Foundation Models for Collider Physics: AI as a Tool for Discovery

Yulei Zhang

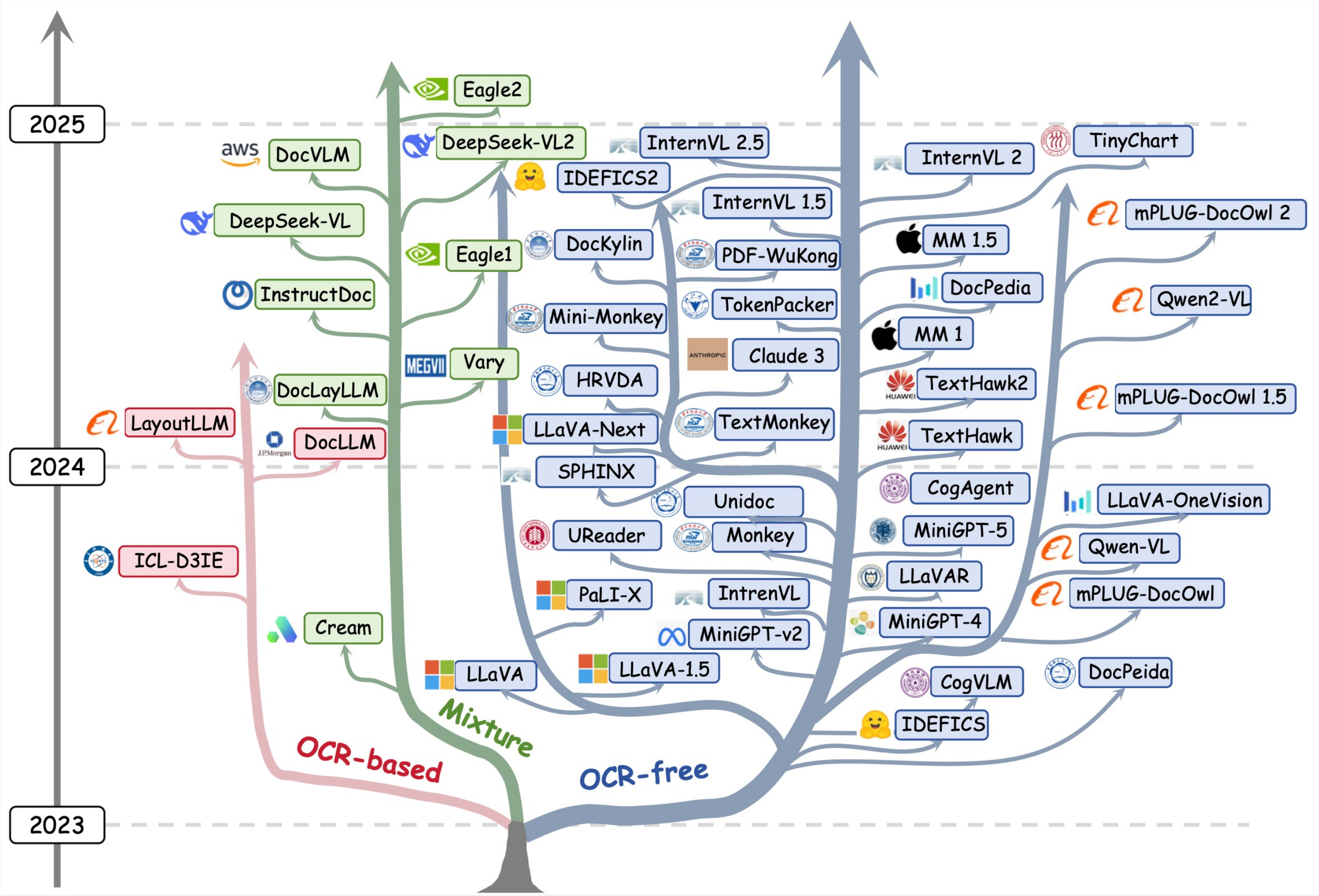
University of Washington, Seattle

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

September 25th, 2025

2025 Postdoctoral Frontier Symposium
in Physics and Astronomy

Large Language Model is changing the world…



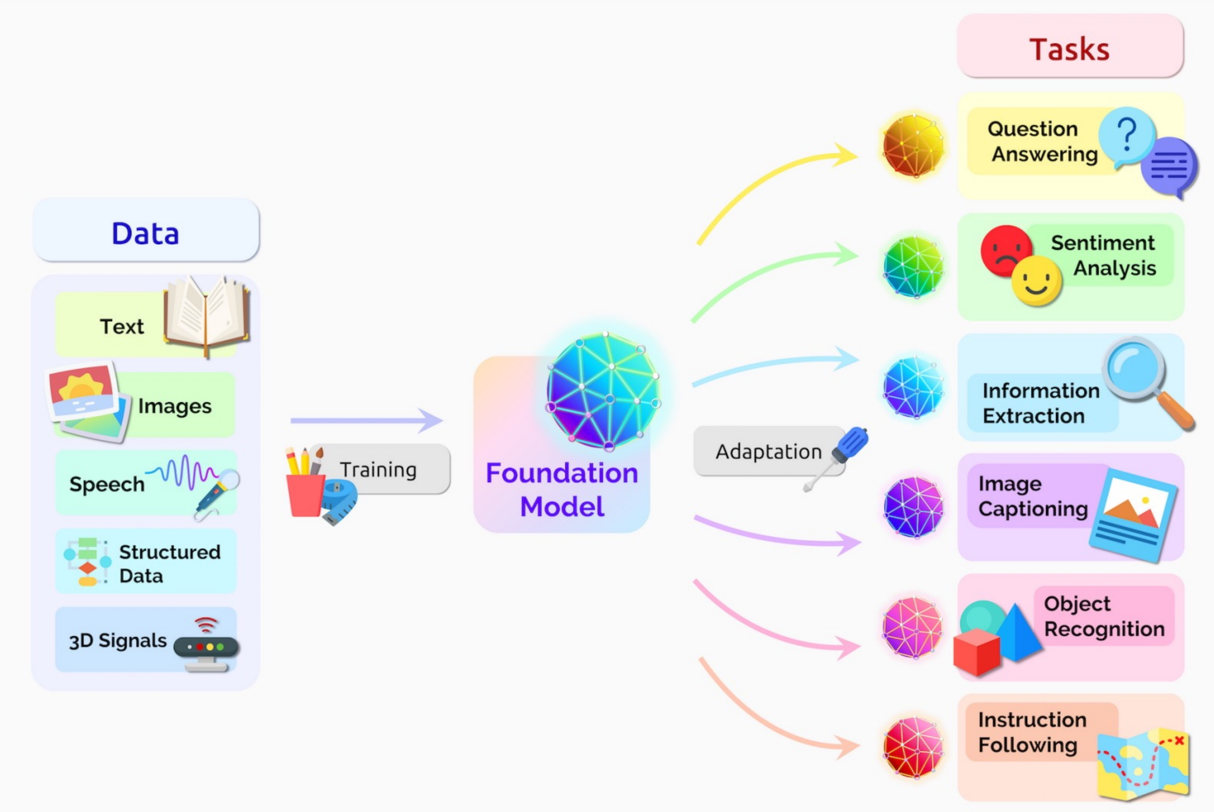
- **Transformers** (self-attention, multi-head attention)
- **Positional encoding**
- **Sparse / Mixture-of-Experts (MoE)** layers
- **Scaling laws** (compute, data, parameters)
- **Pretraining on large dataset**
- **Fine-tuning** (task-specific adaptation)
- **Chain-of-Thought (CoT)** prompting
- …

➢ **Foundation Model**

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**:

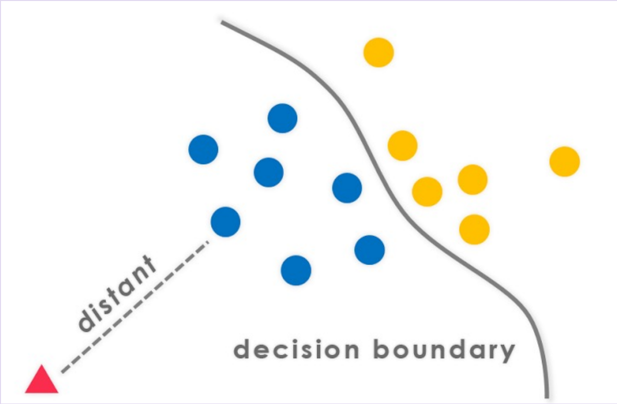
- Pretrain once, reuse anywhere

🌐 **Multimodal potential**:

- Works across data types

Understanding Today's ML Tasks in HEP

Discriminative



Physics Process Classification
Classification for underlying physics process

Resonance Segmentation
Map objects to queries

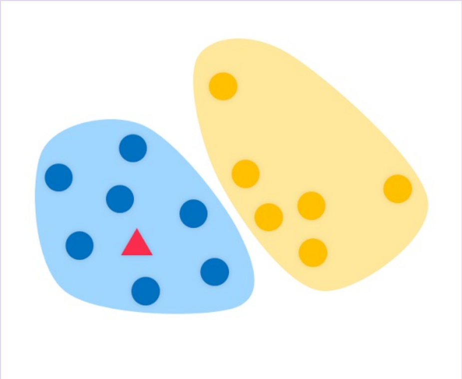
Resonance Detection
Prob. of existence of resonance A

Resonance Mapping
Mapping objects to resonance A

Resonance Classification
Class for each queries

And more…

Generative

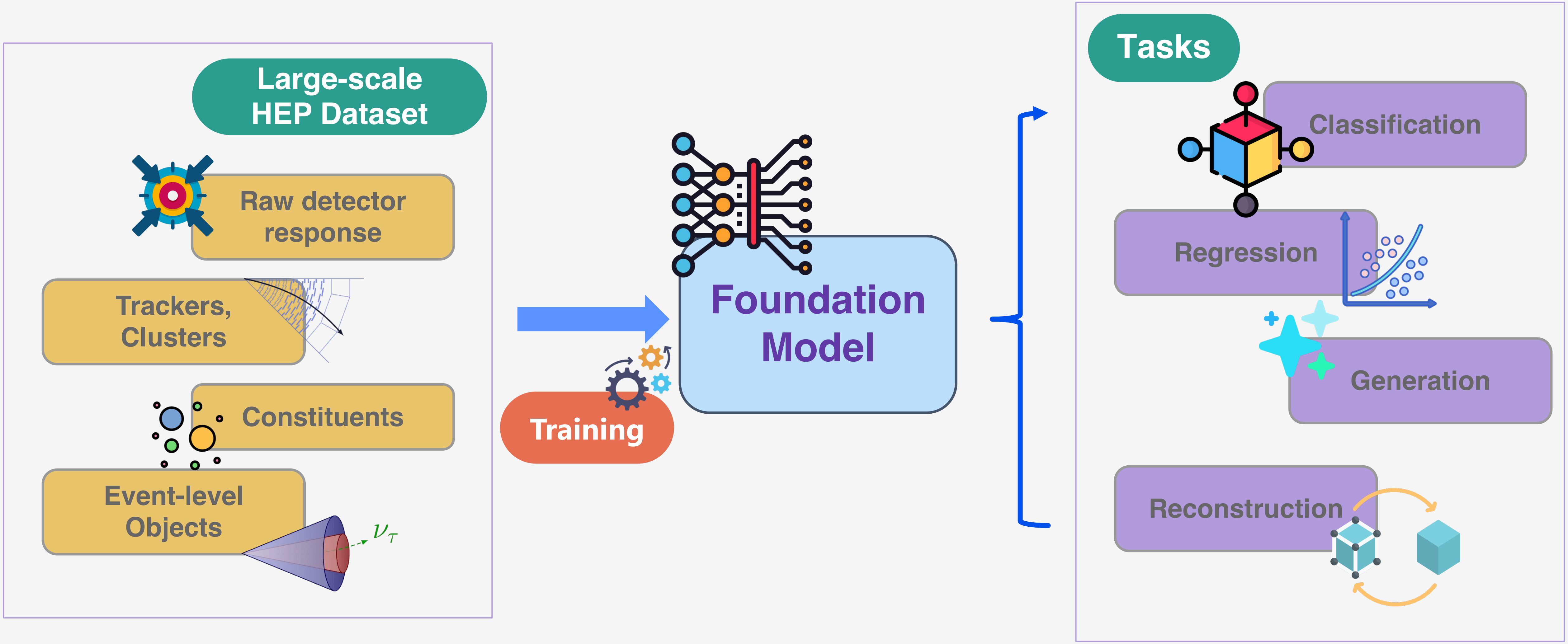


Invisible Particle Prediction

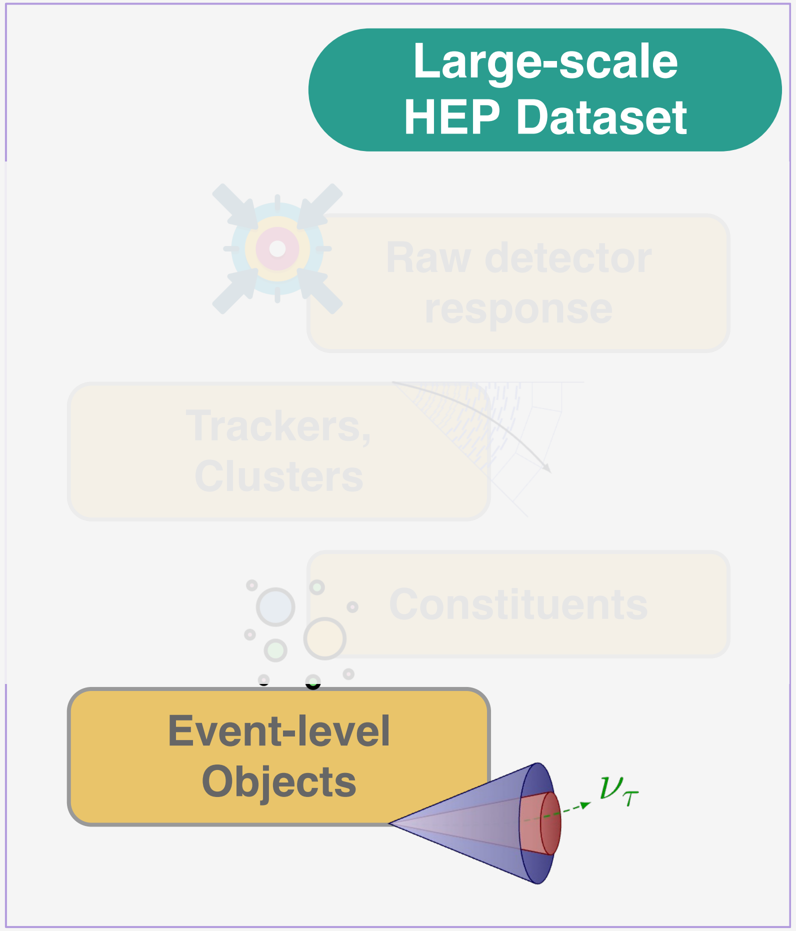
Physics event Generation

Can we **unify** these diverse tasks **under one framework**?

Can We build a Foundation Model for HEP?



Can We build a Foundation Model for HEP?



reconstructed & calibrated objects

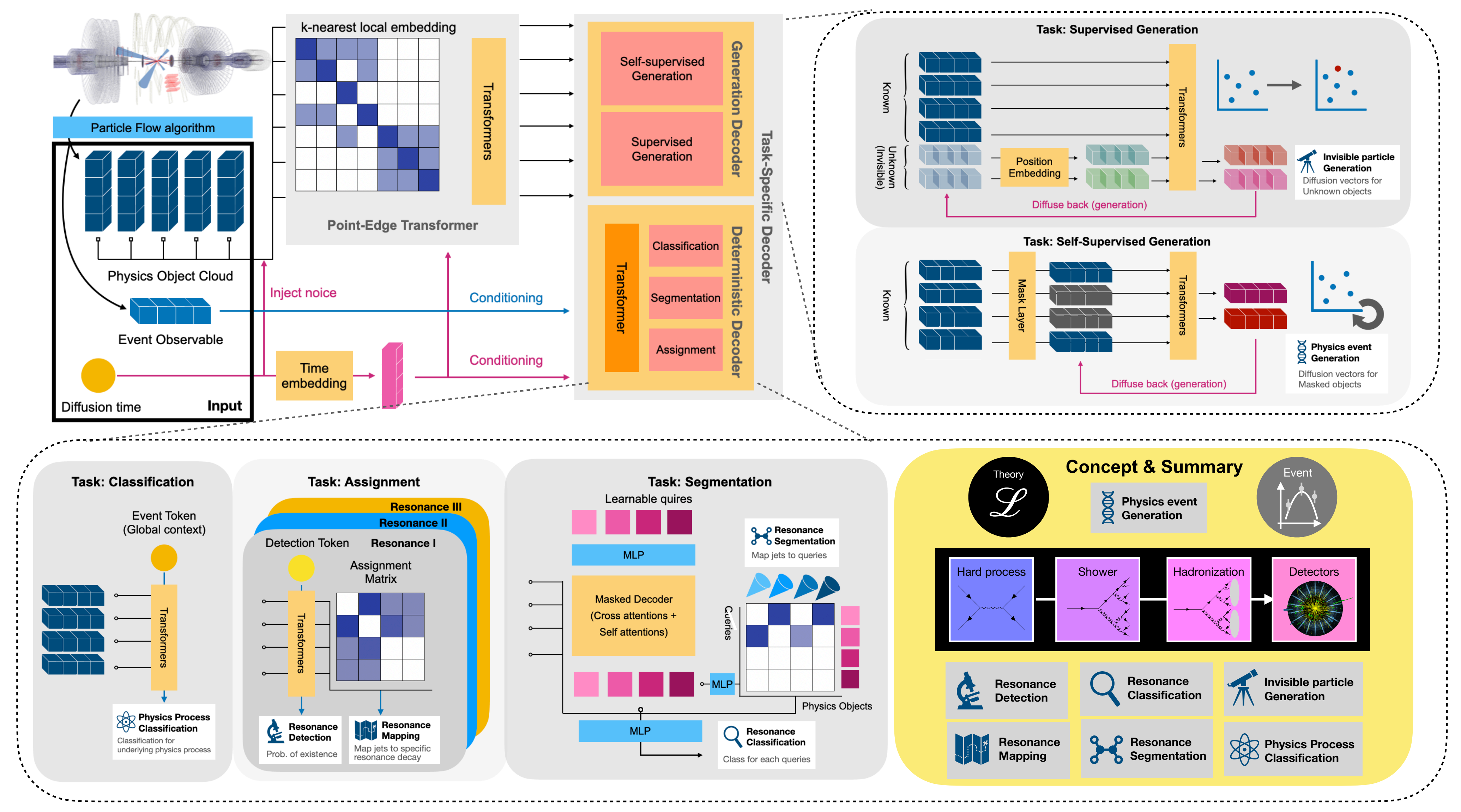
Pros:

- Already **calibrated** and **validated by the experiment**.
- Topology is clear and interpretable (jets, leptons, MET, etc.).
- Closest to final physics interpretation → straightforward connection to analysis.
- ***Lower*** computational cost (compact representation).
- **Less dependent** on **detector design** or **experiment-specific effects** → more portable across experiments and closer to physics truth.

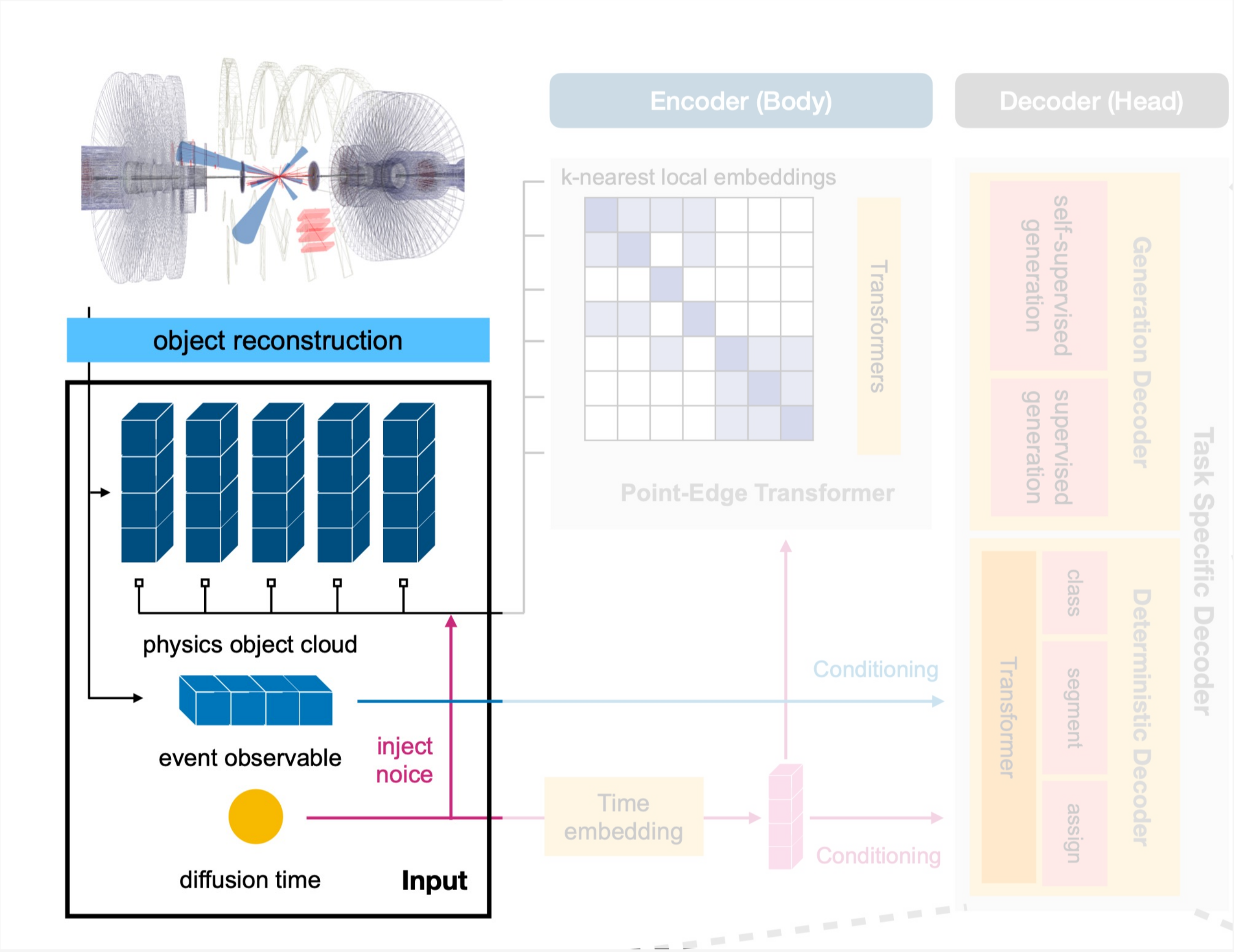
Cons:

- Strongly **dependent on reconstruction algorithms** (jet clustering, tau ID, lepton calibration, etc.).
- Potential **biases** and **information loss** introduced during reconstruction.

EveNet: Our Answer to Event-Level Foundation Models



EveNet: Our Answer to Event-Level Foundation Models



🔢 Input Representation

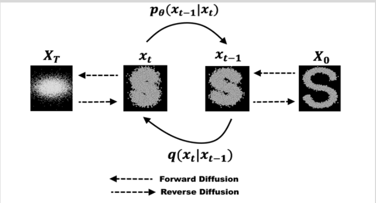
- 📌 **Particle Cloud (Up to 18 Particles per Event):**
  - Each particle is encoded with 7 features: **4-momentum, isbJet, isLepton**, and **charge.**
- 🌐 **Global Features / Event Observables:**
  - Missing transverse energy
  - Number of leptons, number of jets
  - Invariant mass of visible objects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

Un-perturbed PC

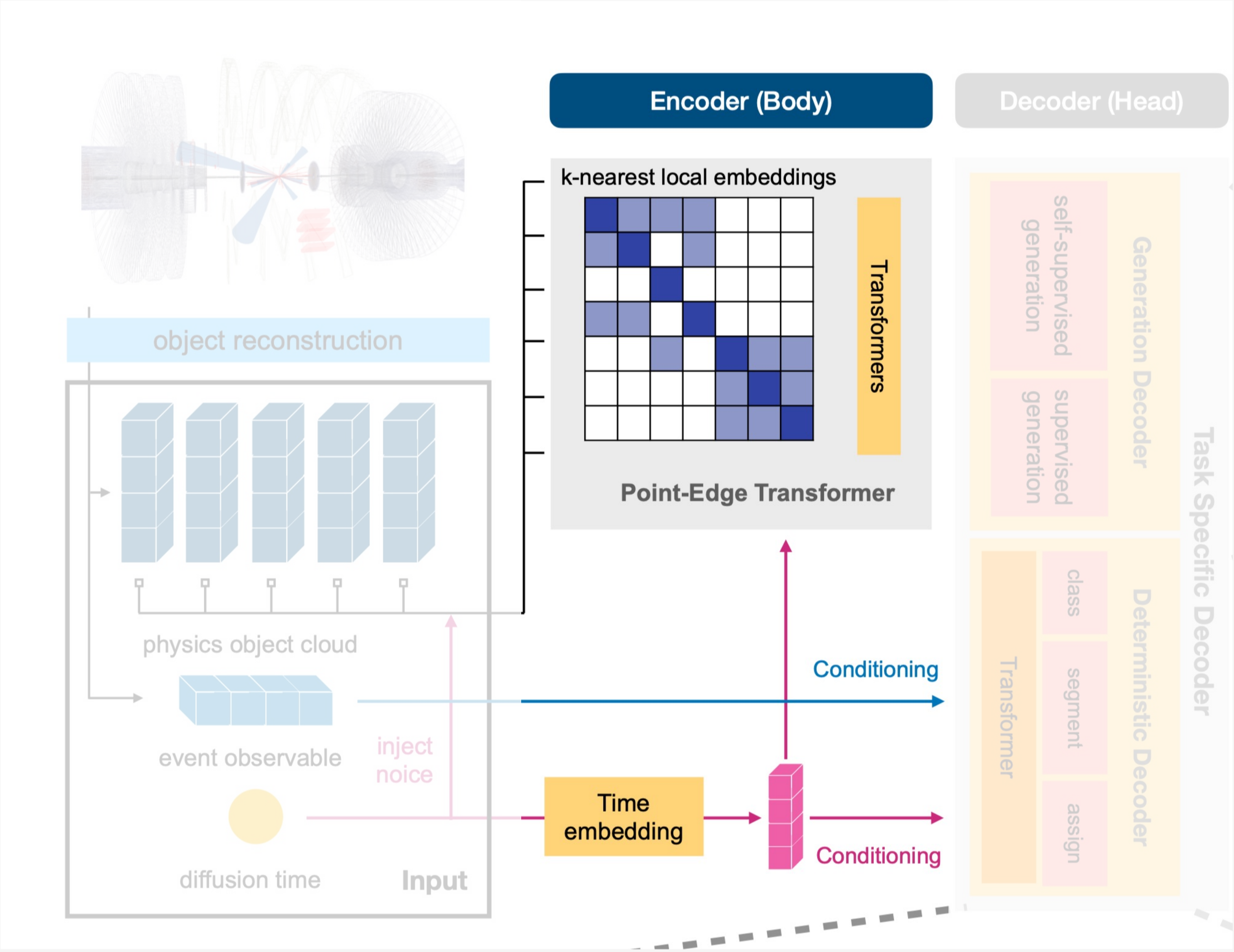
for deterministic tasks…

Perturbed PC

for diffusion model and noise tolerance training



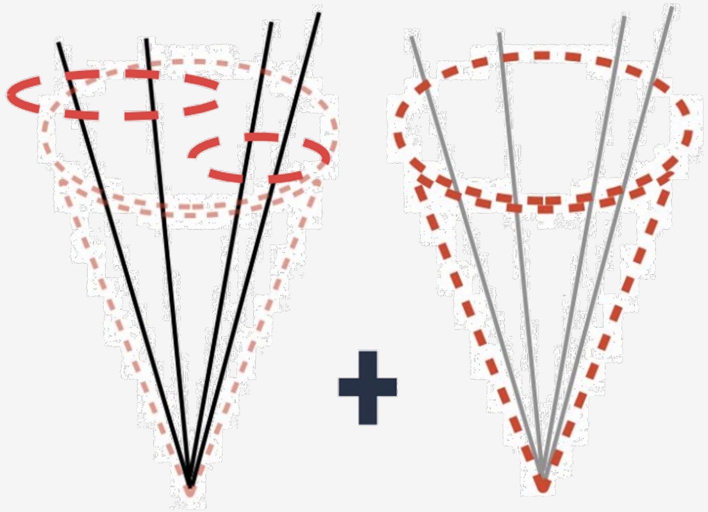
EveNet: Our Answer to Event-Level Foundation Models



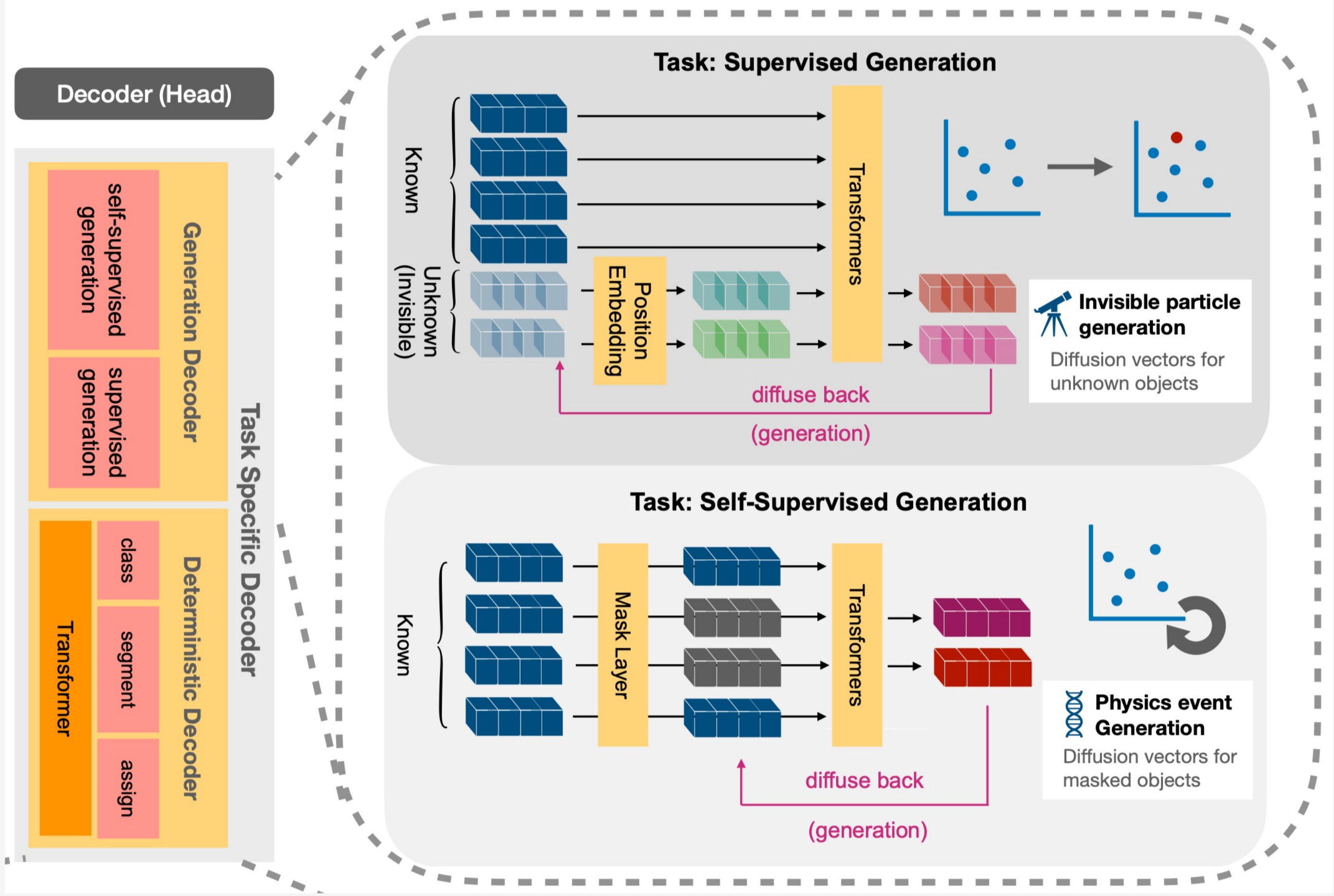
Core Idea: One strong body + many small heads

🧠 **Encoder - Point-Edge Transformer：**

- Inspired by OmniLearn [2404.16091]
- Models **both particles and their relationships** as a graph (points + edges)
- Captures **inter-particle interactions** and **global event structure**



EveNet: Our Answer to Event-Level Foundation Models



Core Idea: One strong body + many small heads

🧠 Decoder – Generation Head：

Supervised Generation

- Use known objects as input to predict missing ones (e.g., neutrinos).
- Diffusion models capture high-dimensional probability densities → predict the most likely kinematics.

Self-supervised Generation

- Mask part/all of the inputs and reconstruct them with a diffusion model.
- Learns underlying event structure without requiring labels.

EveNet: Our Answer to Event-Level Foundation Models

Core Idea: One strong body + many small heads

🧠 Decoder – Discriminative Heads：

Segmentation

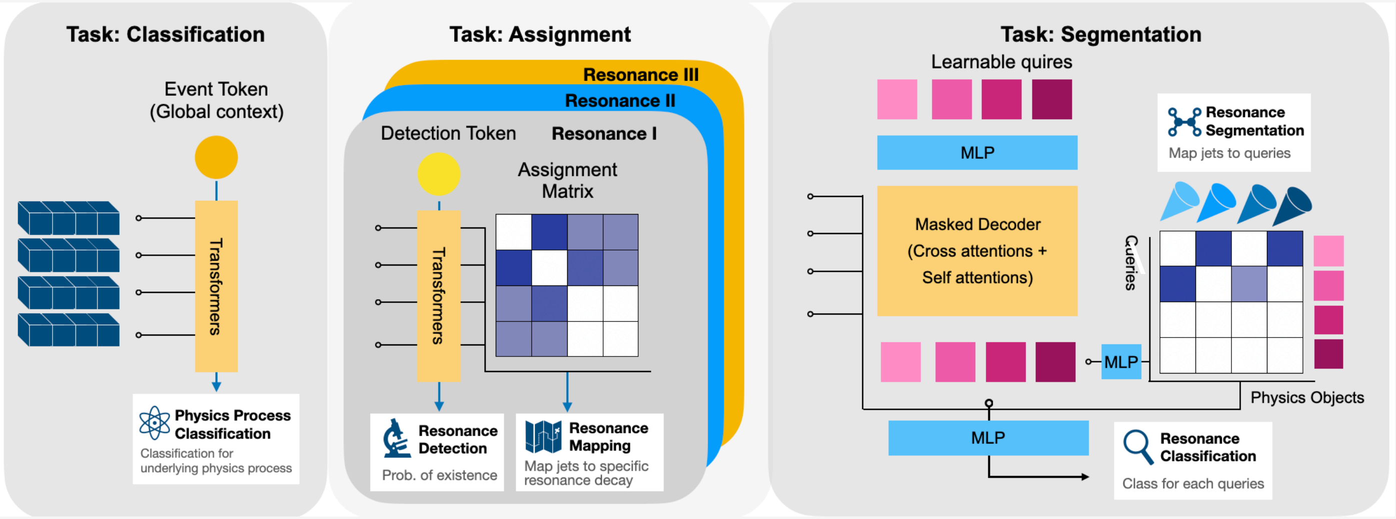
- Inspired by Meta AI's segmentation networks
  - The model performs set prediction (queries → predict class & mask), preserving permutation symmetry.
  - Naturally extendable from objects to substituents without changing the model design.

Classification

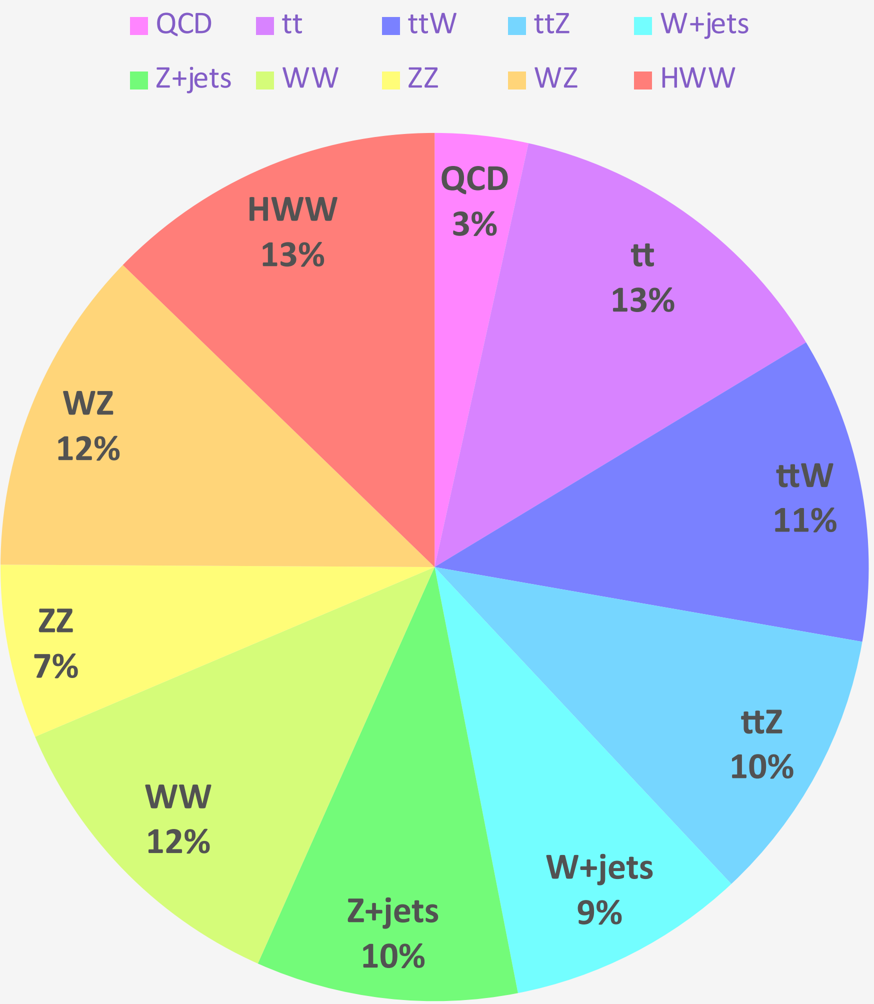
- Multi-Class event classifiers (with regression)

Assignment

- Symmetry-aware mapping of objects to truth partons (requires known decay topology).
- High accuracy for well-defined processes, but rigid, costly, not generalizable.



EveNet Recipe: How we pre-train?



Dataset

- 10 SM processes, **3B raw → 500M preprocessed events**.
- Diverse processes → learn classification & point cloud generation.
- Complex channels (ttV, VV, HWW) → drive segmentation & neutrino generation.

Training Strategy

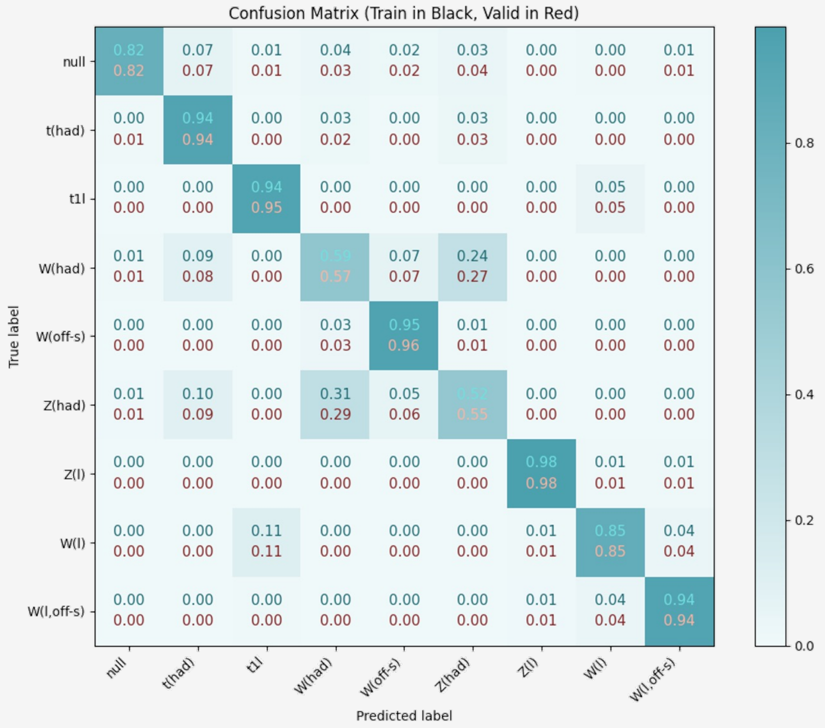
- **Stage 1 (*self-supervised*)**:
  - Only generation head active → learns unlabeled point cloud representation.
  - Gradual masking schedule: start with 10% masked → ramp to 100% (forces learning full event topology).
- **Stage 2 (*full training*)**:
  - All heads active → multi-task optimization.
  - Assignment head is off due to high computational cost
- Stability: **EMA** + **warm-up & cosine decay schedule**.

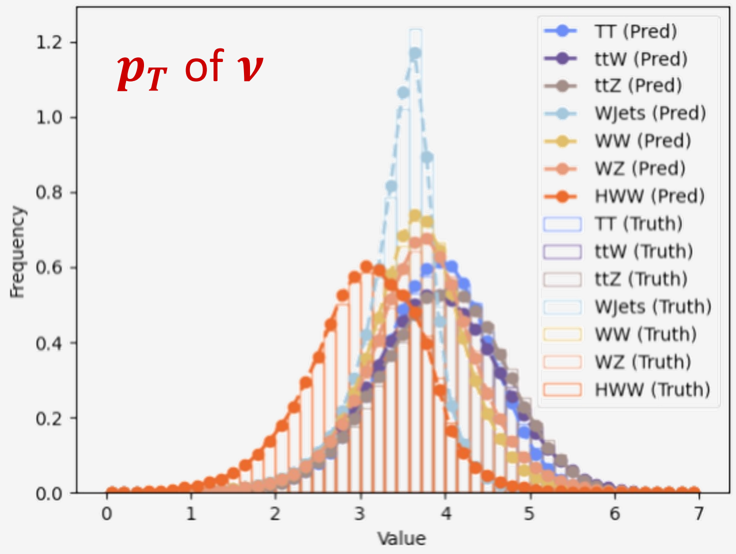
🍽️ EveNet Tasting: First results from pretraining

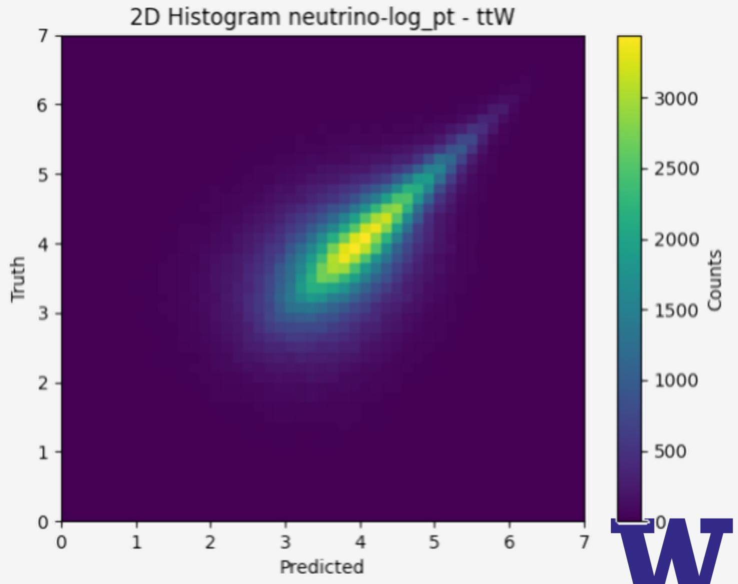
Multi-label Classification

Segmentation

Detected Resonance







Supervised Generation

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

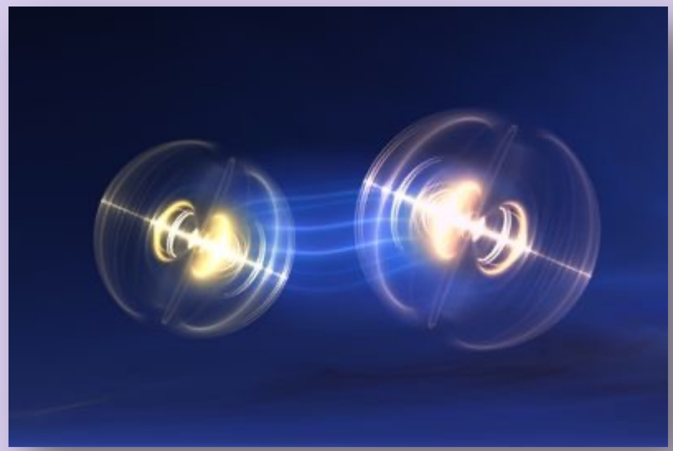
🔋 **Scaling Up EveNet with Perlmutter**

- 🚀 **Training Setup:**
  - **128 nodes**
  - **512 GPUs**
  - **16,384 CPU cores**
- ⚙️ **EveNet Model:**

  Encoder + Decoder

  - **Lite**: 20M + 3M (today's result)
  - **Standard**: 83M + 17M (in progress)

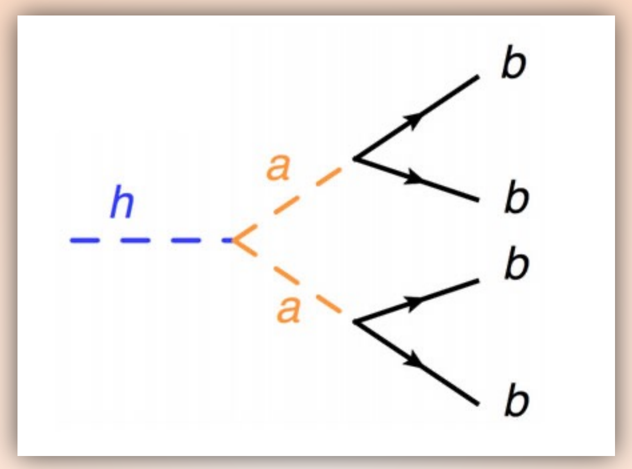
Downstream Applications of EveNet in Physics Analyses



Quantum Entanglement

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

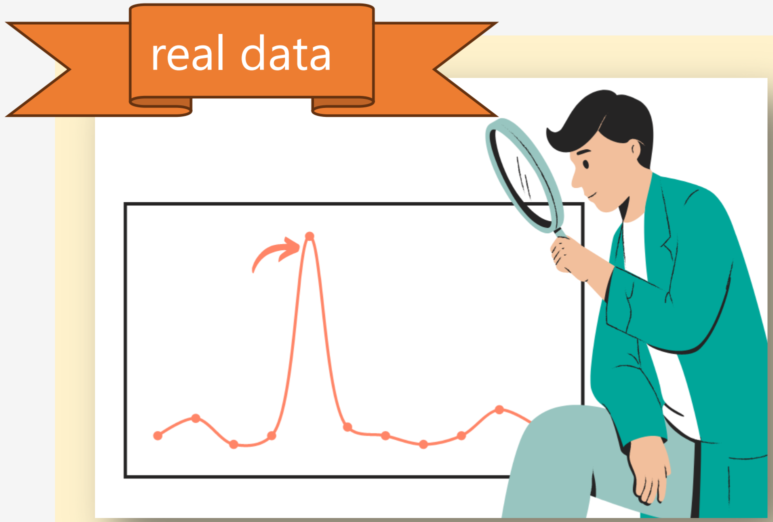
Assignment & Generation



Search for new physics

$H \to aa \to bbbb$

Assignment & Classification



Anomaly Detection

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

Event Generation

In-distribution
($t\bar{t}$ present in pretraining dataset)

Near out-of-distribution
(new signal, bkgd. overlaps)

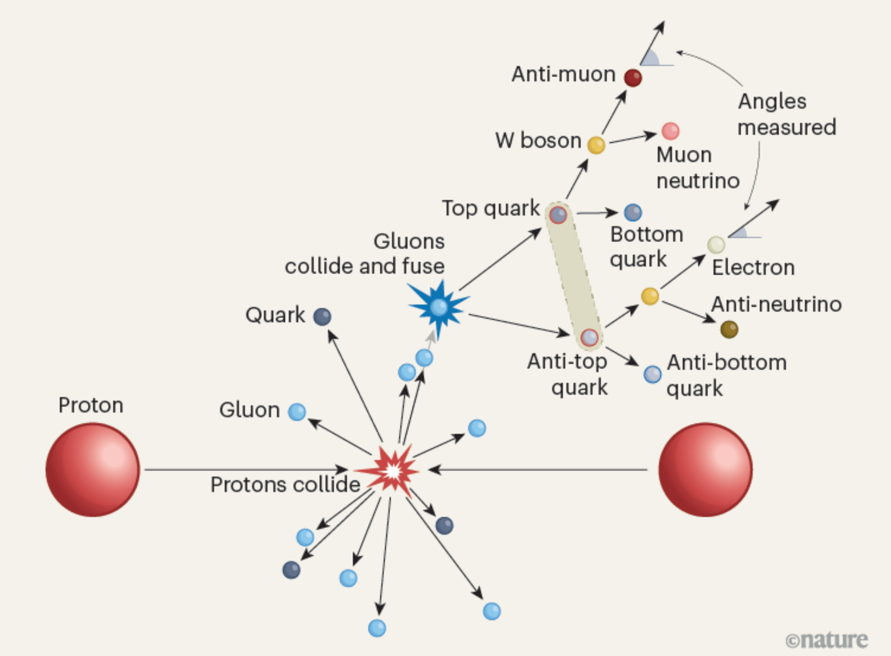
Fully out-of-distribution
(data-driven, different CME)

*Easy - **familiar** physics and energy.*

*Hard - **unseen** physics and **shifted** energy regime.*

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
- **Samples:** $pp \to t\bar{t} \to b\bar{b}\ell\nu\ell\nu$ (threshold region)
- **Methodology**:
  - **Network:** EveNet-Lite
  - **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 $D = -(C_{kk} + C_{rr} + C_{nn})$



The model is jointly trained on the **Assignment** and **Truth Generation** tasks.

Quantum Entanglement: Results

Assignment Efficiency

- [1]**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.
- [2]**All Events**: The **full set of events**, including both matchable and unmatchable ones.

	F.T. (CLS + Gen)		Scratch		*Improvement [%]*		F.T. (CLS + Seg + Gen)		F.T. (SSL)	
	Eff. [%] (M[1])	Eff. [%] (A[2])	Eff. (M)	Eff. (A)	*M*	*A*	Eff. (M)	Eff. (A)	Eff. (M)	Eff. (A)
1.0	**82.17**	**71.2**	80.48	69.71	***2.10***	***2.14***	81.95	71.02	80.83	70.05
0.7	**82.29**	**71.32**	78.85	68.28	***4.36***	***4.45***	81.69	70.78	80.37	69.66
0.3	**81.91**	**70.98**	77	66.64	***6.38***	***6.51***	81.51	70.6	79.15	68.51
0.1	**81.74**	**70.82**	45.64	39.38	***79.10***	***79.84***	81.48	70.56	49.53	42.72

Quantum Entanglement: Results

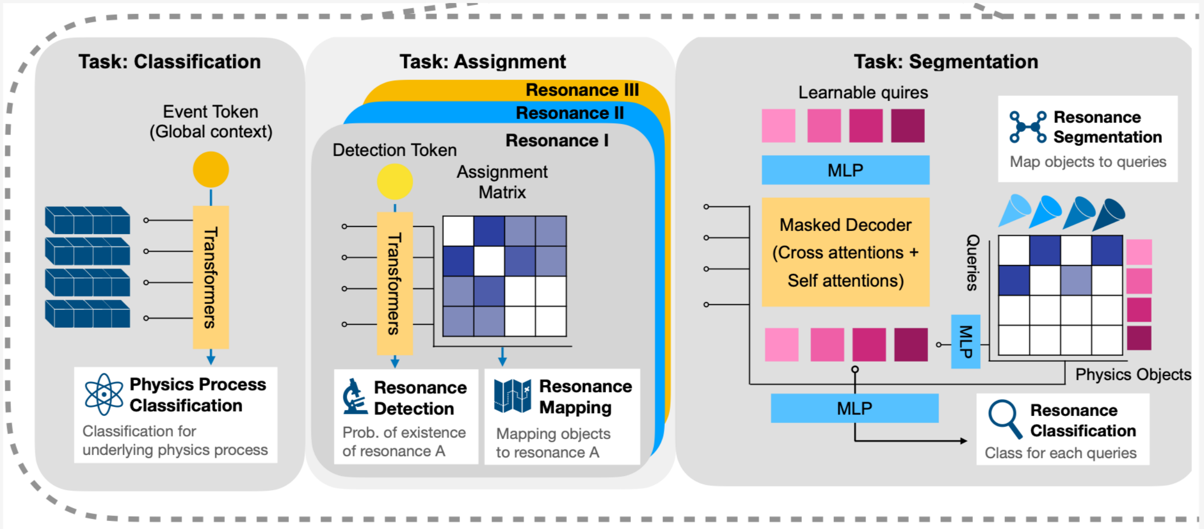
Unfolded Precision for spin correlation matrix and D

- Reference paper: Eur. Phys. J. C (2022) 82:285, assuming $139\ \mathrm{fb}^{-1}$
- The observable $D = -C_{kk} - C_{rr} - C_{nn}$ is sensitive to QE, with $D > 1$ indicating the QE.
- **Relative precision** with $\epsilon = \sigma_D/(D-1)$, *Paper:* $\epsilon_D \approx 5.26\%$

	F.T. (CLS + Gen)	Scratch	*Improvement [%]*	F.T. (CLS + Seg + Gen)	F.T. (SSL)
1.0	**1.21**	1.37	***11.88***	1.24	1.37
0.7	**1.20**	1.45	***17.13***	1.23	1.40
0.3	**1.19**	1.54	***22.29***	1.23	1.48
0.1	**1.20**	1.89	***36.19***	1.24	1.91

Search for New Physics: Overview

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



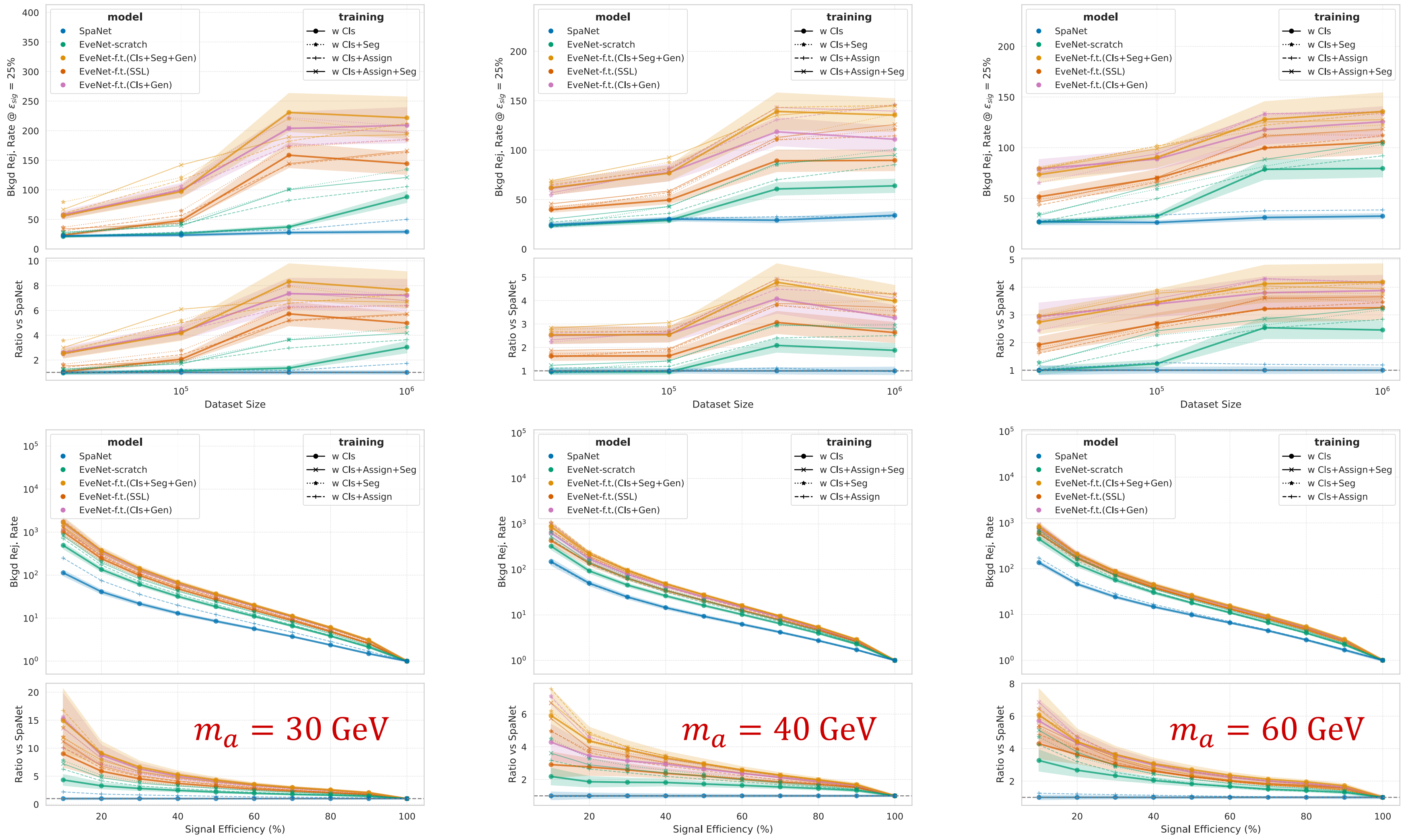
The model is jointly trained on the **Assignment** and **Classification** tasks.

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

Search for New Physics: Results

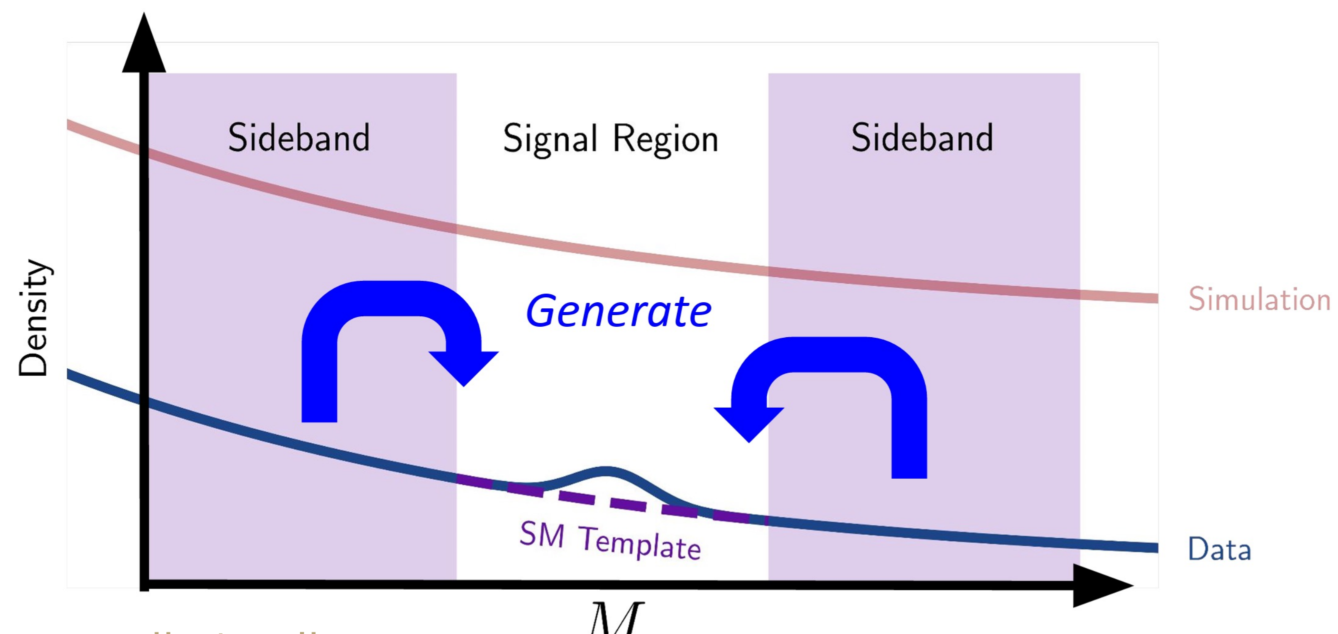
Classification

1. Inversed ROC
2. Bkgd. Rejection rate @ signal efficiency of 25%



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.



Radha's talk

Strategy Overview:

1. SR and SB definition ($m_{\mu\mu}$): $SR = [9, 10.6]$ GeV, $SB = [5, 9]$ & $[10.6, 16]$ GeV
2. **Background Modeling: ensemble of EveNet diffusion models**
3. **Weak supervision**: training XGBoost to separate generated events and data events

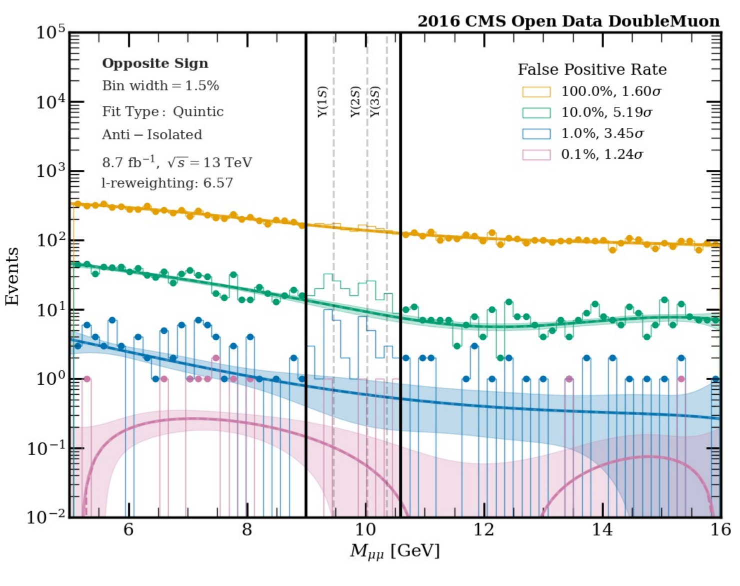
Anomaly Detection: Results

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

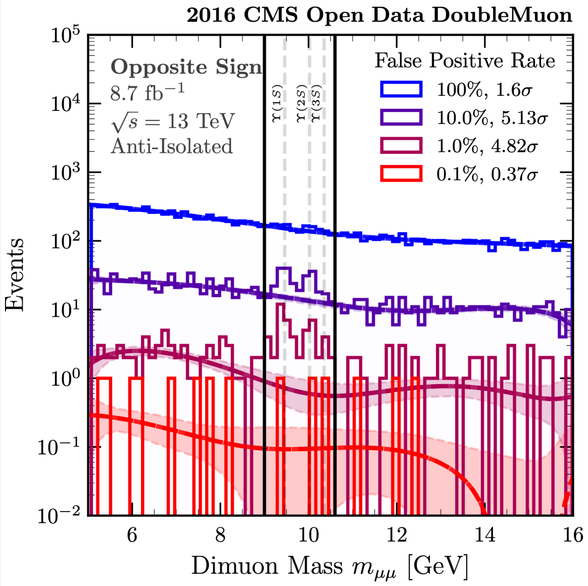
Final Significance (ℓ-reweighting)

- **paper:** $\mathbf{6.4\sigma}$
- **EveNet-Pretrain**: $\mathbf{6.54 \pm 0.24\sigma}$
- ~~**EveNet-Scratch**~~: ~~$\mathbf{7.04 \pm 0.37\sigma}$~~ (mass sculpting ❌)

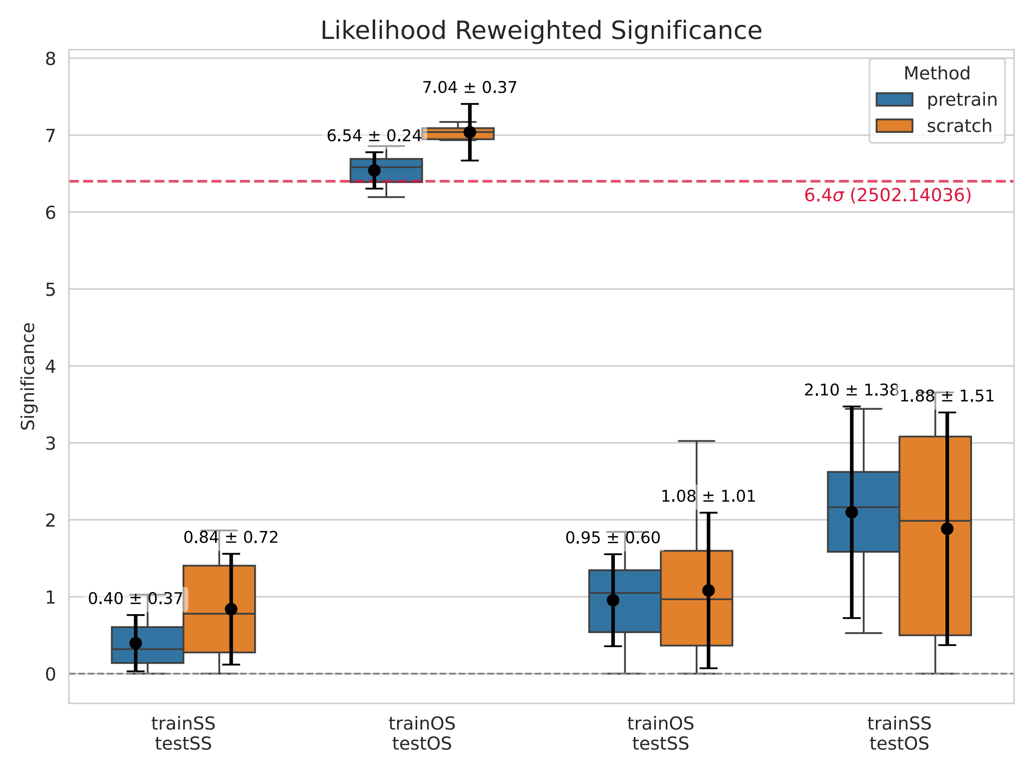
Note: *the energy regime here is even different from the main samples in pretrain*



EveNet-Pretrain



Paper's result



What We Learned?

Q: Can a foundation model in HEP really adapt to new tasks?

A: Yes! We added a new "Assignment" head (**not present during pretraining**) for QE and New Physics searches. With pretrained weights, the model immediately performed strongly → **extended to new heads and tasks** (***Homogenization***).

Q: How does it handle new physics or even new energies?

A: We tested progressively:

- **QE ($t\bar{t}$):** fully in-distribution, same CME.
- **Exotic Higgs:** out-of-distribution signal, but same CME.
- **Anomaly Detection:** fully data-driven, different CME.

In all cases, the model retained strong performance → proof of ***transferable representations*** across processes and energy scales.

Q: Can it go multimodal?

A: Yes. The current heads (especially Segmentation) can naturally extend to multimodal inputs like **tracks + clusters** or **constituents + objects**, enabling clustering and resonance reconstruction. That's the ***multimodal potential***.

❓ **Emergence**:

- New behaviors from scale

✅ **Homogenization**:

- One model, many tasks

✅ **Transferable representations**:

- Pretrain once, reuse anywhere

⚙️ **Multimodal potential**:

- Works across data types

What's Next?

Q: But what about "Emergence"?

A: In LLMs, "emergence" is a debated concept. In HEP, we need to **rethink what emergence *means***, since our evaluation metrics are very different. Defining this is an open research question.

Q: Does bigger mean better?

A: Our current model has $\sim 20M$ parameters. We're scaling to **100M** ($\sim 16{,}000$ GPU hours on 500M events). We want to test whether larger capacity yields new physics insights.

Q: Could this change how we do HEP analyses?

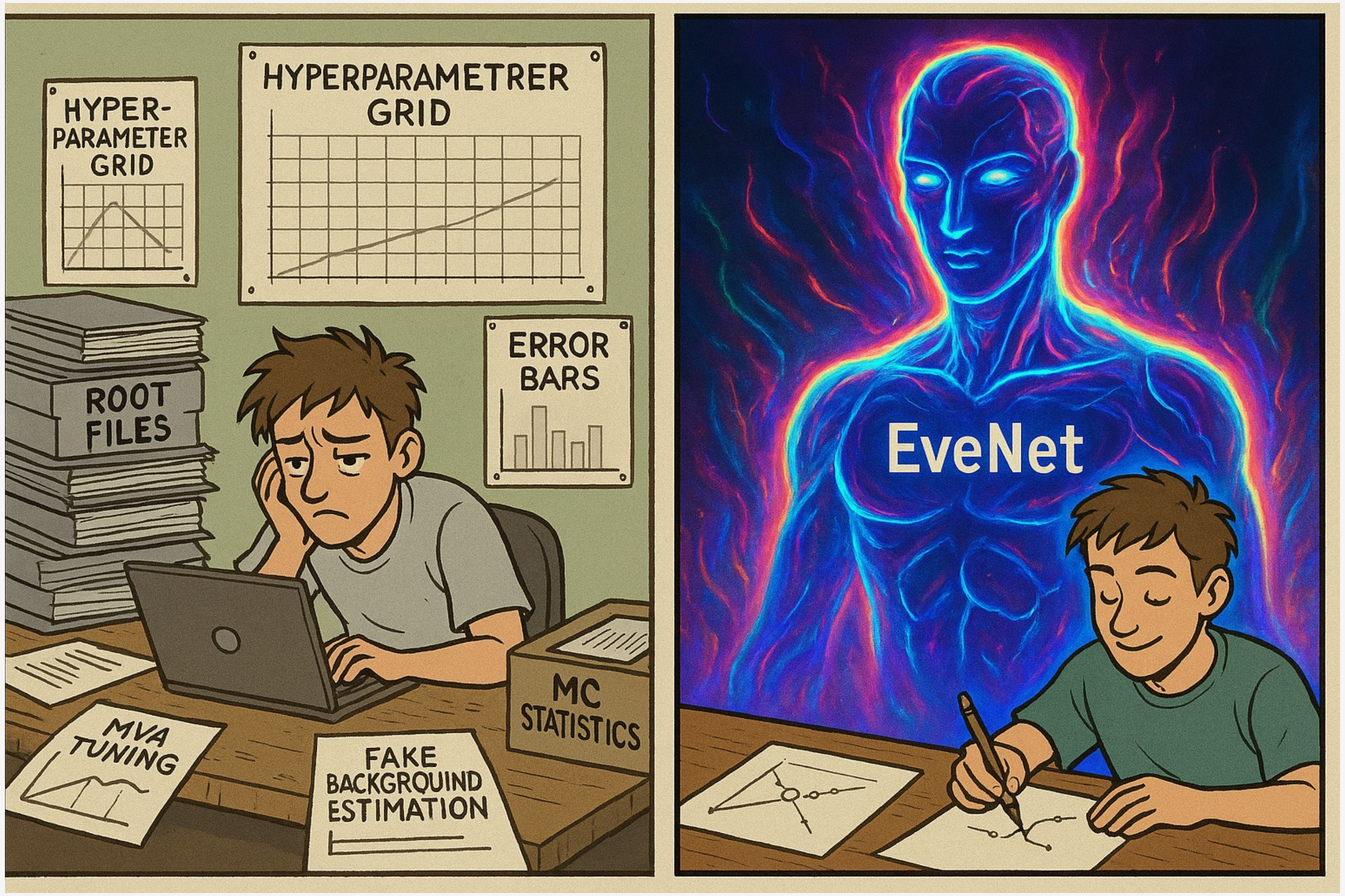
A: Yes. A strong general foundation model means **less need for bespoke, highly-tuned models** for every analysis. This could pave the way toward **auto-analysis** in HEP → a step toward AGI-like helpers for particle physics.

Q: Any other ideas?

A: Plenty! For example, exploring **Mixture-of-Experts (MoE)** architectures to improve scalability and specialization. Foundation model is the pillar for **Reasoning Large Model** (RLM, *DeepSeek* 🤩).

EveNet: Powering the Next Physics Breakthrough

a foundation model to solve all HEP problems

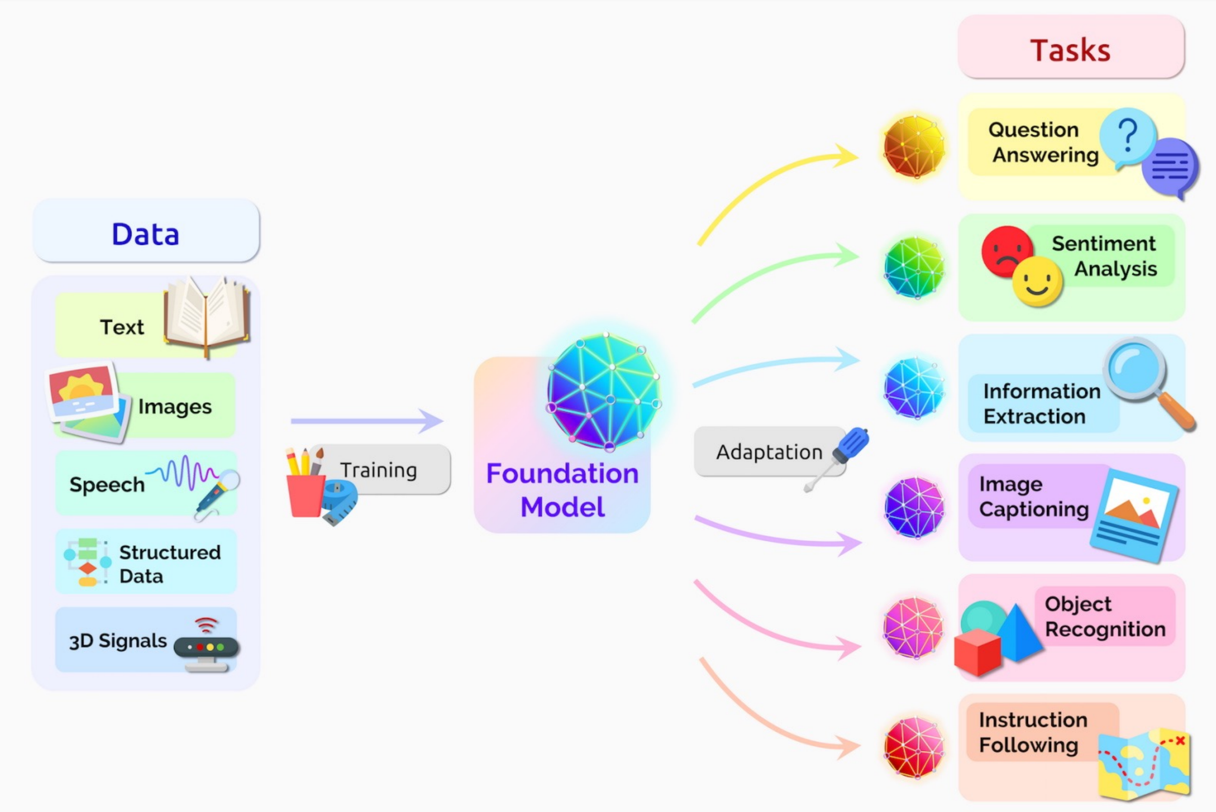


Backup

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

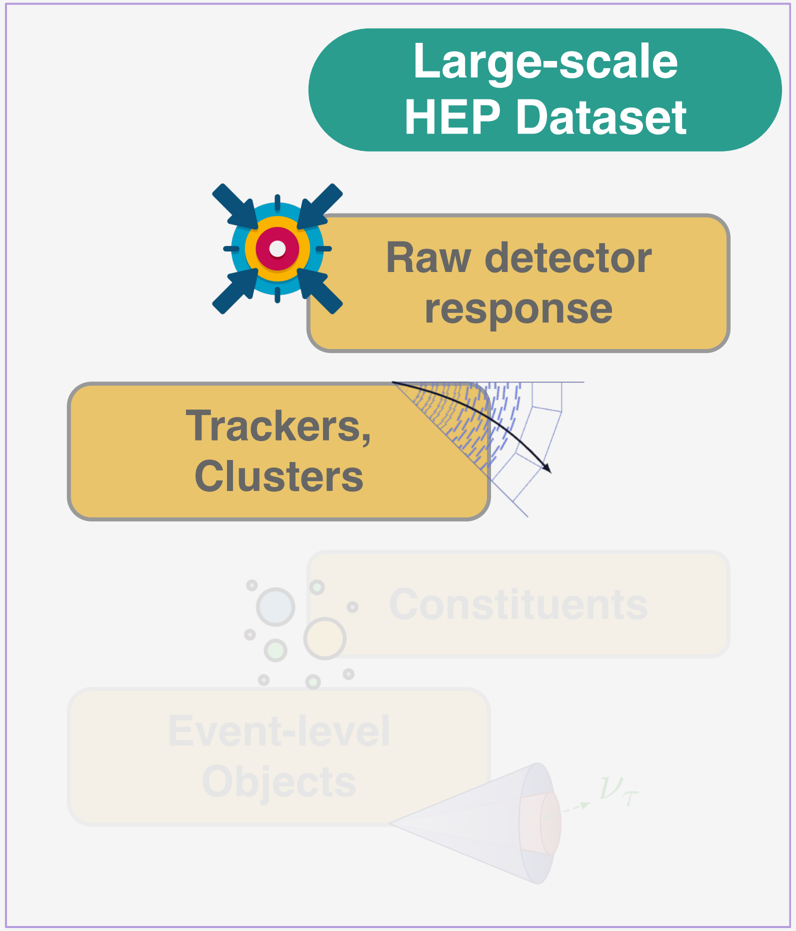
🔄**Transferable representations**:

- Pretrain once, reuse anywhere

🌐 **Multimodal potential**:

- Works across data types

Can We build a Foundation Model for HEP?



Primitive…

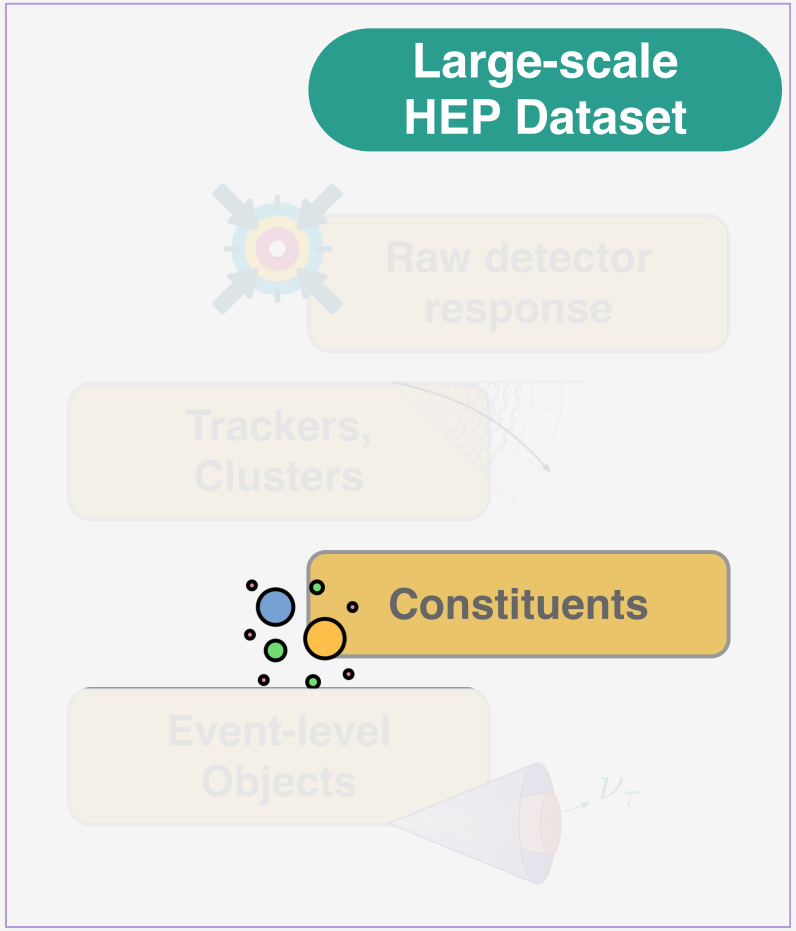
Pros:

- Most **information-rich**; closest to detector truth.
- **Minimal bias** from reconstruction choices.
- Potential for ML to learn **end-to-end physics object reconstruction** directly from detector signals.

Cons:

- **Very high dimensionality** → massive computational and storage requirements.
- Harder to interpret results in physics terms.
- Needs careful handling of detector effects, noise, and calibrations.
- Farther from final physics observables, so **less transparent** to analysts.

Can We build a Foundation Model for HEP?



Particle Flow objects, uncalibrated

<u>Pros</u>:

- **Richer** information than event-level; captures **substructure of jets and taus**.
- **Less dependent** on final reconstruction/ID algorithms.
- More **flexibility** for ML to discover new features (e.g., jet substructure, pile-up mitigation).

<u>Cons</u>:

- Still subject to PF reconstruction algorithms (track–cluster linking).
- **Higher dimensionality** → increased computational cost.
- **Not yet fully calibrated** → may need additional corrections before use in physics interpretation.

Machine Learning + HEP

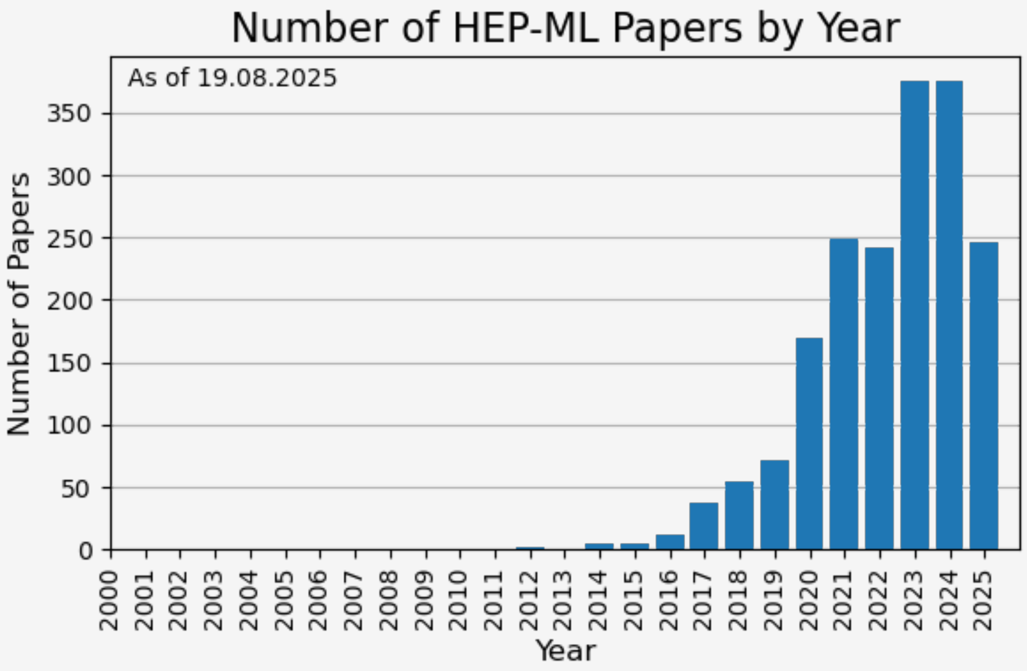
A Living Review of Machine Learning for Particle Physics

- There is a growing interest in applying machine learning techniques to HEP.
- The concept of foundation models has also been explored in recent studies.

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 🧐

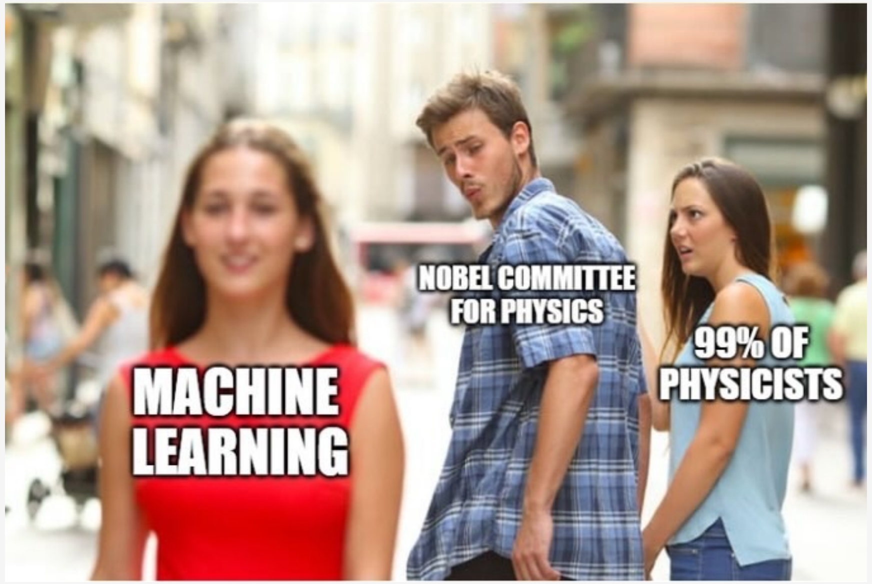
- Extensively pre-trained for general-purpose representations.
- **Lightly fine-tuned** for task-specific applications.
- Especially **effective** in scenarios with **limited** training data.

Foundation Model 🥳

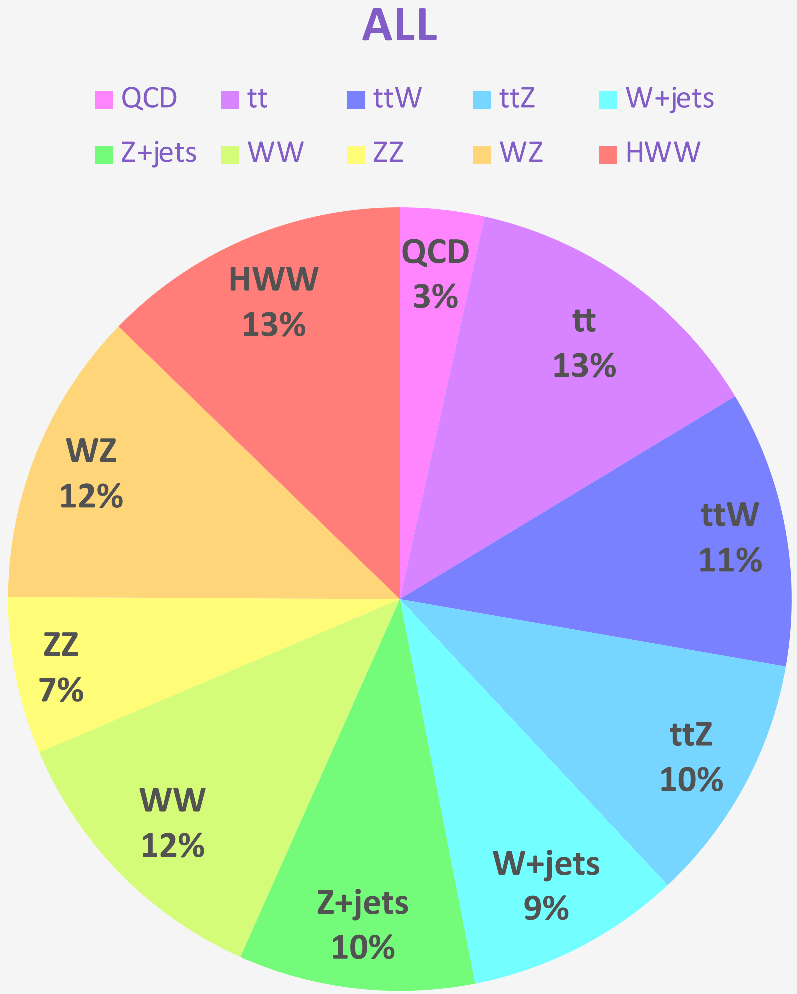
- Enables **a unified understanding** of HEP events.
- Designated to generalize across a wide range of tasks.

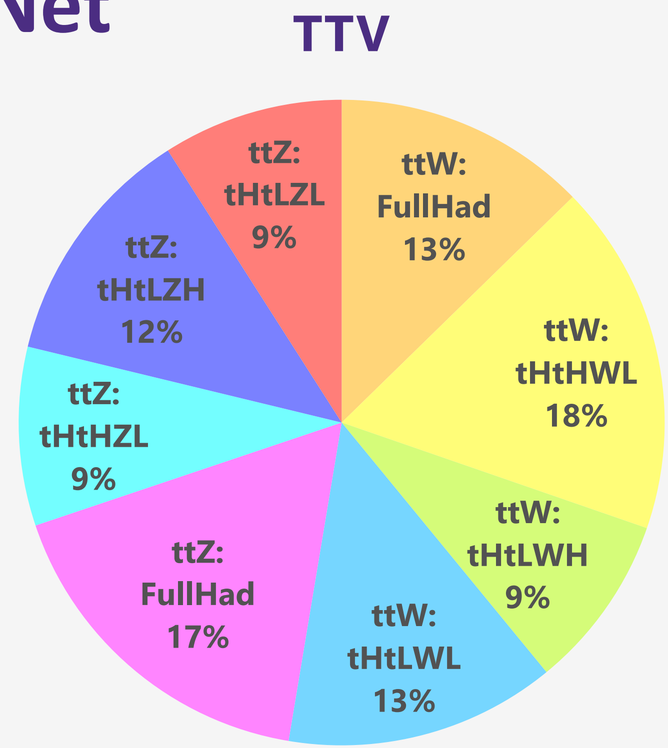
🔧 Core Ingredients of an Event-Level Foundation Model in HEP

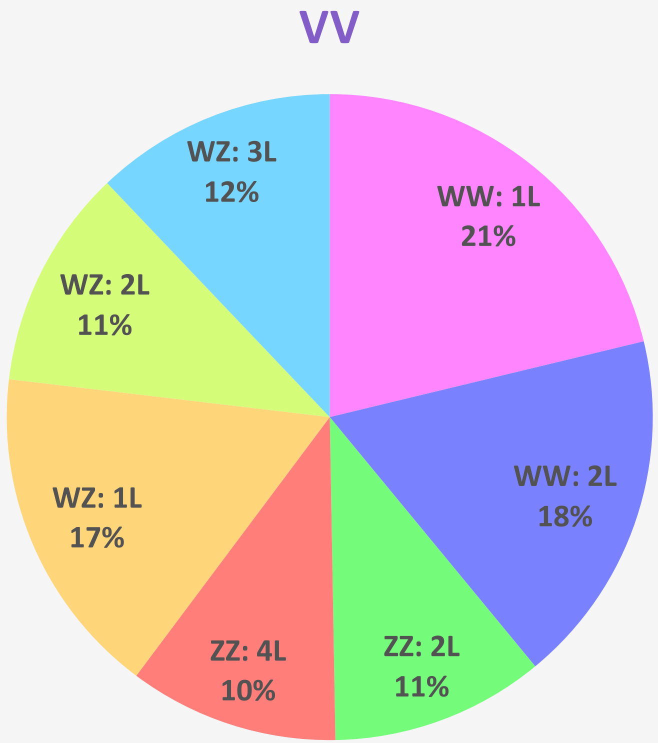
- 🧠 **Generalist Embedding**: Shared event-level representation
- 🧩 **Multi-task Learning**: One model, many objectives
- 🔍 **Self-Supervised Pretraining**: Learns from data structure
- 📈 **Scalability**: Improves with more data + compute
- 🔁 **Transferability**: Fine-tune for new tasks easily



How do we pre-train EveNet







- 🔧 **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.

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: Results

Unfolded Uncertainty for spin correlation matrix and D

- Reference paper: Eur. Phys. J. C (2022) 82:285, assuming $139\ \mathrm{fb}^{-1}$
- The observable $D = -C_{kk} - C_{rr} - C_{nn}$ is sensitive to QE, with $D > 1$ indicating the QE.
- **Relative precision** with $\epsilon = \sigma_D/(D-1)$, *Paper:* $\epsilon_D \approx 5.26\%$

	Recon ν + Recon pairing			Recon ν + Truth pairing			Truth ν + Recon pairing		
	Fine-tuned	Scratch	**imp. [%]**	Fine-tuned	Scratch	**imp. [%]**	Fine-tuned	Scratch	**imp. [%]**
1.0	1.24	1.37	**9.78**	1.25	1.38	**9.41**	0.69	0.69	**0.06**
0.7	1.23	1.45	**15.21**	1.24	1.48	**15.98**	0.69	0.70	**0.99**
0.3	1.23	1.54	**20.06**	1.24	1.55	**19.70**	0.69	0.70	**2.59**
0.1	1.24	1.89	**34.45**	1.25	1.70	**26.62**	0.69	0.80	**14.61**

Quantum Entanglement: Observations

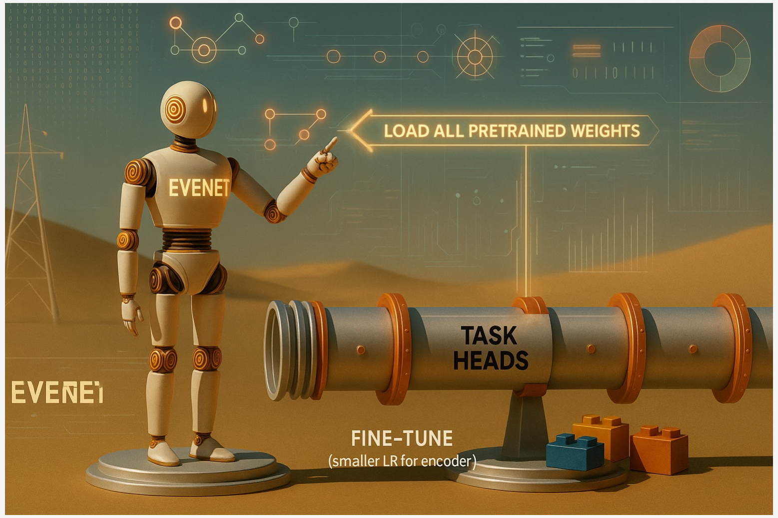
- ⚙️ **Pretrained model shows improved assignment performance**, increasing matching efficiency by:
  - **+2.5%** for **matchable events**
  - **+2.1%** for **all events**
- 📉 **Uncertainty reduction**:
  - **Absolute improvement** of $\sim 12.5\%$ in precision over the scratch model
  - **Relative precision 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

EveNet Playbook: A guide to downstream applications

Option 1: Plug-and-Play

(Pretrained Encoder + Heads)

- Keep input/output format as in pretraining.
- Load pretrained weights for encoder (except normalizers).
- Optionally load head weights as initialization.
- Turn on task heads you need (can combine multiple).



Option 2: Customized Player

(Pretrained Encoder + Your Own Decoder)

- Input features of point cloud should match pretrained ones.
- Global features can be customized.
- Encoder outputs embedded event representation.
- You design your own decoder/heads for specific tasks.
- Only load pretrained encoder weights.

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 **ensemble of EveNet diffusion models**
     - **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: 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**.

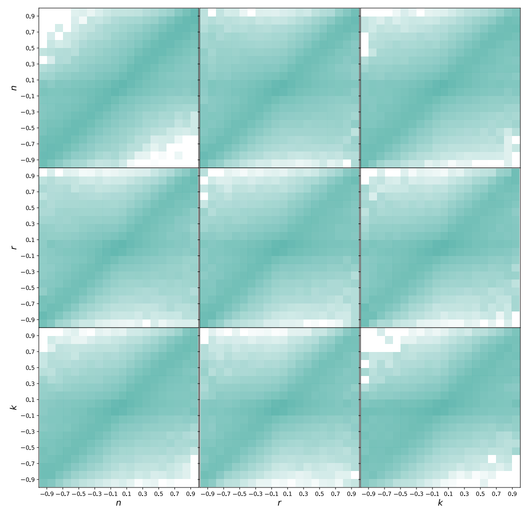
⚙️ **Generation Efficiency**:

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

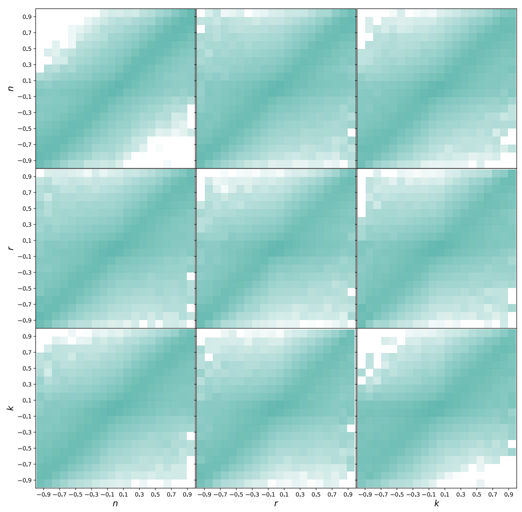
🧠 **Analysis-Specific Limitation**:

- 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.

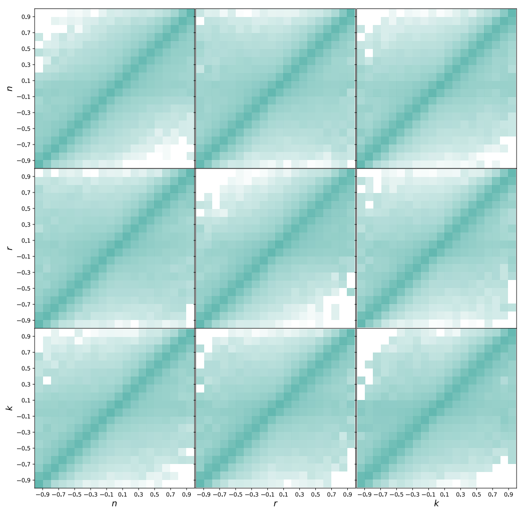
QE response matrix (Fine-tuned)



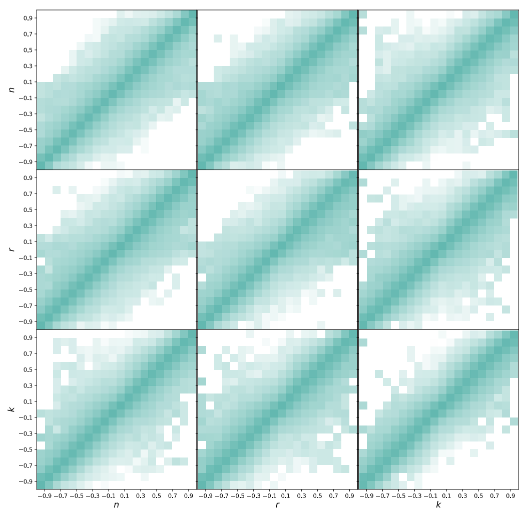
Recon-ν + Recon Pairing



Recon-ν + Truth Pairing

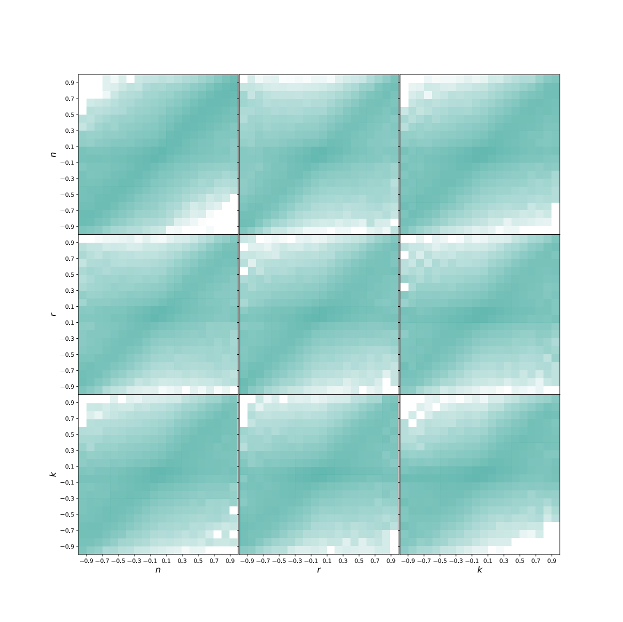


Truth-ν + Recon Pairing

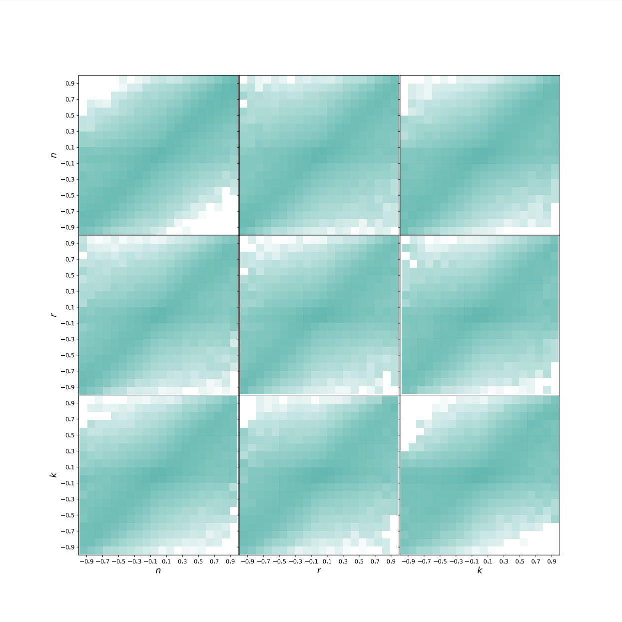


Truth- ν + Truth Pairing

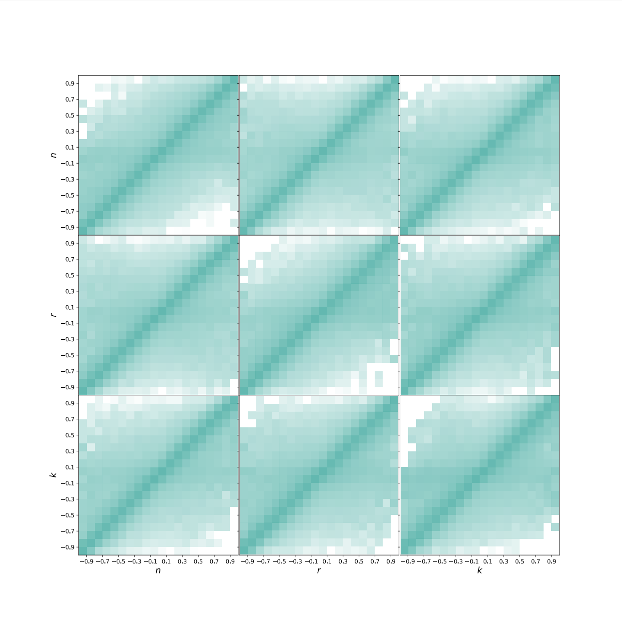
QE response matrix (Scratch)



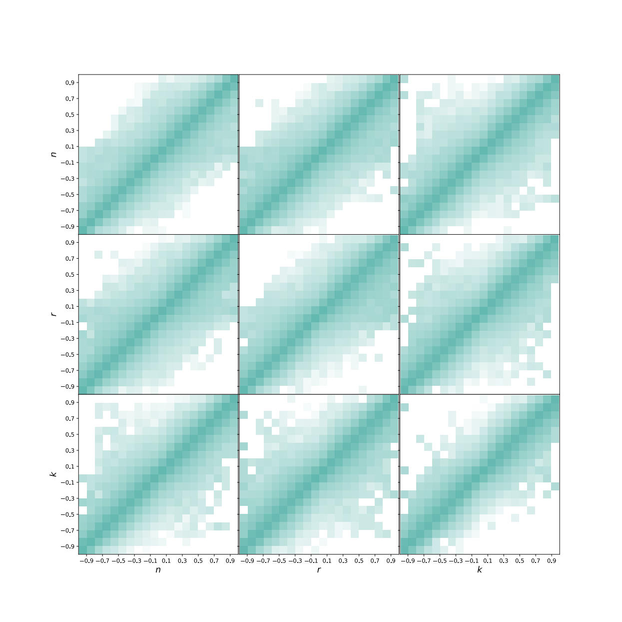
Recon-ν + Recon Pairing



Recon-ν + Truth Pairing



Truth-ν + Recon Pairing



Truth- ν + Truth Pairing

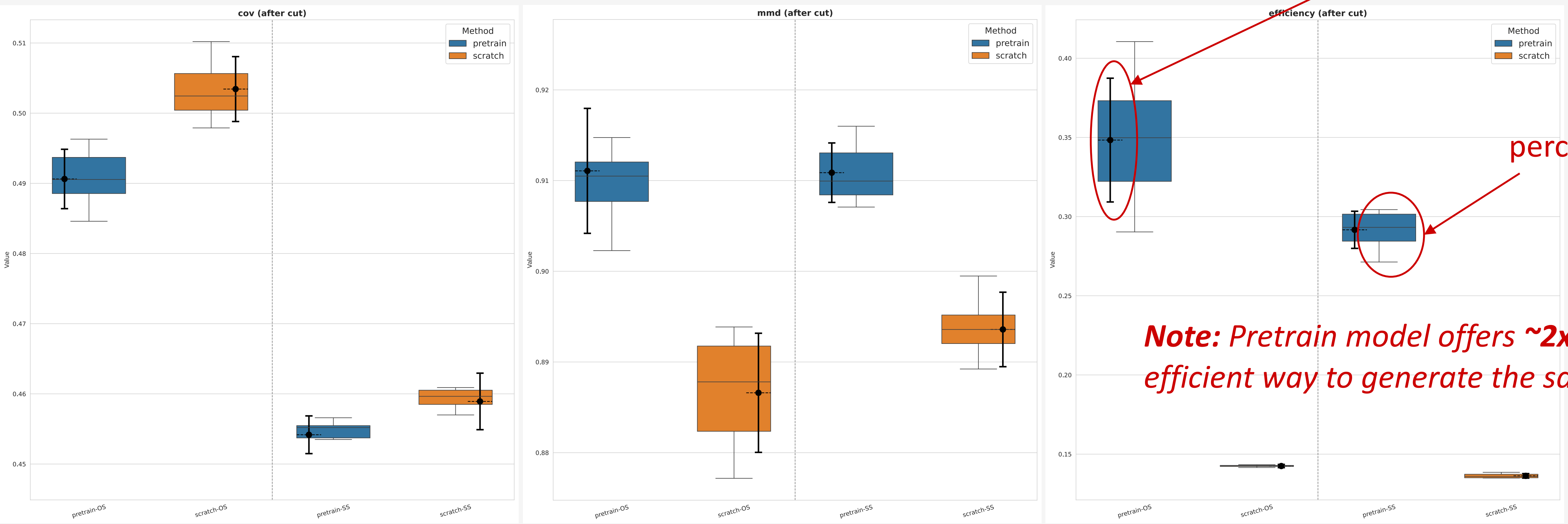
Anomaly Detection: Results

🌟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

$\mu \pm \sigma$



***Note:** Pretrain model offers **~2x** more efficient way to generate the samples*

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/segmentation head**.

Quantum Entanglement

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

Assignment & Generation

+2% assignment, **10% uncertainty reduction**, **$\sim$75% 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 training a **100M-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!**