Jet SIFT-ing

for New Dark Sector Physics

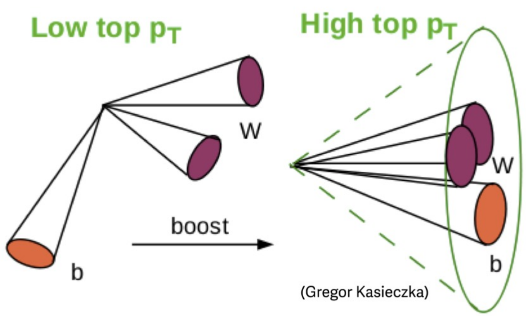
Joel W. Walker
Sam Houston State University

Part I (SIFT Algorithm) with:
Andrew Larkoski (UCLA), Denis Rathjens (CMS), and Jason Veatch (ATLAS)
arXiv: 2302.08609 – to appear in Physical Review D

Part II (Hadronic Dark Sector Mass Reconstruction) with:
William Shepherd, James Floyd, Camryn Sanders, and Jonathan Mellenthin
(hidden valley application is preliminary)

CETUP*
The Institute for Underground Science at SURF
June 19, 2023

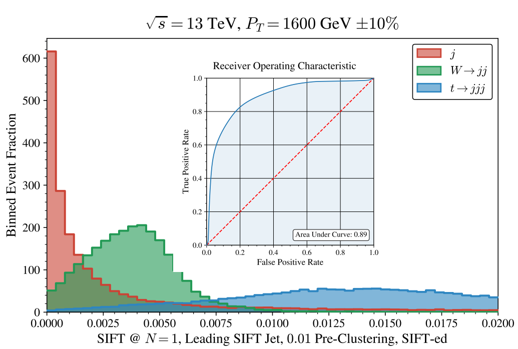
SIFT: Scale-Invariant Filtered Tree

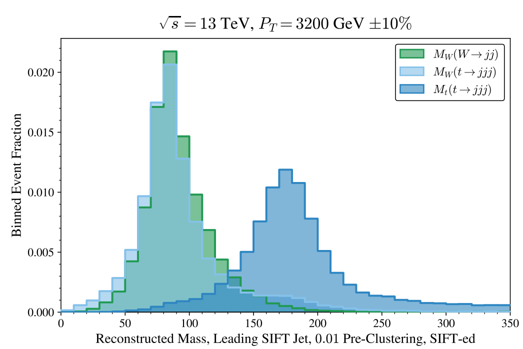


- Massive resonances decay into hard prongs
- Jet definitions with fixed cones impose a scale
- Boosted objects collimate and structure is lost
- Substructure recovery techniques are complex
- Can we avoid losing resolution in the first place?
- Select proximal objects w/ scale-invariant measure

- 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

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

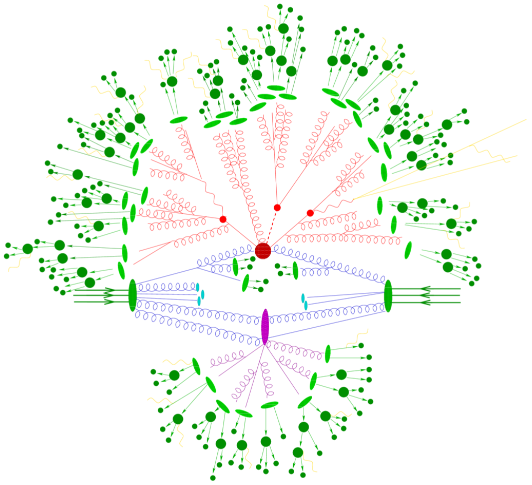




- *N*-subjet Tree holds superposition of projections onto *N*=1,2,3 prongs
- Hard prongs are preserved to end
- The measure history discriminates *N*=1,2,3 typically above 90% AUC
- Faithful kinematic reconstruction

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.



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

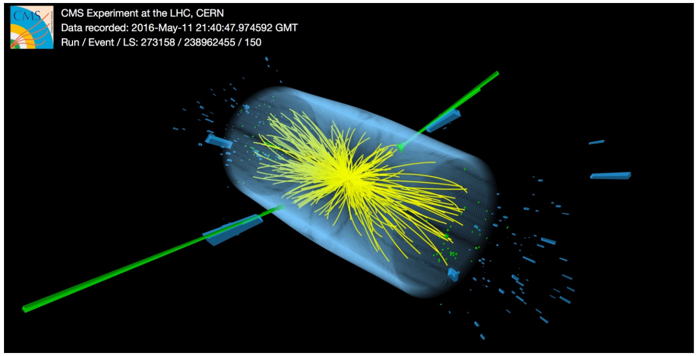
Image: Stefan Höche

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

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

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



$$E_{\mathrm{T}} \equiv \sqrt{M^2 + \vec{P}_{\mathrm{T}} \cdot \vec{P}_{\mathrm{T}}} = \sqrt{E^2 - P_z^2}$$
$$\lim_{M=0} \Rightarrow |\vec{P}_{\mathrm{T}}|$$

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 to examine the SIFT numerator:

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

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

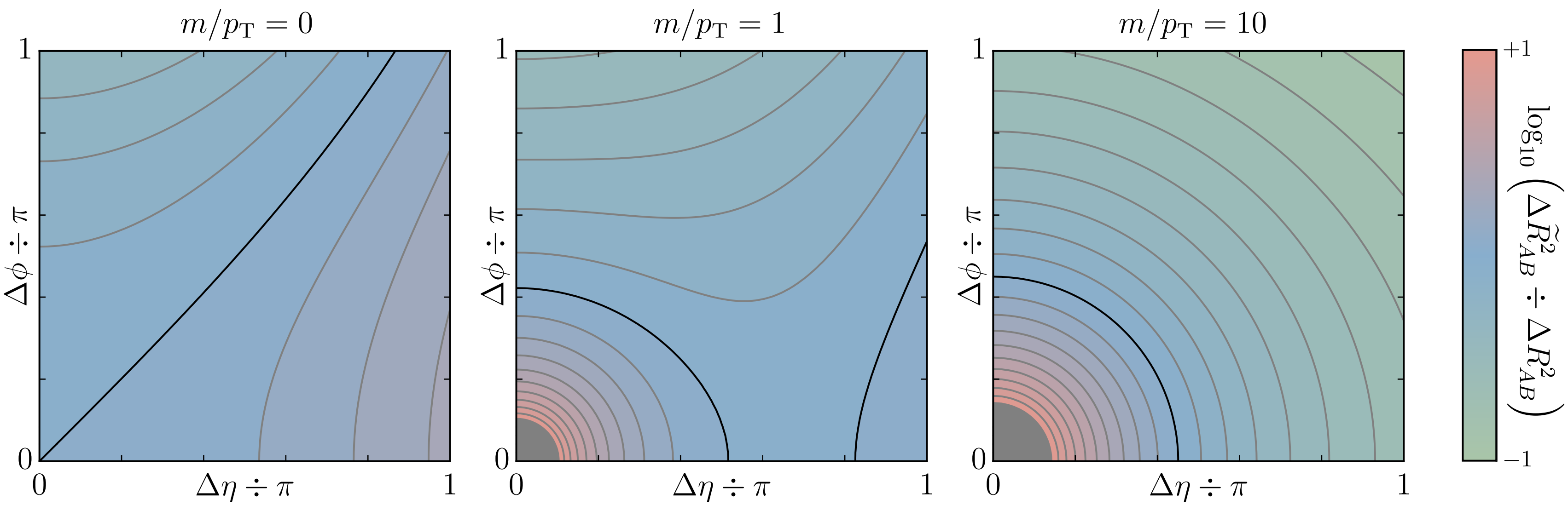
$$\begin{aligned} & E^A E^B - p_z^A p_z^B \\ = \; & E_{\mathrm{T}}^A E_{\mathrm{T}}^B \times \left(\cosh y^A \cosh y^B - \sinh y^A \sinh y^B \right) \\ = \; & E_{\mathrm{T}}^A E_{\mathrm{T}}^B \times \cosh \Delta y_{AB} \end{aligned}$$

- 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



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

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

- 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

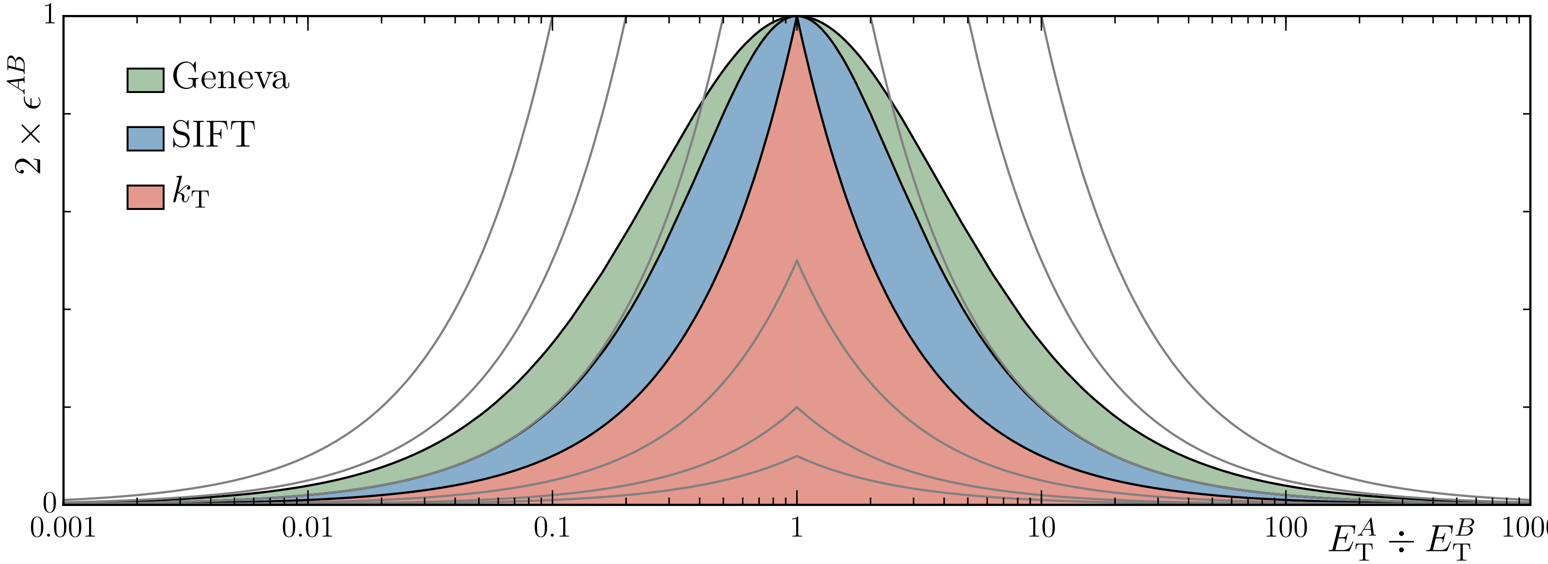
- The remainder of the metric refers only to transverse energy RATIOS
- This factor has a symmetry under $E_T \to 1/E_T$
- It asymptotically mimics BOTH kt and anti-kt clustering, preferencing the clustering of pairs with hierarchically DISPARATE transverse scales
- It has the benefit of being ANALYTIC

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

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

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

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

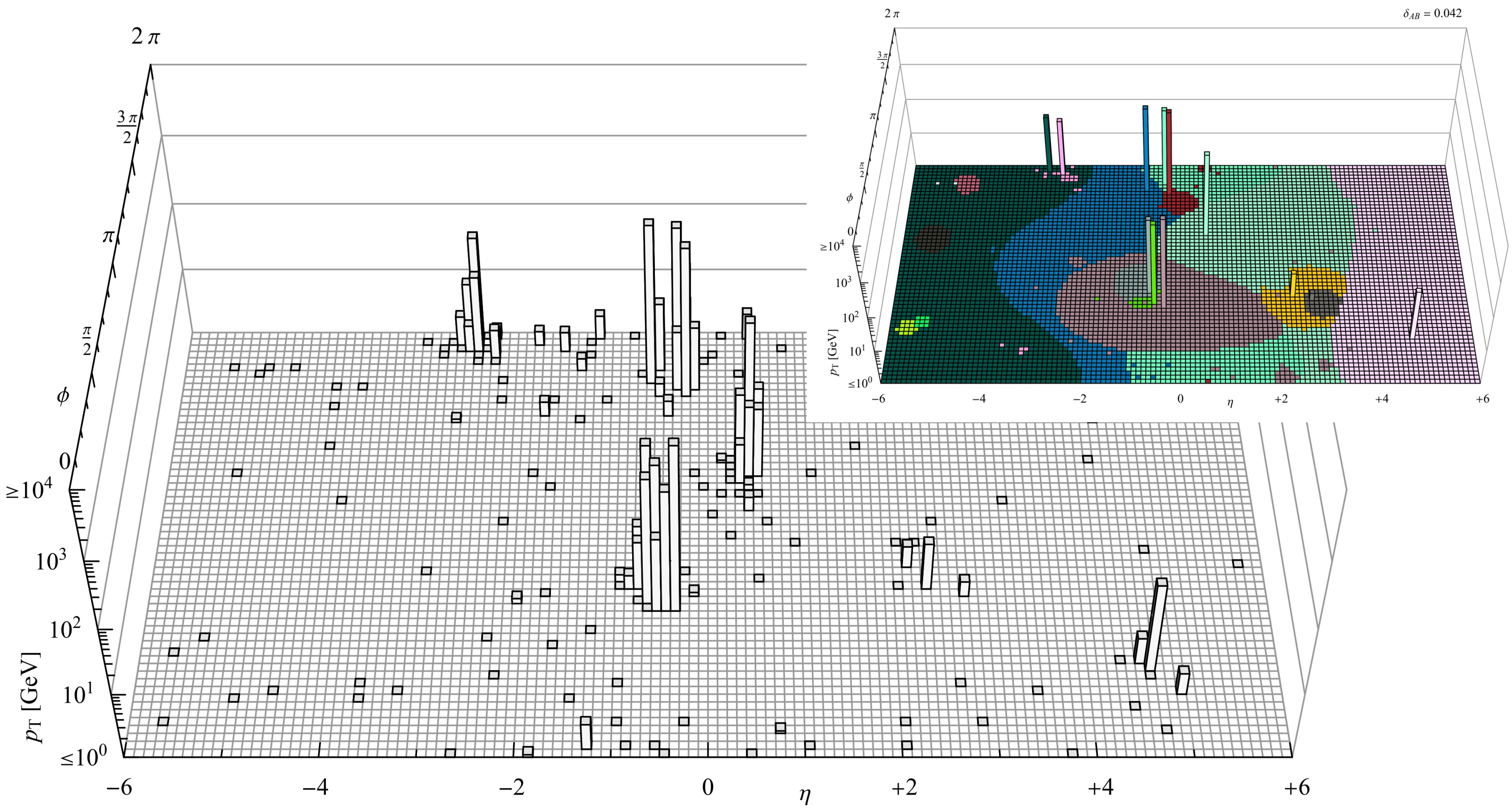
$$\delta_{AB} = \epsilon^{AB} \times \Delta\widetilde{R}^2_{AB}$$
$$= \frac{\cosh \Delta y_{AB} - \xi^A \xi^B \cos \Delta\phi_{AB}}{\cosh \Delta u_{AB}}$$

- 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

Several problems remain beyond the measure (read on for the solutions …)

- 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

pp to TTbar (pT ~ 800 GeV)
Scale Invariant Clustering with Ghost Radiation



- 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

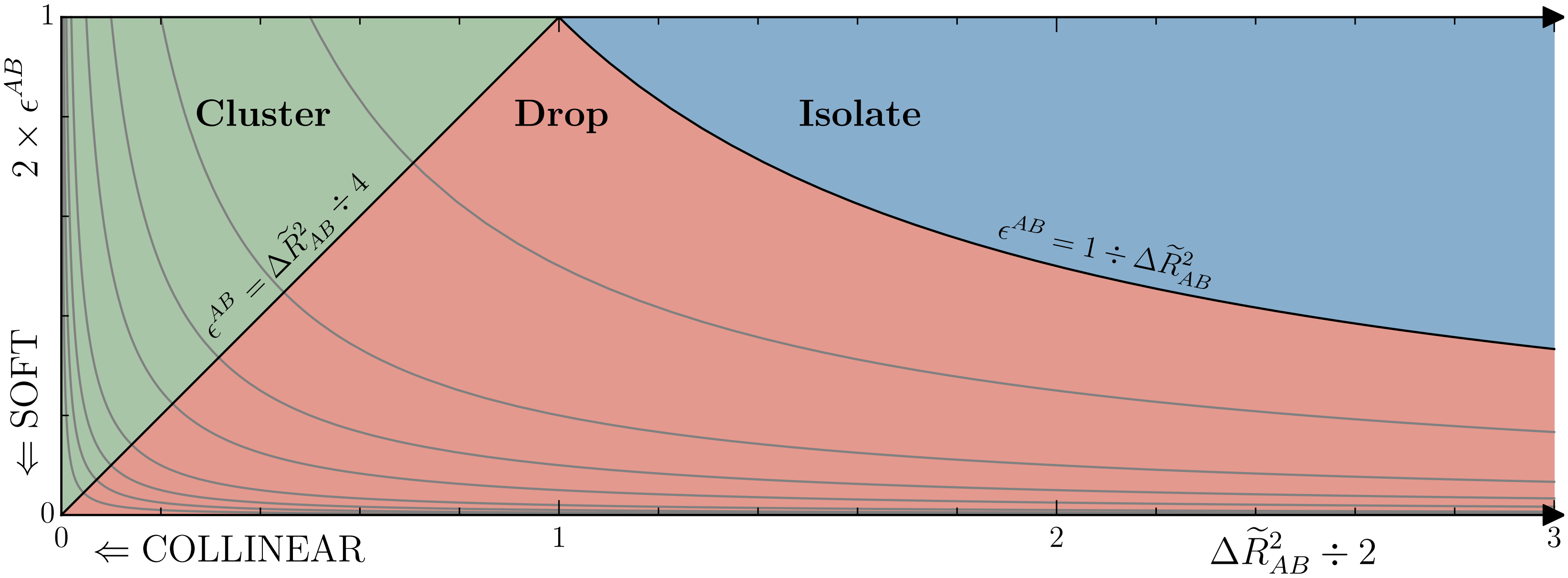
$$(\Delta\widetilde{R}^2_{AB} \gg 1) \text{ and } (\epsilon_{AB} \simeq 1)$$

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

$$\text{Isolate:} \qquad \{1\} \le \delta_{AB}$$

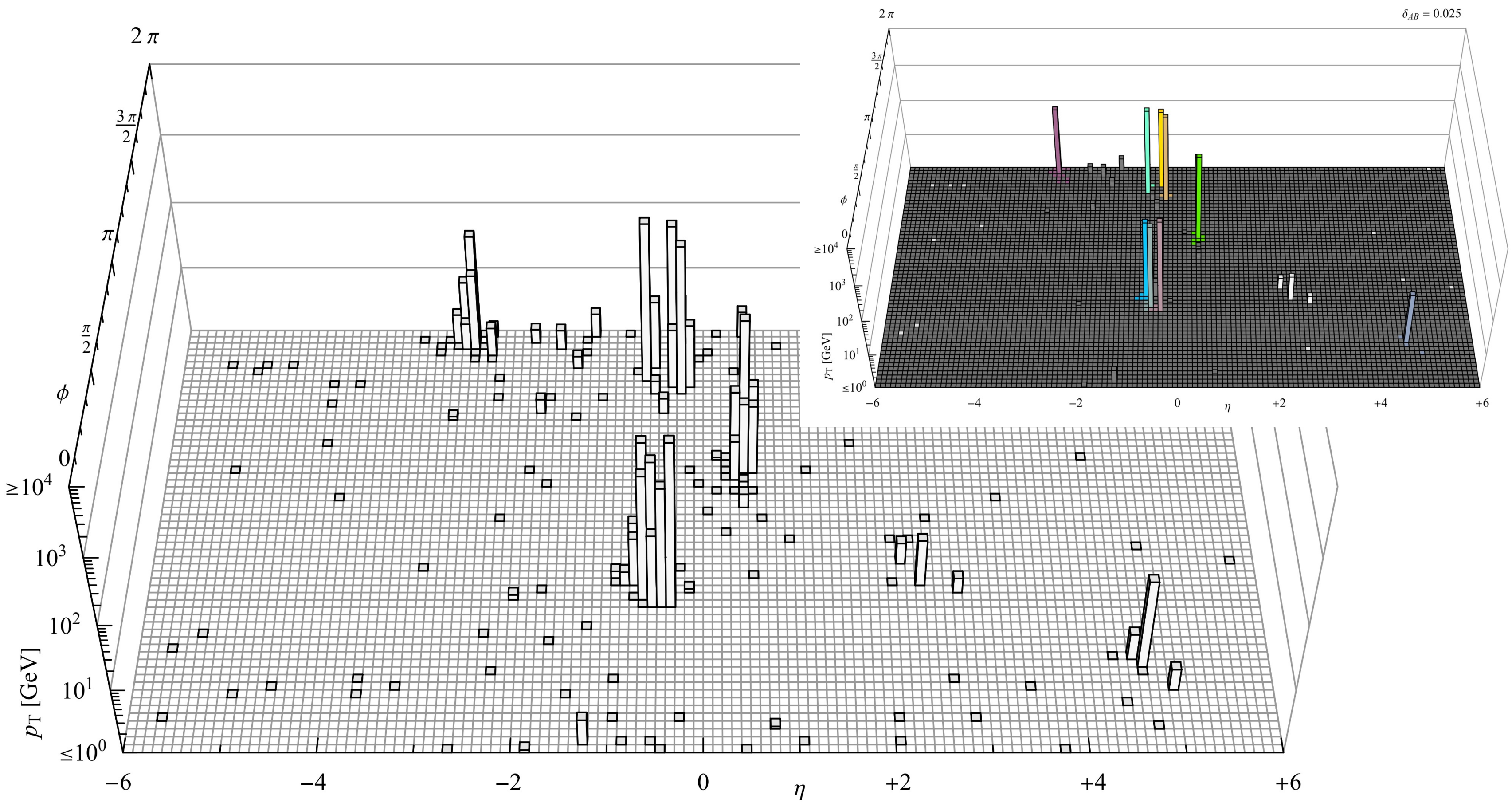
$$\text{Drop:} \qquad \{\,(2\,\epsilon_{AB})^2 \le 1\} \le \delta_{AB} < \{1\}$$

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 "$y = x$"

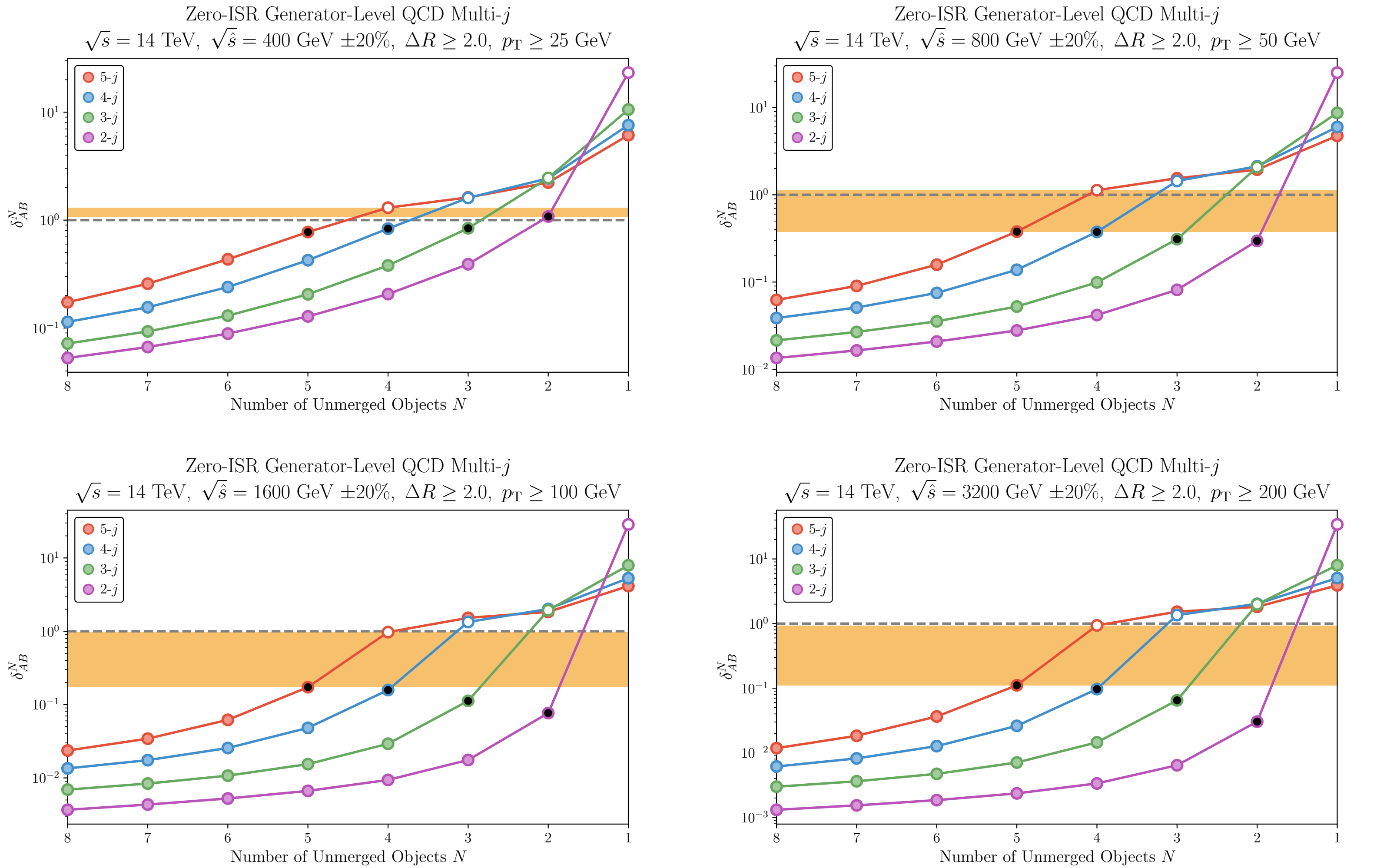
pp to TTbar (pT ~ 800 GeV)
Filtered Scale Invariant Clustering with Ghost Radiation



- See Video “E” Posted at Indico

Evolution of the Measure

- The measure “jumps” when it crosses the natural joint count
- The transition to isolation for $\delta \geq 1$ is supported by simulation



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

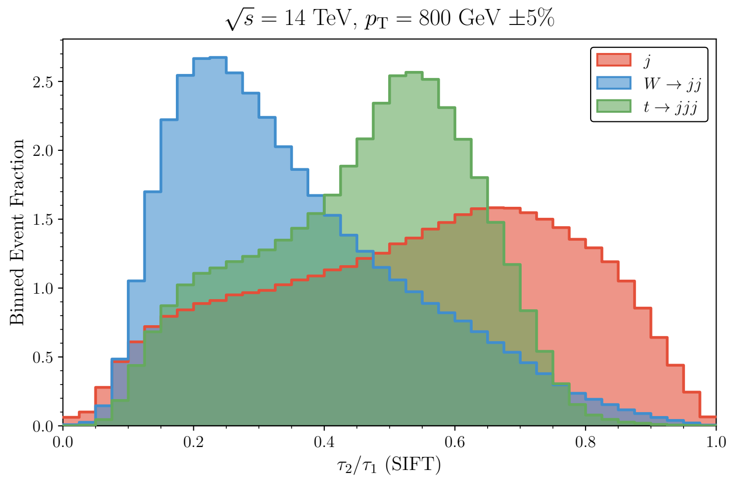
$$\text{Given } N \text{ axes } \hat{n}_k, \quad \tau_N = \frac{\sum_{i \in J} p_{T,i} \min(\Delta R_{ik})}{\sum_{i \in J} p_{T,i} R_0}$$

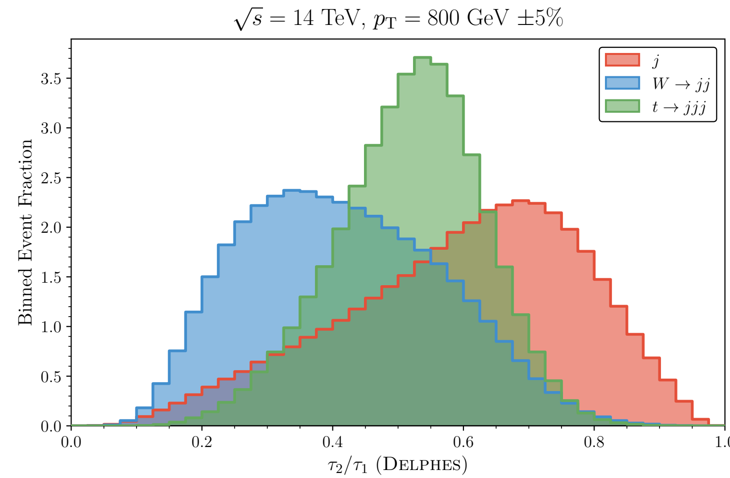
$$r_N = \frac{\tau_N}{\tau_{N-1}}$$

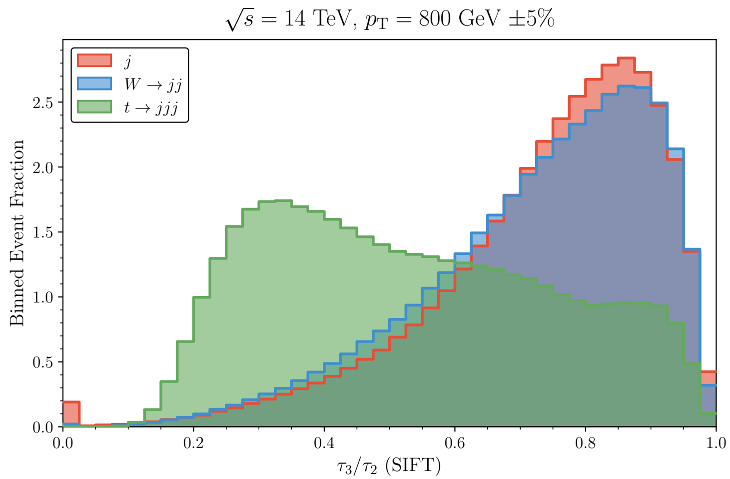
- It is interesting to ask if structure tagging can be incorporated into clustering
- To compare and assess performance, we simulate 1, 2 (W > j j), and 3 (t > j j j) jet event samples, at a range of transverse scales

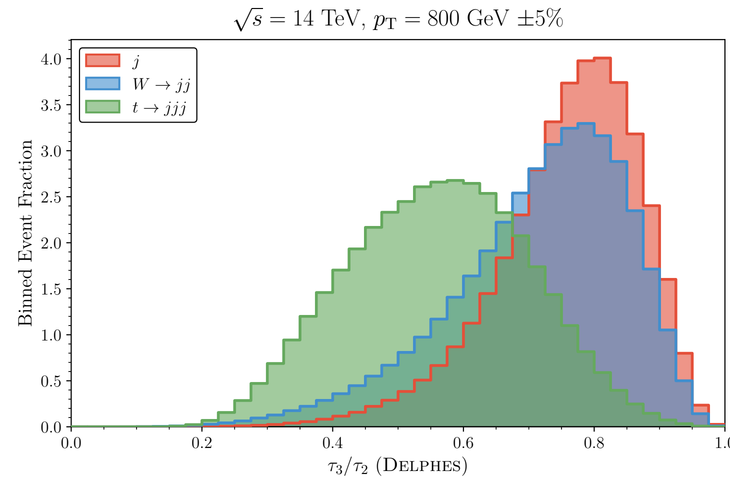
τ_2/τ_1 and τ_3/τ_2 with SIFT Axes

- SIFT is very good for N-subjettiness axis finding (Delphes versions on right)



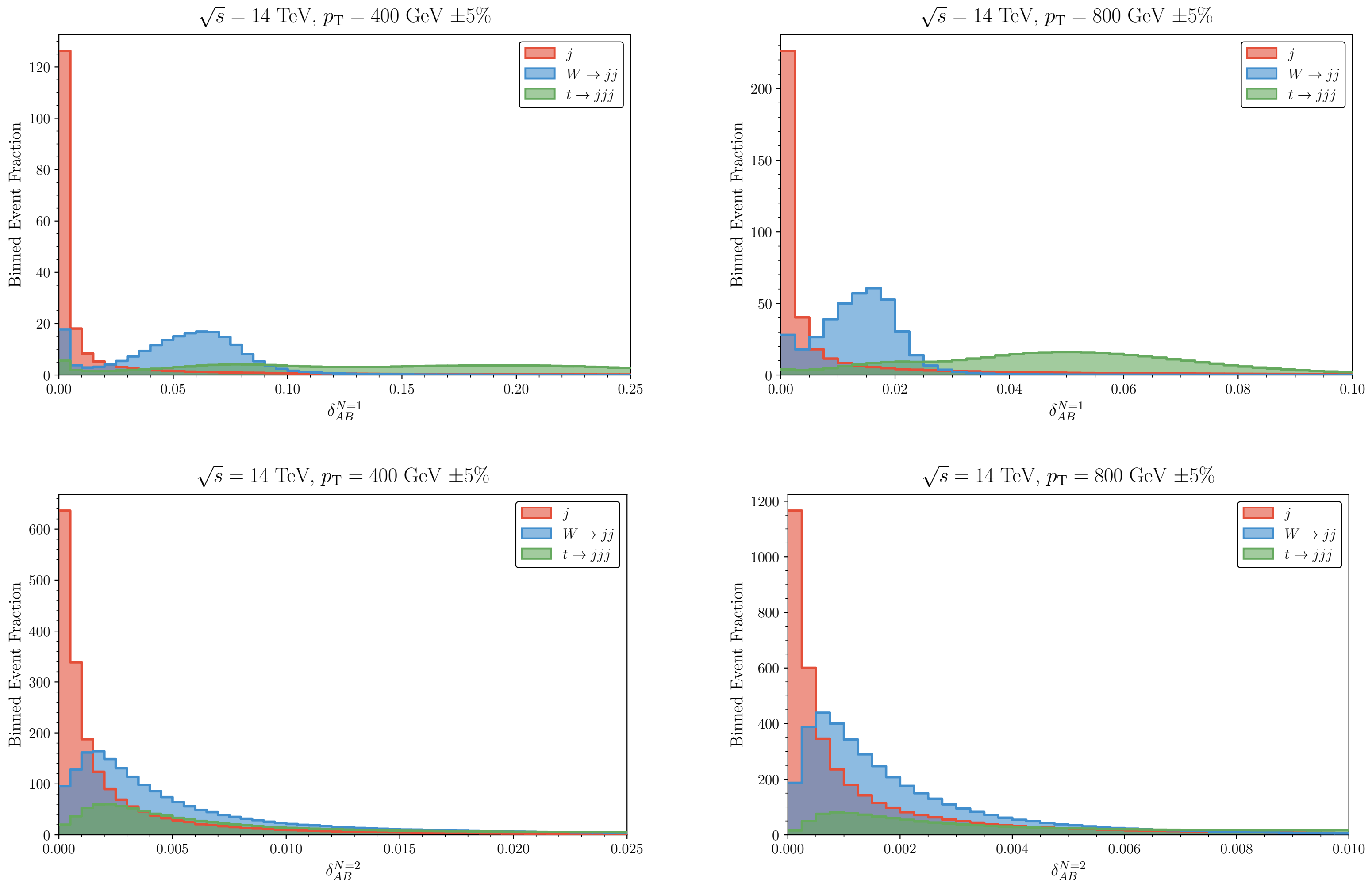






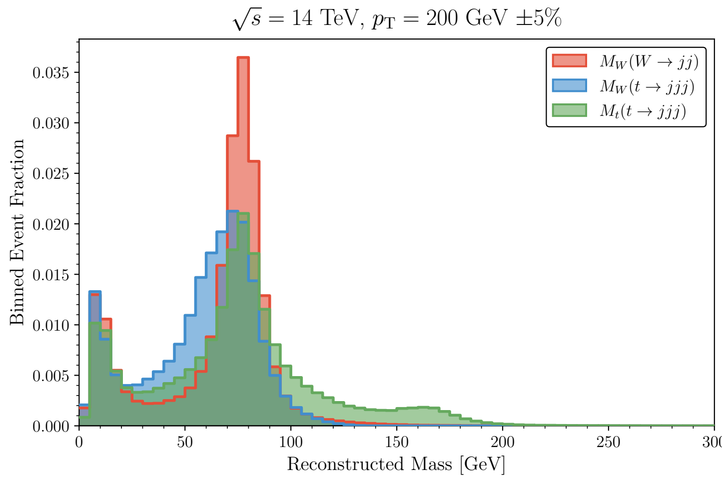
SIFT Measure at Final Mergers

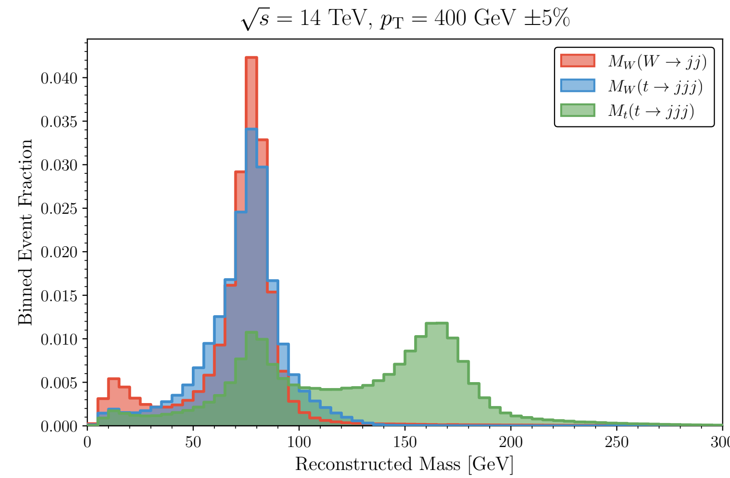
- We are also interested in whether the SIFT measure tracks jettiness DIRECTLY
- It seems not only to do so, but to excel specifically at large boost

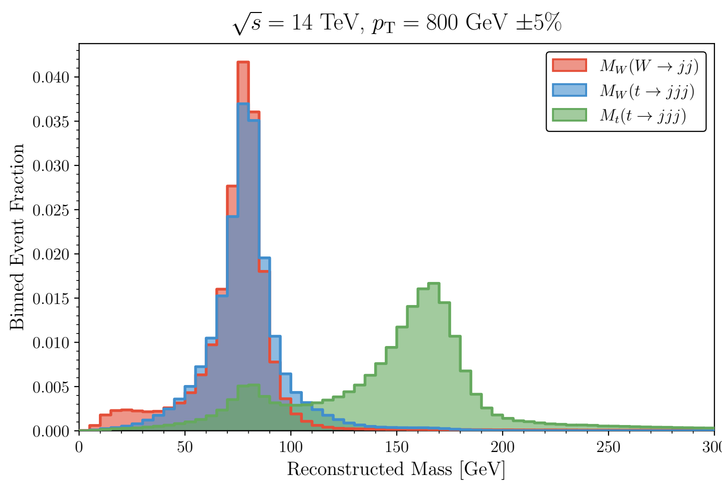


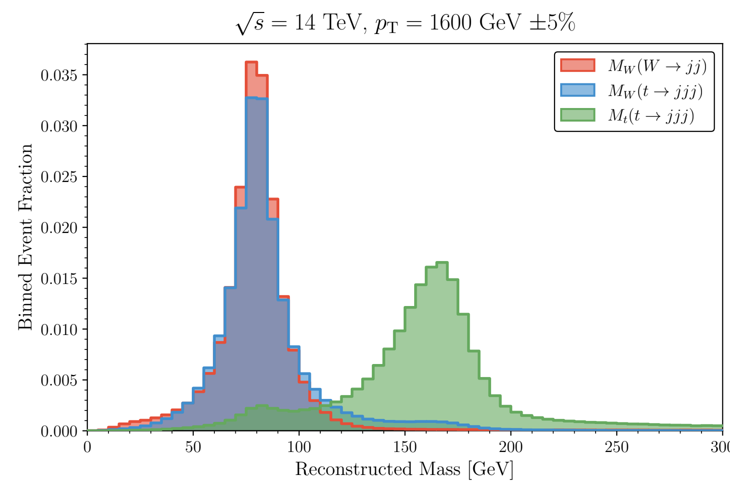
W & top Mass Reconstruction

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





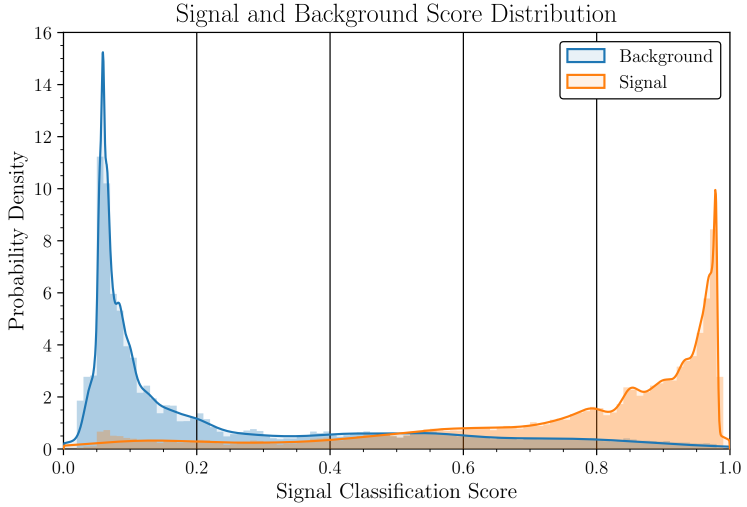




Assessing Performance

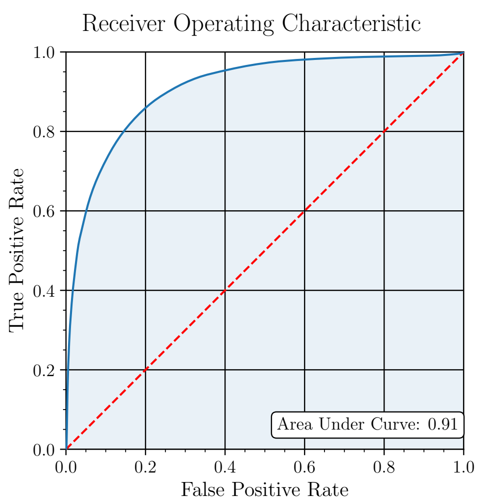
- A Boosted Decision Tree lets us compare information density in an unbiased way
- The BDT is also completely transparent, since it amounts simply to cascaded binary selection cuts (branchings) with assigned scores
- We feed the BDT Delphes N-subjettiness ratios up to 5/4
- We also provide it with the 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

1/2 and 3/2 Discrimination with BDT



$p_T^{\mathrm{GeV}\pm5\%}$	$\tau_{\mathrm{DELPHES}}^{N+1/N}$	$\tau_{\mathrm{SIFT}}^{N+1/N}$	δ_{AB}^{N}	$\delta+\tau$
100	0.62	0.68	0.69	0.70
200	0.91	0.86	0.88	0.89
400	0.89	0.85	0.91	0.92
800	0.82	0.79	0.92	0.93
1600	0.77	0.74	0.91	0.92
3200	0.78	0.76	0.88	0.90

TABLE III. Area under curve ROC scores for discrimination of resonances with hard 1- and 2-prong substructure using a BDT trained on various sets of event observables.



$p_T^{\mathrm{GeV}\pm5\%}$	$\tau_{\mathrm{DELPHES}}^{N+1/N}$	$\tau_{\mathrm{SIFT}}^{N+1/N}$	δ_{AB}^{N}	$\delta+\tau$
100	0.61	0.61	0.63	0.65
200	0.63	0.60	0.71	0.72
400	0.82	0.74	0.90	0.90
800	0.85	0.80	0.94	0.95
1600	0.77	0.77	0.97	0.97
3200	0.77	0.79	0.98	0.99

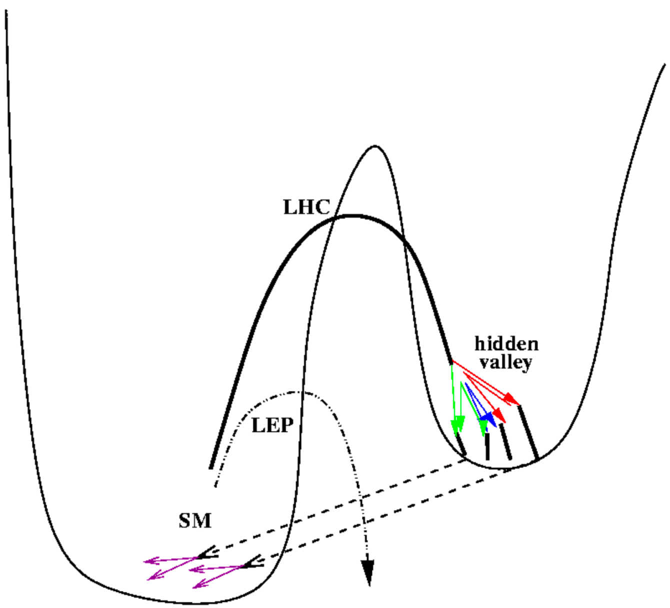
TABLE IV. Area under curve ROC scores for discrimination of resonances with hard 2- and 3-prong substructure using a BDT trained on various sets of event observables.

Application: Dark Sector Mass Reconstruction (Hidden Valley)

- The Hidden Valley Scenarios were described by Strassler and Zurek leading up to the start of collisions at the LHC (hep-ph/0604261)

" A unexpected place …
… of beauty and abundance …
… discovered only after a long climb … "

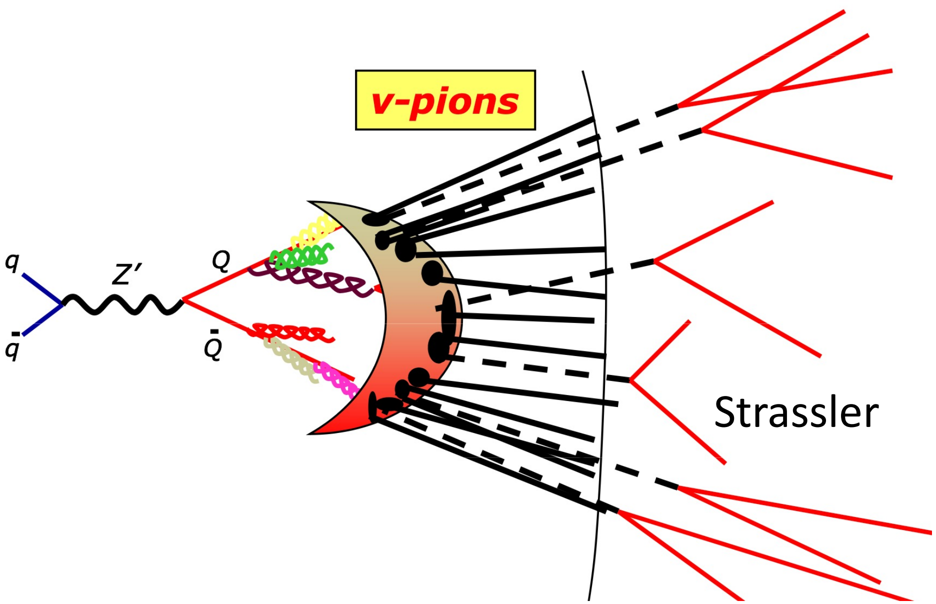
- Characterized by new light physics that is weakly coupled to the SM
- A heavy intermediary presents a high energy barrier to access the new sector
- Strong dynamics & confinement are typical
- A mass gap allows decays back to the SM



Strassler

Hidden Valley Strong Dynamics

- Classic signatures include a heavy dilepton resonance and/or displaced vertices
- We are interested here in a more challenging scenario (0806.2835 Strassler)
- The mediator is a few-TeV *Z'* coupled to the SM by kinetic mixing
- Heavy v-Quarks are pair produced and they shower / hadronized
- Flavor-diagonal pions (10's to 100's of GeV) can decay back to the SM and shower / hadronize AGAIN … helicity-suppression favors b's, taus
- Off-diagonal pions (SM NEUTRAL!!) are stable (DM candidates) … the result is semi-visible jets



Strassler

The Combinatoric Problem

- Mass is accessible if v-Pions are isolated and decay to 1 thick or 2 thin jets
- However, jet definitions & analysis have to be tuned to cross regimes
- As the count of proximal Pions increases, a severe combinatoric BG emerges

To identify this signal, it seems likely that tagging of individual jets is not enough. By definition, the number of heavy-flavor-tagged jets cannot be larger than the number of jets. But the number of B mesons can greatly exceed the number of tagged jets, as suggested in Figs. 16 and 17. In other words, although these events do not have an exceptional number of taggable jets, often four or less in the A cases, they do have an unusual number of B mesons. *Thus to distinguish the signal from background, it is essential to detect as many vertices from the B mesons as possible.*

Simply plotting dijet invariant masses, where the jets are selected at random, cannot reveal the v-pion resonance. The huge combinatoric background, the fact that many jets contain multiple b-quarks, and relatively poor resolution for jet momentum and energy would eliminate any signal.

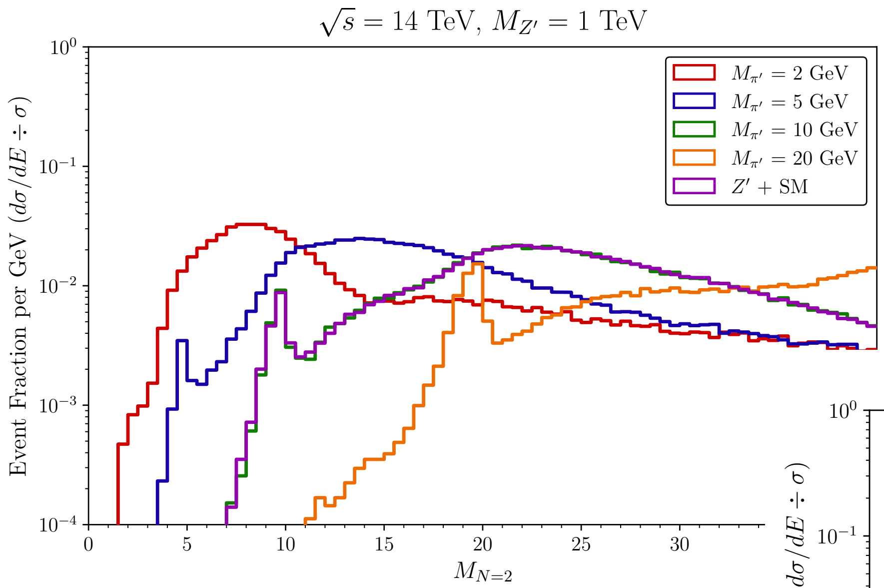
SIFT-ing for Dark Matter

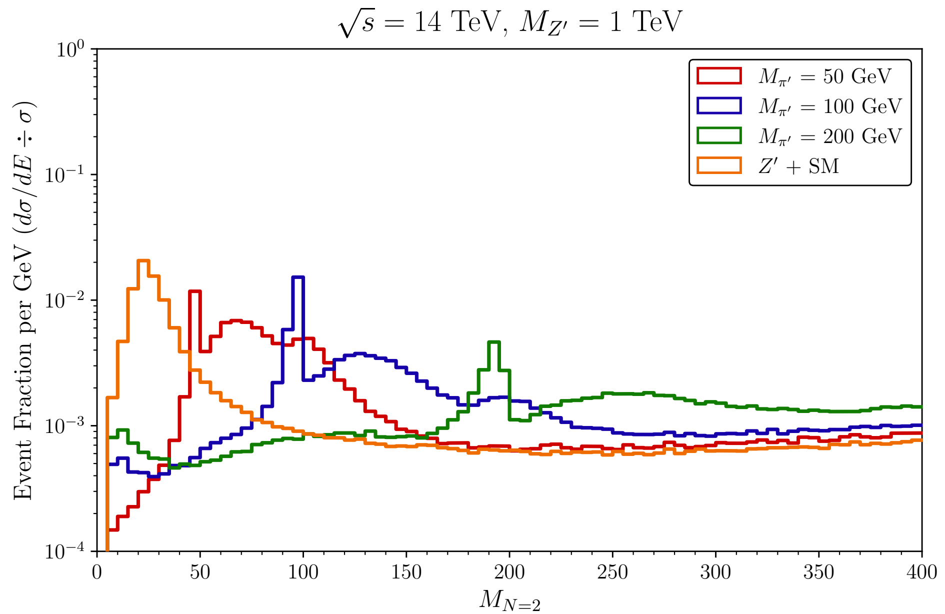
- SIFT, with filtering but without dropping, may be ideal here
- It considers the event as a whole (no cones) with multi-scale sensitivity
- It creates a well-defined sequential SLICE through the combinatorics
- Since hard prongs are merged last, the final mergers are expected to hold relevant physical masses
- We can look for resonances in the distribution of the mass for the Nth pair of merged objects …

> It is conceivable that
> the v-pion resonance can be better identified with a more sophisticated variable than single jet mass, looking more carefully at the substructure of the jets. (It is even possible that, with so many v-pions per event, and with a bit more statistics than available here, the v-pion can be discovered through its rare tree-level decay to muon pairs or its loop-induced decay to photon pairs.) More generally, it is important to study further how best to look for resonances in very-high-multiplicity signals, such as case B1.

A Proof of Concept

- We simulate with Pythia8, omitting ISR and detector effects for a first trial





For Continuing Work …

- We need to develop CUTS to isolate the signal prior to extracting masses
- SUBSTRUCTURE is key here & the SIFT measure with machine learning may work
- We need to evaluate visibility over backgrounds for benchmark models
- We need to carefully compare results with traditional clustering approaches
- We need to include ISR, detector effects, and pileup
- We need to look at setting expected limits

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
- There are a great variety of potential applications, including SIFT-ing the Dark Sector

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