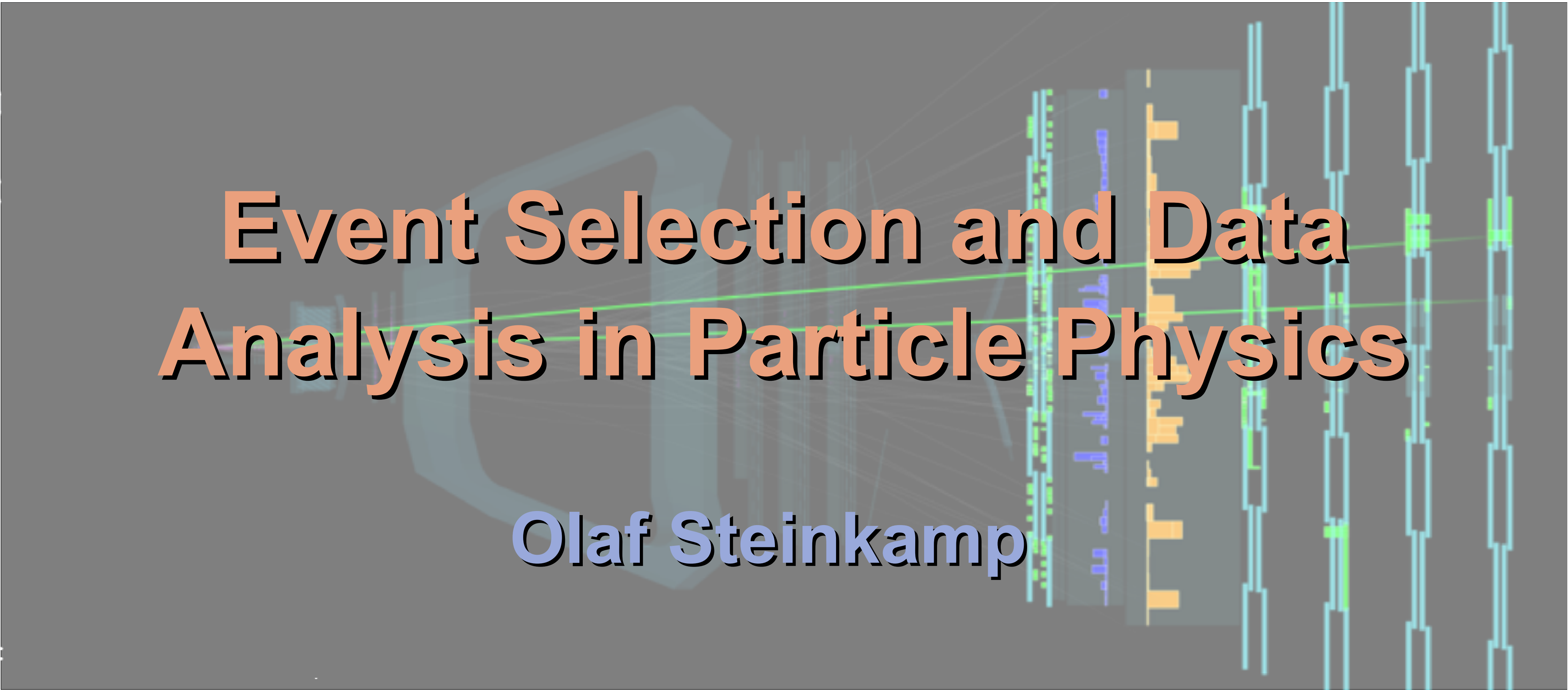
Universität Zürich^UZH

Young Physicists' Forum

UZH, April 22, 2017



Event Selection and Data Analysis in Particle Physics

Olaf Steinkamp

Physik-Institut der Universität Zürich
Winterthurerstrasse 190 CH-8057 Zürich
olafs@physik.uzh.ch

Overview

Part I

What are we trying to measure ?

What do our experiments look like ?

Part II

“Online” event selection

Part III

“Offline” analysis

Overview

Part I

What are we trying to measure ?

What do our experiments look like ?

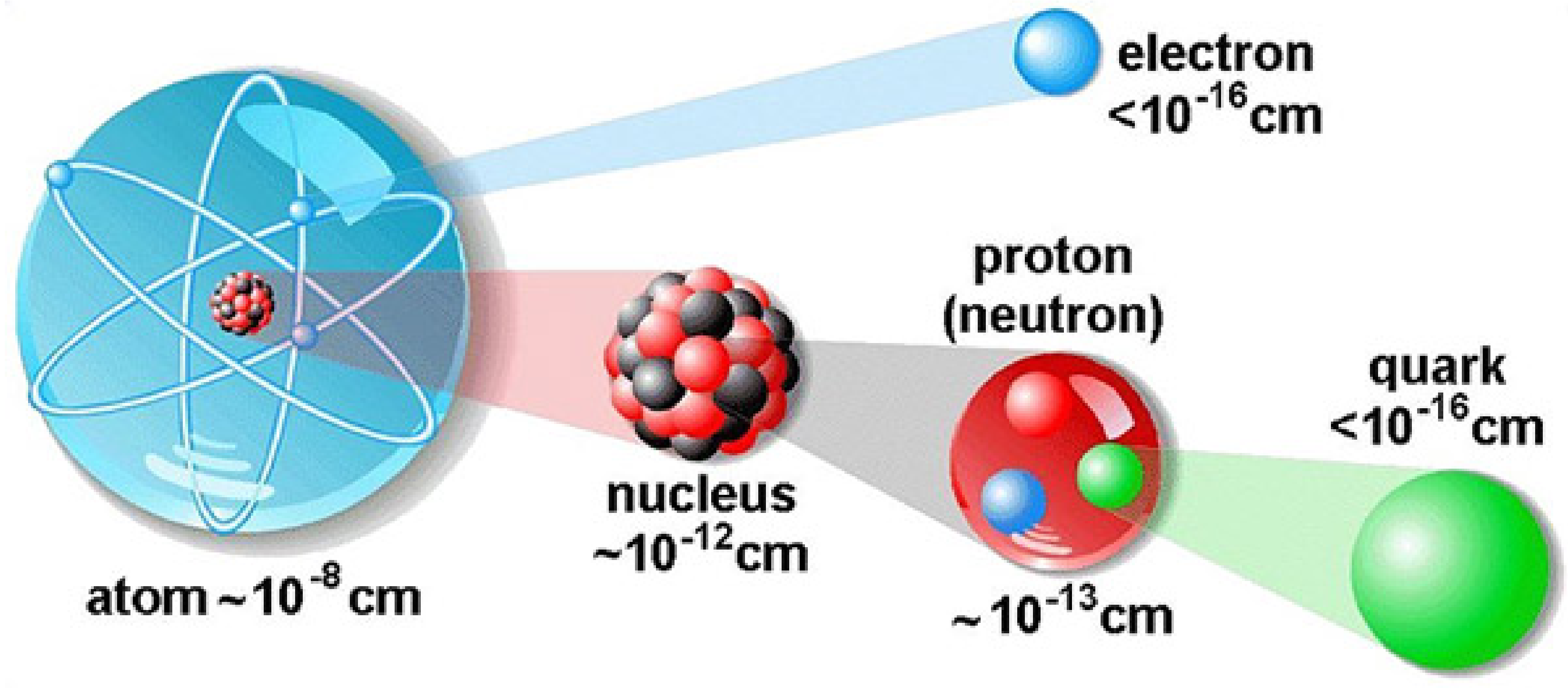
Part II

“Online” event selection

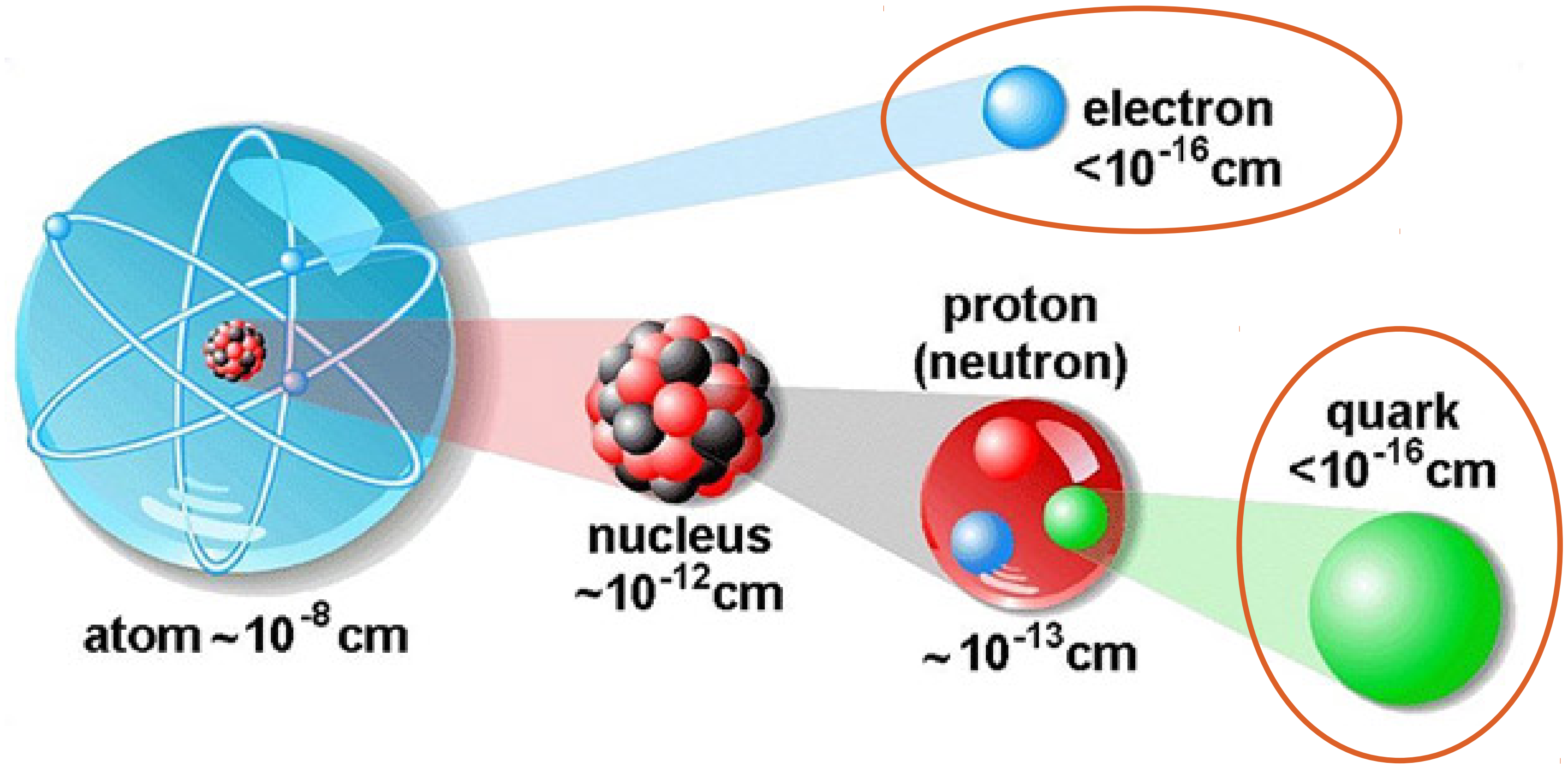
Part III

“Offline” analysis

Elementary Particle Physics



Elementary Particle Physics



Atomic physics,
Chemistry

Nuclear
physics

Particle
physics

Elementary Particle Physics

The "Standard Model" of Particle Physics describes elementary particles and their interactions through three of the four known fundamental forces

electromagnetic interaction
weak interaction
strong interaction

Formulated in the 1960's, it is a Quantum Field Theory based on

relativistic kinematics
quantum mechanics
group theory

Elementary Particle Physics

Describes many observed phenomena with stunning precision

e.g. electric dipole moment of the electron

$(g^{SM}- 2) / 2 = 0.00115965218164(76)$

$(g^{exp}- 2) / 2 = 0.00115965218073(28)$

Many of it's predictions have been confirmed by experiments

e.g. observation of Higgs Boson at CERN,
40 years after its existence was predicted
to accommodate mass of elementary particles

Elementary Particle Physics

Does not explain some fundamental observations, e.g.

Matter – Antimatter asymmetry in the Universe

Dark Matter

Does not incorporate gravity

Holy grail of particle physics today:

Search for phenomena "Beyond the Standard Model"

Elementary Particle Physics

Two approaches:

"Direct Searches"

New elementary particles

"Indirect Searches"

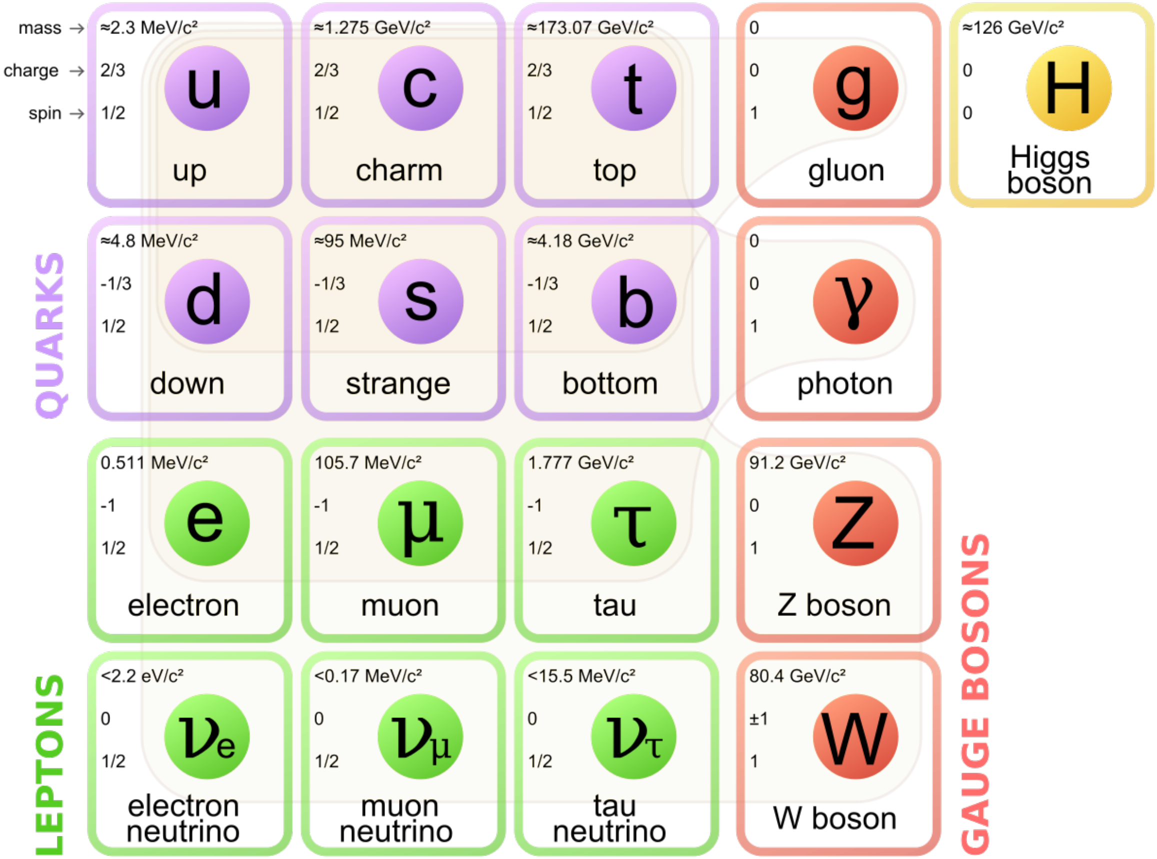
Confront precise predictions with precise measurements

In both cases, studying very rare phenomena:

Need effective and efficient selection algorithms
to separate signal from large backgrounds

Need to handle huge data volumes

Standard Model

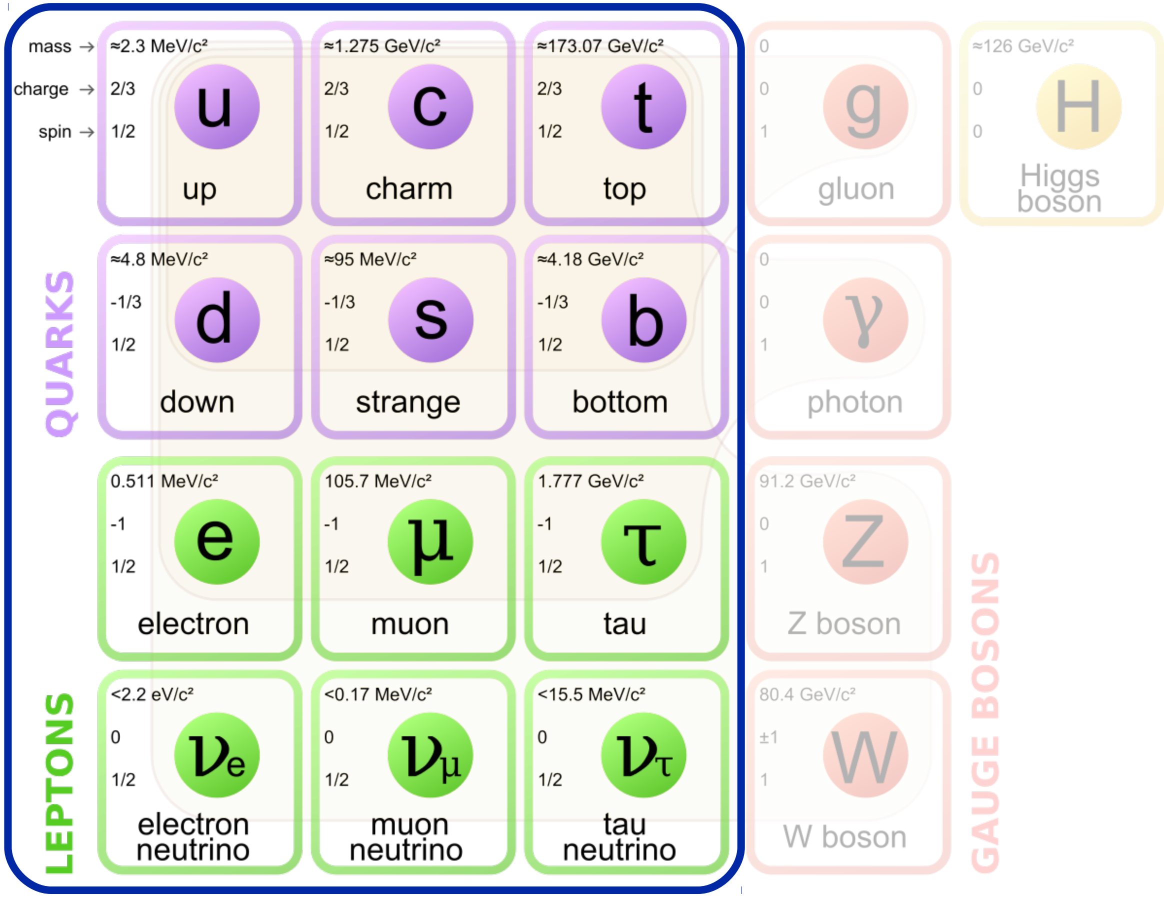


	mass → charge → spin →	≈2.3 MeV/c^2 2/3 1/2 u up	≈1.275 GeV/c^2 2/3 1/2 c charm	≈173.07 GeV/c^2 2/3 1/2 t top	0 0 1 g gluon	≈126 GeV/c^2 0 0 H Higgs boson
QUARKS		≈4.8 MeV/c^2 -1/3 1/2 d down	≈95 MeV/c^2 -1/3 1/2 s strange	≈4.18 GeV/c^2 -1/3 1/2 b bottom	0 0 1 γ photon	
		0.511 MeV/c^2 -1 1/2 e electron	105.7 MeV/c^2 -1 1/2 μ muon	1.777 GeV/c^2 -1 1/2 τ tau	91.2 GeV/c^2 0 1 Z Z boson	GAUGE BOSONS
LEPTONS		<2.2 eV/c^2 0 1/2 ν_e electron neutrino	<0.17 MeV/c^2 0 1/2 ν_μ muon neutrino	<15.5 MeV/c^2 0 1/2 ν_τ tau neutrino	80.4 GeV/c^2 ±1 1 W W boson	

source: <https://en.wikipedia.org/wiki/Standard_Model>

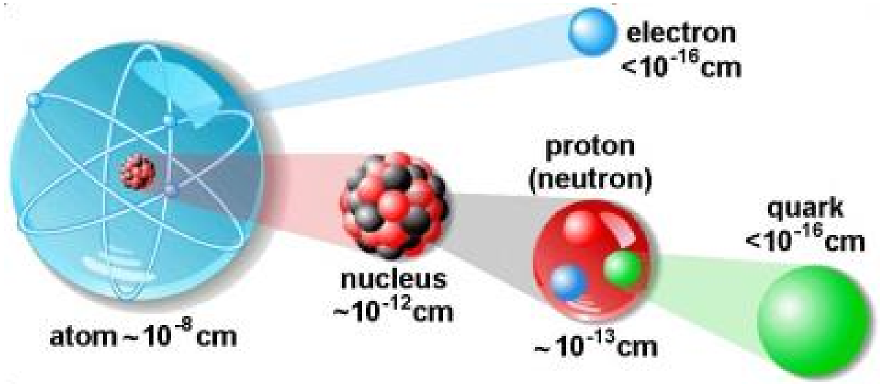
Elementary particles

Standard Model

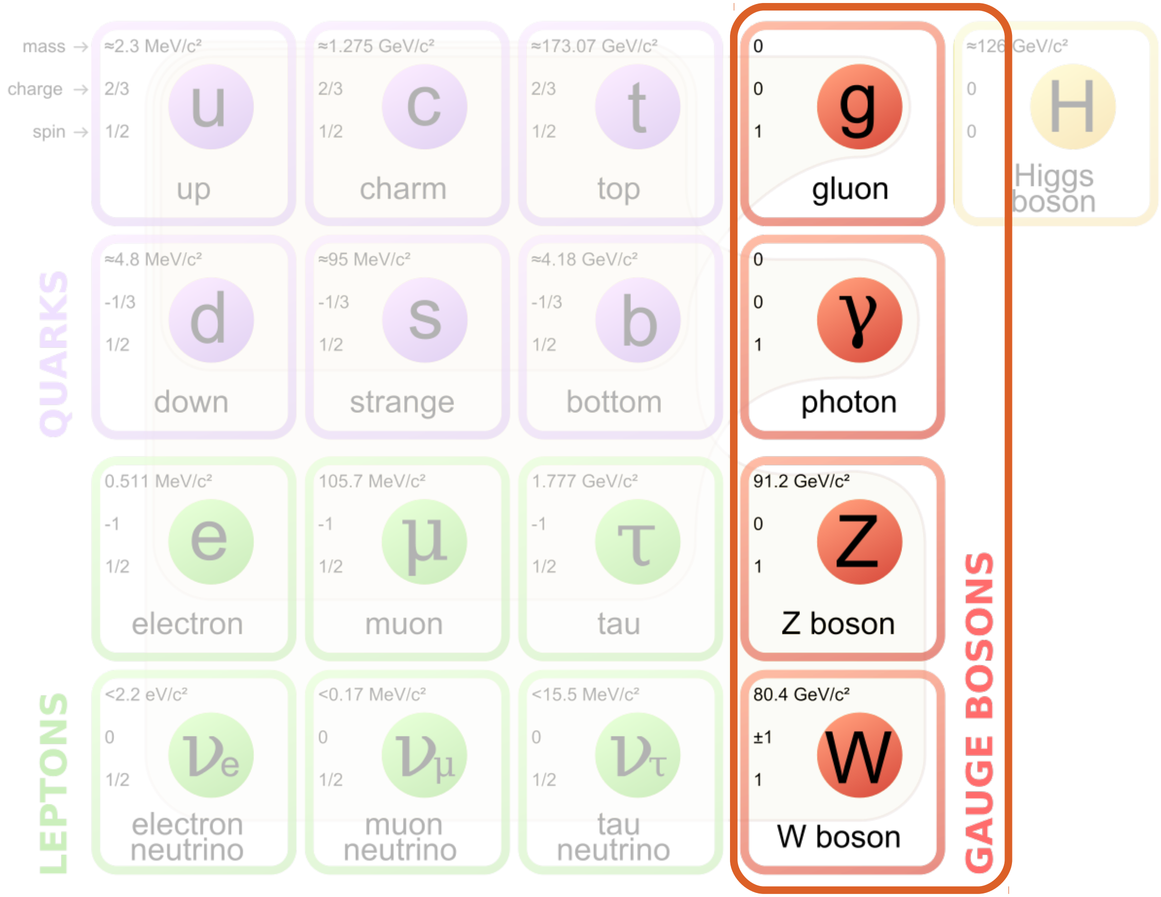


mass → charge → spin →	≈2.3 MeV/c^2 2/3 1/2 u up	≈1.275 GeV/c^2 2/3 1/2 c charm	≈173.07 GeV/c^2 2/3 1/2 t top	0 0 1 g gluon	≈126 GeV/c^2 0 0 H Higgs boson
QUARKS	≈4.8 MeV/c^2 -1/3 1/2 d down	≈95 MeV/c^2 -1/3 1/2 s strange	≈4.18 GeV/c^2 -1/3 1/2 b bottom	0 0 1 γ photon	
	0.511 MeV/c^2 -1 1/2 e electron	105.7 MeV/c^2 -1 1/2 μ muon	1.777 GeV/c^2 -1 1/2 τ tau	91.2 GeV/c^2 0 1 Z Z boson	GAUGE BOSONS
LEPTONS	<2.2 eV/c^2 0 1/2 ν_e electron neutrino	<0.17 MeV/c^2 0 1/2 ν_μ muon neutrino	<15.5 MeV/c^2 0 1/2 ν_τ tau neutrino	80.4 GeV/c^2 ±1 1 W W boson	

Quarks, Leptons → matter

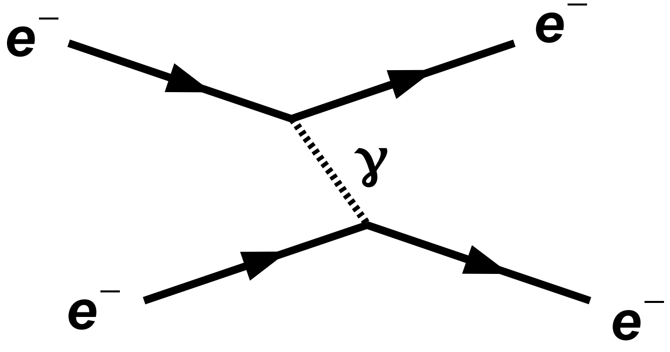


Standard Model

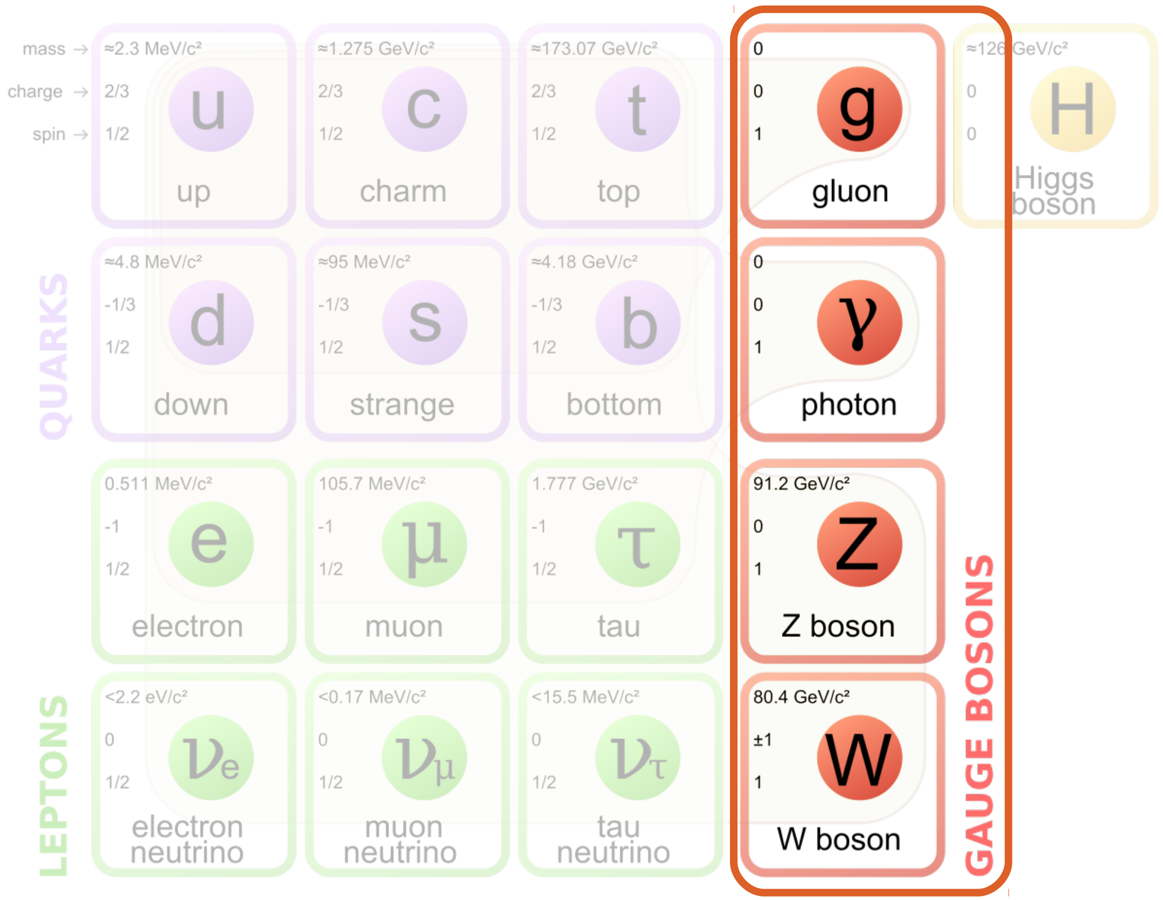


mass → charge → spin →	≈2.3 MeV/c^2 2/3 1/2 u up	≈1.275 GeV/c^2 2/3 1/2 c charm	≈173.07 GeV/c^2 2/3 1/2 t top	0 0 1 g gluon	≈126 GeV/c^2 0 0 H Higgs boson
QUARKS	≈4.8 MeV/c^2 -1/3 1/2 d down	≈95 MeV/c^2 -1/3 1/2 s strange	≈4.18 GeV/c^2 -1/3 1/2 b bottom	0 0 1 γ photon	
	0.511 MeV/c^2 -1 1/2 e electron	105.7 MeV/c^2 -1 1/2 μ muon	1.777 GeV/c^2 -1 1/2 τ tau	91.2 GeV/c^2 0 1 Z Z boson	GAUGE BOSONS
LEPTONS	<2.2 eV/c^2 0 1/2 ν_e electron neutrino	<0.17 MeV/c^2 0 1/2 ν_μ muon neutrino	<15.5 MeV/c^2 0 1/2 ν_τ tau neutrino	80.4 GeV/c^2 ±1 1 W W boson	

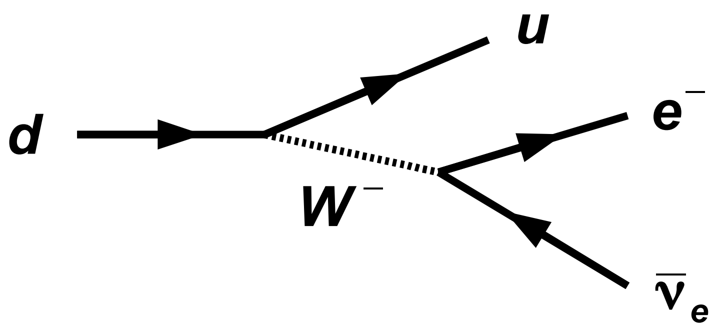
Gauge Bosons → interactions



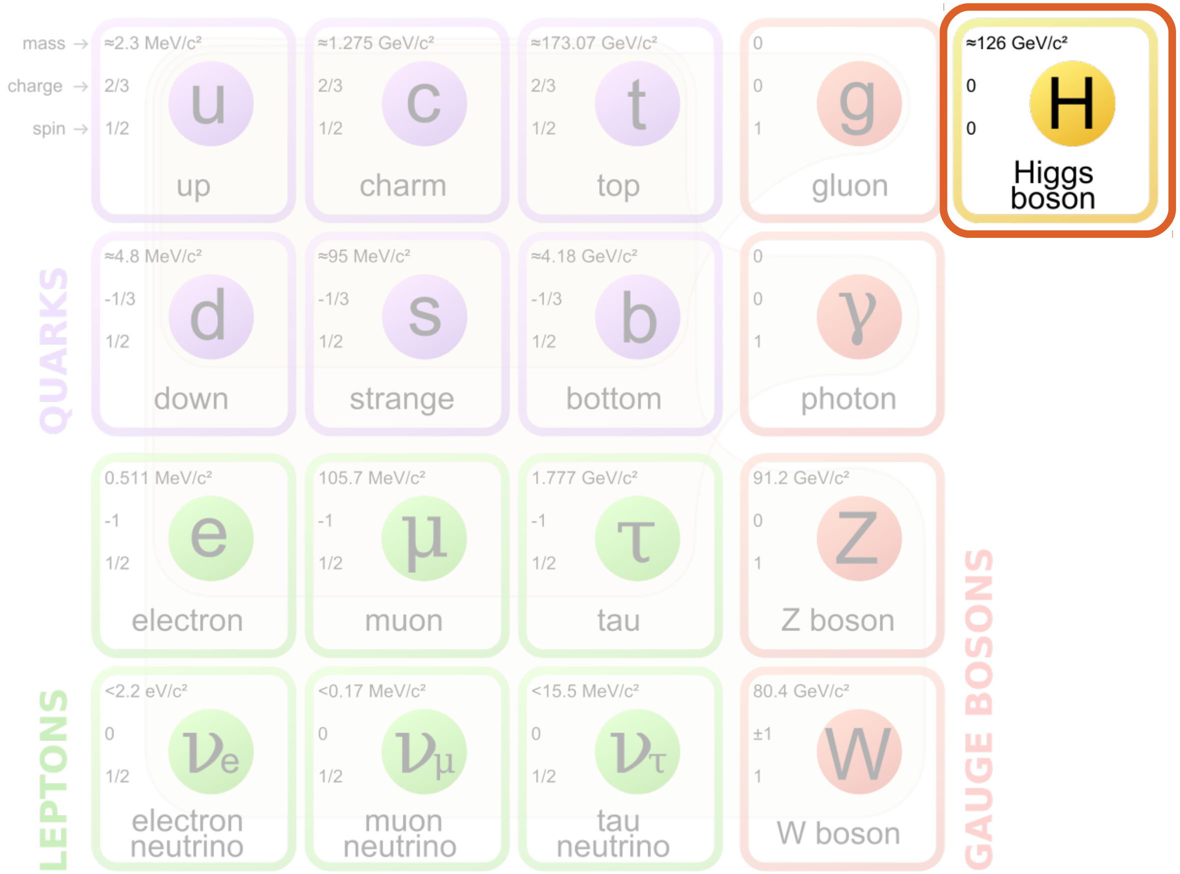
Standard Model



Gauge Bosons → interactions

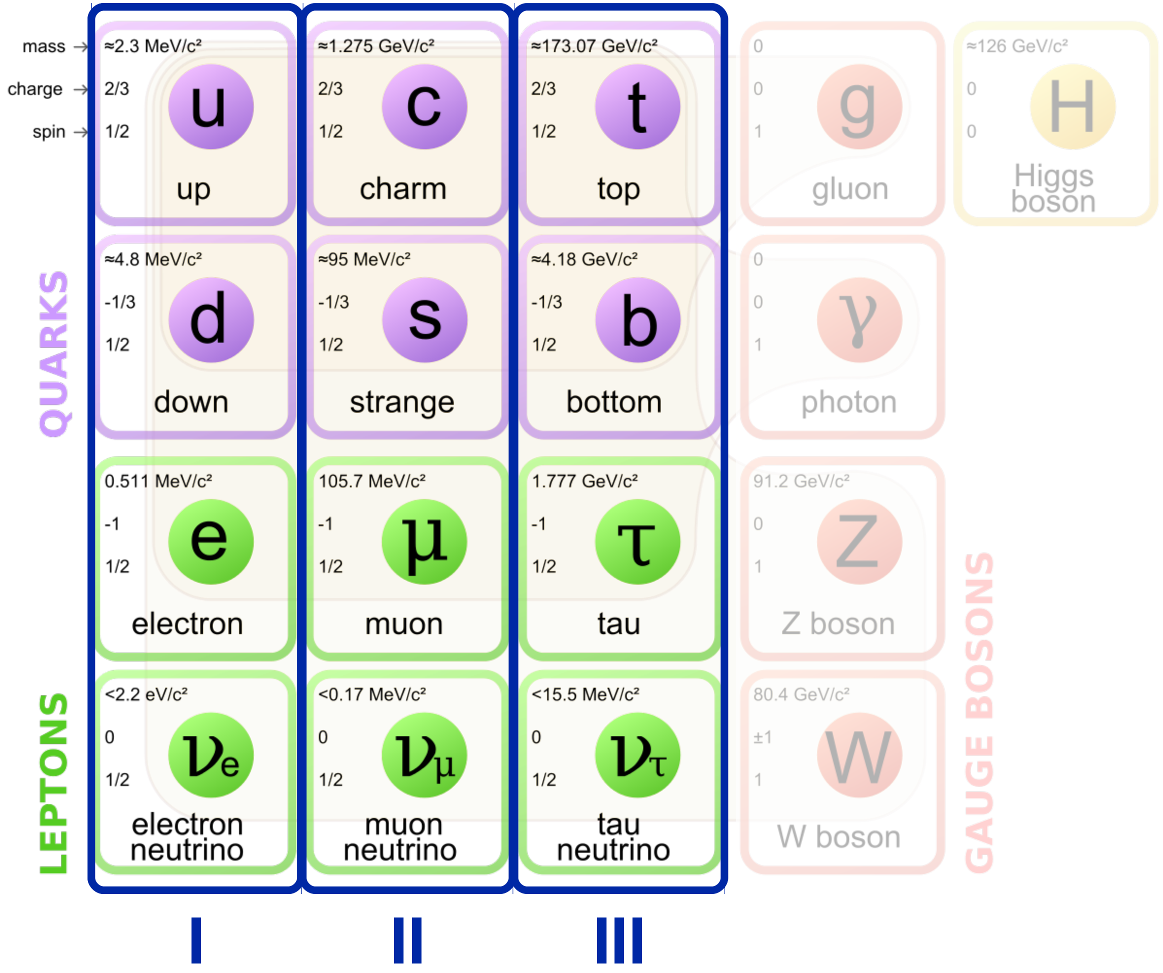


Standard Model



Higgs Boson → mass

Standard Model



Three “generations” of matter particles

2nd and 3rd generation are heavier siblings of 1st generation

Fundamental Questions

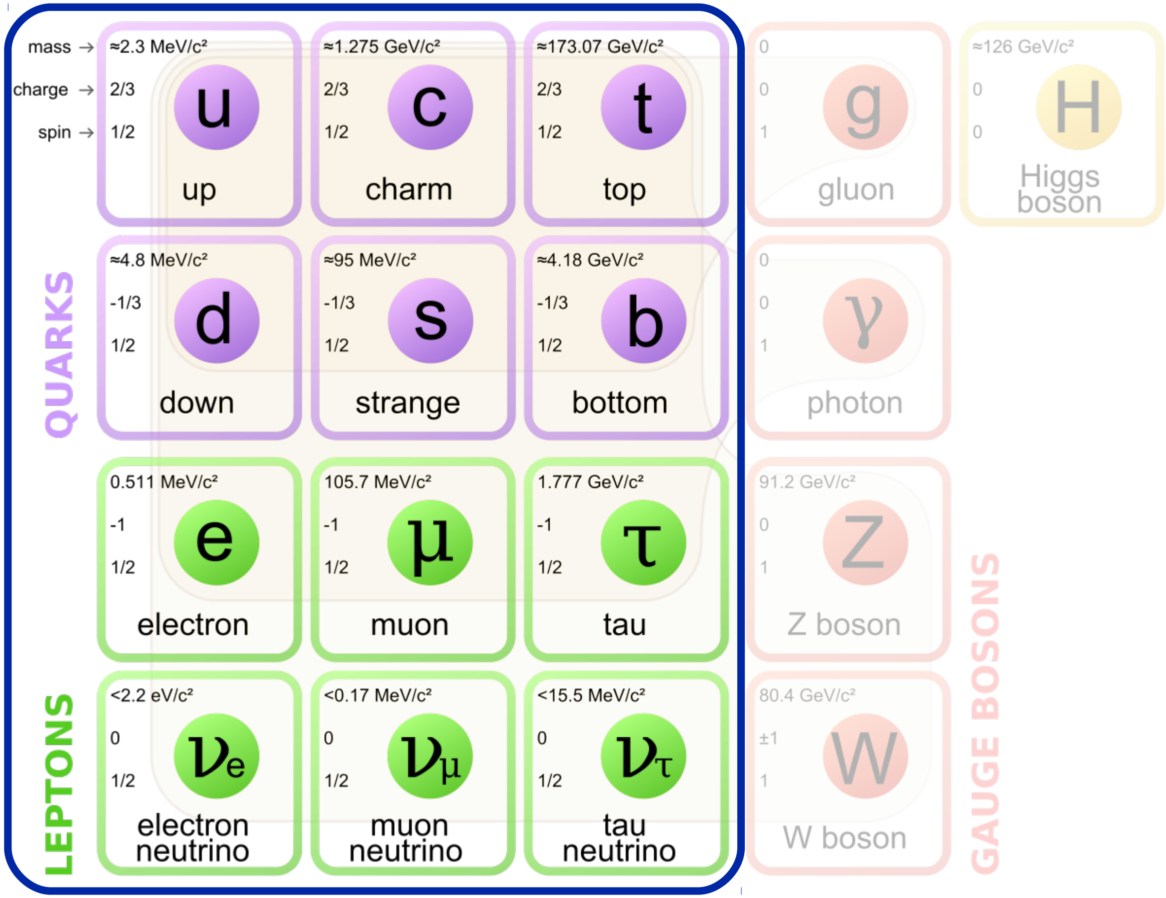
Who ordered THAT ?

(Isidor Isaac Rabi on the discovery of the muon)

photo source: <https://en.wikipedia.org/wiki/Isidor_Isaac_Rabi>

WHY THREE GENERATIONS ?

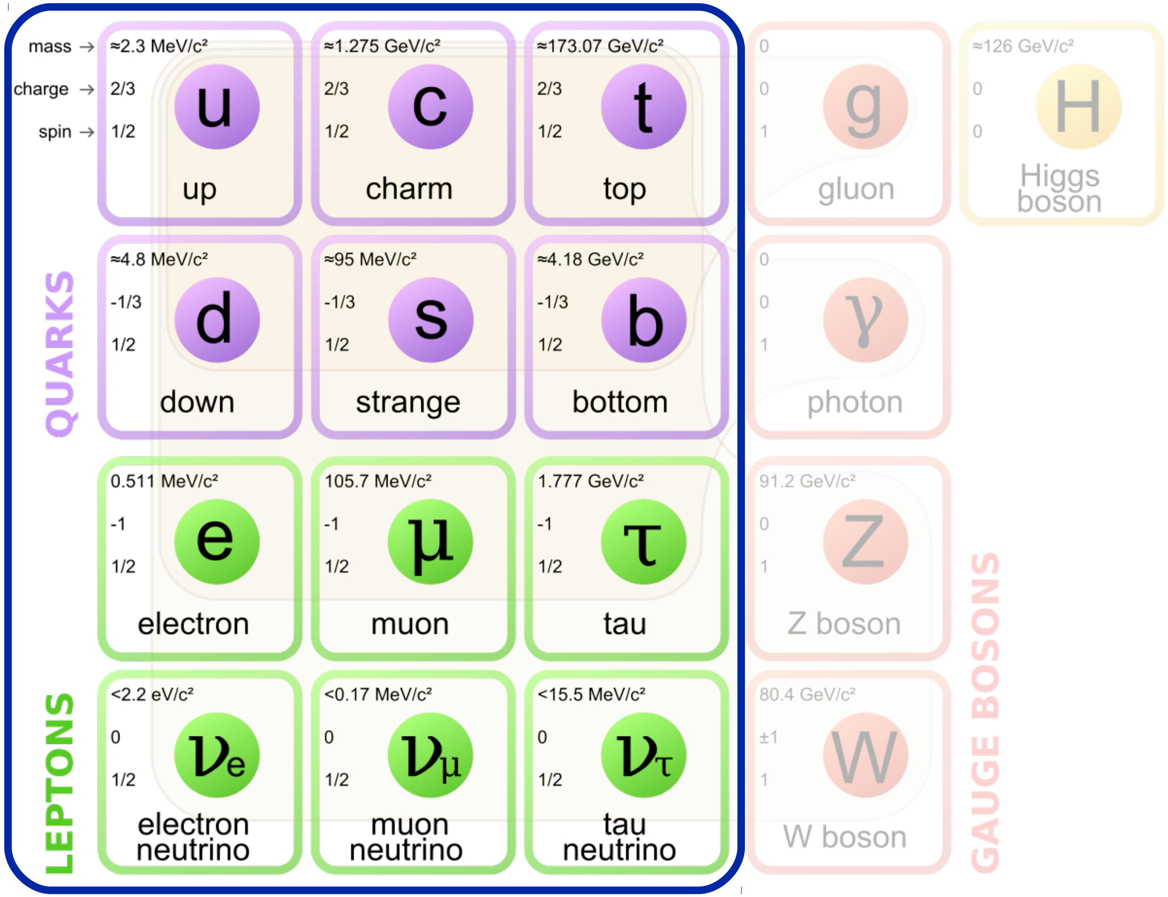
Standard Model



For each matter particle there is a corresponding antiparticle

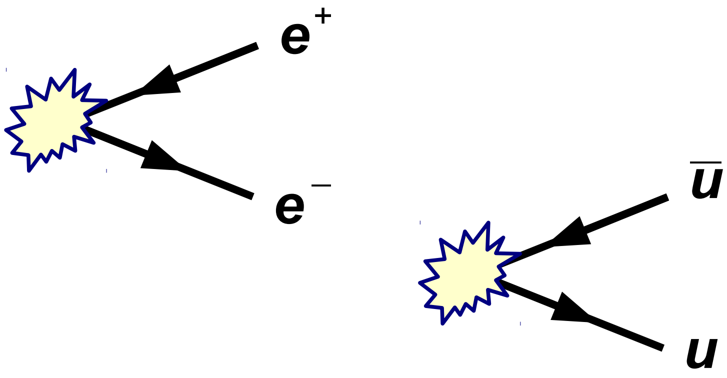
- **Same mass as the particle, but opposite charge**

Standard Model

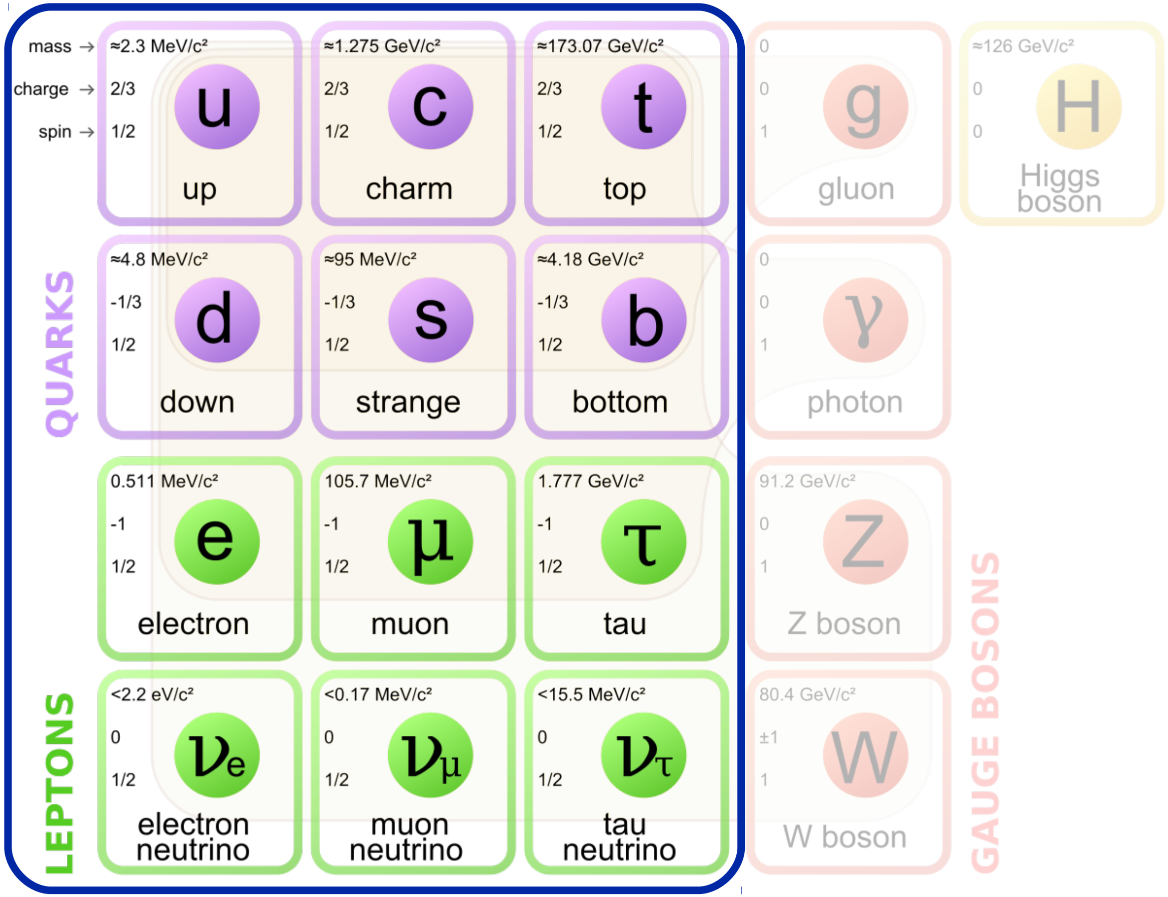


Particle – antiparticle pairs can be produced out of energy

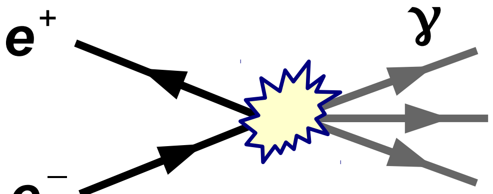
→ $E = m c^2$

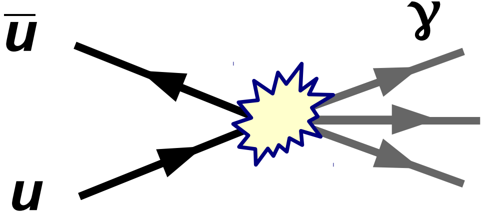


Standard Model



mass → charge → spin →	≈2.3 MeV/c^2 2/3 1/2 u up	≈1.275 GeV/c^2 2/3 1/2 c charm	≈173.07 GeV/c^2 2/3 1/2 t top	0 0 1 g gluon	≈126 GeV/c^2 0 0 H Higgs boson
QUARKS	≈4.8 MeV/c^2 -1/3 1/2 d down	≈95 MeV/c^2 -1/3 1/2 s strange	≈4.18 GeV/c^2 -1/3 1/2 b bottom	0 0 1 γ photon	
	0.511 MeV/c^2 -1 1/2 e electron	105.7 MeV/c^2 -1 1/2 μ muon	1.777 GeV/c^2 -1 1/2 τ tau	91.2 GeV/c^2 0 1 Z Z boson	GAUGE BOSONS
LEPTONS	<2.2 eV/c^2 0 1/2 ν_e electron neutrino	<0.17 MeV/c^2 0 1/2 ν_μ muon neutrino	<15.5 MeV/c^2 0 1/2 ν_τ tau neutrino	80.4 GeV/c^2 ±1 1 W W boson	





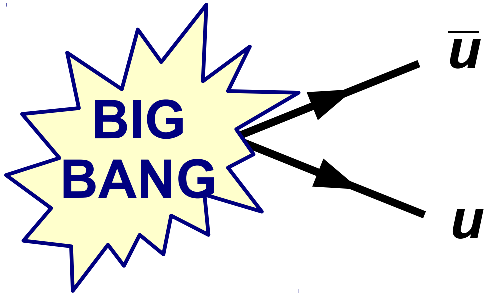
Particle – antiparticle pairs annihilate into energy (e.g. gamma rays)

→ $E = m c^2$

Fundamental Questions

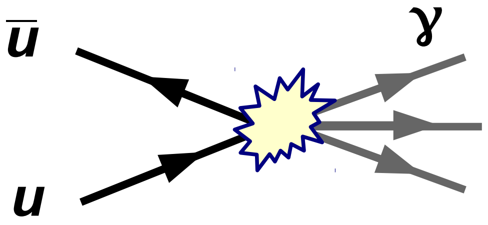
source: <https://www.nasa.gov/mission_pages/hubble/multimedia/index.html>

WHY DO WE SEE SO MUCH MORE MATTER THAN ANTIMATTER IN THE UNIVERSE ?

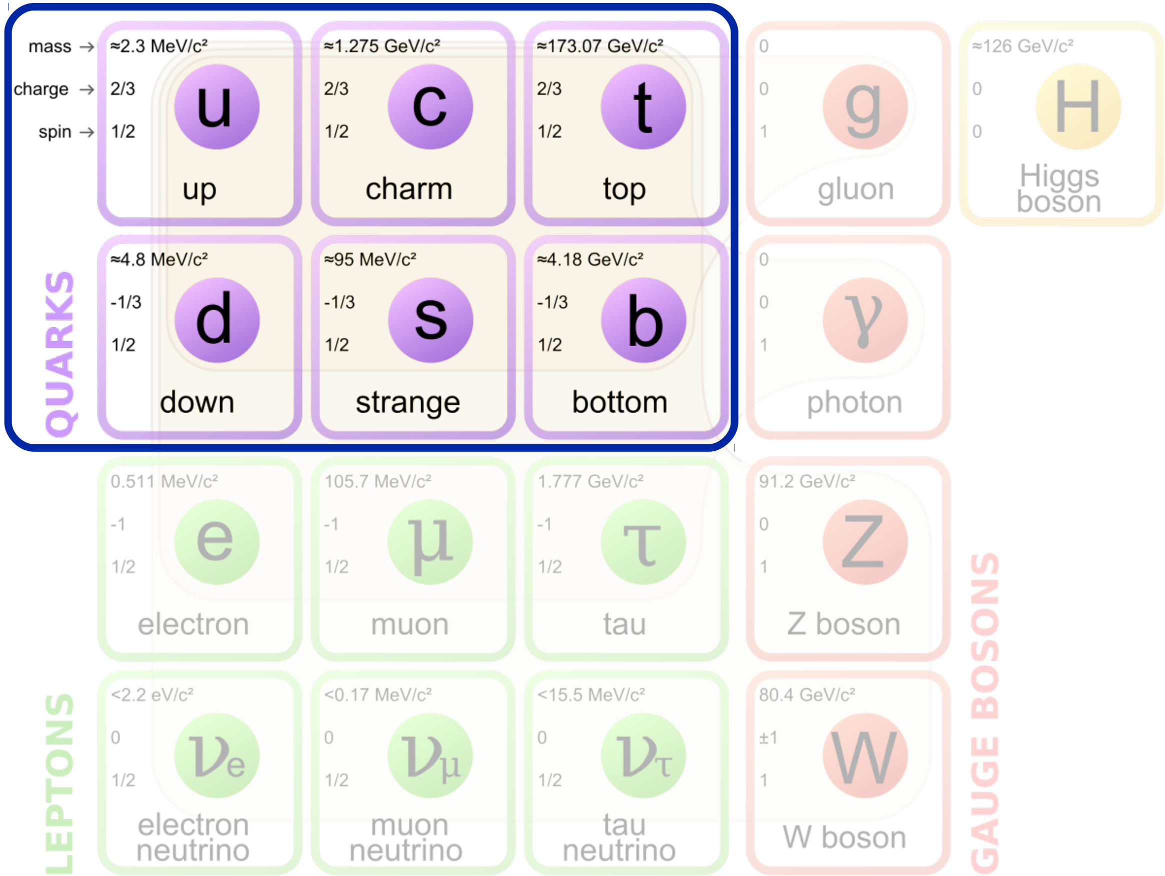


Fundamental Questions

WHY DOES ANYTHING EXIST AT ALL ?

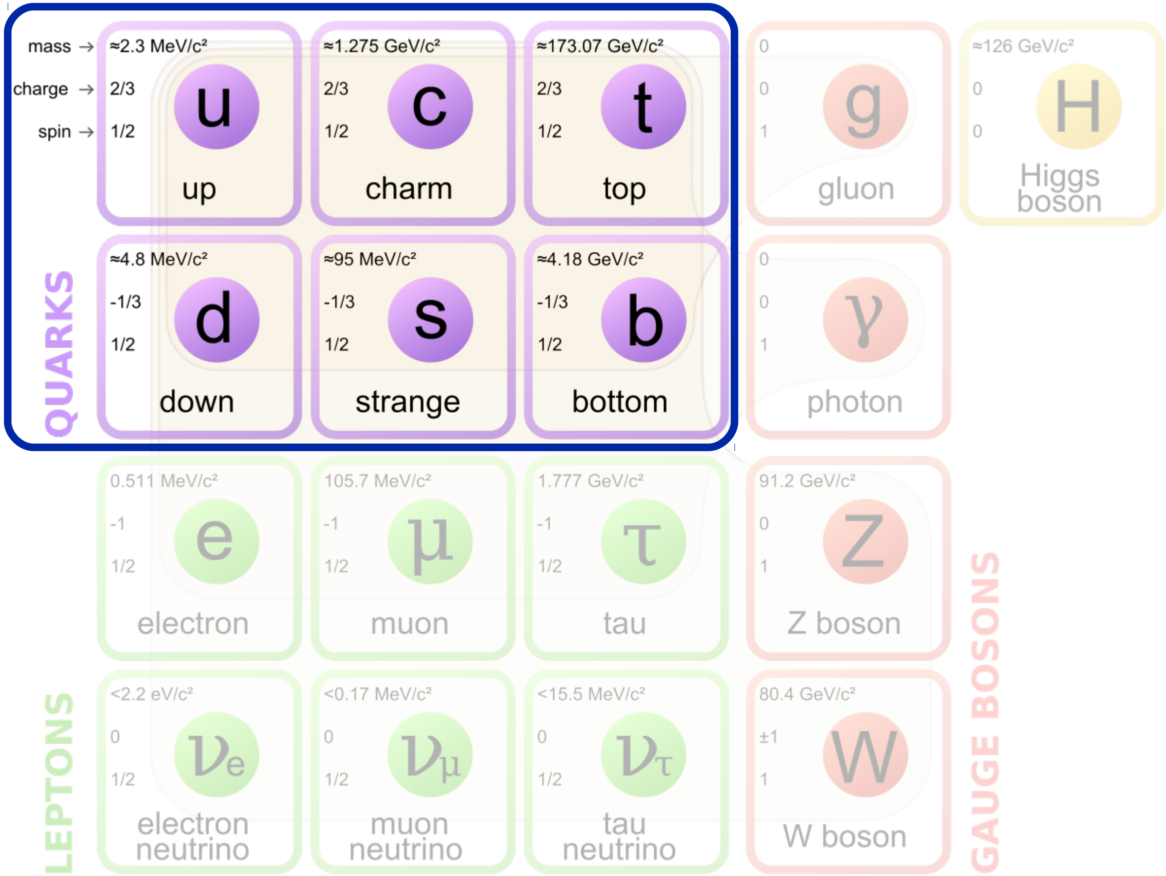


Standard Model



Quarks are not observed as free particles

Standard Model

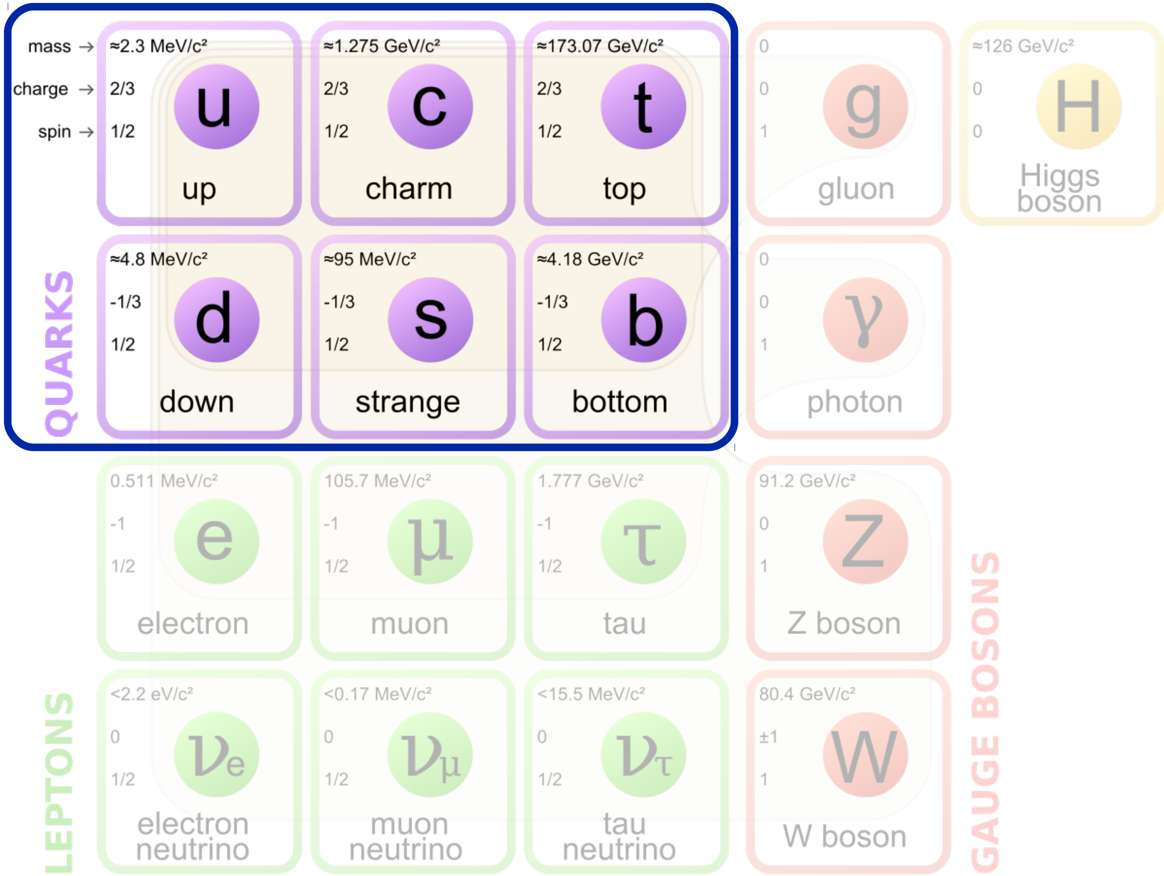


Particles that are observed consist of

- **three quarks (e.g. proton), or**
  - **a quark and an antiquark**

("exotic" combinations: Tetraquarks, Pentaquarks)

Standard Model

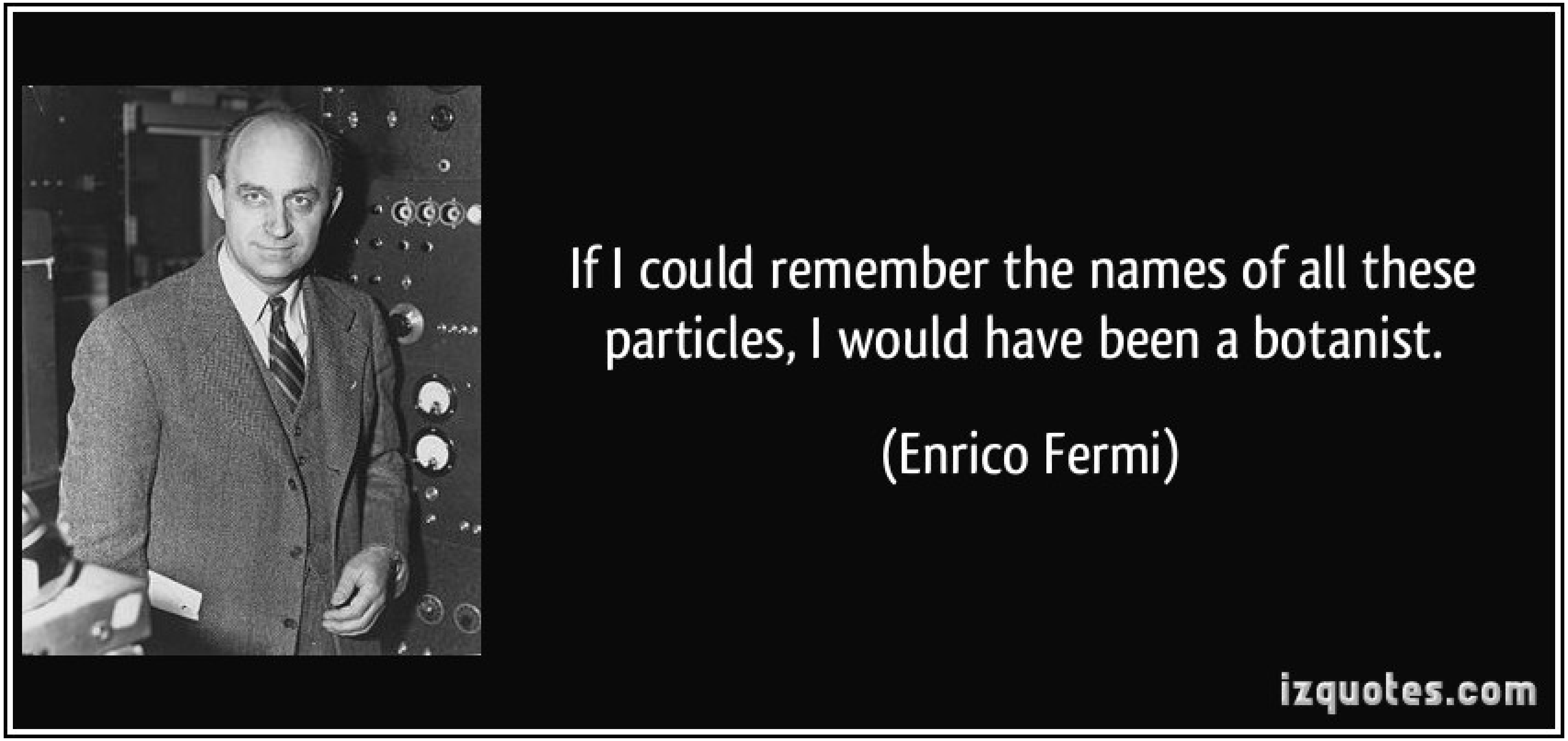


Particles that are observed consist of

- **three quarks (e.g. proton), or**
- **a quark and an antiquark**

("exotic" combinations: Tetraquarks, Pentaquarks)

Many possible combinations: "particle zoo"



If I could remember the names of all these particles, I would have been a botanist.

(Enrico Fermi)

izquotes.com

source: <http://izquotes.com/quote/389467>

Reconstructing an Event

Most particles in the particle zoo are very short-lived

Very few are stable or live long enough to leave a trace in a particle detector

electrons and muons

protons (uud)
pions ($u\bar{d}$) and kaons ($u\bar{s}$)

charged

photons
neutrons (udd)

neutral

… and their antiparticles

Reconstructing an Event

Short-lived particles can be reconstructed indirectly by measuring their long-lived decay products

Relativistic kinematics

$$\boxed{E = m \cdot c^2}$$

Reconstructing an Event

Short-lived particles can be reconstructed indirectly by measuring their long-lived decay products

Relativistic kinematics

$$\boxed{E = m}$$

“natural units”:
$c = 1$

Reconstructing an Event

Short-lived particles can be reconstructed indirectly by measuring their long-lived decay products

Relativistic kinematics

$$\boxed{E^2 = m^2 + p^2}$$

Reconstructing an Event

Short-lived particles can be reconstructed indirectly by measuring their long-lived decay products

Relativistic kinematics

$$\boxed{m^2 = E^2 - p^2}$$

Reconstructing an Event

Short-lived particles can be reconstructed indirectly by measuring their long-lived decay products

Relativistic kinematics

$$\boxed{m^2 = E^2 - p^2}$$

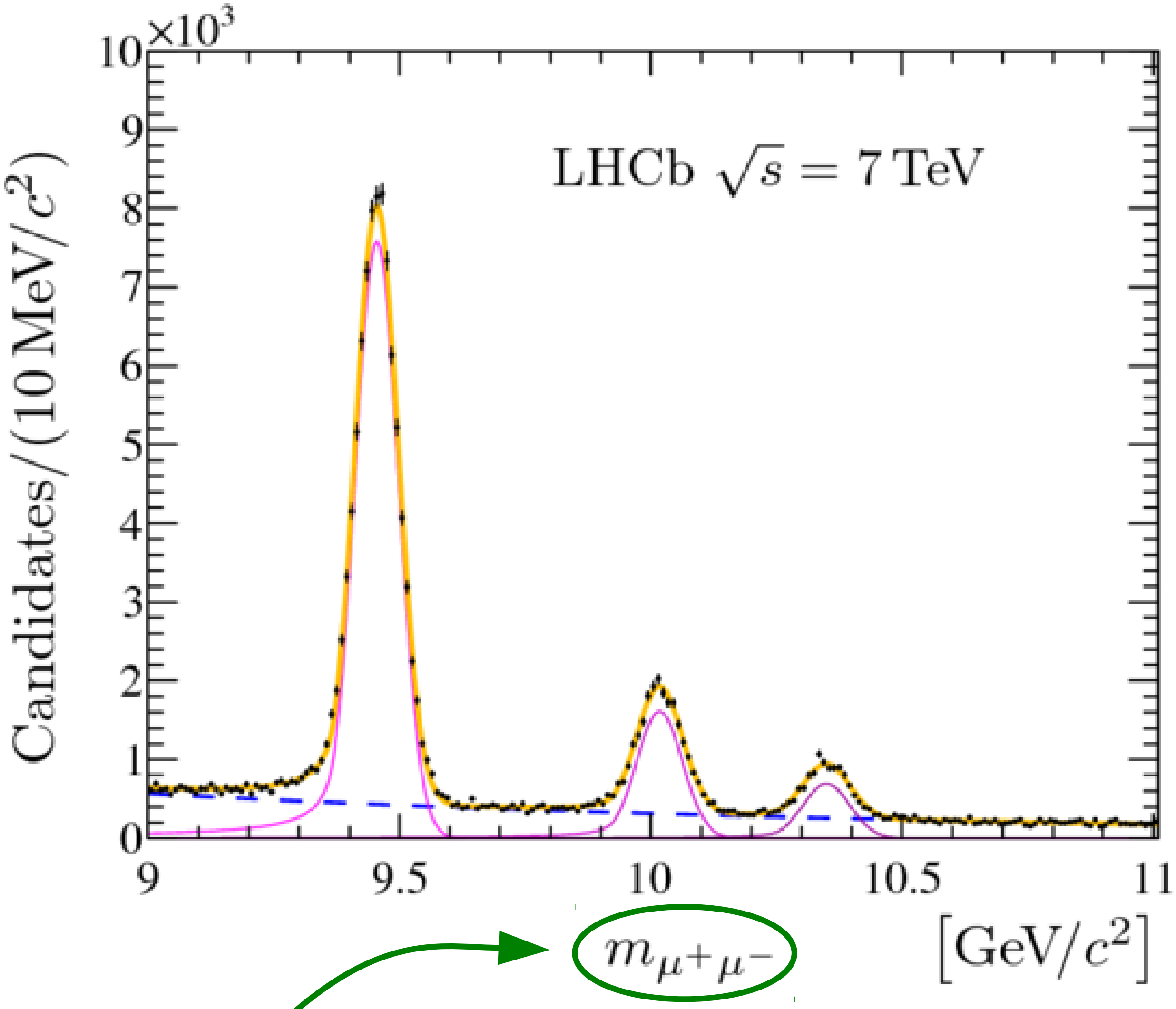
Energy and momentum conservation in the decay

$$\boxed{M^2 = \left(\sum_i E_i \right)^2 - \left| \sum_i \vec{p}_i \right|^2}$$

with

$$E_i^2 = m_i^2 + p_i^2$$

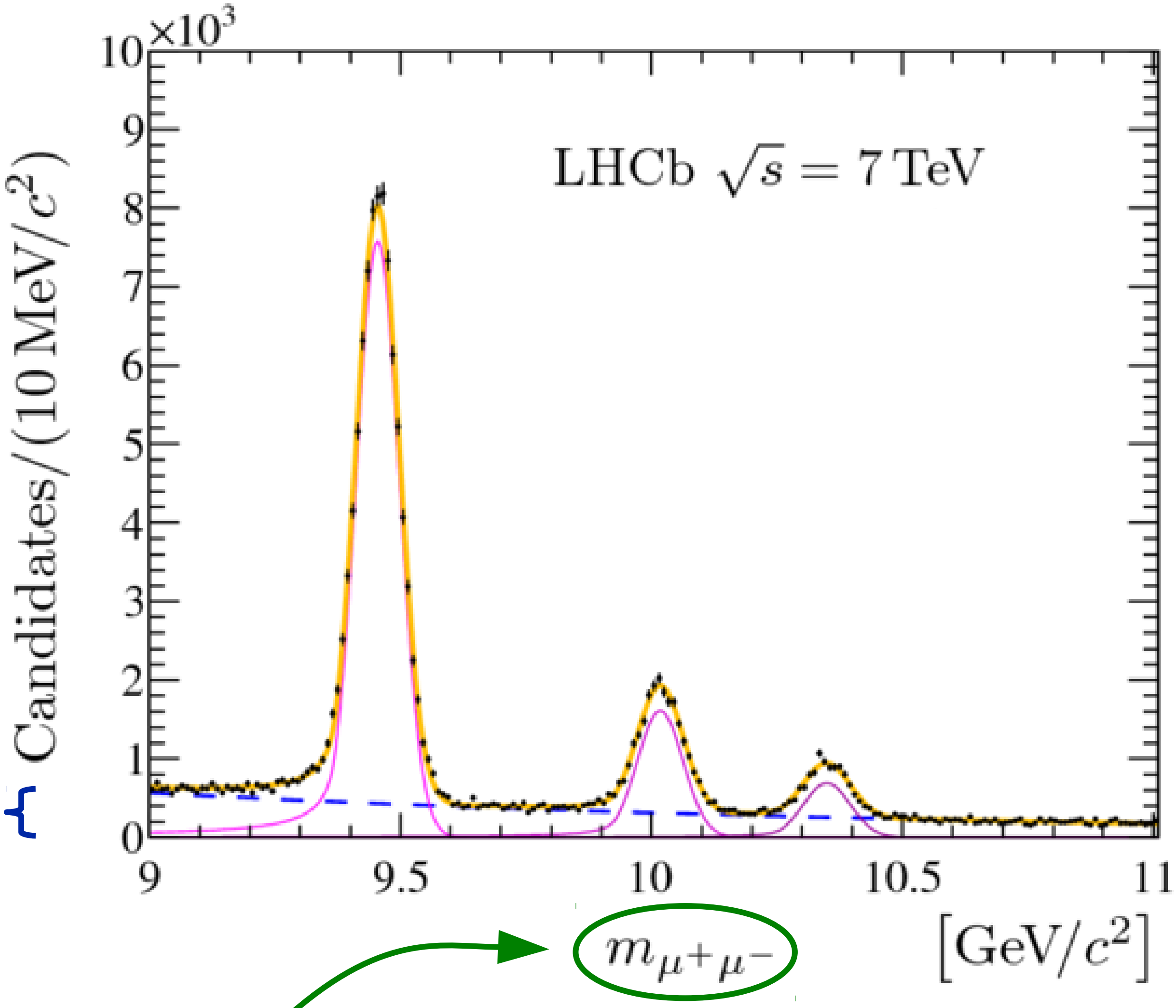
Reconstructing an Event



$$M^2 = \left(E_1 + E_2\right)^2 - \left|\vec{p}_1 + \vec{p}_2\right|^2$$

[The LHCb collaboration, R.Aaij et al., J.High Energy Phys.(2015) 103]

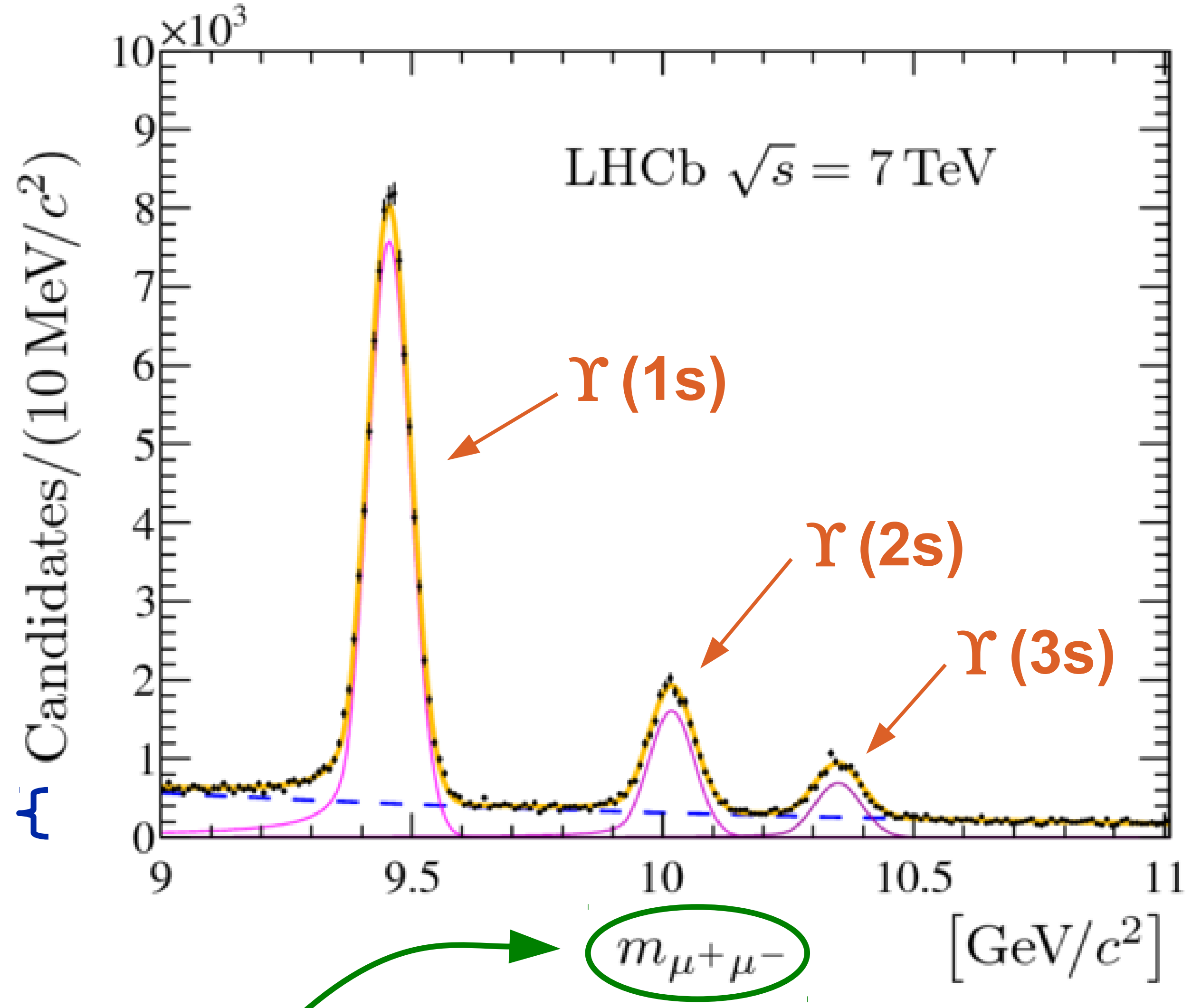
Reconstructing an Event



Random combinations of a muon and an antimuon

$$M^2 = \left(E_1 + E_2\right)^2 - \left|\vec{p}_1 + \vec{p}_2\right|^2$$

Reconstructing an Event



Υ (1s)

Υ (2s)

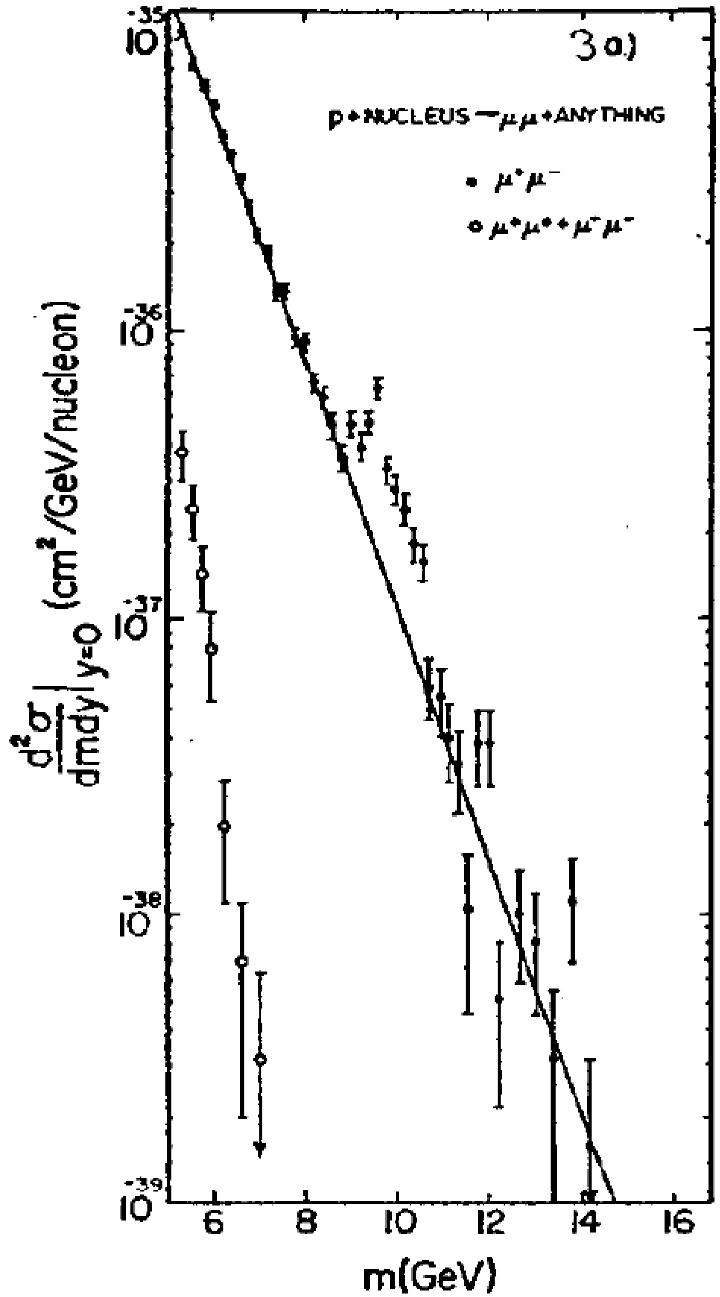
Υ (3s)

Υ particles:
short-lived
bound states
of *b* quark
and $\bar{b}$ quark

Random
combinations
of a muon
and an
antimuon

$$M^2 = \left(E_1 + E_2\right)^2 - \left|\vec{p}_1 + \vec{p}_2\right|^2$$

Discovery of ϒ Particles in 1977



OBSERVATION OF A DIMUON RESONANCE AT 9.5 GeV
IN 400 GeV PROTON-NUCLEUS COLLISIONS

S. W. Herb, D. C. Hom, L. M. Lederman,
J. C. Sens, H. D. Snyder, and J. K. Yoh
Columbia University, New York, New York 10027

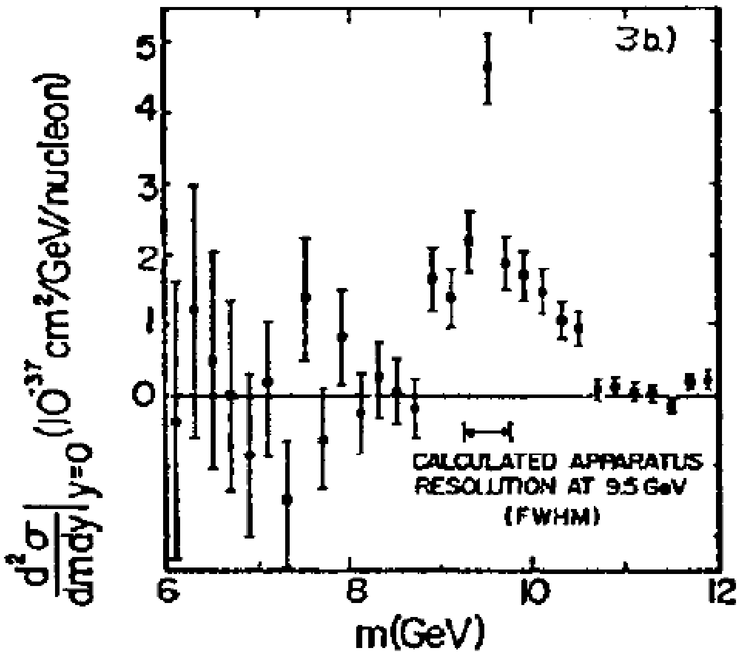
and

J. A. Appel, B. C. Brown, C. N. Brown
W. R. Innes, K. Ueno, and T. Yamanouchi
Fermi National Accelerator Laboratory, Batavia, Illinois 60510

and

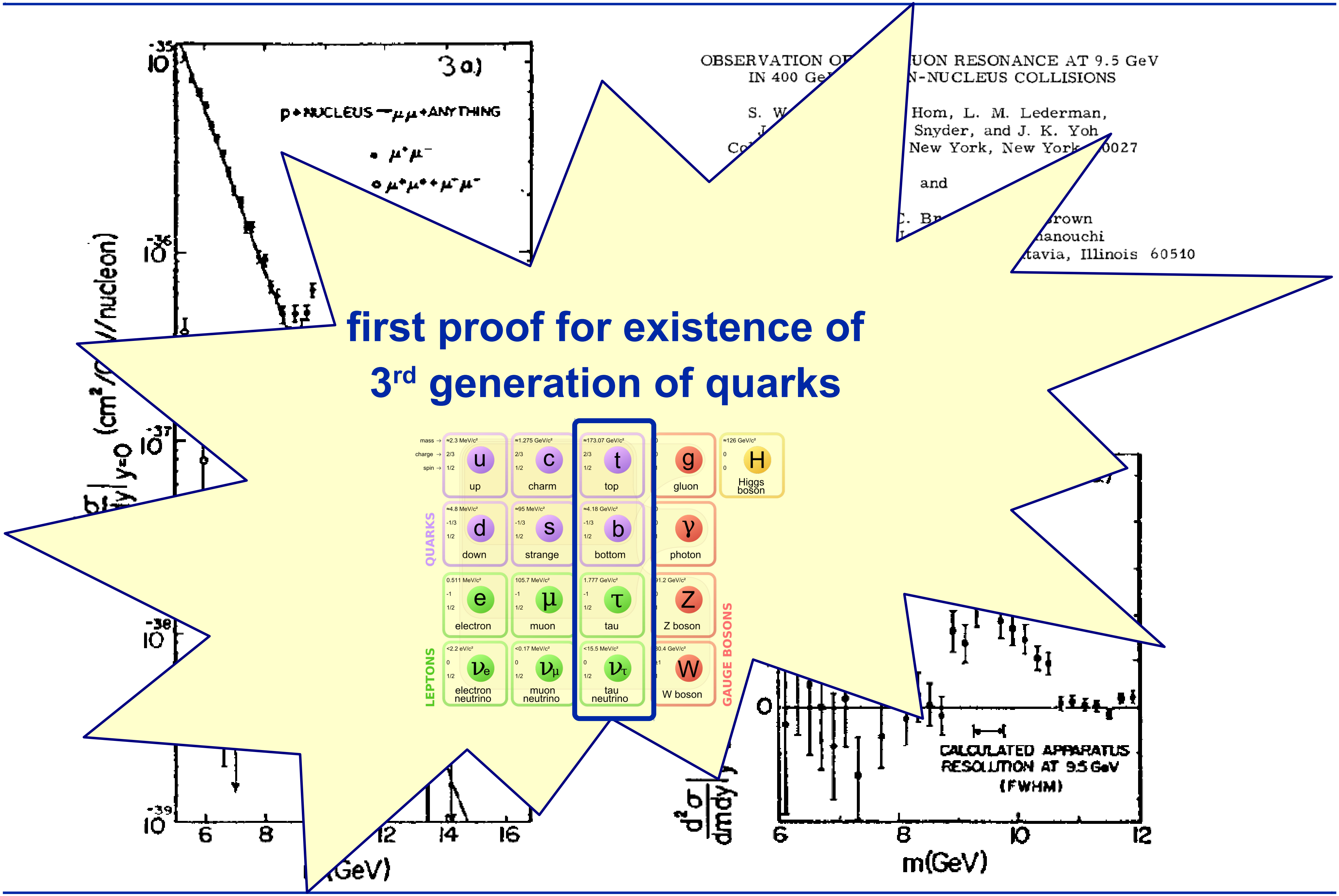
A. S. Ito, H. Jöstlein, D. M. Kaplan,
and R. D. Kephart
State University of New York at Stony Brook
Stony Brook, New York 11794

July 1977



[S.W.Herb et al., Fermilab-Pub-77/58-EXP]

Discovery of Υ Particles in 1977



OBSERVATION OF ... UON RESONANCE AT 9.5 GeV
IN 400 Ge... N-NUCLEUS COLLISIONS

S. W... Hom, L. M. Lederman,
J... Snyder, and J. K. Yoh
Co... New York, New York ...0027

and

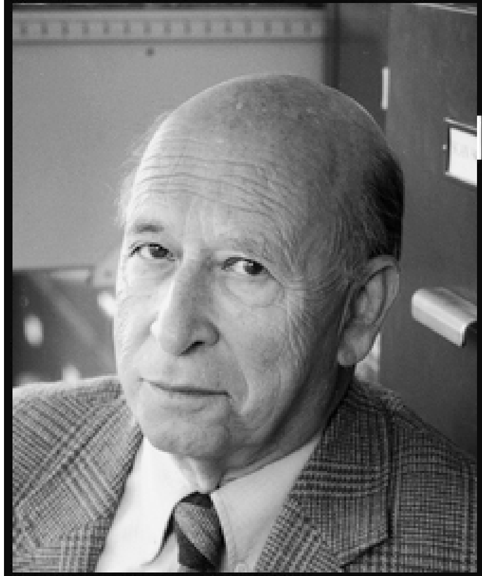
...C. Br... rown
... anouchi
... tavia, Illinois 60510

first proof for existence of
3rd generation of quarks

**Yesterday's sensation
is today's calibration channel**

(Richard P. Feynman)

photo source: <https://en.wikipedia.org/wiki/Richard_Feynman>

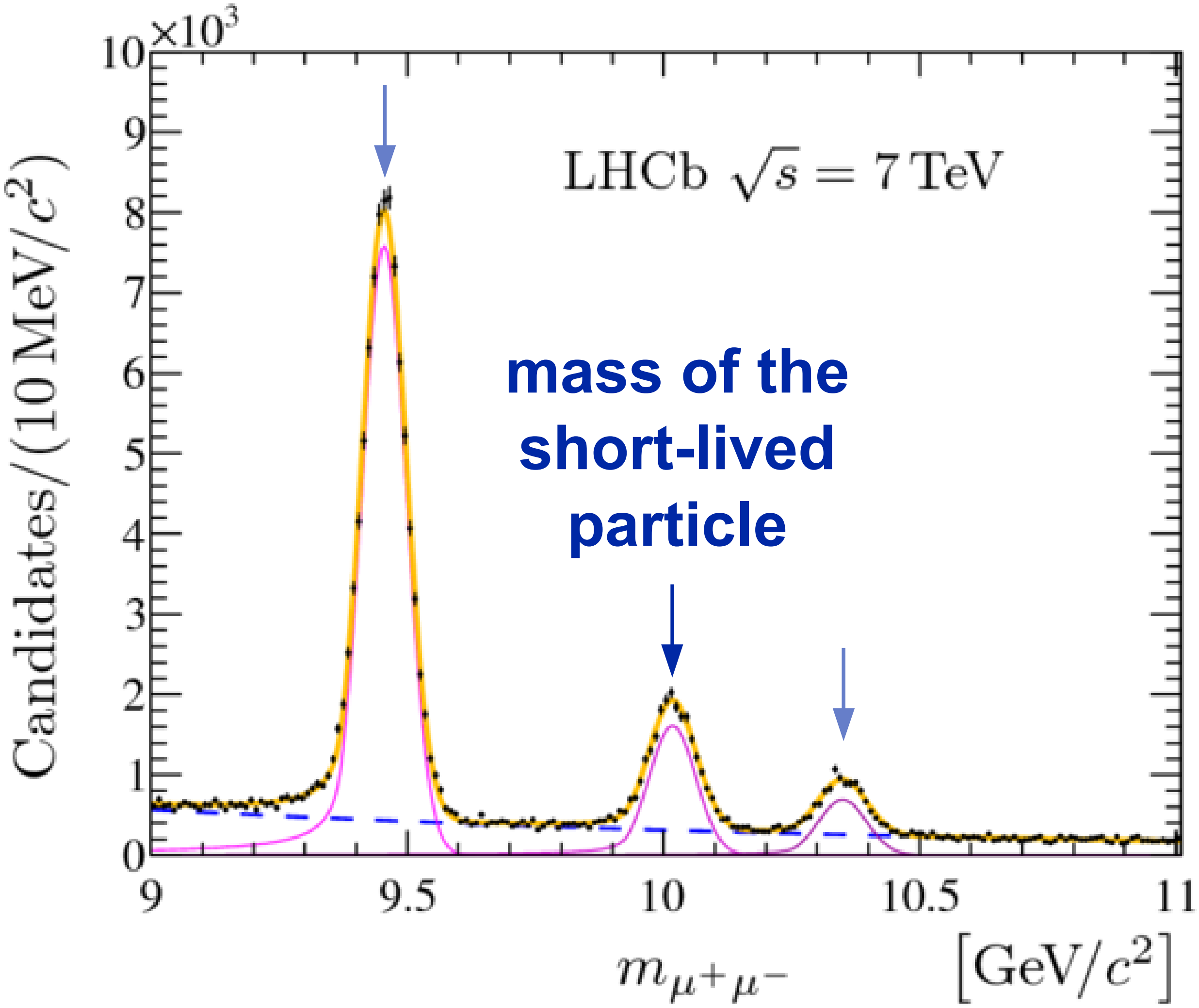


… and tomorrow's background

(Valentine Telegdi)

photo source: <https://en.wikipedia.org/wiki/Valentine_Telegdi>

… Today's Calibration Channel

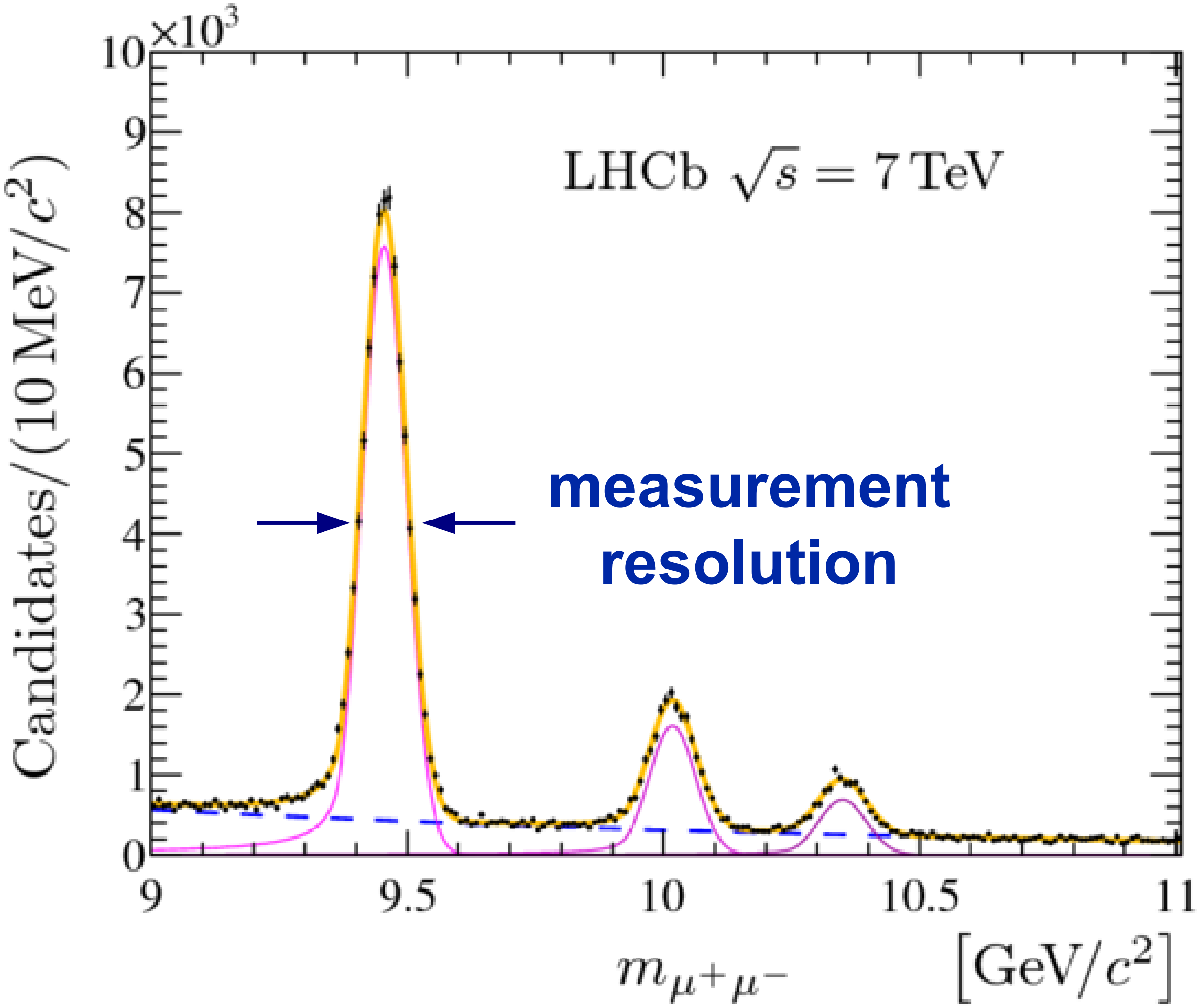


mass of the short-lived particle

Distribution should peak at the known mass values:

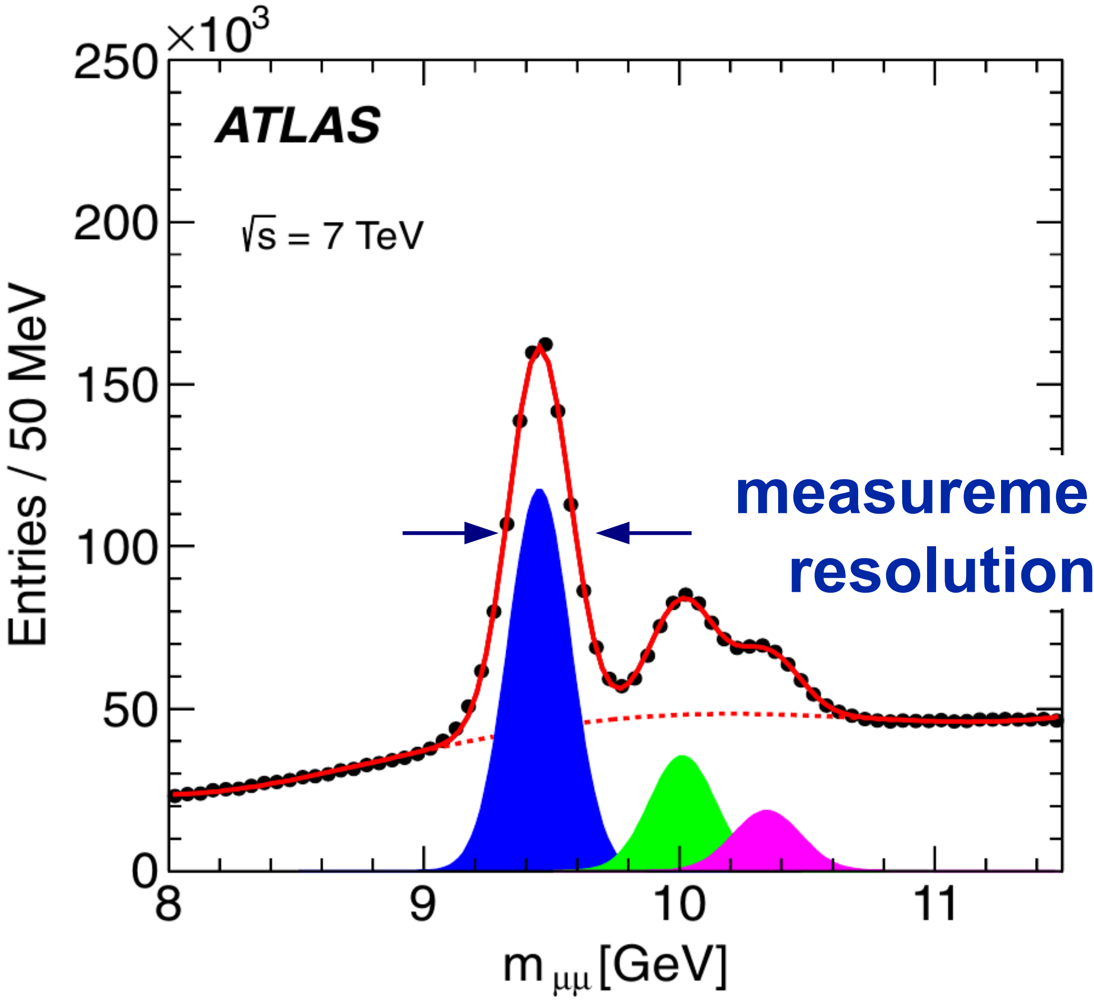
Verify / calibrate your momentum reconstruction

… Today's Calibration Channel



Width of the peaks measures precision of the momentum reconstruction

… Today's Calibration Channel



measurement resolution

[G.Aad et al. (ATLAS Collaboration), Phys.Rev.D 87 (2013) 052004]

Width of the peaks measures precision of the momentum reconstruction

Reconstructing an Event

To reconstruct an event, need to:

- Measure the flight directions of long-lived particles
- Measure the magnitudes of their momenta
- Determine which type of particle they are (to know their mass and energy)

Reconstructing an Event

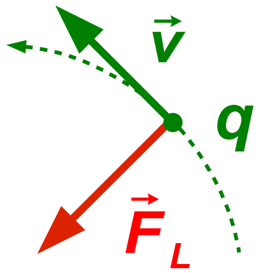
To reconstruct an event, need to:

- Measure the flight directions of long-lived particles
- **Measure the magnitudes of their momenta**
- Determine which type of particle they are
  (to know their mass and energy)

Measuring Momentum

Apply a magnetic field → Lorentz force on moving particle

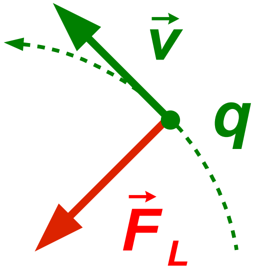
$$\boxed{\vec{F}_L = q \cdot \vec{v} \times \vec{B}}$$



Measuring Momentum

Apply a magnetic field → Lorentz force on moving particle

$$\boxed{\vec{F}_L = q \cdot \vec{v} \times \vec{B}}$$



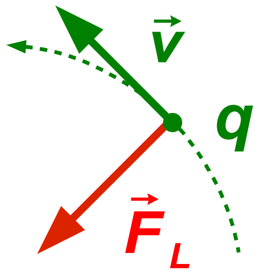
Particle forced onto a circular trajectory

$$\frac{m \cdot v^2}{r} = q \cdot v \cdot B \quad \Rightarrow \quad \boxed{p = q \cdot B \cdot r}$$

Measuring Momentum

Apply a magnetic field → Lorentz force on moving particle

$$\boxed{\vec{F}_L = q \cdot \vec{v} \times \vec{B}}$$



Particle forced onto a circular trajectory

$$\frac{m \cdot v^2}{r} = q \cdot v \cdot B \quad \Rightarrow \quad \boxed{p = q \cdot B \cdot r}$$

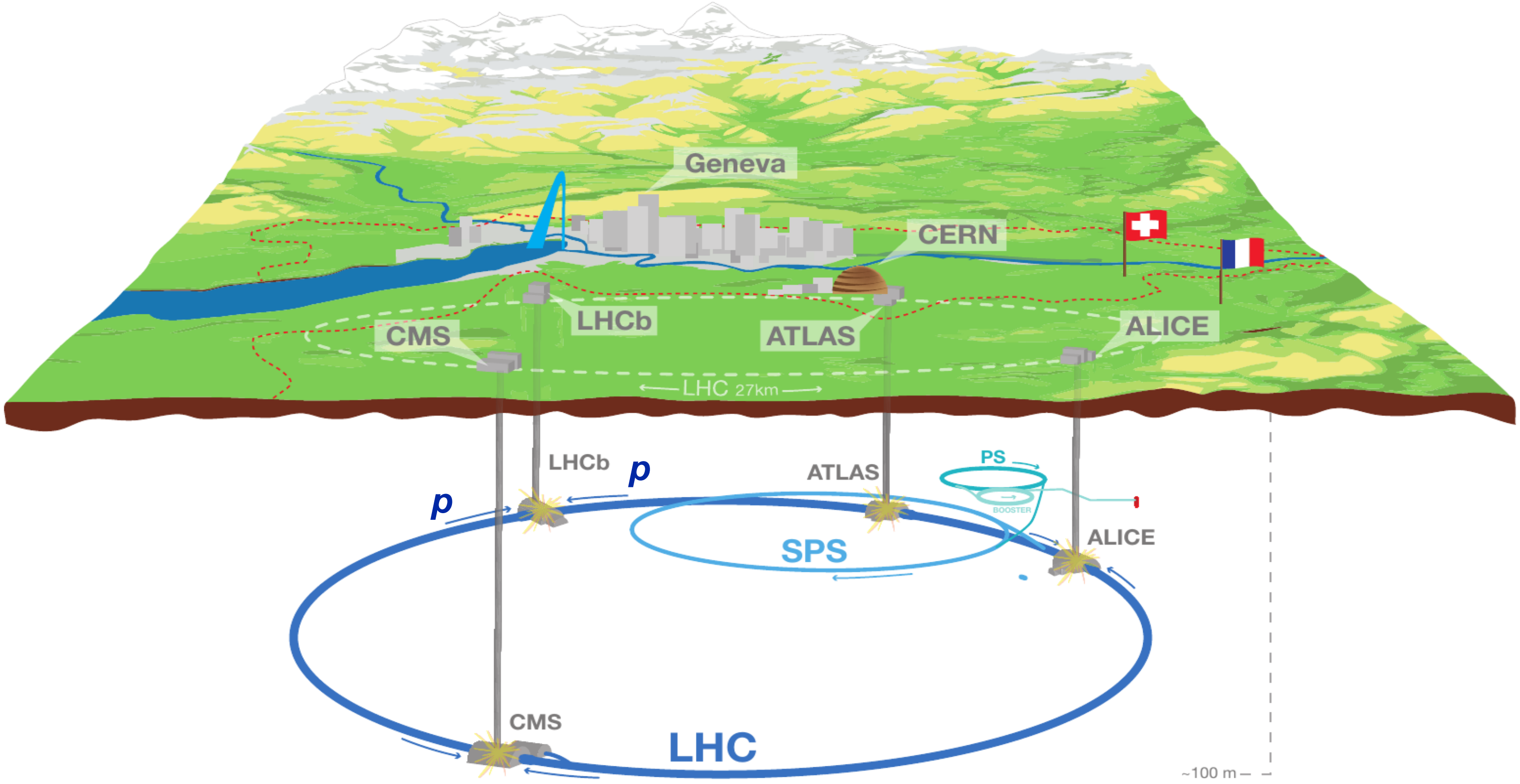
Precision of the momentum measurement

$$\boxed{\frac{\Delta p}{p} \propto \frac{\sigma_x}{\sqrt{N}} \cdot \frac{p}{B \cdot L^2}}$$

precision & number of position measurements

length over which the trajectory is measured

Large Hadron Collider



source: <http://cds.cern.ch/record/1708849>

27 km long ring in a tunnel 100 m below ground

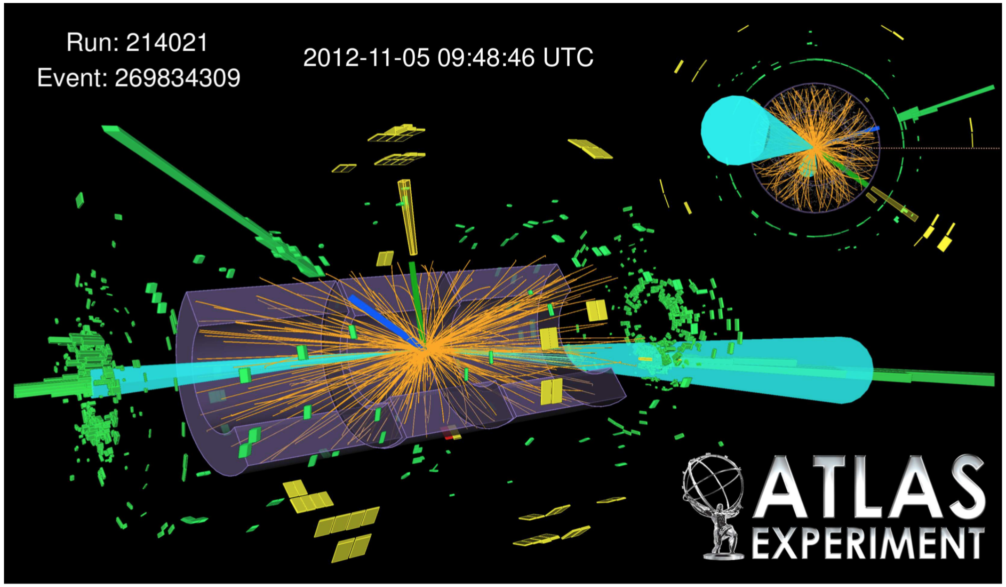
Two proton beams: clockwise and anti-clockwise

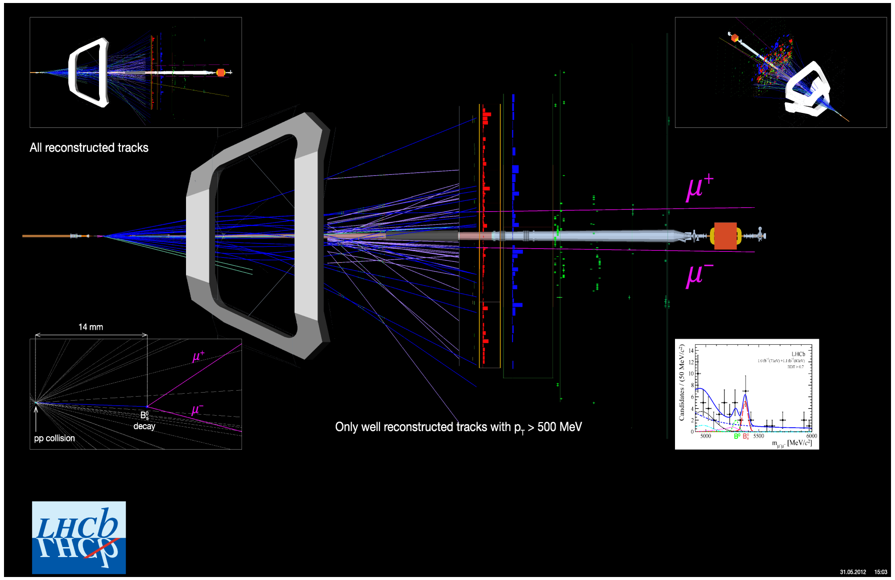
→ proton collisions every 25 ns (40 million times per second)

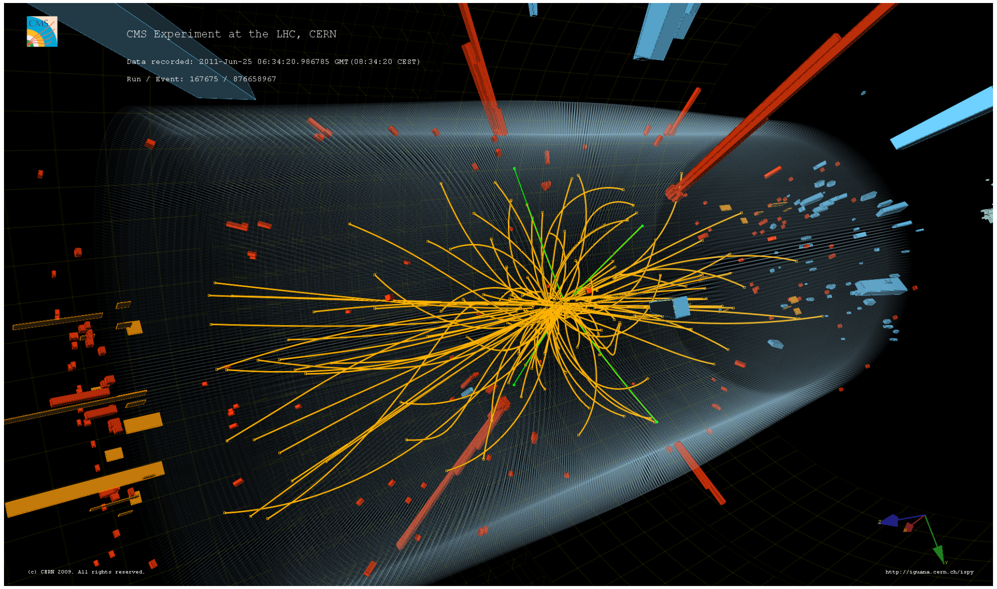
Four collision points: four large experiments

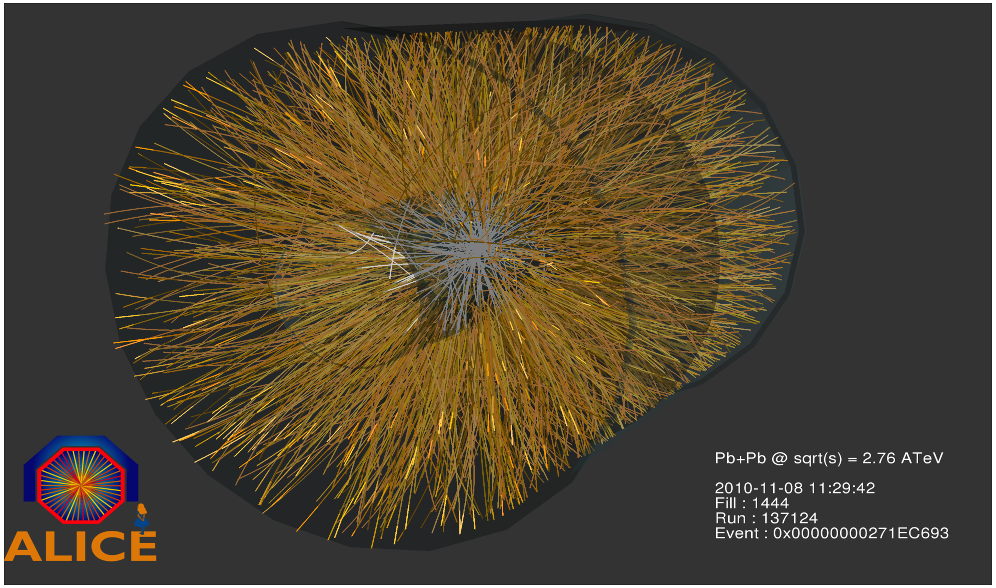
→ more than 10'000 physicists, more than 100 nationalities

LHC Experiments

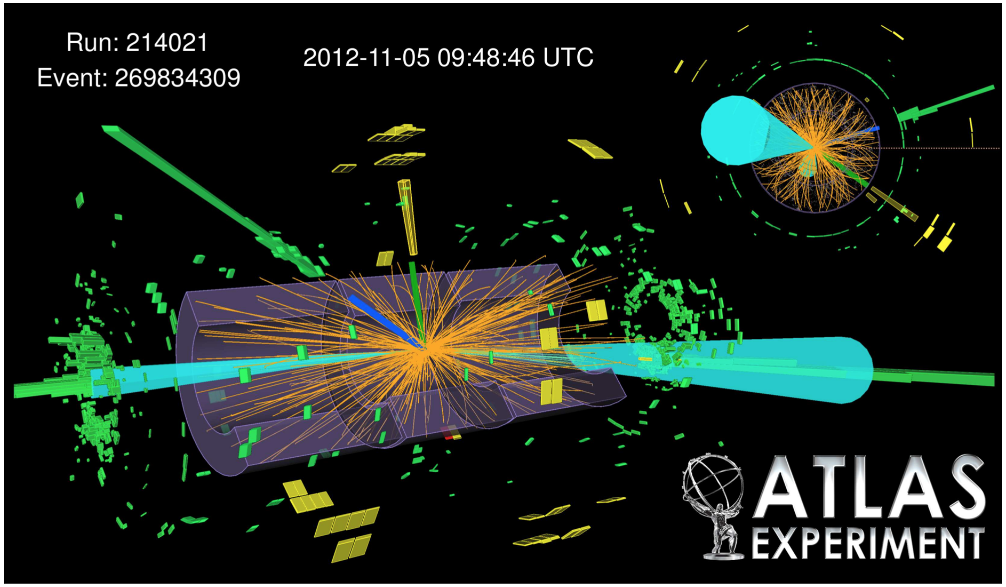


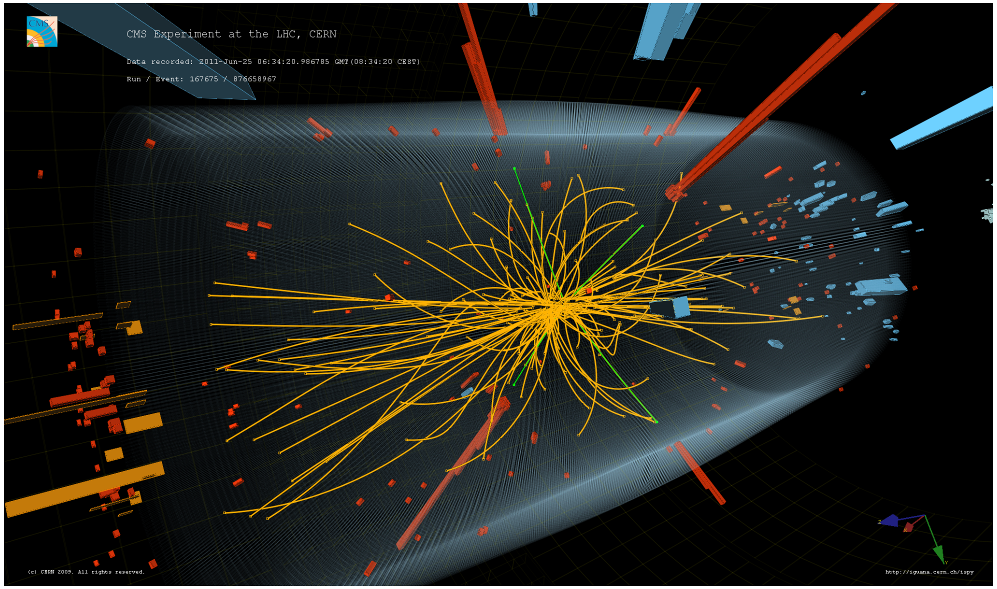


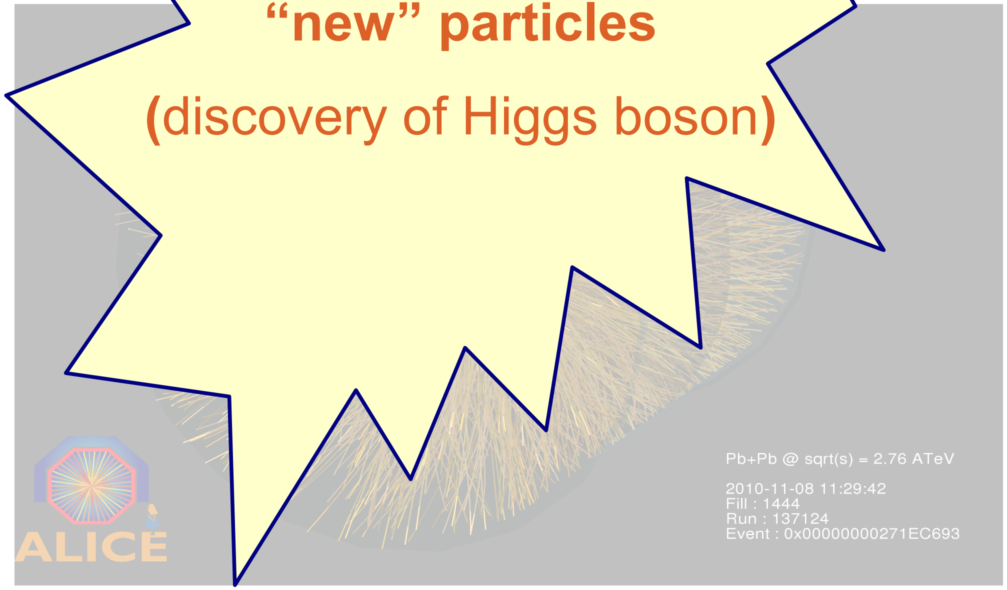




LHC Experiments





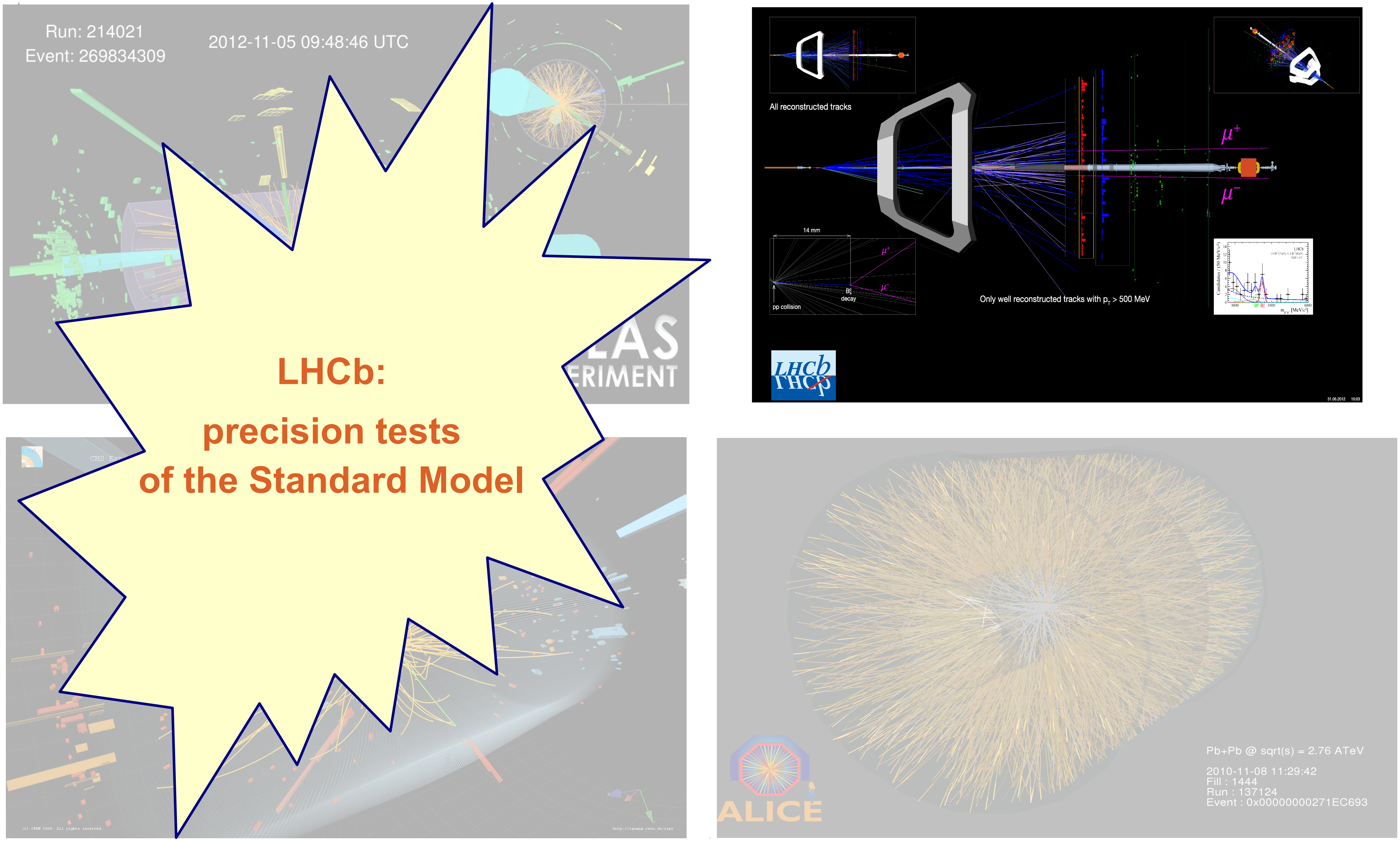


ATLAS / CMS:

search for
“new” particles

(discovery of Higgs boson)

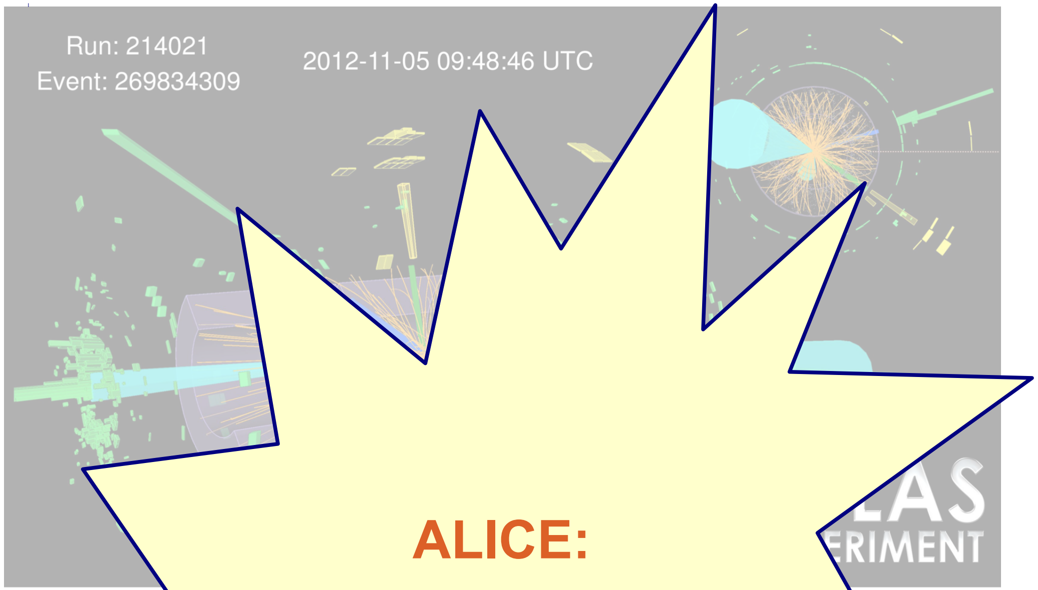
LHC Experiments

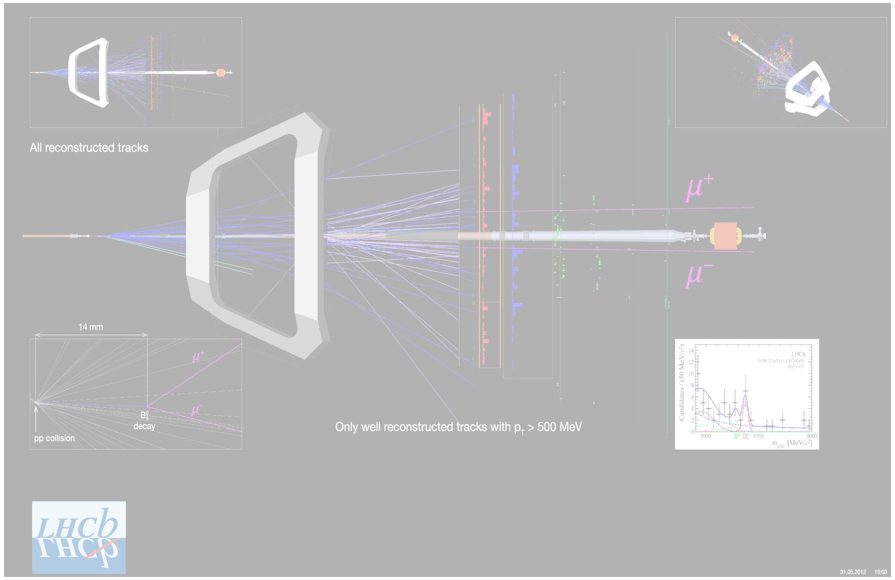


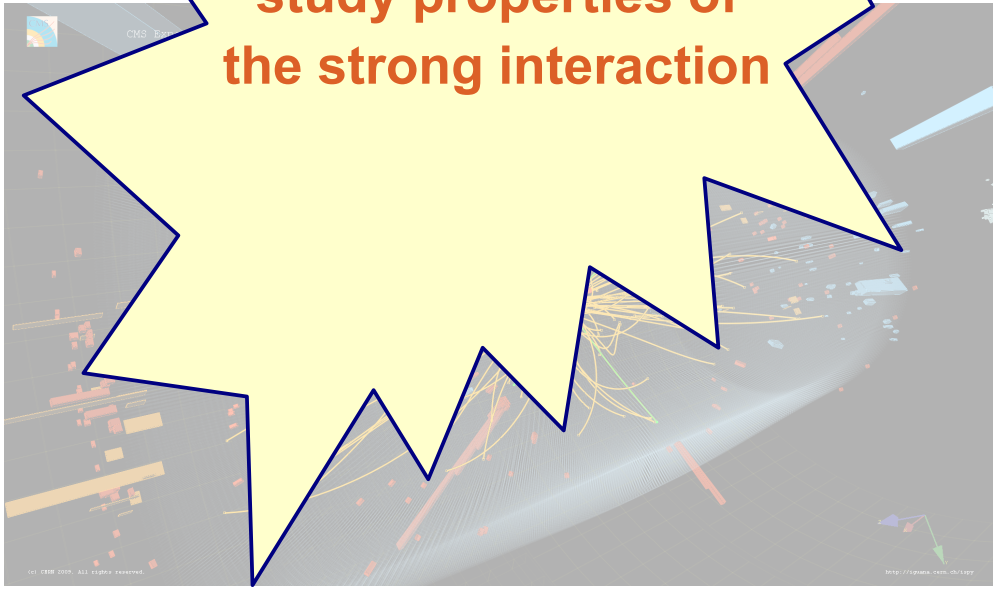
LHCb:

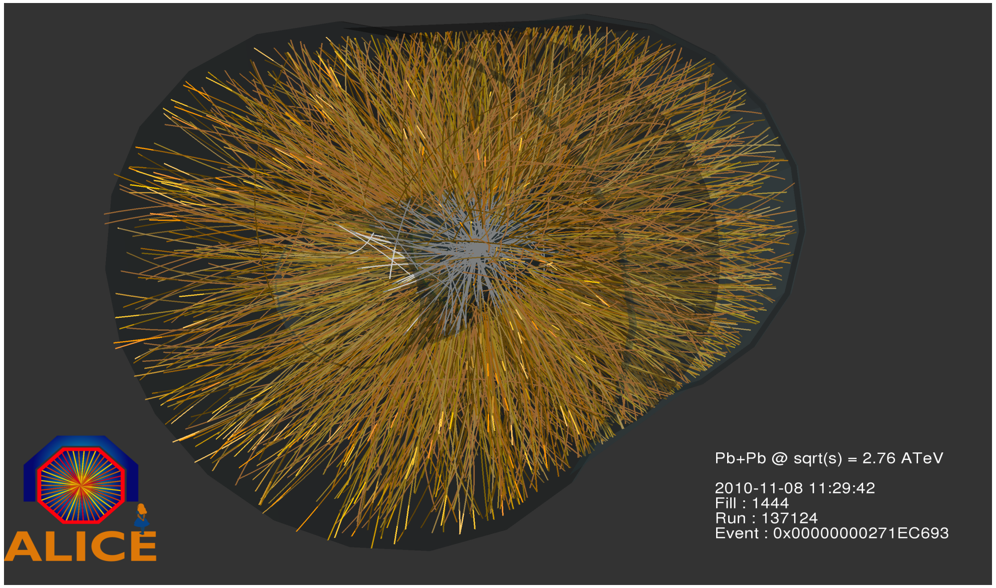
precision tests of the Standard Model

LHC Experiments





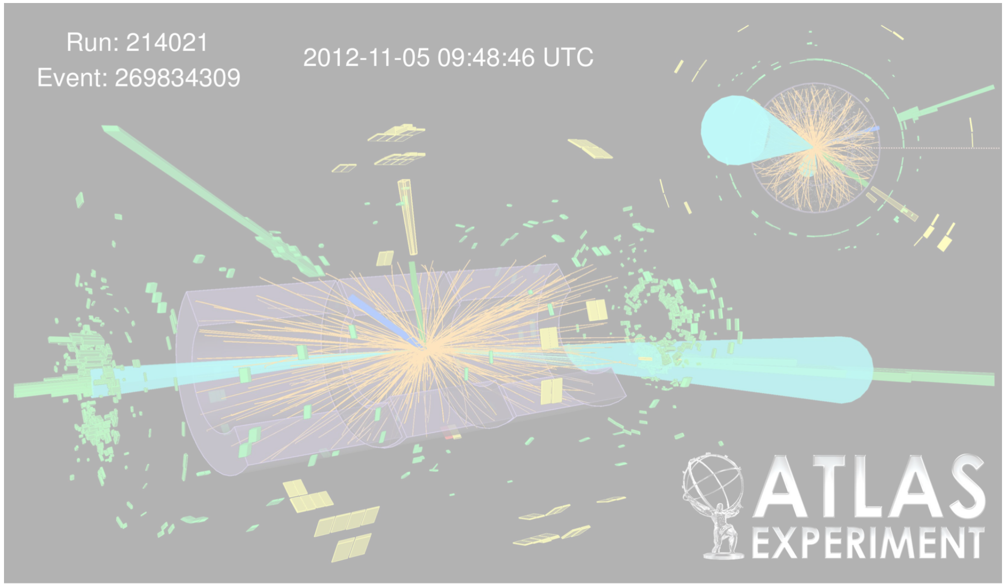


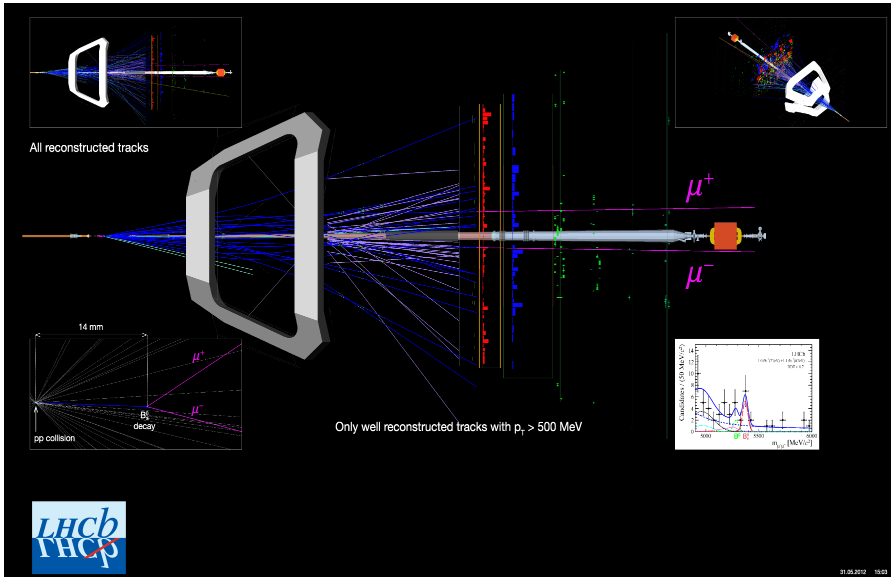


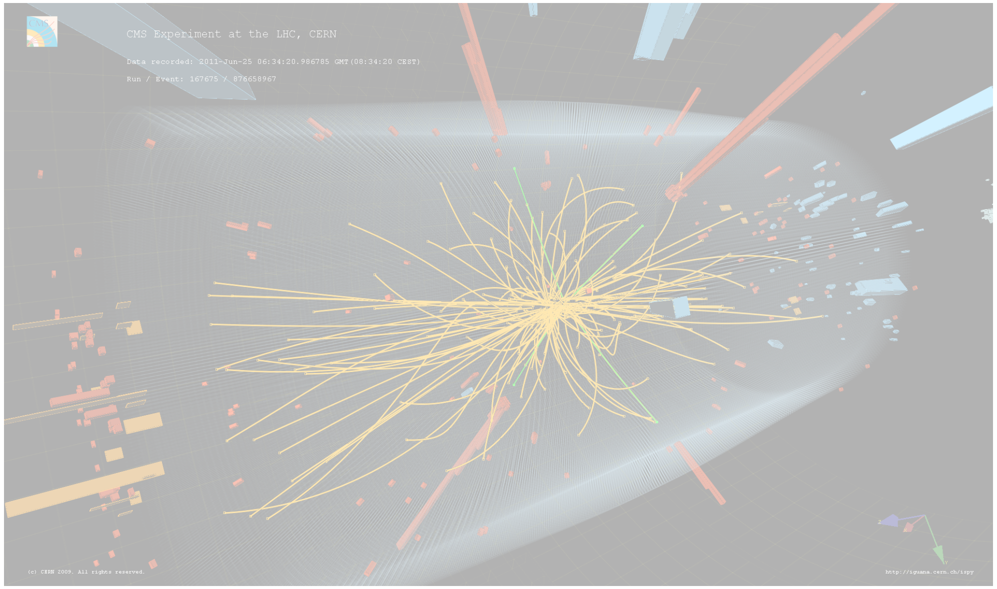
ALICE:

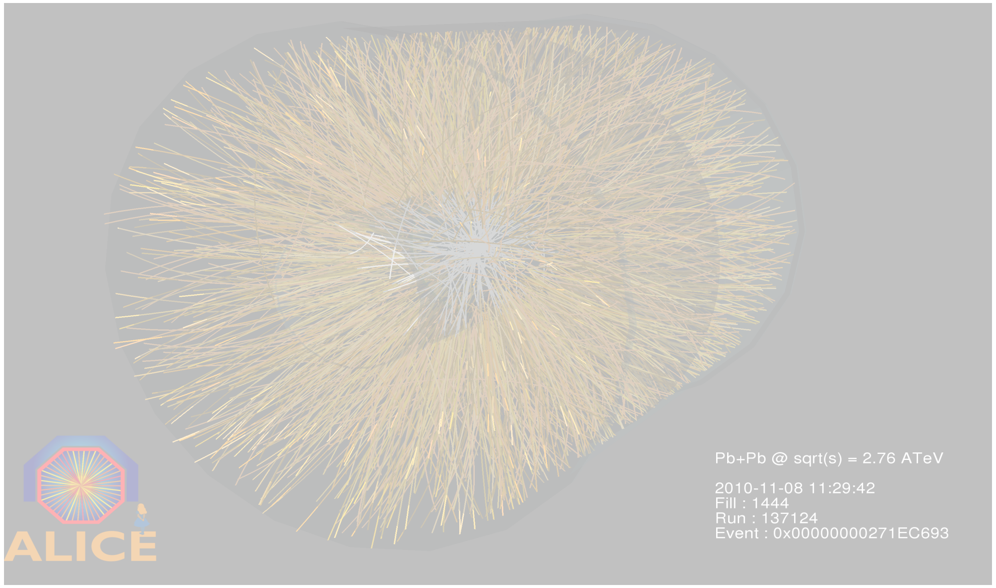
study properties of the strong interaction

LHC Experiments





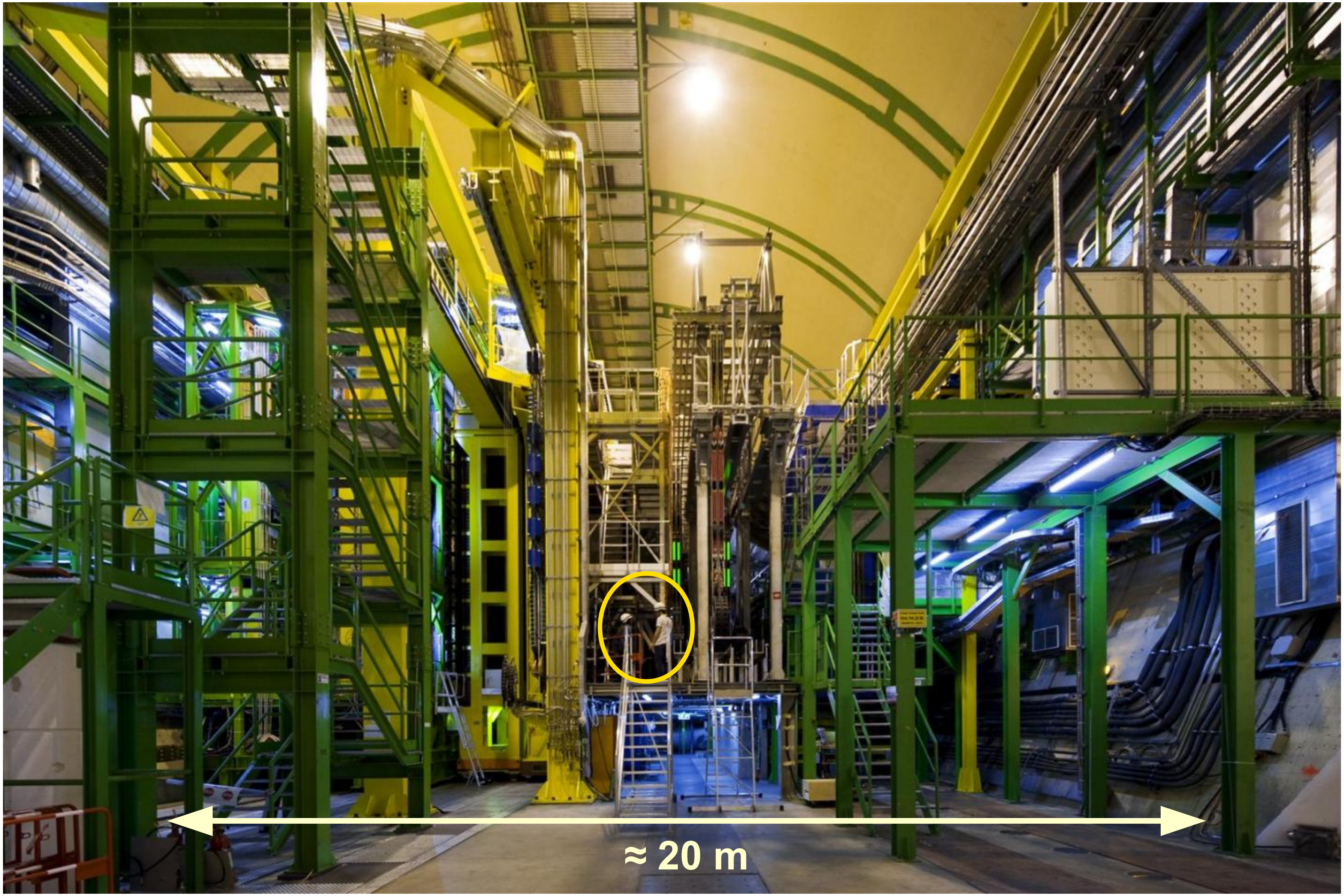




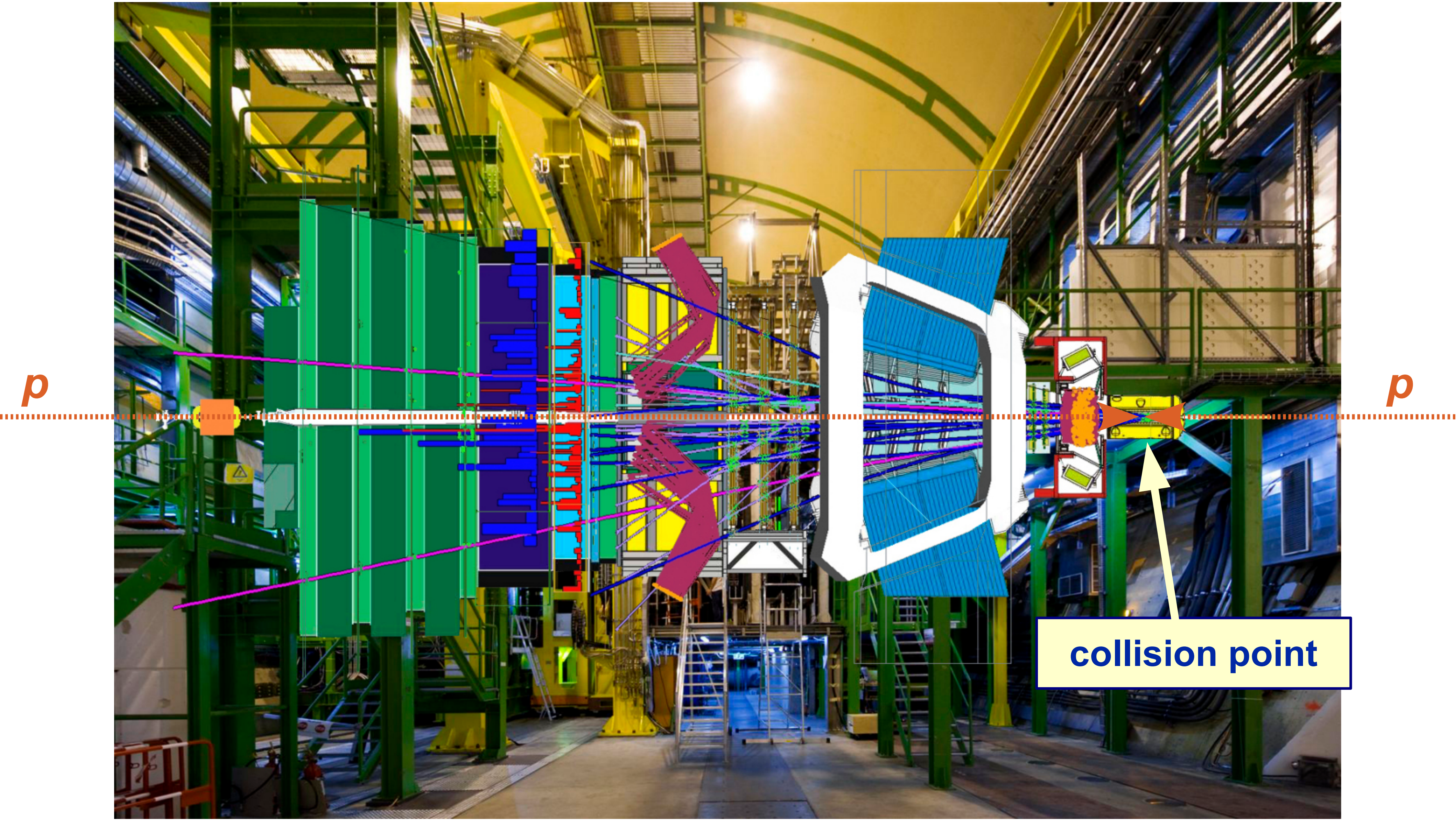
LHCb Experiment

source: <https://cds.cern.ch/record/1463546>

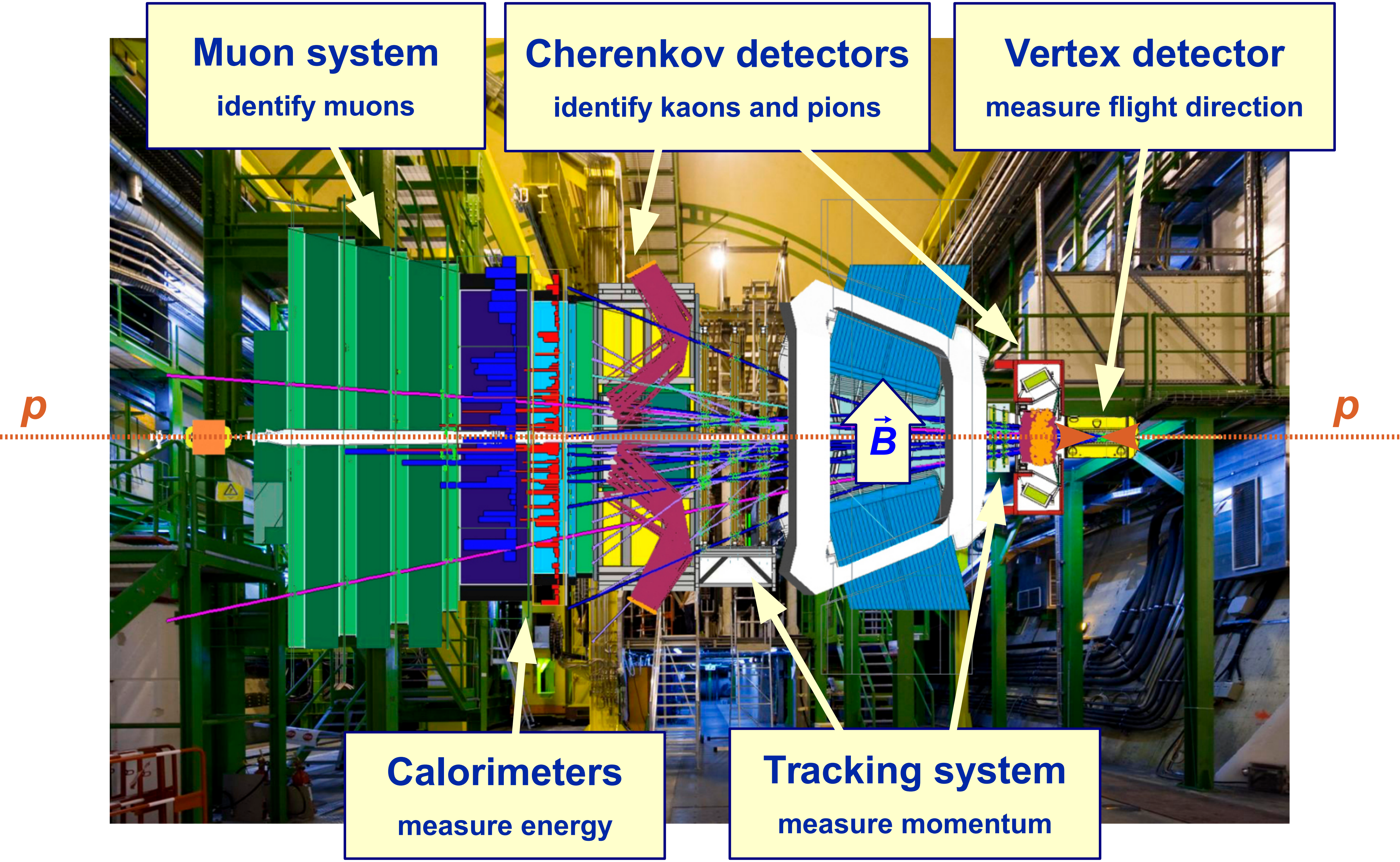
LHCb Experiment



LHCb Experiment



LHCb Experiment



Muon system

identify muons

Cherenkov detectors

identify kaons and pions

Vertex detector

measure flight direction

Calorimeters

measure energy

Tracking system

measure momentum

Overview

Part I

What are we trying to measure

What do our experiments look like

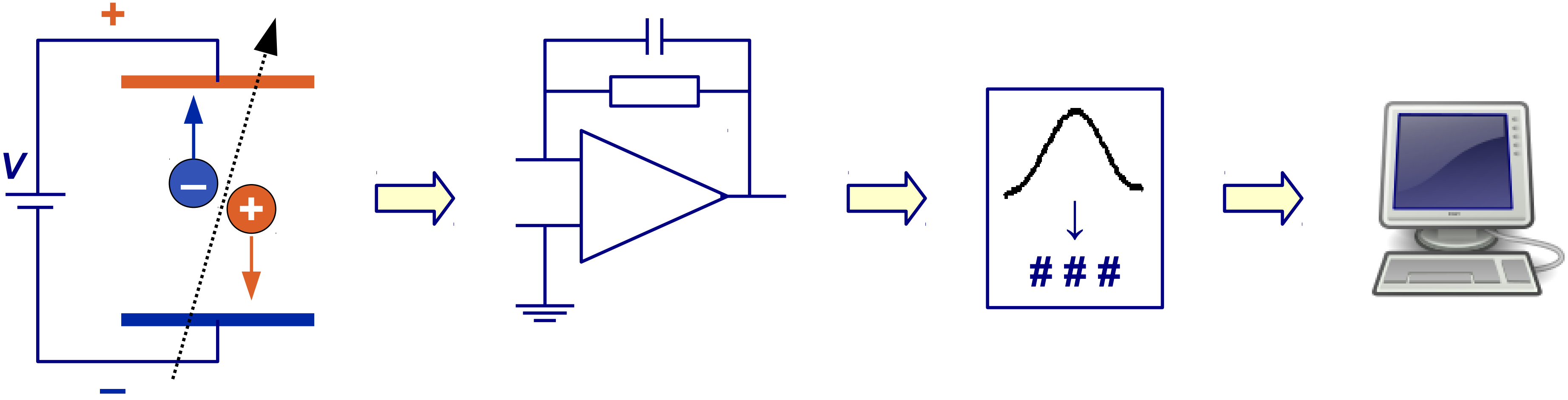
Part II

“Online” event selection

Part III

“Offline” analysis

Readout



particle deposits
energy in detector
→ voltage pulse

amplification

digitization

storage,
reconstruction,
analysis

Millions of detector channels → large amount of data

ATLAS / CMS: 1-2 MB per event

LHCb: about 100 kB per event

Data Volume

1.5 MB per event × 40 million events per second

≈ 60 TB / sec

> 10'000 DVDs every second

≈ 6 billion phone calls

Facebook: ≈ 600 TB / day

→ Can't afford to store everything !

Data Volume

Facebook
180 PB/yr

(1 PB = 1000 TB)

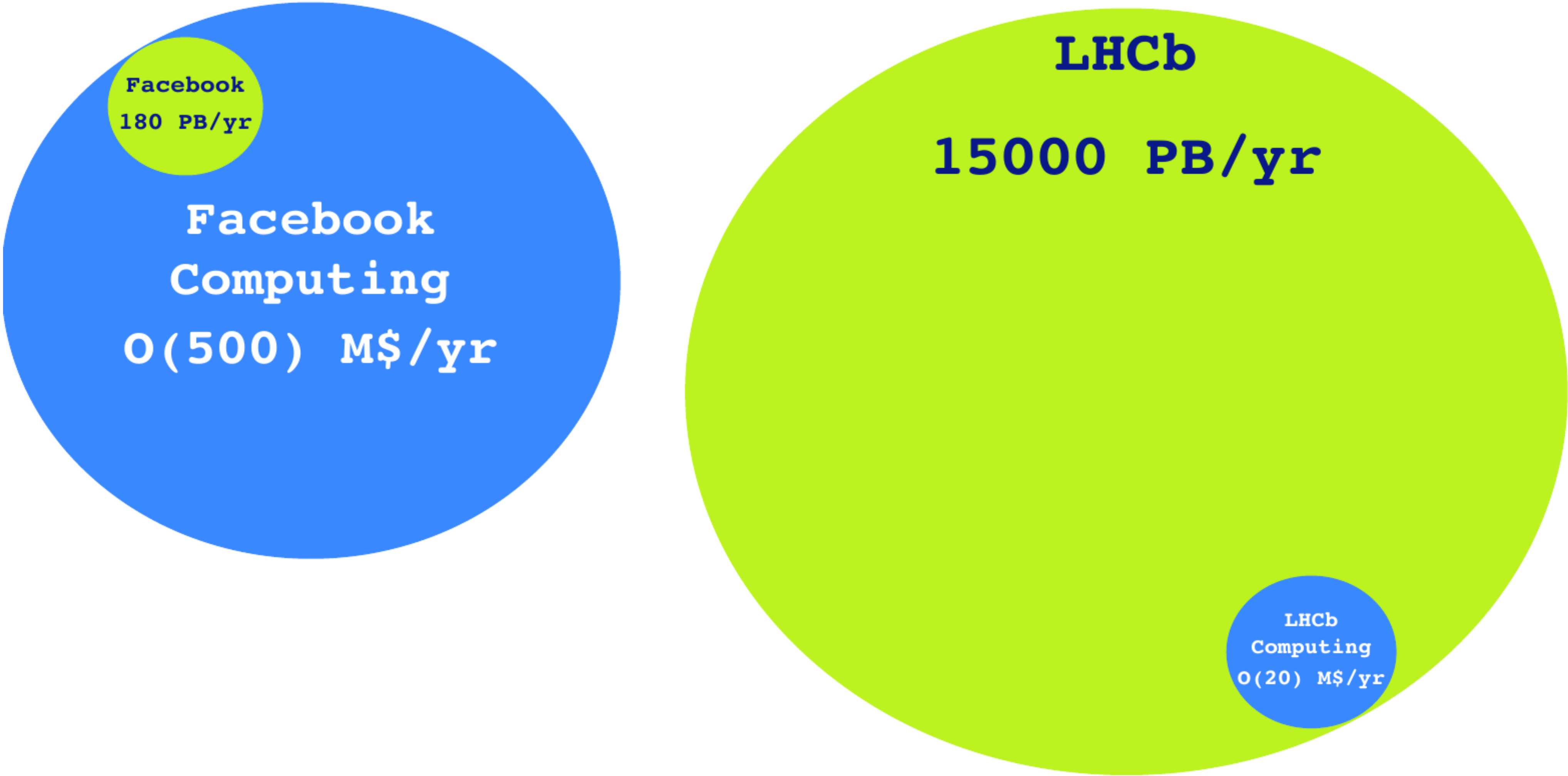
Data Volume

Facebook
180 PB/yr

LHCb
15000 PB/yr

(1 PB = 1000 TB)

Data Volume



Facebook
180 PB/yr

Facebook
Computing
O(500) M$/yr

LHCb
15000 PB/yr

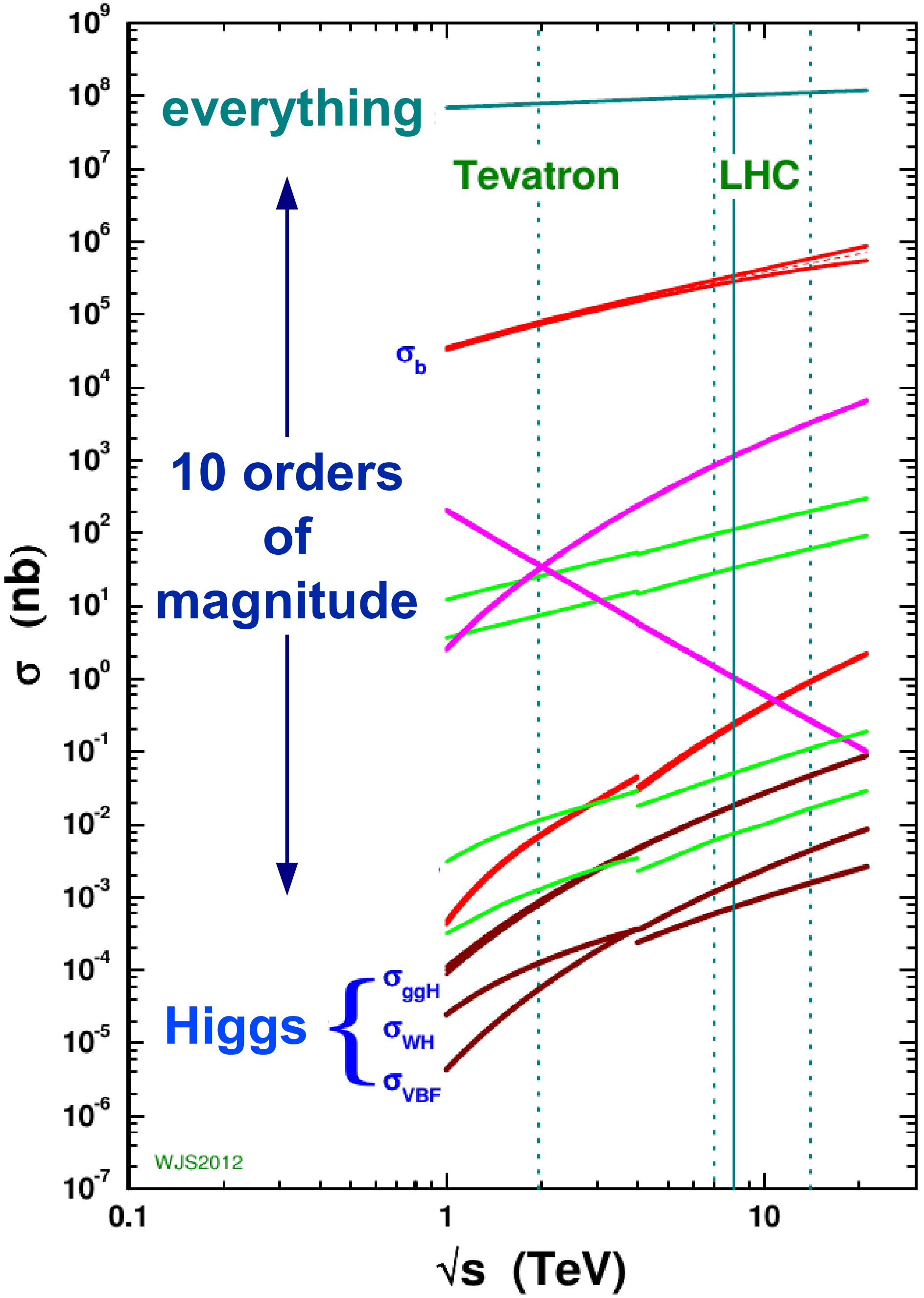
LHCb
Computing
O(20) M$/yr

(1 PB = 1000 TB)

Data Volume

But: we are interested in studying rare processes !

e.g. On average about one Higgs boson / sec



plot: <http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>

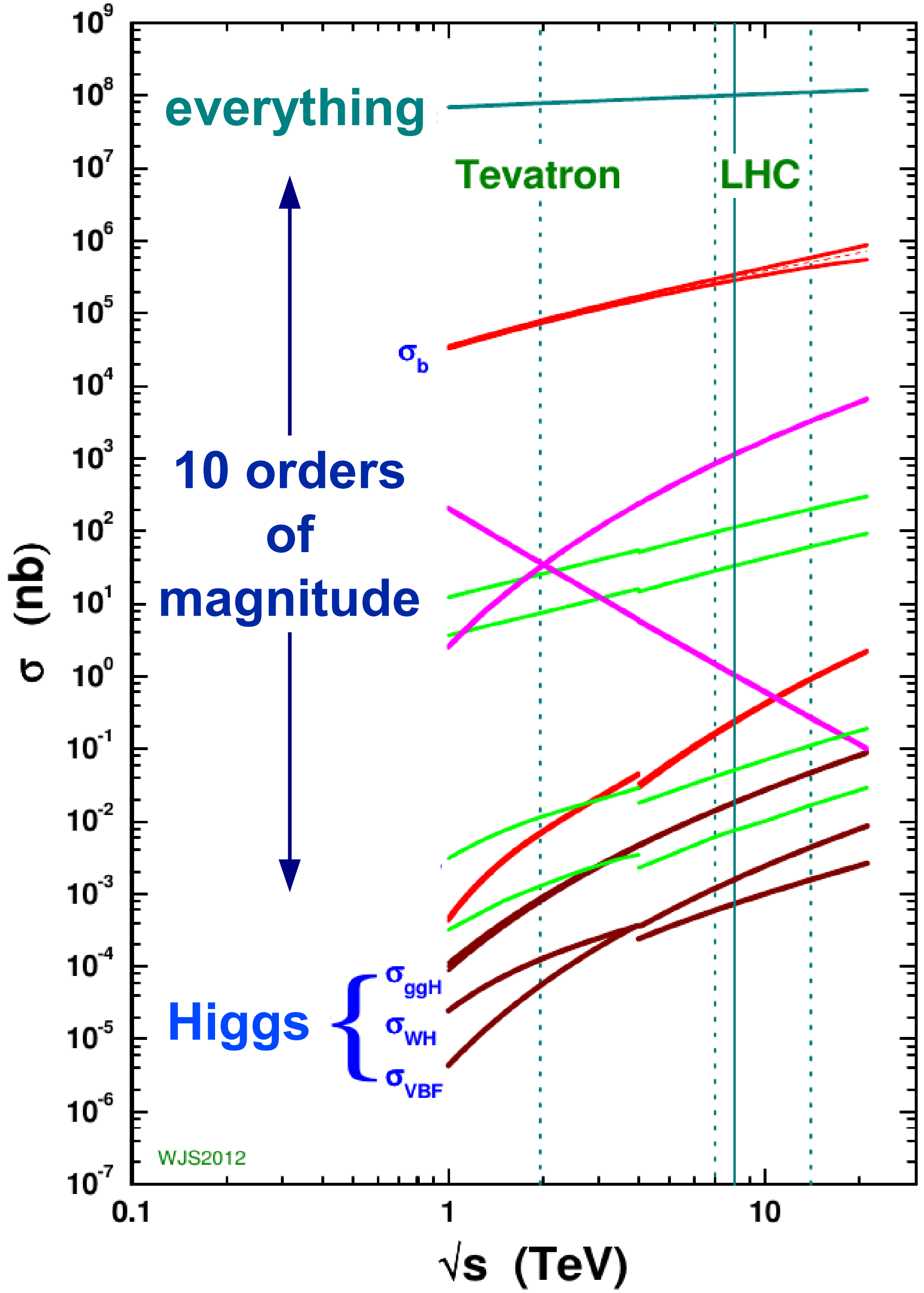
Data Volume

But: we are interested in studying rare processes !

e.g. On average about one Higgs boson / sec

Some more events are useful as “today's calibration channel”

But most events are just “today's background”



plot: <http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>

Data Volume

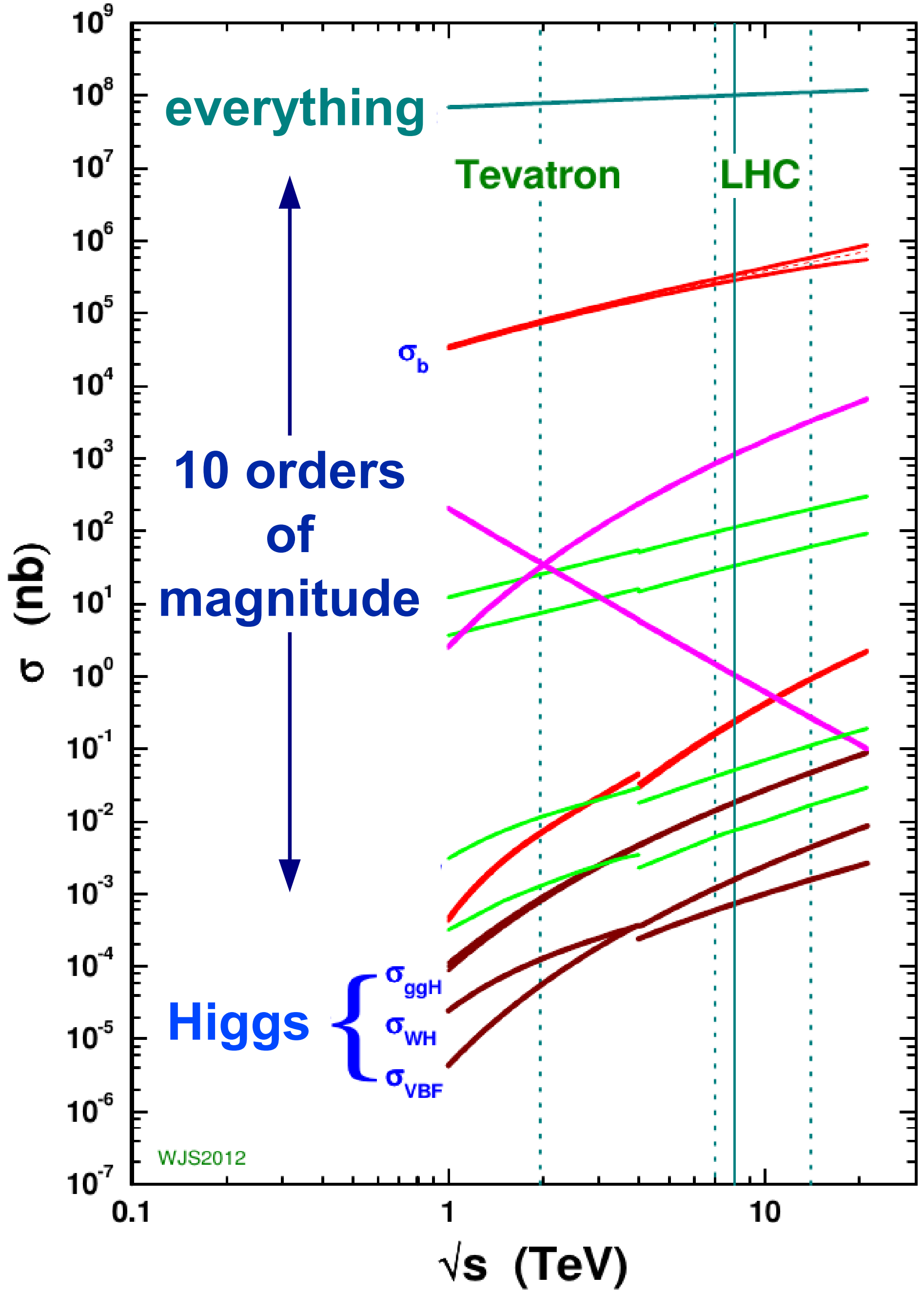
But: we are interested in studying rare processes !

e.g. On average about one Higgs boson / sec

Some more events are useful as “today's calibration channel”

But most events are just “today's background”

→ **No need to store everything !**



plot: <http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>

Data Volume

But: we are interested in studying rare processes !

e.g. On av t

one Higgs b

Some

as “today's

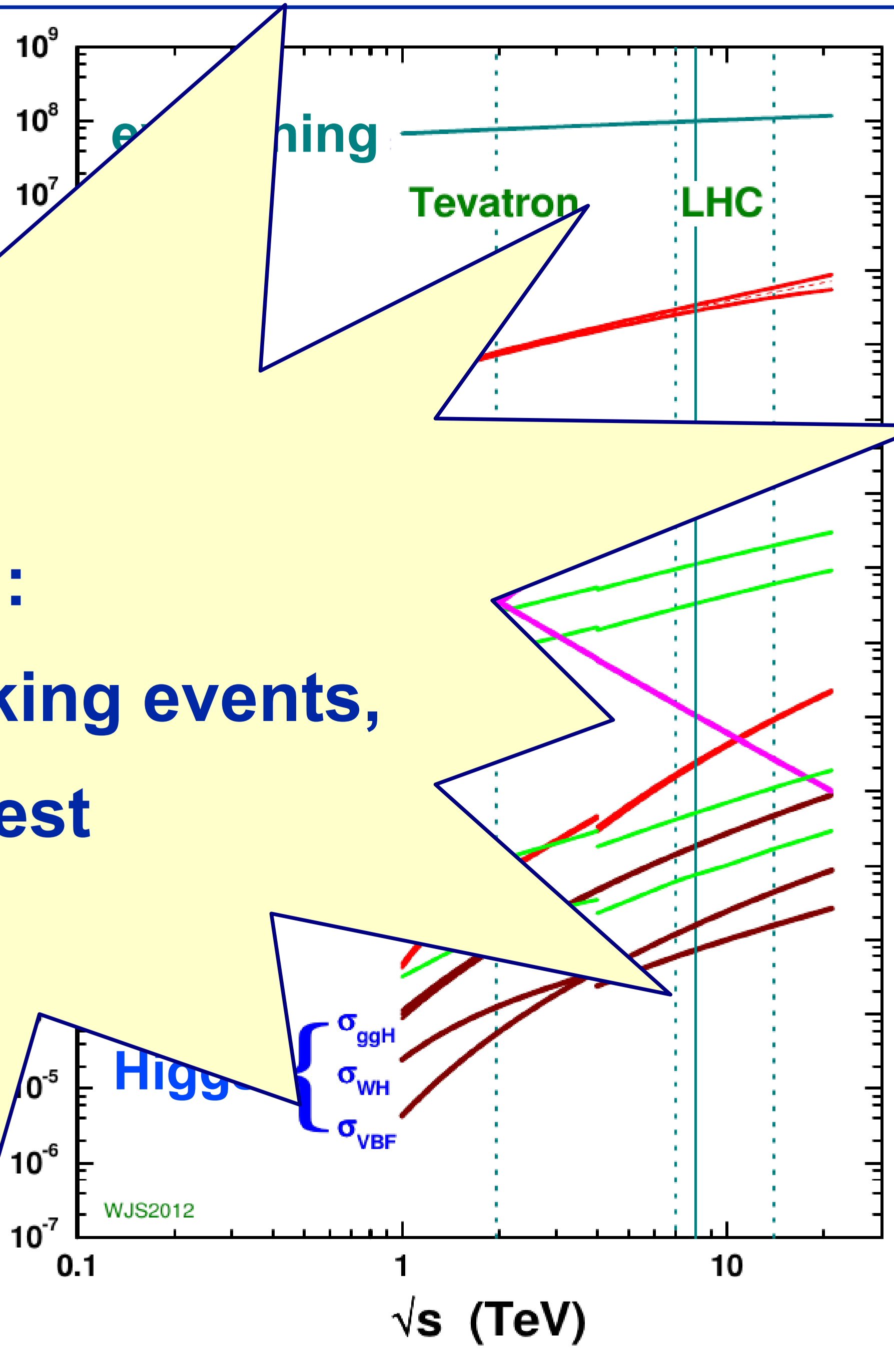
“toc

→ No need to s ery g !

“TRIGGER”:

select interesting-looking events,

discard the rest



plot: <http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>

Trigger Requirements

Needs to have high efficiency for selecting the interesting events

Obviously: want to minimize losses in statistics

Also: losses can cause biases on physics results

Needs to provide high rejection for uninteresting high-rate processes

Determined by how much data you can afford to store

Needs to be flexible

Operating conditions can change

Want to be able to implement new ideas

Trigger Requirements

Needs to be fast

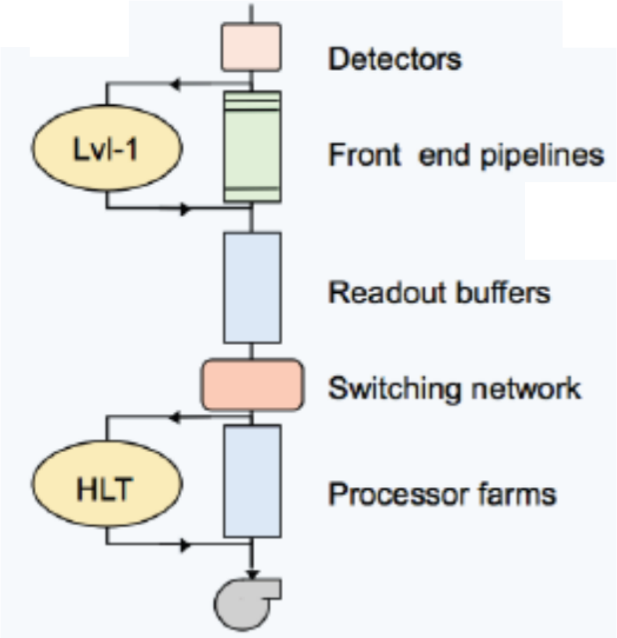
Data need to be stored temporarily while the decision is made:
the more time the algorithm needs to come to a decision,
the more events need to be stored in parallel

→ **cost of storage space**

Events come in at a constant rate of 40 MHz:
the more time the algorithm needs to come to a decision,
the more copies of the algorithm have to be executed in parallel

→ **cost of computing**

Trigger Levels



Low-level triggers:

Full input rate → fast, simple, crude decisions

Usually implemented in custom-made hardware

Higher-level triggers:

Reduced input rate → more time for more sophisticated decisions

Usually implemented in software running on large CPU farms

Trigger Levels

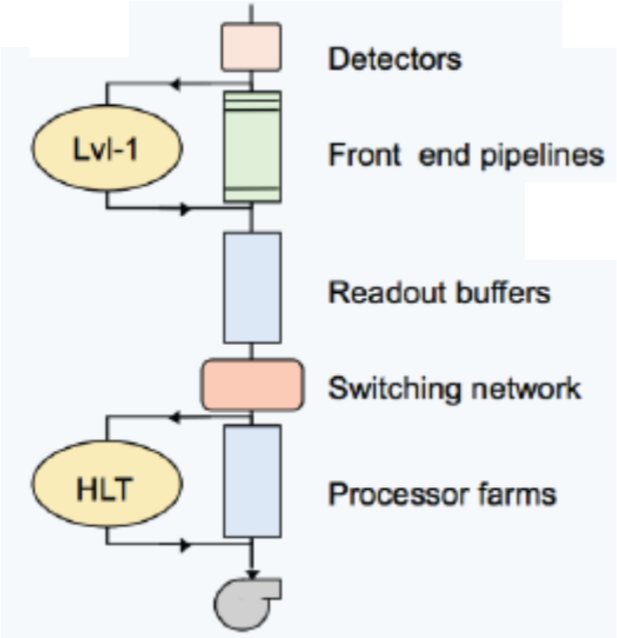
40 MHz

4 µs

1 MHz | 100 kHz

~ 100 ms

12 kHz | 1 kHz



example:
LHCb | CMS

Low-level triggers:

Full input rate → fast, simple, crude decisions

Usually implemented in custom-made hardware

Higher-level triggers:

Reduced input rate → more time for more sophisticated decisions

Usually implemented in software running on large CPU farms

Low-Level Triggers

Need to be fast and simple

The particles we are interest in studying are usually heavy

ATLAS/CMS: Higgs boson 125 GeV = 125 × the mass of a proton

LHCb: particles containing *b* quarks ≈ 5 GeV = 5 × the mass of a proton

→ **Their decay products tend to have large energy and momenta**

Typical signatures to look for:

Large energy deposits in the calorimeters

Muons with large momentum

Low-Level Triggers

Need to be fast and simple

The particles we are interest in studying are usually heavy

ATLAS/CMS: Higgs boson 125 GeV = 125 × the mass of a proton

LHCb: particles containing *b* quarks ≈ 5 GeV = 5 × the mass of a proton

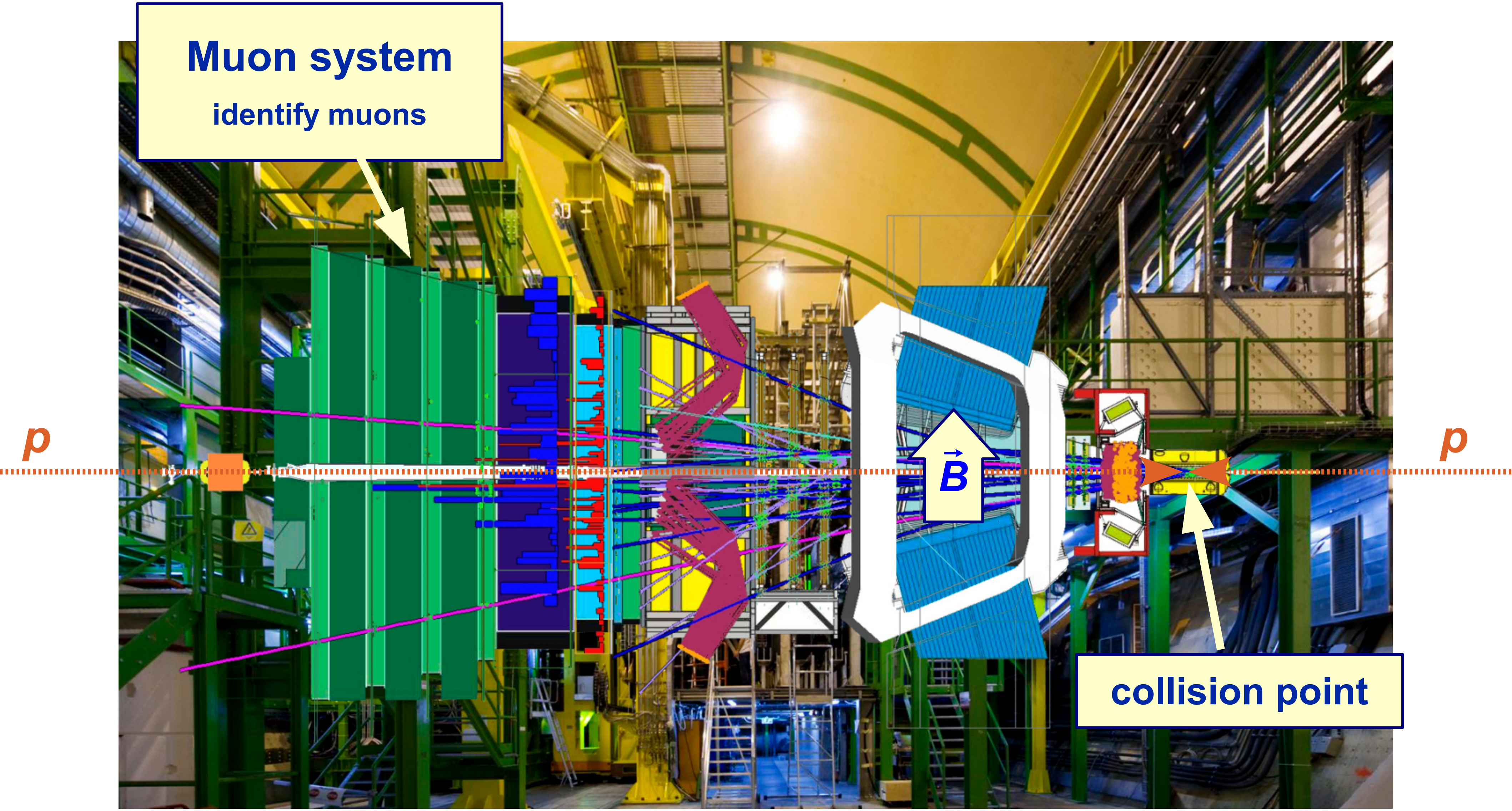
→ **Their decay products tend to have large energy and momenta**

Typical signatures to look for:

Large energy deposits in the calorimeters

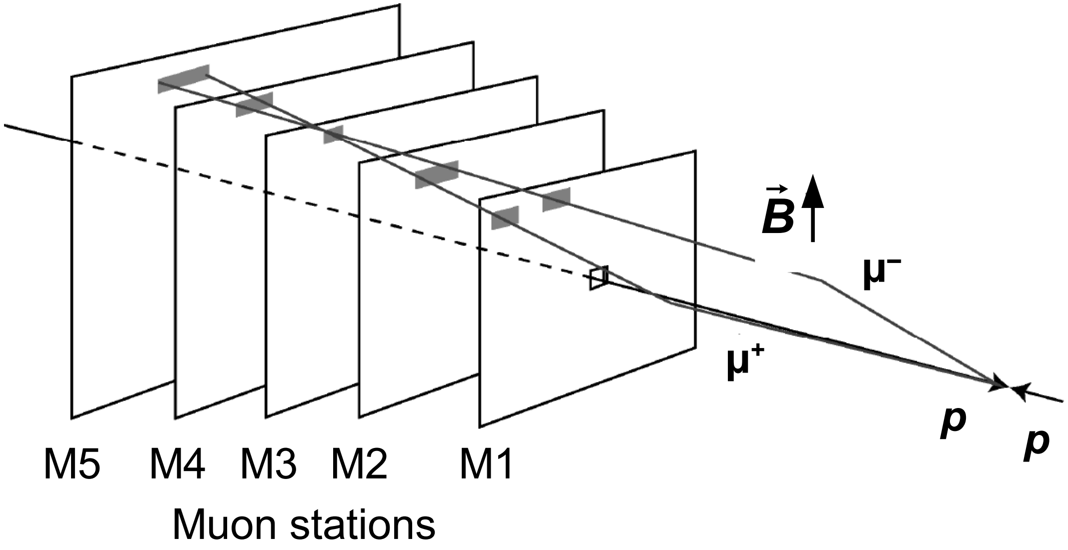
Muons with large momentum

LHCb “L0 Muon” Trigger



Muons are easy to identify: put lots of material, whatever you see in a detector behind this material must be a muon

LHCb "L0 Muon" Trigger



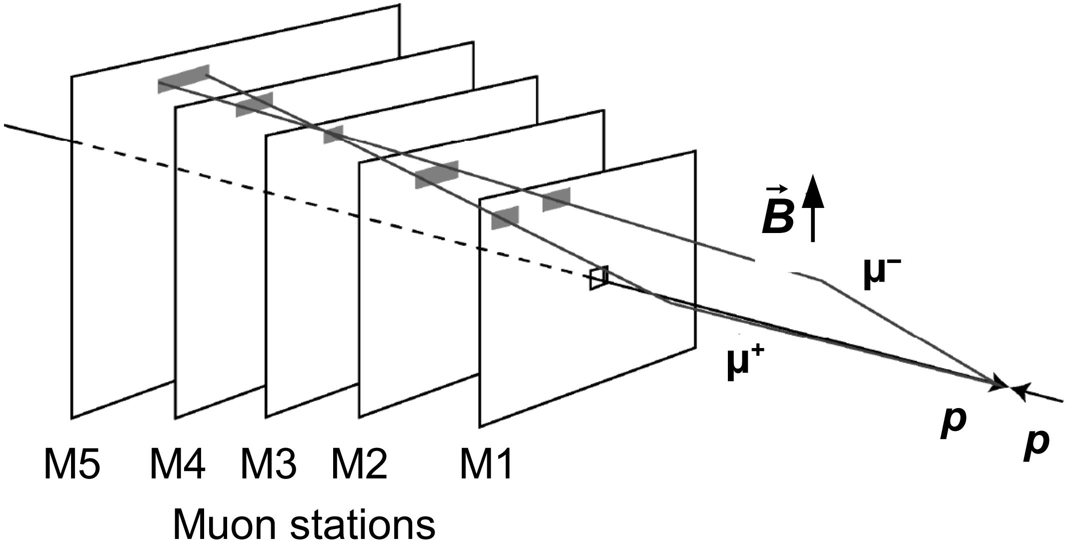
LHCb muon system:

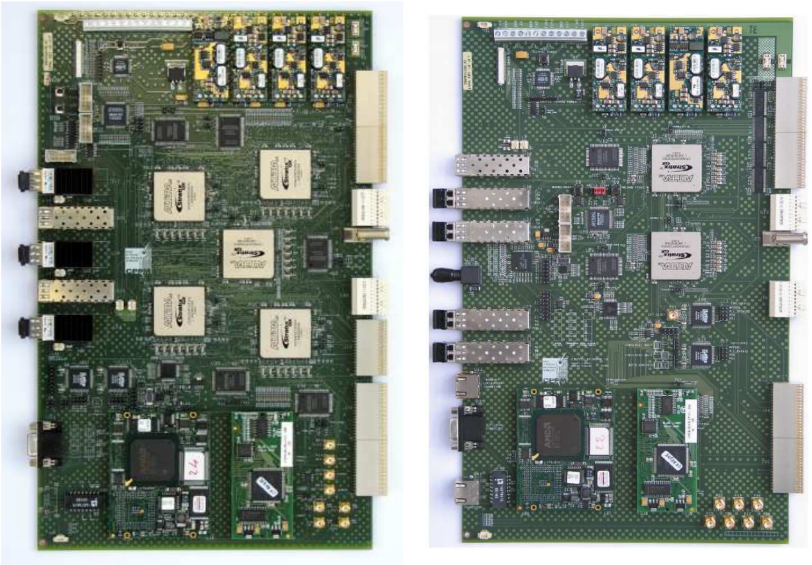
Five detection layers, separated by 80 cm thick walls of iron

First-level muon trigger:

Require coincidence of hits in several detection layers,
pointing back to the *pp* collision point
→ Small deviation in magnetic field → Muon with high momentum

LHCb “L0 Muon” Trigger





Algorithm implemented in custom-designed electronics boards, using FPGA and RAM chips and high-speed data links

Start with a hit in station M3

Extrapolate straight line to M2, M4, M5

Define search windows around the extrapolated position

If matching hits are found, look up the corresponding momentum
(stored in large look-up tables in RAM chips)

Higher-Level Trigger

Usually a simplified version of the offline reconstruction software, running on a large computer farm, close to the experiment

Higher-Level Trigger

Usually a simplified version of the offline reconstruction software, running on a large computer farm, close to the experiment

LHCb: 58'000 CPU cores

40'000 reconstruction jobs running in parallel

400 different trigger selections for different physics analyses

LHCb has so far published 370 physics papers, with many more to come

Higher-Level Trigger

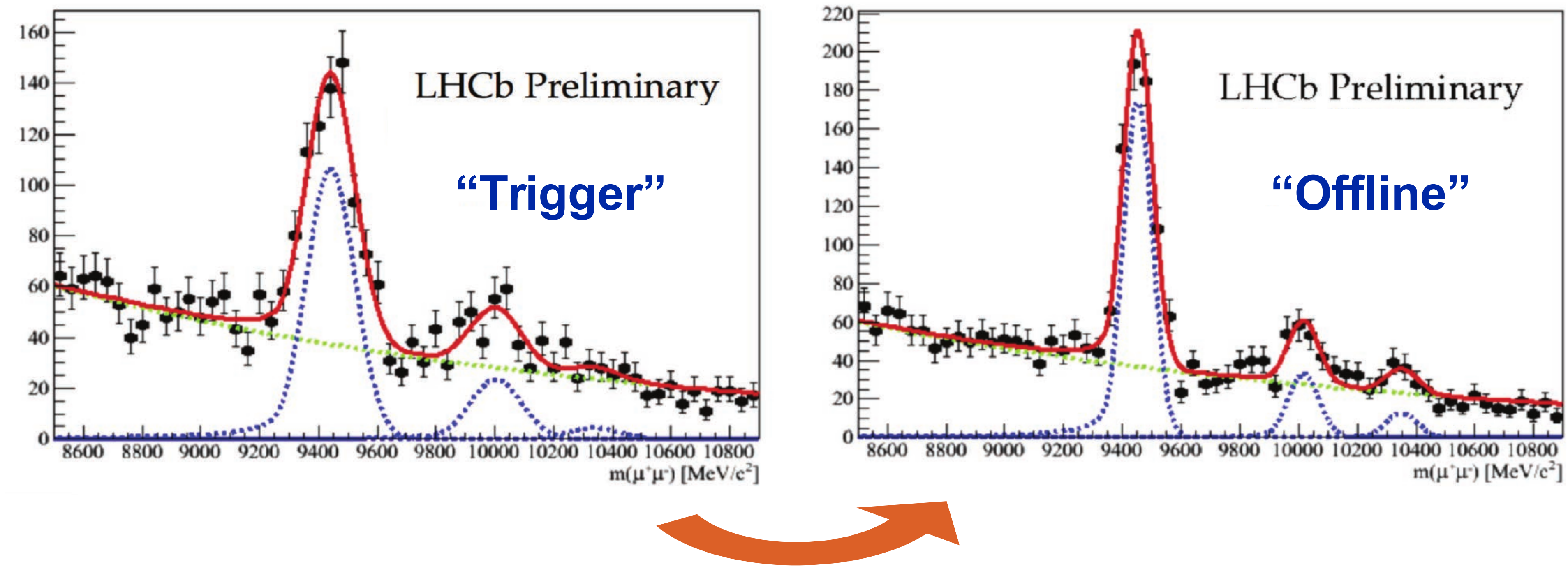
Usually a simplified version of the offline reconstruction software, running on a large computer farm, close to the experiment

Simplifications necessary to meet CPU time constraints, but result in poorer resolution

Higher-Level Trigger

Usually a simplified version of the offline reconstruction software, running on a large computer farm, close to the experiment

Simplifications necessary to meet CPU time constraints, but result in poorer resolution



≈ factor 2 improvement due to better calibration

Higher-Level Trigger

Usually a simplified version o the offline nstruction software, running on a large comp rm the expe ment

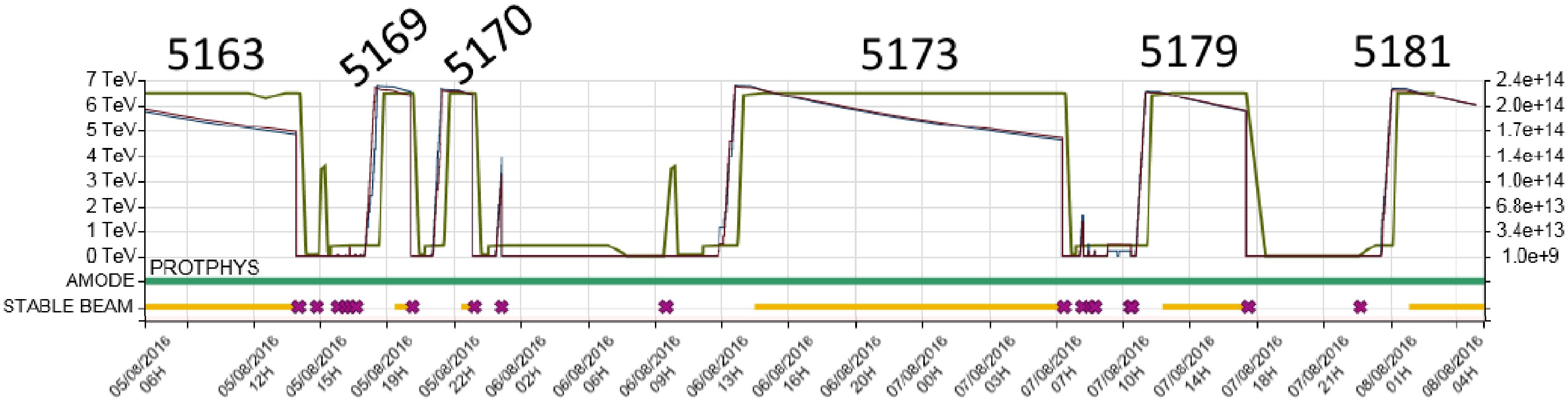
Simplific necess m traints,

Brilliant NEW IDEA !

implemented in LHCb for 2015 / 2016 data taking:

≈ fa or 2 improvement due to better calibration

LHCb High-Level Trigger

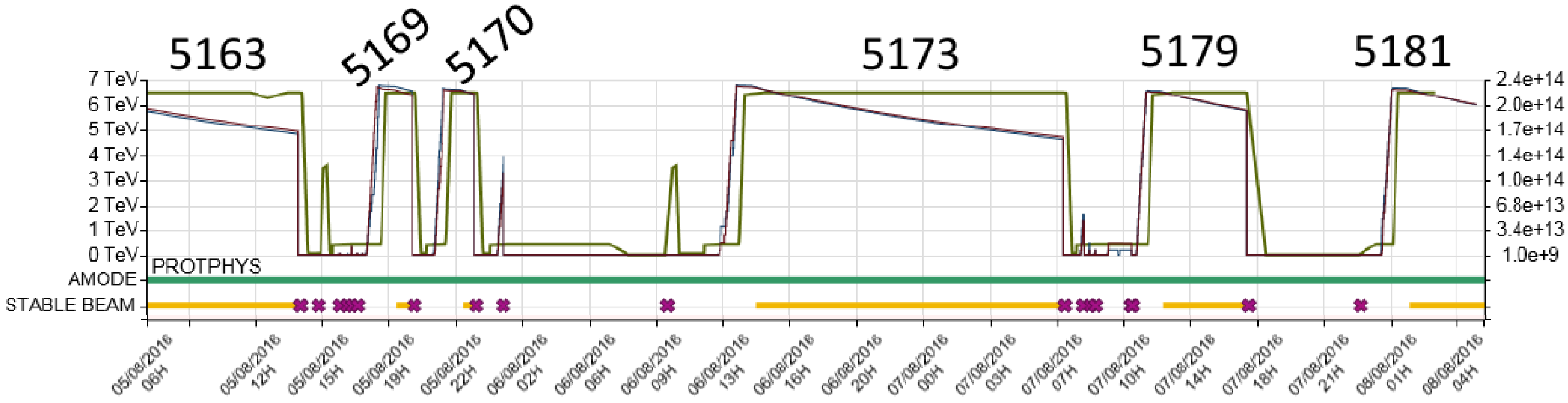


≈ 3 days in August 2016

LHC is colliding protons only ≈ 35 % of the time

Inject fresh proton beams, maintenance work, technical problems, …

LHCb High-Level Trigger



≈ 3 days in August 2016

LHC is colliding protons only ≈ 35 % of the time

Inject fresh proton beams, maintenance work, technical problems, …

Split High-Level Trigger into two parts

Buffer events that pass a first selection on local disks

Use CPU farm to perform final selection
when there are no collisions and no new data is coming in

Calibration with “offline quality” before final selection

10'000 TByte disk space
→ buffer data up to a week

Overview

Part I

What are we trying to measure

What do our experiments look like

Part II

“Online” event selection

Part III

“Offline” analysis

Offline Analysis

After trigger selections, experiments are saving of the order of 1'000 events per second for offline analysis

Reconstructing one event takes about 10-20 sec

Might need to run the reconstruction more than once,
if improved reconstruction algorithms become available

In addition: need to generate huge numbers of simulated events to study reconstruction and selection efficiencies etc.

Generating one simulated event can take up to 15 minutes

In addition: physics analyses

usually fast, but many users, many analysis jobs

→ **need tens of thousands of processors**

Offline Analysis

After trigger selections, experiments are saving of the order of 1'000 events per second for offline analysis

≈ 10^{10} events per year × 1.5 MB per event

→ need storage space for ≈ 15'000 TB of data per experiment per year

Offline Analysis

After trigger selections, experiments ar ving of the order of 1'000 events per ond fo e analysis

^{10} events er e

→ nee

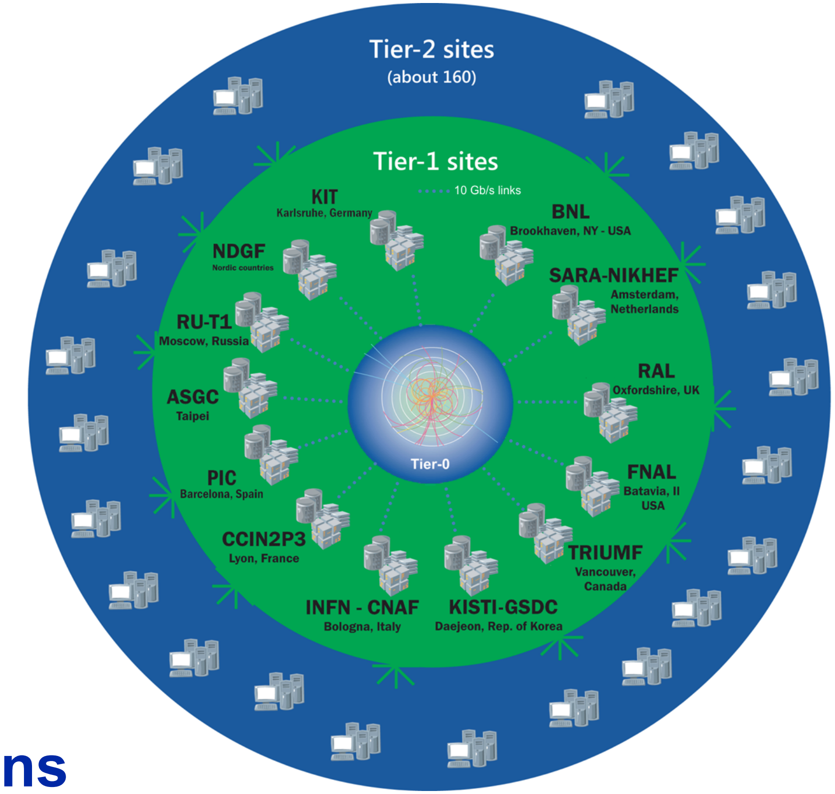
Solution: the LHC Computing Grid

170 computing centres in 42 countries, connected by high-speed data links

2 million jobs run every day

Offline Analysis

Hierarchical structure



Tier 0: CERN + Wigner centre Budapest

→ first copy of the raw data

→ first pass reconstruction

→ reprocessing during LHC down-times

Tier 1: Fourteen large computer centres

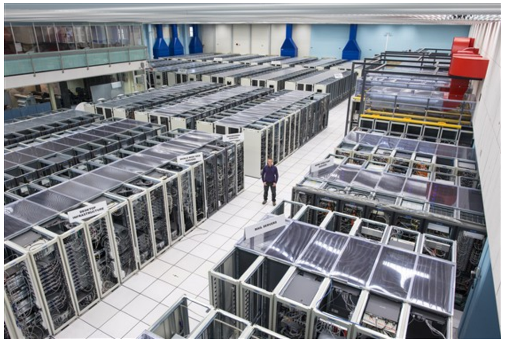
→ fraction of raw and reconstructed data

→ large-scale reprocessing

Tier 2: Large clusters at Scientific Institutions

→ specific analysis tasks

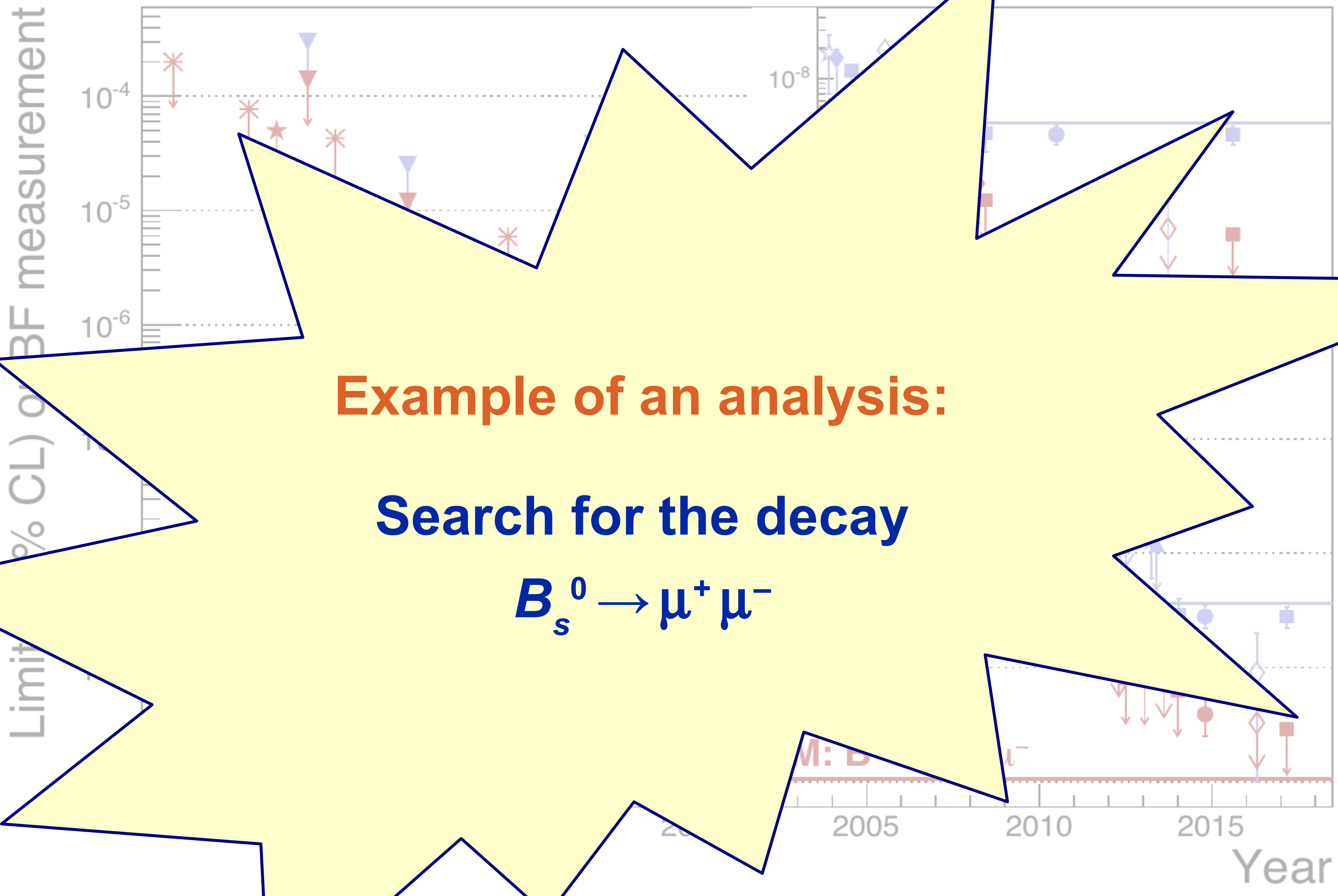
→ production/reconstruction of simulated events



Tier 3: Local clusters in a University Department

→ what the user connects to

Example: $B_s^0 \rightarrow \mu^+\mu^-$

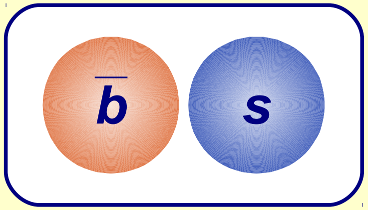


Example of an analysis:

Search for the decay

$B_s^0 \rightarrow \mu^+\mu^-$

Example: $B_s^0 \rightarrow \mu^+ \mu^-$

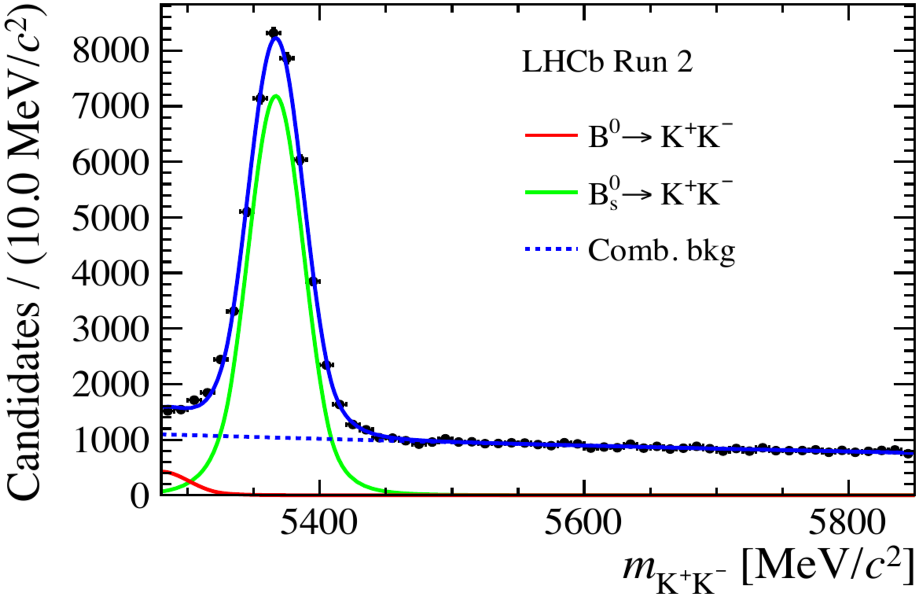


mass ≈ 5366 MeV/c^2

mean lifetime ≈ 1.470 ps

≈ 5 × the mass of a proton

travels ≈ 1 cm in LHCb before it decays



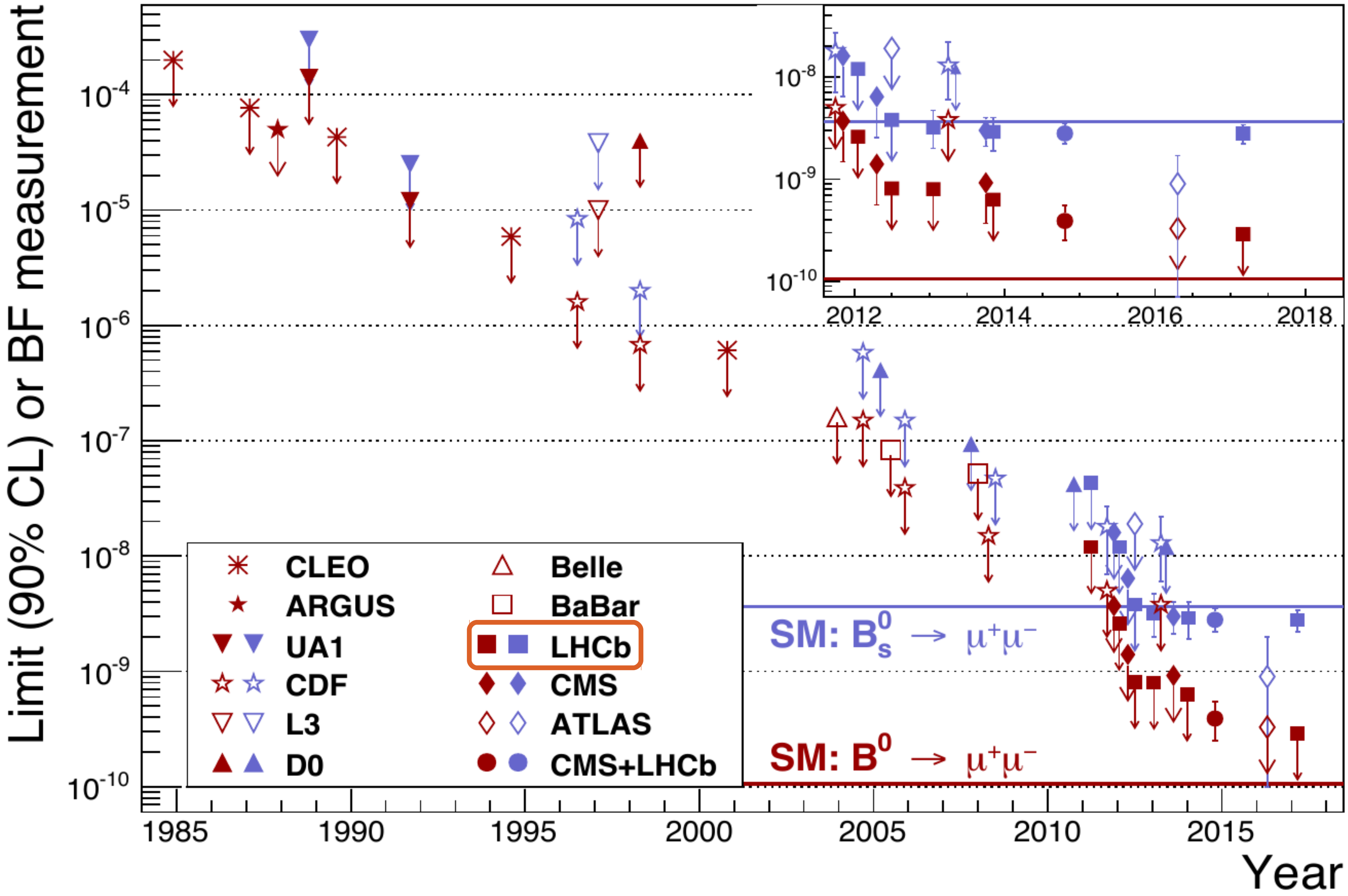
Decay into $\mu^+ \mu^-$ strongly suppressed in the Standard Model

Predicted "Branching Fraction" ≈ 3.5×10^{-9}

(i.e. only about three out of a billion B_s^0 mesons should decay into $\mu^+ \mu^-$)

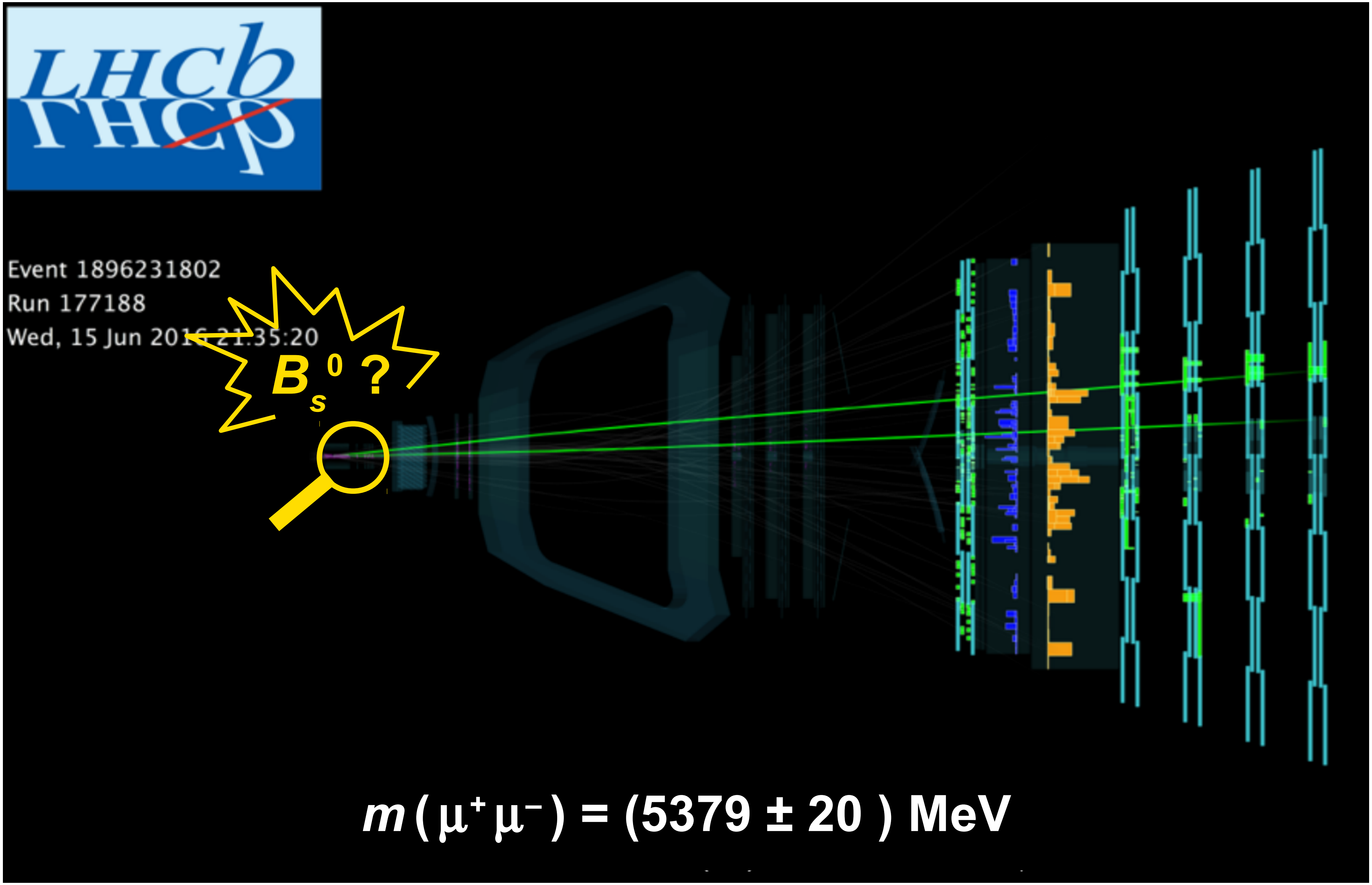
But other theories (e.g. SuperSymmetry) predict this to be much larger

Example: $B_s^0 \rightarrow \mu^+ \mu^-$



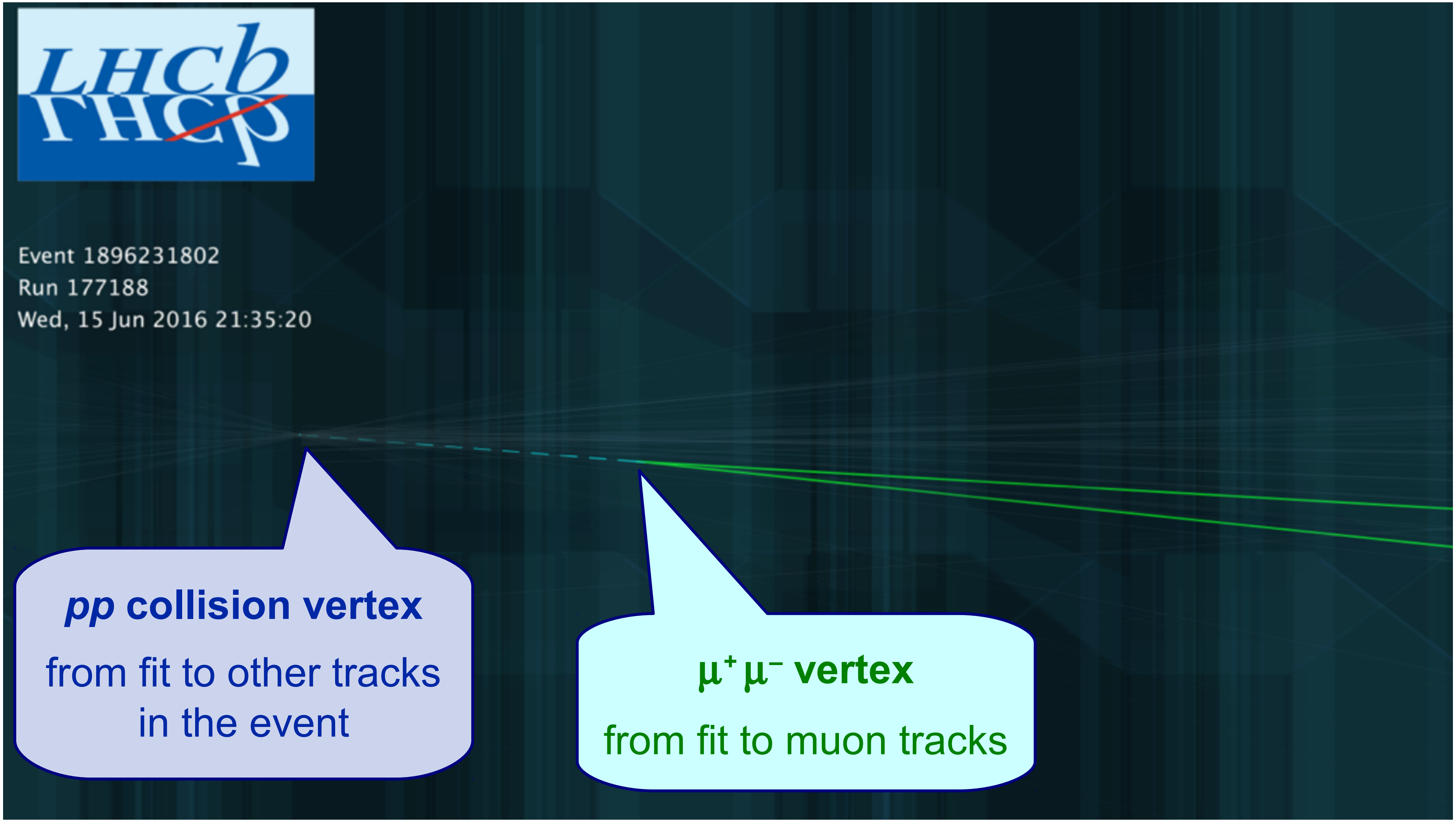
(also $B^0 \rightarrow \mu^+ \mu^-$, even more suppressed)

Example: $B_s^0 \rightarrow \mu^+\mu^-$



$m(\mu^+\mu^-) = (5379 \pm 20)$ MeV

Example: $B_s^0 \rightarrow \mu^+ \mu^-$



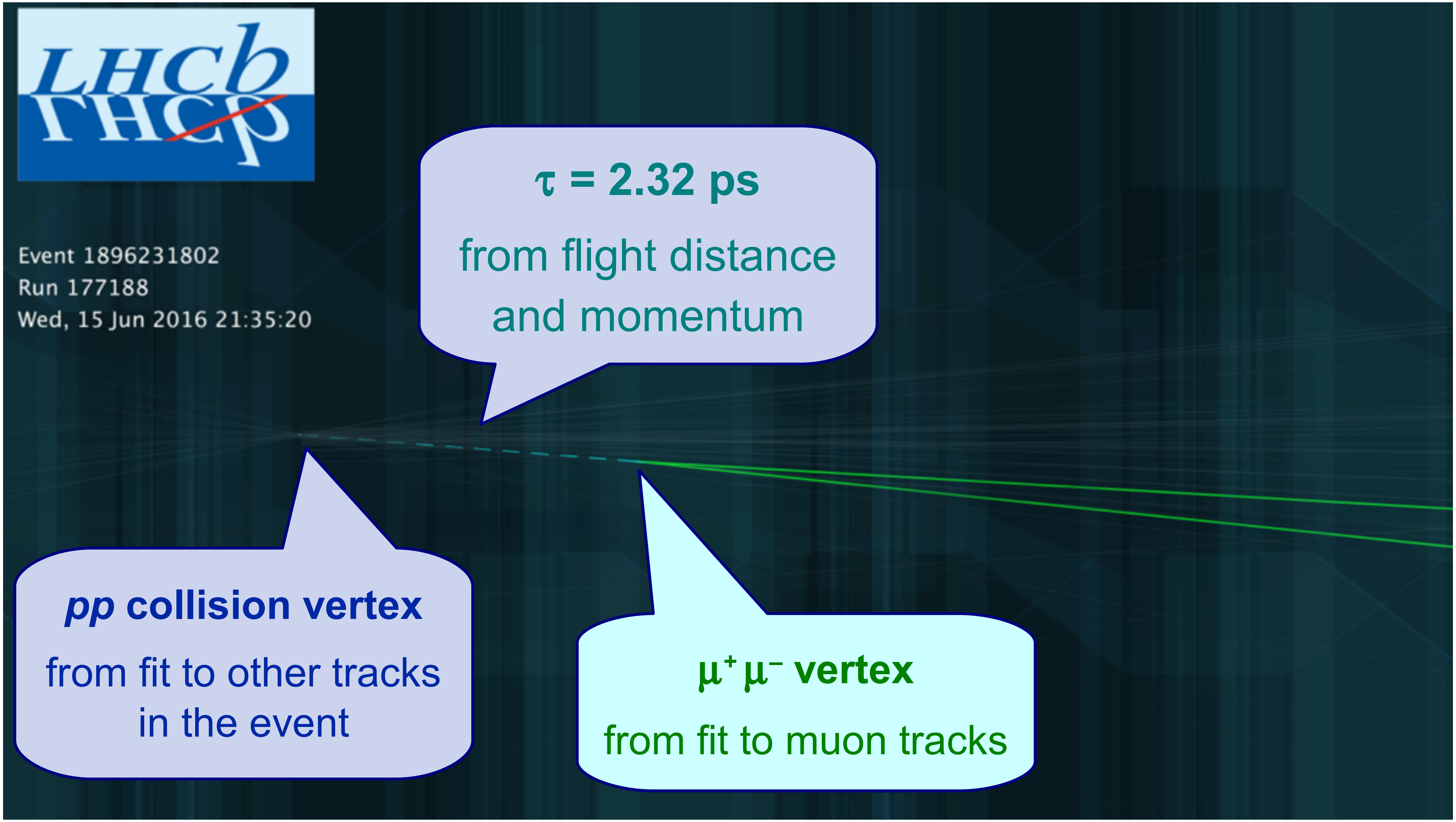
***pp* collision vertex**

from fit to other tracks in the event

$\mu^+\mu^-$ vertex

from fit to muon tracks

Example: $B_s^0 \rightarrow \mu^+ \mu^-$



τ = 2.32 ps

from flight distance and momentum

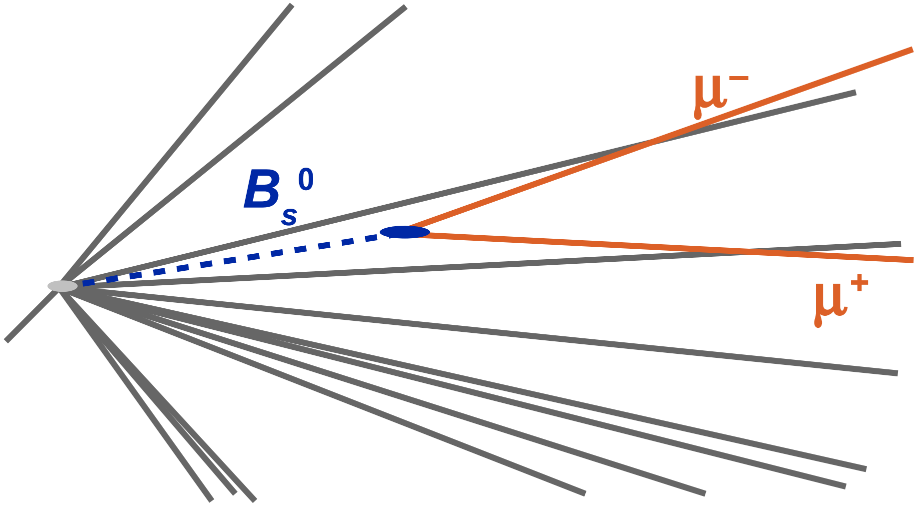
***pp* collision vertex**

from fit to other tracks in the event

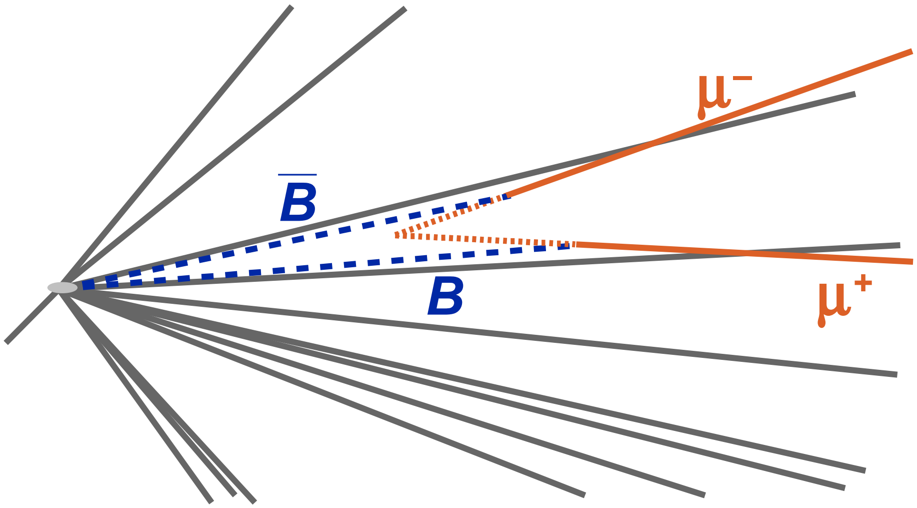
$\mu^+ \mu^-$ vertex

from fit to muon tracks

Example: $B_s^0 \rightarrow \mu^+ \mu^-$



signal



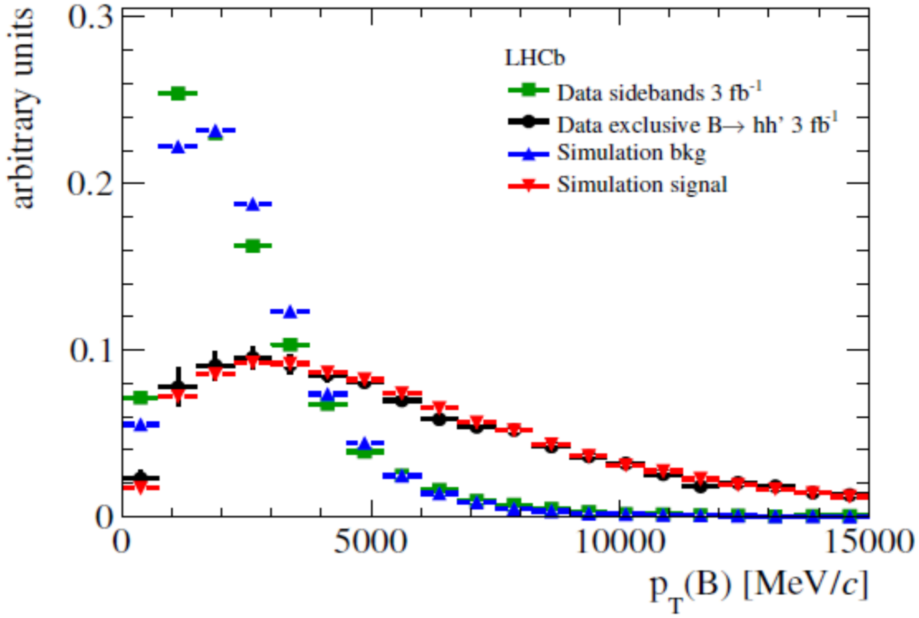
background

Use twelve discriminating variables

to distinguish between signal and background

(but do NOT use $m\,(\mu^+\mu^-)$ yet **)**

Example: $B_s^0 \to \mu^+ \mu^-$



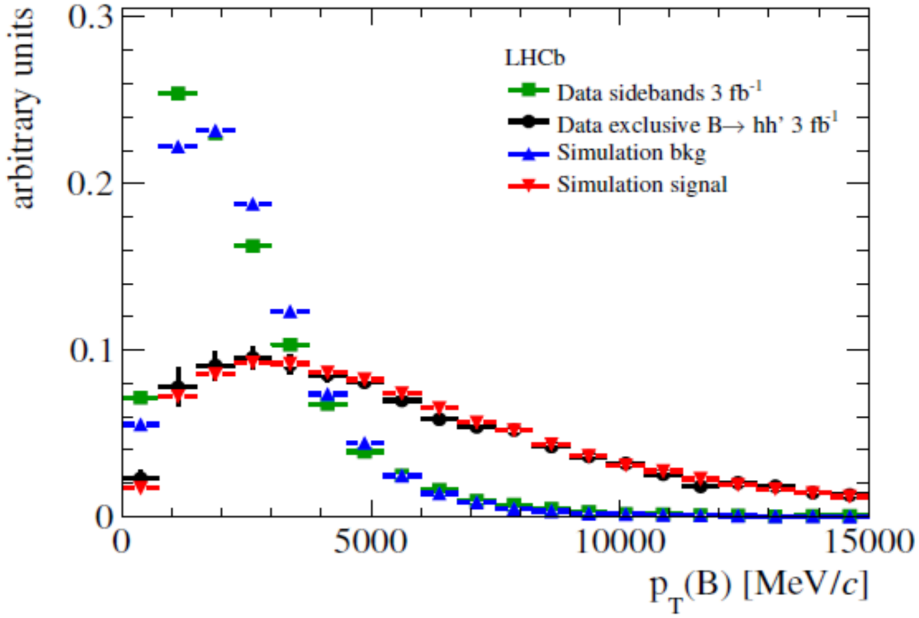
— **signal simulation**

— **background simulation**

— **"signal" in data**
($B_s^0 \to K^+ K^-$ and similar **)**

— **"background" in data**
($m(\mu^+\mu^-) > m(B_s^0)$ **)**

Example: $B_s^0 \to \mu^+\mu^-$



signal simulation

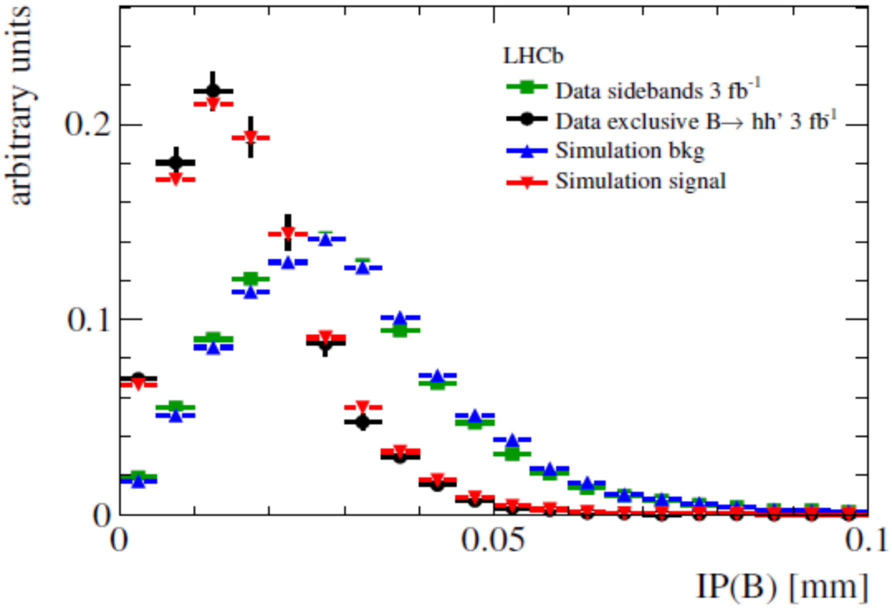
background simulation

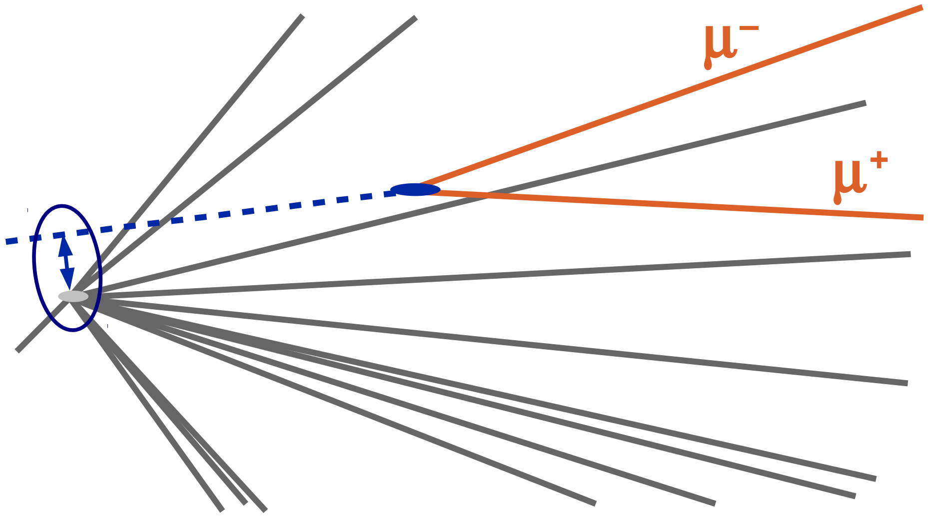
“signal” in data

($B_s^0 \to K^+K^-$ and similar)

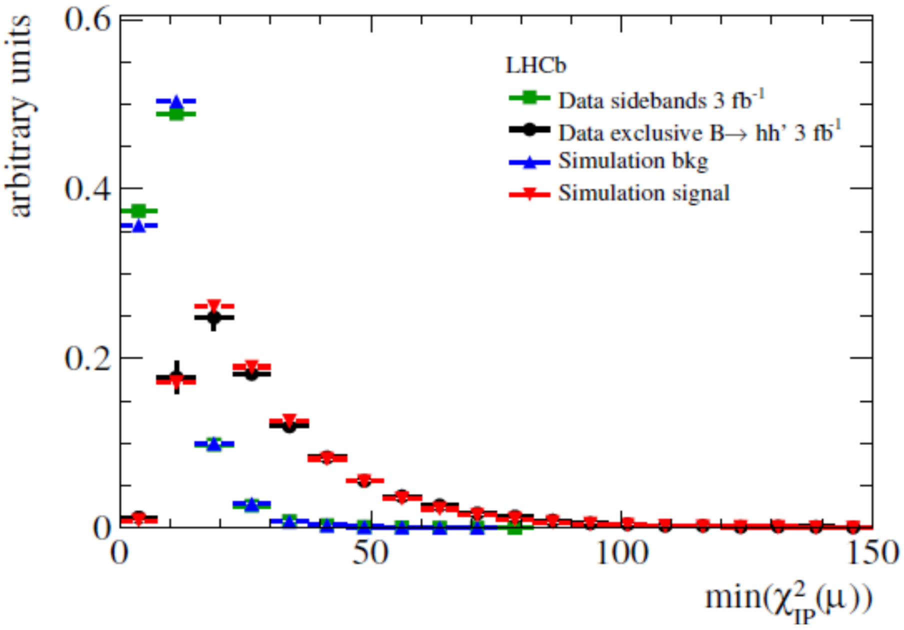
“background” in data

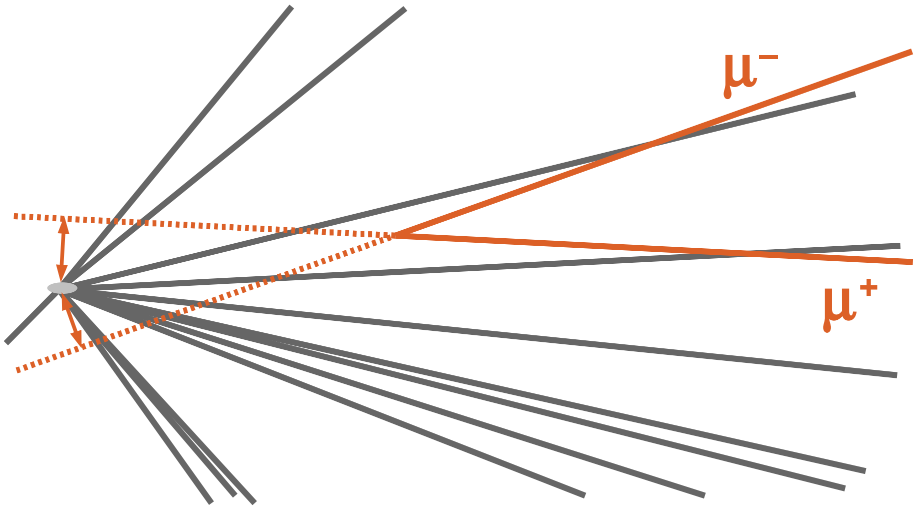
($m(\mu^+\mu^-) > m(B_s^0)$)



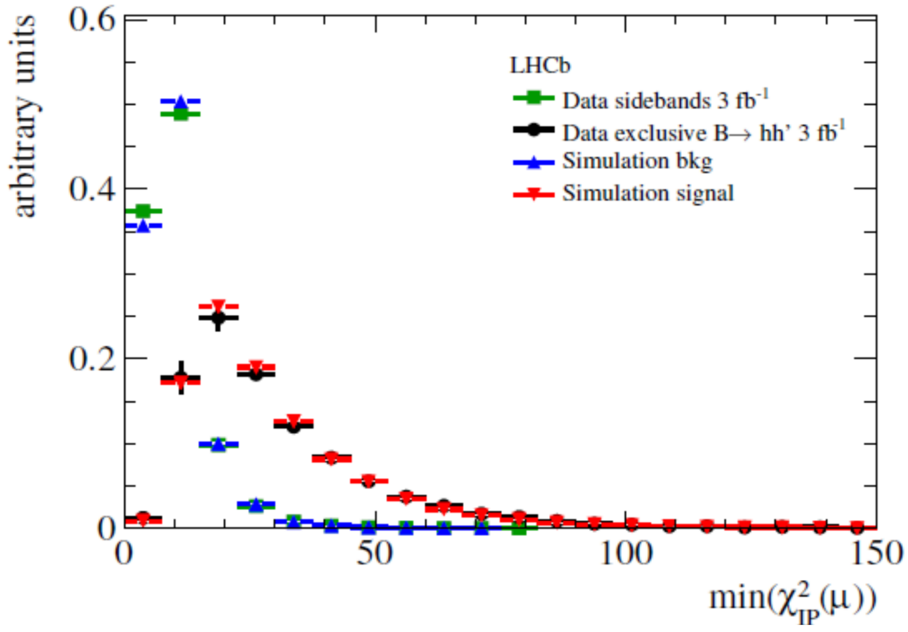


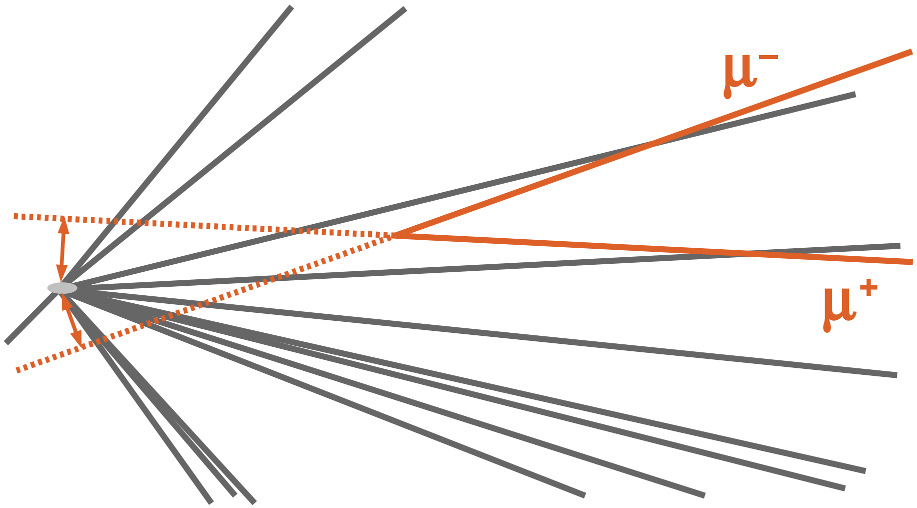
Example: $B_s^0 \rightarrow \mu^+ \mu^-$

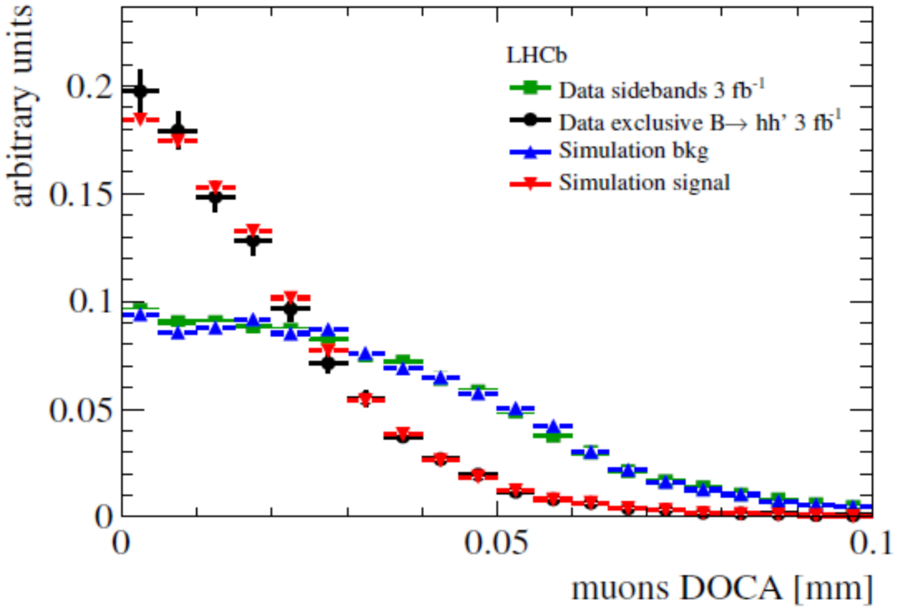


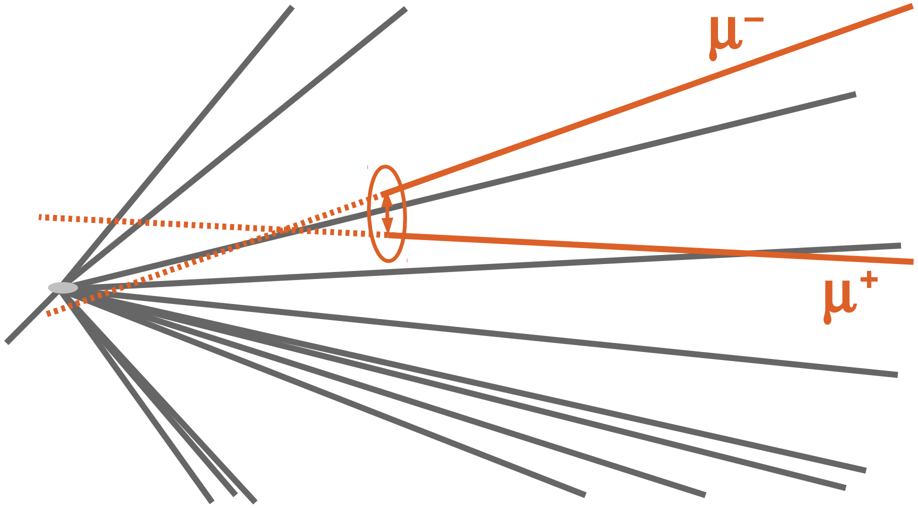


Example: $B_s^0 \to \mu^+ \mu^-$







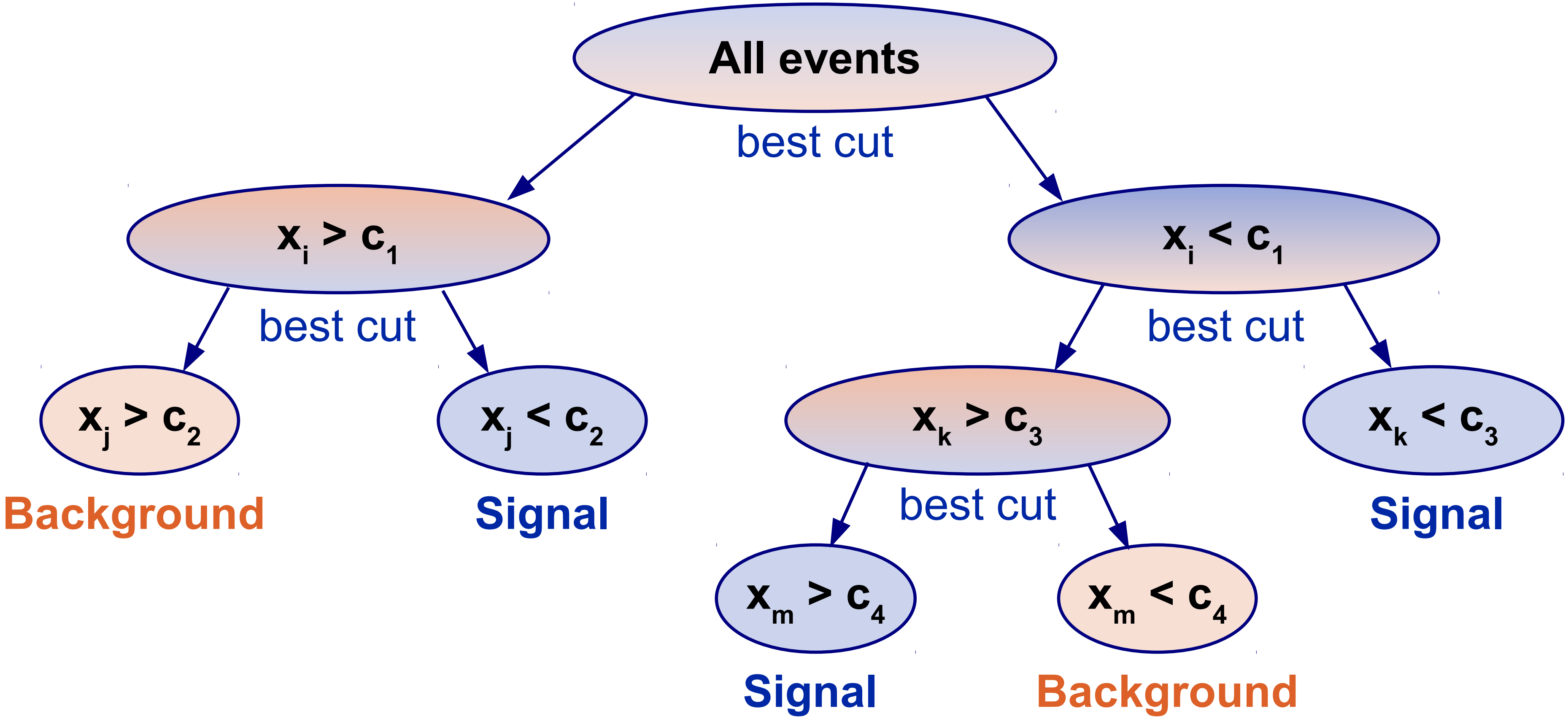


Example: $B_s^0 \rightarrow \mu^+ \mu^-$

For best discrimination:
combine all 12 variables into a single multivariate discriminator

→ Boosted Decision Tree

Decision Tree



All events

best cut

$x_i > c_1$

$x_i < c_1$

best cut

$x_j > c_2$

Background

$x_j < c_2$

Signal

best cut

$x_k > c_3$

$x_k < c_3$

Signal

best cut

$x_m > c_4$

Signal

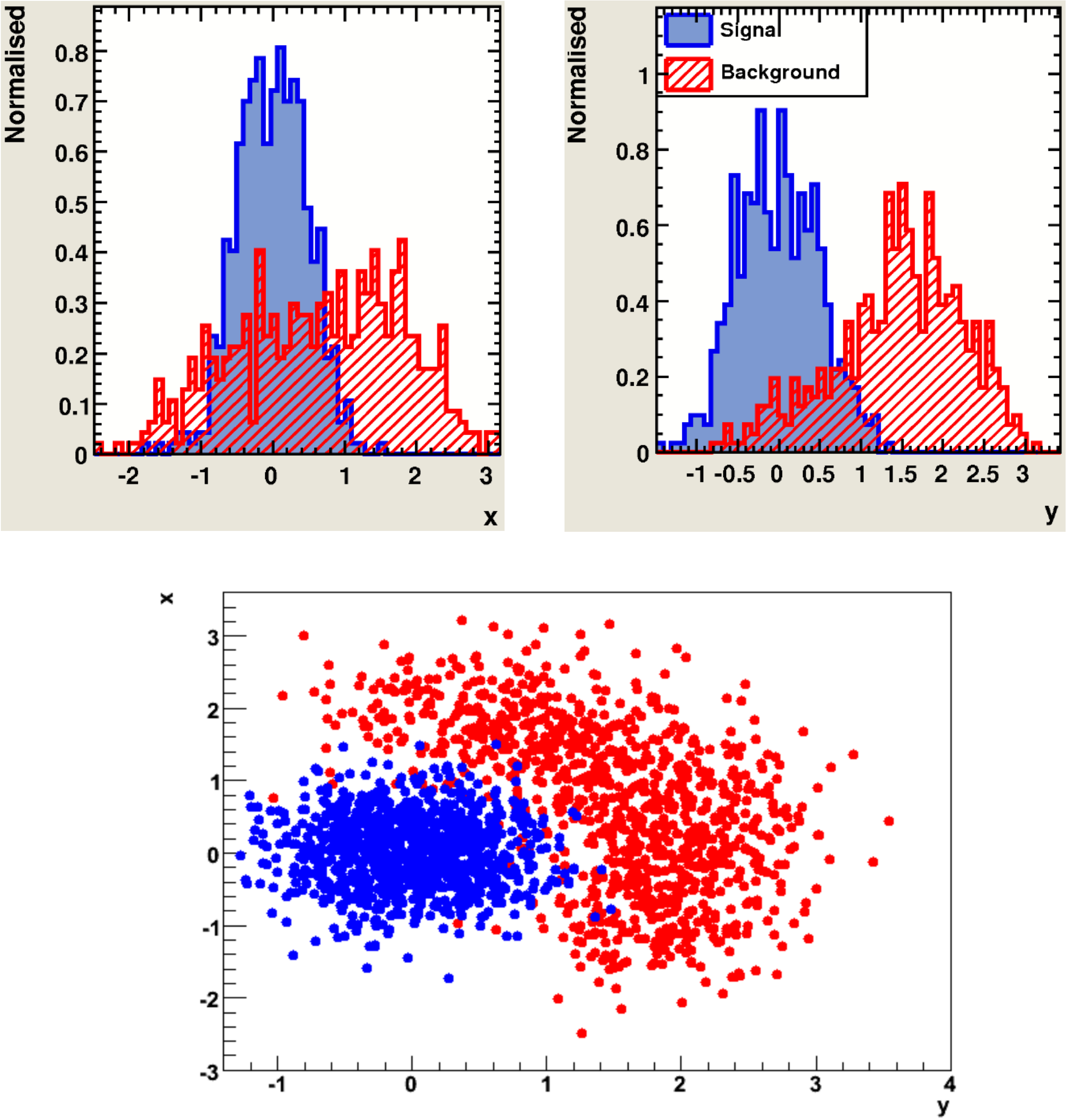
$x_m < c_4$

Background

At each step, determine the most discriminating cut for the given sample

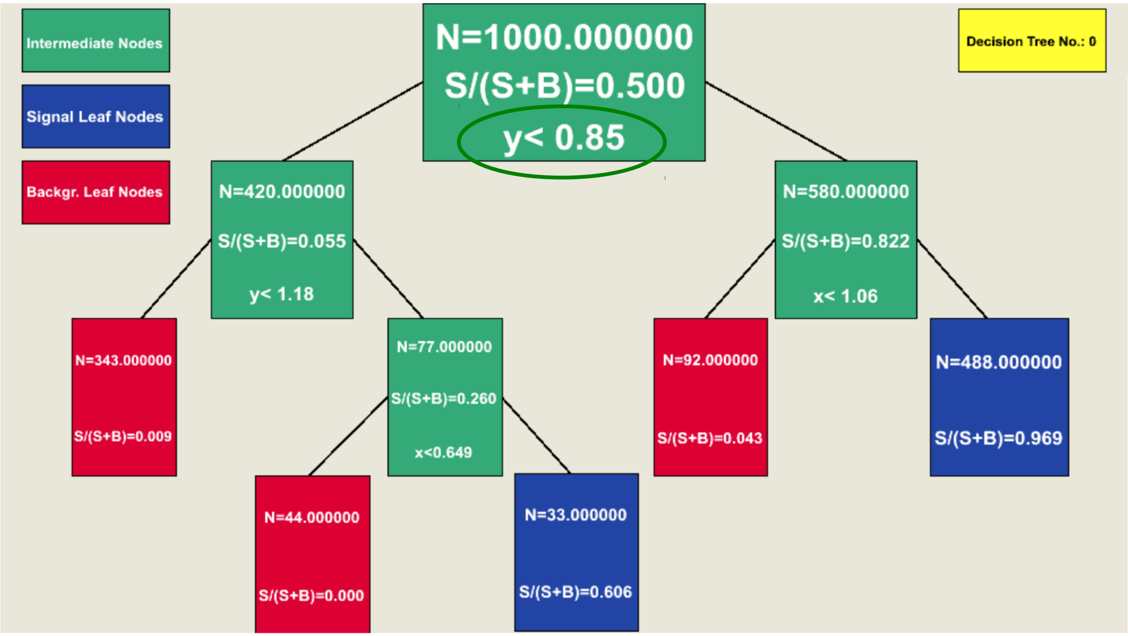
The same variable can be used several times

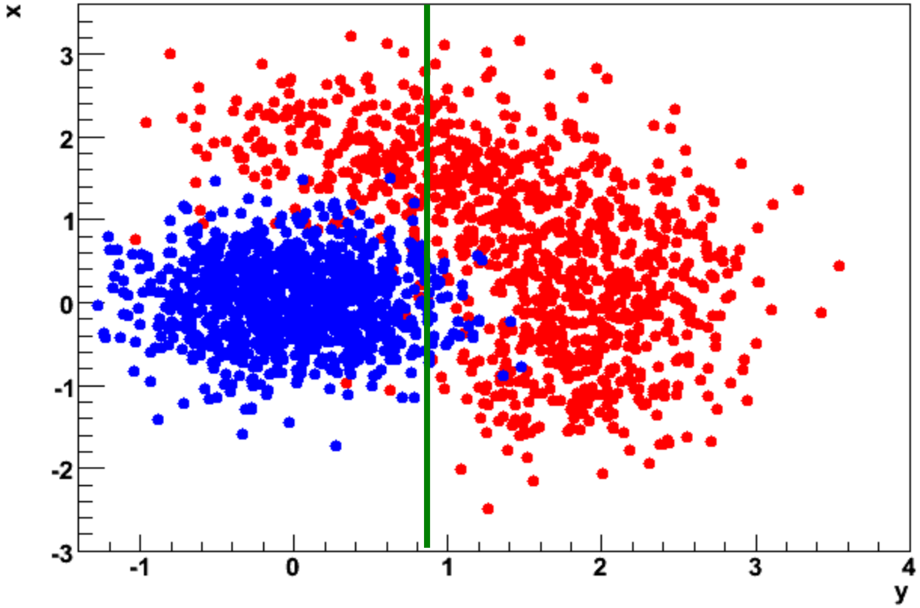
Decision Tree



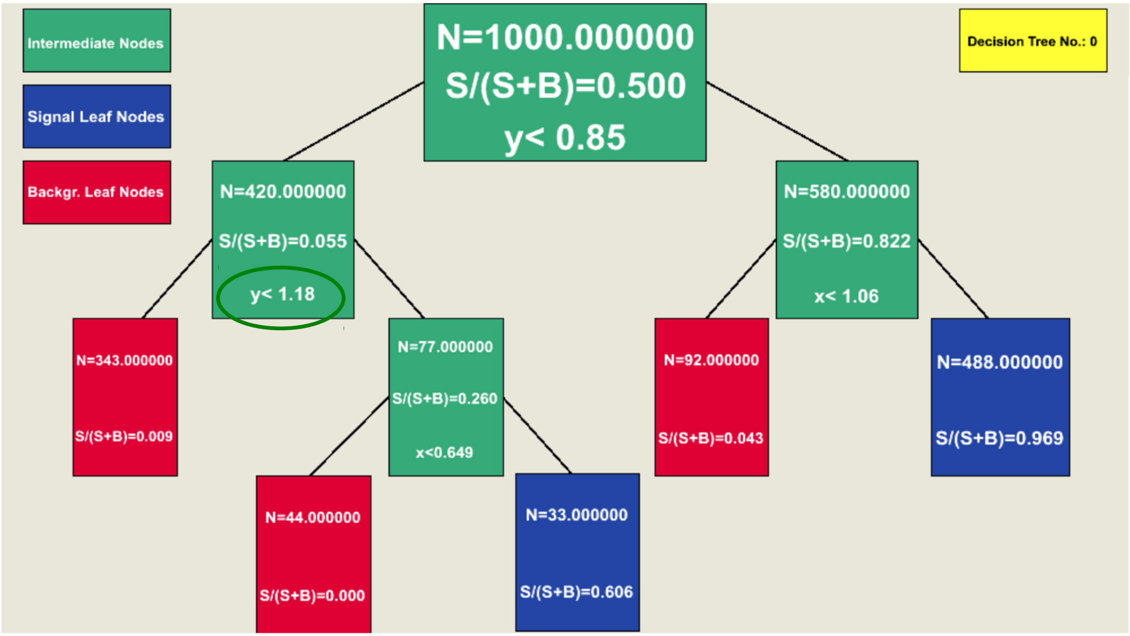
[Yann Coadou (CPPM Marseille), presented at esipap 2014]

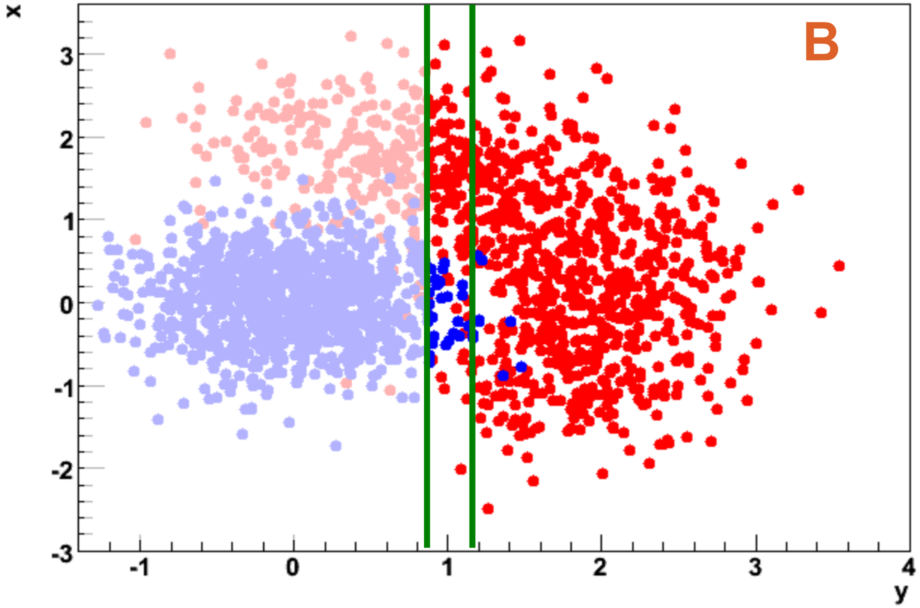
Decision Tree



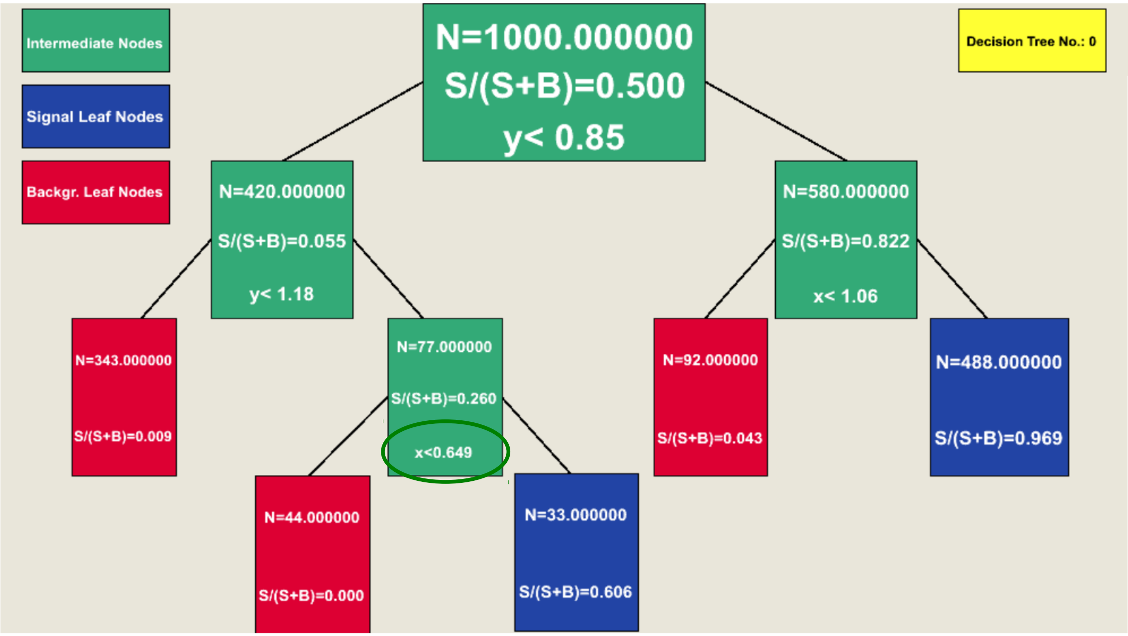


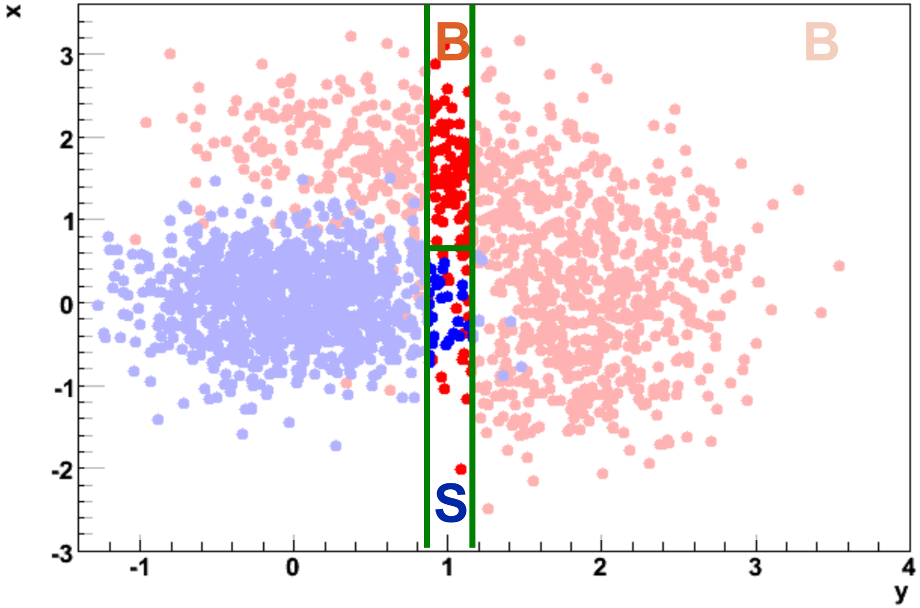
Decision Tree



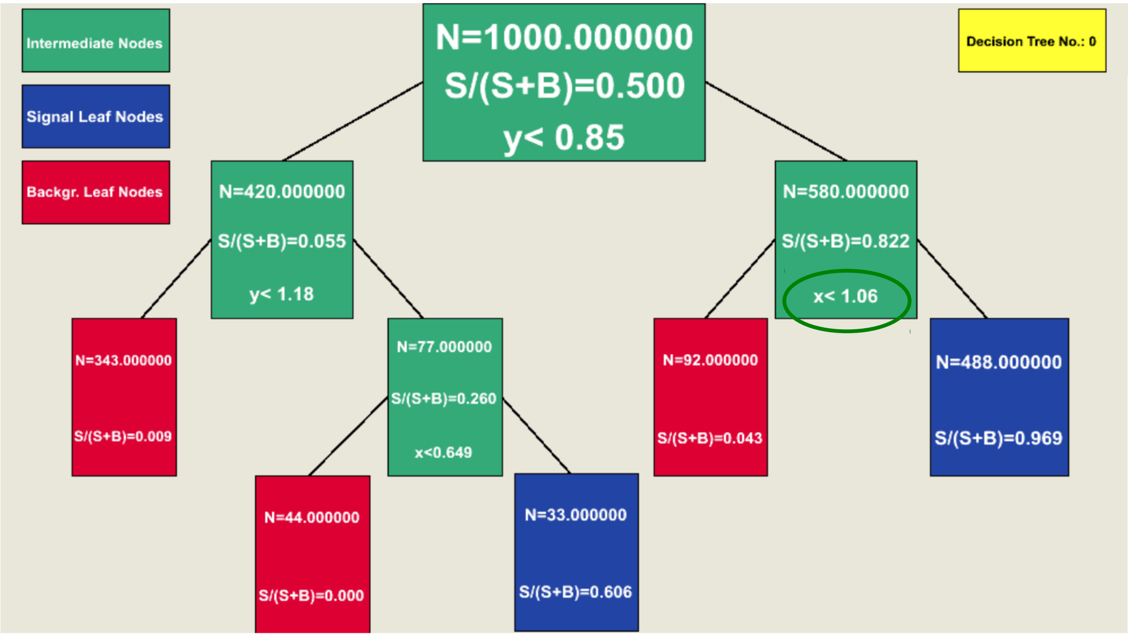


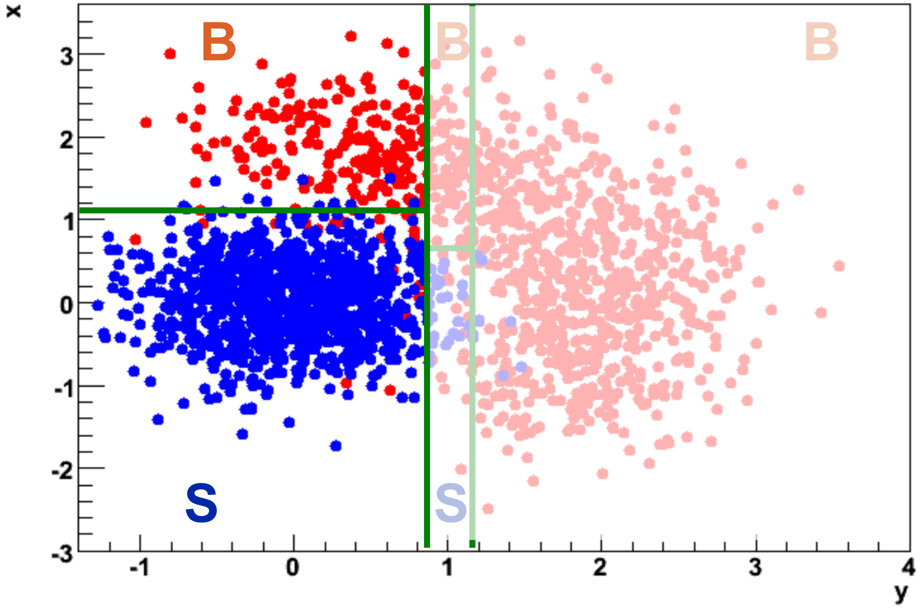
Decision Tree



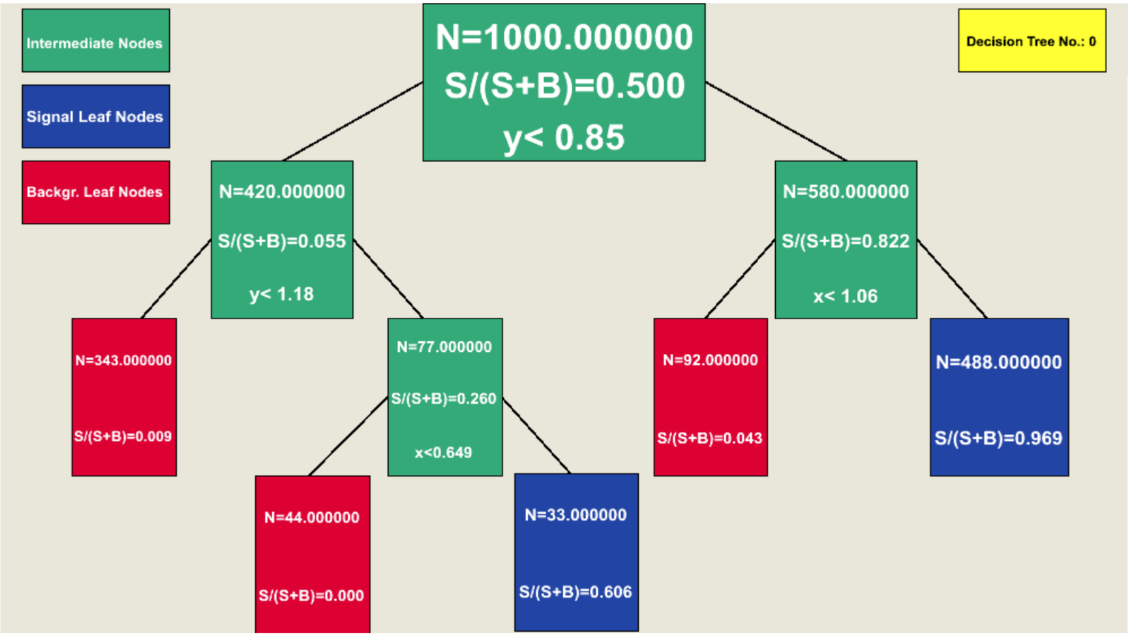


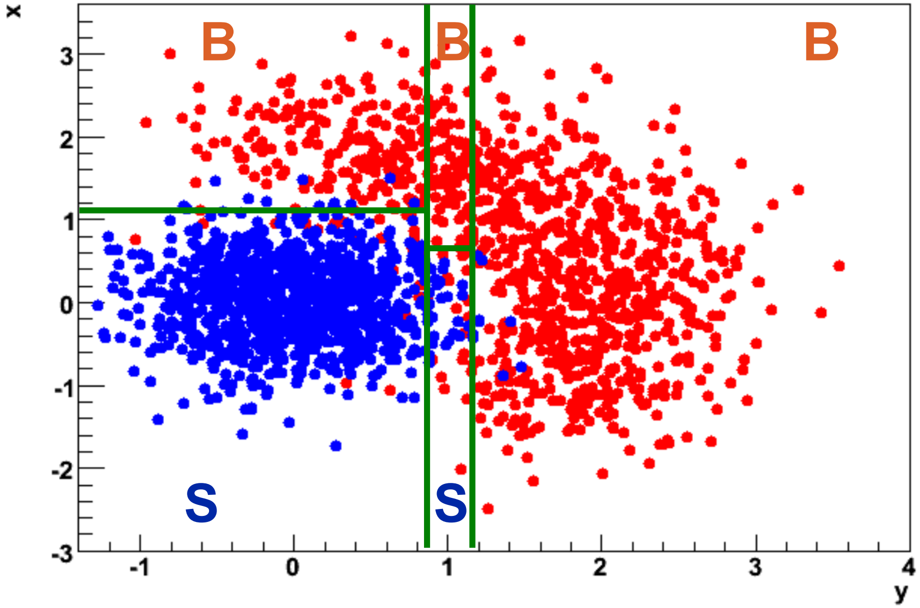
Decision Tree





Decision Tree





Boosted Decision Tree

Advantage: exploits correlations between variables

Disadvantage: not stable

Small fluctuations in the data can make a big difference in the outcome

Solution: “Boosting”

Build many different trees ($O(1000)$),
calculate weighted average over all trees

Automated algorithm:

Apply larger weights to misclassified events,
build new Decision Tree on reweighted events

AdaBoost Algorithm

For Tree *m*

$$W_m(x_i) \;=\; \text{weight of event } x_i$$

$$\text{err}_m \;=\; \sum_{x_i \text{ misclassified}} W_m(x_i)$$

$$\alpha_m \;=\; \text{const.} \times \ln\left(\frac{1-\text{err}_m}{\text{err}_m}\right)$$

Reweight misclassified events for next Tree

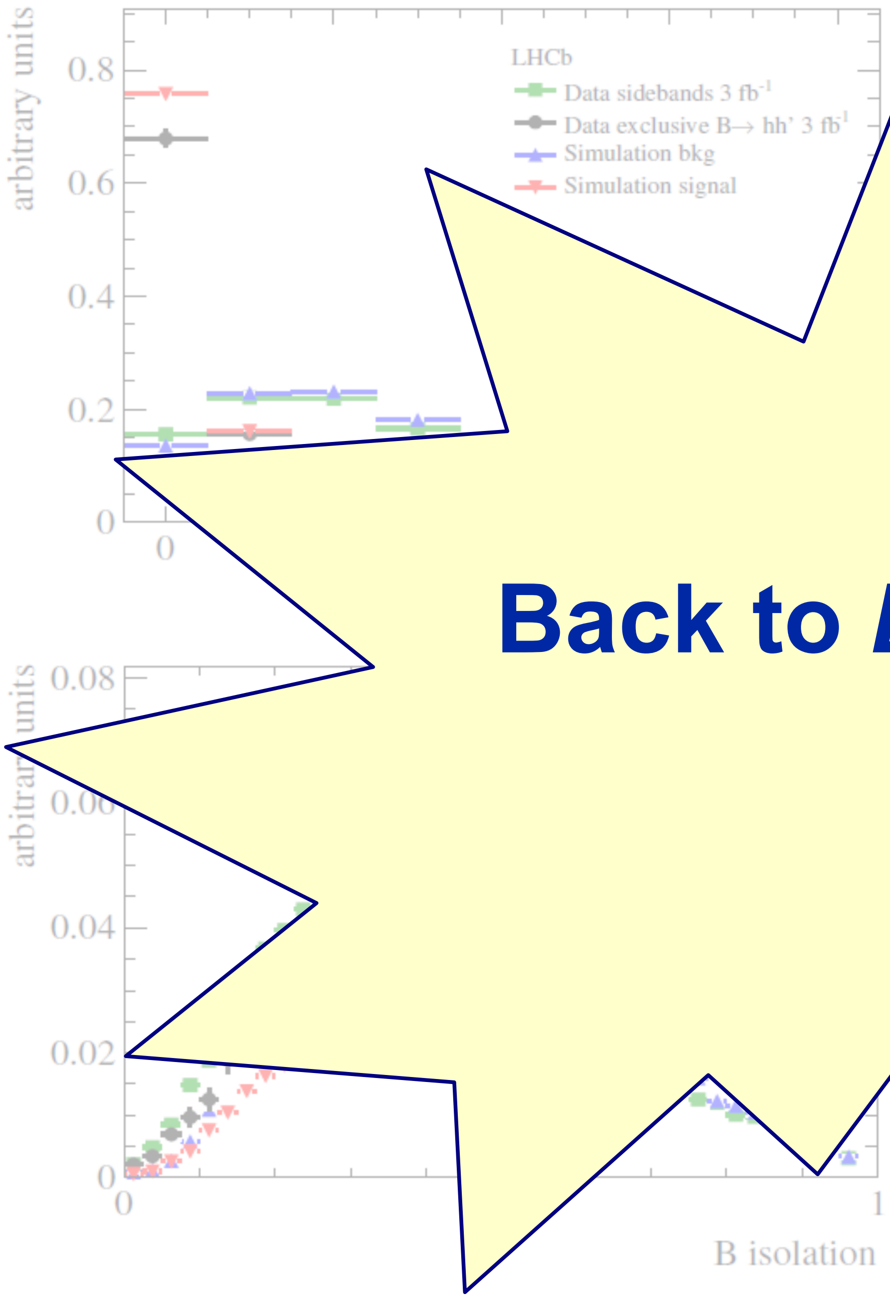
$$W_{m+1}(x_i) \;=\; W_m(x_i) \times e^{\alpha_m}$$

Final classifier for event x_i

$$\boldsymbol{T(x_i) \;=\; \sum_{m=1}^{N_{\text{trees}}} \alpha_m \times T_m(x_i)}$$

$$T_m(x_i) \;=\; \begin{cases} 1 & \text{if } x_i \text{ is classified as signal in Tree } m \\ 0 & \text{if } x_i \text{ is classified as background in Tree } m \end{cases}$$

Example: $B_s^0 \to \mu^+\mu^-$



Back to $B_s^0 \to \mu^+\mu^-$

Example: $B_s^0 \rightarrow \mu^+ \mu^-$

Train BDT on simulated samples of signal and background events

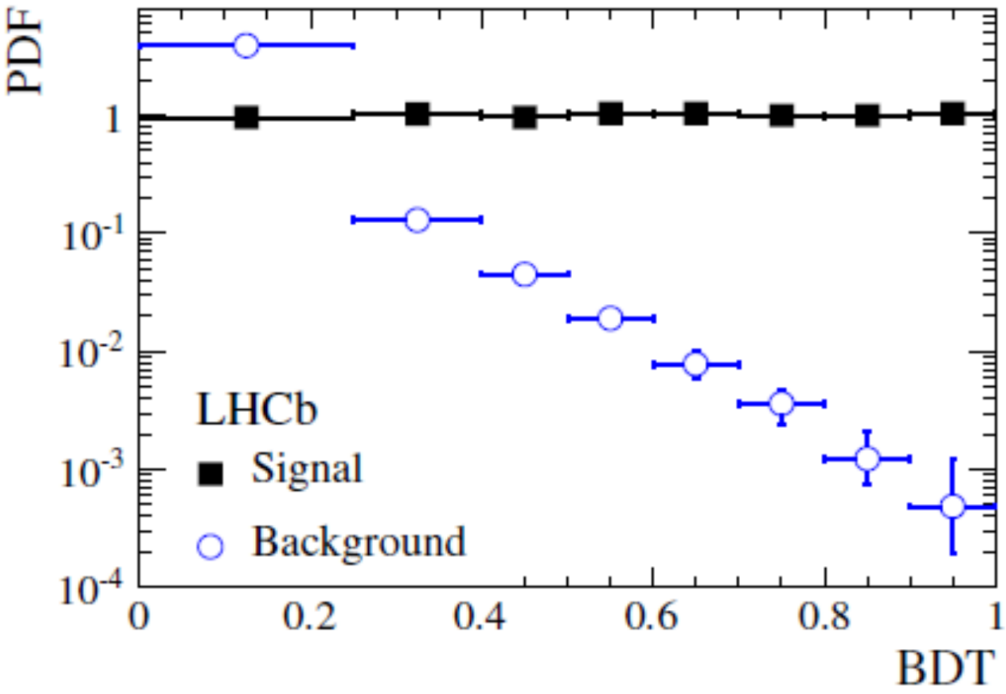
(where we know for each event whether it is signal or background)

Recalibrate BDT response on control channels in data

$B^0 \rightarrow K^+ \pi^-$ as proxy for signal

Events with $m(\mu^+ \mu^-) > m(B_s^0)$ as proxy for background

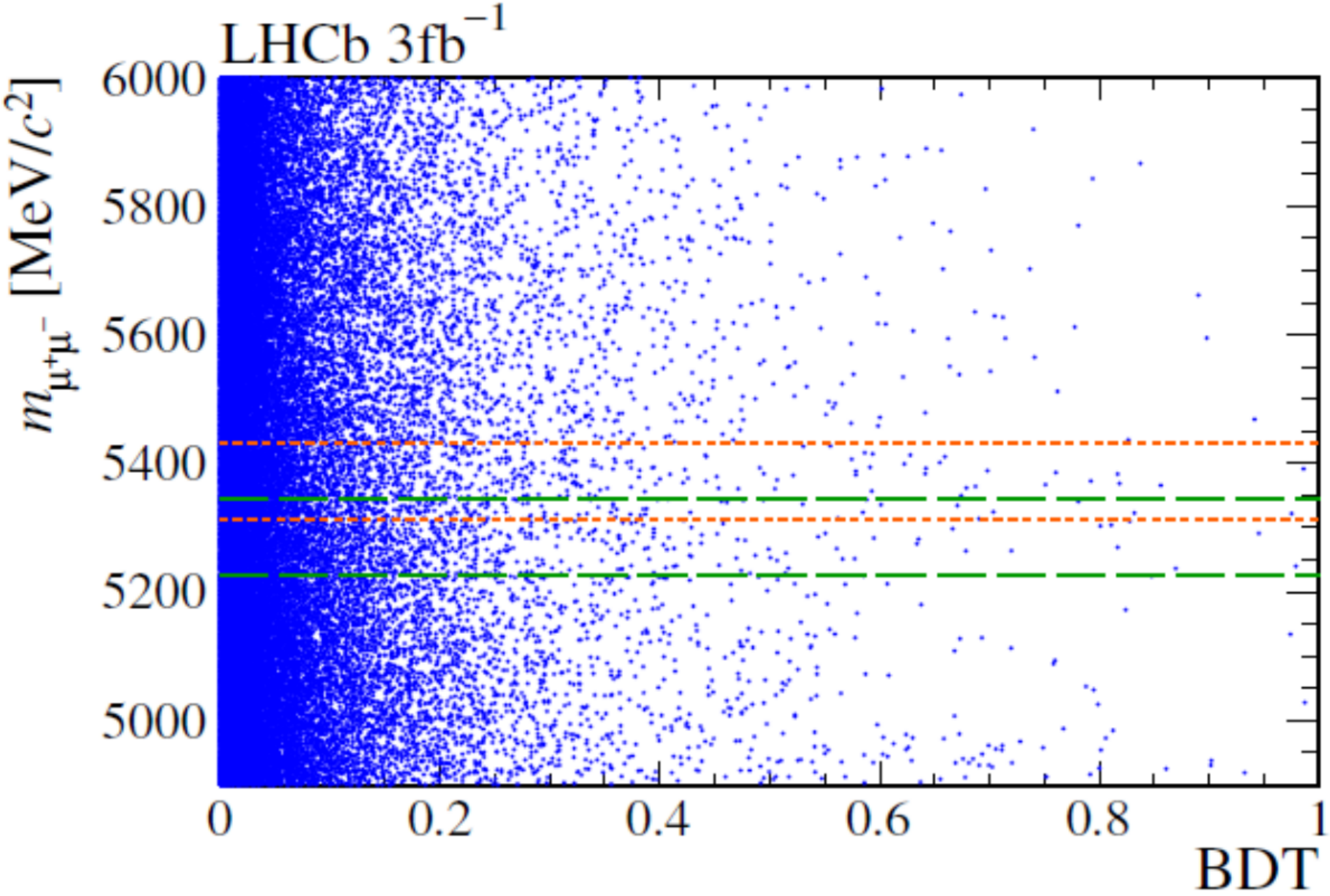
Aim at flat response for signal, peaking at 0 for background



(note logarithmic scale !)

Example: $B_s^0 \to \mu^+\mu^-$

Fit two-dimensional distribution of BDT classifier and $m(\mu^+\mu^-)$ to extract the number of signal candidates



B_s^0 mass region

Example: $B_s^0 \to \mu^+\mu^-$

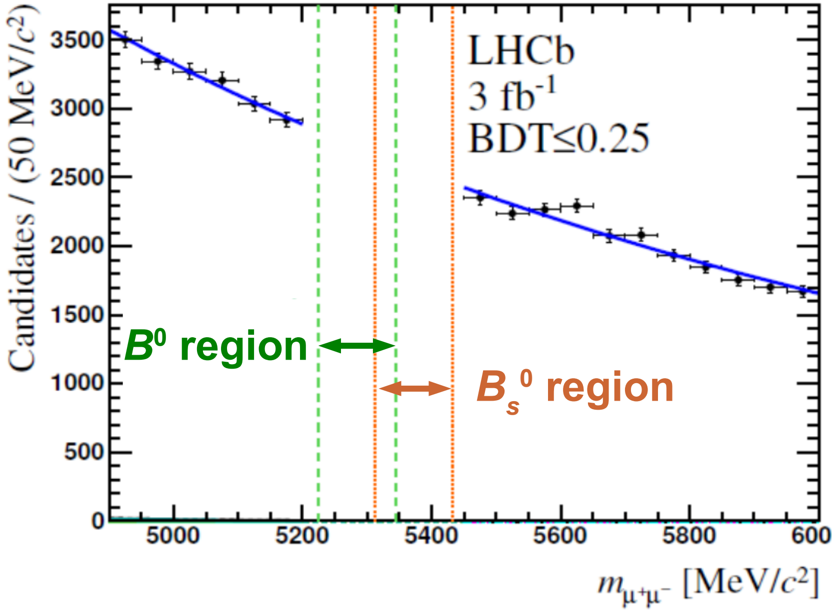
Need to model the expected distribution in $m(\mu^+\mu^-)$

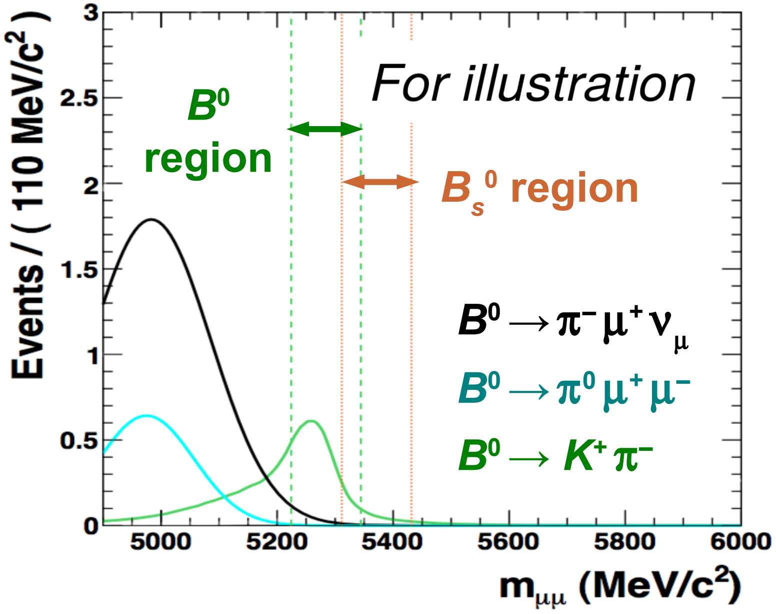
Background from random $\mu^+\mu^-$ combinations

→ Exponential distribution

Background from wrongly reconstructed *B* meson decays

→ from simulated events



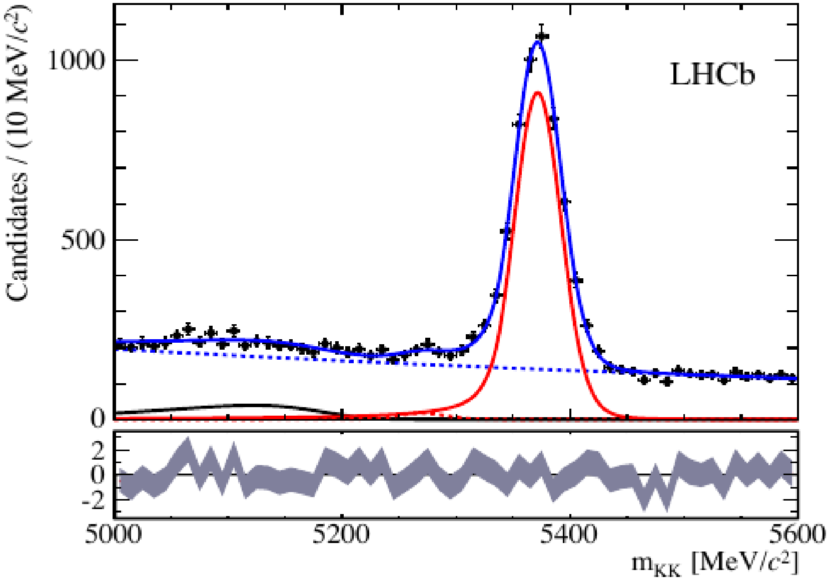


Example: $B_s^0 \to \mu^+\mu^-$

Need to model the expected distribution in $m(\mu^+\mu^-)$

Signal: expected position of the peak from large samples of reconstructed $B_s^0 \to K^+K^-$ and similar decays

$B_s^0 \to K^+K^-$

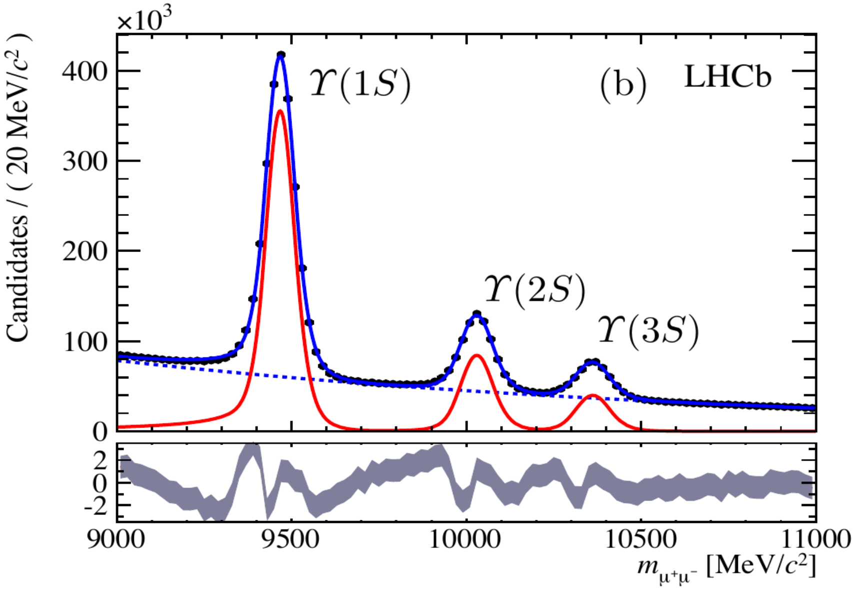


Example: $B_s^0 \to \mu^+\mu^-$

Need to model the expected distribution in $m(\mu^+\mu^-)$

Signal: expected width of the peak from large samples of reconstructed $\Upsilon \to \mu^+\mu^-$ decays and others

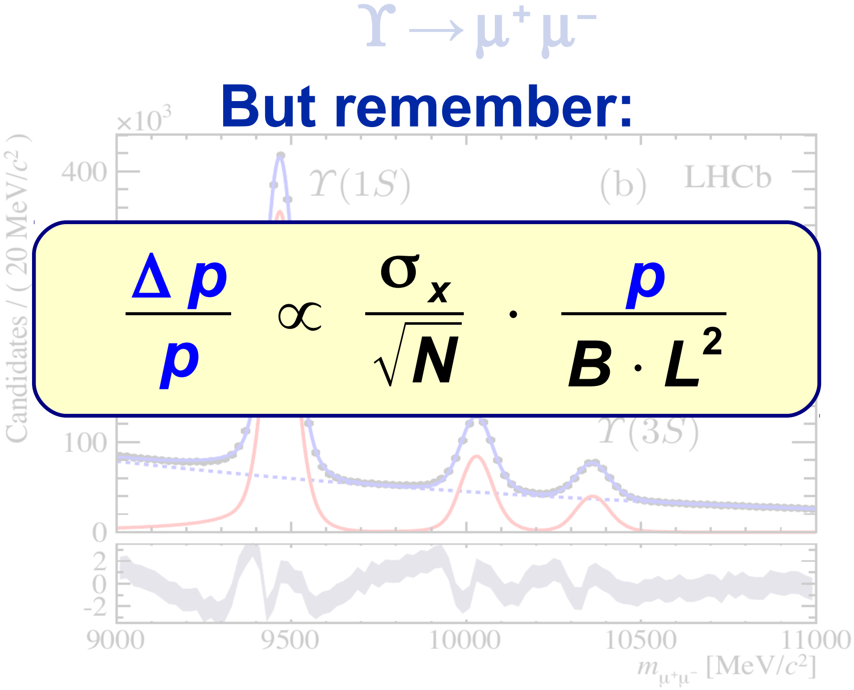
$\Upsilon \to \mu^+\mu^-$



Example: $B_s^0 \to \mu^+\mu^-$

Need to model the expected distribution in $m(\mu^+\mu^-)$

Signal: expected width of the peak from large samples of reconstructed $\Upsilon \to \mu^+\mu^-$ decays and others



But remember:

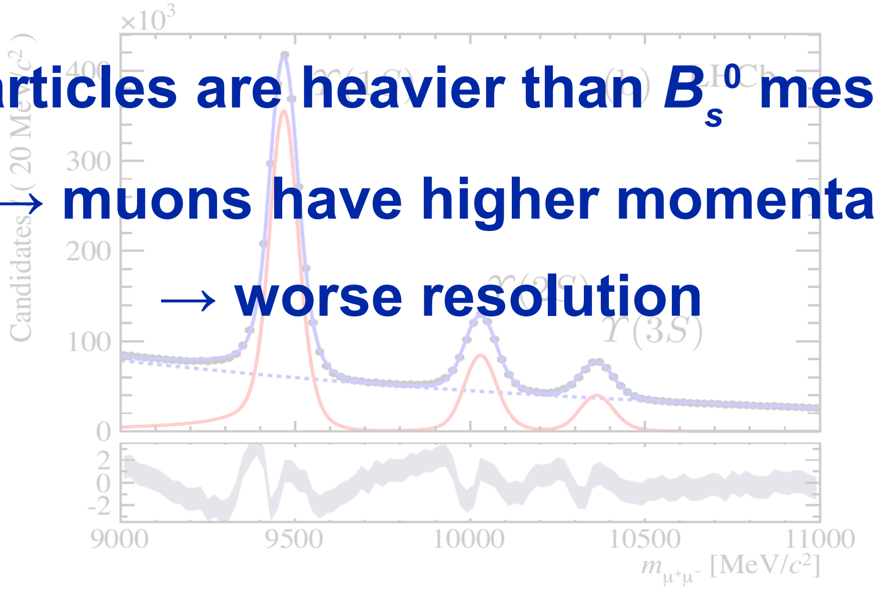
$$\frac{\Delta p}{p} \propto \frac{\sigma_x}{\sqrt{N}} \cdot \frac{p}{B \cdot L^2}$$

Example: $B_s^0 \rightarrow \mu^+\mu^-$

Need to model the expected distribution in $m(\mu^+\mu^-)$

Signal: expected width of the peak from large samples of reconstructed $\Upsilon \rightarrow \mu^+\mu^-$ decays and others

$\Upsilon \rightarrow \mu^+\mu^-$



Υ particles are heavier than B_s^0 mesons

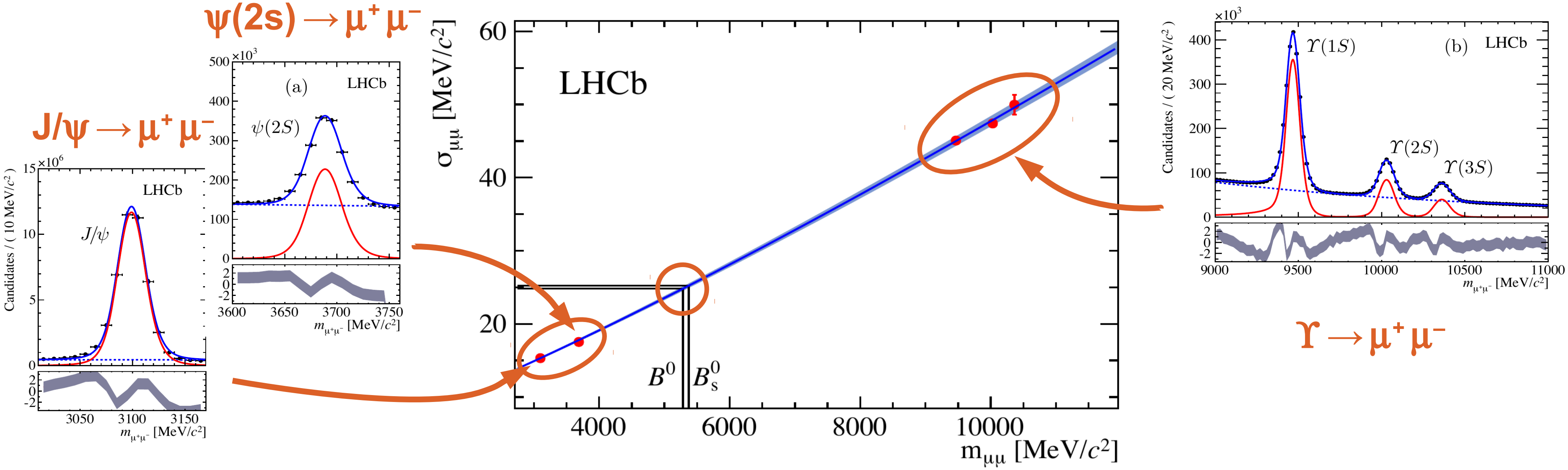
→ muons have higher momenta

→ worse resolution

Example: $B_s^0 \rightarrow \mu^+ \mu^-$

Need to model the expected distribution in $m(\mu^+\mu^-)$

Signal: expected width of the peak from large samples of reconstructed $\Upsilon \rightarrow \mu^+\mu^-$ decays and others


$\psi(2s) \rightarrow \mu^+\mu^-$

$J/\psi \rightarrow \mu^+\mu^-$

$\Upsilon \rightarrow \mu^+\mu^-$

Yesterday's sensation
is today's calibration channel

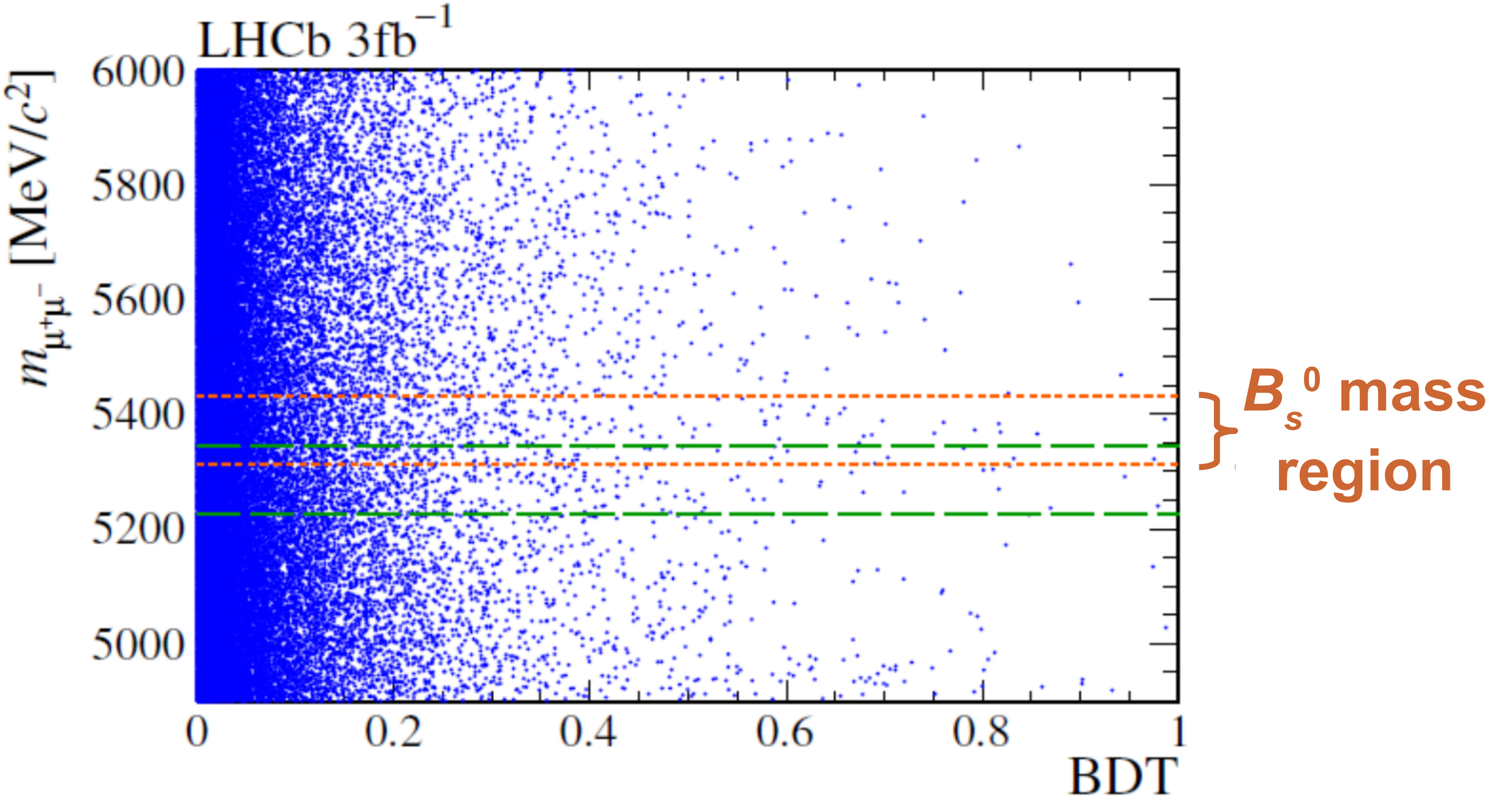
(Richard P. Feynman)

photo source: <https://en.wikipedia.org/wiki/Richard_Feynman>

Example: $B_s^0 \rightarrow \mu^+\mu^-$

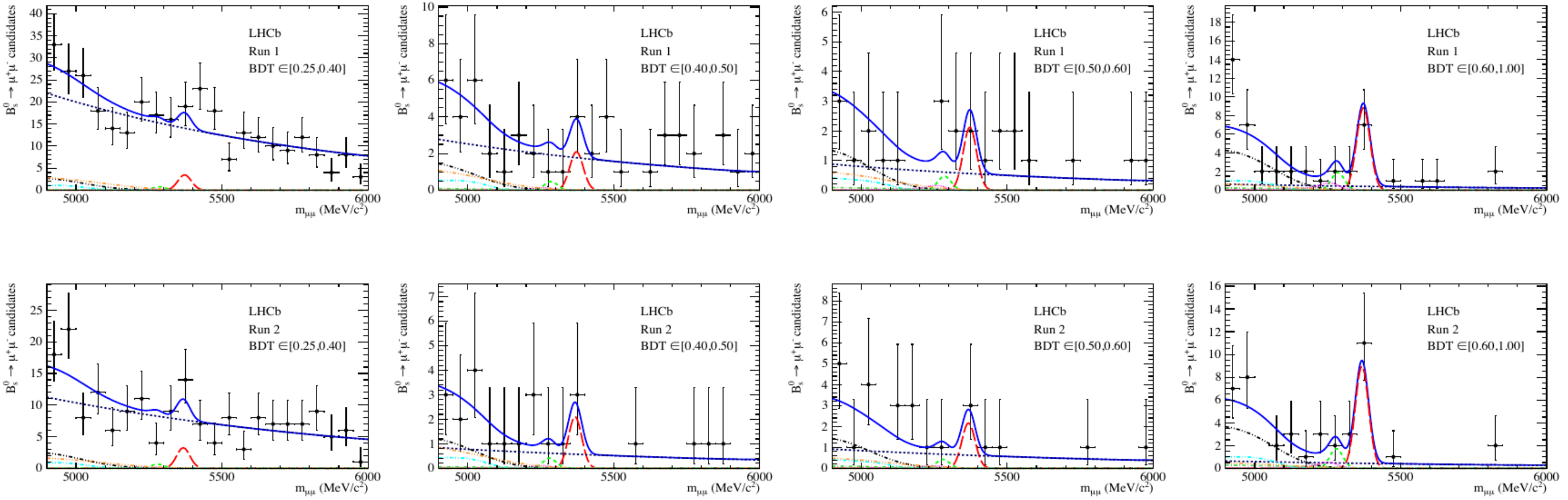
Have models for the expected signal and background distributions in the BDT classifier and in $m(\mu^+\mu^-)$

→ Maximum Likelihood fit to the measured distribution to estimate the number of signal events in our sample



Example: $B_s^0 \rightarrow \mu^+ \mu^-$

Hurray, find an excess of events around B_s^0 mass at high BDT

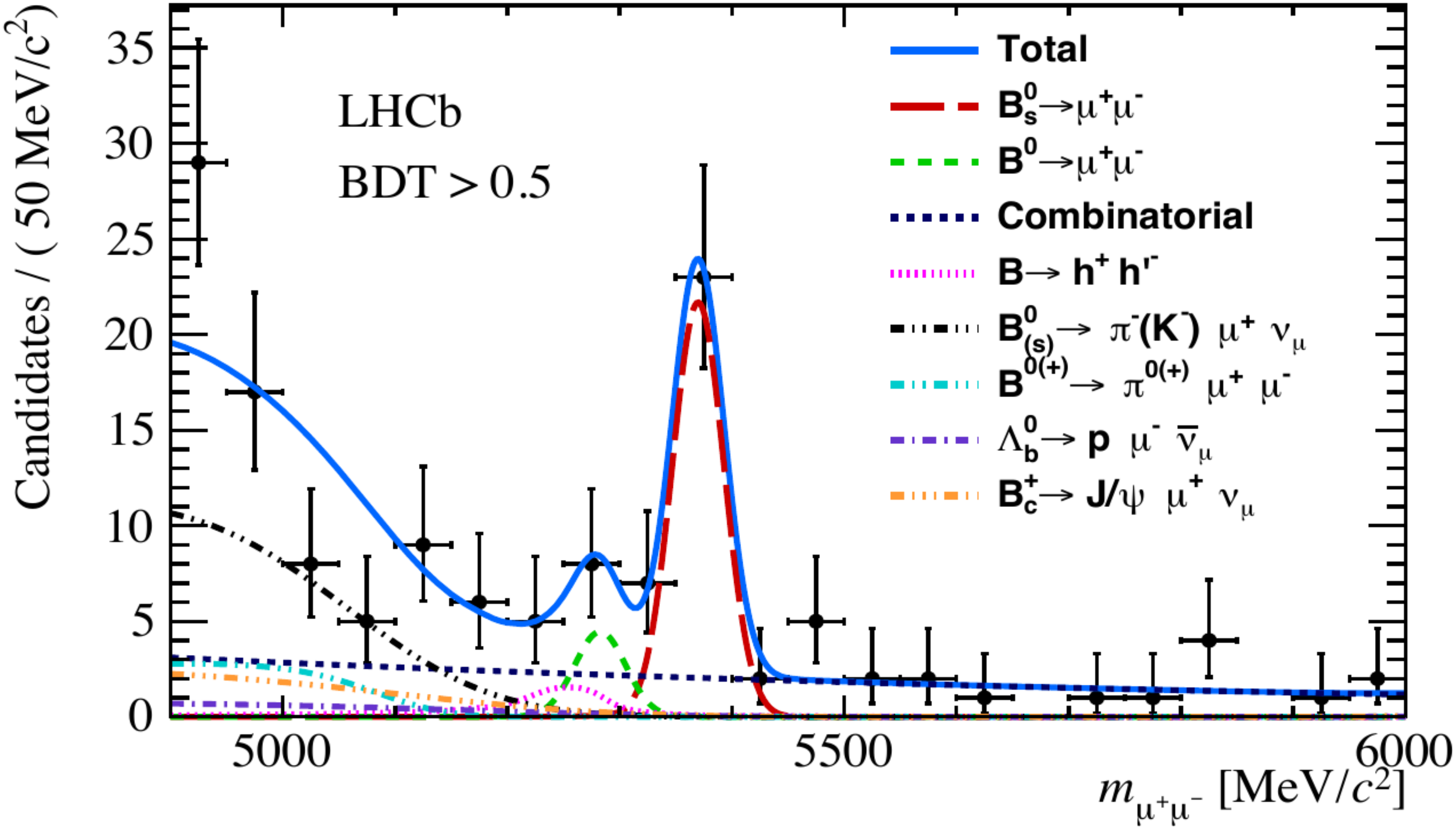


0.25 < BDT < 0.40 0.40 < BDT < 0.50 0.50 < BDT < 0.60 0.60 < BDT < 1.00

background-like ⟷ **signal-like**

Example: $B_s^0 \rightarrow \mu^+ \mu^-$

Hurray, find an excess of events around B_s^0 mass at high BDT



Example: $B_s^0 \rightarrow \mu^+ \mu^-$

<u>Two remaining questions</u>:

Is the excess statistically significant ?

i.e. how large is the probability that it could be caused
by a random fluctuation in the distribution of background events ?

If the excess is "real", how large is the Branching Fraction ?

i.e. how does the extracted number of $B_s^0 \rightarrow \mu^+ \mu^-$ candidates
translate into a probability for a B_s^0 to decay into $\mu^+ \mu^-$?

Example: $B_s^0 \rightarrow \mu^+ \mu^-$

Two remaining questions:

Is the excess statistically significant ?

i.e. how large is the probability that it could be caused
by a random fluctuation in the distribution of background events ?

If the excess is “real”, how large is the Branching Fraction ?

i.e. how does the extracted number of $B_s^0 \rightarrow \mu^+ \mu^-$ candidates
translate into a probability for a B_s^0 to decay into $\mu^+ \mu^-$?

Example: $B_s^0 \rightarrow \mu^+\mu^-$

Number of selected signal events ← what we have

=

Number of *pp* collisions analysed

×

Probability that a *pp* collision produces a B_s^0

×

Probability that the B_s^0 decays to $\mu^+\mu^-$

×

Probability that the μ^+ and the μ^- leave a trace in the detector

×

Efficiency of the trigger selection

×

Efficiency of the reconstruction algorithms

×

Efficiency of the offline selection criteria

Example: $B_s^0 \rightarrow \mu^+ \mu^-$

Number of selected signal events ← what we have

=

Number of *pp* collisions analysed

×

Probability that a *pp* collision produces a B_s^0

×

Probability that the B_s^0 decays to $\mu^+ \mu^-$ ← what we want

×

Probability that the μ^+ and the μ^- leave a trace in the detector

×

Efficiency of the trigger selection

×

Efficiency of the reconstruction algorithms

×

Efficiency of the offline selection criteria

Example: $B_s^0 \to \mu^+\mu^-$

Number of selected signal ents

=

Number of ysed

Pro

ne detector

tion

ec uction algorithms

×

ciency of the offline selection criteria

More precise

to do a relative measurement

Example: $B_s^0 \rightarrow \mu^+\mu^-$

Select a reference decay mode for which the

Branching Fraction (decay probability)
is large and well known from earlier measurements

Trigger, reconstruction and selection efficiencies
are as similar as possible to those for $B_s^0 \rightarrow \mu^+\mu^-$

$$\boxed{\mathrm{BF}\left(B_s^0 \rightarrow \mu^+\mu^-\right) = \underbrace{\mathrm{BF}\left(\mathrm{ref}\right)}_{\text{known}} \times \underbrace{\frac{N\left(B_s^0 \rightarrow \mu^+\mu^-\right)}{N\left(\mathrm{ref}\right)}}_{\text{measured}} \times \underbrace{\frac{\epsilon\left(\mathrm{ref}\right)}{\epsilon\left(B_s^0 \rightarrow \mu^+\mu^-\right)}}_{\substack{\text{to be determined,} \\ \text{should be close to 1}}}}$$

Example: $B_s^0 \to \mu^+ \mu^-$

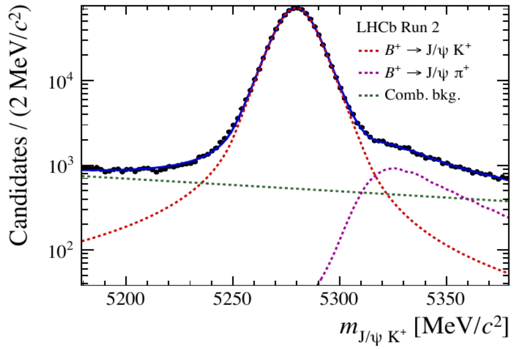
Select a reference decay mode for which

The Branching Fraction (decay probability)
is large and well known from earlier measurements

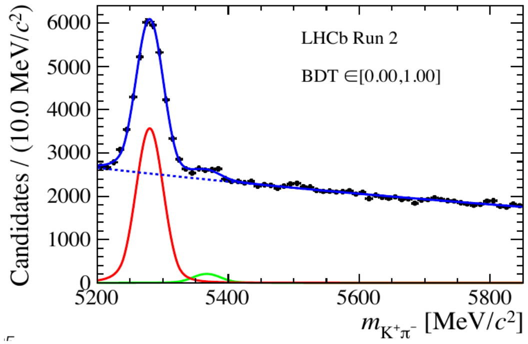
The trigger, reconstruction and selection efficiencies
are as similar as possible to those for $B_s^0 \to \mu^+ \mu^-$

Actually use two reference decays

$B^+ \to J/\psi\, K^+$ with $J/\psi \to \mu^+ \mu^-$
similar trigger but an additional particle



$B^0 \to K^+ \pi^-$
same number of particles, but different trigger



The two give consistent results

Example: $B_s^0 \rightarrow \mu^+ \mu^-$

Two remaining questions:

Is the excess statistically significant ?

i.e. how large is the probability that it could be caused
by a random fluctuation in the distribution of background events ?

If the excess is "real", how large is the Branching Fraction ?

i.e. how does the extracted number of $B_s^0 \rightarrow \mu^+ \mu^-$ candidates
translate into a probability for a B_s^0 to decay into $\mu^+ \mu^-$?

Example: $B_s^0 \to \mu^+ \mu^-$

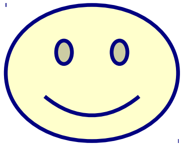
Maximum Likelihood fit:

Finds optimum values for the fit parameters such that the probability to obtain the observed distribution is maximized

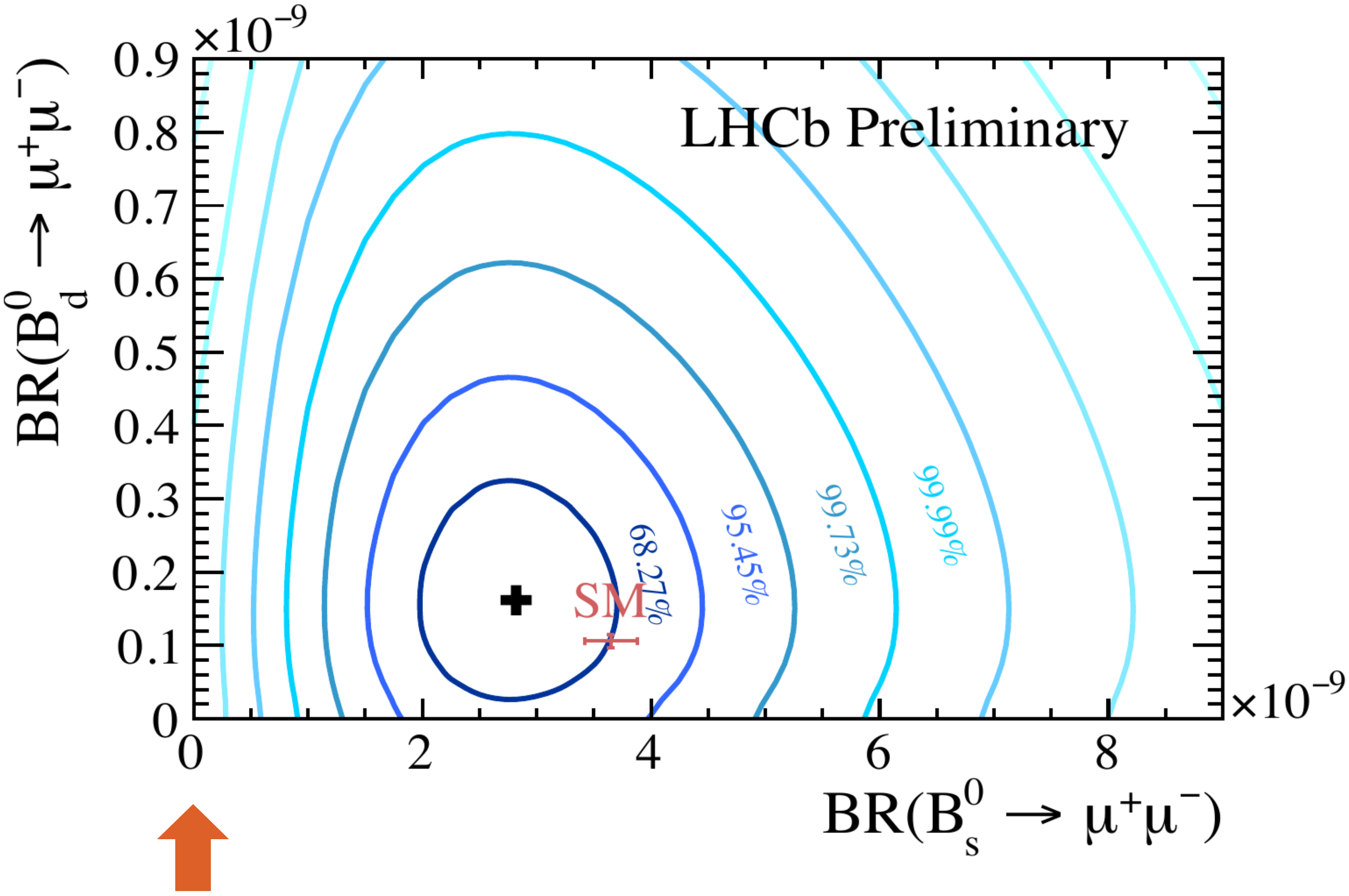
Likelihood profile:

Change the values of the parameters from those found by the fit and re-calculate the probability

➡ **Sign up for my course “Datenanalyse” here at UZH in Fall semester**



Example: $B_s^0 \to \mu^+\mu^-$



$\mathrm{BR}(B_s^0 \to \mu^+\mu^-) = 0$ is excluded with 99.9999999999997 % probability

Result of the measurement is in good agreement with the prediction from the Standard Model

Overview

Part I

What are we trying to measure

What do our experiments look like

Part II

“Online” event selection

Part III

“Offline” analysis