



Skolkovo Institute of Science and Technology

Skolkovo Institute of Science and Technology

ON APPLICATIONS OF VARIATIONAL QUANTUM CIRCUITS

Doctoral Thesis

by

ANDREY KARDASHIN

DOCTORAL PROGRAM IN COMPUTATIONAL AND DATA
SCIENCE AND ENGINEERING

Supervisor

Professor, Jacob Daniel Biamonte, PhD, DSc

Moscow — 2023

© Andrey Kardashin, 2023

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate: Andrey Kardashin

Supervisor: Prof. Jacob Daniel Biamonte

Abstract

Quantum algorithms promise significant computational speed-up compared to their classical counterparts. For these algorithms to be practical, however, large-scale fault-tolerant quantum computers are required. An important step towards such hardware is noisy intermediate-scale quantum (NISQ) devices, which are already available today. NISQ computers are of especial usefulness in executing algorithms implemented in the so-called variational quantum computing approach. In such algorithms, a quantum computer is used to prepare a quantum state, which is obtained as some easy-to-prepare initial state acted by a parametrized unitary operator. Commonly, these operators are represented by sequences of quantum gates, which may depend on several parameters. Such unitary operators are called variational quantum circuits, as their parameters can be varied for obtaining a desired quantum state. This thesis studies applications of such circuits in different fields, including condensed matter physics and machine learning. Inter alia, described is the mathematical apparatus of quantum computing, and the variational model of quantum computations is introduced. Additionally in the thesis, several existing variational quantum algorithms are reviewed and formulated in the tensor networks notation.

Publications

Main author

1. A. Kardashin, A. Vlasova, A. Pervishko, D. Yudin, and J. Biamonte. Quantum-machine-learning channel discrimination. *Physical Review A*, 106:032409, Sep 2022
2. A. Kardashin, A. Pervishko, J. Biamonte, and D. Yudin. Numerical hardware-efficient variational quantum simulation of a soliton solution. *Physical Review A*, 104(2):L020402, 2021
3. A. Kardashin, A. Uvarov, and J. Biamonte. Quantum machine learning tensor network states. *Frontiers in Physics*, 8:644, 2021
4. A. Kardashin, A. Uvarov, D. Yudin, and J. Biamonte. Certified variational quantum algorithms for eigenstate preparation. *Physical Review A*, 102(5):052610, 2020

Co-author

1. O. Borzenkova, G. Struchalin, A. Kardashin, V. Krasnikov, N. Skryabin, S. Straupe, S. Kulik, and J. Biamonte. Variational simulation of Schwinger's Hamiltonian with polarization qubits. *Applied Physics Letters*, 118(14):144002, 2021
2. A. Uvarov, A. Kardashin, and J. Biamonte. Machine learning phase transitions with a quantum processor. *Physical Review A*, 102(1):012415, 2020

Contents

List of symbols and abbreviations	7
List of figures	8
Introduction	16
1 Theoretical background	24
1.1 Hilbert space and Dirac notation	24
1.2 Introduction to quantum computing	27
1.2.1 Quantum states	27
1.2.2 Transformation of quantum states	31
1.2.3 Entanglement	32
1.2.4 Observables and measurement	34
1.3 Tensor diagrams	36
1.3.1 Tensor network states	40
1.4 Quantum circuits	43
1.4.1 Quantum gates	45
1.4.2 Parametrization of a quantum state	48
1.5 Hamiltonian simulation	49
2 Variational quantum computing	52
2.1 Variational quantum eigensolver	53
2.1.1 Experimental realizations of VQE	56
2.1.2 Soliton solution: a case study	61

3	Variational algorithms and tensor diagrams	69
3.1	State preparation verification for VQE	70
3.1.1	Variational Hamiltonian diagonalization	73
3.2	Variational state preparation	78
3.3	Variational quantum compiling	81
3.4	Variational linear system solver	82
3.5	Variational quantum state diagonalization	85
3.5.1	Alternative variational state diagonalization	88
3.5.2	Alternative variational Hamiltonian diagonalization	90
3.6	Remarks on variational quantum algorithms	91
4	Quantum channel discrimination with variational quantum circuits	93
4.1	Variational quantum channel discrimination	93
4.1.1	Discrimination strategies: parallel and sequential	95
4.1.2	Variational formulation of the discrimination strategies	99
4.1.3	Numerical experiments	103
4.1.4	Remarks on quantum channel discrimination	110
4.2	Variational quantum channel classification	112
4.2.1	Numerical experiments	115
4.2.2	Remarks on quantum machine learning	121
	Conclusion	124
	Bibliography	128
A	Cost functions of the variational quantum state diagonalization algorithm	138
B	Swap test	141
B.1	Classic swap test	141
B.2	Destructive swap test	144
C	Hadamard test	146
C.1	Modified Hadamard test	147

List of symbols and abbreviations

\mathbb{B}	Boolean set, $\{0, 1\}$
\mathbb{N}	Field of natural numbers
\mathbb{R}	Field of real numbers
\mathbb{C}	Field of complex numbers
\mathcal{H}	Hilbert space
$\dim(\mathcal{H})$	Dimensionality of a Hilbert space \mathcal{H}
$\mathbb{1}$	Identity operator
A^*	Complex conjugate of an operator
A^\dagger	Hermitian conjugate (conjugate transpose) of an operator
$\ A\ _p$	Operator p -norm of a linear operator
$ \psi\rangle$	A vector from a Hilbert space, unit norm pure quantum state
$\mathcal{L}(\mathcal{H})$	The set of linear operators acting in \mathcal{H}
$\text{Herm}(\mathcal{H})$	The set of Hermitian operators acting in \mathcal{H}
$\text{U}(\mathcal{H})$	The set of unitary operators acting in \mathcal{H}
$\mathcal{D}(\mathcal{H})$	The set of density operators acting in \mathcal{H}
$\sigma_x, \sigma_y, \sigma_z$	Pauli operators
X, Y, Z	Pauli gates
MPS	Matrix product state
NISQ	Noise Intermediate-scale Quantum
TFIM	Transverse field Ising model
VQA	Variational Quantum Algorithm
VQE	Variational quantum eigensolver
HEA	Hardware-efficient ansatz
BFGS	Broyden-Fletcher-Goldfarb-Shanno algorithm

List of figures

- 1-1 A four-qubit example of the checkerboard ansatz of four layers. On the right shown is also the two-qubit operator of five parameters which is put into a block. In it, $R_\sigma(\theta) = e^{-i\theta\sigma}$ and $R_{\sigma\tau}(\theta) = e^{-i\theta\sigma\otimes\tau}$ with $\sigma, \tau \in \{\sigma_x, \sigma_y, \sigma_z\}$ specifying the Pauli operator and $\theta \in [0, 2\pi)$ being the optimization parameters. 49
- 2-1 A layer of a hardware-efficient ansatz for 4 qubits with $p = 12$ variational parameters. Here, $R_\sigma(\theta) = e^{-i\theta\sigma}$ with $\sigma \in \{\sigma_x, \sigma_y, \sigma_z\}$ specifying the Pauli operator and $\theta_j \in [0, 2\pi)$ being the optimization parameters. The controlled-X operators “flip” the target qubit depending on the state of the control qubit, see (1.31). 57
- 2-2 Eigenvalue (a) and order parameter (b) versus bare mass m of Shwinger’s Hamiltonian (2.10). Solid lines stand for the analytical solution, cyan points (violin plot) show the results of simulations, and red points are obtained from the experiment. For each m , the data for simulations are obtained as the average over 30 trials. For the experiment, all attempts are shown as distinct points. The figures are taken from [1]. 60

2-3	<p>Noise simulations for dephasing noise (2.13) acted on one qubit (a), (c) and both qubits (b), (d). Panels (a) and (b) show the dependence of the ground state energy of Schwinger’s Hamiltonian (2.10) on the bare mass m, and panels (c) and (d) indicate the dependence of the order parameter (2.12). Red lines correspond to simulations without noise, and the color blur corresponds to the increase of noise strength ϵ from 0.1 to 1 in 0.1 steps. Distinct points stand for experimental results, and solid areas show theoretical predictions for points with different noise level. The plots are taken from [1].</p>	61
2-4	<p>Ground state energy (a) and the overlap of the approximate ground state vector with the exact one (b) of the Hamiltonian (2.16) for $N = 10$ found by VQE. In panel (a), the dashed green line stands for the exact ground state energy. Each data point is obtained as an average over 5 optimizations initialized with random parameters of the ansatz. The error bars show the standard deviation. The figures are taken from [2].</p>	65
2-5	<p>Concurrence C_{ij} (2.24) between the ith and jth qubits of the ground state of the DMI Hamiltonian (2.16). Panels (a) and (b) show the concurrence calculated for the VQE solution C_{ij}^{VQE} obtained with six ansatz layers and the exact ground state C_{ij}^{exact}, respectively. Panel (c) gives the ratio $C_{ij}^{\text{VQE}}/C_{ij}^{\text{exact}}$. As can be seen, the more distant two spins are, the worse the entanglement is captured by VQE. The figures are taken from [2].</p>	67
2-6	<p>Magnetic texture of the ground state of the Hamiltonian (2.16) of $N = 10$ spin one-half particles versus the number of layers of the hardware-efficient ansatz. Each line corresponds to the magnetic moments of spins in the ground state found by VQE with a specific number of layers of the ansatz, and each arrow shows the magnetic moment of a given spin in the xy plane (z components are equal to zero). The uppermost line of magnetic moments is given by the analytical solution 2.20. The figure is taken from [2].</p>	68

3-1	Reduction of the term $ \langle\psi Q \psi\rangle ^2$ from (3.2) to the VQE cost function of the form $\langle\Psi M \Psi\rangle$, where M and $ \Psi\rangle$ are defined in (3.3) and (3.4), respectively.	72
3-2	The overlap gain (left) and convergence rate (right) for the transverse field Ising model Hamiltonian (a) and the Schwinger model Hamiltonian (b) during the optimization of the functions F_H and F_Q defined in (3.1) and (3.2), respectively. For each Hamiltonian, 300 random sets of initial parameters for the hardware-efficient ansatz were generated. The figures are taken from [3].	74
3-3	Derivations used to reduce the cost function (3.8) to VQE. The vector $ \Psi\rangle$ and the operator S are defined in (3.9) and (3.11), respectively. .	75
3-4	Numerical experiments conducted for diagonalizing random 5-qubit Hamiltonians with the method described by (3.8). The unitary U was parametrized by HEA of different number of layers l . On panel (a) depicted is the error of diagonalization, the sum of the squared absolute values of the off-diagonal entries of the diagonalized matrix. Each data point is obtained as an average over ten random Hamiltonians; the shaded area indicates the standard deviation. On panel (b) plotted is the gain (3.7) of the cost function in (3.8). The lines of different colors stand for different number of layers l in the ansatz. .	77
3-5	Reduction of the 2-norm distance (3.16) to the cost function of VQE for the Hamiltonian $M = (Z \otimes \mathbb{1} \otimes \mathbb{1}) \left(\Phi^+\rangle\langle\Phi^+ \otimes \mathbb{1} \right) (Z \otimes \mathbb{1} \otimes \mathbb{1})$ and the variational state $ \Psi\rangle = CV_{23}LU_{23} \Phi^+\rangle \otimes 0\rangle$	80
3-6	Comparison of the gains (3.7) of the functions (3.15) and (3.16). The number of qubits was set to $n = 6$, the target vector $ \alpha\rangle$ was generated randomly, and the variational state $ \psi\rangle$ was prepared by the hardware-efficient ansatz given in Fig. 2-1 with $l = 6$ layers, the gain was obtained as the average over ten independent runs, and the optimization algorithm was BFGS.	81
3-7	Reduction of the term $ (V, U)_{\text{HS}} ^2$ from (3.17) to the VQE cost function of the form $\langle\Psi M \Psi\rangle$ with $M = \Phi^+\rangle\langle\Phi^+ $ and $ \Psi\rangle = (U \otimes \bar{V}) \Phi^+\rangle$. 83	83

3-8	Reduction of the term $ \langle b A x\rangle ^2$ from (3.18) to the VQE cost function of the form $\langle x M x\rangle$	85
3-9	Reduction of a tensor network which computes $\text{Tr } \rho^2$ to a tensor network which computes $(1 + \text{Tr } \rho^2)/2$. The obtained tensor network is equivalent to the VQE for the Hamiltonian H and state $ \Psi\rangle$ defined in (3.24) and (3.25), respectively.	88
3-10	Tensor network formulation and reduction to VQE for $\text{Tr } D[\rho_U]^2$. The Hamiltonian and the variational state for VQE are given in (3.26) and (3.27), respectively.	89
4-1	A schematic of the parallel strategy for channel discrimination with p allowed channel applications. First, Alice takes two registers, P of p qubits and R of r qubits, and prepares a composite $(p + r)$ -qubit state ρ_{PR}^{in} . Then the qubits of the register P are sent to Bob, who applies the channel Φ_y on each qubit, and sends the qubits back to Alice; meanwhile, the register R remains unaffected. Finally, Alice measures all qubits of the output state with the POVM Π . The figure is taken from [4].	96
4-2	A schematic of the sequential channel discrimination strategy. First, Alice prepares a $(1 + r)$ -qubit state ρ_{PR}^{in} of one qubit of the register P and r qubits of the register R . The qubit of P is sent to Bob who applies a channel Φ_y , while the register R stays with Alice. The output state is then sent from Bob back to Alice who modifies it with a channel \mathcal{E}_j and sends it back to Bob. Having repeated this procedure p times, Alice measures all the $(1 + r)$ qubits a POVM Π . The figure is taken from [4].	98

- 4-3 Parallel quantum channel discrimination strategy implemented as a variational circuit. The circuit $U(\boldsymbol{\theta}_0)$ prepares the input state of Alice ρ_{PR}^{in} from the initial state $|0\rangle_P^{\otimes p} \otimes |0\rangle_R^{\otimes r}$. After this transformation, Bob applies a channel Φ_y to the qubits of the register R , and the output state is sent to Alice to be measured. To rotate the measurement basis, Alice uses the circuit $U(\boldsymbol{\theta}_1)$. The figure is taken from [4]. 101
- 4-4 Sequential quantum channel discrimination strategy A variational quantum circuit implementing the sequential channel discrimination strategy from Figure 4-2. The Alice’s channels $\mathcal{E} = \{\varepsilon_j\}_{j=1}^{p-1}$ are replaced by the parametrized unitaries $\{U(\boldsymbol{\theta}_j)\}_{j=1}^{p-1}$, where p is the number of allowed applications of the channel Φ_y . The input state ρ_{PR}^{in} is prepared from $|0\rangle_P \otimes |0\rangle_R^{\otimes r}$ via the unitary transformation $U(\boldsymbol{\theta}_0)$, while $U(\boldsymbol{\theta}_p)$ is used to rotate the measurement basis. This method necessitates $(1+r)$ qubits, with one qubit in the register P and r qubits in the register R . In analogy with the parallel strategy, r might be set to zero. The figure is taken from [4]. 102
- 4-5 Explicit variational quantum circuits implementing (a) the parallel and (b) sequential strategies discriminating the entanglement breaking channels (4.15) with $p = 2$ applications. Since the considered channels map two-qubit states into one-qubit states, we need to add one qubit in some state (in our case, $|0\rangle$) after the first application of Φ_y 105
- 4-6 Results of numerical experiments with variational discrimination of the the depolarizing channels with parameters α_0 and α_1 . Panel (a) shows the achieved probabilities of successful discrimination p_s . In panel (b) shown are the ratios of the success probability p_s to the maximal achievable probability p_\diamond . The data on the panels on the left are obtained with $l = 1$ layer of the hardware-efficient ansatz while the right panels show the data for $l = 2$ layers. 107

- 4-7 Achieved probabilities of successful discrimination p_s between two depolarizing channels with depolarization factors α_0 and α_1 . Panel (a) shows the results obtained for the parallel strategy, while in panel (b) shown are the results for the sequential strategy. Different colors indicate different values of l , the number of layers in the hardware-efficient ansatz. The probabilities p_s averaged over 10 independent optimizations are shown as marks connected by solid lines, while standard deviations are shown as shaded areas. The maximum achievable probability for the parallel strategy p_s^{par} for $p = 2$ is indicated by the black solid line. The figures are taken from [4]. 109
- 4-8 Correlation coefficients between the average achieved success probabilities p_s (shown in Figure 4-7) and the trace of the product of the channel output states $\rho_y = \Phi_y^{\otimes 2}[\rho]$ with $y \in \{0, 1\}$ (blue circles), the trace distance between them (green triangles), and the diamond distance between the channels Φ_y (orange squares). Here, $\Phi_y \equiv \Phi(\alpha_y)$ and the input state was set to $\rho = |00\rangle\langle 00|$. Panels (a) and (b) show the correlations for the parallel and sequential strategies, respectively. The correlations are plotted versus l , the number of layers of the ansatz. Solid lines show the functions $f(l) = l^{-1/a}$ (blue and green) and $g(l) = 1 - e^{-bl}$ (orange), which fit the data points of the corresponding colors. The figures are taken from [4]. 110
- 4-9 A schematic of the variational quantum channel classifier. First, Bob prepares a random state ρ , passes it through a channel Φ_y , and sends the output state $\Phi_y[\rho]$ to Alice. Alice then applies the unitary $U(\boldsymbol{\theta})$, measures the resultant state in the computational basis, and computes the prediction value p defined in (4.23). The predictions p are then used for training the circuit, which is done by minimizing the square distances (4.24) between the predictions p and the true labels y 113

4-10 Accuracy of quantum channel classifier trained for distinguishing between the depolarizing channels with factors α_0 and α_1 . The classifier is based on the ansatz (4.25) and is trained and tested on the data sets of the sizes $N_{\text{train}} = N_{\text{test}} = 1000$. The figure is taken from [4]. 116

4-11 Classification accuracy for the depolarizing channels with α_0 and α_1 versus the size of the training set N_{train} . Panels (a) and (b) show the accuracy obtained for the training and test sets, respectively. The size of the test set was fixed to $N_{\text{test}} = 1000$. The color of the curves indicate the values of the coefficients α_0 and α_1 (see legends), and shaded areas show the standard deviation computed over 100 initial circuit parameters. 116

4-12 A schematic of the modified variational classifier of quantum channels. First, Bob generates two copies of a random state ρ , and passes one of them through a channel Φ_y , outputting the state $\Phi_y[\rho] \otimes \rho$. This state is then sent to Alice who applies the unitary $U(\boldsymbol{\theta})$, measures the resultant state in the computational basis, and computes prediction p defined in (4.27). The circuit is trained by minimizing the square distances (4.24) between the predictions p and the true labels y 117

4-13 Accuracy of the classification of the depolarizing channels with α_0 and α_1 . Panels (a) and (b) show the accuracy of the classifiers built on the circuits U_1 and U_2 defined in (4.28) and (4.29), respectively. The classifiers are trained and tested on the sets of the sizes $N_{\text{train}} = N_{\text{test}} = 1000$. The figures are taken from [4]. 119

4-14 Accuracy of the classification of the three-qubit depolarizing channels with $\alpha_0 = 0$ and $\alpha_1 > \alpha_0$. Panels (a) and (b) show the results for the classifiers consisting of one and two layers of the hardware-efficient ansatz, respectively. The classifiers were trained on the sets of the size $N_{\text{train}} = 1000, 2000$ and 3000 , and tested on a set of the size $N_{\text{test}} = 1000$. The shaded areas indicate the standard deviation obtained over 10 initial circuit parameters. 120

4-15 Training (a) and test (b) accuracy of the classification of random four-qubit channels. The blue and the red points indicate the label 0 and label 1, respectively. The channels were obtained via the Stinespring channel representation (4.12) by generating random 12-qubit unitary matrices. The classifier was trained on the sets of the size $N_{\text{train}} = 100$, and tested on a set of the size $N_{\text{test}} = 1000$ 121

C-1 Quantum circuits for calculating the real and imaginary parts of the complex number $\langle \psi | U | \psi \rangle$, where U is unitary. For the left circuit, the probability of obtaining the result $x \in \{0, 1\}$ after measuring the auxiliary qubit is $p_x^R = \frac{1}{2}(1 + (-1)^x \text{Re} \langle \psi | U | \psi \rangle)$; for the right circuit, this probability is $p_x^I = \frac{1}{2}(1 + (-1)^x \text{Im} \langle \psi | U | \psi \rangle)$ 147

C-2 Quantum circuits for calculating the real and imaginary parts of the complex number $\langle 0 | A^\dagger B | 0 \rangle$, where A and B are unitary. For the left circuit, the probability of obtaining the result $x \in \{0, 1\}$ after measuring the auxiliary qubit is $p_x^R = \frac{1}{2}(1 + (-1)^x \text{Re} \langle 0 | A^\dagger B | 0 \rangle)$; for the right circuit, this probability is $p_x^I = \frac{1}{2}(1 + (-1)^x \text{Im} \langle 0 | A^\dagger B | 0 \rangle)$. . . 148

Introduction

Relevance of the work

The development of quantum physics and quantum technologies contributed to the fast growth in information technology in the XX century in terms of creating hardware, while the principles of representing and processing information, software, remained classical. At the end of the XX century, proposed were fundamentally new approaches based on the use of more subtle effects of quantum physics, which can be called quantum information processing technologies. For the information representation and processing, these approaches use phenomena and properties inherent exclusively in quantum systems, such as quantum entanglement, quantum parallelism, the Heisenberg uncertainty principle, and the impossibility of cloning quantum states.

A great attention to quantum computation methods was drawn after Peter Shor published the article describing a quantum algorithm for factorization of an integer number into a product of prime numbers [5]. The sensation of this algorithm was that if it would be possible to execute it on a quantum computer, then it would be also possible to crack public-key cryptographic systems commonly used, for example, in banking or on the Internet. The basic principle of such cryptographic systems assumes the presence of a hard computational problem for which a fast classical algorithm is unknown. For example, the popular RSA [6] system implements the public key as the product of two large prime numbers. In order to break the RSA cipher, one needs to find the factors of this product. If this number is sufficiently large, then there is no known algorithm which is able to factorize it in reasonable time. That is, the best classical algorithm does it in exponential time [7], while

Shor's method is polynomial.

As stated earlier, the very idea of quantum computation is to utilize certain phenomena of quantum physics as a resource for computation. For instance, the famous Grover's search algorithm [8] is based on the fact that a number of quantum bits can be put into the uniform superposition of zeros and ones, which is impossible for classical bits as their state is definitive. Many of quantum cryptography protocols [9] make use of the non-cloning theorem [10] which states that there is no operation allowing to copy an unknown quantum state. Other quantum cryptography protocols [11] are based on quantum entanglement, a phenomenon which allows for several quantum systems to be prepared in a state such that they cannot be described separately [12, 13]. Entanglement is also used as a resource for some quantum algorithms, including quantum teleportation [14] and superdense coding [15].

Many other useful quantum algorithms are known today. Amongst them is e.g. the quantum phase estimation algorithm [16] which serves as a subroutine for other algorithms, such as quantum algorithm for solving linear systems of equations [17], singular value decomposition [18, 19], and the celebrated Shor's factorization algorithm [5]. Of great practical relevance are also quantum algorithms for semidefinite [20, 21] and dynamical [22] programming, machine learning [23–25], solving differential equations [26–28], and other [29].

The main peculiarity of quantum computations is that the development of theory is currently far ahead of that of technology. The principal obstacle to the creation of a full-scale quantum computer is errors caused by external noise, or decoherence, which destroys the superposition and entanglement of quantum bits, the basic building blocks in quantum computing. Currently, great efforts are put to the development of quantum error correction methods [30, 31].

Another way around the erroneous quantum computation is the development of quantum algorithms suitable for the so-called noisy intermediate-scale quantum (NISQ) computers [32, 33]. The core of such algorithms is formed by the idea of variational quantum computing [34–41]. The algorithms developed in this setting are called variational quantum algorithms (VQA). In these algorithms, one encodes

a problem of interest into a real-valued function which assumes as an argument a parametrized quantum state called the variational state. This function — called the cost, loss, or objective function — is constructed in such a way that the argument, which delivers the minimum to the function, encodes the solution to the original problem. While executing a VQA, a quantum processor computes the cost function, while a classical processor searches for optimal assignments for the parameters of the variational state. That is, the routines of the optimization of a cost function are distributed between a classical and a quantum computer. Due to this, variational quantum algorithms are sometimes called hybrid quantum-classical.

One of the pioneering VQAs was the variational quantum eigensolver (VQE) algorithm [37]. In simple words, this algorithm uses a quantum device for naturally finding a solution to the Schrödinger equation. Namely, given a Hamiltonian, the algorithm finds its ground state energy and vector. The cost function of this algorithm is essentially the expected value of the Hamiltonian in the variational state. In physics, a Hamiltonian describes a quantum system in terms of energy, which is an observable quantity. That is, a quantum computer prepares the variational state and measures its energy according to a given Hamiltonian, while a classical computer finds parameters for the variational state such that the energy of the system is minimized. This process can be considered as a direct application of the Rayleigh–Ritz principle [42]. In the seminal work, the VQE was executed on an optical device which encodes quantum bits into states of photons, and the target Hamiltonian was describing the bond-dissociation energy of the molecule H-He⁺. It is also noteworthy that measuring an observable with respect to a quantum state is a very basic operation naturally performed by quantum computers, therefore VQAs are based on computing the expected value of an observable.

To describe quantum algorithms, graphical methods for representing equations are often used. The most prominent of such methods is the notation of quantum circuits, which is inspired by the visualization methods for electrical circuit design and tensor networks in physics [43–45]. In a quantum circuit, one indicates the initial quantum state, its transformation in terms of sequences of unitary operators called quantum gates, and the measurement of the resultant state. If a circuit describes a

variational quantum algorithm, then the gates comprising the circuit are indicated to depend on variational parameters. Such circuits are called variational quantum circuits.

Although the quantum circuits notation allows to make pictorial and explanatory representations of quantum algorithms, there is a more general and powerful language called tensor diagrams. Aside from physics and quantum computing, tensor networks methods are also prominent in classical computer science [46, 47] and machine learning [48–50]. So, formulated in terms of tensor networks, variational quantum algorithms could attract more attention of researches from the communities of computer science, computational physics, machine learning etc.

Thesis goals

This thesis aims to study applications of variational quantum circuits. The objectives of the thesis also include the description of variational quantum algorithms, cost functions they are based on, areas of possible applications, obstacles and peculiarities of their execution, and ways of their representation in terms of quantum circuits and tensor diagrams. The goals are achieved by performing the following tasks.

1. Describe an experiment of executing the variational quantum eigensolver (VQE) algorithm on an optical setup. As a target Hamiltonian, choose one that models a real physical system. Determine special features and arising problems characteristic for the chosen hardware and Hamiltonian. Additionally, investigate the performance of VQE in the presence of controlled noise.
2. Perform numerical experiments for the VQE algorithm executed for a Hamiltonian modeling a system the ground state of which is highly entangled. Investigate the performance of the algorithm for this case, make conclusions about abilities to extract properties of interest from the approximated ground state vector.
3. Review several existing variational quantum algorithms and formulate them using the tensor diagrams notation. By leveraging tensor network transforma-

tion rules, show that these variational algorithms can be reduced to the VQE algorithm.

4. Explore possibilities of putting quantum information problems into the framework of variational quantum computing. As a case study, formulate the quantum channel discrimination problem as a variational algorithm and conduct numerical experiments of executing it.
5. Consider variational quantum computing methods as a means for solving classification machine learning tasks. As a case study, determine the properties of quantum classifiers trained for distinguishing between quantum channels.

Statements defended

The thesis defends the following claims.

1. The variational quantum eigensolver (VQE) algorithm can be implemented on an optical setup which encodes states of a qubit into the polarization of a photon. That is, the vertical polarization corresponds to the state $|0\rangle$ while the horizontal polarization stands for $|1\rangle$. In such a setting, one is able to experimentally find the ground state energy and vector for the two-qubit Schwinger Hamiltonian, which depends on several parameters. Even in the presence of noise one is still able to study physical properties of the approximated ground state.
2. VQE can be applied for studying Hamiltonians which describe spin chains with the Dzyaloshinskii-Moriya type interactions. The ground state of such Hamiltonians can be highly entangled, which makes it difficult for VQE to find a precise state vector. That is, even if the ground state energy is found with high precision, the closeness of the approximate state vector to the true one can be small. Nonetheless, one is still able to extract some physical properties of interest from the solution found.
3. Some of the existing variational quantum algorithms can be formulated in the language of tensor networks. Applying this tensor diagrams notation, one can

reduce these algorithms to VQE for a specific Hamiltonian and pure variational state. Among such algorithms are:

- variational quantum linear solver;
 - variational quantum state preparation algorithms;
 - variational quantum state diagonalization algorithm;
 - variational eigenstate verification algorithms.
4. Variational quantum computing approach can be utilized for solving optimization problems from the quantum information theory. Namely, by replacing the optimization over quantum states and measurement operators by the optimization over the parameters of variational circuits, one is able to solve these problems on a quantum computer. One of such problems, quantum channel discrimination, can be solved such that the fundamental theoretical bounds on the solution quality are achieved.
 5. Variational quantum circuits can be utilized for solving classification tasks for data represented by labeled quantum states. Particularly, one can train a variational circuit for distinguishing between quantum states acted by labeled quantum channels.

Scientific novelty

1. The VQE algorithm solves the ground state problem for a given Hamiltonian. We showed that even if the obtained solution is of low precision, one is still able to estimate physical properties of interest. Concretely, we showed that one can observe critical points of a physical model in the presence of noise, and retrieve from a low-fidelity ground state vector the magnetic texture of highly entangled spins.
2. Having found a solution using VQE, one may be interested in the closeness of the obtained solution to the true ground state. We proposed a new method for such a state certification, and compared it to the existing one.

3. Researchers of many fields of science use the tensor networks notation for representing equations, and formulating algorithms and methods. We formulated several existing variational quantum algorithms in the language of tensor networks. By using this notation, we also reduced the considered algorithms to VQE for specific Hamiltonians and variational states.
4. For the problem of binary quantum channel discrimination, theoretical works are typically dedicated to obtaining bounds on probabilities of successful distinguishing, but not telling how to achieve these bounds. We give a recipe for solving this problem using the notion of variational quantum circuits. This implies that if a problem from the quantum information theory is formulated as a task of optimization over quantum states and measurement operators, then it can be put into the framework of variational quantum computing.

Theoretical and practical significance

The results related to VQE are useful for physics of condensed matter, e.g. experimental study of physical properties of different model Hamiltonians. The tensor network formulations of variational algorithms are potentially useful in theoretical physics, classical computing and machine learning fields. Solving the binary quantum channel discrimination problem in the variational framework suggests the applicability of the devised technique on NISQ devices. The results on solving the channel discrimination problem could be important for quantum target detection.

Presentations and validation of the results

The main results of the work are based on articles published in peer-reviewed journals. Some of the results have been reported in the following scientific conferences:

1. International Conference on Quantum Technologies (July 12-16, 2021, Moscow, online, poster);
2. International Conference on Quantum Technologies (July 15-19, 2019, Moscow, poster).

The validity of the work is supported by numerical experiments and rigorous mathematical proofs, where applicable.

Chapter 1

Theoretical background

This chapter introduces the mathematical apparatus used throughout the thesis: main definitions, basics of quantum computing, quantum circuits, and tensor networks notation.

1.1 Hilbert space and Dirac notation

In quantum mechanics, it is postulated that the state of a quantum system is described by a density operator ρ acting in a Hilbert space \mathcal{H} . The vector space \mathcal{H} with a scalar (inner) product $\langle \psi, \varphi \rangle$ is called a Hilbert space if this space is complete with respect to the metric induced by the scalar product in this space, $\|\psi - \varphi\| = \sqrt{\langle \psi - \varphi, \psi - \varphi \rangle}$, $\psi, \varphi \in \mathcal{H}$. The completeness of a space means that any fundamental (Cauchy) sequence in this space converges to an element of this space. For complex Hilbert spaces, we assume that the scalar product is linear in the second argument and is anti-linear in the first argument:

$$\langle \psi + b\varphi, c\eta \rangle = \bar{a}c\langle \psi, \eta \rangle + \bar{b}c\langle \varphi, \eta \rangle,$$

where $a, b, c \in \mathbb{C}$; $\psi, \varphi, \eta \in \mathcal{H}$; and \bar{a} means the complex conjugate of a .

From here on — if not stated otherwise — we will work with finite-dimensional complex Hilbert spaces of dimensionality d , so commonly we work with $\mathcal{H} = \mathbb{C}^d$. In finite dimensions, an element $\psi \in \mathcal{H}$ is represented by a column-vector with the

components ψ_j . For the scalar product we then have

$$\langle \psi, \varphi \rangle = \sum_{j=1}^d \bar{\psi}_j \varphi_j = \psi^\dagger \varphi, \quad (1.1)$$

where $\psi^\dagger \equiv \bar{\psi}^T$ is a complex-conjugated row-vector, the Hermitian conjugate of ψ . The vectors ψ^\dagger are often called covectors to ψ .

Let us now introduce the Dirac (bra-ket) notation. From now on, we will denote vectors ψ and their covectors ψ^\dagger as

$$\begin{aligned} \psi &\equiv |\psi\rangle, \\ \psi^\dagger &\equiv \langle \psi|. \end{aligned}$$

This notation is very convenient and illustrative in writing the products of vectors and operators. Let us denote the set of linear maps from \mathcal{H} to itself as $\mathcal{L}(\mathcal{H})$; if $\dim(\mathcal{H}) = d$, then $\mathcal{L}(\mathcal{H})$ is essentially $\text{Mat}_{d \times d} \mathbb{C}$, the set of all complex matrices of the size d by d . It is also convenient for us to define the space of all Hermitian operators,

$$\text{Herm}(\mathcal{H}) = \{H \in \mathcal{L}(\mathcal{H}) \mid H = H^\dagger\},$$

where H^\dagger is the Hermitian conjugate of H , and unitary operators

$$\text{U}(\mathcal{H}) = \{U \in \mathcal{L}(\mathcal{H}) \mid U^\dagger U = \mathbb{1}\},$$

where $\mathbb{1}$ is the identity on \mathcal{H} . Now, for $\psi, \varphi \in \mathcal{H}$ and $A \in \mathcal{L}(\mathcal{H})$ the bra-ket notation gives

$$\psi^\dagger \varphi \equiv \langle \psi | \varphi \rangle, \quad (1.2)$$

$$\psi \varphi^\dagger \equiv |\psi\rangle \langle \varphi|, \quad (1.3)$$

$$\psi^\dagger A \varphi \equiv \langle \psi | A | \varphi \rangle, \quad (1.4)$$

Therefore, (1.2) is a scalar product of the form (1.1). The expression (1.3) gives a notation for the ‘‘outer product’’ of vectors; it describes rank-1 operators of the form

$T = |\psi\rangle\langle\varphi|$ which act on a vector $|\eta\rangle \in \mathcal{H}$ as $T|\eta\rangle = |\psi\rangle\langle\varphi|\eta\rangle$. The notation in (1.4) is equivalent to $\langle\psi, A\varphi\rangle = \langle A^\dagger\psi, \varphi\rangle$.

In (1.2), formally [51], $\langle\varphi|$ can be also viewed as an anti-linear function on \mathcal{H} defined as

$$\langle\varphi| : |\psi\rangle \rightarrow \langle\varphi|\psi\rangle \in \mathbb{C}. \quad (1.5)$$

Now consider a bi-linear function $\langle\varphi_1 \otimes \varphi_2|$ of the arguments $|\psi_1\rangle \in \mathcal{H}_1$ and $|\psi_2\rangle \in \mathcal{H}_2$ such that

$$\langle\varphi_1 \otimes \varphi_2| (|\psi_1\rangle, |\psi_2\rangle) = \langle\varphi_1|\psi_1\rangle \langle\varphi_2|\psi_2\rangle.$$

Considering the vector space L of linear combinations of all such functions, one may introduce the inner product

$$\langle\varphi_1 \otimes \varphi_2|\psi_1 \otimes \psi_2\rangle = \langle\varphi_1|\psi_1\rangle \langle\varphi_2|\psi_2\rangle \quad (1.6)$$

and extend it by linearity to L . The the space L is a Hilbert space itself, and it is called the tensor product of spaces \mathcal{H}_1 and \mathcal{H}_2 denoted as $\mathcal{H}_1 \otimes \mathcal{H}_2$, and its elements are $|\psi_1 \otimes \psi_2\rangle \equiv |\psi_1\rangle \otimes |\psi_2\rangle$. In fact, for finite-dimensional spaces, the explicit form of a vector $|\varphi_1 \otimes \varphi_2\rangle$ can be obtained if one treats the tensor product \otimes as the Kronecker product of two (column-) vectors. Indeed, one can show that the components i, j of the resultant vector have the form $(\psi_1 \otimes \psi_2)^{ij} = \psi_1^i \psi_2^j$, which satisfies the requirement for the scalar product (1.6).

Any vector $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ can be represented as

$$|\psi\rangle = \sum_{i,j=1}^D c_{ij} |e_1^i\rangle \otimes |e_2^j\rangle,$$

with $D = \dim(\mathcal{H}_1) \cdot \dim(\mathcal{H}_2)$, $\{e_1^j\}$ being a basis in \mathcal{H}_1 and $\{e_2^j\}$ a basis in \mathcal{H}_2 ; important is the fact that the vectors $\{|e_1^i\rangle \otimes |e_2^j\rangle\}_{i,j=1}^D$ also form a basis in $\mathcal{H}_1 \otimes \mathcal{H}_2$. For the operators A_1 and A_2 acting in \mathcal{H}_1 and \mathcal{H}_2 , respectively, the tensor product is given by

$$(A_1 \otimes A_2)(|\psi_1\rangle \otimes |\psi_2\rangle) = A_1 |\psi_1\rangle \otimes A_2 |\psi_2\rangle$$

and is extended to $\mathcal{H}_1 \otimes \mathcal{H}_2$.

The notion of the tensor product is needed for considering composite quantum systems with corresponding Hilbert spaces. Additionally, it allows to define the scalar product (1.5) as follows. The functions $\langle \varphi |$ also form their own Hilbert space \mathcal{H}^* , which is called dual to \mathcal{H} . It allows to define the scalar product (1.2) as a map such that

$$\langle \cdot | \cdot \rangle : \mathcal{H}^* \otimes \mathcal{H} \rightarrow \mathbb{C}.$$

As the tensor product of Hilbert spaces describes states of a composite quantum system, there is also a way to describe one of its parts. Let $T \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$, then its partial trace over the subspace H_B is the operator $\text{Tr}_{H_B} T \in \mathcal{L}(\mathcal{H}_A)$ such that

$$\langle \psi | \text{Tr}_{H_B}(T) | \varphi \rangle = \sum_{j=1}^{d_B} \langle \psi \otimes b_j | T | \varphi \otimes b_j \rangle,$$

where $d_B = \dim \mathcal{H}_B$, $\psi, \varphi \in \mathcal{H}_A$ and $\{b_j\}_{j=1}^{d_B}$ is a basis in \mathcal{H}_B . For $H_B = \mathbb{C}^{d_B}$, the partial trace can be found as

$$\text{Tr}_{H_B}(T) = \sum_{j=1}^{d_B} \left(\mathbb{1}_A \otimes \langle b_j | \right) T \left(\mathbb{1}_A \otimes | b_j \rangle \right),$$

where $\mathbb{1}_A$ is the identity operator on \mathcal{H}_A . The operator $T_A \equiv \text{Tr}_{H_B}(T)$ is often called the reduced operator.

1.2 Introduction to quantum computing

1.2.1 Quantum states

As was mentioned, in quantum mechanics, for a quantum system, there is a corresponding Hilbert space \mathcal{H} . Generally, the state of this system is described by a so-called density operator acting in a Hilbert space \mathcal{H} .

Definition 1 (Set of density operators) *The set of density operators on \mathcal{H} is*

$$\mathcal{D}(\mathcal{H}) = \{ \rho \in \text{Herm}(\mathcal{H}) \mid \rho \geq 0, \text{Tr } \rho = 1 \}. \quad (1.7)$$

In this definition, the inequality $\rho \geq 0$ means that $\lambda \geq 0 \quad \forall \lambda \in \text{spec}(\rho)$. Since we agreed that $\mathcal{H} = \mathbb{C}^d$ for some $d \in \mathbb{N}$, we sometimes will also call such operators density matrices.

By the spectral theorem, a density operator ρ — as any Hermitian operator — can be decomposed as

$$\rho = \sum_j \lambda_j |\psi_j\rangle\langle\psi_j|, \quad (1.8)$$

where λ_j and $|\psi_j\rangle$ are the eigenvalues and the corresponding eigenvectors of ρ , respectively. In the equation above, we assumed that all the eigenvalues λ are distinct. In case of finite dimensionality d and degenerate spectrum, one rewrites (1.8) as

$$\rho = \sum_{j=1}^d \lambda_j M_j,$$

where

$$M_j = \sum_{j:\lambda_j=\lambda} |\psi_j\rangle\langle\psi_j|$$

for distinct eigenvalues λ_j .

Of particular interest are density operators of so-called pure states.

Definition 2 (Pure state) *The state ρ is called pure if $\rho^2 = \rho$, i.e. if ρ is a projector.*

This means that for a pure state ρ there is only one term in the decomposition (1.8), which implies that there is some $\psi \in \mathcal{H}$ such that $\rho|\psi\rangle = |\psi\rangle$. It gives $\rho = |\psi\rangle\langle\psi|$, which is a projector onto $|\psi\rangle$. Therefore, density operators of pure states acting in \mathcal{H} can be represented as unit vectors from this space. The purity of a state $\rho = \sum_j \lambda_j |\psi_j\rangle\langle\psi_j|$ can be quantified by

$$P(\rho) = \text{Tr} \rho^2 = \sum_j \lambda_j^2, \quad (1.9)$$

which equals 1 for pure states and is less than 1 otherwise. The states which are not pure are called mixed. Worthy of note is also that the vectors $|\psi\rangle$ and $|\varphi\rangle = e^{ia} |\psi\rangle$ are considered to describe the same quantum state for any $a \in \mathbb{R}$, as they give the same density operator $|\psi\rangle\langle\psi|$.

Pure and mixed states can be understood from geometric perspective. A set of density operators $\mathcal{D}(\mathcal{H})$ is a convex subset of the space of all linear Hermitian operators in \mathcal{H} [51]. It can be proven that the extreme points $\rho_j \in \mathcal{D}(\mathcal{H})$ called pure states are precisely one-dimensional projectors of the form $|\psi_j\rangle\langle\psi_j|$. For a state $\rho \in \mathcal{D}(\mathcal{H})$ one can write a unique convex combination

$$\rho = \sum_{j=1}^d \lambda_j |\psi_j\rangle\langle\psi_j|, \quad \sum_{j=1}^d \lambda_j = 1,$$

which resembles the spectral decomposition (1.8). In other words, any (mixed) state is a convex combination of pure states.

Of particular interest are states which describe two-level systems, e.g. spin projection of an electron, energy levels of an ion, polarization of a photon etc. Such states are called quantum bits or qubits.

Definition 3 (Quantum bit) *A quantum bit, or qubit, is a density operator acting in a Hilbert space of dimensionality 2.*

Let us fix in $\mathcal{H} = \mathbb{C}^2$ the basis $\{|0\rangle = (1, 0)^T, |1\rangle = (0, 1)^T\}$ called the computational or standard basis. Then, taking into account the definition of a density matrix (1.7), one may establish that the state of a qubit can be parametrized as

$$\begin{aligned} \rho &= \frac{1}{2}(\mathbb{1} + a_x\sigma_x + a_y\sigma_y + a_z\sigma_z), \\ 0 &\leq a_x^2 + a_y^2 + a_z^2 \leq 1. \end{aligned} \tag{1.10}$$

Here, the Pauli operators σ_j are defined as

$$\begin{aligned}\mathbb{1} &= |0\rangle\langle 0| + |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \sigma_x &= |0\rangle\langle 1| + |1\rangle\langle 0| = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \sigma_y &= -i|0\rangle\langle 1| + i|1\rangle\langle 0| = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \\ \sigma_z &= |0\rangle\langle 0| - |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.\end{aligned}$$

If the state is pure, i.e. $\rho = |\psi\rangle\langle\psi|$, one finds the following parametrization for the state vector:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad \alpha, \beta \in \mathbb{C},$$

with the requirement $\text{Tr} \rho = 1$ giving $\langle\psi|\psi\rangle = 1$ and hence $|\alpha|^2 + |\beta|^2 = 1$.

As for a system of n qubits, the associated Hilbert space is $\mathcal{H} = \mathbb{C}^{\otimes 2n} \equiv \mathbb{C}_2^{\otimes n}$ with the computational basis being $\{|0\rangle, |1\rangle\}^{\otimes n}$. One can establish that

$$\mathcal{L}(\mathbb{C}_2^{\otimes n}) \equiv \text{Mat}_{2^n \times 2^n} \mathbb{C} = \text{span}_{\mathbb{C}} \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}, \quad n \in \mathbb{N}, \quad (1.11)$$

and therefore one can parametrize a density operator $\rho \in \mathcal{D}(\mathbb{C}_2^{\otimes n})$ as follows:

$$\begin{aligned}\rho &= \frac{1}{2^n} \sum_{K \in \{0,x,y,z\}^{\times n}} a_K S_K, \\ S_K &= \bigotimes_{j \in K} \sigma_j, \\ a_K &\in \mathbb{R}, \quad a_{\{0\}^{\times n}} = 1, \quad \sum_{K \in \{0,x,y,z\}^{\times n}} a_K^2 \leq 1.\end{aligned} \quad (1.12)$$

As an instance, for $n = 1$ the equation (1.12) reduces to (1.10), and if $n = 2$ it would give

$$\rho = \frac{1}{4} \sum_{i,j=0}^3 a_{ij} \sigma_i \otimes \sigma_j, \quad a_{00} = 1.$$

1.2.2 Transformation of quantum states

Let us consider an affine map $\Phi : \mathcal{D}(\mathcal{H}) \rightarrow \mathcal{D}(\mathcal{H})$, i.e. a map taking density operators to density operators such that it preserves the “weights” of states in mixes:

$$\begin{aligned} \Phi \left[\sum_j p_j \rho_j \right] &= \sum_j p_j \Phi[\rho_j], \\ p_j \geq 0, \quad \sum_j p_j &= 1, \quad \rho_j \in \mathcal{D}(\mathcal{H}). \end{aligned} \tag{1.13}$$

In a case when $\Phi : \mathcal{D}(\mathcal{H}_A) \rightarrow \mathcal{D}(\mathcal{H}_B)$, i.e. when the input and output Hilbert spaces are different, one needs a more general notion of quantum transformations. Such a map is called a quantum channel.

Definition 4 (Quantum channel) *A quantum channel is a map $\Phi : \mathcal{D}(\mathcal{H}_A) \rightarrow \mathcal{D}(\mathcal{H}_B)$ such that*

$$(1) \quad \Phi \left[\sum_i c_i \rho_i \right] = \sum_i c_i \Phi[\rho_i] \tag{linearity}$$

$$(2) \quad \text{Tr} \Phi[\rho] = \text{Tr} \rho \tag{trace preservation}$$

$$(3) \quad \left(\Phi \otimes \text{Id} \right) [\rho_{AE}] \geq 0 \tag{complete positivity}$$

In (1), $\rho_i \in \mathcal{D}(\mathcal{H}_A)$, $0 \leq c_i \leq 1$, and $\sum_i c_i = 1$. In (2), $\rho \in \mathcal{D}(\mathcal{H}_A)$. In (3), $\rho_{AE} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_E)$ and $\mathbb{1}$ is the identity map on $\mathcal{D}(\mathcal{H}_E)$.

Any quantum channel $\Phi : \mathcal{D}(\mathcal{H}_A) \rightarrow \mathcal{D}(\mathcal{H}_B)$ admits the so-called (not unique) Kraus representation:

$$\Phi[\rho] = \sum_{j=1}^D V_j \rho V_j^\dagger,$$

where $D \leq \dim(\mathcal{H}_A) \cdot \dim(\mathcal{H}_B)$ and the operators $V_j : \mathcal{H}_A \rightarrow \mathcal{H}_B$ are such that

$$\sum_{j=1}^D V_j^\dagger V_j = \mathbb{1}. \tag{1.14}$$

In the special case when $D = 1$ and $\mathcal{H}_A = \mathcal{H}_B$ we obtain

$$\Phi[\rho] = U \rho U^\dagger,$$

where U is unitary, i.e. $U^\dagger U = \mathbb{1}$. Interestingly, by Wigner's theorem, this also follows directly from (1.13) if one supposes that the map Φ is bijective, although in this case the operator U is allowed to be also antiunitary, i.e. such that $\|U\psi\| = \|\psi\|$ and $U \sum_j c_j \psi_j = \sum_j c_j^* U \psi_j$, $c_j \in \mathbb{C}$, $\psi_j \in \mathcal{H}$.

Now consider a one-parameter group of unitary operators U_t with $t \in \mathbb{R}$ such that

1. $U_0 = \mathbb{1}$,
2. $U_t U_s = U_{t+s}$
3. the map $t \rightarrow U_t$ is continuous.

The Stone's theorem states that in this case

$$U_t = e^{-itH}, \quad (1.15)$$

where H is a Hermitian operator. We will use this result later when describing the so-called quantum gates, unitary operators used for transforming (commonly, pure) quantum states $|\varphi\rangle \rightarrow |\psi\rangle = U|\varphi\rangle$.

1.2.3 Entanglement

There is an important class of quantum states which comprises the so-called entangled states. Consider a pure state $|\psi\rangle$ from the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ the subspaces of which contain the corresponding orthogonal bases $\{|e_1^j\rangle\}_{j=1}^{d_1}$ and $\{|e_2^j\rangle\}_{j=1}^{d_2}$, where $d_k = \dim \mathcal{H}_k$, $k \in \{1, 2\}$. It is known that this state is always representable in the form of the so-called Schmidt decomposition:

$$|\psi\rangle = \sum_{j=1}^{d_m} \sqrt{\lambda_j} |e_1^j\rangle \otimes |e_2^j\rangle, \quad (1.16)$$

where $d_m = \min\{d_1, d_2\}$ and $\lambda_j \geq 0$.

Definition 5 (Entanglement of a pure state) *A pure state $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ is called separable if it can be represented as $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$, where $|\psi_1\rangle \in \mathcal{H}_k$ with $k \in \{1, 2\}$. Otherwise, the state is called entangled.*

One can quantify the entanglement of a pure state by the number of terms in its Schmidt decomposition (1.16) [52]. That is, a pure state is not entangled (or separable) if there is only one term in this decomposition. More generally, one needs an entanglement measure, a function $E(|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2)$ with the following properties:

1. $E(|\psi\rangle) \geq 0$ for any state vector $|\psi\rangle$, and the equality is achieved only when $|\psi\rangle$ is separable;
2. $E((U_1 \otimes U_2)|\psi\rangle) = E(|\psi\rangle)$ with $U_1 \in U(\mathcal{H}_1)$ and $U_2 \in U(\mathcal{H}_2)$, i.e. the entanglement does not increase under local transformations.

One of such functions is

$$E(|\psi\rangle) = H(\text{Tr}_2 |\psi\rangle\langle\psi|) = H(\text{Tr}_1 |\psi\rangle\langle\psi|),$$

where

$$H(\rho) = -\text{Tr} \rho \log \rho = \sum_j \lambda_j \log \lambda_j$$

is called the von Neumann entropy.

Now let us consider mixed states.

Definition 6 (Entanglement of a mixed state) *A mixed state $\rho \in \mathcal{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is called separable if it can be decomposed as*

$$\rho = \sum_j p_j \rho_1^j \otimes \rho_2^j,$$

where $\sum_j p_j = 1$ and ρ_k^j are (in general, mixed) states from $\mathcal{D}(\mathcal{H}_k)$, $k \in \{1, 2\}$. Otherwise the state is called entangled.

Quantifying the entanglement of mixed states is more complicated. One way to do it is computing the entanglement of formation

$$E(\rho) = \min \sum_j p_j H(\text{Tr}_2 \rho_j),$$

where the minimum is taken over all possible decompositions for $\rho = \sum_j p_j \rho_j$. Information on other entanglement measures can be found e.g. in [53].

1.2.4 Observables and measurement

In quantum mechanics, an observable is commonly described by a Hermitian operator $A = A^\dagger$ which can be decomposed in the form (1.8), i.e.

$$A = \sum_a a E_a,$$

where a and $E_a = |e_a\rangle\langle e_a|$ are the eigenvalues and the corresponding eigenprojectors of A . A result of the measurement of an observable A in a state ρ is $a \in \text{spec}(A)$, and the probability of obtaining a specific value as an outcome is described by the Born-von Neumann statistical postulate:

$$p_\rho^A(a) = \text{Tr } \rho E_a. \quad (1.17)$$

Summing over all possible outcomes a , one obtains the expected value of A measured in the state ρ ,

$$E_\rho(A) = \sum_a a p_\rho^A(a) = \sum_a a \text{Tr } \rho E_a = \text{Tr } \rho A. \quad (1.18)$$

For a pure state $\rho = |\psi\rangle\langle\psi|$, it simplifies to

$$E_\psi(A) = \langle\psi|A|\psi\rangle.$$

An important peculiarity of the measurement of an observable with respect to some state is that this state generally changes in accordance with the obtained measurement outcome. That is, if one measures an observable $A = \sum_a a E_a$ with respect to a state ρ and obtains an outcome a , the post-measurement state would be

$$\rho' = \frac{1}{p_\rho^A(a)} E_a \rho E_a,$$

for $p_\rho^A(a) = \text{Tr } \rho E_a > 0$. Such a transformation is demanded for making the measurements repeatable: After measuring the resultant state again, we must obtain the very same outcome. This causes another issue: We need to conduct a measurement experiment many times to estimate the probabilities $p_\rho^A(a)$. Practically, we perform

N measurements, count the number of times N_a we obtained the outcome a , and infer the probabilities $p_\rho^A(a)$ from the frequencies N_a/N . From these estimated probabilities we then obtain the expected value of A as given in (1.18). To estimate the expected value of an observable with a precision ϵ , one generally needs to conduct $N = O(1/\epsilon^2)$ measurements [54, 55].

It also must be noted that the probability distribution $p_\rho^A(a)$ in (1.17) is affine, i.e. for a mix of states $\rho = \sum_j p_j \rho_j$ we have

$$p_\rho^A(a) = \sum_j p_j p_{\rho_j}^A(a)$$

This property allows to generalize the notion of quantum measurement. [51]

Theorem 1 *Let there be a function from quantum states $\rho \in \mathcal{D}(\mathcal{H})$ to probability distributions p_ρ on some finite set of outcomes \mathcal{A} . Then if this function is affine, then there is a unique set of operators $\{\Pi_a \mid a \in \mathcal{A}\}$ called a resolution of the identity such that*

$$\Pi_a = \Pi_a^\dagger \geq 0, \quad \sum_{a \in \mathcal{A}} \Pi_a = \mathbb{1}$$

for which

$$p_\rho(a) = \text{Tr } \rho \Pi_a. \tag{1.19}$$

Vice versa, for a resolution of the identity in \mathcal{H} , the function (1.19) defines an affine map from quantum states to probability distributions on \mathcal{A} .

Often the resolution of the identity $\{\Pi_a \mid a \in \mathcal{A}\}$ is called the positive operator-valued measure (POVM). If $\mathcal{A} \subset \mathbb{R}$, and if the POVM is orthogonal, namely, if $\Pi_x \Pi_y = \delta_{xy} \Pi_x$ for all $x, y \in \mathcal{A}$, the expression

$$A = \sum_{a \in \mathcal{A}} a \Pi_a$$

establishes a bijection between Hermitian operators on \mathcal{H} with $\text{spec}(A) = \mathcal{A}$ and orthogonal resolutions of identity $\{\Pi_a \mid a \in \mathcal{A}\}$ in this \mathcal{H} . This traces us back to the spectral decomposition (1.8). Non-orthogonal POVMs — similarly to quantum states — allow to consider mixed observables; they are of particular usefulness

quantum state tomography and quantum error correction.

1.3 Tensor diagrams

In quantum computing, objects like vectors and linear operators can be represented in the computational basis. For example, let $|\psi\rangle \in \mathcal{H}_j$, $\dim \mathcal{H}_j = d_j$, and $\{|j\rangle\}_{j=1}^{d_j}$ be the computational basis in \mathcal{H}_j , then

$$|\psi\rangle = \sum_{j=1}^{d_j} \psi^j |j\rangle,$$

where $\psi^j = \langle j|\psi\rangle$. Similarly, for a covector $\langle\varphi| \in \mathcal{H}_j^*$ we have

$$\langle\varphi| = \sum_{j=1}^{d_j} \varphi_j \langle j|$$

with $\varphi_j = \langle\varphi|j\rangle$. An operator $A : \mathcal{H}_j \rightarrow \mathcal{H}_k$ can be decomposed as

$$A = \sum_{j=1}^{d_j} \sum_{k=1}^{d_k} A^k_j |k\rangle\langle j|,$$

where $A^k_j = \langle k|A|j\rangle$ and $d_k = \dim \mathcal{H}_k$.

In the equations above, the vector $|\psi\rangle$ and the operator A can be thought as tensors with the respective components ψ^j and A^k_j in the computational bases of the corresponding spaces. Indeed, the action of the operator A to the vector $|\psi\rangle$ reminds the tensor contraction:

$$A|\psi\rangle = \sum_{j=1}^{d_j} \sum_{k=1}^{d_k} A^k_j \psi^j |k\rangle = \sum_{k=1}^{d_k} \varphi^k |k\rangle = |\varphi\rangle \in \mathcal{H}_k.$$

With this idea in mind, the objects ψ^j and A^k_j can be also thought as tensors with the abstract tensor notation applied. In this notation, the indices indicate not the components in a particular basis, but serve as placeholders and indicate the type of

a tensor. For example, let $T : \mathcal{H}_j \rightarrow \mathcal{H}_k \otimes \mathcal{H}_l$ have the representation

$$T = \sum_{j=1}^{d_j} \sum_{k=1}^{d_k} \sum_{l=1}^{d_l} T^{kl}_j |kl\rangle\langle j|.$$

With the abstract index notation, we can denote the action of $T \equiv T^{kl}_j$ on $|\psi\rangle \equiv \psi^j$ as

$$T|\psi\rangle \equiv T^{kl}_j \psi^j = \eta^{kl} \equiv |\eta\rangle \in \mathcal{H}_k \otimes \mathcal{H}_l,$$

where we applied the Einstein summation convention which allows to remove the sum signs and which implies summation over repeating indices.

In the tensor networks language, tensors have their specific depictions. In this section, basing on the works [43, 44], we introduce the notation of tensor diagrams. A vector $|\psi\rangle \in \mathcal{H}_i$ and a covector $\langle\varphi| \in \mathcal{H}_i^*$ are commonly depicted as a triangle with an outcoming wire,

$$|\psi\rangle = \psi^i = \begin{array}{c} \text{---} i \text{---} \\ \triangle \psi \end{array}$$

$$\langle\varphi| = \varphi_i = \begin{array}{c} \triangle \varphi \\ \text{---} i \text{---} \end{array}$$

To show a tensor T , one uses squares with in- and outcoming wires; for instance, a tensor $T : \mathcal{H}_i \rightarrow \mathcal{H}_j \otimes \mathcal{H}_k$ is depicted as

$$T^{jk}_i = \begin{array}{c} \text{---} k \text{---} \\ \square T \\ \text{---} j \text{---} \end{array} \text{---} i \text{---}$$

The tensor contraction is shown as connecting the wires of the corresponding tensors. For example, let $\psi, \varphi \in \mathcal{H}$ and $T \in \mathcal{L}(\mathcal{H})$; then

$$\langle\psi|T|\varphi\rangle = \psi_i T^i_j \varphi^j = \begin{array}{c} \triangle \psi \text{---} i \text{---} \\ \square T \\ \text{---} j \text{---} \triangle \varphi \end{array}$$

Placing tensors $A \in \mathcal{L}(\mathcal{H}_1)$ and $B \in \mathcal{L}(\mathcal{H}_2)$ together in parallel means their tensor

product:

$$\begin{aligned}
 (A \otimes \mathbb{1})(\mathbb{1} \otimes B) &= (\mathbb{1} \otimes B)(A \otimes \mathbb{1}) = (A \otimes B) \\
 &= \begin{array}{c} \text{---} \boxed{A} \text{---} \\ \text{---} \boxed{B} \text{---} \end{array} = \begin{array}{c} \text{---} \boxed{A} \text{---} \\ \text{---} \boxed{B} \text{---} \end{array} = \begin{array}{c} \text{---} \boxed{A} \text{---} \\ \text{---} \boxed{B} \text{---} \end{array}
 \end{aligned}$$

The equations above also show that an identity operator $\mathbb{1}$ is depicted as a straight line. In the tensor networks language, a special notation is given to the so-called (unnormalized) Bell states:

$$|\Phi^+\rangle = \sum_j |j\rangle \otimes |j\rangle = \begin{array}{c} \text{---} \cup \end{array} \quad (1.20)$$

$$\langle\Phi^+| = \sum_j \langle j| \otimes \langle j| = \begin{array}{c} \text{---} \cap \end{array} \quad (1.21)$$

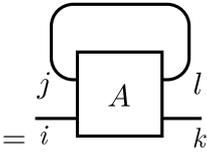
With the notion of the Bell state, one can compute the trace of an operator. For example, the trace of $A \in \mathcal{L}(\mathcal{H})$ can be computed as the expected value of the operator $\mathbb{1} \otimes A \in \mathcal{L}(\mathcal{H} \otimes \mathcal{H})$ in the state $|\Phi^+\rangle \in \mathcal{H} \otimes \mathcal{H}$,

$$\text{Tr } A = \langle\Phi^+|\mathbb{1} \otimes A|\Phi^+\rangle$$

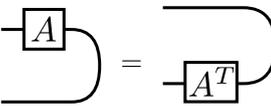
$$= \begin{array}{c} \text{---} \cup \\ \boxed{A} \\ \text{---} \cap \end{array}$$

Therefore, taking the trace of an operator is graphically indicated as connecting the wires of this operator. The partial trace over a subspace can be also computed in the same way, i.e. via connecting the wires of the corresponding subspaces. That

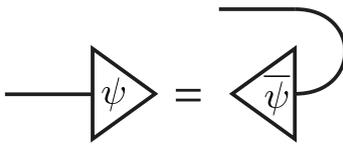
is, if $A \in \mathcal{L}(\mathcal{H}_i \otimes \mathcal{H}_j)$, then the partial trace over \mathcal{H}_i is

$$\text{Tr}_{\mathcal{H}_i} A = \left(\langle \Phi^+ | \otimes \mathbb{1} \right) (\mathbb{1} \otimes A) \left(| \Phi^+ \rangle \otimes \mathbb{1} \right)$$


From this equation one can also establish the so-called ricochet trick:

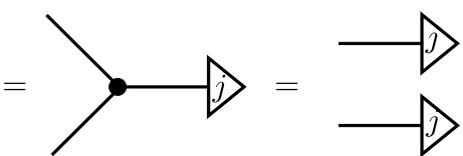
$$(A \otimes \mathbb{1}) | \Phi^+ \rangle = (\mathbb{1} \otimes A^T) | \Phi^+ \rangle \tag{1.22}$$


One can also find the following interesting property of the Bell state:



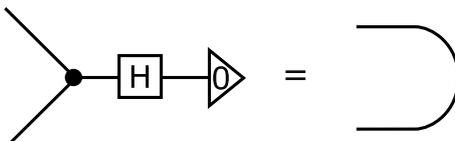
$$\tag{1.23}$$

Important in the tensor networks language also is the COPY which acts as

$$\text{COPY } |j\rangle = |j\rangle \otimes |j\rangle$$


$$\tag{1.24}$$

where $j \in \{0, 1\}$. This tensor is used for constructing the so-called controlled gates which will be discussed further. Additionally, the unnormalized Bell state (1.20) can be prepared as



where H is called the Hadamard gate which acts as $H |0\rangle = \sum_j |j\rangle$. Finally, let us

introduce the swap tensor SW which for the case of qubits is defined to be:

$$\begin{aligned}
 \text{SW} &= \frac{1}{2} \sum_{j=0}^3 \sigma_j \otimes \sigma_j & (1.25) \\
 &= \sum_{k,l=0}^1 |k\rangle\langle l| \otimes |l\rangle\langle k| \\
 &= \text{X}
 \end{aligned}$$

Presented in this section notation will be used in Chapter 3 for formulating variational quantum algorithms and making derivations. Also, in Section 1.4, we will consider quantum circuits, a similar notation for describing quantum computations.

1.3.1 Tensor network states

The notation of tensor network diagrams discussed in Section 1.3 is closely connected to the concept of tensor network states [56, 57]. Overall, tensor network states is a way of representing quantum states classically. Although we do not utilize this concept in this thesis, we nevertheless consider it worthy of mention. In what follows, we describe one of the most prominent tensor network states, the matrix product state (MPS).

Consider a pure n -qubit state $|\psi\rangle \in \mathbb{C}_2^{\otimes n}$ represented in the computational basis,

$$|\psi\rangle = \sum_{i,j,\dots,k \in \{0,1\}} \psi_{ij\dots k} |ij\dots k\rangle, \quad (1.26)$$

where we used the notation $|ij\rangle \equiv |i\rangle \otimes |j\rangle$. This decomposition may have up to 2^n coefficients $\psi_{ij\dots k} \in \mathbb{C}$, meaning that generally we need memory of an exponential capacity to classically store such a state. At the meantime, an n -qubit state can be represented as a so-called matrix product state.

Definition 7 (Matrix product state) *Matrix product state (MPS) is a represen-*

tation of an n -qubit state vector in the form

$$|\psi\rangle = \sum_{i,j,\dots,k \in \{0,1\}} \text{Tr} \left(A_i^{[1]} A_j^{[2]} \cdots A_k^{[n]} \right) |ij \dots k\rangle. \quad (1.27)$$

Comparing (1.26) and (1.27), one notices that we can represent the decomposition coefficients $\psi_{ij\dots k}$ by products of the matrices $A_i^{[\alpha]}$, where $\alpha \in \{1, 2, \dots, n\}$ and $i \in \{0, 1\}$. If these matrices are of the size χ by χ , then one needs $2n\chi^2$ numbers to store an n -qubit state. For example, one can represent the so-called Greenberger–Horne–Zeilinger (GHZ) state

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}} (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$$

as an MPS by setting $A_0^{[\alpha]} = |0\rangle\langle 0|$ and $A_1^{[\alpha]} = |1\rangle\langle 1|$.

One way of finding an MPS representation of a given pure n -qubit state $|\psi\rangle \in \mathbb{C}_2^{\otimes n}$ is the consecutive Schmidt decomposition (1.16) [52, 58]. First, omitting the tensor product signs, we find the Schmidt decomposition of $|\psi\rangle \equiv |\psi^{[1,2,\dots,n]}\rangle$ for the bipartition $[1|2, \dots, n]$, i.e. for the “cut” which separates the first qubit from the rest of the qubits:

$$\begin{aligned} |\psi^{[1,2,\dots,n]}\rangle &= \sum_{a_1 \in \{0,1\}} \Lambda_{a_1}^{[1]} |a_1^{[1]}\rangle |a_1^{[2,\dots,n]}\rangle \\ &= \sum_{a_1 \in \{0,1\}} \sum_{i \in \{0,1\}} A_{a_1,i}^{[1]} \Lambda_{a_1}^{[1]} |i^{[1]}\rangle |a_1^{[2,\dots,n]}\rangle, \end{aligned}$$

where $\Lambda_{a_1}^{[1]}$ are the Schmidt coefficients, and in the second equality we decomposed the Schmidt vectors $|a_1^{[1]}\rangle$ in the computational basis with the decomposition coefficients

being $A_{a_1,i}^{[1]} = \langle i^{[1]} | a_1^{[1]} \rangle$. Now consider the bipartition $[1, 2 | 3, \dots, n]$:

$$\begin{aligned}
 |\psi^{[1,2,\dots,n]}\rangle &= \sum_{a_2 \in \{0,1\}^{\times 2}} \Lambda_{a_2}^{[2]} |a_2^{[1,2]}\rangle |a_2^{[3,\dots,n]}\rangle \\
 &= \sum_{a_1 \in \{0,1\}} \sum_{a_2 \in \{0,1\}^{\times 2}} \sum_{j \in \{0,1\}} A_{a_1,a_2,j}^{[2]} \Lambda_{a_2}^{[2]} |a_1^{[1]}\rangle |j^{[1]}\rangle |a_2^{[3,\dots,n]}\rangle \\
 &= \sum_{a_1 \in \{0,1\}} \sum_{a_2 \in \{0,1\}^{\times 2}} \sum_{i,j \in \{0,1\}} A_{a_1,i}^{[1]} A_{a_1,a_2,j}^{[2]} \Lambda_{a_2}^{[2]} |i^{[1]}\rangle |j^{[2]}\rangle |a_2^{[3,\dots,n]}\rangle,
 \end{aligned}$$

where in the second line we decomposed the Schmidt vectors $|a_2^{[1,2]}\rangle$ in the basis of vectors $|a_1^{[1]}\rangle |j^{[2]}\rangle$ with the coefficients $A_{a_1,a_2,j}^{[2]} = \left(\langle a_1^{[1]} | \langle j^{[2]} | \right) |a_2^{[1,2]}\rangle$, and in the third line we again decomposed the Schmidt vectors $|a_1^{[1]}\rangle$ in the computational basis. This procedure is repeated until reaching the last bipartition, $[1, 2, \dots, n-1 | n]$. For it, we decompose the states $\Lambda_{a_{n-1}}^{[n-1]} |a_{n-1}^{[n]}\rangle$ in the computational basis with the coefficients $A_{a_{n-1},k}^{[n]} = \Lambda_{a_{n-1}}^{[n-1]} \langle k^{[n]} | a_{n-1}^{[n]}\rangle$. This gives

$$|\psi^{[1,2,\dots,n]}\rangle = \sum_{i,j,\dots,k \in \{0,1\}} A_i^{[1]} A_j^{[2]} \dots A_k^{[n]} |i^{[1]}\rangle |j^{[2]}\rangle \dots |k^{[n]}\rangle.$$

Note that in the equation we obtained, $A_i^{[1]}$ is a row vector and $A_k^{[n]}$ is a column vector, so we do not need to take the trace of the matrix product as done in (1.27). It is also worth mentioning that one can apply tensor diagrams notation introduced in Section 1.3 to obtain an MPS as demonstrated in [59].

One may notice that while taking consecutive Schmidt decompositions towards the $\lfloor n/2 \rfloor$ -th qubit, the number of Schmidt coefficients grows exponentially. What can be done is keeping only χ largest Schmidt coefficients after each decomposition, and discarding the rest. Earlier we mentioned that by the number of the Schmidt coefficients one can quantify the entanglement of a bipartition. That is, if there are χ terms in the Schmidt decomposition of a state, one needs about $\log_2 \chi$ two-qubit operations (e.g. CX gates discussed in Section 1.4) to prepare this state [52]. Therefore, the more Schmidt coefficients we keep, the more entanglement our MPS can support.

With $\chi = 2^{\lfloor n/2 \rfloor}$, an MPS is known to be able to approximate any n -qubit

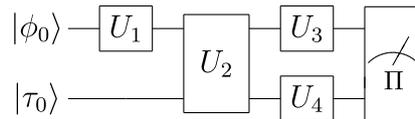
state. Interestingly, in practice, one usually does not need such a large expressibility. That is, the states of quantum systems described by a number of physically relevant models follow one of the so-called area laws [60, 61]: The entanglement grows proportionally to the “area” of the bipartition which separates two subsets of interacting qubits, i.e. the number of qubits located along the boundary of the cut. More information about physical models obeying or disobeying area laws, and also about other tensor network states, can be found in e.g. [57].

1.4 Quantum circuits

Quantum computations, i.e. transformations and measurements of quantum states, can be graphically represented using the quantum circuits notation. Suppose we want to first prepare the following pure two-qubit state:

$$|\psi\rangle = (U_3 \otimes U_4) U_2 (U_1 \otimes \mathbb{1}) (|\phi_0\rangle \otimes |\tau_0\rangle),$$

where U_1, U_3, U_4 are single-qubit operators, U_2 is a two-qubit operator, and $|\psi_0\rangle \otimes |\tau_0\rangle$ is some separable initial state of two qubits. Also suppose that we then want to measure it with the POVM Π . This whole process can be depicted as follows:



This graphical representation is called the quantum circuit notation. The blocks which are called gates indicate unitary operators. The notation of quantum circuits has many similarities with the tensor network notation introduced in Section 1.3. That is, blocks indicate tensors, the lines which connect blocks indicate tensor contraction, and blocks and lines placed in parallel mean tensor products. In quantum circuits though, the lines are commonly treated as individual qubits, and on the left side of the circuit is put the initial state being transformed (in the case above, it is $|\psi_0\rangle \otimes |\tau_0\rangle$). Quantum circuits also often contain measurements which are depicted by $\text{---}\boxed{\text{---}}\text{---}$, where one can indicate the POVM used (in our case, it is

II). In quantum computations, it is commonly assumed that we can measure only individual qubits with the POVM elements $\Pi = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ with the corresponding outcomes $\mathcal{A} = \{-1, +1\}$ (although sometimes the set of outcomes is changed to $\mathcal{A} = \{0, 1\}$, for compatibility with classical computing). In fact, this corresponds to measuring the observable $\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|$. Simultaneously measuring all individual qubits of an n -qubit state means measuring the observable $\sigma_z^{\otimes n}$. Intuitively, one may think that this is a measurement with the POVM elements

$$\Pi = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n}$$

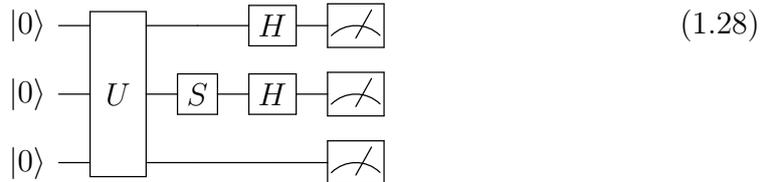
and the outcome “ -1 ” is attributed to the projectors onto $|j_1 j_2 \dots j_n\rangle \equiv |j_1\rangle \otimes |j_2\rangle \otimes \dots \otimes |j_n\rangle$, $j_k \in \{0, 1\}$, such that $\sum_{k=1}^n j_k \bmod 2 = 1$, and the outcome “ $+1$ ” to the projectors such that $\sum_{k=1}^n j_k \bmod 2 = 0$. This follows from the spectral decomposition of the Hermitian operator

$$\sigma_z^{\otimes n} = \sum_{a \in \{0,1\}^n} (-1)^{|a|} |a\rangle\langle a|.$$

For instance, if one measures the observable $A = \sigma_z \otimes \sigma_z$ with respect to a two-qubit state $|\psi\rangle \in \mathbb{C}_2^{\otimes 2}$, then according to (1.18) one has

$$E_\psi(A) = p_\psi^A(+1, -1) - p_\psi^A(-1, -1) - p_\psi^A(-1, +1) + p_\psi^A(+1, +1).$$

To measure other observables, one puts proper gates before the measurement blocks. For example, the following circuit describes the measurement of the observable $\sigma_x \otimes \sigma_y \otimes \sigma_z$ in the state $U|0\rangle^{\otimes 3}$:



Here, we applied the commonly used Hadamard $H = (\sigma_x + \sigma_z)/\sqrt{2}$ and phase $S = \sqrt{\sigma_z} = |0\rangle\langle 0| + i|1\rangle\langle 1|$ gates. One can establish that $H|0\rangle = (|0\rangle + |1\rangle)/\sqrt{2} \equiv |+\rangle$

and $H|1\rangle = (|0\rangle - |1\rangle)/\sqrt{2} \equiv |-\rangle$ are the eigenvectors of σ_x ; the same can be shown for $|\odot\rangle \equiv SH|0\rangle$ and $|\ominus\rangle \equiv SH|1\rangle$ to be the eigenvectors of σ_y . That is, by putting proper gates before the measurement, one “rotates” the measurement basis (or, changes the POVM elements). In the circuit above, we also took into account a common assumption that the initial states of individual qubits are $|0\rangle$.

1.4.1 Quantum gates

In the previous section, we mentioned that qubit transformations in quantum circuit notation are called gates. In quantum computing, a quantum gate is, in general, a unitary operator acting in some Hilbert space \mathcal{H} . Often the notion of a gate is restricted to an element of the special unitary group,

$$\text{SU}(d) = \{U \mid U \in \text{U}(\mathcal{H}), \dim(\mathcal{H}) = d, \det(U) = 1\}.$$

The unitary operators U and $V = e^{i\alpha}U$ with $\alpha \in \mathbb{R}$ are equal up to the factor $e^{i\alpha}$ called the global phase, which is physically not observable since $|e^{i\alpha}| = 1$. Indeed, one easily finds that $V\rho V^\dagger = U\rho U^\dagger$ for any density operator ρ .

Alongside with the mentioned earlier gates H and S , commonly used are the Pauli gates $X \equiv \sigma_x$, $Y \equiv \sigma_y$ and $Z \equiv \sigma_z$. Importantly, the Pauli operators can be used to generate any element from $\text{SU}(2)$. It is known that any its element can be represented as

$$R_{\mathbf{n}}(\alpha) = e^{-i\alpha \mathbf{n} \cdot \boldsymbol{\sigma}} \quad (1.29)$$

where $\alpha \in \mathbb{R}$, $\mathbf{n} = (n_x, n_y, n_z) \in \mathbb{R}^3$ and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, so that $\mathbf{n} \cdot \boldsymbol{\sigma} = n_x\sigma_x + n_y\sigma_y + n_z\sigma_z$ (note the similarity of (1.29) with (1.15)). Of particular interest are also the gates called the Pauli rotations (or x -, y - and z -rotations) of the form $R_j(\alpha) = e^{-i\alpha\sigma_j}$ with $j \in \{x, y, z\}$. In practical quantum computing, for performing an arbitrary single-qubit operation, it is enough to implement any two of these rotations on a hardware. Indeed, one can show that any $U \in \text{SU}(2)$ can be decomposed as

$$U = R_k(\alpha) R_l(\beta) R_k(\gamma) \quad (1.30)$$

for any $k, l \in \{x, y, z\}$ such that $k \neq l$.

As we mentioned in Section 1.2.3, entanglement cannot be generated by local (single-qubit) transformations, so there is a need for at least two-qubit gates for it. One, of course, may think of Pauli rotations of the form $R_{kl}(\alpha) = e^{-i\alpha \sigma_k \otimes \sigma_l}$. For example, the operation R_{xx} implemented via the so-called Mølmer-Sørensen gate [62] is natural for quantum computers based on trapped ions [63]. However, historically researchers were interested in quantum gates which are reminiscent of those in classical computing. An option is a two-qubit gate which transforms the state of one qubit conditioned on the state of another. That is, if one treats the qubit states of the computational basis $\{|0\rangle, |1\rangle\}$ as logical 0 and 1, it is possible to devise a gate CU which acts as follows:

$$\begin{aligned} CU(|0\rangle \otimes |\psi\rangle) &= |0\rangle \otimes |\psi\rangle, \\ CU(|1\rangle \otimes |\psi\rangle) &= |1\rangle \otimes U|\psi\rangle. \end{aligned}$$

Formally, such a gate can be written as

$$CU = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes U = \begin{array}{c} \text{---} \bullet \text{---} \\ | \\ \text{---} \boxed{U} \text{---} \end{array} \quad (1.31)$$

In the general case of n qubits, a controlled- U gate with the i th control and j th target qubits takes the form

$$CU_{ij} = \mathbb{1}^{\otimes(i-1)} \otimes |0\rangle\langle 0| \otimes \mathbb{1}^{\otimes(n-i)} + \mathbb{1}^{\otimes(j-1)} \otimes |1\rangle\langle 1| \otimes \mathbb{1}^{\otimes(j-i-1)} \otimes U \otimes \mathbb{1}^{\otimes(n-j)},$$

where $i < j$. Putting the Pauli X gate instead of U in (1.31), one obtains the CX gate called the controlled-not (or CNOT) gate, which is very prominent in quantum computing. The CX gate “flips” the state of the second qubit if the first one is $|1\rangle$, and does nothing if it is $|0\rangle$. This gate is of special usefulness when combined with

the Hadamard gate H , as it allows to produce a so-called Bell state

$$\begin{aligned}
 |\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle) \\
 &= |0\rangle \text{---} \boxed{H} \text{---} \bullet \text{---} \\
 &\quad |0\rangle \text{---} \text{---} \boxed{X} \text{---}
 \end{aligned} \tag{1.32}$$

which is a maximally entangled state of two qubits. Of great use is also the controlled U gate which acts oppositely,

$$\begin{aligned}
 LU(|0\rangle \otimes |\psi\rangle) &= |0\rangle \otimes U|\psi\rangle, \\
 LU(|1\rangle \otimes |\psi\rangle) &= |1\rangle \otimes |\psi\rangle,
 \end{aligned}$$

and is formally defined as

$$LU = |1\rangle\langle 1| \otimes \mathbb{1} + |0\rangle\langle 0| \otimes U = \text{---} \circ \text{---} = \text{---} \boxed{X} \text{---} \bullet \text{---} \boxed{X} \text{---} \tag{1.33}$$

Single-qubit rotations plus the CX gate form the so-called universal gate set. This means that any many-qubit unitary operation can be decomposed into a finite sequence of the gates from this set. If one possesses universal single-qubit rotations in a universal gate set, then this decomposition can be performed exactly. Moreover, as a consequence of the celebrated Solovay-Kitaev theorem, any single-qubit gate can be efficiently approximated to a given precision as a product of the Hadamard H , phase S , and the so-called $\pi/8$ -gate $T = e^{-i\pi\sigma_z/8}$. That is, for a $U \in \text{SU}(2)$ and some $\epsilon > 0$ we have

$$\left\| U - \prod_{j=1}^N U_j \right\| \leq \epsilon, \quad U_j \in \{H, S, T\},$$

where $N = O(\log^c(1/\epsilon))$ for some constant c . More details on universal gate sets and the proof of the Solovay-Kitaev theorem can be found e.g. in Section 4.5 and Appendix 3 of [64].

1.4.2 Parametrization of a quantum state

In Section 1.2.2 we mentioned that in the simplest case, one obtains one pure state from another via a unitary transformation, i.e.

$$|\varphi\rangle \longrightarrow |\psi\rangle = U |\varphi\rangle.$$

Then later in Section 1.4.1 we discussed that this unitary operator U can be exactly decomposed into a sequence of gates as

$$U = \prod_{j=1}^p U_j, \quad (1.34)$$

where $U_j \in \{R_k, R_l, CX\}$ with $k \neq l$. There we also showed that a gate can be parametrized, i.e. $U_j \equiv U_j(\theta_j)$ with $\theta_j \in \mathbb{R}$ (e.g. a Pauli rotation gate $R_j(\alpha) = e^{-i\alpha\sigma_j}$ with $j \in \{x, y, z\}$). This allows us to rewrite (1.34) as

$$U(\boldsymbol{\theta}) = \prod_{j=1}^p U_j(\theta_j), \quad (1.35)$$

where $\boldsymbol{\theta} = \{\theta_j\}_{j=1}^p \in \mathbb{R}^p$. In (1.35), we also assumed that each gate U_j contains only one parameter, and all non-parametrized gates (e.g. CX) are “embedded” into the parametrized ones. This leads us to the idea that a pure state $|\psi\rangle$ can be prepared by acting with a parametrized unitary operator $U(\boldsymbol{\theta})$ on some initial state $|\psi_0\rangle$, i.e.

$$|\psi\rangle \longrightarrow |\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta}) |\psi_0\rangle.$$

A parametrized in such a way probe state is called the variational state. This parametrization is widely used in variational quantum algorithms described later in Chapter 2.

The parametrized operator $U(\boldsymbol{\theta})$ is called the variational quantum circuit, and it can have a specific structure of placing the gates in the circuit. In this case, such a unitary is called an ansatz. More formally, we define the notion of the ansatz as follows.

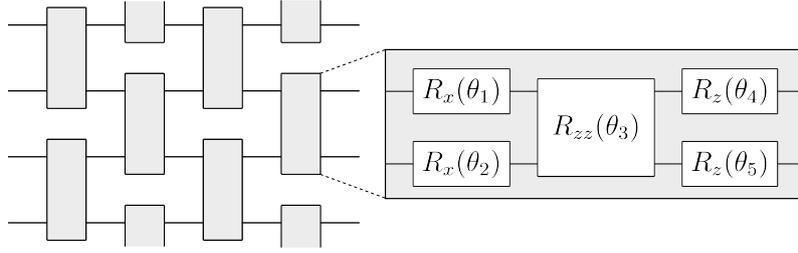


Figure 1-1: A four-qubit example of the checkerboard ansatz of four layers. On the right shown is also the two-qubit operator of five parameters which is put into a block. In it, $R_\sigma(\theta) = e^{-i\theta\sigma}$ and $R_{\sigma\tau}(\theta) = e^{-i\theta\sigma\otimes\tau}$ with $\sigma, \tau \in \{\sigma_x, \sigma_y, \sigma_z\}$ specifying the Pauli operator and $\theta \in [0, 2\pi)$ being the optimization parameters.

Definition 8 (Ansatz) An ansatz $A(\boldsymbol{\theta})$ is a parametrized unitary operator which has the form of a quantum circuit $V(\boldsymbol{\theta}_j)$ concatenated M times,

$$A(\boldsymbol{\theta}) = \prod_{j=1}^M V(\boldsymbol{\theta}_j),$$

where $\bigcup_{j=1}^M \boldsymbol{\theta}_j = \boldsymbol{\theta}$ and

$$V(\boldsymbol{\theta}_j) = \prod_{k=1}^L U_k(\theta_k)$$

with $\theta_k \in \mathbb{R}$ and $\bigcup_{k=1}^L \theta_k = \boldsymbol{\theta}_j$. The operator V is called the ansatz layer and the number M is called the ansatz depth.

The notion of ansatz is very important in applied variational quantum computing, e.g. in quantum chemistry [65–67] and quantum materials science [68–70]. In Figure 1-1 we show an example of the so-called checkerboard ansatz [71–73]. In Section 2.1.1, we will meet the so-called hardware-efficient ansatz, which — as can be established from its name — is implementable on any quantum computer provided that it is able to perform single-qubit rotations and any entangling gate.

1.5 Hamiltonian simulation

One of possible applications of quantum computers is the Hamiltonian simulation. Consider the evolution of a pure state $|\psi_0\rangle$ in time t under a unitary transformation:

$$|\psi(t)\rangle = U(t) |\psi_0\rangle. \quad (1.36)$$

Recalling the Stone's theorem (1.15), we get

$$|\psi(t)\rangle = e^{-iHt} |\psi_0\rangle.$$

Differentiating this equation with respect to t , one obtains the time-dependent Schrödinger's equation for pure states:

$$i \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle. \quad (1.37)$$

The Hermitian operator H is called the Hamiltonian, and it is considered as an “energy” operator. That is, measured in a quantum state, the expected value of a Hamiltonian is the energy of the quantum system in this state. The structure of a Hamiltonian is supposed to describe the quantum system under study. The equation (1.37) is very important in physics as it governs the dynamics of many models of physical quantum systems. For example, a chain of n interacting 1/2-spins placed in a transverse magnetic field can be described by the Ising Hamiltonian

$$H_{\text{Ising}} = J \left(\sum_{j=1}^n X_j X_{j+1} + h \sum_{j=1}^n Z_j \right), \quad (1.38)$$

where $J \in \mathbb{R}$ is the coupling strength between neighbouring spins and $h \in \mathbb{R}$ is the transverse field; in the equation above, we made use of the notation

$$A_j \equiv \mathbb{1}^{\otimes(j-1)} \otimes A \otimes \mathbb{1}^{\otimes(n-j)}$$

and applied the periodic boundary conditions $X_n X_{n+1} \equiv X_1 X_n$.

Obviously, (1.36) is a solution to the Schrödinger equation (1.37). It appears that a gate-based quantum computer could naturally simulate a unitary evolution of a given quantum state under a given Hamiltonian. The obstacles here are, first, that one has to be able to prepare the initial state $|\psi_0\rangle$, and, second, apply to it the unitary $U(t) = e^{-itH}$ as a gate. For the latter, in Section 1.4.1 we mentioned that any unitary can be decomposed into a sequence of single-qubit rotations and CX -gates. But the problem here is that the unitary $U(t)$ itself is unknown, known

only is its generating Hamiltonian H . To circumvent this, one may decompose the given Hamiltonian as a sum of L terms,

$$H = \sum_{k=1}^L H_k,$$

such that the gates of the form $U_k(t) = e^{-itH_k}$ are implementable on a given quantum hardware. For example, if one decomposes the Ising Hamiltonian (1.38) in such a way, the required gates would be $U_k(t) = e^{-ithZ_k}$ and $U_k(t) = e^{-itJX_kX_{k+1}}$; here, the first gate is just a single-qubit Pauli rotation, and the second gate is a Mølmer-Sørensen gate which is natural for trapped ions computers.

And here comes another obstacle in Hamiltonian simulation. The problem is that in general $H_kH_l \neq H_lH_k$, therefore

$$\exp\left\{-it \sum_{k=1}^L H_k\right\} \neq \prod_{k=1}^L \exp\{-itH_k\}.$$

However, one can approximate the exponential of a sum using e.g. the first-order p -step Suzuki-Trotter formula

$$\exp\left\{-it \sum_{k=1}^L H_k\right\} = \left(\prod_{k=1}^L \exp\left\{-\frac{it}{p}H_k\right\}\right)^p + O(L^2t^2/p^2) \quad (1.39)$$

As an instance, for the Ising Hamiltonian (1.38) one obtains

$$e^{-itH_{\text{Ising}}} \approx \left(\prod_{j=1}^n e^{-itJX_jX_{j+1}/p} \prod_{j=1}^n e^{-ithZ_j/p}\right)^p$$

For details on higher-order approximations, or on Hamiltonian simulation in general, see [74].

Chapter 2

Variational quantum computing

There is an important peculiarity in quantum computing: The input and output for a quantum algorithm are commonly quantum states. The output state is often such that it encodes the solution to the problem the algorithm is devised to solve. For example, the output of the Grover's algorithm [8] is a state $|x\rangle$ which encodes the solution to the equation $f(x) = 1$, where $f : \mathbb{B}^n \rightarrow \mathbb{B}$ and $x \in \mathbb{B}^n$ for $n \in \mathbb{N}$. In Section 1.4.2 we also described how one can parametrize a quantum state $|\psi\rangle \rightarrow |\psi(\boldsymbol{\theta})\rangle$ with $\boldsymbol{\theta} \in \mathbb{R}^p$, $p \in \mathbb{N}$.

This forms the core of so-called the variational quantum computing approach. In variational quantum algorithms, for some Hilbert space \mathcal{H} one defines a so-called cost function, the common form of which is

$$f : \text{Herm}(\mathcal{H}) \times \mathcal{D}(\mathcal{H}) \rightarrow \mathbb{R},$$

i.e. it maps pairs of Hermitian observables and quantum states into real numbers. A cost function is supposed to be designed in such a way that the minimization of it leads to the solution of a problem of interest. Most commonly, this function is the expected value of some observable in some (usually, pure) state. As a pure state can be parametrized by a set of real numbers $\boldsymbol{\theta}$, the function becomes parametrized as well, so $f \equiv f(\boldsymbol{\theta})$. Therefore, a variational quantum algorithm is designed in such a

way that a solution for the minimization problem

$$\boldsymbol{\theta}^{\text{opt}} \in \arg \min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) \quad (2.1)$$

encodes the solution of a problem of interest. In [39], the authors identify four criteria desirable for a cost function f :

1. the exact optimum of $f(\boldsymbol{\theta})$ corresponds a the solution of the original problem;
2. the function $f(\boldsymbol{\theta})$ can be efficiently estimated on a quantum computer;
3. the parameters $\boldsymbol{\theta}$ must be efficiently trainable.
4. the smallness of values of $f(\boldsymbol{\theta})$ indicate the closeness to the solution (operational meaning);

As mentioned earlier, generally, this function is based on measuring an observable in the variational state of the form (1.4.2). That is, the function $f(\boldsymbol{\theta})$ itself is evaluated (more accurately, estimated) on a quantum computer. At the same time, the parameters $\boldsymbol{\theta}$ are optimized on a classical computer using an existing minimization algorithm (e.g. gradient descent). Here rises the main peculiarity of variational quantum algorithms: they are hybrid quantum-classical in the sense that one needs a quantum and a classical computer to execute such an algorithm. In the next section, we describe the variational quantum eigensolver, one of the most prominent representatives of variational quantum algorithms.

2.1 Variational quantum eigensolver

In Section 1.5 we mentioned the time-dependent Schrödinger's equation. Let us now consider its stationary variant: for a given Hamiltonian $H \in \text{Herm}(\mathcal{H})$ one seeks to find a $|\psi\rangle \in \mathcal{H}$ such that

$$H |\psi\rangle = E |\psi\rangle, \quad (2.2)$$

which essentially is an eigenvalue-eigenvector problem. Often one is interested in the smallest eigenvalue and its corresponding eigenvector of a given Hamiltonian

H , and these are called the ground energy E_g and ground state $|\psi_g\rangle$, respectively. Physically, this corresponds to the state of a system at rest, and to its energy in this state. That is, in solving the stationary Schrödinger's equation (2.2), one wants to obtain

$$E_g = \min_{\psi} \langle \psi | H | \psi \rangle, \quad (2.3)$$

$$|\psi_g\rangle = \arg \min_{\psi} \langle \psi | H | \psi \rangle. \quad (2.4)$$

Namely, by varying the probe state ψ , one wants it to become the ground state of a given Hamiltonian H . This is actually a direct application of the Rayleigh–Ritz method of approximating an eigenvalue of a matrix [75]. In Section 1.4.2, we discussed that a pure quantum state can be parametrized via the action of a parametrized unitary operator on an initial state, i.e. $|\psi\rangle \rightarrow |\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_0\rangle$ with $U(\boldsymbol{\theta}) \in U(\mathcal{H})$ and $\boldsymbol{\theta} \subset \mathbb{R}$. Therefore, the minimization problems (2.3) and (2.4) can be rewritten as

$$E_g = \min_{\boldsymbol{\theta}} \langle \psi(\boldsymbol{\theta}) | H | \psi(\boldsymbol{\theta}) \rangle, \quad (2.5)$$

$$|\psi_g\rangle \equiv |\psi(\boldsymbol{\theta}^{\text{opt}})\rangle, \quad \boldsymbol{\theta}^{\text{opt}} = \arg \min_{\boldsymbol{\theta}} \langle \psi(\boldsymbol{\theta}) | H | \psi(\boldsymbol{\theta}) \rangle. \quad (2.6)$$

That is, we transitioned from varying over state vectors $\psi \in \mathcal{H}$ to varying over real numbers $\boldsymbol{\theta} \subset \mathbb{R}$.

The variational state can be prepared by acting with gates from a universal set implemented on a given hardware. But what about measuring the expected value of a given Hamiltonian in this state? In Section 1.4 we mentioned that a measurement is commonly carried out in the computational basis, i.e. for each qubit measured is the σ_z observable. In circuit (1.28), we showed how one can rotate the basis in order to measure σ_x and σ_y . When it comes to measuring an arbitrary Hermitian observable H , one may notice that, similarly to (1.11), the space of all Hermitian $2^n \times 2^n$ -matrices is

$$\text{Herm}(\mathbb{C}_2^{\otimes n}) = \text{span}_{\mathbb{R}}\{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}, \quad n \in \mathbb{N}.$$

Therefore, any Hermitian matrix of the size $2^n \times 2^n$ can be written as

$$H = \sum_{K \in \{0,x,y,z\}^{\times n}} a_K S_K, \quad (2.7)$$

$$a_K \in \mathbb{R}, \quad S_K = \bigotimes_{j \in K} \sigma_j.$$

In other words, an observable on n qubits can be represented as a sum of tensor products of Pauli operators called Pauli strings. With this decomposition for an n -qubit Hermitian operator H , we can rewrite (2.5) and (2.6) as

$$E_g = \min_{\boldsymbol{\theta}} \left\{ \sum_{K \in \{0,x,y,z\}^{\times n}} a_K \langle \psi(\boldsymbol{\theta}) | S_K | \psi(\boldsymbol{\theta}) \rangle \right\}, \quad (2.8)$$

$$|\psi_g\rangle \equiv |\psi(\boldsymbol{\theta}^{\text{opt}})\rangle, \quad \boldsymbol{\theta}^{\text{opt}} = \arg \min_{\boldsymbol{\theta}} \left\{ \sum_{K \in \{0,x,y,z\}^{\times n}} a_K \langle \psi(\boldsymbol{\theta}) | S_K | \psi(\boldsymbol{\theta}) \rangle \right\}. \quad (2.9)$$

As we see in (2.8), to calculate the expected value of an arbitrary Hamiltonian H , we represent it as a sum of Pauli strings S_K as shown in (2.7), compute the expected values of each individual Pauli string, and sum these expectations with the corresponding coefficients a_K . In Section 1.2.4, we mentioned that for estimating an expected value with a precision ϵ , one needs to perform $N = O(1/\epsilon^2)$ measurements. Therefore, for M Pauli strings, the required number of measurements would be $N = O(M/\epsilon^2)$ [54]. The issue here is that in the worst case we have $M = 4^n$ Pauli strings to measure. Therefore, it is commonly assumed that a Hamiltonian has a number of terms polynomial in the number of qubits n [38], which often holds for physically relevant Hamiltonians (for example, the Ising Hamiltonian (1.38) consists $2n$ terms). Additionally, one may consider representing a given Hamiltonian in the form

$$H = \sum_j H_j,$$

where each

$$H_j = \sum_K a_K S_K, \quad a_K \in \mathbb{R},$$

consists of commuting Pauli strings, i.e. $[S_K, S_{K'}] \equiv S_K S_{K'} - S_{K'} S_K = 0$. The

reason of doing this is that if two observables commute, then they share the same eigenbasis and therefore — with some effort — can be measured simultaneously [76].

2.1.1 Experimental realizations of VQE

The first application of VQE

The equations (2.8) and (2.9), in fact, incarnate the idea of the variational quantum eigensolver. Historically, this algorithm was first described and experimentally implemented in the work [37]. There, the authors found the bond dissociation energy of the molecule He-H⁺. The target Hamiltonian H was the full configuration interaction Hamiltonian for this system, and this Hamiltonian was adopted for qubits via the Jordan-Wigner transformation [77] (for details, see the original work [37] and its supplemental material). As a quantum computer, used was a photonic chip which implements qubits as path-encoded photons. Such a setup allows to apply CX gates and arbitrary single-qubit rotations via beamsplitters and phaseshifters.

Hardware-efficient ansatz and superconducting qubits

Another interesting work on the implementation of VQE is [78]. There, the authors proposed the notion of the hardware efficient ansatz (HEA) for VQE executed on a quantum computer based on superconducting qubits. The idea of HEA is to alternate arrays of single-qubit operations with many-qubit, or entangling operations. An important property of this ansatz is that it does not matter how the many-qubit operators are implemented and what form they have, important only is the ability of introducing a sufficient amount of entanglement into the quantum state. This is where the name “hardware-efficient” comes from: single-qubit operations are assumed to be implementable on modern-day quantum computers, while it does not matter how implemented is the entangling operator. In Figure 2-1, shown is an example of one layer of HEA for 4 qubits. In this Figure, as one may notice, the many-qubit entangling operator is implemented as a cascade of the CX gates. One can put several layers of this ansatz to increase its expressive power, i.e. the ability of preparing quantum states of wider ranges (in other words, to “cover” wider regions

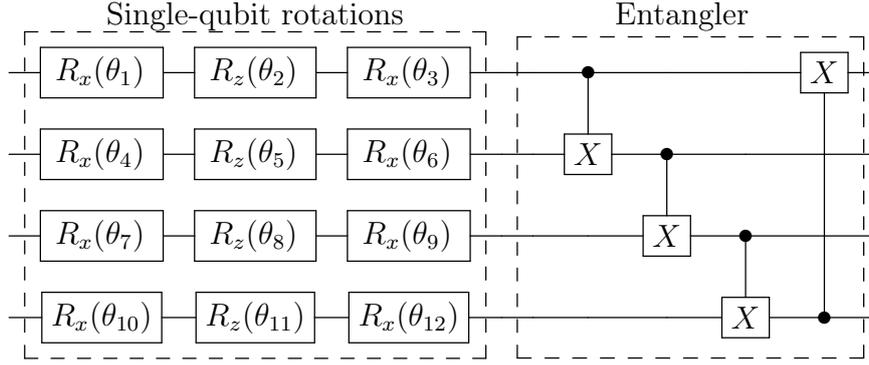


Figure 2-1: A layer of a hardware-efficient ansatz for 4 qubits with $p = 12$ variational parameters. Here, $R_\sigma(\theta) = e^{-i\theta\sigma}$ with $\sigma \in \{\sigma_x, \sigma_y, \sigma_z\}$ specifying the Pauli operator and $\theta_j \in [0, 2\pi)$ being the optimization parameters. The controlled-X operators “flip” the target qubit depending on the state of the control qubit, see (1.31).

of the Hilbert space).

Ion traps qubits

In [79], the authors performed the VQE algorithm for Schwinger’s Hamiltonian adopted for N qubits via the Jordan-Wigner transformation [77]. In Pauli basis, this Hamiltonian is written as

$$H_{\text{Sch}} = w \sum_{j=1}^{N-1} [\sigma_+^j \sigma_-^{j+1} + \sigma_-^{j+1} \sigma_+^j] + \frac{m}{2} \sum_{j=1}^N (-1)^j \sigma_z^j + g \sum_{j=1}^N L^j, \quad (2.10)$$

where $\sigma_\pm = \sigma_x \pm i\sigma_y$ and

$$L^j = \epsilon_0 - \frac{1}{2} \sum_{l=1}^j [\sigma_z^l + (-1)^l].$$

This Hamiltonian describes electron-positron pair creation and annihilation and their interaction. The first term is responsible for the interaction of an electron and a positron, the second depends on the bare mass m of the particles, and the third stands for the energy of the electric field with ϵ_0 being its strength.

In the original experiment, the assignments for the parameters were $w = g = 1$ and $\epsilon_0 = 0$, and the parameter m was varied so that the ground state was found for different values of this parameter. The physical realization of the quantum computer was the ion traps. A hardware-efficient ansatz was used for preparing the variational

state. The entangling block of the HEA was implemented as $U_{\text{ent}} = e^{-i\theta H_{XY}}$, where

$$H_{XY} = \sum_{j=1}^N J_i (\sigma_+^j \sigma_-^{j+1} + \sigma_-^{j+1} \sigma_+^j) + B \sum_{j=1}^N \sigma_x^j,$$

which is called the XY Hamiltonian; it describes the interaction of ions when they are shun by a laser beam of a specific configuration (for details, see the original work [79]).

Photon polarization qubits

In [1], we also experimentally applied the VQE algorithm for the two-qubit version of the Schwinger's Hamiltonian (2.10). There, the qubit states were encoded into the polarization of a photon. That is, the vertical polarization encodes the $|0\rangle$ while the horizontal polarization encodes $|1\rangle$. On such an optical setup, one can naturally implement the following transformation:

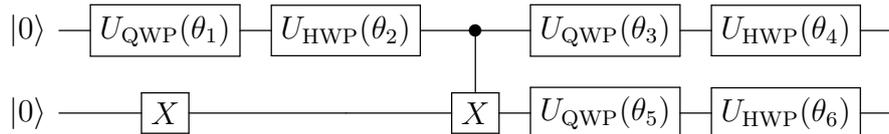
$$U(\theta, \delta) = V(\theta)D(\delta)V^\dagger(\theta), \quad (2.11)$$

$$V(\theta) = \mathbb{1} \cos(\theta) - i\sigma^y \sin(\theta), \quad D(\delta) = e^{i\delta|1\rangle\langle 1|}.$$

The unitary (2.11) describes the half-wave plate (HWP) and the quarter-wave plate (QWP) operators defined as

$$U_{\text{HWP}}(\theta) \equiv U(\theta, \pi), \quad U_{\text{QWP}}(\theta) \equiv U(\theta, \pi/2),$$

which are used to prepare a two-qubit variational state produced by the following circuit:



The ground state energy of Schwinger's Hamiltonian (2.10) found by VQE in this setting is shown in Figure 2-2a. As can be seen, experimentally obtained energies are found with good precision.

For zero background field, $\epsilon_0 = 0$, the Schwinger model (2.10) is known to experience a phase transition at the point $m \approx -0.7$ [79, 80]. This point is also special for the so-called order parameter observable [79]

$$O = \frac{1}{2n(n-1)} \sum_{j>i} \left(1 + (-1)^i \sigma_z^i\right) \left(1 + (-1)^j \sigma_z^j\right). \quad (2.12)$$

In our case of $n = 2$ qubits, this observable is simply the projector $O = |01\rangle\langle 01|$. For large negative m , the ground state is expected to be $|\psi_g\rangle = |01\rangle$, which translates to $\langle \psi_g | O | \psi_g \rangle = 1$; for large positive m , the ground state is $|\psi_g\rangle = |10\rangle$ giving $\langle \psi_g | O | \psi_g \rangle = 0$. The value $m = -0.5$ can be considered as a critical point of the order parameter O , as at this point the ground state is precisely

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle),$$

which is called the singlet state, and which gives $\langle \Psi^- | O | \Psi^- \rangle = 0.5$. An important feature of this state is that $\forall U \in \text{U}(\mathbb{C}_2)$ one has $(U \otimes U) |\Psi^-\rangle = |\Psi^-\rangle$, i.e. this state is invariant under local transformations of the form $U \otimes U$. This means that there could be a region in the space of variational parameters such that it has a “flat valley”, in which the energy does not change. Figure 2-2 supports this statement: in the vicinity of the point $m = -0.5$, at which the energy landscape is flat for some of the variational parameters, the variance of the results is higher than that for more distant points. Additionally, some energies are wrongly found to be $-3/2$, which is the ground state energy for $m = -0.5$.

Inter alia, in [1] we also investigated how controlled dephasing noise affects the results obtained by VQE. In the described setup, one can artificially introduce noise

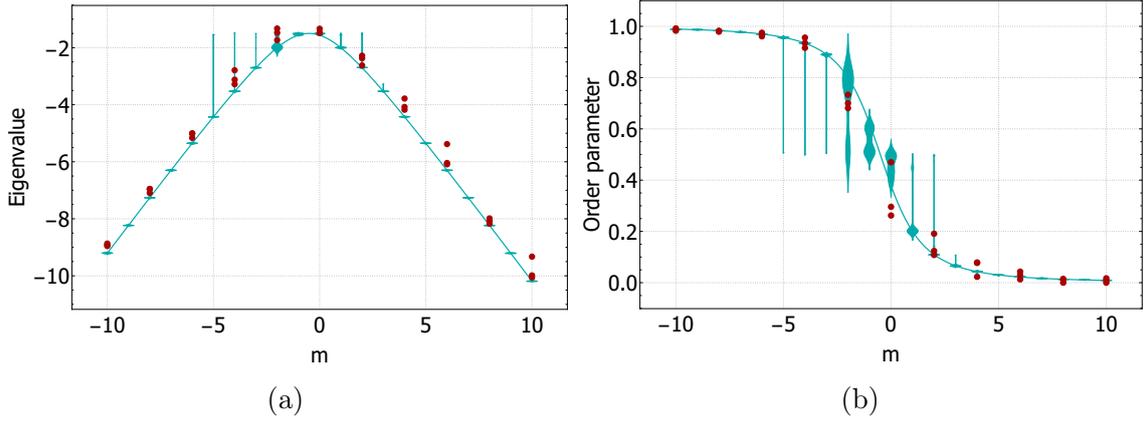


Figure 2-2: Eigenvalue (a) and order parameter (b) versus bare mass m of Shwinger's Hamiltonian (2.10). Solid lines stand for the analytical solution, cyan points (violin plot) show the results of simulations, and red points are obtained from the experiment. For each m , the data for simulations are obtained as the average over 30 trials. For the experiment, all attempts are shown as distinct points. The figures are taken from [1].

described by the following channel:

$$\mathcal{E}(\rho) = \sum_{j=1}^2 E_j \rho E_j^\dagger, \quad (2.13)$$

$$E_i = V(\theta) D_i(\delta) V^\dagger(\theta), \quad (2.14)$$

$$D_1(\delta) = \sqrt{\frac{2-\epsilon}{2}} \begin{pmatrix} e^{i\delta} & 0 \\ 0 & 1 \end{pmatrix}, \quad D_2(\delta) = \sqrt{\frac{\epsilon}{2}} \begin{pmatrix} e^{i\delta} & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.15)$$

Physically, this noise comes from placing a liquid crystal variable retarder (LCVR) before the measurement apparatus. LCVRs allow to change the phase of the specific polarization of a photon depending on the external applied voltage. In (2.13)-(2.15), the parameter θ controls the axis angle of the LCVR, δ is called the mean optical retardance, and the value $0 \leq \epsilon \leq 1$ is the noise strength. For the experiment in [1], we set $\theta = \pi/4$ and $\delta = 2\pi$, while the noise strength ϵ was varied.

As can be seen in Figure 2-3, even in the presence of dephasing noise we can detect the critical point of the order parameter in the Schwinger model, i.e. the point at which the expected value of (2.12) rapidly changes. Moreover, one can also observe that the profile of the obtained ground state energy curves remains the same.

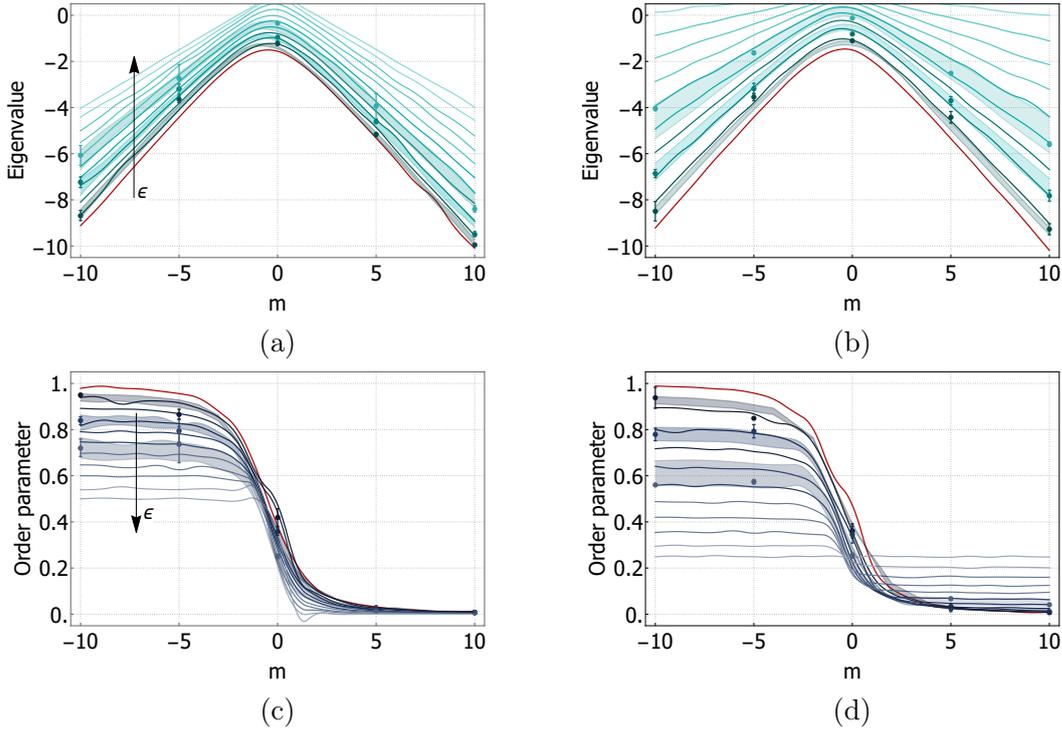


Figure 2-3: Noise simulations for dephasing noise (2.13) acted on one qubit (a), (c) and both qubits (b), (d). Panels (a) and (b) show the dependence of the ground state energy of Schwinger’s Hamiltonian (2.10) on the bare mass m , and panels (c) and (d) indicate the dependence of the order parameter (2.12). Red lines correspond to simulations without noise, and the color blur corresponds to the increase of noise strength ϵ from 0.1 to 1 in 0.1 steps. Distinct points stand for experimental results, and solid areas show theoretical predictions for points with different noise level. The plots are taken from [1].

2.1.2 Soliton solution: a case study

In this section, we discuss the solution obtained by VQE for a Hamiltonian with a highly entangled ground state, which we studied in [2]. Consider the Hamiltonian of a one-dimensional chain of N quantum spins \mathbf{S}_j with the Dzyaloshinskii-Moria interaction (DMI):

$$H_{\text{DMI}} = -J \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j - \sum_{\langle i,j \rangle} \mathbf{D}_{ij} \cdot (\mathbf{S}_i \times \mathbf{S}_j) + \sum_{j=1}^N \mathbf{B} \cdot \mathbf{S}_j. \quad (2.16)$$

Here, the first term describes the so-called direct exchange interaction which for $J > 0$ favors collinear ordering of spins; the index j labels the spins along their position on the z axis, and $\langle i, j \rangle$ implies summation over nearest neighbors. The second term

specifies the DMI, which destroys homogeneity of collinear magnetic ordering: it promotes spin canting between neighboring sites; the strength of DMI is determined by the Dzyaloshinskii vector \mathbf{D}_{ij} . The third term stands for the transverse magnetic field \mathbf{B} . In this field, the competition between DMI and exchange interaction results in a non-collinear configuration of spins [81–83]. The Hamiltonian H_{DMI} defined in (2.16) describes a reliable model of a wide class of so-called chiral magnets [84–86].

Considering spin one-half particles, one has the spin operators of the form $\mathbf{S} = \boldsymbol{\sigma}/2$ with $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ being the Pauli vector. The state of a spin can be described by the vector $|S\rangle = (\cos \frac{\theta}{2}, e^{i\varphi} \sin \frac{\theta}{2})^T$, which gives

$$\begin{aligned} \langle S | \hat{\mathbf{S}}_i | S \rangle &= \mathbf{n}_i / 2, \\ \mathbf{n}_i &= (\cos \varphi_i \sin \theta_i, \sin \varphi_i \sin \theta_i, \cos \theta_i). \end{aligned}$$

In what follows, we assume that the Dzyaloshinskii vector is aligned along the z axis, so $\mathbf{D}_{ij} = D\hat{\mathbf{e}}_z$, where D determines the strength of DMI; we also set the magnetic field to be aligned along the x axis, i.e. $\mathbf{B} = B\hat{\mathbf{e}}_x$ with B being expressed in energy units.

In the basis $|S_1, S_2, \dots, S_N\rangle$, the quantum Hamiltonian (2.16) can be mapped to a classical Heisenberg-type model of interacting spins,

$$H = -\frac{J}{4} \sum_{\langle i,j \rangle} \mathbf{n}_i \cdot \mathbf{n}_j - \frac{D}{4} \sum_{\langle i,j \rangle} (\mathbf{n}_i \times \mathbf{n}_j)_z - \frac{B}{2} \sum_{j=1}^N n_j^x. \quad (2.17)$$

Considering the continuous description of this model, i.e. $\mathbf{n}_i = \mathbf{n}(z)$ so that $\varphi_i = \varphi(z)$ and $\theta_i = \theta(z)$ now depend on the coordinate z , one can describe it in terms of magnetization specified by the vector field

$$\mathbf{n}(z) = (\cos \varphi(z) \sin \theta(z), \sin \varphi(z) \sin \theta(z), \cos \theta(z)).$$

Let a be the distance between a pair of neighboring spins and $\mathbf{n}_{i+1} = \mathbf{n}(z+a)$, and assume $\mathbf{n}(z+a) \approx \mathbf{n}(z) + a\mathbf{n}'(z) + a^2\mathbf{n}''(z)/2$. In (2.17), we can now replace summation by integration, $\sum_j \rightarrow \frac{1}{a} \int_0^L dz$, where L is the total length of the spin

chain, and in the lowest order in a we can write

$$H = \frac{aJ}{8} \int_0^L dz [\theta'^2 + \varphi'^2 \sin^2 \theta - k_0 \varphi' \sin^2 \theta + 2m^2 \cos \varphi \sin \theta], \quad (2.18)$$

where $k_0 = D/(aJ)$ and $m^2 = 2B/(a^2J)$. One can show that the ground state energy of the Hamiltonian (2.18) is obtained when $\theta = \pi/2$, and φ is given by the sine-Gordon equation [81–83],

$$\varphi'' + m^2 \sin \varphi = 0. \quad (2.19)$$

Integrating this equation, one gets

$$\varphi = 2 \operatorname{am}(mz/\kappa, \kappa), \quad (2.20)$$

where κ is called the elliptic modulus to be found later, and $\operatorname{am}(u, \kappa)$ is the Jacobi amplitude determined by $\operatorname{sn} u = \sin \operatorname{am}(u, \kappa)$ with $\operatorname{sn} u$ defining the elliptic sine. The solution (2.20) can be shown to correspond to a so-called soliton lattice with the spatial periodicity,

$$\ell = \frac{2\kappa}{m} \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - \kappa^2 \sin^2 \varphi}} = \frac{2\kappa}{m} K(\kappa), \quad (2.21)$$

where $K(\kappa)$ is the complete elliptic integral of the first kind. Putting (2.20) into (2.18) we obtain the energy of the soliton lattice over a period

$$\varepsilon = \frac{am^2J}{2} \left(\frac{2}{\kappa^2} \frac{E(\kappa)}{K(\kappa)} - \frac{1}{\kappa^2} - \frac{\pi}{2m} \frac{k_0}{\kappa K(\kappa)} \right), \quad (2.22)$$

where we introduced the complete elliptic integral of the second kind $E(\kappa) = \int_0^{K(\kappa)} \operatorname{dn}^2 z dz$ with $\operatorname{dn} z$ being the delta amplitude. The value of κ which minimizes the energy ε can be obtained from

$$\pi \kappa k_0 = 4mE(\kappa), \quad (2.23)$$

which is derived using the fact that $\kappa E'(\kappa) = E(\kappa) - K(\kappa)$ and $\kappa K'(\kappa) = E(\kappa)/(1 -$

$\kappa^2) - K(\kappa)$.

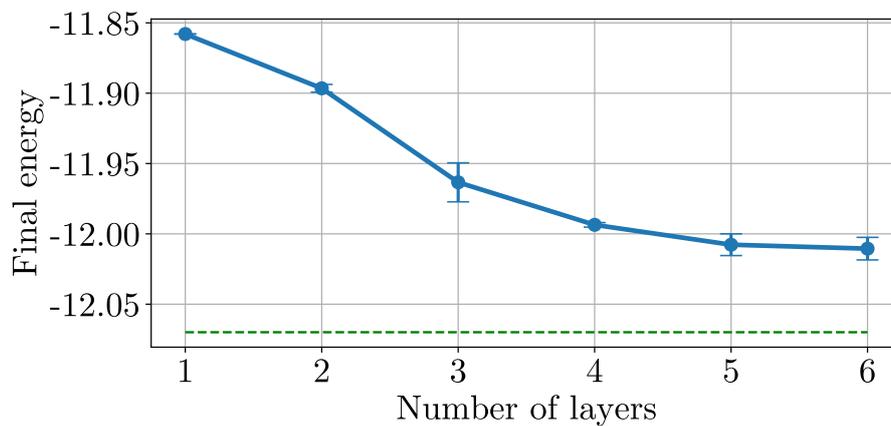
In [2], we applied the VQE algorithm for the Hamiltonian (2.16). For our numerical experiments, we set $D/J = 0.63$, and $B/J = 3.36 \times 10^{-3}$, which, as will be seen later, results in spiral arrangement of the spins. Solving (2.23) gives $\kappa \approx 0.256$ which corresponds to $N = \ell/a \approx 10$. That is, we have to consider $N = 10$ qubits to properly address one period of the soliton lattice. In our numerical experiments, we represent the variational state by the hardware-efficient ansatz shown in Figure 2-1, but with the CX replaced by CR_y gates. To simulate quantum circuits, we used the Qiskit package [87]. As the optimization method, we applied the BFGS algorithm [88] restricting the maximum number of iterations to 50,000.

In Figure 2-4, we show the ground state energy and fidelity of the ground state (i.e. the overlap of this state with the exact one) of the Hamiltonian (2.16) found by VQE versus the number of ansatz layers l . As one might expect, the more layers of the ansatz we have (i.e. the more expressive ansatz is), the more accurate is the found energy. Indeed, we observe discrepancy between the approximated result and the exact one to be less than 1%. Moreover, we apply a simple criterion discussed in [1, 89] for quantifying the precision of a VQE solution. That is, if E_0 and E_1 are the respective exact ground state and the first excited energies, and E^{VQE} is the ground state energy found by VQE, then this solution can be accepted if $\delta = (E^{\text{VQE}} - E_0)/(E_1 - E_0) < 1$. With $l = 6$ layers of the hardware-efficient ansatz, we obtained a solution with the quality $\delta \approx 0.6841$.

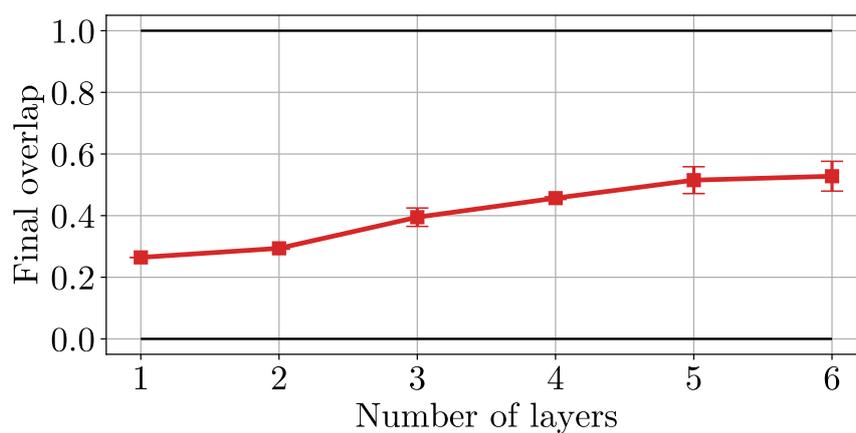
Despite that the ground state energy is found with high precision, the performance of VQE in terms the state fidelity is rather poor. This can be attributed to the entanglement properties of the ground state describing the spin configuration of the Hamiltonian (2.16). Let ρ_{ij} be the reduced density operator of the two qubits i and j . The entanglement of a two-qubit state can be quantified by the concurrence

$$C_{ij} = \max\{0, \sqrt{\lambda_1} - \sqrt{\lambda_2} - \sqrt{\lambda_3} - \sqrt{\lambda_4}\}, \quad (2.24)$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ are the eigenvalues of the non-Hermitian operator $R_{ij} = \rho_{ij} \tilde{\rho}_{ij}$ in increasing order. Here, $\tilde{\rho}_{ij} = (\sigma_y \otimes \sigma_y) \rho_{ij}^* (\sigma_y \otimes \sigma_y)$ is the spin-flipped density



(a)



(b)

Figure 2-4: Ground state energy (a) and the overlap of the approximate ground state vector with the exact one (b) of the Hamiltonian (2.16) for $N = 10$ found by VQE. In panel (a), the dashed green line stands for the exact ground state energy. Each data point is obtained as an average over 5 optimizations initialized with random parameters of the ansatz. The error bars show the standard deviation. The figures are taken from [2].

matrix with asterisk indicating the complex conjugation.

In Figure 2-5, we plot the concurrence C_{ij} of the ground state of the Hamiltonian (2.16) of $N = 10$ qubits. The numerical results show that the highest concurrence is observed for nearest-neighbouring spins. As we see, the solution obtained with VQE captures the entanglement of these spins, while the concurrence between distant qubits is not very well reproduced. In [2], we also showed that this lack of accuracy does not arise from trainability issues. Specifically, with some assignments for D and B for the Hamiltonian (2.16), the nearest-neighbouring spins can be highly entangled, yet with no entanglement present between the next-neighbouring ones. The ground state in this case can be found with high precision, and the performance of VQE with the hardware-efficient ansatz seems to be dependent on how the entanglement is “smeared” between spatially separated sites of the spin chain. We therefore conclude that a soliton solution found analytically describes a ground state with long-range entanglement, and this state cannot be well-reproduced with VQE without a sufficiently large number of ansatz layers. This limits the application range of VQE executed with a shallow ansatz for Hamiltonians describing short-range spin configurations.

Despite that the ground state vector found by VQE is of low fidelity, one nonetheless is still able to extract some physical properties of interest from it. To illustrate this, in Figure 2-6 we show the magnetic texture of the ground state plotted versus the number of ansatz layers used in VQE. In this figure, arrows positioned along horizontal lines represent the directions of spins in the chain. The moments are placed in the xy plane since their z components are zeros, which is in full agreement with the analytical results. In Figure 2-6, shown also is the spin configuration given by the analytical solution $\theta = \pi/2$ and $\varphi = 2\text{am}(mz/\kappa, \kappa)$. We expect that increasing the number of layers should lead to a solution close to the analytical one. What is interesting here is that the directions of the spins given by the VQE solutions are not correct, but nonetheless we capture the spiral structure of magnetic moments. That is, even using a shallow-circuit ansatz, the magnetic texture can be studied at least qualitatively.

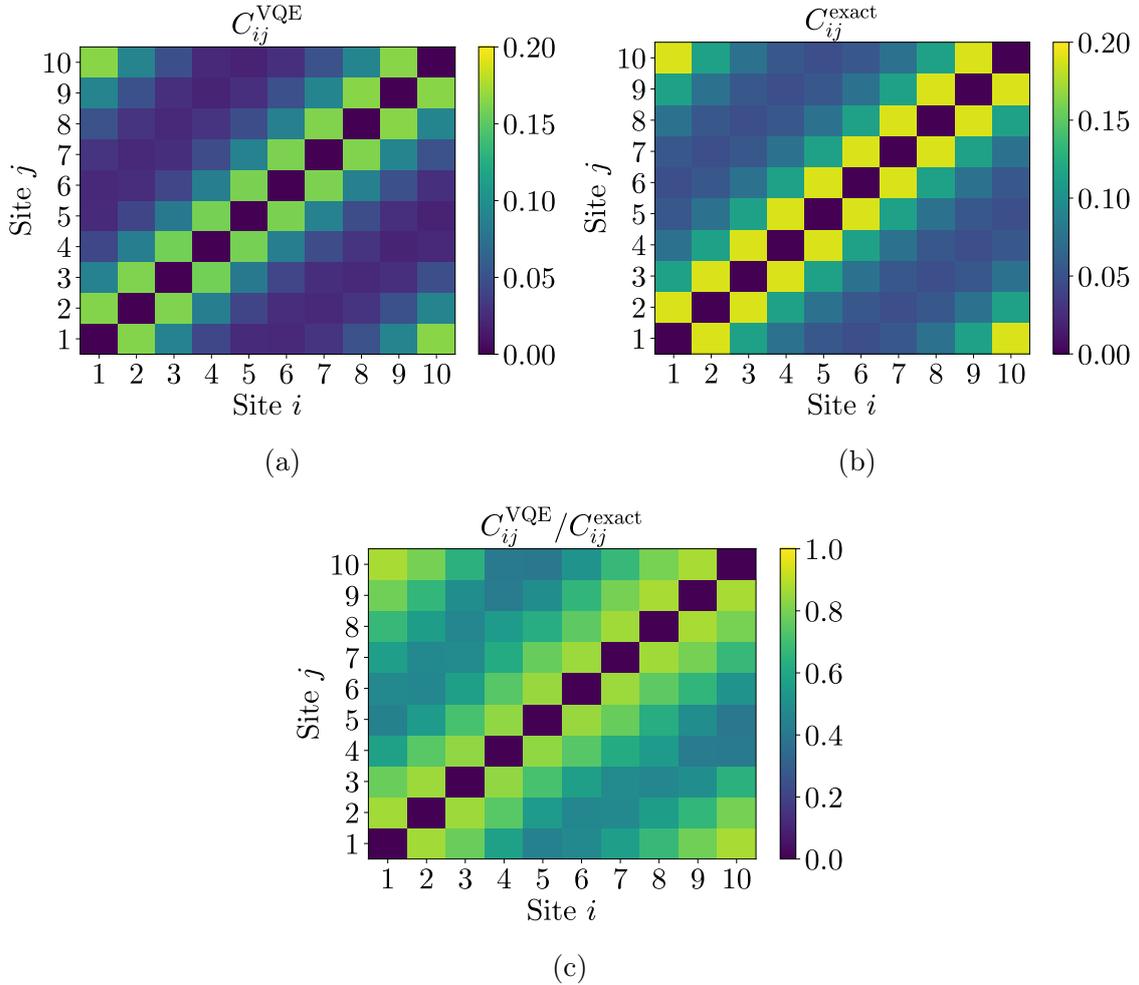


Figure 2-5: Concurrence C_{ij} (2.24) between the i th and j th qubits of the ground state of the DMI Hamiltonian (2.16). Panels (a) and (b) show the concurrence calculated for the VQE solution C_{ij}^{VQE} obtained with six ansatz layers and the exact ground state C_{ij}^{exact} , respectively. Panel (c) gives the ratio $C_{ij}^{\text{VQE}}/C_{ij}^{\text{exact}}$. As can be seen, the more distant two spins are, the worse the entanglement is captured by VQE. The figures are taken from [2].

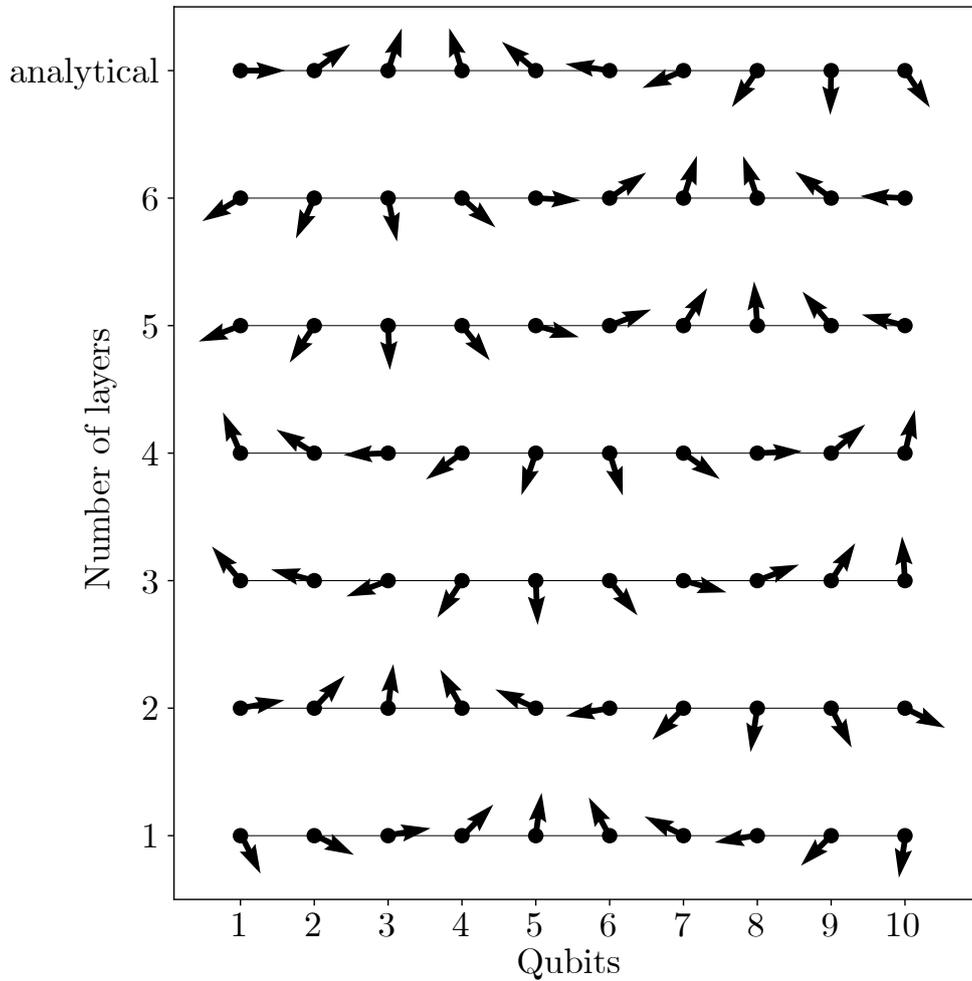


Figure 2-6: Magnetic texture of the ground state of the Hamiltonian (2.16) of $N = 10$ spin one-half particles versus the number of layers of the hardware-efficient ansatz. Each line corresponds to the magnetic moments of spins in the ground state found by VQE with a specific number of layers of the ansatz, and each arrow shows the magnetic moment of a given spin in the xy plane (z components are equal to zero). The uppermost line of magnetic moments is given by the analytical solution 2.20. The figure is taken from [2].

Chapter 3

Variational algorithms and tensor diagrams

In Chapter 2, we discussed the variational quantum eigensolver (VQE) algorithm, one of the most prominent and promising methods of quantum computing for NISQ devices. The core of this algorithm is formed by computing the expected value of a Hermitian operator H in a pure state $|\psi\rangle$, i.e. $\langle\psi|H|\psi\rangle$. In this chapter, we consider some other existing variational algorithms based on computing different expected values.

The algorithms will be formulated in a common way with equations, and also in the notation of tensor networks introduced in Section 1.3. The tensor network language is a way to represent tensor contraction operations in a pictorial and very explanatory way. In this notation, the manipulation with equations and indices of tensors is replaced by manipulating the wires and blocks representing tensors and contraction operations. Tensor networks are widely used in e.g. theoretical physics, classical computer science, and machine learning. So, the formulation in this language would help researchers from other fields of science understand quantum algorithms. Additionally, the tensor diagrams notation allows to notice interesting and often important details about the algorithm formulated. More specifically, *inter alia*, we will show that the variational quantum algorithms we review in this chapter can be reduced to VQE executed for a certain Hamiltonian and a pure variational state.

3.1 State preparation verification for VQE

We begin reviewing variational quantum algorithms with a method for state preparation verification. In [79], of particular interest is the proposed method for certification of the preparation of a desired state (the authors considered the ground state of Schwinger’s Hamiltonian (2.10)). For it, at each minimization iteration of the VQE algorithm, it was proposed to compute the the following quantity:

$$F_H(\psi) = \langle \psi | H^2 | \psi \rangle - \langle \psi | H | \psi \rangle^2, \quad (3.1)$$

where H is a Hamiltonian on n qubits. Essentially, the expression (3.1) is the variance of the energy measured for the Hamiltonian H in the state $|\psi\rangle$. One can prove that $F_H(\psi) = 0$ if and only if $H|\psi\rangle = \lambda|\psi\rangle$, i.e. if $|\psi\rangle$ is an eigenvector of H . Since the ground energy of the target Hamiltonian H is commonly unknown, then, having measured $\langle \psi | H | \psi \rangle$ at each step of the VQE algorithm, one does not know how close is the current obtained solution vector to the exact one. At the same time, the closeness of F_H to zero can serve as an indicator of the closeness of the approximate ground state vector to the true one. It must be also pointed out that it is relatively easy to compute this function. Indeed, the term $\langle \psi | H | \psi \rangle^2$ is trivially obtained after each step of VQE, so one needs only to compute $\langle \psi | H^2 | \psi \rangle$ by decomposing the Hermitian operator H^2 in Pauli basis (see (2.7)) and measuring it as an ordinary observable.

An alternative way of certifying the preparation of an eigenstate was proposed by us in [3]. There, we define the function

$$F_Q(\psi) = |\langle \psi | Q | \psi \rangle|^2, \quad (3.2)$$

where $Q = e^{-iHt}$ with H being a target Hamiltonian for VQE and $t \in \mathbb{R}$. Similarly to F_H , the function F_Q outputs unity if and only if its argument ψ is an eigenvector of H . In this formulation, one can use the tensor network notation to obtain an interesting equivalent expression. In Fig. 3-1, we apply the property (1.23) for

obtaining the formula

$$F_Q(\psi) = \langle \Psi | M | \Psi \rangle,$$

where

$$M = |\Phi^+\rangle\langle\Phi^+|, \quad (3.3)$$

$$|\Psi\rangle = Q |\psi\rangle \otimes |\bar{\psi}\rangle. \quad (3.4)$$

with

$$|\Phi^+\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^d |j\rangle \otimes |j\rangle$$

being a maximally entangled state of n qubits, which is a generalization of the Bell state (1.32) for $d = 2^n$. Obviously, the Hamiltonian M has the vector $|\Phi^+\rangle$ as its unique highest-energy state with the eigenvalue 1, and the rest of the eigenvectors have the the eigenvalue 0. Therefore, one can run the VQE algorithm to find the eigenvector which corresponds to the maximal eigenvalue of M , which is equivalent to maximizing the cost function (3.2); the probe state for VQE should be of the form (3.4).

Interestingly, to execute the VQE, there is no need to decompose the Hamiltonian (3.3) in the Pauli basis. Since we know the how to prepare the state $|\Phi^+\rangle$, see (1.32), we can directly measure the observable M on a conventional quantum computer. If $|\Phi^+\rangle = F(|\mathbf{0}\rangle \otimes |\mathbf{0}\rangle)$, where $|\mathbf{0}\rangle = |0\rangle^{\otimes n}$, then it is needed only to prepare the state $|\varphi\rangle = F^\dagger(Q |\psi\rangle \otimes |\bar{\psi}\rangle)$ and measure it in the computational basis. The probability of obtaining only zeros as a result of this measurement is

$$\begin{aligned} p_0 &= \text{Tr} |\varphi\rangle\langle\varphi| |\mathbf{0}\rangle\langle\mathbf{0}|^{\otimes 2} = |\langle\mathbf{0}|^{\otimes 2} |\varphi\rangle|^2 \\ &= |(\langle\mathbf{0}| \otimes \langle\mathbf{0}|) F^\dagger(Q |\psi\rangle \otimes |\bar{\psi}\rangle)|^2 = |\langle\Phi^+ | \Psi\rangle|^2 \\ &= \langle \Psi | M | \Psi \rangle = |\langle\psi | Q |\psi\rangle|^2. \end{aligned}$$

The calculations simplify even more if one assumes the variational state to be of the form $|\psi\rangle = U^\dagger Q U |\mathbf{0}\rangle$, where U is a variational ansatz. Indeed, consider the Hamiltonian $M = |\mathbf{0}\rangle\langle\mathbf{0}|$, for which the vector $|\mathbf{0}\rangle$ is the unique highest-energy

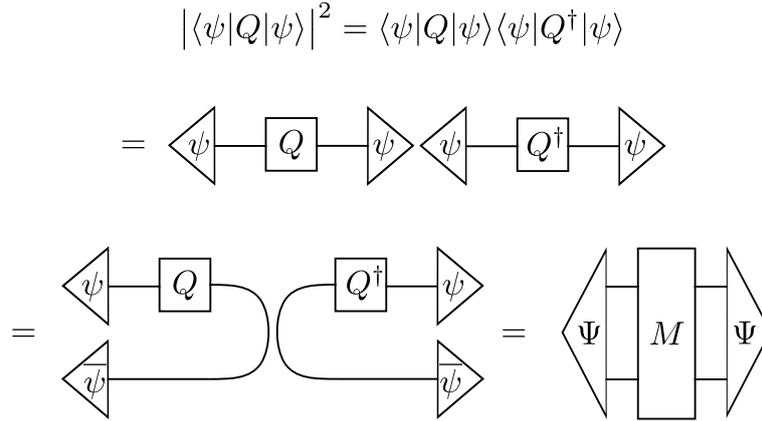


Figure 3-1: Reduction of the term $|\langle \psi | Q | \psi \rangle|^2$ from (3.2) to the VQE cost function of the form $\langle \Psi | M | \Psi \rangle$, where M and $|\Psi\rangle$ are defined in (3.3) and (3.4), respectively.

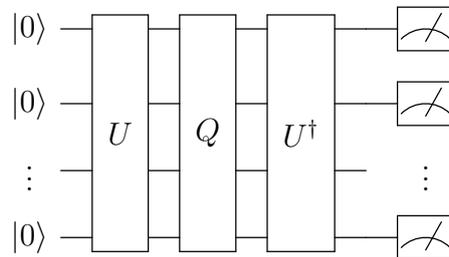
eigenstate. Then the cost function (3.2) can be written as

$$F_Q(\psi) = \langle \psi | M | \psi \rangle = |\langle \mathbf{0} | U^\dagger Q U | \mathbf{0} \rangle|^2,$$

i.e. it is equivalent to VQE executed finding the eigenstate of the highest-energy of M using the variational state of the specified form. And again, the absolute value squared can be estimated as the probability to obtain zeros as a result of the measurement of the state $|\psi\rangle$,

$$p_{\mathbf{0}} = \text{Tr} |\psi\rangle\langle\psi| |\mathbf{0}\rangle\langle\mathbf{0}| = |\langle \mathbf{0} | \psi \rangle|^2 = |\langle \mathbf{0} | U^\dagger Q U | \mathbf{0} \rangle|^2. \quad (3.5)$$

The quantum circuit which performs this computation is the following:



By optimizing the functions F_H and F_Q , it is possible to obtain one of the eigenstates of a given Hamiltonian. In Figure 3-2 we show the performance of F_H and F_Q in finding an eigenstate quantified by two metrics. The first metric is the

convergence rate, which we call the percentage of the instances converged to the values of the overlap

$$\mathcal{O} = |\langle \psi | \varphi \rangle|^2 \quad (3.6)$$

at least 0.999, where $|\psi\rangle$ is the variational state and $|\varphi\rangle$ is an exact eigenstate; this required value of the overlap is relatively large and is comparable to recent experimental results on high-fidelity state preparation [90, 91]. The second metric we use is the gain defined as

$$\mathcal{G}_{\mathcal{O}} = \mathbb{E}_{\mathcal{O}} \left[\frac{\mathcal{O}_{\text{conv}} - \mathcal{O}_{\text{init}}}{\mathcal{O}_{\text{opt}} - \mathcal{O}_{\text{init}}} \right], \quad (3.7)$$

where $\mathcal{O}_{\text{init}}$ is the value the overlap (3.6) calculated at the very beginning of the optimization, \mathcal{O}_{opt} is the function value obtained after convergence, and \mathcal{O}_{opt} is the true optimal value of the function; the notation $\mathbb{E}_{\mathcal{O}}$ means averaging over all instances of the initial ansatz parameters [92]. In fact, the gain can be used to characterize the performance of any cost function f which is supposed to be optimized, and it is used as a convergence metric in machine learning. To quantify the performance of F_H and F_Q by the mentioned metrics, we considered the four-qubit Ising Hamiltonian (1.38) with $J = h = 1$, and the four-qubit Schwinger Hamiltonian (2.10) with $w = g = 1$, $\epsilon_0 = 0$ and $m = -0.7$. To parametrize the probe state $|\psi\rangle$, we used the hardware-efficient ansatz. As can be seen in Figure 3-2, the function F_Q outperforms the function F_H in the case of Ising Hamiltonian, and for the Schwinger Hamiltonian the situation is opposite.

3.1.1 Variational Hamiltonian diagonalization

Interestingly, the cost function (3.2) can be also used to find a unitary which diagonalizes a Hamiltonian. That is, given $H \in \text{Herm}(\mathbb{C}^d)$, one can variationally find a unitary $U \in \text{U}(\mathbb{C}^d)$ such that the operator UHU^\dagger is diagonal in some basis. Indeed, the function

$$f_h(U) = \frac{1}{d} \sum_{j=1}^d |\langle j | U^\dagger e^{-iHt} U | j \rangle|^2 \quad (3.8)$$

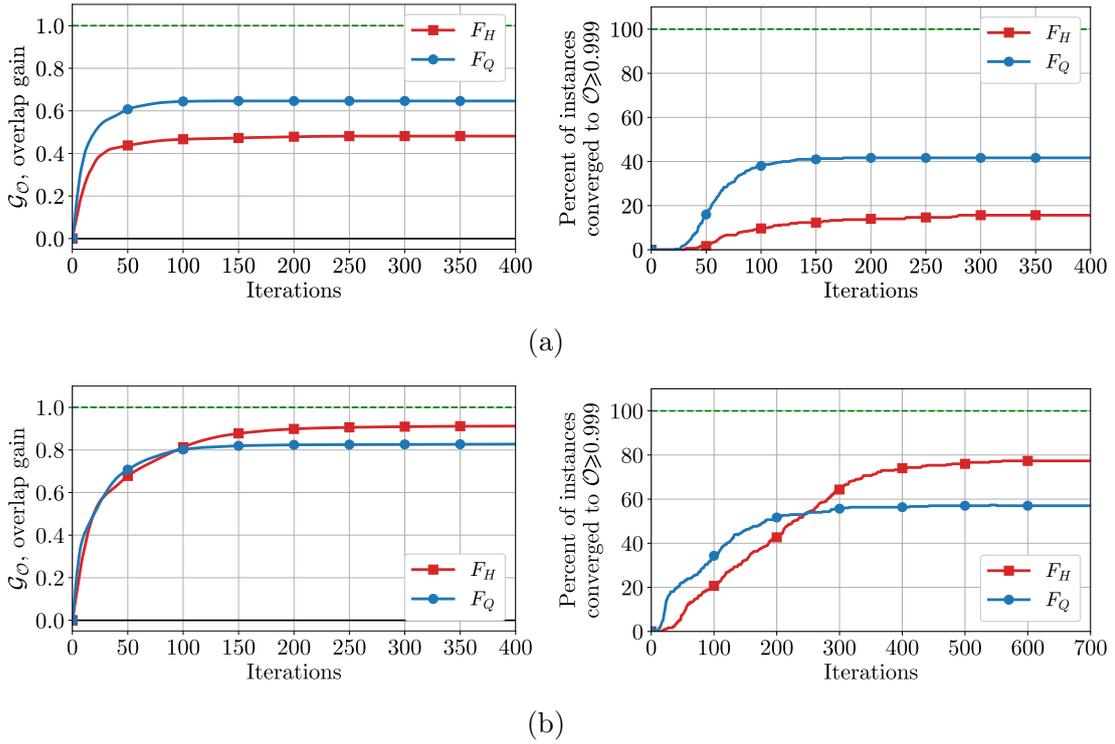


Figure 3-2: The overlap gain (left) and convergence rate (right) for the transverse field Ising model Hamiltonian (a) and the Schwinger model Hamiltonian (b) during the optimization of the functions F_H and F_Q defined in (3.1) and (3.2), respectively. For each Hamiltonian, 300 random sets of initial parameters for the hardware-efficient ansatz were generated. The figures are taken from [3].

returns unity if and only if the operator $U^\dagger e^{-iHt} U$ is diagonal with respect to the computational basis $\{|j\rangle\}_{j=1}^d$. Therefore, one finds a diagonalizing unitary U^{diag} as a solution to the following optimization problem:

$$U^{\text{diag}} = \arg \max_U f_h(U).$$

Obviously, if the operator $U^\dagger e^{-iHt} U$ is diagonal in $\{|j\rangle\}_{j=1}^d$, then is the operator $U^\dagger H U$, as they have the same eigenvectors.

The main shortcoming of the described method is that it necessitates conducting d measurement experiments for a given $H \in \text{Herm}(\mathbb{C}^d)$. That is, given a Hamiltonian acting on n qubits, one has to compute $d = 2^n$ probabilities in (3.8), which is exponential. This obstacle can be overcome by applying the tensor network notation in the same manner as done in the previous section. Letting again $Q = e^{-iHt}$, and

$$\begin{aligned}
 & \frac{1}{d^2} \sum_{j=1}^d |\langle j|U^\dagger Q U|j\rangle|^2 = \frac{1}{d^2} \sum_{j=1}^d \langle j|U^\dagger Q U|j\rangle \langle j|U^\dagger Q^\dagger U|j\rangle \\
 &= \frac{1}{d^2} \sum_{j=1}^d \left\langle j \left| \begin{array}{c} \boxed{U^\dagger} \quad \boxed{Q} \quad \boxed{U} \end{array} \right. j \right\rangle \left\langle j \left| \begin{array}{c} \boxed{U^\dagger} \quad \boxed{Q^\dagger} \quad \boxed{U} \end{array} \right. j \right\rangle \\
 &= \frac{1}{d^2} \sum_{j=1}^d \left(\begin{array}{c} \boxed{U^\dagger} \quad \boxed{Q} \quad \boxed{U} \quad \begin{array}{c} \diagdown \quad \diagup \\ j \quad j \end{array} \\ \boxed{U^\dagger} \quad \boxed{Q^\dagger} \quad \boxed{U} \quad \begin{array}{c} \diagup \quad \diagdown \\ j \quad j \end{array} \end{array} \right) \\
 &= \left(\begin{array}{c} \boxed{U^\dagger} \quad \boxed{Q^\dagger} \quad \boxed{U} \\ \boxed{U^\dagger} \quad \boxed{Q^\dagger} \quad \boxed{U} \end{array} \right) \\
 &= \left\langle \Psi \left| \begin{array}{c} \boxed{S} \end{array} \right. \Psi \right\rangle
 \end{aligned}$$

Figure 3-3: Derivations used to reduce the cost function (3.8) to VQE. The vector $|\Psi\rangle$ and the operator S are defined in (3.9) and (3.11), respectively.

using the derivations shown in Figure 3-3, one can rewrite the cost function (3.8) as

$$f_h(U) = \langle \Psi | M | \Psi \rangle,$$

where

$$|\Psi\rangle = (U^\dagger Q U \otimes \mathbb{1}) |\Phi^+\rangle, \quad (3.9)$$

$$M = \sum_{j=1}^d |j\rangle\langle j| \otimes |j\rangle\langle j|. \quad (3.10)$$

Although we reduced the quantum algorithm to VQE executed for the variational state $|\Psi\rangle$ and Hamiltonian M , the problem still remains: the Hamiltonian M consists of an exponential number of terms in the Pauli basis. This issue can be resolved as follows.

In Fig. 3-3, we introduced the notation

$$S = \frac{1}{d} \sum_{j=1}^d |j\rangle\langle j| \otimes |j\rangle\langle j| \equiv \left(\begin{array}{c} \text{---} \quad \text{---} \\ \diagdown \quad \diagup \\ \bullet \quad \bullet \\ \diagup \quad \diagdown \\ \text{---} \quad \text{---} \end{array} \right), \quad (3.11)$$

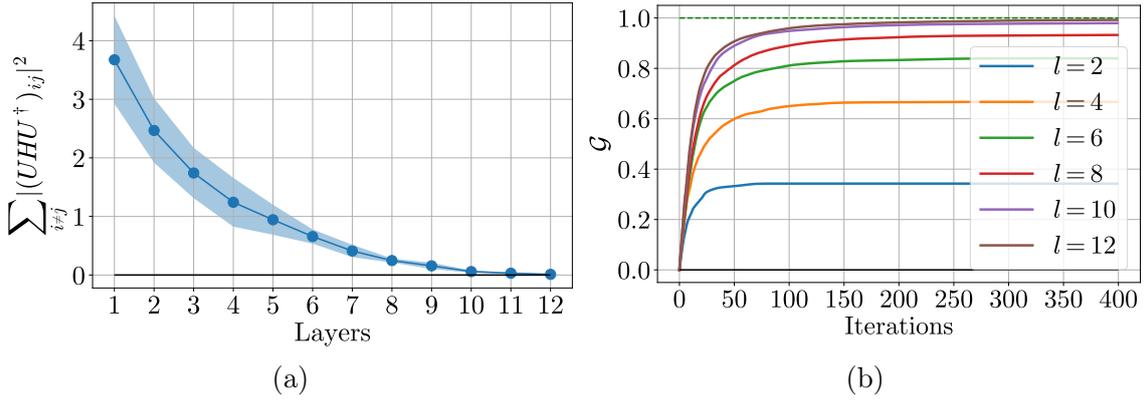


Figure 3-4: Numerical experiments conducted for diagonalizing random 5-qubit Hamiltonians with the method described by (3.8). The unitary U was parametrized by HEA of different number of layers l . On panel (a) depicted is the error of diagonalization, the sum of the squared absolute values of the off-diagonal entries of the diagonalized matrix. Each data point is obtained as an average over ten random Hamiltonians; the shaded area indicates the standard deviation. On panel (b) plotted is the gain (3.7) of the cost function in (3.8). The lines of different colors stand for different number of layers l in the ansatz.

One can show that the probability of obtaining zeros as a result of the measurement is

$$p_0 = \frac{1}{d^2} \sum_{j=1}^d |\langle j | U^\dagger Q U | j \rangle|^2,$$

from which one straightforwardly obtains $f_h(U) = p_0 d$. Therefore, if one finds a unitary U which maximizes the probability p_0 , then we can be sure that the operator $U^\dagger H U$ is diagonal in the computational basis.

We conducted numerical experiments on minimizing the cost function (3.8) randomly generated Hamiltonians of $n = 5$ qubits. In Figure 3-4 we plot the error of the diagonalization and the gain (3.7) of the cost function in (3.8). The error is given as the sum of the squared absolute values of the off-diagonal entries of the matrix $U^\dagger H U$. To parametrize the unitary U , we applied the hardware-efficient ansatz with the number of layers l ranging from 1 to 12. As we see in the figure, the more layers we have, the lower the error is. With $l = 12$ ansatz layers we achieve the error close to zero, meaning that the operator $U^\dagger H U$ is close to be diagonal in the computational basis.

3.2 Variational state preparation

Consider the following problem: given a (possibly, unknown) n -qubit state $|\alpha\rangle$, find a unitary U such that $|\psi\rangle = U|\mathbf{0}\rangle$ and $|\alpha\rangle = e^{ir}|\psi\rangle$, where $r \in \mathbb{R}$ and $|\mathbf{0}\rangle = |0\rangle^{\otimes n}$. In other words, given a black-box state $|\alpha\rangle$, find an operator U which prepares this state from $|\mathbf{0}\rangle$ up to a phase e^{ir} . This task can be performed variationally, and two cost functions can be suggested for it. The first cost function is the overlap between the states,

$$f_1(U) = |\langle\alpha|U|\mathbf{0}\rangle|^2, \quad (3.15)$$

which we already encountered in (3.6). This quantity is equal to unity if and only if $|\alpha\rangle = e^{ir}|\psi\rangle$, which means the equivalence up to a phase, and which satisfies our requirements. The second cost function is the squared distance between $|\alpha\rangle$ and $|\psi\rangle = U|\mathbf{0}\rangle$ induced by the 2-norm,

$$\begin{aligned} f_2(U) &= \|\ |\alpha\rangle - U|\mathbf{0}\rangle \|^2_2 \\ &\equiv \left(\langle\alpha| - \langle\mathbf{0}|U^\dagger \right) \left(|\alpha\rangle - U|\mathbf{0}\rangle \right) \\ &= \langle\alpha|\alpha\rangle - \langle\alpha|U|\mathbf{0}\rangle - \langle\mathbf{0}|U^\dagger|\alpha\rangle + \langle\mathbf{0}|U^\dagger U|\mathbf{0}\rangle \\ &= 2(1 - \operatorname{Re}\langle\alpha|U|\mathbf{0}\rangle), \end{aligned} \quad (3.16)$$

where we used the property $(z + \bar{z})/2 = \operatorname{Re}z$ for $z \in \mathbb{C}$. This cost function is minimized when $|\alpha\rangle = |\psi\rangle$, which means strict equivalence of the corresponding components of the vectors, and this condition is stronger than that for the cost function f_1 .

It turns out that the both cost functions can be calculated on a quantum computer. The cost f_1 can be estimated via the so-called swap test which is described in Appendix B in detail. This routine requires an auxiliary qubit and the ability to apply the controlled-swap gate. However for our case, a simpler way of computing the cost function f_1 is the same as in (3.5), i.e. it can be reduced to the probability of obtaining the result “0” for all qubits after measuring the state $U^\dagger|\alpha\rangle$ in the computational basis. The second cost function f_2 can be calculated using the Hadamard test which is discussed in Appendix C. To use this method, one needs to be able

to apply the gates U and V conditioned on the state of an auxiliary qubit, i.e. the method requires the application of the controlled unitaries CU and CV , which can be done even for unknown operators [93].

For f_1 , the reduction to VQE is trivial: one could use VQE to find the maximal-energy eigenstate of the Hamiltonian $M = |\mathbf{0}\rangle\langle\mathbf{0}|$ using the variational state $|\psi\rangle = U^\dagger|\alpha\rangle$. More interesting is the reduction of f_2 which is given in Figure 3-5. This derivation is inspired by the quantum circuit for the Hadamard test described in detail in Appendix C. In the derivations presented, we first notice that

$$\begin{array}{c} \text{---} \circ \text{---} \\ | \\ \text{---} \text{---} \\ \text{---} \square U \text{---} \\ | \\ \text{---} \bullet \text{---} \\ | \\ \text{---} \text{---} \\ \text{---} \square V \text{---} \end{array} = |0\rangle\langle 0| \otimes V + |1\rangle\langle 1| \otimes U.$$

Then, by applying the phase gates, we change the sign of the second term,

$$(S \otimes \mathbf{1})(|0\rangle\langle 0| \otimes V + |1\rangle\langle 1| \otimes U)(S \otimes \mathbf{1}) = |0\rangle\langle 0| \otimes V - |1\rangle\langle 1| \otimes U.$$

Finally, we trace out the first qubit to obtain the non-unitary operator we need,

$$\text{Tr}_1 [|0\rangle\langle 0| \otimes V - |1\rangle\langle 1| \otimes U] = V - U,$$

which gives the two-norm distance, see the first line in Figure 3-5. Recalling that $S = S^T = \sqrt{Z}$, one can use the properties of the Bell states to move the phase gates up on the tensor diagram to obtain the Z gate, as done in the fourth line in Figure 3-5. Thus, we reduced the cost function f_2 to the VQE for the following Hamiltonian and variational state:

$$\begin{aligned} M &= (Z \otimes \mathbf{1} \otimes \mathbf{1}) \left(|\Phi^+\rangle\langle\Phi^+| \otimes \mathbf{1} \right) (Z \otimes \mathbf{1} \otimes \mathbf{1}), \\ |\Psi\rangle &= CV_{23}LU_{23}(|\Phi^+\rangle \otimes |0\rangle), \end{aligned}$$

where LU is the oppositely controlled U defined in (1.33).

The two costs f_1 and f_2 have also different convergence properties. In Figure 3-6, we plot the gain (3.7) of these functions during the optimization. As can be seen,

$$\begin{aligned}
 & \left\| |U|0\rangle - |V|0\rangle \right\|_2^2 = \langle 0|(U - V)(U^\dagger - V^\dagger)|0\rangle \\
 & = \langle 0| \boxed{V - U} \boxed{V^\dagger - U^\dagger} |0\rangle \\
 & = \text{Diagram 1} \\
 & = \text{Diagram 2} \\
 & = \text{Diagram 3} \\
 & = \text{Diagram 4}
 \end{aligned}$$

Figure 3-5: Reduction of the 2-norm distance (3.16) to the cost function of VQE for the Hamiltonian $M = (Z \otimes \mathbb{1} \otimes \mathbb{1}) (|\Phi^+\rangle\langle\Phi^+| \otimes \mathbb{1}) (Z \otimes \mathbb{1} \otimes \mathbb{1})$ and the variational state $|\Psi\rangle = CV_{23}LU_{23}|\Phi^+\rangle \otimes |0\rangle$.

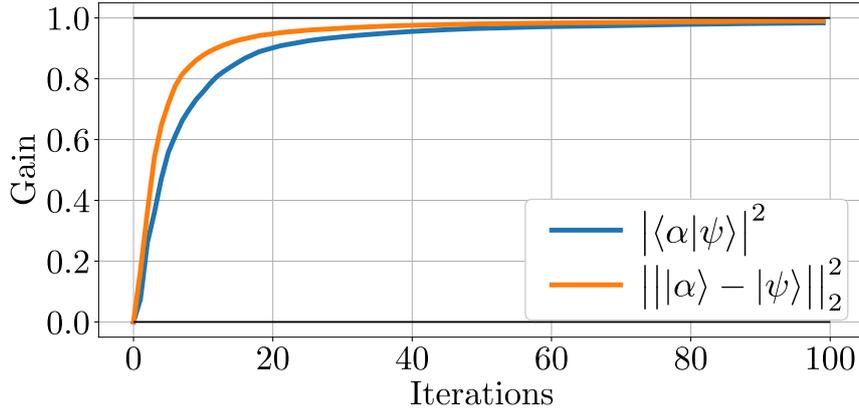


Figure 3-6: Comparison of the gains (3.7) of the functions (3.15) and (3.16). The number of qubits was set to $n = 6$, the target vector $|\alpha\rangle$ was generated randomly, and the variational state $|\psi\rangle$ was prepared by the hardware-efficient ansatz given in Fig. 2-1 with $l = 6$ layers, the gain was obtained as the average over ten independent runs, and the optimization algorithm was BFGS.

the 2-norm distance converges slightly faster than the overlap. Another important advantage of the cost function f_2 is that it has a simpler derivative since it is linear with respect to $\langle \alpha | \psi \rangle$.

3.3 Variational quantum compiling

Consider the following problem: given a unitary U , find a quantum circuit V such that $V = e^{i\alpha U}$, $\alpha \in \mathbb{R}$. This task is often encountered in applied quantum computations and experiments with quantum computing devices [73, 94]. For example, the CX gate is crucial for many quantum algorithms, but it cannot be applied directly on e.g. a hardware based on trapped ions. But this gate can be compiled as a sequence of available operations, e.g. R_x , R_z and R_{xx} in this case. The problem of such compilation can be also faced when it is needed to perform a unitary evolution of a (potentially, unknown) Hamiltonian H in time t , $U = e^{-iHt}$, on the condition that this U can be applied as a black-box. The third possible application of quantum compilation is circuit compression: given an ansatz U consisting of u gates, the task is to find a circuit V which contains $v \leq u$ gates.

Formally, the stated problem can be solved as follows:

$$V^{\text{opt}} = \arg \max_V |(V, U)_{\text{HS}}|^2, \quad (3.17)$$

where we have the Hilbert-Schmidt product $(V, U)_{\text{HS}} = \text{Tr } V^\dagger U$. If $U, V \in \mathcal{U}(d)$ with $d = 2^n$, then $|(V, U)_{\text{HS}}| = d$ if and only if $V = e^{i\alpha}U$, as required. The Hilbert-Schmidt product can be computed basing on the fact that

$$\langle \Phi^+ | V^\dagger U \otimes \mathbb{1} | \Phi^+ \rangle = \frac{1}{d} \text{Tr } V^\dagger U,$$

where

$$|\Phi^+\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^d |j\rangle \otimes |j\rangle$$

is the generalized Bell state.

Let us look at this algorithm from the point of view of tensor diagrams. In Fig. 3-7 we depict the reduction of the cost function in the maximization problem (3.17) to finding the highest-eigenvalue state of the Hamiltonian $M = |\Phi^+\rangle\langle\Phi^+|$ with the variational state $|\Psi\rangle = (U \otimes \bar{V}) |\Phi^+\rangle = (UV^\dagger \otimes \mathbb{1}) |\Phi^+\rangle$. As before, the Hamiltonian M is easy to measure. Suppose $F|0\rangle^{\otimes 2n} = |\Phi^+\rangle$. If one prepares the state $|\Psi\rangle = F^\dagger(V^\dagger U \otimes \mathbb{1})F|0\rangle^{\otimes 2n}$ and measures it in the computational basis, then the probability p_0 of obtaining the ‘‘all-zeros’’ result is precisely $|\text{Tr } V^\dagger U|^2/d^2$. Indeed, if $|\mathbf{0}\rangle = |0\rangle^{\otimes 2n}$, then

$$p_0 = \text{Tr } |\Psi\rangle\langle\Psi| |\mathbf{0}\rangle\langle\mathbf{0}| = |\langle\Psi|\mathbf{0}\rangle|^2 = |\langle\Phi^+| V^\dagger U \otimes \mathbb{1} |\Phi^+\rangle|^2 = \left| \frac{1}{d} \text{Tr } V^\dagger U \right|^2$$

That is, in the considered algorithm, the goal is to find V such that it maximizes this probability.

3.4 Variational linear system solver

In this section, we consider the problem of solving systems of linear equations. This task can be formulated as $Ax = b$, where A and b are the matrix and the vector of

$$\begin{aligned}
 |(V, U)_{\text{HS}}|^2 &= \overline{(V, U)_{\text{HS}}} \cdot (V, U)_{\text{HS}} = \text{Tr}U^\dagger V \cdot \text{Tr}V^\dagger U \\
 &= \text{Diagram 1} \quad \text{Diagram 2} \\
 &= \text{Diagram 3} \quad \text{Diagram 4} = \text{Diagram 5}
 \end{aligned}$$

Figure 3-7: Reduction of the term $|(V, U)_{\text{HS}}|^2$ from (3.17) to the VQE cost function of the form $\langle \Psi | M | \Psi \rangle$ with $M = |\Phi^+\rangle\langle\Phi^+|$ and $|\Psi\rangle = (U \otimes \bar{V}) |\Phi^+\rangle$.

coefficients, respectively, and x is the vector of unknown variables to be found. It turns out that this problem can be solved on a quantum computer [95–97]. Since in quantum mechanics and computations we commonly work with unit-length vectors, let us redefine the vectors b and x as

$$b \equiv \|b\|_2 |b\rangle, \quad x \equiv \|x\|_2 |x\rangle,$$

where the 2-norms $\|b\|_2$ and $\|x\|_2$ can be safely incorporated into the matrix A . This transforms the original equation $Ax = b$ into its “quantum” form, $A|x\rangle = |b\rangle$. The vector $|x^{\text{opt}}\rangle$ which solves this equation can be found as

$$|x^{\text{opt}}\rangle = \arg \max_x |\langle b | A | x \rangle|^2, \quad (3.18)$$

which is essentially the overlap between the vectors $|b\rangle$ and $A|x\rangle$. An important assumption here is that it must be known how to prepare the state $|b\rangle = B|\psi_0\rangle$ from some initial state $|\psi_0\rangle$, i.e. the unitary B must be given. Note also that the matrix A is arbitrary, so the vector $A|x\rangle$ is not of unit norm. Therefore one may divide the cost function in (3.18) by $\langle x | A^\dagger A | x \rangle$ so that the function has an operational meaning: in this case, it is lower-bounded by 0 and upper-bounded by 1. However, for explaining the algorithm it is not crucial.

In the seminal work [96], it is suggested to represent the matrix A as a linear

combination of unitaries,

$$A = \sum_{i=1}^L a_i A_i, \quad a_i \in \mathbb{C}, \quad A_i^\dagger A_i = \mathbb{1}.$$

Let us also suppose that the variational state $|x\rangle$ is prepared by the unitary operator X (not to be confused with the Pauli operator), i.e. $|x\rangle = X|\psi_0\rangle$. This rewrites the optimization problem (3.18) to the form

$$\begin{aligned} |x^{\text{opt}}\rangle &= \arg \max_X \sum_{i=1}^L \left| a_i \langle \psi_0 | B^\dagger A_i X | \psi_0 \rangle \right|^2 \\ &= \arg \max_X \sum_{i,j=1}^L a_i \bar{a}_j \langle \psi_0 | B^\dagger A_i X | \psi_0 \rangle \langle \psi_0 | X^\dagger A_j^\dagger B | \psi_0 \rangle. \end{aligned} \quad (3.19)$$

Two thoughts must be noted here. First, we transitioned from maximization over vectors $|x\rangle$ to maximization over unitaries X (which can be parametrized in accordance with the variational approach, see Section 2.1). Second, in quantum computations we work with observables which are commonly Hermitian operators, while the terms of the form $\langle \psi_0 | B^\dagger A_i X | \psi_0 \rangle$ are, in fact, the expectation values of a unitary operator, so these expectations are complex numbers. Still, it is possible to compute these quantities via the Hadamard test mentioned earlier; more accurately, it is possible to separately compute the real and imaginary parts of these expectation values (see Appendix C for details). In our case, the Hadamard test requires the ability of applying the unitaries B^\dagger , A_i and X conditioned on the state of an auxiliary qubit. Despite that the Hadamard test is a general technique allowing the computation of terms of the form $\langle \psi | U | \psi \rangle$ for arbitrary $|\psi\rangle$ and U , there is a more efficient circuit which directly computes the terms $\langle \psi_0 | B^\dagger A_i X | \psi_0 \rangle \langle \psi_0 | X^\dagger A_j^\dagger B | \psi_0 \rangle$ in (3.19), see Fig. 9(b) in the original article [96].

Let us now apply the language of tensor networks and reduce the cost function in (3.18) to that of the VQE; this is done in Fig. 3-8. The fact that the operator $M = A^\dagger |b\rangle\langle b| A$ is Hermitian is noted by the authors of [96], so the reduction is straightforward. Recall however that in order to measure an expected value of H , one needs to expand it in the Pauli basis (2.7). So the question here is how hard

$$\begin{aligned}
 |\langle b|A|x\rangle|^2 &= \langle x|A^\dagger|b\rangle\langle b|A|x\rangle \\
 &= \langle x| \boxed{A^\dagger} |b\rangle \langle b| \boxed{A} |x\rangle \\
 &= \langle x| \boxed{M} |x\rangle
 \end{aligned}$$

Figure 3-8: Reduction of the term $|\langle b|A|x\rangle|^2$ from (3.18) to the VQE cost function of the form $\langle x|M|x\rangle$.

it is to find such a decomposition, and how many terms it would have. Another interesting fact can be observed if one rewrites (3.19) as follows:

$$|x^{\text{opt}}\rangle = \arg \max_x \sum_{i=1}^L \sum_{j=1}^i \langle x| \left(a_i \bar{a}_j A_i^\dagger |b\rangle\langle b| A_j + a_j \bar{a}_i A_j^\dagger |b\rangle\langle b| A_i \right) |x\rangle$$

In the equation above, the term in brackets is expectantly Hermitian, so the variational linear systems solver also reduces to VQE. However, the efficiency of this approach is yet to be studied.

3.5 Variational quantum state diagonalization

The problem of diagonalizing a density operator is stated as follows. Given a state $\rho \in \mathcal{D}(\mathcal{H})$ with $\dim \mathcal{H} = d$, find $U \in \text{U}(\mathcal{H})$ such that the state $\rho_U \equiv U\rho U^\dagger$ is diagonal in the computational basis $\{|j\rangle\}_{j=1}^d$. In [40], it is suggested to minimize a quantity which characterizes how far is the state ρ_U from being diagonal. Such a metric can be the Hilbert-Schmidt distance

$$d_{\text{HS}}(A, B) = \text{Tr} [(A - B)^\dagger (A - B)].$$

Now consider the channel $D[\rho]$ which sets to zero all the density matrix elements except the ones on the diagonal,

$$D[\rho] = \sum_{j=1}^d \langle j|\rho|j\rangle |j\rangle\langle j|. \tag{3.20}$$

Then the unitary U^{opt} which diagonalizes the state ρ is found as

$$U^{\text{opt}} = \arg \min_U d_{\text{HS}}(\rho_U, D[\rho_U]) \quad (3.21)$$

$$= \arg \min_U \sum_{\substack{a,b=1 \\ a \neq b}}^d |\langle a | \rho_U | b \rangle|^2 \quad (3.22)$$

$$= \arg \min_U \left(\text{Tr} \rho^2 - \text{Tr} D[\rho_U]^2 \right). \quad (3.23)$$

The equivalence of the expressions (3.21), (3.22) and (3.23) is shown in Appendix A.

The most convenient way of solving the stated optimization problem is computing the two terms in (3.23). The first term, $\text{Tr} \rho^2$, can be estimated via the so-called swap test which is described in detail in Appendix B. In Fig. 3-9 we show the tensor network formulation of computing a similar term, $(1 + \text{Tr} \rho^2)/2$, from which the desired quantity can be easily extracted. The first line in this Figure, $\text{Tr} \rho^2 = \text{Tr}(\rho \otimes \rho)SW$ (see equations (B.3-B.7) for the prove) is known as an invariant in the theory of tensor networks [43]. In fact, the swap gate SW defined in (1.25) is Hermitian, which makes it a valid observable to measure. Moreover, $(\rho \otimes \rho)SW = SW(\rho \otimes \rho)$, which makes this expression a Hermitian operator as well, and it would be actually enough in the sense of reducing to VQE. Indeed, the Hamiltonian and the pure “variational” state in this case would be

$$M = \mathbb{1} \otimes SW(\rho \otimes \rho) \otimes \mathbb{1},$$

$$|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{a,b=1}^d |a\rangle \otimes |a\rangle \otimes |b\rangle \otimes |b\rangle,$$

which give $\langle \Psi | M | \Psi \rangle = \frac{1}{d} \text{Tr} \rho^2$. However, of greater use would be a reduction suitable for computing $\text{Tr} \rho \tau$ with $\tau, \rho \in \mathcal{D}(\mathcal{H})$ (for pure states τ and ρ , it gives the overlap between them, see Appendix B). For example, for a given ρ , one could want to find τ such that the trace of their product is maximized. Unfortunately, the operator $SW(\rho \otimes \tau)$ is not Hermitian anymore (see (B.6)), so it cannot be measured. This is why we need further derivations given in the second and the third lines in Fig. 3-9,

which are inspired by the swap test quantum circuit (B.1). The last line is obtained by following the same logic as applied for the derivation of the tensor networks in Fig. 3-5. Essentially, we used the property

$$\begin{array}{c} \triangleleft \text{+} \text{+} \trianglerightarrow \\ \bullet \\ | \\ \square U \end{array} = \begin{array}{c} \text{---} \text{---} \\ \bullet \\ | \\ \square U \end{array} = \text{---} \boxed{\frac{1}{2}(\mathbb{1} + U)} \text{---}$$

but having the swap gate (1.25) instead of U ; we also used the notation $|+\rangle = H|0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. The result is that we obtained the cost function of VQE with the following Hamiltonian and pure “variational” state:

$$M = |+\rangle\langle +| \otimes (\rho \otimes \rho) \otimes \mathbb{1} \otimes \mathbb{1}, \quad (3.24)$$

$$|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{a,b=1}^d \left(|0\rangle \otimes |a\rangle \otimes |b\rangle \otimes |b\rangle \otimes |a\rangle + |1\rangle \otimes |b\rangle \otimes |a\rangle \otimes |b\rangle \otimes |a\rangle \right), \quad (3.25)$$

giving again $\langle \Psi | M | \Psi \rangle = \frac{1}{d} \text{Tr} \rho^2$. As was mentioned, this formulation is more suitable if e.g. one wants to maximize $\text{Tr} \rho \tau_U$, where $\tau_U = U \tau U^\dagger$ is a variational state; the varied unitary U can be incorporated into (3.25), which forms the pure variational state. Note however that in the cost function of (3.23), there is no dependence of the term $\text{Tr} \rho^2$ on the variational unitary U , this term needs to be computed only once.

Now we need to compute the second term in (3.23), $\text{Tr} D[\rho_U]^2$, the trace of the squared variational density operator which has only diagonal entries, see (3.20). The tensor network and the reduction to VQE is not very difficult, and it is depicted in Fig. 3-10. There, we used the definition (3.12). The Hamiltonian and the variational state for VQE are

$$M = \rho \otimes \mathbb{1} \otimes \rho, \quad (3.26)$$

$$|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{a=1}^d U |a\rangle \otimes |a\rangle \otimes U |a\rangle. \quad (3.27)$$

for which we have $\langle \Psi | M | \Psi \rangle = \frac{1}{d} \text{Tr} D[\rho_U]^2$.

$$\begin{aligned}
 \text{Tr} \rho^2 &= \text{Tr}(\rho \otimes \rho) SW = \text{Diagram 1} \\
 &\simeq \frac{1}{2}(1 + \text{Tr} \rho^2) = \text{Tr} \frac{1}{2}(\mathbb{I} + SW)(\rho \otimes \rho) \frac{1}{2}(\mathbb{I} + SW) \\
 &= \text{Diagram 2} \\
 &= \text{Diagram 3} = \text{Diagram 4}
 \end{aligned}$$

Figure 3-9: Reduction of a tensor network which computes $\text{Tr} \rho^2$ to a tensor network which computes $(1 + \text{Tr} \rho^2)/2$. The obtained tensor network is equivalent to the VQE for the Hamiltonian H and state $|\Psi\rangle$ defined in (3.24) and (3.25), respectively.

3.5.1 Alternative variational state diagonalization

Let us look closely at the cost function in (3.22). Applying certain properties of the trace and tensor products, one may find that

$$\begin{aligned}
 \sum_{\substack{a,b=1 \\ a \neq b}}^d |\langle a|\rho|b\rangle|^2 &= \sum_{\substack{a,b=1 \\ a \neq b}}^d \langle a|\rho|b\rangle \langle b|\rho|a\rangle \\
 &= \sum_{\substack{a,b=1 \\ a \neq b}}^d \text{Tr} \rho |a\rangle\langle b| \cdot \text{Tr} \rho |b\rangle\langle a| \\
 &= \sum_{\substack{a,b=1 \\ a \neq b}}^d \text{Tr} (\rho |a\rangle\langle b| \otimes \rho |b\rangle\langle a|) \\
 &= \frac{1}{2} \sum_{\substack{a,b=1 \\ a \neq b}}^d \text{Tr} (\rho \otimes \rho) \left(|a\rangle\langle b| \otimes |b\rangle\langle a| + |b\rangle\langle a| \otimes |a\rangle\langle b| \right) \\
 &= \text{Tr} (\rho \otimes \rho) M,
 \end{aligned}$$

$$\begin{aligned}
 \text{Tr } D[\rho_U]^2 &= \sum_a \langle a | \rho_U | a \rangle^2 = \sum_{a,b} \langle a | \rho_U | b \rangle \langle a | \rho_U | b \rangle \langle a | b \rangle \\
 &= \left(\sum_a \langle a | \otimes \langle a | \otimes \langle a | \right) (\rho_U \otimes \mathbb{1} \otimes \rho_U) \left(\sum_b | b \rangle \otimes | b \rangle \otimes | b \rangle \right) \\
 &= \begin{array}{c} \text{---} \bullet \text{---} \\ \text{---} \bullet \text{---} \end{array} \begin{array}{c} \boxed{U} \text{---} \boxed{\rho} \text{---} \boxed{U^\dagger} \\ \boxed{U} \text{---} \boxed{\rho} \text{---} \boxed{U^\dagger} \end{array} = \begin{array}{c} \text{---} \bullet \text{---} \\ \text{---} \bullet \text{---} \end{array} \begin{array}{c} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \end{array} \begin{array}{c} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \end{array} \begin{array}{c} \text{---} \bullet \text{---} \\ \text{---} \bullet \text{---} \end{array}
 \end{aligned}$$

Figure 3-10: Tensor network formulation and reduction to VQE for $\text{Tr } D[\rho_U]^2$. The Hamiltonian and the variational state for VQE are given in (3.26) and (3.27), respectively.

where we introduced

$$M = \frac{1}{2} \sum_{\substack{a,b=1 \\ a \neq b}}^d \left(|a\rangle\langle b| \otimes |b\rangle\langle a| + |b\rangle\langle a| \otimes |a\rangle\langle b| \right). \quad (3.28)$$

This allows to obtain a method for finding a unitary which diagonalizes a given density operator. Let $\rho \in \mathcal{D}(\mathcal{H})$ with $\dim \mathcal{H} = d$ and

$$f_{\text{sd}}(U) = \text{Tr}(U\rho U^\dagger \otimes U\rho U^\dagger)M, \quad (3.29)$$

where $U \in \text{U}(\mathcal{H} \otimes \mathcal{H})$. By construction, the function $f_{\text{sd}}(U)$ returns zero if and only if the density operator $U\rho U^\dagger$ is diagonal in the computational basis. This means that one can diagonalize a density operator ρ by finding a unitary U^{diag} as

$$U^{\text{diag}} = \arg \min_U f_{\text{sd}}(U).$$

Worthy of note is also the fact that the operator M defined in (3.28) is Hermitian, thus making it a valid observable to measure. Moreover, one can easily find its spectral decomposition. Indeed, consider the eigenvalue problem

$$M |\psi\rangle = \lambda |\psi\rangle.$$

One can verify that the eigenvectors of M are

$$\begin{aligned} |\psi_0^j\rangle &= |j\rangle \otimes |j\rangle, \\ |\psi_{\pm}^{ab}\rangle &= \frac{1}{\sqrt{2}} \left(|a\rangle \otimes |b\rangle \pm |b\rangle \otimes |a\rangle \right), \quad a \neq b, \end{aligned}$$

from which d vectors $|\psi_0^j\rangle$ have to the eigenvalue 0, and $d(d-1)/2$ vectors $|\psi_{\pm}^{ab}\rangle$ correspond to the eigenvalues ± 1 . This gives the following spectral decomposition:

$$M = \sum_{\substack{a,b=1 \\ a \neq b}}^d |\psi_+^{ab}\rangle\langle\psi_+^{ab}| - \sum_{\substack{a,b=1 \\ a \neq b}}^d |\psi_-^{ab}\rangle\langle\psi_-^{ab}| + 0 \sum_{j=1}^d |\psi_0^j\rangle\langle\psi_0^j|. \quad (3.30)$$

As can be seen, the projectors $|\psi_0^j\rangle\langle\psi_0^j|$ have no contribution since their corresponding eigenvalues are 0. The question is, how one can measure the observable M without decomposing it in the Pauli basis? This is an idea for future work.

3.5.2 Alternative variational Hamiltonian diagonalization

It turns out that the method for state diagonalization described in the previous section can be transformed to diagonalize a given Hamiltonian. Consider the following operator on $\mathcal{H} \otimes \mathcal{H}$:

$$\sigma = \frac{1}{d^2} (\mathbb{1} + M), \quad (3.31)$$

where M is defined in (3.28). In fact, this operator is a density operator from $\mathcal{D}(\mathcal{H} \otimes \mathcal{H})$. Indeed, the identity $\mathbb{1}$ just shifts the eigenvalues of M from $\{-1, 0, 1\}$ to $\{0, 1, 2\}$, and the division by d^2 ensures that the resultant operator is of unit trace. This leads us to a method for diagonalizing a Hamiltonian described by the following cost function:

$$\begin{aligned} f_{\text{hd}}(U) &= \frac{1}{d^2} (\text{Tr } H)^2 - \text{Tr} (H \otimes H)(U \otimes U)\sigma(U^\dagger \otimes U^\dagger) \\ &= \frac{1}{d^2} (\text{Tr } H)^2 - \text{Tr} (U^\dagger H U \otimes U^\dagger H U)\sigma \end{aligned} \quad (3.32)$$

Indeed, this function returns zero if the operator $U^\dagger H U$ is diagonal in the computational basis in \mathcal{H} . This suggests that if one is able to prepare the bipartite state σ ,

then by locally varying this state with $U \otimes U$ one can diagonalize the Hamiltonian $H \otimes H$ (and therefore H itself) by minimizing the cost function (3.32).

The main drawback of this method is that the state σ is mixed, so it may be hard to prepare it in a real experiment. To remedy this, one could purify this state, which would double the number of qubits needed (from $2n$ to $4n$ for an n -qubit Hamiltonian), but it will also reduce the method to VQE. The respective Hamiltonian and variational state would be $H \otimes H \otimes \mathbb{1} \otimes \mathbb{1}$ and $|\sigma\rangle$, the purification for σ . This purification can be obtained easily if one notices that

$$\sigma = \frac{2}{d^2} \sum_{\substack{a,b=1 \\ a \neq b}}^d |\psi_+^{ab}\rangle\langle\psi_+^{ab}| + \frac{1}{d^2} \sum_{j=1}^d |\psi_0^j\rangle\langle\psi_0^j|,$$

which is essentially the same decomposition as for M in (3.30) but with the eigenvalues shifted by 1 and normalized by d^2 . This allows the following purification:

$$|\sigma\rangle = \frac{1}{d} \left(\sum_{j=1}^d |j\rangle \otimes |j\rangle \otimes |j\rangle \otimes |j\rangle + \frac{1}{\sqrt{2}} \sum_{\substack{a,b=1 \\ a \neq b}}^d (|a\rangle \otimes |b\rangle + |b\rangle \otimes |a\rangle) \otimes (|a\rangle \otimes |b\rangle + |b\rangle \otimes |a\rangle) \right). \quad (3.33)$$

The question is, how one prepares such a state on a quantum computer? This is an idea for future work as well.

3.6 Remarks on variational quantum algorithms

In this chapter, we considered several existing variational quantum algorithms and formulated them in the tensor networks notation. By applying tensor diagram identities, we reduced the cost functions of these algorithms to that of the VQE algorithm, i.e. to the expected value of a Hamiltonian measured with respect to a (pure) variational state. The question is, is it possible to do it for any variational quantum algorithm?

At the beginning of Chapter 2, we introduced the concept of the cost function

to be a real-valued function of quantum states and observables. In [39], the authors also give the following general expression of a cost function:

$$C(\boldsymbol{\theta}) = \sum_k f_k \left(\text{Tr} [U(\boldsymbol{\theta})\rho_k U^\dagger(\boldsymbol{\theta})H_k] \right), \quad (3.34)$$

where ρ_k are quantum states, H_k are observables, f_k are real-valued functions, and $U(\boldsymbol{\theta})$ is a variational circuit with the parameters $\boldsymbol{\theta}$. A mixed state $\rho_k \in \mathcal{D}(H_k)$ can always be purified: There always exists a Hilbert space \mathcal{H}_E and a pure state $|\Psi_k\rangle \in \mathcal{H}_k \otimes \mathcal{H}_E$ such that $\rho_k = \text{Tr}_E |\Psi_k\rangle\langle\Psi_k|$. More precisely, if $\rho_k = \sum_j p_j |\psi_j\rangle\langle\psi_j|$, then one can take $|\Psi_k\rangle = \sum_j \sqrt{p_j} |\psi_j\rangle \otimes |e_j\rangle$, where $\{|e_j\rangle\}$ is an orthonormal basis in \mathcal{H}_E . Therefore, the expression of the cost function (3.34) can be rewritten as

$$C(\boldsymbol{\theta}) = \sum_k f_k \left(\langle\Psi_k(\boldsymbol{\theta})|\tilde{H}_k|\Psi_k(\boldsymbol{\theta})\rangle \right),$$

where $|\Psi_k(\boldsymbol{\theta})\rangle = (U(\boldsymbol{\theta}) \otimes \mathbb{1}) |\Psi_k\rangle$ and $\tilde{H}_k = H_k \otimes \mathbb{1}$. That is, in expected values, we can always replace mixed states by their purifications, and changing the observables accordingly. Now the question is if we can always find a Hamiltonian and a variational state such that the expectation gives the value of $C(\boldsymbol{\theta})$ for arbitrary functions f_k and any number of terms in (3.34). On the one hand, if the functions are simply $f_k(x) = x$, then we can put $H = N \sum_k \tilde{H}_k \otimes |k\rangle\langle k|$ and $|\Psi(\boldsymbol{\theta})\rangle = \frac{1}{\sqrt{N}} \sum_k |\Psi_k(\boldsymbol{\theta})\rangle \otimes |k\rangle$, where $N = \langle\Psi(\boldsymbol{\theta})|\Psi(\boldsymbol{\theta})\rangle$, which gives $\langle\Psi(\boldsymbol{\theta})|H|\Psi(\boldsymbol{\theta})\rangle = C(\boldsymbol{\theta})$. On the other hand, such a reduction to VQE does not seem possible for any set of functions f_k . For example, we could not do it even for the cost function

$$C(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H^2|\psi(\boldsymbol{\theta})\rangle - \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle^2,$$

discussed in Section 3.1, unless we consider the terms of this expression separately. We therefore argue that generally we cannot reduce any given variational quantum algorithm to VQE, but as we showed in this chapter, it is often possible.

Chapter 4

Quantum channel discrimination with variational quantum circuits

So far we have seen that using the notion of variational circuits one is able to formulate and solve many practical problems on a quantum computer. In this chapter, we show that the variational quantum computing approach can be utilized for solving machine learning tasks. We will see that classification problems for quantum data (i.e. data represented by quantum states) can be solved by means of training the parameters of a variational quantum circuit and a proper post-processing of the results of measurements. Particularly, we will show that the problem of distinguishing between quantum channels can be solved in the variational setting, as we proposed in [4].

4.1 Variational quantum channel discrimination

As was mentioned, with variational quantum circuits one is able to solve the problem of quantum channel discrimination, i.e. to distinguish between a given set of maps sending quantum states to quantum states. This problem is often encountered in quantum communication and quantum information [98–101]. That is, this task is the core of target quantum detection via quantum illumination [102, 103], quantum reading [104], and photonic sensing [105].

The problem of quantum channel discrimination can be conveniently stated as

a game between two parties, Alice and Bob. At the first step of this game, Alice prepares an input state ρ^{in} and sends it to Bob. Bob then randomly selects a channel Φ_j from a collection $\{\Phi_j\}_{j=1}^N$, passes the received state through this channel, and sends the output state $\rho^{\text{out}} = \Phi_y[\rho^{\text{in}}]$ back to Alice. Finally, Alice uses the POVM $\Pi = \{\Pi_j\}_{j=1}^N$ to measure the received state. The result of this measurement is then attributed to y , the label of the channel applied by Bob. The goal of the game for Alice is to find the input state ρ^{in} and the resolution of the identity Π to maximize $p_y = \text{Tr}(\Pi_y \Phi_y[\rho^{\text{in}}])$, the probability to obtain the measurement outcome y equal to the label of the channel applied by Bob. The task of Alice can be stated as the following optimization problem:

$$p_s \equiv \frac{1}{N} \max_{\rho^{\text{in}}, \Pi} \sum_{j=1}^N \text{Tr}(\Pi_j \Phi_j[\rho^{\text{in}}]),$$

i.e. Alice wants to maximize the total probability of successful channel discrimination. The pre-factor $1/N$ comes from the assumption that Bob applies the channels Φ_j with equal probabilities.

Let us consider the situation when there are $N = 2$ channels in the collection $\{\Phi_j\}_{j=0}^{N-1}$, i.e. Alice solves the binary quantum channel discrimination problem for the channels Φ_0 and Φ_1 . Therefore, Alice wants to find an input state ρ^{in} and measurement operators $\Pi = \{\Pi_0, \Pi_1 = \mathbb{1} - \Pi_0\}$ such that they maximize the probabilities of successful discrimination

$$\begin{aligned} p_{00} &= \frac{1}{2} \text{Tr}(\Pi_0 \Phi_0[\rho^{\text{in}}]), \\ p_{11} &= \frac{1}{2} \text{Tr}(\Pi_1 \Phi_1[\rho^{\text{in}}]). \end{aligned}$$

Here, p_{yy} is the probability of obtaining the measurement result $y \in \{0, 1\}$ if Bob applied the channel Φ_y . Note that alongside with the probabilities of obtaining wrong outcomes, p_{01} and p_{10} , the success probabilities sum to unity, $p_{00} + p_{01} + p_{10} + p_{11} = 1$. The task of Alice is therefore to maximize the overall probability of

successful channel discrimination:

$$p_s = \frac{1}{2} \max_{\rho^{\text{in}}, \Pi} \{ \text{Tr}(\Pi_0 \Phi_0[\rho^{\text{in}}]) + \text{Tr}(\Pi_1 \Phi_1[\rho^{\text{in}}]) \}. \quad (4.1)$$

This probability is known to be upper-bounded by the following quantity [106]:

$$p_\diamond = \frac{1}{2} + \frac{1}{4} \|\Phi_0 - \Phi_1\|_\diamond, \quad (4.2)$$

where we have the distance between channels induced by the so-called diamond norm.

Definition 9 (Diamond norm) *Let Φ be a channel acting in $\mathcal{D}(\mathcal{H})$, $\mathbb{1}$ the identity map on $\mathcal{D}(\mathcal{H}_E)$, and ρ a density operator in $\mathcal{D}(\mathcal{H} \otimes \mathcal{H}_E)$. The diamond norm of Φ is defined as*

$$\|\Phi\|_\diamond = \max_{\rho} \|(\Phi \otimes \mathbb{1})[\rho]\|_1, \quad (4.3)$$

where $\|A\|_1 = \text{Tr} \sqrt{A^\dagger A}$.

In the game we described, it is assumed that there is the training stage during which Alice is allowed to find the optimal input state and measurement operators. That is, we assume that for each pair (ρ^{in}, Π) picked by Alice, Bob can provide arbitrary many copies of the state $\rho^{\text{out}} = \Phi_y[\rho^{\text{in}}]$ with the same label $y \in \{0, 1\}$. Moreover, Bob tells to Alice the channel label y for each output state. So, it is assumed that Alice can conduct arbitrarily many measurements of the output state ρ^{out} to exactly compute the probabilities in (4.1). Once Alice has found the optimal ρ^{in} and Π , comes the guessing stage of the game. At this stage, Bob keeps the channel labels y in secret, and Alice is allowed to measure the output state only once. As one may notice, the described game can be considered as a machine learning task of binary classification.

4.1.1 Discrimination strategies: parallel and sequential

In the channel guessing game we described, Bob applies a channel Φ_y only once. In this section, we consider the situation when Alice can ask Bob to apply a chosen

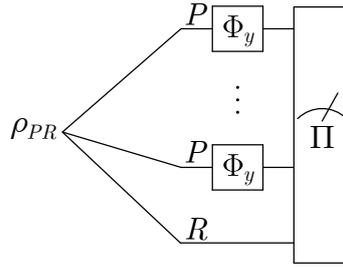


Figure 4-1: A schematic of the parallel strategy for channel discrimination with p allowed channel applications. First, Alice takes two registers, P of p qubits and R of r qubits, and prepares a composite $(p + r)$ -qubit state ρ_{PR}^{in} . Then the qubits of the register P are sent to Bob, who applies the channel Φ_y on each qubit, and sends the qubits back to Alice; meanwhile, the register R remains unaffected. Finally, Alice measures all qubits of the output state with the POVM Π . The figure is taken from [4].

channel Φ_y a finite and fixed number of times p . Provided that for each application the channel label y remains the same, Alice can adjust the discrimination strategy by asking Bob to apply the channel Φ_y in a specific way. In this case, Alice can apply two different approaches for the channel discrimination game: the parallel and the sequential strategy. In the following sections we describe these strategies with the assumption that the channels Φ_y transform qubit states into qubit states.

Parallel strategy

As comes from its name, in the parallel strategy a channel Φ_y is applied p times such that it acts simultaneously on the separate subsystems of a composite input state ρ^{in} . Therefore, the probe state of Alice must contain at least p qubits, which are separately acted by the channels of Bob. Practically, Alice can also have $r \geq 0$ qubits added to the input state, as it allows to have an entangled probe state which can help in solving quantum channel discrimination problems [107–109]. When the output state is received by Alice from Bob, it is measured with a POVM Π on $(p + r)$ qubits. The described discrimination strategy is schematically shown in Figure 4-1

The formal description of the parallel strategy is the following. Suppose that Bob allows Alice to apply the channel Φ_y the number of times p . Then, first, Alice prepares an input state ρ_{PR}^{in} of $(p + r)$ qubits, where p qubits belong to the register P and r qubits to the register R . After that, Alice sends the qubits of the register P

to Bob, who applies a channel Φ_y simultaneously to all the qubits of the register; at the same time, the register R is kept by Alice untouched. From Bob, Alice receives the state

$$\rho^{\text{out}} = (\Phi_y^{\otimes p} \otimes \mathbb{1}^{\otimes r})[\rho_{PR}^{\text{in}}], \quad (4.4)$$

i.e. the channel $\Phi_y^{\otimes p}$ acts only on the qubits of the register P , and the qubits of R are acted trivially by $\mathbb{1}^{\otimes r}$. The output state ρ^{out} is then measured by Alice with the POVM $\Pi = \{\Pi_0, \Pi_1\}$. The expression (4.1) therefore transforms to

$$p_s^{\text{par}} = \frac{1}{2} \max_{\rho_{PR}^{\text{in}}, \Pi} \left\{ \text{Tr}(\Pi_0(\Phi_0^{\otimes p} \otimes \mathbb{1}^{\otimes r})[\rho_{PR}^{\text{in}}]) + \text{Tr}(\Pi_1(\Phi_1^{\otimes p} \otimes \mathbb{1}^{\otimes r})[\rho_{PR}^{\text{in}}]) \right\}. \quad (4.5)$$

Here, alongside with the input state and measurement operators, Alice also implicitly optimizes over r , the size of the auxiliary register R . As we mentioned, these additional qubits allow to have the input state which has entanglement between the qubits of the registers P and R , and which may help achieve probabilities of successful channel discrimination closer to the fundamental upper bound. For the described strategy, this bound is given by

$$p_{\diamond}^{\text{par}}(p) = \frac{1}{2} + \frac{1}{4} \|\Phi_0^{\otimes p} + \Phi_1^{\otimes p}\|_{\diamond} \quad (4.6)$$

for p parallel applications of Bob's channel.

Sequential strategy

The other way to apply a given channel Φ_y is placing it p times in a row. This forms the sequential discrimination strategy, in which Alice's input state ρ^{in} can be single-qubit, and this qubit can be passed through a channel p times step by step. In this setting, each time when Bob applies a channel Φ_y to the input state, the output state is sent back to Alice, who can modify it and send to Bob again. After this procedure is repeated p times, Alice measures the resultant state with a POVM Π . As in the parallel strategy, Alice is allowed to add the auxiliary register R of r qubits to produce a $(1+r)$ -qubit input state. A schematic depiction of the sequential channel discrimination strategy shown in Figure 4-2.

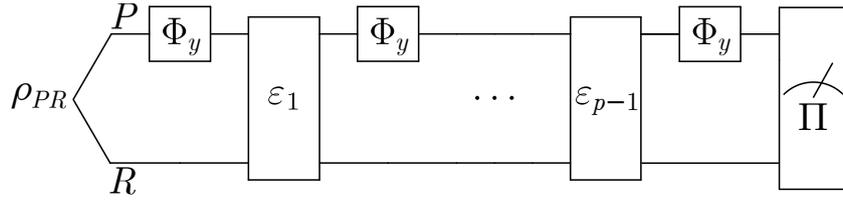


Figure 4-2: A schematic of the sequential channel discrimination strategy. First, Alice prepares a $(1+r)$ -qubit state ρ_{PR}^{in} of one qubit of the register P and r qubits of the register R . The qubit of P is sent to Bob who applies a channel Φ_y , while the register R stays with Alice. The output state is then sent from Bob back to Alice who modifies it with a channel \mathcal{E}_j and sends it back to Bob. Having repeated this procedure p times, Alice measures all the $(1+r)$ qubits a POVM Π . The figure is taken from [4].

Let us describe this strategy more formally. First, Alice prepares the input state ρ_{PR}^{in} of $(1+r)$ qubits, in which one qubit belongs to the register P and the other r qubits are of the register R . The qubit of P is then sent to Bob who applies a channel Φ_y to it, and the register R remains unaffected, so that the output state is $\tilde{\rho}_{PR} = (\Phi_y \otimes \mathbb{1}^{\otimes r})[\rho_{PR}^{\text{in}}]$. Having received this state from Bob, Alice modifies it with a quantum channel ε_1 , which gives $\rho_1 = \varepsilon_1[\tilde{\rho}_{PR}]$. This process is repeated $(p-1)$ times which produces for Alice the state $\rho_{p-1} = \varepsilon_{p-1}[\rho_{p-2}]$. At the very end, Bob applies the channel Φ_y to the qubit of P the p th time, which yields $\rho^{\text{out}} = (\Phi_y \otimes \mathbb{1}^{\otimes r})[\rho_{p-1}]$. We can describe the whole process by the equation

$$\rho^{\text{out}} = \mathcal{C}(\Phi_y, \mathcal{E})[\rho_{PR}^{\text{in}}], \quad (4.7)$$

where $\mathcal{E} = \{\varepsilon_j\}_{j=1}^{p-1}$ are the channels of Alice, and

$$\mathcal{C}(\Phi_y, \mathcal{E}) = (\Phi_y \otimes \mathbb{1}^{\otimes r}) \circ \varepsilon_{p-1} \circ \dots \circ (\Phi_y \otimes \mathbb{1}^{\otimes r}) \circ \varepsilon_2 \circ (\Phi_y \otimes \mathbb{1}^{\otimes r}) \circ \varepsilon_1 \circ (\Phi_y \otimes \mathbb{1}^{\otimes r}),$$

where by \circ we denoted the channel composition, i.e. $(B \circ A)[\rho] \equiv B[A[\rho]]$. The output state ρ^{out} is then measured by Alice with a POVM $\Pi = \{\Pi_0, \Pi_1\}$. The optimization problem (4.1) for this strategy transforms to

$$p_s^{\text{seq}} = \frac{1}{2} \max_{\rho_{PR}^{\text{in}}, \Pi, \mathcal{E}} \left\{ \text{Tr}(\Pi_0 \mathcal{C}(\Phi_0, \mathcal{E})[\rho_{PR}^{\text{in}}]) + \text{Tr}(\Pi_1 \mathcal{C}(\Phi_1, \mathcal{E})[\rho_{PR}^{\text{in}}]) \right\}, \quad (4.8)$$

where alongside with the input state ρ_{PR}^{in} and measurement operators Π , Alice also optimizes over the channels \mathcal{E} .

The sequential strategy incarnates the idea of the so-called quantum comb [105, 110]. For this strategy, however, there is no known upper bound for the maximal success probability similar to that for the parallel strategy (4.6). Nonetheless, compared to the parallel strategy, it is known that the sequential strategy allows to achieve higher successful discrimination probabilities [111, 112]. There is also the so-called indefinite casual order of channel application approach [113], which is a more general discrimination strategy, and it is known to provide even better results. That is, if p^{par} , p^{seq} , and p^{ico} are the upper bounds for successful discrimination probabilities of the parallel, sequential, and indefinite casual order strategies, respectively, then it is proven that there exists a pair of channels Φ_0 and Φ_1 such that

$$p^{\text{par}} < p^{\text{seq}} < p^{\text{ico}}.$$

4.1.2 Variational formulation of the discrimination strategies

In this section, we embed the parallel and sequential discrimination strategies into the framework of variational quantum computing. We will do it by replacing all quantum state transformations with parametrized unitary operators, i.e. by variational quantum circuits. This also requires to accordingly modify the equations (4.5) and (4.8), the optimization problems for p_s^{par} and p_s^{seq} , respectively. A similar work has been done in [114], where the authors considered solving the quantum channel discrimination problem for $p = 1$ using the variational approach.

Parallel strategy

Our embedding of the parallel channel discrimination strategy (4.5) into the framework of variational quantum computing can be expressed by the circuit depicted in Figure 4-3. This circuit prepares the input state of Alice ρ_{PR}^{in} as

$$\rho_{PR}^{\text{in}} = \mathcal{U}(\boldsymbol{\theta}_0)[\rho_0(p, r)],$$

where $\rho_0(p, r) = |0\rangle\langle 0|_P^{\otimes p} \otimes |0\rangle\langle 0|_R^{\otimes r}$, $\mathcal{U}(\boldsymbol{\theta}_0)[\rho] = U(\boldsymbol{\theta}_0)\rho U^\dagger(\boldsymbol{\theta}_0)$, and $U(\boldsymbol{\theta}_0)$ is a variational ansatz with $\boldsymbol{\theta}_0 \subset \mathbb{R}$. After this transformation, as done in (4.4), Bob acts by a channel Φ_y on all the p qubits of the register P of this state, leaving the qubits of the register R unaffected. Then Alice transforms the output state by a unitary $U(\boldsymbol{\theta}_1)$, which actually serves as a rotation of the basis for measurements. Therefore, the state prepared by the circuit before the measurement is

$$\rho^{\text{out}}(\boldsymbol{\theta}, \Phi_y, p, r) = \mathcal{U}(\boldsymbol{\theta}_1) \circ (\Phi_y^{\otimes p} \otimes \mathbb{1}^{\otimes r}) \circ \mathcal{U}(\boldsymbol{\theta}_0)[\rho_0(p, r)], \quad (4.9)$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}_0 \cup \boldsymbol{\theta}_1$. Consider the following resolution of the identity in the Hilbert space of the registers P and R :

$$\mathbb{1} = \sum_{j=1}^d |j\rangle\langle j|,$$

where $\{|j\rangle\}_{j=1}^d$ is the computational basis in \mathbb{C}^d with $d = 2^{p+r}$. Therefore, the simplest way for Alice to choose the POVM operators is

$$\Pi_0 = \sum_{j=1}^{d/2} |j\rangle\langle j|, \quad \Pi_1 = \sum_{j=d/2+1}^d |j\rangle\langle j|. \quad (4.10)$$

The measurement defined by these operators obviously reduces to measurements conducted in the computational basis. Thus, the optimization problem (4.5) for the parallel strategy embedded into the variational quantum computing framework becomes

$$p_s^{\text{par}} = \frac{1}{2} \max_{\boldsymbol{\theta}, r} \left\{ \text{Tr}(\Pi_0 \rho^{\text{out}}(\boldsymbol{\theta}, \Phi_0, p, r)) + \text{Tr}(\Pi_1 \rho^{\text{out}}(\boldsymbol{\theta}, \Phi_1, p, r)) \right\}, \quad (4.11)$$

i.e. we transitioned from the optimization over the input states ρ_{PR}^{in} and measurements Π to the optimization over real numbers $\boldsymbol{\theta}$.

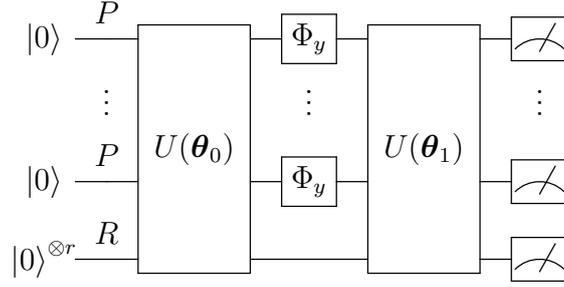


Figure 4-3: Parallel quantum channel discrimination strategy implemented as a variational circuit. The circuit $U(\theta_0)$ prepares the input state of Alice ρ_{PR}^{in} from the initial state $|0\rangle_P^{\otimes p} \otimes |0\rangle_R^{\otimes r}$. After this transformation, Bob applies a channel Φ_y to the qubits of the register R , and the output state is sent to Alice to be measured. To rotate the measurement basis, Alice uses the circuit $U(\theta_1)$. The figure is taken from [4].

Sequential strategy

Let us now consider a quantum circuit embedding of the parallel channel discrimination strategy (4.8) depicted in Figure 4-4. Recall that in this strategy, Alice operates with a set of channels $\mathcal{E} = \{\varepsilon_j\}_{j=1}^{p-1}$ which are used for transforming the probe state after each application of the channel Φ_y by Bob. In a variational circuit, a channel can be applied via the so-called Stinespring representation [43]. For example, to implement $\varepsilon[\rho_{PR}]$, one can add an auxiliary register E for e qubits prepared in a state ρ_E , and then perform a unitary evolution U of the joint state $\rho_{PR} \otimes \rho_E$, after which one traces out the subsystem E . Thus, one has

$$\varepsilon[\rho_{PR}] = \text{Tr}_E [U(\rho_{PR} \otimes \rho_E)U^\dagger]. \quad (4.12)$$

To obtain a general transformation in this representation, one has to have e qubits in the register E such that e is twice more than the total size of the registers P and R [64]. In our case, we have $e = 2(1 + r)$, with one qubit from the register P and r qubits from R .

In the variational circuit in Figure 4-4, we incorporate the register E into R , the register with auxiliary qubits. As we described, the channels $\mathcal{E} = \{\varepsilon_j\}_{j=1}^{p-1}$ are implemented via the Stinespring representation as parametrized unitaries $\{U(\theta_j)\}_{j=1}^{p-1}$. In the circuit, there are also the operators $U(\theta_0)$ and $U(\theta_p)$, which prepare the initial state rotate the measurement basis, respectively. Similar to (4.7), the state prepared

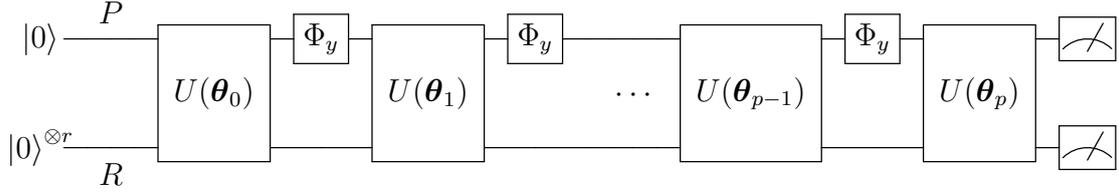


Figure 4-4: Sequential quantum channel discrimination strategy A variational quantum circuit implementing the sequential channel discrimination strategy from Figure 4-2. The Alice's channels $\mathcal{E} = \{\varepsilon_j\}_{j=1}^{p-1}$ are replaced by the parametrized unitaries $\{U(\theta_j)\}_{j=1}^{p-1}$, where p is the number of allowed applications of the channel Φ_y . The input state ρ_{PR}^{in} is prepared from $|0\rangle_P \otimes |0\rangle_R^{\otimes r}$ via the unitary transformation $U(\theta_0)$, while $U(\theta_p)$ is used to rotate the measurement basis. This method necessitates $(1 + r)$ qubits, with one qubit in the register P and r qubits in the register R . In analogy with the parallel strategy, r might be set to zero. The figure is taken from [4].

by the circuit before the measurement is

$$\rho^{\text{out}}(\theta, \Phi_y, p, r) = \mathcal{C}(\theta, \Phi_y, p, r)[\rho_0(r)], \quad (4.13)$$

where $\rho_0(r) = |0\rangle\langle 0|_P \otimes |0\rangle\langle 0|_R^{\otimes r}$ and

$$\begin{aligned} \mathcal{C}(\theta, \Phi_y, p, r) &= \mathcal{U}(\theta_p) \circ (\Phi_y \otimes \mathbb{1}^{\otimes r}) \circ \mathcal{U}(\theta_{p-1}) \circ \dots \\ &\quad \circ \mathcal{U}(\theta_2) \circ (\Phi_y \otimes \mathbb{1}^{\otimes r}) \circ \mathcal{U}(\theta_1) \circ (\Phi_y \otimes \mathbb{1}^{\otimes r}) \circ \mathcal{U}(\theta_0), \end{aligned}$$

with $\theta = \bigcup_{k=0}^p \theta_k$. In the equation above, the channels \mathcal{U} act as in (4.9). Finally, the output state (4.13) is measured in the computational basis, and from the results of this measurement one can reconstruct the results of the measurement conducted with the POVM elements (4.10) with $d = 2^{1+r}$. Therefore, the optimization problem (4.8) becomes

$$p_s^{\text{seq}} = \frac{1}{2} \max_{\theta, r} \{ \text{Tr}(\Pi_0 \rho^{\text{out}}(\theta, \Phi_0, p, r)) + \text{Tr}(\Pi_1 \rho^{\text{out}}(\theta, \Phi_1, p, r)) \}. \quad (4.14)$$

As one may notice, the expressions for p_s^{par} in (4.11) and p_s^{seq} in (4.14) look the same; different only are the structure and the genesis of the output state ρ^{out} .

4.1.3 Numerical experiments

In the following sections, we discuss the results of our numerical experiments for testing the variational quantum channel discrimination approach we described. In our simulations, we fix the number of allowed channel applications to be $p = 1, 2$ and consider different values of r , the size of auxiliary register R . To represent $U(\boldsymbol{\theta}_k)$, we apply the hardware-efficient ansatz, a four-qubit example of which is shown in Figure 2-1. In what follows, we demonstrate the capabilities of our method to discriminate the depolarizing channels with different depolarization factors, and entanglement breaking channels which map two-qubit states into one-qubit states. As the optimization method, we used the L-BFGS-B algorithm [115] for all our numerical simulations, when applicable.

Entanglement breaking channel discrimination

Consider the following entanglement breaking channels

$$\Phi_0[\rho] = \sum_{j=1}^5 A_j \rho A_j^\dagger, \quad \Phi_1[\rho] = \sum_{j=1}^5 B_j \rho B_j^\dagger, \quad (4.15)$$

described by the Kraus operators

$$\begin{aligned} A_1 &= |0\rangle\langle 00|, & A_2 &= |0\rangle\langle 01|, & A_3 &= |0\rangle\langle 10|, \\ A_4 &= \frac{1}{\sqrt{2}} |0\rangle\langle 11|, & A_5 &= \frac{1}{\sqrt{2}} |1\rangle\langle 11|, \end{aligned} \quad (4.16)$$

$$\begin{aligned} B_1 &= |+\rangle\langle 00|, & B_2 &= |+\rangle\langle 01|, & B_3 &= |1\rangle\langle 1+|, \\ B_4 &= \frac{1}{\sqrt{2}} |0\rangle\langle 1-|, & B_5 &= \frac{1}{\sqrt{2}} |1\rangle\langle 1-|, \end{aligned} \quad (4.17)$$

where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. The parallel strategy applied for distinguishing between these channels is known to give $p_{\diamond}^{\text{par}}(p) < 1$ for any finite p , i.e. the success probability (4.6) never reaches unity regardless the number of channel applications [116]. It has also been shown that the sequential strategy with only $p = 2$ applications allows to discriminate the channels perfectly. Computing the upper bounds for the success probabilities (4.6), one obtains $p_{\diamond}^{\text{par}} \approx 0.9268$ for $p = 1$ and $p_{\diamond}^{\text{par}} \approx 0.9771$ for

$p = 2$. To calculate these values, it is needed to compute the diamond norm (4.3), which can be done via semi-definite programming [117]; one can do it numerically using e.g. the CVXPY package [118, 119].

Our numerical experiments are based on training the circuits similar to the ones depicted in Figures 4-3 and 4-4, and the training is performed by solving the maximization problems (4.11) and (4.14), respectively. Recall that the entanglement-breaking channels (4.15) transform two-qubit states into one-qubit states, so the respective circuits must be rebuilt accordingly. For $p = 2$ channel applications, the explicit circuits for the parallel and sequential strategies are shown in Figure 4-5; note that for $p = 1$, the circuits for the both strategies would look the same. In our numerical simulations, for $p = 2$ we achieved the success probability $p_s^{\text{par}} \approx p_\diamond^{\text{par}} \approx 0.9771$ with the use of $l = 5$ layers of the hardware-efficient ansatz for representing the parametrized unitaries $U(\boldsymbol{\theta}_0)$ and $U(\boldsymbol{\theta}_1)$. In the meantime, to achieve $p_s^{\text{seq}} \approx 1$ for the sequential strategy with $p = 2$, we needed only $l = 1$ layer of the ansatz. For both strategies, we did not use the auxiliary register R , i.e. we set $r = 0$. The case $p = 1$ — when there is no difference between the strategies — gives $p_s \approx p_\diamond \approx 0.9268$ with one layer of the ansatz.

The circuits we used to implement the strategies are likely excessively expressive, i.e. the ansätze contain too many parameters. As we mentioned, in [116] it was shown that one needs only a separable input state and a single measurement to perfectly discriminate the channels (4.15) using the sequential strategy with $p = 2$ channel applications. Therefore, the important result for us is that even with the (over)parametrized circuits we applied it is possible to find the input states and measurement bases giving $p_s^{\text{par}} \approx p_\diamond^{\text{par}} \approx 0.9771$ and $p_s^{\text{seq}} \approx 1$. The crucial detail here is that the variational approach we propose needs almost no prior knowledge about the channels Φ_y . That is, the only information we need is the size of the input and output states.

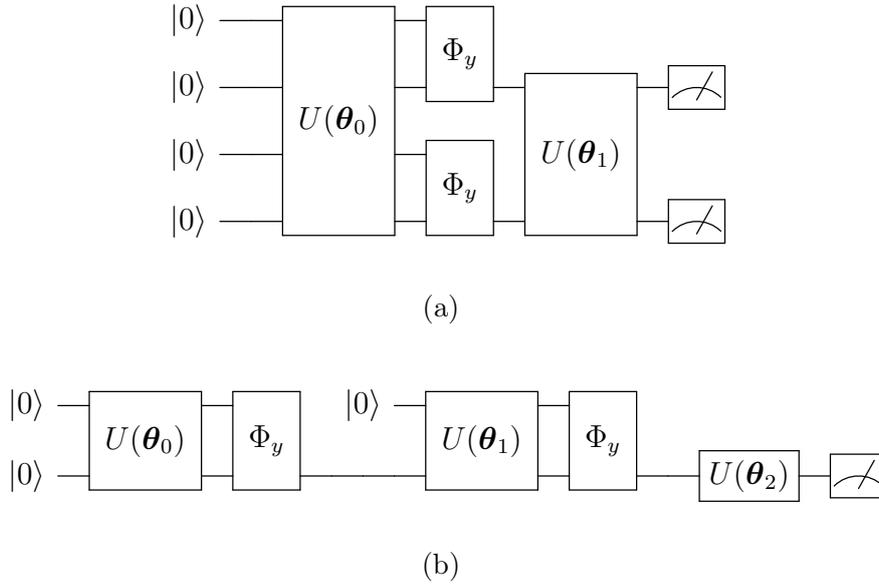


Figure 4-5: Explicit variational quantum circuits implementing (a) the parallel and (b) sequential strategies discriminating the entanglement breaking channels (4.15) with $p = 2$ applications. Since the considered channels map two-qubit states into one-qubit states, we need to add one qubit in some state (in our case, $|0\rangle$) after the first application of Φ_y .

Depolarizing channel discrimination

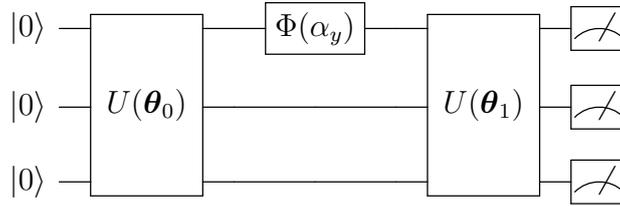
We continue with studying the variational approach applied for the discrimination of the depolarizing channels of the form

$$\Phi(\alpha)[\rho] = (1 - \alpha)\rho + \frac{\alpha}{3}(\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z) \quad (4.18)$$

where α is the depolarization factor, and σ_i stands for a Pauli operator with $i = x, y, z$. In what follows, we train variational circuits to distinguish between a pair of such channels with different factors $0 \leq \alpha_0 \neq \alpha_1 \leq 1$. As for the case of the entanglement-breaking channels, we considered the numbers of channel applications $p = 1, 2$.

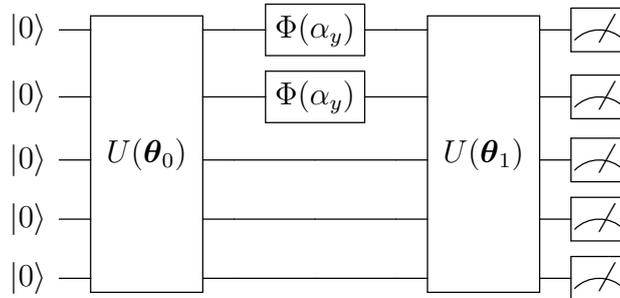
First, let us study the case of $p = 1$, which gives the same quantum circuit for the both strategies. We found that the maximum discrimination efficiency in this

setting is achieved with the following circuit:

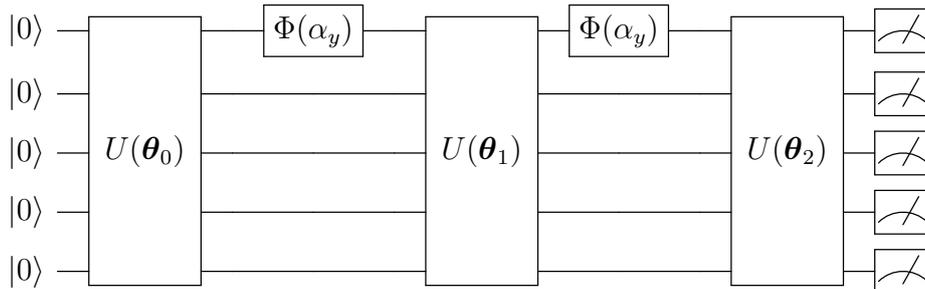


As can be seen, we added $r = 2$ qubits to be the auxiliary register R . The operators $U(\theta_0)$ and $U(\theta_1)$ were represented by the hardware-efficient ansatz of different number of layers l . In Figure 4-6 shown are the achieved probabilities of successful discrimination of the depolarizing channels with parameters α_0 and α_1 . In this Figure, the results are given for the cases $l = 1$ and $l = 2$. As can be seen, two ansatz layers is enough for achieving the success probabilities given by the diamond distance (4.2). Additionally, as one could expect, we observe that the larger is the distance between the parameters α_0 and α_1 — which, in fact, implies the larger diamond distance — the lower is the achieved success probability.

Now let us consider the strategies with $p = 2$. The parallel strategy was implemented as the following circuit:



and the circuit which was applied for the sequential strategy is



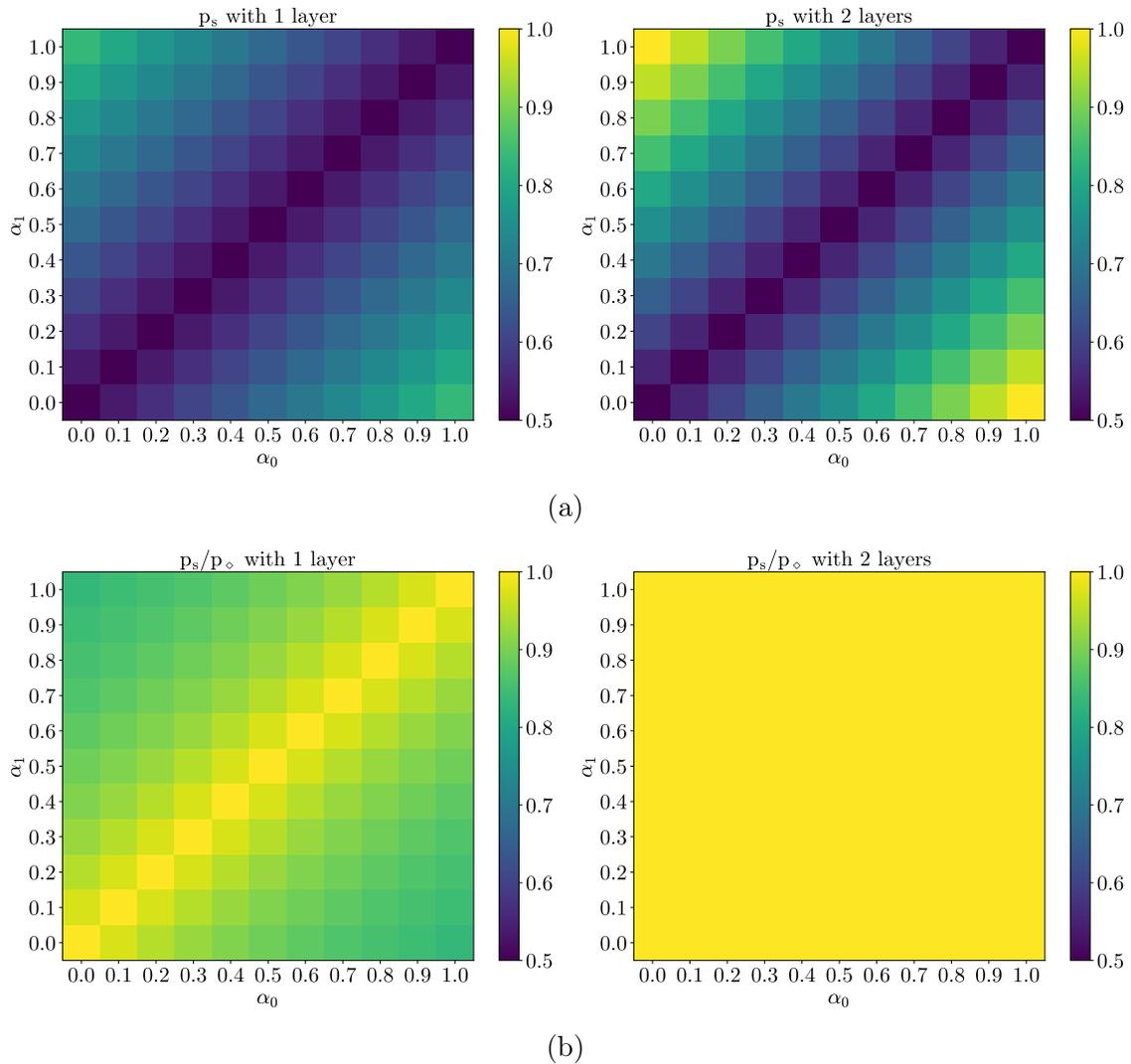
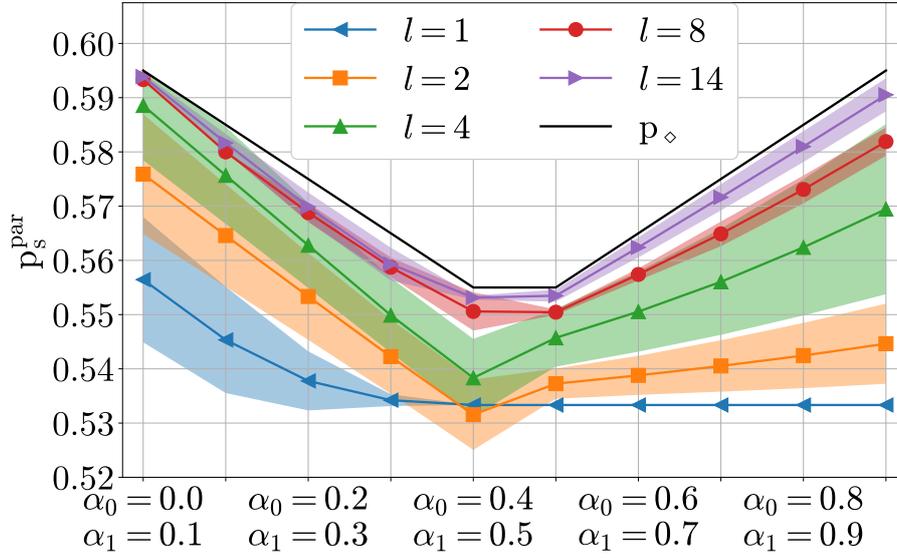


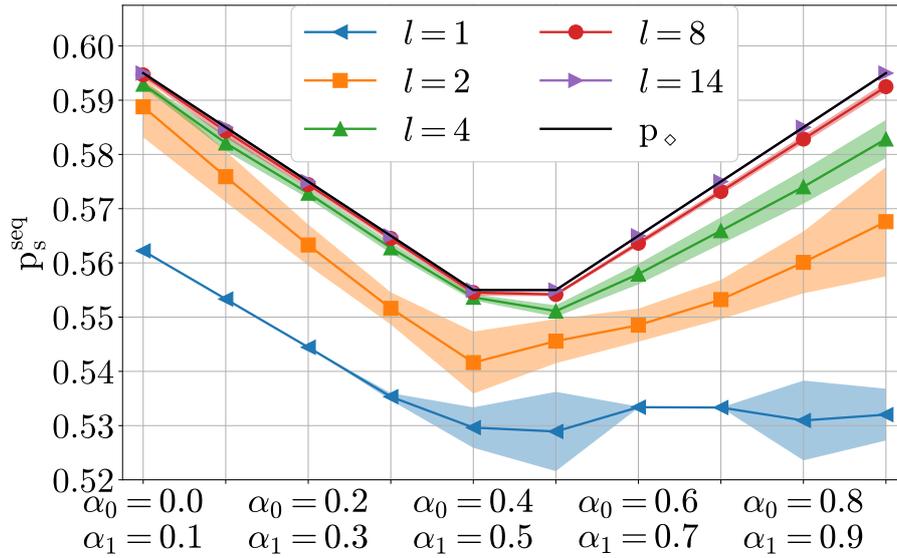
Figure 4-6: Results of numerical experiments with variational discrimination of the depolarizing channels with parameters α_0 and α_1 . Panel (a) shows the achieved probabilities of successful discrimination p_s . In panel (b) shown are the ratios of the success probability p_s to the maximal achievable probability p_\diamond . The data on the panels on the left are obtained with $l = 1$ layer of the hardware-efficient ansatz while the right panels show the data for $l = 2$ layers.

As one may notice, as the size of the auxiliary register R , for the parallel strategy we set $r = 3$, and for the sequential strategy we put $r = 4$, so that the total number of qubits for the both strategies is 5. In Figure 4-7 we show the successful discrimination probabilities (4.11) and (4.14) achieved with the parallel and sequential strategies applied for distinguishing between the depolarizing channels with the factors α_0 and $\alpha_1 = \alpha_0 + 0.1$. This Figure also shows the dependence of the achieved probabilities p_s on l , the number of layers of hardware-efficient ansatz parametrizing the operators $U(\theta_j)$. We observe that the sequential strategy shows higher performance and smaller variance of the results with fewer layers. That is, it is enough to have $l = 14$ layers to achieve the success probability p_\diamond^{par} for all pairs (α_0, α_1) . On the other hand, with the parallel strategy, we cannot reach this result, even with higher numbers of layers as we tried.

In Figure 4-7, one can also observe an interesting effect. The plot for the probabilities p_\diamond^{par} is symmetric, which implies symmetric diamond distances between the channels, $\|\Phi^{\otimes 2}(\alpha_0) - \Phi^{\otimes 2}(\alpha_1)\|_\diamond$. For instance, the pairs $(\alpha_0 = 0.0, \alpha_1 = 0.1)$ and $(\alpha_0 = 0.9, \alpha_1 = 1.0)$ both give the same upper bounds for the success probability, $p_\diamond^{\text{par}} = 0.595$. Despite this, we see that the achieved probabilities p_s for $(\alpha_0 = 0.0, \alpha_1 = 0.1)$ are higher than that for $(\alpha_0 = 0.9, \alpha_1 = 1.0)$. Overall, for the variational channel discrimination method, it is easier to obtain higher p_s for depolarization factor pairs on the left to $\alpha = 0.5$. Interestingly, this agrees with the behaviour of the function $\text{Tr}(\rho_0\rho_1)$, the trace of the product of output states of the channels, $\rho_y \equiv \Phi(\alpha_y)[\rho]$, and the so-called trace distance between these states $\frac{1}{2}\|\rho_0 - \rho_1\|_1$, which serves as a measure of the distinguishability of density operators [120]. In Figure 4-8, we plot the correlation coefficients between the achieved probabilities p_s plotted in Figure 4-7 and the trace of the product of the output states, their trace distances, and the diamond distance between the channels. What we also observe is that the more layers of the ansatz l we use, the less performance of the variational approach depends on the trace of the product and the trace distance. It is also worthy of note that the function $\kappa(\rho_i, \rho_j) = \text{Tr}(\rho_i\rho_j)$ serves as a kernel in kernel-based quantum machine learning [121].



(a)



(b)

Figure 4-7: Achieved probabilities of successful discrimination p_s between two depolarizing channels with depolarization factors α_0 and α_1 . Panel (a) shows the results obtained for the parallel strategy, while in panel (b) shown are the results for the sequential strategy. Different colors indicate different values of l , the number of layers in the hardware-efficient ansatz. The probabilities p_s averaged over 10 independent optimizations are shown as marks connected by solid lines, while standard deviations are shown as shaded areas. The maximum achievable probability for the parallel strategy p_\diamond^{par} for $p = 2$ is indicated by the black solid line. The figures are taken from [4].

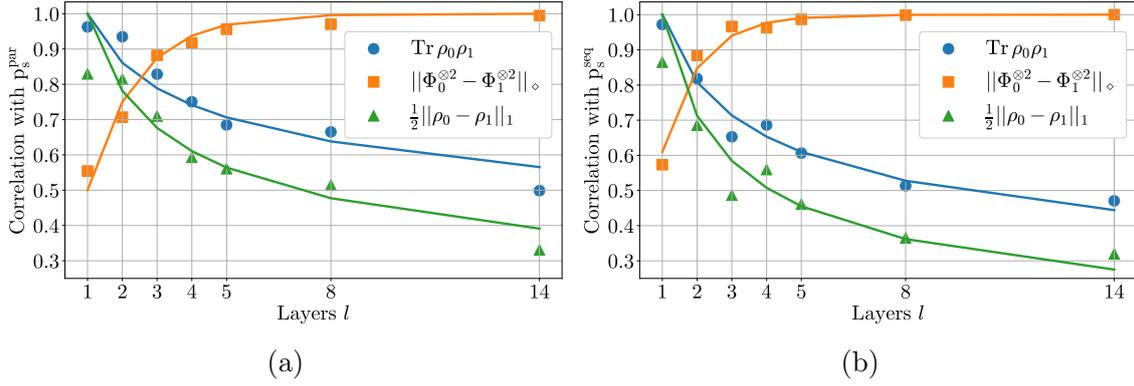


Figure 4-8: Correlation coefficients between the average achieved success probabilities p_s (shown in Figure 4-7) and the trace of the product of the channel output states $\rho_y = \Phi_y^{\otimes 2}[\rho]$ with $y \in \{0, 1\}$ (blue circles), the trace distance between them (green triangles), and the diamond distance between the channels Φ_y (orange squares). Here, $\Phi_y \equiv \Phi(\alpha_y)$ and the input state was set to $\rho = |00\rangle\langle 00|$. Panels (a) and (b) show the correlations for the parallel and sequential strategies, respectively. The correlations are plotted versus l , the number of layers of the ansatz. Solid lines show the functions $f(l) = l^{-1/a}$ (blue and green) and $g(l) = 1 - e^{-bl}$ (orange), which fit the data points of the corresponding colors. The figures are taken from [4].

4.1.4 Remarks on quantum channel discrimination

Let us discuss some aspects of the quantum channel discrimination problem solved with variational circuits. Consider the depolarizing channel defined in (4.18). With this definition, our depolarizing channel is actually a particular realization of the so-called Pauli channel which is well-studied in the context of channel discrimination [122]. That is, it is known that the Bell state $|B\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is optimal for discriminating between any pair of Pauli channels. Once we know an optimal input state ρ , one finds the optimal POVM elements Π_0 and Π_1 to be such that

$$(\Phi_0 \otimes \mathbb{1})[\rho] - (\Phi_1 \otimes \mathbb{1})[\rho] = \sum_{\lambda_j \geq 0} \lambda_j \Pi_0 + \sum_{\lambda_j < 0} \lambda_j \Pi_1, \quad (4.19)$$

where λ_j the eigenvalues of the Hermitian operator $((\Phi_0 \otimes \mathbb{1})[\rho] - (\Phi_1 \otimes \mathbb{1})[\rho])$. Since an optimal input state is generally not known to us, we can find the optimal measurement operators by applying the Naimark extension theorem [123]. Namely, the measurement of a state ρ with a general observable $\Pi = \{\Pi_0, \Pi_1\}$ can be realized

by finding a unitary operator U such that

$$\begin{aligned} p_y &= \text{Tr } \Pi_y \rho \\ &= \text{Tr} \left[(\mathbb{1} \otimes |y\rangle\langle y|_E) U (\rho \otimes |0\rangle\langle 0|_E) U^\dagger \right], \end{aligned}$$

where $\{|y\rangle\langle y|_E\}_{y \in \{0,1\}}$ and $|0\rangle\langle 0|_E$ are an orthogonal POVM and a density operator acting in the auxiliary space $\mathcal{H}_E = \mathbb{C}_2$, respectively.

Therefore, for distinguishing between depolarizing channels (4.18) with $p = 1$ channel application, one may need two auxiliary qubits: One for preparing the input state $\rho = |B\rangle\langle B|$, and one for conducting a general measurement of the output state $\Phi_y[\rho]$. Recall that in our numerical experiments we as well introduced two auxiliary qubits and achieved the upper bound of the success probability (4.2). This signifies that our straightforward variational approach is able to find the optimal input (Bell) state, and implement optimal measurements with the measurement operators which can be found as shown in (4.19).

So, the Bell state $|B\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is optimal for distinguishing between depolarizing channels. It turns out that this state is also optimal for discriminating between any pair of n -qubit channels Φ_y , $y \in \{0, 1\}$, as long as for any Pauli string $\sigma \in \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}$ and any input state ρ there exists a unitary V such that

$$\Phi_y[\sigma \rho \sigma] = V \Phi_y[\rho] V^\dagger. \quad (4.20)$$

This property is called the teleportation covariance of the channel Φ_y , and this name comes from the notion of teleportation unitaries for quantum teleportation [124]. For teleportation covariant channels Φ_0 and Φ_1 , one has [125]

$$\|\Phi_0 - \Phi_1\|_\diamond = \|\chi_0 - \chi_1\|_1,$$

where

$$\chi_y = (\Phi_y \otimes \mathbb{1})[|B\rangle\langle B|]$$

is the so-called Choi state for the channel Φ_y , and $|B\rangle$ is the Bell state. That is,

the diamond-norm distance between teleportation covariant channels — for example, depolarizing channels — is the same as the trace norm distance between the corresponding Choi states.

An example of a channel for which the property of teleportation covariance (4.20) does not hold could be the so-called amplitude damping channel

$$\begin{aligned}\Phi(\gamma)[\rho] &= A(\gamma)\rho A^\dagger(\gamma) + B(\gamma)\rho B^\dagger(\gamma), \\ A(\gamma) &= \sqrt{1-\gamma}|0\rangle\langle 1|, \quad B(\gamma) = |0\rangle\langle 0| + \sqrt{\gamma}|1\rangle\langle 1|.\end{aligned}\tag{4.21}$$

The optimal state for discriminating a pair of channels $\Phi(\gamma_0)$ and $\Phi(\gamma_1)$ in the one-time application setting is known to be [126]

$$|B(t)\rangle = \sqrt{t}|00\rangle + \sqrt{1-t}|11\rangle,$$

where

$$t = \max\left\{0, \frac{1}{2 - \sqrt{\gamma_0} - \sqrt{\gamma_1}}\right\}.$$

Another instance of a channel which is not teleportation covariant is any of the two entanglement breaking channels (4.15). Recall that in our numerical experiments we applied the variational approach for distinguishing between these channels and obtained the known upper bounds on the probability of successful discrimination for the parallel and sequential strategies.

4.2 Variational quantum channel classification

So far we have discussed and tested the variational quantum computing approach applied for solving the quantum channel discrimination problem. In this section, we consider the notion of variational quantum circuits from the point of view of machine learning. More precisely, we will show that a variational quantum circuit can serve as a building block for a quantum classifier, i.e. a means for solving classification tasks for quantum data. Such classifiers have already showed their efficiency in, for example, classifying phases of matter [127–129]. More generally, if one has a

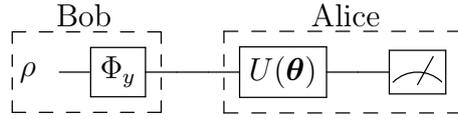


Figure 4-9: A schematic of the variational quantum channel classifier. First, Bob prepares a random state ρ , passes it through a channel Φ_y , and sends the output state $\Phi_y[\rho]$ to Alice. Alice then applies the unitary $U(\boldsymbol{\theta})$, measures the resultant state in the computational basis, and computes the prediction value p defined in (4.23). The predictions p are then used for training the circuit, which is done by minimizing the square distances (4.24) between the predictions p and the true labels y .

quantum classifier trained on labeled data points (e.g. quantum states of different phases), then one can predict the unknown label of a given datum. In this section, we build a quantum classifier trained to distinguish between two depolarizing channels of the form (4.18) with two different depolarization factors α_0 and α_1 .

As we did for the variational channel discrimination discussed in Section 4.1, let us describe the process of building a quantum channel classifier as a game between Alice and Bob. The goal of Alice in this game is to train a variational circuit $U(\boldsymbol{\theta})$ such that having received from Bob a state passed through a channel Φ_y , Alice can predict its label y . In contrast to the variational channel discrimination, in this game Alice has no control on the input state ρ^{in} . This time, this state is prepared by Bob, and each time it is random and mixed. It is also assumed that Alice allowed to measure the output states arbitrarily many times. Schematically, this game is shown in Figure 4-9.

More formally, the game can be described as follows. As the first step, Bob fixes two values of the depolarization factor $0 \leq \alpha_0 \neq \alpha_1 \leq 1$. Then Bob picks a label $y \in \{0, 1\}$, prepares a random (in general, mixed) qubit state ρ , and passes it through the corresponding depolarizing channel, which gives the output state $\rho_y = \Phi(\alpha_y)[\rho]$. Next, Bob sends this state to Alice who uses it as an input for the variational circuit $U(\boldsymbol{\theta})$, which outputs the state

$$\rho(\alpha_y, \boldsymbol{\theta}) = U(\boldsymbol{\theta})\rho_y U^\dagger(\boldsymbol{\theta}). \quad (4.22)$$

For this state, Alice measures the observable σ_z for calculating the value

$$p(\boldsymbol{\theta}) = \frac{1}{2}(1 + \text{Tr}[\rho(\alpha_y, \boldsymbol{\theta}) \sigma_z]), \quad (4.23)$$

which is used to predict the label y . Therefore, the goal of the game is to train the circuit $U(\boldsymbol{\theta})$ such that given a state ρ_y , Alice is able to predict the label y the channel using the prediction value (4.23). That is, for some $0 < b < 1$, Alice concludes that $y = 0$ if obtained is a value $p \leq b$, and $y = 1$ otherwise.

For training the circuit, Alice receives from Bob a training set $\{\rho_{y_j}^j, y_j\}_{j=1}^{N_{\text{train}}}$, where $y_j \in \{0, 1\}$ are true labels; the superscript j reminds that each input state is randomly generated and, therefore, is different for each j . Alice inputs each state from this set into the circuit $U(\boldsymbol{\theta})$ and computes the prediction values $p_j(\boldsymbol{\theta})$ defined in (4.23). Having this done for all N_{train} states, Alice forms the set $\{p_j(\boldsymbol{\theta}), y_j\}_{j=1}^{N_{\text{train}}}$ and minimizes the square distances between the predictions p_j and the true labels y_j :

$$f(\boldsymbol{\theta}) = \sum_{j=1}^{N_{\text{train}}} (y_j - p_j(\boldsymbol{\theta}))^2. \quad (4.24)$$

This gives the optimal parameters $\boldsymbol{\theta}^{\text{opt}}$ for the circuit $U(\boldsymbol{\theta})$ as

$$\boldsymbol{\theta}^{\text{opt}} \in \arg \min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}),$$

from which Alice obtains the final predictions $p_j(\boldsymbol{\theta}^{\text{opt}})$.

With the true labels y_j and sorted predictions p_j , Alice can iteratively find the value b that separates the classes $y = 0$ and $y = 1$. At the iteration t , for the prediction p_t , Alice assigns all elements that are less than or equal to b to belong to the first class, $(p_1, p_2, \dots, p_t) \in \{‘0’\}$ and to the second class otherwise, $(p_{t+1}, p_{t+2}, \dots, p_{N_{\text{train}}}) \in \{‘1’\}$. Therefore, b is equal to the prediction value p_t which gives the best classification accuracy.

Having trained the circuit $U(\boldsymbol{\theta})$ and found the separation value b , Alice is ready predict unknown labels of given data. To test the classifier, Bob sends to Alice a set $\{\rho_{y_j}^j\}_{j=1}^{N_{\text{test}}}$. This set is similar to the training set, but this time Bob does not provide Alice with the true labels y_j . As at the training stage, Alice takes a state $\rho_{y_j}^j$ from

the test set, inputs it into the circuit $U(\boldsymbol{\theta}^{\text{opt}})$, calculates the prediction value p_j , and assigns the label $y_j = 0$ if $p_j \leq b$ or $y_j = 1$ if $p_j \geq b$.

4.2.1 Numerical experiments

In this section, we show the results of our numerical experiments of testing the described approach for building variational channel classifier. Recall that the classifier is trained to distinguish between the depolarizing channels (4.18) with different depolarization factors α_0 and α_1 . As the variational circuit $U(\boldsymbol{\theta})$ we used the ansatz

$$U(\boldsymbol{\theta}) = R_z(\theta_2)R_x(\theta_1). \quad (4.25)$$

To train and test the classifier, we generated the respective data sets $\{\rho_{y_j}^j, y_j\}_{j=1}^{N_{\text{train}}}$ and $\{\rho_{y_j}^j\}_{j=1}^{N_{\text{test}}}$ of the same size, $N_{\text{train}} = 1000$ and $N_{\text{test}} = 1000$. The results of our numerical experiments for this ansatz used as a channel classifier are shown in Figure 4-10. As we can see, the circuit is able to classify the channels with some accuracy. An interesting observation here is that the classifier shows accuracy when one of the depolarization factors is $\alpha = 0.7$ or 0.8 , i.e. near the point 0.75 . We will elaborate on this fact later.

In our work, we also investigated how the the performance of the obtained classifier depends on the number of data points used to train it. In Figure 4-11 we plot the training and test accuracy against the size of the training set N_{train} for different assignments of the coefficients α_0 and α_1 . On the one hand, for the training accuracy, we see that the classifier splits the data points well if there are few of them; the more data points we add to the training set, the lower is the training accuracy. On the other hand, by increasing the size of the training set, we, as could be expected, increase the test accuracy as well, especially for the case $\alpha_0 = 0$ and $\alpha_1 = 0.1$. Interestingly, we need relatively few data points to train the classifier to distinguish between the channels with $\alpha_0 = 0$, $\alpha_1 = 0.75$, which according to Figure 4-10 gives the maximum accuracy.

Now let us make a small modification of our variational classification method. Namely, in the game between Alice and Bob, let us suppose that instead of sending

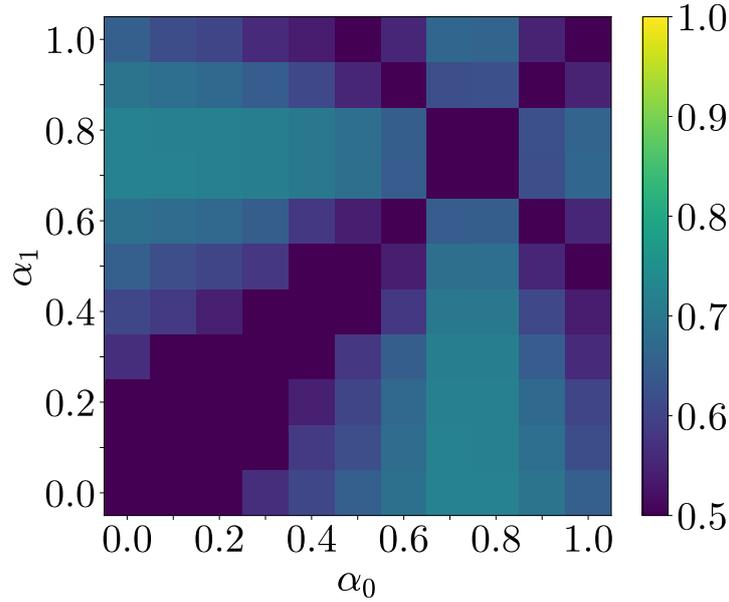


Figure 4-10: Accuracy of quantum channel classifier trained for distinguishing between the depolarizing channels with factors α_0 and α_1 . The classifier is based on the ansatz (4.25) and is trained and tested on the data sets of the sizes $N_{\text{train}} = N_{\text{test}} = 1000$. The figure is taken from [4].

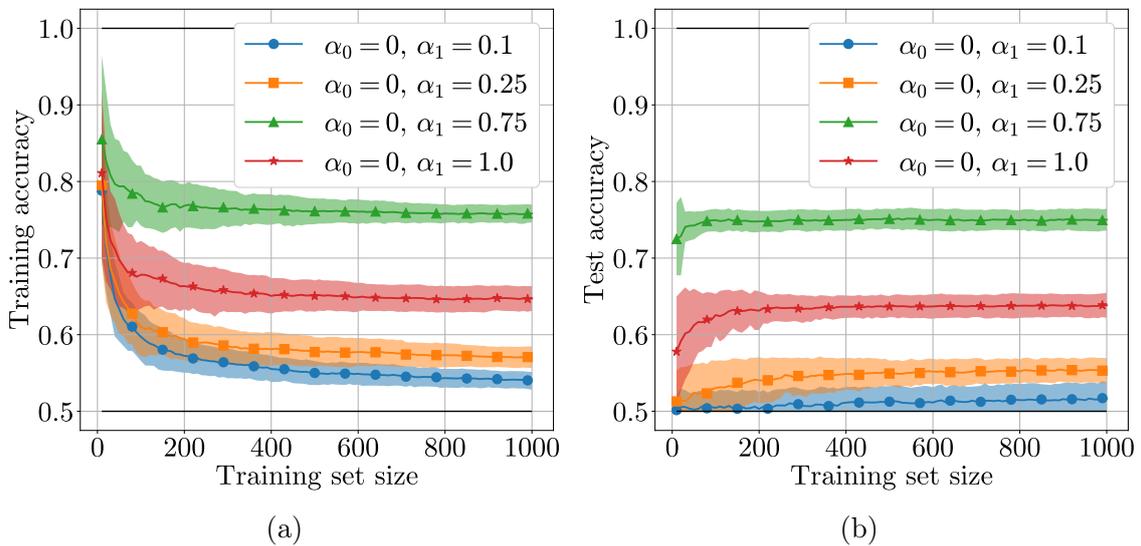


Figure 4-11: Classification accuracy for the depolarizing channels with α_0 and α_1 versus the size of the training set N_{train} . Panels (a) and (b) show the accuracy obtained for the training and test sets, respectively. The size of the test set was fixed to $N_{\text{test}} = 1000$. The color of the curves indicate the values of the coefficients α_0 and α_1 (see legends), and shaded areas show the standard deviation computed over 100 initial circuit parameters.

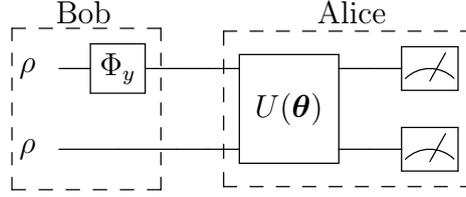


Figure 4-12: A schematic of the modified variational classifier of quantum channels. First, Bob generates two copies of a random state ρ , and passes one of them through a channel Φ_y , outputting the state $\Phi_y[\rho] \otimes \rho$. This state is then sent to Alice who applies the unitary $U(\boldsymbol{\theta})$, measures the resultant state in the computational basis, and computes prediction p defined in (4.27). The circuit is trained by minimizing the square distances (4.24) between the predictions p and the true labels y .

to Alice the output state $\rho_{y_j}^j$, by Bob sent is the state $\rho_{y_j}^j \otimes \rho^j$. That is, alongside with the output state, sent is also a copy of the original, input state. This modified version of the classifier is depicted in Figure 4-12. The expressions (4.22) and (4.23) should be rewritten as, respectively,

$$\rho(\alpha_y, \boldsymbol{\theta}) = U(\boldsymbol{\theta})(\rho_y \otimes \rho)U^\dagger(\boldsymbol{\theta}), \quad (4.26)$$

$$p(\boldsymbol{\theta}) = \frac{1}{2}(1 + \text{Tr}[\rho(\alpha_y, \boldsymbol{\theta})(\sigma_z \otimes \sigma_z)]). \quad (4.27)$$

The training set should be also changed accordingly to $\{\rho_{y_j}^j \otimes \rho^j, y_j\}_{j=1}^{N_{\text{train}}}$. In our numerical experiments with this classifier, we represented the two-qubit unitary $U(\boldsymbol{\theta})$ by the following ansätze:

$$U_1(\boldsymbol{\theta}) = \begin{array}{c} \text{---} R_x(\theta_1) \text{---} R_z(\theta_2) \text{---} R_x(\theta_3) \text{---} \bullet \text{---} \\ \text{---} R_x(\theta_4) \text{---} R_z(\theta_5) \text{---} R_x(\theta_6) \text{---} R_y(\theta_7) \text{---} \end{array} \quad (4.28)$$

$$U_2(\boldsymbol{\theta}) = \begin{array}{c} \text{---} R_x(\theta_1) \text{---} R_z(\theta_2) \text{---} \\ \text{---} R_x(\theta_4) \text{---} R_z(\theta_5) \text{---} \end{array} \quad (4.29)$$

The first ansatz $U_1(\boldsymbol{\theta})$, up to the two-qubit gate, is essentially a single-layered hardware-efficient ansatz shown in Figure 2-1. The second ansatz $U_2(\boldsymbol{\theta})$ is a truncated version of $U_1(\boldsymbol{\theta})$, containing no entangling gates.

The accuracy obtained after training and testing the classifiers based on the

ansätze (4.28) and (4.29) are shown in Figure 4-13. As for the previous case, the sizes of the training and testing sets were $N_{\text{train}} = N_{\text{test}} = 1000$. Among the two classifiers, based on the circuits U_1 and U_2 , the higher accuracy is achieved with the second one, although this circuit is less expressive and has no two-qubit gates. As one can notice, when $\alpha_0 \lesssim 0.75 \lesssim \alpha_1$, this U_2 -based classifier distinguishes the channels perfectly. Moreover, the circuit U_1 shows the highest classification accuracy for the depolarization factors $\alpha = 0.7$ or 0.8 , i.e. near 0.75 . Recall that we observed the same phenomenon for the classifier based on the single-qubit unitary (4.25). Interestingly, it correlates with the fact that the point $\alpha = 3/4 - \epsilon/2$ is the extremum of the function

$$\kappa(\rho_\alpha, \rho_{\alpha+\epsilon}) = \text{Tr}(\rho_\alpha \rho_{\alpha+\epsilon}),$$

for $0 \leq (\alpha + \epsilon) \leq 1$ and any $\rho \neq \mathbb{1}/2$.

Additionally, the point $\alpha = 0.75$ is special for the depolarizing channel, as for this value of the parameter α the output is always the maximally mixed state, $\Phi(\alpha)[\rho] = \mathbb{1}/2$, regardless the input ρ . So, the obtained results suggest that when the classifier is trained only on output states $\Phi(\alpha)[\rho]$, the best performance is achieved when the classifier is tasked to tell the maximally mixed state from any other output state (i.e. when one of the parameters $\alpha_y = 0.75$).

In the meantime, when the classifier is trained on the pairs $\Phi(\alpha)[\rho] \otimes \rho$, i.e. when it is also given the input states, it works best when the output states are from the different sides with respect to $\alpha = 0.75$. This can be explained as follows. For an arbitrary qubit state ρ given by (1.10),

$$\rho = \frac{1}{2} \left[\mathbb{1} + (a_x \sigma_x + a_y \sigma_y + a_z \sigma_z) \right],$$

one finds that

$$\Phi(\alpha)[\rho] = \frac{1}{2} \left[\mathbb{1} + \left(1 - \frac{4}{3} \alpha \right) (a_x \sigma_x + a_y \sigma_y + a_z \sigma_z) \right].$$

Therefore, if one represents the state of a qubit by the vector $\vec{a} = (a_x, a_y, a_z) \in \mathbb{R}^3$,

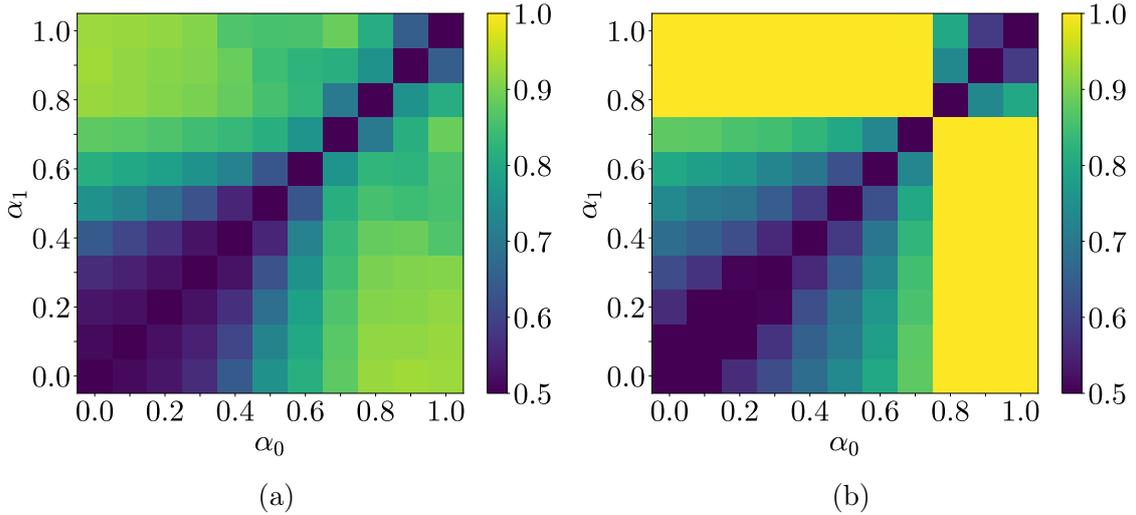


Figure 4-13: Accuracy of the classification of the depolarizing channels with α_0 and α_1 . Panels (a) and (b) show the accuracy of the classifiers built on the circuits U_1 and U_2 defined in (4.28) and (4.29), respectively. The classifiers are trained and tested on the sets of the sizes $N_{\text{train}} = N_{\text{test}} = 1000$. The figures are taken from [4].

which is called the Bloch vector, one sees that under the action of the depolarizing channel $\Phi(\alpha)$, when $0 \leq \alpha < 3/4$ or $3/4 < \alpha \leq 1$, the vector \vec{a} only changes its length. At $\alpha = 3/4$, the length of \vec{a} is zero, which corresponds to the maximally mixed state $\mathbb{1}/2$. Therefore, if one takes $\alpha_0 < 3/4$ and $\alpha_1 > 3/4$, then the corresponding Bloch vectors of $\Phi(\alpha_0)[\rho]$ and $\Phi(\alpha_1)[\rho]$ will be collinear, but pointing to the opposite directions. So, it seems that the classifier trained on the pairs $\Phi(\alpha)[\rho] \otimes \rho$ learns to detect this feature.

So far we have considered the channel classification only for single-qubit depolarizing channels. Let us now test the variational classification approach on the generalized depolarizing channel for n qubits:

$$\Phi(\alpha)[\rho] = (1 - \alpha)\rho + \frac{\alpha}{d^2 - 1} \sum_{j=1}^{d^2-1} S_j \rho S_j, \quad (4.30)$$

where $d = 2^n$ and $S_j \in \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n} / \mathbb{1}^{\otimes n}$ is a Pauli string. In Figure 4-14 we plot the accuracy of the classifier trained to distinguish three-qubit channels (4.30) with $\alpha_0 = 0$ and $\alpha_1 > \alpha_0$. We considered the classifiers with one and two layers of the hardware-efficient ansatz, and both of them were trained on the sets of 1000, 2000 and 3000 entries. As can be seen in Figure 4-14, the classification accuracy

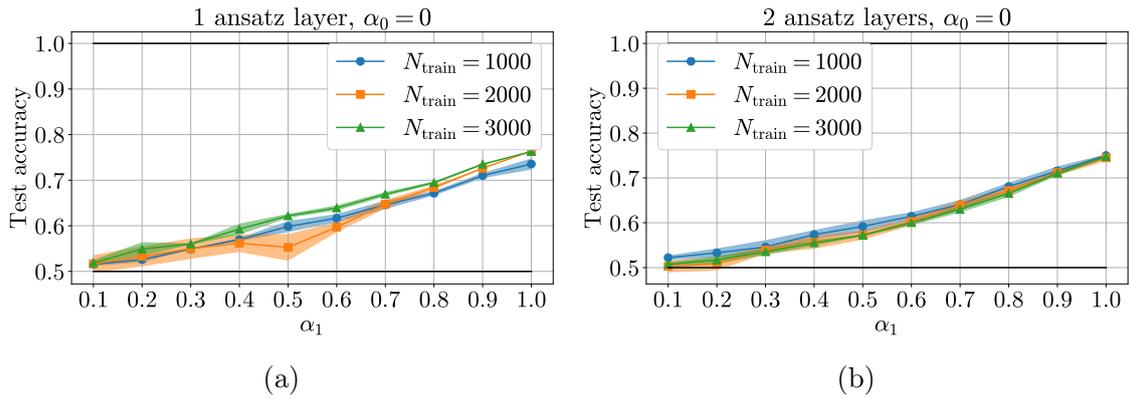


Figure 4-14: Accuracy of the classification of the three-qubit depolarizing channels with $\alpha_0 = 0$ and $\alpha_1 > \alpha_0$. Panels (a) and (b) show the results for the classifiers consisting of one and two layers of the hardware-efficient ansatz, respectively. The classifiers were trained on the sets of the size $N_{\text{train}} = 1000, 2000$ and 3000 , and tested on a set of the size $N_{\text{test}} = 1000$. The shaded areas indicate the standard deviation obtained over 10 initial circuit parameters.

for three-qubit depolarizing channels is comparable to that obtained for the single-qubit channels, see Figure 4-10. Additionally, we do not achieve much improvement by increasing the size of the training set or the expressive power of the variational circuit.

Another test we conducted is the classification of random channels. In Figure 4-15, we show the accuracy of the classifier trained to distinguish between two four-qubit channels obtained by generating random 12-qubit unitary matrices and applying the Stinespring channel representation (4.12). The variational circuit of the classifier consisted of one layer of the hardware-efficient ansatz, and it was trained on only $N_{\text{train}} = 100$ data points. As can be seen, the classification in this case is almost perfect.

So, as we observed, to classify many-qubit depolarizing or random channels, we do not seem to need larger training sets, nor do we need to increase the expressive power of our variational circuits. This, however, should be confirmed for other many-qubit channels, which possibly model some real physical processes.

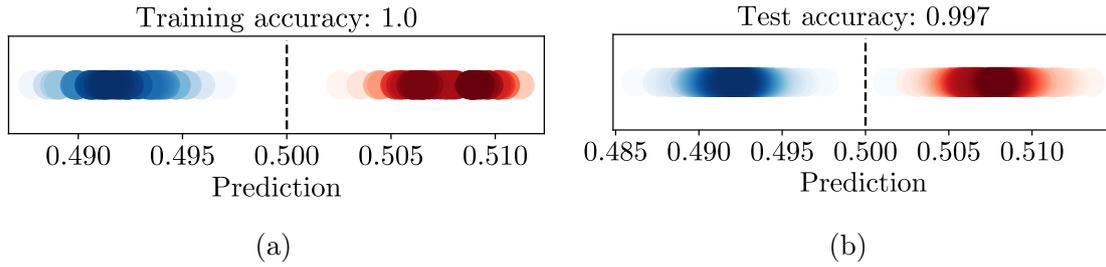


Figure 4-15: Training (a) and test (b) accuracy of the classification of random four-qubit channels. The blue and the red points indicate the label 0 and label 1, respectively. The channels were obtained via the Stinespring channel representation (4.12) by generating random 12-qubit unitary matrices. The classifier was trained on the sets of the size $N_{\text{train}} = 100$, and tested on a set of the size $N_{\text{test}} = 1000$.

4.2.2 Remarks on quantum machine learning

Let us discuss the aspects of machine learning in the context of quantum computing. In [121], the authors distinguish four approaches of machine learning, depending on the nature of the data in question and the computational device used to process it. Namely, there are four classes of machine learning: classical-classical (CC), classical-quantum (CQ), quantum-classical (QC) and quantum-quantum (QQ). The first approach, CC, is the ordinary machine learning for classical data, $\mathbf{x} \in \mathbb{R}^d$, performed using classical computers. In CQ machine learning, one works with classical data encoded into pure quantum states, $\mathbf{x} \mapsto |\mathbf{x}\rangle \in \mathbb{C}^d$, and processed with a quantum computer. QC methods, in their turn, allow to perform machine learning for quantum data using classical algorithms. Finally, QQ approach utilizes quantum computers for working with quantum data.

The problem of quantum channel classification considered in Section 4.2 belongs to the class QQ. Indeed, the data in this problem is represented by quantum states acted by labeled channels, and the data points are processed by applying a variational quantum circuit and performing measurements. A similar problem has been solved in the QC setting in [130]. There, the authors considered the task of multi-classification of quantum channels using classical neural networks. Although one can achieve high classification accuracy with this QC approach, there is an obstacle: To process a data point represented by a density operator $\rho \in \mathcal{D}(\mathbb{C}^d)$ using a classical computer, one needs to obtain and store some representation of ρ . For instance, to

completely describe a state of n qubits, one would need to store $(4^n - 1)$ real numbers, see (1.12). Obtaining these numbers is another task known as quantum state tomography, which can also be enhanced with classical neural networks [131, 132]. Although in many practical cases a quantum state of interest does not have an exponential number of parameters to be estimated [133], the full state tomography is, in general, expensive in terms of measurements.

A similar issue arises in the case of the CQ approach. The necessity of encoding classical data points into the quantum ones can be viewed as a bottleneck of quantum machine learning for classical data. Consider one of the simplest encodings for a data point $\mathbf{x} \in \mathbb{R}^d$, the rotation encoding:

$$\mathbf{x} = \{x_j\}_{j=1}^d \quad \mapsto \quad |\mathbf{x}\rangle = \bigotimes_{j=1}^d \left(\cos(x_j) |0\rangle + \sin(x_j) |1\rangle \right),$$

which can be obtained by acting with the y -rotation gates $R_y(x_j) = \exp(-ix_j\sigma_y)$ on the individual qubits of the state $|0\rangle^{\otimes d}$. Another encoding is the amplitude encoding implemented as

$$\mathbf{x} = \{x_j\}_{j=1}^d \quad \mapsto \quad |\mathbf{x}\rangle = \sum_{j=1}^{\lceil \log_2 d \rceil} x_j |j\rangle,$$

where the vectors $|j\rangle$ form the computational basis in $\mathbb{C}_2^{\otimes \lceil \log_2 d \rceil}$. Compared to the rotation encoding which requires $O(d)$ qubits, the amplitude encoding is advantageous needing only $O(\log_2 d)$ qubits. However, the circuit U which prepares the state $|\mathbf{x}\rangle = U |0\rangle^{\otimes d}$ may consist of an exponential number of gates [134]. More information on data encodings can be found in Chapter 5 in [135].

Having encoded classical data points $\{\mathbf{x}_k\}_{k=1}^m$ into pure quantum states $\{\rho_k = |\mathbf{x}_k\rangle\langle\mathbf{x}_k|\}_{k=1}^m$, one can perform machine learning by training variational circuits as done in Section 4.2. Another option is computing the values of the function $\kappa(\rho_k, \rho_l) = \text{Tr } \rho_k \rho_l = |\langle\mathbf{x}_k|\mathbf{x}_l\rangle|^2$ called the kernel and applying the so-called kernel-based machine-learning methods [136]. On a quantum computer, one can calculate the kernel directly via the swap test routine described in Appendix B. In [134], it is claimed that if we are given m classical data points, the kernel values for them can be estimated on a quantum computer in time linear in m , but it is assumed that we

can efficiently encode our data points into quantum states.

To conclude this section, we argue that machine learning tasks could be solved more efficiently if the nature of the data matches that of the computational device used. That is, quantum (classical) data is processed best by a quantum (classical) computer.

Conclusion

In this thesis, we considered variational quantum algorithms, their formulations and applications. First, we made a brief introduction into the field of quantum computing and its history of development. We noted that one of subareas of quantum computing, variational quantum algorithms, is nowadays fast-developing due to the possibility of execution on contemporary quantum computers.

Then we considered the variational quantum eigensolver (VQE) algorithm and described several cases of its application. In the two case studies we performed, it was found that the performance of VQE is dependent not only on the structure of the variational state, but also on the properties of the target Hamiltonian and especially its ground state. Namely, an important role may be played by e.g. the entanglement properties of the ground state.

Next, we reviewed several variational algorithms (VQA) and formulated them in the language of tensor networks. As the tensor diagrams notation is widely used in theoretical physics, computer science, and machine learning, the obtained formulations of VQA may make them more understandable for researches from the mentioned communities. The tensor network language helped us to show that the considered variational algorithms can be reduced to VQE executed for specific Hamiltonians and variational states. This is an interesting fact suggesting that many variational quantum algorithms can be formulated in such a way that they are based on computing the expected value of an observable with respect to a pure state.

After that, we put the optimization problem of quantum channel discrimination into the framework of variational quantum computing. This suggests that many other optimization problems, which include quantum entities, can be solved this

way, and in this thesis we proposed a recipe for it. For the channel discrimination problem, we found that the performance of the variational approach is dependent not only on the diamond distance between channels — which gives a fundamental bound on the channel distinguishability — but it may depend also on the trace distance between the the channels output states, and also on the trace of their product. Interestingly, the latter serves as a kernel in quantum machine learning.

Finally, we showed that variational circuits are also applicable for binary classification of quantum channels. A big advantage of such a classifier is that it works directly with given data sets of labeled quantum states. In contrast, if one applies classical machine learning for quantum data, it would require explicit matrix representations of density operators, which can be obtained only via expensive state tomography. Additionally, the results of our numerical experiments suggest that the variational circuits required for building a classifier can be quite shallow, and in some cases they do not even need any many-qubit gates.

Results

1. At the certain point of the bare mass m in the 2-qubit Schwinger Hamiltonian, the rapid change of the order parameter can be witnessed [1]. At this point, the ground state is the singlet state, which is maximally entangled and is invariant under local transformations. Due to this, there is a flat valley of the optimization landscape, which makes it hard for VQE to find the true ground state in the vicinity of this point. Additionally, in the presence of controlled dephasing noise, we can detect this critical point of the order parameter.
2. For a spin Hamiltonian with Dzyaloshinskii-Moriya interactions, even when VQE finds the ground state energy with more than 99% precision, the overlap between the found and the true ground state can be small [2]. We attribute this lack of the state fidelity to the entanglement properties of the ground state of the considered Hamiltonian, which cannot be reproduced with a low-depth hardware-efficient ansatz. Although the ground state found by VQE is of low fidelity, one is still able to extract from it some physical features of interest.

For instance, we retrieved the spiral structure of the spins in the chain.

3. Formulated in the tensor networks notation, some variational quantum algorithms can be reduced to the VQE algorithm executed for a specific Hamiltonian and variational state. We obtained explicit reductions for the following algorithms:
 - variational quantum linear solver,
 - variational quantum state preparation algorithms,
 - variational quantum state diagonalization algorithm,
 - variational eigenstate verification algorithm [3].

4. The problem of binary quantum channel discrimination — which is a problem of distinguishing between two given channels, and which is itself an optimization problem — can be put into the framework of variational quantum computing [4]. In this problem, the possibility of applying a given channel several times gives rise to two discrimination strategies with parallel and sequential applications of a channel. We tested both strategies in discriminating depolarizing channels and entanglement breaking channels. Compared to the parallel strategy, the sequential strategy allows to achieve higher channel discrimination accuracy with shorter quantum circuits and lower variance of the results. We also observed that the performance of the variational channel discrimination may depend on the trace distances between the channel output states, and also on the trace of their product.

5. A variational quantum circuit can serve as a binary classifier for distinguishing between quantum channels [4]. We showed that such a classifier can be trained with a data set of relatively few entries. For depolarizing channels, the best classification accuracy is observed when alongside with the channel output states, the classifier is also provided with the input states. For single-qubit depolarizing channels, the classifying circuit does not need to contain entangling gates. We also showed that for classifying many-qubit depolarizing channels, we do not need larger training sets and more expressive variational circuits, as

one might expect. Additionally, the variational classifier distinguishes random many qubit channels with high efficiency.

Bibliography

- [1] O.V. Borzenkova, G.I. Struchalin, A.S. Kardashin, V.V. Krasnikov, N.N. Skryabin, S.S. Straupe, S.P. Kulik, and J.D. Biamonte. Variational simulation of Schwinger’s Hamiltonian with polarization qubits. *Applied Physics Letters*, 118(14):144002, 2021.
- [2] A. Kardashin, A. Pervishko, J. Biamonte, and D. Yudin. Numerical hardware-efficient variational quantum simulation of a soliton solution. *Physical Review A*, 104(2):L020402, 2021.
- [3] A. Kardashin, A. Uvarov, D. Yudin, and J. Biamonte. Certified variational quantum algorithms for eigenstate preparation. *Physical Review A*, 102(5):052610, 2020.
- [4] A.S. Kardashin, A.V. Vlasova, A.A. Pervishko, D. Yudin, and J.D. Biamonte. Quantum-machine-learning channel discrimination. *Physical Review A*, 106:032409, Sep 2022.
- [5] P. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.
- [6] R. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.
- [7] Carl Pomerance. A tale of two sieves. In *Notices of the American Mathematical Society*. Citeseer, 1996.
- [8] L. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, 1996.
- [9] C. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. *arXiv preprint arXiv:2003.06557*, 2020.
- [10] J. Park. The concept of transition in quantum mechanics. *Foundations of physics*, 1(1):23–33, 1970.
- [11] A. Ekert. Quantum cryptography and bell’s theorem. In *Quantum Measurements in Optics*, pages 413–418. Springer, 1992.

- [12] J. Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.
- [13] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.
- [14] C. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Physical review letters*, 70(13):1895, 1993.
- [15] C. Bennett and S. Wiesner. Communication via one-and two-particle operators on einstein-podolsky-rosen states. *Physical review letters*, 69(20):2881, 1992.
- [16] A Kitaev. Quantum measurements and the abelian stabilizer problem. *arXiv preprint quant-ph/9511026*, 1995.
- [17] A. Harrow, A. Hassidim, and S. Lloyd. Quantum algorithm for linear systems of equations. *Physical review letters*, 103(15):150502, 2009.
- [18] A. Gilyén, Y. Su, G. Low, and N. Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 193–204, 2019.
- [19] P. Reberntrost, A. Steffens, I. Marvian, and S. Lloyd. Quantum singular-value decomposition of nonsparse low-rank matrices. *Physical review A*, 97(1):012327, 2018.
- [20] F. Brandao and K. Svore. Quantum speed-ups for solving semidefinite programs. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 415–426. IEEE, 2017.
- [21] J. Van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf. Quantum sdp-solvers: Better upper and lower bounds. *Quantum*, 4:230, 2020.
- [22] A. Ambainis, K. Balodis, J. Iraids, M. Kokainis, K. Prūsis, and J. Vihrovs. Quantum speedups for exponential-time dynamic programming algorithms. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1783–1793. SIAM, 2019.
- [23] J. Biamonte, P. Wittek, N. Pancotti, P. Reberntrost, N. Wiebe, and S. Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.
- [24] S. Lloyd, M. Mohseni, and P. Reberntrost. Quantum algorithms for supervised and unsupervised machine learning. *arXiv preprint arXiv:1307.0411*, 2013.
- [25] Z. Zhao, A. Pozas-Kerstjens, P. Reberntrost, and P. Wittek. Bayesian deep learning on a quantum computer. *Quantum Machine Intelligence*, 1(1):41–51, 2019.

- [26] D. Berry. High-order quantum algorithm for solving linear differential equations. *Journal of Physics A: Mathematical and Theoretical*, 47(10):105301, 2014.
- [27] A. Childs and J. Liu. Quantum spectral methods for differential equations. *Communications in Mathematical Physics*, 375(2):1427–1457, 2020.
- [28] D. Berry, A. Childs, A. Ostrander, and G. Wang. Quantum algorithm for linear differential equations with exponentially improved dependence on precision. *Communications in Mathematical Physics*, 356(3):1057–1081, 2017.
- [29] A. Montanaro. Quantum algorithms: an overview. *npj Quantum Information*, 2(1):1–8, 2016.
- [30] D. Gottesman. *Stabilizer codes and quantum error correction*. California Institute of Technology, 1997.
- [31] D. A Lidar and T. Brun. *Quantum error correction*. Cambridge university press, 2013.
- [32] K. Bharti, A. Cervera-Lierta, T. Ha Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. Kottmann, T. Menke, et al. Noisy intermediate-scale quantum (nisq) algorithms. *arXiv preprint arXiv:2101.08448*, 2021.
- [33] M. Brooks. Beyond quantum supremacy: the hunt for useful quantum computers. *Nature*, 574(7776):19–22, 2019.
- [34] J. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016.
- [35] J. Li, X. Yang, X. Peng, and C. Sun. Hybrid quantum-classical approach to quantum optimal control. *Physical Review Letters*, 118(15):150503, 2017.
- [36] R. Santagati et al. Witnessing eigenstates for quantum simulation of hamiltonian spectra. *Science Advances*, 4(1):eaap9646, 2018.
- [37] A. Peruzzo et al. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5:4213, 2014.
- [38] Jacob Biamonte. Universal variational quantum computation. *Physical Review A*, 103(3):L030401, 2021.
- [39] M. Cerezo et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.
- [40] R. LaRose, A. Tikku, É. O’Neel-Judy, L. Cincio, and P. Coles. Variational quantum state diagonalization. *npj Quantum Information*, 5(1):1–10, 2019.
- [41] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, and D. Jaksch. Variational quantum algorithms for nonlinear problems. *Physical Review A*, 101(1):010301, 2020.

- [42] E. Gross, L. N Oliveira, and W. Kohn. Rayleigh-ritz variational principle for ensembles of fractionally occupied states. *Physical Review A*, 37(8):2805, 1988.
- [43] J. Biamonte. Lectures on quantum tensor networks. *arXiv preprint arXiv:1912.10049*, 2019.
- [44] J. Biamonte and V. Bergholm. Tensor networks in a nutshell. *arXiv preprint arXiv:1708.00006*, 2017.
- [45] R. Penrose. Applications of negative dimensional tensors. *Combinatorial mathematics and its applications*, 1:221–244, 1971.
- [46] A. Cichocki. Era of big data processing: A new approach via tensor networks and tensor decompositions. *arXiv preprint arXiv:1403.2048*, 2014.
- [47] A. Cichocki, A. Phan, Q. Zhao, N. Lee, I. Oseledets, M. Sugiyama, D. Mandic, et al. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations and Trends® in Machine Learning*, 9(4-5):249–429, 2016.
- [48] D. Liu, S. Ran, P. Wittek, C. Peng, R. García, G. Su, and M. Lewenstein. Machine learning by unitary tensor network of hierarchical tree structure. *New Journal of Physics*, 21(7):073059, 2019.
- [49] R. Sengupta, S. Adhikary, I. Oseledets, and J. Biamonte. Tensor networks in machine learning. *European Mathematical Society Magazine*, 101:9, 2022.
- [50] A. Cichocki, A. Phan, Q. Zhao, N. Lee, I. Oseledets, M. Sugiyama, D. Mandic, et al. Tensor networks for dimensionality reduction and large-scale optimization: Part 2 applications and future perspectives. *Foundations and Trends® in Machine Learning*, 9(6):431–673, 2017.
- [51] A. Holevo. *Quantum systems, channels, information*. de Gruyter, 2019.
- [52] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Physical Review Letters*, 91(14):147902, 2003.
- [53] M. Plenio and S. Virmani. An introduction to entanglement theory. In *Quantum information and coherence*, pages 173–209. Springer, 2014.
- [54] Masaya Kohda, Ryosuke Imai, Keita Kanno, Kosuke Mitarai, Wataru Mizukami, and Yuya O Nakagawa. Quantum expectation-value estimation by computational basis sampling. *Physical Review Research*, 4(3):033173, 2022.
- [55] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.
- [56] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of physics*, 326(1):96–192, 2011.
- [57] Jens Eisert. Entanglement and tensor network states. *arXiv preprint arXiv:1308.3318*, 2013.

- [58] Frank Pollmann. Efficient numerical simulations using matrix-product states, 2016.
- [59] Jacob Biamonte, Ville Bergholm, and Marco Lanzagorta. Tensor network methods for invariant theory. *Journal of Physics A: Mathematical and Theoretical*, 46(47):475301, 2013.
- [60] Matthew B Hastings. An area law for one-dimensional quantum systems. *Journal of statistical mechanics: theory and experiment*, 2007(08):P08024, 2007.
- [61] Jens Eisert, Marcus Cramer, and Martin B Plenio. Colloquium: Area laws for the entanglement entropy. *Reviews of modern physics*, 82(1):277, 2010.
- [62] A. Sørensen and K. Mølmer. Entanglement and quantum computation with ions in thermal motion. *Physical Review A*, 62(2):022311, 2000.
- [63] H. Häffner, C. Roos, and R. Blatt. Quantum computing with trapped ions. *Physics Reports*, 469(4):155–203, 2008.
- [64] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.
- [65] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B.P. Lanyon, P. Love, R. Babbush, et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Physical Review X*, 8(3):031022, 2018.
- [66] Y. Li, J. Hu, X.M. Zhang, Zhigang Song, and M.H. Yung. Variational quantum simulation for quantum chemistry. *Advanced Theory and Simulations*, 2(4):1800182, 2019.
- [67] J. Romero, R. Babbush, J.R. McClean, C. Hempel, P.J. Love, and A. Aspuru-Guzik. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology*, 4(1):014008, 2018.
- [68] D. Wecker, M.B. Hastings, N. Wiebe, B.K. Clark, C. Nayak, and M. Troyer. Solving strongly correlated electron models on a quantum computer. *Physical Review A*, 92(6):062318, 2015.
- [69] A.A. Zhukov, S.V. Remizov, W.V. Pogosov, and Yu.E. Lozovik. Algorithmic simulation of far-from-equilibrium dynamics using quantum computer. *Quantum Information Processing*, 17(9):1–26, 2018.
- [70] B. Bauer, S. Bravyi, M. Motta, and G.K.L. Chan. Quantum algorithms for quantum chemistry and quantum materials science. *Chemical Reviews*, 120(22):12685–12717, 2020.
- [71] A.V. Uvarov and J.D. Biamonte. On barren plateaus and cost function locality in variational quantum algorithms. *Journal of Physics A: Mathematical and Theoretical*, 54(24):245301, 2021.

- [72] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature communications*, 12(1):1–12, 2021.
- [73] E. Campos, A. Nasrallah, and J. Biamonte. Abrupt transitions in variational quantum circuit training. *Physical Review A*, 103(3):032607, 2021.
- [74] D. Berry, G. Ahokas, R. Cleve, and B. Sanders. Efficient quantum algorithms for simulating sparse hamiltonians. *Communications in Mathematical Physics*, 270(2):359–371, 2007.
- [75] H. Weber and G. Arfken. *Essential mathematical methods for physicists, ISE*. Elsevier, 2003.
- [76] Tzu-Ching Yen, Vladyslav Verteletskyi, and Artur F Izmaylov. Measuring all compatible operators in one series of single-qubit measurements using unitary transformations. *Journal of chemical theory and computation*, 16(4):2400–2409, 2020.
- [77] C. Hamer, Z. Weihong, and J. Oitmaa. Series expansions for the massive schwinger model in hamiltonian lattice theory. *Physical Review D*, 56(1):55, 1997.
- [78] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. Chow, and J. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.
- [79] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. Joshi, P. Jurcevic, C. A Muschik, P. Silvi, R. Blatt, C. Roos, et al. Self-verifying variational quantum simulation of lattice models. *Nature*, 569(7756):355–360, 2019.
- [80] T.M.R. Byrnes, P. Sriganesh, R.J. Bursill, and C.J. Hamer. Density matrix renormalization group approach to the massive schwinger model. *Physical Review D*, 66(1):013002, 2002.
- [81] Y. Togawa, T. Koyama, K. Takayanagi, S. Mori, Y. Kousaka, J. Akimitsu, S. Nishihara, K. Inoue, A.S. Ovchinnikov, and J. Kishine. Chiral magnetic soliton lattice on a chiral helimagnet. *Physical Review Letters*, 108:107202, Mar 2012.
- [82] J. Kishine and A.S. Ovchinnikov. Chapter One - Theory of monoaxial chiral helimagnet. volume 66 of *Solid State Physics*, page 1. Academic Press, 2015.
- [83] K. Koumpouras, A. Bergman, O. Eriksson, and D. Yudin. A spin dynamics approach to solitons. *Scientific Reports*, 6:25685, May 2016.
- [84] T. Moriya and T. Miyadai. Evidence for the helical spin structure due to antisymmetric exchange interaction in $\text{Cr}_{1/3}\text{NbS}_2$. *Solid State Communications*, 42(3):209, 1982.

- [85] T. Miyadai, K. Kikuchi, H. Kondo, S. Sakka, M. Arai, and Y. Ishikawa. Magnetic properties of $\text{Cr}_{1/3}\text{NbS}_2$. *Journal of the Physical Society of Japan*, 52(4):1394, 1983.
- [86] V. Dyadkin, F. Mushenok, A. Bosak, D. Menzel, S. Grigoriev, P. Pattison, and D. Chernyshov. Structural disorder versus chiral magnetism in $\text{cr}_{1/3}\text{nbs}_2$. *Physical Review B*, 91:184205, May 2015.
- [87] H. Abraham et al. Qiskit: An Open-source Framework for Quantum Computing, 2019.
- [88] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, New York, 2006.
- [89] C. Bravo-Prieto, J. Lumbreras-Zarapico, L. Tagliacozzo, and J. I. Latorre. Scaling of variational quantum circuit depth for condensed matter systems. *Quantum*, 4:272, may 2020.
- [90] Harry Levine, Alexander Keesling, Giulia Semeghini, Ahmed Omran, Tout T Wang, Sepehr Ebadi, Hannes Bernien, Markus Greiner, Vladan Vuletić, Hannes Pichler, et al. Parallel implementation of high-fidelity multiqubit gates with neutral atoms. *Physical review letters*, 123(17):170503, 2019.
- [91] Craig R Clark, Holly N Tinkey, Brian C Sawyer, Adam M Meier, Karl A Burkhardt, Christopher M Seck, Christopher M Shappert, Nicholas D Guise, Curtis E Volin, Spencer D Fallek, et al. High-fidelity bell-state preparation with ca+ 40 optical qubits. *Physical Review Letters*, 127(13):130505, 2021.
- [92] Max Wilson, Rachel Stromswold, Filip Wudarski, Stuart Hadfield, Norm M Tubman, and Eleanor G Rieffel. Optimizing quantum heuristics with meta-learning. *Quantum Machine Intelligence*, 3:1–14, 2021.
- [93] X.Q. Zhou, T.C. Ralph, P. Kalasuwan, M. Zhang, A. Peruzzo, B.P. Lanyon, and J.L. O’Brien. Adding control to arbitrary unknown quantum operations. *Nature communications*, 2(1):413, 2011.
- [94] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A.T. Sornborger, and P.J. Coles. Quantum-assisted quantum compiling. *Quantum*, 3:140, 2019.
- [95] X. Xu, J. Sun, S. Endo, Y. Li, S. Benjamin, and X. Yuan. Variational algorithms for linear algebra. *Science Bulletin*, 66(21):2181–2188, 2021.
- [96] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J Coles. Variational quantum linear solver. *arXiv preprint arXiv:1909.05820*, 2019.
- [97] Y. Subaşı, R. Somma, and D. Orsucci. Quantum algorithms for systems of linear equations inspired by adiabatic quantum computing. *Physical Review Letters*, 122(6):060504, 2019.
- [98] A. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191, 1997.

- [99] G. Wang and M. Ying. Unambiguous discrimination among quantum operations. *Physical Review A*, 73(4):042301, 2006.
- [100] R. Duan, Y. Feng, and M. Ying. Perfect distinguishability of quantum operations. *Physical Review Letters*, 103(21):210501, 2009.
- [101] A. Holevo. *Quantum Systems, Channels, Information: A Mathematical Introduction*. de Gruyter & Co, Berlin, Boston, 2012.
- [102] S. Lloyd. Enhanced sensitivity of photodetection via quantum illumination. *Science*, 321(5895):1463–1465, 2008.
- [103] S. Barzanjeh, S. Guha, C. Weedbrook, D. Vitali, J. H. Shapiro, and S. Pirandola. Microwave quantum illumination. *Physical Review Letters*, 114(8):080503, 2015.
- [104] G. Ortolano, E. Losero, S. Pirandola, M. Genovese, and I. Ruo-Berchera. Experimental quantum reading with photon counting. *Science Advances*, 7(4):eabc7796, 2021.
- [105] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd. Advances in photonic quantum sensing. *Nature Photonics*, 12(12):724–733, 2018.
- [106] G. Benenti and G. Strini. Computing the distance between quantum channels: usefulness of the fano representation. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 43(21):215508, 2010.
- [107] J. ur Rehman, A. Farooq, Y. Jeong, and H. Shin. Quantum channel discrimination without entanglement. *Quantum Information Processing*, 17(10):1–16, 2018.
- [108] D. Puzzuoli and J. Watrous. Ancilla dimension in quantum channel discrimination. In *Annales Henri Poincaré*, volume 18, pages 1153–1184. Springer, 2017.
- [109] W. Matthews, M. Piani, and J. Watrous. Entanglement in channel discrimination with restricted measurements. *Physical Review A*, 82(3):032302, 2010.
- [110] G. Chiribella, G. M. D’Ariano, and P. Perinotti. Quantum circuit architecture. *Physical Review Letters*, 101(6):060401, 2008.
- [111] F. Salek, M. Hayashi, and A. Winter. When are adaptive strategies in asymptotic quantum channel discrimination useful? *arXiv preprint arXiv:2011.06569*, 2020.
- [112] V. Katariya and M. Wilde. Evaluating the advantage of adaptive strategies for quantum channel distinguishability. *Physical Review A*, 104(5):052406, 2021.
- [113] J. Bavaresco, M. Muraio, and M. Quintino. Strict hierarchy between parallel, sequential, and indefinite-causal-order strategies for channel discrimination. *Physical Review Letters*, 127(20):200504, 2021.

- [114] P. Qiu, X. Chen, and Y. Shi. Solving quantum channel discrimination problem with quantum networks and quantum neural networks. *IEEE Access*, 7:50214–50222, 2019.
- [115] R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208, 1995.
- [116] A. Harrow, A. Hassidim, D. Leung, and J. Watrous. Adaptive versus non-adaptive strategies for quantum channel discrimination. *Physical Review A*, 81(3):032339, 2010.
- [117] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.
- [118] A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018.
- [119] S. Diamond and S. Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.
- [120] Mark M Wilde. *Quantum information theory*. Cambridge University Press, 2013.
- [121] R. Mengoni and A. Di Pierro. Kernel methods in quantum machine learning. *Quantum Machine Intelligence*, 1(3):65–71, 2019.
- [122] Massimiliano F. Sacchi. Optimal discrimination of quantum operations. *Physical Review A*, 71(6):062340, 2005.
- [123] Rochisha Agarwal, Soorya Rethinasamy, Kunal Sharma, and Mark M Wilde. Estimating distinguishability measures on quantum computers. *arXiv preprint arXiv:2108.08406*, 2021.
- [124] Stefano Pirandola, Riccardo Laurenza, Carlo Ottaviani, and Leonardo Banchi. Fundamental limits of repeaterless quantum communications. *Nature communications*, 8(1):15043, 2017.
- [125] S. Pirandola, R. Laurenza, C. Lupo, and J. Pereira. Fundamental limits to quantum channel discrimination. *npj Quantum Information*, 5(1), 2019.
- [126] Jason L Pereira and Stefano Pirandola. Bounds on amplitude-damping-channel discrimination. *Physical Review A*, 103(2):022610, 2021.
- [127] A. Shirinyan, V. Kozin, J. Hellsvik, M. Pereiro, O. Eriksson, and D. Yudin. Self-organizing maps as a method for detecting phase transitions and phase identification. *Physical Review B*, 99:041108, Jan 2019.

- [128] A. Uvarov, A. Kardashin, and J. Biamonte. Machine learning phase transitions with a quantum processor. *Physical Review A*, 102(1):012415, 2020.
- [129] A. Berezutskii, M. Beketov, D. Yudin, Z. Zimborás, and J. D. Biamonte. Probing criticality in quantum spin chains with neural networks. *Journal of Physics: Complexity*, 1(3):03LT01, aug 2020.
- [130] A.M. Palmieri, F. Bianchi, M.G.A. Paris, and C. Benedetti. Multiclass classification of dephasing channels. *Physical Review A*, 104(5):052412, 2021.
- [131] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447–450, 2018.
- [132] A. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. Biamonte, and S. Kulik. Experimental neural network enhanced quantum tomography. *npj Quantum Information*, 6:20, 2020.
- [133] M. Cramer, M.B. Plenio, S.T. Flammia, R. Somma, D. Gross, S.D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.K. Liu. Efficient quantum state tomography. *Nature communications*, 1(1):149, 2010.
- [134] M. Schuld. Supervised quantum machine learning models are kernel methods. *arXiv preprint arXiv:2101.11020*, 2021.
- [135] Maria Schuld and Francesco Petruccione. *Supervised learning with quantum computers*, volume 17. Springer, 2018.
- [136] V. Havlíček et al. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.
- [137] H. Kobayashi, K. Matsumoto, and T. Yamakami. Quantum merlin-arthur proof systems: Are multiple merlins more helpful to arthur? In *International Symposium on Algorithms and Computation*, pages 189–198. Springer, 2003.
- [138] M. Schuld, I. Sinayskiy, and F. Petruccione. An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172–185, 2015.
- [139] H. Buhrman, R. Cleve, J. Watrous, and R. De Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16):167902, 2001.
- [140] A. Barenco, A. Berthiaume, D. Deutsch, A. Ekert, R. Jozsa, and C. Macchiavello. Stabilization of quantum computations by symmetrization. *SIAM Journal on Computing*, 26(5):1541–1557, 1997.
- [141] Juan Carlos Garcia-Escartin and Pedro Chamorro-Posada. Swap test and hong-ou-mandel effect are equivalent. *Physical Review A*, 87(5):052330, 2013.
- [142] Sreetama Das, Jingfu Zhang, Stefano Martina, Dieter Suter, and Filippo Caruso. Experimental quantum pattern recognition in ibmq and diamond nvs. *arXiv preprint arXiv:2205.00561*, 2022.

Appendix A

Cost functions of the variational quantum state diagonalization algorithm

In Section 3.5 we claimed that the three cost functions in (3.21)-(3.23) are equivalent,

$$d_{\text{HS}}(\rho_U, D[\rho_U]) = \sum_{a \neq b} |\langle a | \rho_U | b \rangle|^2 = \text{Tr } \rho^2 - \text{Tr } D[\rho_U]^2.$$

Indeed, for the Hilbert-Schmidt distance between the varied state ρ_U and its dephased version $D[\rho_U]$ is

$$\begin{aligned} d_{\text{HS}}(\rho_U, D[\rho_U]) &= \text{Tr} [(\rho - D[\rho])^\dagger (\rho - D[\rho])] \\ &= \text{Tr} [\rho^2 - \rho D[\rho] - D[\rho] \rho + D[\rho]^2] \\ &= \text{Tr } \rho^2 + \text{Tr } D[\rho]^2 - 2 \text{Tr } \rho D[\rho], \end{aligned} \tag{A.1}$$

where we omitted the subscript U for readability. Suppose that $\rho = \sum_i r_i |r_i\rangle\langle r_i|$. The spectral decomposition for the dephased version of this state is

$$\begin{aligned}
 D[\rho] &= \sum_a \langle a|\rho|a\rangle |a\rangle\langle a| \\
 &= \sum_a \langle a| \left(\sum_i r_i |r_i\rangle\langle r_i| \right) |a\rangle |a\rangle\langle a| \\
 &= \sum_a \sum_i r_i |\langle r_i|a\rangle|^2 |a\rangle\langle a|.
 \end{aligned} \tag{A.2}$$

Therefore for the second term in (A.1) we obtain

$$\begin{aligned}
 \text{Tr } D[\rho]^2 &= \text{Tr} \left[\left(\sum_a \sum_i r_i |\langle r_i|a\rangle|^2 |a\rangle\langle a| \right) \left(\sum_b \sum_j r_j |\langle r_j|b\rangle|^2 |b\rangle\langle b| \right) \right] \\
 &= \text{Tr} \left[\sum_{a,b} \sum_{i,j} r_i r_j |\langle r_i|a\rangle|^2 |\langle r_j|b\rangle|^2 \langle a|b\rangle |a\rangle\langle b| \right] \\
 &= \sum_a \sum_{i,j} r_i r_j |\langle r_i|a\rangle|^2 |\langle r_j|a\rangle|^2.
 \end{aligned}$$

The last term in (A.1) reduces to

$$\begin{aligned}
 \text{Tr } \rho D[\rho] &= \text{Tr} \left[\sum_i r_i |r_i\rangle\langle r_i| \sum_a \sum_j r_j |\langle r_j|a\rangle|^2 |a\rangle\langle a| \right] \\
 &= \text{Tr} \left[\sum_a \sum_{i,j} r_i r_j |\langle r_j|a\rangle|^2 |r_i\rangle\langle r_i| |a\rangle\langle a| \right] \\
 &= \sum_a \sum_{i,j} r_i r_j |\langle r_i|a\rangle|^2 |\langle r_j|a\rangle|^2 \\
 &= \text{Tr } D[\rho]^2.
 \end{aligned}$$

Putting this into (A.1) we prove that

$$d_{\text{HS}}(\rho_U, D[\rho_U]) = \text{Tr } \rho^2 - \text{Tr } D[\rho_U]^2,$$

since obviously $\text{Tr } \rho_U^2 = \text{Tr } \rho^2$. Now consider

$$\begin{aligned}
 \sum_{a \neq b} |\langle a | \rho | b \rangle|^2 &= \sum_{a \neq b} \langle a | \rho | b \rangle \langle b | \rho | a \rangle \\
 &= \sum_a \langle a | \rho \left(\sum_{b \neq a} |b\rangle\langle b| \right) \rho | a \rangle \\
 &= \sum_a \langle a | \rho (\mathbb{1} - |a\rangle\langle a|) \rho | a \rangle \\
 &= \sum_a \langle a | \rho^2 | a \rangle - \sum_a |\langle a | \rho | a \rangle|^2 \\
 &= \text{Tr } \rho^2 - \text{Tr } D[\rho_U]^2,
 \end{aligned}$$

where the last term follows directly from tracing the expression (A.2). That is, we have proven the equivalence of the cost functions in (3.21)-(3.23).

Appendix B

Swap test

The swap test routine allows to compare two quantum states. Generally, given the density operators ρ and τ of the same size, the swap test allows to compute $\text{Tr } \rho\tau$ [137]; if the given states are pure, i.e. $\rho = |r\rangle\langle r|$ and $\tau = |t\rangle\langle t|$, then the outcome of the swap test is obviously the overlap $|\langle r|t\rangle|^2$ which is a measure of “closeness” of two states. Moreover, the function $\kappa(\rho_i, \rho_j) = \text{Tr } \rho_i\rho_j$ is actually a kernel which is utilized in quantum machine learning [4, 138]. There are two equivalent swap test routines which we will call the classic and the destructive swap test, and both of them are described in the following sections.

B.1 Classic swap test

The classic swap test proposed in [139, 140] requires an auxiliary qubit and the ability to apply the controlled swap gate CSW . The swap gate SW is a two-qubit operator which —as follows from its name— swaps the states of two qubits,

$$SW(\rho \otimes \tau)SW = \tau \otimes \rho,$$

or, for the case of pure states,

$$SW(|r\rangle \otimes |t\rangle) = |t\rangle \otimes |r\rangle,$$

where the gate SW is defined as

$$\begin{aligned}
 SW &= \frac{1}{2} \sum_{j=0}^3 \sigma_j \otimes \sigma_j \\
 &= \sum_{k,l=0}^1 |k\rangle\langle l| \otimes |l\rangle\langle k| \\
 &= \begin{array}{c} \text{---} \times \text{---} \\ | \\ \text{---} \times \text{---} \end{array} = \begin{array}{c} \text{---} \oplus \text{---} \\ | \\ \text{---} \bullet \text{---} \\ | \\ \text{---} \oplus \text{---} \end{array}
 \end{aligned}$$

where σ_j is a Pauli operator, and the vectors $|k\rangle$ and $|l\rangle$ are from the computational basis. The controlled version of this gate, CSW , is constructed as shown in (1.31).

Consider the circuit

$$\begin{array}{c}
 |0\rangle\langle 0| \text{---} \boxed{H} \text{---} \bullet \text{---} \boxed{H} \text{---} \boxed{\text{---}} \\
 | \\
 \rho \text{---} \times \text{---} \\
 | \\
 \tau \text{---} \times \text{---}
 \end{array} \quad (\text{B.1})$$

where ρ and τ are single-qubit states. In this circuit, the initial state is $\psi_1 = |0\rangle\langle 0| \otimes \rho \otimes \tau$. The first qubit of this state is acted by the Hadamard gate, and the resultant state is

$$\begin{aligned}
 \psi_2 &= (H \otimes \mathbb{1} \otimes \mathbb{1}) \psi_1 (H \otimes \mathbb{1} \otimes \mathbb{1}) \\
 &= |+\rangle\langle +| \otimes \rho \otimes \tau \\
 &= \frac{1}{2} \left(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1| \right) \otimes \rho \otimes \tau.
 \end{aligned}$$

This state is then acted by the controlled swap gate, $\psi_3 = CSW \psi_2 CSW$, which swaps the states of the second and the third qubit conditioned on the state of the first one,

$$\begin{aligned}
 \psi_3 &= \frac{1}{2} \left[|0\rangle\langle 0| \otimes (\rho \otimes \tau) + |0\rangle\langle 1| \otimes (\rho \otimes \tau) SW \right. \\
 &\quad \left. + |1\rangle\langle 0| \otimes SW(\rho \otimes \tau) + |1\rangle\langle 1| \otimes SW(\rho \otimes \tau) SW \right].
 \end{aligned}$$

Finally, the first qubit is again acted by the Hadamard gate, $\psi_4 = (H \otimes \mathbb{1} \otimes \mathbb{1}) \psi_3 (H \otimes$

$\mathbb{1} \otimes \mathbb{1}$), which yields the state

$$\begin{aligned} \psi_4 = & \frac{1}{4} \left[(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1|) \otimes (\rho \otimes \tau) \right. \\ & + (|0\rangle\langle 0| - |0\rangle\langle 1| + |1\rangle\langle 0| - |1\rangle\langle 1|) \otimes (\rho \otimes \tau) SW \\ & + (|0\rangle\langle 0| + |0\rangle\langle 1| - |1\rangle\langle 0| - |1\rangle\langle 1|) \otimes SW(\rho \otimes \tau) \\ & \left. + (|0\rangle\langle 0| - |0\rangle\langle 1| - |1\rangle\langle 0| + |1\rangle\langle 1|) \otimes (\tau \otimes \rho) \right] \end{aligned}$$

where we took into account that $SW(\rho \otimes \tau)SW = \tau \otimes \rho$. The probability of obtaining the outcome $x \in \{0, 1\}$ after measuring the observable $Z \otimes \mathbb{1} \otimes \mathbb{1}$ is

$$\begin{aligned} p_x &= \text{Tr} [\psi_4 |x\rangle\langle x| \otimes \mathbb{1} \otimes \mathbb{1}] \\ &= \frac{1}{4} \text{Tr} \left[\rho \otimes \tau + (-1)^x (\rho \otimes \tau) SW + (-1)^x SW(\rho \otimes \tau) + \tau \otimes \rho \right] \\ &= \frac{1}{4} \left(2 + (-1)^x \text{Tr} \left[(\rho \otimes \tau) SW + SW(\rho \otimes \tau) \right] \right), \end{aligned} \quad (\text{B.2})$$

where we noted that $\text{Tr} [\rho \otimes \tau] = \text{Tr} [\tau \otimes \rho] = 1$ since the both terms are traces of a quantum state. Let us now look at the term with the trace in (B.2),

$$\begin{aligned} & \text{Tr} \left[(\rho \otimes \tau) SW + SW(\rho \otimes \tau) \right] \quad (\text{B.3}) \\ = & \text{Tr} \left[\left(\sum_{i=1}^2 r_i |r_i\rangle\langle r_i| \otimes \sum_{j=1}^2 t_j |t_j\rangle\langle t_j| \right) SW + SW \left(\sum_{i=1}^2 r_i |r_i\rangle\langle r_i| \otimes \sum_{j=1}^2 t_j |t_j\rangle\langle t_j| \right) \right] \quad (\text{B.4}) \end{aligned}$$

$$= \text{Tr} \left[\left(\sum_{i,j=1}^2 r_i t_j |r_i\rangle\langle t_j| \otimes |t_j\rangle\langle r_i| \right) + \left(\sum_{i,j=1}^2 r_i t_j |t_j\rangle\langle r_i| \otimes |r_i\rangle\langle t_j| \right) \right] \quad (\text{B.5})$$

$$= 2 \sum_{i,j=1}^2 r_i t_j \langle r_i | t_j \rangle \langle t_j | r_i \rangle = 2 \text{Tr} \left[\sum_{i,j=1}^2 r_i t_j |r_i\rangle\langle r_i| |t_j\rangle\langle t_j| \right] \quad (\text{B.6})$$

$$= 2 \text{Tr} \left[\sum_{i=1}^2 r_i |r_i\rangle\langle r_i| \sum_{j=1}^2 t_j |t_j\rangle\langle t_j| \right] = 2 \text{Tr} \rho \tau. \quad (\text{B.7})$$

In derivations above, we used the property $\text{Tr} |\alpha\rangle\langle\alpha| = \langle\alpha|\alpha\rangle$ and applied the spectral decompositions $\rho = \sum_{i=1}^2 r_i |r_i\rangle\langle r_i|$ and $\tau = \sum_{j=1}^2 t_j |t_j\rangle\langle t_j|$. Therefore, the probability of getting the outcome $x \in \{0, 1\}$ after executing the swap test circuit (B.1)

is

$$p_x = \frac{1}{2} (1 + (-1)^x \text{Tr } \rho\tau). \quad (\text{B.8})$$

So far we considered the case when ρ and τ are single-qubit states, but the swap test can be easily generalized for the case of many qubits.

Obviously, if the input states are pure, i.e. $\rho = |r\rangle\langle r|$ and $\tau = |t\rangle\langle t|$, then (B.8) reduces to

$$p_x = \frac{1}{2} (1 + (-1)^x |\langle r|t\rangle|^2).$$

Interestingly, the post-measurement state of the swap test circuit in this case is

$$|\psi_{\text{pm}}^x\rangle = \frac{1}{2\sqrt{p_x}} |x\rangle \otimes (|r\rangle \otimes |t\rangle + (-1)^x |t\rangle \otimes |r\rangle),$$

i.e. some information about the input states $|r\rangle$ and $|t\rangle$ is preserved. Also remarkable is the fact that there is no quantum channel which could map this state back to $|r\rangle \otimes |t\rangle$ or $|t\rangle \otimes |r\rangle$ for any $|r\rangle$ and $|t\rangle$ [141], while the reverse process is always possible (recall that this is done in the swap test itself). So, it is impossible to utilize this state again to conduct another swap test.

B.2 Destructive swap test

In the previous section, we considered the swap test method which allows to compute $\text{Tr } \rho\tau$ for any given ρ and τ . As we noted, this approach requires an auxiliary qubit and the ability to apply the three-qubit controlled swap gate CSW . Although several decompositions of this gate in terms of one- and two-qubit gates are known, it is still difficult to implement such an operation on a quantum computer. To circumvent this obstacle, one may apply the so-called destructive swap-test.

Consider the following circuit:



where ρ and τ are n -qubit states, and each line, consequently, represents an n -qubit

register. Suppose that the probability to obtain the result $a \in \mathbb{B}^n$ after measuring the first register and the result $b \in \mathbb{B}^n$ after measuring the second is p_{ab}^d . In [141] it is proven that

$$\sum_{a,b \in \mathbb{B}^n} (a \cdot b \bmod 2) p_{ab}^d = \frac{1}{2}(1 - \text{Tr } \rho\tau), \quad (\text{B.10})$$

where $a \cdot b$ is the scalar product of two bit strings of the size n . It is easy to notice that the expression (B.10) is actually the probability of obtaining the outcome “1” in the classic swap test p_1 defined in (B.8) [142]. Similarly, the probability of the outcome “0” in the classic swap test is connected with the probabilities for the destructive swap test through the expression

$$\sum_{a,b \in \mathbb{B}^n} (1 - a \cdot b \bmod 2) p_{ab}^d = \frac{1}{2}(1 + \text{Tr } \rho\tau),$$

As follows from its name, the destructive swap test destroys almost all information about the input states. This seems to be not a very big loss since, first, it is not obvious how one can utilize the most-measurement state of the classic swap test, and if it is possible at all; second, the destructive swap test has crucial advantages before its classic version: there is no need in an auxiliary qubit, and one does need to be able to apply the CSW gate.

Appendix C

Hadamard test

In quantum computations, we usually dealing with expected values of Hermitian operators. Still, it is possible to separately measure the real and imaginary part of an expectation value of the form $\langle \psi | U | \psi \rangle$, where U is unitary. This is helpful in e.g. computing gradients of several cost functions. In Fig. C-1 shown are the circuits which are designed compute $\text{Re} \langle \psi | U | \psi \rangle$ and $\text{Im} \langle \psi | U | \psi \rangle$, and the whole technique is called the Hadamard test.

Let us consider the circuit depicted in Fig. C-1 on the left. The initial state of this circuit is

$$|\varphi_0\rangle = |0\rangle \otimes |\psi\rangle.$$

Then, the first qubit is acted by the Hadamard gate producing the state

$$|\varphi_1\rangle = \frac{1}{\sqrt{2}} (|0\rangle \otimes |\psi\rangle + |1\rangle \otimes |\psi\rangle).$$

After that, the second qubit is acted by the gate U controlled by the state of the first qubit:

$$|\varphi_2\rangle = \frac{1}{\sqrt{2}} (|0\rangle \otimes |\psi\rangle + |1\rangle \otimes U |\psi\rangle).$$

Finally, the first qubit is once again acted by the Hadamard gate, which gives

$$|\varphi_3\rangle = \frac{1}{2} (|0\rangle \otimes (|\psi\rangle + U |\psi\rangle) + |1\rangle \otimes (|\psi\rangle - U |\psi\rangle)).$$

The probability p_x^R of obtaining the outcome $x \in \{0, 1\}$ after measuring the first

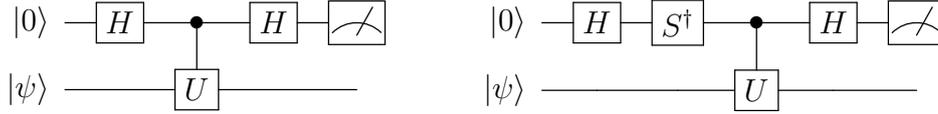


Figure C-1: Quantum circuits for calculating the real and imaginary parts of the complex number $\langle \psi | U | \psi \rangle$, where U is unitary. For the left circuit, the probability of obtaining the result $x \in \{0, 1\}$ after measuring the auxiliary qubit is $p_x^R = \frac{1}{2}(1 + (-1)^x \text{Re} \langle \psi | U | \psi \rangle)$; for the right circuit, this probability is $p_x^I = \frac{1}{2}(1 + (-1)^x \text{Im} \langle \psi | U | \psi \rangle)$.

qubit is

$$\begin{aligned}
 p_x^R &= \text{Tr} |\varphi_3\rangle\langle\varphi_3| |x\rangle\langle x| \otimes \mathbb{1} \\
 &= \frac{1}{4} \left[|\langle x|0\rangle|^2 (\langle \psi | + \langle \psi | U^\dagger) (|\psi\rangle + U |\psi\rangle) + |\langle x|1\rangle|^2 (\langle \psi | - \langle \psi | U^\dagger) (|\psi\rangle - U |\psi\rangle) \right] \\
 &= \frac{1}{4} \left[\delta_{x0} (2 + \langle \psi | U | \psi \rangle + \langle \psi | U | \psi \rangle) + \delta_{x1} (2 - \langle \psi | U | \psi \rangle - \langle \psi | U^\dagger | \psi \rangle) \right] \\
 &= \frac{1}{4} \left[\delta_{x0} (2 + 2 \text{Re} \langle \psi | U | \psi \rangle) + \delta_{x1} (2 - 2 \text{Re} \langle \psi | U | \psi \rangle) \right] \\
 &= \frac{1}{2} \left(1 + (-1)^x \text{Re} \langle \psi | U | \psi \rangle \right).
 \end{aligned}$$

The same can be proven for the circuit on the right in Fig. C-1. That is,

$$p_x^I = \frac{1}{2} \left(1 + (-1)^x \text{Im} \langle \psi | U | \psi \rangle \right).$$

Therefore, the Hadamard test technique in two experiments allows to compute

$$\langle \psi | U | \psi \rangle = \text{Re} \langle \psi | U | \psi \rangle + i \text{Im} \langle \psi | U | \psi \rangle$$

C.1 Modified Hadamard test

The Hadamard test method can be modified for computing a scalar product $\langle \alpha | \beta \rangle$, provided that it is known how to prepare the states $|\alpha\rangle = A |0\rangle$ and $|\beta\rangle = B |0\rangle$. The circuits for computing the real and imaginary part of $\langle \alpha | \beta \rangle$ are shown in Fig. C-2. The logic of obtaining the wanted quantities is the same as for the usual Hadamard test. Indeed, for the circuits on the left and on the right in Fig. C-2, the states

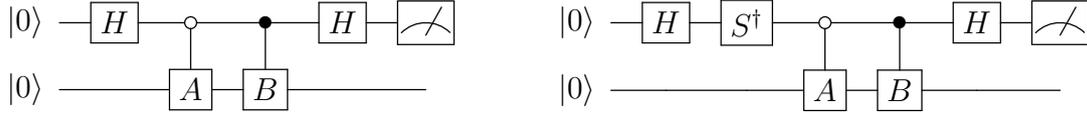


Figure C-2: Quantum circuits for calculating the real and imaginary parts of the complex number $\langle 0|A^\dagger B|0\rangle$, where A and B are unitary. For the left circuit, the probability of obtaining the result $x \in \{0, 1\}$ after measuring the auxiliary qubit is $p_x^R = \frac{1}{2}(1 + (-1)^x \text{Re}\langle 0|A^\dagger B|0\rangle)$; for the right circuit, this probability is $p_x^I = \frac{1}{2}(1 + (-1)^x \text{Im}\langle 0|A^\dagger B|0\rangle)$.

before the measurement are, respectively,

$$\begin{aligned}
 |\varphi_3^R\rangle &= \frac{1}{2} \left(|0\rangle \otimes (A|0\rangle + B|0\rangle) + |1\rangle \otimes (A|0\rangle - B|0\rangle) \right), \\
 |\varphi_3^I\rangle &= \frac{1}{2} \left(|0\rangle \otimes (A|0\rangle - iB|0\rangle) + |1\rangle \otimes (A|0\rangle + iB|0\rangle) \right).
 \end{aligned}$$

The probabilities of obtaining the outcome $x \in \{0, 1\}$ for these states are

$$\begin{aligned}
 p_x^R &= \frac{1}{2} \left(1 + (-1)^x \text{Re}\langle 0|A^\dagger B|0\rangle \right), \\
 p_x^I &= \frac{1}{2} \left(1 + (-1)^x \text{Im}\langle 0|A^\dagger B|0\rangle \right).
 \end{aligned}$$

Therefore, in two experiments of executing the circuits of the modified Hadamard test, one can compute the dot product of two vectors as

$$\langle \alpha|\beta\rangle = \langle 0|A^\dagger B|0\rangle = \text{Re}\langle 0|A^\dagger B|0\rangle + i\text{Im}\langle 0|A^\dagger B|0\rangle.$$