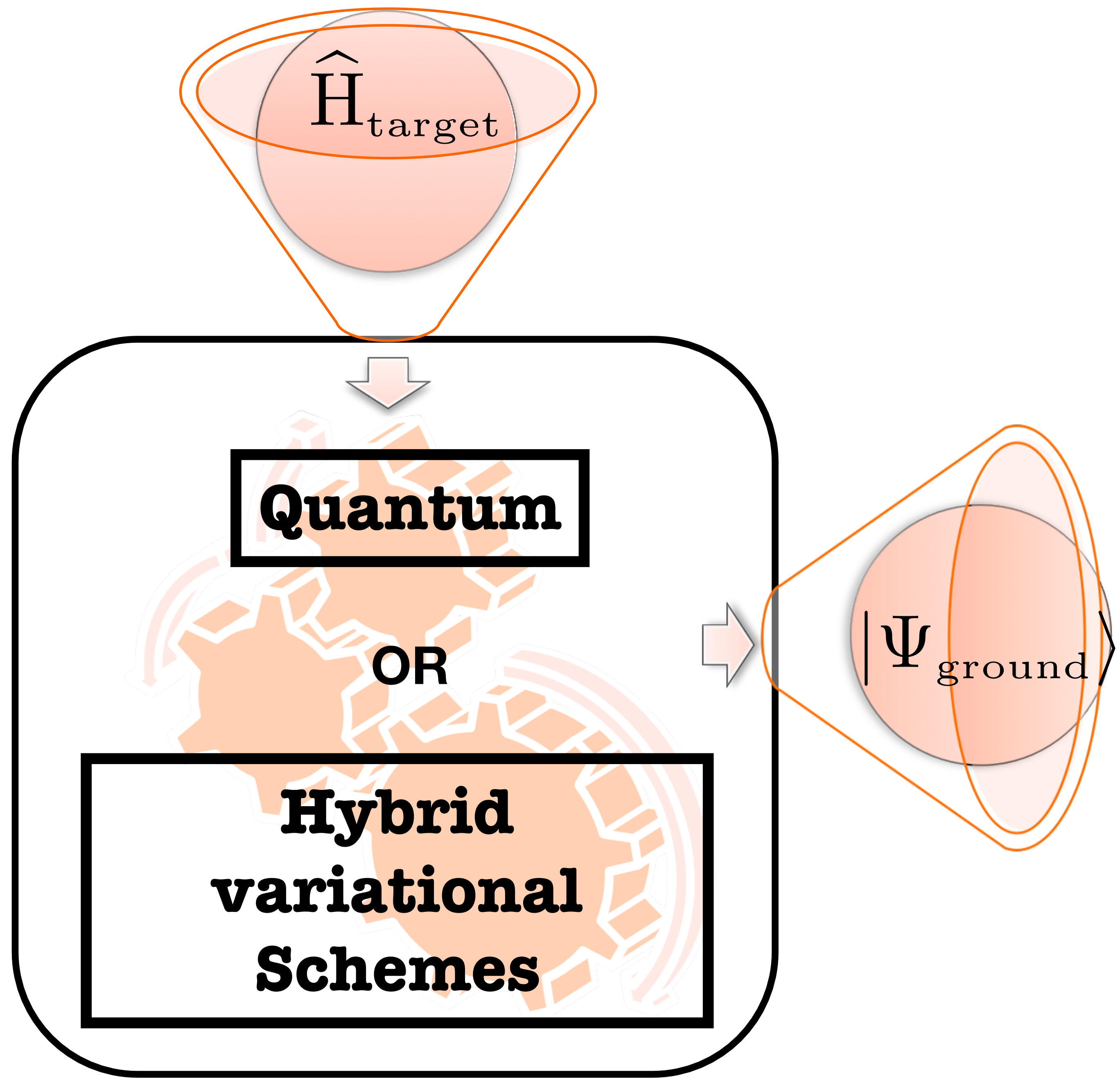
Parameter optimization strategies in variational quantum algorithms

Glen Bigan Mbeng

Seminario INFN, Bari, Italy
18th March, 2025

parityqc
universität innsbruck
beyondC

Quantum ground state preparation



Applications

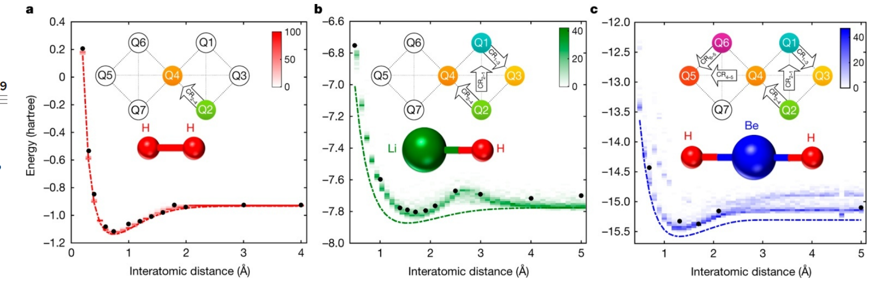
❖**Quantum chemistry**

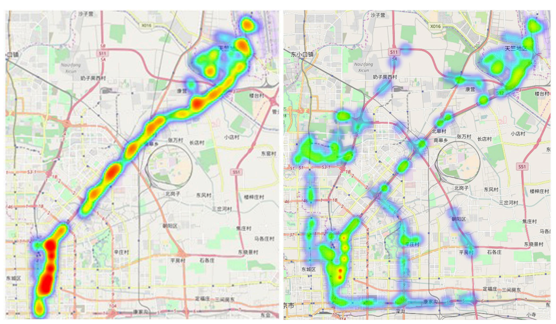
LETTER

doi:10.1038/nature23879

Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets

Abhinav Kandala[1]*, Antonio Mezzacapo[1]*, Kristan Temme[1], Maika Takita[1], Markus Brink[1], Jerry M. Chow[1] & Jay M. Gambetta[1]





Traffic Flow Optimization Using a Quantum Annealer

Florian Neukart[1], Gabriele Compostella[2], Christian Seidel[2], David von Dollen[1], Sheir Yarkoni[3] and Bob Parney[3]*

❖**Computer Science**

❖**Machine learning**

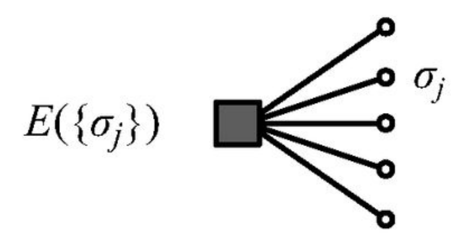
PNAS

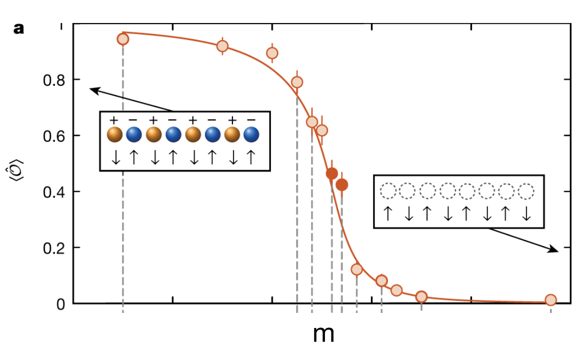
Efficiency of quantum vs. classical annealing in nonconvex learning problems

Carlo Baldassi[a,b,1,2] and Riccardo Zecchina[a,c,1,2]

[a]Bocconi Institute for Data Science and Analytics, Bocconi University, 20136 Milan, Italy; [b]Istituto Nazionale di Fisica Nucleare, Sezione di Torino, 10125 Turin, Italy; and [c]Condensed Matter and Statistical Physics Group, International Centre for Theoretical Physics, 34151 Trieste, Italy

Edited by William Bialek, Princeton University, Princeton, NJ, and approved January 2, 2018 (received for review June 26, 2017)





ARTICLE

https://doi.org/10.1038/s41586-019-1177-4

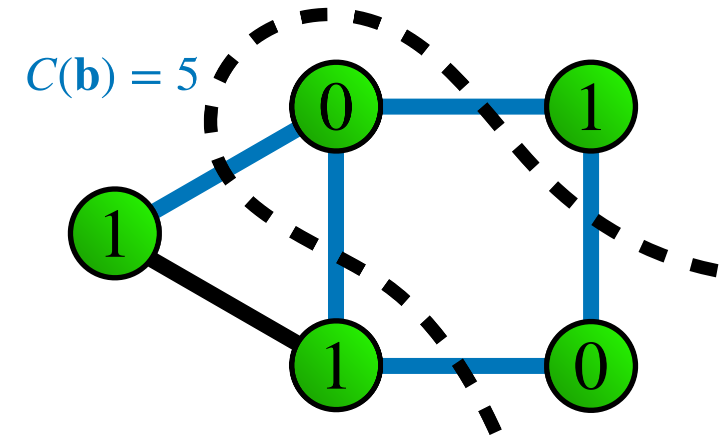
Self-verifying variational quantum simulation of lattice models

C. Kokail[1,2,3], C. Maier[1,2,3], R. van Bijnen[1,2,3], T. Brydges[1,2], M. K. Joshi[1,2], P. Jurcevic[1,2], C. A. Muschik[1,2], P. Silvi[1,2], R. Blatt[1,2], C. F. Roos[1,2] & P. Zoller[1,2]*

❖**Physics**

Quantum Optimization

Combinatorial optimization: Minimization of a single valued function of discrete variables



Examples

MaxCut: $E_{cl}(\mathbf{b}) = -C(\mathbf{b}) = -\sum_{\langle i,j\rangle} (b_i \oplus b_j)$

weighted-MaxCut: $E_{cl}(\mathbf{b}) = -\sum_{\langle i,j\rangle} J_{ij}(b_i \oplus b_j)$

bit: $b = 0,1$

qubit: $|\psi\rangle = a_0|0\rangle + a_1|1\rangle = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$

Pauli matrices: $\hat{\sigma}^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \hat{\sigma}^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \hat{\sigma}^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

Optimal configuration?

$E_{cl}(\mathbf{b}) = -\sum_{\langle i,j\rangle} J_{ij}(b_i \oplus b_j)$

$b_j = (1 - \hat{\sigma}^z)/2$

Ground state $|\psi_{\text{ground}}\rangle$?

$$\hat{H}_z = \frac{1}{2}\sum_{\langle i,j\rangle} J_{ij}(\hat{\sigma}^z_i\hat{\sigma}^z_j - 1)$$

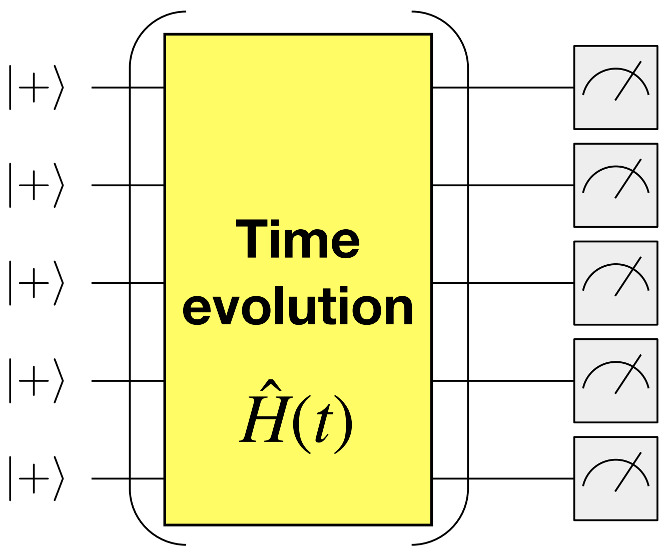
Quantum Optimization

$$\hat{H}_z = \sum_{\langle i,j \rangle} J_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z$$

Approximate ground state $|\psi_{\text{ground}}\rangle$?

Analog quantum optimization

Review: T. Albash and D. A. Lidar, Rev. Mod. Phys. (2018)

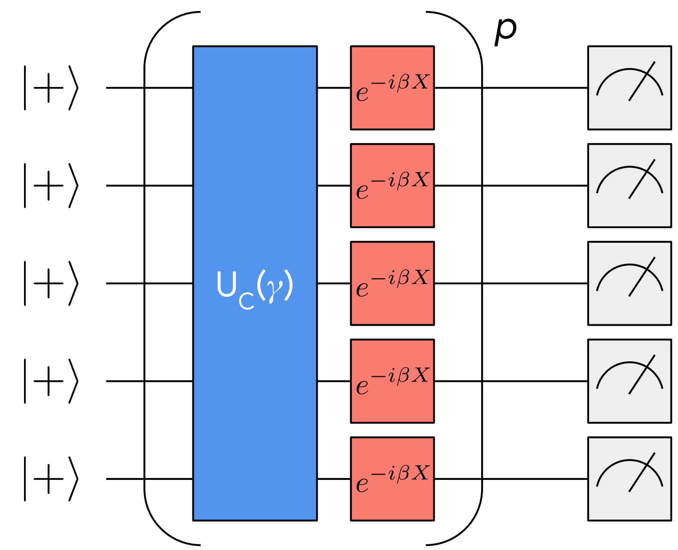


complicated quantum dynamics

$$|\psi_{\text{output}}\rangle = \text{Texp}\left(-\frac{i}{\hbar}\int_0^\tau \hat{H}(t)\mathrm{d}t\right)|\psi_0\rangle$$

Digital quantum optimization

[E. Farhi, *et al*, arXiv:1411.4028 (2014)]



many simple quantum gates

$$|\psi_{\text{output}}\rangle = \ldots \hat{U}_z(\gamma_2)\hat{U}_x(\beta_2)\hat{U}_z(\gamma_1)\hat{U}_x(\beta_1)|\psi_0\rangle$$

Example: quantum search

Grover's problem

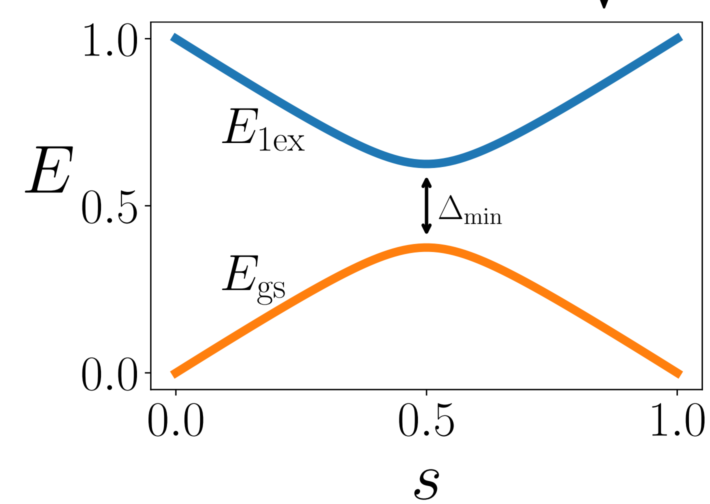
Search an item in an unsorted list of $n = 2^N$ items

Database: binary strings (classical spin configurations) of length N

Marked item: eg. $|\bar{z}\rangle = |\uparrow\downarrow\uparrow\uparrow\downarrow\downarrow\cdots\downarrow\downarrow\uparrow\rangle$

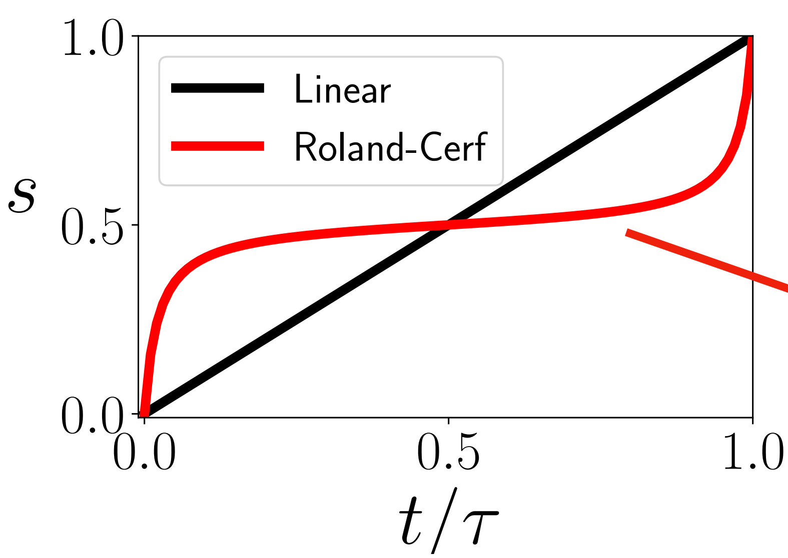
Runtime of classical sequential search: $\tau \propto n$

$$\Delta_{\min} = \Delta(s_c = \frac{1}{2}) = \frac{1}{\sqrt{n}}$$



Quantum Annealing for Grover's problem

$$\widehat{H}(s(t)) = s(t)[\mathbf{1} - |\bar{z}\rangle\langle\bar{z}|] + (1 - s(t))[\mathbf{1} - |+\rangle\langle+|]$$



Linear s(t) (constant speed): $\tau \propto \frac{1}{\Delta^2_{\min}} \propto n$

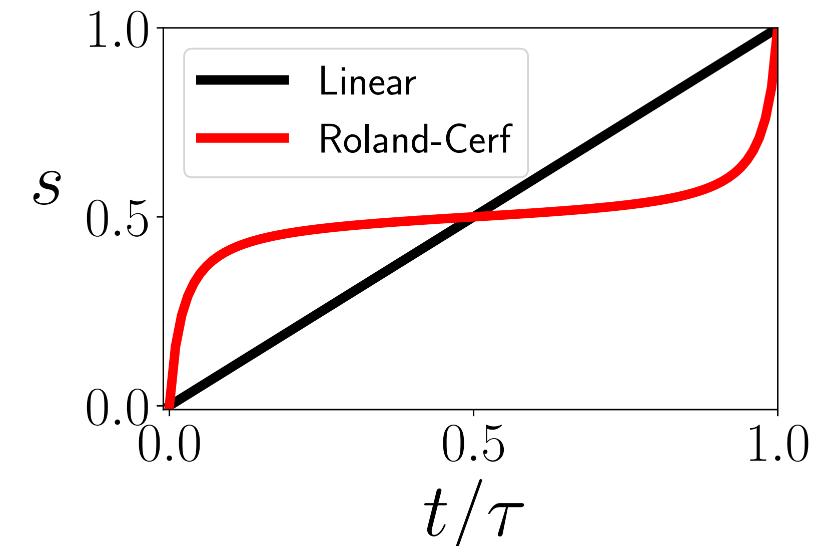
Smart s(t) (slows down near s_c): $\tau \propto \frac{1}{\Delta_{\min}} \propto \sqrt{n}$

[Roland and Cerf, PRA (2002)]

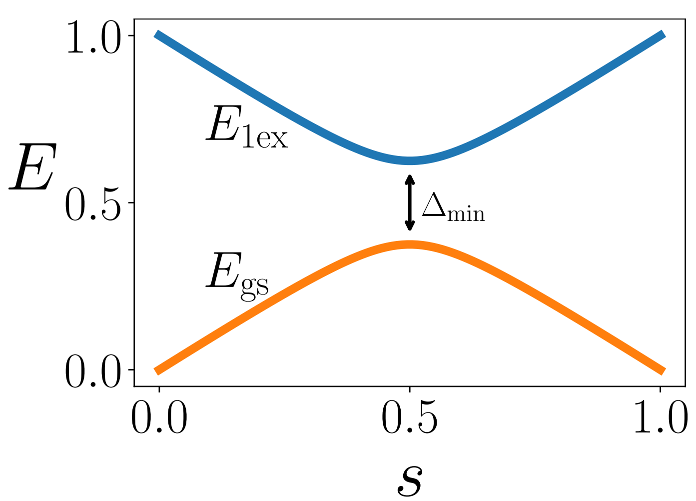
The key role of $s(t)$

Grover's problem

Search an item in an unsorted list of $n = 2^N$ items



Optimal control of the schedule is crucial!



Optimal control of the schedule required spectral information

[Z Jiang, et al PRA (2017)]

How do we do it in general, when spectral information is not available?

[Cubitt et al.,``Undecidability of the spectral gap", Nature (2015)]

Quantum Approximate Optimization

Quantum approximate optimization algorithm

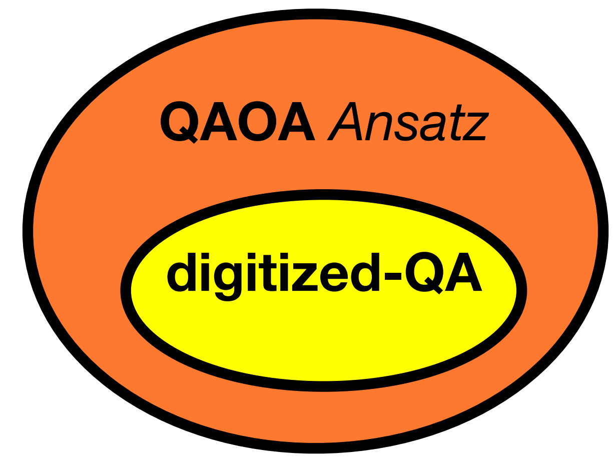
[E. Farhi, *et al*, arXiv:1411.4028 (2014)]
[Blekos, et al Phys. Rep. (2024)]

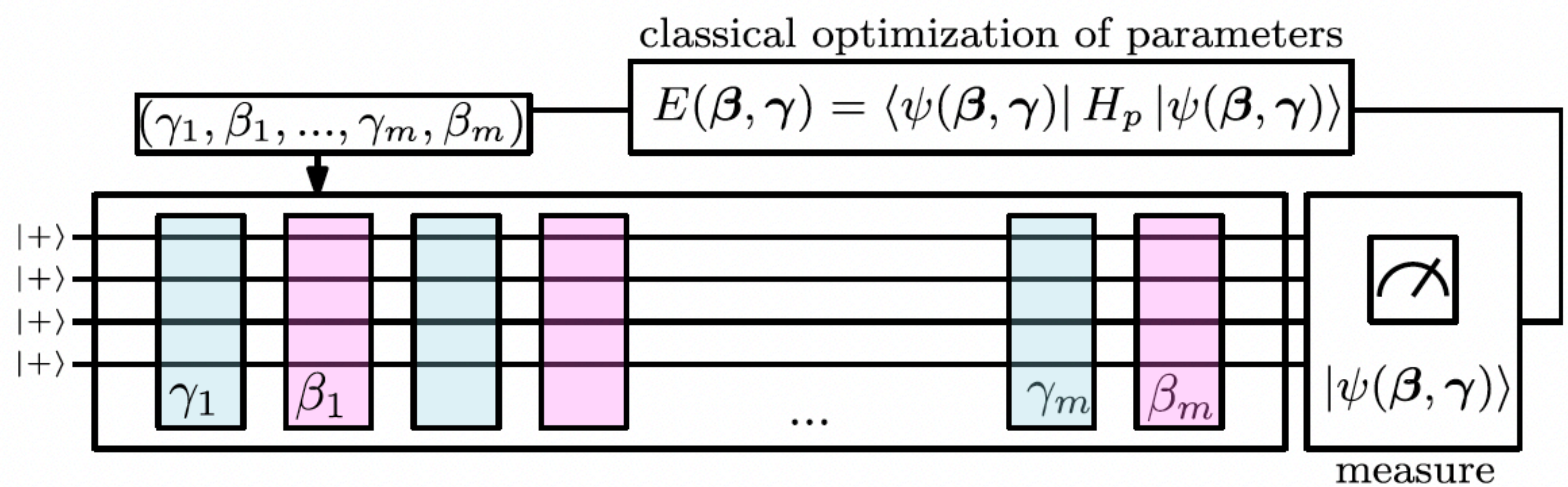
$$U_p(\gamma) = \mathrm{e}^{-i\gamma H_p}$$

$$H_p = H_z = \sum_{i=1}^{N} \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z$$

$$U_x(\beta) = \mathrm{e}^{-i\beta H_x}$$

$$H_x = -\sum_{i=1}^{N} \hat{\sigma}_i^x$$





Barren plateaus in quantum neural networks

ARTICLE

DOI: 10.1038/s41467-018-07090-4 OPEN

Barren plateaus in quantum neural network training landscapes

Jarrod R. McClean[1], Sergio Boixo[1], Vadim N. Smelyanskiy[1], Ryan Babbush[1] & Hartmut Neven[1]

ARTICLE

https://doi.org/10.1038/s41467-021-21728-w OPEN

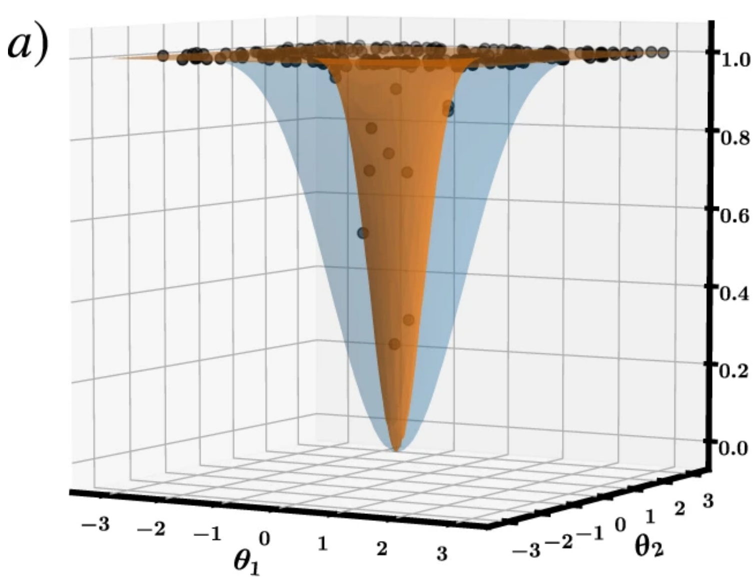
Check for updates

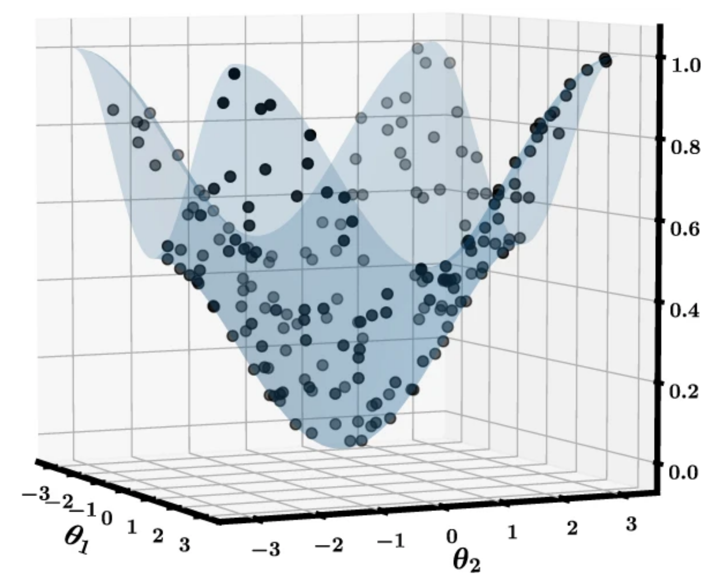
Cost function dependent barren plateaus in shallow parametrized quantum circuits

M. Cerezo[1,2✉], Akira Sone[1,2], Tyler Volkoff[1], Lukasz Cincio[1] & Patrick J. Coles[1✉]

Gradient of the cost function $\nabla E = \left(\frac{\partial E}{\partial \gamma_1}, \frac{\partial E}{\partial \gamma_2} \ldots \frac{\partial E}{\partial \beta_1} \frac{\partial E}{\partial \beta_2} \ldots \right)$

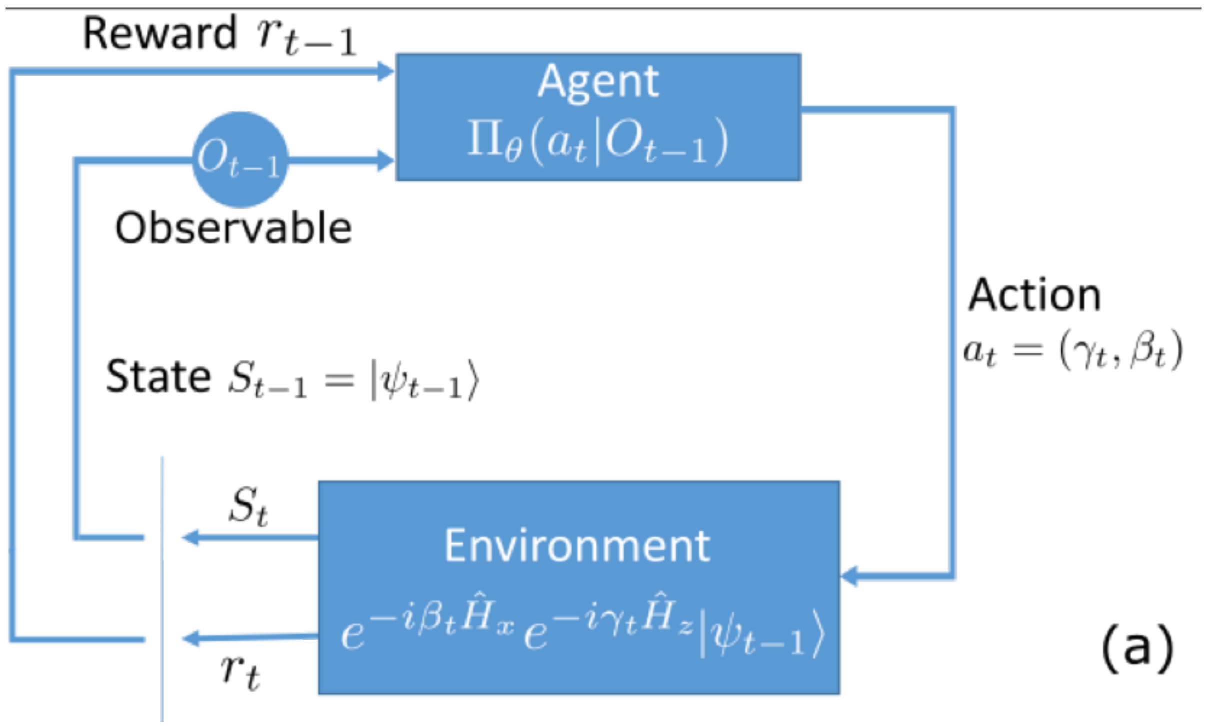
$$\mathrm{Var}\left[\frac{\partial E}{\partial \theta}\right] \to 0 \qquad\qquad \mathrm{Var}\left[\frac{\partial E}{\partial \theta}\right] \neq 0$$





QAOA as Markov decision process

[Wauters, et al., Phys. Rev. Res. (2020)]



Proximal Policy Optimization (PPO)

RL library: https://spinningup.openai.com

Quantum env: https://github.com/mwauters92/QuantumRL

episode → $|\psi_0\rangle$ through $|\psi_P\rangle$
state **S** → instantaneous $|\psi_t\rangle$
observable **O** → $\langle\psi_t|O|\psi_t\rangle$
action **A** → (γ_m, β_m)
reward **R** → $r = -\delta_{t,P}\langle\psi_t|\,\widehat{\mathrm{H}}_{\mathrm{target}}\,|\psi_t\rangle$

Two Observables

$O^{(z)} = \langle\psi_t|\,\widehat{H}_z\,|\psi_t\rangle$

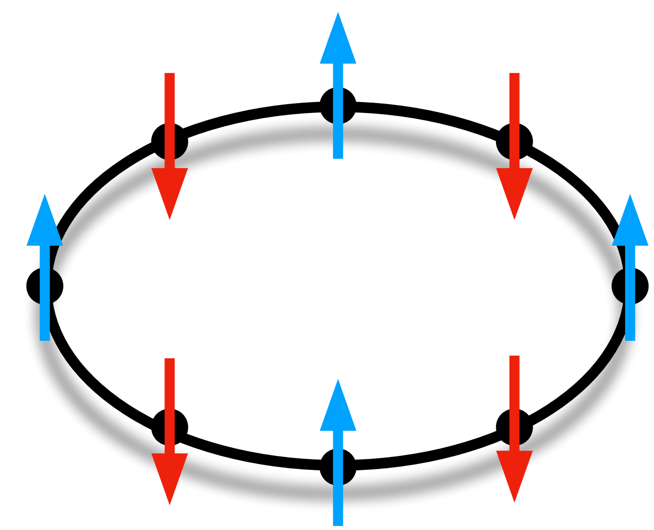
$O^{(x)} = \langle\psi_t|\,\widehat{H}_x\,|\psi_t\rangle$

Require multiple **quantum measurements**!

Quantum Ising Chain/Ring of disagrees

$$\widehat{\mathrm{H}}_{\text{target}} = \widehat{H}_z + \cancel{h\widehat{H}_x}$$

$$\widehat{H}_x = -\sum_{i=1}^{N} \hat{\sigma}_i^x \qquad \widehat{H}_z = \sum_{i=1}^{N} \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z$$



$$\hat{U}_m = \mathrm{e}^{-i\beta_m \widehat{H}_x} \mathrm{e}^{-i\gamma_m \widehat{H}_z}$$

$$|\psi_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta})\rangle = \hat{U}_{\mathrm{P}} \cdots \hat{U}_2 \hat{U}_1 |\psi_0\rangle$$

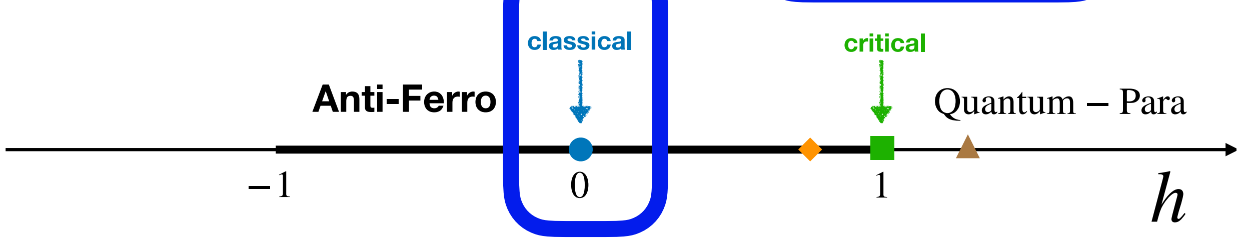
$$E_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta}) = \langle \psi_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta}) | \widehat{H}_{\text{target}} | \psi_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta})\rangle$$

$$E_{\min} = -N \qquad E_{\max} = N$$

residual energy
(relative **error**)

$$\epsilon_{\mathrm{P}}^{\text{res}}(\boldsymbol{\gamma},\boldsymbol{\beta}) = \frac{E_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta}) - E_{\min}}{E_{\max} - E_{\min}} \in [0,1]$$

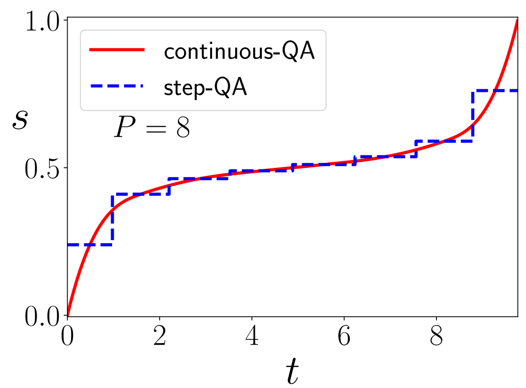
$$h = 0$$



Translationally invariant Ising Chain

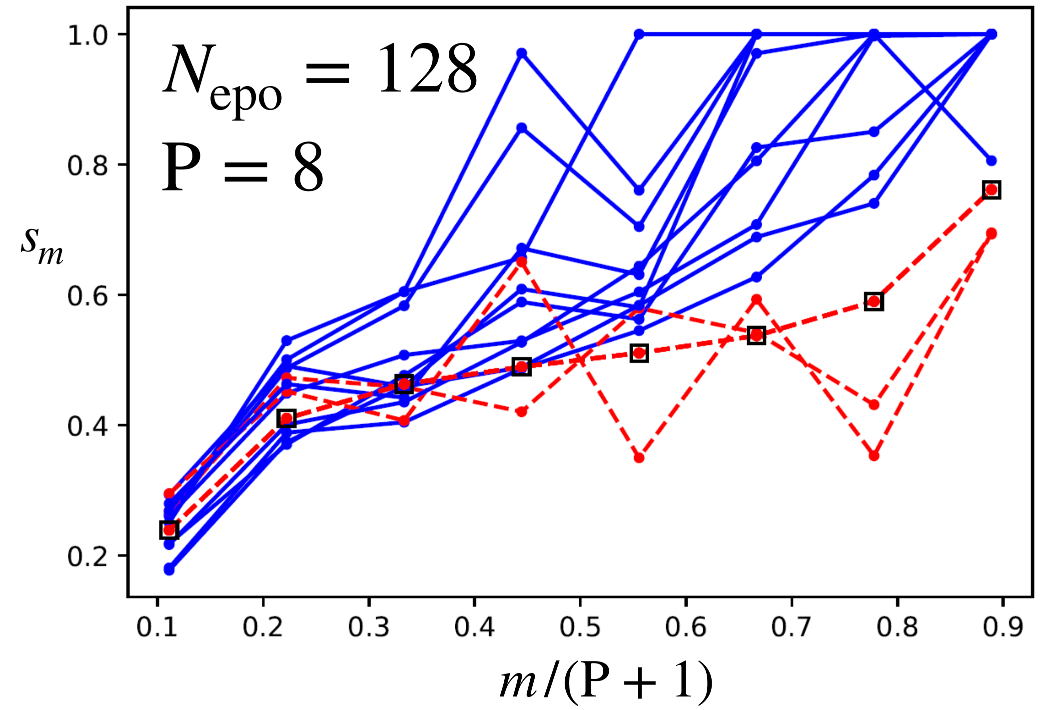
Analyze action choice

$$s_m = \frac{\gamma_m}{\beta_m + \gamma_m}$$



Allows comparison with digital-QA schedule

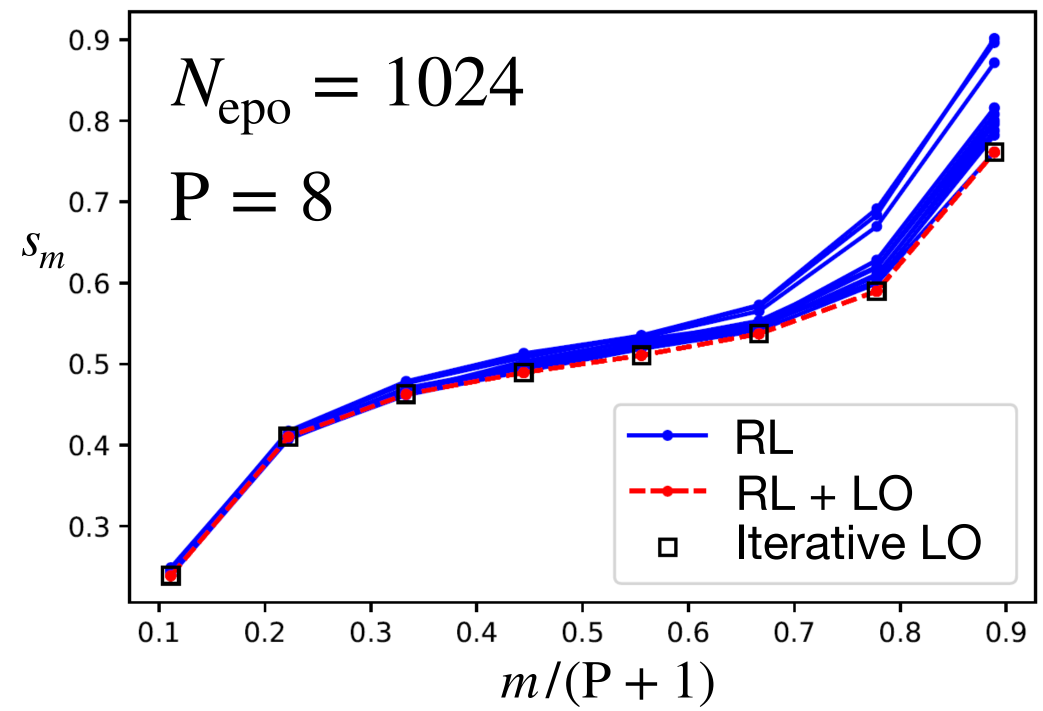
$$\widehat{H}(s_m) = s_m \widehat{\mathrm{H}}_{\text{target}} + (1 - s_m)\widehat{\mathrm{H}}_{\text{driving}}$$



I. Random action choices

II. Large trajectory dispersion. RL+LO fall on the same smooth minimum (iterative LO)

III. RL converges to QAOA iterative solution

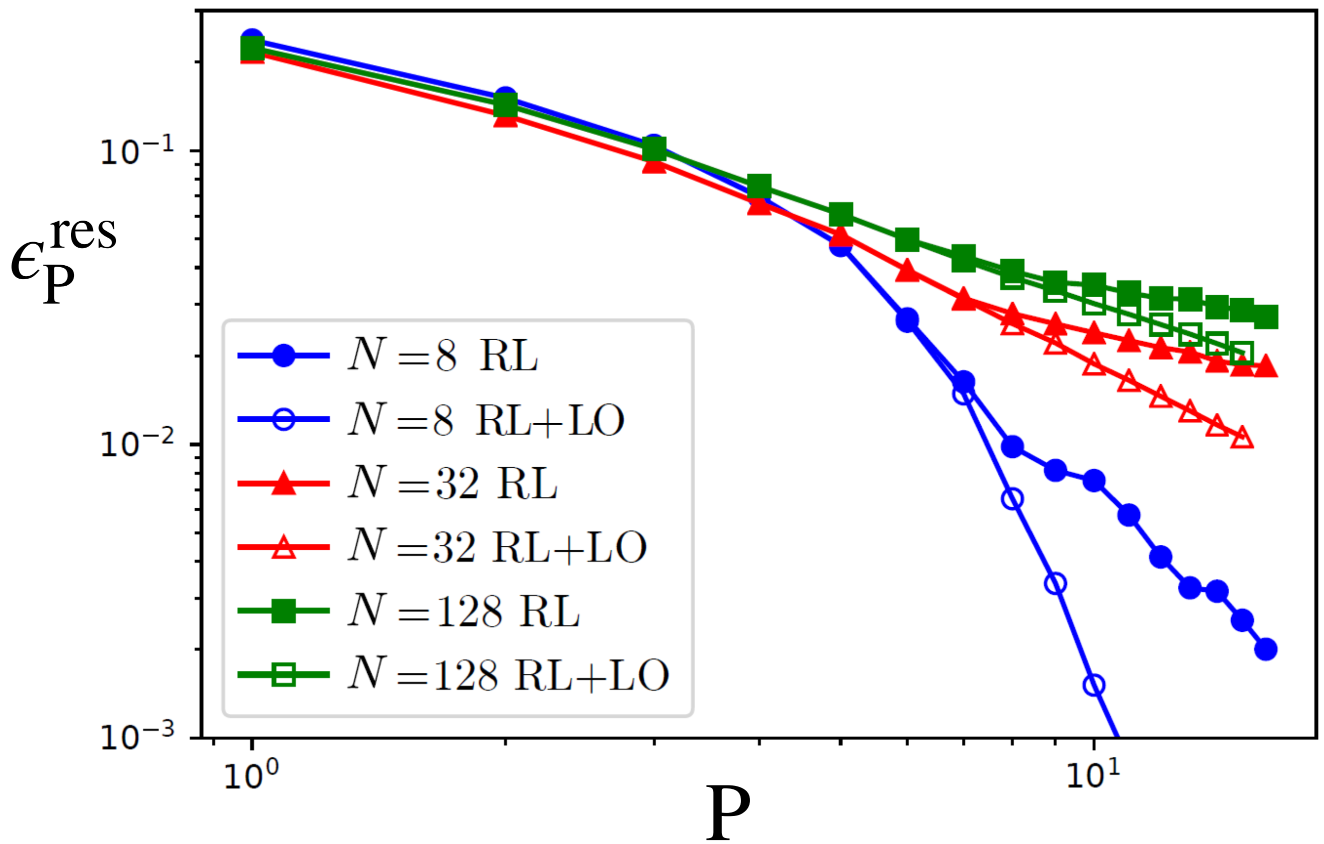


Random Ising Chain

$$\widehat{H}_z = \sum_{i=1}^{N} J_i \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z$$

J_i Uniformly disturbed in $[0,1]$

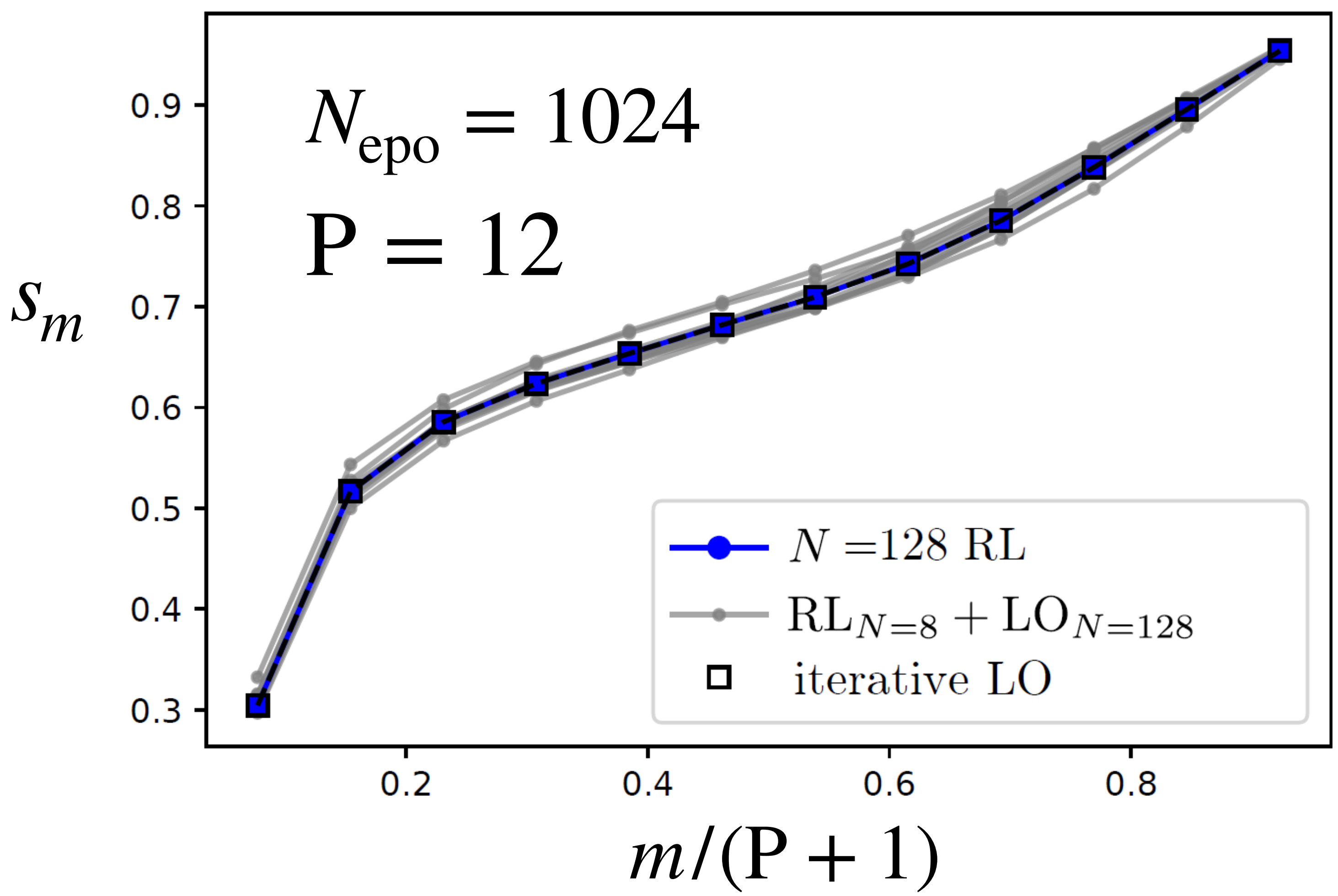
Makes the problem harder



Same phenomenology of uniform TFIM: Local optimization leads to lower energy for P > 6

Open issues:
- No analytical result
- Is solution optimal?
- Better than QA?

Random Ising Chain (transferability)



Policy transferability:

I. Train a small system ($N = 8$) on single disorder instance

II. Get approximate solution $(\boldsymbol{\gamma}^\star, \boldsymbol{\beta}^\star)_{N=8}$

III. Use $(\boldsymbol{\gamma}^\star, \boldsymbol{\beta}^\star)_{N=8}$ to initialize a local optimization on larger system ($N = 128$)

III. RL converges to QAOA iterative solution

Reduces used quantum resources

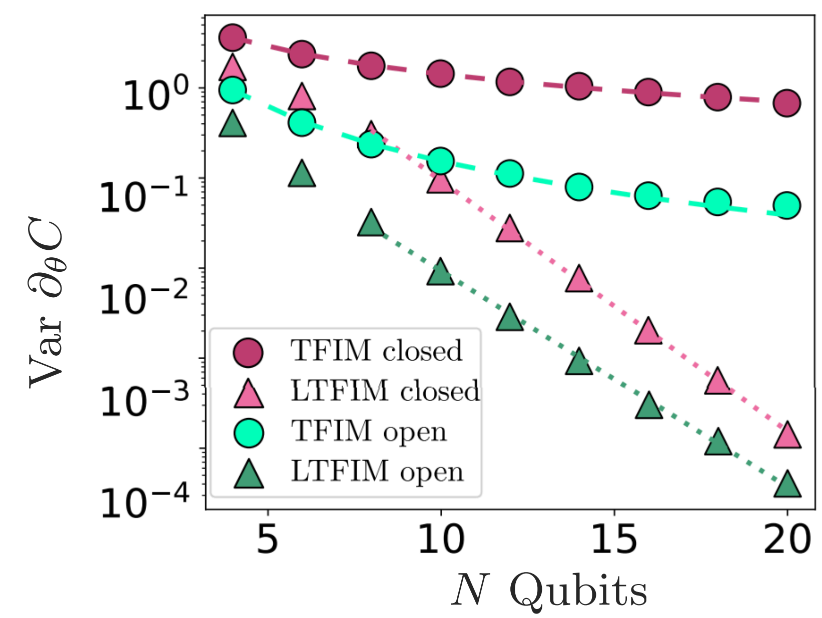
Construction of digitized-QA schedules

Barren plateaus problem

[McClean, et al, PRA (2018)]
[Larocca, et al, Quatum (2022)]

$$\hat{H}_T = \sum_{j=1}^{N} \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z - g_x \sum_{j=1}^{N} \hat{\sigma}_j^x - g_z \sum_{j=1}^{N} \hat{\sigma}_j^z$$

- Gradients vanish exponentially (flat landscape)

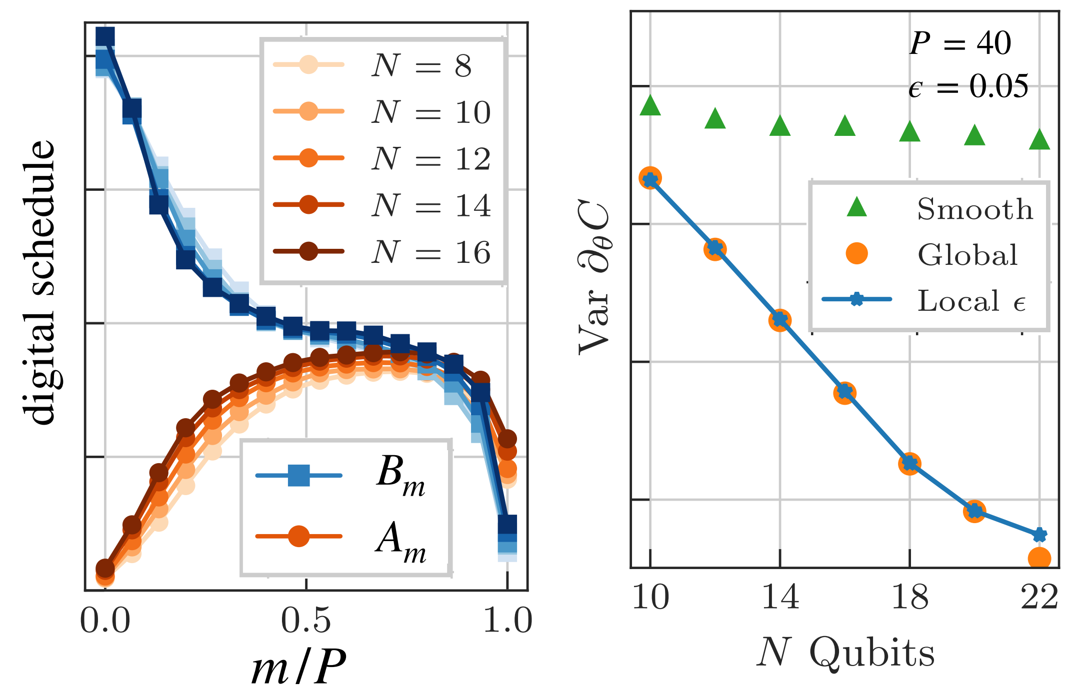


Smooth schedules

[Mele, et al, PRA (2022)]
[Torta, et. Al PRB (2022)]

- Gradients in a neighbourhood of radius ϵ
- Transferred smooth schedule avoids barren plateaus

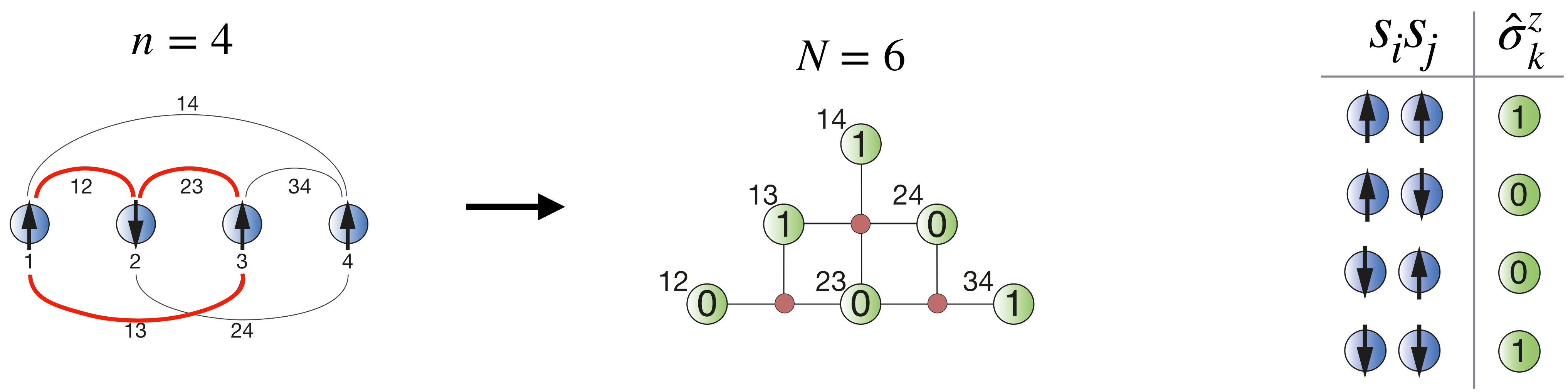
$g_z = g_x = 1$



Hardware challenge: Connectivity

Parity Architecture

[Lechner, Hauke, and Zoller, Sci. Adv. 1, (2015)]



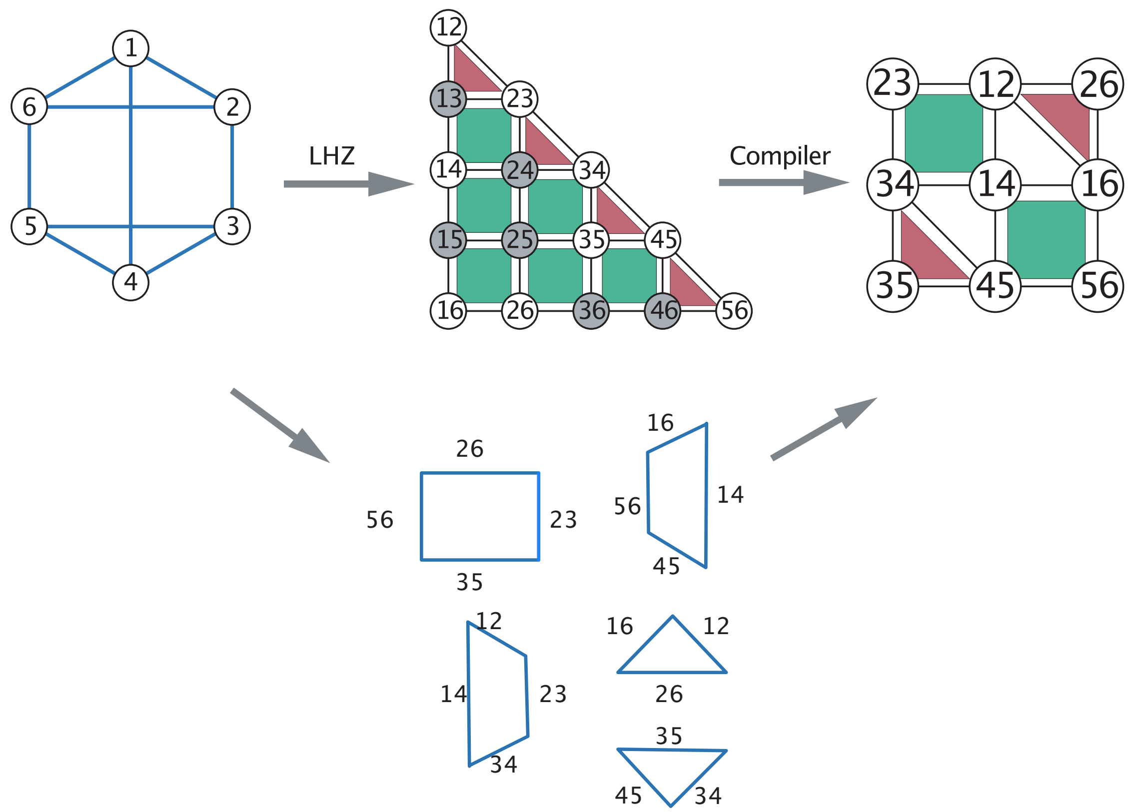
$$H_P = \sum_{i=1}^{n} \sum_{j=1}^{i-1} J_{ij} s_i s_j \longrightarrow H_{\text{LHZ}} = \underbrace{\sum_{i=1}^{N} J_i \hat{\sigma}^z_i}_{\hat{H}_z} + C \underbrace{\sum_{l=1}^{L} (1 - \hat{\sigma}^z_{l_1} \hat{\sigma}^z_{l_2} \hat{\sigma}^z_{l_3} [\hat{\sigma}^z_{l_4}])}_{\hat{H}_C}$$

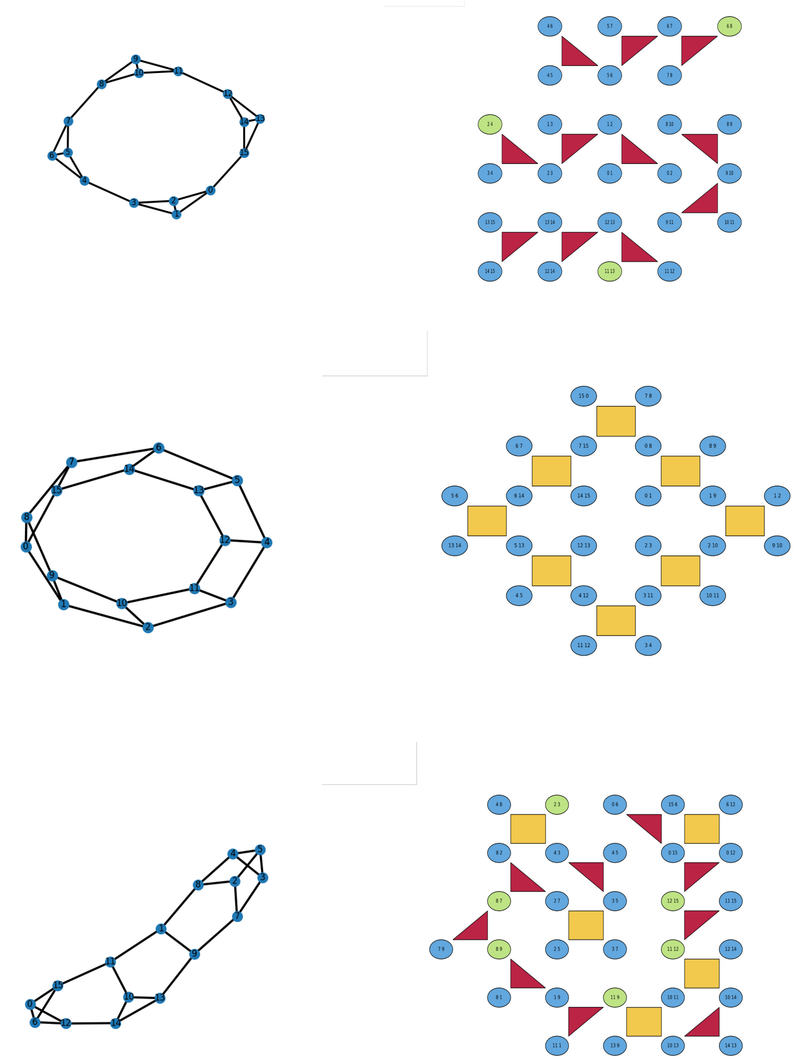
- ❖ The LHZ architecture maps an all-to-all connected spin model to a spin model with only **quasi-local** interactions.
- ❖ The physical qubits encode the **parity** of the logical qubits.
- ❖ **No long-range interactions** but only local 3- or 4-body couplings are necessary.
- ❖ The parity architecture requires nearest-neighbour interactions on a **square lattice**, regardless of the qubit platform.

Parity compilation

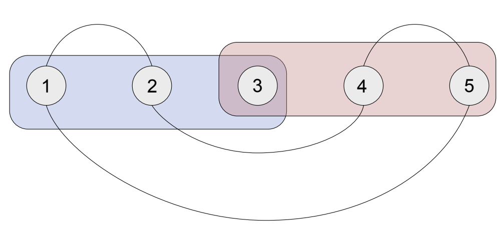
[Ender, et al, Quantum (2023)]

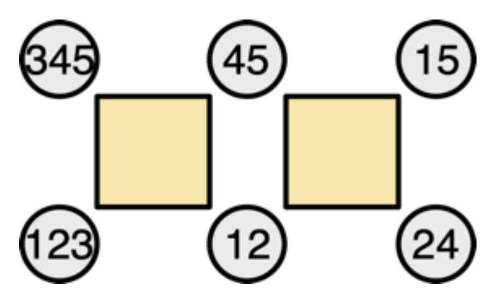
Sparse graphs





Hypergraphs





Parity based QAOA

[Lechner, IEEE Trans. Quantum Eng. (2020)]
[Fellner, et al., PRL (2022)]
[Ender, et al. arXiv (2021)]

$$\tilde{U}_c(\gamma) = e^{-i\Omega \tilde{H}_c}$$

$$\tilde{U}_p(\gamma) = e^{-i\gamma \tilde{H}_P}$$

$$\tilde{U}_x(\beta) = \prod_j e^{-i\beta \tilde{\sigma}_j^x}$$

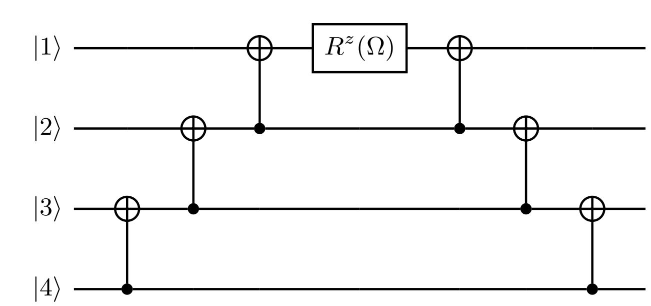
Parity QAOA ansatz:

$$|\psi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Omega})\rangle = \tilde{U}_x(\beta_p)\tilde{U}_P(\gamma_p)\tilde{U}_c(\Omega_p) \cdots \tilde{U}_x(\beta_1)\hat{U}_P(\gamma_1)\tilde{U}_c(\Omega_1)|\psi_0\rangle$$

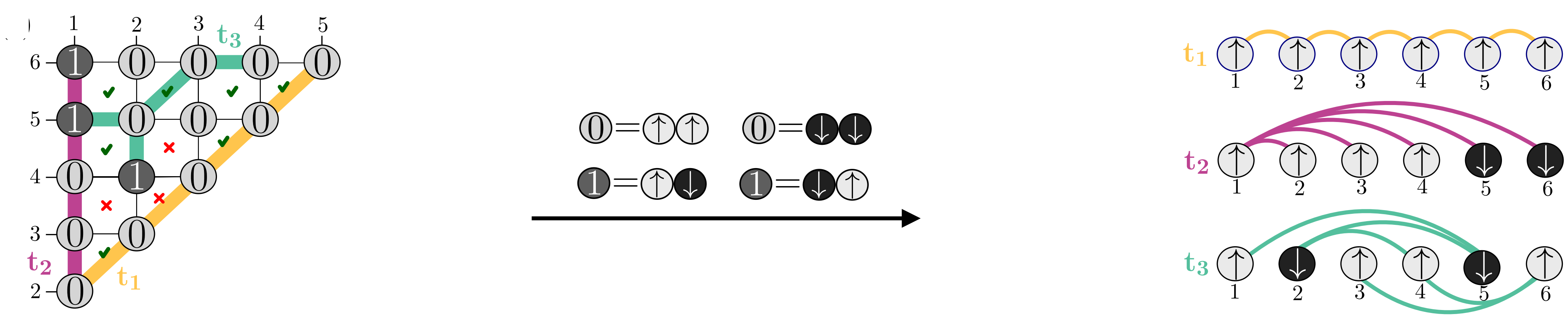
- Fully **parallelizable**
- Generalisation to **k-body terms**
- Uses fewer **CNOT gates**
- **Universal** quantum computing

4-qubit gates

$$e^{-i\Omega \tilde{\sigma}_1^z \tilde{\sigma}_2^z \tilde{\sigma}_3^z \tilde{\sigma}_4^z} =$$

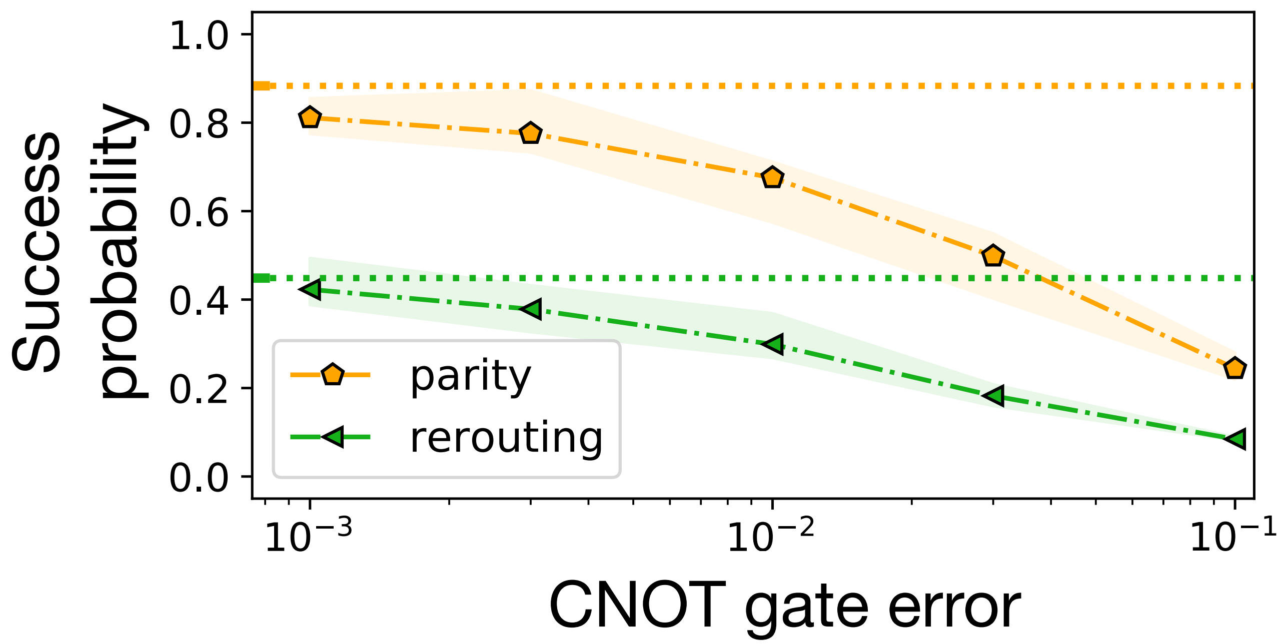


Decoding step: high chance of **measuring invalid states**



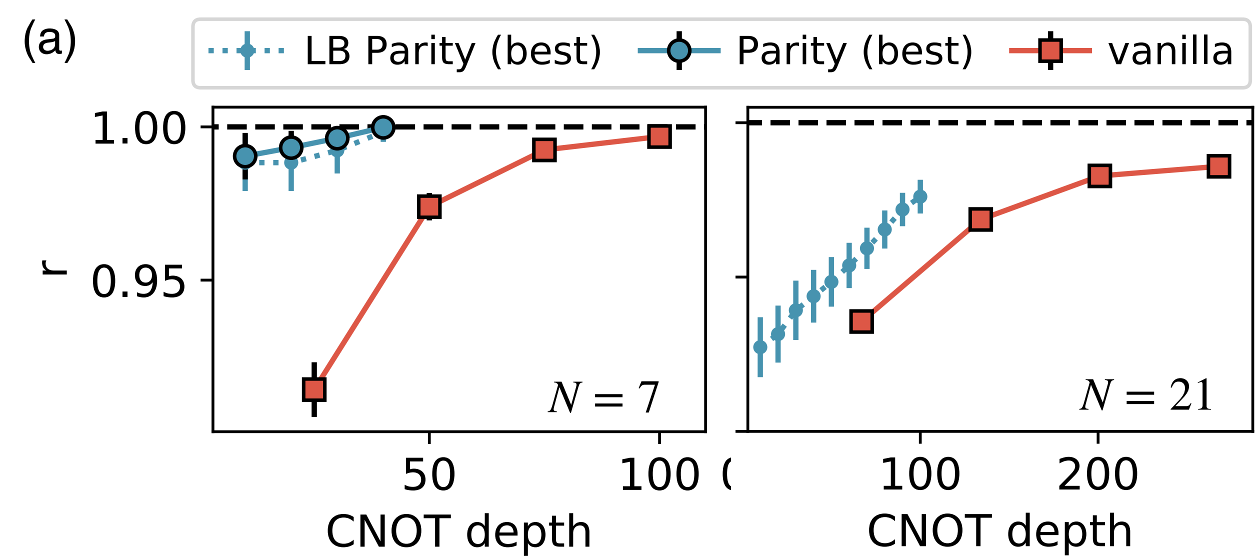
Performance of parity QAOA

[Weidinger, et al PRA 2023]
[Weidinger, et al arXiv 2024]



- $n_l = 6$ classical variables
- $m = 6$ spanning trees
- **Imperfect quantum gates**
- $n_p/n_l = 2.5$ copies for rerouting

- Search is restricted to Clifford circuits
- Computable lower bound on perfromance



Gate sequence

$$\tilde{H}_p = \tilde{H}_z + \tilde{H}_c$$

$$U_z(\gamma) = \mathrm{e}^{-i\gamma \tilde{H}_z}$$

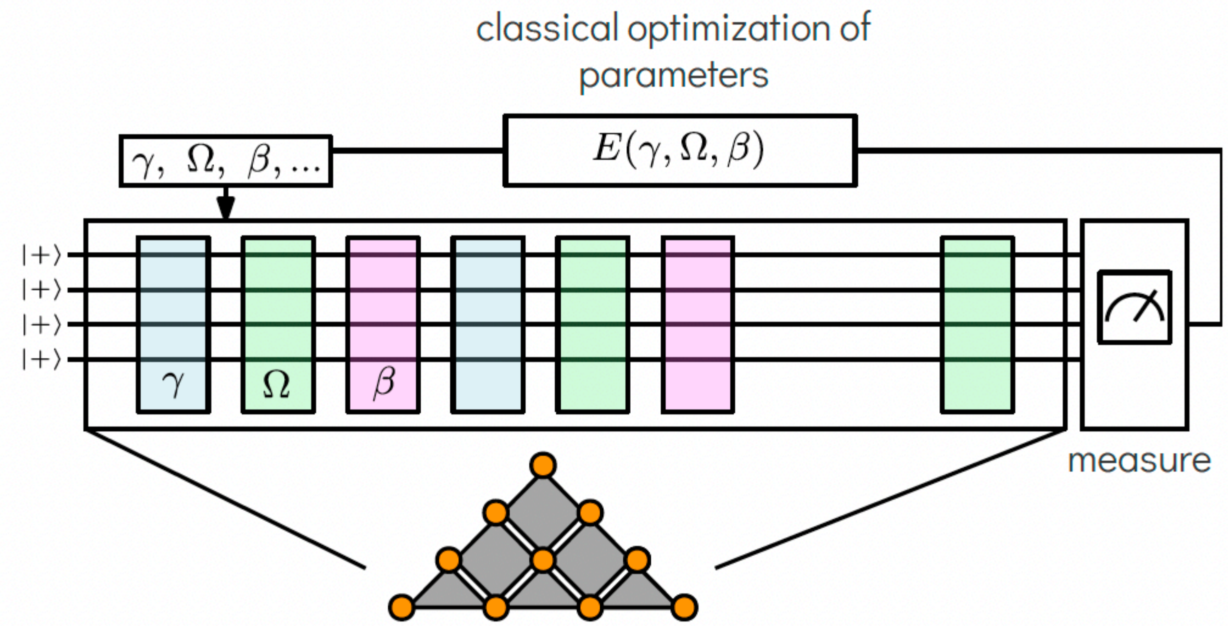
$\hat{\mathrm{Z}}$

$$U_c(\Omega) = \mathrm{e}^{-i\gamma \tilde{H}_c}$$

$\hat{\mathrm{C}}$

$$U_x(\beta) = \mathrm{e}^{-i\beta \tilde{H}_x}$$

$\hat{\mathrm{X}}$



$$A_\lambda^{(l)} = i \sum_{k=1}^{l} \alpha_k \underbrace{[H_a, [H_a, \ldots [H_a,}_{2k-1} \partial_\lambda H_a]]]$$

$$\hat{\mathrm{X}} : \tilde{U}_x(\beta) = \mathrm{e}^{-i\beta \tilde{H}_x}$$

$$\hat{\mathrm{Z}} : \tilde{U}_z(\gamma) = \mathrm{e}^{-i\gamma \tilde{H}_z}$$

$$\hat{\mathrm{C}} : \tilde{U}_c(\Omega) = \mathrm{e}^{-i\Omega \tilde{H}_c}$$

$$\hat{\mathrm{U}}_1 : \tilde{U}_1(\alpha_1) = \mathrm{e}^{-\alpha_1 [\tilde{H}_x, \tilde{H}_z]}$$

$$\hat{\mathrm{U}}_2 : \tilde{U}_2(\alpha_2) = \mathrm{e}^{-\alpha_2 [\tilde{H}_x, \tilde{H}_c]}$$

$$\hat{\mathrm{U}}_3 : \tilde{U}_3(\alpha_3) = \mathrm{e}^{-\alpha_3 [\tilde{H}_x, [\tilde{H}_x, [\tilde{H}_x, \tilde{H}_c]]]}$$

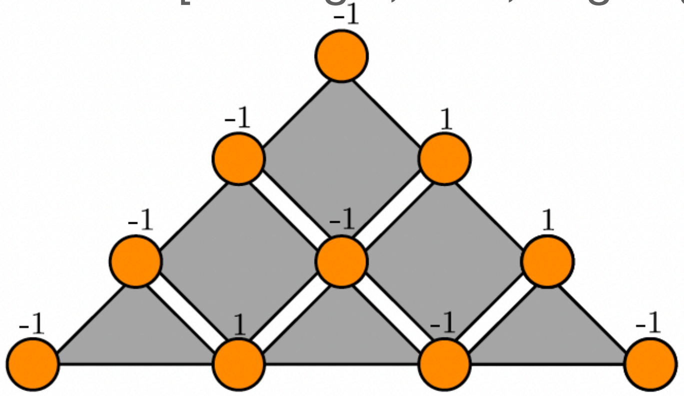
What is the optimal gate sequence?

RL for gate sequences

[Haslinger, et al, ongoing work]

Problem configuration:

- 10 parity qubits
- Local fields $\tilde{J}_i \in \{-1, 1\}$



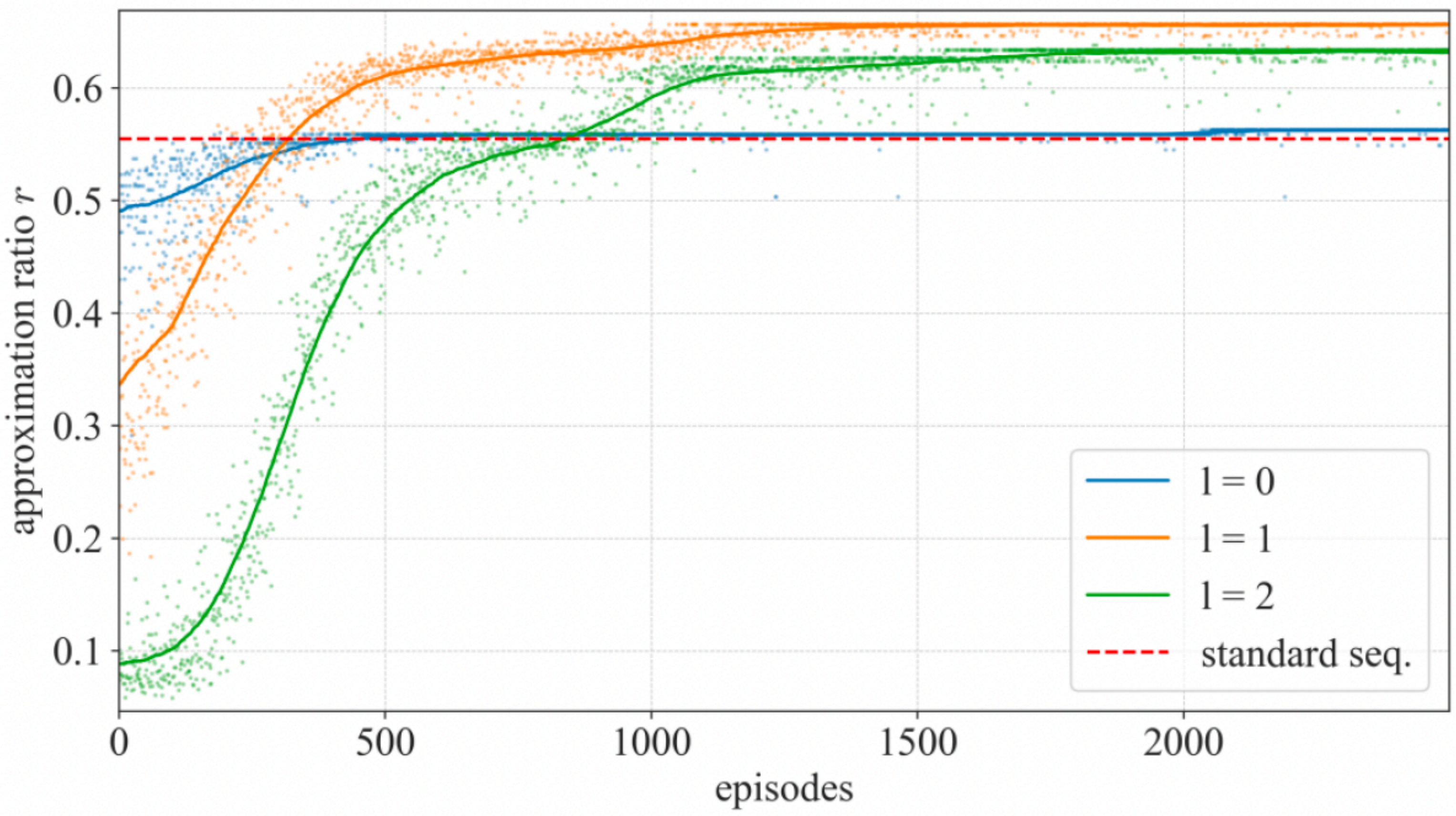
- Learning curves averaged over 10 agents for three gate pools

l = 0: $\{\hat{X}, \hat{Z}, \hat{C}\}$

l = 1: $\{\hat{X}, \hat{Z}, \hat{C}, \hat{U}_1, \hat{U}_2\}$

l = 2: $\{\hat{X}, \hat{Z}, \hat{C}, \hat{U}_1, \hat{U}_2, \hat{U}_3, \hat{U}_4, \hat{U}_5, \hat{U}_6, \hat{U}_7\}$

- Max. gate length $q = 6$
- Energy of discovered sequence is stored

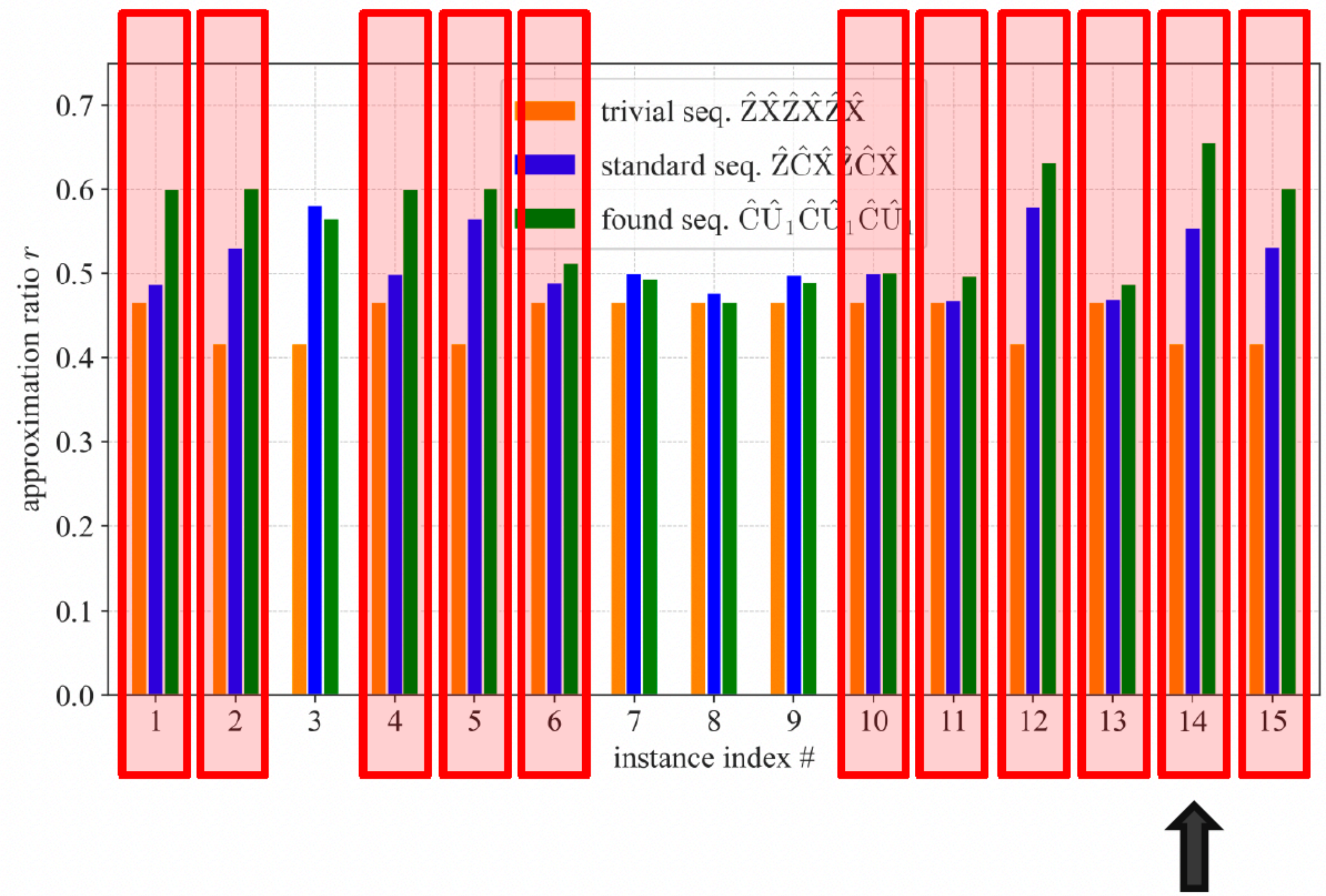


universität innsbruck

Performace across different istances

- 15 hard problem instances
- 100 random QAOA initializations

→ Found sequence outperforms standard sequence in 11 out of 15 problem instances



Summary

- The presence of **barren plateaus** hinders the **performance of QAOA**
- The **reinforcement learning** optimization converges to smooth optimal schedules, avoiding barren plateaus
- Few observables needed for the learning process
- Optimal gates and parameters can be **transferred among different system sizes**