Optimizing AI-based HEP algorithms using HPC and Quantum Computing

J.P. García Amboage*[1,2], E. Wulff[1], M. Girone[1], Tomás F. Pena[2]

[1]CERN, 1211 Geneva 23, Switzerland

[1]CoE RAISE, Wilhelm-Johnen-Straße, 52428 Jülich, Germany

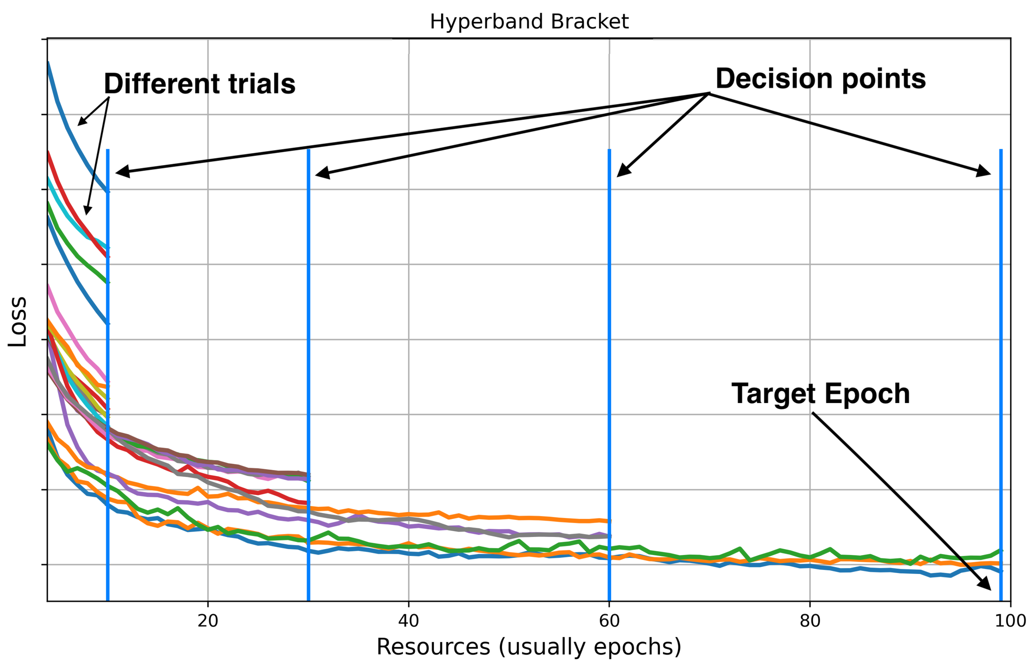
[2]CiTIUS, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain

*juan.pablo.garcia.amboage@cern.ch

1. Hyperparameter Optimization

Hyperparameter Optimization (HPO) can be used to systematically explore the search space of hyperparameter configurations of Deep Learning (DL) models.

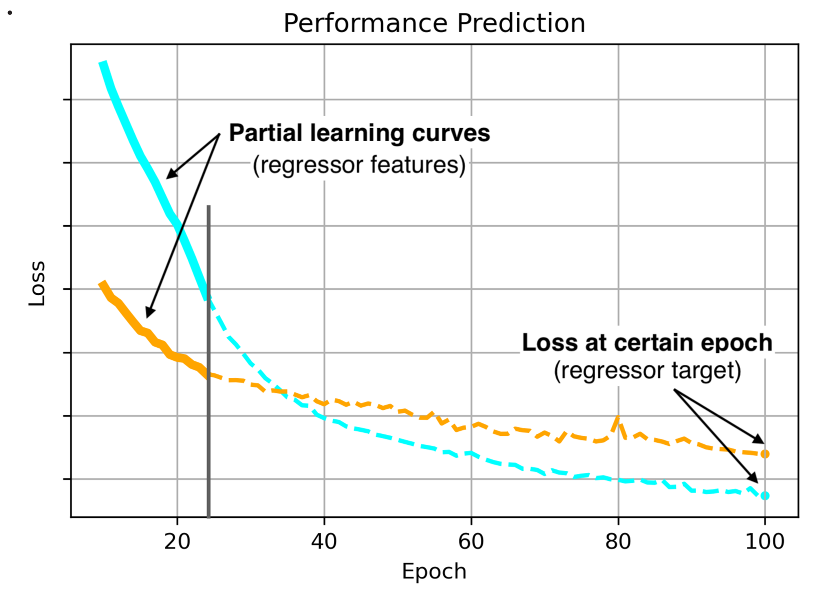
Current state-of-the-art HPO algorithms such as Hyperband [1], ASHA [2], and BOHB [3], rely on a method of early termination. Badly performing trials are automatically terminated allocating more computing resources to more promising ones.



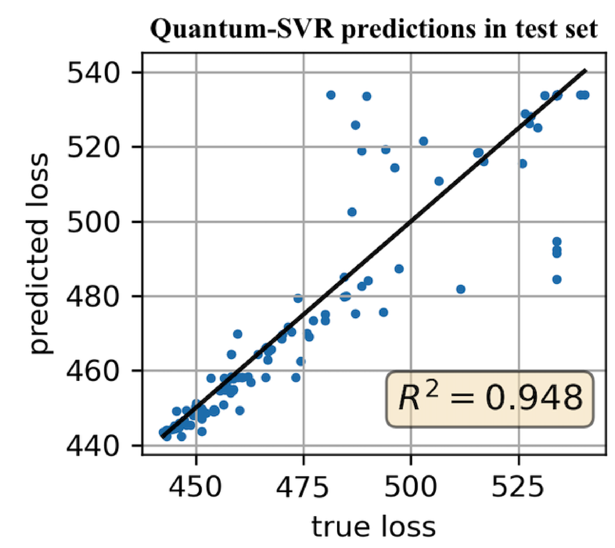
Such methods have been successfully applied to optimize MLPF, a particle flow reconstruction neural network [4]. The validation loss of MLPF was reduced by ∼44% [5].

2. Performance Prediction using Quantum-SVR

Support vector regression (SVR) models can be used to predict the loss of several NN architectures based on their partial learning curves [6].



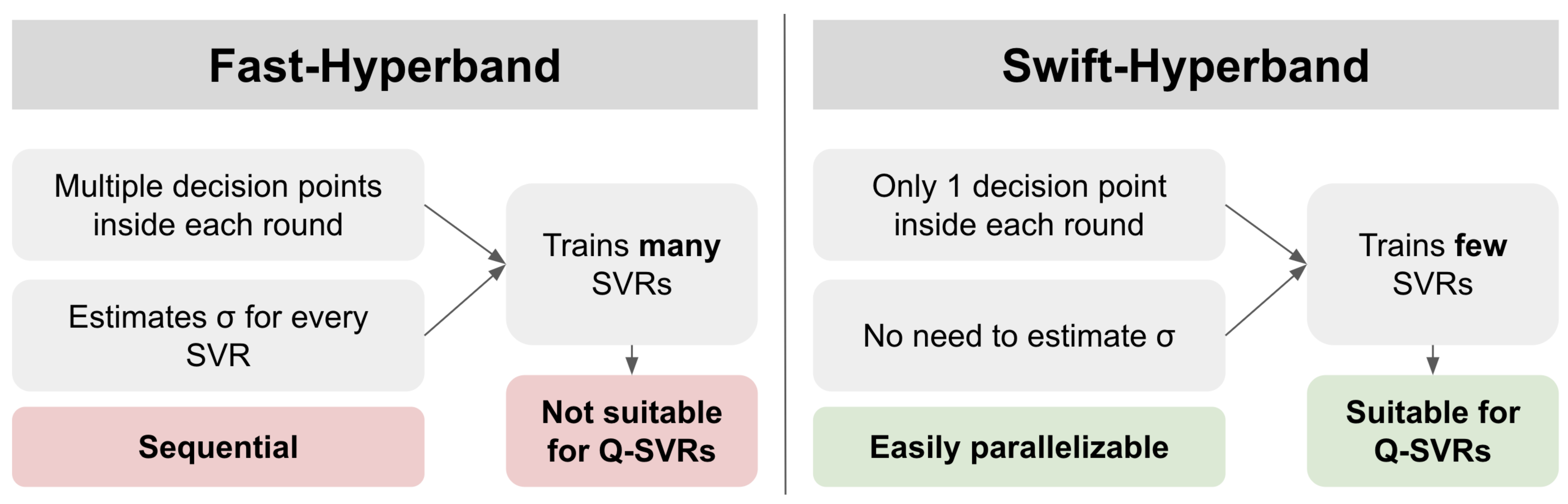
Via CoE RAISE, we accessed the Quantum Annealer at the Jülich Supercomputer Centre, to train Q-SVR models [7] on MLPF learning curves [8].

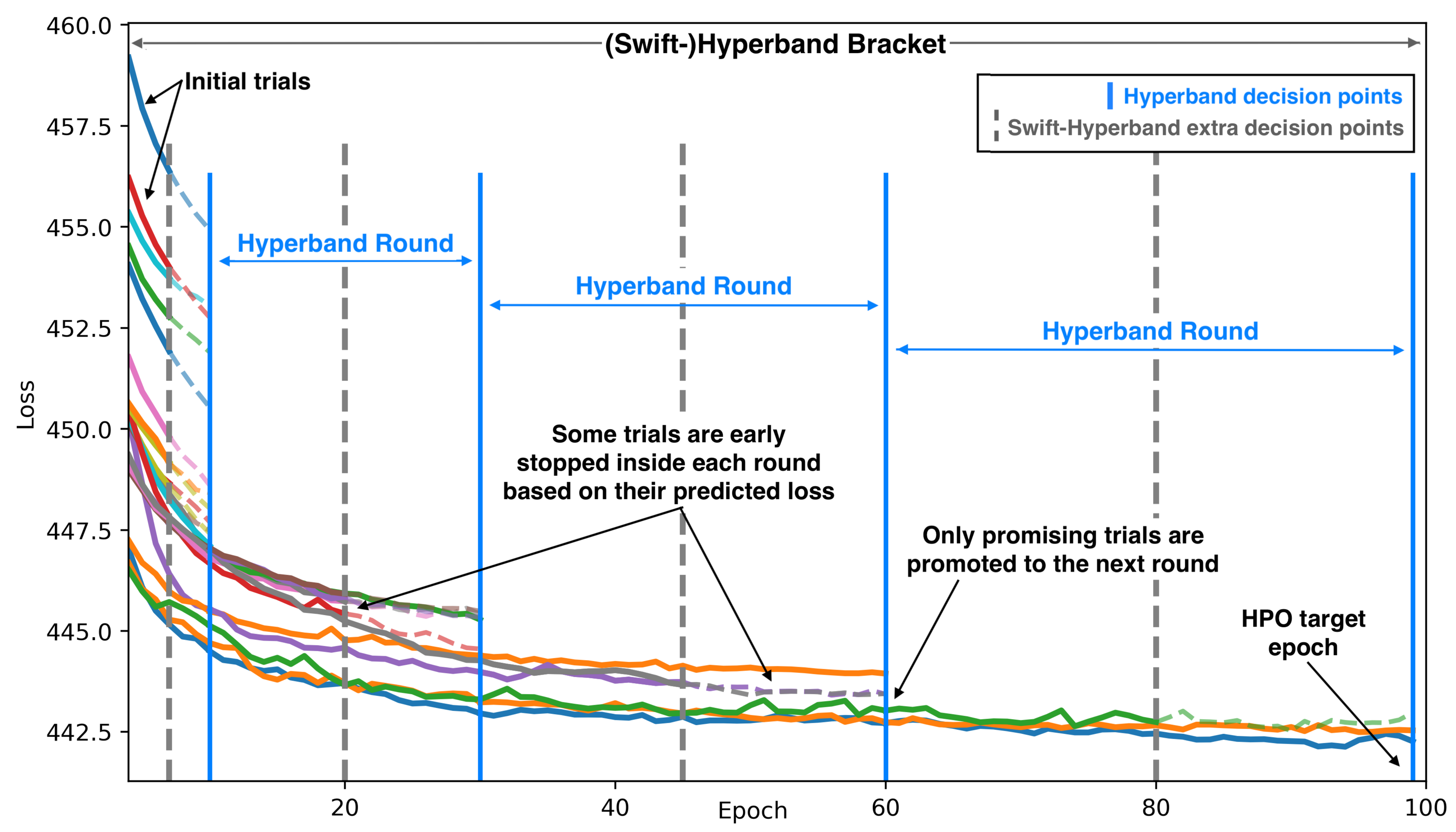


3. Using Performance Prediction to aid Hyperparameter Optimization

Baker et al. proposed the **sequential** algorithm Fast-Hyperband [6], a modified version of Hyperband that adds an additional decision point for every epoch inside each Hyperband round. Performance prediction is used for the extra decision points.

We propose Swift-Hyperband, a new way to integrate performance prediction with Hyperband. Our approach requires training far fewer performance predictors than Fast-Hyperband and is also easily **parallelizable**. Multiple trainings can be carried out simultaneously on different nodes within a round. As a result, **Swift-Hyperband has the potential to use Q-SVRs and benefit from HPC environments**.

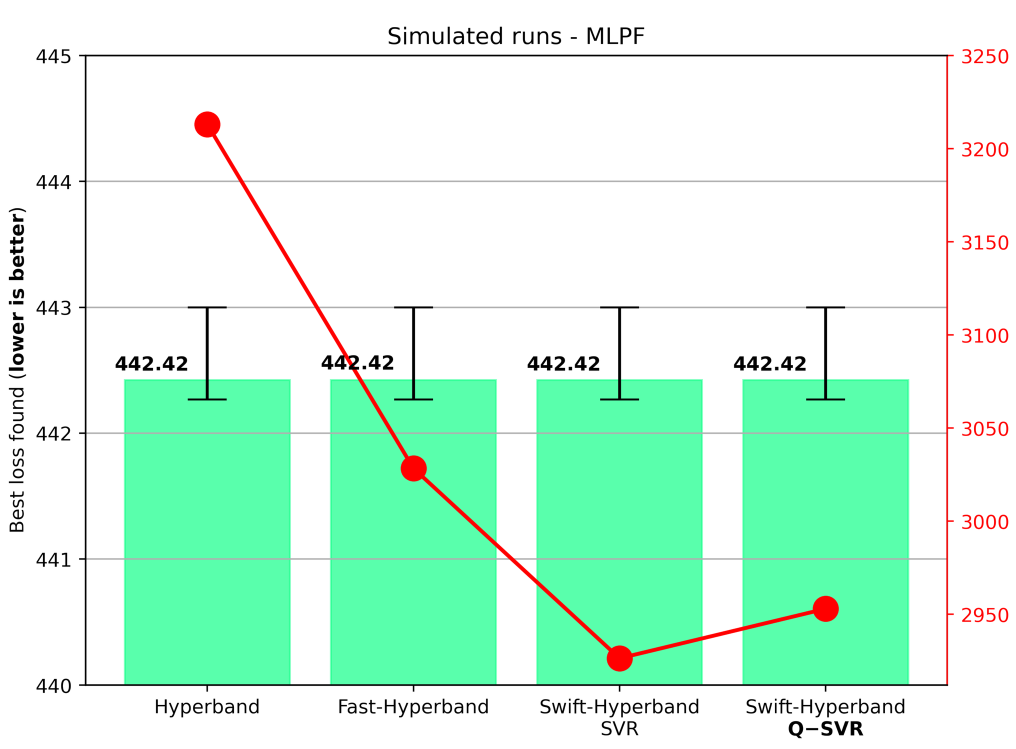


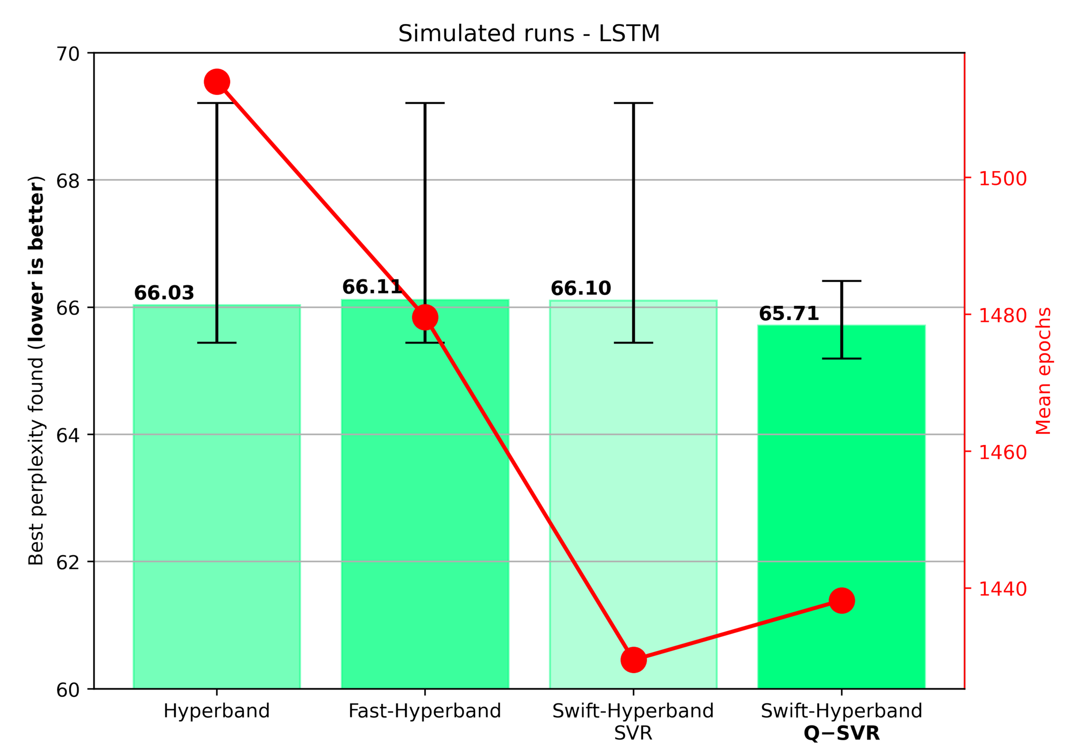


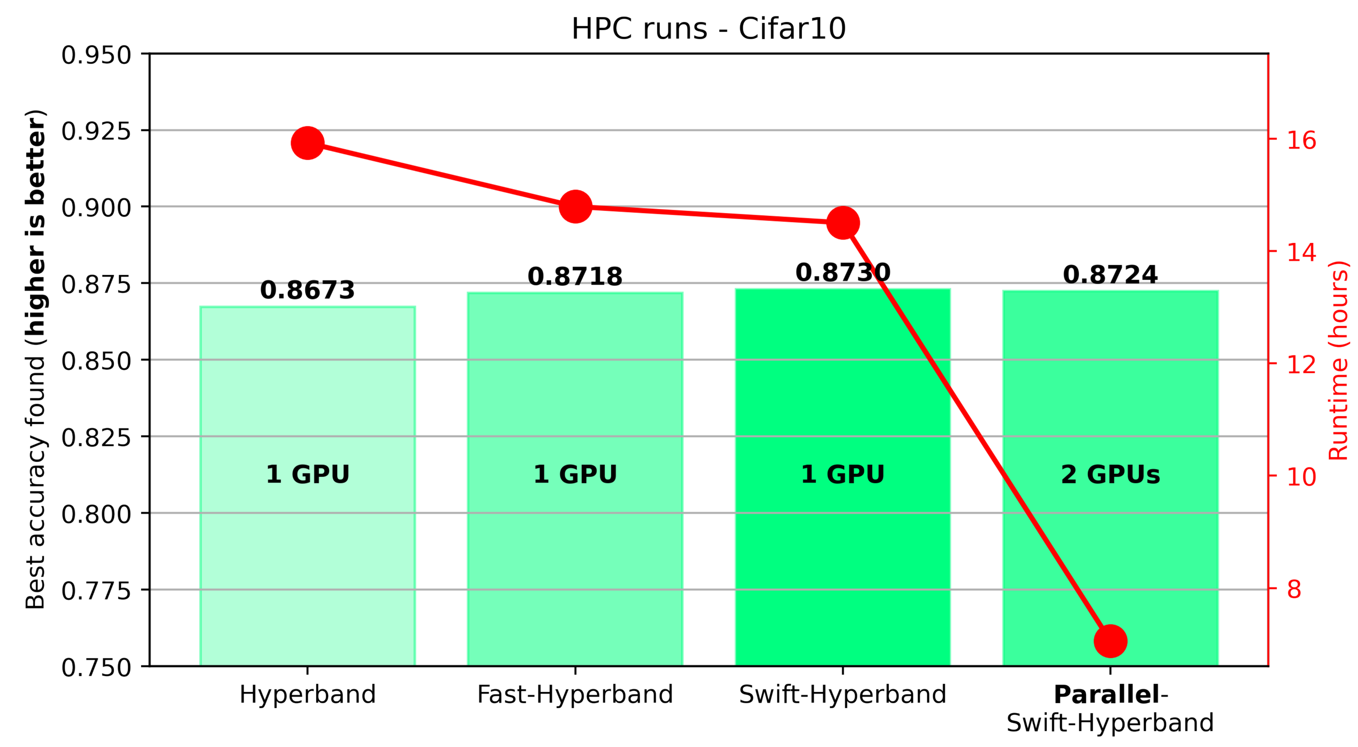
4. Results

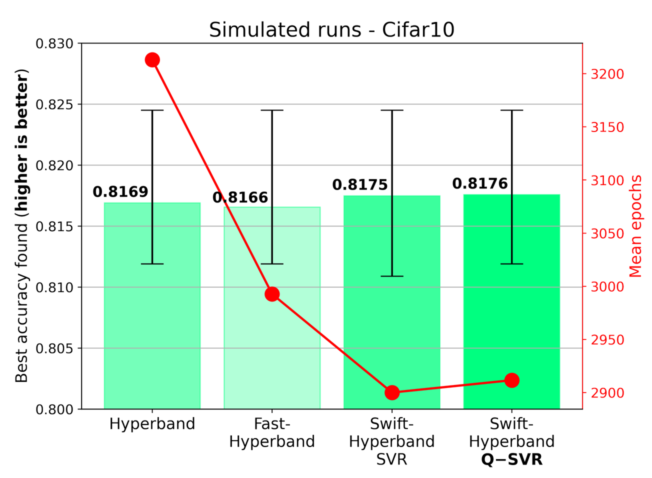
To **compare Hyperband, Fast-hyperband, Swift-Hyperband and Quantum-Swift-Hyperband** (Swift-Hyperband using QSVRs) for different NN architectures we **simulate 10 runs of each algorithm** using existing datasets of learning curves for MLPF [8] and other NNs: an image recognition CNN for Cifar10 modified from [2], an image recognition CNN for SVHN used in [6] and a NLP LSTM for PTB used in [6] .

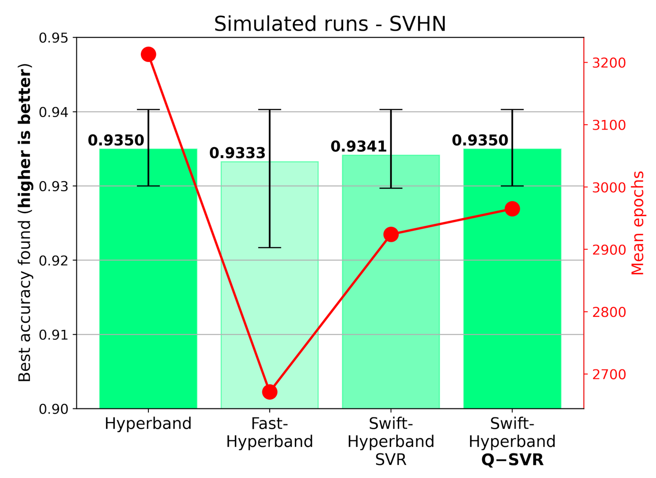
To test the speedup provided by the parallelization and the achieved accuracies we ran Hyperband, Fast-Hyperband, Swift-Hyperband and a parallel version of Swift-Hyperband (using MPI to coordinate one CPU node and two GPU worker nodes). The HPO target was a simple 6-layer CNN (different to the CNN used in the simulated runs) trained on Cifar10 using a 3-dimensional search space consisting of learning rate, weight decay and dropout. This network was chosen because it was relatively fast to train.











Learning Curves Datasets Description

Neural Network	Metric	Search space dimension	Target epoch for HPO
MLPF for Delphes	Loss	7	100
CNN for Cifar10	Accuracy	5	100
CNN for SVHN	Accuracy	9	100
LSTM for PTB	Perplexity	2	60

Both Swift-Hyperband and its version using Q-SVRs achieve accuracies comparable to classical Hyperband while needing considerably fewer epochs in all cases. In comparison to Fast-Hyperband, Swift-Hyperband (SVR and Q-SVR) is faster in all cases except on the SVHN problem. When it comes to the non-simulated runs we observe that all algorithms achieve accuracies around 87%, with both Swift-Hyperband and Parallel-Swift-Hyperband slightly beating Fast-Hyperband.

5. Conclusions

→ **We proposed a new** promising **parallelizable HPO algorithm** integrating Hyperband and performance predictors **that can be used in combination with Q-SVRs**. This leaves the door open for the use of Swift-Hyperband in later hyperparameter optimization cycles of MLPF.

→ **It was shown that**, despite the current limitations of quantum computers, **it is possible to integrate hybrid Quantum/HPC workflows for HPO**.

→ **There is a need for further studies on the speedup** achieved by the parallelization of Swift-Hyperband **when using a greater number of nodes**.

6. References

[1] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 18(1):6765–6816, 2017.

[2] L. Li, K. Jamieson, A. Rostamizadeh, E. Gonina, M. Hardt, B. Recht, and A. Talwalkar. Massively parallel hyperparameter tuning. *CoRR*, abs/1810.05934, 2018.

[3] S. Falkner, A. Klein, and F. Hutter. BOHB: Robust and efficient hyperparameter optimization at scale. In *International Conference on Machine Learning*, pages 1437–1446. PMLR, 2018.

[4] J. Pata, J. Duarte, JR. Vlimant, M. Pierini, and M. Spiropulu. MLPF: efficient machine-learned particle-flow reconstruction using graph neural networks. *The European Physical Journal C*, 81(5), may 2021.

[5] E. Wulff, M. Girone, and J. Pata. Hyperparameter optimization of data-driven AI models on HPC systems, 2022.

[6] Bowen Baker, Otkrist Gupta, Ramesh Raskar, and Nikhil Naik. Accelerating neural architecture search using performance prediction, 2017.

[7] Edoardo Pasetto, Morris Riedel, Farid Melgani, Kristel Michielsen, and Gabriele Cavallaro. Quantum SVR for chlorophyll concentration estimation in water with remote sensing. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[8] Eric Wulff, Maria Girone, David Southwick, Juan Pablo García Amboage, and Eduard Cuba. Hyperparameter optimization, quantum-assisted model performance prediction, and benchmarking of ai-based high energy physics workloads using hpc, 2023.