



Exascale HPC for Scalable AI and Simulation Workflows

Andreas Lintermann[†]

[†] Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH,
Wilhelm-Johnen-Straße, 52425 Jülich, Germany
A.Lintermann@fz-juelich.de

AI technologies are nowadays used in almost every area of research and industry. They automate and accelerate processes, and uncover new, unseen relations in huge datasets. Such technologies have become key to develop new models for complex simulations, e.g., reduced-order models, surrogates, or closure models. Modular and heterogeneous systems with accelerator and/or specialized AI components are found in state-of-the-art Exascale HPC systems, providing the opportunity to deal with the demands of complex and intertwined simulations and AI data processing workflows.

This talk provides an overview of European state-of-the-art HPC technologies and develops an understanding from hardware over application to workflow perspective, with an emphasis on intertwining simulations and AI. It specifically discusses training AI at scale for a physics application, with recommendations on increasing the training performance by intelligently selecting the right parameters for training and optimization. These recommendations are complemented by a presentation of tuning strategies to improve model accuracy by means of hyperparameter optimization techniques that make efficient use of the underlying HPC hardware. It is shown how such technologies embed into machine-learning operations (MLOps) workflows that are capable of running on large-scale HPC infrastructures, and how they might benefit future research applications.