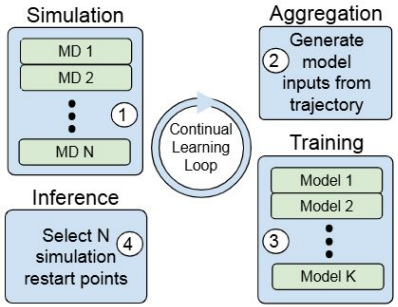
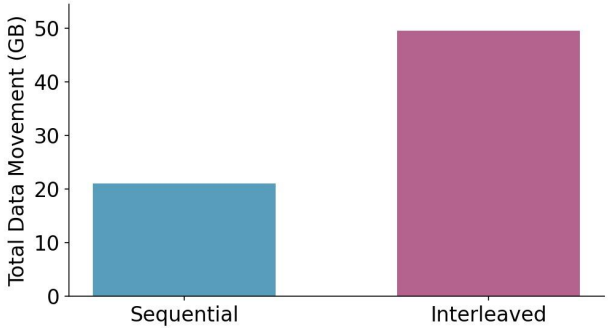
1. Background: Coupling AI to HPC simulation (MD) has been successfully shown to speed up time to results



2. Problem: Naively coupling the simulation and AI creates a large increase in data movement



3. Solution:

Two Core Policies:

1. **Temporal Batching:** Run the AI model only once every **N** simulation steps, amortizing the data movement cost over more computation.
2. **State-Aware Thresholding:** Run the AI only when the simulation's physical state has changed significantly (e.g., RMSD > **T**), avoiding redundant calculations.

4. Expected Results & Impact

- **Efficiency Gains:** We expect to demonstrate a **reduction in total data movement** and a corresponding **increase in application throughput**
- **Architectural Harmony:** We will show via **Roofline analysis**, making it a more efficient match for the underlying GPU architecture.
- **Broader Impact:** This research provides a generalizable strategy for efficiently coupling AI and HPC applications, enabling more scalable and energy-efficient AI-driven scientific discovery.