HELMHOLTZAI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

In-situ ML-based surrogate model training via continual learning and streaming

Varun Makdani
Hoffmann Group
Helmholtz-Zentrum Dresden-Rossendorf

Laser Plasma Acceleration

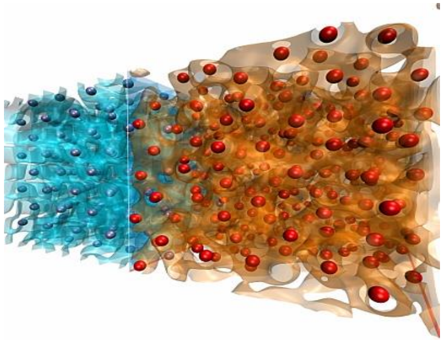
- Building small laser facilities is of interdisciplinary interest
  - medicine
  - aerospace
  - material science

- **For practical usage full understanding of physical processes is necessary**

- Digital research environment is important
  - Planning experiments
  - Small time scales
  - High resolution

Source: HZDR.de

Modern cancer therapy



Source: Jan Vorberger

Matter under extreme condition

Plasma Simulations

System description

$$\nabla \cdot \mathbf{D} = \rho$$
$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$
$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$

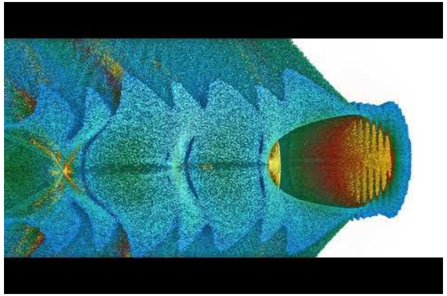
System parameter

Simulation

- High-resolution simulations model complex physical systems
- Simulation runs on large multi-node clusters
- Inferring effect of simulation parameters is expensive
- PIConGPU (Particle-in-Cell) is a fully relativistic, multi-core, simulation code for e.g. Laser plasma acceleration

<u>Aim:</u>

- **Train surrogate model concurrently to a running PIC simulation**
- **Approximation of numerical simulation by Reduced Order Models (ROM).**
  [Talk covered by **Anna Willmann** at **1:00 PM**]



High fidelity simulation

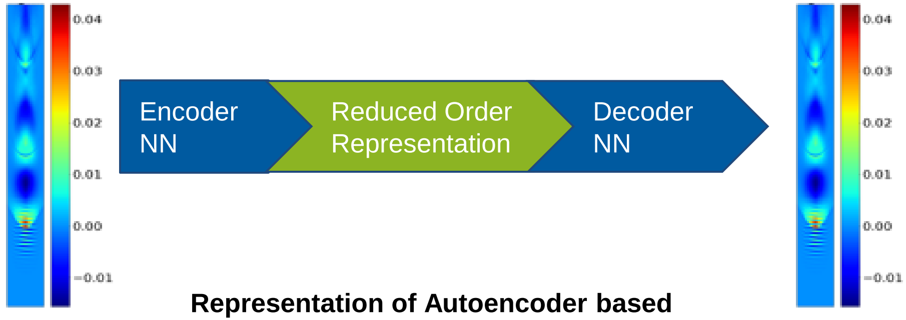
Reduced Order Model (ROM)

Reduced Order Model

- Learns an mapping in a reduced domain
- High memory compression & speed up for forward simulation

Data Generation

- Requires multiple simulation runs
- Massive Amount of Data
- Intensive storage for high fidelity simulation in exa-scale (10^{18}) era
- Data transfer at network speeds
- Access data in streaming fashion



Example of E-field in y-direction

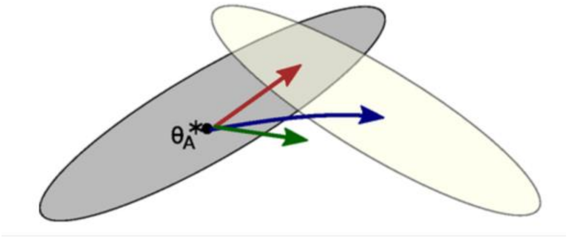
Representation of Autoencoder based ROM

Recreated E-field in y-direction

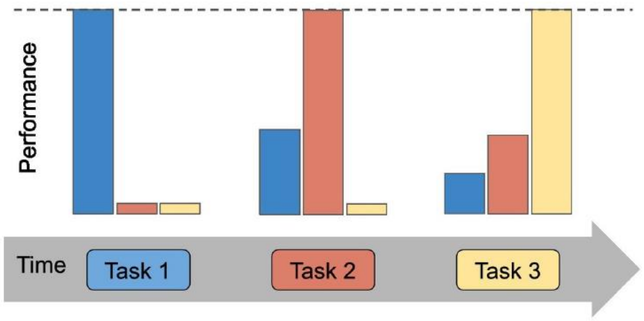
Training ROM

Continual Training

- Learning of **tasks/data** in a continuous way
- Model parameters converge on current task (data)
- High error previous tasks - **Catastrophic Forgetting** in NNs.
- Regularization and constraints on model parameters to avoid forgetting.

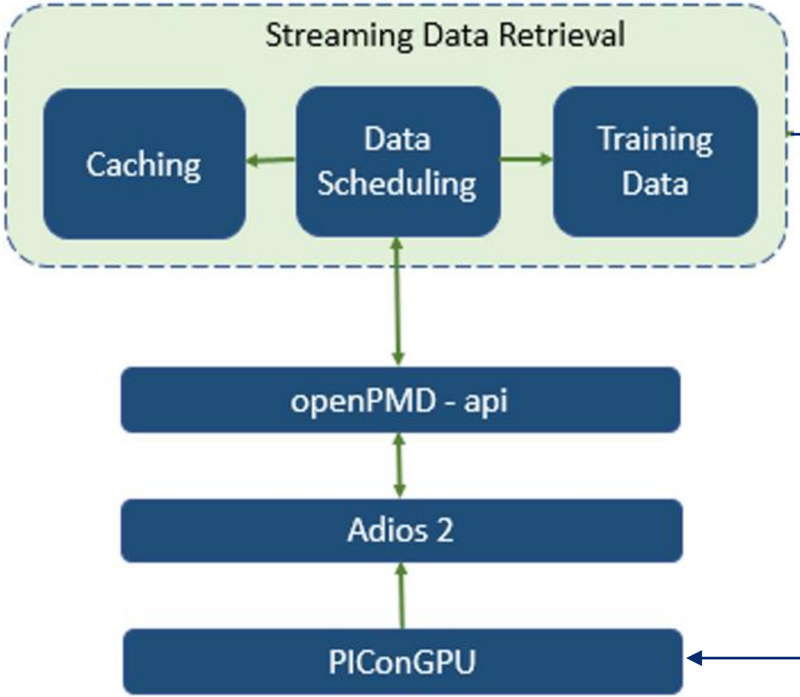


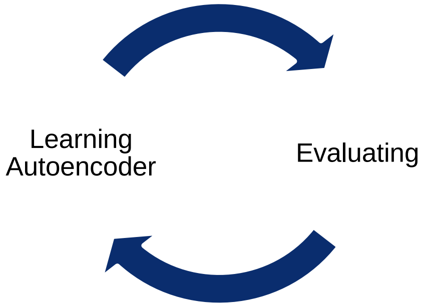
Model Parameters

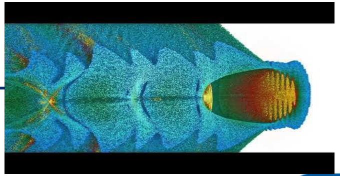


Represntation of Catastrophic Forgetting

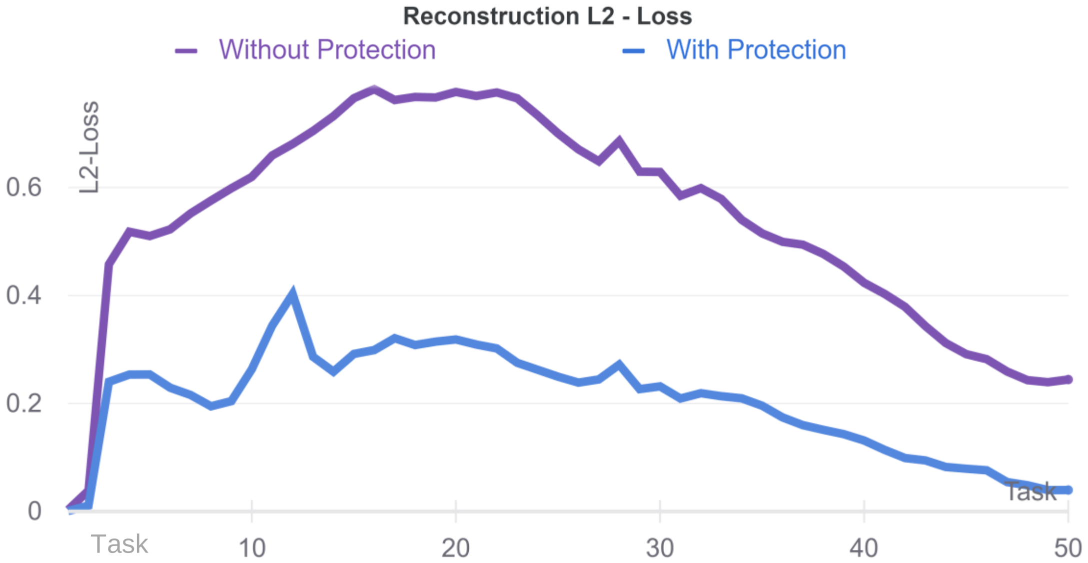
Current Framework ROM







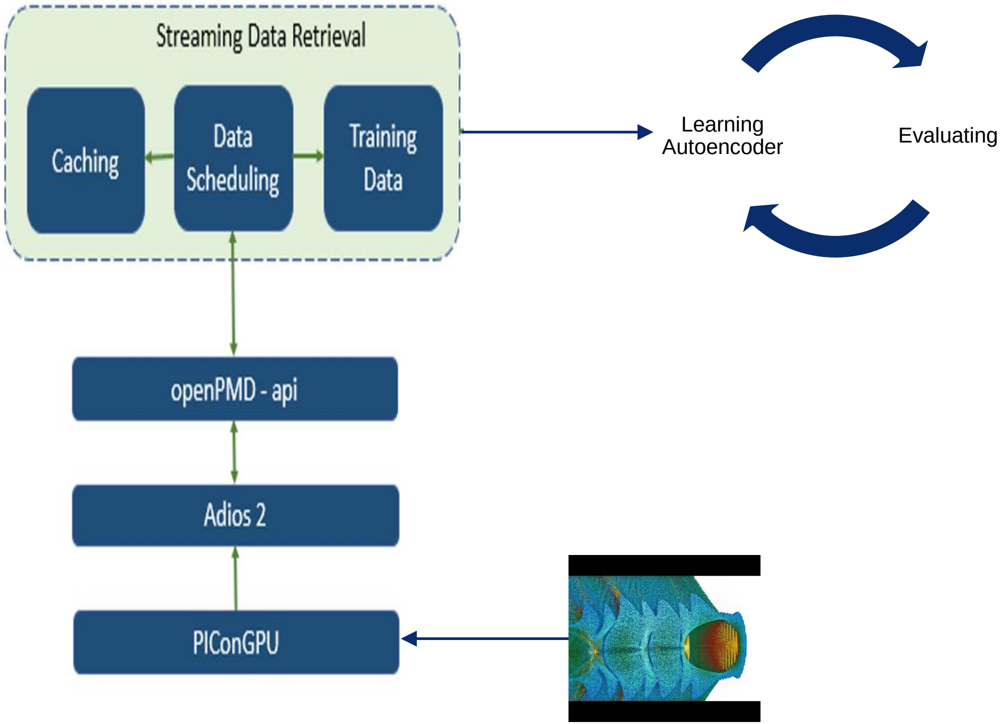
Current Results

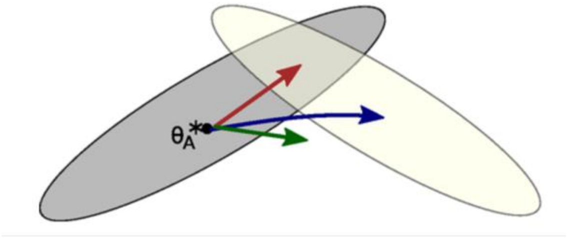


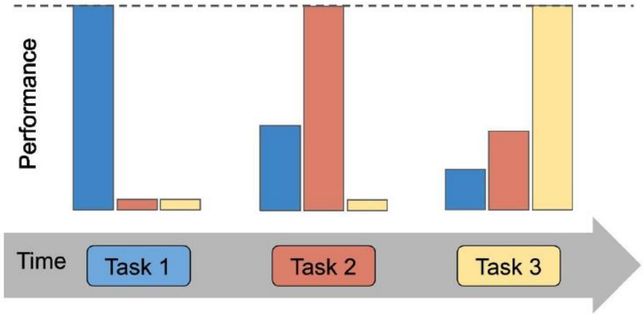
Result after training on entire stream of data in a continual manner.
Reconstruction loss is formulated as:

$$\boldsymbol{loss_t} = \left(\boldsymbol{x_t} - \boldsymbol{decode}\left(\boldsymbol{encode}(\boldsymbol{x_t})\right)\right)^{\mathbf{2}}$$
$$\forall\ \boldsymbol{t} \in \boldsymbol{T}$$

Summary


Streaming Data Retrieval
Caching
Data Scheduling
Training Data
openPMD - api
Adios 2
PIConGPU
Learning Autoencoder
Evaluating


θ_A^*


Performance
Time
Task 1
Task 2
Task 3