Quantum codes from neural networks

Felix Leditzky

(JILA & CTQM, University of Colorado Boulder)

Joint work with Johannes Bausch (University of Cambridge)

arXiv:1806.08781

Beyond IID in information theory

Sydney, Australia

4 July 2019

JILA
CU Boulder and NIST

Center for Theory of Quantum Matter
CTQM

Quantum capacity of quantum channels

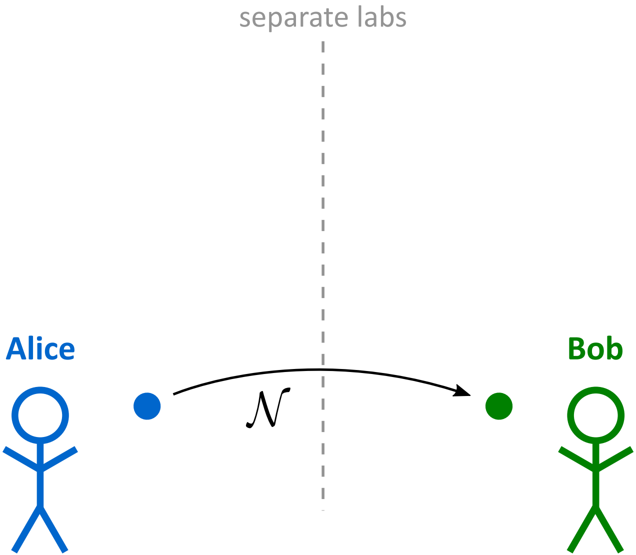
- Most general description of **noise in a quantum system**:
  A quantum channel $\mathcal{N}$ is a linear, completely positive, trace-preserving map.

- Quantum Shannon theory: Interpret $\mathcal{N}: A \to B$ as **noisy communication link** between Alice and Bob.

- Protecting quantum system from noise in information-theoretic terms:
  How much quantum information can Alice send to Bob reliably through $\mathcal{N}$?

- Equivalent formulation:
  How much entanglement can Alice and Bob generate using $\mathcal{N}$ and 1-LOCC?

Quantum information transmission



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

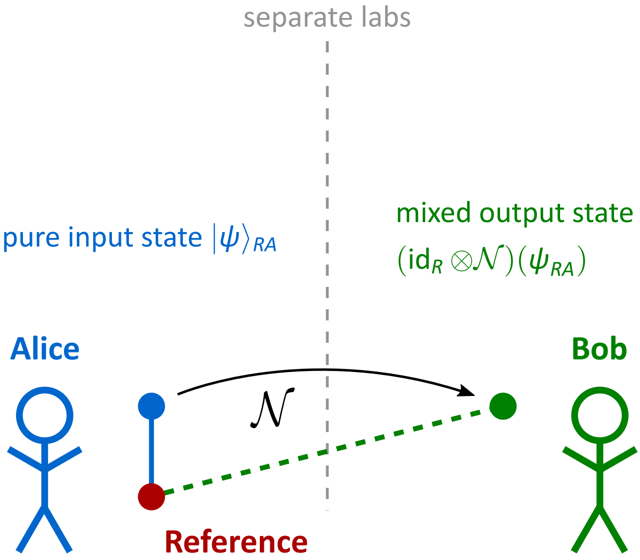
Goal:

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:

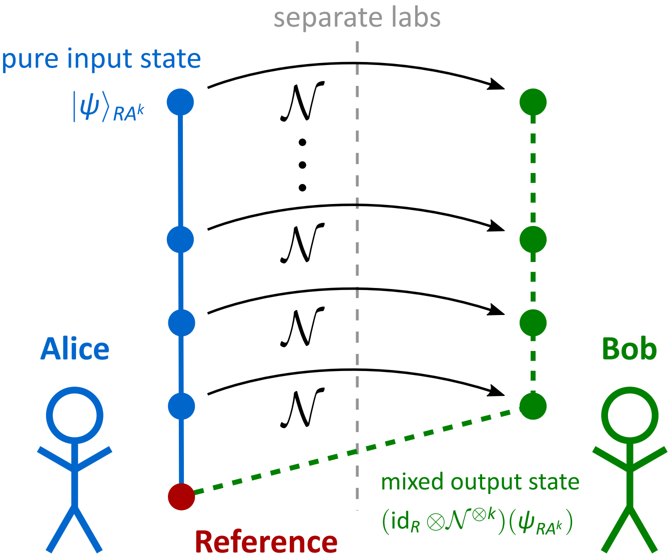
Channel coherent information

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

Can we achieve more?

[Devetak 2005; Devetak, Winter 2005]

Quantum information transmission



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

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

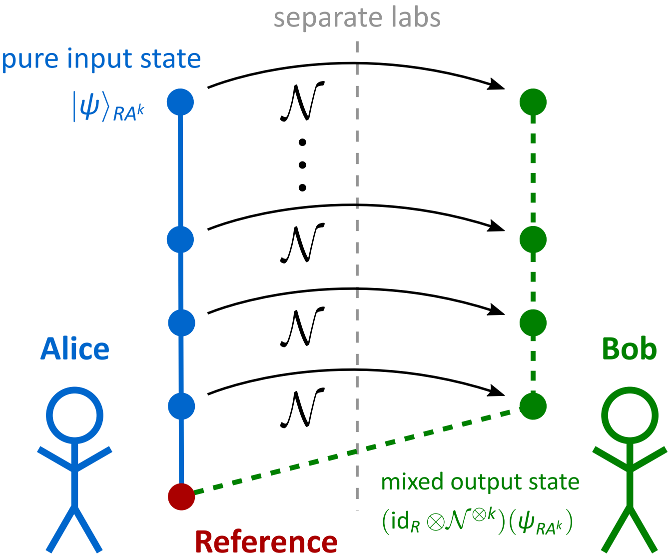
For certain $\mathcal{N}$ and ψ,

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

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

[Shor, Smolin 1996; DiVincenzo et al. 1998]

Quantum information transmission



Quantum capacity:

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

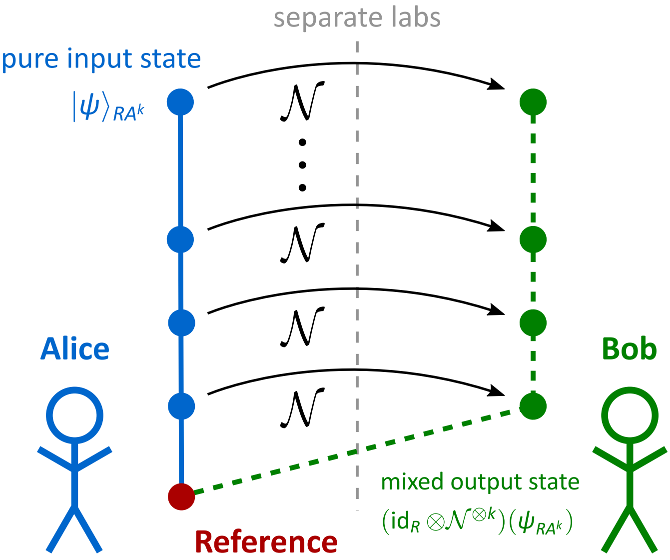
Good: Superadditivity can boost achievable rates, $Q(\mathcal{N}) > \mathcal{I}_c(\mathcal{N})$.

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:

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

Fundamental question:

Highest threshold?*

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

Practical question:

Highest possible rate?

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

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

Talk outline

Table of Contents

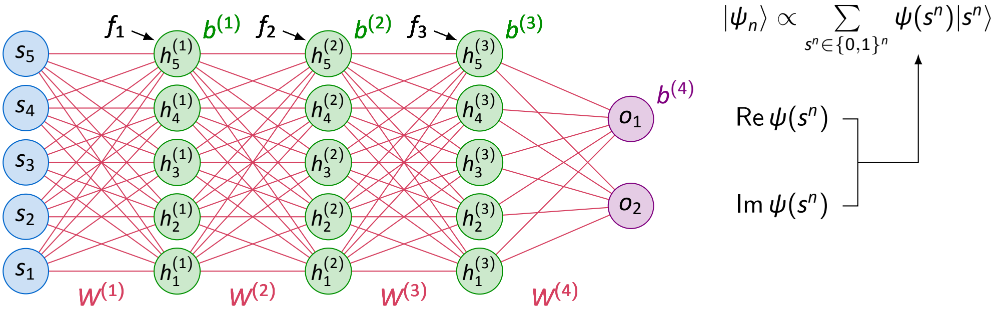
Entanglement in quantum information transmission

- **Goal:** Find good quantum codes ψ_{RA^k} with high rate $\frac{1}{k}\mathcal{I}_c(\psi, \mathcal{N}^{\otimes k})$ (for fixed k).
- **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 such as entanglement.
- Most prominent: tensor networks [Fannes et al. 1992; Verstraete and Cirac 2004]
- More recent: **neural network states** [Carleo and Troyer 2017]
- Network architectures: restricted Boltzmann machines, **feedforward nets**, ...

Quantum states from feedforward nets



- Hidden layers $h^{(m)}$, output layer o: biases $b^{(m)}$ and weight matrices $W_{ij}^{(m)}$.
- $h^{(m)} = f_m\left(W^{(m)}h^{(m-1)} + b^{(m)}\right)$ with **activation function** f_m such as **rectified linear unit** $\mathrm{ReLU}(x) = \max\{0, x\}$ or **hyperbolic tangent** $\tanh(x)$.
- Up to normalization, $\psi(s^n) = o_1 + io_2$ (or $\psi(s^n) = \exp(o_1 + io_2)$).

Neural network states as quantum codes

- NN states are known to be capable of efficiently representing states such as graph states, surface codes, string-bond states and **general stabilizer states**.

  [Gao and Duan 2017; Glasser et al. 2018; Jia et al. 2018; Zhang et al. 2018]

- Versatile ansatz for multipartite entanglement $\longrightarrow$ ansatz for good quantum codes?

- **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 $\sigma_{RB^k} := (\mathrm{id}_R \otimes \mathcal{N}^{\otimes k})(\psi_{RA^k})$.
  3. For the mixed state σ_{RB^k} compute $\mathcal{I}_c(\psi_{RA^k}, \mathcal{N}^{\otimes k}) = S(B^k)_\sigma - S(RB^k)_\sigma$.
  4. Update $\{b_\vartheta, W_\vartheta\}$ using suitable optimization technique.

Optimization procedure

- Machine-learning typically uses gradient-based updates (such as ADAM or Adagrad).
- **Problem:** Coherent information of a very noisy channel has **lots of local maxima** given by product states.

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

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

Table of Contents

Dephrasure channel

- We apply the neural network state ansatz to find quantum codes for interesting channel models.
- Recently introduced: **dephrasure channel**. For $p, q \in [0,1]$, [FL, Leung, Smith 2018]

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

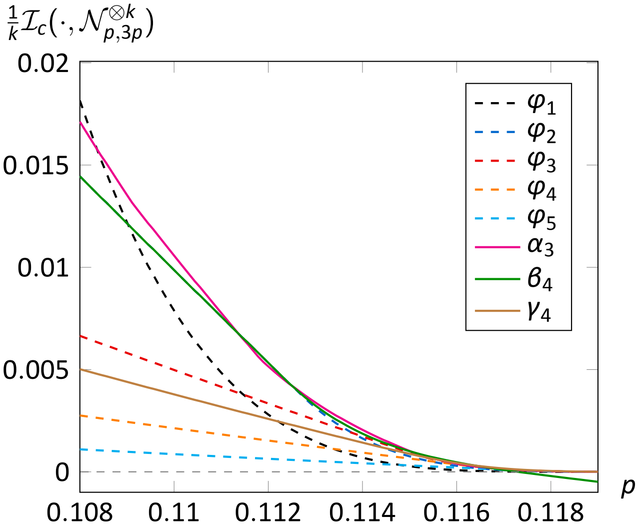
- **deph**asing + **erasure**: first dephase the input with probability p, then erase output with probability q.
- Dephrasure channel exhibits **substantial superadditivity effects**.
- Superadditivity achieved by, e.g., weighted repetition codes

$$\sqrt{\lambda}|0\rangle_R|0\rangle_A^{\otimes k} + \sqrt{1-\lambda}|1\rangle_R|1\rangle_A^{\otimes k}.$$

Dephrasure channel: NN setup

- For certain values of (p, q) and $k = 3, 4$ channel copies:
  NN ansatz finds quantum codes that **outperform all known ones**!
- **NN setup:** FF net with 4 hidden layers.
- **Activation functions:** Cos $\rightarrow$ ReLU $\rightarrow$ ReLU $\rightarrow$ ReLU.
- 182/306 real parameters vs. 128/512 in direct parametrization.
- Periodic activation function is non-standard in ML (sometimes even a no-go).
- However, Cos can easily implement parity check $\longrightarrow$ bias towards **degenerate codes**.
- Investigate dephrasure channel along $(p, 3p)$-diagonal in the (p, q)-plane.

Dephrasure channel: codes



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

p ... dephasing probability

$3p$ ... erasure probability

Weighted repetition code:

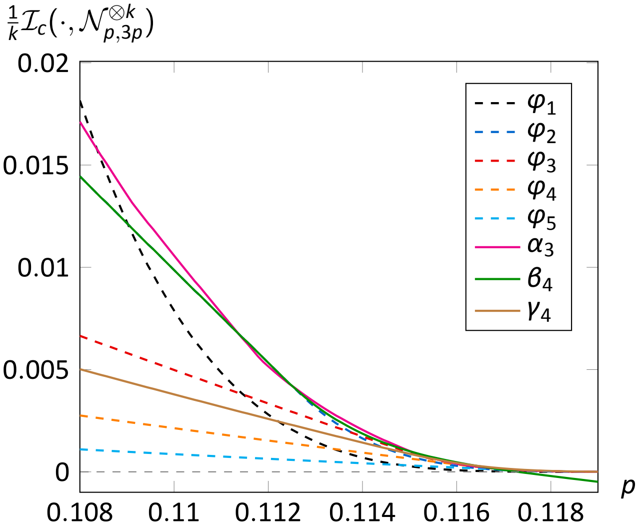
$$|\varphi_n\rangle = \sqrt{\lambda}|0\rangle_R|0\rangle_A^{\otimes k} + \sqrt{1-\lambda}|1\rangle_R|1\rangle_A^{\otimes k}$$

Neural network codes:

α_3, β_4, γ_4

(FF net with Cos-ReLU-ReLU-ReLU)

Dephrasure channel: codes



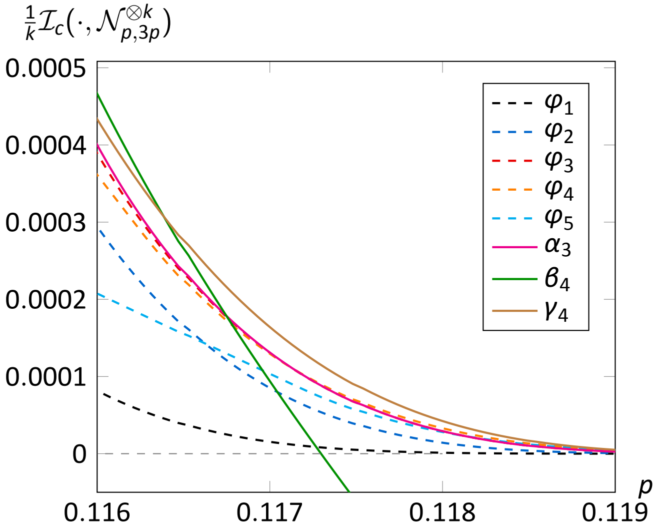
Neural network codes:

$|\alpha_3\rangle = a_1(|000\rangle_R|100\rangle_{A^3} + |110\rangle_R|001\rangle_{A^3}) +$
$a_2(|000\rangle_R|110\rangle_{A^3} + |110\rangle_R|011\rangle_{A^3}) +$
$a_3|001\rangle_R|000\rangle_{A^3} + a_4|011\rangle_R|111\rangle_{A^3}$

$|\beta_4\rangle =$
$b_1(|0000\rangle_R|0101\rangle_{A^4} + |1111\rangle_R|1010\rangle_{A^4}) +$
$b_2(|0100\rangle_R|0000\rangle_{A^4} + |1011\rangle_R|1111\rangle_{A^4})$

$|\gamma_4\rangle = |0110\rangle_R \otimes (c_1|0111\rangle_{A^4} +$
$c_2|1111\rangle_{A^4}) + c_3|1011\rangle_R|1000\rangle_{A^4}$

Dephrasure channel: codes



Neural network codes:

$|\alpha_3\rangle = a_1(|000\rangle_R|100\rangle_{A^3} + |110\rangle_R|001\rangle_{A^3}) + a_2(|000\rangle_R|110\rangle_{A^3} + |110\rangle_R|011\rangle_{A^3}) + a_3|001\rangle_R|000\rangle_{A^3} + a_4|011\rangle_R|111\rangle_{A^3}$

$|\beta_4\rangle = b_1(|0000\rangle_R|0101\rangle_{A^4} + |1111\rangle_R|1010\rangle_{A^4}) + b_2(|0100\rangle_R|0000\rangle_{A^4} + |1011\rangle_R|1111\rangle_{A^4})$

$|\gamma_4\rangle = |0110\rangle_R \otimes (c_1|0111\rangle_{A^4} + c_2|1111\rangle_{A^4}) + c_3|1011\rangle_R|1000\rangle_{A^4}$

Generalized amplitude damping channel

- Generalized amplitude damping channel (GADC) $\mathcal{N}_{\gamma,N}$:
  qubit in contact with thermal bath at non-zero temperature.
- Transition probability $\gamma \in [0,1]$, thermal bath temperature $N \in [0,1]$.
- Kraus operators of $\mathcal{N}_{\gamma,N}$ are:

$$K_1 = \sqrt{1-N}(|0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|) \qquad K_2 = \sqrt{\gamma(1-N)}|0\rangle\langle 1|$$

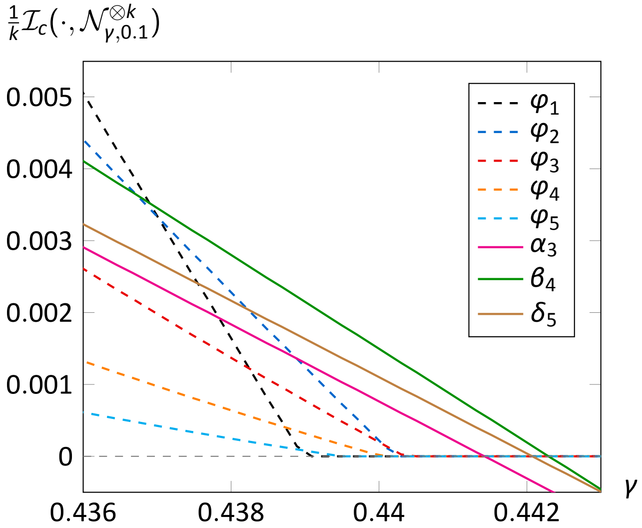
$$K_3 = \sqrt{N}(\sqrt{1-\gamma}|0\rangle\langle 0| + |1\rangle\langle 1|) \qquad K_4 = \sqrt{\gamma N}|1\rangle\langle 0|$$

- $N = 0$: thermal bath at zero temperature $\longrightarrow$ GADC reduces to amplitude damping.
- Realistic noise model in superconducting quantum computing. [Chirolli and Burkard 2008]
- GADC (for $N \notin \{0,1\}$) is neither degradable/antidegradable.

GADC: NN setup

- NN ansatz again finds quantum codes **outperforming all known ones**!
- **NN setup:** FF net with 4 hidden layers.
- **Activation functions:** Cos $\rightarrow$ Tanh $\rightarrow$ Tanh $\rightarrow$ Tanh.
- 182/306 real parameters vs. 128/512 in direct parametrization.
- Investigate GADC for $N = 0.1$ and interesting γ regime.
- Benchmark against weighted repetition codes.

GADC: codes



Channel: $\mathcal{N}_{\gamma,0.1}$

γ . . . transition probability

$N = 0.1$ . . . thermal bath temperature

Weighted repetition code:

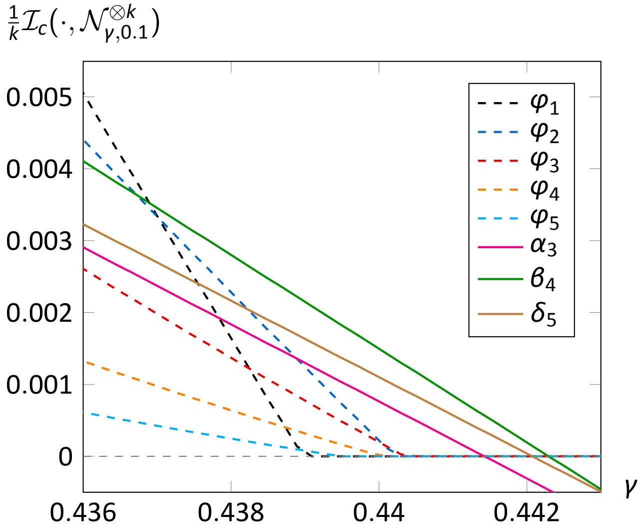
$|\varphi_n\rangle = \sqrt{\lambda}|0\rangle_R|0\rangle_A^{\otimes k} + \sqrt{1-\lambda}|1\rangle_R|1\rangle_A^{\otimes k}$

Neural network codes:

α_3, β_4, δ_5

(FF net with Cos-Tanh-Tanh-Tanh)

GADC: codes



Neural network codes:

$$|\alpha_3\rangle = a_1|000\rangle_R|000\rangle_{A^3} + a_2|110\rangle_R|000\rangle_{A^3} + |111\rangle_R \otimes (a_3|001\rangle_{A^3} + a_4|010\rangle_{A^3} + a_5|100\rangle_{A^3})$$

$$|\beta_4\rangle = |1110\rangle_R \otimes (b_1|0101\rangle_{A^4} + b_2|1010\rangle_{A^4}) + b_3|0001\rangle_R|1111\rangle_{A^4} + b_4|1000\rangle_R|1111\rangle_{A^4}$$

$$|\delta_5\rangle = |00011\rangle_R \otimes (c_1|01010\rangle_{A^5} + c_2|10101\rangle_{A^5}) + c_3|10110\rangle_R|11111\rangle_{A^5} + c_4|11101\rangle_R|11111\rangle_{A^5}$$

Depolarizing channel

- Depolarizing channel $\mathcal{D}_p$: For $p \in [0, 1]$,

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

- **Quantum capacity unknown** in the high-noise regime.
- In the interval $0.1889 \lesssim p \lesssim 0.1904$ and for up to $k = 10$ channel copies, best known codes are simple repetition codes $|0\rangle_R|0\rangle^{\otimes k} + |1\rangle_R|1\rangle^{\otimes k}$.
- **NN state ansatz + PSO reliably finds these codes after few optimization steps.**
- Rep. codes are **typically not found** with most common optimization techniques.
- Interesting: for $k \lesssim 6$ we couldn't find any better codes.

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 over direct parametrization (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 $S_\alpha(\rho) = \frac{1}{1-\alpha} \log \operatorname{Tr} \rho^\alpha$ with $\alpha \in \mathbb{N}$.
- Problem 1: Differences of Rényi entropies are problematic. [Linden et al. 2013]
- 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 $\mathcal{N}^{\otimes k}$.
- Numerically favorable implementations:
  - Sequential Kraus operator application (# op's $O(n)$):

$$\sigma_1 = (\mathrm{id} \otimes \ldots \mathrm{id} \otimes \mathcal{N})(\varphi) \longrightarrow \sigma_2 = (\mathrm{id} \otimes \ldots \otimes \mathrm{id} \otimes \mathcal{N} \otimes \mathrm{id})(\sigma_1) \longrightarrow \ldots$$

  - Transfer matrix formalism: translate 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 well for interesting channels such as **dephrasure channel**, **generalized amplitude damping channel**, and **depolarizing channel**.

Open problems

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

References

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

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

Chirolli, L. and G. Burkard (2008). *Advances in Physics* 57.3, pp. 225–285. eprint: https://doi.org/10.1080/00018730802218067.

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.

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

Glasser, I. et al. (2018). *Physical Review X* 8.1, p. 011006. arXiv: 1710.04045 [quant-ph].

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

Linden, N., M. Mosonyi, and A. Winter (2013). *Proc. Royal Soc. A: Math., Phys. and Eng. Sci.* 469.2158. arXiv: 1212.0248 [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.

Shor, P. and J. A. Smolin (1996). *arXiv preprint*. arXiv: quant-ph/9604006.

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!