Quantum codes from neural networks

Felix Leditzky

(JILA & CTQM, University of Colorado Boulder)

Joint work with Johannes Bausch

arXiv:1806.08781

Quantum Innovators workshop

IQC, University of Waterloo

23 October 2018

JILA
CU Boulder and NIST

Center for Theory of Quantum Matter
CTQM

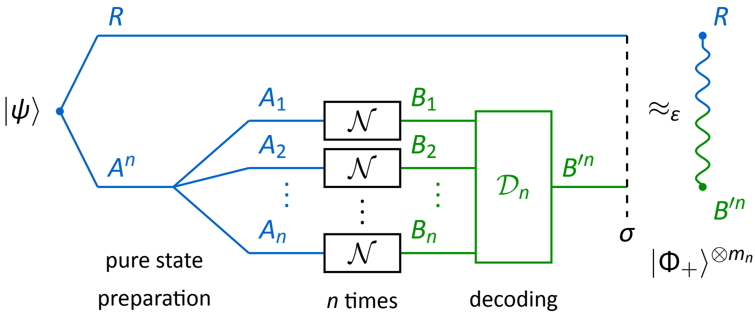
Quantum capacity

- Assume Alice and Bob are in separated labs but can communicate via a **noisy quantum channel** $\mathcal{N} : A \to B$.

- **Quantum information transmission** through $\mathcal{N}$ can be described as **entanglement generation:**
  Use the noisy channel and local operations (encoding, decoding) to generate entanglement between the parties.

- We can allow for **one-way classical communication** without changing the task. [Barnum et al. 2000]

Quantum capacity



- **Goal:** Generate m_n ebits $|\Phi_+\rangle \sim |00\rangle + |11\rangle$ through n **i.i.d. uses** of the quantum channel $\mathcal{N}$.
- **Alice** prepares $|\psi\rangle_{RA^n}$ and sends A^n to **Bob** through $\mathcal{N}^{\otimes n}$.
- **Quantum capacity** $Q(\mathcal{N}) := \sup \left\{ \lim \frac{m_n}{n} \text{ s.t. } \varepsilon \xrightarrow{n\to\infty} 0 \right\}$.

Quantum capacity

- **Coding theorem**: [Lloyd 1997; Shor 2002; Devetak 2005]

$$Q(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} Q^{(1)}(\mathcal{N}^{\otimes n}) = \sup_{n\in\mathbb{N}} \frac{1}{n} Q^{(1)}(\mathcal{N}^{\otimes n}) \qquad (*)$$

  with the **channel coherent information**

$$Q^{(1)}(\mathcal{N}) := \max_{|\psi\rangle_{A'A}} Q^{(1)}(\psi, \mathcal{N})$$

  where $Q^{(1)}(\psi, \mathcal{N}) = I(A'\rangle B)_{(\mathrm{id}\otimes\mathcal{N})(\psi)} = S(B) - S(A'B)$.

- **Regularized formula** $(*)$ in general **intractable to compute**.

- Notorious example: Qubit depolarizing channel, $p \in [0, 1]$

$$\mathcal{D}_p(\rho) := (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z).$$

- Known: $Q(\mathcal{D}_0) = 1$ and $Q(\mathcal{D}_p) = 0$ for $p \geq 0.25$ (no-cloning).

Qubit depolarizing channel

- **Unknown:** $Q(\mathcal{D}_p)$ for $p \in (0, 1/4)$.
- Partial answer for *low noise* ($p \gtrsim 0$):

$$\mathcal{D}_p \approx \mathrm{id} \quad \Longrightarrow \quad Q(\mathcal{D}_p) \approx Q^{(1)}(\mathcal{D}_p) \quad \text{up to } O(p^2 \log p)$$

[FL, Leung, Smith 2017] based on [Sutter et al. 2017]

- **Superadditivity:** $Q^{(1)}(\mathcal{D}_p) = 0$ for $p \geq 0.1894$, but $Q^{(1)}(\mathcal{D}_p^{\otimes 3}) > 0$ for $p \lesssim 0.1901$. [DiVincenzo et al. 1998]
- Achieved by **repetition code** $\sim |0\rangle^{\otimes n} + |1\rangle^{\otimes n}$ (degenerate code).
- Maximum threshold for single rep code at $n = 5$, for $n \geq 10$ by **concatenated codes**. [Smith and Smolin 2007; Fern and Whaley 2008]
- Result: there are $\mathcal{N}$ and $n \in \mathbb{N}$ s.t. $Q^{(1)}(\mathcal{N}^{\otimes n}) > nQ^{(1)}(\mathcal{N})$.

A word on terminology

- Quantum capacity theorem: optimize coherent information $Q^{(1)}(\psi, \mathcal{N}) = I(A'\rangle B)_{(\mathrm{id} \otimes \mathcal{N})(\psi)}$ over pure input state $|\psi\rangle_{A'A}$.
- We refer to $|\psi\rangle$ as a **quantum code** for $\mathcal{N}$.
- Relation to the actual coding strategy: [Hayden et al. 2008]
  - Obtain k copies of $\sigma_{A'B} = (\mathrm{id} \otimes \mathcal{N})(\psi_{A'A})$ by sharing k copies of the **inner code** $|\psi\rangle_{A'A}$ through $\mathcal{N}^{\otimes k}$.
  - The **outer code** used by Alice and Bob to generate entanglement through $\mathcal{N}$ is based on a **Haar-random subspace** R of a suitable typical subspace of $\sigma_{A'B}^{\otimes k}$.
  - The entanglement generation protocol succeeds (up to error ε) if $\frac{1}{k} \log |R|$ is not greater than $I(A'\rangle B)_\sigma$.
- Alternatively: one-way entanglement distillation on $\sigma_{A'B}^{\otimes k}$

[Devetak and Winter 2005]

Entanglement in quantum information transmission

- Quantum capacity theorem: If $r := Q^{(1)}(\psi, \mathcal{N}) > 0$ for some code ψ, then $\exists$ entanglement generation protocol with rate r.
- Using **block codes**, we may try to find quantum codes $|\psi_n\rangle_{A'^n A^n}$ such that $Q^{(1)}(\psi_n, \mathcal{N}^{\otimes n}) > 0$.
- Rate of the protocol: $\frac{1}{n} Q^{(1)}(\psi_n, \mathcal{N}^{\otimes n}) > 0$.
- This can give better rates (cf. repetition code for depolarizing channel) due to **superadditivity of coherent information**.
- Reason for superadditivity is the **multipartite entanglement** present in the quantum code ψ_n.

Entanglement in quantum information transmission

- **Goal:** Find good quantum codes $\psi_{A'^n A^n}$ with high coherent-information rate $\frac{1}{n}Q^{(1)}(\psi_n, \mathcal{N}^{\otimes n})$

  $\longrightarrow$ lower bound on quantum capacity $Q(\mathcal{N})$.

- **Challenge:** Hard to parametrize multipartite entanglement in many-body quantum state

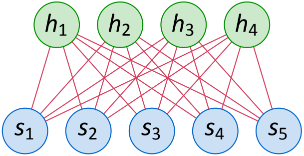
$$(\mathbb{C}^2)^{\otimes k} \ni |\psi_k\rangle = \sum_{s^k \in \{0,1\}^k} \psi(s^k) |s_1\rangle \otimes \ldots \otimes |s_k\rangle.$$

- # degrees of freedom (amplitudes $\psi(s^k)$) is exponential in k.
- **Idea from many-body physics:** Use ansatz for $|\psi_k\rangle$ with poly(k) parameters that retains interesting features.
- Well-known examples: tensor networks (e.g. MPS [Fannes et al. 1992], PEPS [Verstraete and Cirac 2004])

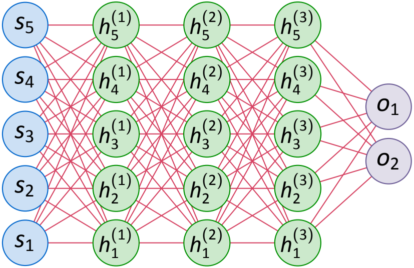
Neural network states

$$(\mathbb{C}^2)^{\otimes k} \ni |\psi_k\rangle = \sum_{s^k \in \{0,1\}^k} \psi(s^k) |s_1\rangle \otimes \ldots \otimes |s_k\rangle.$$

Recent idea: use a neural network to compute $\psi(s^k)$ given the input string s^k. [Carleo and Troyer 2017]

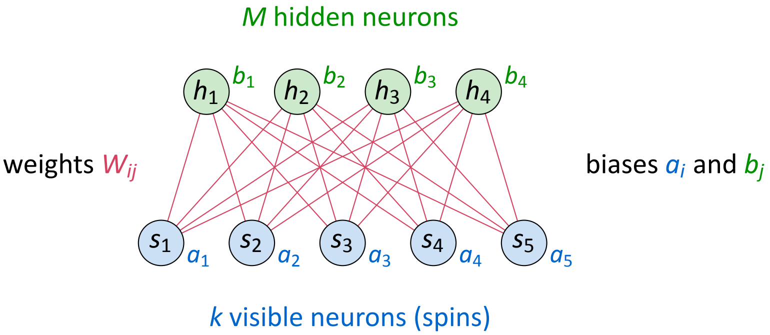


Restricted Boltzmann machine (RBM)



Feedforward net (FF)

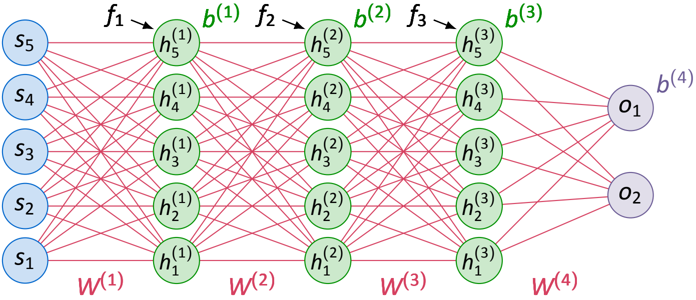
Restricted Boltzmann machines



- $a^k \in \mathbb{C}^k, b^M \in \mathbb{C}^M$: biases for visible neurons s^k and hidden neurons h^k.
- Weight $W_{ij} \in \mathbb{C}$: interaction between s_j and h_i.
- Amplitude $\psi(s^k)$ computed as (up to normalization)

$$\psi(s^k) = \sum_{h^k} \exp\left(\sum_j a_j s_j + \sum_i b_i h_i + \sum_{i,j} W_{ij} h_i s_j\right)$$

Feedforward nets



- Hidden layers $h^{(l)}$, output layer o^2: biases $b^{(l)}$ and weights $W_{ij}^{(l)}$.
- Values of the hidden neurons are set by $(h^{(0)} \equiv s)$

$$h^{(l)} = f_l\left(W^{(l)}h^{(l-1)} + b^{(l)}\right)$$

  where f_l is an **activation function** (e.g., sigmoid or ReLU).
- Amplitude function: $\psi(s^k) = o_1 + io_2$ (or $\psi(s^k) = e^{o_1+io_2}$).

Neural network states

- NN state ansatz has proven successful to describe ground states of interesting Hamiltonians, for example:
  - transverse-field Ising model [Carleo and Troyer 2017]
  - antiferromagnetic Heisenberg model [Carleo and Troyer 2017]
  - free fermions on a lattice [Cai and Liu 2018]
- NN states are known to be capable of efficiently representing interesting classes of states, for example:
  - Graph states [Gao and Duan 2017]
  - Toric code [Gao and Duan 2017]
  - General stabilizer states [Zhang et al. 2018]
  - Surface codes [Jia et al. 2018]
- Versatile ansatz for multipartite entanglement

$\longrightarrow$ use it to find quantum codes!

Optimization procedure

- **Goal:** Maximize coherent information $Q^{(1)}(\psi_n, \mathcal{N}^{\otimes n})$ w.r.t. network parameters $\{b_\vartheta, W_\vartheta\}$ that define ψ_n:

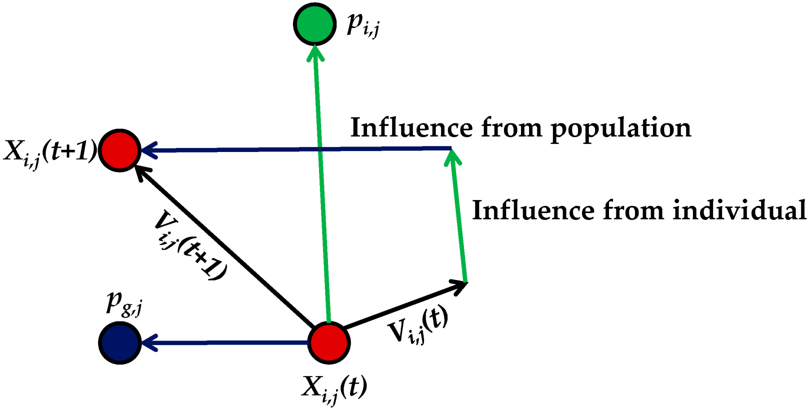
$$\{b_\vartheta, W_\vartheta\} \longmapsto |\psi_n\rangle \longmapsto (\mathrm{id}_n \otimes \mathcal{N}^{\otimes n})(\psi_n) \longmapsto I(A'^n \rangle B^n)$$

- **Tunable parameters of FF**:
  - Number L of hidden layers $h^{(i)}$ with width H_i and activation function f_i.
  - # real parameters: $H_1(2n+1) + \sum_{i=2}^{L-1} H_i(H_{i-1}+1) + 2(H_L+1)$.
- Typical FF-choices for us:
  - $L = 3$, $H_i = 2n$.
  - $f_1(x) = \cos(x)$ and ReLUs $f_i(x) = \max\{0, x\}$ for $i \geq 2$.
- **Tunable parameters for RBM**:
  - Width M of hidden layer (tune this to match FF-model).
  - # real parameters: $2(M + 2n + 2nM)$.

Optimization procedure

- We had to compute the coherent information exactly instead of using the sampling method for the energy in C/T.
- Machine learning typically uses gradient-based updates (such as ADAM or Adagrad).
- However: Coherent information of a high-noise channel has lots of local maxima given by product states.
- Gradient is likely to get stuck $\longrightarrow$ **gradient-free optimization**
- Example: particle swarm optimization (PSO)
  - Send out N particles, each probing the landscape.
  - Each particle records personal best function value, and all know the global swarm best.
  - In each iteration, particle velocity is updated with weights towards personal best, global best, and inertial movement.

Particle swarm optimization

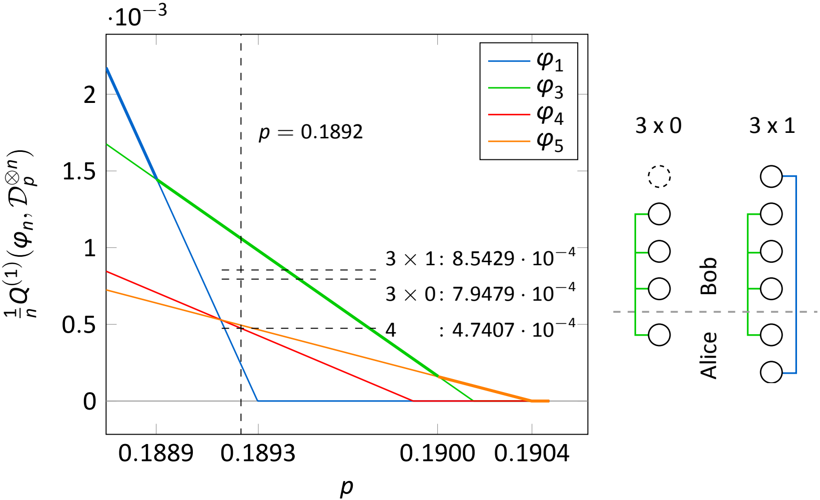


Source: Wang et al., Appl. Sci. 2017, 7(8), 754

Optimal codes for depolarizing channel

Known optimal codes for $\mathcal{D}_p(\rho) = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$

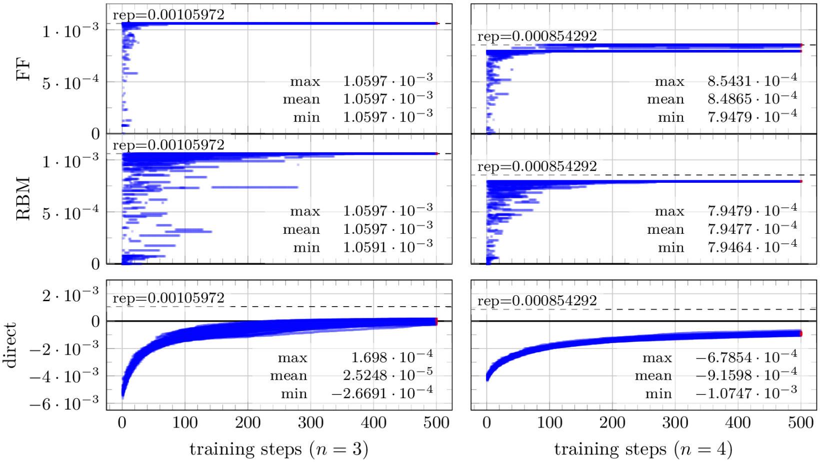
are repetition codes $|\varphi_n\rangle \sim |0\rangle_{A'}|0\rangle_A^{\otimes n} + |1\rangle_{A'}|1\rangle_A^{\otimes n}$ (up to $n \leq 9$)



Main result: depolarizing channel

- Numerical limitation: $n \leq 6$ channels (up to 12-qubit NN states).
- For this talk: $\mathcal{D}_p^{\otimes n}$ with $p = 0.1892$ and $n = 3, 4$.
- FF configuration:
  - 3 hidden layers of width $2n$: Cos $\rightarrow$ ReLU $\rightarrow$ ReLU
  - Cos as activation function is non-standard, but has been used to deal with the Hamiltonian sign problem. [Cai and Liu 2018]
  - 140/234 real parameters.
- RBM configuration:
  - $M = 9$ (for $n = 3, 4$).
  - 138/232 real parameters.
- Direct parametrization (for comparison): 128/512 parameters.
- 80 parallel threads with 100 particles each, 500 PSO iterations.

Main result: depolarizing channel



$$\tfrac{1}{3}Q^{(1)}(\varphi_3, \mathcal{D}_p^{\otimes 3}) = 1.0597 \cdot 10^{-3}$$

$$\tfrac{1}{4}Q^{(1)}(\varphi_3 \otimes \chi_0, \mathcal{D}_p^{\otimes 4}) = 7.9479 \cdot 10^{-4}$$

$$\tfrac{1}{4}Q^{(1)}(\varphi_1 \otimes \varphi_3, \mathcal{D}_p^{\otimes 4}) = 8.5429 \cdot 10^{-4}$$

More results (in the paper)

- We also give analytical constructions of product codes using RBM and FF.
- NN ansatz is not limited or tailored to Pauli channels.
- Test on **dephrasure channel** [FL, Leung, Smith 2018]

$$\mathcal{N}_{p,q}(\rho) = (1-q)((1-p)\rho + Z\rho Z) + q\,\mathrm{Tr}(\rho)|e\rangle\langle e|.$$

- Exhibits superadditivity of coherent information for $n \geq 2$.
- For $n = 3, 4$ the NN ansatz finds codes that **outperform** all codes found in [FL, Leung, Smith 2018] using a direct parametrization, including weighted repetition codes.
- Finally, NN ansatz can also be used to find **absolutely maximally entangled states** (maximally mixed after tracing out half the systems) by maximizing an average linear entropy quantity.

Summary

- NN state ansatz is an **efficient way of finding optimal codes** for quantum information transmission using reinforcement learning.
- For $n \leq 6$ uses of the depolarizing channel, ansatz has **remarkable convergence**.
- Also works well for other channels such as **dephrasure channel**.
- Ansatz can also be used find **AME states**, which are certain types of error-correction codes.
- **Take-away message:** NN ansatz is capable of representing relevant multipartite entanglement for quantum information transmission and quantum error correction.

Limitations/open problems

- Severe numerical limitations due to computation of entropies (involves diagonalization).
- Couldn't tap into polynomial scaling advantage of NN states (for us: parametrization of the entanglement).
- More efficient ways of computing the coherent information, e.g. mimicking sampling trick as in Carleo/Troyer?
- Other bottleneck: compute channel action
  $\longrightarrow$ model this by a neural network as well?
- More generally, for what other information-processing tasks can we use the NN ansatz?

References

Barnum, H. et al. (2000). *IEEE Transactions on Information Theory* 46.4, pp. 1317–1329. arXiv: quant-ph/9809010.

Bausch, J. and F. Leditzky (2018). *arXiv preprint*. arXiv: 1806.08781 [quant-ph].

Cai, Z. and J. Liu (2018). *Physical Review B* 97.3, p. 035116. arXiv: 1704.05148 [cond-mat.str-el].

Carleo, G. and M. Troyer (2017). *Science* 355.6325, pp. 602–606. arXiv: 1606.02318 [cond-mat.dis-nn].

Devetak, I. (2005). *IEEE Transactions on Information Theory* 51.1, pp. 44–55. arXiv: quant-ph/0304127.

Devetak, I. and A. Winter (2005). *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* 461.2053, pp. 207–235. arXiv: quant-ph/0306078.

DiVincenzo, D. P. et al. (1998). *Physical Review A* 57.2, p. 830. arXiv: quant-ph/9706061.

Fannes, M. et al. (1992). *Communications in Mathematical Physics* 144.3, pp. 443–490.

Fern, J. and K. B. Whaley (2008). *Physical Review A* 78.6, p. 062335. arXiv: 0708.1597 [quant-ph].

Gao, X. and L.-M. Duan (2017). *Nature Communications* 8.1, p. 662. arXiv: 1701.05039 [cond-mat.dis-nn].

Hayden, P. et al. (2008). *Open Systems & Information Dynamics* 15.01, pp. 7–19. arXiv: quant-ph/0702005.

Jia, Z.-A. et al. (2018). arXiv: 1802.03738.

Leditzky, F. et al. (2018a). *Physical Review Letters* 120.16, p. 160503. arXiv: 1705.04335 [quant-ph].

— (2018b). *Physical Review Letters* 121.16, p. 160501. arXiv: 1806.08327 [quant-ph].

Lloyd, S. (1997). *Physical Review A* 55.3, p. 1613. arXiv: quant-ph/9604015.

Shor, P. W. (2002). *MSRI Workshop on Quantum Computation*. Berkeley, CA, USA. Available at: http://www.msri.org/realvideo/ln/msri/2002/quantumcrypto/shor/1/.

Smith, G. and J. A. Smolin (2007). *Physical Review Letters* 98.3, p. 030501. arXiv: quant-ph/0604107.

Sutter, D. et al. (2017). *IEEE Transactions on Information Theory* 63.12, pp. 7832–7844. arXiv: 1412.0980 [quant-ph].

Verstraete, F. and J. I. Cirac (2004). *arXiv preprint*. arXiv: cond-mat/0407066.

Zhang, Y.-H. et al. (2018). arXiv: 1809.08631 [quant-ph].

Thank you very much for your attention!