Masked particle modelling

Foundation models for HEP

Sam Klein +

M. Leigh   J. Raine   L. Heinrich   M. Kagan   R. Osadchy   T. Golling

Masked modelling
Images and words

- The <u>BERT</u> pretraining strategy has been very successful for NLP

- So has <u>BEiT</u> for images

- Both based on recovering masked input sequences
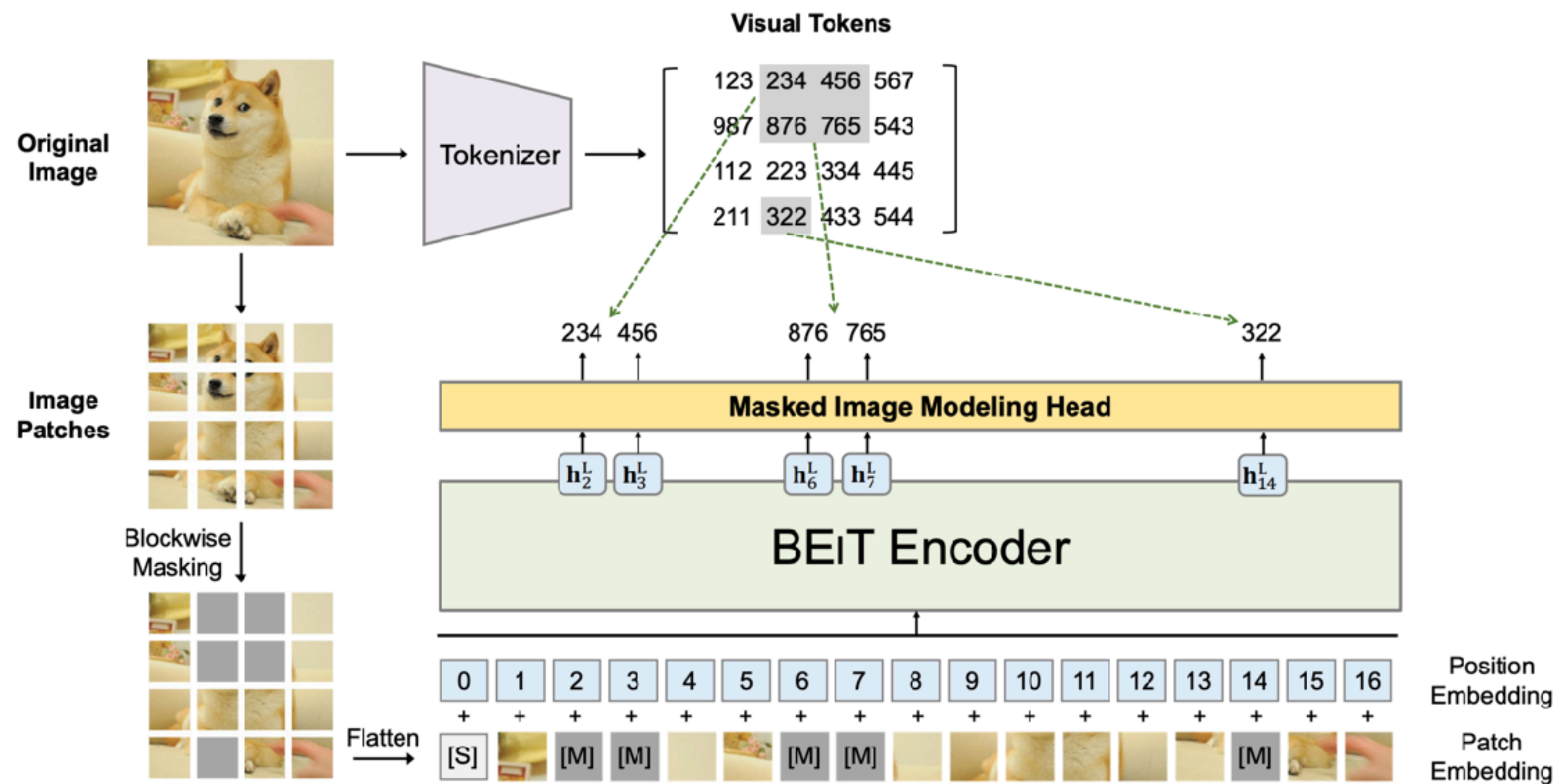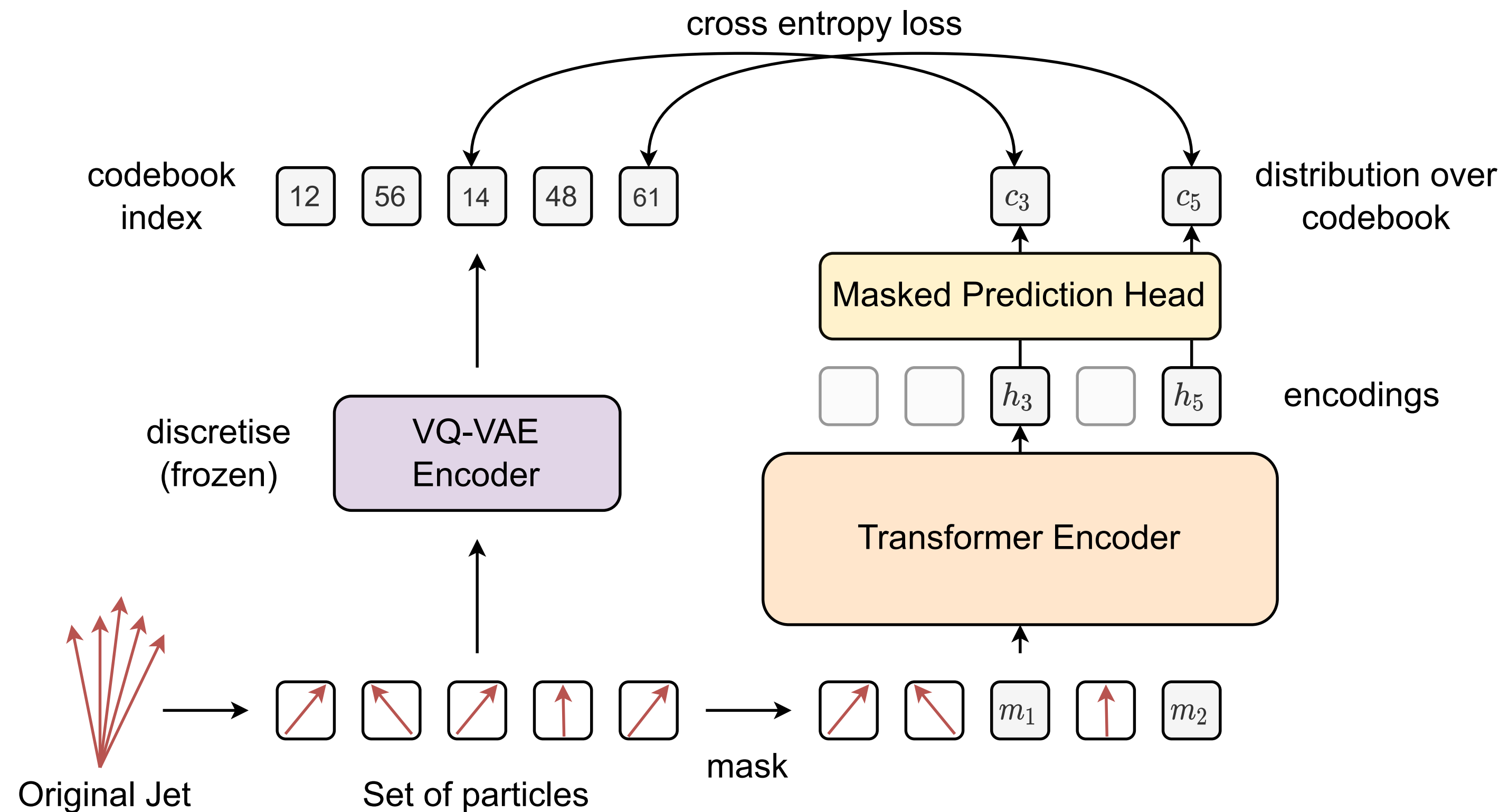

Image from <u>2106.08254</u>

Masked modelling
## Does this work for HEP: Jets

- Like images: continuous inputs

- Like language: 'meaningful' constituents

- Unlike both: no positional information
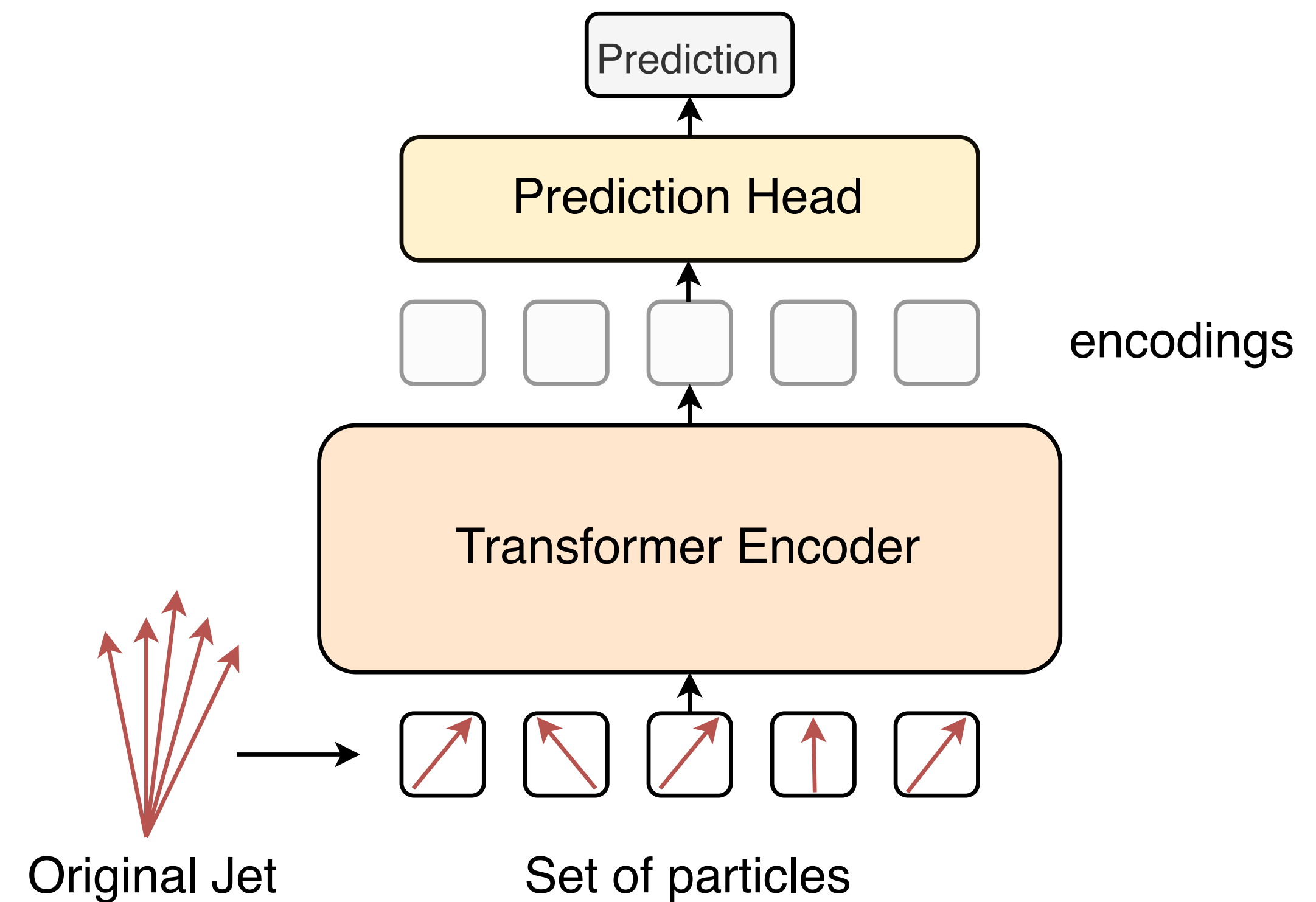
Masked modelling
Performance

- 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 classes

Prediction

Prediction Head

encodings

Transformer Encoder

Original Jet   Set of particles
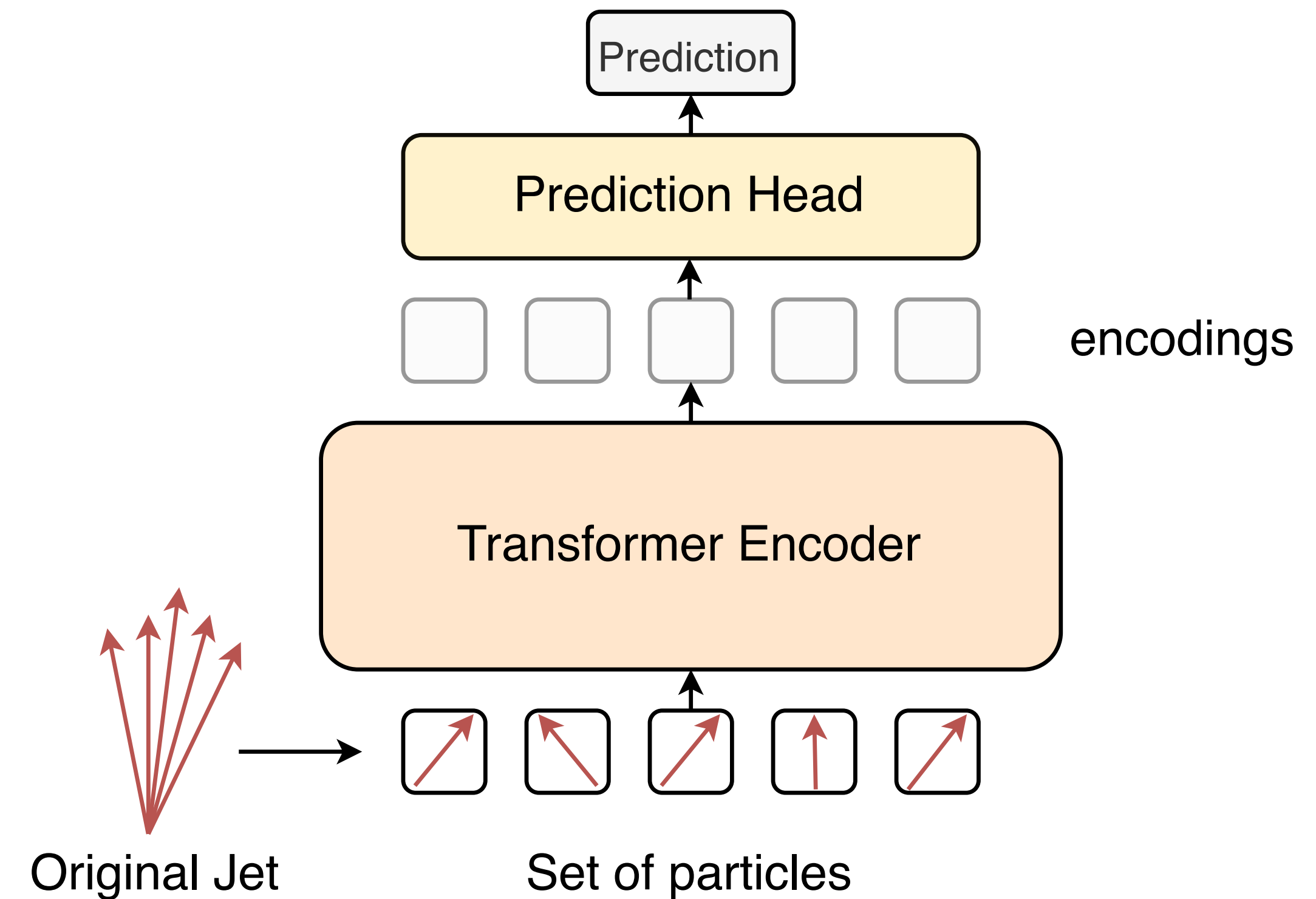
Masked modelling

Training strategies

- Freeze the encoder = fixed backbone

- Train the prediction head and the backbone = fine-tune backbone

- Reinitialise everything from scratch = from scratch

Prediction

Prediction Head

encodings

Transformer Encoder

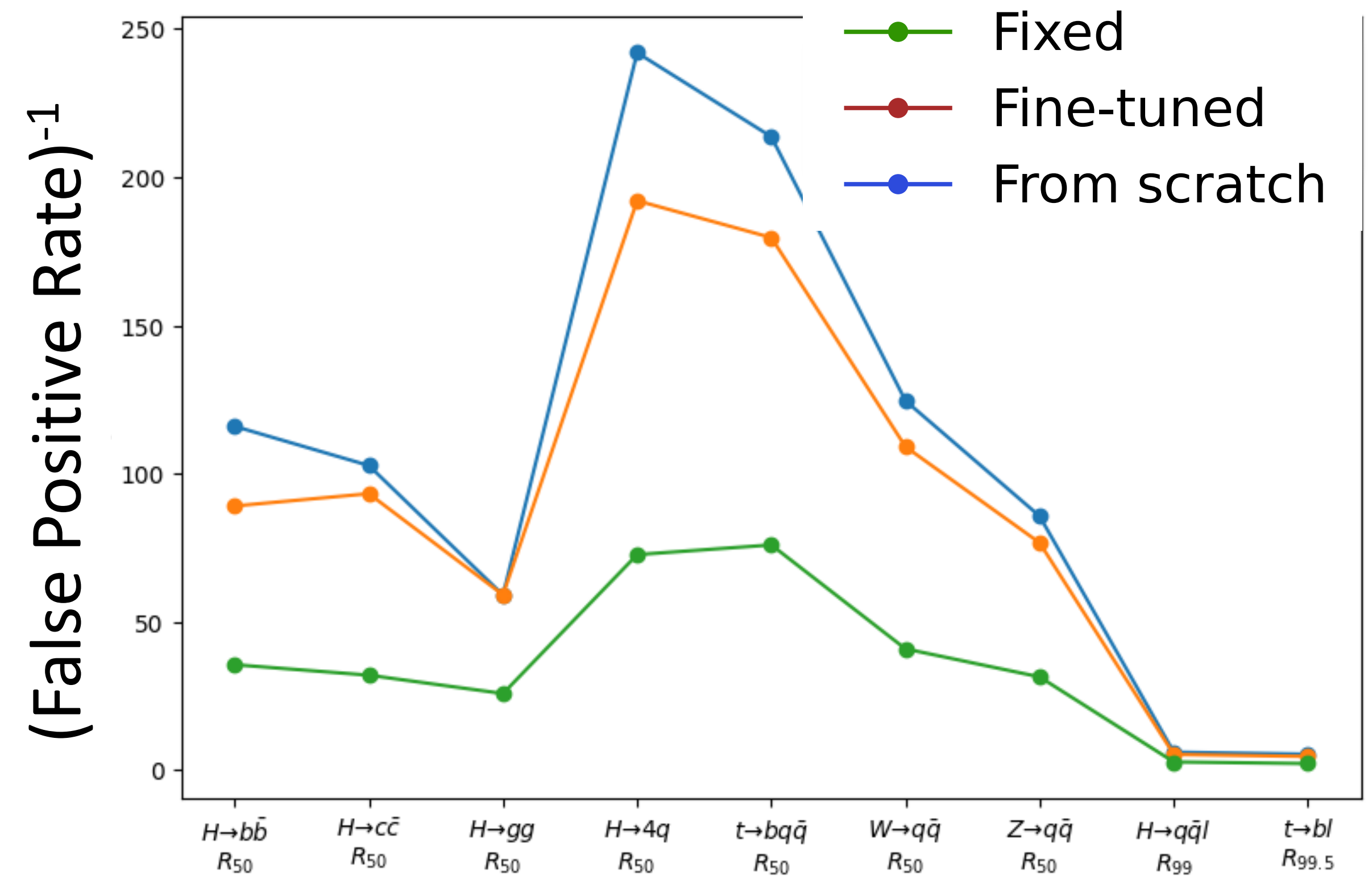Original Jet   Set of particles

Masked modelling
Fine tune on pretraining set

- JetClass contains 10 classes

- Select 1M events and fine tune

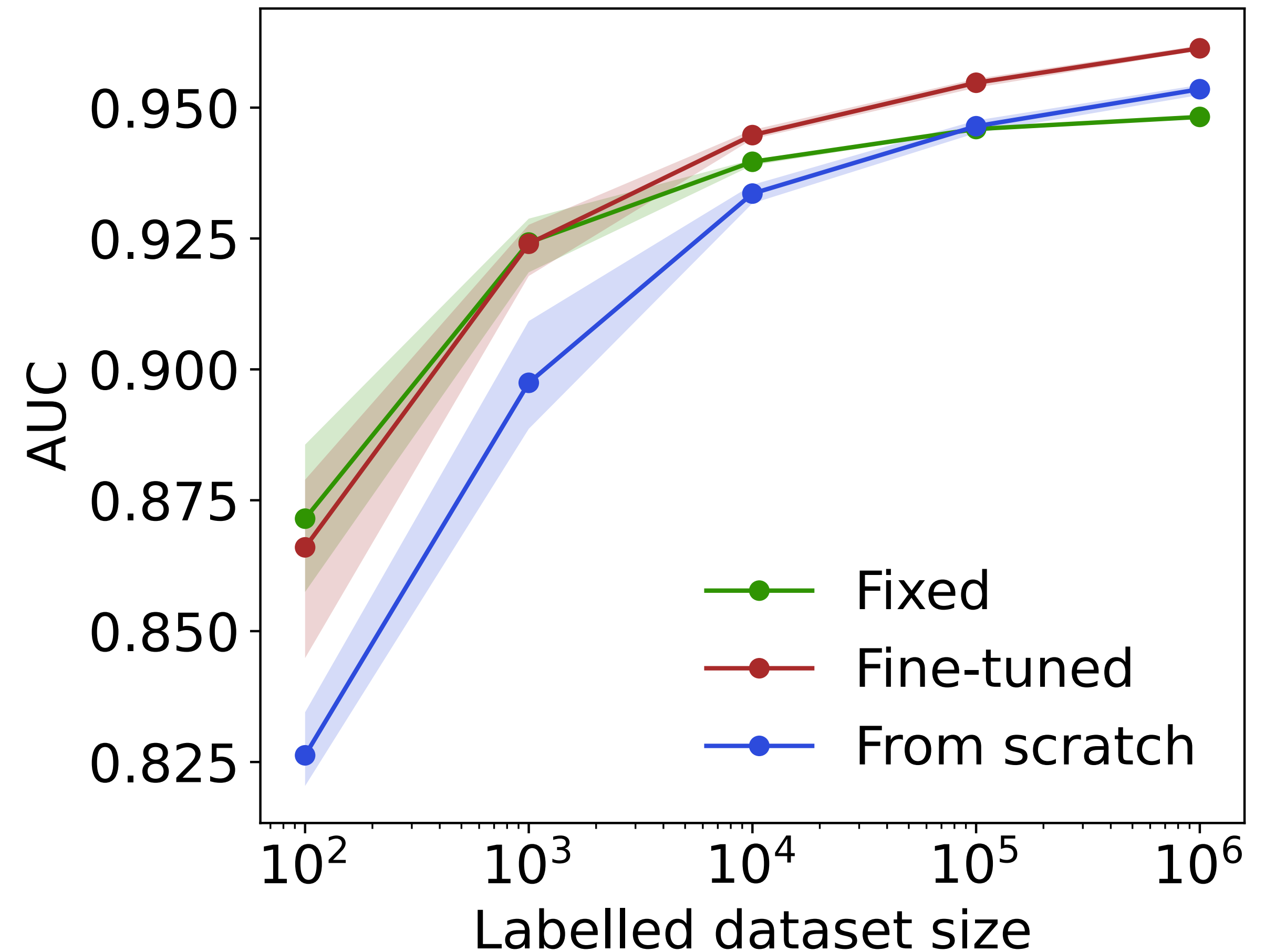- The backbone model outperforms from scratch

10-Class Classification Performance

Masked modelling
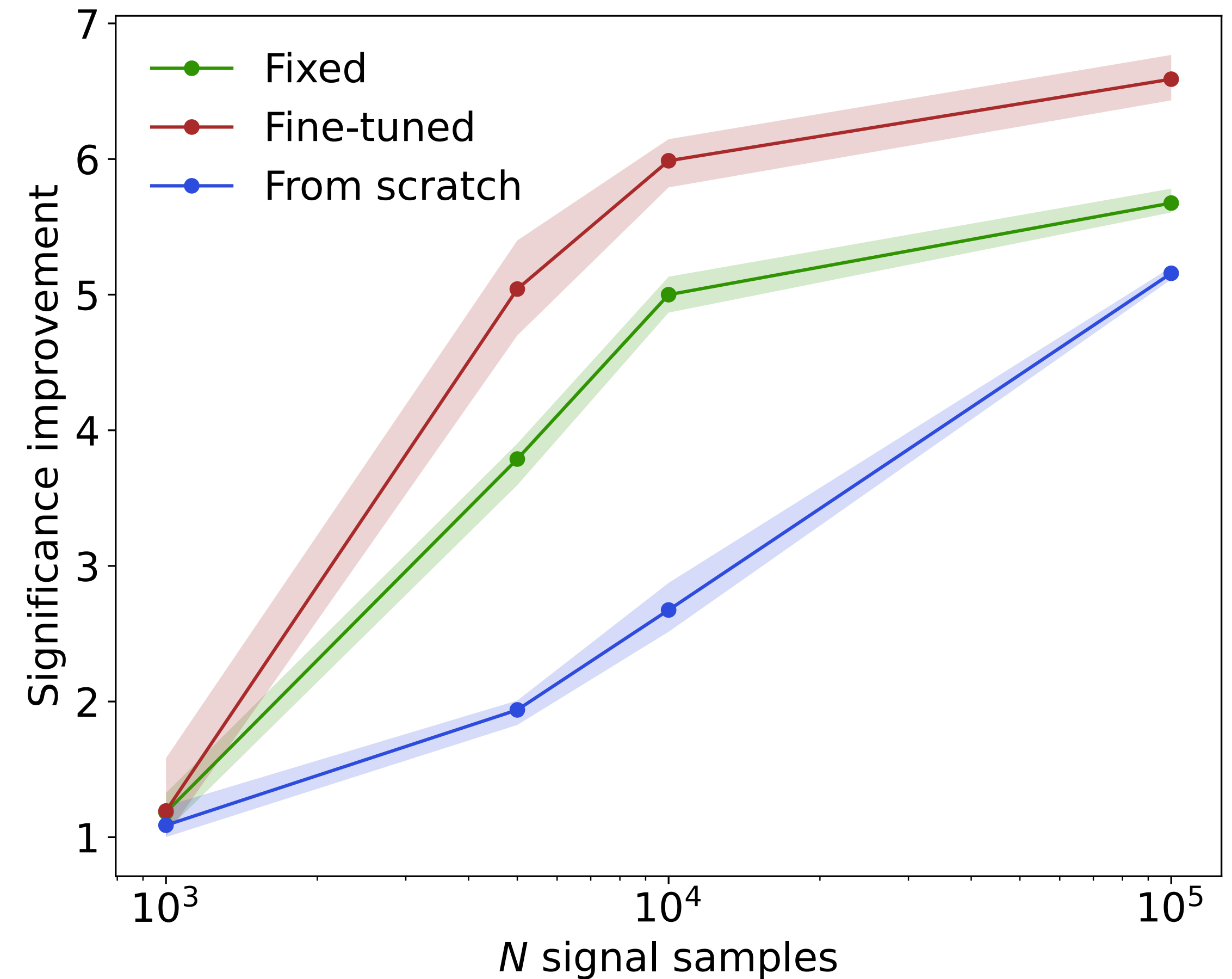## Fine tune on new dataset

- 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
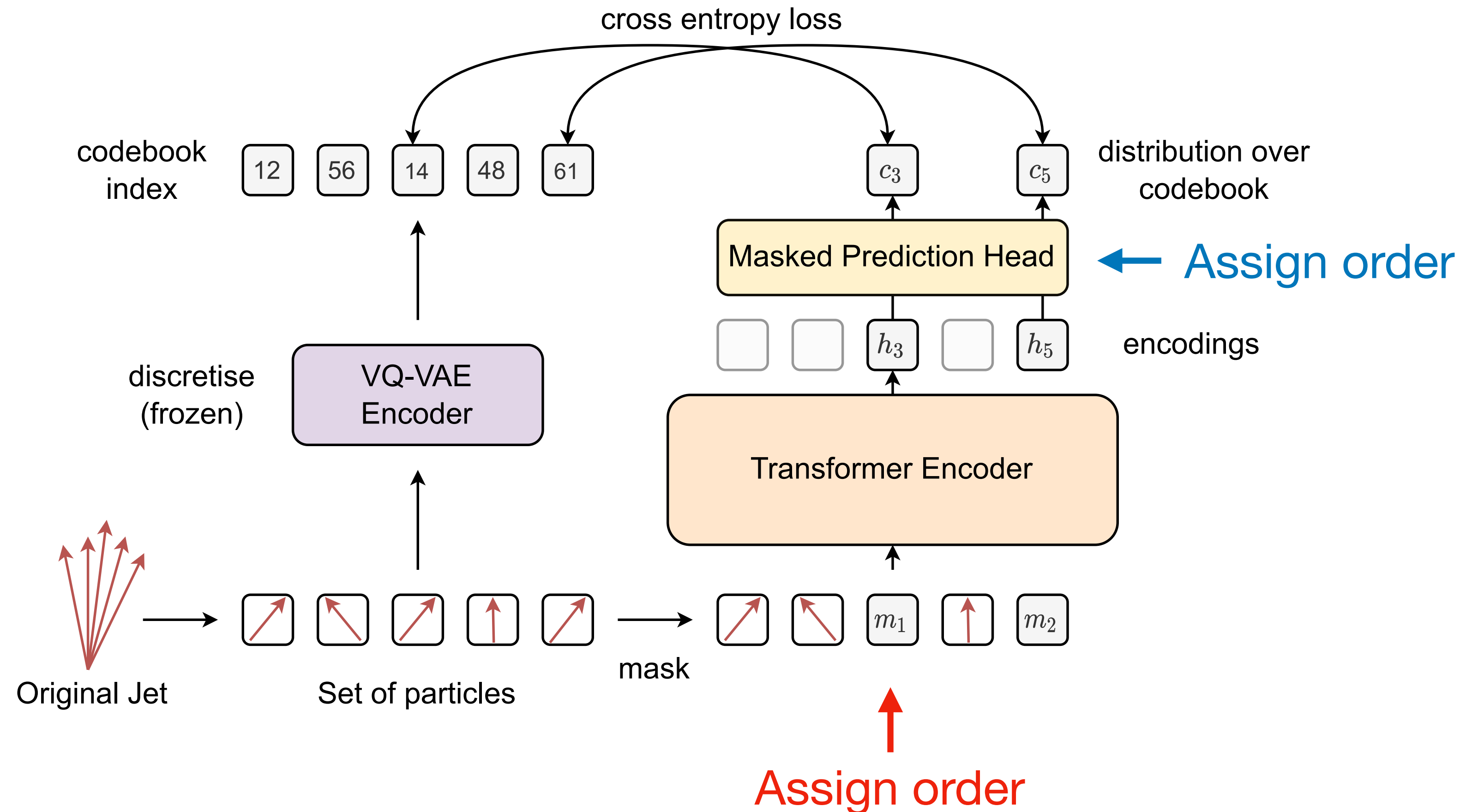
- Pretraining helps!

Summary
Masked particle modelling

- Masked particle modelling is a very useful pretraining task for HEP

- Permutation invariant issue not tackled in other domains

  - Plays important role in HEP

- If we really learned a useful representation then this should be useful for many downstream tasks

Masked modelling
Permutation invariance

- Three approaches to permutation invariance

  - Don't worry about it

  - Input to backbone

  - Input to masked prediction head

cross entropy loss

codebook index   12   56   14   48   61 c_3 c_5 distribution over codebook

Masked Prediction Head   ← Assign order

h_3 h_5 encodings

discretise (frozen)   VQ-VAE Encoder

Transformer Encoder

Original Jet   Set of particles   mask m_1 m_2

Assign order

Masked modelling
Permutation invariance

I'm working to add a
red line to this figure

- Three approaches to permutation invariance

  - Don't worry about it

  - Input to backbone

  - Input to masked prediction head