

# A new quantum machine learning algorithm: split hidden quantum Markov model inspired by quantum conditional master equation

Xiao-Yu Li<sup>1</sup>, Qin-Sheng Zhu<sup>†2</sup>, Yong Hu<sup>2</sup>, Hao Wu<sup>2,3</sup>, Guo-Wu Yang<sup>4</sup>, Lian-Hui Yu<sup>2</sup>, and Geng Chen<sup>4</sup>

<sup>1</sup>School of Information and Software Engineering, University of Electronic Science and Technology of China, Cheng Du, 610054, China

<sup>2</sup>School of Physics, University of Electronic Science and Technology of China, Cheng Du, 610054, China

<sup>3</sup>Institute of Electronics and Information Industry Technology of Kash, Kash, 844000, China

<sup>4</sup>School of Computer Science and Engineering, University of Electronic Science and Technology of China, Cheng Du, 610054, China

**The Hidden Quantum Markov Model (HQMM) has significant potential for analyzing time-series data and studying stochastic processes in the quantum domain as an upgrading option with potential advantages over classical Markov models. In this paper, we introduced the split HQMM (SHQMM) for implementing the hidden quantum Markov process, utilizing the conditional master equation with a fine balance condition to demonstrate the interconnections among the internal states of the quantum system. The experimental results suggest that our model outperforms previous models in terms of scope of applications and robustness. Additionally, we establish a new learning algorithm to solve parameters in HQMM by relating the quantum conditional master equation to the HQMM. Finally, our study provides clear evidence that the quantum transport system can be considered a physical representation of HQMM. The SHQMM with accompanying algorithms present a novel method to analyze quantum systems and time series grounded in physical implementation.**

## 1 Introduction

The significant increase of data and information has emphasized that classical algorithms have dif-

Qin-Sheng Zhu<sup>†</sup>: Corresponding:zhuqinsheng@uestc.edu.cn

iculties in meeting computation demands for efficiency and speed. Therefore, quantum computing, capable of efficient computation, has become a viable solution. Unlike classical computing, quantum computing employs quantum bits' superposition for data storage, reading, and Efficient computing. As a result, quantum computing promises to solve problems that are too complex for classical computers.

Quantum computing has made great improvements in hardware implementations [1, 2] and algorithmic approaches [3, 4], propelled the field of quantum computing into the "*Noisy intermediate-scale quantum (NISQ) algorithms*" era in the last decade. During that era, many hybrid framework [5] algorithms emerged that combined quantum and classical approaches to cope with suboptimal hardware conditions. Initial applications of quantum computing in chemistry [6], Hamiltonian simulation [7], biology [8], pharmaceutical [9], finance [10], materials [11] and various other fields demonstrate its potential advantages over classical computing.

In the field of machine learning, the Hidden Markov Model (HMM) algorithm plays a vitall role and has been extensively and effectively utilized in domains such as stock market forecasting [12, 13], natural language processing [14, 15], protein sequencing [16, 17].

The classical HMM [18] has three main parts: training, decoding, and learning. When the dimensionality of the hidden state is not large, the Baum-Welch algorithm [19], the Viterbi algorithm [20], and the EM algorithm [21] can be

used to solve these problems efficiently. However, when the dimensions of the hidden state and the observation space increase simultaneously, the solution speed and accuracy of the classical algorithms become weak.

Quantum algorithms were introduced in response to address such problems and find more effective solutions. Monras *et.al.* [22] gave the mathematical definition of the hidden quantum Markov process in terms of Kraus operators. This key work offers insight into HQMM and a method for studying this model from the perspective of quantum open systems, as well as creating heuristic quantum computing algorithms. Compared to the classical HMM, the quantum version also functions as a stochastic probability graph model, and it includes the same three main parts. Crucially, the advantages of HQMM are gradually becoming apparent with the study of problem solving the model parameters.

Srinivasan *et. al.*[23] proposed an algorithm based on the Norm Observable Operator Model's learning algorithm, originally presented by Jaeger *et.al.*[24]. This study demonstrates that HQMM offers advantages over traditional algorithms with regard to model complexity and accuracy. Nevertheless, this model is only suitable for situations where the hidden state dimension is relatively small, and it is inclined to succumb to local optimal solutions. Liu *et.al.* [25] analytically demonstrated the superiority of the quantum stochastic model in comparison to the classical model, providing qualitative evidence of the quantum advantage that HQMM possesses. In further work [26], they also highlight that the quantum implementation of the HMM could both mitigate thermal dissipation and achieve an advantage in memory compression. In 2020, Adhikary *et.al.* [27] proposed a learning algorithm grounded in optimization theory on the manifold [28] with goal of solving the Kraus, which provides an algorithmic basis for substantial applications of HQMM. The work of Markov *et.al.* [29] in 2022 shows that HQMM has a significant advantage in state space complexity over stochastic process languages. In 2023, Li *et.al.* [30] presents a new algorithm for modeling the dynamics of Markovian open quantum systems that is cleaner and more efficient than previous algorithms.

Taken together, the above work show that HQMM has potential advantages that need to

be explored. The primary motivation of this paper is to further explore HQMM from a physical perspective, providing new tools for a nuanced understanding of the intricate states of quantum systems while ensuring algorithmic performance.

We developed a new HQMM learning algorithm, namely SHQMM, using the quantum conditional master equation on an open quantum system described by the quantum master equation [31, 32, 33] building upon the work of Clark *et al.* [34]. As a result, SHQMM has the capability to function with open systems and non-unitary quantum algorithms, rendering it applicable to the noise-related issues that arise in quantum computers during NISQ. Another aspect, SHQMM provides an understanding of the quantum state of the system under fine balance condition while guaranteeing algorithmic performance, provides HQMM with interpretability that corresponds to quantum transport systems [35]. Additionally, our work is expected to provide new ideas for the physical implementation of quantum neural network (QNN).

The contribution of this paper is as follows:

1. A new SHQMM learning algorithm is constructed on open quantum systems by using the quantum conditional master equation, which is able to handle open systems and non-unitary quantum algorithms.
2. SHQMM with the introduction of periodic boundary conditions guarantees the performance of the algorithm under fine balance conditions, which can be physically related to quantum transport systems. Model possesses a clear physical representation, well-defined dimensional relationships, and robust interpretability.
3. Numerical experiments show that SHQMM achieves better results on both quantum and classical data sets and is robust to random initialization. Exploration of the quantum system did not degrade the performance of the model.

The paper is organized as follows: Sec. 2 provides an introduction of related work, including the basic concepts of HMM, HQMM and quantum conditional master equation. In Sec. 3, we formally introduce the SHQMM, a novel stochastic probabilistic graphical model. The implementation of SHQMM is demonstrated using a quantum transport system as an illustrative example. Sec. 4 presents the results of numerical experiments on the SHQMM model, demonstrating its

adaptability on both quantum and classical data, robust to random initialization, and performance on DA. In Sec. 5, we compare SHQMM to previous work, analyze its complexity, and discuss the benefits. Finally, in Sec. 6, we summarize our work.

## 2 Related work

### 2.1 Hidden quantum Markov model

The Hidden Markov Model (HMM) is a type of probabilistic graph model that describes the evolutionary properties of Markov dynamics. It consists of two important parameters: the transition matrix  $\mathbf{T}$  and the observation matrix  $\mathbf{C}$ , which are constant matrices. A HMM can be defined as  $\lambda = (\mathbf{T}, \mathbf{C}, x_0)$ , where  $x_0$  is the initial state vector. The update of the hidden state and the observable results can be obtained from Eq.1, which represents a state-emitting (Moore) hidden Markov model.

$$\begin{aligned} x_{t+\Delta t} &= \mathbf{T}x_t \\ y_{t+\Delta t} &= \text{diag}(\mathbf{C}_{(y,:)})x_{t+\Delta t}, \end{aligned} \quad (1)$$

where the variable  $y$  represents an output symbol in the observable space  $O$ .

A Hierarchical Quantum Markov Model (HQMM) can be defined using a set of parameters  $\lambda_Q = (\rho_0, K_y)$ , similar to the classical Markov process. Here,  $\rho_0$  corresponds to the initial state vector  $x_0$  of the classical Markov model, and the Kraus operator  $K_y$  corresponds to the matrices  $\mathbf{T}$  and  $\mathbf{C}$ . In comparison to the classical Markov model, the Kraus operator  $K_y$  plays a dual role as both the evolution state and the observable output result. It satisfies the condition  $\sum_m K_m^\dagger K_m = I$ . When the system is measured (assuming the measurement or read-out result is  $y$ ), the density matrix can be expressed as follows [27]:

$$\rho_y(t + \Delta t) = \frac{\sum_{\omega_y} K_{\omega_y} \rho(t) K_{\omega_y}^\dagger}{\text{Tr}[\sum_{\omega_y} K_{\omega_y} \rho(t) K_{\omega_y}^\dagger]} \quad (2)$$

where  $\omega_y$  denotes the auxiliary dimension of Kraus operator.

The difference between the HMM and the HQMM is shown in Table.1.

To calculate the parameters  $\{K\}$  of the HQMM, Adhikary *et.al.* [27] proposed a maximum likelihood estimation algorithm. This algo-

Table 1: The difference between the HMM and the HQMM

Model	HMM	HQMM
State	state vector $x$	density matrix $\rho$
Transition and Emission	matrix $\mathbf{T}, \mathbf{C}$	Kraus operators $\{K\}$
Steady State	$x^* = \mathbf{T}x^*$	$\rho^* = \sum_{\omega_y} K_{\omega_y} \rho^* K_{\omega_y}^\dagger$
Probability	$\text{Idiag}(\mathbf{C}_{(y,:)})x$	$\text{Tr}(\sum_{\omega_y} K_{\omega_y} \rho K_{\omega_y}^\dagger)$

rithm assumes that a set of observation sequences  $y_1, y_2, y_3, \dots, y_T$  is known, and constructs the maximum likelihood function based on this data. This is a particular case where  $\omega = 1$ :

$$\mathcal{L} = -\ln \text{tr} \left( K_{y_T} \cdots K_{y_2} K_{y_1} \rho_0 K_{y_1}^\dagger K_{y_2}^\dagger \cdots K_{y_T}^\dagger \right). \quad (3)$$

Then the parameter solving problem of the HQMM is transformed into a constrained optimization problem:

$$\begin{aligned} &\text{minimize}_{\{K\}} \mathcal{L}(\{K\}) \\ &\text{subject to} \quad \sum_y K_y^\dagger K_y = I, K_y \in \mathbb{C}^{n \times n}. \end{aligned} \quad (4)$$

Stack  $K_m$  by column to form a new matrix  $\kappa = [K_1, K_2, \dots, K_m]^T$  with dimension  $nm \times n$ , the constraint condition in Eq.4 can be rewritten as

$$\kappa^\dagger \kappa = I, \kappa \in \mathbb{C}^{nm \times n}. \quad (5)$$

According to Ref.[27],  $\kappa$  in Eq.5 lies on the Stiefel manifold, and the following gradient descent method can be used to solve Eq.4:

$$\begin{aligned} G &= \frac{\partial \mathcal{L}}{\partial \kappa}, \\ \kappa &= \kappa - \tau \mathbf{U} (I + \frac{\tau}{2} \mathbf{V}^\dagger \mathbf{U})^{-1} \mathbf{V}^\dagger \kappa. \end{aligned} \quad (6)$$

In Eq.6,  $\mathbf{U} = [G, \kappa]$ ,  $\mathbf{V} = [\kappa, -G]$ , and  $\tau$  is a positive real number.

### 2.2 The quantum conditional master equation

As a general type of equation describing the evolution of open quantum systems, the quantum conditional master equation adopts the specific form of the Lindbladian equation [36], with particular consideration given to the interaction with the environment.

For an open quantum system, the Hamiltonian form can be applied as

$$H = H_S + H_E + H'. \quad (7)$$

$H_S$  and  $H_E$  represent the Hamiltonians of the quantum system and the environment, respectively. The Hamiltonian  $H'$  describes the coupling effect between the quantum system and the environment. In the case of weak coupling between the quantum system and the environment,  $H'$  can be treated as a perturbation. Using the expansion of the second-order cumulant, we can obtain a description of the evolution of the reduced density matrix.

$$\begin{aligned} \dot{\rho}(t) = & -i\mathcal{L}\rho(t) \\ & - \int_0^t d\tau \langle \mathcal{L}'(t)\mathcal{G}(t,\tau)\mathcal{L}'(\tau)\mathcal{G}^\dagger(t,\tau) \rangle \rho(t). \end{aligned} \quad (8)$$

Here, the Liouvillian super operator is defined as  $\mathcal{L} = [H_S, (\dots)]$ ,  $\mathcal{L}' = [H'_S, (\dots)]$ .  $\mathcal{G}(t,\tau) = G(t,\tau) \times (\dots) \times G^\dagger(t,\tau)$ .  $G(t,\tau)$  is the Green's function related to  $H_S$ . The reduced density matrix is obtained by partially tracing the density matrix of the composite system, that is,  $\rho(t) = \text{Tr}_E[\rho_T(t)]$ .

In experiments, measurement results are typically linked to changes in the internal state of a system. Therefore, unlike the method used to derive Eq.8, Li et. al. [35] introduced the "detailed balance" condition to illustrate the relationship among different system states when studying the current ( the measurement result) of a quantum transport system. This allowed them to derive the quantum conditional master equation (QCME) in Equation 12 and to obtain some interesting results. By adopting the approach of the QCME described in reference [35] and considering the detailed balance among different system states, the general QCME can be expressed as following when the environment space is divided into different subspaces as shown in Figure 1.

$$\begin{aligned} \dot{\rho}^{(\mathcal{M}_q)} = & -i\mathcal{L}\rho^{(\mathcal{M}_q)} \\ & - \int_0^t d\tau \text{Tr}_{E^{(\mathcal{M}_q)}}[\mathcal{L}'(t)\mathcal{G}(t,\tau)\mathcal{L}'(\tau)\mathcal{G}^\dagger(t,\tau)\rho_T(t)]. \end{aligned} \quad (9)$$

Here, the proposed initial conditions for the quantum conditional master equation are  $\rho_T(0) \simeq \sum_{\mathcal{M}_q} \rho^{(\mathcal{M}_q)}(0) \otimes \rho_E^{(\mathcal{M}_q)}(0)$ , and  $\rho^{(\mathcal{M}_q)}$

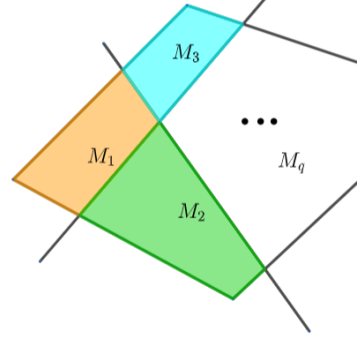


Figure 1: The diagram of dividing the environment space. The Pentagon represents the Hilbert space of the environment, which is divided into some subspace  $M_1, M_2, M_3, \dots, M_q$  corresponding to  $\rho^{(\mathcal{M}_1)}, \rho^{(\mathcal{M}_2)}, \dots, \rho^{(\mathcal{M}_q)}$ .

denotes *conditional density matrix* of the quantum system corresponding to the environment  $\rho_E^{(\mathcal{M}_q)}$  associated with the subspace  $M_q$ . Note that  $\rho^{(\mathcal{M}_q)}$  also satisfies positive semi-definite, and  $\text{Tr}(\rho^{(\mathcal{M}_q)}) \leq 1$ ,  $\text{Tr}[\sum_{\mathcal{M}_q} \rho^{(\mathcal{M}_q)}] = 1$ . The use of quantum conditional master equation enables the representation of the relationship between the different subspaces  $M_q$ , which provides a better understanding of the open quantum system being studied.

## 3 The SHQMM based on QCME

### 3.1 The quantum master equation of quantum transport system

*Since any quantum computing needs an actual physical system to implement, we need search for an open quantum system that can be described by the conditional master equation to establish HQMM. We found that the quantum transport system is suited for implementing our HQMM based on previous work [27, 35].*

As a result, in this section, we present the quantum conditional master equation for the quantum transport system. The Hamiltonian of this quantum system is expressed as follows [35]:

$$\begin{aligned} H = & H_S(a_\mu^\dagger, a_\mu) + \sum_{\alpha=L,R} \sum_{\mu k} \epsilon_{\alpha\mu k} d_{\alpha\mu k}^\dagger d_{\alpha\mu k} \\ & + \sum_{\alpha=L,R} \sum_{\mu k} (t_{\alpha\mu k} a_\mu^\dagger d_{\alpha\mu k} + \text{H.c.}). \end{aligned} \quad (10)$$

where  $H_S$  is the Hamiltonian of the quantum dots system,  $L$  and  $R$  represent the left and right

electrodes respectively,  $d_{\alpha\mu k}^\dagger$  and  $d_{\alpha\mu k}$  represent the creation and annihilation operators of electrons in the electrode, respectively, and  $t_{\alpha\mu k}$  represents the coupling strength between the electrode and the quantum dot system. The master equation for the quantum transport system can be derived through some calculations [35] based on Eq.8:

$$\dot{\rho} = -i\mathcal{L}\rho - \frac{1}{2} \sum_{\mu} \{[a_{\mu}^\dagger, A_{\mu}^{(-)}\rho - \rho A_{\mu}^{(+)}] + \text{H.c.}\}. \quad (11)$$

If the state space where the electrode is located, without any electrons passing through the quantum dot system, is denoted as  $E^{(0)}$ , it is formed by the wave function of the two isolated electrodes on the left and right  $E^{(0)} = \text{span}\{|\psi_L\rangle \otimes |\psi_R\rangle\}$ . If there are  $n$  electrons from the state space where the right electrode passes through the quantum dot to the left electrode, it is denoted as  $E^{(n)}$  ( $n = 1, 2, 3, \dots$ ). Then the electrode state space  $E$  in the equation can be decomposed as  $E = \oplus_n E^{(n)}$ , which leads to the quantum conditional master equation [35] with an initial condition of  $\rho_T(0) \simeq \sum_n \rho^{(n)}(0) \otimes \rho_E^{(n)}(0)$ .

$$\begin{aligned} \dot{\rho}^{(n)} = & -i\mathcal{L}\rho^{(n)} - \frac{1}{2} \sum_{\mu} \{[a_{\mu}^\dagger A_{\mu}^{(-)}\rho^{(n)} + \rho^{(n)} A_{\mu}^{(+)} a_{\mu}^\dagger \\ & - A_{L\mu}^{(-)}\rho^{(n)} a_{\mu}^\dagger - a_{\mu}^\dagger \rho^{(n)} A_{L\mu}^{(+)} - A_{R\mu}^{(-)}\rho^{(n-1)} a_{\mu}^\dagger \\ & - a_{\mu}^\dagger \rho^{(n+1)} A_{R\mu}^{(+)}] + \text{H.c.}\}. \end{aligned} \quad (12)$$

Here,  $\rho^{(n)} = \text{Tr}_{-E^{(n)}}[\rho_T(t)]$  is the conditional density matrix of the quantum dot system [35], which means that there are  $n$  electrons passing through the quantum dot system within time  $t$ . Here, the number of electrons  $n$  corresponds to the subspace  $\mathcal{M}_q$  in Eq. 9.

### 3.2 The relationship between the QCME equation and HQMM

Based on the contents in Sec.2.2 and 3.1, we derive the hidden quantum Markov model from a quantum master equation and propose a novel stochastic graph model from a quantum conditional master equation. After some calculations (the detailed proof and calculation process are shown in supplemental material, we obtain:

(1) For quantum master equation of Eq.8, the evolution density matrix of quantum dot system

is

$$\rho(t + \Delta t) = \sum_{i,\mu} K_{i,\mu} \rho K_{i,\mu}^\dagger. \quad (13)$$

(2) For quantum conditional master equation of Eq.12, the evolution density matrix of quantum dot system is

$$\begin{aligned} \rho^{(n)}(t + \Delta t) = & \sum_{i,\mu} K_{i,\mu} \rho^{(n)} K_{i,\mu}^\dagger \\ & + \sum_{\mu} K_{3,\mu} \rho^{(n-1)} K_{3,\mu}^\dagger \\ & + \sum_{\mu} K_{4,\mu} \rho^{(n+1)} K_{4,\mu}^\dagger, \end{aligned} \quad (14)$$

where  $i = 0, 1, 2$ .

Comparing Eq.2( $\omega_y = 1$ ) and Eq.13, we concluded that there is a close relationship between a quantum Markov model and a quantum master equation. However, the Kraus operators  $K_{i,\mu}$  of Eq.14 are involved with the related  $\rho^{(n)}$ ,  $\rho^{(n-1)}$ , and  $\rho^{(n+1)}$ . This difference arises from the division of the Hilbert space of the environment, which gives rise to a new HQMM called the split hidden quantum Markov model.

### 3.3 Split hidden quantum Markov model

In this section, we present the SHQMM inspired by quantum transport systems, which is the main theoretical improvement in this work. Similar to the HQMM, the SHQMM is defined by applying a set of parameters  $\lambda_{SQ} = (\rho^{(0)}, \rho^{(1)}, \rho^{(2)}, \dots, K_y)$  where  $\text{Tr}(\sum_{i=0} \rho^{(i)}) = 1$ .

Firstly, the evolution conditional density matrix of quantum system  $H_S$  is written as

$$\begin{aligned} \rho^{(\mathcal{M}_q)}(t + \Delta t) = & \sum_y K_y^{(\mathcal{M}_1)} \rho^{(\mathcal{M}_1)}(t) K_y^{(\mathcal{M}_1)\dagger} + \dots \\ & + \sum_y K_y^{(\mathcal{M}_q)} \rho^{(\mathcal{M}_q)}(t) K_y^{(\mathcal{M}_q)\dagger} + \dots, \end{aligned} \quad (15)$$

where,  $q$  denotes the values of subspace for environment  $\mathbf{M}$  and  $\sum_{i,\mu} K_{i,\mu}^\dagger K_{i,\mu} = I$ . The parameter  $y$  represents the read-out of information symbols from the open quantum system. Eq.15 represents a comprehensive expression that correlates to the relationship between  $\rho^{(\mathcal{M}_q)}$ .

Secondly, when we read out or measure a certain value  $y'$  for  $\rho^{(\mathcal{M}_q)}(t)$ , the conditional density matrix  $\rho^{(\mathcal{M}_q)}(t + \Delta t)$  is rewritten as follows:



$$\begin{aligned}\rho_{y'}^{(\mathcal{M}_q)}(t + \Delta t) &= \frac{\rho_{y'}^{(\mathcal{M}_q)}(t + \Delta t)}{\text{Tr}[\sum_{\mathcal{M}_q} \rho_{y'}^{(\mathcal{M}_q)}(t + \Delta t)]}, \\ \rho_{y'}^{(\mathcal{M}_q)}(t + \Delta t) &= K_{y'}^{(\mathcal{M}_1)} \rho^{(\mathcal{M}_1)}(t) K_{y'}^{(\mathcal{M}_1)\dagger} + \dots \\ &\quad + K_{y'}^{(\mathcal{M}_q)} \rho^{(\mathcal{M}_q)}(t) K_{y'}^{(\mathcal{M}_q)\dagger} + \dots.\end{aligned}\quad (16)$$

Thirdly, the probability of obtaining the measurement result  $y'$  is given by:

$$P(y') = \sum_{\mathcal{M}_q} \text{Tr}[\rho_{y'}^{(\mathcal{M}_q)}(t + \Delta t)]. \quad (17)$$

Here, Eq.17 describes the contribution of different  $\rho^{(\mathcal{M}_q)}$  to the probability  $P(y')$ , and this process reveals the concept of "detailed balance" in physics, as described in Eq.15.

**The concretely implement example of our SHQMM** To calculate the parameters of the SHQMM, assuming that a set of sequences

$$\begin{aligned}\rho_T^{(0)} &= K_{y_{T-1}} \rho_{T-1}^{(0)} K_{y_{T-1}}^\dagger + A_{y_{T-1}} \rho_{T-1}^{(1)} A_{y_{T-1}}^\dagger, \\ \rho_T^{(1)} &= R_{y_{T-1}} \rho_{T-1}^{(0)} R_{y_{T-1}}^\dagger + K_{y_{T-1}} \rho_{T-1}^{(1)} K_{y_{T-1}} + A_{y_{T-1}} \rho_{T-1}^{(2)} A_{y_{T-1}}^\dagger, \\ &\dots \\ \rho_T^{(n)} &= R_{y_{T-1}} \rho_{T-1}^{(n-1)} R_{y_{T-1}}^\dagger + K_{y_{T-1}} \rho_{T-1}^{(n)} K_{y_{T-1}} + A_{y_{T-1}} \rho_{T-1}^{(n+1)} A_{y_{T-1}}^\dagger, \\ &\dots \\ \rho_T^{(N_{max})} &= R_{y_{T-1}} \rho_{T-1}^{(N_{max}-1)} R_{y_{T-1}}^\dagger + K_{y_{T-1}} \rho_{T-1}^{(N_{max})} K_{y_{T-1}}.\end{aligned}\quad (18)$$

where  $K_{i,\mu}$ ,  $K_{3,\mu}$  and  $K_{4,\mu}$  of Eq.14 denote  $K_{y_i}$ ,  $R_{y_i}$  and  $A_{y_i}$ , respectively.  $N_{max}$  denotes the maximum value of  $n$ . The probability of  $y_i$  is  $P(y_i) = \sum_n \text{Tr}[K_{y_i} \rho^{(n)}(t) K_{y_i}^\dagger + R_{y_i} \rho^{(n-1)}(t) R_{y_i}^\dagger + A_{y_i} \rho^{(n+1)}(t) A_{y_i}^\dagger]$ .

From Eq.18, the probability of sequences  $y_0, y_1 \dots, y_T$  can be easily obtained

$$\begin{aligned}P_{y_0, y_1 \dots, y_T} &= \text{Tr}(\rho_T) \\ \rho_T &= \rho_T^{(0)} + \rho_T^{(1)} + \dots + \rho_T^{(\dots)}.\end{aligned}\quad (19)$$

To compute the parameters of the SHQMM, we propose a maximum likelihood estimation method, based on the results in [23, 27]. Firstly, we use the probability function to derive all possible Kraus operators, and then use the gradient

$y_0, y_1 \dots, y_T$  are known, the conditional density matrix evolution under the measurement result  $y_i$  is shown in Fig.2 based on transport system. It can be seen that Fig.2 is similar to a neural network and shows the process of forward propagation through time  $t$ . This suggests that SHQMM is promising as a physical realization pathway for QNN, which will be our further research. This demonstrates the connection of quantum state evolution among the different subspaces  $n$  in QCME and the conversion relationship among the probabilities  $\text{Tr}(\rho^{(n)})$ . Compared to previous work on HQMM (the probabilities for the measurement value  $y_i$  depend on  $\rho = \sum_n \rho^{(n)}$ ), the property illustrated in Fig.2 also displays the contribution variance of different  $\rho^{(n)}$  to the probabilities of obtaining the measurement value  $y_i$  at time  $t_i$ . Therefore, our model produces a more stable and robust model structure (as seen in experimental results).

Here, based on the QCME (Eq.14), we can write a probability function using the following equations.

descent algorithm to find the matrix form of the Kraus operator that satisfies the minimum probability function of the given sequence. This turns parameter-solving into an optimization problem.

$$\begin{aligned}\text{minimize}_{\{K, R, A\}} \quad & \mathcal{L}(\{K, R, A\}) = -\ln \text{Tr}(\rho_T) \\ \text{subject to} \quad & \sum_{y \in O} K_y^\dagger K_y + R_y^\dagger R_y + A_y^\dagger A_y = I.\end{aligned}\quad (20)$$

Stack  $K_y$ ,  $R_y$ ,  $A_y$  by column to form a new matrix  $\kappa = [K_1, K_2, \dots, R_1, R_2, \dots, A_1, A_2, \dots]$  with dimension  $3\text{dim}O \cdot m \times m$  ( $m$  is the dimension of the Kraus operator). So the constraint condition in Eq.20 can be rewritten as

$$\kappa^\dagger \kappa = I, \kappa \in \mathbb{C}^{3\text{dim}O \cdot m \times m}. \quad (21)$$

We summarize all the above steps into an algorithm for solving the Kraus operator. The specific steps are shown in the Algorithm1.

---

**Algorithm 1** Leaning SHQMM using gradient descent method on Stiefel manifold

---

**Require:** Training data  $D \in \mathbb{N}^{M \times l}$ ,  $W$  is number of sequences and  $l$  is the length of sequence.

**Ensure:**  $\{\mathbf{K}_i\}_{i=1}^{\dim O}$ ,  $\{\mathbf{R}_i\}_{i=1}^{\dim O}$ ,  $\{\mathbf{A}_i\}_{i=1}^{\dim O} \dots$

- 1: **Initialize:** Complex orthogonal matrix on Stiefel  $\kappa \in \mathbb{C}^{3\dim O \cdot m \times m}$  and  $\rho^{(0)}, \rho^{(1)}, \rho^{(2)}, \dots, \rho^{(N-1)}$  and require that  $\rho^{(i)}$  is positive semi-definite and  $\sum_{i=0}^{N-1} \rho^{(i)} = \rho_{total}$ ,  $\rho_{total}$  is density matrix.
  - 2: **for**  $epoch = 1 : E$  **do**
  - 3:   split the data  $D$  into  $B$  batches  $D_B$
  - 4:   **for**  $batch = 1 : B$  **do**
  - 5:     Compute gradient  $G_i^{\{K\}} = \frac{\partial \mathcal{L}}{\partial K_i^*}$ ,  $G_i^{\{R\}} = \frac{\partial \mathcal{L}}{\partial R_i^*}$ ,  $G_i^{\{A\}} = \frac{\partial \mathcal{L}}{\partial A_i^*}$
  - 6:     Compute the like-hood function  $\mathcal{L}$
  - 7:     Stack  $G_i^{\{K\}}$ ,  $G_i^{\{R\}}$ ,  $G_i^{\{A\}}$  vertically to construct  $G = [G_1^K, \dots, G_O^A]^T$
  - 8:     Construct  $\mathbf{U} = [G|\kappa]$ ,  $\mathbf{V} = [\kappa| -G]$
  - 9:      $G = \beta G_{old} + (1 - \beta)G$
  - 10:     Update  $\kappa = \kappa - \tau \mathbf{U} (I + \frac{\tau}{2} \mathbf{V}^\dagger \mathbf{U})^{-1} \mathbf{V}^\dagger \kappa$
  - 11:   **end for**
  - 12:   Update learning rate  $\tau = \alpha \tau$
  - 13: **end for**
  - 14: Compute the DA function by using the value of probability function  $\mathcal{L}$
  - 15: **return**  $\{\mathbf{K}_i\}_{i=1}^{\dim O}$ ,  $\{\mathbf{R}_i\}_{i=1}^{\dim O}$ ,  $\{\mathbf{A}_i\}_{i=1}^{\dim O}$  and DA
- 

In Algorithm1,  $\tau$ (learning rate),  $\alpha$ (decay factor) and  $\beta$ (momentum parameter) are the hyper-parameters, and DA is a function that describes the quality of the model.

As one of the core evaluation metrics of HQMM, DA [37] is able to measure the performance among models with different structures and functions in a relatively fair way [38], defined as:

$$DA = f\left(1 + \frac{\log_\iota P(D|M)}{l}\right). \quad (22)$$

Where,  $D$  is data,  $M$  is model.  $l$  is the length of the sequence, and  $\iota$  is the number of output symbol in the sequence. The function  $f(\cdot)$  is a non-linear segmented function that can map any argument in  $(-\infty, 1]$  to  $(-1, 1]$ , defined as:

$$f(x) = \begin{cases} x, & x \geq 0, \\ \frac{1 - e^{-0.25x}}{1 + e^{-0.25x}}, & x < 0. \end{cases} \quad (23)$$

The model can perfectly predict the Markov sequence if  $DA = 1$  and the model better than random model for  $DA > 0$ .

In the SHQMM, different models can produce different prediction effects for the same sequences, depending on the number  $\mathcal{M}q$  of initialized conditional density matrices and the connections between them. Eq.18 describes the closest connection between the conditional density matrix and the number  $\mathcal{M}q$ , which is equal to  $n$ .

The optimal solution of HQMM is one of the optimal solutions of Eq.20. Therefore, periodic boundary conditions are applied to the first and last conditional density matrices, similar to the arrangement of atoms in a crystal.

$$\begin{aligned} \rho_T^{(0)} &= K_{y_{T-1}} \rho_{T-1}^{(0)} K_{y_{T-1}}^\dagger + A_{y_{T-1}} \rho_{T-1}^{(1)} A_{y_{T-1}}^\dagger \\ &\quad + R_{y_{T-1}} \rho_{T-1}^{(N_{max})} R_{y_{T-1}}^\dagger, \\ \rho_T^{(N_{max})} &= R_{y_{T-1}} \rho_{T-1}^{(N_{max}-1)} R_{y_{T-1}}^\dagger + K_{y_{T-1}} \rho_{T-1}^{(N_{max})} K_{y_{T-1}}^\dagger \\ &\quad + A_{y_{T-1}} \rho_{T-1}^{(0)} A_{y_{T-1}}^\dagger. \end{aligned} \quad (24)$$

Different from Eq.18, Eq.24 includes periodic boundary conditions. This empirical improvement is due to the fact that the Kraus operators ( $R, A$ ) often converge to zero during the optimization process, which means that the model has essentially degenerated into a HQMM. This improvement ensures the stability of the learning process.

More detailed cases are presented in supplemental material.

### Extend the expandability ability of the SHQMM

If we need to further improve the complexity of the model, we can set the parameters  $N_{max} = 4, 5, 6, \dots$  and apply a more complicated connection defined as  $k$ -local. Thus, the general SHQMM can be defined as a tuple  $\lambda_{SQ} = (\mathbb{C}^m, k\text{-local}, K_y^j, y \in O, j = 2k + 1, \rho_0^{(i)}, i \in [0, N_{max} - 1])$  with the following conditions:

- (1)  $\rho_0^{(i)}$  is conditional density matrix,  $\text{Tr}(\sum_{i=0}^{N_{max}-1} \rho_0^{(i)}) = 1$ .

(2) For every Kraus operator,  $K_y^j : \mathbb{C}^m \rightarrow \mathbb{C}^m$  and  $\sum_{y,j} (K_y^j)^\dagger K_y^j = I$ .

(3) The evolution of  $\rho^{(i)}$  follows Eq.25.

$$\rho^{(i)}(t + \Delta t) = \sum_{j' = -\lfloor \frac{j}{2} \rfloor}^{\lfloor \frac{j}{2} \rfloor} \sum_{y \in O} K_y^{j'+1+\lfloor \frac{j}{2} \rfloor} \rho^{(i-j')}(t) K_y^{j'+1+\lfloor \frac{j}{2} \rfloor \dagger}, \quad (25)$$

where the periodic boundary conditions should be applied and conditional density matrix beyond index range should be zeroed, that is,  $0 \leq i - j' \leq N_{max} - 1$ .  $\lfloor \frac{j}{2} \rfloor$  equal to  $k$ ,  $j' \in \{-\lfloor \frac{j}{2} \rfloor, -\lfloor \frac{j}{2} \rfloor + 1, \dots, \lfloor \frac{j}{2} \rfloor - 1, \lfloor \frac{j}{2} \rfloor\}$ .

(4) The probability of observation symbol  $y$  is

$$p(y) = \sum_{i=0}^{N-1} \text{Tr} \left( \sum_{j' = -\lfloor \frac{j}{2} \rfloor}^{\lfloor \frac{j}{2} \rfloor} K_y^{j'+1+\lfloor \frac{j}{2} \rfloor} \rho^{(i-j')}(t) K_y^{j'+1+\lfloor \frac{j}{2} \rfloor \dagger} \right), \quad (26)$$

where,  $k$ -local represents the relationship between different conditional density matrix and  $j$  represents the number of Kraus operator classes.

In summary, enhancing state correlations and moving beyond simple neighbor connections enhances SHQMM's applicability to diverse time series problems.

**Comparison of properties of SHQMM and HQMM** We use a simple case to illustrate the correlation and difference between SHQMM and HQMM. For 1-local model, by summing the conditional density matrix  $\rho^{(n)}$  with index  $n$  in Eq.18, we obtain:

$$\rho_{y|x}^{(n)}(t + \Delta t) = \frac{\sum_{j' = \lfloor \frac{j}{2} \rfloor}^{\lfloor \frac{j}{2} \rfloor} K_y^{j'+1+\lfloor \frac{j}{2} \rfloor} \rho^{(n-j')}(t) K_y^{j'+1+\lfloor \frac{j}{2} \rfloor \dagger}}{\text{Tr} \left( \sum_{n=0}^{N-1} \sum_{j' = \lfloor \frac{j}{2} \rfloor}^{\lfloor \frac{j}{2} \rfloor} K_y^{j'+1+\lfloor \frac{j}{2} \rfloor} \rho^{(n-j')}(t) K_y^{j'+1+\lfloor \frac{j}{2} \rfloor \dagger} \right)}. \quad (28)$$

Eq.28 expresses, under the principle of conditional probability, the quantum state corresponding to the system when the system is observed as  $x$  at  $t$  and as  $y$  at  $t + \Delta t$ . Table 2 shows the properties of the SHQMM and HQMM.

$$\begin{aligned} \rho(t + \Delta t) &= \sum_y K_y \rho(t) K_y^\dagger + \sum_y R_y \rho(t) R_y^\dagger + \sum_y A_y \rho(t) A_y^\dagger \\ &= \sum_{\omega_y} K_{\omega_y} \rho(t) K_{\omega_y}^\dagger. \end{aligned} \quad (27)$$

The second equal sign in Eq.27 shows a formal relationship between SHQMM and HQMM, but our model has clear physical implications compared to the auxiliary dimension  $\omega_y$  in Eq.2 (HQMM with  $\omega_y = 3$ ). This indicates that the SHQMM is a valid HQMM at the same time.

The differences between SHQMM and HQMM lie in the following aspects:

- SHQMM makes density matrix have a hierarchy structure as shown in Fig.2, and the density matrix evolves through multiple channels.
- Kraus operator  $\{K_y^j\}_{y \in O}^{j=2k+1}$  acts on different conditional density matrix  $\rho^{(n)}$  in SHQMM, whereas in HQMM, Kraus operator  $\{K_{\omega_y}\}_{y \in O}$  acts on total density matrix  $\rho$ .
- SHQMM can be derived from actual physical systems, such as quantum transport systems, as shown in Fig.2.

The SHQMM can reflect the relationship between hidden states and is more suitable for handling more complex data than the HQMM. From a physical system implementation point of view, the quantum conditional master equation may differ from the quantum transport system for other open quantum systems[31, 32, 33], resulting in different split hidden quantum Markov models. The Bayesian rule for SHQMM is:

## 4 Experiment and results

In this section, we applied quantum and classical data to train and test our SHQMM. All experiments were performed on an experimental plat-



Table 2: The properties of HQMM and sHQMM

Model	State	Transition and Emission	Probability	Bayesian Rule	Evolution
SHQMM	$\{\rho^{(i)}\}_{i=0}^{N-1}$	quantum channel $\mathcal{K}_s$	Eq.26	$\rho_{y x}^{(n)}$	Eq.25
HQMM	$\rho$	quantum channel $\mathcal{K}$	$\text{Tr}(\sum_{\omega_y} K_{\omega_y} \rho(t) K_{\omega_y}^\dagger)$	Eq.2	$\sum_{\omega_y} K_{\omega_y} \rho(t) K_{\omega_y}^\dagger$

form outfitted with Intel Core i7 CPU and 16 GB RAM, trials were executed on Python 3.8 with PyTorch and qiskit libraries for modeling and evaluating. The source code and further details are available at [here](#).

**Quantum data** Firstly, we used the quantum data generated by a quantum mechanical process in Ref.[23]. The quantum data has six hidden states and six observational values. The size of the quantum data is  $40 \times 3000$ .

**Training and validation** We use  $20 \times 3000$  data to train our model, generating a total of 20 models. Simultaneously, we used  $10 \times 3000$  data to verify the model, and the remaining data were used to test the model. The results are shown in Fig.3 with hyperparameters  $\tau = 0.95, \alpha = 0.95, \beta = 0.90$  (More calculation results can be found in supplemental material). Fig.3 shows several models  $\lambda_{SQ}$  for single, double and three qubits quantum system which are used to construct the SHQMM under the different parameters  $N$  and  $k$ . It found : (1) With the increase of the qubit number, the value of DA also increases and reaches a stable value at about 20 epochs (three qubits exceed 20 epochs). (2) Apart from the single bit, the connection mode (different k-local) and  $N$  have little effect on the DA value. (3) While a higher number of quantum bits results in a relatively larger value of DA, the standard deviation (STD) decreases as the number of qubits increases during evaluation with alternative data. This suggesting potential overfitting.

To further test the reliability of our model, we will evaluate it from several perspectives.

**Initialize Kraus** In Ref. [27], it was stated that the training outcome of HQMM is susceptible to the initial Kraus operators in smaller models. Thus, this study investigates the effect of the

initial position of Kraus operators on Stiefel manifolds for SHQMM. Eq. 29 is utilized to evaluate the distance between various initial positions.

$$D(\kappa_1, \kappa_2) = \|\kappa_1 \kappa_2^\dagger - I\|_2. \quad (29)$$

When  $\kappa_1 = \kappa_2$ ,  $D = 0$ . The initialization method for the Kraus operator is presented in Algorithm 2. The varying behaviors of the DA are depicted in Figure 4 for different random initialization seeds (RS). Results clearly shows that our proposed SHQMM remains stable regardless of the initial Kraus values, thus confirming its validity.

---

**Algorithm 2** Initial Kraus operator on Stiefel manifolds

---

**Require:** the dimension of Kraus operator  $m$ , the class of Kraus operator  $j$ , the dimension of observable space  $\text{dim}O$  and random seed RS

**Ensure:** Kraus operator  $\{K_y^j\}$

- 1: **Initialize:** random vector  $\vec{v}_1 = \text{random}(m \cdot j \cdot \text{dim}O, 1, \text{RS})$ , zero matrix  $\kappa = \text{zeros}(m \cdot j \cdot \text{dim}O, 2m)$ ,  $\kappa(:, 1) = \frac{\vec{v}_1}{\|\vec{v}_1\|}$
  - 2: **for**  $s = 2 : 2m$  **do**
  - 3:    $\vec{v}_s = \text{random}(m \cdot j \cdot \text{dim}O, 1)$
  - 4:   **if** all column vectors in  $\kappa$  and  $\vec{v}_s$  are orthogonal **do**
  - 5:      $\vec{v}_s = \frac{\vec{v}_s}{\|\vec{v}_s\|}$ ;  $\kappa(:, s) = \vec{v}_s$
  - 6:   **else do**
  - 7:     Schmidt Orthogonalization of  $\kappa$  and  $\vec{v}_s$
  - 8:      $\vec{v}_s = \frac{\vec{v}_s}{\|\vec{v}_s\|}$ ;  $\kappa(:, s) = \vec{v}_s$
  - 9:   **end if**
  - 10: **end for**
  - 11: construct new matrix  $\kappa' = \kappa(:, 1 : m)/\sqrt{2} + \kappa(:, m + 1 : 2m)/\sqrt{2}$
  - 12: split  $\kappa'$  into Kraus operator
- 

**Selection of effective models** Given the various methods available to construct models, selecting the most effective one for a given dataset

is a critical challenge. Typically, the expressiveness of a model is directly linked to the number of its parameters. The number of parameters for SHQMM is as follows:

$$\mathcal{N}_P = m^2 \cdot j. \quad (30)$$

The corresponding results are shown in Fig.5 to obtain best training results, we should change the dimension  $m$  of Kraus operator firstly and then adjust the parameter  $j$  for a given sequence.

### Hyperparameters selection for the model

To obtain the optimal  $DA$ , Algorithm 1 employs three hyperparameters, namely  $\tau$ ,  $\alpha$ , and  $\beta$ . Fig. 6 demonstrates the impact of varying hyperparameters on  $DA$ , indicating that  $DA$  is more sensitive to changes in  $\alpha$  as compared to  $\tau$ . Moreover, the existence of multiple local optima in SHQMM is evident. To identify the global optimum, we investigated the effect of momentum parameter  $\beta$  on  $DA$  for the best case ( $\tau = 0.95, \alpha = 0.95$ ) and the worst case ( $\tau = 0.65, \alpha = 0.65$ ), as presented in Fig.7. It was observed that  $DA$  may reach the global optimum at  $\tau = 0.95, \alpha = 0.95$ , and  $\beta = 0.90$ , and that  $DA$  can be further enhanced by selecting a different  $\beta$ . However, after computing the distance of Kraus solutions between different hyperparameters ( $\tau = 0.95, \alpha = 0.95, \beta = 0.90$  and  $\tau = 0.65, \alpha = 0.65, \beta = 0.60$ ) using Eq. 29, we discovered that their  $DA$  values were comparable despite locating at different positions on the Stiefel manifold.

**Classical data** To conduct a thorough evaluation of the model, classical data generated by a hidden Markov process with transition matrix  $\mathbf{T}$  and emission matrix  $\mathbf{C}$  were utilized to compute the Kraus operator and determine  $DA$ . The results obtained from the classical data are presented in Fig. 8, where hyperparameters were set to  $\tau = 0.95, \alpha = 0.95$ , and  $\beta = 0.90$ .

Similar to the quantum case, the value of  $DA$  also increases and reaches a stable state after approximately 20 epochs (for three qubits, it took 20 epochs to stabilize). The different values of  $k$ -local have little impact on the  $DA$  value, and the standard deviation (STD) continues to decrease as the qubit number increases for the testing data, possibly due to model overfitting. Additional test results can be found in supplemental material.

$$\mathbf{T} = \begin{pmatrix} 0.8 & 0.01 & 0 & 0.1 & 0.3 & 0 \\ 0.02 & 0.02 & 0.1 & 0.15 & 0.05 & 0 \\ 0.08 & 0.03 & 0.1 & 0.4 & 0.05 & 0.5 \\ 0.05 & 0.04 & 0.5 & 0.35 & 0 & 0.5 \\ 0.03 & 0.5 & 0.03 & 0 & 0.6 & 0 \\ 0.02 & 0.4 & 0.27 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 0.2 & 0 & 0.05 & 0.95 & 0.01 & 0.05 \\ 0.7 & 0.1 & 0.05 & 0.01 & 0.05 & 0.05 \\ 0.05 & 0.8 & 0.1 & 0.02 & 0.05 & 0.04 \\ 0.04 & 0.04 & 0.02 & 0 & 0.84 & 0.11 \\ 0.01 & 0.03 & 0.7 & 0.01 & 0.02 & 0.2 \\ 0 & 0.03 & 0.08 & 0.01 & 0.03 & 0.55 \end{pmatrix}.$$

**Result** The numerical experimental results of SHQMM on quantum and classical data are displayed in supplemental material. The current results are likely to be overfitted given the complexity of the relationship between complex model and simple data with lower number and dimensionality of hidden states [26]. Keeping the hyperparameters constant, the training step converges more and more slowly as the number of qubits increases.

SHQMM exhibits good robustness in numerical experiments and maintains stable performance in multiple task scenarios. For different sequences, the model maintains  $\text{STD}(DA) < 0.01$ . Different initial values of the Kraus operator have a negligible effect on the convergence of the model, due to the hierarchical structure of the conditional density matrix.

## 5 Analysis and discussion

In this section, the performance of SHQMM is analyzed from both experimental and theoretical perspectives.

**Performance comparisons** It is imperative to acknowledge the existence of structural differences between different models. Remarkably, the model presented here represents a groundbreaking development in our recognized field, as it is uniquely capable of articulating a meticulous representation of the internal state of a quantum system. Consequently, the task of identifying models that are precisely aligned to the same baseline for performance evaluation becomes a formidable

challenge. To address this, we have wisely chosen to compare models that exhibit maximum congruence in both structure and functionality. Table. 3 compares the DA of quantum data obtained by the SHQMM, HMM and HQMM in [23] as follows:

Table 3: Comparison of HMM, HQMM and SHQMM on DA. The presentation of complete data under different parameters is shown in [23] and supplemental material.

Model	DA
2, 6–HMM( $L$ )	0.0327
2, 6, 1–HQMM ( $L$ )	0.1303
3,AR-SHQMM	$\geq 0.1700$

2, 2-HMM( $L$ ) stands for the classical Hidden Markov model with two hidden states and two observables, and  $L$  stands for the fact that the value is obtained by learning. Similarly, 2, 2, 1-HQMM( $L$ ) denotes the quantum hidden Markov model having two hidden states, two observables, with one Kraus operator per observable. For SHQMM, 3,AR means that the number of hidden states is 3, and AR means that the current model has 3 hidden states and connectivity is adjacent.  $\geq$  indicates that the value did not converge at the number of iterations set.

As can be seen from Table. 3, SHQMM improves DA by about 23.35% over the HQMM algorithm, while the quantum versions of the algorithm all have better performance in model quality over the classical algorithm.

**Complexity analysis** Complexity serves as a theoretical measure encompassing both the computational and time costs associated with an algorithm. To ensure consistency between various models and their original literature in terms of description, we have retained their respective symbol systems in this chapter. The complexity of HMM, HQMM and SHQMM is analyzed as follows:

- Building upon previous research [23], a classical HMM algorithm can be succinctly characterized as  $(n, s)$ , where  $n$  denotes the number of hidden states and  $s$  represents the number of observations. In the context of an HMM with a prediction sequence of length  $T$ , the complexity arises as each state transition involves  $n$  states undergoing transi-

tions, with each state having  $n$  possible transitions, resulting in a complexity denoted as  $O(Tn^2)$ [39].

- Similarly, the HQMM is denoted by the ternary  $(n, s, w)$ . Introducing the parameter  $w$ , alongside  $n$  and  $s$ , accounts for the number of Kraus operators per observable. In a manner akin to HMM, where a HQMM prediction sequence of length  $T$  is considered, there exist  $n$  states undergoing transitions, with each state producing  $w$  observations. Consequently, the complexity of HQMM is denoted as  $O(Tnw)$ . In scenarios of comparable scale, the complexity of HQMM is notably influenced by the parameter  $w$ , which can be independently set based on the application scenario. Generally, in smaller-scale tasks,  $n < w$  may occur, while in longer sequence tasks,  $n > w$  is more likely.
- The model structure of SHQMM is denoted as  $(N_{max}, s, k)$ , where  $N_{max}$  represents the number of hidden states in the model, equivalent to  $n$  in other models.  $k$  represents the connectivity state which is related to the number of Kraus operators. Due to structural constraints,  $2k + 1 \leq N_{max}$ . Similarly to HMM, SHQMM exhibits parallels in complexity analysis. During the transition process among  $N_{max}$  states, each state yields  $2k + 1$  observational outputs. Therefore, the complexity is expressed as  $O(TN_{max}(2k + 1))$ . Notably, due to the constraints of the connection method, imposing an upper limit on the complexity of SHQMM, expressed as  $O(TN_{max}(2k + 1)) \leq O(TN_{max}^2)$ .

**Discussion of Quantum Advantage** Exploring quantum advantages is a shared pursuit among all quantum machine learning algorithms.

In prior research, a substantial body of work has focused on examining the benefits of quantum versions of HMM. Noteworthy contributions, exemplified by [22] and [23], have highlighted the advantages observed in numerical experiments of Quantum HMM (HQMM), particularly in the DA. Additionally, certain studies have undertaken theoretical analyses to elucidate these quantum advantages. Notably, [25] and [26] provide insights into the advantages of HQMM, emphasizing its physical performance and memory

efficiency. Furthermore, [29] illustrates the quantum approach’s superiority in terms of state space complexity. Experiments in this work align with the aforementioned research, further substantiating the merits associated with quantum-based approaches.

According to the analysis of complexity, it is evident that HQMM and SHQMM demonstrate advantages when  $n$  is larger. This implies that quantum solutions are more suitable for scenarios involving long sequence tasks. It should be acknowledged that in numerical experiments on classical computers, the results did not fully reflect this advantage. This is attributed to the additional computational cost incurred by the classical computer’s quantum simulation process.

Nevertheless, the experimental results of our work, [23] and [27] still showcase the potential performance advantages of the quantum approach, namely the ability to convey more information with fewer hidden states. This high expressivity has been mathematically proven in [29], representing a potential advantage of quantum approaches. In terms of encoding efficiency, a substantial amount of information can be encoded onto a smaller number of quantum states, thereby enhancing the utilization of informational resources. This implies that SHQMM can be applied to large-scale tasks such as weather forecasting.

Furthermore, the superiority of SHQMM is underscored by its capacity to correspond to a tangible physical system, specifically the quantum transport system. This attribute imparts a lucid interpretability to SHQMM, enabling continuous tracking of the internal intricacies of the quantum system throughout the evolutionary process further than HQMM. Also, owing to its association with an authentic quantum physical system, SHQMM demonstrates an increased affinity for quantum data, a trait supported by the results of numerical experiments. Together with other quantum versions of HMM models, SHQMM enriches the available selection of models for the study of quantum Markov processes.

It is noteworthy, however, that SHQMM’s performance with some classical datasets appears suboptimal. Further analysis, informed by the insights of Michael *et.al.*’s work [36] thereby constituting a focal point for our forthcoming research endeavors.

## 6 Conclusion

In this paper, the novel stochastic probabilistic graphical model, SHQMM, is introduced as our main theoretical contribution using quantum conditional master equations. An empirical improvement has also been implemented by introducing periodic boundary conditions to ensure the stability of the learning process.

Numerical experiments underscore that the model’s performance on DA remains robust following the introduction of the new structure, displaying heightened insensitivity to the initial state and increased overall robustness.

SHQMM emerges as a valuable tool for elucidating the intricate relationships among hidden states within quantum systems, providing support and additional choices for addressing HMM challenges through quantum methods. The model’s correspondence with quantum transport systems further enhances its appeal, offering promising prospects for physical implementation. Additionally, given the structural similarities, SHQMM assumes a pivotal role as a theoretical foundation for the physical realization of QNN.

## 7 Acknowledgements

This work is supported by the National Key R&D Program of China, Grant No.2018FYA0306703 and Chengdu Innovation and Technology Project, No.2021-YF05-02413-GX and 2021-YF09-00114-GX, the Open Fund of Advanced Cryptography and System Security Key Laboratory of Sichuan Province (Grant No. SKLACSS-202210), Sichuan Province key research and development project, No.2022YFG0315, Natural Science Foundation of Xinjiang Uygur Autonomous Region, Granted No. 2023D01A63.

## References

- [1] Juan I Cirac and Peter Zoller. “Quantum computations with cold trapped ions”. *Physical review letters* **74**, 4091 (1995).
- [2] Emanuel Knill, Raymond Laflamme, and Gerald J Milburn. “A scheme for efficient quantum computation with linear optics”. *nature* **409**, 46–52 (2001).
- [3] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe,



- and Seth Lloyd. “Quantum machine learning”. *Nature* **549**, 195–202 (2017).
- [4] M Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J Coles. “Challenges and opportunities in quantum machine learning”. *Nature Computational Science* **2**, 567–576 (2022).
- [5] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermann Heimonen, Jakob S Kottmann, Tim Menke, et al. “Noisy intermediate-scale quantum (nisq) algorithms (2021)” (2021). [arXiv:2101.08448v1](https://arxiv.org/abs/2101.08448v1).
- [6] Alán Aspuru-Guzik, Roland Lindh, and Markus Reiher. “The matter simulation (r) evolution”. *ACS central science* **4**, 144–152 (2018).
- [7] Iulia M Georgescu, Sahel Ashhab, and Franco Nori. “Quantum simulation”. *Reviews of Modern Physics* **86**, 153 (2014).
- [8] Markus Reiher, Nathan Wiebe, Krysta M Svore, Dave Wecker, and Matthias Troyer. “Elucidating reaction mechanisms on quantum computers”. *Proceedings of the National Academy of Sciences* **114**, 7555–7560 (2017).
- [9] Yudong Cao, Jhonathan Romero, and Alán Aspuru-Guzik. “Potential of quantum computing for drug discovery”. *IBM Journal of Research and Development* **62**, 6–1 (2018).
- [10] Roman Orus, Samuel Mugel, and Enrique Lizaso. “Quantum computing for finance: Overview and prospects”. *Reviews in Physics* **4**, 100028 (2019).
- [11] Pierre-Luc Dallaire-Demers, Jonathan Romero, Libor Veis, Sukin Sim, and Alán Aspuru-Guzik. “Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer”. *Quantum Science and Technology* **4**, 045005 (2019).
- [12] Elizabeth Fons, Paula Dawson, Jeffrey Yau, Xiao-jun Zeng, and John Keane. “A novel dynamic asset allocation system using Feature Saliency Hidden Markov models for smart beta investing”. *Expert Systems with Applications* **163**, 113720 (2021).
- [13] PV Chandrika, K Visalakshmi, and K Sakthi Srinivasan. “Application of Hidden Markov Models in Stock Trading”. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). Pages 1144–1147. (2020).
- [14] Dima Suleiman, Arafat Awajan, and Wael Al Etaiwi. “The use of hidden Markov model in natural arabic language processing: A survey”. *Procedia computer science* **113**, 240–247 (2017).
- [15] Hariz Zakka Muhammad, Muhammad Nasrun, Casi Setianingsih, and Muhammad Ary Murti. “Speech recognition for English to Indonesian translator using hidden Markov model”. In 2018 International Conference on Signals and Systems (ICSigSys). Pages 255–260. IEEE (2018).
- [16] Erik LL Sonnhammer, Gunnar Von Heijne, Anders Krogh, et al. “A hidden Markov model for predicting transmembrane helices in protein sequences”. In LSMB 1998. Pages 175–182. (1998). url: <https://cdn.aaai.org/ISMB/1998/ISMB98-021.pdf>.
- [17] Gary Xie and Jeanne M Fair. “Hidden Markov Model: a shortest unique representative approach to detect the protein toxins, virulence factors and antibiotic resistance genes”. *BMC Research Notes* **14**, 1–5 (2021).
- [18] Sean R Eddy. “What is a hidden markov model?”. *Nature biotechnology* **22**, 1315–1316 (2004).
- [19] Paul M Baggenstoss. “A modified baum-welch algorithm for hidden markov models with multiple observation spaces”. *IEEE Transactions on speech and audio processing* **9**, 411–416 (2001).
- [20] Aleksandar Kavcic and Jose MF Moura. “The viterbi algorithm and markov noise memory”. *IEEE Transactions on information theory* **46**, 291–301 (2000).
- [21] Todd K Moon. “The expectation-maximization algorithm”. *IEEE Signal processing magazine* **13**, 47–60 (1996).
- [22] Alex Monras, Almut Beige, and Karoline Wiesner. “Hidden quantum Markov models and non-adaptive read-out of many-body states” (2010). [arXiv:1002.2337](https://arxiv.org/abs/1002.2337).



- [23] Siddarth Srinivasan, Geoff Gordon, and Byron Boots. “Learning hidden quantum markov models”. In Amos Storkey and Fernando Perez-Cruz, editors, Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics. Volume 84 of Proceedings of Machine Learning Research, pages 1979–1987. PMLR (2018). url: <https://proceedings.mlr.press/v84/srinivasan18a.html>.
- [24] Herbert Jaeger. “Observable operator models for discrete stochastic time series”. *Neural computation* **12**, 1371–1398 (2000).
- [25] Qing Liu, Thomas J. Elliott, Felix C. Binder, Carlo Di Franco, and Mile Gu. “Optimal stochastic modeling with unitary quantum dynamics”. *Phys. Rev. A* **99**, 062110 (2019).
- [26] Thomas J Elliott. “Memory compression and thermal efficiency of quantum implementations of nondeterministic hidden markov models”. *Physical Review A* **103**, 052615 (2021).
- [27] Sandesh Adhikary, Siddarth Srinivasan, Geoff Gordon, and Byron Boots. “Expressiveness and Learning of Hidden Quantum Markov Models”. In International Conference on Artificial Intelligence and Statistics. Pages 4151–4161. (2020). url: <http://proceedings.mlr.press/v108/adhikary20a/adhikary20a.pdf>.
- [28] Bo Jiang and Yu-Hong Dai. “A framework of constraint preserving update schemes for optimization on Stiefel manifold”. *Mathematical Programming* **153**, 535–575 (2015).
- [29] Vanio Markov, Vladimir Rastunkov, Amol Deshmukh, Daniel Fry, and Charlee Stefanski. “Implementation and learning of quantum hidden markov models” (2022). [arXiv:2212.03796v2](https://arxiv.org/abs/2212.03796v2).
- [30] Xiantao Li and Chunhao Wang. “Simulating markovian open quantum systems using higher-order series expansion” (2022). [arXiv:2212.02051v2](https://arxiv.org/abs/2212.02051v2).
- [31] Yoshitaka Tanimura. “Stochastic Liouville, Langevin, Fokker–Planck, and master equation approaches to quantum dissipative systems”. *Journal of the Physical Society of Japan* **75**, 082001 (2006).
- [32] Akihito Ishizaki and Graham R Fleming. “Unified treatment of quantum coherent and incoherent hopping dynamics in electronic energy transfer: Reduced hierarchy equation approach”. *The Journal of chemical physics* **130** (2009).
- [33] Jinshuang Jin, Xiao Zheng, and YiJing Yan. “Exact dynamics of dissipative electronic systems and quantum transport: Hierarchical equations of motion approach”. *The Journal of chemical physics* **128** (2008).
- [34] Lewis A Clark, Wei Huang, Thomas M Barlow, and Almut Beige. “Hidden quantum markov models and open quantum systems with instantaneous feedback”. In ISCS 2014 Interdisciplinary Symposium on Complex Systems. Pages 143–151. (2015).
- [35] Xin-Qi Li, JunYan Luo, Yong-Gang Yang, Ping Cui, and YiJing Yan. “Quantum master-equation approach to quantum transport through mesoscopic systems”. *Physical Review B* **71**, 205304 (2005).
- [36] Michael J Kastoryano, Fernando GSL Brandão, András Gilyén, et al. “Quantum thermal state preparatio” (2023). [arXiv:2303.18224](https://arxiv.org/abs/2303.18224).
- [37] Ming-Jie Zhao and Herbert Jaeger. “Norm-observable operator models”. *Neural computation* **22**, 1927–1959 (2010).
- [38] Sandesh Adhikary, Siddarth Srinivasan, and Byron Boots. “Learning quantum graphical models using constrained gradient descent on the stiefel manifold” (2019). [arXiv:2101.08448v1](https://arxiv.org/abs/2101.08448v1).
- [39] M. S. Vijayabaskar David R. Westhead, editor. “Hidden markov models”. Volume 2, page 18. Humana New York, NY. (2017).

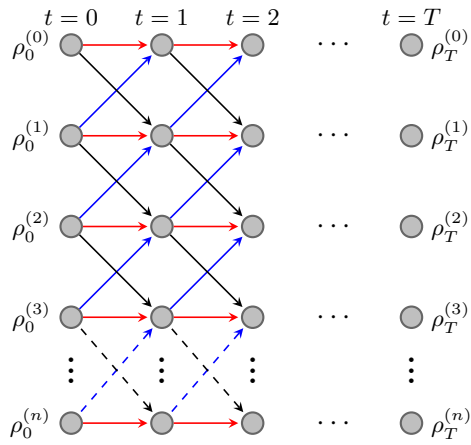


Figure 2: The expanded calculation diagram of  $\rho^n(t)$  for a set of sequences  $y_0, y_1 \dots, y_T$ : The red line represents the Kraus operator  $K_y$  in  $\{K\}$ , the black line represents the Kraus operator  $R_y$  in  $\{R\}$ , and the blue line represents the Kraus operator  $A_y$  in  $\{A\}$

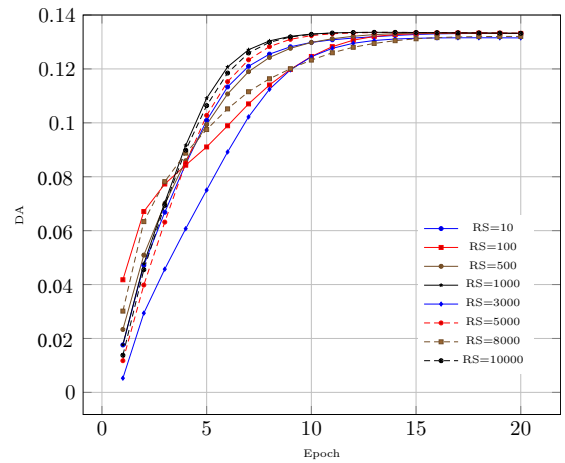


Figure 4: The training result of SHQMM( $N_{max} = 3$ , AR, Single Qubit) in random initialization

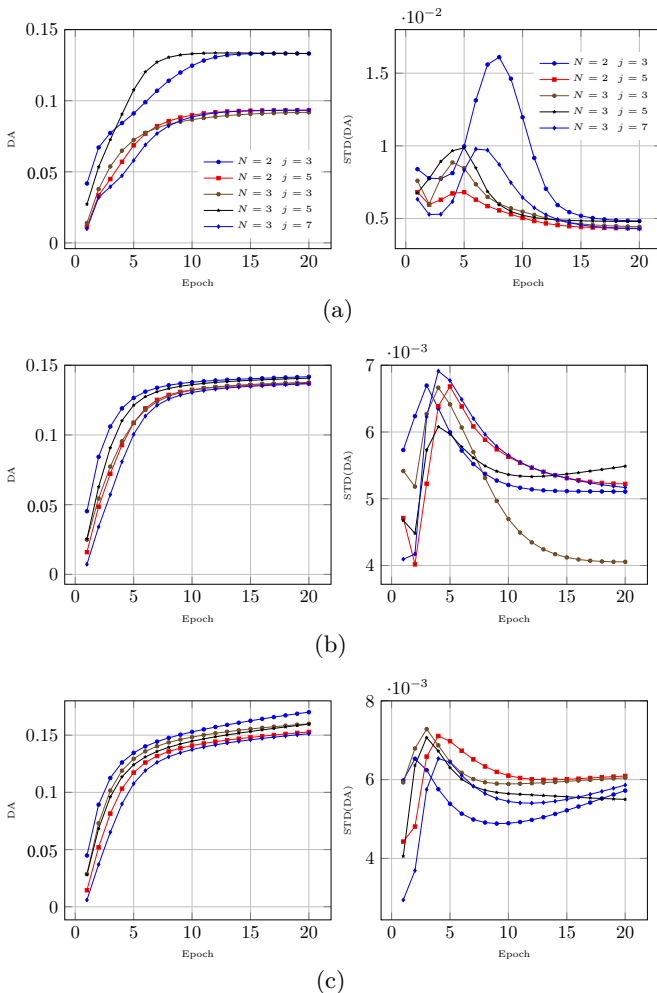


Figure 3: The training result of the different SHQMM for quantum data under different parameters  $N$  and  $j = 2k + 1$ . The subfigures (a), (b), (c) represent the training results for choosing single, double and three qubits quantum system, respectively.

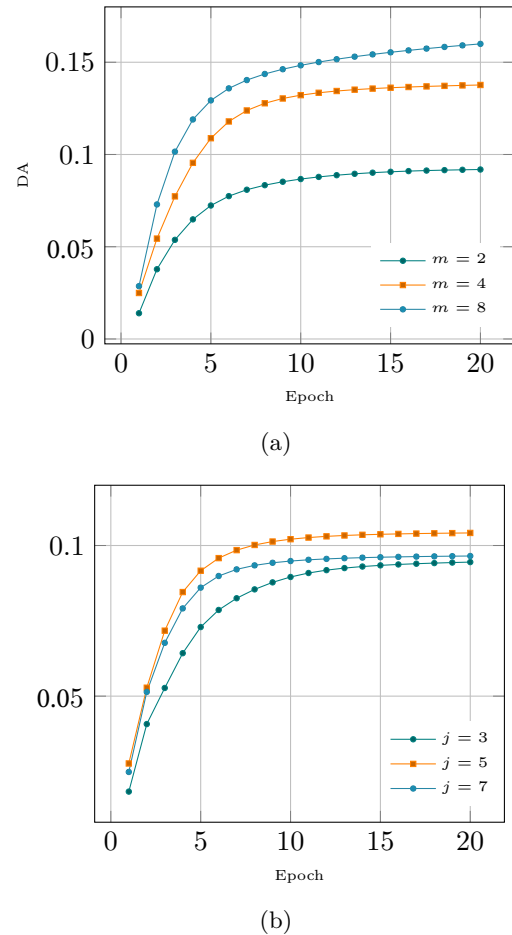


Figure 5: The relationship between the training outcome of SHQMM and the number of parameters. (a) represents the variation of DA with the dimension  $m$  of the Kraus operator. (b) represents the variation of DA with the  $j$  of the conditional density matrix. The training results are more sensitive to  $m$  than  $j$

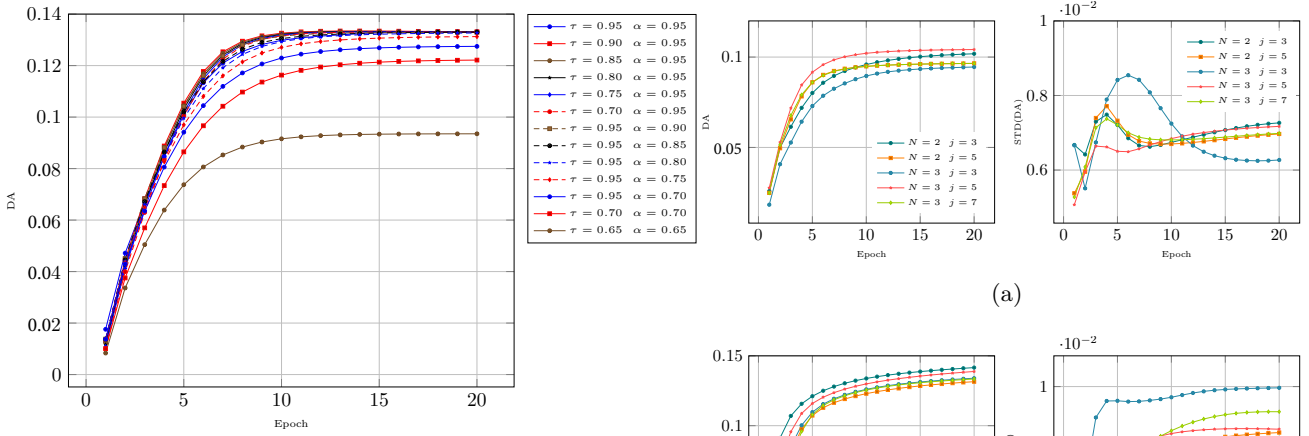
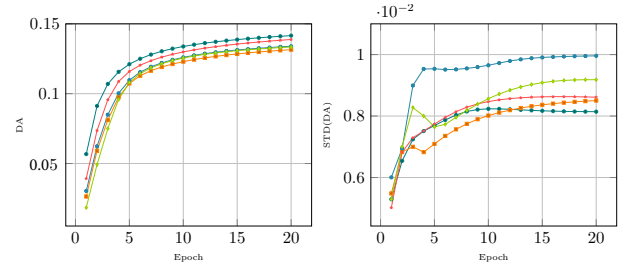
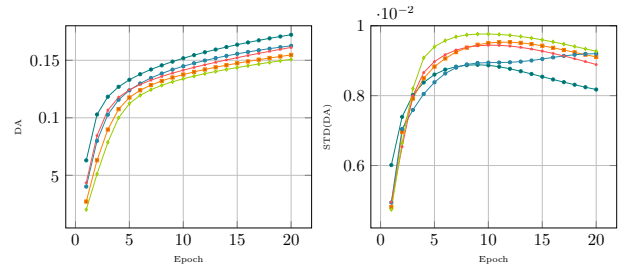


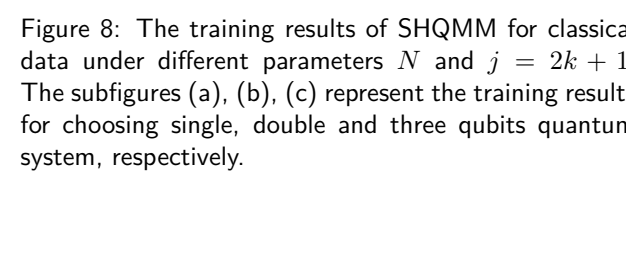
Figure 6: The effect of DA on different hyperparameters tuple  $(\tau, \alpha)$  ( $\beta = 0.90$ ). The best DA is 0.1332 ( $\tau = 0.95, \alpha = 0.95$ ) and the worst DA is 0.0951 ( $\tau = 0.65, \alpha = 0.65$ )



(a)

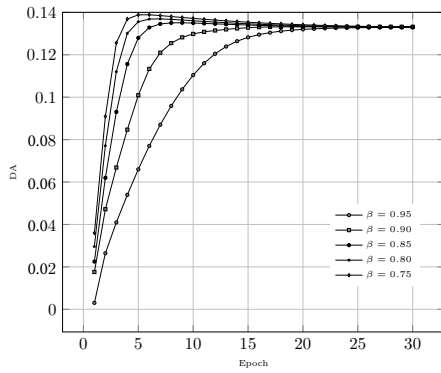


(b)

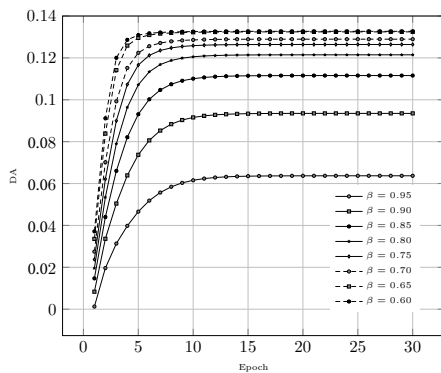


(c)

Figure 8: The training results of SHQMM for classical data under different parameters  $N$  and  $j = 2k + 1$ . The subfigures (a), (b), (c) represent the training results for choosing single, double and three qubits quantum system, respectively.



(a)



(b)

Figure 7: Choose the hyperparameters for sHQMM. (a) shows the DA of  $\tau = 0.95, \alpha = 0.95$ ; (b) shows the DA of  $\tau = 0.65, \alpha = 0.65$ .