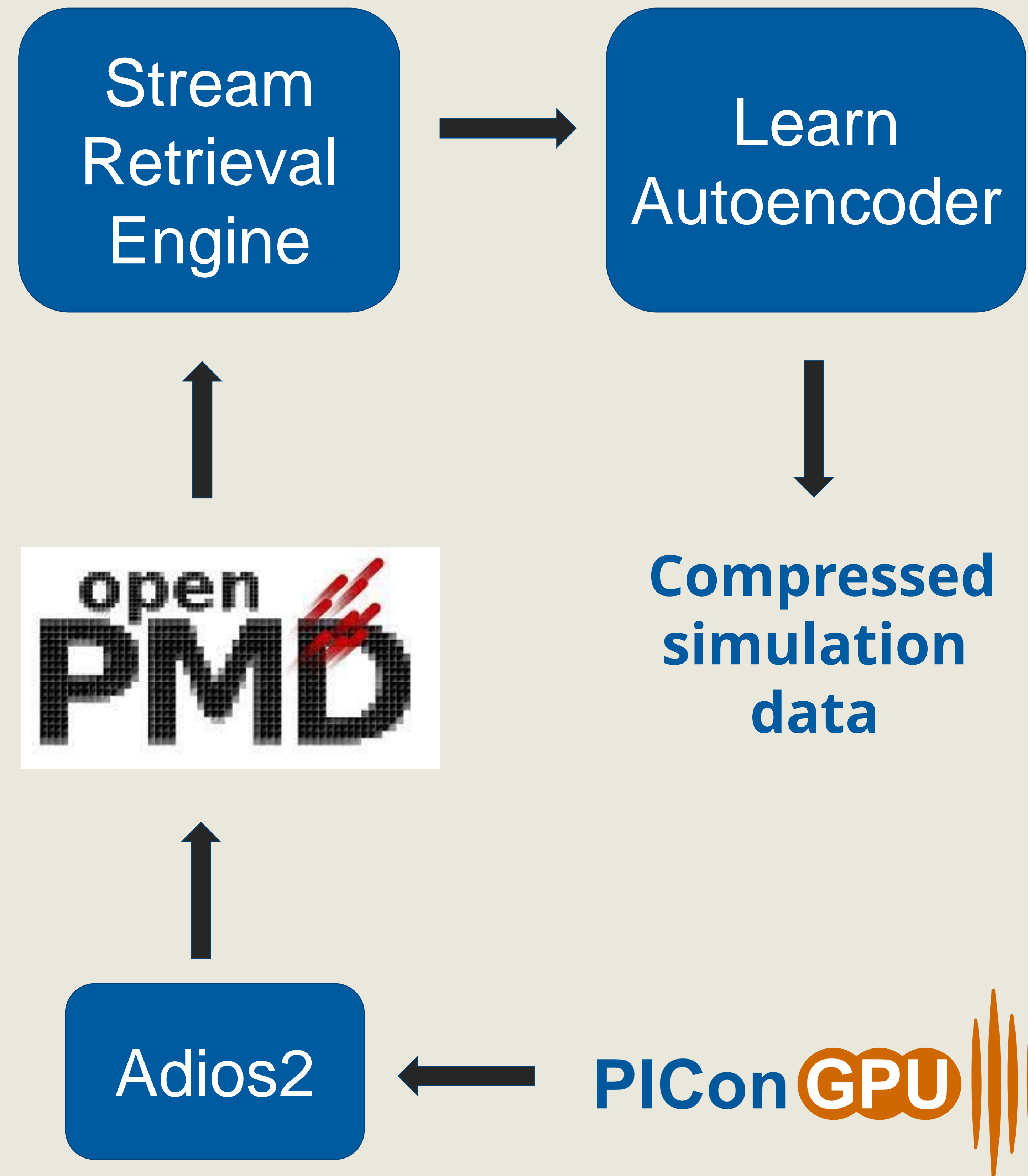


In-Situ ML-based Surrogate Model training via Continual Learning and Streaming

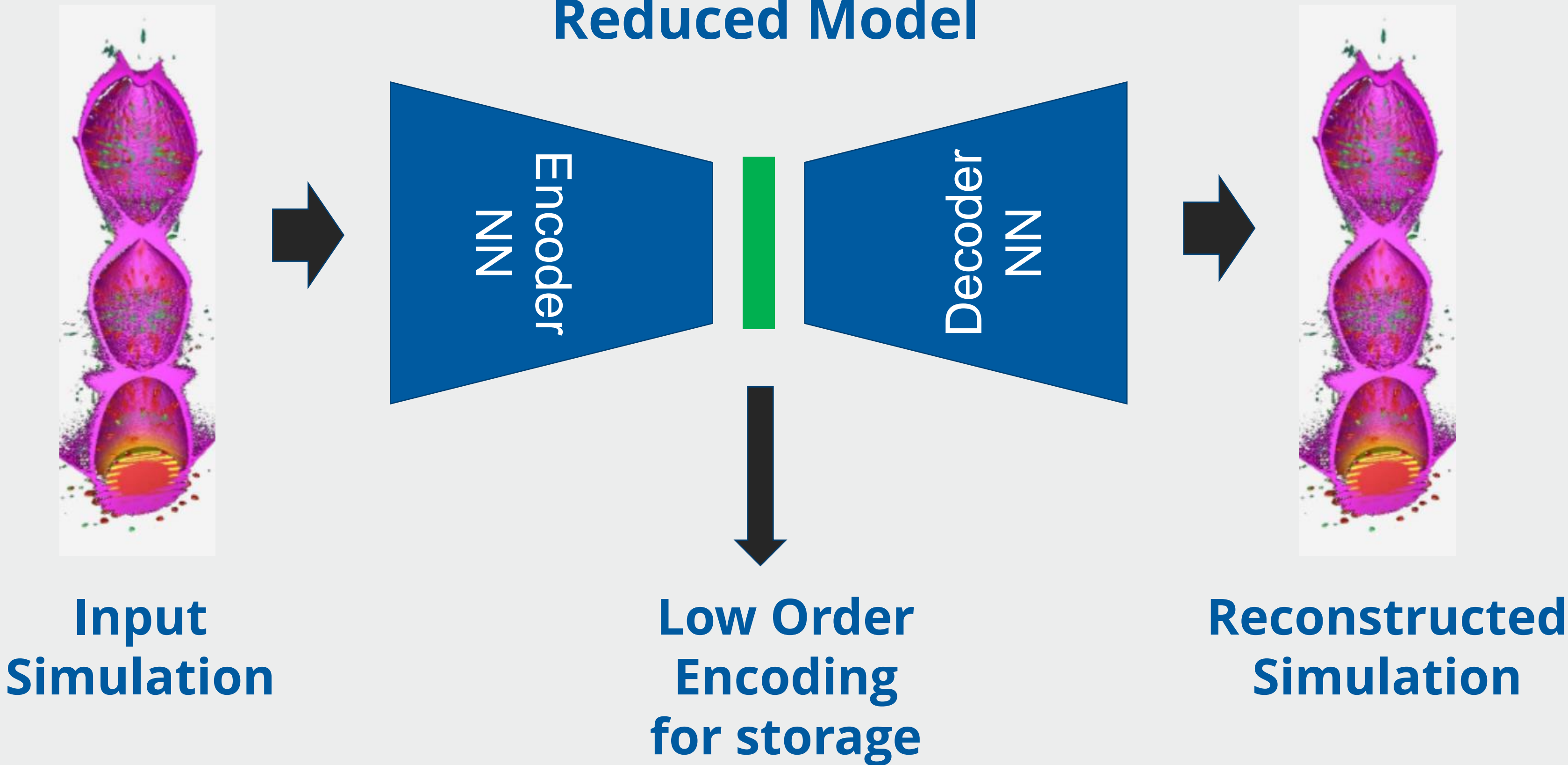
Varun Makdani (HZDR), Franz Poeschel (CASUS), Patrick Stiller (HZDR), Richard Pausch (HZDR), Anna Willmann (HZDR), Dr. Axel Huebl (Lawrence Berkley National Laboratory), Alexander Debus (HZDR), Michael Bussmann (CASUS), Nico Hoffmann (HZDR)

Train a Surrogate model to a concurrent running PIC simulation

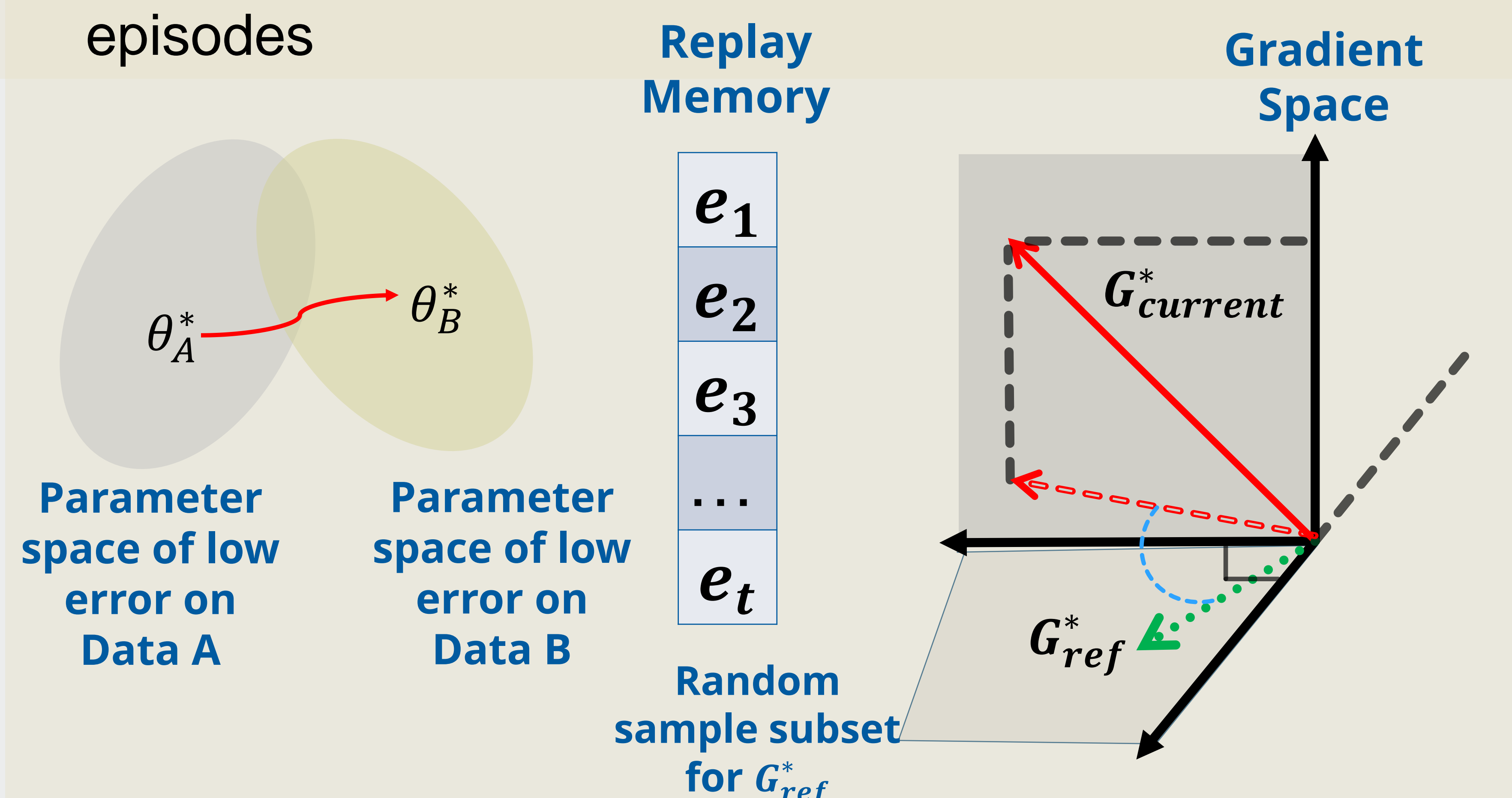
- Data generated from PIC simulations demand massive memory storage for a general surrogate
 - 2-D LWFA Simulations: **5 GB – 100 GB**
 - 3-D LWFA Simulations: **300 GB – 3000 GB (3 TB)**(per simulation)
- Highly compressed representation allows for efficient storage and reconstruction of simulation data



Representation of Autoencoder based Reduced Model



- Simulation accessed as a **stream**, without regaining previously accessed data. Model parameters shift to region of low error on current data, thereby causing **catastrophic forgetting** on previously learned data
- Protect forgetting by **storing encoded samples** from previously learned tasks. Project current data gradients ($G_{current}^*$), in space of gradients (G_{ref}^*) - calculated on a memory subset sampled from replay memory like episodes



- Reconstruction result of 20th data iteration of Electric field of a LWFA simulation (Without vs With protection towards forgetting) after learning 50 data iterations in a stream fashion

