FM4NPP: A Scaling Foundation Model for Nuclear and Particle Physics

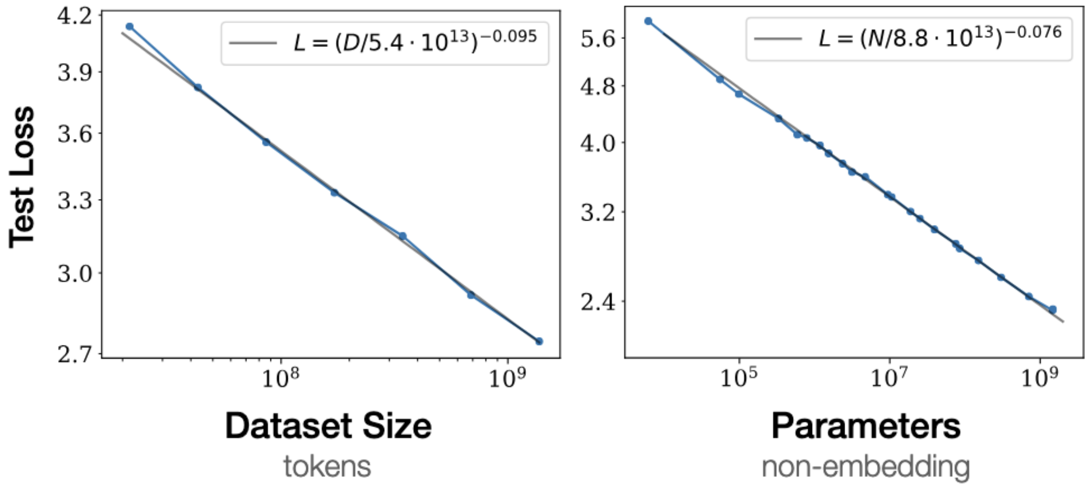
Shinjae Yoo (sjyoo@bnl.gov)
Distinguished Scientist & Chair of Artificial Intelligence Department
Computational Research Division, Brookhaven National Laboratory

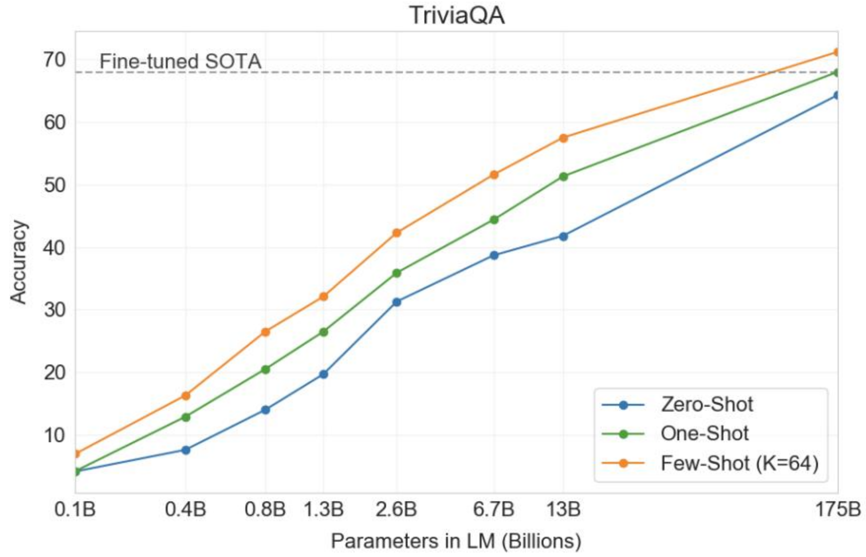
@BrookhavenLab

Technical Presentation to PARTIQAL | July 14, 2026

Main Features of LLM

- **Self-supervised** pre-training on unlabeled data.
- **Neural Scaling Behavior**: the performance scales with the model size and the dataset size.
- **Generalizability**: A single model can be used on multiple downstream tasks.





However, scientific data are not natural languages nor natural images.

Brookhaven National Laboratory

Foundation Model

Foundation Models (FMs) are envisioned as a counterpart to text-based LLM, but they can handle multiple types of data.

- Built on large-scale, primarily unlabeled data
- Capable of handling multiple modalities
- Trained via self-supervised learning on surrogate tasks
- Pre-trained and adaptable to diverse downstream applications
- Achieve state-of-the-art performance across application tasks
- Exhibit strong neural scaling behavior

arXiv > cs > arXiv:2108.07258

Computer Science > Machine Learning

[Submitted on 16 Aug 2021 (v1), last revised 12 Jul 2022 (this version, v3)]

On the Opportunities and Risks of Foundation Models

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx

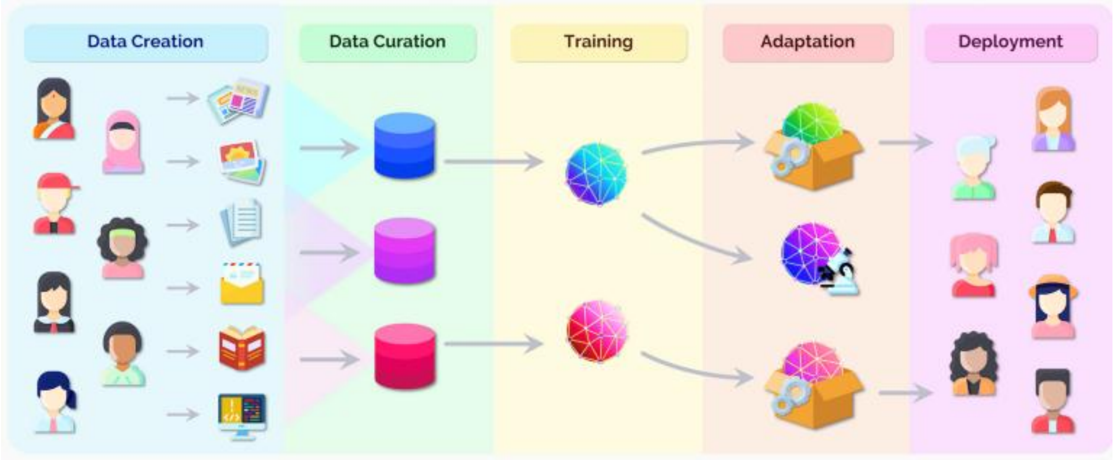


Image credit [1]

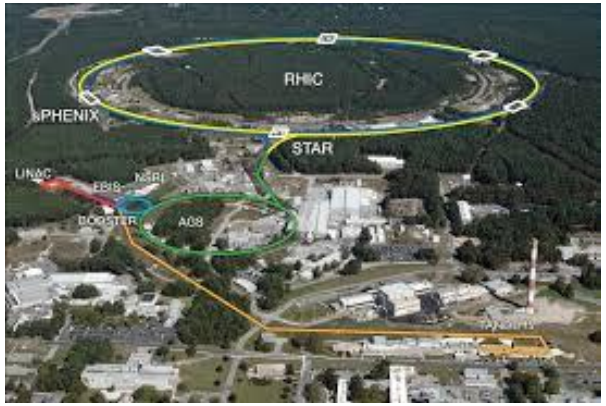
Can we build one for Nuclear and Particle Physics (NPP)?

[1] Bommasani, R. et al. On the Opportunities and Risks of Foundation Models. Preprint at https://doi.org/10.48550/arXiv.2108.07258 (2022).

Brookhaven National Laboratory

A Perfect Fit!

- **Broad impact**: RHIC/EIC, LHC, future circular collider (FCC), etc.



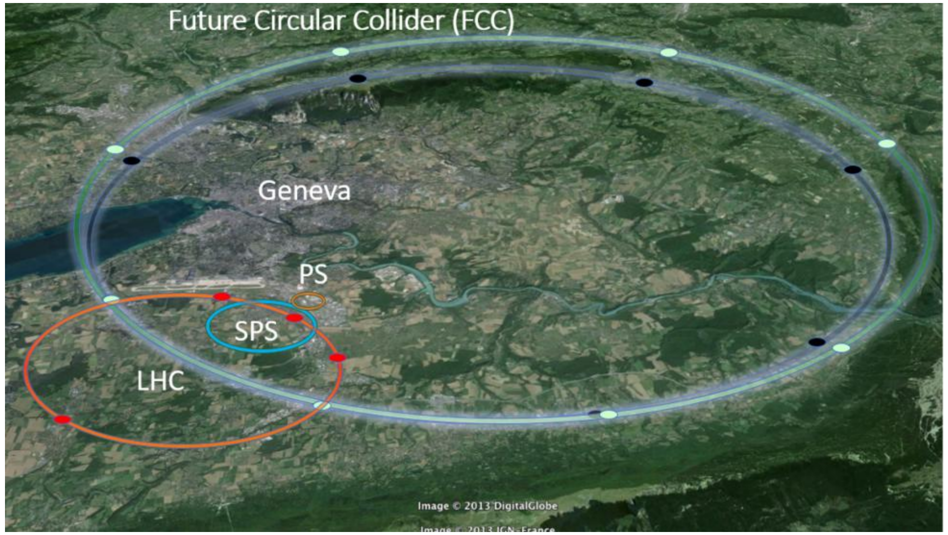
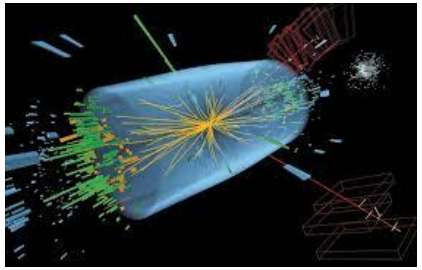
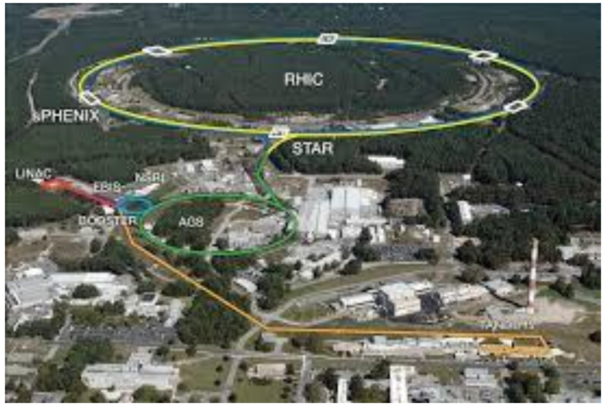


Image Credit: ICMAB

A Perfect Fit!

- **Broad impact**: RHIC/EIC, LHC, future circular collider (FCC), etc.
- These experiments share a fundamental methodology—colliding particles at relativistic speeds to probe subatomic interactions via high-precision, specialized detector systems.





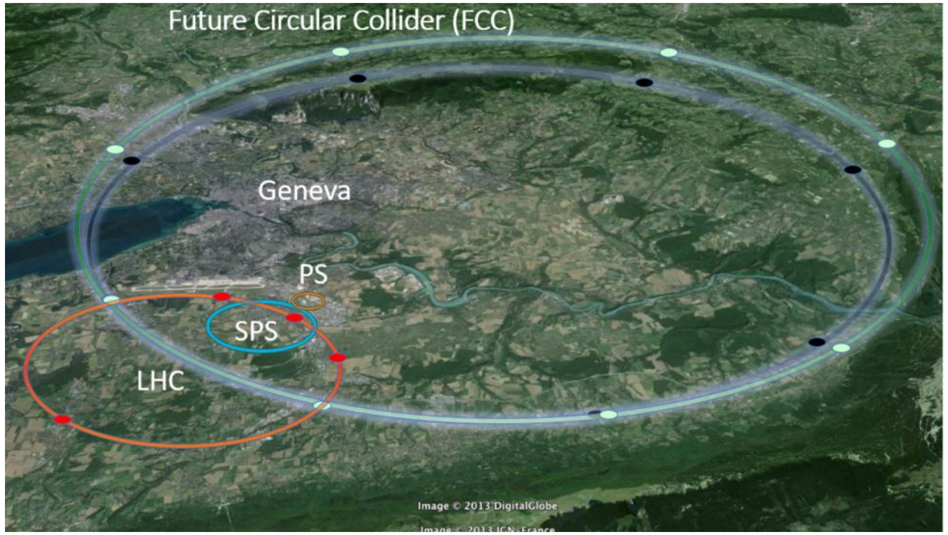
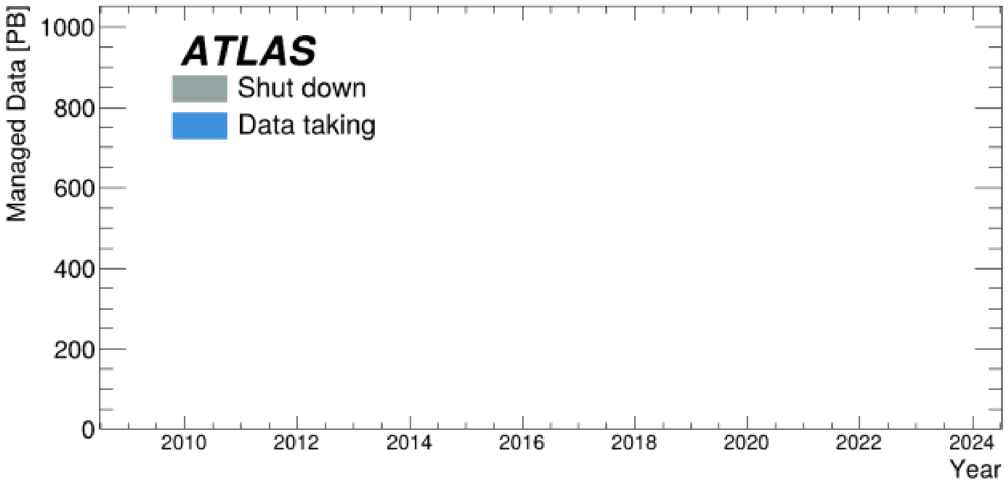


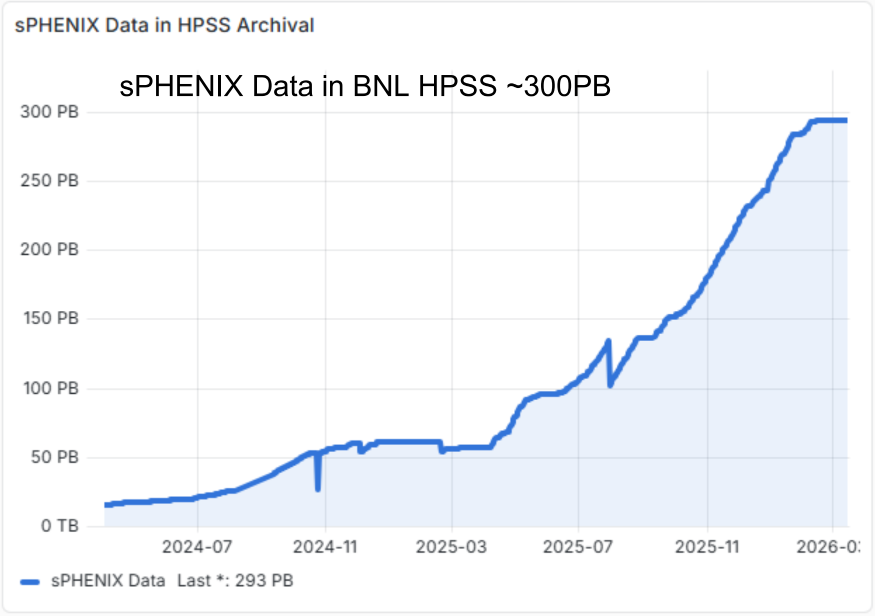
Image Credit: [ICMAB]

A Perfect Fit!

- **Broad impact**: RHIC/EIC, LHC, future circular collider (FCC), etc.
- These experiments share a fundamental methodology—colliding particles at relativistic speeds to probe subatomic interactions via high-precision, specialized detector systems.
- As a result, a large amount of experimental data are generated.



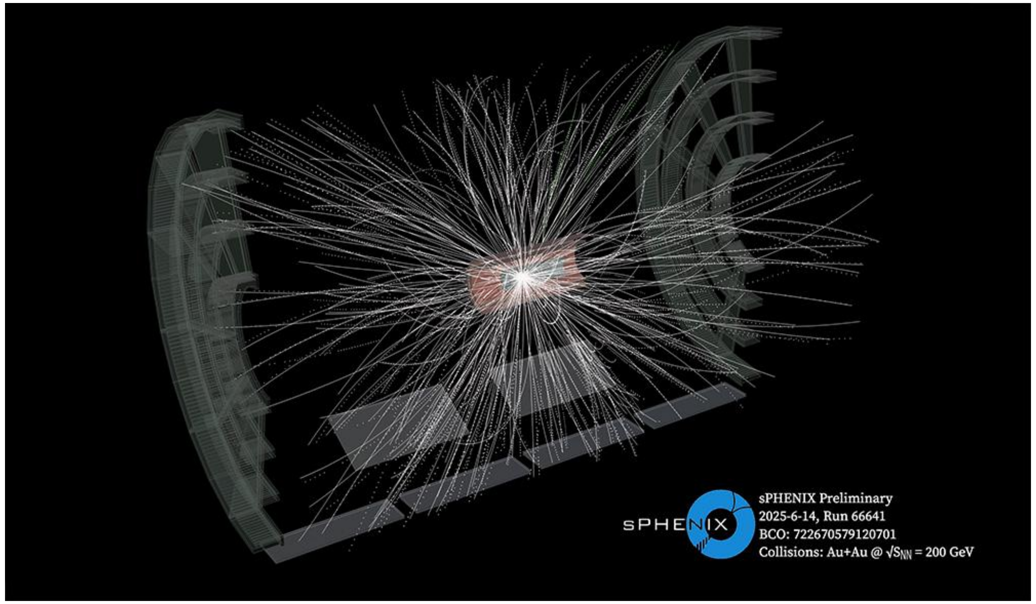
2024/07/10: [ATLAS data-taking milestone: 1EB of data]



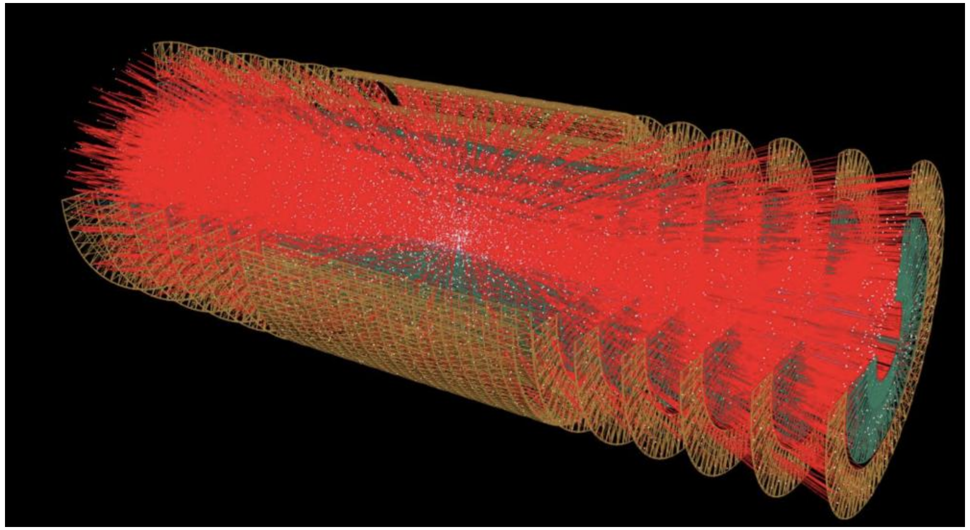
Brookhaven National Laboratory

A Perfect Fit!

- **Broad impact**: RHIC/EIC, LHC, future circular collider (FCC), etc.
- These experiments share a fundamental methodology—colliding particles at relativistic speeds to probe subatomic interactions via high-precision, specialized detector systems
- As a result, a large amount of experimental data are generated.
- The data are extremely challenging to analyze, and require large international collaborations working for years.



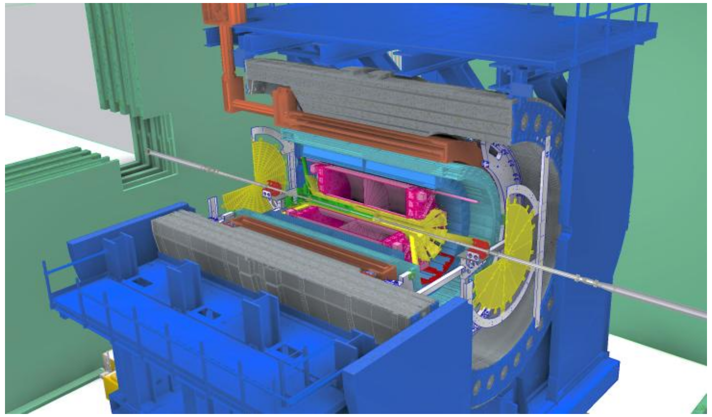
2026/02/02: [Sparse spacepoints in sPHENIX]



TrackML Detector challenge. arXiv:1904.06778

Frontier nuclear physics collaborations: sPHENIX and ePIC Experiments

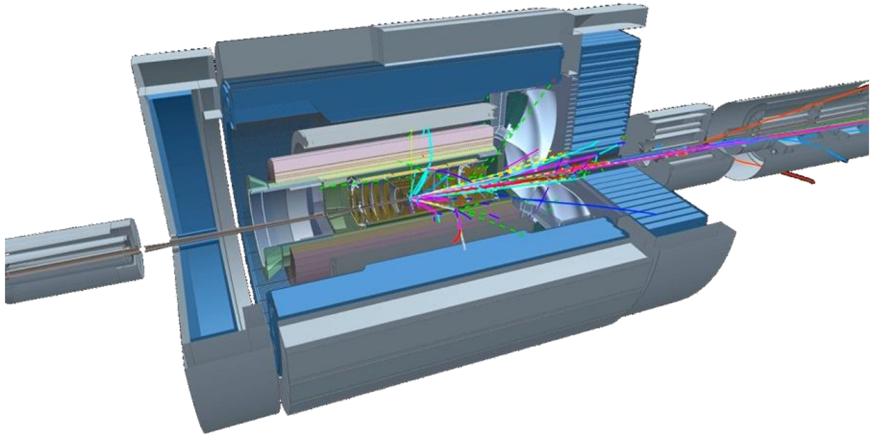
sPHENIX



The latest collider experiment, exploring the primordial soup of quark and gluon ~1us after the big bang.

- **Data Rate:** ~ 300 Gbps (storage)
- **Collaboration:** ~250 people, representing 45 institutes from 17 countries

ePIC (EIC's 1st detector)



A Deep Inelastic Scattering (DIS) experiment. Its goal is to look inside the proton and nuclei to understand how they are built.

- **Estimated Data Rate:** ~ 100 Gbps (storage)
- **Collaboration:** ~1200 people, representing 183 institutes from 26 countries

Brookhaven National Laboratory

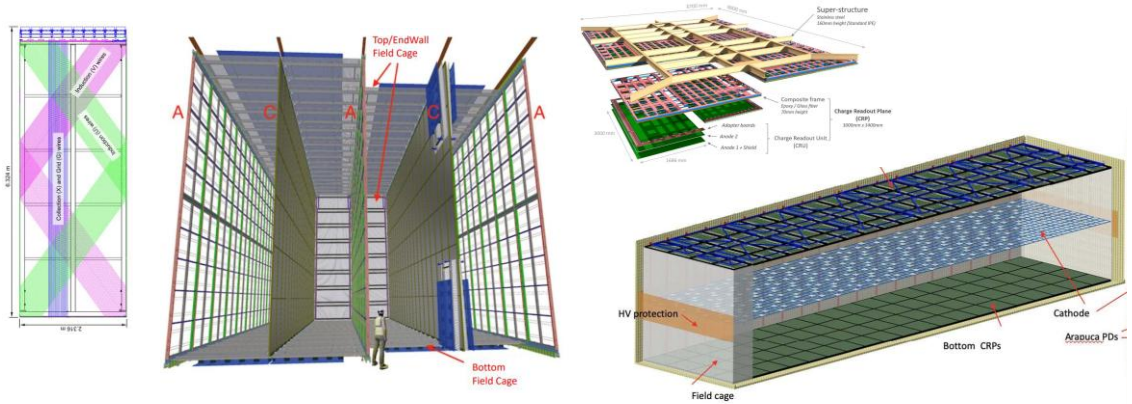
Frontier particle physics collaborations: ATLAS and DUNE Experiments

ATLAS

The energy frontier explorer. Beyond Higgs Boson. Measuring Higgs behaviors more precisely.

- **Data Rate**: ~Pb/s raw data (40 MHz) → ~0.5 Tb/s storage (10 kHz)
- **Collaboration**: >5500 members from ~280 institutes in ~40 countries
- **Technical Challenges**: extreme data rates, high-speed radiation-hard electronics, real-time event selection

DUNE



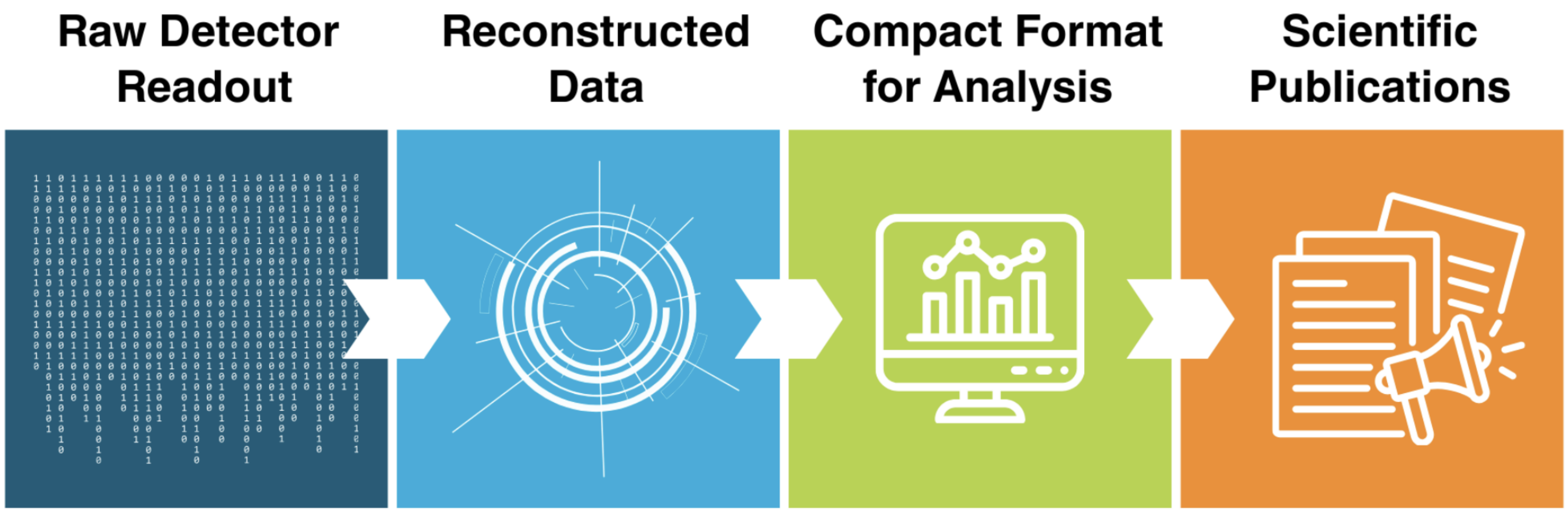
Studying the fundamental nature of matter and the pivotal role of neutrinos in the evolution of the universe.

- **Data Rate**: ~Tb/s continuous streaming, PB-scale annual storage
- **Collaboration**: >1400 members from >200 institutions in >30 countries
- **Technical Challenges**: electronics at 77 K, achieving high signal-to-noise ratio.

Actual Size of the sPHENIX Detector

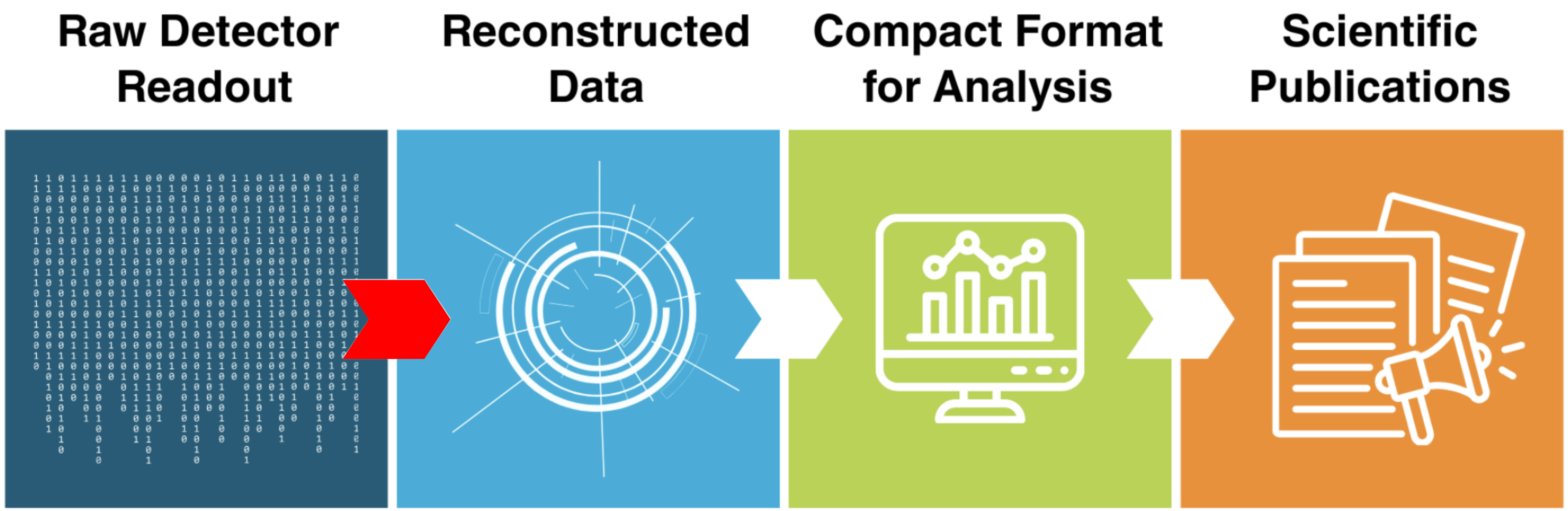
Feb 6, 2026: Under Secretary Dario Gil visiting sPHENIX following success completion of data taking

A typical scientific workflow of NPP



This process from data taking to publication usually take several years!

A typical scientific workflow of NPP

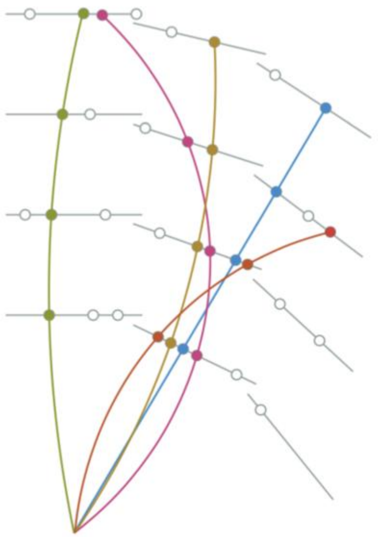


The fundamental challenge lies in the **reconstruction and interpretation** of raw detector signals into meaningful physics objects.

Brookhaven National Laboratory

Some examples of subtasks

- **Particle Tracking:** Pattern recognition to reconstruct trajectories from discrete sensor hits.
- **Particle Identification (PID):** Classifying species using specialized signatures like Cherenkov radiation or energy loss.
- **Noise & Background Tagging:** Distinguishing primary collision products from electronics noise, pile-up, and secondary material interactions.
- **Calorimetry (Clustering):** Grouping adjacent energy deposits to measure the total energy of neutral and charged particles.
- **Vertexing:** Locating Primary Vertices (collision points) and Secondary Vertices (displaced decays) with micron-level precision.
- **Jet Reconstruction:** Clustering particles into collimated sprays to probe the underlying Quark-Gluon dynamics.



Particle tracking is the most fundamental task.

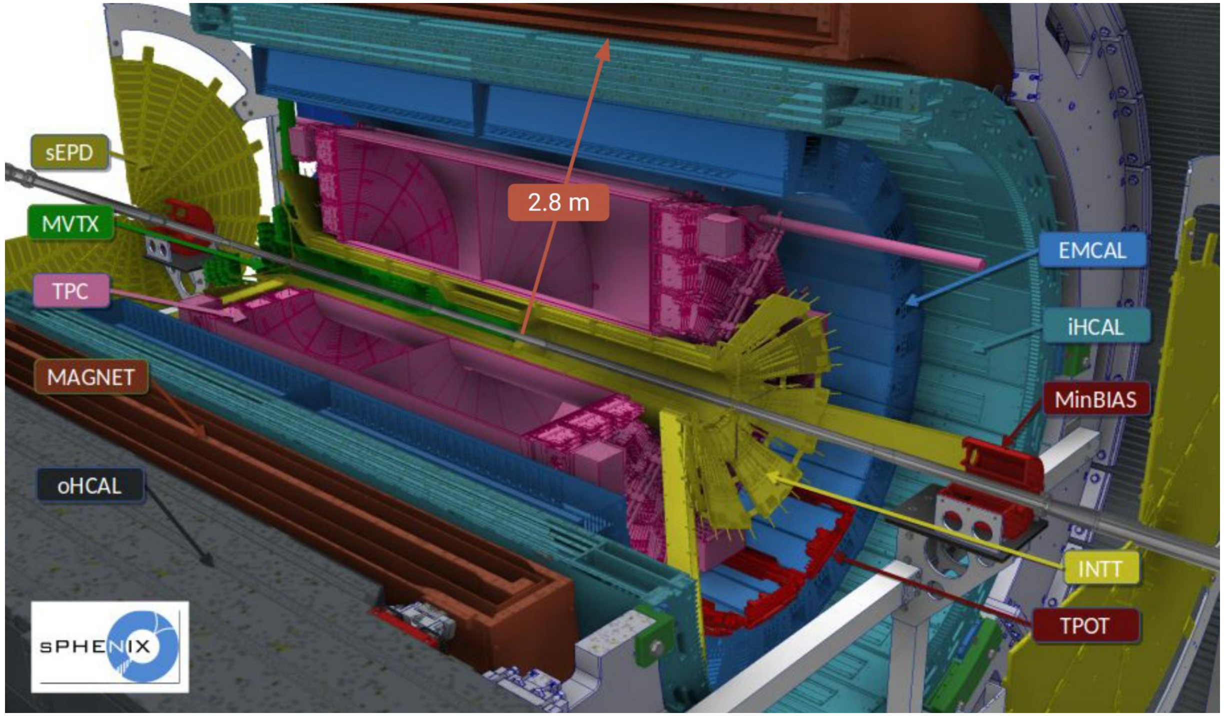
Why we chose sPHENIX?

AI-Ready Data:

- A large amount of realistically simulated data with labels.
- sPHENIX were collecting real data.
- Multiple subdetector systems.
- sPHENIX data are DOE-owned.

Productive Collaboration:

- > 10 co-authored papers since 2021.



Brookhaven National Laboratory

Scientific Goal and Achievements

Proof of concept for a FM4NPP:

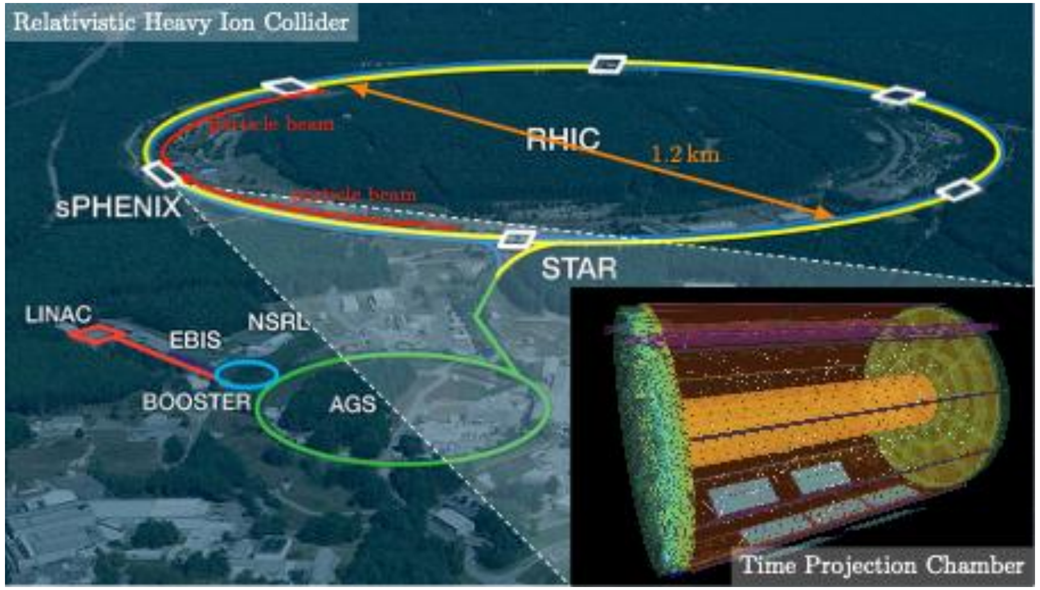
1. Neural scaling behavior
2. Generalizable to downstream tasks

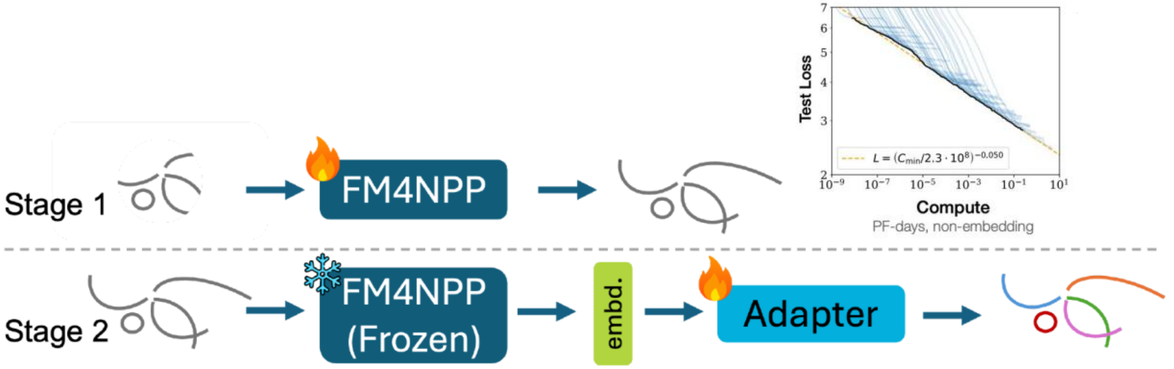
Two-stage approach:

1. Large-scale FM pre-training on unlabeled data
2. Adapt the frozen FM for various tasks

Key Achievements:

1. Identified a self-supervised learning task for NPP sparse spacepoints data.
2. Verified the neural scaling behavior of FM pre-training.
3. Demonstrated the effectiveness of FM approach on a diverse downstream tasks both in terms of performance and data efficiency.





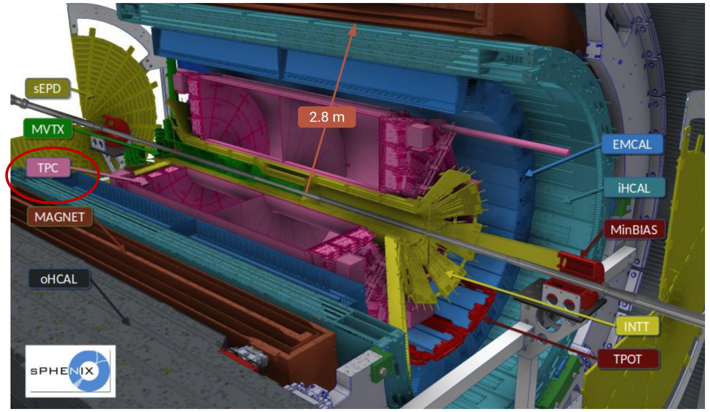
Brookhaven National Laboratory

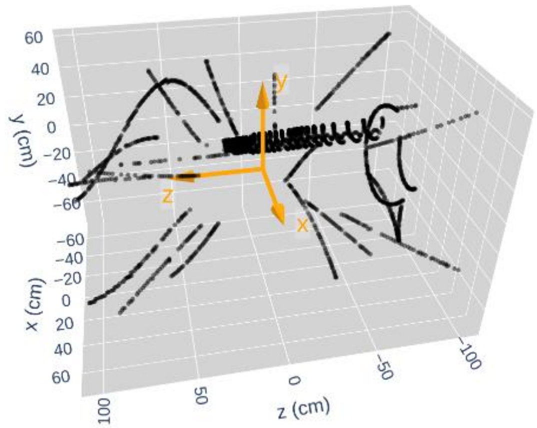
Simulated Dataset

sPHENIX Time Projection Chamber (TPC) spacepoints

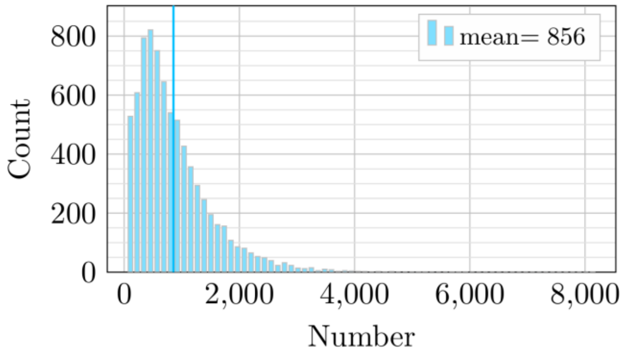
Data:

- **10M** events for self-supervised pre-training
- **75k** events for downstream tasks supervised training.
- p+p collisions at √s = 200 GeV
- Simulated using sPHENIX software tool chains.
- The simulation includes continuous energy loss, multiple scattering, secondary particle production, and decay processes when interacting with the detector material.

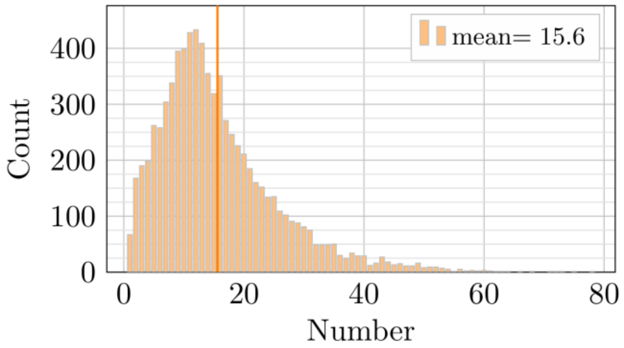




(a) Number of spacepoints distribution



(b) Number of tracks distribution

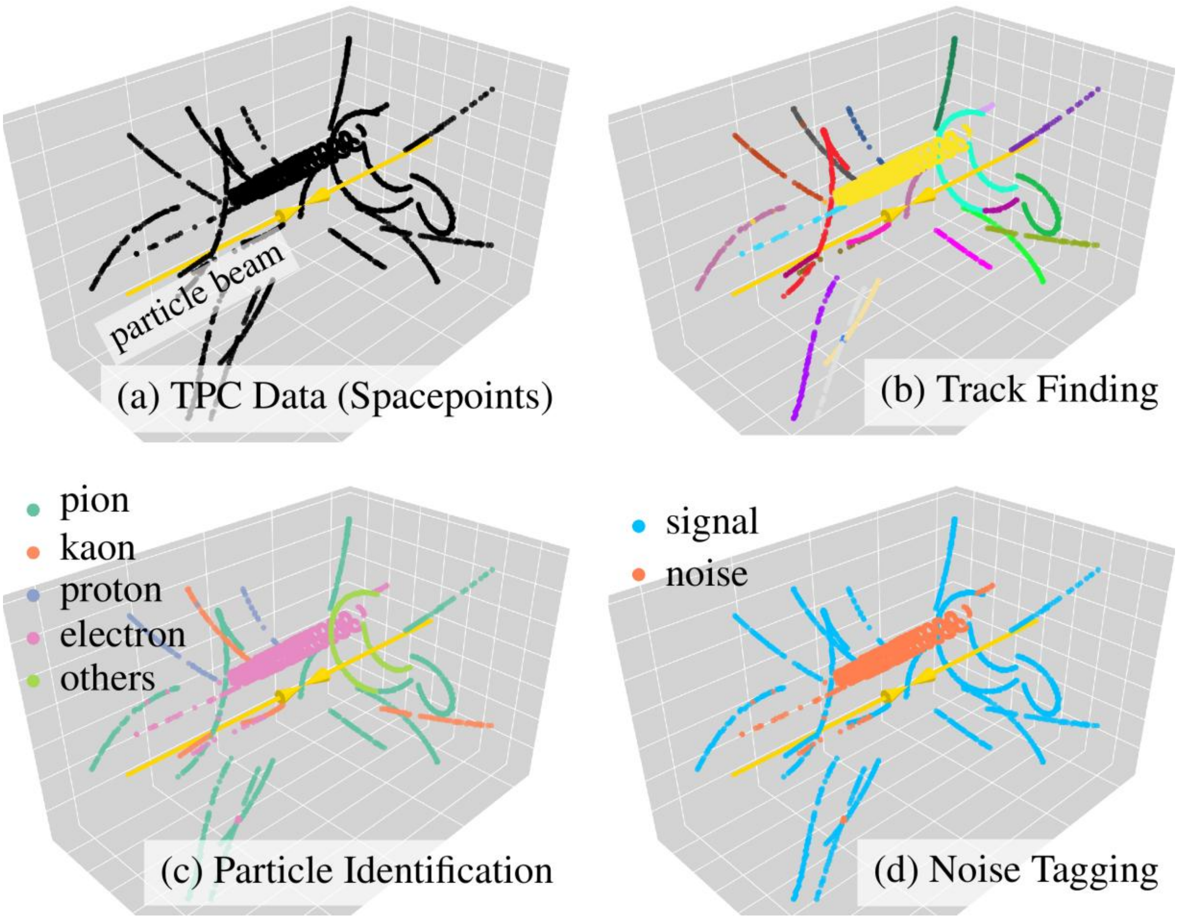


Brookhaven National Laboratory

Downstream Tasks

We identified three downstream tasks to verify the generalizability of the FM:

1. **Track Finding**: group spacepoints from the same particle together
2. **Noise Tagging**: classify spacepoints from low momentum secondary particles (“noise”)
3. **Particle Identification**: classify spacepoints based on their particle species

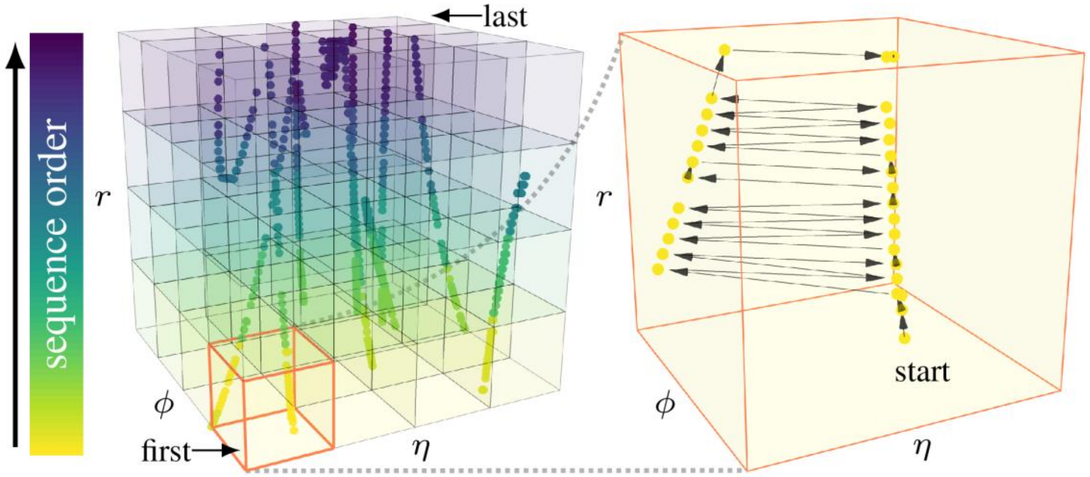


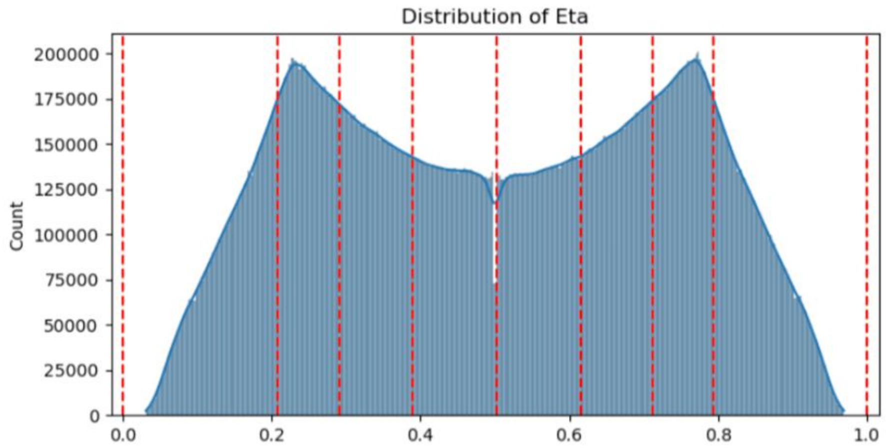
Hierarchical Raster Scan: `3D -> 1D`

HRS serializes spacepoints into a 1D sequence.

- Transform spacepoints from Cartesian (x, y, z) to a cylindrical-polar system (r, ϕ, η).
- Divide the space into boxes, 6x8x8 in (r, η, ϕ).
- Within each box, order the points based on radius.
- Boxes are ordered from inner most box to the outermost box.

(a) 3D grid partition and ordering



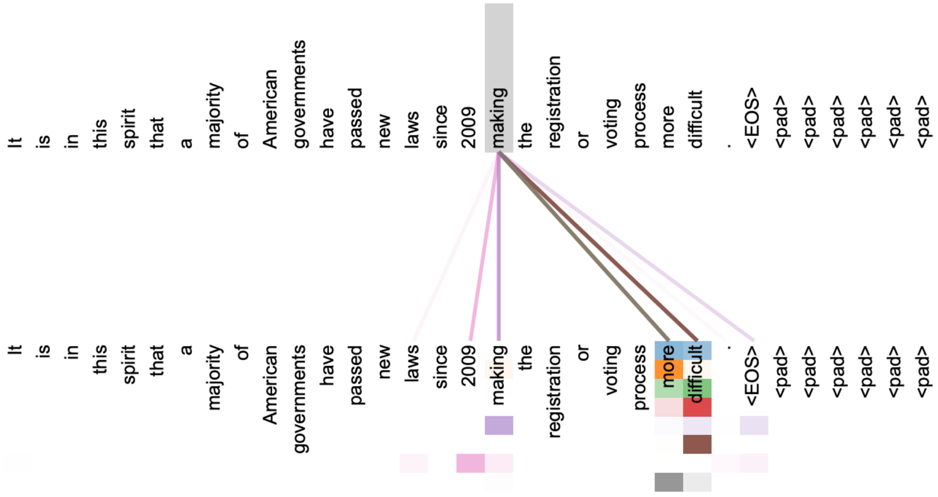


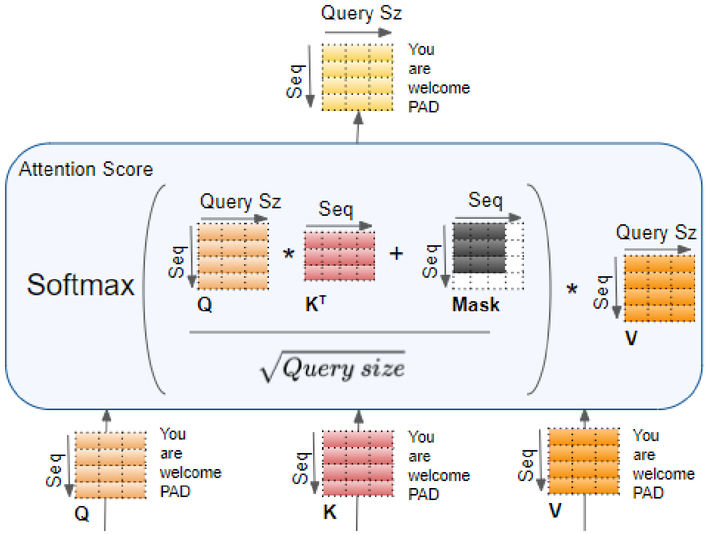
Brookhaven National Laboratory

Picking a backbone model

Transformers

- Attention Mechanism: every token in the sequence can "attend" any other tokens.
- Quadratic Complexity O(n^2): doubling the sequence length quadruples the computation.
- Memory: the KV cache grows as the sequence length.
- Need a large computing budget.

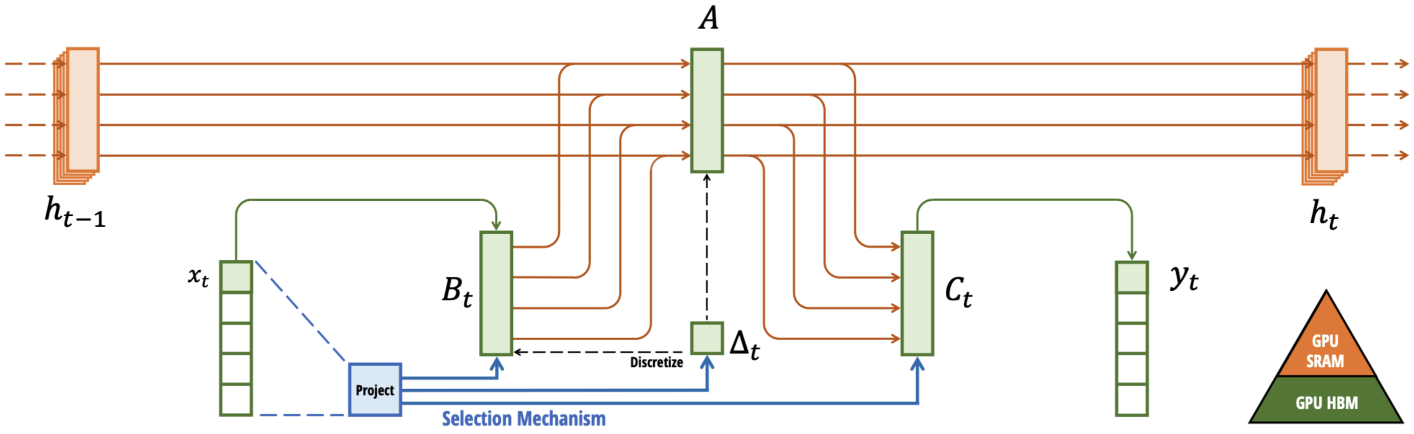


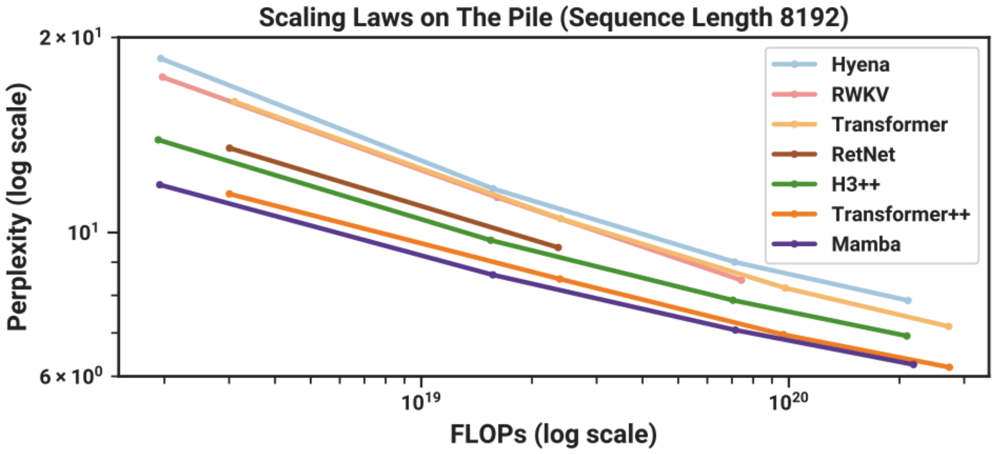


Brookhaven National Laboratory

State Space Model (SSM)

- **Structured State Space Models (SSMs)** improve computational efficiency while maintaining long-range sequence modeling capabilities.
- **Linear Complexity O(n):** computing cost grows linearly with the sequence length.
- **Constant Memory Size**: Fixed-length hidden states unlike growing KV cache in Transformer.
- **Efficient implementation:** Mamba achieves linear time and memory complexity, and hardware-optimized.
- Due to the limited computing budget, we adapted the **Mamba** model [1].

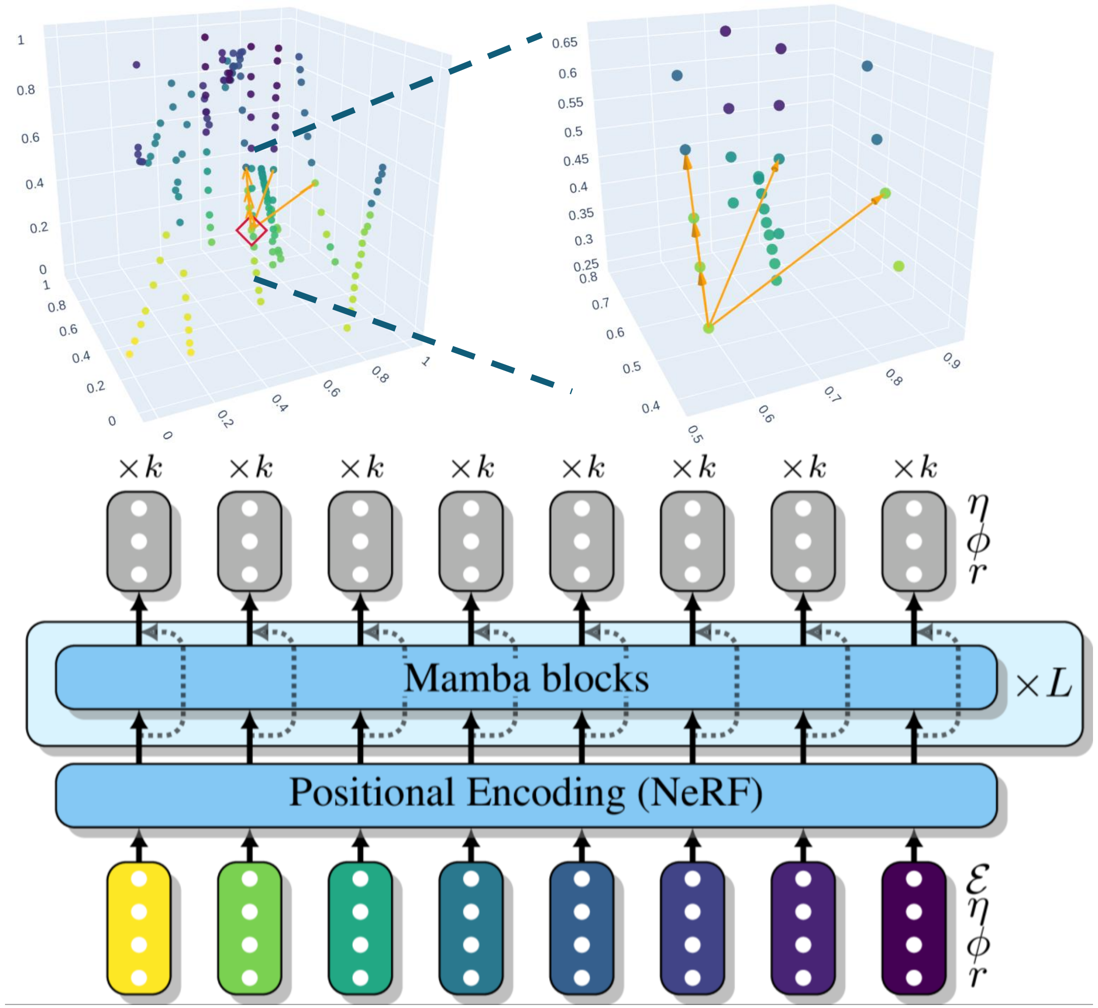




[1] Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. *arXiv:2312.00752*.

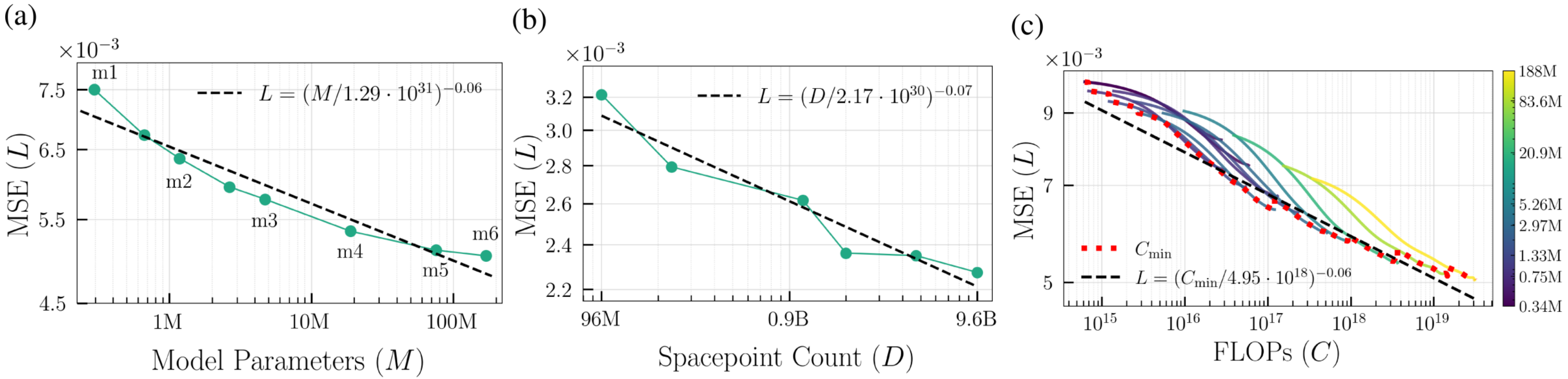
Next k-nearest-neighbor Prediction

- HRS Serialization is not following a track. This means ordinary autoregression or “next token” prediction may be too random.
- Predicting next k-nearest neighbors maybe better as more points may fall into the same track.

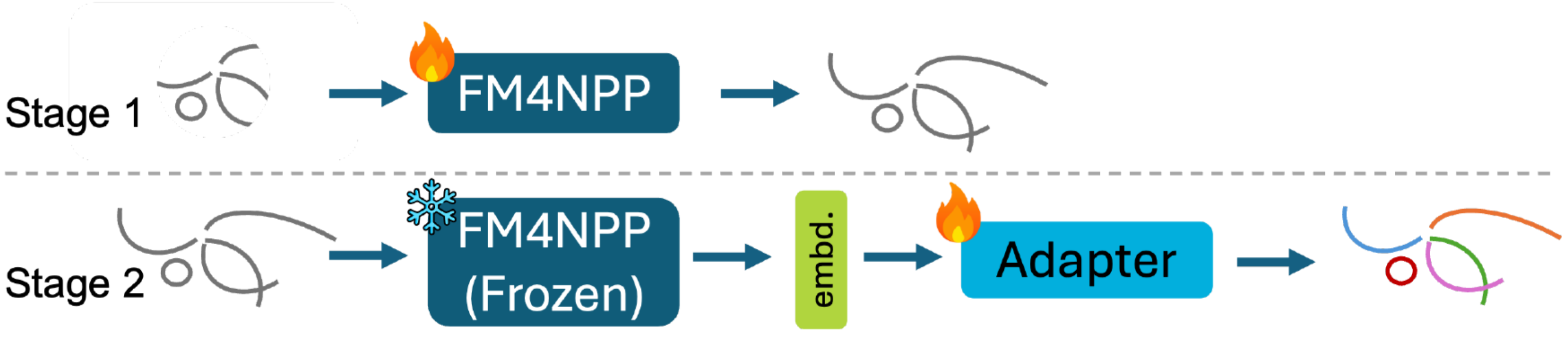


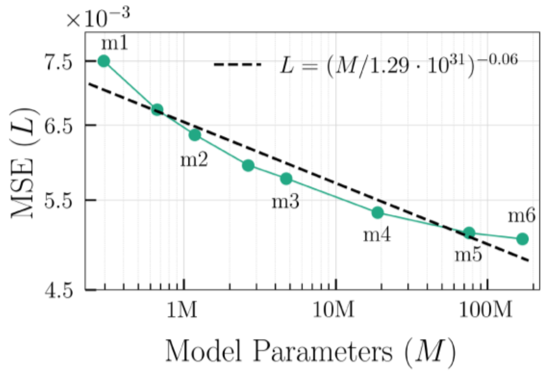
Neural Scaling Behavior

	Model Sizes					
	m1	m2	m3	m4	m5	m6
Model Width	64	128	256	512	1024	1536
Model Params	0.34M	1.3M	5.3M	21M	84M	188M



- Log-log scale of MSE loss versus number of Model Parameters, number of Spacepoints and Compute (FLOPs)

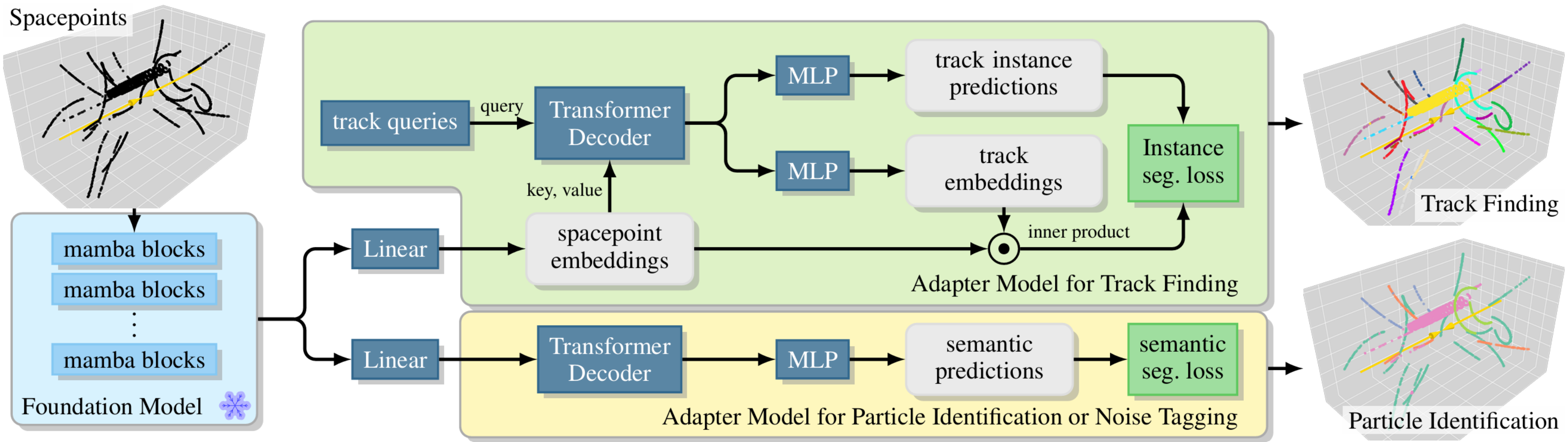




1. ✅ Identified a self-supervised learning task for NPP sparse spacepoints data.

2. ✅ Verified the neural scaling behavior of FM pre-training.

Adapting FM for Downstream Tasks



1. FM weights are frozen.
2. Lightweight Adaptor models are trainable on labeled data.
3. Evaluated on three downstream tasks: Track Finding, Particle Identification and Noise Tagging.

Conventional Particle Tracking Algorithm

- Particle Tracking is to associate points into tracks, Fig. 1.
- Cellular Automaton (CA) algorithm that finds “seeds”
  - Greedy algorithm that associates spacepoints via layer adjacent mutual edges.
  - Find all nearby two-edges among three layers. Keep the most straight ones and prune the rest.
- Kalman Filter extender propagates seed tracks to add possible spacepoints.
- CPU only, serial graph finding algorithm operating layer-by-layer.

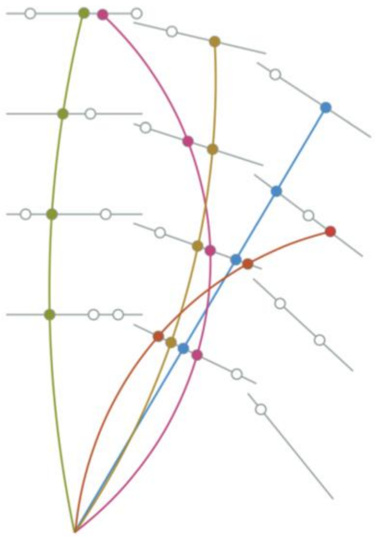


Fig. 1 A simplified view of “tracking” in 2D [1].

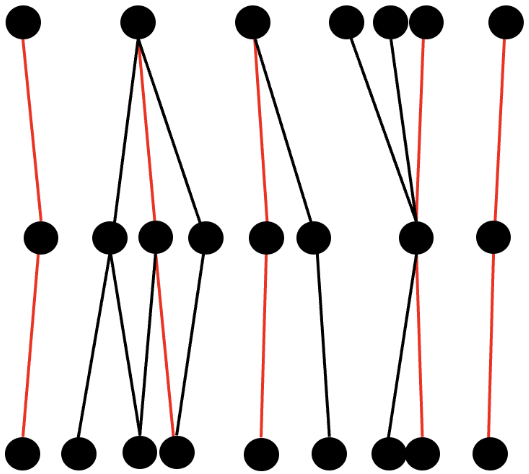


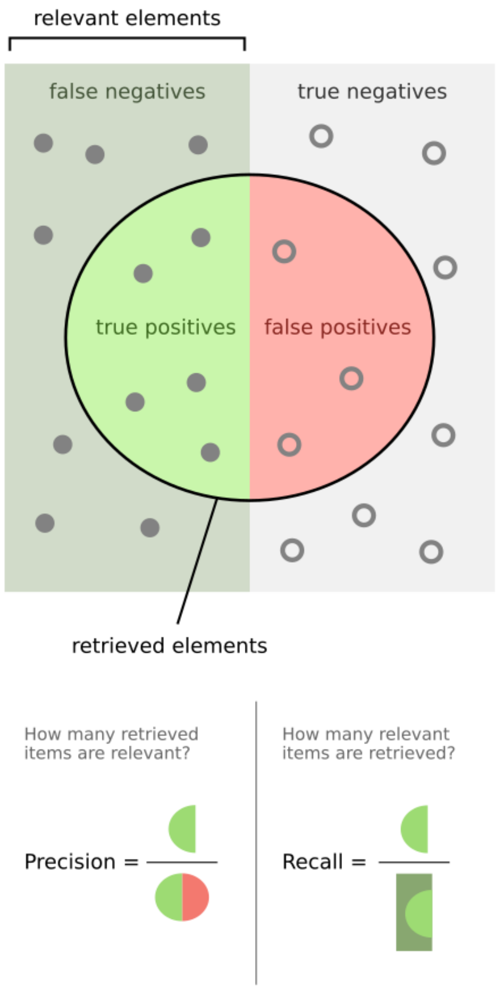
Fig. 2 CA method finds seed tracks [2].

[1] Amrouche, Sabrina, et al. "The tracking machine learning challenge: accuracy phase." *The NeurIPS'18 Competition: From Machine Learning to Intelligent Conversations*. Cham: Springer International Publishing, 2019. 231-264.

[2] Osborn, J.D., Frawley, A.D., Huang, J. *et al.* Implementation of ACTS into sPHENIX Track Reconstruction. *Comput Softw Big Sci* **5**, 23 (2021).

Evaluation Metrics for Tracking

- **Tracking efficiency (recall)**. The fraction of true tracks have been identified. This is the primary metrics in NPP.
- **Tracking purity (precision)**. The fraction of true positives with respect to all predicted tracks.
- **Adjusted Rand Index (ARI)**. Rand Index is a standard metrics for measuring the accuracy of “clustering”. Adjusted for chance of random assignment. ARI ∈[-1, 1].
- Tracks are uniquely matched to particles by the double majority rule, adapted from TrackML [1].



The Rand index, R, is:[1][2]

$$R = \frac{a+b}{a+b+c+d} = \frac{a+b}{\binom{n}{2}}$$

a: # of pairs in the same cluster for both partitions.
b: # of pairs in the different clusters for both partitions.

$$ARI := \frac{R - E[R]}{\max(R) - E[R]}$$

[1] Calafiura, P. "TrackML: A High Energy Physics Particle Tracking Challenge, in the proceedings of the 14th International Conference on e-Science, Amsterdam, Netherlands."

Main Results

1. Our FM4NPP approach outperforms all comparative models on Tracking Task.
2. EggNet (Calafiura, et al., 2024) and Exa.TrkX (Ju, et al. 2021) are graph neural networks (GNN)-based methods.
3. HEPT (Miao, et al. ICML 2024) is based on transformer.
4. We confirm the performance gain is from FM pre-training by comparing with the "AdapterOnly" model.

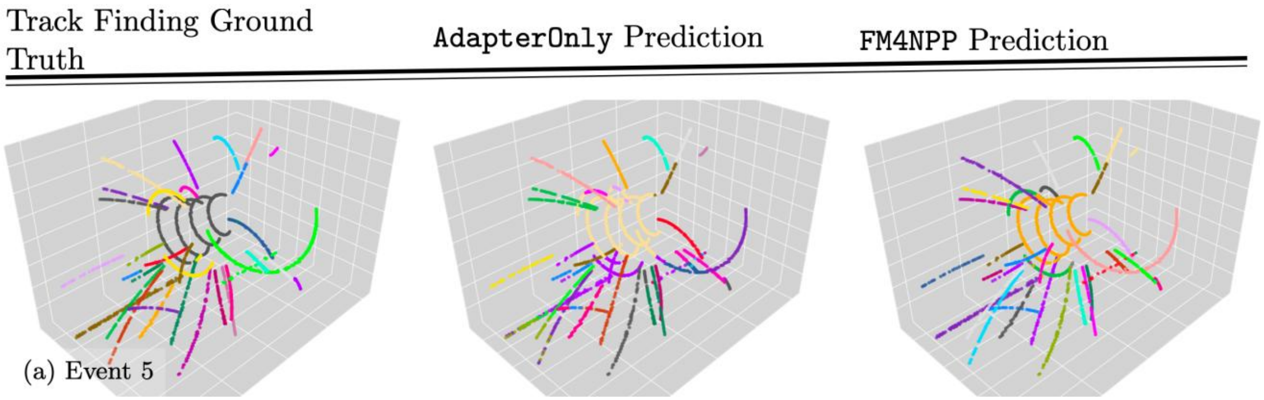


Figure 1. Example Output for Track Finding

model	#trnbl para.	Track Finding ARI↑	Track Finding efficiency↑	Track Finding purity↑
EggNet	0.16M	0.726	74.2%	75.1%
Exa.TrkX	3.86M	<u>0.877</u>	<u>91.8%</u>	66.4%
HEPT	0.31M	0.831	81.2%	<u>78.0%</u>
AdapterOnly	2.39M	0.724	78.0%	64.5%
FM4NPP (m6)	2.39M	**0.945(3)**	**96.1(2)%**	**93.1(1)%**

Brookhaven National Laboratory

Main Results

1. Our FM4NPP approach outperforms all comparative models on all three downstream tasks.
2. Tested against various of GNN models.
3. OneFormer3D (Kolodiazhnyi, et al. CVPR 2024) is the SOTA for point cloud data.

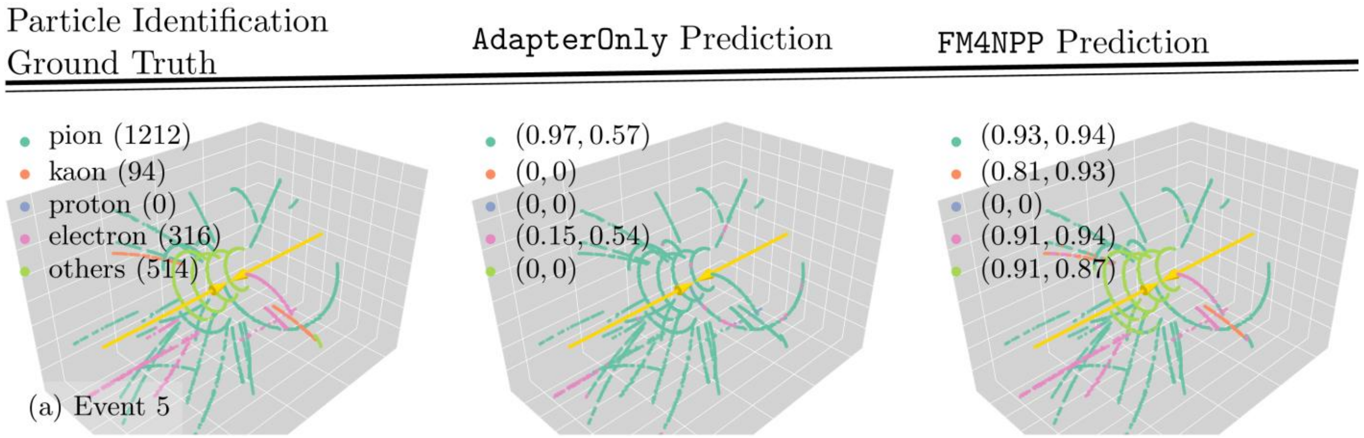


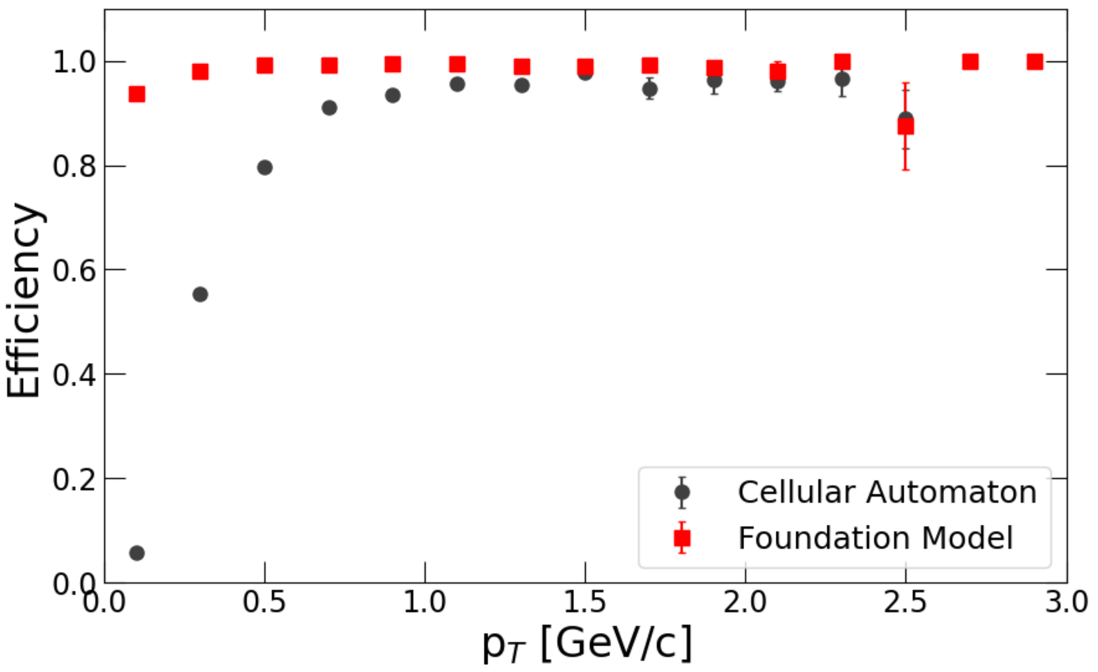
Figure 1. Example Output for Particle Identification

model	#trnbl para.	Particle Identification acc.↑	Particle Identification recall↑	Particle Identification pre.↑	Noise Tagging acc.↑	Noise Tagging recall↑	Noise Tagging pre.↑
`GATConv`	0.91M	0.692	0.397	0.637	0.910	0.673	0.806
`GCNConv`	0.91M	0.689	0.391	0.632	0.910	0.673	0.804
`SAGEConv`	0.91M	0.726	0.456	<u>0.650</u>	0.917	0.723	0.817
`GraphConv`	0.91M	0.7079	0.4176	0.6425	0.9190	0.7213	0.8252
`OneFormer3D`	44.95M	<u>0.770</u>	<u>0.490</u>	0.577	<u>0.965</u>	**0.940**	<u>0.895</u>
`AdapterOnly`	0.74M	0.663	0.339	0.611	0.911	0.622	0.836
`FM4NPP(m6)`	0.74M	**0.904(1)**	**0.765(3)**	**0.878(3)**	**0.971(1)**	<u>0.937(1)</u>	**0.919(1)**

Comparing with Cellular Automaton

High tracking efficiency across p_t.

- A canonical way of showing the goodness of particle tracking is to plot the tracking efficiency w.r.t. the moment p_t.
- Greater the p_t, the more straight the track, easier to find using conventional methods.
- Low p_t particles such as delta electrons, secondary interaction of radiation and detector materials, are extremely difficult to detect using conventional algorithms.
- In the $p_t > 1$ region, FM4NPP reaches a tracking efficiency of 99.6%, exceeding the sPHENIX pipeline's 94.6%.
- FM4NPP is ~ 30 times faster!

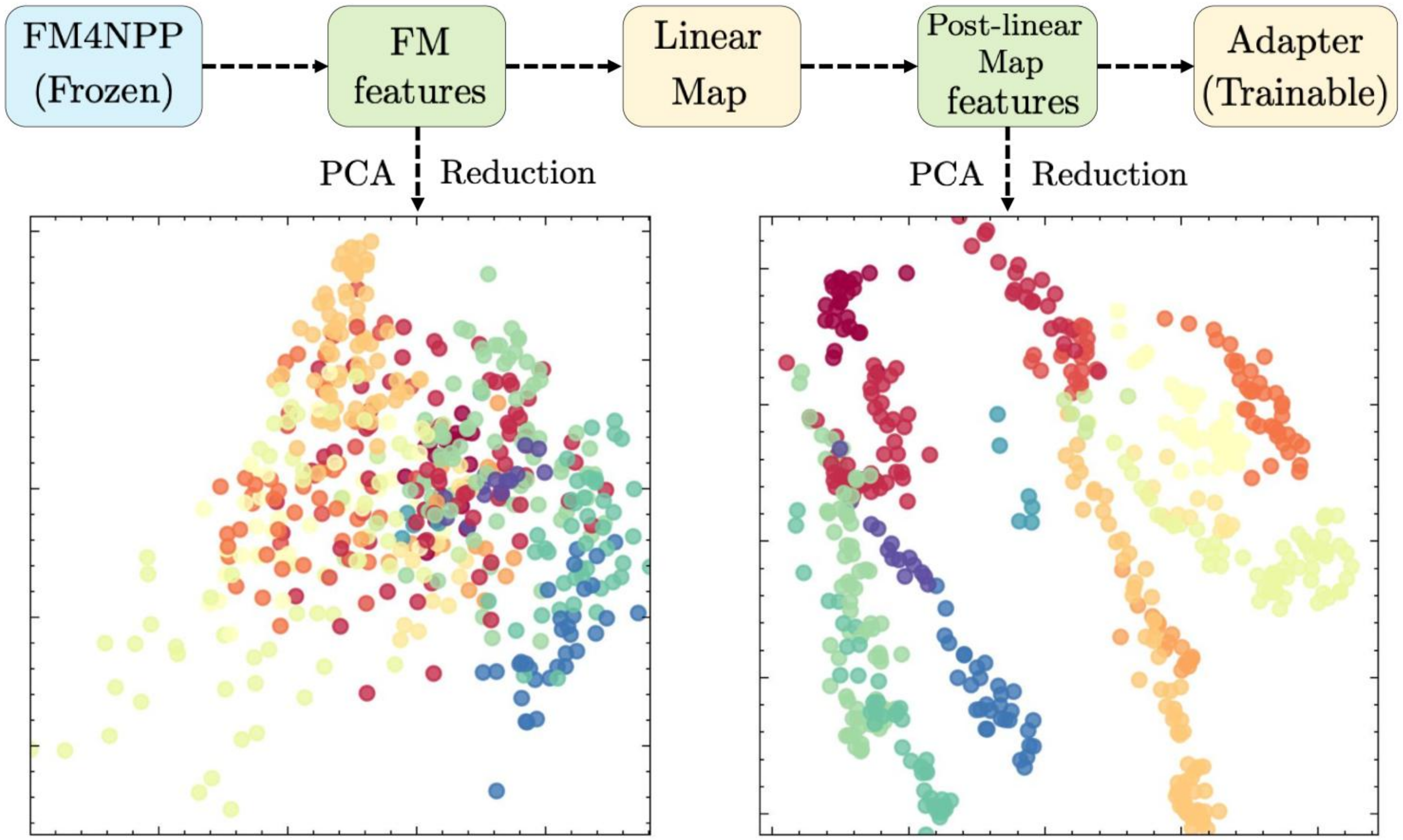


Typically, particle physicists focus on high-momentum tracks with filtering. Here, there is no filtering.

Brookhaven National Laboratory

Why does FM4NPP work?

1. FM backbone is a large "feature extractor" enriching spacepoint with "contextual" information.
2. "output features" from FM backbone first went through a linear map to match feature dimensions.
3. After finetuning, post-linear features exhibit clustering effects prior to transformer adapters, make the task much easier.

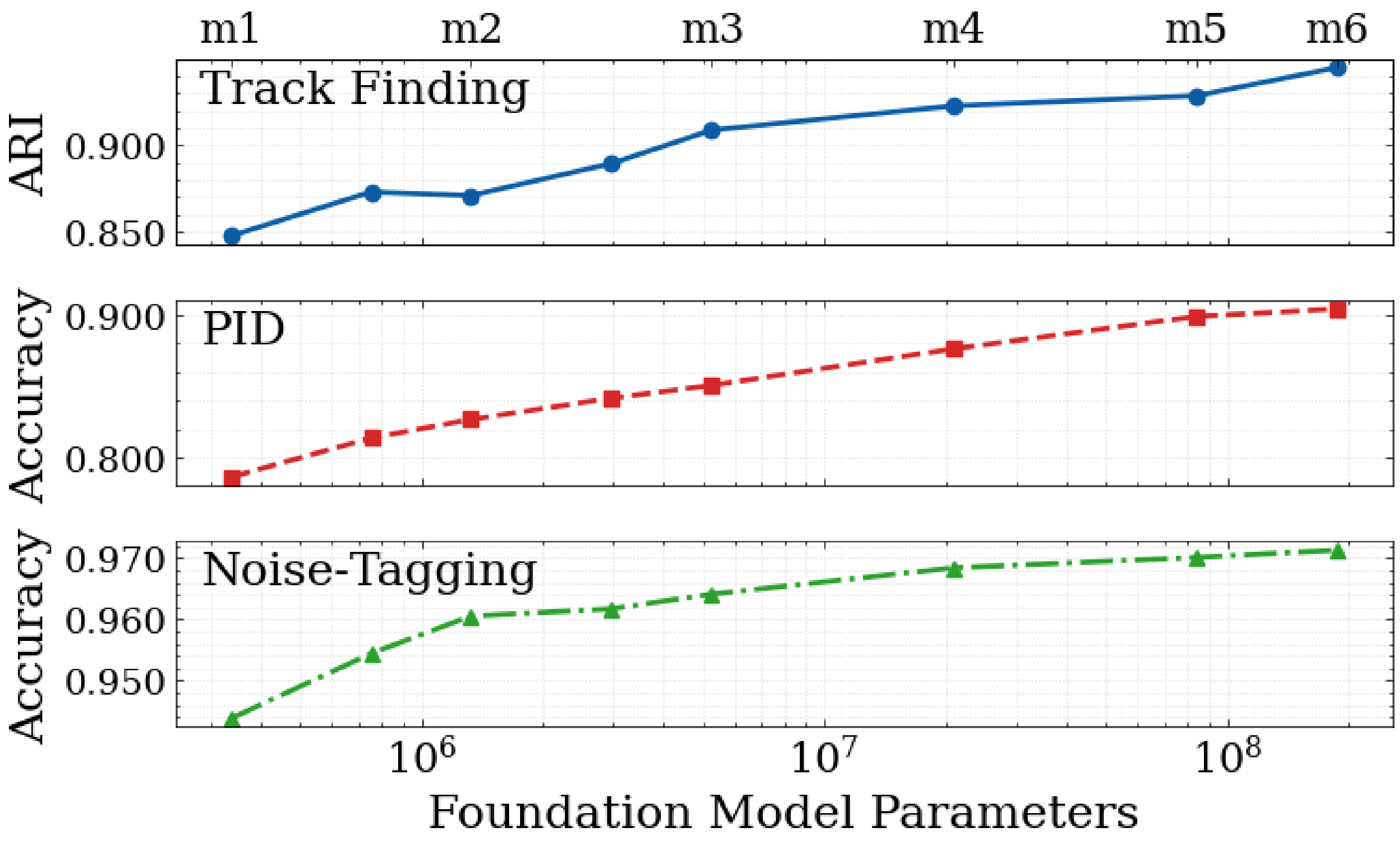


FM Features are Task-Agnostic
But, Task-relevancy is one linear map away!

Brookhaven National Laboratory

Scaling behavior translates to downstream tasks

- Across all three downstream tasks, we use the adapter model of the same size.
- The experiments show the performance of all downstream tasks improve as the FM getting bigger.



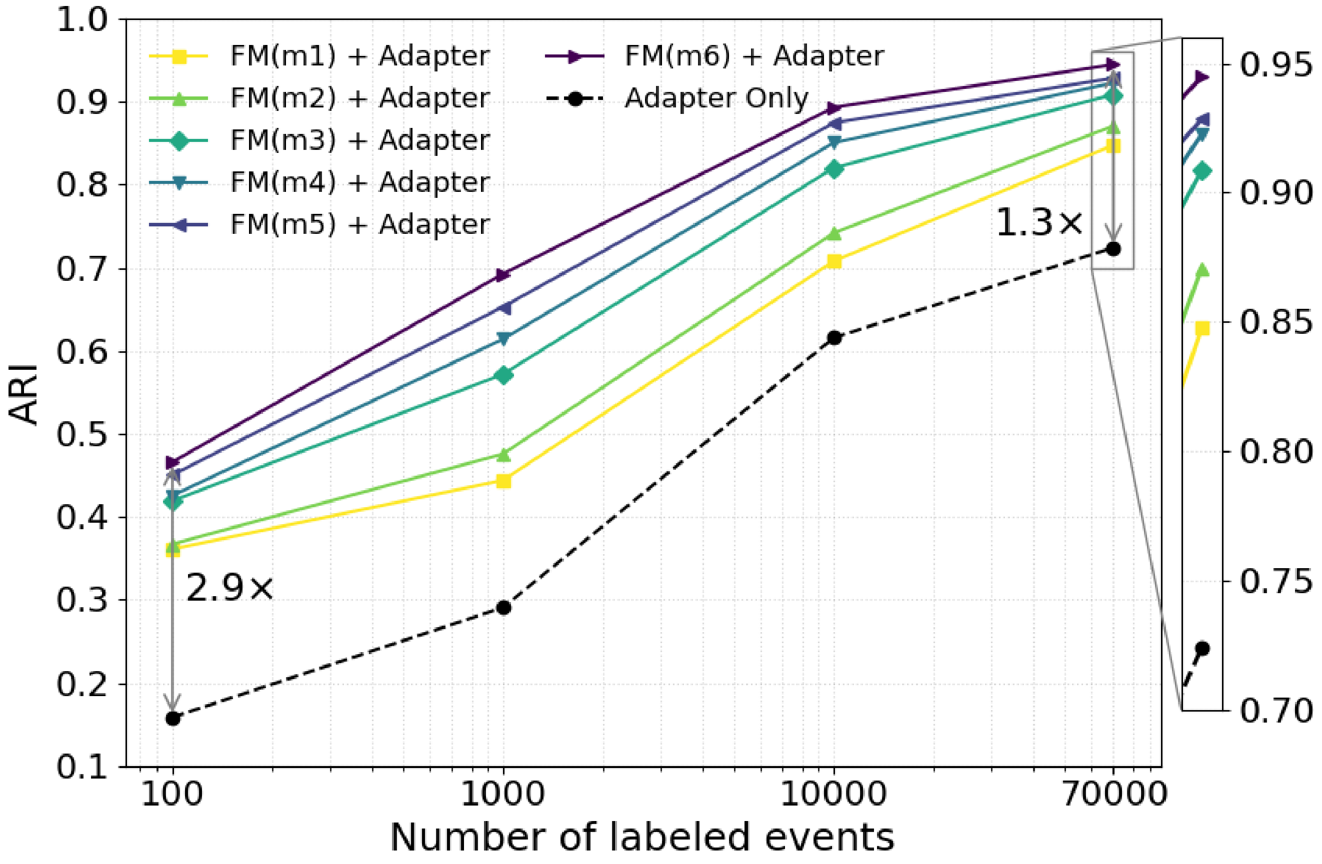
High Data Efficiency

Data Efficiency: model can achieve high accuracy when the labeled dataset is small.

Very often, labeled scientific data are hard to obtain.

We limited the training data points from very small (100) to abundant (70,000).

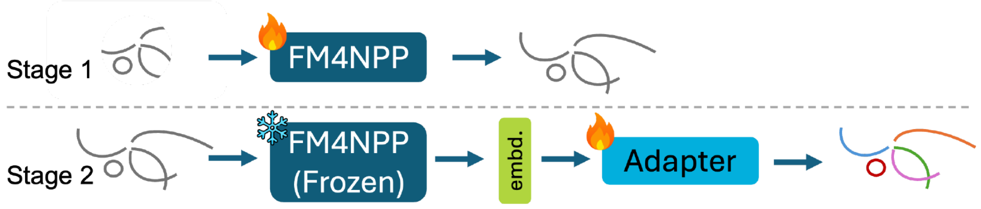
Foundation models of greater sizes have higher data efficiency: 2.9x better than without foundation model.

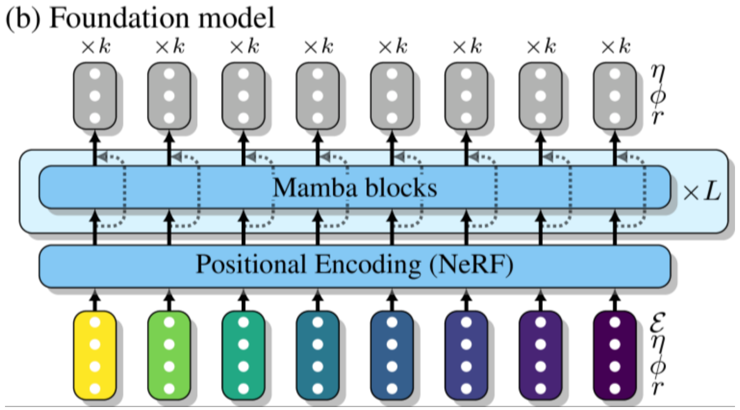


Brookhaven National Laboratory

Conclusions

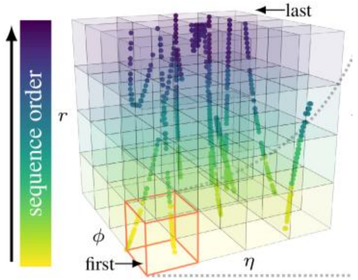
1. We have demonstrated a scalable and adaptable FM4NPP approach that can leverage big data and big computation.
2. The FM4NPP achieves a new state of the art on three downstream tasks.

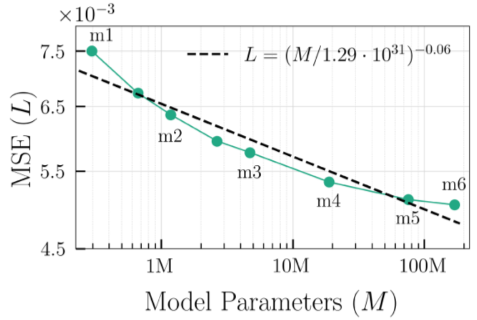




Key Achievements:

1. Identified a self-supervised learning task for NPP sparse spacepoints data.
2. Verified the neural scaling behavior of FM pre-training.
3. Demonstrated the effectiveness of FM approach on a diverse downstream tasks both in terms of performance and data efficiency.





Conclusions

1. We have demonstrated a scalable and adaptable FM4NPP approach that can leverage big data and big computation.
2. The FM4NPP achieves a new state of the art on three downstream tasks.

Key Achievements:

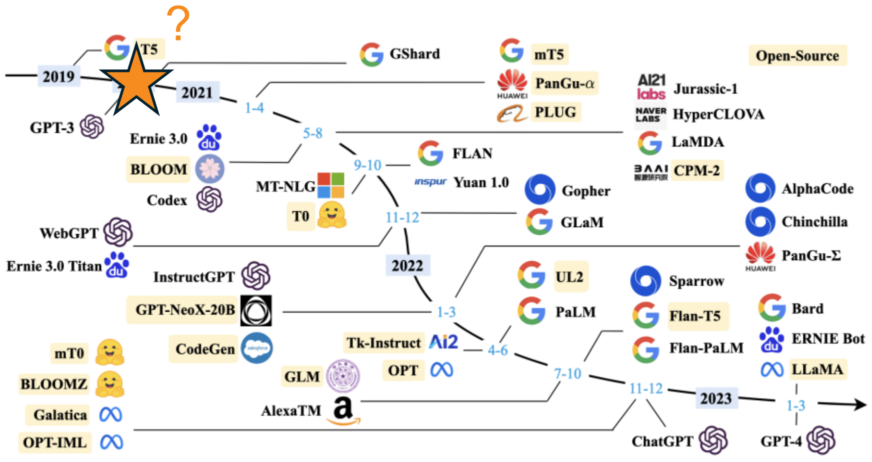
1. Identified a self-supervised learning task for NPP sparse spacepoints data.
2. Verified the neural scaling behavior of FM pre-training.
3. Demonstrated the effectiveness of FM approach on a diverse downstream tasks both in terms of performance and data efficiency.

ICLR 2026 Meta Review Comment:
While reviewers correctly note the scope is limited to sPHENIX TPC rather than all of nuclear and particle physics, the paper demonstrates, for the first time in this domain, that FM-style self-supervised pretraining with scaling laws can work for sparse detector data. By the Bommasani et al. definition (self-supervised pretraining, adaptation to multiple tasks), this qualifies as a foundation model for its target domain, analogous to the authors' "GPT-2 era" positioning.

[1] Park, David K. and Li, Shuhang and Huang, Yi and Luo, Xihaier and Yu, Haiwang and Go, Yeonju and Pinkenburg, Christopher and Lin, Yuewei and Yoo, Shinjae and Osborn, Joseph D. and Huang, Jin and Ren, Yihui, *FM4NPP: A Scaling Foundation Model for Nuclear and Particle Physics*. **ICLR2026**
Openreview: https://openreview.net/forum?id=qaI3cLFsiX

Brookhaven National Laboratory

Future Work



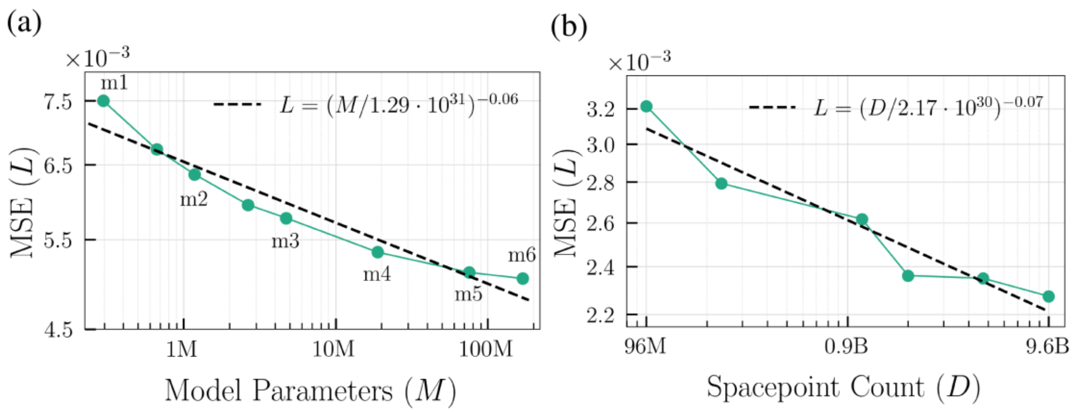
FM4NPP-1 is at a very early stage: small model size and only demonstrated on TPC.

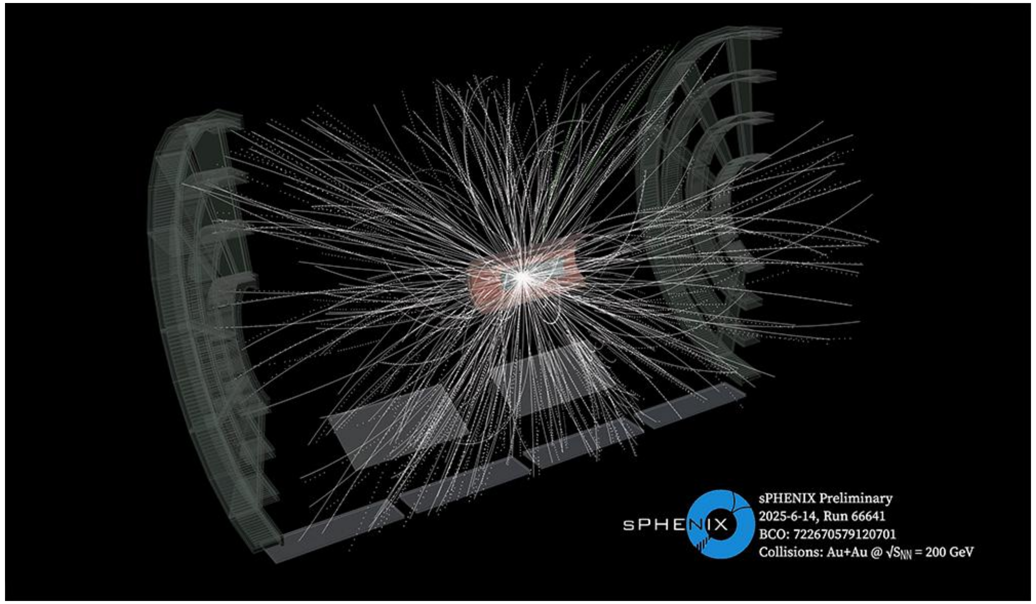
This is very similar to GPT-2 era that auto-regressive pre-training was demonstrated, but others are using BERT.

There are so much can be explored.

Future Work

1. Continue the scaling study to 1B or 10B model parameters, heavy-ion collisions, more data.
    a. heavy-ion collisions are more complex: p+p contains ~1000 spacepoints whereas Au+Au ~8000.
    b. Need more computation power.



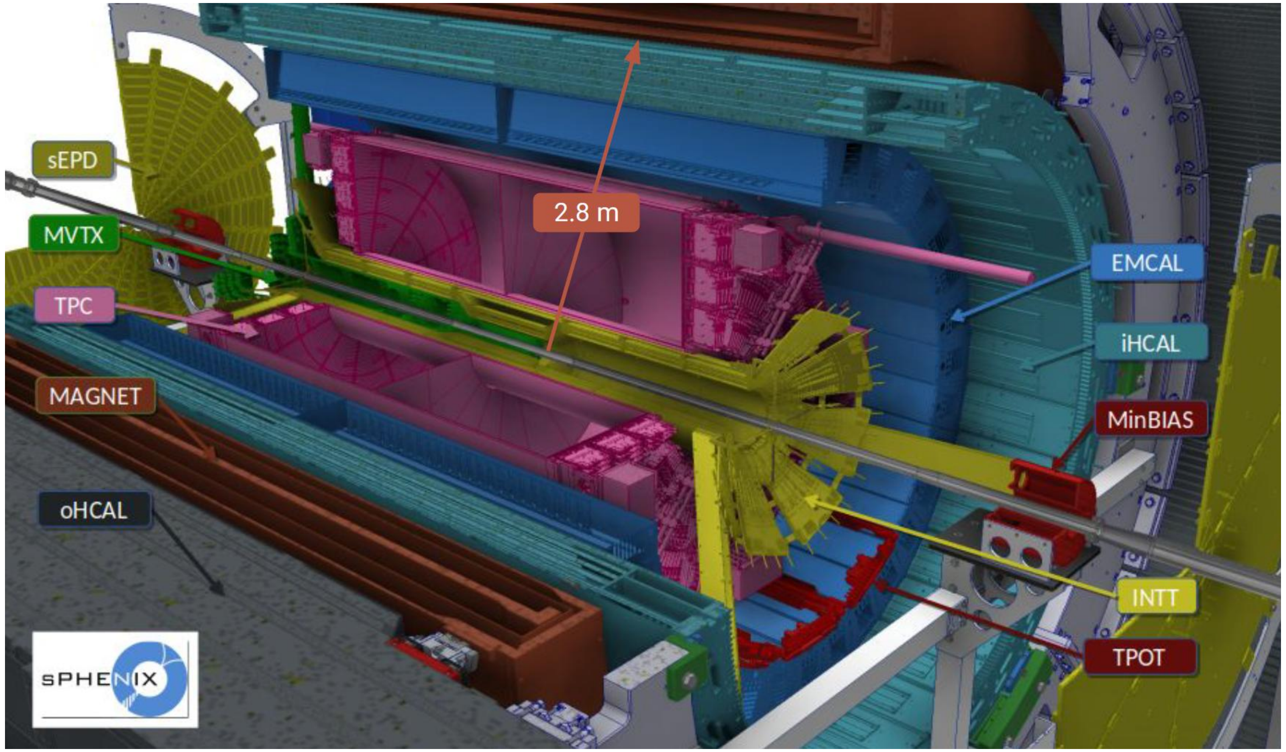


2026/02/02: [Sparse spacepoints in sPHENIX]

Brookhaven National Laboratory

Future Work

1. Continue the scaling study to 1B or 10B model parameters, heavy-ion collisions, more data.
2. Multi-detector data (silicon tracker, calorimeter, etc)
    a. This will unlock more downstream tasks: vertexing, jet finding, etc.

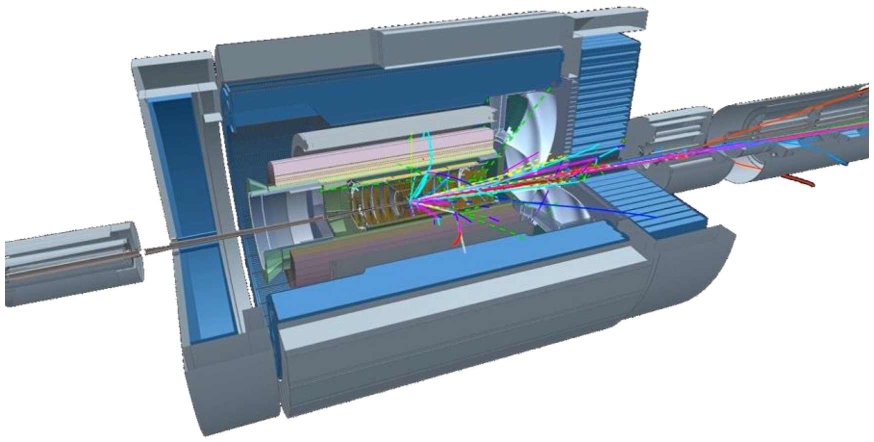


Future Work

1. Continue the scaling study to 1B or 10B model parameters, heavy-ion collisions, more data.
2. Multi-detector data (silicon tracker, calorimeter, etc)
3. Exploring more general tokenization, neural architectures, and self-supervised learning tasks.
   a. Other detectors can also benefit from this work.

ATLAS

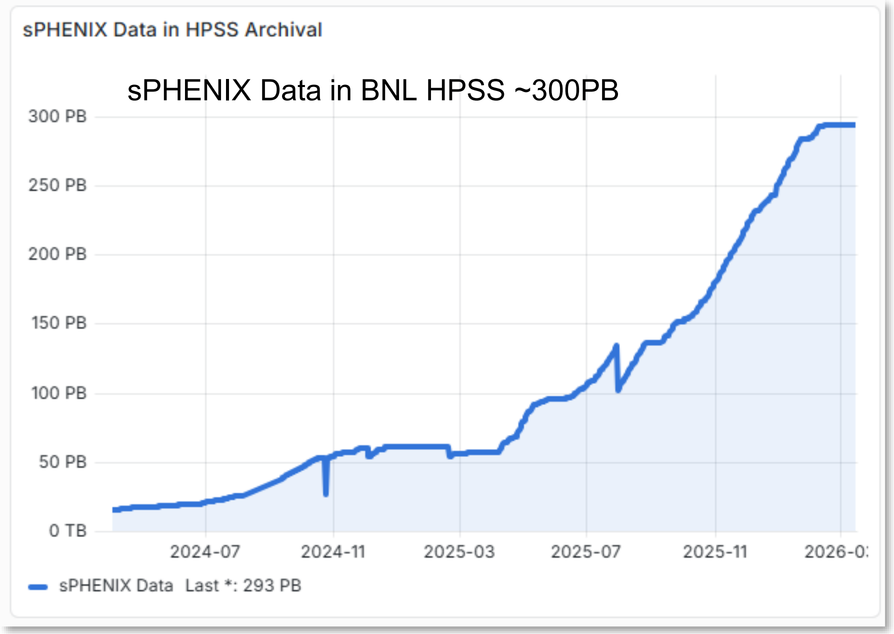
ePIC (EIC's 1st detector)



Brookhaven National Laboratory

Future Work

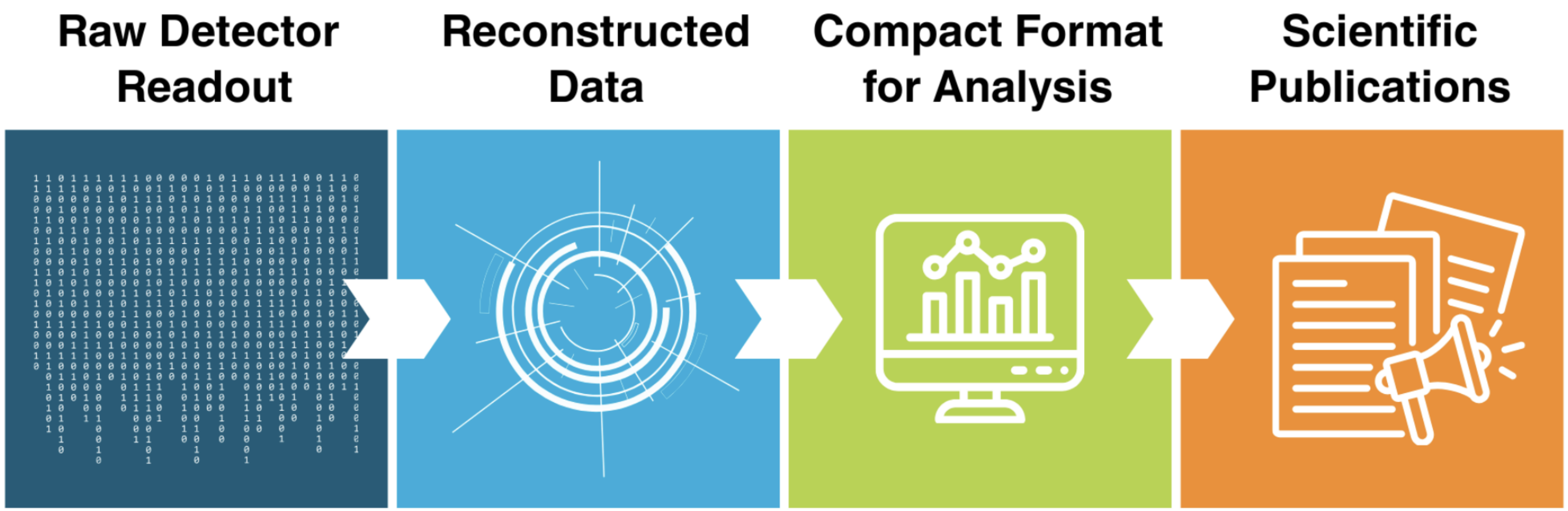
1. Continue the scaling study to 1B or 10B model parameters, heavy-ion collisions, more data.
2. Multi-detector data (silicon tracker, calorimeter, etc)
3. Exploring more general tokenization, neural architectures, and self-supervised learning tasks.
4. Pre-train FM4NPP on authentic experimental data to validate the efficacy and impact of our methodology in real-world physics environments.



Brookhaven National Laboratory

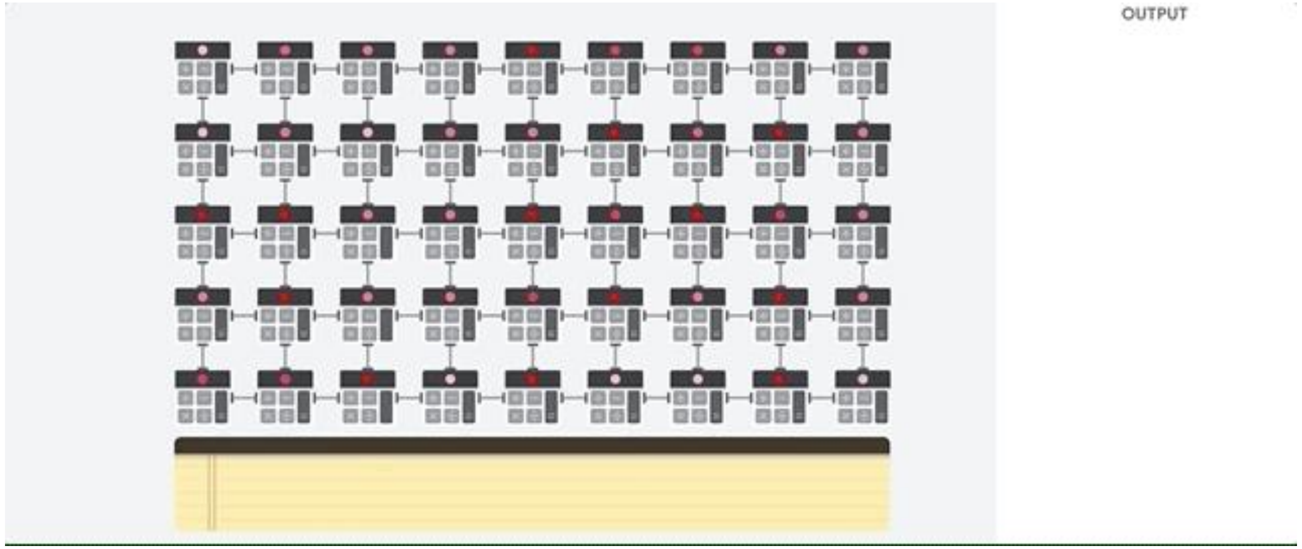
Future Work

1. Continue the scaling study to 1B or 10B model … collisions, …
2. Multi-detec…
3. Exploring … tokenizatio… and self-su…
4. Apply FM4…
5. Connecting with higher-level analysis and reasoning models, enabling end-to-end hypothesis verification.



Future Work

1. Continue the scaling study.
2. Multi-detector data
3. Exploring more general tokenization, neural architectures, and self-supervised learning tasks.
4. Apply FM4NPP to real-world data.
5. End-to-end hypothesis verification.
6. Improve training and inference speed through quantization, distillation, and hardware-aware codesign.

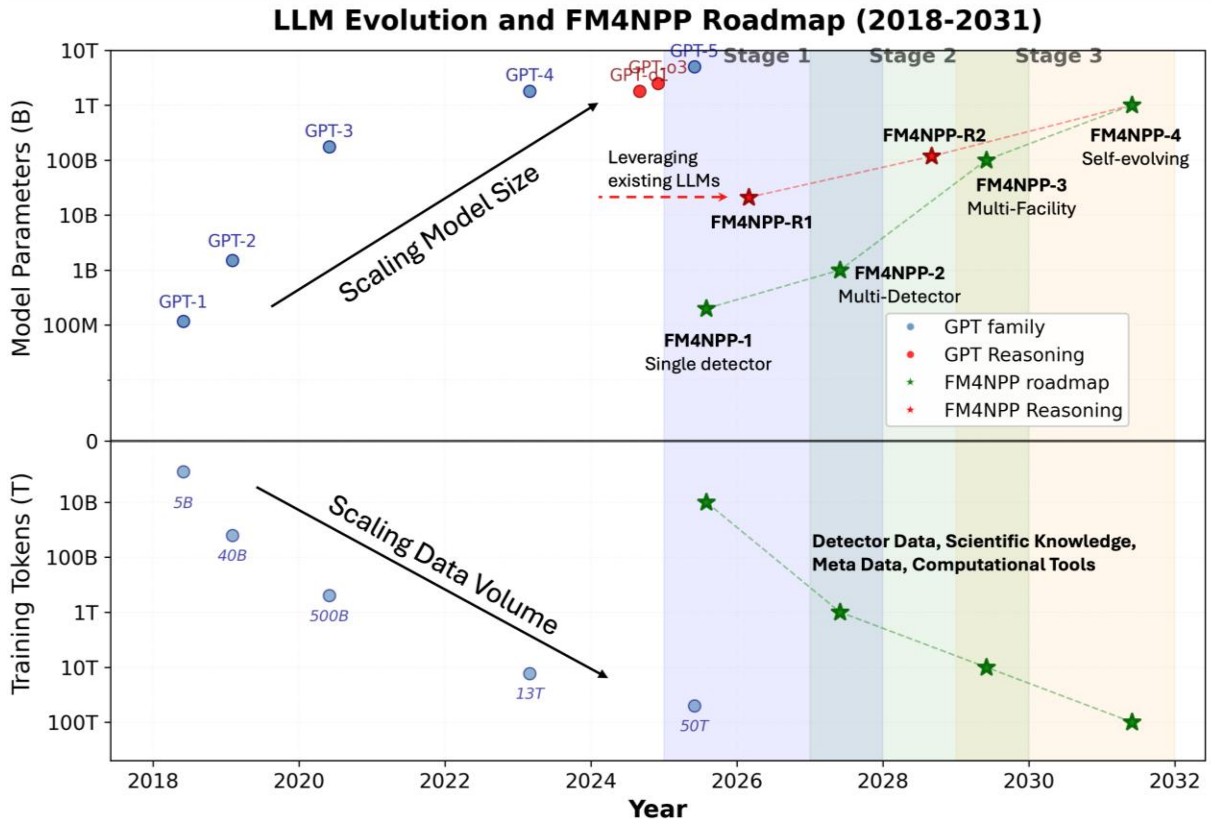


Google's TPU

Brookhaven National Laboratory

Future Work

1. Continue the scaling study larger model and more complex data.
2. Multi-detector data
3. Exploring more general tokenization, neural architectures, and self-supervised learning tasks.
4. Apply FM4NPP to real-world data.
5. End-to-end hypothesis verification.
6. Improve training and inference speed through quantization, distillation, and hardware-aware codesign.



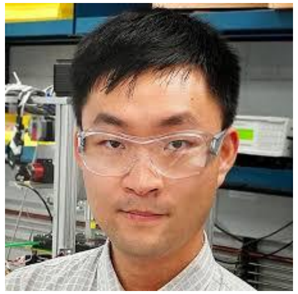
Roadmap paper: http://dx.doi.org/10.2139/ssrn.5467666

Acknowledgement

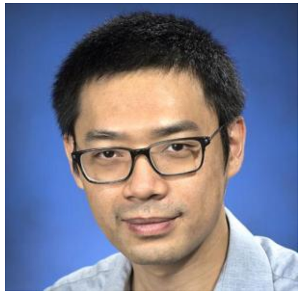
- Yihui Ren, Xihaier Luo and Shinjae Yoo were partially supported by the DOE Office of Science through the Office of Advanced Scientific Computing Research and the Scientific Discovery through Advanced Computing (SciDAC) program.
- Shuhang Li was partially supported by the DOE Office of Science through the Office of Nuclear Physics under Award No.~DE-FG02-86ER40281.

David Park

Shuhang Li



Jin Huang



Yihui Ren

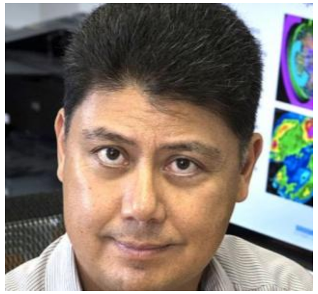
Shinjae Yoo

Acknowledgement

- Yihui Ren, Xihaier Luo and Shinjae Yoo were partially supported by the DOE Office of Science through the Office of Advanced Scientific Computing Research and the Scientific Discovery through Advanced Computing (SciDAC) program.
- Shuhang Li was partially supported by the DOE Office of Science through the Office of Nuclear Physics under Award No.~DE-FG02-86ER40281.
- This work was supported by the Laboratory Directed Research and Development (LDRD) Program at Brookhaven National Laboratory, LDRD 25-045, which is operated and managed for the U.S. Department of Energy (DOE) Office of Science by Brookhaven Science Associates under contract No. DE-SC0012704.

John created LDRD-D to support us.
Nick prioritized CDS computing powers for us.

John Hill (LD)



Nick D'Imperio (ALD)

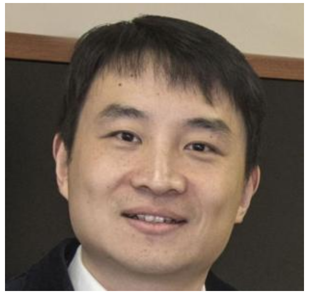
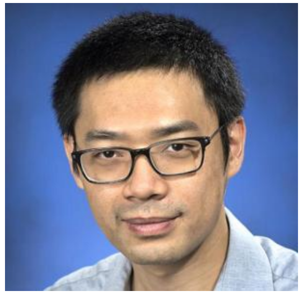
Acknowledgement


- Yihui Ren, Xihaier Luo and Shinjae Yoo were partially supported by the DOE Office of Science through the Office of Advanced Scientific Computing Research and the Scientific Discovery through Advanced Computing (SciDAC) program.
- Shuhang Li was partially supported by the DOE Office of Science through the Office of Nuclear Physics under Award No.~DE-FG02-86ER40281.
- This work was supported by the Laboratory Directed Research and Development (LDRD) Program at Brookhaven National Laboratory, LDRD 25-045, which is operated and managed for the U.S. Department of Energy (DOE) Office of Science by Brookhaven Science Associates under contract No. DE-SC0012704.
- This research also utilized resources of the National Energy Research Scientific Computing Center (NERSC) under the ``GenAI@NERSC'' program. NERSC is a DOE Office of Science User Facility with Award No. DDR-ERCAP0034059. The authors are grateful to the NERSC staff for their support, particularly **Shashank Subramanian** and **Wahid Bhimji**.

We used ~10,000 GPU node hours on NERSC ☺

Thank you!

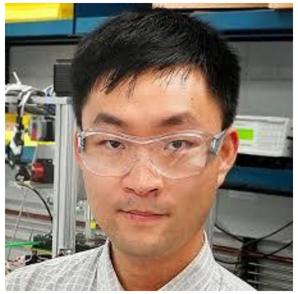


(AI Dept.) David Park, Yi Huang, Xihaier Luo, Yuewei Lin, Yihui (Ray) Ren





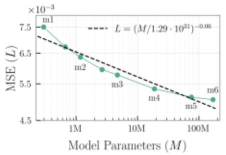
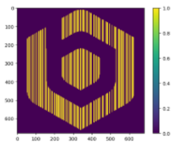
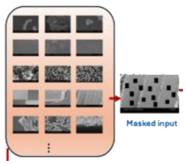


(Phys Dept.) Shuhang Li, Haiwang Yu, Joe Osborn, Yeonju Go, Jin Huang

AI Research Portfolio

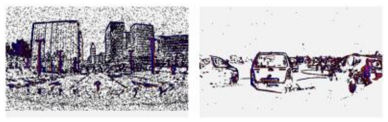
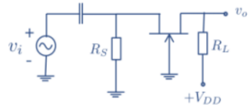
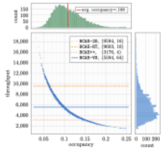
Foundation Models

- ***Novel scientific foundation models (FMs)***. Nuclear and particle physics (FM4NPP), Scanning Electron Microscopy (SEM) images, oceanography, robotics (VLAs). Multi-stage training, knowledge distillation, sparse spatiotemporal data encoding
- ***Multi-agent systems***. Advanced mask design, biological image segmentation tasks. Inner/outer loop evaluations, efficient tool use

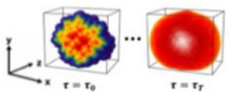
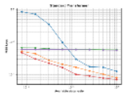
AI Co-Design

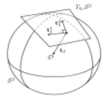
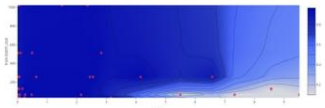
- ***Real-time scientific data requirements.*** BCAE-VS for real-time neural compression of sPHENIX detector streams
- ***AI for novel sensing modalities.*** EvRT-DETR for safeguards object detection with event-based camaras
- ***Chip design and benchmarking***. AutoSizer AI agent for transistor sizing, CircuitSense for analog circuit design benchmarking

AI Theory and Security

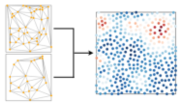
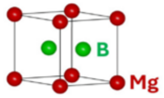
- ***Theory and experimental validation***. NN training as optimization on manifold, Fourier Neural Operators for physics surrogate models
- ***AI security applications and impact***. First FM scaling experiments for multi-application gamma spectroscopy (nuclear security-critical)
- ***Governance & infrastructure***. Platforms for AI, data and tools in inter-lab catalogs (open & CUI)

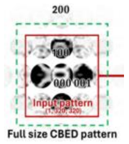



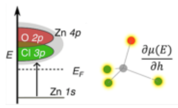



Trustworthy AI

- ***Interpretable scientific surrogates***. FLUID high-fidelity fields from low-fidelity unstructured simulations, GNN prediction of XAS spectra and attribute features to solvation motives.
- ***Uncertainty-aware inverse solvers***. Joint/conditional diffusion inverse solvers (JCDI) for beam diffraction, dynamic power-system load calibration.

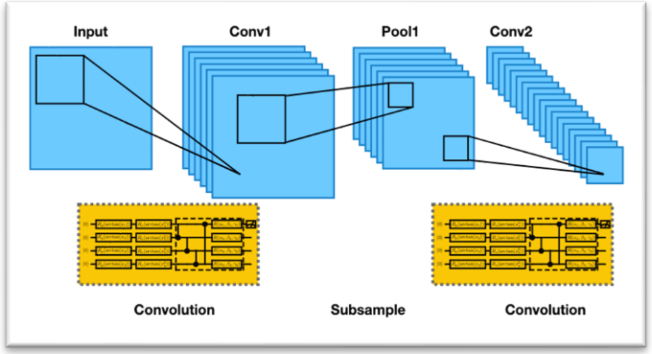





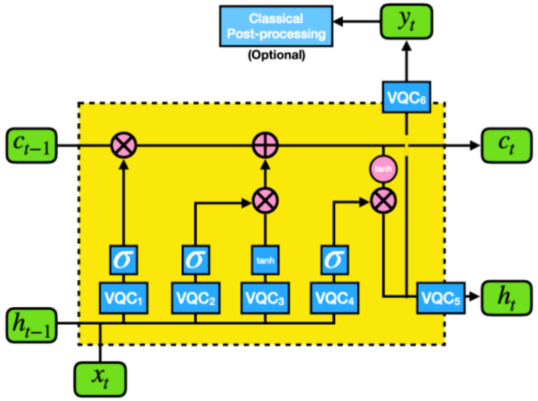



Quantum for AI Overview

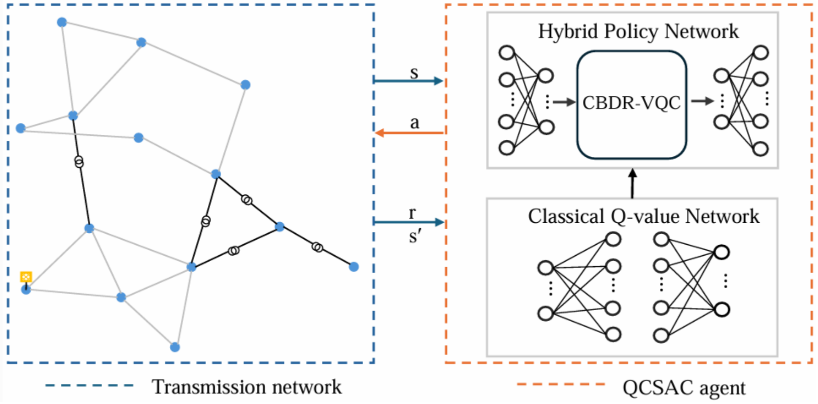
Quantum CNN



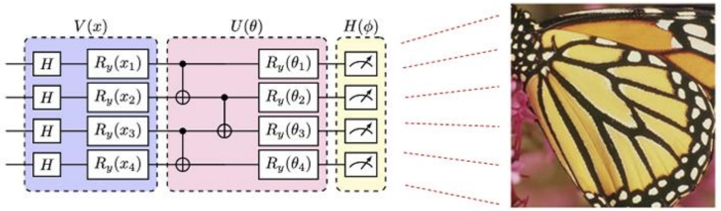
Quantum LSTM



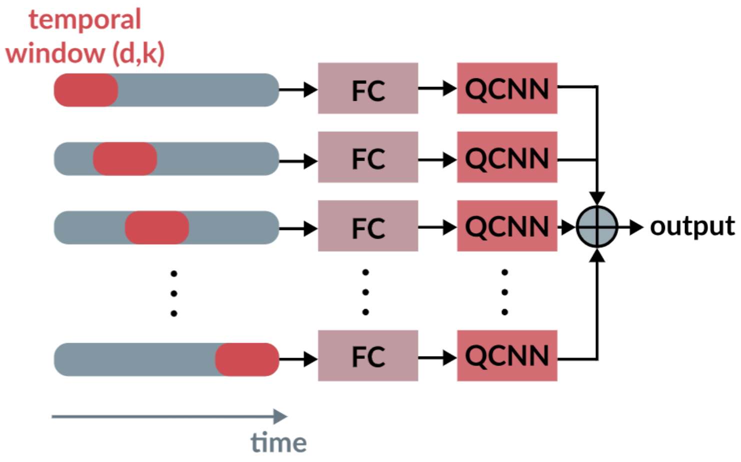
Quantum RL



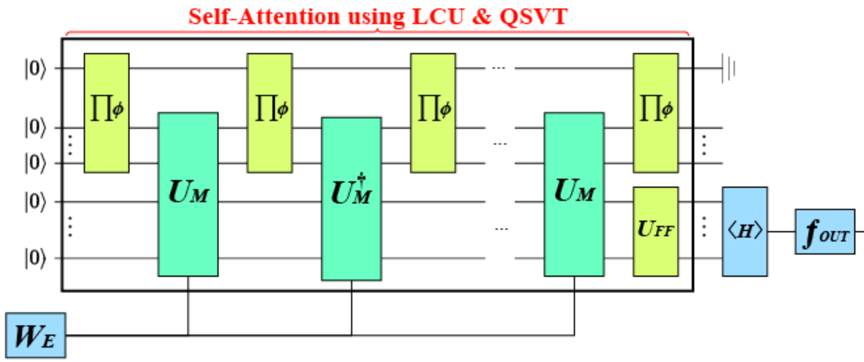
Quantum Super-resolution



Quantum Temporal Convolution



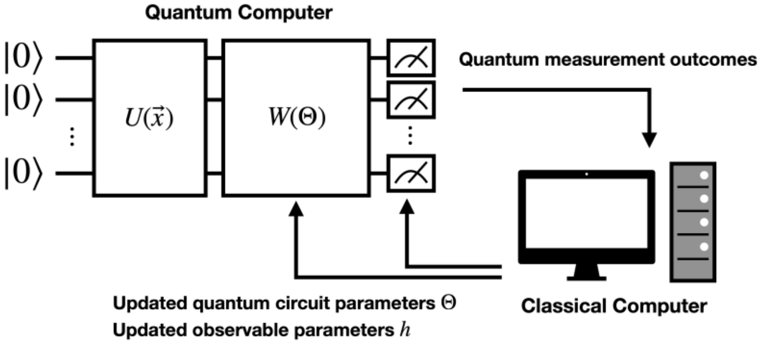
Quantum Transformers



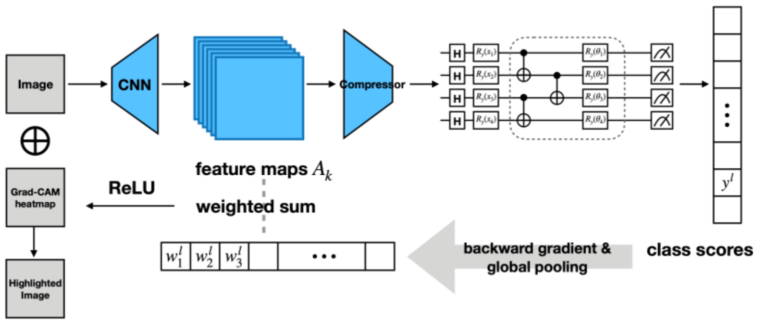
Brookhaven National Laboratory

AI for Quantum Overview

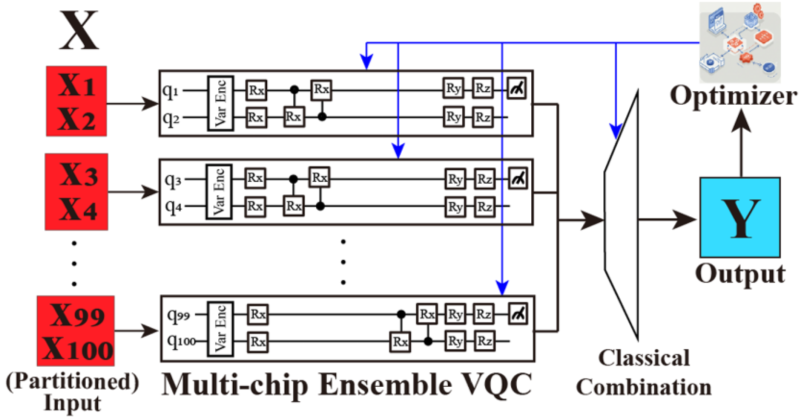
Learning to Measure



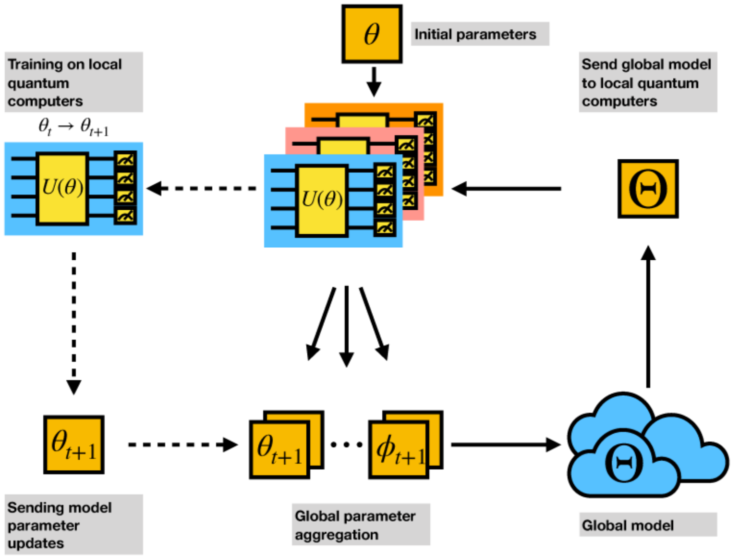
Quantum xAI (QGrad-Cam)



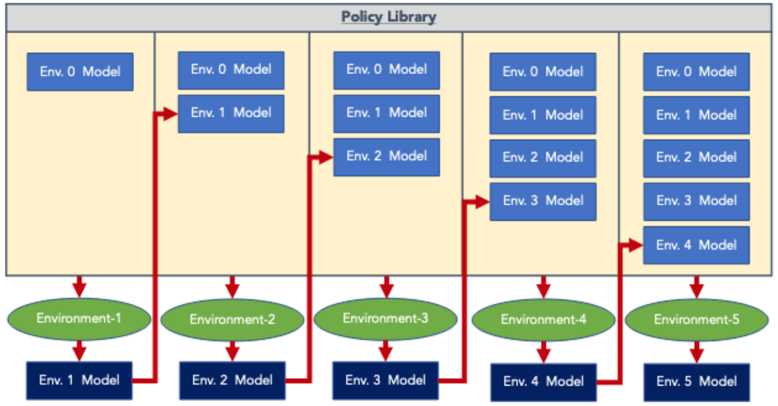
Quantum Multichip Learning



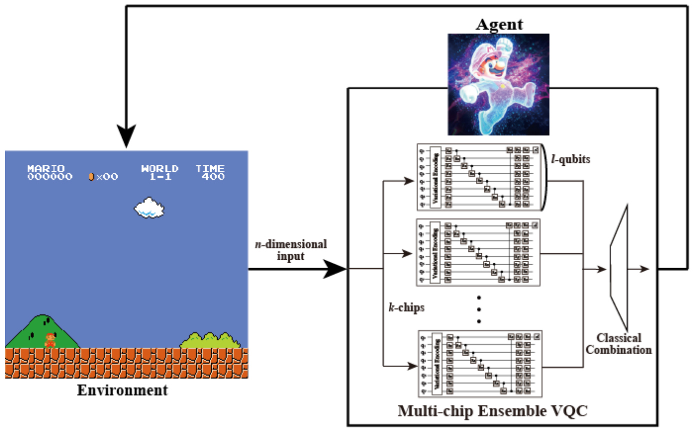
Quantum Federated Learning



Quantum Architecture Search with Continual Learning



Quantum Multichip RL



NeuroX-Fusion: Multimodal Foundation Model for Neuroscience

PI(s)/Facility Lead(s): David Park, Collaborating Institutions: BNL, Princeton, SNU

Brookhaven National Laboratory

Scientific Achievement

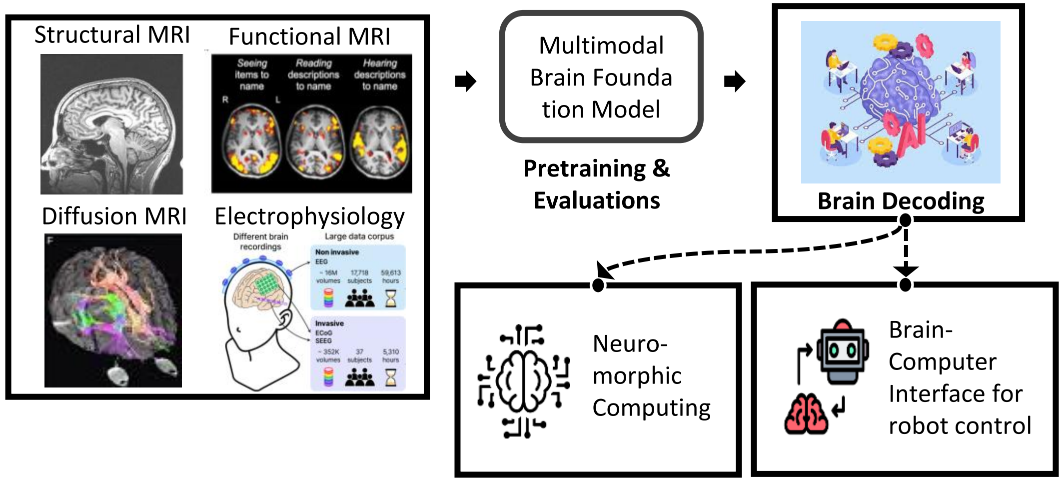
- Builds the first 130B-parameter computable "digital twin" of the brain, unifying spatial MRI with millisecond-scale electrophysiology.
- Delivers two foundation sub-models — NeuroX-MRI (91B) and NeuroX-Ephys (50B) — then fuses them into the unified model.
- Uses an LLM as a shared semantic "Rosetta Stone" to align modalities, bypassing scarce co-registered MRI–Ephys data.

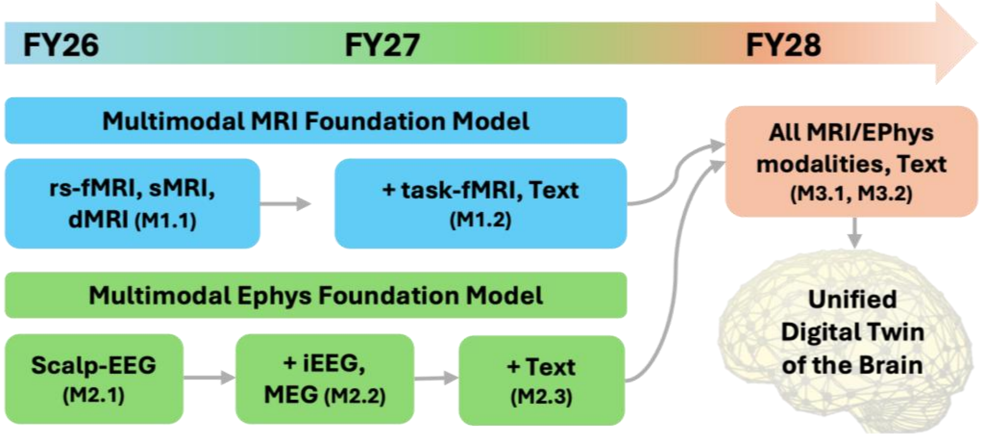
Significance and Impact

- Enables in-silico experiments impossible in humans, shifting neuroscience from static correlation toward dynamic, testable prediction.
- Yields blueprints for neuromorphic computing and a universal, calibration-free neural decoder for BCIs and robot control.
- Multi-institution effort (BNL, Princeton, SNU) with open models and a companion DOE ALCC compute proposal.

Technical Approach

- Self-supervised contrastive learning plus LLM-adapter alignment; a layout-agnostic backbone generalizes across electrode configurations.
- Overcomes Ephys heterogeneity and paired-data scarcity via set-based equivariance and the LLM semantic bridge.
- Trains on Aurora, Frontier, Polaris, Perlmutter using DeepSpeed ZeRO, BF16/FP16, and µTransfer tuning.





Publications

OmniField (ICLR 2026), Mamba-GINR (NeurIPS 2025), SOLID (arXiv 2026). Diffusion Transformer (NeurIPS 2025). DIVER-0 (ICML 2025, Oral), DIVER-1 (arXiv 2025), NeuroMamba (NeurIPS 2025), DANCE (arXiv 2026), INCAMA (arXiv 2026).