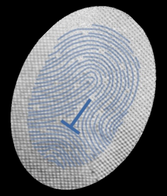
SAND2023-07736C

Exceptional service in the national interest

Sandia National Laboratories

CHiMaD Phase Field XV



Beyond Fingerprinting

Proposed Benchmark Problem for Machine-Learning in Phase Field

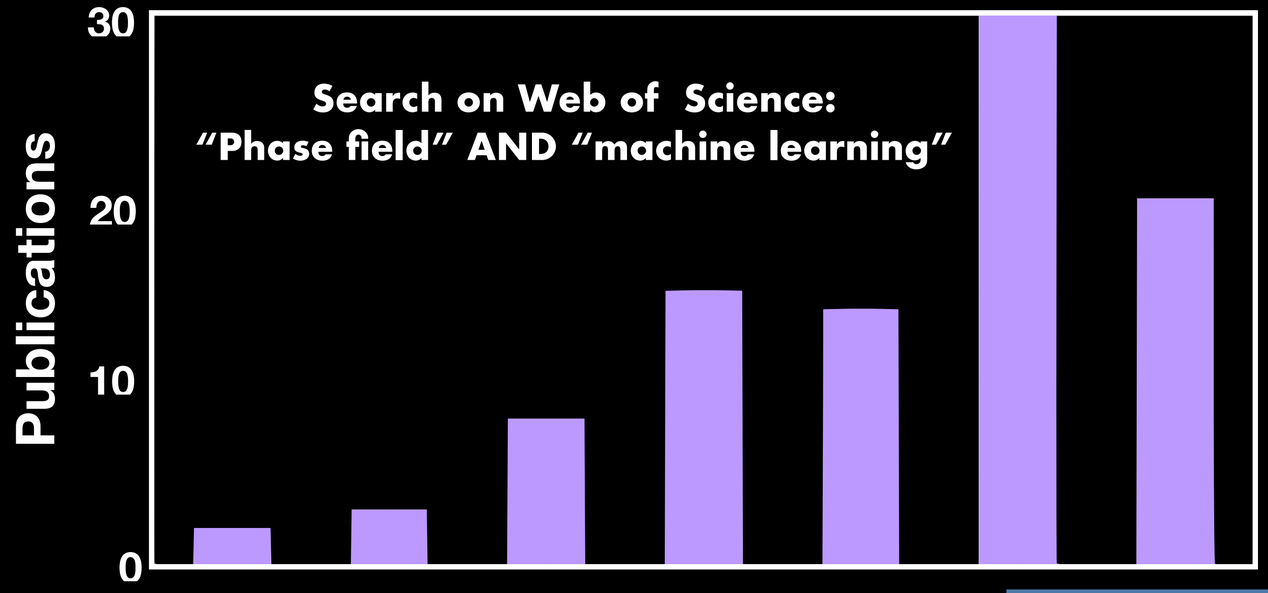
Rémi Dingreville (rdingre@sandia.gov)
Sandia National Laboratories

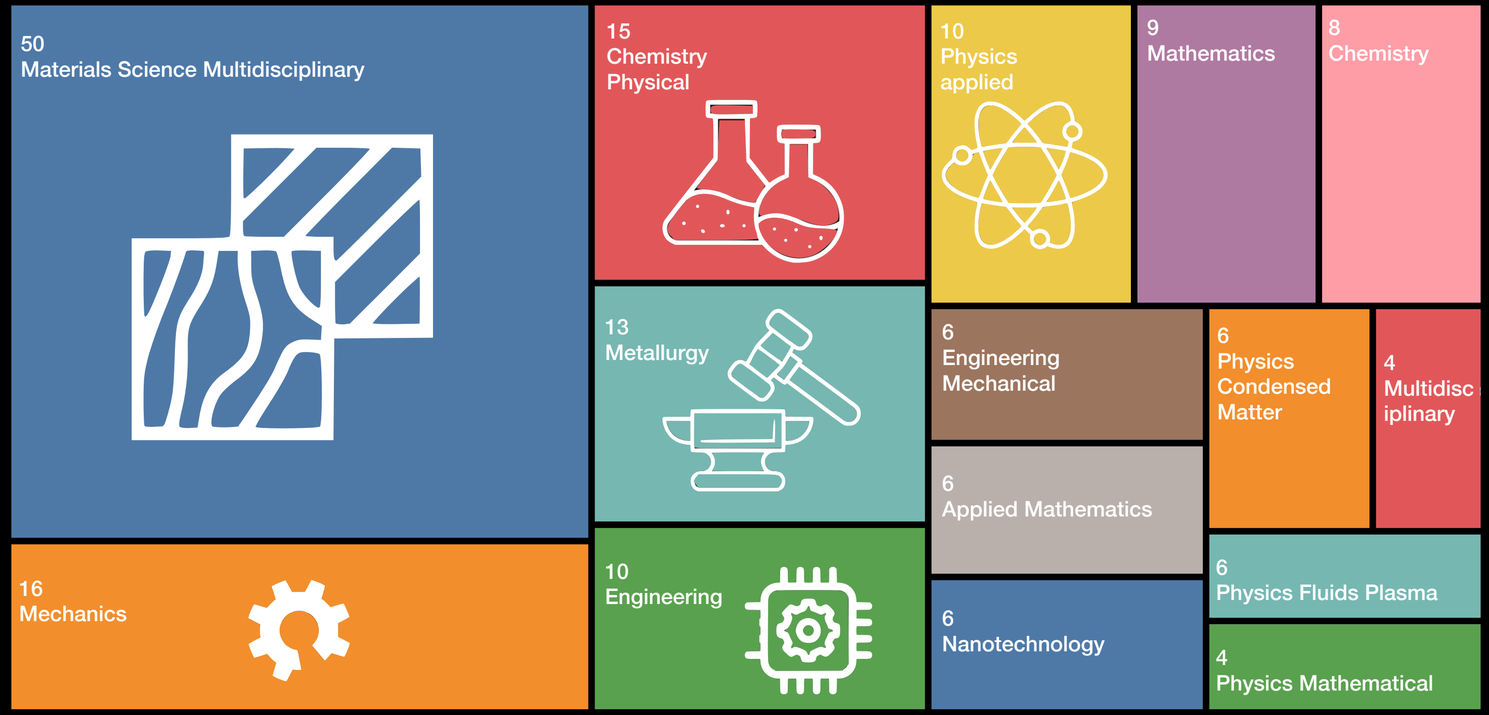
U.S. DEPARTMENT OF ENERGY | NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Over the Past 5 Years, ML has Infiltrated the Phase-Field Method

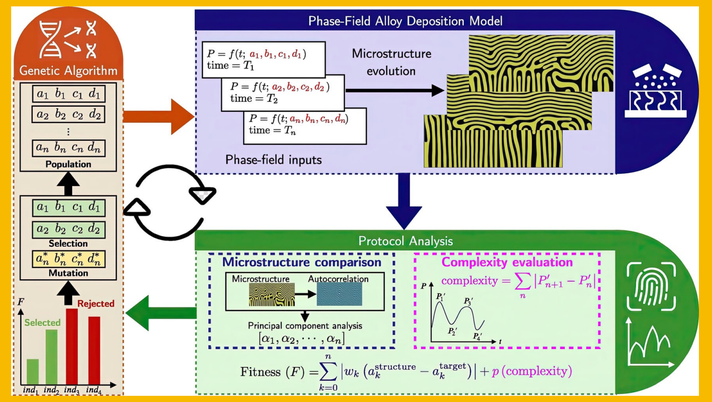




The Effective Use of ML Alg. Creates New Ways to Do "Business" Differently

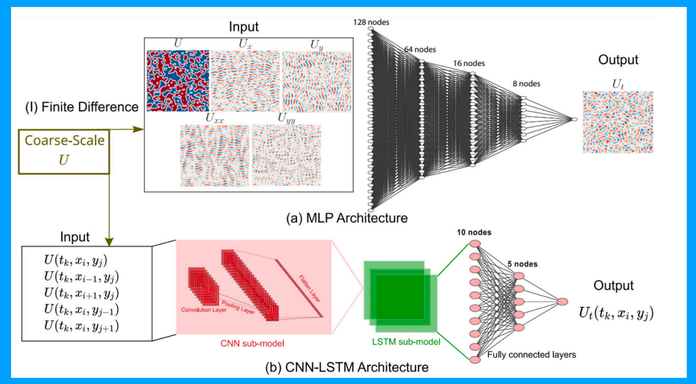
Sandia National Laboratories

Optimization



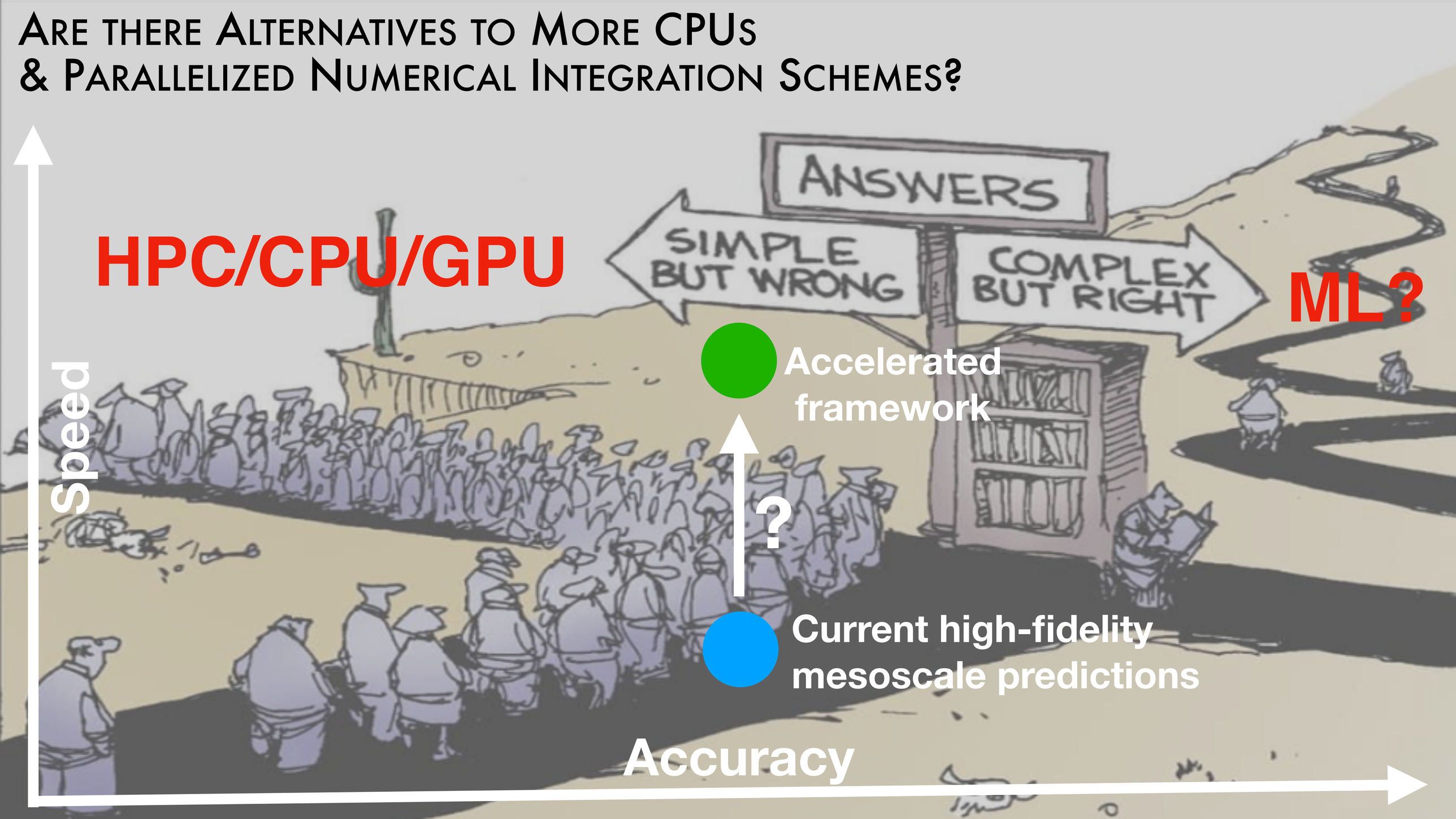
Desai et al., *Mater. Des.*, 2022

Acceleration

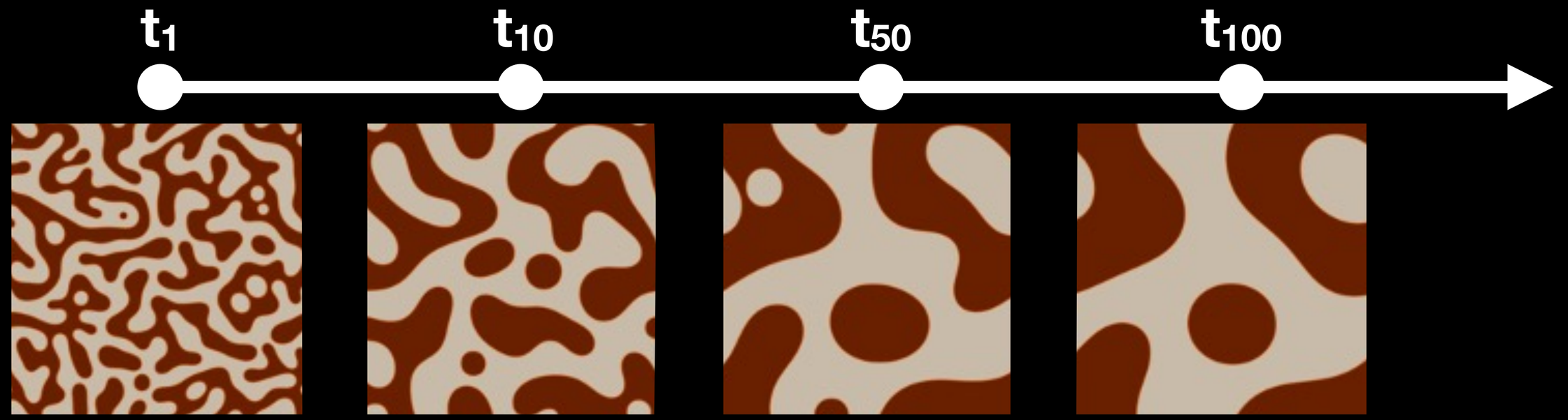


Kiyani et al., *PRE*, 2022

Are there Alternatives to More CPUs & Parallelized Numerical Integration Schemes?



The Stereotypical Problem

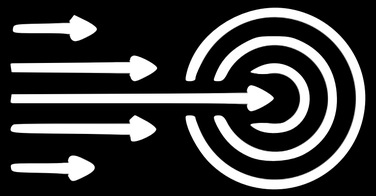


Learning a mapping function:

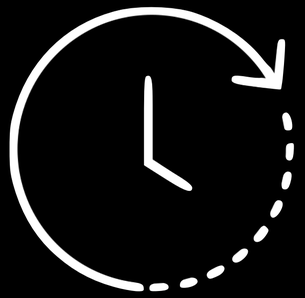
- History of the microstructure, $u(\mathbf{x},t)$
- State of of microstructure at time t, $\phi(\mathbf{x},t)$

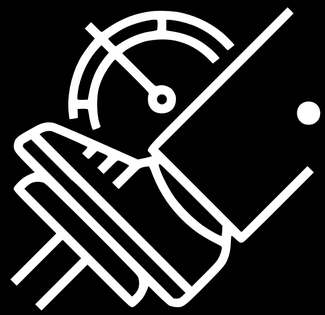
$$\mathcal{G}\left(u(\mathbf{x},t)\right) = \phi\left(\mathbf{x},t\right)$$

What are good metrics to evaluate the performance of a ML-based phase-field capability?

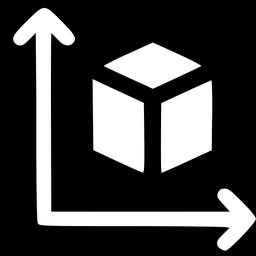


- Spatial accuracy (global vs. local)
- Time accuracy (interpolation vs. extrapolation)

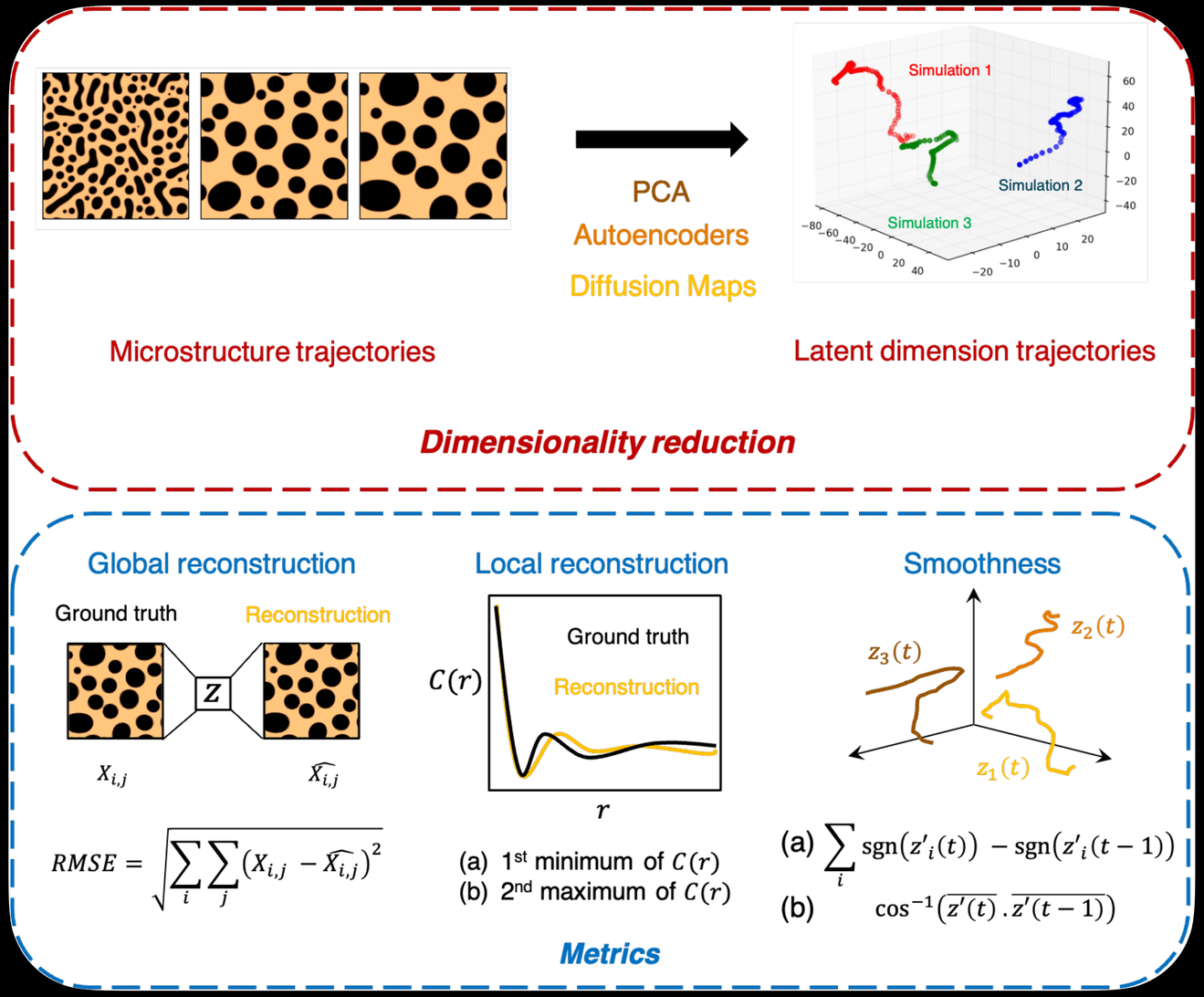




- Speed up & generalization
- Size and time required for training

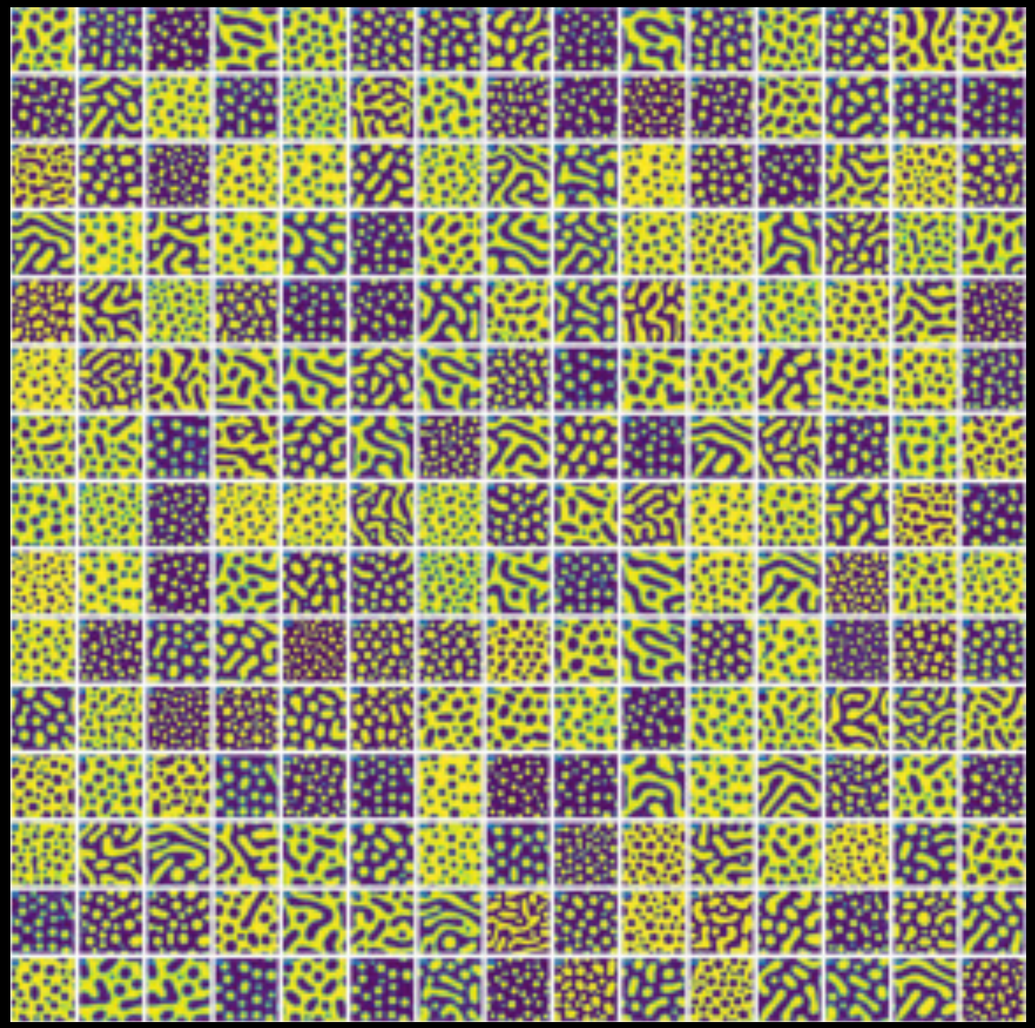


Spatial and temporal accuracies



- **Spatial accuracy:**
  - Global: RMSE on reconstruction or other image-based techniques (PSNR…)
  - Local: Identification of spatial feature
  - Deterministic vs. statistic characterization
- **Time accuracy:**
  - Interpolation vs. extrapolation based on data seen during training
  - Smoothness of the low-dimensional representation?

Training data: the 5 Vs

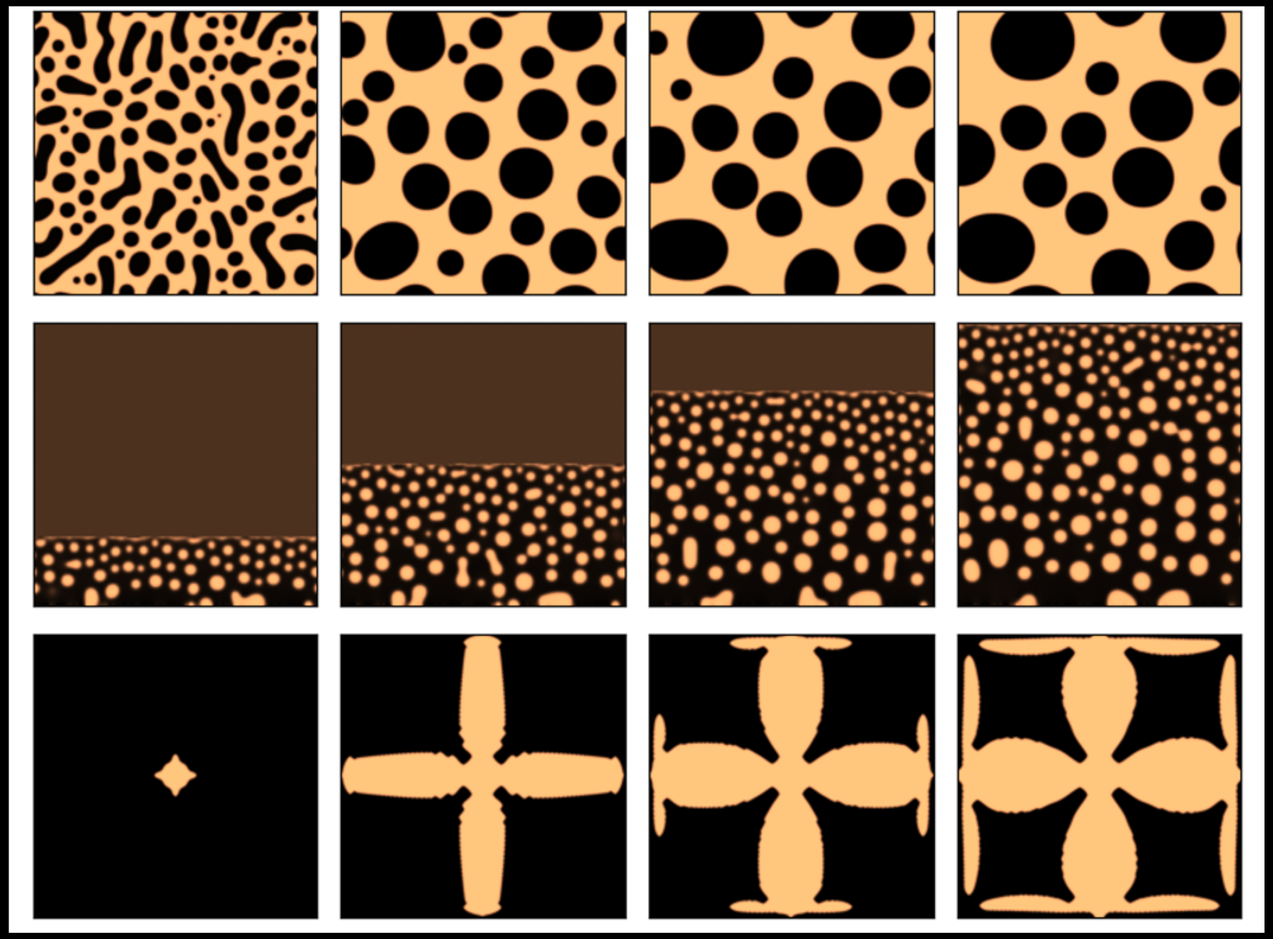


- **Volume:** when is enough enough?
- **Velocity:** how fast can we generate the training data?
- **Variety:** how to change input parameter to have enough flavor?
- **Veracity:** quality and accuracy of the training data being generated
- **Value:** how much added value do we have with one additional data point in our data set?

Speed-up and generalization…

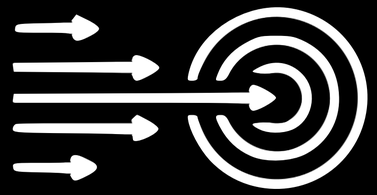
- **CPU (parallelization) vs. GPU**
- **Interoperability:** can I use it with various PF platforms?
- **Representability (unit test):** does the ML-based model correctly represent basic features of the PF model (free energy, conservation of field variables, etc.)?

Formulation of a ML-related BM…

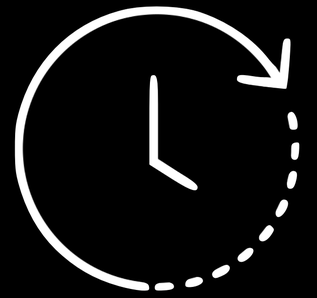


- Spinodal decomposition (BM1)
  - Simple
  - Pre-existing datasets
- Physical vapor deposition
  - Pre-existing datasets
  - Moving boundary and evolving size of computational domain
- Dendrite (BM3)
  - Different length scales
  - Highly sensitive to formulation and numerical scheme
- Other?
  - Oswald ripening (BM2): coupling of conserved and non-conserved dynamics?
  - Nucleation (BM8)

Submission guideline

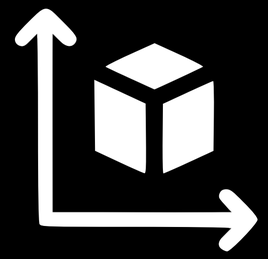


- Spatial accuracy (global vs. local)
  - Reconstruction vs. time
  - Local reconstruction of specific microstructural features



- Time accuracy (interpolation vs. extrapolation)
  - Interpolation performance for within distribution
  - Extrapolation performance outside of distribution

- Speed up & generalization
  - Speedup compared to DNS
  - Free energy evolution
  - Extrapolation to unseen data?



- Size and time required for training
  - How much data has been used
  - Time and resources required for training