CERN openlab

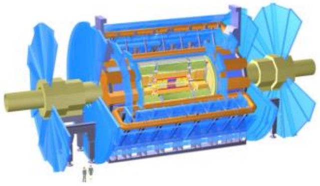
Generalization and self-supervision for physics simulation

Renato Cardoso, Sofia Vallecorsa

Work realized in collaboration with Oracle

Introduction

- Motivation:
  - Utilization of machine learning algorithms at CERN is **increasing** together with the **complexity** of the algorithms
  - Necessity to develop efficient methods for **performance tracking** and **cost optimization**
  - Ensure model **generalizability** and **reusability**.
- Current focus:
  - Start with a Generative model
    - Generative Models are amongst the most important architectures studied at CERN,
    - Need to replace Monte Carlo based physics simulations
  - How much can we extrapolate to other geometries and tasks (example: measuring particle properties)
    - Understand how **foundation model concepts**, such as self-supervised learning and transfer learning, apply to our use case.
    - Understand the **minimal scale of the model** for reaching meaningful results (No need to reach BERT / GPT-4 scale)
    - High number of resources needed for training



What is a foundation model in our study

What are foundation models:

- A model **trained on broad data and adaptable to a range of different downstream tasks**, zero-shot, few-shot learning.
- Self/semi-supervised learning + transfer learning but at scale
  - Large and diverse datasets → powerful representations
- Examples: BERT, GPT, DALL-E
- Stanford CRFM (2021) : On the Opportunities and Risks of Foundation Models [arxiv.2108.07258]

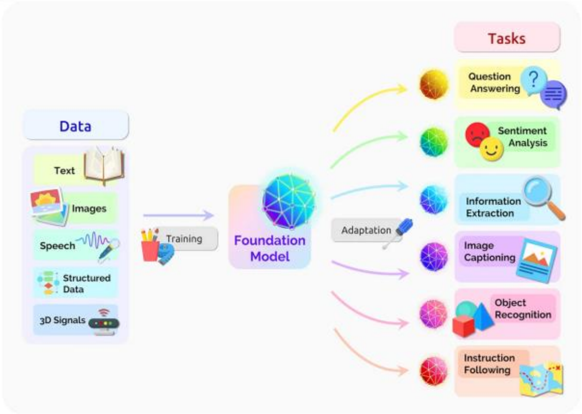


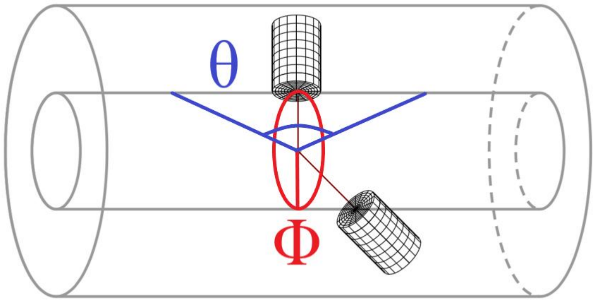
Image obtained from:
On the Opportunities and Risks of Foundation Models

Motivation:

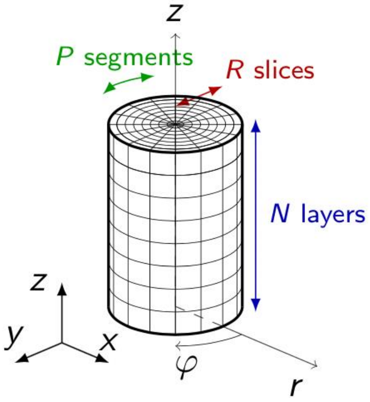
- Foundation model trained on MC data to perform different physics related tasks
  - Simulations - one lengthy training, then fast adaptation to different detector geometries
  - **Reconstruction** - one base model adaptable to different tasks (particle identification, regression on phys. variables, etc.)
- Understand how foundation model concepts (self-supervised learning and transfer learning) apply to our use case:
  - Understand the **minimal scale of the model for reaching meaningful results**

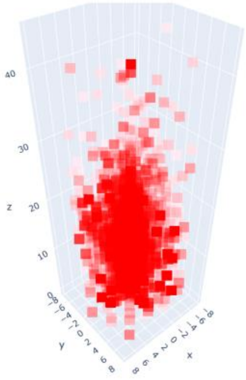
Single particles in calorimeters

- The datasets are simulated using the GEANT4 toolkit (G4)
  - The energy deposits in the calorimeter cells result from the interaction of an incoming primary particle with the calorimeter material. These deposits form a characteristic shape that we call “**shower**”.
  - Simplistic **cylindrical** geometry
  - The mesh aligns with the direction of incident particle
  - The direction, i.e., the angles are recorded.

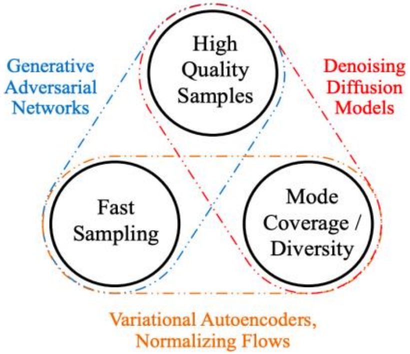


- Dataset Characteristics:
  - **1M samples**
  - 45x16x9 – **6480** voxels
  - Incident energies ranging from **1 GeV to 1 TeV**
  - Φ - **0 to 2π**
  - θ - **0.87 to 2.27**



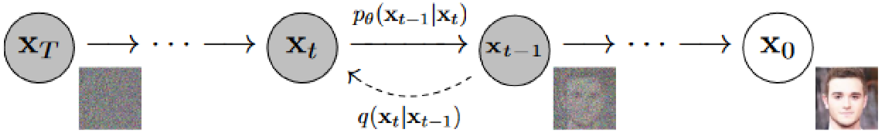


Diffusion + transformers



Start with a Generative model:

- Current state of the art are Diffusion models
  - Lead to higher accuracy and higher diversity
  - Markov chain that gradually adds noise to the until signal is destroyed
- Denoising Diffusion Probabilistic Models [https://arxiv.org/pdf/2006.11239.pdf]



Add transformer blocks:

- A **generalized architecture** that works with any type of data, e.g., text, images, audio, etc.
- **Models long-range dependencies** (Attention mechanism).
- Attention is all you need [https://arxiv.org/abs/1706.03762]

Our model

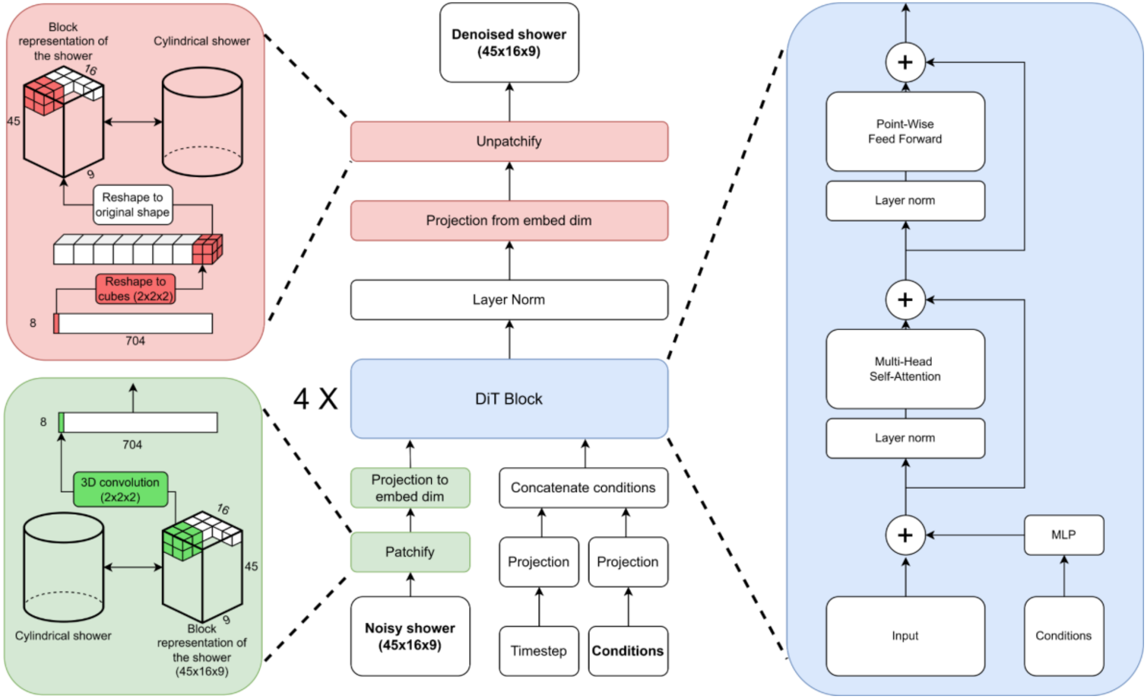
Conditions:

- Energy
- Φ
- θ
- Geometry

The model:

- Diffusion steps = 400
- Embed dim = 144
- Cosine scheduler

https://arxiv.org/abs/2212.09748

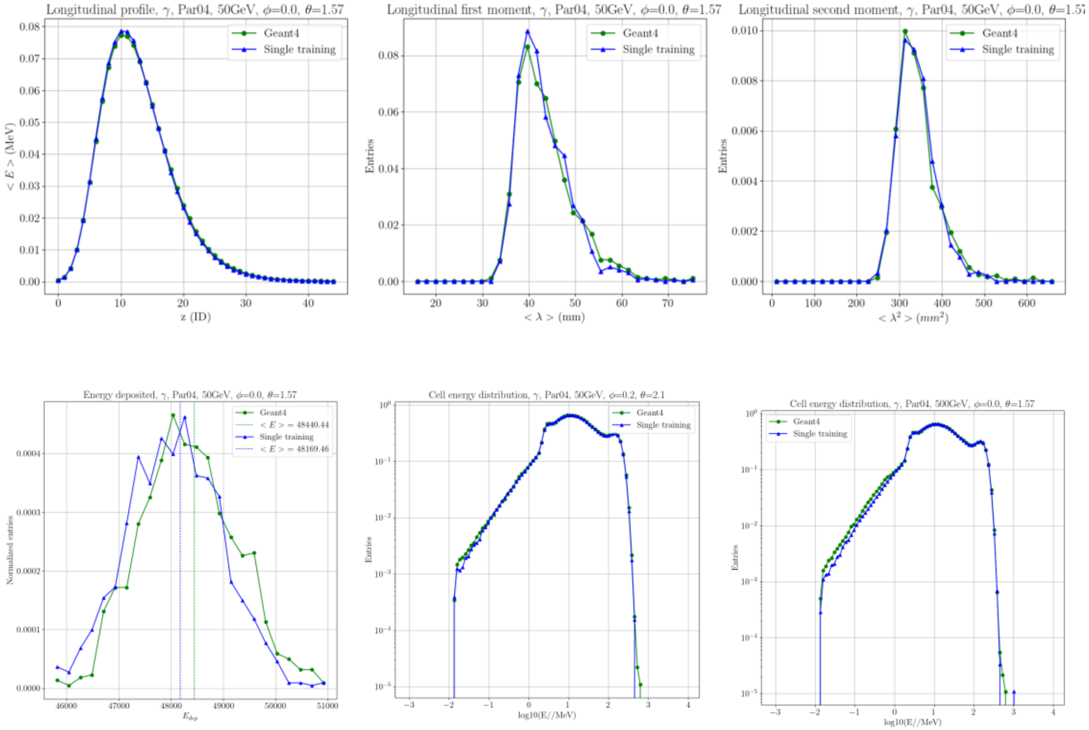


Generation Results

Good accuracy throughout all the profiles

Cell energy shows particular good results compared to other generative models

Trained for 300 epochs on a single V100 GPU. (around 15 minutes per epoch)



So what else can it do?

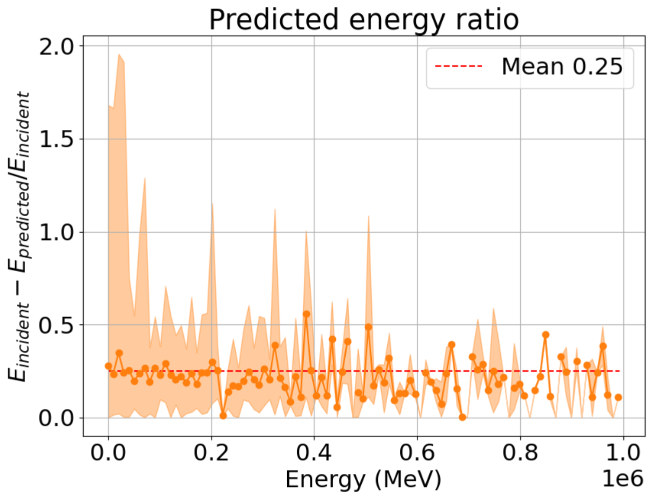
PRELIMINARY

Reconstruction - one base model adaptable to different tasks

- **Energy regression**
- Around **75% accuracy**
- Better than other generative models
  - Diffusion Models Beat GANs on Image Classification https://arxiv.org/pdf/2307.08702.pdf

Energy regression training

- **Few shot training** (around 20000 samples)
- 1 Linear layer for classification (+ Pooling for dimensional reduction)
- Use internal representation (after 1st DiT block)
- Fast adaptation



Conclusion and Future work

Conclusion:

- Development and implementation of a generalizable model for physics simulation
  - Combination of a diffusion model and transformers
  - Diffusion model used for **sampling quality**
  - Transformer architecture use as a **generalized architecture**
- Shown the ability of the model to do downstream task
  - Usage of the model for **energy regression**

Future work:

- Understand what features of the model are used for downstream tasks
- Leverage this model to better model downstream tasks
- Increase the quality of the model proposed