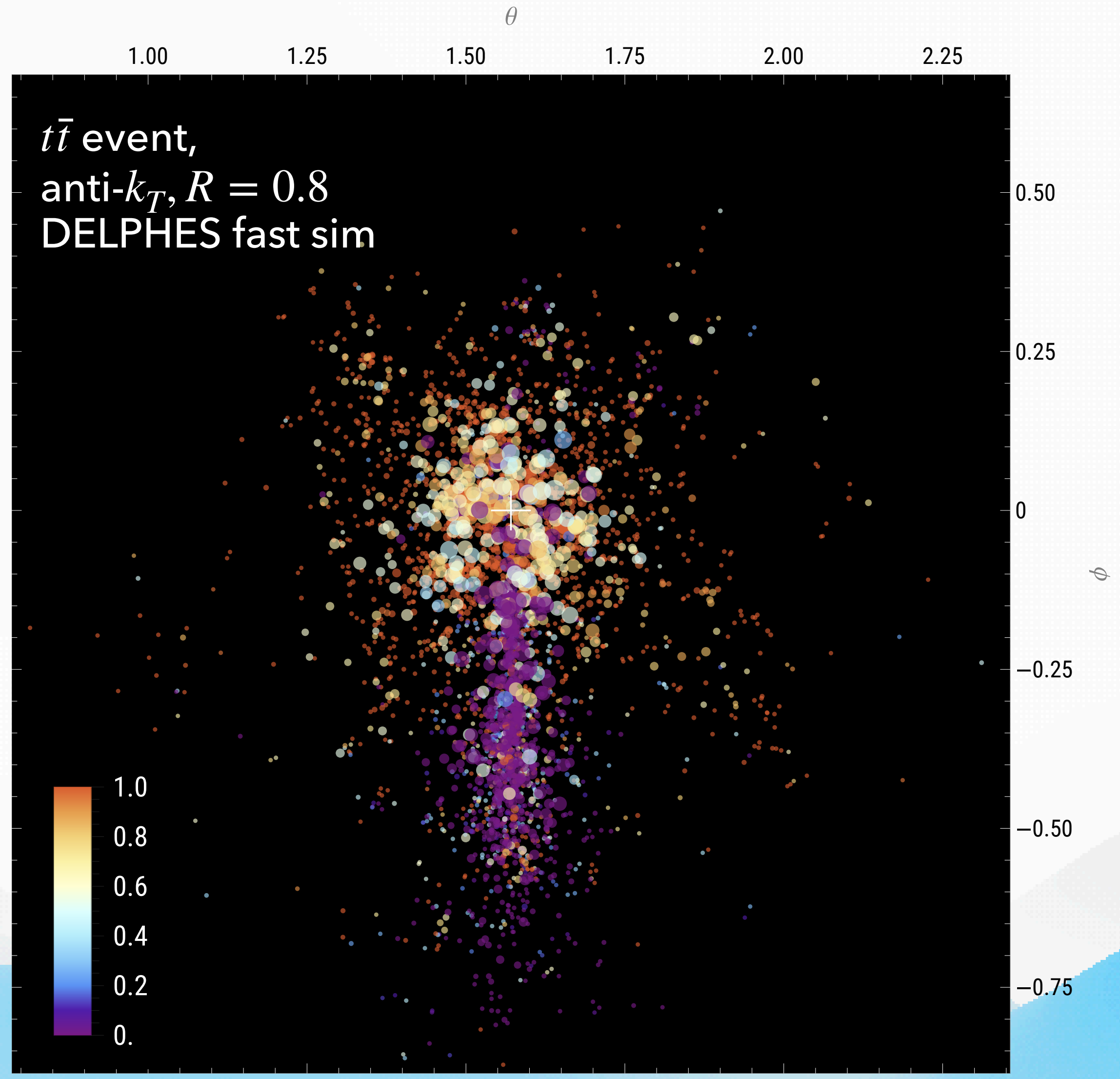
BOOST 2023 Indico

PELICAN

PERMUTATION- AND LORENTZ-EQUIVARIANT NETWORKS FOR PARTICLE PHYSICS



Identifying constituent-level W-boson decay products within a top quark jet!

Alexander Bogatskiy, Timothy Hoffman,
Xiaoyang Liu, David W. Miller, Jan T. Offermann

BOOST 2023

August 1, 2023

ENRICO FERMI INSTITUTE

THE UNIVERSITY OF CHICAGO
DATA SCIENCE INSTITUTE

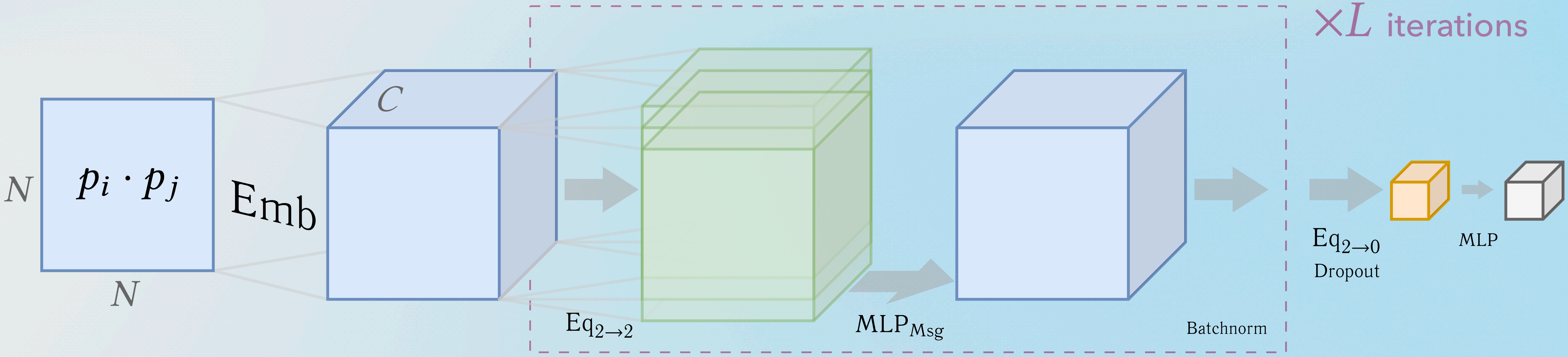
Miller Lab

FLATIRON INSTITUTE

- Tagging jets is fundamentally important for measurements and searches! Neural networks have proven useful for this.
- *Equivariant* networks are strong, robust and lean.
  - Relevant here: permutation & Lorentz equivariance. [a,b]
  - Beyond tagging: subjet structure and constituent info!
- Growing interest in general approaches for Lorentz-equivariant NN's. [c] *(low parameterization, explainability)*
- For today: PELICAN for top-tagging, and measuring W-boson four-momentum *within* top quark jets.

[a] A Bogatskiy, T. Hoffman, D. W. Miller, J. T. Offermann, PELICAN (2022).
[b] S. Gong et. al., An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging (2022).
[c] A. Bogatskiy et. al., Symmetry Group Equivariant Architectures for Physics (2022).

- PELICAN is a network architecture that…
  - is permutation- and Lorentz-equivariant,
  - acts on collections of four-momenta, (c.f. point clouds)
  - and can predict Lorentz-*invariants*, e.g. tagging…
    - ***or Lorentz-covariants*, e.g. p^μ regression!**



THE PELICAN ARCHITECTURE: LORENTZ EQUIVARIANCE

- Lorentz equivariance is achieved by constructing functions of inner products of the four-momenta.

Invariants (output for PELICAN's "scalar form")

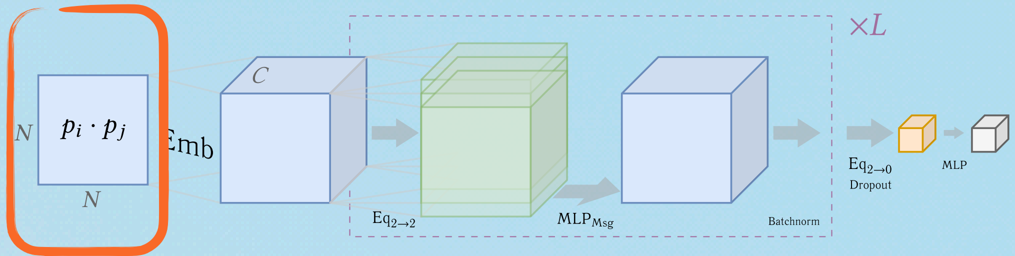
$$I_k\left(p_1, \ldots, p_N\right) = I_k\left(\left\{d_{ij}\right\}_{i,j}\right) \quad , \qquad d_{ij} \equiv p_i^\mu p_{j,\mu}$$

Lorentz-invariant

Covariants (output for PELICAN's "vector form")

$$F^\mu = \sum_k I_k\left(p_1, \ldots, p_N\right) p_k^\mu$$

Lorentz-covariant



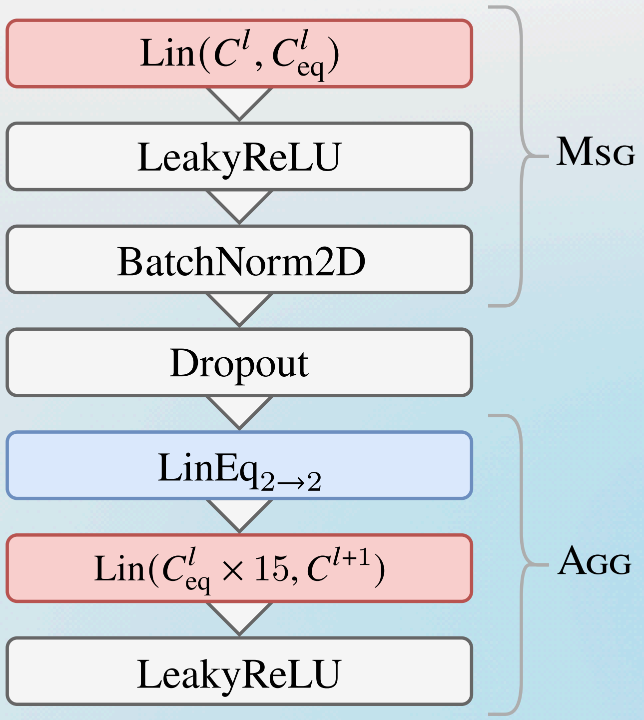
THE PELICAN ARCHITECTURE: PERMUTATION EQUIVARIANCE

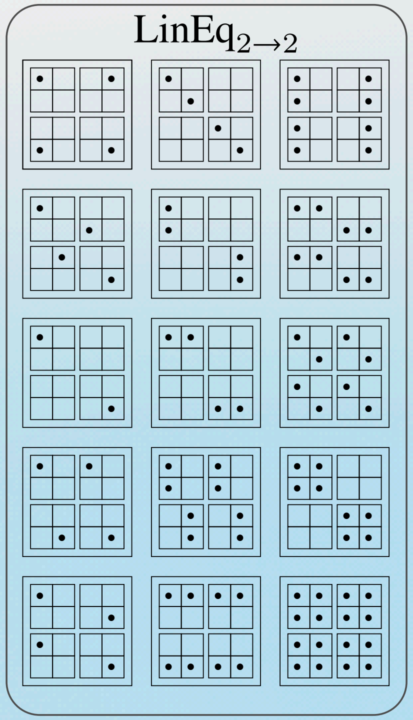
- Operations are equivariant w.r.t. particle permutations.

Equivariant $2 \to 2$ block:

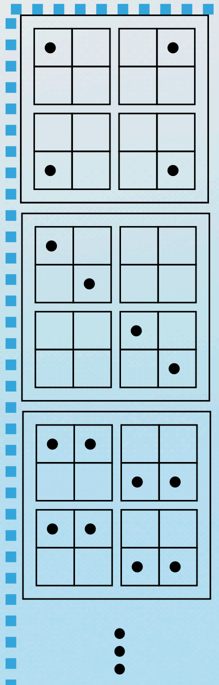
$$T^{(l+1)} = \text{Agg} \circ \text{Msg}\left(T^{l}\right)$$

- 15-dimensional linear permutation-equivariant map.
- Output tensor elements are aggregated from input tensor elements.





where…

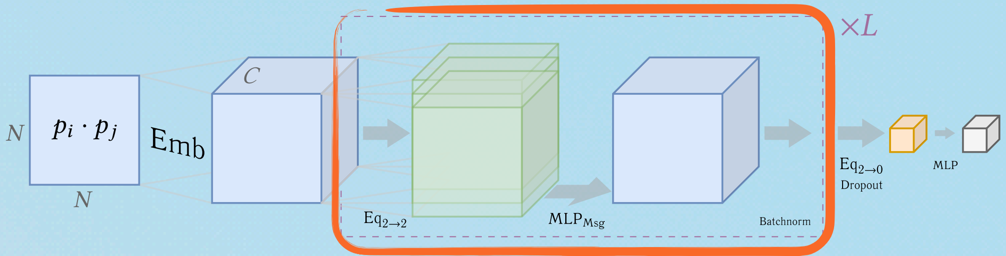


= identity

= trace (on diagonal)

= aggregate rows

et cetera…

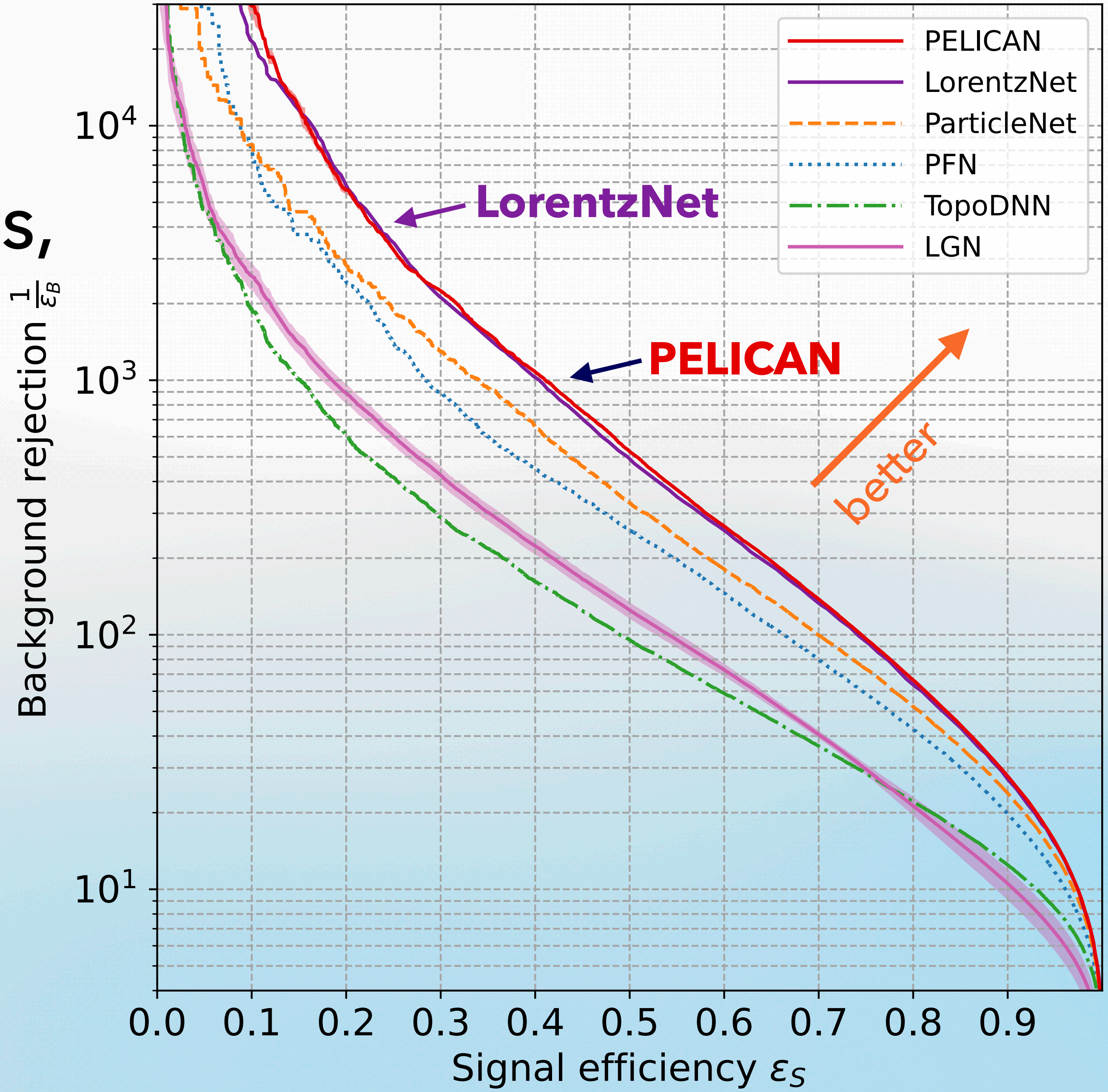


PELICAN: TOP-TAGGING

First task: top tagging!

- Hadronic tops vs. QCD dijets, anti-k_T $R = 0.8$ jets. [d]
- Interesting problem with existing ML benchmarks.
- Use PELICAN's scalar form.
    - State-of-the-art performance, even with low parameterization! [a]

Scalar form: $I\left(p_1, \ldots, p_N\right) = I\left(\left\{d_{ij}\right\}_{i,j}\right)$



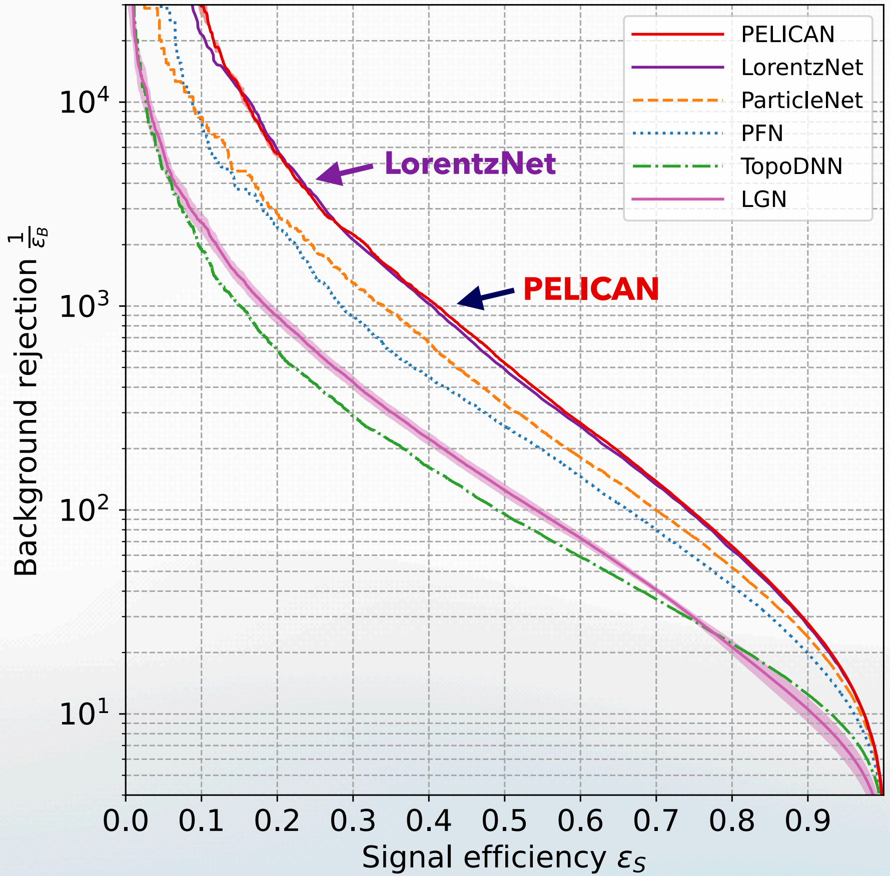
Architecture	Accuracy	AUC	# Params
PELICAN	0.9425(1)	0.9869(1)	45k
LorentzNet	0.942	0.9868	220k
ParticleNet	0.938	0.985	498k
PFN	0.932	0.982	82k
TopoDNN	0.916	0.972	59k
LGN	0.929	0.964	4.5k

[a] A Bogatskiy, T. Hoffman, D. W. Miller, J. T. Offermann, PELICAN (2022).

[d] G. Kasieczka, T. Plehn, J. Thompson, and M. Russel (2019).

PELICAN: TOP-TAGGING

- State-of-the-art performance, even with low parameterization!

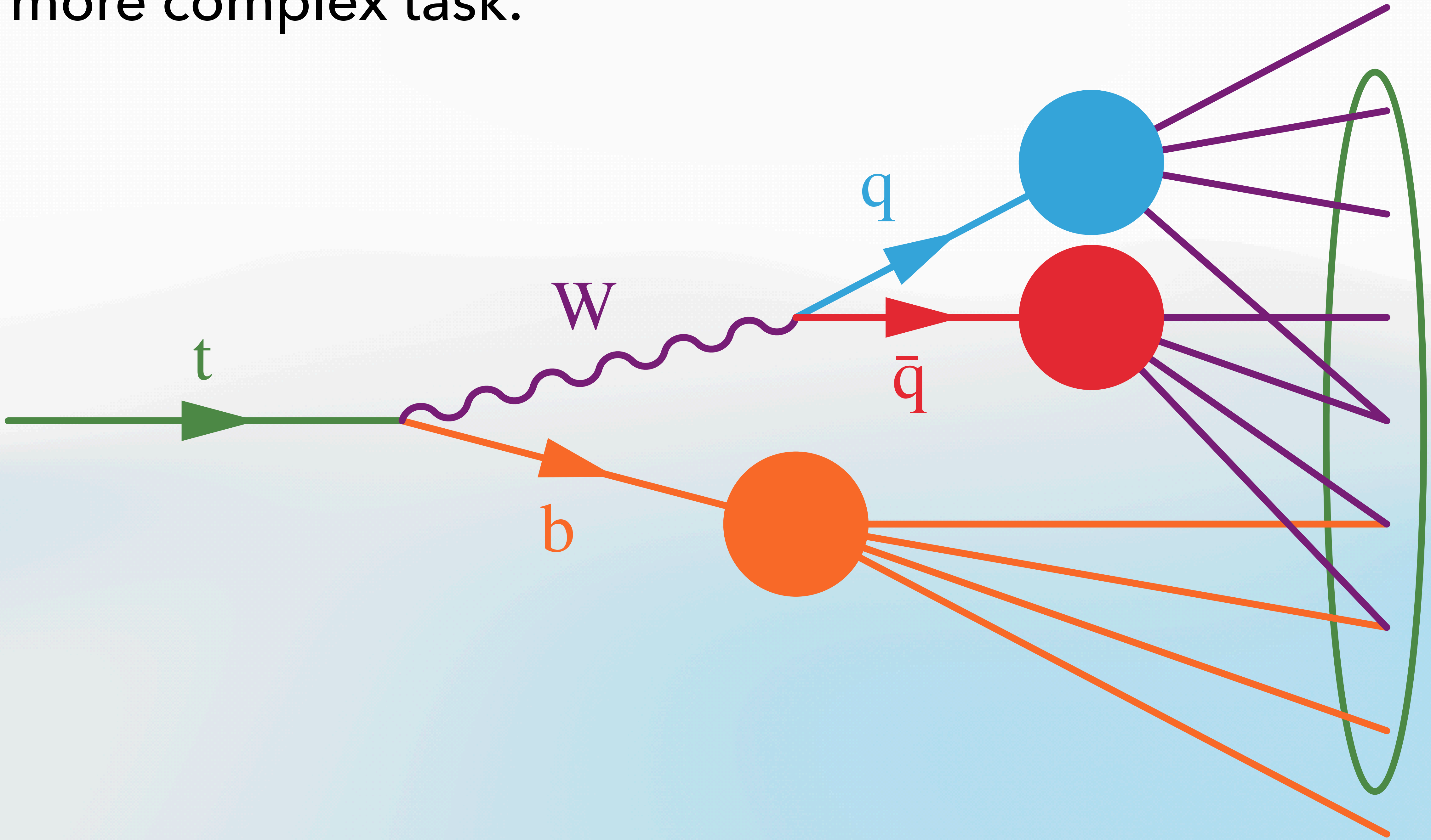


[a] A Bogatskiy, T. Hoffman, D. W. Miller, J. T. Offermann, PELICAN (2022).

[d] G. Kasieczka, T. Plehn, J. Thompson, and M. Russel (2019).

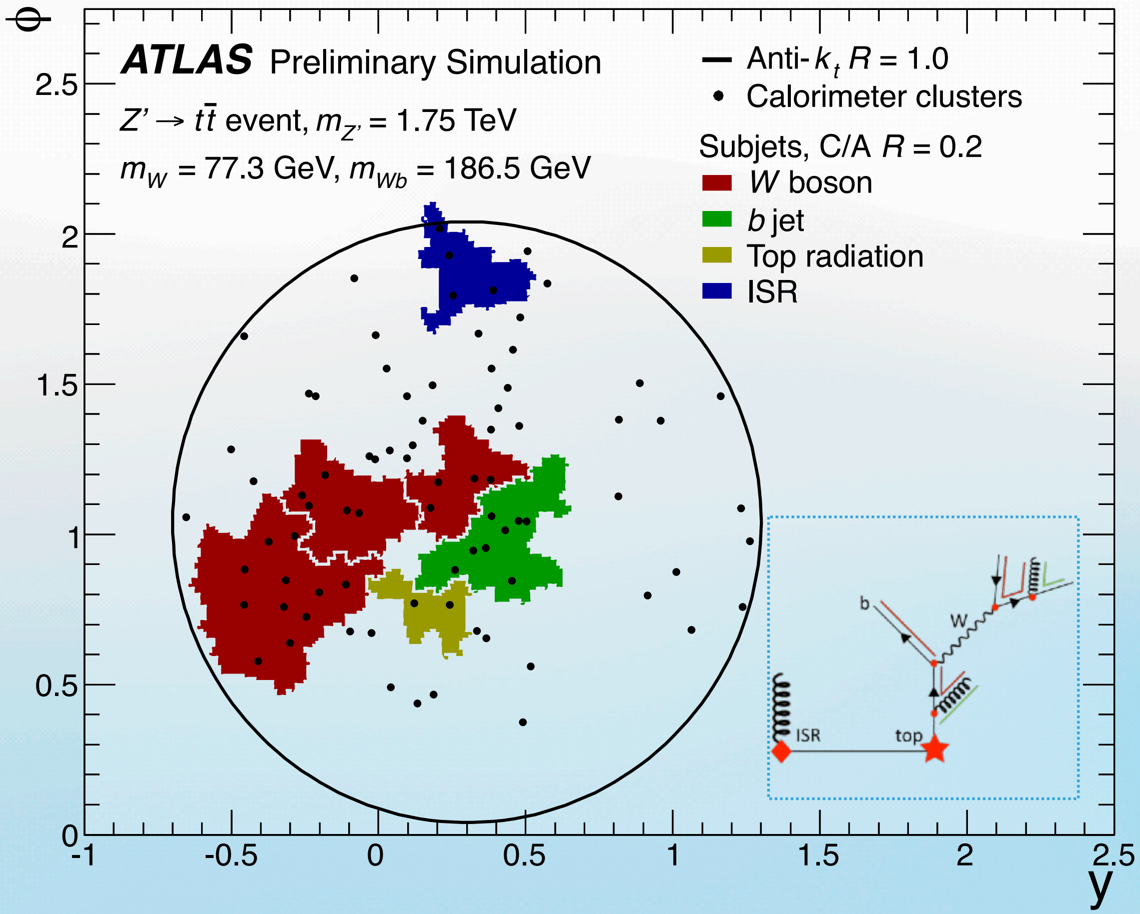
Architecture	Accuracy	AUC	# Params
PELICAN	0.9425(1)	0.9869(1)	45k
LorentzNet	0.942	0.9868	220k
ParticleNet	0.938	0.985	498k
PFN	0.932	0.982	82k
TopoDNN	0.916	0.972	59k
LGN	0.929	0.964	4.5k

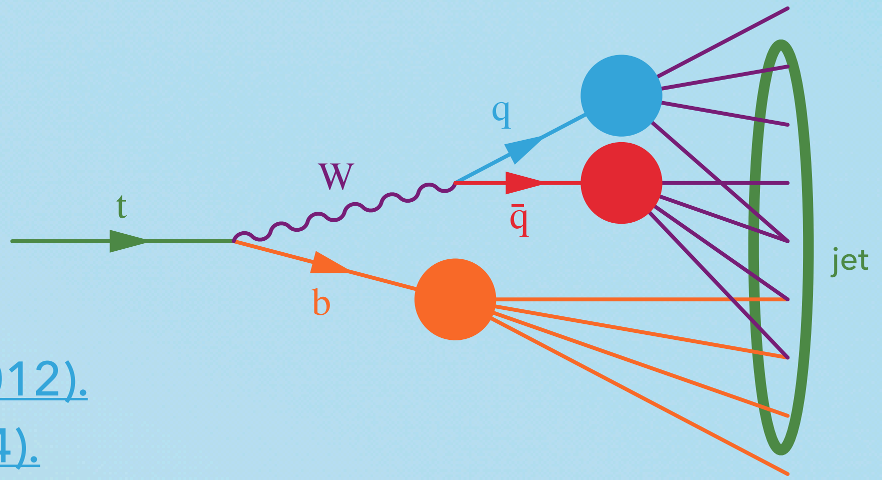
A more complex task:



Given a top quark jet from fully-hadronic top decay, can we measure properties of the *W*-boson, such as its momentum?

- Analogous to *shower deconstruction*: jet components are associated with elements of top decay/shower. [e,f]
- In our task, identifying part of the jet as the W-boson showering products is implicit.
    - Deconstruction works at the subjet level – we will work at the *constituent* level!
- Our goal is to *measure properties* of the W-boson!

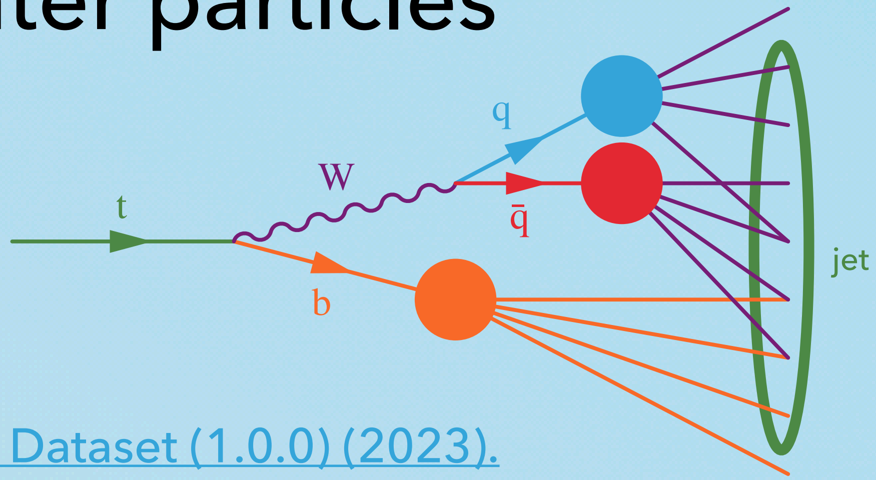




[e] D. E. Soper, M. Spannowsky, Finding top quarks with shower deconstruction (2012).

[f] The ATLAS collaboration, Performance of shower deconstruction in ATLAS (2014).

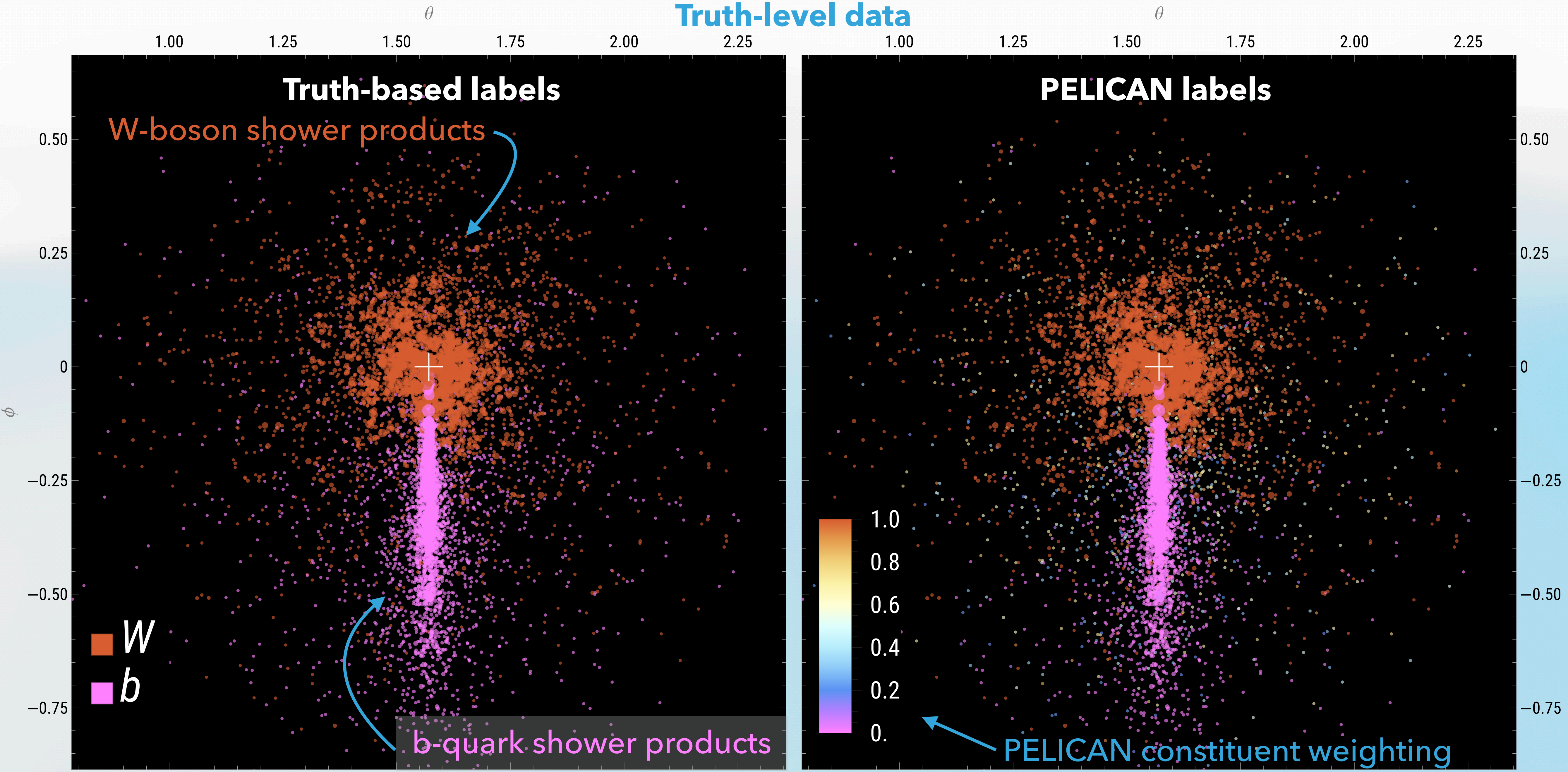
- Use PELICAN's vector form: $F^{\mu} = \sum_{k} I_k\left(p_1, \ldots, p_N\right) p_k^{\mu}$
- Fully-hadronic tops, anti-k_T $R = 0.8$ jets. [g]
  - Nominal & variable true W mass: $m_W^0 \in [64,96]$ GeV.
- Predicted W-boson momentum is a *weighted* sum of jet constituent momenta.
- At truth-level, weights I_k should be binary *(is / is not W daughter)*.
  - W momentum is the sum of its daughter particles' momenta.



[g] J. T. Offermann, T. Hoffman, A. Bogatskiy, Top Jet W-Momentum Reconstruction Dataset (1.0.0) (2023).

PELICAN: MOMENTUM REGRESSION

- Aggregated jet displays, centered on truth-level W.
- PELICAN weights each constituent in producing W 4-momentum – interpretable as likelihood of being W daughter!



Truth-level data

Truth-based labels

W-boson shower products

b-quark shower products

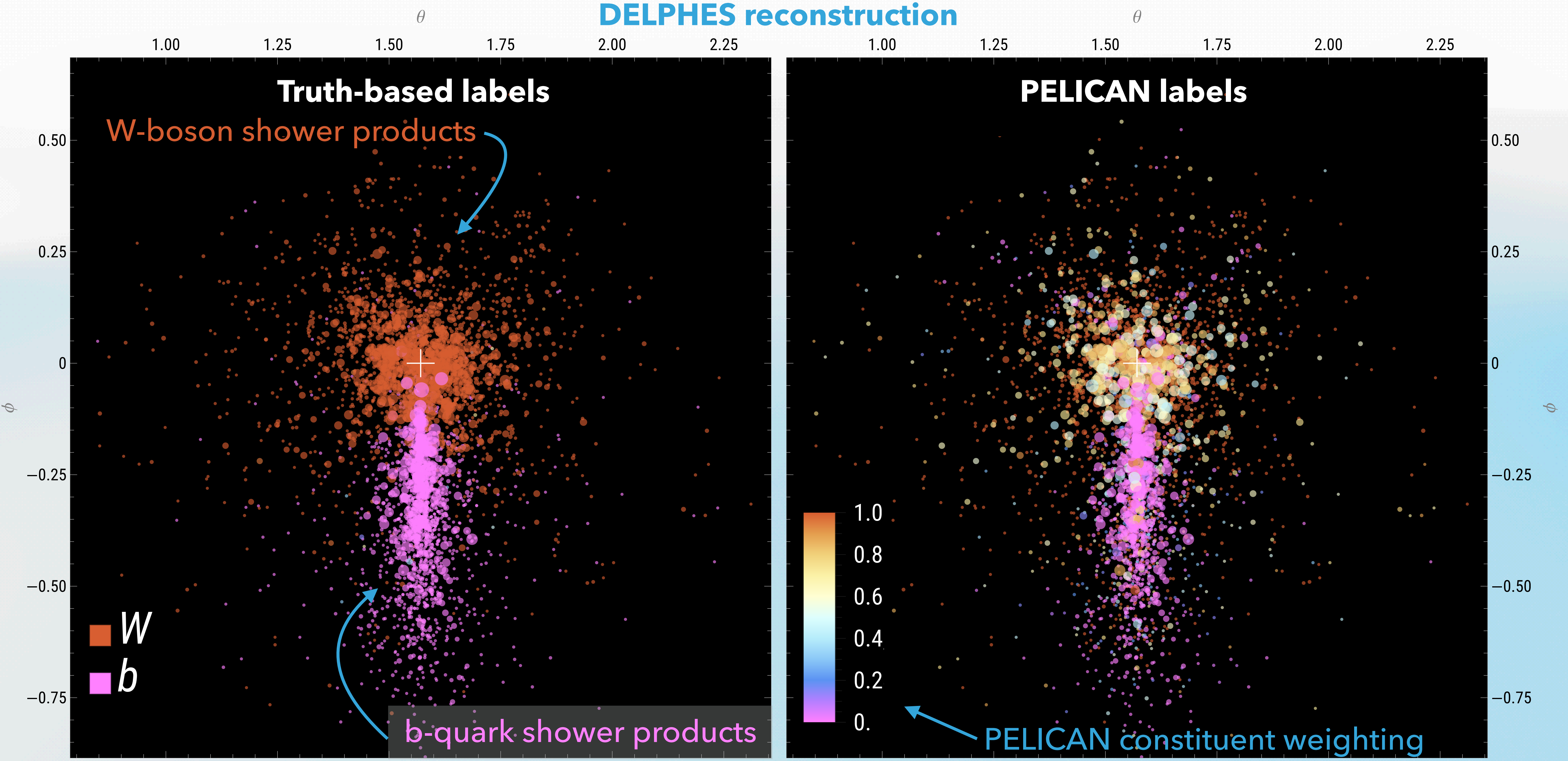
PELICAN labels

PELICAN constituent weighting

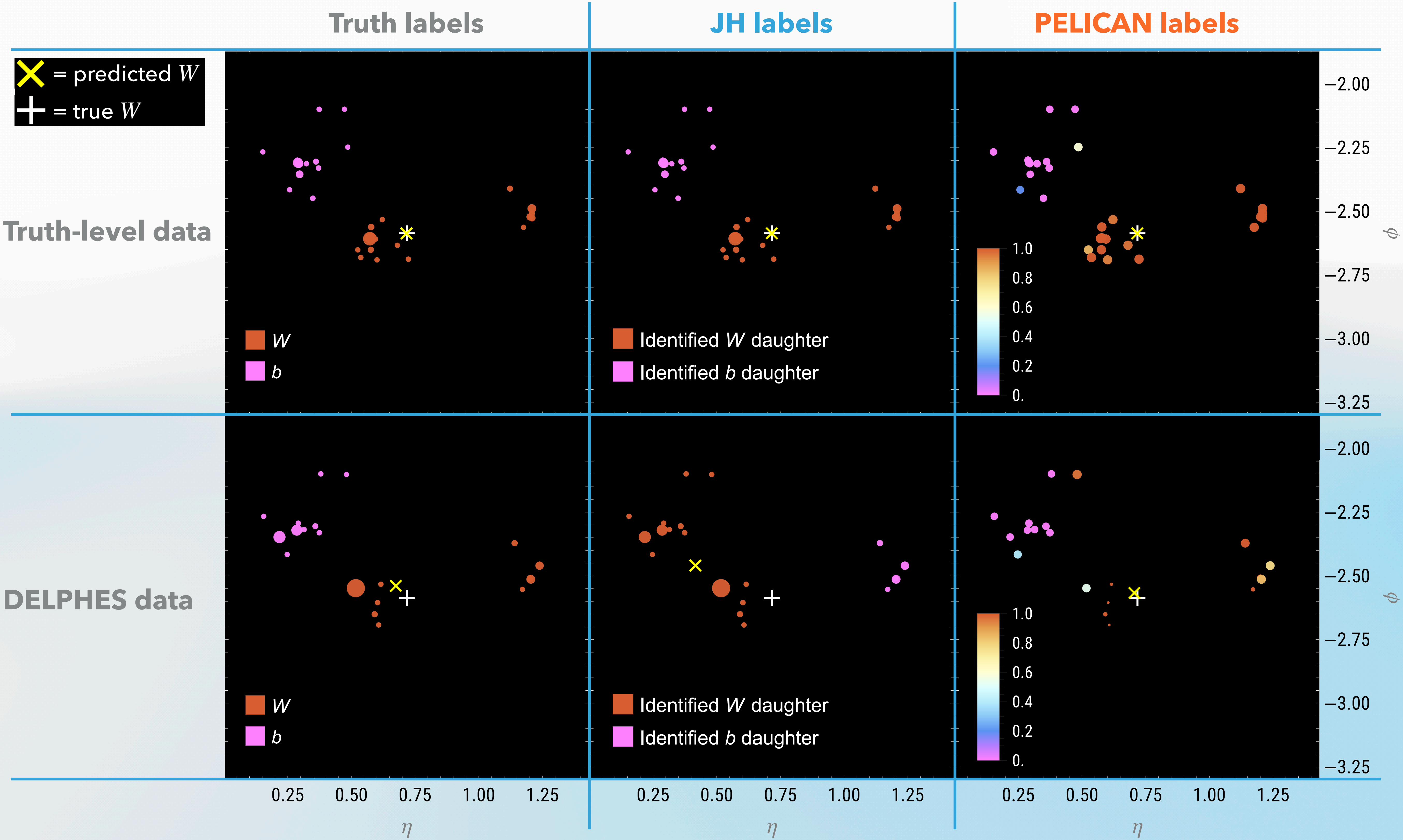
PELICAN: MOMENTUM REGRESSION

- Aggregated jet displays, centered on truth-level W.
- PELICAN weights each constituent in producing W 4-momentum – interpretable as likelihood of being W daughter!

DELPHES reconstruction

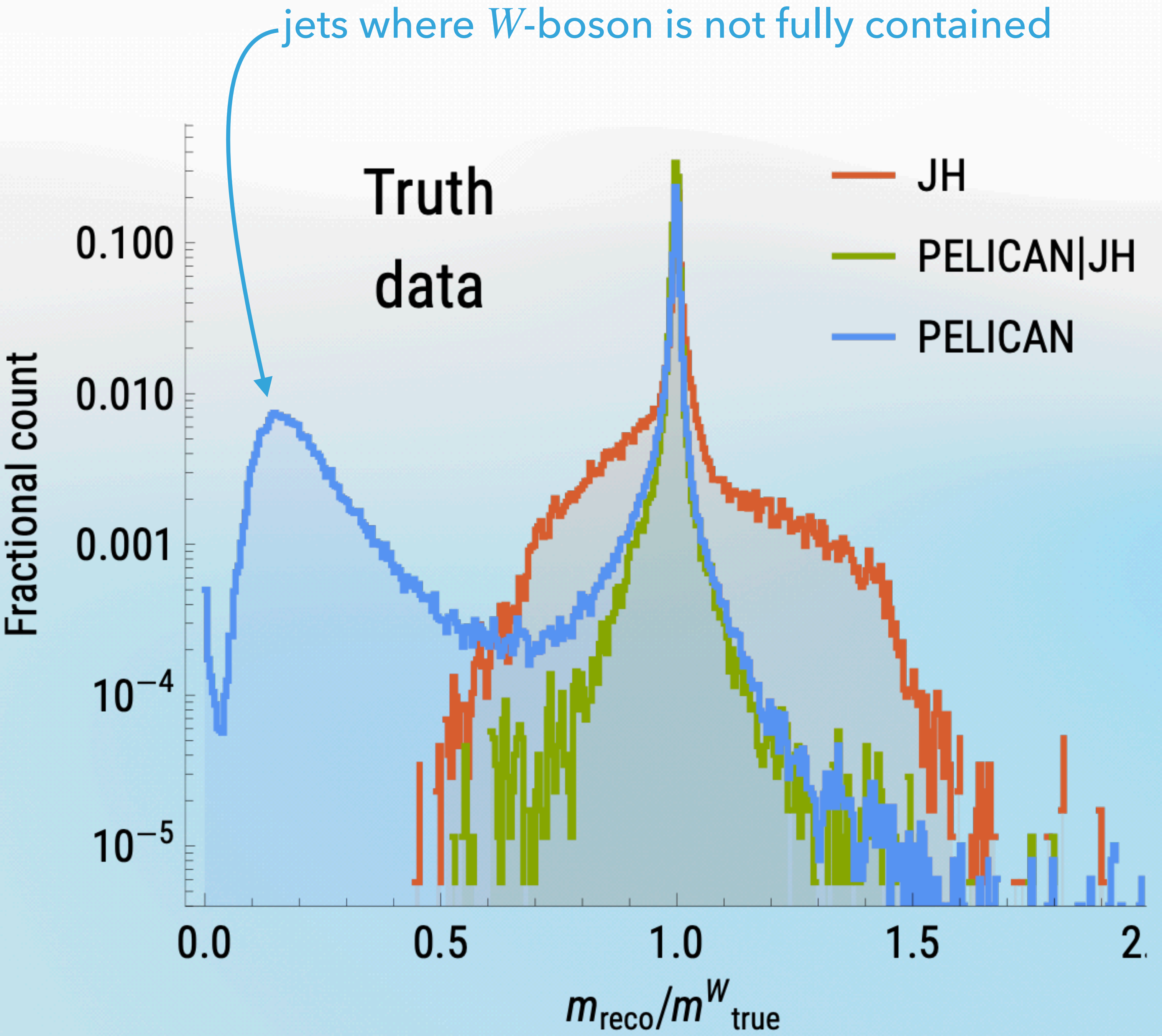


- PELICAN compared with the Johns Hopkins top tagger. [h]



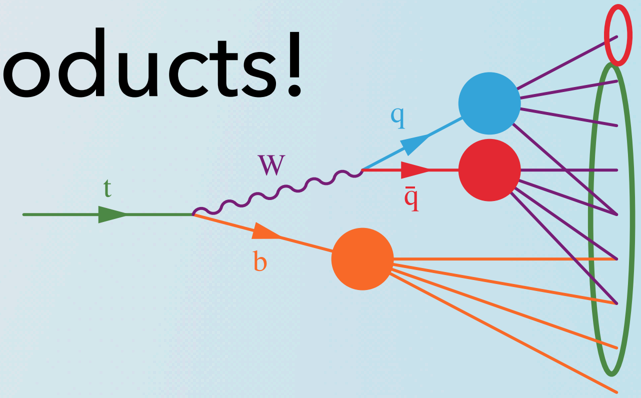
[h] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, Top Tagging (2008).

- PELICAN can reconstruct W momentum on *all* jets.
  - JH predicts W momentum *only* on JH-tagged jets.
  - JH-tagged jets account for only ~35% of the dataset.
  - Evaluating PELICAN on tagged jets gives improved mass resolution!
    - What about the untagged jets?

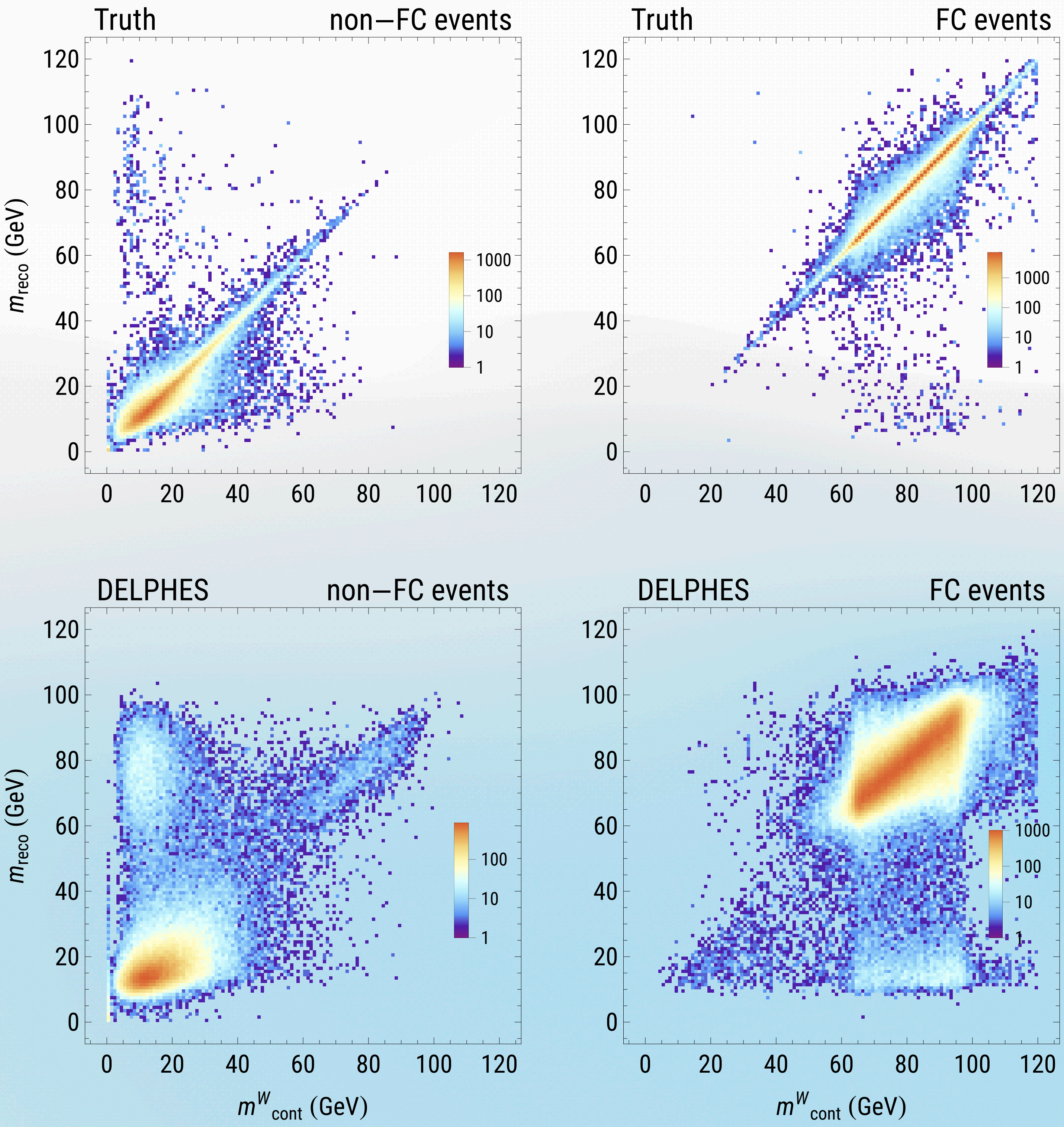

jets where W-boson is not fully contained

PELICAN: MOMENTUM REGRESSION

- Jets may not contain all $W \to qq'$ products.
- Train PELICAN targeting sum momentum of all *contained* W products.
- PELICAN reconstructs the momentum well, and can be used effectively to tag jets containing W decay products!



"FC" = jet contains $\mathrm{W} \to \mathrm{qq}'$



mass of contained W decay/shower products

IR safety

$$\lim_{\epsilon\to 0} f^{(N+1)}\left(p_1, \ldots, p_N, \epsilon p\right) = f^{(N)}\left(p_1, \ldots, p_N\right)$$

C safety

$$C_{12}(p)f = \partial_\lambda f\left(\lambda p, (1-\lambda)p, p_3, \ldots, p_N\right) = 0$$ (where $p^2 = 0$)

- IR- and IRC-safe implementations of PELICAN!
  - IR-safe regressor resolutions $\left(\sigma_{p_T}, \sigma_m, \sigma_{\Delta R}\right)$ worsen 20-35 % (truth-level), 40-50 % (DELPHES reconstruction)
  - IRC-safe regressor resolutions worsen 5-6 × (truth-level).

- PELICAN performs well in top-quark tagging, and W-boson momentum regression, and provides constituent-level information!
- Plenty of other problems where PELICAN can be used!
  - Top quark helicity measurement (in hadronic W channel)
  - Reconstructing resonant decays, complex decay chains
  - Charged particle tracking (add charge info!)
  - Full event reconstruction…

CONCLUSION & ACKNOWLEDGEMENTS

- PELICAN repo: https://github.com/abogatskiy/PELICAN
- New paper: https://arxiv.org/abs/2307.16506

Alexander Bogatskiy Timothy Hoffman David W. Miller Xiaoyang Liu Jan T. Offermann

ENRICO FERMI INSTITUTE

THE UNIVERSITY OF CHICAGO
DATA SCIENCE INSTITUTE

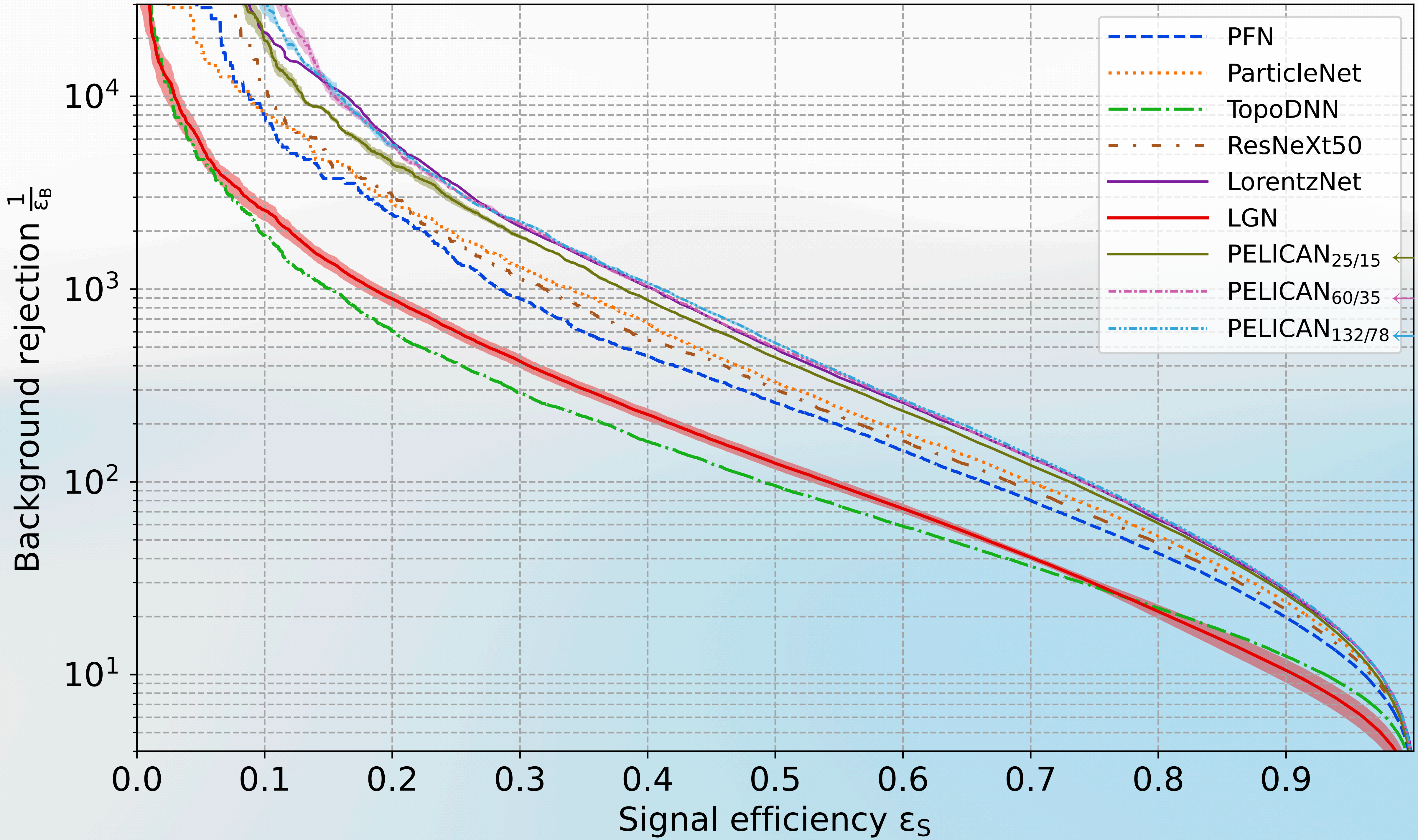
Miller Lab

FLATIRON INSTITUTE

QUESTIONS?

BACKUP

- Latest top-tagging results show that performance saturates quickly w.r.t. number of parameters. [i]

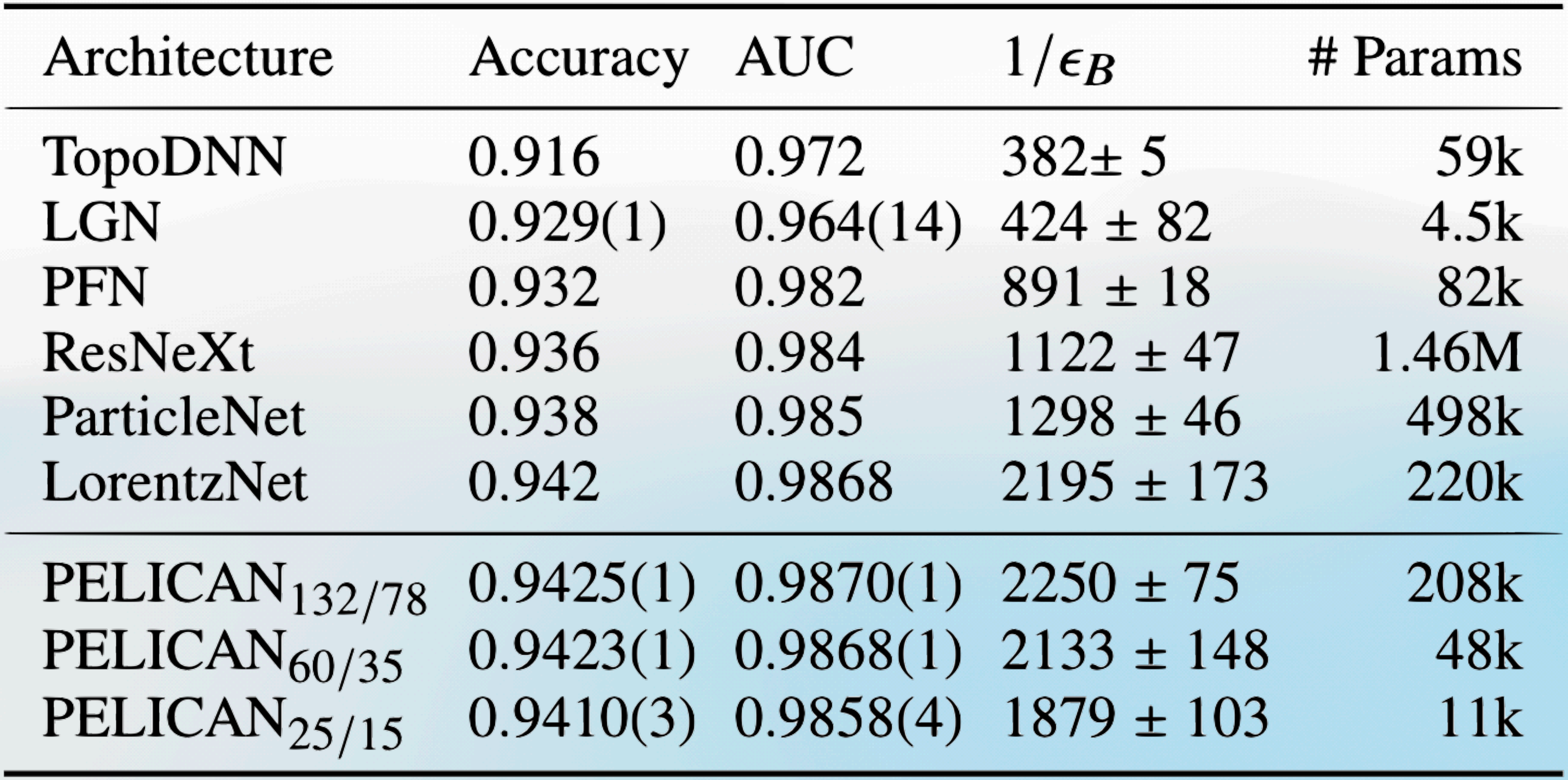


← 11k params
← 48k params
← 208k params

[i] A. Bogatskiy, T. Hoffman, D. W. Miller, J. T. Offermann, X. Liu, Explainable Equivariant Networks: PELICAN (2023).

PELICAN: TOP-TAGGING

- Latest top-tagging results show that performance saturates quickly w.r.t. number of parameters. [i]



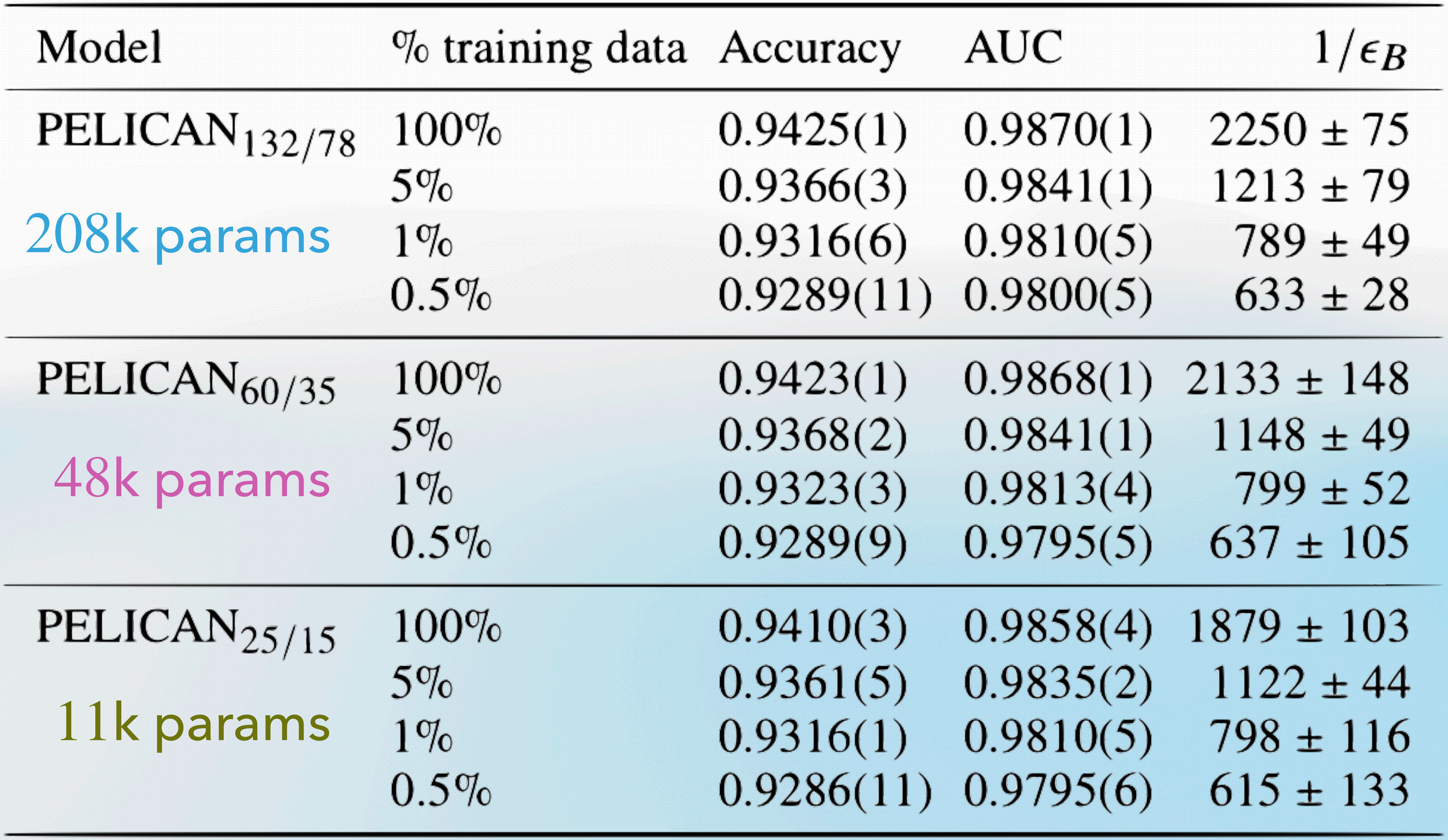
Architecture	Accuracy	AUC	$1/\epsilon_B$	# Params
TopoDNN	0.916	0.972	382± 5	59k
LGN	0.929(1)	0.964(14)	424 ± 82	4.5k
PFN	0.932	0.982	891 ± 18	82k
ResNeXt	0.936	0.984	1122 ± 47	1.46M
ParticleNet	0.938	0.985	1298 ± 46	498k
LorentzNet	0.942	0.9868	2195 ± 173	220k
$\text{PELICAN}_{132/78}$	0.9425(1)	0.9870(1)	2250 ± 75	208k
$\text{PELICAN}_{60/35}$	0.9423(1)	0.9868(1)	2133 ± 148	48k
$\text{PELICAN}_{25/15}$	0.9410(3)	0.9858(4)	1879 ± 103	11k

$\text{PELICAN}_{X/Y}$ = MSG has X input channels, Y output channels

[i] A. Bogatskiy, T. Hoffman, D. W. Miller, J. T. Offermann, X. Liu, Explainable Equivariant Networks: PELICAN (2023).

PELICAN: TOP-TAGGING

- PELICAN is very efficient with training data. [i]

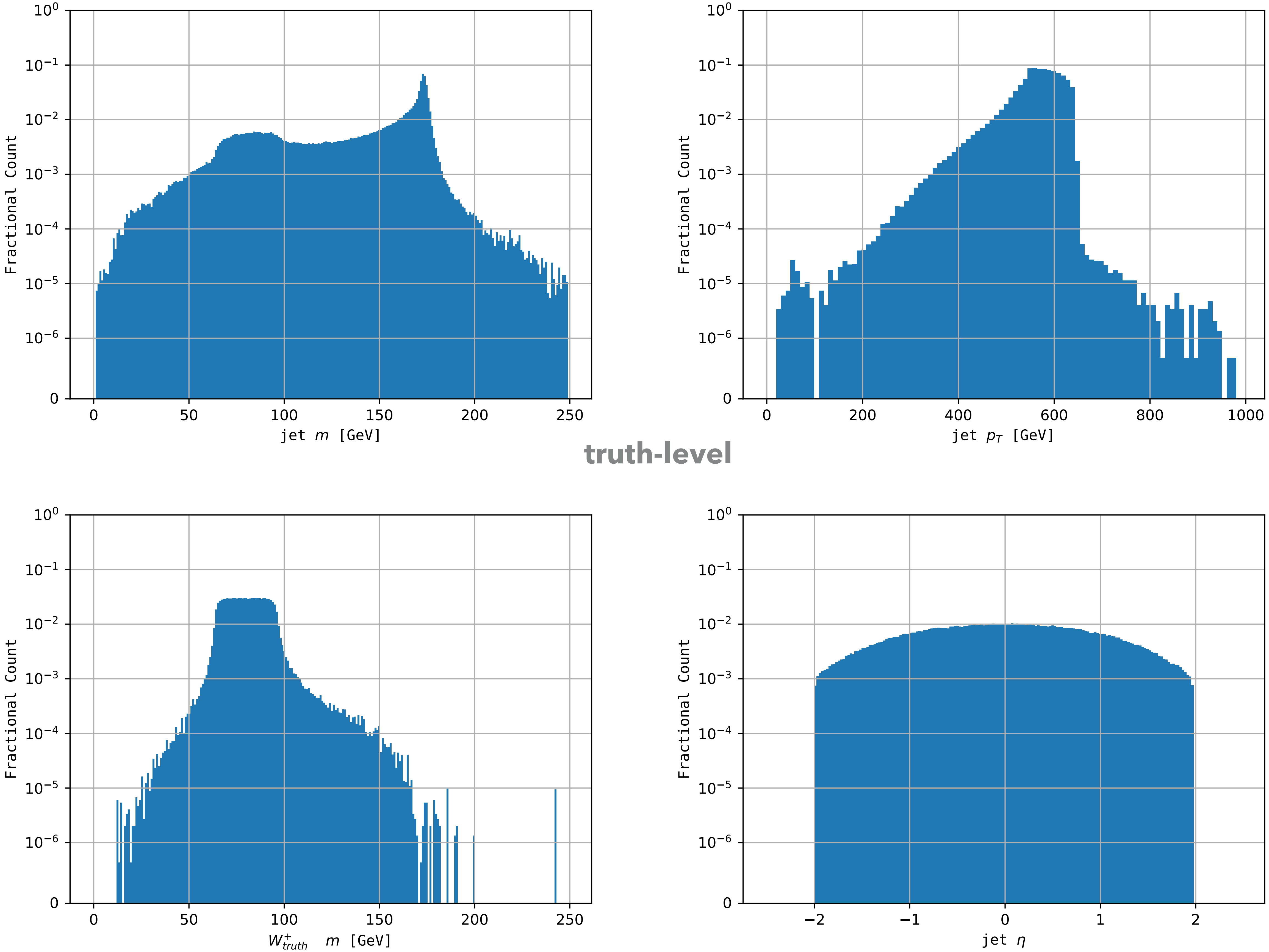


Model	% training data	Accuracy	AUC	$1/\epsilon_B$
$\text{PELICAN}_{132/78}$	100%	0.9425(1)	0.9870(1)	2250 ± 75
	5%	0.9366(3)	0.9841(1)	1213 ± 79
208k params	1%	0.9316(6)	0.9810(5)	789 ± 49
	0.5%	0.9289(11)	0.9800(5)	633 ± 28
$\text{PELICAN}_{60/35}$	100%	0.9423(1)	0.9868(1)	2133 ± 148
	5%	0.9368(2)	0.9841(1)	1148 ± 49
48k params	1%	0.9323(3)	0.9813(4)	799 ± 52
	0.5%	0.9289(9)	0.9795(5)	637 ± 105
$\text{PELICAN}_{25/15}$	100%	0.9410(3)	0.9858(4)	1879 ± 103
	5%	0.9361(5)	0.9835(2)	1122 ± 44
11k params	1%	0.9316(1)	0.9810(5)	798 ± 116
	0.5%	0.9286(11)	0.9795(6)	615 ± 133

[i] A. Bogatskiy, T. Hoffman, D. W. Miller, J. T. Offermann, X. Liu, Explainable Equivariant Networks: PELICAN (2023).

W-BOSON MOMENTUM REGRESSION: DATASET

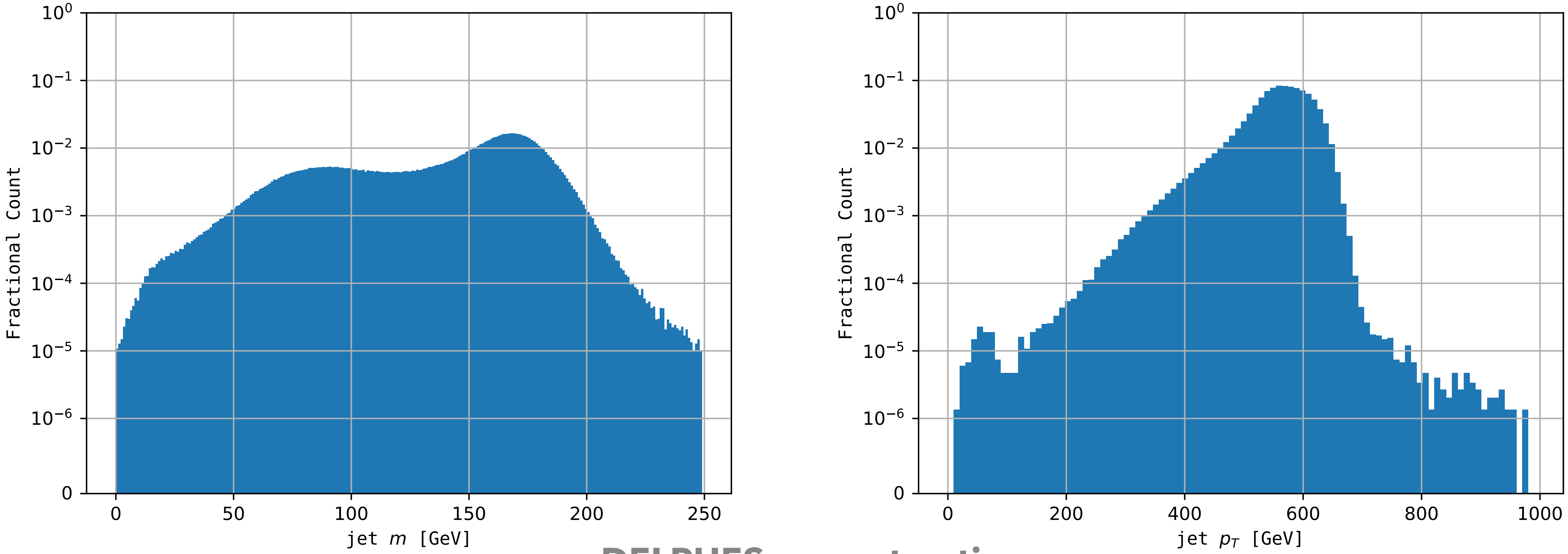
HEPData4ML tool: https://github.com/janTOffermann/HEPData4ML



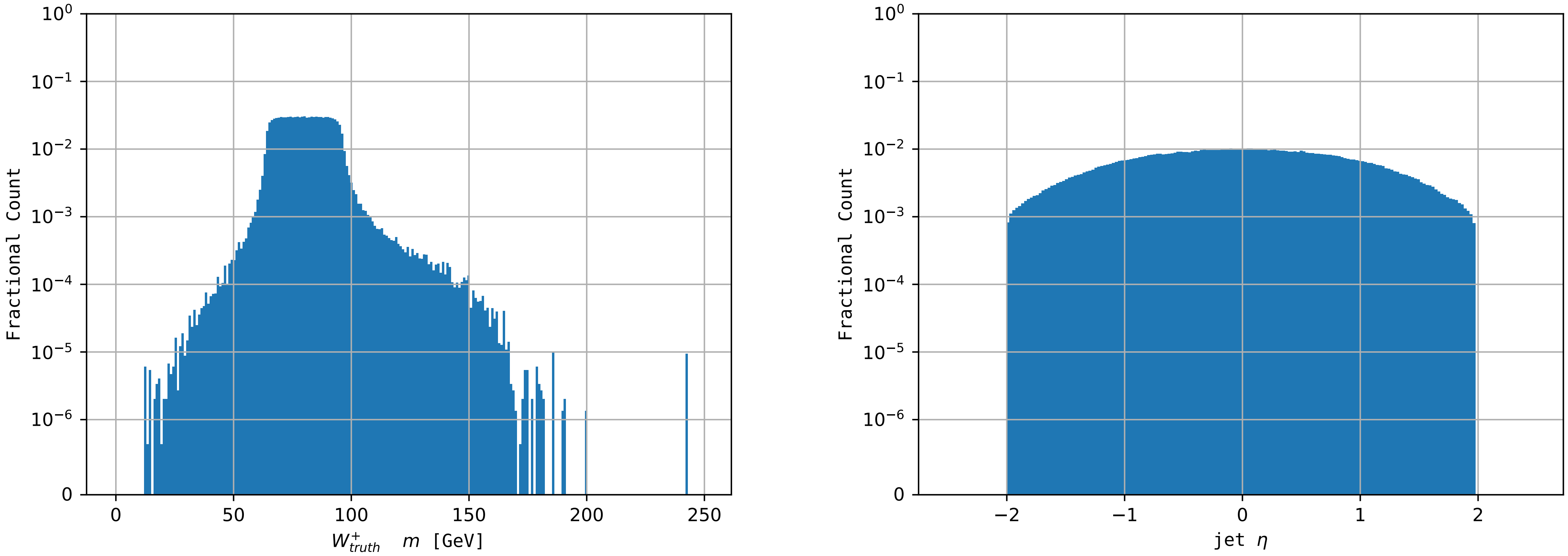
truth-level

W-BOSON MOMENTUM REGRESSION: DATASET

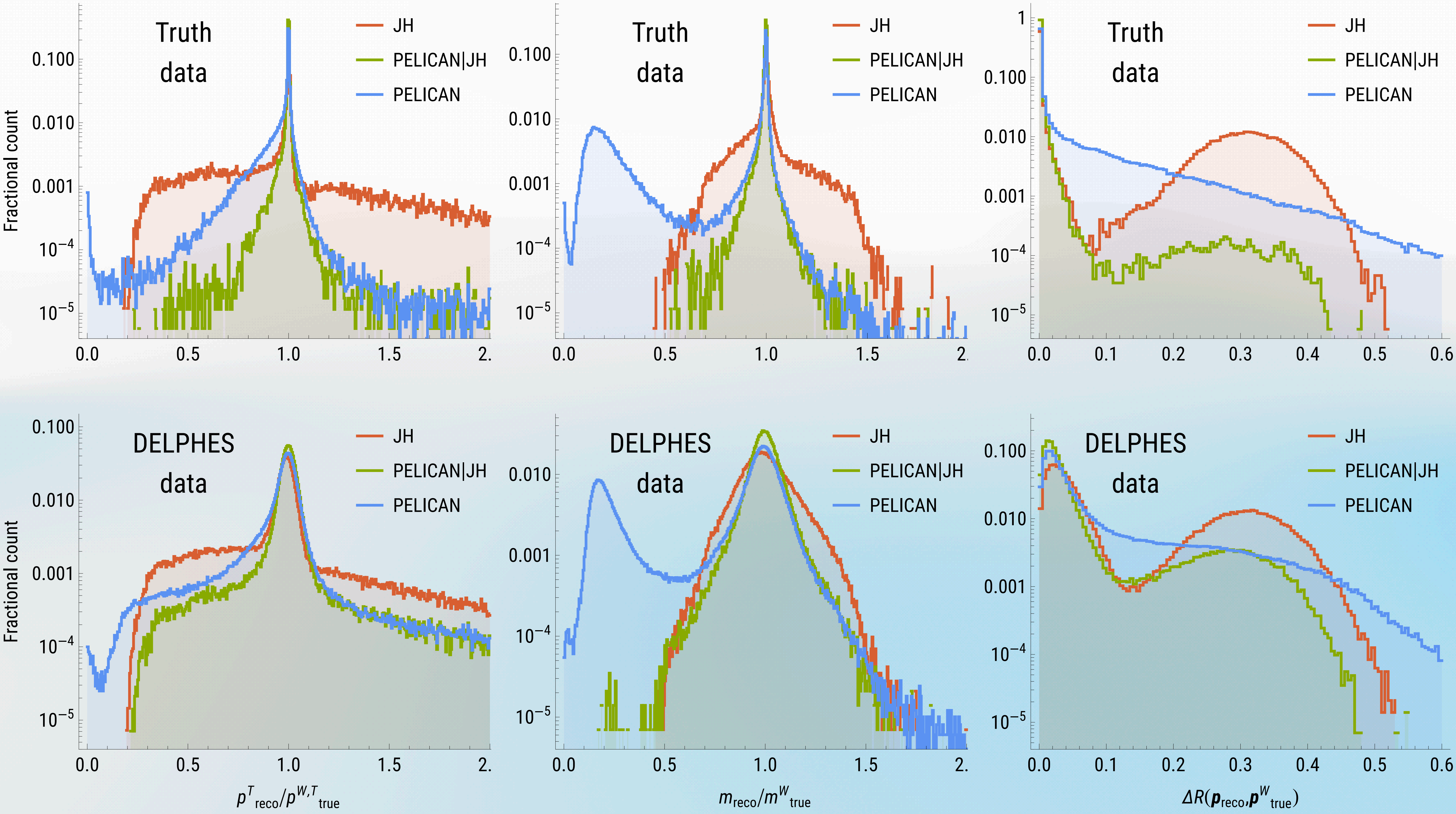
HEPData4ML tool: https://github.com/janTOffermann/HEPData4ML



DELPHES reconstruction

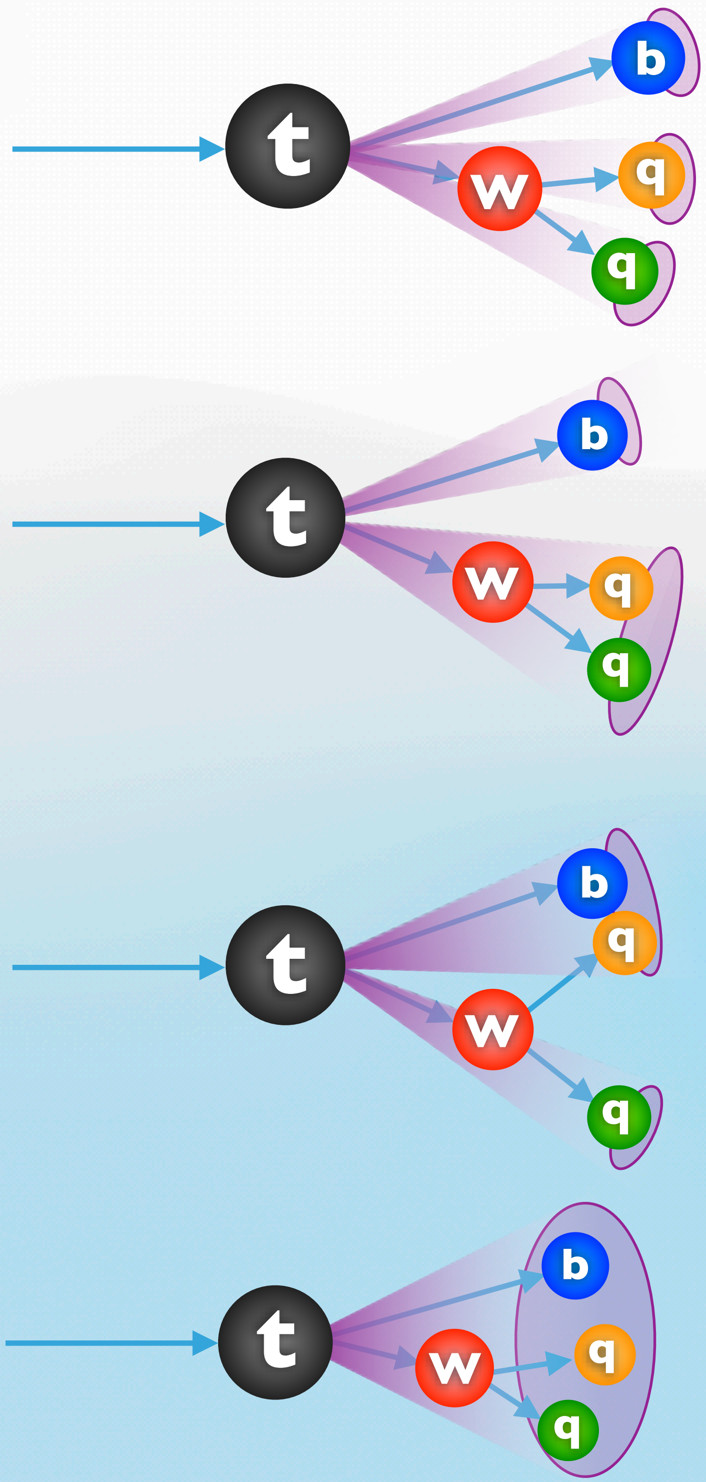


W-BOSON MOMENTUM REGRESSION: RESOLUTIONS



W-BOSON MOMENTUM REGRESSION: CONTAINMENT

- Top quark jets might not contain all the decay products.
- Cannot reconstruct W-boson from jets that do not contain $W \to qq'$ (not enough info!).
- Can train regressor to construct sum of W decay products within the jet.



Unmerged
(q jet + q jet + b jet)

Partially merged
(qq jet + b jet)

Partially merged
(bq jet + q jet)

Fully merged
(bqq jet)

Image Source: Jet Substructure HATS@LPC (June 18-19, 2014) (modified)

W-BOSON MOMENTUM REGRESSION: JOHNS HOPKINS TOP TAGGER

- Re-cluster jet with Cambridge-Aachen algorithm.
- Iteratively de-cluster the CA jet, producing 2 subjets.
- Throw out softer jet if $p_{T,2} / p_T < \delta_p$.
- Continue until:
  - Both objects harder than δ_p. ← both objects are subjets, repeat process on them
  - Both objects softer than δ_p.
  - Two objects are too close, $|\Delta\eta| + |\Delta\phi| < \delta_r$.
  - Only one calorimeter cell is left.

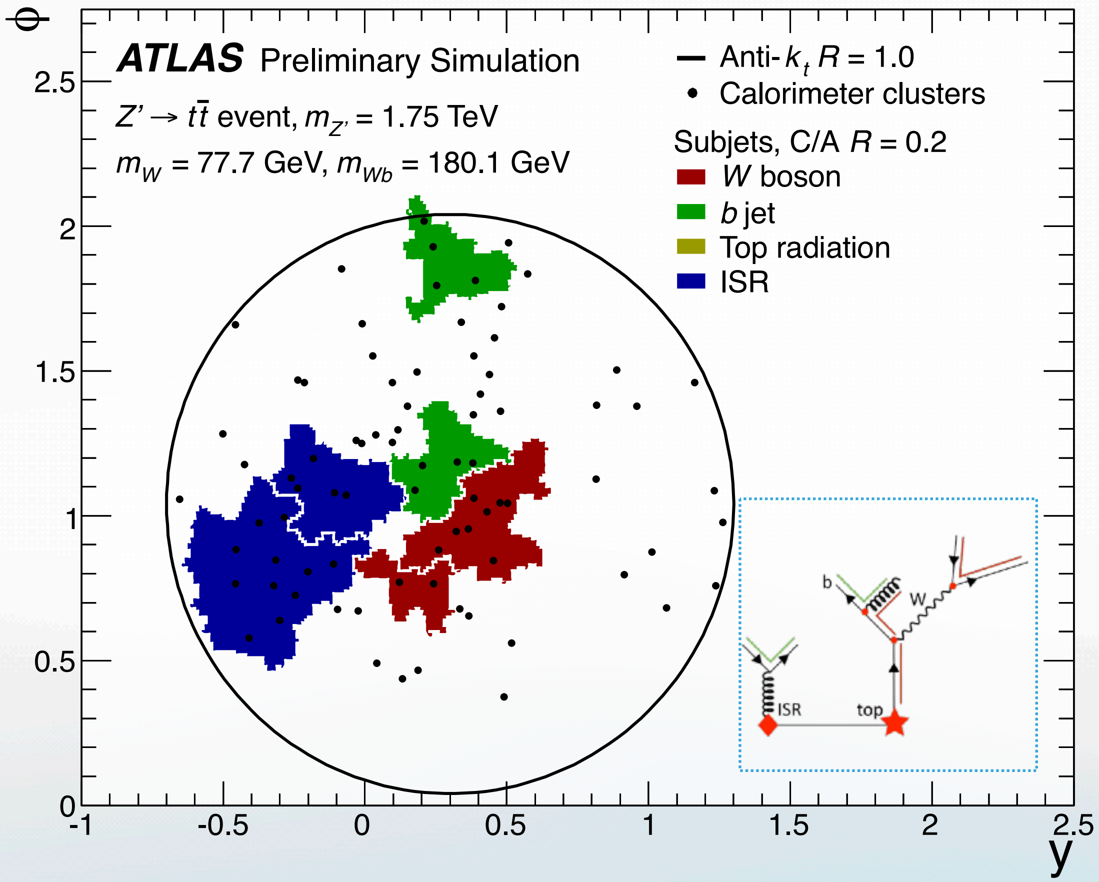
objects are irreducible

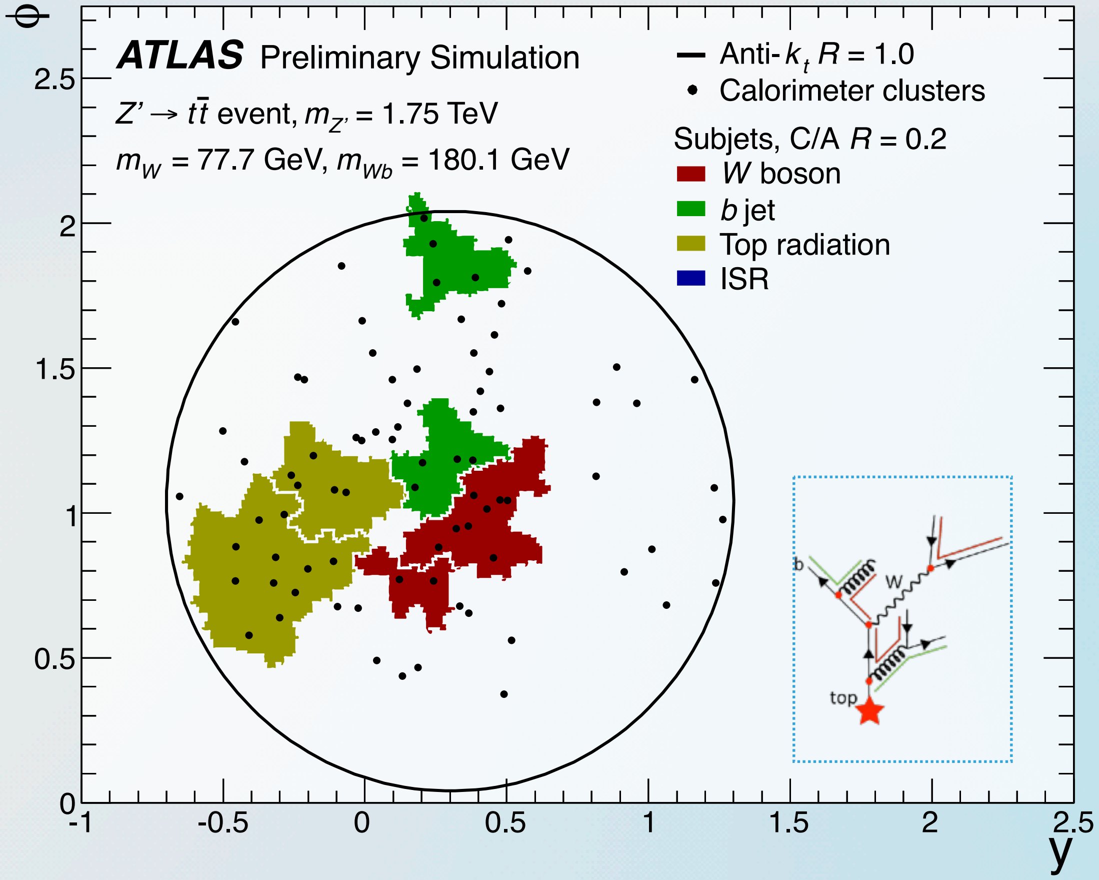
W-BOSON MOMENTUM REGRESSION: JOHNS HOPKINS TOP TAGGER

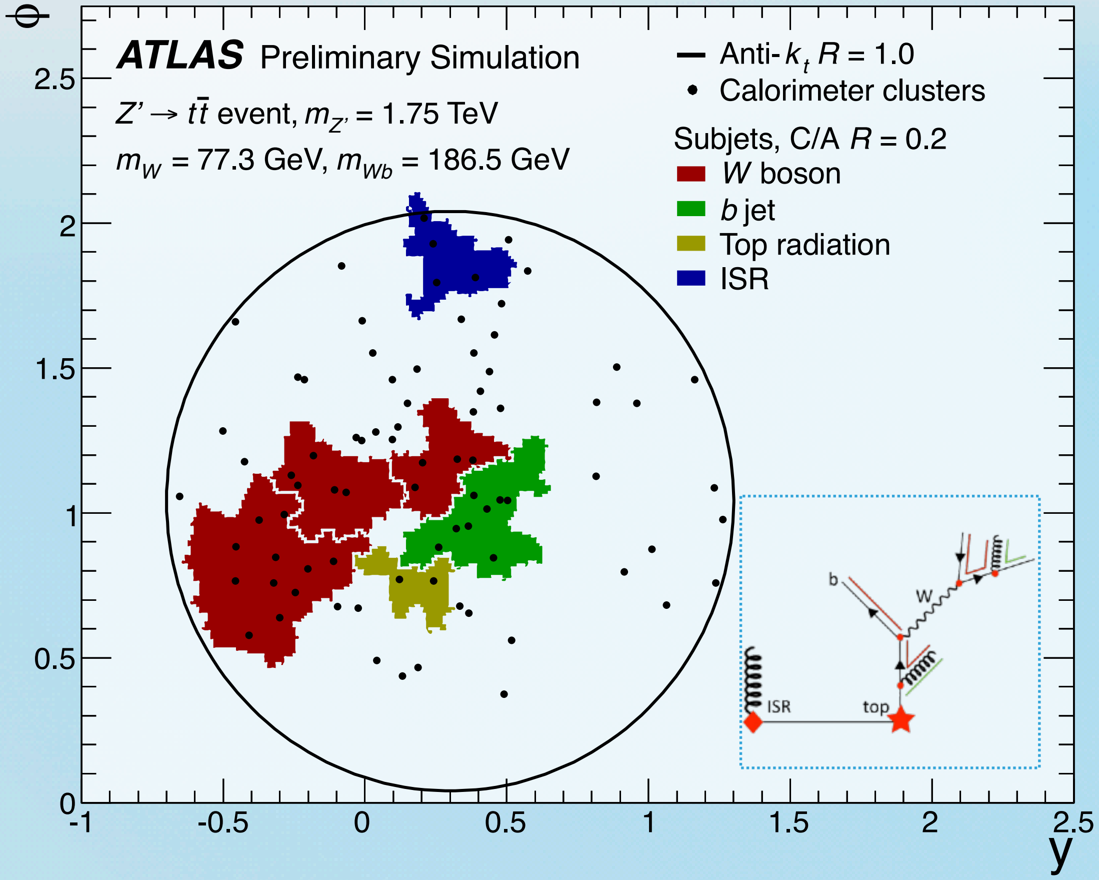
- Reject jets which only yield 2 subjets.
- For jets with 3+ subjets, W candidate will be constructed from the 2 subjets whose mass is closest to m_W.
- Implemented with (default) values and cuts:
  - $\delta_p = 0.1, \delta_r = 0.19$
  - Cuts:
    - $m_t^{\text{reco}} \in \{150{,}200\}$ GeV, $m_W^{\text{reco}} \in \{65{,}95\}$ GeV
    - $\cos\left(\theta_W\right) < 0.7$

W-BOSON MOMENTUM REGRESSION: SHOWER DECONSTRUCTION

- Shower deconstruction is a method for determining the likeliest showering history of a given jet.
- Identification done at the *subjet* level.







PELICAN: ACHIEVING IRC SAFETY

- Achieving IR safety = remove dependence on multiplicity.
  - Remove scaling functions based on # of particles.
  - Set bias to zero in all linear layers.
  - Input embedding maps $0 \to 0$, activation has a fixed point at zero. *(PELICAN already satisfies these)*
  - *Soft masking*: *Last 12 aggregators* followed by multiplication by $J \cdot p_i$, with $J = \sum_{i=1}^{N} p_i$ scaled and clipped to be within $\left[-1,1\right]$.

- Achieving C-safety = output depends on massless inputs only through their sum.
  - Soft mask applied after input layer, multiplication by m_i^2 instead of by $J \cdot p_i$.
  - This *also* gives IR safety → applied C-safety modification gives a fully IRC-safe PELICAN.
- More significant negative effect on performance than just using the IR-safe modifications.