ML model optimization pipeline development

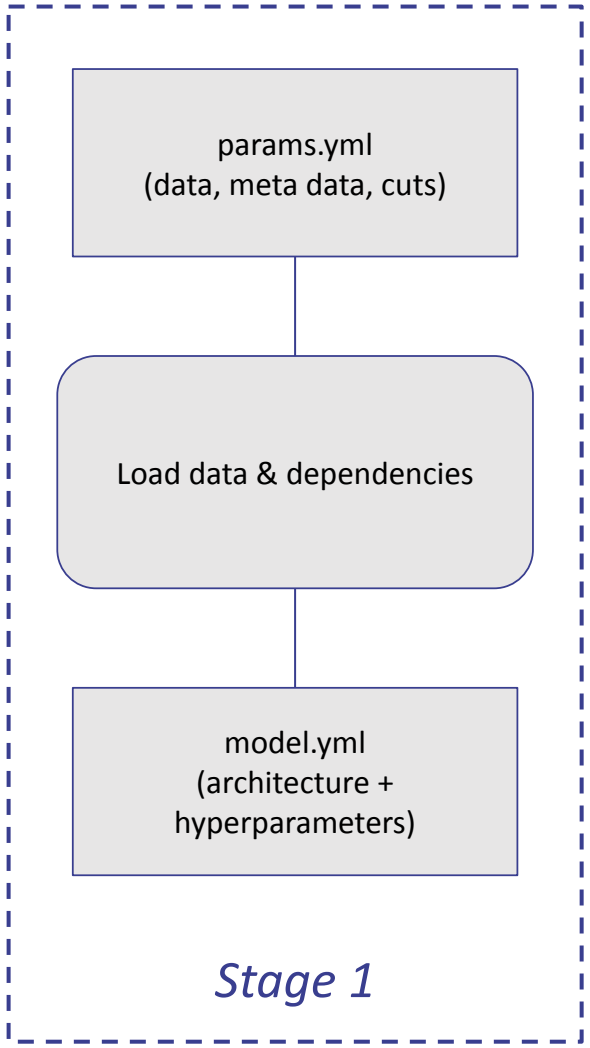
CERN openlab Summer Student Programme Lightning talks

Stylianos Kostakis

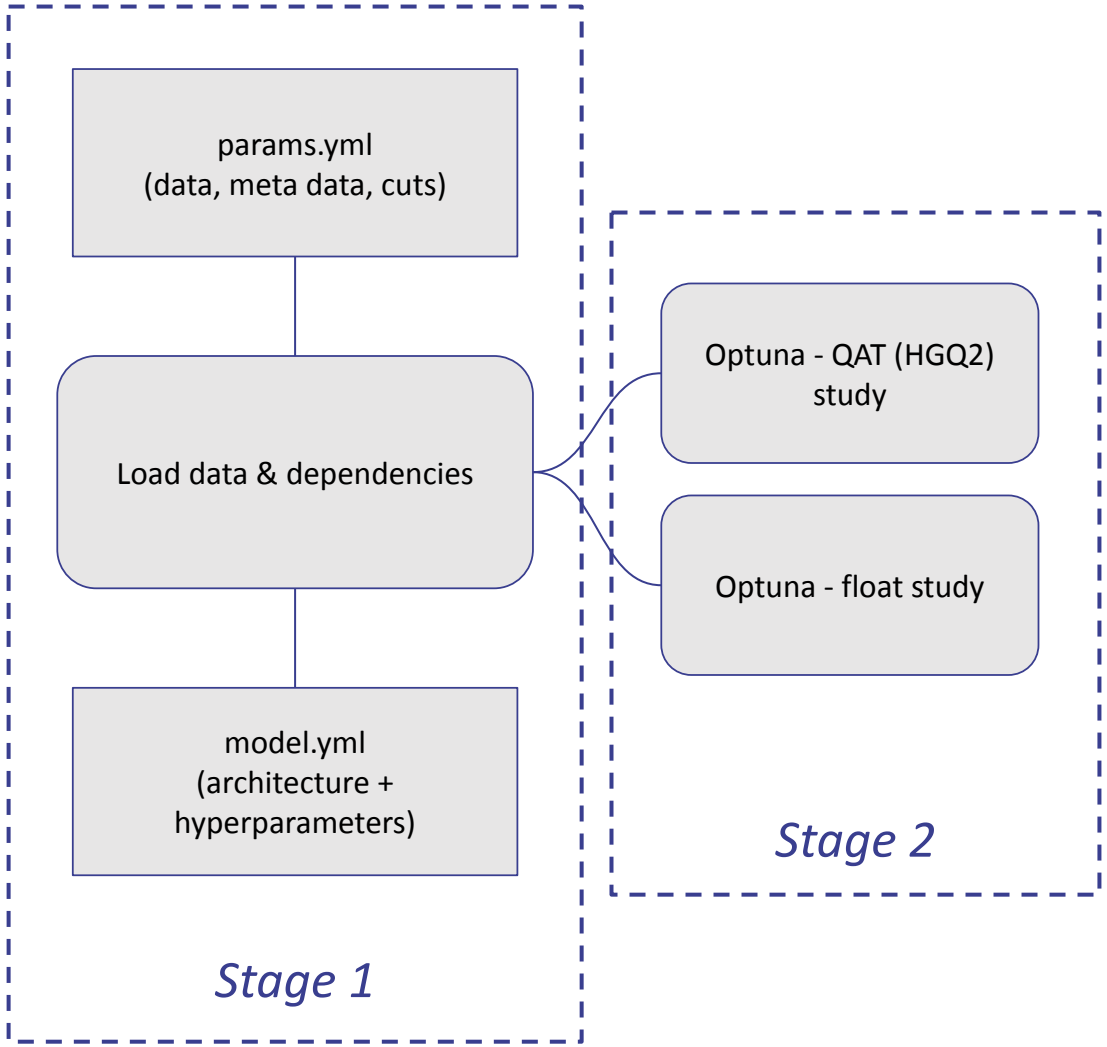
Ioannis Xiotidis

Noah Clarke Hall

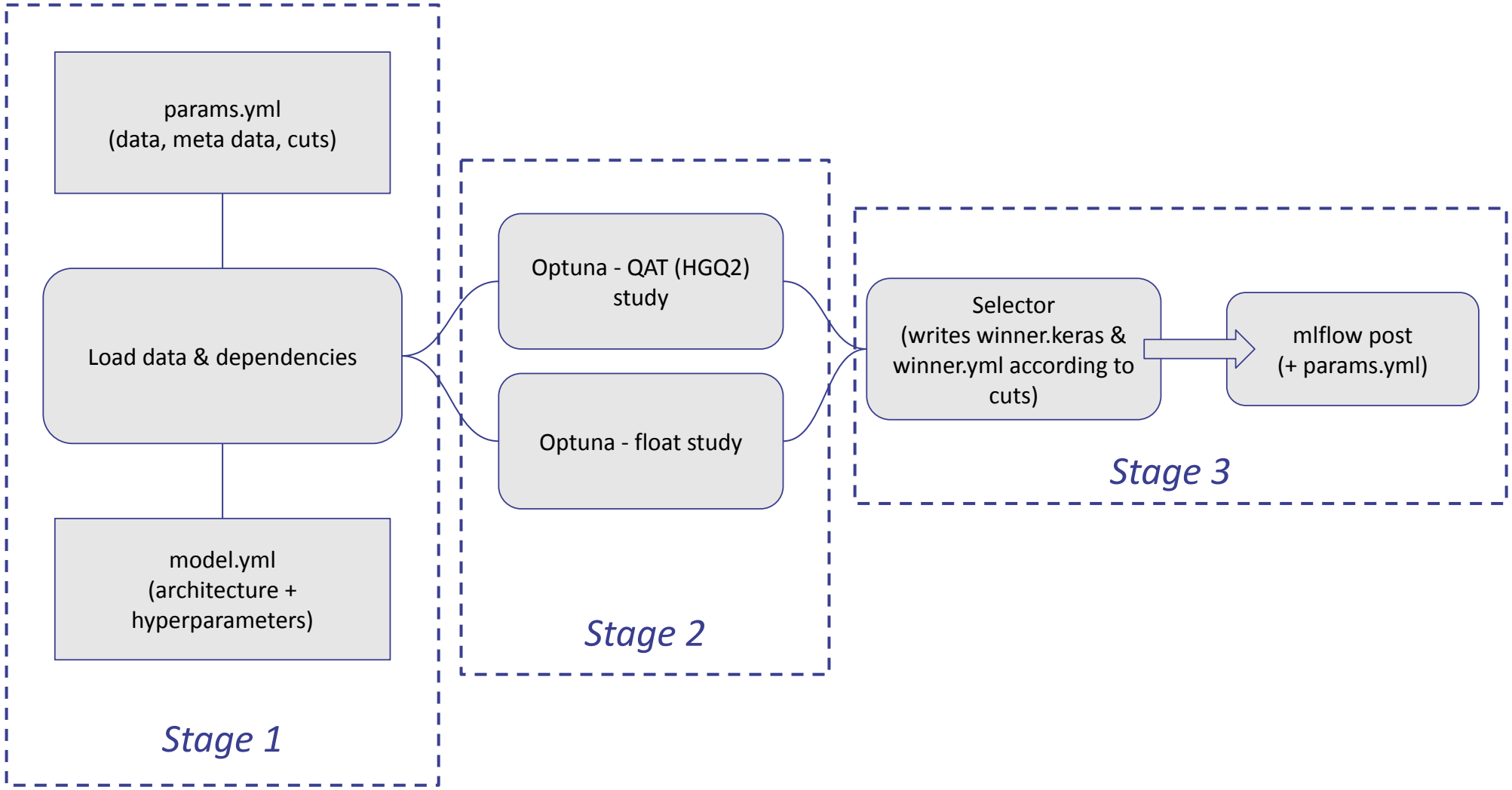
Pipeline Description



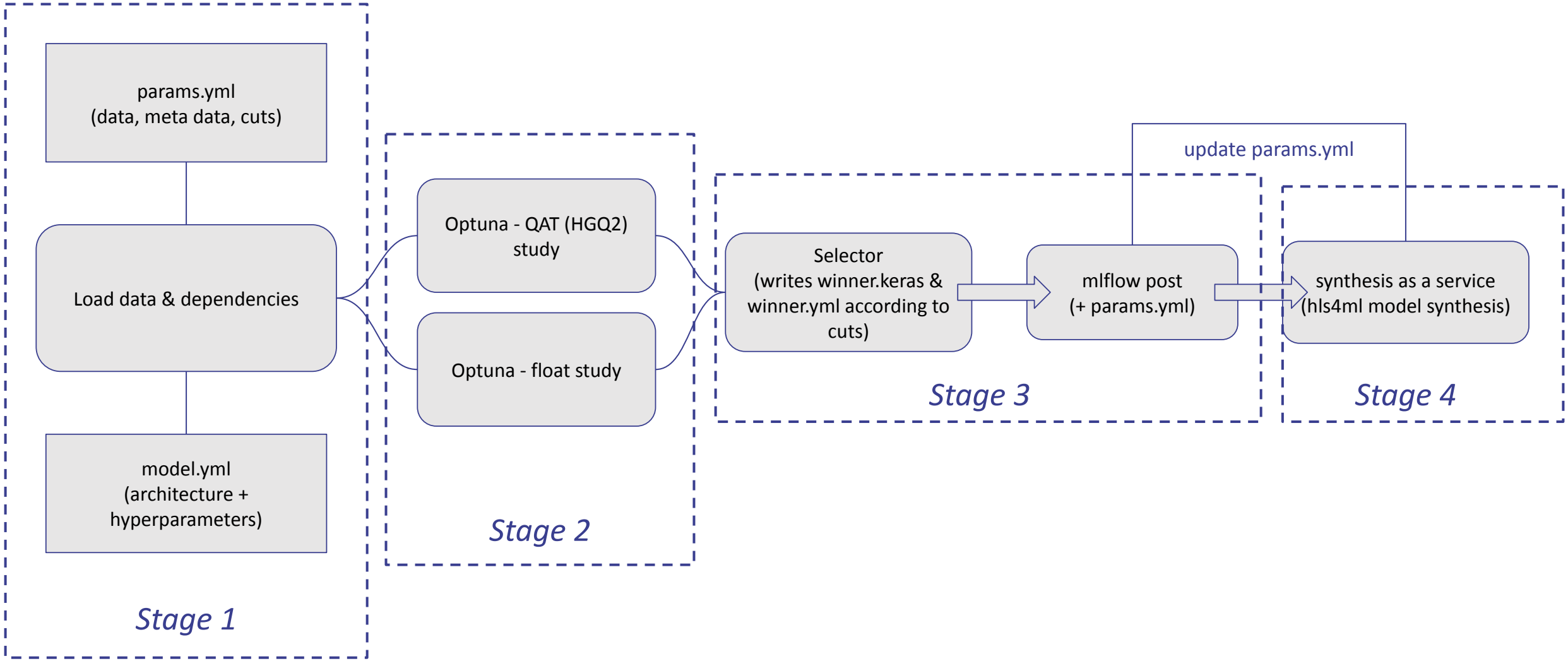
Pipeline Description



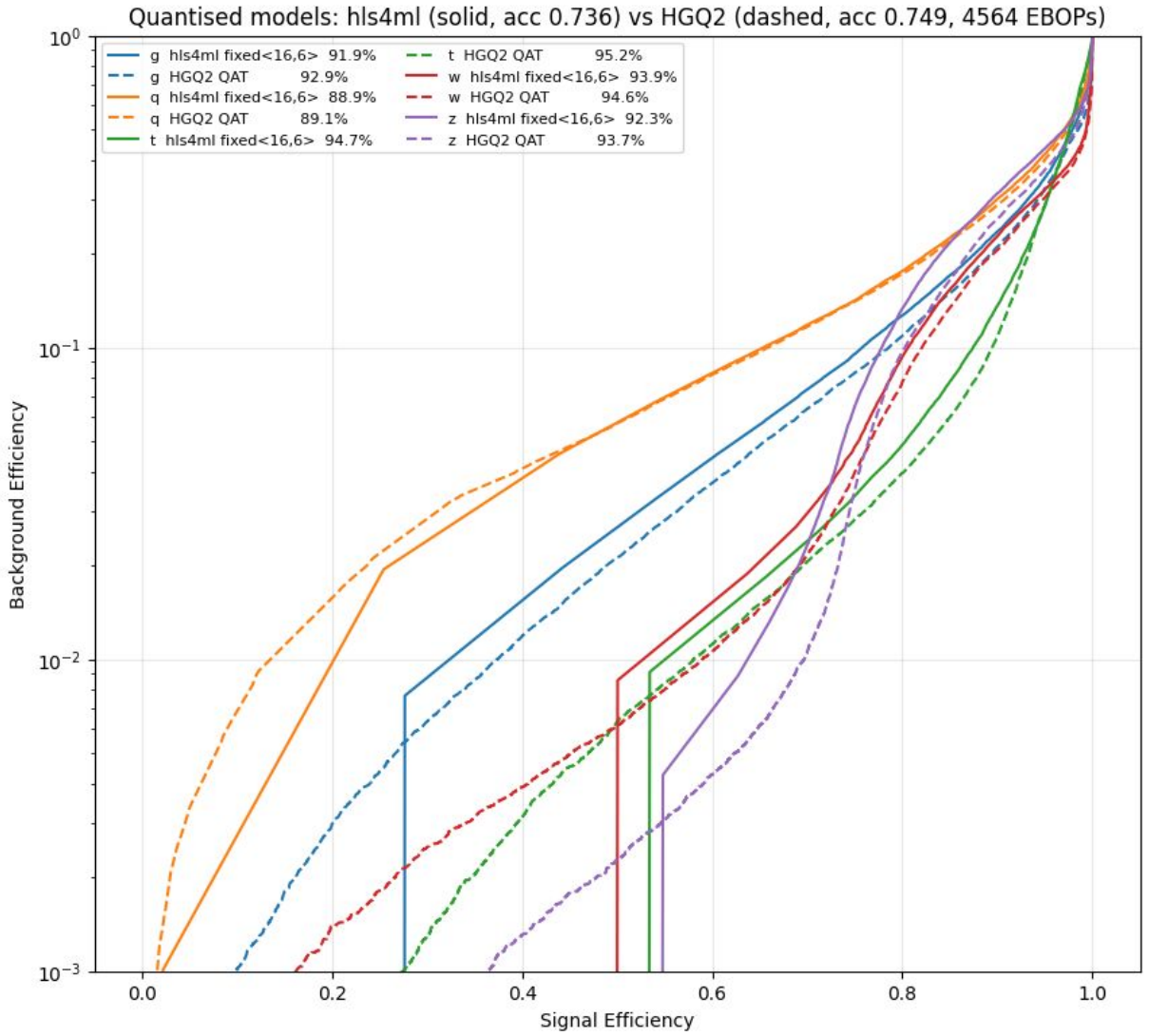
Pipeline Description



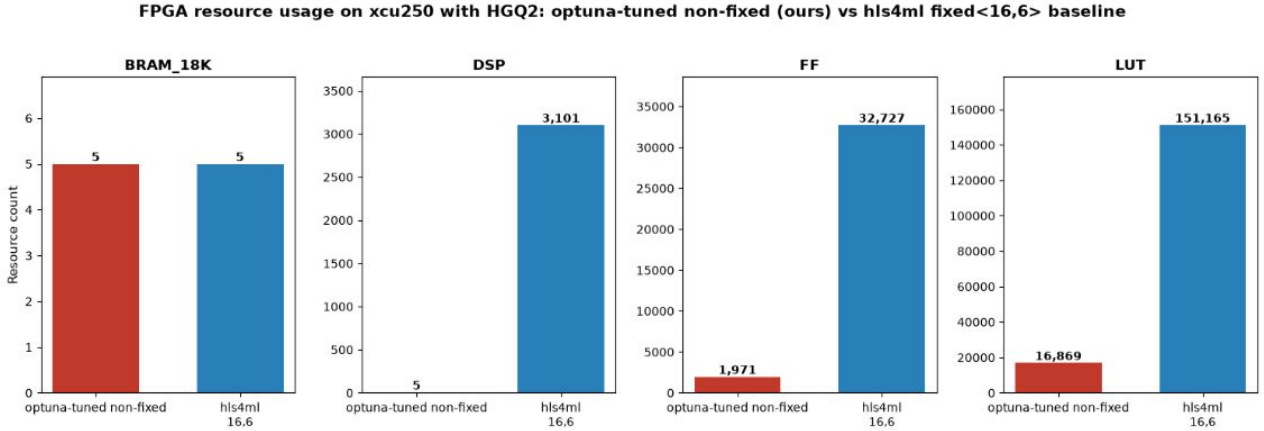
Pipeline Description



Performance



- Same MLP model as the hls4ml tutorial.
- QAT with HGQ2 and hyperparameter scanning with optuna, where model architectures are also treated as hyperparameters.
- Model building on the fly with wp21_ml_utils.
- Non-fixed optuna-tuned quantized model shows better performance & less recourse usage than the hls4ml fixed accuracy <16,6> quantized MLP.
- Resources and meta data update the param.yml file so that winning model can get re-built within the same infrastructure.



Thank you!

ML model optimization pipeline development

Stylianos Kostakis