

# Attention-based Quantum Tomography

Peter Cha,<sup>1</sup> Paul Ginsparg,<sup>2</sup> Felix Wu,<sup>2</sup> Juan Carrasquilla,<sup>3,4</sup> Peter L. McMahon,<sup>5</sup> and Eun-Ah Kim<sup>1</sup>

<sup>1</sup>*Department of Physics, Cornell University, Ithaca, NY 14853, USA*

<sup>2</sup>*Department of Computer Science, Cornell University, Ithaca, NY 14853, USA*

<sup>3</sup>*Vector Institute, MaRS Centre, Toronto, Ontario, M5G 1M1, Canada*

<sup>4</sup>*Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada*

<sup>5</sup>*School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA*

With rapid progress across platforms for quantum systems, the problem of many-body quantum state reconstruction for noisy quantum states becomes an important challenge. Recent works found promise in recasting the problem of quantum state reconstruction to learning the probability distribution of quantum state measurement vectors using generative neural network models. Here we propose the “Attention-based Quantum Tomography” (AQT), a quantum state reconstruction using an attention mechanism-based generative network that learns the mixed state density matrix of a noisy quantum state. The AQT is based on the model proposed in “Attention is all you need” by Vishwani et al (2017) that is designed to learn long-range correlations in natural language sentences and thereby outperform previous natural language processing models. We demonstrate not only that AQT outperforms earlier neural-network-based quantum state reconstruction on identical tasks but that AQT can accurately reconstruct the density matrix associated with a noisy quantum state experimentally realized in an IBMQ quantum computer. We speculate the success of the AQT stems from its ability to model quantum entanglement across the entire quantum system much as the attention model for natural language processing captures the correlations among words in a sentence.

Recent developments in modern quantum devices [1] are bringing quantum systems with a large number of qubits within reach. However, characterization and validation of such quantum systems are essential for assessing progress, beyond mere qubit count, as devices are subject to noise. Comprehensive characterization of quantum systems can be performed through quantum state tomography [2], in which the density matrix describing the noisy quantum  $N_q$ -qubit many-body state is reconstructed from projective measurements on identically prepared copies of the quantum state. However, the exponential-in- $N_q$  Hilbert space of many-body states implies that exact tomography techniques, such as maximum likelihood estimation (MLE), require exponential-in- $N_q$  amount of data as well as an exponential-in- $N_q$  time for processing. Such prohibitive costs limit traditional methods of tomography to small system sizes  $N_q < 10$ . In fact, the tomographic measurement method that is integrated into IBM’s Qiskit library is limited to  $N_q = 3$ . Hence, many experiments rely on indirect methods of error determination, for example variants of randomized benchmarking [3]. Indeed there are efforts to estimate outcome probabilities showing promising scaling [4]. Nevertheless new strategies for the reconstruction of the full density matrix using experimentally accessible measurements for a general quantum state are much needed.

There has been a growing interest in using machine learning tools, such as deep neural networks, to overcome the curse of dimensionality through generative modeling [5–7]. The foundation for this approach was laid in Ref [8], which trained a restricted Boltzmann machine to represent complex quantum many-body states without

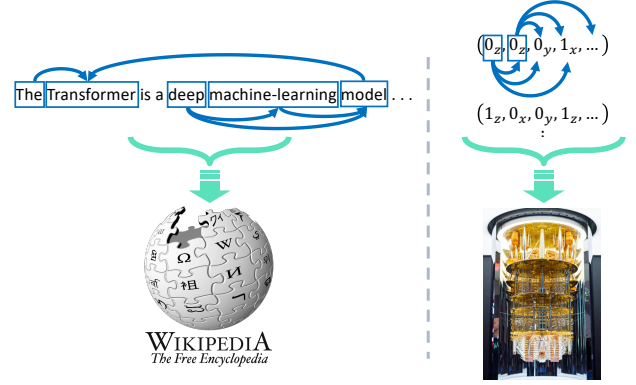


FIG. 1. We illustrate the self-attention mechanism, which is key to the success of the Transformer architecture. Left: The Transformer describes the “state” of Wikipedia after self-attention learns the correlations among words in the sentence specific to Wikipedia. Right: The AQT reconstructs the density matrix of a noisy mixed state (pictured: IBMQ) after self-attention learns the entanglement among qubits as it is reflected in projective measurements of the quantum state.

requiring exponentially many parameters or memory size. However, the expressibility of restricted Boltzmann machines and scalability of training is typically restricted to pure, positive quantum states [5, 8–11], and small mixed states [12], which limits their applicability at the scale of modern noisy quantum computers. In contrast, Ref. [6] demonstrated, using a recurrent neural network (RNN), that generative neural network models trained on informationally complete positive operator-valued measurements (IC-POVM) may be capable of providing a clas-

sical description of a noisy quantum many-body state. However, RNN-based tomography has so far only been demonstrated on classically simulated data, and despite promising indications, its ability to reconstruct a full density matrix has not been demonstrated even in simulation.

Here, we propose “Attention-based Quantum Tomography” (AQT) by adapting the attention-based Transformer architecture, a generative neural network model recently developed for natural language processing tasks [13], for the task of quantum state tomography. We begin by giving an intuition behind the Transformer and the rationale for its suitability for the tomography task. We then demonstrate that AQT outperforms a previous RNN-based approach by an order-of-magnitude reduction in the sample complexity of the reconstruction procedure. We also simulate a simple faulty-qubit model and demonstrate the promise of AQT in the task of mixed-state reconstruction. Next, we deploy AQT on experimental data from IBMQ’s quantum computer, showing strong qualitative agreement with MLE. Finally, we explore scalability by reconstructing a quantum state with a system size that exceeds the reach of the tomographic tools offered publicly by IBMQ.

The rationale behind the AQT is our observation of a promising parallel between the task of natural language processing (NLP) and scalable quantum state tomography (see Fig. 1). For a NLP task, one would like to “learn” the abstraction at the core of the language, or a “state”, from a set of “sentences” that constitute an extraordinarily tiny fraction of the complete set of all possible word combinations. For scalable quantum state tomography, one would like to “learn” the underlying quantum state from a set of projective measurements that constitute a tiny fraction of the complete set of projective measurements. The attention-based NLP models achieve such seemingly impossible “efficient learning” by learning long-range correlations among words in sentences, e.g., recognizing how a word at the end of a sentence may determine the meaning of a word at the beginning. For generative neural networks, learning the long-range correlations present in a sentence directly translates into learning the language. Given that learning the entanglement among qubits is the key to the tomography of a many qubit state, we can anticipate that the self-attention aspect of the Transformer architecture to be well-adapted to the tomography task.

The Transformer [13], which employs the “self-attention” mechanism [14–16], has been shown to be a dramatic step forward in efficiency and accuracy compared to previous state-of-the-art NLP models such as RNN [17, 18]. Before the Transformer, NLP tasks primarily relied on the RNN architecture [19, 20], which incorporates correlations between words by passing an encoded “memory” of the words going back to the beginning of the sentence as each new word was read in se-

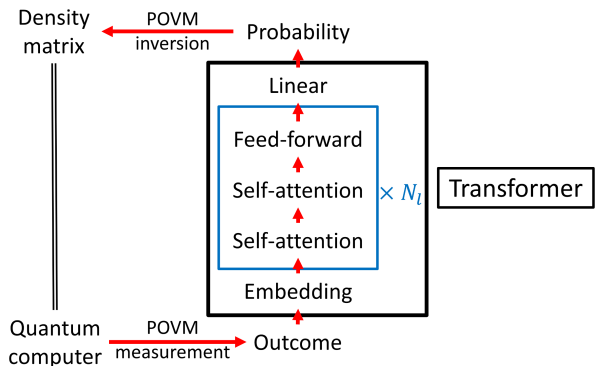


FIG. 2. A flowchart illustration of Attention-based Quantum Tomography (AQT), including a schematic diagram showing the internal structure of the Transformer neural network. In our work, we use  $N_t = 6$  layers and  $d = 256$  dimensional embedding.

quentially. However, the correlations captured in this approach are inherently short-ranged, as the encoded memory in a sequential model such as the RNN suffers from exponential suppression in correlation [21]. The challenge of long-range correlations in semantic modeling were addressed with the Transformer model, which uses “self-attention” to study correlations between all words in a sentence simultaneously, and learns which long-range correlations between which words are key to the meaning of the sentence. For example, the Transformer studies the set of sentences that make up Wikipedia (see Fig. 1) and learns the correlations within the words of each sentence over a series of self-attention layers. Once training is complete, the Transformer has learned an abstract representation of the full “state” of the language as represented by Wikipedia, and serves as a generative model from which sample sentences following the grammar structure, diction, style, etc, of Wikipedia can be drawn.

As many-body quantum states have no natural ordering of qubits and the states can have arbitrarily long-range entanglement, the aspect of the Transformer that learns the long-range correlation within different parts of the data is highly desirable. The idea is that what amounts to the abstract notion of the “state” of a given language that is expressed through the actual sentences maps on to the quantum state that is expressed through projective measurements. As we schematically depict in Fig. 2, the AQT learns the abstract representation of the quantum state from a set of POVM measurements and outputs the reconstructed density matrix. (Additional supplemental data will be posted along with the software in an update.)

Our target state both in experiment and in classical simulations will be the  $N_q$ -qubit Greenberger-Horne-Zeilinger (GHZ) state

$$|\text{GHZ}\rangle_{N_q} = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N_q} + |1\rangle^{\otimes N_q}), \quad (1)$$

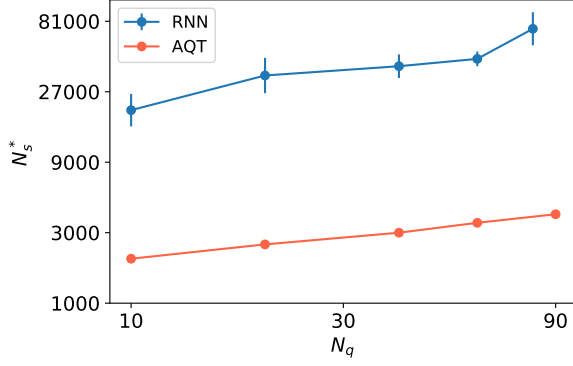


FIG. 3. Log-log plot demonstrating scaling of necessary sample size  $N_s^*$  for fixed classical fidelity  $F_C = 0.99$  in number of qubits  $N_q$  using RNN [6] and the AQT.

with system sizes ranging from  $N_q = 3$  to 90 qubits. We choose the GHZ because it is a pure state of interest for quantum communication protocols and we can benchmark our results against others in the literature, including those that do not reconstruct the full density matrix [4, 6]. In this work, we use the Pauli-4 POVM [6], which is easily measured in the IBMQ quantum computers.

A comprehensive measure of reconstruction is the quantum fidelity

$$F_Q(\rho_0, \rho_1) = \left( \text{Tr} \left[ \sqrt{\sqrt{\rho_0} \rho_1 \sqrt{\rho_0}} \right] \right)^2, \quad (2)$$

where  $\rho_0$  is the target density matrix against which we compare the reconstructed density matrix  $\rho_1$ . Quantum fidelity, however, in general requires full density matrix reconstruction [22]. In order to benchmark our results against the earlier works using neural networks, we first investigate the classical fidelity, which can be used when the state reconstruction only yields measurement probabilities:

$$F_C(p_0, p_1) = \sum_{\vec{a}} \sqrt{p_0(\vec{a}) p_1(\vec{a})}. \quad (3)$$

Here the sum is over all IC-POVM outcomes  $\vec{a} = (a_1, a_2, \dots, a_{N_q})$ ,  $a_i \in \{1, 2, \dots, N_a\}$  and  $p_0$  and  $p_1$  represent the measurement statistics of an IC-POVM over states  $\rho_0$  and  $\rho_1$ , respectively. Even though the classical fidelity contains a number of terms exponential in  $N_q$ , it is possible to estimate  $F_C(p_0, p_1)$  efficiently by sampling from the generative model representing  $p_1$ , i.e.,  $F_C(p_0, p_1) = \sum_{\vec{a}} p_1(\vec{a}) \sqrt{\frac{p_0(\vec{a})}{p_1(\vec{a})}} = \sum_{\vec{a} \sim p_1} \sqrt{\frac{p_0(\vec{a})}{p_1(\vec{a})}}$ . This choice is enabled by the Transformer architecture, which allows for both exact sampling from  $p_1$  and the exact calculation of  $p_1(\vec{a})$  for any choice of  $\vec{a}$  in polynomial time in  $N_q$ . However it should be noted that the classical fidelity only provides an upper bound on the quantum fidelity [6], and the discrepancy can be substantial [4].

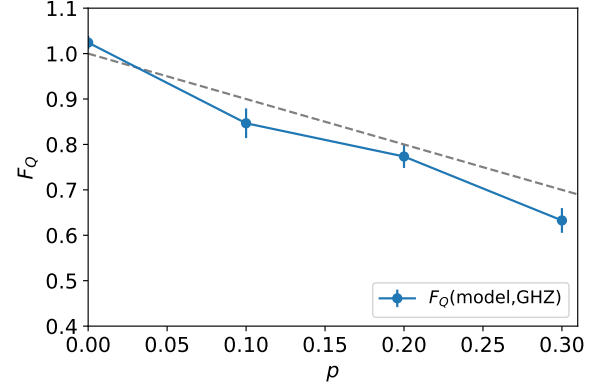


FIG. 4. Quantum fidelity  $F_Q$  of states reconstructed using AQT, where each state was created with error rate  $p$ , which is the error parameter in the simulation to be characterized by the reconstruction. From  $F_Q$  we can read off the error rate as  $p = 1 - F_Q$ . The expected result  $F_Q = 1 - p$  is plotted in dashed grey.

We first benchmark AQT against previous state-of-art neural quantum state tomography using RNN. Ref [6] studied an  $N_q$ -qubit GHZ state, with  $N_q = 10$  to 90, using classically sampled measurements, and demonstrated that  $N_s^*(N_q)$ , the minimum size of training data for which the RNN can achieve a classical fidelity of 0.99, scales linearly with  $N_q$ . In Fig. 3, we demonstrate a similar linear scaling in  $N_s^*$  vs  $N_q$  using the AQT, indicating that these natural language processing models can indeed learn the measurement statistics of a quantum state from an amount of data that grows sub-exponentially with the system size. More importantly the AQT exhibits an order-of-magnitude improvement in the sample complexity of learning the GHZ state compared to the RNN, as well as a smaller slope  $dN_s^*/dN_q$ .

We now investigate the AQT's performance on a mixed state with a built-in simulated error. We consider a 3-qubit GHZ system and assume there is one faulty qubit, which we pick to be qubit-0. We assume that the faulty qubit flips ( $0 \leftrightarrow 1$ ) with probability  $p$ . More precisely, this represents the mixed state

$$\rho_{err} = (1 - p)|\text{GHZ}\rangle_3 \langle \text{GHZ}|_3 + p|\psi\rangle_3 \langle \psi|_3, \quad (4)$$

where  $|\psi\rangle_3 = \frac{1}{\sqrt{2}}(|100\rangle + |011\rangle)$ . For the small number of qubits that we study, we are able to compute the exact quantum fidelity. First, we consider the fidelity between the reconstructed state and the noisy state in Eq. 4, for which we find  $F_Q(\rho_{\text{model}}, \rho_{\text{err}}) = 1$  within statistical error. This demonstrates that the AQT is sufficiently expressive to support a successful training procedure. To facilitate comparison to an experimental setting where  $p$  is *a priori* unknown, we compute the fidelity of the “realized” density matrix  $\rho_{\text{model}}$  to the “target” density matrix  $\rho_{\text{GHZ}}$ , which is the error-free pure GHZ state. The numerical results for  $p = 0.0 \sim 0.3$  displayed in Fig. 4 are

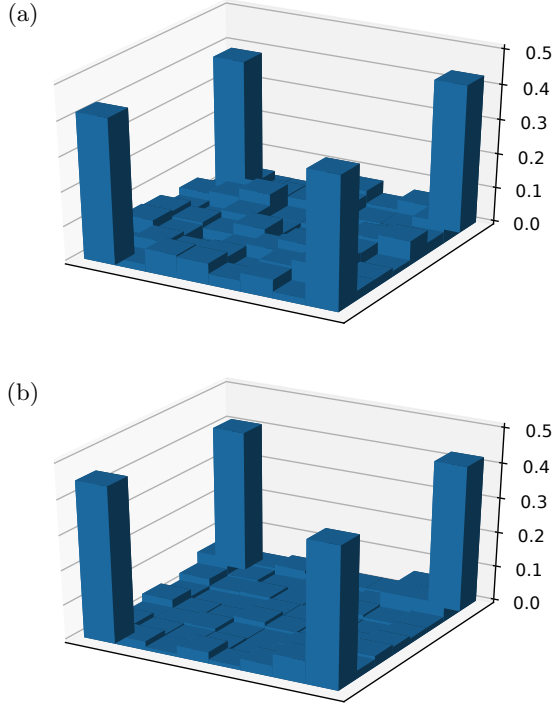


FIG. 5. Benchmarking AQT (a) to MLE tomography offered by IBM’s Qiskit library (b) for a noisy 3-qubit GHZ state data generated on the IBMQ\_OURENSE quantum computer. Each bar represents the absolute value of a density matrix (DM) element.

consistent with the expectations from the built-in error.

Next, we benchmark the AQT against the MLE algorithm that is built into the IBM Qiskit library by performing tomography using the two approaches on the measurements taken on IBMQ\_OURENSE on a 3-qubit system (Fig. 5). For the reconstruction we took 100 measurements in each of  $3^3 = 27$  possible measurement configurations, for a total of 2,700 measurements. Fig 5 shows two reconstructed density matrices using the usual graphical representation. Here, each bar represents a matrix element, in general complex, with the bar height set by its absolute value. The tall bar to the left is the density matrix element  $|000\rangle\langle 000|$ , the bar to the rear is  $|000\rangle\langle 111|$ , and so on. Note that the matrix elements represented by the bars in the rear and front are related by complex conjugation. The AQT-reconstructed density matrix is in strong qualitative agreement with the MLE reconstruction, capturing the error in realizing the GHZ state on the quantum computer. From the Transformer reconstruction, we find an exact quantum fidelity to the target pure GHZ state of  $F_Q = 0.81$ , while the MLE reconstruction has fidelity 0.84. These results give a mutually consistent estimation of the reliability of the IBMQ\_OURENSE quantum computer. The advantage of AQT as compared to exact tomographic methods such as

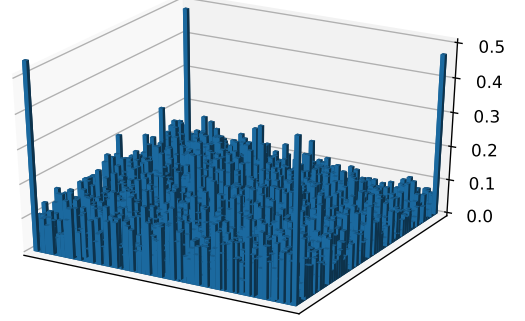


FIG. 6. Reconstructed density matrix of the 6-qubit GHZ state, using classically generated data. Each bar represents the absolute value of a matrix element.

MLE is that AQT can be scaled to larger systems.

To explore the scalability of the AQT, we reconstruct the density matrix for a 6-qubit GHZ state, already beyond the tomography functionality offered in Qiskit (see Fig. 6, and the density matrix included as supplemental file with this submission). We use classically generated data sampled from the noise-free GHZ state rather than data from IBMQ. (This is due in part to limited number of POVM measurements publicly accessible with IBMQ.) For the 6-qubit reconstruction, we use a total of 20,000 measurements. The reconstructed IC-POVM probability distribution  $p_1$  (see Eq. (3)) is in excellent agreement with the GHZ state, as expected from Fig. 3. Namely, achieving classical fidelity of  $F_C = 0.99$ , which directly measures the accuracy of  $p_1$  reconstruction, for a 6-qubit state requires even fewer than the 3,000 measurements required for the 10-qubit state. On the other hand, the reconstruction of the full density matrix  $\rho_1$  shows noise even with 20,000 measurements, though it is still in reasonable agreement with the GHZ state. In general, an accurate reconstruction of  $\rho_1$  requires much more data and computing time than an accurate reconstruction of  $p_1$ , since even small errors in  $p_1$  are amplified into large errors in  $\rho_1$ . This is a restatement of the well-known fact that classical fidelity is an upper bound on quantum (exact) fidelity. Error-scaling analysis in the number of samples is in general NP-hard [23] and remains an open question in AQT.

In summary, we proposed the AQT which adopts elements of the Transformer, a generative deep neural network for NLP, to the task of quantum state tomography. The AQT outperformed earlier neural quantum tomography based on the RNN architecture on an identical task, demonstrating a significant enhancement in the sample complexity of the reconstruction. This suggests that the AQT provides a nontrivial inductive bias suitable for the reconstruction of entangled states such as the ones considered in our experiments. We constructed

a qubit-error model and showed that AQT provides a reliable estimate of *a priori* unknown quantum mixed states and error rates in our specific setting. We then demonstrated for the first time that a machine learning based tomographic technique can reliably reconstruct the noisy density matrix of a quantum computer, by reconstructing the 3-qubit GHZ state realized by a quantum computer provided publicly by IBMQ. Furthermore, using AQT we have reconstructed a 6-qubit GHZ state, which is a tomography task of a size beyond the reach of the tomography functionality in IBM Qiskit's software.

To the best of our knowledge, AQT represents the first machine-learning based approach to successfully reconstruct density matrices describing the states produced in an experimentally realized quantum computer. AQT holds much promise for future progress. This work has been largely based on the GHZ state, facilitating a comparison with previous works without full density matrix reconstruction. Nevertheless the AQT is not inherently limited to a special pure state, and an examination of how  $N_s^*$  scales with  $N_q$  in states with more complex entanglement will provide much insight into machine-learning based tomography. Tests on a bigger experimental system and other architectures will help us determine the full scalability of the AQT. Furthermore, whether the AQT approach can build on the initial insight from our elementary error model towards more sophisticated error modeling and assessment to complement gate-set tomography [24, 25] would be also an interesting direction.

**Acknowledgements:** We thank Roger Melko, Giuseppe Carleo, Giacomo Torlai, Mikhail Lukin, Markus Greiner for useful discussions. PC, FW, and E-AK are supported by NSF HDR-DIRSE award number OAC-1934714 and in part by the Cornell Center for Materials Research with funding from the NSF MRSEC program (DMR-1719875). JC acknowledges support from Natural Sciences and Engineering Research Council of Canada (NSERC), the Shared Hierarchical Academic Research Computing Network (SHARCNET), Compute Canada, Google Quantum Research Award, and the Canadian Institute for Advanced Research (CIFAR) AI chair program.

- 
- [1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, *Nature* **574**, 505 (2019).
  - [2] M. Paris and J. Rehacek, *Quantum state estimation*, Vol. 649 (Springer Science & Business Media, 2004).
  - [3] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, *Physical Review A* **93**, 012301 (2016).
  - [4] H.-Y. Huang, R. Kueng, and J. Preskill, "Predicting

- many properties of a quantum system from very few measurements," (2020), arXiv:2002.08953.
- [5] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nature Physics* **14**, 447 (2018).
- [6] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, *Nature Machine Intelligence* **1**, 155 (2019).
- [7] J. Carrasquilla, D. Luo, F. Prez, A. Milsted, B. K. Clark, M. Volkovs, and L. Aolita, "Probabilistic simulation of quantum circuits with the transformer," (2019), arXiv:1912.11052 [cond-mat.str-el].
- [8] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [9] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. Zibrov, M. Endres, M. Greiner, V. Vuletic, and M. Lukin, *Nature* **551** (2017), 10.1038/nature24622.
- [10] G. Torlai, B. Timar, E. P. L. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, *Phys. Rev. Lett.* **123**, 230504 (2019).
- [11] I. J. S. D. Vlugt, D. Iouchtchenko, E. Merali, P.-N. Roy, and R. G. Melko, "Reconstructing quantum molecular rotor ground states," (2020), arXiv:2003.14273.
- [12] G. Torlai and R. G. Melko, *Phys. Rev. Lett.* **120**, 240503 (2018).
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, in *NIPS* (2017).
- [14] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, arXiv preprint arXiv:1406.1078 (2014).
- [15] J. Cheng, L. Dong, and M. Lapata, arXiv preprint arXiv:1601.06733 (2016).
- [16] A. P. Parikh, O. Täckström, D. Das, and J. Uszkoreit, arXiv preprint arXiv:1606.01933 (2016).
- [17] S. Hochreiter and J. Schmidhuber, *Neural computation* **9**, 1735 (1997).
- [18] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, arXiv preprint arXiv:1412.3555 (2014).
- [19] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, *et al.*, arXiv preprint arXiv:1609.08144 (2016).
- [20] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, and J. Dean, arXiv preprint arXiv:1701.06538 (2017).
- [21] H. Shen, arXiv:1905.04271 [cond-mat, stat] (2019), arXiv:1905.04271 [cond-mat, stat].
- [22] Recent work [4] showed it is possible to estimate quantum fidelity without density matrix reconstruction in the special case when the target state is a pure state.
- [23] D. Suess, ukasz Rudnicki, T. O. Maciel, and D. Gross, "Error regions in quantum state tomography: computational complexity caused by geometry of quantum states," (2016), arXiv:1608.00374.
- [24] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, arXiv preprint arXiv:1310.4492 (2013).
- [25] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, *Nature communications* **8**, 1 (2017).