Lecture 7:

- Bayes' Theorem
- Blind Analysis
- Data "Correction"
- Statistical Optimisation
- Redundancy

Bayes' Theorem

$$P(\mathrm{A\ and\ B}) = P(\mathrm{B})P(\mathrm{A}|\mathrm{B}) = P(\mathrm{A})P(\mathrm{B}|\mathrm{A})$$

$$P(A\,|\,B) = \frac{P(B\,|\,A)P(A)}{P(B)}$$

note:
$P(A\,|\,B) \neq P(B\,|\,A)$

If there are multiple versions of A to choose from, then

$$P(B) = \sum_j P(B\,|\,A_j)P(A_j)$$

$$P(A_i\,|\,B) = \frac{P(B\,|\,A_i)P(A_i)}{\sum_j P(B\,|\,A_j)P(A_j)}$$

likelihood of the data given the hypothesis

prior probability

hypothesis

$$P(\mathrm{H}_i|\mathrm{D}) = \frac{P(\mathrm{D}|\mathrm{H}_i)P(\mathrm{H}_i)}{\sum_j P(\mathrm{D}|\mathrm{H}_j)P(\mathrm{H}_j)}$$

data

posterior probability

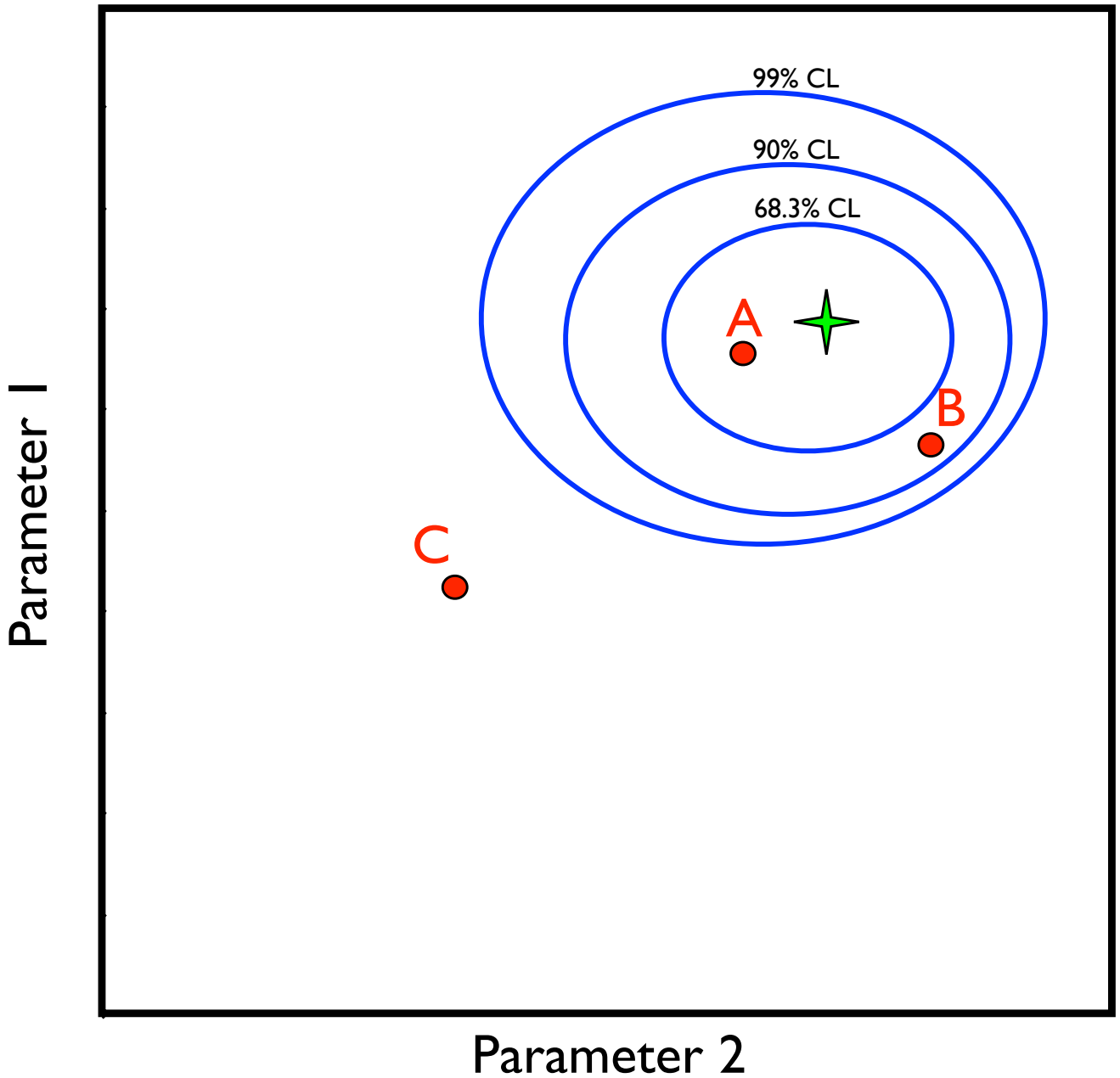
YOUR understanding of whether any one hypothesis is favoured more than any other **prior** to looking at the data

YOUR confidence that a particular hypothesis is true given the data **and** any prior understanding

Relative probability ratio between two different hypothesis given the same observed data

$$\frac{P(H_i|D)}{P(H_k|D)} = \underbrace{\frac{P(D|H_i)}{P(D|H_k)}}_{\text{likelihood ratio}} \underbrace{\frac{P(H_i)}{P(H_k)}}_{\text{“odds” ratio}}$$

Consider a single experiment in which 2 parameters are measured (✦) and compared with predictions from 3 different theoretical models (A, B, C)



Different Definitions of Probability in relation to models:

Bayesian:
Degree of belief. Given a single measurement, ascribe "betting odds" to the phase space of possible models. Requires an assumed context for the comparison of these models (prior). There is no relevance to the "statistical coverage of a confidence interval," because there is only one measurement (which is not repeated over and over again).

Frequentist:
Frequency of occurrence given a hypothetical ensemble of 'identical' experiments. Individual measurements are not used to assess the validity of a model. There is no such thing as a "probability" for a model parameter to lie within derived bounds - either it does or it doesn't. However, if everyone played the same game, the correct model would be bounded a known fraction of the time.

1. Informative:

Permits known, physical constraints to be imposed (e.g. energies and masses must be greater than zero; the position of observed events must be inside the detector, *etc.*) and allows known attributes of the physical system to be taken into account (e.g. energies are being sampled from some particular spectrum; the relative probabilities for different event classes are drawn from some given distribution, *etc.*).

The probabilities of different hypotheses are the same in what metric?

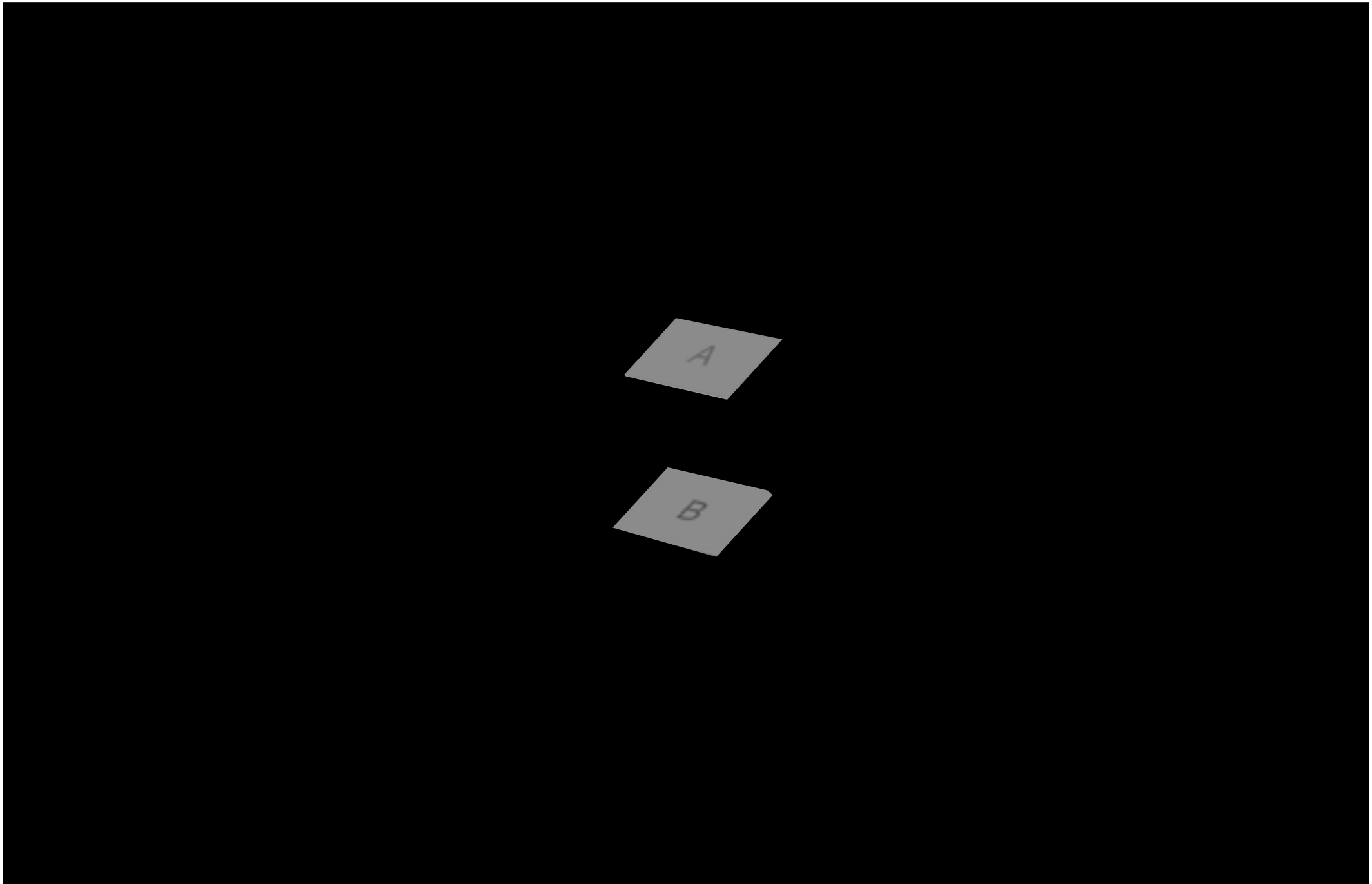
All values of **A** are equally likely $\neq$ All values of $\mathbf{A}^2$ are equally likely

2. Non-Informative:

(A Case of Too Much History!)

When there is no clear *a priori* preference, you must still choose a context to be used for comparing models.

Your brain inherently makes Bayesian inferences:

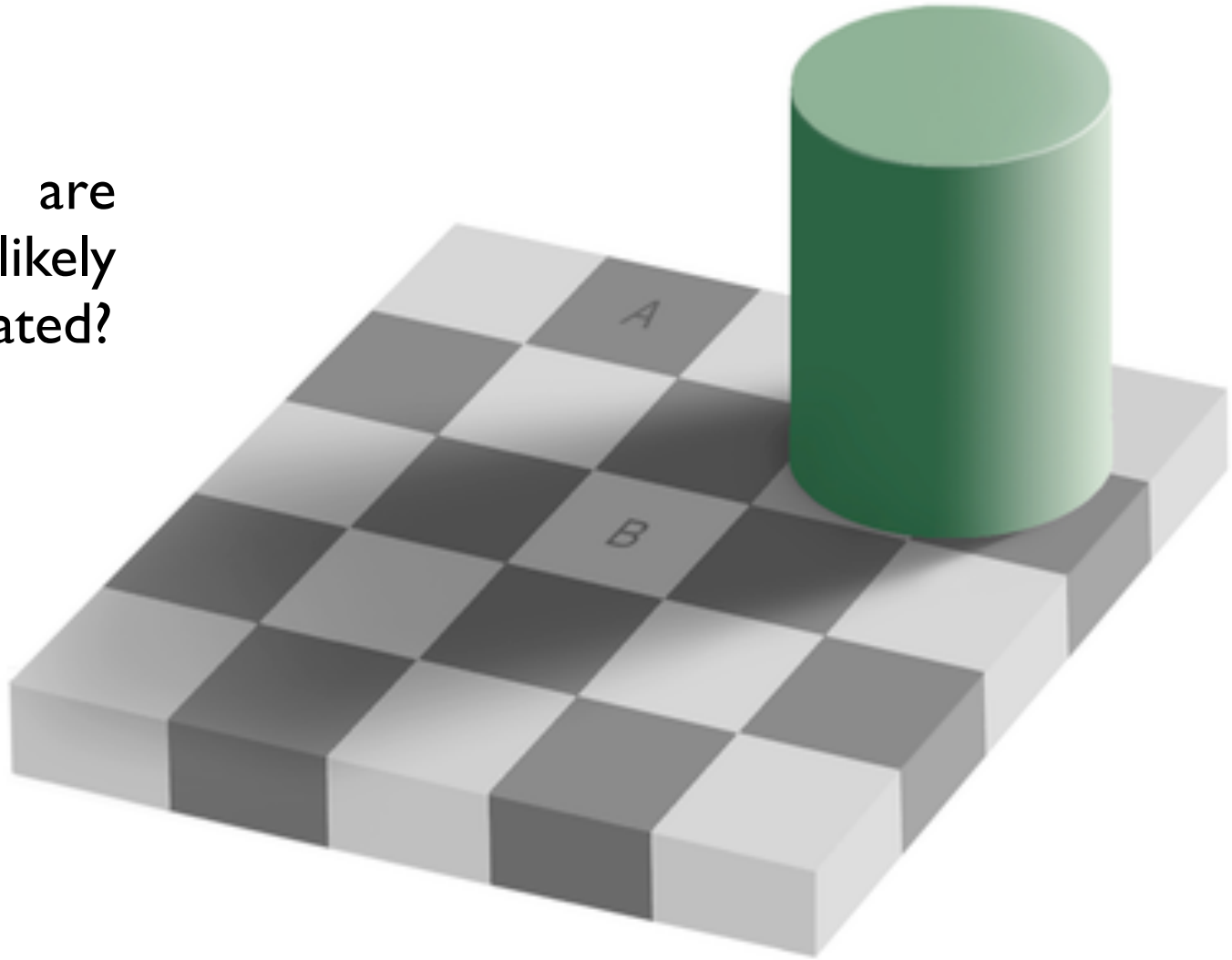


Your brain inherently makes Bayesian inferences:

Context is necessary to relate data to model parameters

(visual observation) (optical properties of surface)

Prior: How are the squares likely being illuminated?



The <u>model</u> is of central importance to enable predictions

Blue and Black

White and Gold

Any inference about models based on an observation is an inherently Bayesian undertaking as it requires an assessment of the posterior probability $P(H_i|D)$ and, thus, **<u>requires</u>** the choice of a prior!

This is ~~often not~~ rarely appreciated! The assumption that the relative likelihoods for two hypotheses alone is the same as the betting odds for which hypothesis is correct tacitly assumes an odds ratio of 1.

If there is an ambiguity in the choice of prior that can lead to notably different conclusions, you should show this!

Science & Environment

Cosmic inflation: 'Spectacular' discovery hailed

By Jonathan Amos
Science correspondent, BBC News

17 March 2014 Science & Environment

Scientists say they have extraordinary new evidence to support a Big Bang Theory for the origin of the Universe.

Researchers believe they have found the signal left in the sky by the super-rapid expansion of space that must have occurred just fractions of a second after everything came into being.

It takes the form of a distinctive twist in the oldest light detectable with telescopes.

The work will be scrutinised carefully, but already there is talk of a Nobel.

"This is spectacular," commented Prof Marc Kamionkowski, from Johns Hopkins University.

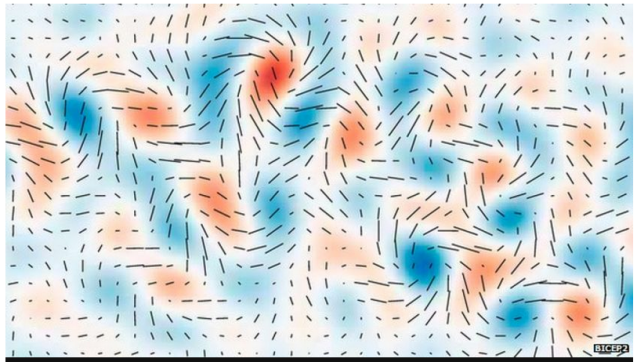
"I've seen the research; the arguments are persuasive, and the scientists involved are among the most careful and conservative people I know," he told BBC News.

> **Nature did not have to be so kind and the theory didn't have to be right**
> Prof Alan Guth, Inflation pioneer

The breakthrough was announced by an American team working on a project known as **BICEP2**.

This has been using a telescope at the South Pole to make detailed observations of a small patch of sky.

The aim has been to try to find a residual marker for "inflation" - the idea that the cosmos experienced an exponential growth spurt in its first trillionth, of a trillionth of a trillionth of a second.



Gravitational waves from inflation put a distinctive twist pattern in the polarisation of the CMB

Science & Environment

Cosmic inflation: BICEP 'underestimated' dust problem

By Jonathan Amos
Science correspondent, BBC News

22 September 2014 Science & Environment

BICEP's South Pole telescope targeted what the team hoped was a relatively clean part of the sky

One of the biggest scientific claims of the year has received another set-back.

Example:

As the result of a random blood test, you are diagnosed with "Saturday Night Fever," a disease suffered by 0.5% of the population that results in convulsions when exposed to anything associated with John Travolta. The blood test reliably diagnoses the disease in 80% of cases and yields a false positive 5% of the time. Should you avoid listening to BeeGees albums?

$$P(SNF \mid B) = \frac{P(B \mid SNF)P(SNF)}{P(B \mid SNF)P(SNF) + P(B \mid no\ SNF)(P(no\ SNF)}$$

$$= \frac{(0.8)(0.005)}{(0.8)(0.005) + (0.05)(0.995)} \qquad = 0.074$$

What if the reason you went to your GP for a blood test was that you got splitting headaches whenever someone mentioned the word "Grease?"

These are basically the same numbers as for COVID-19 (early Oct 2020).

<u>What if you feel ill and get a positive test?</u>

Say the average person is typically ill 10 days per year, so the odds of currently being ill from the common cold is ~10/365 = 0.027. With social distancing, reduce this by a factor of ~10 to 0.0027. So, the fraction of people feeling ill that have COVID-19 is perhaps something like 0.005/(0.005+0.0027) = 0.65 (this, then, is the prior instead of 0.005).

$$P(CV19 \mid +T) = \frac{P(+T \mid CV19)P(CV19)}{P(+T \mid CV19)P(CV19) + P(+T \mid no\ CV19)P(no\ CV19)}$$

$$= \frac{(0.8)(0.65)}{(0.8)(0.65) + (0.05)(0.35)} \quad = 0.97$$

Priors are important!

Example 2:

Atmospheric neutrinos result from the decay of charged pions produced by hadronic interactions in the atmosphere. The characteristic decay sequences are:

$$\pi^- \to \mu^- + \overline{\nu}_\mu \qquad\qquad \pi^+ \to \mu^+ + \nu_\mu$$
$$\hookrightarrow e^- + \nu_\mu + \overline{\nu}_e \qquad\qquad \hookrightarrow e^+ + \overline{\nu}_\mu + \nu_e$$

You are detecting these neutrinos coming from directly overhead with an underground water Cherenkov detector. From the fuzziness of the ring pattern of observed light from a particular event, simulations tell you that 70% of ν_e's will produce a ring at least this fuzzy, whereas only 50% of ν_μ's will do this. What is the probability that this event is a ν_e ?

$$P(\nu_e|R) = \frac{P(R|\nu_e)P(\nu_e)}{P(R|\nu_e)P(\nu_e) + P(R|\nu_\mu)P(\nu_\mu)}$$

$$= \frac{(0.7)(1/3)}{(0.7)(1/3) + (0.5)(2/3)} \quad = 0.41$$

Bernstein – von Mises Theorem

In the limit of an infinitely large data set, the posterior probability is independent of the exact form of the prior probability.

(the likelihood function that multiplies the prior crushes it's impact away from the region of interest)

For example, if you instead asked for the probability for a large number Cherenkov events to be ν_e out of a big data set, the information contained in the distribution of ring fuzziness within the data itself carries more weight than the form of any previously assumed prior.

Priors carry greater weight for weaker data sets

Sally Clark: A Statistical Conviction

Sentenced to life imprisonment in 1999 for murdering her two children because the odds of suffering two cot deaths was estimated to be far too small. No other aspects of possible motive or evidence came into play. The conviction was finally overturned after a second appeal in 2003.

The Case as presented:

Odds of 1 cot death: 1 in 1303

For affluent, non-smoking, mother over 26 years old, odds drop to: 1 in 8500

Odds of 2 cot deaths: $(1/8500)^2$ = 1 in 73 million

Ignored factors such as that both children were boys, which statistically increases the risk of cot death. Additionally, the same factors cited also make the statistical probability of homicide smaller. Stick with 1 in 1303 .

Are probabilities independent? Any evidence of possible genetic links in cot death? Yes: Studies suggest odds of 2nd cot death increase by a factor of 10-20. Assume a factor of 10.

So, the overall odds are roughly (1/1303)x(1/130) = 1 in 169390

Still pretty small!

What's the question?

"If she were innocent, what are the odds that this would happen?"

"Prosecutor's Fallacy"

Correct question:

"Given these two deaths, what are the odds that she committed murder?"

There are ~650000 births in England and Wales each year. Assume roughly half of these are for a 2nd child. What's the expectation value for the number of second cot deaths per year?

$$Np = 325000 \times (1/169390) = 1.9$$

(Expectation perfectly consistent with data)

Another way:

$$P(\text{Innocent}|\text{Deaths}) = \frac{P(\text{Deaths}|\text{Innocent})P(\text{Innocent})}{P(\text{Deaths}|\text{Innocent})P(\text{Innocent}) + P(\text{Deaths}|\text{Guilty})P(\text{Guilty})}$$

Statistically, there are ~30 children murdered by mothers each year in England and Wales. The number of double murders must be less... assume ~1 per year:

$$P(\text{Guilty}) = \frac{1}{325000} \simeq 3 \times 10^{-6}$$

$$P(\text{Innocent}) = 1 - P(\text{Guilty}) = 0.999997$$

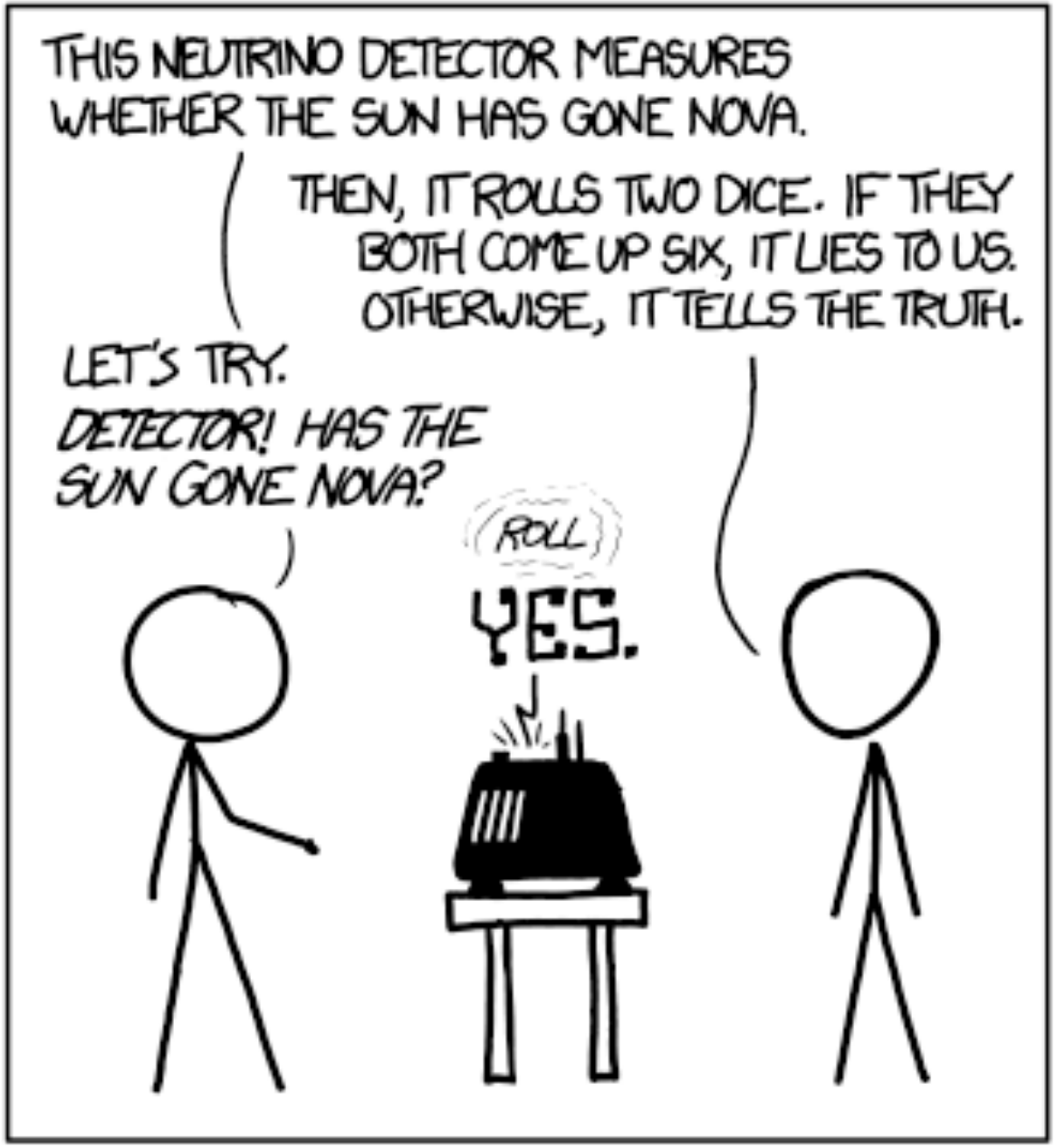
$$P(\text{Deaths}|\text{Innocent}) = \frac{1}{169000} \simeq 6 \times 10^{-6}$$

$$P(\text{Deaths}|\text{Guilty}) = 1$$

$$P(Innocent\,|\,Deaths) = \frac{(6 \times 10^{-6})(0.99999)}{(6 \times 10^{-6})(0.99999) + (1)(3 \times 10^{-6})} \simeq 0.67$$

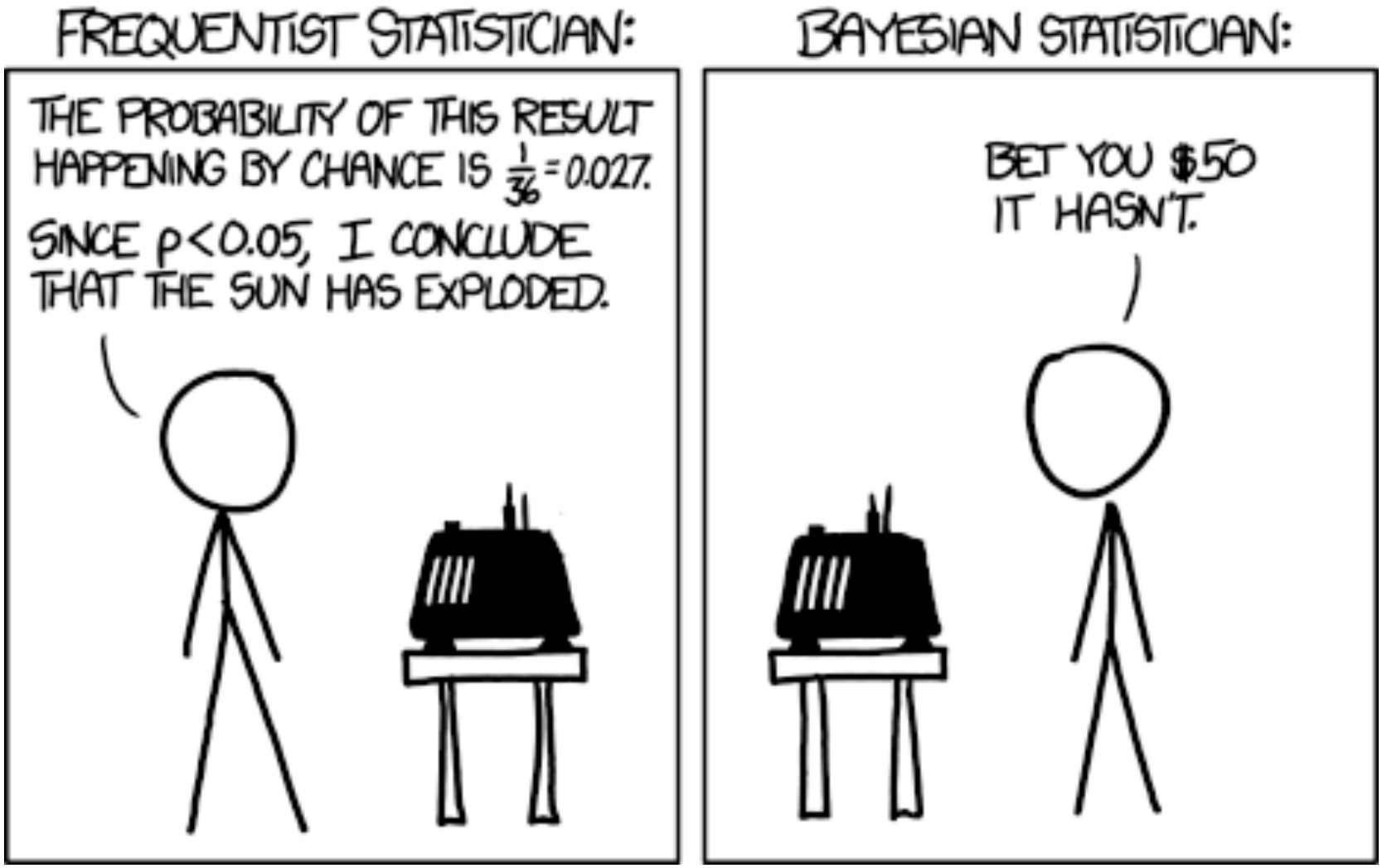
http://xkcd.com/1132/

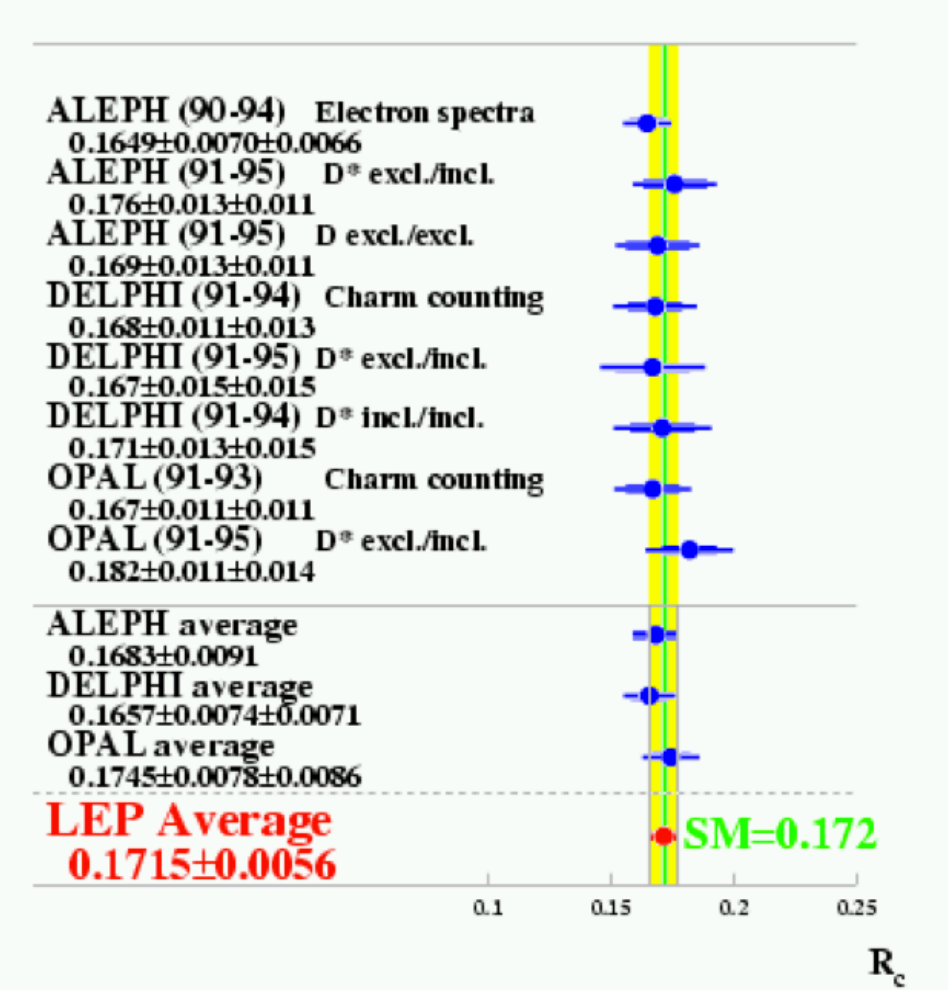
DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)

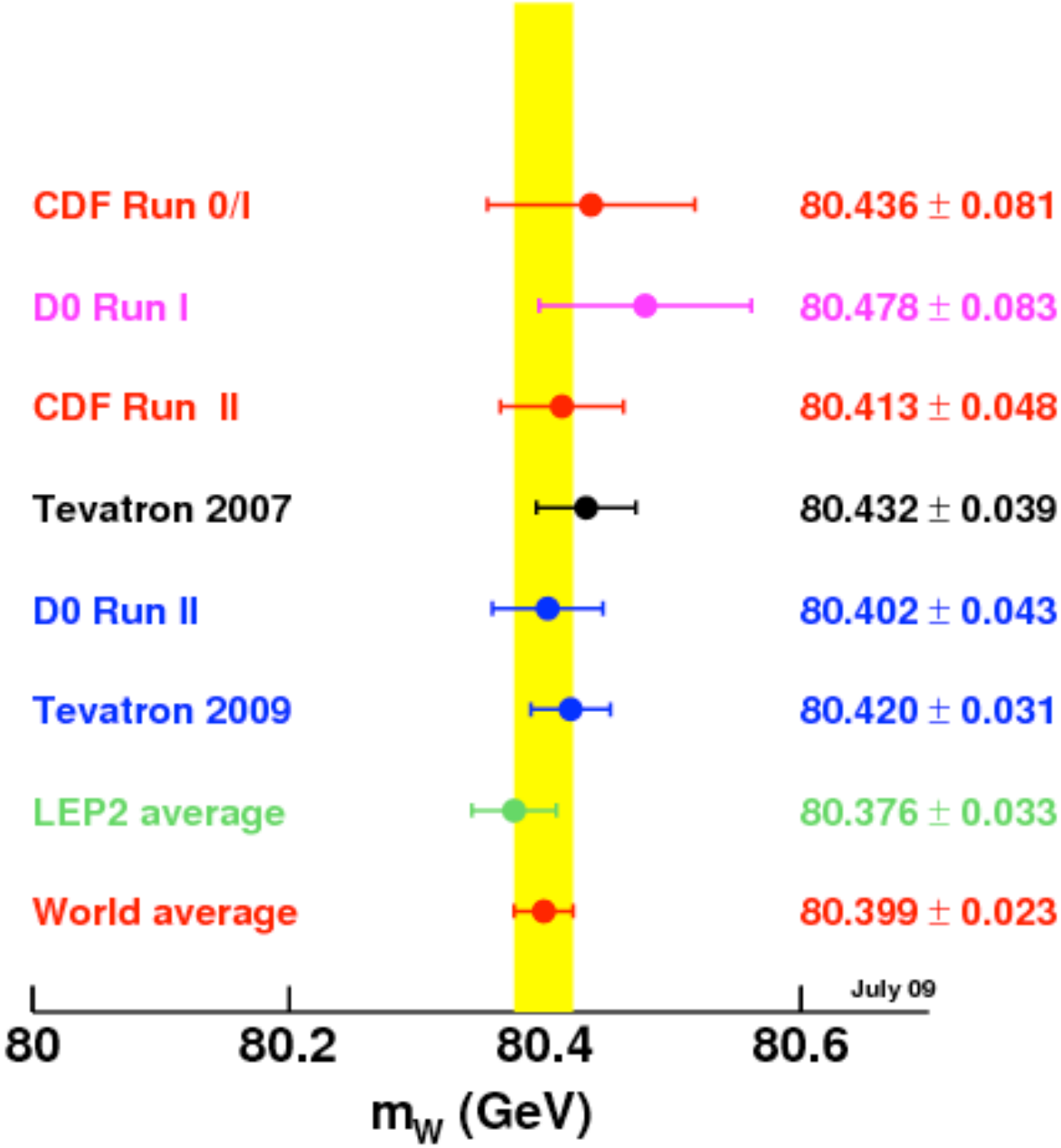


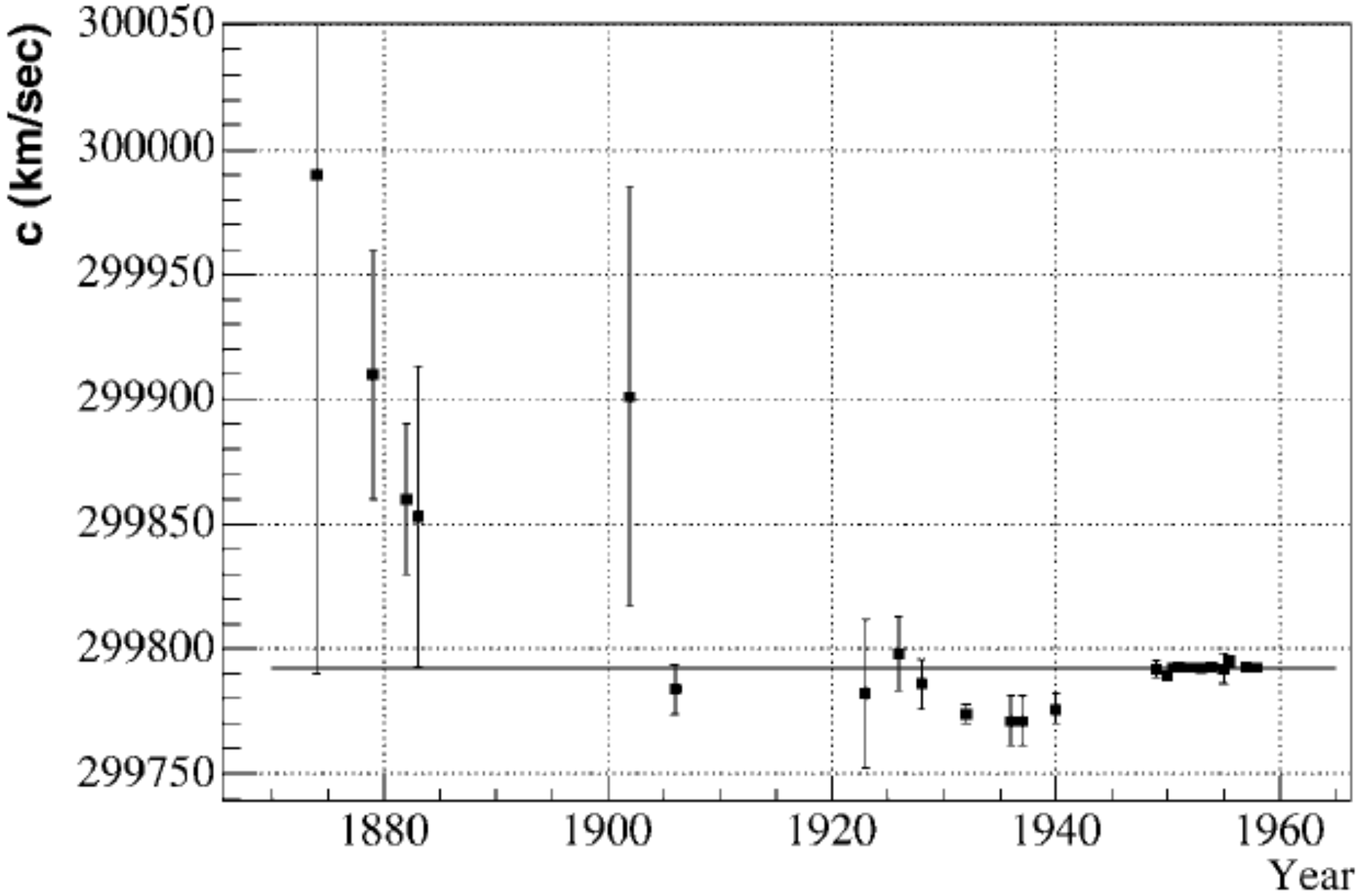
FREQUENTIST STATISTICIAN:

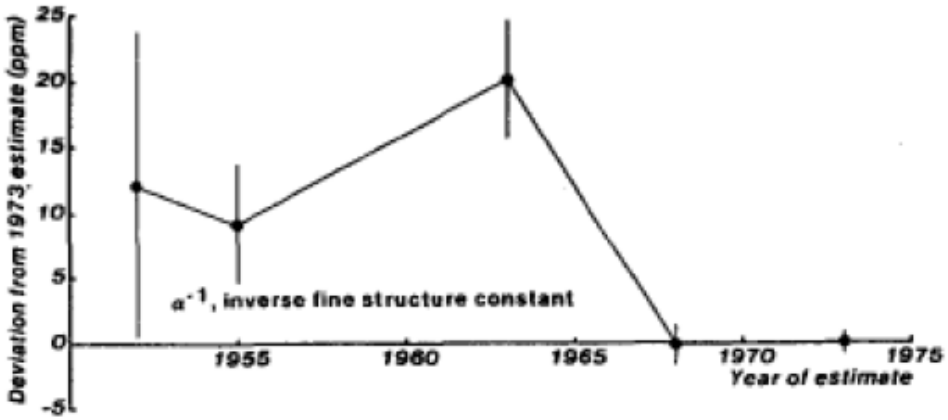
BAYESIAN STATISTICIAN:

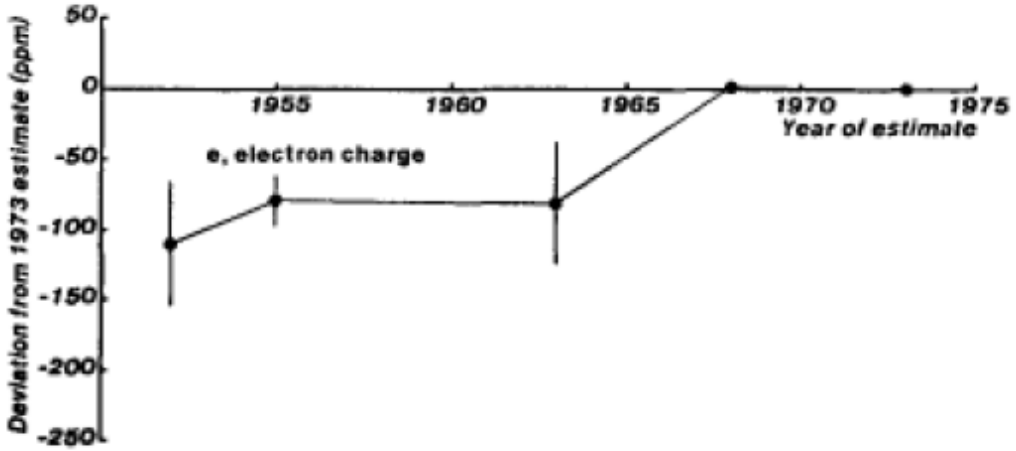

FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36}$ = 0.027.
SINCE p<0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

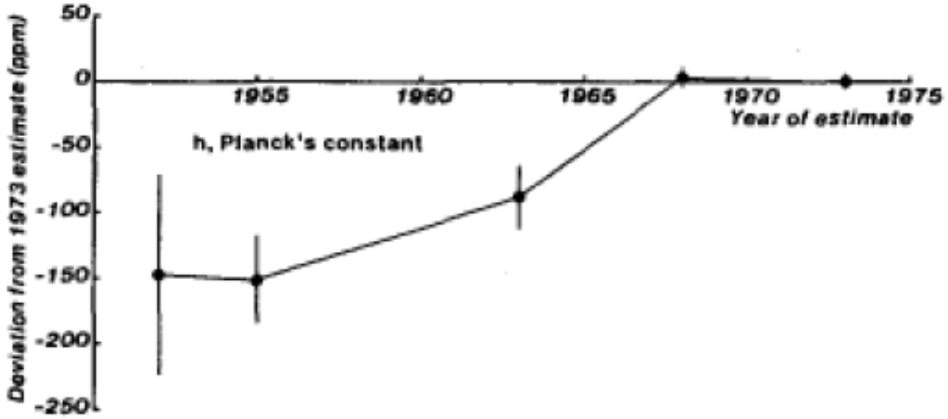

ALEPH (90-94) Electron spectra
0.1649±0.0070±0.0066
ALEPH (91-95) D* excl./incl.
0.176±0.013±0.011
ALEPH (91-95) D excl./excl.
0.169±0.013±0.011
DELPHI (91-94) Charm counting
0.168±0.011±0.013
DELPHI (91-95) D* excl./incl.
0.167±0.015±0.015
DELPHI (91-94) D* incl./incl.
0.171±0.013±0.015
OPAL (91-93) Charm counting
0.167±0.011±0.011
OPAL (91-95) D* excl./incl.
0.182±0.011±0.014
ALEPH average
0.1683±0.0091
DELPHI average
0.1657±0.0074±0.0071
OPAL average
0.1745±0.0078±0.0086
LEP Average
0.1715±0.0056
SM=0.172
0.1
0.15
0.2
0.25
R_c

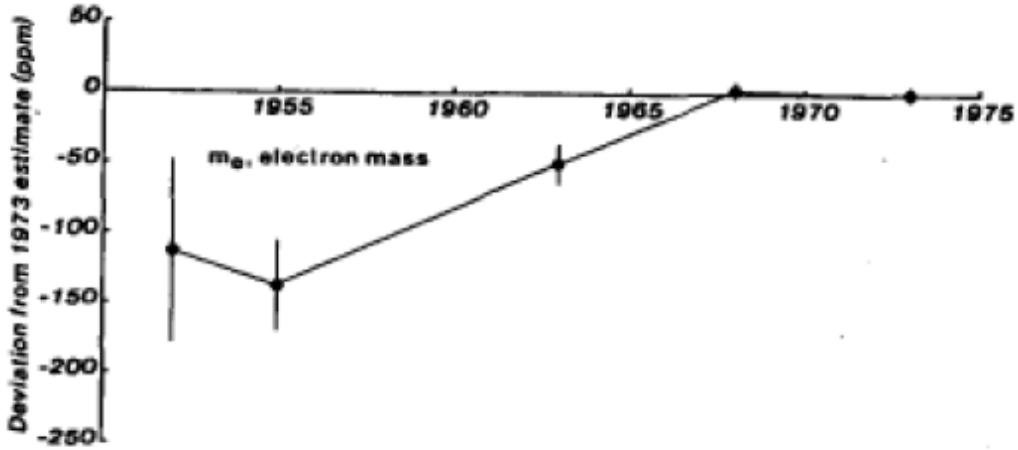

CDF Run 0/I
80.436 ± 0.081
D0 Run I
80.478 ± 0.083
CDF Run II
80.413 ± 0.048
Tevatron 2007
80.432 ± 0.039
D0 Run II
80.402 ± 0.043
Tevatron 2009
80.420 ± 0.031
LEP2 average
80.376 ± 0.033
World average
80.399 ± 0.023
July 09
80
80.2
80.4
80.6
m_W (GeV)

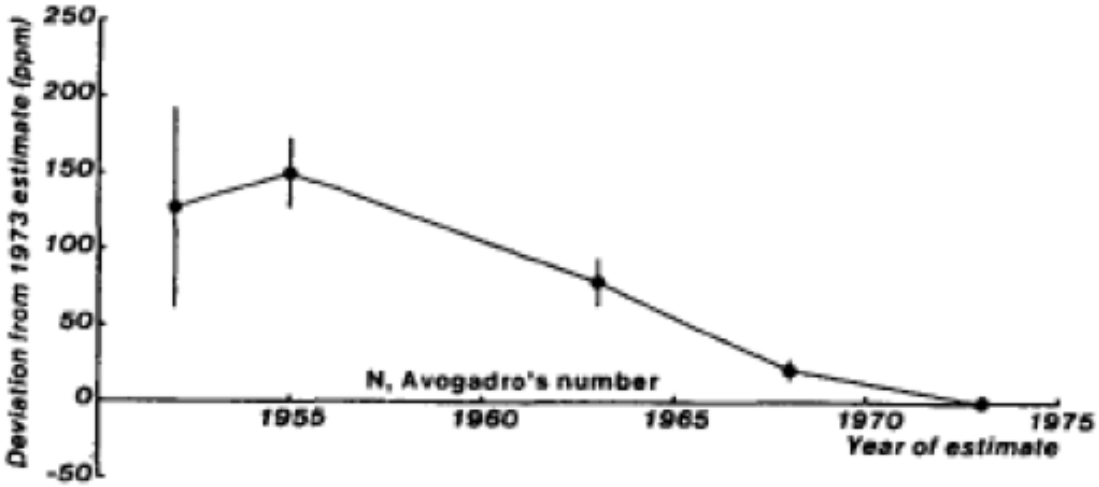

c (km/sec)
300050
300000
299950
299900
299850
299800
299750
1880
1900
1920
1940
1960
Year

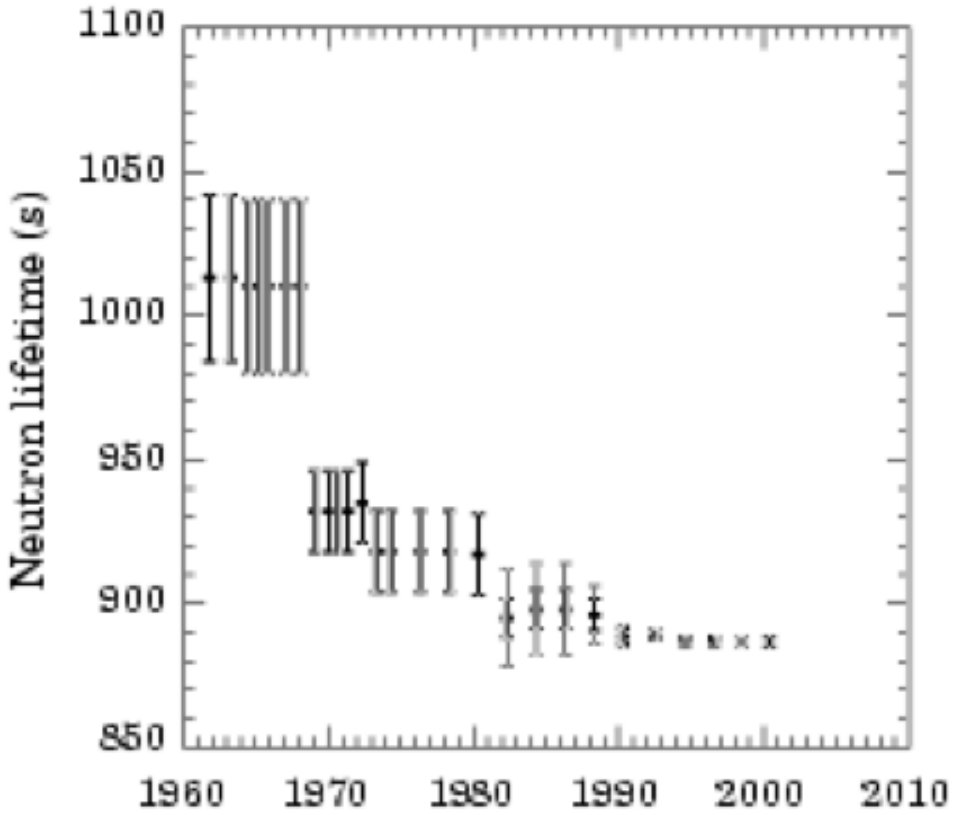

Deviation from 1973 estimate (ppm)
25
20
15
10
5
0
-5
1955
1960
1965
1970
1975
Year of estimate
α⁻¹, inverse fine structure constant

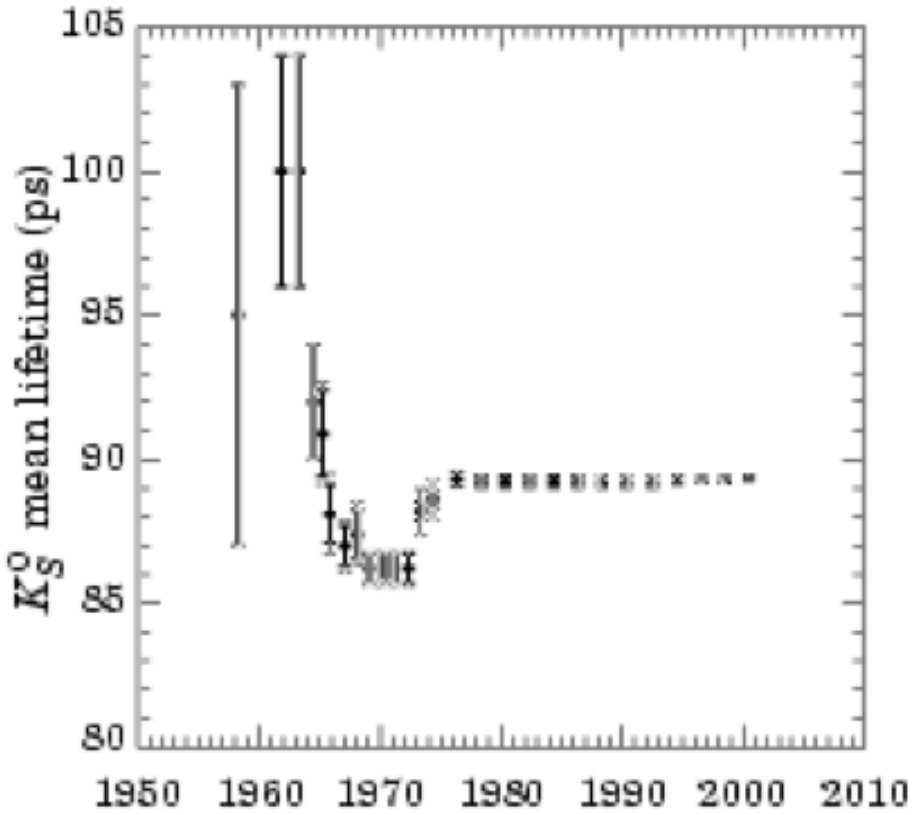

Deviation from 1973 estimate (ppm)
50
0
-50
-100
-150
-200
-250
1955
1960
1965
1970
1975
Year of estimate
e, electron charge


Deviation from 1973 estimate (ppm)
50
0
-50
-100
-150
-200
-250
1955
1960
1965
1970
1975
Year of estimate
h, Planck's constant

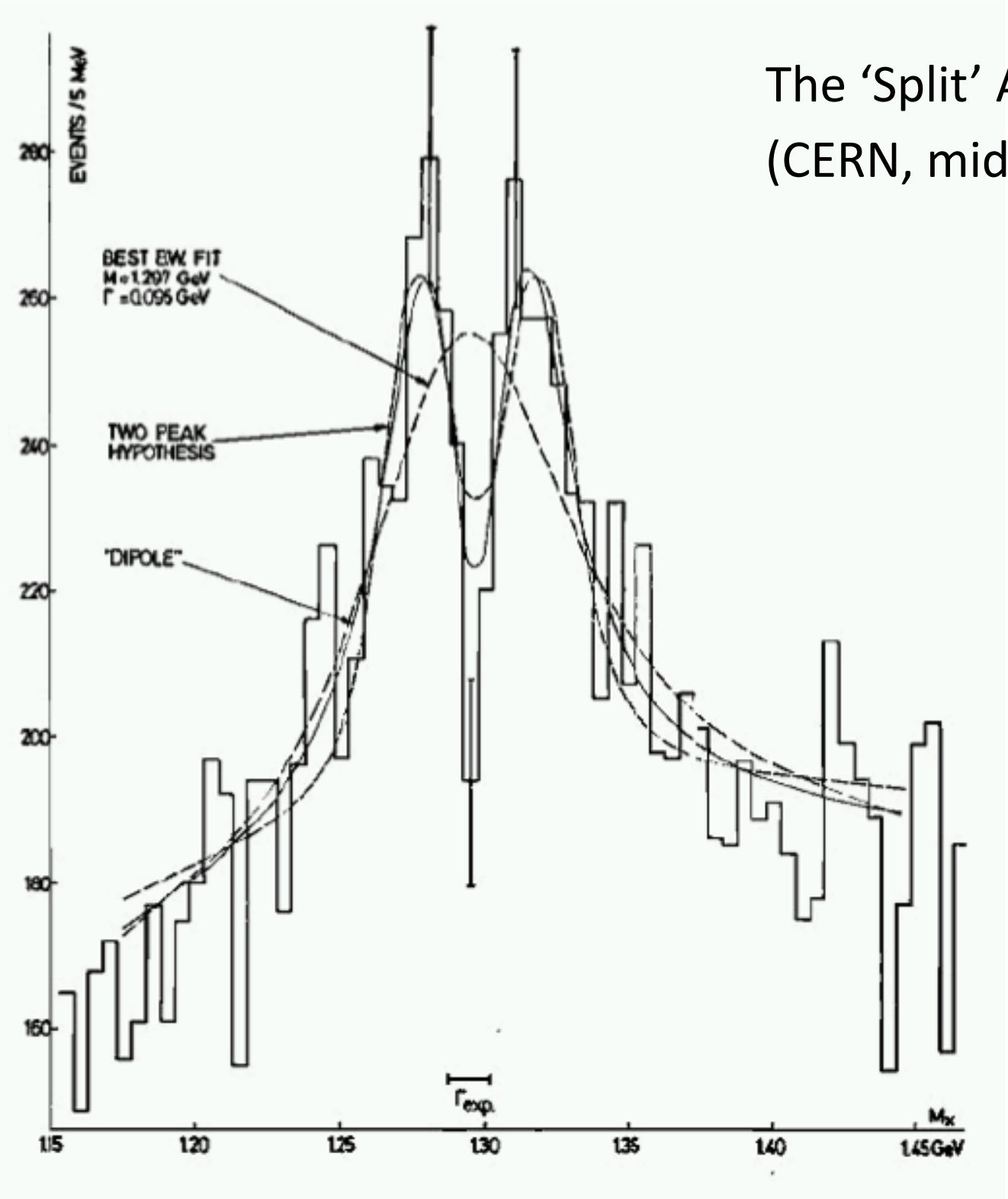

Deviation from 1973 estimate (ppm)
50
0
-50
-100
-150
-200
-250
1955
1960
1965
1970
1975
mₑ, electron mass


Deviation from 1973 estimate (ppm)
250
200
150
100
50
0
-50
1955
1960
1965
1970
1975
Year of estimate
N, Avogadro's number


Neutron lifetime (s)
1100
1050
1000
950
900
850
1960
1970
1980
1990
2000
2010


K_S^0 mean lifetime (ps)
105
100
95
90
85
80
1950
1960
1970
1980
1990
2000
2010

The 'Split' A_2 Meson (CERN, mid 1960's)

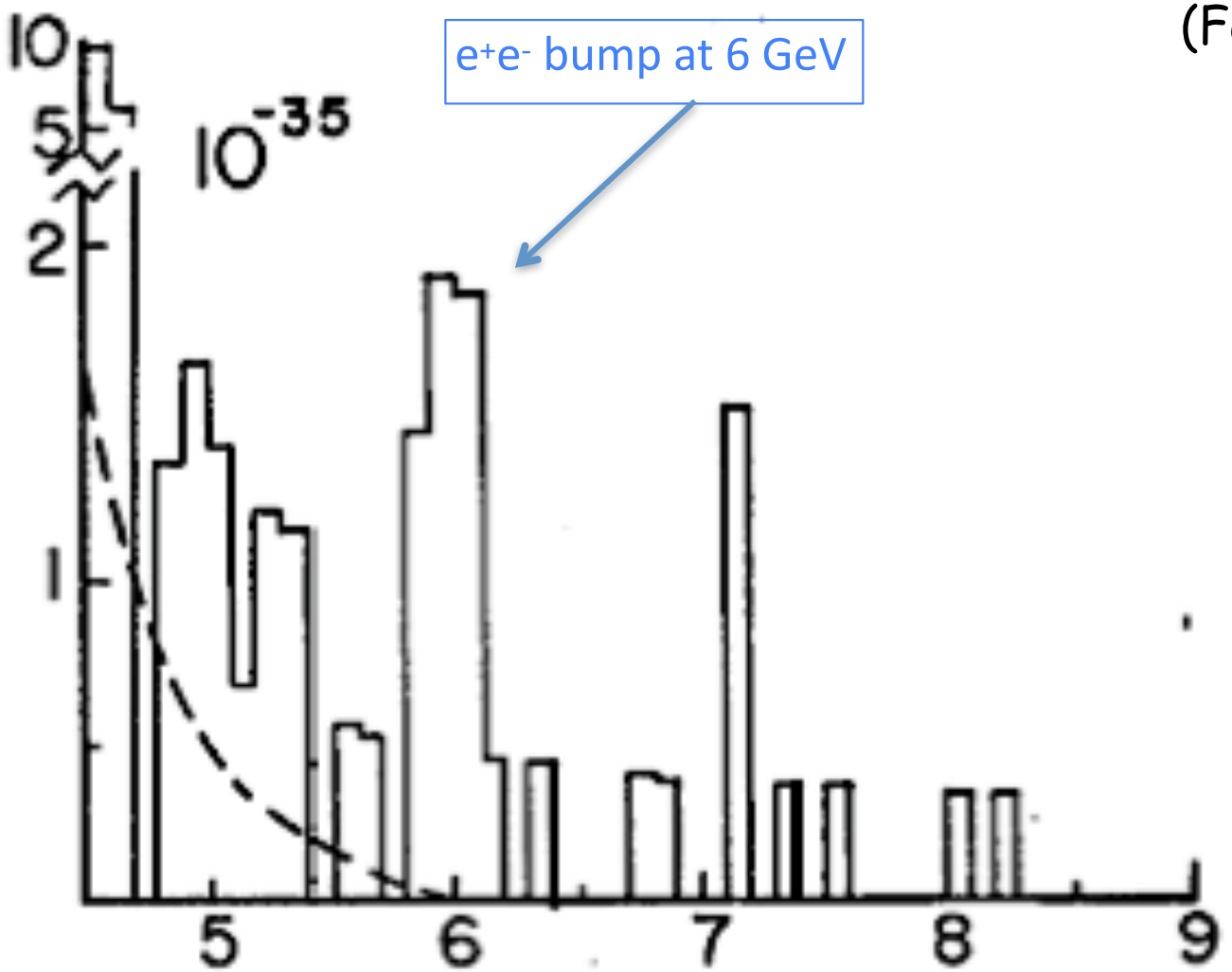


biased data selection

Υ

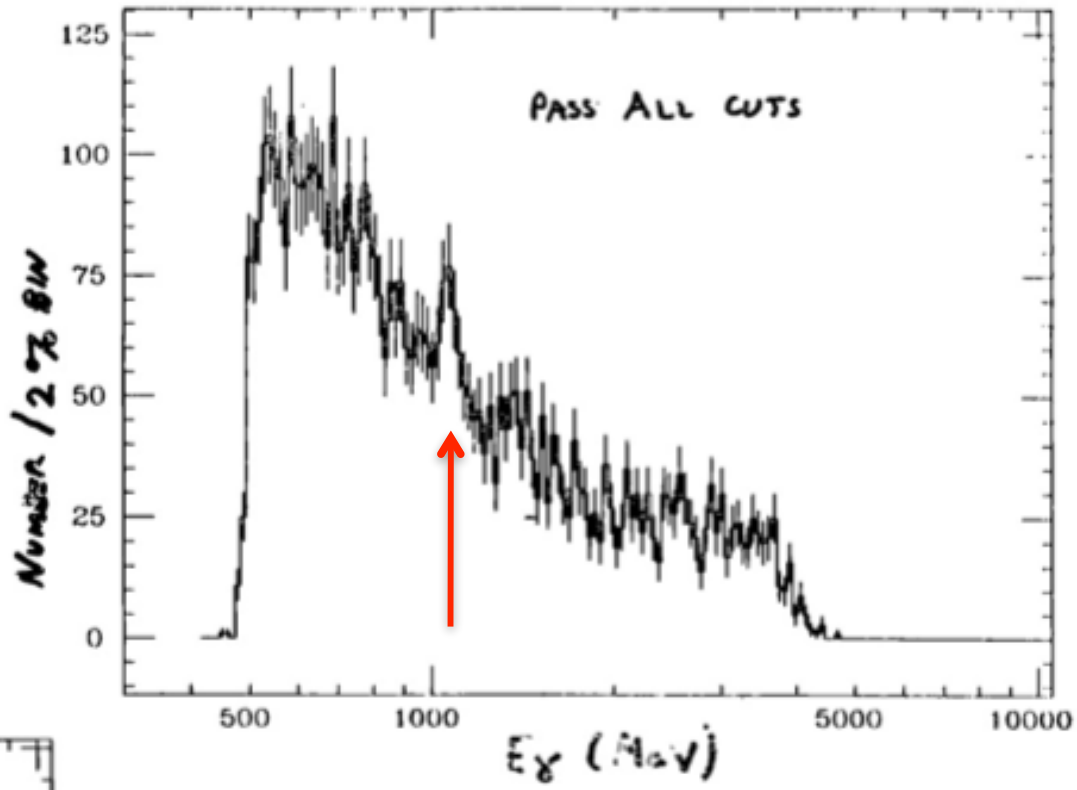
"Oops-Leon"

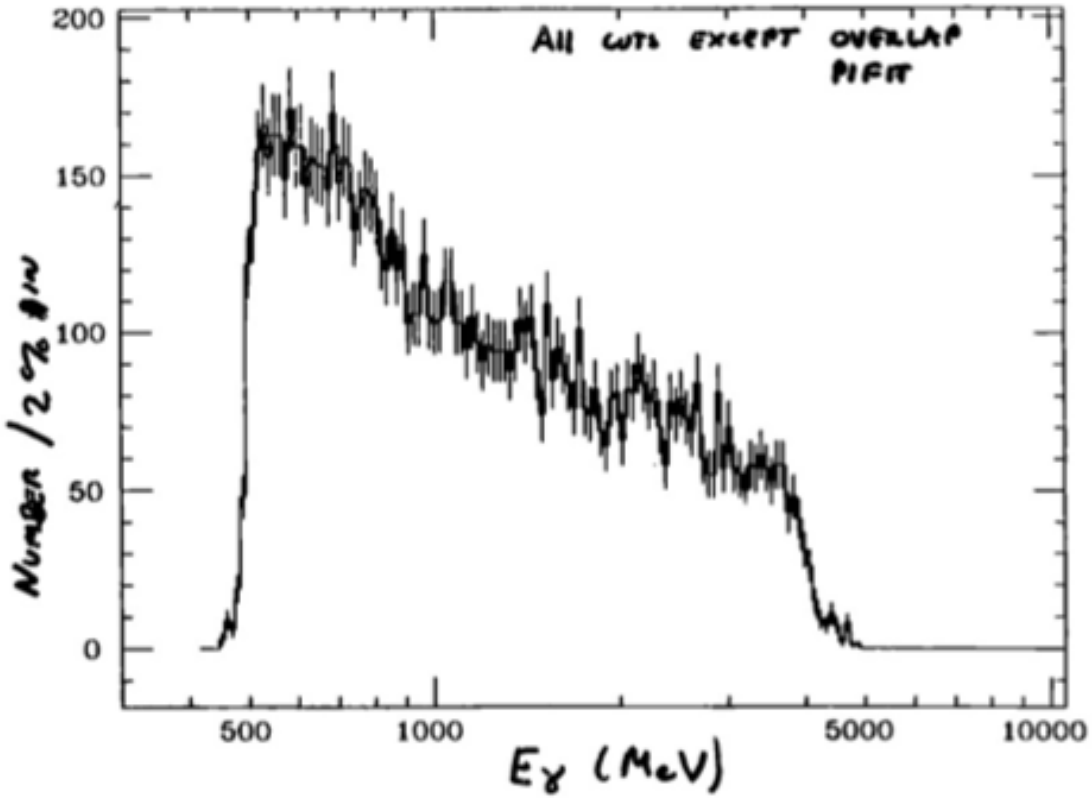
(Fermilab, 1976)



Statistical fluctuation

The ζ Particle
(DESY, 1984)





Biased data cuts

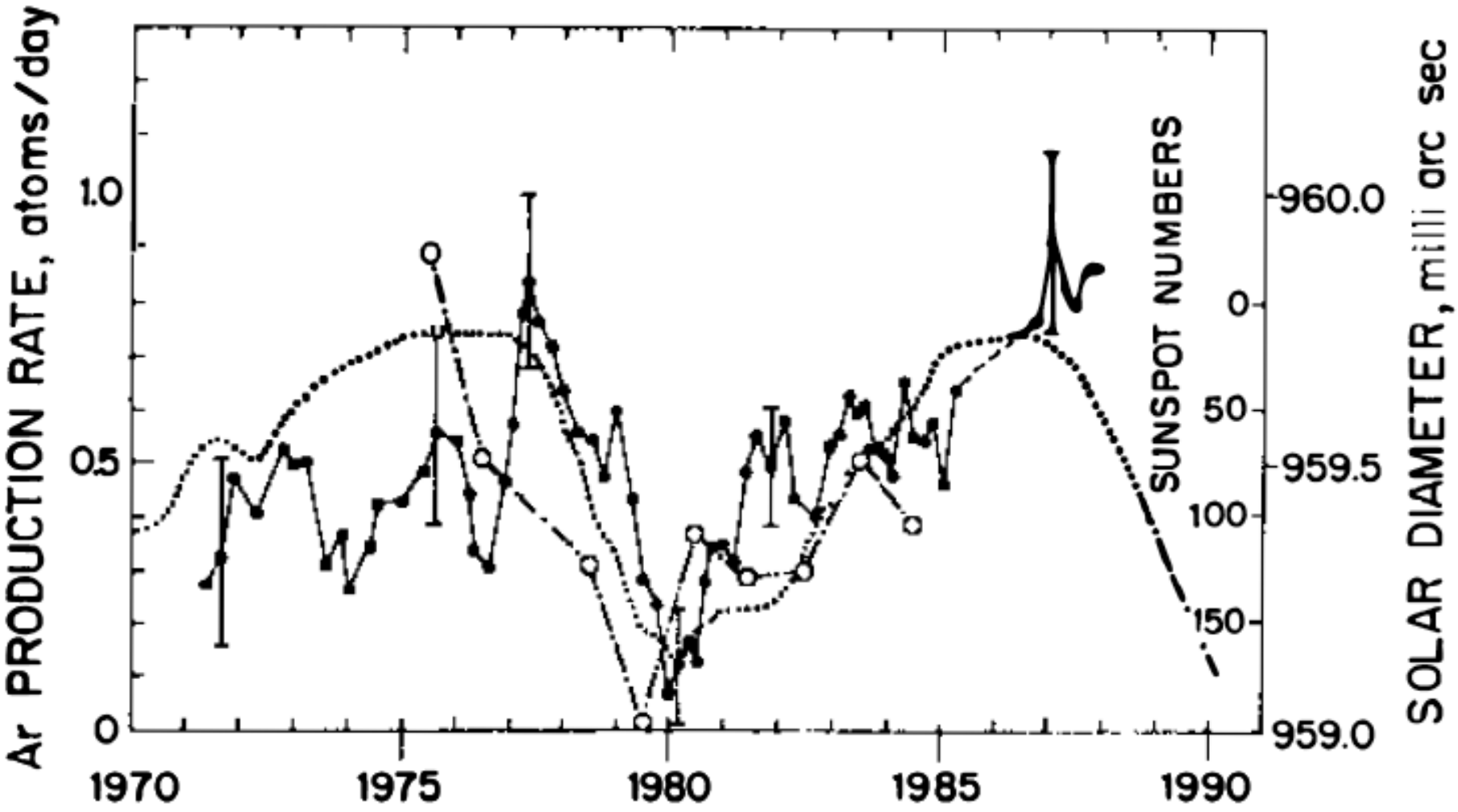
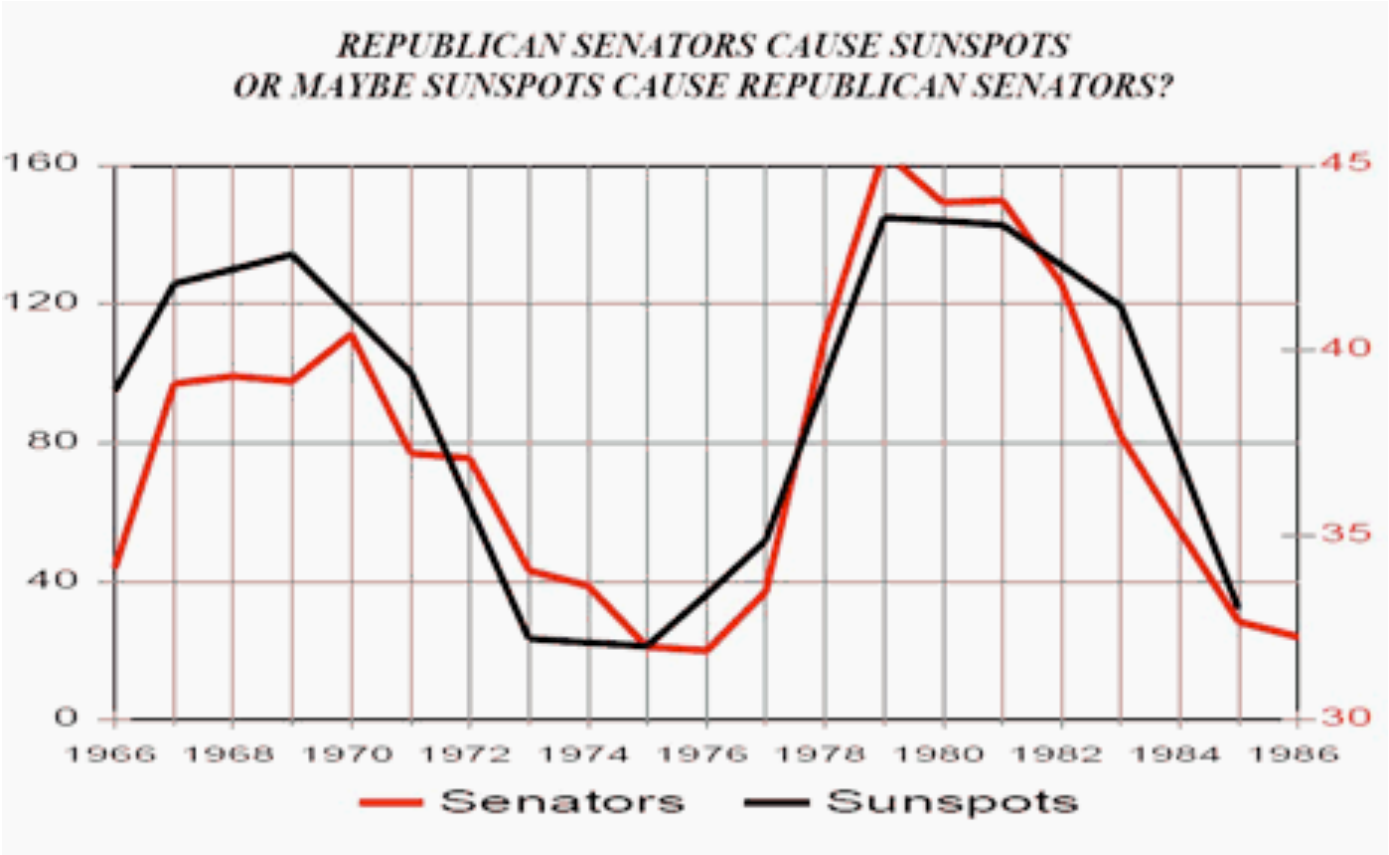


Figure 5 Plots of five-point running average of ^{37}Ar production and smoothed sunspot numbers against time in years (from 130). Solid circles, ^{37}Ar production; dotted curve, sunspot numbers; open circles, solar diameter.


REPUBLICAN SENATORS CAUSE SUNSPOTS
OR MAYBE SUNSPOTS CAUSE REPUBLICAN SENATORS?
160
120
80
40
0
45
40
35
30
1966
1968
1970
1972
1974
1976
1978
1980
1982
1984
1986
Senators
Sunspots

CERN, 2011

NewScientist

Surprise LHC blip hints at Higgs – again

22:49 22 July 2011

"...The combined statistical significance, taking all three types of excess reported by ATLAS into account, is 2.8 sigma, slightly below the 3 sigma threshold (equivalent to a 1-in-370 chance of being due to a fluke) that a measurement must pass to count as "evidence" for something new: only 5 sigma data, equivalent to a 1-in-1.7 million chance of being due to a fluke, gains "discovery" status.

The other main detector at the LHC, called CMS, has found an excess in a similar range, between 130 and 150 GeV, reports Nature. The size of that excess is roughly 2 sigma, writes physicist Adam Falkowski on the Resonances blog.

If all this sounds a tad familiar, rewind back to April, when four physicists claimed to have found hints of the Higgs in ATLAS data in a study abstract leaked online. A subsequent official analysis by the collaboration of 700 physicists who run ATLAS concluded that result was an error. Unlike that claim, the new excesses have been vetted by the ATLAS and CMS collaborations respectively."

---

guardian.co.uk

Higgs boson signals fade at Large Hadron Collider

Cern scientist says he sees 'no striking evidence of anything that could resemble a discovery' in hunt for Higgs boson

Ian Sample
guardian.co.uk, Monday 22 August 2011 17.10 BST
Article history

Bias

and Experimental Design

“Blind” Analysis Techniques

Goal: To remove the ability to unconsciously tune on statistical fluctuations and/or adjust analyses towards a particular outcome by hiding the final result until the full analysis (incl. assessment of uncertainties) is fixed.

At which point you then “open the box” and take what life brings you!

Rules of the Game

- **Agree on an appropriate blindness scheme in advance**
- **Make sure no one breaks it**
- **Agree on the criteria necessary to “open the box”**
- **State the blindness scheme up front in any publication**
- **Agree to show exactly what results from box-opening and then justify any alterations**

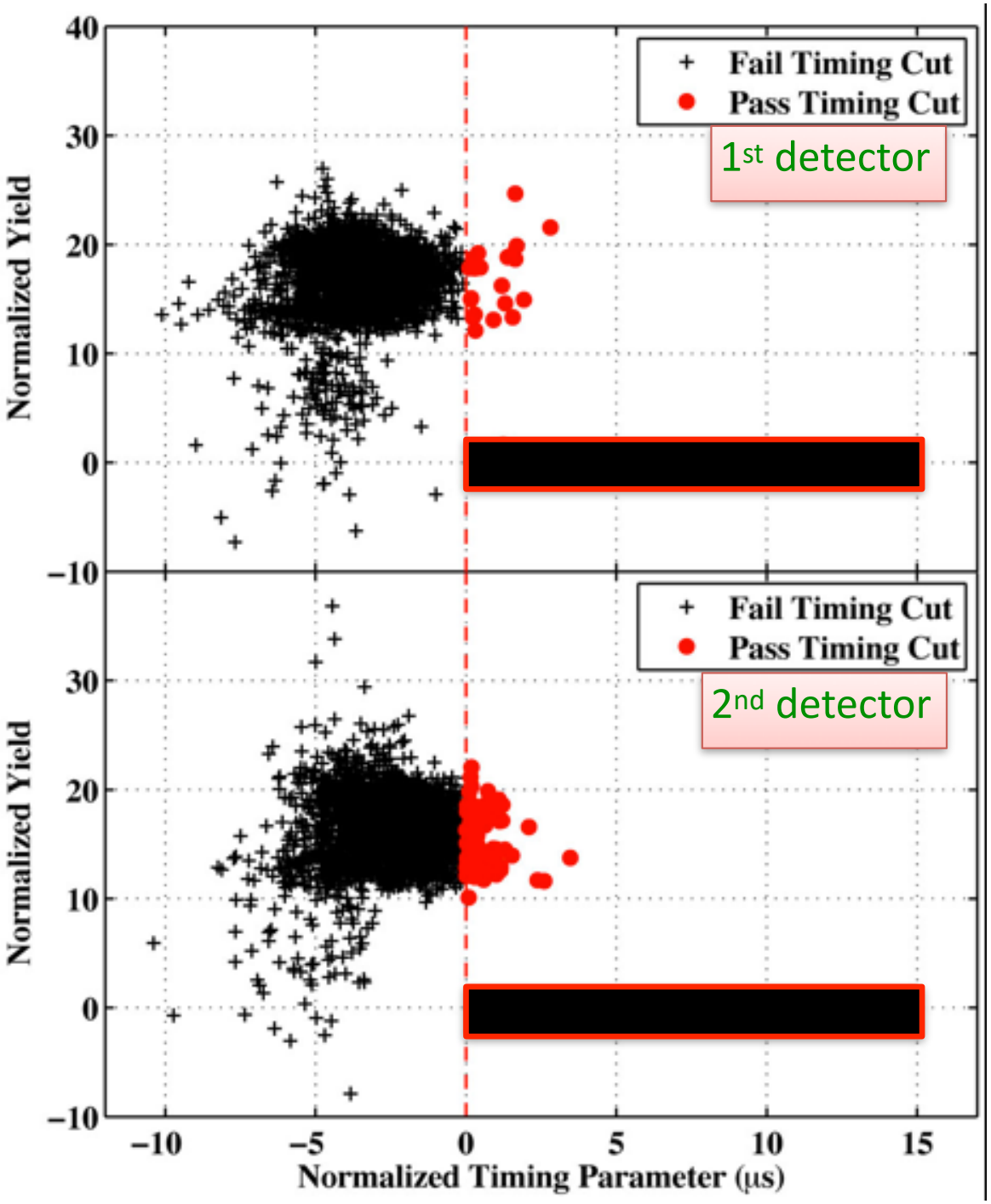
Signal Box Method

CDMS results on search for Dark Matter (Dec, 2009)

Expected summed background in both detectors: **0.9 ± 0.2**

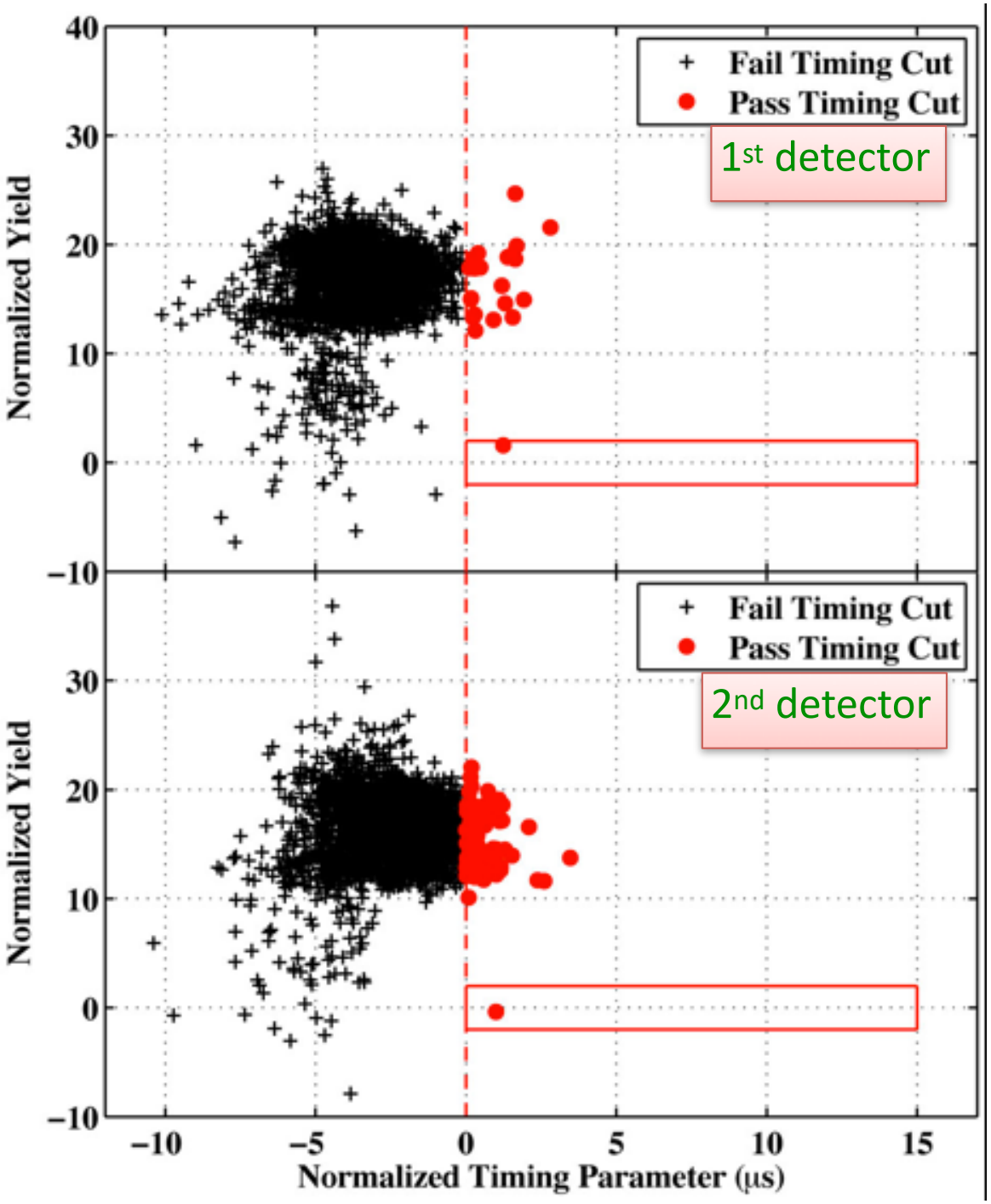
RESULTS:

[redacted]



Signal Box Method

CDMS results on search for Dark Matter (Dec, 2009)



Expected summed background in both detectors: **0.9 ± 0.2**

RESULTS:

2

(consistent with background fluctuations)

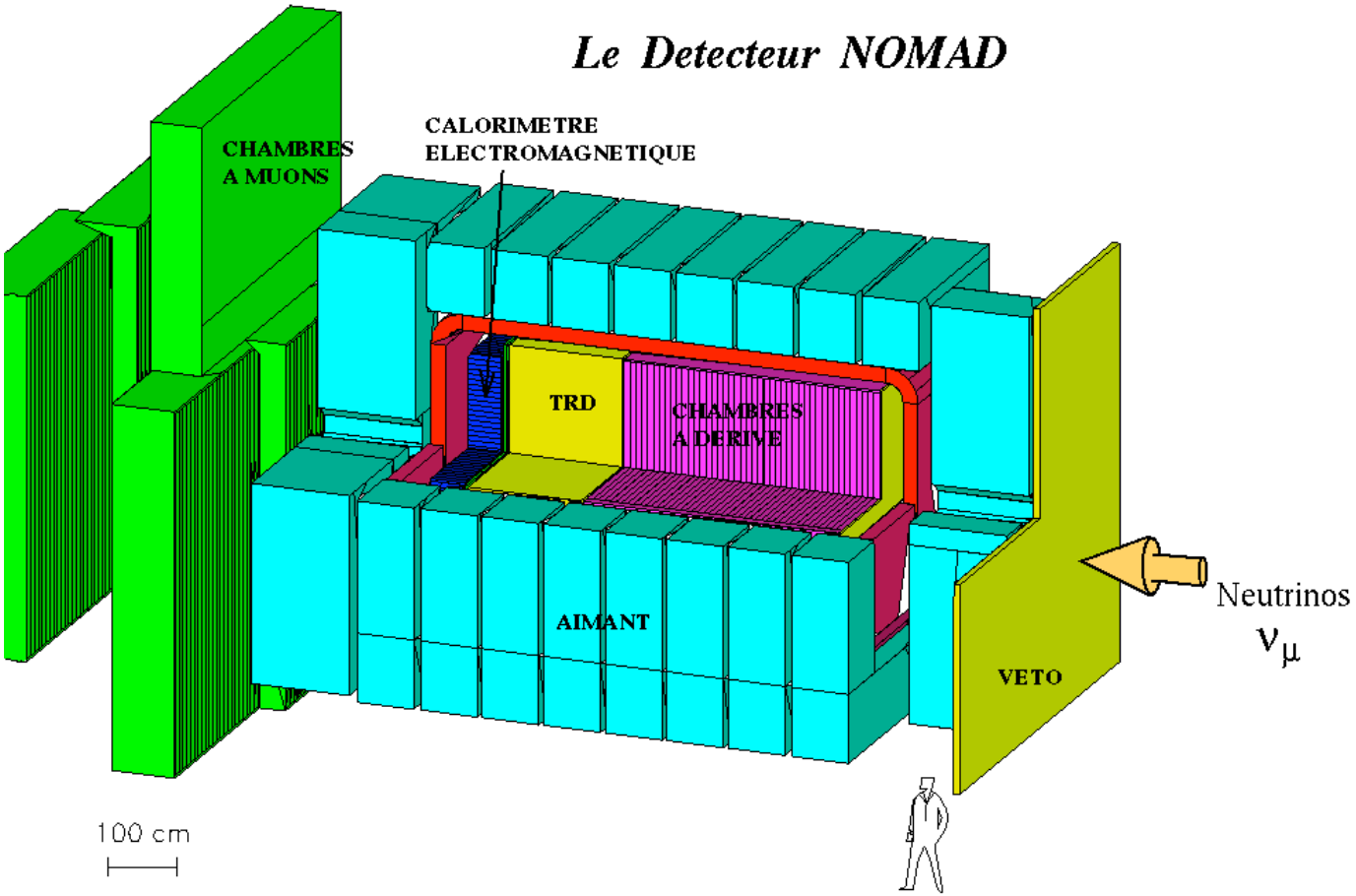
Divided Data Sample

NOMAD Search for

$\nu_\mu - \nu_\tau$ oscillations

(Feb, 1999)

Used 20% of data to confirm background predictions and define search window, then impose signal box method on remaining 80% of the data



RESULTS:

Expected background in signal box: **6.5 ± 1.1**

[redacted]

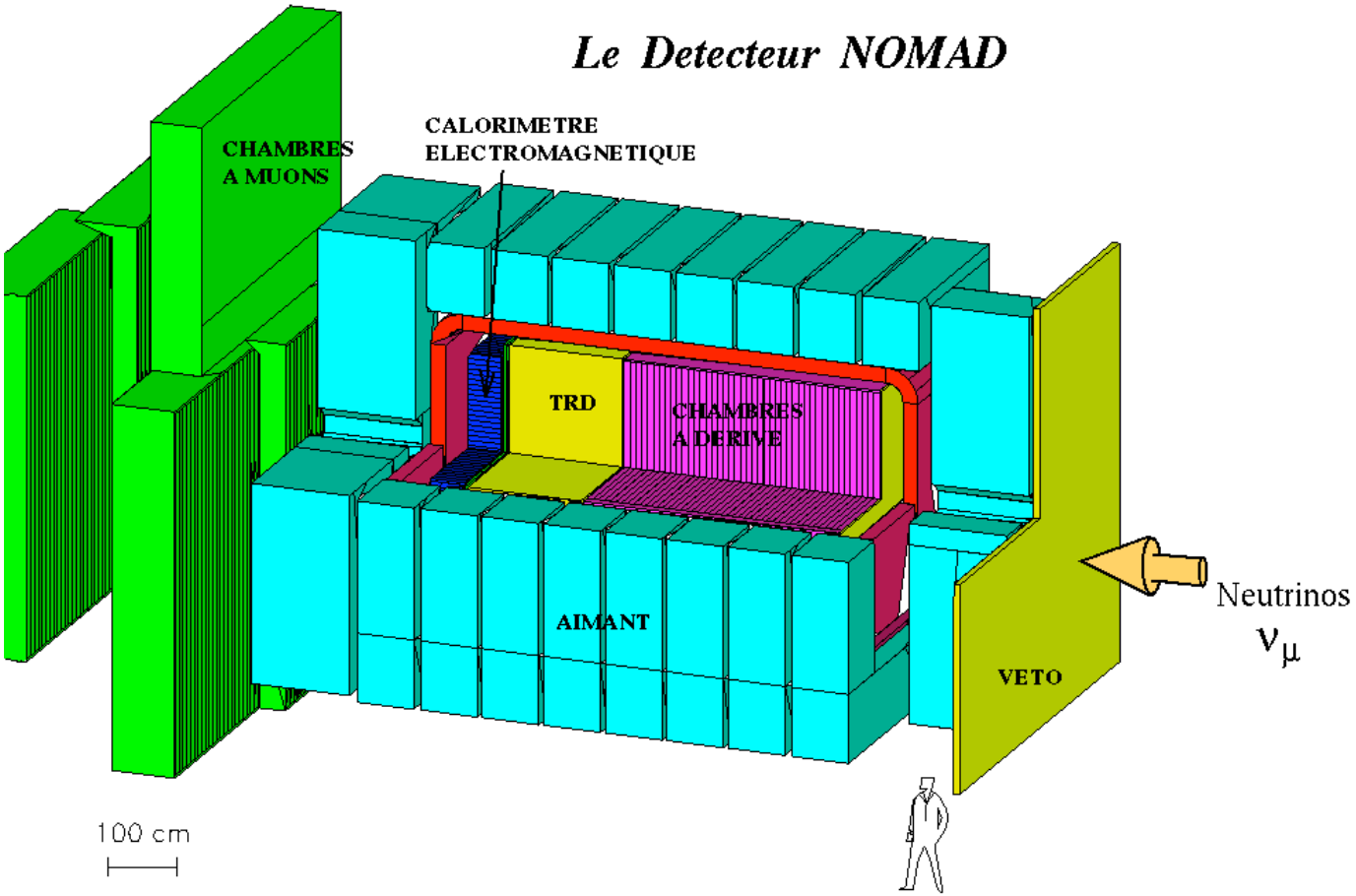
Divided Data Sample

NOMAD Search for

$\nu_\mu - \nu_\tau$ oscillations

(Feb, 1999)

Used 20% of data to confirm background predictions and define search window, then impose signal box method on remaining 80% of the data



RESULTS:

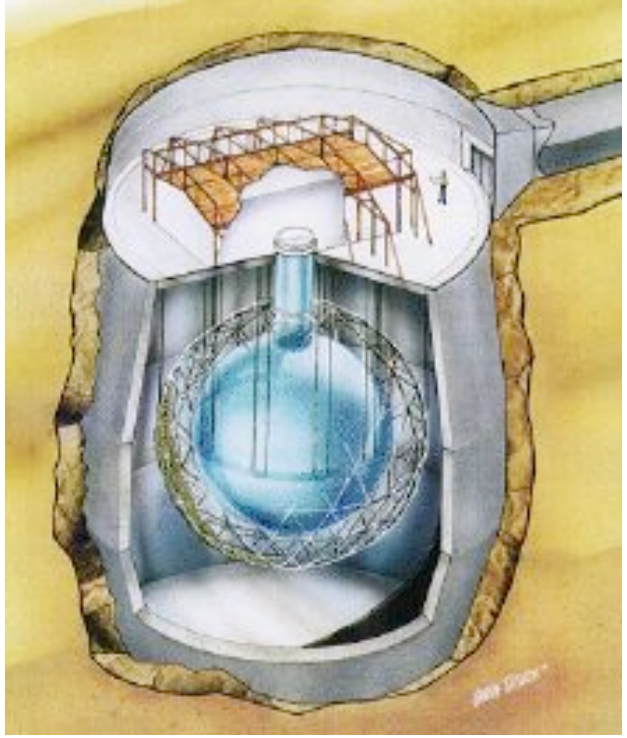
5

(consistent with background fluctuations)

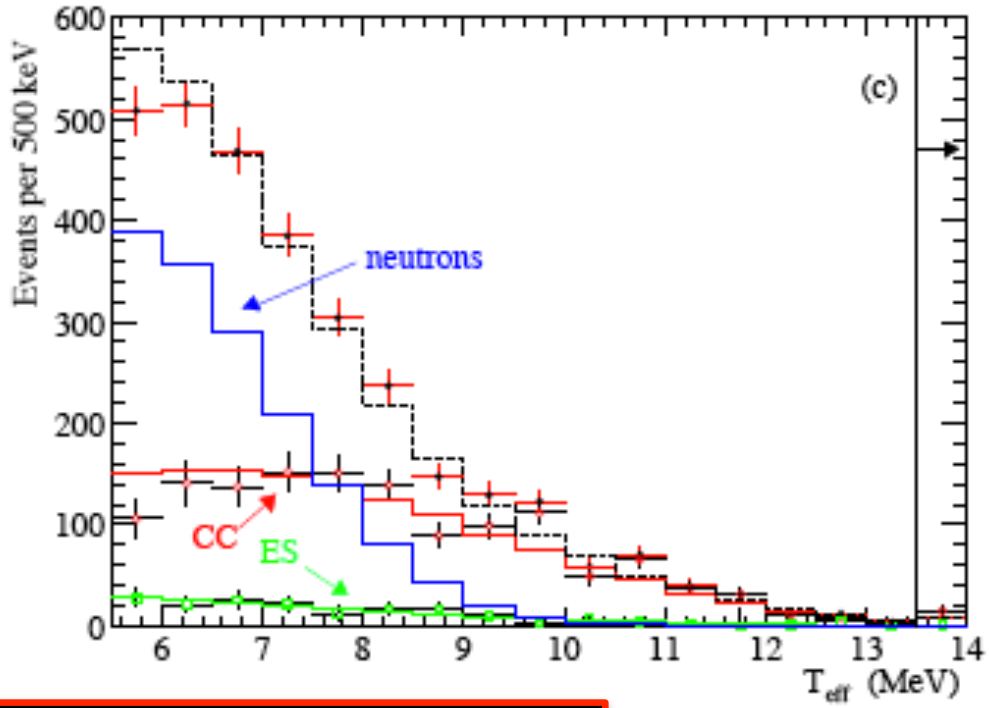
Expected background in signal box: **6.5 ± 1.1**

Hidden Parameters

SNO Measurement of total solar neutrino flux (Sept, 2003)



Excluded a hidden fraction of the final data set (unknown flux normalisation), included hidden admixture of tagged background neutrons, scaled simulation NC cross section by hidden factor

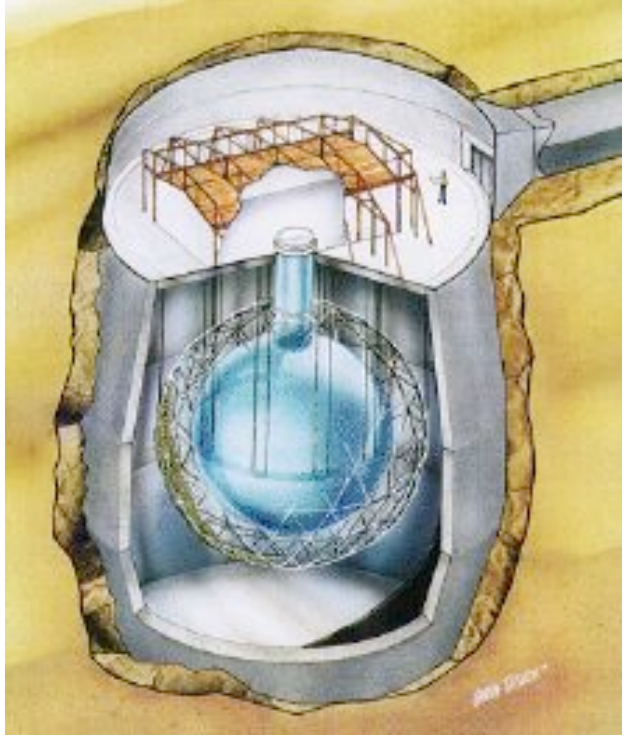


RESULTS:

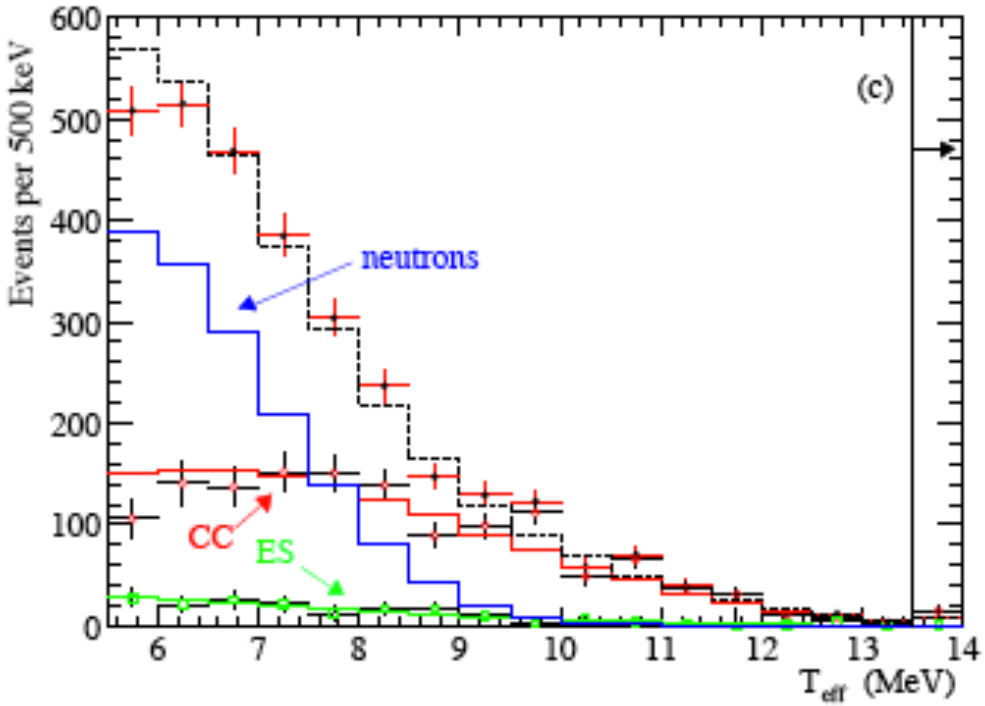
[redacted]

Hidden Parameters

SNO Measurement of total solar neutrino flux (Sept, 2003)



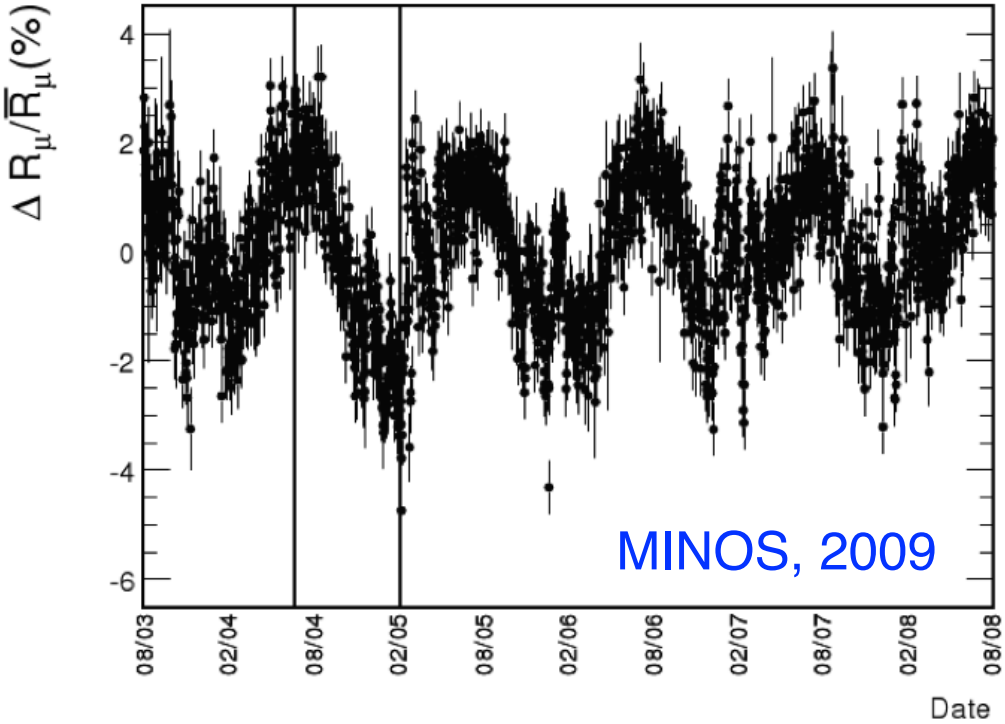
Excluded a hidden fraction of the final data set (unknown flux normalisation), included hidden admixture of tagged background neutrons, scaled simulation NC cross section by hidden factor

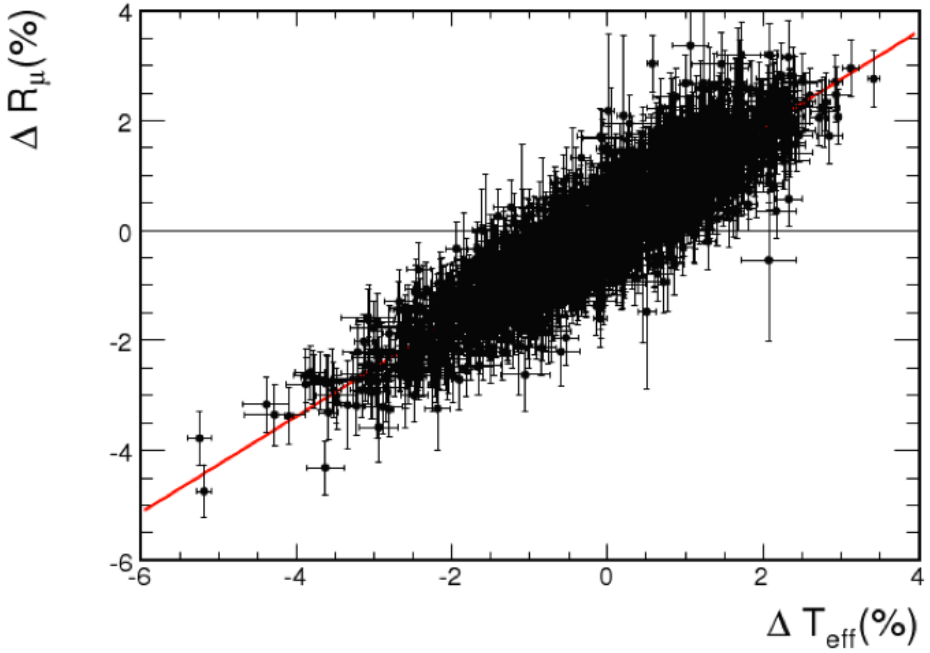


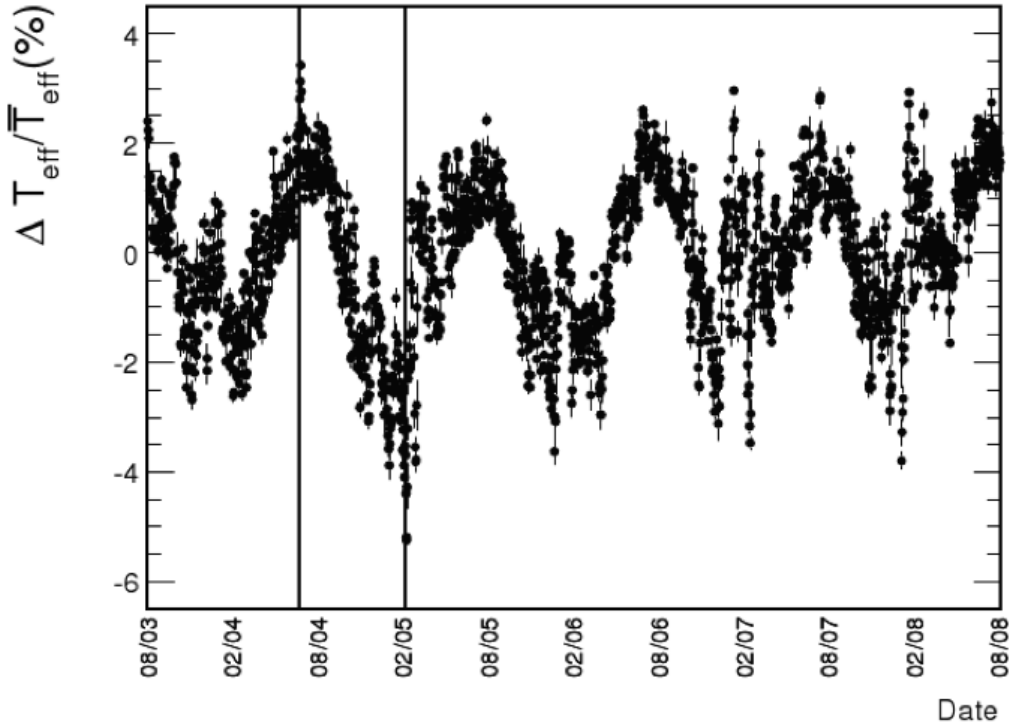
RESULTS:

$$\Phi_\nu = 5.21 \pm 0.27_{(stat)} \pm 0.38_{(sys)} \times 10^6/cm2/s$$

Post-Data Corrections





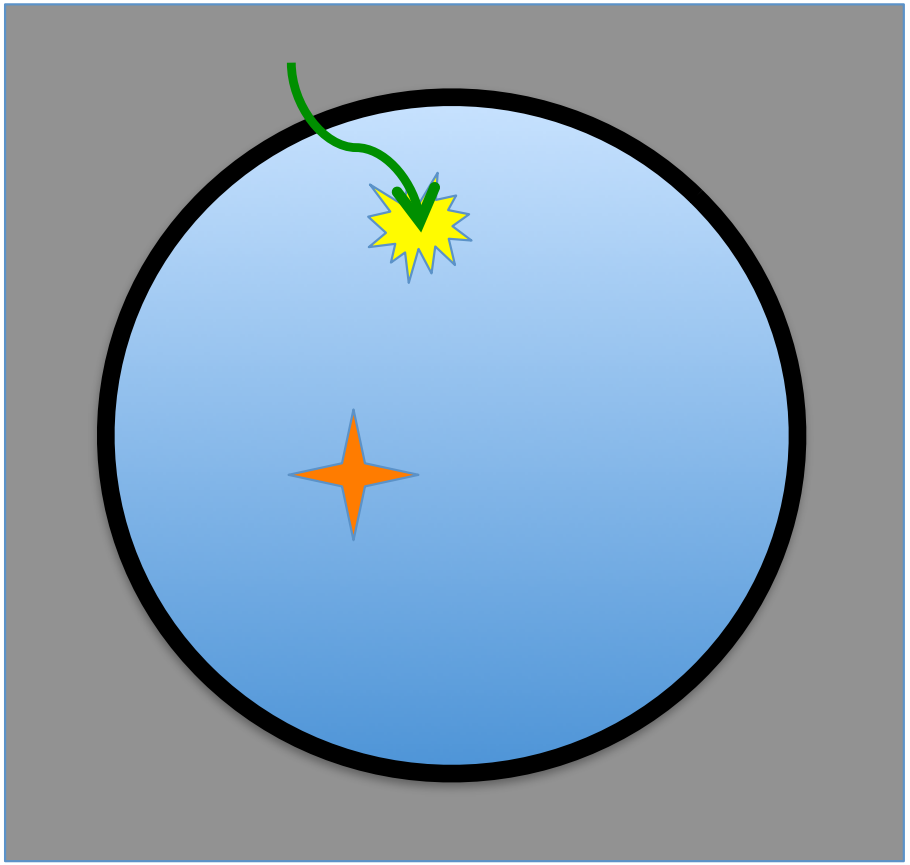


Correlation can be used
to correct model prediction

Statistical Optimisation

Working Backwards

Assume that both the signal and background levels are proportional to the detector mass, M, and running time, T. Find an expression for the maximum background level that can be tolerated to achieve a 3σ detection as a fraction of the expected signal for a given model. How does the sensitivity change as a function of M and T ?



$$B = fS$$

$$1\sigma = \sqrt{B} = \sqrt{fS}$$

under H0

Thus, for a 3σ signal: $$3\sqrt{fS} = S$$

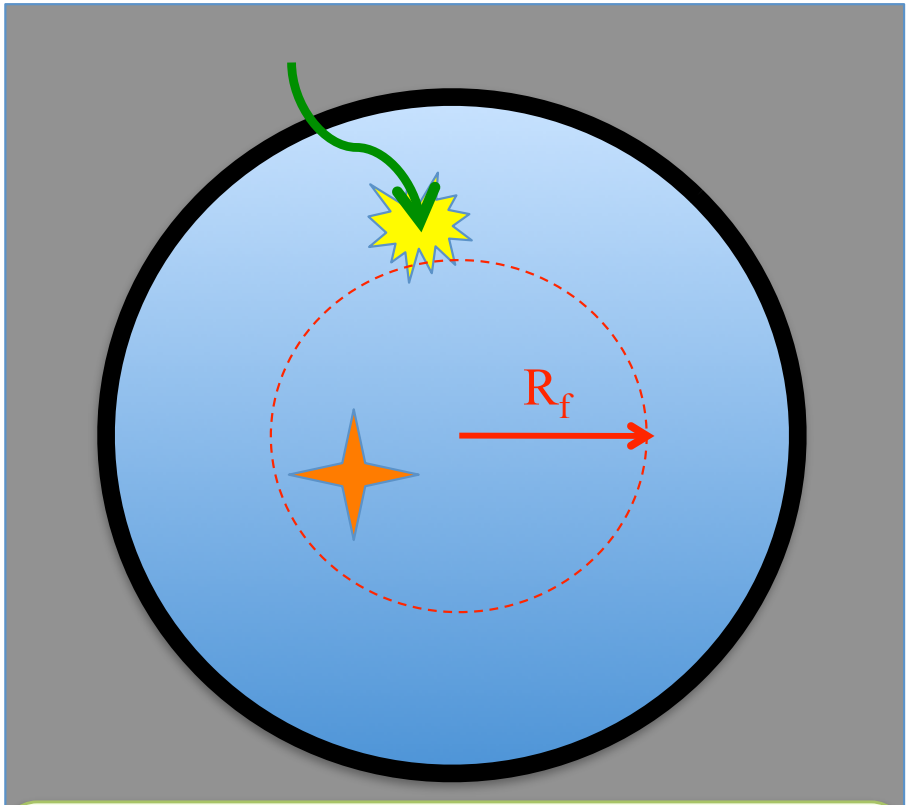
(able to tolerate more background for larger signal) $$f = \frac{S}{9}$$

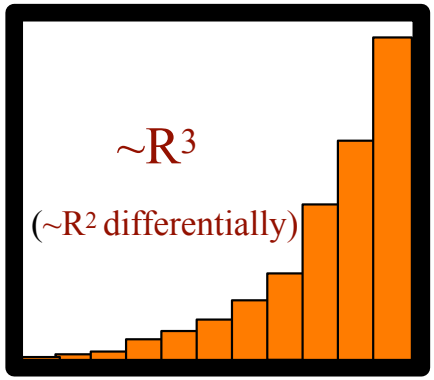
or $$B = \frac{S^2}{9}$$

$$\text{Significance } (\sigma\text{'s}) = \frac{S}{\sqrt{B}}$$

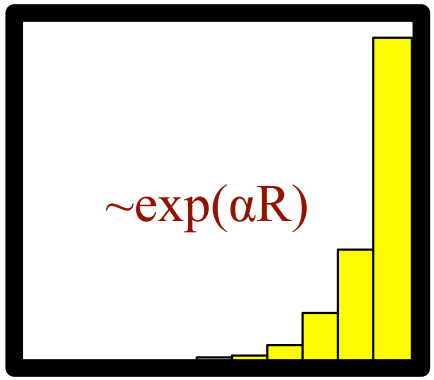
$$= \frac{\alpha MT}{\sqrt{f\alpha MT}} \propto \sqrt{MT}$$

Example of Statistical Optimisation





"Radius"



"Radius"

Assume that we are in the "large N" limit and expected the number of counts to be dominated by background events.

We wish to exclude the worst of the background by choosing a radius to define a "fiducial volume," within which will look for an excess of events as evidence of a signal.

What choice of fiducial radius will give the best sensitivity for the search?

$$\frac{S}{\sqrt{B}} \sim \frac{R^3}{\sqrt{\exp(\alpha R)}} = R^3 e^{-\alpha R/2}$$

maximise:

$$3R^2 e^{-\alpha R/2} - \frac{\alpha}{2} R^3 e^{-\alpha R/2} = 0$$

$$3R^2 = \frac{\alpha}{2} R^3 \qquad R = \frac{6}{\alpha}$$

From the plot, it looks like backgrounds fall by ~1/e when R changes by 10% of the detector radius... so α ~10

$$R_f = 0.6 R_d$$

Redundancy & Estimating Systematics

Sudbury Neutrino Observatory (SNO)

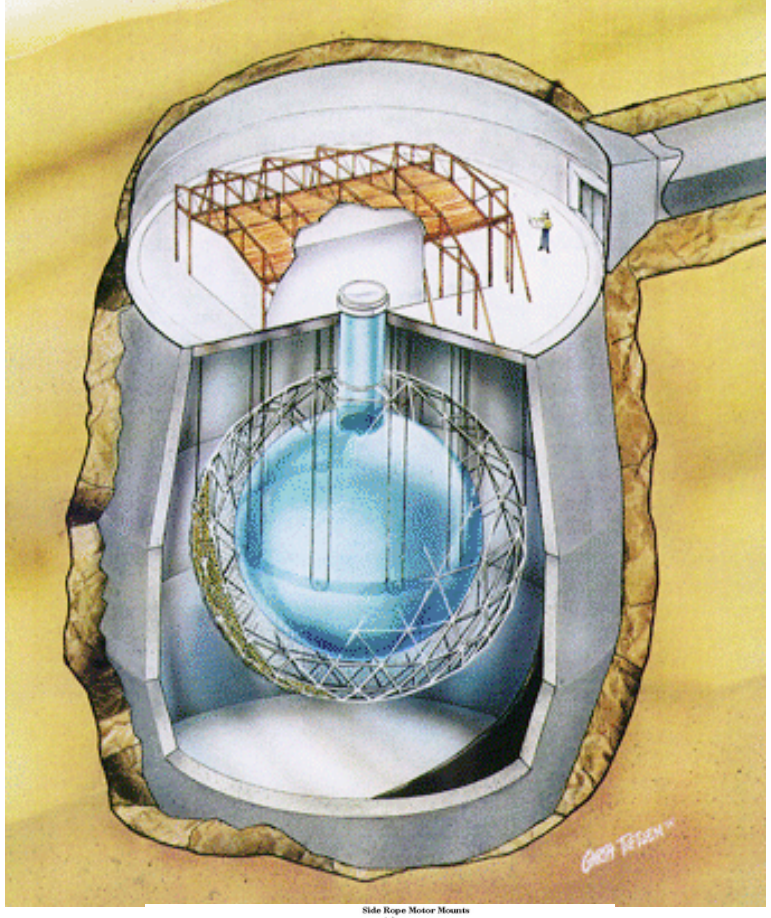
3 Different Operational Phases

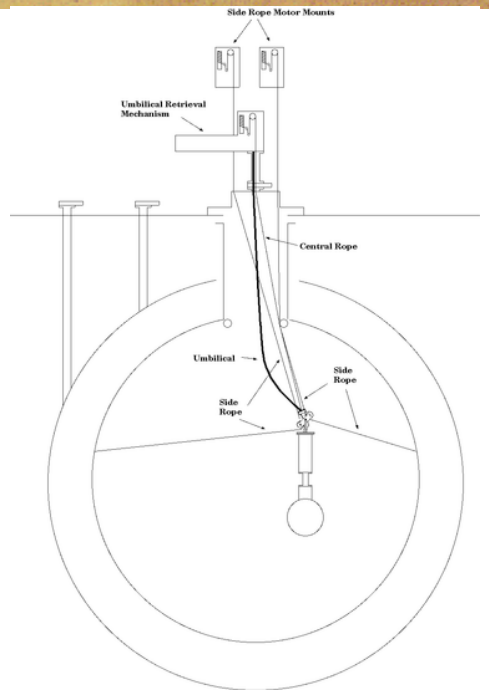
Found that estimated systematic uncertainty in possible position-dependent energy resolution was larger for the 2nd phase, which should have performance at least as good as 1st phase(?!)

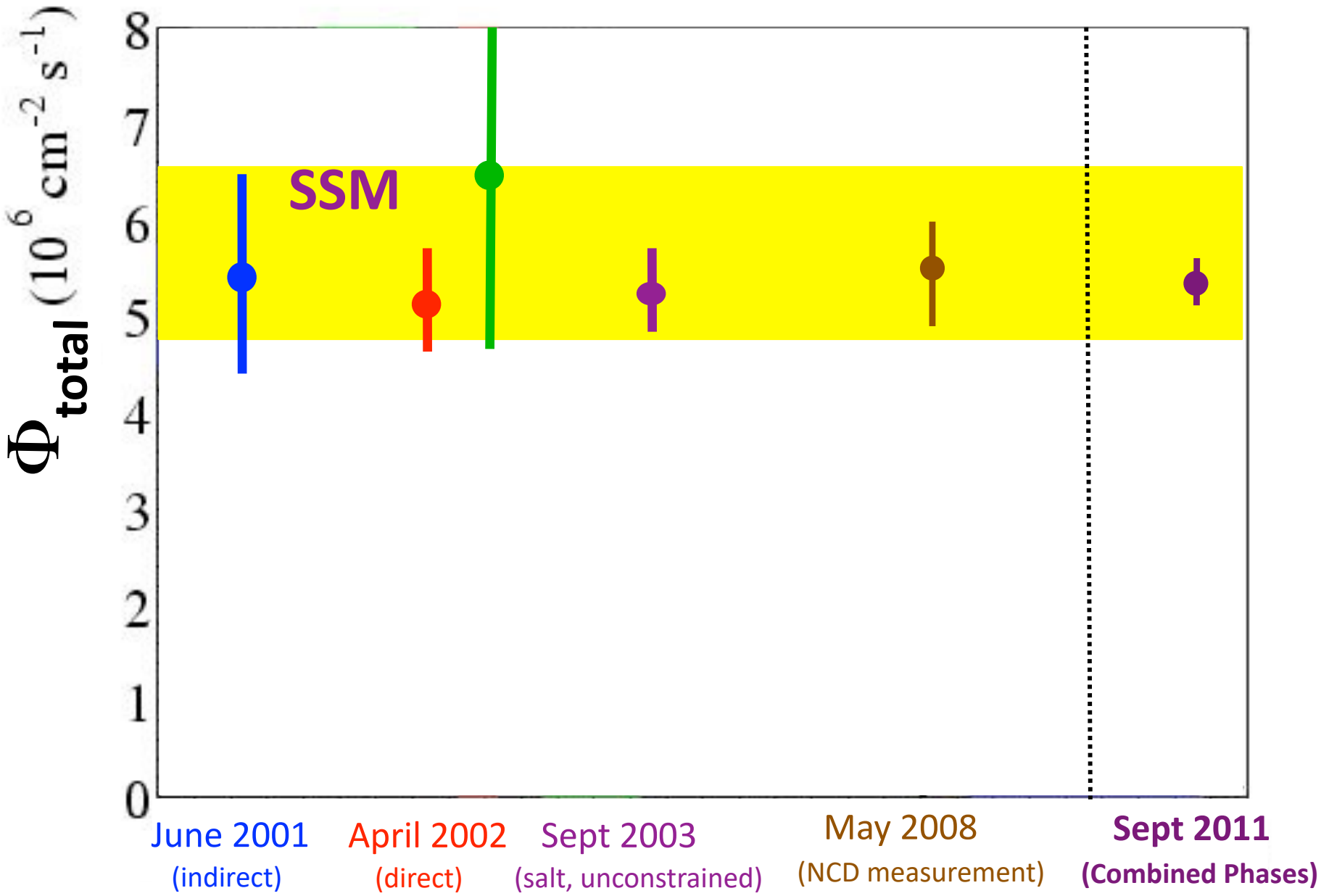
Realised that fewer calibrations had been done in 1st phase, so there was less data to compare!

If you don't look, you don't see!!

(Some groups seem to have elevated this to a strategy for getting small errors!)







3 Experimental Techniques,
at Least 2 Analyses/Technique + Combined Cross-checks

“The Unexpected Hanging”

You will hang
within the next 5 days, but
the exact day will be a “surprise”
in that you will not know whether
the night before is your last!

~~Mon~~ ~~Tue~~ ~~Wed~~ ~~Thu~~ ~~Fri~~

“The Unexpected Hanging”

You will hang
within the next 5 days, but
the exact day will be a “surprise”
in that you will not know whether
the night before is your last!

“The Unexpected Hanging”