

# Masked particle modelling

Foundation models for HEP

EuCAIFCon 24

[2401.13537]

Sam Klein +



M. Leigh



J. Raine



L. Heinrich



M. Kagan



R. Osadchy



T. Golling

# Foundation models

## Why build them?

- Goal is to learn generic and robust representations
  - Allows models to be efficiently trained on **small datasets**
  - **Same** model can be reused for **many** downstream tasks
  - Save on resources

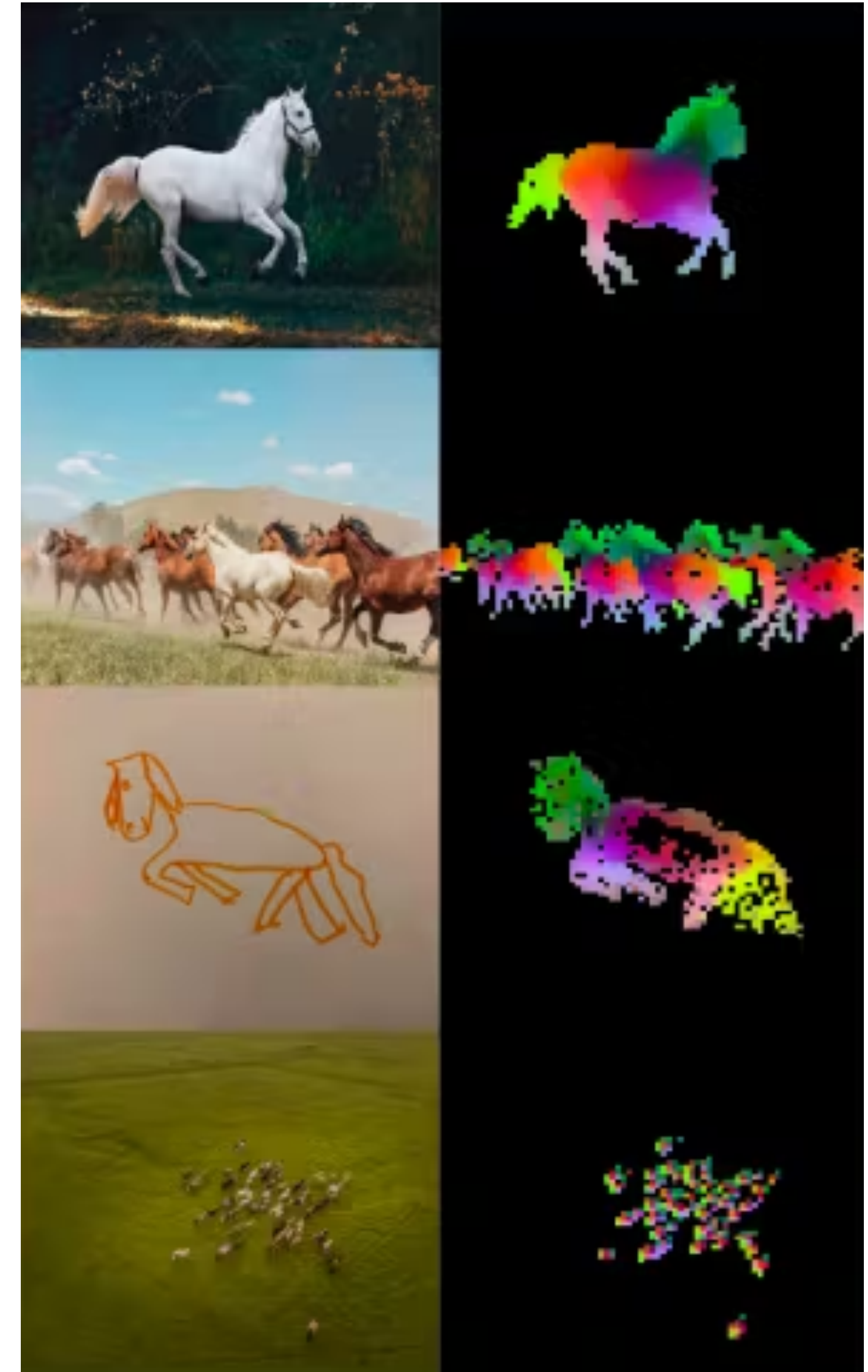
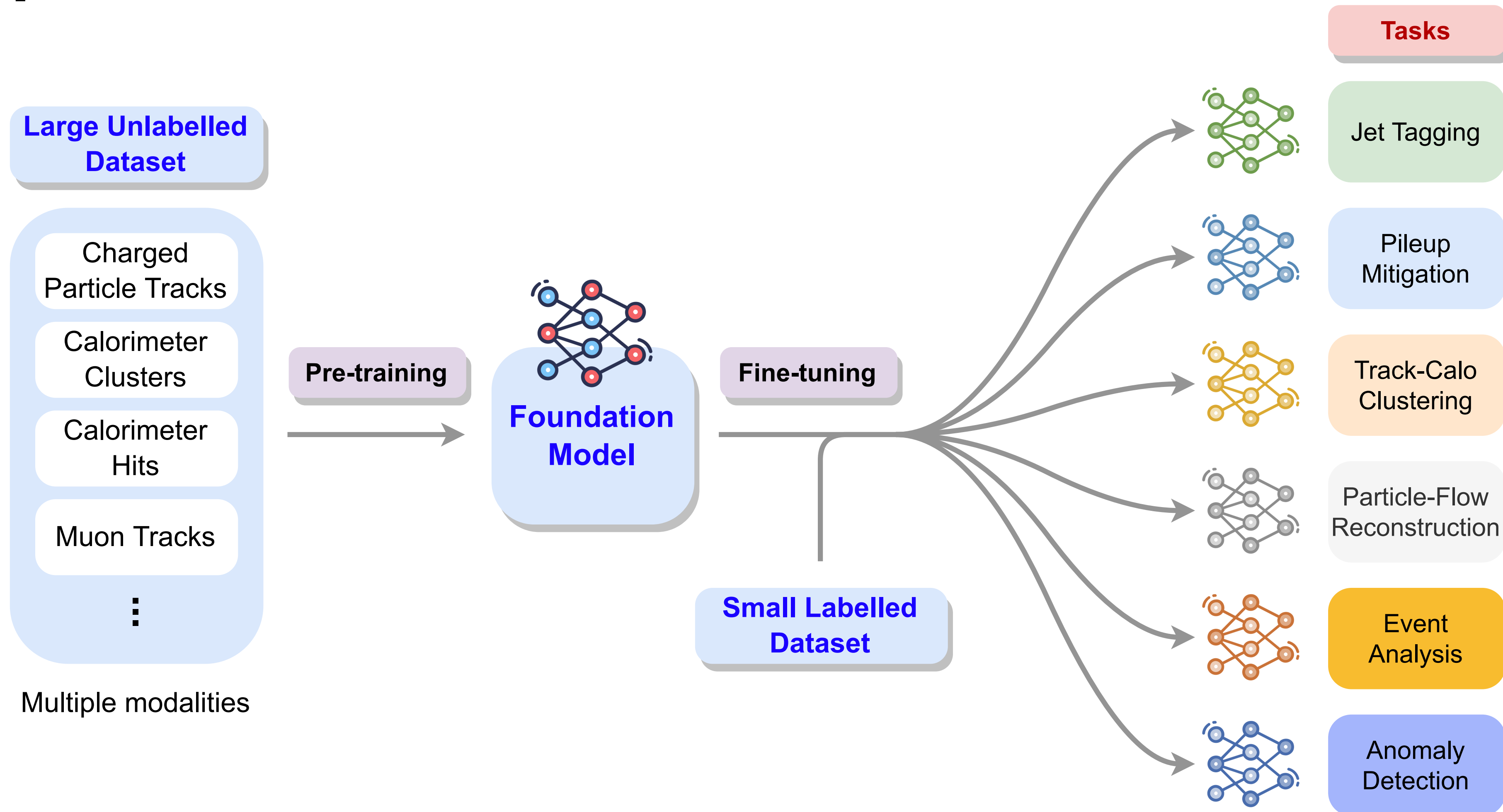


Image from DINOv2

# Foundation models

## In HEP?

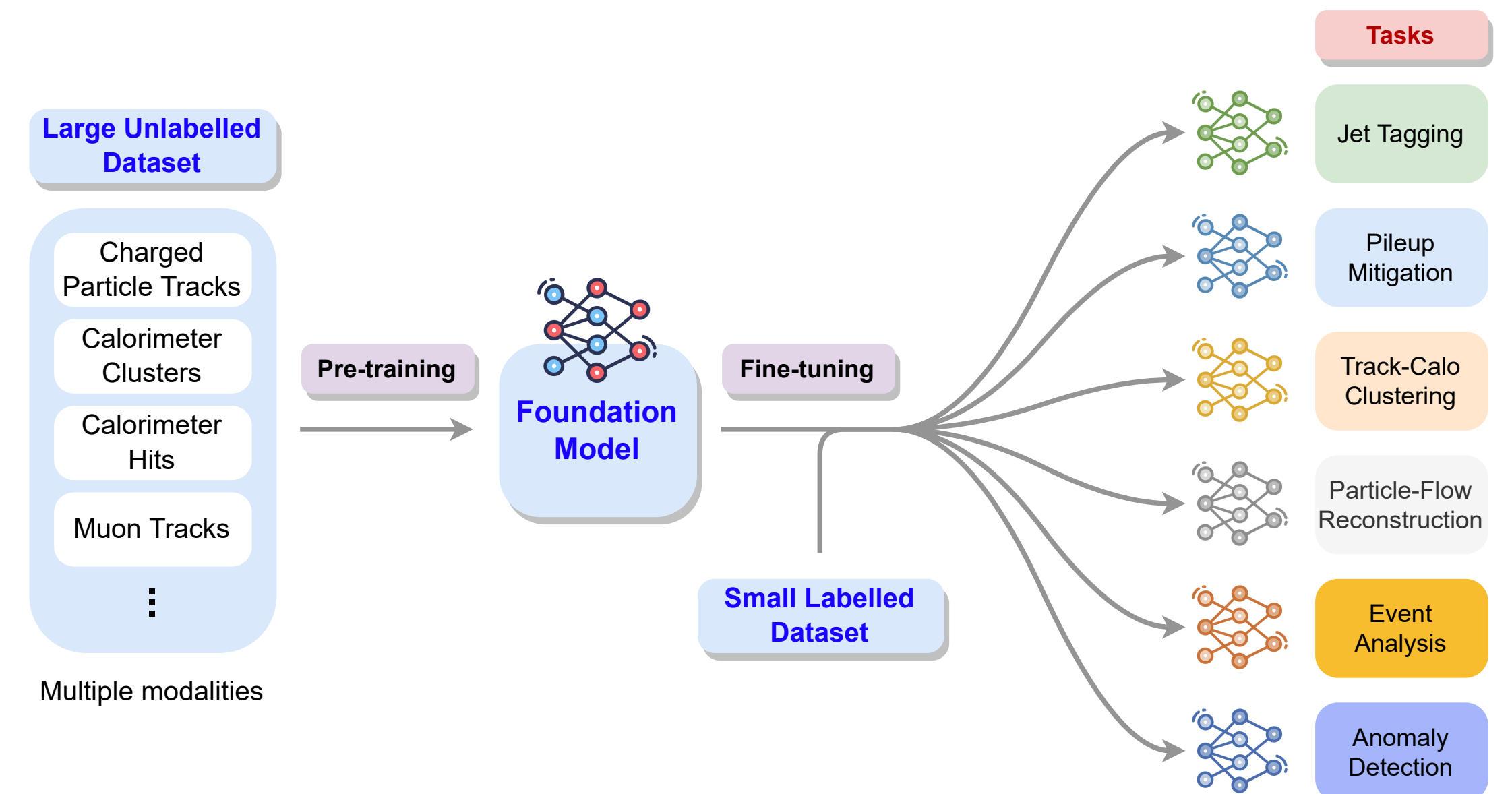




# Foundation models

## In HEP?

- Reduce dependence on large simulated datasets for supervised learning
- Help mitigate uncertainties related to domain shift?
- The problem: existing SSL strategies are **data type specific**, so we need new methods!



# Masked modelling

## Images and words

- The BERT pretraining strategy has been very successful for NLP
- So has BEiT for images
- Both based on recovering masked input sequences

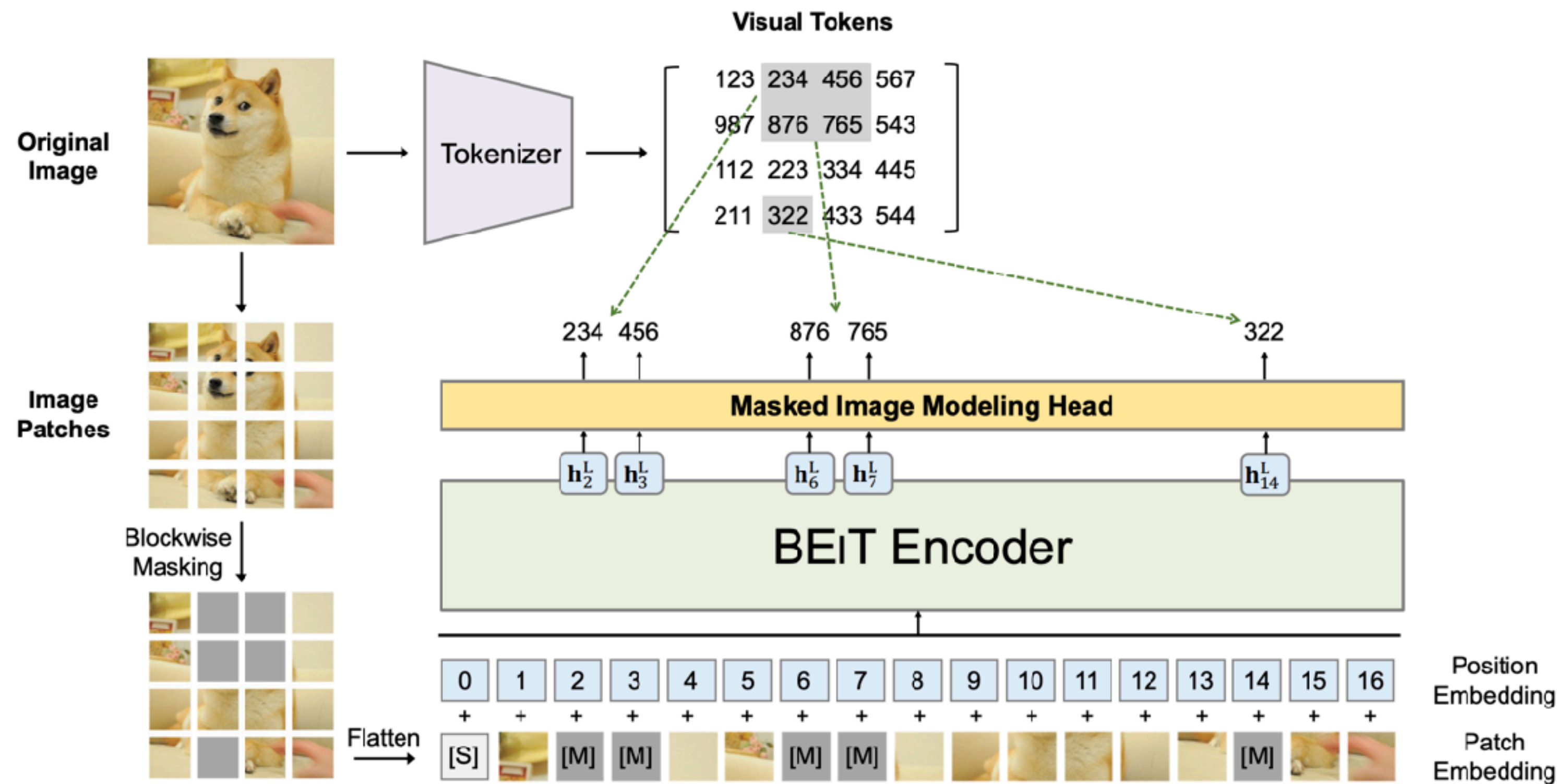
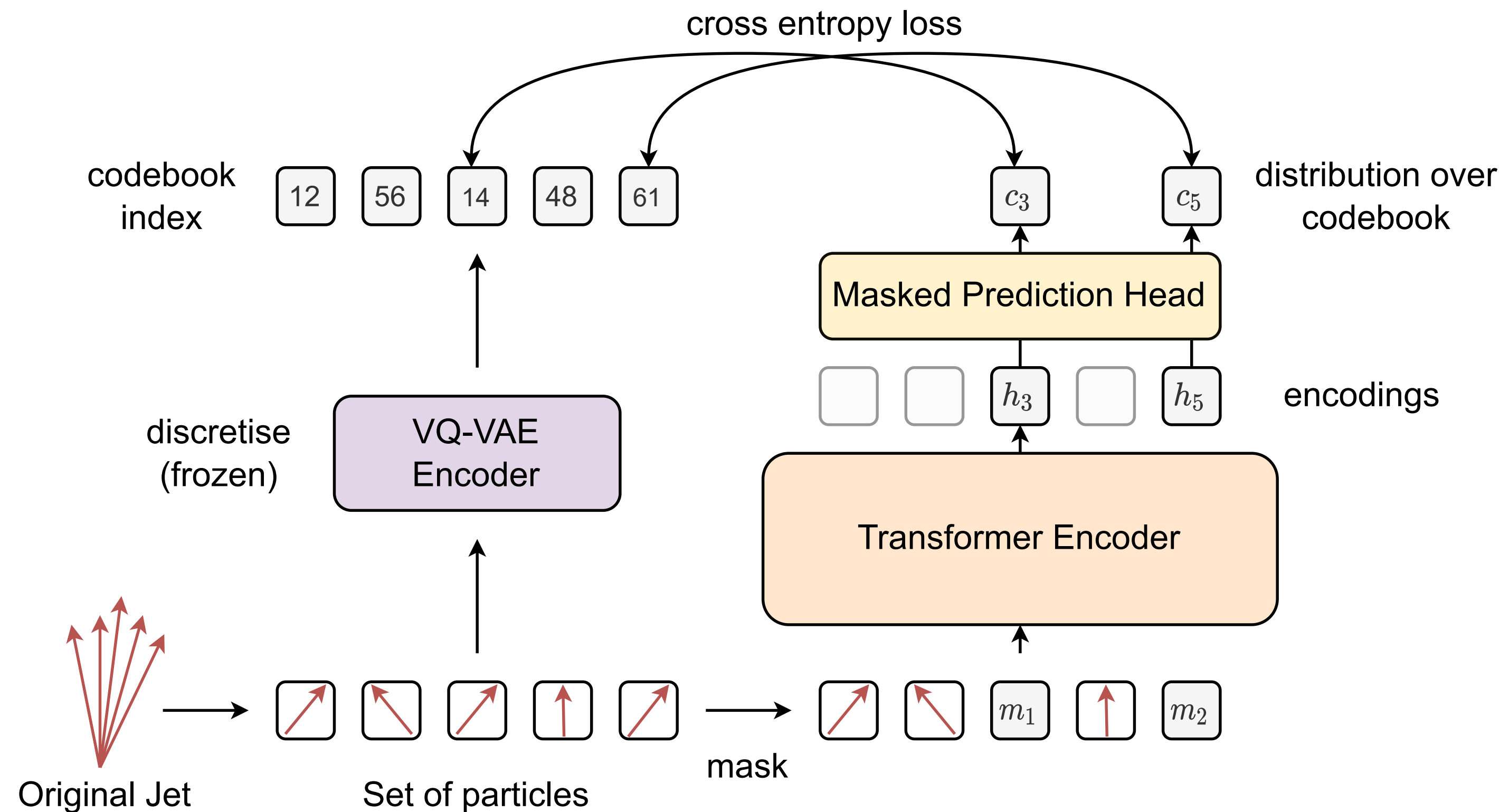


Image from [2106.08254](https://www.istockphoto.com/photo/210608254)

# Masked modelling

## Does this work for HEP: Jets

- Like images: continuous inputs
- Like language: 'meaningful' constituents
- Unlike both: no positional information
- No public massive dataset
  - Use jetClass 100M



# Masked modelling

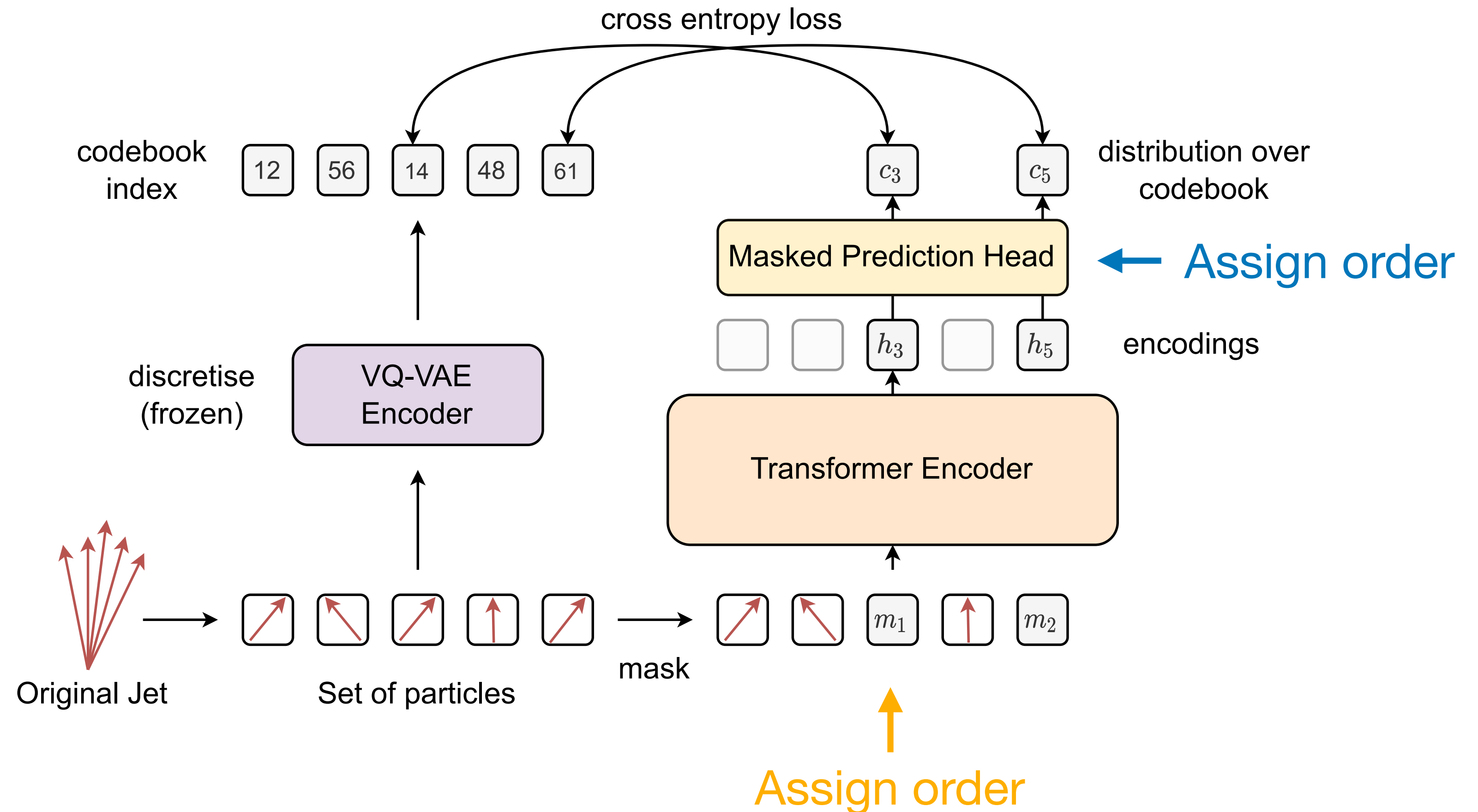
## Permutation invariance

- Three approaches to permutation invariance

- Don't worry about it

- Input to backbone

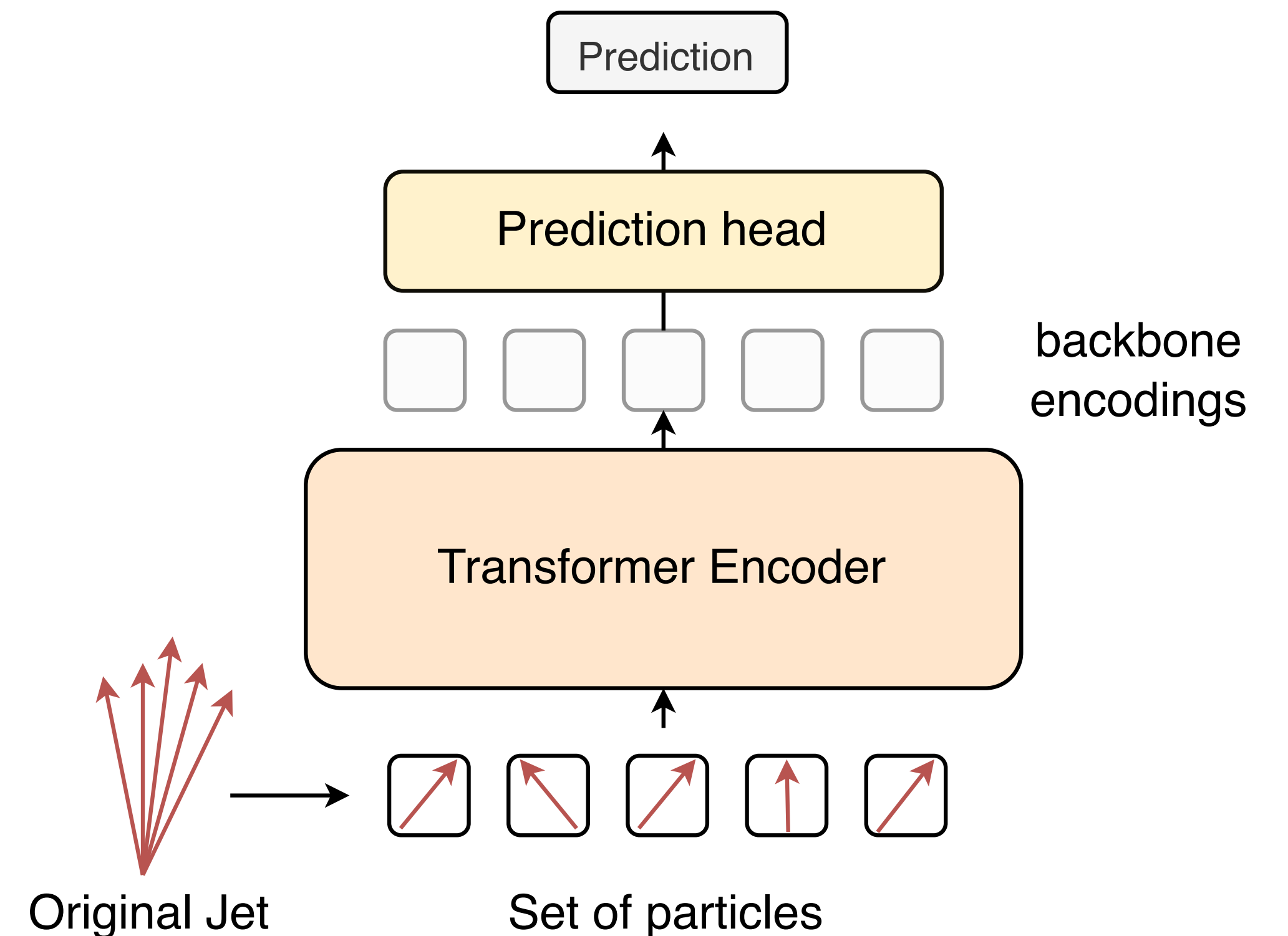
- Input to masked prediction head



# Masked modelling

## Permutation invariance

- Three approaches to permutation invariance
- Which one to pick?
- JetClass has 10 classes
- Use linear separation





# Masked modelling

## Permutation invariance

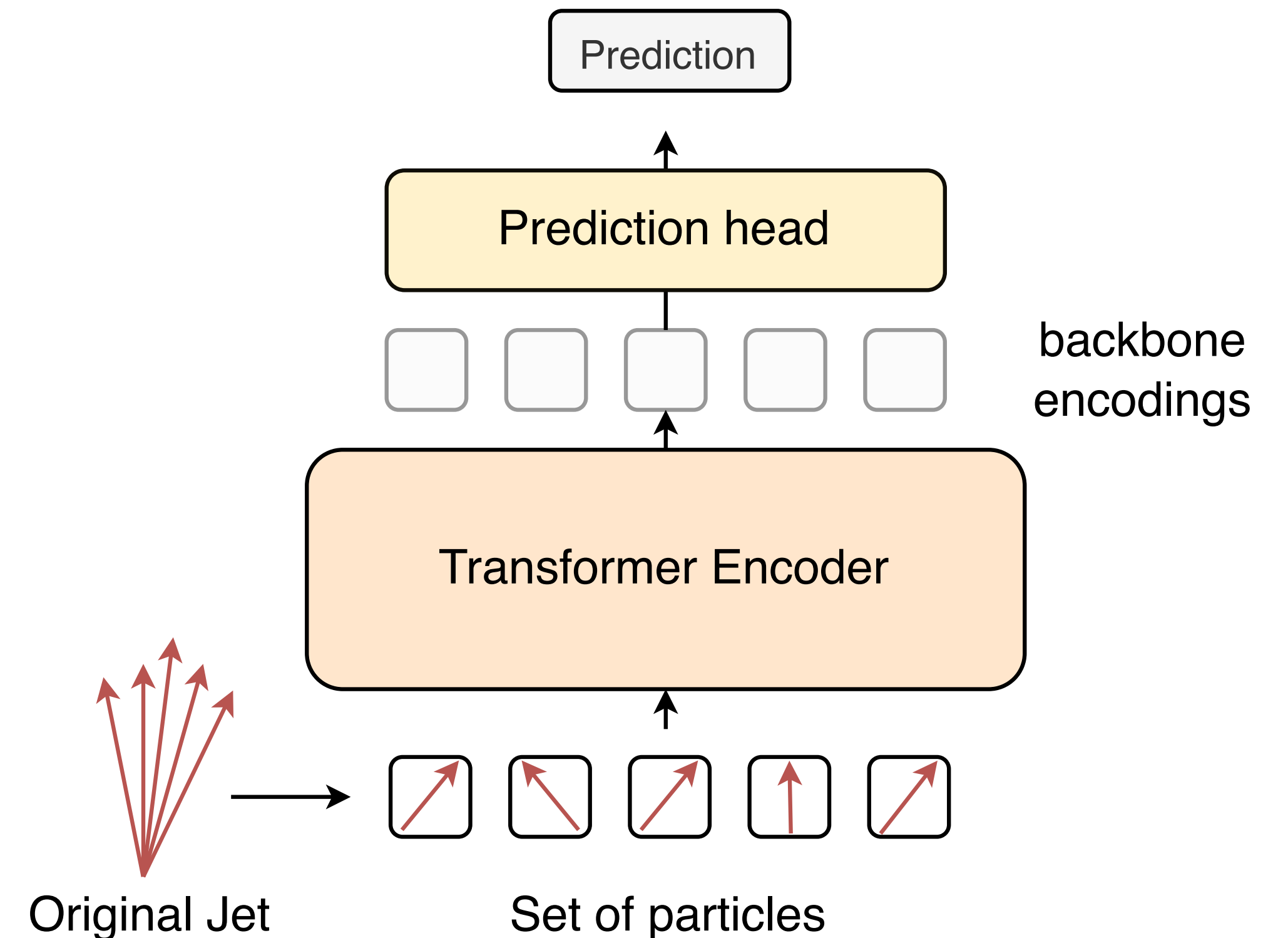
- Ordering at the pretraining head does the best
- Ordering at the input leads to overfitting

	No order	Order input	Order head
Linear Accuracy	54.1%	53.4%	<b>56.8%</b>

# Masked modelling

## Performance

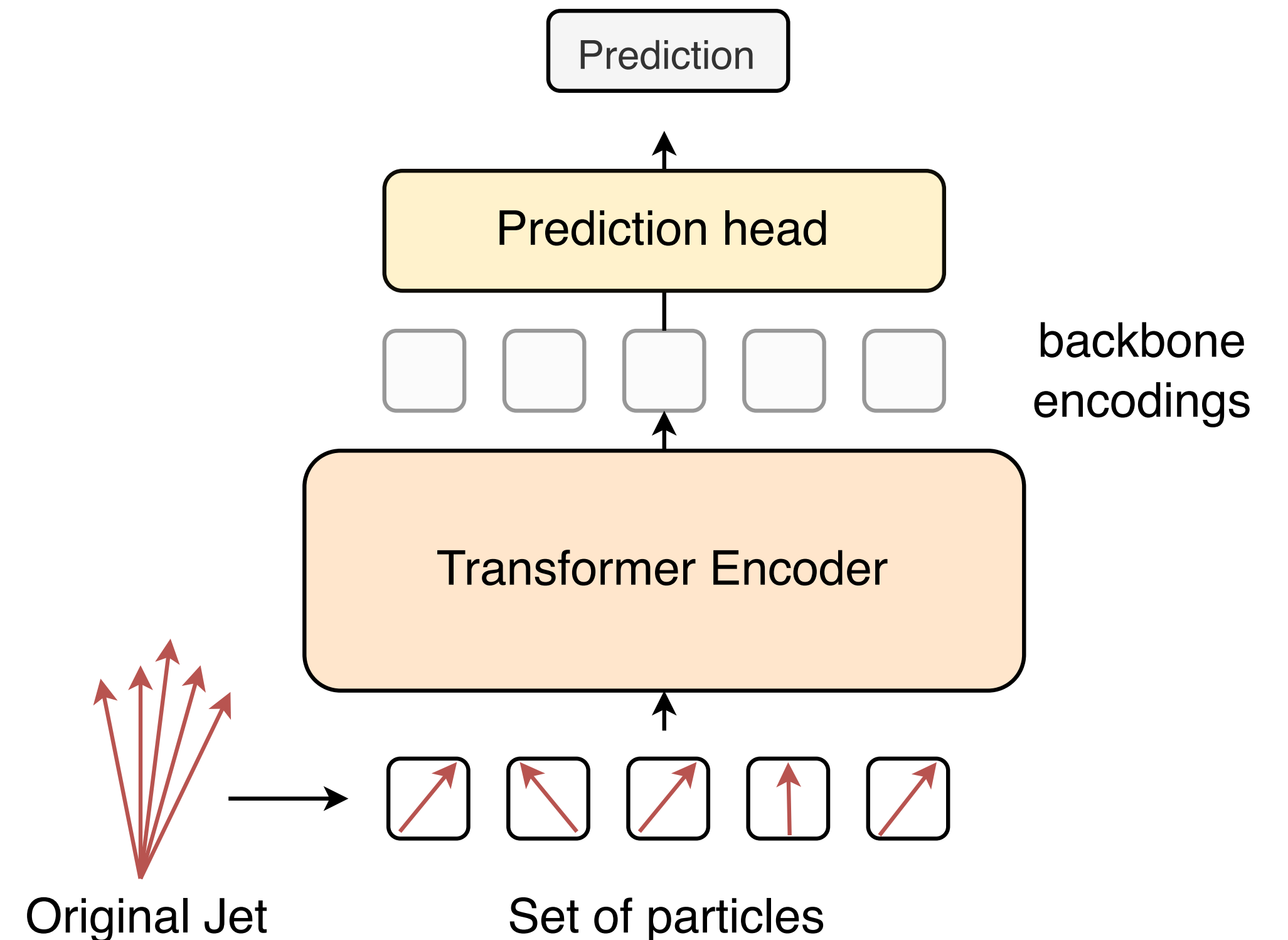
- How to quantify the performance of a pretrained model?
  - Array of downstream tasks — fine tuning
- Pretraining on 100M Jets from JetClass
- Fine tuning on array of different jet level tasks



# Masked modelling

## Training strategies

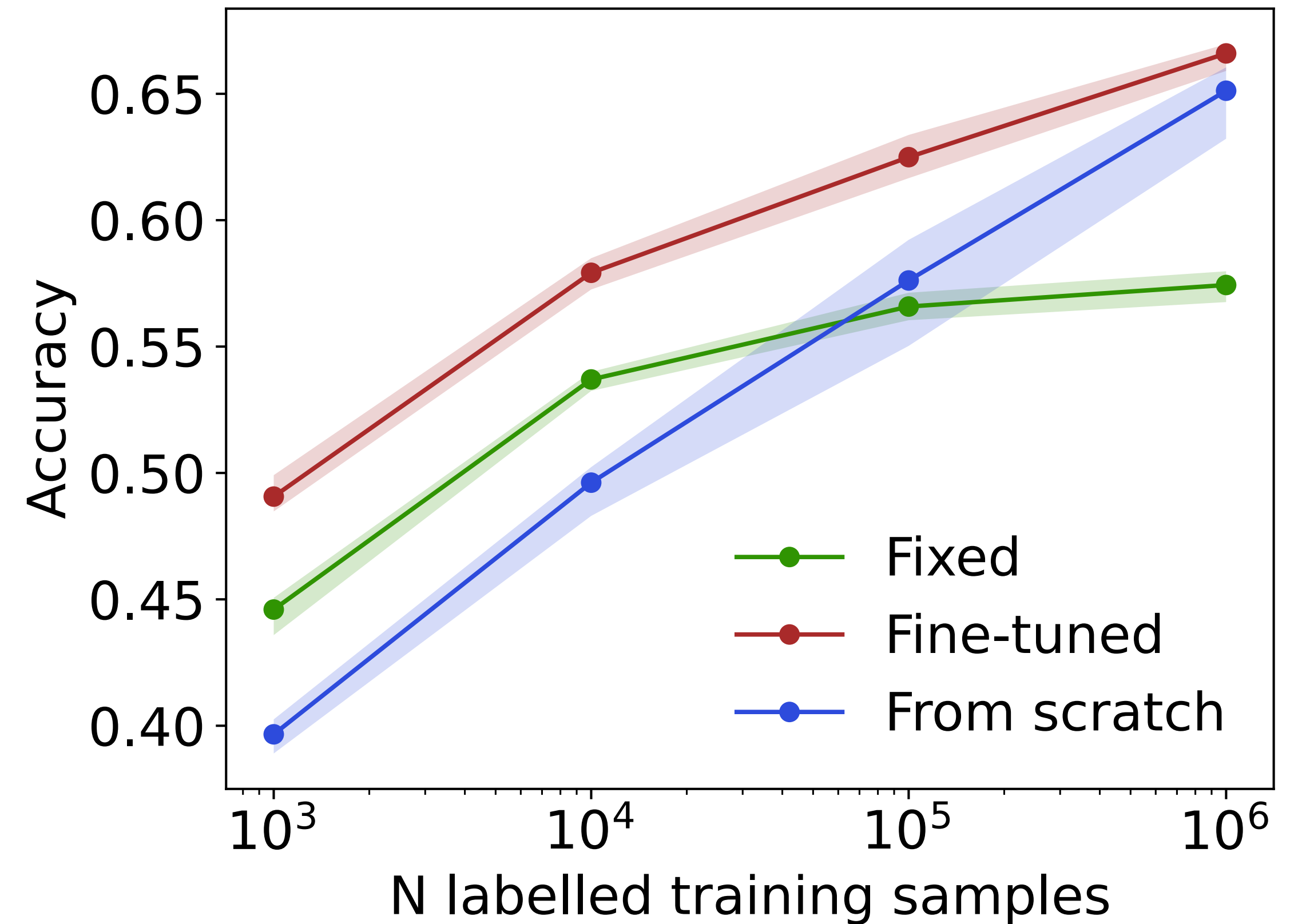
- **fixed backbone:**  
Freeze the encoder
- **fine-tune backbone:**  
Train the prediction head and the backbone
- **from scratch:**  
Reinitialise model from scratch



# Masked modelling

## Fine tune on pretraining set

- JetClass contains 10 classes
- Select N events and fine tune
- The backbone model outperforms from scratch

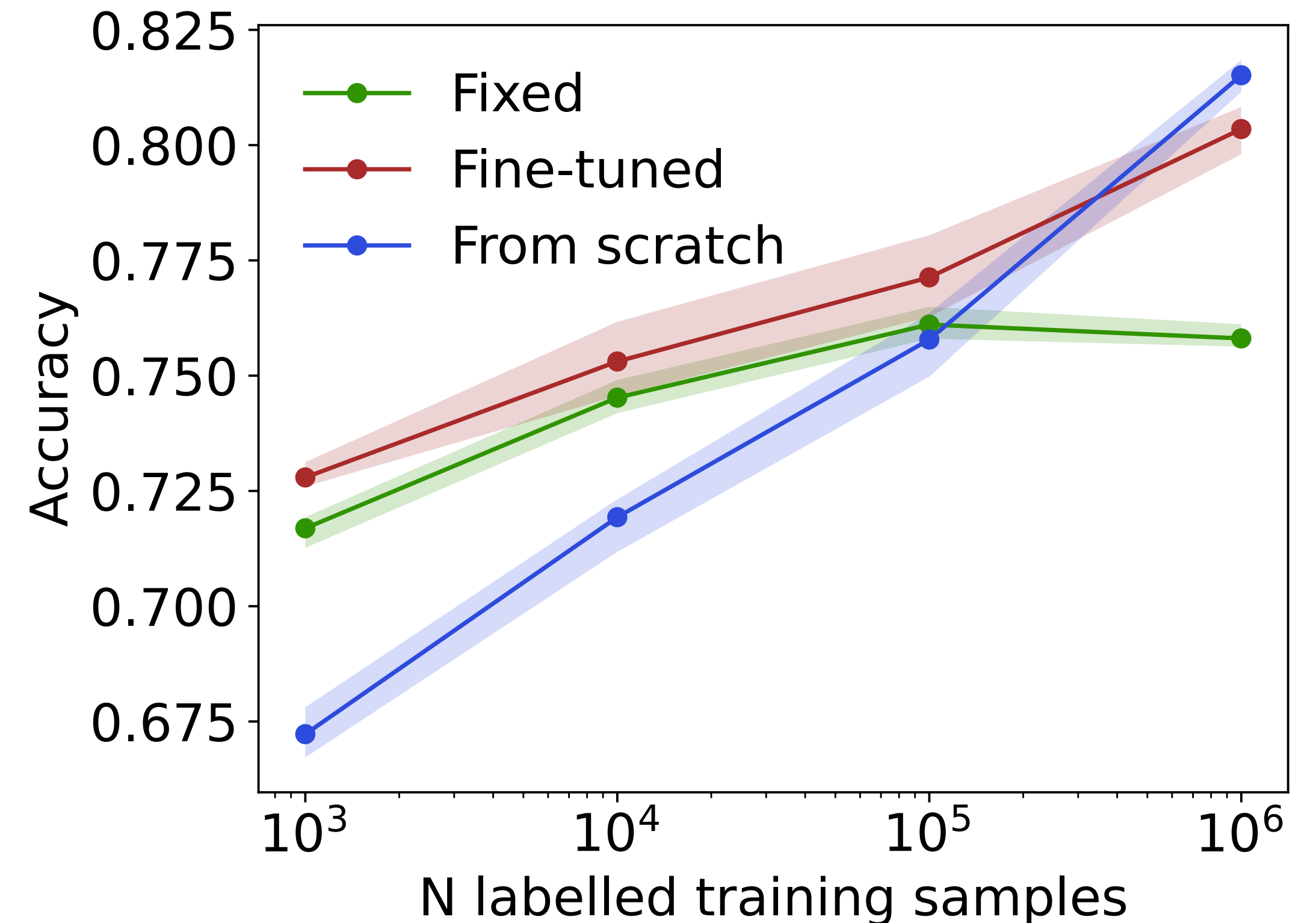




# Masked modelling

## Fine tune on new dataset

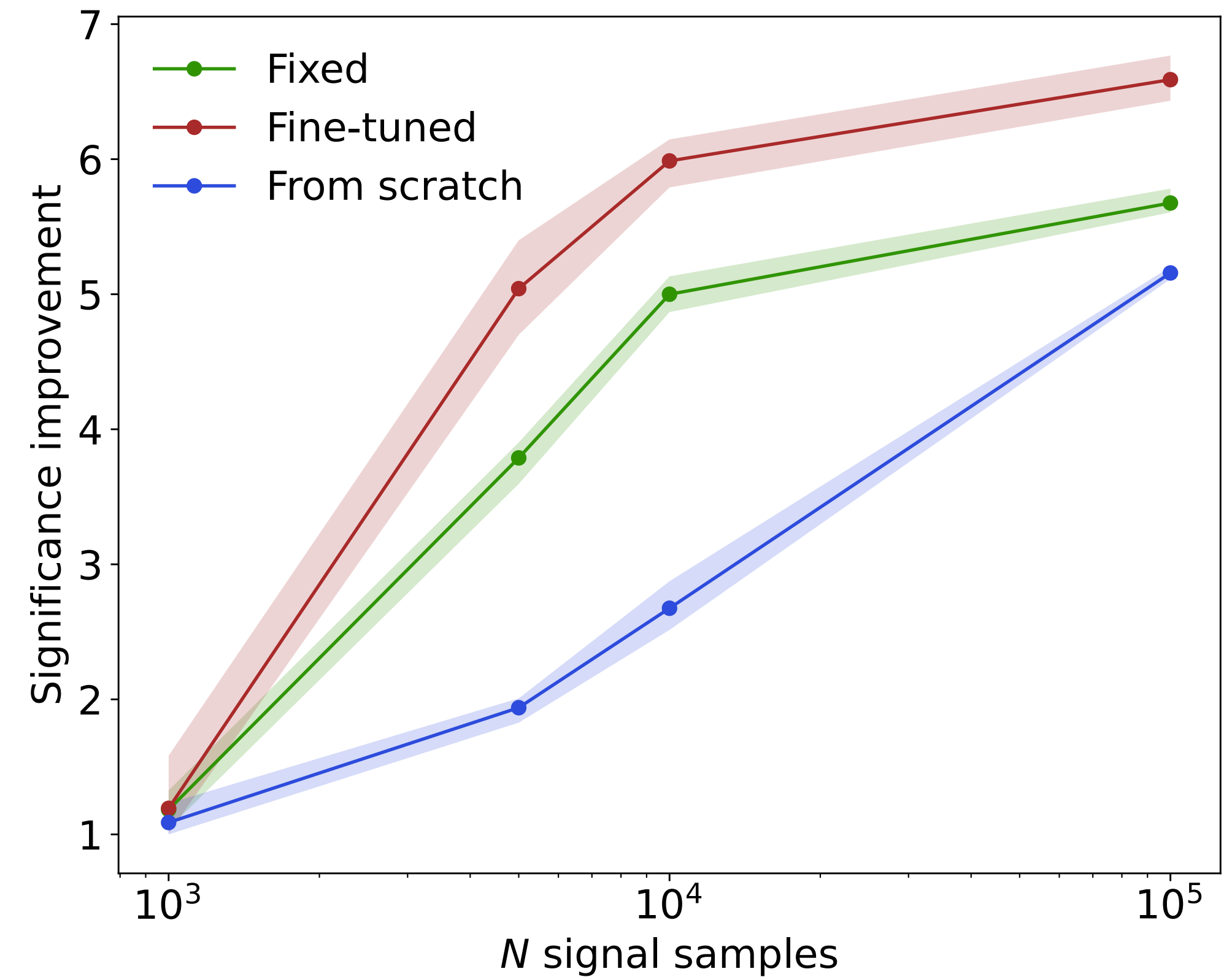
- The learned features are generically useful
- The performance gain applies to data generated with a different simulator
- Change card to Atlas and fine-tune (JetClass is CMS)



# Masked modelling

## Fine tune on weak supervision

- Take two QCD samples
- Add  $x$  top jets to one sample and label 'signal'
- Fine-tune model on noisy labels
- Pretraining helps!



# Summary

## Masked particle modelling

- Masked particle modelling is a useful pretraining task for HEP
- Simple and easy to set up (when using a kNN)
  - Can be applied to low level data cheaply
- Foundation models can and should be built for HEP
- Permutation invariant issue not tackled in other domains
  - Plays important role in HEP