BOOST 2013 Workshop Summary

Experimental results: where we are now, how we arrived here, and where we might go from here

David Miller
on behalf of all of the attendees of BOOST 2013

Enrico Fermi Institute

THE UNIVERSITY OF CHICAGO

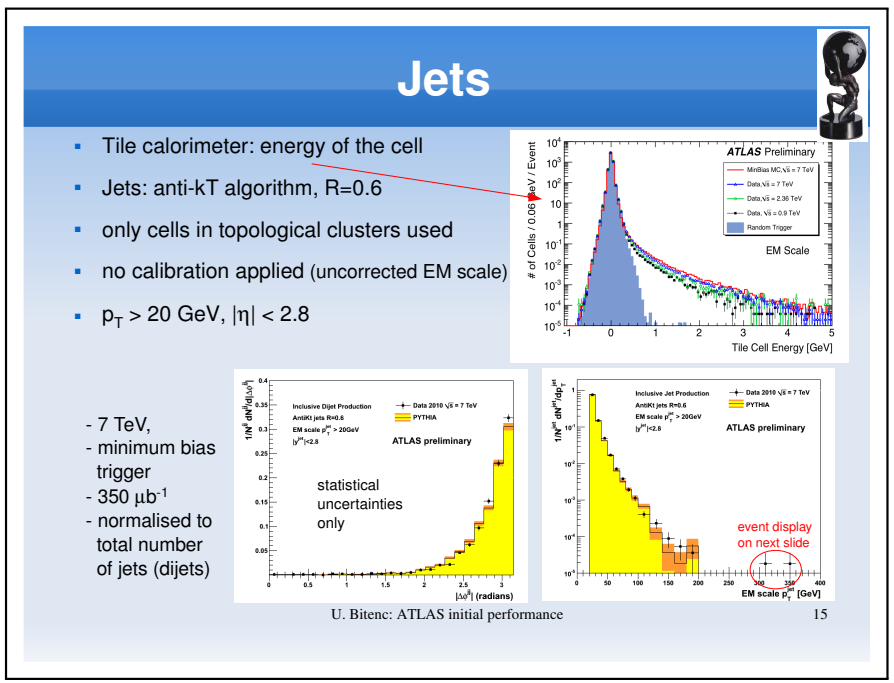
August 16, 2013

Roadmap

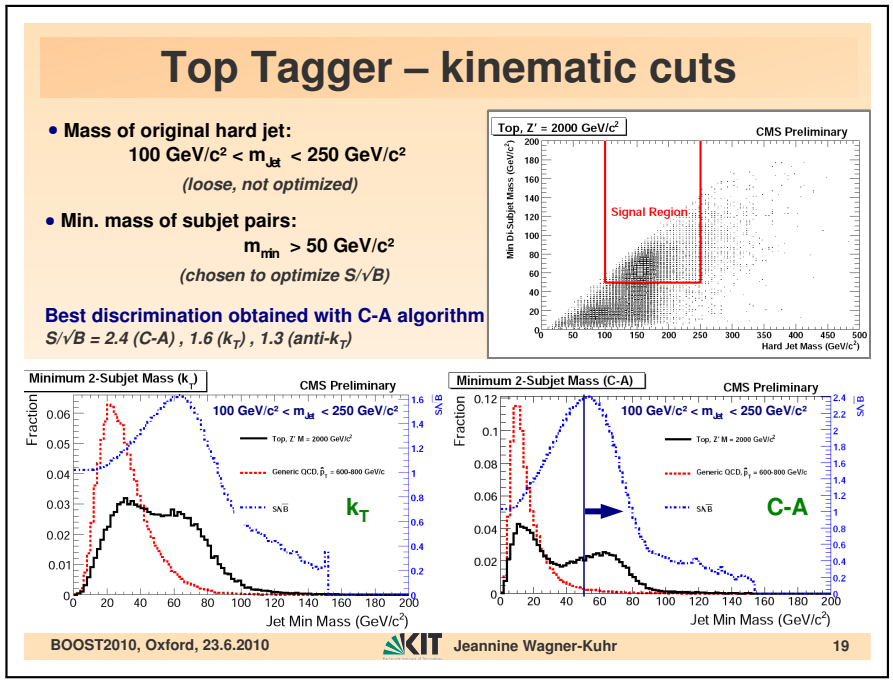
1. *A Look Back at the Experimental Progression of Substructure at BOOST*
   - BOOST 2010: These aren't your daddy's jets
   - BOOST 2011: "First" data
   - BOOST 2012: Kids in a candy store
   - BOOST 2013: Bringing substructure into the mainstream
2. *Precision Jet Substructure Measurements and Experimental Techniques*
   - Detailed calibration and validation schemes
3. *Jet Substructure at Work: Tagging*
   - New Tagging Techniques and Comparisons
4. *Jet Substructure at Work: Searching*
   - New New Physics Searches
5. *High Luminosity Issues and Mitigation Techniques*
   - Experimental Reality
   - Approaches to and performance of pile-up removal
   - Pile-up Jet Tagging
6. *Looking Towards the Future*
   - Where do our techniques break down?
   - Very high luminosity
   - Summary and conclusions

Outline

BOOST 2010: Getting ready for $\sqrt{s} = 7$ TeV LHC data



– Urban Bitenc



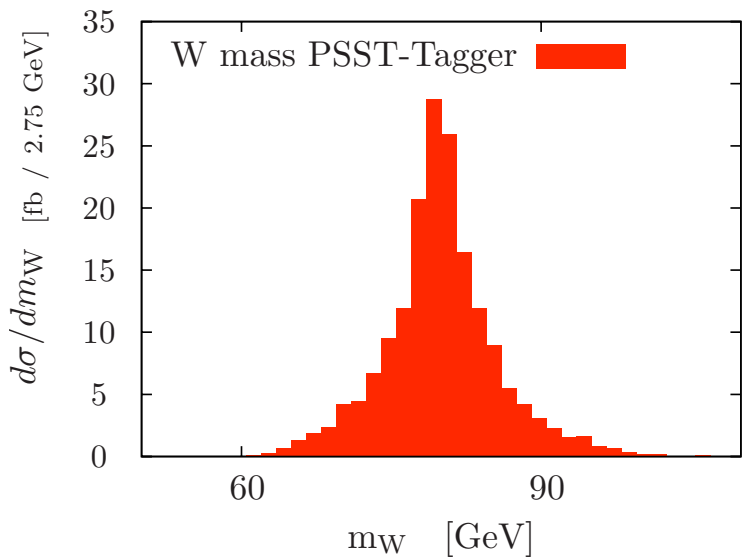
– Jeannine Wagner-Kuhr

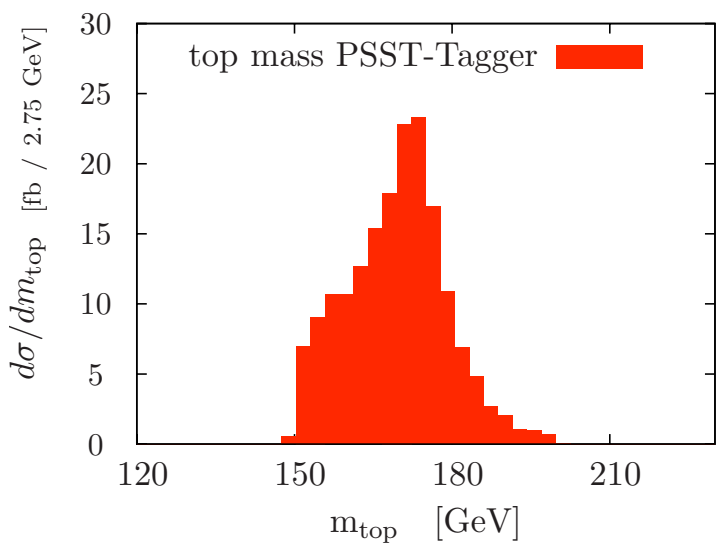
Only a few hundred $\mu\mathbf{b}^{-1}$ of data for BOOST 2010, but we already had grand plans...

BOOST 2010: Getting ready for $\sqrt{s} = 7$ TeV LHC data

Boosted 2012

Cleanly isolated samples of boosted tops and Ws from ttbar samples



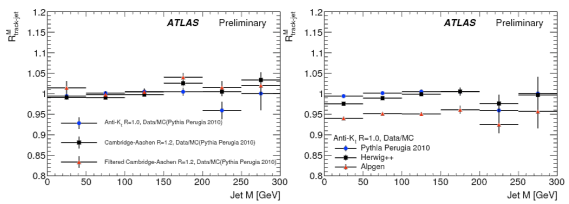


Looks like someone knew what they were talking about

BOOST 2011: Calibrating, Validating, and Exploiting Jet Substructure

Calibrating Jet Substructure

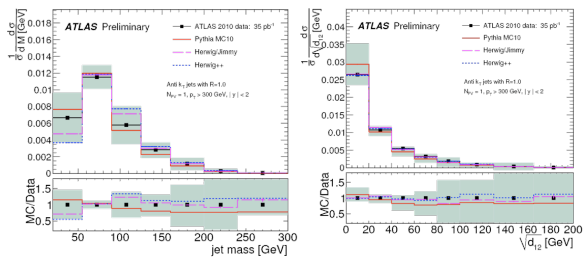
- Mass calibration from comparing tracking and calorimeter (double ratio) within tracking volume
  Lily Asquith



26 May 2011 Princeton, J.Butterworth

Validating QCD Expectations

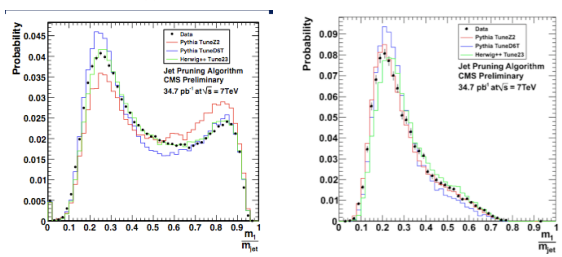
- ATLAS results (anti-k_T jet mass, y scale)
  Adam Davison



26 May 2011 Princeton, J.Butterworth

Validating QCD Expectations

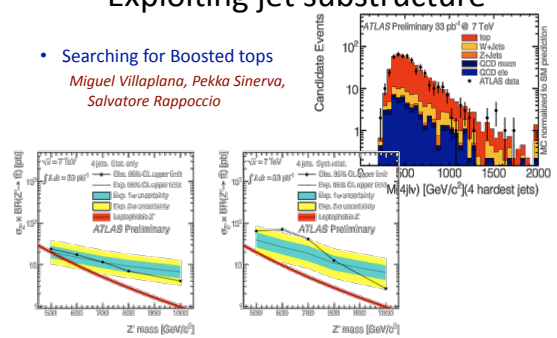
- CMS results
  – *Salvatore Rappoccio*



26 May 2011 Princeton, J.Butterworth

Exploiting jet substructure

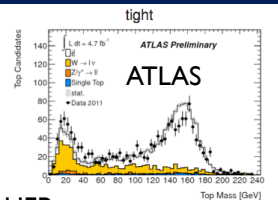
- Searching for Boosted tops
  Miguel Villaplana, Pekka Sinerva, Salvatore Rappoccio

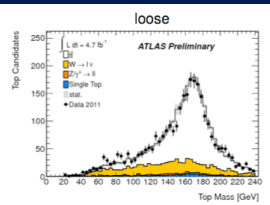


26 May 2011 Princeton, J.Butterworth

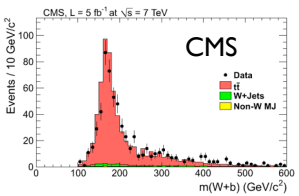
BOOST 2012: Prove that we know what we're doing and we're not afraid to use it

Boosted top in data





HEP top tagger

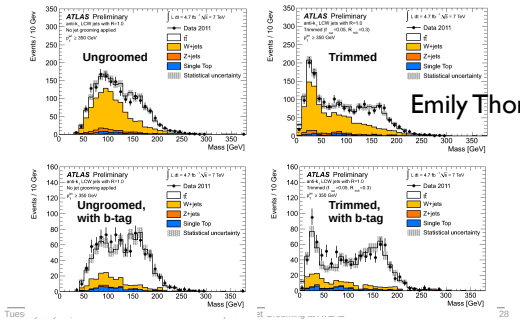


W-tagger + jet

ATLAS : Effect of grooming in top

SM ttbar events in data

Jet Mass (leading p_T anti-kt R=1.0 jet, p_T > 350 GeV)

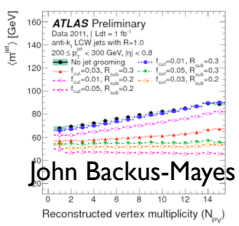


Emily Thompson

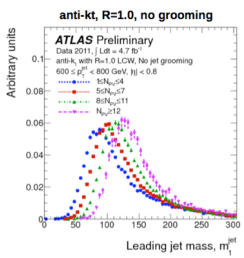
20

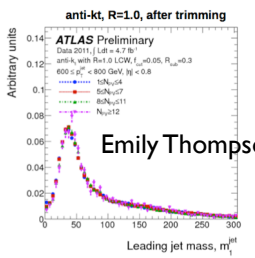
Pileup

- Grooming helps pileup dependence dramatically!



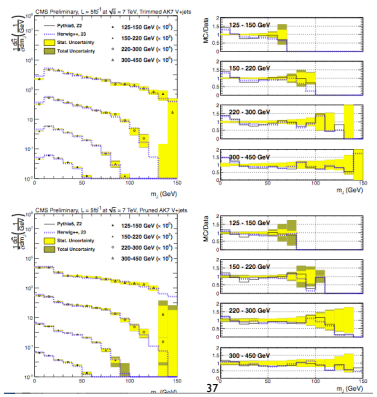
John Backus-Mayes





Emily Thompson

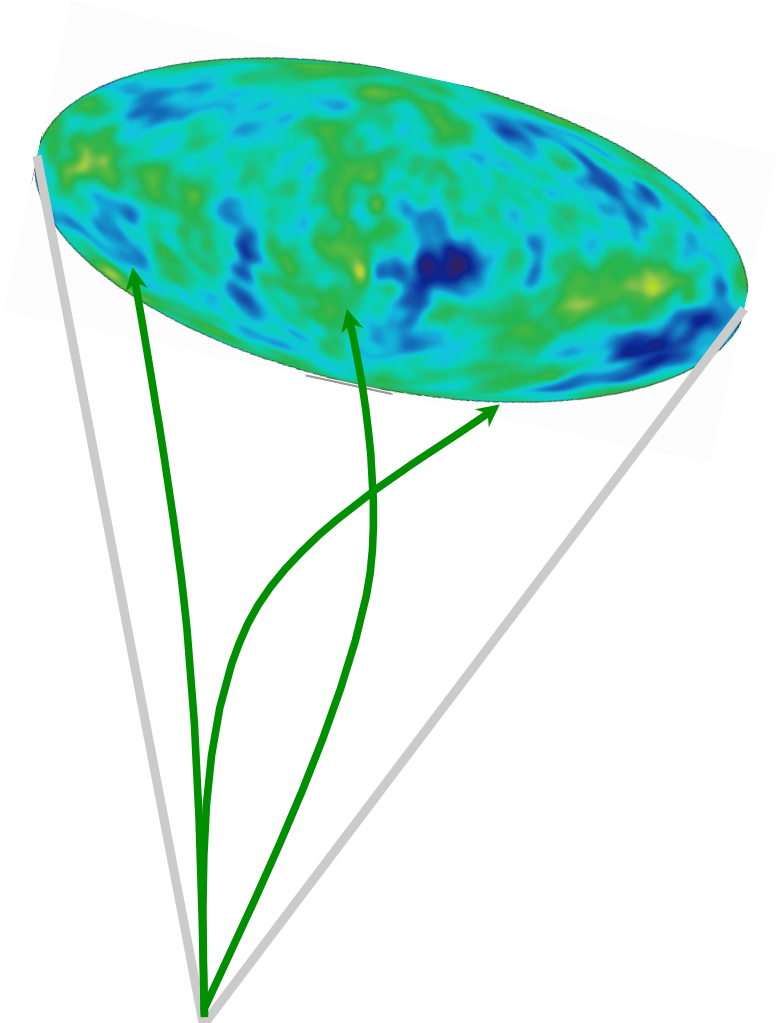
CMS : Jet substructure



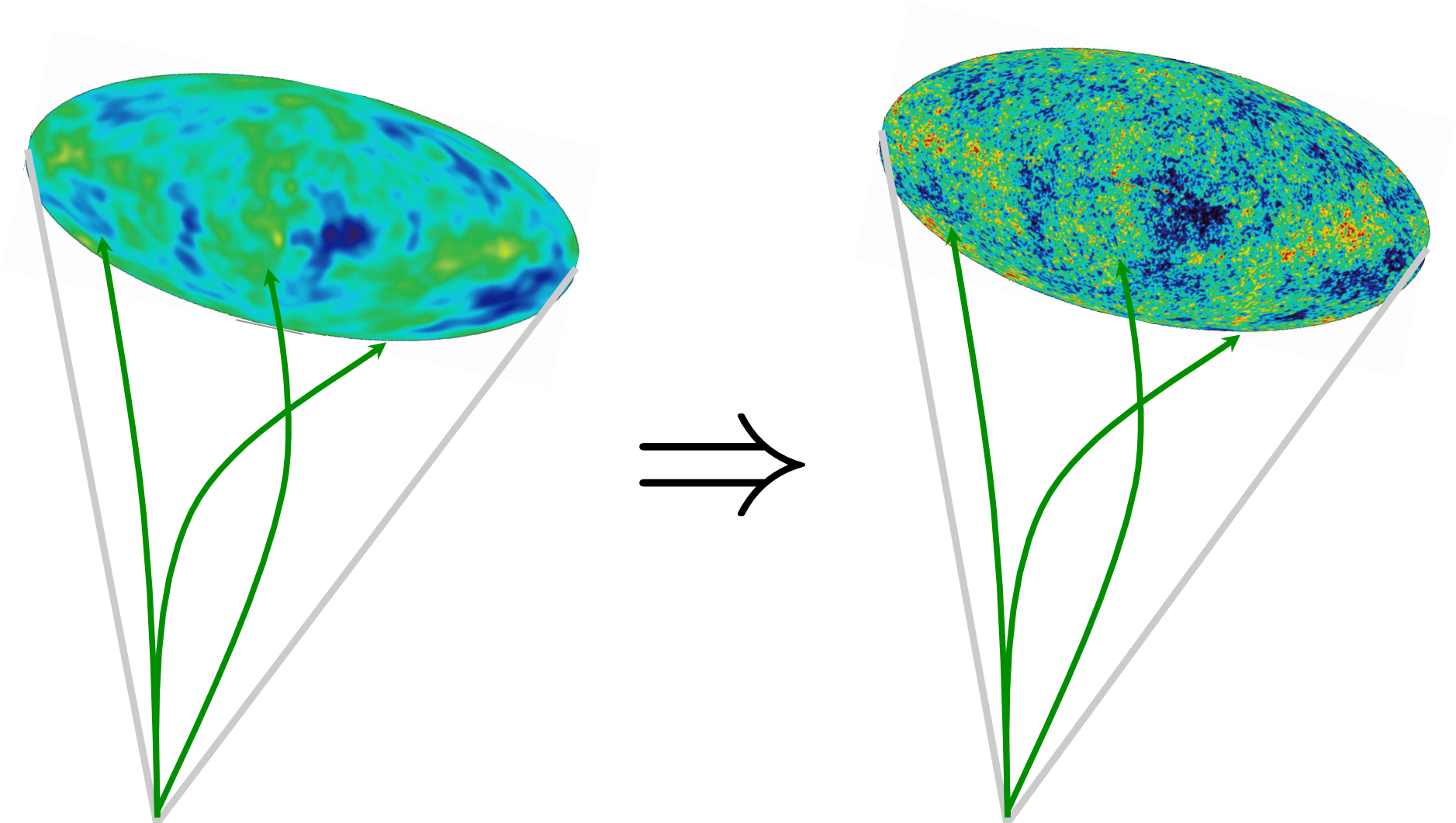
Nhan Tran

37

BOOST 2013: Beginning an era of precision substructure?



BOOST 2013: Beginning an era of precision substructure?



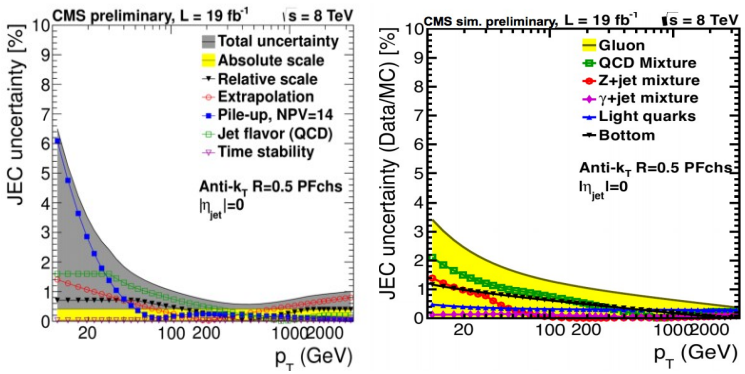
Outline

Extensive calibrations performed in data and MC for substructure and boosted objects

08/13/13 Philip Harris BOOST 49

Quarks/Gluons and JEC (JES)

- Jet energy corrections (scale) change w/flavor
- New approach in CMS : propagate flav ratio
  - Assume di-jet flavor ratio propagate herwig-pythia difference to JEC

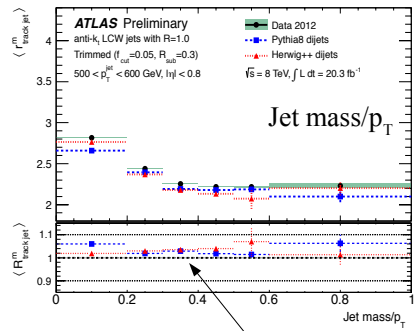


– Phil Harris

Jet Mass Ratios

https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2013Jms

Why jet m/p_T?
→ less susceptible to JES
→ each bin covers a broad range of jet masses and p_Ts.



JMS uncertainty determination:

$$\delta_{\text{Pythia/Herwig}} = \text{abs}(R^{m}_{\text{track jet}} - 1)$$
$$\delta_{\text{track}} = \max[\delta_{\text{Pythia}}, \delta_{\text{Herwig}}]$$
$$\delta_{\text{tracking efficiency}} = 0.027$$
$$\delta_{\text{track merging}} = 0.02$$
$$\epsilon_{\text{tot}} = \sqrt{\delta^2_{\text{track}} + \delta^2_{\text{tracking eff}} + \delta^2_{\text{track merging}}}$$

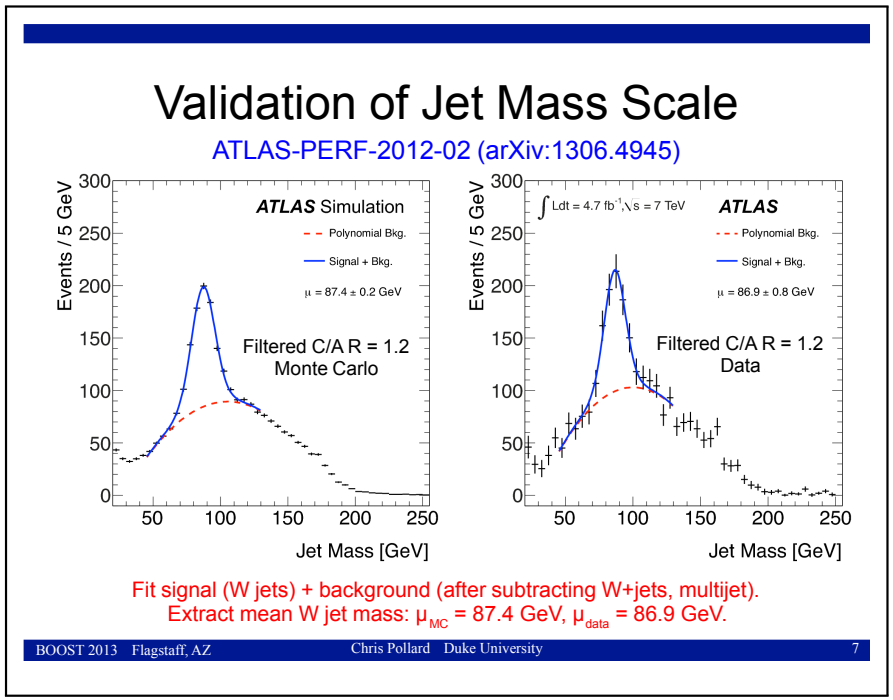
For a top-jet with p_T = 500 GeV, the JMS uncertainty is 4-5%

BOOST 2013 Flagstaff, AZ Chris Pollard Duke University 12

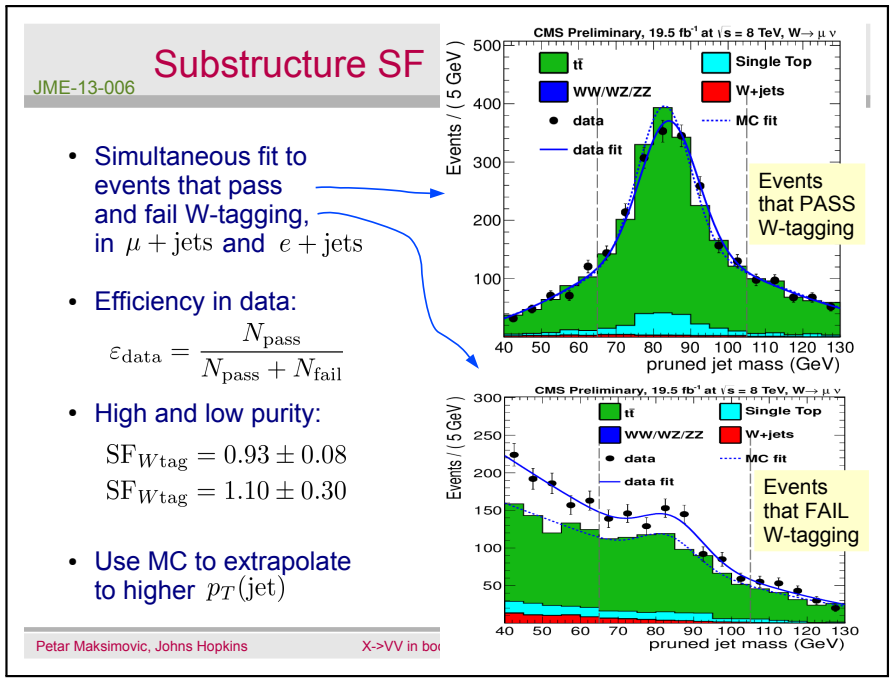
– Chris Pollard

- ATLAS and CMS JES **uncertainties have reached the 1% mark**, and keep sinking
- Proof that accounting for quarks and gluons is critical at low p_T
- Decdicated jet mass uncertainties further proof that precision physics can be done with these objects

Complex, precise calibration schemes using boosted objects



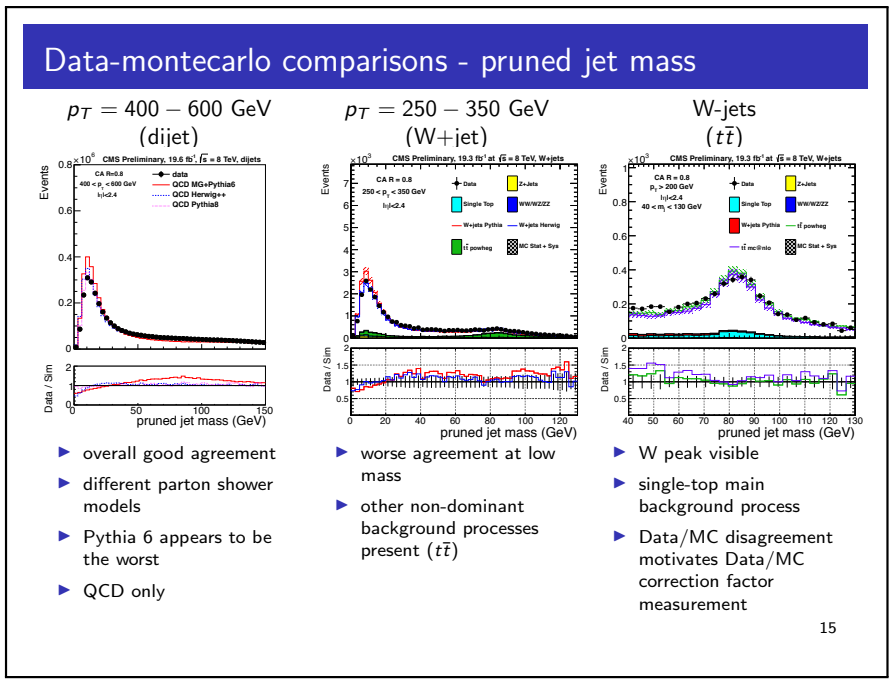
– Chris Pollard



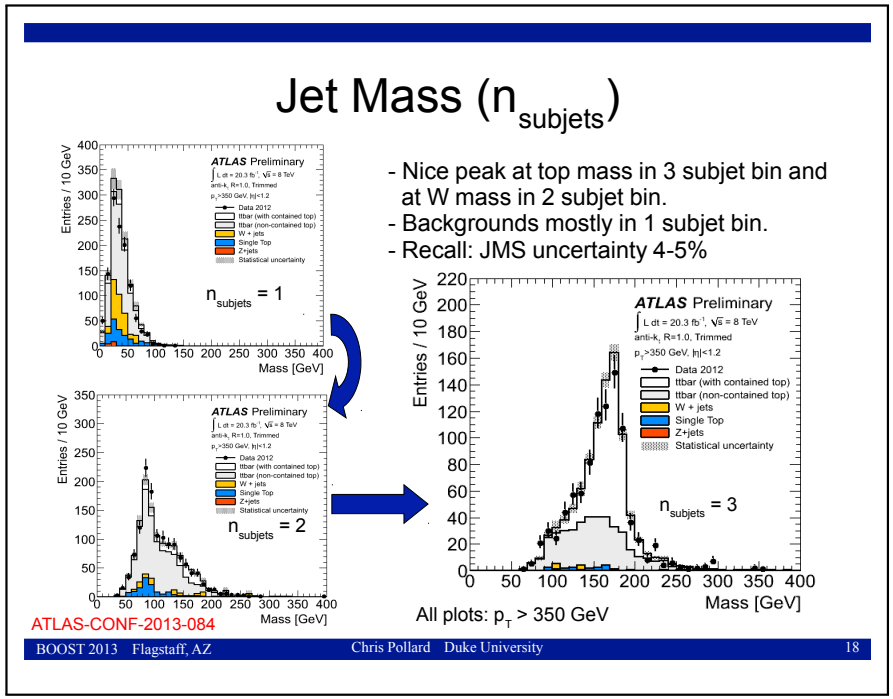
– Petar Maksimovic

- Extensive **data-driven calibration** for boosted objects
- Multiple jet algorithms in use in many analyses means that **precise and complete procedures in place** for providing the collaborations with experimental uncertainties
- Tagging efficiencies then carried out for individual analyses (typically)

Detailed experimental validation by ATLAS and CMS



– Emuanele Usai

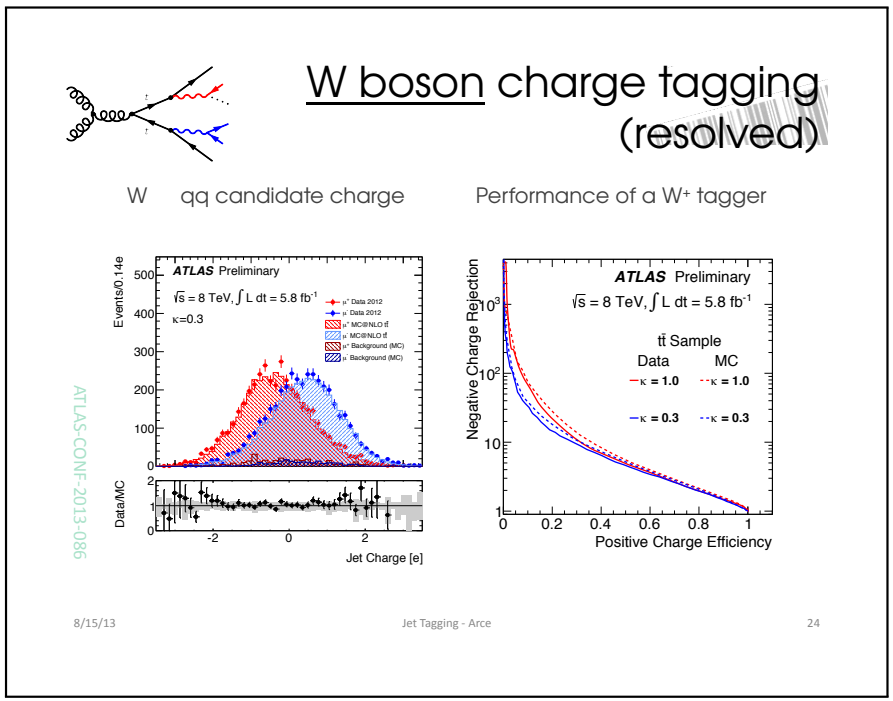


– Chris Pollard

- **Most comprehensive, detailed, impressive set of data/MC comparisons of substructure observables yet.**
- **Masses, mass drops, color-flow variables, subjet tagging, (sub-)jet counting, charge, correlations, uncertainties, ...**
- *All of which ATLAS and CMS have tested in various data samples, and usually between several MC generators, parton showers, and tunes*

Outline

Charge Tagging For QCD and W's



– Ayana Arce

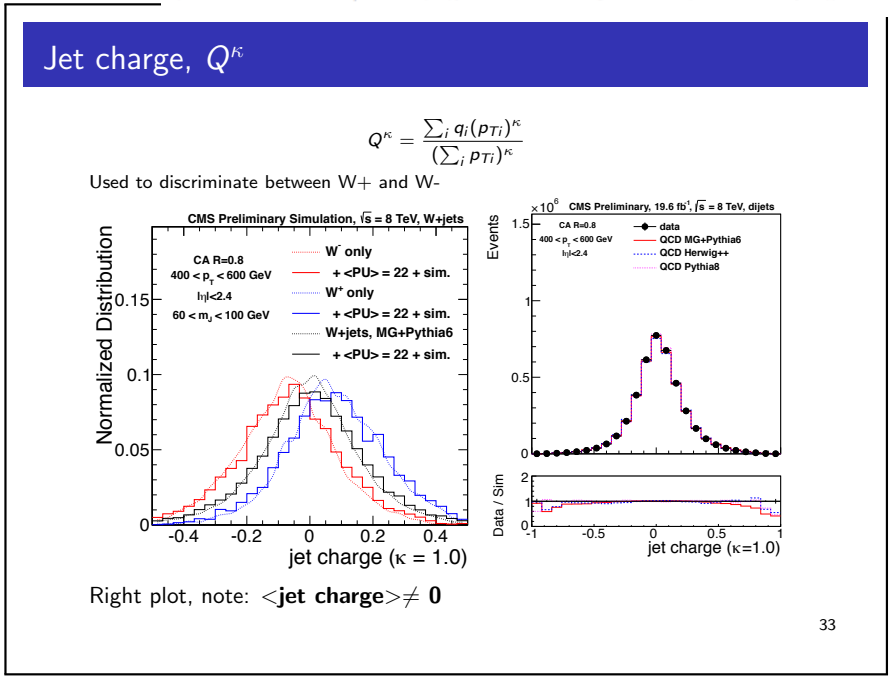
arXiv.org > hep-ph > arXiv:1209.2421

High Energy Physics – Phenomenology

Jet Charge at the LHC

David Krohn, Tongyan Lin, Matthew D. Schwartz, Wouter J. Waalewijn

(Submitted on 11 Sep 2012 (v1), last revised 14 Jun 2013 (this version, v2))



– Emuanele Usai

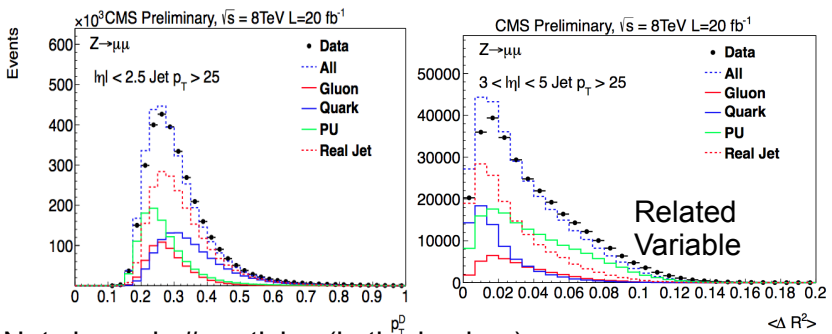
- **Amazing example of the importance of BOOST:** from "proposal" by Krohn, Lin, Schwartz, Waalewijn to measurement by ATLAS and CMS in less than 1 year (Sep 2012→Aug 2013)
- Excellent modeling by MC's and performance in the experiment

Quark-Gluon Tagging

08/13/13 Philip Harris BOOST 25

Quark/Gluon Disc. Concept

- Similar concept to the pileup jet id
  - Must rely on the shower shape to discriminate
- Concept
  - Several of these variables overlap with PU Jet id



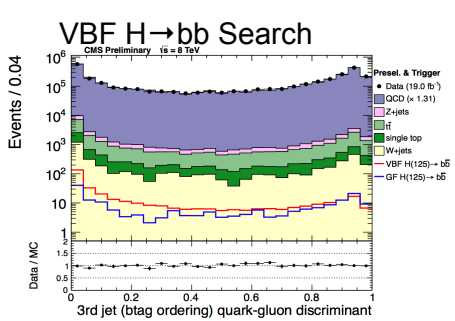
Not shown is # particles (in the backup)

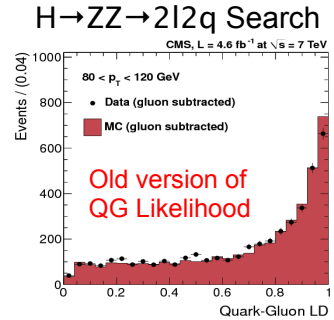
– Phil Harris

08/13/13 Philip Harris BOOST 26

QG Performance + Usage

- QG discrimination used in VBF selection
  - Reduces the QCD/Pileup bkgs for forward jets
- QG discrimination used in Z boson tagging
- Reduction of 60% gluon for 80% quark eff





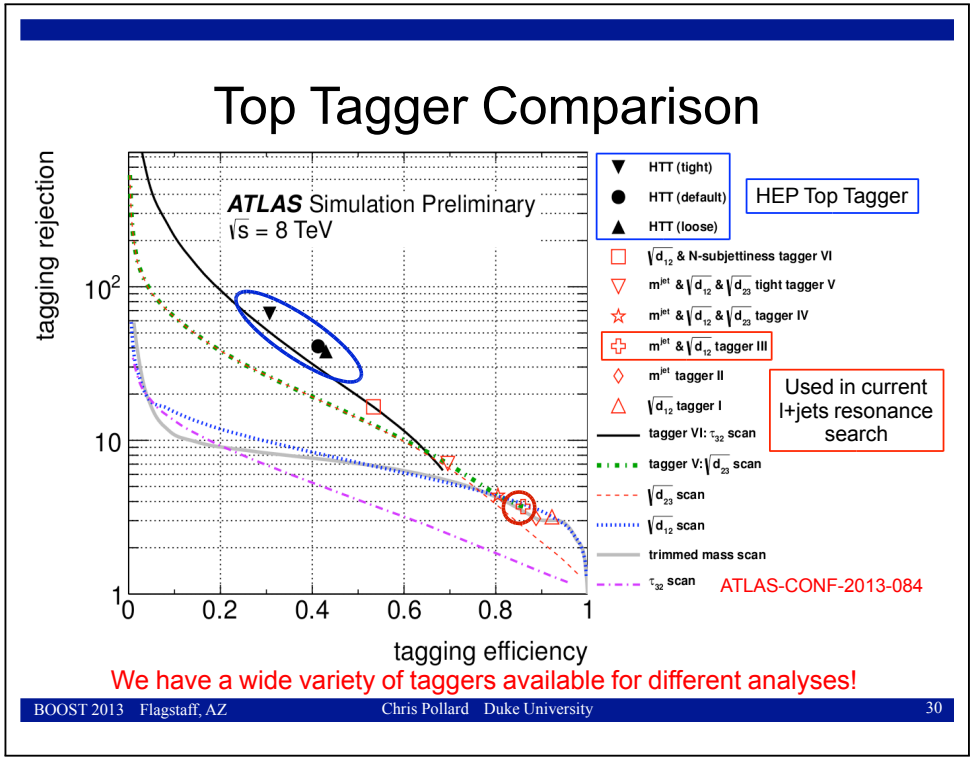
– Phil Harris

- Extensively **discussed both ATLAS and CMS results at BOOST 2012**
- **New observables for tagging** described by CMS this year
- Very important impact on Higgs searches
- **See this as crucial arena for discussion between ATLAS/CMS and theory; significant impact of calculations and theory guidance**

Near-Exhaustive Top and W Tagger Comparisons

But much more work to do!

- **We have seen the implementations, observables, and applications of top and *W* taggers evolve significantly over the past few years**
  - New taggers and observables than when plans were laid out in 2012
  - Some observables slightly better modeled, or less sensitive to pile-up
  - High efficiency vs low fake rate optimizations
- **Important job for both experimentalists and theorists to help organize this playing field**



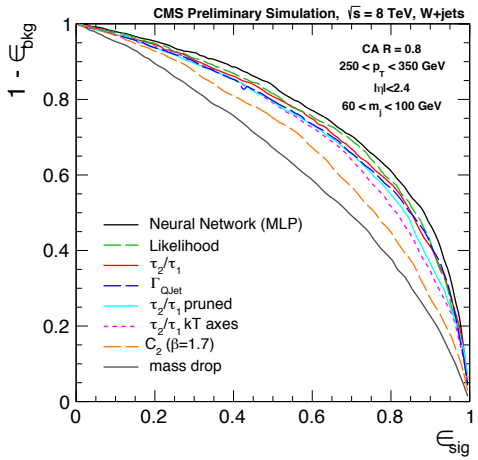
– Chris Pollard

- Nonetheless, we do need to understand in more detail these correlations
- *I personally appreciate the efforts to understand the underpinning of these curves and the taggers and groomers that make this plot possible*

Detailed Studies of Tagging Efficiencies, Correlations, Optimization

Optimization

$60 < m_{\text{pruned}} < 100$ GeV



Mistag vs efficiency

- background: QCD, signal: $H \to WW$
- pruned jet mass cut $60 < m < 100$ GeV
- τ_2/τ_1 best single variable discriminator
- neural network trained with TMVA shows improvement over τ_2/τ_1
- other variables correlated with τ_2/τ_1

MVA variables

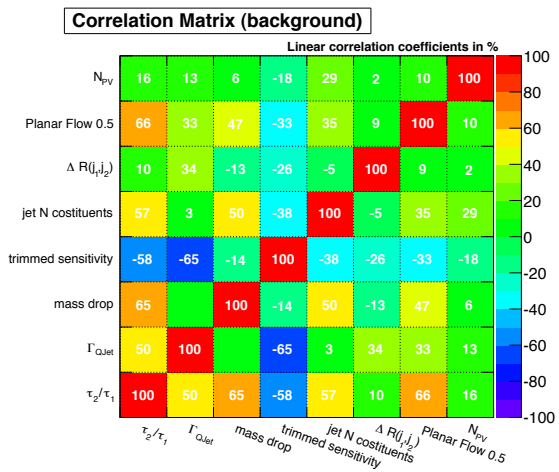
- Pruned mass drop m_{pr}^{drop}
- Q-jet volatility Γ_{Qjet}
- N-subjettines τ_2/τ_1
- Planar flow R = 0.5
- Number of jet constituents
- Subjet ΔR
- Trimmed grooming sensitivity
- Number of primary vertices N_{PV}

24

– Emuanele Usai

Correlation matrix (background)

Correlation matrix for the input variables given to the MVA evaluated on background (W+jets Pythia $p_T > 180$ GeV)

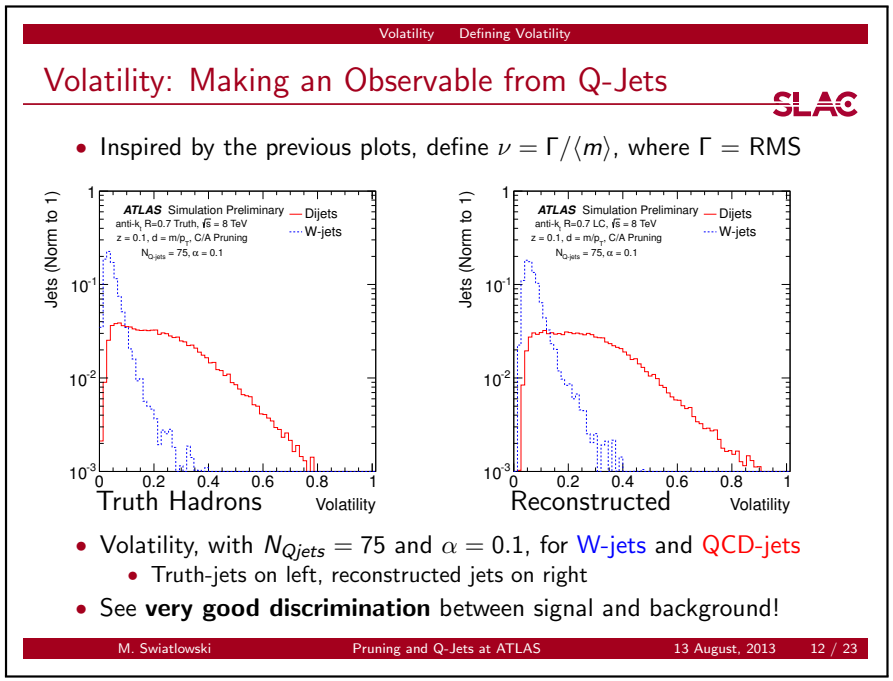


26

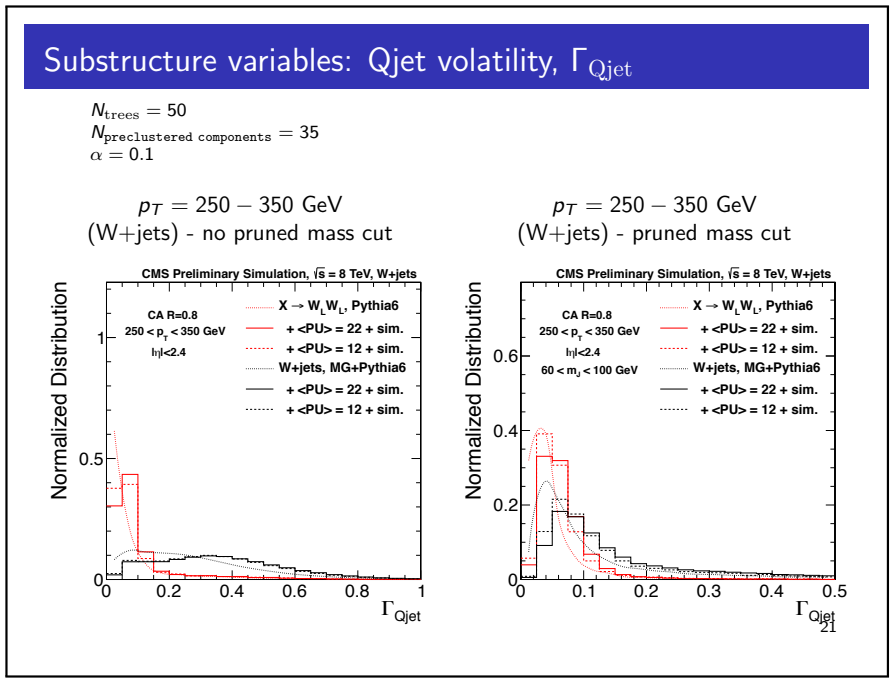
– Emuanele Usai

- *W*-jet tagging forms the **canonical proving ground on which to test these complex tools**
- **Color-singlet** provides stable foundation for issues of contamination, correlations, and optimization of ROC curves
- This year we have **started to see very detailed comparisons including correlation matrices** like this one

Extending Tagging with Q-Jets



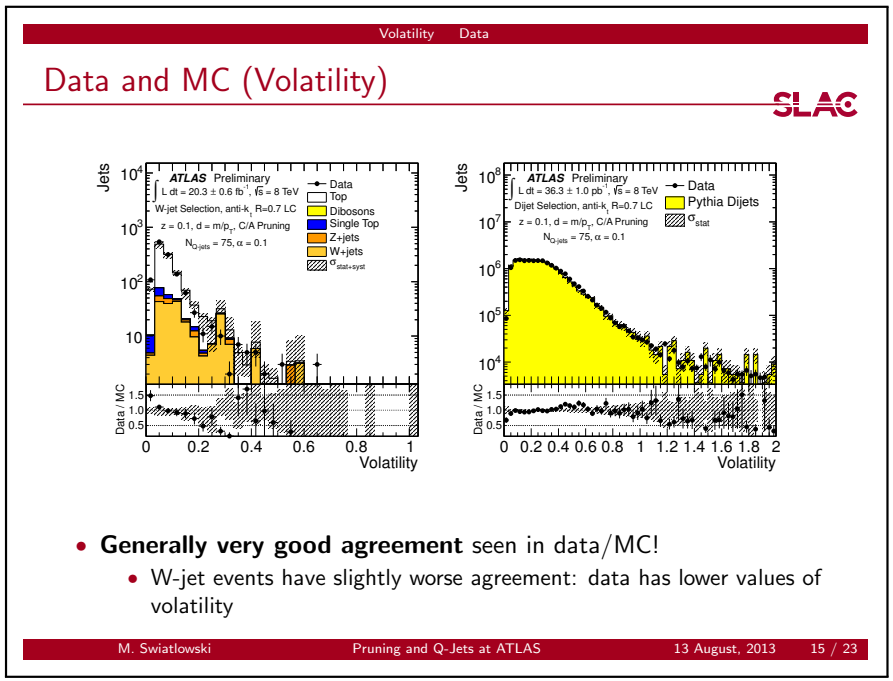
– Max Swiatlowski



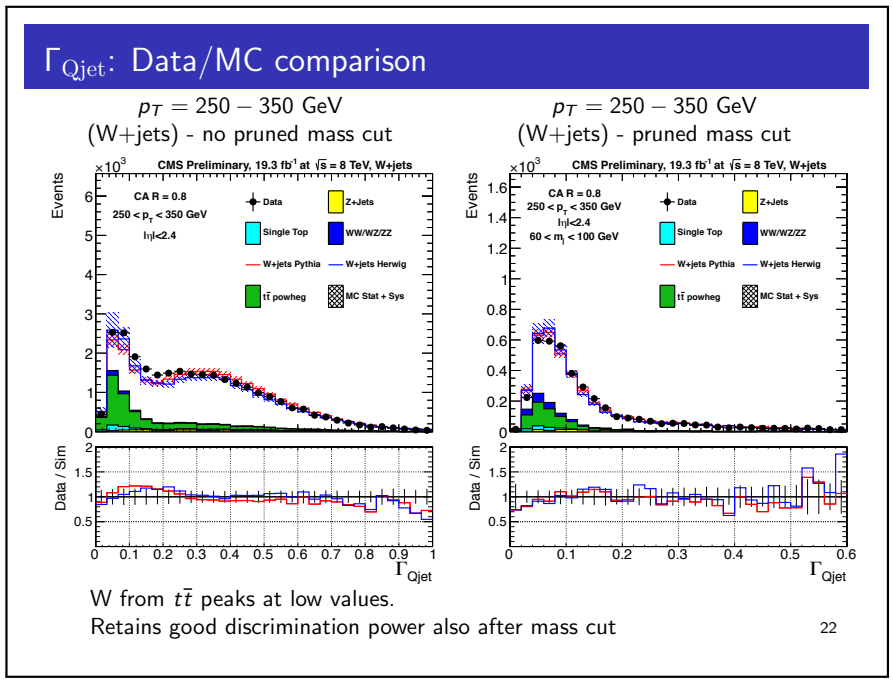
– Emuanele Usai

- *Q*-jets are a perfect example: highly non-trivial concept (and implementation)
  - Again a new observable...from concept to experimental reality in 20 months
- See **excellent *S* vs *B* discrimination** with full detector simulation
- Important stability with pile-up

Extending Tagging with Q-Jets



– Max Swiatlowski



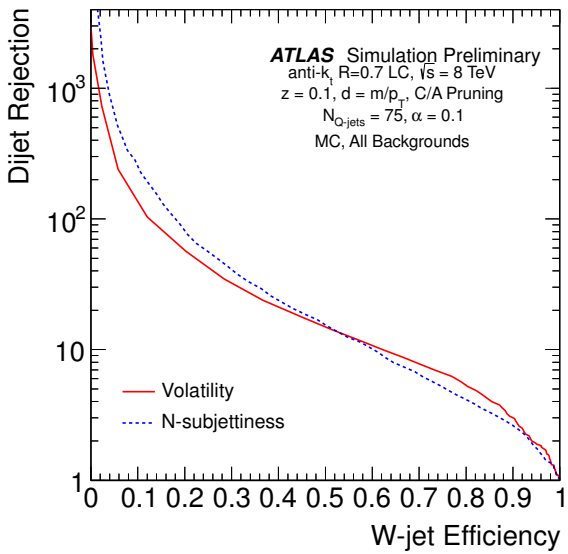
– Emuanele Usai

- Moreover, it's **well-described by MC models for multiple event samples** for both signal and background
- Retains good discrimination after mass cuts and sample purification

Extending Tagging with Q-Jets

Efficiency/Rejection with N-subjettiness

SLAC

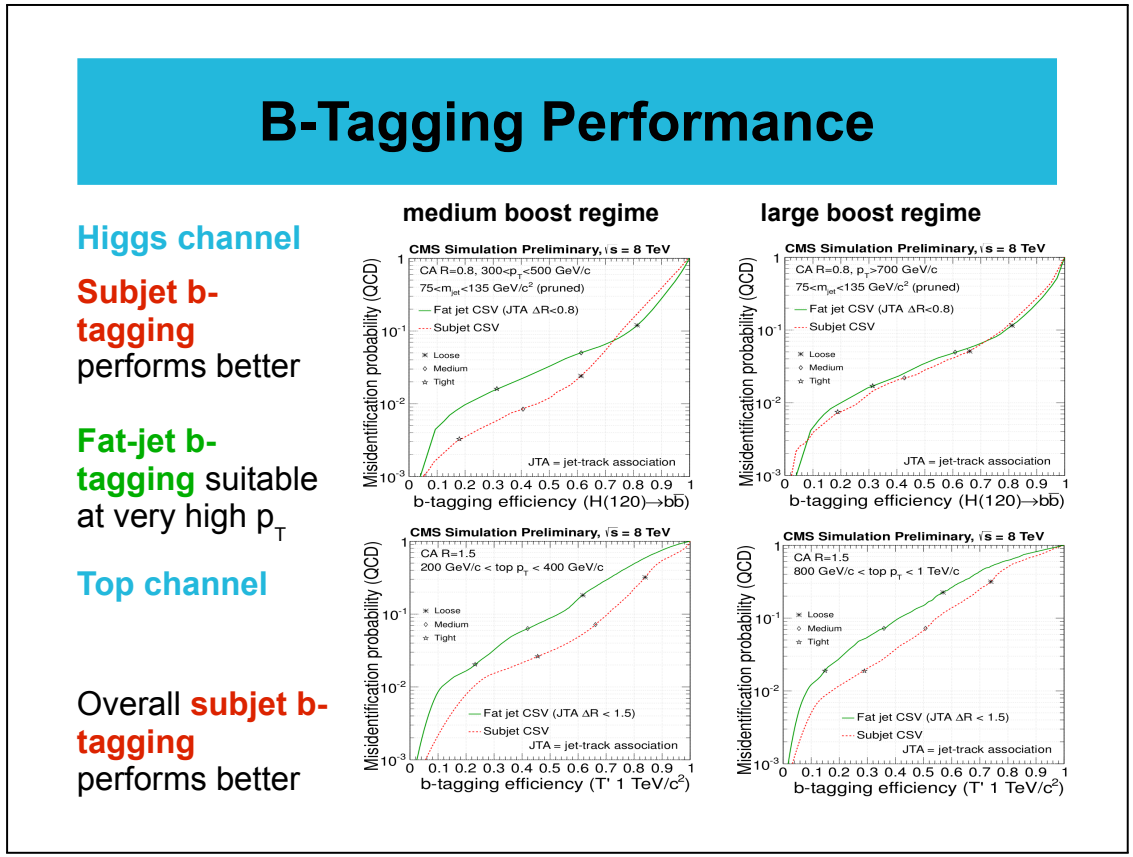


- Now, compare the full ROC curve for both variables
- **See generally similar performance**
  - At high efficiency, volatility is a little stronger
  - At low efficiency, N-subjettiness is a little stronger
- Next step: a combination, exploiting the strengths of each

- **Homework for Boston Jet Physics Workshop and BOOST 2014:**
  Can we understand this cross-over point in detail?

Subjet b-Tagging

- Extensive, in-depth discussion of subjet *b*-tagging from CMS
- **Extremely important topic as we move forward to 2015**
- Need to outline the **requirements and the test samples, as well as benchmarks for the resulting performance very clearly**

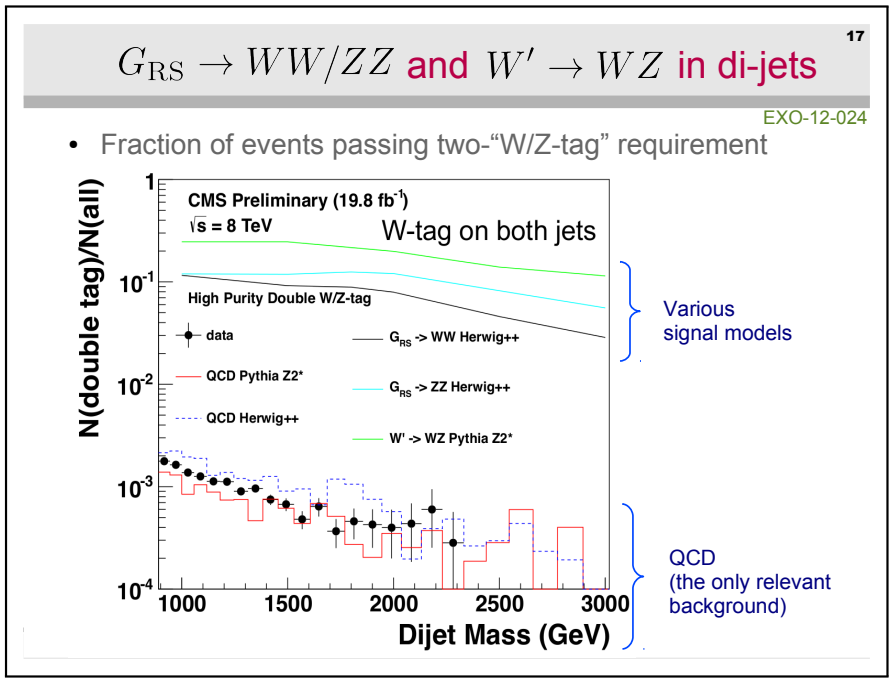


– Ivan Marchesi

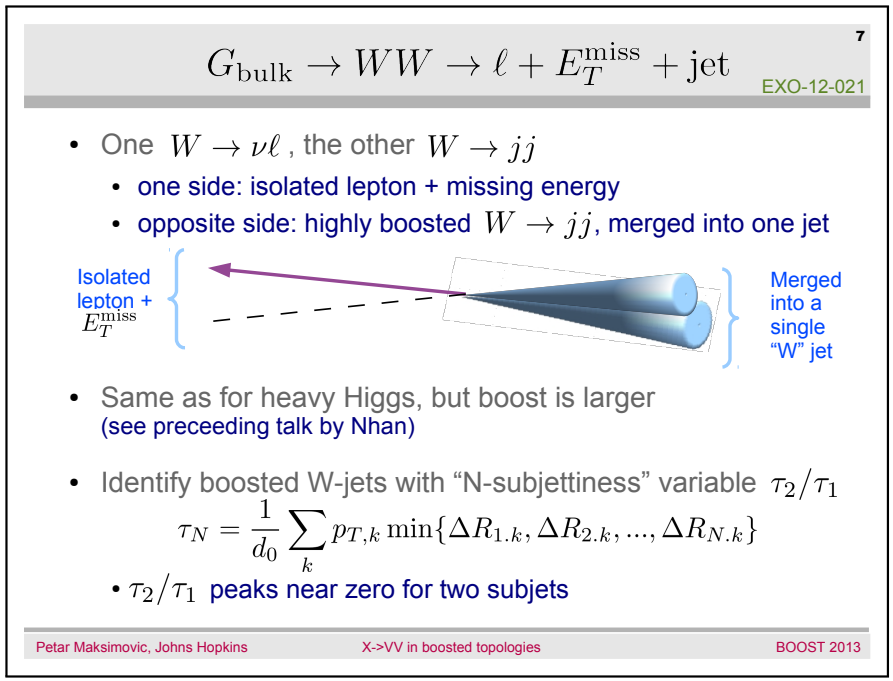
- Is track-sharing an issue? How can we overcome it?
- Do we need a new approach to SV finding and tagging?
- Do we need the calorimeter at all?

Outline

Searches Using Intrinsically Boosted Systems: W's and Z's



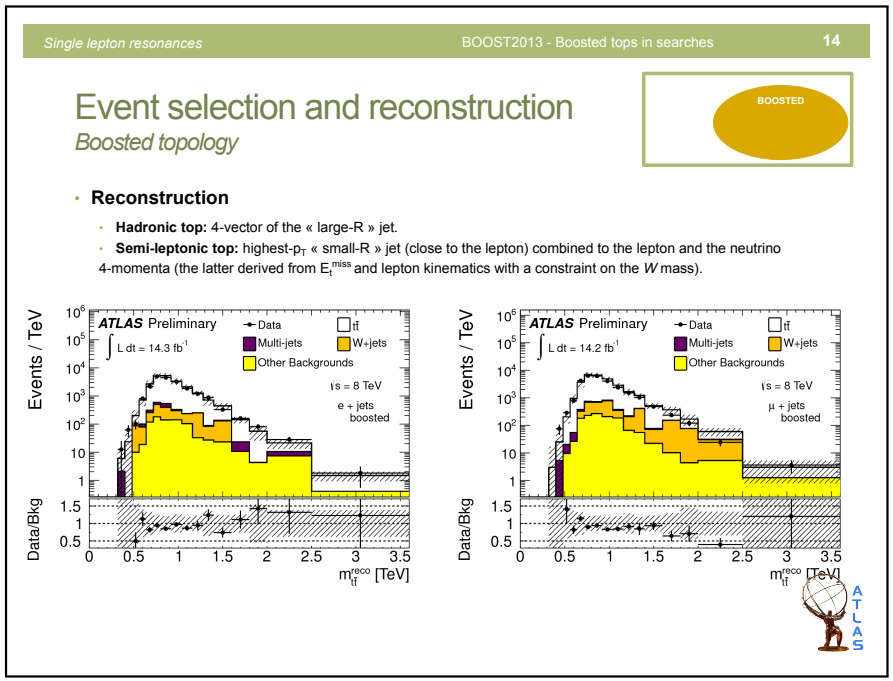
– Petar Maksimovic



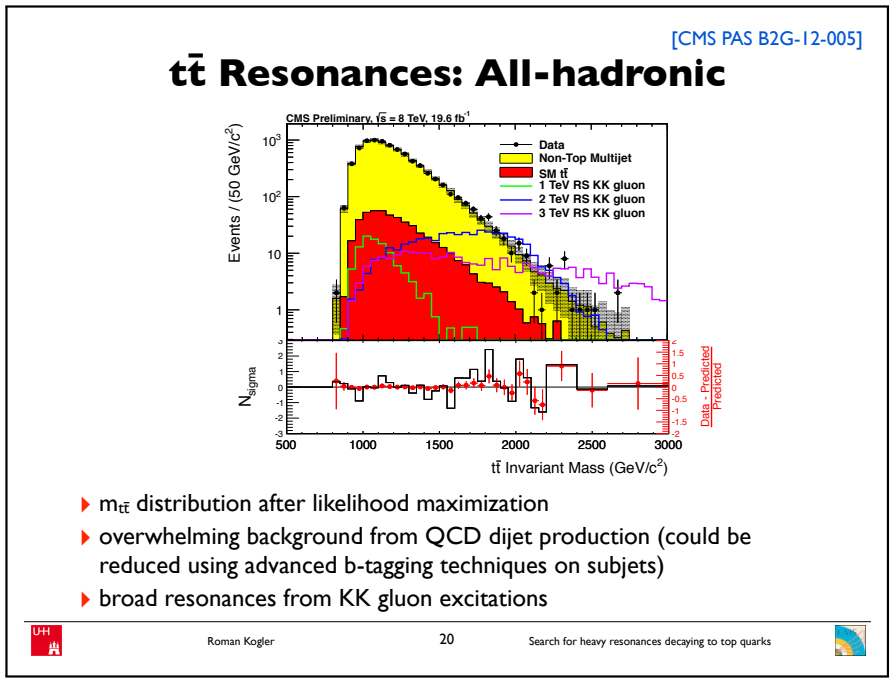
– Petar Maksimovic

- Models that predict high-mass particles and couplings to vector bosons and tops are of course the **bread and butter of this workshop**
- Several new and updated searches for such signals
- But, most important (in my mind): **excellent data-driven background estimations techniques for boosted hadronic objects**

Searched Using Intrinsically Boosted Systems: $t\bar{t}$pairs



– Loïc Valery



– Roman Kogler

- The search continues for the benchmark models that, in part, started this enterprise
- See several updates of background estimations approaches and systematic uncertainty reductions
- **Question for the future (and later on in this summary): When do we hit the wall at high p_T and $M_{t\bar{t}}$?**

Searched Using Intrinsically Boosted Systems: $t + W$

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Reconstructed T Mass

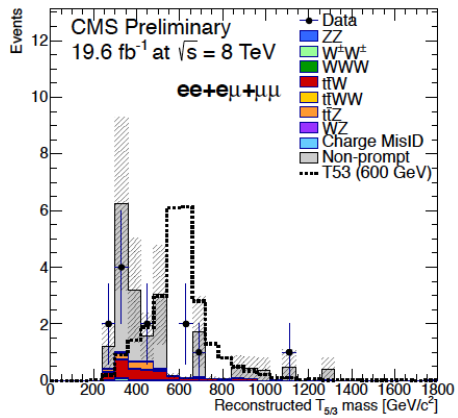
Use substructure information to reconstruct events

Slightly different selection than for limit setting:
≥ 5 constituents of T decay must all be jets in this case

Reconstructed T mass:
Invariant mass of top and W candidates

Top candidates are:
A) top-tagged CA8 jets
B) W-tagged CA8 jet + ak5 jet

W candidates are:
A) W-tagged CA8 jets
B) two ak5 jets with an invariant mass Within m_W ± 20 GeV



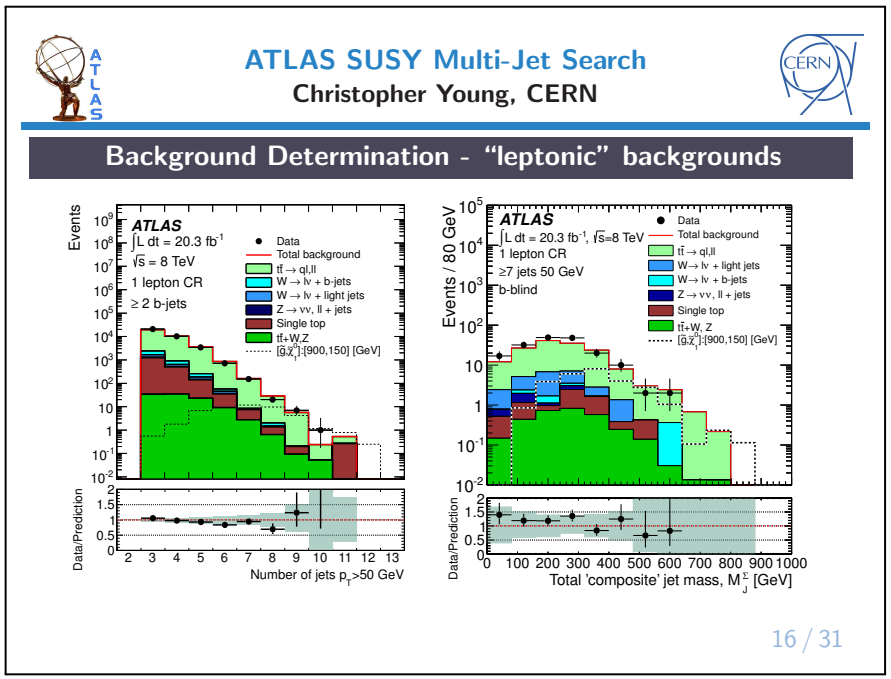
Future discovery → use distributions to distinguish from other exotic particles

Rebekka Sophie Höing 15

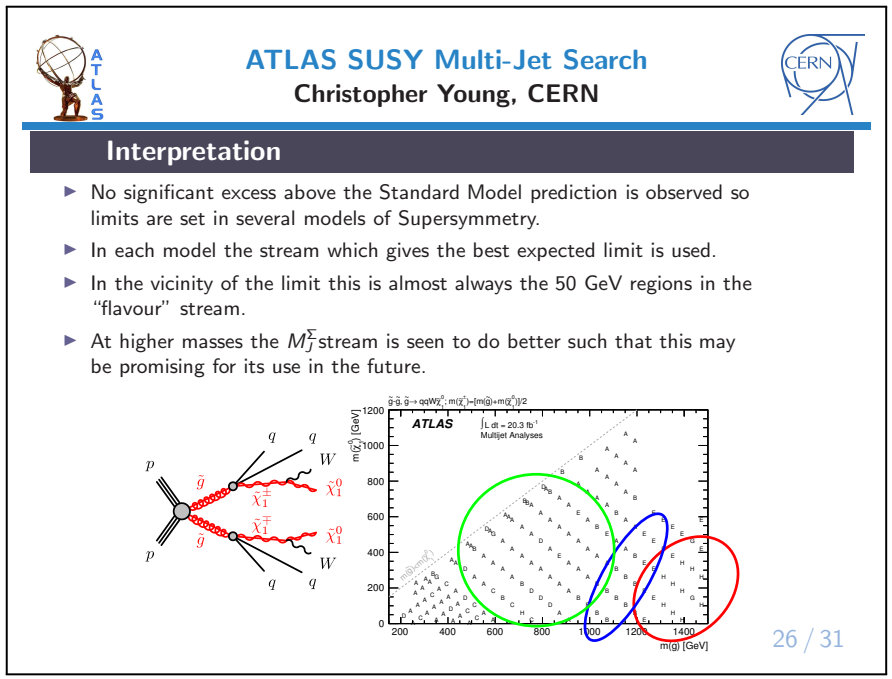
– Rebekka Hoing

- New searches and applications for BSM this year!
- Top partner limits and other exotics are pushing the mass region where boosted objects are inevitable
- If you can't beat 'em, join 'em!

Using Jet Substructure for Multijet Event Shapes



– Chris Young



– Chris Young

- New ideas put forth around time of BOOST 2012 now being seen in the experiment
- Active discussion surrounding the question: **to what extent is it useful to use the techniques we've been discussing for non-intrinsically boosted final states?**
- So far, no **"kill app"** in this context, but indications that the technique is worthwhile to consider

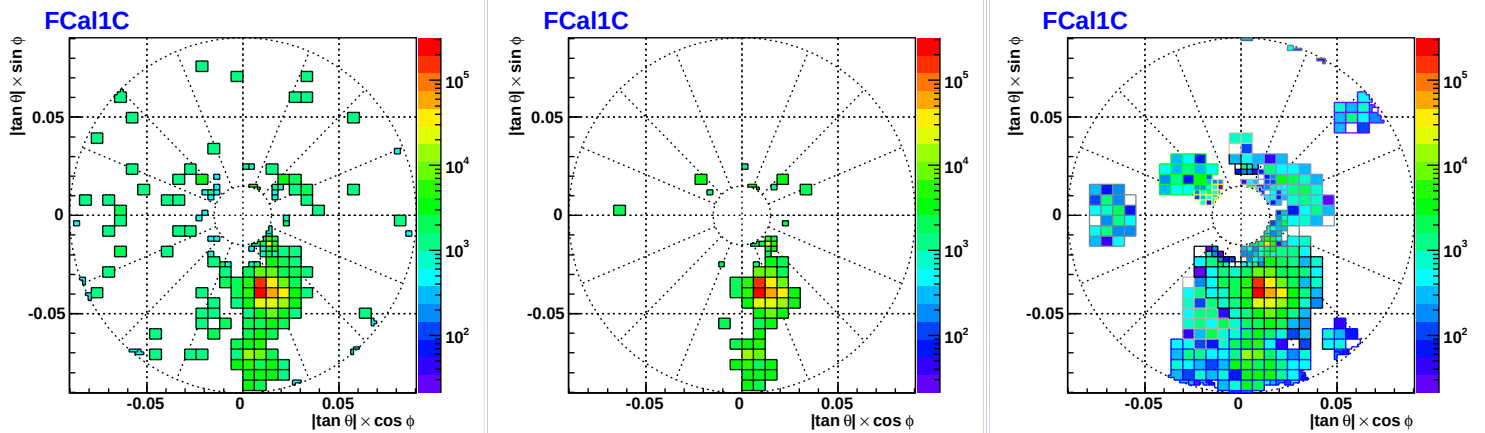
Outline

Experimental Reality

Jets

- jets in ATLAS are made out of "topological clusters"
  - 3d energy blobs of neighboring calorimeter cells around seed cell with $|E| > 4\sigma$
  - direct seed neighbors with $|E| > 2\sigma$ become seeds too
  - re-clustering of this reduced cell set around local maxima

$|E| > 2\,\sigma_{\text{noise}}$ $|E| > 4\,\sigma_{\text{noise}}$ 4/2/0 topological clusters



- $2\,\sigma$ cut is removing cells from the signal region
- $4\,\sigma$ cut shows seeds for the cluster maker
- after clustering all cells in the signal regions are kept
- cluster splitter finds hot spots

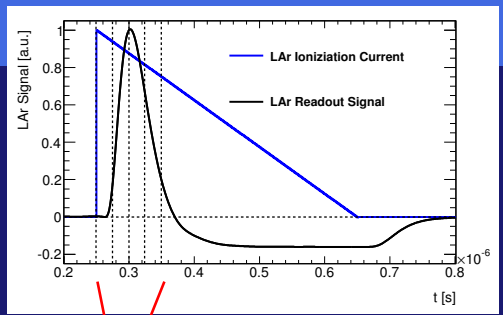
S. Menke, MPP München ◄ Pile-Up in Jets in ATLAS ► BOOST, 12-16. Aug 2013, Flagstaff, AZ 8

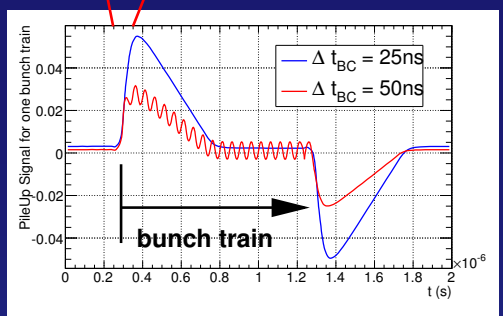
– Sven Menke

Experimental Reality

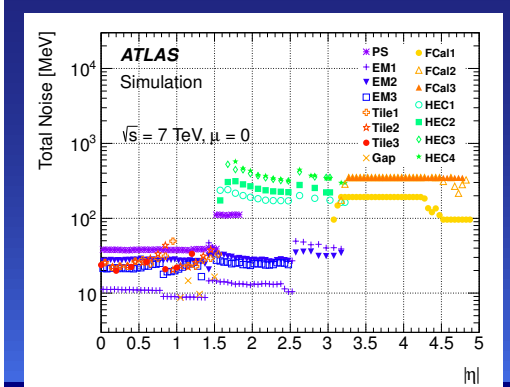
Jets ► Noise Thresholds

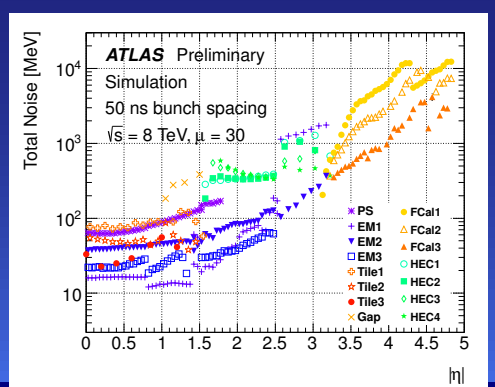
- Noise is $\sigma_{\text{elec}} \oplus \sigma_{\text{pile-up}}$
  - σ_{elec} relevant for no pile-up
  - $\sigma_{\text{pile-up}}$ grows with $\sqrt{\mu}$
- a 20% increase in noise means 20× more clusters for fixed thresholds!
- thresholds and filter weights are adjusted to expected maximum $\langle\mu\rangle$
  - modified weights slightly increase σ_{elec} and decrease $\sigma_{\text{pile-up}}$ from $\sqrt{\mu}$ scaling
- average cell-level bias
  - $f(\text{BCID}) \lesssim O(0.3\sigma_{\text{pile-up}})$ corrected in 2012

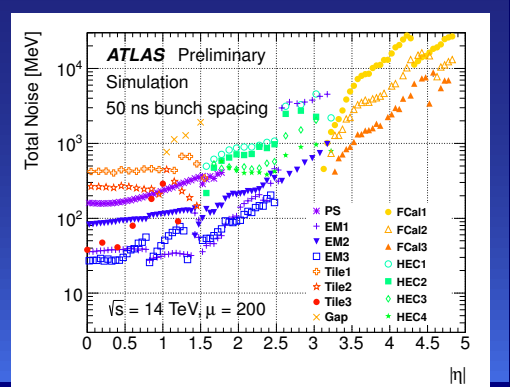




2010 $\mu = 0$ | 2012 $\mu = 30$ | 2022 $\mu = 200$?







S. Menke, MPP München ◄ Pile-Up in Jets in ATLAS ► BOOST, 12-16. Aug 2013, Flagstaff, AZ 9

– Sven Menke

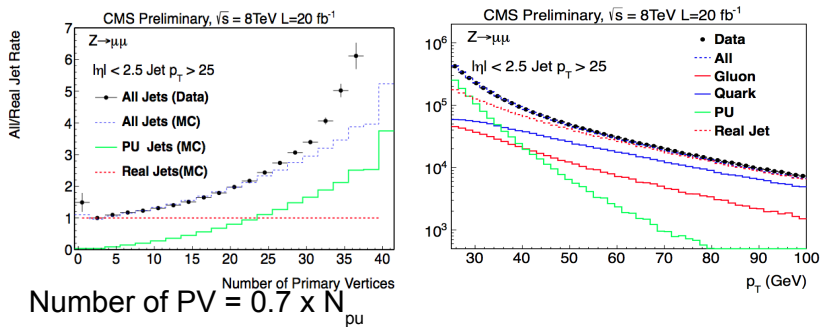
Addressing pile-up in the data

08/13/13 Philip Harris BOOST 8

What does it look like in Data?

$P(\text{overlap}|pT) \approx C\ N^2_{pu}\ a^2_{jet} pT^{-6.2}$

- Expect pileup to grow quadratically
- Expect pileup jets to fall off more rapidly
  - Pileup jets remain a few% level problem up to 70 GeV

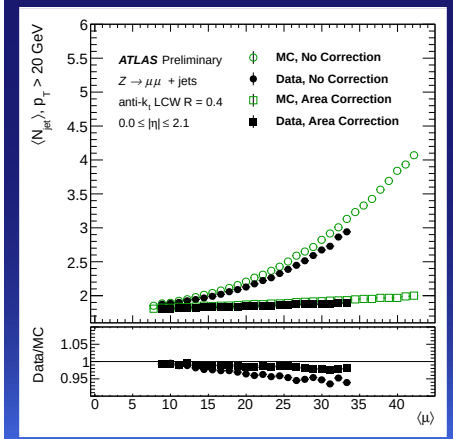


Number of PV = 0.7 x N_{pu}

– Phil Harris

Jet Area ► Jet Multiplicity

- pile-up subtraction with jet area method reduces amount of pile-up inside jets
- but also suppresses jets with $p_\perp^{jet} < \rho \times A_{jet}$



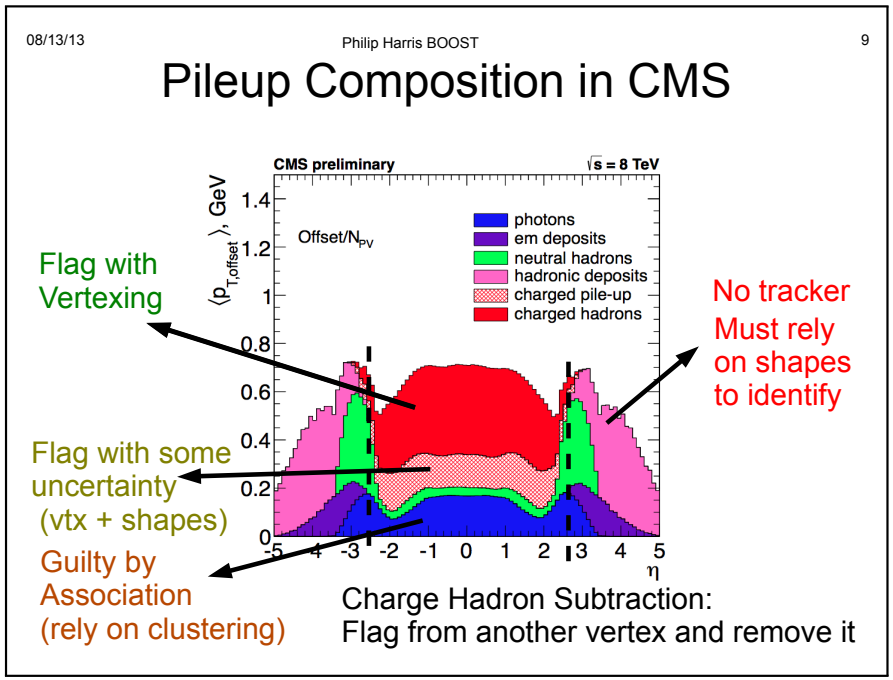
- jet multiplicity as function of $\langle\mu\rangle$ in $Z \to \mu\mu$ + jets events before and after pile-up correction
  - pile-up correction improves data/MC agreement
  - reduced μ dependency
- still some jets remain due to localized pile-up fluctuations

S. Menke, MPP München ◄ Pile-Up in Jets in ATLAS ► BOOST, 12-16. Aug 2013, Flagstaff, AZ 14

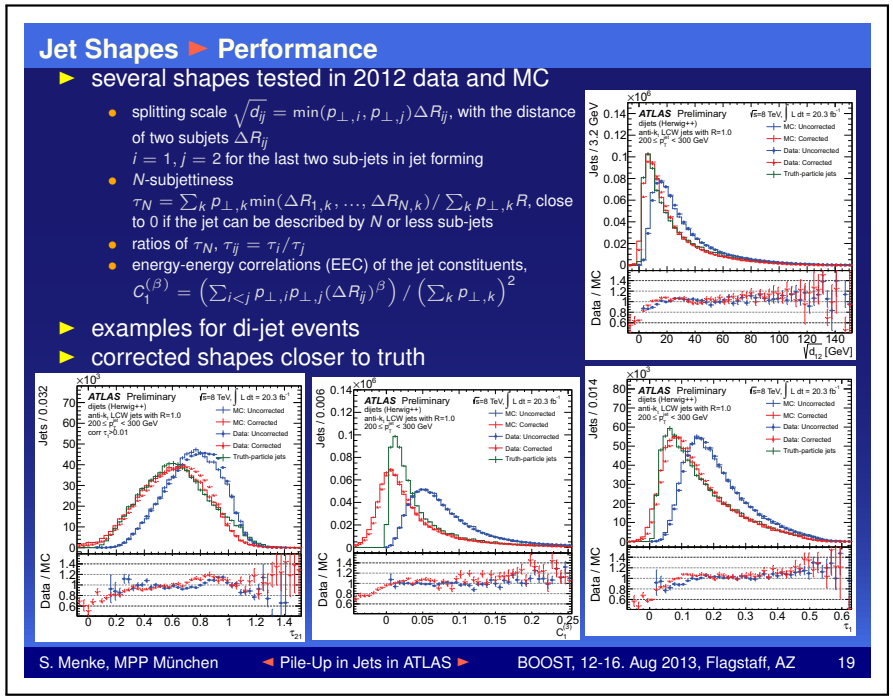
– Sven Menke

- Significant effort to address pile-up issue in ATLAS and CMS...**largely with great success!**
- Some departures from naive exceptions persist, but able to be addressed by experimental techniques derived from the data

Pile-up corrections for jet shapes



– Phil Harris



– Sven Menke

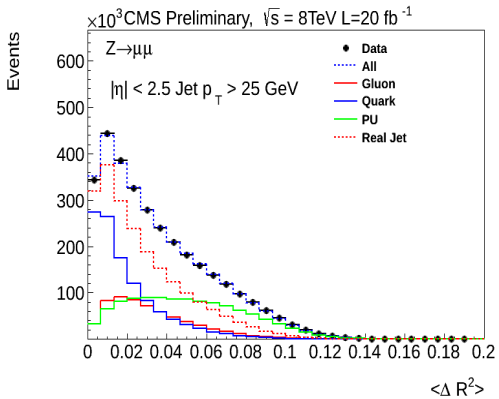
- Particle flow reconstruction **coupled with charged hadron subtraction performs extremely well in CMS and largely addresses the first order problem** on its own
- Dedicated **shape subtraction corrections** in ATLAS demonstrate the potential to address these issues without folding in tracks (yet) and **potentially easing the task of unfolding to particle level**

Using Substructure To Identify Pile-up Jets

08/13/13 Philip Harris BOOST 13

Pileup Jet Id Algorithm:Shapes

- 13 variables for the full discrimination
  - QG disc: *pTD*,#Charged particles,#Neutral particles
  - 6 Shape variables



Most Effective Jet shape variable

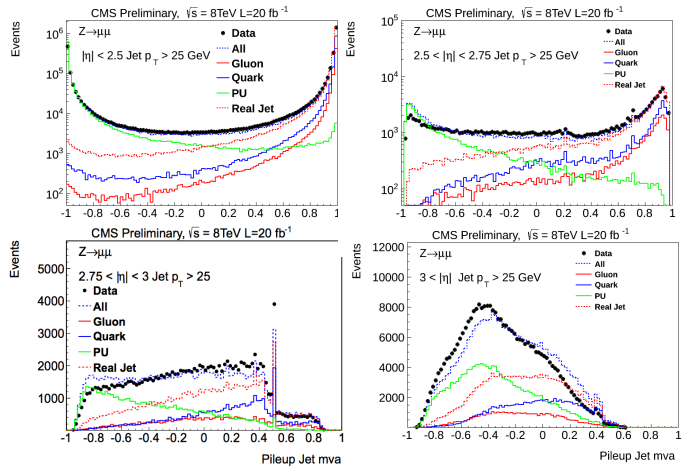
$$\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R_i^2 p_{Ti}^2}{\sum_i p_{Ti}^2}$$

– Phil Harris

08/13/13 Philip Harris BOOST 16

Pileup Jet Id in Data

- Fraction of pileup grows with higher |η|



– Phil Harris

- **Tracking agnostic pile-up jet ID in CMS** is an enormous development
- Used already in VBF analysis in CMS and presumably to more soon (or now)
- See some **small disagreements in forward region**

Outline

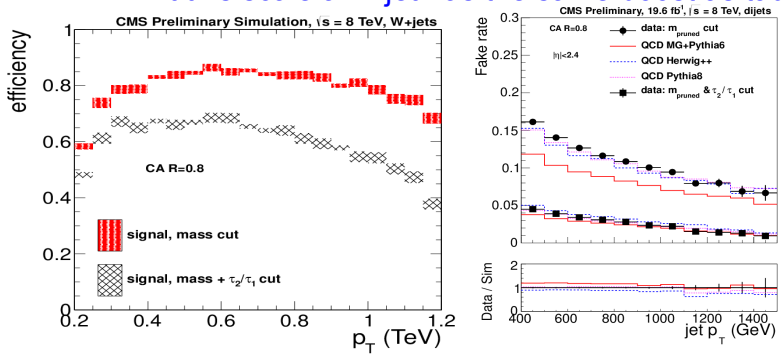
Very high p_T jet substructure

- See indications that at **very very high** p_T some of our current techniques, with their current implementations, may begin to suffer in important ways
- Several ideas 'out there' for handling this:
  - ECAL-only reco for better angular resolution of substructure
  - More detailed track-based measurements for improved intrinsic resolution
- **Would be nice to see some of this at BOOST 2014!**

08/13/13 Philip Harris BOOST 33

Jet Substructure at high *pT*

- At high *pT* (1.2 TeV) substructure starts to break
  - Things merge into one another
    - Detector cannot resolve the subjets (even w/tracking)
  - Shape variables are difficult to separate
  - Such high *pT* may call for alternative approaches
    - At this scale a W jet has the same boost as tau



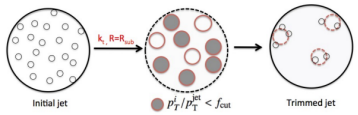
– Phil Harris

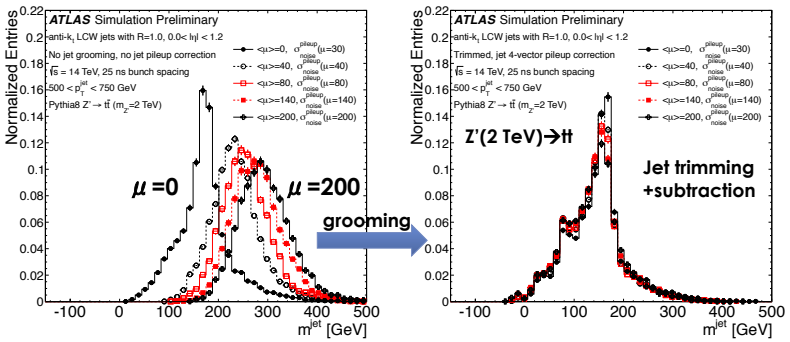
Very high luminosity

Jet substructure

- **Test performance of grooming algorithms using 2012 based optimization:**
  - Trimming R_{kt}=0.3, f=5%

Jet trimming





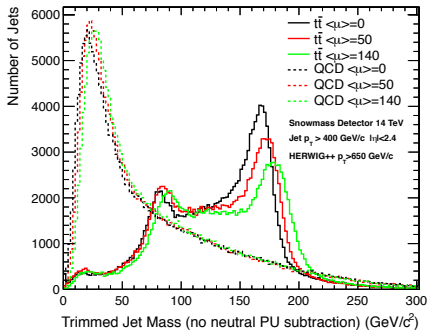
- **Trimming with 2012 parameter optimization works at μ =200**
  - Jet mass distribution stable with μ up to very high luminosity

20

– Ariel Schwartzman

Grooming conclusion

- Trimmed (CA R=0.2 p_T^{frac}=0.05) chosen for Snowmass Ntuples
- Jet mass more stable with increasing pileup
- Small jet area → need much smaller PU correction (in this case none is applied)



James Dolen BOOST 2013 22

– Jim Dolan

- **We will be facing pile-up similar to this before too long**
- **Extremely important to make predictions, and even more important to make accurate predictions**
- **General features sometimes borne-out in toy models, but detailed, quantitative studies often disagree in important ways**

Summary and conclusions

- A *huge* amount of work from the experiments has been shown, yet again this year
- It seems that the level of precision reached with these new, complex, and almost completely physics-driven techniques has essentially reached that of tools that have been in use for decades.
- ATLAS and CMS are approaching the calibration, validation, and exploitation of substructure and boosted objects in a prolific, yet careful and methodical manner.
- We have learned an enormous amount from our theoretical collegues, and the richness of the results we now have in hand should only bolster that collaboration.

A huge THANK YOU is due to our hosts, Peter Loch, the University of Arizona, and the BOOST 2013 Committee

This has been yet another successful BOOST workshop, and there will be many more to come!

Outline

Additional Material