



Contribution ID: 29

Type: YSF oral presentation

## Simulation-based Self-supervised Learning (S3L)

*Monday 30 October 2023 17:30 (10 minutes)*

Self-Supervised Learning (SSL) is at the core of training modern large ML models, providing a scheme for learning powerful representations in base models that can be used in a variety of downstream tasks. However, SSL training strategies must be adapted to the type of training data, thus driving the question: what are powerful SSL strategies for collider physics data? In the talk, we present a novel simulation-based SSL (S3L) strategy wherein we develop a method of “re-simulation” to drive data augmentation for contrastive learning. We show how an S3L-trained base model can learn powerful representations that can be used for downstream discrimination tasks and can help mitigate uncertainties.

### Brainstorming idea [title]

What are the right SSL strategies for HEP data?

### Brainstorming idea [abstract]

Which is the best way to learn powerful representations of HEP data, and what are its most promising applications? We will brainstorm about HEP-typical problems like classification, regression, and clustering and how self-supervised learning strategies can help solving them.

**Primary authors:** MAIER, Benedikt (KIT - Karlsruhe Institute of Technology (DE)); KRUPA, Jeffrey (Massachusetts Institute of Technology); PIERINI, Maurizio (CERN); KAGAN, Michael (SLAC National Accelerator Laboratory (US)); WOODWARD, Nathaniel Sherlock (Massachusetts Inst. of Technology (US)); HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US))

**Presenters:** MAIER, Benedikt (KIT - Karlsruhe Institute of Technology (DE)); KRUPA, Jeffrey (Massachusetts Institute of Technology)

**Session Classification:** Young Scientist Forum