Molecular Quantum Transformer

Yuichi Kamata,* Quoc Hoan Tran, Yasuhiro Endo, and Hirotaka Oshima
Quantum Laboratory, Fujitsu Research, Fujitsu Limited, Kawasaki, Kanagawa 211-8588, Japan

The Transformer model, known for its powerful attention mechanism, has achieved state-of-the-art performance in various artificial intelligence tasks but faces challenges such as high computational cost and memory usage. Researchers are exploring quantum computing to enhance the Transformer's design. We propose the Molecular Quantum Transformer (MQT), which utilizes quantum circuits to implement the attention mechanism for modeling interactions in molecular quantum systems. MQT can efficiently calculate ground state energies for all configurations of molecules. In the calculation of ground state energies for H_2, LiH, BeH_2, and H_4, MQT outperforms the classical Transformer, demonstrating a potential advantage of the quantum structure. Our method offers an alternative to existing quantum algorithms like quantum phase estimation and variational quantum eigensolver, opening new avenues in quantum chemistry and materials science.

Background.— The Transformer model [1] has been recognized as a remarkable advancement in artificial intelligence. Its key power lies in its "attention mechanism", which discerns the relative importance of different parts of its input and the connection strengths between them. This mechanism has been successfully applied to both natural language processing and visual tasks, delivering state-of-the-art performance across a variety of datasets. Despite its many advantages and successes, the current implementation of the Transformer has several notable disadvantages, such as high computational cost, significant memory usage, the requirement for large datasets, and a large number of training parameters. Considering these aspects, researchers are actively seeking improved designs for the Transformer.

In the marriage of quantum computing and machine learning to establish the quantum machine learning (QML) research field, there is hope to build the Transformer model on quantum computers to solve problems impractical for classical computers. Several implementations have been proposed in this area, including quantum self-attention neural networks [2, 3] and quantum vision Transformers [4]. However, these implementations still have very limited performance on tasks involving classical data, such as text and image classification.

Since there is no clear evidence that classical data, such as text and images, inherently require quantum effects, the research community is shifting to using QML methods that utilize quantum effects on data originating from quantum systems. In this line, we propose a design for a quantum Transformer model called Molecular Quantum Transformer (MQT) to calculate molecular ground state energies. MQT leverages the attention mechanisms through quantum circuits to model interactions within molecular quantum systems more efficiently.

Remarkably, with access to random molecular configurations during each training iteration, MQT can be used to obtain the ground state energies for all molecular configurations, thereby reconstructing the potential energy surface (PES) of molecules. Our approach serves as an alternative to current algorithms for calculating ground state energies, such as Quantum Phase Estimation (QPE) and Variational Quantum Eigensolver (VQE). For instance, independently running QPE or VQE for each configuration would incur significant computational costs. Instead, estimating the PES for all molecular configurations is managed by the efficient processing capabilities of MQT, which can effectively capture and model complex interactions in quantum systems.

Molecular Quantum Transformer (MQT).— The MQT model for calculating the ground state energy of molecules is illustrated in Fig. 1. As a preprocessing step, we use PennyLane's built-in functions [5] to compute the qubit representation with the number of qubits n_q and the Hamiltonian H from the molecular structure. The input token $electron_*$-$nucleus_*$ represents the electronic state as a $n \times m$ two-dimensional array of 8-dimensional features. Here, n and m denote the number of electrons and nuclei, respectively. Consequently, the array of all tokens forms an $n \times m \times 8$ matrix. Similar to token preparation in Ref. [6], in the token matrix, each 8-dimensional feature input vector $\boldsymbol{x}$ is calculated from a vector concatenating the relative positions of electrons to each nucleus and their distances, processed through a fully-connected (FC) layer. The input $\boldsymbol{x}$ is then processed using a quantum transformer of L layers ($L = 6$ in our experiments). The output $\boldsymbol{y}^{(L)}$ of the final layer is then exponentially attenuated by the distance r_{ae} between the nuclei and electrons as $e^{-r_{\mathrm{ae}}} \times \boldsymbol{y}^{(L)}$. An FC layer applies this attenuated feature to obtain the n_q qubits state vector. After converting the features of each token into state vectors, the average of all these vectors is calculated to obtain a n_q qubits state vector. It is then added with the Hartree-Fock state vector to generate the final electronic state vector via amplitude embedding. The expectation value of H is then calculated from this amplitude embedding representation to be minimized.

The architecture of the quantum Transformer component in MQT is depicted in Fig. 2(a). First, the input features of classical data are converted into quantum states through an embedding layer. Query, Key, and Value transforma-

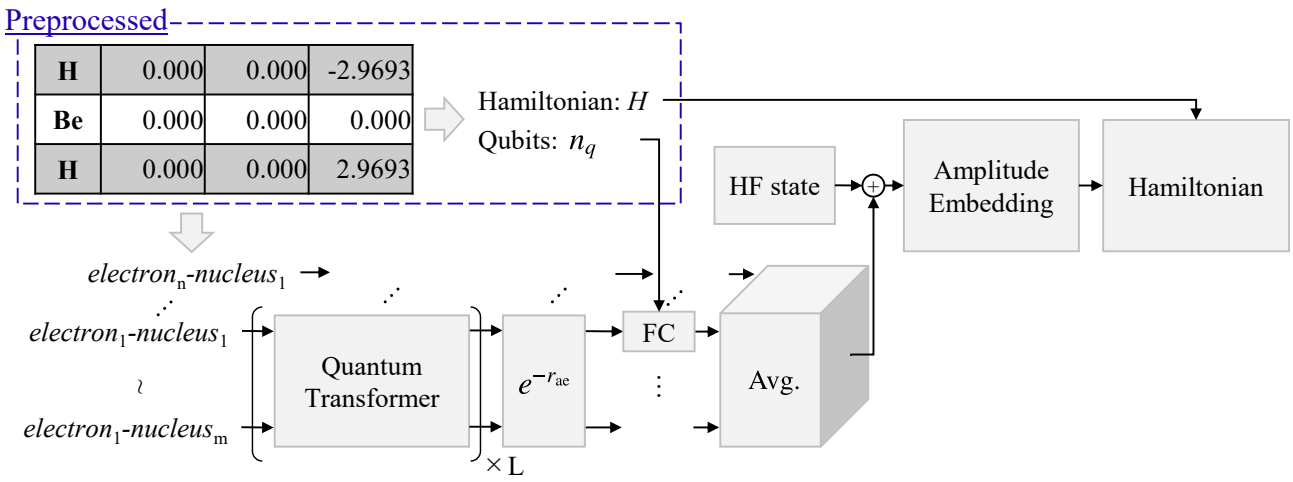

FIG. 1. The Molecular Quantum Transformer (MQT) model for the calculation of the ground state energy

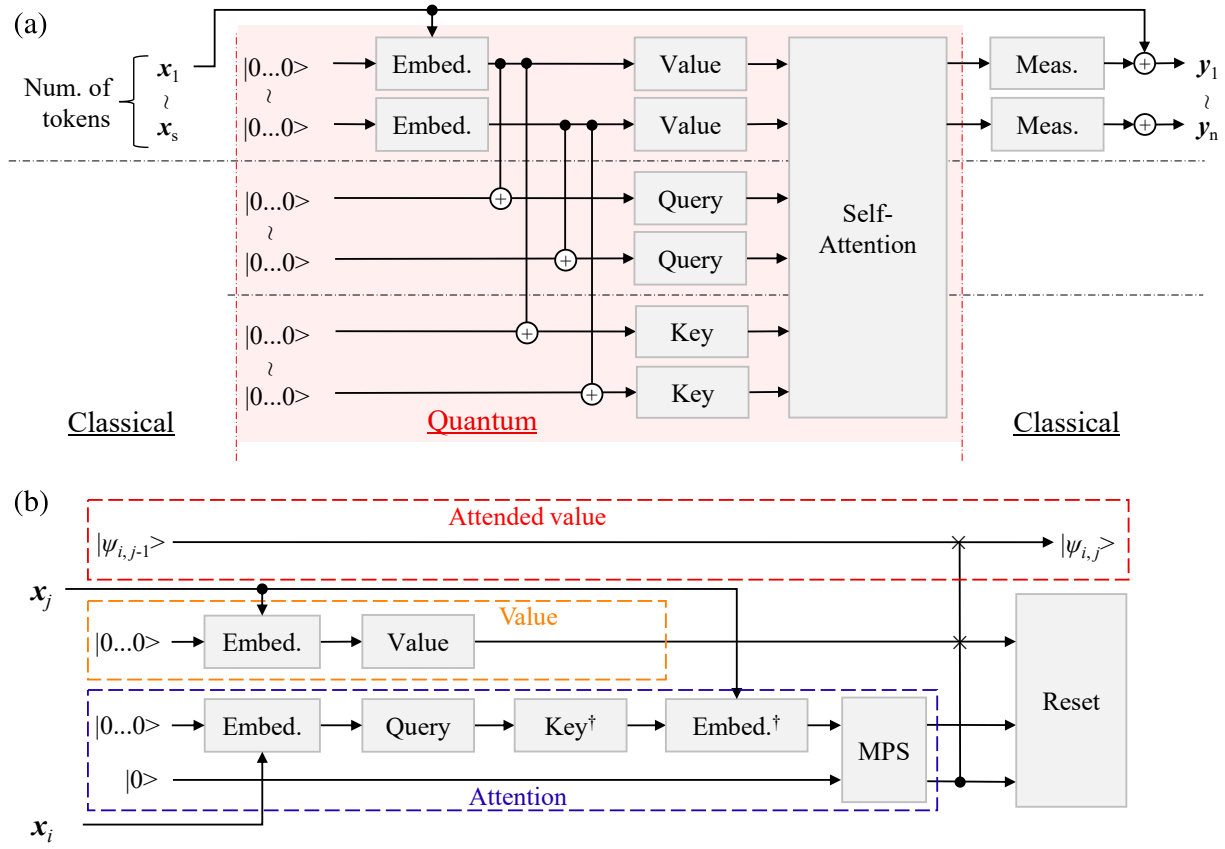

FIG. 2. (a) Quantum Transformer using the quantum attention mechanism. (b) The update of Attended value.

tions, constructed using quantum ansatz, are then performed to calculate the Attended value using the self-attention mechanism. The result is measured and converted back into classical features, which are added to the input features via a residual connection to generate the features for the next layer's input.

The detailed calculation of the Attended value update ($|\psi_{i,j-1}\rangle \rightarrow |\psi_{i,j}\rangle$) is shown in Fig. 2(b). First, for a selected input token $\boldsymbol{x}_i$ from the token sequence, an angle embedding by Y rotation gate R_Y (Embed.) is performed using the second auxiliary token register. Sequentially, a Query transformation and an adjoint matrix of Key transformation, both of which are constructed by a single layer of strongly entangling layer (StrEnt) ansatz,

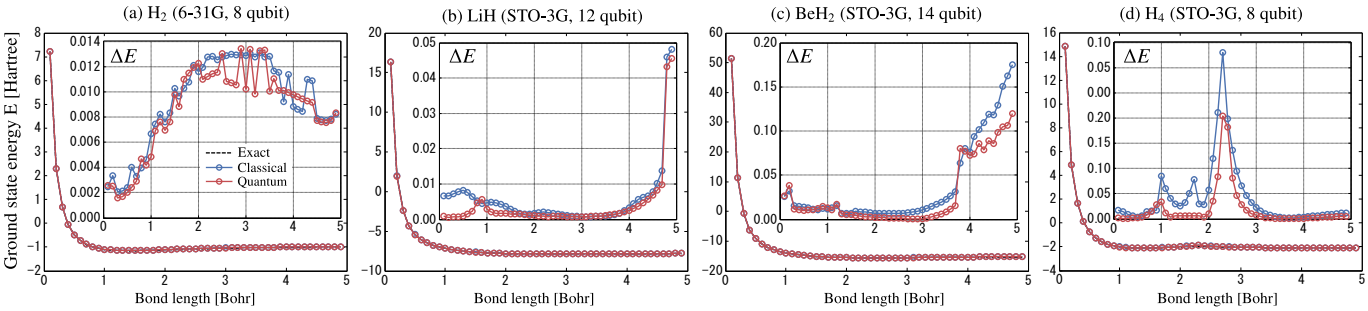

FIG. 3. Potential energy curves and estimation errors (ΔE) for varying interatomic bond lengths in (a) H_2 (b) LiH, (c) BeH_2, and (d) H_4 molecules using the quantum (red lines) and classical (blue lines) Transformers.

and an adjoint matrix of the angle embedding (Embed.†) for another selected input token $\boldsymbol{x}_j$ from the token sequence are applied to compute the Hadamard product of the query and key. This product is then converted into a 1-qubit Attention representation using a 2-bond matrix product state (MPS) ansatz implemented with learnable R_Y and an additional auxiliary qubit. Next, a Value transformation for $\boldsymbol{x}_j$ is performed using the first auxiliary token register with the angle embedding (Embed.) and six layers StrEnt ansatz. Finally, a controlled SWAP gate is used to apply the feature update corresponding to the Attention representation and the value to the Attended value.

Results.— In the following numerical experiments to estimate the potential energy curves, we compare the performance of MQT with the corresponding implementation using the classical Transformer (with 8 hidden dimension and single attention head) instead of the quantum Transformer module depicted in Fig. 1.

In the second quantization of Hamiltonian, we use the Bravyi-Kitaev transformation for the molecules H_2, LiH, and BeH_2, and Jordan-Wigner transformation for H_4. In each iteration of the Transformer, the token matrix is created from the configuration with a random bond length in (0.0, 5.0) [Bohr]. After the training, the potential energy curves for bond lengths from 0.1 to 4.9 [Bohr] with 0.1 [Bohr] increments were estimated using the trained model. We use AdamW optimizer with a weight decay rate of 0.001. The learning rate is set to 0.008 for the classical Transformer in the case of H_2; and 0.004 otherwise. In the optimization, we also employ the method in Ref. [7] to run 4 processes in parallel on a single GPU, with each process performing 2500 iterations, totaling 10^4 iterations.

Figure 3 depicts the energy estimation compared to the theoretically calculated ground state energy of the Hamiltonian for each molecule. In the STO-3G basis set, MQT demonstrates lower estimation errors than the classical Transformer for (b) LiH, (c) BeH_2, and (d) H_4. In the 6-31G basis set for (a) H_2, the estimated errors of the MQT are almost equivalent to those of the classical Transformer (note that the scaling of the y-axis in each plot is different). From these results, we can conclude that the MQT performs as well as, or even better than, the classical Transformer, demonstrating the potential benefits of incorporating quantum structure into the Transformer.

Conclusion.— We propose the Molecular Quantum Transformer (MQT), which leverages quantum circuits to revolutionize the calculation of ground state energies for all molecular configurations, offering potential improvements in accuracy over classical Transformers. This approach could open new avenues in quantum chemistry and materials science, where accurate understanding of ground state energies is crucial.

* kamata.yuichi@fujitsu.com

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, in *Advances in Neural Information Processing Systems*, Vol. 30 (Curran Associates, Inc., 2017).
[2] J. Shi, R.-X. Zhao, W. Wang, S. Zhang, and X. Li, arXiv 10.48550/arxiv.2207.07563 (2022).
[3] G. Li, X. Zhao, and X. Wang, Sci. China Inf. Sci. **67**, 142501 (2024).
[4] E. A. Cherrat, I. Kerenidis, N. Mathur, J. Landman, M. Strahm, and Y. Y. Li, Quantum **8**, 1265 (2024).
[5] V. Bergholm and et al., arXiv 10.48550/arXiv.1811.04968 (2022).
[6] I. von Glehn, J. S. Spencer, and D. Pfau, in *The Eleventh International Conference on Learning Representations* (2023).
[7] B. Recht, C. Re, S. Wright, and F. Niu, in *Advances in Neural Information Processing Systems*, Vol. 24 (Curran Associates, Inc., 2011).