Reinforcement Learning using Quantum Boltzmann Machines on the D-Wave System

Joint work with A. Levit, N. Ghadermarzi, J. S. Oberoi, E. Zahendinejad, and P. Ronagh (arXiv:1612.05695, 1706.00074)

Daniel Crawford | 1QBit

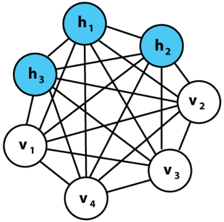
September 27, 2017

Outline

Reinforcement learning
Quantum Boltzmann machines
Quantum Monte Carlo simulations
Results

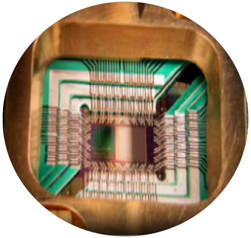
1QBit

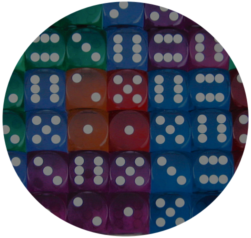
RL



GBM

QBM





QMC
Simulations

1QBit

Markov Decision Process

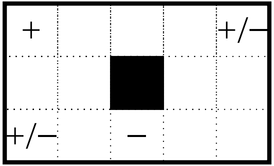
- Markov decision process

  states and actions: (finite) sets S and A;

  ***controlled* Markov chain:** defined by a transition kernel $\mathbb{P}(s' \in S | s \in S, a \in A)$;

  immediate reward structure: given via $r : S \times A \to \mathbb{R}$;

  discount factor: a constant $\gamma \in [0, 1)$.

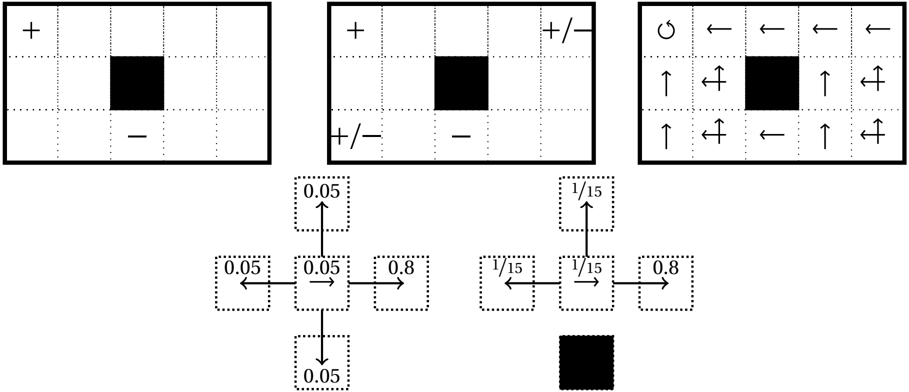


- The stationary policy

$$\pi : S \longrightarrow A$$

reduces the MDP into a time-homogeneous Markov chain, Π, with kernel $\mathbb{P}(s'|s, \pi(s))$.

Example: Grid-World Problem



Markov Decision Problem

- Goal: to solve the optimization problem

$$\pi^* = \underset{\pi}{\operatorname{argmax}}\, V(\pi, s).$$

- The discounted reward function

$$V(\pi, s) = \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i \, r\left(\Pi_i^s, \pi(\Pi_i^s)\right)\right].$$

- Bellman recursion:

$$\begin{aligned} V(\pi, s) &= \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i \, r\left(\Pi_i^s, \pi(\Pi_i^s)\right)\right] \\ &= \mathbb{E}[r\left(\Pi_0^s, \pi(\Pi_0^s)\right)] \\ &\quad + \gamma\, \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i \, r\left(\Pi_{i+1}^s, \pi(\Pi_{i+1}^s)\right)\right] \\ &= \mathbb{E}[r\left(s, \pi(s)\right)] + \gamma \sum_{s' \in S} \mathbb{P}(s' | s, \pi(s))\, V(\pi, s') \end{aligned}$$

1QBit

Q-function

- Maps a triplet (π, s, a) to the expected value of the reward of the Markov chain that begins with taking action a at initial state s and continuing according to π:

$$Q(\pi, s, a) = \mathbb{E}[r(s, a)] + \mathbb{E}\left[\sum_{i=1}^{\infty} \gamma^i \, r(\Pi_i^s, \pi(\Pi_i^s))\right].$$

- Reconstructs π^* and $V^*(s) = V(\pi^*, s)$:

$$V^*(s) = \max_a Q^*(s, a),$$

for $Q^*(s, a) = \max_\pi Q(\pi, s, a)$ and

$$\pi^*(s) = \operatorname*{argmax}_a Q^*(s, a).$$

1QBit

Parametrization of the Temporal Difference

- The goal is to minimize

$$Q_{n+1}(s,a) - Q_n(s,a) = \mathbb{E}[r(s,a)] + \gamma \sum_{s'} \mathbb{P}(s'|s,a) \max_a Q_n(s',a) - Q_n(s,a)$$

Parametrization of the Temporal Difference

- The goal is to minimize

$$Q_{n+1}(s,a) - Q_n(s,a) = \mathbb{E}[r(s,a)] + \gamma \sum_{s'} \mathbb{P}(s'|s,a) \max_a Q_n(s',a) - Q_n(s,a)$$

↳ temporal difference (E_{TD})

1QBit

Parametrization of the Temporal Difference

- The goal is to minimize

$$Q_{n+1}(s,a) - Q_n(s,a) = \mathbb{E}[r(s,a)] + \gamma \sum_{s'} \mathbb{P}(s'|s,a) \max_a Q_n(s',a) - Q_n(s,a)$$

↑ temporal difference (E_{TD})

- Local parametrization

$$Q(s,a) = Q(s,a;\boldsymbol{\theta})$$

1QBit

Parametrization of the Temporal Difference

- The goal is to minimize

$$Q_{n+1}(s,a) - Q_n(s,a) = \mathbb{E}[r(s,a)] + \gamma \sum_{s'} \mathbb{P}(s'|s,a) \max_a Q_n(s',a) - Q_n(s,a)$$

↑ temporal difference (E_{TD})

- Local parametrization

$$Q(s,a) = Q(s,a;\boldsymbol{\theta})$$

- Descent along

$$-\nabla_{\boldsymbol{\theta}} (E_{TD})^2 = -E_{\mathrm{TD}} \nabla_{\boldsymbol{\theta}} E_{\mathrm{TD}}$$

$$\xrightarrow{\text{SGD}} \left(r_n(s_n,a_n) + \gamma \max_{a_{n+1}} Q(s_{n+1},a_{n+1}) - Q(s_n,a_n) \right) \frac{\partial}{\partial \boldsymbol{\theta}} Q(s_n,a_n)$$

1QBit

Quantum Boltzmann Machines

GBMs, DBMs, and QBMs as function approximators

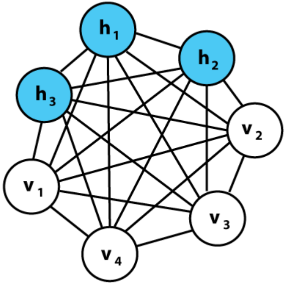
1QBit

Quantum Boltzmann Machines

Amin et al., 2016

- Classical GBM

$$\mathcal{E}(\mathrm{v},\mathrm{h}) = -\sum_{v\in V,\, h\in H} w^{vh} v h - \sum_{\{v,v'\}\subseteq V} w^{vv'} v v' - \sum_{\{h,h'\}\subseteq H} w^{hh'} h h'$$



1QBit

Quantum Boltzmann Machines

- Classical GBM

$$\mathscr{E}(\mathrm{v},\mathrm{h}) = -\sum_{v\in V,\, h\in H} w^{vh} v h - \sum_{\{v,v'\}\subseteq V} w^{vv'} v v' - \sum_{\{h,h'\}\subseteq H} w^{hh'} h h'$$

- Clamped GBM (fixed assignment v of the visible binary variables)

$$\mathscr{E}_{\mathrm{v}}(\mathrm{h}) = -\sum_{v\in V,\, h\in H} w^{vh} v h - \sum_{\{v,v'\}\subseteq V} w^{vv'} v v' - \sum_{\{h,h'\}\subseteq H} w^{hh'} h h'$$

- Clamped QBM

$$\mathscr{H}_{\mathrm{v}} = -\sum_{v\in V,\, h\in H} w^{vh} v \sigma^z_h - \sum_{\{v,v'\}\subseteq V} w^{vv'} v v' - \sum_{\{h,h'\}\subseteq H} w^{hh'} \sigma^z_h \sigma^z_{h'} - \Gamma \sum_{h\in H} \sigma^x_h$$

Free Energy of a QBM

- Equilibrium free energy

$$F(\mathrm{v}) := -\frac{1}{\beta} \ln Z_{\mathrm{v}} = \langle \mathscr{H}_{\mathrm{v}} \rangle + \frac{1}{\beta} \mathrm{tr}(\rho_{\mathrm{v}} \ln \rho_{\mathrm{v}}).$$

where $\beta = \frac{1}{K_B T}$, $Z_{\mathrm{v}} = \mathrm{tr}(e^{-\beta \mathscr{H}_{\mathrm{v}}})$, $\rho_{\mathrm{v}} = \frac{1}{Z_{\mathrm{v}}} e^{-\beta \mathscr{H}_{\mathrm{v}}}$

- Entropy $-\mathrm{tr}(\rho_{\mathrm{v}} \ln \rho_{\mathrm{v}})$
- Gibbs measure

$$\langle \mathscr{H}_{\mathrm{v}} \rangle = \frac{1}{Z_{\mathrm{v}}} \mathrm{tr}(\mathscr{H}_{\mathrm{v}} e^{-\beta \mathscr{H}_{\mathrm{v}}})$$

Free Energy as a Function Approximator

$$Q(s,a) \approx -F(\mathrm{s},\mathrm{a}) = -F(\mathrm{s},\mathrm{a};\boldsymbol{\omega}) \qquad \longleftarrow \text{QBM coupling strength}$$

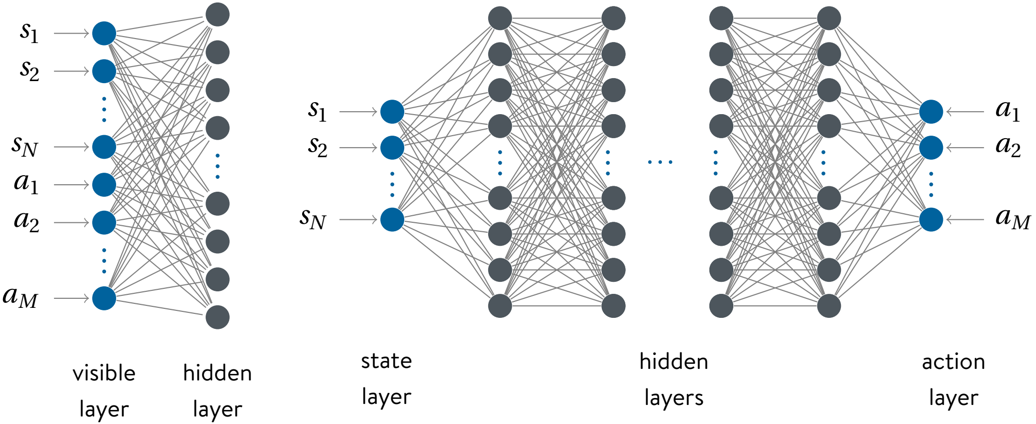
$$\Delta\omega = -\varepsilon(r(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a))\frac{\partial F}{\partial \omega}$$

$$\begin{aligned}\frac{\partial F(\mathrm{s},\mathrm{a})}{\partial \omega} &= \frac{1}{Z_{\mathrm{s,a}}}\frac{\partial}{\partial \omega}\mathrm{tr}(e^{-\beta \mathcal{H}_{\mathrm{s,a}}}) \\ &= -\frac{1}{Z_{\mathrm{s,a}}}\mathrm{tr}(\beta e^{-\beta \mathcal{H}_{\mathrm{s,a}}}\frac{\partial}{\partial \omega}\mathcal{H}_{\mathrm{s,a}}) \\ &= -\beta\left\langle \frac{\partial}{\partial \omega}\mathcal{H}_{\mathrm{s,a}}\right\rangle\end{aligned}$$

$$\begin{aligned}\Delta\omega^{vh} &= \varepsilon(r(s,a) \\ &\qquad -\gamma \min_{a'} F(s',a') + F(s,a))v\langle\sigma_h^z\rangle \\ \Delta\omega^{hh'} &= \varepsilon(r(s,a) \\ &\qquad -\gamma \min_{a'} F(s',a') + F(s,a))\langle\sigma_h^z\sigma_{h'}^z\rangle\end{aligned}$$

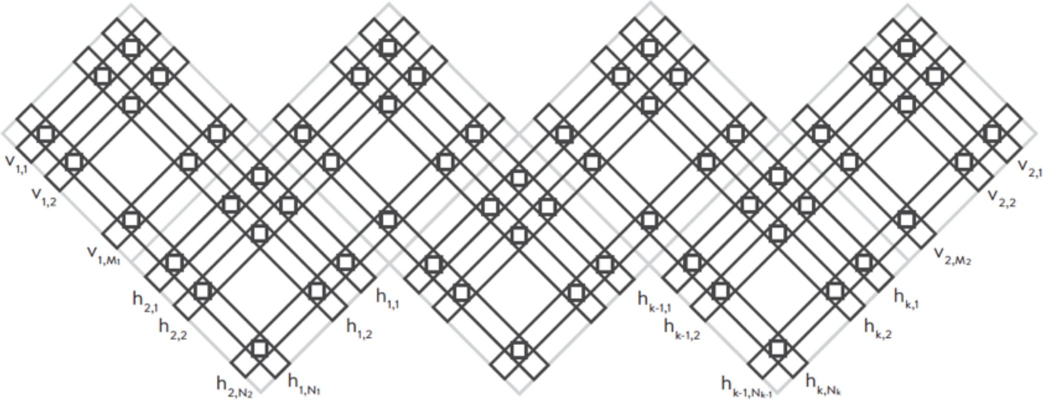
1QBit

RBM and DBM Layout



1QBit

A DBM as a Layout of Superconducting Qubits



1QBit

Quantum Monte Carlo Simulations

Approximating free energy and spin expectations

1QBit

Suzuki–Trotter Expansion

The effective Hamiltonian of an Ising model with transverse field

Transverse field Ising Hamiltonian:

$$H = -\sum_{(i,j)} J_{ij}\sigma_i^z\sigma_j^z - \Gamma\sum_{i=1}^{N}\sigma_i^x .$$

Suzuki–Trotter Expansion

The effective Hamiltonian of an Ising model with transverse field

Transverse field Ising Hamiltonian:

$$H = -\sum_{(i,j)} J_{ij} \sigma_i^z \sigma_j^z - \Gamma \sum_{i=1}^{N} \sigma_i^x .$$

The partition function is approximated as

$$Z \approx \left(\frac{1}{2} \sinh \frac{2\beta\Gamma}{M} \right)^{\frac{nM}{2}} \operatorname{tr}\left[e^{(-\beta \mathscr{H}_{\mathrm{eff}})} \right]$$

with an *effective* Hamiltonian

$$\mathscr{H}_{\mathrm{eff}}(\sigma) = -\sum_{i,j,k} \frac{J_{ij}}{M} \sigma_{ik} \sigma_{jk} - \sum_{i,k} \frac{1}{2\beta} \ln \coth\left(\frac{\beta\Gamma}{M} \right) \sigma_{ik} \sigma_{ik+1} .$$

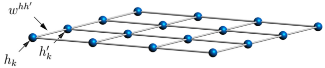
Approximating the Expected Values

Theorem [Suzuki, 1976] Spin correlations $\langle \sigma_i^z \sigma_j^z \rangle$ are calculated as those of the corresponding $(d+1)$-dimensional Ising model:

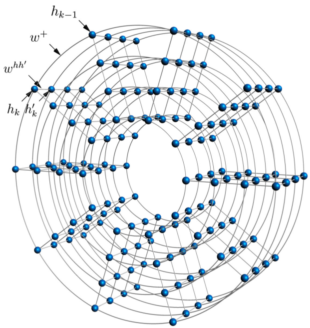
$$\langle \sigma_i^z \sigma_j^z \rangle = \frac{1}{Z_{d+1}} \lim_{n \to \infty} \sum_{\sigma} \sigma_i \sigma_j e^{\left(-\beta \mathscr{H}_{\text{eff}}(\sigma)\right)}.$$

1QBit

Simulated Quantum Annealing



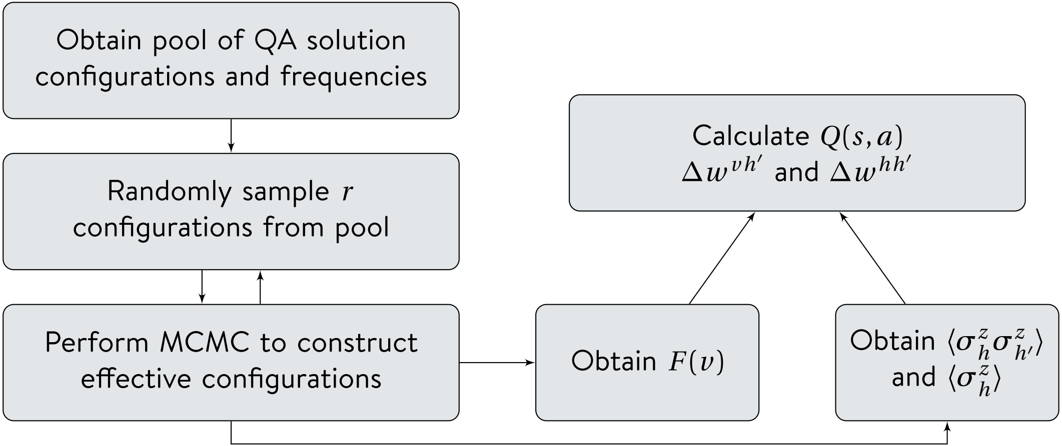
(a) Original transverse field Ising Hamiltonian



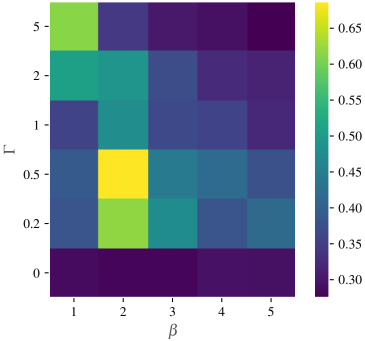
(b) Effective classical Hamiltonian in one dimension higher

The dynamics are governed by the MCMC method using Metropolis acceptance probabilities, as the transverse field strength Γ is slowly decreased to zero.

Using QA Samples to Build Effective Configurations



Issue: Effective Temperature and Transverse Field



$$\mathcal{H}_{\text{eff}}(\sigma) = -\sum_{i,j,k} \frac{J_{ij}}{M} \sigma_{ik} \sigma_{jk} - \sum_{i,k} \frac{1}{2\beta} \ln\coth\left(\frac{\beta\Gamma}{M}\right) \sigma_{ik} \sigma_{ik+1}$$

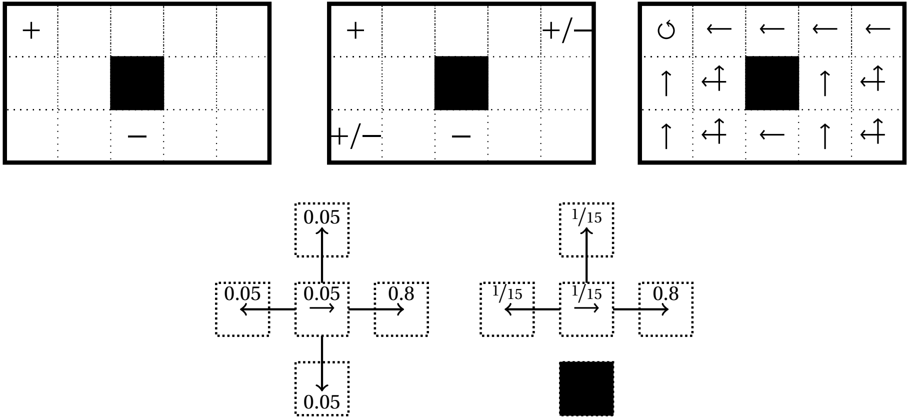
1QBit

Experimental Results

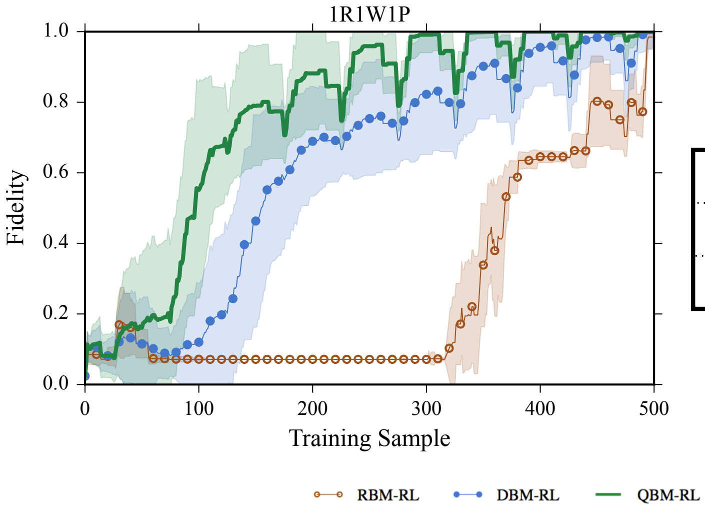
Reinforcement learning of a maze

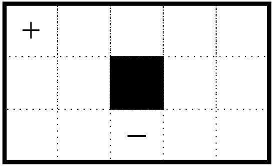
1QBit

Grid-World Problem



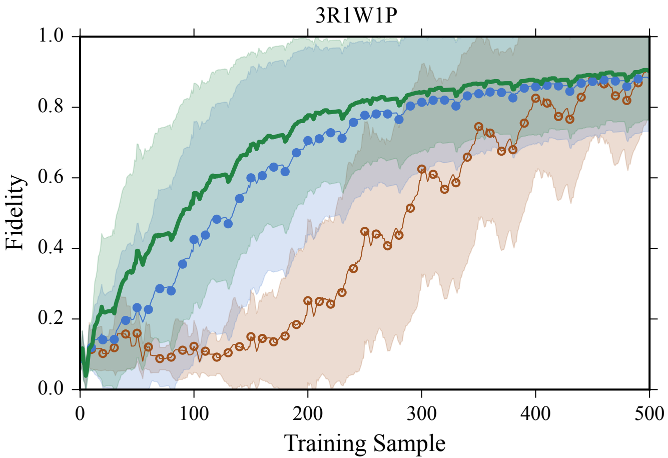
Results

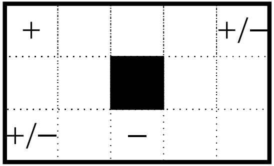




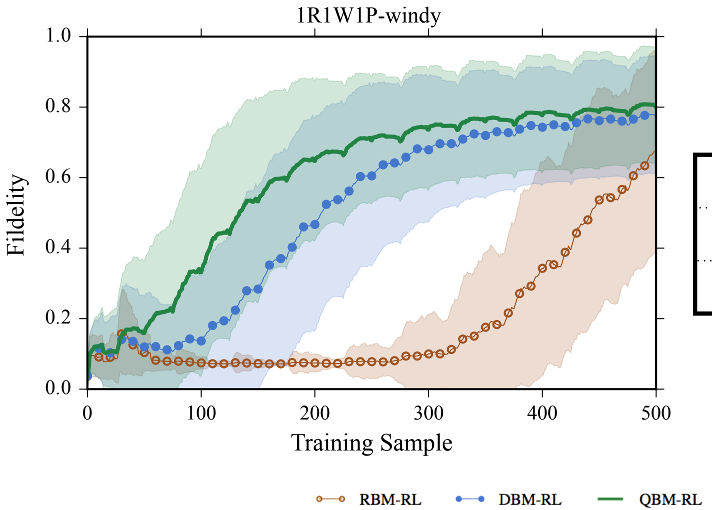
1QBit

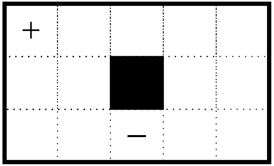
Results – Stochastic Rewards





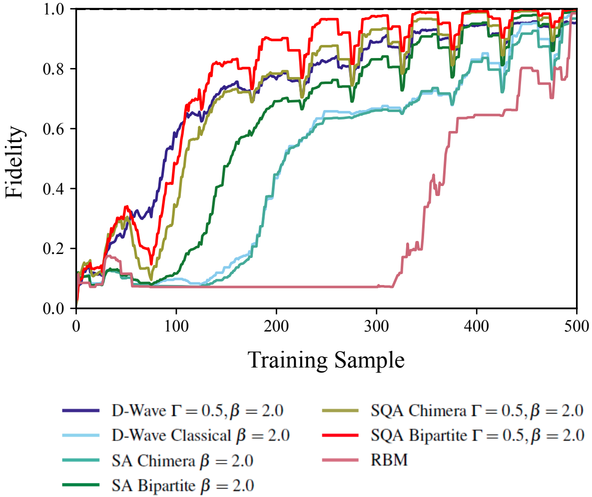
Results – Stochastic Transitions

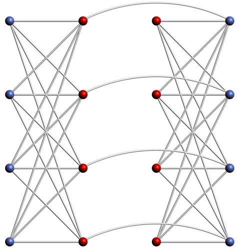




1QBit

Results





Chimera Graph

Thank you!

Daniel.Crawford@1qbit.com