Replicability crisis in Science?
Department of Statistical Sciences,
University of Padova 18-22 September 2023

From Science in Theory to Science in Practice: Particle Physics Challenges in Fulfilling Replicability Standards

Tommaso Dorigo

INFN, Sezione di Padova

About me

- FIRST RESEARCHER AT INFN, PADOVA

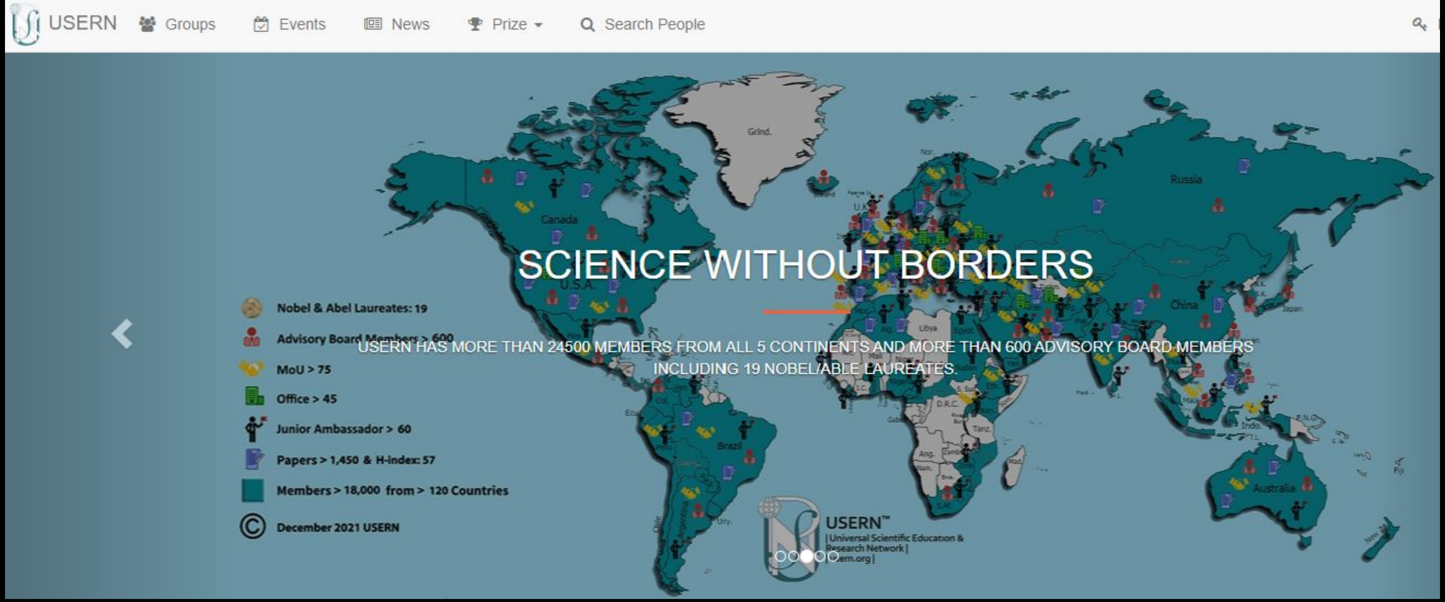
INFN
PADOVA
Istituto Nazionale di Fisica Nucleare
Sezione di Padova

About me

- FIRST RESEARCHER AT INFN, PADOVA
- PROFESSOR AT THE STATISTICAL SCIENCES DEPARTMENT, UNIV. PADOVA, 2018-
- PRESIDENT OF USERN ORGANIZATION, 2022- HTTPS://USERN.TUMS.AC.IR

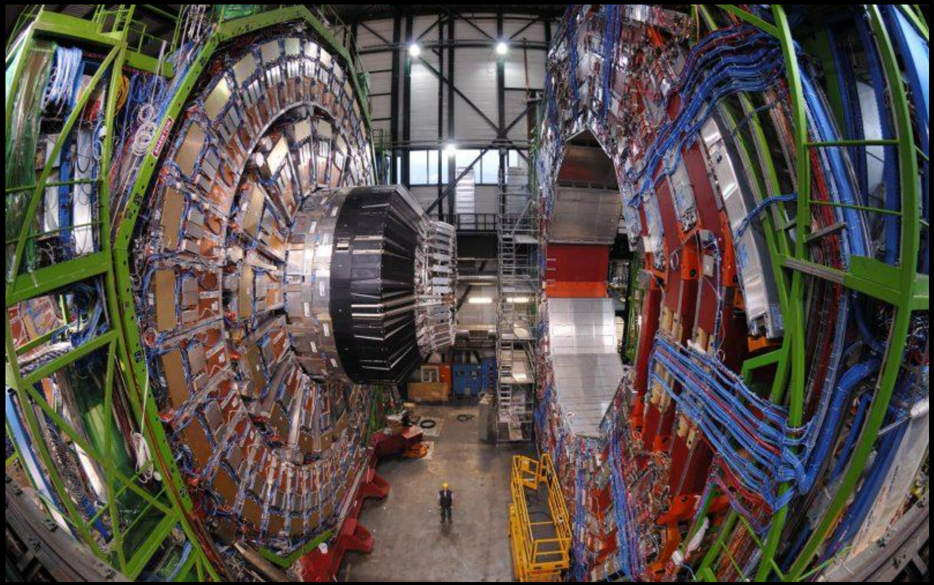
About me

- FIRST RESEARCHER AT INFN, PADOVA
- PROFESSOR AT THE STATISTICAL SCIENCES DEPARTMENT, UNIV. PADOVA, 2018-
- PRESIDENT OF USERN ORGANIZATION, 2022 - [HTTPS://USERN.ORG](https://usern.org)
- MEMBER OF CMS EXPERIMENT AT CERN, 2002-



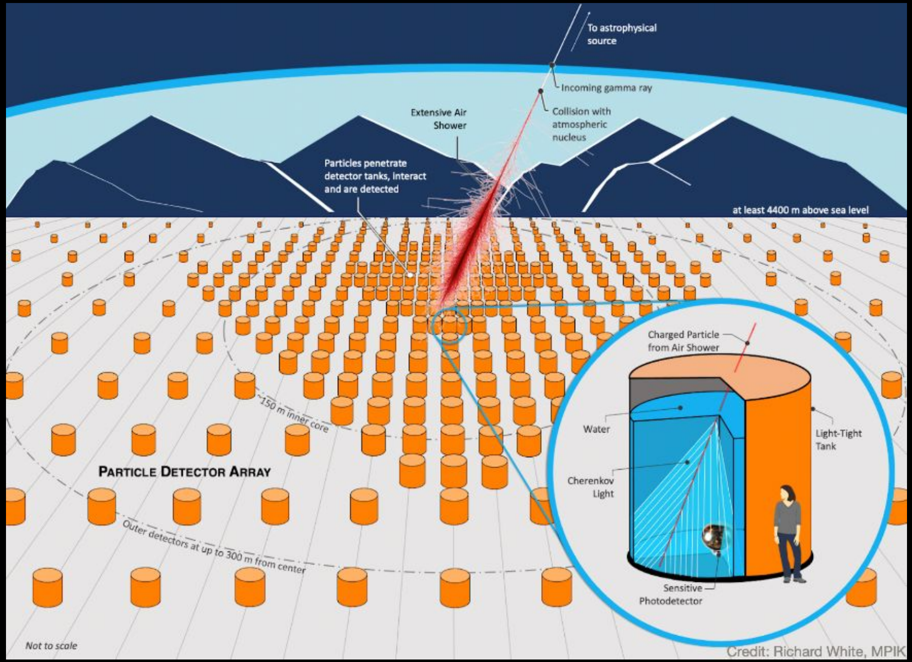
About me

- FIRST RESEARCHER AT INFN, PADOVA
- PROFESSOR AT THE STATISTICAL SCIENCES DEPARTMENT, UNIV. PADOVA, 2018-
- PRESIDENT OF USERN ORGANIZATION, 2022 - [HTTPS://USERN.ORG](https://usern.org)
- MEMBER OF CMS EXPERIMENT AT CERN, 2002-
  - MEMBER OF THE CMS STATISTICS COMMITTEE, 2009- (AND CHAIR, 2012-2015)
- MEMBER OF SWGO EXPERIMENT, 2022-



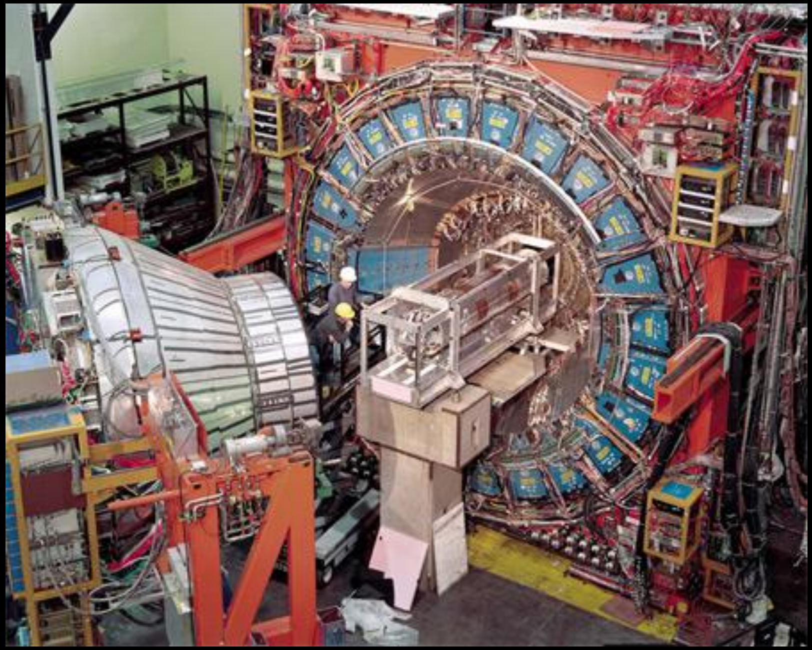
About me

- FIRST RESEARCHER AT INFN, PADOVA
- PROFESSOR AT THE STATISTICAL SCIENCES DEPARTMENT, UNIV. PADOVA, 2018-
- PRESIDENT OF USERN ORGANIZATION, 2022 - HTTPS://USERN.ORG
- MEMBER OF CMS EXPERIMENT AT CERN, 2002-
  - MEMBER OF THE CMS STATISTICS COMMITTEE, 2009- (AND CHAIR, 2012-2015)
- MEMBER OF SWGO EXPERIMENT, 2022-
- FOUNDER AND CHAIR OF MODE COLLABORATION



About me

- FIRST RESEARCHER AT INFN, PADOVA
- PROFESSOR AT THE STATISTICAL SCIENCES DEPARTMENT, UNIV. PADOVA, 2018-
- PRESIDENT OF USERN ORGANIZATION, 2022 - [HTTPS://USERN.ORG](https://usern.org)
- MEMBER OF CMS EXPERIMENT AT CERN, 2002-
  - MEMBER OF THE CMS STATISTICS COMMITTEE, 2009- (AND CHAIR, 2012-2015)
- MEMBER OF SWGO EXPERIMENT, 2022-
- FOUNDER AND CHAIR OF MODE COLLABORATION ([MODE-COLLABORATION.GITHUB.IO](https://mode-collaboration.github.io)), 2020-
- PREVIOUSLY (1992-2010) MEMBER OF CDF EXPERIMENT AT THE TEVA



About me

- PREVIOUSLY (1992-2010) MEMBER OF CDF EXPERIMENT AT THE TEVATRON
  - INTEREST IN STATISTICS DATES BACK TO EARLY ANALYSES IN CDF, BUT BECOMING SAPIENT IN STATISTICAL MATTERS IS A LIFELONG TASK

About me


science 2.0
JOIN THE REVOLUTION

On The Multiverse

By Tommaso Dorigo | August 24th 2023 07:45 AM | 1555 reads | Print | E-mail

RSS Share / Save Tweet Like

Tommaso Dorigo

Search This Blog

I recently read a book by Martin Rees, "On the future". I found it an agile small book packed full with wisdom and interesting considerations on what's in the plate for humanity in the coming decades, centuries, millennia, billions of years. And I agree with much of what he wrote in it, finding also coincidental views on topics I had built my own judgement independently in the past.

One of the things Rees notices is that life in the cosmos is probably overwhelmingly "electronic", as he writes - or as I would put it, artificial, although I understand his wording is more descriptive than mine in a sense. This realization came to me quite naturally, by pondering on the logical necessity for advanced civilizations to produce general artificial intelligence, and on the intrinsic robustness of artificial

- DOING PHYSICS OUTREACH IN A BLOG SINCE 2005. THE BLOG IS NOW AT HTTP://WWW.SCIENCE20.COM/QUANTUM_DIARIES_SURVIVOR
  - THE BLOG CONTAINS OVER 1500 ARTICLES ON SCIENCE OUTREACH

Ways to contact me

- EMAIL: TOMMASO.DORIGO@GMAIL.COM (PREFERRED WAY, QUICK TO ANSWER)
- TWITTER: DORIGO
- PHONE: +39 / 3666995594
- OFFICE PHONE: +39 / 0499677230

Contents

- The context: Particle physics in ~~three~~ ~~six~~ eight slides
- The special nature of subnuclear physics
  - Air-tight alpha values
  - Who needs peer review?!
  - Bias, conservativeness, and undercoverage
  - The dark side
- One full-fledged example of reproducibility from the CMS experiment
- Concluding remarks

Particle physics in eight slides

Statistical practice in particle physics has its peculiarities

"Particles, particles, particles."

Particle physics in eight slides

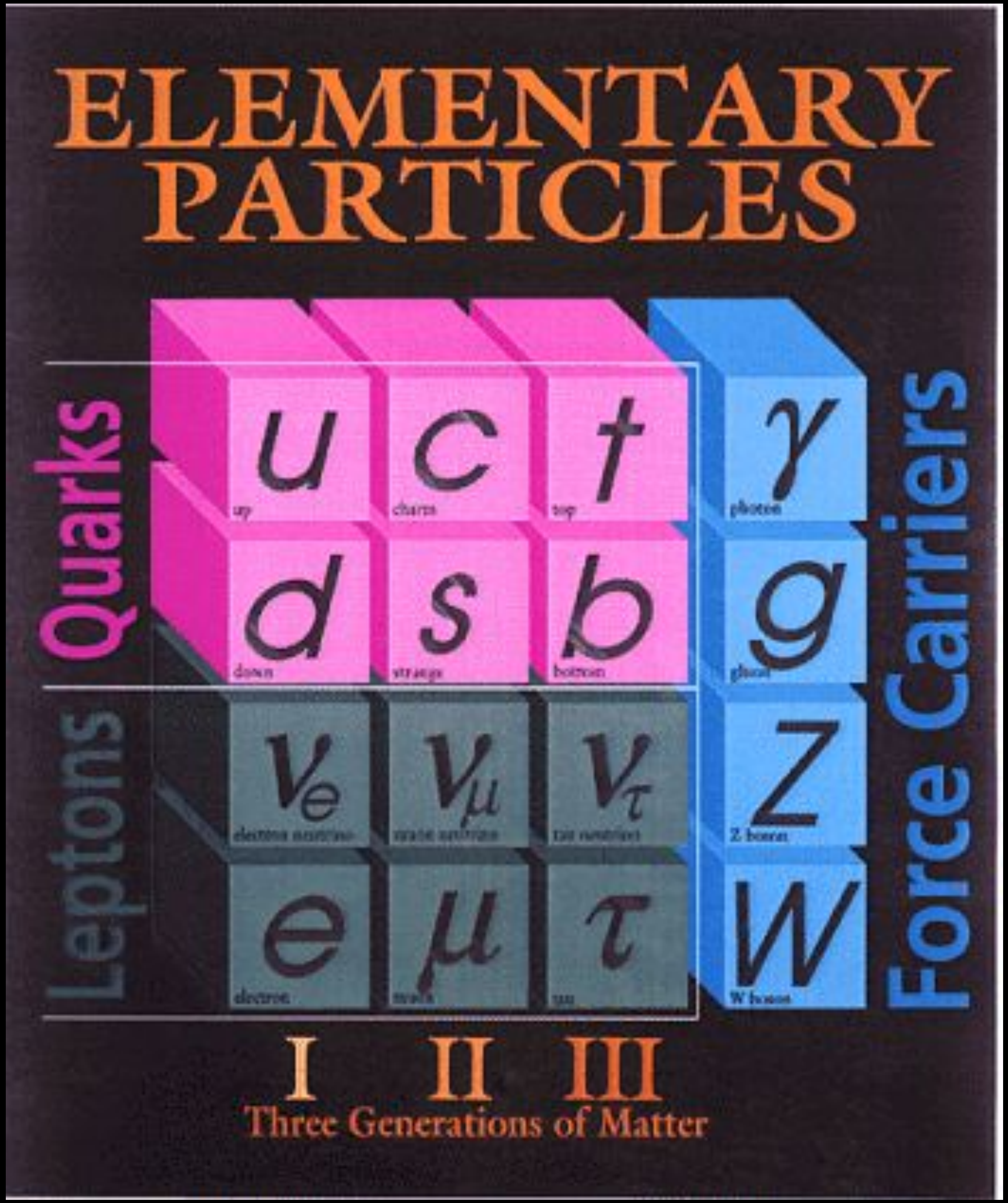
Statistical practice in particle physics has its peculiarities

It is a good idea to first explain in general what we do, if we want to discuss how we do it

"Particles, particles, particles."

The Standard Model

A misnomer – **not** a model but a full-blown theory allowing for high-precision prediction of subatomic processes



The Standard Model

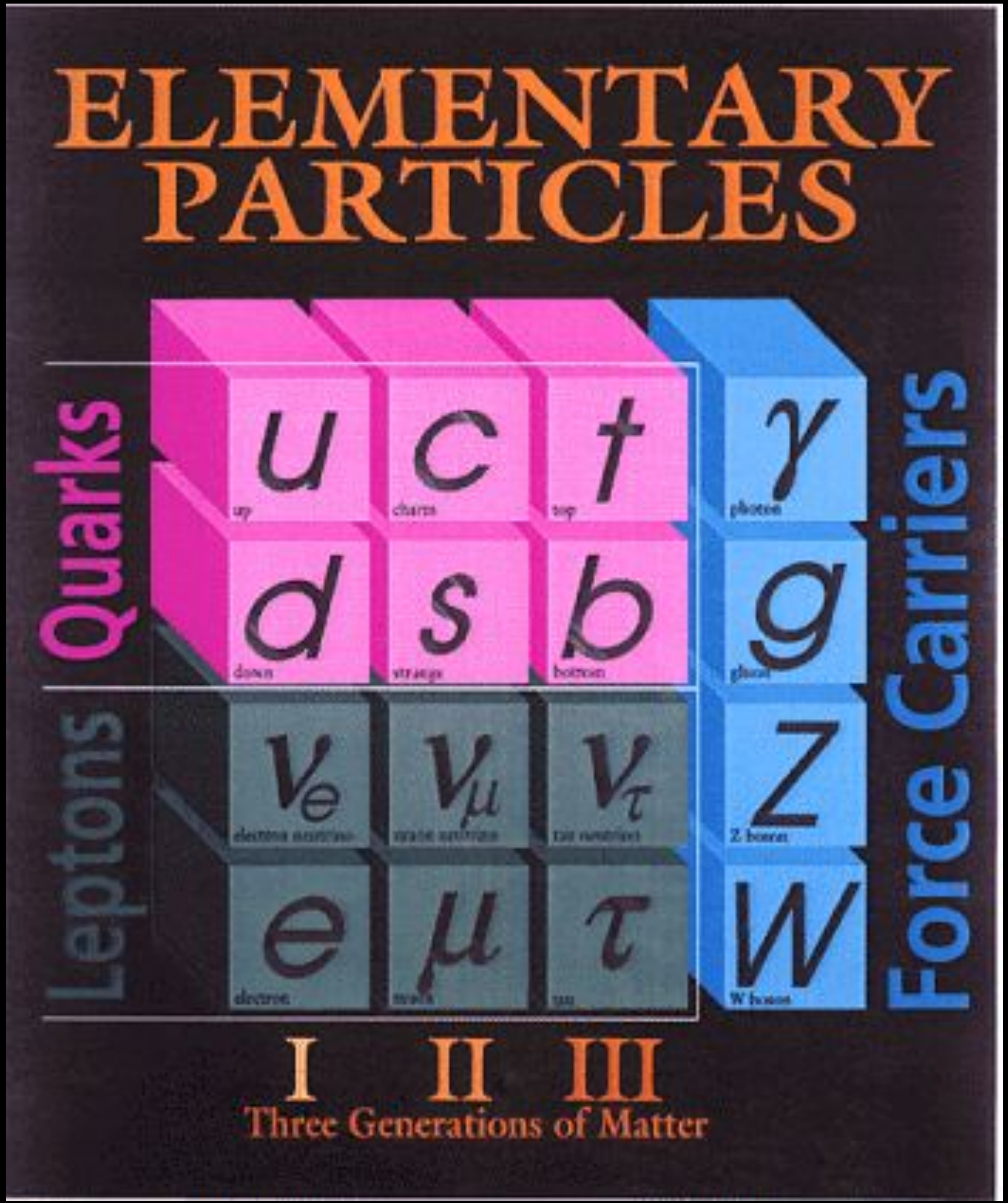
A misnomer – **not** a model but a full-blown theory allowing for high-precision prediction of subatomic processes

Matter is constituted by three families of **quarks**, and three families of **leptons**

Strong interactions between quarks are mediated by 8 gluons, g

Electromagnetic interactions between charged particles are mediated by the photon, γ

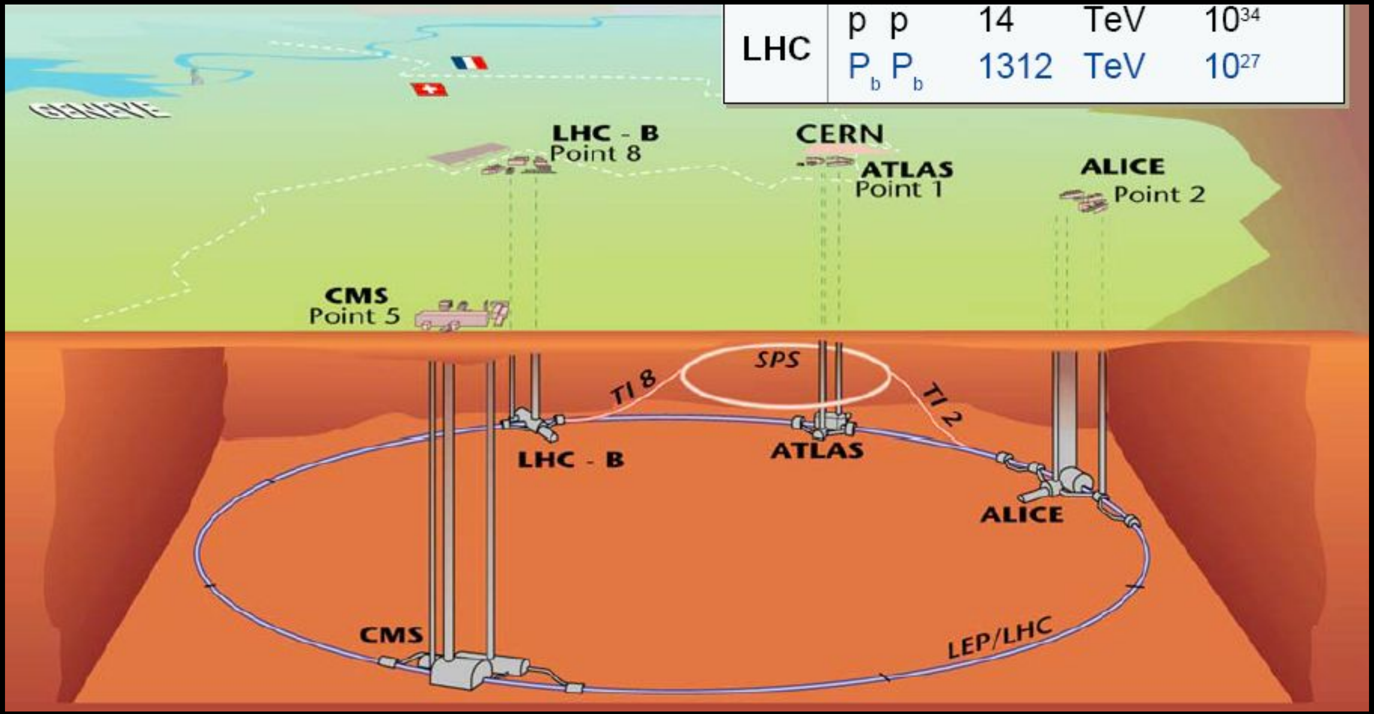
The weak force is mediated by W and Z



The Large Hadron Collider

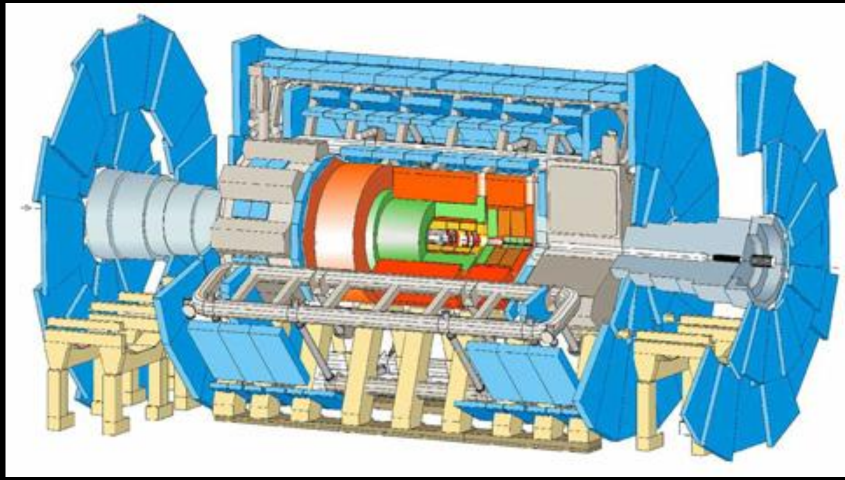
LHC is the largest and most powerful particle accelerator, built to investigate matter at the shortest distances

A giant microscope in the subnuclear world!



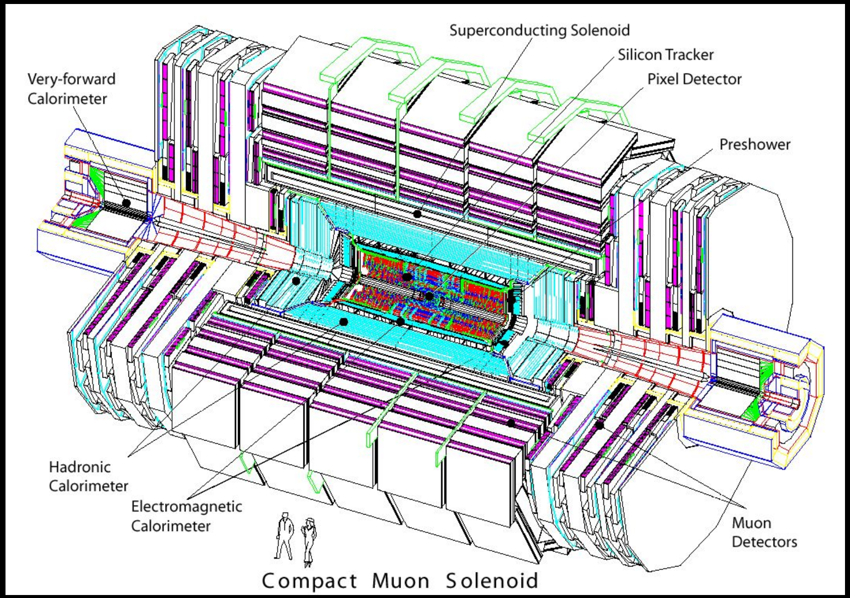
The Large Hadron Collider

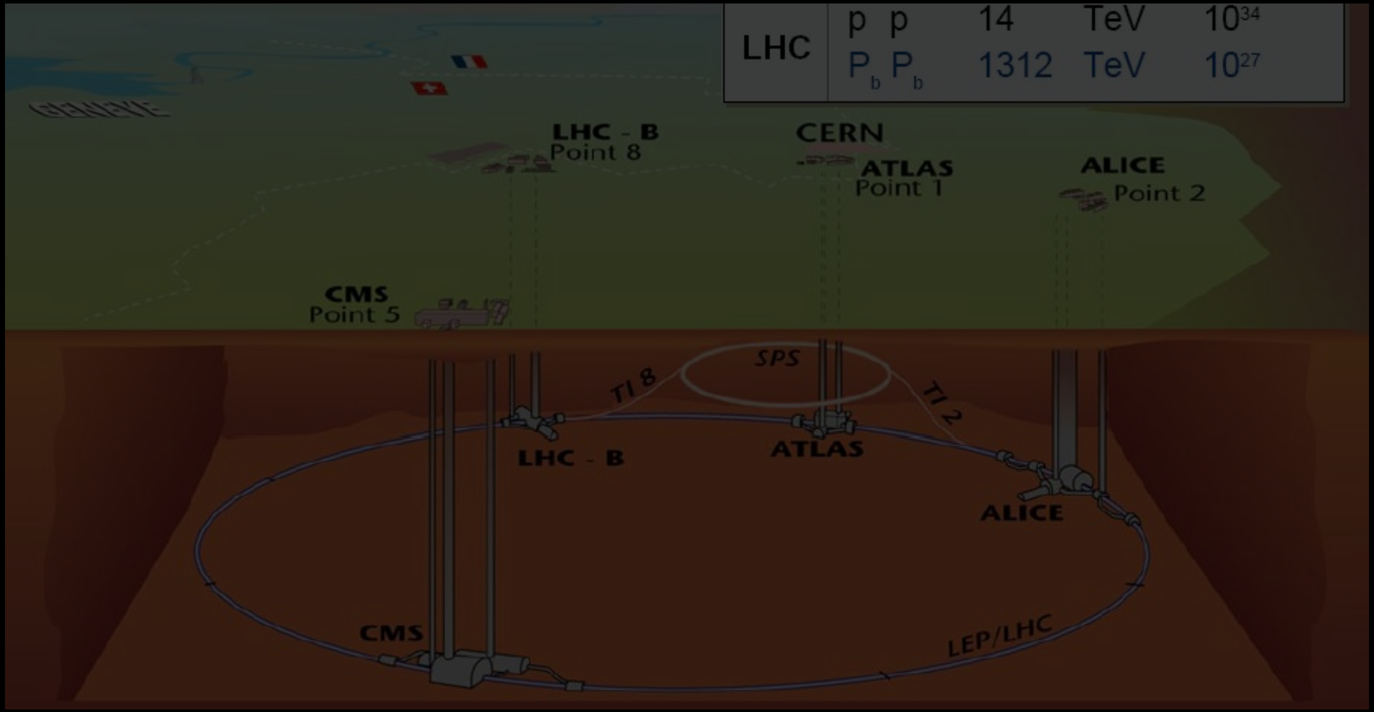
CMS and ATLAS: two huge apparatus that watch like electronic eyes as collisions take place in their core, at 25 nanosecond intervals



ATLAS

CMS

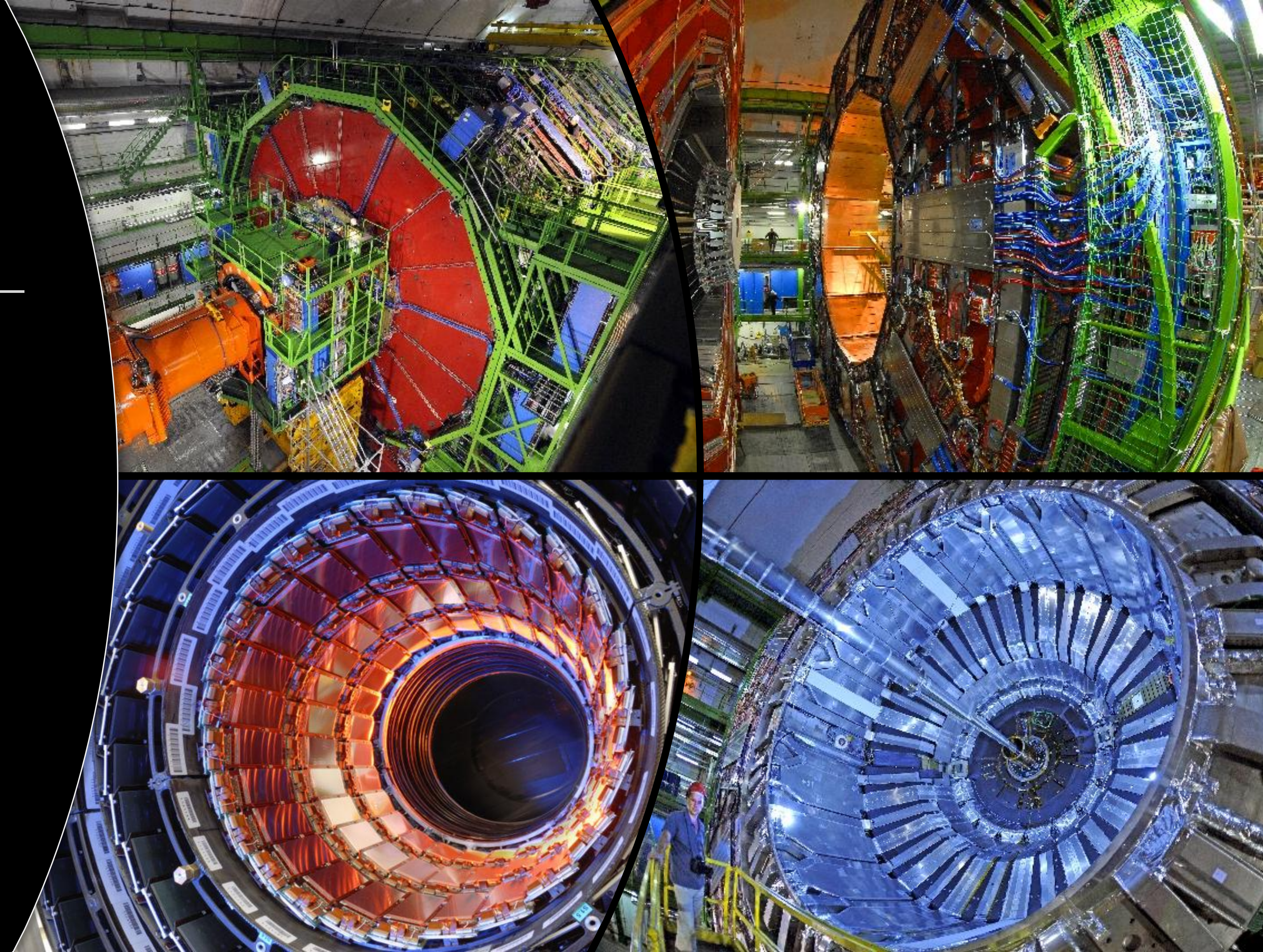




CMS

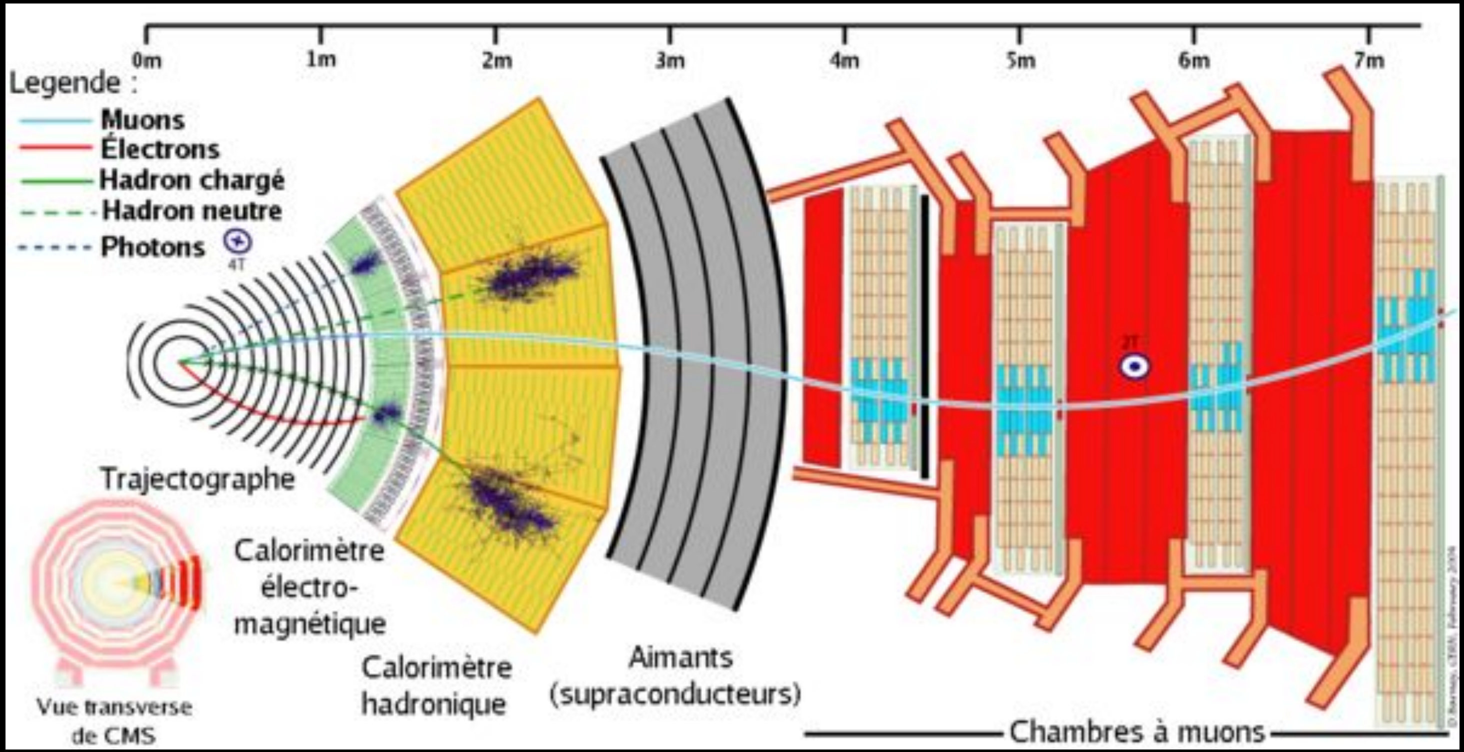
CMS (Compact Muon Solenoid) was built with the specific goal of finding the Higgs boson

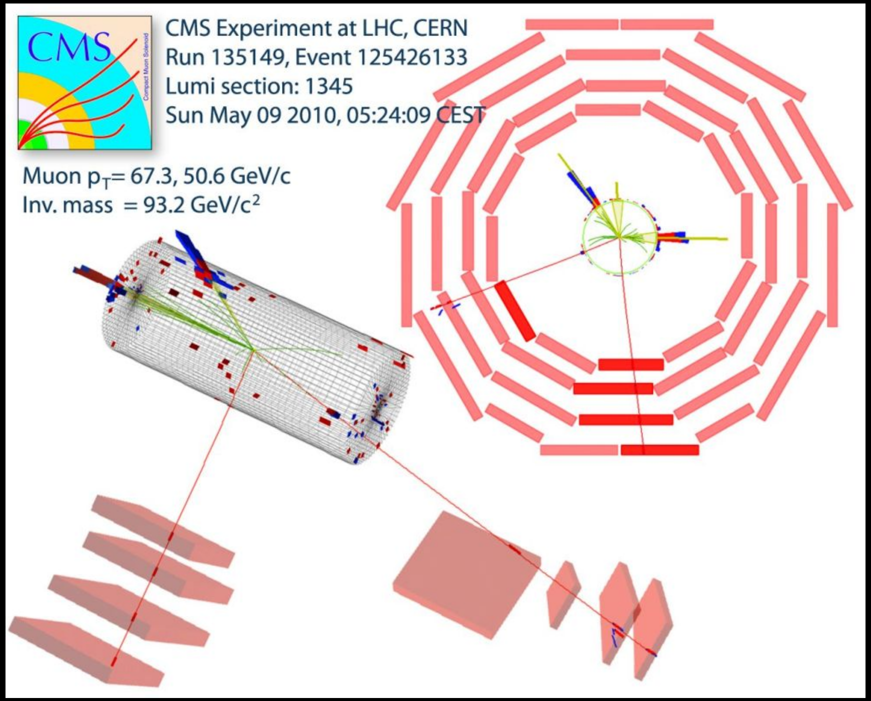
Along with ATLAS, it is arguably the most complex machine ever built by humankind

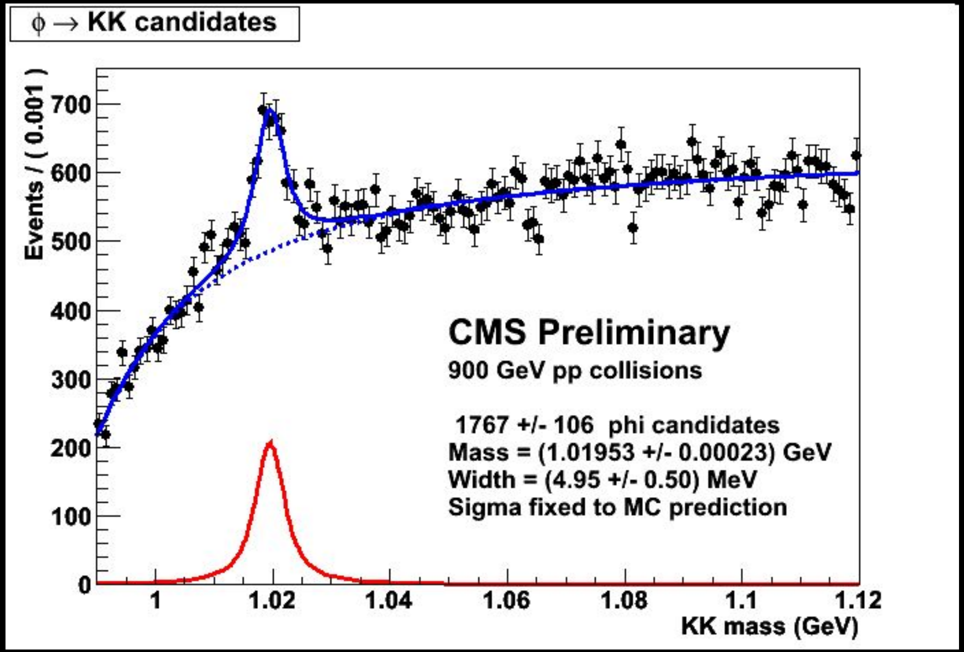


How we detect particles

Particles of different properties leave distinguishable signals in different subsystems







How we see a collision

A reconstruction of the electronic signals provides us a «view» of the created objects. Using their characteristics we build high-level variables which we compare to theoretical models, for measurements and searches

What it is that we do

We have a theory that works wonders – but we believe it is incomplete and to some extent unsatisfactory.

What it is that we do

We have a theory that works wonders – but we believe it is incomplete and to some extent unsatisfactory.

- So we look for new physics processes: things that the standard model does not predict
  - New matter particles
  - New force carriers

What it is that we do

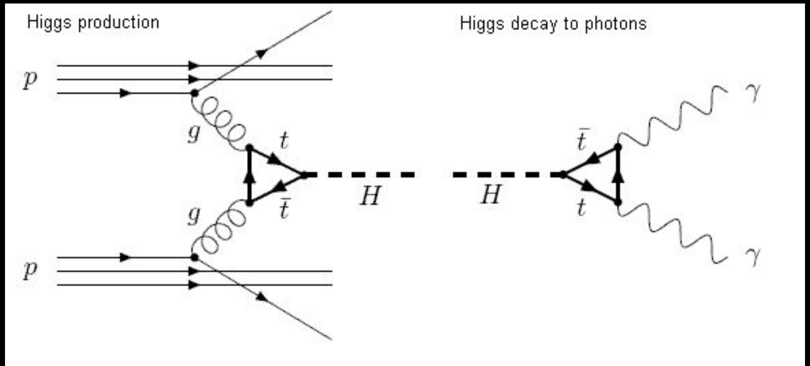
We have a theory that works wonders – but we believe it is incomplete and to some extent unsatisfactory.

- So we look for new physics processes: things that the standard model does not predict
  - New matter particles
  - New force carriers

- We also measure with precision known processes, in the attempt of finding a significant difference with model calculations

Example: new particle searches

The typical search for a new particle involves a model which predicts the particle, from which we can derive a simulation of its production

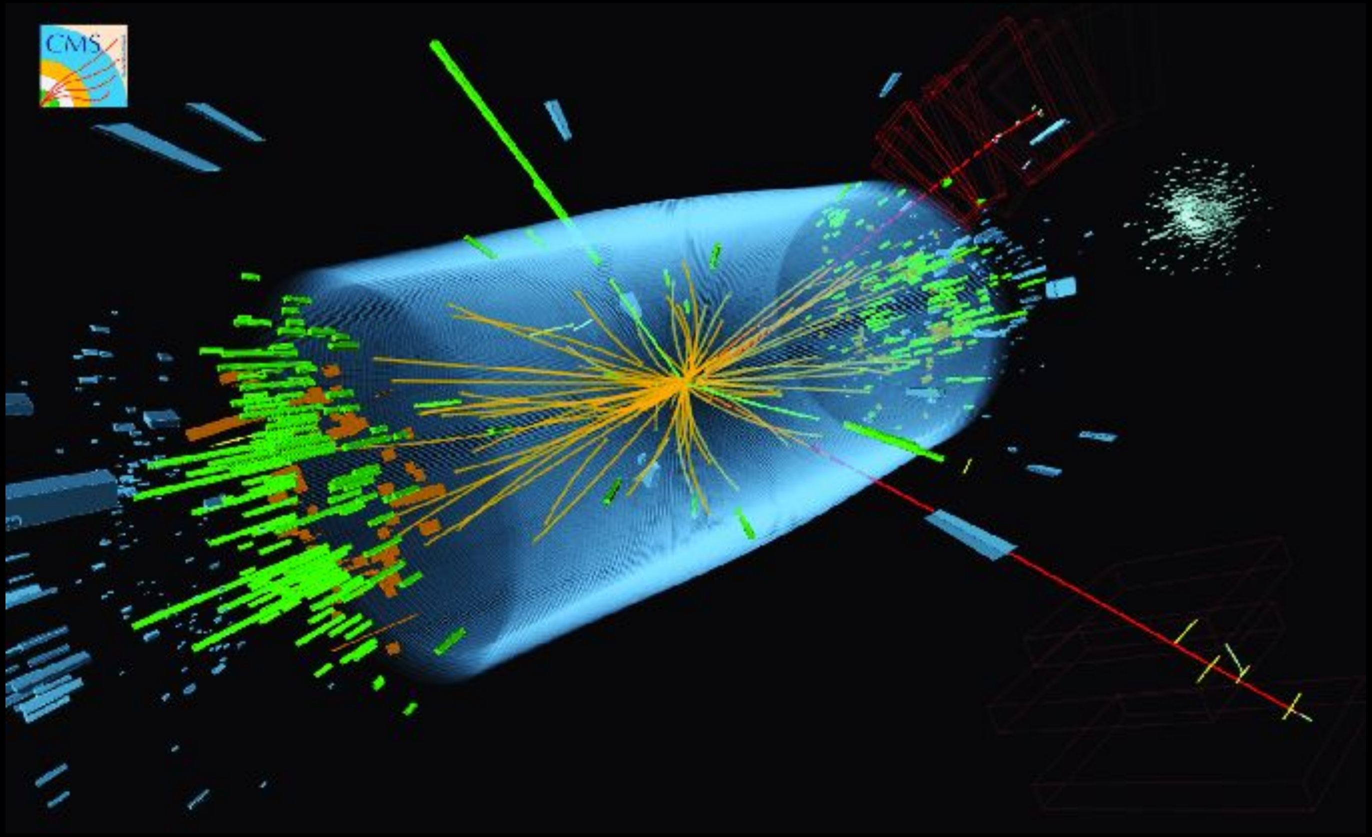


Example: new particle searches

The typical search for a new particle involves a model which predicts the particle, from which we can derive a simulation of its production

- Monte Carlo methods allow us to produce high-fidelity simulated datasets that teach us how the signal looks like

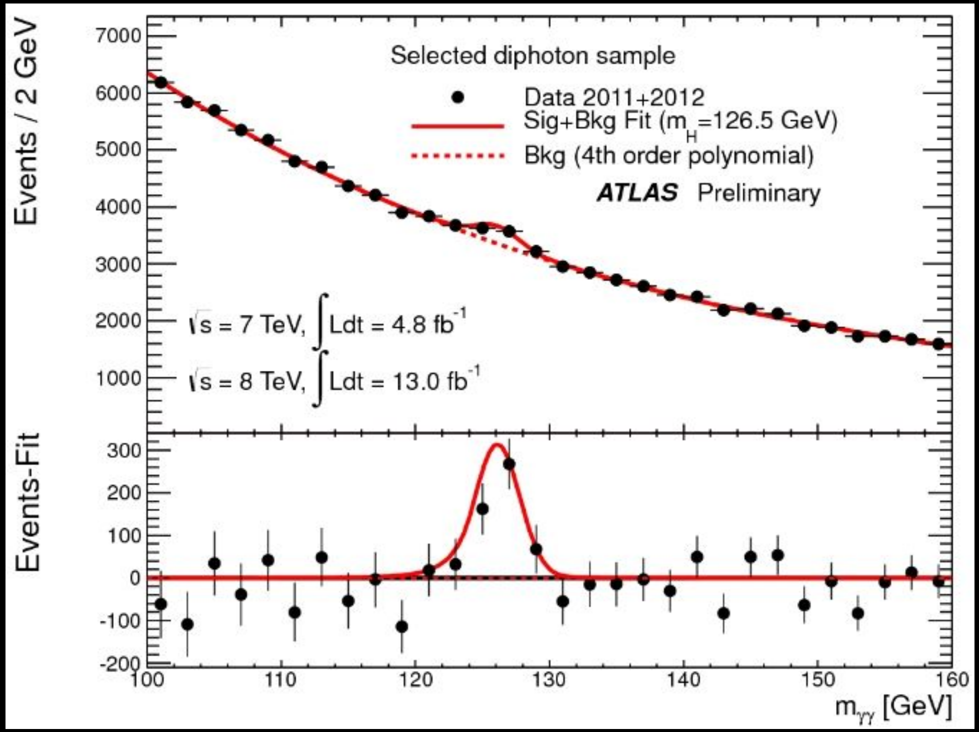
- A data selection isolates a sample



Example: new particle searches

The typical search for a new particle involves a model which predicts the particle, from which we can derive a simulation of its production

- Monte Carlo methods allow us to produce high-fidelity simulated datasets that teach us how the signal looks like
- A data selection isolates a sample where we try to evidence the particle – typically a narrow bump on a smooth background
- A test of hypotheses allows to derive

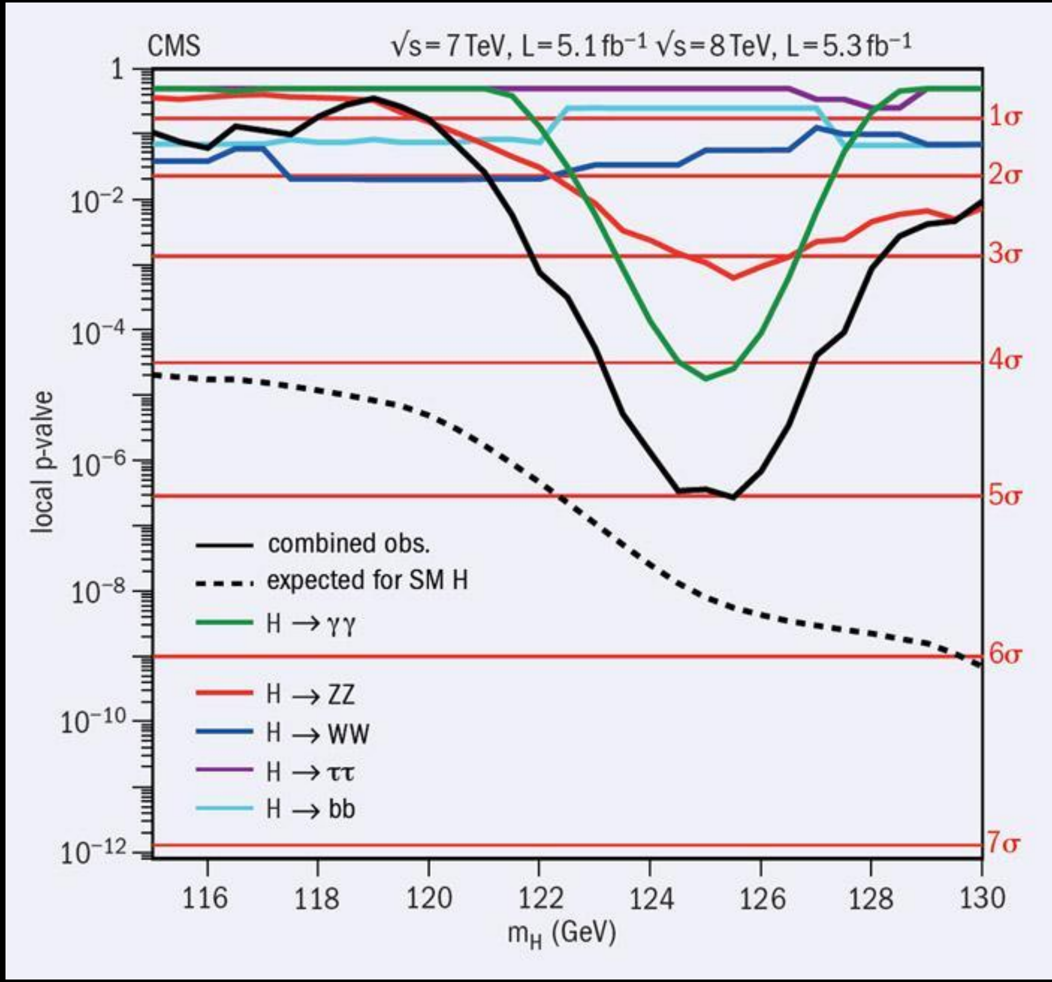


Example: new particle searches

The typical search for a new particle involves a model which predicts the particle, from which we can derive a simulation of its production

- Monte Carlo methods allow us to produce high-fidelity simulated datasets that teach us how the signal looks like

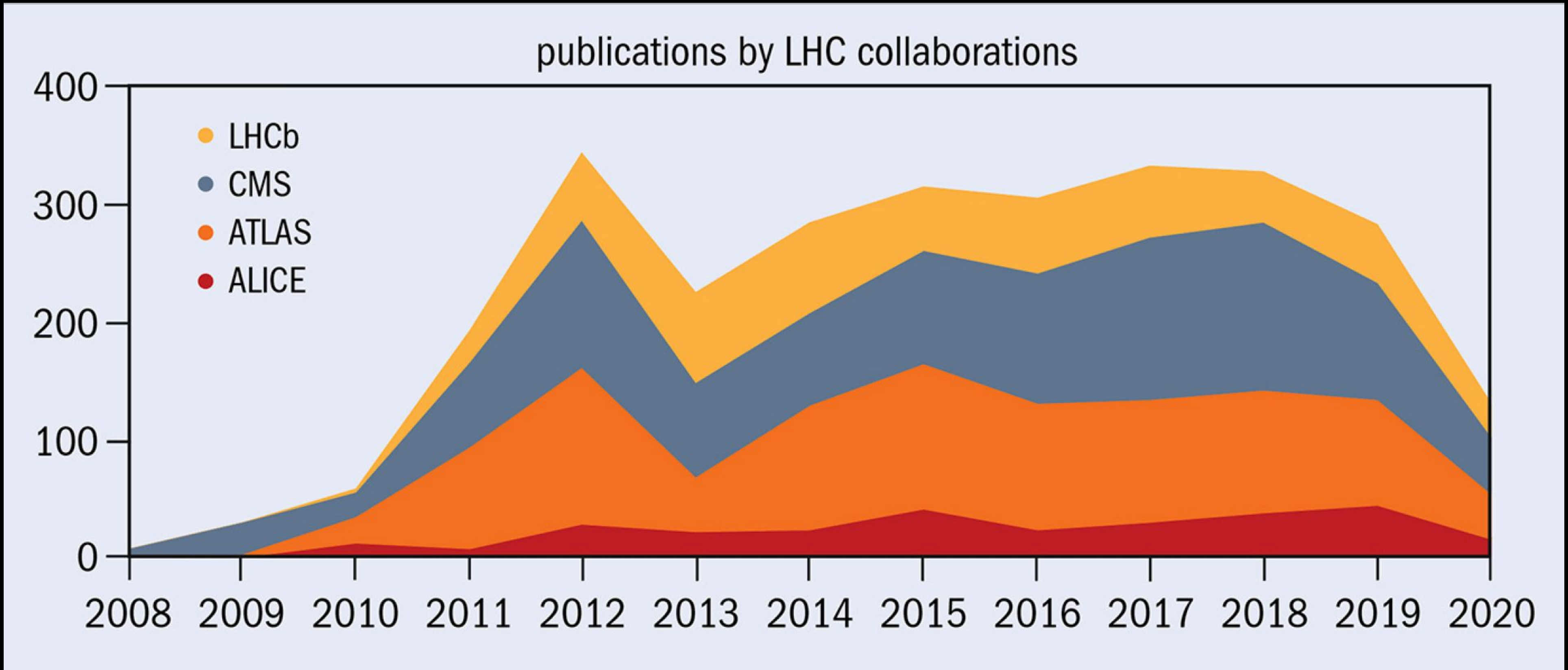
- A data selection isolates a sample where we try to evidence the particle – typically a narrow bump on a smooth background



A test of hypotheses finally allows to derive p(data|H0)

And what if there is no signal ?

If we do not see a signal, we do not despair. We can exclude the new physics model, and still publish a highly cited paper!

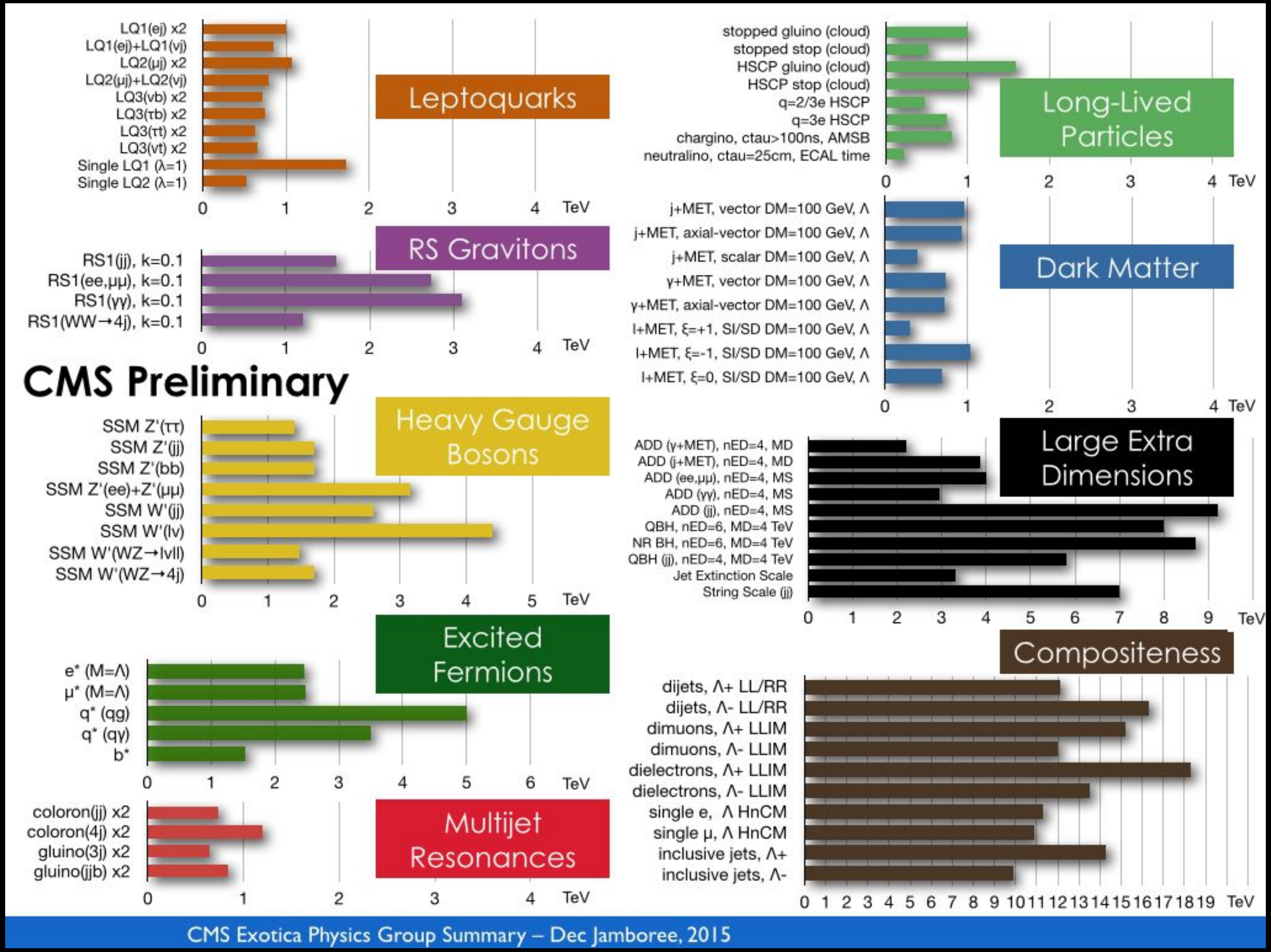


And what if there is no signal ?

If we do not see a signal, we do not despair. We can exclude the new physics model, and still publish a highly cited paper!

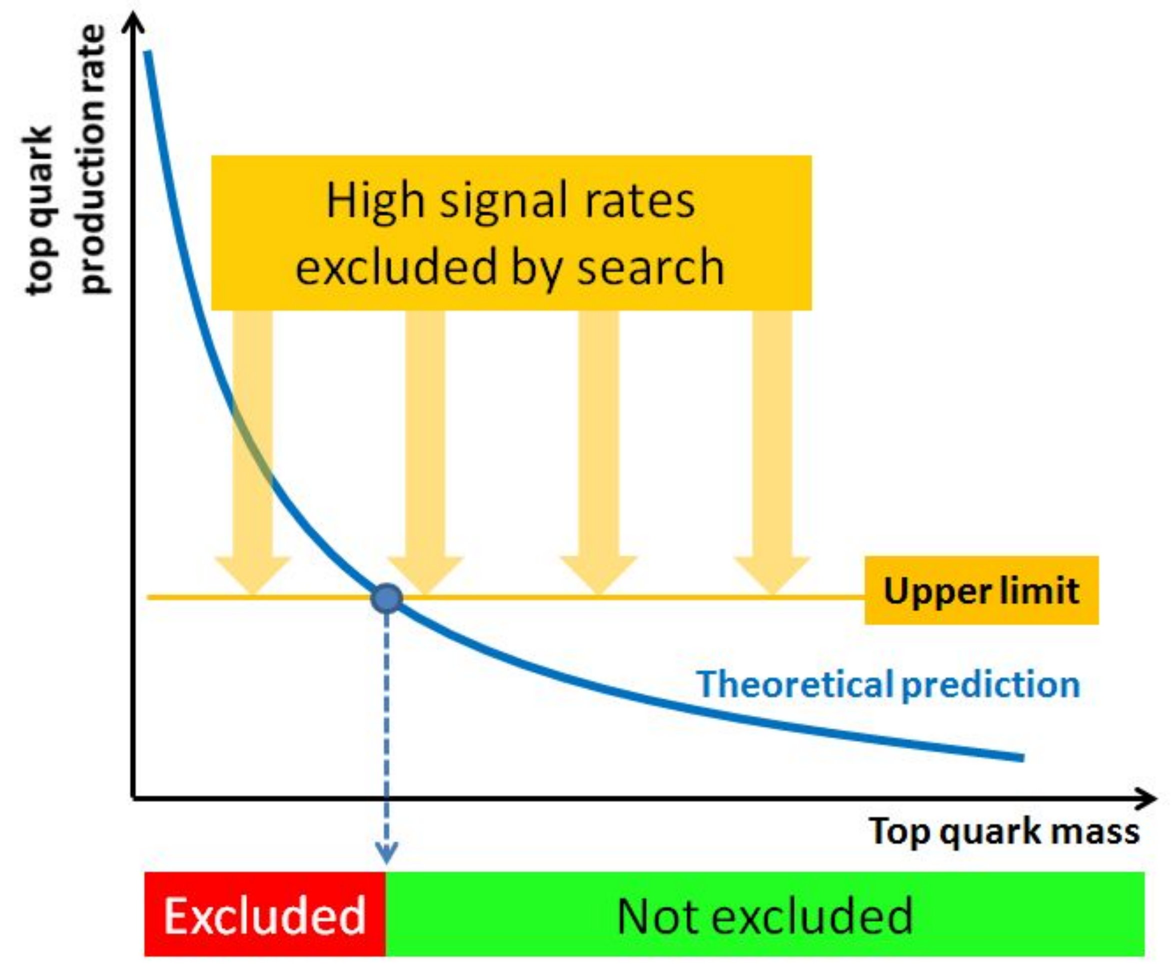
More often the theory depends on some unknown parameter (theorists are cunning!), so we exclude ranges of its value

- Typically this is the mass of the particle



And what if there is no signal ?

We can e.g. derive lower limits on the particle mass from upper limits on the signal strength, by comparing those to a theoretical model



The Golden Age of Particle Physics

- New microscopes
- The first idea of Reines and Cowan
- Bump hunters and the birth of the five-sigma criterion

The Special Nature of Subnuclear Physics: 1 – Evanescence of the Studied Object

If you can't look inside, throw something at it

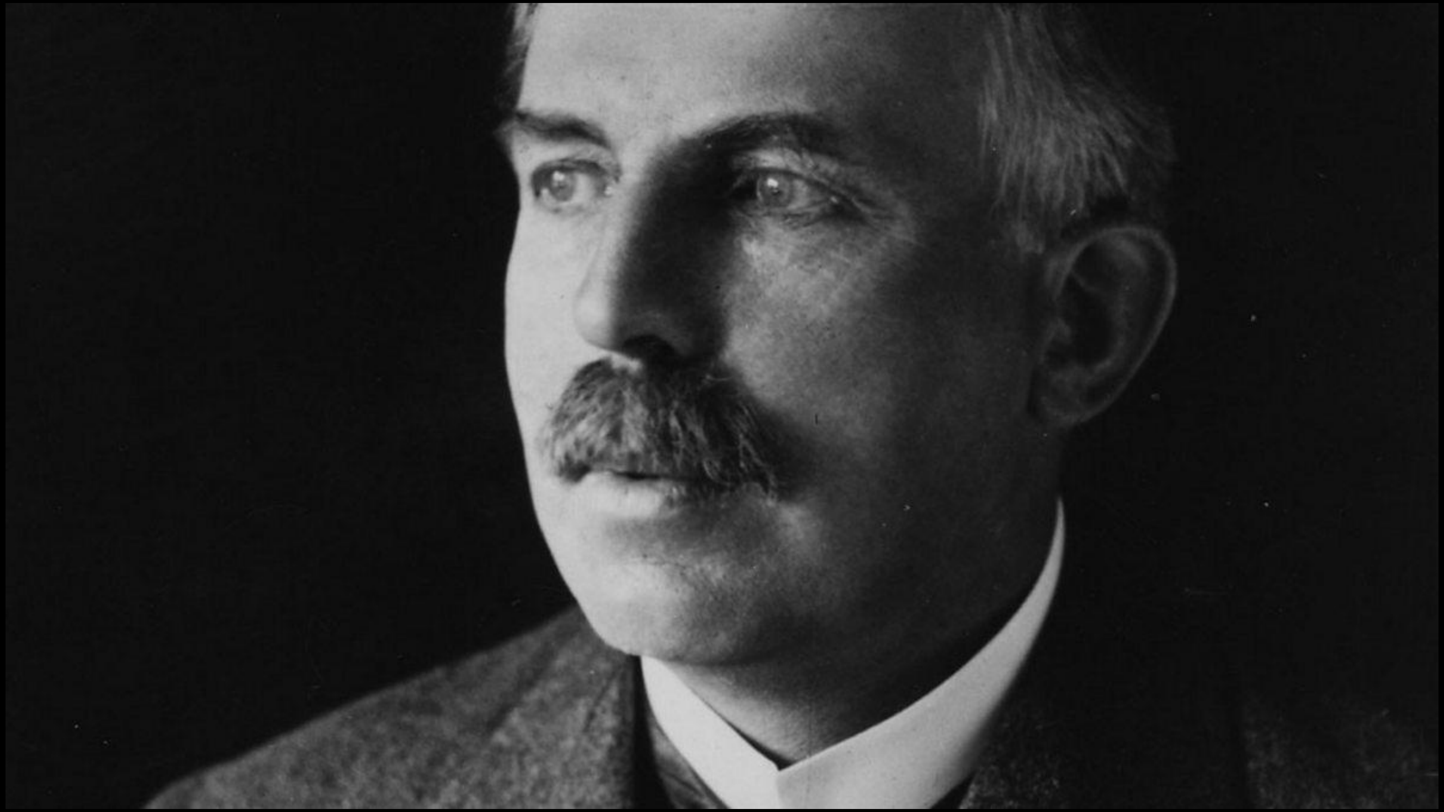
Particles cannot be seen directly

Particles decay before you can say tau

Particles are small – you need many!

Ernest Rutherford and the nucleus

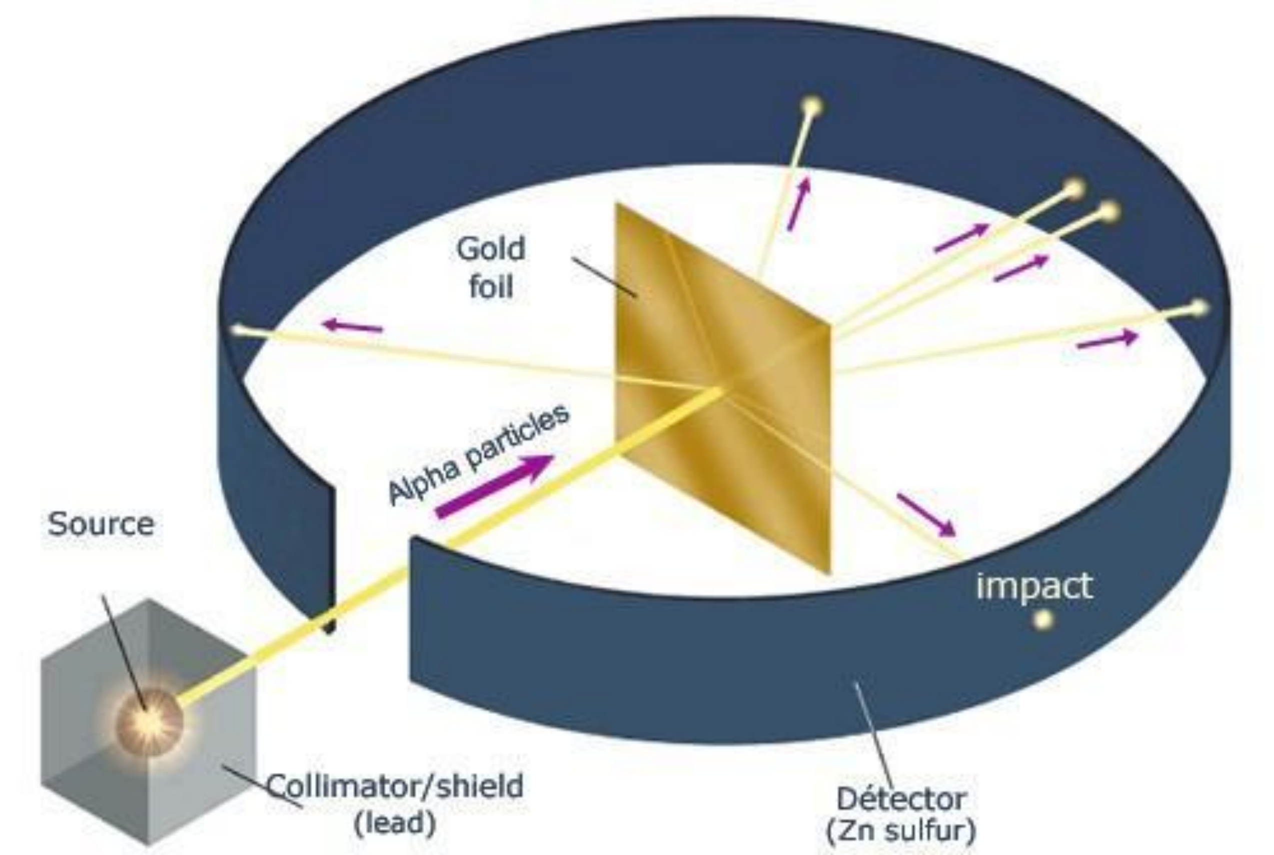
Rutherford must be credited of two extraordinary accomplishments:



Ernest Rutherford and the nucleus

Rutherford must be credited of two extraordinary accomplishments:

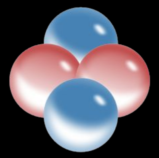
1) He invented a new way to look inside objects

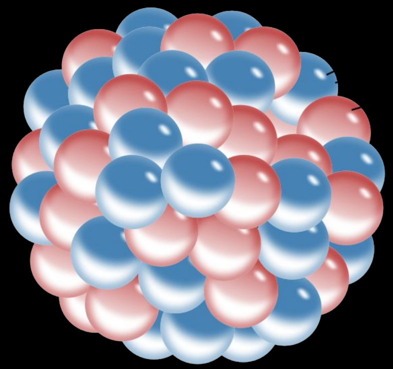


Ernest Rutherford and the nucleus

Rutherford must be credited of two extraordinary accomplishments:

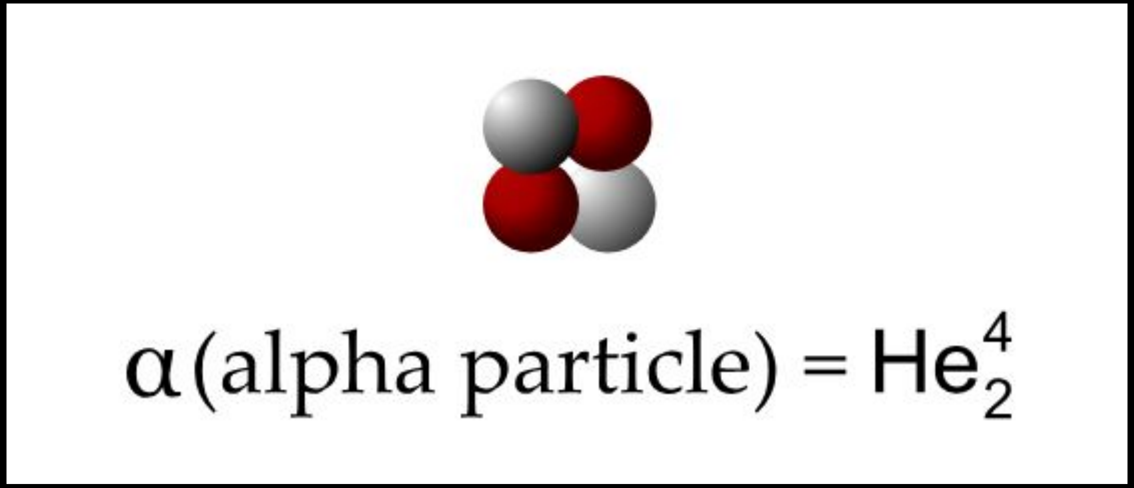
1) He invented a new way to look inside objects





2) He produced an unsurpassed feat in scientific replicability:

he discovered X using X !



Particles decay

The only reason why protons, electrons, photons, and neutrinos are stable is that there is nothing lighter than them they can turn into

Particles decay

The only reason why protons, electrons, photons, and neutrinos are stable is that there is nothing lighter than them they can turn into

Other subnuclear particles decay with rates ranging from 0.0000000000000000000000001 (the W and Z bosons, the top quark) to 1000 seconds (the neutron)

Particles decay

The only reason why protons, electrons, photons, and neutrinos are stable is that there is nothing lighter than them they can turn into

Other subnuclear particles decay with rates ranging from 0.0000000000000000000000001 (the W and Z bosons, the top quark) to 1000 seconds (the neutron)

In most cases we do not get to "see" the direct effects of a particle propagation in the detector, but only *infer* their existence from their decay products

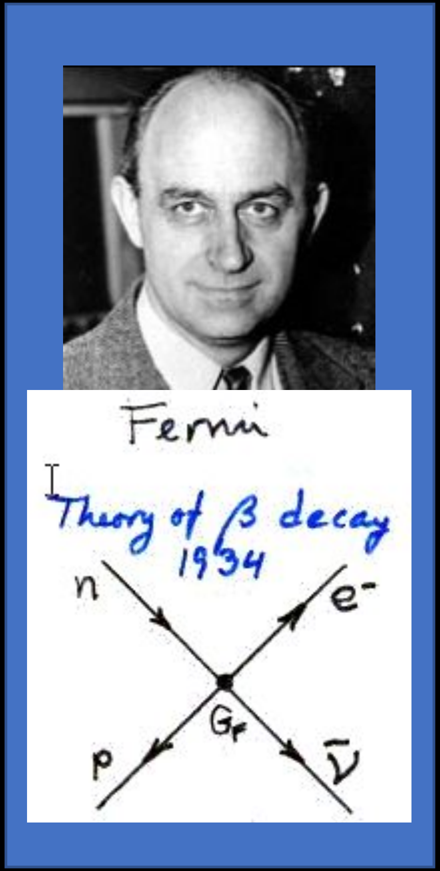
□ reliance on a model

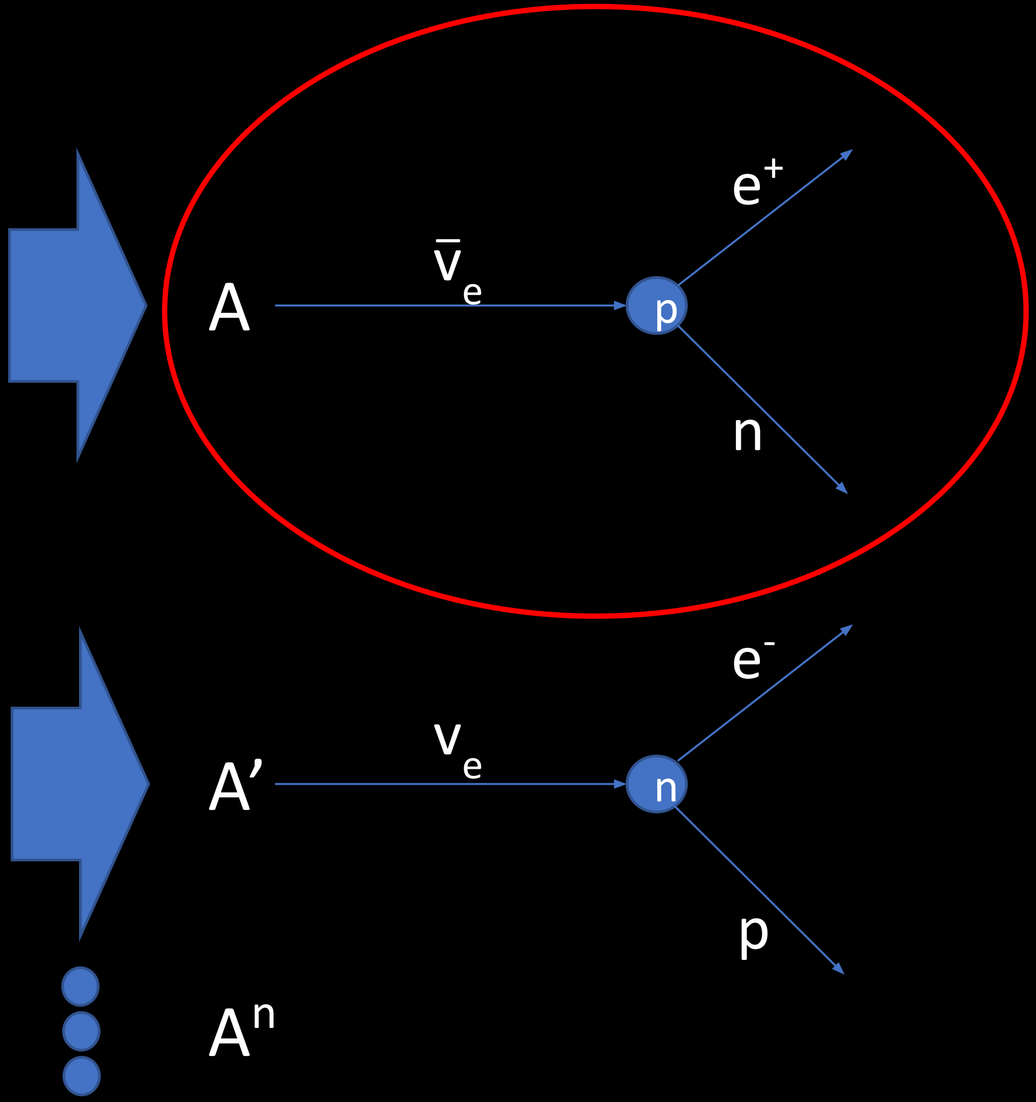
□ issues of repeatability

The Neutrino: making a repeatable experiment

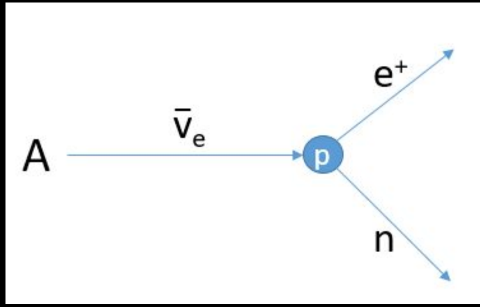
Wolfgang Pauli hypothesized the existence of the neutrino in 1930, and Fermi successively constructed a theory of beta decay, implying many reactions must exist.

To test Fermi's theory and the neutrino existence we may focus on a reaction, say A, and devise a means to study it.





Given a well defined goal (observing a particular reaction) there are few viable options to produce it and observe it



Predict the rate of events

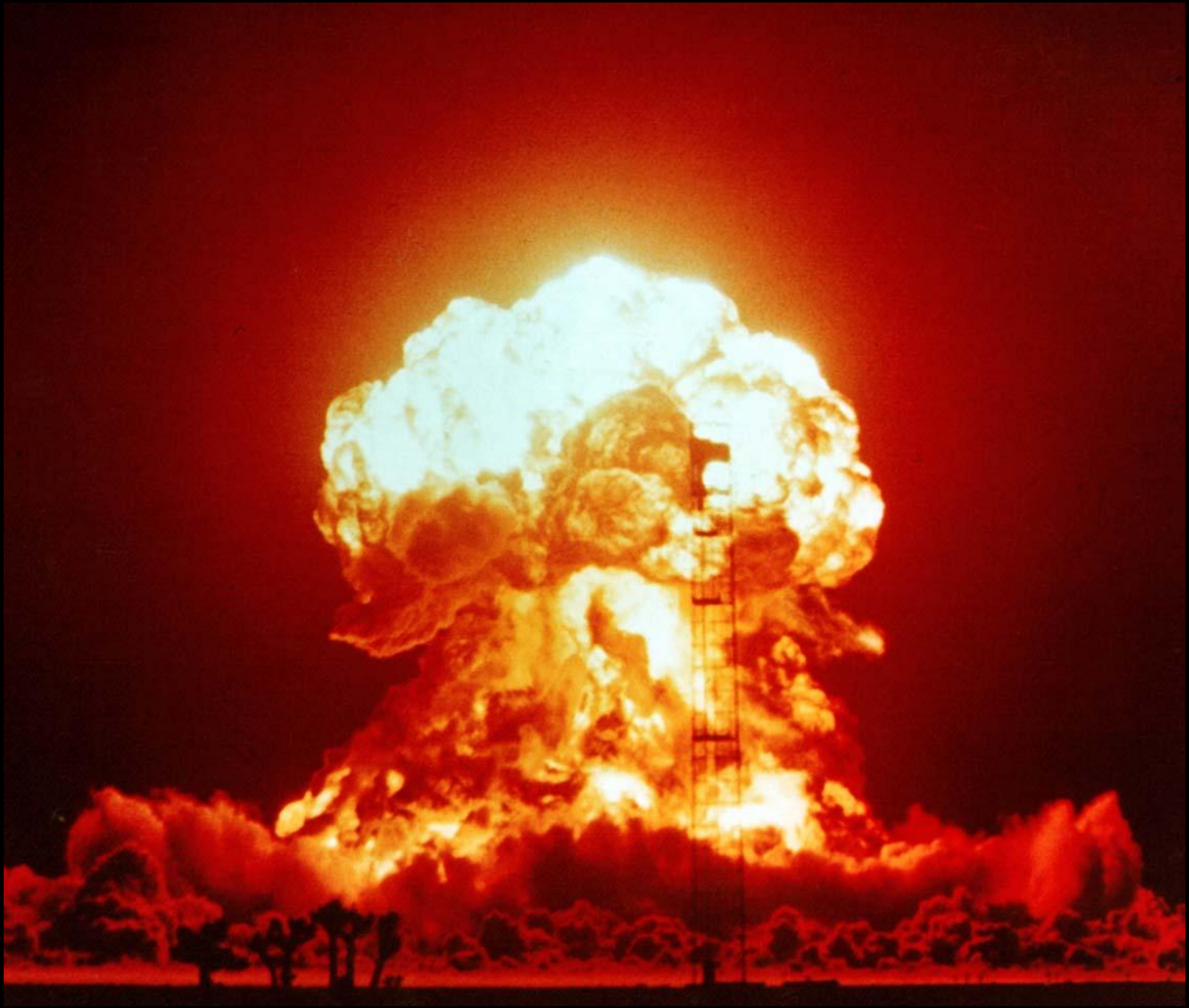
$\sigma = 6 * 10^{-44}\ cm^{-2}$

 Rates are minuscule!

In the Reines-Cowan experiment (1956), one early question was how to produce an intense flux of neutrinos

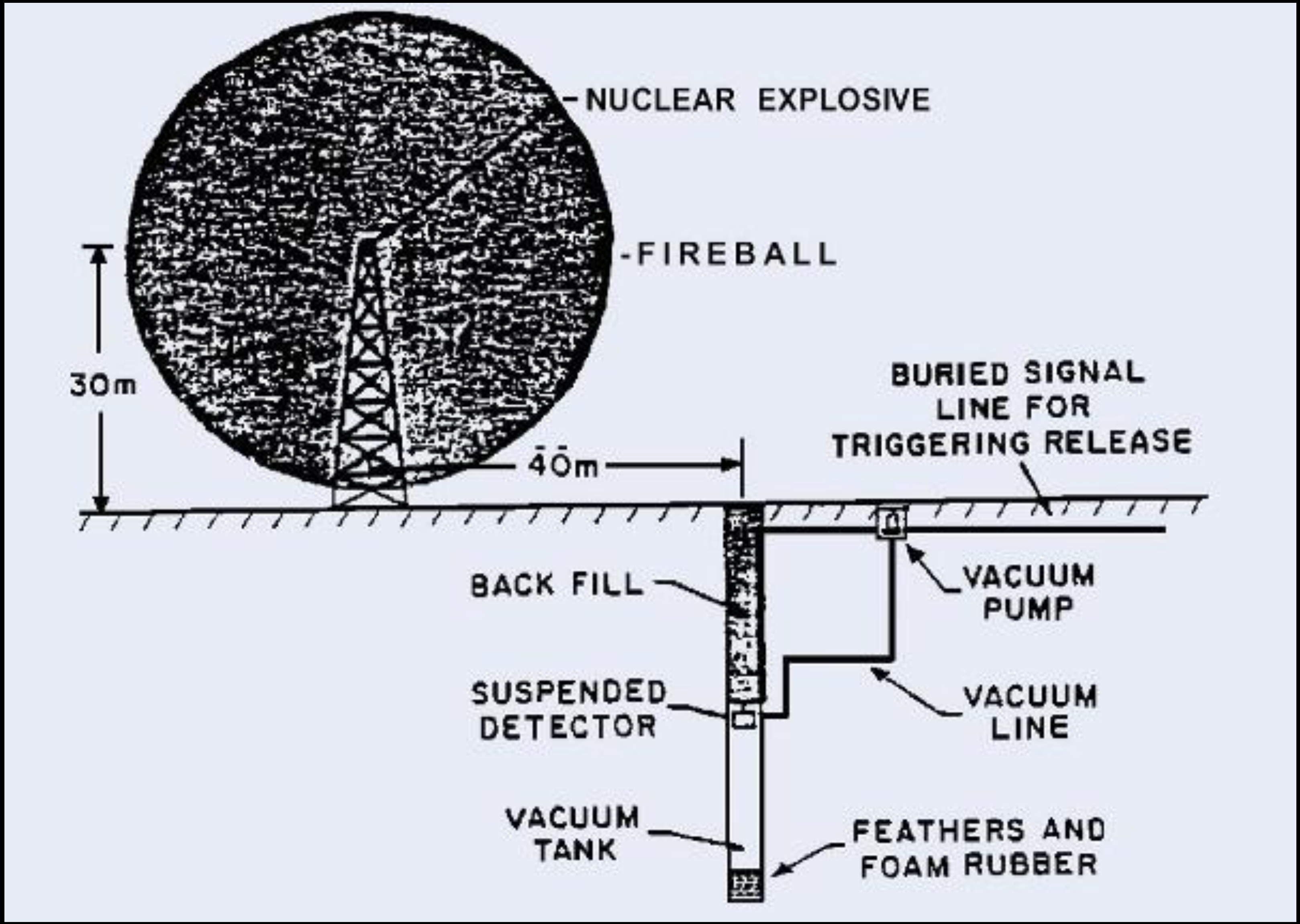
 Need a LARGE flux!

… like the one from this kind of gadget?



I thought the original idea of the Reines-Cowan experiment belonged to a workshop on replicability

Can’t think of an experiment whose replicability is less practical!



The Special Nature of Subnuclear Physics: 2 - Stochasticity

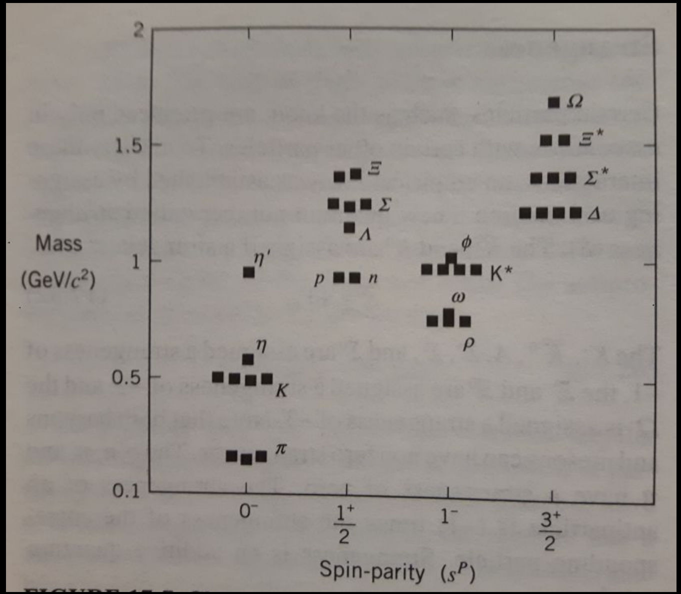
Quantum physics is the quintessential stochastic data generation mechanism

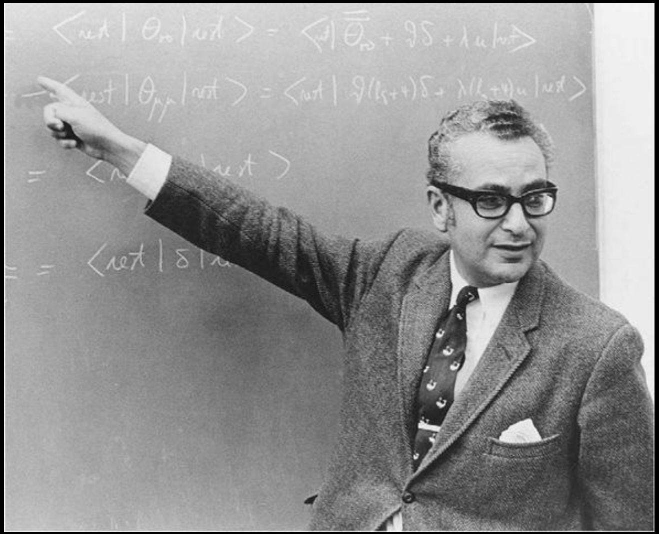
One-event discoveries: e.g., the Omega minus

The era of bump hunts and the birth of the five-sigma criterion

The static quark model and the Omega discovery

Between 1961 and 1964 Murray Gell-Mann and independently George Zweig, note how all hadrons until then discovered can be classified into «supermultiplets» of bodies sharing the same symmetry properties under representations of the SU(3) group

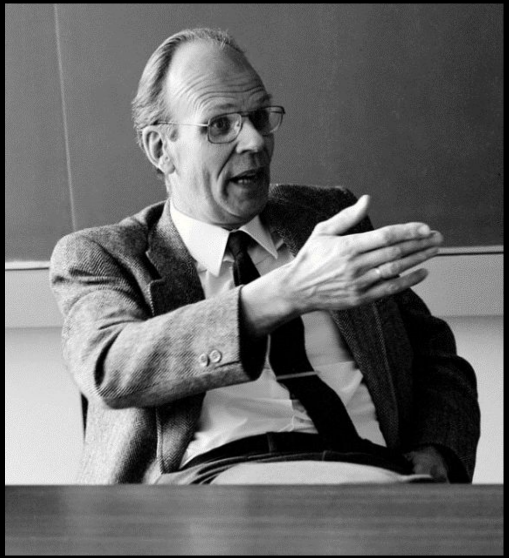


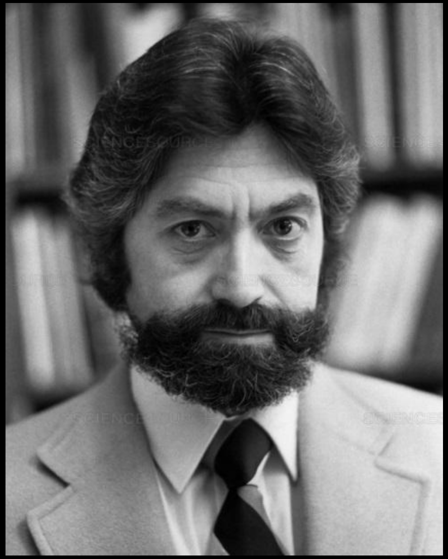


They both write up their ideas, but Zweig is prevented from publishing because of an extreme position taken by the CERN theory division head, L. van Howe

Zweig's recollection

... I wanted to send a paper to the Physical Review, but the head of the Theory Division, Leon Van Hove, wouldn't allow it. He told me that all reports from CERN had to be published in European journals, even though American institutions paid my salary, overhead, and publication costs. When I asked the theory secretary, Madame Fabergé, to retype the paper for publication, she politely refused, saying that Van Hove had instructed her not to type any of my papers. This was a real problem because I didn't know how to type, and didn't have a typewriter (remember, I was trained as a typesetter, not a typist).
I was scheduled to give a theory seminar at CERN titled ``Dealer's choice: Aces are Wild". Van Hove cancelled the seminar, and I was not allowed to reschedule it. When Van Hove and Kokedee published a book four years later reprinting articles on the quark model they did not include either of the CERN reports. Van Hove deliberately and systematically tried to keep my work from public view. ..

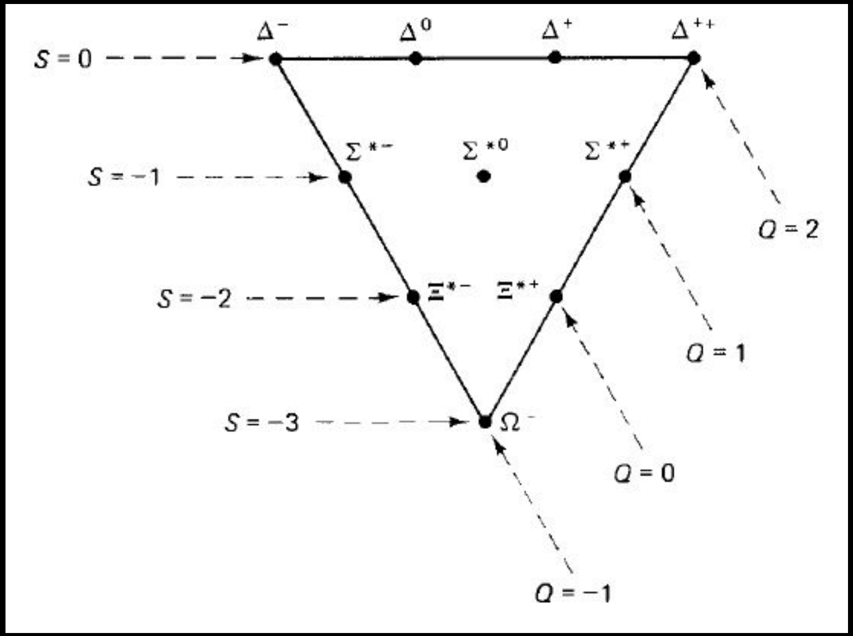


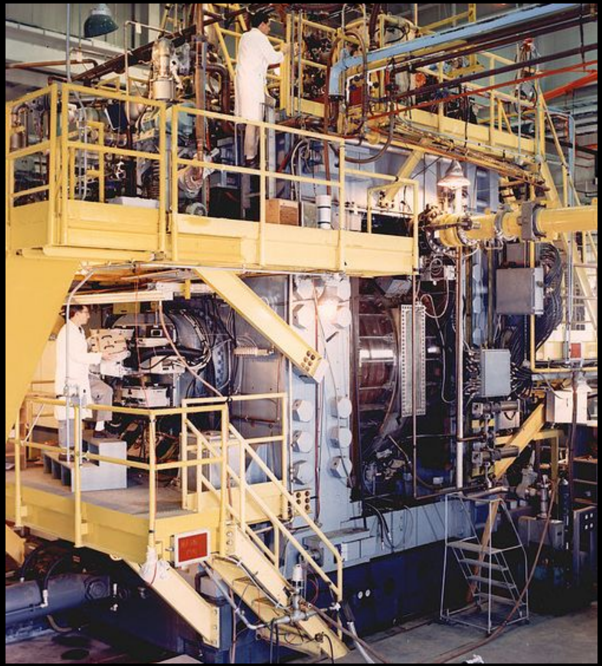


The discovery of Ω^-

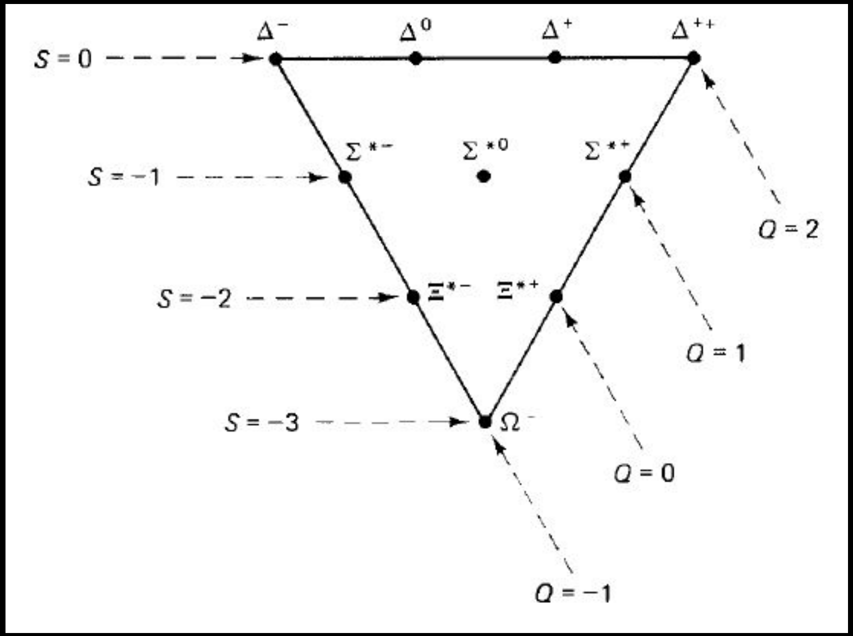
In 1964, at the Brookhaven labs a 5 GeV K^- beam interacts with protons in the target, generating the following sequence:

$$K^- p \rightarrow \Omega^- K^+ K^0 \rightarrow \Xi^0 \pi^- \rightarrow \Lambda \pi^0$$





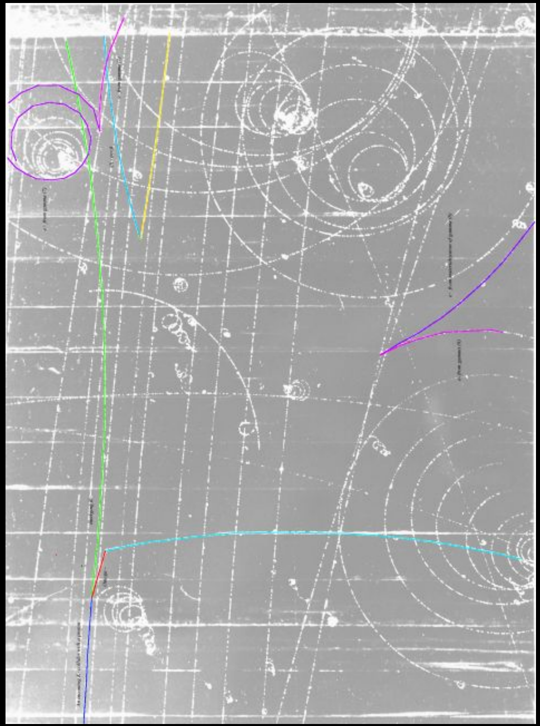
The discovery of Ω^-

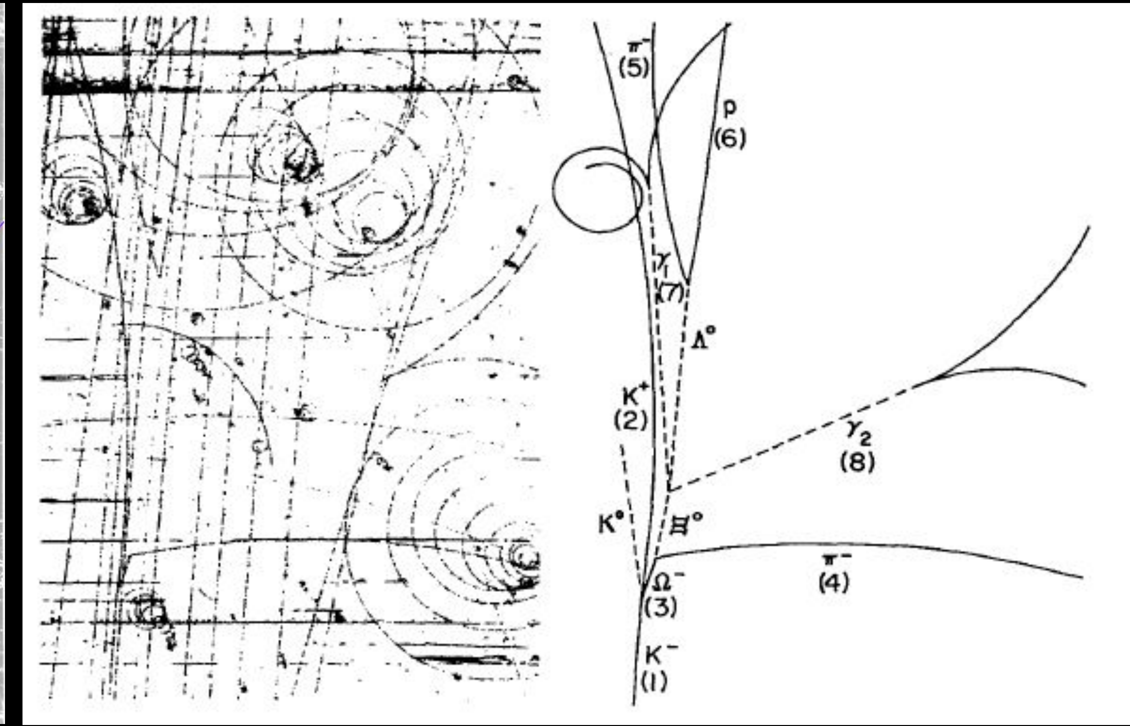


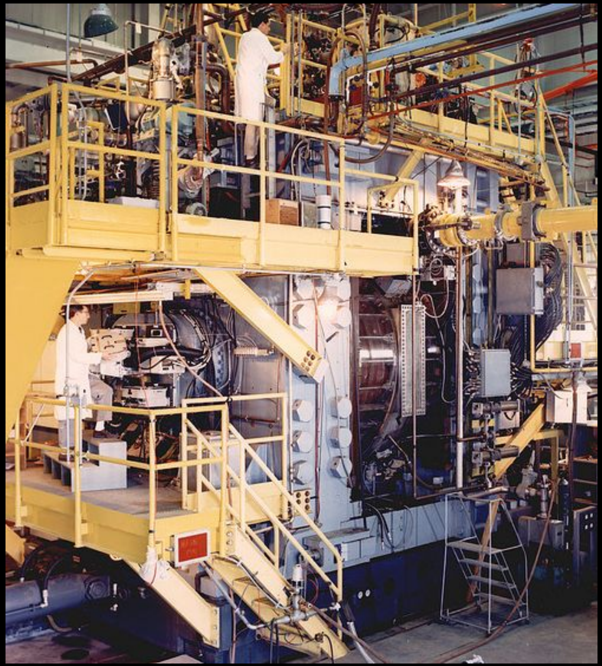
In 1964, at the Brookhaven labs a 5 GeV K⁻ beam interacts with protons in the target, generating the following sequence:

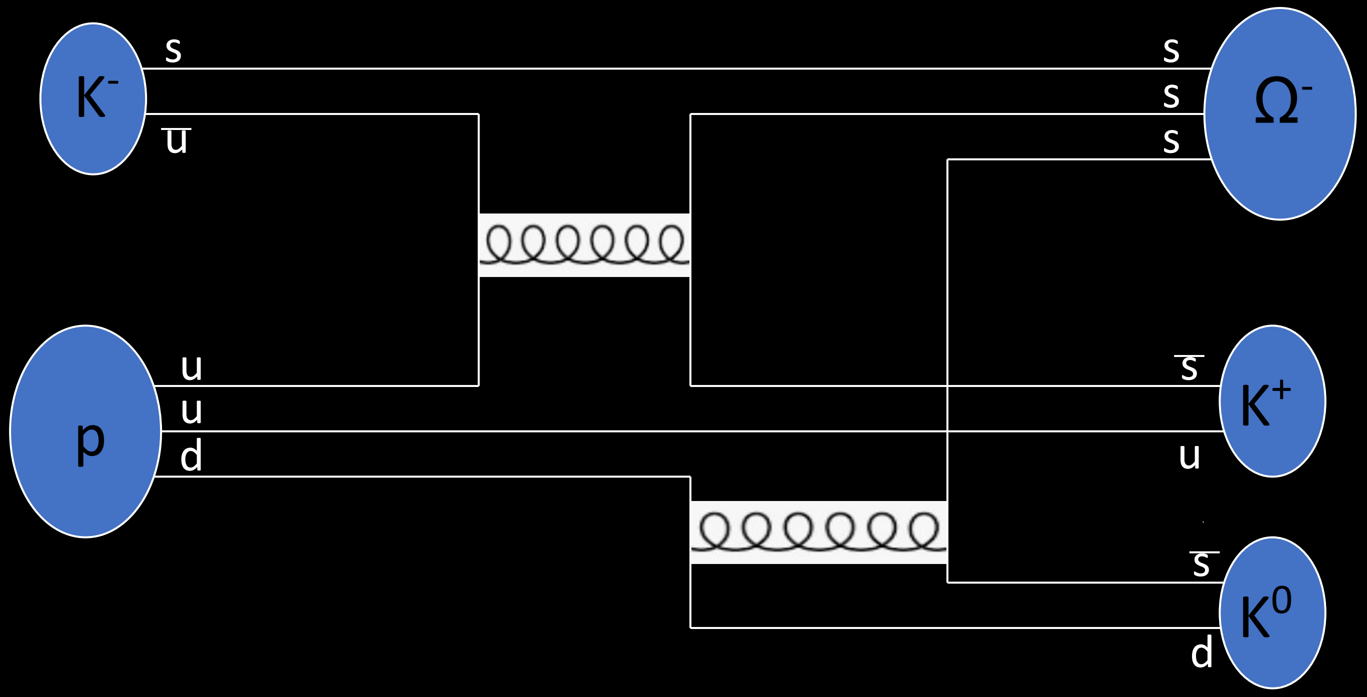
$$K^- p \rightarrow \Omega^- K^+ K^0 \rightarrow \Xi^0 \pi^- \rightarrow \Lambda \pi^0$$

The lambda then decays to proton and pion, and the two photons emitted by pion decay convert (such luck!) into electron-positron pair within the detector volume. Kinematic closure allows to compute the mass of the hypothetical opega particle, in excellent agreement with predictions.



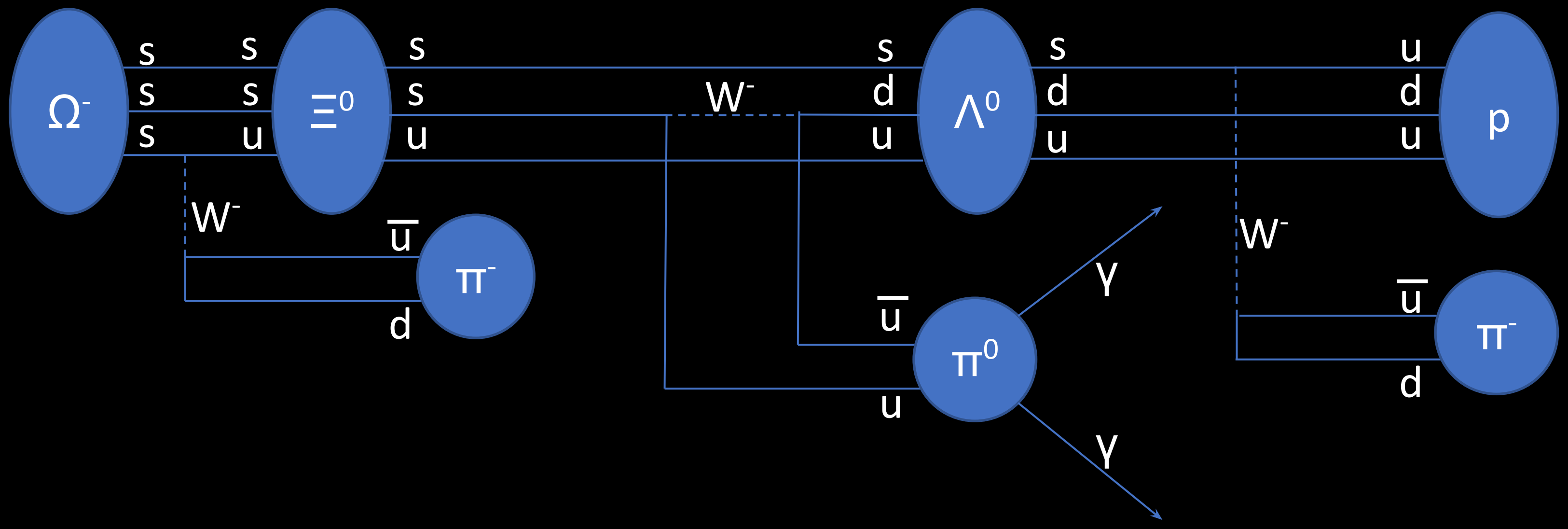






The observed production and decay processes are quite complex and the possibility to see the same reaction are slim.

Replicability is not guaranteed in any given time, although the same reaction will eventually be seen again if sufficient data is collected…



The Birth of the Five-Sigma Criterion

Arthur H. Rosenfeld (Univ. Berkeley)

Statistical Significance: What it is

Statistical significance reports the probability that an experiment obtains data at least as discrepant as those actually observed, under a given "null hypothesis“ H_0

- In physics H_0 *usually* describes the currently accepted and established theory

Statistical Significance: What it is

Statistical significance reports the probability that an experiment obtains data at least as discrepant as those actually observed, under a given "null hypothesis“ H_0

- In physics H_0 *usually* describes the currently accepted and established theory

Given some **data X** and a suitable **test statistic T** (a function of X), one may obtain a **p**-value as the probability of obtaining a value of T at least as extreme as the one observed, if H_0 is true. A way to do that is e.g. Wilks' theorem (discussed later).

p can then be converted into the corresponding number of "sigma," *i.e.* standard deviation units from a Gaussian mean. This is done by finding **x** such that the integral from **x** to infinity of a unit Gaussian N(0,1) equals **p**:

$$\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}\,dt = p$$

Statistical Significance: What it is

Statistical significance reports the probability that an experiment obtains data at least as discrepant as those actually observed, under a given "null hypothesis“ H_0

- In physics H_0 *usually* describes the currently accepted and established theory

Given some **data X** and a suitable **test statistic T** (a function of X), one may obtain a **p**-value as the probability of obtaining a value of T at least as extreme as the one observed, if H_0 is true. A way to do that is e.g. Wilks' theorem (discussed later).

p can then be converted into the corresponding number of "sigma," *i.e.* standard deviation units from a Gaussian mean. This is done by finding **x** such that the integral from **x** to infinity of a unit Gaussian N(0,1) equals **p**:

$$\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}\,dt = p$$

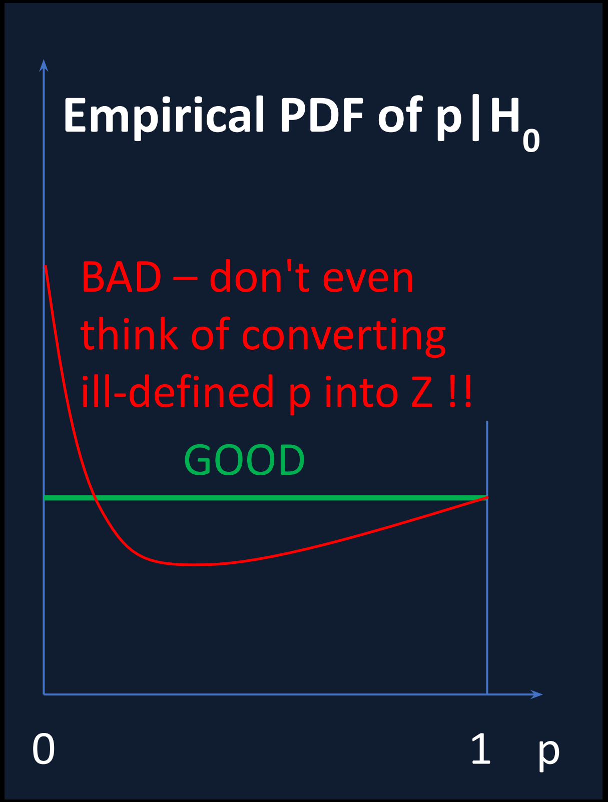
According to the above recipe, a **15.9%** probability is a one-standard-deviation effect; a **0.135%** probability is a three-standard-deviation effect; and a **0.0000285%** probability corresponds to five standard deviations - "**five sigma**" in jargon.

Notes

The convention is to use a "one-tailed" Gaussian: we do not care about departures of x from the mean in the *un-interesting direction*

The whole construction rests on a proper definition of the p-value. Any shortcoming of the properties of p (*e.g*. a tiny non-flatness of its PDF under the null hypothesis) totally invalidates the meaning of the derived Nσ

<ins>Again, note: the "*probability of the data*" is not used</ins>. What is used is the probability of a subset of the possible outcomes of the experiment, defined by the outcome actually observed (**as much** or **more** extreme)



Far-Out Hadrons

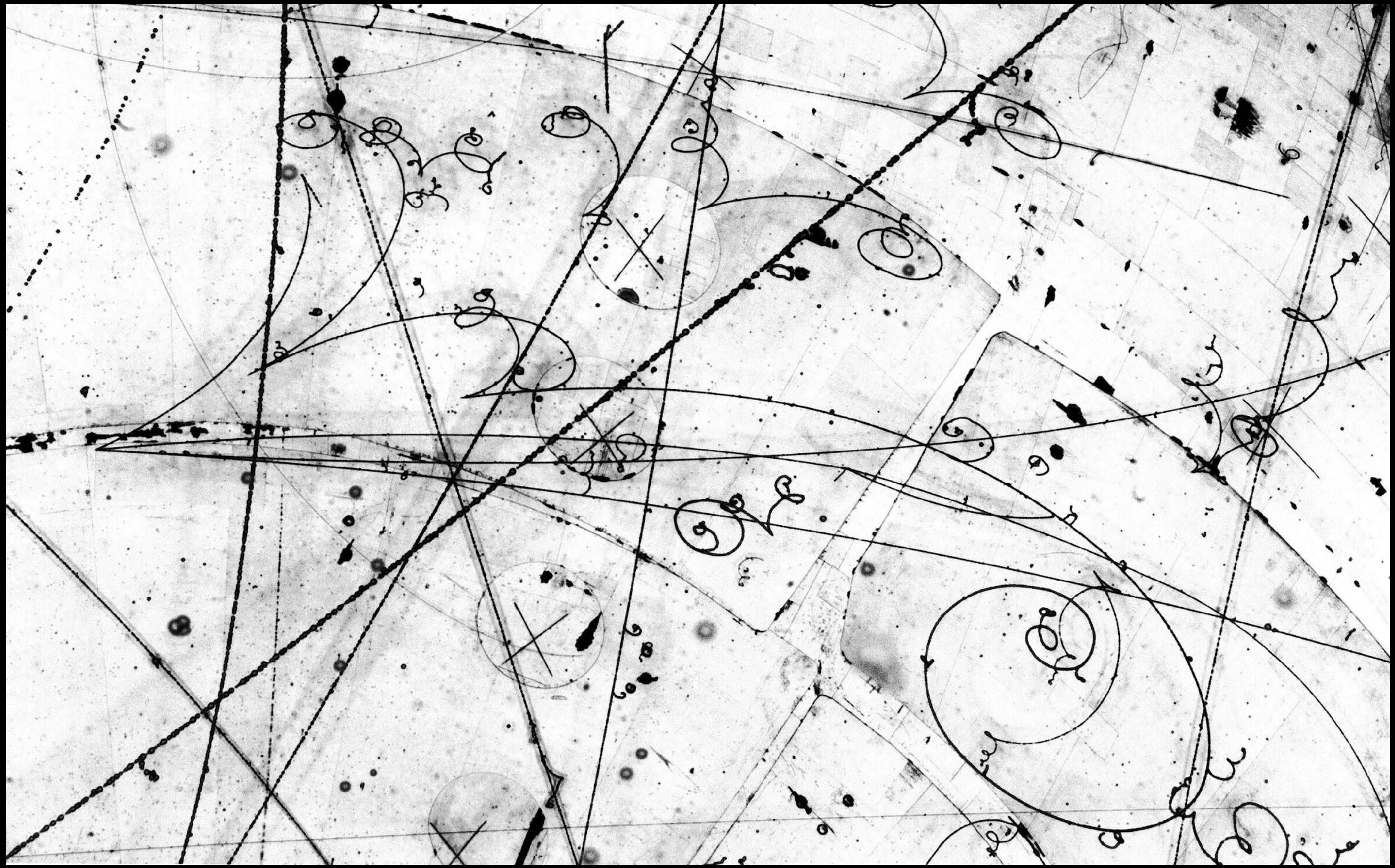
- In 1968 Arthur Rosenfeld wrote a paper titled "*Are There Any Far-out Mesons or Baryons?*“. In it, he demonstrated that the number of claims of discovery of exotic particles published in scientific magazines **agreed with the number of statistical fluctuations** that one would expect in the analyzed datasets.

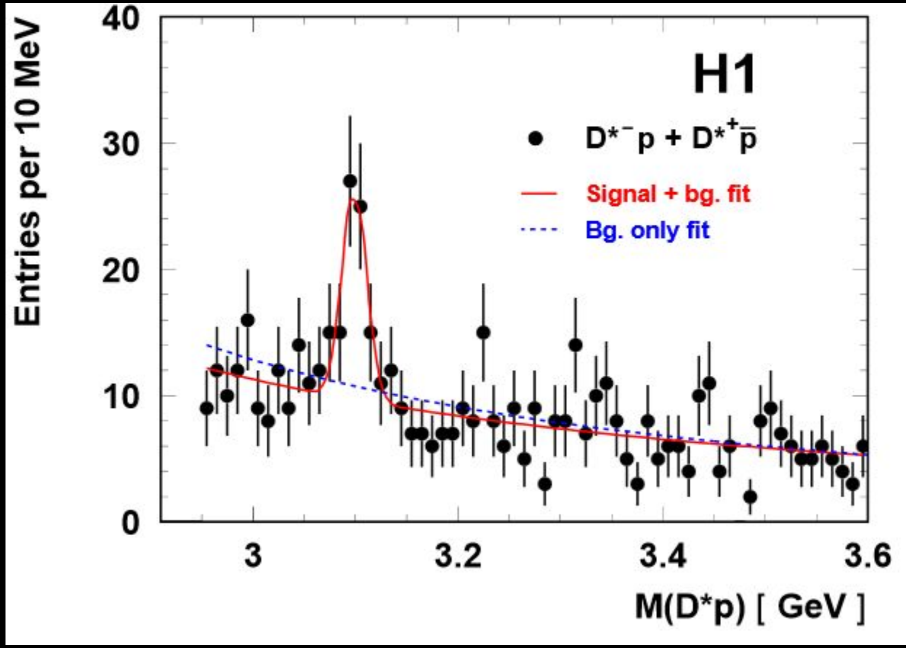
  (“Far-out hadrons” are particles not fitting in SU(3) multiplets)

- Rosenfeld pointed his finger at large trial factors coming into play due to the massive use of combinations of observed particles in deriving mass spectra containing potential resonances:

  > *“[...] This reasoning on multiplicities, extended to all combinations of all outgoing particles and to all countries, leads to an estimate of 35 million mass combinations calculated per year. How many histograms are plotted from these 35 million combinations? A glance through the journals shows that a typical mass histogram has about 2,500 entries, so the number we were looking for, h is then 15,000 histograms per year […]"*

Bubble chamber images and bump hunts





More Rosenfeld

“[...] Our typical 2,500 entry histogram seems to average 40 bins. This means that therein a physicist could observe 40 different fluctuations one bin wide, 39 two bins wide, 38 three bins wide... This arithmetic is made worse by the fact that when a physicist sees 'something', he then tries to enhance it by making cuts...”

“In summary of all the discussion above, I conclude that each of our 150,000 annual histograms is capable of generating somewhere between 10 and 100 deceptive upward fluctuations [...]”.

That was indeed a problem! Rosenfeld concluded:

*“To the theorist or phenomenologist the moral is simple: **wait for nearly 5σ effects**. For the experimental group who has spent a year of their time and perhaps a million dollars, the problem is harder... go ahead and publish... but they should realize that any bump less than about 5σ **calls for a repeat of the experiment.**”*

Gerry Lynch and GAME

- Rosenfeld's article also cites the half-joking, half-didactical effort of his colleague Gerry Lynch at Berkeley:

  "My colleague Gerry Lynch has instead tried to study this problem 'experimentally' using a 'Las Vegas' computer program called Game […]

  When a friend comes showing his latest 4-sigma peak,

  You draw a smooth curve […] (based on the hypothesis that the peak is just a fluctuation) [and] call for 100 Las Vegas histograms […]

  You and your friend then go around the halls, asking physicists to pick out the most surprising histogram in the printout. Often it is one of the 100 phoneys, rather than the real '4-sigma' peak."

The proposal to raise to 5-sigma of the threshold above which a signal could be claimed was an earnest attempt at reducing the flow of claimed discoveries, which distracted theorists and caused confusion.

The Special Nature of Subnuclear Physics: 3 – Experiment cost and lifetime

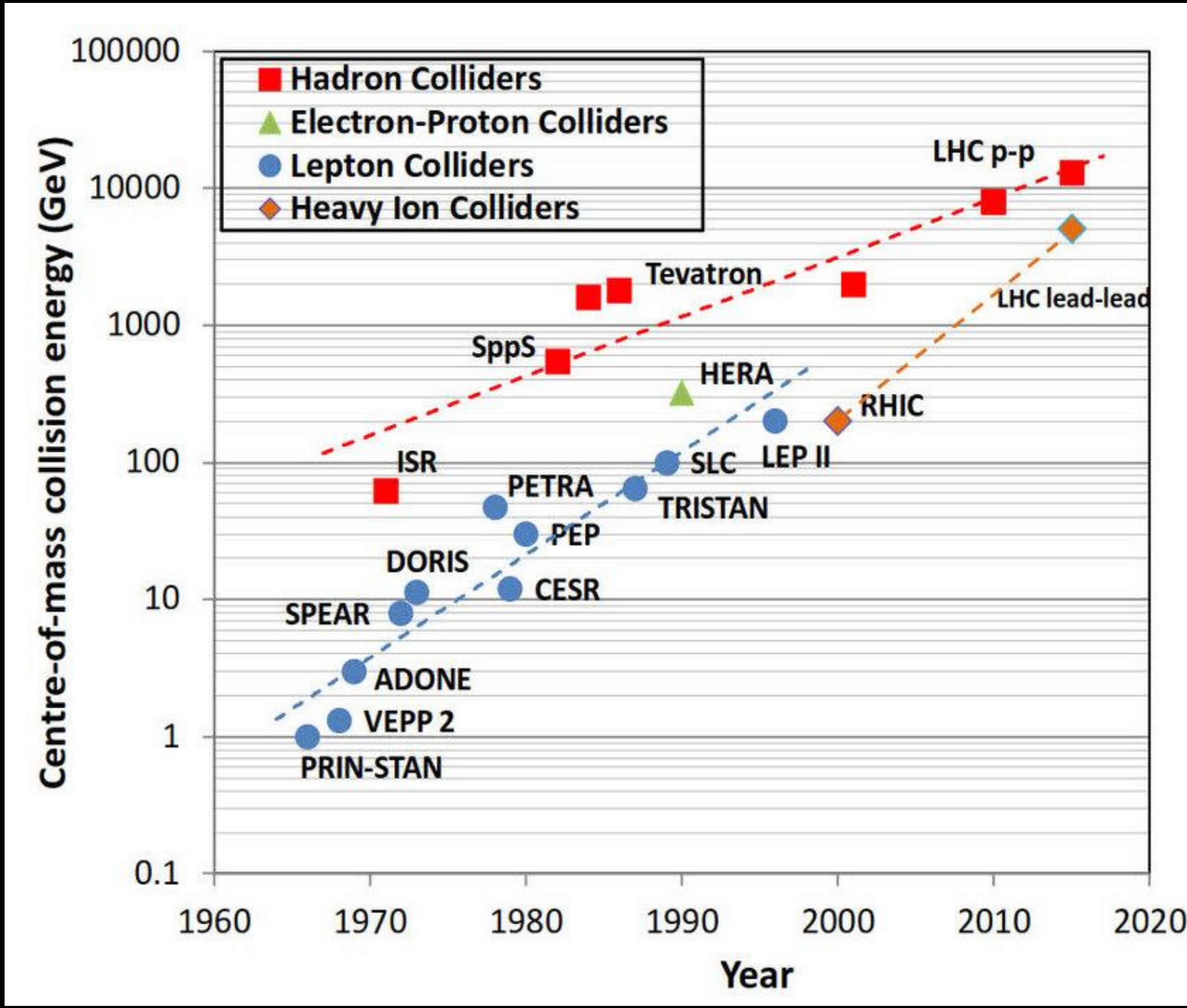
Detector cost

Two experiments are better than one

Getting old while you build it

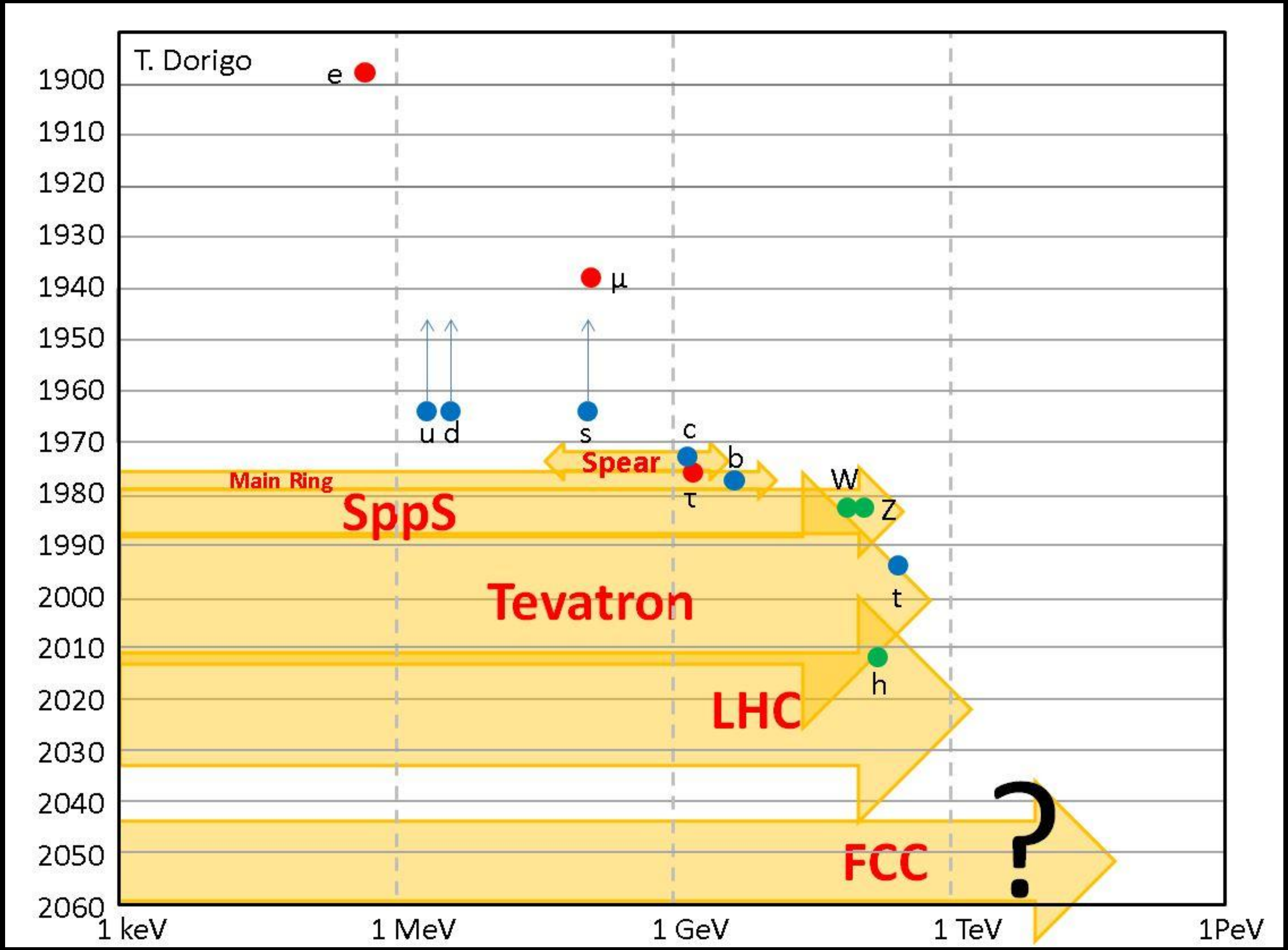
Collision energy

The graph shows the energy of collisions in large accelerator experiments versus time



Particle discoveries versus time

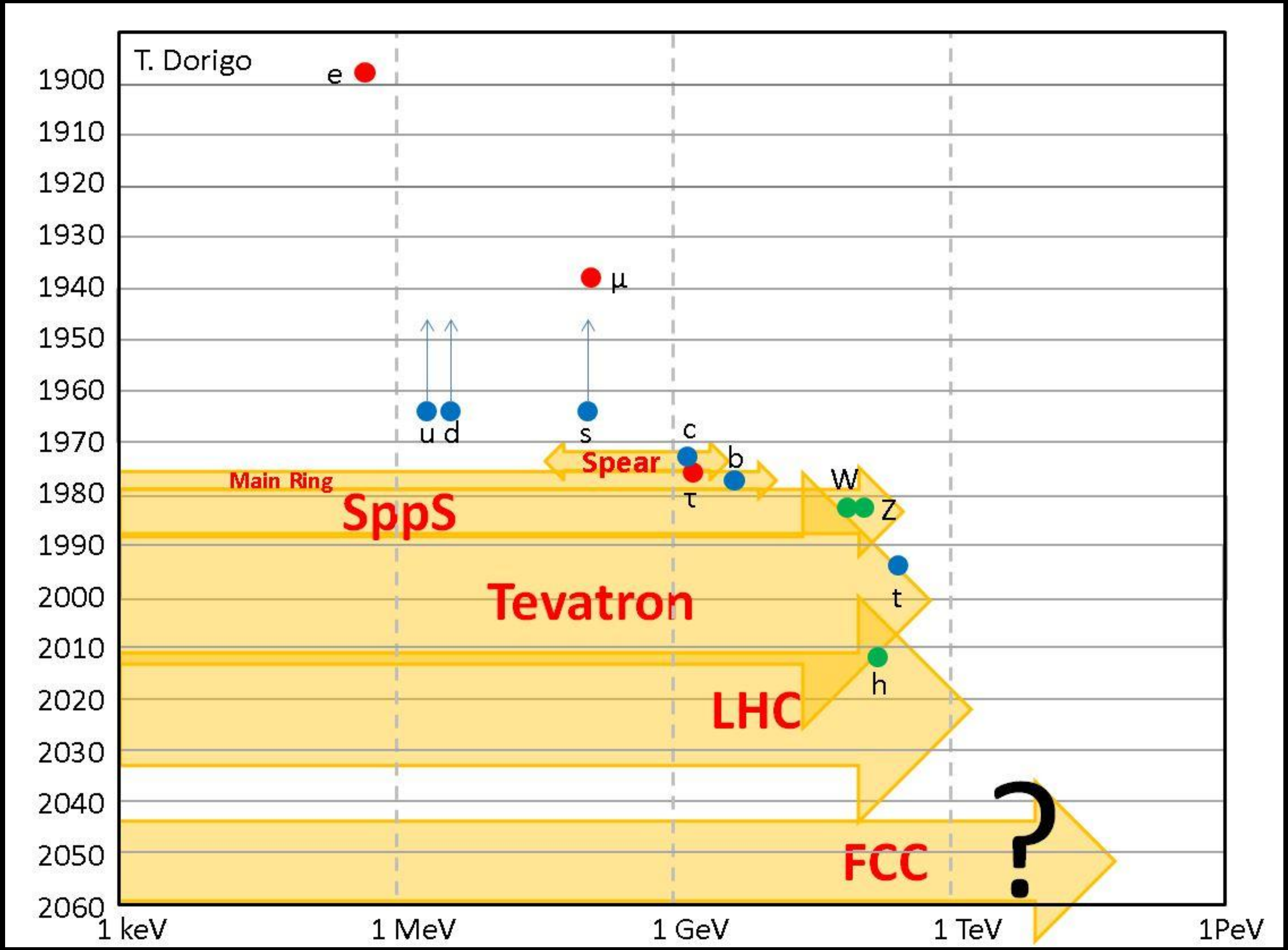
Whenever a new energy range was probed, new discoveries were made



Particle discoveries versus time

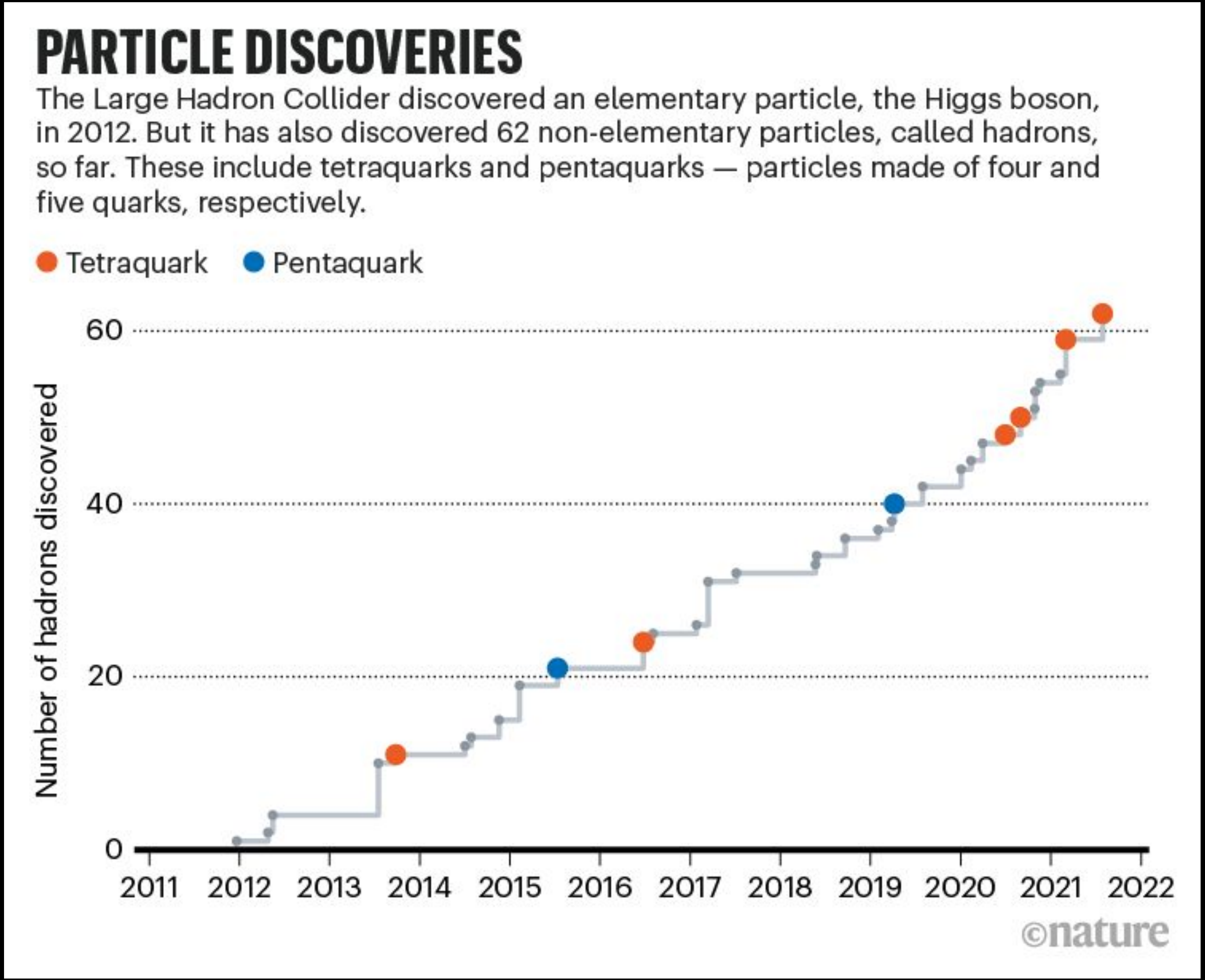
Whenever a new energy range was probed, new discoveries were made

In particle physics, the issue today is not reproducibility, but producibility!



More to it than the Higgs:

Composite hadrons



Colliders have >=2 experiments

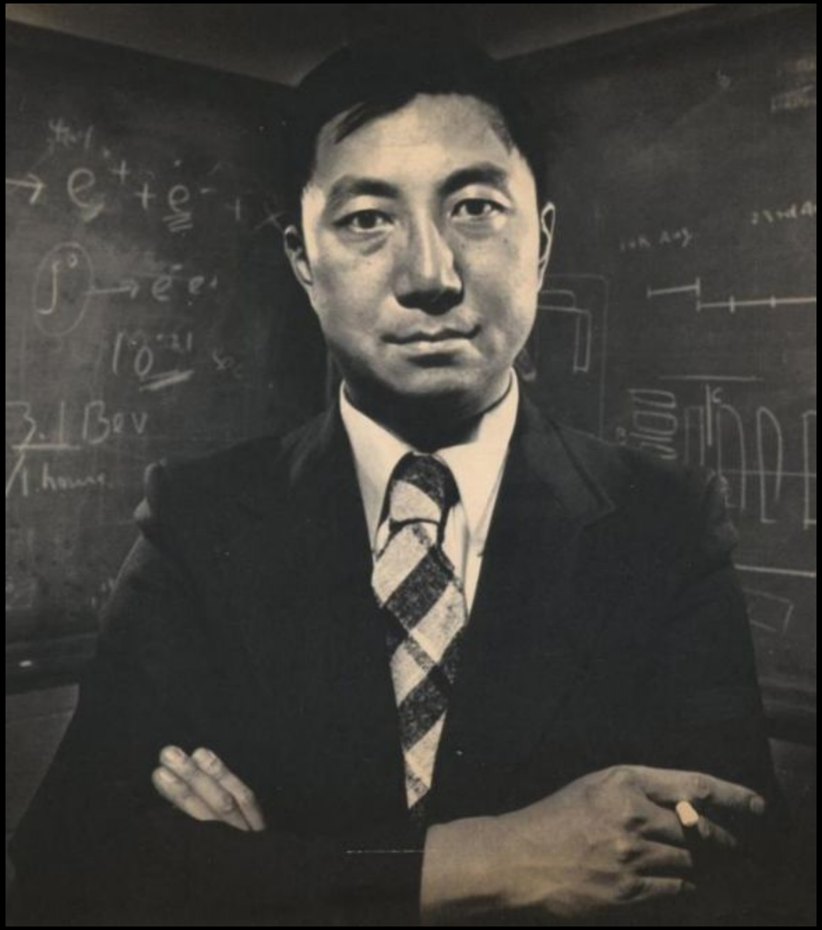
Not a general rule, but a sound practice – experiments may fail!

The practice was initiated by Rubbia (UA1, UA2 was a “backup”).
Redundance rather than reproducibility was the issue

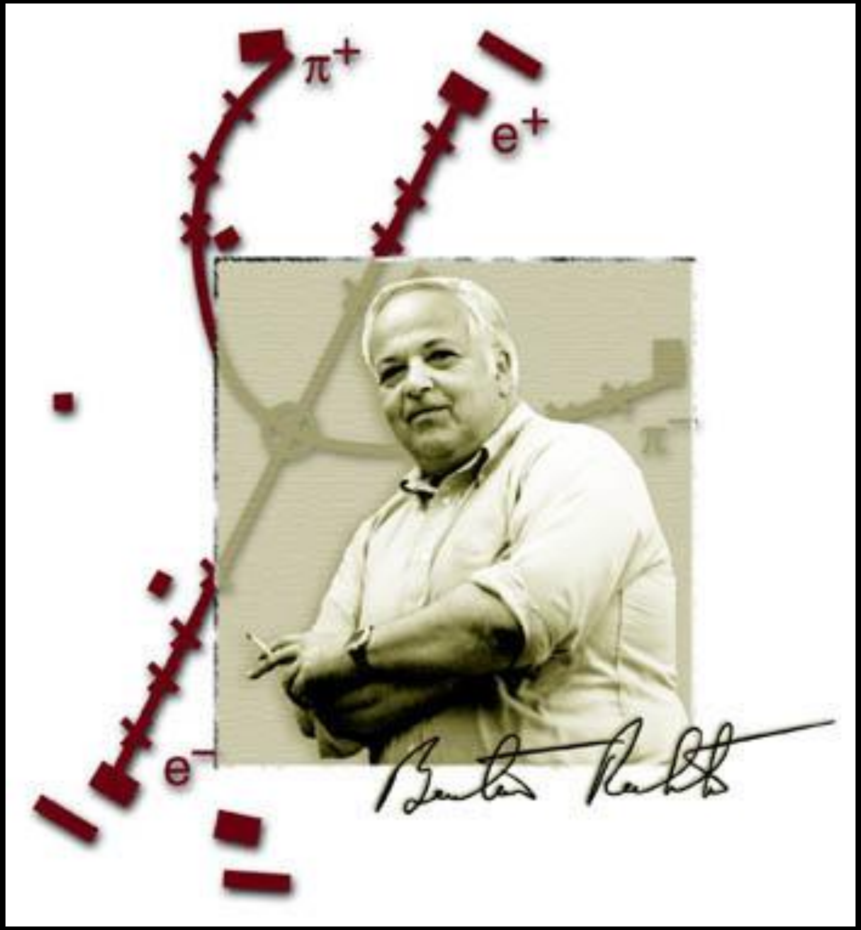
Anyway, a good idea that continues to be used.

And not necessarily a hindrance to the chance to be the only discoverer!

When the claim is steep, you pause…

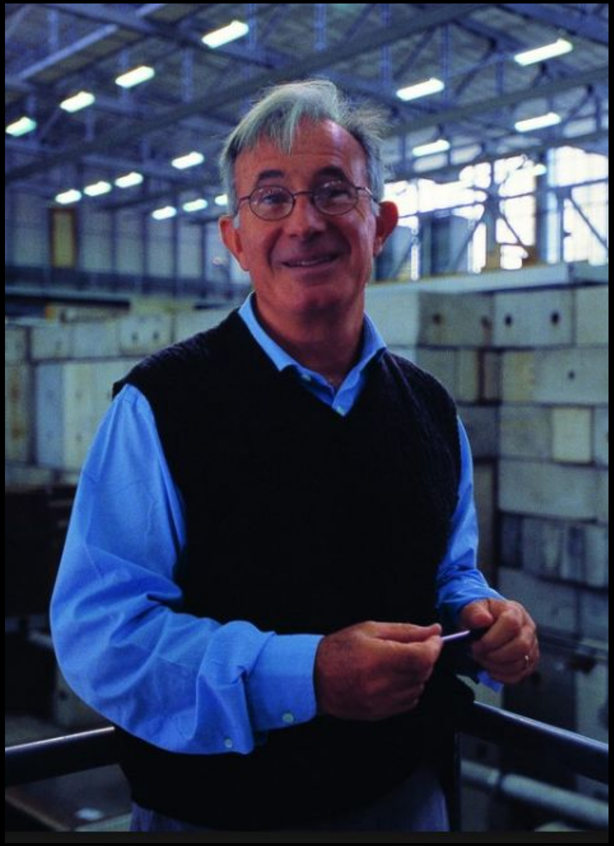


Samuel Ting



Burton Richter

Carlo Rubbia



Luigi Di Lella

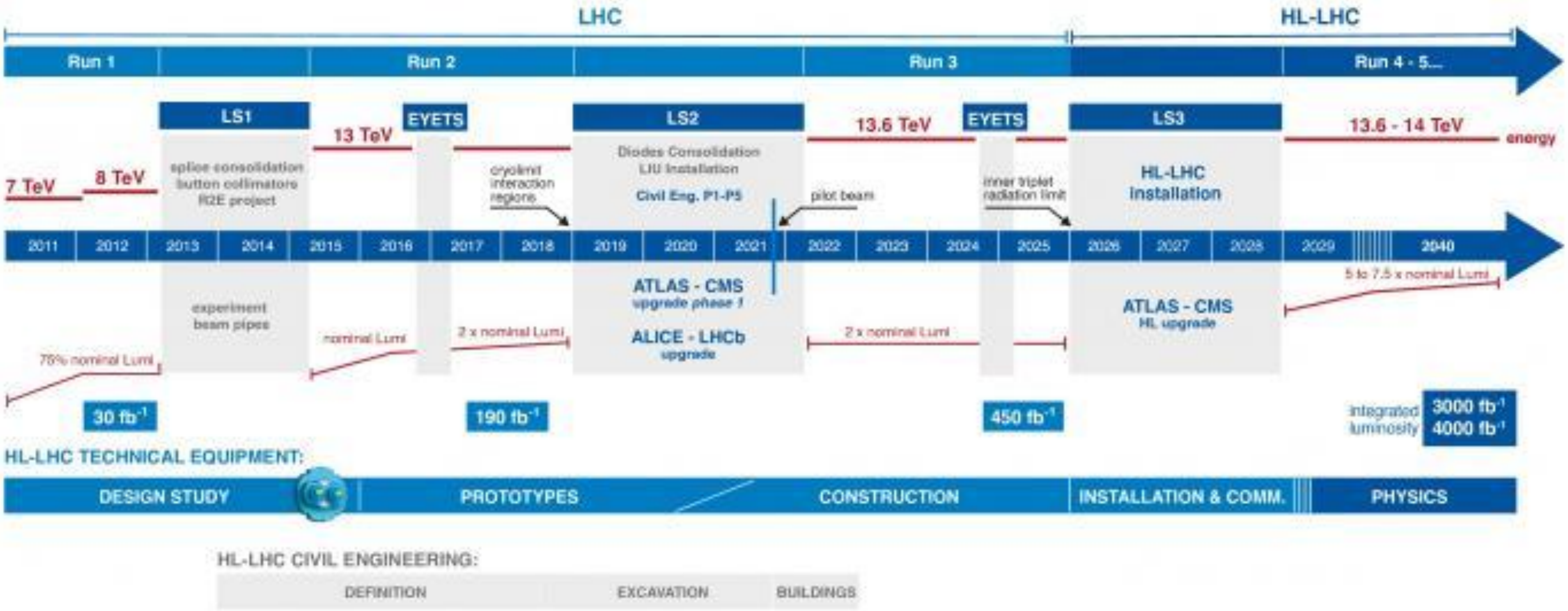
Simon van der Meer

From Blueprint to Stockholm

The scale of today’s large collider experiments creates a shear between design and fruition time

- Another challenge to replicability

LHC / HL-LHC Plan



Measurements Adrift

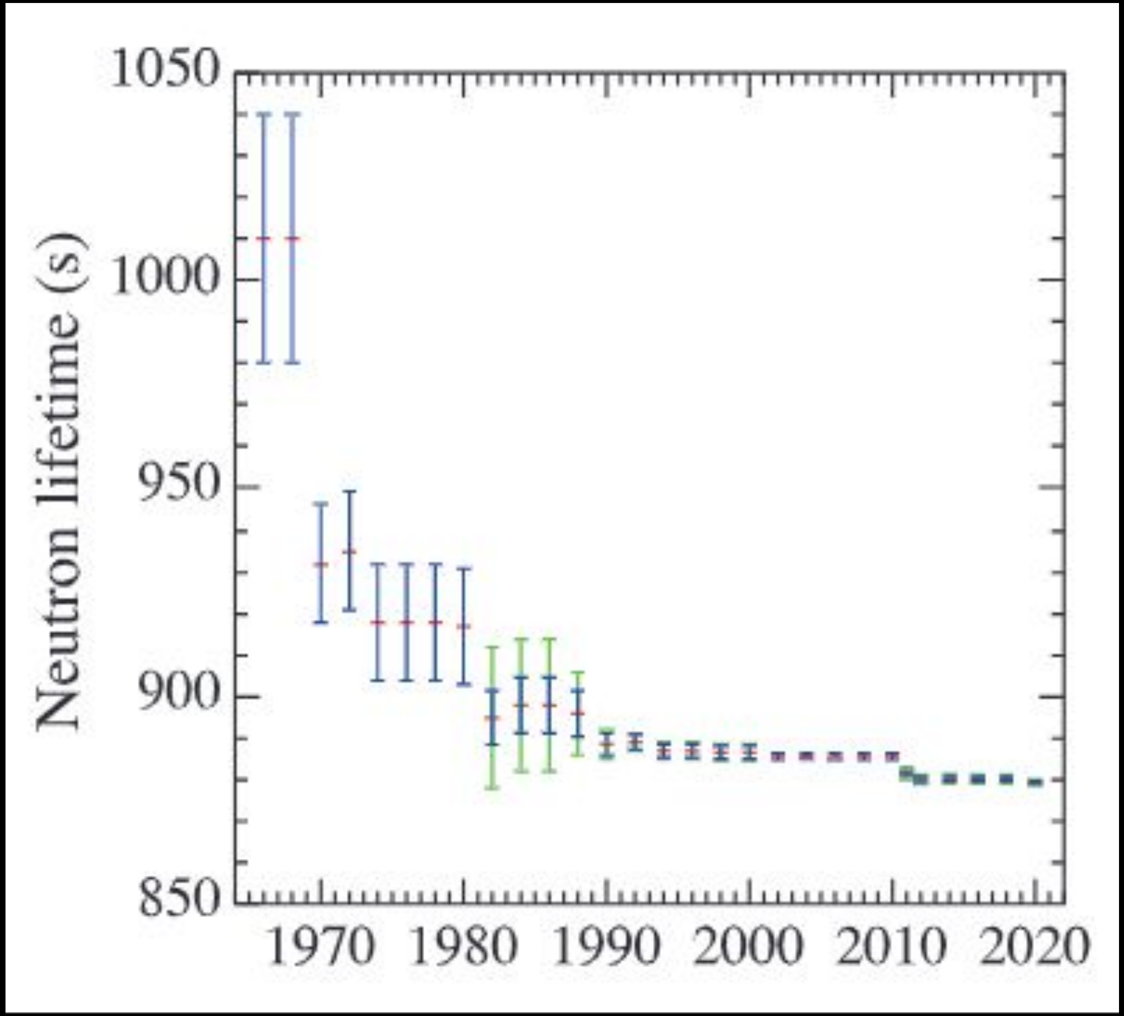
The history of particle physics is studded with juicy stories of measurements gone awry – but these are good examples of success of the replicability idea

Measurements Adrift

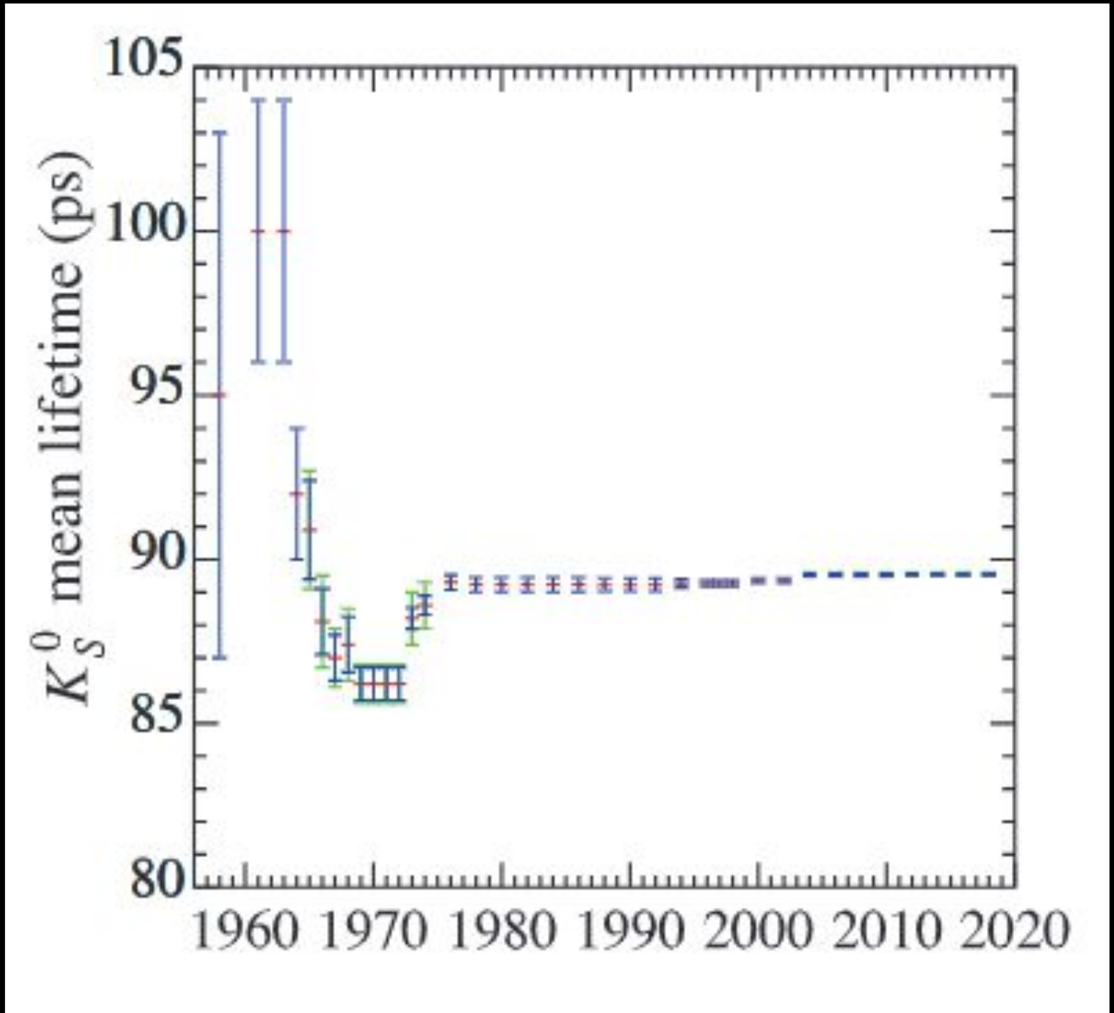
The history of particle physics is studded with juicy stories of measurements gone awry – but these are good examples of success of the reproducibility procedures

- Maybe more interesting: when constants of nature show a trend versus time

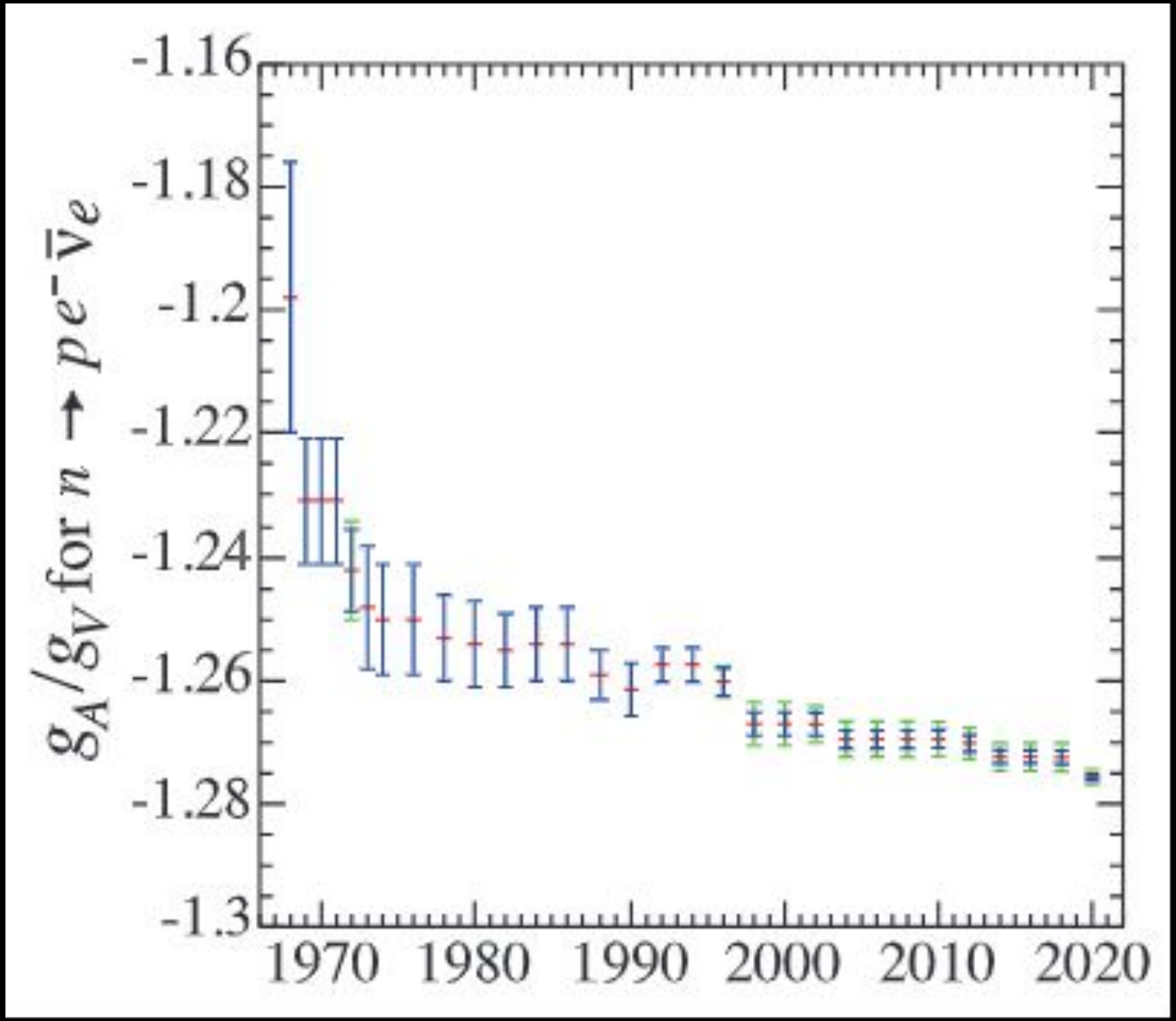
The lifetime of the neutron



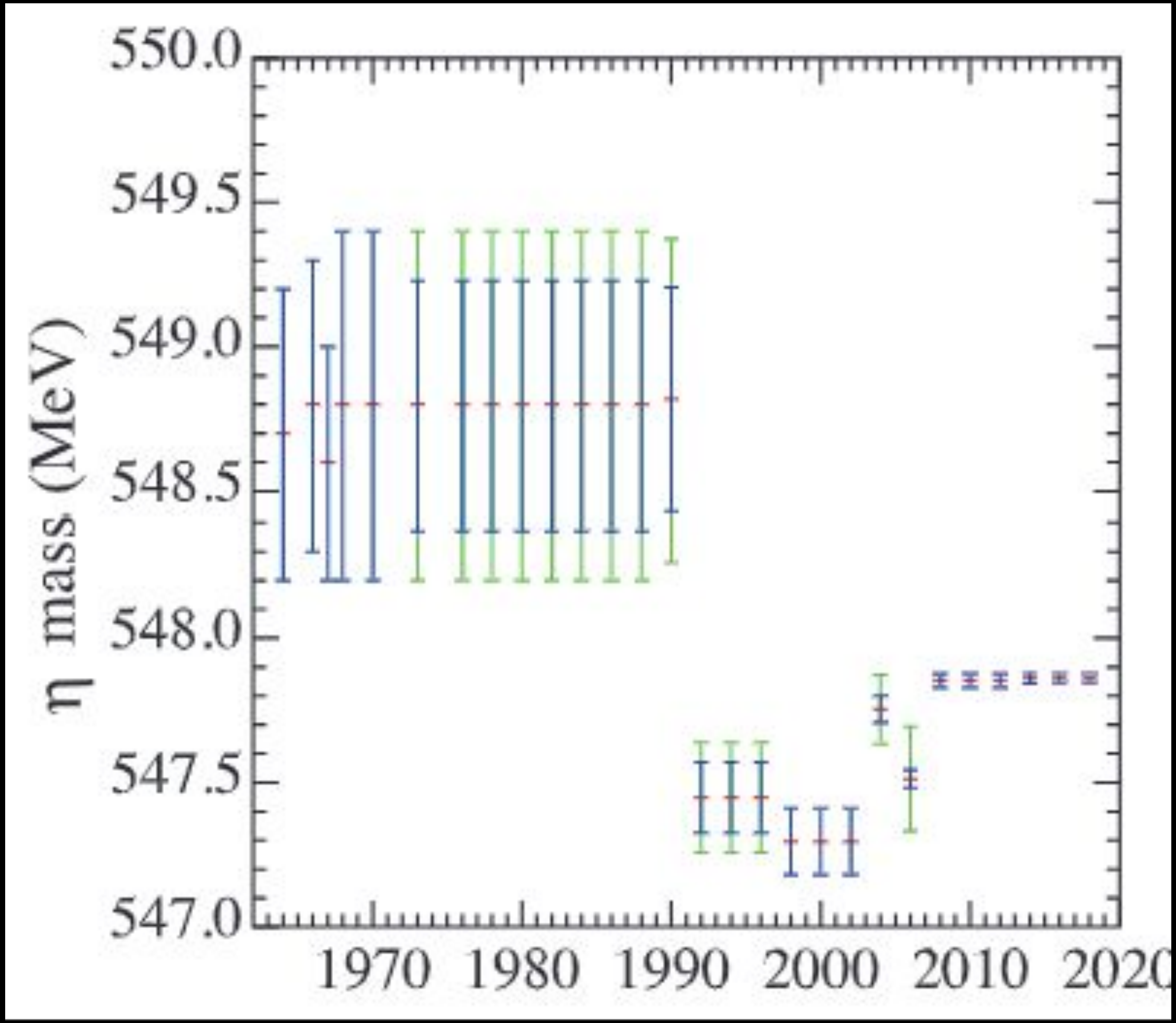
The lifetime of the neutral kaon



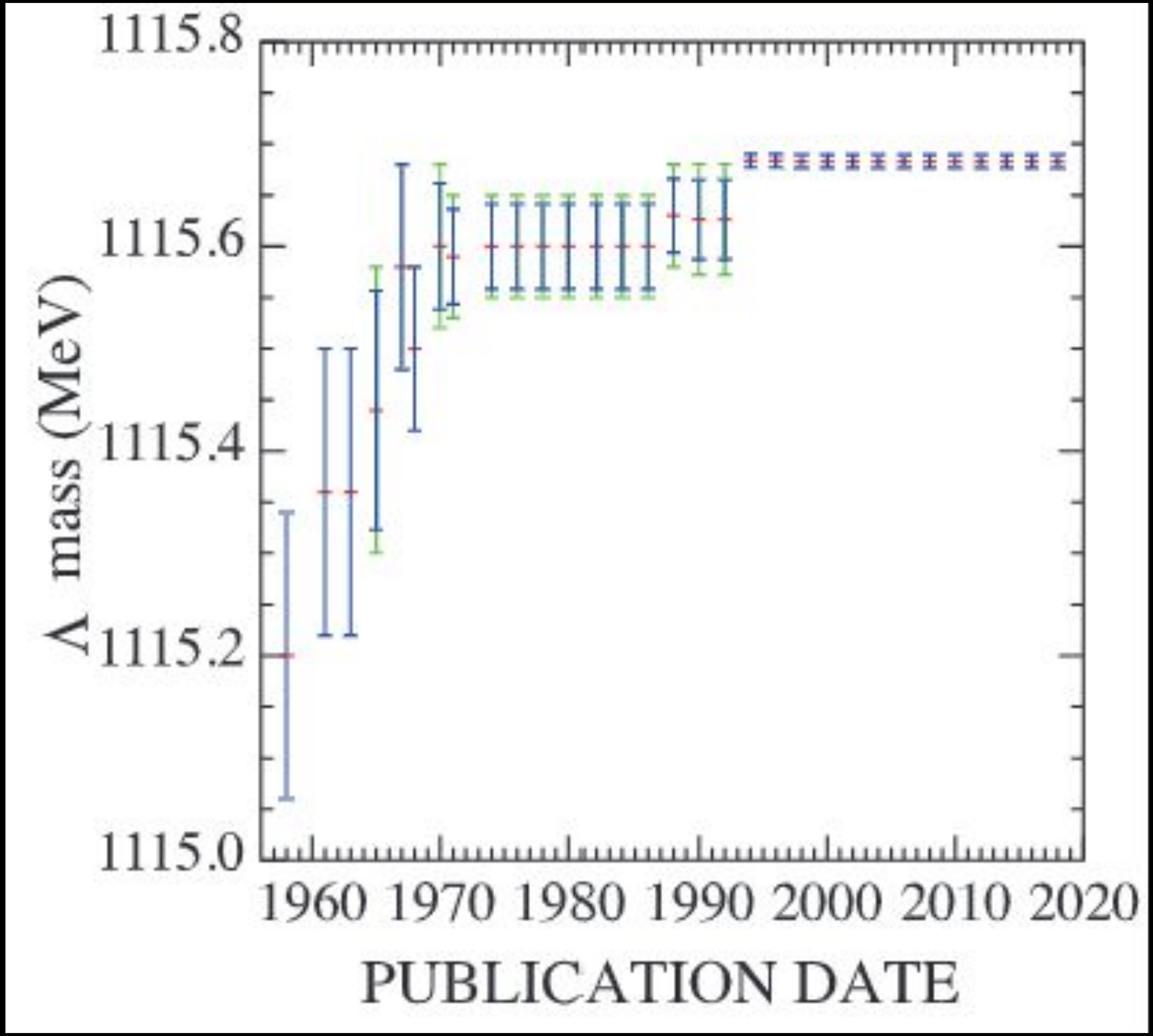
The Axial/Vector coupling ratio



The mass of the Eta meson



The mass of the Lambda baryon



The Particle Data Group to the Rescue

Since a couple of decades, the PDG (a 1000+ page handbook containing ALL measurements of particle properties and related quantities, published yearly) has in place an inflationary mechanism for uncertainty bars

The Particle Data Group to the Rescue

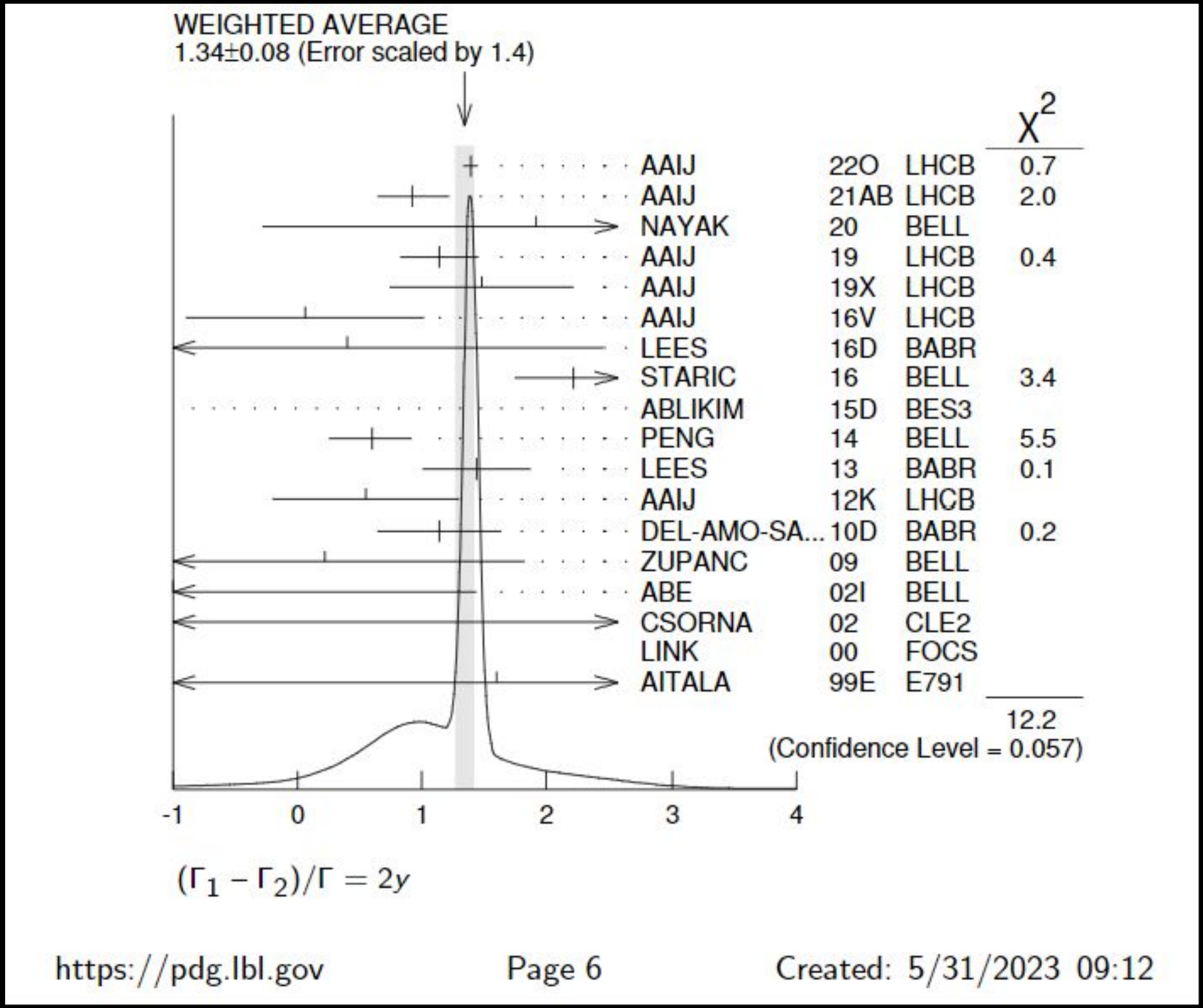
Since a couple of decades, the PDG (a 1000+ page handbook containing ALL measurements of particle properties and related quantities, published yearly) has in place an inflationary mechanism for uncertainty bars

This "PDG method" is applied to weighted averaging of different experimental measurements of the same physical quantity.

It consists in scaling up the uncertainty bars of results that look suspiciously off, or of the weighted average overall, to improve the probability of the set of measurements

The central value is unaffected by the procedure

Example: the D0 meson Oscillation Parameter



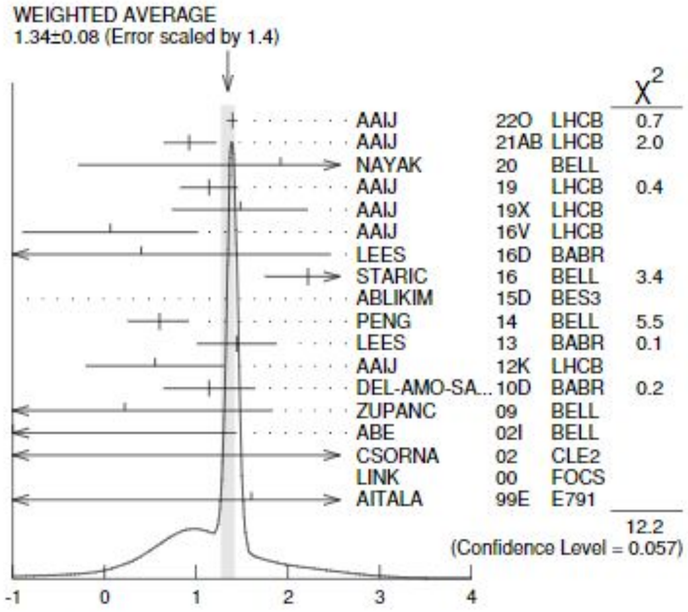
Due to the strong phase difference between $D^0 \to K^+\pi^-$ and $\overline{D}^0 \to K^+\pi^-$, we exclude from the average those measurements of y' that are inferred from the D^0-$\overline{D}^0$ mixing ratio $\Gamma(K^+\pi^-$ via $\overline{D}^0)$ / $\Gamma(K^+\pi^-)$ given near the end of this D^0 Listings. OUR AVERAGE assumes CP conservation, though this is disfavored by data.

In the absence of CP violation, the experimental measurement of the observable y_{CP} corresponds to y. In the presence of CP violation, y_{CP} is approximately equal to y up to a CP-violating weak phase cosine, that is very close to unity, and higher order CP-violating effects. Many experiments measure y_{CP} with respect to a reference channel, such that the quoted results below are often $y_{CP} \pm y_{CP}$(reference). We denote the actual observable and the reference channel below for those experiments with sufficient sensitivity such that y_{CP}(reference) cannot be neglected.

Some early results have been omitted. See our 2006 *Review* (Journal of Physics **G33** 1 (2006)).

"OUR EVALUATION" comes from CPV allowing averages provided by the Heavy Flavor Averaging Group, see the note on "D^0-$\overline{D}^0$ Mixing."

VALUE (units 10^{-2})	*EVTS*	*DOCUMENT ID*		*TECN*	*COMMENT*
1.394 ± 0.056 OUR EVALUATION					
1.34 ± 0.08 OUR AVERAGE Error includes scale factor of 1.4. See the ideogram below.					
$1.392 \pm 0.052 \pm 0.026$		[1] AAIJ	22O	LHCB	$y_{CP} - y_{CP}(K\pi)$, pp at 13 TeV
$0.92 {}^{+0.30}_{-0.28}$	30.6M	[2] AAIJ	21AB	LHCB	pp at 13 TeV
$1.92 \pm 1.82 {}^{+1.29}_{-1.24}$	91k	[3] NAYAK	20	BELL	$D^0 \to K^0_S \omega$
$1.14 \pm 0.26 \pm 0.18$		[4] AAIJ	19	LHCB	pp at 7, 8 TeV
1.48 ± 0.74	2.3M	[5] AAIJ	19X	LHCB	$D^0 \to K^0_S \pi^+\pi^-$
		[6] AAIJ	18K	LHCB	pp at 7, 8, 13 TeV
$0.06 \pm 0.92 \pm 0.26$		[7] AAIJ	16V	LHCB	pp at 7 TeV
$0.4 \pm 1.8 \pm 1.0$		[8] LEES	16D	BABR	e^+e^-, 10.6 GeV
$2.22 \pm 0.44 \pm 0.18$		[9] STARIC	16	BELL	$e^+e^- \to \Upsilon(nS)$
$-4.0 \pm 2.6 \pm 1.4$		[10] ABLIKIM	15D	BES3	e^+e^- at $\psi(3770)$
		[11] KO	14	BELL	$e^+e^- \to \Upsilon(nS)$
$0.60 \pm 0.30 {}^{+0.10}_{-0.17}$		[12] PENG	14	BELL	$e^+e^- \to \Upsilon(nS)$
		[13] AALTONEN	13AE	CDF	$p\overline{p}$ at 1.96 TeV
$1.44 \pm 0.36 \pm 0.24$		[14] LEES	13	BABR	$e^+e^- \to \Upsilon(4S)$
$0.55 \pm 0.63 \pm 0.41$		[15] AAIJ	12K	LHCB	pp at 7 TeV
$1.14 \pm 0.40 \pm 0.30$		[16] DEL-AMO-SA...	10D	BABR	e^+e^-, 10.6 GeV
$0.22 \pm 1.22 \pm 1.04$		[17] ZUPANC	09	BELL	$e^+e^- \approx \Upsilon(4S)$
$-1.0 \pm 2.0 {}^{+1.4}_{-1.6}$	18k	[18] ABE	02I	BELL	$e^+e^- \approx \Upsilon(4S)$
$-2.4 \pm 5.0 \pm 2.8$	3393	[19] CSORNA	02	CLE2	$e^+e^- \approx \Upsilon(4S)$
$6.84 \pm 2.78 \pm 1.48$	10k	[18] LINK	00	FOCS	γ nucleus
$+1.6 \pm 5.8 \pm 2.1$		[18] AITALA	99E	E791	$K^-\pi^+$, K^+K^-
• • • We do not use the following data for averages, fits, limits, etc. • • •					
		[20] AAIJ	17AO	LHCB	Repl. by AAIJ 18K
		[21] AAIJ	13CE	LHCB	Repl. by AAIJ 17AO
		[22] AAIJ	13N	LHCB	Repl. by AAIJ 13CE
$2.32 \pm 0.44 \pm 0.36$		[23] AUBERT	09AI	BABR	See LEES 13
$-0.12 {}^{+1.10}_{-1.28} \pm 0.68$		[24] AUBERT	09AN	BABR	e^+e^- at 10.58 GeV
$1.4 {}^{+4.8}_{-5.4}$		[25] LOWREY	09	CLEO	e^+e^- at $\psi(3770)$
1.70 ± 1.52	13k	[26] AALTONEN	08E	CDF	$p\overline{p}$, $\sqrt{s} = 1.96$ TeV
$2.06 \pm 0.66 \pm 0.38$		[27] AUBERT	08U	BABR	See AUBERT 09AI
$1.94 \pm 0.88 \pm 0.62$	4k	[26] AUBERT	07W	BABR	$e^+e^- \approx 10.6$ GeV
$2.62 \pm 0.64 \pm 0.50$	160k	[28] STARIC	07	BELL	Repl. by STARIC 16
$0.74 \pm 0.50 {}^{+0.20}_{-0.31}$	534k	[29] ZHANG	07B	BELL	Repl. by PENG 14
-0.7 ± 4.9	4k	[26,30] ZHANG	06	BELL	e^+e^-
$-3.0 {}^{+5.0}_{-4.8} {}^{+1.6}_{-0.8}$		[29] ASNER	05	CLEO	$e^+e^- \approx 10$ GeV
-0.3 ± 5.7		[26,30] LI	05A	BELL	See ZHANG 06
$-5.2 {}^{+18.4}_{-16.8}$		[26,30] LINK	05H	FOCS	γ nucleus
$1.6 \pm 0.8 {}^{+1.0}_{-0.8}$	450k	[31] AUBERT	03P	BABR	See AUBERT 08U
$1.6 {}^{+6.2}_{-12.8}$		[26,30] AUBERT	03Z	BABR	e^+e^-, 10.6 GeV
$-5.0 {}^{+2.8}_{-3.2} \pm 0.6$		[26] GODANG	00	CLE2	e^+e^-



$(\Gamma_1 - \Gamma_2)/\Gamma = 2y$

[1] AAIJ 22O is the combination of the measurement in the $D^0 \to \pi^-\pi^+$ channel and the one in the $D^0 \to K^-K^+$ channel, that are $(1.314 \pm 0.106 \pm 0.032)$% and $(1.416 \pm 0.060 \pm 0.028)$% respectively, assuming fully correlated systematics except those related to peaking backgrounds.

[2] AAIJ 21AB analysis of $D^0 \to K^0_S \pi^+\pi^-$ events allows for CP violation (none seen).

[3] NAYAK 20 reports $(1.92 \pm 1.82 \pm 1.24 {}^{+0.34}_{-0.00}) \times 10^{-2}$ where the last uncertainty is due to possible presence of CP-even decays in the data sample. Extracts $y_{CP} = (\Gamma_{CP+} - \Gamma_{CP-}) / (\Gamma_{CP+} + \Gamma_{CP-})$ in $D^0 \to K^0_S \omega$ versus $\overline{D}^0 \to K^0_S \omega$, by measuring the decay lifetime of $D^0 \to K^0_S \omega$ with $\omega \to \pi^+\pi^-\pi^0$ versus $D^0 \to K^-\pi^+$. We list $2y_{CP} = 2y$ $(= \Delta\Gamma/\Gamma)$ in the absence of CP violation.

[4] Based on 3 fb^{-1} of data collected at $\sqrt{s} = 7$, 8 TeV. Measures the lifetime difference between $D^0 \to K^-K^+$ and $D^0 \to \pi^-\pi^+$ (CP even) decays and $D^0 \to K^-\pi^+$ (CP mixed) decays, or $y_{CP} = (\Gamma_{CP+} - \Gamma_{CP-}) / (\Gamma_{CP+} + \Gamma_{CP-})$. The D^0 mesons are required to originate from semimuonic decays of B mesons. We list $2y_{CP} = \Delta\Gamma/\Gamma$.

[5] AAIJ 19X D^0 come from D^{*+} and $\overline{B} \to D^0 \mu^- X$ decays (and c.c.) in pp collisions at 7 and 8 TeV. Measurement allows for CP violation (none seen).

[6] The result was established with D^0 from prompt and secondary D^*. Based on 5 fb^{-1} of data collected at $\sqrt{s} = 7$, 8, 13 TeV. Assumes no CP violation. Reported $x'^2 = (3.9 \pm 2.7) \times 10^{-5}$ and $y' = (5.28 \pm 0.52) \times 10^{-3}$, where $x' = x \cos(\delta) + y \sin(\delta)$, $y' = y \cos(\delta) - x \sin(\delta)$ and δ is the strong phase between the $D^0 \to K^+\pi^-$ and $\overline{D}^0 \to K^+\pi^-$.

[7] Model-independent measurement of the charm mixing parameters in the decay $D^0 \to K^0_S \pi^+\pi^-$ using 1.0 fb^{-1} of LHCb data at $\sqrt{s} = 7$ TeV.

[8] Time-dependent amplitude analysis of $D^0 \to \pi^+\pi^-\pi^0$.

[9] An improved measurement of $\overline{D}^0 - D^0$ mixing and a search for CP violation in D^0 decays to CP-even final states K^+K^- and $\pi^+\pi^-$ using the final Belle data sample of 976 fb^{-1}.

[10] ABLIKIM 15D uses quantum correlations in $e^+e^- \to D^0\overline{D}^0$ at the $\psi(3770)$.

[11] Based on 976 fb^{-1} of data collected at $Y(nS)$ resonances. Assumes no CP violation. Reported $x'^2 = (0.09 \pm 0.22) \times 10^{-3}$ and $y' = (4.6 \pm 3.4) \times 10^{-3}$, where $x' = x \cos(\delta) + y \sin(\delta)$, $y' = y \cos(\delta) - x \sin(\delta)$ and δ is the strong phase between $D^0 \to K^+\pi^-$ and $\overline{D}^0 \to K^+\pi^-$.

[12] The time-dependent Dalitz-plot analysis of $D^0 \to K^0_S \pi^+\pi^-$ is emplored. Decay-time information and interference on the Dalitz plot are used to distinguish doubly Cabibbo-suppressed decays from mixing and to measure the relative phase between $D^0 \to K^{*+}\pi^-$ and $\overline{D}^0 \to K^{*+}\pi^-$. This value allows CP violation and is sensitive to the sign of Δm.

[13] Based on 9.6 fb^{-1} of data collected at the Tevatron. Assumes no CP violation. Reported $x'^2 = (0.08 \pm 0.18) \times 10^{-3}$ and $y' = (4.3 \pm 4.3) \times 10^{-3}$, where $x' = x \cos(\delta) + y \sin(\delta)$, $y' = y \cos(\delta) - x \sin(\delta)$ and δ is the strong phase between the $D^0 \to K^+\pi^-$ and $\overline{D}^0 \to K^+\pi^-$.

[14] Obtained $y_{CP} = (0.72 \pm 0.18 \pm 0.12)$% based on three effective D^0 lifetimes measured in $K^\mp\pi^\pm$, K^-K^+, and $\pi^-\pi^+$. We list $2y_{CP} = \Delta\Gamma/\Gamma$.

[15] Compared the lifetimes of D^0 decay to the CP eigenstate K^+K^- with D^0 decay to π^+K^-. The values here assume no CP violation.

[16] DEL-AMO-SANCHEZ 10D uses $540{,}800 \pm 800$ $K^0_S \pi^+\pi^-$ and $79{,}900 \pm 300$ $K^0_S K^+K^-$ events in a time-dependent amplitude analyses of the D^0 and $\overline{D}^0$ Dalitz plots. No evidence was found for CP violation, and the values here assume no such violation.

[17] ZUPANC 09 uses a method based on measuring the mean decay time of $D^0 \to K^0_S K^+K^-$ events for different K^+K^- mass intervals.

The Special Nature of Subnuclear Physics:
4 – Who needs peer review?

Internal digestion

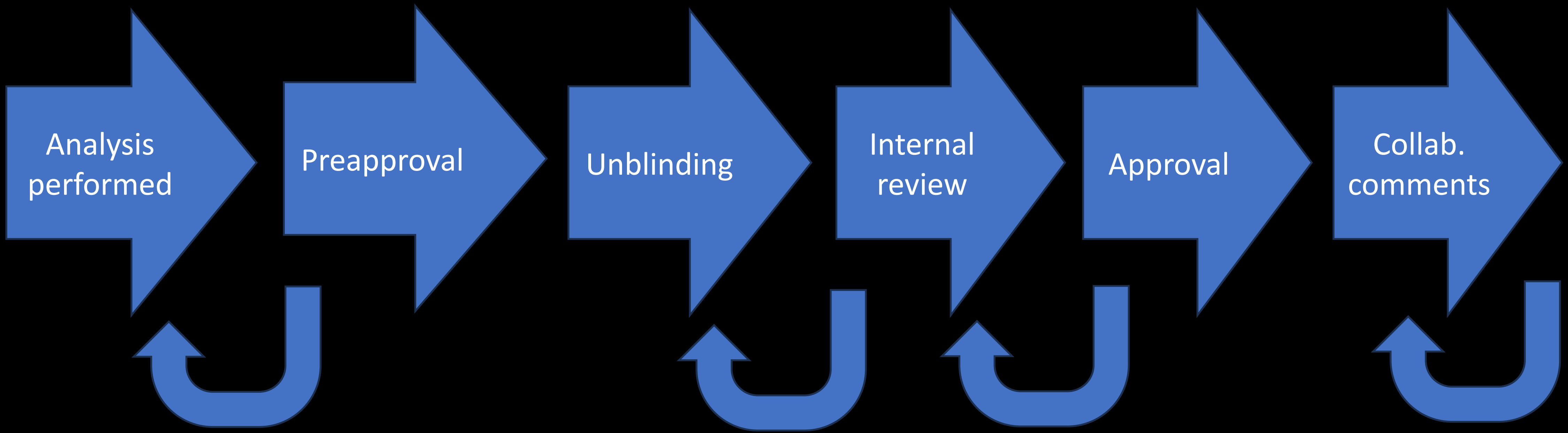
Your glory, my reputation

A few stories from the CDF experiment

Internal review: ARC, Godparents, and lobbies

To become public, results must go through an internal approval procedure

This may last for months, or years in special cases



Internal review: ARC, Godparents, and lobbies

To become public, results must go through an internal approval procedure

This may last for months, or years in special cases

Every experiment has some key measurements every want to contribute to

How to be recognized as a main contributor?

Internal review: ARC, Godparents, and lobbies

Large collaborations involve physicists from many countries, with different needs to feed their career goals

- Ego also an important component

Internal review: ARC, Godparents, and lobbies

Large collaborations involve physicists from many countries, with different needs to feed their career goals

- Ego also an important component

- Fear of publishing incorrect results or wrong claims reigns supreme

- “Be conservative” often uncertainties are overblown for no reason

The CDF experiment at the Tevatron

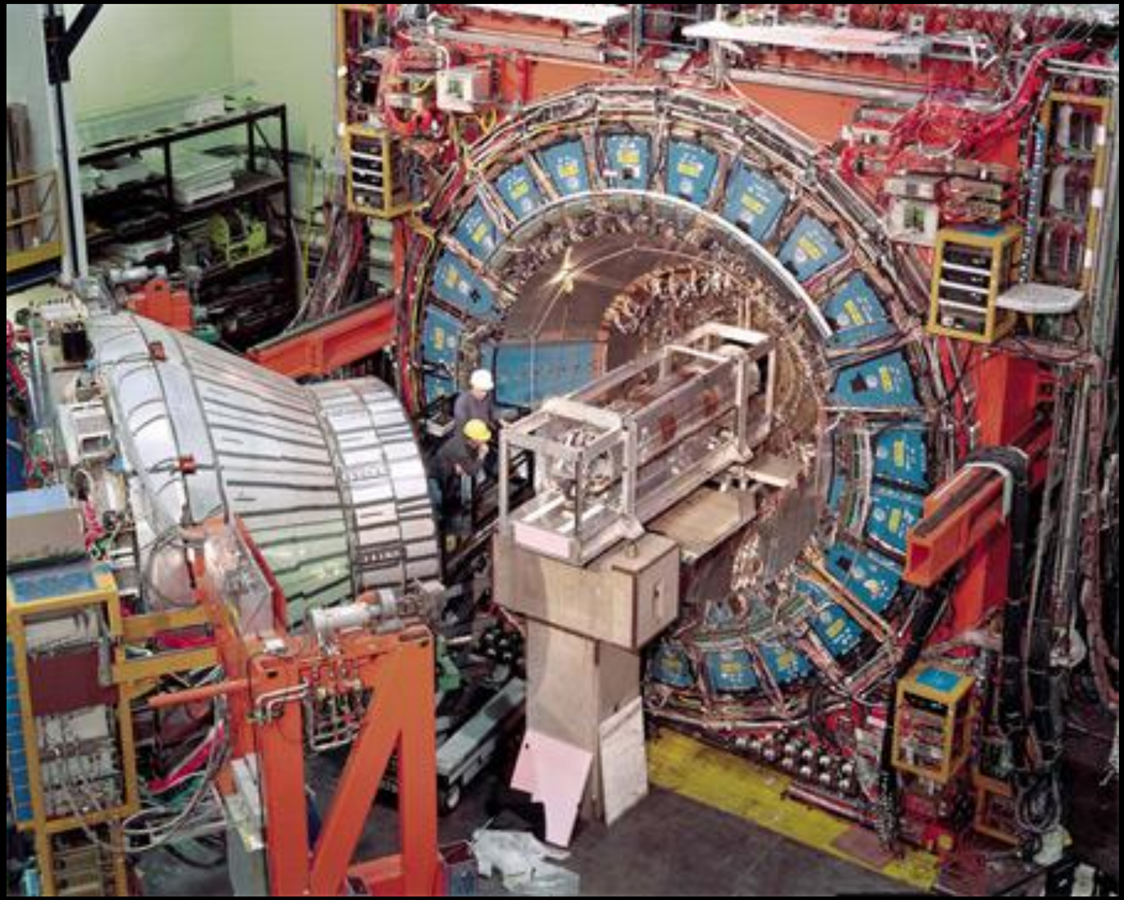
Built in 1980-85, operated from 1986 to 2012

Exploiting the Tevatron's high collision energy
to produce and observe the last quark, top

Produced hundreds of world-class results

600-strong collaboration,
mainly americans / Italians / Japanese at first

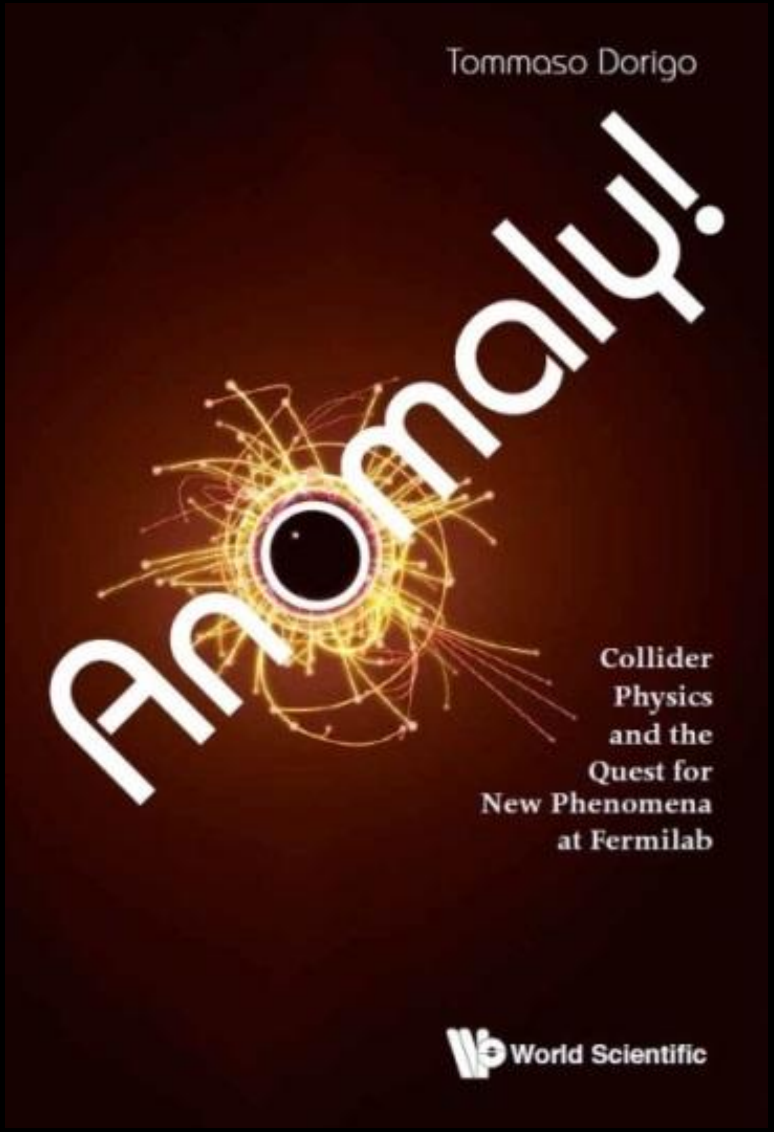
Complex sociology!



A book that tells those stories

The internal sociology of CDF, and the clashes of different experimental groups in search for the top quark or for new discoveries, is quite rich I wrote a book about it

Available in pdf for free, just email me



… But Uncertainty Bars Still Undercover!

- The complexity of physics analyses in particle physics trumps the cautious approach of physicists to interval estimation

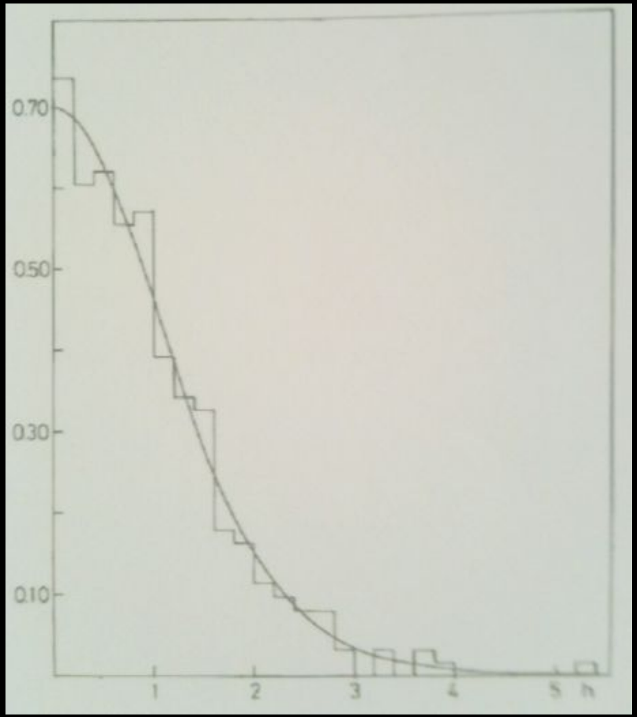
- A couple of results are worth mentioning

A study of residuals in HEP

A study of the measurement of particle properties in 1975 revealed that residuals were **not Gaussian**. Matts Roos *et al.* considered the difference between true and measured values of kaon and hyperon mean life and mass measurements, and concluded that these seemed to all have a similar shape, well described by a Student distribution $S_{10}(h/1.11)$:

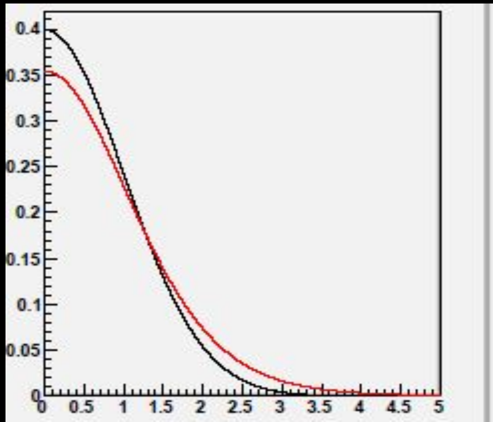
$$S_{10}\left(\frac{x}{1.11}\right)=\frac{315}{256\sqrt{10}}\left(1+\frac{x^2}{12.1}\right)^{-5.5}$$

While one cannot extrapolate to 5-sigma the behaviour observed by Roos in the bulk of the distribution, one may consider this as evidence that the uncertainties evaluated in experimental HEP may have a significant non-Gaussian component

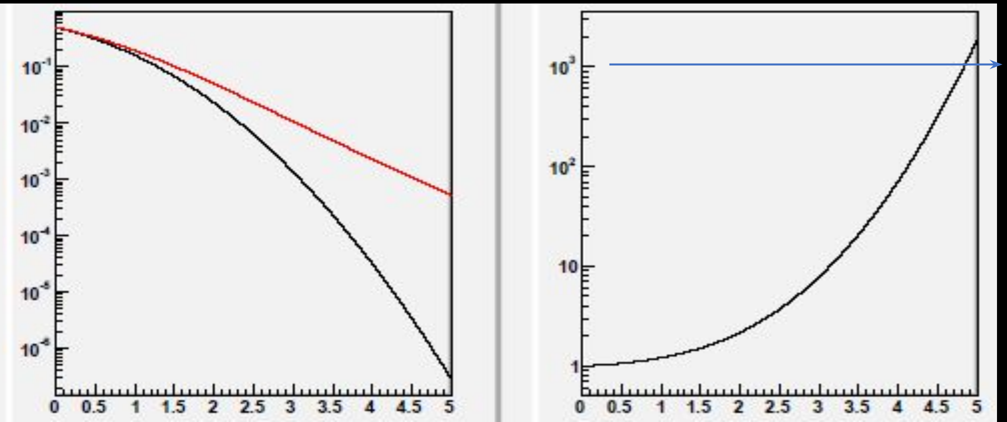


The distribution of residuals of 306 measurements

Black: a unit Gaussian;
red: the $S_{10}(x/1.11)$ function



Left: 1-integral distributions of the two functions.
Right: ratio of the 1-integral values as a function of z



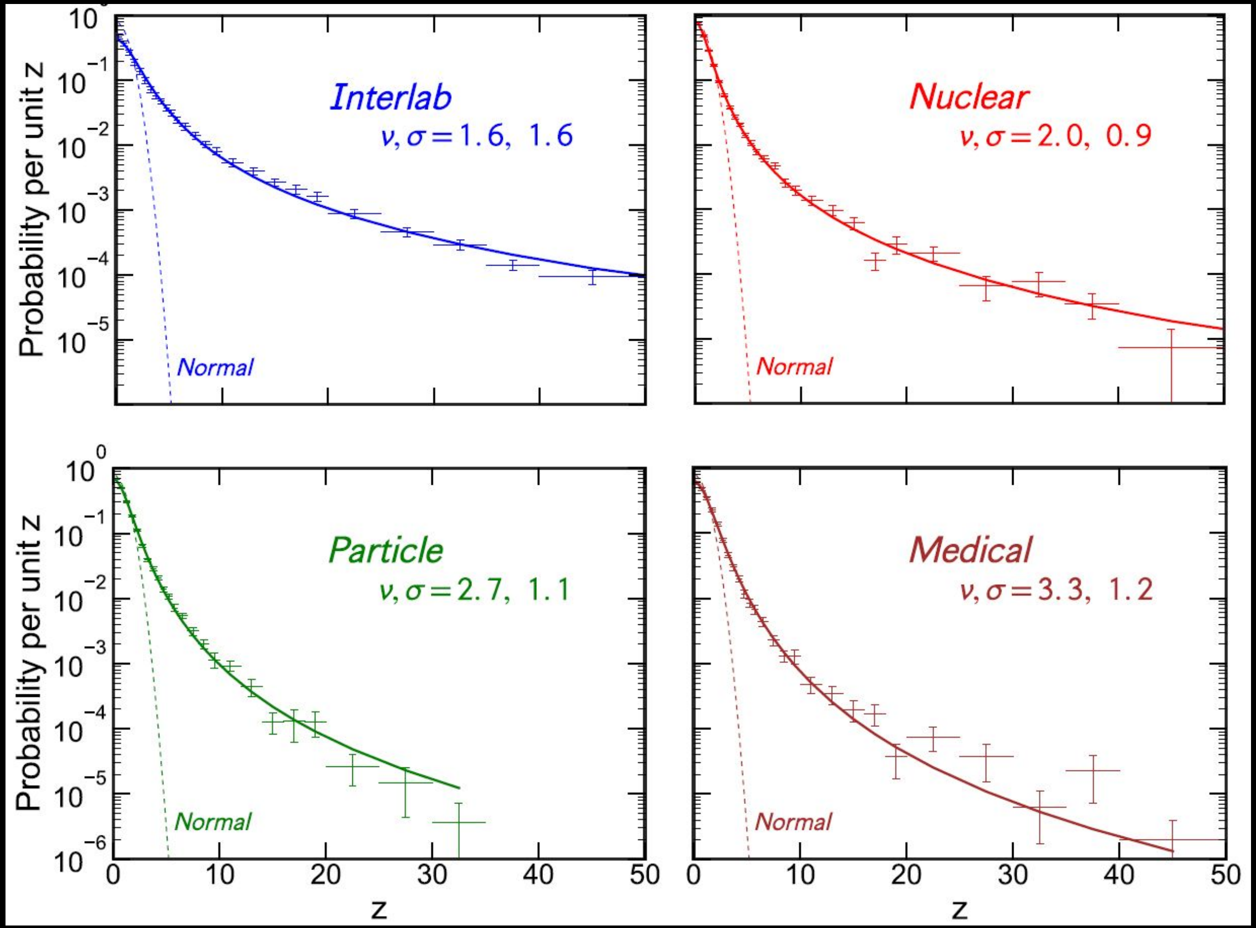
x1000!

A Bigger, Meaner Study of Residuals

David Bailey (U. Toronto) recently published a study where use of large datasets is made (all of RPP, Cochrane medical and health database, Table of Radionuclides)

- 41,000 measurements of 3200 quantities studied

- The methodology is similar to that of Roos et al., but some shortcuts are made, and data input automation prevents more vetting (e.g. correlations not properly accounted for)



Results are quite striking - we seem to have ubiquitous Student-t distributions in our Z values, with large tails – almost Cauchy-like.

One LHC analysis that was reproduced

I know of only one LHC analysis whose full reproduction was ever attempted, out of 2000+ analyses published by ATLAS and CMS in the past 13 years….

By me and

One LHC analysis that was reproduced

I know of only one LHC analysis whose full reproduction was ever attempted, out of 2000+ analyses published by ATLAS and CMS in the past 13 years….

By me and a PhD student of mine!

One LHC analysis that was reproduced

In 2019 I published with P. de Castro a method to make a NN classifier aware of systematic uncertainties affecting the modeling of input data, realigning the classification step with the final measurement goals

ELSEVIER

Computer Physics Communications
Volume 244, November 2019, Pages 170-179

INFERNO: Inference-Aware Neural Optimisation

Pablo de Castro, Tommaso Dorigo

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.cpc.2019.06.007

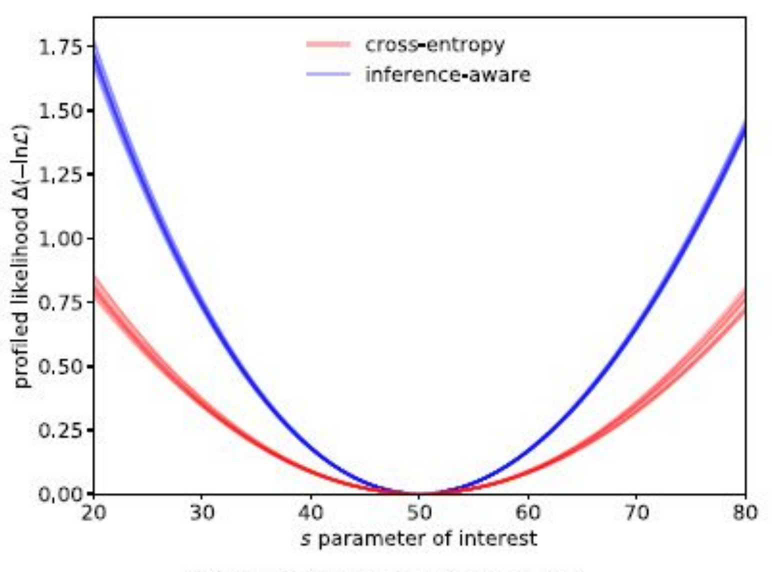
Abstract

Complex computer simulations are commonly required for accurate data modelling in many scientific disciplines, making statistical inference challenging due to the intractability of the likelihood evaluation for the observed data. Furthermore, sometimes one is interested on inference drawn over a subset of the generative model parameters while taking into account model uncertainty or misspecification on the remaining nuisance parameters. In this work, we show how non-linear summary statistics can be constructed by minimising inference-motivated losses via stochastic gradient descent such that they provide the smallest uncertainty for the parameters of interest. As a use case, the problem of confidence interval estimation for the mixture coefficient in a multi-dimensional two-component mixture model (i.e. signal vs background) is considered, where the proposed technique clearly outperforms summary statistics based on probabilistic classification, a commonly used alternative which does not account for the presence of nuisance parameters.

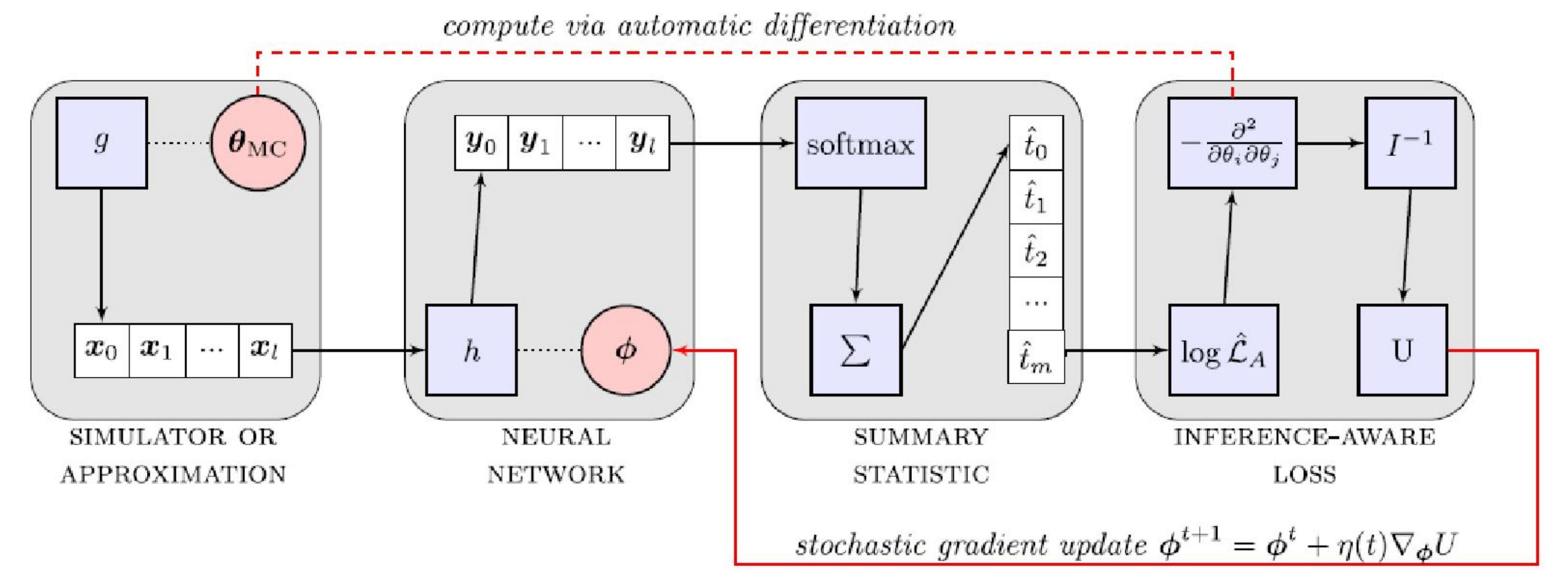
One LHC analysis that was reproduced

The study used a synthetic dataset to prove the power of the method.

Meaningful application to real data entailed serious modeling work, and finding one use case where systematic uncertainties were dominating in a NN-based dimensionality reduction analysis



(b) profile-likelihood comparison

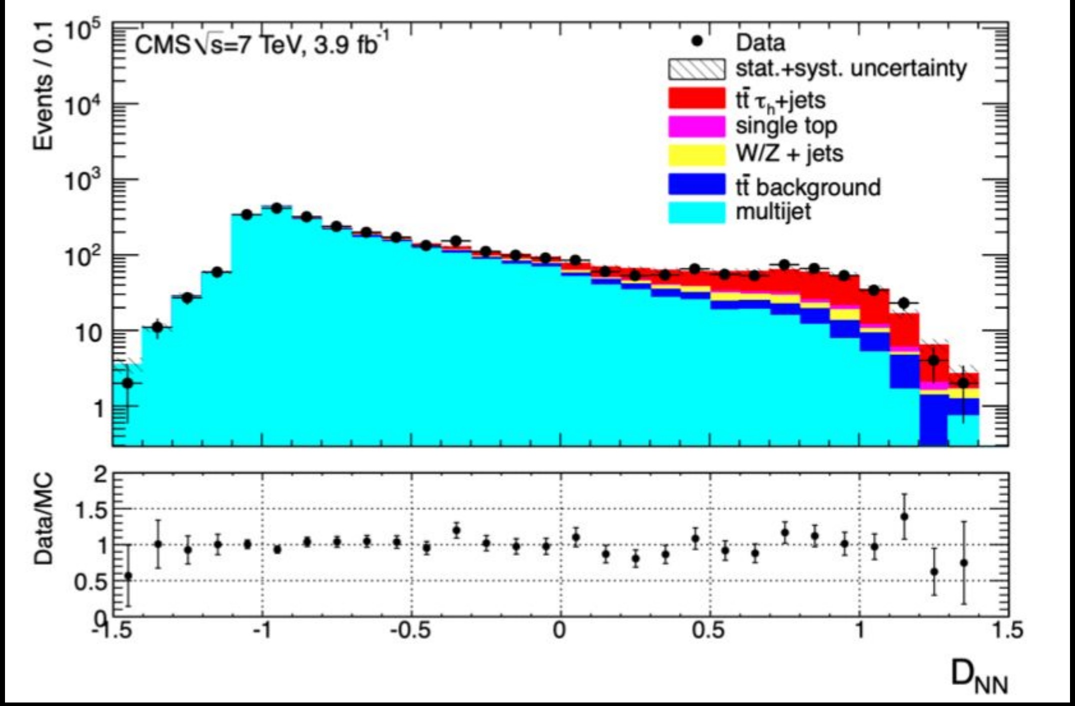


Replication of tt → $T_{h\text{ + jets (TOP-11-004)}}$ with CMS Open Data

The possibility to replicate a full-blown LHC analysis is due to the publishing of 2011-2012 data by the ATLAS and CMS collaborations. However, analysis of those data is still experts-only, and even experts take many months to get things lined up properly!

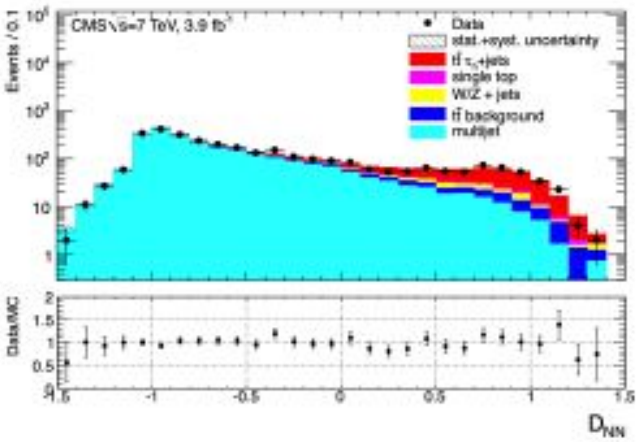
Original result for the top quark pair cross section:

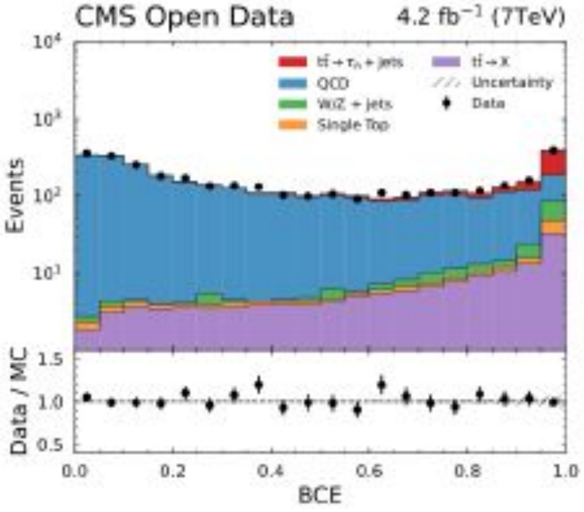
σ(tt+X)=156 ± 12 (stat.) ± 33 (sys.) ± 3 (lumi) pb



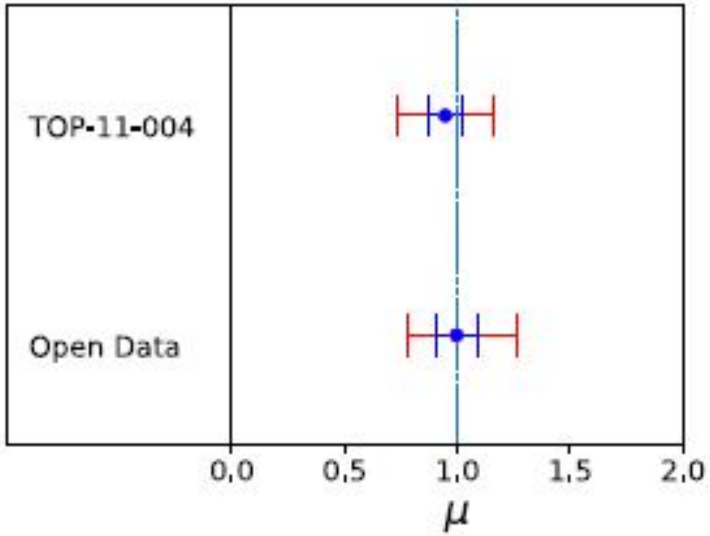
Not so easy to reproduce a result!

https://arxiv.org/abs/1301.5755





Source	Original	CMS Open Data
Signal $t\bar{t} \to \tau_h +$ jets	383	348
Background $t\bar{t} \to X$	151	134
WZ+Jets	83	64
Single-Top	41	30
QCD	2392	2690
Total backgrounds	2667	2918
Data	3050	3323



- Small differences in selected number of events → different CMSSW versions / different b-tagging / different JES/JER corrections / different selections
- Signal strengths comparable
- Different systematics due to different recipes

In summary

Particle physics is a very special science: replicability here must be taken in a wide sense

- experiments have a scale (in time, size, $) that makes replicas impractical, even when it does not involve blowing up an atomic bomb
- the scale of collaborations, the depth of internal review, and the type-I error rate chosen in the field make replicability less important than elsewhere

Repeating a measurement, even with the same data, is a challenge!!

Thank you for your time