Substructure at ATLAS

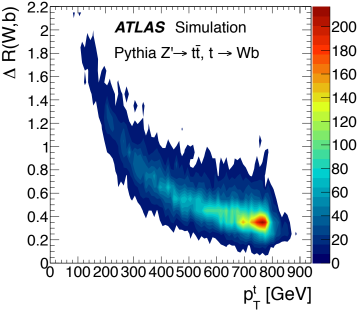
Berkeley Search Workshop 2014

Maximilian Swiatlowski

27 January, 2014

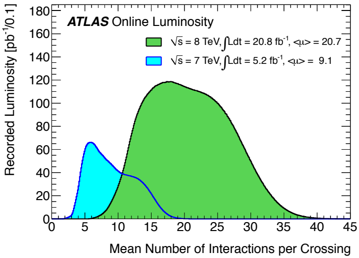
The Challenges of Jet Substructure

- Jet substructure is a **huge opportunity** for searches at the LHC
  - We have available an entirely new class of observables compared to the Tevatron!
- Many reasons to consider substructure when looking for new physics
  1. Signals are **often boosted–** standard techniques can break down!
  2. Detectors have incredibile **granularity and resolution–** small structure is easier to resolve!
  3. **New physics is hiding–** need as many handles as we can get!



The Challenges of Jet Substructure

- But with great opportunity, come **great challenges**
- ATLAS has been spent the past 3+ years trying to address these challenges, and this talk will be structured around them:
    1. Understanding the theory landscape
    2. Dealing with pileup
    3. Testing methods in data
    4. Actually applying substructure to searches
- Not going to go into great detail on the methods, but feel free to ask!



Understanding the Theory Landscape

The Theory Landscape

SLAC

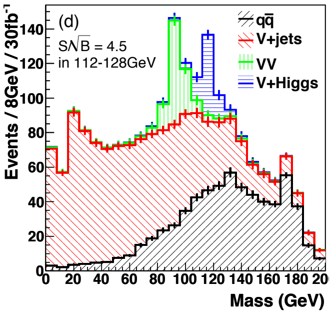
- From an experimental perspective...

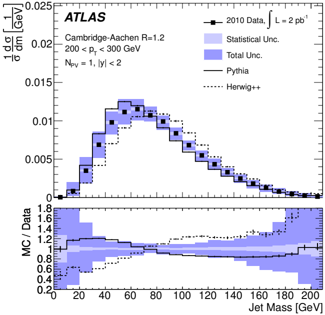
Shower Deconstruction
BDRS
Filtering
Calculations??
Telescoping Jets
Q-Jets
Pull
Trimming
Cleansing
N-subjettiness
TOM
BBQ?
Pruning
Quark-Gluon
Subjet Counting
Mass Drop
Accidental Substructure

- How can you process all of these ideas?
- How do you begin to tell what's useful?

Processing New Ideas

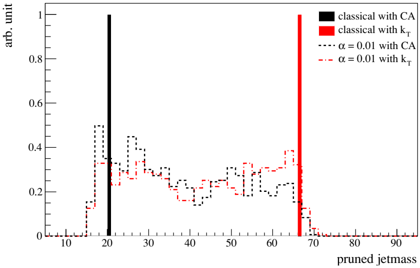
- Substructure as we know it began in 2008– the seminal BDRS paper
  - A new idea at the time: use Zh and Wh production at high boson p_T and merged Higgs-jets to search for $h \to b\bar{b}$
  - This field began because of a new idea for a **search**!
- First ATLAS results (with data) on similar large-R jets: May 2011
  - Not a bad turn around, given we had to start running a detector!
  - But three years is still a long time...

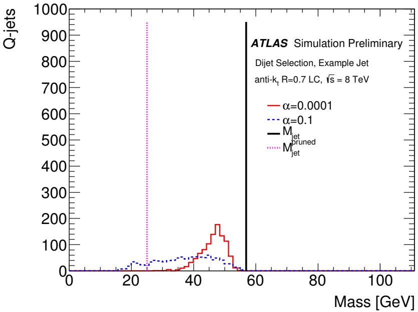




Current Processing Time: Q-jets

- More recent example: look at Q-jets (multiple re-interpretations of one jet):

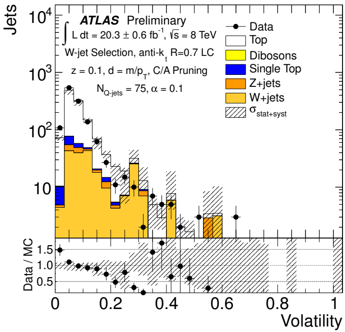




- Went from theory paper (June 2012) to experimental result (August 2013) in about one year
- **Much faster turn-around time for trying new ideas!**

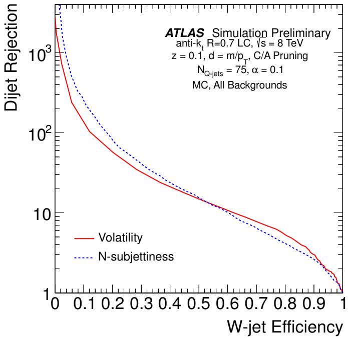
How Do We Get Fast Results?

- Many factors contribute to our ability to study new algorithms and ideas quickly
  1. **Well understood data samples**: especially boosted W-jets, top jets, QCD jets
  2. Existing code infrastructure means less time spent on "basics"
  3. The collaboration is much more comfortable with new ideas because of the success of previous studies



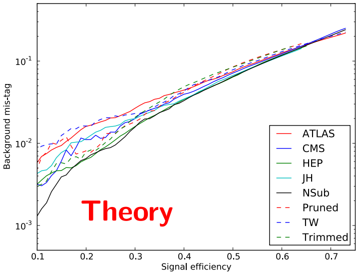
How Do You Choose?

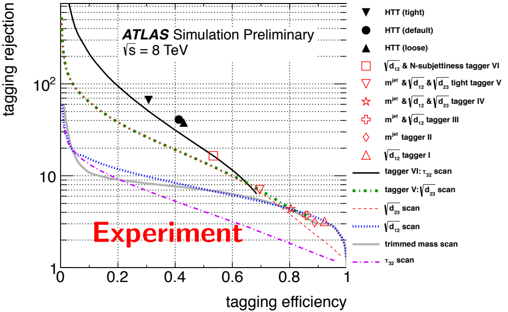
- Being able to study new algorithms quickly is great! But we do not study them just for their own sake...
  - Importantly, need to **compare performance** of all these ideas
- New ideas and improved performance are great, but there is a need for **stability and consistency** so that searches can actually use these techniques



- Study Q-jets in detail, but we see it **performs the same** as older techniques (n-subjettiness)

How Do You Choose: Top Tagging





- Theory results suggest that n-Subjettiness and HEP Top Tagger are best: **also seen in ATLAS**
- But important to note that peformance **changes with efficiency**
  - Some final states might require higher efficiency, and a different tagger!

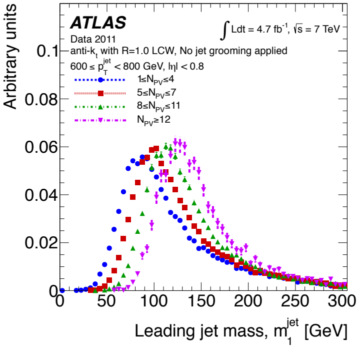
Understanding the Theory Landscape

- **New ideas are constantly emerging,** and ATLAS is constantly implementing them
  - Turn-around on new ideas is a year or less, oftentimes
  - Dialogue with theorists has been **extremely productive** for everyone
- We are beginning to understand **what works best**: but we tread carefully, as this is final state dependent
- **What is missing?**
  - *No public ATLAS or CMS results on boosted Higgs!*
  - Hard to pin down exactly why... not enough cross-section? Need new techniques? Backgrounds too difficult?

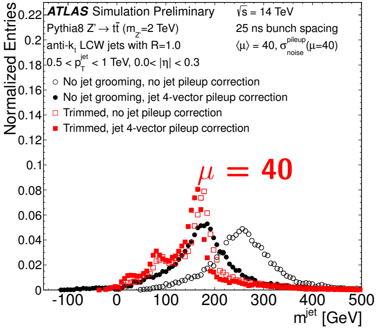
Pileup

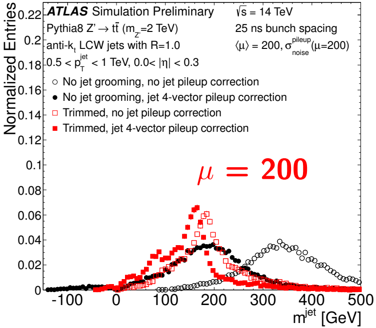
Pileup is a Concern

- Using large-R jets means we are **very suspectible to extra radiation**
    - Which is a shame, because the LHC plans to throw a lot of radiation at us
- Even in 2011 conditions, see a **large dependence of mass on pileup**
- How can we hope to do physics in these conditions?



Area Corrections and Grooming are Magic

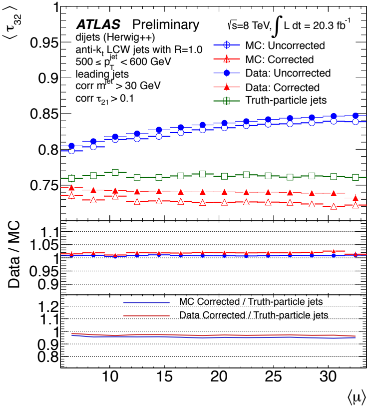




- Areas correction: measures susceptibility of jets to radiation by measuring ambient energy in events
- Trimming: remove soft subjets from large-R jets
- Even with 200 interactions, these two techniques **restore the jet mass spectrum**

What About Jet Shapes?

- Can easily extend the areas correction to work for **jet shapes**
- Corrected distributions show **no slope vs. more interactions**
- We can use the same technology to correct for many other shapes as well (and even event-shapes!)



Are We Still Worried About Pileup?

- The answer is both yes and no:
  - **Yes!** We do not know if the MC is correct! Low p_T jets have problems!
  - **No!** The MC has not lied yet, and 2012 results are good! Substructure does not really care about low p_T!
- For now, it seems like jet substructure observables will survive to very high pileup conditions
  - Other elements of searches, like E_T^{miss}, may not!
  - Validation of these techniques will be **critical** in 2015

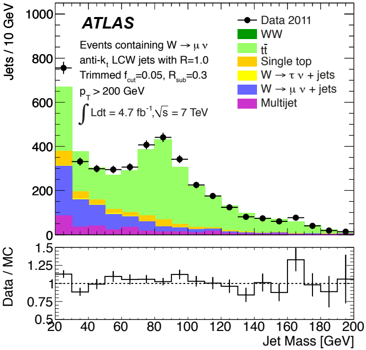
Comparing Data/MC

Validating Substructure

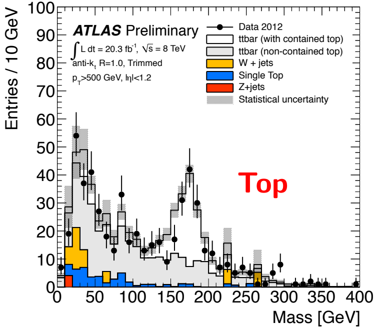
- Much of ATLAS's efforts with substructure so far have been towards **validation**:
  - Understanding data/MC agreement
  - Studying performance in different samples
  - Measuring **resolution** and **uncertainties**
- The challenge: understand to what extent jet substructure **actually works** in the data
  - Don't just look at QCD backgrounds, but **real signals** as well

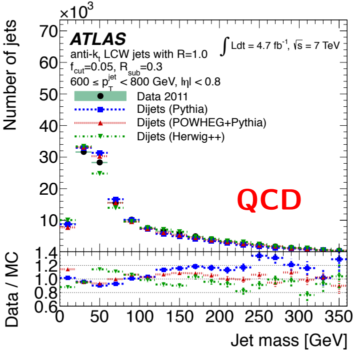
Validating Mass

- First thing to understand: in a well known sample, how accurate is our modeling of the jet mass?
  - The simplest jet substructure observable...
- Can use a clean sample of W-jets to measure how well data and MC agree– use this to derive uncertainties



Jet Mass (with Trimming)

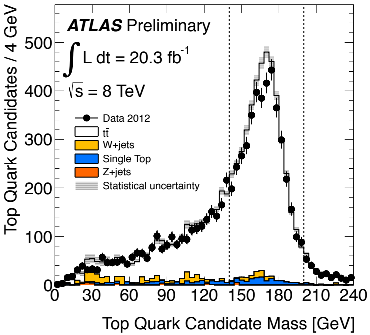




- Can see a **beautiful top peak** in the top sample: spectrum is very well modeled
- Lots of variation in MC for QCD: Powheg and Herwig++ do best

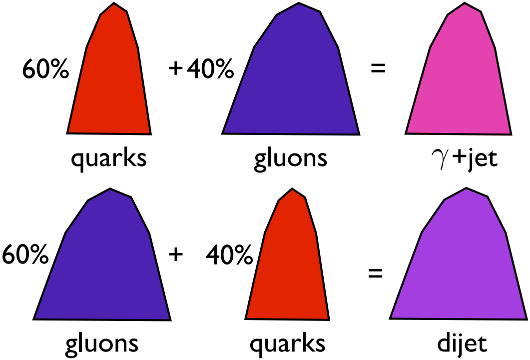
HEP Top Tagger: Very High Purity

- Trimming and n-subjettiness often used in semi-leptonic searches: need high signal efficiency
- But what about all-hadronic searches– where you need high signal purity?
- HEP Top Tagger provides a much cleaner sample, and is also very well modeled



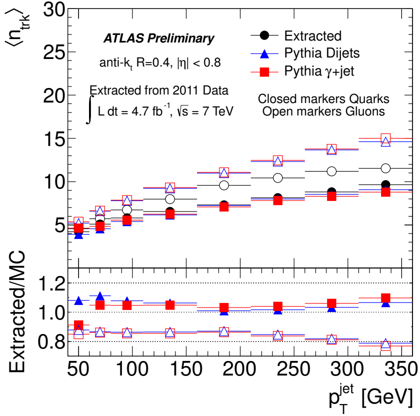
Template Methods in Quark-Gluon Tagging

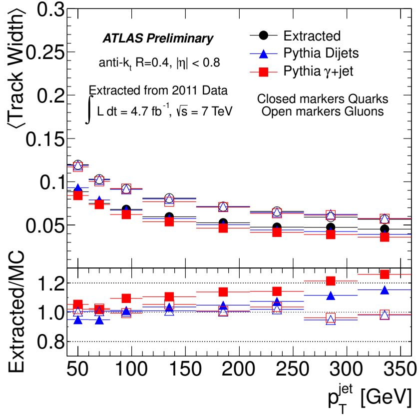
- Significant data/MC disagreement for variables sensitive to quark/gluon tagging required the use of a data-driven templates



- Take percentages from MC, measure γ+jet and dijet in data: solve for quark and gluon distributions in data
- More information on method in backup

Templates with Data





- Data **disagrees with Pythia** in n_{trk}, leading to worse separation than expected
- Track Width has better agreement, though not good at high p_T

Validating Substructure

- In top and W measurements, **remarkable agreement** between data and MC
  - In both QCD and signal samples!
- Quark/gluon tagging has more trouble: gluon properties in particular are not well modeled
  - Need to help theorists understand these issues!
- Has the challenge been met?
  - W and top samples used as much as possible to understand “real” substructure
    - But the p_T spectrum is very limited– not related to **real signals**
  - Where else is it interesting to measure structure?

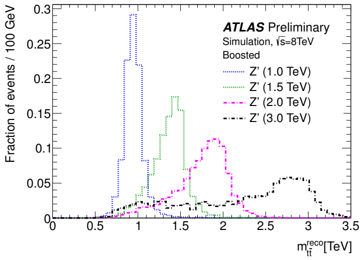
Searches

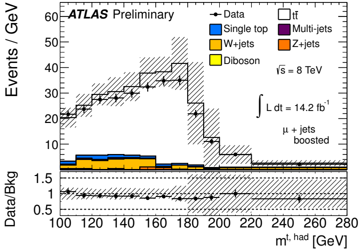
Searches: Actually Using Substructure

- Now that we understand the tools, we should actually be using these in searches
    - Recall: the entire field is motivated by a search (BDRS and Higgs)
- ATLAS has a number of analyses that apply our techniques

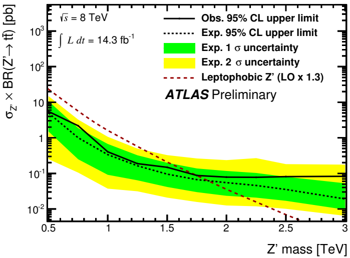
$t\bar{t}$ Resonances

- One of the canonicial searches utilizing substructure
  - High mass Z' decays to high p_T top quarks: **too boosted for a resolved analysis to see**
  - Reconstruct Z' mass with semi-leptonic and hadronic top candidates

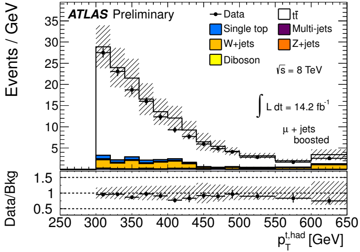




$t\bar{t}$ Resonances: Limits

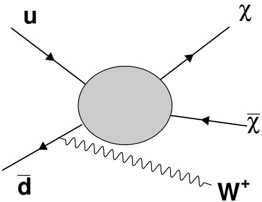


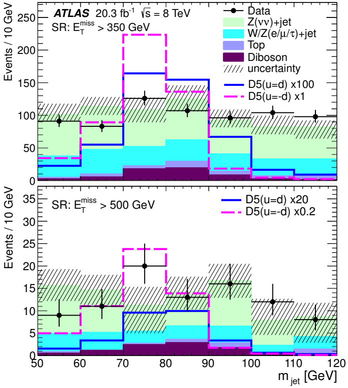
- Search utilizes very high p_T top quarks: how well do we understand large-R jets, especially mass, here?
    - Validation regions are at lower p_T
- Limits are significantly better than a resolved analysis would be!



Dark Matter with Substructure

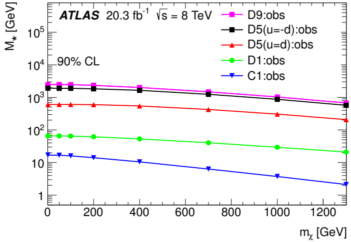
- Another search with substructure at the very heart: look for **dark matter** produced in association with ISR W-jet
  - Semi-leptonic and resolved jets have huge backgrounds! Go to very high p_T to avoid backgrounds: **requires substructure**

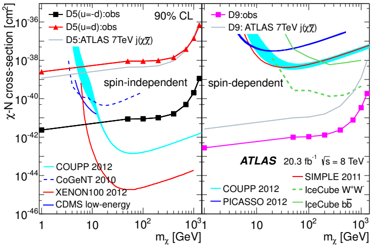




Mono-W Limits

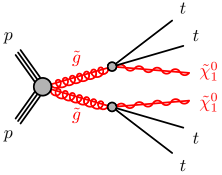
- Set limits on effective theory mass scale as a function of dark matter mass
- Making some assumptions on the effective operators of dark matter, can even show limits **compared to direct-detection experiments**

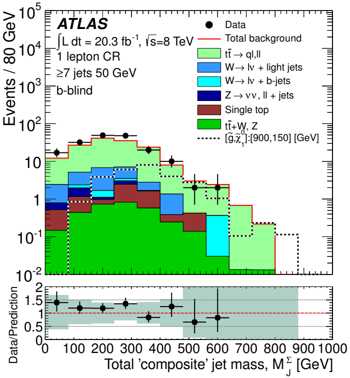




Total Jet Mass in SUSY

- A completely different approach: high multiplicity SUSY final state, lots of jets and E_T^{miss}
- Large jets here may not coincide with tops or gluinos: **but summing their mass is sensitive**
- This signal region is not the most effective at the end, but may be useful for 13 TeV





The Potential For Searches

- So far, not too many searches have utilized substructure!
  - Many more are planned: direct stop, RPV SUSY multijets, $W' \to tb$, $W' \to WZ$, exotic 4 top models, etc.
- What is preventing analyses from being released?
  - **Access to information**: SUSY multijets had to use "composite" jets, because $R = 1.0$ jets were not available
  - **More handles for optimization**: New observables mean more room for optimization, which takes time
  - **Analyzers need to help with performance**: A smaller set of users means uncertainties, etc., require more work to derive
  - **This is not low hanging fruit**: The simple analyses have already been done, so all that's left is the fun stuff :-)
- What existing searches would benefit from substructure?

Conclusions

Summary

- Substructure is a **new opportunity** for searches at the LHC
- ATLAS has been working to **test new techniques**, **measure performance**, and **study data/MC agreement**
  - These techniques are no longer “experimental”: they are commissioned and well understood!
- Data/MC agreement is, in general, **remarkably good**
  - Large samples of boosted W and top jets make validation straightforward
  - Notably, quark/gluon discrimination does not work as well as expected
    - The one island of significant disagreement?
- Techniques are already used in many searches, but **even more are in progress**
- 13 TeV will make boosted objects **even more important**: looking forward to new developments!

Thank You For Your Attention!

Backup

Extracting Templates

- Goal: to better understand quark/gluon shapes in data, extrapolate **data** to 100% purity with fractions from MC
- Ideally, solve for q/g on bin-per-bin basis from:

$$h^{\gamma+j} = P_Q^{\gamma+j} q + P_G^{\gamma+j} g$$
$$h^{dijet} = P_Q^{dijet} q + P_G^{dijet} g$$

P_Q = percentage quark
h = histogram value
q/g = templates
$(\gamma + jet)/(dijet)$ = different sample

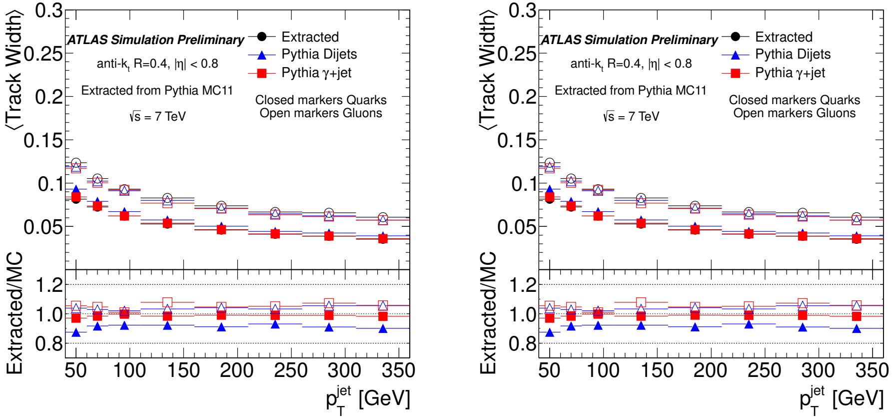
- But, need to account for b and c fractions (for now, taken from MC):

$$h^{\gamma+jet} = P_Q^{\gamma+jet} q + P_G^{\gamma+jet} g + P_B^{\gamma+jet} b + P_C^{\gamma+jet} c$$
$$h^{dijet} = P_Q^{dijet} q + P_G^{dijet} g + P_B^{dijet} b + P_C^{dijet} c$$

From Data
From MC
Solving for This

- Then, compare pure data shapes to pure MC shapes (used for training tagger)

Testing Method in MC



- MC-labeled distributions in γ+jet and dijets agree very well with templates derived in MC
  - Disagreement at low p_T will be discussed at length soon
- Gives us confidence that the algorithm is doing something sensible