EveNet: Towards a Generalist Event Transformer for Unified Understanding and Generation of Collider Data

Yulei Zhang

University of Washington, Seattle

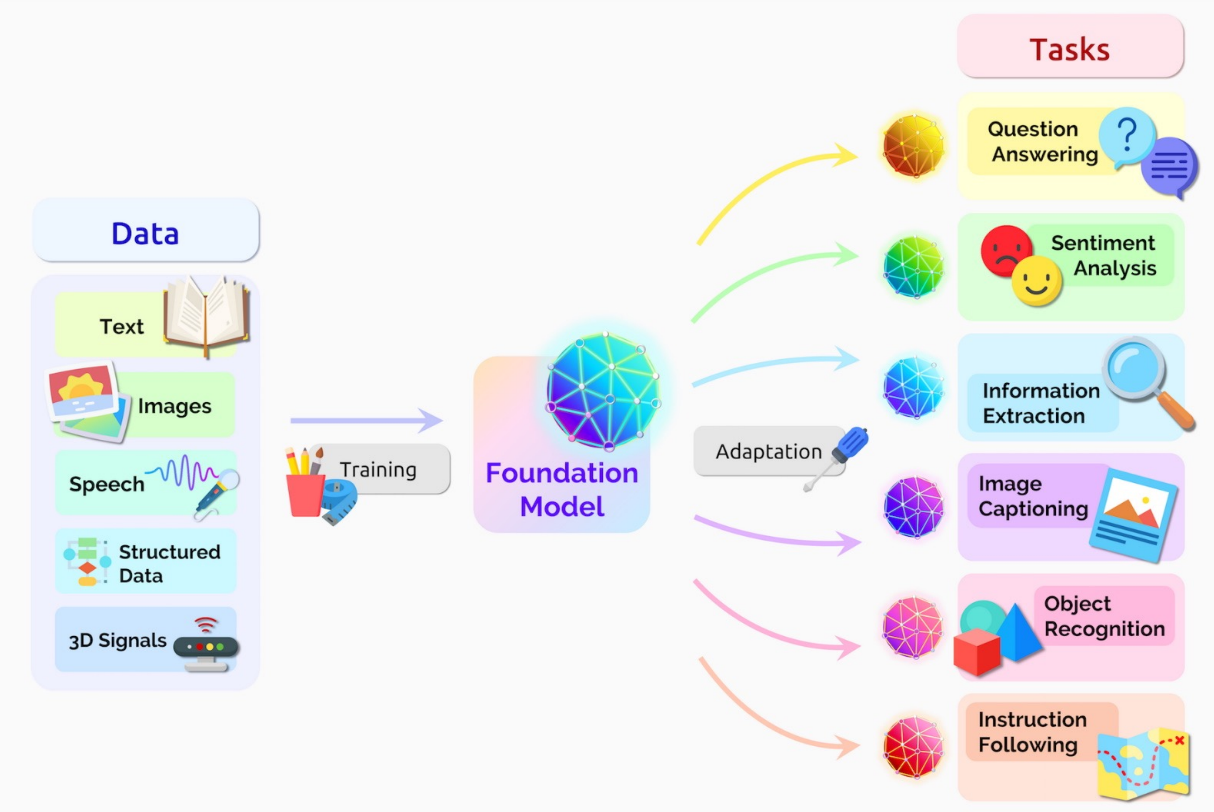
Team: Ting-Hsiang Hsu, **Qibin** Liu, **Yuan-Tang** Chou, **Wei-Po** Wang, **Yue** Xu, **Haoran** Zhao, **Bai-Hong** Zhou, **Shu** Li, **Benjamin** Nachman, **Shih-Chieh** Hsu, **Vinicius** Massami Mikuni, **Yulei** Zhang

July 10th, 2025

Foundation Model

arXiv: 2108.07258

A *foundation model* is a model trained on broad data at scale that can be adapted (fine-tuned) to a wide range of downstream tasks. It is *not* a fully complete model in itself, but a *foundation* — a starting point for building task-specific models.



NVIDIA blog: What are foundation models?

📈**Emergence**:

- New behaviors from scale

🧩**Homogenization**:

- One model, many tasks

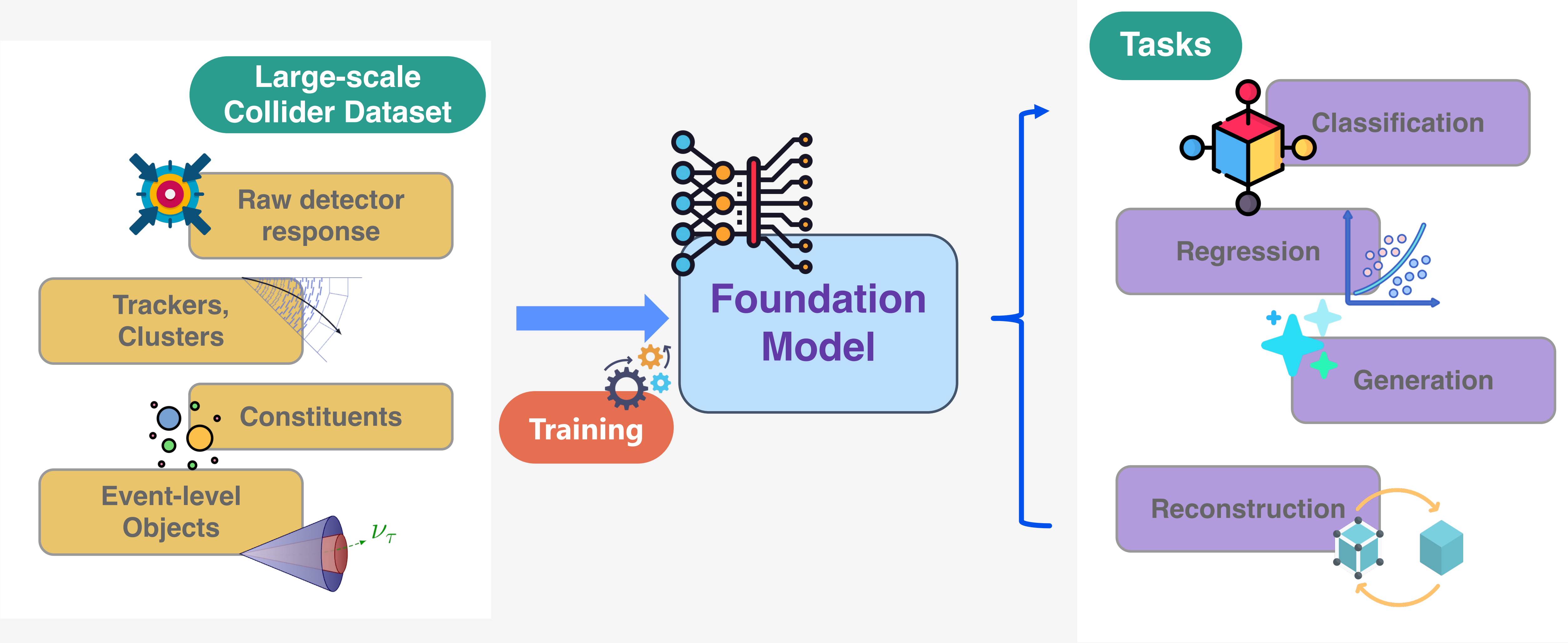
🔄**Transferable representations**:

- Pretrain once, reuse anywhere

🌐 **Multimodal potential**:

- Works across data types

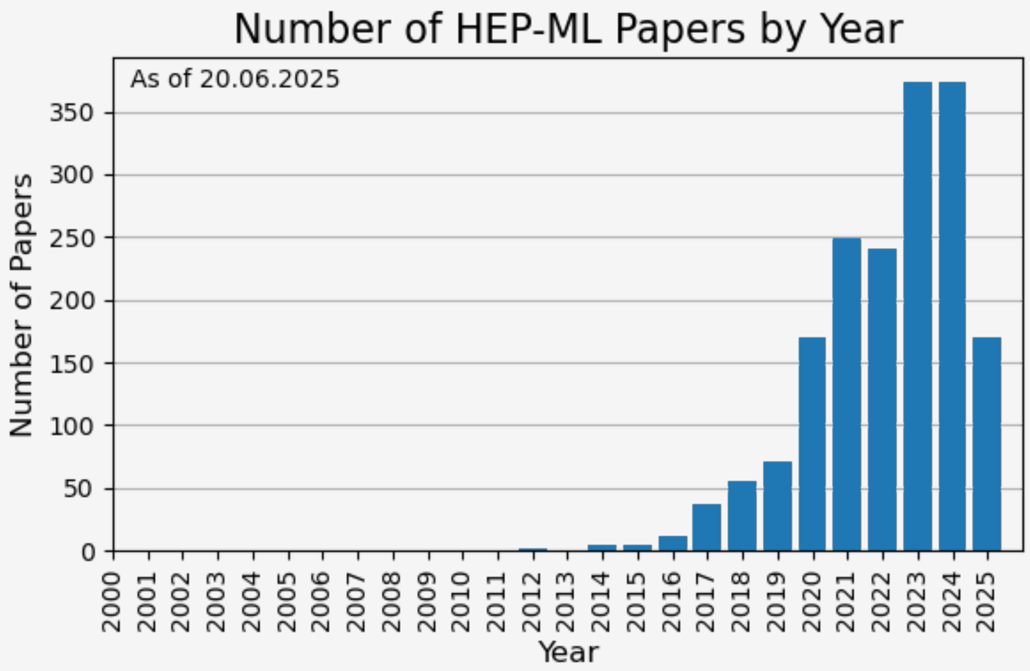
Foundation Model in HEP



Machine Learning + HEP

A Living Review of Machine Learning for Particle Physics

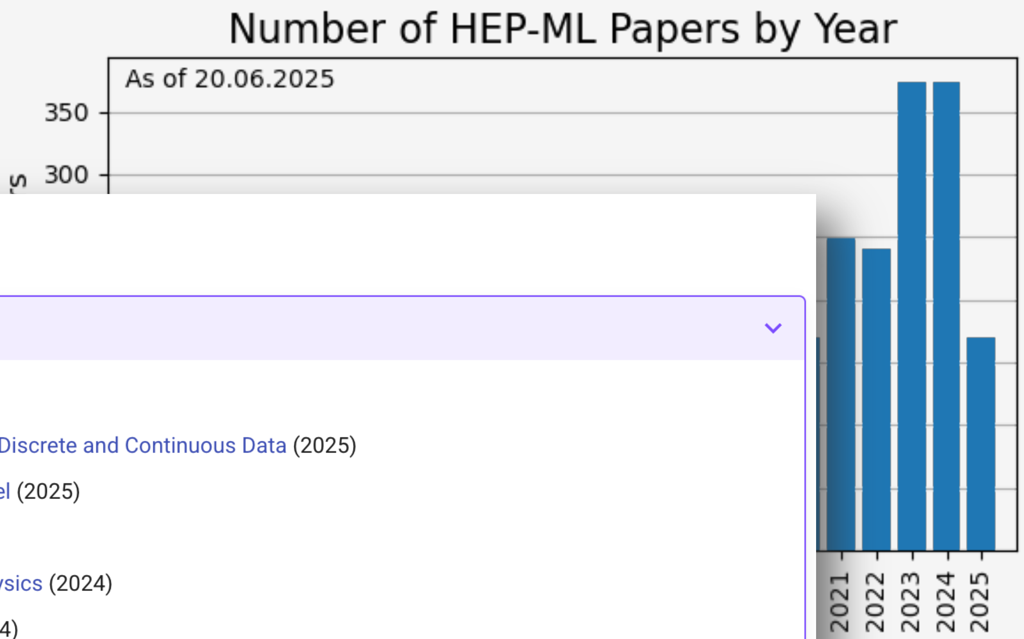
- Classification
- Regression
- Decorrelation Methods
- Equivariant Networks, PINNs, KANs
- Generative Models
- Anomaly Detection
- **Foundation Models**
- Simulation-based Inference
- …



Machine Learning + HEP

A Living Review of Machine Learning for Particle Physics

- Classification
- Regression
- Decorrelation Methods
- Equivariant Networks, PI
- Generative Models
- Anomaly Detection
- **Foundation Models**
- Simulation-based Inferer
- …



Foundation Models, LLMs.

Foundation Models, LLMs.

- Large Language Models -- the Future of Fundamental Physics? (2025)
- Towards Foundation Models for Experimental Readout Systems Combining Discrete and Continuous Data (2025)
- Reconstructing hadronically decaying tau leptons with a jet foundation model (2025)
- A Method to Simultaneously Facilitate All Jet Physics Tasks [DOI] (2025)
- Aspen Open Jets: Unlocking LHC Data for Foundation Models in Particle Physics (2024)
- Pretrained Event Classification Model for High Energy Physics Analysis (2024)
- Towards a foundation model for heavy-ion collision experiments through point cloud diffusion (2024)
- Bumblebee: Foundation Model for Particle Physics Discovery (2024)
- Is Tokenization Needed for Masked Particle Modelling? (2024)
- OmniLearn: A Method to Simultaneously Facilitate All Jet Physics Tasks [DOI] (2024)
- Xiwu: A Basis Flexible and Learnable LLM for High Energy Physics (2024)
- Physics Event Classification Using Large Language Models [DOI] (2024)
- Re-Simulation-based Self-Supervised Learning for Pre-Training Foundation Models [DOI] (2024)
- OmniJet-α: The first cross-task foundation model for particle physics [DOI] (2024)
- Finetuning Foundation Models for Joint Analysis Optimization [DOI] (2024)

Can We build an Event-Level Foundation Model?

Could we resolve all event-level tasks with a single model?

Pre-trained Model 🧐

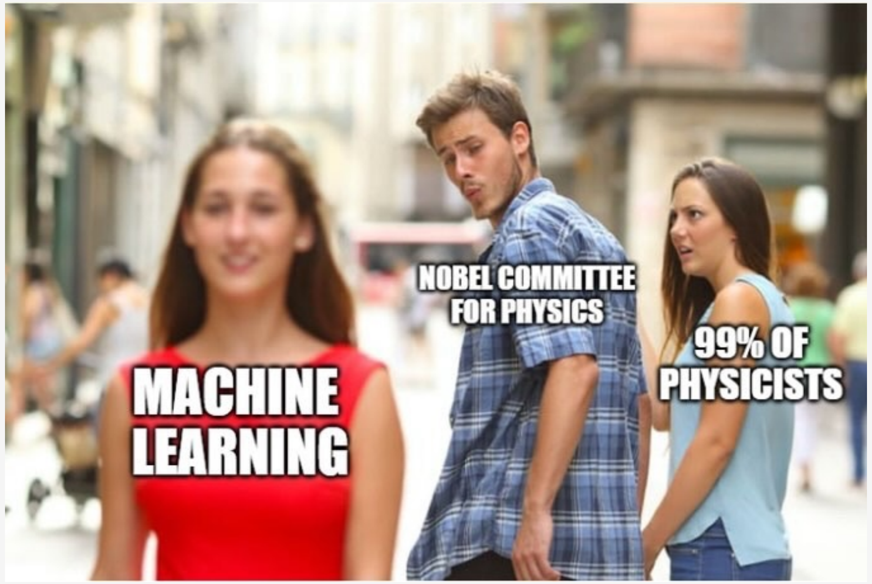
- Extensively pre-trained for general-purpose representations.
- **Lightly fine-tuned** for task-specific applications.
- Especially **effective** in scenarios with **limited** training data.

Foundation Model 🥳

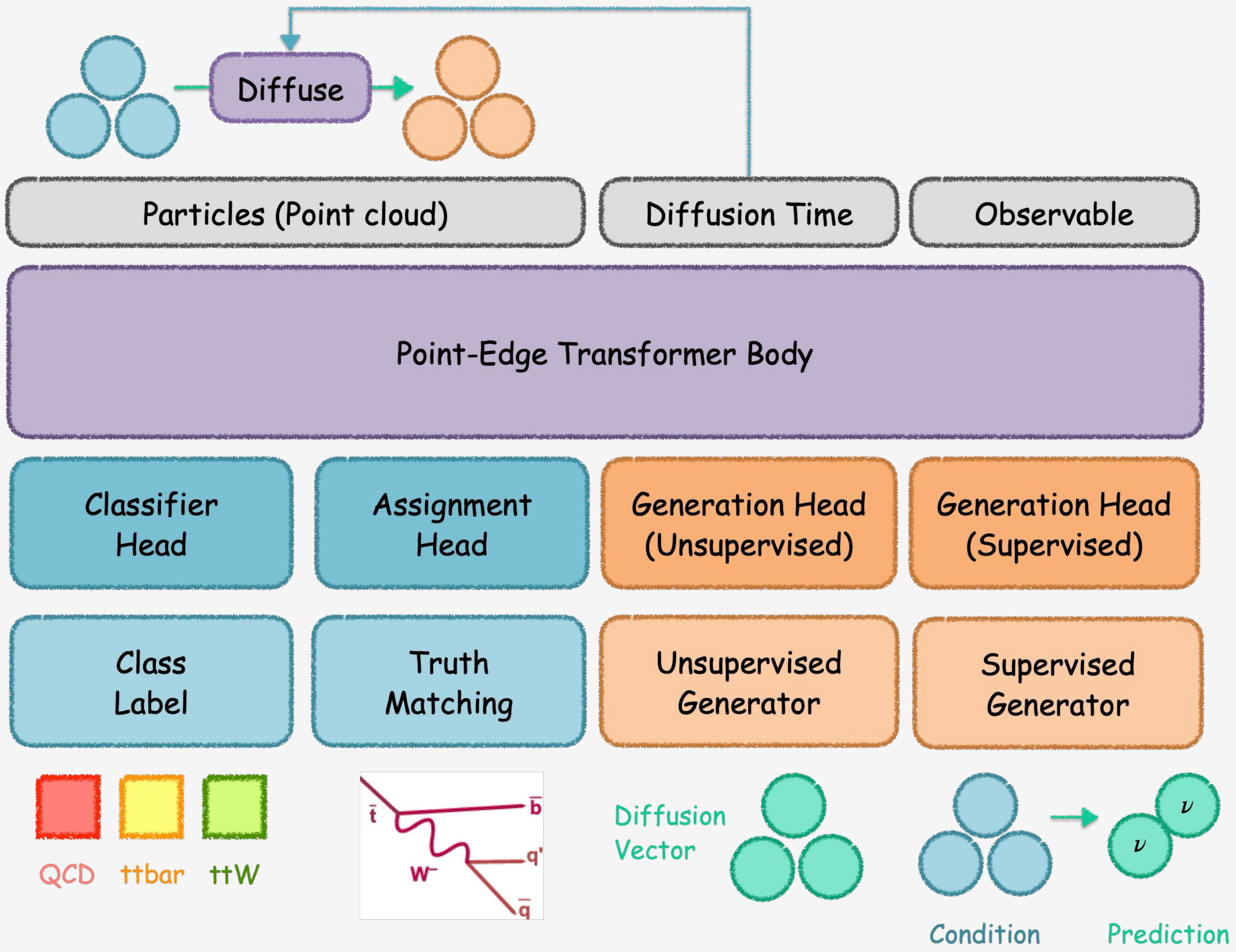
- Enables **a unified understanding** of HEP events.
- Designated to generalize across a wide range of tasks.

🔧 Core Ingredients of an Event-Level Foundation Model in HEP

- 🧠 **Generalist Embedding**: Shared event-level representation
- 🧩 **Multi-task Learning**: One model, many objectives
- 🔍 **Self-Supervised Pretraining**: Learns from data structure
- 📈 **Scalability**: Improves with more data + compute
- 🔁 **Transferability**: Fine-tune for new tasks easily



EveNet: Our Answer to Event-Level Foundation Models



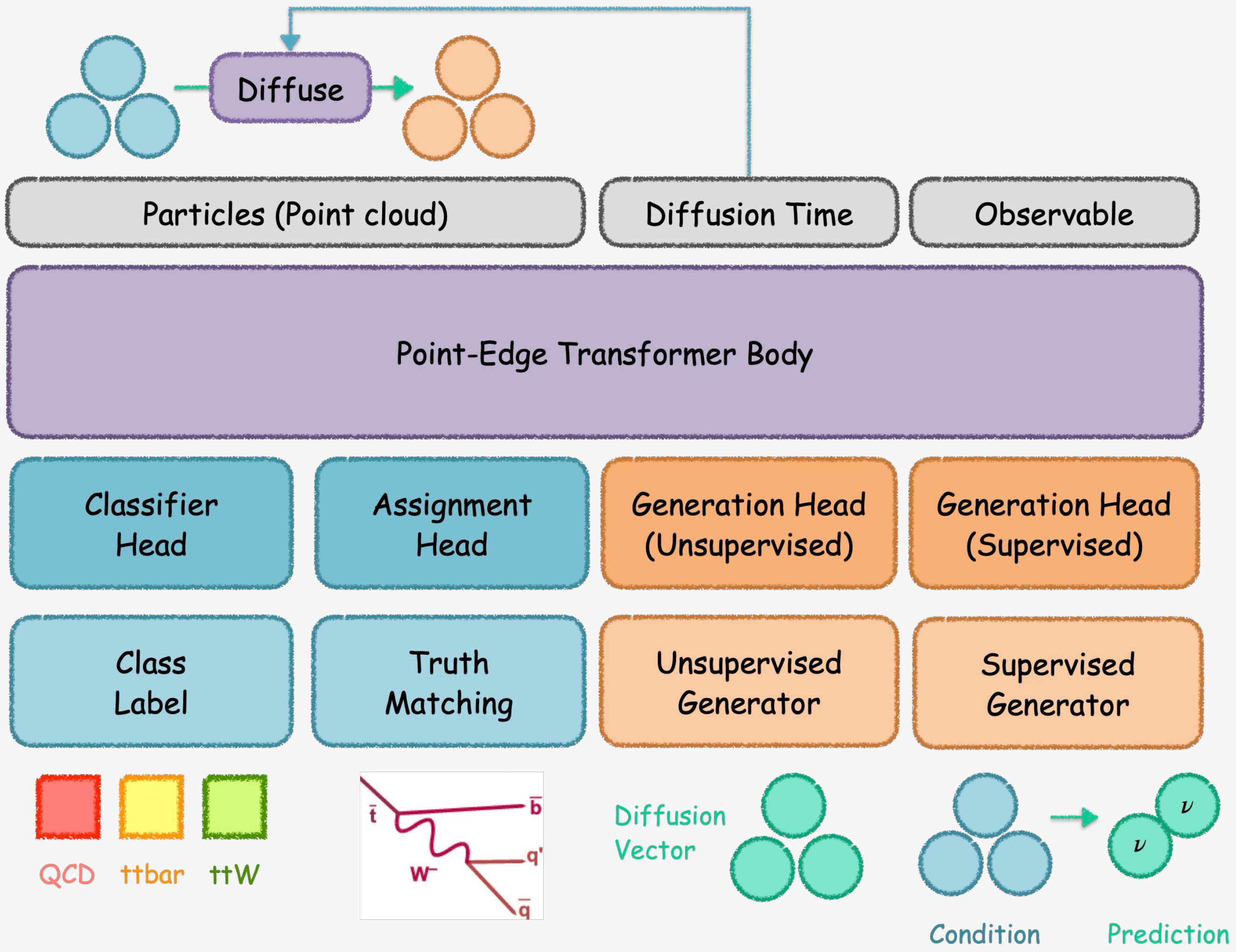
🧠 Body - Point-Edge Transformer:

- Models **both particles and their relationships** as a graph (points + edges)
- Captures **inter-particle interactions** and **global event structure**

🧱 Heads (multi-task outputs)

- **Classification**: Multi-class event classifier
- **Assignment**: Symmetry-aware mapping of objects to truth-level partons
- **Generation (unsupervised)**: Reconstructs masked particles via a diffusion model
- **Generation (supervised)**: Predicts missing objects (e.g., neutrinos)

What EveNet Sees and Learns?

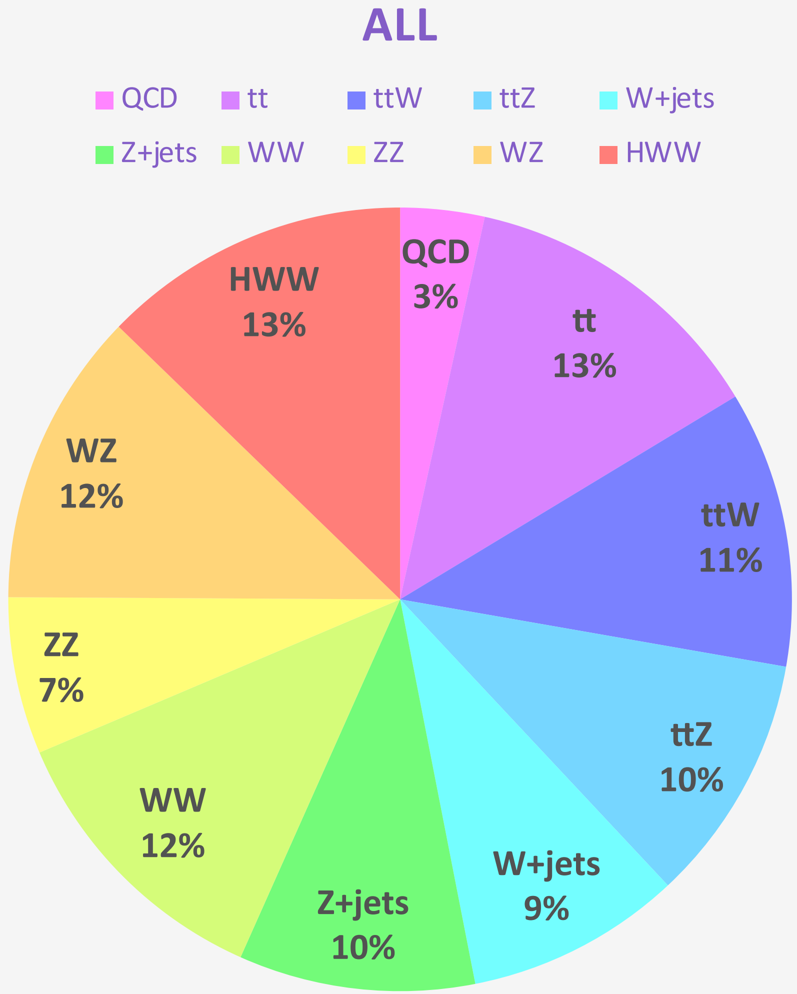


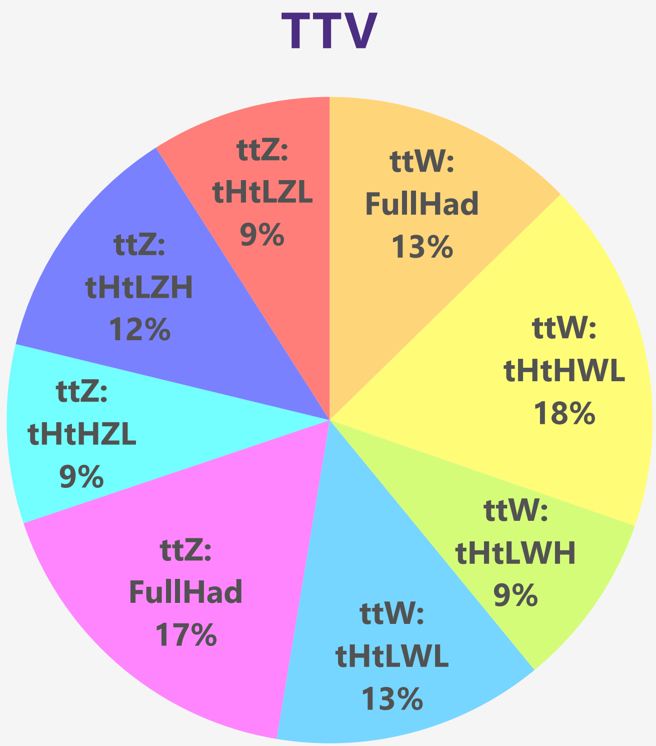
🔢 Input Representation

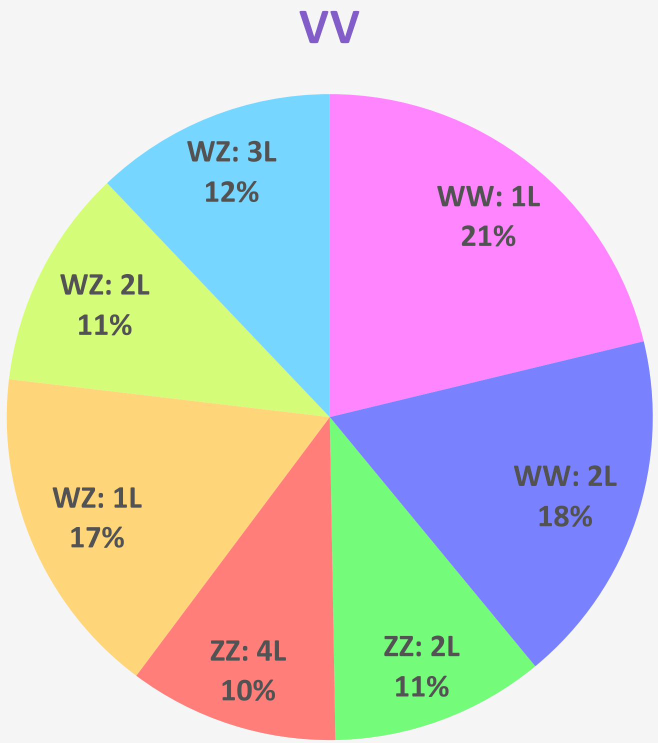
- 📌 **Particle Cloud (Up to 18 Particles per Event):**
  - Each particle is encoded with 7 features: **4-momentum, isbJet, isLepton**, and **charge.**
  - → (7 features × 18 particles, zero-padded)
- 🌐 **Global Features / Event Observables:**
  - Missing transverse energy
  - Number of leptons, number of jets
  - Invariant mass of visible objects
  - Scalar sums like **HT, ST**, etc.

The heads shown here are illustrative examples, designed to guide training, but **can be easily extended or replaced** for new downstream tasks, **while the body remains the core foundation**.

Dataset







- 🔧 **All processes** help learn diverse point cloud patterns for **classification** and **point cloud generation**.
- 🧠 ttV, VV, and HWW focus on harder tasks like **assignment** and **neutrino generation** due to their complex final states.

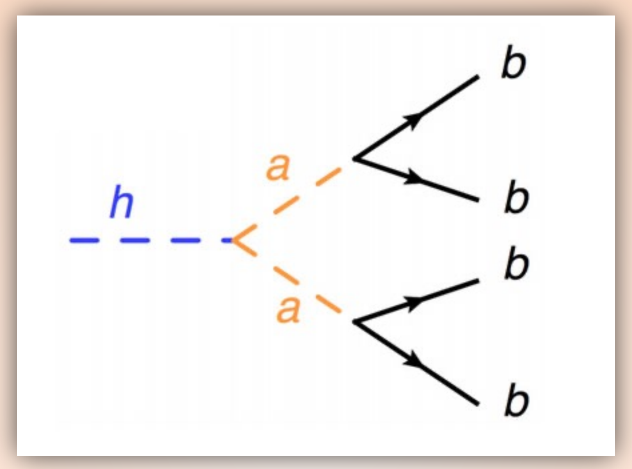
EveNet Wouldn't Train Itself—Thank You, Perlmutter!

🔋 **Scaling Up EveNet with Perlmutter**

- 🚀 **Training Setup:**
  - **Model:** EveNet-Pretrain (**40M**)
  - **128 nodes**
  - **512 GPUs**
  - **16,384 CPU cores**
- 📊 **Data Scale:**
  - **3000 million raw events**
  - **500 million effective events** after processing

    (*Only 100M used in this talk*)
  - Trained for **10 full epochs**

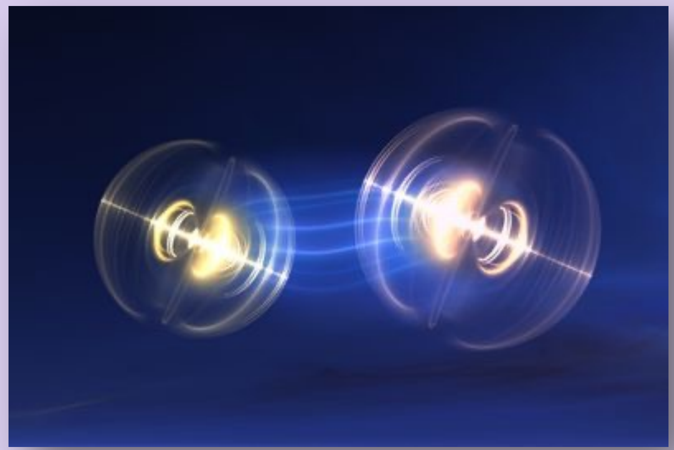
Downstream Applications of EveNet in Physics Analyses



Search for new physics

$H \to aa \to bbbb$

Assignment & Classification



Quantum Entanglement

$pp \to t\bar{t} \to b\bar{b}\ell\nu\ell\nu$

Assignment & Generation

Anomaly Detection

$\Upsilon \to \mu^+\mu^-$

Event Generation

🔧 ***No hyperparameter tuning*** *was applied in any of our tests*

Search for New Physics: Overview

- 🧬 **Exotic Higgs Decay ($H \rightarrow aa \rightarrow bbbb$)**: A challenging **4-b final state** sensitive to **b-tagging inefficiency** and **jet misassignments** → ideal for testing EveNet's **assignment** and **classification** capabilities.
- **Samples:**
  - **Signal:** $H \rightarrow aa \rightarrow bbbb$ **(**$m_a = 30, 40, 60$ **GeV)**
  - **QCD:** $bbbb, bbbj, bbjj$
- **Methodology**:
  - **Network:** EveNet ($\sim$40M parameters) vs. SPANet (same hidden dim, $\sim$40M parameters)
  - **Pretrain weights**: True vs. False
  - **Training Dataset size:** 10k / 30k / 100k / 300k / 1M (signal portion: 10%)
  - **Assignment head (as Aux Task):** True vs. False

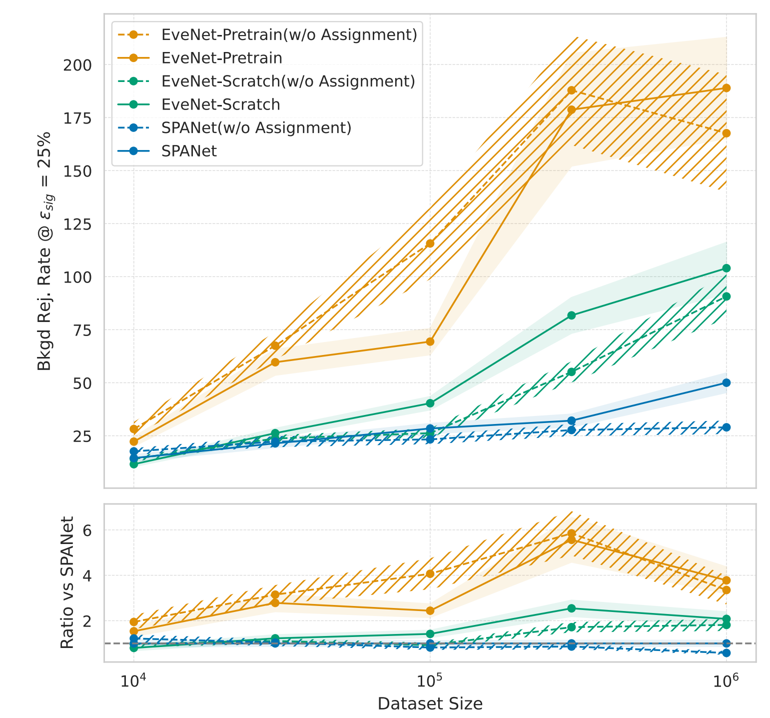
🌟*The signal samples used here were **not included** in pretraining,*

which tests EveNet's ability to generalize to unseen new physics signatures.

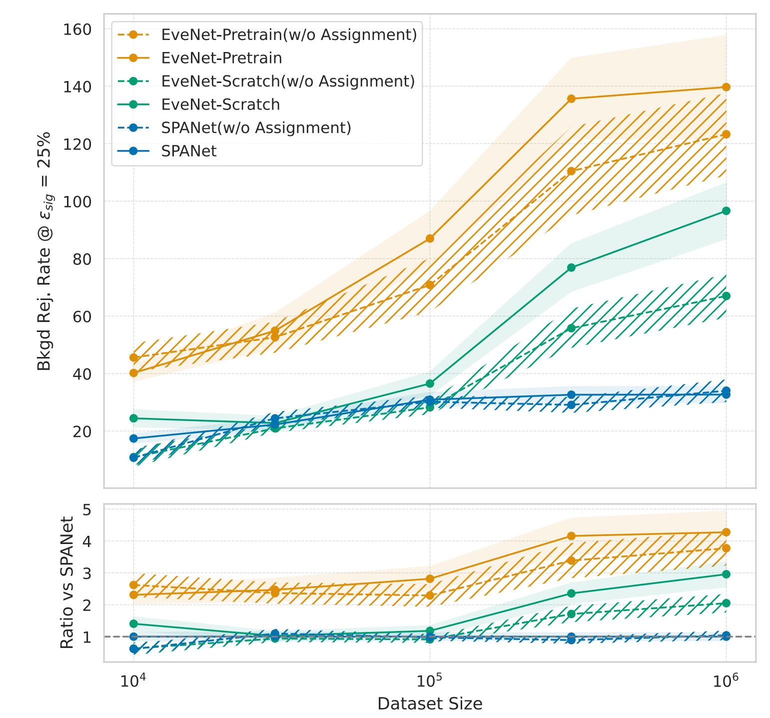
Search for New Physics: Results

🎯 Final performance is reported as **background rejection at 25% signal efficiency**, reflecting the metric most relevant for new physics searches.

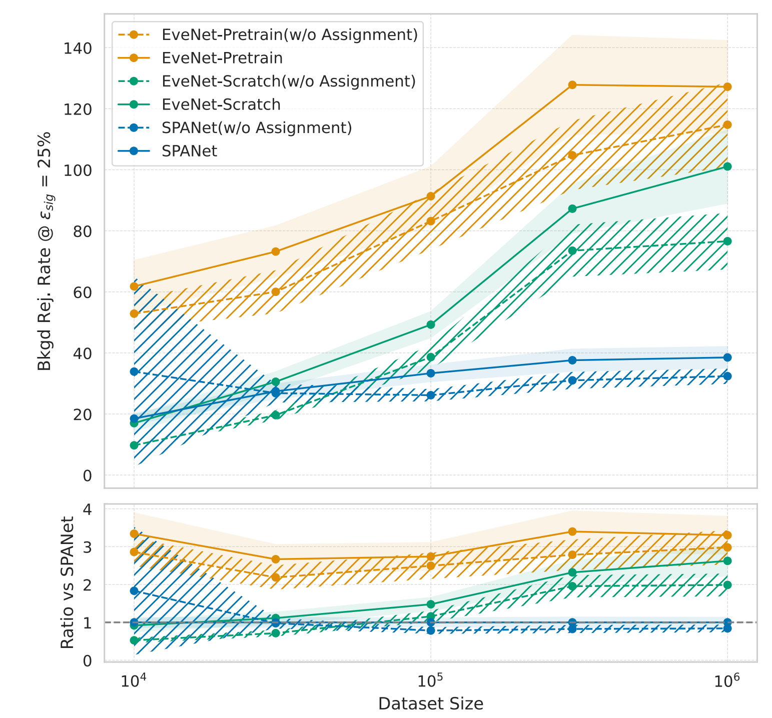
- *This focus aligns with standard practices, where the sensitivity is driven by events in **the highest MVA score bins**.*



$m_a = 30$ GeV



$m_a = 40$ GeV



$m_a = 60$ GeV

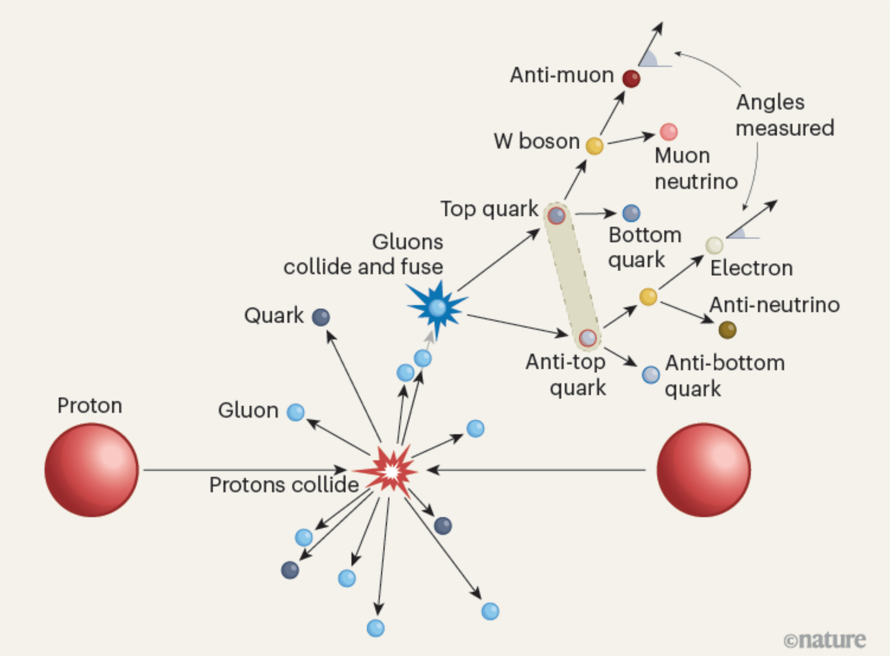
Search for New Physics: Observations

- 📊 **EveNet shows strong scalability**:
  - Performs well even on **small training datasets**.
  - Continues to improve with **increasing data volume.**
  - **Pretrained model performs well even without assignment head, unlike SPANet or scratch models.**
- ⚖️ **Compared to SPANet**:
  - EveNet offers better **scalability** and **robust generalization** out of the box.
  - SPANet may require **additional tuning** to match performance at larger scales.
  - Performance improves **2–4×** with the pretrained EveNet.

Quantum Entanglement: Overview

- 🔗 **Quantum Entanglement ($pp \to t\bar{t} \to b\bar{b}\ell\nu\ell\nu$)**: A complex 2-lepton final state with multiple neutrinos and combinatorial jet ambiguity → ideal for testing EveNet's **assignment** and **neutrino generation** capabilities.

- **Samples:** $pp \to t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ (threshold region)
- **Methodology**:
  - **Network:** EveNet ($\sim$40M parameters)
  - **Pretrain weights**: True vs. False
  - **Dataset size:** 3.6M for training, 2.4M for evaluation
- **Metrics**:
  - $t \to b\ell$ **pairing efficiency**
  - **Uncertainty** from unfolded spin correlation matrix and $D = -(C_{kk} + C_{rr} + C_{nn})$



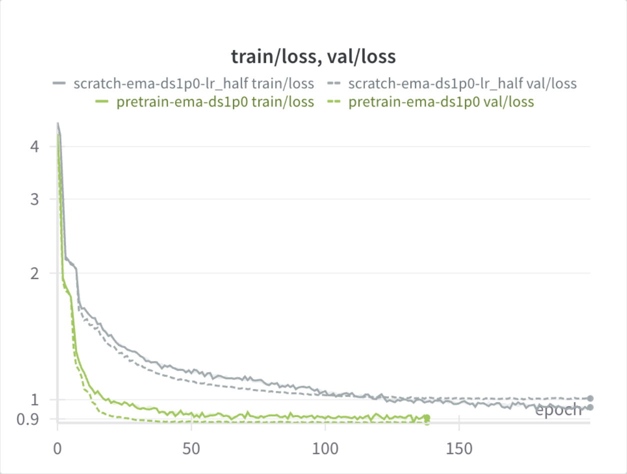
Quantum Entanglement: Results

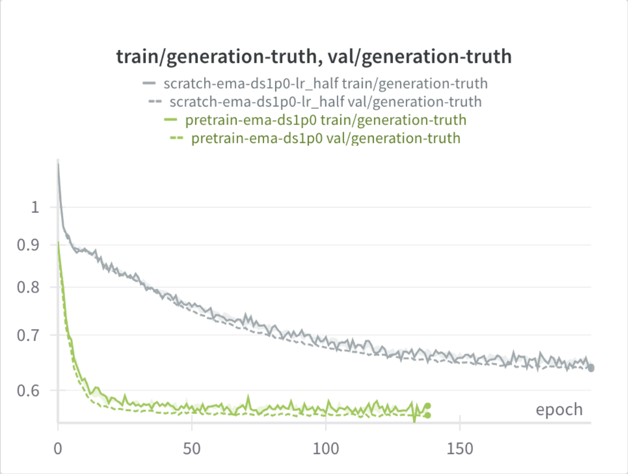
Assignment Efficiency

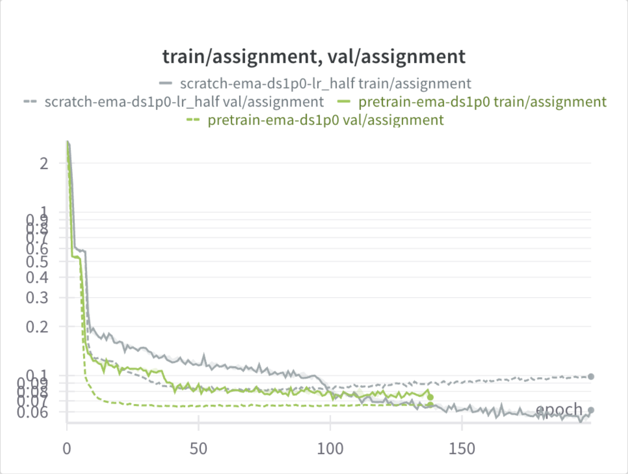
- **Matchable:** Events where a **ground-truth assignment** exists; i.e., the event topology allows a **well-defined mapping** between reconstructed objects (e.g., jets) and true partons.
- **All Events**: The **full set of events**, including both matchable and unmatchable ones.

Model	Matchable Events	All Events	Efficiency [%]
Scratch	172,681 / 241,986	172,681 / 287,950	71.36 / 59.97
Pretrain	178,909 / 241,986	178,909 / 287,950	**73.93 / 62.13**

Training Loss







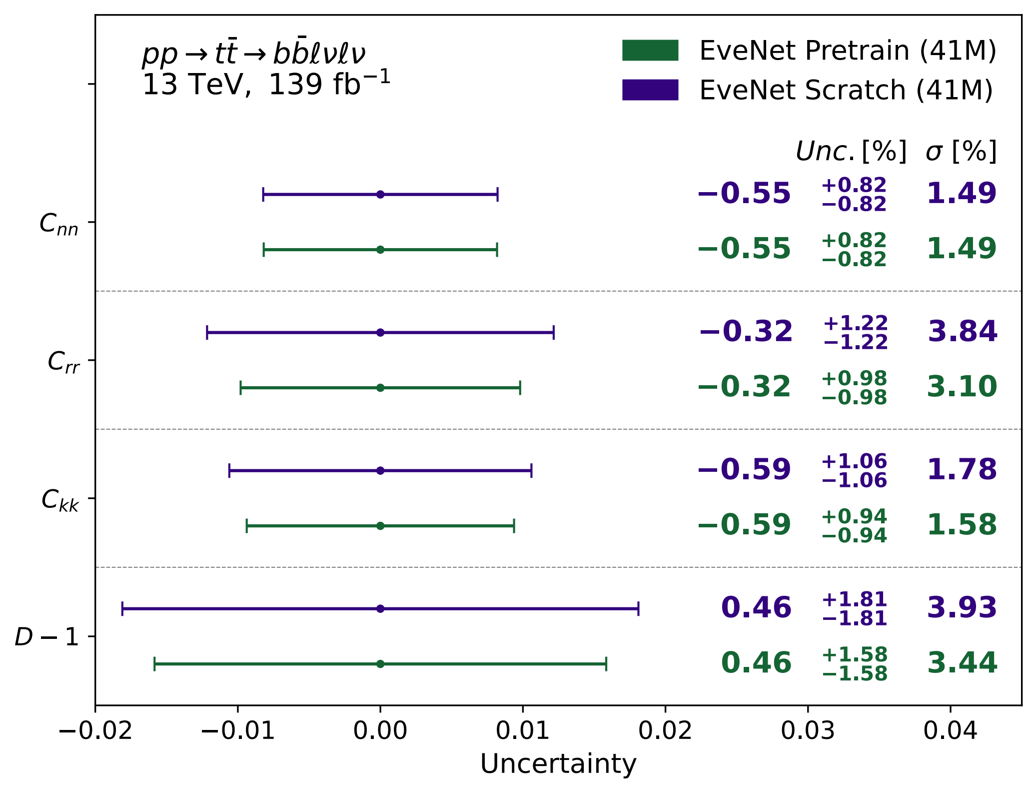
Quantum Entanglement: Results

Unfolded Uncertainty for spin correlation matrix and D

- Reference paper: Eur. Phys. J. C (2022) 82:285, assuming 139 fb^{-1}
- The observable $D = -C_{kk} - C_{rr} - C_{nn}$ is sensitive to quantum entanglement, with $D > 1$ indicating the quantum entanglement.
- Absolute uncertainty improvement (*pretrain vs. scratch*):

$$\frac{\sigma^s - \sigma^p}{\sigma^s} \approx \mathbf{12.5\%}$$

- Relative precision with $\epsilon = \sigma_D/(D-1)$
  - **Pretrain: $\epsilon_D \approx \mathbf{3.43\%}$**
  - Scratch: $\epsilon_D \approx 3.93\%$
  - Paper: $\epsilon_D \approx 5.26\%$



Quantum Entanglement: Observations

- **Pretrained model shows improved assignment performance**, increasing matching efficiency by:
  - **+2.5%** for **matchable events**
  - **+2.1%** for **all events**
- **Uncertainty reduction**:
  - **Absolute improvement** of $\sim 12.5\%$ in precision over the scratch model
  - **Relative precision improvement** of $\sim 35\%$ over the pheno paper result
- **Rapid and stable convergence**:
  - Pretrained model **converges faster** for both assignment and generation heads
  - Reduces risk of **overfitting** in the assignment task

Anomaly Detection: Overview

- ***Reference paper***: 2502.14036 (*To test EveNet's generative capability, we extend an existing anomaly detection method **using normalizing flows** by replacing it **with diffusion-based generation** of full 4-momentum)*
- **Dataset**: CMS Open Data (2016 DoubleMu primary dataset) targeting Υ resonances in di-muon final states.
- **Goal**: Perform **model-independent bump hunting** in the invariant mass spectrum using **diffusion-based generative models** to interpolate background.
- **Strategy Overview:**
  1. Signal region (SR) and Sideband (SB) definition ($m_{\mu\mu}$): $SR = [9, 10.6]$ GeV, $SB = [5, 9]$ & $[10.6, 16]$ GeV
  2. **Background Modeling** Replace NF (CATHODE in paper) with an <u>ensemble of EveNet diffusion models</u>
     - **Global Generation:** Conditioned on mass, generate H_T and $\Delta R_{\mu\mu}$
     - **PC generation:** Conditioned on mass, H_T and $\Delta R_{\mu\mu}$, generate muons with features: 4-momentum and ip3d
     - **Quality selection**: Recalculate every global information from the point cloud directly and re-apply analysis cut i.e., windows cut on the generated events.
  3. **Weak supervision**: training XGBoost to separate generated events and data events
  4. **Significance extraction**: cut-and-count and likelihood-reweighting

Anomaly Detection: Results

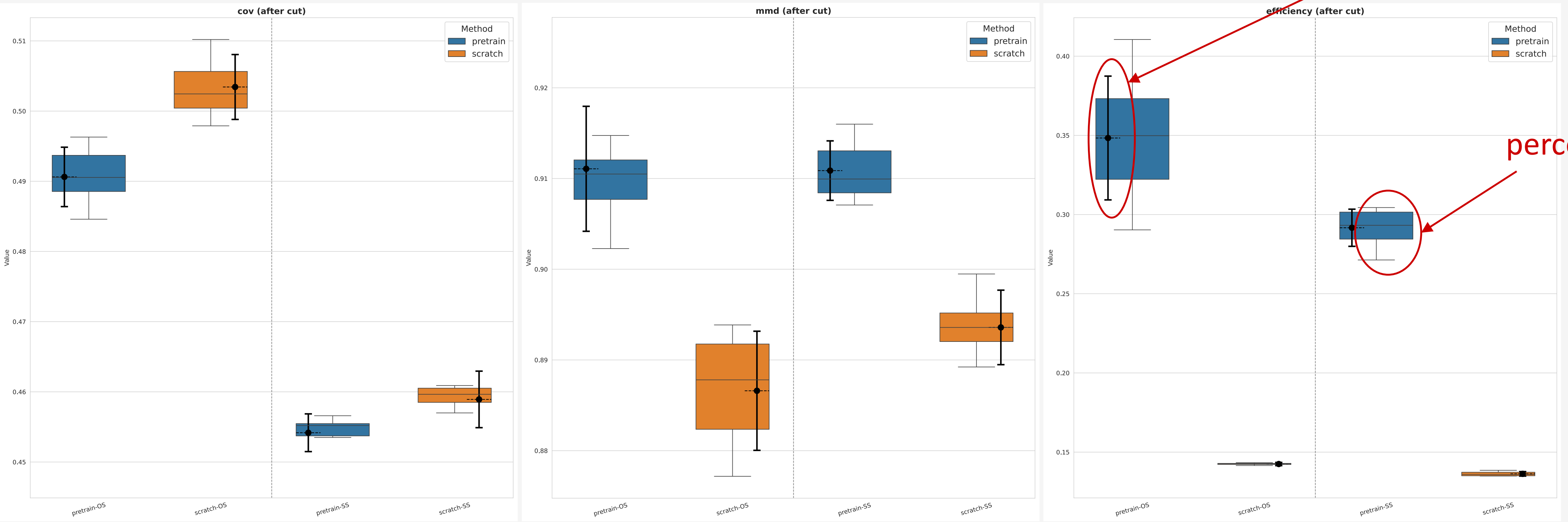
🌟All *results are performed 8 times with different random seeds to test the spread*

Generation Quality (**arXiv: 2106.11535**)

- **Coverage**: measuring the diversity of the samples in Y relative to X
- **MMD**: the average distance between matched samples, measuring the quality of samples
- **Efficiency**: quality selection efficiency for generated events

$\mu \pm \sigma$

percentiles

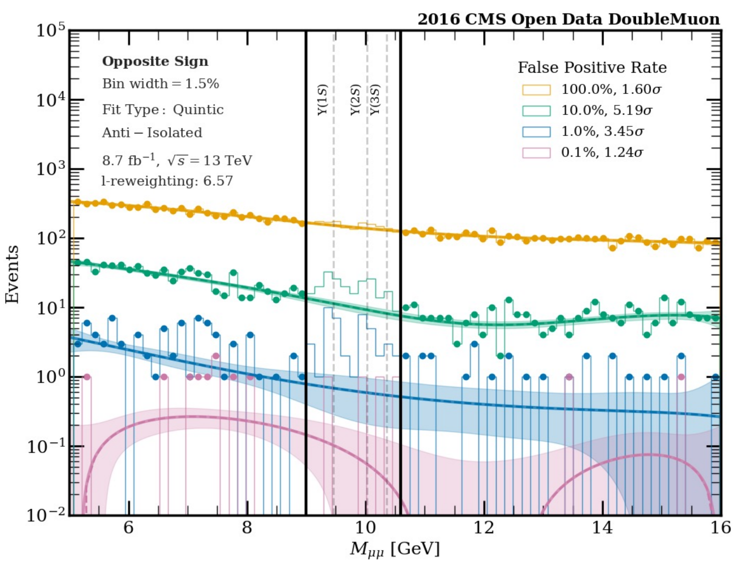


Anomaly Detection: Results

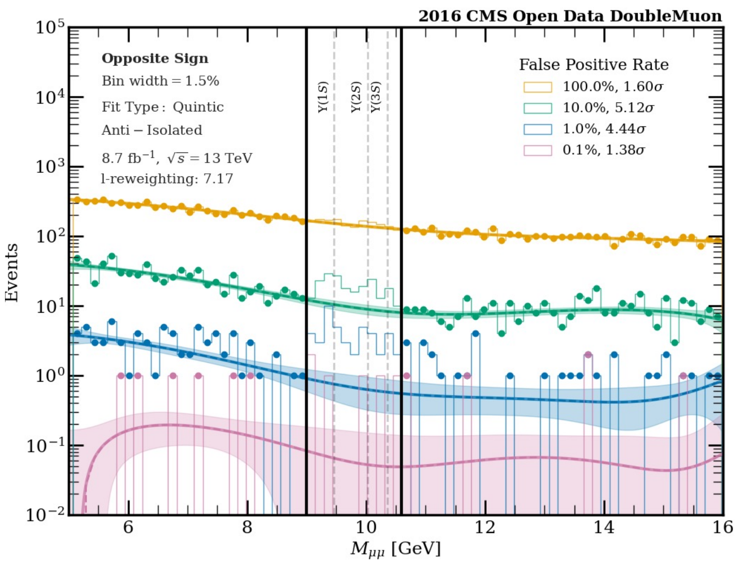
⭐ *All results are performed 8 times with different random seeds to test the spread*

Final Significance (ℓ-reweighting)

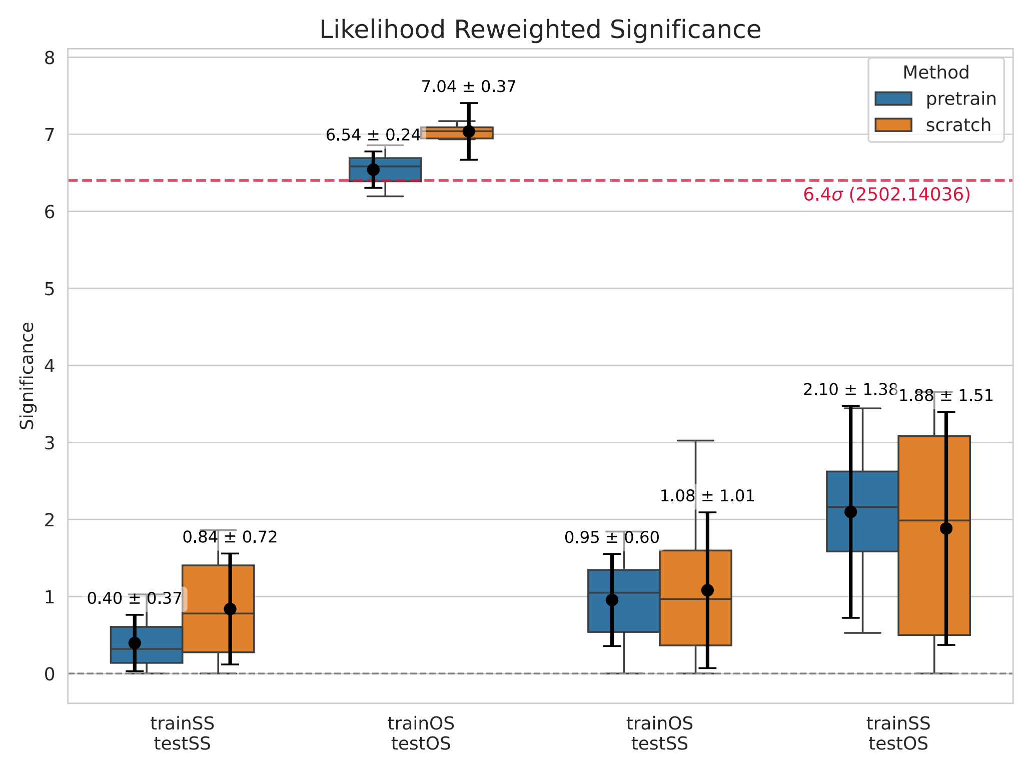
- **paper:** $\mathbf{6.4\sigma}$
- **EveNet-Pretrain:** $\mathbf{6.54 \pm 0.24\sigma}$
- **EveNet-Scratch:** $\mathbf{7.04 \pm 0.37\sigma}$



EveNet-Pretrain



EveNet-Scratch



Anomaly Detection: Observations

📊 **Final Significance**:

- Both **pretrained** and **scratch** models achieve comparable or better results than the original CATHODE benchmark.
- **Scratch model** slightly outperforms the paper baseline.
- **Pretrained model** performs slightly below, but with **smaller variance** across 8 random seeds.
- No mass sculpting observed in **same-sign control region**.

⚙️ **Generation Efficiency**:

- **Pretrained model converges faster** and achieves **2.5× higher quality selection efficiency** than the scratch model.

🧠 **Analysis-Specific Limitation**:

- Slight underperformance of the pretrained model is likely due to the use of **ip3D**, a feature not present in pretraining.
- With a **lower learning rate** on body during fine-tuning, pretrained models adapt **more slowly to unseen features** like ip3D.
- For **4-momentum-related distributions**, the pretrained model consistently produces **higher-quality samples** than scratch.

Summary

Pre-trained Model 🧐 → **Foundation Model** 🥳

Our current study shows that **pretraining enables transferable and multi-task representations** across diverse HEP tasks.

- Pretrained EveNet demonstrates **strong scalability**, **fast convergence**, and **robust generalization** across diverse HEP tasks, without the need for hyperparameter tuning or task-specific design.

Search for new Physics

$H \to aa \to bbbb$

Assignment & Classification

Up to **2–4× gain** on bkgd. Rej. Rate @ $\epsilon_{sig} = 25\%$, strong performance even **without assignment head**.

Quantum Entanglement

$pp \to t\bar{t} \to b\bar{b}\ell\nu\ell\nu$

Assignment & Generation

+2.5% assignment, **12.5% uncertainty reduction**, **$\sim 35\%$ better** than prior work

Anomaly Detection

$\Upsilon \to \mu^+\mu^-$

Event Generation

Matches or exceeds baseline; **2.5× more efficient** generation and better **4-momentum modeling**.

Summary

Pre-trained Model 🧐 → **Foundation Model** 🥳

🚀 **Next Milestone: Scaling for Emergence and Multimodal**

- To explore **emergent capabilities**, we are preparing a **150M-parameter model** trained on up to **1.5B effective events**, aiming to push EveNet into the true foundation model regime.
- **Multimodal Potential Ahead**: Future extensions include integrating **jet constituents**, **tracker hits**, and **heterogeneous data forms** to explore **multimodal learning** in HEP.

📦 **Dataset Sharing:** We have **3B raw events in Parquet format** and are happy to share them for benchmarking or related studies.

📄 **Paper Coming Soon**: We are finalizing the draft, and the **arXiv link will be shared shortly!**