Imperial College London

First Year Project 2016: Introduction to $H \rightarrow \tau\tau$ searches

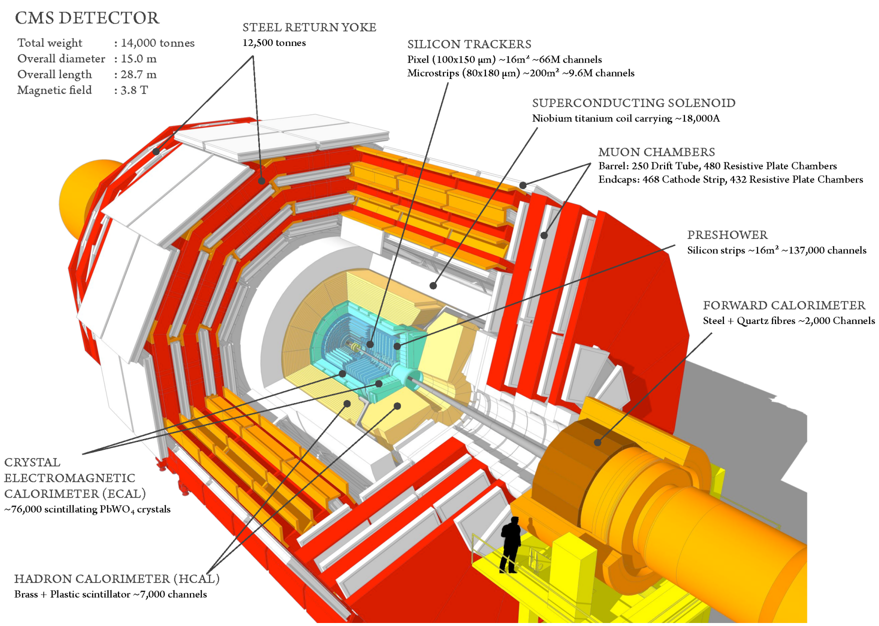
Rebecca Lane

05/05/16

The LHC

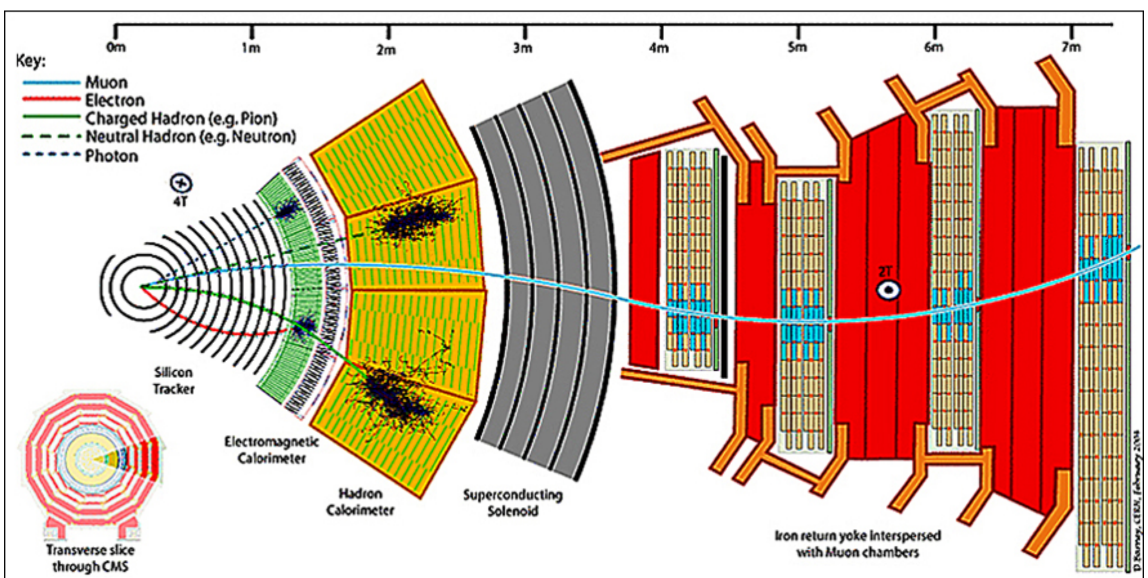
- Worlds largest hadron collider.
- Circulates protons in opposite directions around a 27km ring, allowing them to collide in 4 different positions around the ring, where detectors are placed to study the collisions.
- Collisions occur at centre of mass energies of up to 14 TeV – higher than anything previously reached.

The CMS detector

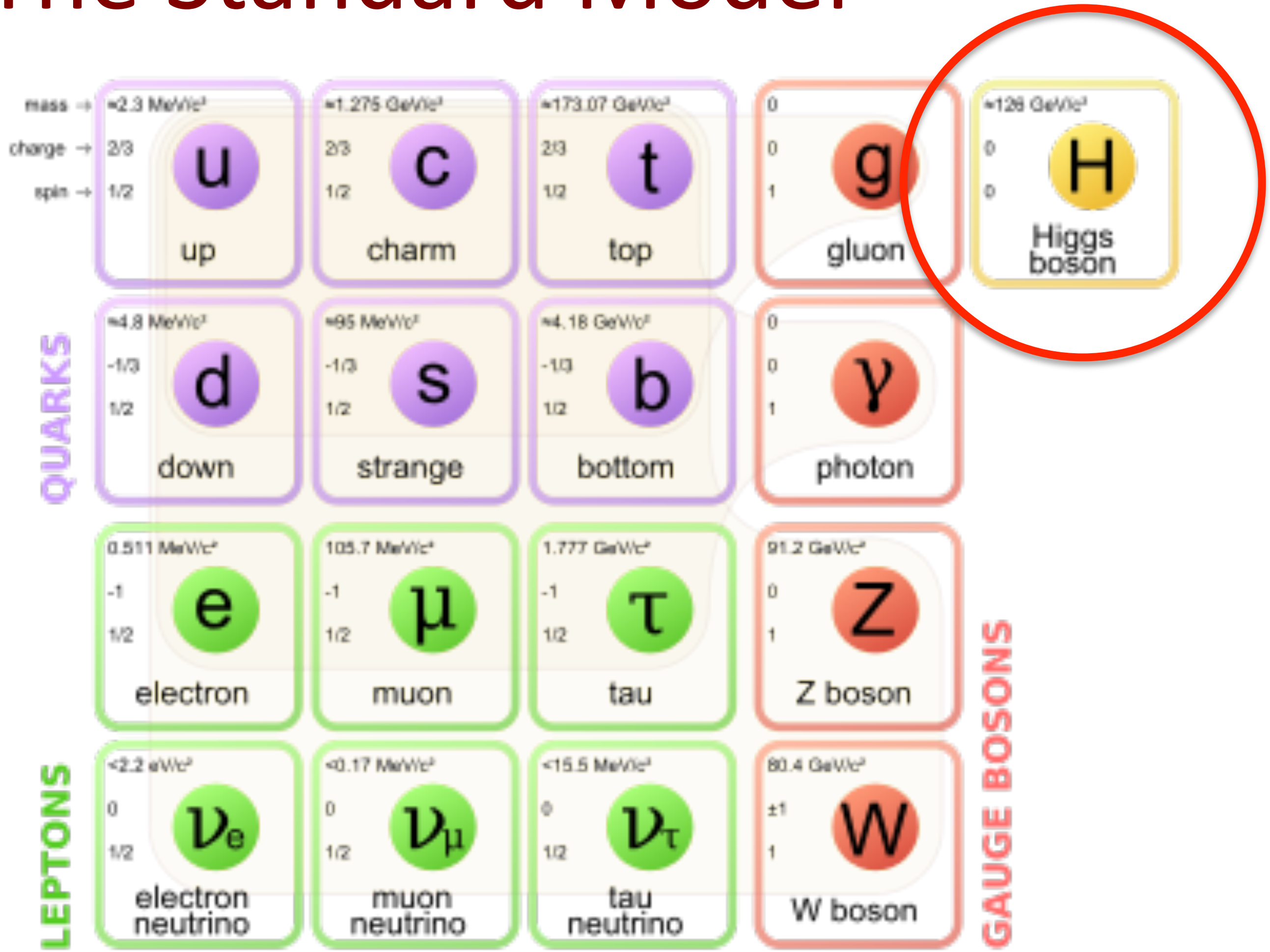


"hermetic" or "4pi" detector, designed to completely surround the collision point and measure every particle coming from it.

Constructed in layers each designed to do a different job. Different types of particles will make it different distances through the layers, allowing them to be identified by their "signatures" in the various sub-detectors.



The Standard Model



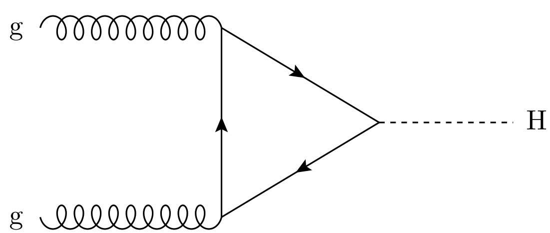
- Mathematics-based theory to explain the building blocks of the universe
- Collects together different types of particles
- All particles discovered over last ~80 years of work – the most recent discovery was in 2012 at the LHC – The Higgs boson.

Searching for the Standard Model Higgs

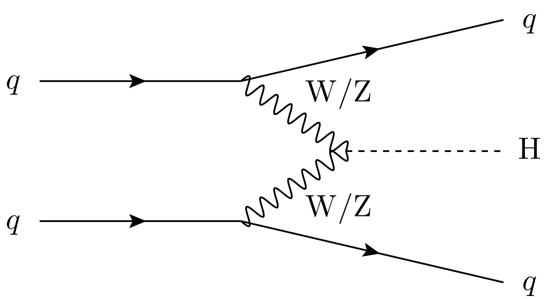
Searching for the Higgs boson requires information about how it is produced and how it decays.

PRODUCTION

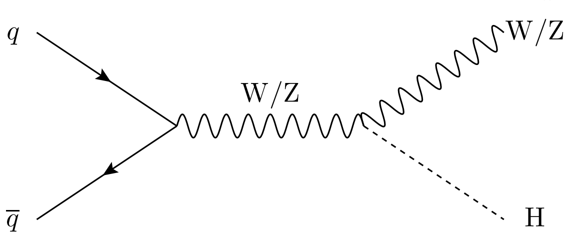
Protons contain quarks and gluons, allowing for the following 3 dominant methods of production:



"gluon fusion"

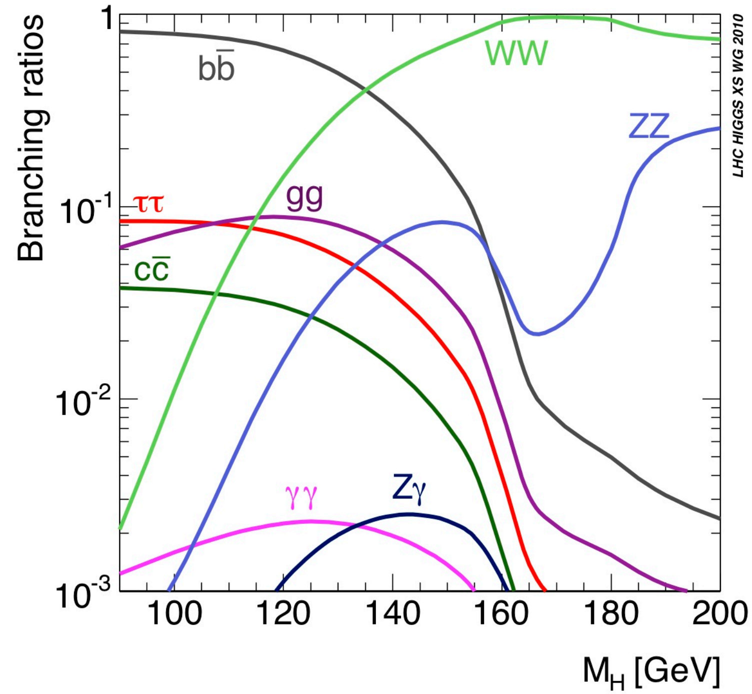


"vector boson fusion (VBF)"

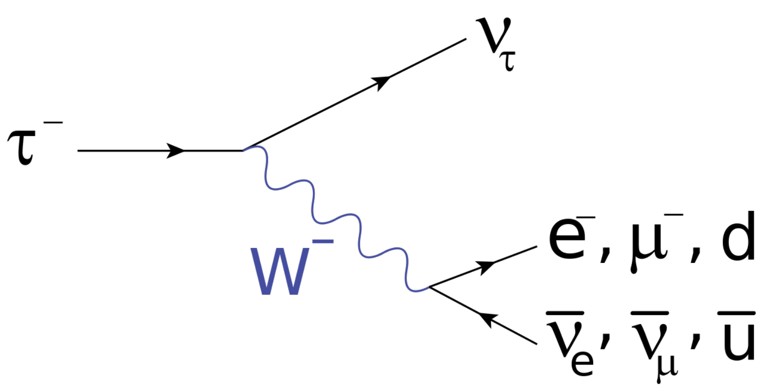


"associated production"

DECAY



How to search for H→ττ at CMS



τ_h = hadronic tau decay

Possible final states: **$e\tau_h$ $\mu\tau_h$ $e\mu$ $\tau_h\tau_h$ $\mu\mu$** and **ee**

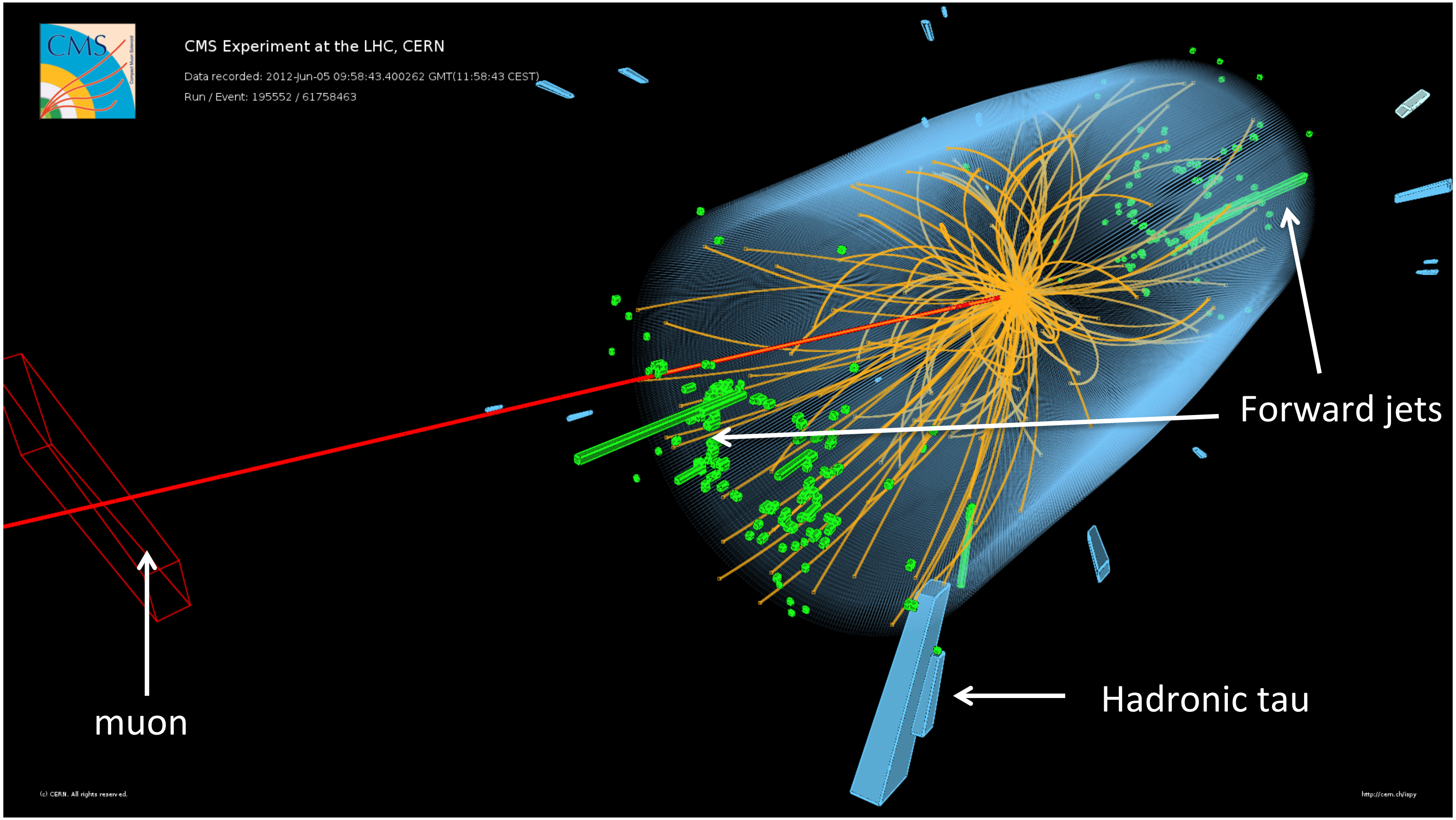
Studied at IC

→ Split events into 6 **"channels"** for final states

→Each final state will have different background compositions – thus the selection is optimised separately for each channel, with common properties across channels where possible.

For this project: focus on the **$\mu\tau_h$ channel**

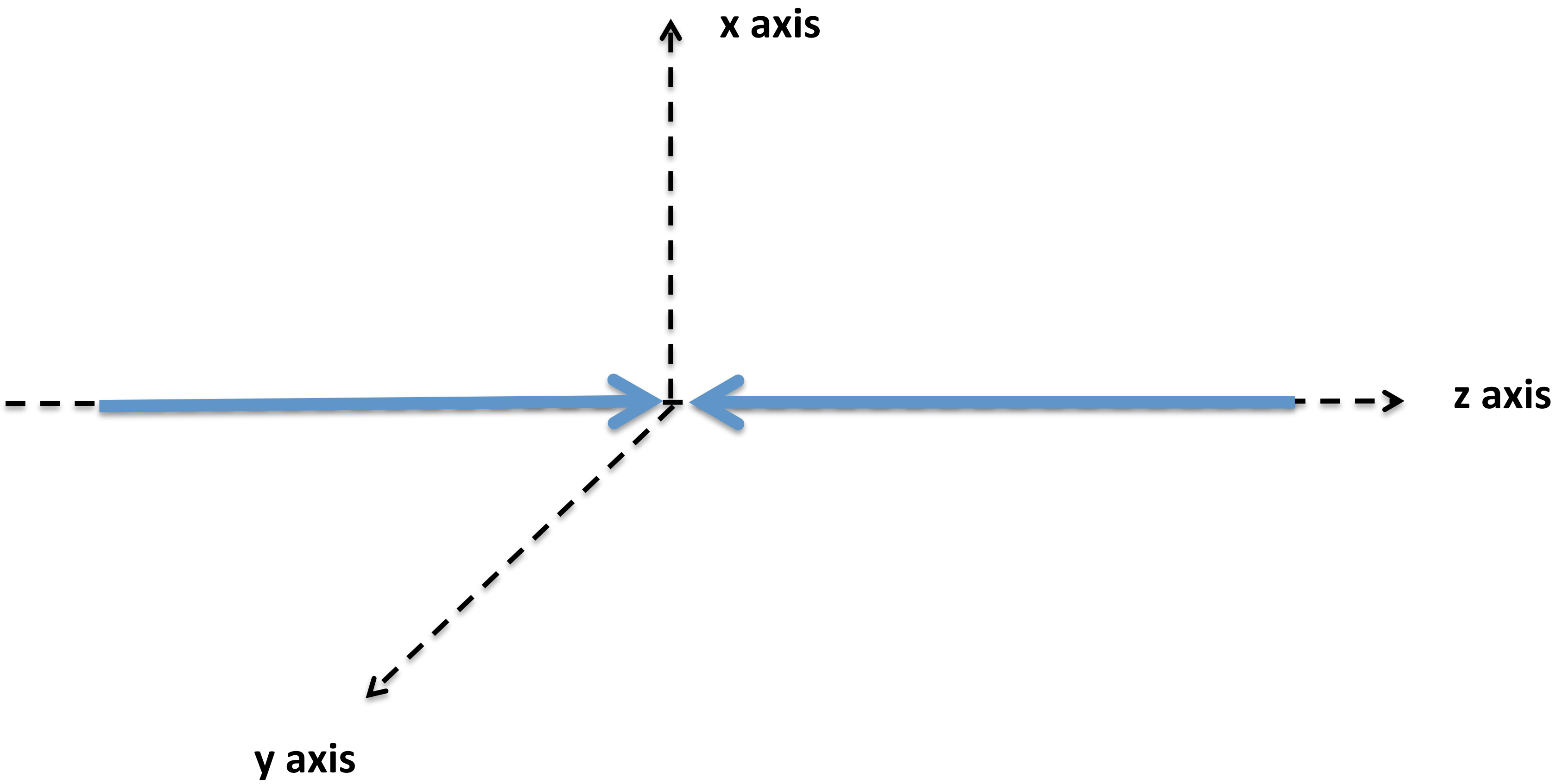
How do we find these events?



Look for "signatures" consistent with the expected production and decay – this is a VBF mu+tau event

Tricks of the trade

- Use of transverse variables and detector geometry



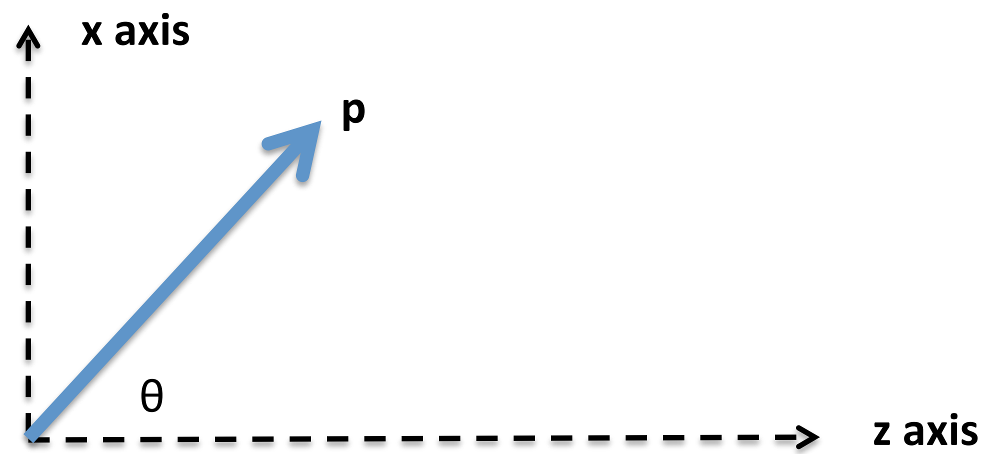
Tricks of the trade

- Use of transverse variables and detector geometry

Scattering angle θ

Conventionally given in terms of pseudorapidity:

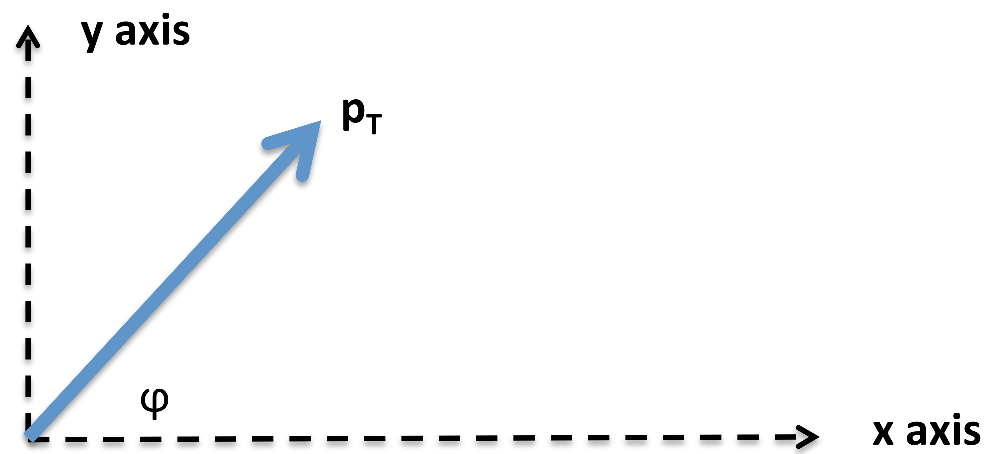
$\eta = -\ln(\tan(\theta)/2)$



Tricks of the trade

- Use of transverse variables and detector geometry

"Transverse" plane



→Conventionally we use p_T, η, φ, E as coordinates rather than p_x , p_y , p_z , E

→Can you think of a reason why transverse variables are more useful?

Tricks of the trade

- Knowing what other "background" processes we could see – use of Monte Carlo simulation

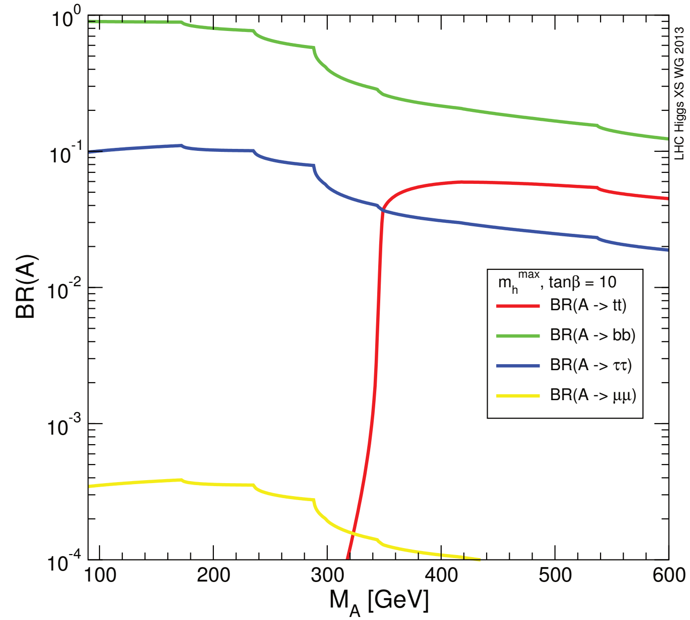
- Also knowing what our signal would look like if we found it

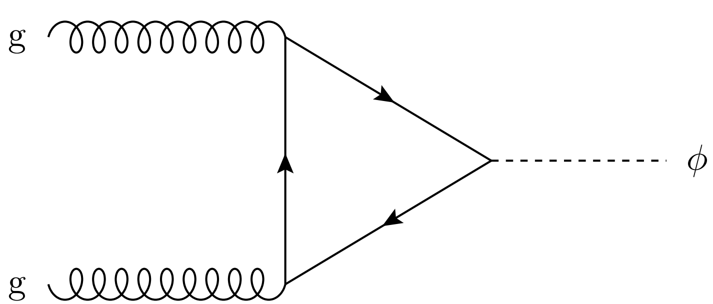
- Computer generated events which:
  - Require understanding of the structure of the protons and the possible quark and gluon interactions
  - Require understanding of the possible decays of any particles which could be created from this "hard scatter"
  - Includes additional soft scatters, or "pileup" which can result from other protons in the bunches
  - Simulates the detector itself and what signatures these interactions could give inside it.

Beyond the SM Higgs

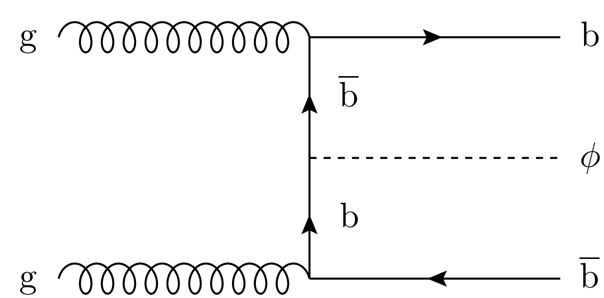
- Supersymmetry: popular BSM theory predicting new particles at heavier masses.
- The Higgs sector consists of 2 Higgs doublets → resulting in 5 physical Higgs states: **$H^{\pm}$, h, A, H**.
- Of the neutral bosons, A is pseudo scalar and **H and h are scalar** (and in certain MSSM models **can be SM H-like**).
- Different dominant production modes

Possible higher BR to taus:





gg→φ (similar to SM)



gg→bbφ

Format of collision data

- Data collected by CMS (and the Monte Carlo samples used to simulate it) are vast in size and stored at multiple different sites around the world.
- The data is stored in a common format called ROOT.
- ROOT is an enormous C++ class allowing for manipulation of this data in ways which are commonly useful to particle physicists:
  - Data is typically stored in "trees" where one entry is one collision and each "branch" of the tree is a different properly of the event (e.g. number of hits in the ECAL, p_T of the highest track…)
  - There are different possible stages of this data, starting from the raw detector information and ending with fully reconstructed candidate particles. I am going to give you the data in the format such that the candidates have already been reconstructed for you, so you will see things like p_T of muon candidate 1 etc.
- Ultimately the challenge for all CMS physicists is to process this data in an efficient way and define a search strategy for whatever signature we are looking for, whilst trying to minimise the number of background events we pick up.

The Projects

1) Basic ROOT – get used to the format of particle physics data by performing some simple tasks using Higgs signal Monte Carlo. Using the Monte Carlo sample for a type of Higgs signal, figure out how to design a selection in order to select the most signal-like di-tau pairs. Plot the invariant mass of the pair you have selected.

2) Having designed your selection, I will give you a new file containing Monte Carlo for one of the major backgrounds to this search. Try varying some requirements in your selection (pt, ID and isolation) in order to maximise the number of selected signal events selected compared with background.

3) Investigate additional cuts related to other event information which can help to enhance the selection of signal over background.

4) Depending on time, see if you can use machine learning to do an even better job than you can do by yourself!

Information

Instructions and ROOT files can always be found here:

http://www.hep.ph.ic.ac.uk/~rcl11/FirstYearProject/

We will discuss details by email as much as possible – since this is a computing based project it is often useful if you send me snippets of code when you have problems so I can test them out myself.

Feel free to grab me in my office if I am there (I am often not though, so email is your friend if you can't find me in person ☺)