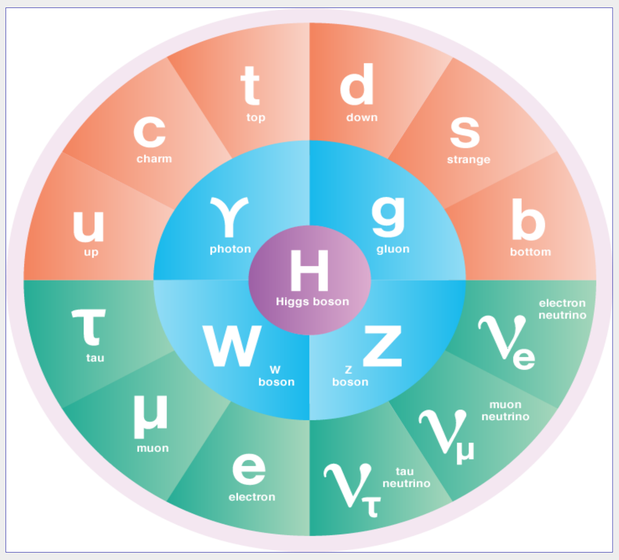
Identification of b quark jets at the CMS Experiment in the LHC

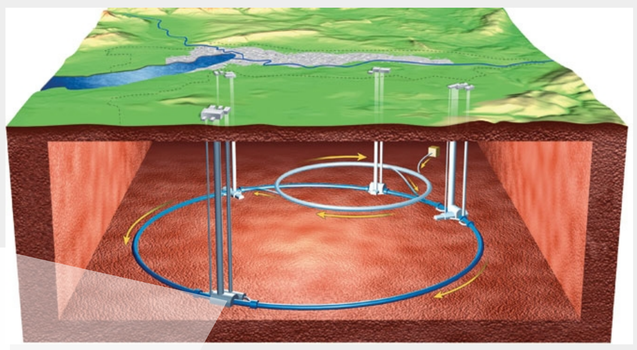
Kevin Deroover on behalf of the CMS collaboration
kevin.deroover@cern.ch

Many new physics searches as well as Standard Model measurements rely on the accurate identification of jets originating from b quarks

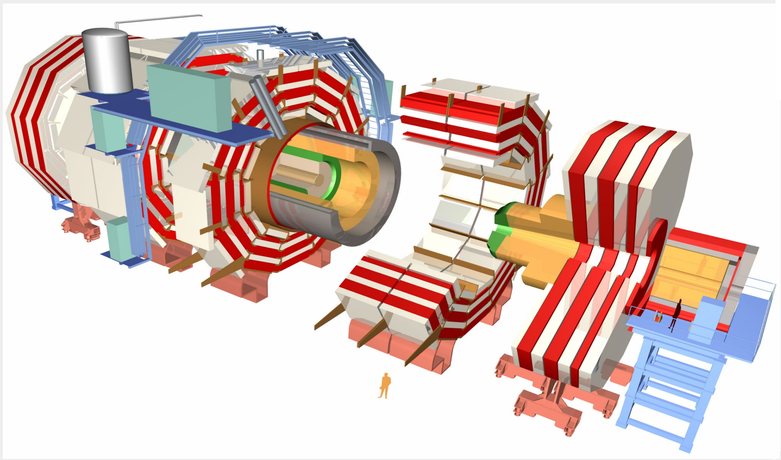


The **Standard Model** of particle physics describes our current understanding of what the most fundamental building blocks of nature are and how they interact.

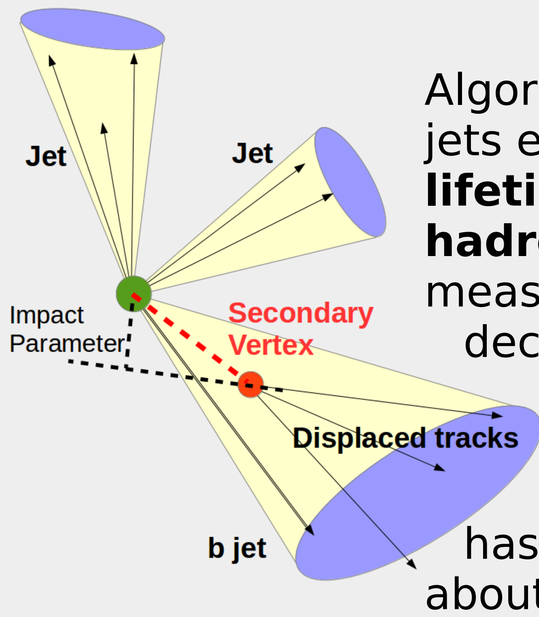
Precise measurements of the Standard Model, as well as searches for new physics, are performed at the **Large Hadron Collider** (LHC). The LHC collides protons at the highest energies (13 TeV) ever.



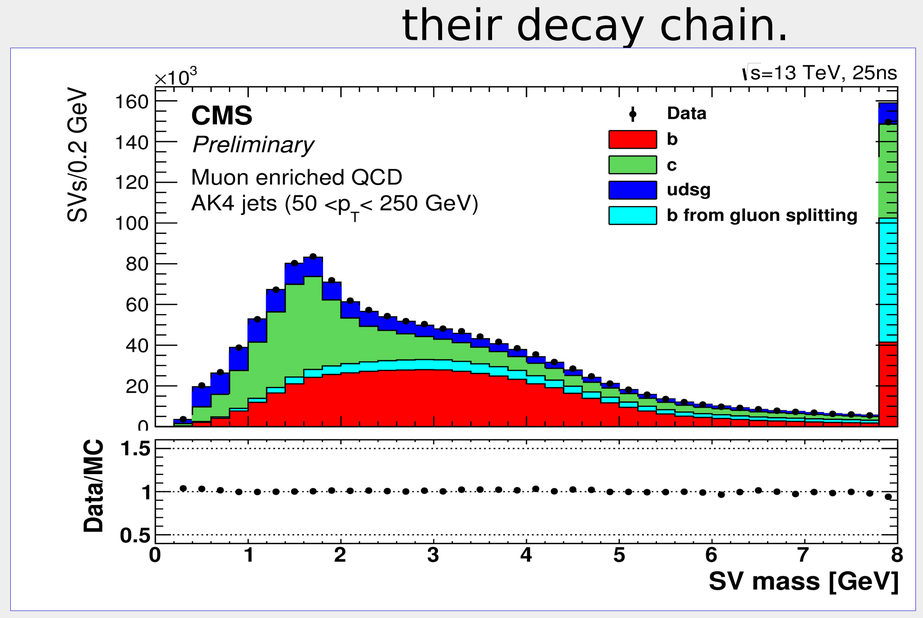
With large detectors like the **Compact Muon Solenoid**, we detect the particles that emerge from the collisions.



Quarks hadronize and produce a jet of particles



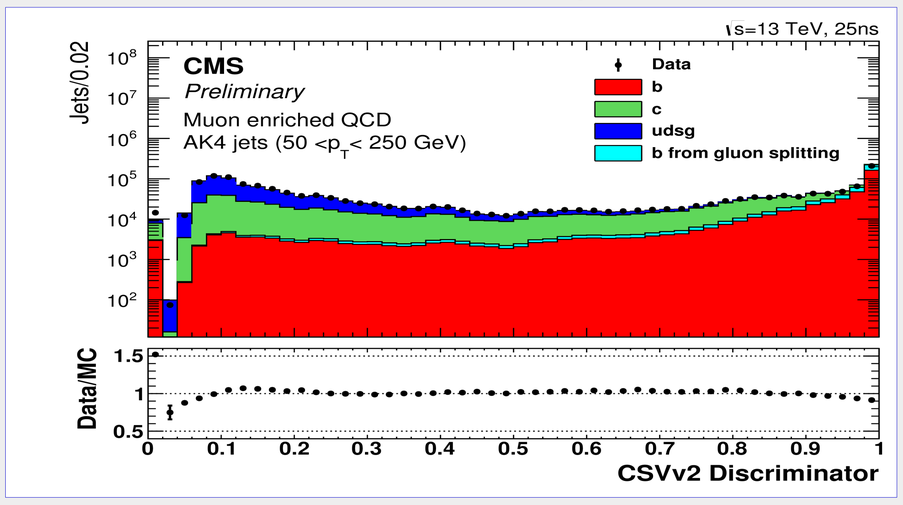
Algorithms to identify b jets exploit the **long lifetime of the B hadron**, resulting in a measurable **displaced** decay with respect to the primary interaction vertex. A B hadron also has a probability of about 20% to have a **muon or electron** in their decay chain.



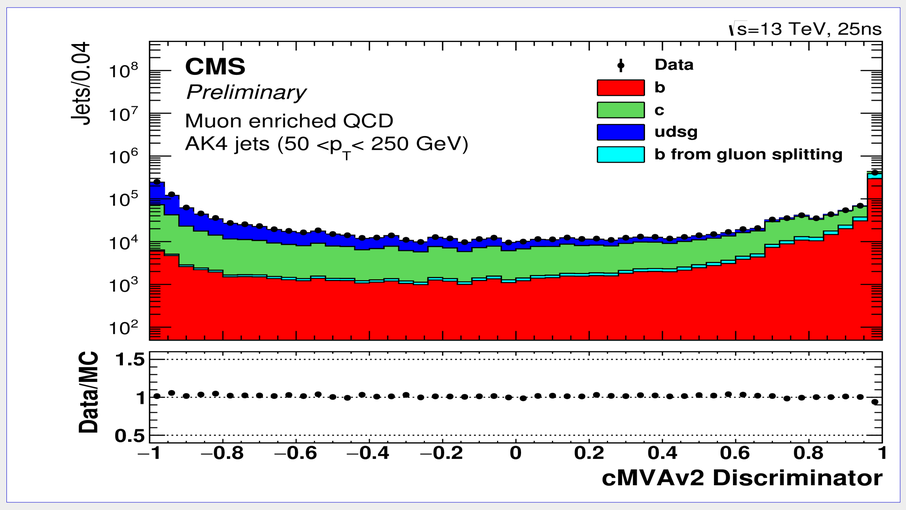
The mass of a reconstructed **secondary vertex** is one of the most powerful discriminating variables.

Algorithms for b jet identification

- Combined Secondary Vertex (CSV): Combines track- and secondary vertex variables in a multivariate way.



- Combined MVA (cMVA): Combines, using boosted decision trees, various discriminators including CSVv2 and exploiting soft muon and electron information.

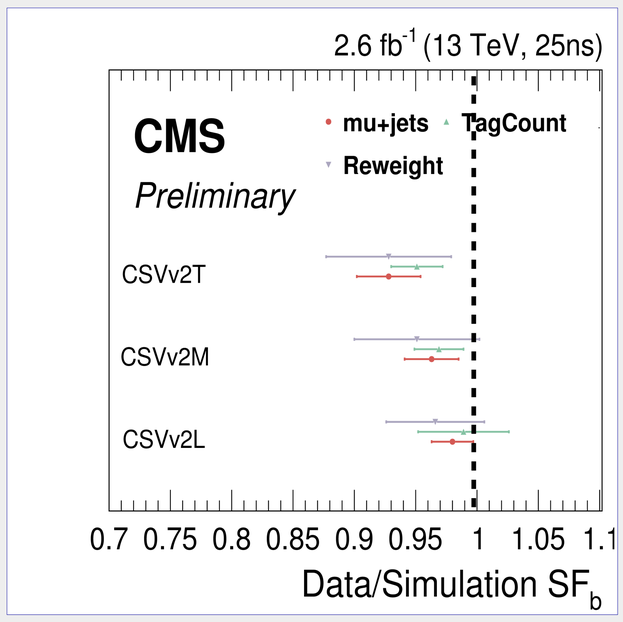


Simulations do not describe the data perfectly → Scale factors!

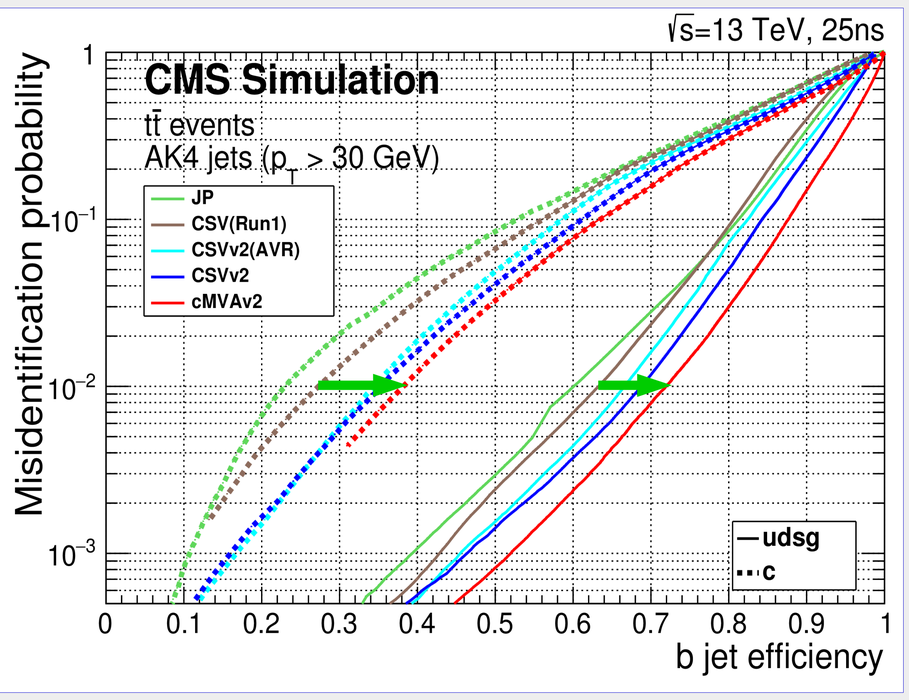
Scale factors need to be calculated to correct simulated events for the observed **discrepancies** between data and simulations. This scale factor is measured for a certain threshold on the discriminator as

$$SF_b = \frac{\epsilon_b^{data}}{\epsilon_b^{sim}}$$

The efficiency for correctly identifying b jets ϵ_b in data is measured with various techniques for three working points of the discriminator (Loose, Medium & Tight). The selected sample of jets for these **measurements are enriched** in b jets.



A direct comparison of the algorithms on simulated top quark pair events shows the b jet identification algorithms have **improved significantly** since LHC Run 1. The cMVA algorithm even has an **absolute improvement of 8%**, achieving about 72% b jet efficiency for 1% misidentification probability.



We can increase the sensitivity of precision measurements and enhance the discovery potential of many new physics searches!

Reference: “Identification of b quark jets at the CMS Experiment in the LHC Run 2”
BTV-15-001