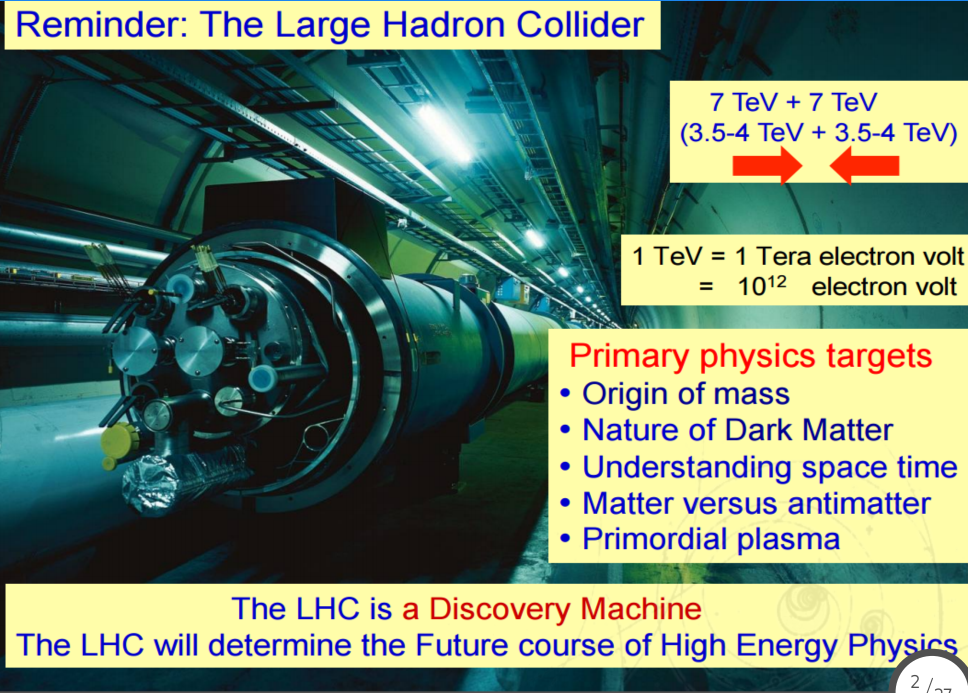
$\tau \longrightarrow \mu\mu\mu$ in LHCb

Marcin Chrząszcz
mchrzasz@cern.ch

University of Zurich^UZH

Kaggle Seminar, San Francisco
August 21, 2015

Reminder: The Large Hadron Collider



7 TeV + 7 TeV
(3.5-4 TeV + 3.5-4 TeV)

1 TeV = 1 Tera electron volt
= 10^{12} electron volt

Primary physics targets
- Origin of mass
- Nature of Dark Matter
- Understanding space time
- Matter versus antimatter
- Primordial plasma

The LHC is a Discovery Machine
The LHC will determine the Future course of High Energy Physics

LHC did very well

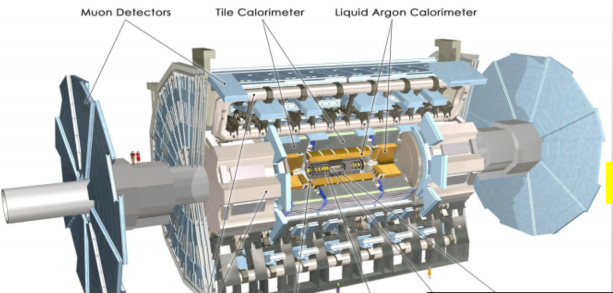
2011: luminosity 3.5 . 10^{33} cm^{-2} s^{-1} ⇒ >5 fb^{-1} collected in total

2012: luminosity 7.6 . 10^{33} cm^{-2} s^{-1} ⇒ >20 fb^{-1} collected in total

Next pp collisions in 2015. Shutdown for 'energy upgrade'

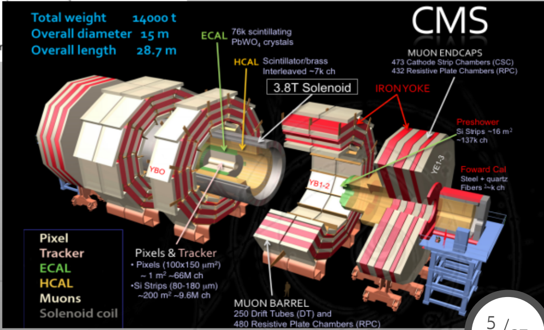
Experiments at the LHC

The Higgs Hunters @ the LHC



LHC: pp collisions at 7/8 TeV

The ATLAS experiment



The CMS experiment

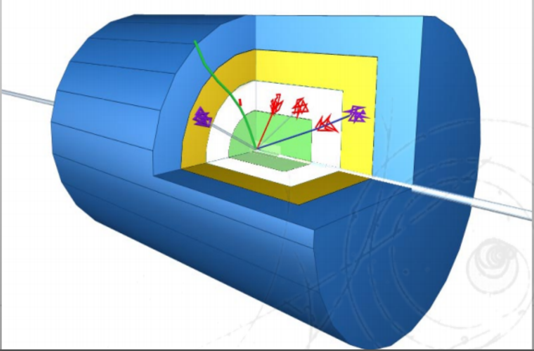
These experiments use different technologies for their detector components

Schematic of a LHC Detector

Physics requirements drive the design!

Analogy with a cylindrical onion:

Technologically advanced detectors comprising many layers, each designed to perform a specific task.
Together these layers allow us to identify and precisely measure the energies and directions of all the particles produced in collisions.



Such an experiment has ~ 100 Million read-out channels!!

The experiments are in good shape!

ATLAS

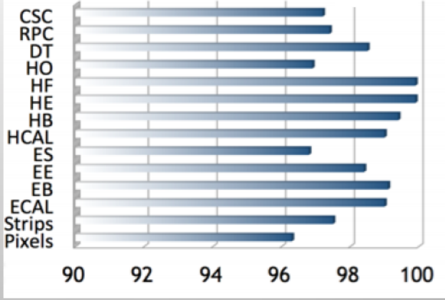
Fraction of non-operational detector channels: (depends on the sub-detector)

Few permil (most cases) to 4%

Data-taking efficiency = (recorded lumi)/(delivered lumi):~ 94.6%

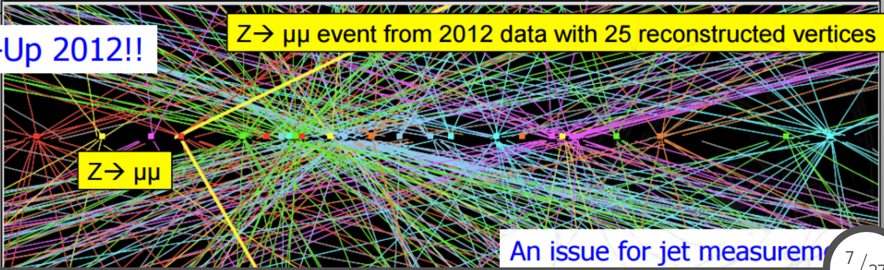
Good-quality data fraction, used for analysis : ~ 93.6%
(will increase further with data reprocessing)

CMS



Pile-Up 2012!!

Z→ μμ event from 2012 data with 25 reconstructed vertices



An issue for jet measurem

Data Taking Challenges

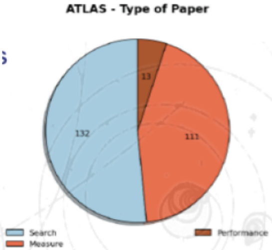
- Collider: 20M bunch crossings per second
- ~ 30 events per bunch crossing: pile-up
- Trigger on 400 events/sec (+ another 400-600 Hz of parked data in CMS): keep the interesting (incl. unknown) physics
- Total data volume in eg ATLAS: 5 billion detector events, 120 PB of data (simulation and data). Several billion Monte Carlo events (produce ~ 10^9 events/2 months)
- ATLAS+CMS > 500 papers so far

  > ~600 papers for all experiments

No attempt to cover everything ☺ but

examples to illustrate the LHC

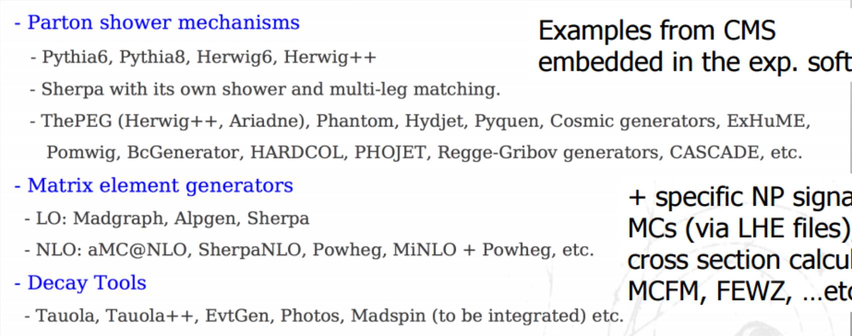
Most examples from CMS/ATLAS



Monte Carlos are very important

For our daily work…

- Guidance on the background to the signal (direct or training)
- Signal acceptance
- Systematic error evaluation
- Comparison with theory



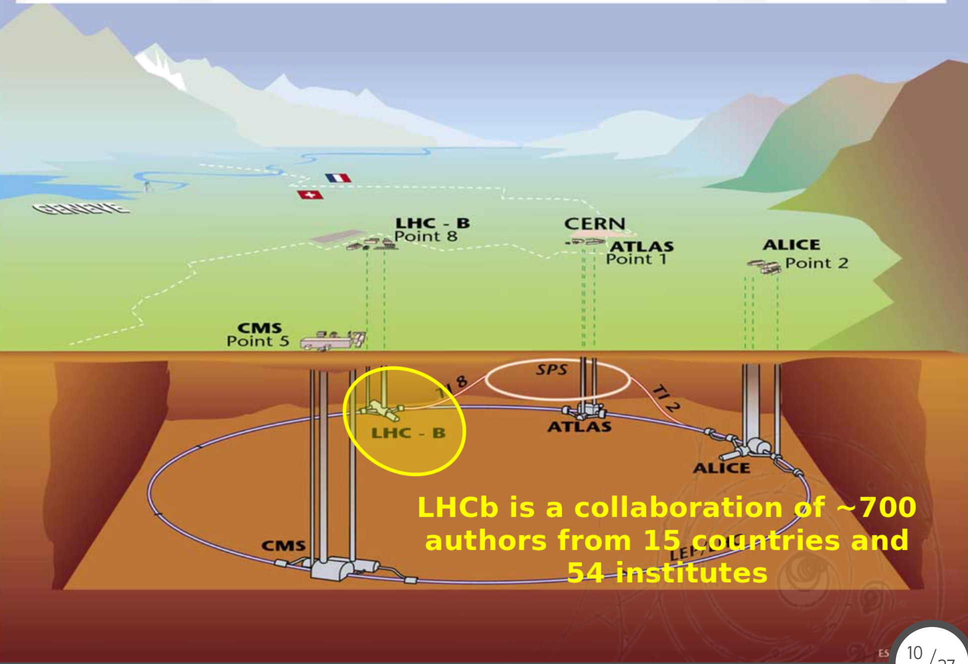
- Parton shower mechanisms
  - Pythia6, Pythia8, Herwig6, Herwig++
  - Sherpa with its own shower and multi-leg matching.
  - ThePEG (Herwig++, Ariadne), Phantom, Hydjet, Pyquen, Cosmic generators, ExHuME, Pomwig, BcGenerator, HARDCOL, PHOJET, Regge-Gribov generators, CASCADE, etc.
- Matrix element generators
  - LO: Madgraph, Alpgen, Sherpa
  - NLO: aMC@NLO, SherpaNLO, Powheg, MiNLO + Powheg, etc.
- Decay Tools
  - Tauola, Tauola++, EvtGen, Photos, Madspin (to be integrated) etc.

Examples from CMS embedded in the exp. software

+ specific NP signal MCs (via LHE files), cross section calculators MCFM, FEWZ, …etc…

Experimentalists always want more, better (HO), faster, well tuned MC…

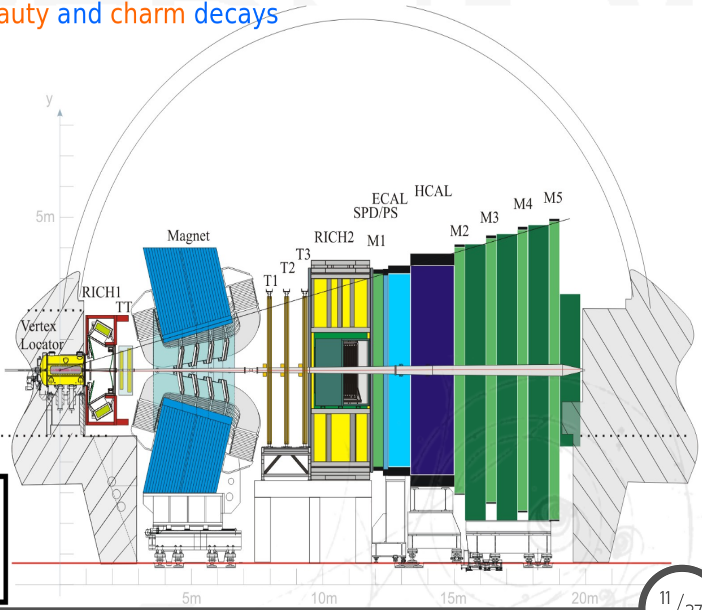
Overall view of the LHC experiments.



LHCb is a collaboration of ~700 authors from 15 countries and 54 institutes

Dedicated flavour physics experiment

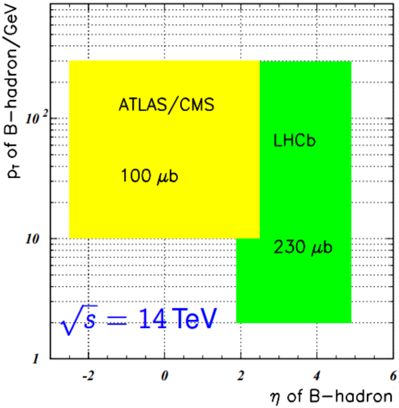
- forward precision spectrometer
- optimised for beauty and charm decays

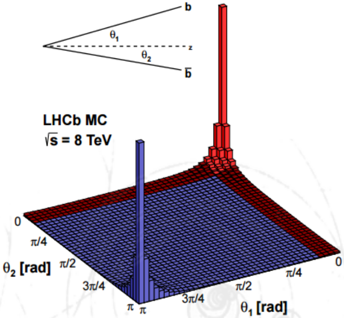


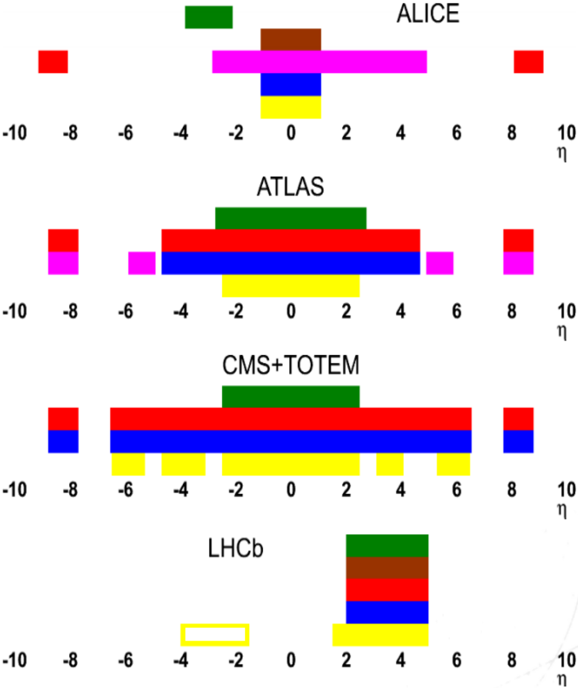
Magnet spectrometer

→ optimization for B-Physics – but can do much more

- forward angular coverage → large boosts: B decay lengths O(1 cm)
- focus on vertex reconstruction and particle identification
- phase space coverage down to low p_T, small x_{Bj} and large η
- flexible and highly selective trigger

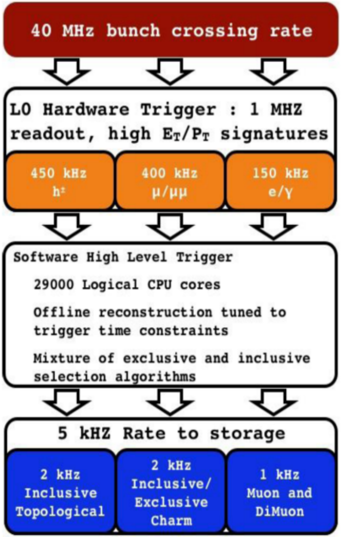






- ALICE
  - central
  - forward muon coverage
- ATLAS & CMS
  - central detectors
- LHCb
  - forward detector
  - tracking, particle-ID and calorimetry in full acceptance

🟫 hadron PID
🟩 muon system
🟪 lumi counters
🟥 HCAL
🟦 ECAL
🟨 tracking



40 MHz bunch crossing rate

L0 Hardware Trigger : 1 MHZ readout, high E_T/P_T signatures

450 kHz $h^{\pm}$

400 kHz $\mu/\mu\mu$

150 kHz e/γ

Software High Level Trigger

29000 Logical CPU cores

Offline reconstruction tuned to trigger time constraints

Mixture of exclusive and inclusive selection algorithms

5 kHZ Rate to storage

2 kHz Inclusive Topological

2 kHz Inclusive/ Exclusive Charm

1 kHz Muon and DiMuon

ca. 50 kB/event

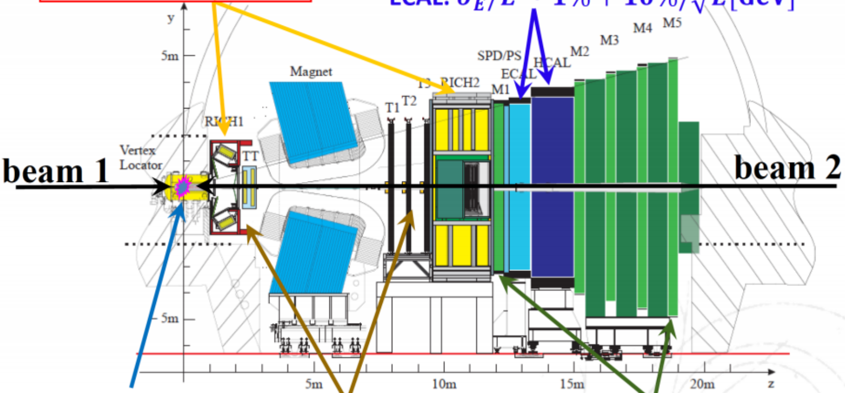
→ *allow selection of rare processes*

- Level-0 Trigger: hardware
  - → fully synchronous at 40 MHz
  - → use calorimeters and muon system
  - → selection of high-p_T particles
    - ◆ $p_T(\mu) > O(1)$ GeV/c
    - ◆ $p_T(h, e, \gamma) > O(3)$ GeV/c
- High-Level Trigger: software
  - → HLT1: add VELO information
    - ◆ impact parameter- and lifetime cuts
  - → HLT2: global event reconstruction
    - ◆ exclusive & inclusive selections
- up to $O(30)$ kHz “deferred” triggering

Pseudorapidity acceptance
$2 < \eta < 5$

RICH1 & RICH2
$\epsilon(K \to K) \sim 95\%$
$\pi \to K$ mis-id: $\sim 5\%$

Calorimeters
ECAL: $\sigma_E/E \sim 1\% + 10\%/\sqrt{E[\mathrm{GeV}]}$



VELO
$\sigma_{\mathrm{IP}} \sim 20\ \mu\mathrm{m}$
for high-p_T tracks

Tracking System
$\Delta p/p = 0.4\%@5\ \mathrm{GeV}/c$
to $0.6\%@100\ \mathrm{GeV}/c$

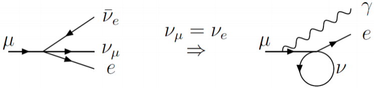
Muon System
$\epsilon(\mu \to \mu) \sim 97\%$
$\pi \to \mu$ mis-id: $1 \sim 3\%$

Int.J.Mod.Phys.A 30(2015)1530022

Lepton Flavour/Number Violation

Lepton Flavour Violation(LFV):

After μ^- was discovered (1936) it was natural to think of it as an excited e^-.



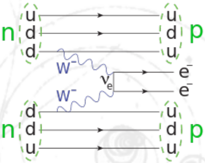
- Expected: $B(\mu \to e\gamma) \approx 10^{-4}$
- Unless there is a nother ν.

I.I.Rabi:

> "Who ordered that?"

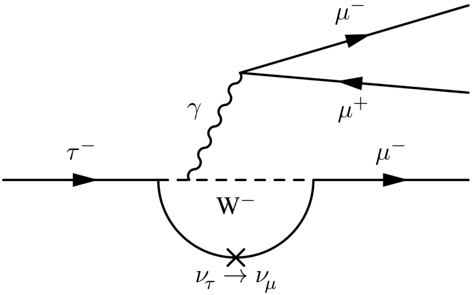
- Up to this day charged LFV is being searched for in various decay modes.
- LFV was already found in neutrino sector (oscillations).

Lepton Number Violation (LNV)



- Even with LFV, lepton number can be a conserved quantity.
- Many NP models predict LNV (Majorana neutrinos)
- LNV searched in s-called neutrinoless double β decays.

Status of searches for



- Charged Lepton Flavour Violation process.
- The Standard Model contribution: penguin diagram with neutrino oscillation

Current limits (90 % CL)

BaBar	3.3×10^{-8}
Belle	2.1×10^{-8}

Predictions

SM	$O(10^{-40})$
var. SUSY	10^{-10}
non universal Z'	10^{-8}
mSUGRA+seesaw	10^{-9}

and many more...

τ production

- τ's in LHCb come from five main sources:

Mode	7 TeV	8 TeV
Prompt $D_s \to \tau$	$71.1 \pm 3.0\,\%$	$72.4 \pm 2.7\,\%$
Prompt $D^+ \to \tau$	$4.1 \pm 0.8\,\%$	$4.2 \pm 0.7\,\%$
Non-prompt $D_s \to \tau$	$9.0 \pm 2.0\,\%$	$8.5 \pm 1.7\,\%$
Non-prompt $D^+ \to \tau$	$0.18 \pm 0.04\,\%$	$0.17 \pm 0.04\,\%$
$X_b \to \tau$	$15.5 \pm 2.7\,\%$	$14.7 \pm 2.3\,\%$

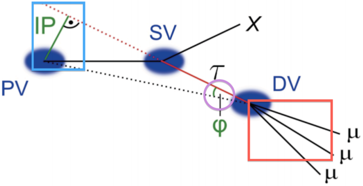
- Pythia produces them in wrong propotions
- Channels were produced seperatly and added in the given proporitons.

$\mathcal{B}(D^+ \to \tau)$

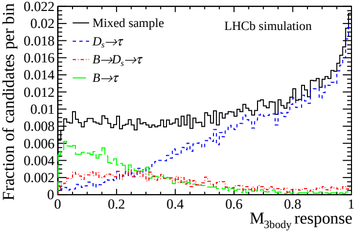
- There is no measurement of $\mathcal{B}(D^+ \to \tau)$.
- One can calculate it from: $\mathcal{B}(D^+ \to \mu\nu_\mu)$ + helicity suppression + phase space, `hep-ex:0604043`.
- $\mathcal{B}(D^+ \to \tau\nu_\tau) = (1.0 \pm 0.1) \times 10^{-3}$.

Signal and background discrimination

- Two multivariate classifiers, $\mathcal{M}_{3body}$ and $\mathcal{M}_{\mathcal{PID}}$.

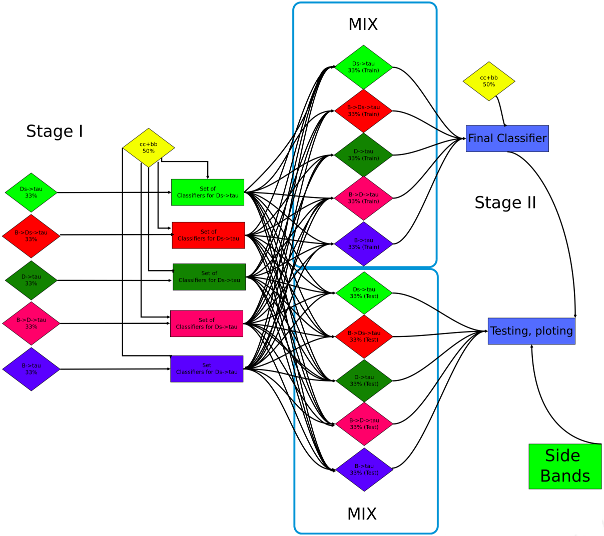


- $\mathcal{M}_{3body}$ trained using vertex and track fit quality, vertex displacement, vertex pointing, vertex isolation and τ p_T.
- Used Blending Technique (see the next slide).



- Trained on signal and background MC.
- Calibrated on $D_s \to \phi(\mu\mu)\pi$ sample.

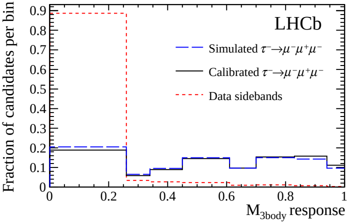
Blending technique



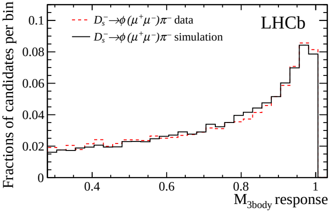
- Each of the τ lepton production channel have a different signature in terms of kinematic distributions.
- Signal blending technique improved the discriminating power by $6\ \%$

Calibration

- Assume all differences between $\tau \to \mu\mu\mu$ and $D_s \to \phi\pi$ come from kinematics (mass, resonance, decay time), which is correct in MC.
- Get correction $D_s \Longrightarrow \tau$ from MC.
- Apply corrections to $D_s \to \phi\pi$ on data.
- Publication in preparation.



- $D_s \to \phi\pi$ decay well modelled in MC.

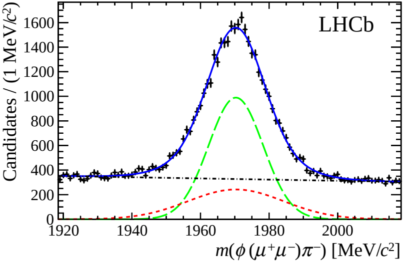


Relative normalisation

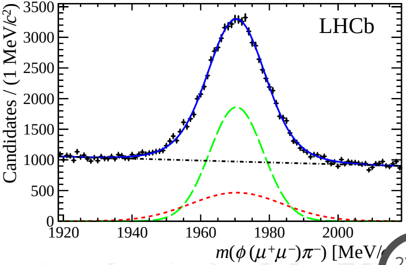
$$\boxed{\mathcal{B}(\tau \to \mu\mu\mu) = \frac{\mathcal{B}(D_s \to \phi\pi)}{\mathcal{B}(D_s \to \tau\nu_\tau)} \times f_{D_s}^{\tau} \times \frac{\varepsilon_{\text{norm}}}{\varepsilon_{\text{sig}}} \times \frac{N_{\text{sig}}}{N_{\text{norm}}} = \alpha \times N_{\text{sig}}}$$

- where ε stands for trigger, reconstruction, selection efficiency.
- $f_{D_s}^{\tau}$ is the fraction of τ coming from D_s.
- norm = normalisation channel $D_s \to \phi\pi$
  i.e. $(83 \pm 3)\,\%$ for 2012 data.

2011



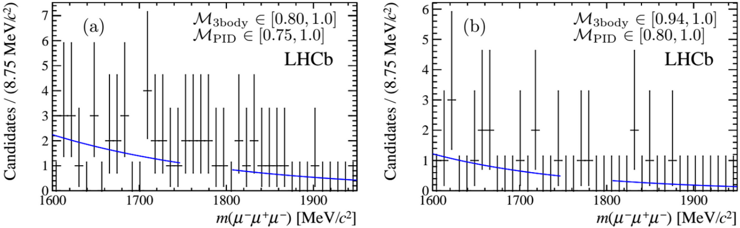
2012



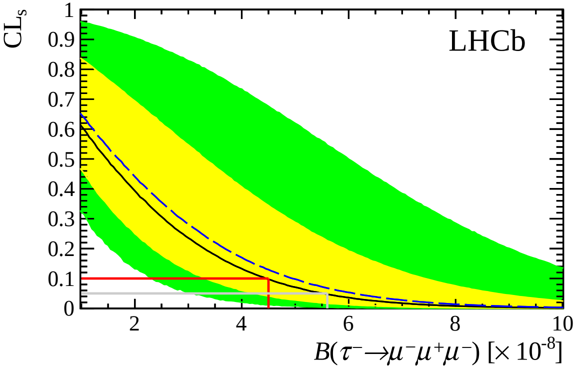
Remaining backgrounds

- Fit exponential to invariant mass spectrum in each likelihood bin.
- Don't use the $\pm 30\ \mathrm{MeV}$ region.

Example of most sensitive regions in 2011 and 2012



Results



Limits(PHSP):

Observed(Expected)

$4.6\ (5.0) \times 10^{-8}$ at 90% CL

$5.6\ (6.1) \times 10^{-8}$ at 95% CL

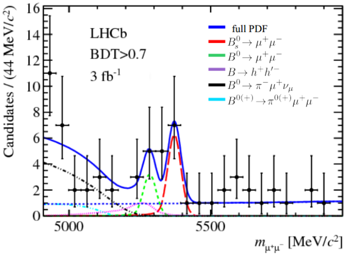
Dalitz distribution	$\times 10^{-8}$
$\varrho_V^{(LL)(LL)}$	4.2 (4.7)
$\varrho_V^{(LL)(RR)}$	4.1 (4.6)
$\varrho_{rad}^{(LR)}$	6.8 (7.6)
$\varrho_{mix}^{(LL)(LL)}$	4.4 (5.1)
$\varrho_{mix}^{(LL)(RR)}$	4.6 (5.0)

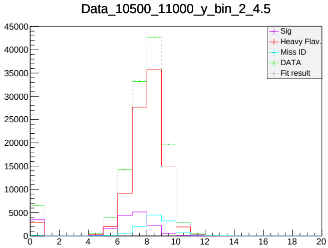
Why are we not putting the mass in the classifier?

⇛ Why don't we put mass in the classifier?

⇉ Many reasons:

- Our normalization channel is in different mass range!
- Mass resolution is wrongly modelled in MC.
- Easily to interpret:





Data agreement check, why do we bother?

$\Rightarrow$ It all boils down to our equation:

$$\boxed{\mathcal{B}(\tau \to \mu\mu\mu) = \frac{\mathcal{B}(D_s \to \phi\pi)}{\mathcal{B}(D_s \to \tau\nu_\tau)} \times f^{\tau}_{D_s} \times \frac{\varepsilon_{\text{norm}}}{\varepsilon_{\text{sig}}} \times \frac{N_{\text{sig}}}{N_{\text{norm}}} = \alpha \times N_{\text{sig}}}$$

There are 3 variables that we need to terminate: ε_{sig}, $\varepsilon_{\text{norm}}$ and N_{norm}.

- $\varepsilon_{\text{norm}}$; determine from data, by a cut and count method.
- N_{norm}; determined from data by a simple fit.
- ε_{sig}; calibrated on data:

$$\varepsilon_{\text{sig}} = \varepsilon^{\text{MC}}_{\text{sig}} \frac{\varepsilon^{\text{DATA}}_{\text{norm}}}{\varepsilon^{\text{MC}}_{\text{norm}}}$$

The hack that is used here is: ε_{sig} is ok, but N_{norm} is smaller, so alpha is bigger $\Rightarrow$ worse sensitivity.

Wrap up

1. Physics has a different application of ML than computer science.
2. There are physics consequance of what you use!
3. Blindly taking all varaibles is the bad solution.

Backup