PELICAN: deep jets with permutation and Lorentz equivariance

Alexander Bogatskiy
Research Fellow, Flatiron Institute

Joint work with: David Miller, Timothy Hoffman, Jan Offermann, Xiaoyang Liu (UChicago)

FLATIRON INSTITUTE

THE UNIVERSITY OF CHICAGO

ML4Jets 2022

Regular Article - Experimental Physics | Open Access | Published: 13 July 2016

Jet-images — deep learning edition

Luke de Oliveira, Michael Kagan, Lester Mackey, Benjamin Nachman & Ariel Schwartzman

A preprint version of th

ABSTRACT

Building on the notion of a
high energy particles, or jet
learning techniques based
Modern deep learning algo
motivated feature driven a
these features are learned
performance. This interpla
learning algorithms is gene
discover new particles and
jets.

Regular Article - Theoretical Physics | Open Access | Published: 25 January 2017

Deep learning in color: towards automated quark/gluon jet discrimination

Patrick T. Komiske, Eric M. Meto

A preprint version of the a

ABSTRACT

Artificial intelligence offers th
collider physics. To establish i
convolutional neural network
designed by physicists. Our ap
image, with intensity given by
by adding color to the images,
momentum in charged partic
charged particle counts. Over
variables. We also find that, w
the neural networks are surp
observables. This suggests tha
imperfect simulations.

Regular Article - Theoretical Physics | Open Access | Published: 18 October 2018

Pulling out all the tops with computer vision and deep learning

Sebastian Macaluso

A preprint versio

ABSTRACT

We apply computer v
(CNN) — to build a hi
of Kasieczka et al) ha
performance to state-
introduce a number o
image preprocessing,
BDTs based on high-l
wide range of tagging
QCD background reje
(non-merged) top jets
straightforwardly ext
here may be useful to

Regular Article - Experimental Physics | Open Access | Published: 02 May 2017

Deep-learning top taggers or the end of QCD?

Gregor Kasieczka, Tilman Plehn, Michael Russell & Torben Schell

A preprint version o

ABSTRACT

Machine learning based
from the LHC. Top taggi
DeepTop approach and
a network architecture t
Model production chan
multivariate QCD-based
performance, establishi
multivariate hypothesis-

Neural Message Passing for Jet Physics

Isaac Henrion, Johann Brehmer, Joan Bruna, Kyunghun Cho, Kyle Cranmer
Center for Data Science
New York University
New York, NY 10012
{henrion*, johann.brehmer, bruna, kyunghyun, kyle.cranmer*}@nyu.edu

Gilles Louppe
Department of Computer Science
University of Liège
Belgium
g.louppe@ulg.ac.be

Gaspar Rochette
Department of Computer Science
École Normale Supérieure
Paris, France
gaspar.rochette@ens.fr

Regular Article - Theoretical Physics | Open Access | Published: 07 January 2019

QCD-aware recursive neural networks for jet physics

Gilles Louppe, Kyunghyun Cho, Cy

Journal of High Energy Physics 20

857 Accesses | 78 Citations | 3

A preprint version of the art

ABSTRACT

Recent progress in applying ma
between calorimeters and imag
networks built instead upon an
four-momenta are like words a
algorithms is like the parsing of a sentence.
momenta of a variable-length set of particle
event-by-event basis. Our experiments high
specific jet embeddings and show that recu
and data efficient than previous image-base
individual jets (sentences) to full events (pa
level classifier operating on all the stable pa

[Submitted on 24 Nov 2017]

Long Short-Term Memory (LSTM) networks with jet constituents for boosted top tagging at the LHC

Shannon Eg

Multivariat
identificati
Recent De
activation
network. T
this work t
networks i
constituen
achieves a
than a fact
trained on

Regular Article | Published: 17 July 2019

Pileup mitigation at the Large Hadron Collider with graph neural networks

J. Arjona Martínez, O. Cerri, M. Spi

The European Physical Journal Plu

116 Accesses | 30 Citations | 2

Abstract.

At the Large Hadron Collider, t
collaborations occur in coincide
usually referred to as pileup. Pi
event reconstruction as pileup
observables. We present a class
particles coming from high-tra
from pileup collisions. This mo
Bertolini *et al.*, JHEP **10**, 059 (
Thanks to an extended basis of
considered network architectur
with respect to state-of-the-art

Special Article - Tools for Experiment and Theory | Open Access | Published: 18 July 2019

Learning representations of irregular particle-detector geometry with distance-weighted graph networks

Shah Rukh Qasim, Jan Kieseler, Yutaro Iiyama & Maurizio Pierini

The European Physical Journal C 79, Articl

2903 Accesses | 49 Citations | 38 Altm

A preprint version of the article is av

Abstract

We explore the use of graph networks t
of particle reconstruction. Thanks to th
networks can exploit the full detector g
and arbitrarily complex detector geome
network architectures, dubbed GARNET
reconstruction task. The performance o
simulated particle interactions on a toy
inspired by the endcap calorimeter to b
Luminosity LHC phase. We study the c
calorimetric particle reconstruction, an
approaches. The proposed algorithms p

Unveiling CP property of top-Higgs coupling with graph neural networks at the LHC

Jie Ren [a, b, c], Lei Wu [a], Jin Min Yang [b, c, d]

Show more

Share

https://doi.org

Under a Creat

Abstract

The top-H
The precis
beyond the
Neural Ne
semi-lepto
constructe
that the pu
LHC, with

Jet tagging via particle clouds

Huilin Qu and Loukas Gouskos

Phys. Rev. D **101**, 056019 – Published 26 March 2020

Article | References | Citing Articles (73) | PDF | HTML | Export Citation

ABSTRACT

How to represent a jet is at the core of machine learning on jet physics. Inspired by the notion of point clouds, we propose a new approach that considers a jet as an unordered set of its constituent particles, effectively a "particle cloud." Such a particle cloud representation of jets is efficient in incorporating raw information of jets and also explicitly respects the permutation symmetry. Based on the particle cloud representation, we propose ParticleNet, a customized neural network architecture using Dynamic Graph Convolutional Neural Network for jet tagging problems. The ParticleNet architecture achieves state-of-the-art performance on two representative jet tagging benchmarks and is improved significantly over existing methods.

[Submitted on 27 Jul 2017 (v1), last revised 23 Apr 2018 (this version, v3)]

Deep-learned Top Tagging with a Lorentz Layer

Anja Butter, Gregor Kasieczka, Tilman Plehn, Michael Russell

Regular Article - Theoretical Physics | Open Access | Published: 15 January 2019

Energy flow networks: deep sets for particle jets

Patrick T. Komiske, Eric M. Metodiev & Jesse Thaler

Journal of High Energy

1334 Accesses | 101

A preprint versi

ABSTRACT

A key question for m
and learn from collid
particles, we build u
or "point clouds". Ad
introduce Energy Fl
We also develop Part
inclusion of addition
feature a per-particle
overall event-level la
existing event repres
demonstrate the pow
the collider task of di
performance compared to existing methods. We also show how the learned event representation can be directly visualized, providing insight into the inner workings of the model. These architectures lend themselves to efficiently processing and analyzing events for a wide variety of tasks at the Large Hadron Collider. Implementations and examples of our architectures are available online in our EnergyFlow package.

Open Access

Acc

Equivariant energy flow networks for jet tagging

Matthew J. Dolan and Ayodele Ore
Phys. Rev. D 103, 074022 – Published 27 April 2021

[Submitted on 8 Jun 2020]

Lorentz Group Equivariant Neural Network for Particle Physics

Alexander Bogatskiy, Brandon Anderson, Jan T. Offermann, Marwah Roussi, David W. Miller, Risi Kondor

We present a neural network architecture that is fully equivariant with respect to transformations under the Lorentz group, a fundamental symmetry of space and time in physics. The architecture is based on the theory of the finite-dimensional representations of the Lorentz group and the equivariant nonlinearity involves the tensor product. For classification tasks in particle physics, we demonstrate that such an equivariant architecture leads to drastically simpler models that have relatively few learnable parameters and are much more physically interpretable than leading approaches that use CNNs and point cloud approaches. The competitive performance of the network is demonstrated on a public classification dataset [27] for tagging top quark decays given energy-momenta of jet constituents produced in proton-proton collisions.

[Submitted on 27 Jul 2017 (v1), last revised 23 Apr 2018 (this version, v3)]

Deep-learned Top Tagging with a Lorentz Layer

Anja Butter, Gregor Kasieczka, Tilman Plehn, Michael Russell

Regular Article - Theoretical Physics | Open Access | Published: 15 January 2019

Energy flow networks: deep sets for particle jets

Patrick T. Komiske, Eric M. Metodiev & Jesse Thaler

Journal of High Energy

1334 Accesses | 101

A preprint versi

ABSTRACT

A key question for m
and learn from collid
particles, we build u
or "point clouds". Ad
introduce Energy Flo
We also develop Par
inclusion of addition
feature a per-particle
overall event-level la
existing event repres
demonstrate the pow
the collider task of di
performance compared to existing methods. We also show how the learned eve
representation can be directly visualized, providing insight into the inner worki
model. These architectures lend themselves to efficiently processing and analyzi
wide variety of tasks at the Large Hadron Collider. Implementations and examp
architectures are available online in our EnergyFlow package.

Open Access

Acc

Equivariant energy flow networks for jet tagging

Matthew J. Dolan and Ayodele Ore
Phys. Rev. D **103**, 074022 – Published 27 April 2021

[Submitted on 8 Jun 2020]

Lorentz Group Equivariant Neural Network for Particle Physics

Alexander Bogatskiy, Brandon Anderson, Jan T. Offermann, Marwah Roussi, David W. Miller, Risi Kondor

We present a neural network architecture that
the Lorentz group, a fundamental symmetry o
the theory of the finite-dimensional represent
involves the tensor product. For classification
equivariant architecture leads to drastically si
and are much more physically interpretable th
approaches. The competitive performance of t
dataset [27] for tagging top quark decays give
proton collisions.

[Submitted on 5 Jul 2021]

Particle Convolution for High Energy Physics

Chase Shimmin

We introduce the Partic
equivariant neural netw
particle convolution lay
Energy Flow network a
operator is promoted t
implemented for variou
of rotation about the je
tasks, q/g tagging and
achieves performance
Moreover, we show tha
significantly outperform
We speculate that by g
convolutional symmetr
current state-of-the-a

[Submitted on 20 Jan 2022 (v1), last revised 29 May 2022 (this version, v5)]

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong, Qi Meng, Jue Zhang, Huilin Qu, Congqiao Li, Sitian Qian, Weitao Du, Zhi-Ming Ma, Tie-Yan Liu

Deep learning m
particle physics
be an important
in many applica
spacetime symm
incorporated in
design is compu
high-order tens

Particle Transformer for Jet Tagging

Huilin Qu, Congqiao Li, Sitian Qian *Proceedings of the 39th International Conference on Machine Learning*, PMLR 162:18281-18292, 2022.

Abstract

Jet tagging is a critical yet challenging classification task in particle physics. While deep learning has

Symmetries in jet data

Symmetries in jet data

- Euclidean symmetries in jet images (CNN’s)

Symmetries in jet data

- Euclidean symmetries in jet images (CNN’s)
- Permutations of particles (particle clouds, Deep Sets, GNN, MPNN)

Symmetries in jet data

- Euclidean symmetries in jet images (CNN's)
- Permutations of particles (particle clouds, Deep Sets, GNN, MPNN)
- Rotational, boost, Lorentz symmetries

LGN

LGN

- Linear: a restricted linear mixing operations acting on data stored as a collection of irreducible representations $\alpha A + \beta B$

LGN

- **Linear**: a restricted linear mixing operations acting on data stored as a collection of irreducible representations $\alpha A + \beta B$
- **Nonlinear**: a tensor product combined with a Clebsch-Gordan decomposition

$$\otimes : R_1 \times R_2 \rightarrow R_3$$

$$\rho_1(g) \otimes \rho_2(g) = C_{\rho_1,\rho_2} \left[\bigoplus_{\rho} \bigoplus_{1}^{\mu(\rho)} \rho(g) \right] C^{\dagger}_{\rho_1,\rho_2}$$

LGN

- Linear: a restricted linear mixing operations acting on data stored as a collection of irreducible representations $\alpha A + \beta B$
- Nonlinear: a tensor product combined with a Clebsch-Gordan decomposition
- A covariant universal approximation theorem motivates the sufficiency of these operations for outputs belonging to arbitrary finite-dimensional representations.

$$\otimes : R_1 \times R_2 \to R_3$$

$$\rho_1(g) \otimes \rho_2(g) = C_{\rho_1,\rho_2} \left[\bigoplus_{\rho} \bigoplus_{1}^{\mu(\rho)} \rho(g) \right] C^{\dagger}_{\rho_1,\rho_2}$$

What constraints does full Lorentz+Permutation symmetry impose?

Invariants of the Lorentz group

VECTOR INVARIANTS 45

6. Second example: unimodular group $SL(n)$

We shall take up the question of *invariants of the unimodular group* $SL(n)$ at once for any number of covariant or Latin vectors $x, y, \cdots$ and any number of contravariant or Greek vectors $\xi, \eta, \cdots$.

THEOREM (2.6.A).

$$[xy \cdots z], \qquad (\xi x), \qquad [\xi\eta \cdots \zeta]$$

is a complete table of typical basic invariants for the unimodular group.

[H. Weyl, The Classical Groups, 1939]

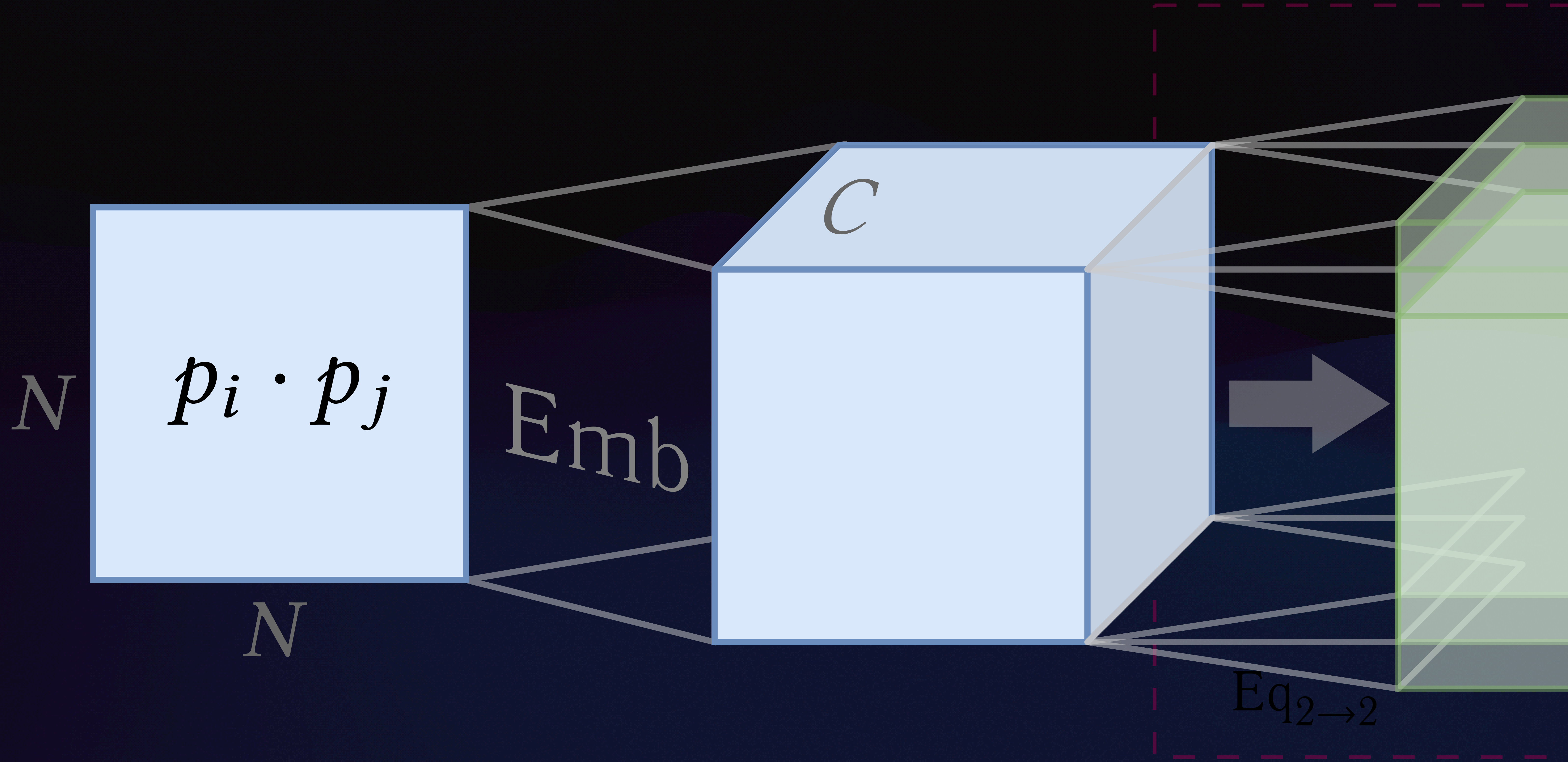
Invariants of the Lorentz group

All Lorentz-invariant symmetric functions of a set of 4-vectors p_i^μ can be expressed as functions of only the dot products $d_{ij} = \eta_{\mu\nu} p_i^\mu p_j^\nu$

$$I(p_1, \ldots, p_N) = I\left(\{p_i \cdot p_j\}_{i,j} \right)$$

$$p_i \cdot p_j = p_i^0 p_j^0 - \overrightarrow{p_i} \cdot \overrightarrow{p_j}$$

Input to the network: matrix of dot products $d_{ij} = p_i \cdot p_j$



Permutation equvariant architecture with only edge data?

PELICAN

PELICAN

Permutation Equivariant, Lorentz Invariant/Covariant Aggregator Net

Equivariant Aggregators

Equivariant Aggregators

- Basic approach — Deep Sets $\rho\left(\sum_i \phi\left(x_i\right)\right)$

Equivariant Aggregators

- Basic approach — Deep Sets $\rho\left(\sum_i \phi\left(x_i\right)\right)$
- Common approach — GNN/MPNN: $v_i' = f(v_i, \sum_j m_{ij}(v_i, v_j))$

Equivariant Aggregators

- Basic approach — Deep Sets $\rho\left(\sum_i \phi\left(x_i\right)\right)$
- Common approach — GNN/MPNN: $v_i' = f(v_i, \sum_j m_{ij}(v_i, v_j))$
- For pure "edge data", look for the most general permutation-equivariant mapping

$T_{ij} \mapsto T_{ij}'$

$$F\left(\pi \circ T_{i_1 i_2 \cdots i_r}\right) = \pi \circ F\left(T_{i_1 i_2 \cdots i_s}\right), \quad \pi \in S_N .$$

Equivariant Aggregators

- Basic approach — Deep Sets $\rho\left(\sum_i \phi\left(x_i\right)\right)$
- Common approach — GNN/MPNN: $v_i' = f(v_i, \sum_j m_{ij}(v_i, v_j))$
- For pure "edge data", look for the most general permutation-equivariant mapping

$T_{ij} \mapsto T'_{ij}$

$$F\left(\pi \circ T_{i_1 i_2 \cdots i_r}\right) = \pi \circ F\left(T_{i_1 i_2 \cdots i_s}\right), \quad \pi \in S_N .$$

- The space of such linear mappings between $N \times N$ matrices is 15-dimensional

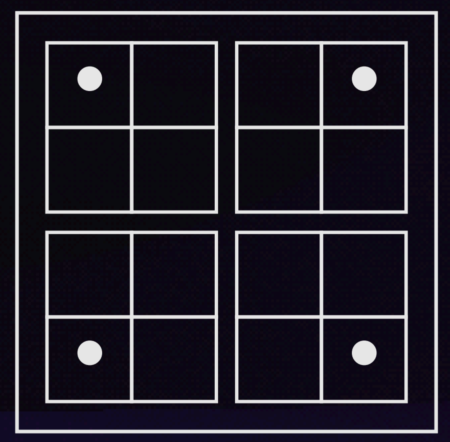
Equivariant Aggregators

Let's illustrate equivariant mappings $T_{ij} \mapsto T'_{ij}$ as rank 4 binary tensors $(N = 2)$

[Maron 2018, Pan&Kondor 2022]

Equivariant Aggregators

Let's illustrate equivariant mappings $T_{ij} \mapsto T'_{ij}$ as rank 4 binary tensors $(N = 2)$



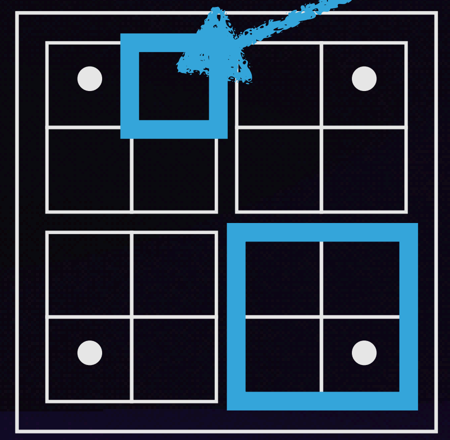
[Maron 2018, Pan&Kondor 2022]

Equivariant Aggregators

Let's illustrate equivariant mappings $T_{ij} \mapsto T'_{ij}$ as rank 4 binary tensors $(N = 2)$

T_{12}

Identity map

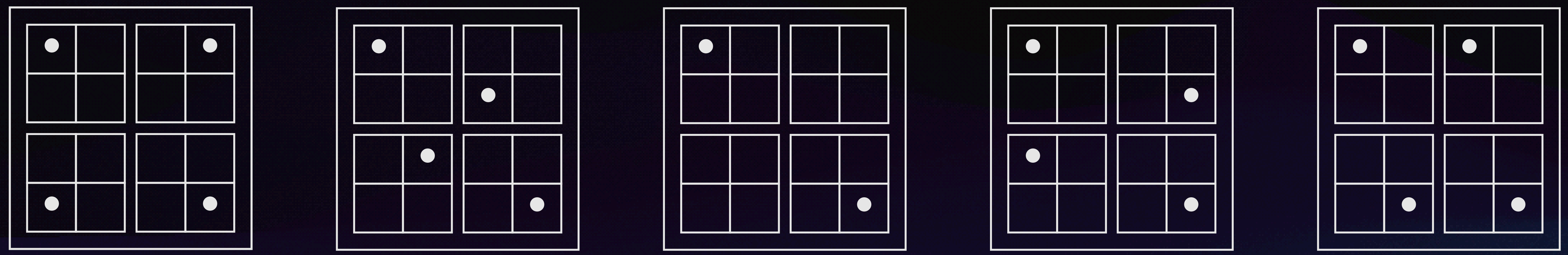


$$T'_{22} = \mathrm{Agg}(\{T_{22}\})$$

[Maron 2018, Pan&Kondor 2022]

Equivariant Aggregators

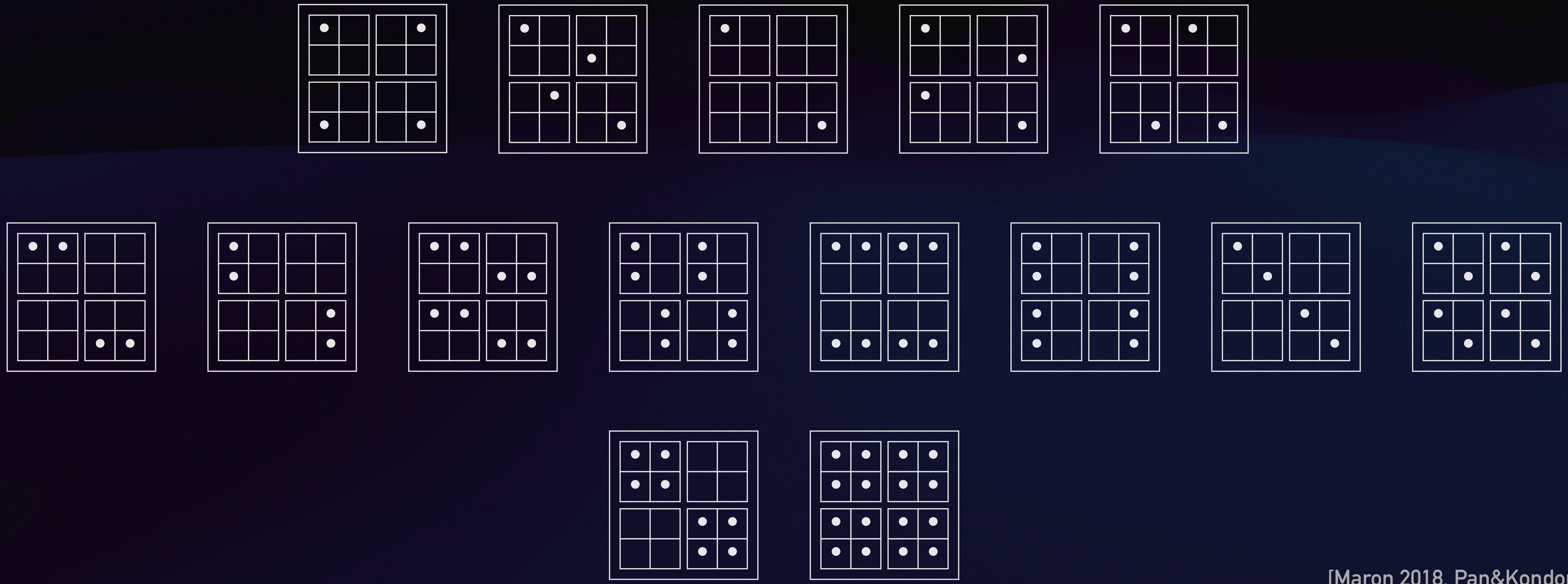
Let's illustrate equivariant mappings $T_{ij} \mapsto T'_{ij}$ as rank 4 binary tensors $(N = 2)$



[Maron 2018, Pan&Kondor 2022]

Equivariant Aggregators

Let's illustrate equivariant mappings $T_{ij} \mapsto T'_{ij}$ as rank 4 binary tensors $(N = 2)$

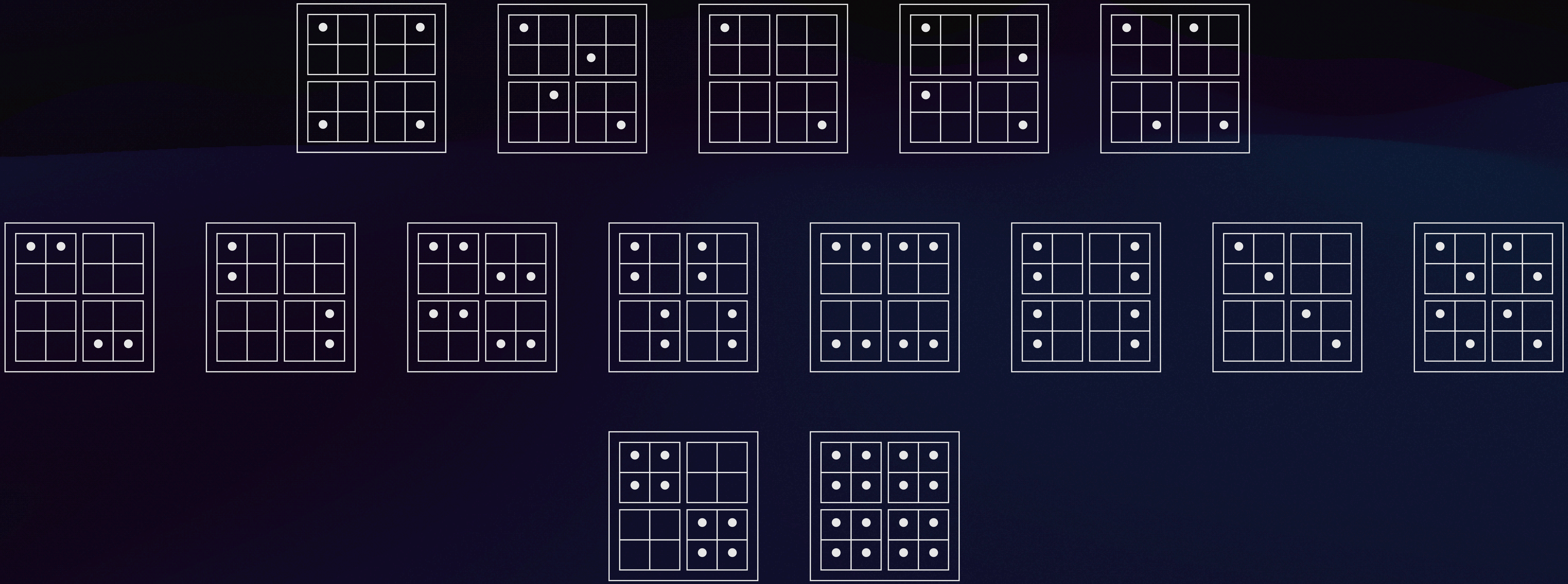


[Maron 2018, Pan&Kondor 2022]

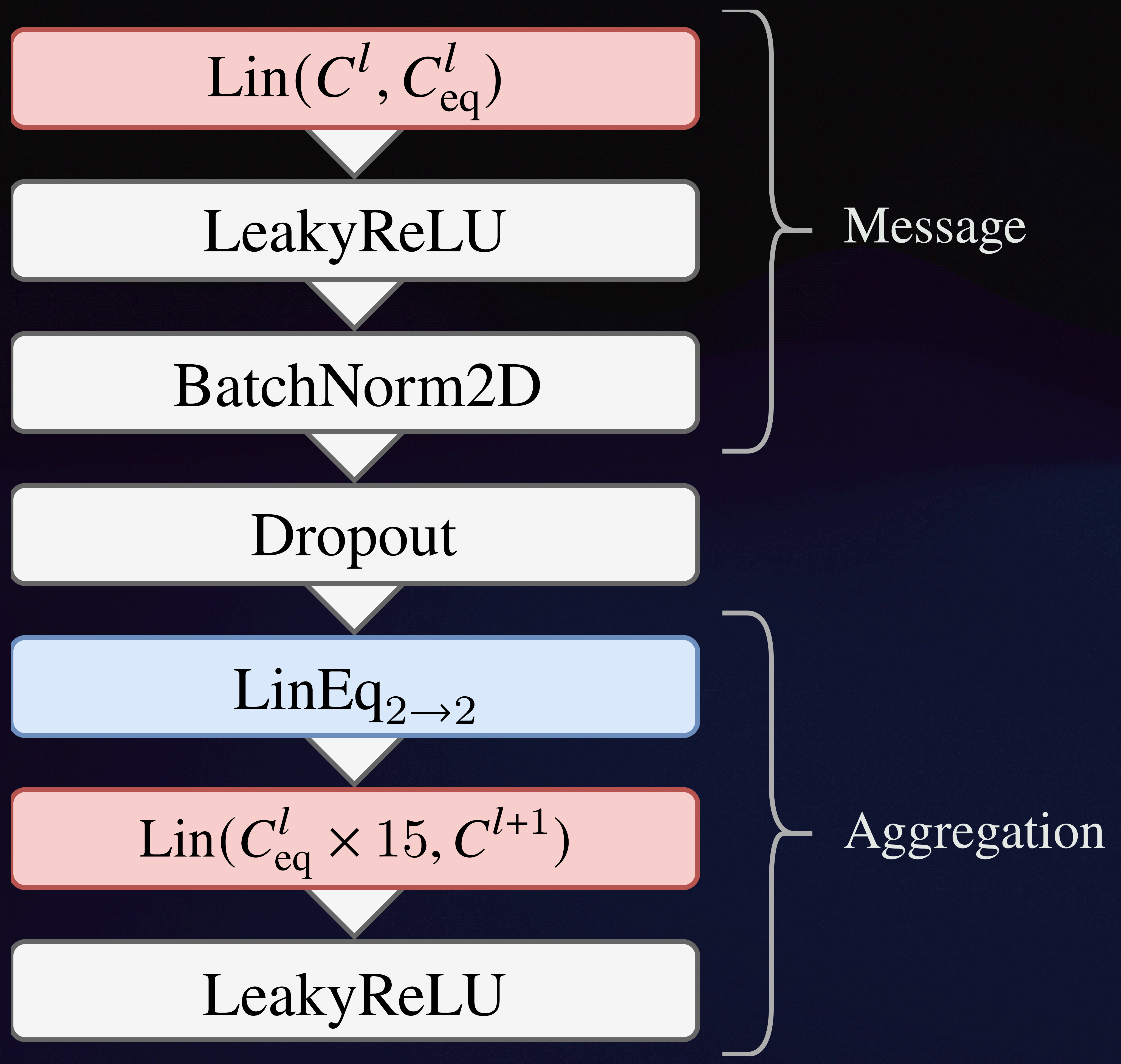
Equivariant Aggregators

The aggregation function can be any symmetric function of a set of inputs.

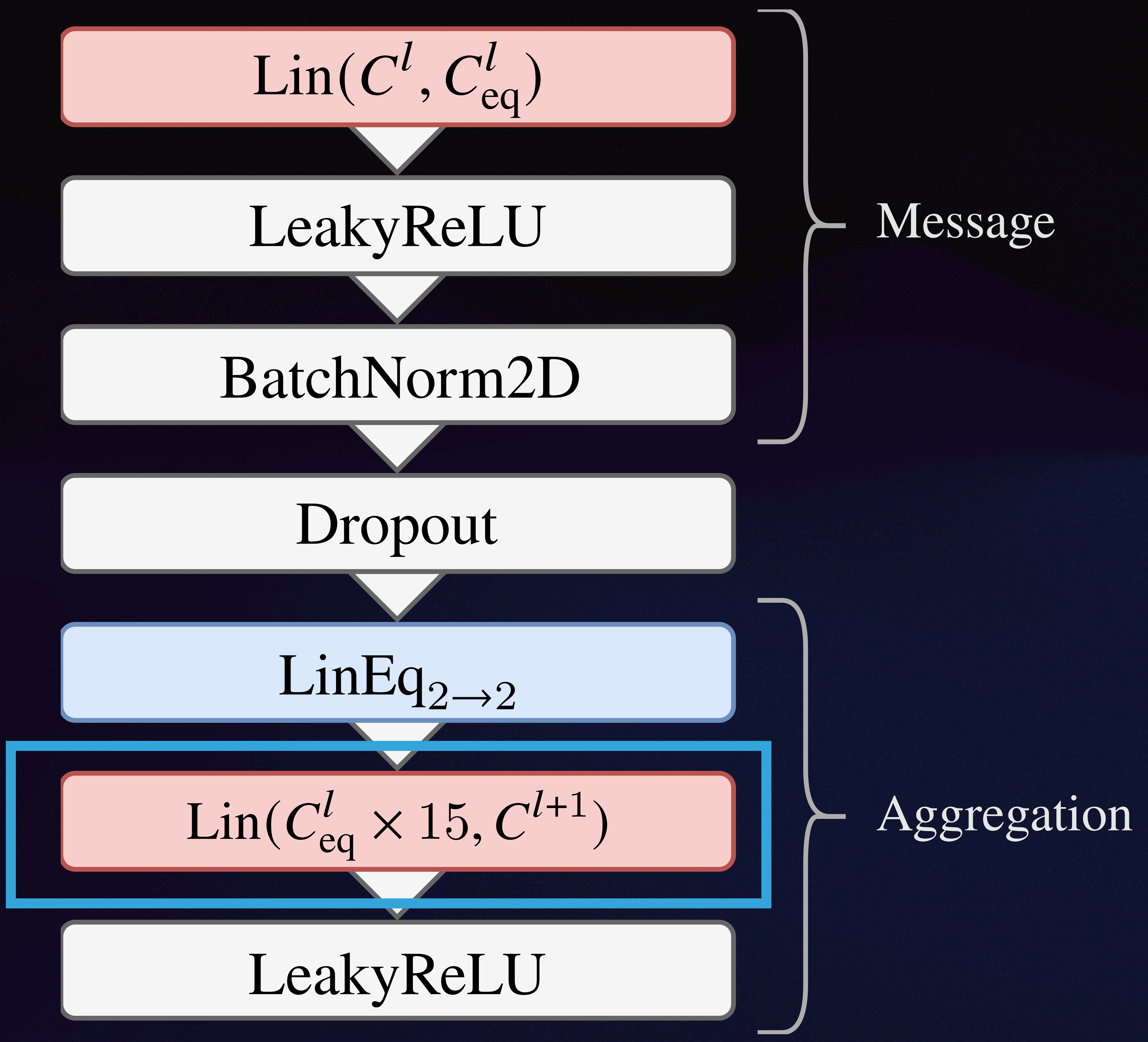
We choose weighted means $N^{\alpha} \sum_{i=1}^{N} x_i$ with learnable α.



PELICAN equivariant block



PELICAN equivariant block

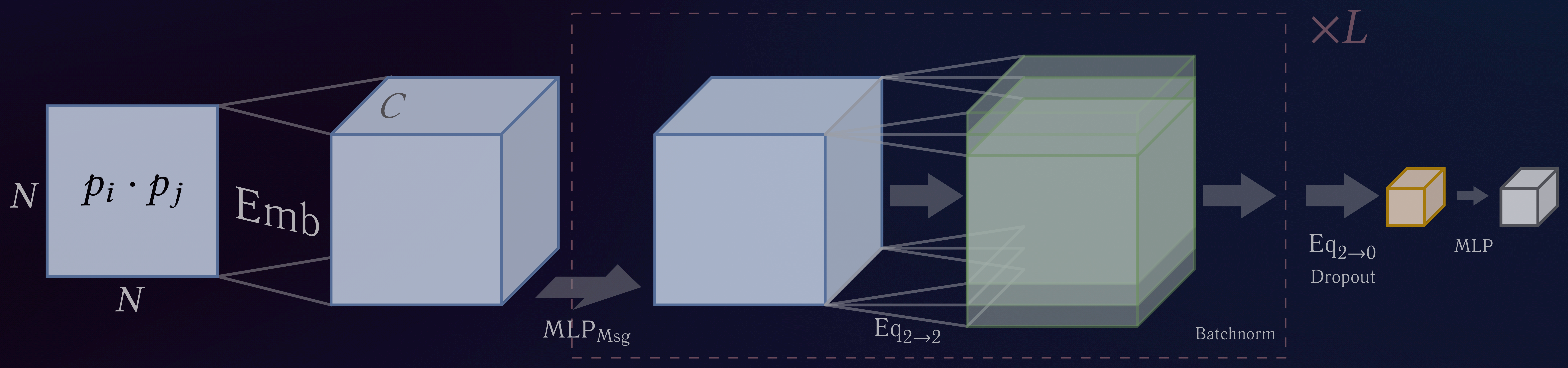


Linear mixing of the outputs of 15 aggregators

PELICAN Classifier

Classifier

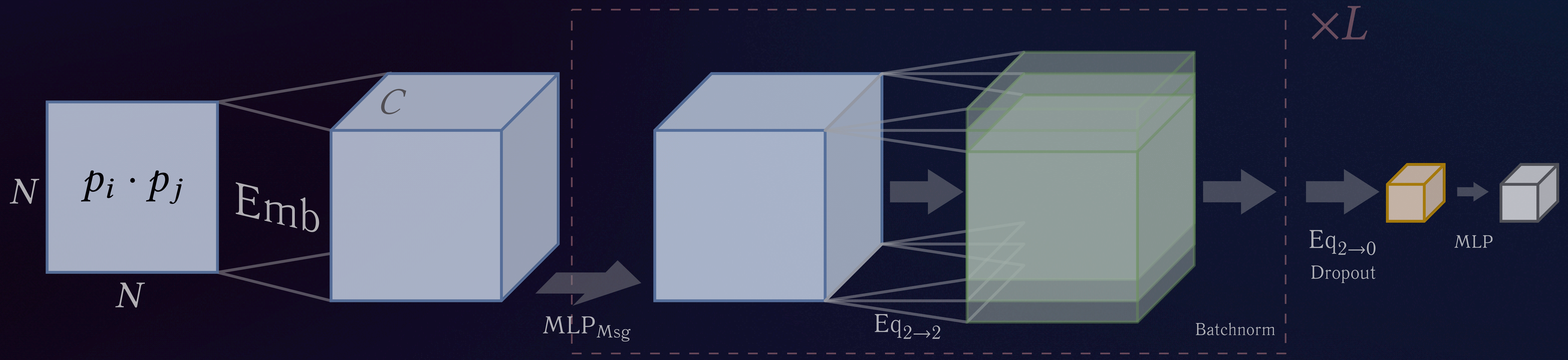
$$p_i \rightarrow d_{ij} \rightarrow \text{Embed} \rightarrow \left[\text{Eq}_{2\to2}\right]^L \rightarrow \text{Eq}_{2\to0} \rightarrow \text{MLP}$$



PELICAN Classifier

Classifier

$$p_i \rightarrow d_{ij} \rightarrow \text{Embed} \rightarrow \left[\text{Eq}_{2\to 2}\right]^L \rightarrow \boxed{\text{Eq}_{2\to 0}} \rightarrow \text{MLP}$$

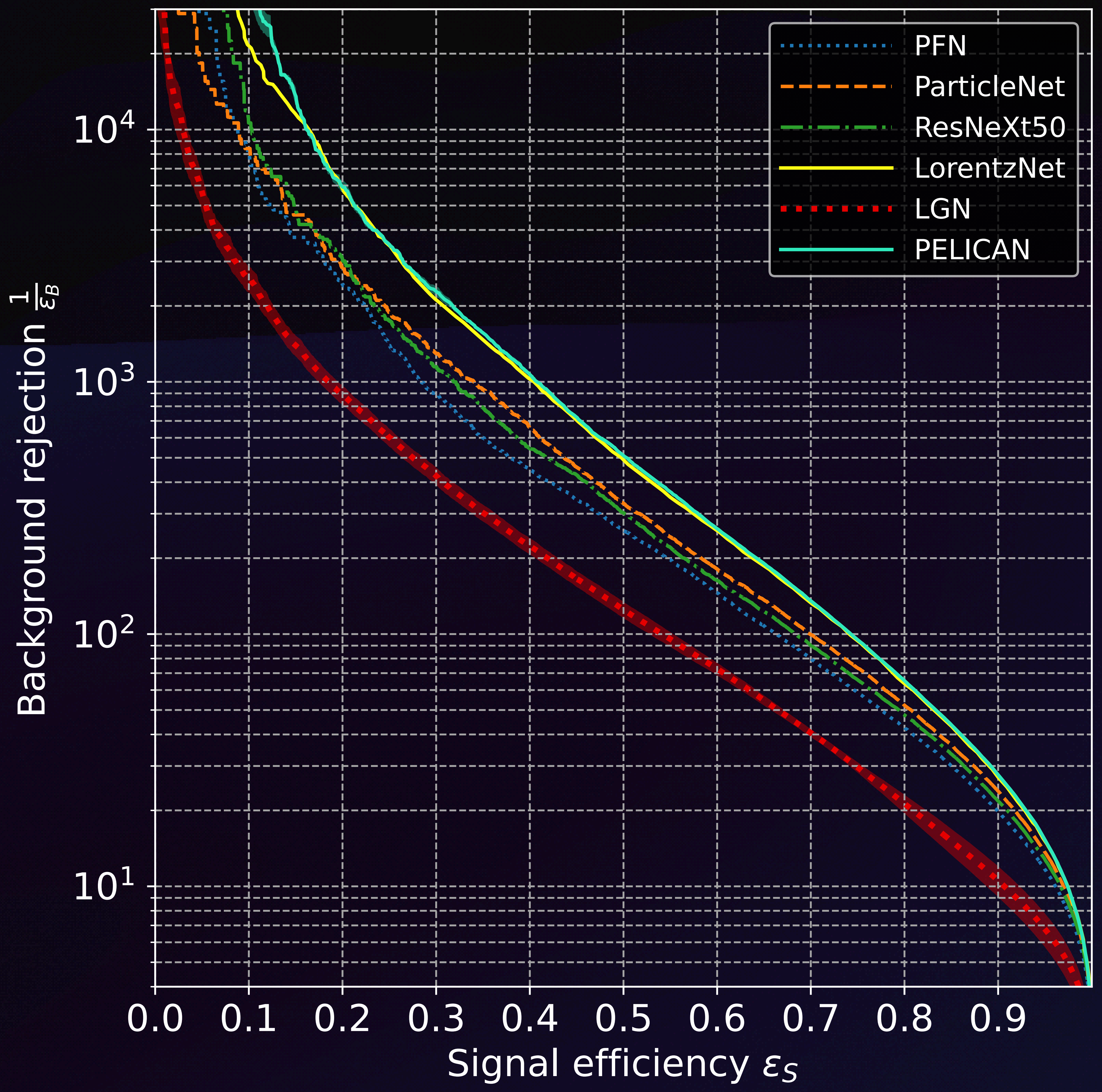


Top-tagging with PELICAN


Jet
b
μ^+
ν_μ
W^+
t
p
p
$\bar{t}$
W^-
$\bar{c}$
Jet
b
$\bar{b}$
Jet
Jet
$p = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix}$

Top-tagging performance

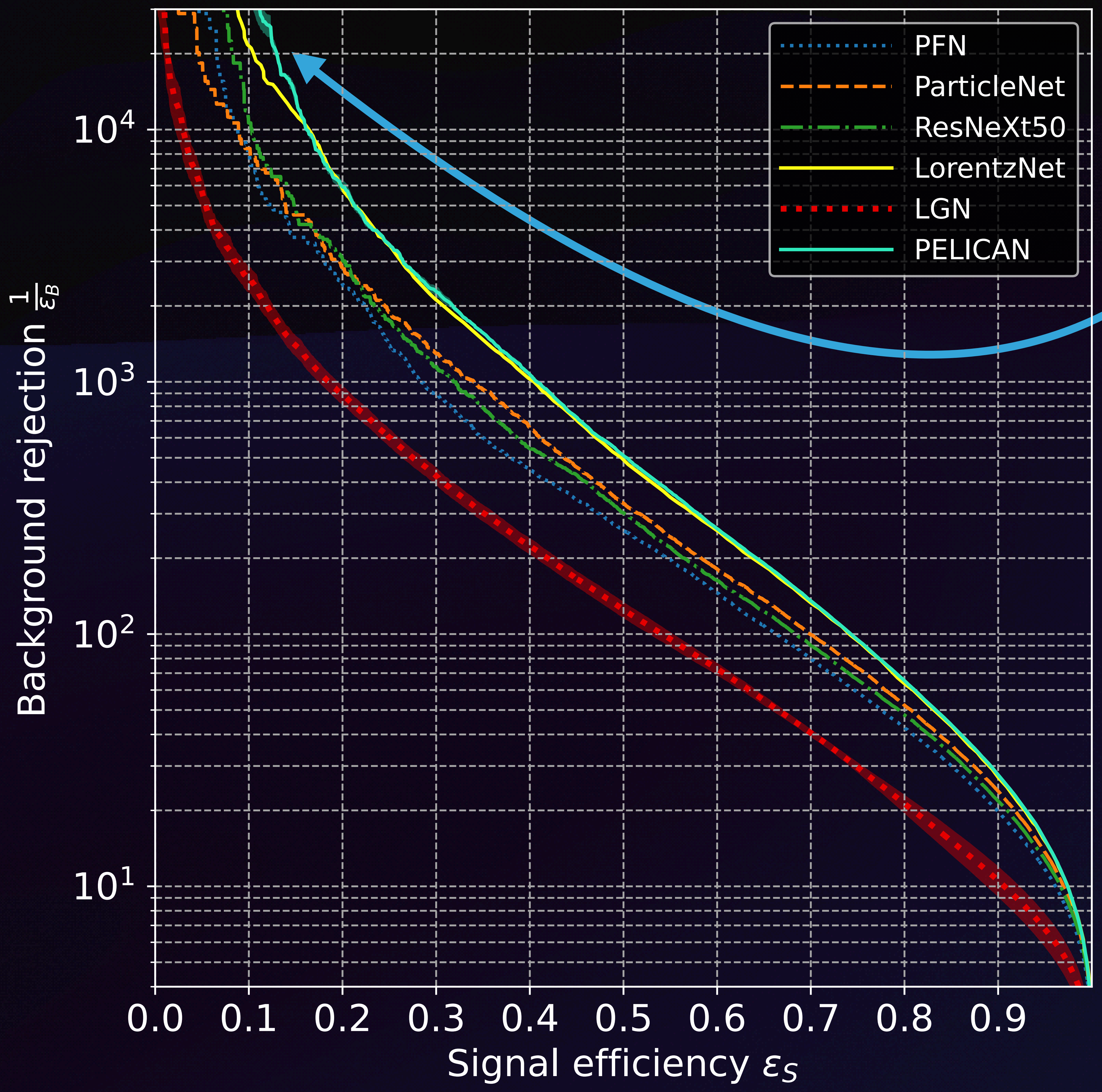
- State of the art top-tagger with 8x fewer params of the previous best tagger
- Exact invariance massively improves sample efficiency



Architecture	Accuracy	AUC	$1/\epsilon_B$	# Params
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
PELICAN	0.9425(1)	0.9869(1)	2289±204	46k

Top-tagging performance

- State of the art top-tagger with 8x fewer params of the previous best tagger
- Exact invariance massively improves sample efficiency

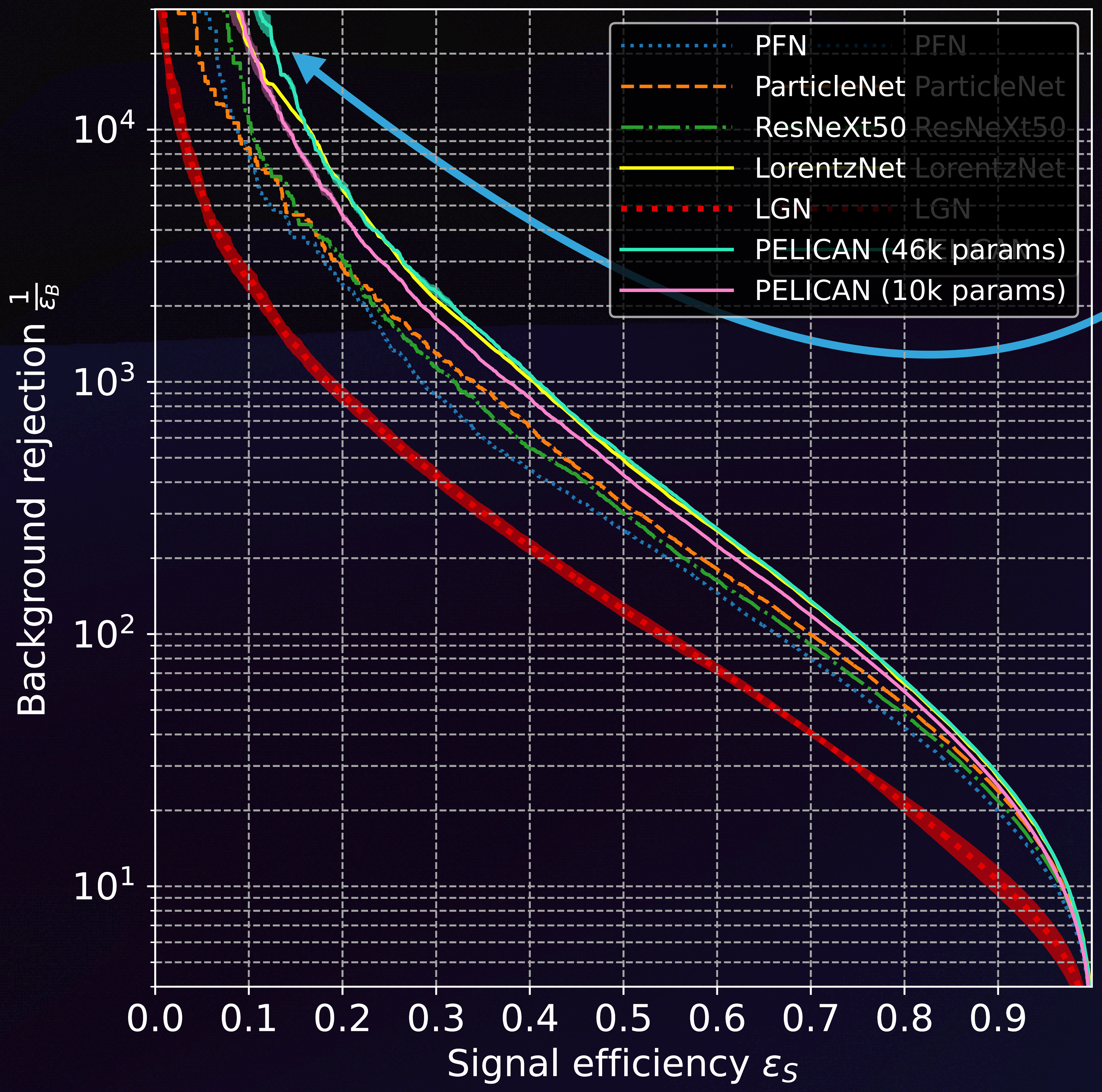


PELICAN

Architecture	Accuracy	AUC	$1/\epsilon_B$	# Params
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
PELICAN	0.9425(1)	0.9869(1)	2289±204	46k

Top-tagging performance

- State of the art top-tagger with 8x fewer params of the previous best tagger
- Exact invariance massively improves sample efficiency

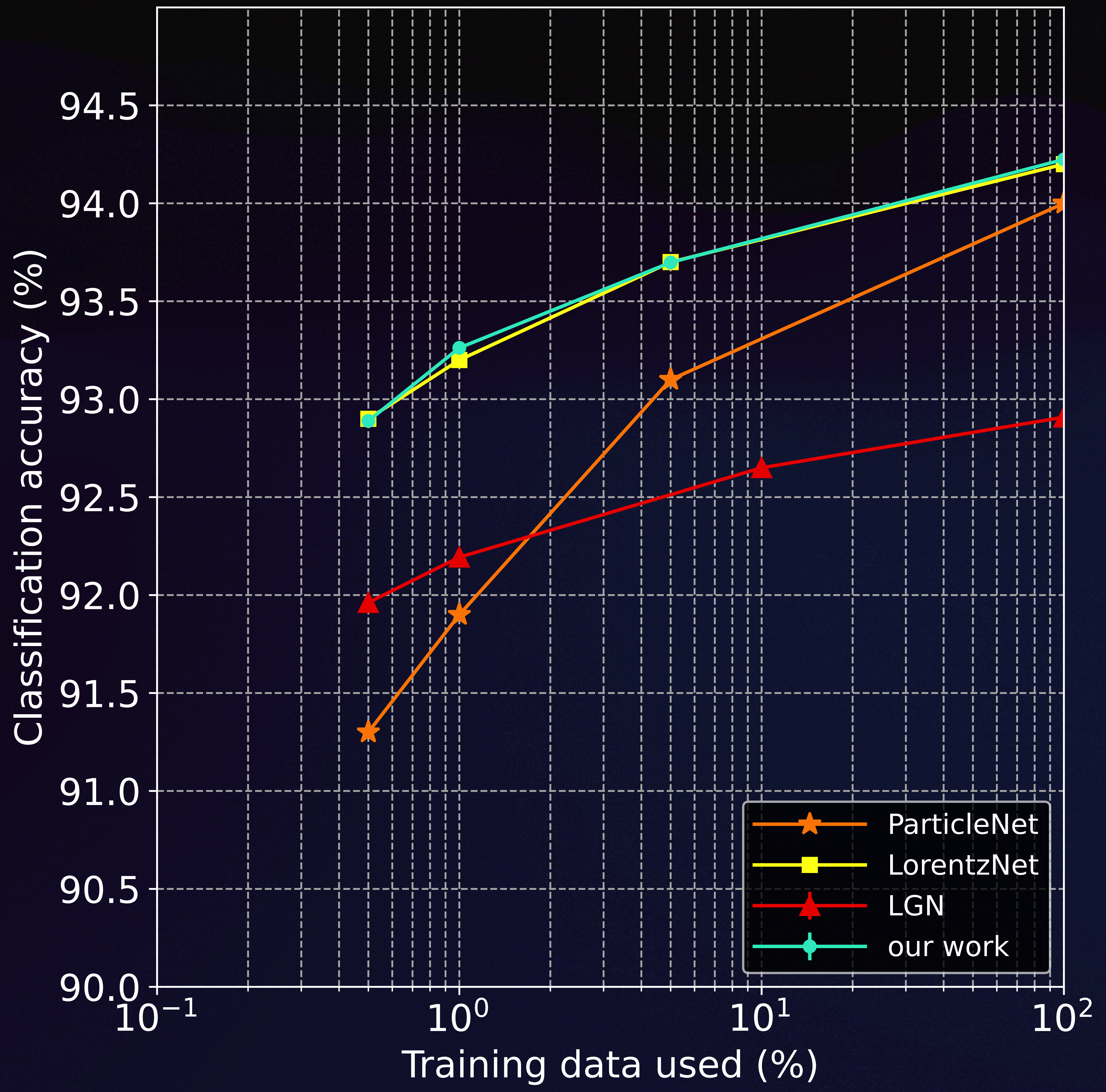


PELICAN

Architecture	Accuracy	AUC	$1/\epsilon_B$	# Params
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
PELICAN	0.9425(1)	0.9869(1)	2289±204	46k

Sample efficiency

- Exact invariance improves sample efficiency



Regression with PELICAN

Lorentz equivariance

All Lorentz-equivariant mappings $F^\mu(p_1, \ldots, p_N)$ have the form

$$F(p_1, \ldots, p_N) = \sum_i c_i p_i,$$

where $c_i(p_1, \ldots, p_N)$ are Lorentz-invariant, i.e. only functions of d_{ij}.

Lorentz equivariance

All Lorentz-equivariant mappings $F^\mu(p_1, \ldots, p_N)$ have the form

$$F(p_1, \ldots, p_N) = \sum_i c_i p_i,$$

where $c_i(p_1, \ldots, p_N)$ are Lorentz-invariant, i.e. only functions of d_{ij}.

What about permutation invariance?

Lorentz-equivariant

$$F^{\mu}(p_1, \ldots, p_N)$$

Lorentz-equivariant

$$F^{\mu}(p_1, \ldots, p_N)$$

Permutation-equivariant

$$c_i\left(\{d_{jk}\}\right)$$

Lorentz-equivariant

$$F^\mu(p_1, \ldots, p_N)$$

Permutation-equivariant

$$c_i\left(\{d_{jk}\}\right)$$

$$F^\mu(p_1, \ldots, p_N) = \sum_i c_i p_i^\mu$$

PELICAN 4-momentum Regressor

4-momentum regressor

$$p_i \rightarrow d_{ij} \rightarrow \text{Embed} \rightarrow \left[\text{Eq}_{2\rightarrow 2}\right]^L \rightarrow \boxed{\text{Eq}_{2\rightarrow 1}} \rightarrow \text{MLP} \rightarrow \{c_i\}_{i=1}^N$$

PELICAN 4-momentum Regressor

4-momentum regressor

PELICAN weights

$$p_i \to d_{ij} \to \text{Embed} \to \left[\text{Eq}_{2\to 2}\right]^L \to \boxed{\text{Eq}_{2\to 1}} \to \text{MLP} \to \boxed{\{c_i\}_{i=1}^N}$$

PELICAN 4-momentum Regressor

4-momentum regressor

PELICAN weights

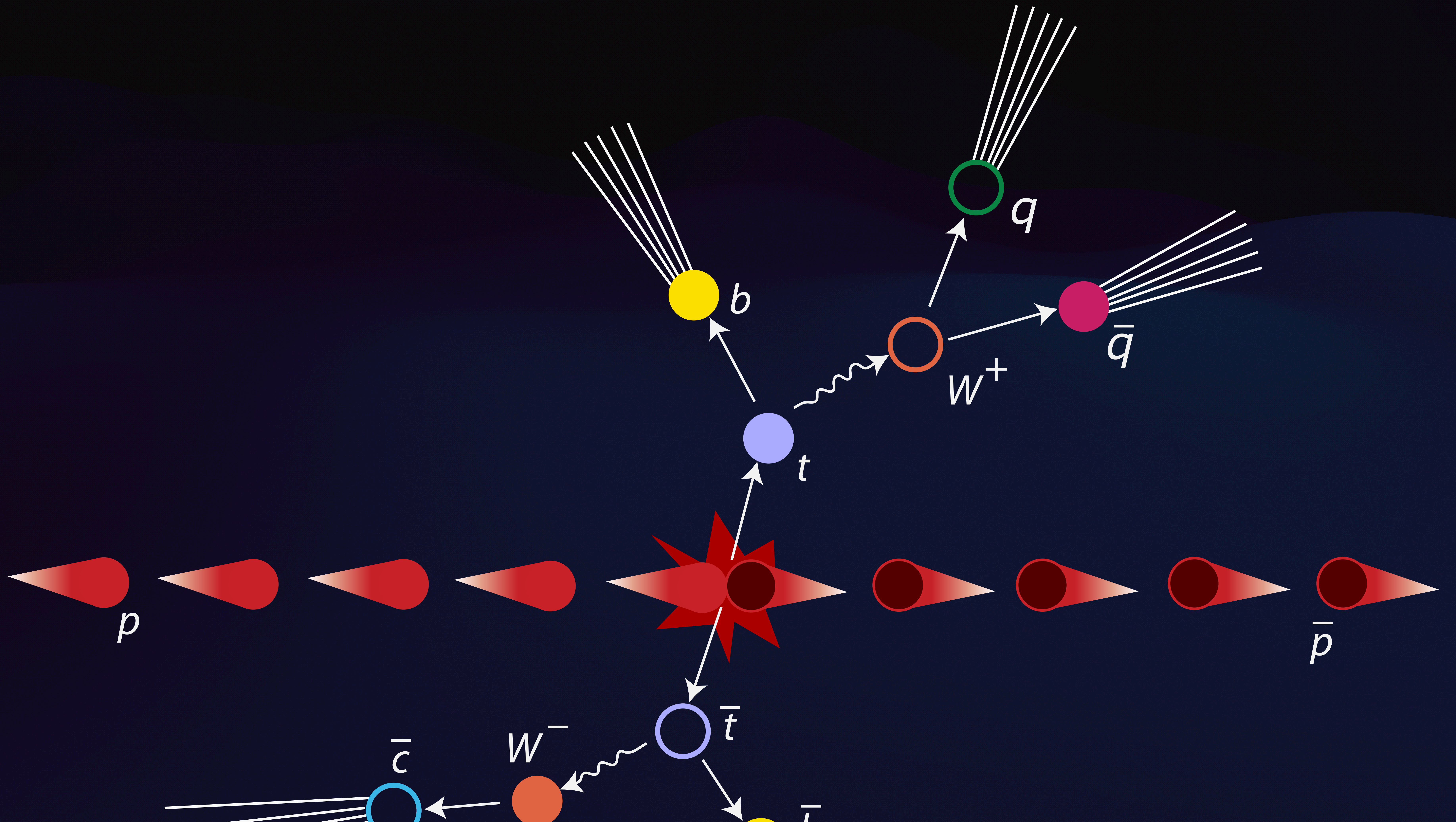
$$p_i \to d_{ij} \to \text{Embed} \to \left[\text{Eq}_{2\to 2}\right]^L \to \boxed{\text{Eq}_{2\to 1}} \to \text{MLP} \to \boxed{\{c_i\}_{i=1}^N}$$

The 4-momentum is reconstructed as $p_{\text{predict}} = \sum_{i=1}^{N} c_i p_i$

The loss function is a linear combination of $|m_{\text{predict}} - m_{\text{target}}|$ and $|\vec{p}_{\text{predict}} - \vec{p}_{\text{target}}|$

Hadronic W reconstruction inside of a top jet

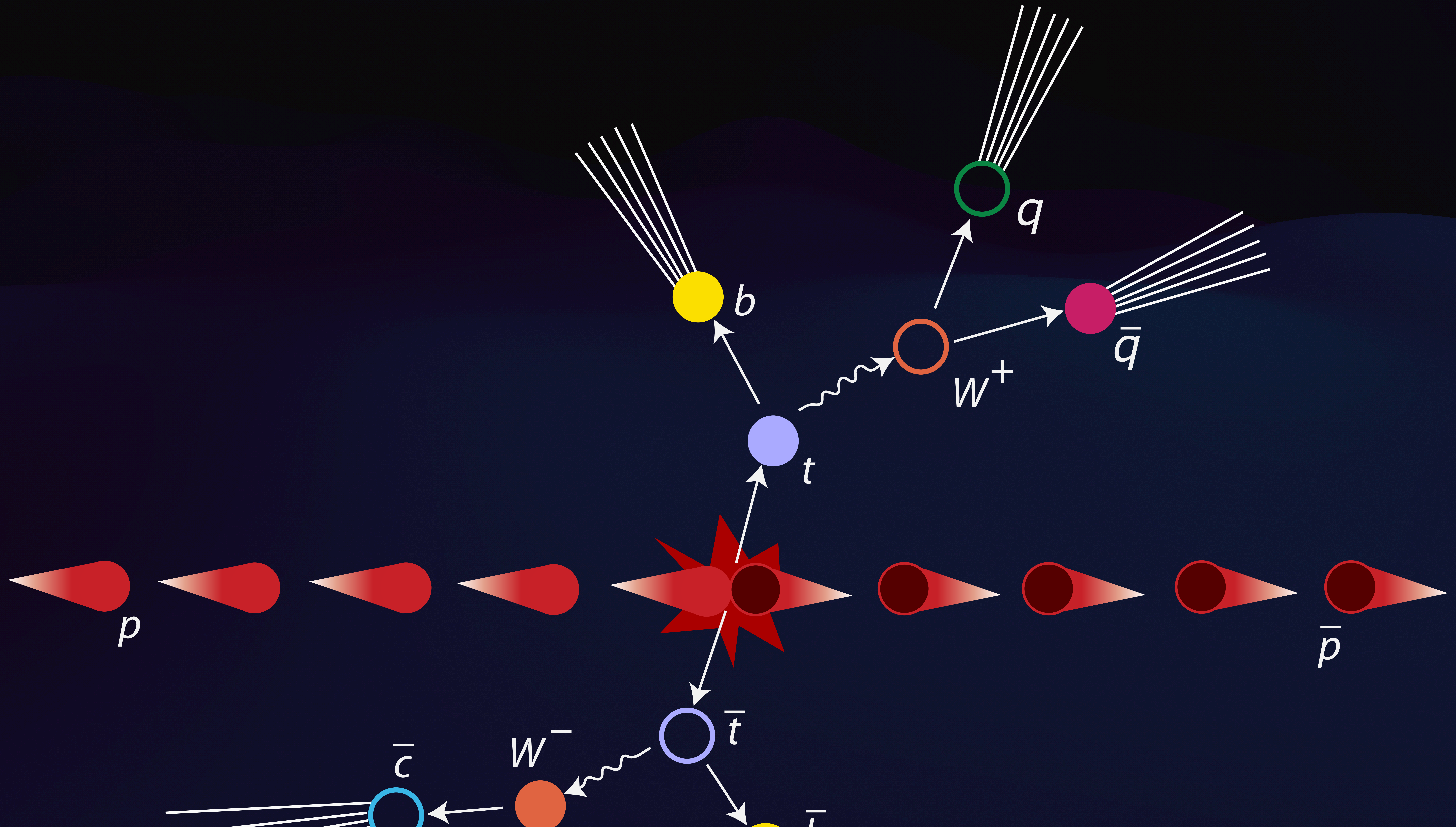
Dataset similar to the top-tagging one:



Hadronic W reconstruction inside of a top jet

Dataset similar to the top-tagging one:

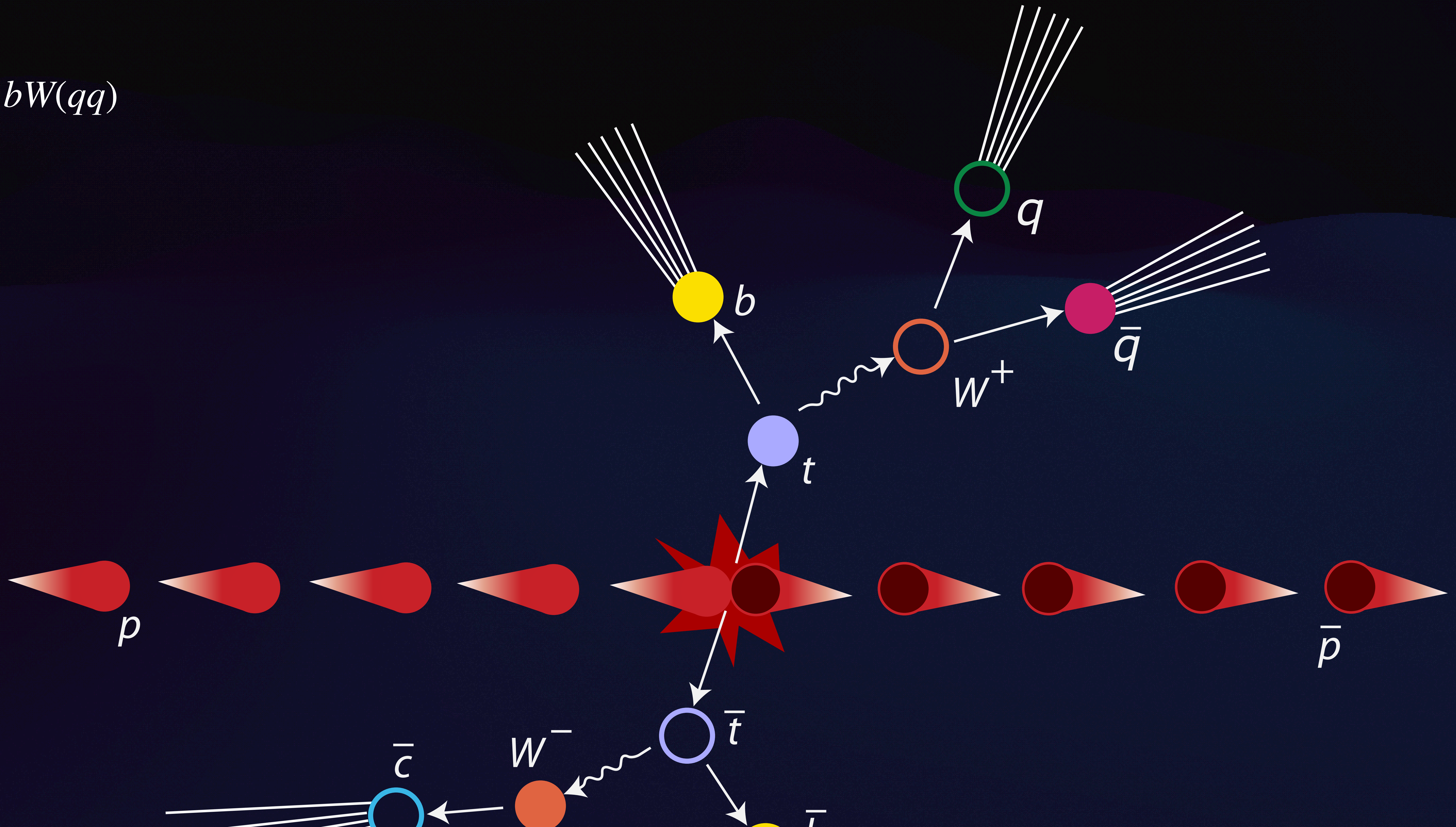
1. Only top jets



Hadronic W reconstruction inside of a top jet

Dataset similar to the top-tagging one:

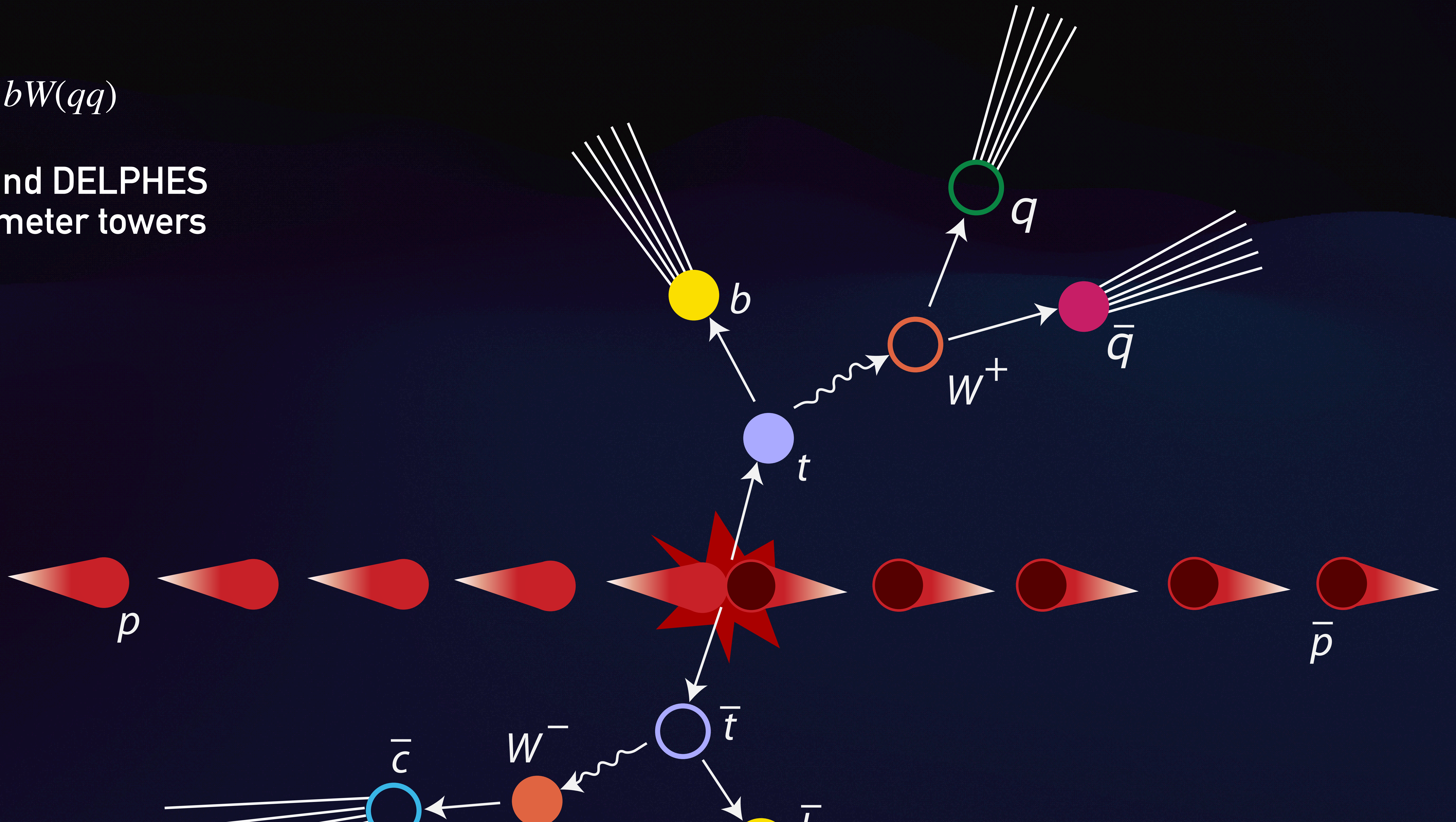
1. Only top jets
2. Only hadronic decays $t \rightarrow bW(qq)$



Hadronic W reconstruction inside of a top jet

Dataset similar to the top-tagging one:

1. Only top jets
2. Only hadronic decays $t \to bW(qq)$
3. Two versions: truth level and DELPHES reconstructed from calorimeter towers



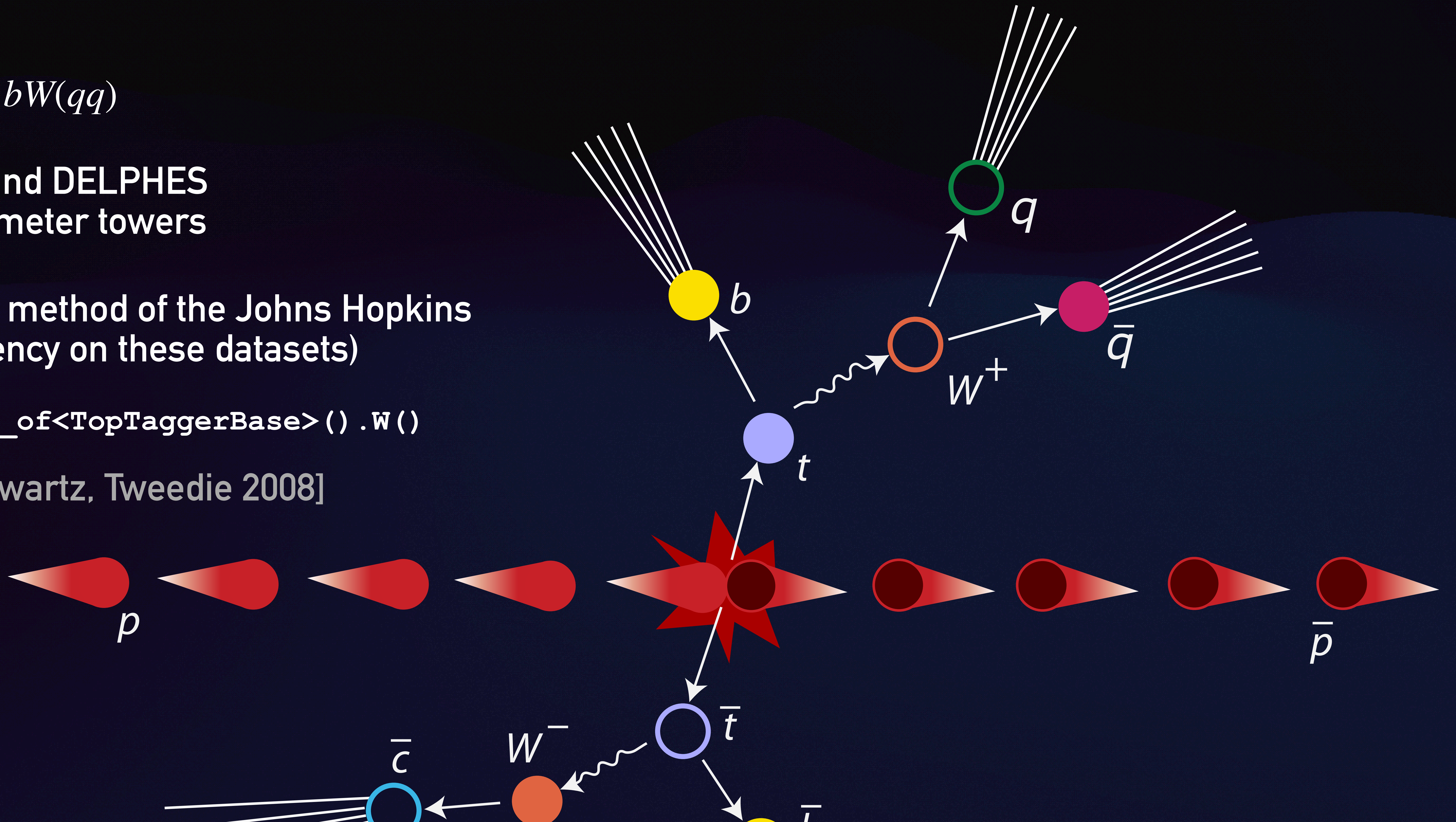
Hadronic W reconstruction inside of a top jet

Dataset similar to the top-tagging one:

1. Only top jets
2. Only hadronic decays $t \to bW(qq)$
3. Two versions: truth level and DELPHES reconstructed from calorimeter towers
4. Non-ML baseline: built-in method of the Johns Hopkins top-tagger (37%/31% efficiency on these datasets)

```
top_candidate.structure_of<TopTaggerBase>().W()
```

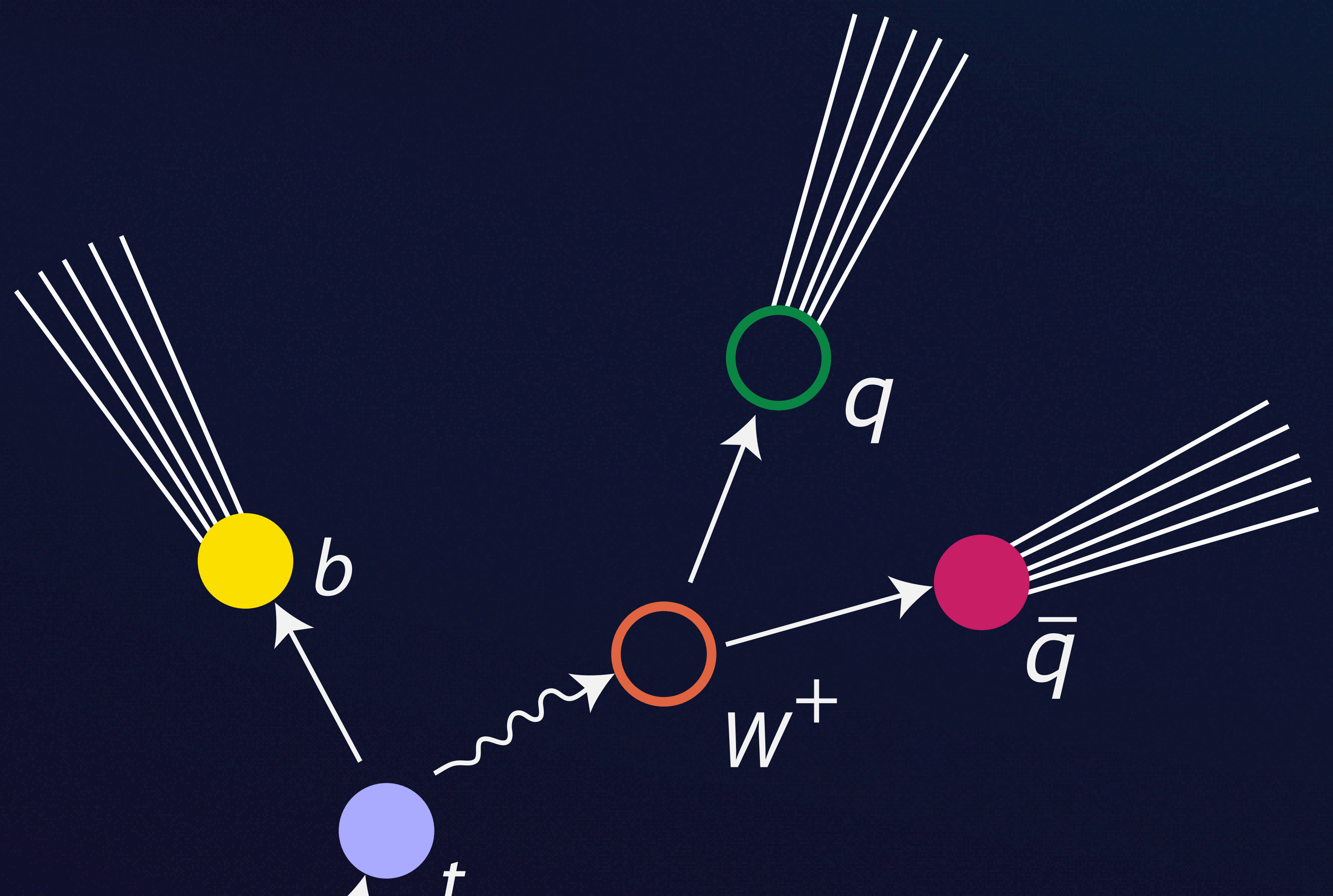
[Kaplan, Rehermann, Schwartz, Tweedie 2008]



Hadronic W reconstruction inside of a top jet

JH = Johns Hopkins Top Tagger
[Kaplan, Rehermann, Schwartz, Tweedie 2008]

PELICAN | JH = PELICAN on JH-tagged events

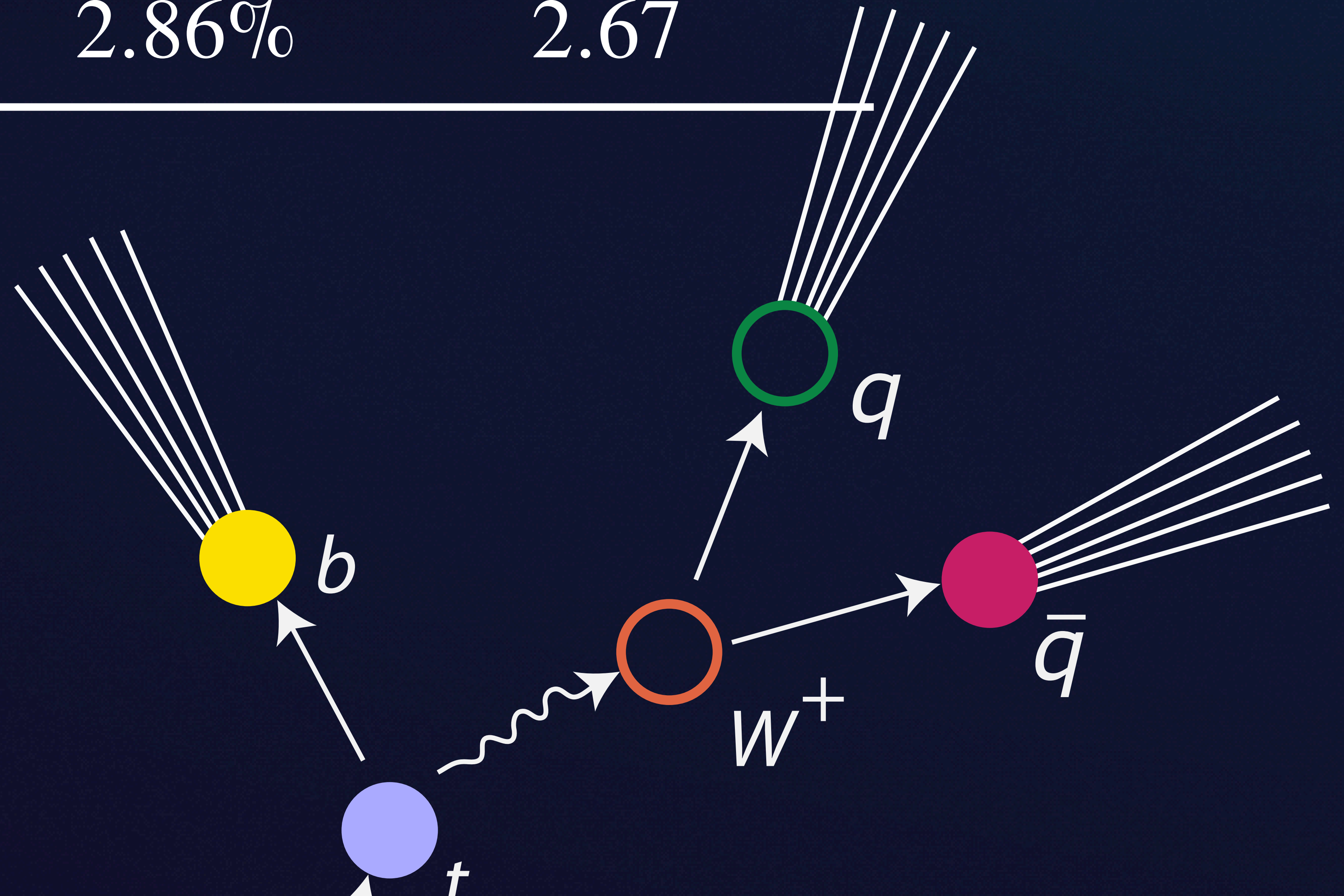


Hadronic W reconstruction inside of a top jet

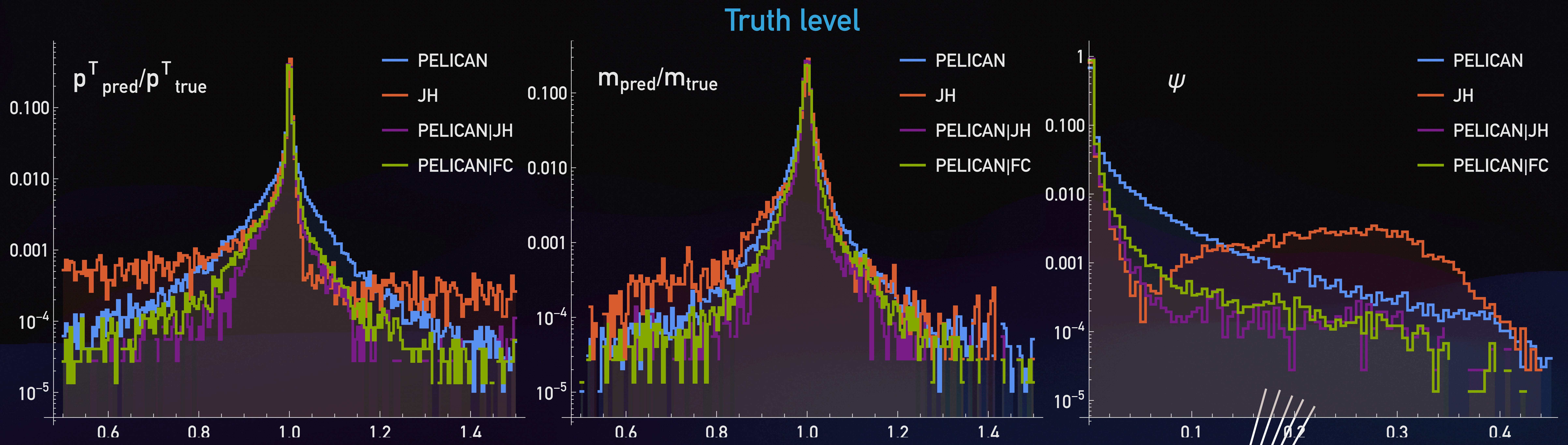
	Method	σ_{p_T} (%)	σ_m (%)	σ_ψ (centirad)
Without DELPHES	JH	0.65%	1.27%	0.156
	PELICAN	0.82%	1.20%	0.384
	PELICAN|JH	0.27%	0.59%	0.088
With DELPHES	JH	10.3 %	8.3 %	8.73
	PELICAN	5.51%	3.22%	4.16
	PELICAN|JH	3.81%	2.86%	2.67

JH = Johns Hopkins Top Tagger
[Kaplan, Rehermann, Schwartz, Tweedie 2008]

PELICAN | JH = PELICAN on JH-tagged events

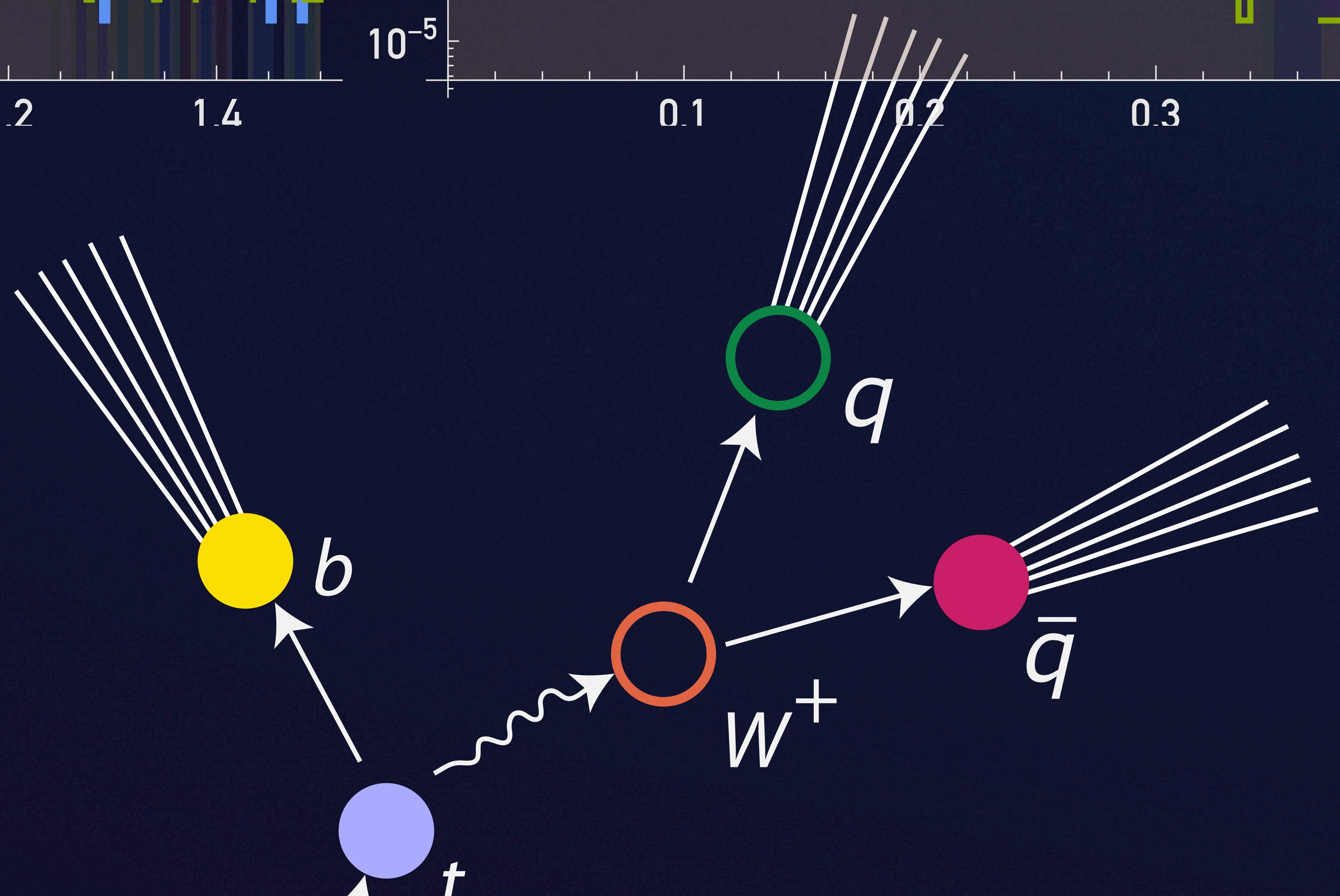


Hadronic W reconstruction inside of a top jet

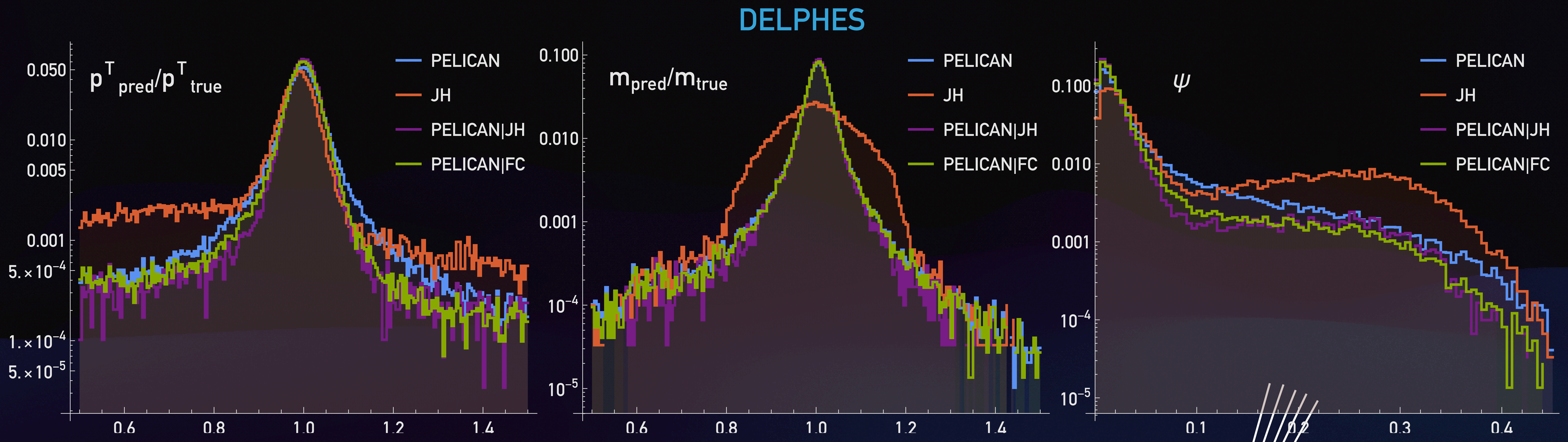


JH = Johns Hopkins Top Tagger
[Kaplan, Rehermann, Schwartz, Tweedie 2008]

PELICAN | JH = PELICAN on JH-tagged events

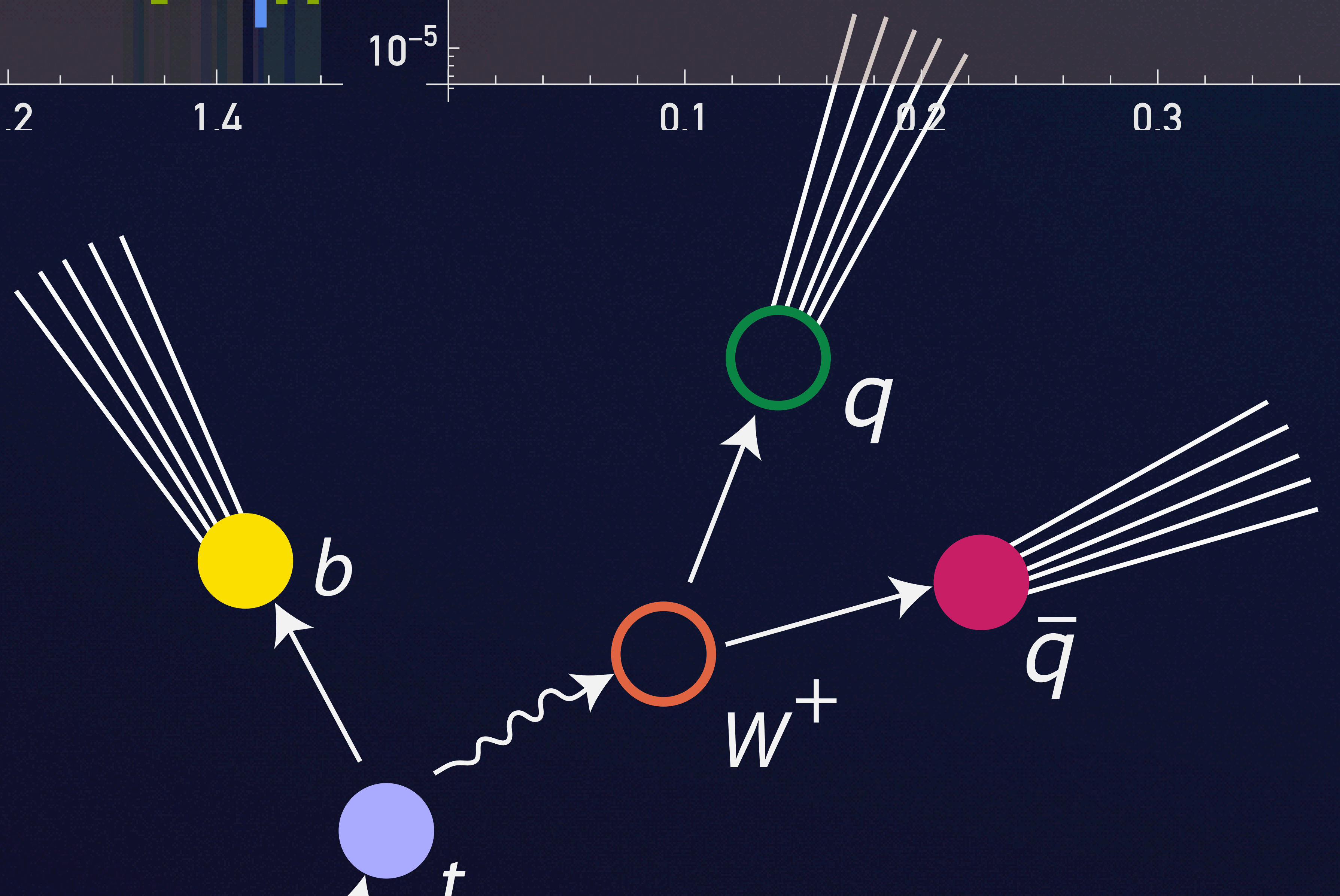


Hadronic W reconstruction inside of a top jet



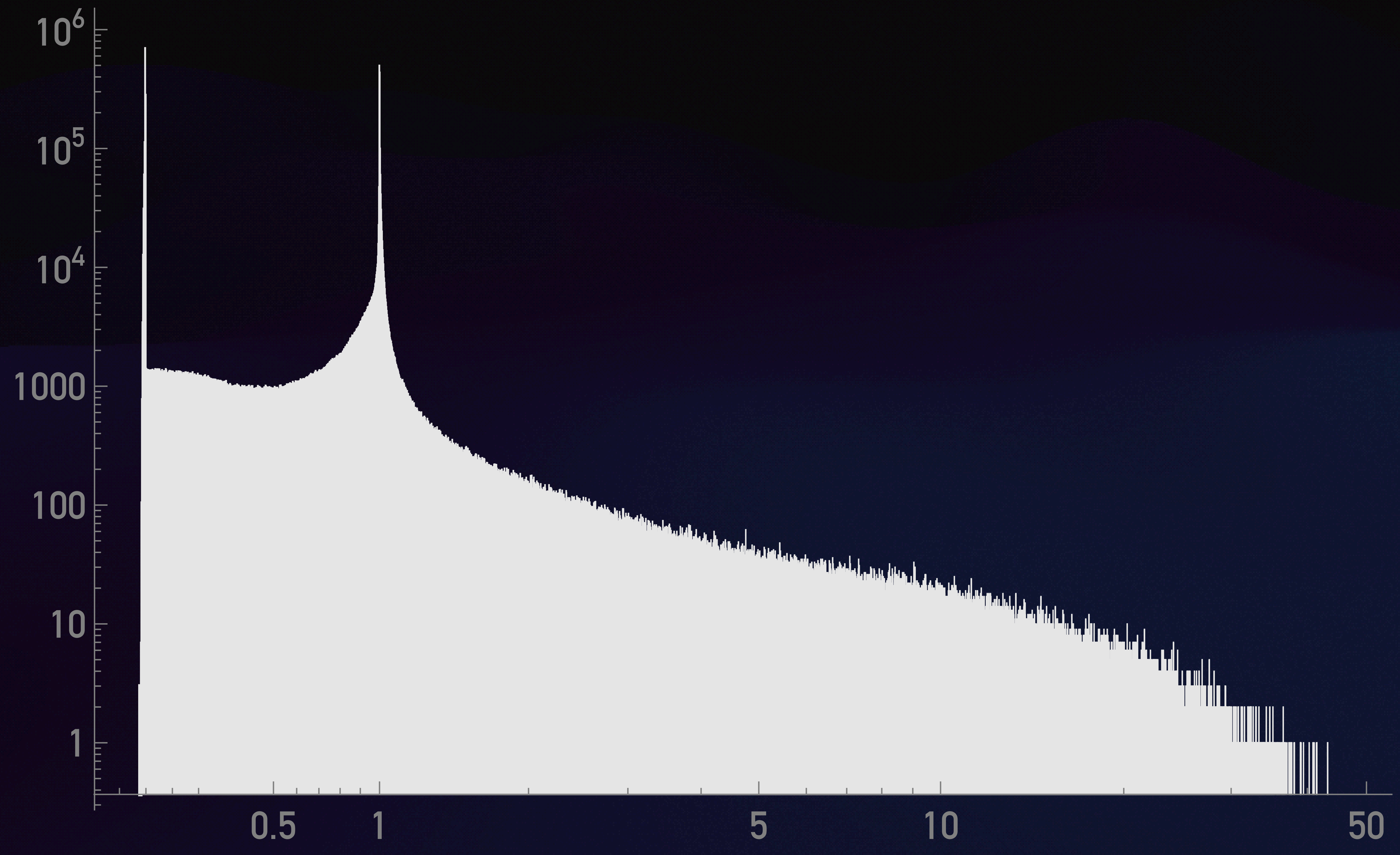
JH = Johns Hopkins Top Tagger
[Kaplan, Rehermann, Schwartz, Tweedie 2008]

PELICAN | JH = PELICAN on JH-tagged events



Explaining PELICAN

Distribution of output PELICAN weights on truth-level dataset
(~100k events)

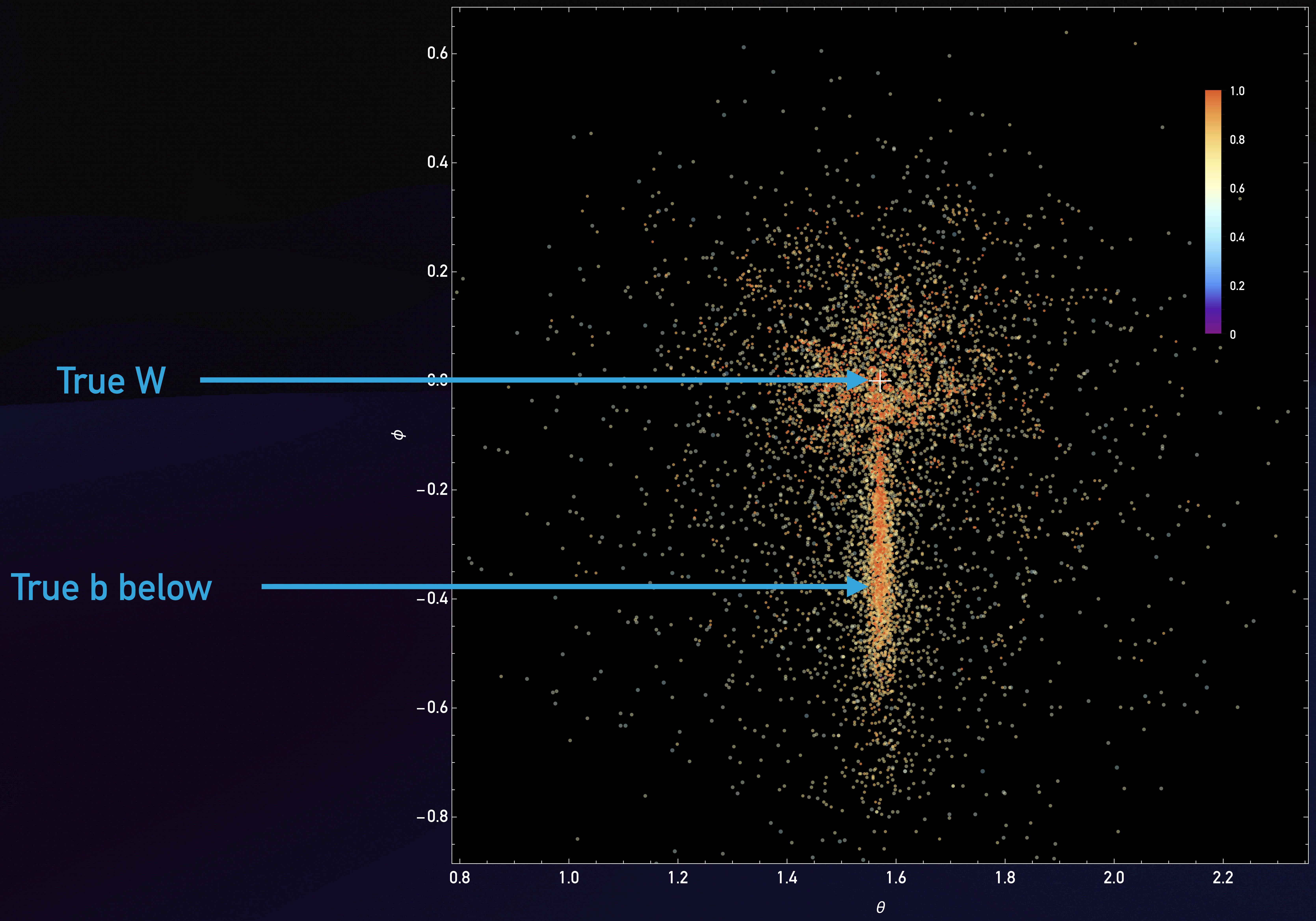


$\{c_i\}_{i=1}^{N}$

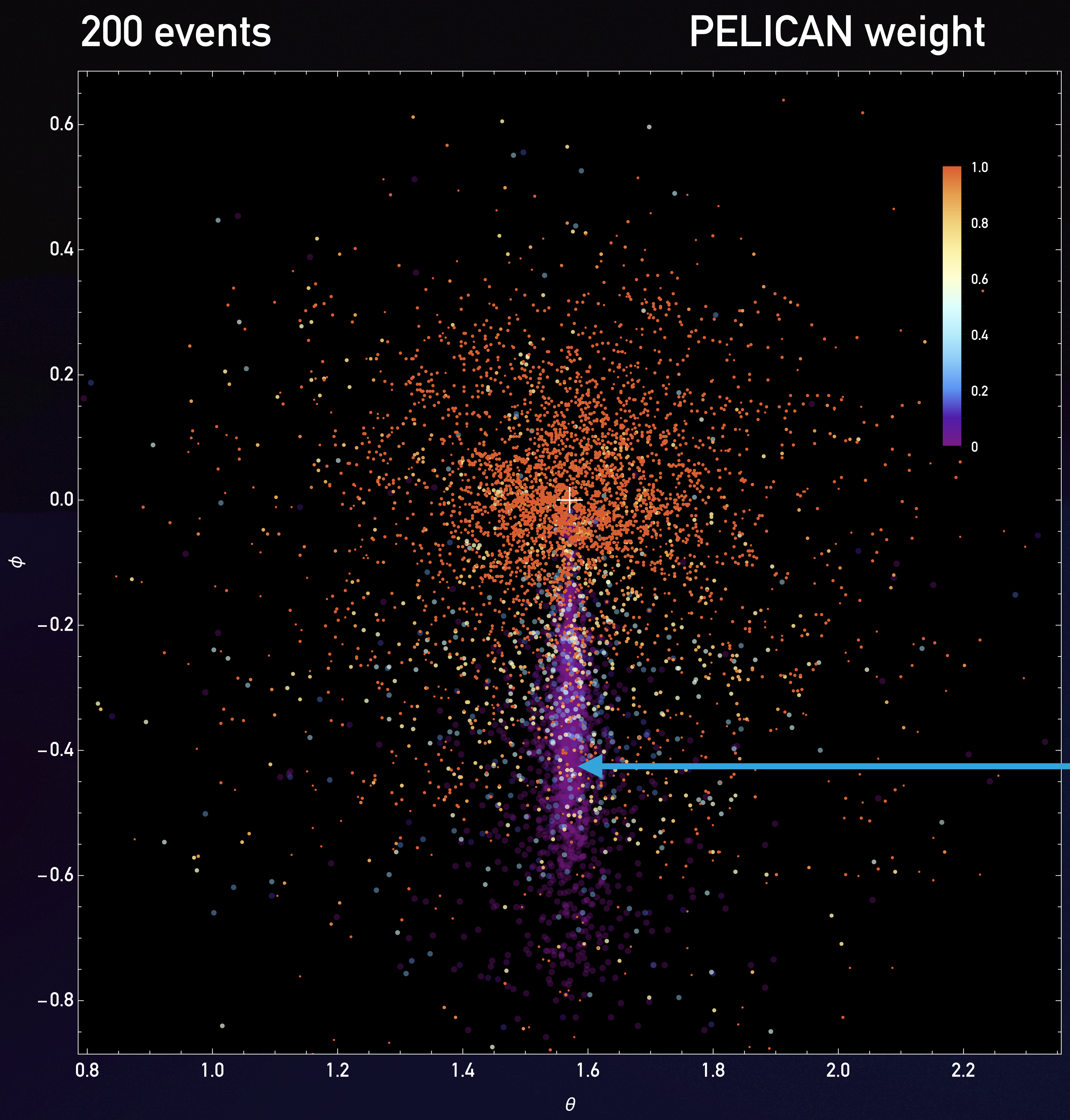
Explaining PELICAN

200 events

Energy of constituents

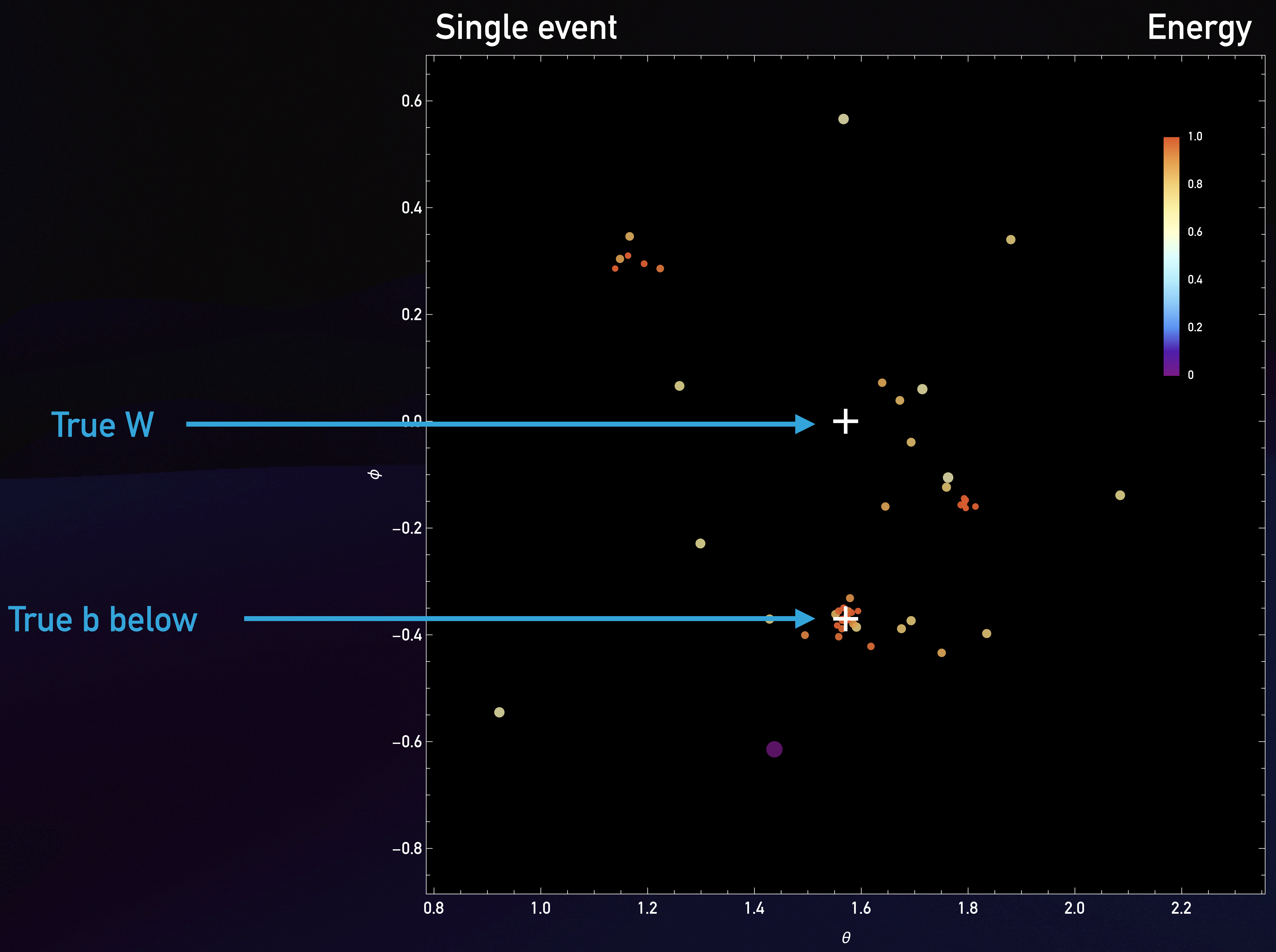


Explaining PELICAN



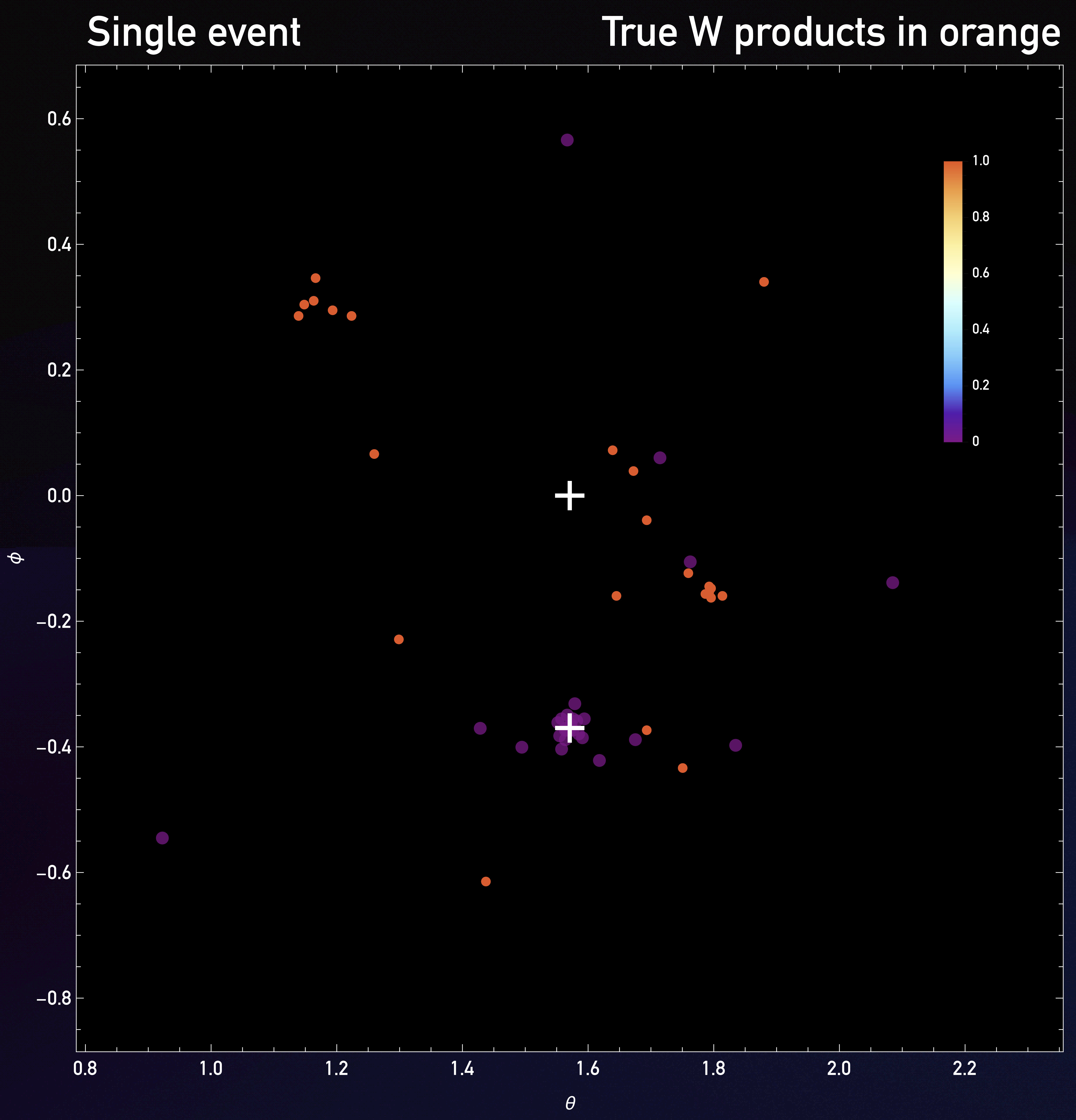
PELICAN learns to separate out the b-quark cluster

Truth level

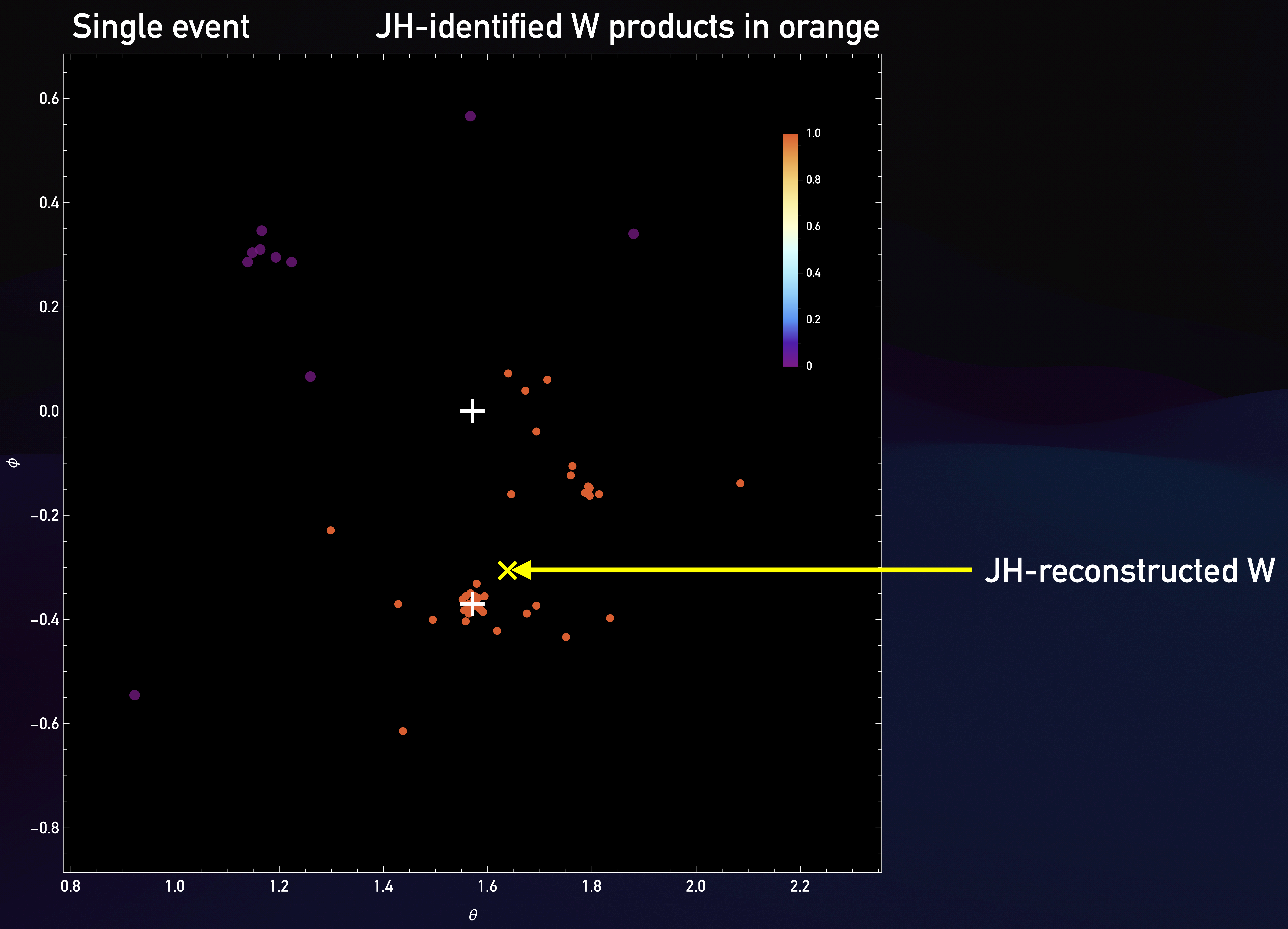


Single event
Energy
True W
True b below
1.0
0.8
0.6
0.4
0.2
0
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
ϕ
θ

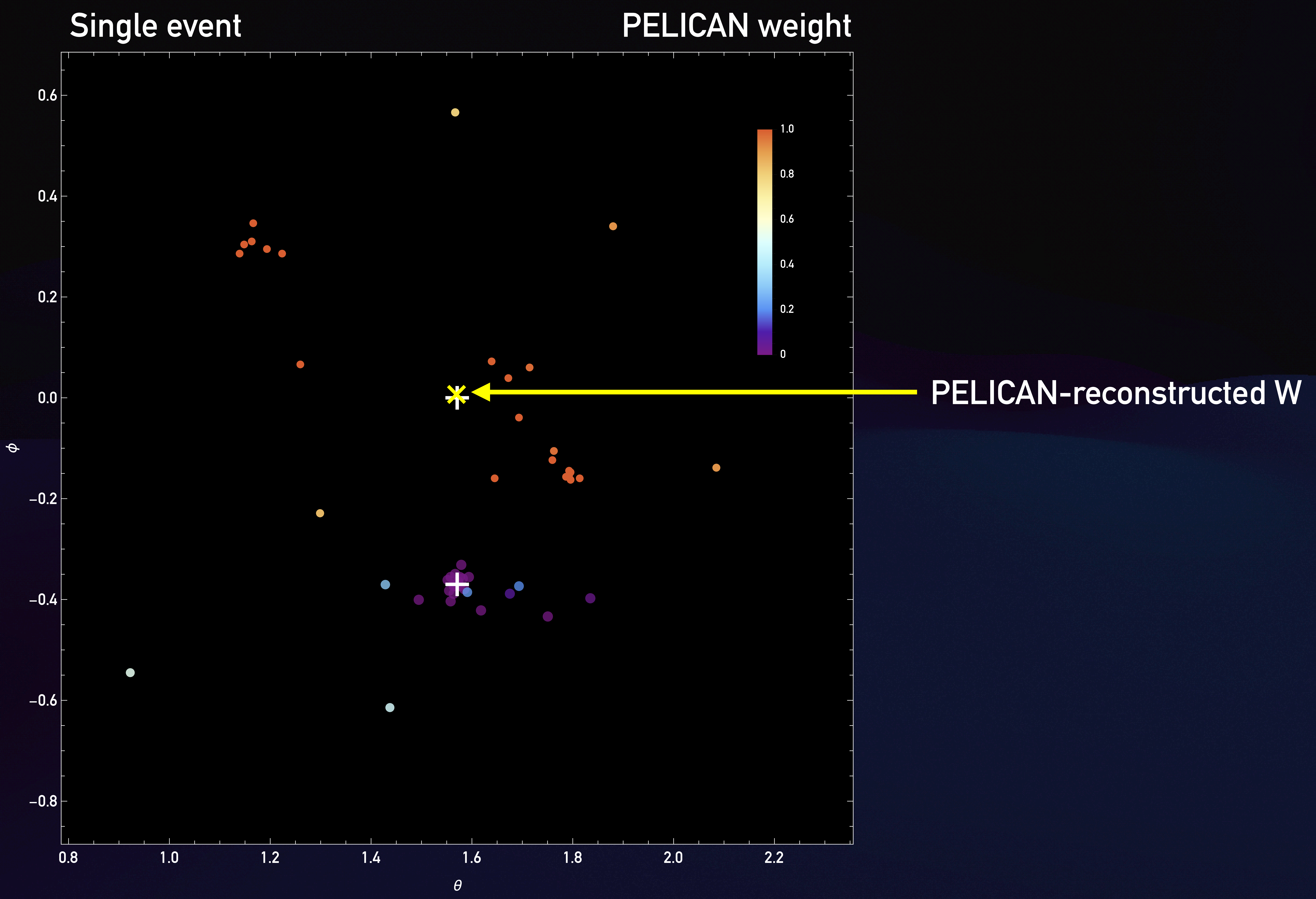
Truth level


Single event
True W products in orange
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
ϕ
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
θ
1.0
0.8
0.6
0.4
0.2
0

Truth level

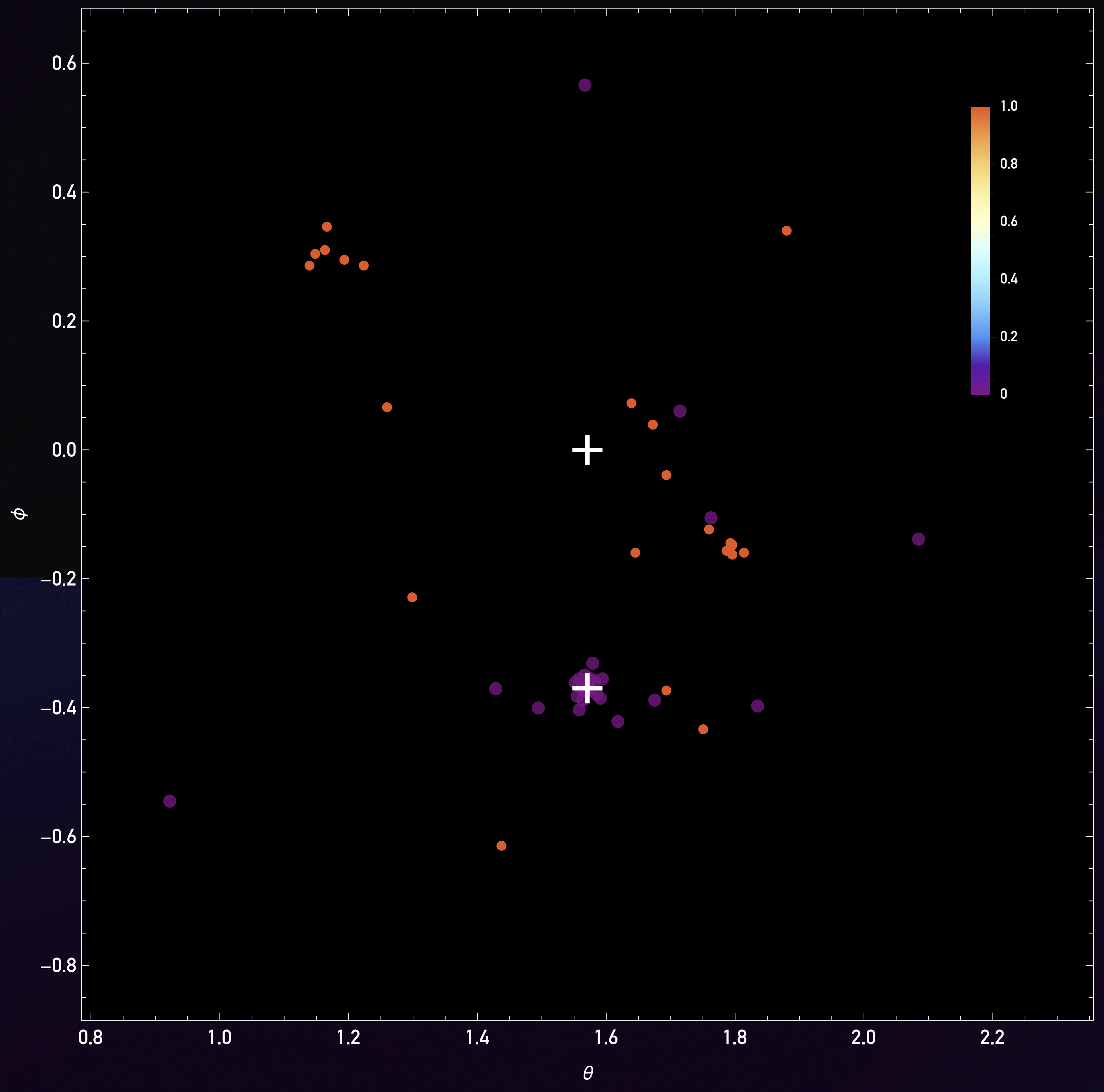

Single event
JH-identified W products in orange
JH-reconstructed W
1.0
0.8
0.6
0.4
0.2
0
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
ϕ
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
θ

Truth level

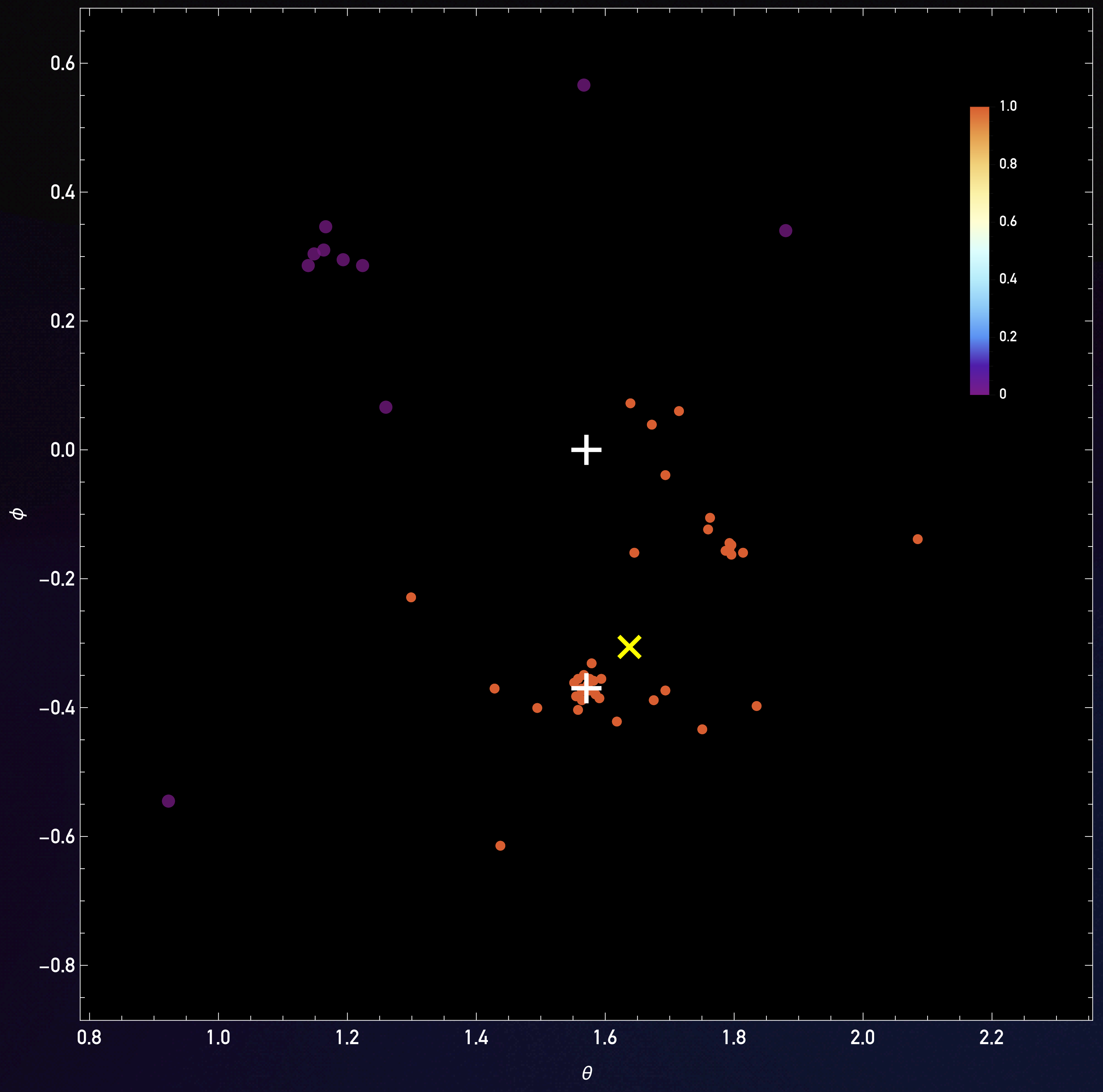

Single event
PELICAN weight
PELICAN-reconstructed W
θ
ϕ
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
1.0
0.8
0.6
0.4
0.2
0

Truth level

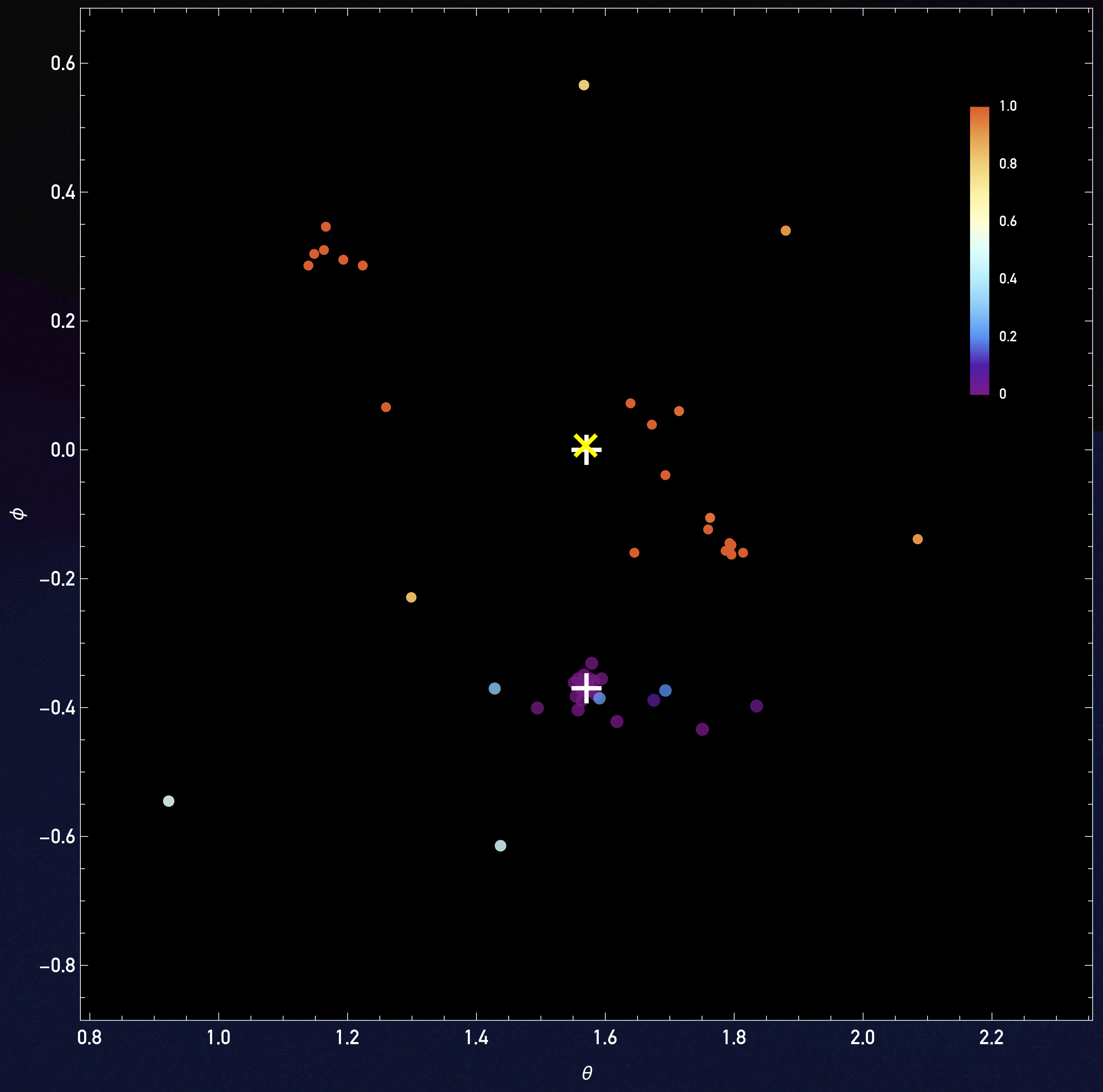
Truth



JH

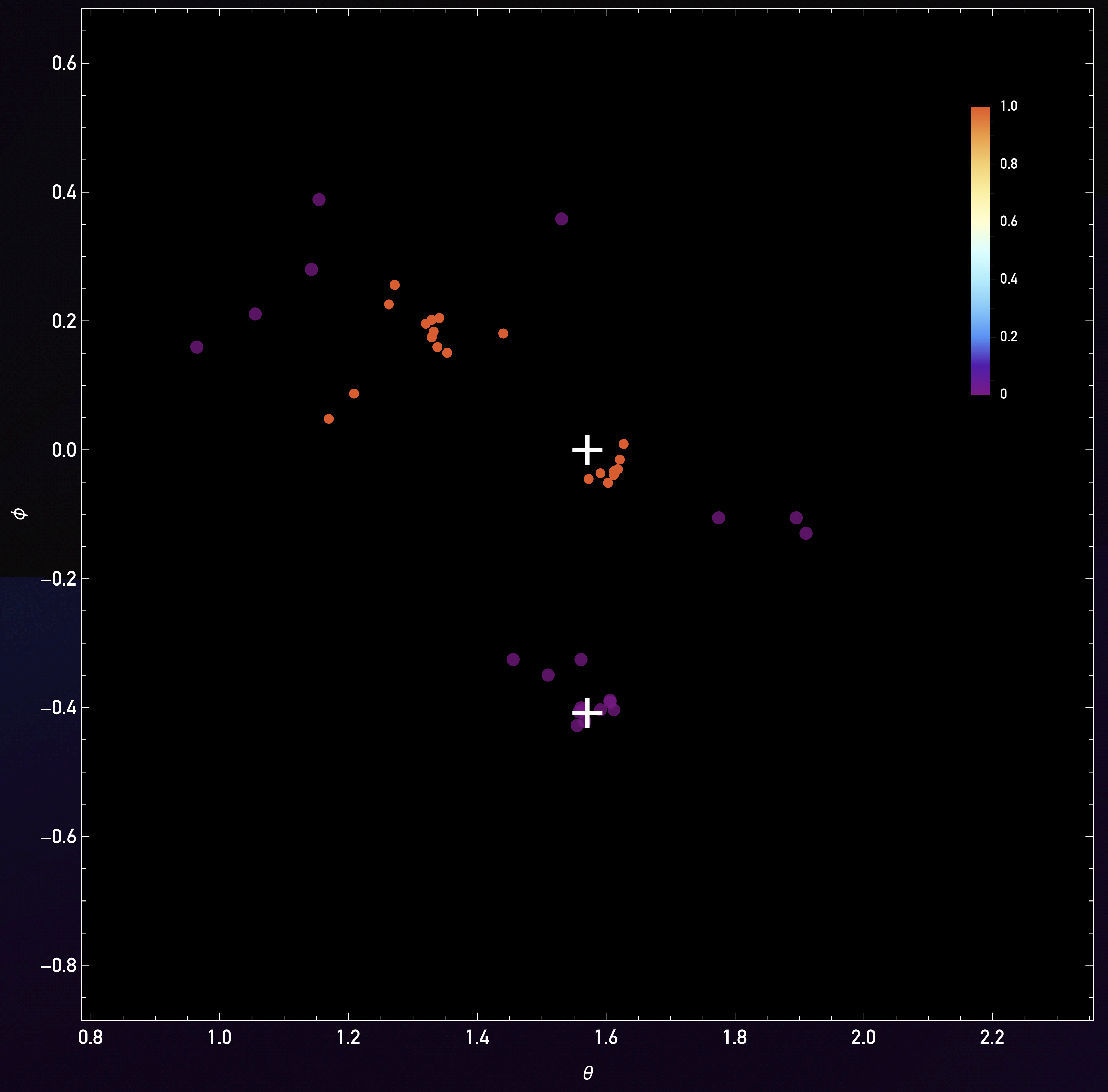


PELICAN weight

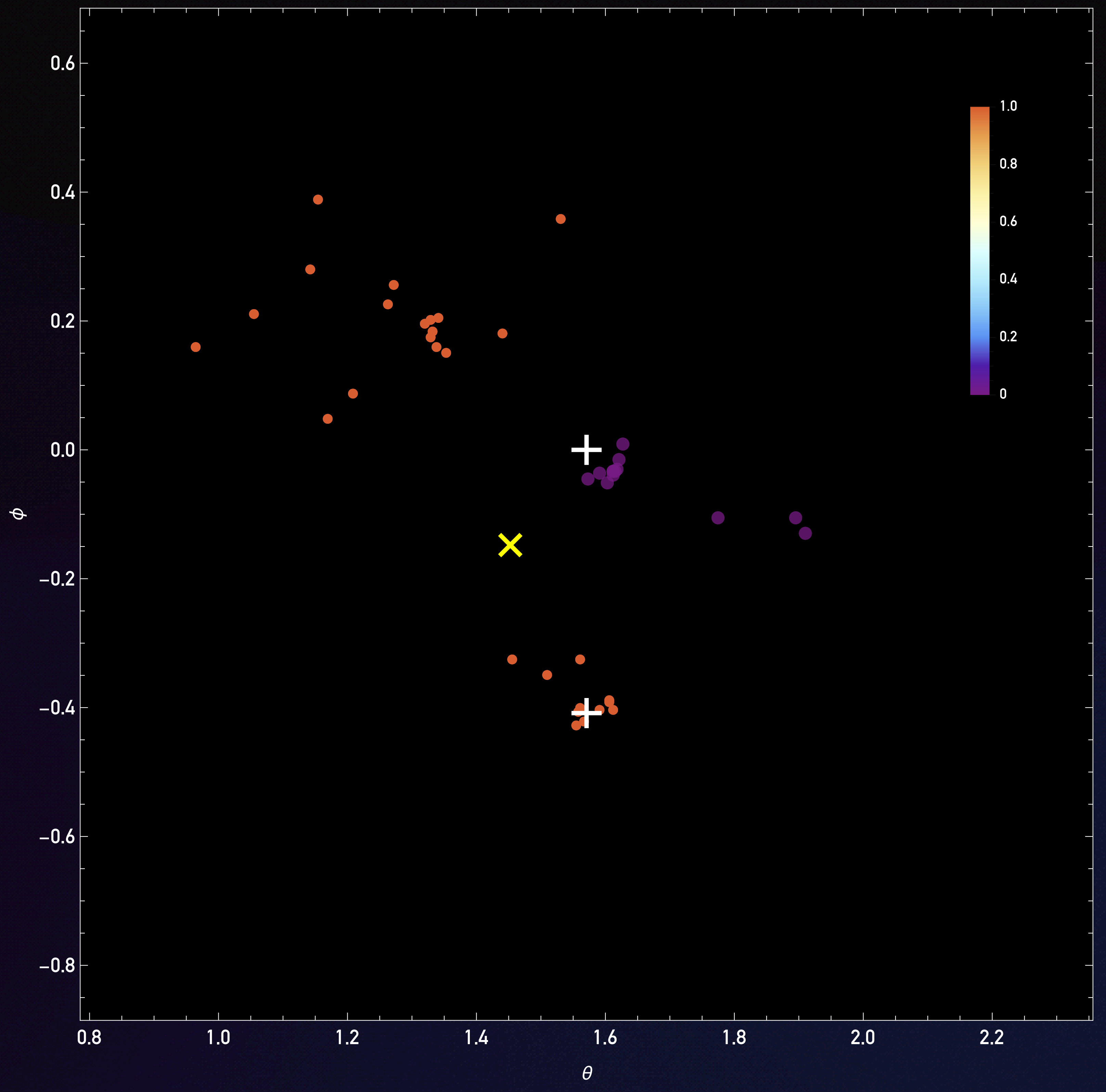


Truth level

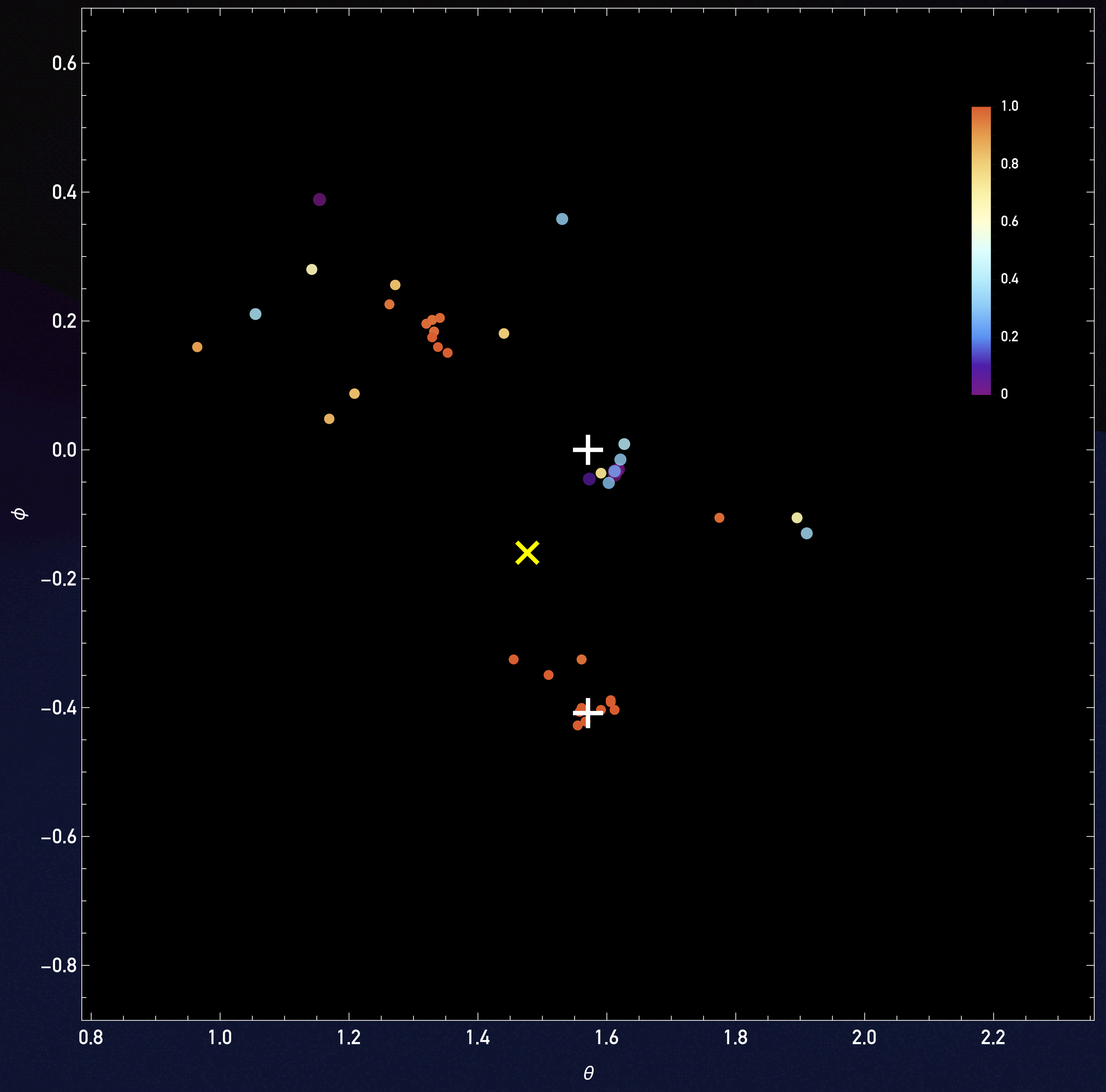
Truth



JH

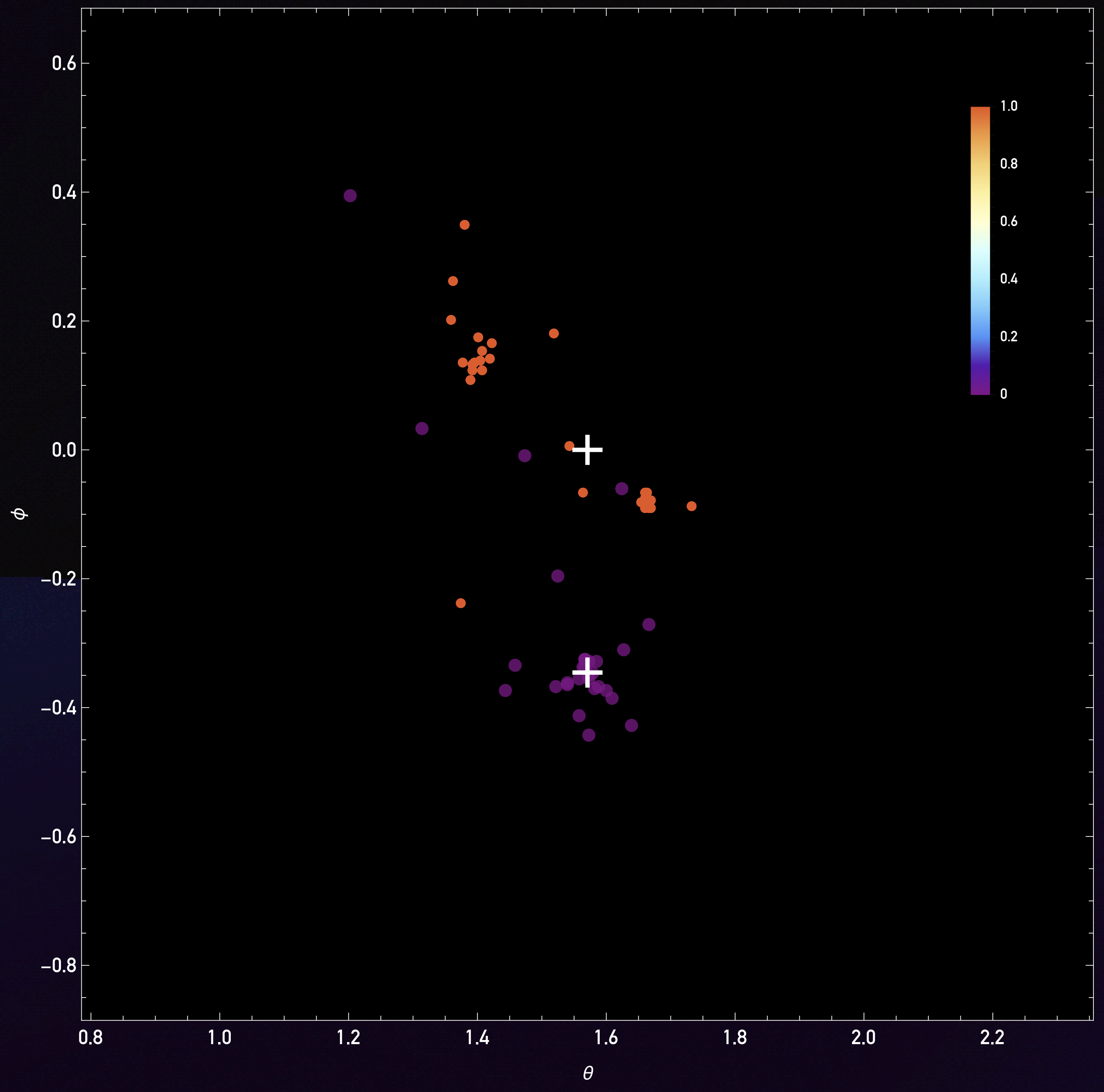


PELICAN weight

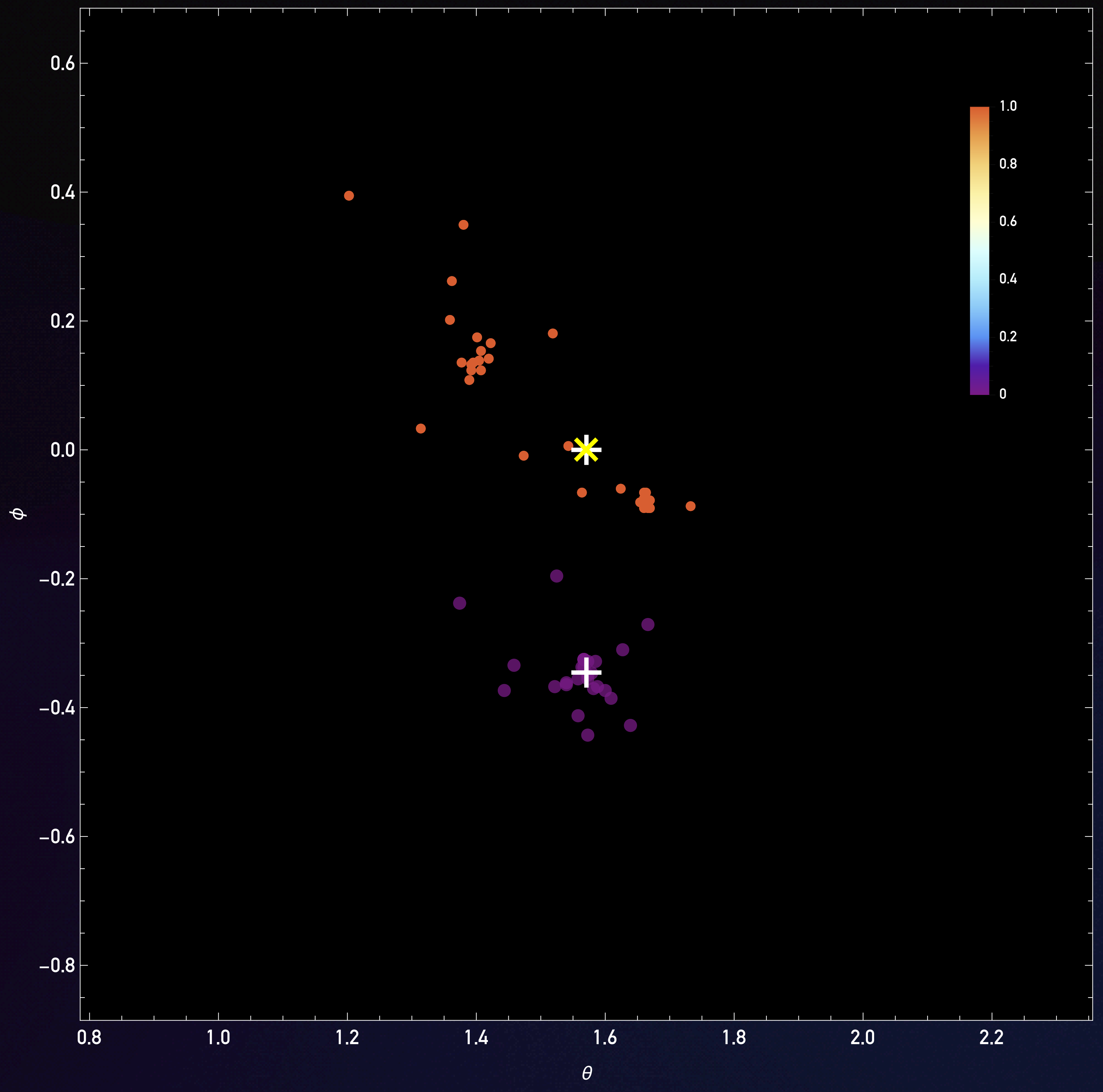


Truth level

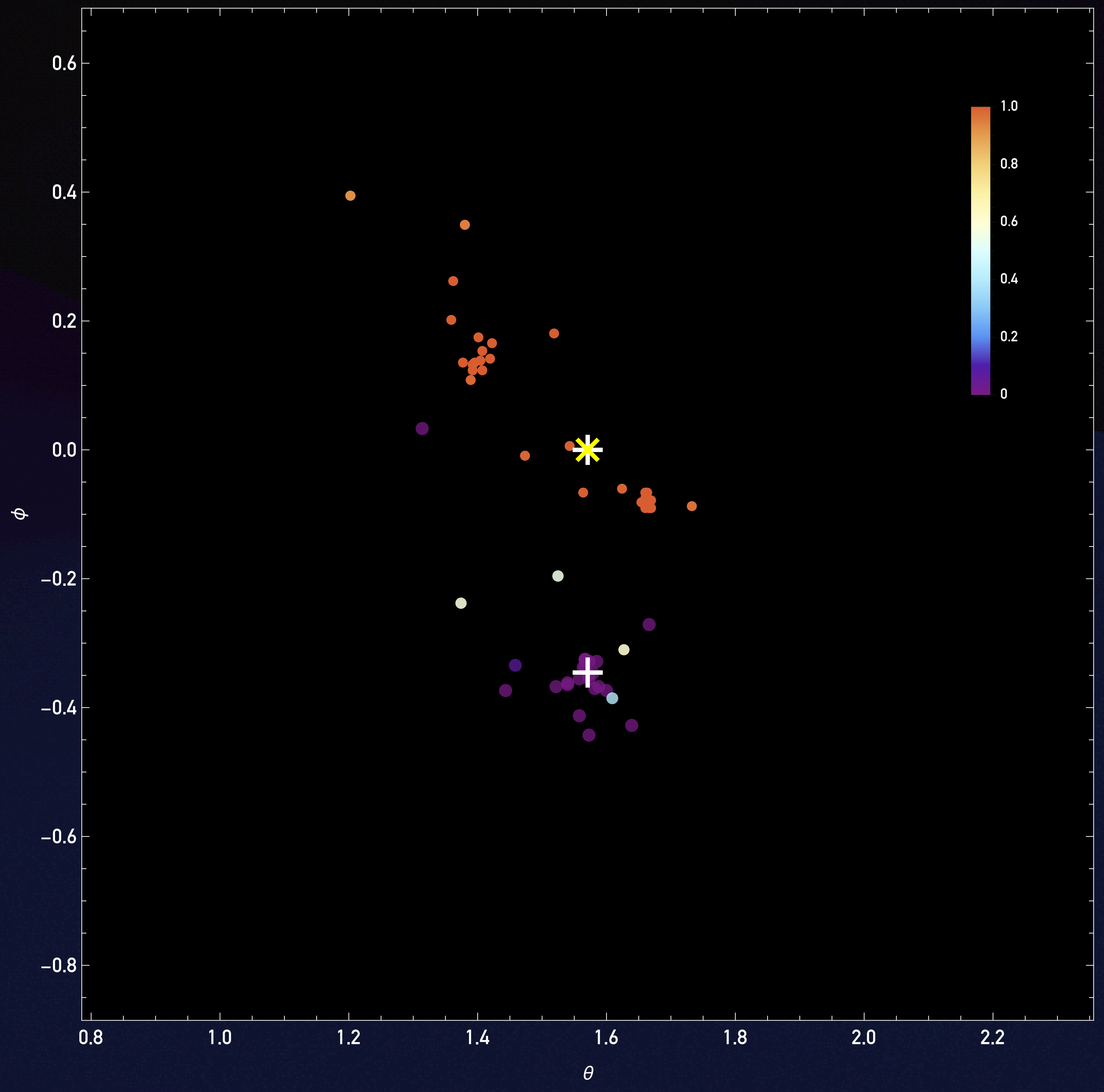
Truth



JH



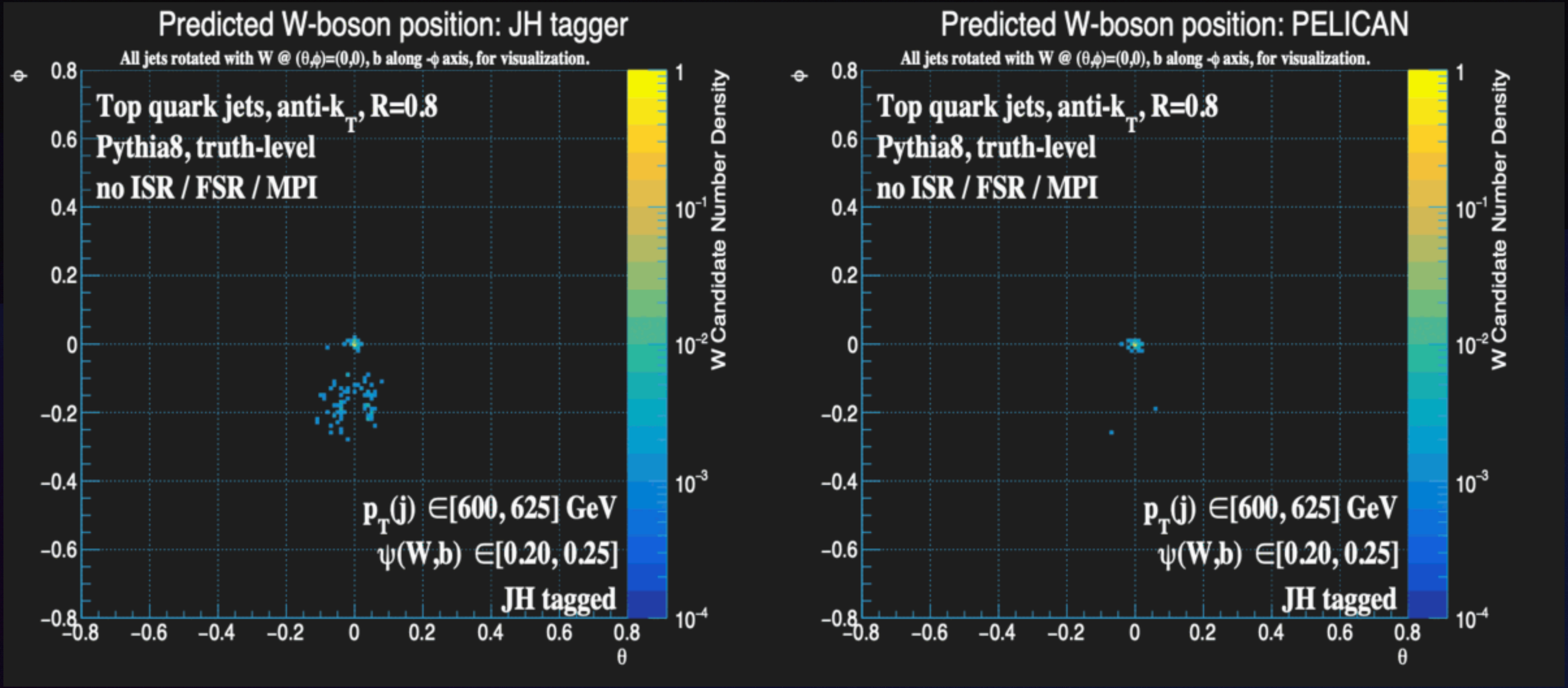
PELICAN weight



Truth level

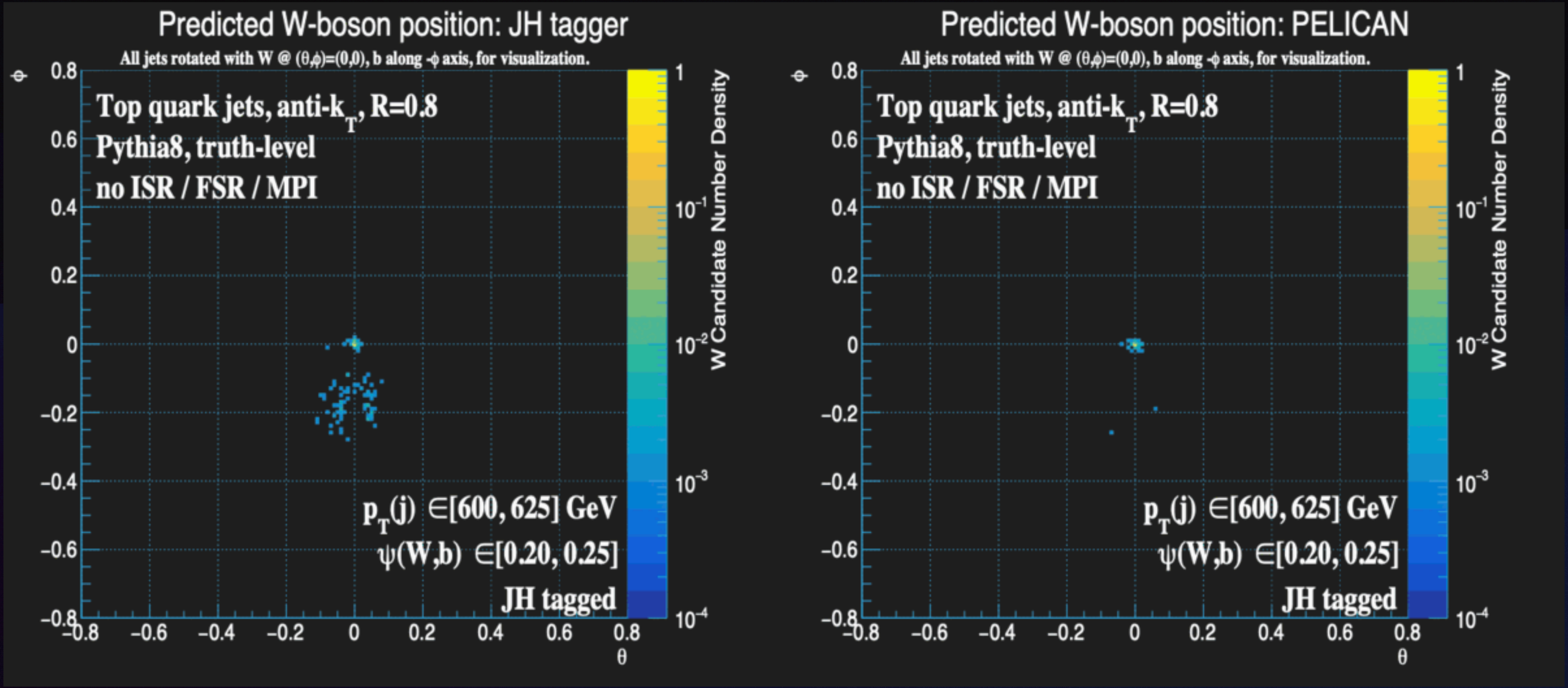
And these are only events successfully tagged by JH (~37%)!

Truth level



PELICAN is consistently better at identifying the W clusters

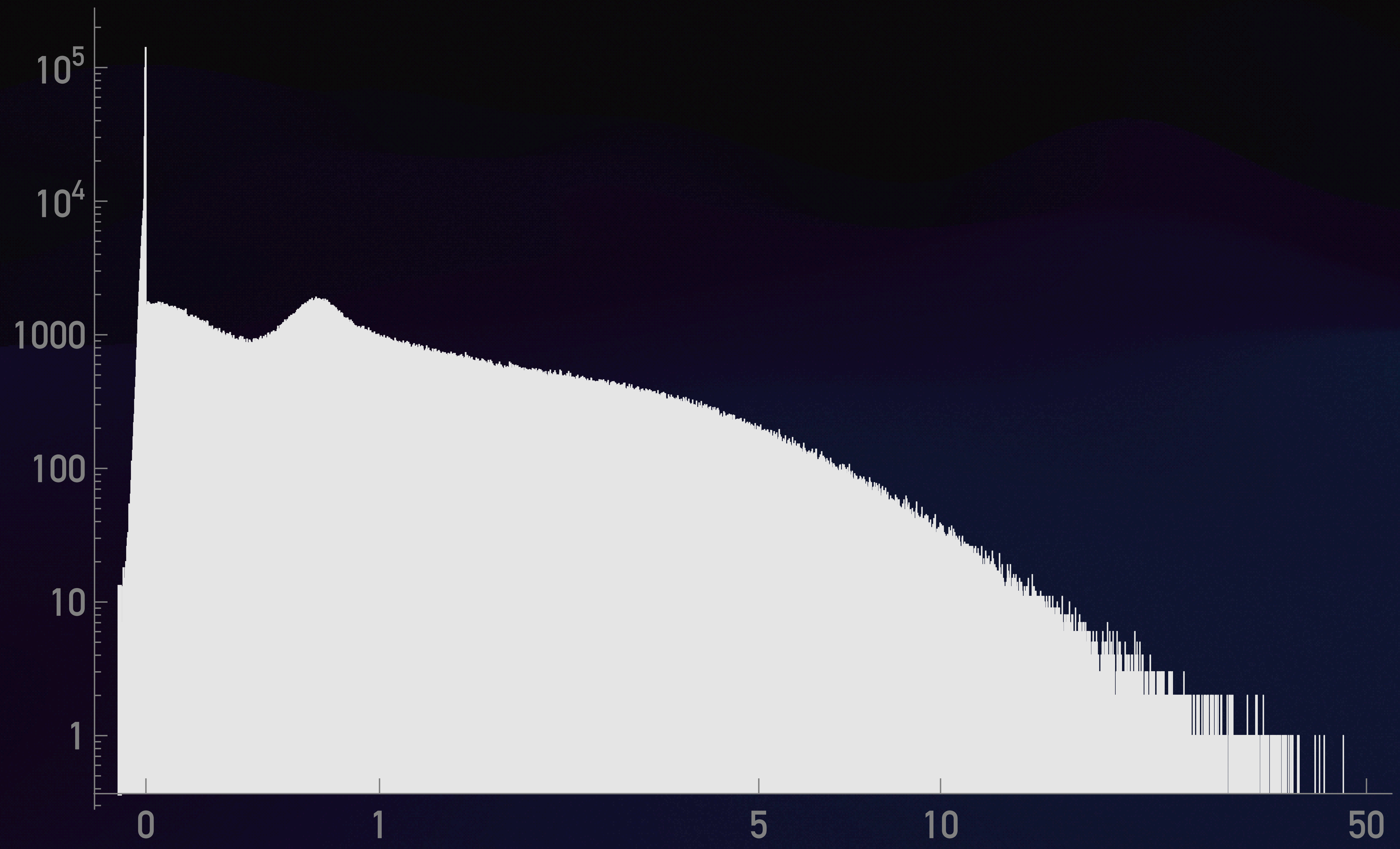
Truth level



PELICAN is consistently better at identifying the W clusters

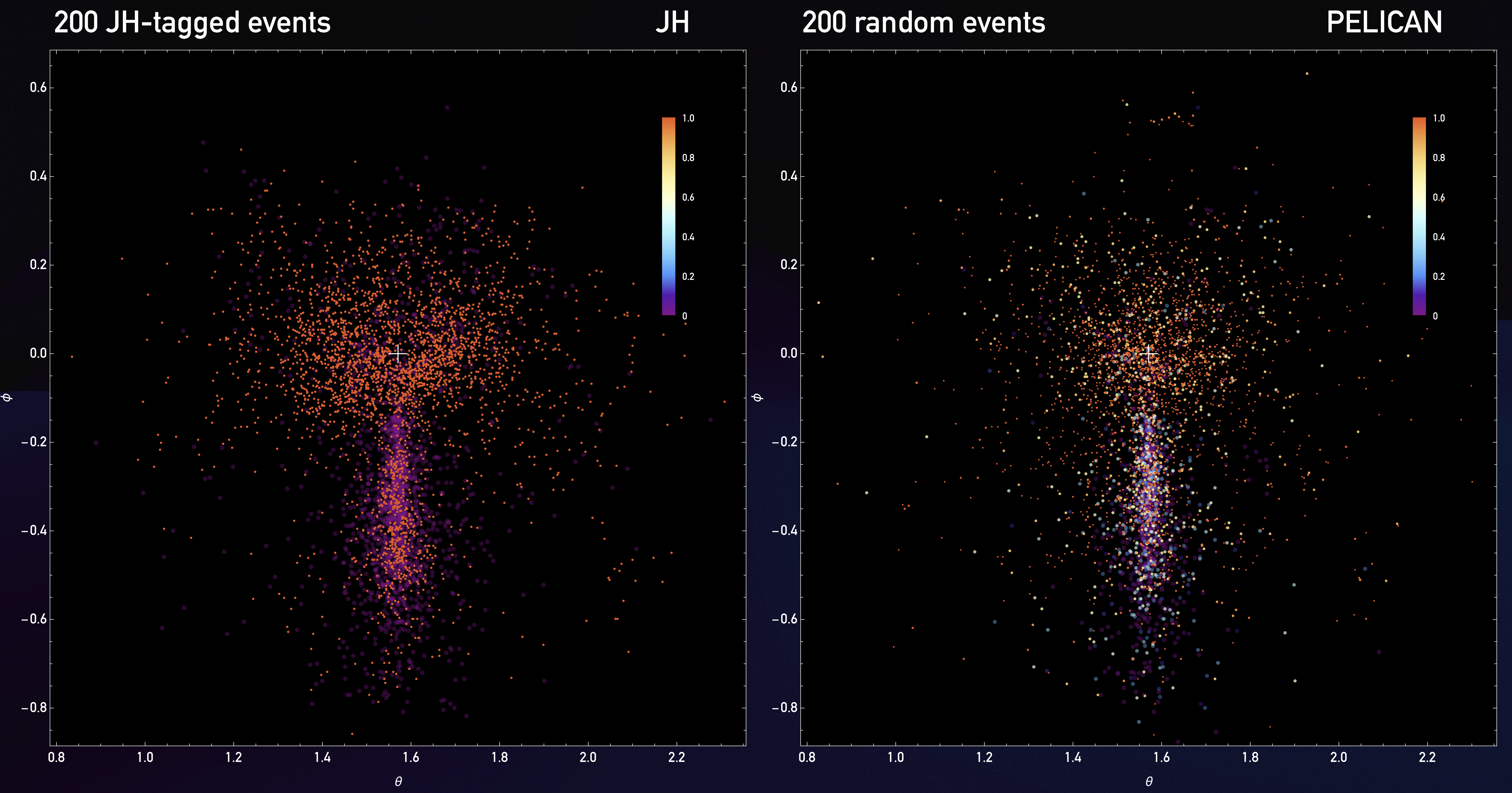
DELPHES dataset

Distribution of output PELICAN weights on DELPHES dataset
(~100k events)

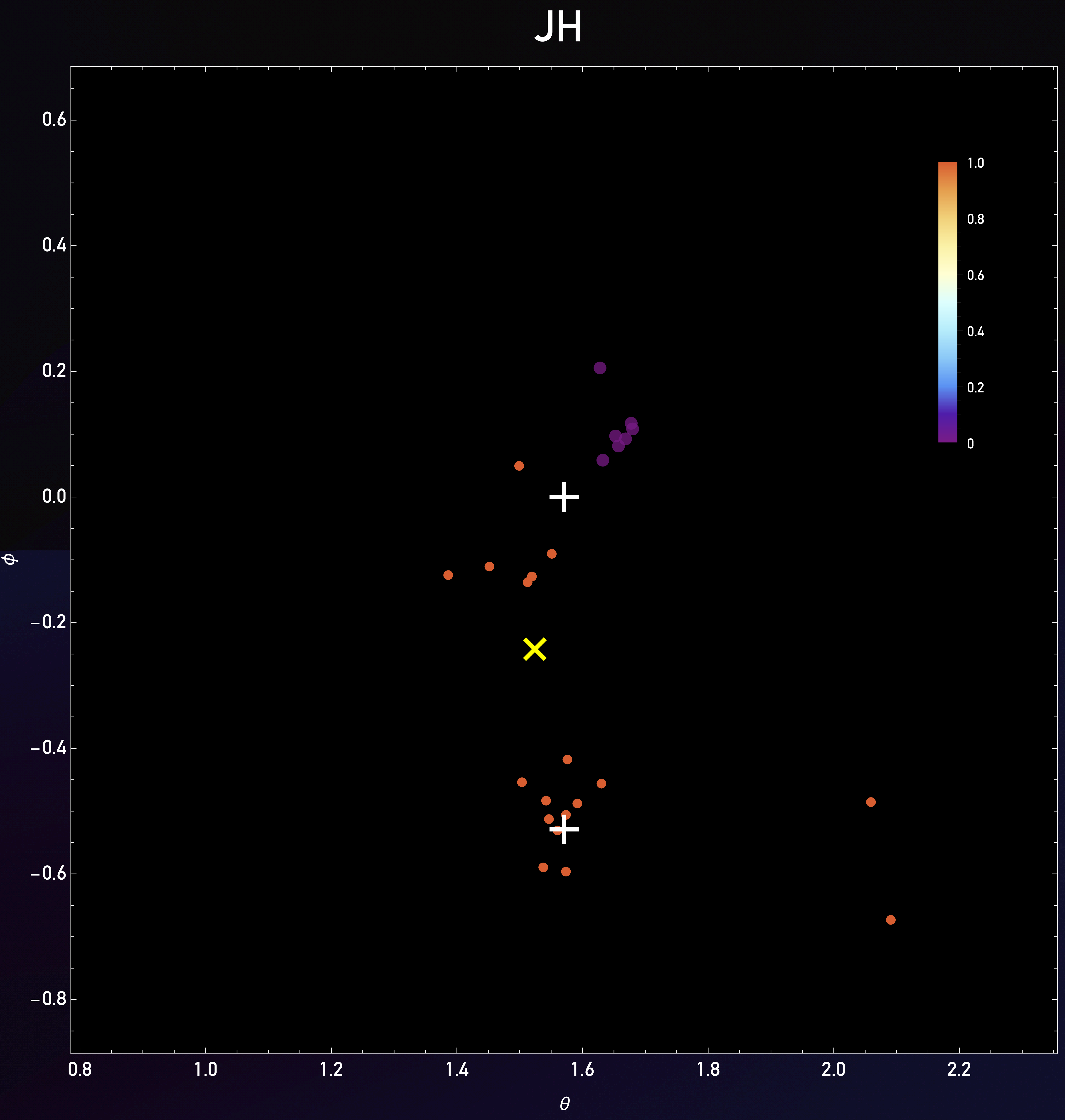


$$\{c_i\}_{i=1}^{N}$$

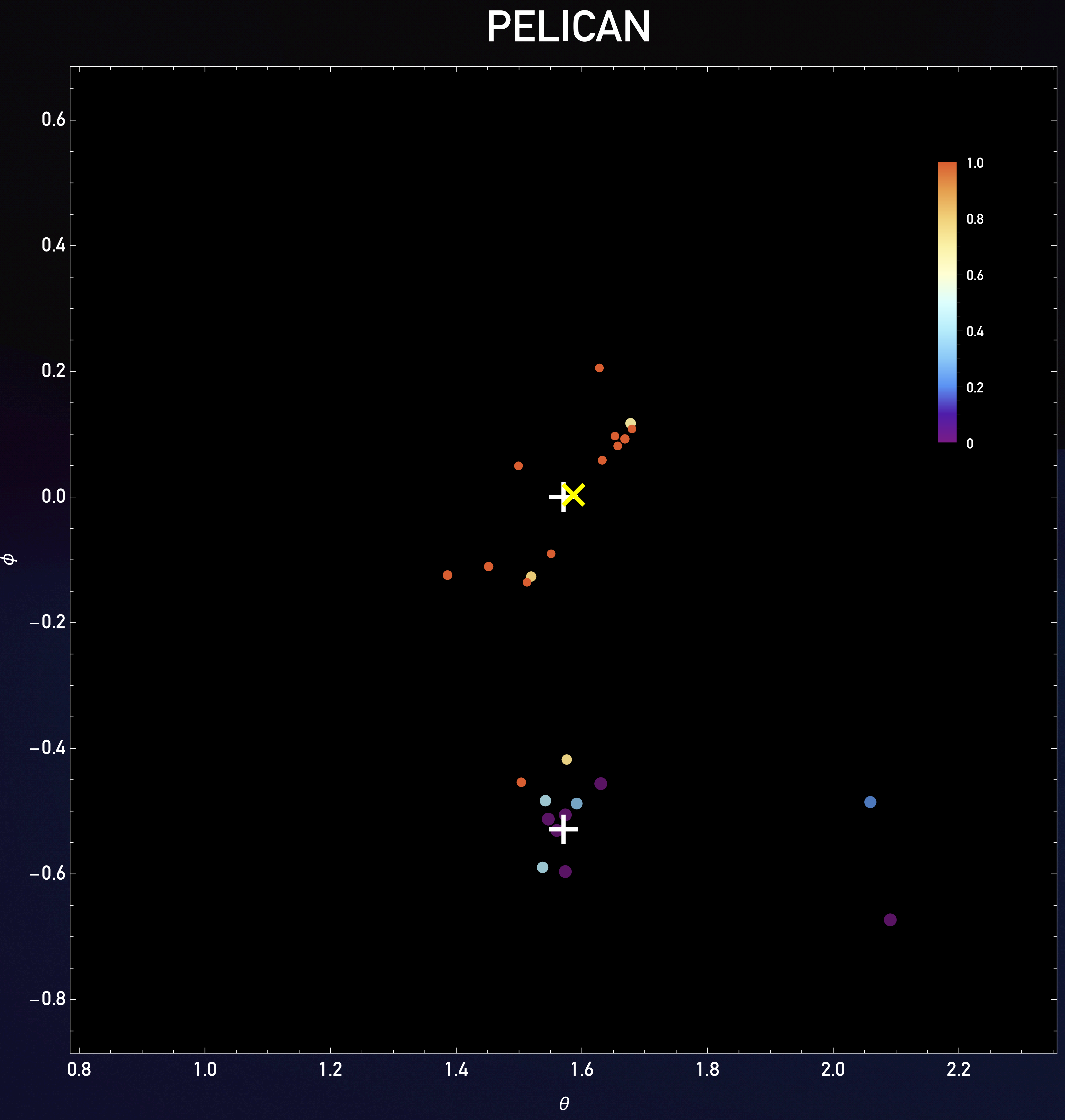
DELPHES


200 JH-tagged events
JH
200 random events
PELICAN
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
ϕ
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
θ
1.0
0.8
0.6
0.4
0.2
0

DELPHES



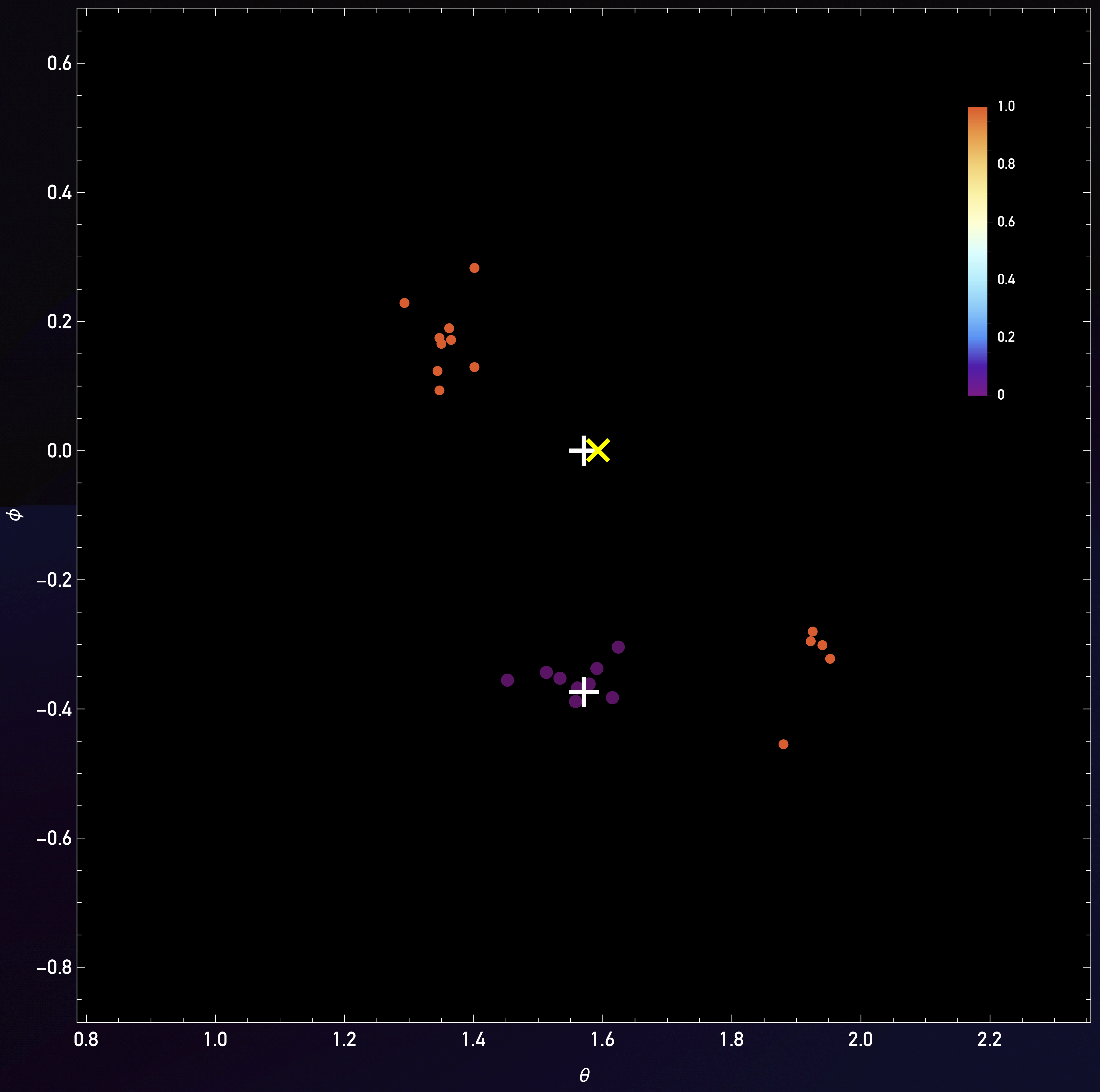
JH
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
ϕ
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
θ
1.0
0.8
0.6
0.4
0.2
0



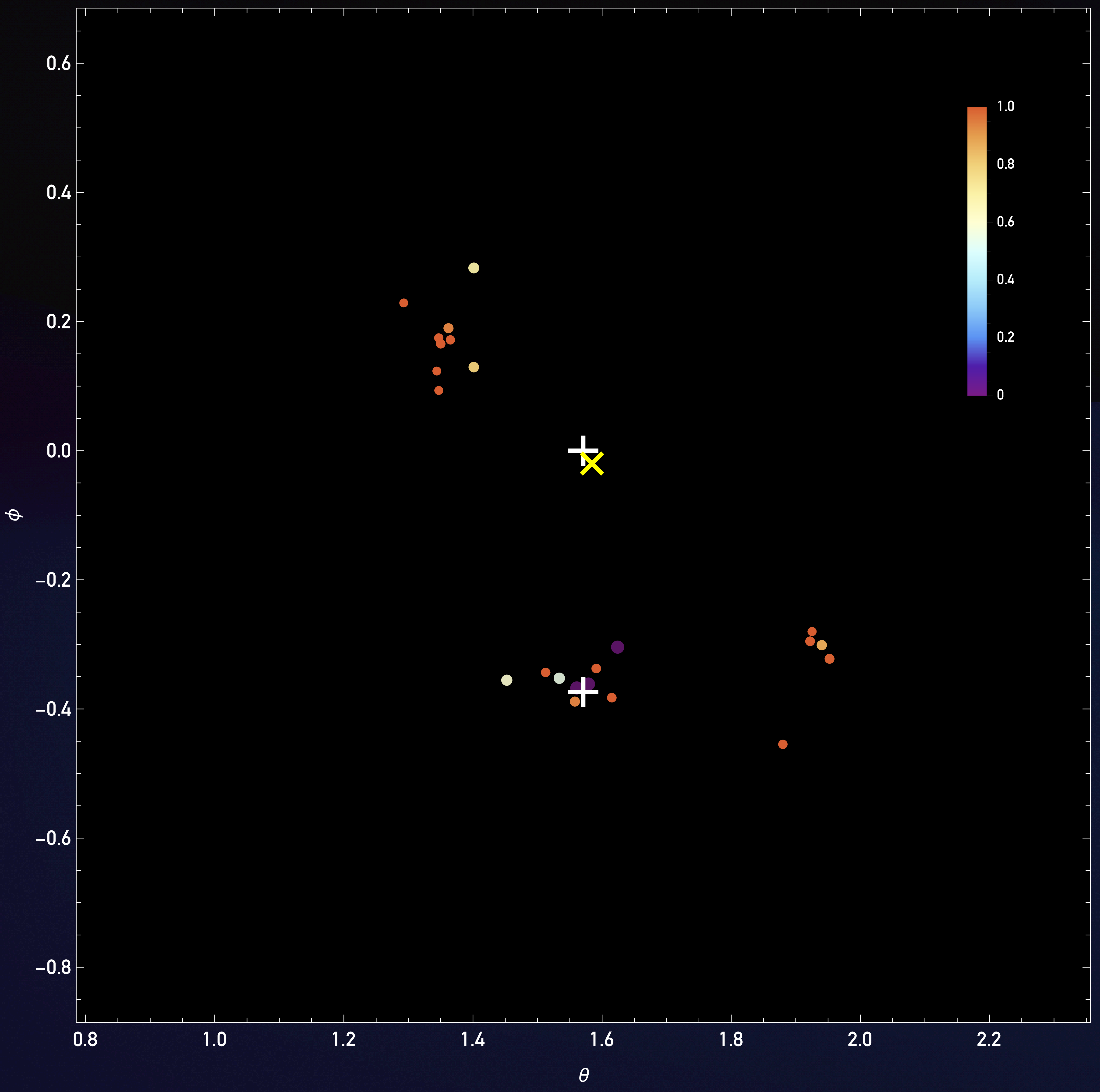
PELICAN
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
ϕ
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
θ
1.0
0.8
0.6
0.4
0.2
0

DELPHES

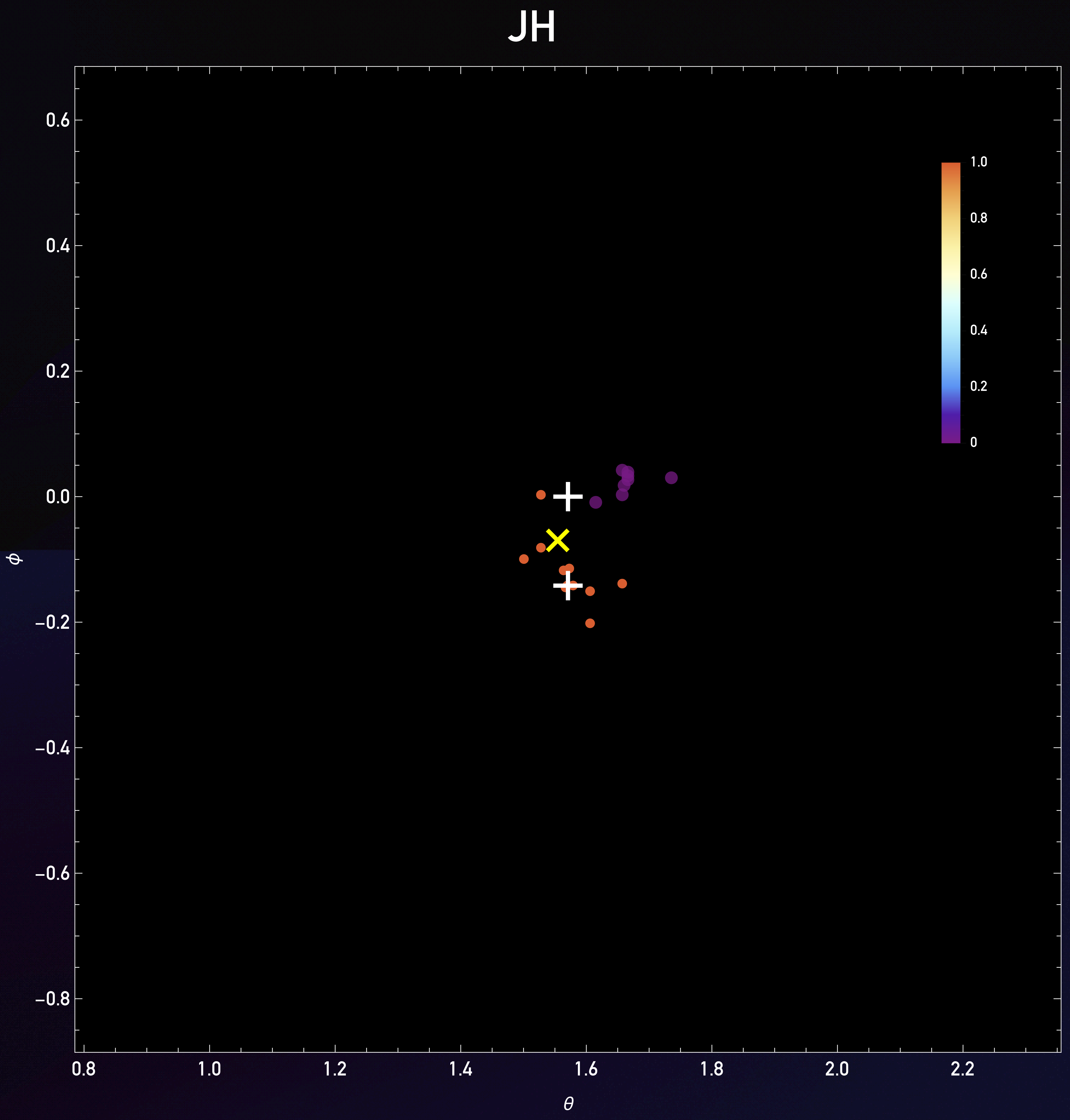
JH

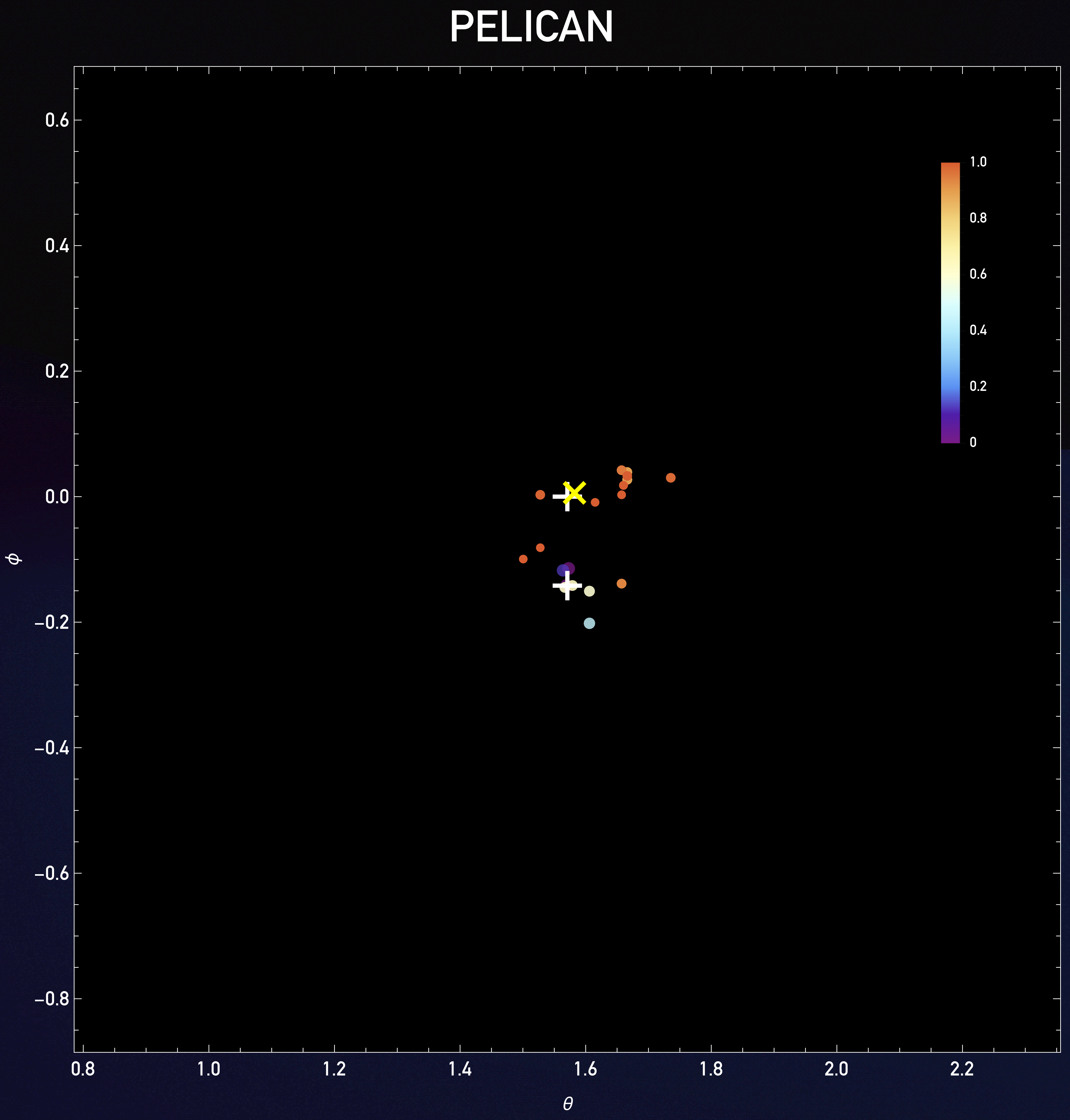


PELICAN

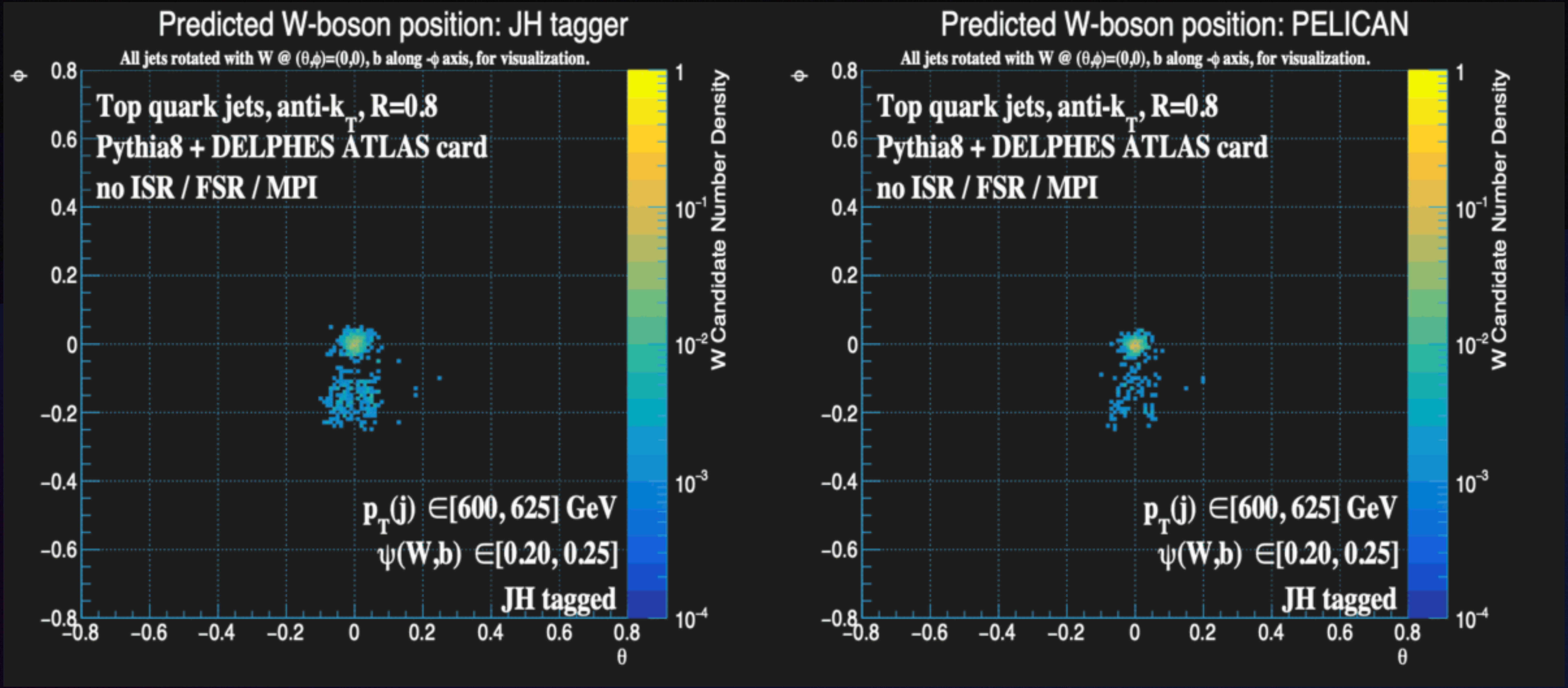


DELPHES


JH
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
θ
ϕ
1.0
0.8
0.6
0.4
0.2
0

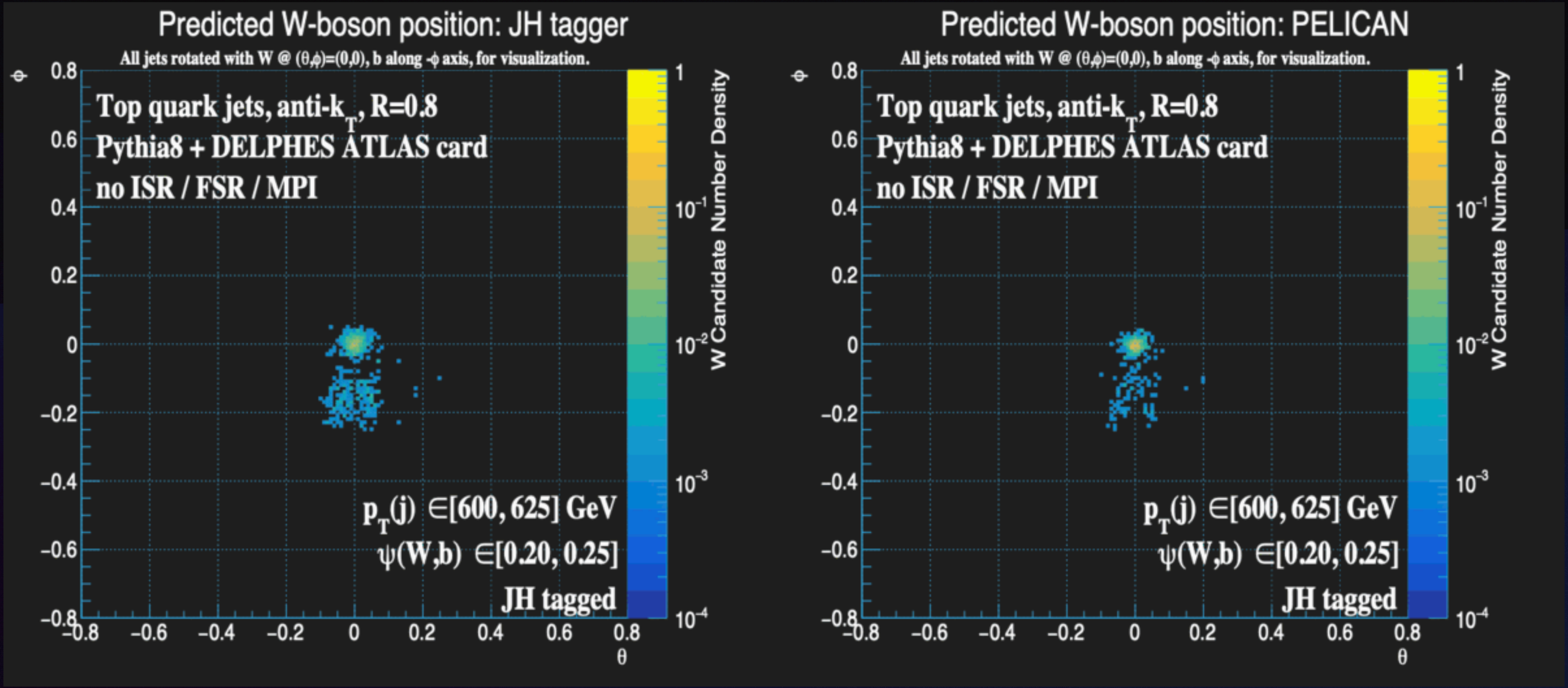

PELICAN
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
θ
ϕ
1.0
0.8
0.6
0.4
0.2
0

DELPHES



PELICAN is consistently better at identifying the W clusters

DELPHES



PELICAN is consistently better at identifying the W clusters

Explaining PELICAN

Explaining PELICAN

- What is the function of non-binary weights in the DELPHES case?

Explaining PELICAN

- What is the function of non-binary weights in the DELPHES case?
- Hypothesis:

Explaining PELICAN

- What is the function of non-binary weights in the DELPHES case?
- Hypothesis:
  - to undo the effect of multiple decay products hitting the same calorimeter cell

Explaining PELICAN

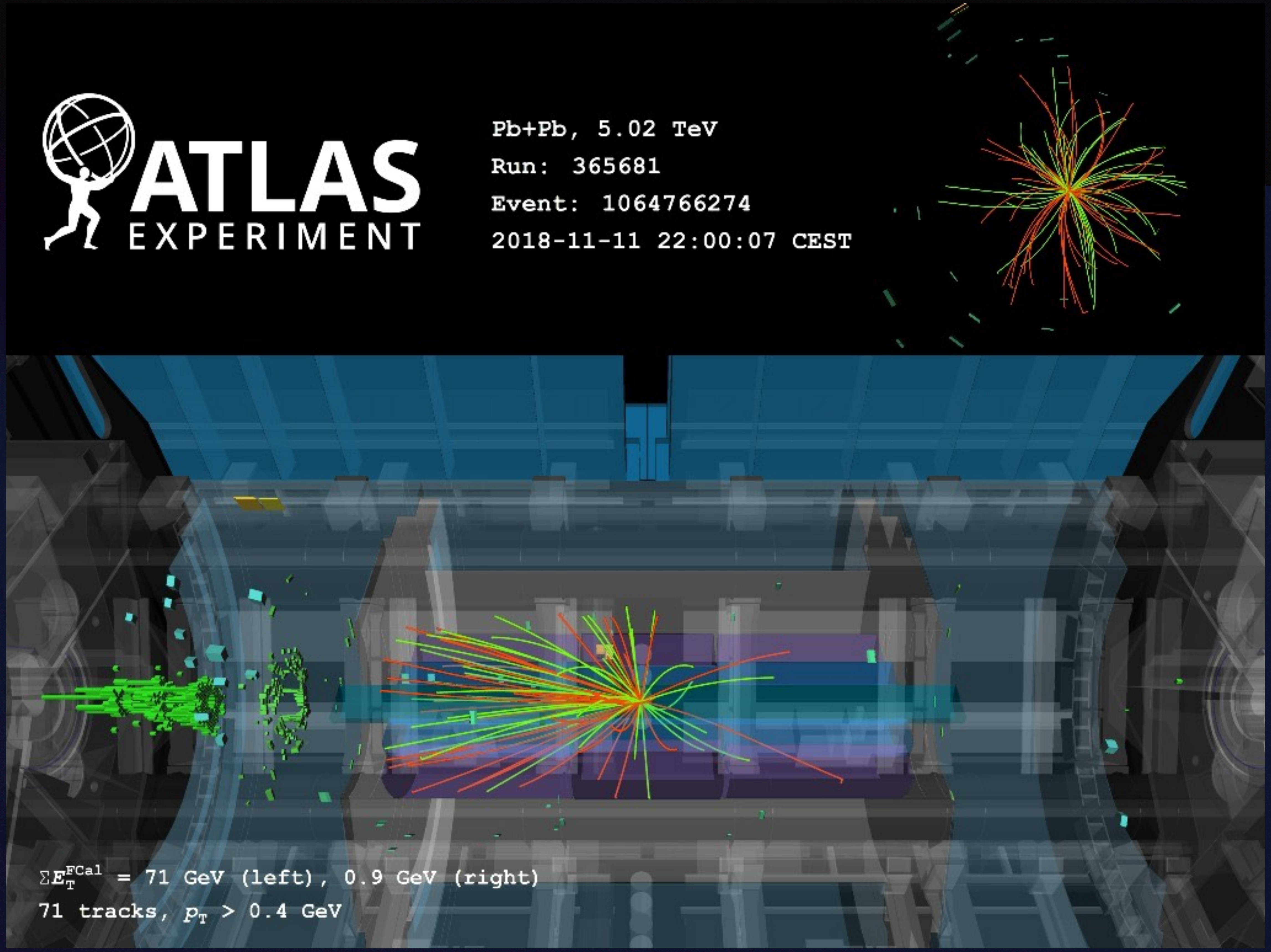
- What is the function of non-binary weights in the DELPHES case?
- Hypothesis:
  - to undo the effect of multiple decay products hitting the same calorimeter cell
  - to correct the mass of the reconstruction

Explaining PELICAN

- What is the function of non-binary weights in the DELPHES case?
- Hypothesis:
  - to undo the effect of multiple decay products hitting the same calorimeter cell
  - to correct the mass of the reconstruction
- Define $r_i = E_i^{W_{\text{true}}}/E_i$. Note that $E^{W_{\text{true}}} = \sum_i r_i E_i$.

  Therefore expect high correlation $c_i \sim r_i$

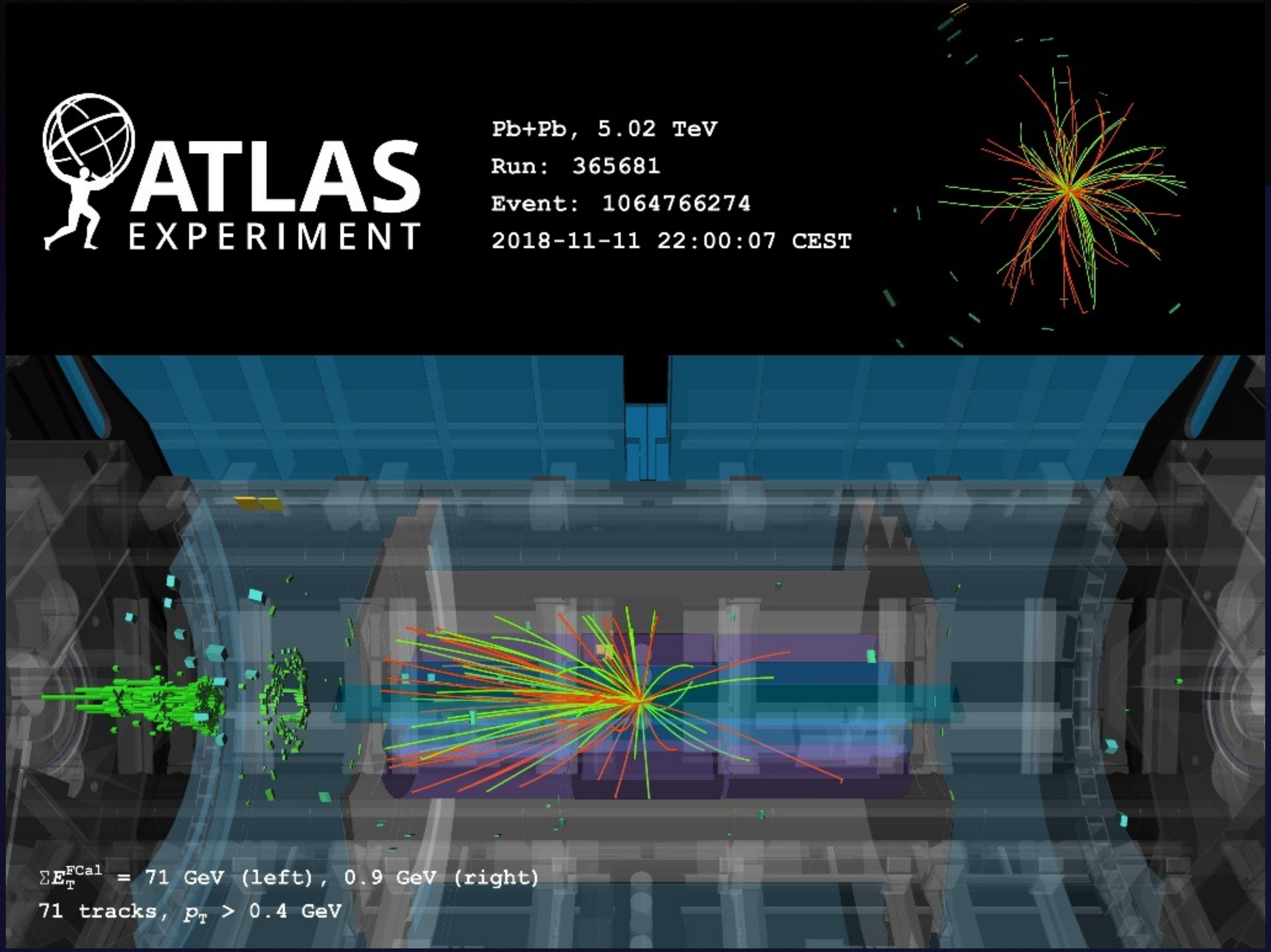
Detector breaks symmetry?



ATLAS Experiment © 2022 CERN

Detector breaks symmetry?

Don't beam axes break SO(3) symmetry?

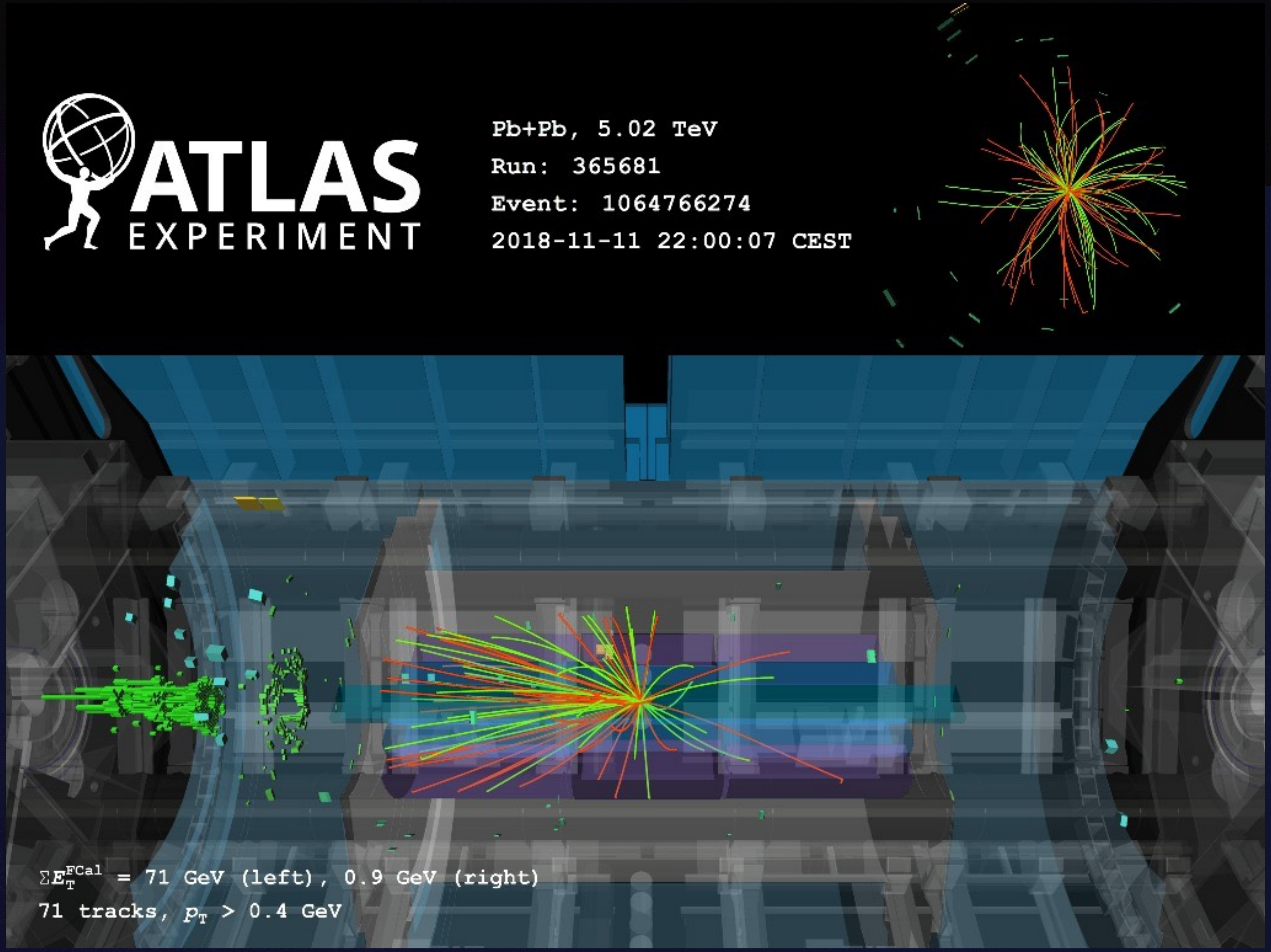


ATLAS Experiment © 2022 CERN

Detector breaks symmetry?

Don't beam axes break SO(3) symmetry?

No! Rather than giving up symmetry, provide complete inputs.



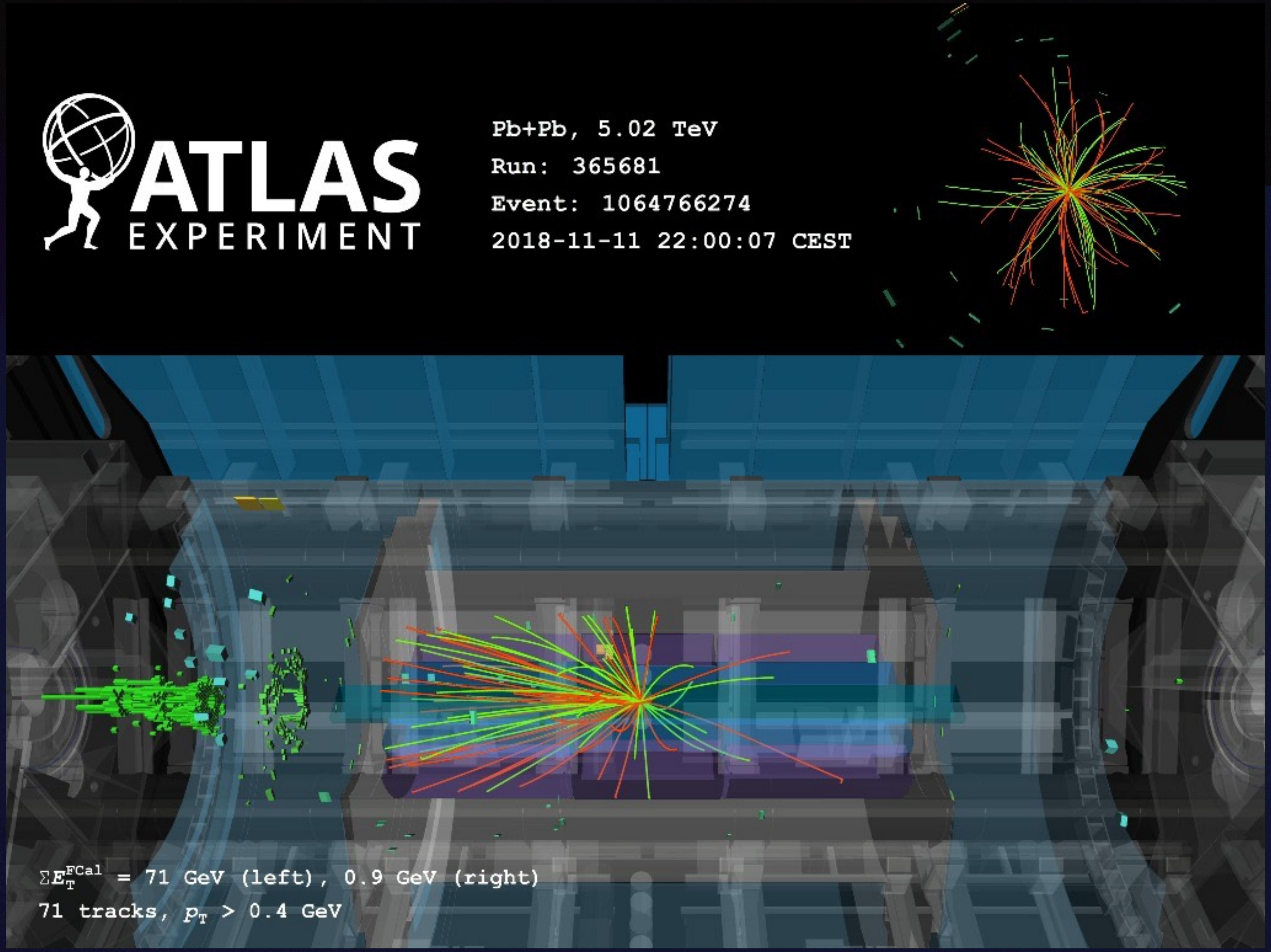
ATLAS Experiment © 2022 CERN

Detector breaks symmetry?

Don't beam axes break SO(3) symmetry?

No! Rather than giving up symmetry, provide complete inputs.

$$p_1, \ldots, p_N$$



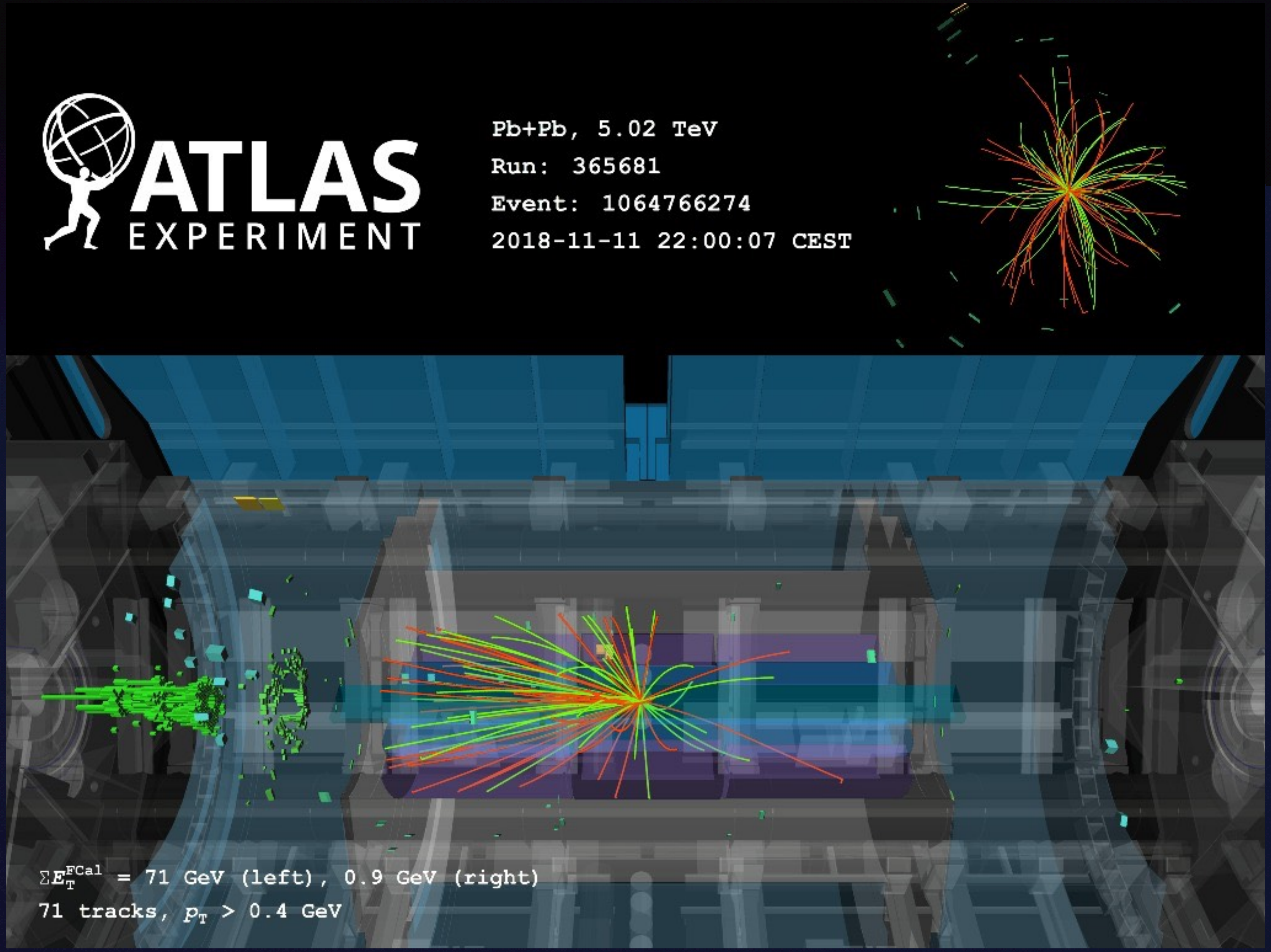
Detector breaks symmetry?

Don't beam axes break SO(3) symmetry?

No! Rather than giving up symmetry, provide complete inputs.

$$e_{z+}, e_{z-}, p_1, \ldots, p_N$$

$$e_{z+} = (1,0,0,1),\ e_{z-} = (1,0,0,-1)$$



ATLAS Experiment © 2022 CERN

Detector breaks symmetry?

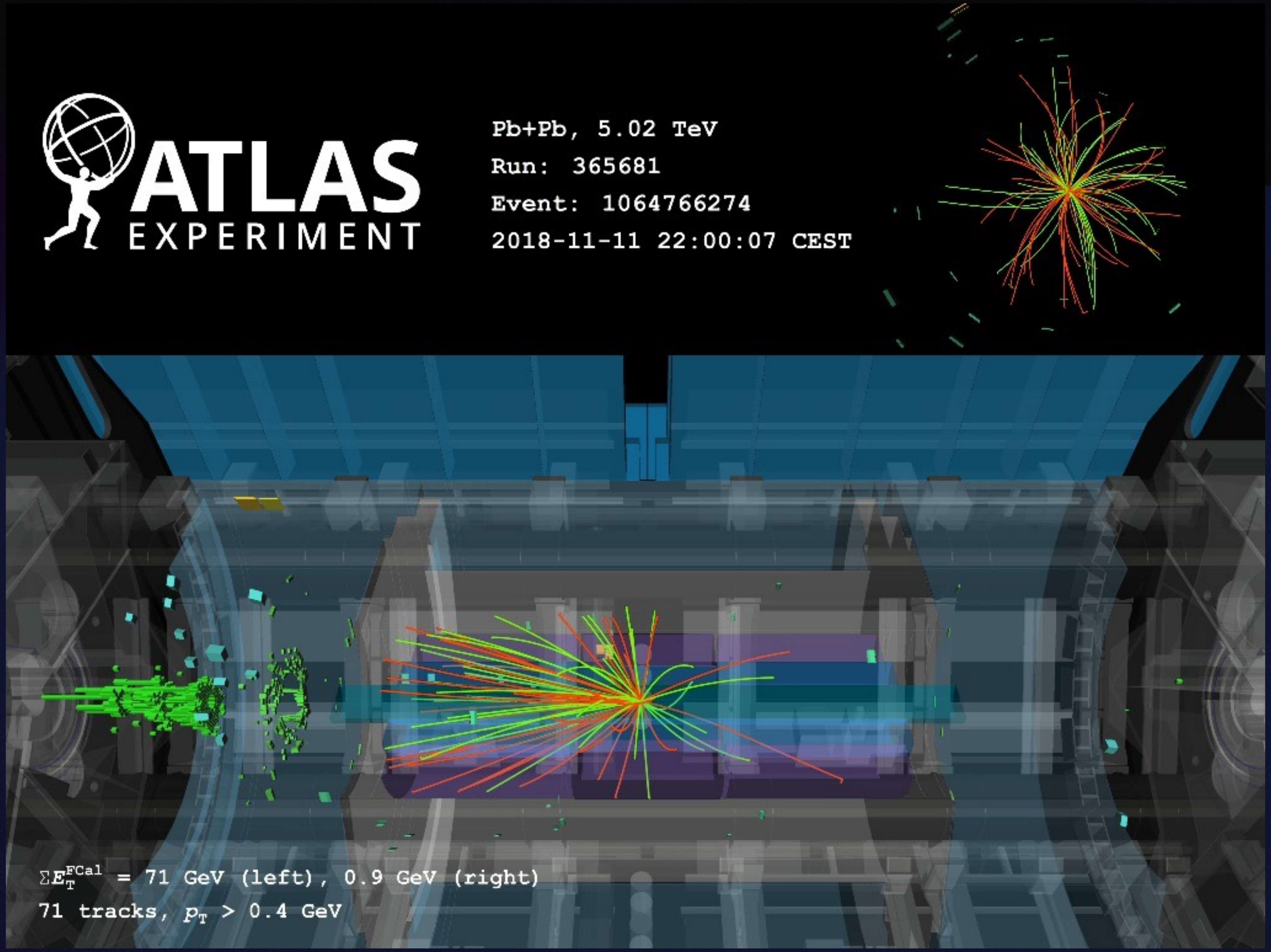
Don't beam axes break SO(3) symmetry?

No! Rather than giving up symmetry, provide complete inputs.

$$e_{z+}, e_{z-}, p_1, \ldots, p_N$$

$$e_{z+} = (1,0,0,1),\ e_{z-} = (1,0,0,-1)$$

Now PELICAN knows E_i, m_i, p_i^T, p_i^z.

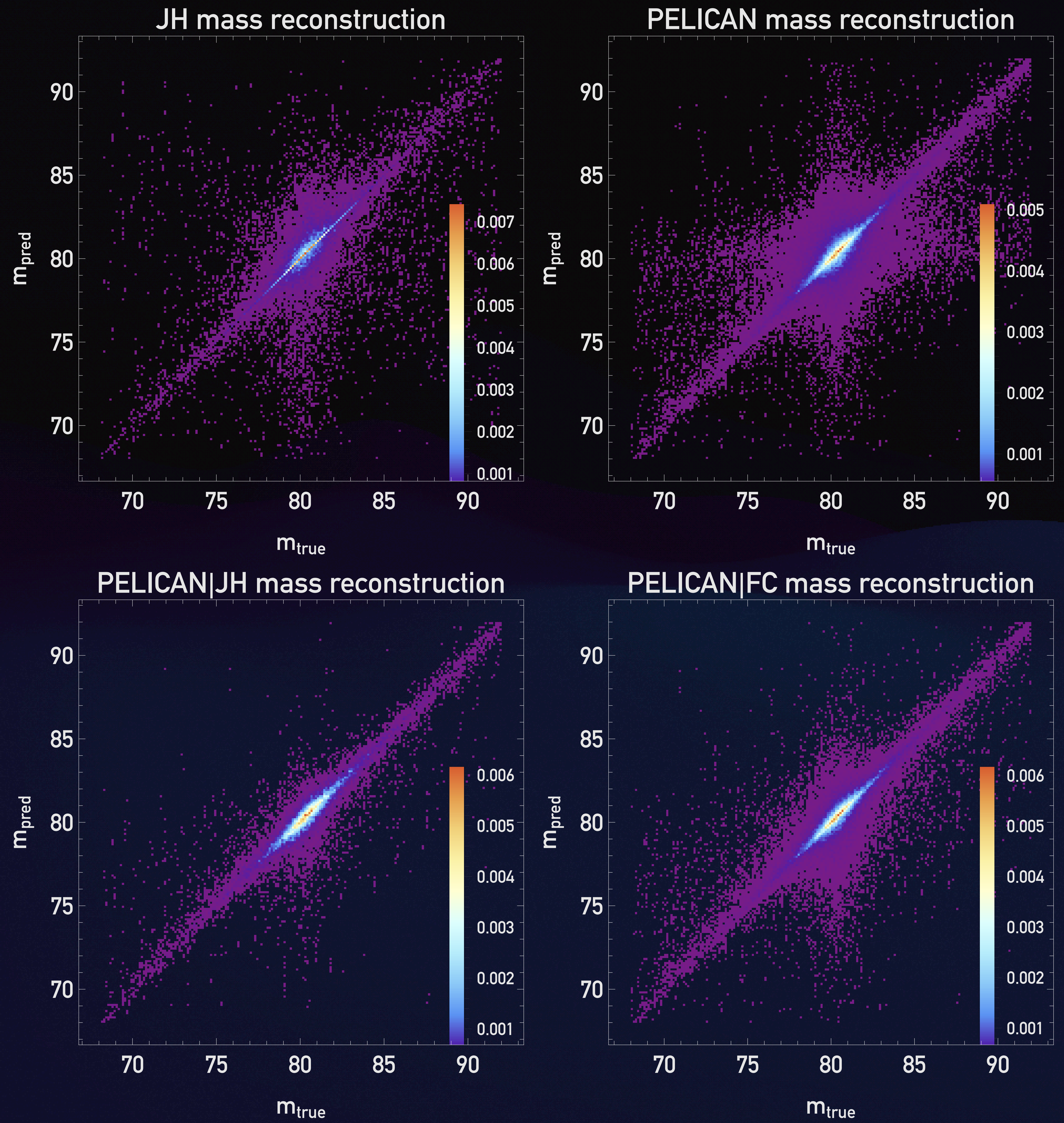


ATLAS Experiment © 2022 CERN

Mass reconstruction

W mass reconstruction

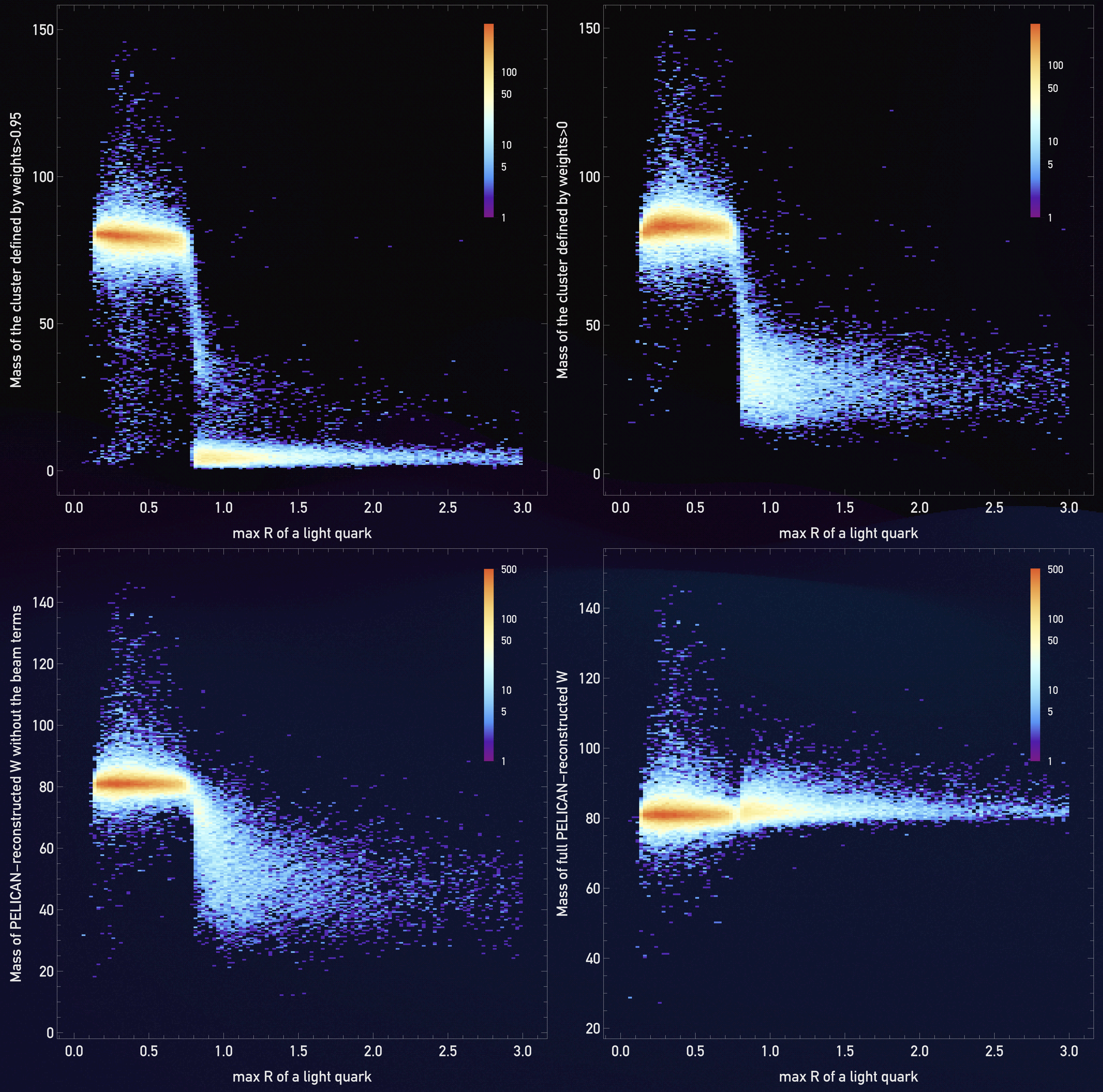
Truth level



W mass reconstruction

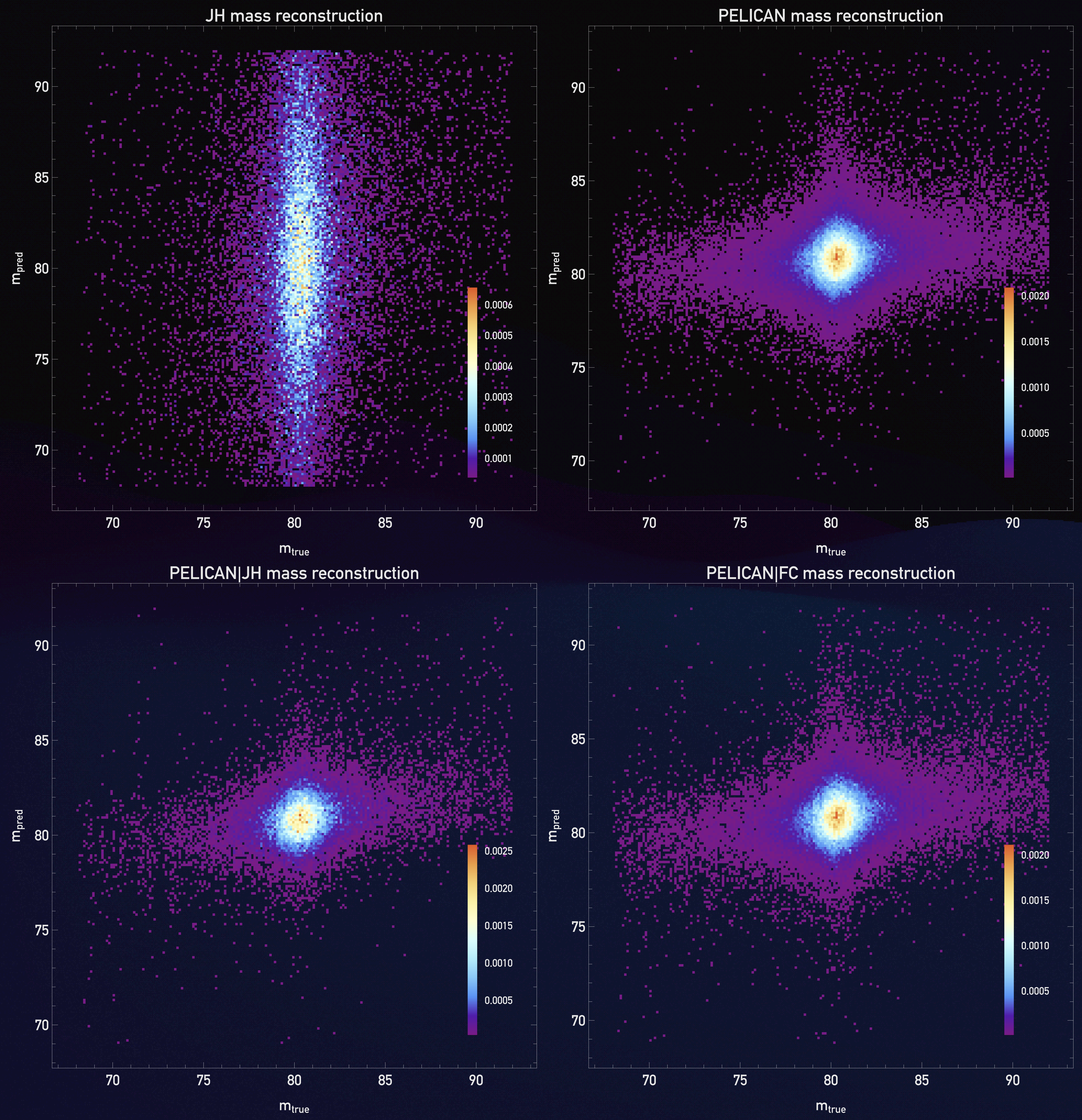
Truth level

PELICAN uses beams to adjust mass



W mass reconstruction

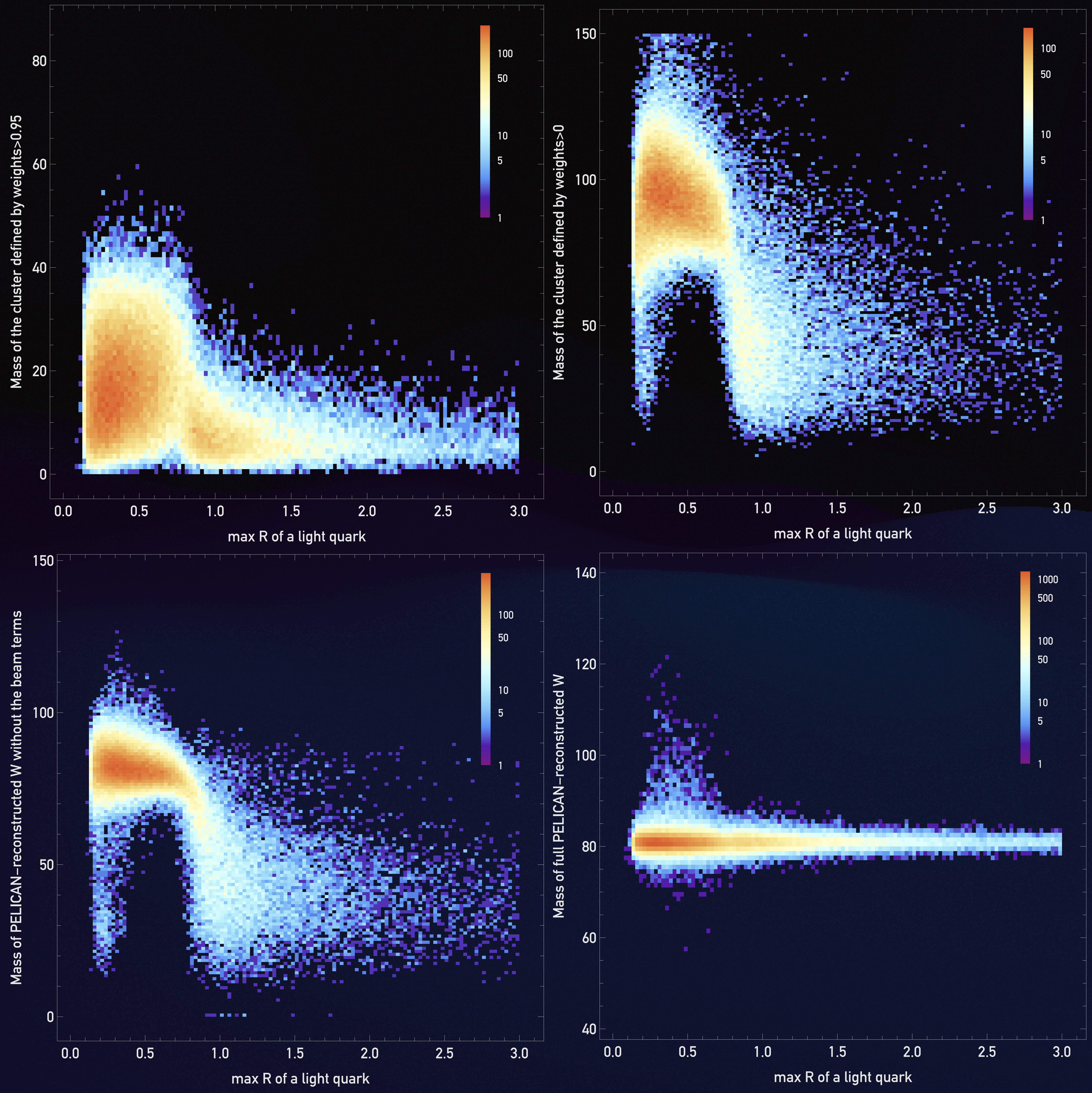
Delphes



W mass reconstruction

Delphes

PELICAN uses beams to adjust mass

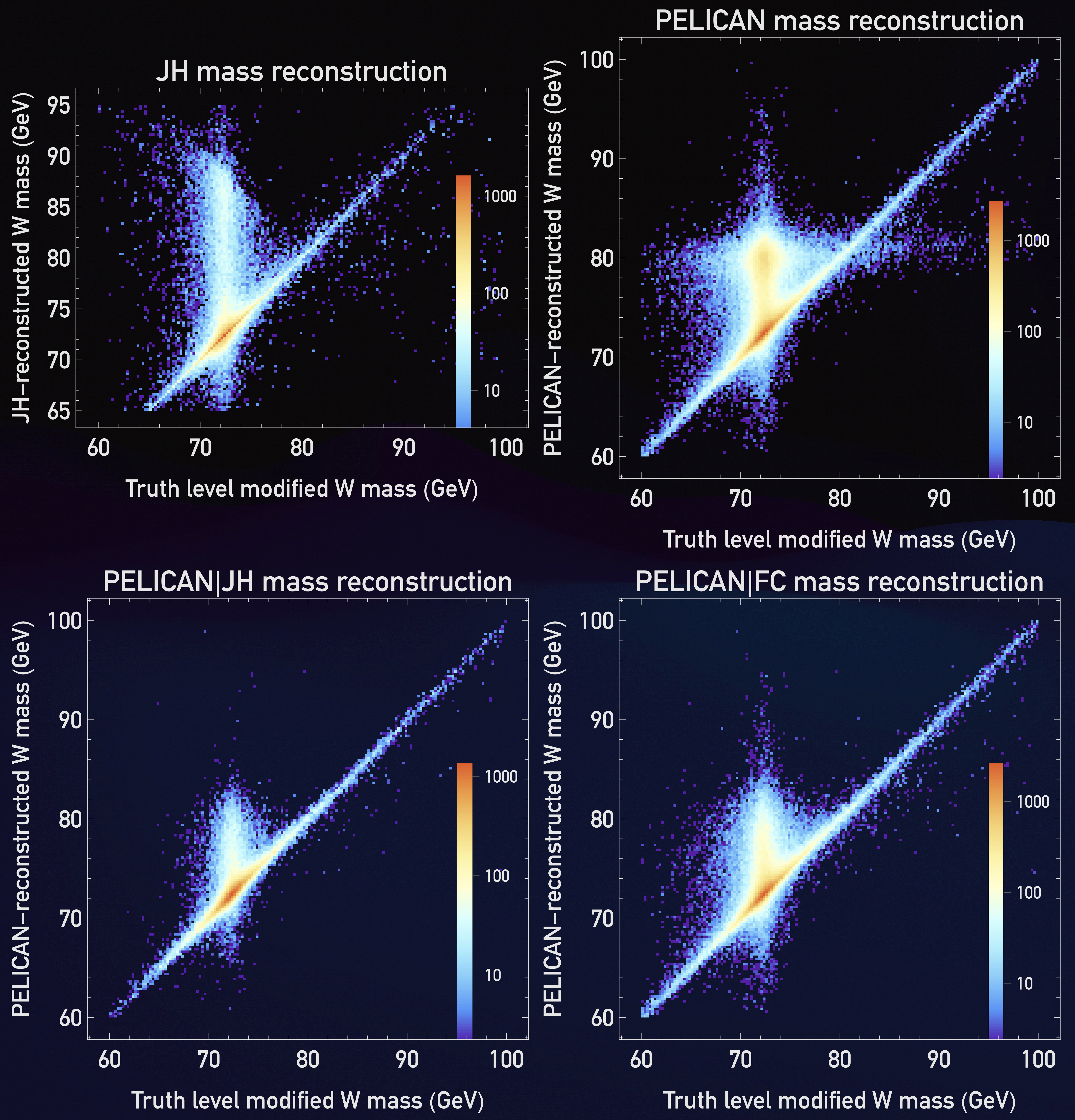


Mass regression

Transfer Learning: W mass measurement

Truth level

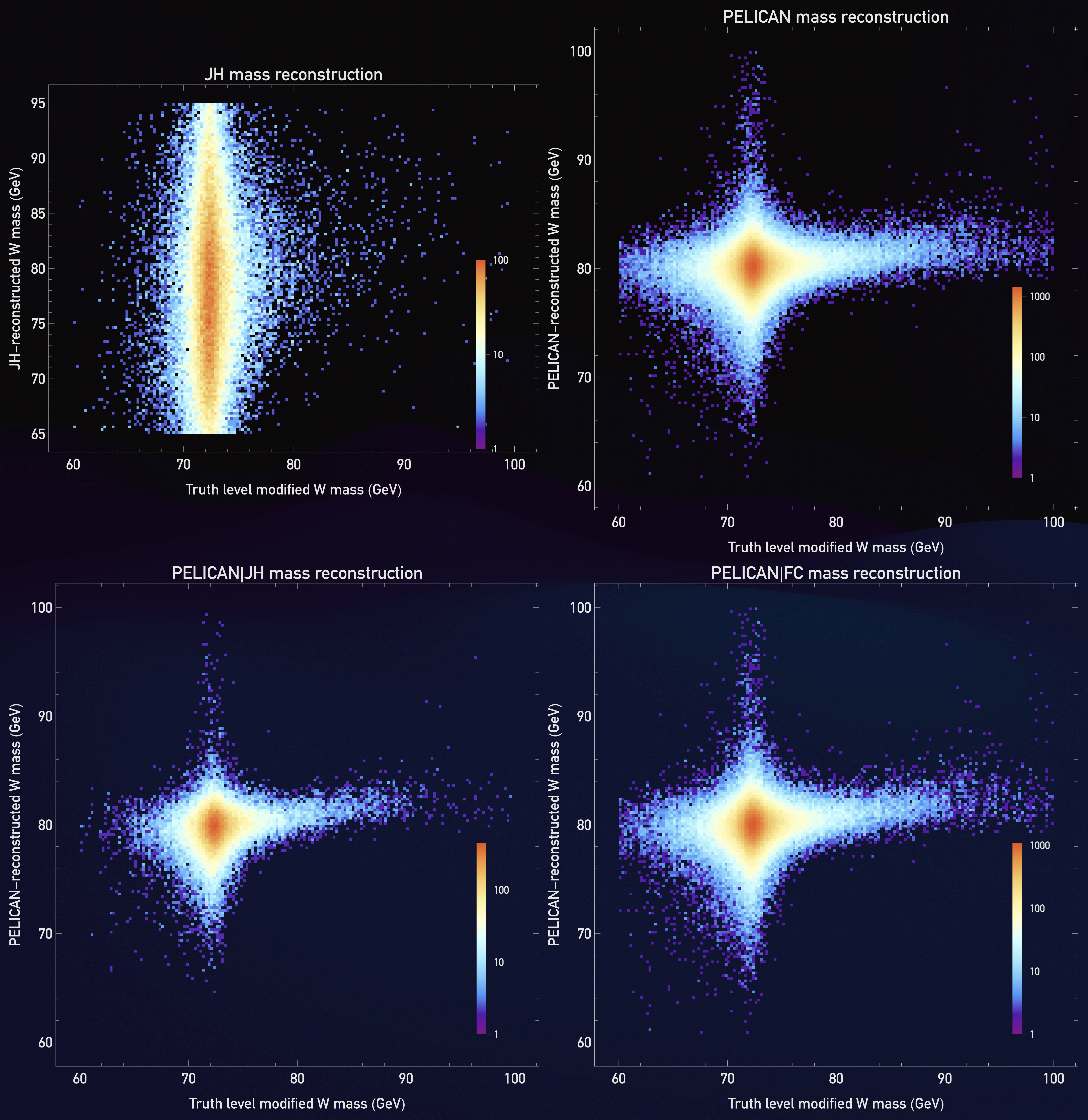
Trained on 80GeV jets
Evaluated on 72GeV jets



Transfer Learning: W mass measurement

Delphes

Trained on 80GeV jets
Evaluated on 72GeV jets



Other applications

Other applications

- IRC safe PELICAN

Other applications

- IRC safe PELICAN
- Looking inside jets: parent reconstruction

Other applications

- IRC safe PELICAN
- Looking inside jets: parent reconstruction
- Integrated pipeline: tagging → regression

Other applications

- IRC safe PELICAN
- Looking inside jets: parent reconstruction
- Integrated pipeline: tagging $\rightarrow$ regression
- Polarization tagging (helicity angle distributions)

Other applications

- IRC safe PELICAN
- Looking inside jets: parent reconstruction
- Integrated pipeline: tagging → regression
- Polarization tagging (helicity angle distributions)
- Jet combinatorics?