Jet SIFT-ing

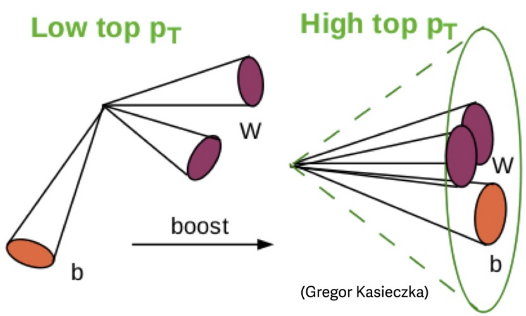
A Scale-Invariant Jet Clustering Algorithm for the Substructure Era

Joel W. Walker
Sam Houston State University

with
Andrew Larkoski (UCLA), Denis Rathjens (CMS), and Jason Veatch (ATLAS)
arXiv coming soon …

ML4Jets
Rutgers University
November 1-4, 2022

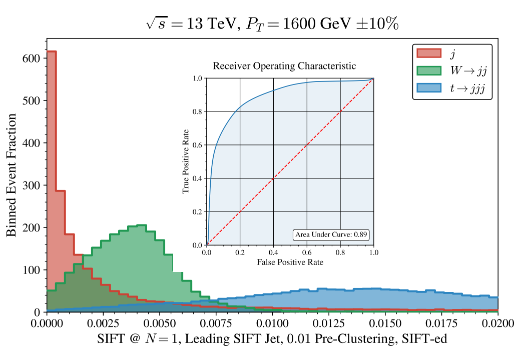
SIFT: Scale-Invariant Filtered Tree

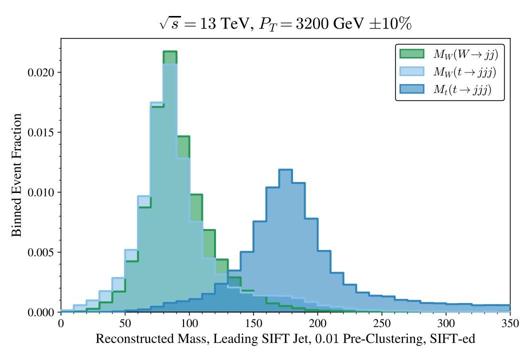


- Massive resonances decay into hard prongs
- Jet definitions with fixed cones impose a scale
- Boosted objects collimate and structure is lost
- Substructure recovery techniques are complex
- Can we avoid losing resolution in the first place?
- Select proximal objects w/ scale-invariant measure

- Candidate pairs are merged, dropped, or isolated, according to criteria integrated into the SI measure
- SIFT unifies: a) large-radius jet finding, b) filtering of soft wide radiation, and c) substructure axis finding into a single-pass prescription for low/high boosts

$$\delta_{AB} \equiv \frac{\Delta M_{AB}^2}{E_{\mathrm{T}A}^2 + E_{\mathrm{T}B}^2}$$

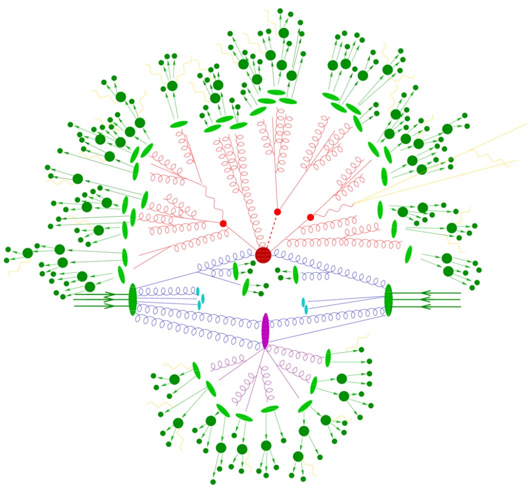




- *N*-subjet Tree holds superposition of projections onto *N*=1,2,3 prongs
- Hard prongs are preserved to end
- The measure history discriminates *N*=1,2,3 typically above 90% AUC
- Faithful kinematic reconstruction

Standard kT Jet Clustering Algorithms

- Debris from showering & hadronization must be reassembled in a manner that preserves correlation with the underlying hard (partonic) event
- 3 related algorithms reference an input angular width R_0 & differ by an index n
- Objects wider than R_0 will never be clustered; Objects inside cone always merge
- $n = 0$, or “Cambridge/Aachen” favors objects with high angular adjacency
- $n = +1$, or “kT” additionally favors clustering where one of the pair is soft
- $n = -1$, or “Anti-kT” prioritizes clustering where one of the pair is hard
- Anti-kT is now the default jet clustering tool at LHC, with $R_0 \sim 0.5$
- It is robust against “soft” and “collinear” jet perturbations and has regular jet shapes which are favorable for calibration against pileup, etc.



$$\delta_{AB} \equiv \min\left[P_{\mathrm{T}A}^{2n}, P_{\mathrm{T}B}^{2n}\right] \times \left(\frac{\Delta R}{R_0}\right)^2$$

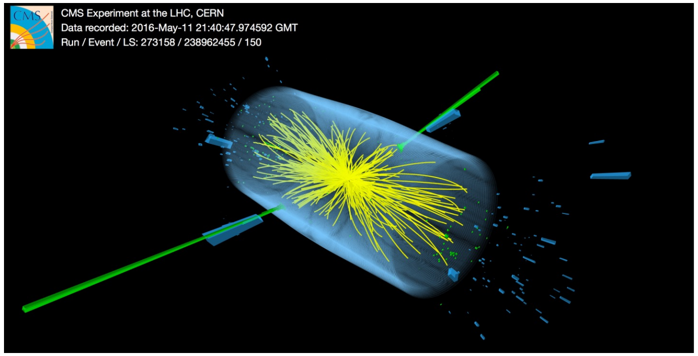
Image: Stefan Höche

A Scale-Invariant Distance Measure

- It is worth asking whether alternative techniques could provide intrinsic resiliency to boosted event structure; this requires dropping the input scale R_0
- It would be good to "asymptotically" recover key behaviors of Anti-kT
- Numerator should favor angular collimation; we propose ΔM^2, similar to JADE
- Denominator should suppress soft pairings; we propose ΣE_T^2, similar to Geneva
- Result is dimensionless, Lorentz invariant (longitudinally in the denominator), and free from references to external / arbitrary scales

$$\delta_{AB} \equiv \frac{\Delta M_{AB}^2}{E_{\mathrm{T}A}^2 + E_{\mathrm{T}B}^2}$$

$$\Delta m_{AB}^2 \equiv (p_A^\mu + p_B^\mu)^2 - m_A^2 - m_B^2 \; = 2 p_A^\mu p_\mu^B \simeq 2E^A E^B \times (1 - \cos \Delta\theta_{AB}) \simeq E^A E^B \Delta\,\theta_{AB}^2$$



$$E_{\mathrm{T}} \;\equiv\; \sqrt{M^2 + \vec{P}_{\mathrm{T}} \cdot \vec{P}_{\mathrm{T}}} \;=\; \sqrt{E^2 - P_z^2}$$

$$\lim_{M=0} \;\Rightarrow\; |\vec{P}_{\mathrm{T}}|$$

Image: CMS

Comparison to the Geneva Measure

$$y_{ij} = \frac{8}{9} \frac{E_i E_j (1 - \cos\theta_{ij})}{(E_i + E_j)^2}$$

- Though motivated for new reasons, our measure is similar to "Geneva"
- In addition to normalization, there are three primary differences:
  - Sum of squares rather than square of sum (minor change)
  - Transverse cylindrical coordinates are referenced, as suitable for hadron collider rather than electron collider applications (relevant change)
  - Mass of merger candidates is accounted for (significant change)
- The more novel updates are not to the measure, but relate instead to:
  - Filtering of stray radiation and a related halting criterion
  - The concept of an *N*-subject Tree (superposition of axis candidates)

Moving Toward a Geometric Measure

- An efficient algorithm needs something like a "GEOMETRIC" neighbor finding
- We need to refer to the collider coordinates of A & B directly ($\Delta\eta_{AB}, \Delta\phi_{AB}$, etc.)
- For massive A & B, it will actually be rapidity Δy_{AB} that is relevant
- Boost from the $P_z = 0$ frame into the lab:

$$\Delta m^2_{AB} = 2 \times \left(E^A E^B - p_z^A p_z^B - p_\mathrm{T}^A p_\mathrm{T}^B \cos \Delta\phi_{AB} \right)$$

$$\begin{pmatrix} E \\ p_z \end{pmatrix} = \begin{pmatrix} \cosh y & \sinh y \\ \sinh y & \cosh y \end{pmatrix} \begin{pmatrix} E_\mathrm{T} \\ 0 \end{pmatrix} = \begin{pmatrix} E_\mathrm{T} \cosh y \\ E_\mathrm{T} \sinh y \end{pmatrix}$$

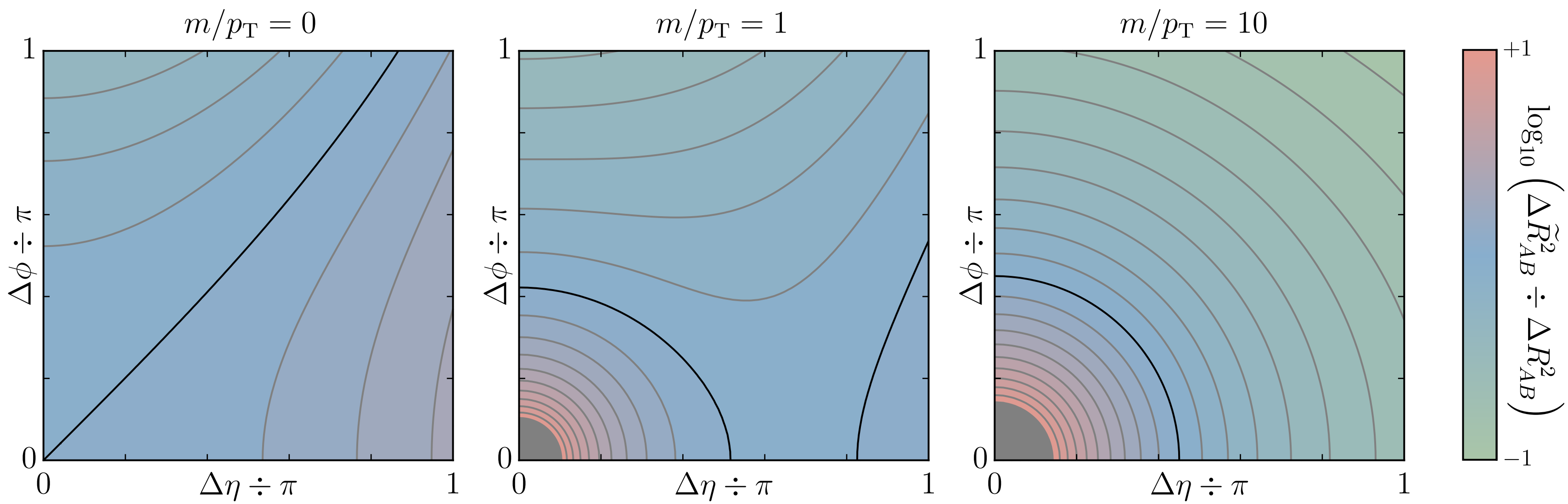
$$\begin{aligned} & E_\mathrm{T}^A E_\mathrm{T}^B - p_z^A p_z^B \\ = & E_\mathrm{T}^A E_\mathrm{T}^B \times \left(\cosh y^A \cosh y^B - \sinh y^A \sinh y^B \right) \\ = & E_\mathrm{T}^A E_\mathrm{T}^B \cosh \Delta y_{AB} \end{aligned}$$

- The difference between E_T & P_T (i.e. MASS) means that we cannot perfectly factorize kinematics from geometrics
- The role of ξ is to deemphasize azimuthal differences in the non-relativistic limit

$$\Delta m^2_{AB} = 2\, E_\mathrm{T}^A E_\mathrm{T}^B \times \left(\cosh \Delta y_{AB} - \xi^A \xi^B \cos \Delta\phi_{AB} \right)$$

$$\xi \equiv \frac{p_\mathrm{T}}{E_\mathrm{T}} = \left(1 - \frac{m^2}{E_\mathrm{T}^2} \right)^{+1/2} = \left(1 + \frac{m^2}{p_\mathrm{T}^2} \right)^{-1/2}$$

Comparative Angular Response



$$\begin{aligned}\Delta\widetilde{R}^2_{AB} &\equiv \frac{\Delta m^2_{AB}}{E_{\mathrm{T}}^A E_{\mathrm{T}}^B} \\ &= 2 \times \left(\cosh \Delta y_{AB} - \xi^A \xi^B \cos \Delta\phi_{AB}\right) \\ &\simeq \Delta\eta^2_{AB} + \Delta\phi^2_{AB} \equiv \Delta R^2_{AB}\end{aligned}$$

$$\Delta\widetilde{R}^2_{AB} \Rightarrow \Delta R^2_{AB} + \left\{1 - \frac{\Delta R^2_{AB}}{2}\right\} \times \left\{\left(\frac{m_A}{p_{\mathrm{T}}^A}\right)^2 + \left(\frac{m_B}{p_{\mathrm{T}}^B}\right)^2\right\} + \cdots$$

- The ΔR^2 measure is recovered for zero mass & small angular separations
- Hyperbolic cosine differs from cosine in that all Taylor terms are POSITIVE … rapidity separations dominate azimuth
- Massive or low-pT objects resist clustering, even at small angles; this is a type of BEAM MEASURE

Geometrizing the Denominator

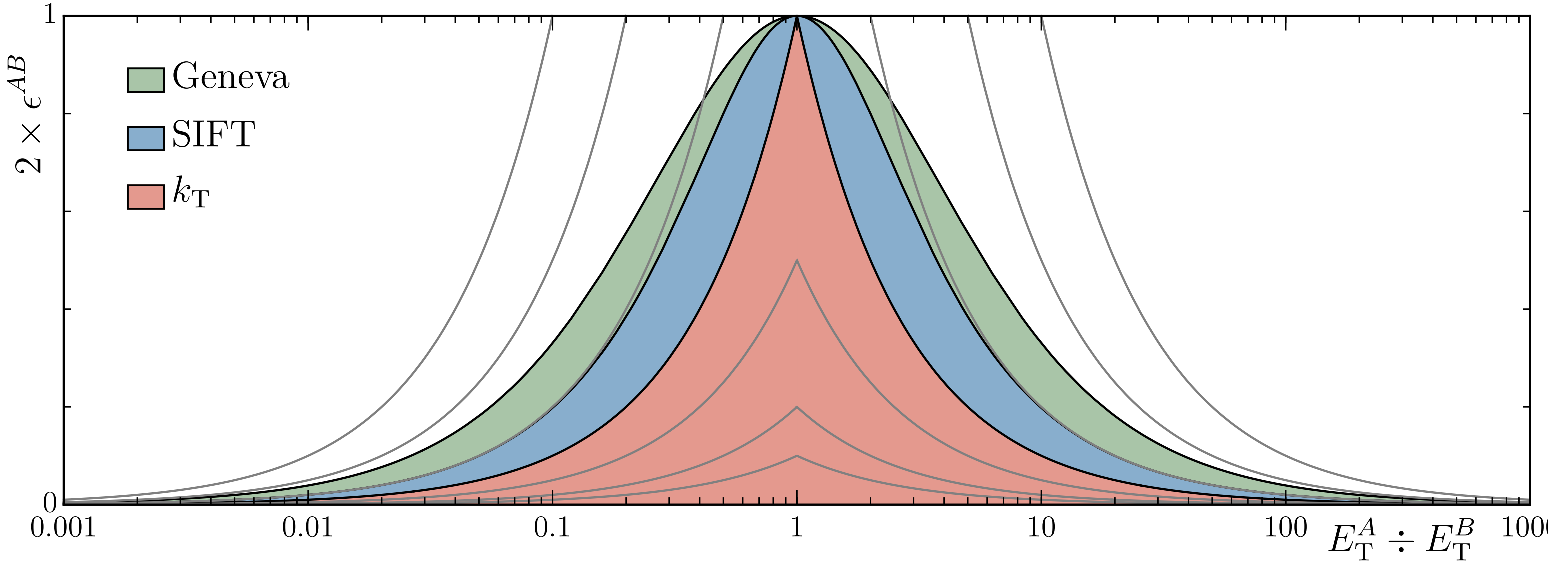
- The remainder of the metric refers only to transverse energy RATIOS
- This factor has a symmetry under $E_T \to 1/E_T$
- It asymptotically mimics BOTH kt and anti-kt clustering, preferencing the clustering of pairs with hierarchically DISPARATE transverse scales
- It has the benefit of being ANALYTIC

$$\epsilon^{AB} \equiv \frac{E_{\mathrm{T}}^{A} E_{\mathrm{T}}^{B}}{(E_{\mathrm{T}}^{A})^2 + (E_{\mathrm{T}}^{B})^2} = \left\{ \left(\frac{E_{\mathrm{T}}^{A}}{E_{\mathrm{T}}^{B}} \right) + \left(\frac{E_{\mathrm{T}}^{B}}{E_{\mathrm{T}}^{A}} \right) \right\}^{-1}$$

$$u \equiv \ln \left(E_{\mathrm{T}} / [\mathrm{GeV}] \right)$$

$$\epsilon^{AB} = \left(e^{+\Delta u_{AB}} + e^{-\Delta u_{AB}} \right)^{-1} = \left(2 \cosh \Delta u_{AB} \right)^{-1}$$

Comparative Energy-Momentum Response



$$\epsilon^{AB} = \left\{ \min\left(\frac{E_{\mathrm{T}}^{A}}{E_{\mathrm{T}}^{B}}\right) + \max\left(\frac{E_{\mathrm{T}}^{A}}{E_{\mathrm{T}}^{B}}\right) \right\}^{-1}$$

$$\simeq \left\{ \max\left(\frac{E_{\mathrm{T}}^{A}}{E_{\mathrm{T}}^{B}}\right) \right\}^{-1}$$

$$= \min\left(\frac{E_{\mathrm{T}}^{A}}{E_{\mathrm{T}}^{B}}\right) \simeq \frac{\min(E_{\mathrm{T}}^{A}, E_{\mathrm{T}}^{B})}{E_{\mathrm{T}}^{A} + E_{\mathrm{T}}^{B}}$$

$$\delta_{AB}^{k_{\mathrm{T}}} \underset{\sim}{\propto} \left(\frac{E_{\mathrm{T}}^{A} E_{\mathrm{T}}^{B}}{E_{0}^{2}}\right)^{n} \times \min\left[\frac{E_{\mathrm{T}}^{A}}{E_{\mathrm{T}}^{B}}, \frac{E_{\mathrm{T}}^{B}}{E_{\mathrm{T}}^{A}}\right]$$

- SIFT & Geneva are scale invariant here
- The kT algorithms SCALE the overall response by a power of the geometric mean of transverse energies
- Grey contours are 0.1, 0.2, 0.5, 2, 5, 10, with reverse ordering for anti-kT

All Together: the SIFT Measure

$$\delta_{AB} = \epsilon^{AB} \times \Delta \widetilde{R}^2_{AB}$$
$$= \frac{\cosh \Delta y_{AB} - \xi^A \xi^B \cos \Delta \phi_{AB}}{\cosh \Delta u_{AB}}$$

- The measure is a simple product of energy and angular-type factors
- Clustering preferences pairs that are (relatively) soft and/or collinear
- Since mutually hard (relative to other available radiation) members will defer clustering, prongy structure is preserved to the end and easily accessed
- However, extraneous wide and soft radiation is assimilated very early
- This distorts the kinematic reconstruction (mass especially)
- Moreover, there is no sense of when to *stop* clustering
- These failures are potentially fatal, precluding practical application!

Lepton to TTbar 2.5 TeV Scale Invariant Clustering with Ghosts

- See Video “A” Posted at Indico

FILTERING Stray Radiation

- We know, at least, how to deal with soft, wide-angle radiation
- Take a cue from “Soft Drop” (2014 Larkoski, Marzani, Soyez, Thaler)
- This “Grooming” removes contaminants like ISR, UE, and pileup
- SD iteratively DECLUSTERS C/A, dropping softer object unless & until:

$$\frac{\min(P_{TA}, P_{TB})}{P_{TA} + P_{TB}} > z_{\text{cut}} \left(\frac{\Delta R_{AB}}{R_0}\right)^{\beta}$$

- Typically, z_{cut} is $\mathcal{O}(0.1)$, and $\beta > 0$ for grooming
- We propose an analog to be applied within the original clustering itself, expressible in the scale invariant language

$$\text{Cluster:} \quad \frac{\Delta \widetilde{R}^2_{AB}}{2} < \left\{ \left(2\,\epsilon^{AB}\right) \leq 1 \right\}$$

- With factors of 2 in their “natural” places the maximal effective cone size is $\sqrt{2}$
- This is a DYNAMIC boundary, and the angular size reduces for imbalanced scales

Dropping vs. Isolating

- This leaves the question of what to do when clustering FAILS …
- There are two distinct ways to fail the filtering criterion, to be handled differently
- The scale disparity can be too extreme (soft radiation) at O(1) angular separation

$$(\epsilon_{AB} \ll 1) \text{ and } (\Delta \widetilde{R}^2_{AB} \simeq 1)$$

- In this case the metric product is small … DROP the softer member
- Or, the angular separation can be too large (wide angle) with comparable scales

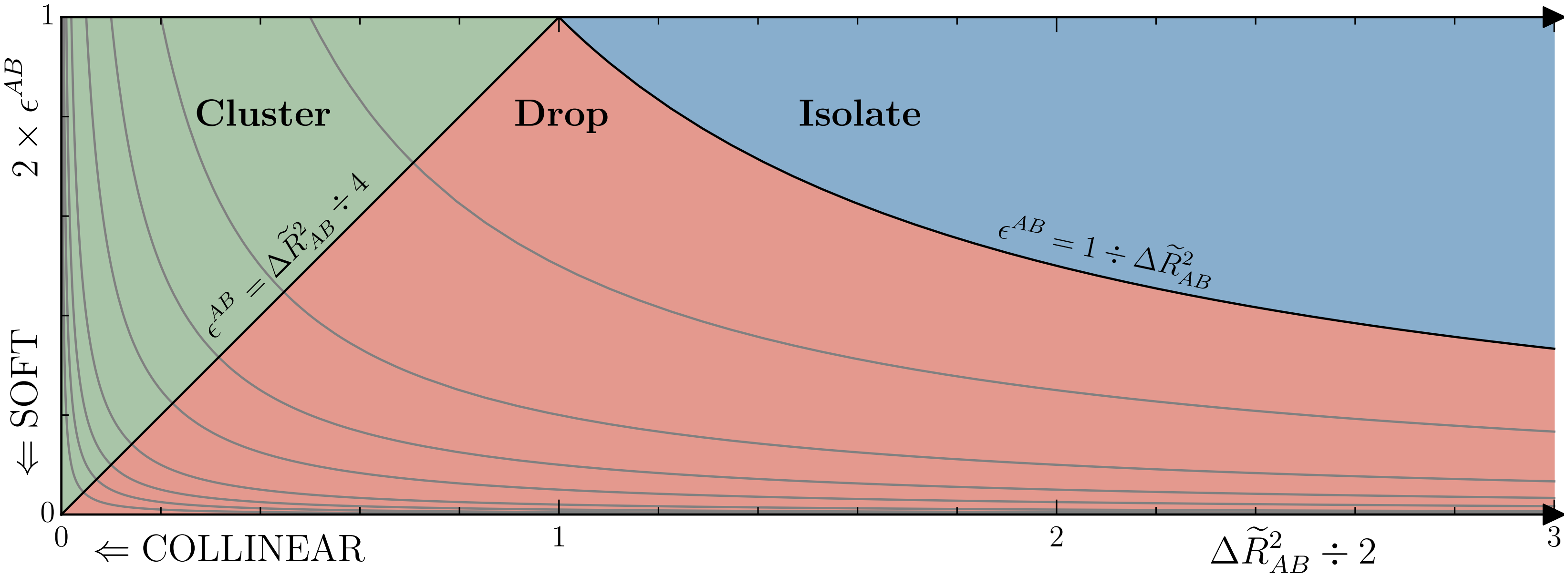
$$(\Delta \widetilde{R}^2_{AB} \gg 1) \text{ and } (\epsilon_{AB} \simeq 1)$$

- In this case the metric product is large … ISOLATE both objects

$$\text{Isolate:} \quad \{1\} \leq \delta_{AB}$$

$$\text{Drop:} \quad \left\{ (2\,\epsilon_{AB})^2 \leq 1 \right\} \leq \delta_{AB} < \{1\}$$

Clustering Phase Diagram



- The unification of clustering, filtering, and isolation also provides natural halting
- Grey contours "$y = \delta/x$" mark constant values of the measure
- Isolation occurs above $\delta = 1$; this amounts finding of variable large-radius jets
- The same factors separate clustering from dropping at "$y = x$"

Lepton to TTbar 2.5 TeV SIFT Filtered Clustering with Ghosts

- See Video “B” Posted at Indico

The *N*-Subjet TREE

- We observe that:
  - hard structures are preserved
  - wide concentrations of hard objects are isolated
  - soft wide radiation is dropped

- However, hard prongs within a variable radius jet do still cluster
- How do we fix the interior halting criterion to avoid losing structure?
- The most interesting alternative is to not halt at all …
- We learn more about whether the prongs “want” to merge by merging!
- Hard prongs are the final objects to be merged, and we retain a superposition of projections onto all numbers *N* of prongs – suitable for computing N-subjettiness
- The record of structure is also directly imprinted on the measure history

Tagging Jet Substructure

- N-Subjettiness τ_N is the leading tool for characterizing how well a given event matches an N-prong hypothesis (axes chosen separately)
- The best discrimination comes from the ratio r_N, e.g. how much more 3-prong-like is the event than 2-prong like
- However, this procedure is also substantially complicated

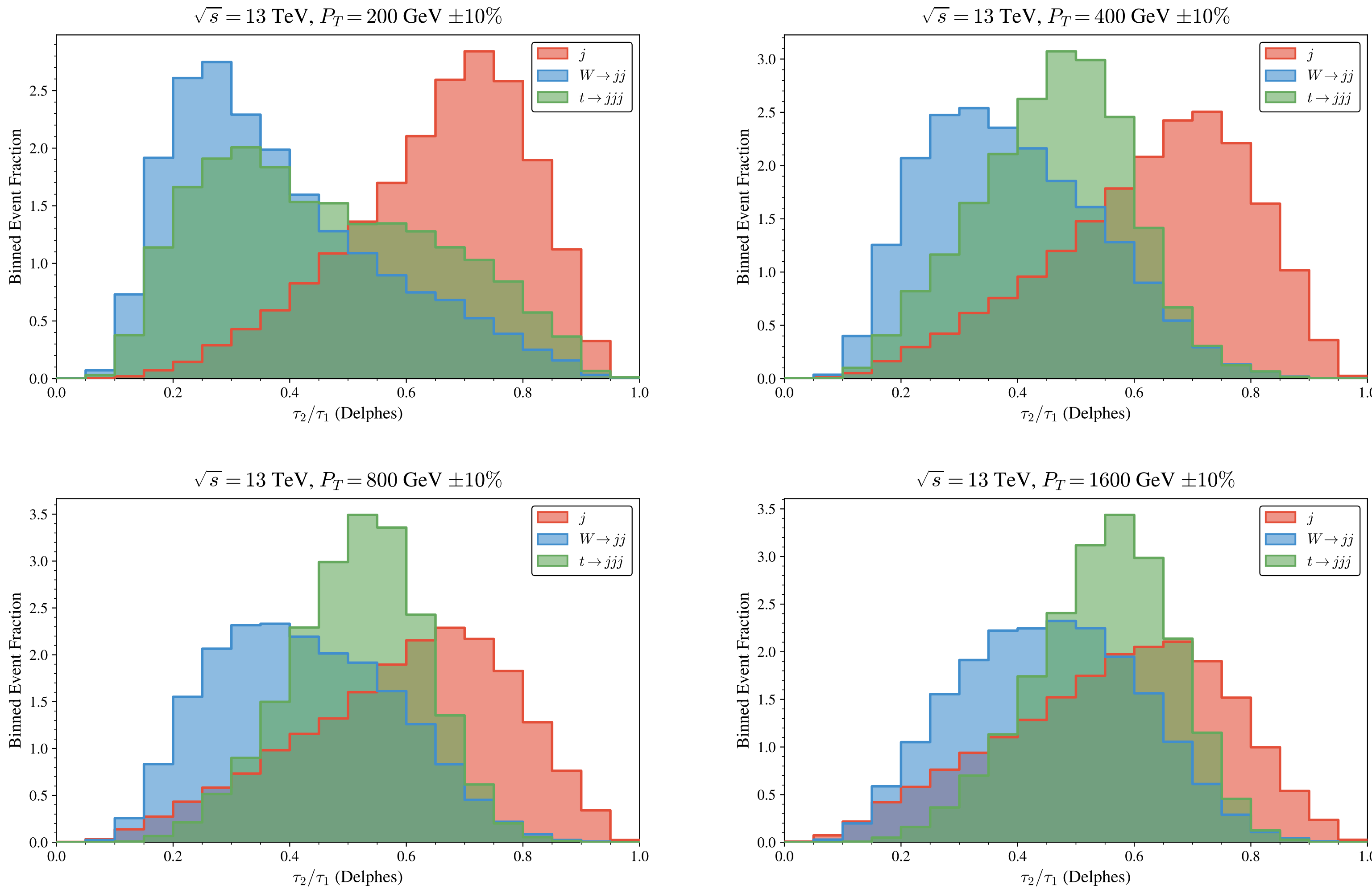
$$\text{Given } N \text{ axes } \hat{n}_k, \quad \tau_N = \frac{\sum_{i \in J} p_{T,i} \min(\Delta R_{ik})}{\sum_{i \in J} p_{T,i} R_0}$$

$$r_N = \frac{\tau_N}{\tau_{N-1}}$$

- It is interesting to ask if structure tagging can be incorporated into clustering
- To compare and assess performance, we simulate 1, 2 (W > j j), and 3 (t > j j j) jet event samples, at a range of transverse scales

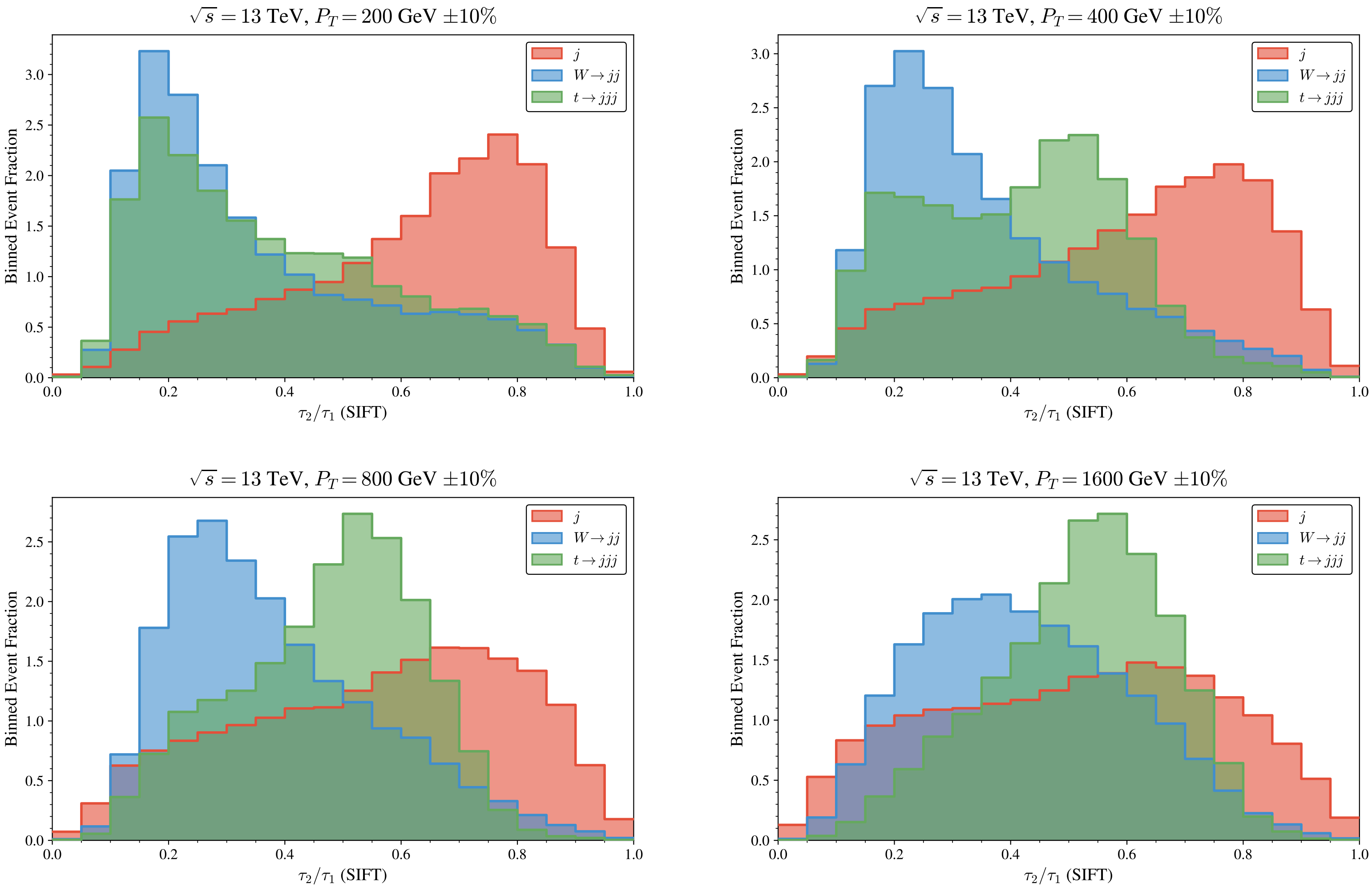
τ_2/τ_1 as implemented by Delphes

- The expected 2/1 discrimination is validated, but seen to degrade at high boost



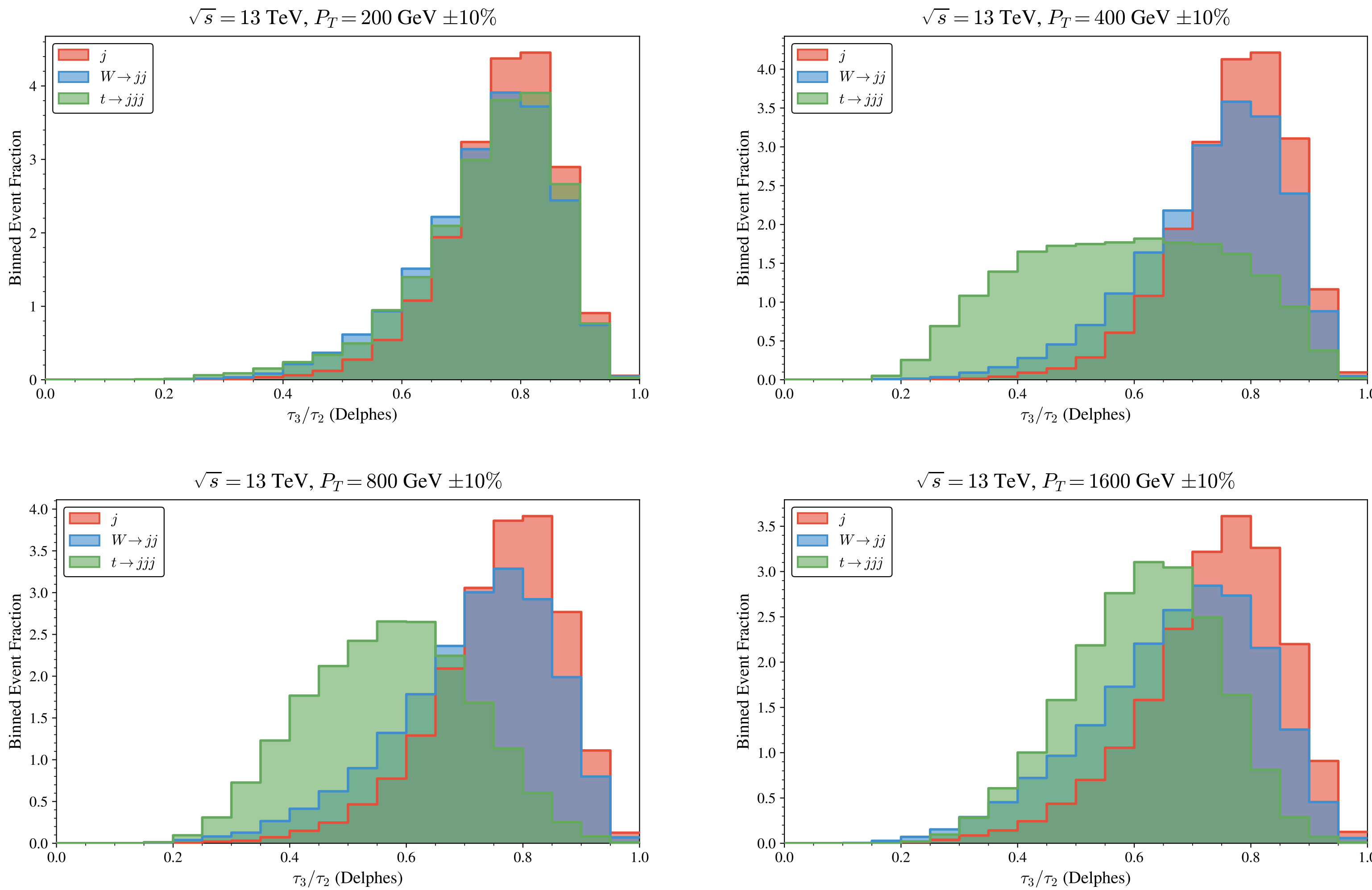
τ_2/τ_1 with SIFT Axes

- SIFT is also very good for N-subjettiness axis finding

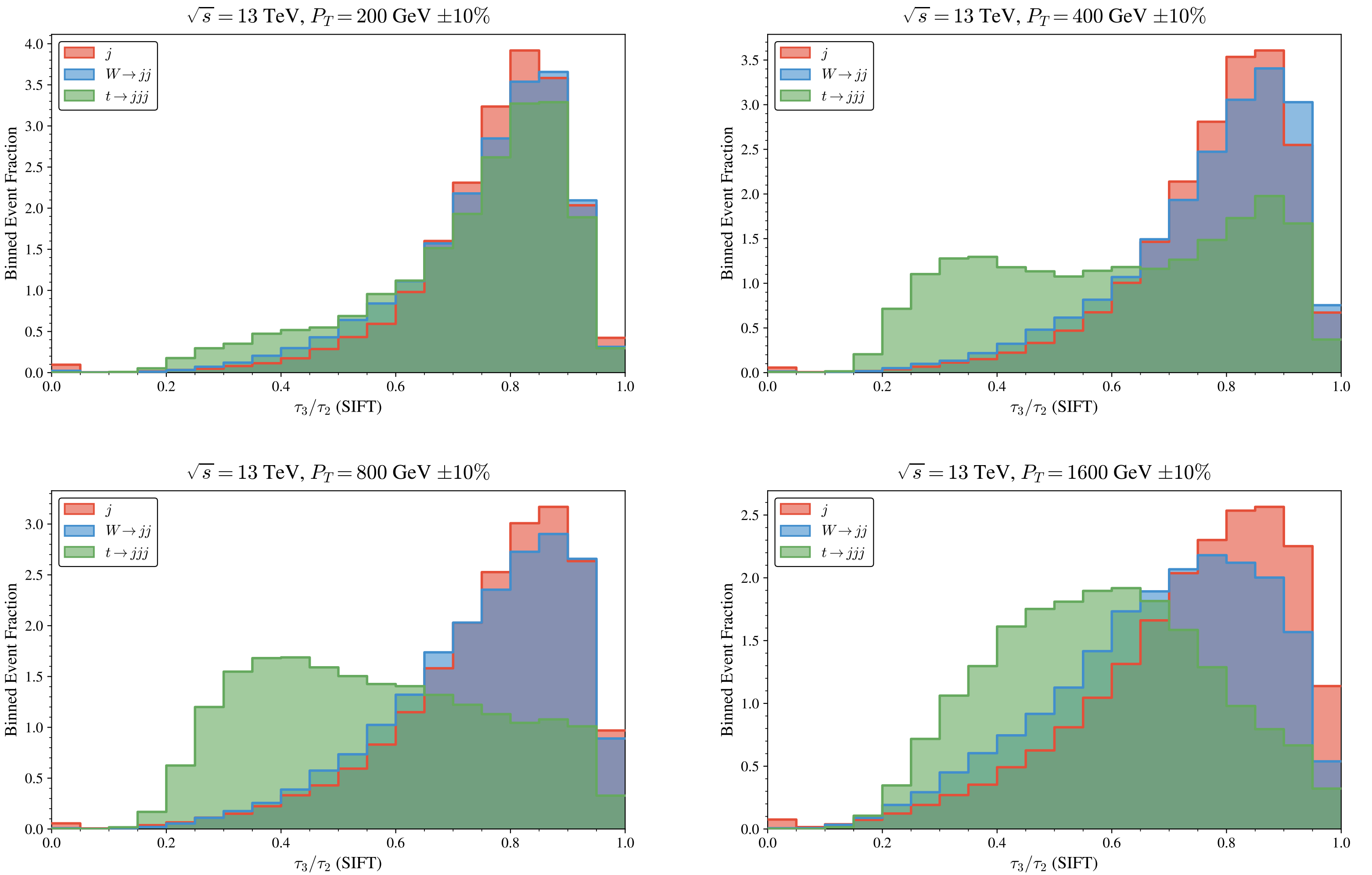


τ_3/τ_2 (Delphes)

- The next ratio is confirmed to separate t/W effectively, but likewise degrades

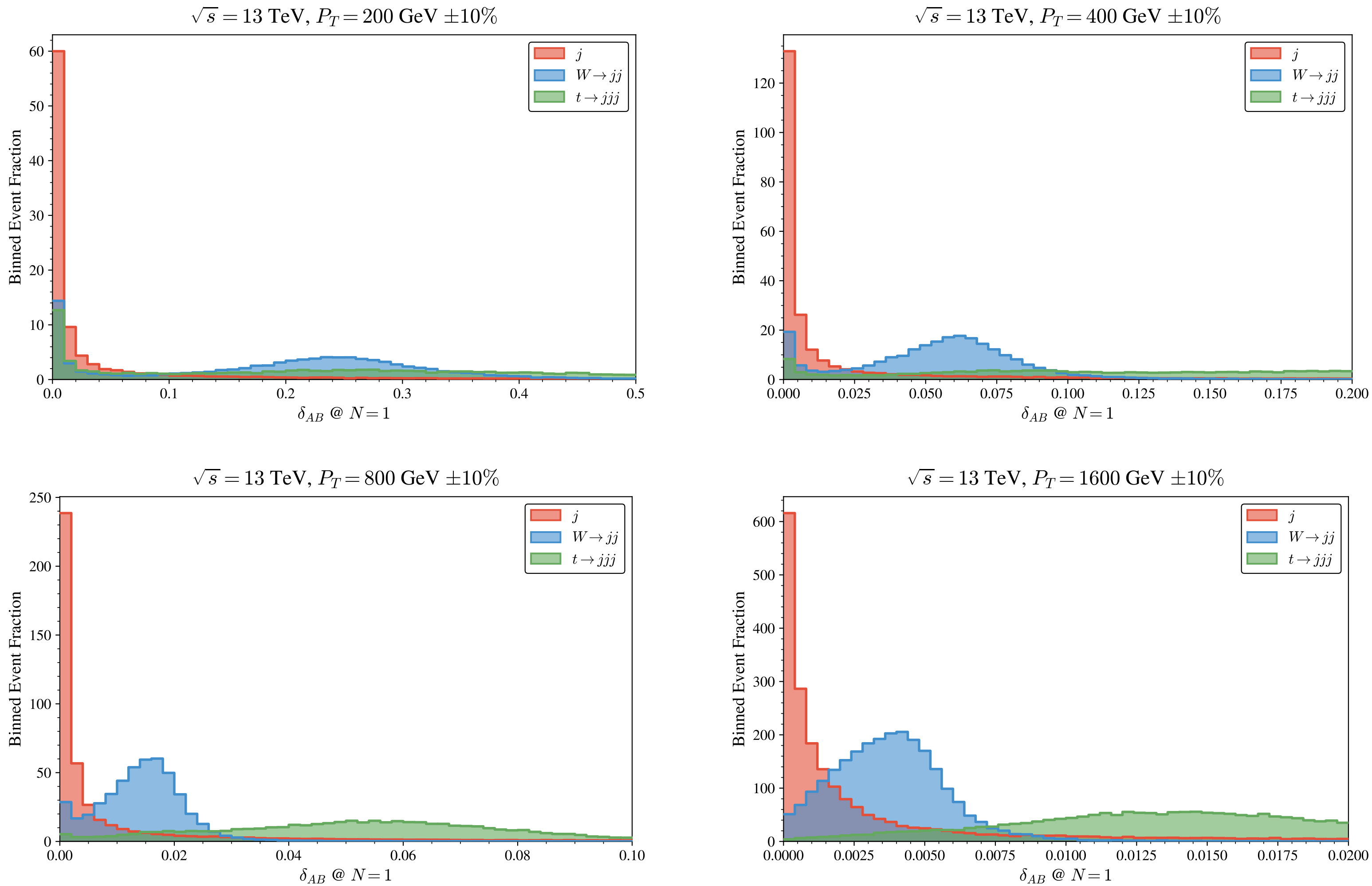


τ_3/τ_2 with SIFT Axes



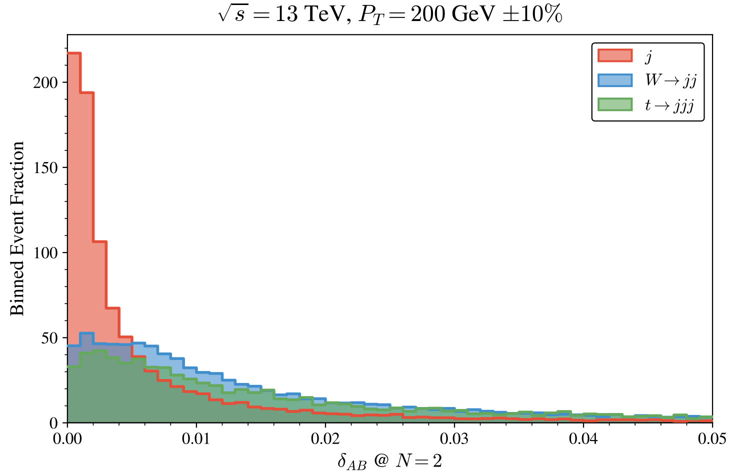
SIFT(1)

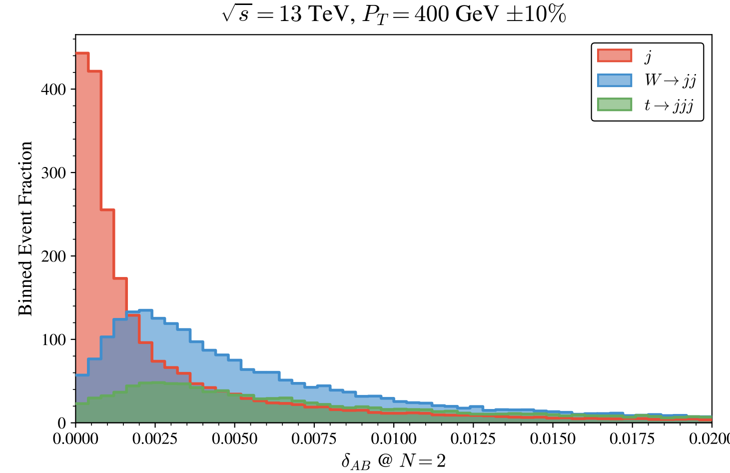
- We are also interested in whether the SIFT measure tracks jettiness DIRECTLY
- It seems not only to do so, but to excel specifically at large boost

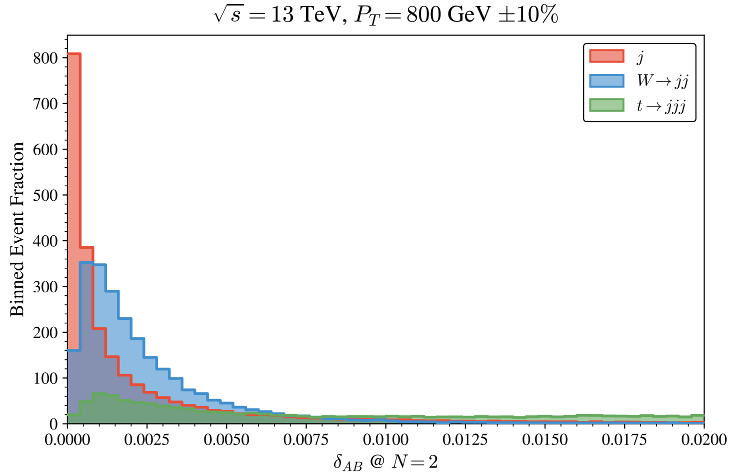


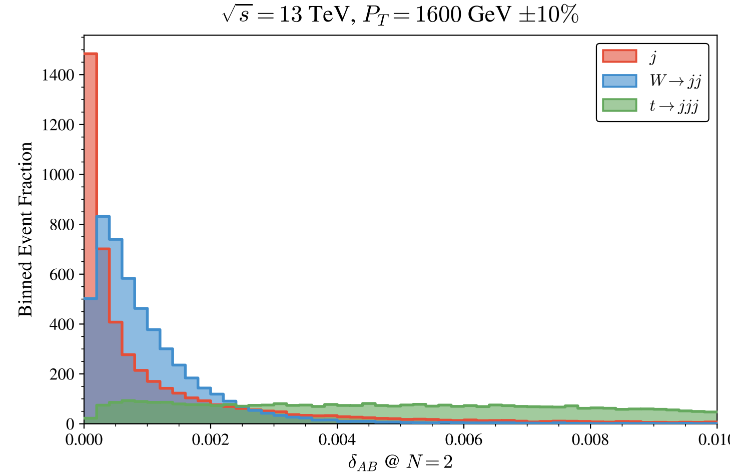
SIFT(2)

- The last several mergers hold the most information and are complementary



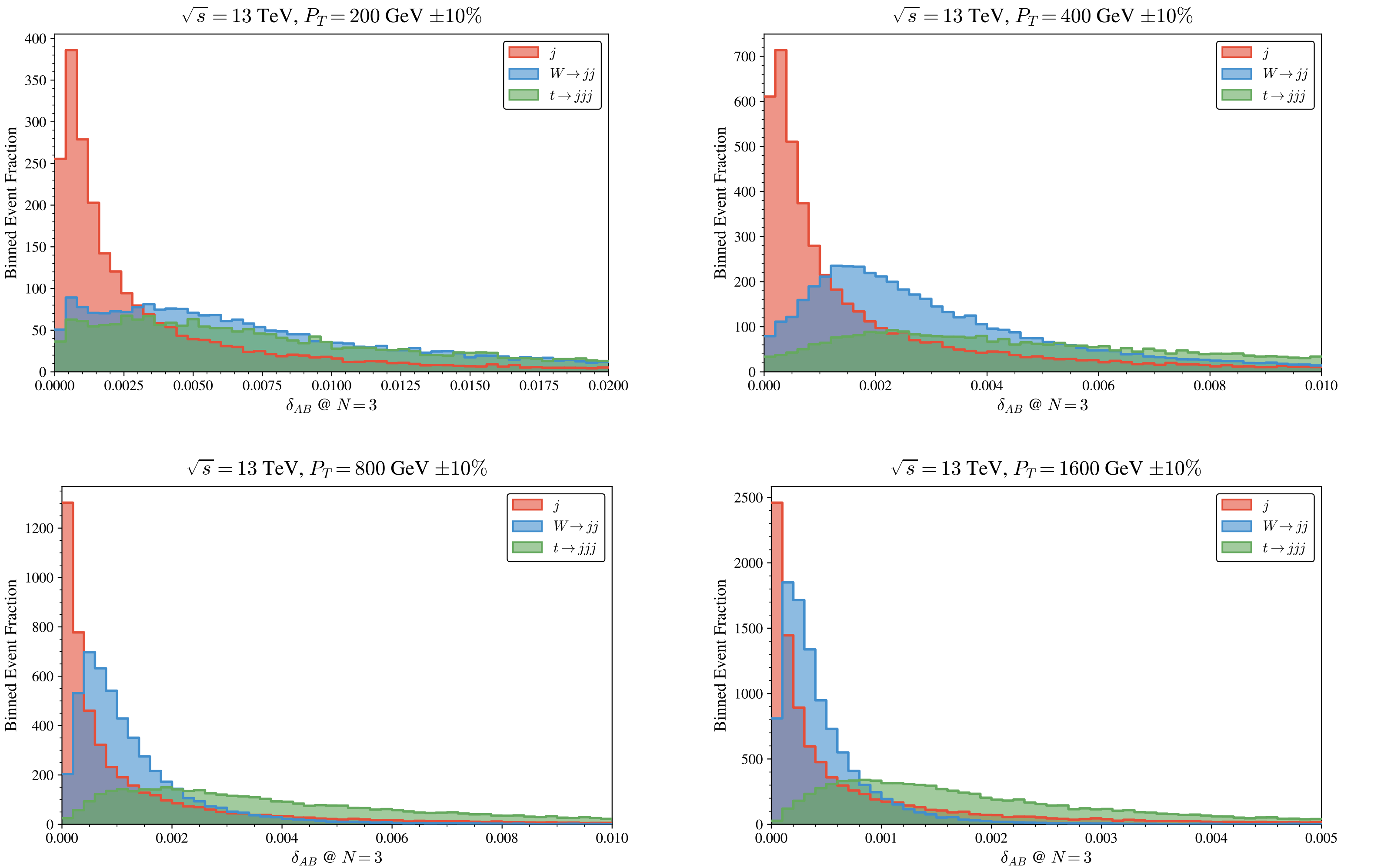






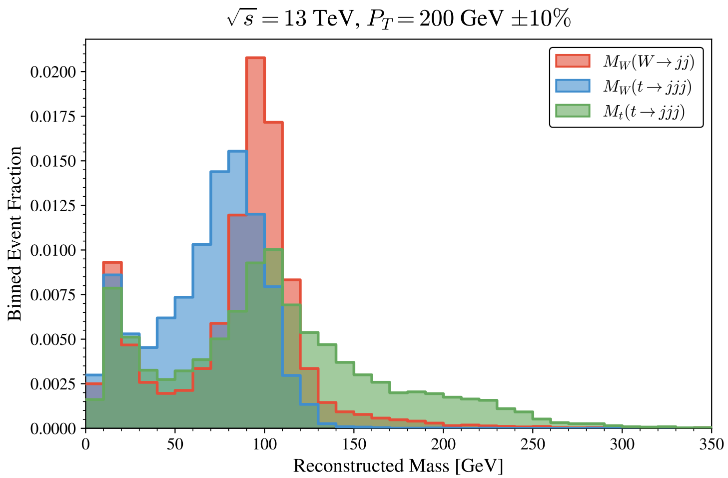
SIFT(3)

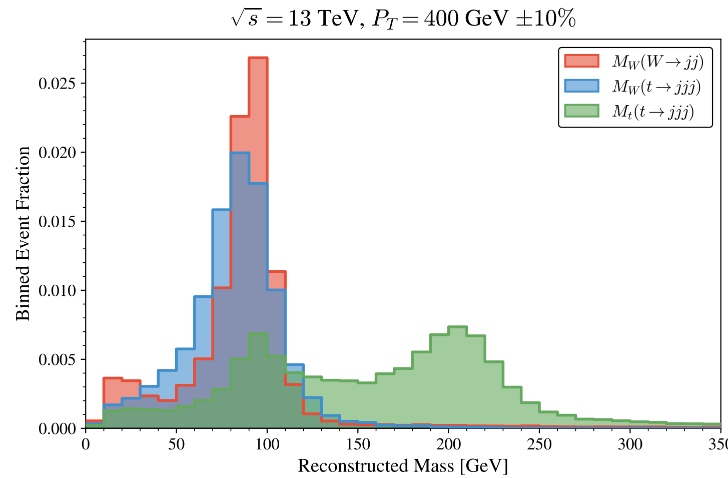
- In general, the measure is large if you have merged at least N hard prongs

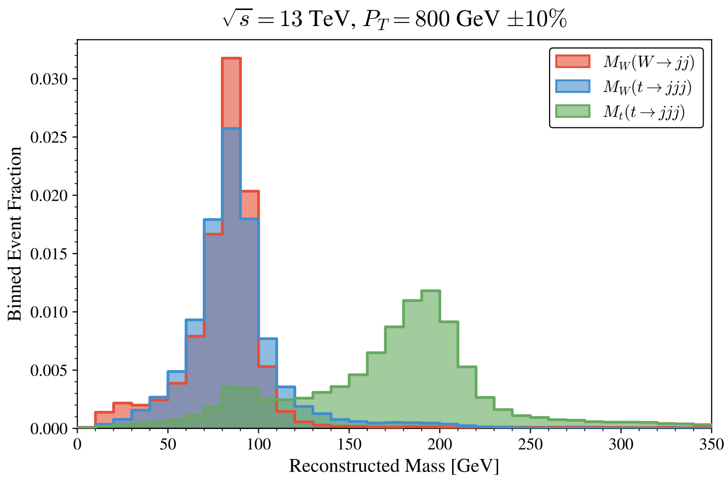


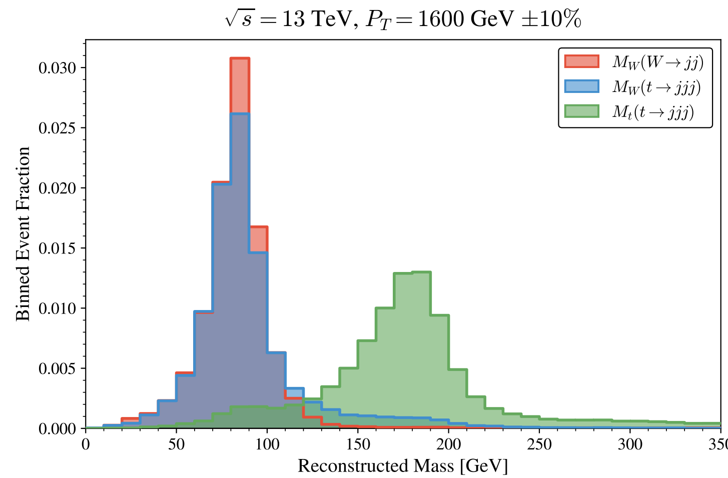
W/t Mass Reconstruction

- The included filtering gives sharp accurate mass reconstruction at large boost







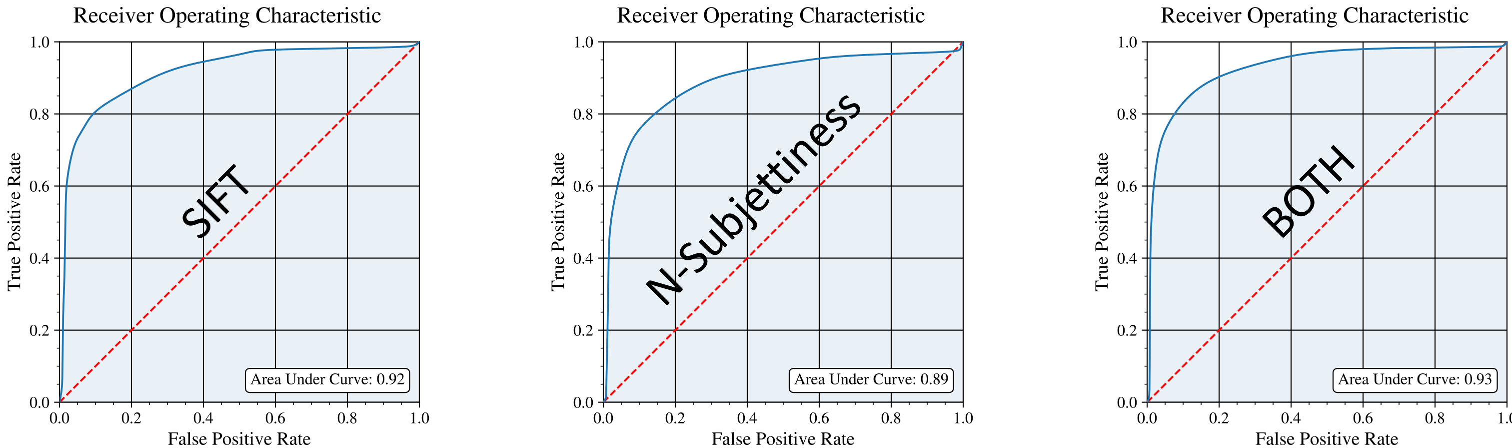


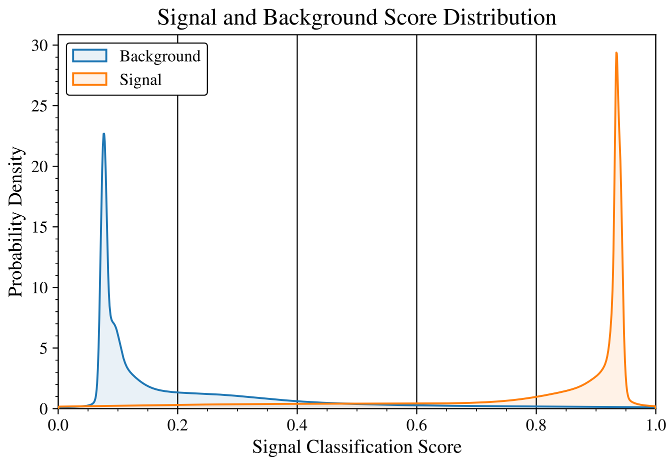
Assessing Performance

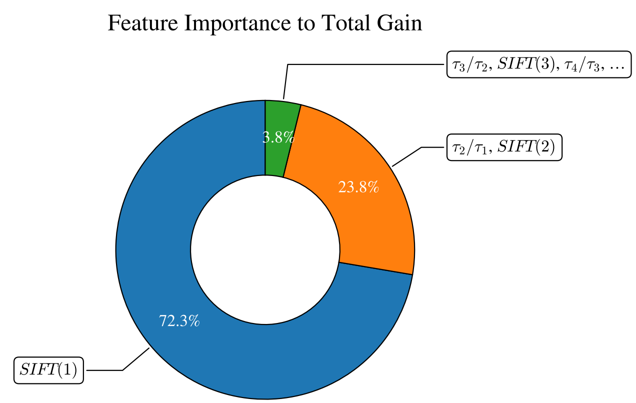
- A Boosted Decision Tree lets us compare information density in an unbiased way
- The BDT is also completely transparent, since it amounts simply to cascaded binary selection cuts (branchings) with assigned scores
- We feed the BDT Delphes N-subjettiness ratios up to 5/4
- We also provide it with the final values of the SIFT measure
- We compare outcomes in isolation, and with both data sets provided together
- We compare the power of 2/1 and 3/2 discrimination at a range of scales

2/1 Discrimination with BDT @200 GeV

- Performance is comparable at low boost, but the BDT gives the edge to SIFT

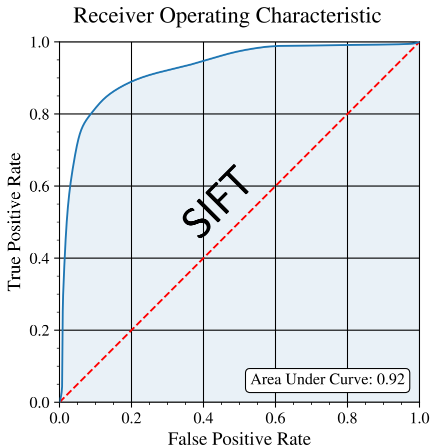


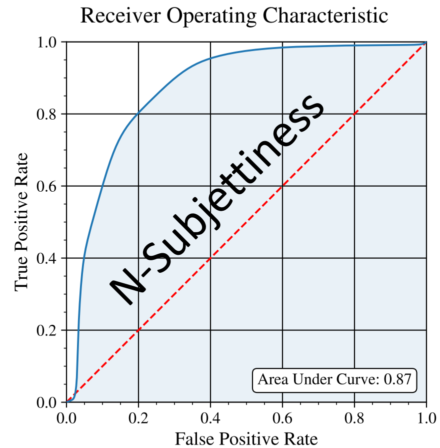


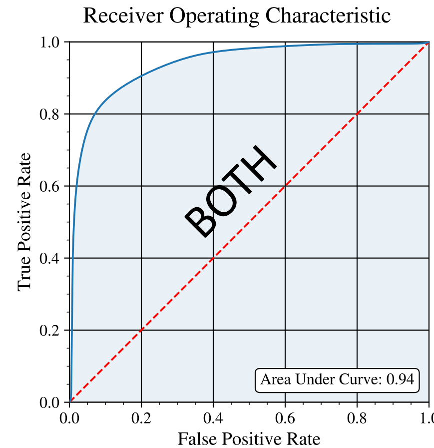


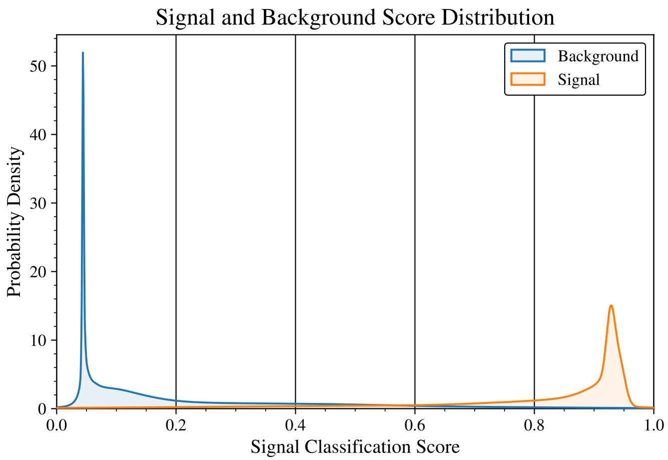
2/1 Discrimination with BDT @400 GeV

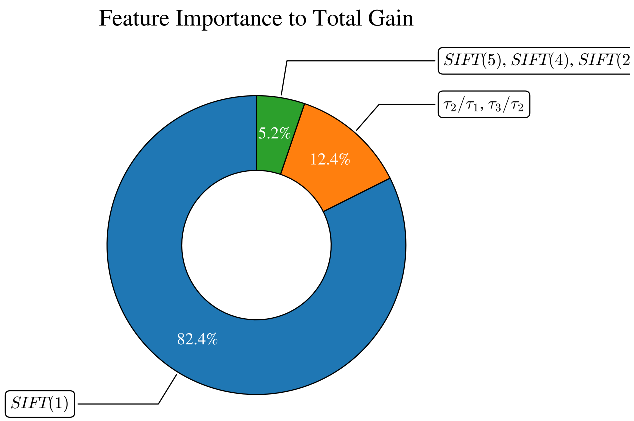
- The pattern continues as we climb in energy





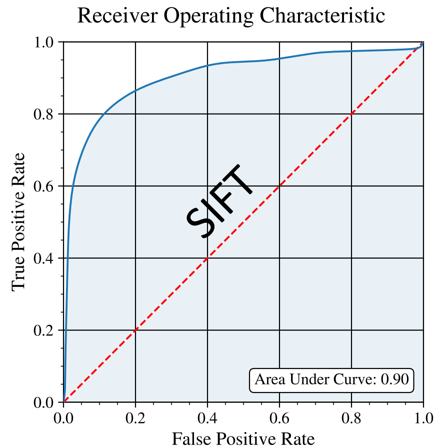


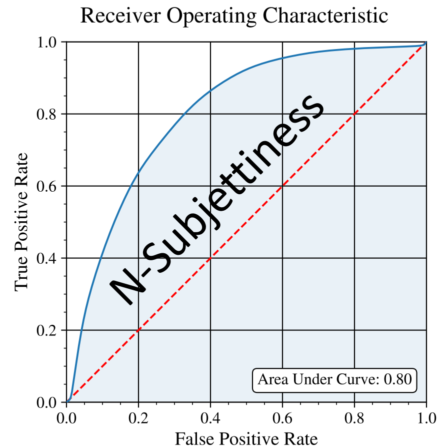


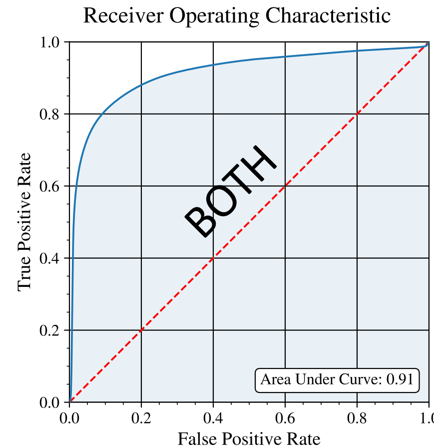


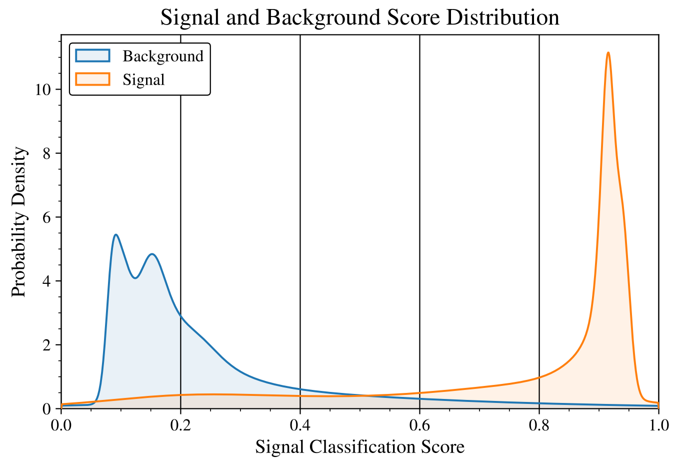
2/1 Discrimination with BDT @800 GeV

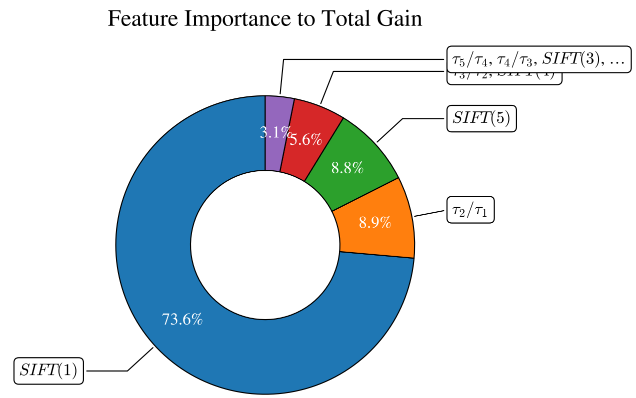
- At larger transverse scales, the edge tips decisively to the SIFT measure





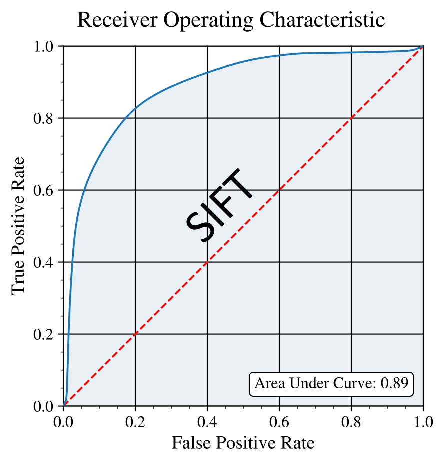


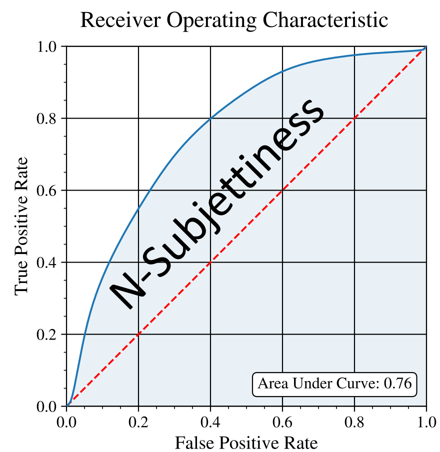


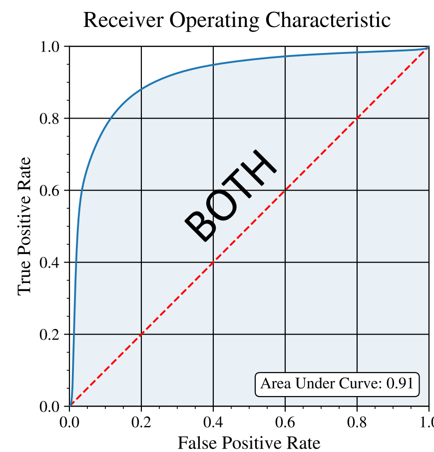


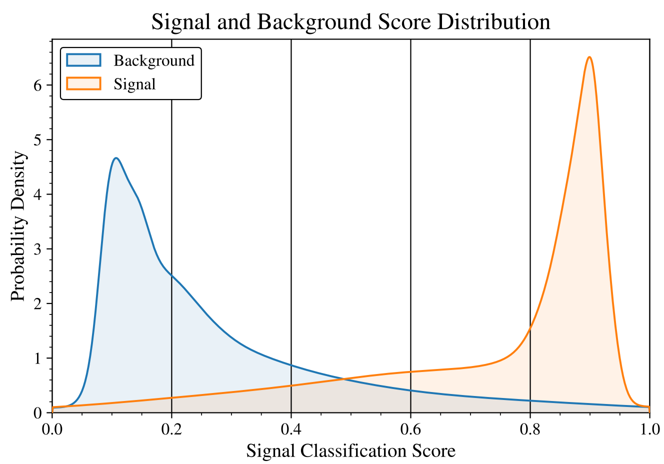
2/1 Discrimination with BDT @1.6 TeV

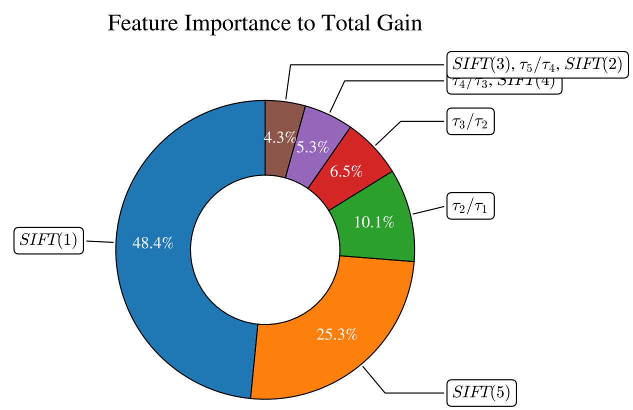
- At extreme boost, we observe a substantial advantage for SIFT





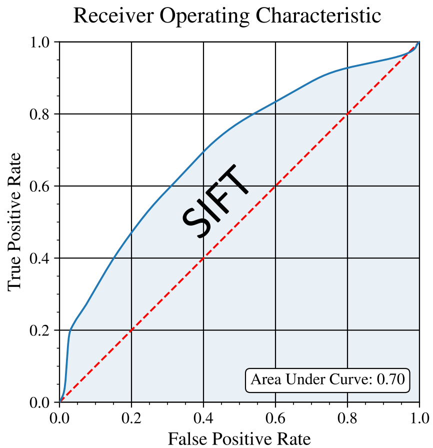


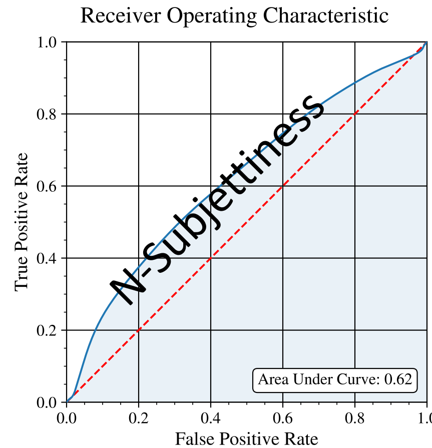


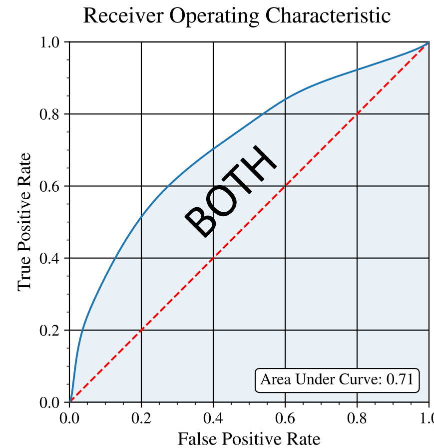


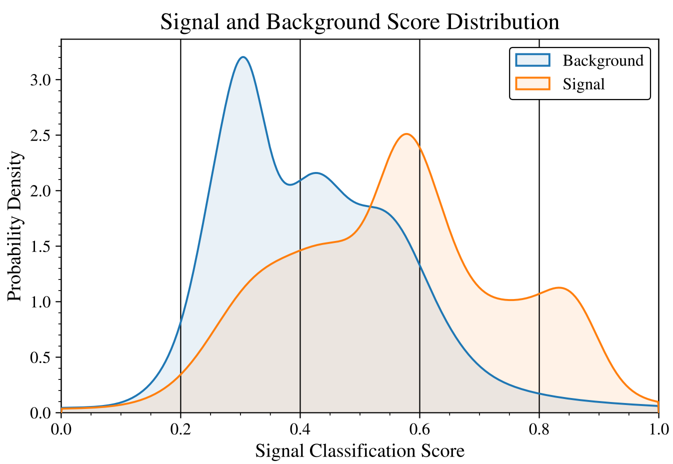
3/2 Discrimination with BDT @200 GeV

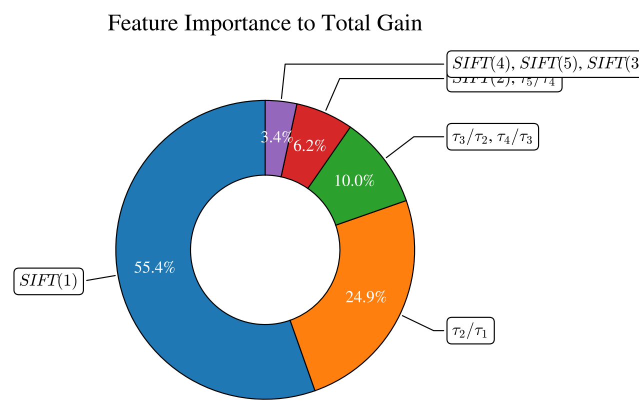
- The same pattern emerges in the 3/2 discrimination





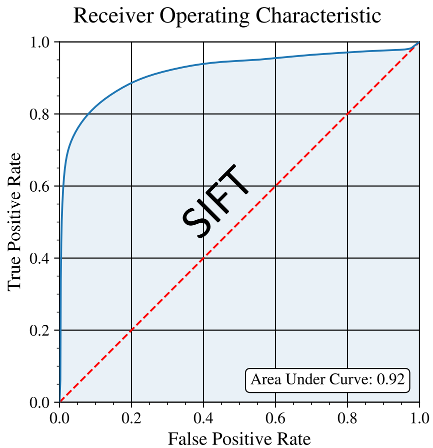


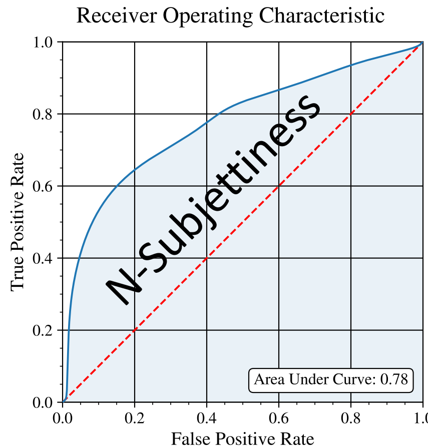


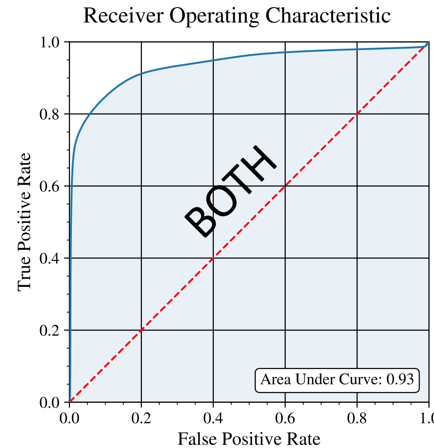


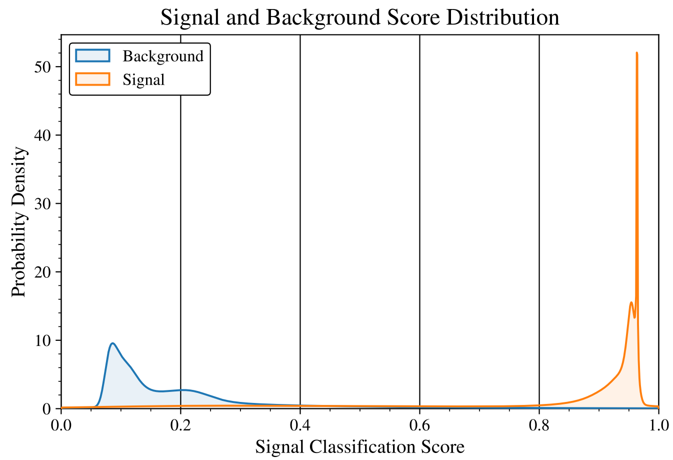
3/2 Discrimination with BDT @400 GeV

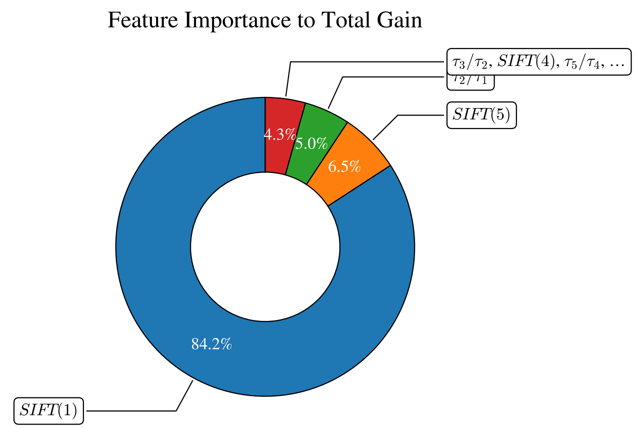
- In all cases, the combined power is greater than either method alone





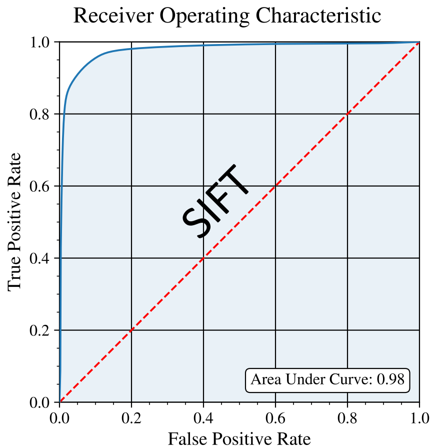


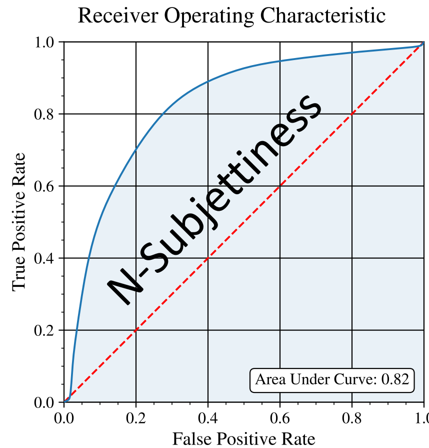


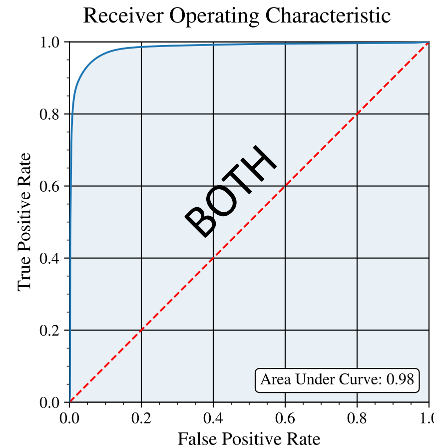


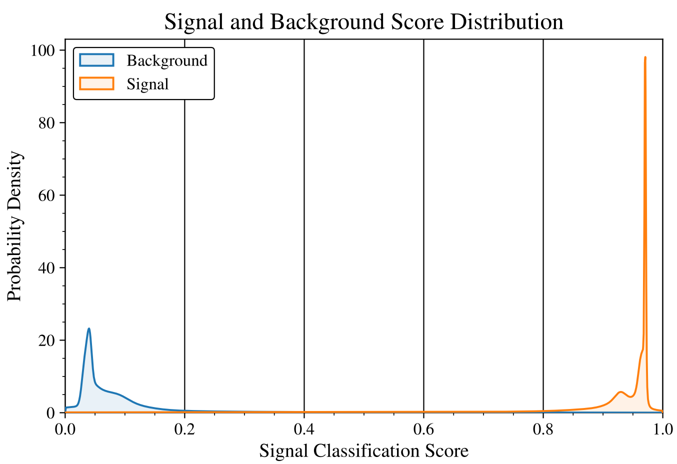
3/2 Discrimination with BDT @800 GeV

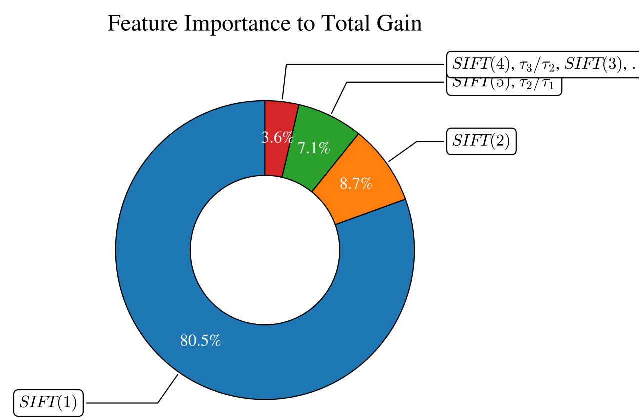
- Again, the main advantages emerge at large scales, where SIFT excels





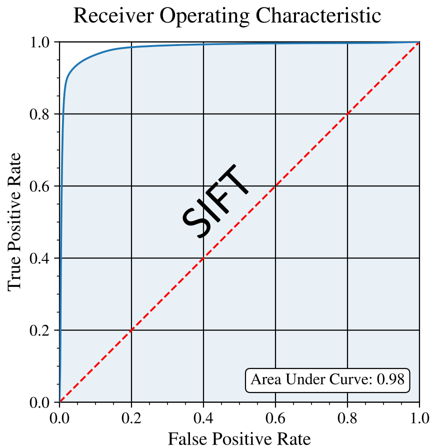


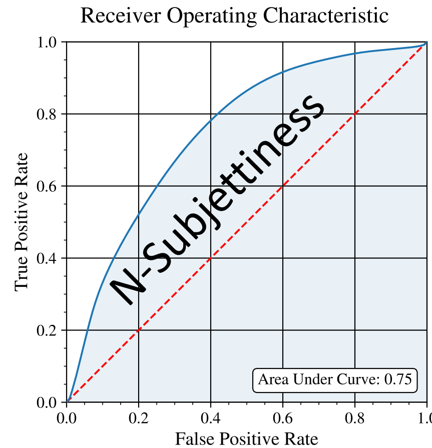


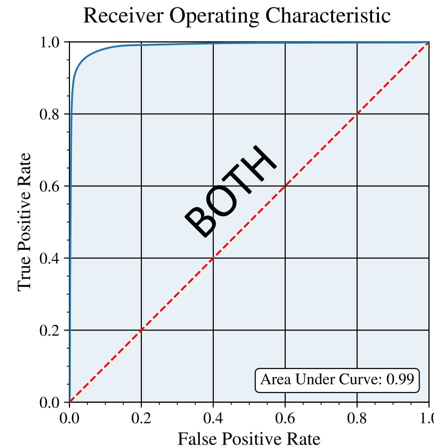


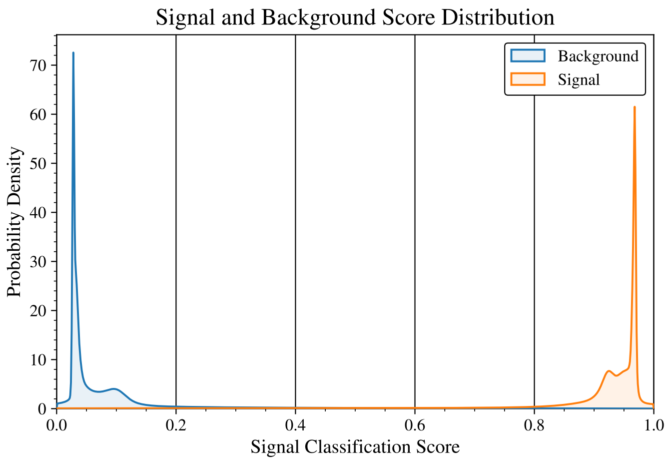
3/2 Discrimination with BDT @1.6 TeV

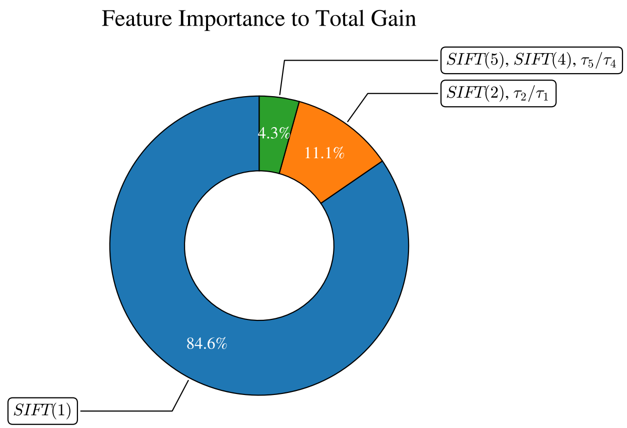
- We also find that SIFT is less sensitive to course graining of the object coordinates











Summary and Conclusions

- SIFT is a SCALE INVARIANT clustering algorithm designed to avoid losing substructure
- FILTERING of soft-wide radiation and variable-radius isolation is fully integrated
- The measure history & TREE of *N*-subjet axis candidates encode structure on the fly

Thank You!

Mathematica movie-generating notebook is available by request to jwalker@shsu.edu

Software Advertisement

- All data analysis for this project was performed with the indicated set of tools
- The package is available for download & public use from GitHub:
- https://github.com/joelwwalker/AEACuS
- I will help you!

PROCEEDINGS OF SCIENCE

Automated collider event selection, plotting, & machine learning with AEACuS, RHADAManTHUS, & MInOS

Joel W. Walker[a,*]

[a]*Department of Physics and Astronomy, Sam Houston State University, Box 2267, Huntsville, TX 77341, USA*

E-mail: jwalker@shsu.edu

A trio of automated collider event analysis tools are described and demonstrated, in the form of a quick-start tutorial. AEACuS interfaces with the standard MadGraph/MadEvent, Pythia, and Delphes simulation chain, via the Root file output. An extensive algorithm library facilitates the computation of standard collider event variables and the transformation of object groups (including jet clustering and substructure analysis). Arbitrary user-defined variables and external function calls are also supported. An efficient mechanism is provided for sorting events into channels with distinct features. RHADAManTHUS generates publication-quality one- and two-dimensional histograms from event statistics computed by AEACuS, calling MatPlotLib on the back end. Large batches of simulation (representing either distinct final states and/or oversampling of a common phase space) are merged internally, and per-event weights are handled consistently throughout. Arbitrary bin-wise functional transformations are readily specified, e.g. for visualizing signal-to-background significance as a function of cut threshold. MInOS implements machine learning on computed event statistics with XGBoost. Ensemble training against distinct background components may be combined to generate composite classifications with enhanced discrimination. ROC curves, as well as score distribution, feature importance, and significance plots are generated on the fly. Each of these tools is controlled via instructions supplied in a reusable cardfile, employing a simple, compact, and powerful meta-language syntax.

PoS(CompTools2021)027

Collider Variables & Coordinates

- Transverse components (perpendicular to the beam) are very important (invariant under longitudinal boosts, P_T total is zero)
- Differences in orientation characterized by ΔR, referring also to azimuth angle ϕ
- The pseudorapidity η is a proxy for the polar (beam) angle θ, defined such that differences $\Delta\eta$ are (almost) invariant under longitudinal boosts
- This invariance is exact for the rapidity y (difference is handling of MASS)

$$P_{\mathrm{T}} \equiv \sqrt{P_x^2 + P_y^2}$$

$$\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

$$\eta \equiv \frac{1}{2}\ln\left(\frac{|\vec{P}| + P_z}{|\vec{P}| - P_z}\right) \equiv -\ln\tan\left(\frac{\theta}{2}\right)$$

$$y \equiv \frac{1}{2}\ln\left(\frac{E + P_z}{E - P_z}\right) \equiv \ln\left(\frac{\sqrt{\cosh^2\eta + \frac{M^2}{P_{\mathrm{T}}^2}} + \sinh\eta}{\sqrt{1 + \frac{M^2}{P_{\mathrm{T}}^2}}}\right)$$

FIG. 1: The pseudorapidity η (bold, orange) is plotted as a function of the polar angle θ. For comparison, the longitudinal rapidity y (fine, blue) is also shown for various values of M/P_{T}, equal to $\{1/2, 1, 2, 5, 10, 20\}$ from top to bottom.

Pseudorapidity η vs. Polar Angle θ