Delft University of Technology

A quantum information theoretic analysis of reinforcement learning-assisted quantum architecture search

Sadhu, Abhishek; Sarkar, Aritra; Kundu, Akash

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

A quantum information theoretic analysis of reinforcement learning-assisted quantum architecture search

Abhishek Sadhu[1,2] · Aritra Sarkar[3] · Akash Kundu[4,5,6]

Abstract
In the field of quantum computing, variational quantum algorithms (VQAs) represent a pivotal category of quantum solutions across a broad spectrum of applications. These algorithms demonstrate significant potential for realising quantum computational advantage. A fundamental aspect of VQAs involves formulating expressive and efficient quantum circuits (namely ansatz), and automating the search of such ansatz is known as quantum architecture search (QAS). Recently reinforcement learning (RL) techniques is utilized to automate the search for ansatzes, know as RL-QAS. This study investigates RL-QAS for crafting ansatz tailored to the variational quantum state diagonalisation problem. Our investigation includes a comprehensive analysis of various dimensions, such as the entanglement thresholds of the resultant states, the impact of initial conditions on the performance of RL-agent, the phase transition behaviour of correlation in concurrence bounds, and the discrete contributions of qubits in deducing eigenvalues through conditional entropy metrics. We leverage these insights to devise an entanglement-guided admissible ansatz in QAS to diagonalise random quantum states using optimal resources. Furthermore, the methodologies presented herein offer a generalised framework for constructing reward functions within RL-QAS applicable to variational quantum algorithms.

1 Introduction

Quantum computing represents a paradigm-shifting approach to computation, leveraging the principles of quantum mechanics to process information in ways fundamentally different from classical computers. At its core, this technology employs quantum bits, or qubits, which can exist in superposition states, enabling parallel computation paths. The information in qubits is transformed via quantum circuits consisting of quantum logic gates that are capable of acting on these computation paths in tandem. The quantum circuits represent a corresponding quantum algorithm that can effectuate a computational advantage for a target application (Harrow 2017; Montanaro 2016; Sadhu et al. 2023). These algorithms are either constructed manually, based on conventional logical reasoning or automatically by composing the quantum gates to replicate a required input–output behaviour via training. The latter approach is termed quantum architecture search (QAS) (Zhang et al. 2022; Du et al. 2022; Kuo et al. 2021; Zhang et al. 2021.)

QAS typically consists of two parts. First, a template of the circuits is built, called the ansatz. The ansatz can have parametric quantum gates, e.g. rotation angles. Then, these parameters are determined via the variational principle using a classical optimiser in a feedback loop. Algorithms constructed via this technique are called variational quantum algorithms (VQA) (McClean et al. 2016; Cerezo et al. 2021). QAS can also be used to determine non-parametric circuits as an approach for quantum program synthesis (Sarkar 2024). In this research, however, we will focus on QAS in the context of VQA.

The ansatz design is critical, as it directly influences the expressivity and efficiency of the quantum solution. Given the vast potential configuration space of quantum circuits, the process of identifying a suitable ansatz is a challenging problem. In Ostaszewski et al. (2021); Kundu et al. (2024); Patel et al. (2024); Du et al. (2022), the search for the quantum circuits that solve various VQA has been automated using reinforcement learning (RL) techniques (Sutton and Barto 2018). We particularly call ansatzes capable of solving a VQA an *admissible ansatz*. The automated search for such an admissible ansatz using the RL-agent is termed

Extended author information available on the last page of the article

reinforcement learning-assisted quantum architecture search (RL-QAS). In this setting, after a certain number of episodes, the RL-agent returns possible configurations of quantum ansatz that solve the problem. For E_{tot} episodes, the agent returns E_s successful episodes, and in $E_{tot} - E_s$ episodes, the agent fails to find a configuration within a specific predefined accuracy and depth. It has been shown in Ostaszewski et al. (2021); Wu et al. (2023); Kundu et al. (2024); Patel et al. (2024); Kundu (2024) that classical machine learning-driven QAS algorithms outperform the state-of-art structures of ansatz and effectively optimise the number of parameters, depth, and impact of noise in VQAs.

The use-cases of such admissible ansatzes cover applications in quantum information theory (LaRose et al. 2019; Cerezo et al. 2021; Tan and Volkoff 2021; Cerezo et al. 2020; Kundu and Miszczak 2022), chemistry (Kandala et al. 2017; Mustafa et al. 2022; Delgado et al. 2021) and combinatorial optimization (Khairy et al. 2020; Liu et al. 2022; Glos et al. 2022). Yet, a thorough investigation of such admissible ansatzes is still missing in recent literature. It has been previously shown (Patel et al. 2024) that in RL-QAS methods, it is crucial to feed proper information about the problem to the agent to enhance the agent's performance. One of the ways of providing this information is by engineering a reward function. In RL, the formulation of the reward function and its engineering can be sparse or dense. The choice of optimal reward function depends on trial and error by investigating the performance of the agent under various reward configurations.

In this article, we numerically investigate the admissible ansatzes proposed by the RL-agent from the perspective of quantum information theory (Wilde 2013). For the investigation, we specifically focus on the reinforcement learning enhanced variational quantum state diagonalisation (RL-VQSD) problem (Kundu et al. 2024). In the VQSD algorithm (LaRose et al. 2019), the problem-inspired ansatz is based on density matrix exponentiation (Lloyd et al. 2014), but its depth and number of gates increase exponentially with the number of qubits. Hence, searching for an optimal ansatz (with a minimum number of gates and depth) that diagonalises a quantum state is an open research problem. In RL-VQSD, the search for the optimal configuration of ansatz is done by using a double deep-Q network (DDQN). See Sect. 3 for further details on the VQSD algorithm, the agent and environment specifications. The RL-agent proposes $E_s < E_{tot}$ admissible ansatzes that solve the VQSD problem. The primary motivation of our work is to filter potential admissible ansatzes based on concurrence, using the highest and lowest entanglement bounds. This helps identify the range of optimal ansatzes and determine if increasing entanglement can improve the agent's performance. Another goal is to see

if we can find an admissible ansatz that both diagonalizes and enhances the entanglement of the vacuum state.

Our investigation shows that the concurrence of the E_s ansatz lies between an upper and lower bound, independent of the different initialisation of the double deep-Q network (DDQN), where each initialisation of DDQN corresponds to a different random initialisation of the weights. Moreover, through the numerical analysis, we show that the upper and lower bounds are anti-correlated with respect to the increment in the initial state entanglement. The anti-correlation turns into a mild, and later strong, correlation when the initial state entanglement goes beyond 0.322, indicating a phase transition in correlation between the upper and lower bounds of the concurrence. The phase transition provides us with an in-depth insight into the relation between the input and the output of the VQSD algorithm. Further, our investigation reveals that the RL-agent can generate high concurrence admissible ansatzes with fewer 2-qubit gates and circuit depth compared to ansatzes with very low concurrence. This is used as a guiding principle to design a two-part ansatz, where an entanglement enhancing initial block adjusts the input state's concurrence thereby enhancing the performance of the RL-agent by a factor of 2. Furthermore, we quantify the contribution of individual qubits of the RL-ansatz using conditional quantum entropy. Using such a measure, we observe a correlation between entanglement and quantum entropy, revealing why RL-VQSD can efficiently find the largest eigenvalues but fails to find the smallest ones.

The structure of the paper is as follows. We present an overview of our main results in Sect. 2. In Sect. 3, we present an overview of the reinforcement learning enhanced variational quantum state diagonalisation algorithm. We present an analysis of the sampled 2-qubit Haar random quantum states in Sect. 4. We present the main results in Sect. 5. Specifically, in Sect. 5.1, we analyse the upper and lower bounds on the concurrence of the admissible ansatzes that solve the VQSD problem. In Sect. 5.1.1, we present the variation of the entanglement bounds for different weight initialisation of the DDQN. We observe the correlation properties of the entanglement bounds dependent on the concurrence of the initial state in Sect. 5.1.2. In Sect. 5.1.3, we present the dependence of the entanglement bounds on the number of gates and circuit depth of the ansatz. The analysis of the contribution of individual qubits to the admissible ansatz is provided in Sect. 5.2. In particular, we present the contribution of each qubit for change in the entanglement of the ansatz in Sect. 5.2.1. In Sect. 5.2.2, we explain why RL-VQSD can efficiently find the largest eigenvalues of the initial state but not the smallest eigenvalues. We provide concluding remarks and discuss possible future directions in Sect. 6.

2 Contributions

Through the following points, we summarise the contribution of the paper.

1. We numerically demonstrate that the entanglement of the states generated by the RL-agent-proposed admissible ansatzes lies within a lower and upper concurrence bound. The upper and lower bounds of concurrence remain consistent across various weight initialisation of the DDQN. Furthermore, we show that these bounds are anti-correlated with respect to increasing entanglement until concurrence surpasses 0.322; after this point, the anti-correlation turns into correlation. Hence, we introduce the concurrence 0.322 as the phase transition point.
2. Investigation of ansatzes in upper and lower bounds reveals that the optimal ansatz for VQSD generates strongly entangled states. The RL-agent requires fewer gates and circuit depth to generate such states to produce strongly entangled admissible ansatzes than weakly entangled ansatzes. We then utilise this information to minimise RL-agent and environment (which is the configuration of ansatz) interaction, thereby optimising the convergence time of the RL-VQSD. Specifically, we propose a two-part ansatz, where an entanglement enhancing initial block adjusts the input state's concurrence thereby enhancing the performance of the RL-agent by a factor of 2.
3. We numerically evaluate individual qubits' contribution in diagonalising 2-qubit quantum states using conditional quantum entropy. Our investigation shows that the qubits have equal contributions on average in diagonalising random quantum states.
4. We further investigate the correlation in conditional entropy between qubits for different inferred eigenvalues. It requires mild correlation for the first two largest eigenvalues and mild anti-correlation for the smallest eigenvalues. Most admissible ansatzes lie in a mild correlation regime, explaining VQSD's limitation in finding the smallest eigenvalues.

3 The reinforcement learning enhanced variational quantum state diagonalisation

The VQSD algorithm (LaRose et al. 2019) for a quantum state ρ comprises three subroutines: (1) In the training subroutine, the parameters $\vec{\theta}$ of a quantum gate sequence $U(\vec{\theta})$ are optimised for a given state ρ. Ideally, after optimisation, the sequence $U(\vec{\theta}_{\text{opt}})$ satisfies

$$\rho' = U(\vec{\theta}_{\text{opt}})\rho U(\vec{\theta}_{\text{opt}})^{\dagger} = \rho_{\text{diag}},$$

where ρ_{diag} is ρ diagonalised in its eigenbasis and $\vec{\theta}_{\text{opt}}$ are the optimal angles. Classical gradient-based methods such as SPSA and Gradient Descent or gradient-free optimisation (Powell 1994, 2006) can be utilised in the training process. (2) In the eigenvalue readout subroutine, using the optimised unitary $U(\vec{\theta}_{\text{opt}})$ and one copy of state ρ, one can extract all the eigenvalues for low-rank states or the largest eigenvalues for full-rank states. This is achieved by measuring ρ' in the computational basis, $\mathbf{b} = b_1 b_2 \dots b_n$, as follows:

$$\lambda' = \langle \mathbf{b}|\rho'|\mathbf{b}\rangle,$$

where λ' are inferred eigenvalues. (3) In the final step, we prepare the eigenvectors associated with the largest eigenvalues. If $\mathbf{b}'$ is a bit string associated with λ', then the inferred eigenvectors $|v'_{\mathbf{b}'}\rangle$ are obtained as follows:

$$|v'_{\mathbf{b}'}\rangle = U(\theta_{\text{opt}})^{\dagger}|\mathbf{b}'\rangle = U(\theta_{\text{opt}})^{\dagger}\left(X^{b_1}\otimes\dots\otimes X^{b_n}\right)|\mathbf{0}\rangle.$$

In the reinforcement learning enhanced variational quantum state diagonalisation, an RL-agent is utilised to find the optimal configuration of $U(\vec{\theta})$. The agent contains a double deep-Q network (DDQN) (Hasselt et al. 2016) with ϵ-greedy policy. The DDQN optimises the stability of Q-learning by using two networks to decouple the action selection a' in the subsequent state s' from its current reward r evaluation, leading to less overestimated Q-values and more stable and reliable learning outcomes.

The key update equation is

$$Q(s, a; \theta) \leftarrow Q(s, a; \theta)$$
$$+\alpha[r + \gamma Q(s', \arg\max_{a'} Q(s', a'; \theta); \theta') - Q(s, a; \theta)] \quad (1)$$

where $Q(s, a; \theta)$ is the predicted Q-value for state s and action a, parameterised by weights θ of the online network. α is the learning rate, and γ is the discount factor. The settings of the parameters are similar to the ref. (Kundu et al. 2024).

The optimal policy is obtained by optimising the weights of the DDQN using the ADAM optimiser. In the RL setup, the RL-state encodes the ansatz and is defined by a tensor-based encoding presented introduced in Patel et al. (2024). After each environment and agent interaction, the reward function

$$R = \begin{cases} +\mathcal{R} & \text{for } C_t(\vec{\theta}) < \zeta + 10^{-5} \\ -\log\left(C_t(\vec{\theta}) - \zeta\right) & \text{for } C_t(\vec{\theta}) > \zeta \end{cases}, \quad (2)$$

is calculated. The goal of the RL-agent is to reach the minimum error for a predefined threshold ζ (which is a hyper-

Fig. 1 The difference in conditional entropy converges towards the zero as the entanglement of the sampled input state increases. Meanwhile, as the input state gets closer to the maximally entangled state (with concurrence 1), the magnitude of the largest eigenvalue increases while the other eigenvalue decreases
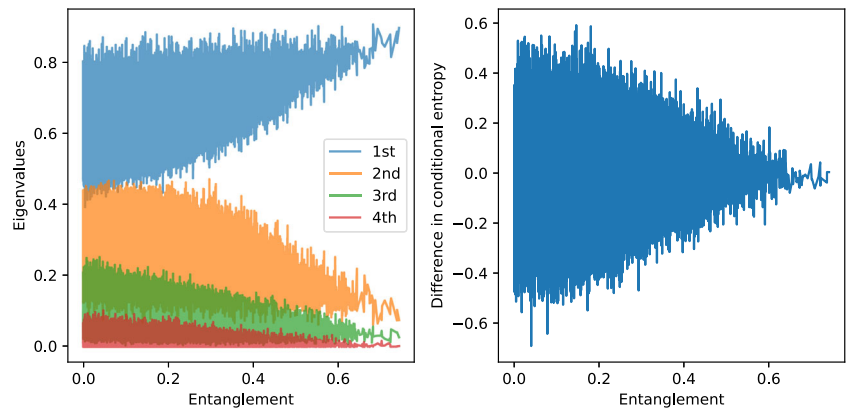
parameter of the model) by optimising the parameters of the $U(\vec{\theta})$ using the COBYLA optimiser. The cost function

$$C_t(\vec{\theta}) = \mathrm{Tr}(\rho^2) - \mathrm{Tr}(\rho_t(\vec{\theta})^2). \tag{3}$$

At each step, t is calculated for the ansatz which outputs a state $\rho_t(\vec{\theta})$.

4 Analysis of the Haar random input quantum states

In this section, we analyse the 2-qubit Haar random quantum states. The Haar random states can be efficiently simulated using polynomial-time algorithms, as demonstrated in prior research (Alagic et al. 2020). This holds even in scenarios where adversaries have unbounded capabilities with black-

box oracle access (Chen et al. 2024). Here, we sample 100,000 Haar random quantum states and observe how the eigenvalues and the qubits' conditional entropy change with the state's entanglement.

In Fig. 1, see that the difference in conditional entropy in qubits of the input state varies evenly on either side of $S_{AB}^{\rho_{AB}} = 0$ where A and B are subsystems containing one qubit each and ρ_{AB} is the sampled Haar random state. Meanwhile, the area of the distribution of $S_{AB}^{\rho_{AB}}$ on either side of zero entropy converges towards zero as the input state entanglement increases. Furthermore, the magnitude of the input quantum state's largest eigenvalue increases with the increase in entanglement, and all the other eigenvalues converge towards 0. It also shows that even for 100,000 seeds, the probability of sampling states with concurrence greater than 0.6 is very low, justifying the low density of points after 0.6 in Fig. 2.
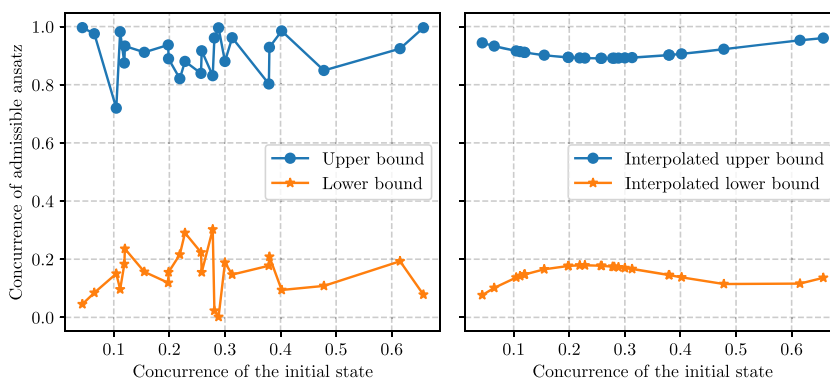
Fig. 2 Reciprocal behaviour in the upper and lower bounds of concurrence with respect to the increasing concurrence in the input random state. For the first eleven points, the Pearson correlation coefficient (PCC) is -0.982, denoting a strong anti-correlation between the upper and lower bounds of concurrence. The strong anti-correlation turns into a mild correlation after concurrence of the initial state surpasses 0.322. The correlation after the first eleven points changes from -0.055 to 0.049. There is a phase transition from weak anti-correlation to weak correlation

5 Main results

In this section, we investigate the upper and lower bounds of concurrence of the ansatzes produced by the RL-agent in diagonalising 2-qubit random quantum states sampled from `IBM qiskit`'s `random_density_matrix` module. For the simulation, we consider multiple such quantum states (specified in the caption of figures), and for each state, we use $10,000$ episodes of RL-VQSD algorithm to diagonalise the state. The algorithm consists of DDQN with ϵ-greedy policy where the ϵ decays in each step by a factor of 0.99995 from its initial value 1, down to a minimal value 0.05. Furthermore, the network is trained by ADAM optimizer (Kingma and Ba 2014) with a learning rate of 0.0001. The neural network consists of 5 hidden layers with 1000 neurons and a buffer size of 15,000. The configuration of the RL-VQSD is provided in detail in Sect. 3.

5.1 Entanglement

5.1.1 DDQN weight independent entanglement bound

To benchmark the performance of the RL-agent for a specific problem, it is ideal to take multiple initialisations of the DDQN (see Sect. 3 for a brief description).

To efficiently investigate the upper and lower bounds of concurrence (Hill and Wootters 1997; Wootters 1998) of the admissible ansatz, we calculate the *concurrence of the admissible ansatz*, which is defined by the entanglement induced by the admissible ansatz when initialised with $|00\rangle$. We first sample different random quantum states, then diagonalise them using RL-VQSD (Kundu et al. 2024) for five different initialisations of DDQN weights. Thereafter, the agent suggests different possible configurations for the RL-ansatz, which diagonalises the sampled random quantum states. For a specific state, we collect such valid quantum circuits and find the concurrence of the state obtained after evolving through the RL-ansatz. It should be noted that finding the concurrence of the evolved state indicates how much the RL-ansatz induces entanglement to the input sampled state. See Sect. 4 for further analysis of the sampled random quantum states, where we show that in `qiskit`, `random_density_matrix` even for $100,000$ seeds, the probability of sampling states with concurrence greater than 0.6 is very low.

The results are presented in Table 1. We observe that the optimal configuration of the RL-ansatz always gives us a state with concurrence close to one. Hence, the deviation in the maximum concurrence over different weights of the DDQN is more significant than the deviation in the minima of the concurrence. As seen in Table 1, the deviation in the max concurrence is in the order of 10^{-3}. The low standard deviation implies that there is a negligible deviation in concurrence

Table 1 The average and standard deviation of the final state's maximum (max. concurrence) and minimum concurrence (min. concurrence) after evolution through the circuit for five different initialisations of the DDQN weights for different quantum seed states

State no	Max. concurrence		Min. concurrence	
	avg	std. div	avg	std. div
1	0.874	0.002	0.183	0.004
2	0.849	0.002	0.108	0.028
3	0.975	0.001	0.084	0.006
4	0.936	0.002	0.118	0.009
5	0.880	0.001	0.188	0.013
6	0.911	0.003	0.156	0.024
7	0.933	0.001	0.235	0.013
8	0.928	0.005	0.208	0.013
9	0.831	0.011	0.302	0.010

To obtain the results, we run RL-VQSD to diagonalise 2-qubit random quantum states for $10,000$ episodes. The agent then proposes $E_s < 10,000$ admissible ansatzes, which generate E_s states after the random quantum state evolves them. The max. (min.) concurrence is attributed to the ansatz that generates a state with the highest (lowest) concurrence following the evolution of the random quantum state

across different initialisation of the DDQN for a specific state. This indicates that the bounds on concurrence are invariant for different configurations of the RL-ansatz. Hence, proving that *the entanglement bounds of the RL-ansatz are independent of initialisation of DDQNs*. Due to this observation, it is sufficient to consider a single initialisation of DDQN to benchmark the outcomes in the remaining simulations.

In the following subsection, we investigate how the upper and lower bounds on the entanglement of the RL-ansatz depend on the entanglement of the input quantum state.

5.1.2 Anti-correlation between entanglement bounds

Corresponding to a predefined threshold, the agent proposes different architectures of quantum circuits that diagonalise the quantum state. To observe the amount of entanglement generated by the RL-ansatz, we calculate the concurrence of the state after passing through the RL-ansatz, just before the dephasing operation. By sorting out the quantum circuit with maximum and minimum entanglement, we observe their variation with the entanglement of the input state. It has been proved in the previous section that the upper and lower bounds of entanglement are independent of different initialisation of DDQNs; hence, for further benchmarking, we can constrain ourselves to just one initialisation of the DDQN weights.

In Fig. 2, we illustrate the variation of the entanglement bounds with the entanglement of the input quantum state. To investigate the dependency, we use the Pearson correlation coefficient (PCC) (Rodgers and Nicewander 1988) as a quan-

tifier of correlation. See Sect. 1 for a brief description. We see that the upper and lower bounds of entanglement are initially strongly anti-correlation when the input state entanglement is below 0.322 and then strongly correlate afterwards, indicating a *phase transition* in the upper and lower bounds of entanglement.

Intuitively, when the entanglement of the starting state is low, i.e. $k \le 0.322$, it becomes easier for the dephasing subroutine of the VQSD algorithm to eliminate the off-diagonal terms. Hence, the agent proposes an admissible ansatz to either convert the state into a separable state (eliminating the need for a dephasing gate at the end) or increase its entanglement to the maximum level, thereby making full use of the dephasing gate's potential. This dynamic explains the observed anti-correlation between the entanglement bounds.

As noted previously, when the entanglement of the starting state is high (in the region $k \ge 0.322$), it becomes easier for the dephasing subroutine to eliminate the off-diagonal terms. Thus, for states with high entanglement, the agent proposes an admissible ansatz to maximise the entanglement of the starting state. Hence, we observe the positive relation between the entanglement bounds.

Next, we obtain the variation of PCC between the upper and lower bounds on the concurrence of the initial state. Specifically, we introduce a parameter $k \in (i, j)$ to split the range of initial state concurrence $\mathfrak{u} := [i, j]$ into two intervals: $\mathfrak{u}_1 := [i, k]$ and $\mathfrak{u}_2 := [k, j]$. Also, let us denote the set of lower and upper bounds on the concurrence of the ansatz as u and l, respectively. For both the ranges $\mathfrak{u}_1$ and $\mathfrak{u}_2$, we introduce the following correlation function:

$$\eta^k_{w_{ik}} = w_{ik}\text{PCC}_{ik} + (1 - w_{ik})\text{PCC}_{kj}, \qquad (4)$$

where $w_{ik} \in \{0, 1\}$ and $\text{PCC}_{ik}, \text{PCC}_{kj}$ are the value of PCC between the upper and lower bounds of concurrence of the ansatz for the intervals $\mathfrak{u}_1$ and $\mathfrak{u}_2$, respectively. For the sake of simplicity, we denote PCC_{ik} as PCC^{ul}_{ik} and also PCC_{kj} as PCC^{ul}_{kj}.

Notes 1 e observe that depending on the values of w_{ik}, $\eta^k_{w_{ik}}$ can take values either PCC_{ik} or PCC_{kj}. For $\eta^k_{w_{ik}} = \text{PCC}_{ik}$, we obtain the amount of (anti-)correlation between the lower and upper bounds of concurrence of the ansatz for the interval $\mathfrak{u}_1$. While for $\eta^k_{w_{ik}} = \text{PCC}_{kj}$, we obtain the amount of (anti-)correlation for the interval $\mathfrak{u}_2$. We leave proving $\eta^k_{w_{ik}}$ is a valid measure of correlation for arbitrary values of $w_{ik} \in (0, 1)$ as an open problem.

We present in Fig. 3 the variation of η^k_0 (in blue) and η^k_1 (in red) for different values of k. We observe that $\eta^k_1 \approx -1 \ \forall k \in (i, j)$. This indicates strong cumulative anti-correlation between the upper and lower bounds of concurrence. For η^k_0, we observe a gradual change of PCC_{kj}
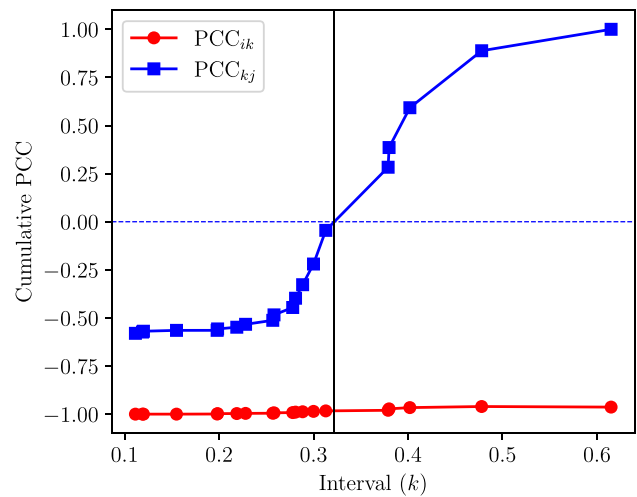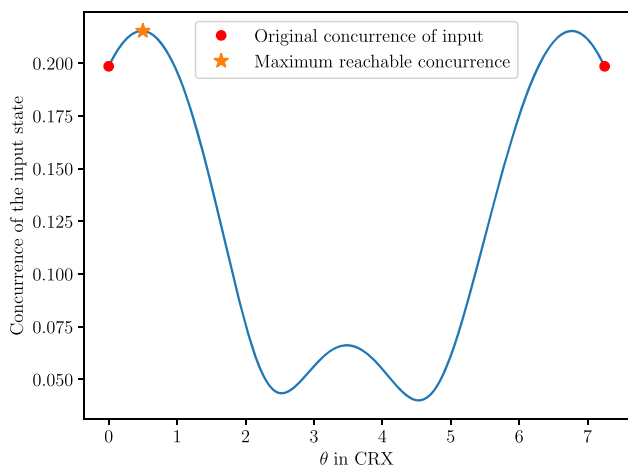

Fig. 3 We present the phase transition of correlation between the upper and lower bounds of concurrence for the admissible RL-VQSD ansatz. To show this, we split the total range of the concurrence of the input state in two intervals $\mathfrak{u}_1 := [i, k]$ and $\mathfrak{u}_2 := [k, j]$, where k is a free parameter and plot the variation of the Pearson correlation coefficient between the upper and lower bound on the concurrence of the ansatz corresponding to these two intervals. We denote by PCC_{ik} and PCC_{kj} the correlation coefficients in the intervals $\mathfrak{u}_1$ and $\mathfrak{u}_2$, respectively. The phase transition point is observed at $k = 0.322$ where the concurrence bounds change from mild anti-correlation to mild correlation for the interval $\mathfrak{u}_2$ while remaining strongly anti-correlated in the interval $\mathfrak{u}_1$

from -0.582 to 1. We note that η^k_0 changes from -0.0451 to 0.279 for $k \in [0.313, 0.379]$ with $\eta^k_0 = 0$ at $k = k_* = 0.322$.
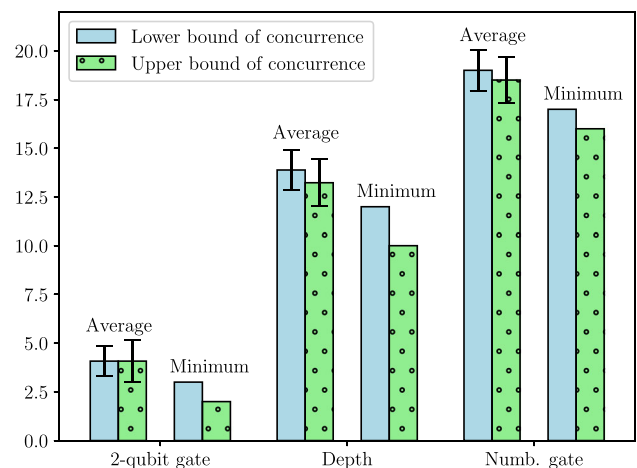
Fig. 4 The RL-agent can generate an admissible ansatz with strong entanglement with fewer 2-qubit gates (2-qubit gate) and circuit depth (Depth) and total number of gates (Numb. gate) compared to the ansatzes with weak entanglement. As the main aim of RL-VQSD is to diagonalise unitary with a small number of gates and depth, our analysis shows that if we particularly focus on the admissible ansatzes near the upper bound of concurrence, we can not only reduce the computational cost of RL-VQSD severely but also find the optimal admissible ansatzes. The numbers in the bars correspond to the amount of entanglement generated by the RL-ansatz when initialised in the vacuum state

Fig. 5 The concurrence of input state with θ of $CRX(\theta)$ gate in ENTANGLEMENT ENHANCING increases the concurrence of the input random state from 0.198 to 0.215 when $\theta = 0.5$

We call this point the *phase transition point of correlation between upper and lower bounds of concurrence* for the VQSD ansatz. At values of $k < k_*$, there is anti-correlation between the upper and lower bounds of concurrence of the ansatz for both the intervals $\mathbb{u}_1$ and $\mathbb{u}_2$. While for values of $k > k_*$, there is (anti-)correlation between the upper and lower bounds of concurrence of the ansatz for both the intervals $(\mathbb{u}_2)$ $\mathbb{u}_1$.

The investigation in Figs. 2 and 3 suggests that the structure of the quantum state in terms of its entanglement properties can significantly influence the effectiveness of certain ansatz strategies. Specifically, when the entanglement is low, i.e. below 0.322, the variability in the entanglement bounds can be exploited. For wide entanglement bounds, the flexibility of the ansatz allows for the generation of states that are either highly entangled or nearly separable, starting from $|0\rangle^{\otimes 2}$.

5.1.3 Enhancing the performance of agent based on initial state entanglement

Here, we provide a detailed investigation of the admissible ansatzes produced by the RL-agent in terms of ansatz depth,

the total number of 2-qubit gates, and the total number of gates in the upper and lower bounds of concurrence.

Our investigation reveals that the admissible ansatzes corresponding to the upper bound of concurrence have, on average, a shorter depth of 13.23 and a smaller number of gates of 18.5 compared to the admissible ansatzes corresponding to the lower bound of concurrence where the depth is 13.9 and number of gates is 19 over 25 random quantum states. Moreover, Fig. 4 makes it more prominent that the minimum number of gates and the depth of the admissible ansatzes are smaller for the upper bound of concurrence than the lower. This observation is crucial in reducing the subspace of all possible admissible ansatzes. The main motivation of RL-QAS in VQAs is to find ansatzes that solve a VQA with a minimum number of gates and circuit depth; in the case of RL-VQSD, this observation can help us to narrow down the search for an optimal admissible ansatz in the two following ways.

Entanglement dependent initialisation of RL-state

In RL-VQSD (Kundu et al. 2024), the RL-state that encodes the ansatz initiates from an empty circuit (i.e. in circuit Eq. 5, the ENTANGLEMENT ENHANCING block is absent and $U_{\text{RL-agent}} = \mathbb{1}$). We showed that the optimal admissible ansatz, on average, lies in the upper bound of concurrence, motivating us to initialise the quantum state (that is diagonalised using RL-VQSD) closest to the upper bound of entanglement. To do this, the ansatz is divided into two parts

$$\text{ENTANGLEMENT ENHANCING BLOCK} \qquad (5)$$
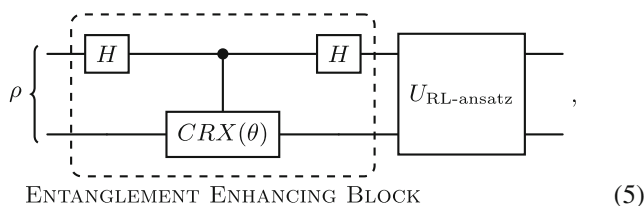
where the first, ENTANGLEMENT ENHANCING (EE) block, helps the diagonalising state to achieve the highest entanglement before it is fed to the second part, $U_{\text{RL-agent}}$, which the RL-agent decides. For example, for seed 27 (see Sect. 4 for details), we get a random quantum state with concurrence 0.198 (for seed 27, and see Appendix 4 for details on how

Table 2 Turning on the ENTANGLEMENT ENHANCING (EE) mode (denoted as Enhanced ($\theta = 0.5$)) enhances the learning of the performance of the RL-agent in the absence of the EE mode (denoted as Default ($\theta = 0$))

| EE mode (θ)	Tot. succ ep	Tot. rwd coll	1st 100 succ. ep. (avg.)		
			1q gate	2q gate	depth
Enhanced ($\theta = 0.5$)	858	4.45×10^5	15.41	4.02	14.6
Default ($\theta = 0$)	387	2.14×10^5	14.71	5.29	15.59

In the first 3500 episodes, the Enhanced ($\theta = 0.5$) collects 2.08 times more reward (tot. rwd. coll.) and 2.22 times more successful episodes (Tot. succ. ep.) than Default ($\theta = 0$). Moreover, the average number of 2-qubit (2q gate) and depth over the first 100 successful episodes reduced in the case of Enhanced ($\theta = 0.5$) settings compared to Default ($\theta = 0$), showing enhanced learning of the RL-agent for entanglement enhanced input state

to generate Haar random state), and using the EE, we can increase the input state concurrence to 0.215 at $\theta = 0.5$ as shown in Fig. 5.

Our investigation reveals that turning on the EE block with $\theta = 0.5$ enhances the performance of the RL-agent by 2-times compared to the default setting at $\theta = 0$. The results are summarised in Table 2 where we see that the number of successful episodes and the reward acquired (where the reward function is in the form defined in Eq. 2 by the agent in the RL-VQSD process) increase by 2-fold while the EE block is activated for $\theta = 0.5$ compared to default settings at $\theta = 0$.

Meanwhile, in Table 3, we show that the EE block also helps to optimise the cost function and the minimum number of 2-qubit gates in the optimal admissible ansatz compared to the default settings.

It should be noted that due to the sake of the interpretability of the EE block, we chose the block resembling a maximally entangled circuit for a 2-qubit system, where the entangler (i.e. the CX gate is parameterised and can induce no entanglement (for $\theta = 0$) and maximal entanglement (for $\theta = 0.5$) to $|00\rangle$ state. The second Hadamard (H) gate on the first qubit is introduced to cancel out the first H in the absence of the entangler at $\theta = 0$. We can automate the search for gateset in the EE block utilising an RL-agent and diagonalise the input state simultaneously by decomposing the RL-QAS into two subroutines. Suppose the total number of steps in an episode is S, then in the first subroutine for $s_{EE} < S$ steps the agent's task would be to find a gateset (consists of maximum s_{EE} gates) that maximises the entanglement of the input state by utilising a reward function R_{EE}. One can formulate R_{EE} similar to ref. (Kuo et al. 2021). Wherein in the second subroutine, for the remaining steps $S - s_{EE}$, the agent focuses on finding possible structures of $U_{RL\text{-}ansatz}$ to diagonalise the output state of the EE block using the reward defined in Eq. 2.

The aforementioned observations lead us to the conclusion that maximising the entanglement of the input state through the initialisation strategy of Fig. 5 enhances the learning process of the RL-agent and facilitates the further optimisation of the admissible ansatzes compared to the start-of-art (Kundu 2024).

Table 3 Turning on the ENTANGLEMENT ENHANCING (EE) mode (denoted as Enhanced ($\theta = 0.5$)) optimises the number of minimum number of 2-qubit gates (Min. 2q gate) in RL-ansatz better than in the absence of the EE mode (denoted as Default ($\theta = 0$))

EE mode (θ)	Cost function	Min. 1q gate	Min. 2q gate
Enhanced ($\theta = 0.5$)	5.12×10^{-7}	14	2
Default ($\theta = 0$)	6.08×10^{-7}	11	5

Moreover, the Enhanced ($\theta = 0.5$) mode minimises the cost function better than the Default ($\theta = 0$) mode

5.2 Conditional quantum entropy

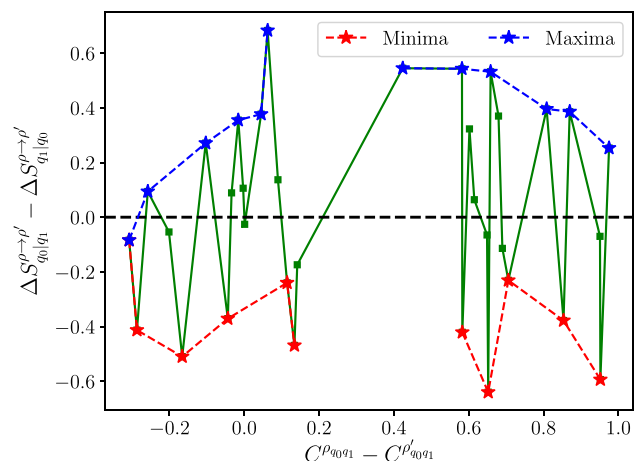### 5.2.1 Contribution of individual qubits for change in entanglement

We quantify the contribution of each qubit using the conditional quantum entropy $S_{q_0|q_1}^{\rho'_{q_0q_1}}$ (Cerf and Adami 1997, 1999; Wilde 2013). The conditional entropy of an individual qubit for an N qubit quantum system indicates the amount of uncertainty remaining about that qubit after measuring the whole system. Hence, it is essential to know how much information the state of qubits provides about the state of the N qubits. For a detailed discussion and notation, see Sect. 1.

Let the change in the conditional entropy of the qubit $q_0(q_1)$ after evolving through the VQSD ansatz be given by $\Delta S_{q_0|q_1}^{\rho \to \rho'}(\Delta S_{q_1|q_0}^{\rho \to \rho'})$. We quantify the relative contribution of individual qubits via the quantity:

$$\Delta_{q_0}^{q_1} := \Delta S_{q_0|q_1}^{\rho \to \rho'} - \Delta S_{q_1|q_0}^{\rho \to \rho'}. \tag{6}$$

We note that $\Delta_{q_0}^{q_1} > 0$ implies a greater change in conditional entropy of q_0 as compared to q_1, and hence the ansatz has a greater contribution from q_0 as compared to q_1. Following a similar reasoning, $\Delta_{q_0}^{q_1} < 0$ implies q_1 has greater contribution as compared to q_0, while $\Delta_{q_0}^{q_1} = 0$ implies both q_0 and q_1 contribute equally to the ansatz.

We present the variation of $\Delta_{q_0}^{q_1}$ for the change in concurrence due to the ansatz given by $\Delta_c := C^{\rho_{q_0q_1}} - C^{\rho'_{q_0q_1}}$ in Fig. 6. We observe that the distribution of $\Delta_{q_0}^{q_1}$ is bounded in the interval $[-0.7, 0.7]$ and is uniform with respect to

Fig. 6 We analyse the relative change in the contribution of the first and second qubits as a function of changes in entanglement. To quantify the relative contribution of individual qubits to the RL-VQSD ansatz, we use $\Delta_{q_0}^{q_1}$ (see Eq. (6)). For the change in concurrence due to the ansatz, the distribution of $\Delta_{q_0}^{q_1}$ is uniform with respect to $\Delta_{q_0}^{q_1} = 0$. The cumulative weight of all data points in the region $\Delta_{q_0}^{q_1} < 0$ is 4.854, while that for data points in the region $\Delta_{q_0}^{q_1} > 0$ is 5.524. This indicates an equal contribution from both qubits to the RL-VQSD ansatz across the range of changes in entanglement

$\Delta_{q_0}^{q_1} = 0$. The magnitude of the cumulative weight of all data points in the region $\Delta_{q_0}^{q_1} < 0$ is 4.854, while that for data points in the region $\Delta_{q_0}^{q_1} > 0$ is 5.524. This indicates equal contribution from both the qubits to the RL-VQSD ansatz for the range of change in entanglement. In the region of $\Delta_{q_0}^{q_1} > 0$, the relative contribution of q_0 first increases and then decreases. However, no such pattern is observed for the region of $\Delta_{q_0}^{q_1} < 0$. When $\Delta_c > 0.4$, the value of $\Delta_{q_0}^{q_1}$ monotonically decreases with increase in Δ_c. Also, for $\Delta_c < 0.2$, $\Delta_{q_0}^{q_1}$ increases monotonically with increase in Δ_c.

### 5.2.2 Problem with variational quantum state diagonalisation

In this section, we investigate the contribution of individual qubits to the performance of the RL-VQSD algorithm. Through this investigation, we get a deeper insight into why the variational quantum state diagonalisation algorithm can not find the smallest eigenvalues.

We present the main observations in Fig. 7, where we plot the four eigenvalues of 35, 2-qubit Haar random mixed quantum states with the correlation in conditional entropy among the qubits. For an individual quantum state, the RL-agent proposes **s** different structures of the admissible ansatzes. Then, we get ρ'_i states where $i \in [0, \mathbf{s}]$. Afterwards, we calculate the conditional entropy of individual qubits for each state and take the median.[1] Then, we calculate the correction in conditional entropy among the qubits using PCC.

In Fig. 7, we observe that a mild anti-correlation between the qubits is required to find the two largest eigenvalues. As the magnitude of the largest eigenvalues decreases, the mild anti-correlation turns into a strong correlation between the qubits. This indicates that the qubits require mild to strong correlation to achieve the two largest eigenvalues. Meanwhile, we require a mild to strong anti-correlation between the qubits to find the smallest eigenvalues of the same states. Then, as the magnitude of the smallest eigenvalue increases, the strong anti-correlation turns into mild anti-correlation.

We recall that the correlation is calculated among the qubits based on the conditional entropy of individual qubits. As it is impossible to have two different kinds of correlation (such as mild correlation and mild anti-correlation simultaneously) between the qubits in the same ansatz structure, we can not find the largest and smallest eigenvalues utilising the same ansatz. For example, we mark the eigenvalues of a state with ★, where we observe that to find the largest two eigenvalues, the RL-ansatz needs to induce a mild correlation between the two qubits; meanwhile, to obtain the smallest eigenvalues, we require mild anti-correlation.

---

[1] Calculating the median over the mean is justified by the broad range of conditional entropy distribution, which makes the median a more appropriate measure of central tendency.
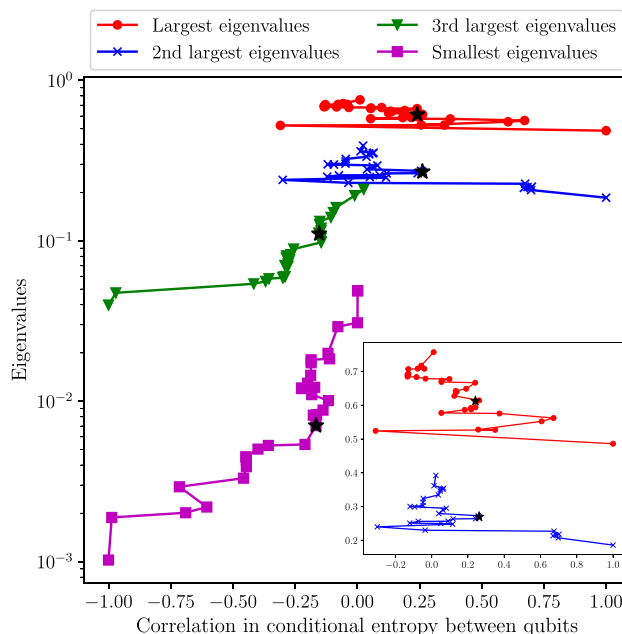
Fig. 7 On average over 35 random quantum states the correlation in conditional entropy between the first and second qubits decreases as the magnitude of the largest and second largest eigenvalues increases. Meanwhile, the same correlation between the qubits increases with the magnitude of the third largest and the smallest eigenvalues. The RL-ansatz proposes specific structures of admissible ansatzes, and for each such ansatz, it is not possible to have both types of correlations simultaneously. Hence, our investigation suggests that it is not feasible to find the smallest eigenvalues using the same ansatz that provides a good approximation of the largest eigenvalues. The correlation of conditional entropy between qubits is calculated using the Pearson correlation coefficient (PCC) and is the median over all configurations of RL-ansatz proposed by the agent. The ★ signifies the eigenvalues corresponding to the same state

The average correlation in conditional entropy among qubits falls within the range of mild correlation for the first two largest eigenvalues across 35 random quantum states. Conversely, the conditional entropy among qubits exhibits a mild anti-correlation for the smallest eigenvalues. This finding, together with the observation in Fig. 8, indicates that with a specific configuration of an admissible ansatz, it is interactable to simultaneously identify the largest and smallest eigenvalues. It should be noted that for the 35 random quantum states, the average correlation in conditional entropy lies in the mild correlation zone, indicating that RL-VQSD can efficiently find the largest few eigenvalues. However, it fails to provide a good approximation for the smallest eigenvalues simultaneously.

6 Discussion

In this article, we use tools from quantum information theory to analyse the admissible ansatzes proposed by the RL-agent

for solving reinforcement learning-based quantum architecture search problems, and we particularly investigate the RL-agent proposed ansatzes in solving recently proposed RL-VQSD problem.

We observe that the concurrence of the admissible ansatz ranges between an upper and lower bound, which are independent of the initialisation of the weights of the deep-Q network. The upper and lower bounds are initially anti-correlated with respect to the initial state entanglement. The anti-correlation turns into a mild and eventually strong correlation as the entanglement of the initial state surpasses beyond a phase transition point at concurrence 0.322. We also see that the optimal configuration of the admissible ansatz lies in the upper bound of concurrence, which has a smaller circuit depth and requires fewer gates than the lower bound. These observations provide insight regarding the relation between the entanglement of the initial state and the entanglement generated by the RL-VQSD admissible ansatzes. We effectively utilise these observations to greatly reduce the RL-agent and environment interaction, i.e. the computational time needed to solve the RL-VQSD problem. Specifically, we propose an entanglement enhancing block that adjusts the input state's concurrence, thereby enhancing the performance of the RL-agent by a factor of 2. Additionally, the admissible ansatzes not only diagonalise the random quantum states with high accuracy but can be used to generate close to maximum entangled states from the vacuum state.

Furthermore, we quantify the contribution of each qubit in the admissible ansatz using conditional quantum entropy. For the admissible RL-VQSD ansatz, we observe an equal contribution from each qubit for the full range of change
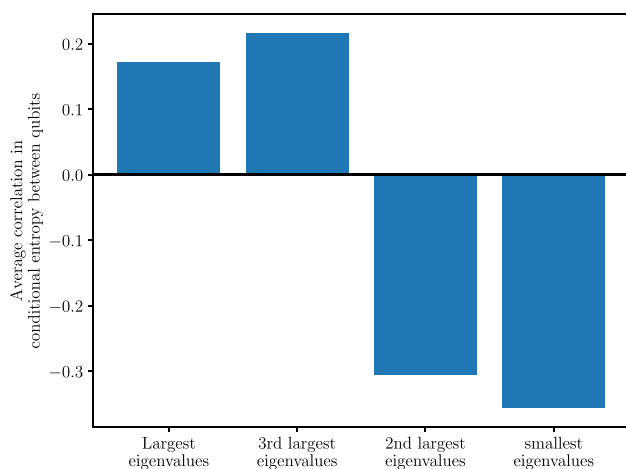
Fig. 8 The average correlation in conditional entropy among qubits is in the range of mild correlation for 35 random quantum states for the first two largest eigenvalues, whereas the conditional entropy among qubits requires a mild anti-correlation for the smallest eigenvalues. This observation, along with the observation in Fig. 7, certify that with a specific configuration of an admissible ansatz, we cannot simultaneously find the largest and the smallest eigenvalues

in the entanglement of the state due to the ansatz. Focusing on the task of obtaining the eigenvalues of the starting state, we observe that a mild correlation between the qubits of the admissible ansatz is required to obtain the two largest eigenvalues, while mild anti-correlation between the qubits is needed to find the smallest eigenvalues of the same state. Noting that it is impossible to have both mild correlation and anti-correlation between the qubits of the same ansatz, we cannot simultaneously obtain both the largest and smallest eigenvalues from the RL-VQSD admissible ansatz. From the observation, most of the admissible ansatz has mild correlation among the qubits, which explains why it is easier for the VQSD algorithm to find the largest eigenvalues than the smallest ones.

For future work, one could conduct a comprehensive study of the implication of introducing correlations such as entanglement (Horodecki et al. 2009), nonlocality (Brunner et al. 2014; Sadhu and Das 2023), or steering (Uola et al. 2020; Sadhu and Das 2024)-based reward functions of the RL-agent in QAS. Specifically, given that most of the admissible ansatzes have high values of concurrence, it is expected that an entanglement or non-locality-based reward function may further reduce the time required to find the optimal configurations. Also, we may expect this to provide an optimal quantum circuit that provides both the eigenvalues of a random quantum state with higher accuracy within a smaller time.

In this work, we present an intuitive explanation of the phase transition behaviour of the bounds on the entanglement of the admissible ansatz. However, we point out that the phase transition point is problem-specific. It would be interesting to see if the phase transition behaviour still holds for other classes of algorithms, a problem we leave for future research.

Furthermore, using a similar approach, presented in our paper, one can rigorously investigate the RL-ansatz in the task of minimising the entanglement entropy of a quantum many-body system to identify quantum phase transition (An et al. 2021) and finding the ground state of chemical hamiltonian (Patel et al. 2024; Ostaszewski et al. 2021).

Appendix A: Preliminaries

We assume that every quantum system say S has an associated Hilbert space $\mathscr{H}_S$ of finite dimension d_S. The state of a quantum system S is represented by a density operator ρ_S defined on $\mathscr{H}_S$ and satisfies the conditions (a) Positivity: $\rho_S \geq 0$, (b) Hermitian: $\rho_S = \rho_S^\dagger$, and (c) Trace Preserving: $\text{Tr}[\rho_S] = 1$. We denote the set of density operators of S as $\mathscr{D}(\mathscr{H}_S)$. A pure state is a rank-one density operator given by $\rho_S := |\psi\rangle\langle\psi|_A$, where $|\psi\rangle_S \in \mathscr{H}_S$ is a unit-norm vector in $\mathscr{H}_S$. We denote the density operator of a composite sys-

tem $S_1 S_2$ as $\rho_{S_1 S_2} \in \mathscr{H}_{S_1 S_2}$ with $\text{Tr}_{S_2} \rho_{S_1 S_2} = \rho_{S_1} \in \mathscr{H}_{S_1}$ being the reduced state of S_1. The state $\rho_{S_1 S_2} \in \mathscr{H}_{S_1 S_2}$ is called separable if it cannot be expressed in the form of $\rho_{S_1 S_2} = \sum_x p_x \rho_{S_1}^x \otimes \rho_{S_2}^x$, where $\{\rho_{S_1}^x\}_x$ and $\{\rho_{S_2}^x\}_x$ are sets of pure states, $p_x \in [0, 1]$ and $\sum_x p_x = 1$. The states that cannot be expressed in the above form are said to be entangled (Horodecki et al. 2009). Next, we briefly discuss a relevant measure of entanglement for this work.

Concurrence

We quantify the amount of entanglement present in a bipartite quantum state $\rho_{AB} \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$ using concurrence (Hill and Wootters 1997; Wootters 1998) which is defined as

$$C^{\rho_{AB}} = \max\{0, \lambda_1^e - \lambda_2^e - \lambda_3^e - \lambda_4^e\} \quad (7)$$

where λ_s^e are the square root of the eigenvalues of $\rho_{AB}\tilde{\rho}_{AB}$ in descending order with $\tilde{\rho}_{AB} = (\sigma_y \otimes \sigma_y)\rho_{AB}^*(\sigma_y \otimes \sigma_y)$ being the spin-flipped state of ρ_{AB} and σ_y being a Pauli spin matrix. For the state $\rho_{AB} = |\psi\psi|_{AB}$, the concurrence is defined as

$$\mathcal{C}^{|\psi\rangle_{AB}} = \sqrt{2[1 - \text{Tr}\rho_A^2]}, \quad (8)$$

where $\rho_A = \text{Tr}_B \rho_{AB}$ is a subsystem of the combined state. It is easy to see that the concurrence of a pure bipartite qubit maximally entangled state is one, while that of a bipartite qubit separable state is zero. In the following subsection, we discuss a measure to quantify the contribution of individual qubits in a quantum circuit.

Conditional quantum entropy

We quantify the contribution of each qubit in a quantum circuit using the conditional quantum entropy (Cerf and Adami 1997; Wootters 1998), which is defined as follows:

Definition 1 [Conditional quantum entropy (Wilde 2013)] For the state $\rho'_{AB} \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$, the conditional quantum entropy $S_{A|B}^{\rho'_{AB}}$ is equal to the difference of the joint quantum entropy $S_{AB}^{\rho'_{AB}}$ and the marginal entropy $S_B^{\rho'_{AB}}$:

$$S_{A|B}^{\rho'_{AB}} := S_{AB}^{\rho'_{AB}} - S_B^{\rho'_{AB}}. \quad (9)$$

Operationally, the conditional entropy $S_{A|B}^{\rho'_{AB}}$ of the state for the state ρ'_{AB} signifies how mixed the marginal state ρ'_A (ρ'_B) is as compared ρ'_{AB}. For more details of the following theorems and observations, please refer (Wilde 2013).

Observation 1 *For the state $\rho'_{AB} \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$, consider a purification $|\psi\rangle_{ABE} \in \mathscr{H}_A \otimes \mathscr{H}_B \otimes \mathscr{H}_E$. It then follows that*

$$S_{A|B}^{\rho'_{AB}} = S_E^{\psi} - S_B^{\psi}. \quad (10)$$

The conditional quantum information thus measures the difference in the entropy of the states $\rho_E \in \mathscr{D}(\mathscr{H}_E)$ and $\rho_B \in \mathscr{D}(\mathscr{H}_B)$.

In the following subsection, we discuss a measure to quantify the amount of correlation between two random variables.

Correlation coefficient

We quantify the amount of correlation between two random variables r_1 and r_2 using the Pearson correlation coefficient (PCC) (Rodgers and Nicewander 1988) defined as

$$\text{PCC}^{r_1 r_2} = \frac{\text{E}[(r_1 - \mu_{r_1})(r_2 - \mu_{r_2})]}{\sigma_{r_1}\sigma_{r_2}}, \quad (11)$$

where μ_{r_j}, σ_{r_j} are the mean and standard deviation of the distribution r_j with $j \in \{1, 2\}$ and E[.] is an operator that returns the expectation value of a random variable. Note that the Pearson correlation coefficient (PCC) is symmetric: $\text{PCC}^{r_1 r_2} = \text{PCC}^{r_2 r_1}$. The value of PCC ranges in between -1 and $+1$. A correlation value of $+1(-1)$ implies perfect correlation (anti-correlation) between the two random variables r_1 and r_2, while zero implies no linear dependence between the variables.

Sampling a random quantum state

If n is the number of qubits and s is the random seed to sample, then we can sample a state from `random_density_matrix` of IBM qiskit's `quantum_info` module as follows:

```
random_density_matrix(2**n, seed=s).  (12)
```

The state used in 5.1.3 is sampled using $n = 2$ and $s = 27$.

References

Harrow AW, Montanaro A (2017) Quantum computational supremacy. Nat 549:203–209

Montanaro A (2016) "Quantum algorithms: an overview". npj Quantum Inf 2:1–8

Sadhu A, Somayajula MA, Horodecki K, Das S (2023) "Practical limitations on robustness and scalability of quantum internet". arXiv preprint arXiv:2308.12739

Zhang SX, Hsieh CY, Zhang S, Yao H (2022) Differentiable quantum architecture search. Quantum Sci Technol 7:045023

Du Y, Huang T, You S, Hsieh MH, Tao D (2022) "Quantum circuit architecture search for variational quantum algorithms". npj Quantum Inf 8:62

Kuo EJ, Fang YLL, Chen SYC (2021) "Quantum architecture search via deep reinforcement learning," arXiv preprint arXiv:2104.07715

Zhang SX, Hsieh CY, Zhang S, Yao H (2021) Neural predictor based quantum architecture search. Mach Learn: Sci Technol 2

McClean JR, Romero J, Babbush R, Aspuru-Guzik A (2016) The theory of variational hybrid quantum-classical algorithms. New J Phys 18:023023

Cerezo M, Arrasmith A, Babbush R, Benjamin SC, Endo S, Fujii K, McClean JR, Mitarai K, Yuan X, Cincio L et al (2021) Variational quantum algorithms. Nat Rev Phys 3:625–644

Sarkar A (2024) Automated quantum software engineering. Autom Softw Eng 31:1–17

Ostaszewski M, Trenkwalder LM, Masarczyk W, Scerri E, Dunjko V (2021) Reinforcement learning for optimization of variational quantum circuit architectures. Adv Neural Inf Process Syst 34:18182–18194

Kundu A, Bedelek P, Ostaszewski M, Danaci O, Patel YJ, Dunjko V, Miszczak JA (2024) Enhancing variational quantum state diagonalization using reinforcement learning techniques. New J Phys 26:013034

Patel YJ, Kundu A, Ostaszewski M, Bonet-Monroig X, Dunjko V, Danaci O (2024) "Curriculum reinforcement learning for quantum architecture search under hardware errors," arXiv preprint arXiv:2402.03500 (2024)

Sutton RS, Barto AG (2018) Reinforcement learning: an introduction (MIT press, 2018)

Wu W, Yan G, Lu X, Pan K, Yan J (2023) "Quantumdarts: differentiable quantum architecture search for variational quantum algorithms". in International Conference on Machine Learning (PMLR, 2023). pp 37745–37764

Kundu A (2024) "Reinforcement learning-assisted quantum architecture search for variational quantum algorithms". arXiv preprint arXiv:2402.13754

LaRose R, Tikku A, O'Neel-Judy E, Cincio L, Coles PJ (2019) "Variational quantum state diagonalization". npj Quantum Inf 5:57

Tan KC, Volkoff T (2021) Variational quantum algorithms to estimate rank, quantum entropies, fidelity, and fisher information via purity minimization. Phys Rev Res 3:033251

Cerezo M, Poremba A, Cincio L, Coles PJ (2020) Variational quantum fidelity estimation. Quantum 4:248

Kundu A, Miszczak JA (2022) Variational certification of quantum devices. Quantum Sci Technol 7:045017

Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow JM, Gambetta JM (2017) Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. Nat 549:242–246

Mustafa H, Morapakula SN, Jain P, Ganguly S (2022) "Variational quantum algorithms for chemical simulation and drug discovery," in 2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT) (IEEE, 2022) pp 1–8

Delgado A, Arrazola JM, Jahangiri S, Niu Z, Izaac J, Roberts C, Killoran N (2021) Variational quantum algorithm for molecular geometry optimization. Phys Rev A 104:052402

Khairy S, Shaydulin R, Cincio L, Alexeev Y, Balaprakash P "Learning to optimize variational quantum circuits to solve combinatorial problems". in Proceedings of the AAAI conference on artificial intelligence, Vol. 34. pp 2367–2375

Liu X, Angone A, Shaydulin R, Safro I, Alexeev Y, Cincio L (2022) Layer VQE: a variational approach for combinatorial optimization on noisy quantum computers. IEEE Trans Quantum Eng 3:1–20

Glos A, Krawiec A, Zimborás Z (2022) "Space-efficient binary optimization for variational quantum computing," npj Quantum Inf 8:39

Wilde MM (2013) Quantum information theory (Cambridge university press, 2013)

Lloyd S, Mohseni M, Rebentrost P (2014) Quantum principal component analysis. Nat Phys 10:631–633

Powell MJD (1994) "A direct search optimization method that models the objective and constraint functions by linear interpolation". in Advances in Optimization and Numerical Analysis. Mathematics and Its Applications, Vol. 275, (ed) by Gomez S, Hennart JP. (Springer,1994)

Powell MJD (2006) "A fast algorithm for nonlinearly constrained optimization calculations," in Numerical Analysis: Proceedings of the Biennial Conference Held at Dundee, June 28–July 1, 1977 (Springer, 2006) pp 144–157

Hasselt HV, Guez A, Silver D (2016) "Deep reinforcement learning with double q-learning". in Proceedings of the AAAI conference on artificial intelligence, Vol. 30

Alagic G, Majenz C, Russell A (2020) "Efficient simulation of random states and random unitaries," in Advances in Cryptology– EUROCRYPT 2020: 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10–14, 2020, Proceedings, Part III 39 (Springer, 2020) pp 759–787

Chen B, Coladangelo A, Sattath O (2024) "The power of a single haar random state: constructing and separating quantum pseudorandomness," arXiv preprint arXiv:2404.03295

Kingma DP, Ba J (2014) "Adam: a method for stochastic optimization," arXiv preprint arXiv:1412.6980

Hill SA, Wootters WK (1997) Entanglement of a pair of quantum bits. Phys Rev Lett 78:5022–5025

Wootters WK (1998) Entanglement of formation of an arbitrary state of two qubits. Phys Rev Lett 80:2245–2248

Rodgers JL, Nicewander WA (1988) Thirteen ways to look at the correlation coefficient. Am Stat 42:59–66

Cerf NJ, Adami C (1997) Negative entropy and information in quantum mechanics. Phys Rev Lett 79:5194–5197

Cerf NJ, Adami C (1999) Quantum extension of conditional probability. Phys Rev A 60:893–897

Horodecki R, Horodecki P, Horodecki M, Horodecki K (2009) Quantum entanglement. Rev Mod Phys 81:865–942

Brunner N, Cavalcanti D, Pironio S, Scarani V, Wehner S (2014) Bell nonlocality. Revi Mod Phys 86:419

Sadhu A, Das S (2023) Testing of quantum nonlocal correlations under constrained free will and imperfect detectors. Phys Rev A 107:012212

Uola R, Costa ACS, Nguyen HC, Gühne O (2020) Quantum steering. Rev Mod Phys 92:015001

Sadhu A, Das S (2024) "Quantum steering under constrained free-will". arXiv preprint arXiv:2406.13494

An ,Z Cao C, Xu CQ, Zhou DL (2021) "Learning quantum phases via single-qubit disentanglement". arXiv preprint arXiv:2107.03542

Authors and Affiliations

Abhishek Sadhu[1,2] · Aritra Sarkar[3] · Akash Kundu[4,5,6]

✉ Akash Kundu
   akash.kundu@helsinki.fi

   Abhishek Sadhu
   abhisheks@rri.res.in

   Aritra Sarkar
   a.sarkar-3@tudelft.nl

[1] Raman Research Institute, Bengaluru, India

[2] Centre for Quantum Science and Technology (CQST), International Institute of Information Technology, Hyderabad, Telangana, India

[3] QuTech, Delft University of Technology, Delft, The Netherlands

[4] Joint Doctoral School, Silesian University of Technology, Gliwice, Poland

[5] Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Gliwice, Poland

[6] QTF Centre of Excellence, Department of Physics, University of Helsinki, Helsinki, Finland