Decoding Circuit-Level Noise with Machine Learning

John Blue[1,2], Harshil Avlani[1], Zhiyang He (Sunny)[3], Liu Ziyin[1,4], Isaac L. Chuang[1]

[1]Department of Physics, Massachusetts Institute of Technology
[2]The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
[3]Department of Mathematics, Massachusetts Institute of Technology
[4]Physics & Informatics Laboratories, NTT Research

Motivation and Procedure

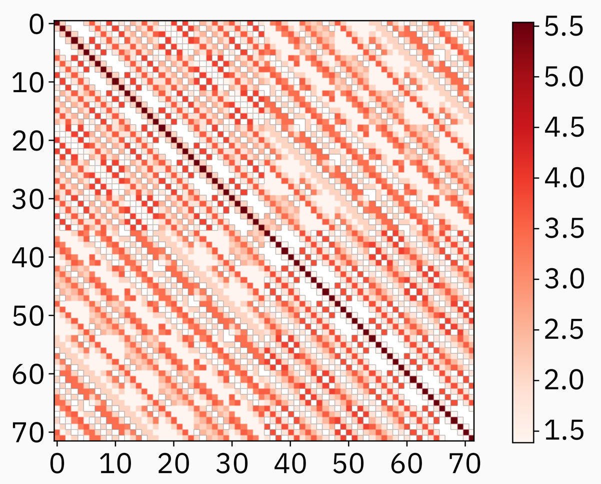
- Recently proposed quantum low-density-parity-check (qLDPC) codes [2, 9], using the belief propagation with ordered statistics decoding (BP-OSD) [7] have achieved pseudo-thresholds comparable with the surface code, but with much better encoding rates.
- However, the OSD step has a cubic runtime, leading to a desire for faster decoding algorithms.
- Inspired by the success of machine learning (ML) decoders for decoding the surface code [1, 3] and data qubit noise for qLDPC codes, [6], in this work we investigate using a transformer-based ML decoder for circuit level noise on qLDPC codes, focusing on the specific example of Bivariate Bicycle (BB) codes [2].

Masked Self-Attention

To include information about the code/syndrome measurement circuit, we use masked self-attention inspired by [4].

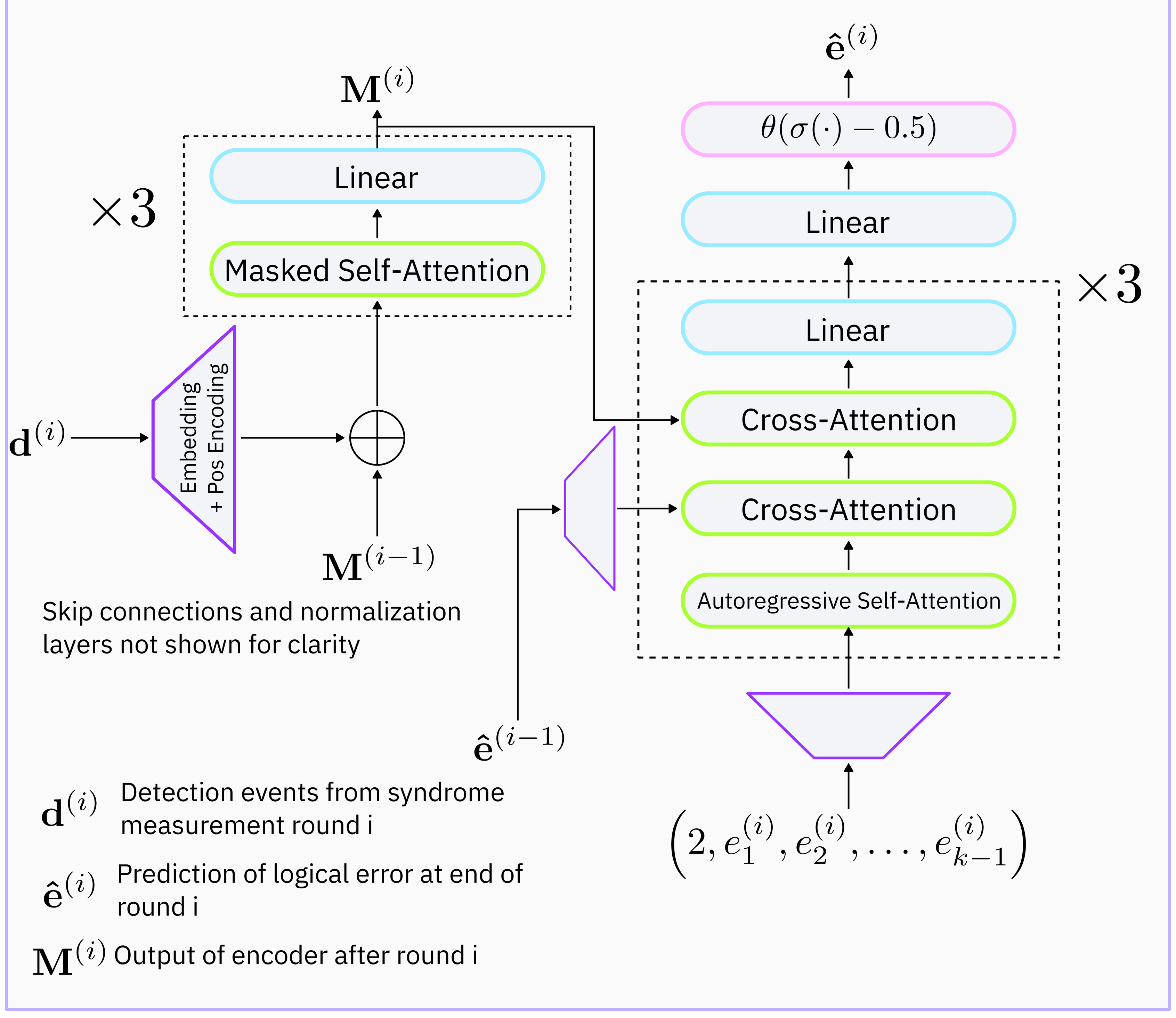
$$\alpha_{i,j} = \text{softmax}_j\left(\frac{\langle Q(x_i), K(x_j)\rangle}{\sqrt{k}} + M_{i,j}\right) \quad D_{i,j} = \begin{cases} 1 & \text{if error } j \text{ flips detector } i \\ 0 & \text{otherwise} \end{cases}$$

$$M = \log(DD^T) =$$



Example mask for [[72,12,6]] bivariate bicycle code (white indicates value of -∞).

Model Architecture



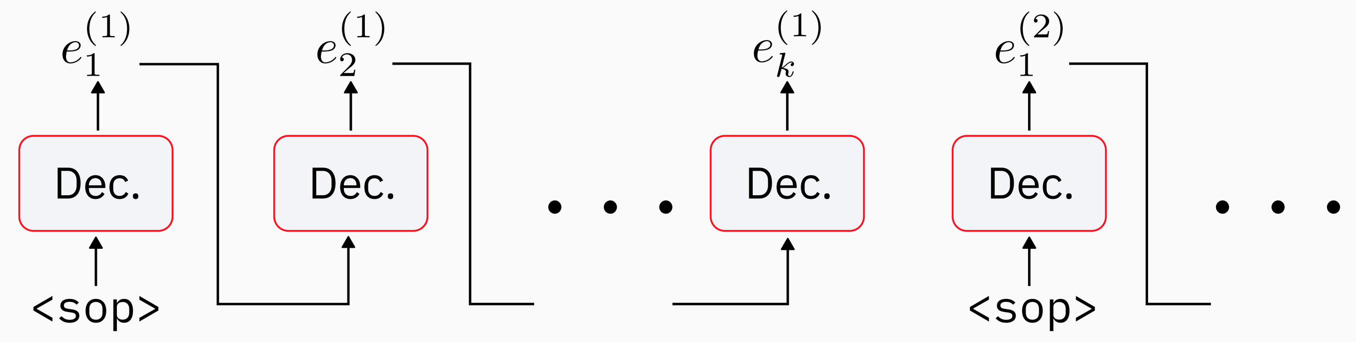
Acknowledgements

The authors would like to thank Andrew Cross, Ted Yoder, and Patrick Rall for helpful discussions throughout this project. This work is supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, http://iaifi.org/).

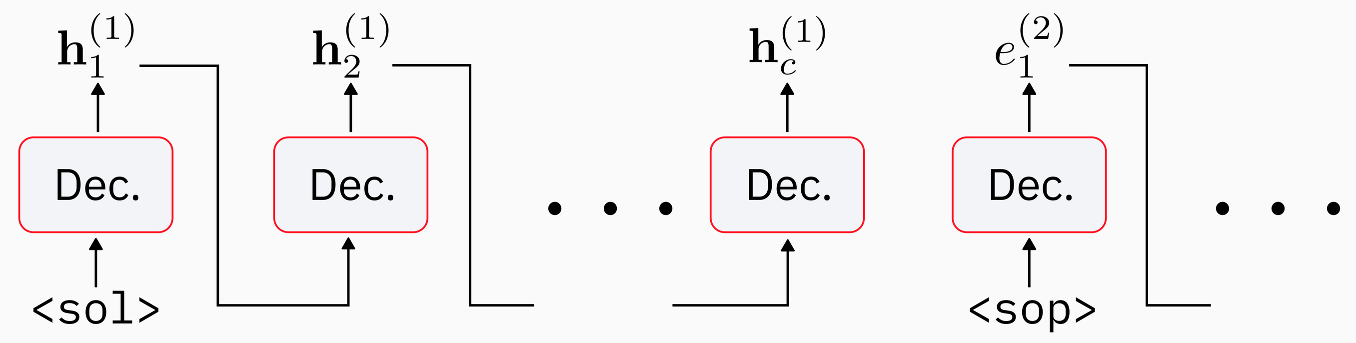
Training

We use a multi-step training process, similar to [5].

For the first step of training, the decoder autoregressively predicts each logical error for each round of syndrome measurement.



For the next step of training, we remove the final linear and sigmoid layers from the decoder for the first round of syndrome measurement, and generate c latent space predictions instead of k logical error predictions. The loss is computed using the predictions from the last N - 1 rounds.

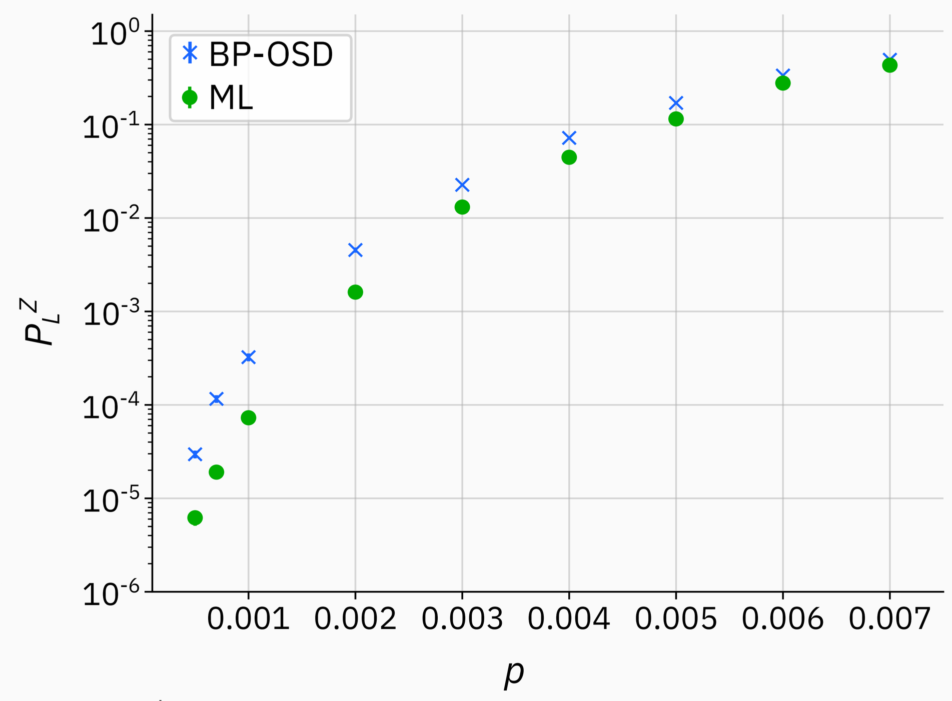


This process is repeated until logical error predictions are only made after the last round of syndrome measurement.

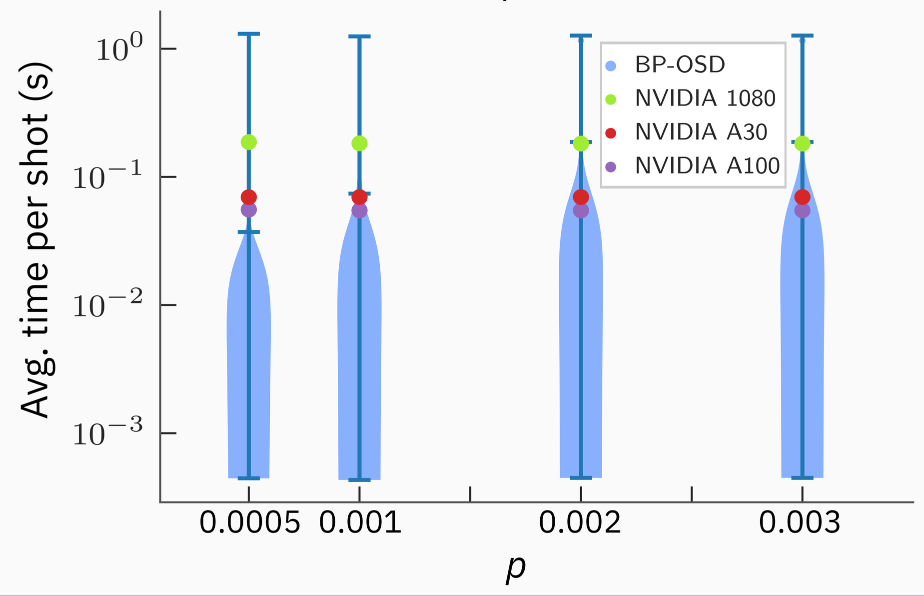
Results

To evaluate the model, we compare the logical error rates and runtimes to an open source BP-OSD implementation [8].

Logical vs physical error rate for the ML decoder and BP-OSD,evaluated on the [[72,12,6]] bivariate bicycle code with six rounds of noisy syndrome measurement followed by a single round of noiseless syndrome measurement. Error bars indicate one standard deviation. Third-order OSD was used for BP-OSD. Note that the ML decoder uses both X and Z check detection events, while BP-OSD only uses X check detection events.



Time per shot for the ML decoder on several different GPUs, and BP-OSD. The distribution of times for BP-OSD are shown via the violin plots, whereas the average time per shot of the ML decoder is shown as a horizontal line. The BP-OSD implementation was run on a AMD EPYC 9654 96-Core Processor (2.4 GHz).



Conclusions and Future Directions

The ML decoder outperforms BP-OSD on the [[72,12,6]] BB code, both in terms of logical error rate and run time. Next steps could include extending these results to larger BB codes, other qLDPC codes, as well as streaming decoding, as well as investigating noise models beyond simple depolarizing noise.

References

[1]: Johannes Bausch et al. "Learning high-accuracy error decoding for quantum processors". In: Nature (2024). doi: 10.1038/s41586-024-08148-8.
[2]: Sergey Bravyi et al. "High-threshold and low-overhead fault-tolerant quan- tum memory". In: Nature 627.8005 (Mar. 2024), pp. 778–782. issn: 1476- 4687. doi: 10.1038/s41586-024-07107-7
[3]: Hanyan Cao et al. qecGPT: decoding Quantum Error-correcting Codes with Generative Pre-trained Transformers. 2023. arXiv: 2307.09025 [quant-ph].
[4]: Yoni Choukroun and Lior Wolf. Error Correction Code Transformer. 2022. arXiv: 2203.14966 [cs.LG]
[5]: Shibo Hao et al. Training Large Language Models to Reason in a Continuous Latent Space. 2024. arXiv: 2412.06769 [cs.CL]
[6]: Arshpreet Singh Maan and Alexandru Paler. *Machine Learning Message- Passing for the Scalable Decoding of QLDPC Codes*. 2024. arXiv: 2408. 07038 [quant-ph].
[7]: Pavel Panteleev and Gleb Kalachev. "Degenerate Quantum LDPC Codes With Good Finite Length Performance". In: Quantum 5 (Nov. 2021), p. 585. issn: 2521-327X. doi: 10.22331/q-2021-11-22-585.
[8]: Joschka Roffe et al. "Decoding across the quantum low-density parity-check code landscape". In: Physical Review Research 2.4 (Dec. 2020). doi: 10. 1103/physrevresearch.2.043423.
[9]: Qian Xu et al. *Constant-Overhead Fault-Tolerant Quantum Computation with Reconfigurable Atom Arrays*. 2023. arXiv: 2308.08648 [quant-ph].