Improved Neural Quantum State Tomography with Teacher Forcing and Classical Shadow

Taihei Takahashi (Keio Univ・Yamamoto Lab)

Abstract

We propose "Pauli NQS," a novel approach to address the critical issue of learning instability in conventional Neural Quantum State (NQS) tomography. The objective of NQS tomography with clasical shadows is to reduce the cost of storing big shadow data and computation for estimating expectation values. Our autoregressive Transformer model directly takes sequences of Pauli operators as input and outputs their corresponding measurement probabilities. To overcome the inherent instability and slow convergence of traditional autoregressive models, we introduce Teacher Forcing, which uses reliable estimates of conditional probabilities as guidance during training.
We demonstrate the effectiveness of our method on the 3-qubit Greenberger–Horne–Zeilinger (GHZ) state. The results show that the Pauli NQS model trained with Teacher Forcing converges significantly faster and achieves a more stable, lower loss compared to the model trained without it. This work establishes a robust and efficient framework for NQS-based tomography, paving the way for reliable applications on larger-scale quantum systems.

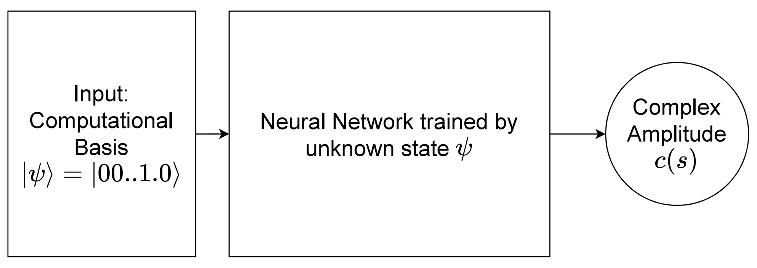
Introduction | Building on NQS, Shadows, and Transformers

Background①

Neural Quantum State (NQS)[0]

Simulating quantum systems is plagued by the "curse of dimensionality." Carleo and Troyer [5] introduced **Neural Quantum States (NQS)**, using a neural network represents the complex wavefunctions of the quantum state ψ which estimates $\{c(s)\}$ where:

$$|\psi\rangle = \sum_{s\in B} c(s)|s\rangle.$$



B is the set of binary vectors.

Background②

Classical Shadows[1]

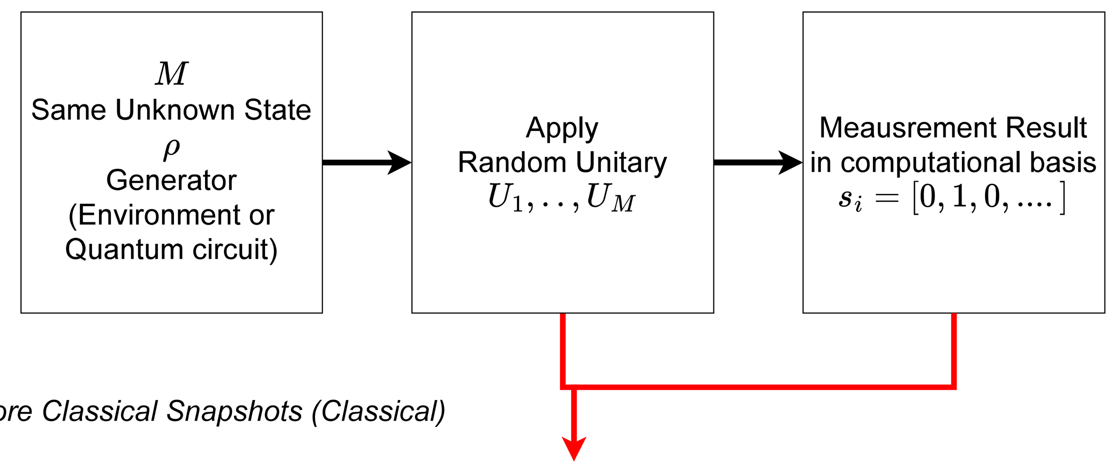
Objective : Estimate measurement expectation values $\langle O_1\rangle_\rho, ..., \langle O_N\rangle_\rho$ of

- N observables $\{O_1, ..., O_N\}$ (Each O_k is the sum of polynomial number Pauli operators).
- M copies of Unknown state ρ
- Small error ε.

Classical Shadow is one of the most efficient algorithm (information theoretically proved) w.r.t. M, N, ε.

Method :

Step 1: Random Measurements on the State ρ (Quantum)



Step 2: Restore Classical Snapshots (Classical)

$$\rho_i = \mathrm{ClassicalSnapshot}(U_i, s_i) = U_i|s_i\rangle\langle s_i|U_i^\dagger.$$

Step 3: Estimate Expectation Values (Classical)

$$\langle O_k\rangle_\rho \sim \sum_{i=1}^{M} \langle O_k\rangle_{\rho_i}.$$

In the case where O_k can be decomposed with polynomial number of Pauli observables, it can use *Stabilizer Formalism* to compute $\langle O_k\rangle_{\rho_i} = tr[O_k\rho_i]$ because ρ_i is a stabilizer state.

Result :

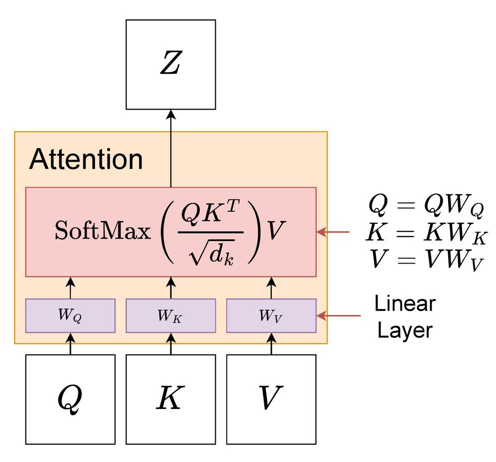
Classical Shadow achieves $M \sim O\left(\frac{\log N}{\varepsilon^2}\right)$ scaling.

Background③

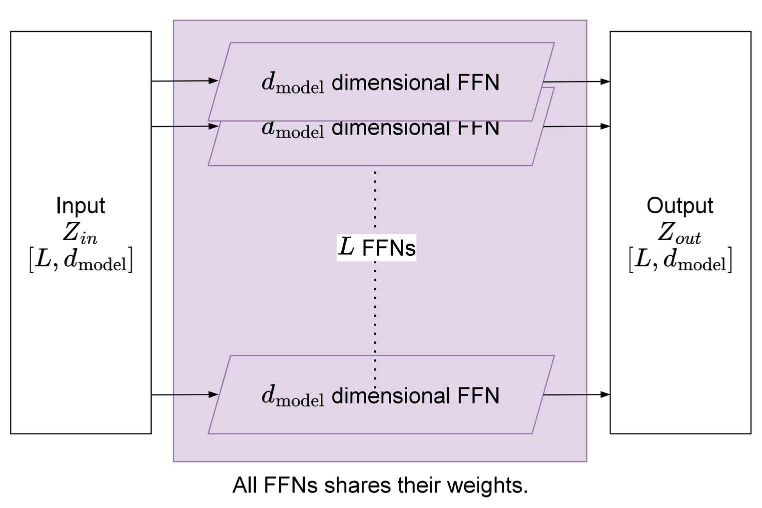
Transformer[2]

Objective : Translate sequence to sequence (Seq2Seq) with smallest loss value.

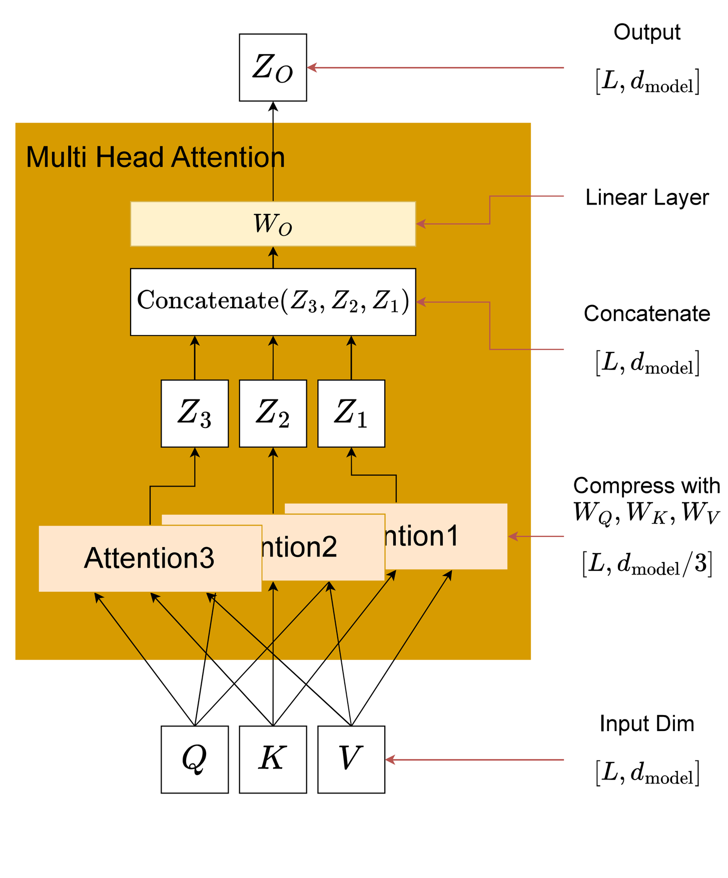
QKV Attention Mechanism



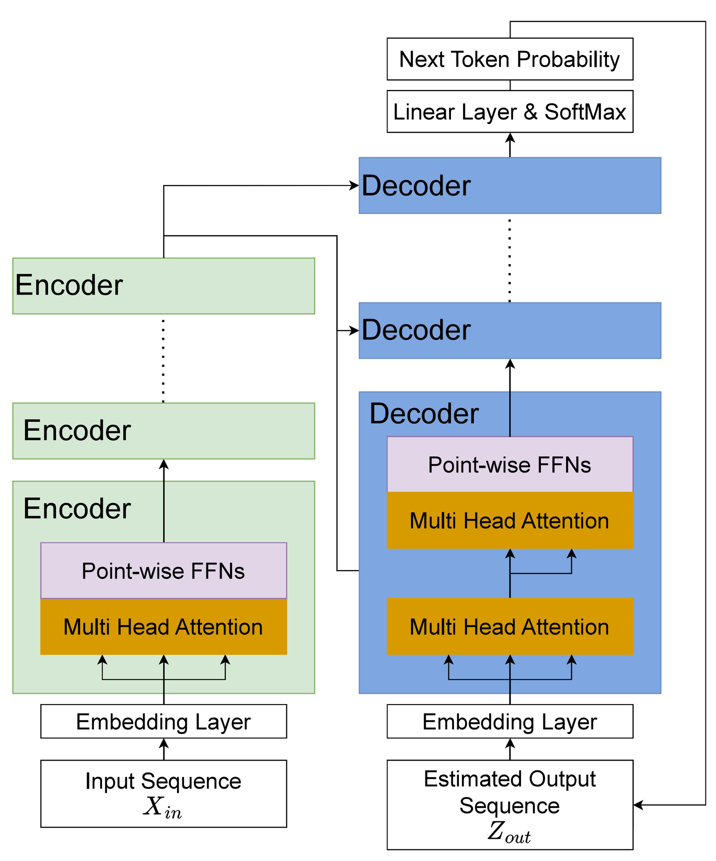
Point-wise FeedForward Network



Multi-Head Attention



Transformer

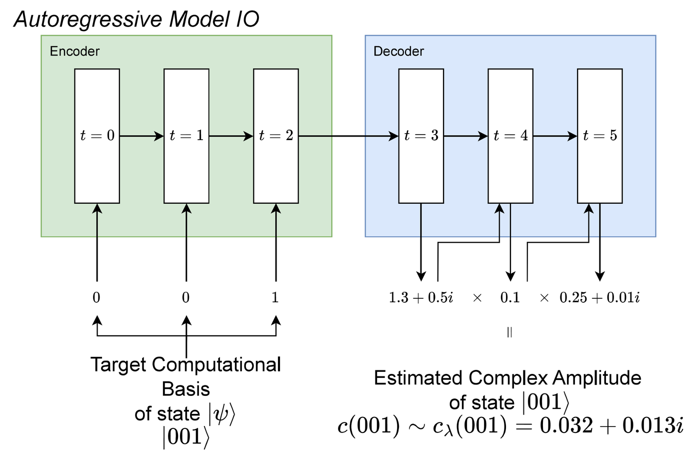


Background④

Shadow NQS[3]

Objective : Estimate complex amplitude $c(s)$ where $|\psi\rangle = \sum_s c(s)|s\rangle$. with autoregressive model and **Classical Shadow**

Autoregressive Model IO



Shadow Weight

$$p^{\mathrm{sh}}(\phi) = \frac{w(\phi)}{\sum_{\phi\in\tilde{\mathcal{D}}} w(\phi)}, \quad w(\phi) = |\langle\phi|\hat{\rho}|\phi\rangle|,$$

$\tilde{\mathcal{D}}$: set of snapshots.
$\hat{\rho}$: Shadow State
(sum of snapshot).

Loss Function

$$\mathcal{L}_{SCE}(\lambda) = -\sum_{\phi\in\tilde{\mathcal{D}}} p^{\mathrm{sh}}(\phi)\ln p_\lambda(\phi)$$

$$p_\lambda(\phi) = \|\langle\psi_\lambda, \phi\rangle\|^2$$

We can compute $p_\lambda(\phi)$ as

$$p_\lambda(\phi) = \boxed{\|\sum_s\langle\psi_\lambda, s\rangle\langle s, \phi\rangle\|^2}.$$

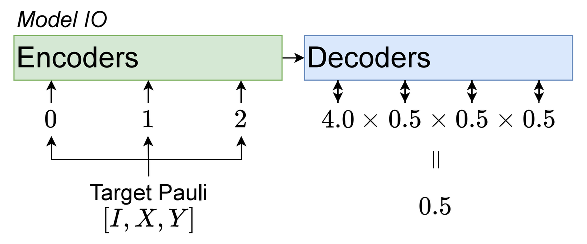
This sum has exponential 2^n terms so it uses Monte Carlo Estimation.
This makes training unstable.

Method

Pauli NQS

Objective : Estimate expectation value $\langle P\rangle_\psi$ of Pauli P with transformer model and **Classical Shadow.**

Model IO



Expectation Ratio

The output of decoders of each step must be Expectation Ratio. One wants to estimate the expectation value of $P = P_0P_1...P_N$ with Pauli NQS, the network should outputs estimators $\{NQS_i\}_{i=0}^N$ as follows:

$$NQS_i \sim \frac{\sum_{L_{i+1},...,L_N\in[I,X,Y,Z]}\langle\bigotimes_{k=1}^{i} P_k \otimes \bigotimes_{k=i+1}^{N} L_k\rangle_\psi}{\sum_{L_i,...,L_N\in[I,X,Y,Z]}\langle\bigotimes_{k=1}^{i-1} P_k \otimes \bigotimes_{k=i}^{N} L_k\rangle_\psi}.$$

Example : Expectation Ratio

In the case we have ψ and Target Pauli is XXX, Then

$$NQS_0 = \sum_{s_1,s_2,s_3\in[IXYZ]}\langle s_1\otimes s_2\otimes s_3\rangle. \quad NQS_1 = \frac{\sum_{s_2,s_3\in[IXYZ]}\langle X\otimes s_2\otimes s_3\rangle}{\sum_{s_1,s_2,s_3\in[IXYZ]}\langle s_1\otimes s_2\otimes s_3\rangle}. \quad NQS_2 = \frac{\sum_{s_3\in[IXYZ]}\langle X\otimes X\otimes s_3\rangle}{\sum_{s_2,s_3\in[IXYZ]}\langle X\otimes s_2\otimes s_3\rangle}. \quad NQS_3 = \frac{\langle X\otimes X\otimes X\rangle}{\sum_{s_3\in[IXYZ]}\langle X\otimes X\otimes s_3\rangle}.$$

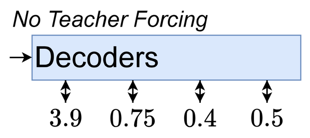
$$NQS_0 \times NQS_1 \times NQS_2 \times NQS_3 = \langle X\otimes X\otimes X\rangle.$$

Instead of modeling computational basis amplitudes, our **Pauli NQS** model directly learns the measurement probabilities of **Pauli operator strings.** Because we cannot learn more information obtained from Classical Shadow method, it is reasonable to learn NQS in the strength of Classical Shadow. Pauli expectation value learning aligns more naturally with the Classical Shadow framework, which efficiently estimates expectations of Pauli observables.

Teacher Forcing and Loss

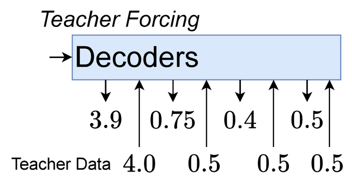
Objective : Use teacher data to improve initial learning stabiility

No Teacher Forcing



Decoders uses its output as an input.

Teacher Forcing



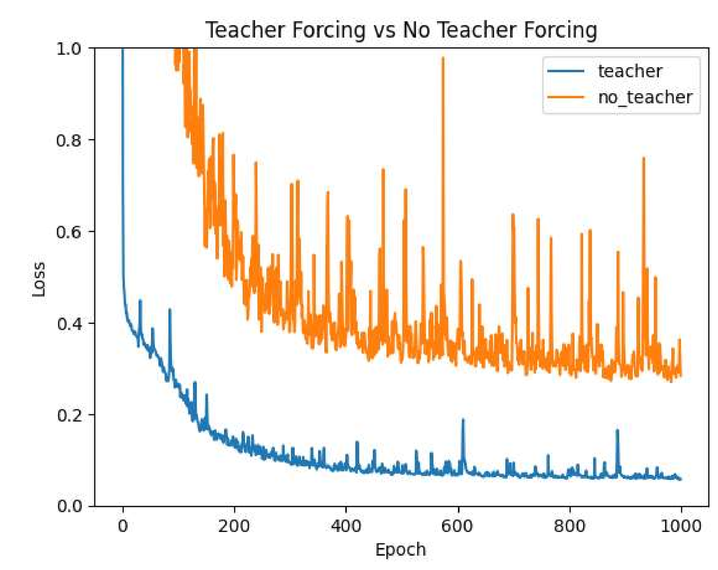
Instead of using its output, use teacher data as an input with some probability.
In this case it descide the probability based on the variance of teacher data.

Loss Function

$$\mathcal{L}_\lambda(P) = \|\prod_i NQS_i - \langle P\rangle_\psi\|^2 + \alpha\sum_i \|\langle NQS_i - \langle P_{i|0,..,i-1}\rangle\rangle\|^2$$

The first term $\langle P\rangle_\psi$ is directly obtained by Classical Shadow. Therefore, its variance is respectively small.
The second term $\langle P_{i|0,..,i-1}\rangle$ is sum of exponential number terms.
Therefore, we use Monte Carlo estimation and loss function unreliable.
To control the variance of the loss function, we add hyper-parameter α.

Results



Performance comparison for 3-qubit small problems.
Compared to without Teacher Forcing, with TF, low cost values are found earlier and learning is more stable.

	Shadow NQS	Pauli NQS
Computational Basis Complex Amplitude Estimation	Designed to estimate complex amplitude of computational basis.	Needs Monte Carlo estimation of the sum of exponential terms. (Big variance)
Pauli Expectation Estimation	Needs Monte Carlo estimation of the sum of exponential terms. (Big variance)	Designed to estimate Pauli observable expectation values

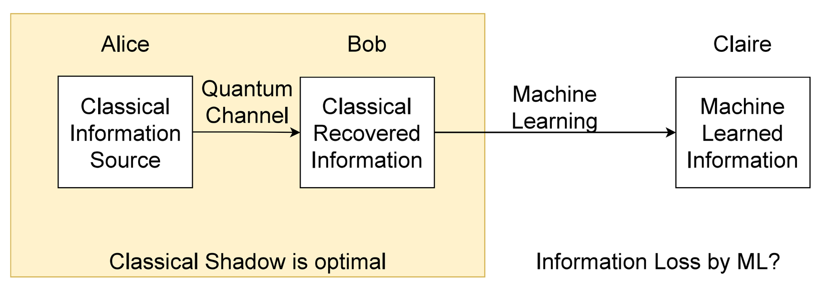
Because of the difference of the design, Shadow NQS and Pauli NQS has different strength. This makes difficult to compare fairly.
- Pauli Observable will be decomposed with exponential terms of computational basis.
- Computational Basis Projector will be decomposed with exponential terms of Paulis.

Discussion and Future work

Discussion

Theoretical limitation

Classical Shadow is optimal method w.r.t. the number of target observables N in the meaning of distinguishabiity of classical informations embedded by Alice.



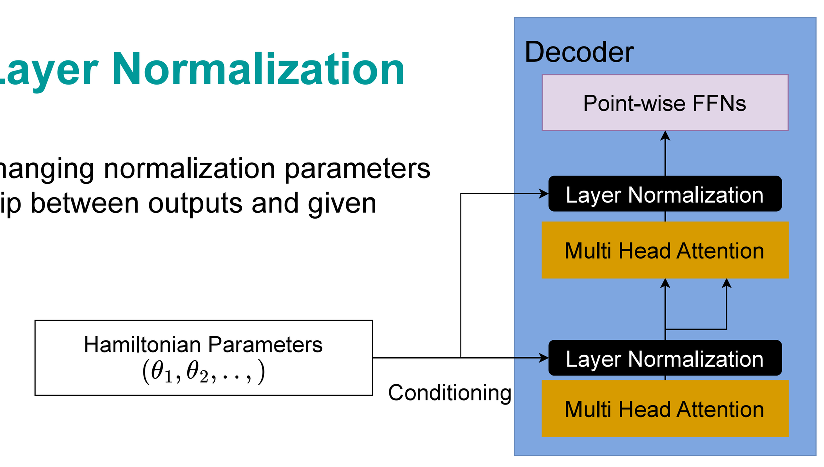
However, we are not sure how much information be lost by the process of machine learning. Can we recover optimality through machine learning without computing of Classical Shadow? What is the nessesity condition?

Future Work

Simulation Model with Conditional Layer Normalization

"Conditional Layer Normalization" can control outputs by changing normalization parameters γ and β used in Layer Normalization by learning relationship between outputs and given conditioning parameters.

We can make a Pauli NQS network a generative neural network which estimates the output state after the time evolution with Hamiltonian of given parameters.



Refferences

[0] Giuseppe Carleo, Matthias Troyer, Aug/2016・Solving the Quantum Many-Body Problem with Artificial Neural Networks.
[1] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, John Preskill・Jun/2021・Provably efficient machine learning for quantum many-body problems.
[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin・Jun/ 2017・Attention Is All You Need.
[3] Wirawat Kokaew, Bohdan Kulchytskyy, Shunji Matsuura, Pooya Ronagh・May/2024・Bootstrapping Classical Shadows for Neural Quantum State Tomography.

25' JSKeio