Jet substructure and superstructure for new physics searches

Andrew Lowe

16 April 2013

Overview

- Thank you for giving me the opportunity to be here and speak to you today
- Will try to avoid being too technical
- Will start from the basics, at the top, and "drill down" to the specifics of my work
- I will talk about ATLAS, which is the experiment that I've been involved with for more than 10 years, but much of what I'll describe is equally applicable to other experiments at the LHC
- The techniques I've worked on have the potential for significant impact on many new physics searches at the LHC
- Not possible to cover everything that I have looked at, or all possible physics scenarios
- Picked the search for a Higgs boson in decays to *b*-quarks ($H \rightarrow b\overline{b}$) as a benchmark physics process to demonstrate the utility of this work for new physics searches in general
 - I also looked at Hidden Valley Higgs ($H \rightarrow \pi_v \pi_v \rightarrow b\overline{b}b\overline{b}$) and decays of colour octet scalars (S_8) and excited bosons (W^*) to dijets and found several discriminants that would be useful for these searches, but not shown here today

The structure of matter

- Particle physics is the study of subatomic particles and the fundamental forces that act between them
- Present-day particle physics research represents man's most ambitious and organised effort to answer the question: "What is the universe made of?"
- The current theoretical description is known as the Standard Model (SM)
- Developed throughout the mid to late 20th century
- Extremely successful model
 - We have now observed all the particles predicted by the model, including the Higgs boson
 - Explains a wide variety of experimental results
- But for various reasons we know that it is not a complete theory
 - No gravity, no viable dark matter particle, doesn't explain neutrino oscillations or matter/antimatter asymmetry of universe

The Standard Model



- Building blocks of elementary particles is shown left
- Particles of matter: quarks and leptons
 - Fermions (half-integer spin)
 - For every matter particle there is a corresponding antimatter particle with the same mass but opposite charge
- Three families/generations
 - Why 3? We don't know!
- Three kinds of force carriers
 - Bosons (integer spin)
 - Photon: electromagnetism
 - W and Z bosons: weak nuclear force
 - Gluons: strong nuclear force
- The Higgs boson is responsible for giving mass to some fundamental particles

Quarks and leptons

Leptons:

- The muon and tau can be thought of as heavier versions of the electron, which is the most familiar example of a lepton
- Neutrinos are extremely weakly interacting particles that travel essentially at the speed of light

Quarks:

- Free quarks not observed in nature
- Always bound together in groups of three, called *baryons*, or in quark-antiquark pairs, called *mesons*
 - Examples of baryons are the ubiquitous proton and neutron of ordinary matter
- Baryons and mesons collectively known as *hadrons*
- Quarks possess an attribute whimsically called colour
 - Three quark colours: red, green, blue
 - Antiquarks carry complementary colours: anti-red, anti-green, anti-blue
- The strong force is mediated by gluons, which couple to colour but also carry colour
 - They interact not only with quarks, but also with each other
 - Strength of the force between quarks increases with the distance between them
- All quark bound states are colourless

Hadronisation and jets

- Hadronisation is the process of formulation of hadrons out of quarks and gluons
 - Quarks and gluons are collectively known as partons
- Dynamics of the part of the SM that describes the strong nuclear force quantum chromodynamics (QCD) – is not yet fully understood
- Modelled and parameterised phenomenologically
- During hadronisation, quarks and gluons combine with quarks and antiquarks that spontaneously pop out of the vacuum to form hadrons
- This results in spray of hadrons contained in a narrow cone that is called a jet
 - Exact definition of what constitutes a jet in the detector depends a lot on the specific algorithm used to build jets from the hadrons that are detected
 - Jet finding is the *approximate* attempt to reverse-engineer the process of hadronisation
 - Several different approaches and algorithms exist
 - Cone algorithms (e.g. SISCone)
 - Sequential clustering algorithms (e.g. kT, Anti-kT, Cambridge/Aachen)

Conceptual picture



The Higgs boson

- Observe very large differences in mass between quark and lepton generations
- Photon is massless, but W and Z are massive
 - What is the source of this disparity?
- The Higgs field is responsible for imparting mass to the fundamental fermions and the W and Z
- In the same way that photons are quanta of the EM field, Higgs bosons are quanta of the Higgs field
 - Electrically neutral and *scalar* (that is, spin-0)
- Higgs boson completes the set of particles predicted by the SM
- So important to fundamental physics that a 40 year search for the Higgs boson was launched, culminating in the construction of the Large Hadron Collider (LHC)

The Large Hadron Collider



- World's largest and highest energy particle accelerator
- Built by CERN over a 10-year period at a cost of ~\$10billion
- 27km in circumference, buried 50-175m underground
- Collides protons and heavy ions (pp, p-Pb, Pb-Pb)
- Design beam energy = 7 TeV
 - Peak beam energy 3.5 TeV so far
- Four large experiments at beam intersection points:
 - ALICE
 - ATLAS
 - CMS
 - LHCb

The ATLAS detector



ATLAS consists of several highly granular and hermetic concentric subdetector systems oriented coaxially with respect to the LHC beamline and centered around the nominal proton-proton collision point

Search for a Higgs boson in decays to *b*-quarks



- Higgs boson couples to mass
- Strength of coupling proportional to the mass of the particle with which it interacts
- Higgs will decay with highest probability into particle-antiparticle pair with the highest mass that is kinematically possible
 - No direct couplings to photons or gluons
- For a Higgs boson mass of ~125 GeV, it will decay with highest probability into a $b\overline{b}$ pair, as shown left
- b-quarks hadronise to form b-jets
- *B*-hadrons have lifetime and fly a bit before decaying, resulting in a decay vertex that is displaced from the collision vertex \rightarrow *b*-tag

Motivation for present work

- Higgs $\rightarrow b\overline{b}$ widely considered a poor search channel because of overwhelming backgrounds
- Observation of a signal is very difficult
- At hadron colliders like the LHC, QCD jet production is the dominant high transverse momentum process
- Physics processes of interest usually have cross-sections that are many orders of magnitude smaller
- Consequently, QCD jet production is the main background for many new physics searches, predictions for which often feature jets in the final state as a signature
- The ability to distinguish jets in interesting new physics events from those in QCD background events would, therefore, be extremely valuable and likely to have a significant impact on the search for new physics

Jet substructure and superstructure

- Report on work by myself on the use of hadronic event shapes and jet substructure
- Inspired by recent flurry of activity in this field
- Availability of fine-grained calorimetry in ATLAS allows for jets to be studied in much greater detail than was previously possible
 - O(10 100) calorimeter cells per typical QCD jet, ~10 times CDF or D0
- Can measure more than just bulk properties of jets?
- Have investigated the potential benefit of hundreds of different event shapes and jet shapes for both Higgs and exotics offline analyses
 - Approach was to cover as much territory as possible
 - Some variables have theoretical motivation, others don't

Datasets and colour key for plots

• Signal:

- WH120Inubb (PYTHIA)
- Boosted WH120Inubb (HERWIG)
- Background:
 - Wbb (ALPGEN)
 - QCD multijets
 - bb/cc and require one of the b or c to decay to e/ μ with pT > 15 GeV and $|\eta| < 2.5$ (PYTHIA)
 - Also: tt (MC@NLO)
- 7 TeV samples with pile-up and realistic detector conditions
 - Pile-up: average 23 pp interactions superposed on each collision event at LHC design luminosity
- All plots normalised to unit area

Feynman diagrams for signal and *Wbb* background



Jet reconstruction

- Full detector simulation
- Jet reconstruction algorithms: Anti- k_{τ} (4, 6, 10) and Cambridge/Aachen (12)
 - Mostly showing just Anti- k_{T} 4 today; don't need fat jets for most variables (some improvements with fatter jets)
- Only calorimeter information for discriminants studied, no tracking used
 - Jet constituents = topological clusters
- $p_{_{\rm T}}$ > 25 GeV, $|\eta|$ < 2.5

Colour flow



- Jets are not totally independent objects; they are *colour-connected*
- Term comes from the way SU(3) group indices are contracted in amplitudes
- Jets initiated by colour-singlets are colour-connected with each other; those by quarks or gluons, to the proton remnants that travel down the beam pipe
- Provides the event with an observable and characterisable *superstructure*
- Such information, if observable, would be complementary to information in event kinematics and may temper otherwise irreducible backgrounds
- Colour connections for $H \rightarrow bb$ (left) and $g \rightarrow bb$ (right)

Radiation patterns for events with different colour configurations but identical kinematics



- Average showered p_{τ} density in (y, ϕ) for an ensemble of events with fixed parton-level kinematics
 - Figures taken from PRL 105, 022001 (2010) by Gallicchio and Schwartz
- * $H \rightarrow bb$ (left) and $g \rightarrow bb$ (right)
- For a colour-singlet, the radiation is mostly found in the region between the two jets
- For the background-like colour configuration, the radiation is pulled towards the beam
- Difference is independent of event kinematics and, if observable, would enable extra discrimination that is orthogonal to cuts on kinematical variables

New colour-connection discriminant



- Beam colour flow = sum of E_{T} -weighted squares of distance ϕ_i of jet constituents from jet
- Inspired by jet dipolarity observable introduced in arXiv:1102.1012 by Hook, Jankowiak, Wacker, but my own invention
- Complementary:
 - Jet dipolarity measures degree of colour-connectivity between jets
 - BCF measures degree of colour-connectivity of jet with proton remnants in beam

New discriminant in pictures



- Can work either as a jet shape or an event shape
 - Event shape = sum of values of discriminant for leading and next-to-leading jet
- Sensitive to jet size; works better with larger size jets
 - Larger jet radius captures more of the differences between radiation patterns
- Expected features:
 - Signal: less radiation between jets and beam → lower weighting to jet constituents in this region → small values of discriminant
 - Background: more radiation between jets and beam → higher weighting to jet constituents in this region → large values of discriminant

Results (Anti- k_{T} 10 jets)

Seems to offer some discrimination



Width radiation variable

$$g = \sum_{i \in J} \frac{p_T^i |r_i|}{p_T^J}$$

- Distribution of particles within a jet can be useful for distinguishing jets initiated by different flavours of quark or by a gluon
- Width: sum of $p_{\rm T}$ -weighted radial distances of each jet constituent from the jet axis
 - Especially interesting because it may be possible to calculate in QCD
- Studies of width already done in ATLAS

ATLAS jet width distributions



- Distribution of jet width for isolated Anti-k_τ jets with R=0.6 tagged as light quark-jets and gluon-jets in the MC simulation with the PYTHIA AMBT1 tune and full-detector simulation
- Plot from ATLAS-CONF-2011-053

Jet width for q/g tagging (Anti- k_{T} 6 jets)



obtained by ATLAS and shown on previous slide

Width of next-to-leading jet

- Introduced in Multivariate discrimination and the Higgs + W/Z search, JHEP 1104:069 (2011), in which the width ("girth") of the next-toleading jet was identified as important.
- Expect strong dependence on pile-up
- Note that, exceptionally, results here are shown for "fat" Anti-kT10 jets
 - Anti-kT6 still OK, but discrimination power appears to decrease with narrower jets

Distribution of width of next-to-leading jet



- Width of the next-to-leading tagged b-jet for ZH signal (solid blue) and Zbb background (hashed red) at the LHC
- Hard-parton level + smearing
- Events satisfy selection cuts and Higgs mass-window cut (details in PRL paper)
- Plot from updated version of PRL paper that will soon replace version on arXiv; received from authors via private communication

Width of next-to-leading jet for Anti- k_{T} 10 jets



Seems to offer reasonable discrimination, especially wrt Wbb

Width of next-to-leading jet and pile-up



As expected, some dependence on pile-up. Strength of dependence has been observed to increase with jet size. Reminder of colour key: WH120Inubb, Boosted WH120Inubb, Wbb, QCD multijets, ttbar



 $|y^*| = |y_1 - y_2|/2$

Already used in Exotic Dijet offline analyses

- For example: ATLAS-CONF-2012-088 (Search for New Phenomena in the Dijet Mass Distribution using 5.8 fb-1 of pp Collisions at $\sqrt{s} = 8$ TeV collected by the ATLAS Detector)
- Require events with $|y^*| < 0.6$ to optimise analysis sensitivity
- Used in dijet trigger in 2012 menu to provide an unprescaled trigger that is fully efficient for the phase space region $m_{ij} > 2$ TeV, $|y^*| < 1.7$

$|y^*|$ for Anti- k_T 4 jets



Suggests that a cut on $|y^*|$ (~0.6) could enhance signal

y* and pile-up



Little or no dependence on pile-up

Fox-Wolfram moment H_2/H_0

$$H_l = \sum_{i,j} \frac{|\mathbf{p}_i| |\mathbf{p}_j|}{E_{\text{vis}}^2} P_l(\cos \theta_{ij}) ,$$

- θ_{ij} is the opening angle between constituents *i* and *j*, E_{vis} is the total visible energy of the event, and $P_{I}(x)$ are the Legendre polynomials; autocorrelations (*i* = *j*) are included in the sum
- The ratio of the second and zeroth moments is sometimes called R₂, and was used in the past in the context of e+e- colliders (LEP, BaBar...)
- Indicative of the collimation of the event topology; values closer to zero indicate a more spherical event

Fox-Wolfram moment H_2/H_0 for Anti- k_T 4 jets



Some weak discrimination power here

Fox-Wolfram moment H_2/H_0 and pile-up



Very little dependence on pile-up

 H_{T}'

- $H_{\tau}' = H_{\tau} (E_{\tau_1} + E_{\tau_2})$, where H_{τ} is the total visible transverse energy in the event
- Introduced in Measurement of the ttbar cross section using high-multiplicity jet events, D0 Collaboration, Phys.Rev D82, 032002, (2010)

H_{T}' for Anti- $k_{T}4$ jets



Suggests that discriminant might be useful for reducing ttbar background Have observed that narrow jets work better than fatter jets Reminder of colour key: WH120Inubb, Boosted WH120Inubb, Wbb, QCD multijets, ttbar

H_{τ} for different sized jets

Anti- k_{T} 6

Anti-*k*,10



Discrimination power degraded as jets get fatter

H_{T}' and pile-up



Significant dependence on pile-up

Mass democracy and betaflow

- Introduced in Unburied Higgs, arXiv:1006.1650 [hep-ph]
 - Designed for Higgs search when dominant decay is to gluons via light uncoloured resonances (thus, it is "buried"); $H \rightarrow aa \rightarrow 4g$
- "Mass democracy": $\alpha = \min[m_1/m_2, m_2/m_1]$
 - Signal events are distinguished by the symmetry of their decay products; for the background there is no reason for QCD radiation to produce democratic jets
- "Betaflow"*: $\beta = E_{T_3}/(E_{T_1} + E_{T_2})$
 - Sensitive to differences in colour structure
 - For signal events, colour is seen late in the Higgs decay process
 - Background jets are initiated by hard coloured particles that are colour-connected to the rest of the event; moreover, there is more phase space for QCD radiation
 - Expect typical values to be larger for QCD background
- * My name for this variable, to distinguish it from umpteen other b's; not sure if authors of the paper have a name for this

Mass democracy for Anti- k_{T} 4 jets



Some discrimination power here

Mass democracy for different sized jets

Anti-*k*₁**6**

Anti-*k*,10



Fatter jets may well work better

Mass democracy and pile-up



Very little dependence on pile-up

Betaflow β result for Anti- k_{T} 4 jets



Discrimination power somewhat weak, but appears to be still there

Betaflow β for different sized jets

Anti-*k*₁**6**

Anti-*k*,10



Discrimination power better with fatter jets

Betaflow β and pile-up



Little or no dependence on pile-up

$\Delta \phi$ versus $\Delta \eta$ Anti- k_{T} 4 jets



Analysis of $\Delta \eta / \Delta \phi$ scatter plot

- For low Higgs p_T , $p_T < m_H$, the signal lives in bands clustered around $|\Delta \phi| \sim \pi$
- For $p_T \ge m_H$, the signal lives in a ring of $\Delta R \sim 2m_H/p_T$
- Purely a consequence of spherically-symmetric Higgs decay boosted transverse to the beam
- In contrast, for background, there is a bias for one of the jets to have large rapidity

Twist angle: $\tau = \tan^{-1} (\Delta \phi / \Delta y)$





- Twist introduced in arXiv:1010.3698v2 [hep-ph], Multivariate discrimination and the Higgs + W/Z search
- Twist is a longitudinal-boost-invariant version of the rotation of the H/b/b plane with respect to the beam/H plane
- For sake of clarity, the case shown above has no longitudinal boost
- If the $(\Delta \eta, \Delta \phi)$ plane is interpreted as 2D Cartesian coordinates, the polar coordinate combinations are ΔR and τ
- Expect signal to prefer $\tau = \pi/2$ and background to prefer $\tau = 0$

Expected behaviour of Twist



- Blue: ZH signal
- Red: Zbb background
- Plot from hep-ph paper
- MADGRAPH hard partons
- Showered with PYTHIA
- Jets reconstructed with FASTJET
- Detector-level cuts (no full detector simulation, no smearing)
- Ideal situation with no pile-up

Twist result for Anti- k_{T} 4 jets



Distributions very close to expected behaviour

Helicity angle in $b\overline{b}$ CoM frame (θ_{h})



From arXiv:1010.3698v2 [hep-ph], *Multivariate discrimination and the Higgs* + *W/Z search* Reminder: blue = ZH signal, red = Zbb background

Cosine of helicity angle of leading jet for Anti- k_{T} 4 jets



Distributions show expected trend in general (Biased toward positive values because the higher p_T jet in each pair is chosen) Reminder of colour key: WH120Inubb, Boosted WH120Inubb, Wbb, QCD multijets, ttbar

Acoplanarity, $\phi_{aco} = |\pi - |\Delta \phi|| + |\pi - \Sigma \theta|$ for Anti- k_{τ} 4 jets



Note: during construction, *all* angles are wrapped into $-\pi$ to $+\pi$

Response to light-quark/gluon jets

- Jets labelled as originating from a gluon have a different response to those labelled as originating from a light quark
- Different observable final state properties
 - Studied at LEP
- Quarks and gluons have different colour charges
- Jets labelled as gluon jets tend to have more particles, and those particles tend to be softer than in the case of lightquark jets
- The harder particles in light-quark jets tend to penetrate further into the calorimeter
- The magnetic field of the Inner Detector amplifies the broadness of gluon jets

Jet labelling procedure

- Jets labelled using the partons in the generator event record
- Scheme 1: The highest energy parton that points to the jet (i.e. with $\Delta R < 0.4$, 0.6, 1.0, and 1.2 for jets with R = 0.4, 0.6, 1.0 and 1.2 respectively) determines the flavour of the jet
 - Identical to that used in ATLAS-CONF-2011-053, Light-quark and Gluon Jets in ATLAS: Calorimeter Response, Jet Energy Scale Systematics, and Sample Characterization
- Scheme 2: A jet is labelled as a *b*-jet if a *b* quark with $p_T > 5$ GeV is found in a cone of size $\Delta R = 0.3$ around the jet direction. The various labelling hypotheses are tried in this order: *b* quark, *c* quark and τ lepton. When no heavy flavour quark nor τ lepton satisfies these requirements, the jet is labelled as a light-jet. No attempt is made to distinguish between *u*, *d*, *s* quarks and gluon.
 - Identical to that used for measuring *b*-tagging performance in ATLAS CSC Note and elsewhere (standard analysis tool available to do this)
- Hybrid: Use second scheme to find b-, c-, and τ-jets, then use first scheme to sort remaining light jets into light-quark jets and gluon jets
 - This is what I use when showing differences in response for *b*-, *c*-, *τ*-, light-quark and gluon jets, because I want to show utility of discriminants for both *b*/*c*/*τ*-tagging and q/g-tagging
- These labelling procedures are not unambiguous and are not *strictly identical* for different Monte Carlo generators
- Definitions are not theoretically sound, but studies have shown that for most generators and for isolated jets, truth labelling is identical to matrix-element-based labelling for >95% of jets

Flavour tagging discriminant (Anti- $k_{T}10$ jets)



Number of jet constituents



Possible future work

Improve method

- Look for new discriminants
- Look for ways to ameliorate the effects of pile-up
- May get better performance from jets built from tracks rather than calorimeter clusters

Monte Carlo and real data comparison

 Measurements of the discriminants obtained with real data over a range of kinematic and topological regions would be compared with equivalent measurements obtained with a variety of Monte Carlo (MC) event generators to assess how well the data are modelled

Validation study with a known physics process in real data

- Apply the most promising discriminants to improve an analysis with a known physics process using real data
- To employ these kinds of discriminants in a search for new physics it will be essential to:
 - Develop methods to estimate the efficiency and fake rate from LHC data
 - Optimise selection cuts for efficiency and purity for signal events
 - Determine the significance gain obtained

New physics search

 Would be done in the context of an existing analysis by employing these discriminants to extend the amount of information available from each jet, allowing a tighter selection and thereby improving the sensitivity to new physics

Summary

- These are very exciting times for high-energy physics
- The LHC provides an unprecedented opportunity to explore a new energy frontier where new physics might be discovered
- The next few years are likely to see important discoveries made at the LHC
- These will answer old questions and probably raise new ones that will dominate the direction of the field for years to come
- This work has the potential to make a considerable impact on many new physics searches
- Represents a small but important part of a jigsaw puzzle
- Will address an important problem whose solution will bring us one step closer to realising a long-term dream



Bonus slides...

Number of Anti-*k*₁4 jets



$\theta_1 - \theta_2 / \theta_1 + \theta_2$ (Anti-*k*T4 jets)



Betaflow β for q/g tagging



Particle detection in ATLAS



Particle physics in popular culture 😳

Nyugati railway station, August 2012



Anything here look familiar? \rightarrow



Simulated Higgs boson event in the ATLAS detector



About the speaker (career timeline)

