A new Scale-Invariant Jet Algorithm for the Substructure Era

Joel W. Walker
Sam Houston State University

a work in progress
with special thanks to
Andrew Larkoski and Denis Rathgens

PHENO
University of Pittsburgh
May 9, 2017

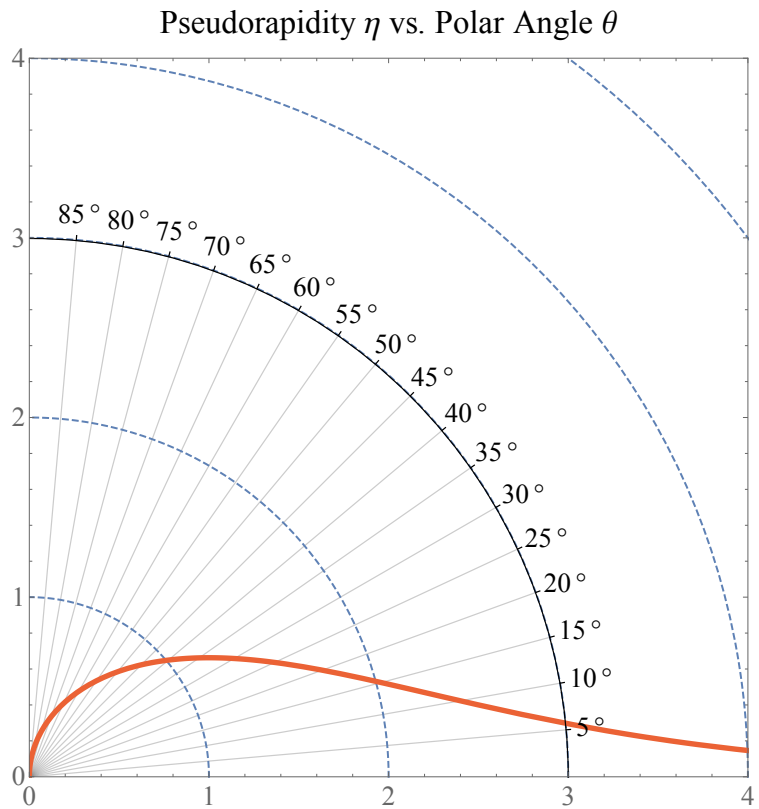
Collider Variables & Coordinates

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

$$\eta \equiv -\ln\tan(\theta/2)$$

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

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



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

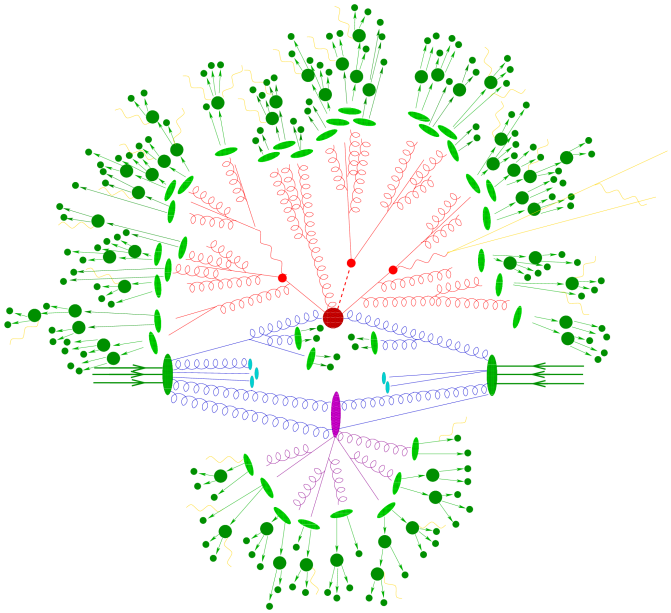


Image: Stefan Höche

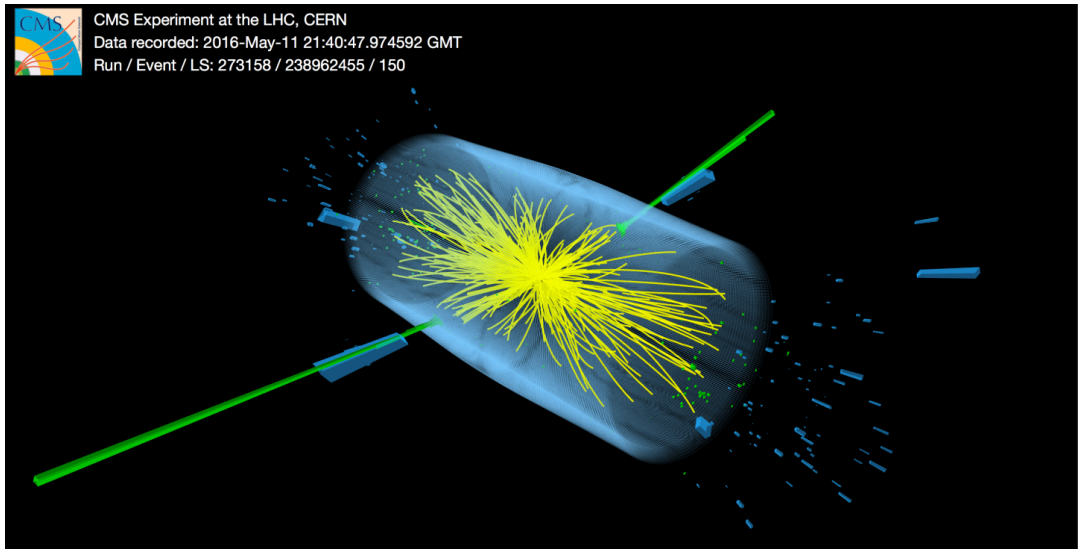


Image: CMS

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 member of the pair is hard
- Anti kT is now the default jet clustering tool at LHC, with R_0 = (0.4,0.5)
- It is robust against "soft" and "colinear" jet perturbations and has regular jet shapes which are favorable for calibration against pileup, etc.

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

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

$$\frac{M_{AB}^2 - M_A^2 - M_B^2}{\sum_{A,B} E_{\mathrm{T}i}^2}$$

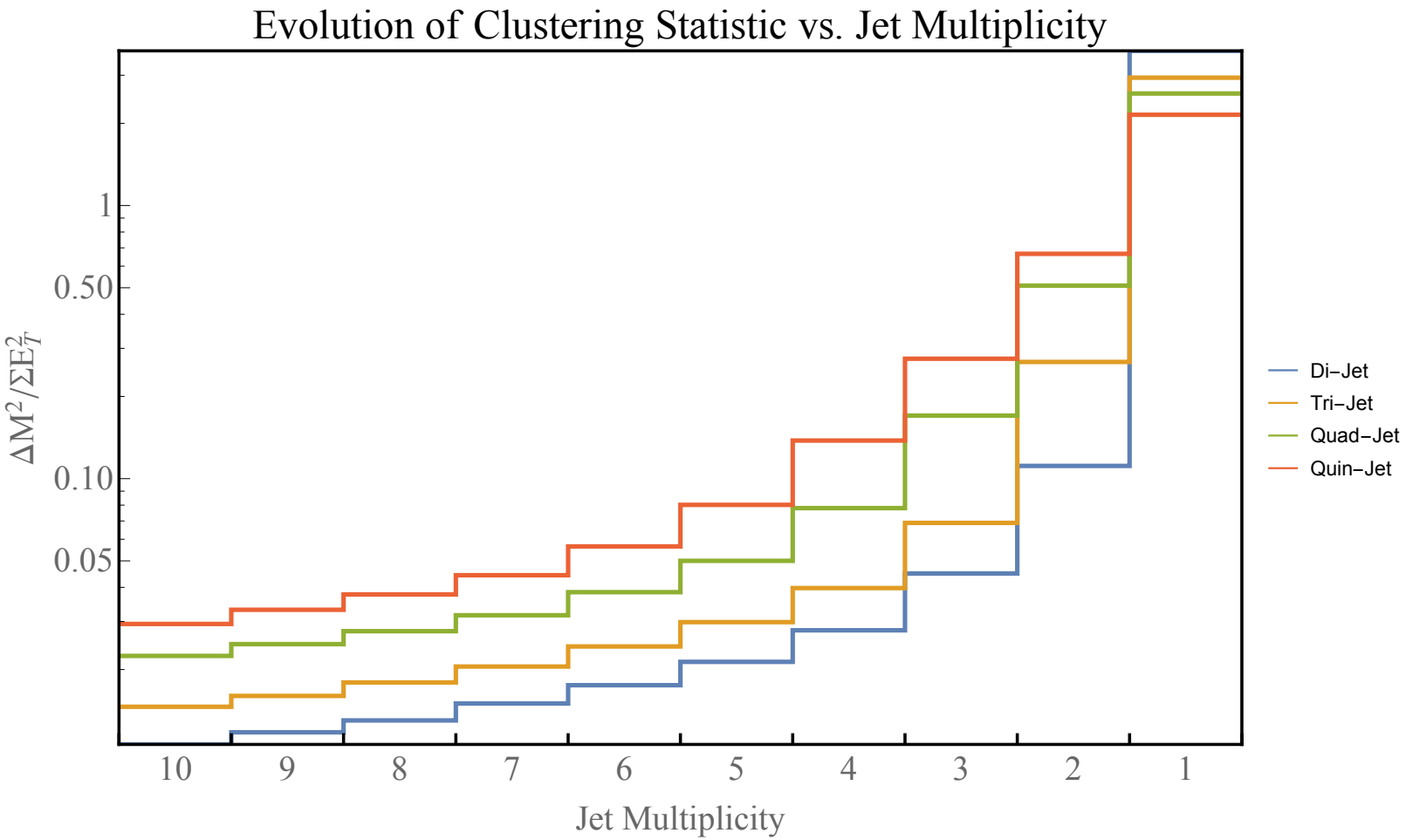
$$\begin{aligned} M^{\mathrm{A,B}} &\equiv \sqrt{\left(P_\mu^{\mathrm{A}} + P_\mu^{\mathrm{B}}\right)\left(P_{\mathrm{A}}^\mu + P_{\mathrm{B}}^\mu\right)} \\ &= \sqrt{M_{\mathrm{A}}^2 + M_{\mathrm{B}}^2 + 2\left(E^{\mathrm{A}}E^{\mathrm{B}} - \vec{P}^{\mathrm{A}} \cdot \vec{P}^{\mathrm{B}}\right)} \\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

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

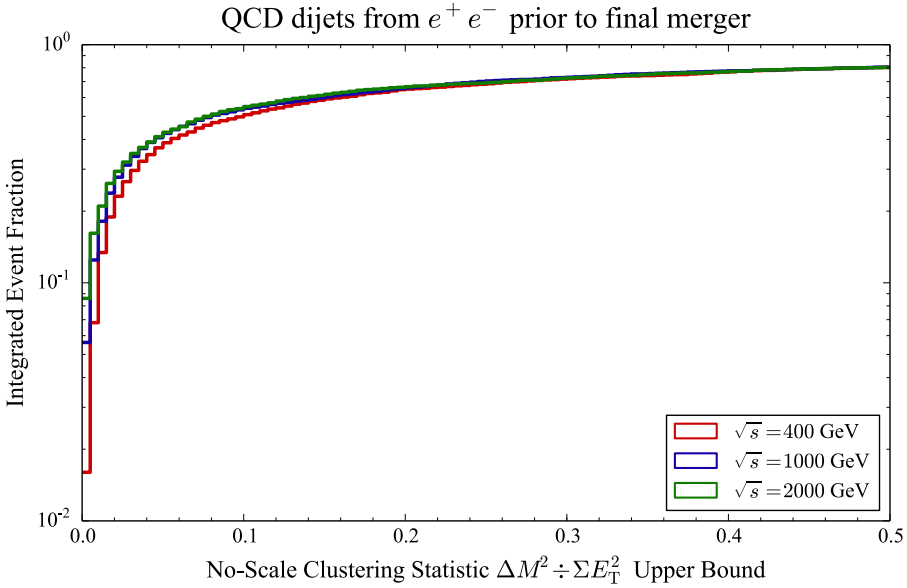
Clustering Visualization (Dijet)

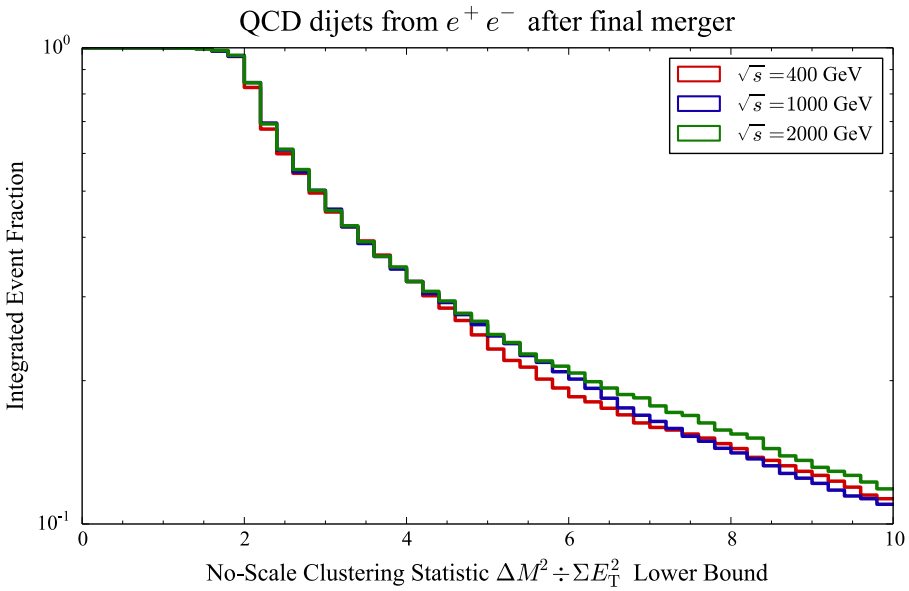
Clustering Visualization (Tops)

Visualization of Statistic Jump at Clustering



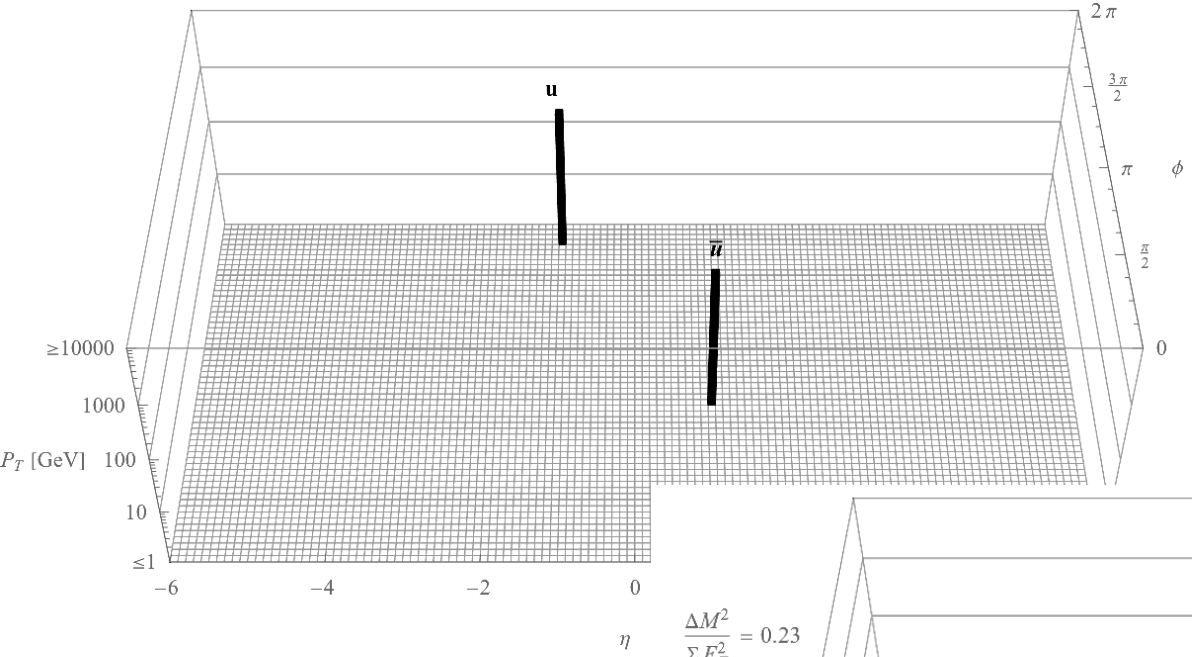
Test of Pre/Post Merger Statistic for Di-jets

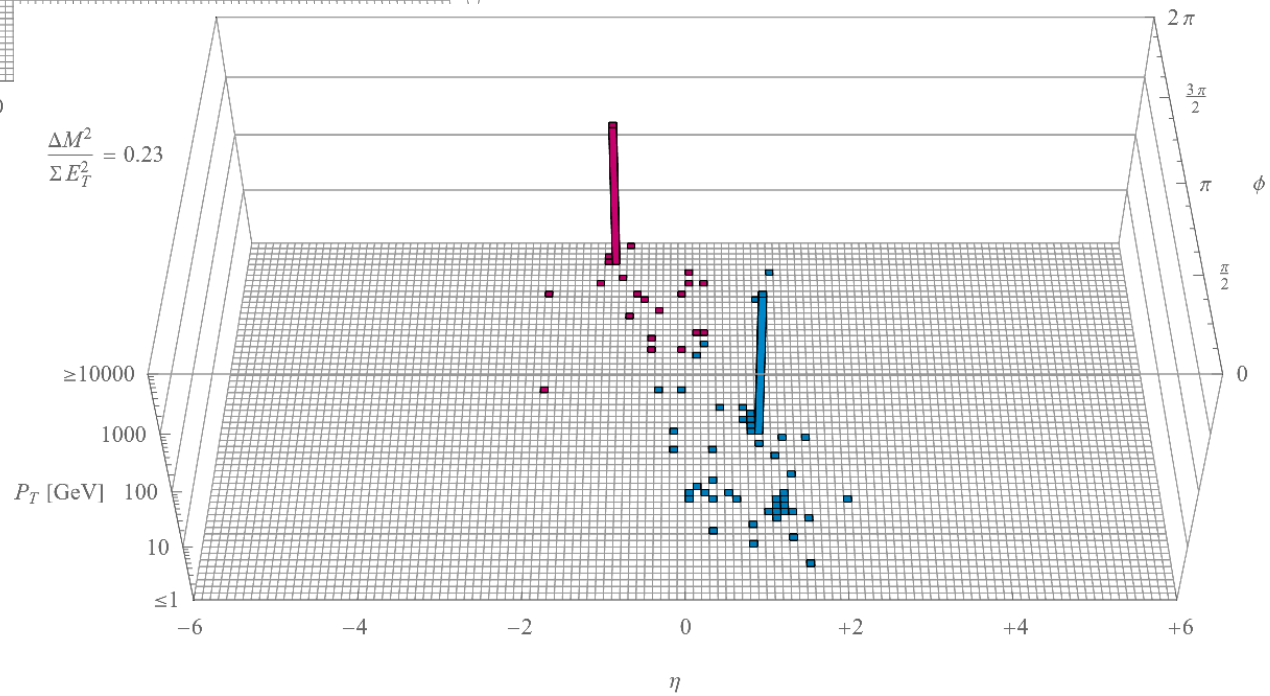




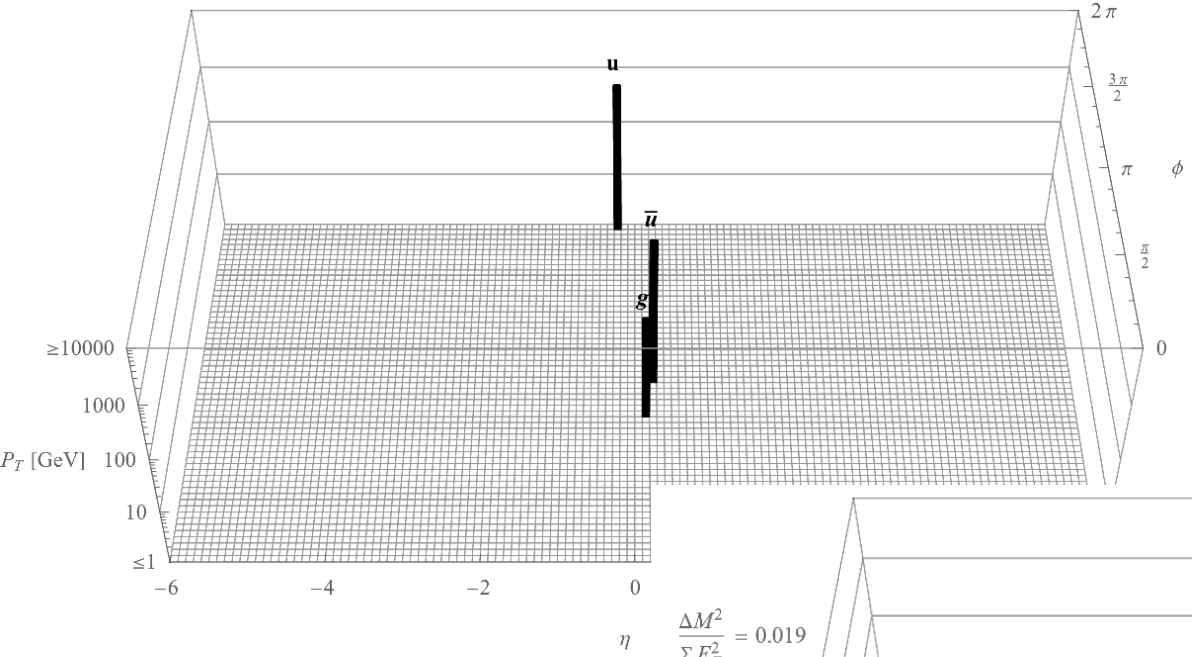
- 95% clustered prior to 0.1
- 95% of final clusters exceed 2.0

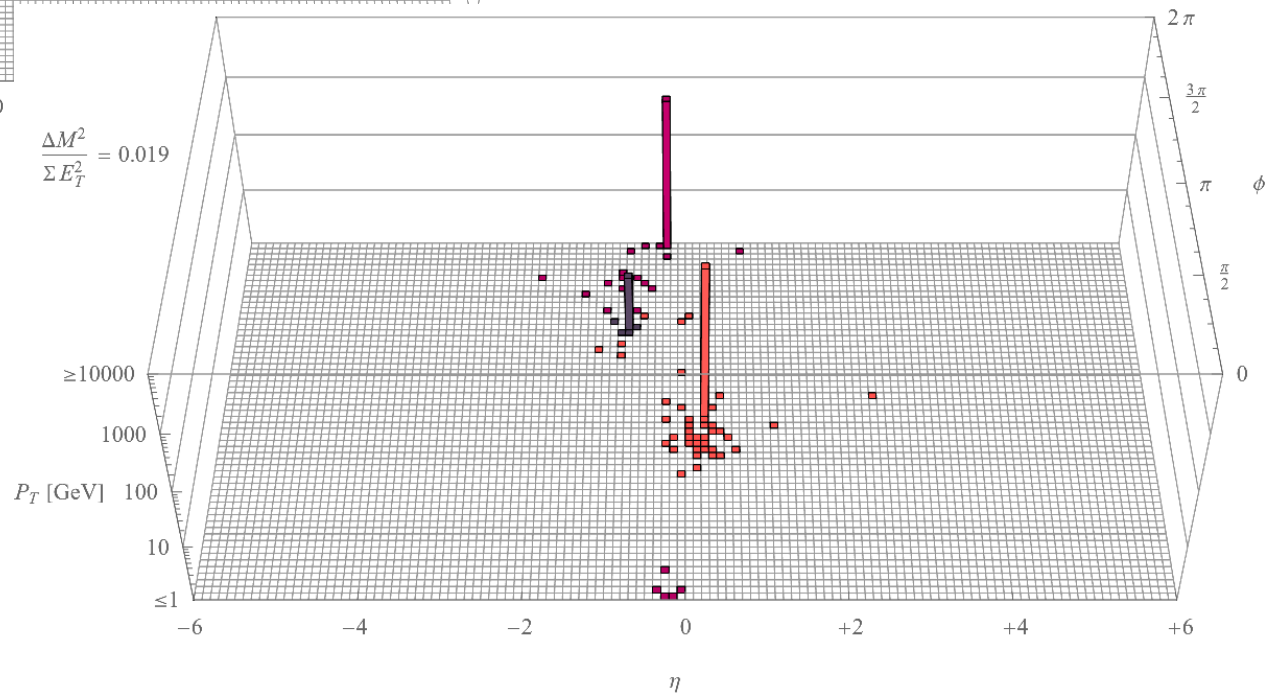
Matching of final 2 objects with Di-jet Truth, 1 TeV



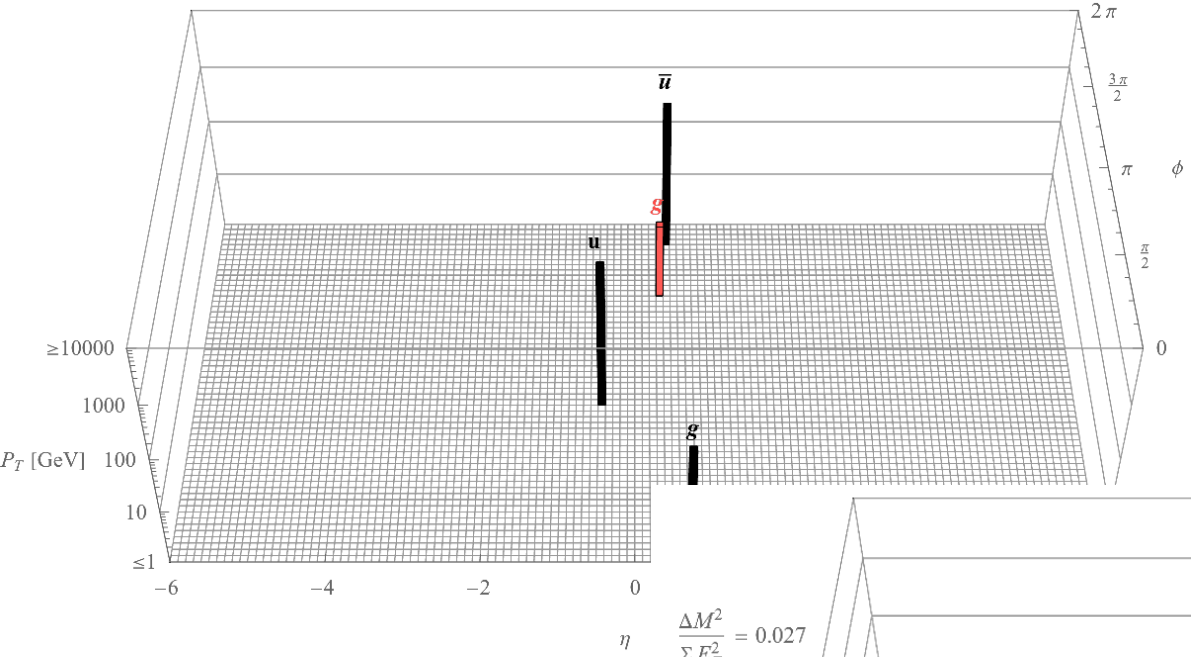


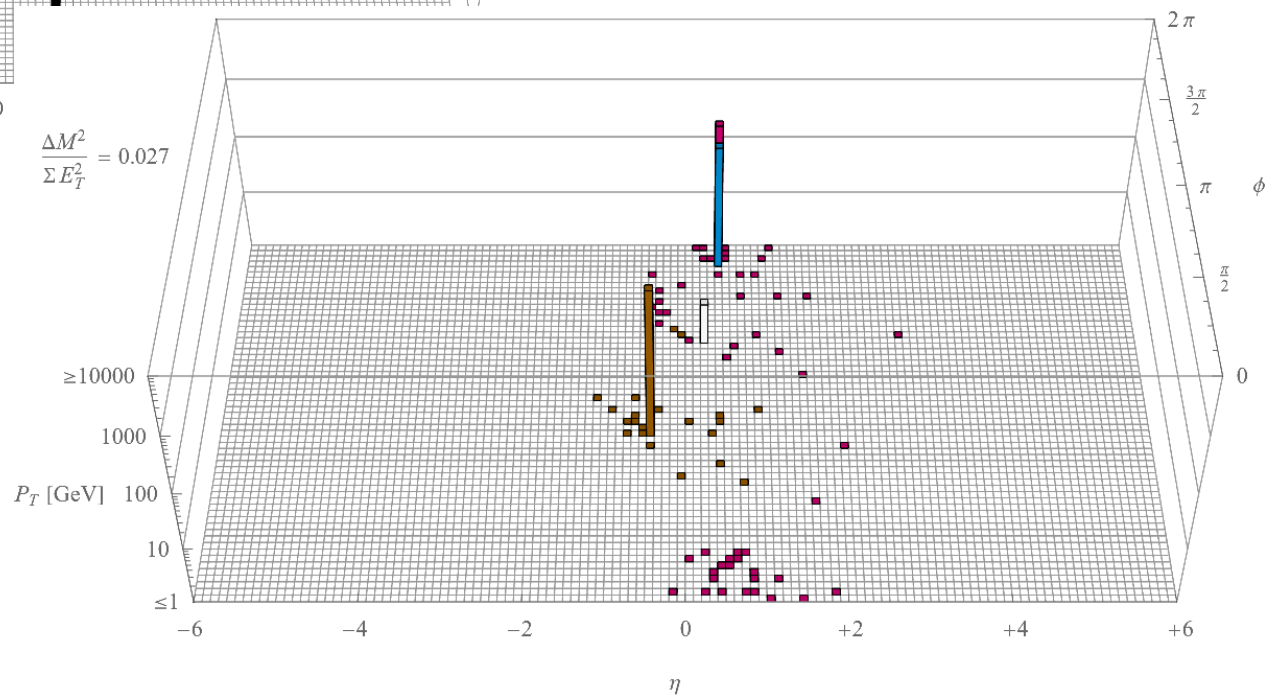
Matching of final 3 objects with Tri-jet Truth, 1 TeV



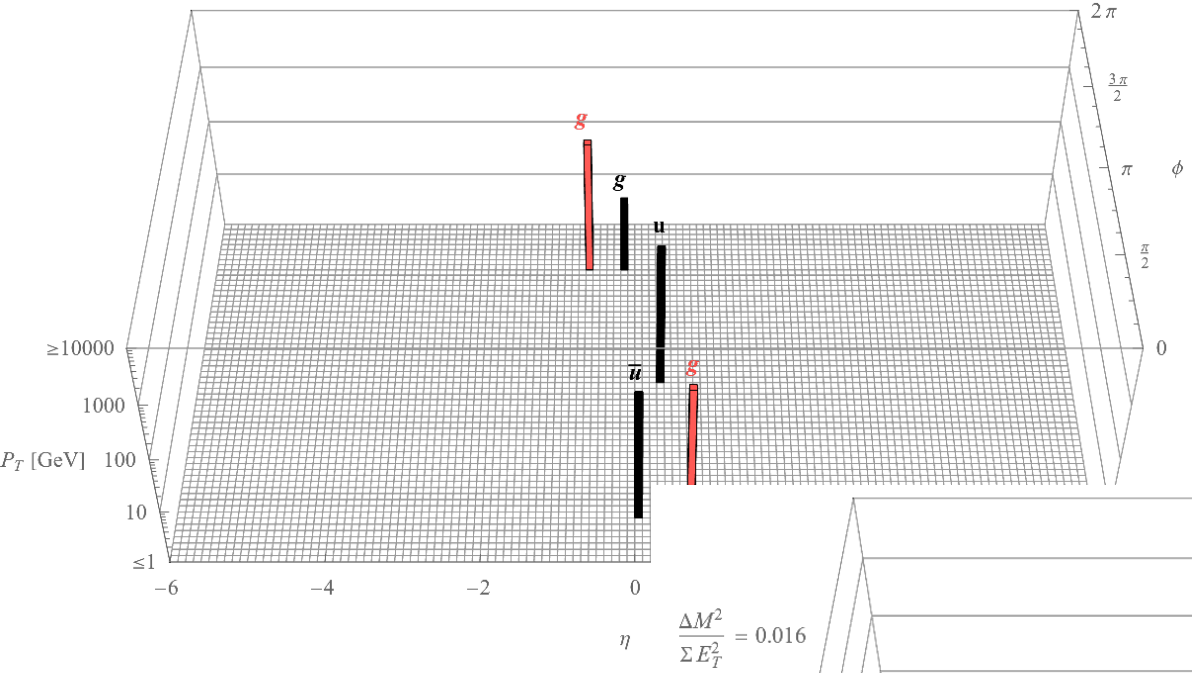


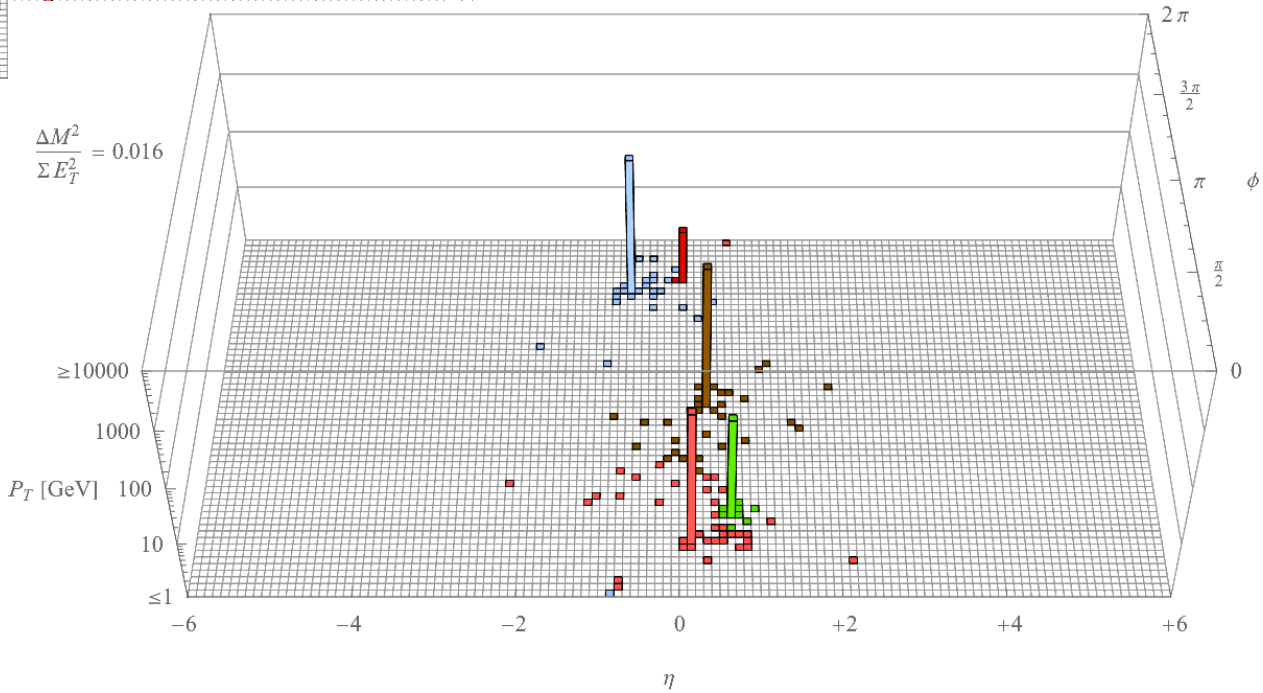
Matching of final 4 objects with Quad-jet Truth, 1 TeV



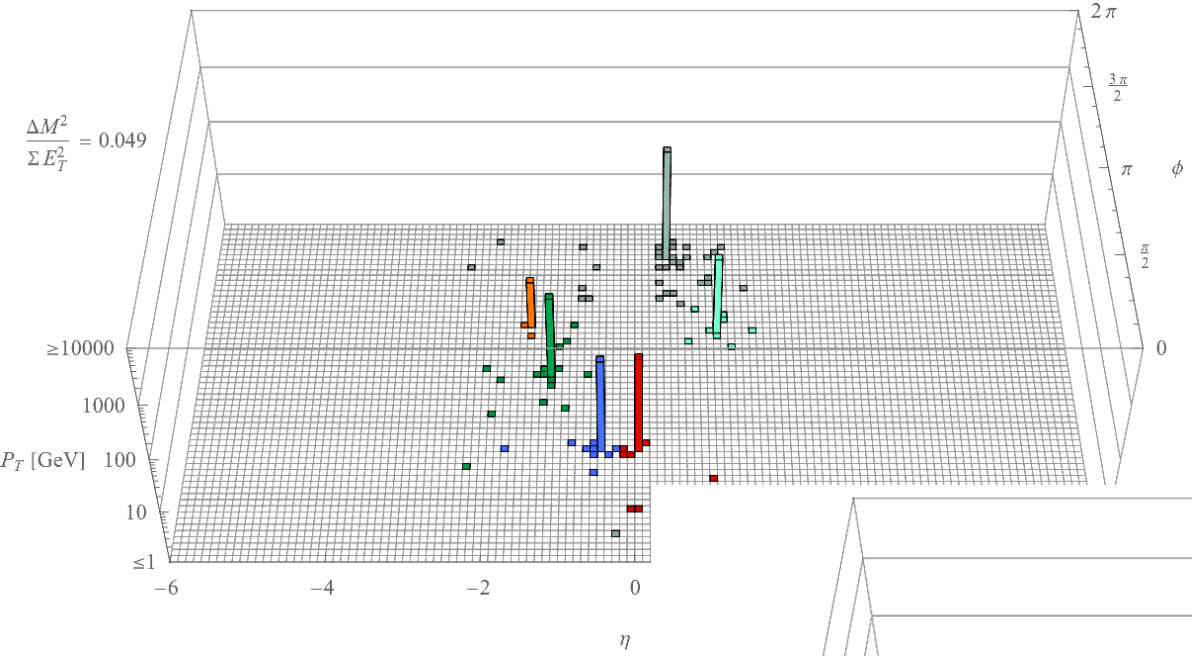


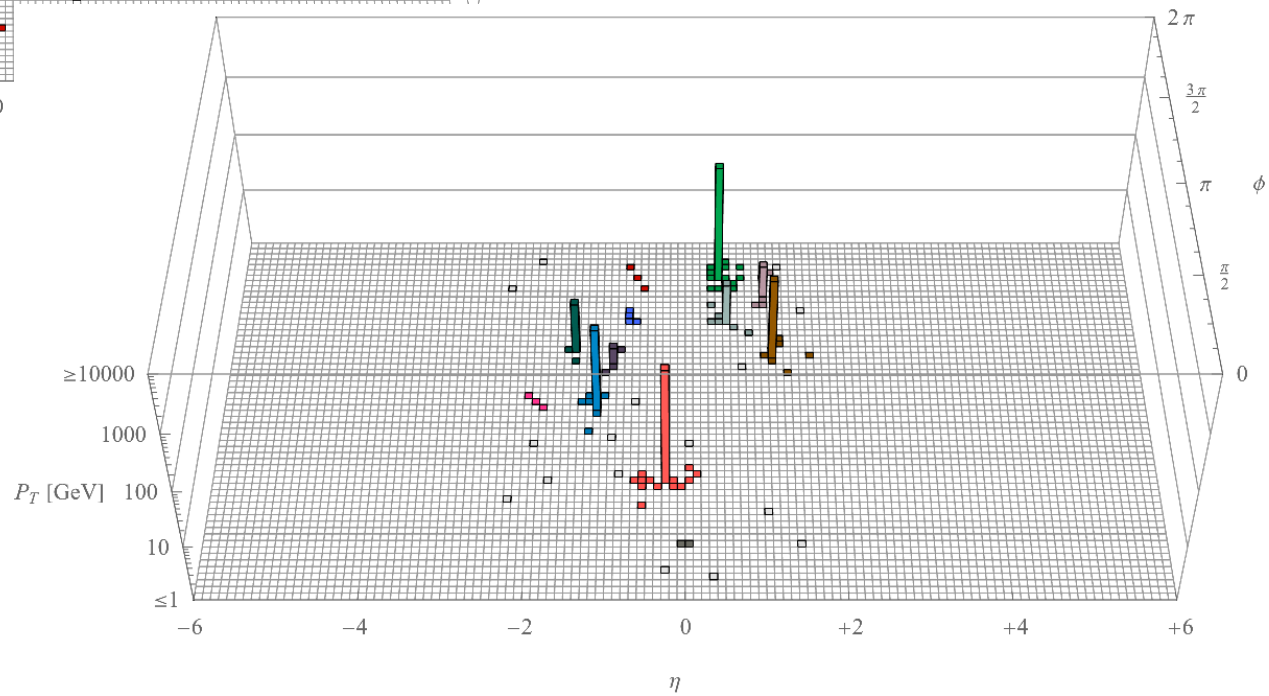
Matching of final 5 objects with Quin-jet Truth, 1 TeV



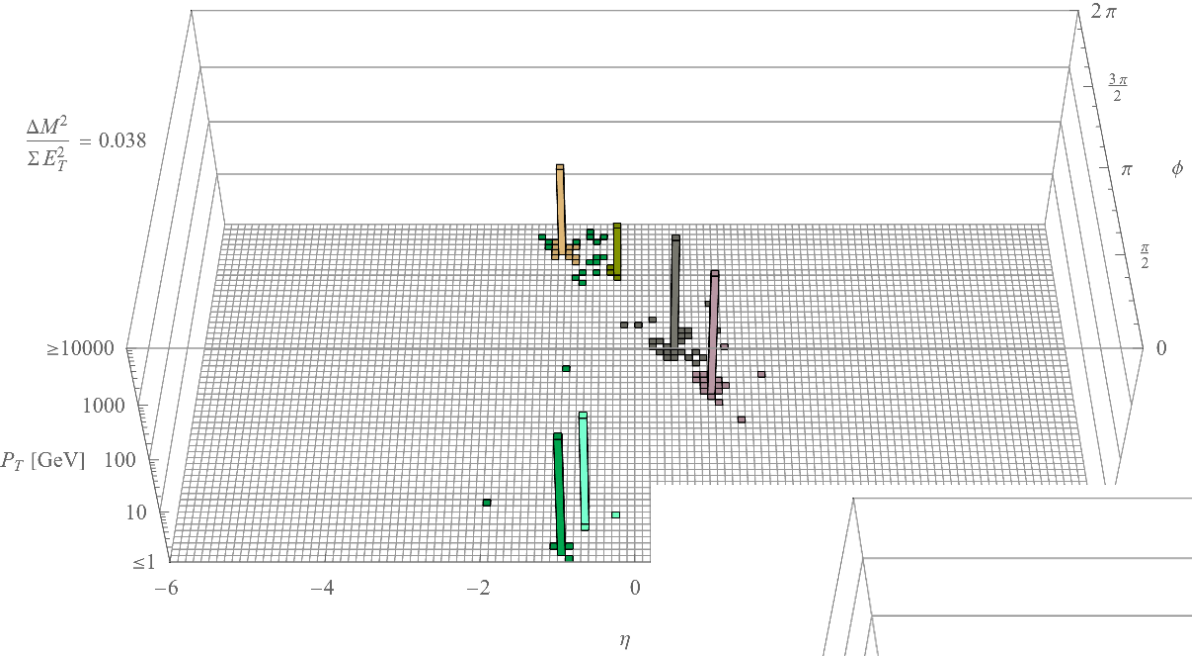


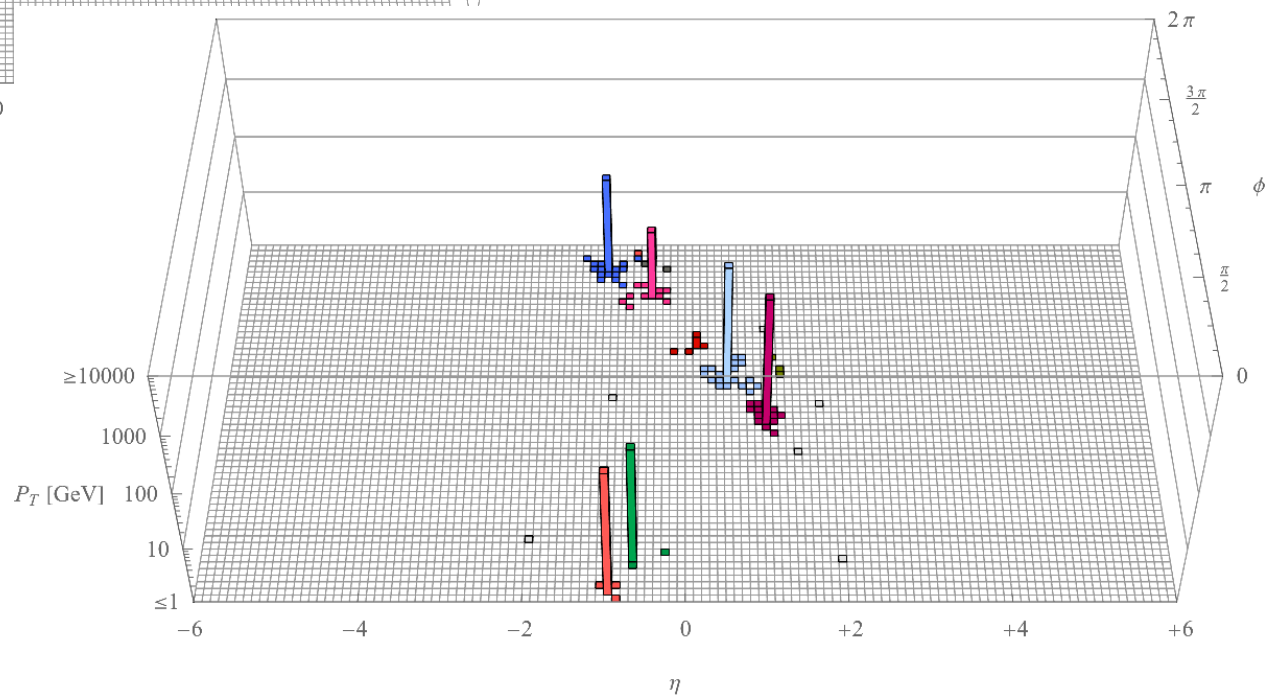
Comparison of final 6 objects vs AK5, 400 GeV



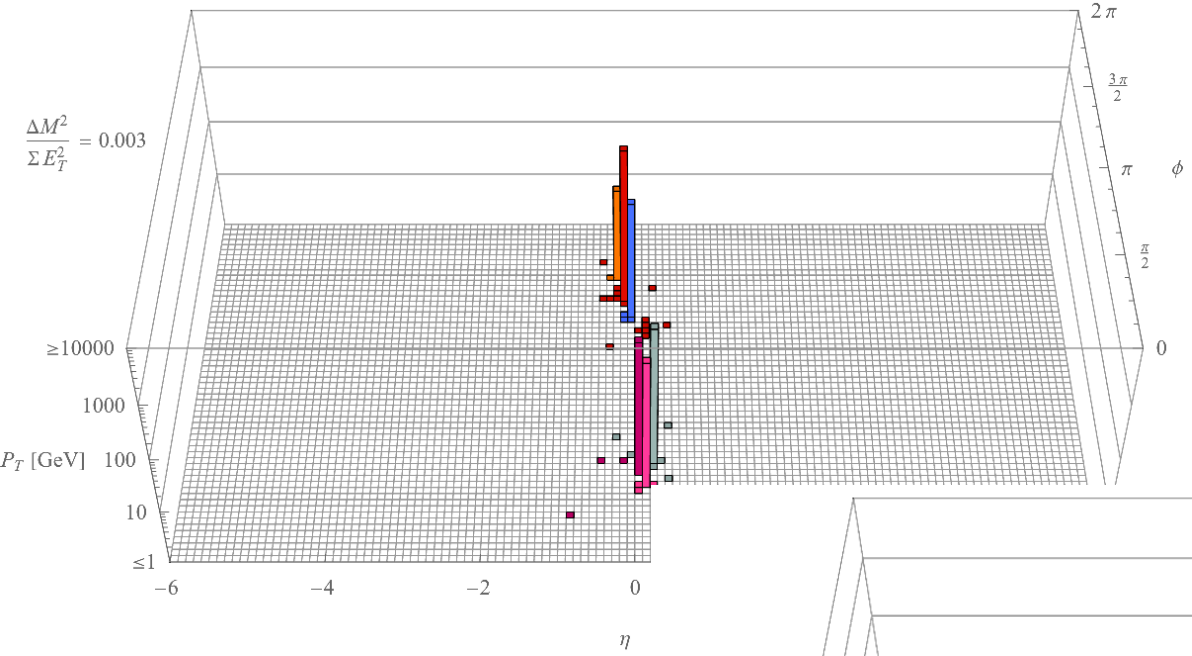


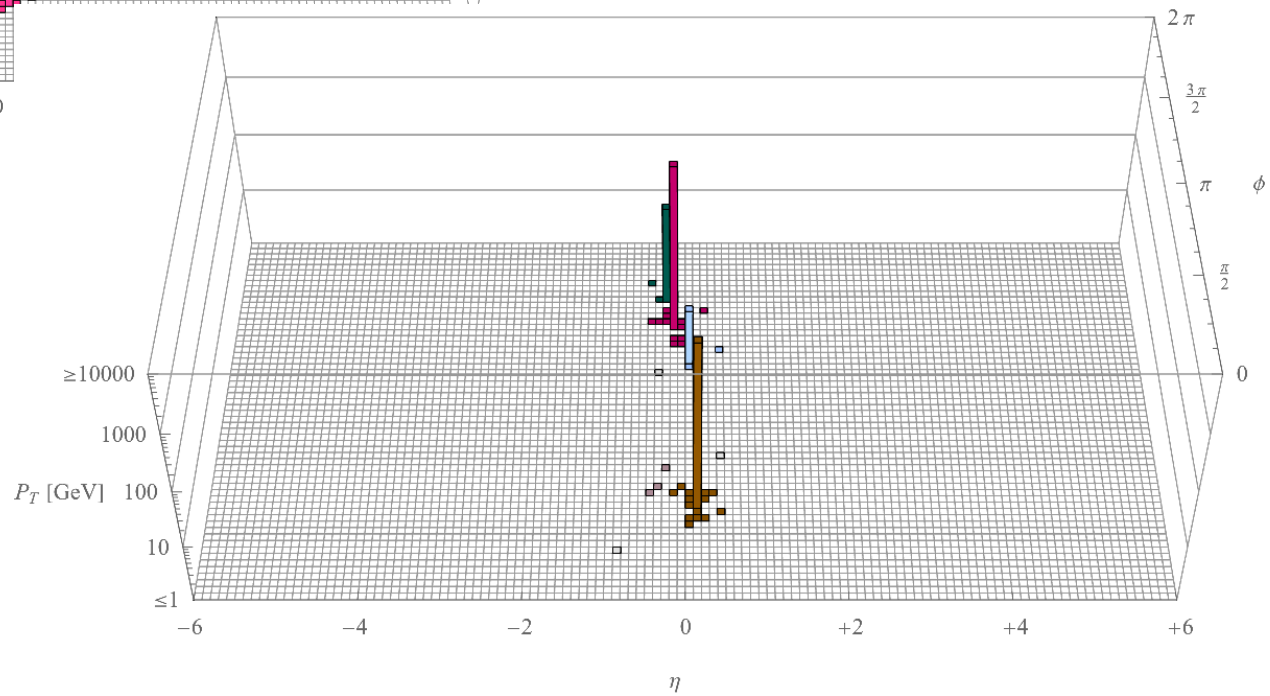
Comparison of final 6 objects vs AK5, 1 TeV



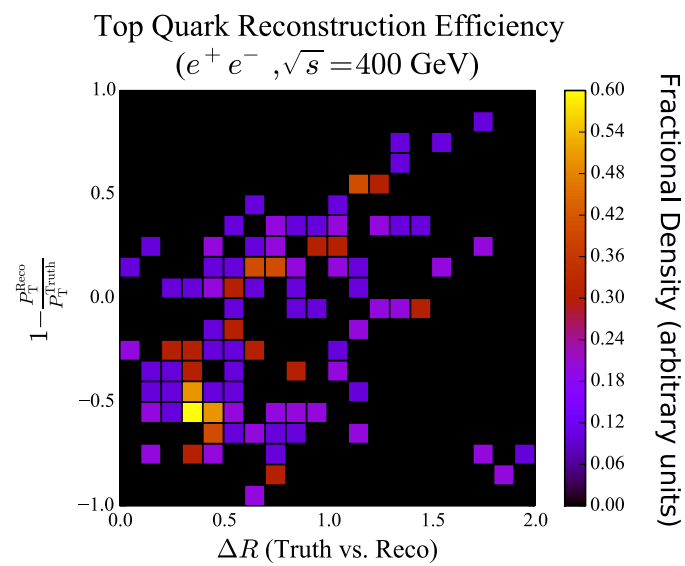


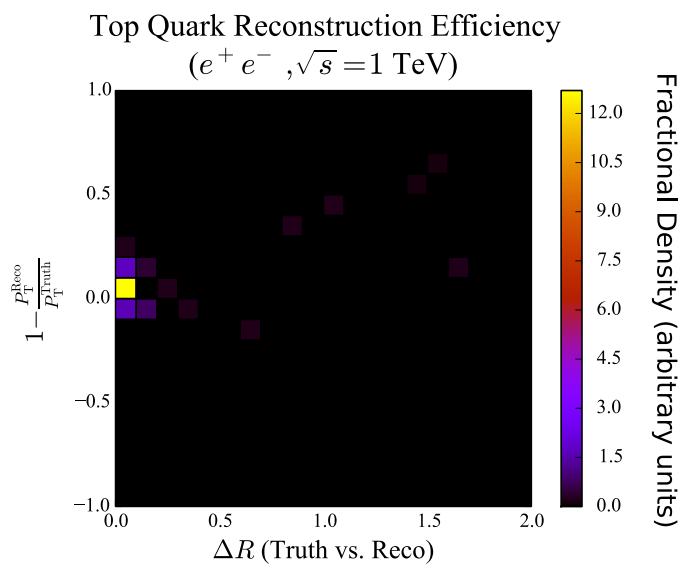
Comparison of final 6 objects vs AK5, 2 TeV

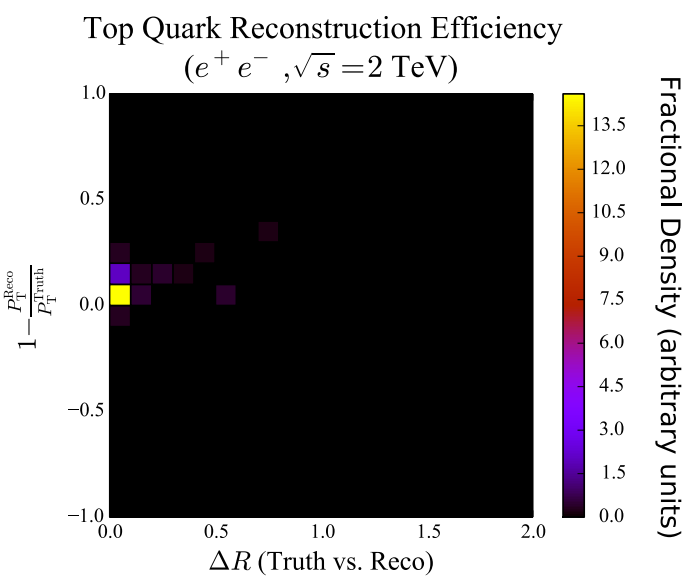


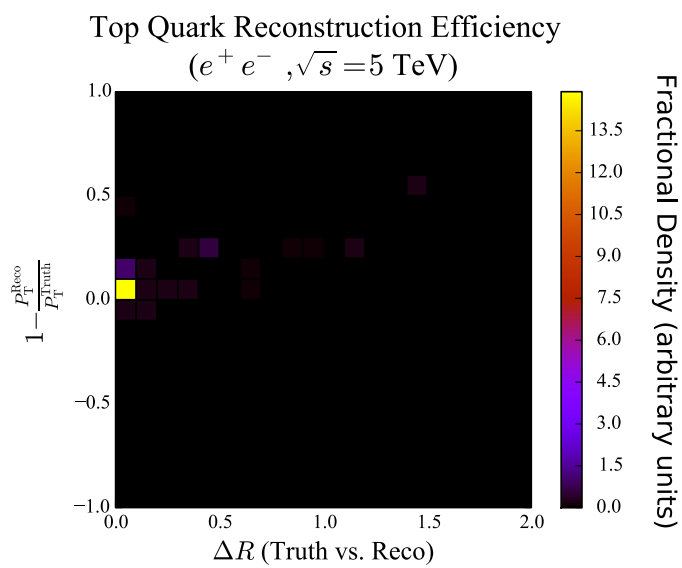


Matching of final 2 objects with Truth-Level Tops

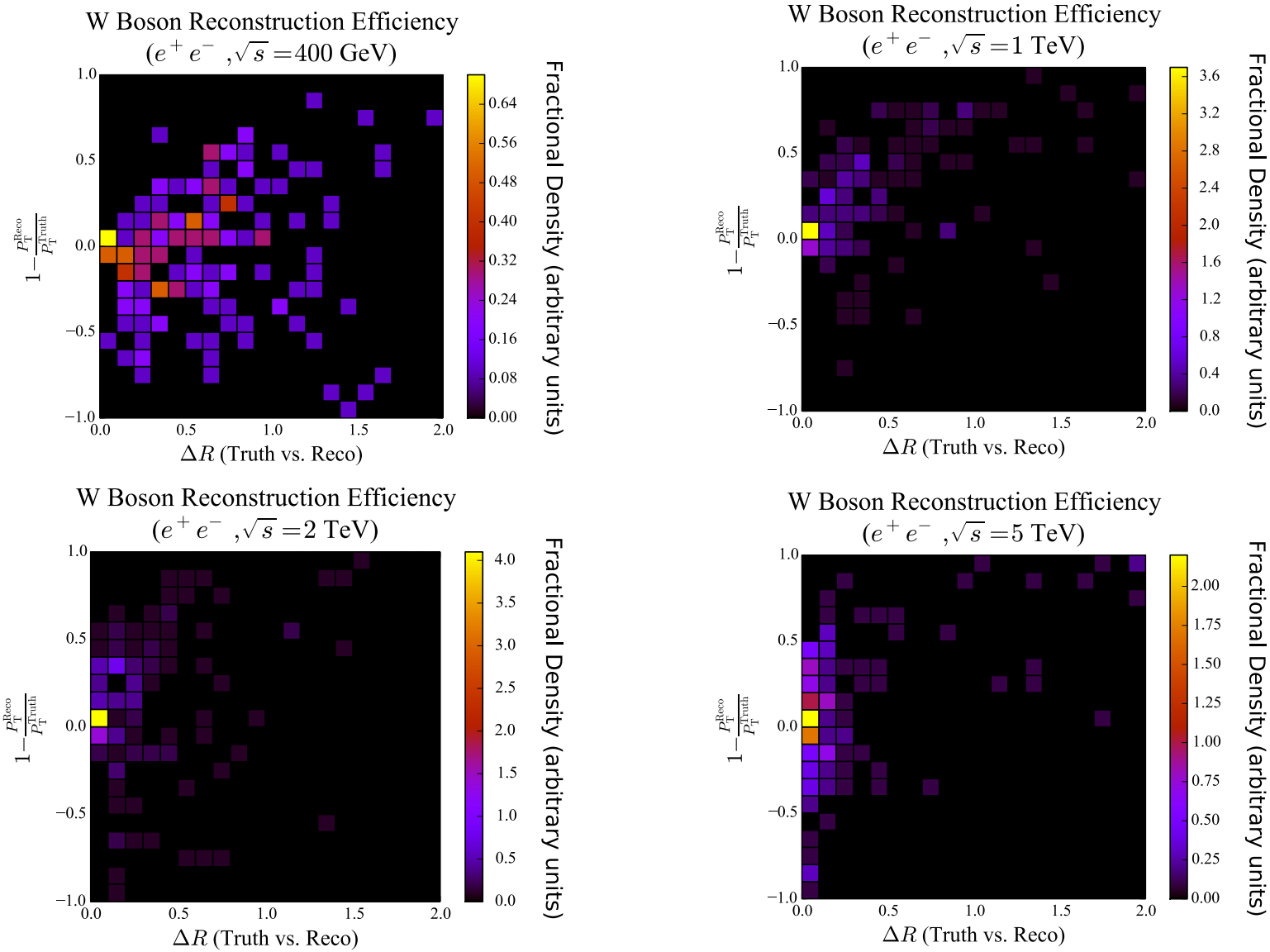




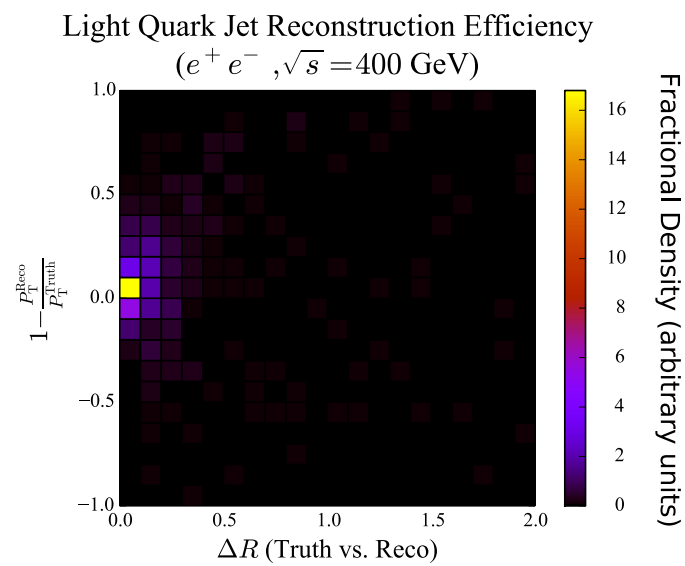


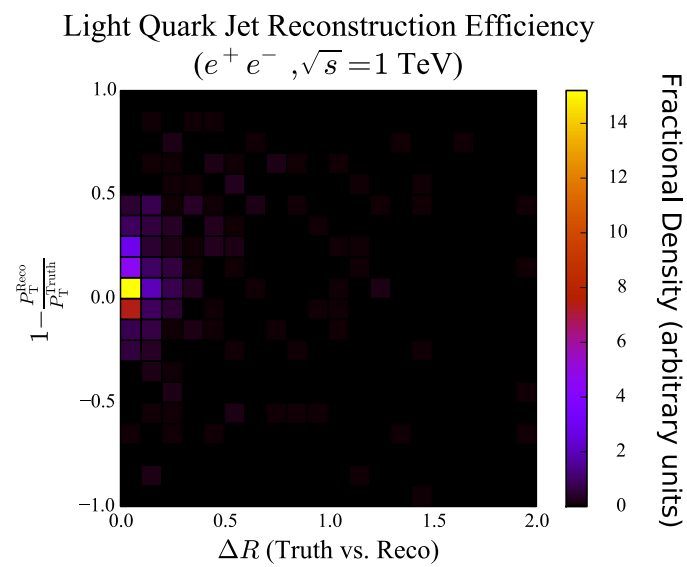


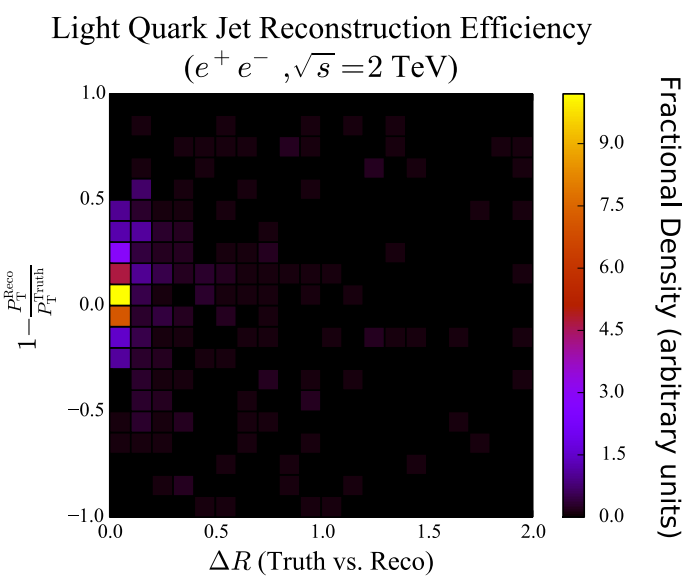
Matching of final 4 objects with Truth-Level W's (&b's)

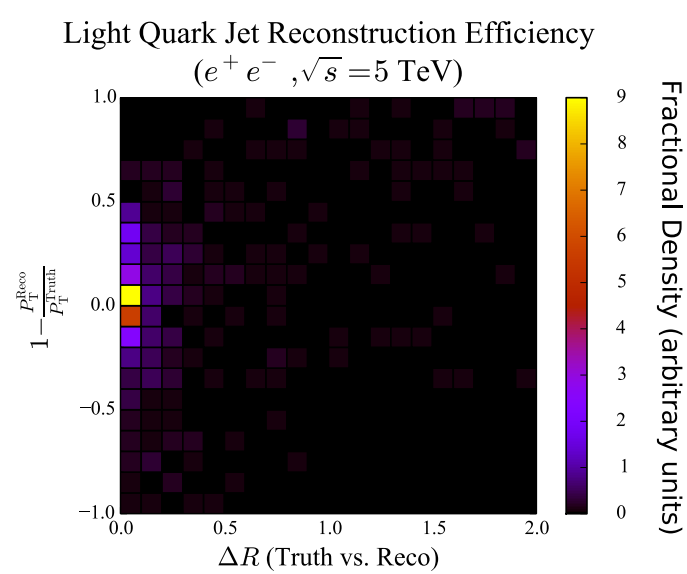


Matching of final 6 objects with Truth-Level Quarks









Thank You