Hewlett Packard Enterprise

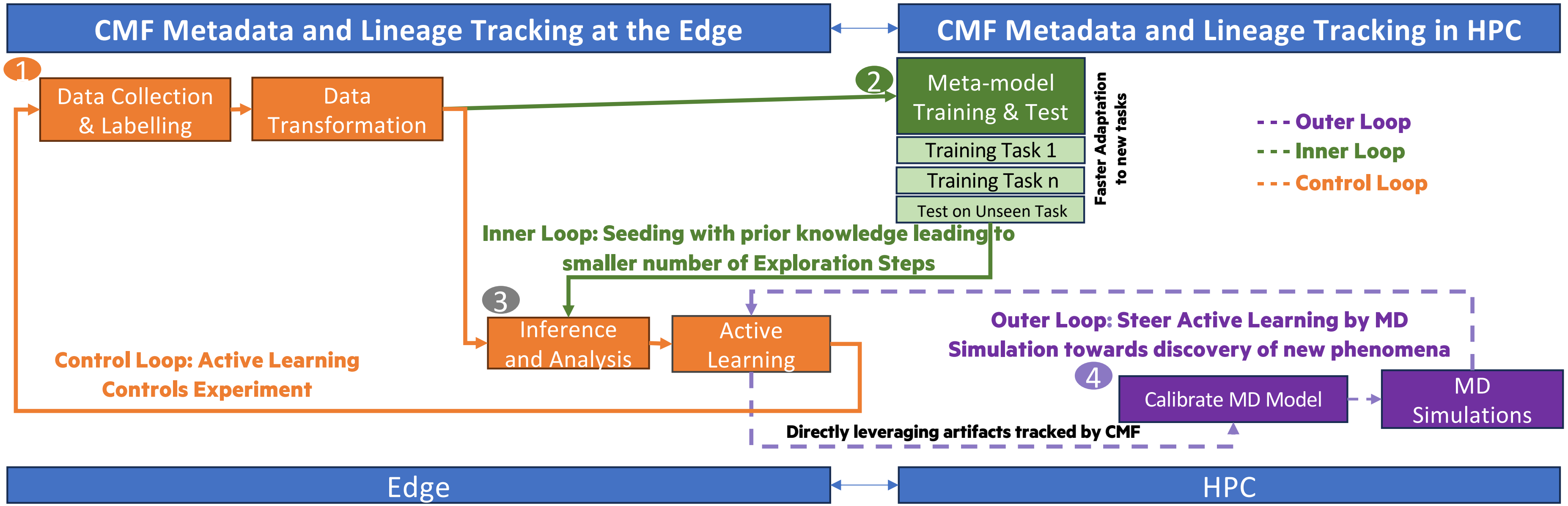
Common Federation Framework for Autonomous Instrumentation & Algorithmic Steering

Larry Kaplan, Senior Distinguished Technologist
Aalap Tripathy, Master Technologist

March 12th, 2024

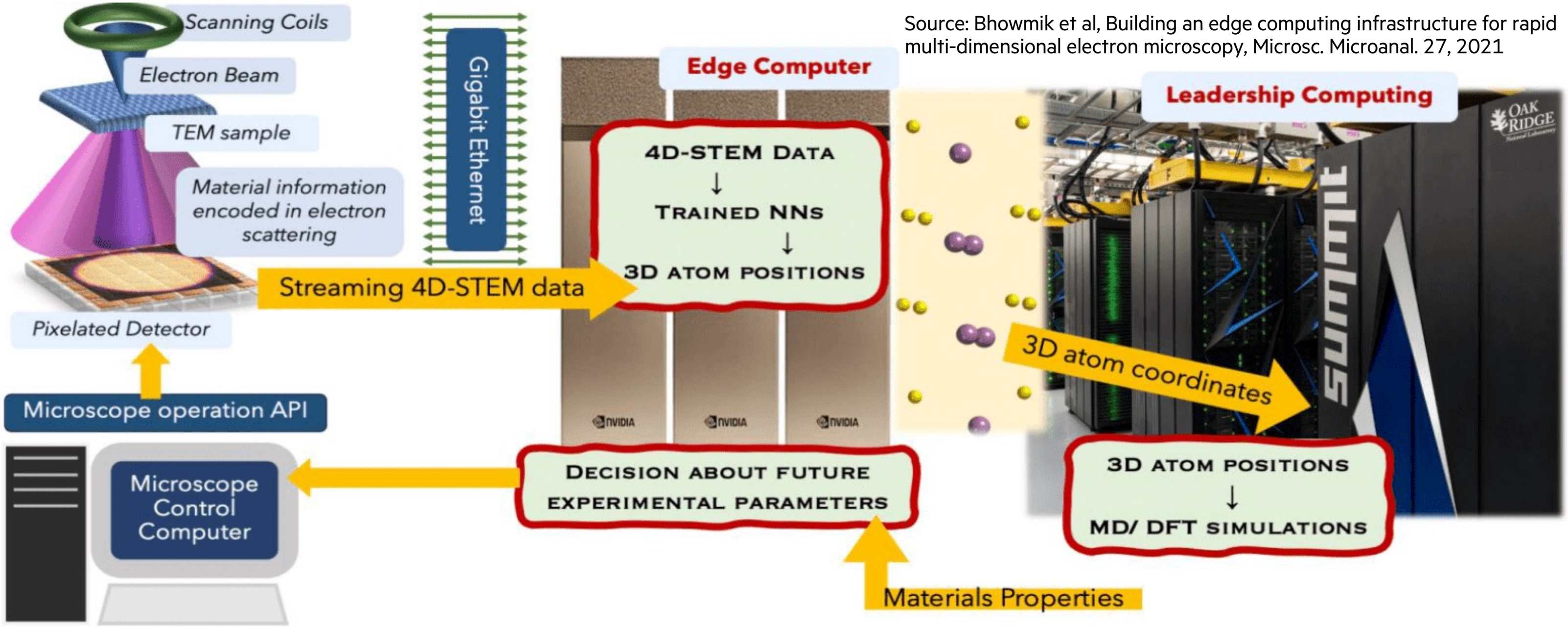
Outline

- Enabling experiment steering in a federated workflow incorporating HPC simulation, AI training & inferencing, edge instrumentation
- Improving a spectrum reconstruction algorithm to be faster, more accurate, and generalizable
- Using log and metadata generated in this workflow to autonomously steer a science experiment



A Representative Edge-to-Exascale Autonomous Instrumentation Workflow

Shared Services GitHub | JupyterHub | Globus | Constellation | DataFed



Source: Bhowmik et al, Building an edge computing infrastructure for rapid multi-dimensional electron microscopy, Microsc. Microanal. 27, 2021

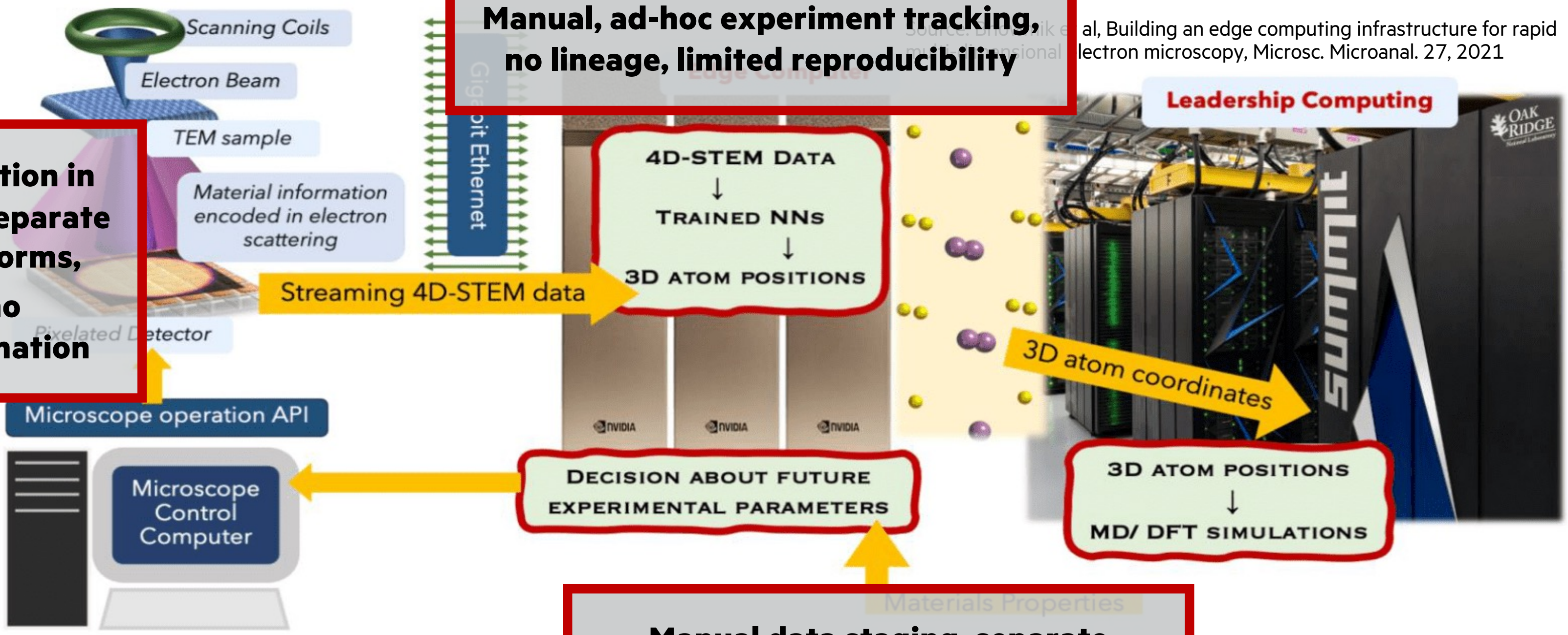
Challenges in Edge-to-Exascale Autonomous Instrumentation Workflows

Shared Services | GitHub | JupyterHub | Globus | Constellation | DataFed

Manual, ad-hoc experiment tracking, no lineage, limited reproducibility

al, Building an edge computing infrastructure for rapid
lectron microscopy, Microsc. Microanal. 27, 2021

Execution in silos, separate platforms, no automation

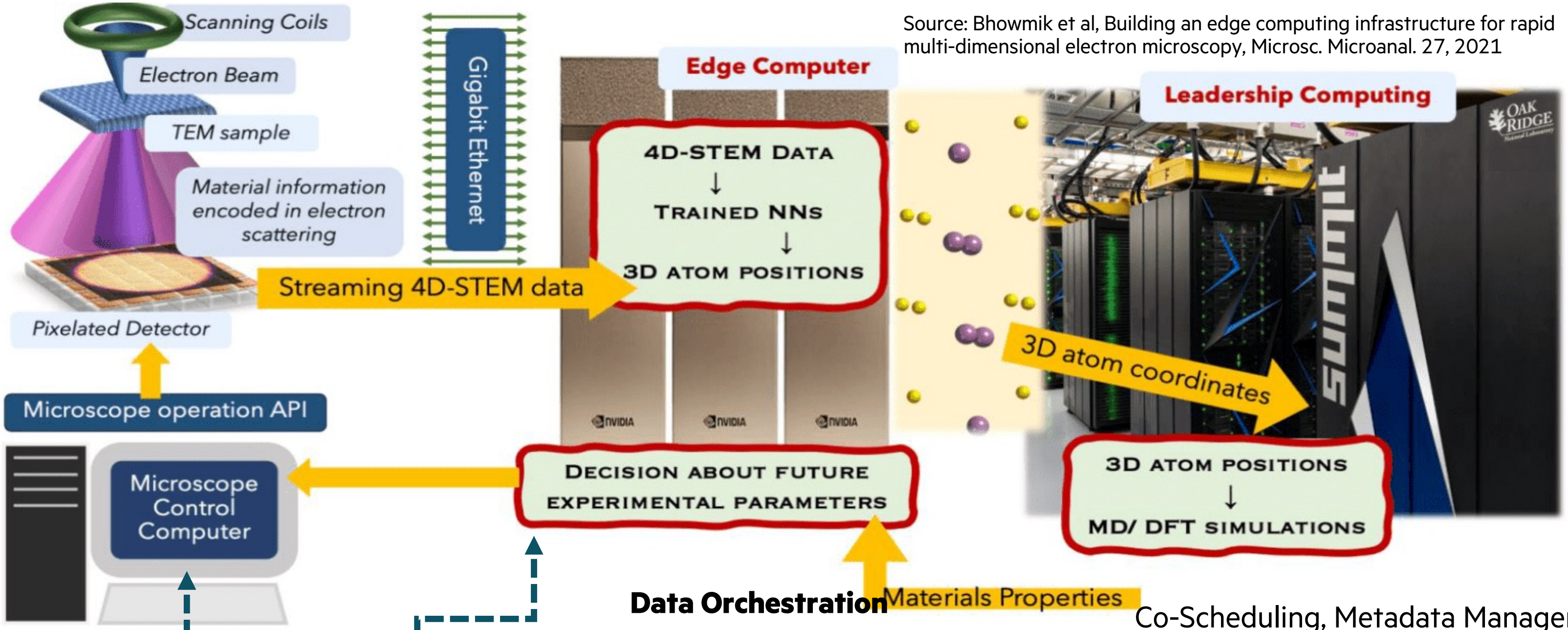


Manual data staging, separate orchestration, difficult interaction with HPC job scheduler

Automation of Edge-to-Exascale Autonomous Instrumentation Workflow

Shared Services GitHub | JupyterHub | Globus | Constellation | DataFed | Metadata Hub (3)

Source: Bhowmik et al, Building an edge computing infrastructure for rapid multi-dimensional electron microscopy, Microsc. Microanal. 27, 2021



Experiment Data and AI Pipeline Metadata

Data Orchestration
- Streaming
- Staging
- Store-forward
- DTN connection

Co-Scheduling, Metadata Management, Experiment Tracking and Optimization

(2) Data Foundation (Common Metadata Framework)

(1) Edge-to-Exascale SDK (Common Federation Framework)

References: - G. Saranathan et.al., Towards Rapid Autonomous Electron Microscopy with Active Metalearning, SC 23, https://doi.org/10.1145/3624062.3626085
- G. Saranathan et.al., Enrichment and Acceleration of Edge to Exascale Computational Steering STEM Workflow using CMF, submitted CUG 2024

Challenges in Autonomous Instrumentation & Steering

Long & costly experiment time

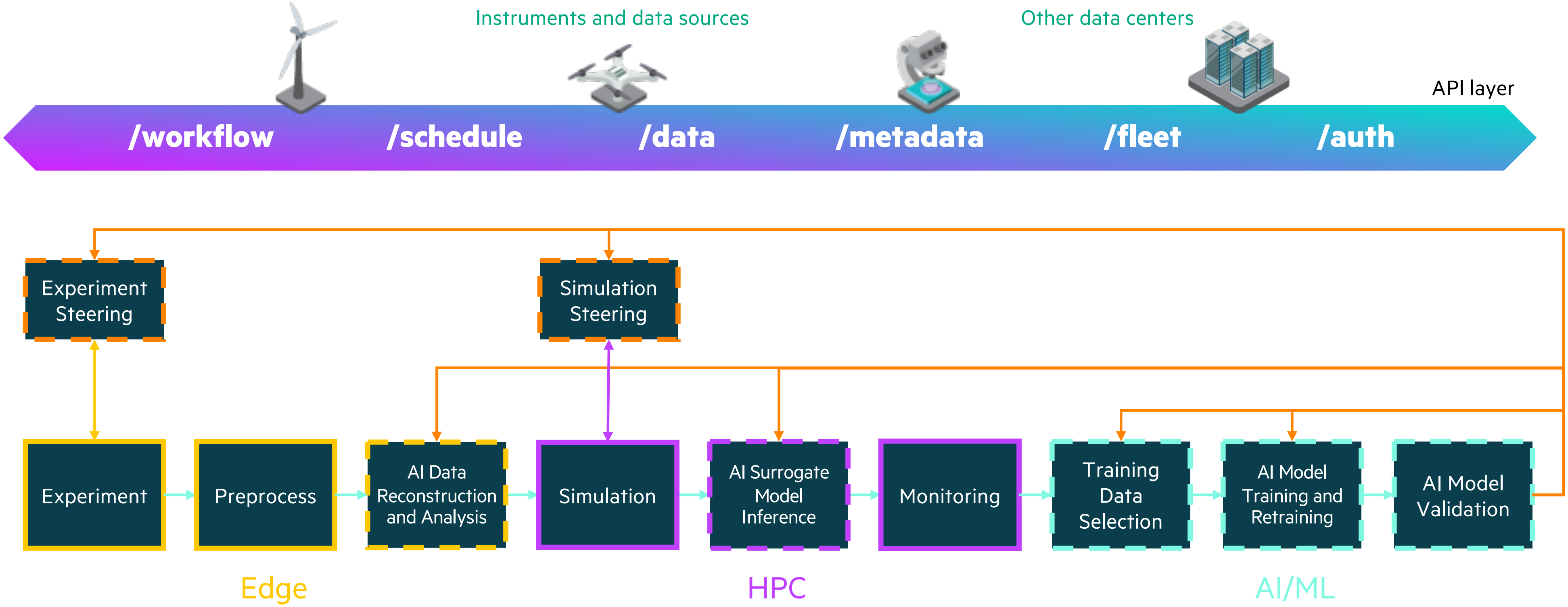
Sample exposure time

Algorithmic generalization time efficiency

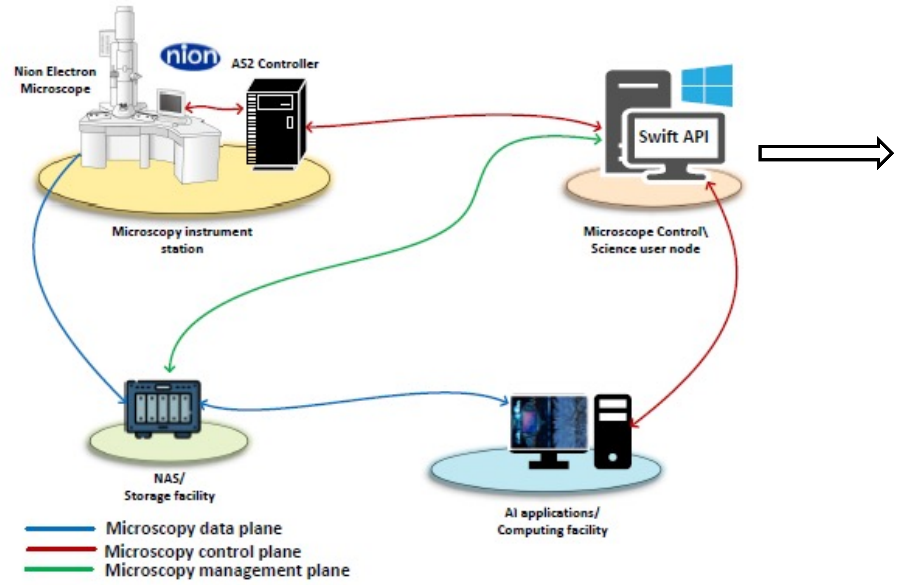
Lack of automation, siloed & ad-hoc execution

Federated Workflow Deployment SDK

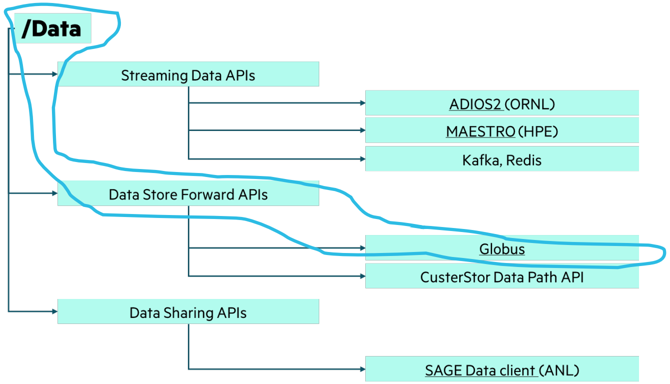
Enables federated hybrid workflows on data from Edge to Extreme-Scale to Cloud



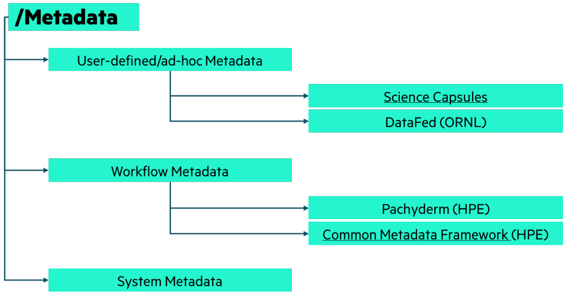
Orchestrating End-to-end Data Transfers and Metadata Tracking



Source: Enabling Autonomous Electron Microscopy for Networked Computation and Steering (2021)



Proposed /data API's in the Common Federation Framework

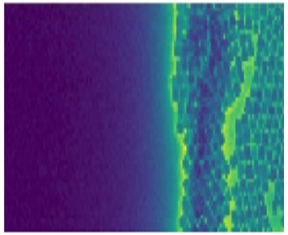


Proposed /metadata APIs in the Common Federation Framework

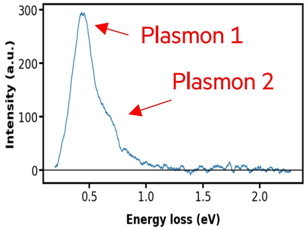
- Executing data streaming, store-forward jobs in edge instrumentation, storage and compute domains
- Enabling ad-hoc and workflow generated metadata to be logged and queried across compute domains

Steering Microscopy Experiments with Active Meta Learning

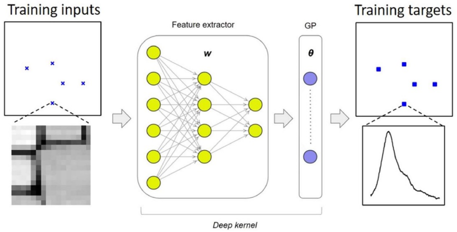
1. Structural imaging is cheap



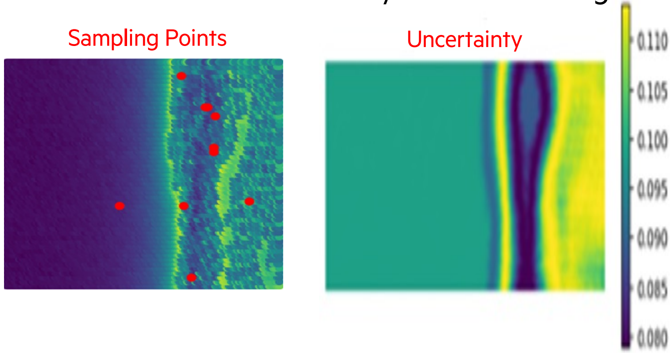
2. Functional analysis is expensive



3. Neural Network reconstructs Structure-Function relationship



4. Neural network is trained by Active Learning



Problem Statement

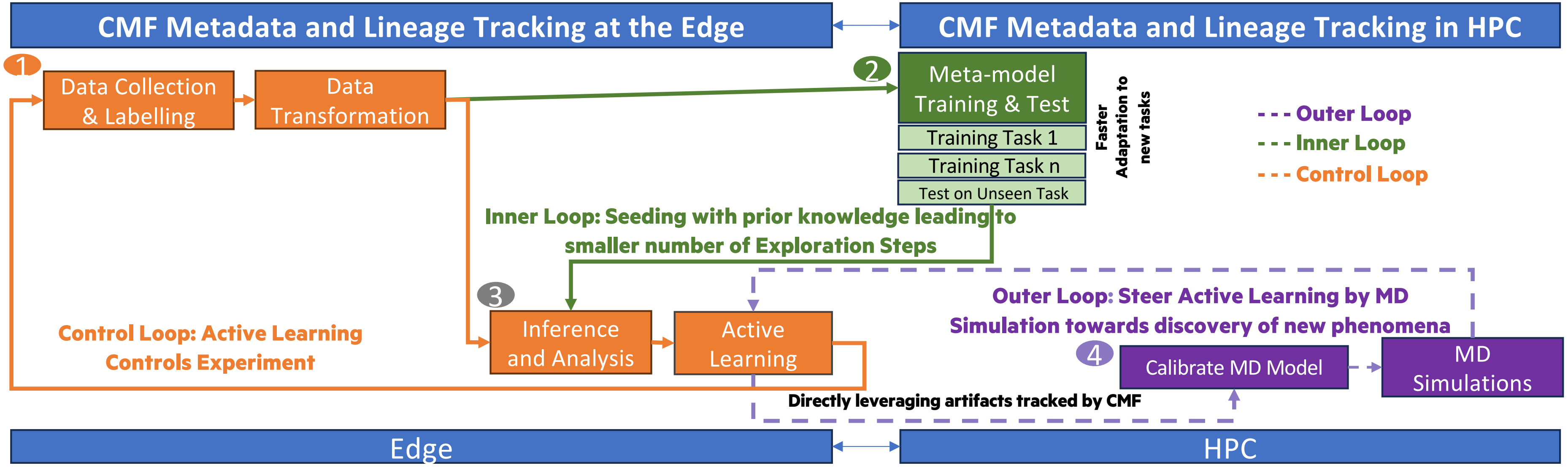
1. Prior knowledge not utilized in model development
2. Functional analysis disconnected from physics modeling

Our Solution

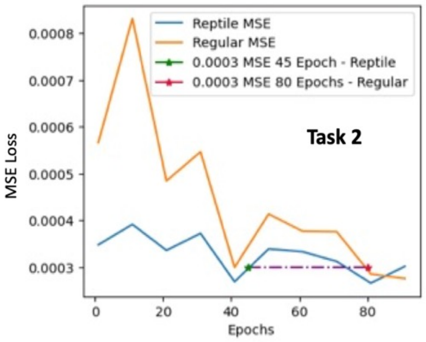
1. Seed Active Learning with a meta-model trained on results from multiple microscopy experiments and sites
2. Drive spectrum-to-function assignment and Active Learning by Molecular Dynamic (MD) simulations. In turn, calibrate MD by experiment

References: G. Saranathan et.al., Towards Rapid Autonomous Electron Microscopy with Active Metalearning, SC 23
G. Saranathan et.al., Enrichment and Acceleration of Edge to Exascale Computational Steering STEM Workflow using CMF, submitted CUG 2024

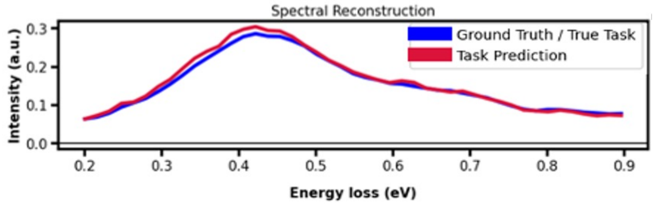
Preliminary Results and Next Steps



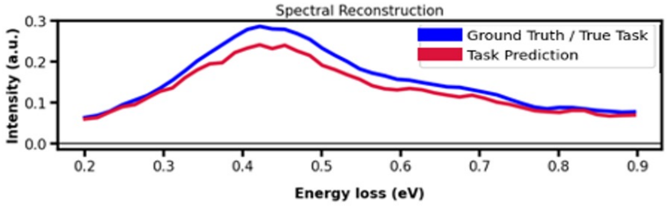
Full Training: Reptile vs. Regular



Active Learning Reptile:

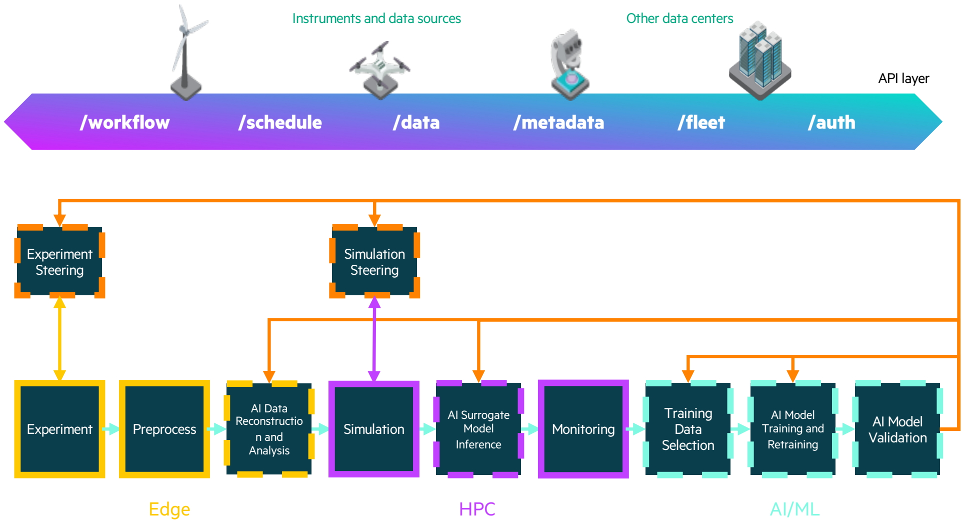


Active Learning Regular:



Results

1. Mapped a real-world Edge + HPC + AI complex workflow involving data, fleet, authentication, and scheduling challenges using **Common Federation Framework (CFF)**
2. Demonstrated algorithmic enhancements to improve efficiency, increase generalization of experiment steering using **Active Meta-learning**
   - Achieves 30-40% reduction in training epochs across domains
   - **Reduces Training Time** while maintaining accuracy
   - Enables efficient training of AI models **across diverse scientific experiments with limited data**
3. Ongoing improvements from workflow metadata capture, lineage tracking, and forensic analysis using **Common Metadata Framework (CMF)**



Future Challenges and Opportunities

- Work in integrating MD simulation to enable simulation steering and integration with real-world instruments is ongoing
- Synchronizing different timescales of execution for simulation, experimentation, AI model training, and active learning workflows
- Learn from prior experiment results and data characteristics to accelerate meta-model optimization after adding new data sets
- Investigate simulation steering (calibration of simulation parameters) from results of active learning
- Consider necessary APIs and their organization for the Common Federation Framework (CFF)

Thank you

Larry Kaplan – lkaplan@hpe.com
Key Contributors:
Gayathri.Saranathan, Martin.Foltin, AnnMary.Roy, Suparna.Bhattacharya, Rangan.Sukumar, Paolo.Faraboschi}@HPE.com
{ZiatdinovMa, GhoshA, RoccaprioreKM}@ornl.gov