Quantum codes from neural networks

Felix Leditzky

(JILA & CTQM, University of Colorado Boulder)

Joint work with Johannes Bausch (Univ. of Cambridge)

arXiv:1806.08781

KITP program `machine19`

Santa Barbara, 5 March 2019

JILA CU Boulder and NIST

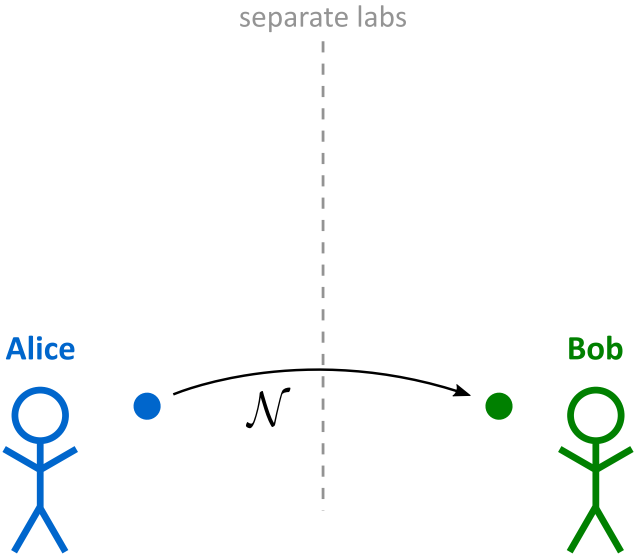
Center for Theory of Quantum Matter CTQM

Table of Contents

Quantum channels

- Most general description of **noise in a quantum system**:
  A quantum channel $\mathcal{N}$ is a linear, completely positive, trace-preserving map.
- Physical interpretation: unitary evolution on system + environment.
  Stinespring Theorem: $\mathcal{N}(\rho_A) = \mathrm{Tr}_E\left(U_{AE}(\rho_A \otimes |0\rangle\langle 0|_E)U_{AE}^\dagger\right)$ for some unitary U_{AE}.
- Interpret $\mathcal{N}: A \to B$ as **noisy communication link** between Alice and Bob.
- Protecting quantum system from noise can be put in information-theoretic terms:
  How much information (quantum, classical, ...) can Alice send to Bob through $\mathcal{N}$?
- Equivalently, how much entanglement can Alice and Bob establish between themselves? $\longrightarrow$ quantum capacity.

Quantum information transmission



Quantum channel $\mathcal{N} : A \to B$

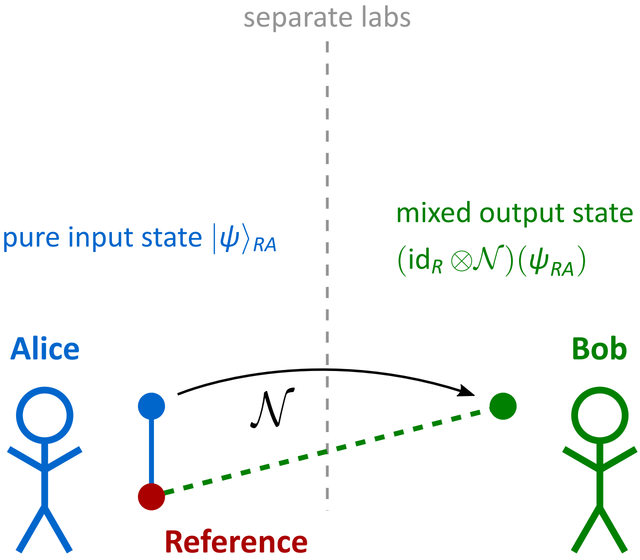
Goal:

Transmit quantum information from Alice to Bob.

Strategy:

Share (mixed) entangled state via $\mathcal{N}$ and distill EPR pairs using local operations and **forward** classical communication $A \to B$.

Quantum information transmission



Rate of the distillation protocol:

coherent information

$$\mathcal{I}_c(\psi_{RA}, \mathcal{N}) := S(\mathcal{N}(\psi_A)) - S(\mathrm{id}_R \otimes \mathcal{N}(\psi_{RA})).$$

Optimizing over **quantum codes** ψ:

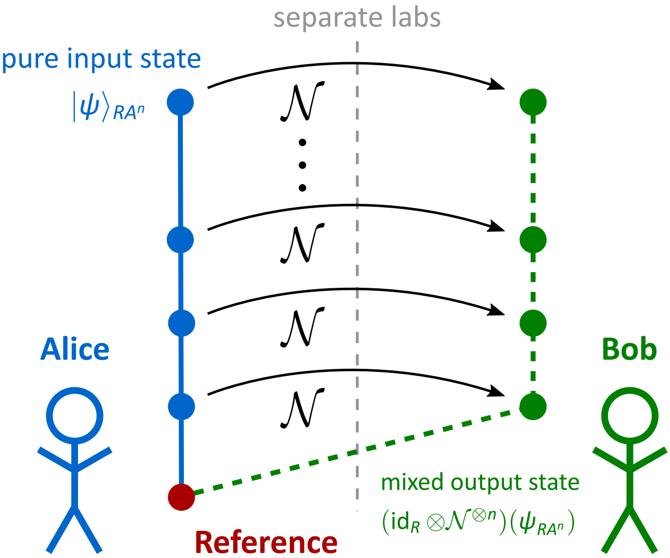
Channel coherent information

$$\mathcal{I}_c(\mathcal{N}) := \sup_\psi \mathcal{I}_c(\psi_{RA}, \mathcal{N}).$$

Can we achieve more?

[Devetak 2005; Devetak, Winter 2005]

Quantum information transmission



Idea: Use n channels in parallel to share multipartite state $|\psi\rangle_{RA^n}$.

Distillation rate: $\frac{1}{n}\mathcal{I}_c(\psi_{RA^n}, \mathcal{N}^{\otimes n})$

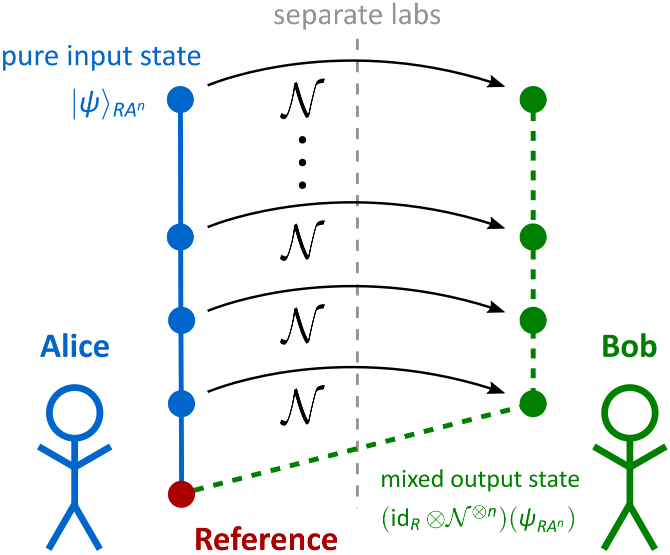
For certain $\mathcal{N}$ and ψ,

$$\frac{1}{n}\mathcal{I}_c(\psi_{RA^n}, \mathcal{N}^{\otimes n}) > \mathcal{I}_c(\mathcal{N}).$$

This is called **superadditivity** of coherent information.

[DiVincenzo et al. 1998]

Quantum information transmission



Quantum capacity:

$$Q(\mathcal{N}) = \sup_{n\in\mathbb{N}} \frac{1}{n}\mathcal{I}_c(\mathcal{N}^{\otimes n})$$

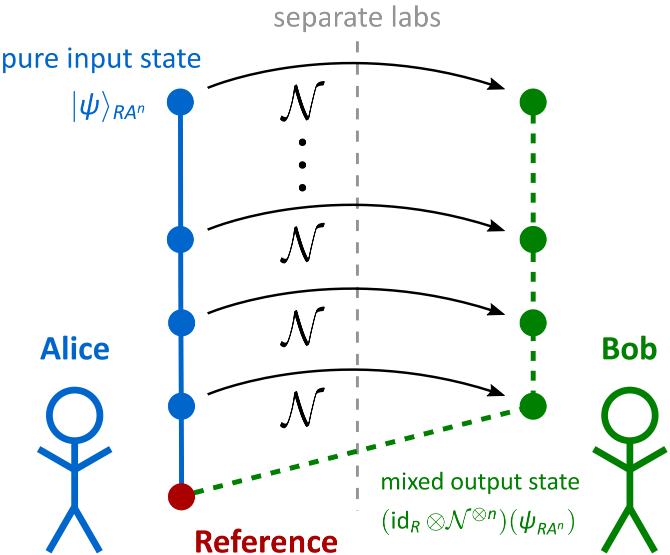
Good: Superadditivity can boost achievable rates, $Q(\mathcal{N}) > \mathcal{I}_c(\mathcal{N})$.

Bad: In general, quantum capacity **intractable to compute**.

Challenge: Find good codes achieving superadditivity.

[Lloyd 1997; Shor 2002; Devetak 2005]

Quantum information transmission



Quantum capacity:

$$Q(\mathcal{N}) = \sup_{n \in \mathbb{N}} \frac{1}{n} \mathcal{I}_c(\mathcal{N}^{\otimes n})$$

Practical question:

Highest possible rate?

Maximize $\frac{1}{n}\mathcal{I}_c(\mathcal{N}^{\otimes n})$.

Fundamental question:

Highest threshold?*

Assert $Q(\mathcal{N}) > 0$.

* for $\mathbb{R} \ni r \longmapsto \mathcal{N}_r$.

Qubit depolarizing channel

$$\mathcal{D}_p(\rho) := (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z) \quad \text{for } p \in [0,1].$$

[DiVincenzo et al. 1998]
[Smith, Smolin 2007]
[Fern, Whaley 2008]
[Sutter et al. 2017]
[FL, Leung, Smith 2018]

Superadditivity known for $0.1889 \leq p \leq 0.1912$.

Achieved by **repetition codes** $|0\rangle_R \otimes |0\rangle_A^{\otimes n} + |1\rangle_R \otimes |1\rangle_A^{\otimes n}$ and (Shor-like) concatenated codes.

Largest magnitude of superadditivity $\sim 10^{-3}$.

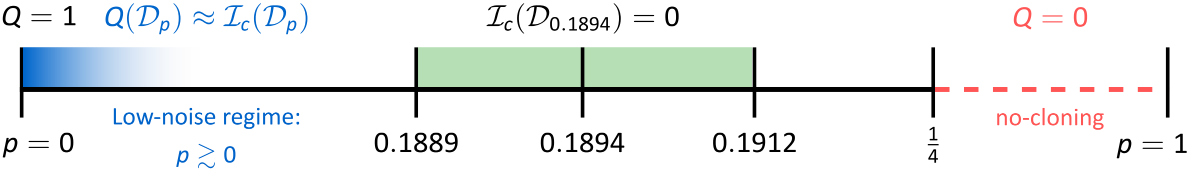


Table of Contents

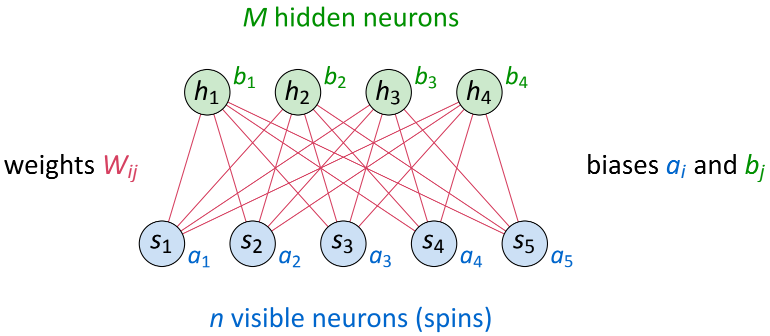
Entanglement in quantum information transmission

- **Goal:** Find good quantum codes ψ_{RA^k} with high rate $\frac{1}{k}\mathcal{I}_c(\psi, \mathcal{N}^{\otimes k})$ to obtain lower bound on quantum capacity $Q(\mathcal{N})$.
- **Challenge:** Hard to parametrize multipartite entanglement in many-body quantum state with exponentially many degrees of freedom ($n = 2k$):

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

- **Idea from many-body physics:** Use ansatz for $|\psi_n\rangle$ with poly(n) parameters that retains interesting features.
- **Options:**
  - tensor networks [Fannes et al. 1992; Verstraete and Cirac 2004]
  - **neural network states** [Carleo and Troyer 2017]

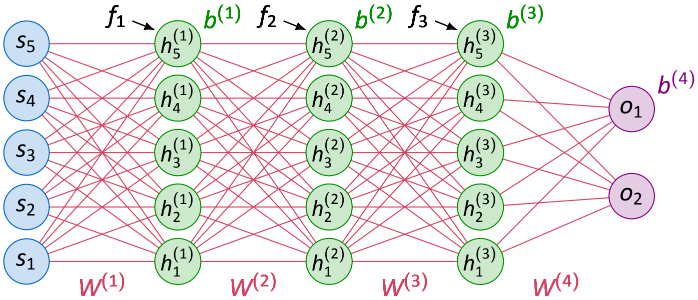
Restricted Boltzmann machines



- $a^n \in \mathbb{C}^n$, $b^M \in \mathbb{C}^M$: biases for visible neurons s^n and hidden neurons h^M.
- Weight $W_{ij} \in \mathbb{C}$: interaction between s_j and h_i.
- Up to normalization, amplitude $\psi(s^n)$ given by

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

Feedforward nets



- Hidden layers $h^{(m)}$, output layer o^2: biases $b^{(m)}$ and weights $W_{ij}^{(m)}$.
- Values of the hidden neurons are set by $h^{(m)} = f_m\left(W^{(m)}h^{(m-1)} + b^{(m)}\right)$ with activation function f_m (e.g., sigmoid or ReLU).
- Up to normalization, $\psi(s^k) = o_1 + io_2$ (or $\psi(s^k) = \exp(o_1 + io_2)$).

Neural network states

- NN states are known to be capable of efficiently representing:
  - Graph states, toric code [Gao and Duan 2017]
  - Surface codes [Jia et al. 2018]
  - **General stabilizer states** [Zhang et al. 2018]

- Versatile ansatz for multipartite entanglement

  $\longrightarrow$ use it to **find good quantum codes**!

- Apply this to find quantum codes for:
  - Pauli channels such as **depolarizing channel**;
  - non-Pauli error models such as **dephrasure channel**.

Optimization procedure

- **Goal:** Maximize coherent information $\mathcal{I}_c(\psi, \mathcal{N}^{\otimes k})$ w.r.t. network parameters $\{b_\vartheta, W_\vartheta\}$ that define ψ_{RA^k}:
  1. Compute $|\psi\rangle_{RA^k}$ for given weights $\{b_\vartheta, W_\vartheta\}$.
  2. Compute channel action $\sigma_{RB^k} := (\mathrm{id}_R \otimes \mathcal{N}^{\otimes k})(\psi)$.
  3. For the mixed state σ_{RB^k} compute $\mathcal{I}_c(\psi, \mathcal{N}^{\otimes k}) = S(B^k)_\sigma - S(RB^k)_\sigma$.
  4. Update $\{b_\vartheta, W_\vartheta\}$.
- Typical parameter choices for FF-nets:
  - Three hidden layers of width $n = 2k$ each.
  - Activation functions: $f_1(x) = \cos(x)$ and ReLUs $f_i(x) = \max\{0, x\}$ for $i \geq 2$.
- Typical parameter choices for RBMs:
  - Hidden-layer width $M \sim n = 2k$, tuned to match # degrees of freedom of FF-net.

Optimization procedure

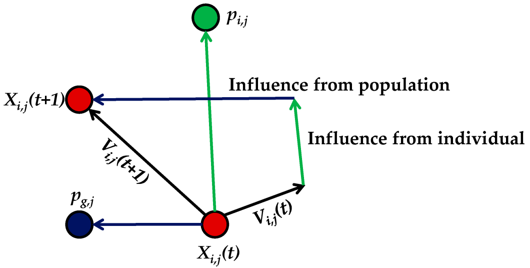
- **Problem:** Coherent information of a high-noise channel has **lots of local maxima** given by product states.

$$\mathcal{I}_c(|\chi\rangle_R \otimes |\varphi\rangle_A, \mathcal{N}) = S(B)_{\mathcal{N}(\varphi)} - S(RB)_{\chi\otimes\mathcal{N}(\varphi)}$$
$$= S(B)_{\mathcal{N}(\varphi)} - S(R)_\chi - S(B)_{\mathcal{N}(\varphi)} = 0.$$

- These maxima are not interesting for us, and gradient is likely to get stuck in them.
- Use **gradient-free optimization** instead.
- Good choices: **particle swarm optimization (PSO)**, artificial bee colonization (ABC), pattern/direct search (DS)

Particle swarm optimization

- Send out N particles, each probing the landscape.
- Each particle records personal best function value.
- All particles know global best.
- Weighted velocity update:
  - Towards personal best;
  - towards global best;
  - current direction ($\triangleq$ inertia).



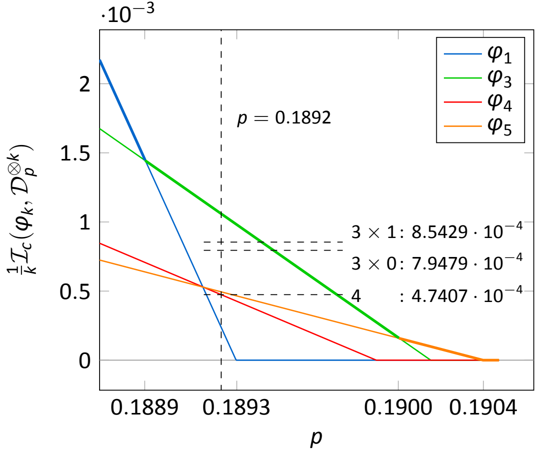
Source: Wang et al., Appl. Sci. 2017, 7(8), 754

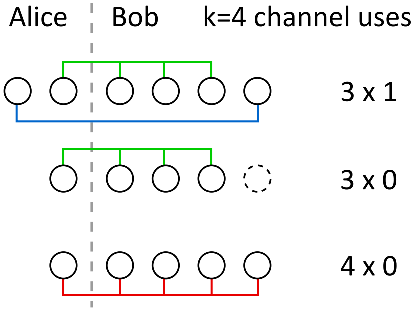
Table of Contents

Optimal codes for depolarizing channel

Known optimal codes for $\mathcal{D}_p(\rho) = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$ (up to $k \leq 9$):

repetition codes $|\varphi_k\rangle \sim |0\rangle_R|0\rangle_A^{\otimes k} + |1\rangle_R|1\rangle_A^{\otimes k}$

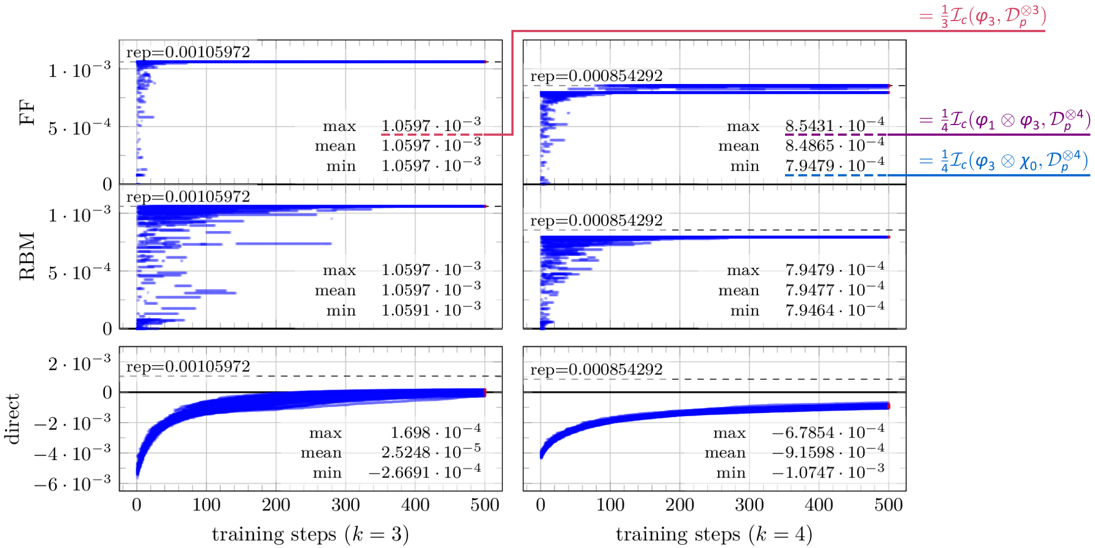




Main result: depolarizing channel

- Numerical limitation: $k \leq 6$ channels (up to 12-qubit NN states).
- For this talk: $\mathcal{D}_p^{\otimes k}$ with $p = 0.1892$ and $k = 3, 4$.
- FF configuration:
  - 3 hidden layers of width $2k$: Cos $\to$ ReLU $\to$ ReLU
  - 140/234 real parameters.
- RBM configuration:
  - $M = 9$ (for $k = 3, 4$).
  - 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



Periodic activation function in first layer

- **Observation:** Much better convergence with cosine activation function in first layer.
- Many-body physics: Cos in first layer seems to help with sign problem. [Cai and Liu 2018]
- **Quantum information folklore:** good quantum codes are **degenerate**.

  ($\Leftrightarrow$ different (Pauli) errors have same error syndrome)
- Good example: repetition code

$$Z_1 Z_2 Z_3 \; Z_4 \qquad\qquad Z_1 Z_2 Z_3 \; Z_4$$

$$|\varphi_n\rangle \sim |0000\ldots0\rangle + |1111\ldots1\rangle$$

$$Z_1 Z_2 Z_4 \qquad\qquad Z_1 Z_2 Z_4$$

$$Z_1 Z_2 Z_3 |\varphi_n\rangle = |\widetilde{\varphi}_n\rangle \sim |0\rangle^{\otimes k} - |1\rangle^{\otimes k}$$

$$Z_1 Z_2 Z_4 |\varphi_n\rangle = |\widetilde{\varphi}_n\rangle$$

$$Z_4 |\varphi_n\rangle = |\widetilde{\varphi}_n\rangle$$

- Cos followed by ReLU can easily **detect parity** in input string

  $\longrightarrow$ bias towards degenerate codes.

Finding quantum codes for non-Pauli channels

- NN ansatz for quantum codes yields best known codes for depolarizing channel.
- Depolarizing channel belongs to class of **Pauli channels**.
- Is there a channel for which NN ansatz finds better codes than the known ones?
- **Dephrasure channel:** For $p, q \in [0,1]$, [FL, Leung, Smith 2018]

$$\mathcal{N}_{p,q}(\rho) := (1-q)\left[(1-p)\rho + pZ\rho Z\right] \oplus q\,\mathrm{Tr}(\rho)|e\rangle\langle e|.$$

- **deph**asing + **erasure**: first dephase with probability p, then erase with probability q.

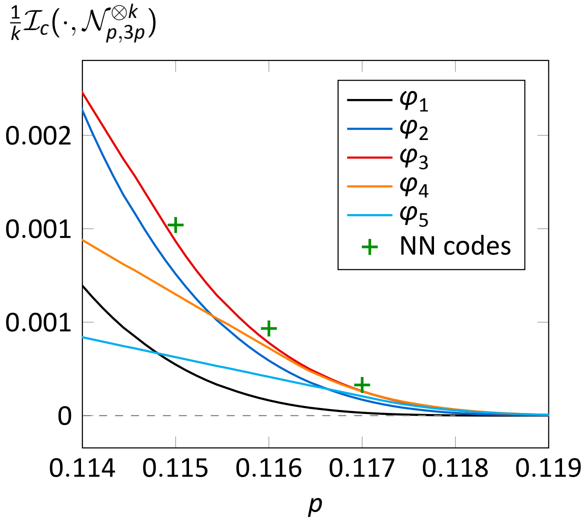
Dephrasure channel

- Dephrasure channel exhibits **substantial superadditivity effects**.
- Known optimal codes are weighted repetition codes (similar to depolarizing channel).
- For certain values of (p, q) and $k = 3, 4$ uses:

  NN ansatz finds quantum codes that **outperform all known ones**!
- Example: $k = 4, (p, q) = (0.116, 0.348)$.
- NN setup: FF net with Cos $\to$ ReLU $\to$ ReLU $\to$ ReLU.

$$\begin{aligned}|v\rangle_{RA^4} = b_1(&|0000\rangle_R|0101\rangle_{A^4} + |1111\rangle_R|1010\rangle_{A^4}) \\ &+ b_2(|0100\rangle_R|0000\rangle_{A^4} + |1011\rangle_R|1111\rangle_{A^4}).\end{aligned}$$

Dephrasure channel



Channel: $\mathcal{N}_{p,3p}$

p . . . dephasing probability

$3p$ . . . erasure probability

Weighted repetition code:

$$|\varphi_n\rangle = \sqrt{\lambda}|0\rangle_R|0\rangle_A^{\otimes n} + \sqrt{1-\lambda}|1\rangle_R|1\rangle_A^{\otimes n}$$

Neural network codes:

FF net with Cos-ReLU-ReLU-ReLU

Table of Contents

Numerical bottlenecks

- Target function $\mathcal{I}_c(\psi, \mathcal{N}^{\otimes k})$ is an **entropic quantity**.

  $\longrightarrow$ Monte Carlo sampling method of [Carleo and Troyer 2017] not applicable.

- Worse: Need to diagonalize large matrix ($2^{2k} \times 2^{2k}$) to compute entropies.

- Infeasible in optimization methods for $k \gtrsim 7$.

- Keeps us from tapping into the scaling advantage of NN states (poly(k) vs. exp(k)).

- **However:** NN states seem to be a **good ansatz for entanglement in quantum codes**.

Numerical bottlenecks

Possible remedy 1

- Find an easy to compute “indicator function” for positivity of coherent information?
- Natural candidate: Rényi entropies $S_\alpha(\rho) = \frac{1}{1-\alpha} \log \operatorname{Tr} \rho^\alpha$.
- Problem 1: Rényi versions of differences of entropies are problematic.
- Problem 2: In high-noise regime, superadditivity effects have **tiny magnitude**.

Possible remedy 2

- Switch to optimization techniques that minimize the number of function evaluations?
- Find a smarter gradient-based technique?

Numerical bottlenecks

- Another problem: computing the channel action $\mathcal{N}^{\otimes k}$.
- Most favorable implementations:
  - Sequential Kraus operator application:
    $$\sigma_1 = (\mathrm{id} \otimes \ldots \mathrm{id} \otimes \mathcal{N})(\varphi) \longrightarrow \sigma_2 = (\mathrm{id} \otimes \ldots \otimes \mathrm{id} \otimes \mathcal{N} \otimes \mathrm{id})(\sigma_1) \longrightarrow \ldots$$
  - Transfer matrix formalism that translates channel application to matrix multiplication.
- Both approaches involve the handling of large dense matrices.
- Can we model/approximate the channel action using a neural network?
- Related: circuit decompositions of quantum channels. [Iten et al. 2017; Shen et al. 2017]

Table of Contents

Conclusion

- Good quantum codes for quantum information transmission have **non-trivial multipartite entanglement**.
- **Hard to find** both analytically/algebraically and in numerical optimization.
- Neural network states: efficient representation of interesting entangled states.
- In conjunction with global optimization techniques, **NN states yield good superadditive quantum codes**.
- Works for interesting channels such as **depolarizing channel** and **dephrasure channel**.

Open problems

- Overcome the numerical limitations in our applications to go to higher dimensions:
  - diagonalizing large matrices;
  - computing entropies;
  - compute channel action.
- Extend ansatz to use NN density operators? [Torlai and Melko 2018]
- Identify other applications of NN states in quantum information-theoretic contexts.
- Use more sophisticated ML techniques (autoencoders, adversarial networks).

References

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). *Proc. Royal Soc. A: Math., Phys. and Eng. Sci.* 461.2053, pp. 207–235. arXiv: quant-ph/0306078.

DiVincenzo, D. P., P. W. Shor, and J. A. Smolin (1998). *Physical Review A* 57.2, p. 830. arXiv: quant-ph/9706061.

Fannes, M., B. Nachtergaele, and R. F. Werner (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].

Iten, R., R. Colbeck, and M. Christandl (2017). *Phys. Rev. A* 95 (5), p. 052316. arXiv: 1609.08103 [quant-ph].

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

Leditzky, F., D. Leung, and G. Smith (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.

Shen, C. et al. (2017). *Phys. Rev. B* 95 (13), p. 134501. arXiv: 1611.03463 [quant-ph].

Shor, P. W. (2002). *MSRI Workshop on Quantum Computation*. Berkeley, CA, USA.

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

Sutter, D., V. B. Scholz, A. Winter, and R. Renner (2017). *IEEE Transactions on Information Theory* 63.12, pp. 7832–7844. arXiv: 1412.0980 [quant-ph].

Torlai, G. and R. G. Melko (2018). *Physical Review Letters* 120 (24), p. 240503. arXiv: 1801.09684 [quant-ph].

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

Zhang, Y.-H., Z.-A. Jia, Y.-C. Wu, and G.-C. Guo (2018). arXiv: 1809.08631 [quant-ph].

Thank you very much for your attention!