Curriculum Learning for Qubit Mapping Across Hardware Topologies

Aleksandr Govenko

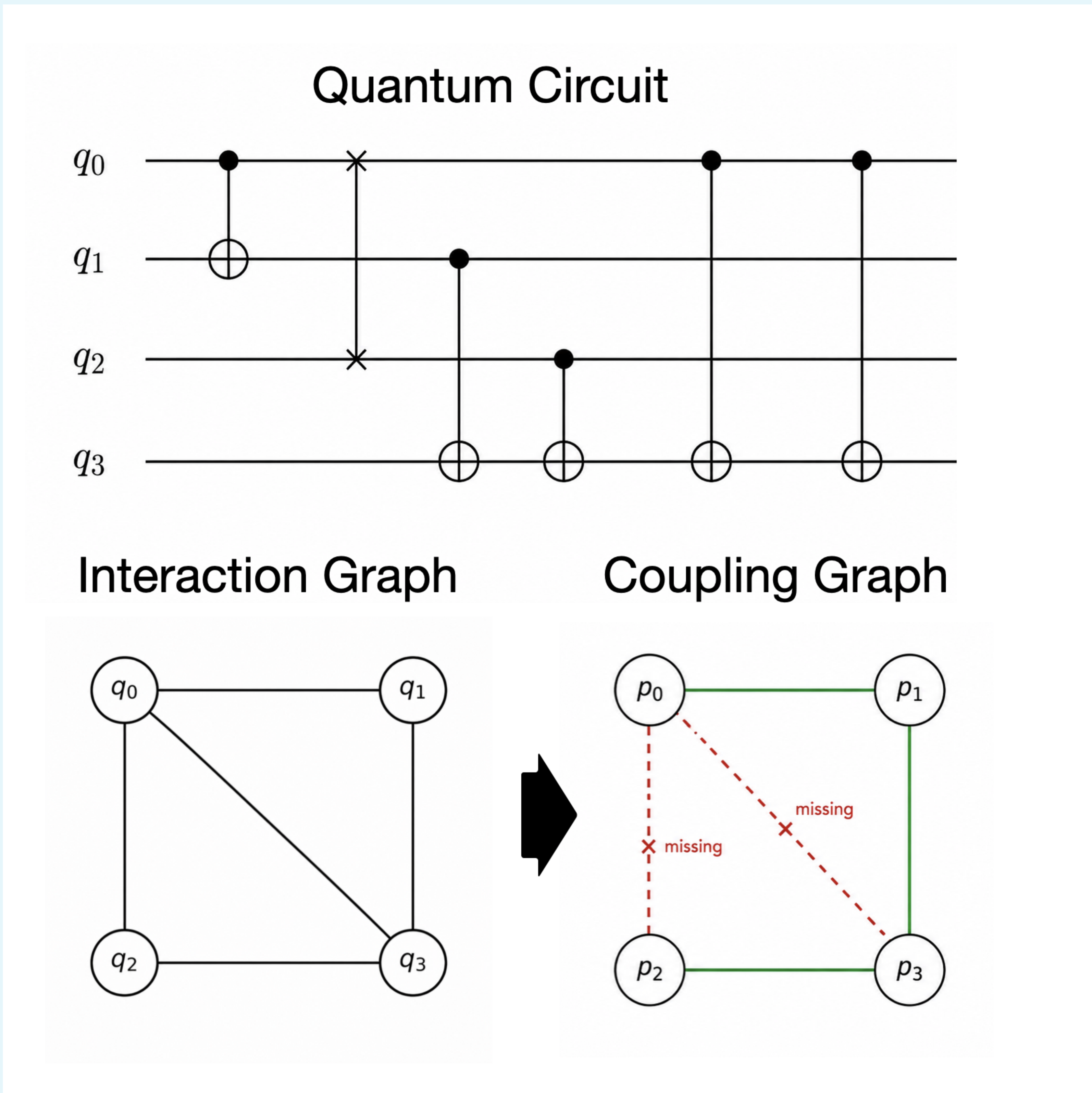
Supervisors: Sebastian Feld, Akash Kundu, Matthijs Spaan EEMCS, Delft University of Technology agovenko@tudelft.nl

June 2026

1. Problem

- Quantum circuits must be compiled onto physical hardware before execution
- Hardware qubits are sparsely connected — not all qubit pairs can interact directly
- Mismatched interactions require SWAP gates, adding noise and reducing fidelity
- Current approach: train a separate RL agent per device topology — expensive, with poor cross-topology transfer

Figure 1: Problem illustration

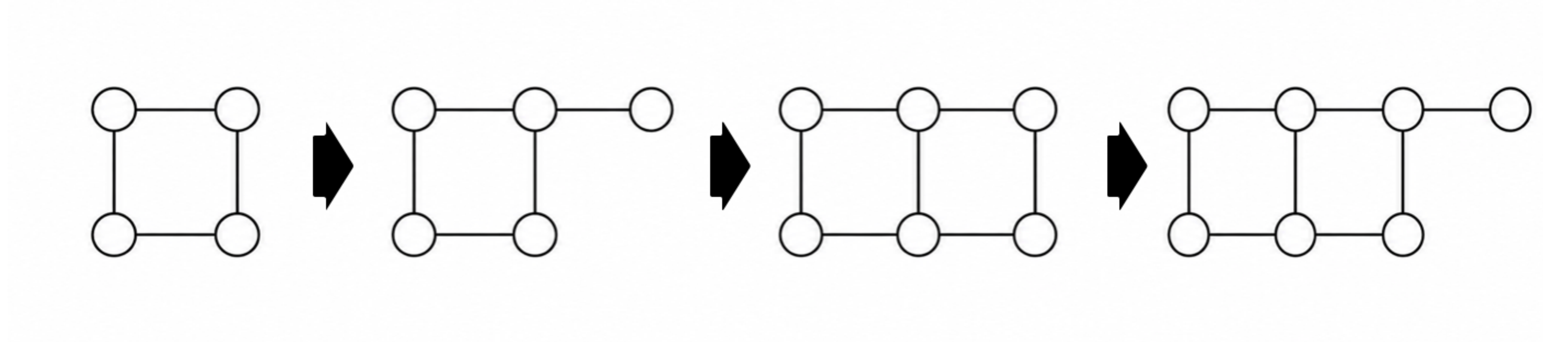


Research Question

Can curriculum learning over hardware topologies produce an RL agent for initial qubit mapping that generalises to unseen devices?

2. Method

- MaskablePPO agent in QGym's `InitialMapping` environment
- Train on grid-like topologies of increasing size (4–8 qubits)
- Advance to next topology when mean reward $\tilde{r} > 0.5$
- Replay ratio: fraction of training spent on new vs. earlier topologies after advancing



Curriculum variants — differ in replay ratio after advancing:

- 100:0 — train only on new (frontier) topology
- 80:20 — 80% frontier, 20% earlier topologies
- 50:50 — equal split

Baselines: single topology (line8), all q8 topologies (q8-only), random topology order — no curriculum

Does ordered exposure to simpler topologies help?

Conclusion

- Curriculum provides better initialisation for larger topologies
- Multi-topology training enables zero-shot transfer to unseen topologies
- Replay ratio controls generalisation vs. frontier specialisation

4. Limitations and Future Work

- All experiments ≤ 8 qubits; real devices have 50–100+
- Random interaction graphs, not real circuit benchmarks
- No hardware noise or fidelity modelling
- Adaptive curriculum scheduling (e.g. PLR, PAIRED) may outperform fixed thresholds
- Evaluate on standard benchmarks (MQT Bench)

3. Results

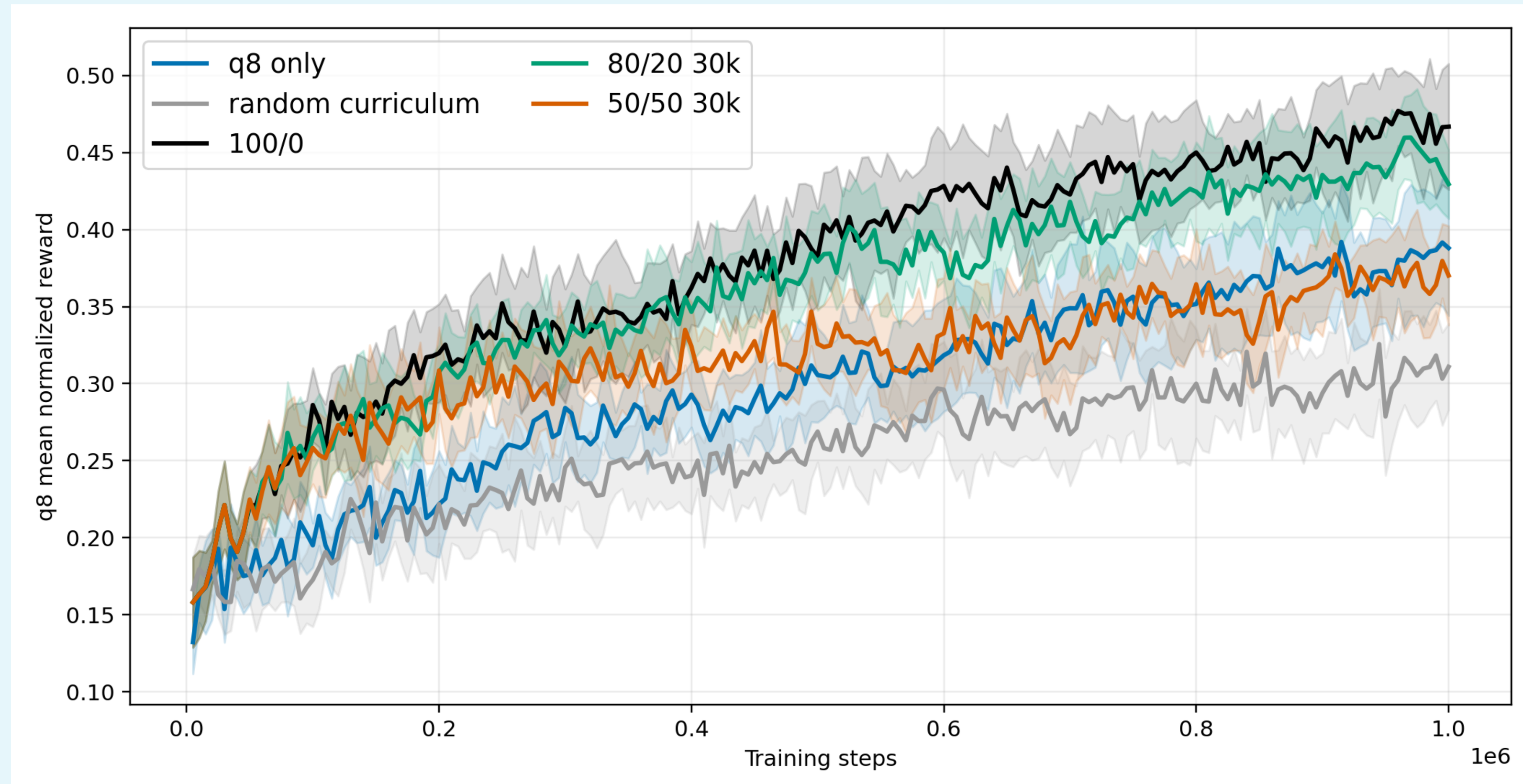
Mean normalised reward $\tilde{r}$ (10 seeds). Held-out: topologies never seen during training. 8-qubit frontier: largest training topology.

Condition	Held-out	8-qubit frontier
100/0, 30k warmup	0.467 ± 0.035	0.462 ± 0.036
80/20, 30k warmup	**0.502** ± **0.032**	0.432 ± 0.024
50/50, 30k warmup	0.494 ± 0.031	0.369 ± 0.037
q8-only baseline	0.389 ± 0.046	0.384 ± 0.043
fixed-line8 baseline	0.421 ± 0.043	0.428 ± 0.043
random curriculum	0.487 ± 0.032	0.315 ± 0.038

All curriculum variants outperform narrow-training baselines on held-out topologies. Warmup length has no meaningful effect.

More replay → better generalisation, weaker frontier performance.

Figure 2: 8 qubit reward over training (mean ± std, 10 seeds)



100/0 and 80/20 converge faster and higher than q8-only despite fewer q8-specific steps — positive transfer from smaller topologies.

Code & reproducibility

`https://github.com/alvov26/topology-curriculums`

TU Delft — Delft University of Technology