Jet SIFT-ing

A Scale-Invariant Jet Clustering Algorithm for the Substructure Era

Joel W. Walker
Sam Houston State University

a work in progress
with
Andrew Larkoski (Reed), Denis Rathjens (CMS), and Jason Veatch (ATLAS)

Collider, Dark Matter, and Neutrino Conference
Mitchell Institute for Fundamental Physics and Astronomy
May 21-23, 2018

SIFT: Scale-Invariant Filter Tree

Outline of Presentation

- SCALE INVARIANT
  - Jet Clustering Background
  - Motivation for Scale Invariance
  - Algorithm Implementation
  - Algorithm Visualization
  - Algorithm Testing
- FILTER
  - Integrated Grooming
  - Remove Soft Co-Linear Radiation
  - A Natural Halting Condition
- TREE
  - Fast Algorithms
  - Multidimensional Trees

SIFT: SCALE-INVARIANT Filter Tree

- Traditional Jet Clustering imposes a fixed cone size, and thus a fixed scale on events
- Boosted objects tend to collimate and fall into a single jet radius
- Substructure techniques are essential for recovering information inside the jet
- However, these techniques are often complicated, with de- and re-clustering

- We propose as SCALE INVARIANT approach which is intrinsically suitable for tagging substructure AS the jet is being assembled

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)

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

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

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

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

Pseudorapidity η vs. Polar Angle θ

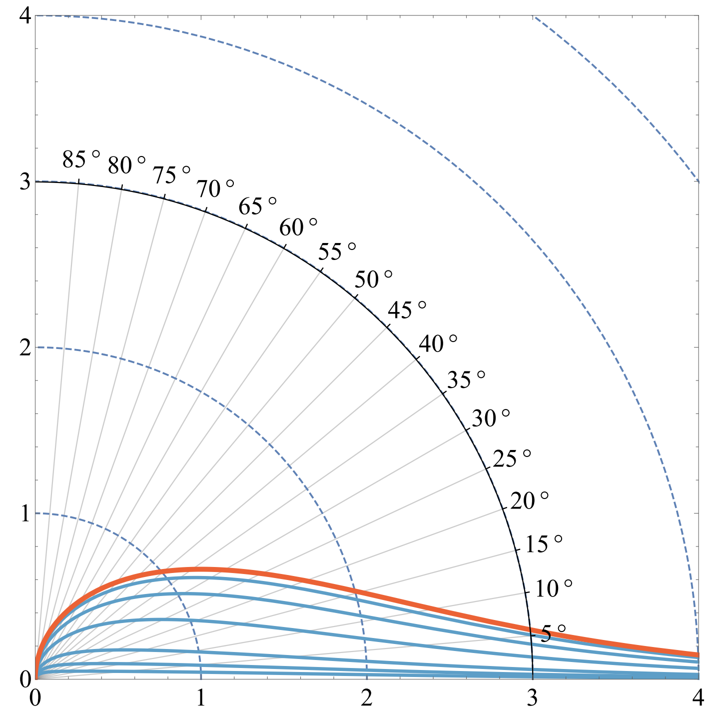


FIG. 1: The pseudorapidity η (bold, orange) is plotted as a function of the polar angle θ. For comparison, the longitudinal rapidity y (fine, blue) is also shown for various values of M/P_{T}, equal to $\{1/2, 1, 2, 5, 10, 20\}$ from top to bottom.

Formation of Hadronic Jets

- The hard partonic event may result in the production of colored objects (at Feynman diagram level, e.g. MadGraph)
- These objects rapidly ”shower”, radiating quarks & gluons (e.g. Pythia)
- QCD confinement implies that strongly charged particles cannot exist as free objects at large separations; they must convert “hadronize” (e.g. Lund color strings in Pythia) into color-neutral particles such as pions, K mesons, etc.
- Color strings may convolve descendants of partonic objects with each other and even with the underlying beam; this is partially mitigated in a lepton collider

3.25

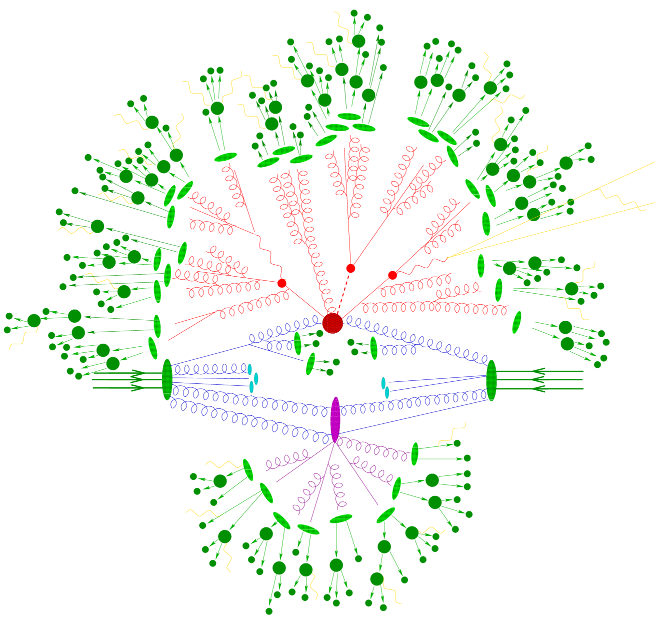


Image: Stefan Höche

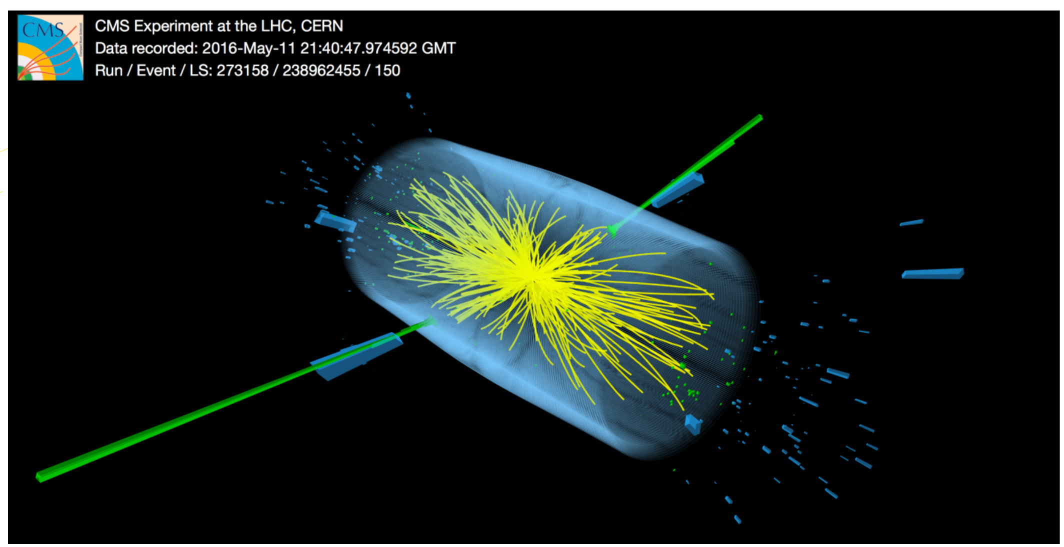


Image: CMS

Standard Jet Clustering Algorithms

- Hadronized objects need to be recombined 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 more widely separated than R_0 will never be clustered
- $n = 0$, or "Cambridge/Aachen" clusters objects with high angular adjacency
- $n = +1$, or "kT" additionally favors clustering of soft pairs first
- $n = -1$, or "Anti-kT" prioritizes clustering where one of the pair is hard
- Anti-kT is now the default jet clustering tool at LHC, with $R_0 = 0.4$
- It is robust against "soft" and "colinear" jet perturbations and has regular jet shapes which are favorable for calibration against pileup, etc.

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

Jet Substructure

- Highly boosted mothers will tend to yield very collimated daughters
- In hadronic top quark decays t $\Rightarrow$ W/b $\Rightarrow$ u/d/b with COM energy above a TeV, the likelihood of resolving only 2 or even 1 discrete object increases
- For example, within, a “fat” (large $R_0 \gtrsim 1$), N-Subjettiness τ_N can characterize how well the 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
- Variable cone sizes have also been considered to cope with loss of structure

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

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

A Scale-Invariant Jet Algorithm

- It may be 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 the favorable behavior of Anti-kT
- Numerator should favor angular collimation; we propose ΔM^2, similar to JADE
- Denominator should suppress soft pair clustering; we propose a sum of E_T
- Result is dimensionless, Lorentz invariant (longitudinally in the denominator), and free from references to external / arbitrary scales

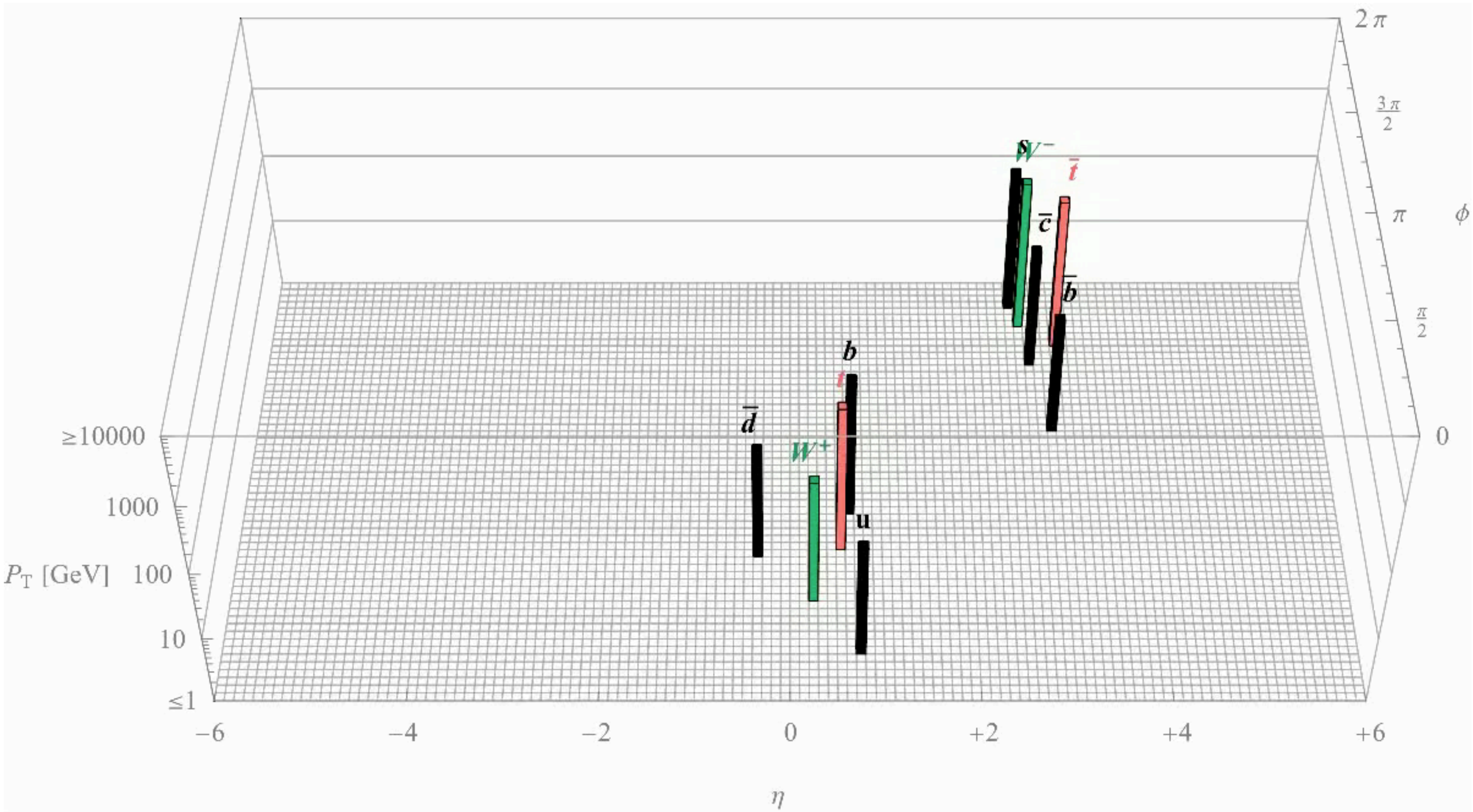
$$\delta_{AB} \equiv \frac{\Delta M^2_{AB}}{E^2_{\mathrm{T}A} + E^2_{\mathrm{T}B}}$$

$$\begin{aligned}
M^{\mathrm{A,B}} &\equiv \sqrt{\left(P^{\mathrm{A}}_\mu + P^{\mathrm{B}}_\mu\right)\left(P^\mu_{\mathrm{A}} + P^\mu_{\mathrm{B}}\right)} \\
&= \sqrt{M^2_{\mathrm{A}} + M^2_{\mathrm{B}} + 2\left(E^{\mathrm{A}}E^{\mathrm{B}} - \vec{P}^{\mathrm{A}} \cdot \vec{P}^{\mathrm{B}}\right)} \\
\lim_{M_{\mathrm{A}}=M_{\mathrm{B}}=0} &\Rightarrow \sqrt{2\,|\vec{P}^{\mathrm{A}}||\vec{P}^{\mathrm{B}}|\left(1 - \cos\Delta\varphi^{\mathrm{B,A}}\right)}
\end{aligned}$$

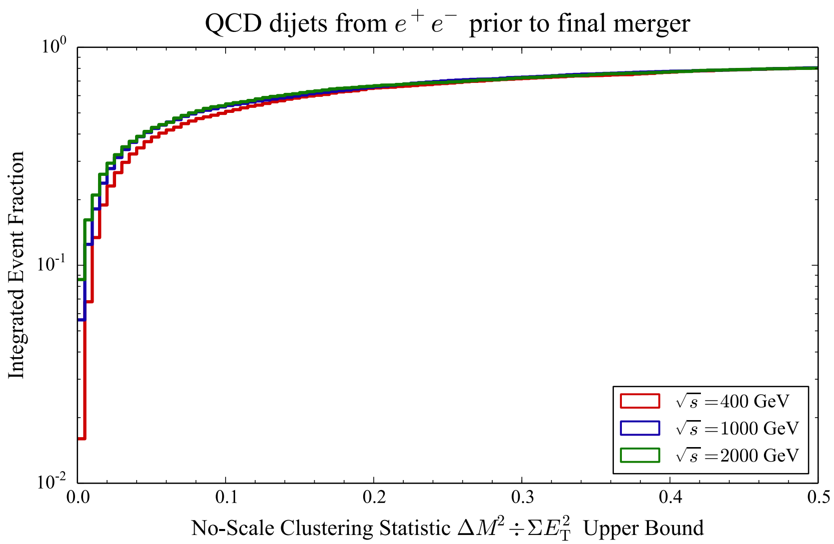
$$\begin{aligned}
E_{\mathrm{T}} &\equiv \sqrt{M^2 + \vec{P}_{\mathrm{T}} \cdot \vec{P}_{\mathrm{T}}} = \sqrt{E^2 - P^2_z} \\
\lim_{M=0} &\Rightarrow |\vec{P}_{\mathrm{T}}|
\end{aligned}$$

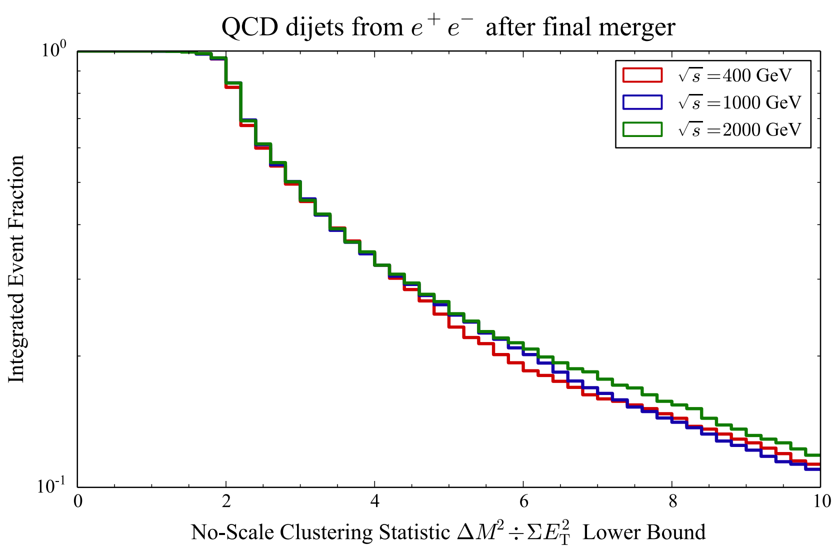
Hadronic TTbar Scale-Invariant Clustering

https://youtu.be/u9Z4qDuXL84



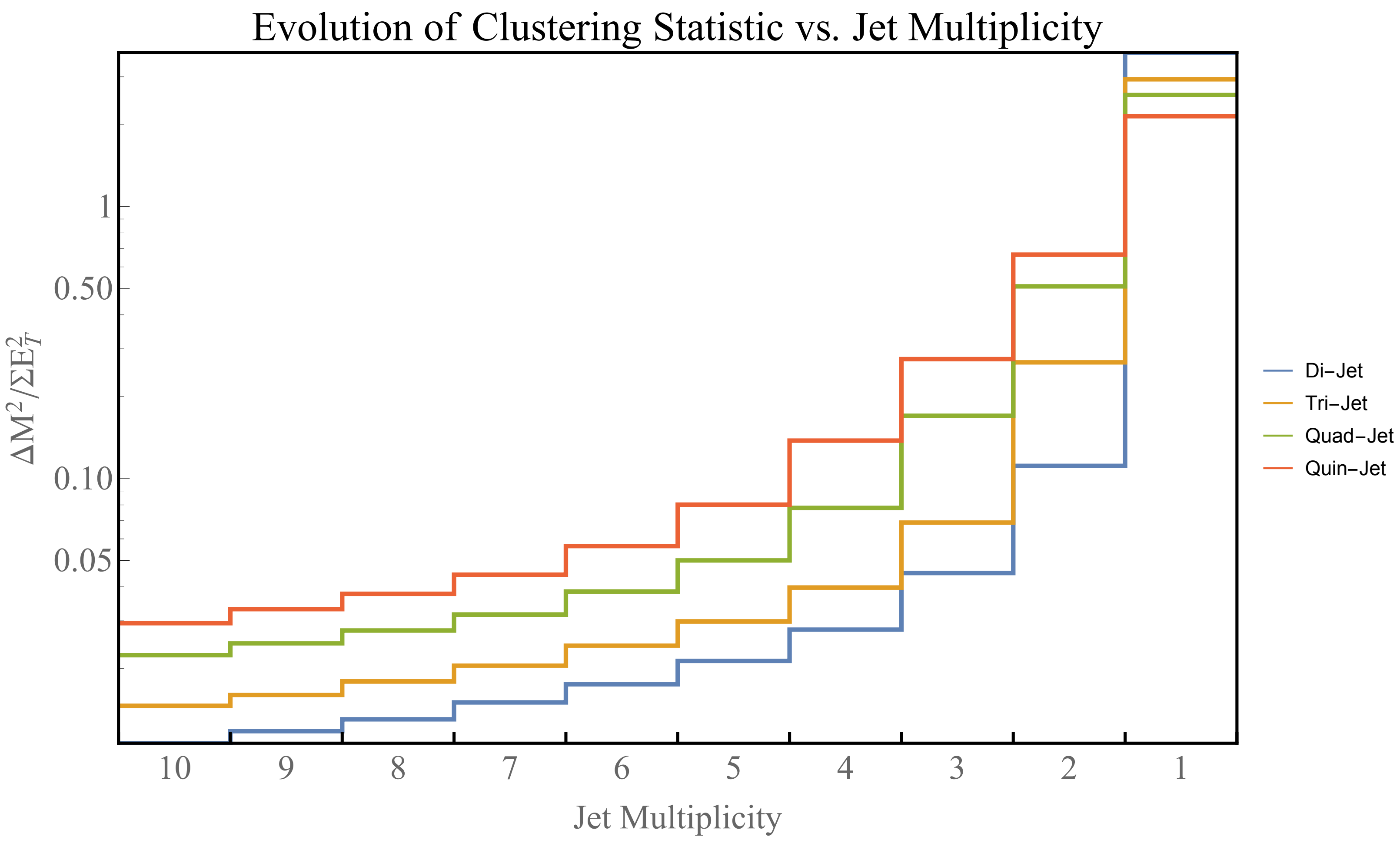
Test of Pre/Post Merger Statistic for Di-jets





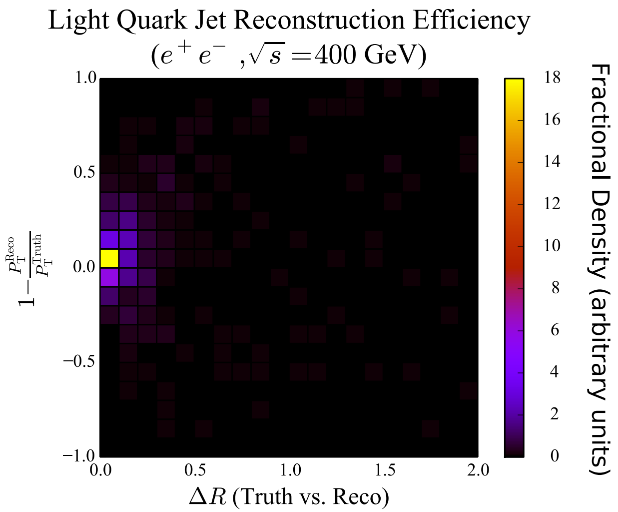
- 95% of pairs reconstructed prior to 0.1
- 95% of final final mergers are after 2.0
- Results are invariant wrt beam energy

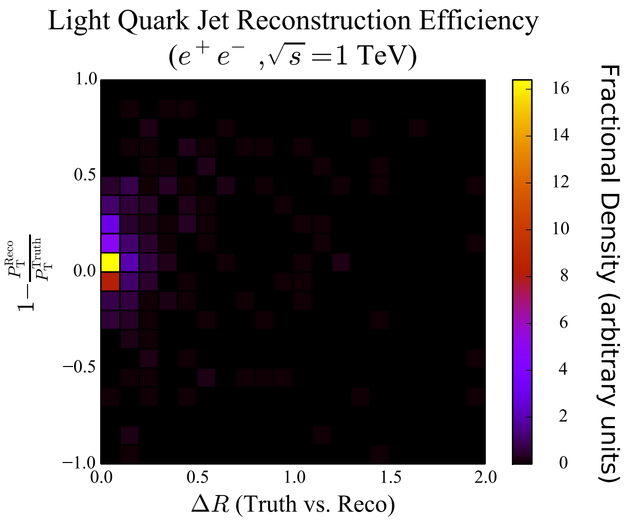
Visualization of Statistic Jump at Clustering

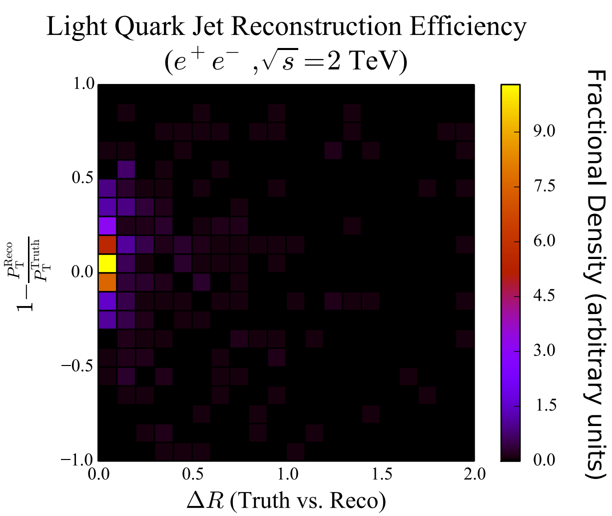


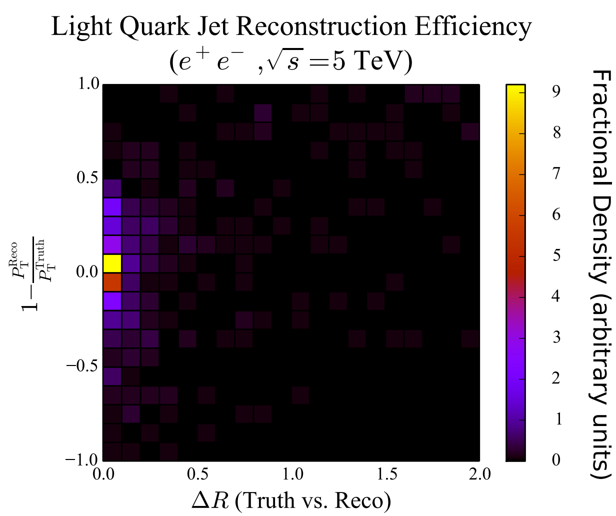
- The event jettiness count is intrinsically imprinted on the clustering history

Matching of final 6 objects with Truth-Level Quarks



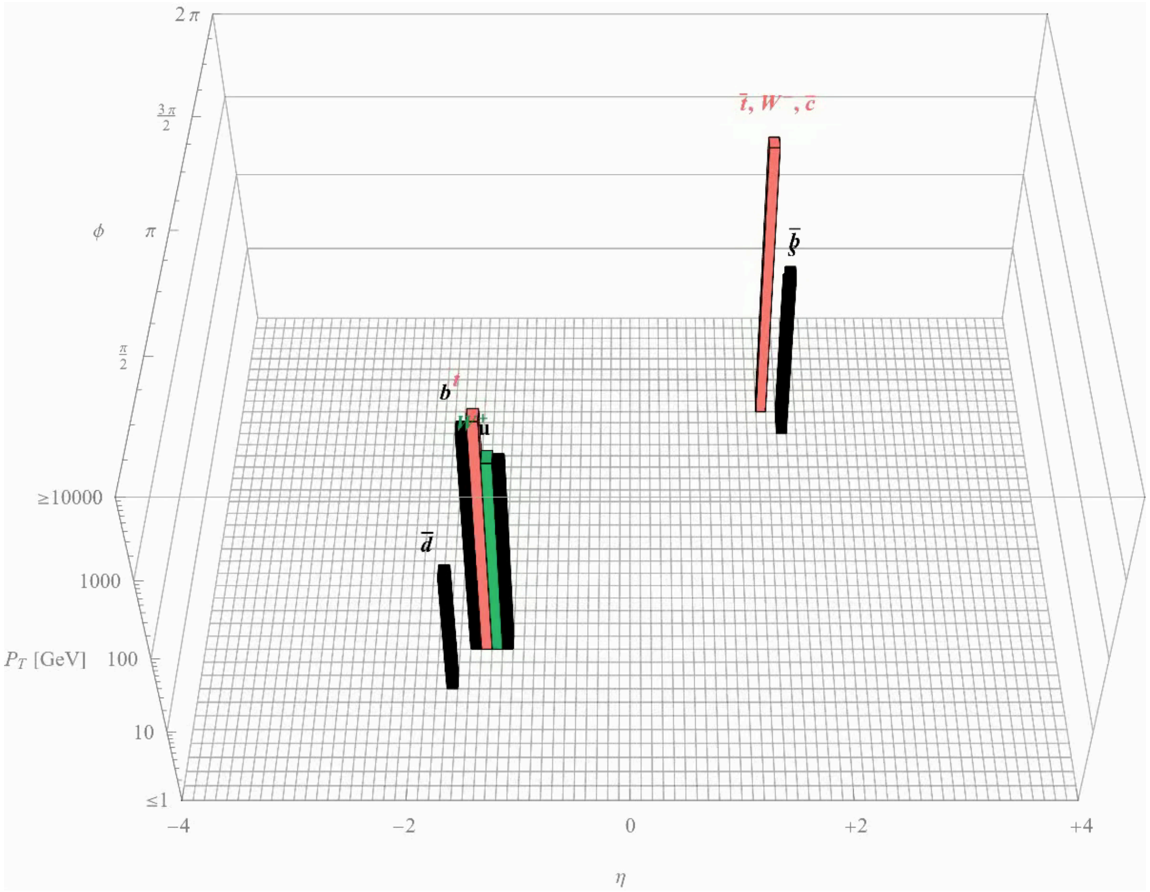






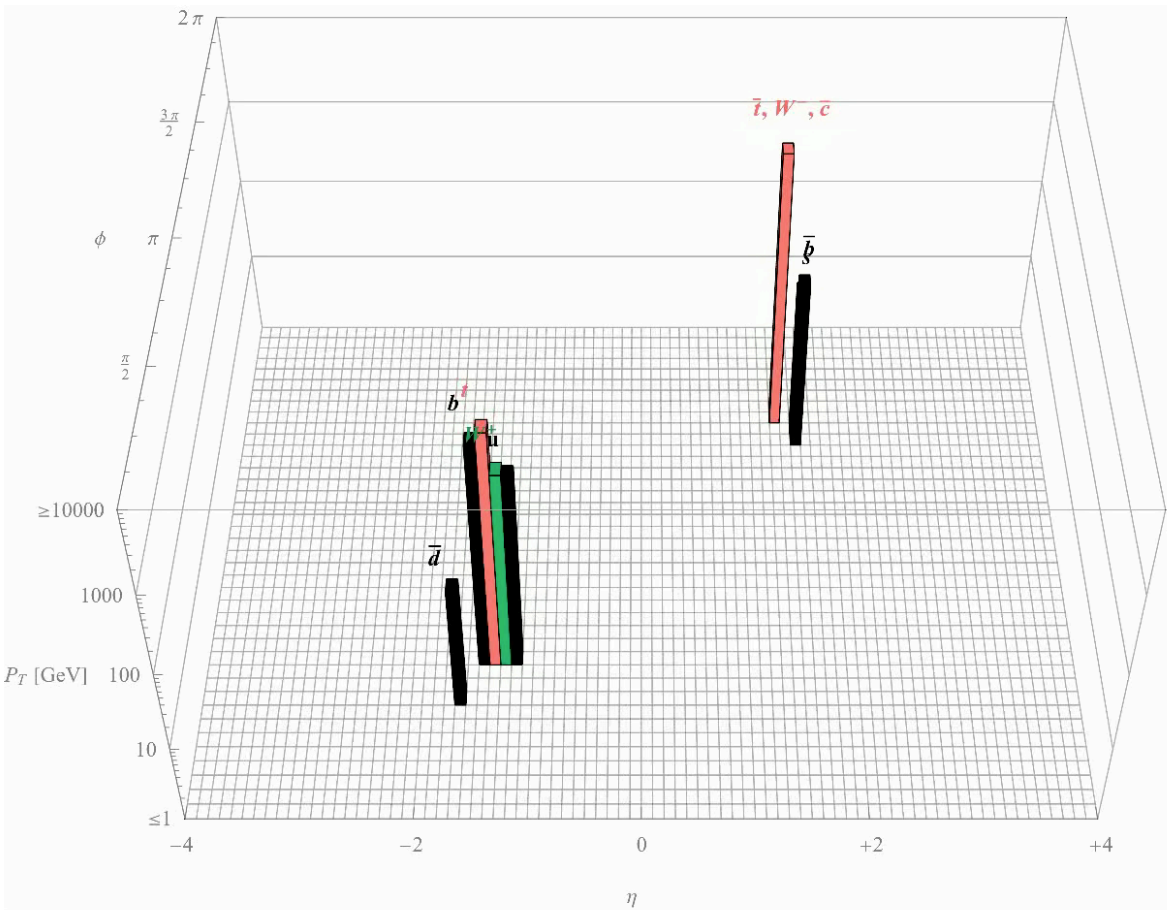
Lepton to TTbar 2.5 TeV Anti-KT 0.5 with Ghosts

https://youtu.be/1fhbhlDrORA



Lepton to TTbar 2.5 TeV Scale Invariant Clustering with Ghosts

https://youtu.be/kxUmgv1HHMs



SIFT: Scale-Invariant FILTER Tree

- Running to termination can lead to merging of stray radiation
- Take a cue from "Soft Drop" (2014 Larkoski, Marzani, Soyez, Thaler)
- This procedure "Grooms" a jet by removing soft, wide-angle radiation to mitigate contamination from ISR, UE, and pileup
- SD iteratively DECLUSTERS C/A, dropping softer object unless & until:

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

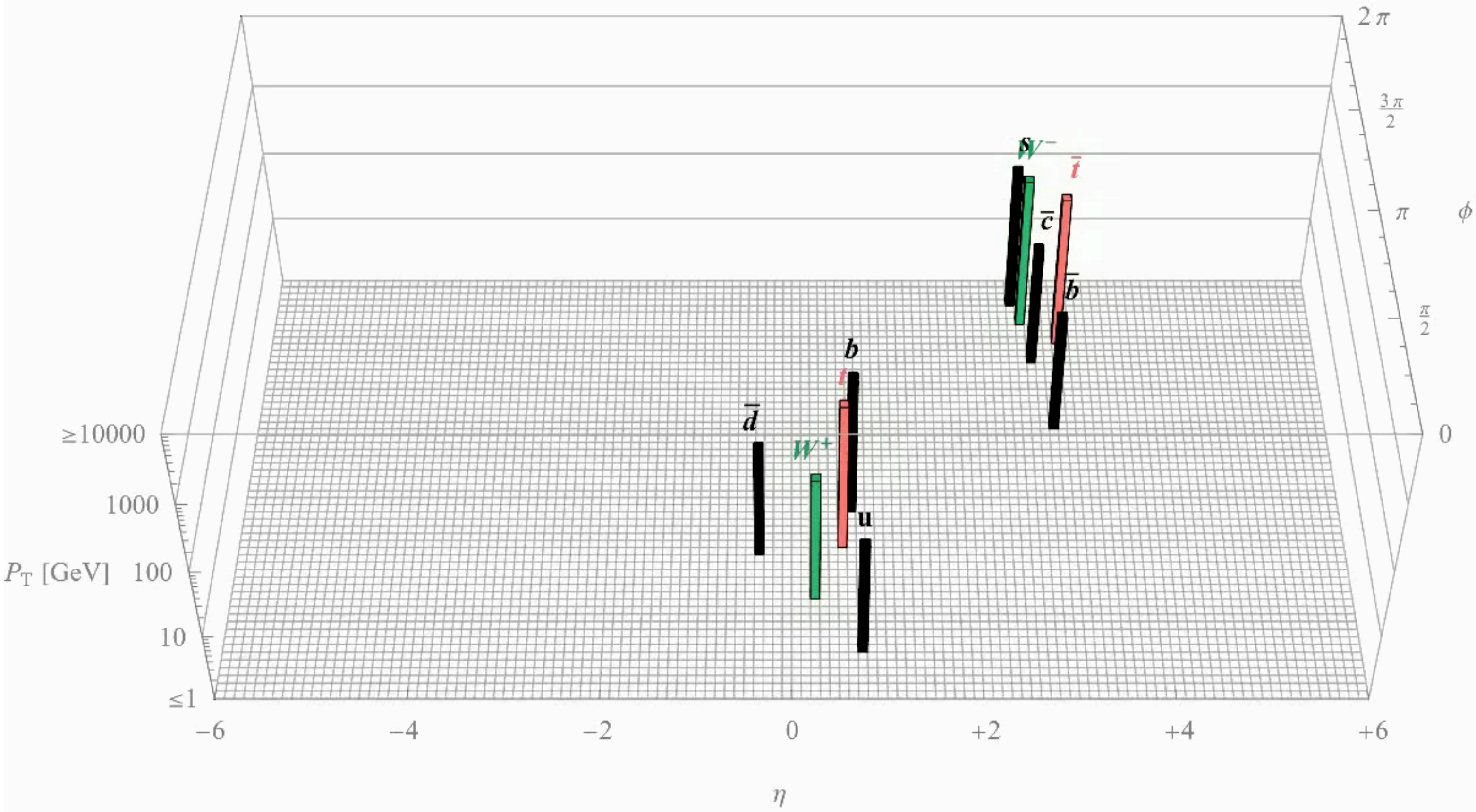
- Typically, z_{cut} is $\mathcal{O}(0.1)$, and $\beta > 0$ for grooming
- We propose a scale-invariant analog which is applied within the original clustering itself.

$$\frac{E_{TA} E_{TB}}{E_{TA}^2 + E_{TB}^2} > \frac{\Delta M_{AB}^2}{2 E_{TA} E_{TB}} \implies \delta_{AB} \equiv \frac{\Delta M_{AB}^2}{E_{TA}^2 + E_{TB}^2} < \frac{2 E_{TA}^2 E_{TB}^2}{(E_{TA}^2 + E_{TB}^2)^2}$$

- The softer object is considered isolated unless it passes this FILTER
- This provides a natural halting condition to prevent total assimilation
- Curiously, the dynamic threshold is symmetric under $E_T \to 1/E_T$

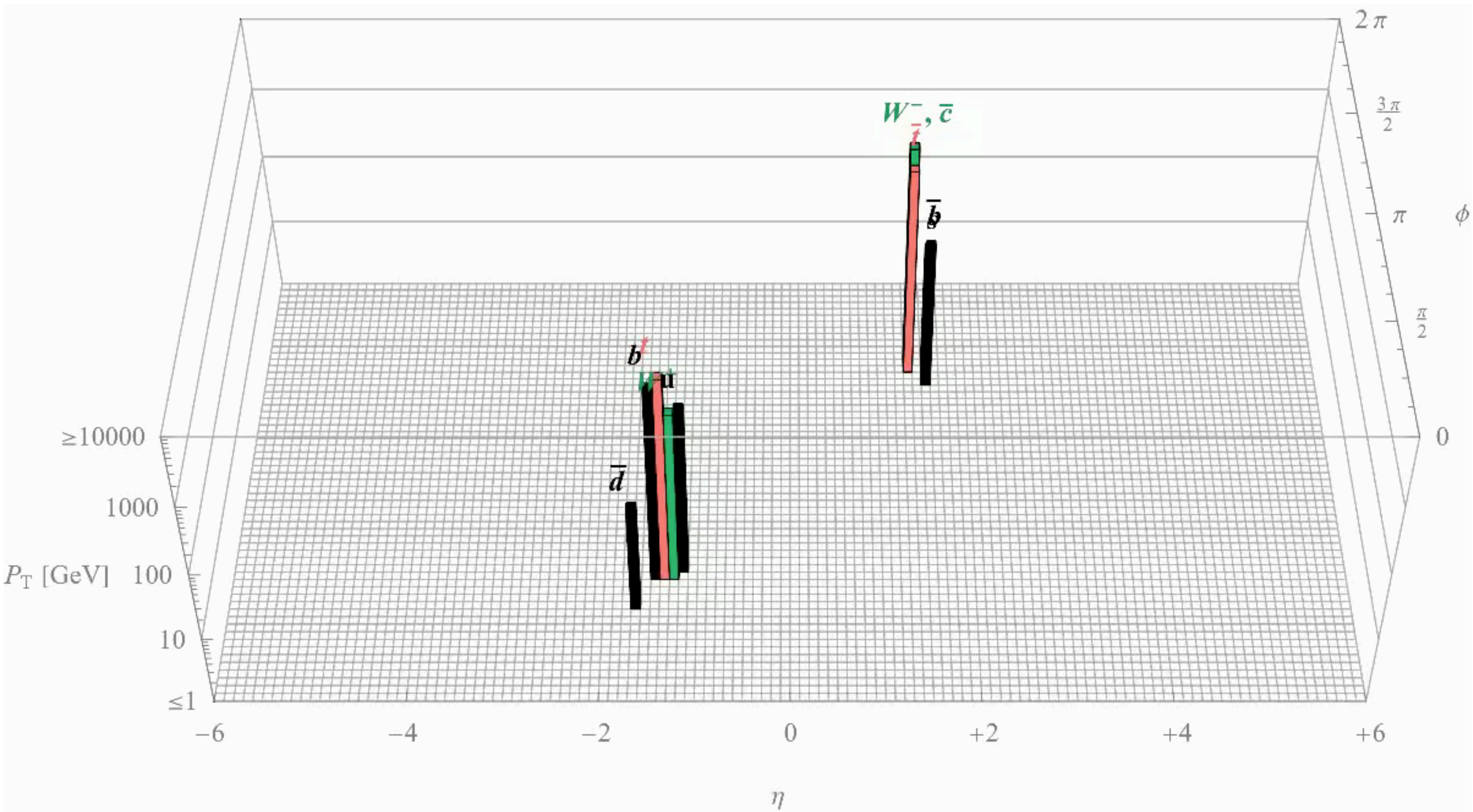
Hadronic TTbar Scale-Invariant Clustering with Filtering

https://youtu.be/rDsBeEBTimw



Lepton to TTbar 2.5 TeV SIFT Filtered Clustering with Ghosts

https://youtu.be/G1XB5sQaolk



SIFT: Scale-Invariant Filter TREE

- A jet clustering algorithm is USELESS practically unless it is FAST
- Critical issue is the scaling dimension with number *N* of constituents
- A naïve implementation is CUBIC $\mathcal{O}(N^3)$ because there are *N* mergers with a scan over *N* x *N* possible pairings at each stage. TOO SLOW!
- Why is FastJet (Cacciari, Salam, Soyez) FAST?
- FJ Lemma trims to $\mathcal{O}(N^2)$ by scanning only GEOMETRIC nearest neighbors
- How? The magic of "min of a min" facilitates factorization
- GLOBAL min of δ_{AB} has the property that *B* minimizes ΔR_{AB} if $P_{TA}^{2n} < P_{TB}^{2n}$

$$\delta_{AB} \equiv \min(P_{TA}^{2n}, P_{TB}^{2n}) \times \left(\frac{\Delta R_{AB}}{R_0}\right)^2$$

- Then, with a FAST $\mathcal{O}(\log N)$ algorithm for caching neighbors, the combined runtime can be "linearithmic" $\mathcal{O}(N \log N)$. GOLD STANDARD!
- Signature of $\mathcal{O}(\log N)$ algorithms is halving of problem size with each cycle
- Example is "bisection" method of traversing a sorted list
- The FAST approach to finding nearest neighbors can use a TREE

Can SIFT be FAST?

- If yes, there needs to be something like a "GEOMETRIC" measure
- As originally expressed, the metric is not even written in terms of coordinates
- For massless *A* & *B*, $\Delta M_{AB}^2 = 2P_A^\mu P_\mu^B \Rightarrow 2P_A P_B(1 - \cos\Delta\theta) \approx P_A P_B(\Delta\theta^2 - \Delta\theta^4/12)$
- But, we need to refer to the collider coordinates of *A* & *B* directly ($\Delta\eta_{AB}, \Delta\phi_{AB}$, etc.)
- Conjecture: for massive *A* & *B*, it will actually be Δy_{AB} that is relevant
- Boost from the $P_z = 0$ frame into the lab:

$$\begin{pmatrix} E \\ P_z \end{pmatrix} = \begin{pmatrix} \cosh y & \sinh y \\ \sinh y & \cosh y \end{pmatrix} \begin{pmatrix} E_T \\ 0 \end{pmatrix} = \begin{pmatrix} E_T \cosh y \\ E_T \sinh y \end{pmatrix}$$

$$\begin{aligned} 2P_A^\mu P_\mu^B &= 2(E_A E_B - P_z^A P_z^B - P_T^A P_T^B \cos\Delta\phi_{AB}) \\ &= 2(E_T^A E_T^B [\cosh y^A \cosh y^B - \sinh y^A \sinh y^B] - P_T^A P_T^B \cos\Delta\phi_{AB}) \\ &= 2(E_T^A E_T^B \cosh\Delta y^{AB} - P_T^A P_T^B \cos\Delta\phi_{AB}) \end{aligned}$$

- We are getting WARM. BUT the difference between E_T & P_T (i.e. MASS) means that we CANNOT perfectly factorize kinematics from geometrics
- Nevertheless, we can proceed. BUT, we must seek neighbors in a 3D or 4D space
- The FastJet engine (Voronoi Tesselation) is 2D. We need a custom engine.
- NOTE: hyperbolic cosine differs from cosine in that all Taylor terms are POSITIVE

Building an D-Dimensional Tree

- "Balanced KD-Tree" framework (2003 Procopiuc, Agarwal, Arge, Vitter) is suitable
- The forking property of a tree allows $\mathcal{O}(\log N)$ traversal
- Each descending "row" of the tree sorts on the next cyclic coordinate index
- To stay "balanced" we never add objects to a tree after initial construction
- We maintain a "forest" of trees of doubling size, as needed
- Protocols for pruning, grafting, and merging leaves must be built in
- Be sure to not reinject $\mathcal{O}(N^2)$ scaling in these updates. Non-Trivial!
- Protocols for neighbor finding under a user defined metric must be built in
- Use "templating" to allow input from user-defined data structures
- Cyclic indices: extend by half principal domain either way & build "image" leaves
- Status: working D-dimensional $\mathcal{O}(N \log N)$ implementation exists / tested on Anti-kt
- Currently, this is being ported to C++ for increased speed in the "coefficient"

Conclusions and Ongoing Work

- SIFT is a SCALE INVARIANT clustering algorithm designed specifically for substructure
- FILTER-ing of soft and co-linear radiation can be done as the jet is clustered
- Organization of the data structure in a balanced TREE can make clustering fast

- The clustering history holds *information* – it may be better to not halt at fixed radius.
- Could the algorithm be applied to existing fat jets for exclusive clustering?
- What is the jet-energy resolution width, and does it vary with P_T?
- How does SIFT fare with pileup subtraction?
- How does the absolute mass of reconstructed particles connect?
- Is the distilled clustering history amenable to machine learning applications?
- Can SIFT intrinsically confront the problem of tagging boosted objects?

Thank You

(movie notebook available upon request to jwalker@shsu.edu)