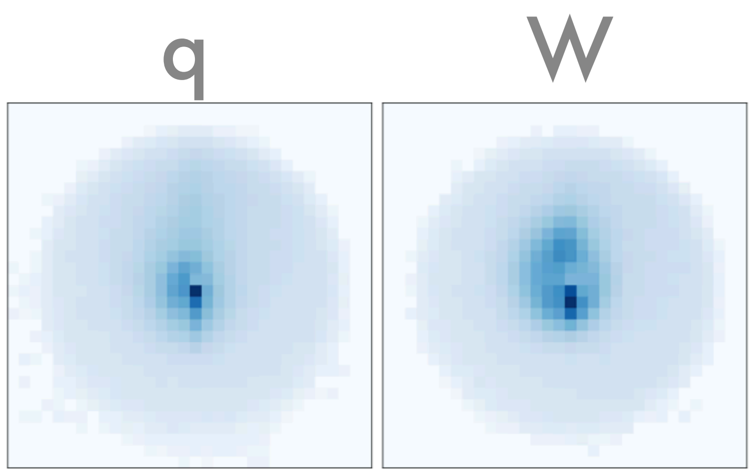
Physics-inspired ML For boson tagging

Daniel Whiteson, UC Irvine

Aug 2023

Motivation

Deep networks find new power for jet substructure



1603.09349

But these are fairly simple objects.

- Can we extend this to many bosons?
- Can we interpret the trained ML?
- Can we incorporate physics knowledge?

2202.00723

- Can we extend this to many bosons?

Resolving Extreme Jet Substructure

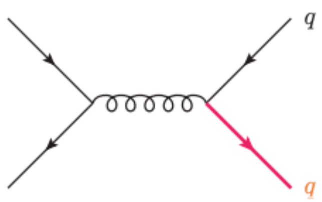
Yadong Lu[a] **Alexis Romero**[b] **Michael James Fenton**[b] **Daniel Whiteson**[b] **Pierre Baldi**[c]

[a] *Department of Statistics, University of California, Irvine, CA, USA 92627*
[b] *Department of Physics and Astronomy, University of California, Irvine, CA, USA 92627*
[c] *Department of Computer Science, University of California, Irvine, CA, USA 92627*

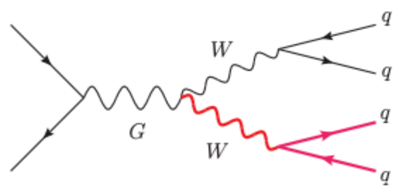
E-mail: yadongl1@uci.edu, alexir2@uci.edu, m.fenton@uci.edu, daniel@uci.edu, pfbaldi@uci.edu

- Fat jets with up to eight hard subjets!

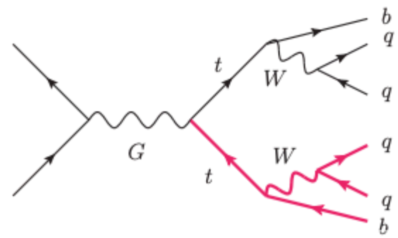
Higgses and Ws



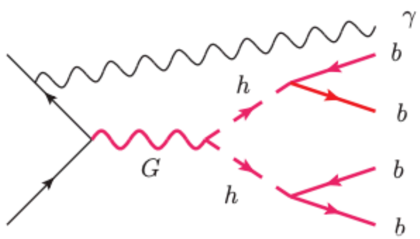
(a) $N = 1$



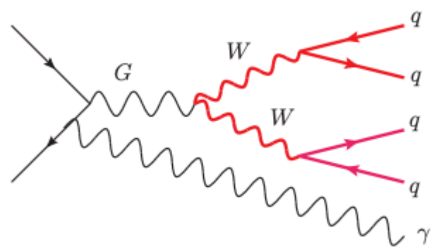
(b) $N = 2$



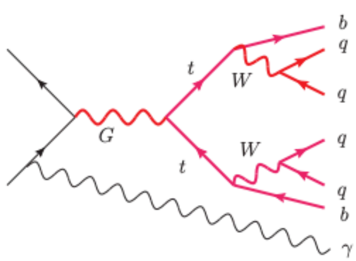
(c) $N = 3$



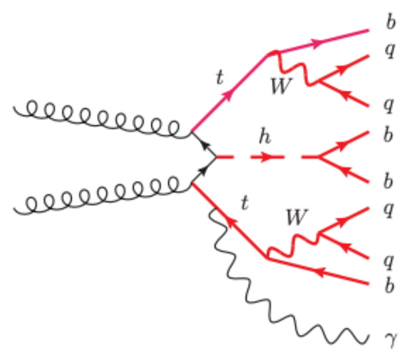
(d) $N = 4b$



(e) $N = 4q$



(f) $N = 6$



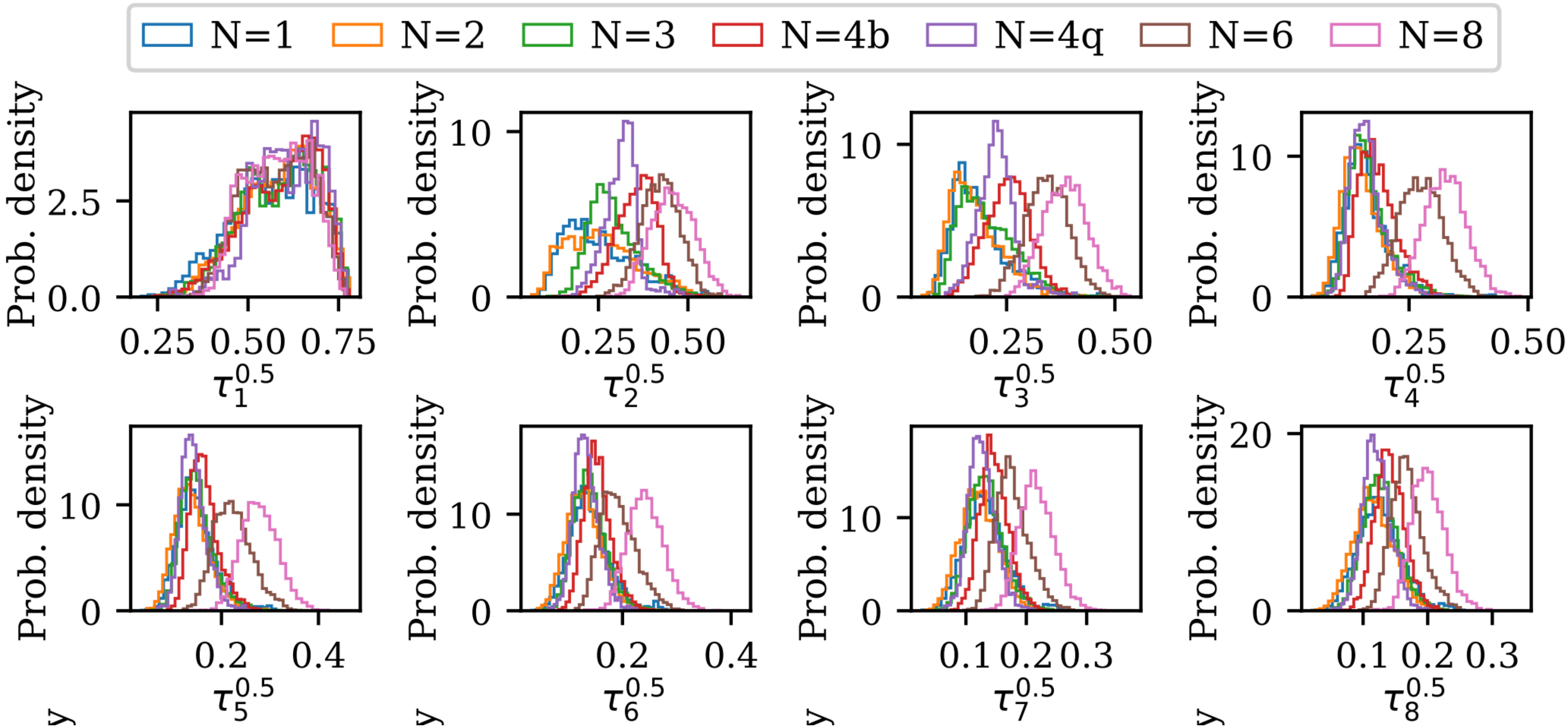
(g) $N = 8$

<u>Jets in range</u>

p_T=[1.0,1.2] TeV

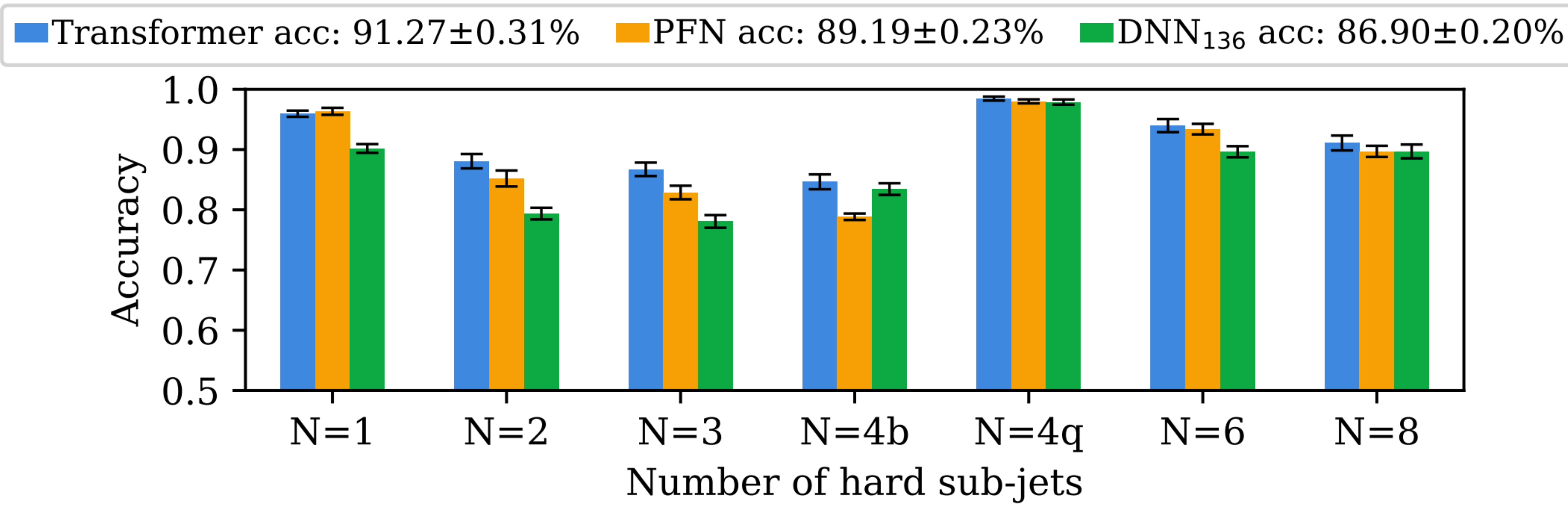
Mass=300-700 GeV

Physics observables



And many (136 total) more!

Performance

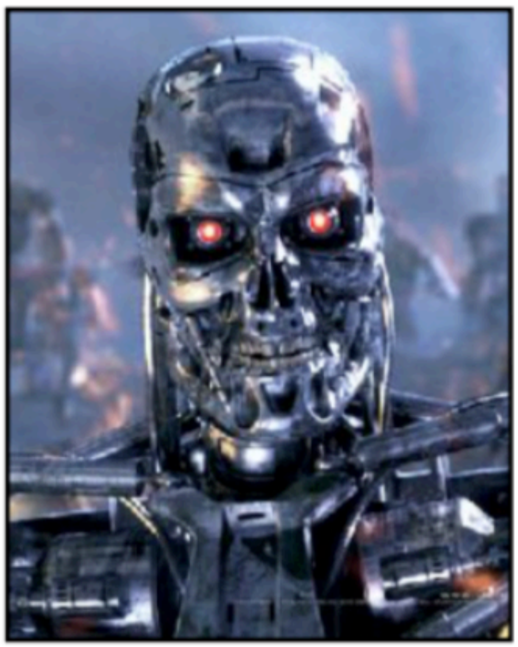


Lower level beats higher level

Gap smaller for higher N

What has it learned?

The Machines



How to interpret?

Mapping Machine-Learned Physics into a Human-Readable Space

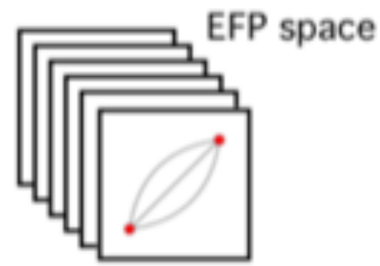
Taylor Faucett,[1] Jesse Thaler,[2,3] and Daniel Whiteson[1]

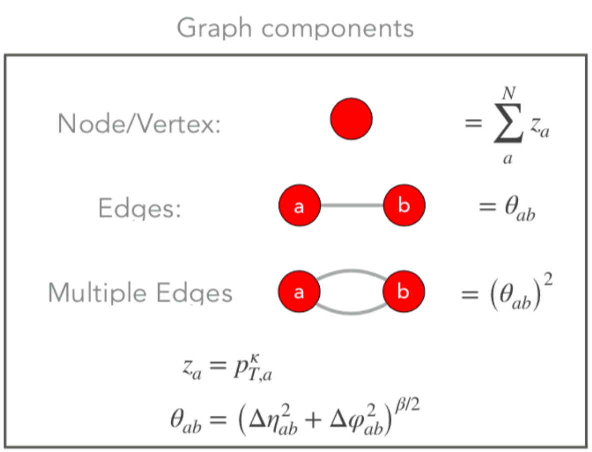
https://arxiv.org/abs/2010.11998

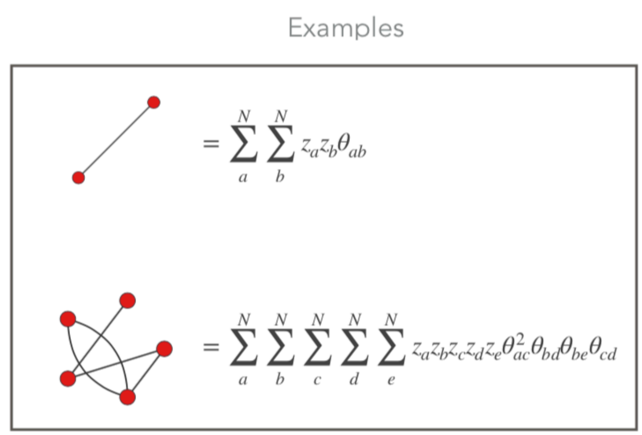
How?

<u>I. Define space of interpretable observables</u>

- provides context
- defines problem
- does NN live in this space?
- Can it be compactly represented?
- Yes or No are both interesting!







How?

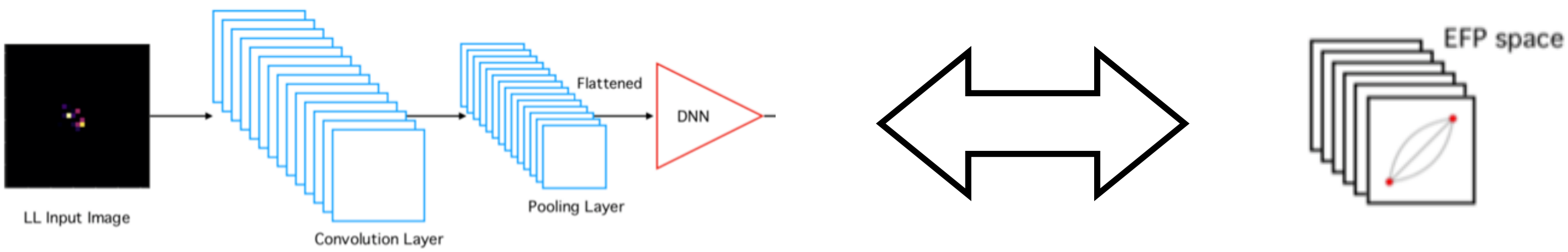
<u>I. Define space of interpretable observables</u>

- provides context
- defines problem
- does NN live in this space?
- Can it be compactly represented?
- Yes or No are both interesting!

<u>II. Define mapping metric</u>

- how do you compare two solutions?
- can’t use functional identity or linear correlation

Mapping



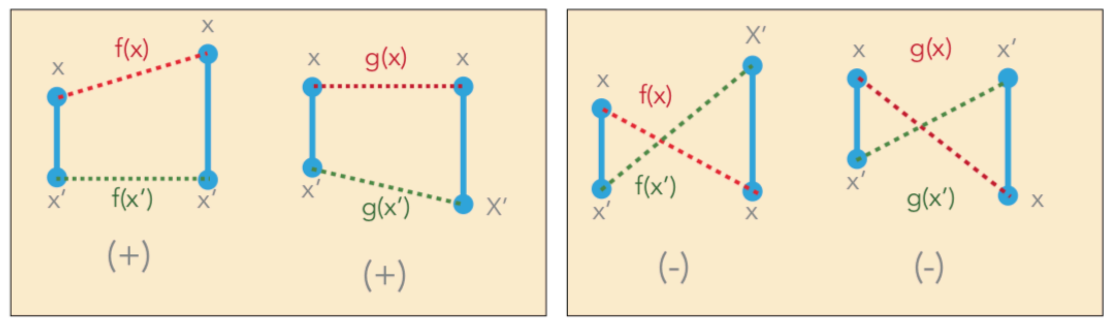
Function sameness

Complete equivalence
not the idea

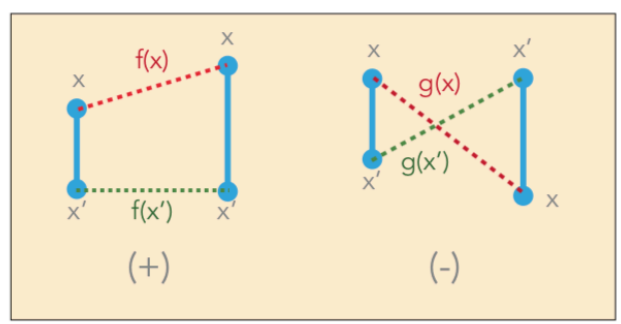
Any 1:1 transformation
of function has no impact
in our context

Only care about the
ordering of points
not the actual function
values

Similar Orderings



Dissimilar Orderings



Muons

Learning to Isolate Muons 2102.02278

Julian Collado,[1] Kevin Bauer,[2] Edmund Witkowski,[2] Taylor Faucett,[2] Daniel Whiteson,[2] and Pierre Baldi[1]

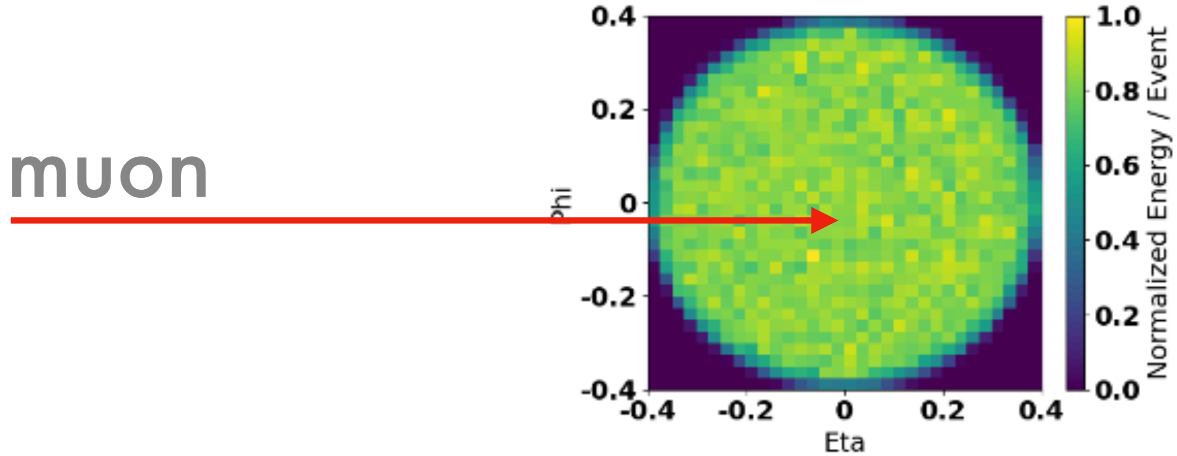
[1]*Department of Computer Science, University of California, Irvine, CA, 92697*
[2]*Department of Physics and Astronomy, University of California, Irvine, CA 92697*
(Dated: February 2, 2021)

Learning to Isolate Muons in Data 2306.15737

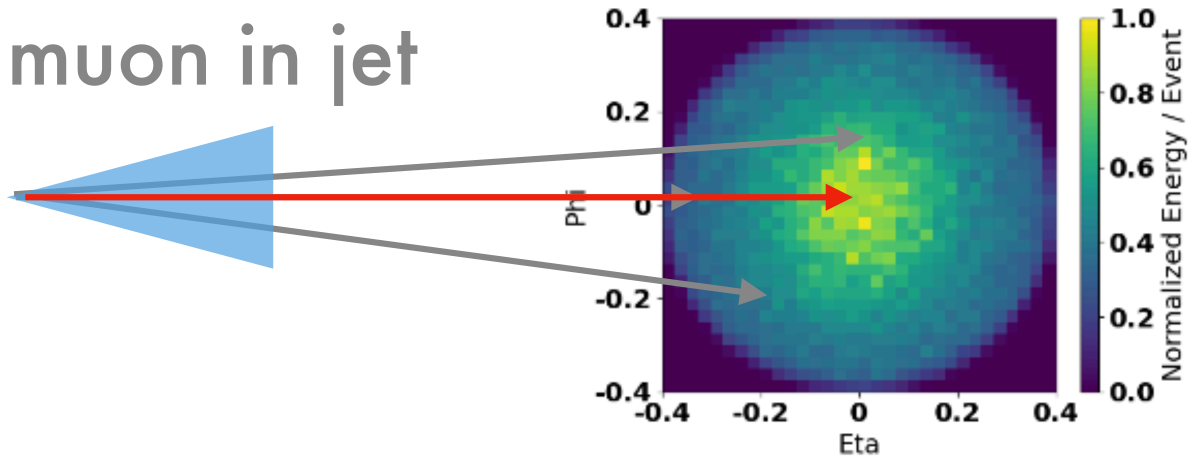
Edmund Witkowski,[1,*] Benjamin Nachman,[2,3,†] and Daniel Whiteson[1,‡]

[1]*Department of Physics and Astronomy, University of California, Irvine, CA 92697*
[2]*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[3]*Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*
(Dated: June 29, 2023)

Muons

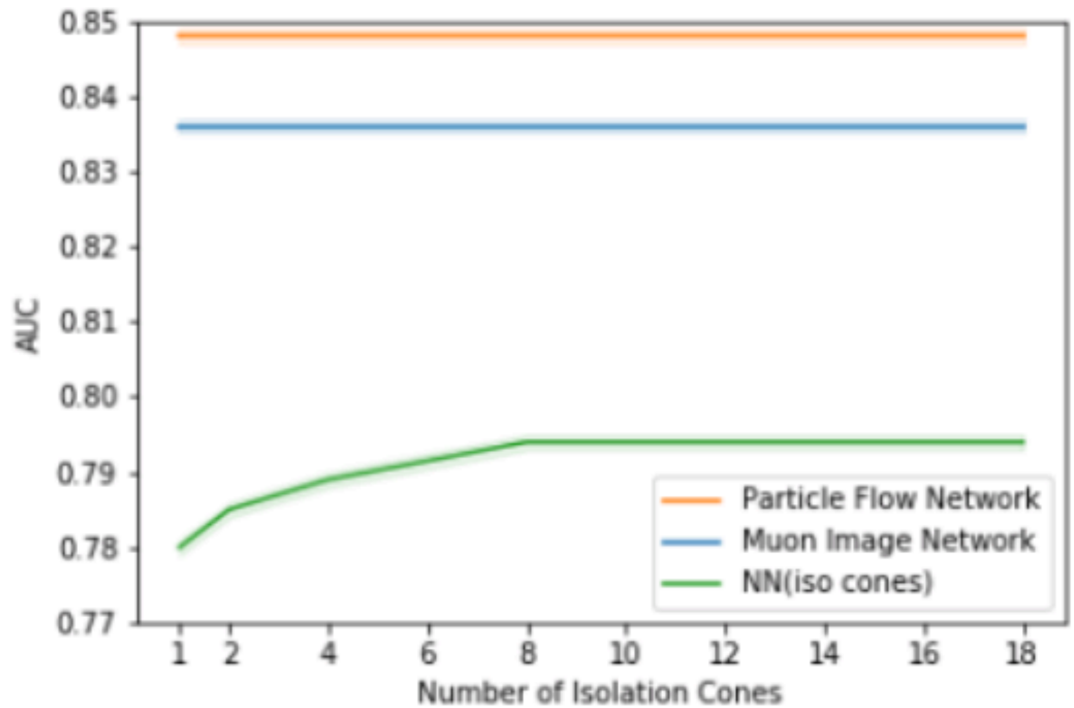


(a) Mean Prompt Muon



(b) Mean Non-prompt Muon

Results



Calo-cell networks

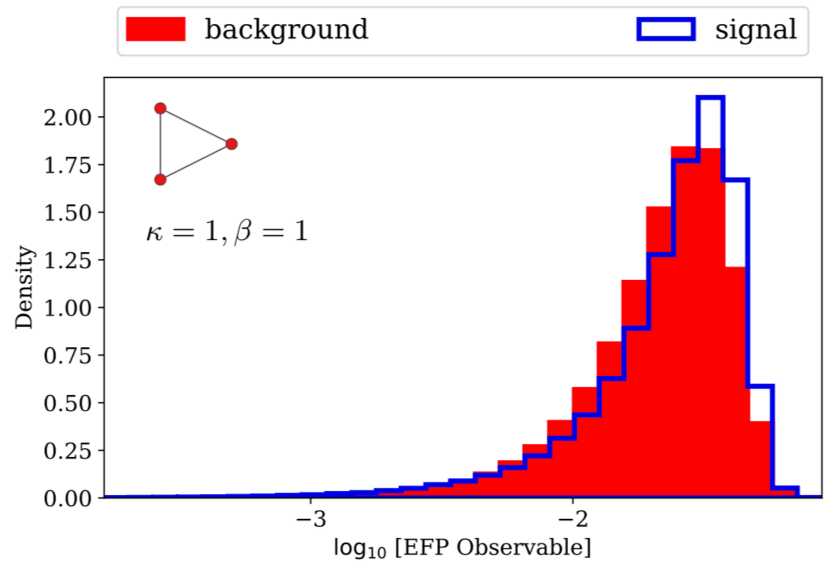
Iso networks

More iso cones improves performance
Isolation cannot match calo-cell networks

Useful observable

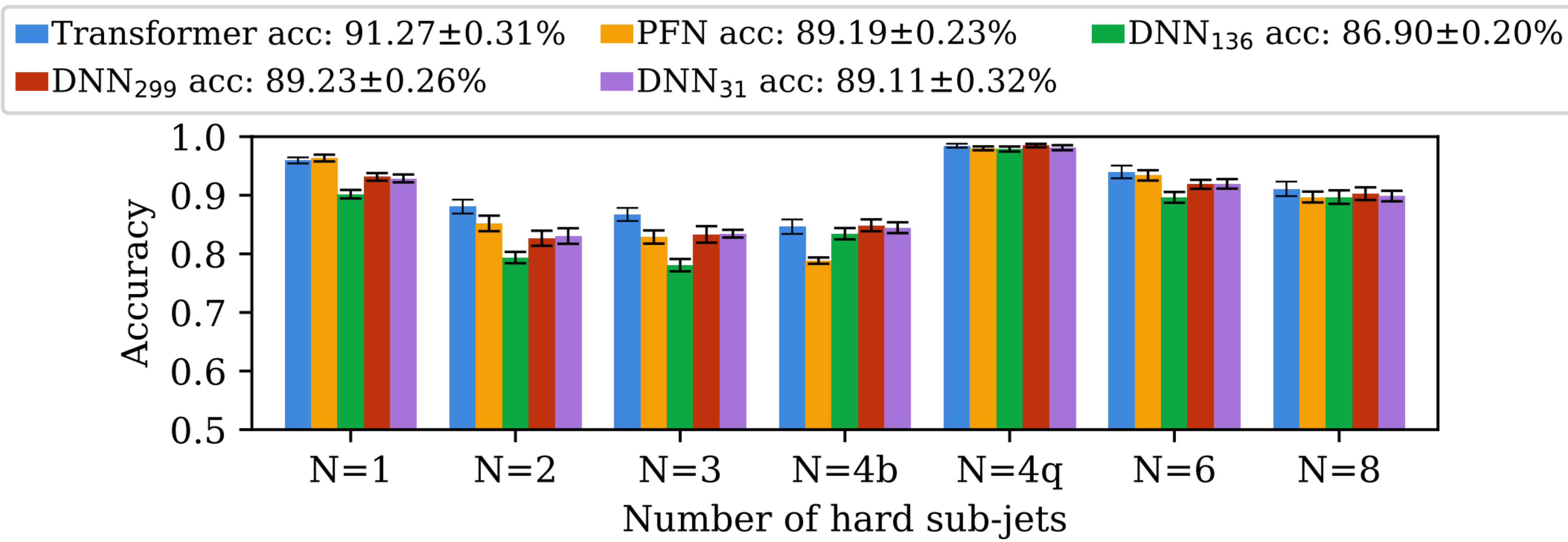
This observable helps!

$$= \sum_{a,b,c=1}^{N} z_a z_b z_c \theta_{ab} \theta_{bc} \theta_{ca}$$



Back to bosons

1. Build a lot of EFPs
2. Add to our high-level set (299)
3. Prune the ones we don't need (31)

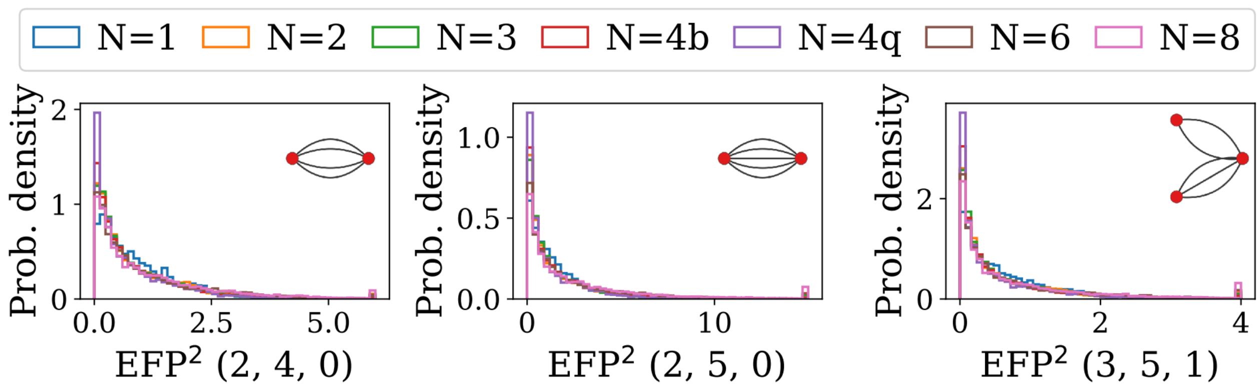


Full set helps close the gap.
Pruned set maintains performance.

What does it mean?

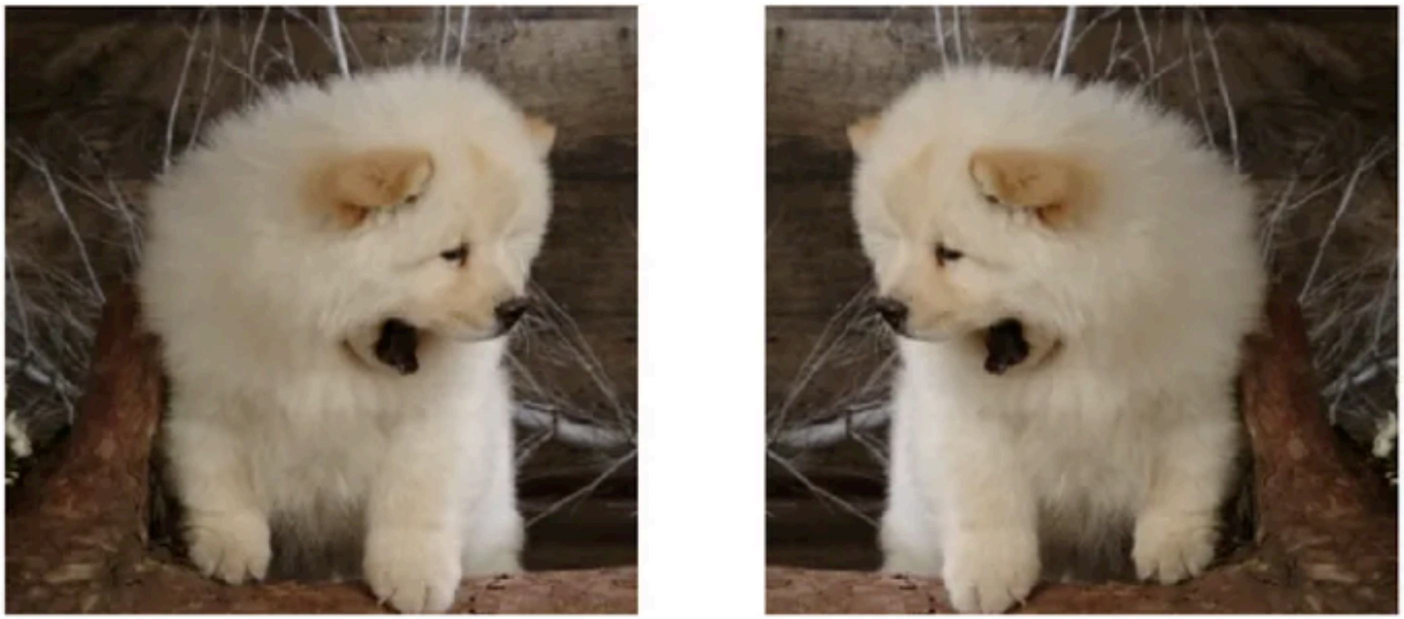
Rank them by importance

Rank	Observable	Accuracy drop
1	EFP^2 (2, 4, 0)	$50.60 \pm 1.95\%$
2	EFP^2 (2, 5, 0)	$44.81 \pm 2.28\%$
3	EFP^2 (3, 5, 1)	$41.68 \pm 2.35\%$
4	τ_1^1	$41.49 \pm 1.06\%$
5	EFP^1 (2, 4, 0)	$38.81 \pm 1.41\%$
6	$\mathrm{EFP}^{0.5}$ (4, 3, 1)	$37.99 \pm 1.02\%$
7	$\mathrm{EFP}^{0.5}$ (2, 2, 0)	$37.26 \pm 1.38\%$
8	EFP^2 (4, 5, 0)	$35.37 \pm 0.70\%$
9	τ_2^1	$34.97 \pm 0.63\%$
10	$\mathrm{EFP}^{0.5}$ (2, 5, 0)	$33.66 \pm 2.02\%$
11	τ_3^1	$30.26 \pm 1.08\%$
12	Norm. Jet Mass	$29.44 \pm 0.92\%$
13	$\tau_2^{0.5}$.	$29.38 \pm 0.90\%$
14	EFP^2 (4, 5, 2)	$27.66 \pm 1.66\%$
15	EFP^1 (3, 5, 3)	$27.58 \pm 0.88\%$
16	τ_4^1	$26.74 \pm 0.96\%$



Add physics in advance

Take advantage of symmetries



These are the same dog.
Should have same representation,
network output

Add physics in advance

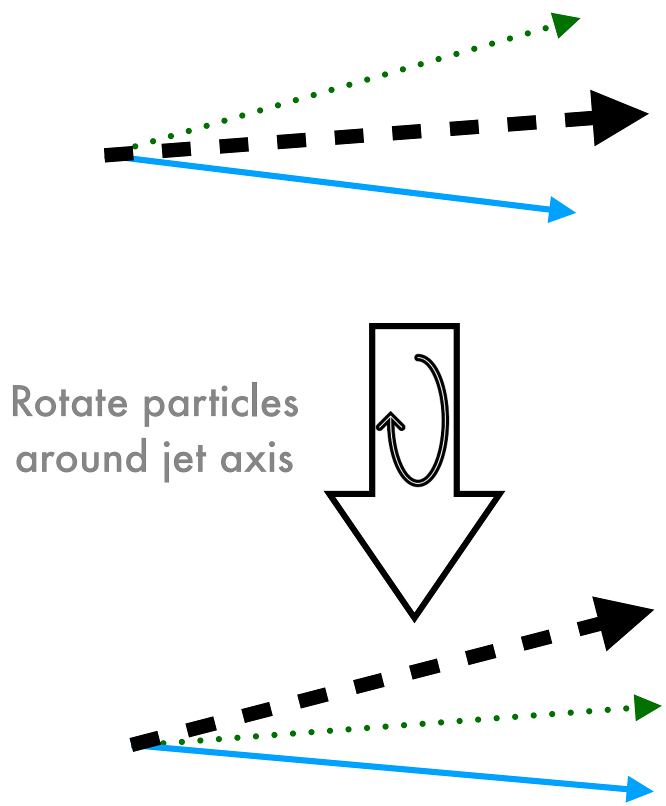
Take advantage of symmetries

<u>Learn with less data</u>

1) Build symmetry constraints into networks, or
2) Augment dataset with symmetric copies

Rotated jets

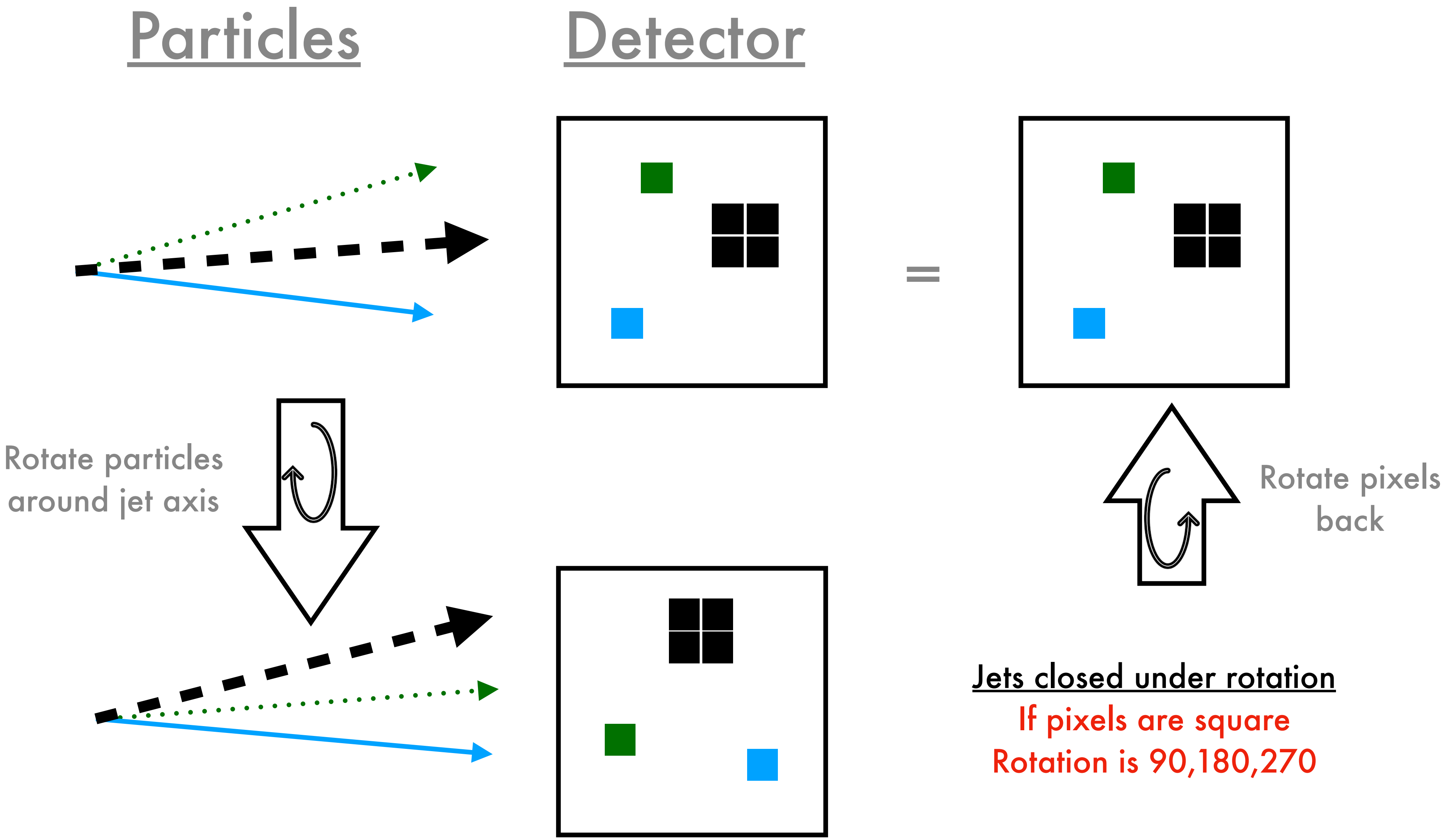
Particles



Jet rotation around axis has no effect on physics.

These are the same jet.
Should have same representation, network output

Rotated jets

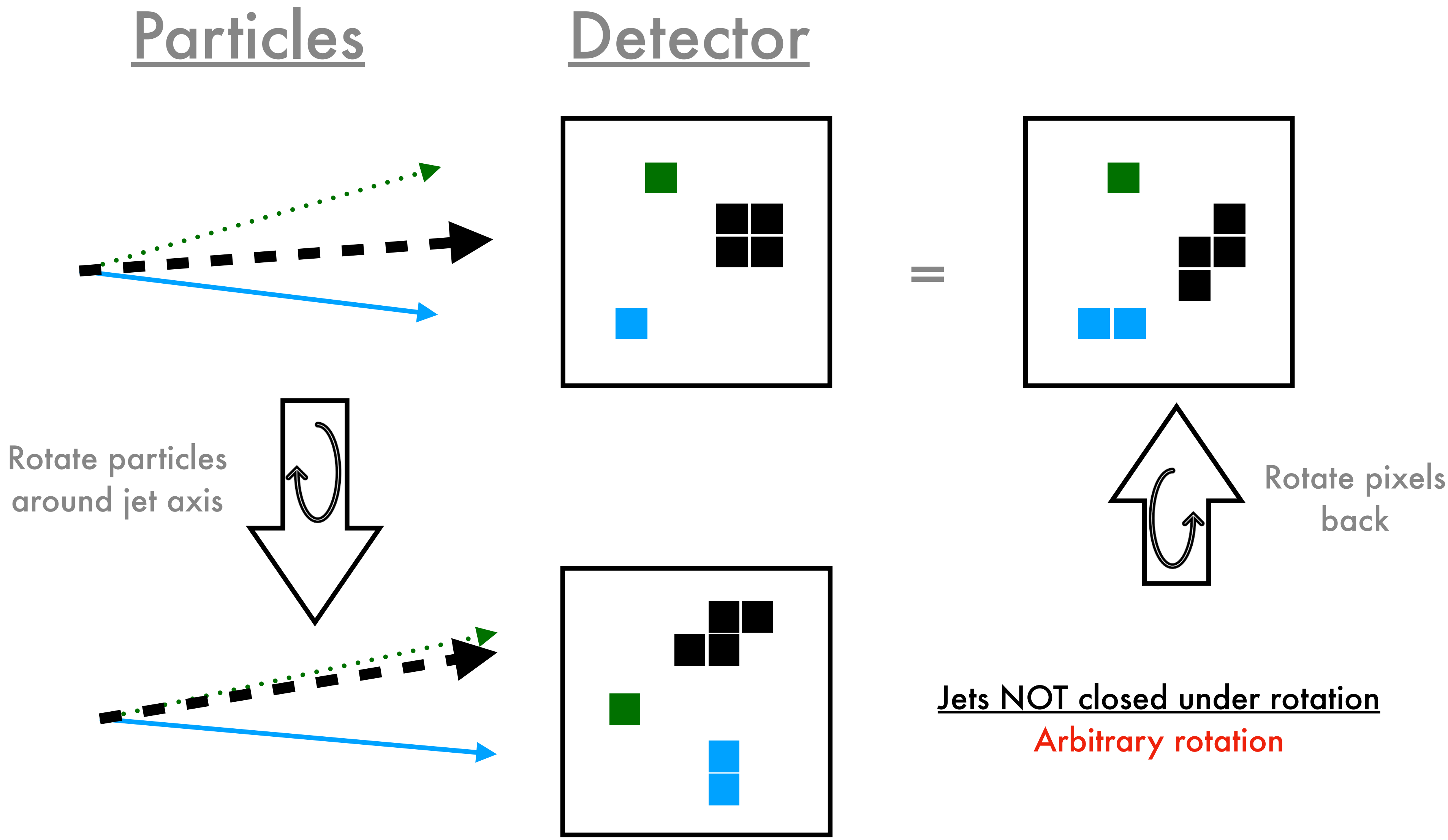


<u>Jets closed under rotation</u>

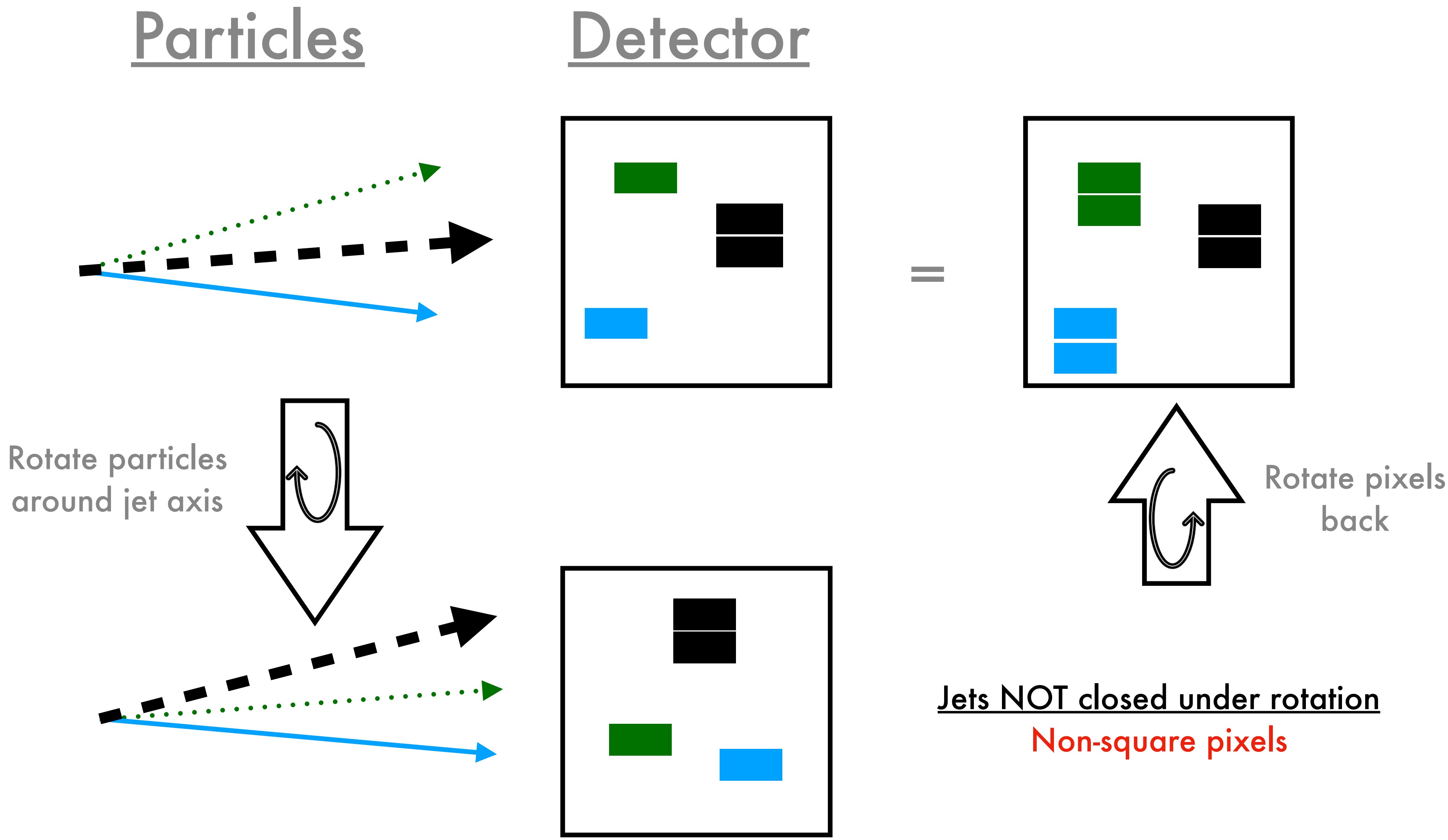
If pixels are square

Rotation is 90,180,270

Rotated jets



Rotated jets



Symmetry

The true, hidden symmetry is pre-detector

The detector breaks the symmetry

We cannot learn the symmetry from augmented post-detector data

We should not impose the symmetry on post-detector data

Upcoming paper

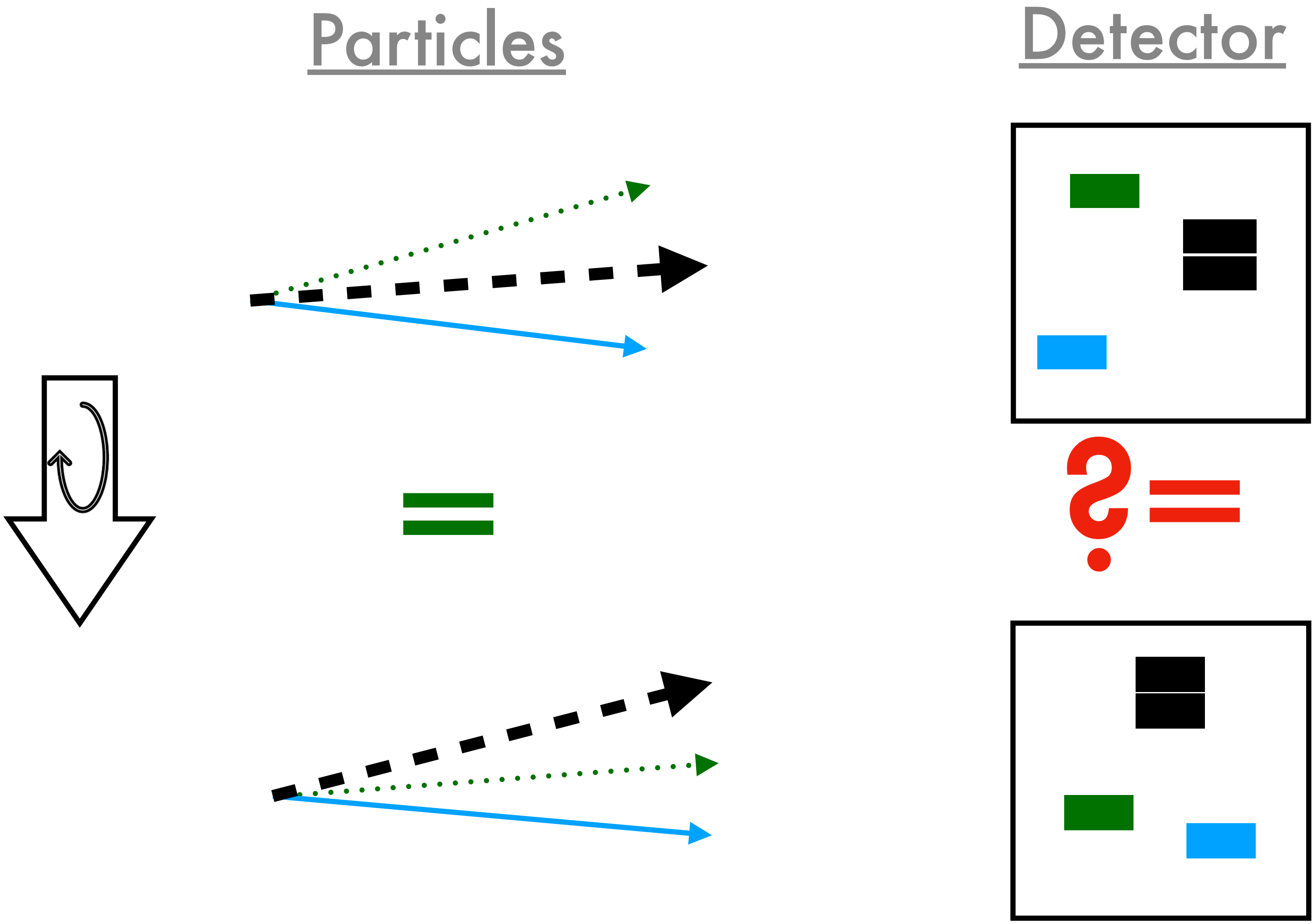
Learning with Hidden Symmetries

Edmund Witkowski[1] and Daniel Whiteson[1]

[1] *Department of Physics and Astronomy, UniversityofCalifornia, Irvine, CA92697*

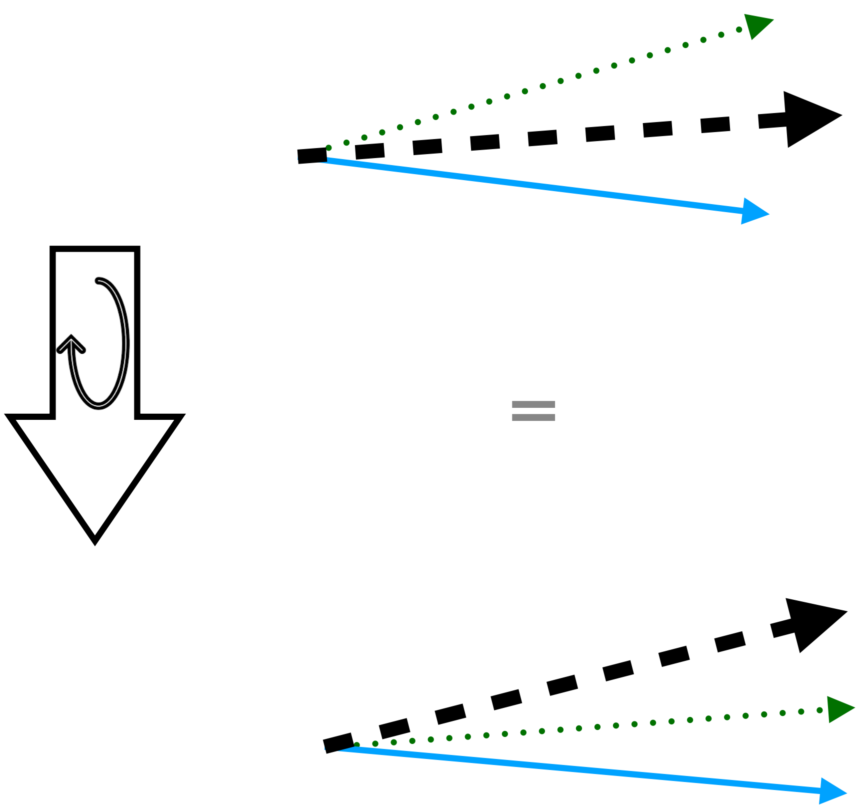
(Dated: August 29, 2023)

Using hidden symmetries



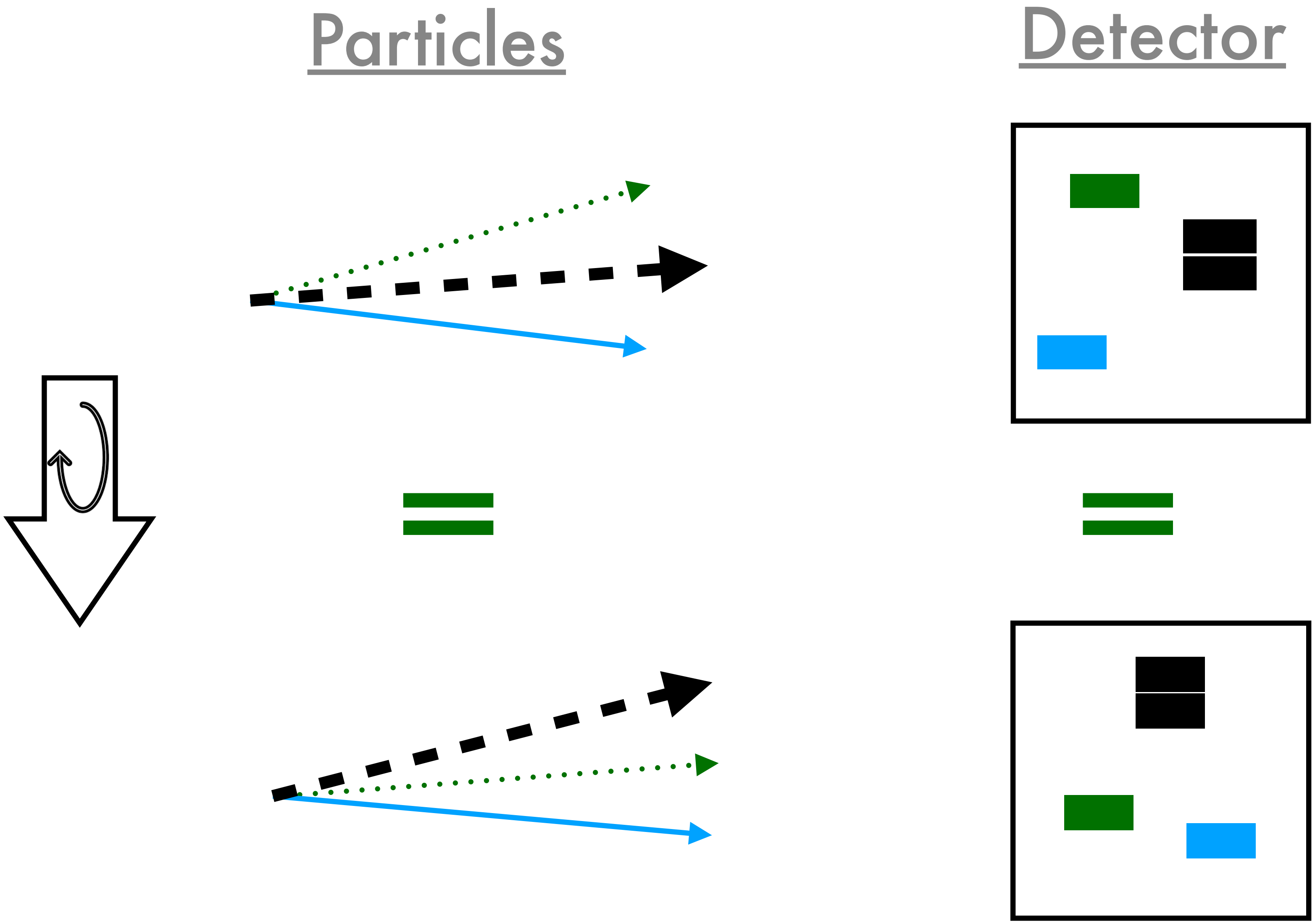
Using hidden symmetries

Particles



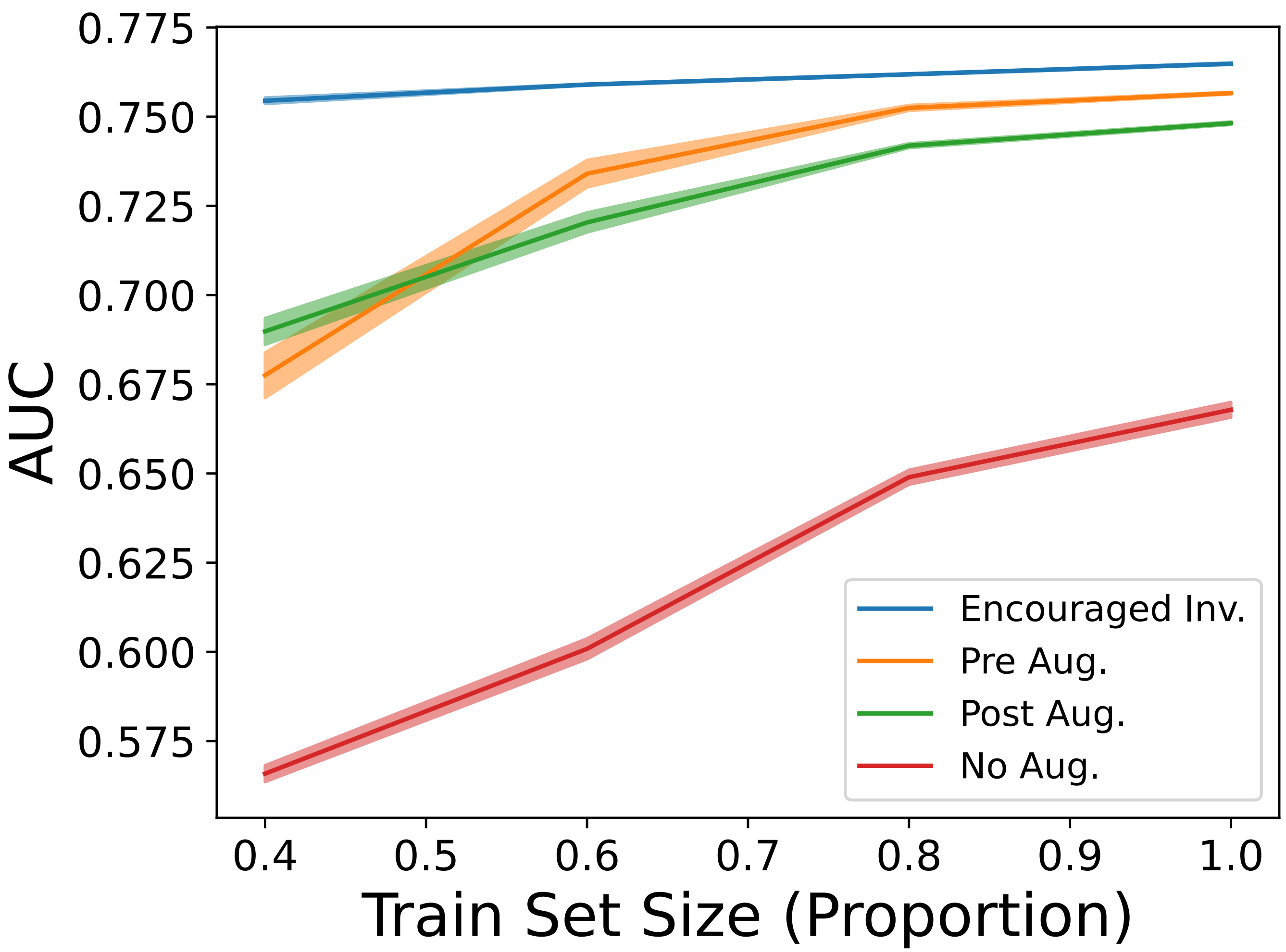
Encourage network to
Treat these as the same jet
Even if they look different
In the detector

Using hidden symmetries



Results

Simple toy dataset:
grid of non-square cells



Encouraged invariance helps even more!

Pre-detector augmentation helps more

Post-detector augmentation helps

Summary

We can extend jet tagging to many bosons

We can even decode what the network learns

We can incorporate hidden symmetries to boost learning