Learning Powerful Jet Representations via Self-Supervision

Qibin Liu, **Shudong Wang**, Congqiao Li, Huilin Qu

ML4Jets 2024, LPNHE, Paris

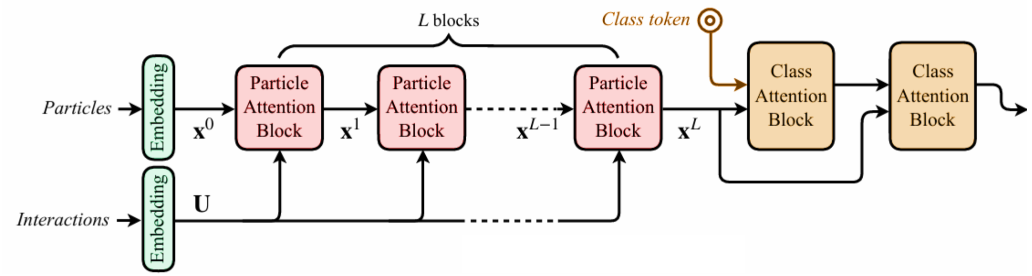
7 November, 2024

Introduction

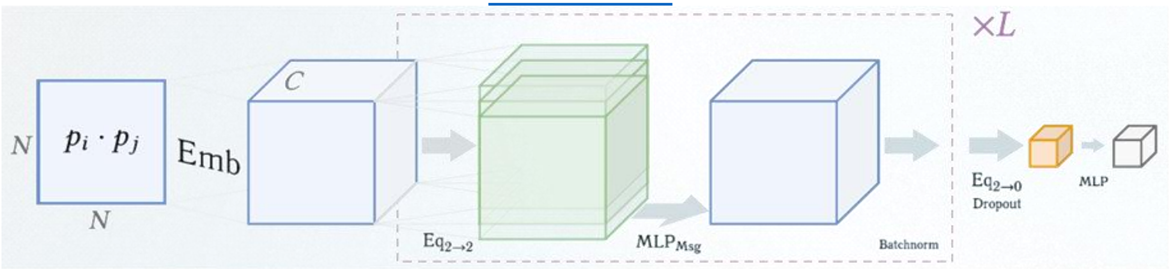
Machine Learning has been extensively used for Jets

- In recent years…

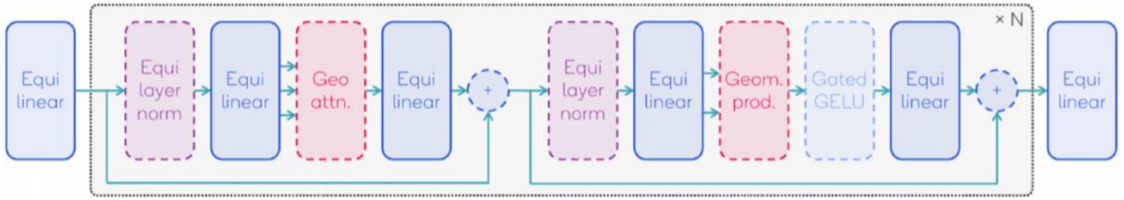
Particle Transformer



PELICAN



L-GATr



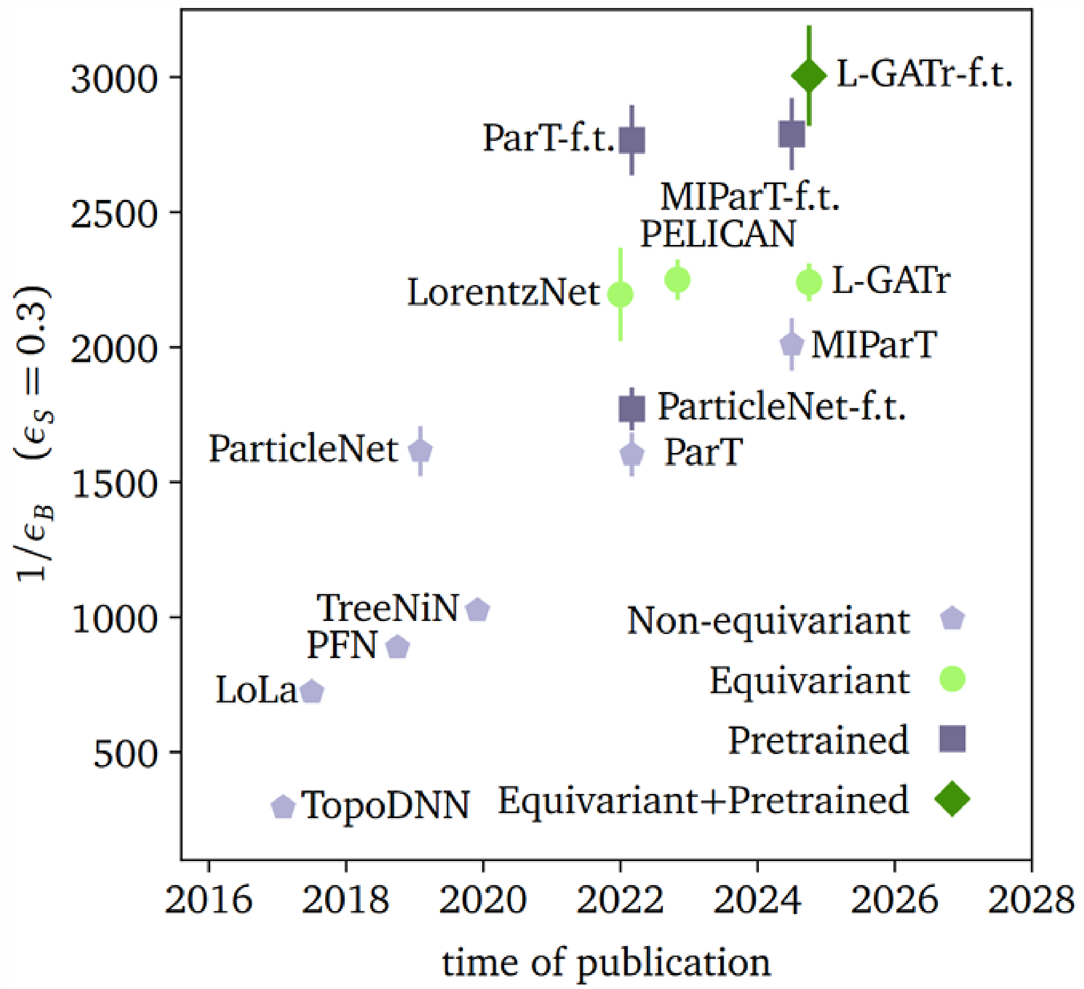


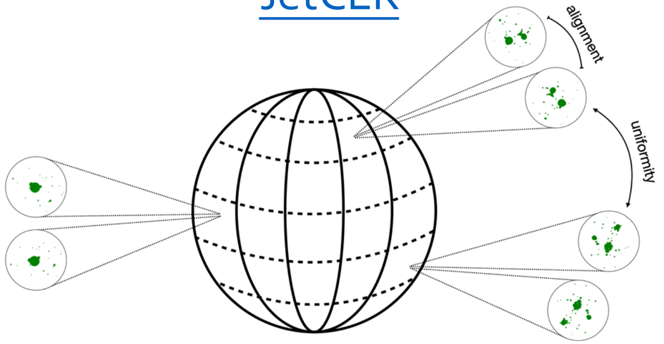
Figure Credit: Víctor's talk

Self-supervised Learning

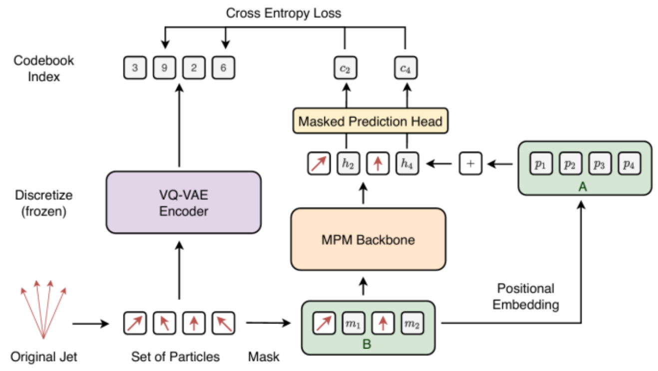
Learn from data directly?

- Training with no label required
  - Could **learn from data** directly and efficiently!
  - **Relying less on MC**, has the potential to better control uncertainty!
- Extract physics behind jets
  - Encourage algorithms to **learn physics**, rather than obsessed with minor details
    - Parton shower, hadronization, detector effects…
- Learn comprehensive jet representations suitable for various applications
  - Jet tagging, generation, reconstruction, anomaly detection…
  - Prospect: an approach to construct **Foundation Models**
- Self-supervised learning on jets
  - JetCLR, MPM, OmniJet-α …

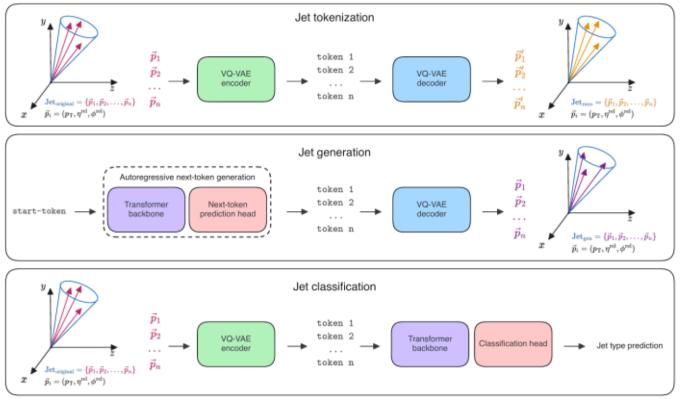
JetCLR



MPMv1



OmniJet-α



Architectures of Self-supervised Learning

There's more than one way...

- Joint-Embedding Architecture (Contrastive)
  - Minimize/maximize distances between representations of similar/dissimilar jets
  - SimCLR, JetCLR, AnomalyCLR, DarkCLR, RS3L
- Generative Architecture
  - Directly generate partial or full jets
  - MPM, OmniJet-α, MPMv2
- Joint-Embedding Predictive Architecture
  - Complete the representation of jets
  - J-JEPA (Jet-based Joint-Embedding Predictive Architecture)
  - Our approach, ***Particle Joint-Embedding Predictive Architecture (P-JEPA)***, inspired by I-JEPA
  - ***Could directly train on data***
  - ***No discrete tokenization needed***
  - ***Take full inputs, not only kinematics***

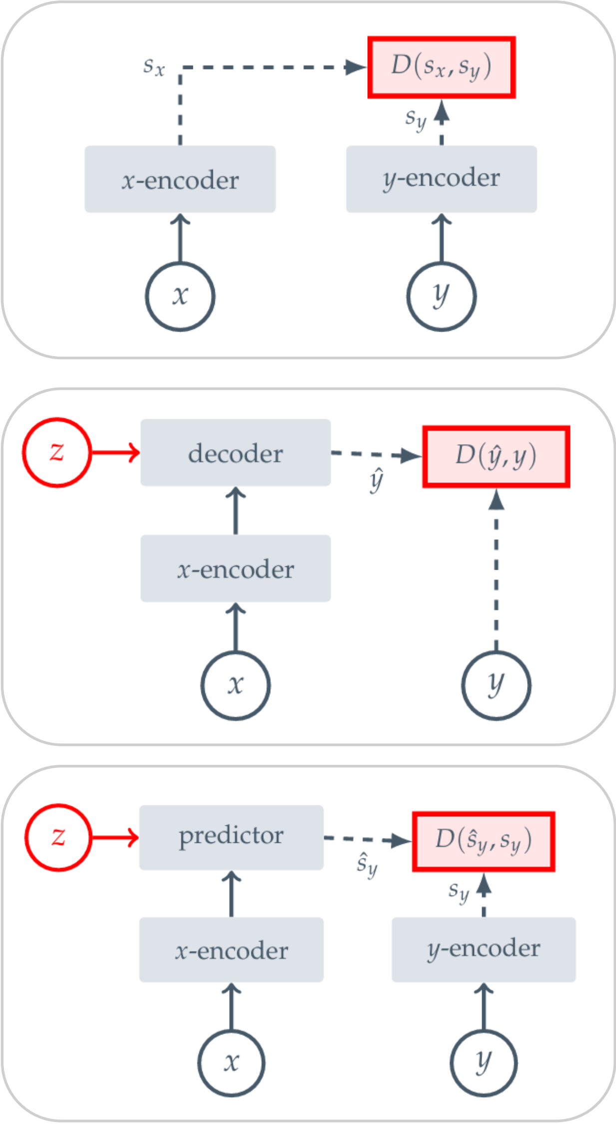
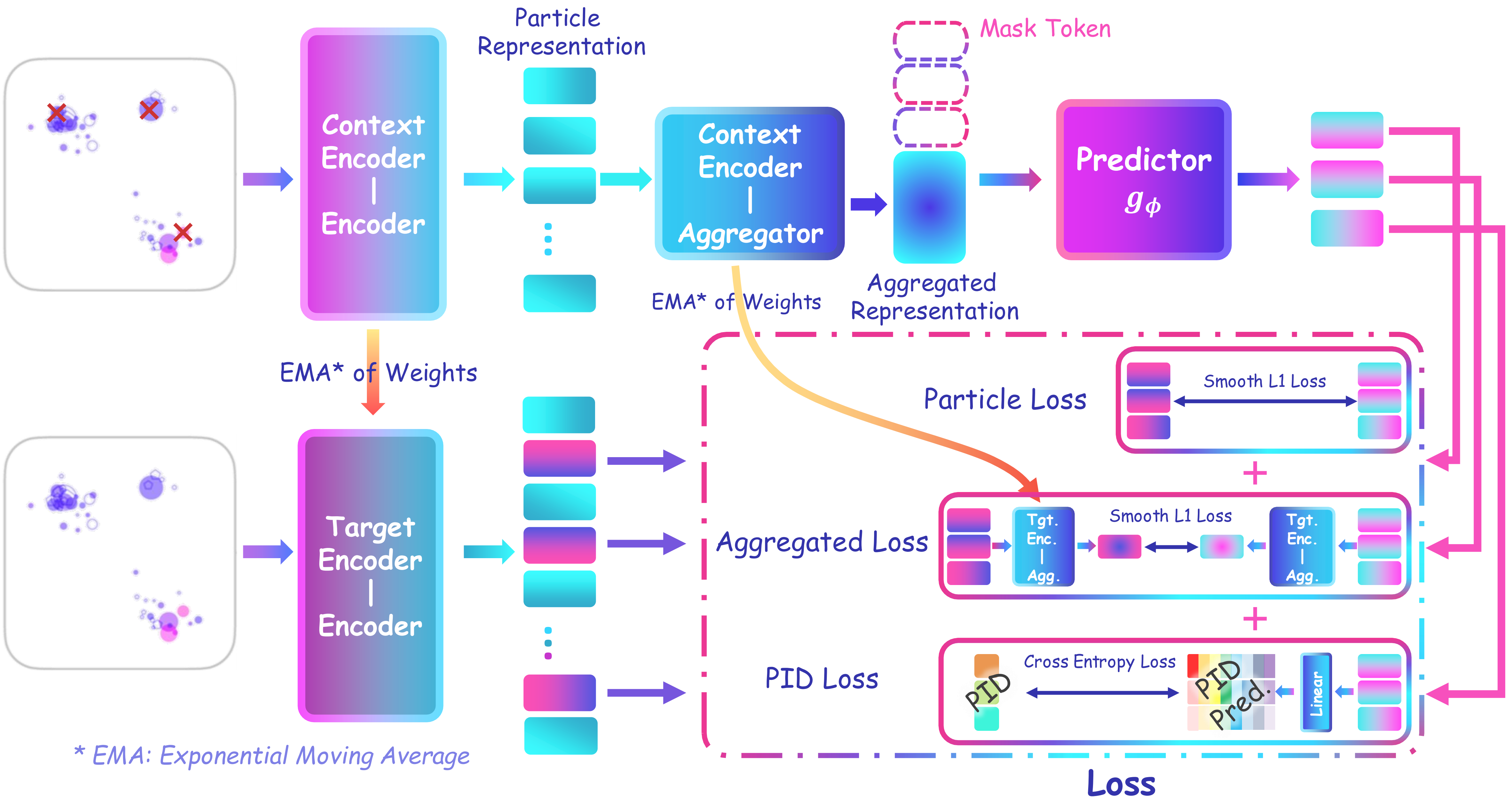


Figure Credit: 2301.08243

Bring the Concept to Life

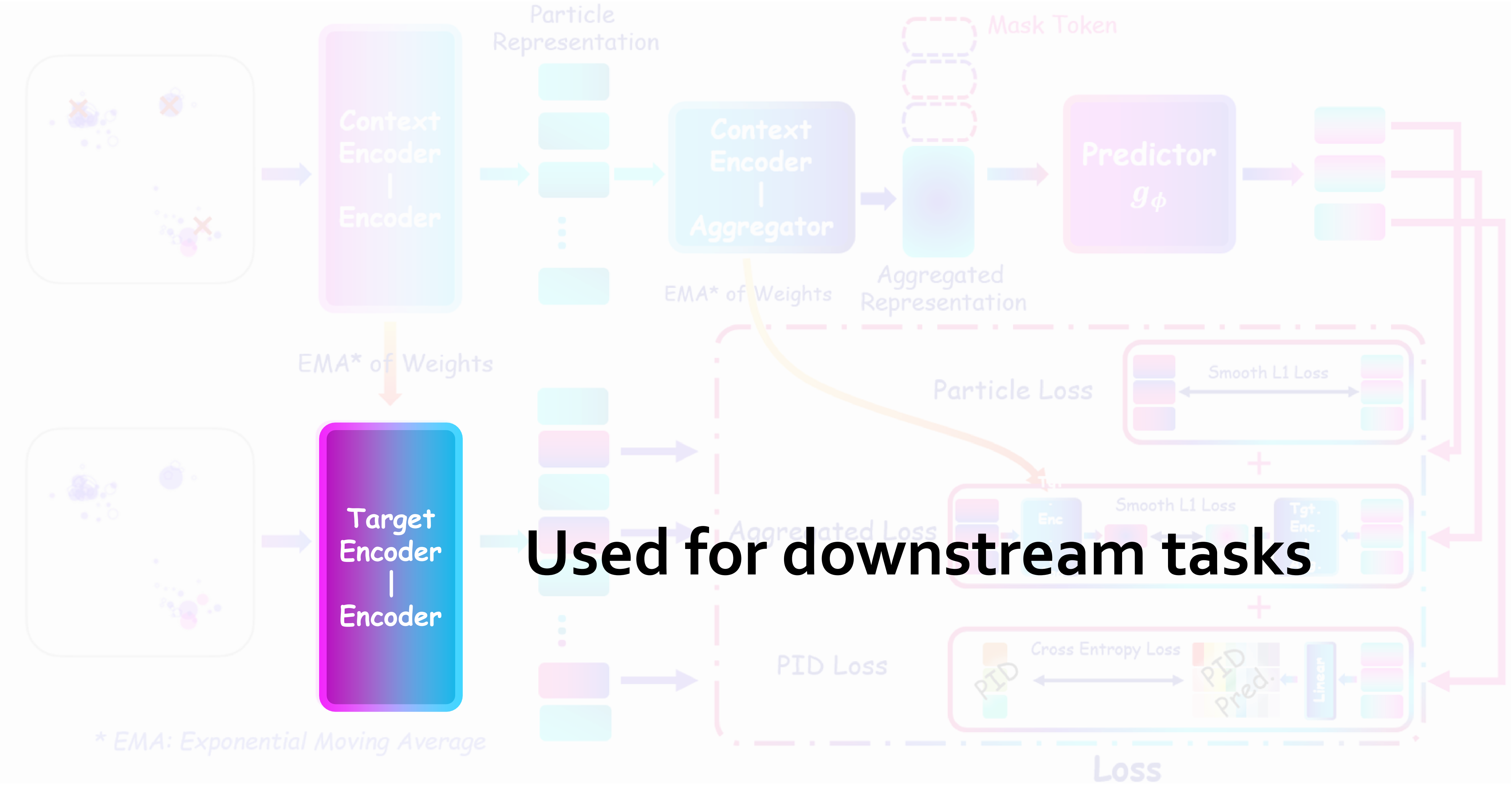
Implementation of the P-JEPA network

P-JEPA



shudong.wang@cern.ch

P-JEPA

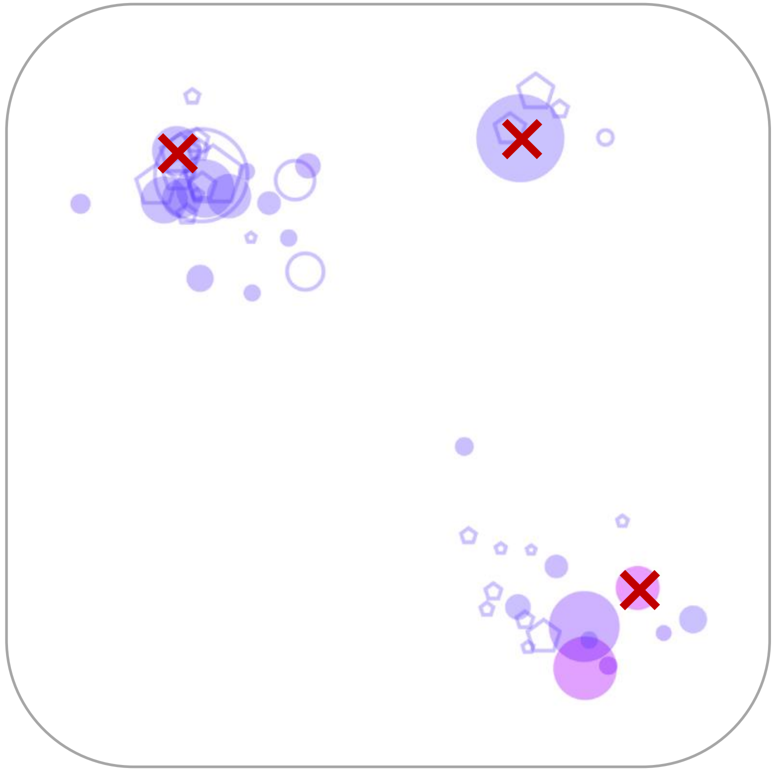


Used for downstream tasks

Particle Masking

Defining Context and Target for the Training

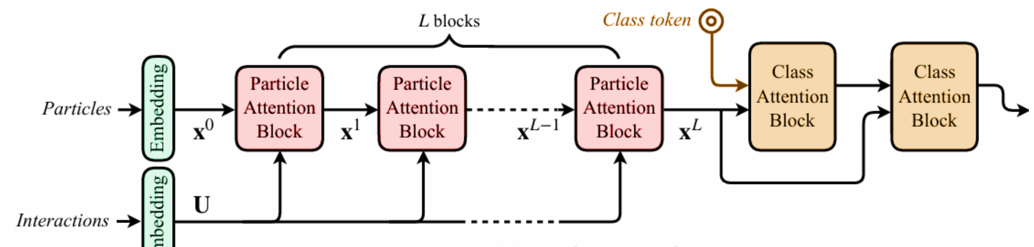
- Jets are consist of particles
  - Own features: kinematics, charge, PID, track info
  - Correlation between each other: angular distance, invariant mass…
- Masking sets the task for training
  - Random masking of **30% - 50%** of particles in a jet
  - Remaining particles provide "context" information for prediction
- Learning jet representations through predicting masked particles' representations



Context Encoder and Predictor

Encode Context Into Latent Space and Predict the Representation of Masked Particles Based on It

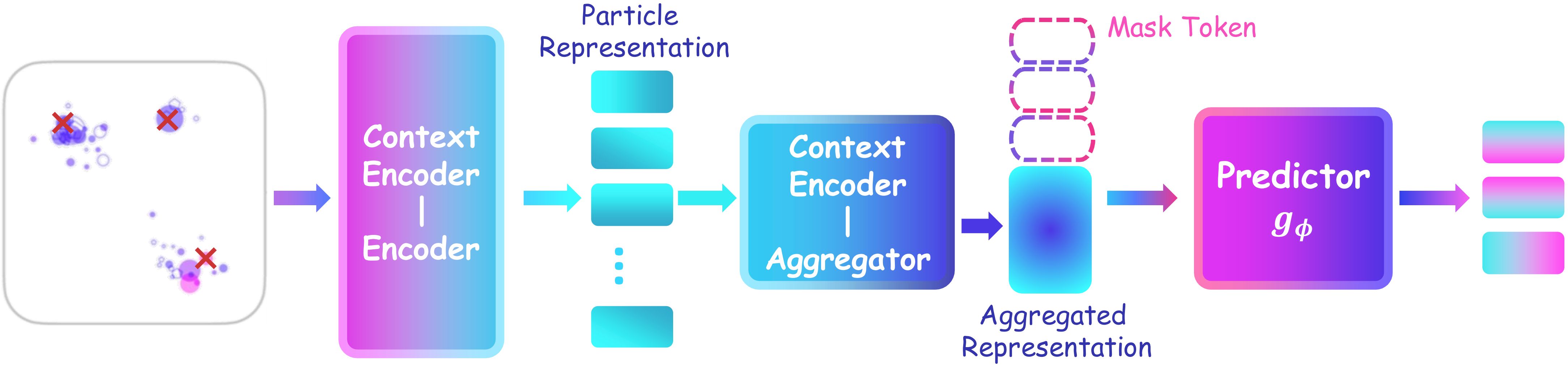
- ParticleTransformer as backbone



(a) Particle Transformer

	Context Encoder	Predictor
Embed Dims	(512, 512, 512)	192
Pair Embed Dims	(64, 64, 64)	/
Num Heads	8	6
Num Blocks	16	4
Num Class Blocks	2	/

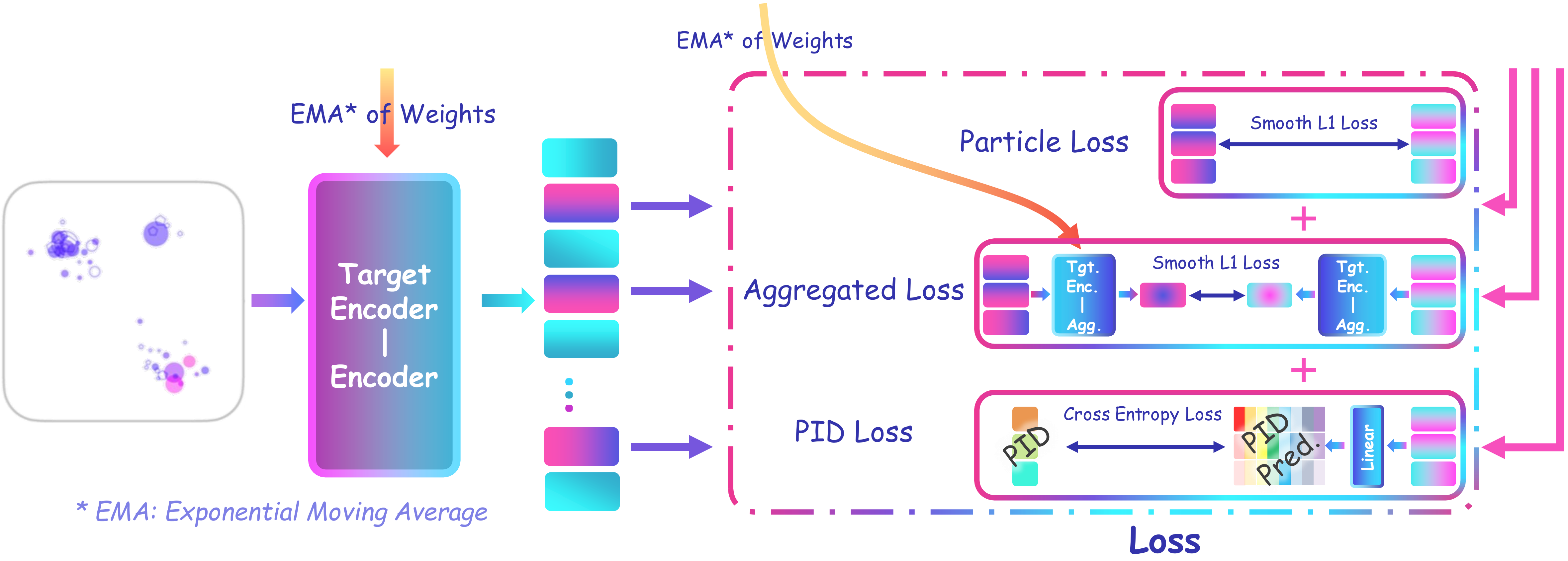
- Predictor is narrower and shallower than encoders
  - Predict masked particles' representations using context and auxiliary info (mask token)
- Encoder and predictor are trained simultaneously



Target Encoder and Loss

Encode All Particles Into Latent Space and Compare with the Predicted Representation

- Context Encoder and Target Encoder share the same architecture
  - Weights of Target Encoder are updated via an exponential moving average (EMA) of the Context Encoder's weights



* EMA: Exponential Moving Average

- Loss = Particle Loss + Aggregated Loss + PID Loss

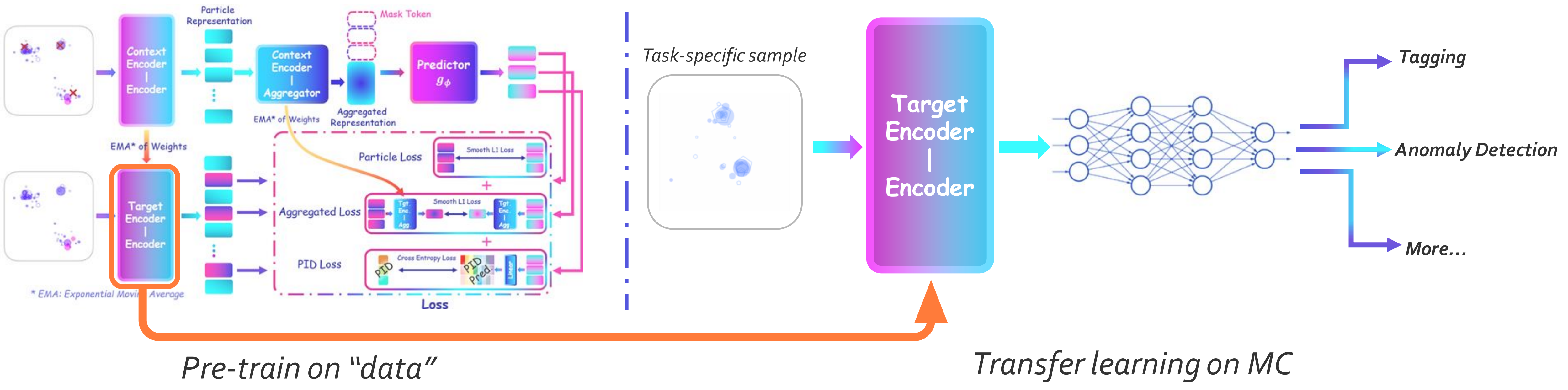
Does it work?

Experiments and Preliminary Results

Pre-training and Transfer Learning

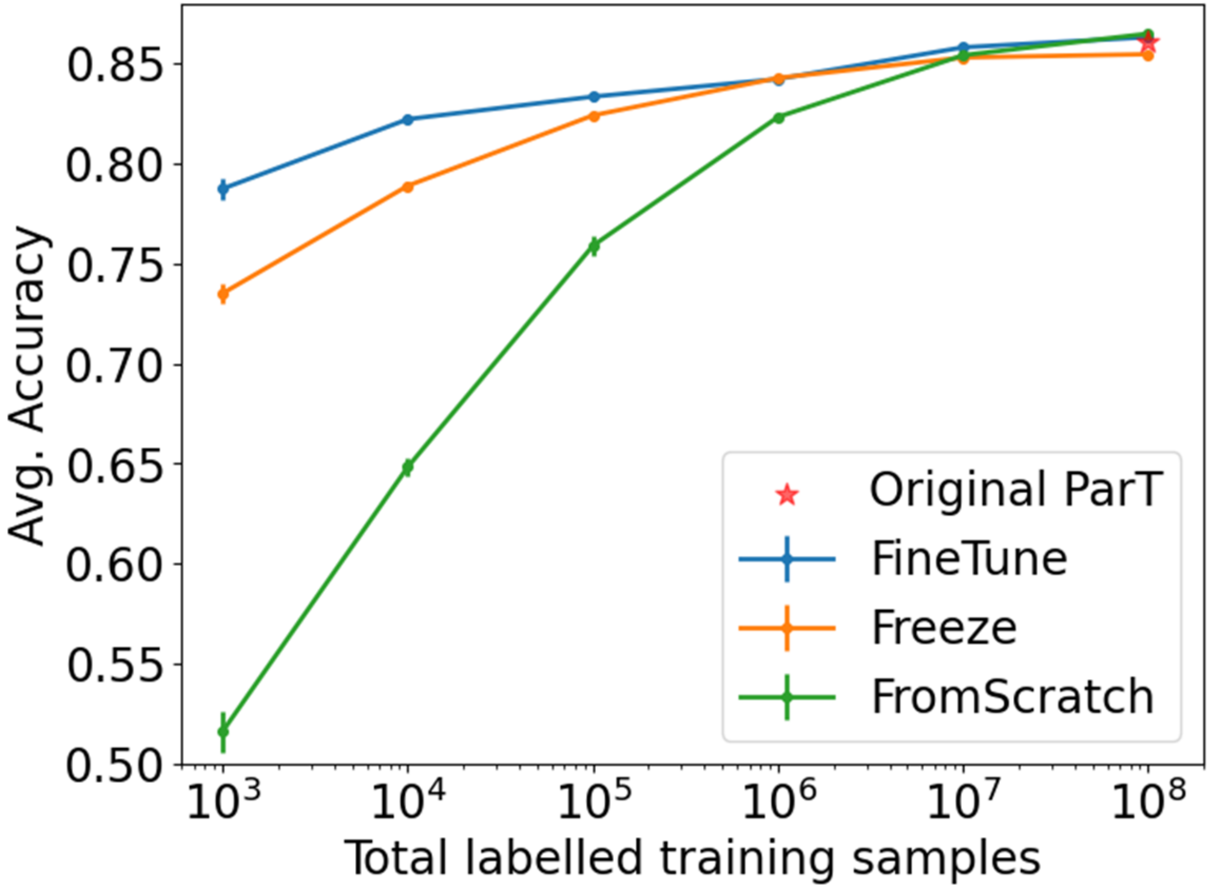
General knowledge goes a long way

- Performance evaluated with pre-training + transfer learning pipeline
- Foundation P-JEPA model pre-trained on "data"
  - Full **JetClass** training dataset (100M jets), but not using labels
- Transfer learning to specific task
  - Different downstream models share the same target encoder (jet representation)



Application: Jet Tagging

- Transfer learning for jet tagging
  - 10-class jet classification on **JetClass**



FineTune:

Encoder allowed **slightly updating** when tagging task is trained

Freeze:

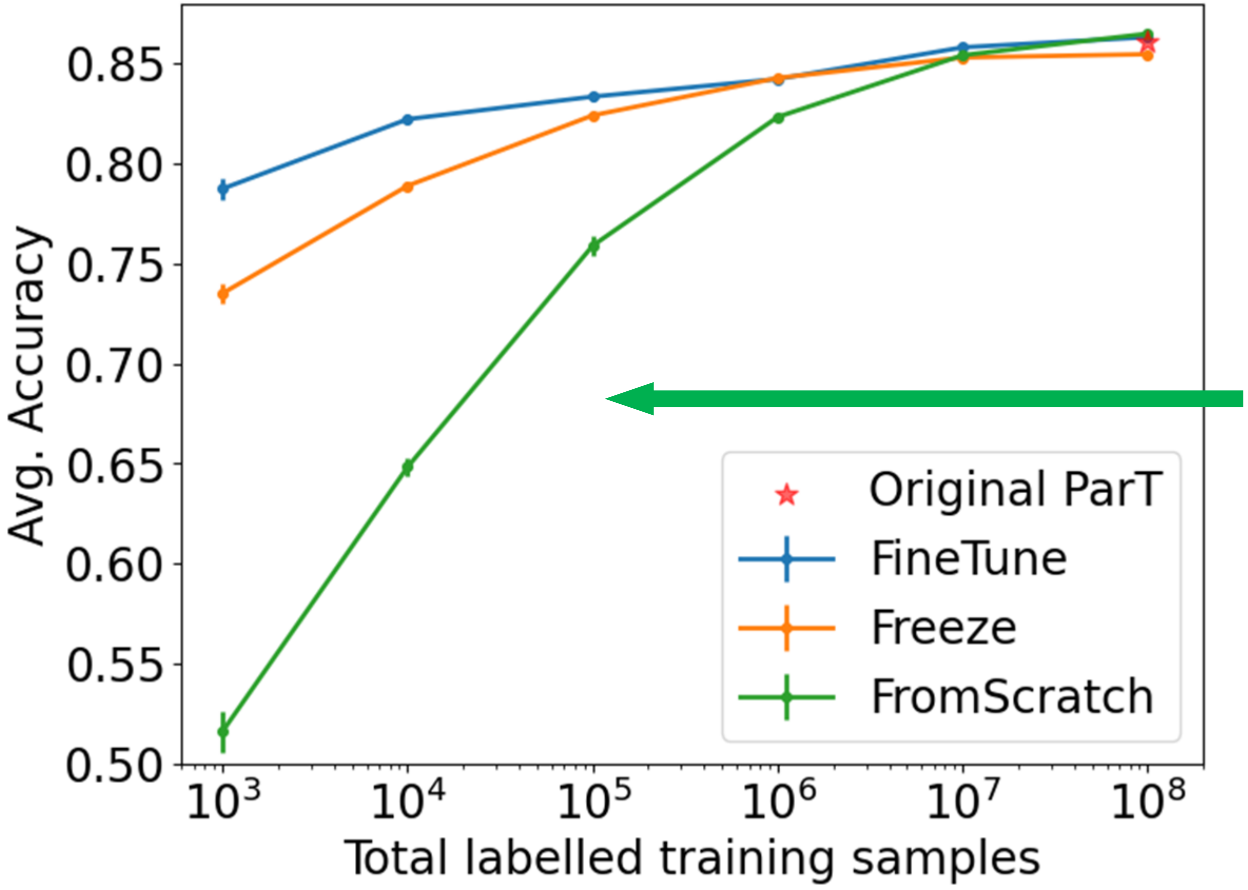
Encoder **fixed** when jet tagging task is trained

FromScratch:

Identical network architecture but training started with **randomly initialized weights**

Application: Jet Tagging

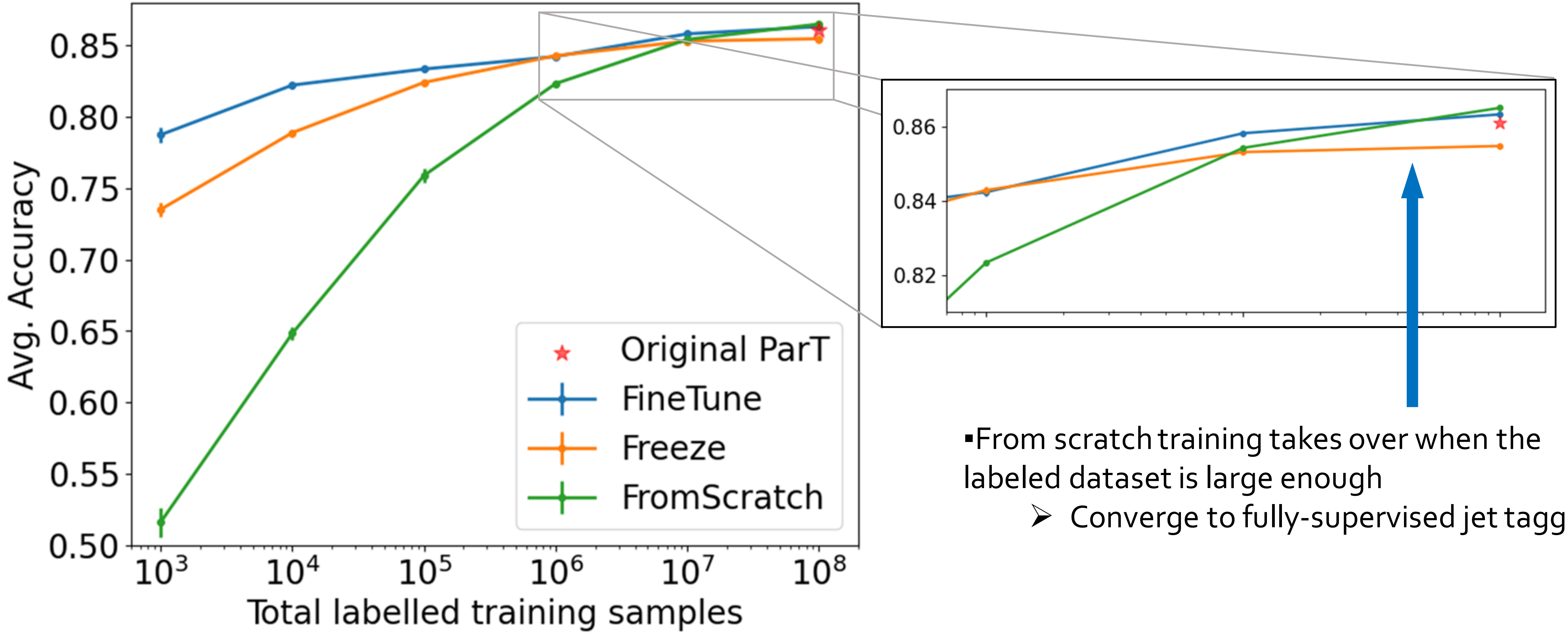
- Transfer learning for jet tagging
  - 10-class jet classification on **JetClass**



- Pre-training + transfer learning gives a significant performance boost with very limited number of labeled samples **(as low as 100 jet/class)!**
  - ➢ Benefit from jet representation learned via **self-supervised learning**

Application: Jet Tagging

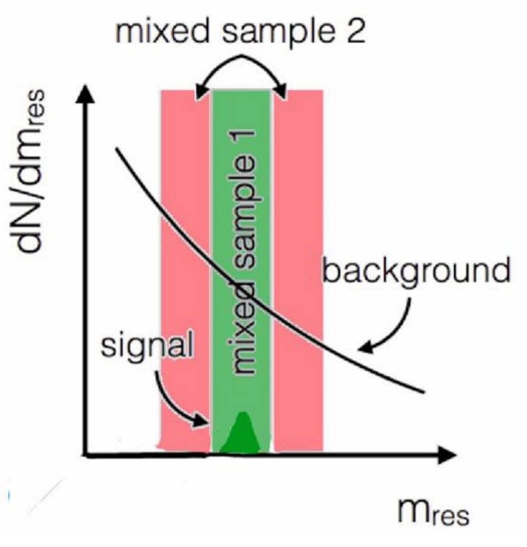
- Transfer learning for jet tagging
  - 10-class jet classification on **JetClass**

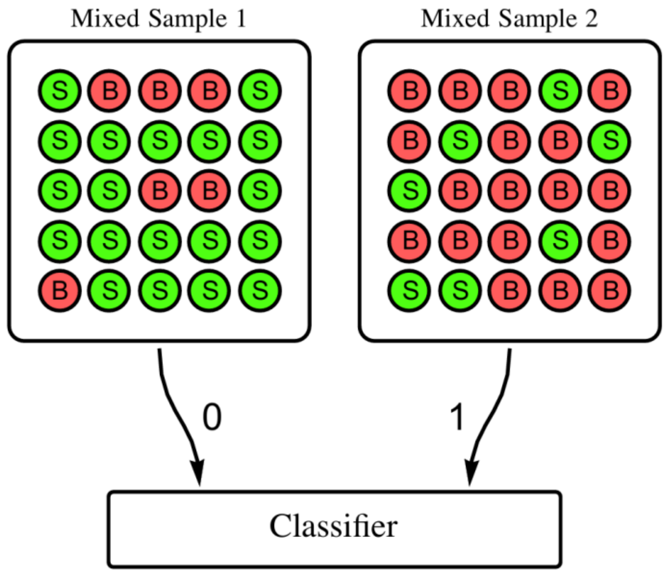


- From scratch training takes over when the labeled dataset is large enough
  - ➢ Converge to fully-supervised jet tagging

Application: Anomaly Detection

- Test the effectiveness of pre-trained jet representations on anomaly detection
  - Model independent search for new physics signals
  - Share same framework of AD study in [Sophon](), originated from [CWoLa]() (classification without labels)
    - CWoLa: allow to detect anomalies purely from data
    - train a classifier for mass window vs mass sideband (mixed sample 1 vs 2)

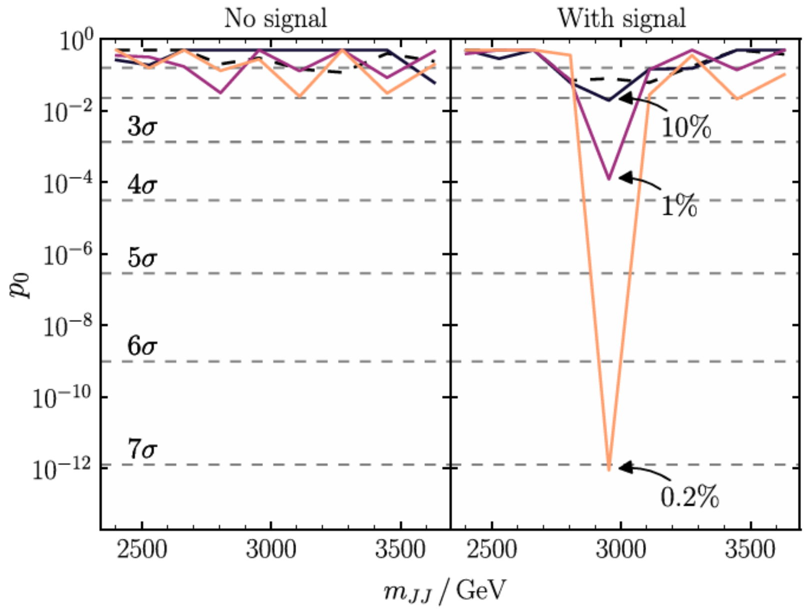




[*Figure Credit*]()

can discover W’ → Wϕ → WWW signals

see 2σ → 7σ improvement



[*PRL, 121 (2018) 24, 241803*]()

Application: Anomaly Detection

- **Sophon** (Signature Oriented Pre-training for Heavy-resonance ObservatioN)
  - A model pre-trained on the comprehensive JetClass-II dataset (188 classes) with supervision
  - Use the Sophon model by performing transfer learning or constructing discriminants from selected output scores

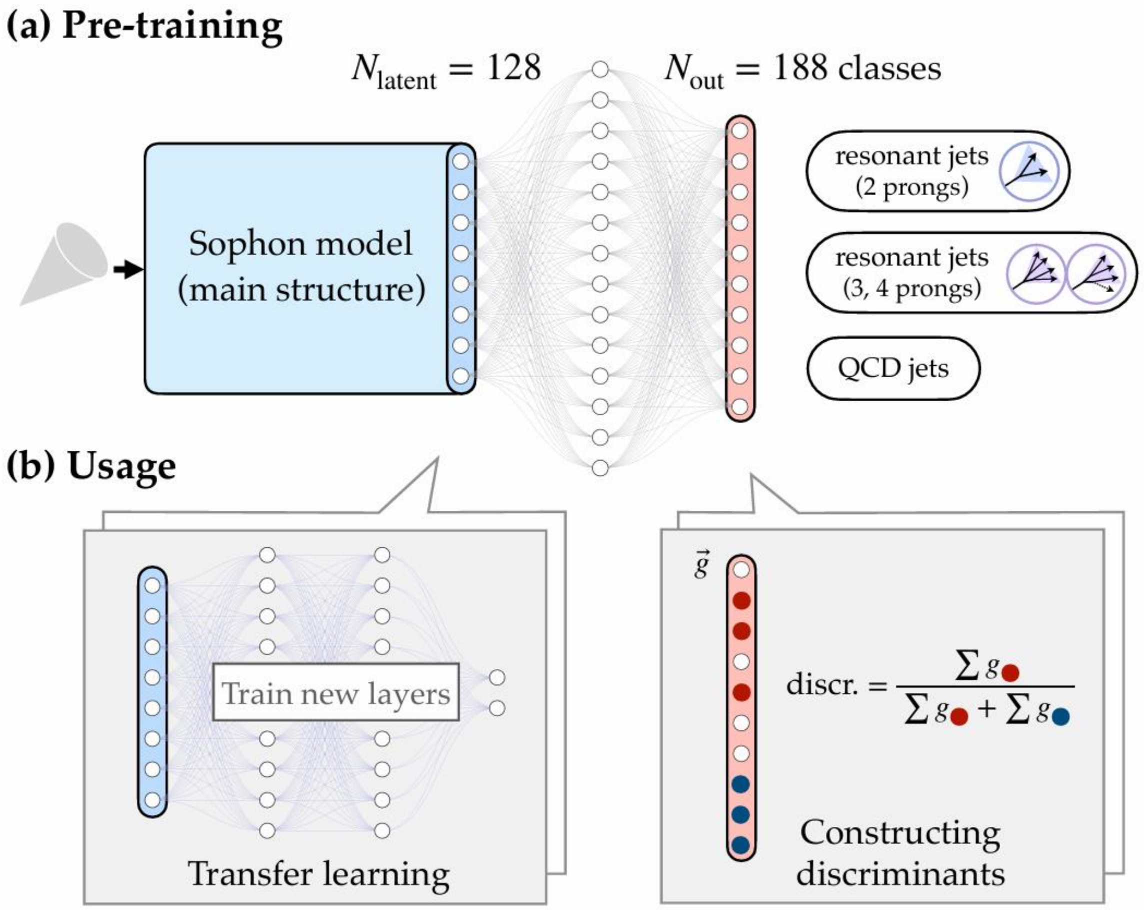
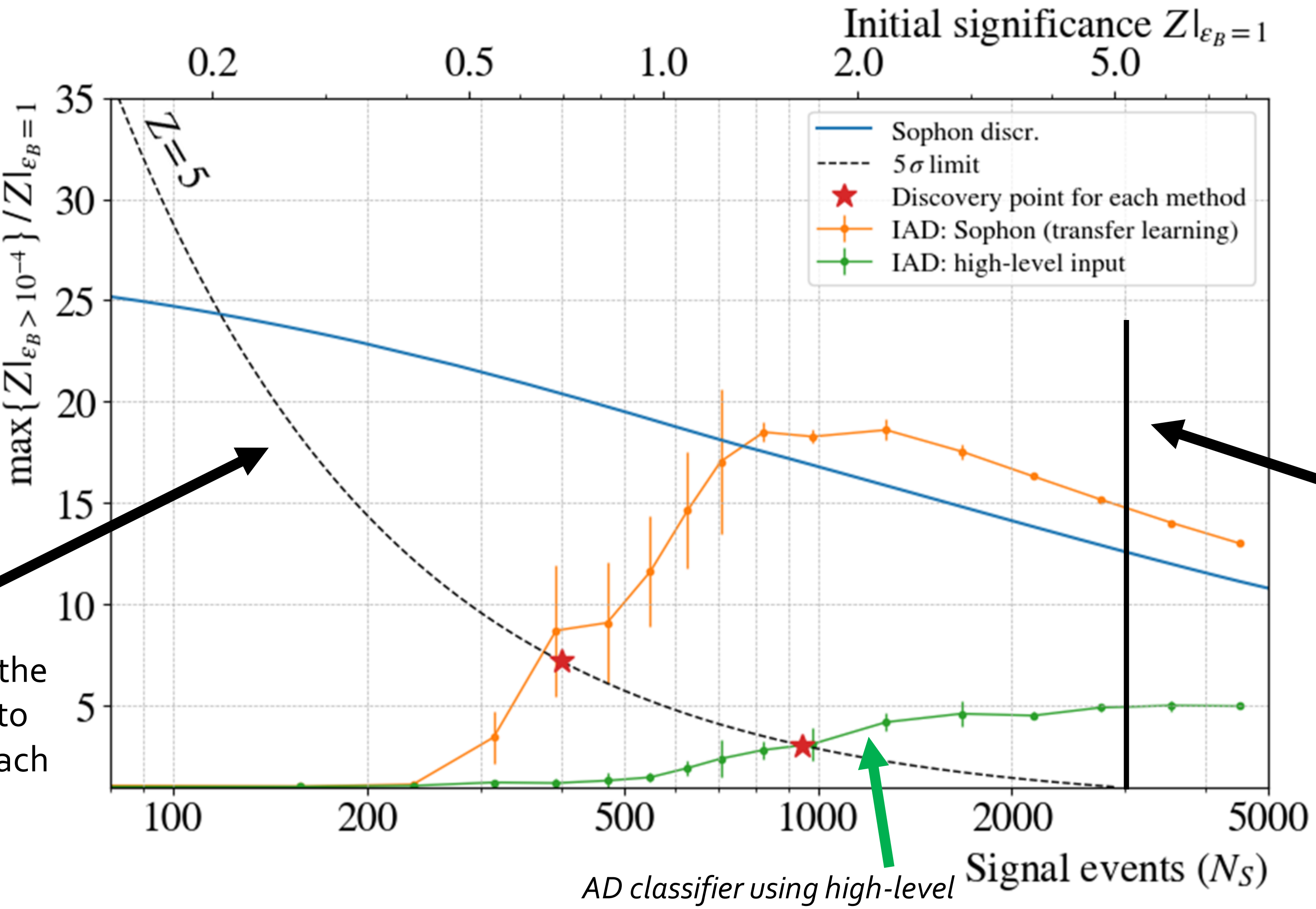


Figure Credit

shudong.wang@cern.ch

Application: Anomaly Detection

- **Sophon** (Signature Oriented Pre-training for Heavy-resonance ObservatioN)



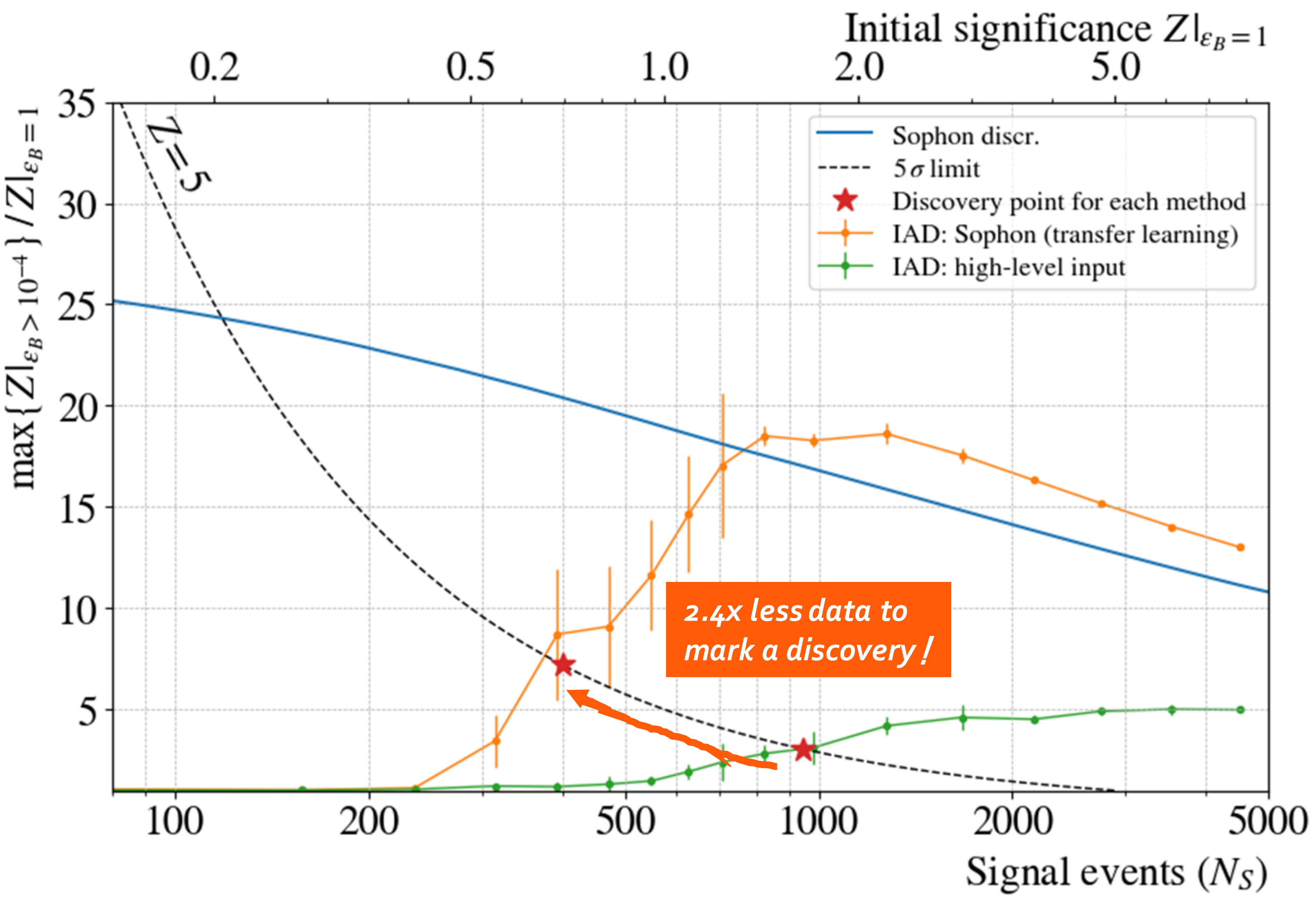
"How much does the significance need to be increased to reach the 5σ discovery"

"If signal events reach this point, **with initial Z=5**, then we have already discovered the signal without needing to make a cut"

AD classifier using high-level observables
[PRL, 121 (2018) 24, 241803]

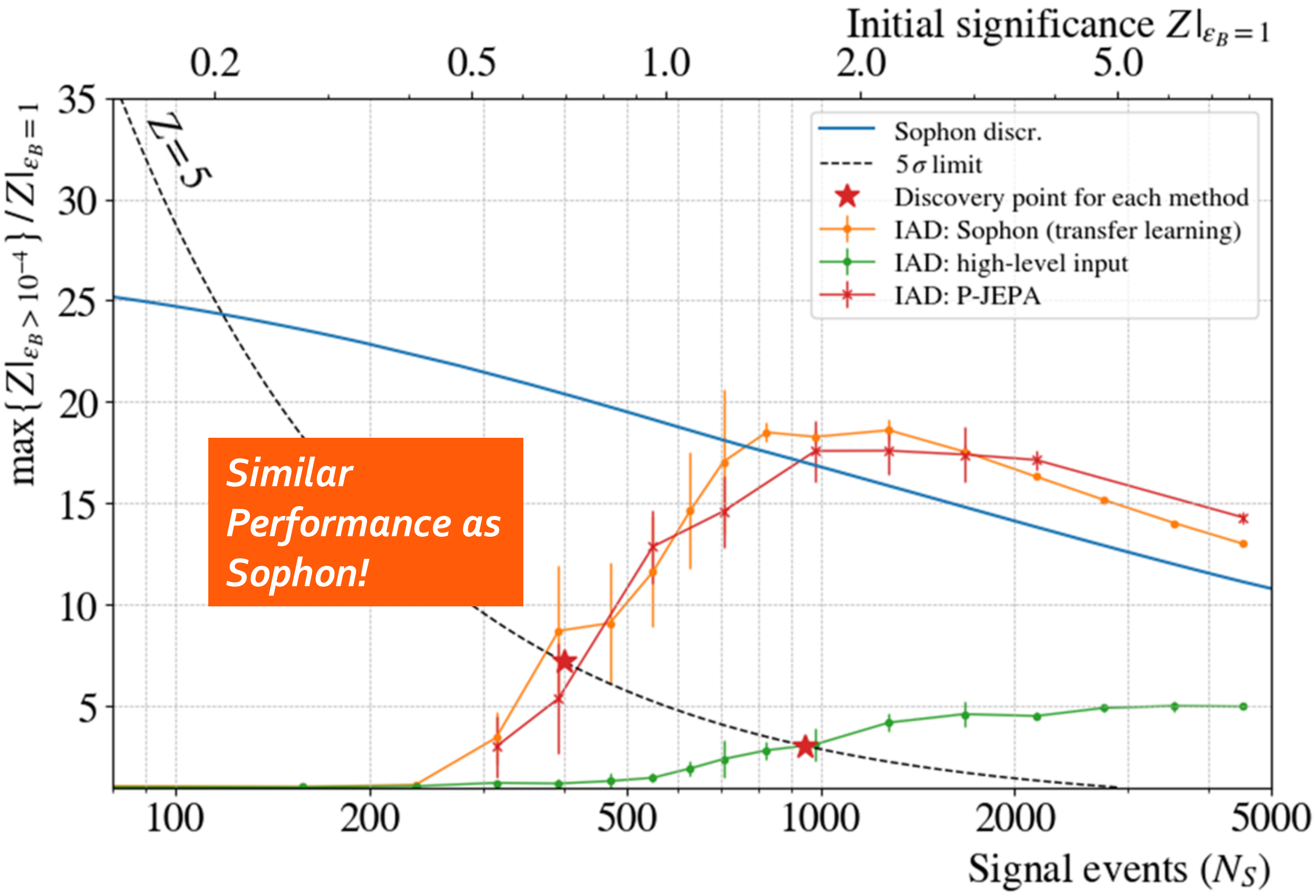
Application: Anomaly Detection

- **Sophon** (Signature Oriented Pre-training for Heavy-resonance ObservatioN)



Application: Anomaly Detection

- Using the output of P-JEPA target encoder as input to train the AD classifier



Summary

- Proposed the P-JEPA network for self-supervised learning on jets
- Promising performance showed on downstream jet tagging and anomaly detection tasks
- Main take away:
  - Effective jet representations can be learned from unlabeled dataset!

Outlook

- Using novel & performant algorithm as backbone (e.g. L-GATr)
- P-JEPA as a Foundation Model (more downstream tasks to be tested)
- Uncertainty-free or calibration-free jet tagging (ultimate goal though still long way to go)

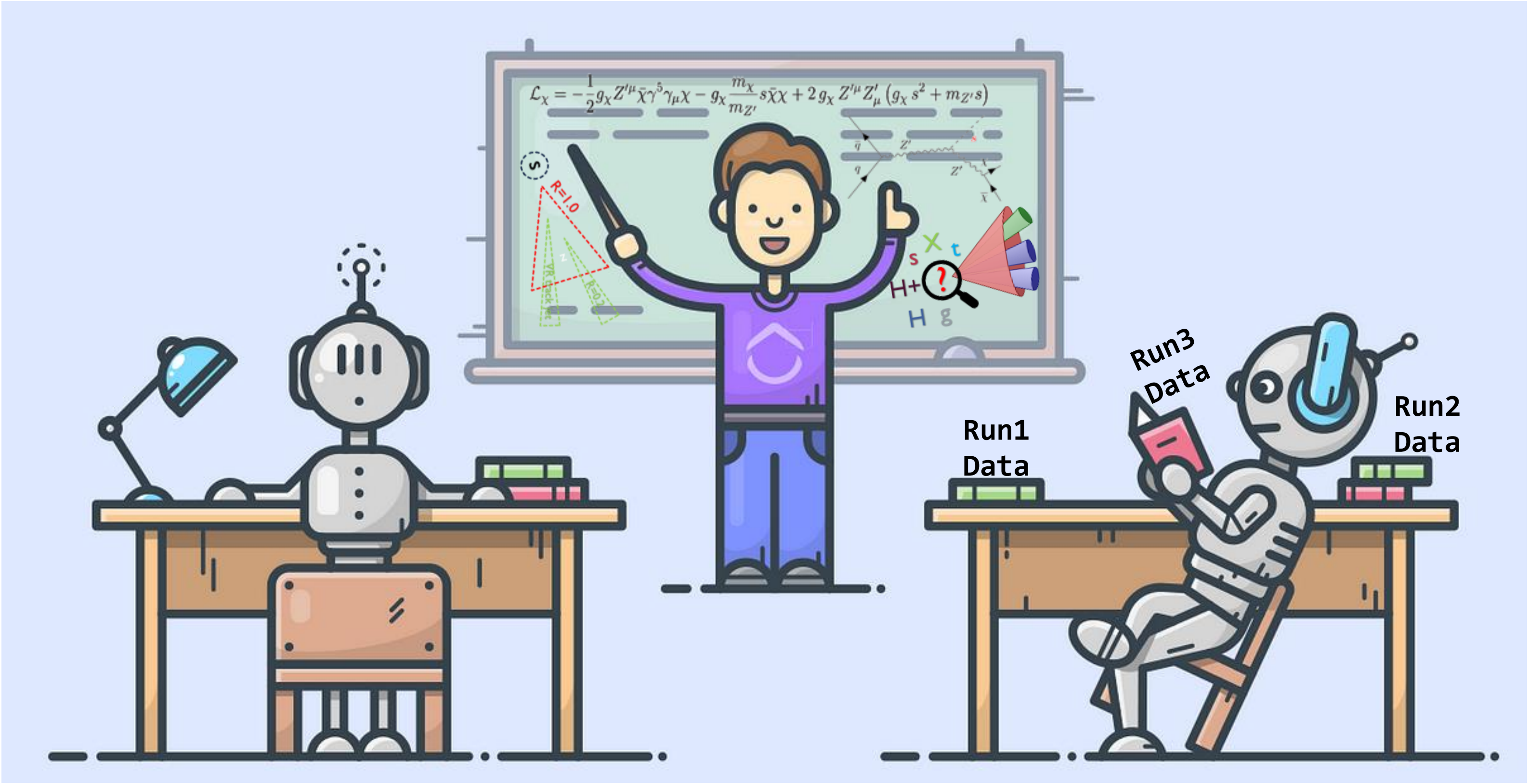


Figure modified from @Srinivas Rao

shudong.wang@cern.ch