E4 COMPUTER ENGINEERING

WHEN PERFORMANCE MATTERS

SmartHPC-QC: evaluating the impact of malleability for HPC-QC integration

Gabriella Bettonte, PhD

QRUCH Workshop – ISC 2025
June 13th, 2025

ACKNOWLEDGEMENTS

Project funded by the European Union

NextGenerationEU, ICSC National Centre, CN00000013, MUR Act n. 1031 - 17/06/2022

E4 COMPUTER ENGINEERING

E4 Computer Engineering designs and manufactures highly technological solutions for **HPC** Clusters, **Cloud**, **Data Analytics**, **Artificial Intelligence**, **Hyper-Converged infrastructure** and **Quantum Computing** for the Academic and Industrial markets. We have been collaborating for years with the main research centers at national and international level (CINECA, CERN, ECMWF, LEONARDO) and we are involved in national and **European projects** in the HPC, Quantum Computing, and AI fields.

Each E4 solution is **UNIQUE**, like every one of our customers; **TESTED** in every single component; **VALIDATED** to verify the actual performance of each system and **SERVICED** by technicians who provide assistance in the most extensive and complex Italian and European computing infrastructures.

E4 ANALYTICS

LET YOUR DATA PAY YOUR GROWTH

Through the sister company E4 Analytics, E4 works to integrate **Artificial Intelligence** and **Data Science** in organizations that undertake the **Digital Transformation** of their business to improve products/processes and optimize resources. We operate at the intersection between business and technology, supporting the customer in the adoption of customized and secure **GenAI solutions**: with E4 Analytics, company data become a **resource for creating value**, enhancing **innovation** and **competitiveness** in the marketplace.

QUANTUM COMPUTING ROAD MAP

2024

January
Organizer
Workshop HPC-QC
Munich
~100 Attendes

May
Submitted Proposal
Spoke 1 e 10
- Smart HPC-QC → **WON!**
- MoSeGaD → **WON!**

June
Submitted Proposal
Spoke 10
QuacK → **WON!**

September
Partner of:
Professional Master's
2nd LevelHPQC –
High-Performance and
Quantum Computing

November
EuroHPC
Quantum Center of Excellence
Project **WON!**

December
High Performance
Computing and
Quantum Computing
– Seventh Edition
1 Talk

2025

January 2025
Organizer
2nd edition
Workshop HPC-QC
& **2 Talk**
Barcelona

March 2025
Organizer
Industrial Workshop
Emerging Technologies in
HPC: The Rise of Quantum
Computing **& 1 talk**
Turin

April 2025
ARGUS Project **WON!**
Advanced Reconnaissance
and Geospatial Unified
Surveillance

HPC-QC MIDDLEWARE FOR A SMART USAGE OF COMPUTATIONAL RESOURCES

Goals:

- To **advance the state-of-the-art of quantum resource allocation and scheduling** by integrating QPUs with HPC resource schedulers (SLURM).

- To define HPC-QC interactions for different workload types.

STRONG LINKS WITH A WEALTH OF EUROPEAN AND NATIONAL PROJECTS

European-Driven HPCQC

EuroHPC Joint Undertaking Mandate: Integrate QC into HPC Centers

EuroHPC Joint Undertaking

Six centers selected for hosting QC systems

- Germany: LRZ
- Spain: BSC
- Czechia: IT4Innovations (LUMI-Q)
- Italy: Cineca
- Poland: PSNC
- France: Genci

OUR RECENT PUBLICATION

Assessing the Elephant in the Room in Scheduling for Current Hybrid HPC-QC Clusters

Paolo Viviani[1,*] Roberto Rocco[2], Matteo Barbieri[2], Gabriella Bettonte[2], Elisabetta Boella[2], Marco Cipollini[4], Jonathan Frassineti[3], Fulvio Ganz[2], Sara Marzella[3], Daniele Ottaviani[3], Simone Rizzo[2], Alberto Scionti[1], Chiara Vercellino[1], Giacomo Vitali[1,4], Olivier Terzo[1], Bartolomeo Montrucchio[4] and Daniele Gregori[2]

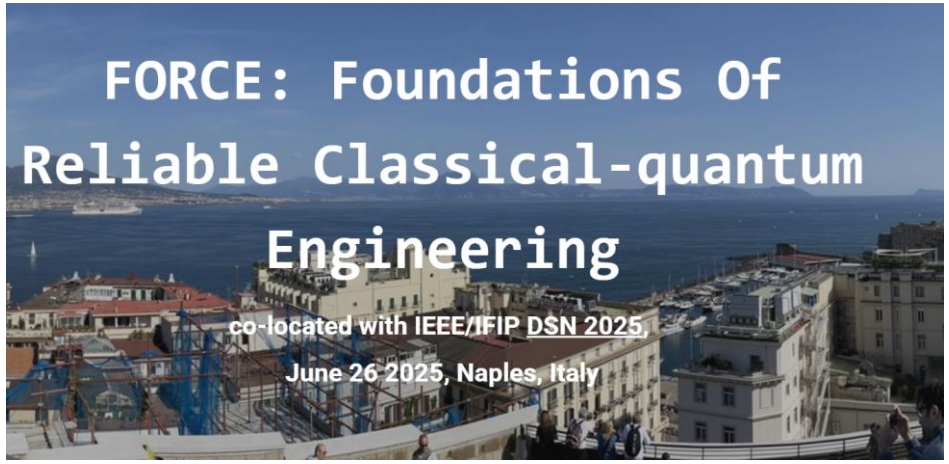
[1]*LINKS Foundation, Torino, Italy*
[2]*E4 Computer Engineering, Scandiano, Italy*
[3]*CINECA, Casalecchio di Reno, Italy*
[4]*Politecnico di Torino, Torino, Italy*

[*]*paolo.viviani@linksfoundation.com*

Abstract: Quantum computing resources are among the most promising candidates for extending the computational capabilities of High-Performance Computing (HPC) systems. As a result, HPC–quantum integration has become an increasingly active area of research. While much of the existing literature has focused on software stack integration and quantum circuit compilation, key challenges such as hybrid resource allocation and job scheduling—especially relevant in the current Noisy Intermediate-Scale Quantum era—have received less attention. In this work, we highlight these critical issues in the context of integrating quantum computers with operational HPC environments, taking into account the current maturity and heterogeneity of quantum technologies. We then propose a set of conceptual strategies aimed at addressing these challenges and paving the way for practical HPC-QC integration in the near future.

https://arxiv.org/pdf/2504.10520

ACCEPTED AT:



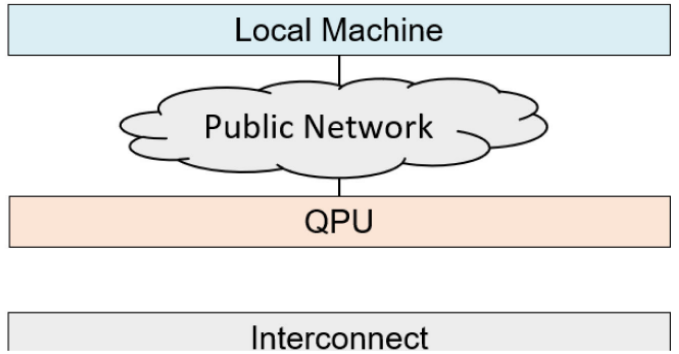
QPUS IN HPC TODAY: RESOURCE OR BOTTLENECK?

QPUs in the future:

- Many qubits
- Hopefully, fault tolerant
- Directly attached to CPUs with high-speed connections, similarly to GPUs
- One interface independent on technology

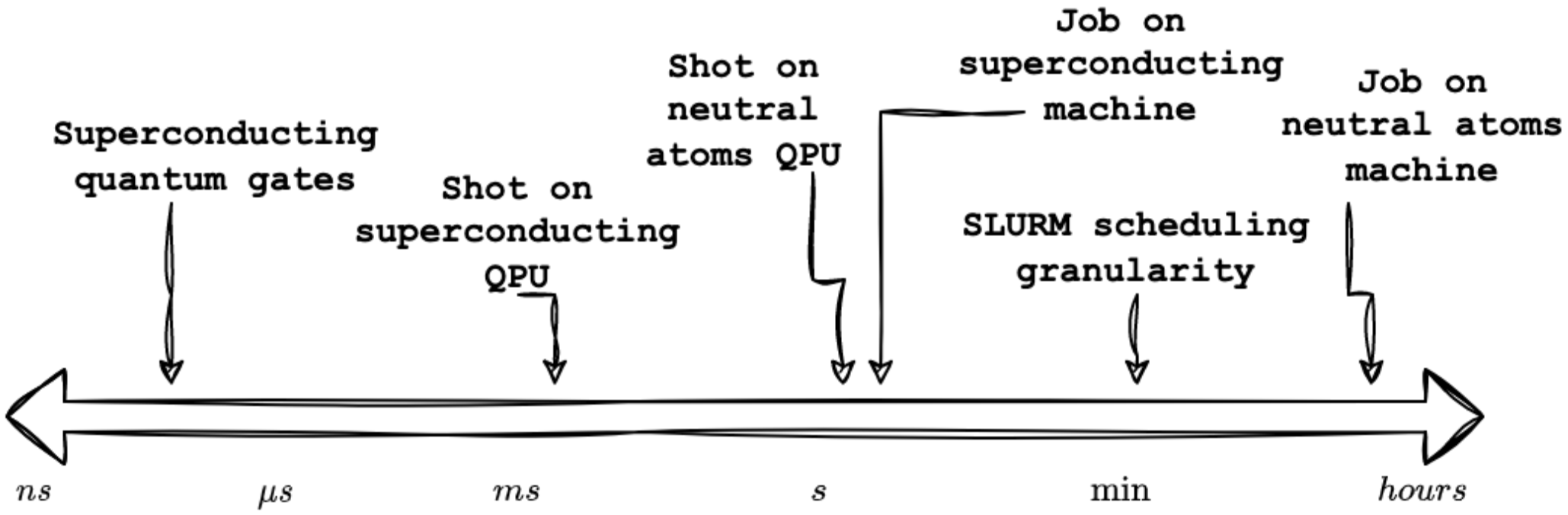
QPUs as of today:

- Limited amount of qubits
- Limited reliability, need for fault-handling mechanisms
- Noise-sensitive
- Attached via ethernet
- Every QPU has its own features and interfaces
- Small amount of quantum computers compared to the number of HPC nodes



Humble et al., IEEE Micro 41, 15 (2021).

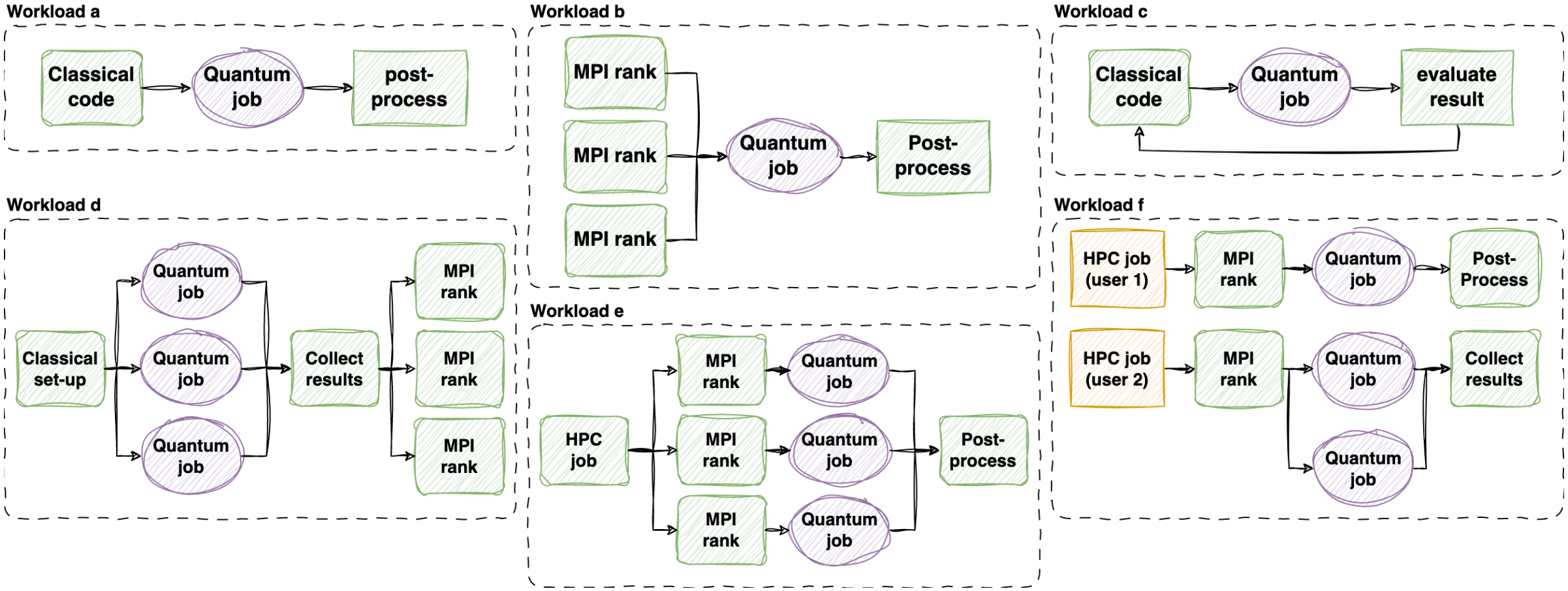
DIFFERENT QPUS HAVE DIFFERENT EXECUTION TIMES



© P. Viviani

Viviani et al., accepted at FORCE (2025).

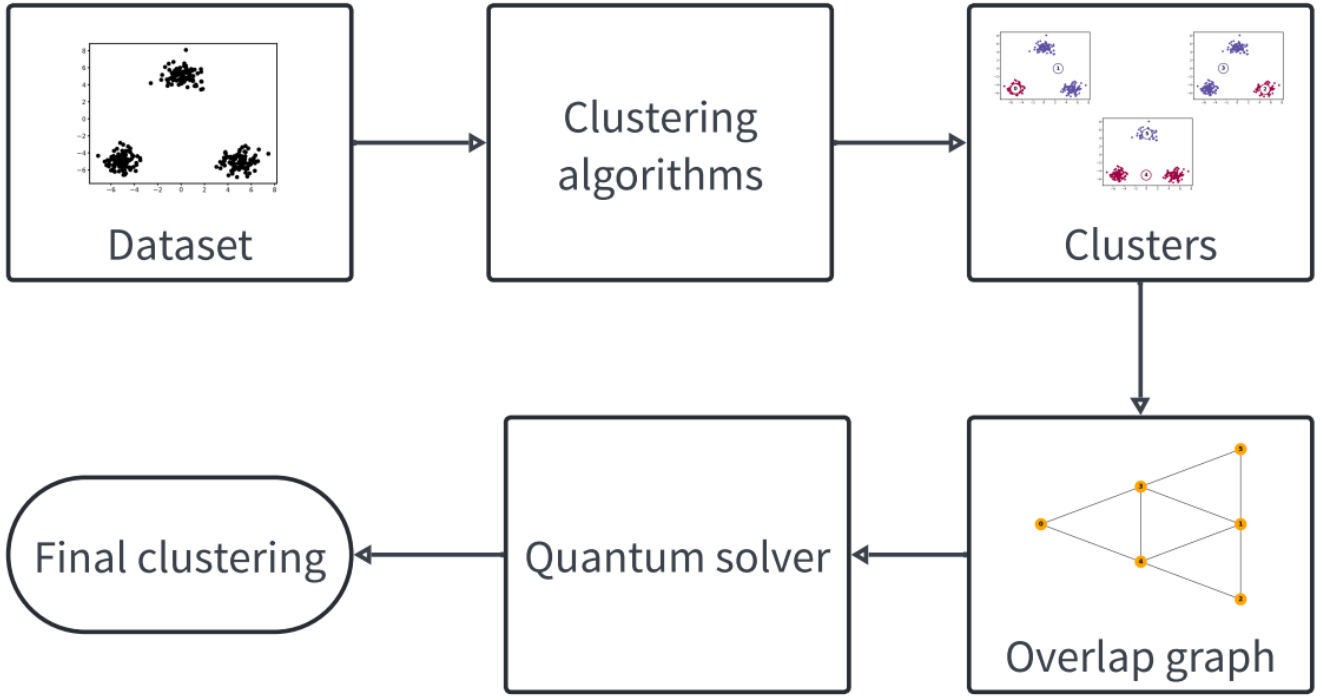
DIFFERENT HYBRID WORKLOADS CAN EXIST AND CO-EXIST



Viviani et al., accepted at FORCE (2025).

OUR USE CASE: CLUSTERING AGGREGATION

- Core idea: map aggregation of multiple clustering methods to a Quadratic Unconstrained Binary Optimization problem and solve it using a QPU
- Every algorithm has its pros and cons, the aggregation can improve results [1]
- Classic-Quantum approaches on quantum hardware already tested [2]

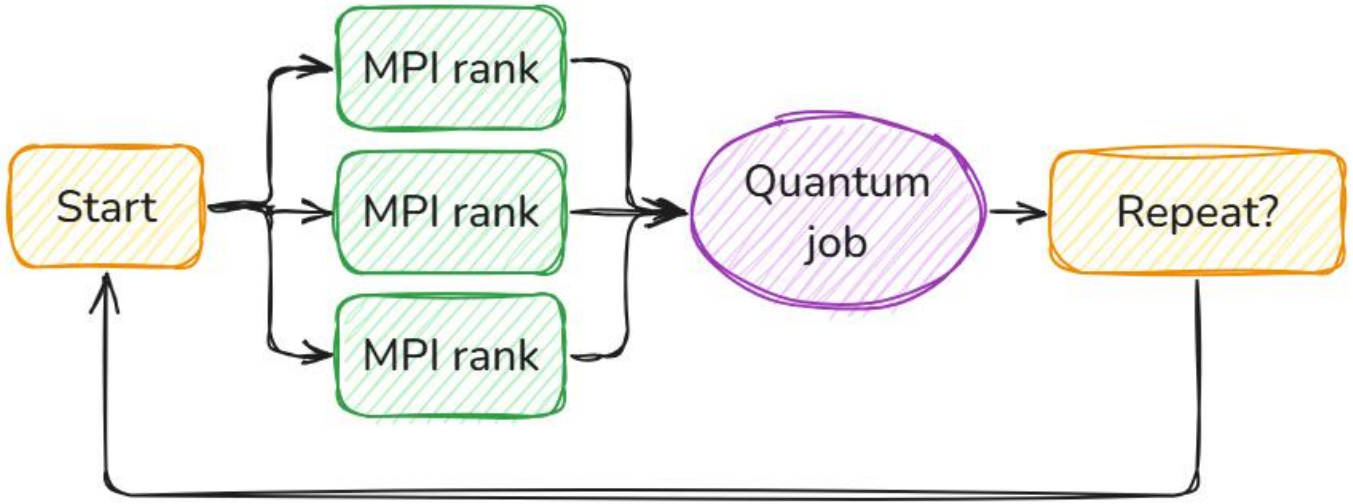


Scotti et al., arXiv [quant-ph] (2024).

[1] "Clustering Aggregation as Maximum-Weight Independent Set", Li et al., NIPS 2012,
[2] "A clustering aggregation algorithm on neutral-atoms and annealing quantum processors", Scotti et al., arXiv:2412.07558

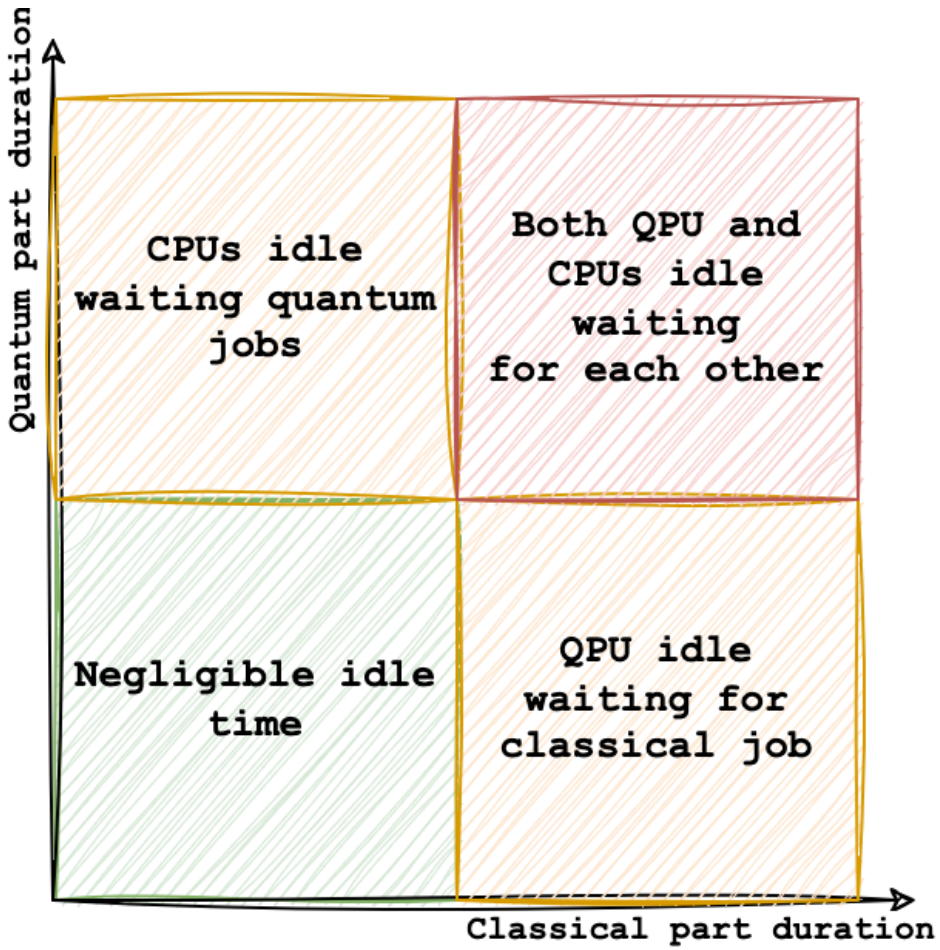
OUR USE CASE: QUANTUM OFFLOADING FROM A PARALLEL JOB

- Classical code runs on a SLURM compute partition
- Quantum code runs on a SLURM quantum partition (co-located QPU or emulator)
- Resources can be temporarily deallocated from one partition while the other partition is active



© S. Rizzo

DIFFERENCES IN CLASSICAL AND QUANTUM DURATIONS MATTER

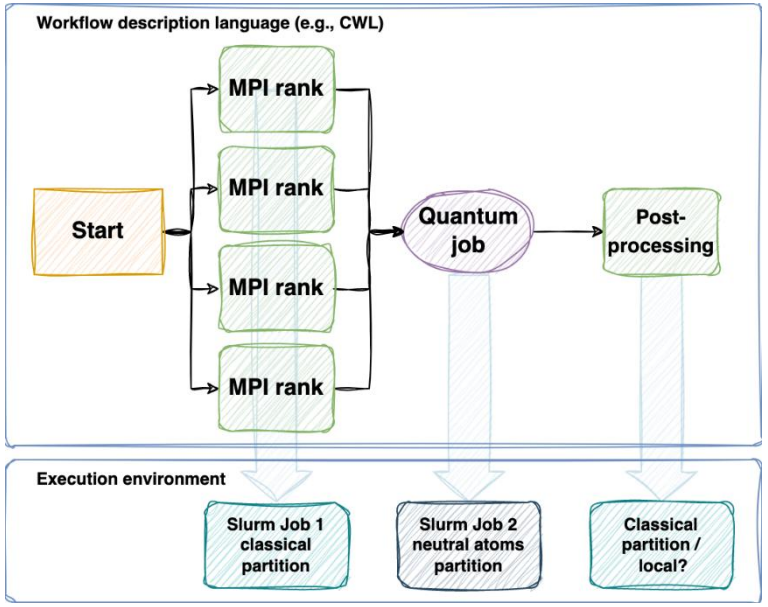


© P. Viviani

Viviani et al., accepted at FORCE (2025).

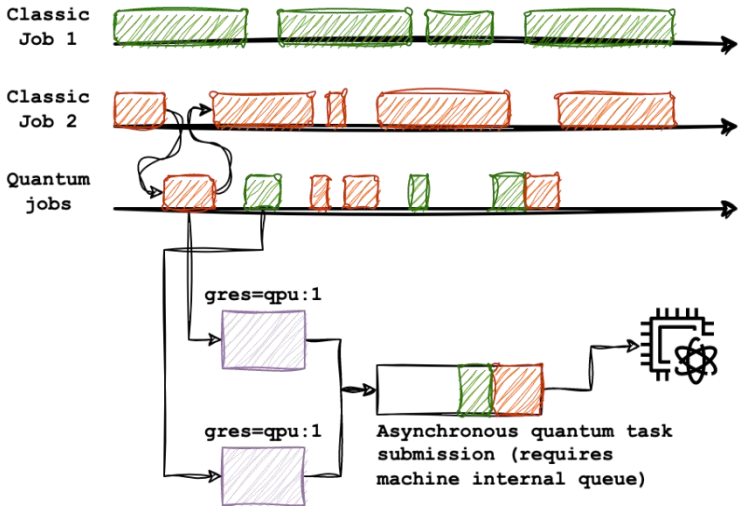
THREE POSSIBLE SOLUTIONS DEPENDING ON THE HPC-QC WORKLOAD

Workflow



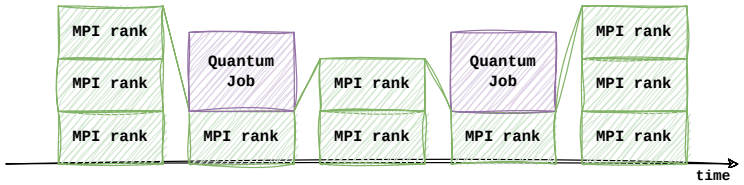
© P. Viviani

Virtual QPU



© P. Viviani

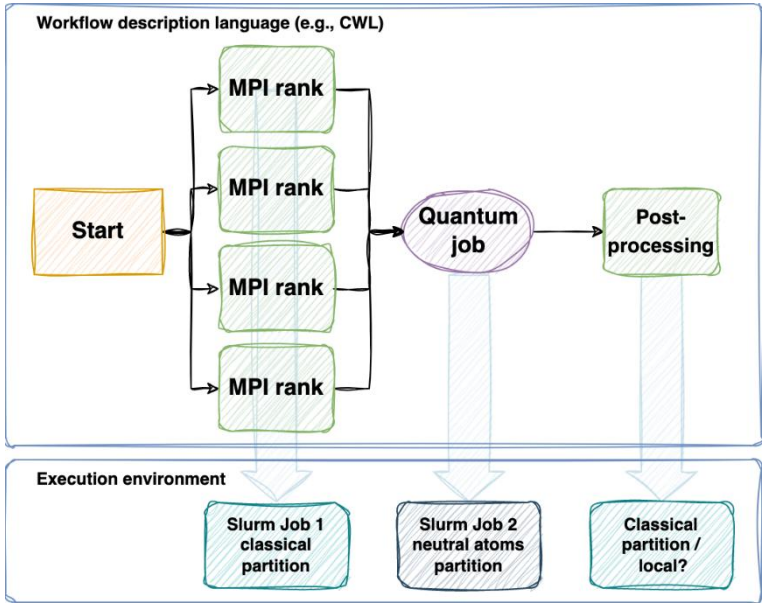
Malleability



© R. Rocco

Viviani et al., accepted at FORCE (2025).

WORKFLOW FOR LOOSELY COUPLED JOBS

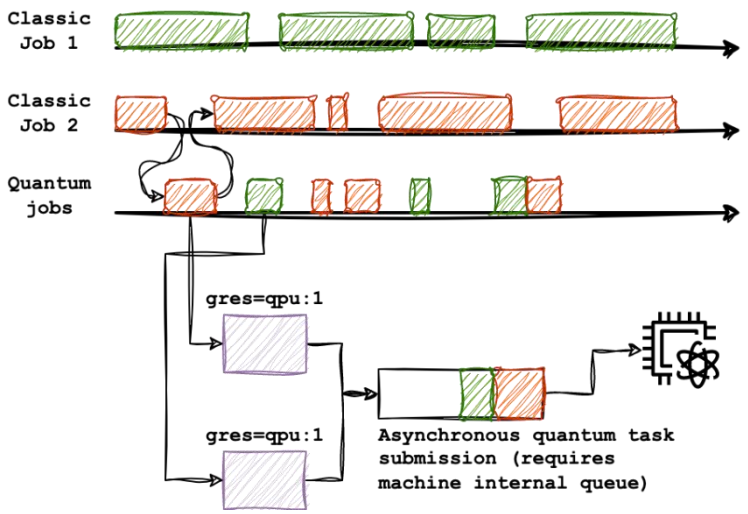


© P. Viviani

- Ideal when quantum portion of a hybrid job lasts long (e.g. > 30 min.)
- Quantum and classical jobs scheduled in an independent way, but with a single workflow
- Using workflow managers, such as StreamFlow
- QPU allocated by SLURM exclusively

Viviani et al., accepted at FORCE (2025).

VQPU WHEN CLASSICAL PART OF HYBRID JOB IS MUCH LONGER THAN QUANTUM PART

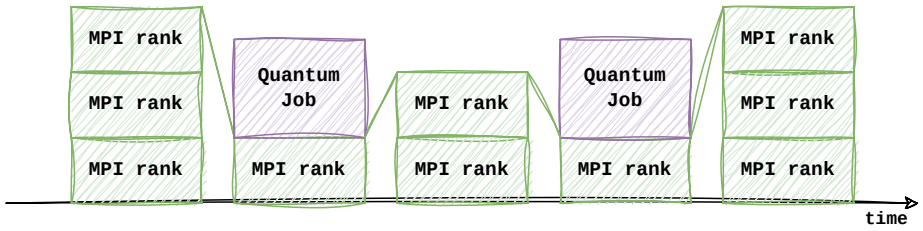


© P. Viviani

- Possible to allocate more QPUs than available with gres
- Internal QPU queue manages the quantum workload
- Maximum number of concurrent quantum/hybrid job submissions must be fixed
- Maximum waiting time for quantum job defined

Viviani et al., accepted at FORCE (2025).

MALLEABILITY FOR REDUCING CLASSICAL AND QUANTUM QUEUE TIMES

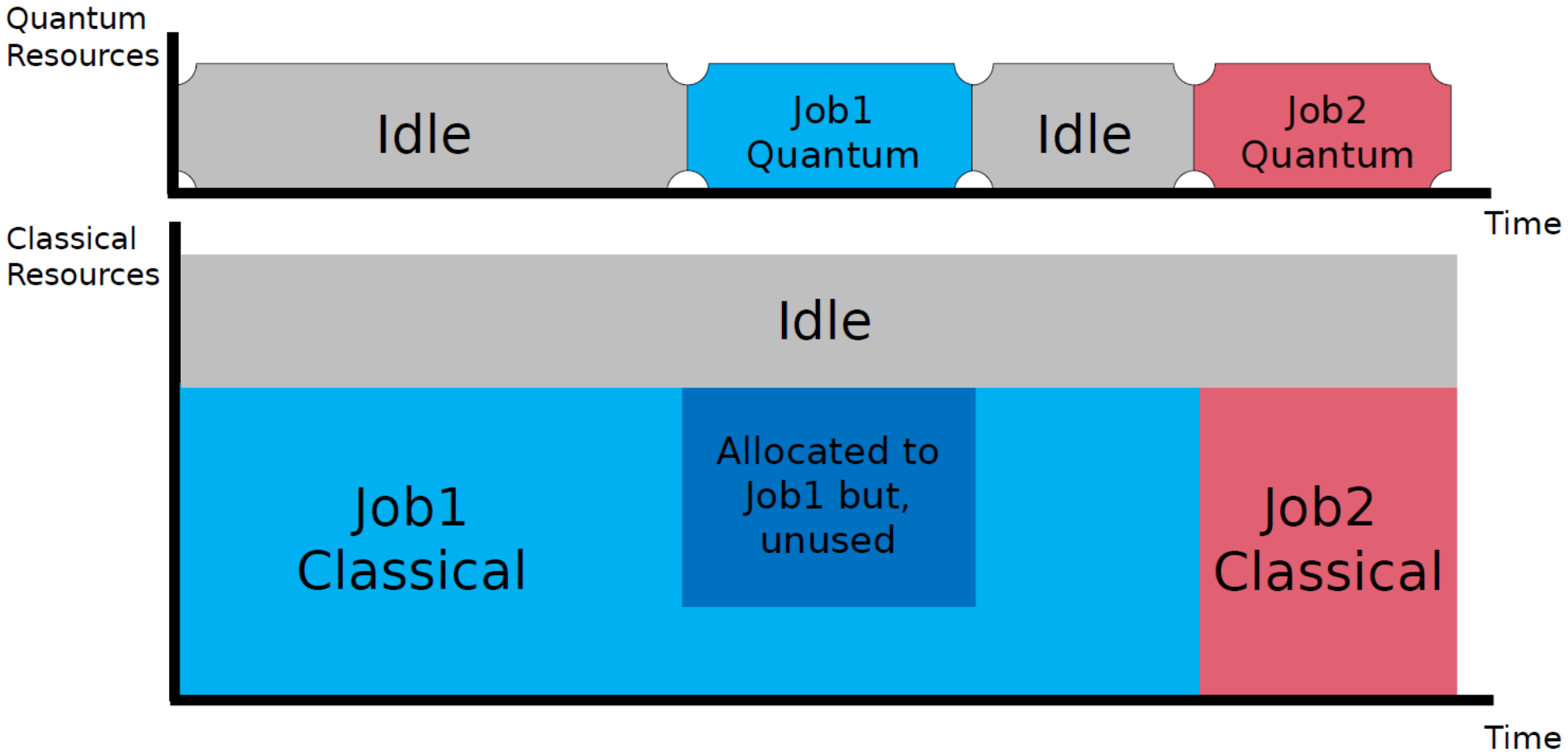


© R. Rocco

- Ideal when classical and quantum parts of a hybrid job have approximately same duration
- Allow for varying at runtime number of resources allocated for a specific job
- Could improve energy efficiency and allocation inefficiency

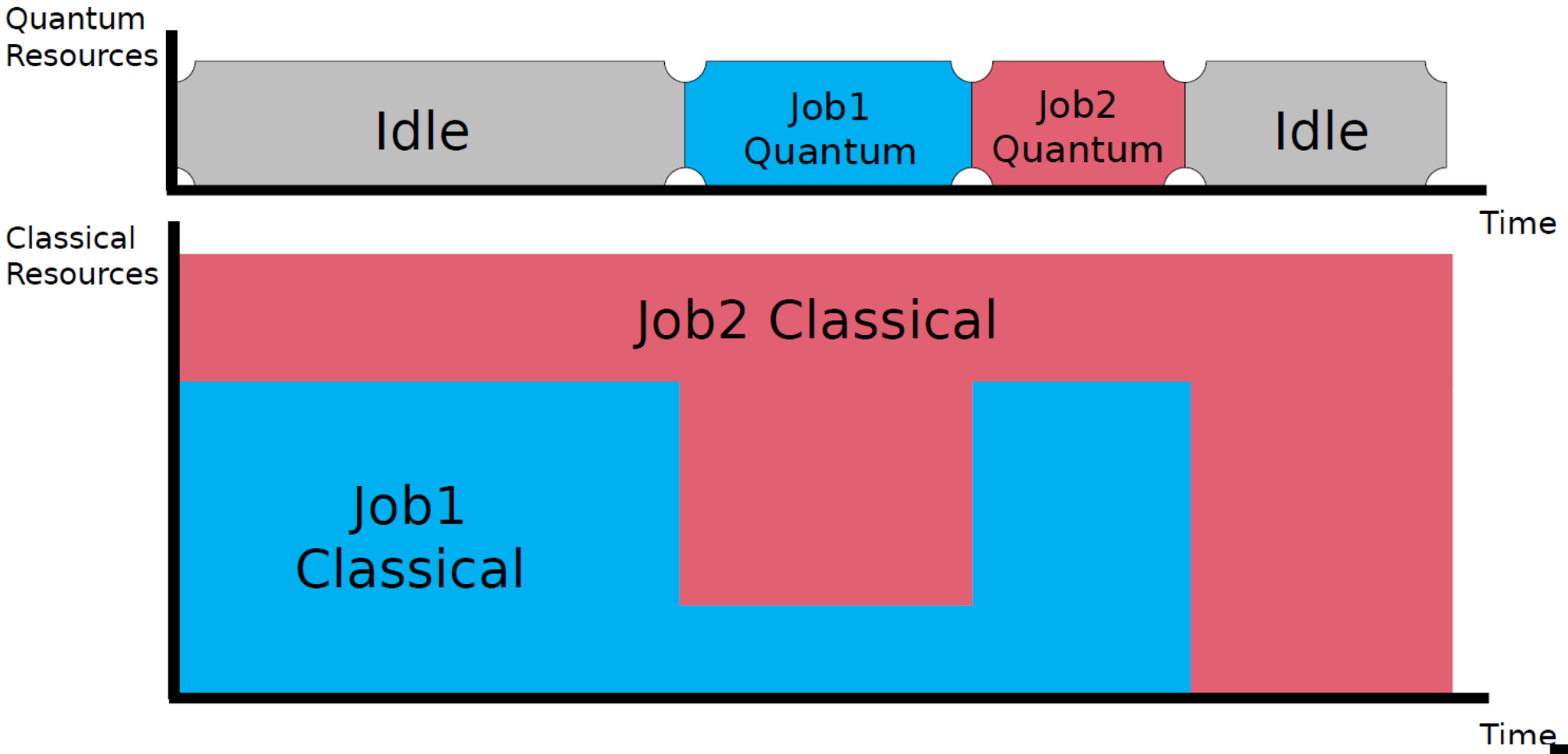
Viviani et al., accepted at FORCE (2025).

TRADITIONAL RESOURCE MANAGEMENT



Credit to Sergio Iserte and Petter Sandås from BARCELONA SUPERCOMPUTING CENTER

DYNAMIC RESOURCE MANAGEMENT



Credit to Sergio Iserte and Petter Sandås from BARCELONA SUPERCOMPUTING CENTER

MULTIPLE MALLEABILITY SOLUTIONS

Adaptive MPI:

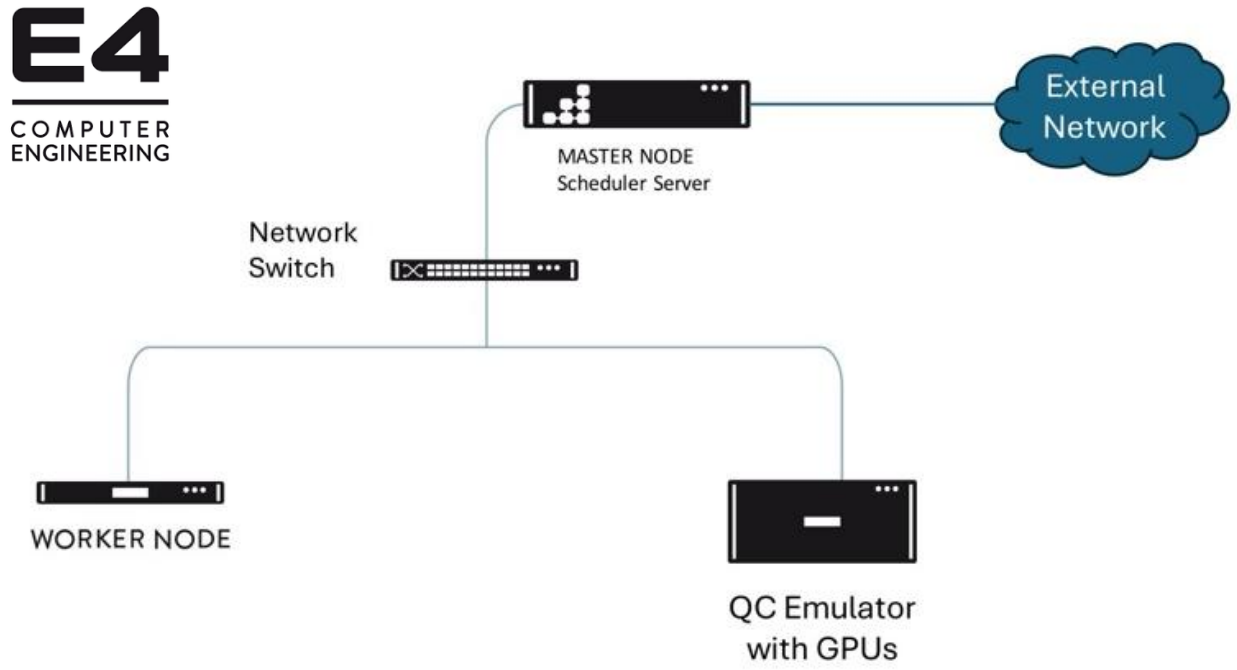
- Automatic migration of resources done via virtualisation of physical resources
- Data movement and checkpoint managed by the library
- No OOTB integration with SLURM

DMRlib:

- A selection of MPI-like primitives to ease malleability usage
- Data movement and checkpoint managed by the library
- Integration with SLURM
- More intrusive in the code

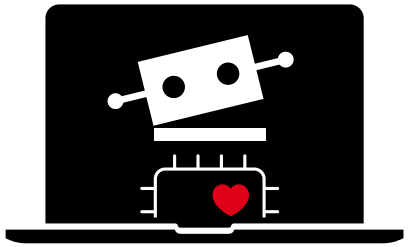
Other possible solutions: FlexMPI, MPI Sessions, ParaStation

OUR TESTBED PLATFORM



SUMMARY & PERSPECTIVES

- HPC-QC integration is a hot topic with many unsolved challenges
- With SmartHPC-QC, we chose to focus on the scheduling policies of hybrid jobs
- In our opinion, there is no universal solution to the scheduler problem today
- We identified three possible approaches depending on the workload of the hybrid job and the underlining quantum technology: workflow, VQPU and malleability



E4 COMPUTER ENGINEERING

WHEN PERFORMANCE MATTERS

THANKS!

Gabriella Bettonte
mail: gabriella.bettonte@e4company.com