Does equivariance make better models?

Alexander Bogatskiy

Research Fellow, Flatiron Institute

BOOST 2024, Genova, Italy

Fundamental Science models

Minimally parametrized
Interpretable
Generalizable

Fundamental Science models

Minimally parametrized
Interpretable
Generalizable

Overparametrized
Black box
Limited generalization

Machine Learning models

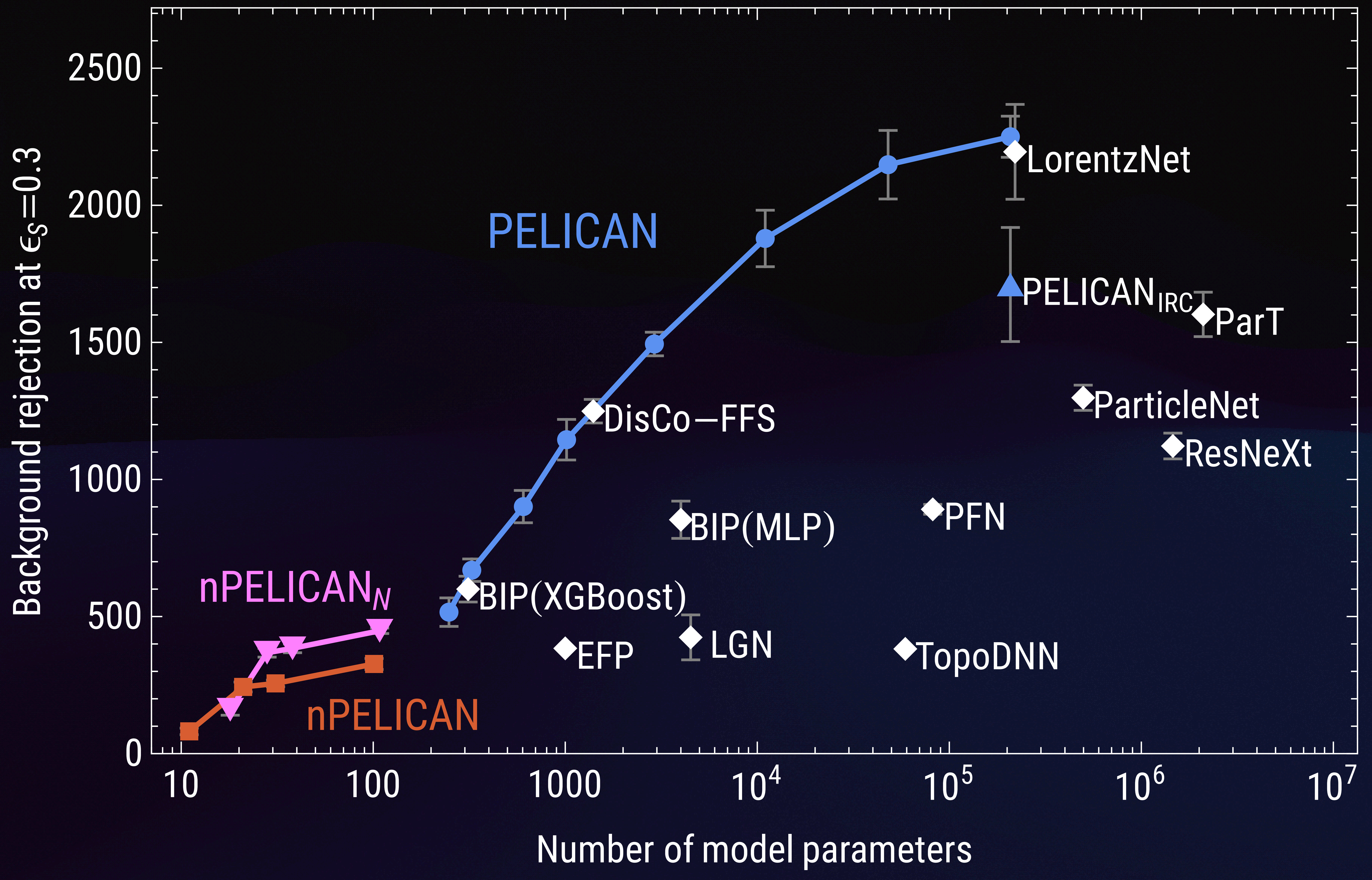
Fundamental Science models

Minimally parametrized
Interpretable
Generalizable

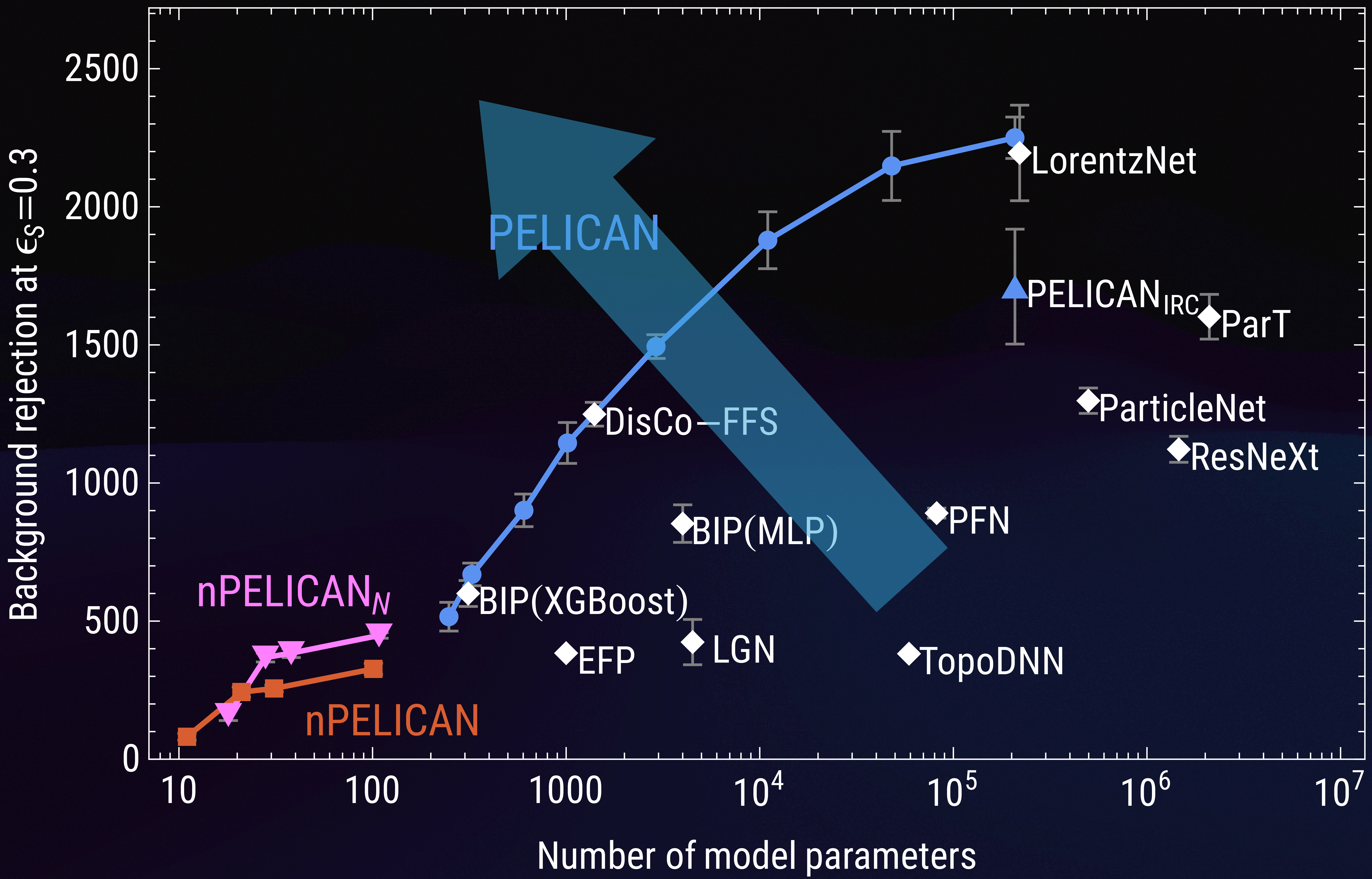
Overparametrized
Black box
Limited generalization

Machine Learning models

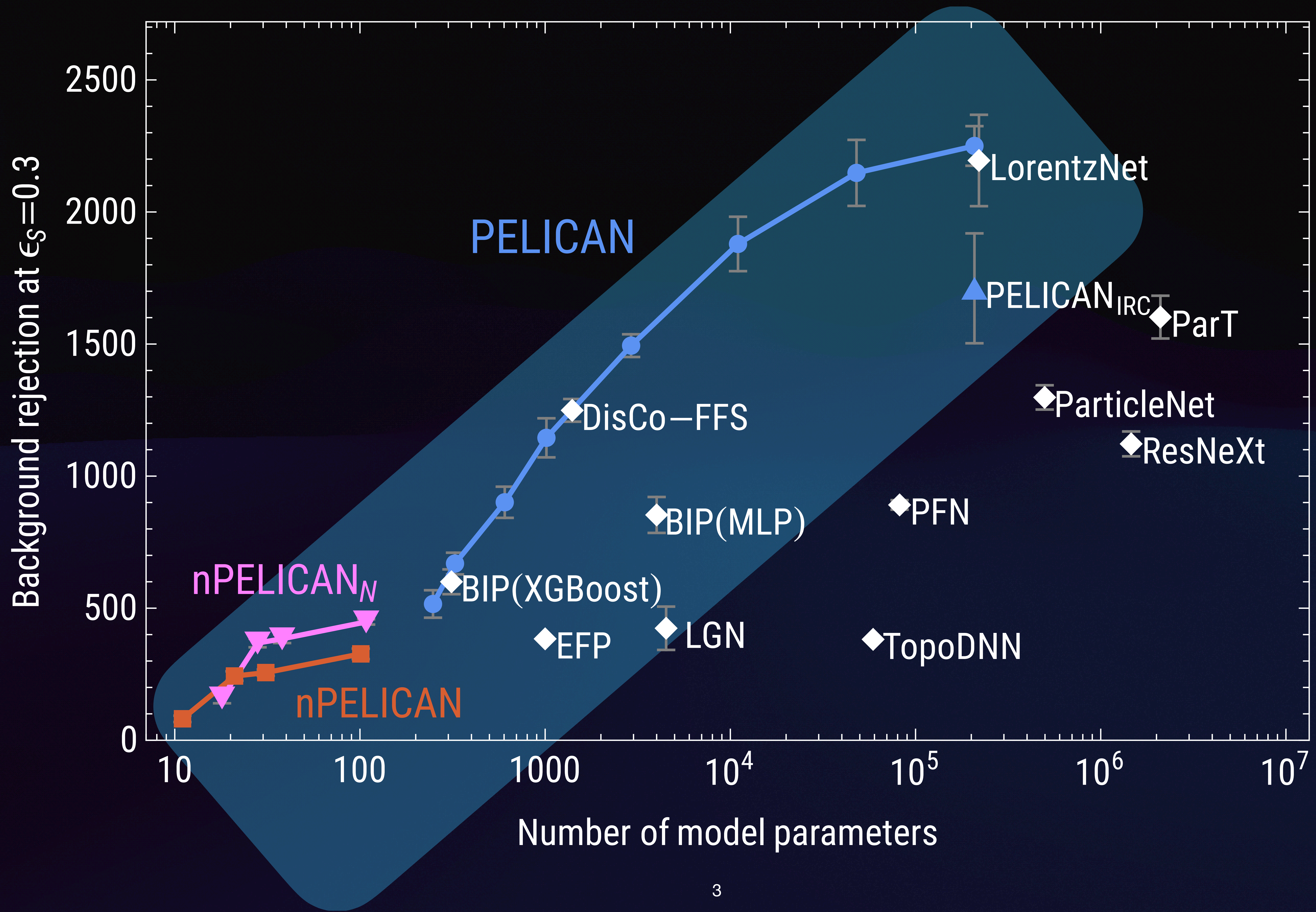
Comparison of top-taggers (binary classification of jets)



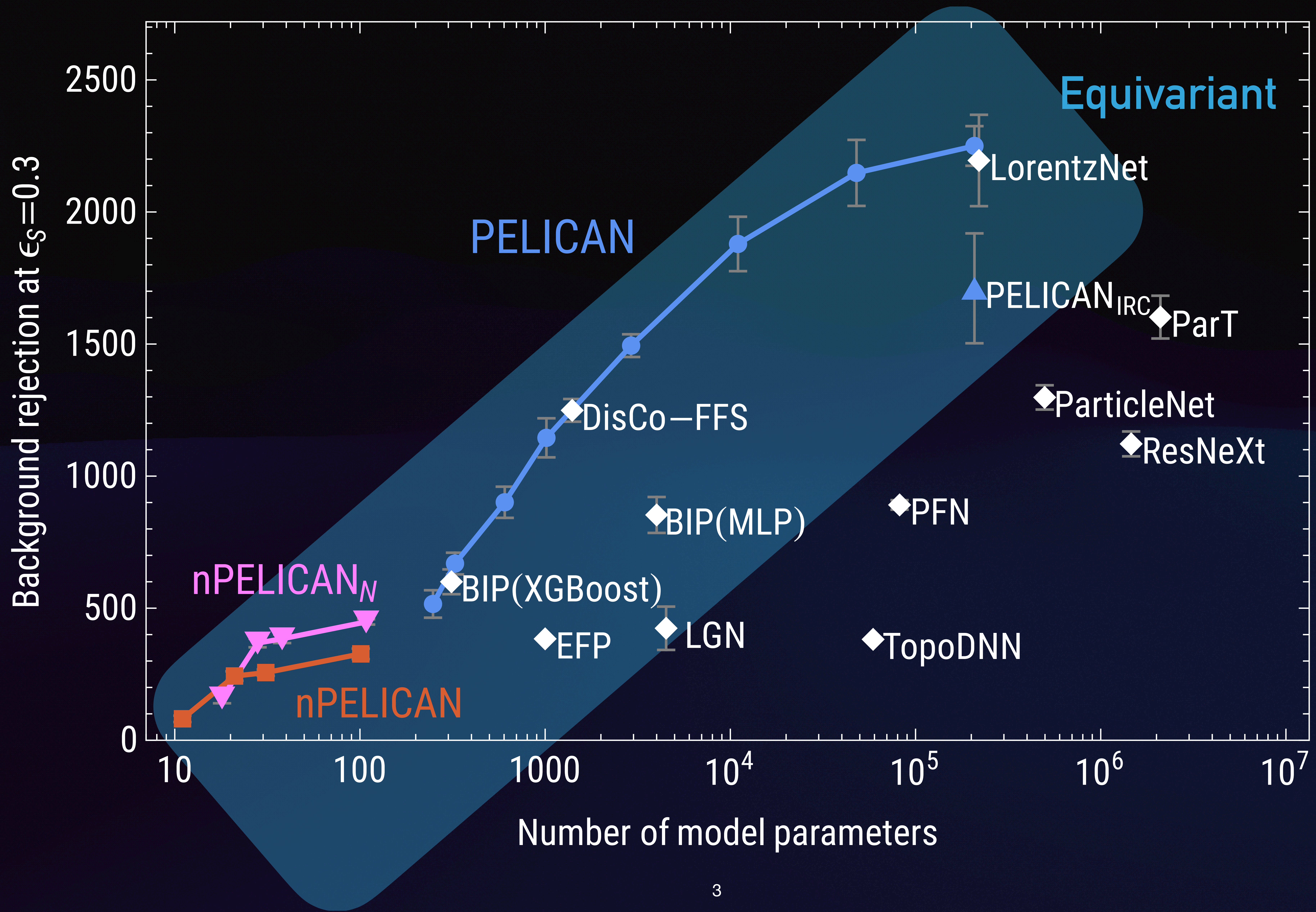
Comparison of top-taggers (binary classification of jets)



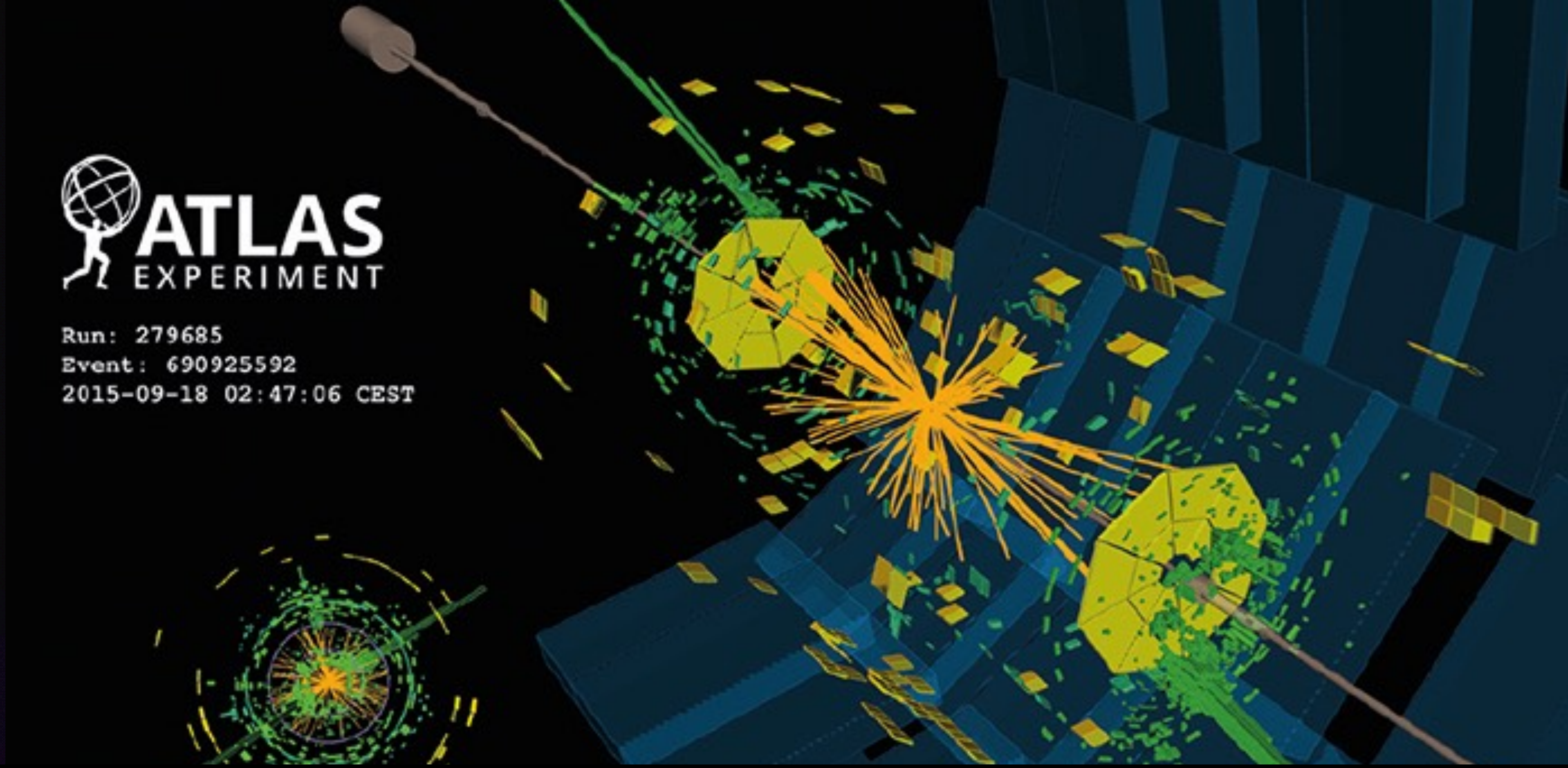
Comparison of top-taggers (binary classification of jets)

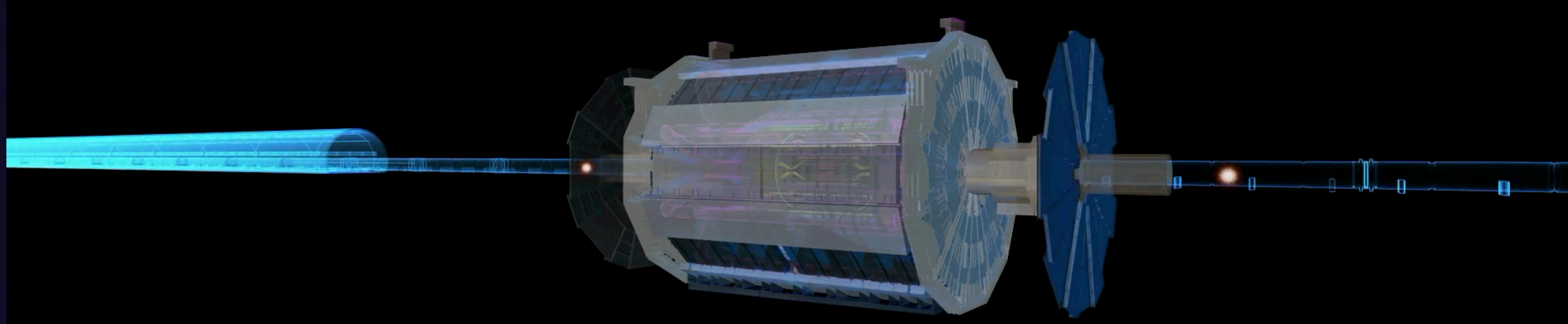


Comparison of top-taggers (binary classification of jets)



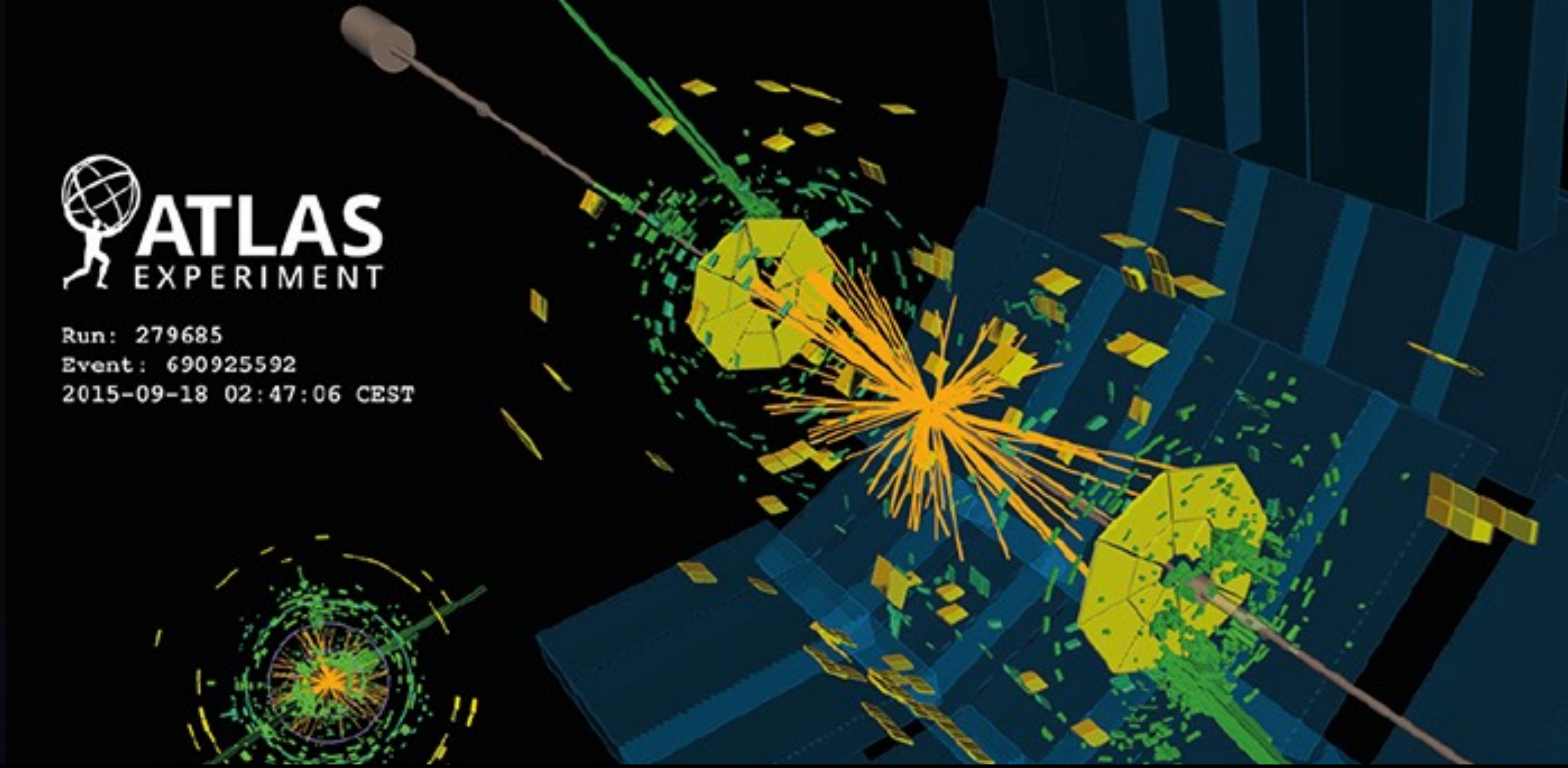
Symmetries in jet data

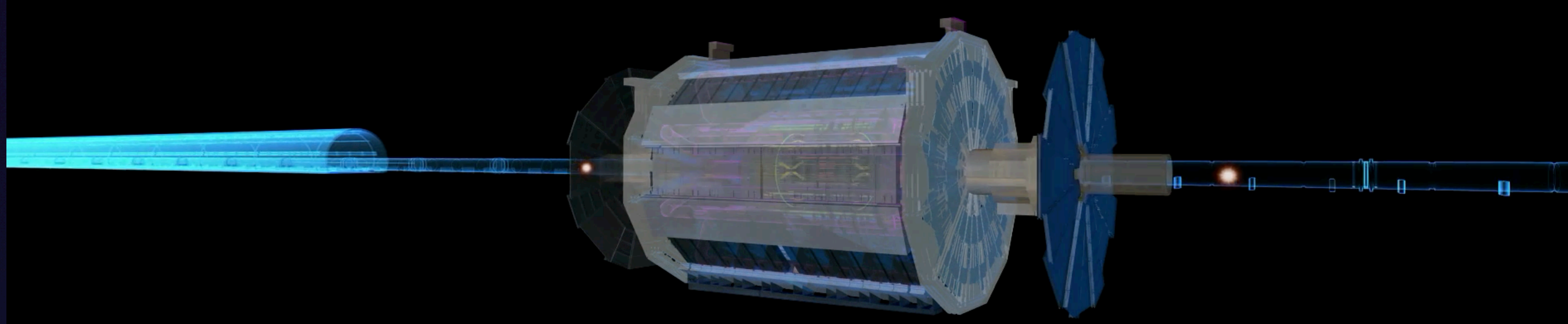




ATLAS Experiment © CERN

Symmetries in jet data

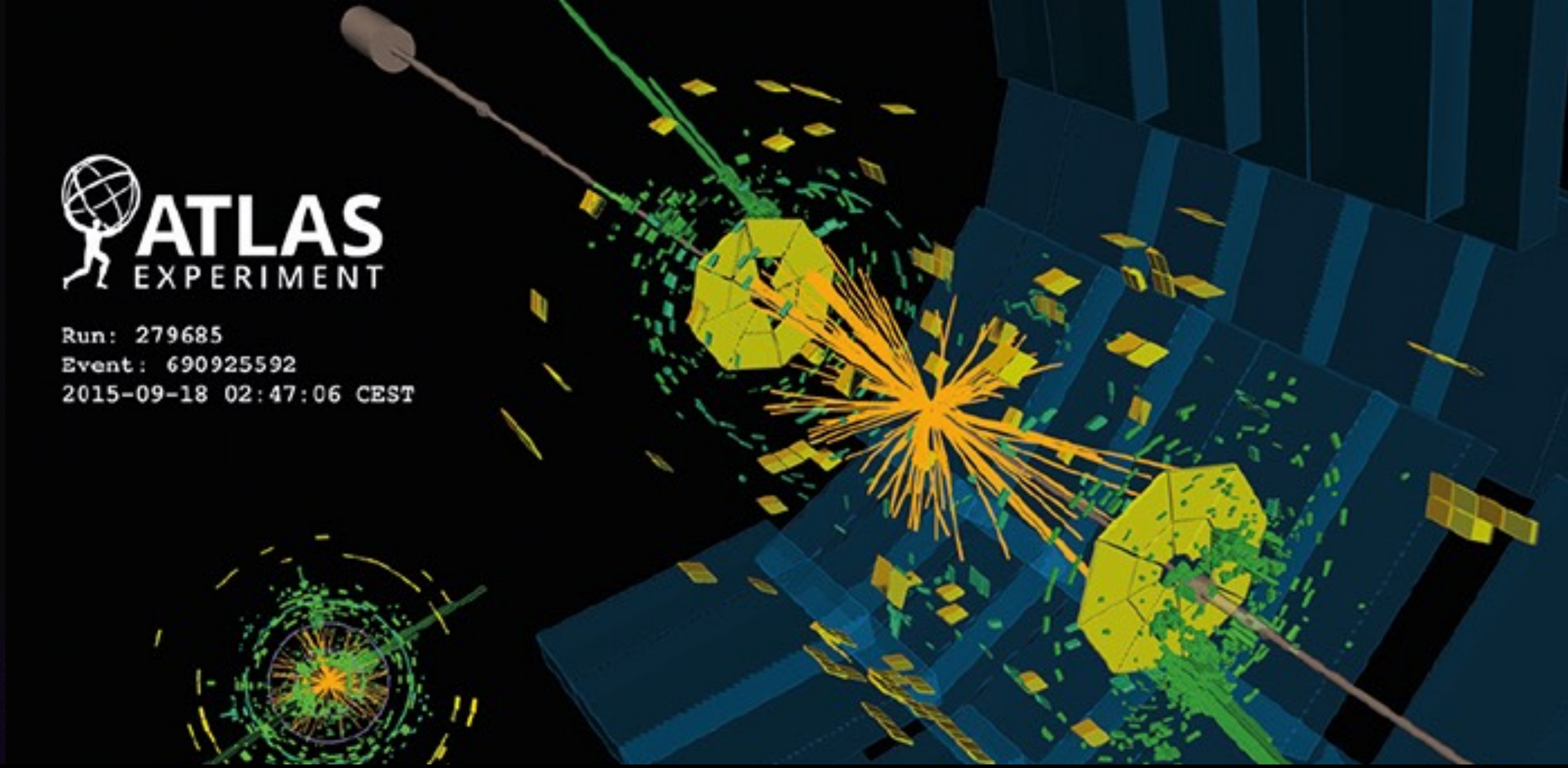


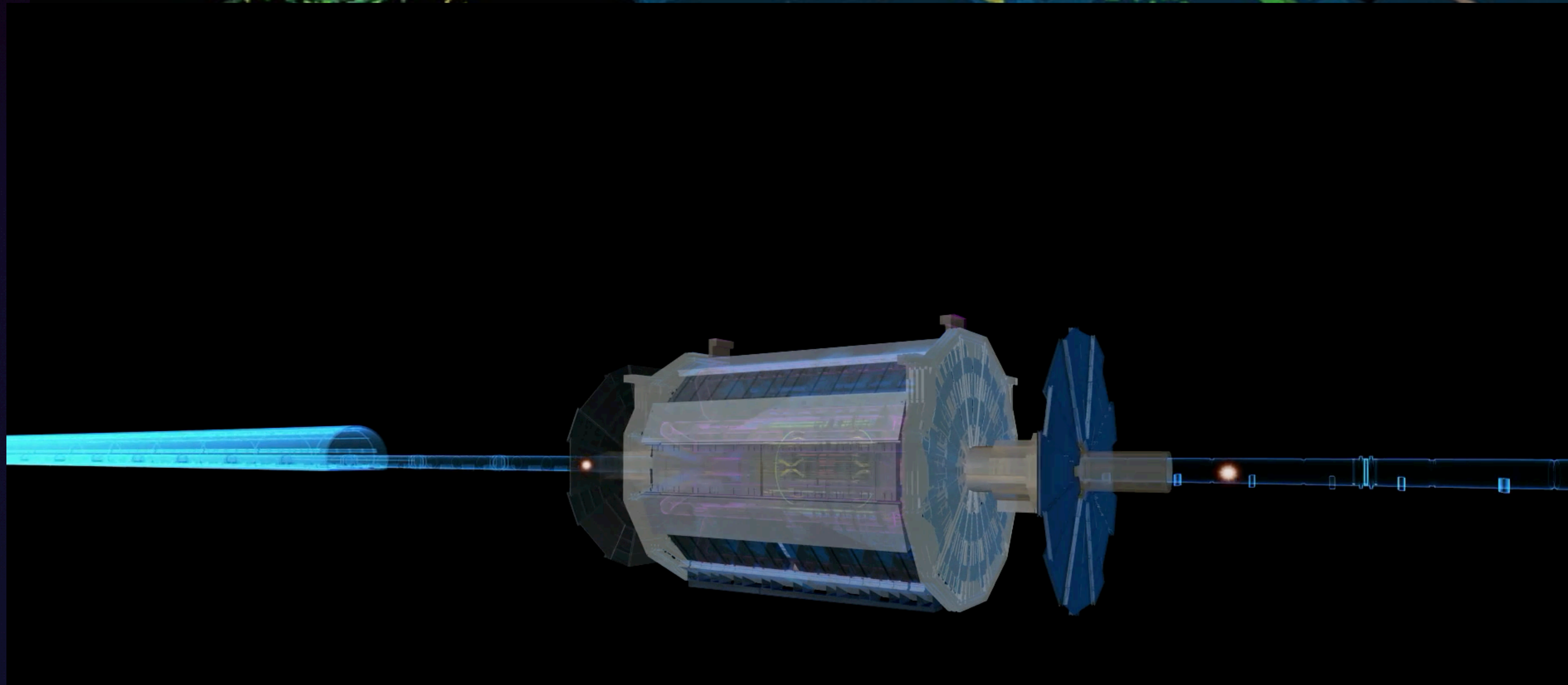


ATLAS Experiment © CERN

Symmetries in jet data

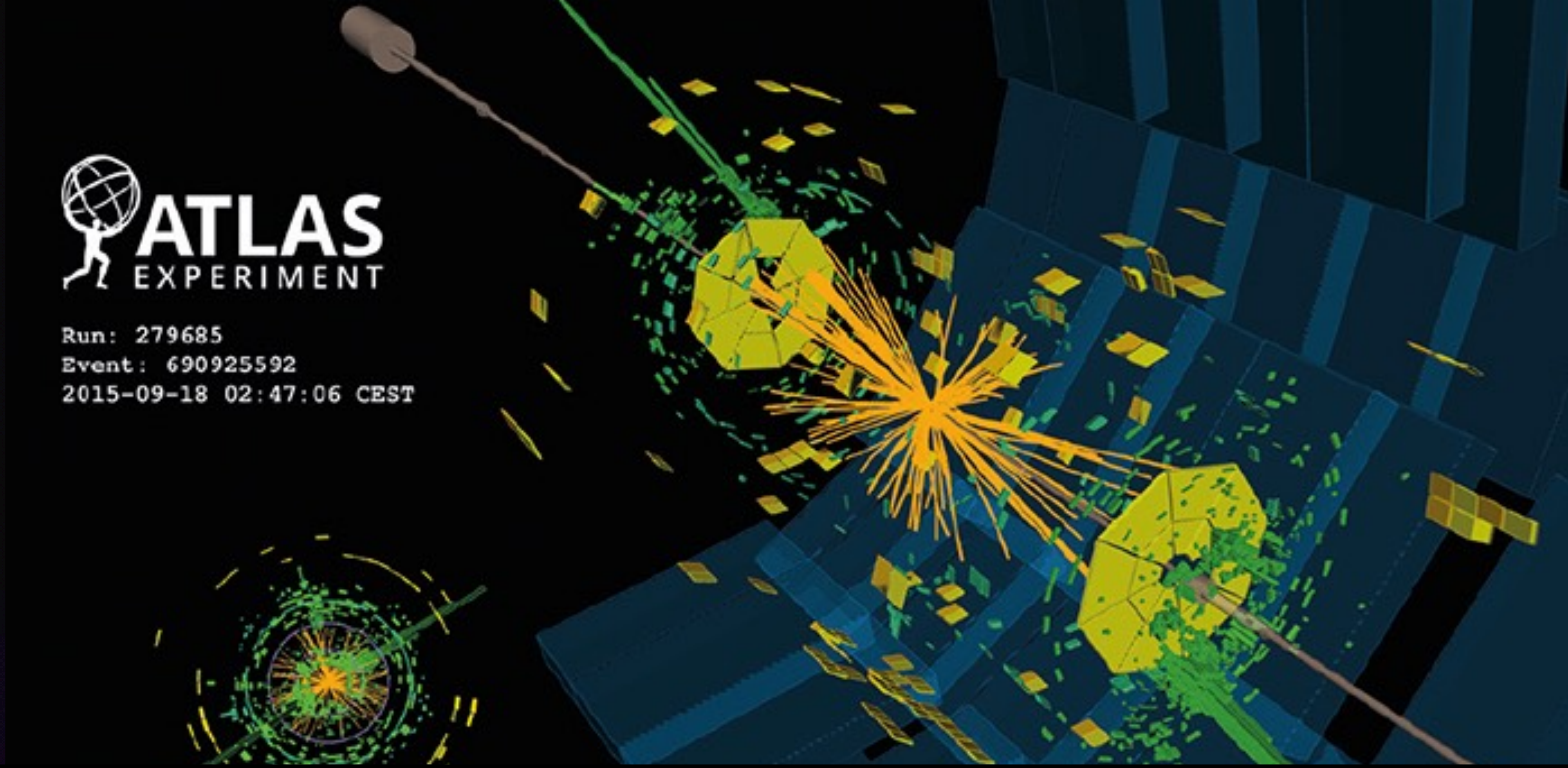
- Permutations of constituents





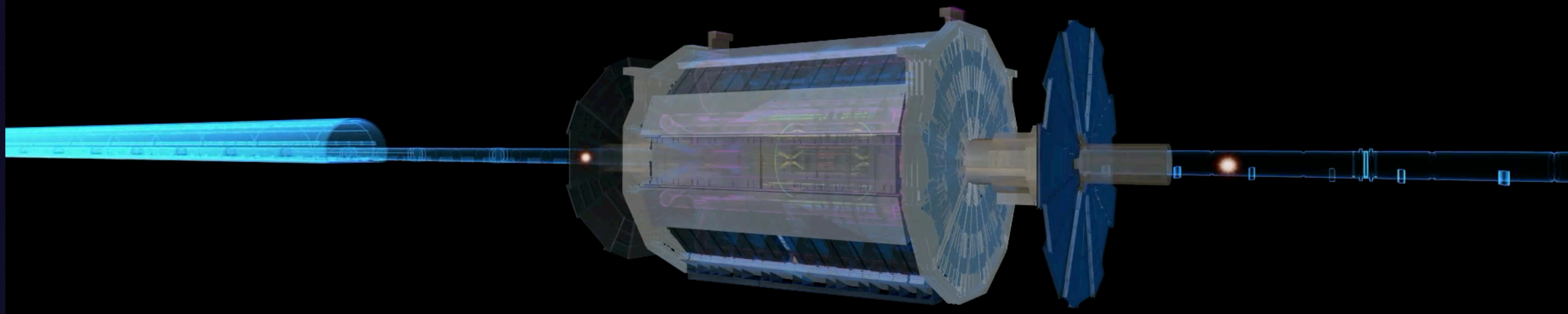
ATLAS Experiment © CERN

Symmetries in jet data



- Permutations of constituents

- Rotational, boost, Lorentz symmetries



ATLAS Experiment © CERN

To isolate effects of equivariance:

To isolate effects of equivariance:

1. Take a particular equivariant model

To isolate effects of equivariance:

1. Take a particular equivariant model

2. “Break” its equivariance in a “minimal” way

To isolate effects of equivariance:

1. Take a particular equivariant model
2. "Break" its equivariance in a "minimal" way
   - Change only the input layer

To isolate effects of equivariance:

1. Take a particular equivariant model

2. “Break” its equivariance in a “minimal” way

- Change only the input layer

- Same architecture, same hyper parameters and optimization

To isolate effects of equivariance:

1. Take a particular equivariant model
2. “Break” its equivariance in a “minimal” way
    - Change only the input layer
    - Same architecture, same hyper parameters and optimization
3. Compare to an equivariant model that accesses the same information content in the data.

To isolate effects of equivariance:

High Energy Physics – Phenomenology

[Submitted on 16 Aug 2022 (v1), last revised 7 Mar 2024 (this version, v3)]

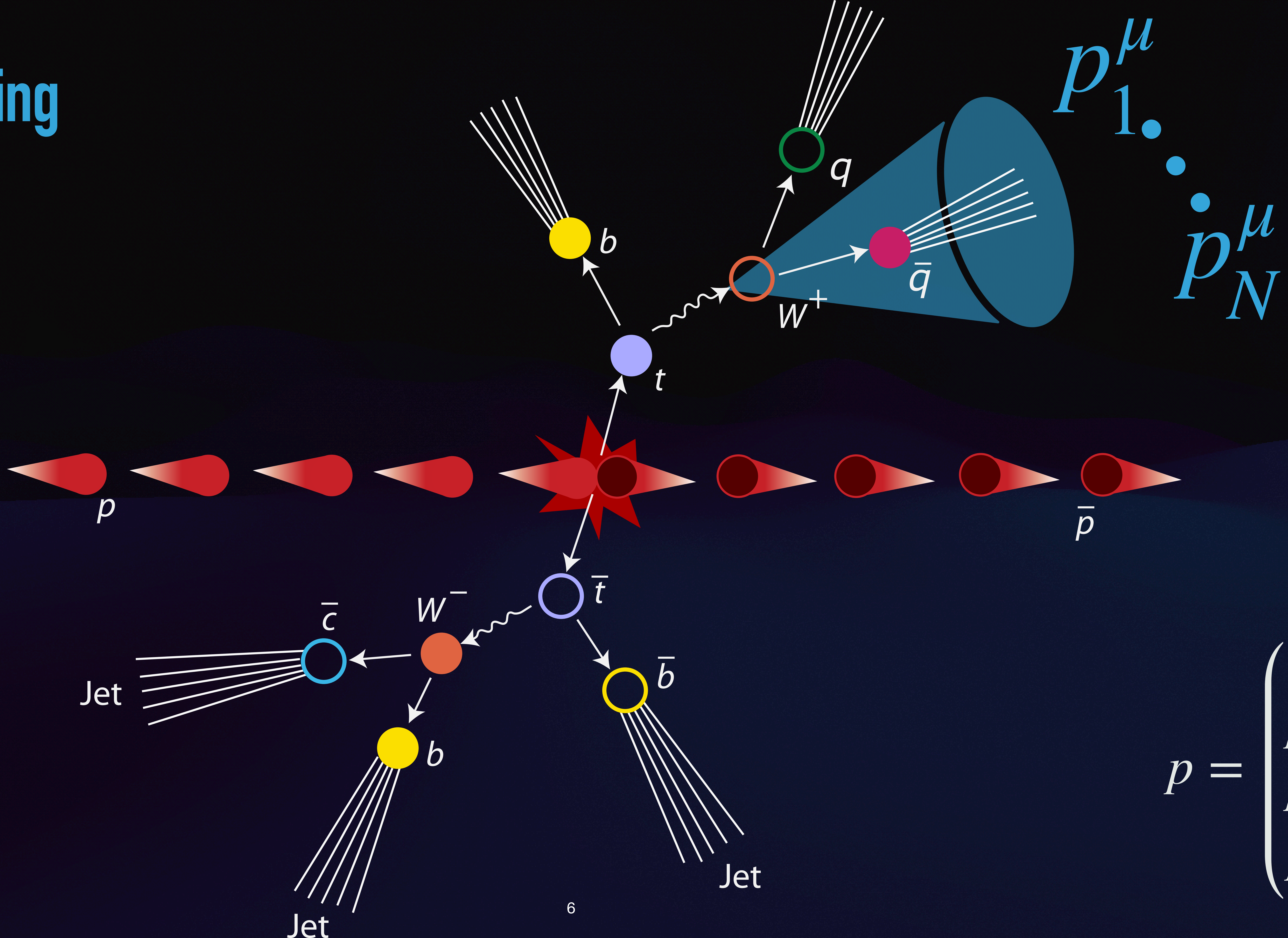
Does Lorentz-symmetric design boost network performance in jet physics?

Congqiao Li, Huilin Qu, Sitian Qian, Qi Meng, Shiqi Gong, Jue Zhang, Tie-Yan Liu, Qiang Li

In the deep learning era, improving the neural network performance in jet physics is a rewarding task as it directly contributes to more accurate physics measurements at the LHC. Recent research has proposed various network designs in consideration of the full Lorentz symmetry, but its benefit is still not systematically asserted, given that there remain many successful networks without taking it into account. We conduct a detailed study on the Lorentz-symmetric design. We

1. Take a particular equivariant model
2. “Break” its equivariance in a “minimal” way
   - Change only the input layer
   - Same architecture, same hyper parameters and optimization
3. Compare to an equivariant model that accesses the same information content in the data.

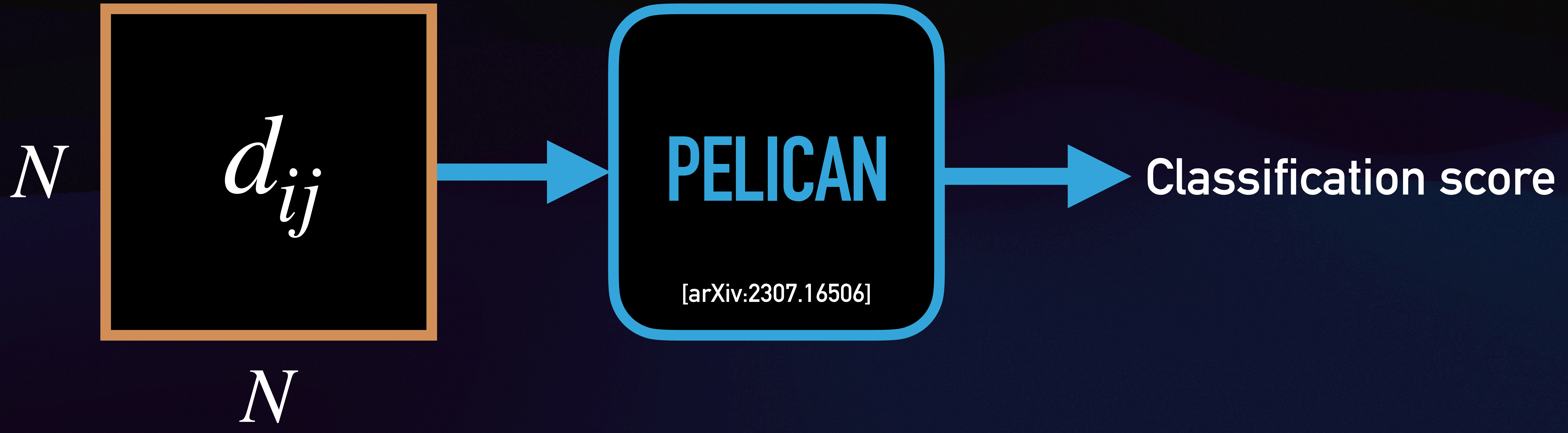
Jet tagging



$$p = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix}$$

PELICAN

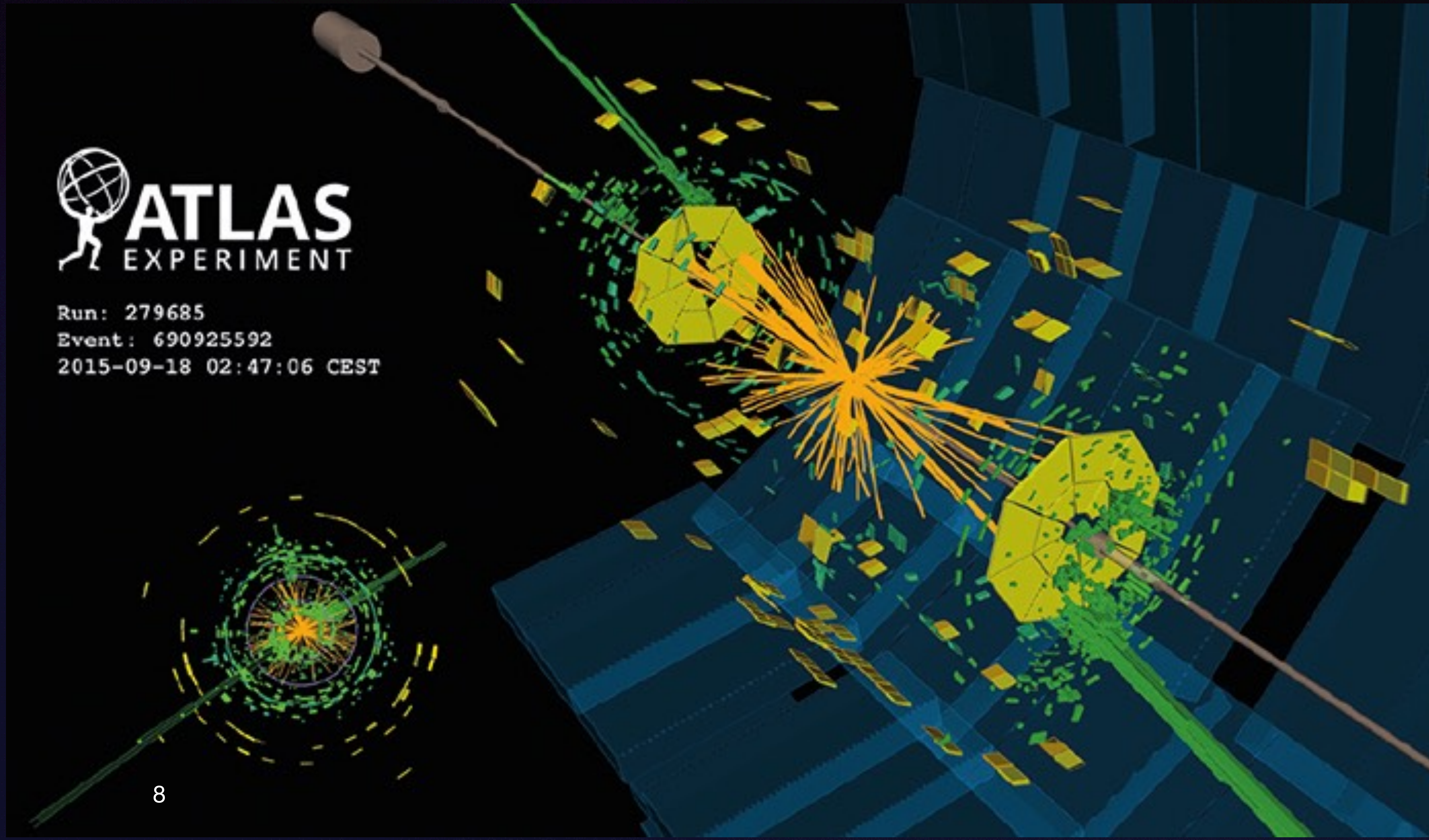
Inputs: $d_{ij} = p_i \cdot p_j = E_i E_j - \vec{p}_i \cdot \vec{p}_j$



Invariant w.r.t. the Lorentz group $\mathrm{SO}^+_{1,3}$

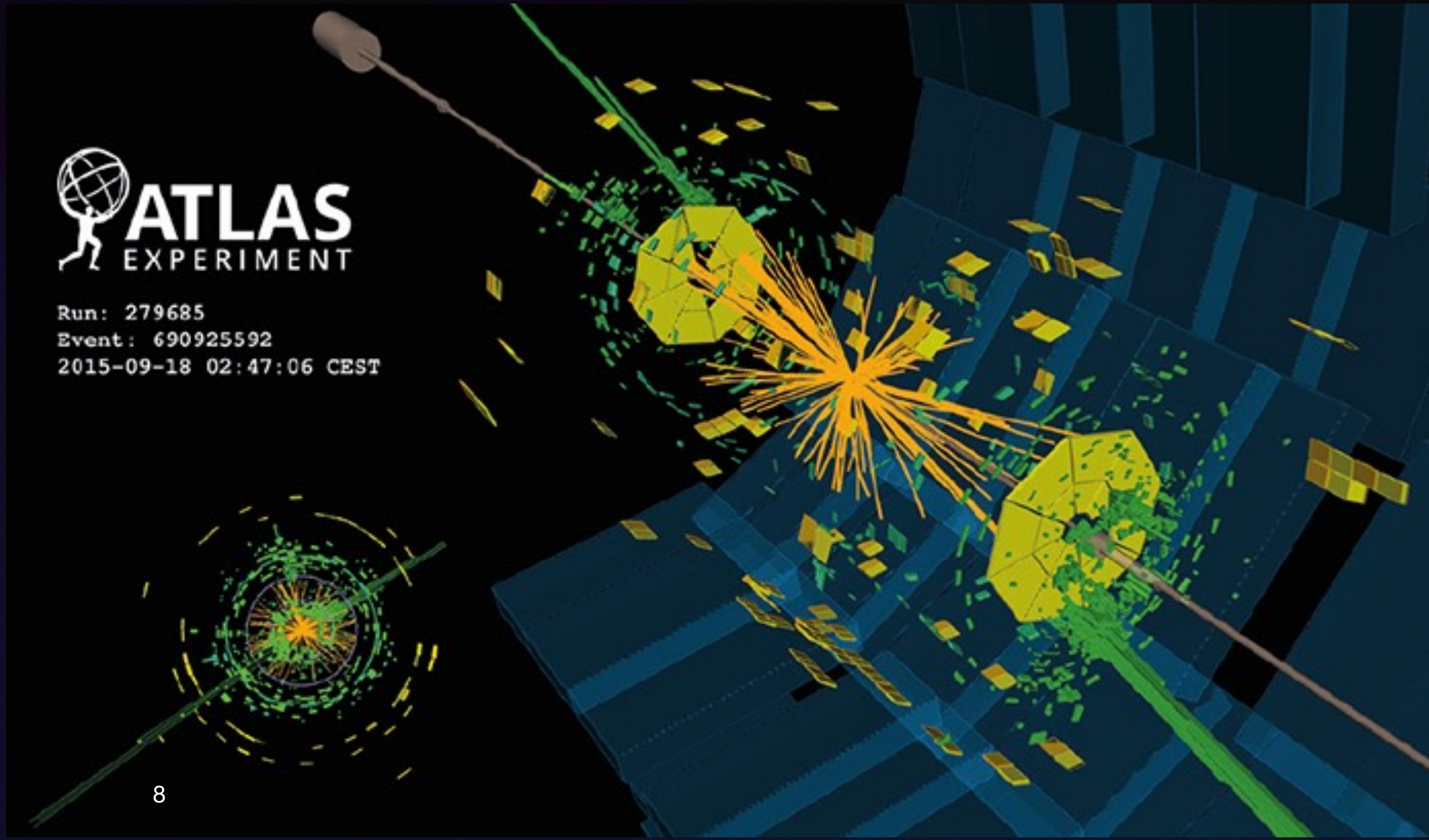
State-of-the-art across multiple tagging and regression tasks

What information is “not invariant”?



What information is "not invariant"?

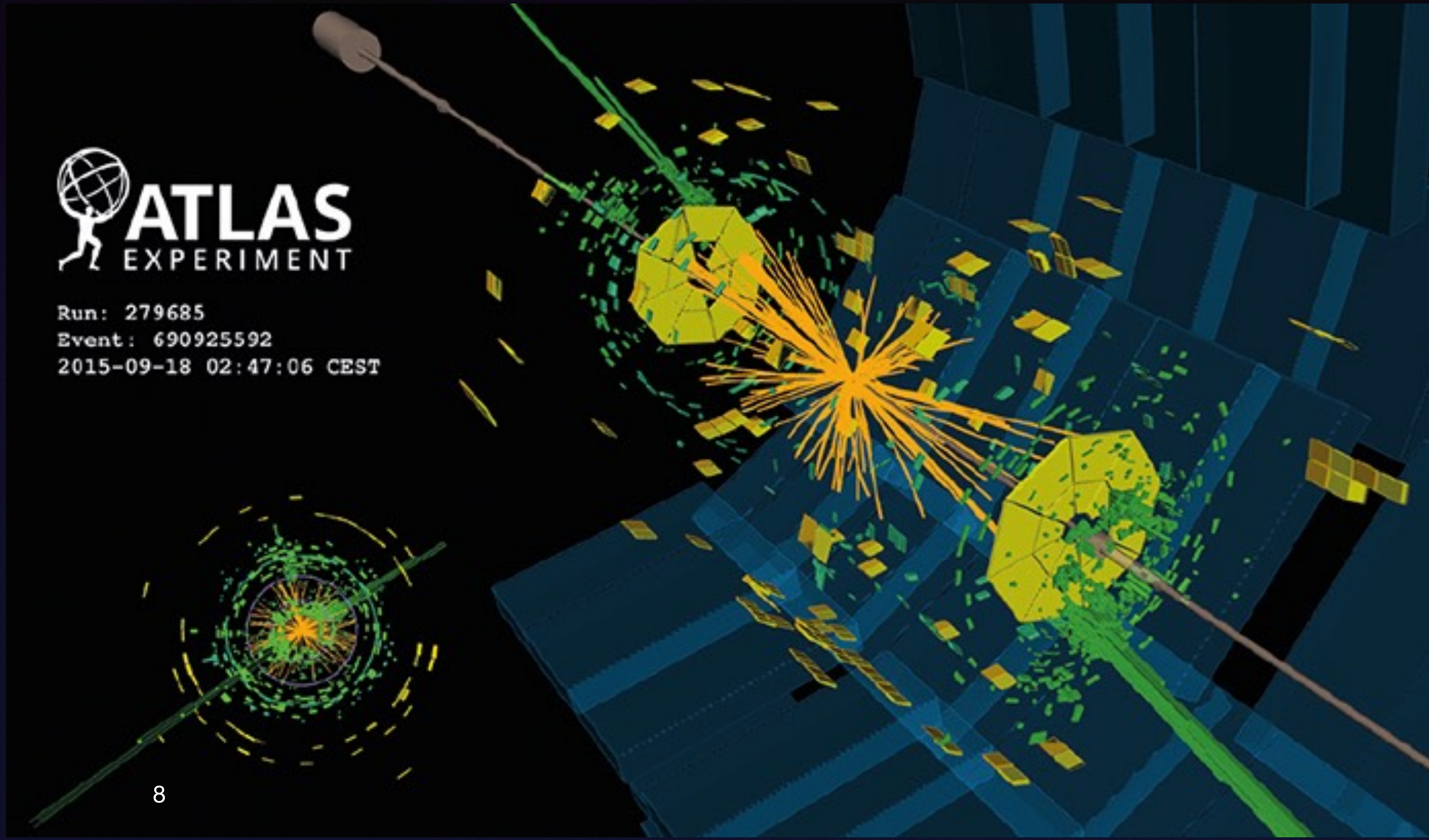
1. Detector orientation: the z-axis is special. Explicit dependence on E, p_z.



What information is "not invariant"?

1. Detector orientation: the z-axis is special. Explicit dependence on E, p_z.

2. Finite resolution introduces potential non-isotropy w.r.t. φ-rotations.



Two methods

Promote non-invariant data to an input of an invariant architecture

"Spurion method"

(thanks to Jesse Thaler)

Add non-invariant data as a direct "scalar" input

"Input method"

Two methods

Promote non-invariant data to an input of an invariant architecture

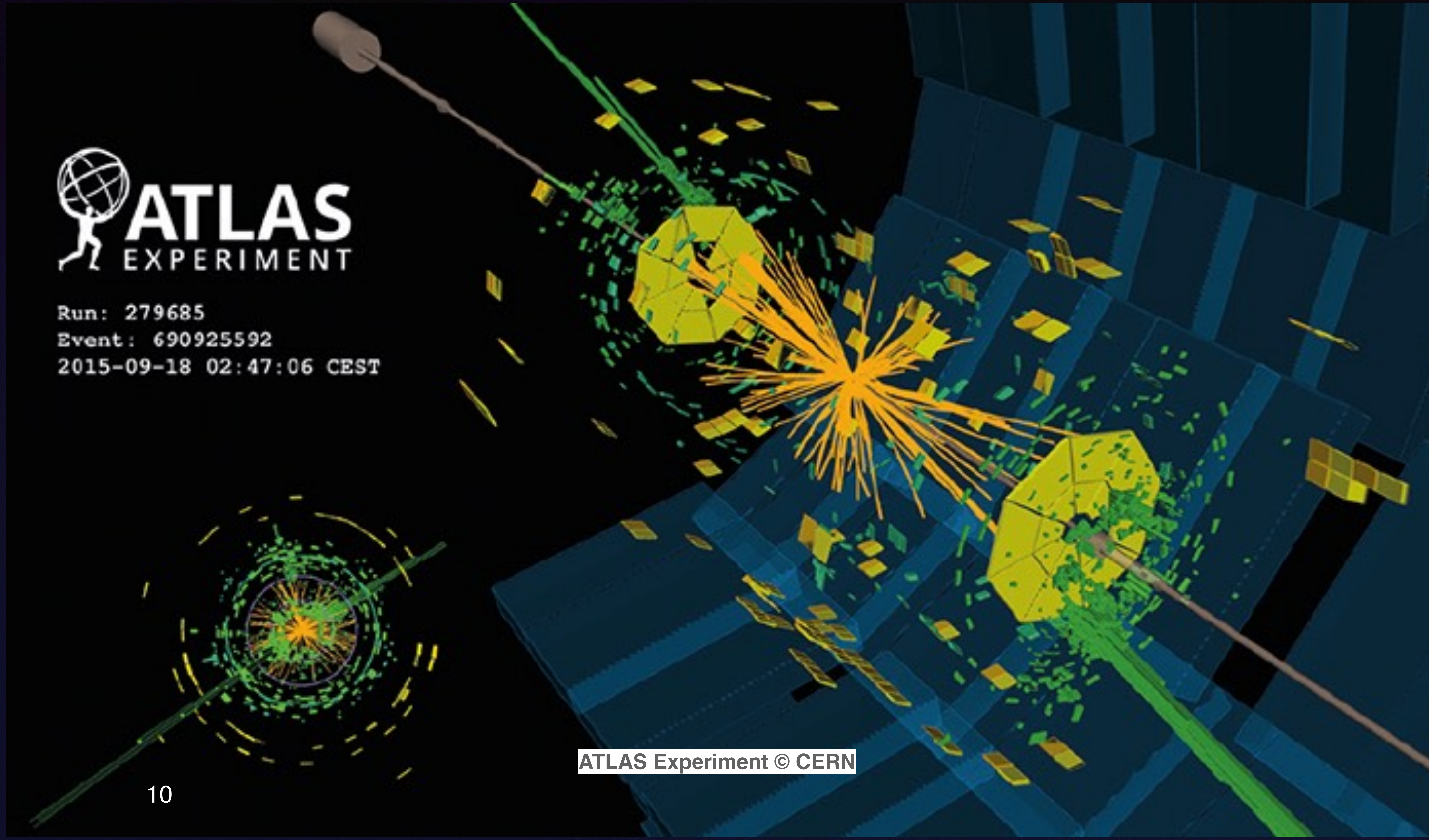
"Spurion method"
(thanks to Jesse Thaler)

Add non-invariant data as a direct "scalar" input

"Input method"

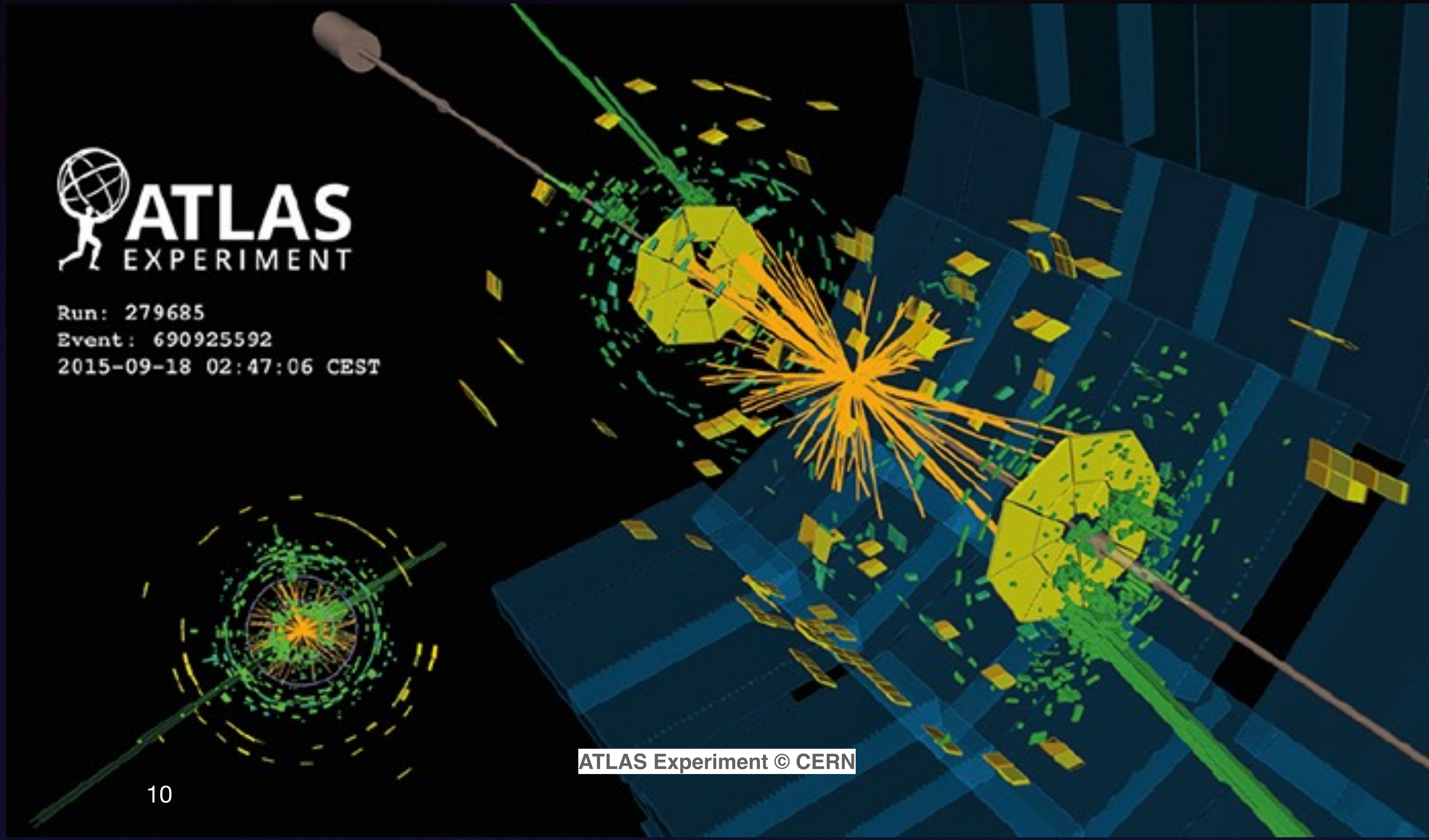
If the spurion method produces better models, equivariance can be claimed to be a source of a performance boost.

Workaround: spurions



Workaround: spurions

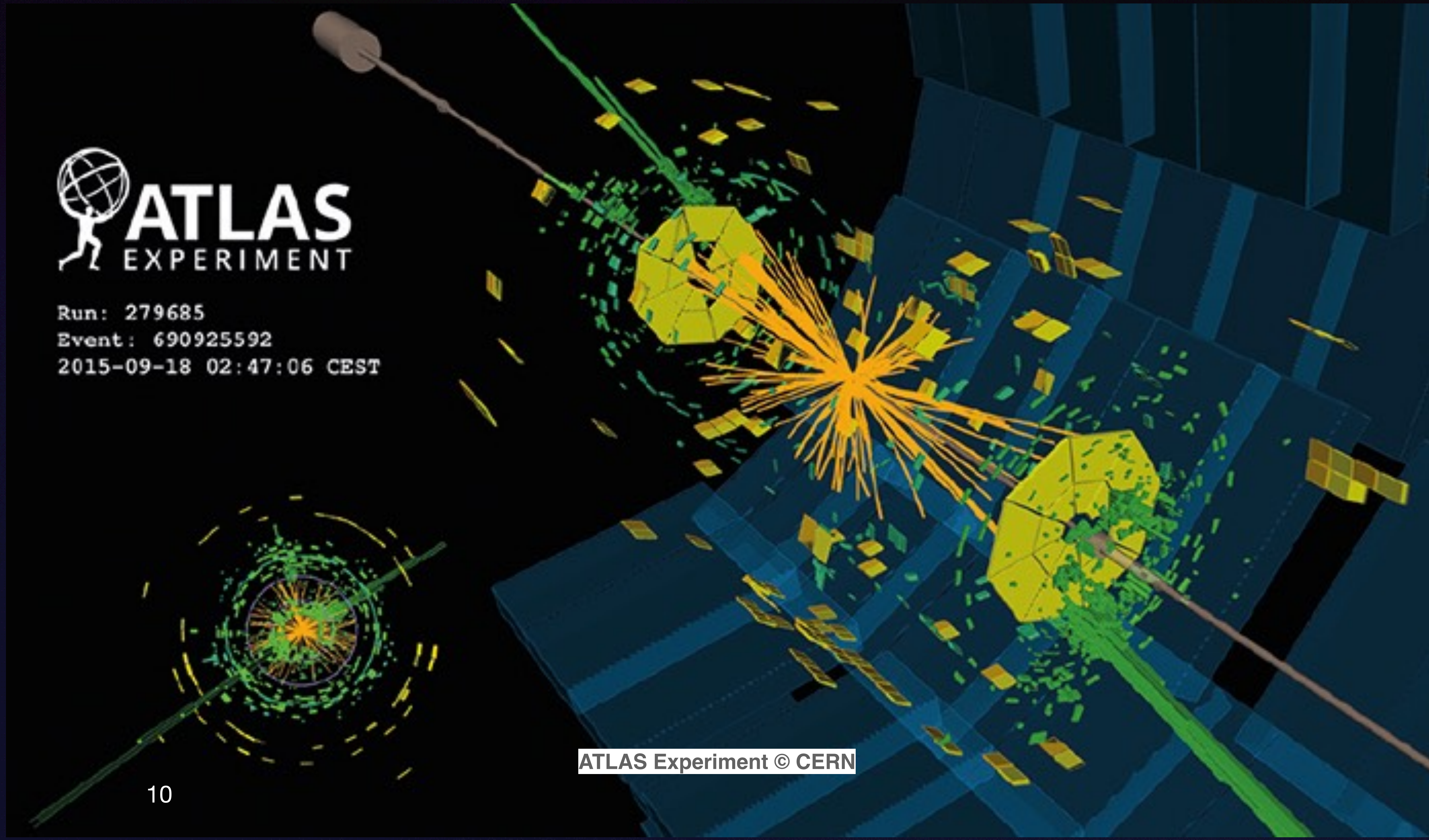
1. Treat detector geometry as part of input



Workaround: spurions

1. Treat detector geometry as part of input

2. To each event append fixed 4-vectors like $e_{\pm} = (1,0,0,\pm 1)$



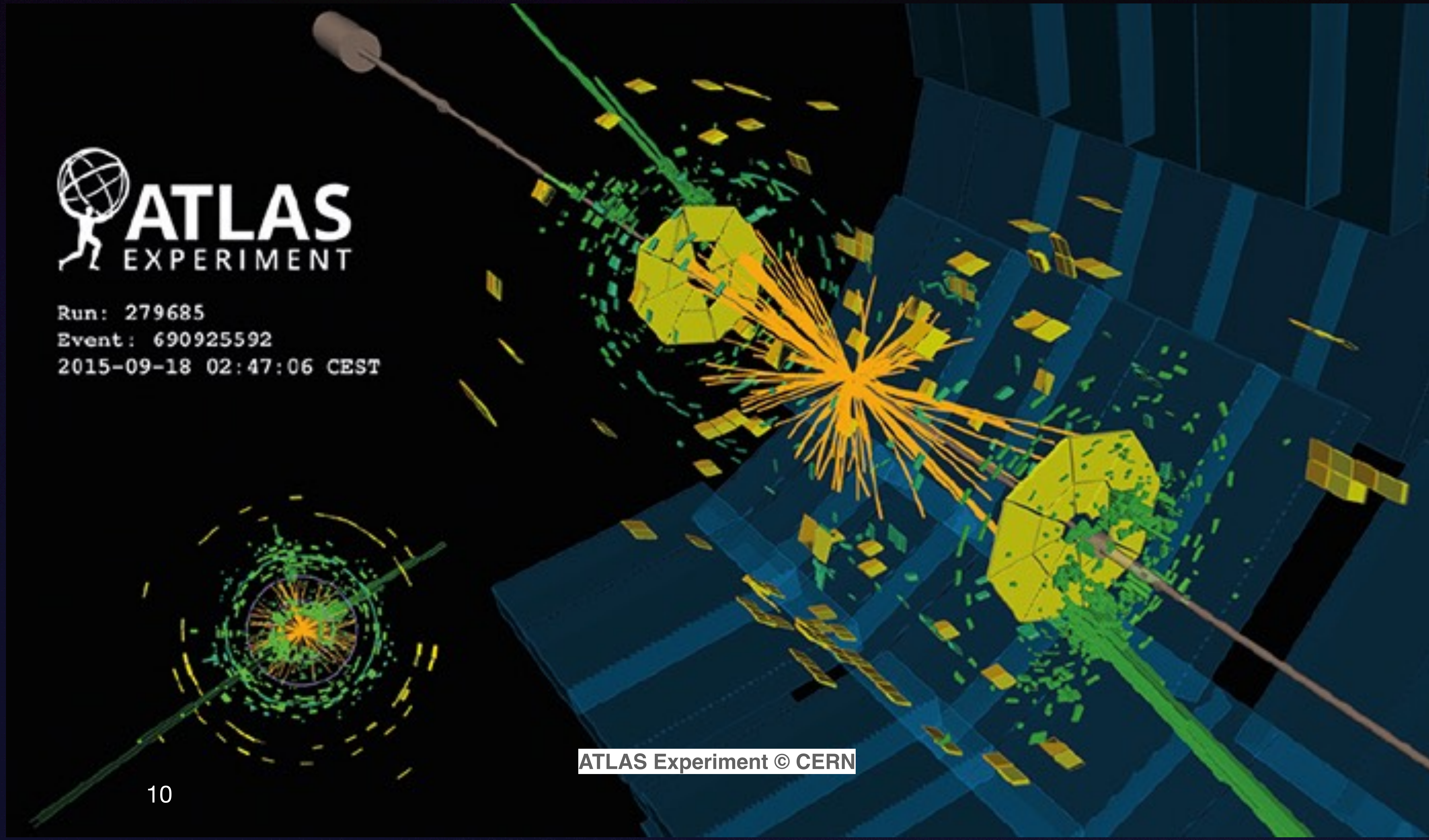
Workaround: spurions

1. Treat detector geometry as part of input

2. To each event append fixed 4-vectors like $e_\pm = (1,0,0,\pm 1)$

$$\cancel{p_1, p_2, \ldots, p_N}$$

$$e_+, e_-, p_1, p_2, \ldots, p_N$$



Workaround: spurions

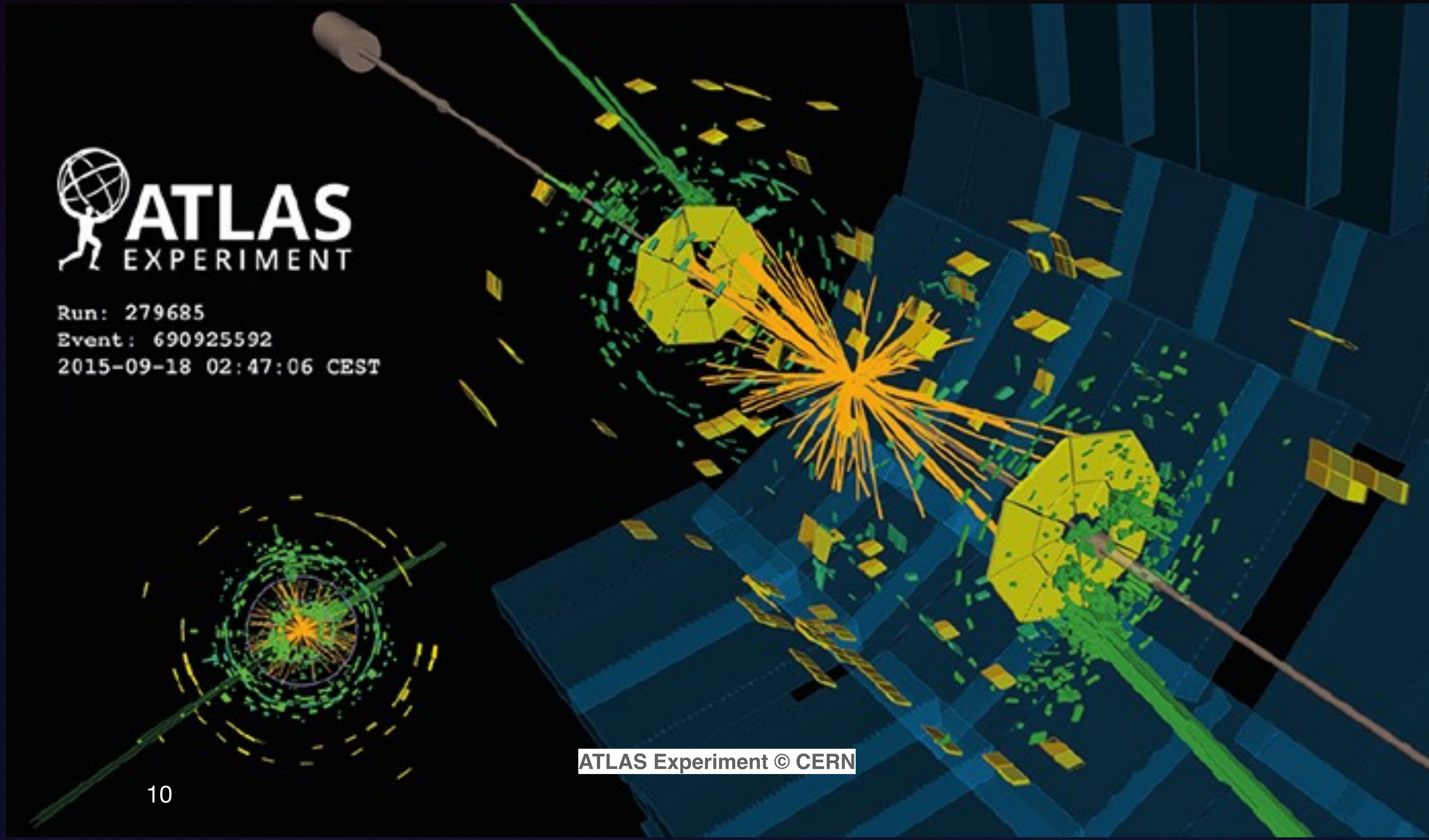
1. Treat detector geometry as part of input

2. To each event append fixed 4-vectors like $e_{\pm} = (1,0,0,\pm 1)$

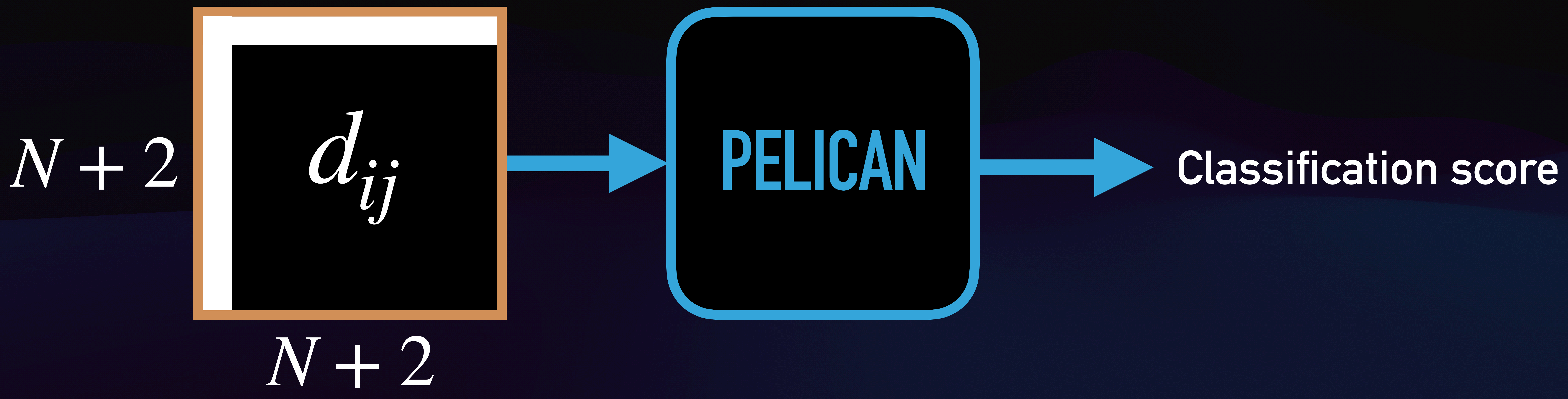
$$\cancel{p_1, p_2, \ldots, p_N}$$

$$e_+, e_-, p_1, p_2, \ldots, p_N$$

Global Lorentz is preserved, while on the jet only SO_2 remains.

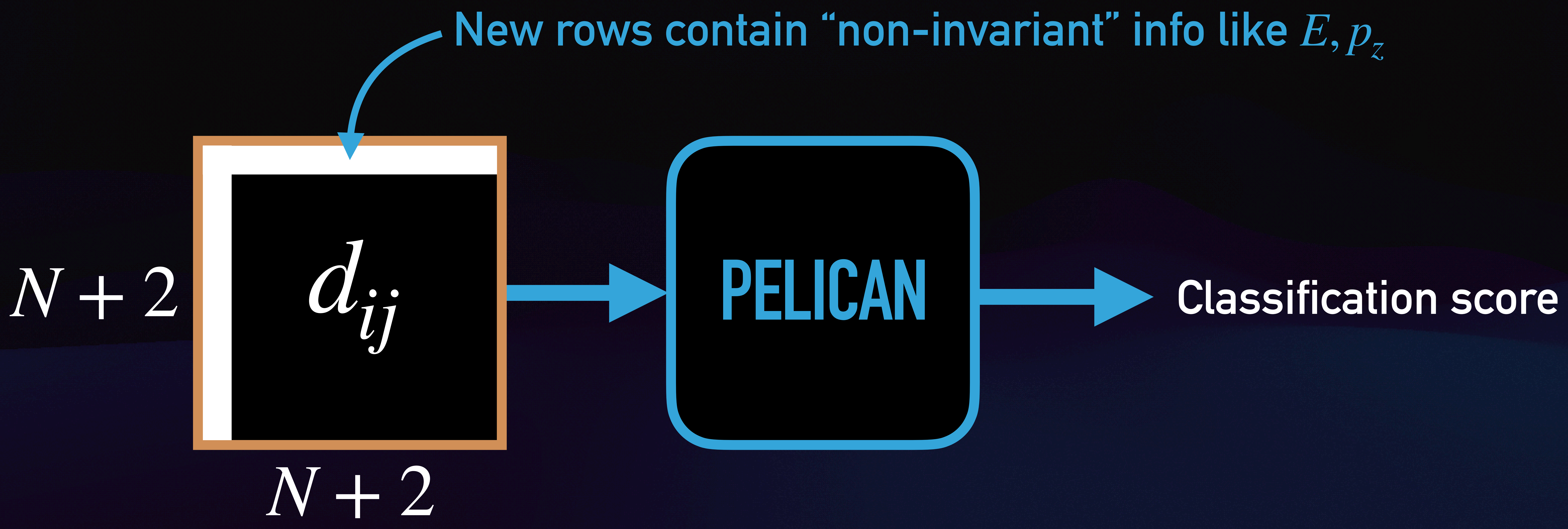


Workaround: spurions



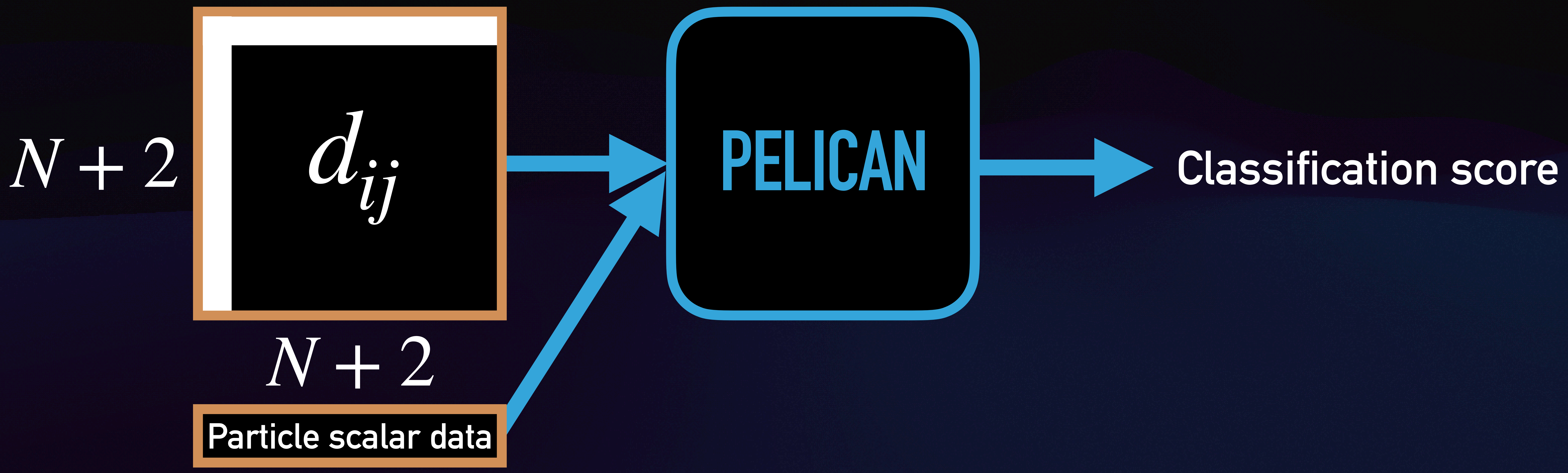
The architecture remains exactly the same and globally Lorentz invariant!

Workaround: spurions

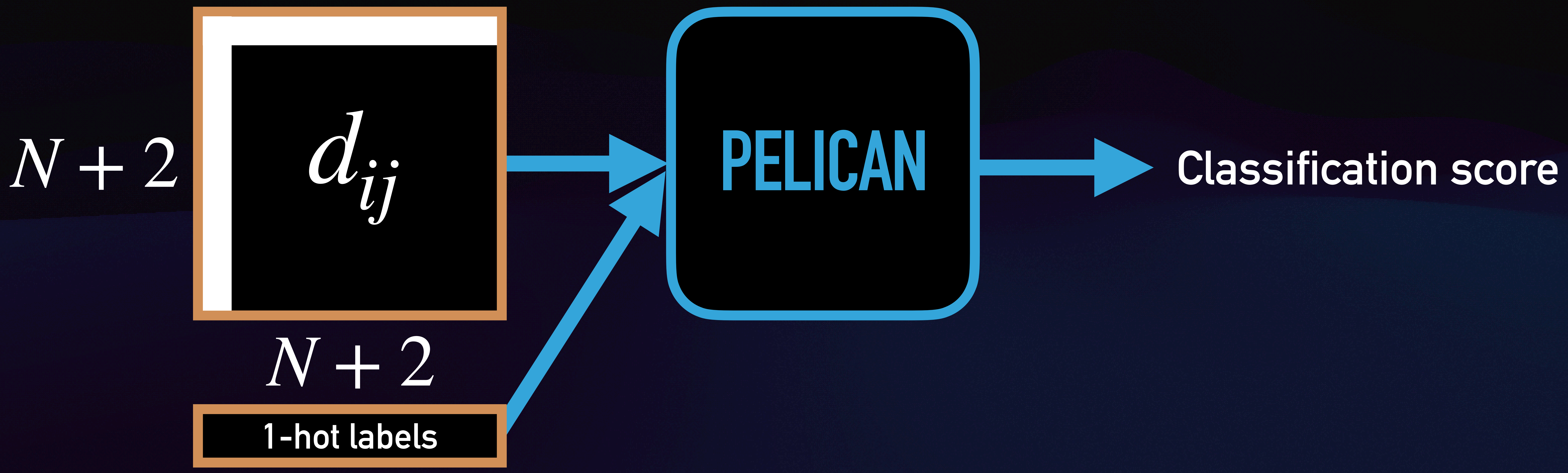


The architecture remains exactly the same and globally Lorentz invariant!

Workaround: spurions

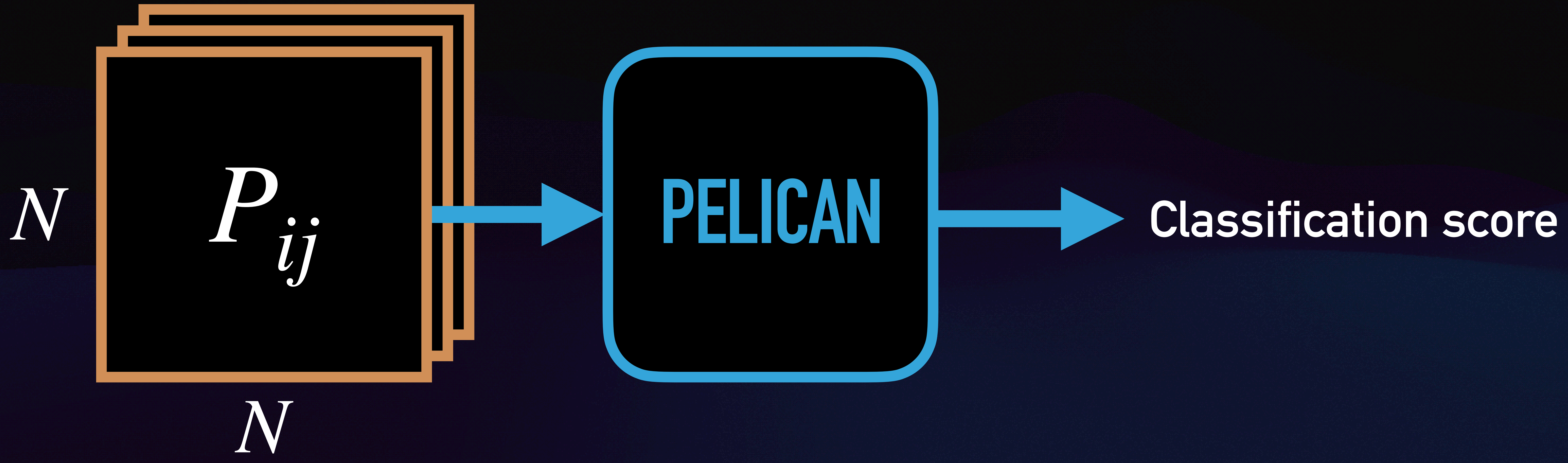


Workaround: spurions



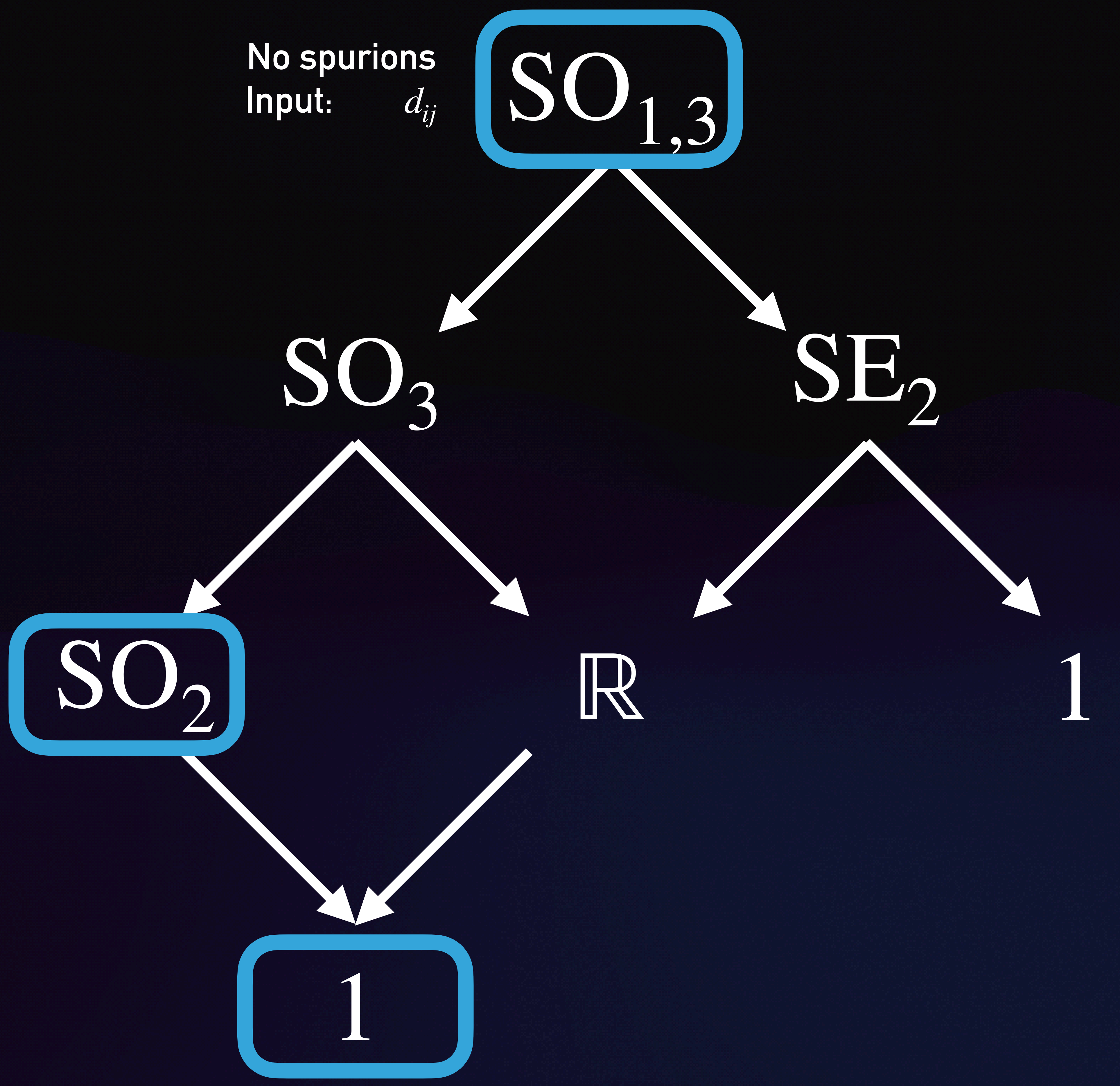
Each spurion gets a unique label, different from the common label of a jet constituent.

Alternative: extra inputs

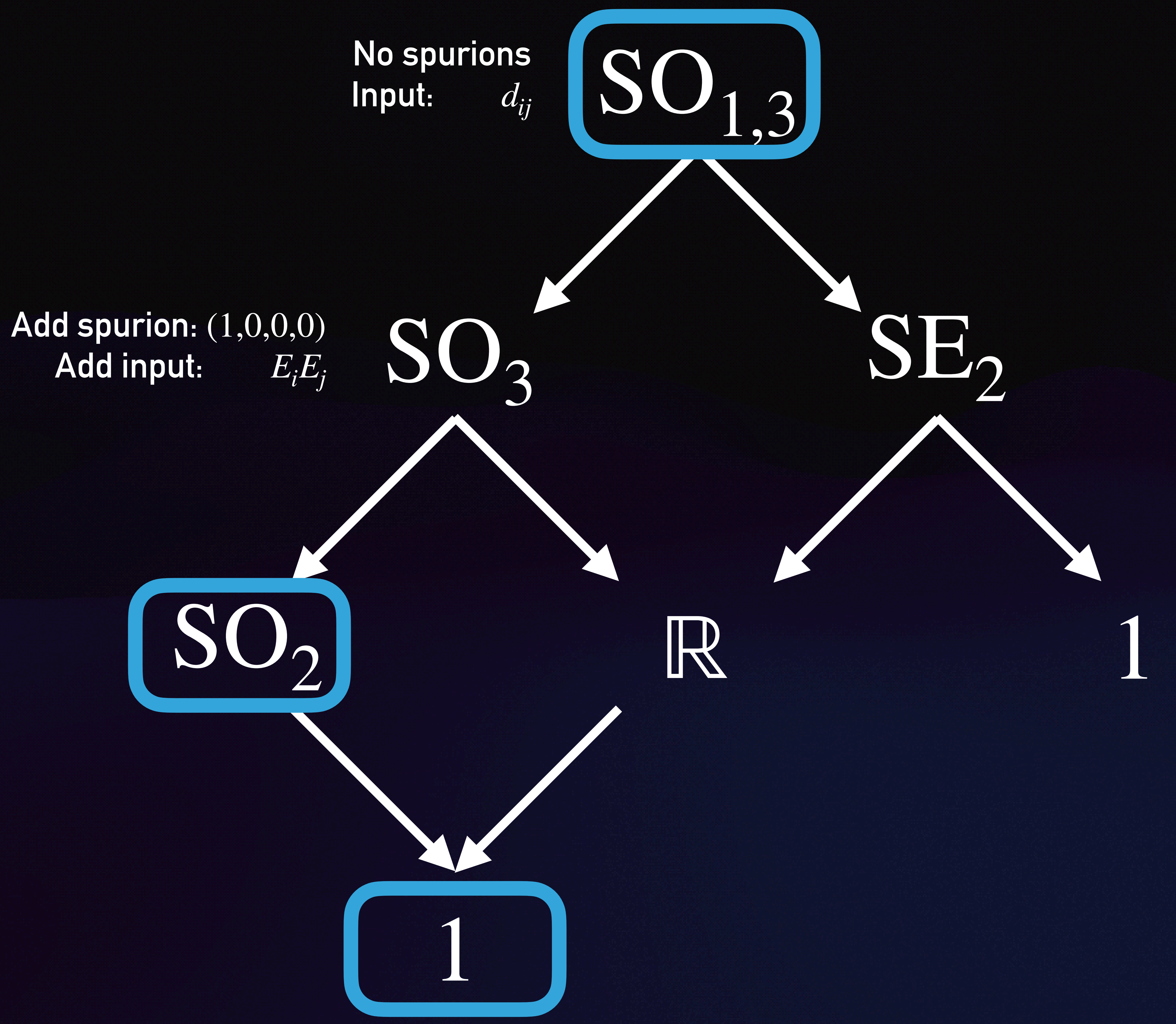


P_{ij} a reduction of (p_i, p_j) to the relevant symmetry group

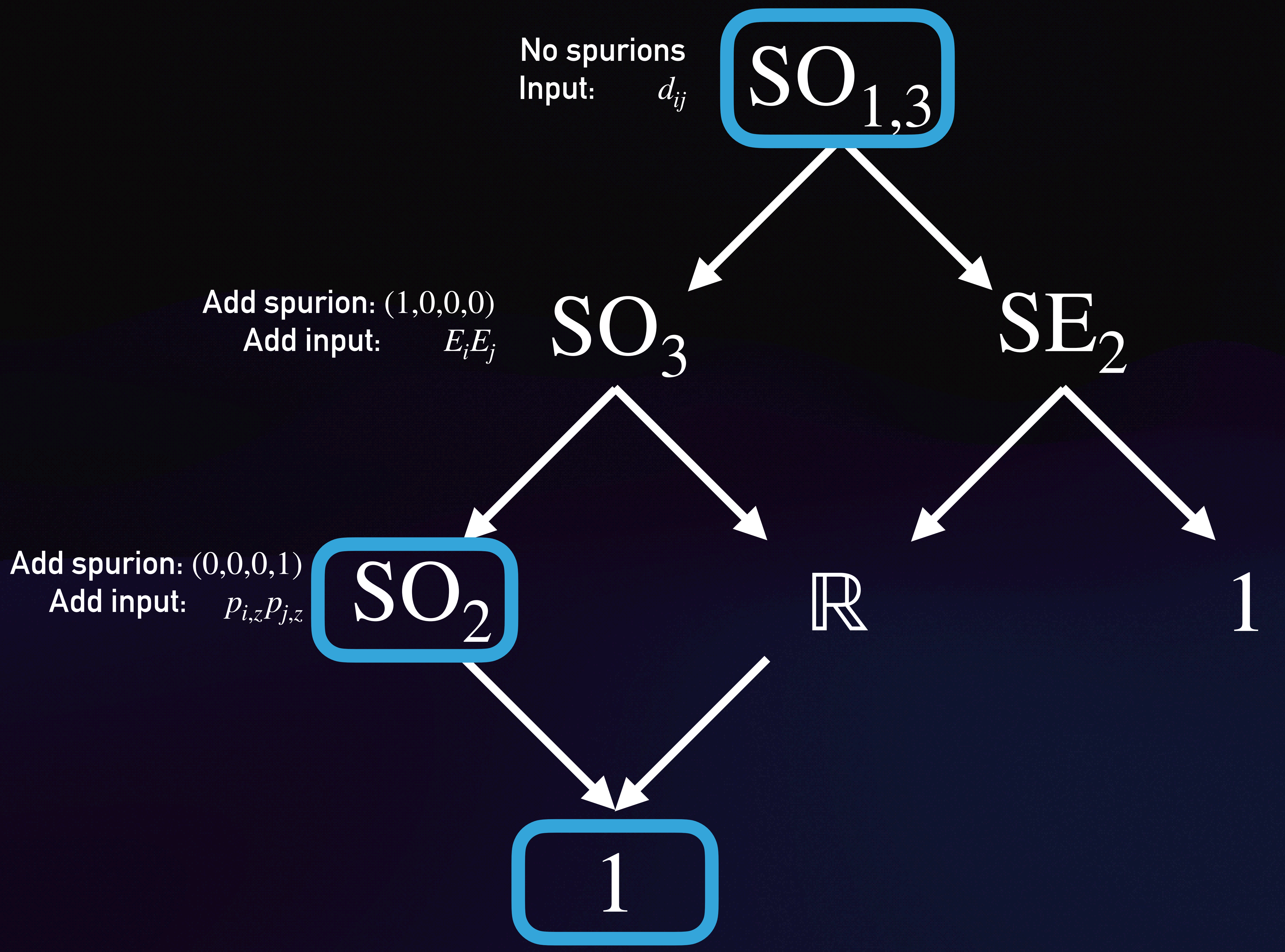
Hierarchy of symmetries



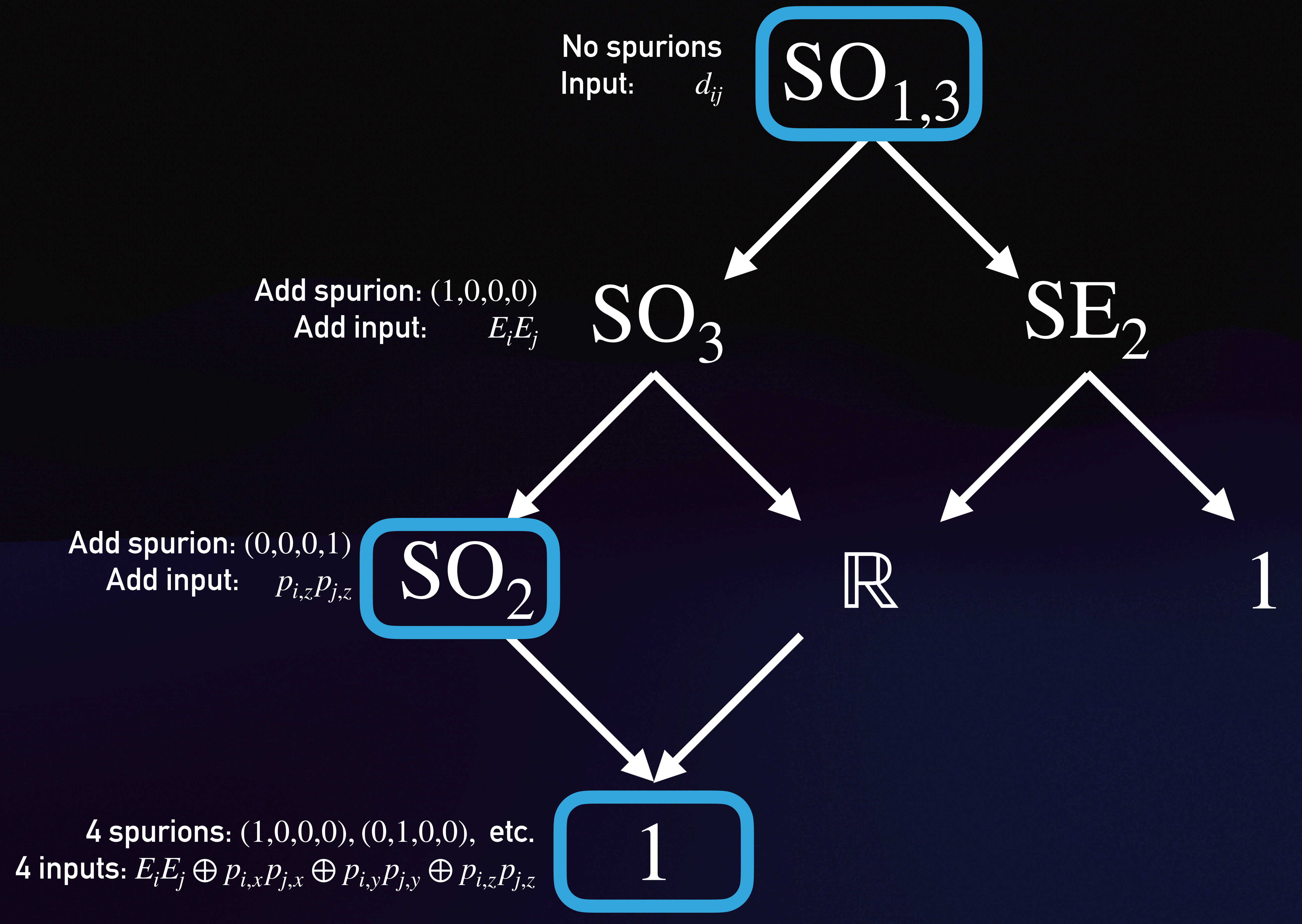
Hierarchy of symmetries



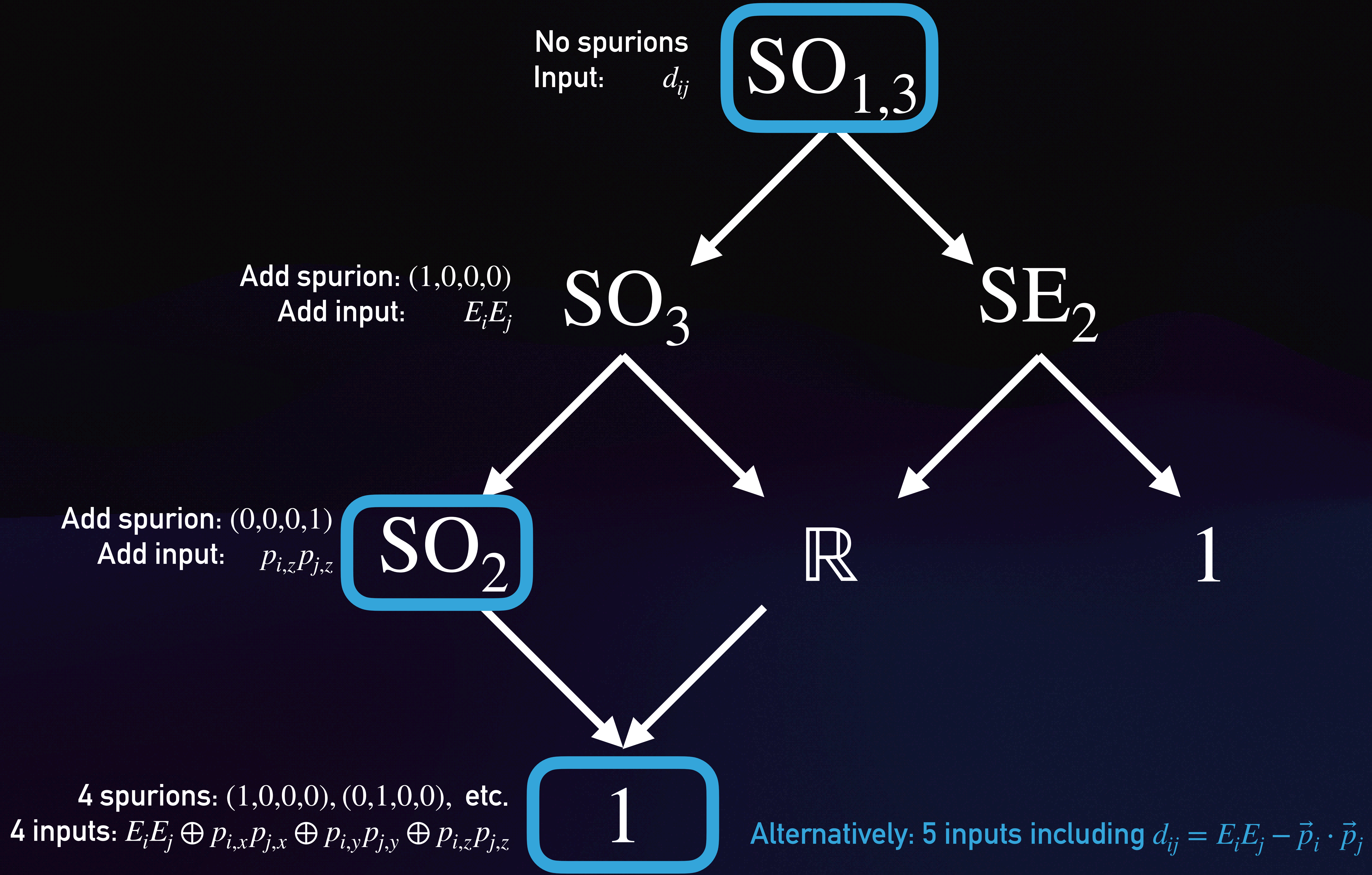
Hierarchy of symmetries



Hierarchy of symmetries



Hierarchy of symmetries



Two methods

Spurion method

k spurions, full Lorentz symmetry

Input method

$k+1$ inputs, partial/no symmetry

Two methods

Spurion method	Input method
k spurions, full Lorentz symmetry	$k+1$ inputs, partial/no symmetry

- Both methods access the same information in the data.

Two methods

Spurion method	Input method
k spurions, full Lorentz symmetry	$k+1$ inputs, partial/no symmetry

- Both methods access the same information in the data.
- Models differ only in the input layer; almost the same model size.

Two methods

Spurion method	Input method
k spurions, full Lorentz symmetry	$k+1$ inputs, partial/no symmetry

- Both methods access the same information in the data.
- Models differ only in the input layer; almost the same model size.
- Fix random seeds: initial weights beyond input layer are identical.

Two methods

Spurion method	Input method
k spurions, full Lorentz symmetry	$k+1$ inputs, partial/no symmetry

- Both methods access the same information in the data.
- Models differ only in the input layer; almost the same model size.
- Fix random seeds: initial weights beyond input layer are identical.
- Input layers are initialized to produce identical outputs.

Two methods

Spurion method	Input method
k spurions, full Lorentz symmetry	$k+1$ inputs, partial/no symmetry

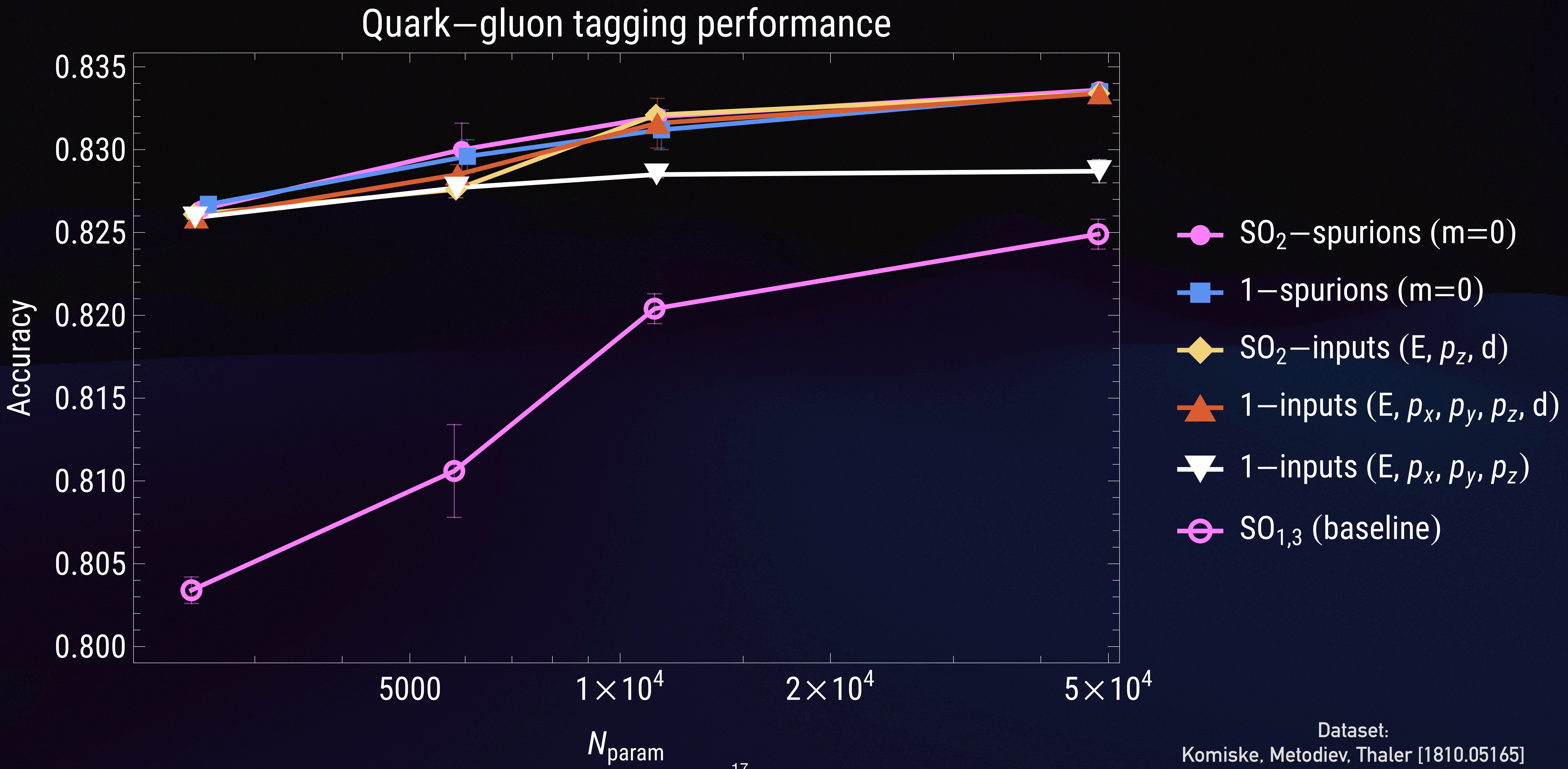
- Both methods access the same information in the data.
- Models differ only in the input layer; almost the same model size.
- Fix random seeds: initial weights beyond input layer are identical.
- Input layers are initialized to produce identical outputs.
- Outputs of all models are identical on the first batch.

Two methods

Spurion method	Input method
k spurions, full Lorentz symmetry	$k+1$ inputs, partial/no symmetry

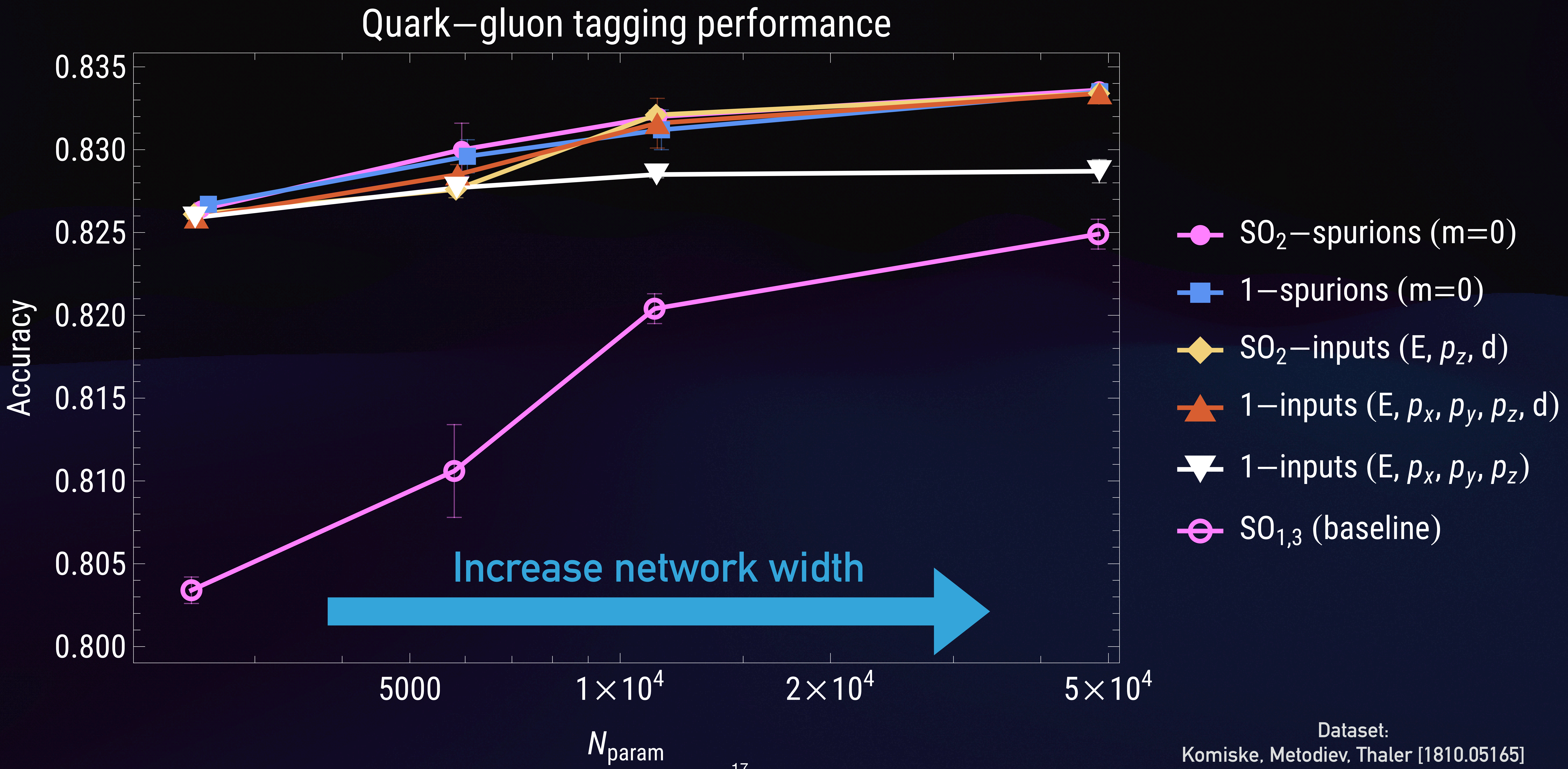
- Both methods access the same information in the data.
- Models differ only in the input layer; almost the same model size.
- Fix random seeds: initial weights beyond input layer are identical.
- Input layers are initialized to produce identical outputs.
- Outputs of all models are identical on the first batch.
- Input method has extra dimensions to optimize in.

Quark/gluon tagging



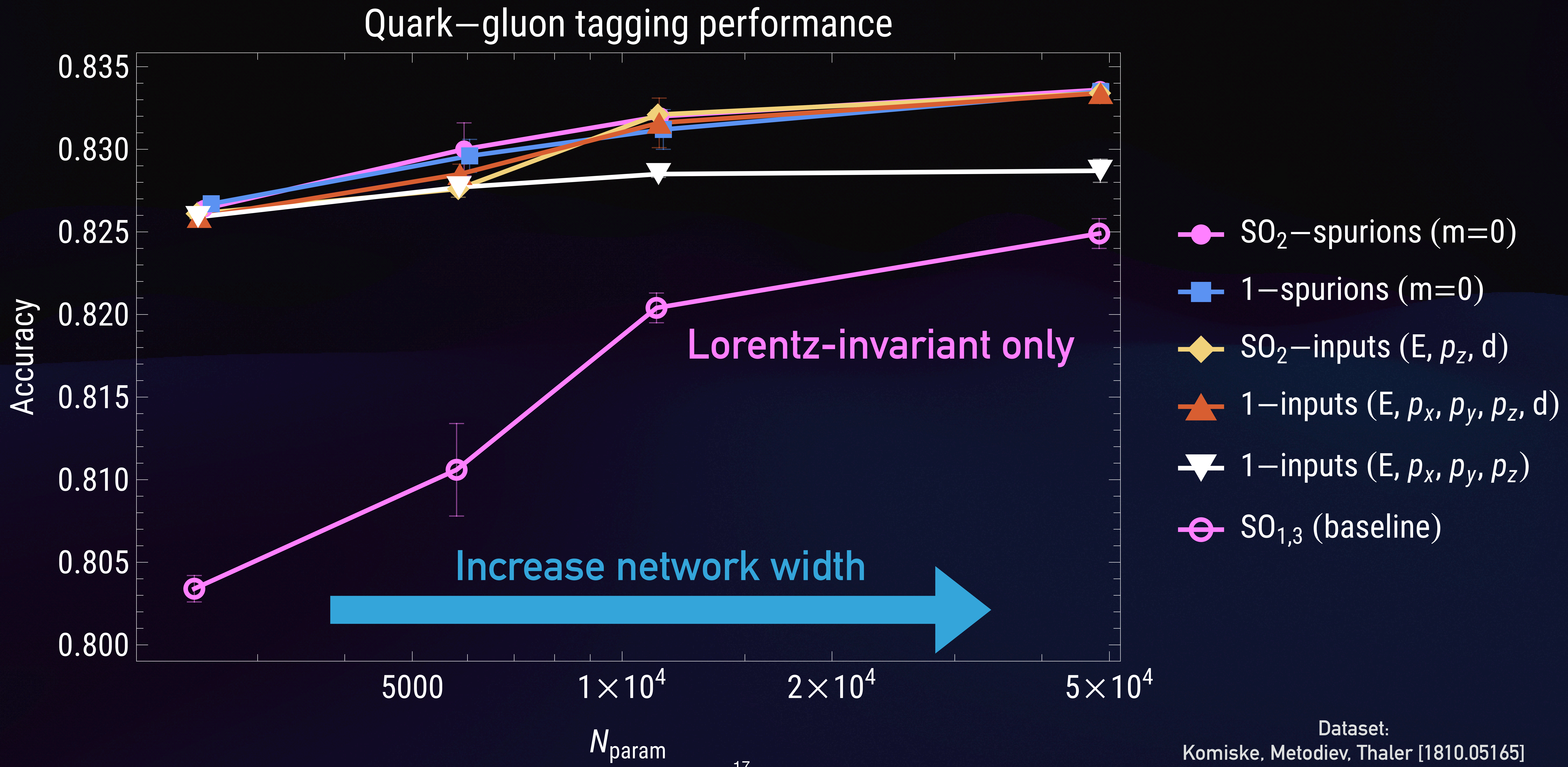
Dataset:
Komiske, Metodiev, Thaler [1810.05165]

Quark/gluon tagging



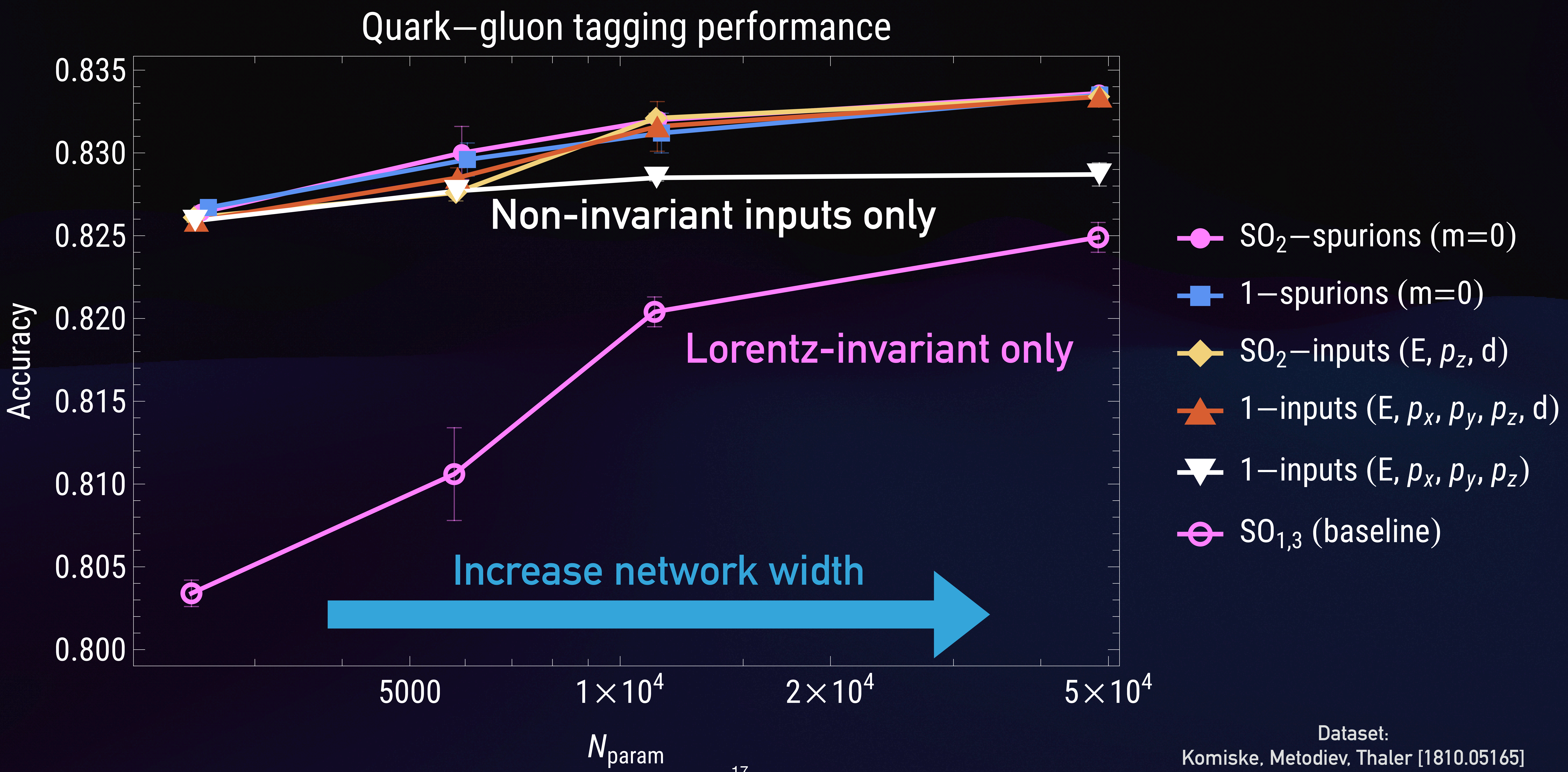
Dataset:
Komiske, Metodiev, Thaler [1810.05165]

Quark/gluon tagging



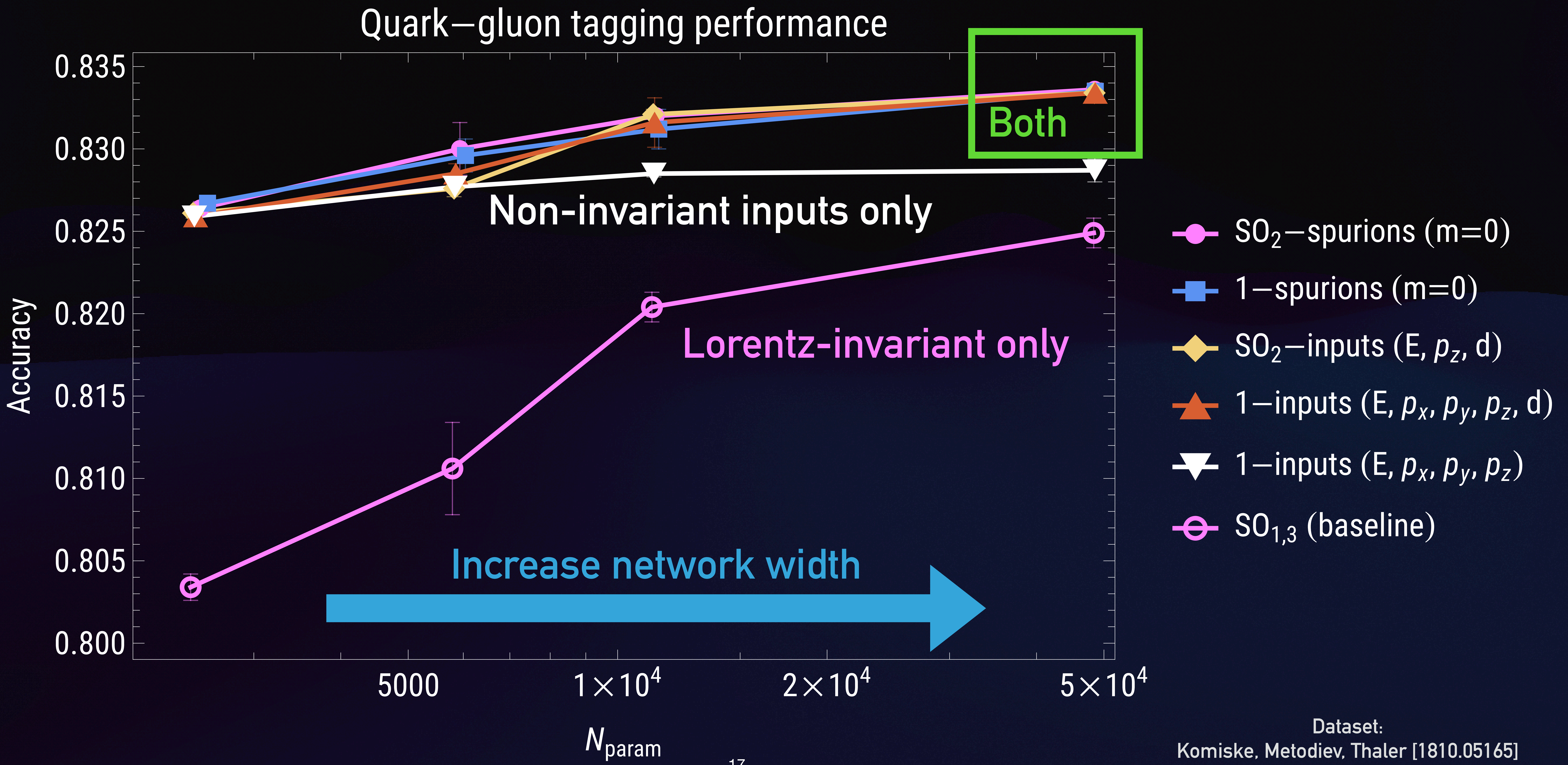
Dataset:
Komiske, Metodiev, Thaler [1810.05165]

Quark/gluon tagging



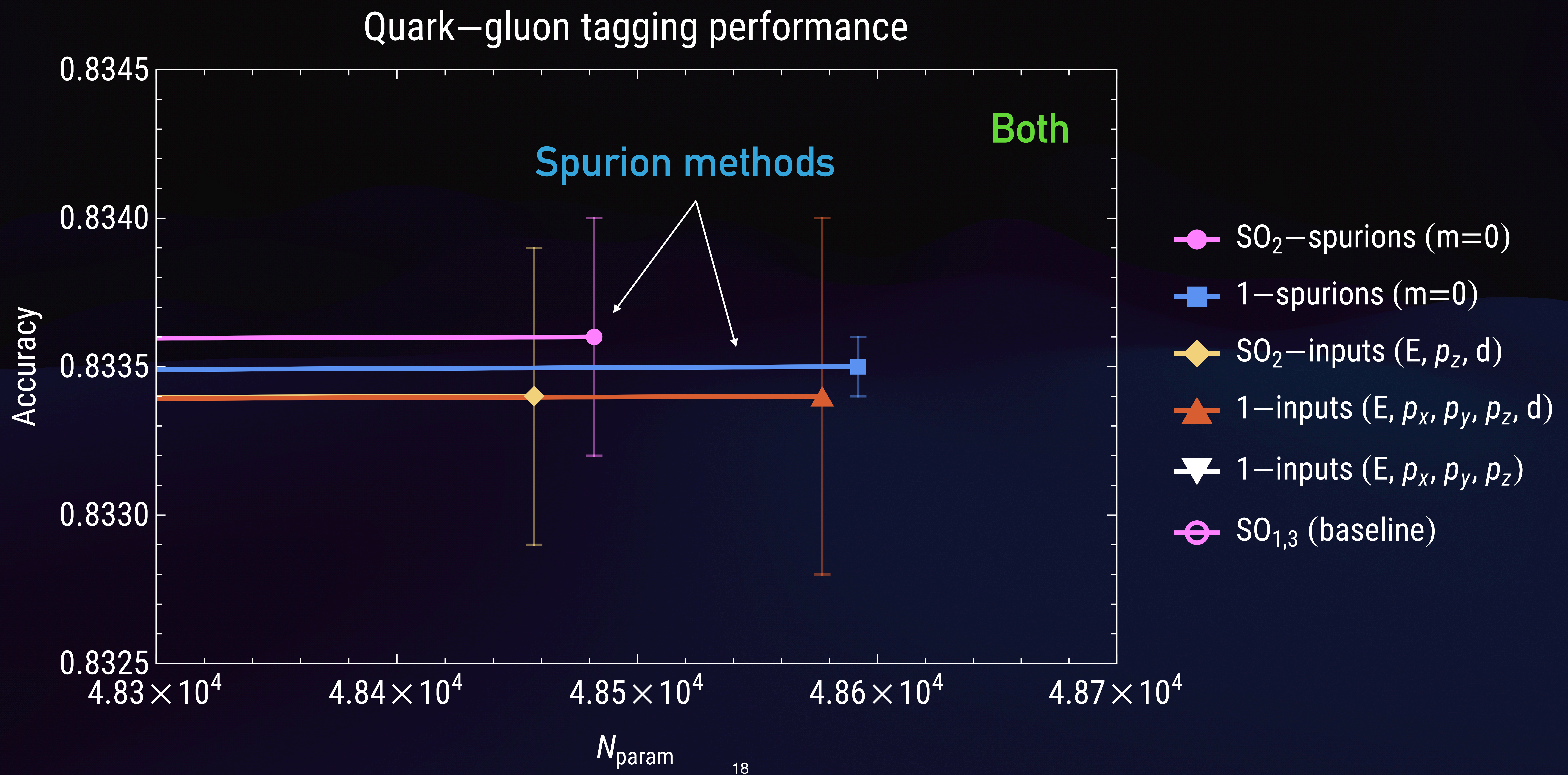
Dataset:
Komiske, Metodiev, Thaler [1810.05165]

Quark/gluon tagging

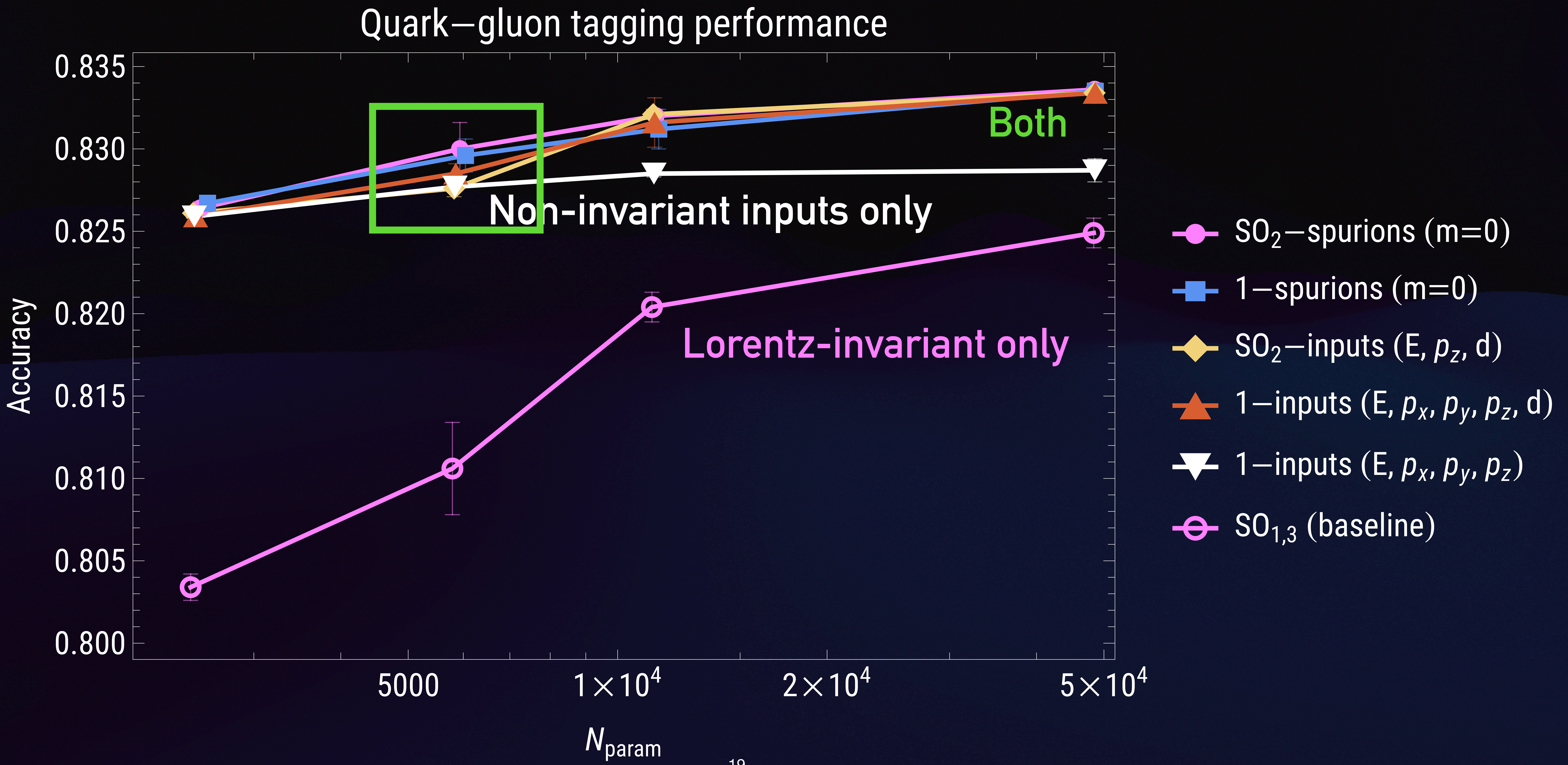


Dataset:
Komiske, Metodiev, Thaler [1810.05165]

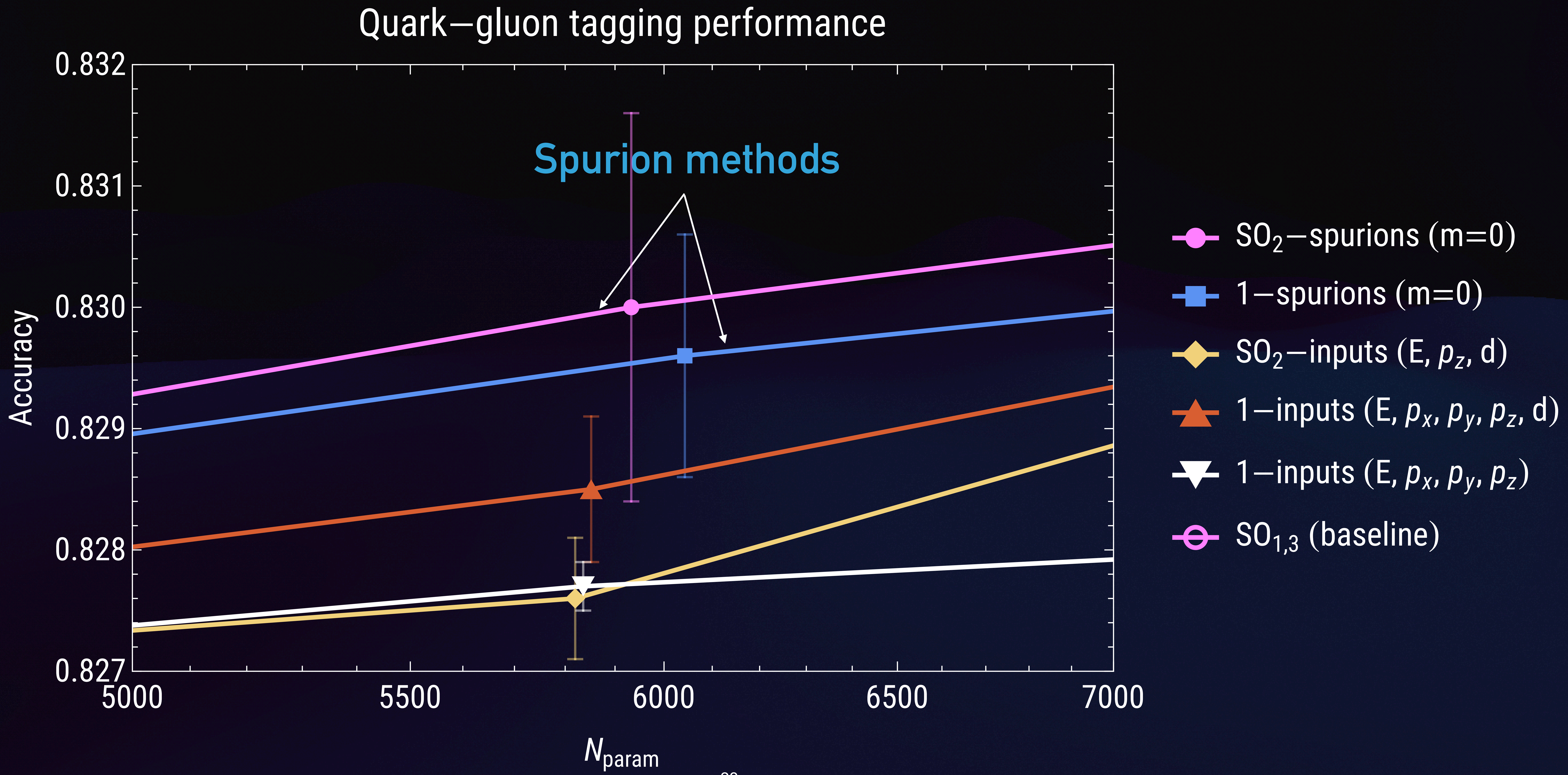
Quark/gluon tagging



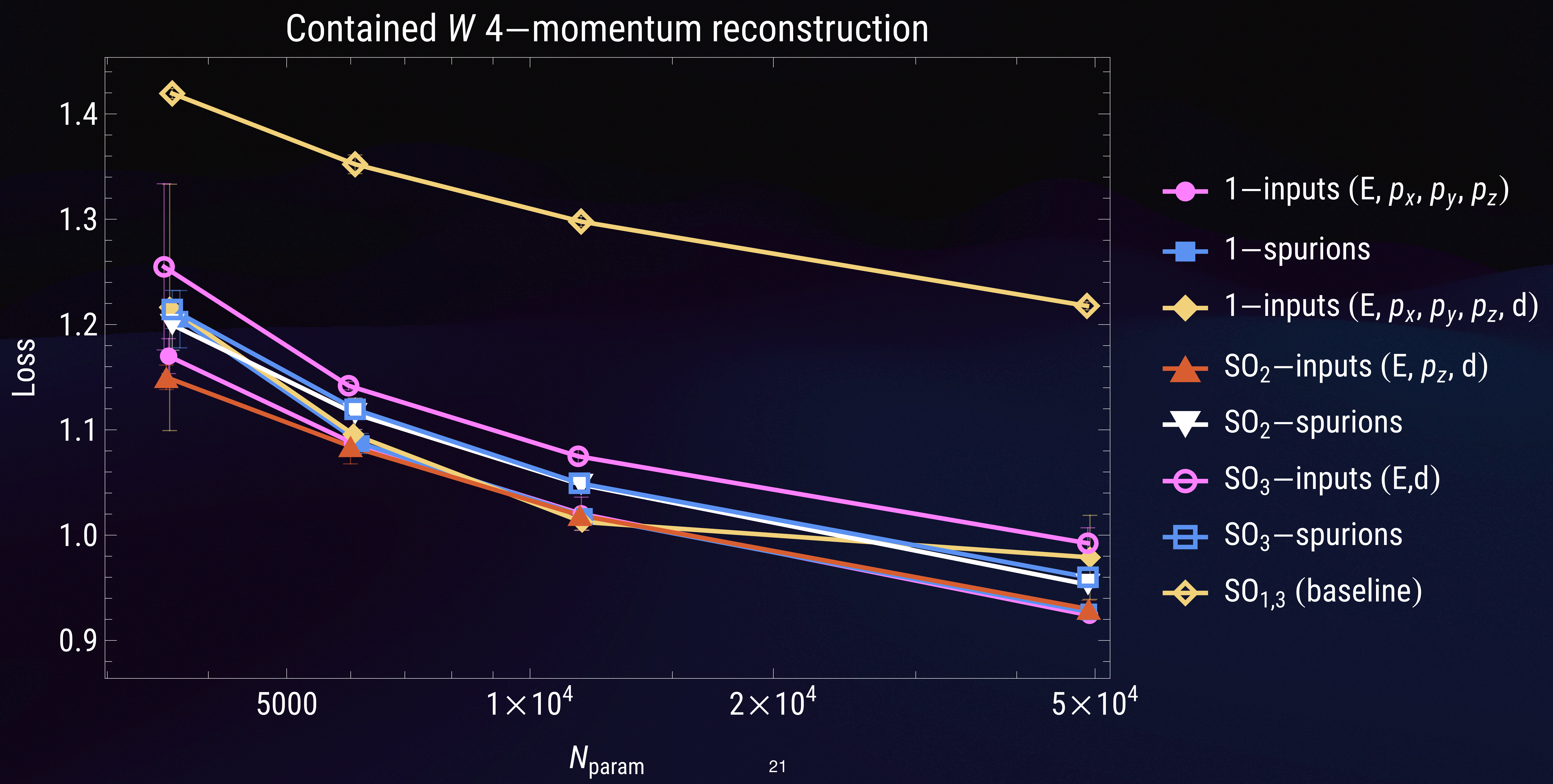
Quark/gluon tagging



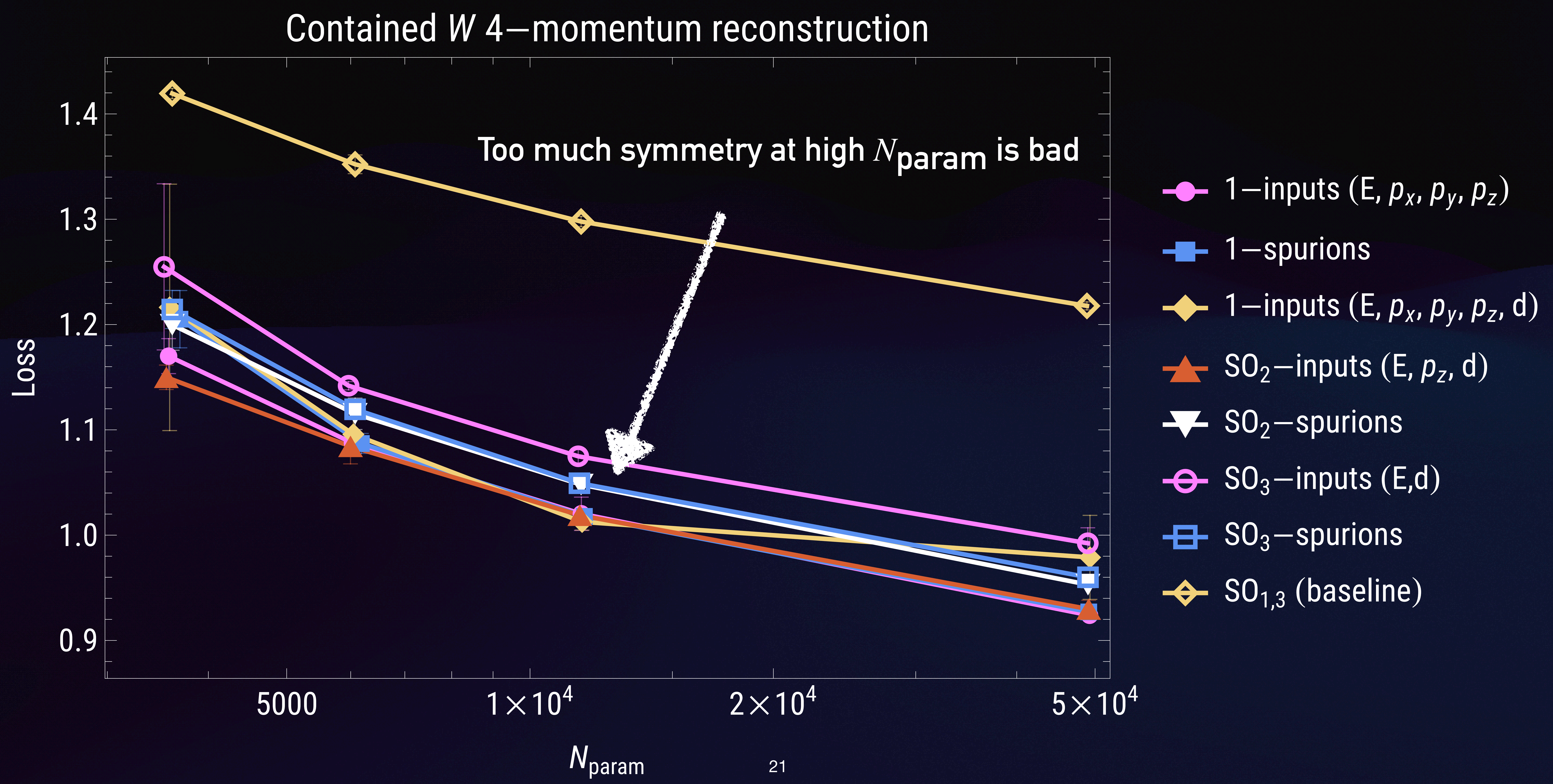
Quark/gluon tagging



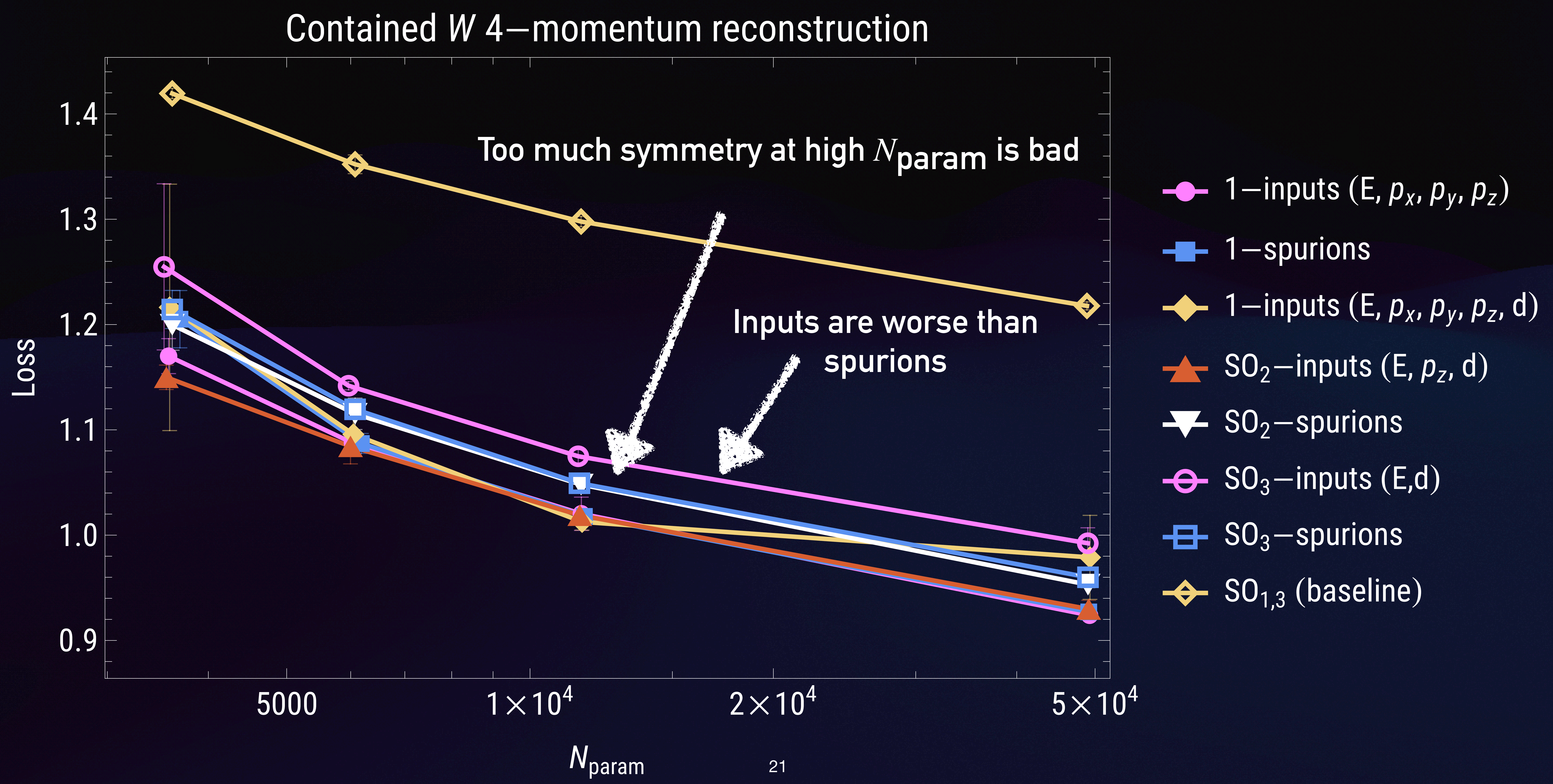
W momentum reconstruction



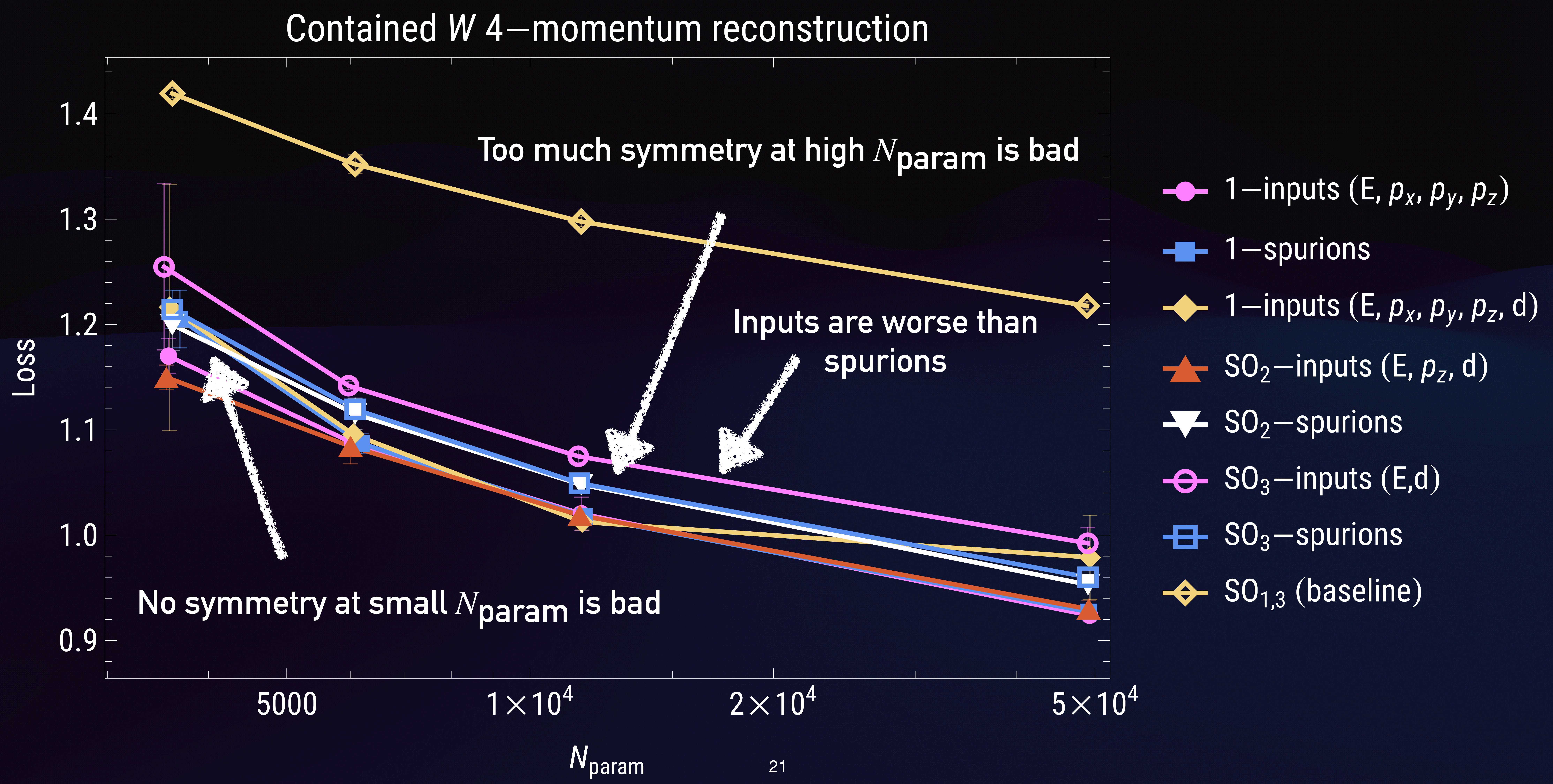
W momentum reconstruction



W momentum reconstruction



W momentum reconstruction



Conclusions

Conclusions

1. Invariant inputs help (consistent with LorentzNet findings).

Conclusions

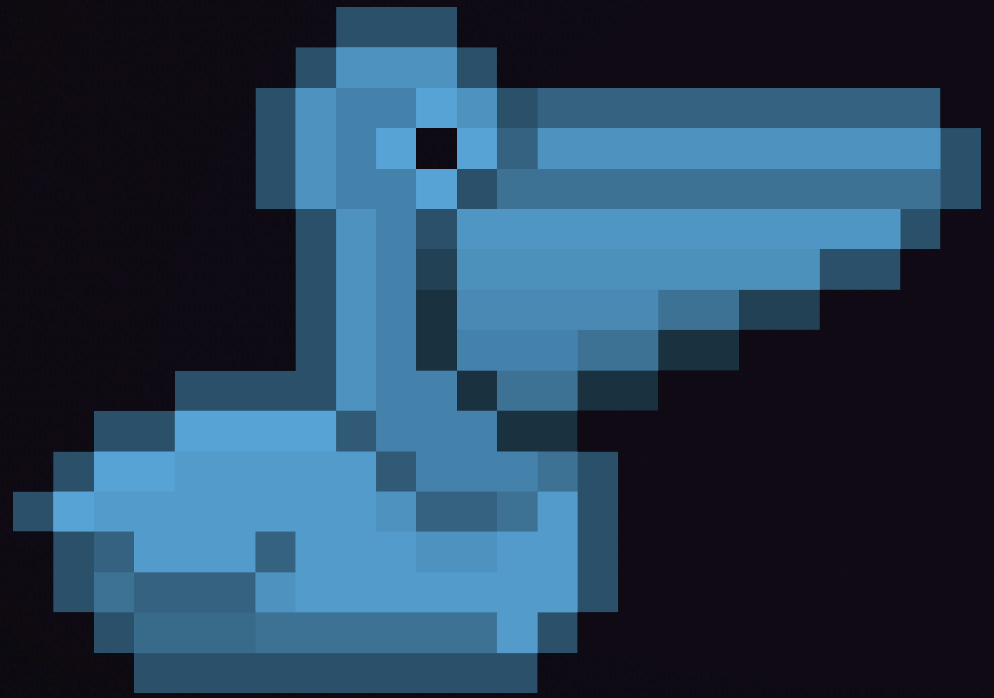
1. Invariant inputs help (consistent with LorentzNet findings).
2. Stronger effect at higher model complexity.

Conclusions

1. Invariant inputs help (consistent with LorentzNet findings).
2. Stronger effect at higher model complexity.
3. Equivariance edges out augmented non-invariant inputs.

Conclusions

1. Invariant inputs help (consistent with LorentzNet findings).
2. Stronger effect at higher model complexity.
3. Equivariance edges out augmented non-invariant inputs.
4. At small model size, more symmetry is better.



arXiv

2211.00454

2307.16506

2310.16121