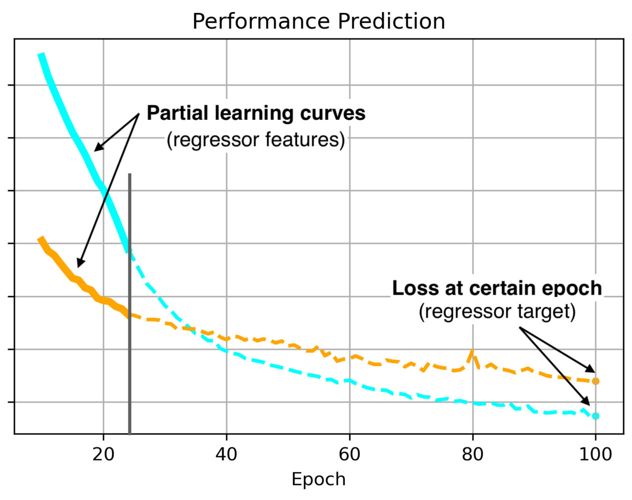
Distributed Hybrid Quantum-Classical Performance Prediction for Hyperparameter Optimization

J.P. García Amboage[†,*,1], E. Wulff[*,1], M. Aach[*,2,3], M. Girone[1], A. Delilbasic[2,3], E. Pasetto[2], R. Sarma[2], M. Riedel[2,3], A. Lintermann[2]

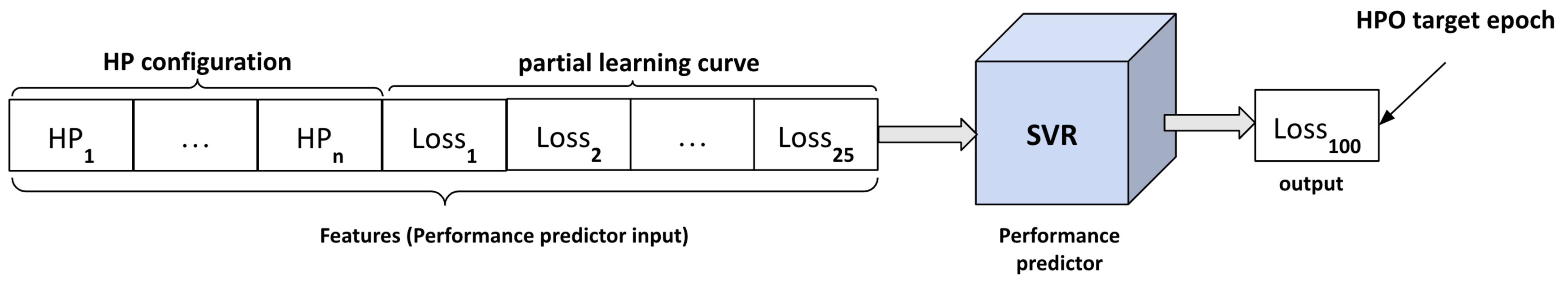
[†] jugarcia@student.ethz.ch
[*] Authors contributed equally to this work
[1] CERN, The European Organization for Nuclear Research
[2] Jülich Supercomputing Centre
[3] University of Iceland

1. Introduction

- The performance of neural networks is highly sensitive to the choice of Hyperparameters (HPs).
- Popular HP optimization (HPO) algorithms such as Hyperband [1] or ASHA [2] rely on early termination, where under-performing trials are automatically terminated to free up compute resources for more promising trials.
- Since the training process is non-linear, the ranking of trials at one point during the training does not necessarily hold at a later point.

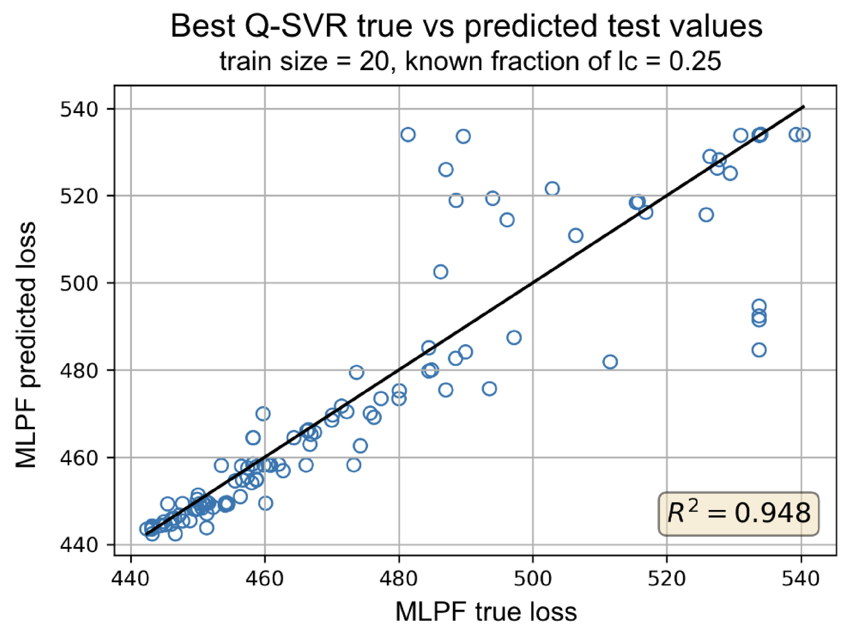


- A potential solution to this problem is to use a model performance predictor such as an SVR to predict future model performance improvements, from a partially trained model learning curve [3].



2. Quantum SVR

- The Support Vector Regression (SVR) model can be modified to be trained on a Quantum Annealer by formulating it as a QUBO problem [4].



General QUBO problem formulation

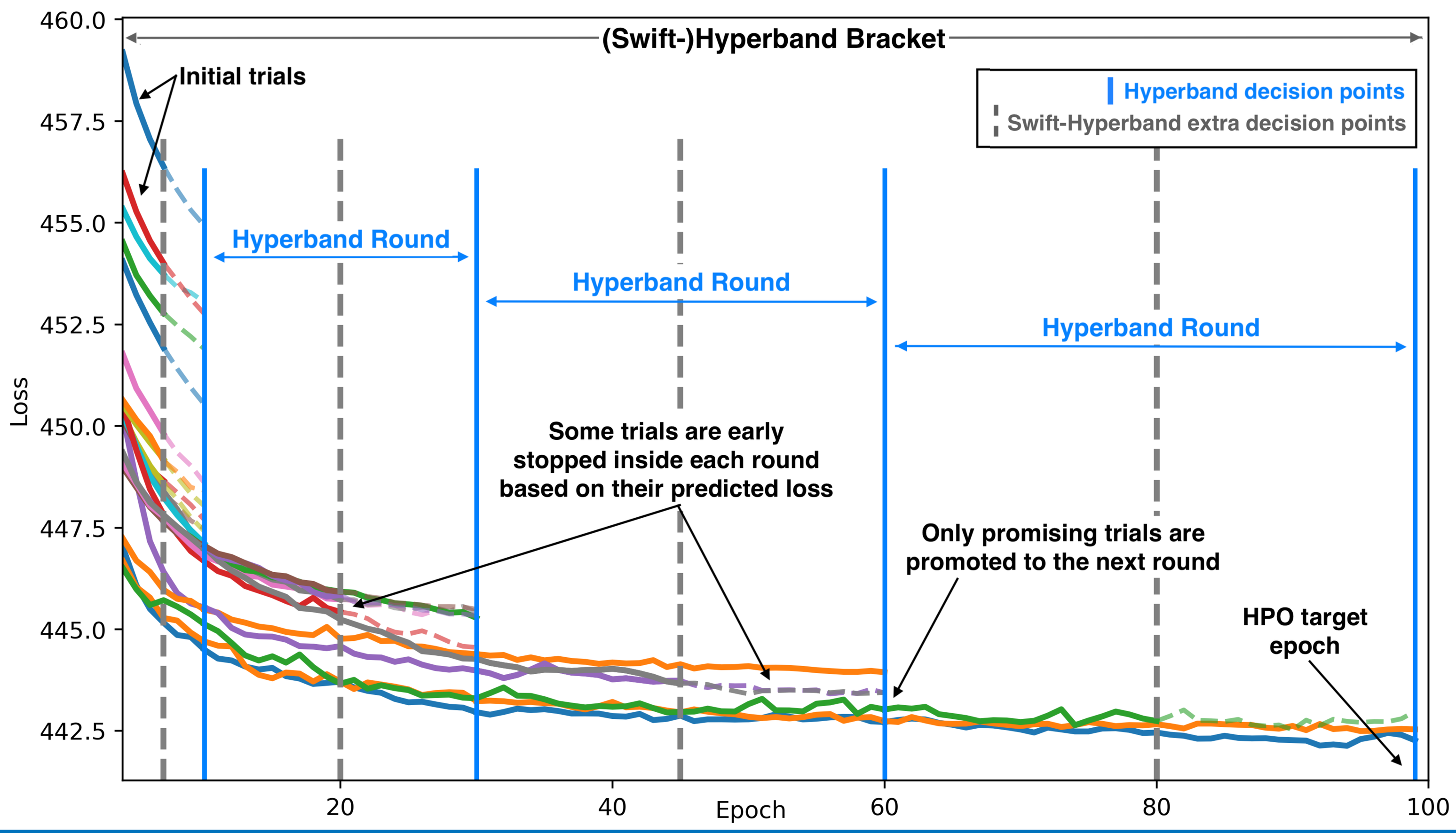
$$\text{Minimize: } f(x) = \sum_{i<j}^{N} Q_{i,j} x_i x_j + \sum_{i}^{N} Q_{i,i} x_i$$

$x_i \in \{0,1\}$ and Q is a $N \times N$ upper triangular matrix.

- The complexity order of training a classical SVR is cubic on the number of training samples. However, training the Q-SVR on the Quantum Annealer is linear in the number of training samples [5].
- Via CoE RAISE, the Quantum Annealer at the Jülich Supercomputer Centre was used to train and evaluate Q-SVR models for predicting the performance of MLPF [6], a particle flow reconstruction neural network.

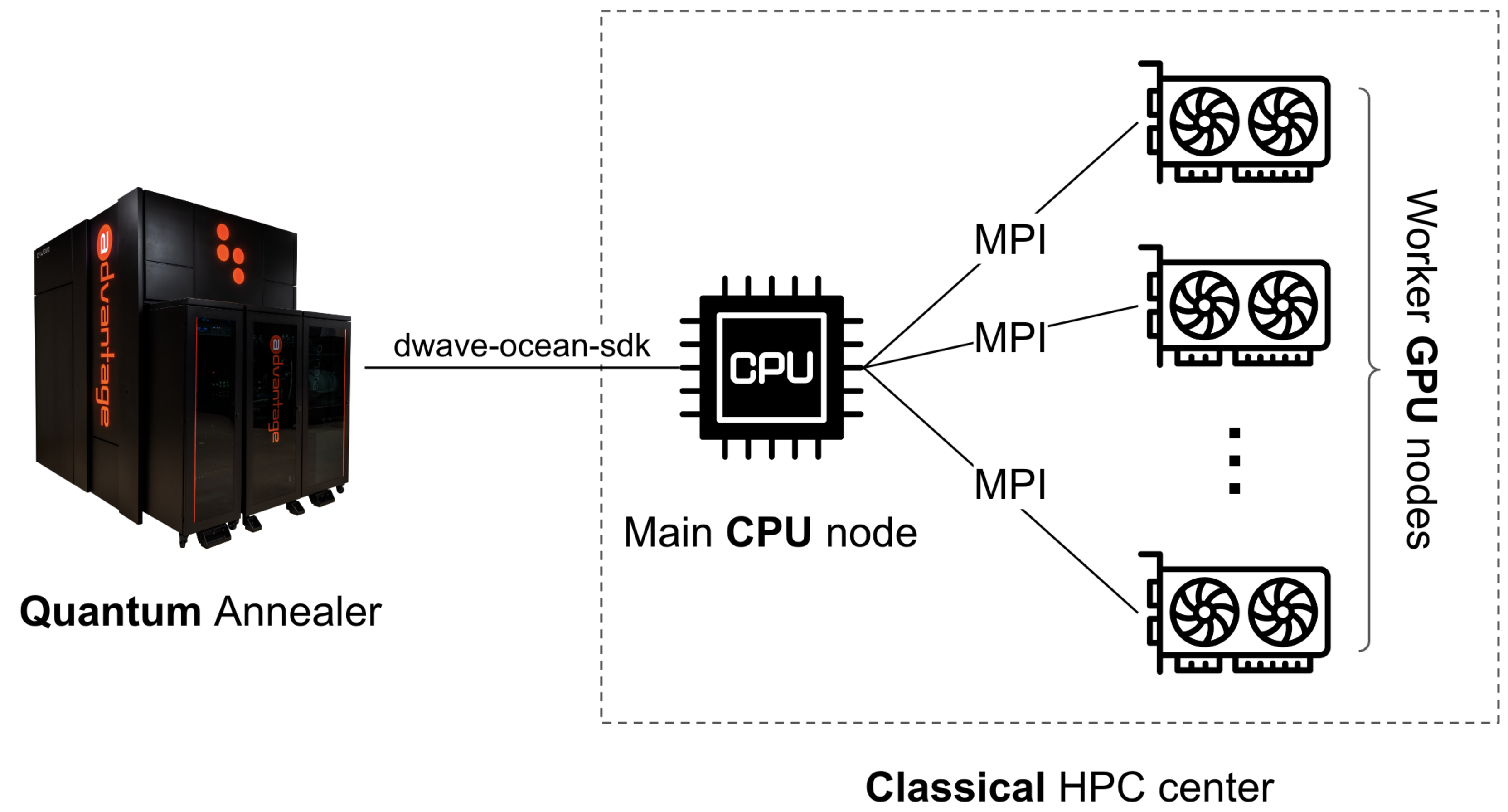
3. Swift-Hyperband

- Swift-Hyperband [7] and Fast-Hyperband [3] are enhanced versions of the HPO algorithm Hyperband that incorporate performance predictors at multiple decision points.
- Swift-Hyperband requires fewer performance predictors than Fast-Hyperband.
- Each round of Swift-Hyperband can be parallelized as the partial and full trainings can be executed at the same time.



4. Hybrid Quantum-Classical workflow

- We implemented a distributed version of Swift-Hyperband that uses multiple GPU nodes for training the target model and which connects to a Quantum Annealer for training the performance predictor.
- MPI and the *dwave-ocen-sdk* are used to handle communication among nodes and with the Quantum Annelaer.

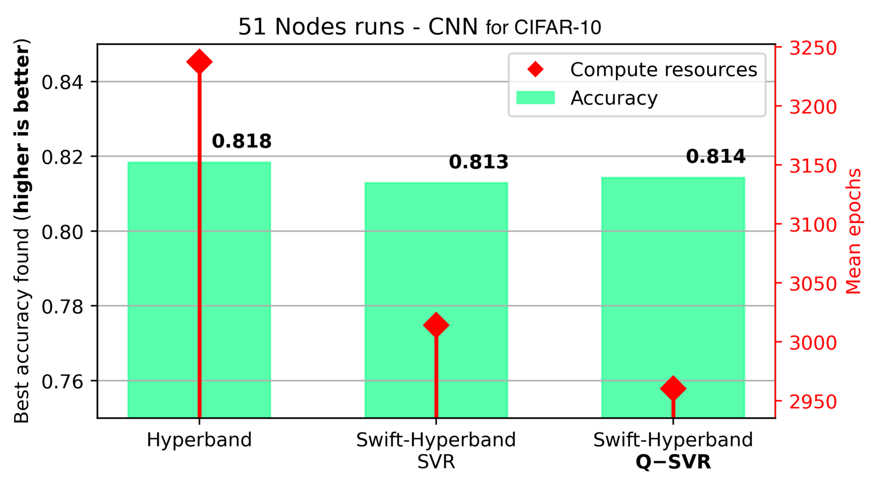


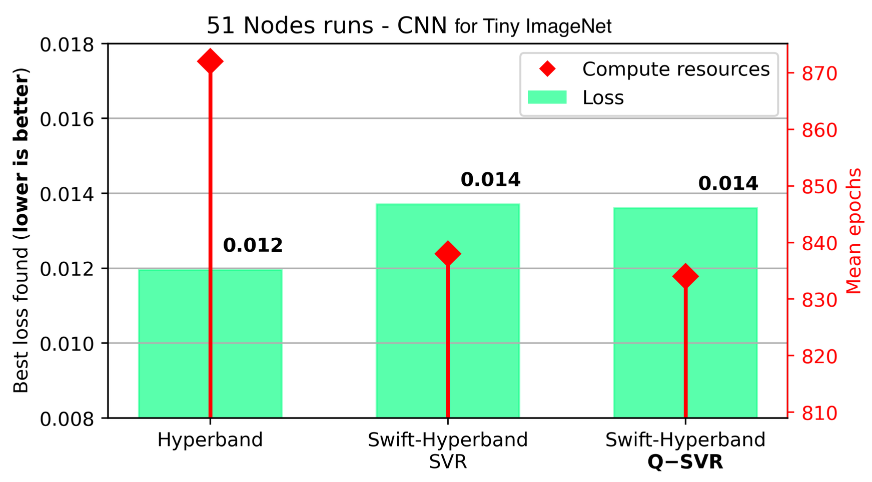
5. Results

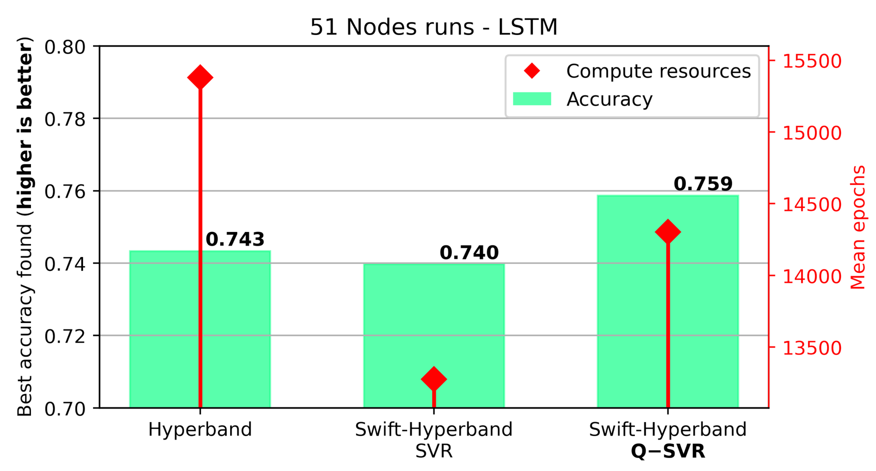
- We compare Swift-Hyperband using Q-SVR and classical SVR with Hyperband for several target models from different domains: computer vision (CV), natural language processing (NLP) and high energy physics (HEP).
- The results show that Swift-Hyperband, both in case of SVRs and Q-SVRs, achieves **similar results** as classical Hyperband in terms of the target model performance while consuming **less computational resources** in all cases.

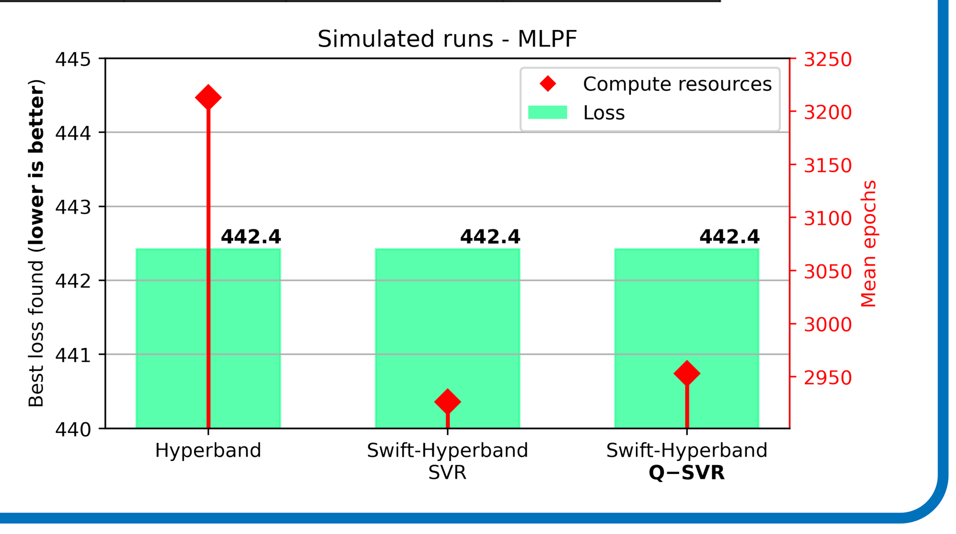
Summary of the benchmarking cases used to test the HPO algorithms

NN architecture	Dataset	Domain	Evaluation metric	# HPs for HPO	Target epoch for HPO	# GPU Nodes for HPO
CNN	CIFAR-10	CV	accuracy	5	100	50
CNN	Tiny ImageNet	CV	loss	3	35	50
LSTM	bABI, task 17	NLP	accuracy	4	300	50
MLPF	Delphes	HEP	loss	7	100	simulated









6. Conclusions

- The presented work along with the obtained results highlight the potential of hybrid Quantum-Classical machine learning algorithms in various fields of application.
- Further lines of work include integrating the presented quantum-classical workflow and performance prediction with other HPO algorithms and studying the applicability of Swift-Hyperband to even more neural network architectures and use cases.

7. 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] Bowen Baker, Otkrist Gupta, Ramesh Raskar, and Nikhil Naik. Accelerating neural architecture search using performance prediction, 2017.
[4] 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.
[5] Prasanna Date, Davis Arthur, and Lauren Pusey-Nazzaro. QUBO formulations for training machine learning models. *Scientific Reports*, 11(1), may 2021.
[6] 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.
[7] JP García Amboage, E Wulff, M Girone, and Tomás F Pena. Optimizing AI-based HEP algorithms using HPC and Quantum Computing, 2023.

- This work was performed in the Center of Excellence (CoE) Research on AI and Simulation-Based Engineering at Exascale (RAISE) receiving funding from EU's Horizon 2020 Research and Innovation Framework Programme H2020-INFRAEDI-2019-1 under grant agreement no. 951733.
- The authors gratefully acknowledge the Jülich Supercomputing Centre for funding this project by providing computing time on the D-Wave Advantage™ System JUPSI through the Jülich Unified Infrastructure for Quantum computing (JUNIQ).