Title: Reinforcement Learning assisted Quantum Optimization

Speakers: Matteo Wauters

Series: Machine Learning Initiative

Date: June 05, 2020 - 11:00 AM

URL: http://pirsa.org/20060019

Abstract: We propose a reinforcement learning (RL) scheme for feedback quantum control within the quantum approximate optimization algorithm (QAOA). QAOA requires a variational minimization for states constructed by applying a sequence of unitary operators, depending on parameters living in a highly dimensional space. We reformulate such a minimum search as a learning task, where a RL agent chooses the control parameters for the unitaries, given partial information on the system. We show that our RL scheme learns a policy converging to the optimal adiabatic solution for QAOA found by Mbeng et al. arXiv:1906.08948 for the translationally invariant quantum Ising chain. In presence of disorder, we show that our RL scheme allows the training part to be performed on small samples, and transferred successfully on larger systems. Finally, we discuss QAOA on the p-spsin model and how its robustness is enhanced by reinforce learning. Despite the possibility of finding the ground state with polynomial resources even in the presence of a first order phase transition, local optimizations in the p-spsin model suffer from the presence of many minima in the energy landscape. RL helps to find regular solutions that can be generalized to larger systems and make the optimization less sensitive to noise.

References

https://arxiv.org/abs/2003.07419

https://arxiv.org/abs/2004.12323

Reinforcement Learning assisted Quantum Optimization

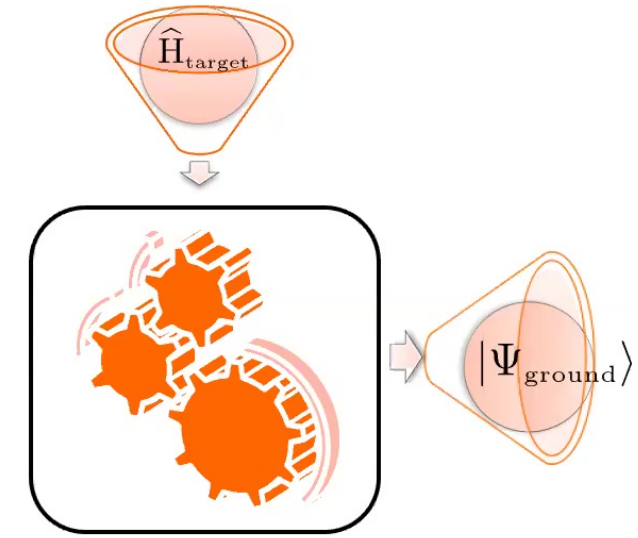
[Preliminary account in: arXiv:2004.12323]

SISSA

Matteo M. Wauters

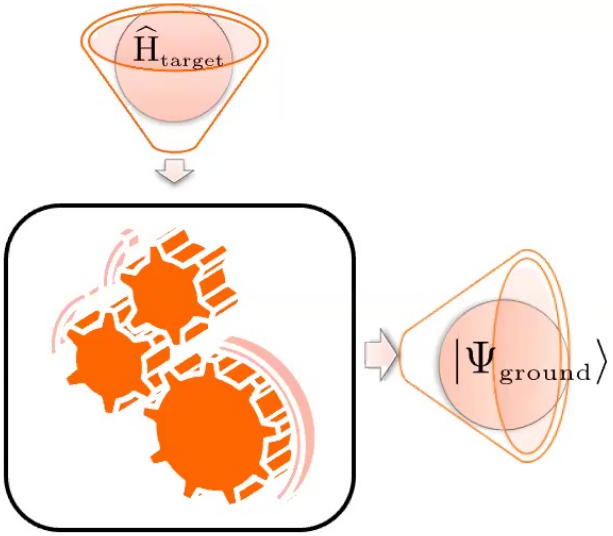
SISSA, Trieste

The dream of an H-to-Psi machine



Quantum
or
Hybrid
(Quantum+Classical)

The dream of an H-to-Psi machine



Quantum
or
Hybrid
(Quantum+Classical)

- **Material/Drug design**
- **Condensed Matter problems**

- **Classical Optimization problems (SAT, TSP,…)**

Algorithms:

- **Adiabatic (QA/AQC) ⇒ continuous evolution**

- **Digital (QAOA, Bang-Bang) ⇒ discrete, stepwise evolution**

Continuous or discrete?

$$\widehat{\mathrm{H}}_{\mathrm{target}} = \widehat{H}_z = \sum_{\langle i,j \rangle} J_{ij}\, \hat{\sigma}_i^z \hat{\sigma}_j^z \qquad \widehat{\mathrm{H}}_{\mathrm{driving}} = \widehat{H}_x = -\sum_i \hat{\sigma}_i^x$$

$$|\psi_0\rangle = |+\rangle = \frac{1}{\sqrt{2^n}} \prod_{i=1}^{n} (|\uparrow\rangle_i + |\downarrow\rangle_i)$$

Equal weight quantum superposition

Continuous or discrete?

$$\widehat{\mathrm{H}}_{\mathrm{target}} = \widehat{H}_z = \sum_{\langle i,j \rangle} J_{ij}\, \hat{\sigma}_i^z \hat{\sigma}_j^z \qquad \widehat{\mathrm{H}}_{\mathrm{driving}} = \widehat{H}_x = -\sum_i \hat{\sigma}_i^x$$

$$|\psi_0\rangle = |+\rangle = \frac{1}{\sqrt{2^n}} \prod_{i=1}^{n} (|\uparrow\rangle_i + |\downarrow\rangle_i)$$

Equal weight quantum superposition

Quantum annealing/adiabatic quantum computing

$$\widehat{H}(t) = s(t)\widehat{H}_z + (1 - s(t))\widehat{H}_x$$

$$s(t) \in [0,1], \quad s(0) = 0, \quad s(\tau) = 1$$

Total evolution time $\tau \gg 1$

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\psi(t)\rangle = \widehat{H}(s(t)) |\psi(t)\rangle$$

Adiabatic theorem $\Rightarrow |\psi(\tau)\rangle \simeq$ GS of $\widehat{H}_z$

✓ Convergence guaranteed in $\tau \to \infty$ limit

Continuous or discrete?

$$\widehat{\mathrm{H}}_{\mathrm{target}} = \widehat{H}_z = \sum_{\langle i,j \rangle} J_{ij}\, \hat{\sigma}_i^z \hat{\sigma}_j^z \qquad \widehat{\mathrm{H}}_{\mathrm{driving}} = \widehat{H}_x = -\sum_i \hat{\sigma}_i^x$$

$$|\psi_0\rangle = |+\rangle = \frac{1}{\sqrt{2^n}} \prod_{i=1}^{n} (|\uparrow\rangle_i + |\downarrow\rangle_i)$$

Equal weight quantum superposition

Quantum annealing/adiabatic quantum computing

$$\widehat{H}(t) = s(t)\widehat{H}_z + (1 - s(t))\widehat{H}_x$$

$$s(t) \in [0,1], \quad s(0) = 0, \quad s(\tau) = 1$$

Total evolution time $\tau \gg 1$

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\psi(t)\rangle = \widehat{H}(s(t)) |\psi(t)\rangle$$

Adiabatic theorem $\Rightarrow |\psi(\tau)\rangle \simeq$ GS of $\widehat{H}_z$

- ✓ Convergence guaranteed in $\tau \to \infty$ limit
- ✓ Easy to implement

- ✗ Sensitive to small gaps
- ✗ Hard schedule optimization
- ✗ Time dependent $|\psi(t)\rangle$ hard to compute

$$|\psi(\tau)\rangle = \mathrm{Texp}\left(-\frac{i}{\hbar}\int_0^\tau \mathrm{d}t'\, \widehat{H}(s(t'))\right)|\psi_0\rangle$$

Continuous or discrete?

Stepwise evolution + Trotter decomposition

$$|\psi(\tau)\rangle = \mathrm{Texp}\left(-\frac{i}{\hbar}\int_0^\tau \mathrm{d}t'\,\widehat{H}(s(t'))\right)|\psi_0\rangle \quad \Rightarrow \quad |\psi(\tau)\rangle_{\mathrm{step}} = \prod_{m=1}^{\leftarrow \mathrm{P}} \mathrm{e}^{-\frac{i}{\hbar}\widehat{H}(s_m)\Delta t_m}|\psi_0\rangle$$

$$\mathrm{e}^{-\frac{i}{\hbar}\widehat{H}(s_m)\Delta t_m} \longrightarrow \mathrm{e}^{-i\beta_m \widehat{H}_x}\mathrm{e}^{-i\gamma_m \widehat{H}_z} + O(\Delta t_m^2)$$

Continuous or discrete?

Stepwise evolution + Trotter decomposition

$$|\psi(\tau)\rangle = \mathrm{Texp}\left(-\frac{i}{\hbar}\int_0^\tau \mathrm{d}t'\,\widehat{H}(s(t'))\right)|\psi_0\rangle \Rightarrow |\psi(\tau)\rangle_{\mathrm{step}} = \prod_{m=1}^{\leftarrow \mathrm{P}} \mathrm{e}^{-\frac{i}{\hbar}\widehat{H}(s_m)\Delta t_m}|\psi_0\rangle$$

$$\mathrm{e}^{-\frac{i}{\hbar}\widehat{H}(s_m)\Delta t_m} \longrightarrow \mathrm{e}^{-i\beta_m \widehat{H}_x}\mathrm{e}^{-i\gamma_m \widehat{H}_z} + O(\Delta t_m^2)$$

digital-QA: from $s(t)$ to s_m

$$\begin{cases} \gamma_m = s_m \dfrac{\Delta t_m}{\hbar} \\ \beta_m = (1-s_m)\dfrac{\Delta t_m}{\hbar} \end{cases} \Rightarrow \begin{cases} \dfrac{\Delta t_m}{\hbar} = \gamma_m + \beta_m \\ s_m = \dfrac{\gamma_m}{\gamma_m + \beta_m} \end{cases}$$

QAOA: γ_m and β_m are variational parameters to optimize with classical methods.

Minimize $\langle\psi(\boldsymbol{\gamma},\boldsymbol{\beta})|\widehat{H}_z|\psi(\boldsymbol{\gamma},\boldsymbol{\beta})\rangle \Rightarrow \boldsymbol{\gamma}^*, \boldsymbol{\beta}^*$

QAOA

✓ Insensitive to small gaps

Continuous or discrete?

Stepwise evolution + Trotter decomposition

$$|\psi(\tau)\rangle = \mathrm{Texp}\left(-\frac{i}{\hbar}\int_0^\tau \mathrm{d}t'\,\widehat{H}(s(t'))\right)|\psi_0\rangle \Rightarrow |\psi(\tau)\rangle_{\mathrm{step}} = \prod_{m=1}^{\leftarrow \mathrm{P}} \mathrm{e}^{-\frac{i}{\hbar}\widehat{H}(s_m)\Delta t_m}|\psi_0\rangle$$

$$\mathrm{e}^{-\frac{i}{\hbar}\widehat{H}(s_m)\Delta t_m} \longrightarrow \mathrm{e}^{-i\beta_m \widehat{H}_x}\mathrm{e}^{-i\gamma_m \widehat{H}_z} + O(\Delta t_m^2)$$

digital-QA: from $s(t)$ to s_m

$$\begin{cases} \gamma_m = s_m \dfrac{\Delta t_m}{\hbar} \\ \beta_m = (1-s_m)\dfrac{\Delta t_m}{\hbar} \end{cases} \Rightarrow \begin{cases} \dfrac{\Delta t_m}{\hbar} = \gamma_m + \beta_m \\ s_m = \dfrac{\gamma_m}{\gamma_m + \beta_m} \end{cases}$$

QAOA: γ_m and β_m are variational parameters to optimize with classical methods.
Minimize $\langle\psi(\boldsymbol{\gamma},\boldsymbol{\beta})|\widehat{H}_z|\psi(\boldsymbol{\gamma},\boldsymbol{\beta})\rangle \Rightarrow \boldsymbol{\gamma}^*, \boldsymbol{\beta}^*$

QAOA

- ✓ Insensitive to small gaps
- ✓ Easier schedule control (also dQA)
- ✓ More general than dQA

- ✗ Local minima in parameter space
- ✗ Requires many measures for local optimization

Continuous or discrete?

Stepwise evolution + Trotter decomposition

$$|\psi(\tau)\rangle = \mathrm{Texp}\left(-\frac{i}{\hbar}\int_0^\tau \mathrm{d}t'\,\widehat{H}(s(t'))\right)|\psi_0\rangle \quad \Rightarrow \quad |\psi(\tau)\rangle_{\mathrm{step}} = \prod_{m=1}^{\leftarrow \mathrm{P}} \mathrm{e}^{-\frac{i}{\hbar}\widehat{H}(s_m)\Delta t_m}|\psi_0\rangle$$

$$\mathrm{e}^{-\frac{i}{\hbar}\widehat{H}(s_m)\Delta t_m} \longrightarrow \mathrm{e}^{-i\beta_m \widehat{H}_x}\mathrm{e}^{-i\gamma_m \widehat{H}_z} + O(\Delta t_m^2)$$

digital-QA: from $s(t)$ to s_m

$$\begin{cases} \gamma_m = s_m \dfrac{\Delta t_m}{\hbar} \\ \beta_m = (1-s_m)\dfrac{\Delta t_m}{\hbar} \end{cases} \Rightarrow \begin{cases} \dfrac{\Delta t_m}{\hbar} = \gamma_m + \beta_m \\ s_m = \dfrac{\gamma_m}{\gamma_m + \beta_m} \end{cases}$$

QAOA: γ_m and β_m are variational parameters to optimize with classical methods.

Minimize $\langle\psi(\boldsymbol{\gamma},\boldsymbol{\beta})|\widehat{H}_z|\psi(\boldsymbol{\gamma},\boldsymbol{\beta})\rangle \Rightarrow \boldsymbol{\gamma}^*, \boldsymbol{\beta}^*$

QAOA

- ✓ Insensitive to small gaps
- ✓ Easier schedule control (also dQA)
- ✓ More general than dQA

- ✗ Local minima in parameter space
- ✗ Requires many measures for local optimization
- ✗ Hard to improve systematically

Merging dQA and QAOA

- QAOA > dQA
- Exist optimal bang-gang schedule $s(t)$
  - Belong to QAOA *Ansatz*
  - Non-smooth, diabatic

PHYSICAL REVIEW X 7, 021027 (2017)

Optimizing Variational Quantum Algorithms Using Pontryagin's Minimum Principle

Zhi-Cheng Yang,[1] Armin Rahmani,[2,3] Alireza Shabani,[4] Hartmut Neven,[4] and Claudio Chamon[1]

Merging dQA and QAOA

- QAOA > dQA
- Exist optimal bang-gang schedule $s(t)$
  - Belong to QAOA *Ansatz*
  - Non-smooth, diabatic
- Iterative optimization of QAOA
  - Optimal dQA solution
  - Adiabatic

PHYSICAL REVIEW X 7, 021027 (2017)

Optimizing Variational Quantum Algorithms Using Pontryagin's Minimum Principle

Zhi-Cheng Yang,[1] Armin Rahmani,[2,3] Alireza Shabani,[4] Hartmut Neven,[4] and Claudio Chamon[1]

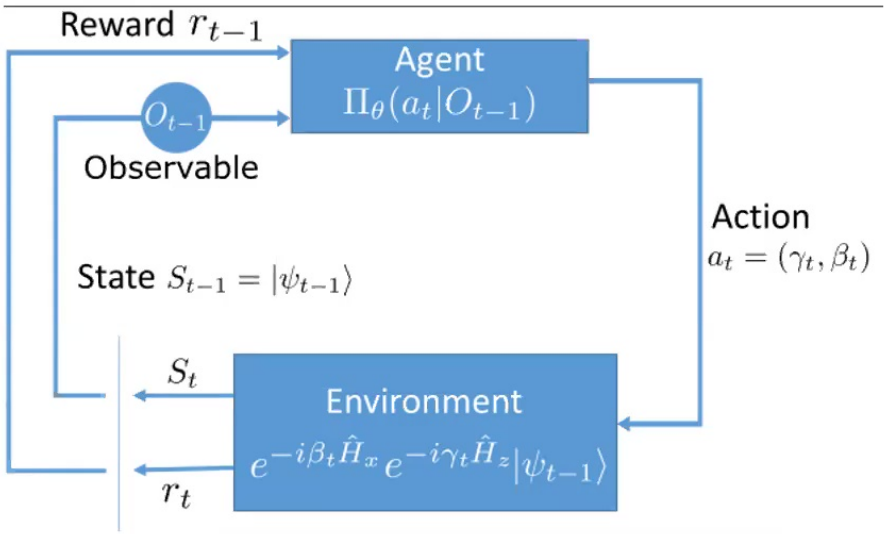
arXiv:1911.12259

Optimal quantum control with digitized Quantum Annealing

Glen Bigan Mbeng,[1,2] Rosario Fazio,[3,4] and Giuseppe E. Santoro[1,3,5]
[1]*SISSA, Via Bonomea 265, I-34136 Trieste, Italy*
[2]*INFN, Sezione di Trieste, I-34136 Trieste, Italy*
[3]*Abdus Salam ICTP, Strada Costiera 11, 34151 Trieste, Italy*
[4]*Dipartimento di Fisica, Università di Napoli "Federico II", Monte S. Angelo, I-80126 Napoli, Italy*
[5]*CNR-IOM Democritos National Simulation Center, Via Bonomea 265, I-34136 Trieste, Italy*

Our goal: Reinforcement Learning can find smooth optimal schedules

QAOA as a RL process

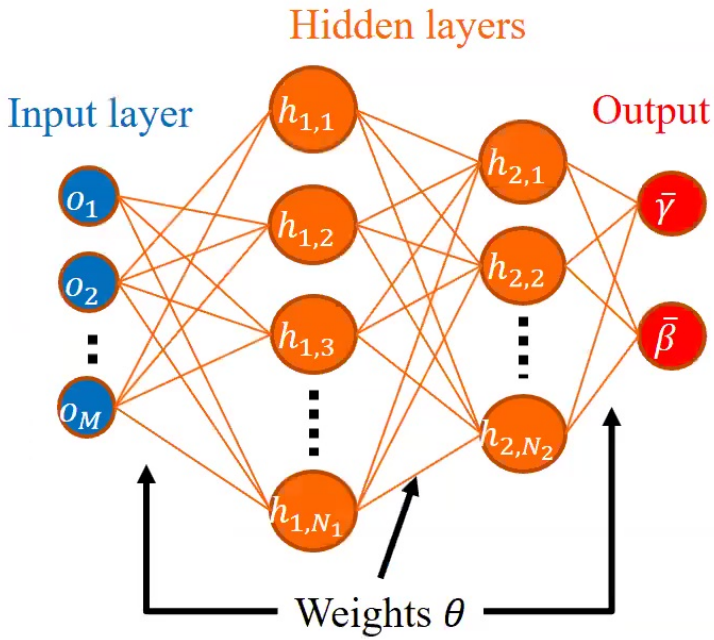


Goal:
Maximize reward

At each step $t = 1, \ldots, P$

- Measure $O_{t-1} = \langle\psi_{t-1}| \hat{O} |\psi_{t-1}\rangle$
- Compute action $a_t = (\gamma_t, \beta_t)$ from policy $\Pi_\theta(a|O)$
- Update state $|\psi_t\rangle = e^{-i\beta_t \hat{H}_x} e^{-i\gamma_t \hat{H}_z} |\psi_{t-1}\rangle$
- Compute new reward r_t: measure quality of variational Ansatz

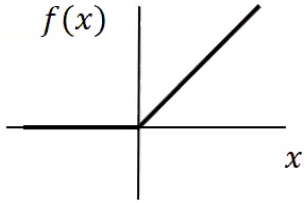
Choosing the policy



Gaussian policy

$$\Pi_\theta(\gamma|o) = e^{-\frac{(\gamma-\bar{\gamma})^2}{2\sigma_\gamma^2}}$$

$$\Pi_\theta(\beta|o) = e^{-\frac{(\beta-\bar{\beta})^2}{2\sigma_\beta^2}}$$

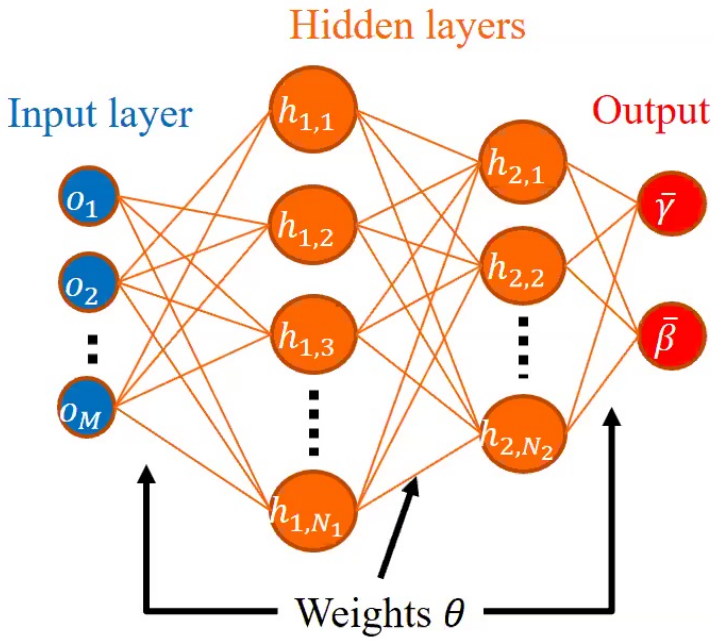


Our choice:

$O_t = \left(\langle\psi_t|\hat{H}_x|\psi_t\rangle, \langle\psi_t|\hat{H}_z|\psi_t\rangle\right)$

NN: 32,16 fully connected hidden, ReLu activation function

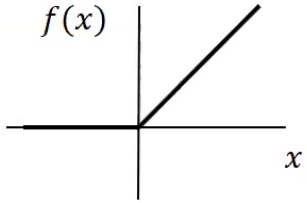
Choosing the policy



Gaussian policy

$$\Pi_\theta(\gamma|o) = e^{-\frac{(\gamma-\bar{\gamma})^2}{2\sigma_\gamma^2}}$$

$$\Pi_\theta(\beta|o) = e^{-\frac{(\beta-\bar{\beta})^2}{2\sigma_\beta^2}}$$



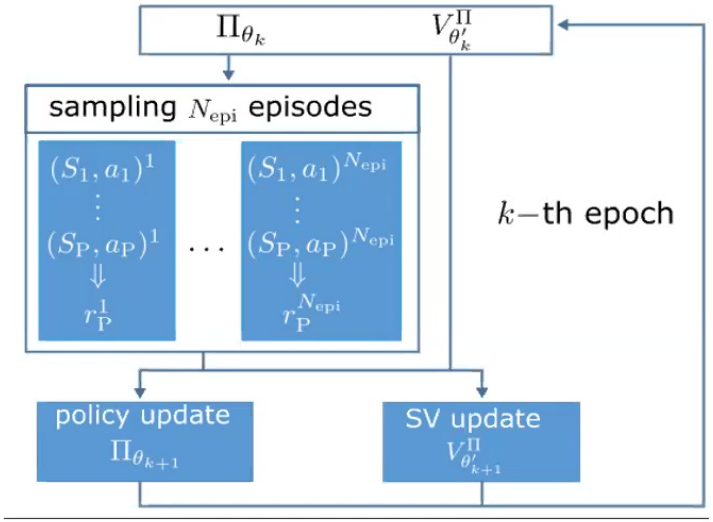
Our choice:

$O_t = \left(\langle\psi_t|\hat{H}_x|\psi_t\rangle, \langle\psi_t|\hat{H}_z|\psi_t\rangle\right)$

NN: 32,16 fully connected hidden, ReLu activation function

https://spinningup.openai.com

Learning process



Proximal Policy Optimization (PPO) algorithm from OpenAI Spinningup library

$k = 1, \ldots, N_{\text{epo}}$ epochs

In each epoch:

- sample N_{epi} digital evolutions (episodes)
- Update policy evaluation $V^{\Pi}_{\theta'}(o) = \mathbb{E}^{\Pi}(r_P | o_t = o)$
- Update policy Π_θ

Learning parameters

- N_{epi} =100 episodes per epoch
- $N_{\text{epo}} = 1024$ epochs
- Both policy $\Pi_\theta(a|o)$ and state-value function $V^{\Pi}_{\theta'}(o)$ parametrized by 32,16 NN
- Reward $r_t = -\delta_{t,P}\langle\psi_t|\widehat{H}_T|\psi_t\rangle$

Technical interlude: PPO details

NN updates:

- Compute gradients $\partial_\theta \Pi_\theta(a|o), \partial_{\theta'} V^\Pi_{\theta'}(o)$ ⇒ backpropagation (chain rule)
- Decide what to do with gradient ⇒ PPO algorithm

Value function $V^\Pi(o) = \mathbb{E}(R|o_0 = o)$

Expected future reward with policy Π, starting from o

Update: minimize $\left|V^\Pi(o) - V^\Pi_{\theta'}(o)\right|$

$R = r_1 + \lambda r_2 + \lambda^2 r_3 + \cdots$
$\lambda \leq 1$ discount factor
$V^\Pi(o)$ **real** VF, estimated from experience
$V^\Pi_{\theta'}(o)$ **approximated** by NN

Policy $\Pi_\theta(a|o)$: probability of taking action a when observing o

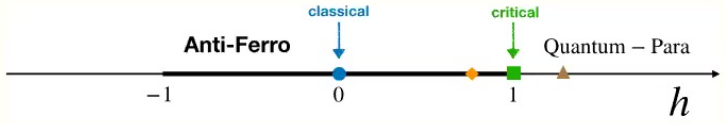
Given action $a : o \to o', r$ ⇒ define **advantage** $A^\Pi(o,a) = \left(r + V^\Pi(o')\right) - V^\Pi(o))$

Update: maximize $\frac{\Pi_\theta(a|o)}{\Pi_{\theta_k}(a|o)} A^{\Pi_{\theta_k}}(o,a)$

θ_k are the old parameters
Update is **clipped** to control maximum change

Quantum Ising chain

[... on an N-ring with PBC]



$$\widehat{H}_{\mathrm{T}} = \widehat{H}_z + h\widehat{H}_x$$

$$\widehat{H}_z = \sum_{j=1}^{N} \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z$$

$$\widehat{H}_x = -\sum_{j=1}^{N} \hat{\sigma}_j^x$$

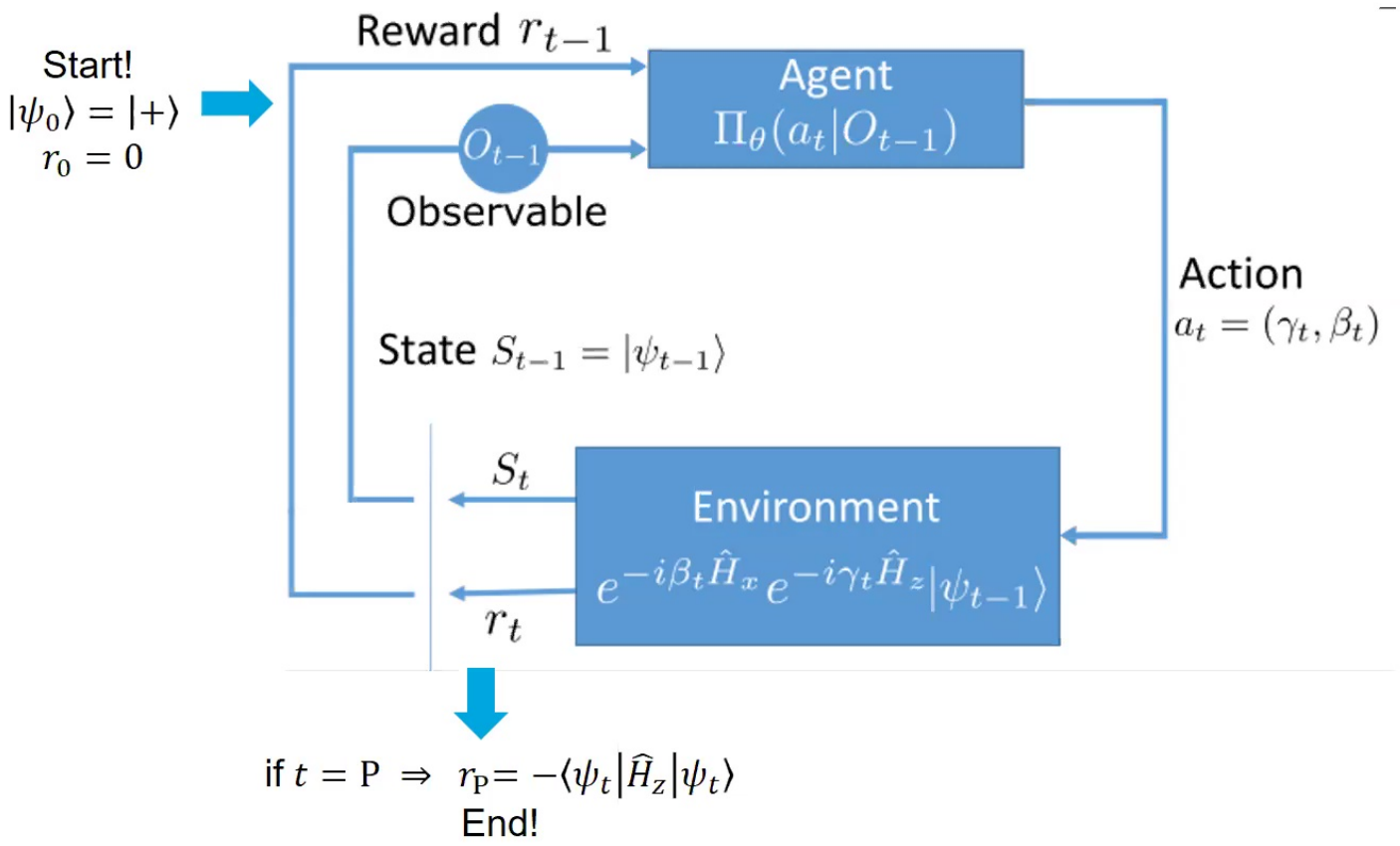
$$|\psi_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta})\rangle = \widehat{U}_{\mathrm{P}} \cdots \widehat{U}_1 |\psi_0\rangle \qquad \widehat{U}_m = \mathrm{e}^{-i\beta_m \widehat{H}_x} \mathrm{e}^{-i\gamma_m \widehat{H}_z}$$

$$E_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta}) = \langle\psi_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta})|\widehat{H}_{\mathrm{T}}|\psi_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta})\rangle \qquad \text{Reward } r_{\mathrm{P}} = -E_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta})$$

Residual energy

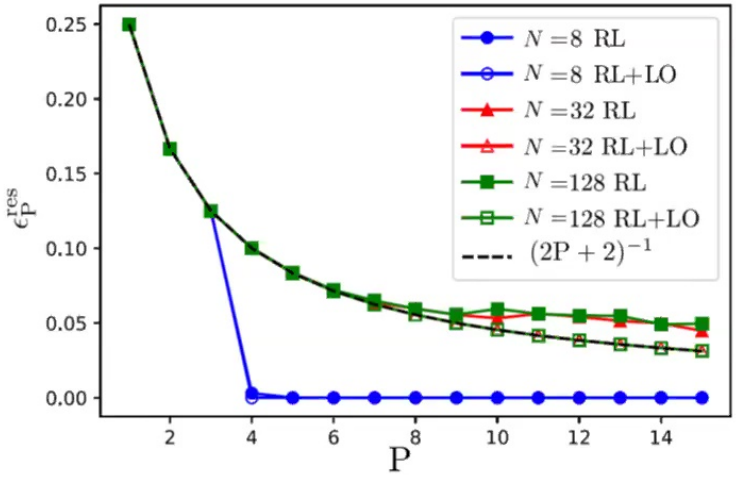
$$\epsilon_{\mathrm{P}}^{\mathrm{res}}(\boldsymbol{\gamma},\boldsymbol{\beta}) = \frac{E_{\mathrm{P}}(\boldsymbol{\gamma},\boldsymbol{\beta}) - E_{\mathrm{gs}}}{E_{\mathrm{max}} - E_{\mathrm{gs}}} \in [0,1]$$

Recap: QAOA as a RL process



Quantum Ising chain

[... on an N-ring with PBC]



Test of the trained agent

- $P \leq 7$: RL alone replicates QAOA solution.
- $P > 7$: RL follows QAOA solution but needs local optimization for exact match

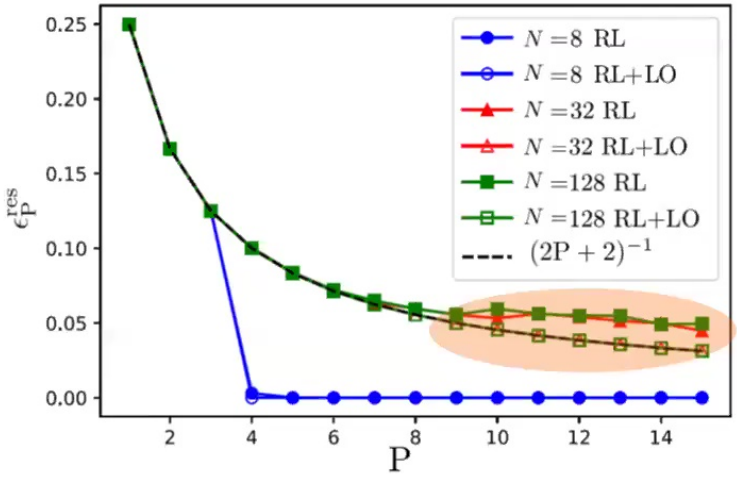
Lower bound for TFIM in arbitrary dimension, exact in 1d

$$\epsilon_P^{res}(\gamma,\beta) \geq \begin{cases} \frac{1}{2P+2} & 2P < N \\ 0 & 2P \geq N \end{cases}$$

Mbeng et al. arXiv:1906.08948

Quantum Ising chain

[... on an N-ring with PBC]



Lower bound for TFIM in arbitrary dimension, exact in 1d

$$\epsilon_P^{res}(\gamma,\beta) \geq \begin{cases} \frac{1}{2P+2} & 2P < N \\ 0 & 2P \geq N \end{cases}$$

Mbeng et al. arXiv:1906.08948

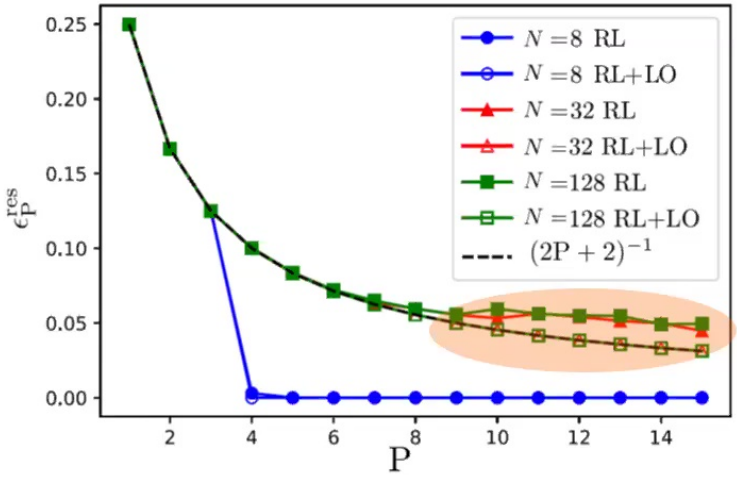
Test of the trained agent

- $P \leq 7$: RL alone replicates QAOA solution.
- $P > 7$: RL follows QAOA solution but needs local optimization for exact match

- Convergence in training epochs not reached
- Need of fine tuning of training parameters
- Stochastic policy ⇒ larger trajectory dispersion for longer episodes

Quantum Ising chain

[... on an N-ring with PBC]



Lower bound for TFIM in arbitrary dimension, exact in 1d

$$\epsilon_P^{res}(\gamma,\beta) \geq \begin{cases} \frac{1}{2P+2} & 2P < N \\ 0 & 2P \geq N \end{cases}$$

Mbeng et al. arXiv:1906.08948

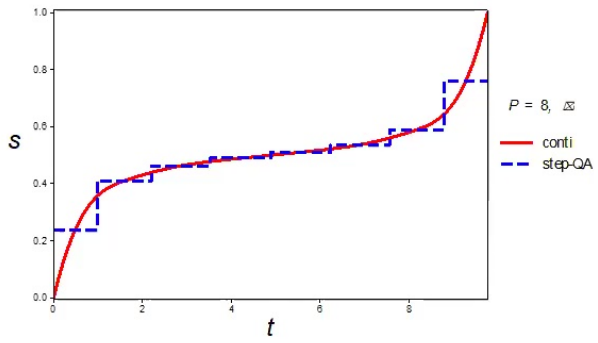
Test of the trained agent

- $P \leq 7$: RL alone replicates QAOA solution.
- $P > 7$: RL follows QAOA solution but needs local optimization for exact match

- Convergence in training epochs not reached
- Need of fine tuning of training parameters
- Stochastic policy ⇒ larger trajectory dispersion for longer episodes

Recap: QAOA and dQA

- QA/AQC: $\widehat{H}(t) = s(t)\widehat{H}_z + \big(1 - s(t)\big)\widehat{H}_x$
  - $s(t) \in [0,1]$
  - Evolution time τ

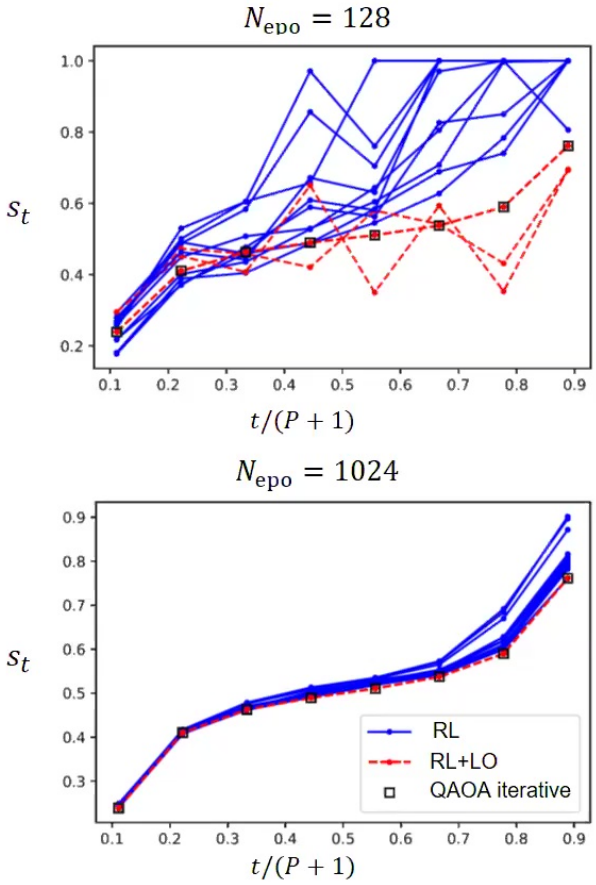


- dQA: $s(t) \to s_t$
  - $t = 1, \dots, \mathrm{P}$
  - $e^{-\frac{i}{\hbar}\widehat{H}_t \Delta t} \to e^{-\frac{i}{\hbar}\beta_t \widehat{H}_x} e^{-\frac{i}{\hbar}\gamma_t \widehat{H}_z}$
  - $s_t = \frac{\gamma_t}{\gamma_t + \beta_t}$

- QAOA: $(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*) = \arg\min_{(\boldsymbol{\gamma},\boldsymbol{\beta})} E_{\mathrm{P}}(\boldsymbol{\gamma}, \boldsymbol{\beta})$
  - Classical minimization in 2P-dimensional parameter space
  - $\exists\ (\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)$ such that s_t^* is the discretization of continuous schedule + adiabatic

Quantum Ising chain

[... on an N-ring with PBC]



Action choice analysis

$$s_t = \frac{\gamma_t}{\gamma_t + \beta_t}$$

Allows comparison with digital-QA schedule

$$\hat{H}_t = s_t \hat{H}_T + (1 - s_t)\hat{H}_x$$

Training epochs

i. Random action choices

ii. Large trajectory dispersion. RL+LO fall on the same smooth minimum (QAOA iter.)

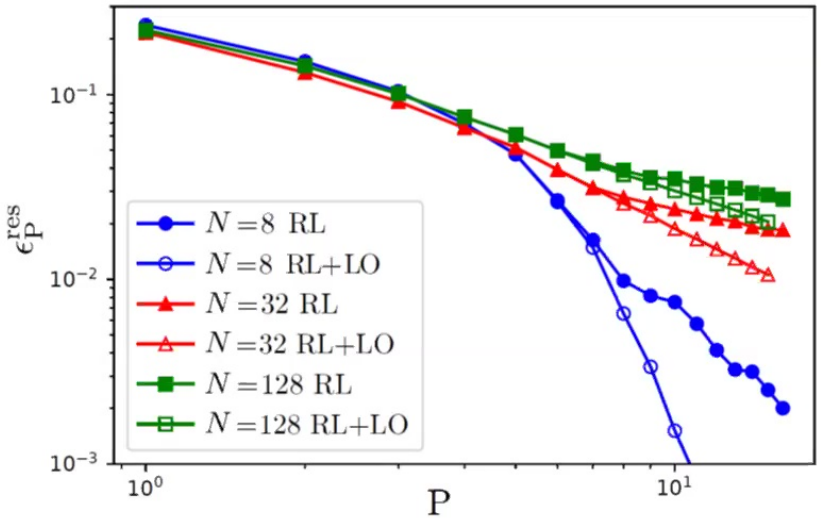
iii. RL converges to QAOA iterative solution

Quantum Ising chain

With random couplings

$$\hat{H}_z = \sum_{i=1}^{N} J_i \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z \quad \text{with } J_i \in [0,1], \text{ uniformly distributed}$$

Harder problem for optimization than uniform Ising model



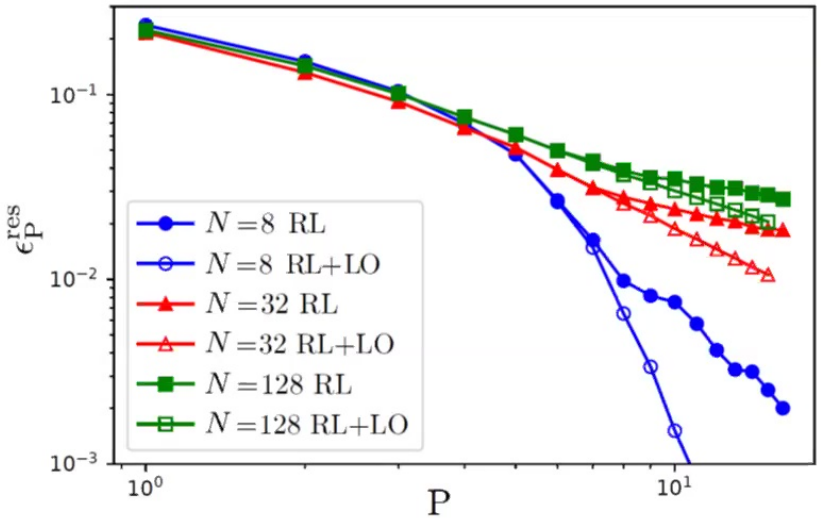
Same phenomenology of uniform TFIM: local optimization leads to lower energy for $P \geq 8$

Quantum Ising chain

With random couplings

$$\hat{H}_z = \sum_{i=1}^{N} J_i \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z \quad \text{with } J_i \in [0,1], \text{ uniformly distributed}$$

Harder problem for optimization than uniform Ising model



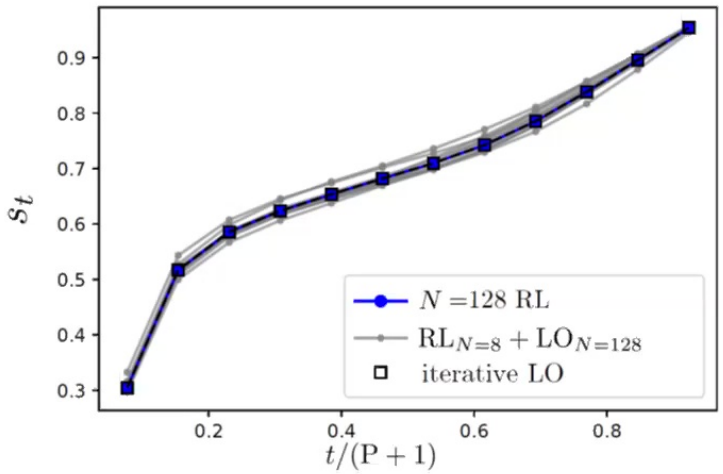
Same phenomenology of uniform TFIM: local optimization leads to lower energy for $P \geq 8$

Open issues:

- ? No analytical result
- ? Is solution optimal?
- ? Better than QA?

Quantum Ising chain

With random couplings



Same smooth parameter choice in RL and iterative local optimization

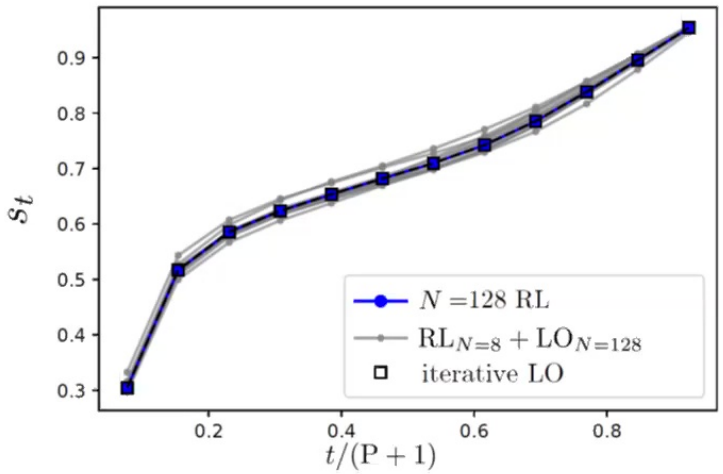
Policy transferability:

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

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

Quantum Ising chain

With random couplings



Same smooth parameter choice in RL and iterative local optimization

Policy transferability:

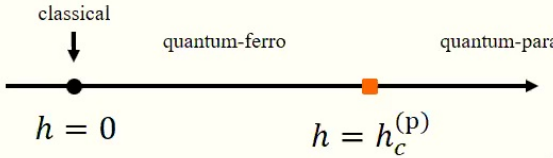
i. Train a small system ($N = 8$) on single disorder instance

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

iii. Use $(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)_{N=8}$ as ansatz to initialize a local optimization on larger system ($N = 128$) with any disorder instance

Possible solution to measurement problem

Fully connected Ising model



$$\widehat{H}_T = \widehat{H}_z + h\widehat{H}_x$$

$$\widehat{H}_z = -\frac{1}{N^{\mathrm{p}-1}}\left(\sum_{i=1}^{N}\hat{\sigma}_i^z\right)^{\mathrm{p}} \qquad \widehat{H}_x = -\sum_{j=1}^{N}\hat{\sigma}_j^x$$

p = 2

- Second order phase transition

  $h_c^{(2)} = 2$

- Minimum energy gap $\Delta \sim N^{-\frac{1}{3}}$
- Annealing time $\tau \sim N^{\frac{2}{3}}$

p > 2

- First order phase transition

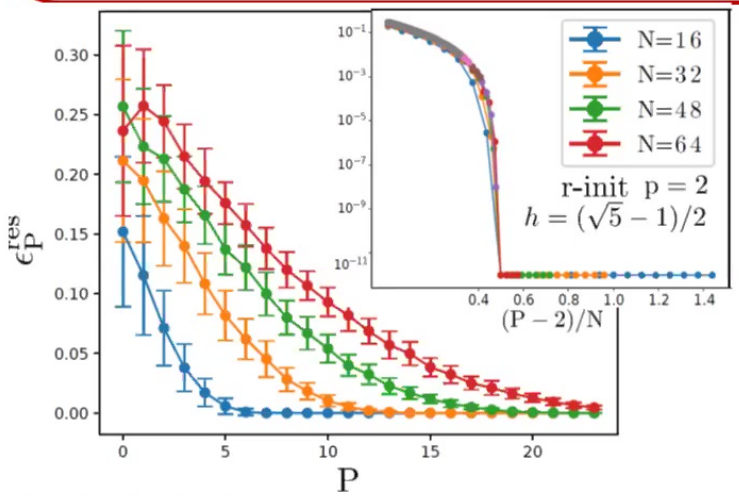
  $1 < h_c^{(\mathrm{p})} < 2$

- Minimum energy gap $\Delta \sim e^{-\alpha_{\mathrm{p}}N}$
- Annealing time $\tau \sim e^{2\alpha_{\mathrm{p}}N}$

Fully connected Ising model

arXiv:2003.07419

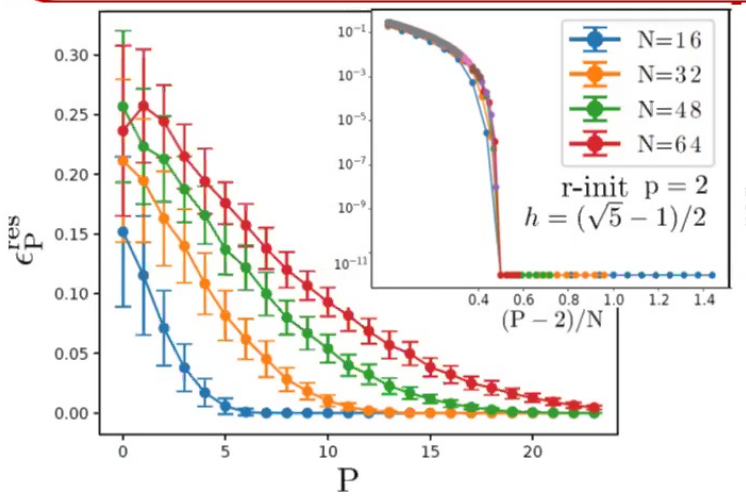
- Results with QAOA (local optimization) only:
- The $h = 0$ ground state $|\Uparrow\rangle + |\Downarrow\rangle$ prepared exactly with P=1 (if N odd) or P=2 (if N even) with $\gamma \sim N^{p-1}$
- Any state can be prepared exactly with $P = \frac{N}{2} + 2$ (p even) or $P = N + 1$ (p odd).

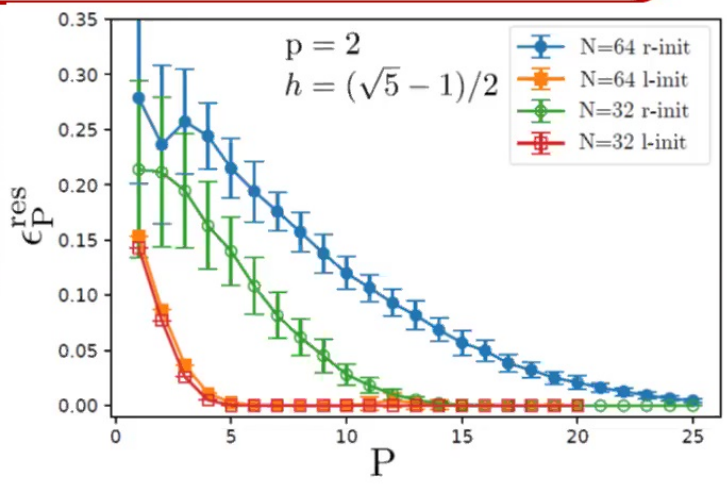


Fully connected Ising model

arXiv:2003.07419

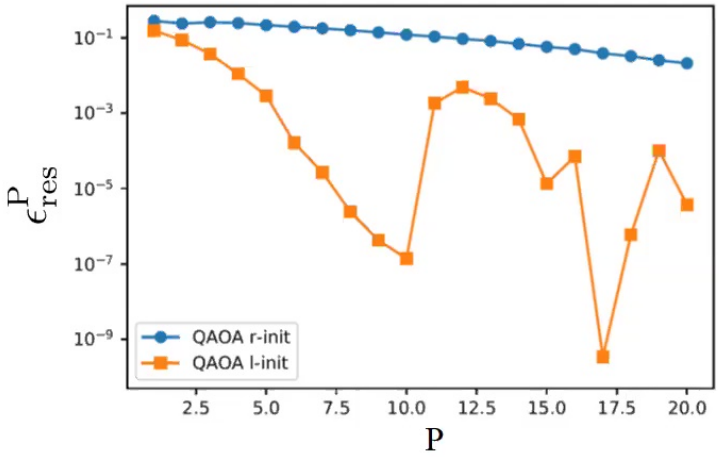
- Results with QAOA (local optimization) only:
- The $h = 0$ ground state $|\Uparrow\rangle + |\Downarrow\rangle$ prepared exactly with P=1 (if N odd) or P=2 (if N even) with $\gamma \sim N^{p-1}$
- Any state can be prepared exactly with $P = \frac{N}{2} + 2$ (p even) or $P = N + 1$ (p odd).
- Exact ground state with polynomial resources.
- Many local minima ⇒ local optimizations depend strongly on initial parameters.
- Hard to find smooth control parameters s_t.



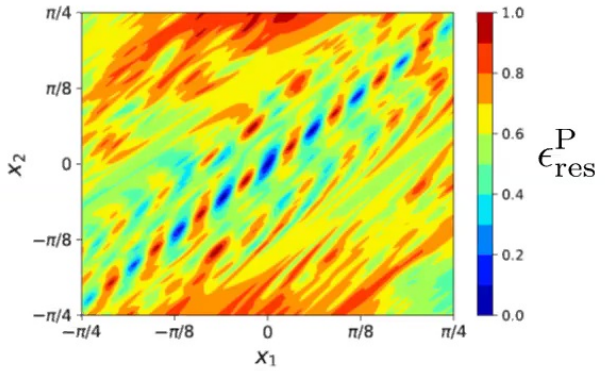


Fully connected Ising model

$$N = 64, \qquad h = \frac{\sqrt{5}-1}{2}$$



Rugged energy landscape
Strong dependence from initial parameter set

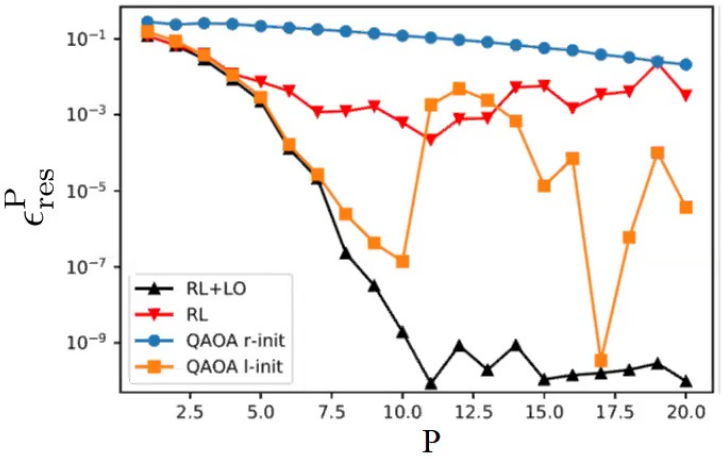


r-init $\quad (\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0) \in [0, \pi)$

l-init
$$\begin{cases} \gamma_t^0 = \dfrac{s_t \Delta t_t}{\hbar} \\ \beta_t^0 = \dfrac{\Delta t_t}{\hbar}\left(1 - s_t(1-h)\right) \end{cases}$$

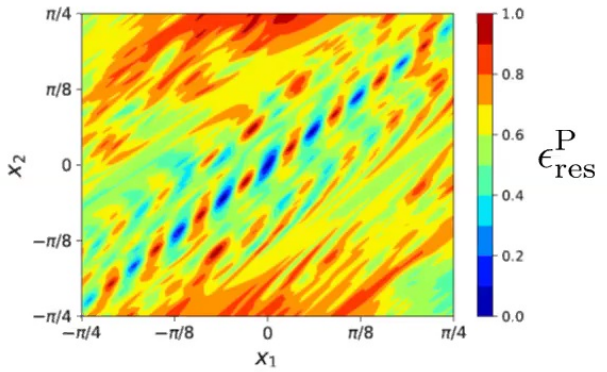
Fully connected Ising model

$$N = 64, \qquad h = \frac{\sqrt{5}-1}{2}$$



- RL good only for small P
- RL + LO improves and stabilizes results over local optimization alone

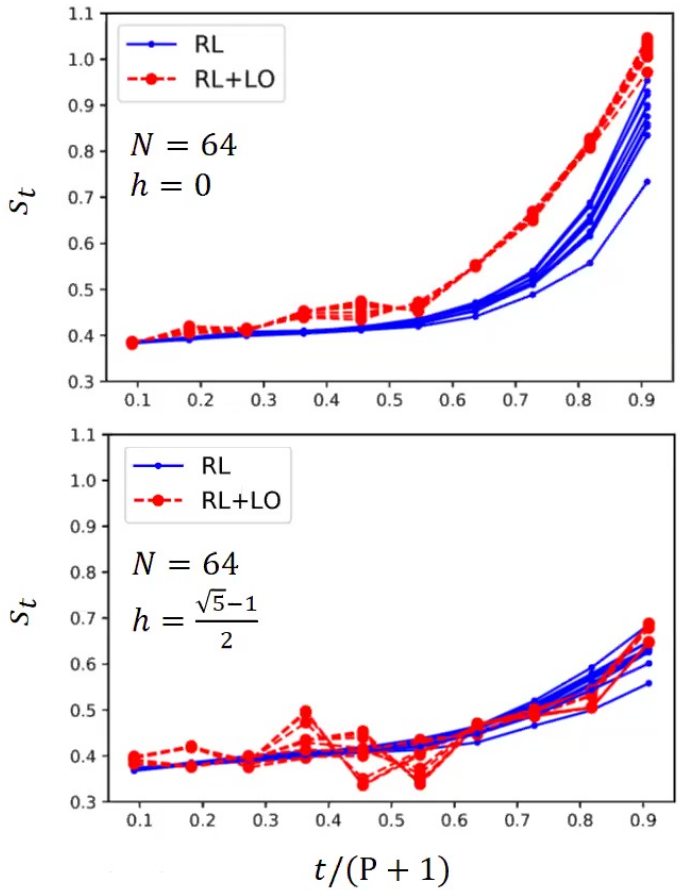
Rugged energy landscape
Strong dependence from initial parameter set



r-init $\quad (\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0) \in [0, \pi)$

l-init
$$\begin{cases} \gamma_t^0 = \dfrac{s_t \Delta t_t}{\hbar} \\ \beta_t^0 = \dfrac{\Delta t_t}{\hbar}\left(1 - s_t(1-h)\right) \end{cases}$$

Fully connected Ising model



Reinforcement learning

✓ suggests smooth action choices

✗ Low performance for large P

RL+LO

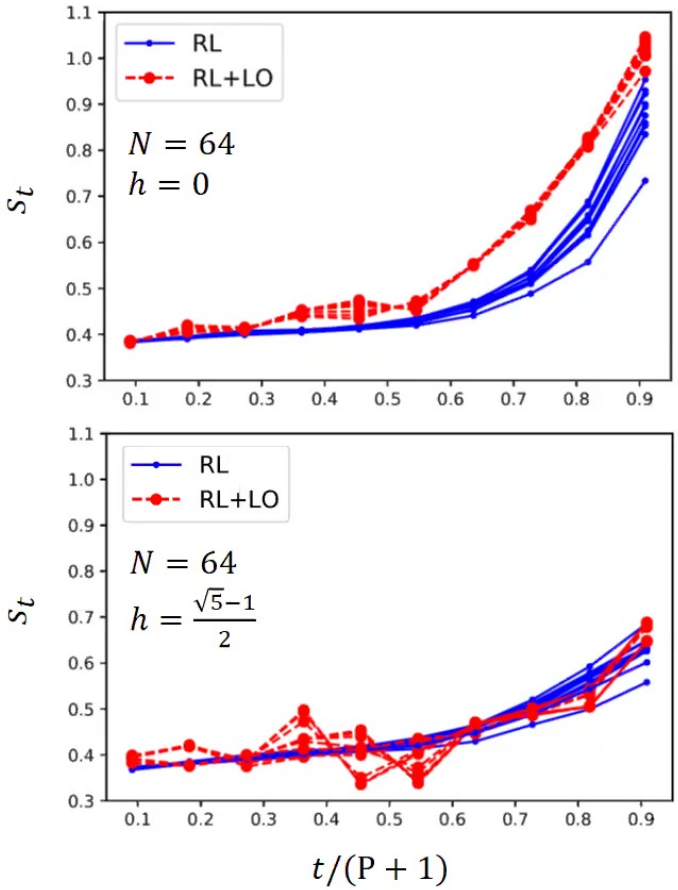
✓ dramatically enhances performance

✗ Optimized parameters less smooth

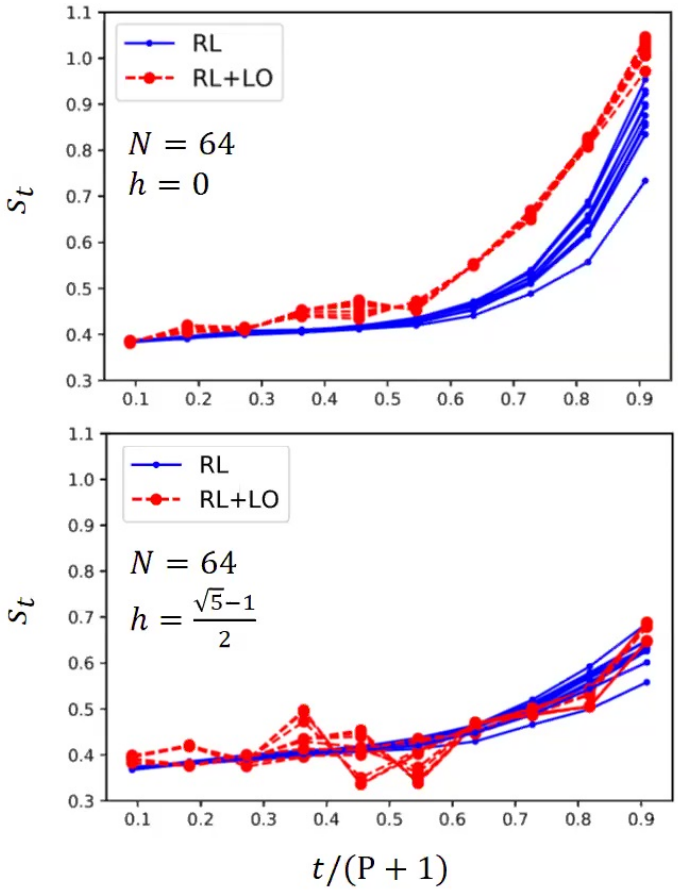
✗ Different local minima

- Is schedule adiabatic?
- Bad local minima due to non-local interactions?

Fully connected Ising model



Fully connected Ising model



Reinforcement learning

✓ suggests smooth action choices

✗ Low performance for large P

RL+LO

✓ dramatically enhances performance

✗ Optimized parameters less smooth

✗ Different local minima

- Is schedule adiabatic?
- Bad local minima due to non-local interactions?

Summary & outlook

- Parameters for QAOA ansatz can be learned with RL approach
  - ✓ Smooth action choices
  - ✓ Few observables needed
  - ✓ Converges to adiabatic optimal schedules
  - ✓ Strategies can be transferred among different system sizes
  - ✗ Requires observation on the system during the evolution

- Open issues
  - ? Solve measurement problem: weak measures, ancillary qubits,…?
  - ? Robustness to noise
  - ? Characterize collapse of learned policy during the training
  - ? More complex Hamiltonians (DMRG/MPS?)

Reinforcement Learning assisted Quantum Optimization

[Preliminary account in: arXiv:2004.12323]

Giuseppe E. Santoro
The boss

Thank you!

Emanuele Panizon
The RL guy

Glen B. Mbeng
The QAOA expert