Jet SIFT-ing

with ML for New Dark Sector Physics

Joel W. Walker
Sam Houston State University

Part I: Machine Learning & Operation of a BDT
Proceedings of Science *DOI: https://doi.org/10.22323/1.409.0027*

Part II: Hadronic Dark Sector Mass Reconstruction
William Shepherd, James Floyd, Camryn Sanders, and Jonathan Mellenthin

Part III: The SIFT Jet Clustering Algorithm
Andrew Larkoski (UCLA), Denis Rathjens (CMS), and Jason Veatch (ATLAS)
arXiv: 2302.08609 – *Phys.Rev.D* 108 (2023) 1, 016005

CETUP*
The Institute for Underground Science at SURF
June 24, 2024

From Jesse Thaler – PHENO 24

"…but what is the machine actually learning?"

My evolving perspective:

The desire for human interpretability often arises when we imperfectly specify the task we want to accomplish

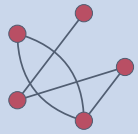
A more actionable definition of interpretability: identifying low-rank structures in high-dimensional datasets

From Jesse Thaler – PHENO 24

Three Lessons since Pheno 2019

Highlighting the power of active interpretability

Apologies that examples are all from my own work



If you have a catalog of trusted observables, you can translate a black-box algorithm on low-level inputs into a simple classifier on high-level features

$$\left\langle \Phi^{a_1} \Phi^{a_2} \right\rangle_{\mathcal{P}}$$

If there are simple operations like multiplication and sums that don't really require "interpretation", you can bake those into your machine learning architecture

$$\|\Phi(\hat{p}_1) - \Phi(\hat{p}_2)\| \leq L\|\hat{p}_1 - \hat{p}_2\|$$

If there is a property you want your network to have, make sure to impose algorithmic guardrails, otherwise the machine might pursue undesirable optimization

My Experience

- Neural Networks set the performance limit for S/B discrimination, etc.
- They can model any non-linear functional transformation with enough depth and the correct training objectives.
- This is time intensive, and the "black box" problem is real.
- My preferred use case is to build the domain knowledge (physics) into a curated set of applicable observables explicitly at the front end.
- Then, simpler machine learning (a Boosted Decision Tree) can tell you which observables are most relevant and perform an optimal separation.
- We have found that performing cuts by hand is biased & fine tuned (essential sensitivity adjustments in the tails may not appear until you have already cut very hard and are very easy to overshoot).

(Dutta, Fantahun, Fernando, Ghosh, Horne, Kumar, Palmer, Sandick, Stengel, Snedeker, JWW)

What is a BDT?

- Boosted Decision Trees are a type of Supervised Machine Learning
- "Hypothesis **Boosting**" is a technique for combining a number of "weak learners" (here shallow **Decision Trees**) into a "strong learner"
- Each tree separates signal (class 1) from background (class 0) via successive forks at selected split points on one data feature at a time
- Each terminal leaf carries a score, totaled over trees for a result on (0,1)
- Later trees focus on misclassifications from earlier trees (boosting!)

Why BDTs for Physics?

- Binary classification problems (Signal vs. Bg) are common
- We want to maximize discrimination power
- We want to eliminate bias and work efficiently
- We want to incorporate domain knowledge & expertise
- We want to understand what the machine learning learned

BDTs balance POWER with TRANSPARENCY

What is XGBoost?

- XGBoost (Extreme Gradient Boosting) by Tianqi Chen is a popular, innovative, widely available, and very fast BDT implementation
- Trees are built "greedily" (no backtracking), with the splitting feature, splitting value, and leaf score selected to optimize an objective $\mathcal{L}$
- This is guided by first (gradient) and second (Hessian) derivatives of the loss-function with respect to the class estimator of the nth object

$$g_n \equiv \frac{\partial \mathcal{L}_L}{\partial \hat{y}_n}$$
$$h_n \equiv \frac{\partial^2 \mathcal{L}_L}{\partial \hat{y}_n^2}$$

$$\delta \mathcal{L}_L \simeq \sum_{n=1}^{N} \left\{ g_n \delta \hat{y}_n + h_n \frac{\delta \hat{y}_n^2}{2} \right\}$$
$$\partial(\delta \mathcal{L}) / \partial(\delta \hat{y}_n) = 0$$
$$\delta_0 \hat{y}_n = -g_n / h_n$$
$$(-\delta_0 \mathcal{L}_L) \simeq \textstyle\sum_n g_n^2 / (2 h_n)$$

XGBoost Logic

- Events with common features flow similarly through the decision tree, and "vote" for the score carried by their destination node.
- Events that are currently missorted (large slopes) get bigger votes.
- These factors are also scalable by per-event physics weights.

$$G_\ell \equiv \sum_{n=1}^{N} g_n \times \delta_{\ell,\ell'(\vec{x}_n)}$$

$$H_\ell \equiv \sum_{n=1}^{N} h_n \times \delta_{\ell,\ell'(\vec{x}_n)}$$

XGBoost Details

- The algorithm "works backwards". If leaf assignments were KNOWN, then one could calculate the leaf score s_ℓ that optimizes the gain.
- The max $-(\delta_0 \mathcal{L})$ split is selected (features AND value are SCANNED)
- "Regulators" limit overtraining

$$\mathcal{L}_\Omega = \gamma L + \alpha \sum_{\ell=1}^{L} |s_\ell| + \lambda \sum_{\ell=1}^{L} \frac{s_\ell^2}{2}$$

$$\delta\mathcal{L} \simeq \sum_{\ell=1}^{L} \left\{ \gamma + \alpha|s_\ell| + G_\ell s_\ell + (H_\ell + \lambda)\frac{s_\ell^2}{2} \right\}$$

$$s_0^\ell = -\frac{G_\ell \pm \alpha}{H_\ell + \lambda}$$

$$(-\delta_0\mathcal{L}) \simeq \sum_{\substack{\ell=1 \\ (|G_\ell|>\alpha)}}^{L} \left\{ \frac{(G_\ell \pm \alpha)^2}{2\,(H_\ell + \lambda)} - \gamma \right\}$$

Binary Logistic Regression

- The binary logistic objective yields continuous classification scores on 0 to 1
- It is an explicit function with explicit derivatives, so the gradient and Hessian are calculated & known.
- A key idea is to map between an infinite space where scores are summed and a finite space where they are reported.

"logistic" function $p = \dfrac{1}{1+e^{-y}} \in \{0, 1/2, 1\}$

"logit" function $y = \ln\left(\dfrac{p}{1-p}\right) \in \{-\infty, 0, +\infty\}$

$$\mathcal{L}_L = -\sum_{n=1}^{N}\left\{p_n \ln(\hat{p}_n) + (1-p_n)\ln(1-\hat{p}_n)\right\}$$

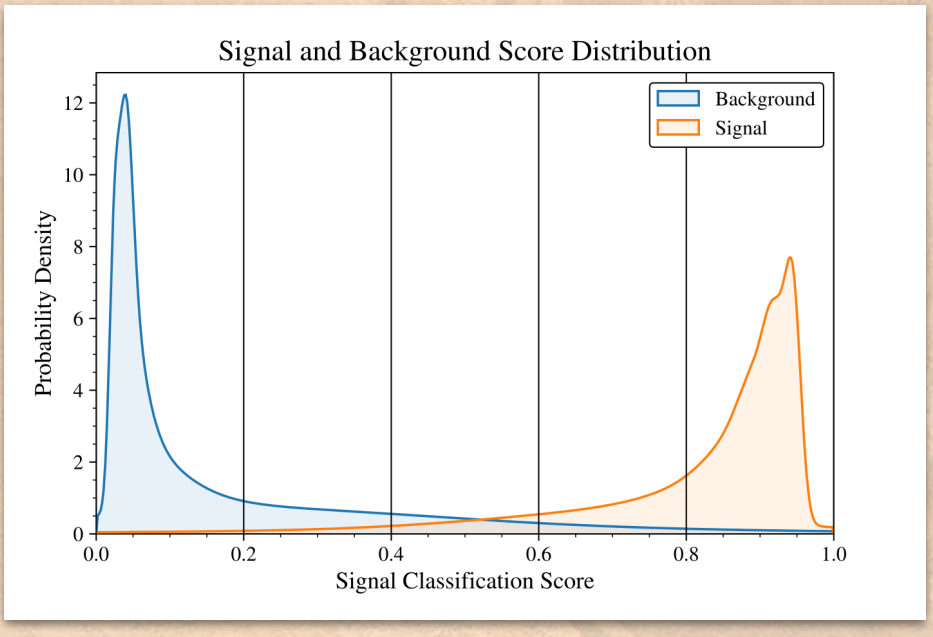
$$\begin{aligned} g_n \equiv \frac{\partial \mathcal{L}_L}{\partial \hat{y}_n} &= \frac{\partial \mathcal{L}_L}{\partial \hat{p}_n} \times \frac{\partial \hat{p}_n}{\partial \hat{y}_n} \\ &= -p_n \times (1-\hat{p}_n) + \hat{p}_n \times (1-p_n) \\ &= \hat{p}_n - p_n \end{aligned}$$

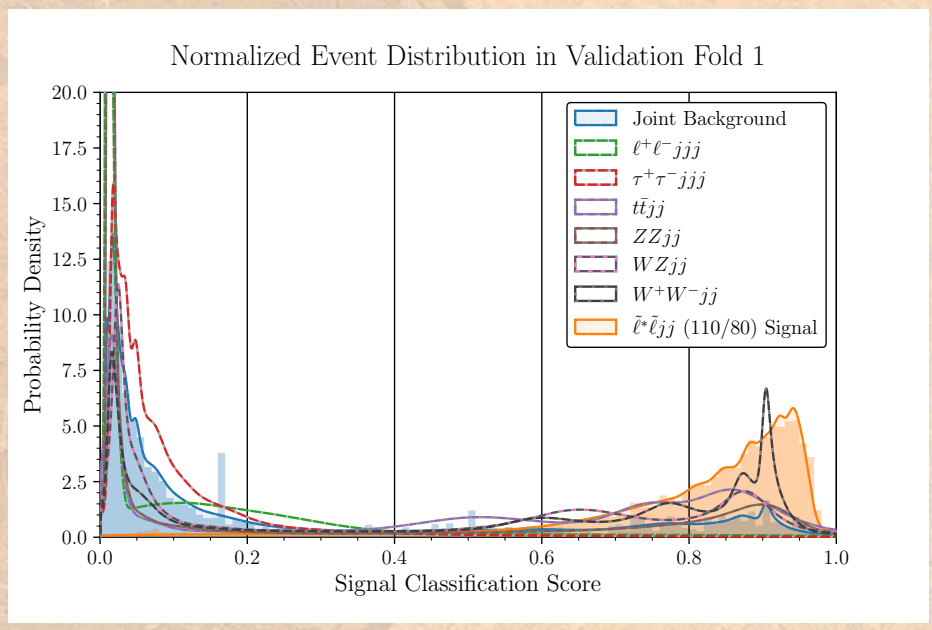
$$\begin{aligned} h_n \equiv \frac{\partial^2 \mathcal{L}_L}{\partial \hat{y}_n^2} &= \left(\frac{\partial^2 \mathcal{L}_L}{\partial \hat{y}_n \partial \hat{p}_n} = 1\right) \times \frac{\partial \hat{p}_n}{\partial \hat{y}_n} \\ &= \hat{p}_n \times (1-\hat{p}_n) \end{aligned}$$

Note on Tuning / Weights

- A number of hyper-parameters, including the objective-level regulators γ, α, and λ are available to confront "Bias-Variance" issues
- In short, one must not add complexity without benefit (like χ^2/DOF)
- XGBoost also allows specification of maximal tree depth and count, with handles for "early stopping" or "pruning" when learning slows
- We balances data sets sent for training by separately normalizing the signal and background cross section to unity
- This stabilizes optimal numerical values of various hyper parameters, eliminates tension between intensive/extensive scaling, and induces a natural $\mathcal{O}(1)$ scale for the gradient, Hessian, and regulators

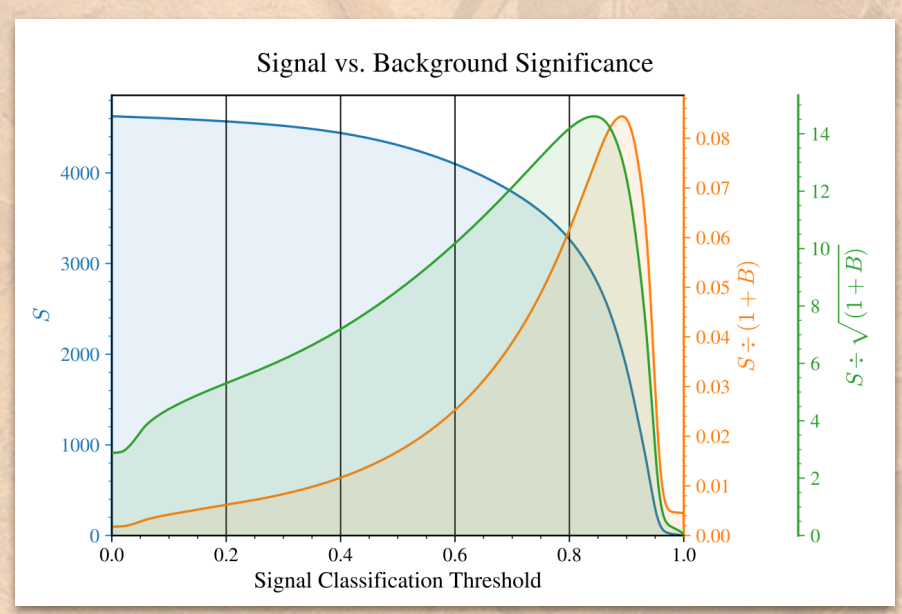
MInOS Output

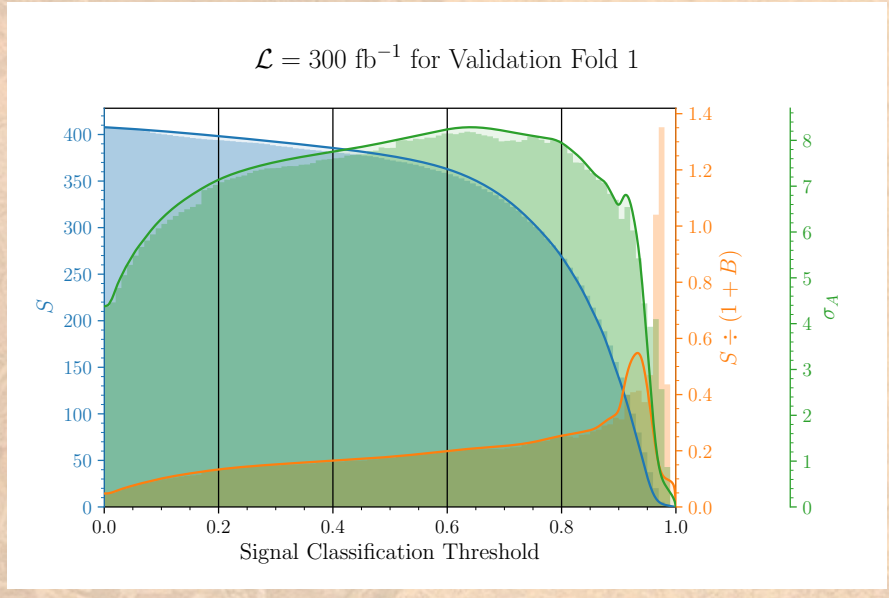




- Signal & Background Probability Density Visualizes Separation

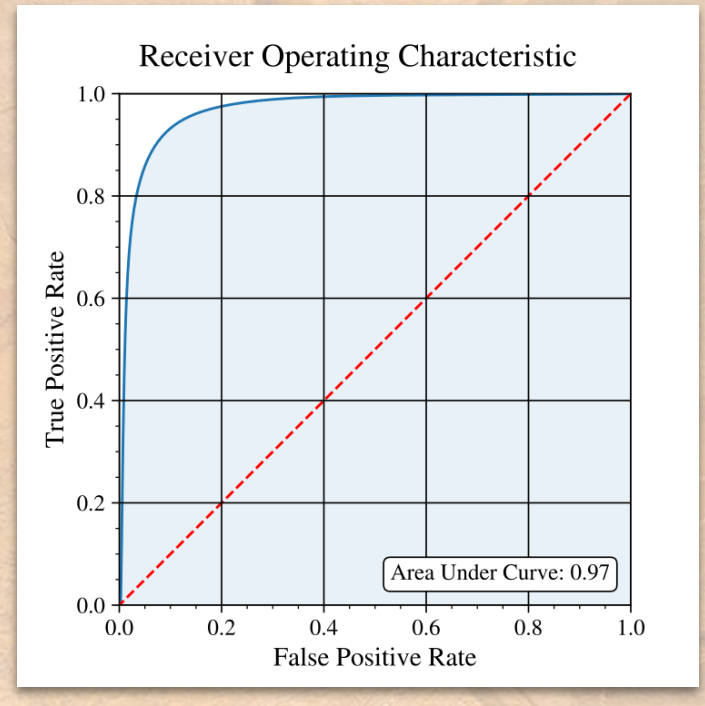
MInOS Output

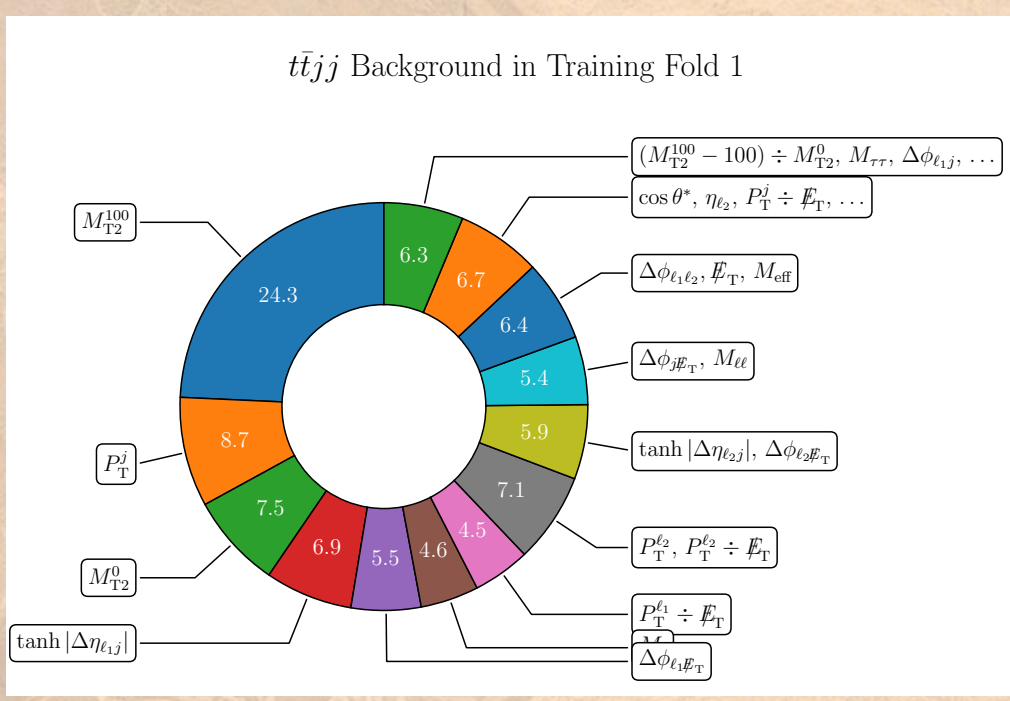




- Survival fraction of S, B as a function of the classification threshold are used to show achievable significance (at specified luminosity)

MInOS Output





- The ROC curve is a standard metric of S/B separability
- A feature importance chart clarifies what is going on inside the BDT

SMG Takeaway Messages

- We considered a small-mass-gap scenario that is challenging at the LHC.
- We were able to improve on a prior study with manual event selections.
- The older study was very sensitive to the sequence of applied cuts.
- The older study did not separate training from validation (bias).
- We doubled significance and increased the S/B by 50%.
- We found good stability across different train/test folds.
- We learned that it is beneficial to help the ML by making "obvious" cuts by hand so that its attention can be focused on the surviving tails.

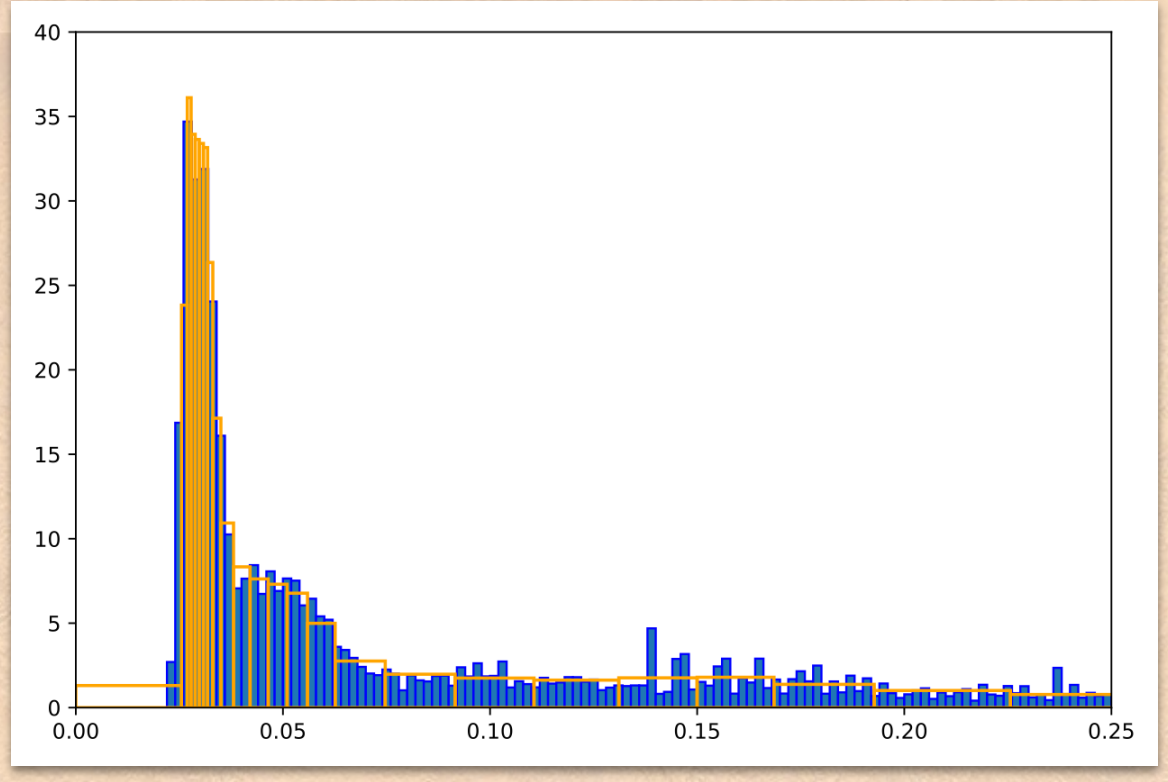
(Dutta, Fantahun, Fernando, Ghosh, Horne, Kumar, Palmer, Sandick, Stengel, Snedeker, JWW)

Data Smoothing

- The statistical population of events grows sparse with harder cuts
- Smoothing may better approximate the reality of continuum data
- Naive interpolation (e.g. cubic spline) can induce unphysical artifacts
- We want to retain sharpness where clustering is real while washing out jitter where statistical event densities are low
- A proprietary multi-step solution is adopted to meet these goals

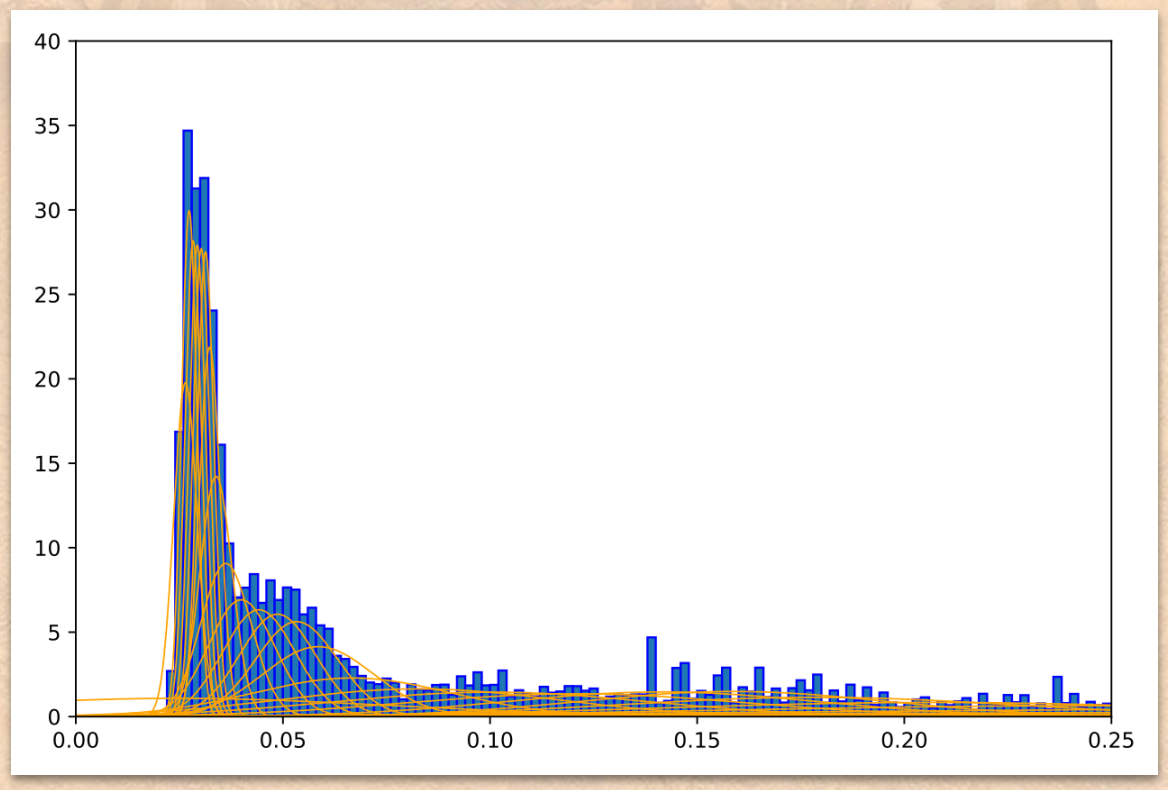
Data Smoothing

- First, we do variable-width binning with equal areas (cross sections)



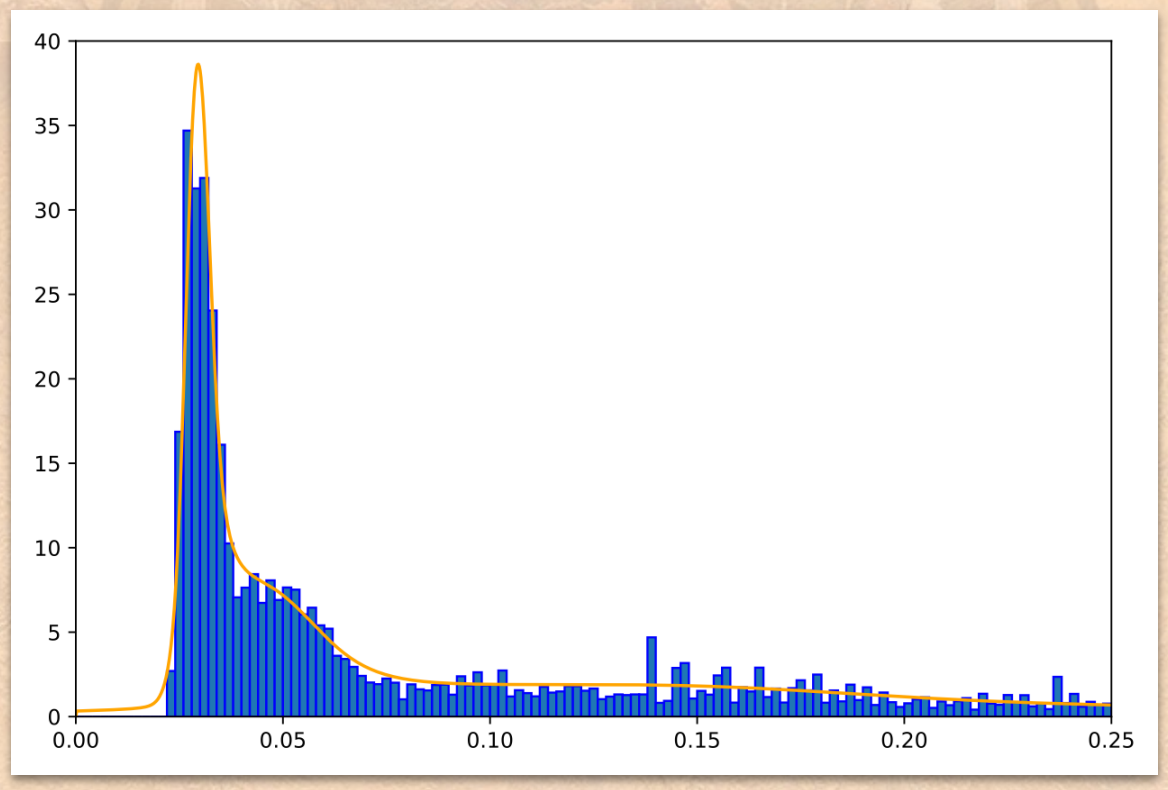
Data Smoothing

- Then, we populate narrow fixed bins via Gaussians ∝ prior widths



Data Smoothing

- Finally, we sum and scale to generate a smooth density of area one



Software Advertisement

- A unified pipeline for computing observables, plotting distributions, and applying BDT ML
- Per-event weights for independent and oversampled sim is handled automatically
- Code is fully decoupled from instructions via a unified analysis description meta-language
- The package is available for download & public use from GitHub:
- https://github.com/joelwwalker/AEACuS

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.

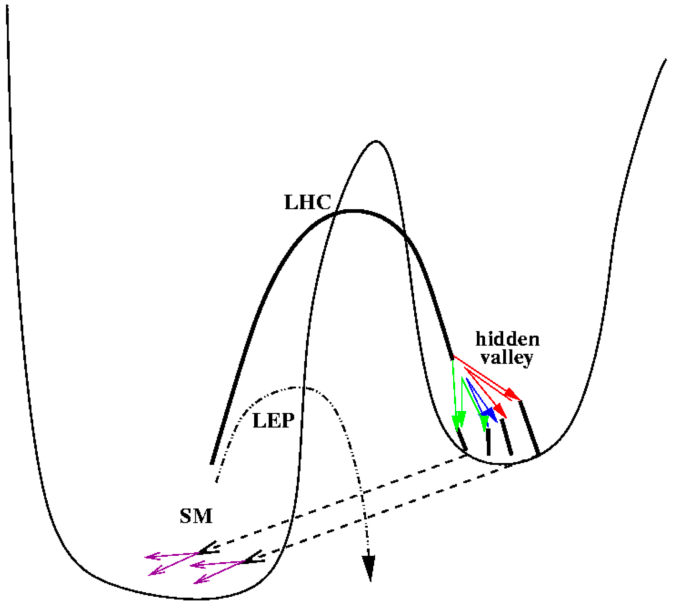
PoS(CompTools2021)027

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 … "

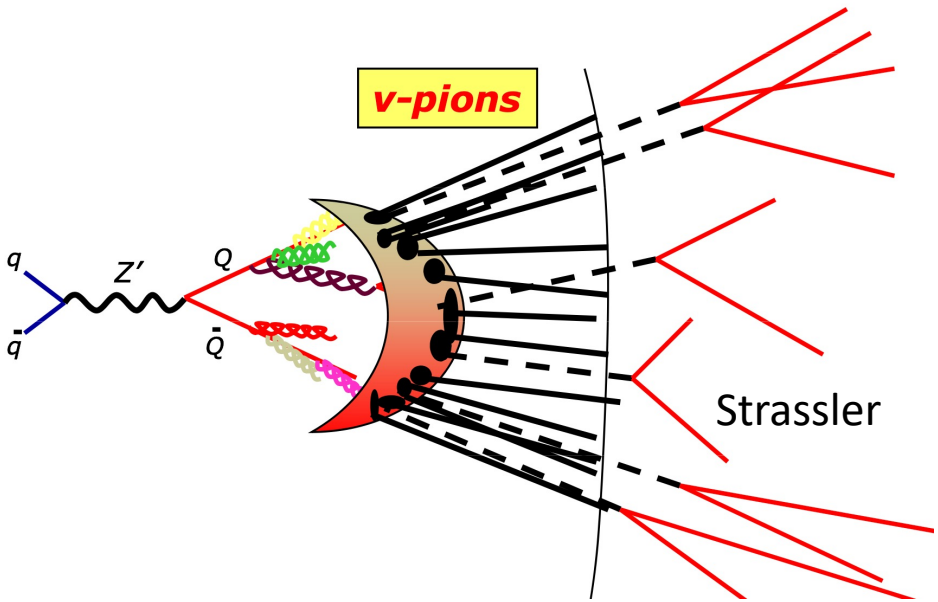
- Characterized by new light physics that is weakly coupled to the SM
- A heavy intermediary presents a high energy barrier to access the new sector
- Strong dynamics & confinement are typical
- A mass gap allows decays back to the SM



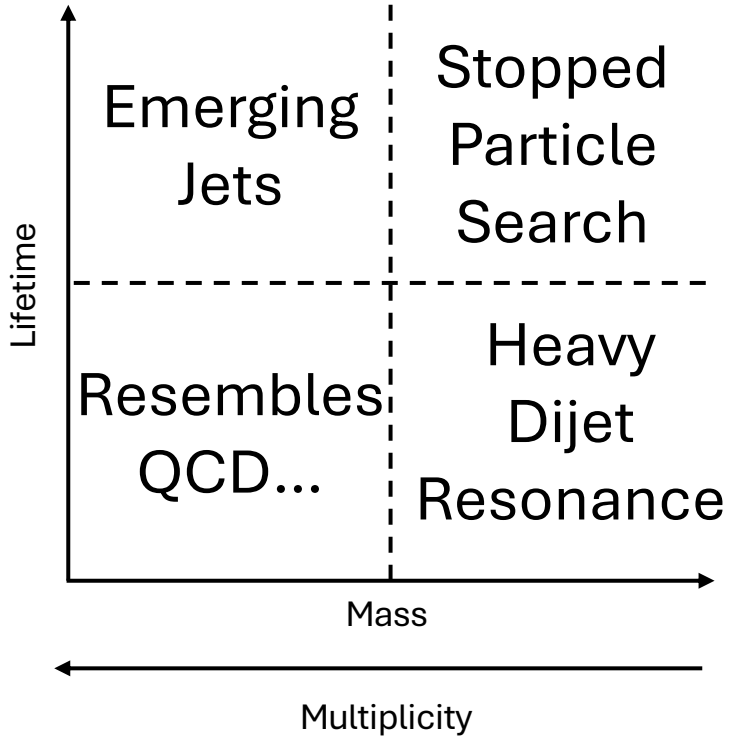
Strassler

Hidden Valley Strong Dynamics

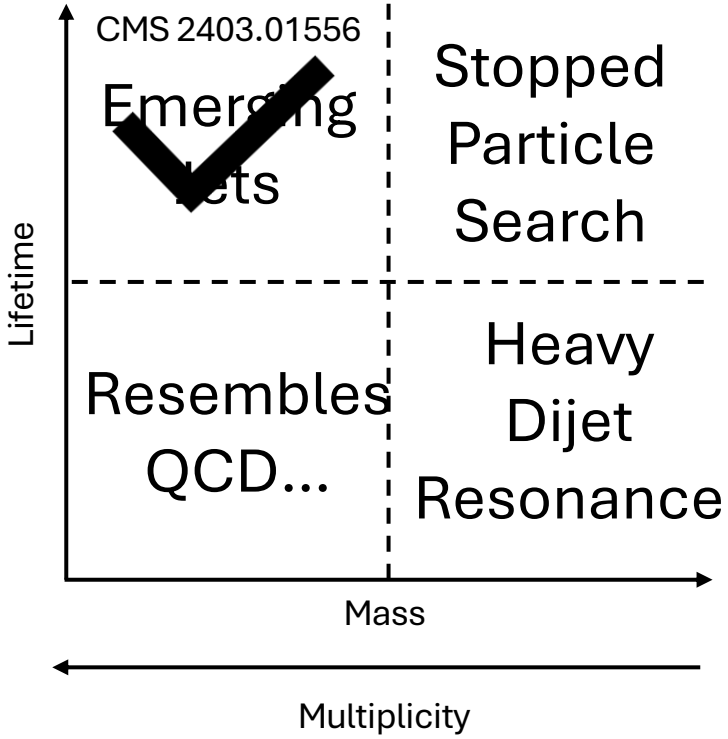
- Classic signatures include a heavy dilepton resonance and/or displaced vertices
- We are interested here in a more challenging scenario (0806.2835 Strassler)
- The mediator is a few-TeV *Z'* coupled to the SM by kinetic mixing
- Heavy v-Quarks are pair produced and they shower / hadronized
- Flavor-diagonal 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

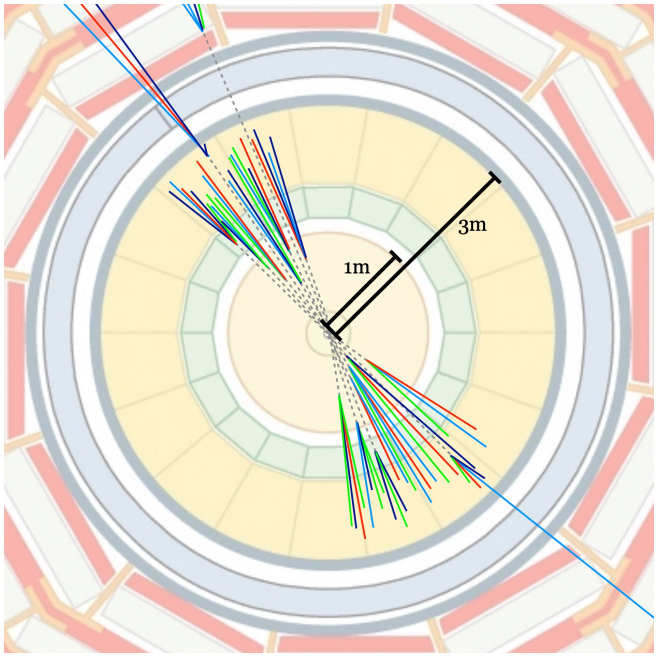


Hidden Valley Signatures

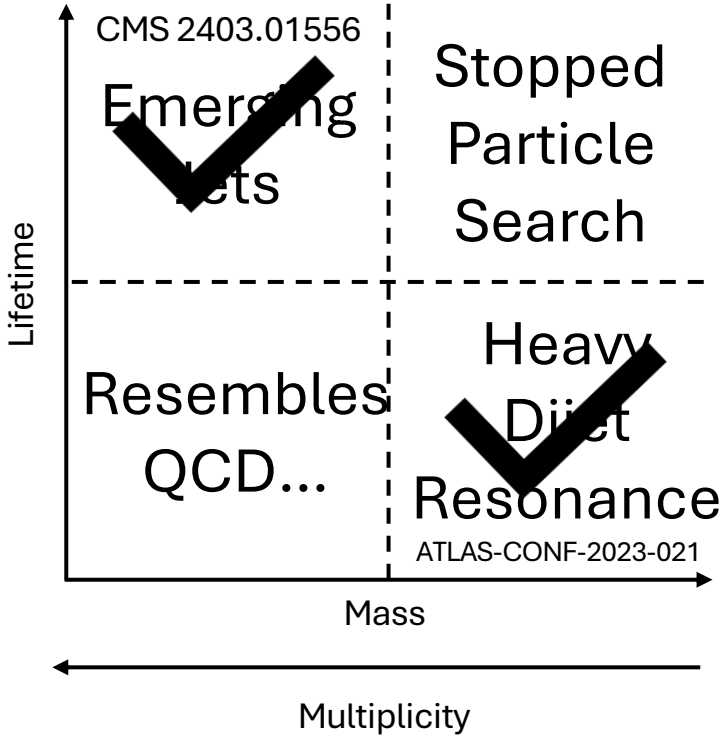


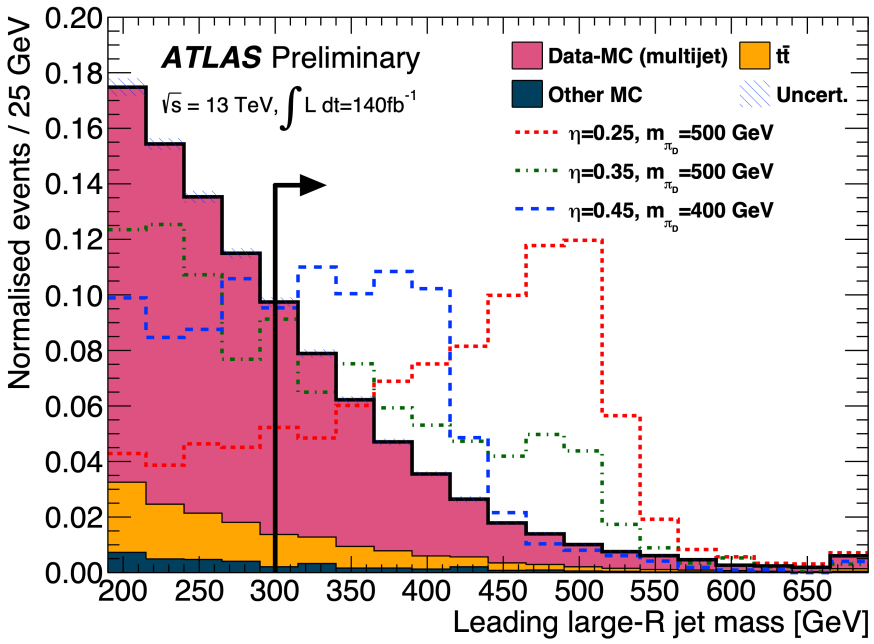
Hidden Valley Signatures



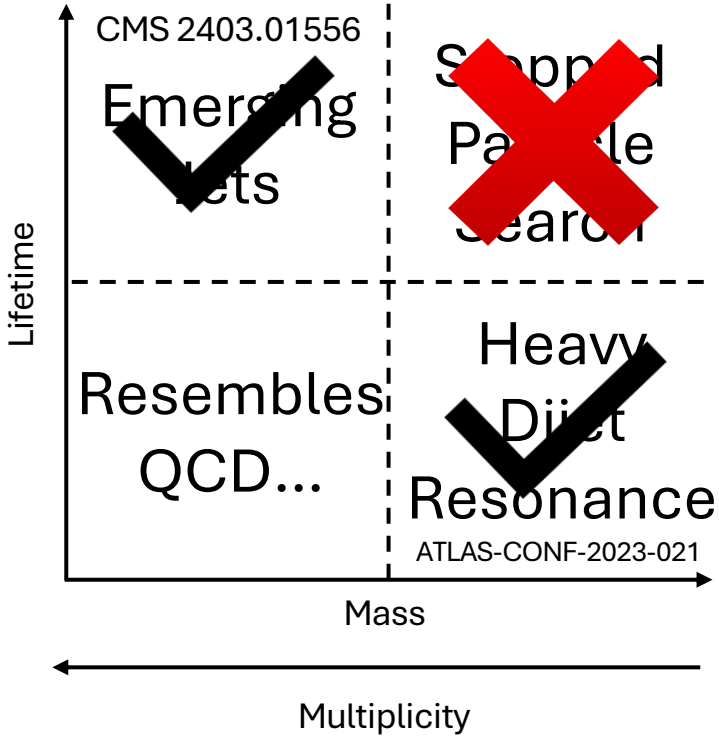


Hidden Valley Signatures



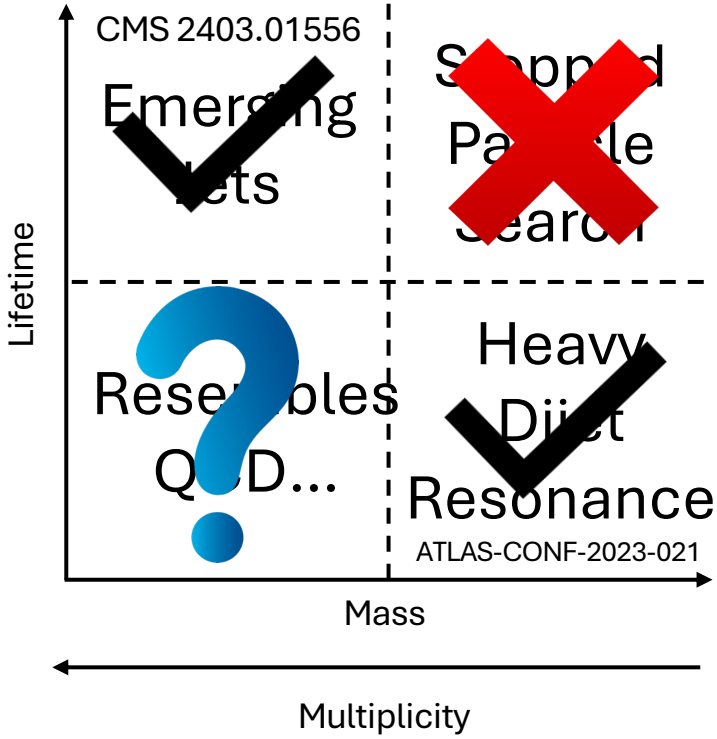


Hidden Valley Signatures



- Large mass with long lifetime means very small couplings
- Negligible particle production at LHC

Hidden Valley Signatures



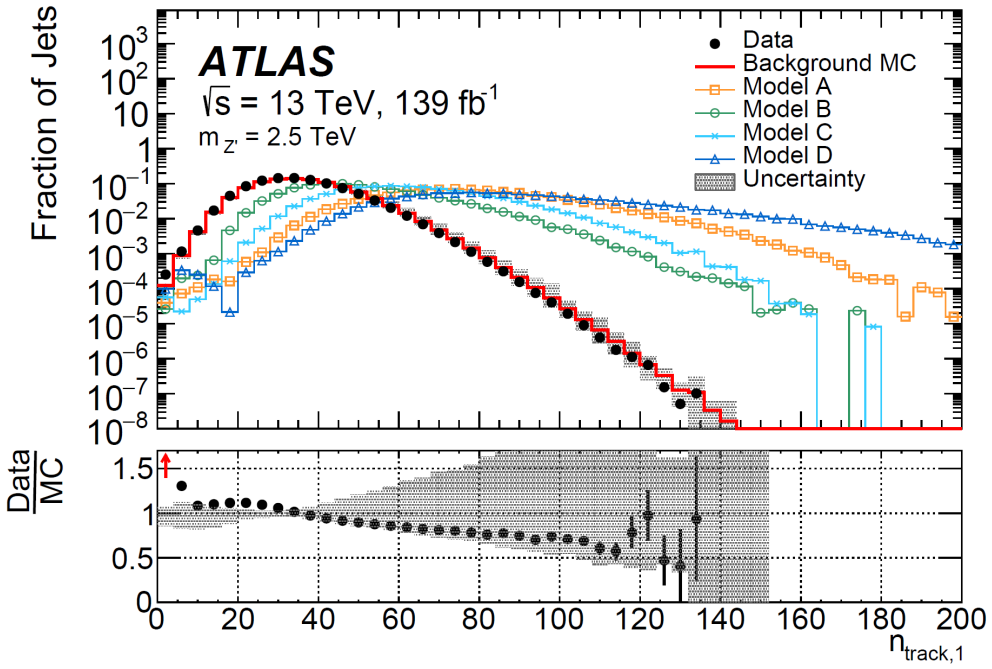
What we want to focus on

- Resembles QCD -
  - This explains difficulty of searches in this area.
  - Our goal is to discern just how closely it resembles QCD and distinguish it
- Combinatoric background –
  - Due to the large number of possible pairings, reconstructing the physical dark pion mass is very challenging

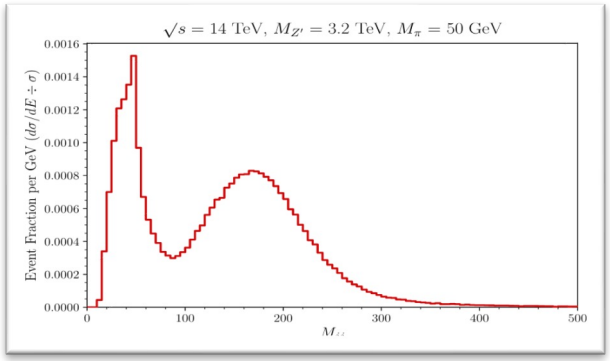
ATLAS Search for Dark Jets

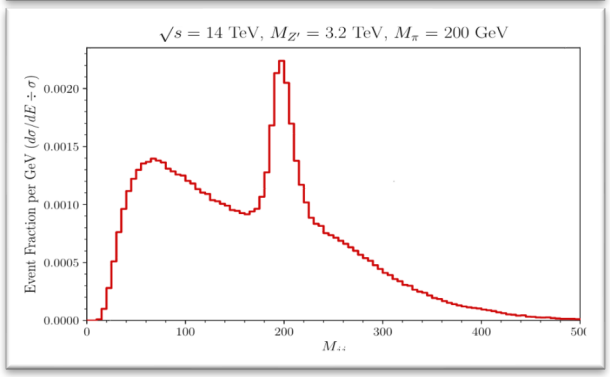
- Search for Z-prime production with decays into 'dark quarks' at ATLAS
- Selects signal events by anomalously large number of tracks in jets at a given P_T
- Aims to reconstruct Z' mass, but not to see the dark pions directly
- A limit was set on the production cross-section times BR as a function of the Z' mass

arXiv:2311.03944

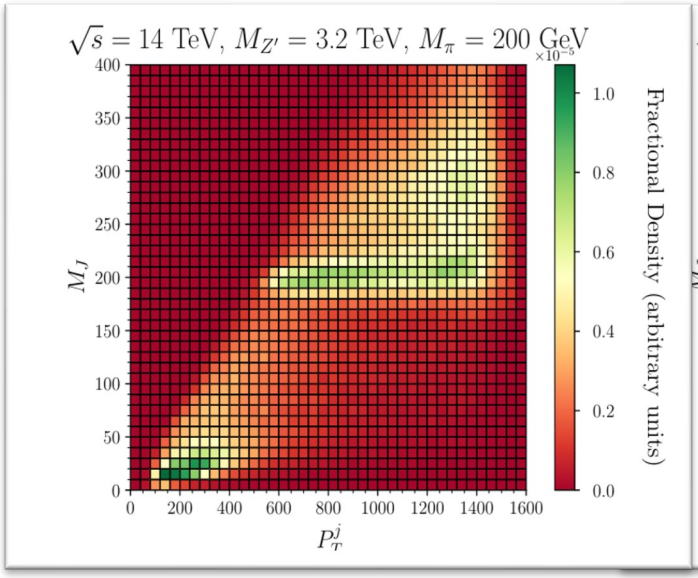


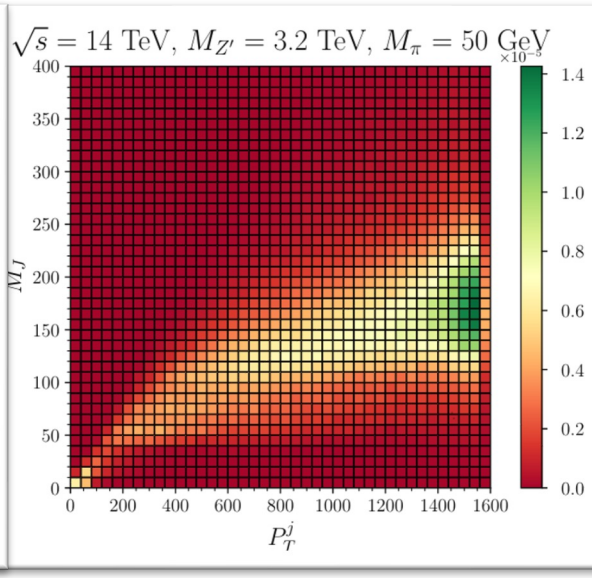
DiJet Resonance Techniques





A la Strassler arXiv:0806.2385





- ❑Assumed that dark pions decay into bottom quarks
- ❑Reconstructed resonances in dijet masses or monojet masses
- ❑This reconstruction works much better for higher masses

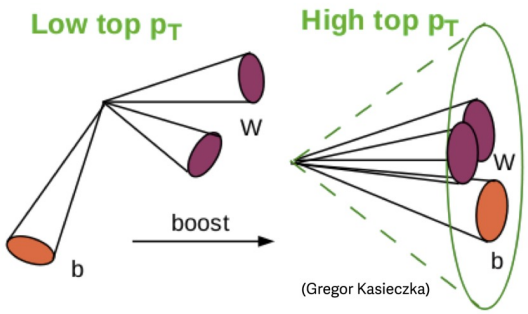
The Combinatoric Problem

- 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.

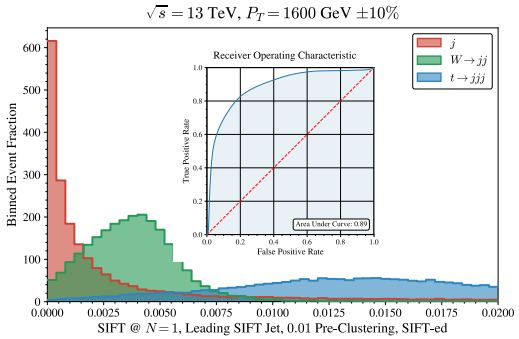
SIFT: Scale-Invariant Filtered Tree

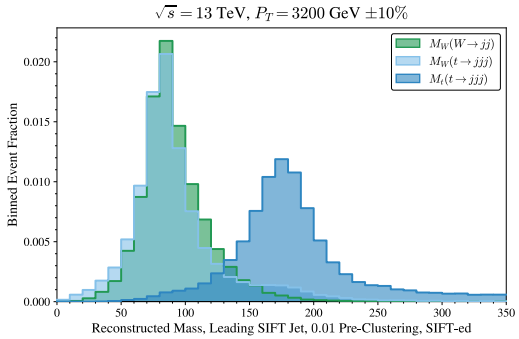


- Massive resonances decay into hard prongs
- Jet definitions with fixed cones impose a scale
- Boosted objects collimate and structure is lost
- Substructure recovery techniques are complex
- Can we avoid losing resolution in the first place?
- Select proximal objects w/ scale-invariant measure

- 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

SIFT-ing for Dark Matter

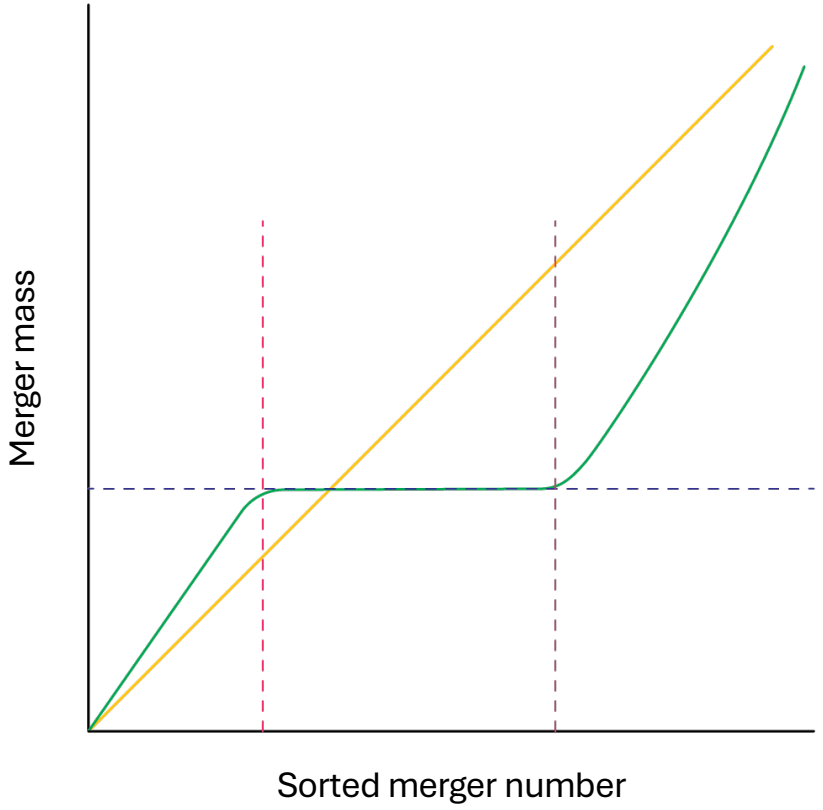
- SIFT, with filtering but without dropping, may be ideal here
- It considers the event as a whole (no cones) with multi-scale sensitivity
- It creates a well-defined sequential SLICE through the combinatorics
- Since hard prongs are merged last, the final mergers are expected to hold relevant physical 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.

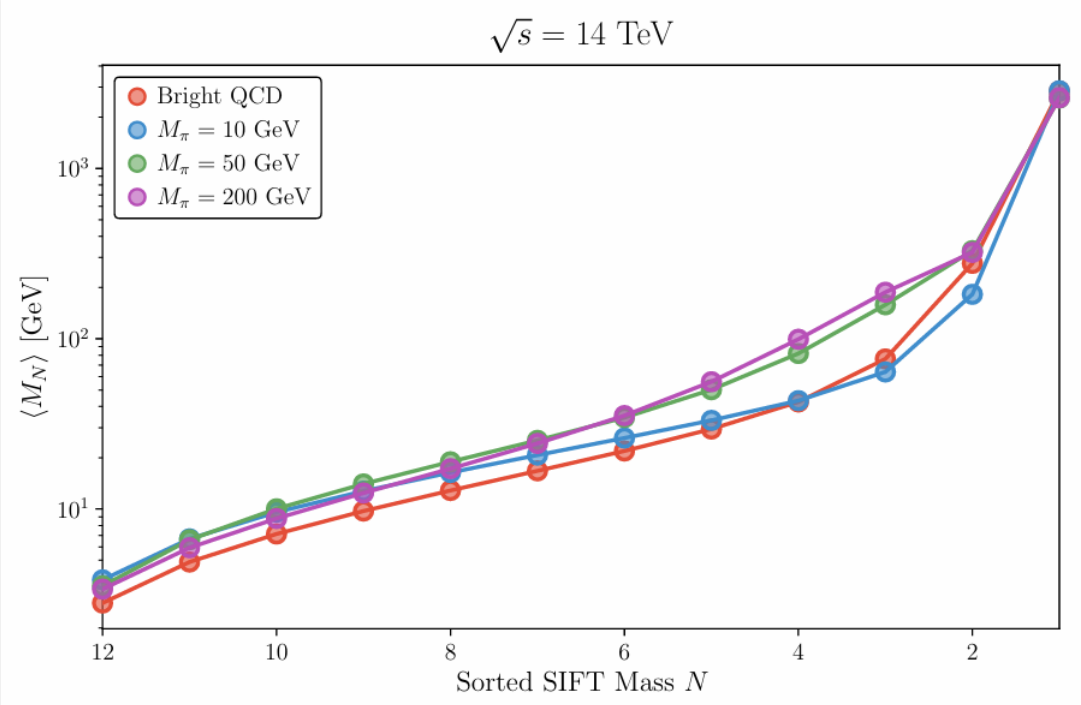
Plateau Investigation of New Scalse (PIONS)

- Cartoon of a log plot of jet algorithm merger masses, sorted by mass

- **Yellow** line shows a typical QCD showering
- **Green** shows a Hidden Valley plus QCD showering
- **Red** upright lines are separated by the hypothetical plateau length N
  - This length is a parameter defining the variable
- Slope of the **Blue** line is the variable we want to search in
  - We calculate the rms of discretized derivatives between mergers to yield the variable Π_N

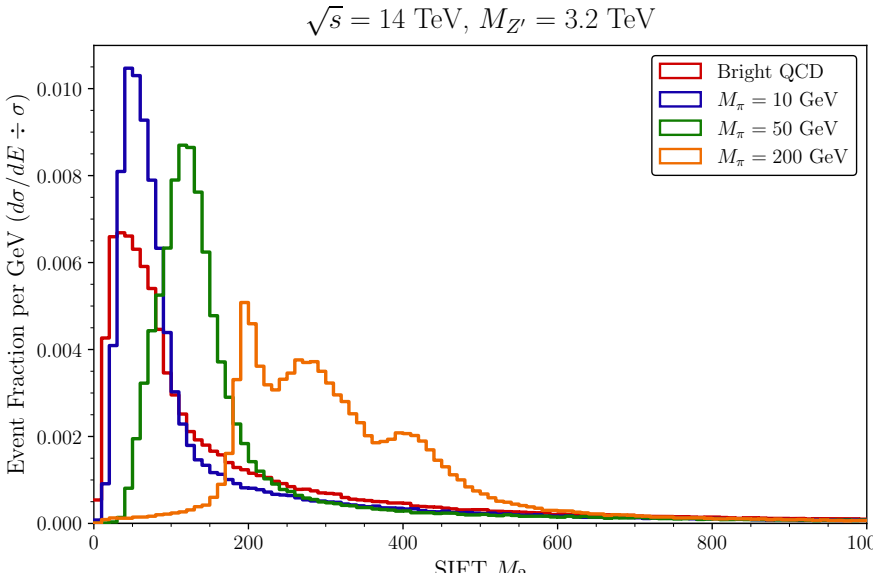


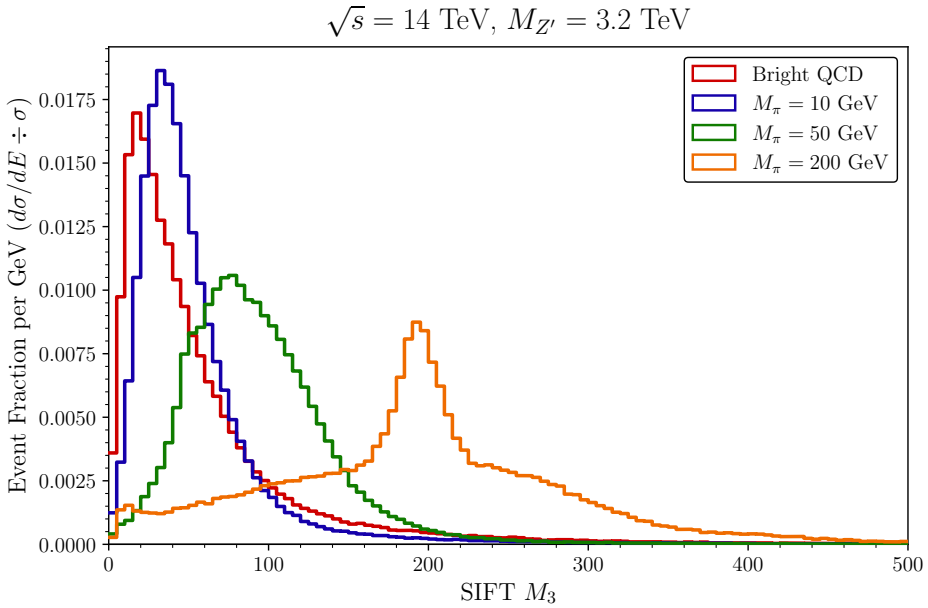
Exploring Merger Masses in Simulation

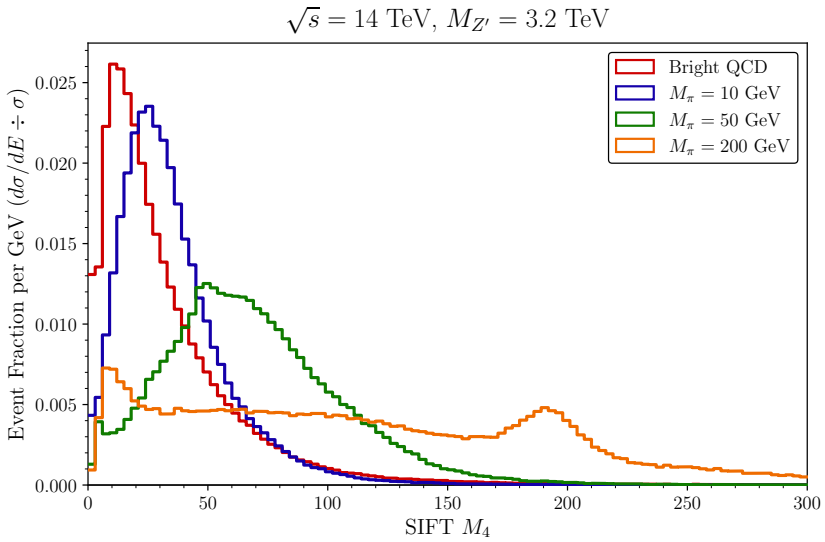


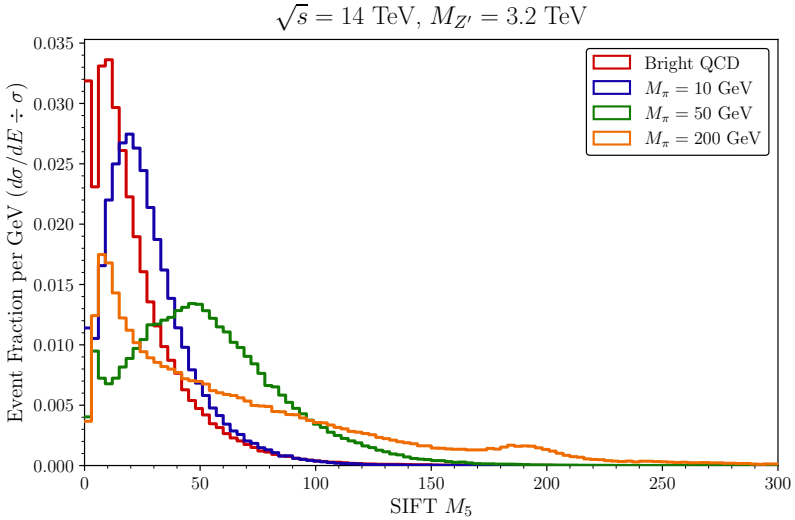
- This is the sorted merger mass plot averaged over a large number of events.
- Note flattening in blue curve relative to red

SIFT does a good job of highlighting Masses



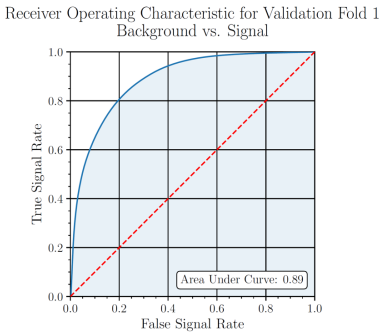


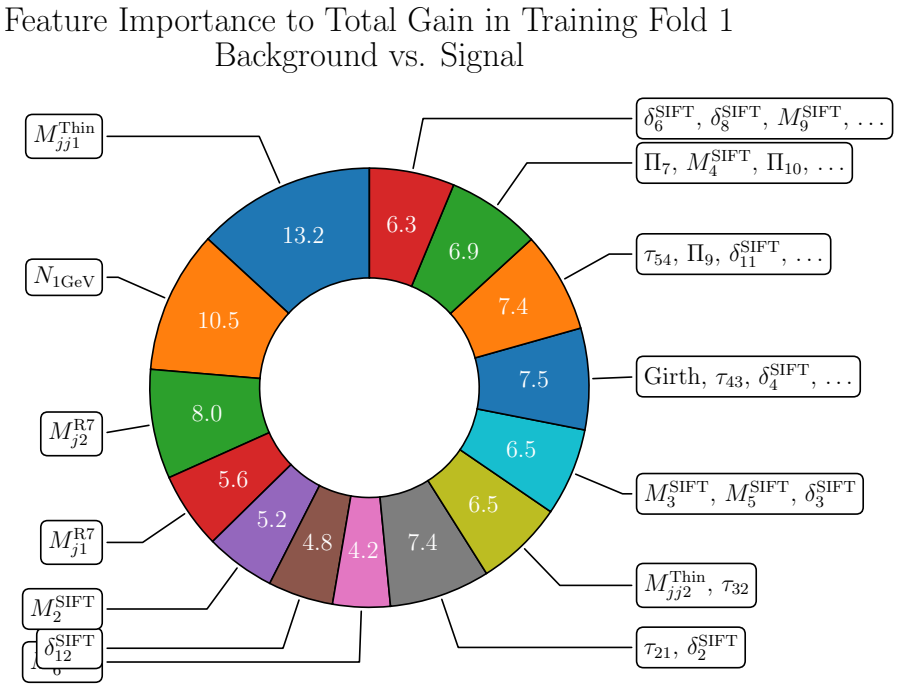




BDT Results, $M_\pi = 10$ GeV – All Variables

- A kitchen-sink analysis utilizes dijet masses, monojet masses, particle counts, and SIFTy variables
- Uses information very similar to ATLAS analysis, improved by addition of the two other approaches
- AUROC score of 89%





Classification Scores

M_π	Strassler All	SIFT All	Classic QCD	Kitchen Sink
10	79	83	77	89
25	89	94	87	96
50	95	98	91	99
120	98	99	93	100
200	99	99	92	100
500	96	99	73	99

- SIFTy technique alone is outcompeting dijet resonance techniques of Strassler and broadly-classified QCD variables alone
- Putting it all together, we have strong classification power throughout the explored parameter space
  - Makes explicit the complementarity of these approaches

Conclusions for Dark Sector Study

- We can achieve this level of discrimination between QCD showers and new dark QCD showers with prompt decays
- We can do that in this regime, where it seems other techniques do not work as well
  - Information from these techniques is complimentary
- This is a valuable expansion of the reach of the LHC into the hidden valley parameter space

SIFT: 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

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

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

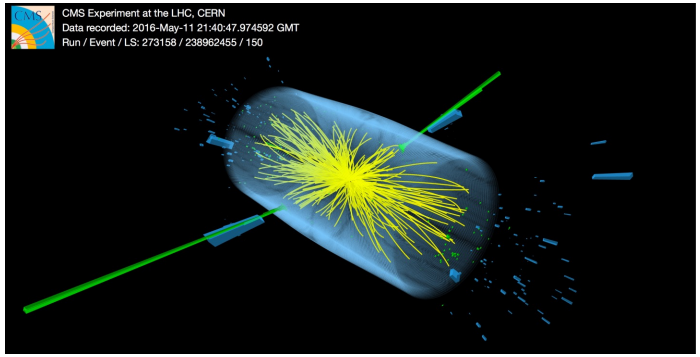


Image: CMS

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

Geometrizing 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

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

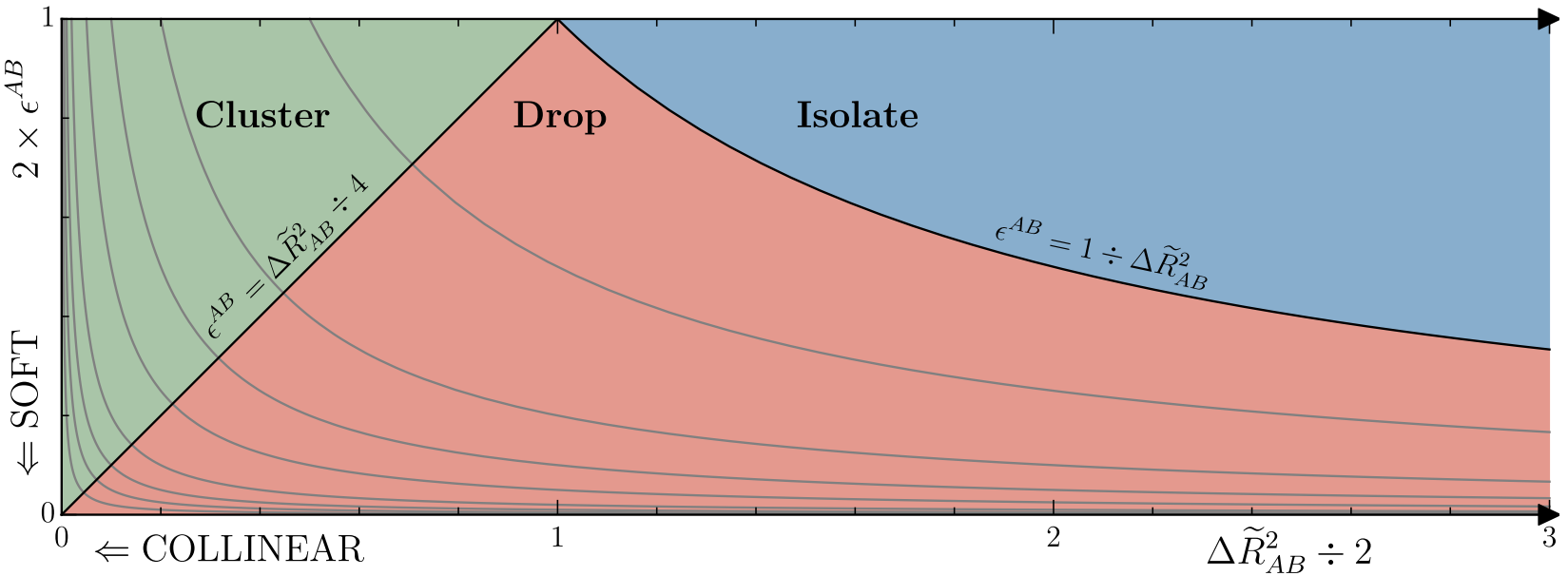
$$(\Delta\widetilde{R}^2_{AB} \gg 1) \text{ and } (\epsilon_{AB} \simeq 1)$$

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

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

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

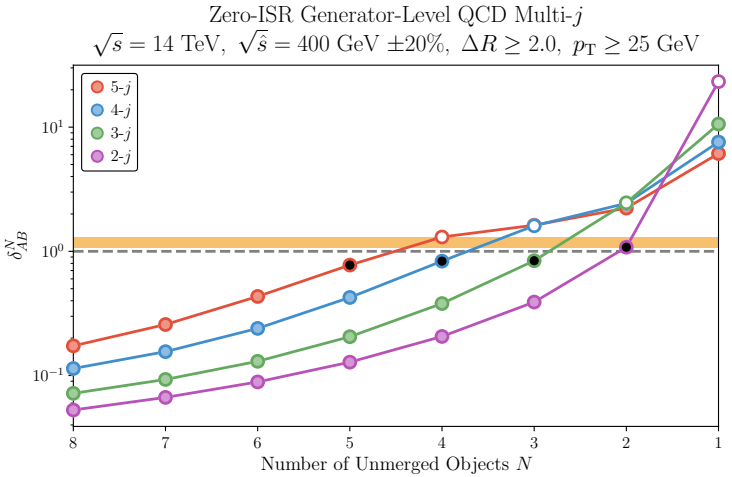
Clustering Phase Diagram

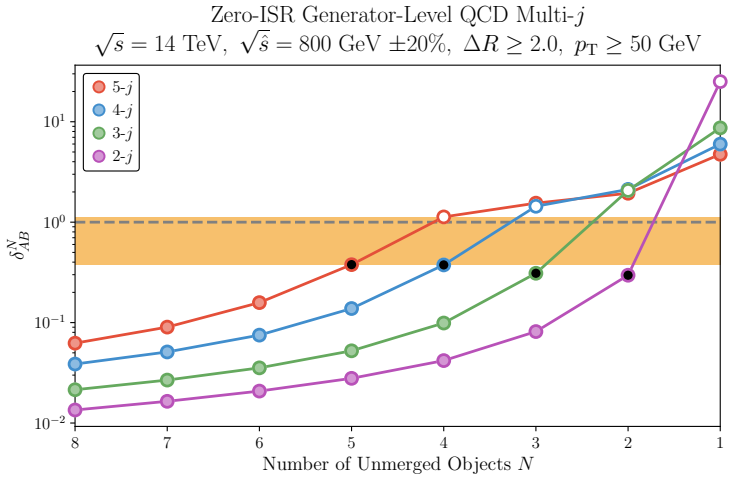


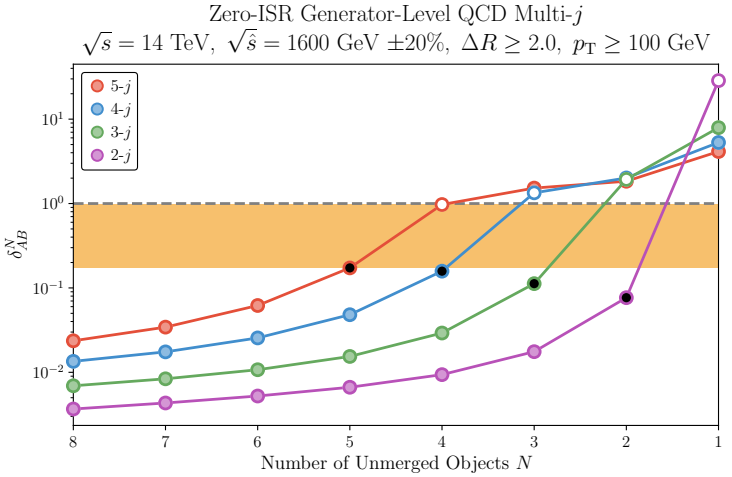
- The unification of clustering, filtering, and isolation also provides natural halting
- Grey contours "$y = \delta/x$" mark constant values of the measure
- Isolation occurs above $\delta = 1$; this amounts finding of variable large-radius jets
- The same factors separate clustering from dropping at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

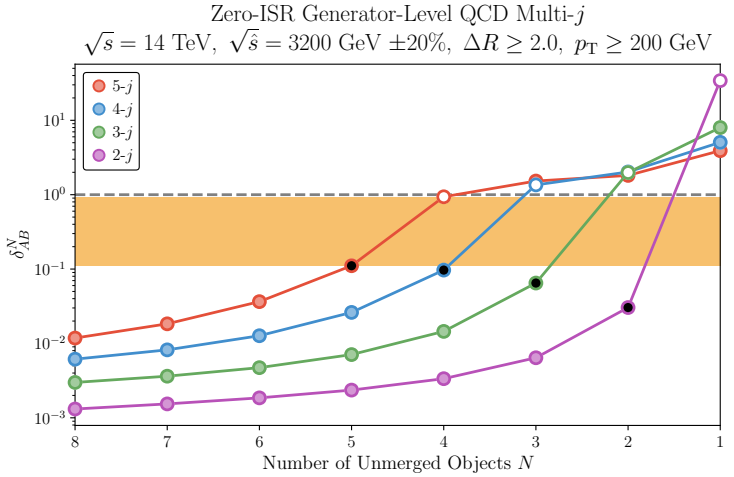
Evolution of the Measure

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









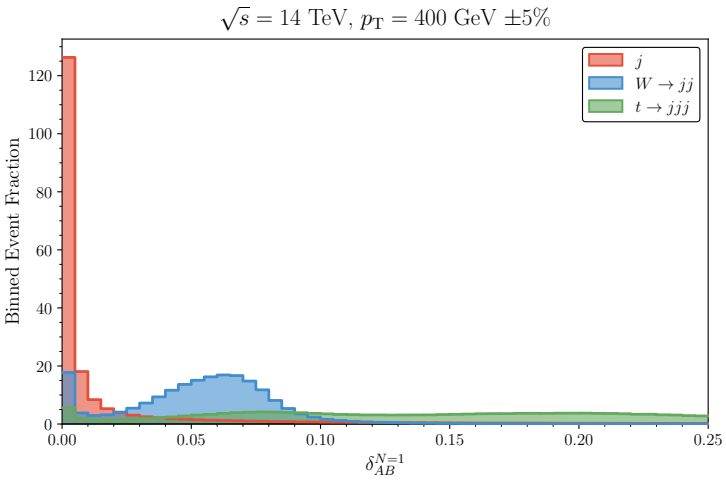
The *N*-Subjet TREE

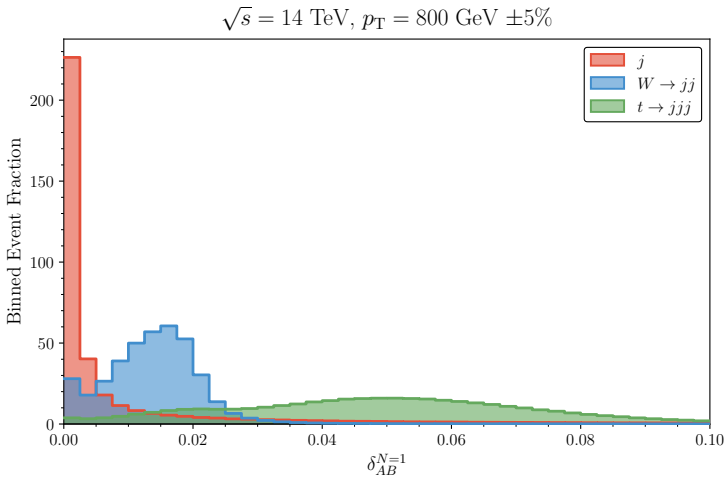
- We observe that:
  - hard structures are preserved
  - wide concentrations of hard objects are isolated
  - soft wide radiation is dropped

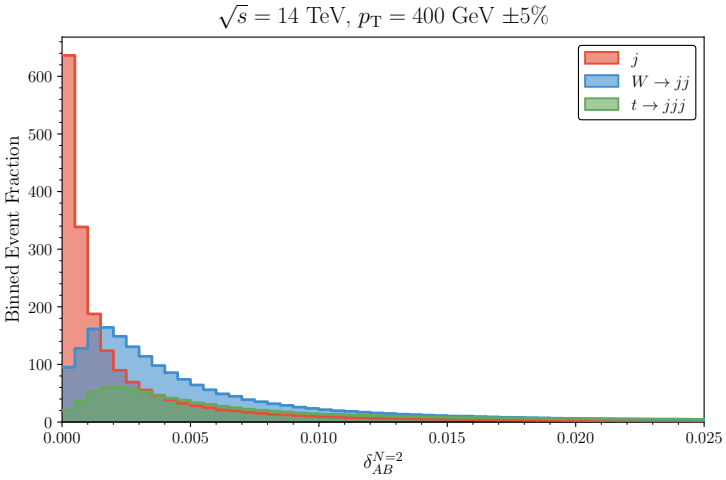
- However, hard prongs within a variable radius jet do still cluster
- How do we fix the interior halting criterion to avoid losing structure?
- The most interesting alternative is to not halt at all …
- We learn more about whether the prongs “want” to merge by merging!
- Hard prongs are the final objects 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

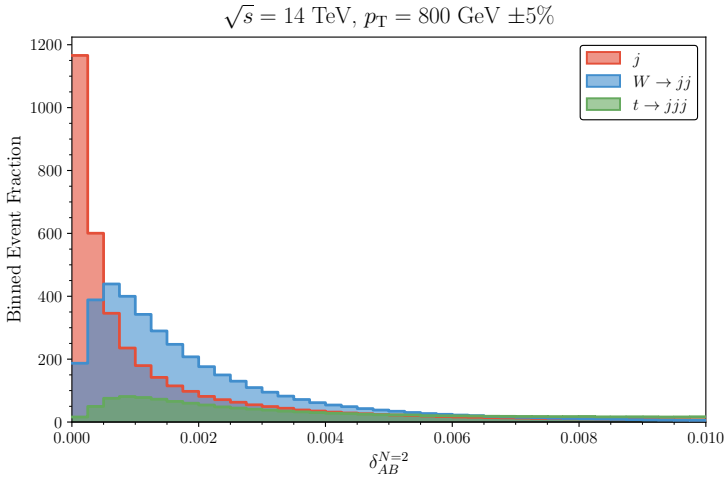
SIFT Measure at Final Mergers

- 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



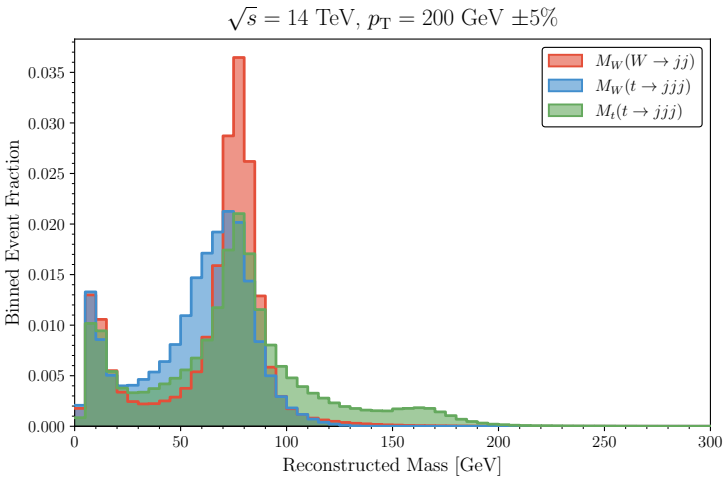


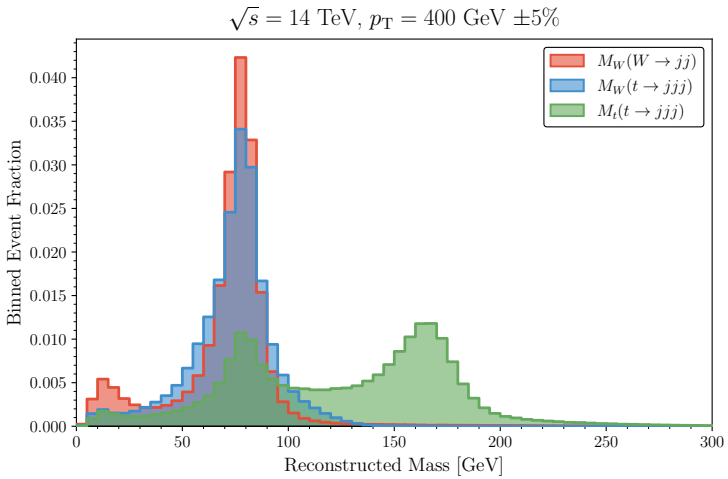


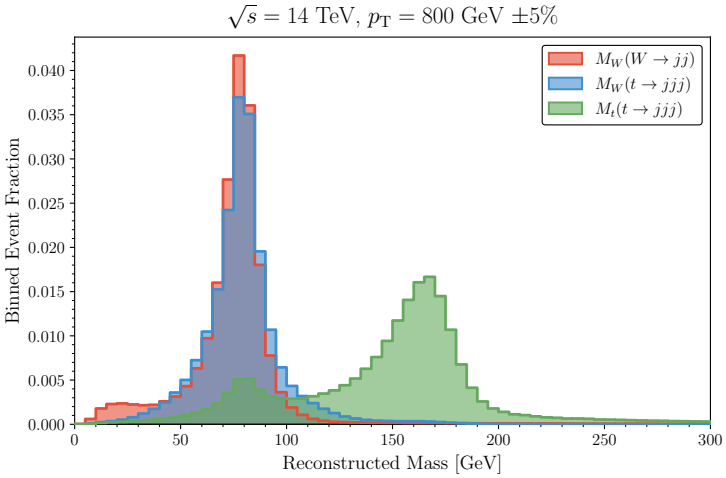


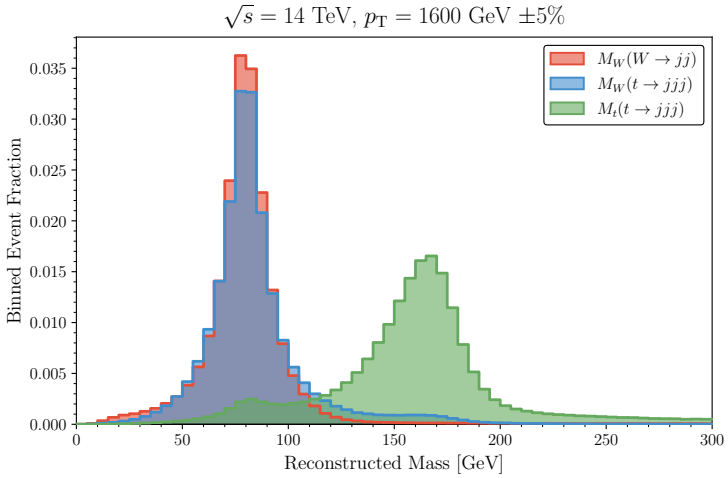
W & top Mass Reconstruction

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

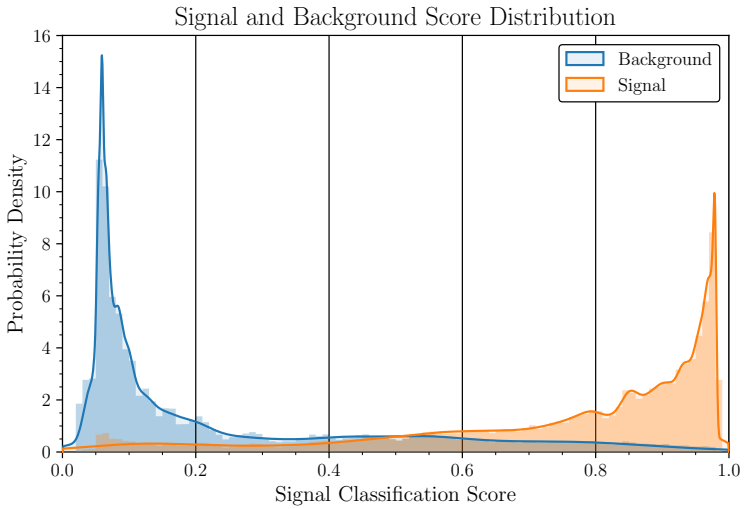






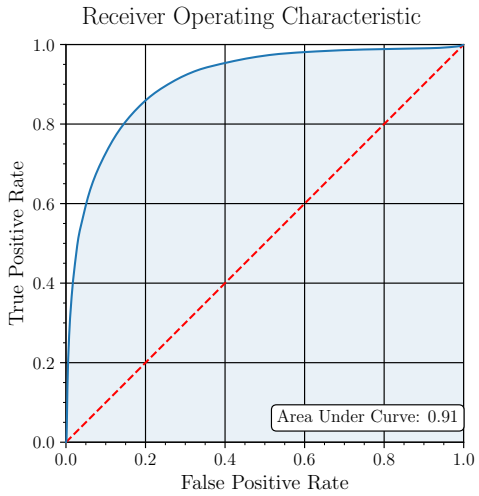


1/2 and 3/2 Discrimination with BDT



$p_T^{\text{GeV}\pm5\%}$	$\tau_{\text{DELPHES}}^{N+1/N}$	$\tau_{\text{SIFT}}^{N+1/N}$	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.



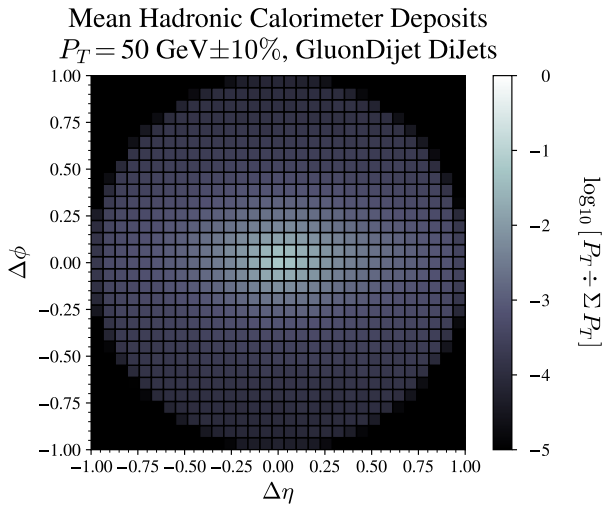
$p_T^{\text{GeV}\pm5\%}$	$\tau_{\text{DELPHES}}^{N+1/N}$	$\tau_{\text{SIFT}}^{N+1/N}$	δ_{AB}^{N}	$\delta+\tau$
100	0.61	0.61	0.63	0.65
200	0.63	0.60	0.71	0.72
400	0.82	0.74	0.90	0.90
800	0.85	0.80	0.94	0.95
1600	0.77	0.77	0.97	0.97
3200	0.77	0.79	0.98	0.99

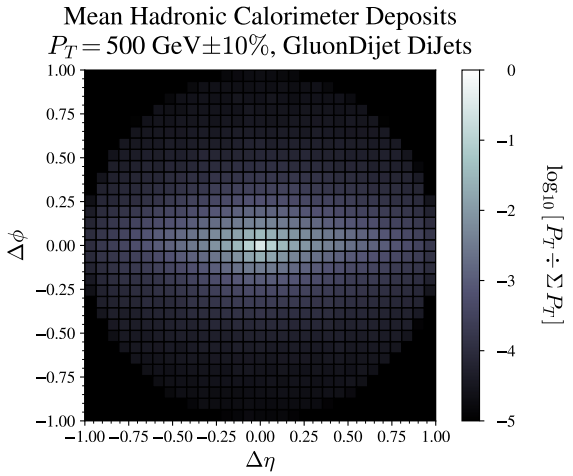
TABLE IV. Area under curve ROC scores for discrimination of resonances with hard 2- and 3-prong substructure using a BDT trained on various sets of event observables.

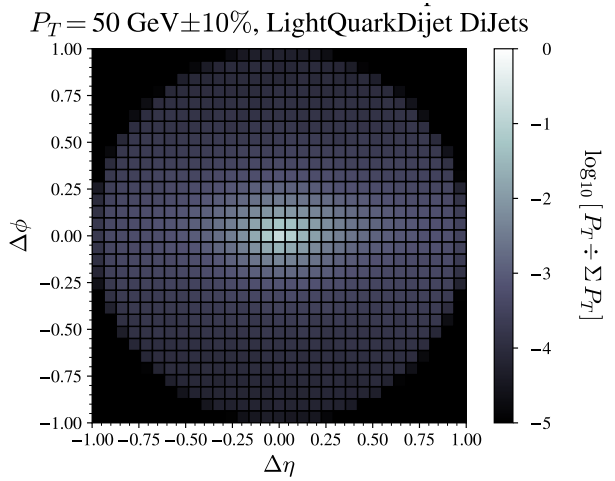
Other Applications: Q/G discrimination

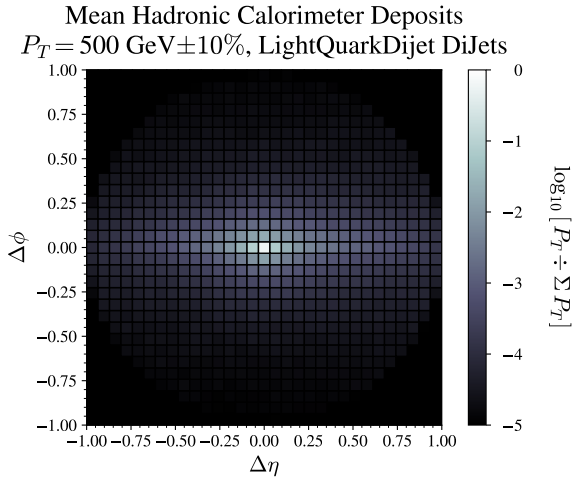
- An image-based approach to discriminating quark- and gluon- initiated jets
- Plots are an ENSEMBLE after normalization and rotation, etc.

(Dutta, Kamon, Kim, P.R. Kumar, B. Lei, B. Mallick, S. Sinha, JWW)









Other Applications: Q/G discrimination

- Neural networks can hit ~ 80% discrimination by AUC

Table 2: AUC values% in each energy level.

Particle \ *Energy*	25	50	100	250	500	750	1000	1250	1500	2000
ResNet	77.77	76.05	79.32	80.43	80.94	81.42	81.88	81.67	81.75	82.09
ResNet + side	80.42	**76.64**	79.71	80.46	79.34	80.96	80.41	78.34	78.28	78.49
XGBoost (side only)	75.15	70.69	74.39	75.32	75.22	75.56	75.88	75.51	75.67	76.04
BART (side only)	78.26	74.21	77.43	78	77.85	78.08	78.28	78.24	78.18	78.49
Autoencoder + BART	**80.44**	76.39	**80.05**	80.74	**81.45**	**81.75**	**82.38**	**81.9**	**81.93**	**82.58**
Autoencoder + XGBoost	77.96	73.29	77.51	78.3	79.1	79.15	80.03	79.7	79.49	80.35
PCA + BART	77.51	67.21	70.68	71.53	71.9	72.05	72.55	72.68	73.29	74.07
PCA + XGBoost	75.36	64.94	68.34	69.18	70.13	70.5	71.42	71.94	71.53	72.34
Autoencoder	80.34	76.29	79.69	**80.93**	**81.30**	**81.64**	**81.72**	**81.81**	81.76	**82.54**
Variational Autoencoder	**80.65**	76.36	79.81	80.83	81.23	81.56	81.63	81.76	**81.95**	82.49
Autoencoder + side	80.38	**76.52**	**79.87**	**80.89**	80.81	81.3	81.56	81.13	81.56	82.03

- SIFT + BDT hits 80-83% at the 50 GeV benchmark

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

Physics Bootcamp

Mathematical Methods for First-Year Physics and Engineering

JAMES B. DENT AND JOEL W. WALKER
SAM HOUSTON STATE UNIVERSITY

FOR PUBLICATION BY
CAMBRIDGE UNIVERSITY PRESS
DRAFT: June 23, 2024

- Math for the first 1-2 years of physics, from our perspective, in context
- I: Arithmetic, II: Algebra, III: Trig & Vectors, IV: Differentiation, V: Integration
- For HS students and incoming majors, during (or before) the 1st semester
- We're distributing samples now – please email to get on the list!
- jwalker@shsu.edu & jbdent@shsu.edu