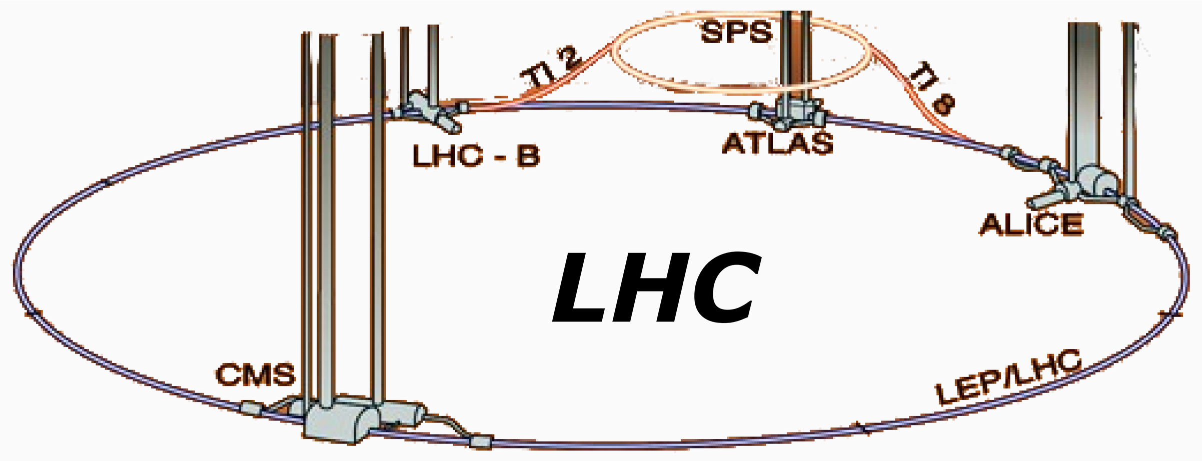
Identification of c quark jets at the CMS experiment

Seth Moortgat on behalf of the CMS Collaboration
seth.moortgat@cern.ch



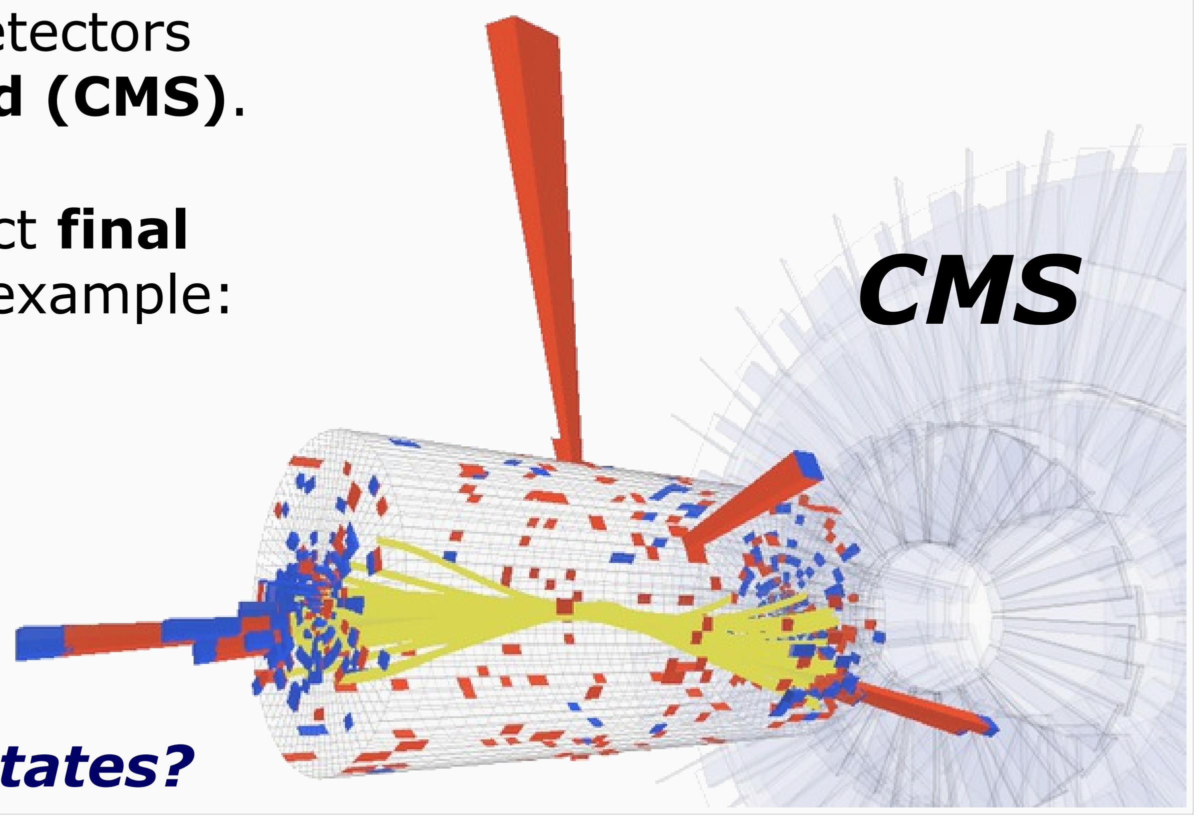
1. Introduction

The **Large Hadron Collider (LHC)** is a particle accelerator that collides protons at a center-of-mass energy of up to 13 TeV.

The particles that emerge from these collisions are detected by large detectors like the **Compact Muon Solenoid (CMS)**.

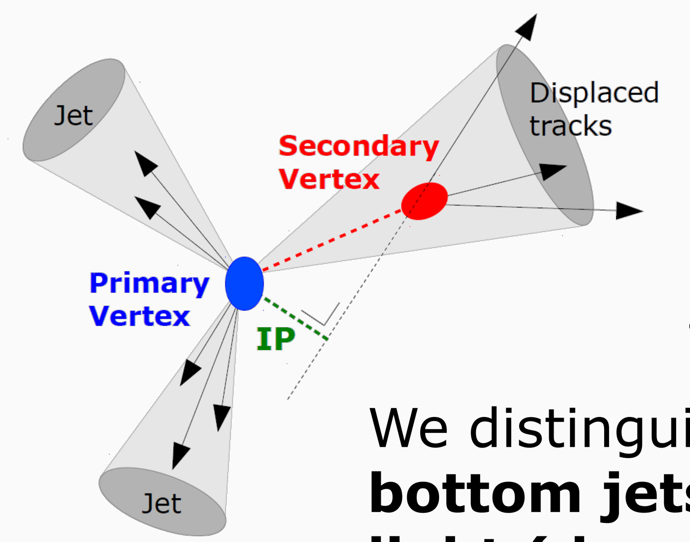
Many (new) physics models predict **final states with charm quarks,** for example:

$$SUSY: \tilde{t} \rightarrow \tilde{\chi}^0 + c$$
$$FCNC: \ t \rightarrow c + Z/\gamma/\cdots$$
$$Higgs: \ h \rightarrow c + \bar{c}$$



Can we identify the charm quarks in these final states?

2. Heavy-flavor taggers



Hadronic particles undergo a fragmentation process and create **particle showers ('jets')** inside the CMS detector.

We distinguish three kinds of jets: **bottom jets, charm jets and light (down, up, strange or gluon) jets.**

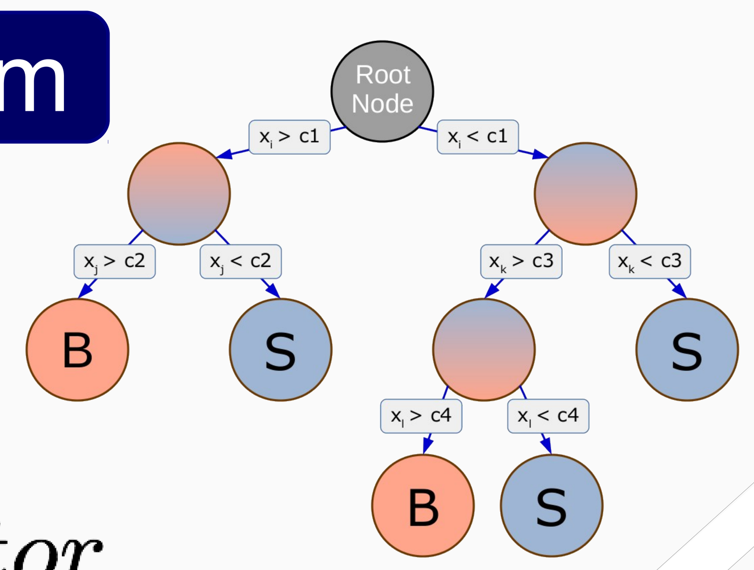
The charm-tagging algorithm relies on the lifetime and (semi-)leptonic decays of hadrons containing charm quarks. We distinguish charm jets from bottom jets (more displaced vertices and tracks) and from light jets (no displacement).

Therefore we exploit jet properties related to:

Secondary vertices | **Displaced Tracks** | **Soft leptons**

3. A machine learning algorithm

- **Learn the properties of charm jets** by giving the algorithm labeled examples from simulated collision events.
- Make predictions on unlabeled data.



$$\mathbb{R}^d \rightarrow \mathbb{R}: \{\vec{x}\} \mapsto Discriminator$$

d = dimensionality of the input-variable space
= number of jet properties used in the training

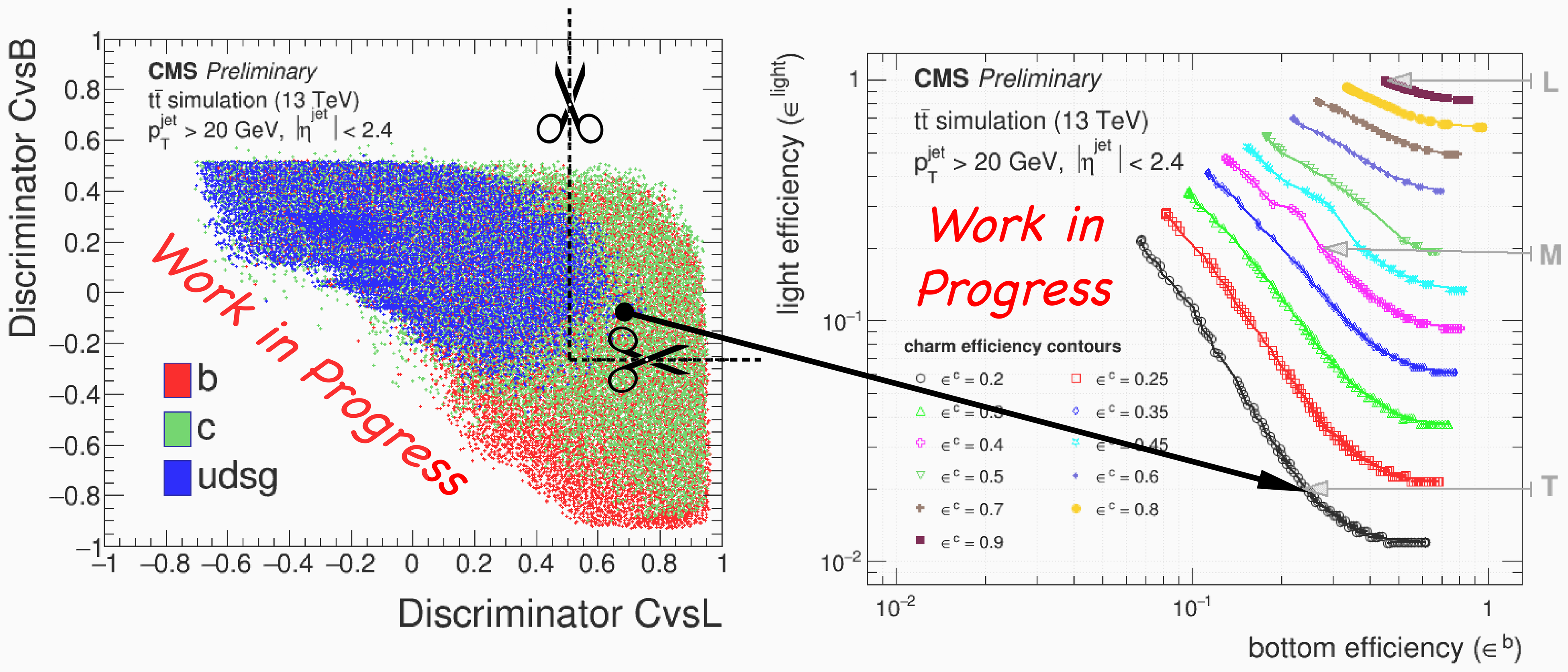
We use a **boosted decision tree (BDT)**
→ consecutive cuts split nodes into final leaves
→ penalize misclassified jets in next tree

4. Performance on simulation

Use jets from simulated top quark pair ($t\bar{t}$) events in the trained charm tagger BDT and build **discriminator distributions** for the different jet flavors.

- **CvsB:** distinguish charm jets from bottom jets
- **CvsL:** distinguish charm jets from light jets

Apply a 2D selection criterion to isolate charm jets. Each threshold has a **charm selection efficiency** (ε) and a light or bottom **mistag efficiency.**



5. Scale factors: data ↔ simulation

Define three **working points**:

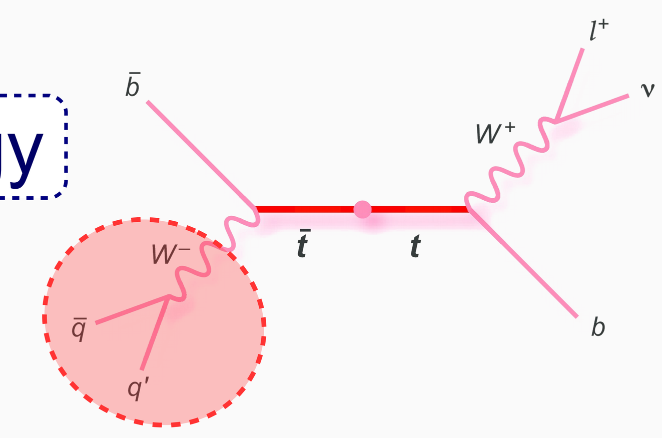
Work in Progress

WP	ε^c	ε^b	ε^{light}
Loose (L)	0.90	0.45	0.99
Medium (M)	0.39	0.26	0.19
Tight (T)	0.20	0.24	0.02

$$SF \equiv \frac{\epsilon(DATA)}{\epsilon(SIM)}$$

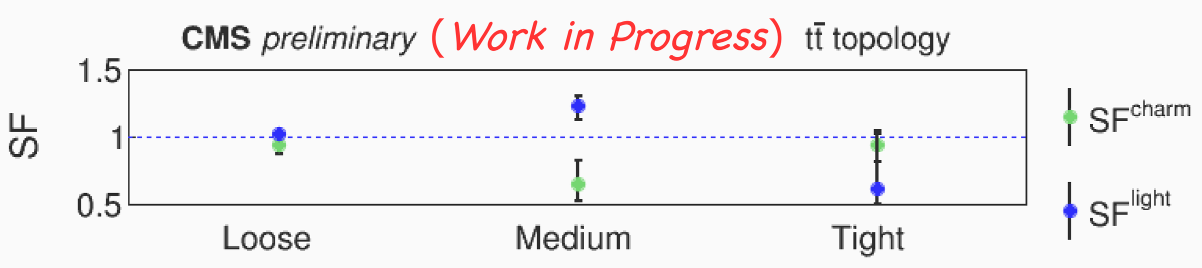
Measure possible discrepancies between simulations and real collision data.

Top-quark pair topology

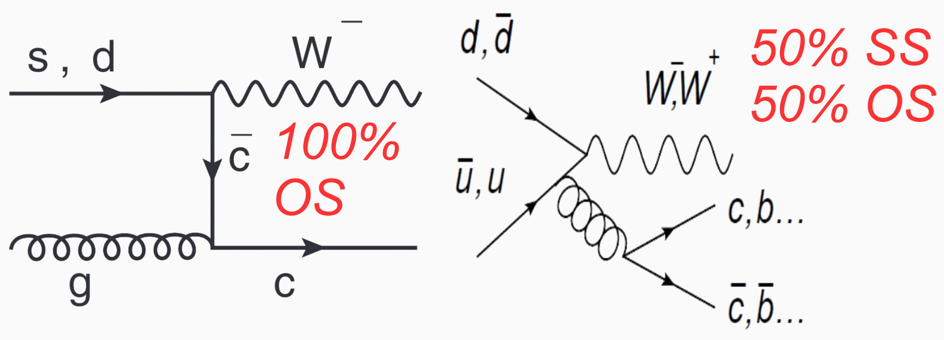


Select jets from hadronically decaying W bosons in semi-leptonic $t\bar{t}$ events. Hadronic W boson decays are known to yield ~25% charm quarks.

→ Fit to a likelihood-discriminant to measure SF_{charm} and SF_{light}.



W + charm topology



Select W+charm events exploiting the opposite electric charge of the W boson and the charm quark (OS). Backgrounds have 50% opposite sign and 50% same sign (SS) electric charge.

6. Conclusion and outlook

- **First charm tagger** for the CMS experiment available
- A **more performant version** has already been developed for the next data taking period
- Looking forward to applications in **future CMS analyses**

"Identification of c quark jets at the CMS experiment", CMS PAS-16-001 in preparation
In collaboration with Juan Pablo Fernandez, Kittikul Kovitanggoon, Gerrit Van Onsem, Mauro Verzetti and the CMS Collaboration