \psi_i\, Z_\mu\, . \qquad (10.1)
\end{aligned}$$

$\theta_W \equiv \tan^{-1}(g'/g)$ is the weak angle; $e = g \sin\theta_W$ is the positron electric charge; and $A \equiv B\cos\theta_W + W^3 \sin\theta_W$ is the (massless) photon field. $W^\pm \equiv (W^1 \mp i W^2)/\sqrt{2}$ and $Z \equiv -B\sin\theta_W + W^3\cos\theta_W$ are the massive charged and neutral weak boson fields, respectively. T^+ and T^- are the weak isospin raising and lowering operators. The

Particle Data Group, Barnett et al

"Clear statement of how the world works"

Additional term goes here

Experiment

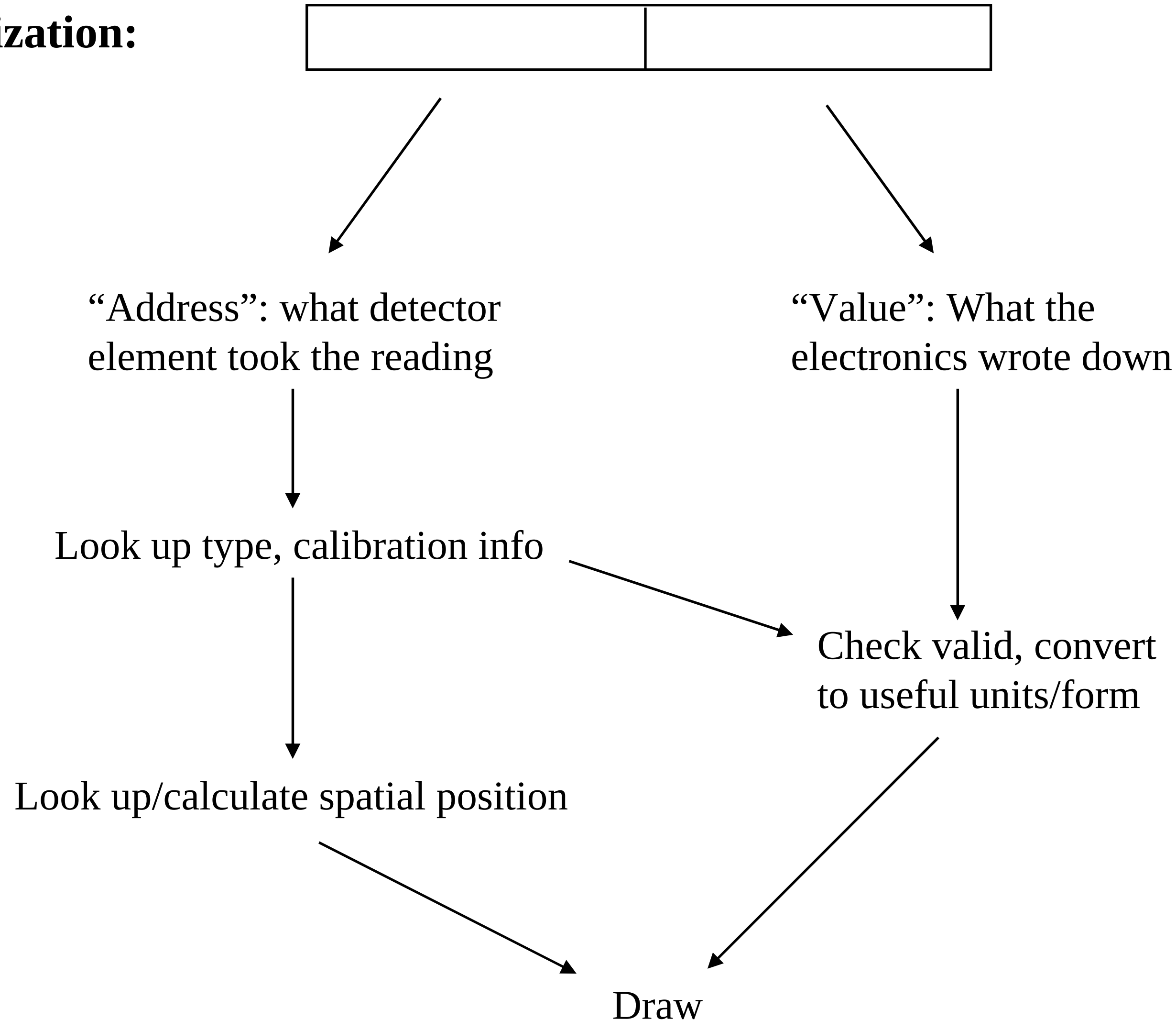
```
0x01e84c10:  0x01e8 0x8848 0x01e8 0x83d8 0x6c73 0x6f72 0x7400 0x0000
0x01e84c20:  0x0000 0x0019 0x0000 0x0000 0x01e8 0x4d08 0x01e8 0x5b7c
0x01e84c30:  0x01e8 0x87e8 0x01e8 0x8458 0x7061 0x636b 0x6167 0x6500
0x01e84c40:  0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84c50:  0x01e8 0x8788 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
```

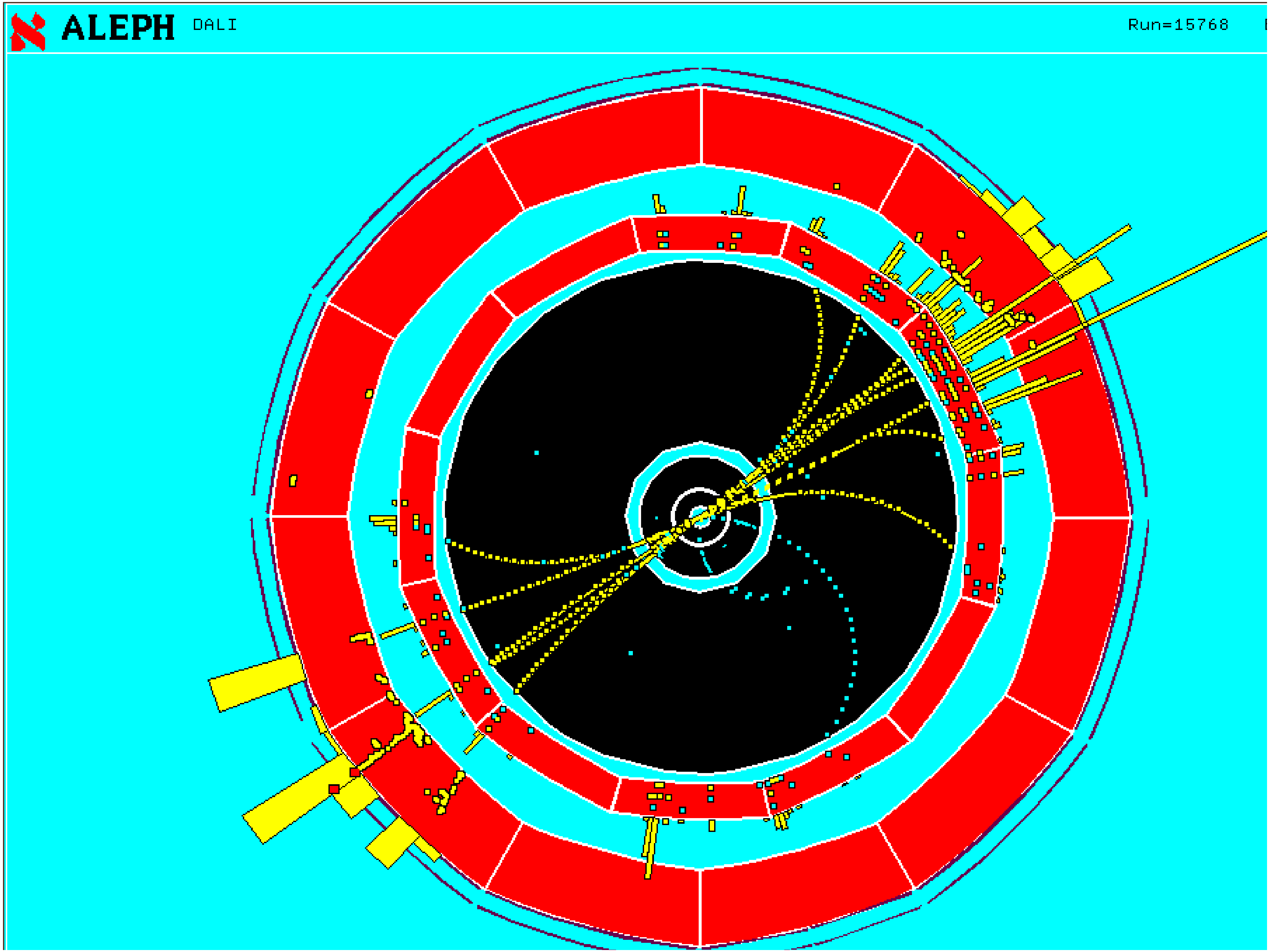
1/30th of an event in the BaBar detector

- Get about 100 events/second

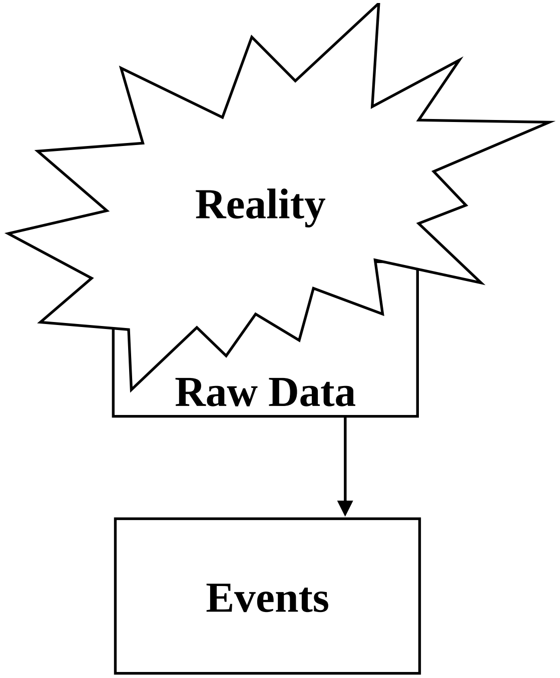
What does the data mean?

Digitization:



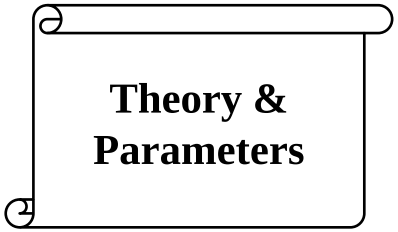


ALEPH
DALI
Run=15768



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

A unique happening:
Run 21007, event 3916 which contains a Z -> xx decay



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

Phenomenology

A good theory contains very few numbers

But it can predict a large number of reactions

Getting those predictions from the theory is called “phenomenology”

10.4. *W* and *Z* decays

The partial decay width for gauge bosons to decay into massless fermions $f_1\overline{f}_2$ is

$$\Gamma(W^+ \to e^+\nu_e) = \frac{G_F M_W^3}{6\sqrt{2}\pi} \approx 226.5 \pm 0.3 \text{ MeV} \quad , \qquad (10.41a)$$

$$\Gamma(W^+ \to u_i\overline{d}_j) = \frac{C G_F M_W^3}{6\sqrt{2}\pi} |V_{ij}|^2 \approx (707 \pm 1)|V_{ij}|^2 \text{ MeV} \quad , \quad (10.41b)$$

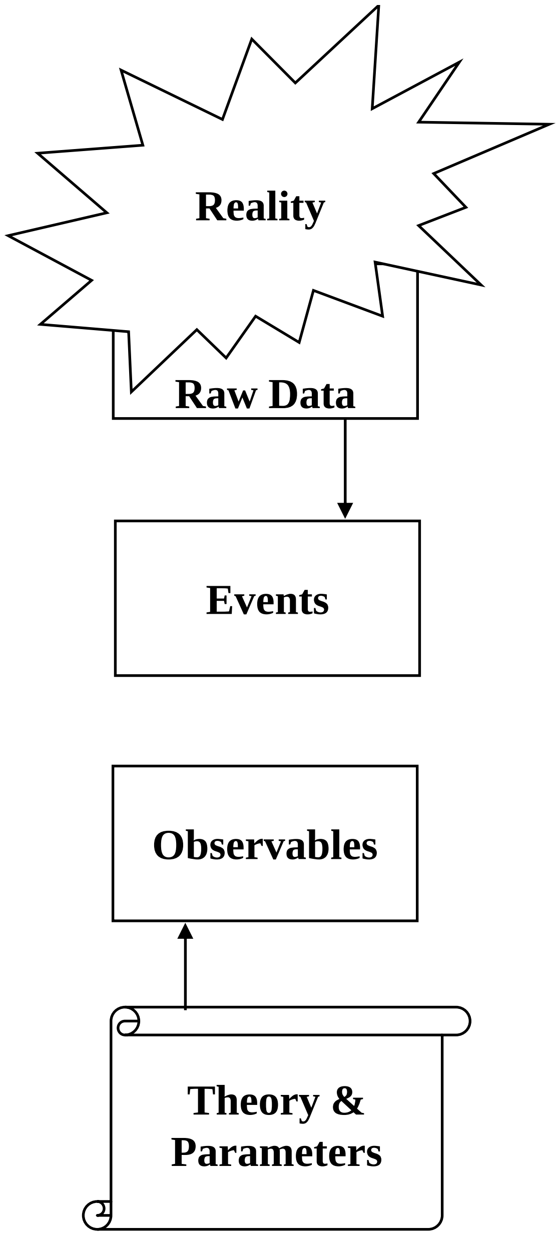
$$\Gamma(Z \to \psi_i\overline{\psi}_i) = \frac{C G_F M_Z^3}{6\sqrt{2}\pi}\left[g_V^{i2} + g_A^{i2}\right] \qquad (10.41c)$$

$$\approx \begin{cases} 300.3 \pm 0.2 \text{ MeV } (u\overline{u}), & 167.24 \pm 0.08 \text{ MeV } (\nu\overline{\nu}), \\ 383.1 \pm 0.2 \text{ MeV } (d\overline{d}), & 84.01 \pm 0.05 \text{ MeV } (e^+e^-), \\ 375.9 \mp 0.1 \text{ MeV } (b\overline{b}). \end{cases}$$

From Particle Data Book

Our modified theory predicts a different rate for Z->μμ

- This gives us a way to prove or disprove it!



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

A unique happening:
Run 21007, event 3916 which contains a Z -> xx decay

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

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

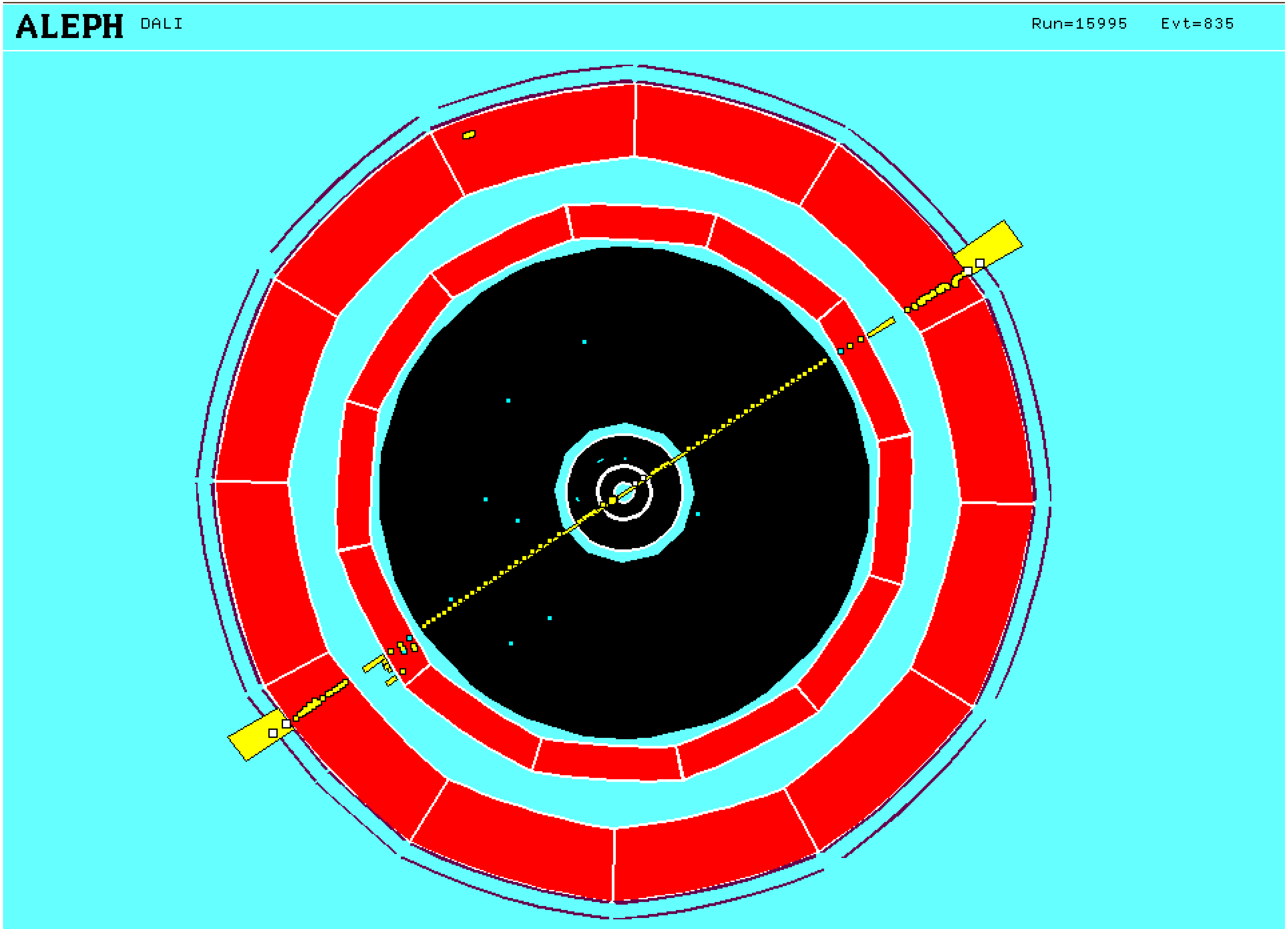
A simple analysis: What's BR(Z->μ+μ-)?

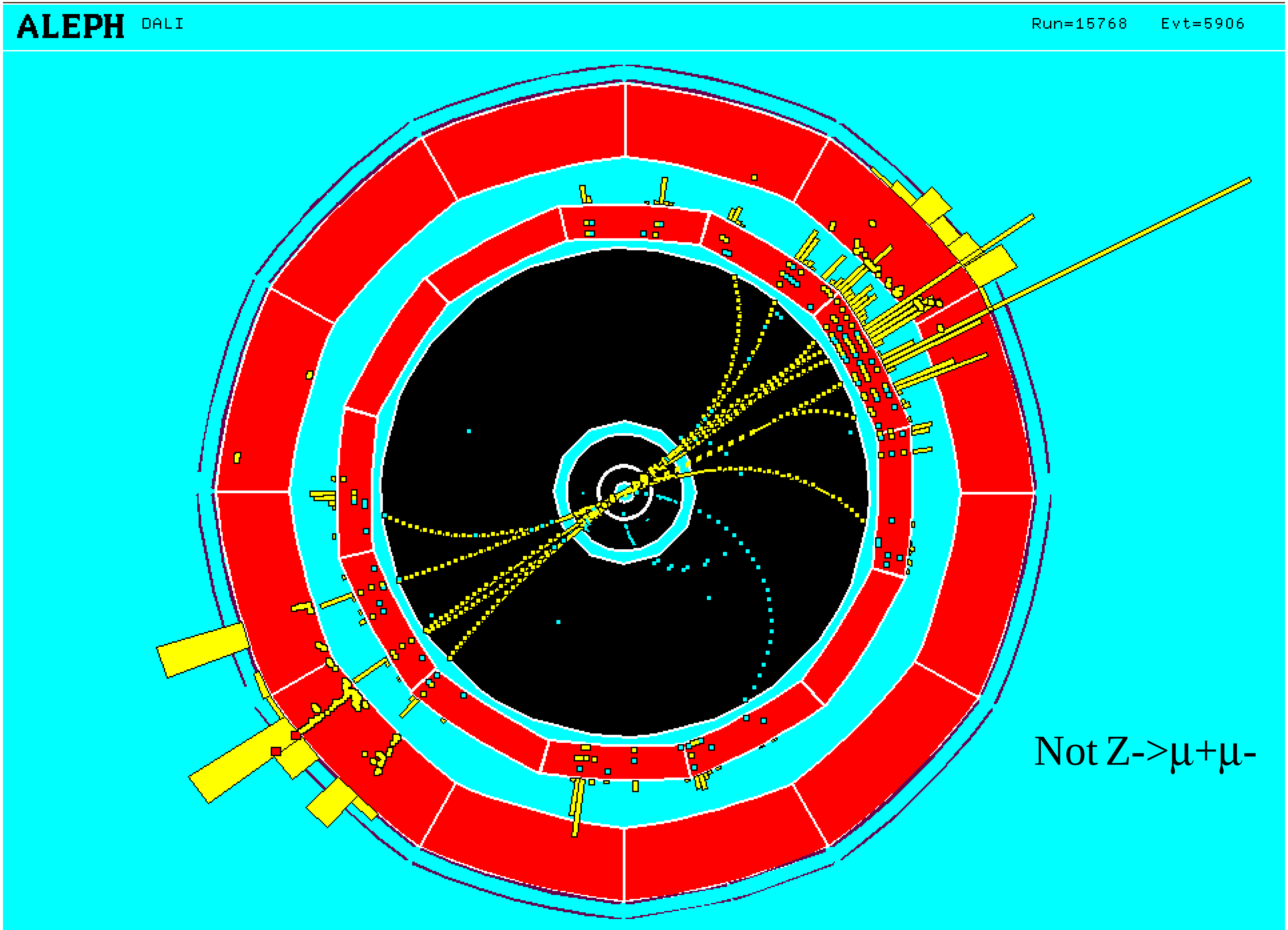
Measure:

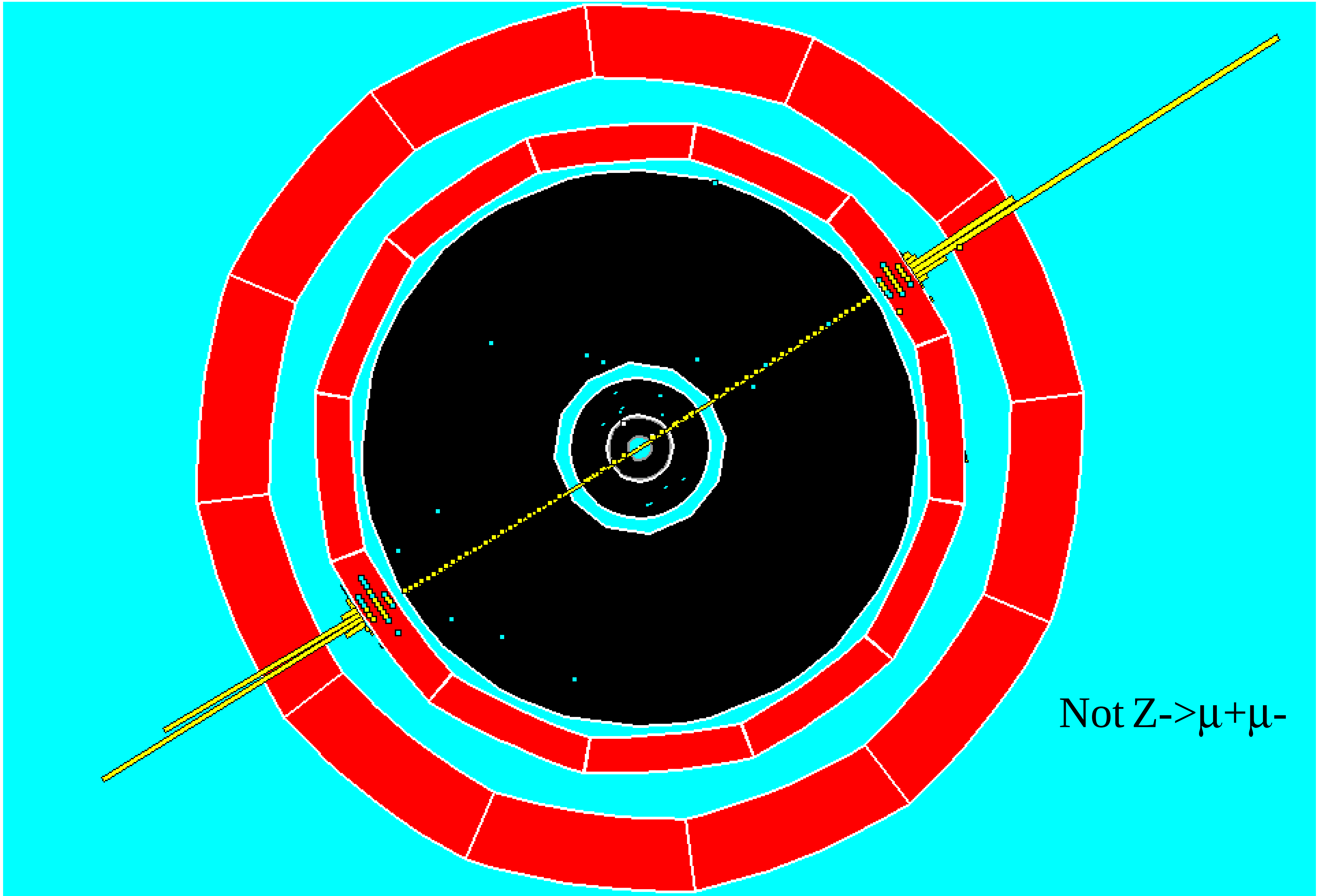
$$BR\left(Z^0 \rightarrow \mu^+\mu^-\right) = \frac{\text{Number of } \mu^+\mu^- \text{ events}}{\text{Total number of events}}$$

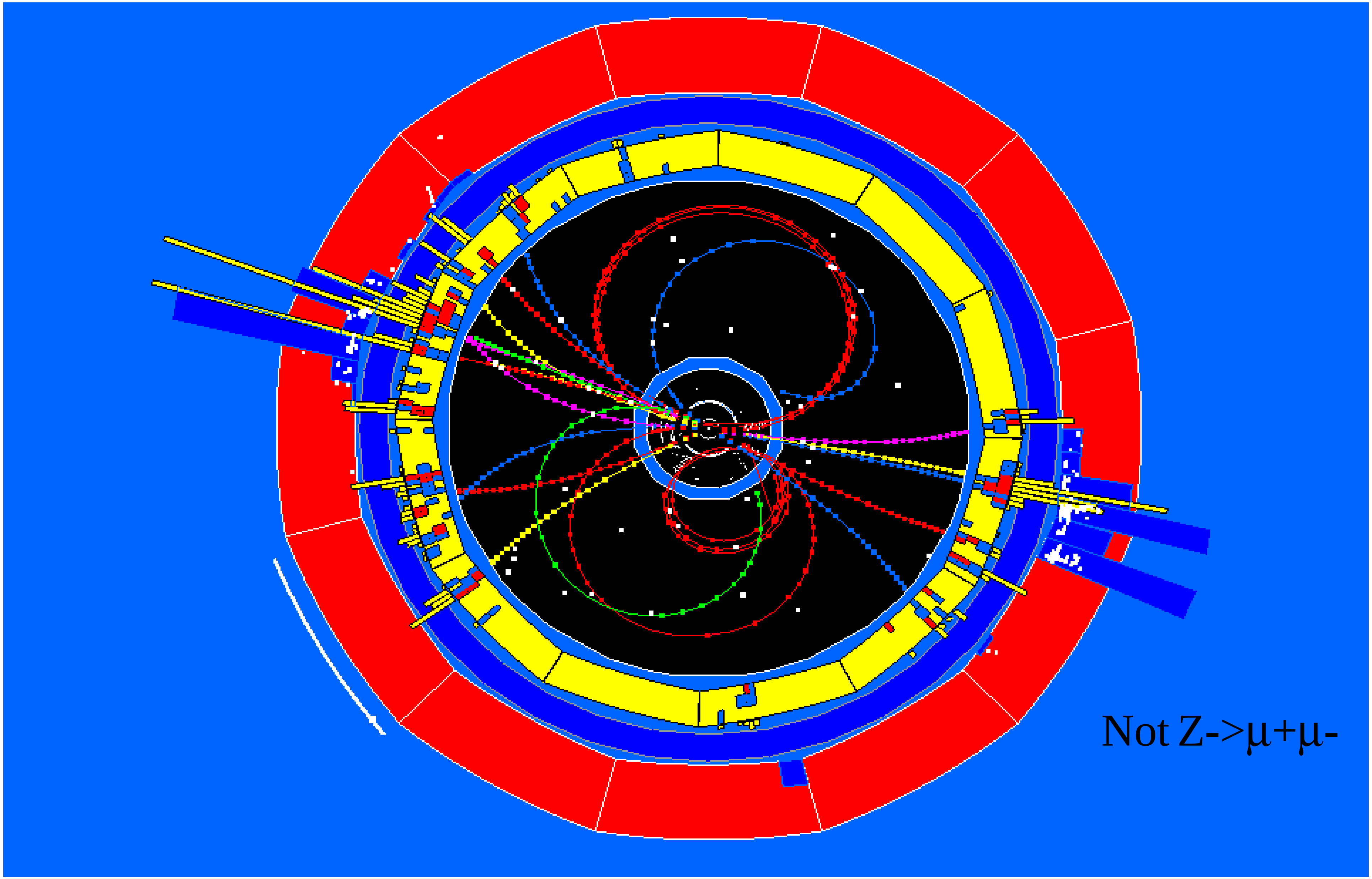
Take a sample of events, and count those with a $\mu^+\mu^-$ final state.

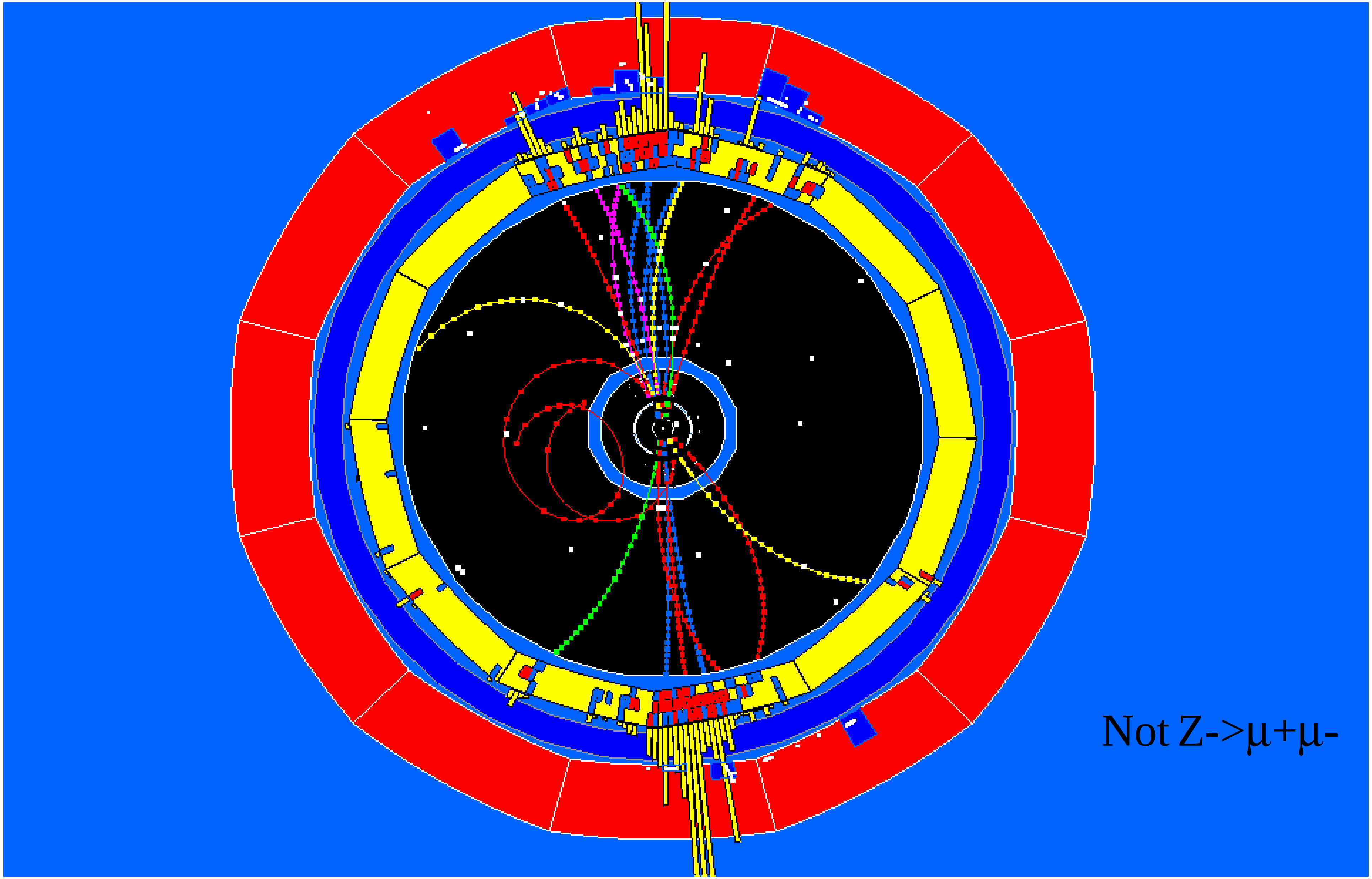
- Two tracks, approximately back-to-back with the expected |p|
  - Empirically, other kinds of events have more tracks
- Right number of muon hits in outer layers
  - Muons are very penetrating, travel through entire detector
- Expected energy in calorimeter
  - Electrons will deposit most of their energy early in the calorimeter; muons leave little

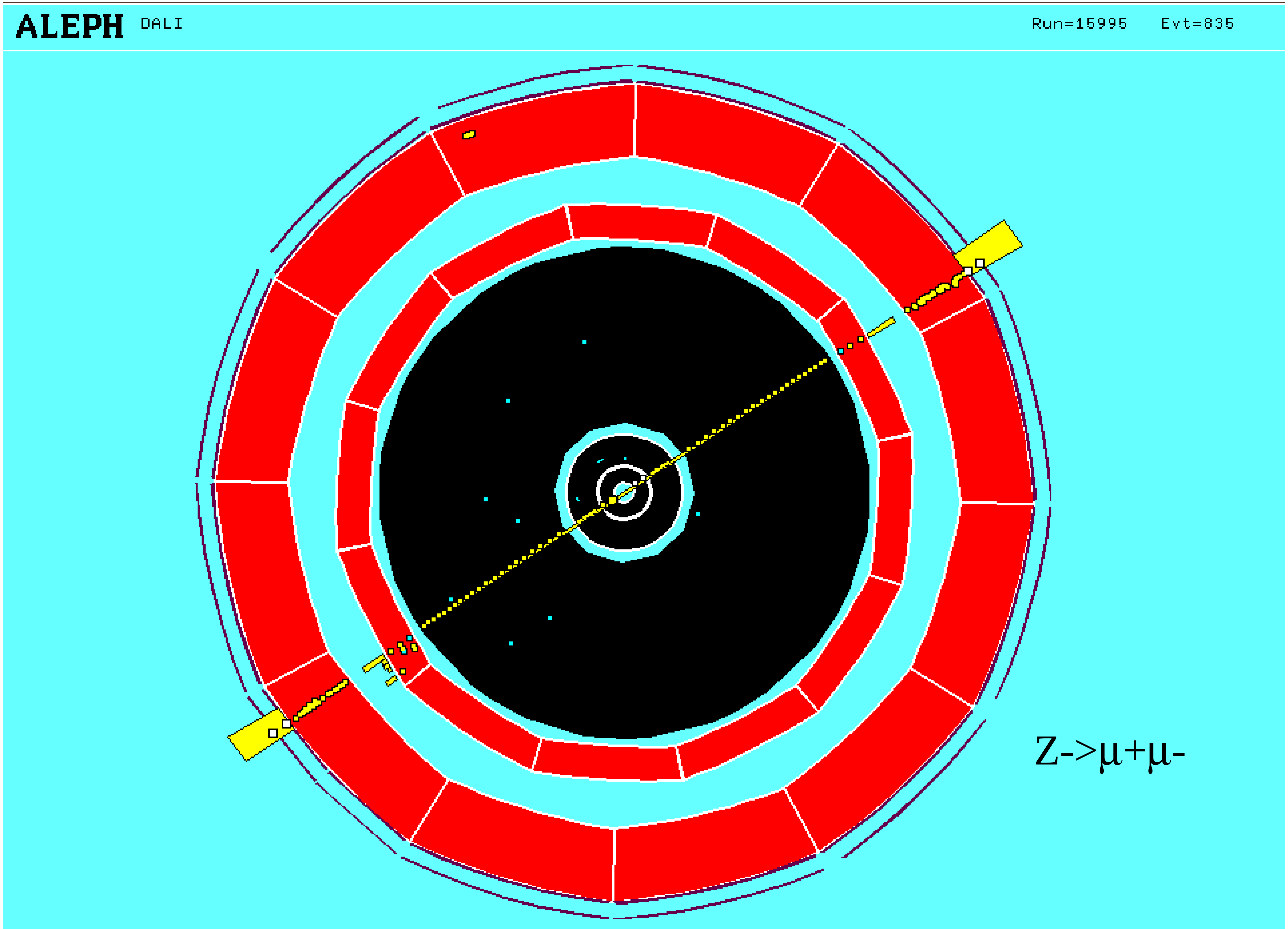

ALEPH
DALI
Run=15995
Evt=835

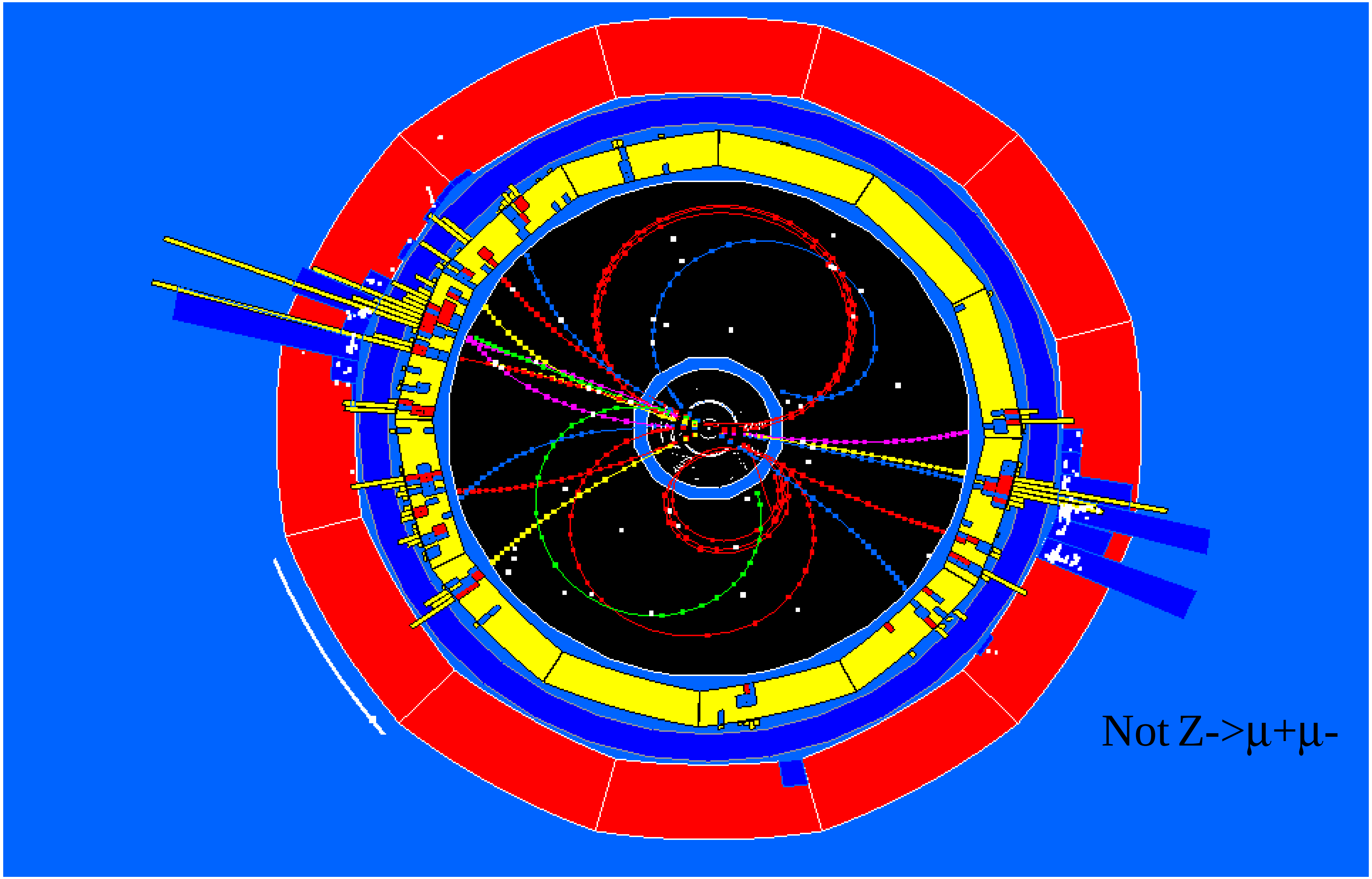

ALEPH
DALI
Run=15768 Evt=5906
Not Z->μ+μ-

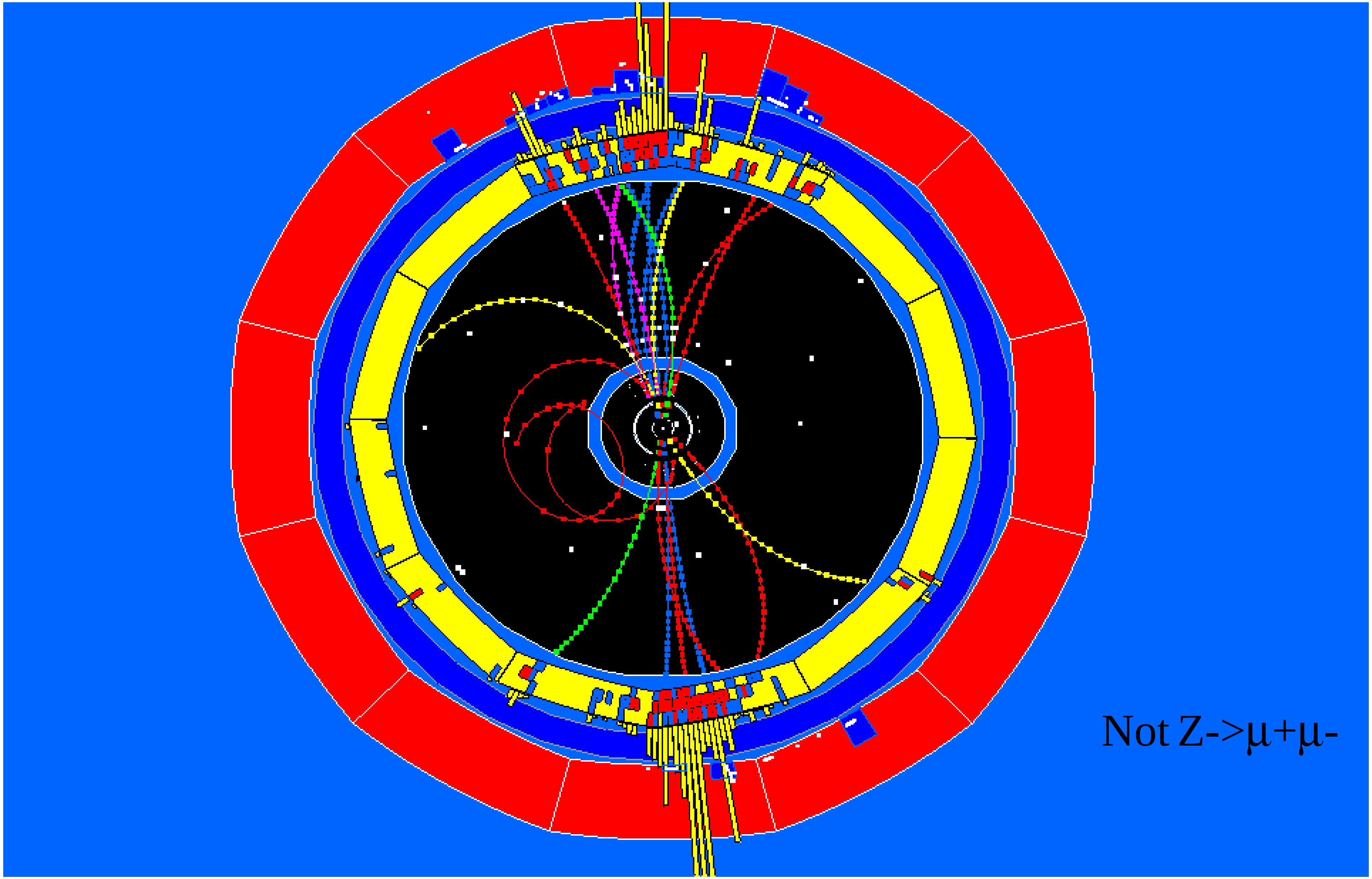

Not Z->μ+μ-

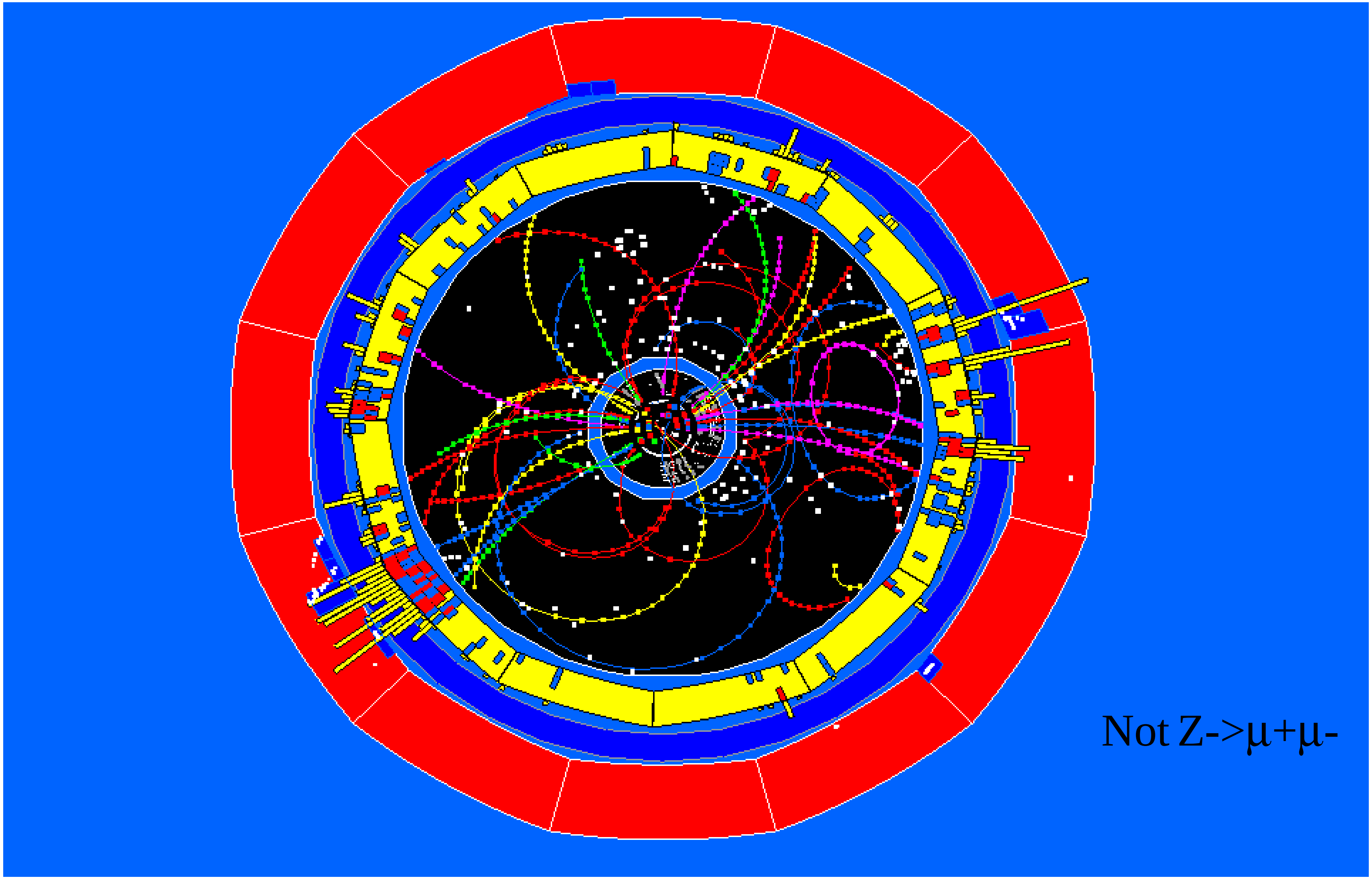

Not Z->μ+μ-

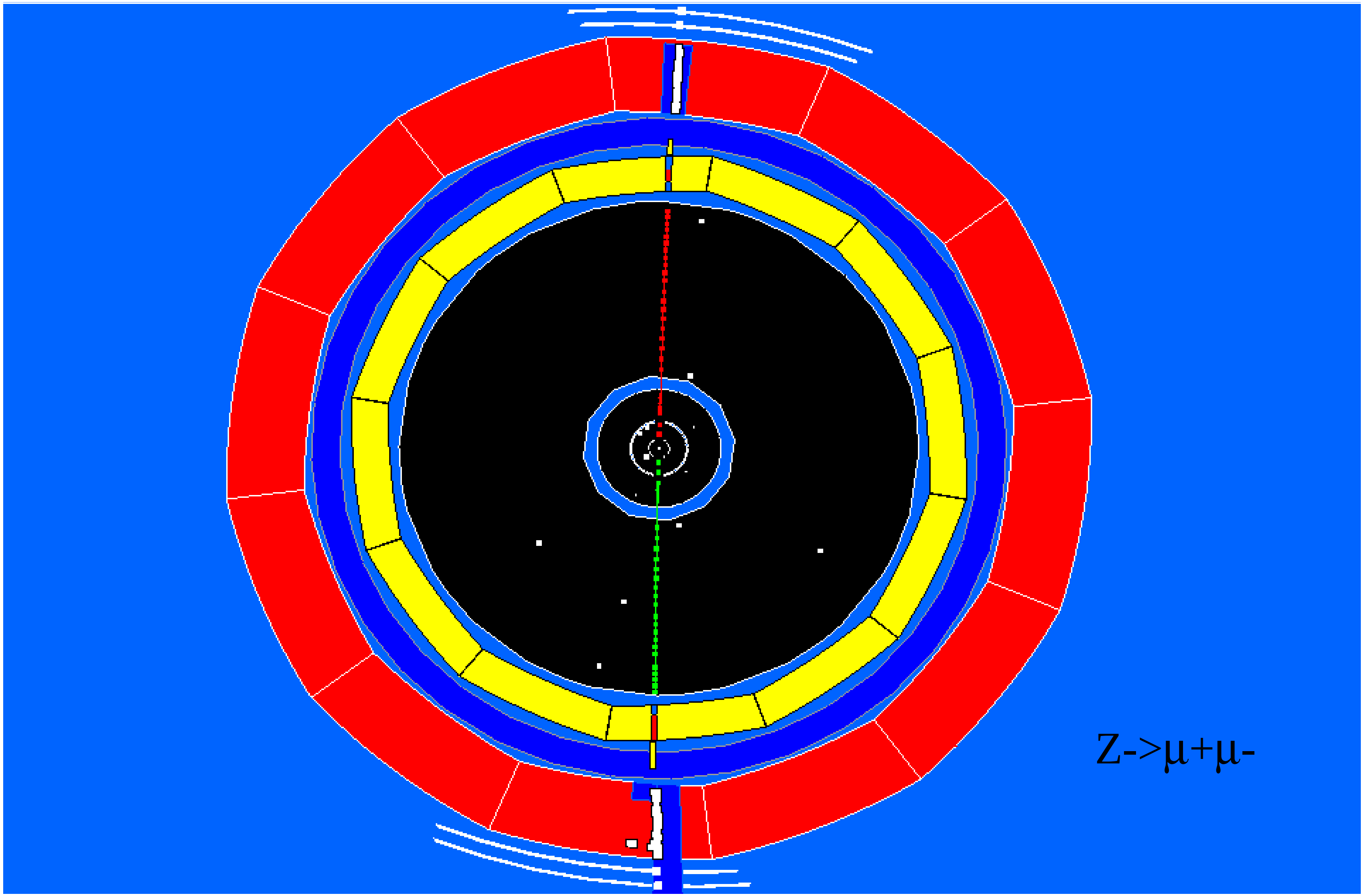

Not Z->μ+μ-

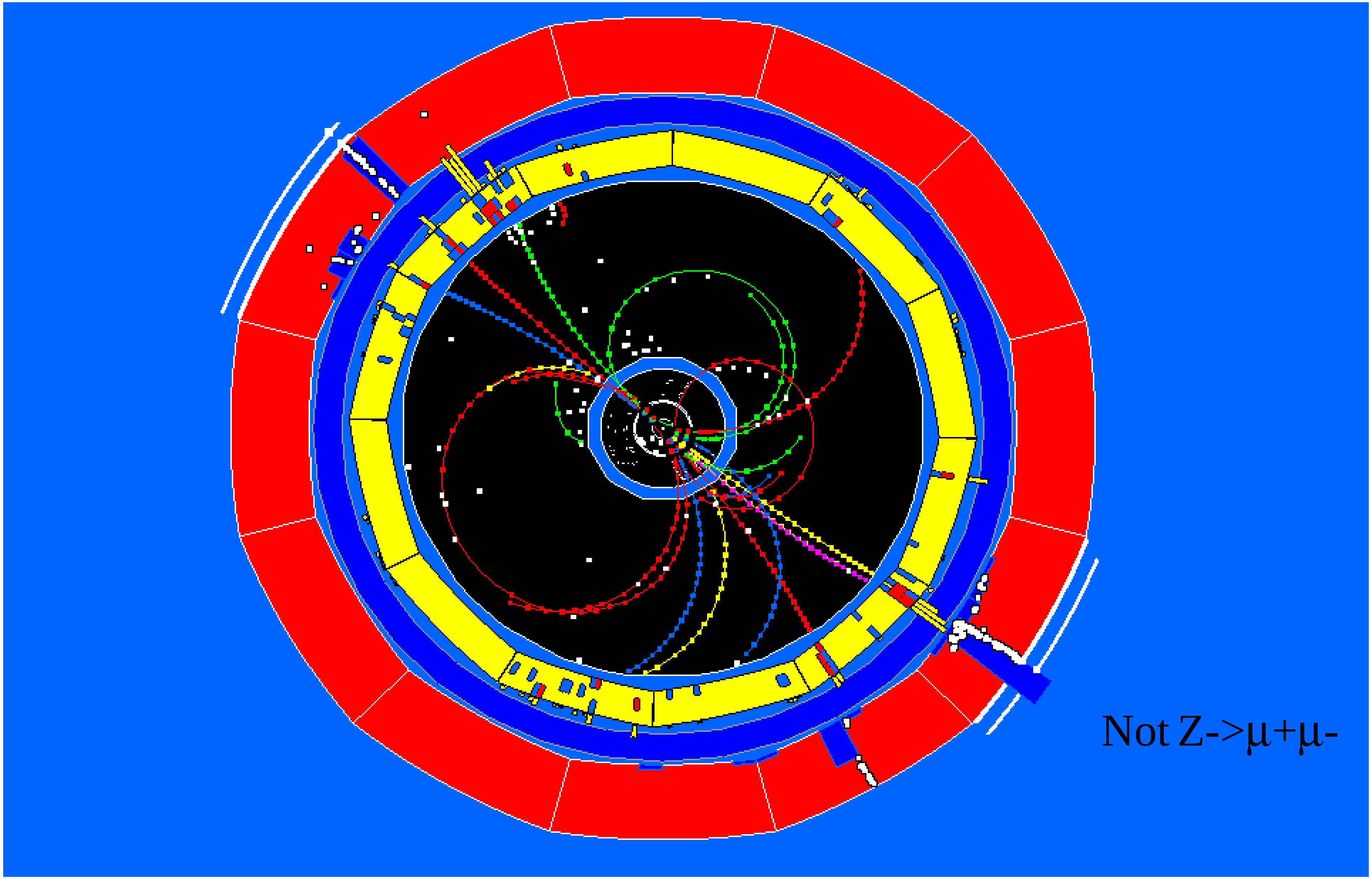

ALEPH
DALI
Run=15995
Evt=835
Z->μ+μ-

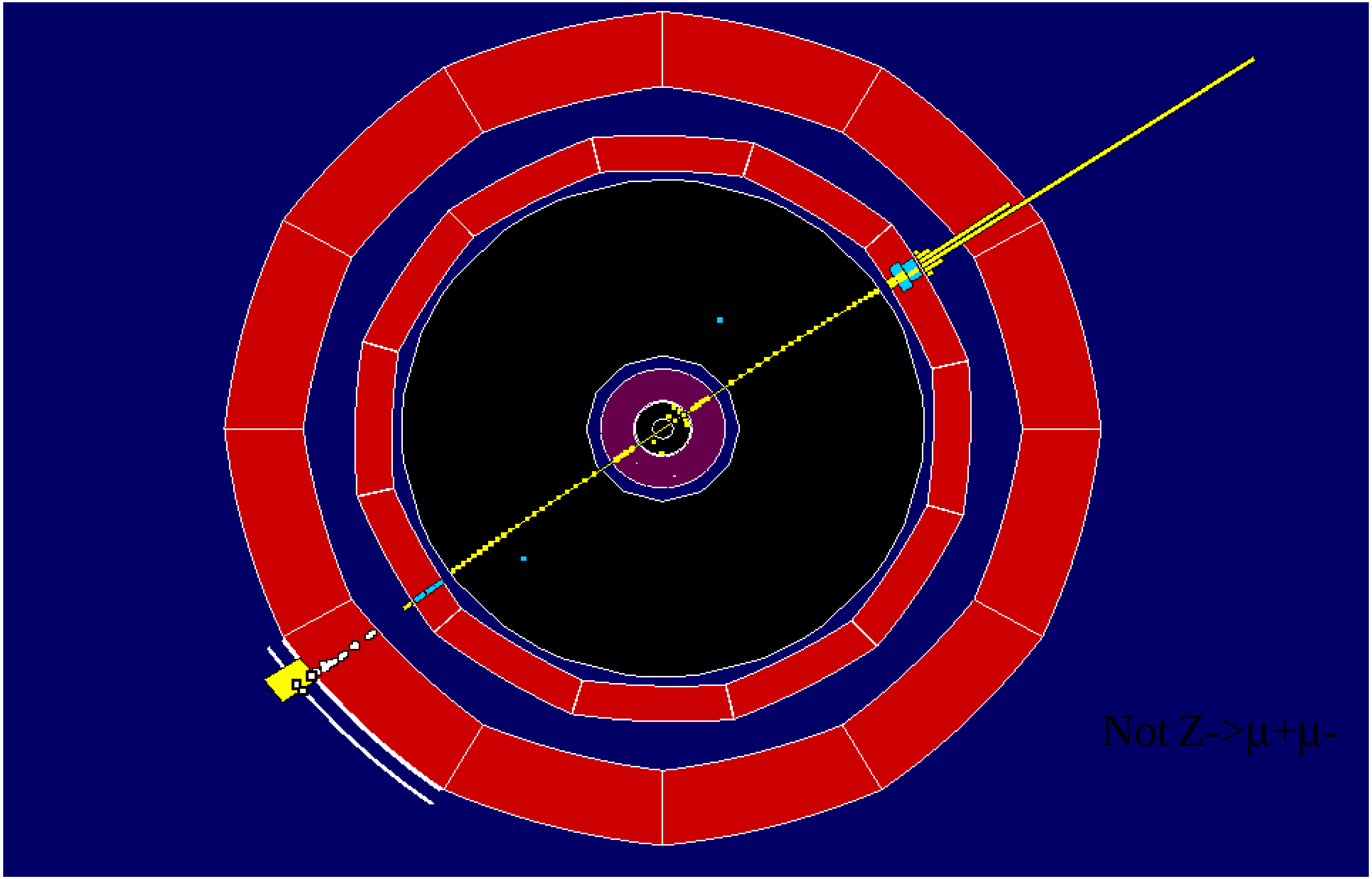

Not Z->μ+μ-

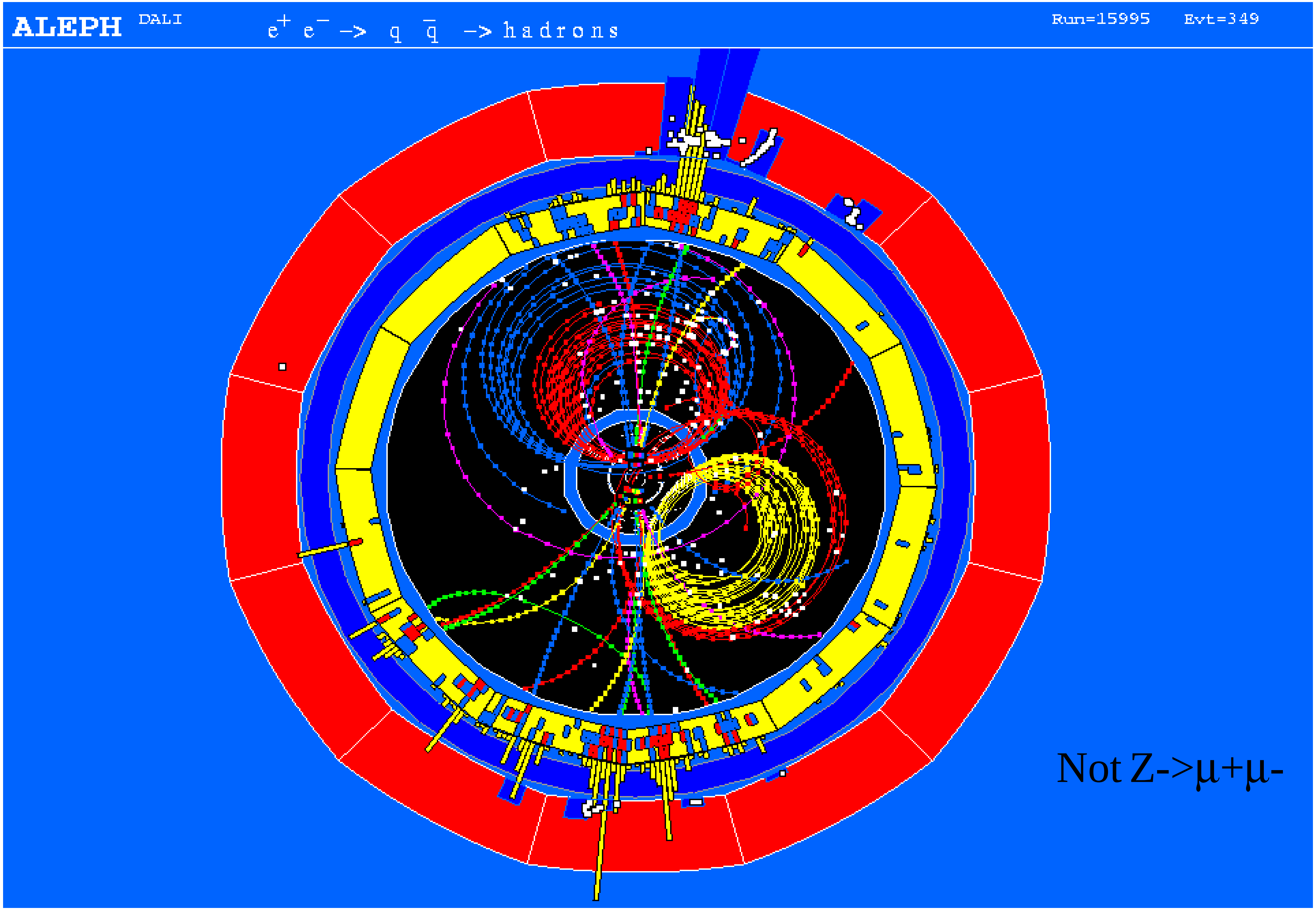

Not Z->μ+μ-

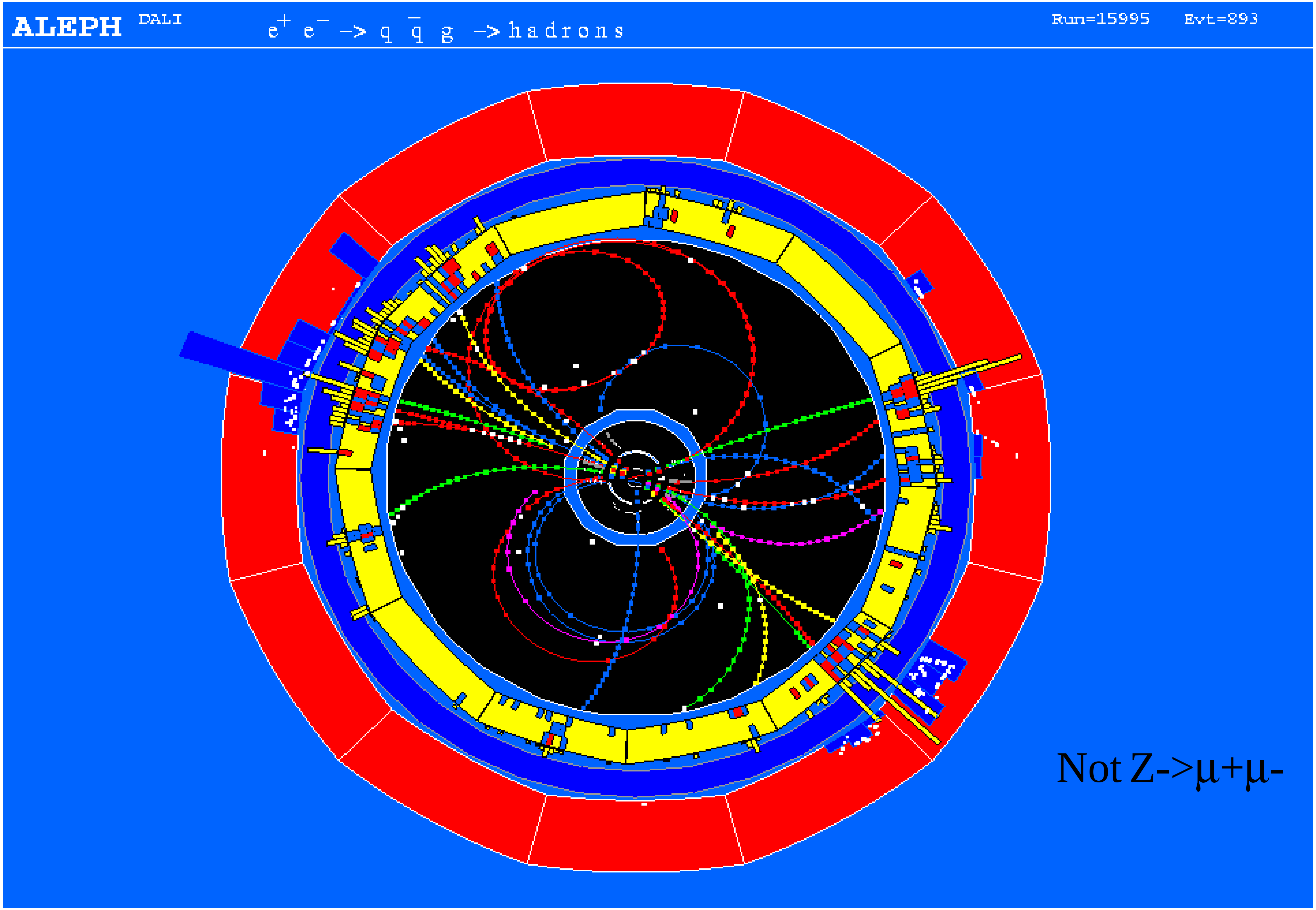

Not Z->μ+μ-


Z->μ+μ-

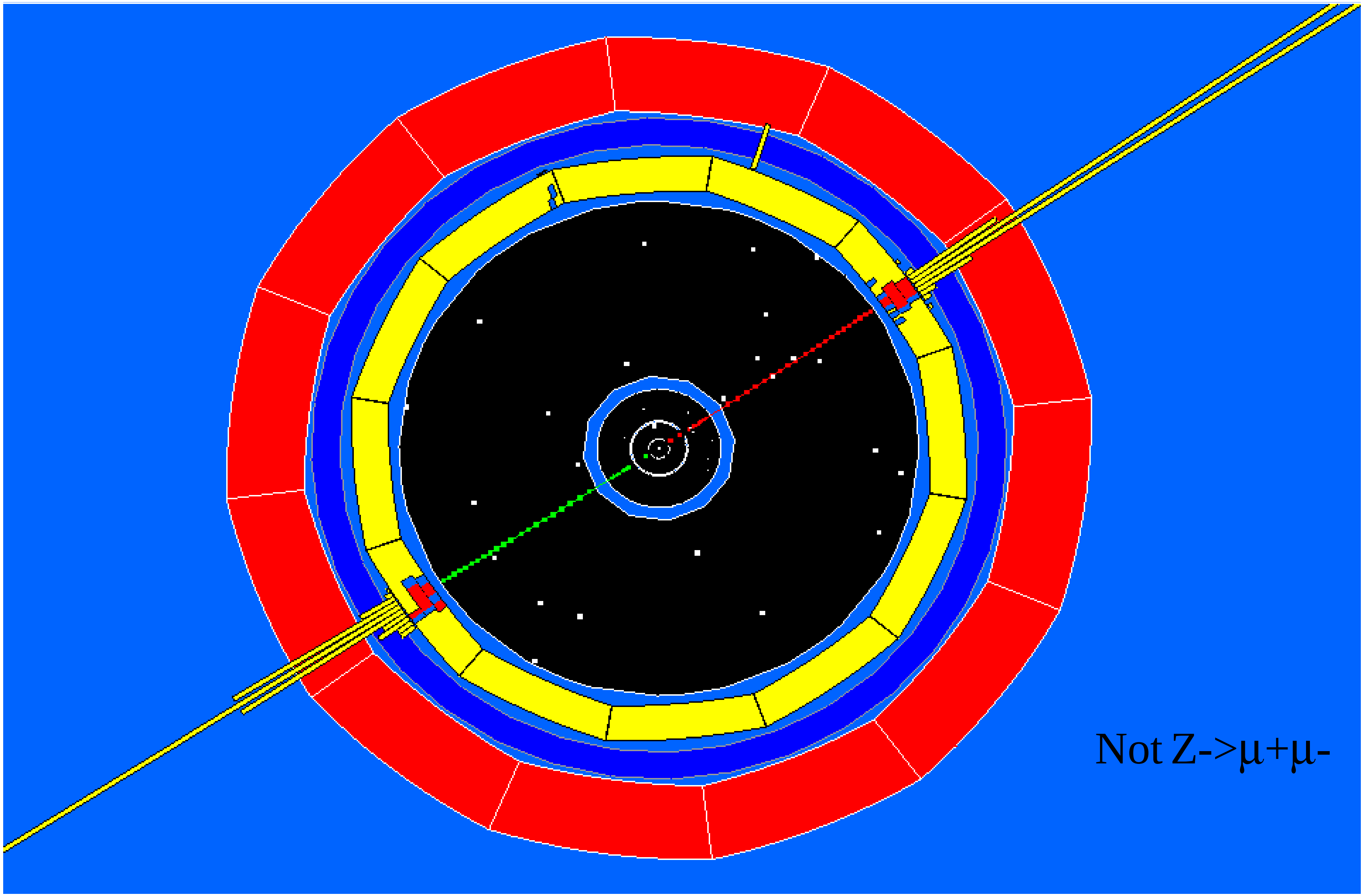

Not Z->μ+μ-

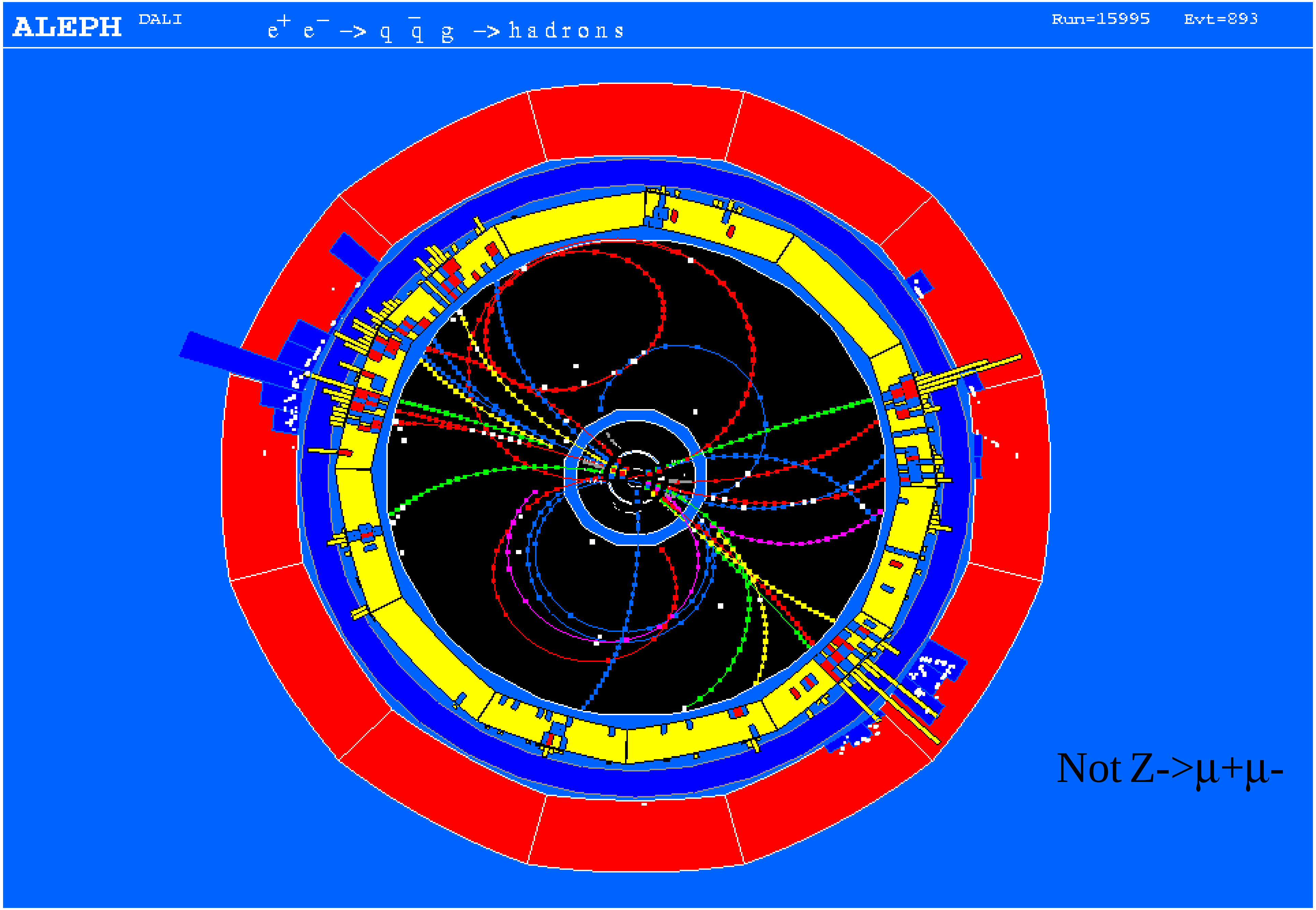

Not Z->μ+μ-

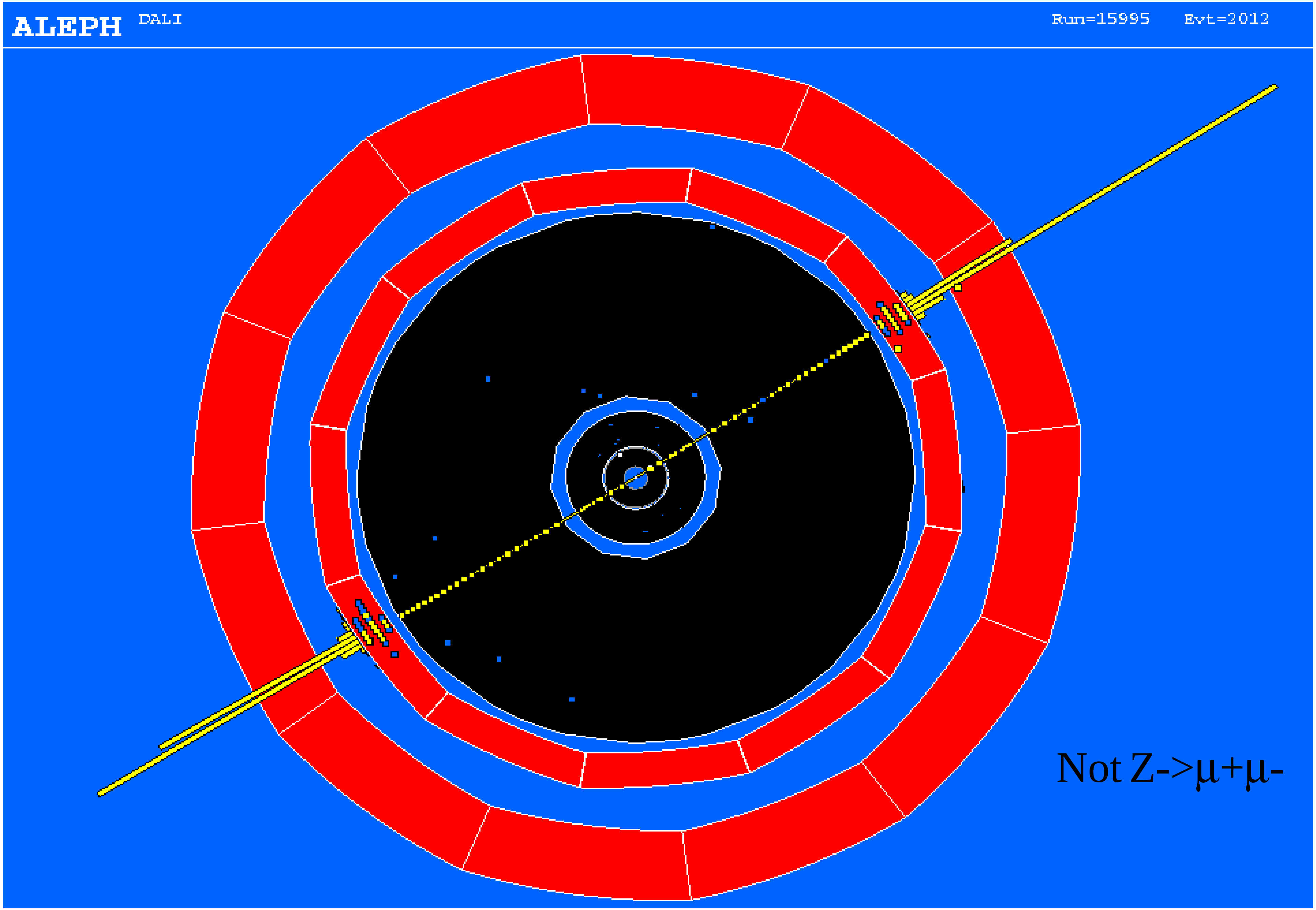

ALEPH DALI
e+ e− -> q q̄ -> hadrons
Run=15995 Evt=349
Not Z->μ+μ-



Not Z->μ+μ-


Not Z->μ+μ-


ALEPH
DALI
e+ e- -> q q̄ g -> hadrons
Run=15995
Evt=893
Not Z->μ+μ-


ALEPH
DALI
Run=15995
Evt=2012
Not Z->μ+μ-

Summary so far

We have a result: BR(Z->μ+μ-) = 2/45

But there's a lot more to do!

Statistical error

- We saw 2 events, but it could easily have been 1 or 3
- Those fluctuations go like the square-root of the number of events:

$$BR\left(Z^0 \to \mu^+\mu^-\right) = \frac{N_{\mu\mu}}{N_{total}} \pm \frac{\sqrt{N_{\mu\mu}}}{N_{total}}$$

- To reduce that uncertainty, you need lots of events

  Need to record lots of events in the detector, and then process them

Systematic error

- What if you only see 50% of the μ+μ- events?

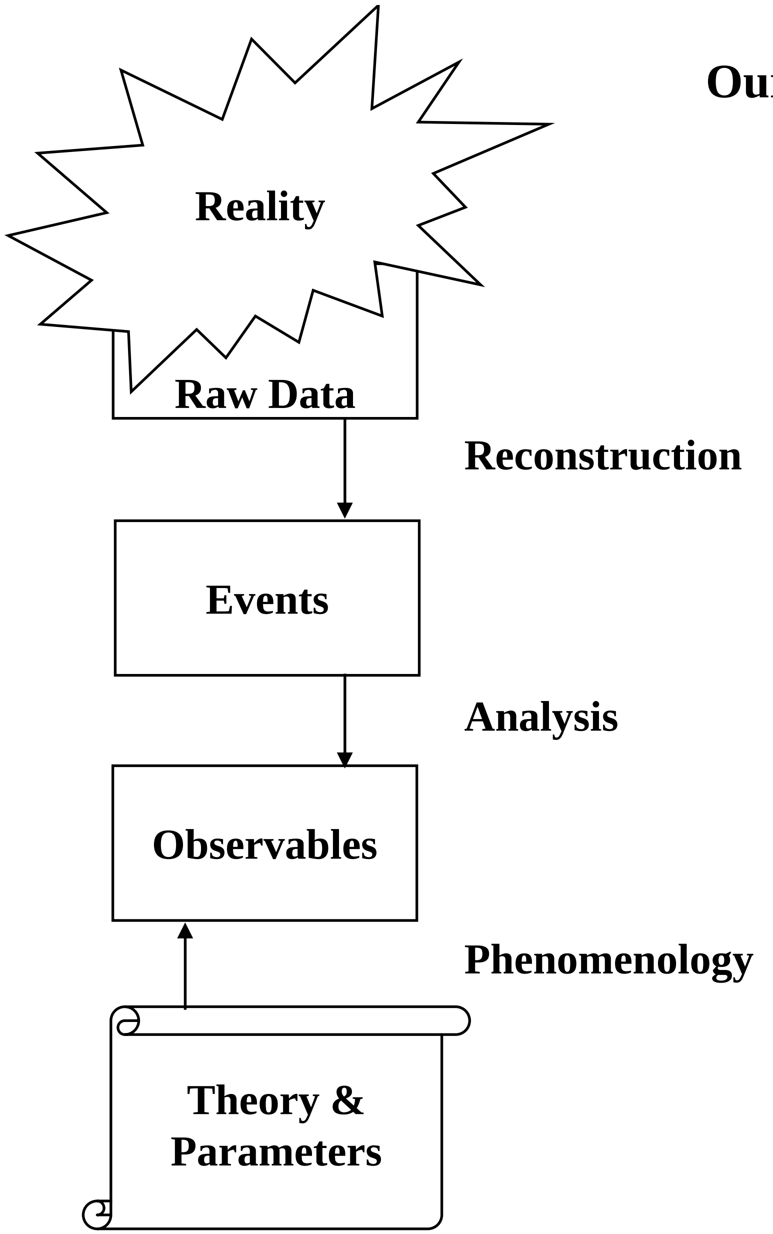
  Due to detector imperfections, poor understanding, etc?

$$N_{\mu\mu_{\text{seen}}} = \varepsilon N_{\mu\mu}$$

$$BR\left(Z^0 \to \mu^+\mu^-\right) = \frac{N_{\text{seen}}/\varepsilon}{N_{total}}$$

$$\varepsilon = 0.50 \pm 0.05$$

Our model so far…



We "confront theory with experiment" by comparing what we measured, with what we expected from our hypothesis.

The process in practice:

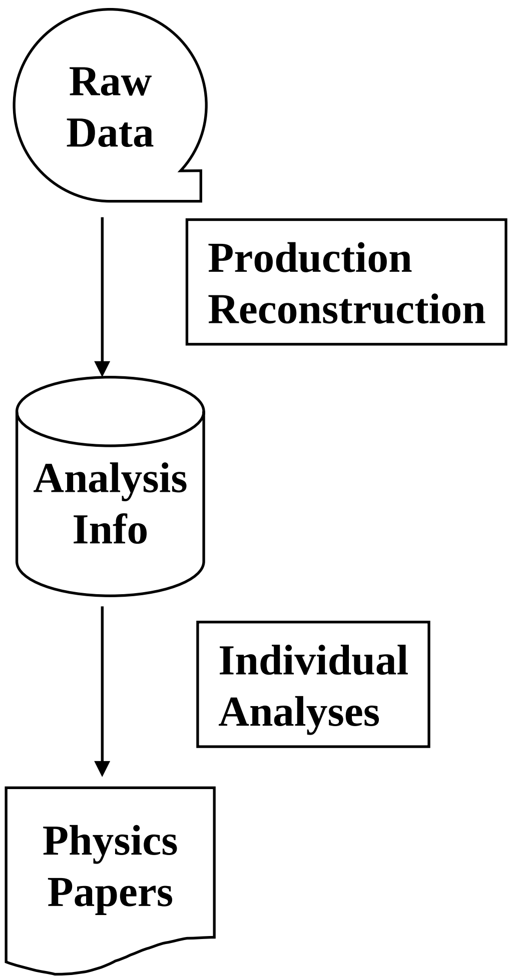
The reconstruction step is usually done in common

- "Tracks", "particle ID", etc are general concepts, not analysis-specific. Common algorithms make it easier to understand how well they work.
- Common processing needed to handle large amounts of data. Data arrives every day, and the processing has to keep up.

Analysis is a very individual thing

- Many different measurements being done at once
- Small groups working on topics they're interested in
- Many different timescales for these efforts

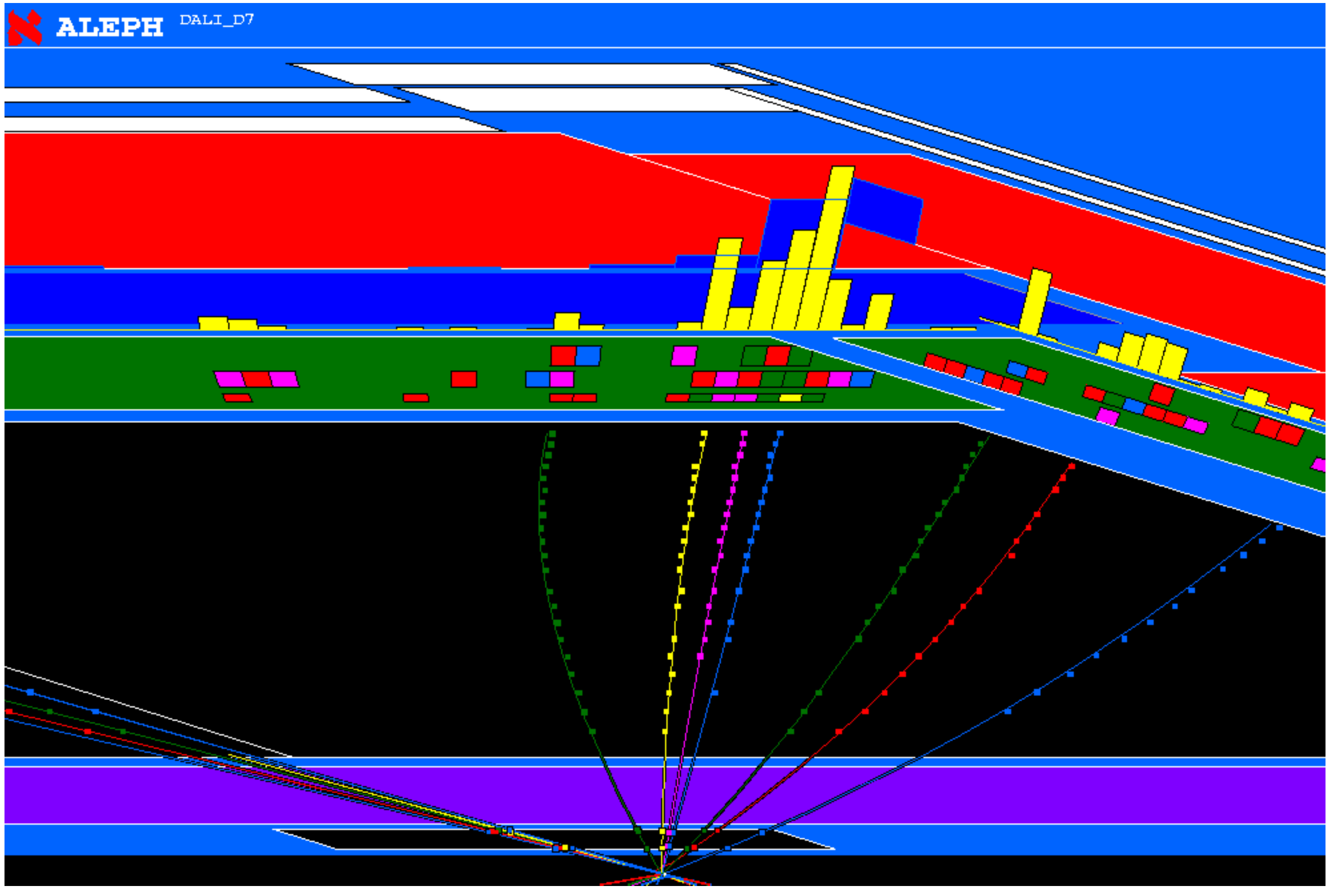
Collaborations build "offline computing systems" to handle all this.



Reconstruction: Calorimeter Energy

Goal is to measure particle properties in the event

- "Finding" stage attempts to find patterns that indicate what happened
- "Fitting" stage attempts to extract the best possible measurement from those patterns.



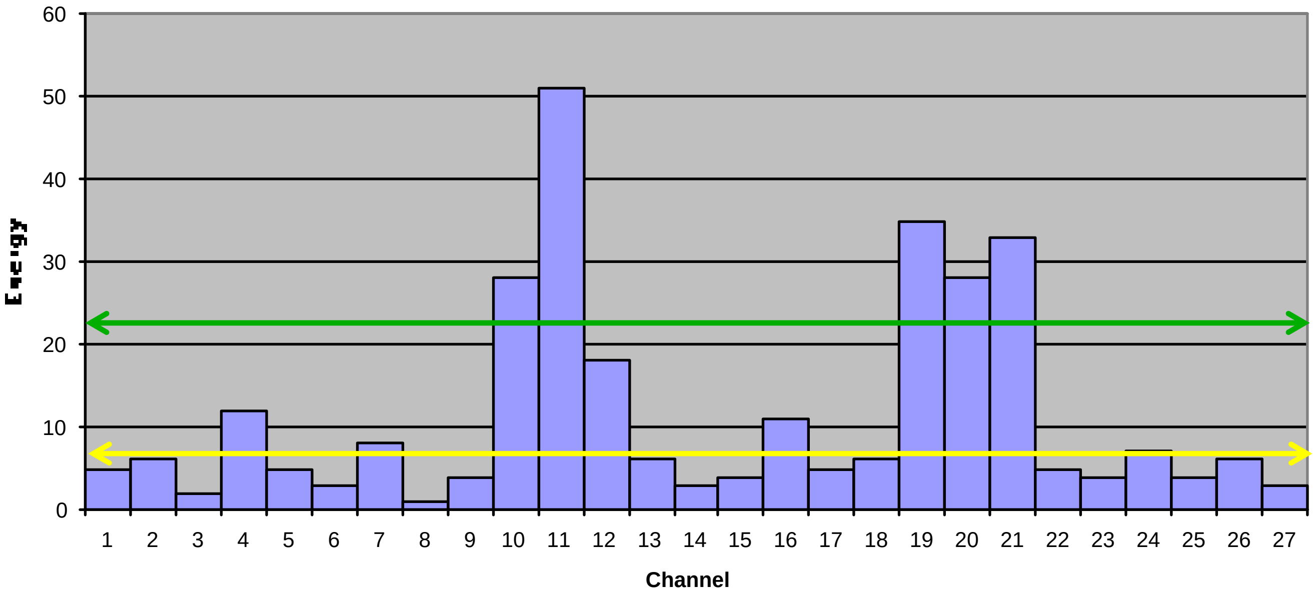
Finding

Clusters of energy in a calorimeter are due to the original particles

- Clustering algorithm groups individual channel energies
- Don't want to miss any; don't want to pick up fakes

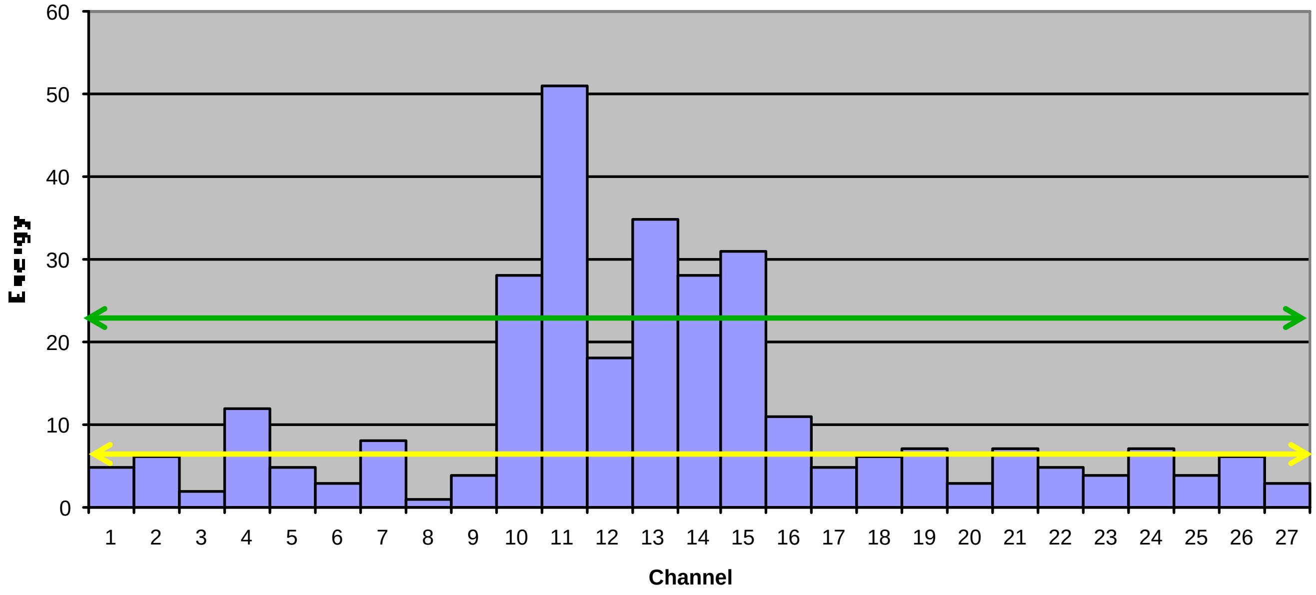
Many algorithms exist

- Scan for one or more channels above a high threshold as "seeds"
- Include channels on each side above a lower threshold:



Not perfect! Doesn't use prior knowledge about event, cluster shape, etc

One lump or two?



Hard to tune thresholds to get this right.

Perhaps a smarter algorithm would do better?

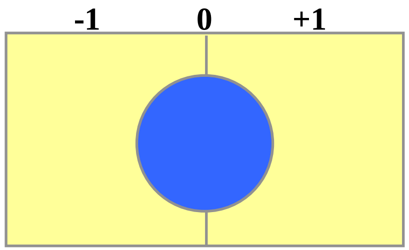
Fitting

From the clusters, fit for energy and position

- Complicated by noise & limited information

Simple algorithm: Sum of channels for energy, average for position

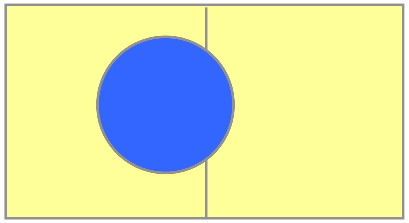
50%, 50%



Cluster at 0, evenly split

$$\frac{-0.5+0.5}{2}=0.0$$

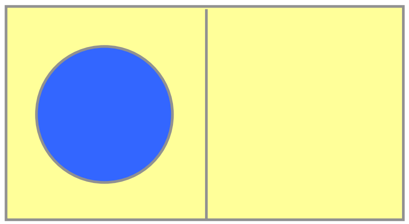
85%, 15%



Cluster at -0.5, unevenly split

$$\frac{-0.85+0.15}{2}=-0.30$$

100%, 0%



Cluster at -1

$$\frac{-1.0+0.0}{2}=-1.0$$

Empirical corrections are important!

Once you understand an effect, you can correct for it

But you need data …

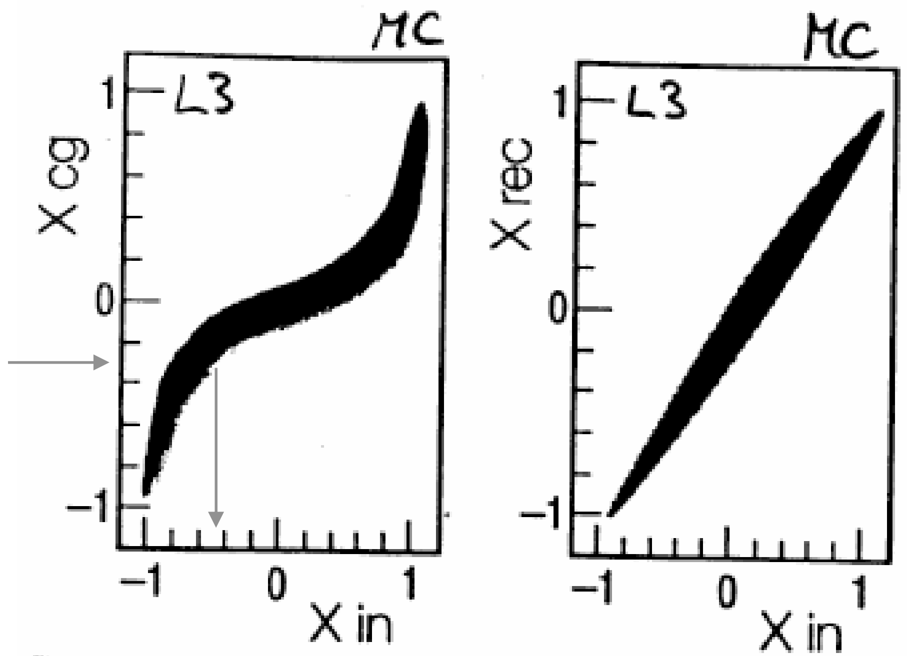


Figure 8 Correlation between the positions measured with *(a)* the center of gravity method (X_{cg}) and *(b)* the reconstructed positions (X_{rec}) vs the actual positions (X_{in}). The results are derived from 5000 $Z \rightarrow e^+e^-$ decays simulated by the GEANT Monte Carlo in the L3 BGO calorimeter (44).

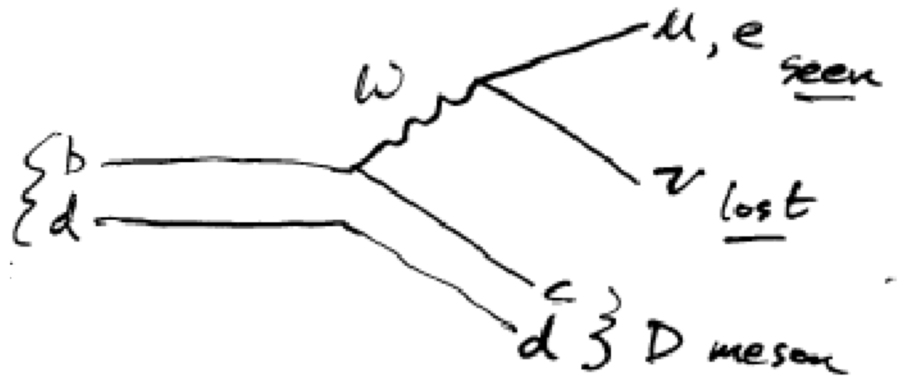
Analysis: Lifetime measurement

Why bother?

Standard model contains 18 parameters, a priori unknown

Particle lifetimes can be written in terms of those

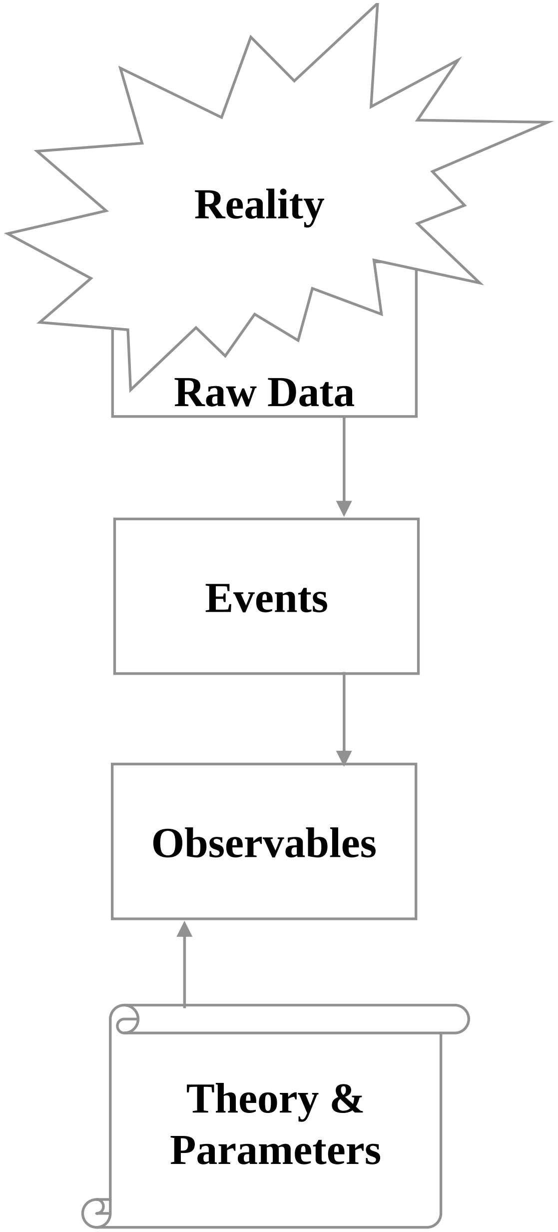
$$\Gamma_Q^{sl} \equiv \Gamma(Q \to q l \upsilon) = \frac{G_F^2}{192\pi^3} m_Q^5 f \left|V_{Qq}\right|^2$$



"Measure once to determine a parameter

Measure in another form to check the theory"

Measure lots of processes to check overall consistency



A model of how physics is done.

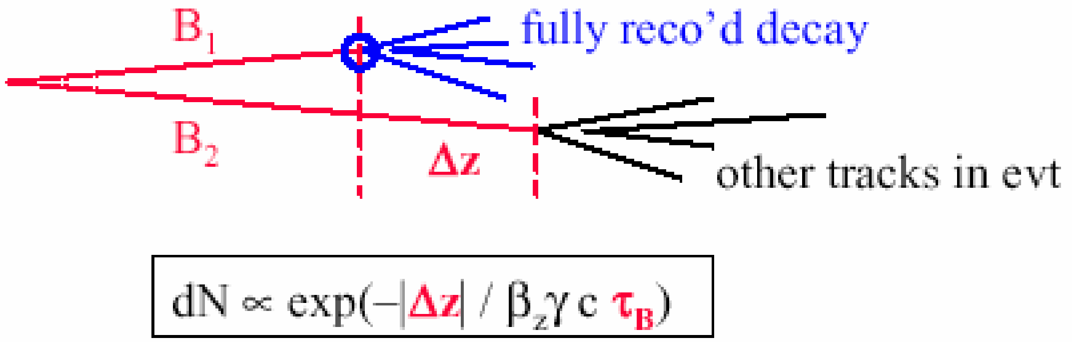
The imperfect measurement of a (set of) interactions in the detector

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

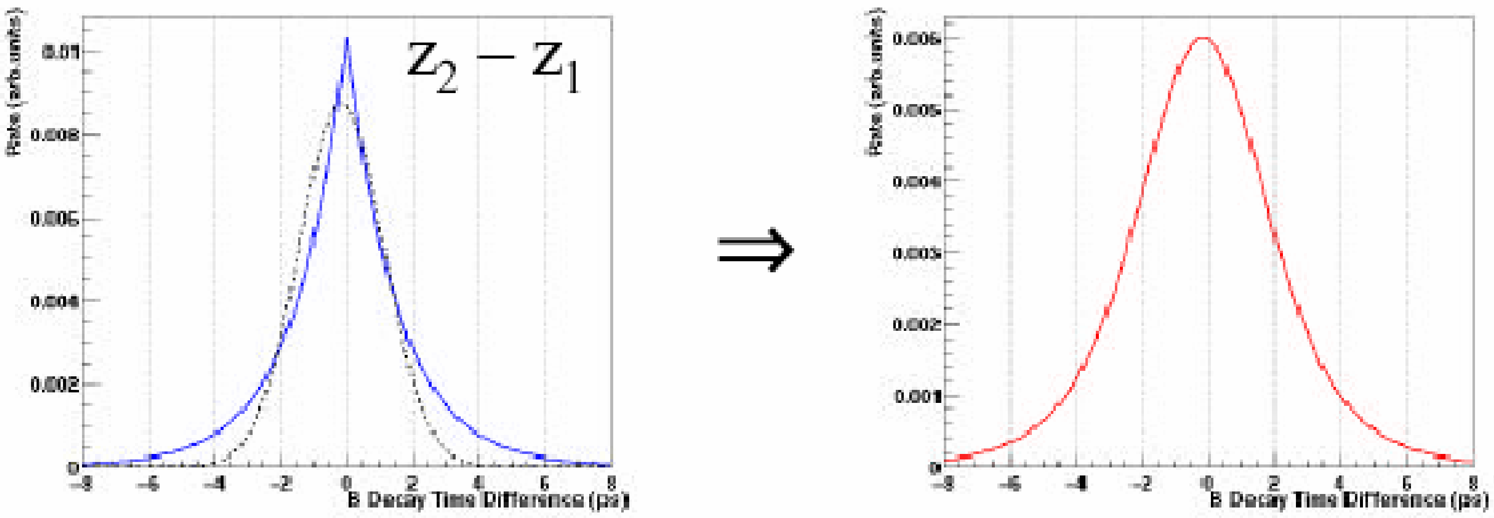
Specific lifetimes, probabilities, masses, branching ratios, interactions, etc

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

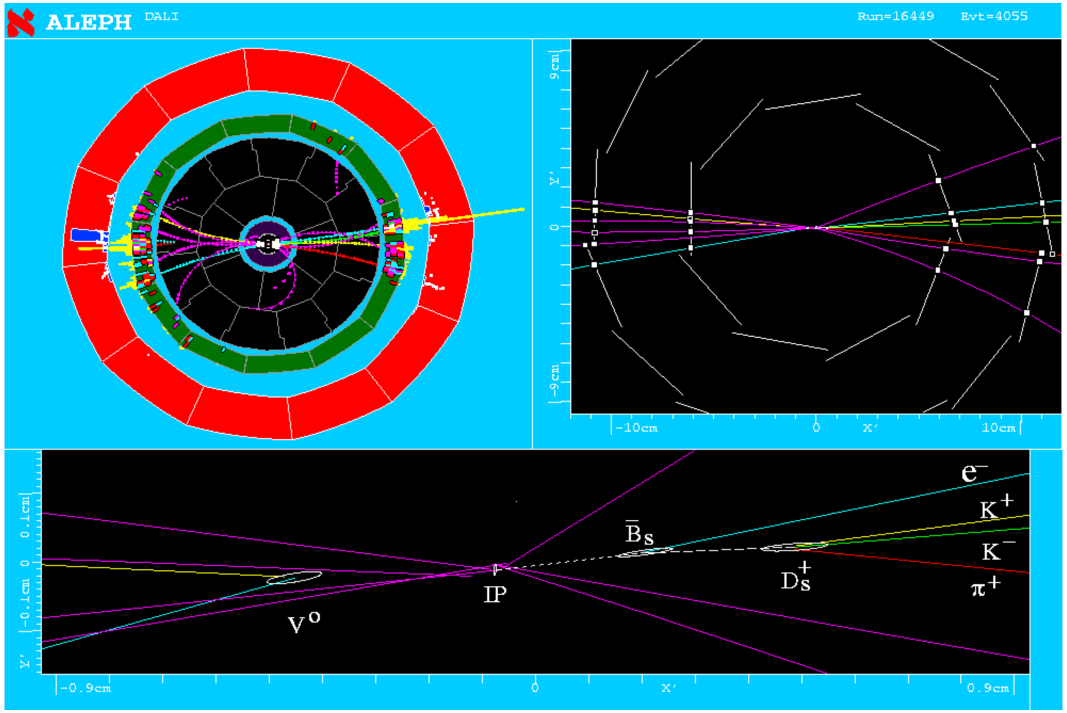
B lifetime: What we measure at BaBar:

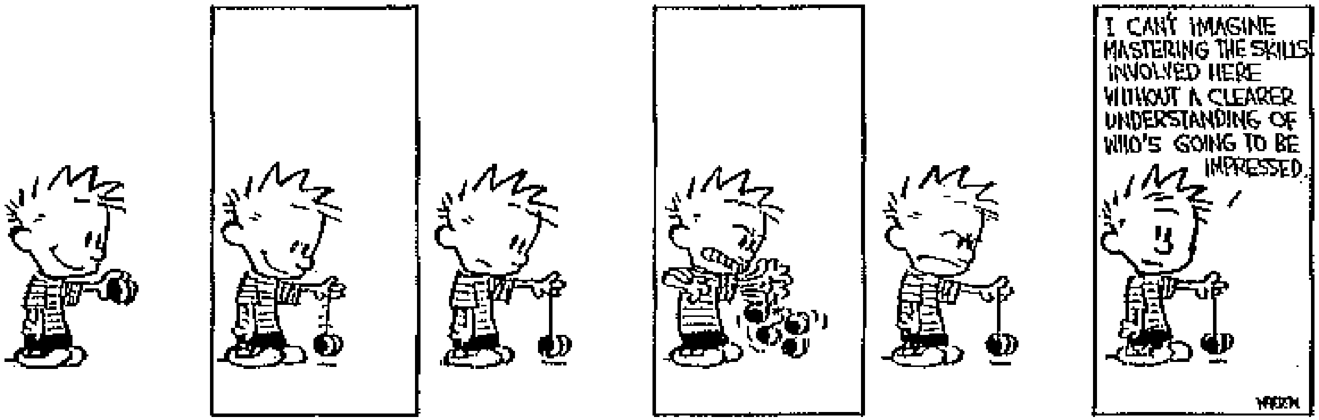


Unfortunately, we can't measure Δz perfectly:



This is why so much effort is put into "tracking"





You also have to find the B vertex

To reconstruct a B, you need to look for a specific decay mode

(Un)fortunately, there are lots!

B0-> D*+ pi-
D*+ rho-
D*+ a1-
D+ pi-
D+ rho-
D+ a1-
J/Psi K*0bar

Each involves additional long-lived particles, which have to be searched for:

D*+ -> D0 pi+
D*0 -> D0 pi0

D0 -> K- pi+, K- pi+ pi0,
K- pi+ pi- pi+, K0S pi+ pi-

D+ -> K- pi+ pi+, K0S pi+

K0S -> pi+ pi-

a1- -> rho0(-> pi+ pi-) pi-
rho- -> pi- pi0
pi0 -> gamma gamma

Psi(2S) -> J/Psi pi+ pi-, mu+ mu-, e+ e-
J/Psi -> mu+ mu-, e+ e-

K*0bar -> K- pi+,

One case: find B–>J/Ψ K*

Neither J/Ψ nor K* is a long-lived particle

- Detector doesn't see them, only their decay products K^*–>Kπ

Take all pairs of possible particles, and calculate their mass

$$m^2 = E^2 - p^2 = (E_1 + E_2)^2 + (\vec{p}_1 + \vec{p}_2)^2$$

If its not the K* mass,

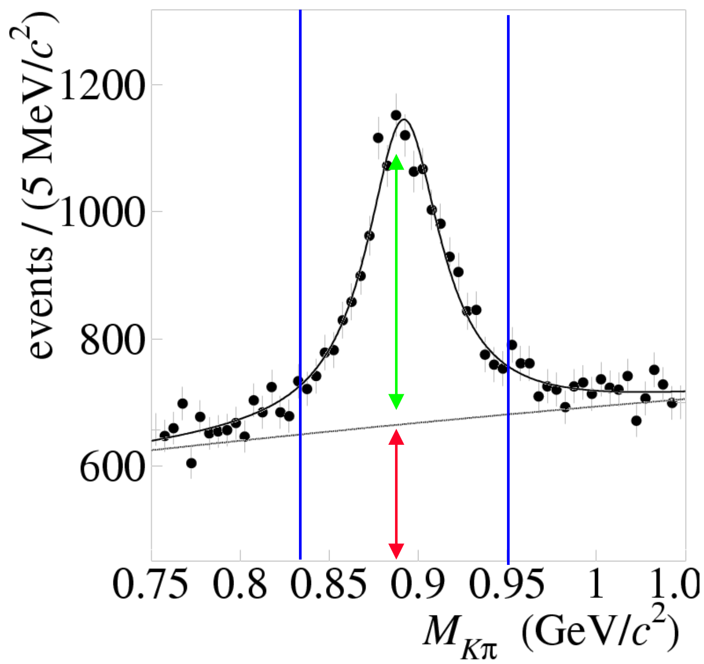
that combination can't be a K^*–>Kπ

If it is the K* mass, it

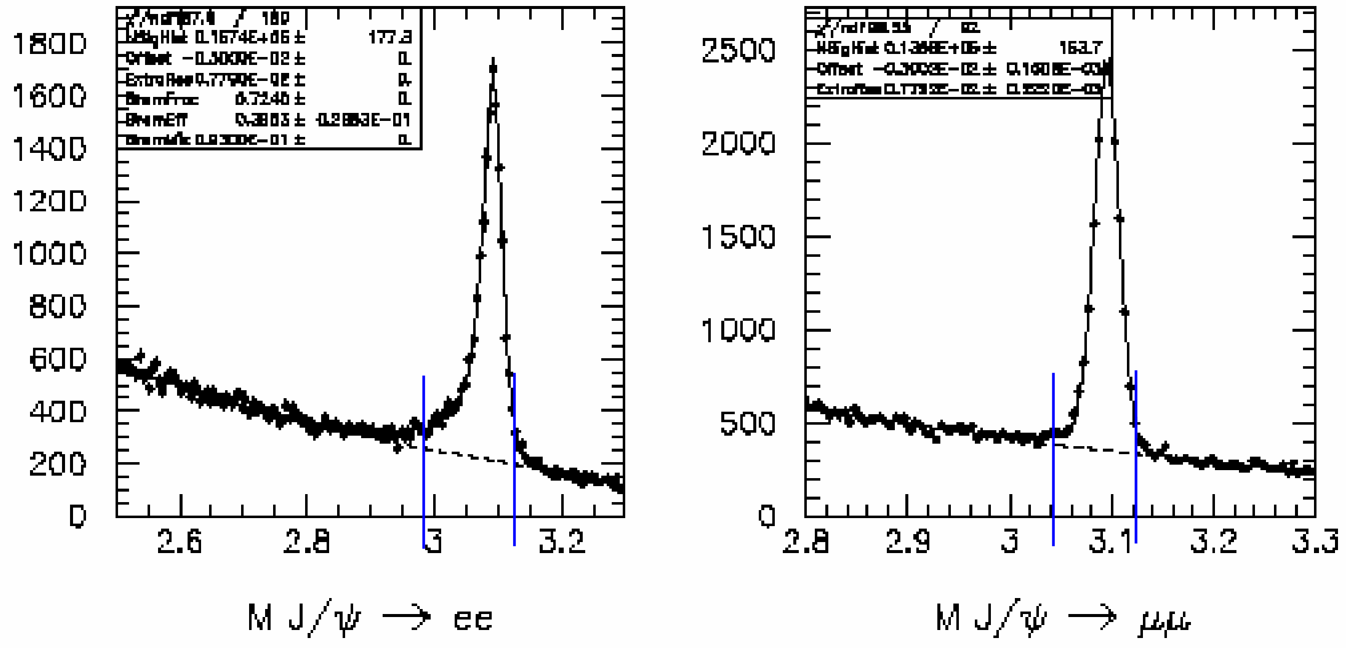
<u>might</u> be a K*

Signal/Background ratio

is critical to success!



Next, look for J/Ψ–>e+e- and J/Ψ–>μ+μ-



Why not J/Ψ–>hadrons? Too many wrong combinations!

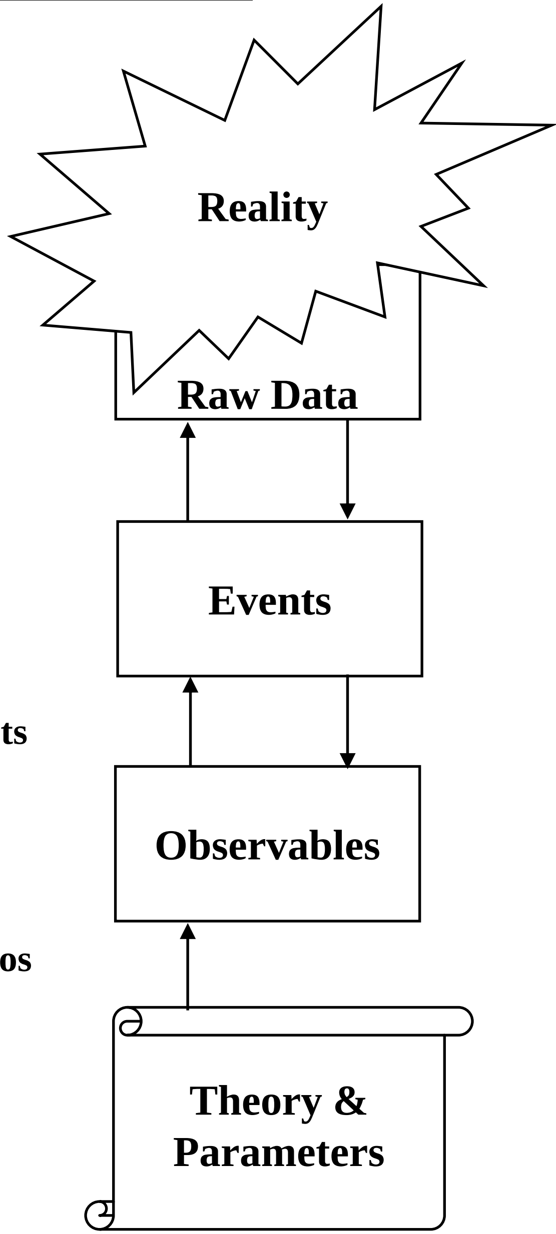
- Only a few e/m in an event, so only a few combinations
- About 10 hadrons, so about 50 combinations of two

  Some are bound to at about the right mass!

Note peaks not same size, shape

- Do we understand our efficiency?

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

How do you know it is correct?

Divide and conquer

- A very detailed simulation can reproduce even unlikely problems
- By making it of small parts, each can be understood
- Some aspects are quite general, so detailed handling is possible

Why does it matter?

- We “cut on” distributions
- Example: Energy (e.g. signal) from particle in a Si detector

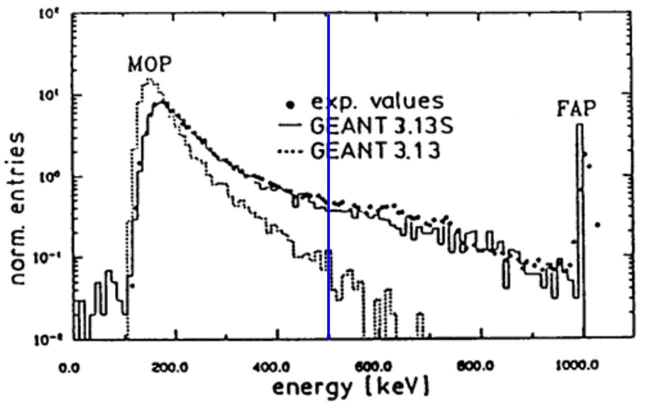


Fig. 15 Comparison of measured and simulated energy deposition in 530 μm silicon for 1 MeV electrons (experimental points see [30]).

Take only particles to left of blue line

Dots are data in test beam
Two solid lines are two simulation codes

One simulation doesn’t provide the right efficiency!

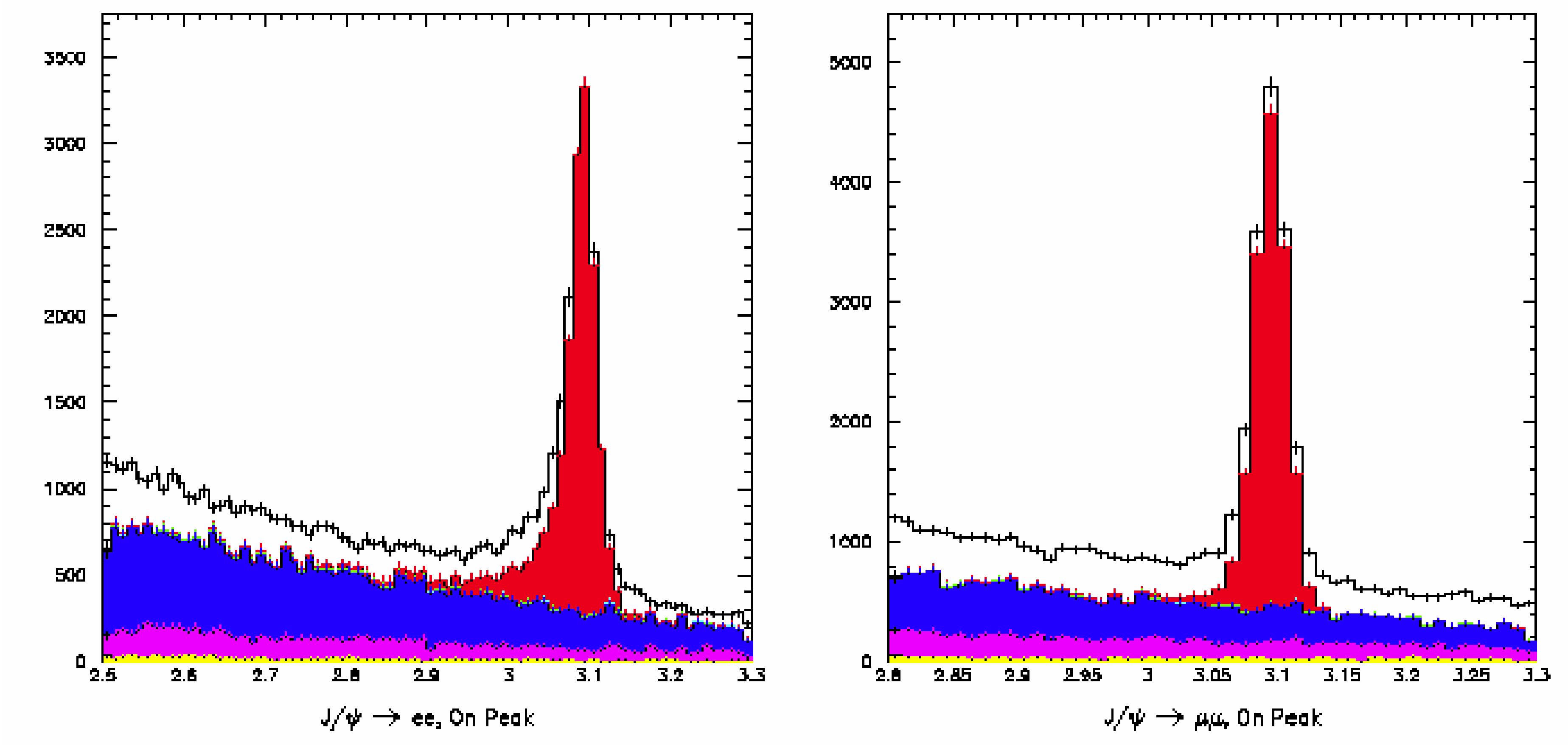


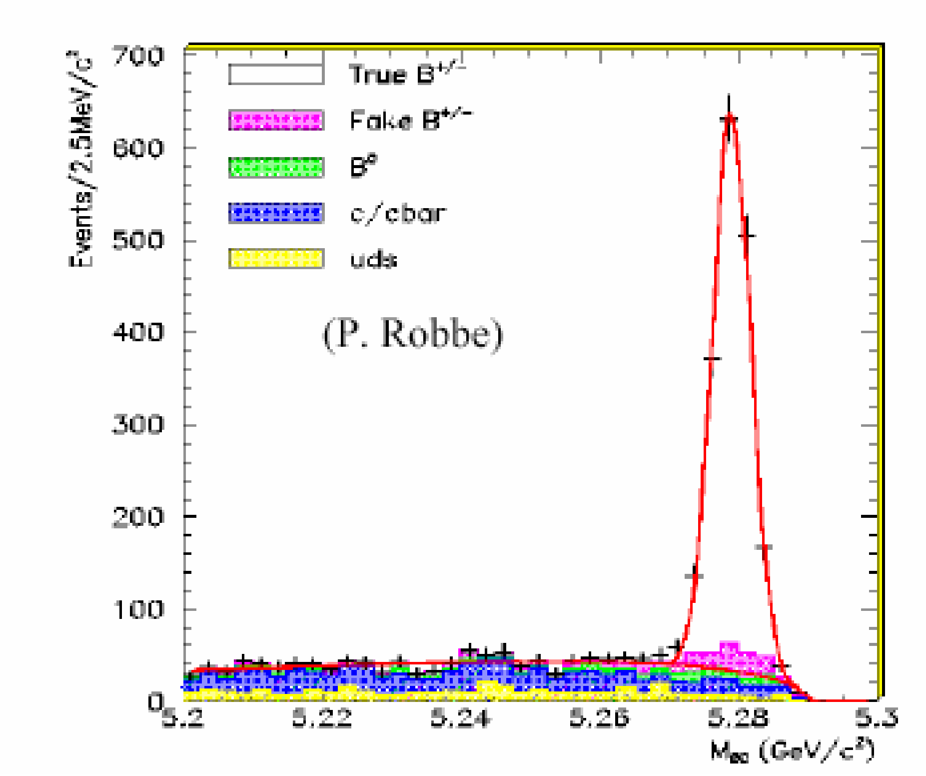
Figure 18: Observed mass distribution superimposed with uds, cc, generic $B\overline{B}$ and signal MC events for (a) $J/\psi \to e^+e^-$ and (b) $J/\psi \to \mu^+\mu^-$.

The tricky part is understanding the discrepancies….

Finally, put together parts to look for B–>J/Ψ K*

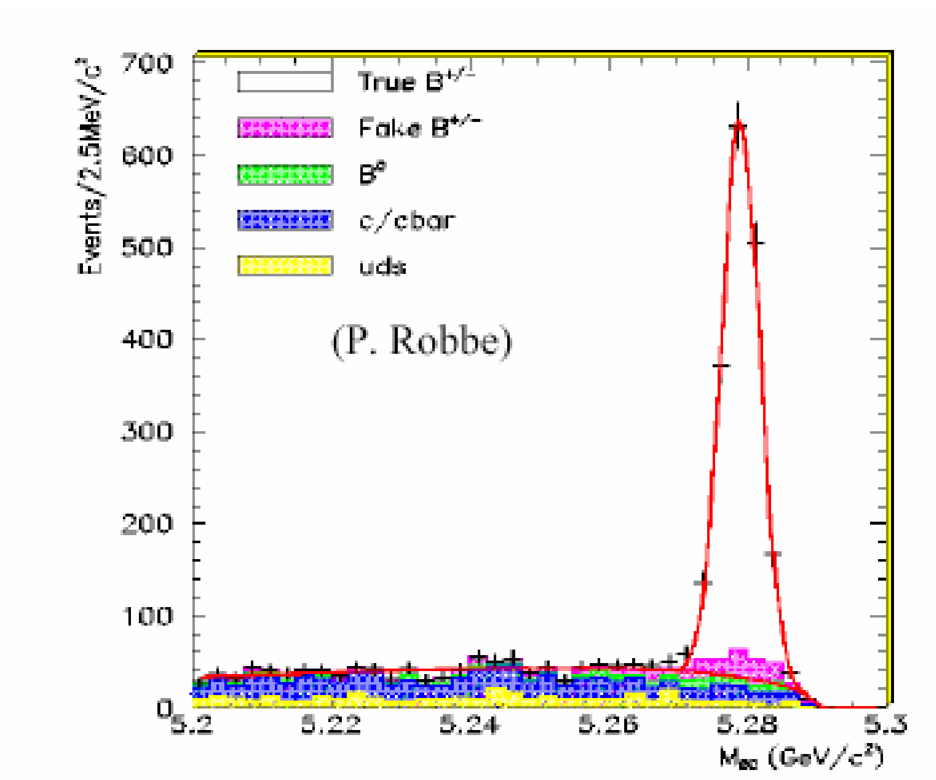
Details:

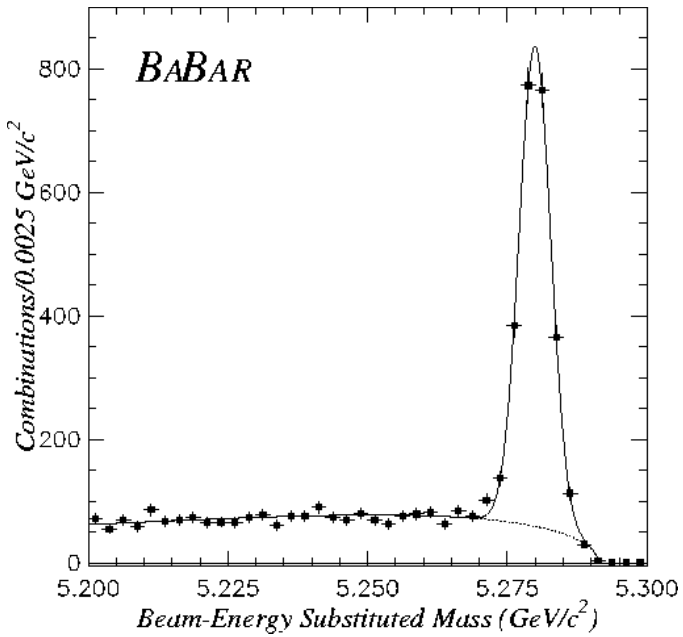
- Background under peak?
- Systematic errors on efficiency
- ….



When you get more data, you need to do a better job on the details

You don't know which are the wrong events!



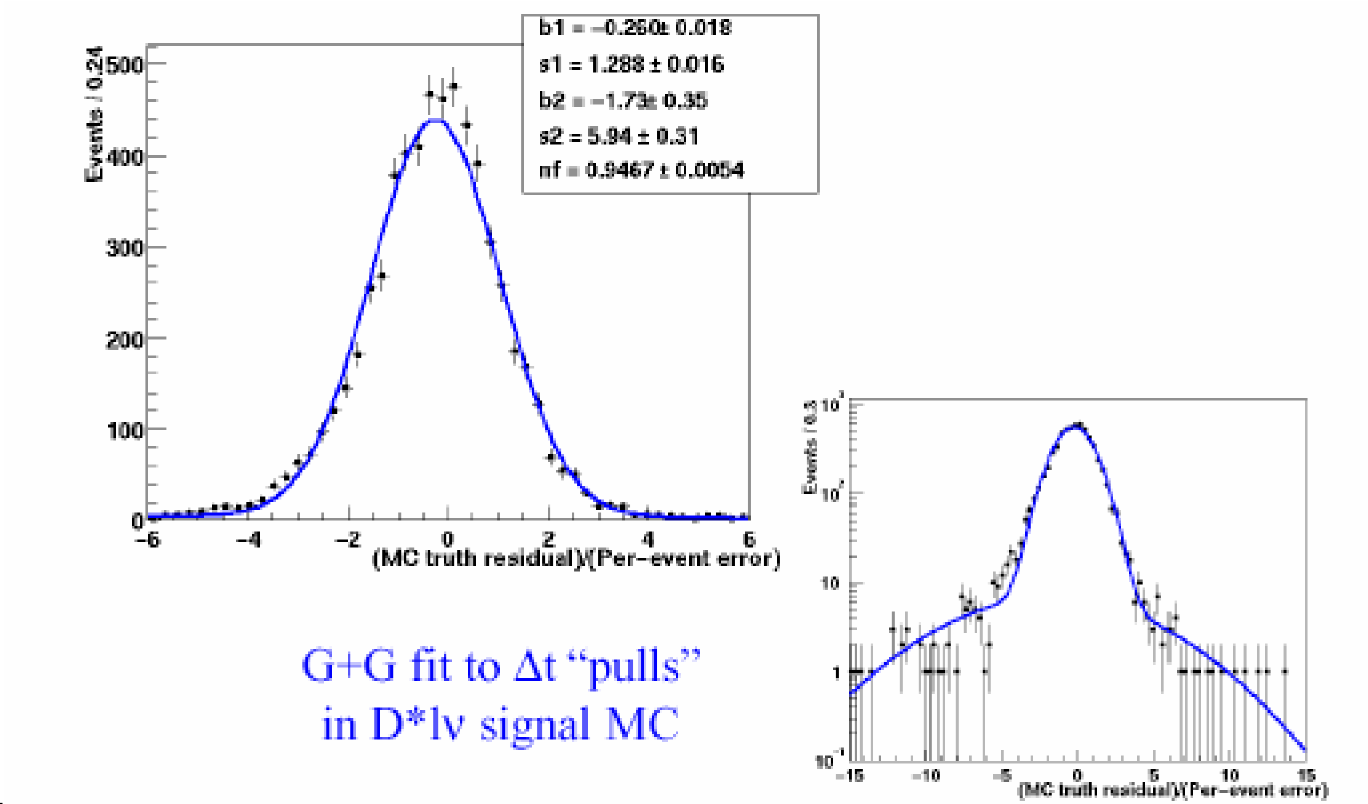


Have to correct for effects of these when calculating the result

Including a term in systematic error for limited understanding

Next, have to understand the resolution:

Studies of resolution seen in Monte Carlo simulation:



But how do you know the simulation is right?

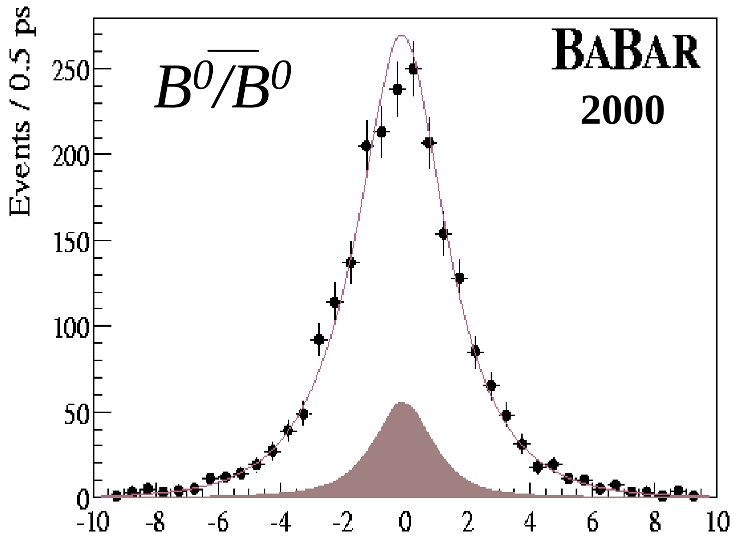
- Find ways to compare data and Monte-Carlo predictions
- Watch for bias in your results!

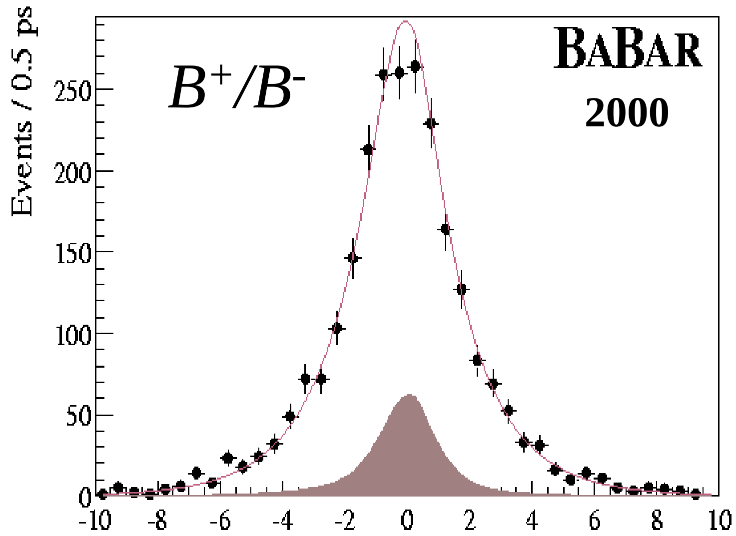
Combined fit to the data gives the lifetime:

You can't extract a lifetime from one event - it's a distribution property

$$N(t) = f(t;\tau) \otimes G(a,b,c,d) + b(t;e,f,g)$$

Try different values until you 'best' fit the data





$\tau_{B0} = 1.506 \pm 0.052\ \text{(stat)} \pm 0.029\ \text{(syst) ps}$ [PDG= 1.548 ± 0.032]

$\tau_{B+} = 1.602 \pm 0.049\ \text{(stat)} \pm 0.035\ \text{(syst) ps}$ [PDG= 1.653 ± 0.028]

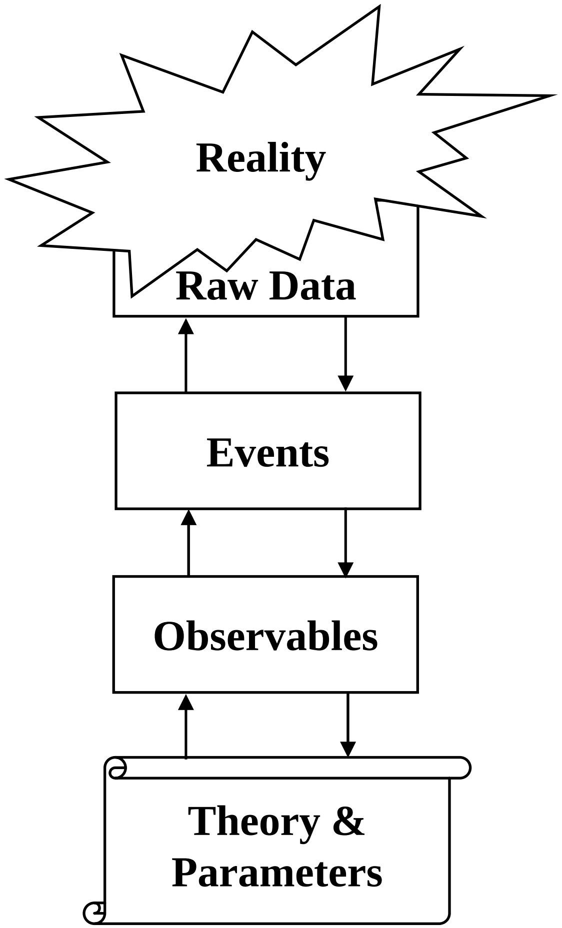
Note that systematic errors are not so much smaller than statistical ones:
2001 data reduces the statistical error; only improved understanding reduces systematic

Summary so far

We seen some simple analyses

We have a model of the steps involved

We're starting to see details of how its done



More detailed examples tomorrow!