ATLAS High Mass Resonance Analysis

Ben Sheff[1], Zirui Wang[2], Magda Chelstowska[2]

[1]University of California Berkeley, [2]University of Michigan

April 6, 2016

Intro

Background

- The LHC is a discovery machine
  - Smash protons at center-of-mass energy 13 TeV (13000 times the mass of a proton)
  - Elaborate detectors wrap around collision sites
- Proton collisions
  - Lots of energy available for particle creation
  - Spray of fundamental particles
  - Look for decay results of interesting particles
- It works
  - Big success story of the Higgs discovery at 125 GeV
- Our focus is on the new excess found at 750 GeV

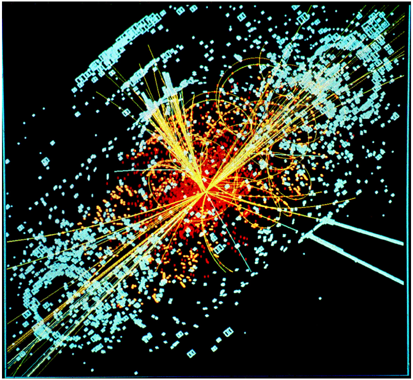
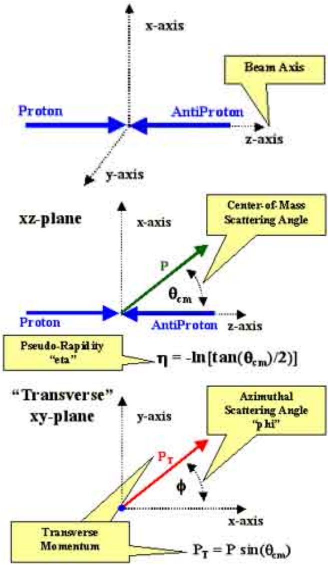


Figure: Simulated Collision Event in CMS, from http://cdsweb.cern.ch/record/628469

Commonly Used Terms

- Pseudorapidity (η): Coordinate describing angle of the particle relative to the beam axis
- Phi (ϕ): Azimuthal angle of the particle around the beam axis
- Transverse Momentum (p_{T}): Component of the particle's momentum perpendicular to the beam axis
- Photon Conversion: a process where a photon turns into an electron and a positron, usually tagged with a "type" referring to where the change happened
- Jets: Narrow cones of high energy piercing through the detector, usually indicative of a quark
- Monte Carlo: an algorithm for generating known physics processes and running them through a simulated detector. Useful for emulating signal data



Isolation overview

- We focus on $H \to \gamma\gamma$ process
  - Good energy resolution
  - Relatively easy to select for
- Only photons produced in $H \to \gamma\gamma$
  - Few other nearby particles
  - Little energy detected outside the photon itself
- Quantify track isolation summing transverse momentum (p_{T}) in a cone around the particle trajectory
  - High isolation means particle isn't very isolated

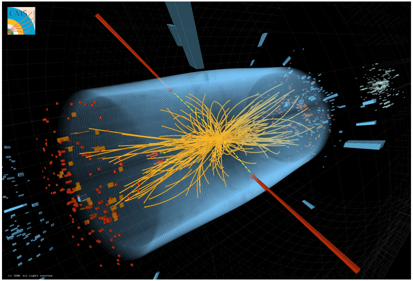


Figure: LHC Collision Rendition from CERN http://i.livescience.com/images/i/000/022/658/boson.jpg?1323784920

My work

1. Derive isolation working point uncertainties
2. Cross-checks on correlations between isolation variables
3. Validation and investigation of new ntuples
4. Investigation of alternative Monte Carlo sample for $bb\gamma\gamma$ analysis

Isolation

Isolation Studies

- Can cut out high isolation (non-isolated) particles, since they're less likely to be from $H \rightarrow \gamma\gamma$
    - Cut itself has already been created and tuned
    - For high mass $\gamma\gamma$ resonance is still an open point
    - some ambiguity about systematic uncertainties
    - Systematics may be dependent on some photon qualities
- My role is to derive isolation working point uncertainties in a more robust fashion
    - Check effects of transverse momentum (p_{T}), pseudorapidity (η), and photon conversion type

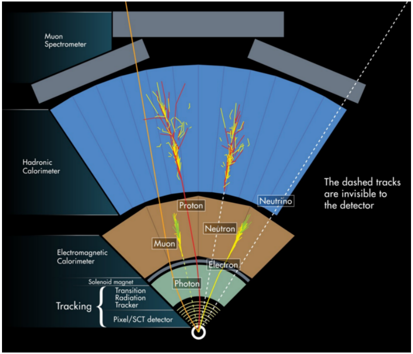
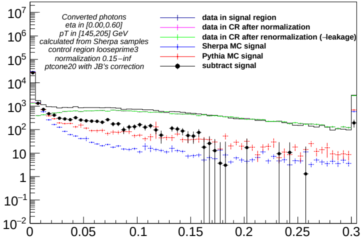
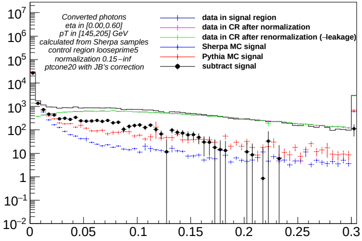


Figure: Layers of the ATLAS Detector. http://collider.physics.ox.ac.uk/img/layers.png

Methodology

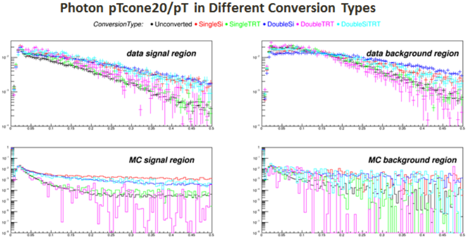
- Find uncertainty of the efficiency of the cut
  - How much of the signal is captured by the cut
  - Calculate separately for different cuts along p_T, η, and conversion
- Compare Monte Carlo-generated signal model with actual signal
- Need to find "actual" signal (photon spectrum)
  - Model background by data from control region after removing signal leakage
  - Remove the background to get signal
- Need scale background distribution to data, but how?
  - Fit a sum of Monte Carlo signal and background to get background scaling
  - Scale background to exactly match data at high isolation

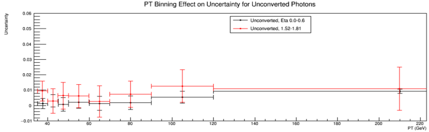


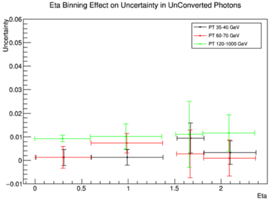


Splitting up the data

- Different p_T, η, and conversion type lead to different trends
- Uncertainty varies significantly with each axis
- 8 p_T regions
- 10 η regions
  - geometric considerations
  - Detector changes
- Converted vs unconverted
  - Similarities among converted photons
  - Low statistics for specific types

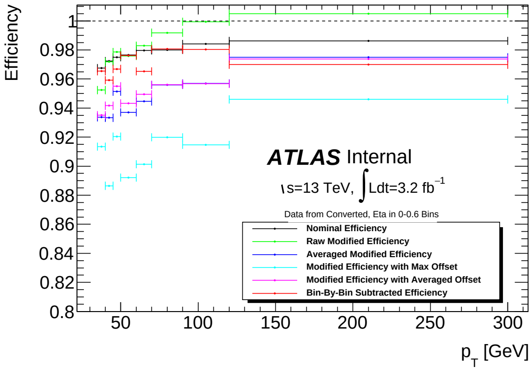






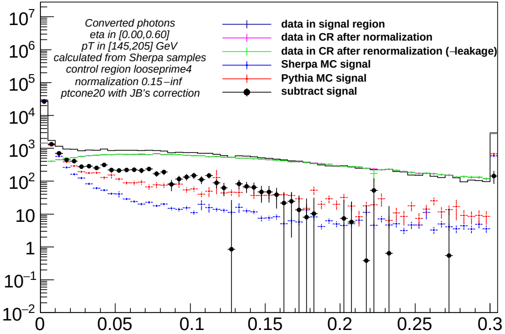
Trouble on the horizon

- Looking closer, the efficiency of our "actual" signal was sometimes > 1
  - Fitted background above data in some places
  - Background $>$ data leads to negative events in signal
- Can offset the signal and iterate fit to remove new background
  - Offset signal to remove all negative events
  - Offset signal by an average of the negative events
- Can also try fitting background on directly to high isolation data and subtract it manually (called bin-by-bin subtraction)



The horizon runs away

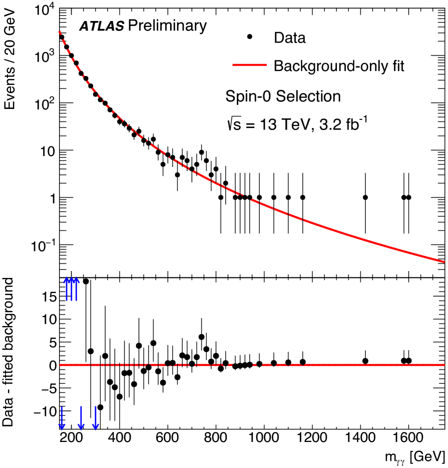
- Along the way, sent out some results
  - Full signal+background model fit went into an analysis package update
  - Moriond support note released including results from directly fitted background approach and some preliminary analysis
- Still deciding on an approach to use
  - Some question of mis-modeled signal and/or background
  - Peculiar trends in background causing unphysical results



Correlations

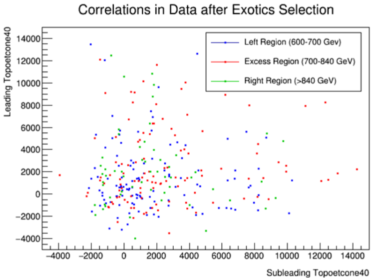
Why do this again?

- Excess of events at high mass
- Need to refine uncertainties for this high mass regime
  - Detailing systematics involved between 700 and 840 GeV
- Plenty of other work important to the cause



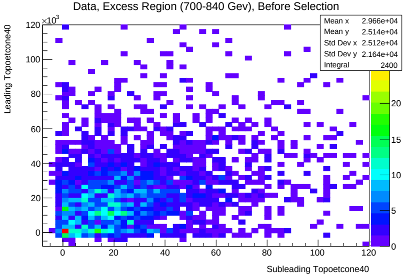
Isolation Correlations

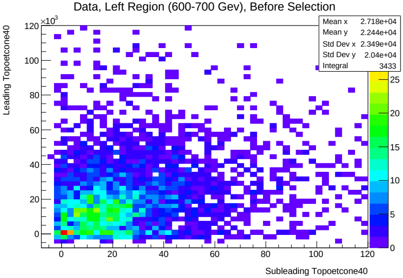
- Some interest in correlations between isolation variables
  - Been focused on track isolation (ptcone20)
  - Calorimeter isolation (topoetcone40)
  - Different cone sizes (topoetcone40, topoetcone20)
  - Comparing leading and subleading photons
- Focused on the excess region, 700-840 GeV
- Not much prior work done

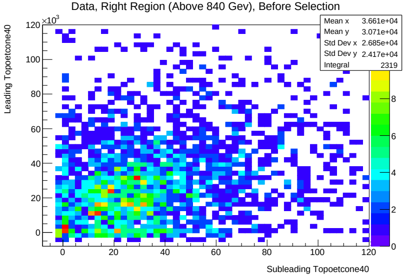


Plotting correlations

Between statistical concerns and poor ability to recognize correlations by eye, plots were not terribly informative.







Correlation Coefficient Matrices

- Made code to generate correlation coefficient matrices
- Little correlation between variables
  - As expected, strong correlation between the same variable on the same photon with different radii
  - Other correlations usually low
  - Correlations seem to rise with low statistics, causing ambiguity
- Some increased correlation at high pileup

Data Exotic Selection Right of the Excess (59 Events)

X	topoetcone40_leading	topoetcone20_leading	ptcone20_leading	topoetcone40_subleading	topoetcone20_subleading	ptcone20_subleading
topoetcone40_leading	1	0.89674	0.347233	0.0265723	-0.0328839	-0.140545
topoetcone20_leading	0.89674	1	0.357375	0.0548048	0.00851499	-0.14442
ptcone20_leading	0.347233	0.357375	1	-0.104892	-0.093166	-0.099327
topoetcone40_subleading	0.0265723	0.0548048	-0.104892	1	0.846317	0.25583
topoetcone20_subleading	-0.0328839	0.00851499	-0.093166	0.846317	1	0.246613
ptcone20_subleading	-0.140545	-0.14442	-0.099327	0.25583	0.246613	1

New Ntuple

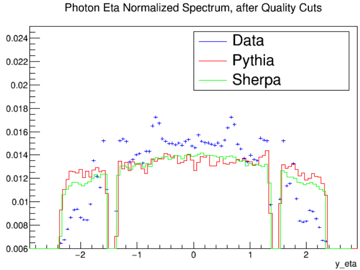
Validating an Ntuple

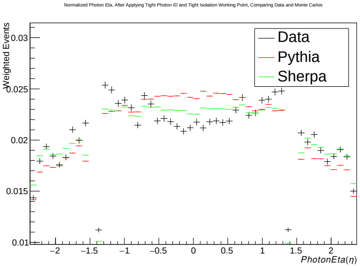
Ntuples are collections of data that remove the huge amounts of information you don't want, and are decorated with well defined higher order parameters

- Can't run on the full data all the time
  - Would have to manually re-evaluate high order parameters, like filters
  - *Way* too much of it
- New one produced to include:
  - More data
  - Updated weighting
  - More parameters per event
  - Various other updates
- Tailored for single photon studies
- Needs to be validated
  - check that patterns in old ntuples are present in the new one
  - check how new weighting and different cuts interplay with other variables

Comparing Data and Monte Carlo η

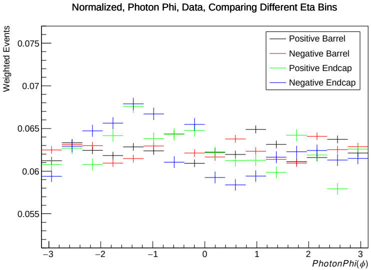
- Checking how well data matches Monte Carlo (MC) in this Ntuple
- Poor initial correspondence in pseudorapidity between data and Monte Carlo
- Discrepancy likely comes from jets, which tend to be closer to the center
  - Supported by checking tightID photons
- Some discrepancy remains
  - May be a small weighting issue

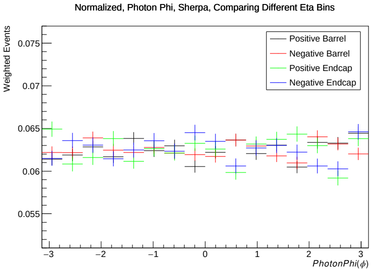




Comparing Data and Monte Carlo Φ

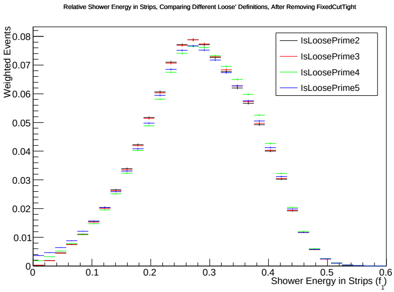
- Saw some asymmetries previously, wanted to check if they remained in new ntuple
- Compare different eta regions
  - Negative endcap in particular seems somewhat skewed toward negative ϕ, around $-\frac{\pi}{2}$
- Not really seen in Monte Carlo information
- Absent from Jet ϕ distributions as well

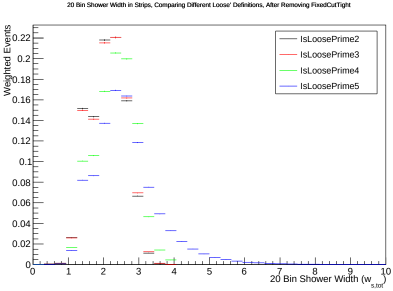




Background Cuts

- Background region is passed through a loose filter, called Loose', and a number
  - New Ntuple gives access to different Loose' definitions
  - Loose' definitions give different particle definitions
- Focused on shower shape variables to try and see Loose' cuts
  - High energy photons make a shower of particles on impact with detector
  - Focused on photons not passing tight isolation working points
- Different definitions of Loose' had little effect on most variables
- Some were more dramatic
  - Most of the more dramatic ones dealt with shower width along η
  - Some difference in relative size of secondary energy peak in calorimeter as well





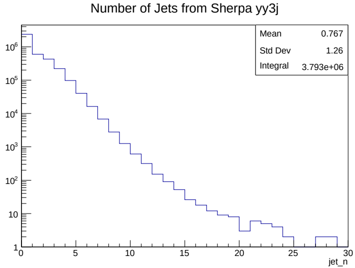
MC Sample

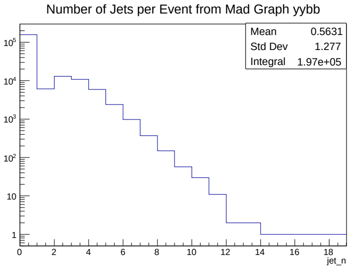
Need for new $bb\gamma\gamma$ Monte Carlo

- To analyze quarks, need to look at jets
  - Between filters and decay chains, nearly all jets are quarks
  - Jets are tagged based on flavour of the quark involved
  - For $bb\gamma\gamma$ analysis, want events with two leading bottom jets in particular
- Analysis thus far based in MadGraph sample
  - Verified jet distributions
  - Correct proportion of heavy flavour jets
  - Relatively low statistics once filtered down
- Production already done for a really high statistics Sherpa sample
  - Includes millions of events tagged as $\gamma\gamma$+jets
  - Intended for photon analysis, so jet flavour ratios might be off
- Need to verify some basic spectra match up between Sherpa and MadGraph
  - Flavour ratios need to match
  - Spectra of basic p_{T}, η, ϕ
  - Spectra of Dijet properties, Lorentz invariant quantities over leading and subleading jets

Feasibility of the Plan

There are many more events in the Sherpa MC
Most of the Sherpa events have 0 or 1 jets, but the 2+ jet events still have substantial statistics





Flavour proportion checks

- Looked at comparing the presence of different flavour combinations for leading two jets
  - Removed events with fewer than two jets
  - Normalized to the number of total events
- Nice symmetry between leading and subleading
- Decent agreement by eye between the two MCs

Flavour comparison, Sherpa $\gamma\gamma 3J$ MC

Reconstructed Jet Information

Leading Flavour ↓	Light	Charm	Bottom	τ
Light	8.29×10^{-1}	7.00×10^{-2}	9.88×10^{-3}	4.90×10^{-6}
Charm	6.61×10^{-2}	1.37×10^{-2}	9.00×10^{-4}	0
Bottom	8.50×10^{-3}	7.03×10^{-4}	1.61×10^{-3}	0
τ	2.45×10^{-6}	1.23×10^{-6}	0	0

Truth Jet Information

Leading Flavour ↓	Light	Charm	Bottom	τ
Light	7.44×10^{-1}	8.73×10^{-2}	1.56×10^{-2}	3.28×10^{-6}
Charm	9.79×10^{-2}	3.34×10^{-2}	2.23×10^{-3}	0
Bottom	1.43×10^{-2}	1.67×10^{-3}	3.94×10^{-3}	0
τ	1.64×10^{-6}	3.28×10^{-6}	0	0

Flavour comparison, MadGraph $\gamma\gamma jj$ MC

Reconstructed Jet Information

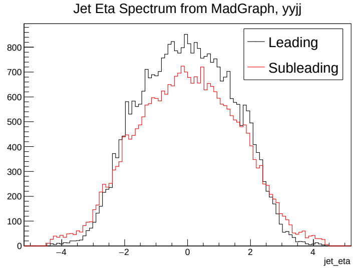
Leading Flavour ↓	Light	Charm	Bottom	τ
Light	7.76×10^{-1}	8.89×10^{-2}	7.29×10^{-3}	3.46×10^{-5}
Charm	8.92×10^{-2}	3.01×10^{-2}	1.14×10^{-3}	0
Bottom	5.98×10^{-3}	6.57×10^{-4}	3.46×10^{-4}	0
τ	3.46×10^{-5}	0	0	0

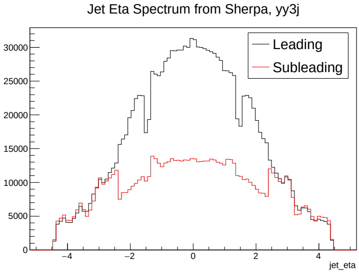
Truth Jet Information

Leading Flavour ↓	Light	Charm	Bottom	τ
Light	7.85×10^{-1}	8.14×10^{-2}	7.20×10^{-3}	3.76×10^{-4}
Charm	8.59×10^{-2}	3.19×10^{-2}	8.83×10^{-4}	3.07×10^{-5}
Bottom	5.94×10^{-3}	9.06×10^{-4}	3.84×10^{-4}	0
τ	2.23×10^{-4}	1.54×10^{-5}	0	0

Comparing Sherpa and MadGraph η

- Sherpa forward jets, both leading and subleading jets have the same spectrum
- Sherpa central jets are partnerless about half the time
  - Leading and subleading still have similar trends
- MadGraph has no odd behavior at gap region
  - Cut built into MadGraph removing most forward jets
- Similar behavior for p_{T} and ϕ, plots available in backup

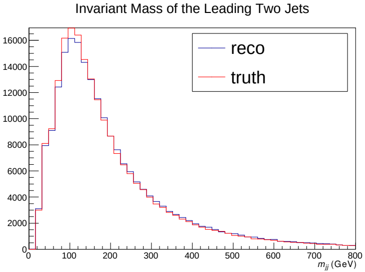




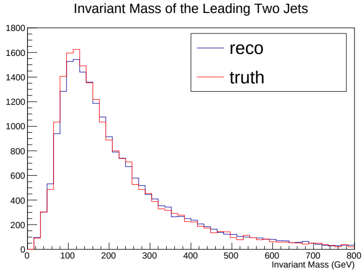
Invariant Mass of Leading Jets Comparison

- Cut to events with at least two matched jets
  - Removed any reconstructed jets that were over 0.2 away from all truth jets in $\phi - \eta$ space
- Reconstructed and truth jets match nicely
- Similar curve shape between Sherpa and MadGraph

Sherpa:



MadGraph:



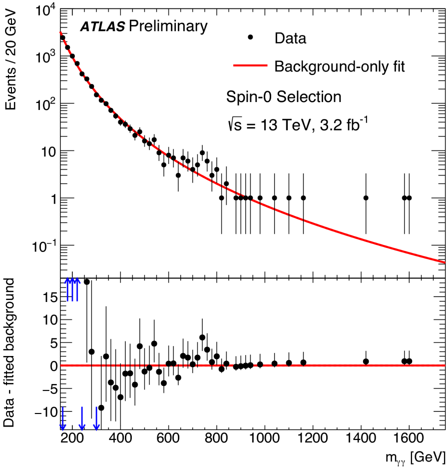
Results

Results

- Isolation Studies
  - Preliminary results have higher uncertainties than previous working points
  - Further investigation needed to more robustly approximate actual signal distribution in data
  - Results added to official analysis package update and Moriond support note
- Correlation Studies
  - No correlation between leading and subleading photon
  - Some possible correlation between calorimeter and track isolation in high pileup events
  - Results added to Moriond support note
- New Ntuple validation
  - Seems fine
  - Different control region definitions shown to relate to different widths of photon shower along η
- $bb\gamma\gamma$ Monte Carlo sample
  - Sherpa seems okay to use thus far
  - A few peculiarities, especially in the η spectrum that may be built in cuts in MadGraph
  - More investigation needed to check cross section normalization and direct spectrum comparisons

The Future

- Isolation studies will be refined for the next paper
- Some finishing touches on the Ntuple and Sherpa checks
- Apply the new methods and results to the next set of run 2 data



Touring Europe

Thanks!

Backup

Data from Excess Region (700-840 GeV), without applying exotics selection, at different pileup values:

Low Pileup ($\mu < 8$), (50 Events)

X	topoetcone40_leading	topoetcone20_leading	ptcone20_leading	topoetcone40_subleading	topoetcone20_subleading	ptcone20_subleading
topoetcone40_leading	1	0.757107	0.308363	0.107886	-0.0110961	-0.123711
topoetcone20_leading	0.757107	1	0.343881	0.11471	0.0589021	-0.137677
ptcone20_leading	0.308363	0.343881	1	-0.0102904	-0.0419078	-0.118195
topoetcone40_subleading	0.107886	0.11471	-0.0102904	1	0.942406	0.544297
topoetcone20_subleading	-0.0110961	0.0589021	-0.0419078	0.942406	1	0.584073
ptcone20_subleading	-0.123711	-0.137677	-0.118195	0.544297	0.584073	1

Normal Pileup ($8 < \mu < 15$), (2471 Events)

X	topoetcone40_leading	topoetcone20_leading	ptcone20_leading	topoetcone40_subleading	topoetcone20_subleading	ptcone20_subleading
topoetcone40_leading	1	0.817707	0.305609	0.252143	0.196576	0.116333
topoetcone20_leading	0.817707	1	0.357947	0.284189	0.22147	0.125553
ptcone20_leading	0.305609	0.357947	1	0.118084	0.0816183	0.0651018
topoetcone40_subleading	0.252143	0.284189	0.118084	1	0.674158	0.287076
topoetcone20_subleading	0.196576	0.22147	0.0816183	0.674158	1	0.462231
ptcone20_subleading	0.116333	0.125553	0.0651018	0.287076	0.462231	1

High Pileup ($\mu > 15$), (50 Events)

X	topoetcone40_leading	topoetcone20_leading	ptcone20_leading	topoetcone40_subleading	topoetcone20_subleading	ptcone20_subleading
topoetcone40_leading	1	0.331653	0.137619	0.159416	0.207343	0.126085
topoetcone20_leading	0.331653	1	0.592444	0.41922	0.513225	0.452457
ptcone20_leading	0.137619	0.592444	1	0.435233	0.2762	0.0426803
topoetcone40_subleading	0.159416	0.41922	0.435233	1	0.903706	0.722139
topoetcone20_subleading	0.207343	0.513225	0.2762	0.903706	1	0.750632
ptcone20_subleading	0.126085	0.452457	0.0426803	0.722139	0.750632	1

Sherpa Monte Carlo Sample in Excess Region (700-840 GeV), without applying exotics selection, at different pileup values:

Low Pileup ($\mu < 8$), (5700 Effective Events)

X	topoetcone40_leading	topoetcone20_leading	ptcone20_leading	topoetcone40_subleading	topoetcone20_subleading	ptcone20_subleading
topoetcone40_leading	1	0.761852	0.0085064	0.0208207	0.0271315	-0.00460648
topoetcone20_leading	0.761852	1	0.0110731	0.0118587	0.0256807	-0.000275445
ptcone20_leading	0.0085064	0.0110731	1	0.00467398	-0.00113252	-0.000127438
topoetcone40_subleading	0.0208207	0.0118587	0.00467398	1	0.577833	0.0236234
topoetcone20_subleading	0.0271315	0.0256807	-0.00113252	0.577833	1	0.0352392
ptcone20_subleading	-0.00460648	-0.000275445	-0.000127438	0.0236234	0.0352392	1

Normal Pileup ($8 < \mu < 15$), (65000 Effective Events)

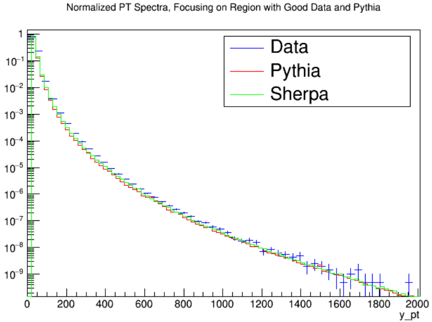
X	topoetcone40_leading	topoetcone20_leading	ptcone20_leading	topoetcone40_subleading	topoetcone20_subleading	ptcone20_subleading
topoetcone40_leading	1	0.716855	0.00305372	0.0157408	0.0130908	-0.00222131
topoetcone20_leading	0.716855	1	0.00104489	0.0116283	0.0110757	-0.000983864
ptcone20_leading	0.00305372	0.00104489	1	-0.000782011	9.21E-06	-0.000357667
topoetcone40_subleading	0.0157408	0.0116283	-0.000782011	1	0.594916	0.000541029
topoetcone20_subleading	0.0130908	0.0110757	9.21E-06	0.594916	1	0.00157228
ptcone20_subleading	-0.00222131	-0.000983864	-0.000357667	0.000541029	0.00157228	1

High Pileup ($\mu > 15$), (65000 Effective Events)

X	topoetcone40_leading	topoetcone20_leading	ptcone20_leading	topoetcone40_subleading	topoetcone20_subleading	ptcone20_subleading
topoetcone40_leading	1	0.712655	0.00060086	0.0187898	0.0114936	-0.000972022
topoetcone20_leading	0.712655	1	0.000347915	0.0200151	0.012907	-0.000407202
ptcone20_leading	0.00060086	0.000347915	1	-0.000239845	-0.000701451	-2.89E-05
topoetcone40_subleading	0.0187898	0.0200151	-0.000239845	1	0.613147	0.0024292
topoetcone20_subleading	0.0114936	0.012907	-0.000701451	0.613147	1	0.00257018
ptcone20_subleading	-0.000972022	-0.000407202	-2.89E-05	0.0024292	0.00257018	1

Comparing Data and Monte Carlo p_T

- Checking how well the data matches Monte Carlo (MC) in this Ntuple
- First looked into transverse momentum (p_T), where there's pretty good agreement until about 2 TeV
- Disagreement at high p_T likely due to MC predicting fractions of events
- MC spectra agree with each other farther, limited to 3 TeV
  - Pythia is generated up to 3 TeV
- Plot below cuts off high p_T since data isn't very good beyond 2 TeV



Flavour comparison, MadGraph $bb\gamma\gamma$ MC

Reconstructed Jet Information

Leading Flavour ↓	Light	Charm	Bottom	τ
Light	8.88×10^{-2}	3.32×10^{-3}	2.30×10^{-1}	8.90×10^{-5}
Charm	2.25×10^{-3}	1.78×10^{-4}	5.04×10^{-3}	0
Bottom	2.73×10^{-1}	1.08×10^{-2}	3.86×10^{-1}	2.97×10^{-5}
τ	2.97×10^{-5}	0	5.93×10^{-5}	0

Truth Jet Information

Leading Flavour ↓	Light	Charm	Bottom	τ
Light	3.89×10^{-2}	1.89×10^{-3}	1.54×10^{-1}	8.94×10^{-5}
Charm	2.21×10^{-3}	3.13×10^{-4}	8.56×10^{-3}	0
Bottom	2.57×10^{-1}	1.40×10^{-2}	5.23×10^{-1}	4.11×10^{-4}
τ	7.15×10^{-5}	0	2.77×10^{-4}	0

Flavour comparison, MadGraph $\gamma\gamma bj$ MC

Reconstructed Jet Information

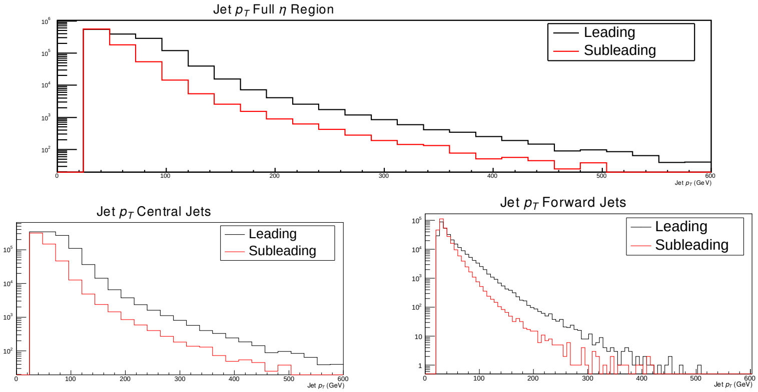
Leading Flavour ↓	Light	Charm	Bottom	τ
Light	2.46×10^{-1}	1.70×10^{-2}	3.12×10^{-1}	3.34×10^{-5}
Charm	1.47×10^{-2}	2.97×10^{-3}	2.98×10^{-2}	0
Bottom	3.03×10^{-1}	3.16×10^{-2}	4.27×10^{-2}	0
τ	6.68×10^{-5}	0	0	0

Truth Jet Information

Leading Flavour ↓	Light	Charm	Bottom	τ
Light	1.87×10^{-1}	9.92×10^{-3}	3.20×10^{-1}	1.65×10^{-4}
Charm	1.15×10^{-2}	2.98×10^{-3}	3.54×10^{-2}	1.57×10^{-5}
Bottom	3.48×10^{-1}	4.10×10^{-2}	4.35×10^{-2}	1.41×10^{-4}
τ	9.40×10^{-5}	0	1.33×10^{-4}	0

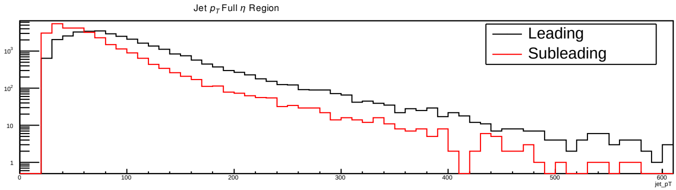
Sherpa p_{T} Spectra

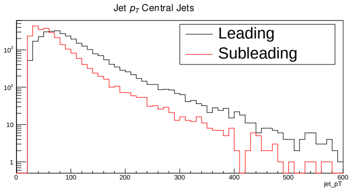
- Discrepancies seem to match up with η issues
- Uncut p_{T} spectrum curves differ by a fairly steady factor

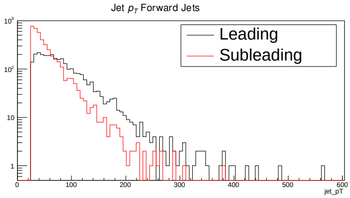


MadGraph p_T Spectra

- As with η, funny differences between forward and central disappear
- Shapes seem similar

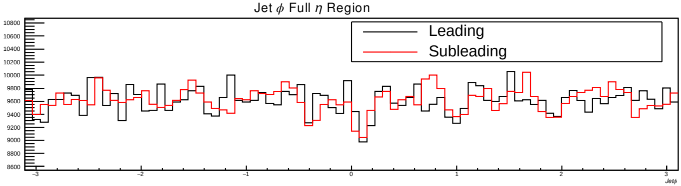


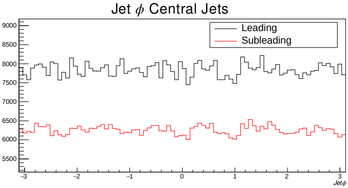


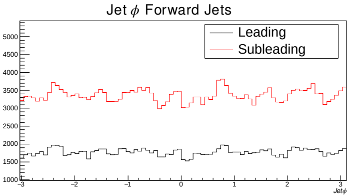


Sherpa Φ Spectra

- Cut to only events with 2+ jets
- Leading and subleading jets match nicely
- Phi is very symmetric, as should be expected

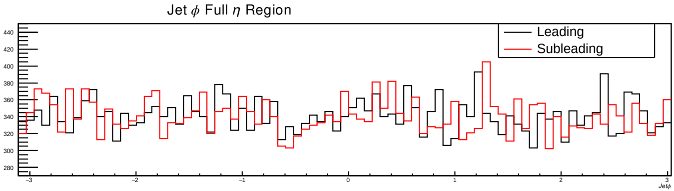


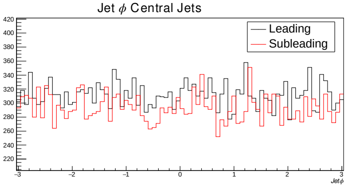


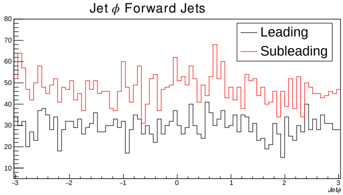


MadGraph Φ Spectra

- Same cuts as for Sherpa
- Leading and subleading jets match nicely, and still symmetric
- Asymmetries don't match up well, but should be statistical anyway

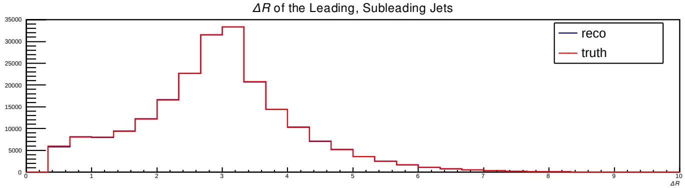


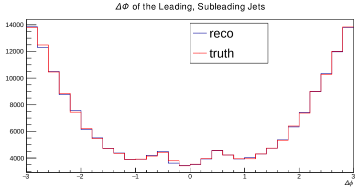


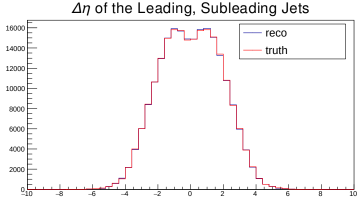


Sherpa Leading vs. Subleading Momentum Difference

- Cut to events with at least two matched jets
  - Removed any reconstructed jets that were over 0.2 away from all truth jets in $\phi\eta$ space
- Reconstructed and truth jets match nicely







MadGraph Leading vs. Subleading Momentum Difference

- Same cuts as for Sherpa
- Fairly good agreement between MCs, other than small bump of small $\Delta\phi$ events
  - Reconstructed and truth jets match up nicely here as well