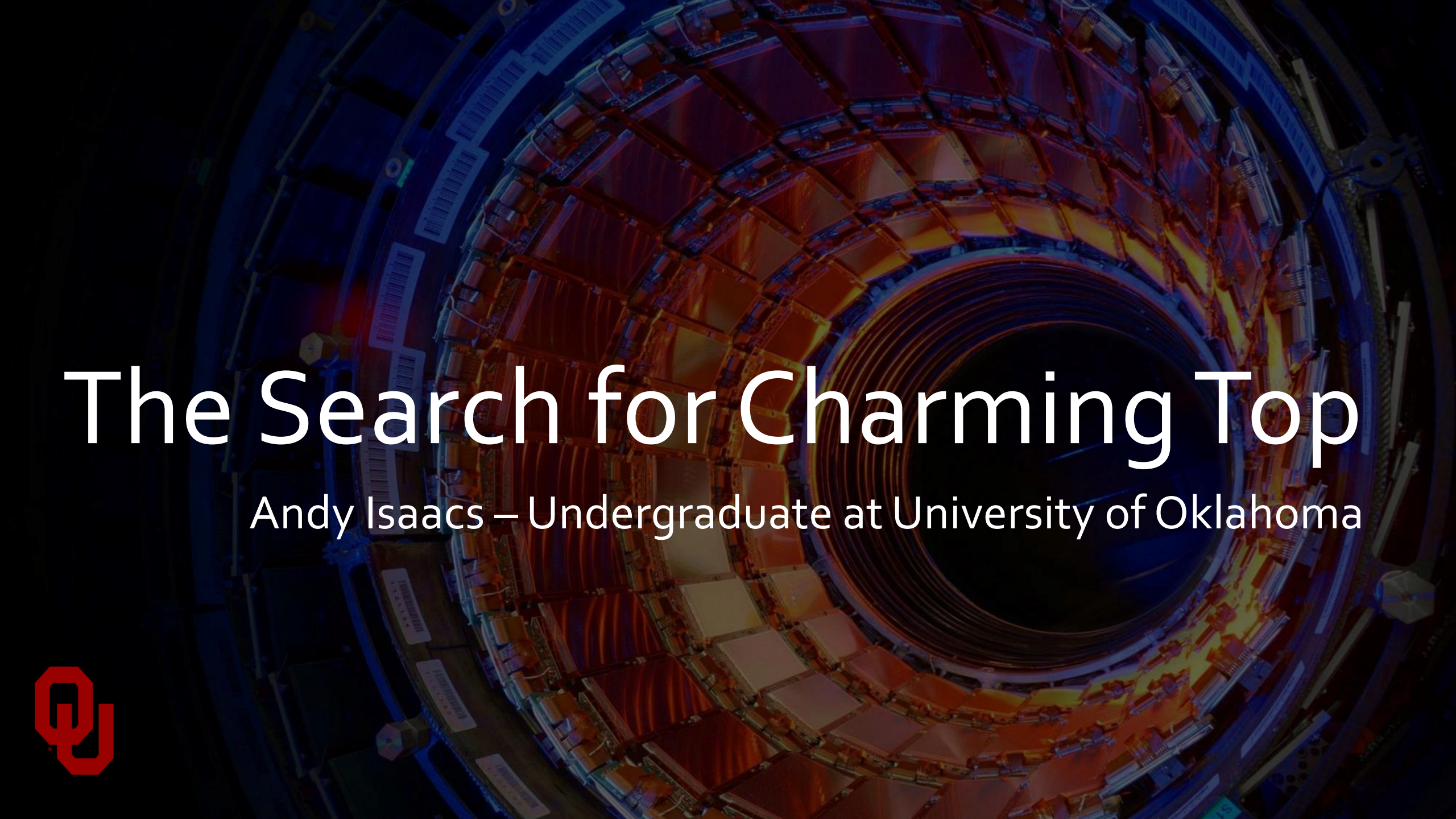
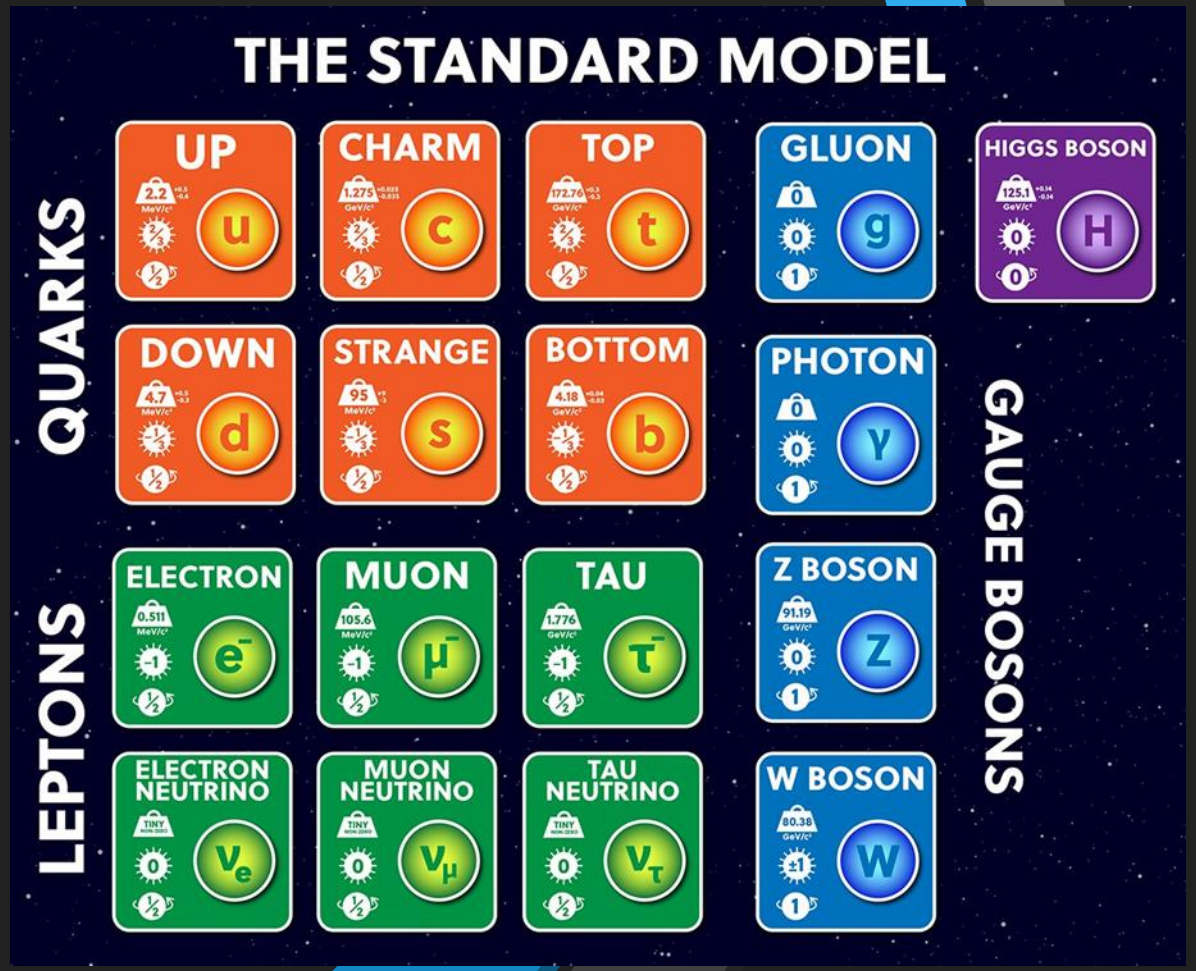


The Search for Charming Top

Andy Isaacs – Undergraduate at University of Oklahoma

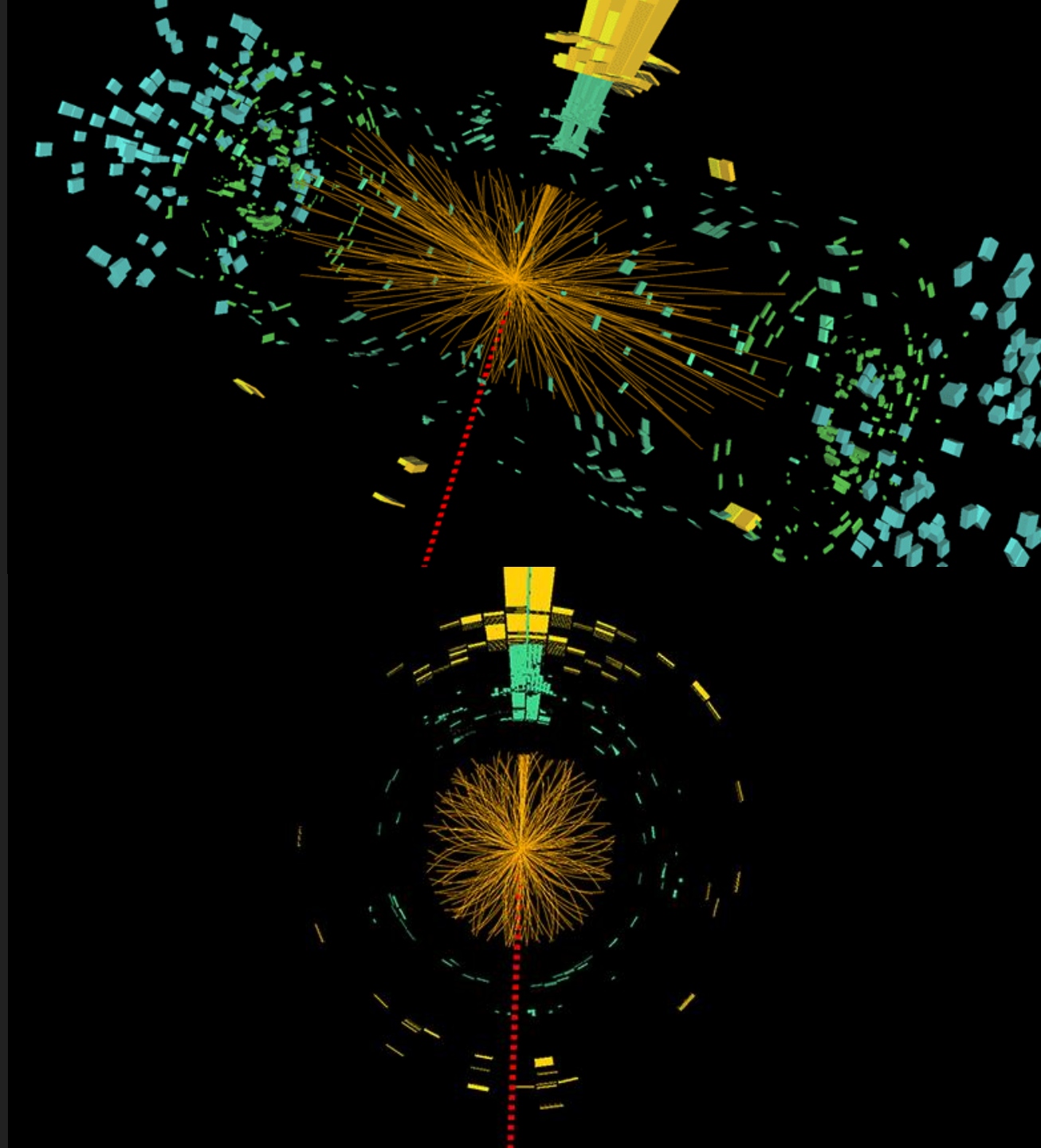
Standard Model

- Separates known particles into categories
- Quarks and Leptons carry matter
- Gauge Bosons carry force
- Interaction between matter particles is an exchange of force carrying boson
- SM provides cross-section - probability that a certain process can occur
- Tested by colliding hadrons to observe resulting matter & force particles



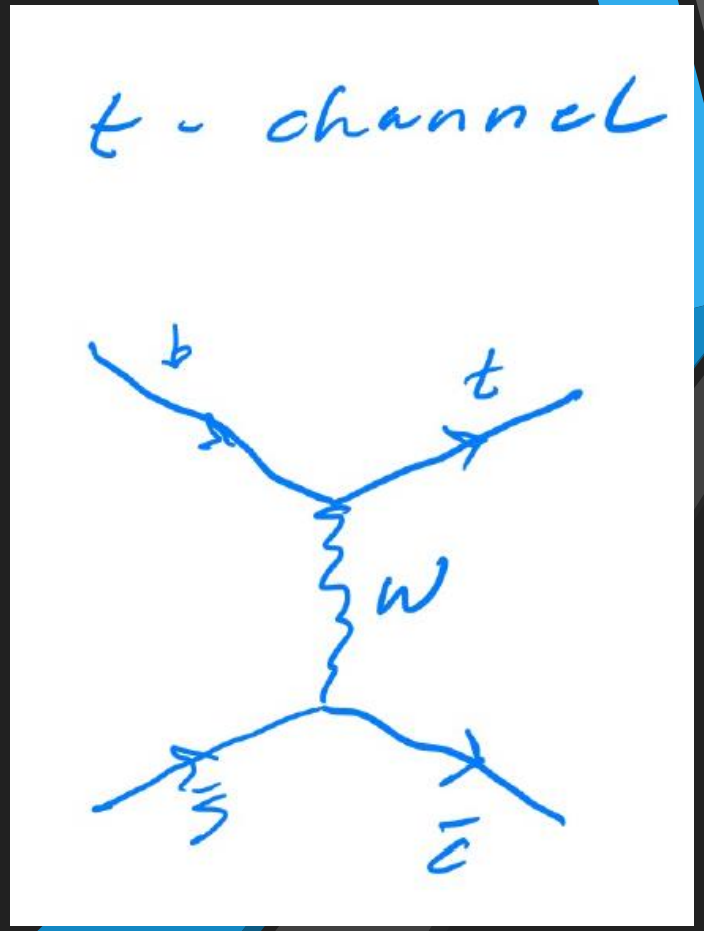
LHC & ATLAS Detector

- Protons collide at 13.6 TeV inside ATLAS detector
- Resulting showers of quarks, leptons, & bosons are recorded
- Additional information including mass, transverse momentum, eta, phi, missing transverse energy, etc.
- Tools to relate physical world with theoretical models
- SM predicts cross section that certain channels will occur based on particles involved



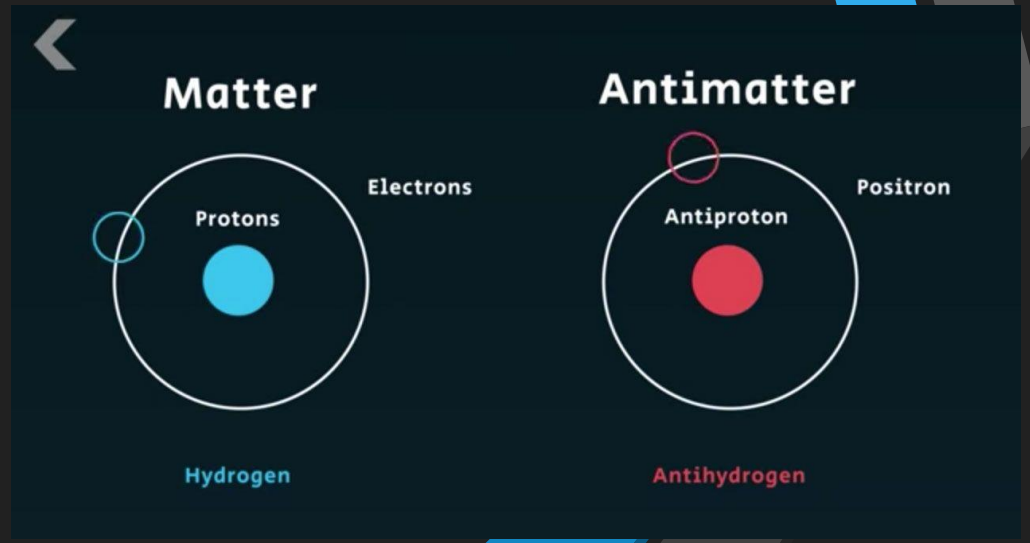
Charming Top

- Collisions resulting in the emission of a top quark and a charm quark (with exchange of W boson) have not been observed
- Observation is accepted at a significance of 5 sigma - statistical analysis quantity that relates signal to background ($S/\sqrt{B}$)
- Massive data from ATLAS needs to separate charming top from everything else that could resemble it
- How likely is it that background fluctuated to look like signal?
- If < 0.00003%, charming top has been observed



Why Is This Important?

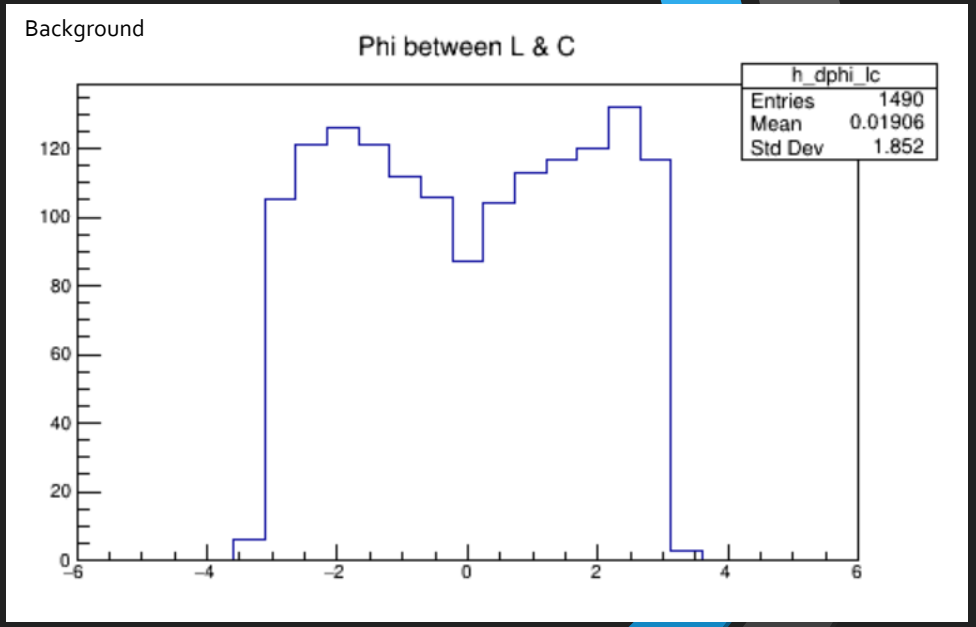
- Standard Model is a model – observations that disagree with predictions could point to new physics
- Why is Higgs Boson so light when interactions should make it heavy? Why does matter/antimatter asymmetry appear to exist?
- If experimental cross section points to new physics, this might help us understand previously unanswered questions



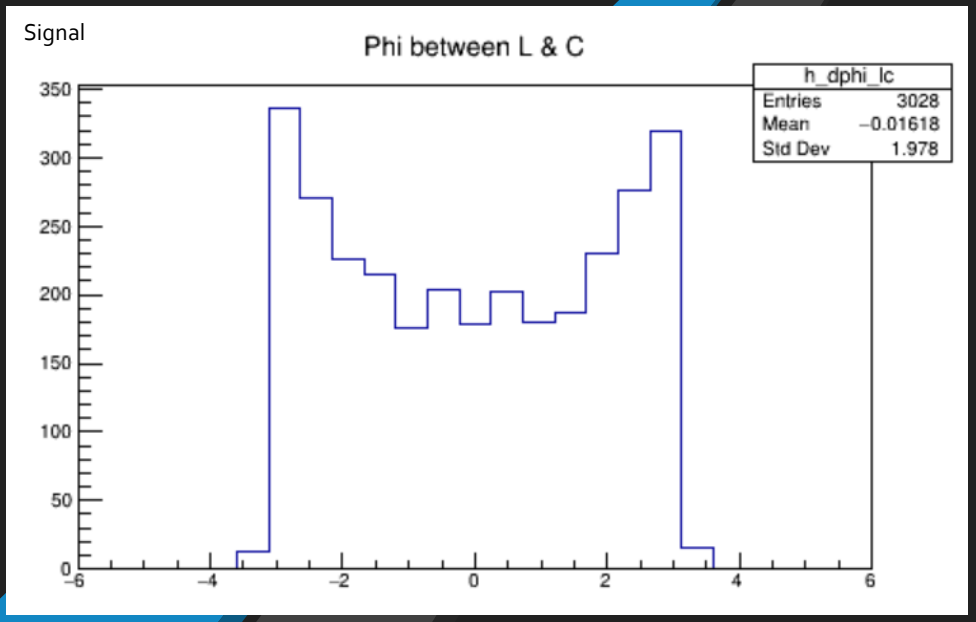
Data Analysis

- Monte Carlo generated data
- Histograms – type of plot that shows probability that variable exists at certain value
- Important to understand why peaks exist at certain points (mass of top)
- Amount of signal vs. background determines significance
- Must figure out how to isolate regions where signal dominates (cuts, shapes)

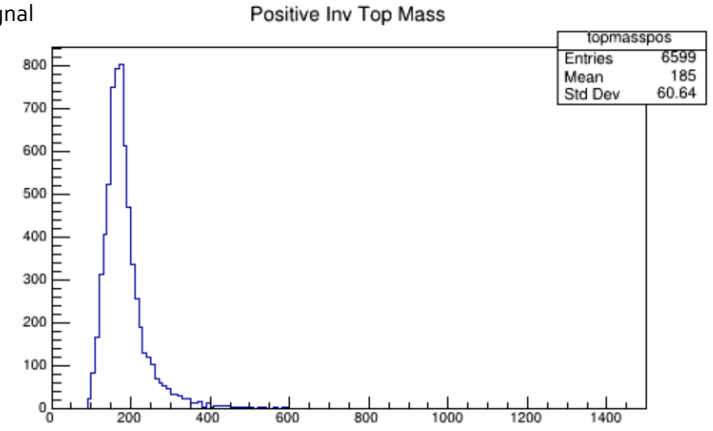
Background

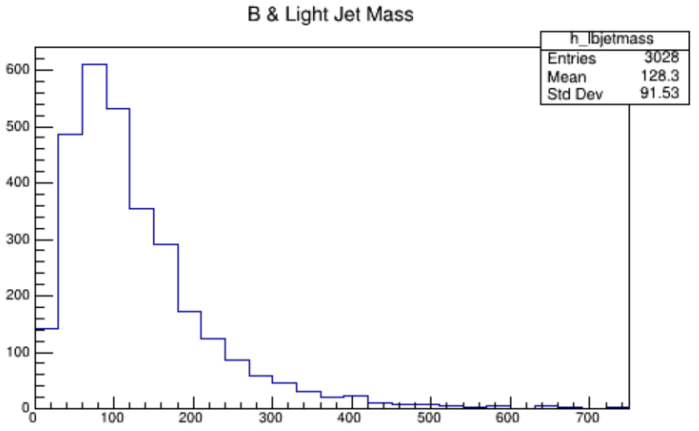


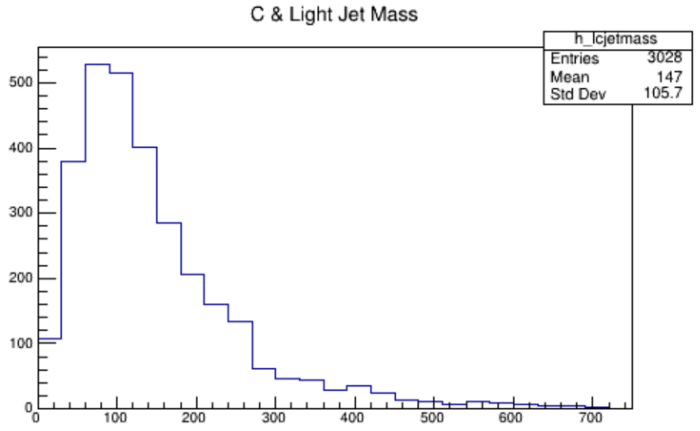
Signal

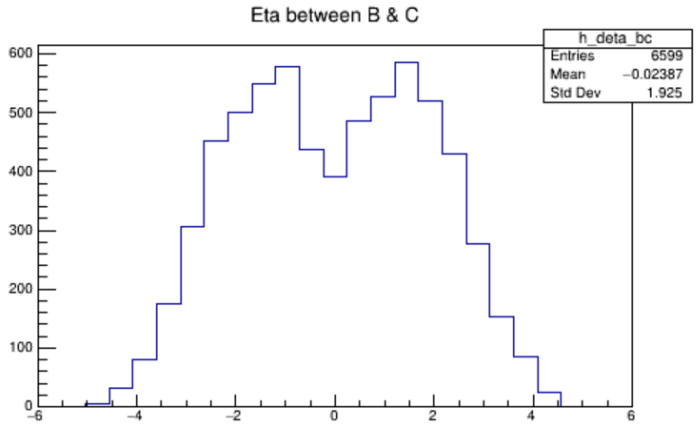


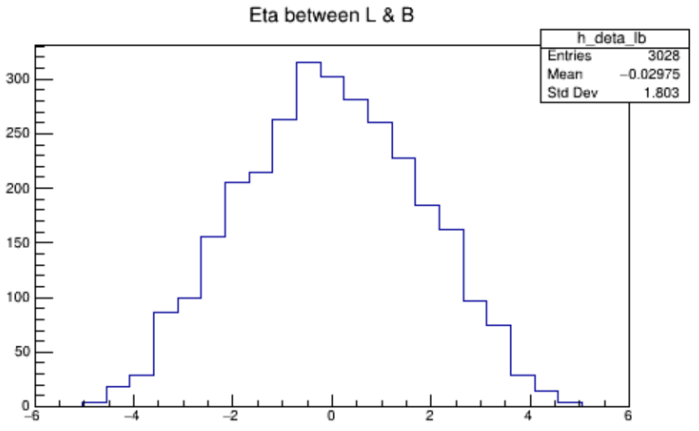
Signal

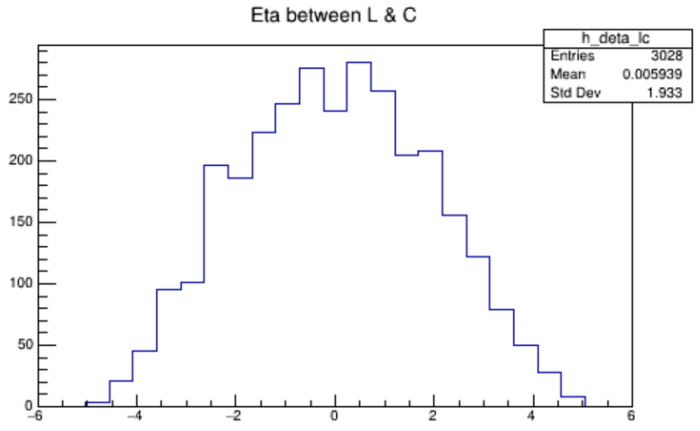

Positive Inv Top Mass
topmasspos
Entries 6599
Mean 185
Std Dev 60.64

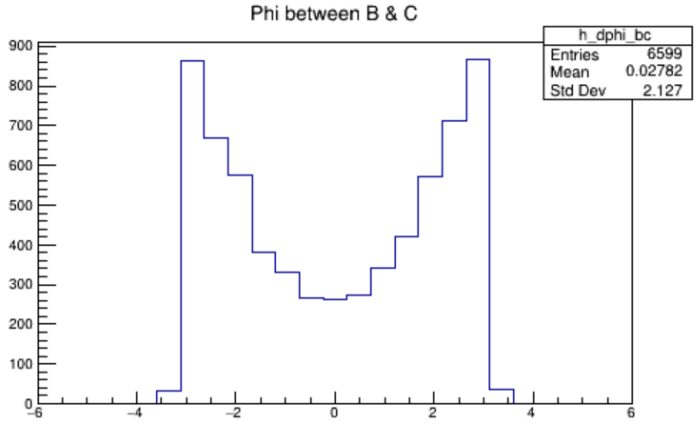

B & Light Jet Mass
h_lbjetmass
Entries 3028
Mean 128.3
Std Dev 91.53

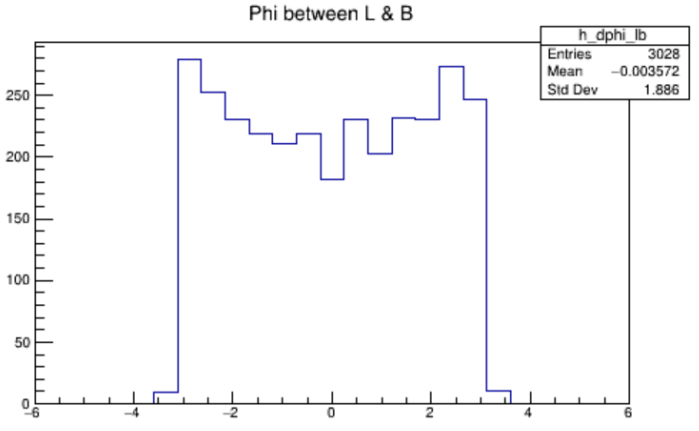

C & Light Jet Mass
h_lcjetmass
Entries 3028
Mean 147
Std Dev 105.7

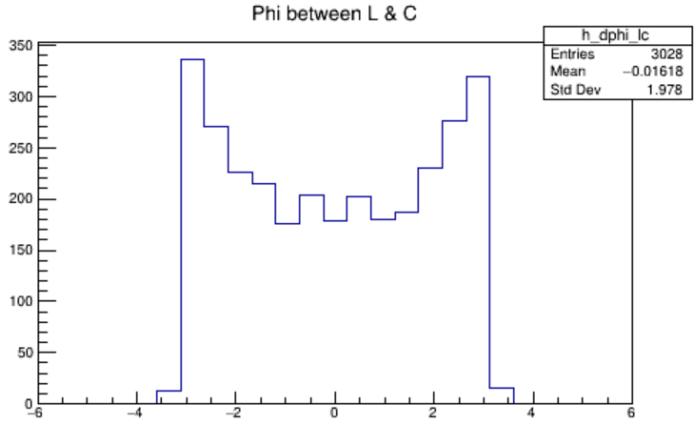

Eta between B & C
h_deta_bc
Entries 6599
Mean −0.02387
Std Dev 1.925


Eta between L & B
h_deta_lb
Entries 3028
Mean −0.02975
Std Dev 1.803

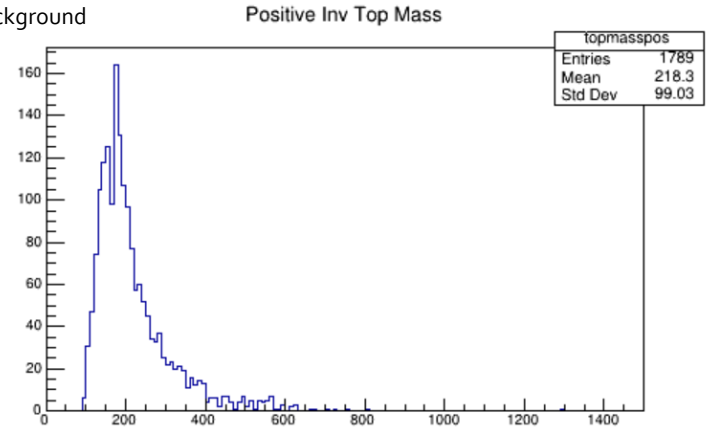

Eta between L & C
h_deta_lc
Entries 3028
Mean 0.005939
Std Dev 1.933

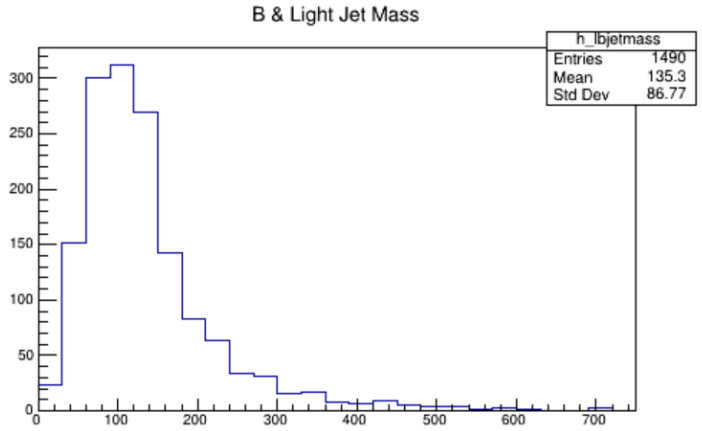

Phi between B & C
h_dphi_bc
Entries 6599
Mean 0.02782
Std Dev 2.127

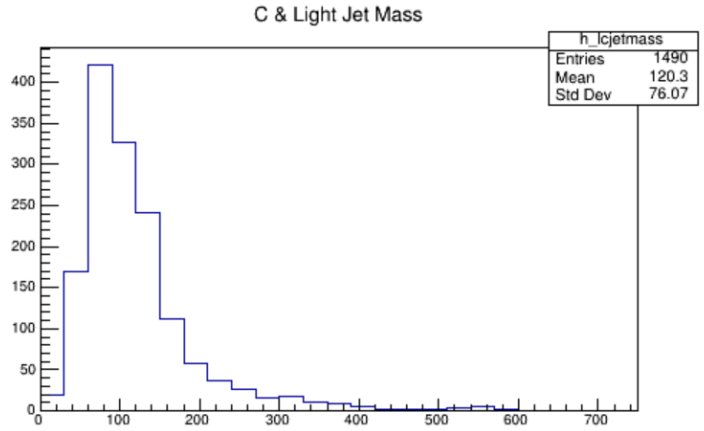

Phi between L & B
h_dphi_lb
Entries 3028
Mean −0.003572
Std Dev 1.886

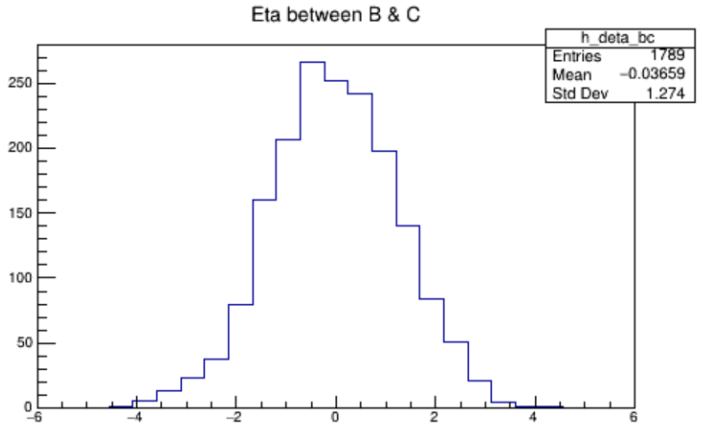

Phi between L & C
h_dphi_lc
Entries 3028
Mean −0.01618
Std Dev 1.978

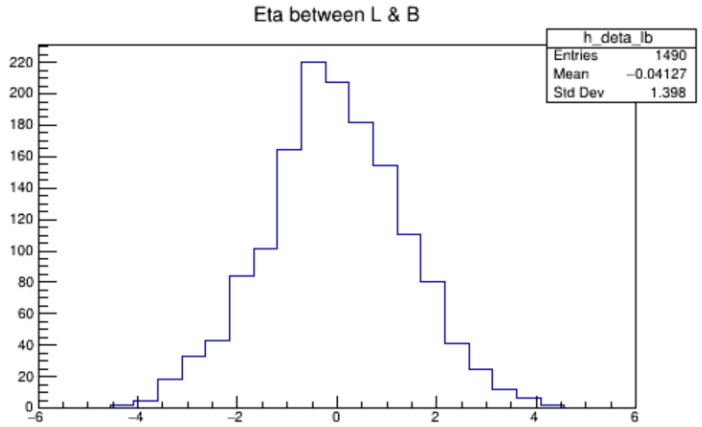
Background

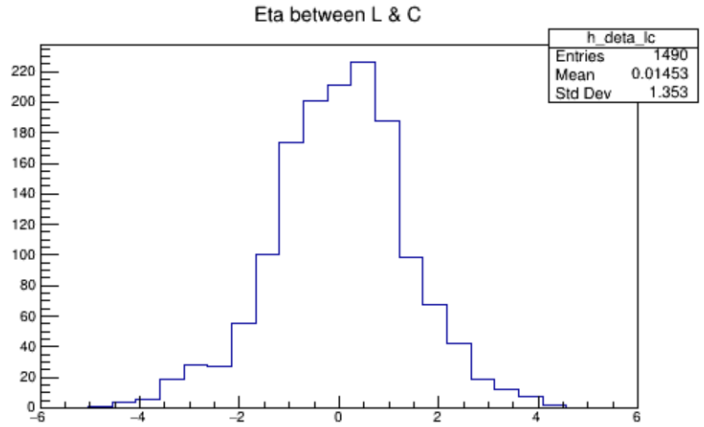

Positive Inv Top Mass
topmasspos
Entries 1789
Mean 218.3
Std Dev 99.03

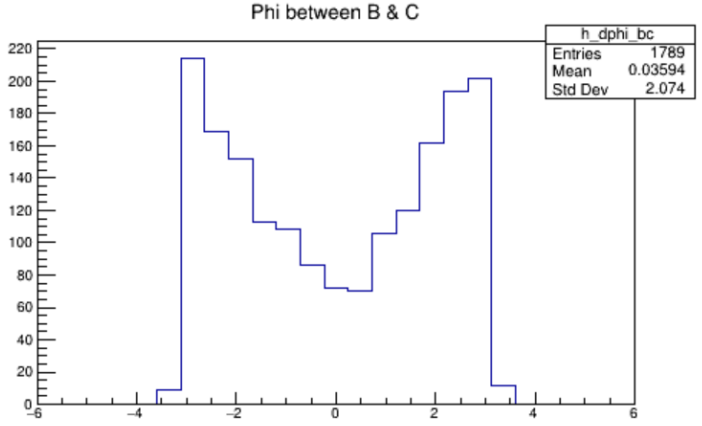

B & Light Jet Mass
h_lbjetmass
Entries 1490
Mean 135.3
Std Dev 86.77

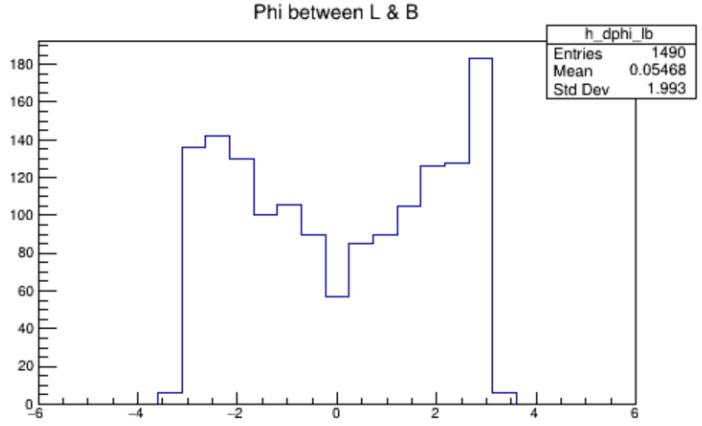

C & Light Jet Mass
h_lcjetmass
Entries 1490
Mean 120.3
Std Dev 76.07

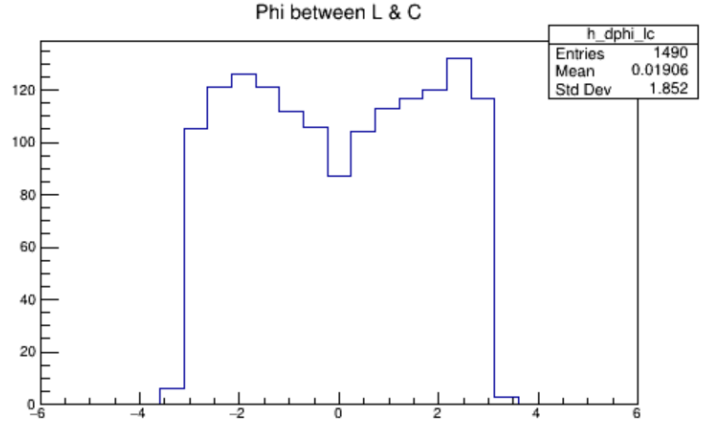

Eta between B & C
h_deta_bc
Entries 1789
Mean −0.03659
Std Dev 1.274


Eta between L & B
h_deta_lb
Entries 1490
Mean −0.04127
Std Dev 1.398


Eta between L & C
h_deta_lc
Entries 1490
Mean 0.01453
Std Dev 1.353


Phi between B & C
h_dphi_bc
Entries 1789
Mean 0.03594
Std Dev 2.074

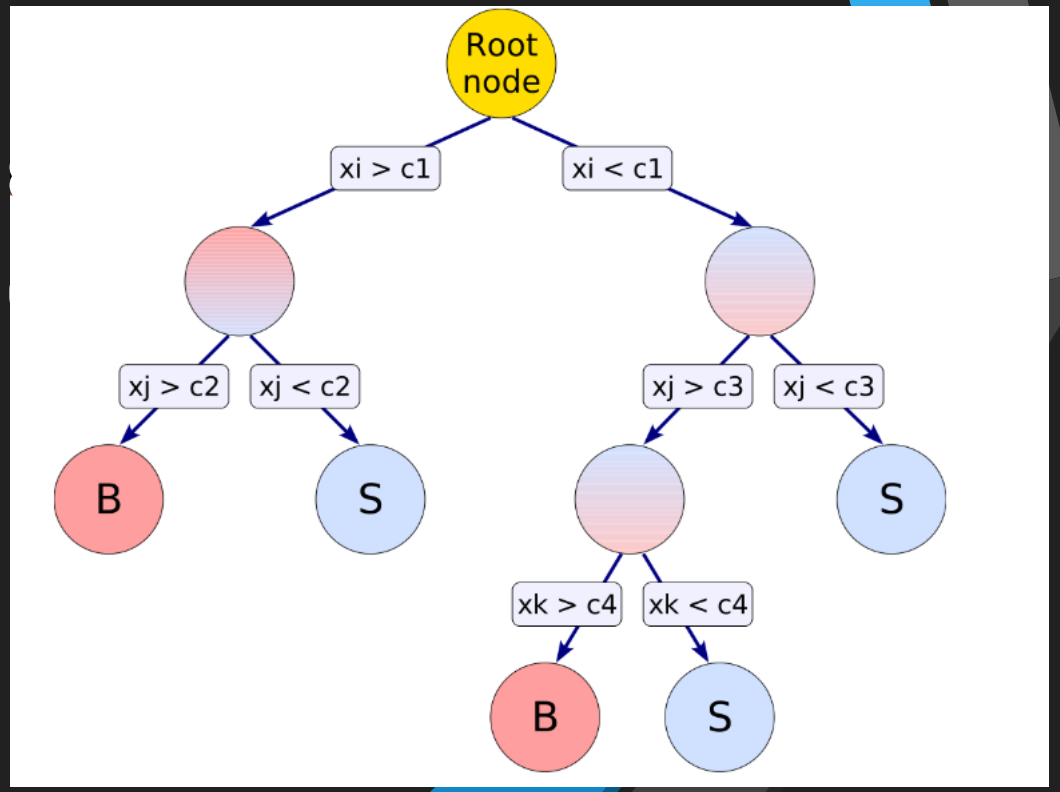

Phi between L & B
h_dphi_lb
Entries 1490
Mean 0.05468
Std Dev 1.993


Phi between L & C
h_dphi_lc
Entries 1490
Mean 0.01906
Std Dev 1.852

Variable	Definition
$m(\ell\nu b)$	top-quark mass reconstructed from the charged lepton, neutrino, and b-tagged jet
$m(jb)$	invariant mass of the b-tagged and untagged jet
$m_{\mathrm{T}}(\ell E_{\mathrm{T}}^{\mathrm{miss}})$	transverse mass of the reconstructed W boson
$\lvert\eta(j)\rvert$	modulus of the pseudorapidity of the untagged jet
$m(\ell b)$	invariant mass of the charged lepton (ℓ) and the b-tagged jet
$\eta(\ell\nu)$	rapidity of the reconstructed W boson
$\Delta R(\ell\nu b, j)$	ΔR of the reconstructed top quark and the untagged jet
$\cos\theta^*(\ell, j)$	cosine of the angle θ^* between the charged lepton and the untagged jet in the rest frame of the reconstructed top quark
$\Delta p_{\mathrm{T}}(\ell\nu b, j)$	Δp_{T} of the reconstructed top quark and the untagged jet
$\Delta R(\ell, j)$	ΔR of the charged lepton and the untagged jet

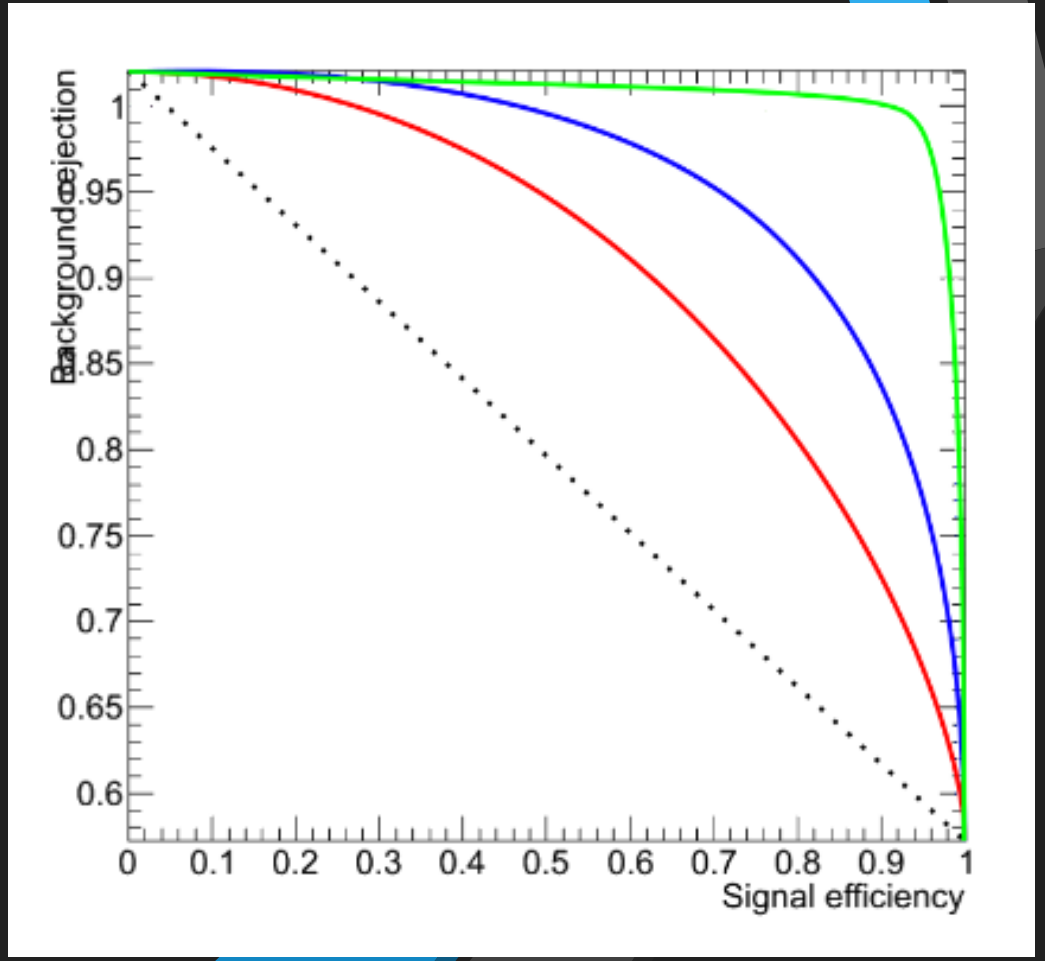
BDT

- Boosted Decision Tree – method of multivariate analysis
- Consecutive set of questions/nodes with only two possible answers
- Final answer given after max number of nodes
- Running sets of ten variables through BDT and looking for ROC and feature importance
- Help determine which variables best for separating signal and background



ROC Curves

- Receiver Operating Characteristic (ROC) curve illustrates performance of classifier
- In code, AUC value assigned to variable from 0.5 – 1.0
- AUC score provided for set of 10 variables as a whole and for each individual variable
- AUC range from 0.7 – 0.8 acceptable (more realistic)
- Paired with a "feature importance" score – this is relative to the set of ten
- Allows us to determine which variables best for signal separation



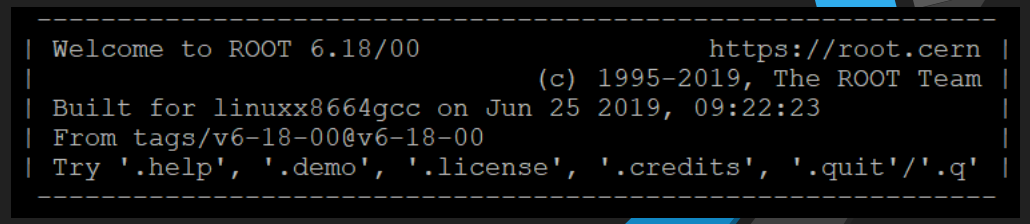
Results

- Out of 303 variables, ran thirty trials of ten variables each, with grouping loosely based on variable type (mass, phi, transverse momenta)
- Best run of 10 variables with AUC score of ~ 0.71
- For cross-section measurement of charming top before BDT, calculation for significance ~ 2
- After BDT, calculation for significance ~ 6
- Next step is to run with BDT again, but include systematics, which will most likely reduce significance

mass_lbc	5.12022781372070
mass_lbcvN	3.47165250778198
mass_lbcvX	3.27979612350463
pt_lvP	2.88116383552551
deltaR_lc	2.16968655586242
deltapt_lb	1.35062956809997
deltapt_bvP	1.32800805568695
e_l	1.32090557861328
mass_bvX	0.96435546875000
Total AUC	0.70704109943044

Conclusion

- When calculating cross section for very rare event channel, separating signal from background is key
- Tools to do this include histograms, BDTs, and ROC/Feature Importance values
- This summer, I have learned histogram generation, statistical analysis techniques, and machine learning methods
- How to work with an HEP research group



Acknowledgments

- Dr. Abbott – Research Advisor
- Veena Balakrishnan – Graduate Research Assistant
- Daniel Wilbern – Graduate Research Assistant
- REU Colleagues

Images

- https://indico.scc.kit.edu/event/48/contributions/3410/attachments/1690/2312/BDT_KSETA_Freudenstadt.pdf
- https://www.universetoday.com/140769/the-large-hadron-collider-has-been-shut-down-and-will-stay-down-for-two-years-while-they-perform-major-upgrades/
- https://atlas.cern/updates/briefing/precision-search-dark-matter
- https://www.flickr.com/photos/95869671@N08/51148317732
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.popularmechanics.com%2Fscience%2Fa25908%2Fhow-to-make-and-trap-antimatter%2F&psig=AOvVaw3JK3B8GKgrAjIhut1LdQxY&ust=1690563784530000&source=images&cd=vfe&opi=89978449&ved=0CAwQjRxqFwoTCKjcm4Gvr4ADFQAAAAAdAAAAABAE