Quantum-machine-learning channel discrimination

A. S. Kardashin,* A. V. Vlasova, A. A. Pervishko, D. Yudin, and J. D. Biamonte†
Skolkovo Institute of Science and Technology, Moscow 121205, Russia

(Received 21 June 2022; accepted 16 August 2022; published 7 September 2022)

In the problem of quantum channel discrimination, one distinguishes between a given number of quantum channels, which is done by sending an input state through a channel and measuring the output state. This work studies applications of variational quantum circuits and machine-learning techniques for discriminating such channels. In particular, we explore (i) the practical implementation of embedding this task into the framework of variational quantum computing, (ii) training a quantum classifier based on variational quantum circuits, and (iii) applying the quantum kernel estimation technique. For testing these three channel discrimination approaches, we considered a pair of entanglement-breaking channels and the depolarizing channel with two different depolarization factors. For the approach (i), we address solving the quantum channel discrimination problem using widely discussed parallel and sequential strategies. We show the advantage of the latter in terms of better convergence with less quantum resources. Quantum channel discrimination with a variational quantum classifier (ii) allows one to operate even with random and mixed input states and simple variational circuits. The kernel-based classification approach (iii) is also found effective as it allows one to discriminate depolarizing channels associated not with just fixed values of the depolarization factor, but with ranges of it. Additionally, we discovered that a simple modification of one of the commonly used kernels significantly increases the efficiency of this approach. Finally, our numerical findings reveal that the performance of variational methods of channel discrimination depends on the trace of the product of the output states. These findings demonstrate that quantum machine learning can be used to discriminate channels, such as those representing physical noise processes.

DOI: 10.1103/PhysRevA.106.032409

I. MOTIVATION

A quantum channel is a linear completely positive trace-preserving map which transforms quantum states into quantum states. The problem of quantum channel discrimination, i.e., distinguishing between a given finite set of channels, is ubiquitous in quantum information and quantum communication [1–4]. Solving this task forms the core of various applications, including but not limited to photonic sensing [5], target quantum detection via quantum illumination [6,7], and quantum reading [8]. Within a general approach for implementing quantum channel discrimination one sends an input state as specified by its density operator ρ^{in} through a channel Φ_y randomly selected from a collection $\{\Phi_j\}_{j=1}^N$. The output state $\rho^{\text{out}} = \Phi_y[\rho^{\text{in}}]$ may be equivalently expressed in terms of positive-operator-valued measures (POVM) [9]. The POVM is constituted by a set of non-negative Hermitian operators $\Pi = \{\Pi_j\}_{j=1}^N$ that add up to the identity. The probability for the quantum system to be in a particular state $\Phi_y[\rho]$ is determined by the expectation value of the POVM operator corresponding to this state $\text{Tr}(\Pi_y \Phi_y[\rho^{\text{in}}])$ [10,11], whereas quantum channel discrimination should return the index y which labels the channel having been applied.

Depending on the available computational resources and the properties of a given channel Φ_y, one might adopt various strategies for discrimination [12–15]. In particular, it is expected that quantum channels can be discriminated more efficiently if it is allowed to apply a given channel several times. For different channels and discrimination strategies, the efficiency of discrimination as well as the associated quantitative measures have previously been considered [16–23]. Of particular interest and technological relevance [24] are the parallel and sequential strategies [24]. Given the opportunity to apply a channel a fixed number of times, the parallel strategy implies that a channel is applied on distinct quantum subsystems (e.g., qubits) simultaneously, whereas in the sequential strategy a channel acts on the same subsystem step by step.

The problem of quantum channel discrimination can be seen as an optimization task in the space of parameters specified by the input state ρ^{in} and the POVM Π. This allows one to put this task into the framework of variational quantum computing [25–32] and quantum machine learning [33–37]. In the variational scheme, a quantum processor is used to prepare a family of probe states with a polynomial number of parameters, while minimizing a given loss function within this family of states is achieved by a means of classical optimization algorithms. In this case, one minimizes the use of quantum resources by delegating as much computation as possible to the classical device. Machine learning as implemented with variational quantum circuits [38–40] allows one to solve classification tasks for quantum-embedded classical data [41], learn phases of quantum matter [42–46], and discriminate quantum states [47,48]. The two latter tasks are peculiar in a sense that they are solved for quantum data by quantum

*andrey.kardashin@skoltech.ru
†http://quantum.skoltech.ru

means, which makes them a part of the rapidly advancing quantum-quantum machine learning [49,50], and the problem of channel discrimination can be classified likewise.

In this paper, we highlight the use of variational quantum circuits and machine-learning techniques for quantum channel discrimination. We start our analysis with formulating this task as an optimization problem, followed by a direct application of variational quantum scheme. Then we discuss how one can use variational circuits for binary classification of quantum channels. After that, we demonstrate that such a binary classifier can be also based on a kernel, a specific real-valued function of output states of channels $\mathcal{K}(\Phi_i[\rho], \Phi_j[\rho])$ for an input state ρ. We test these three methods of channel discrimination by distinguishing between a pair of entanglement-breaking channels [51,52], and the qubit depolarizing channel [53,54].

II. QUANTUM CHANNEL DISCRIMINATION

The problem of binary quantum channel discrimination can equivalently be viewed as the game of two parties, Alice and Bob. This game is consisted of the active stage and the training stage. At the beginning of the active stage, Alice prepares the state ρ^{in} and sends it to Bob. Then, Bob, in secret, randomly and with equal probabilities chooses a channel $\Phi_y \in \{\Phi_0, \Phi_1\}$ and applies it to Alice's state, so that the output state is $\rho^{\text{out}} = \Phi_y[\rho^{\text{in}}]$. Having received this state from Bob, Alice measures the output state with the POVM elements $\Pi = \{\Pi_0, \Pi_1 = \mathbb{1} - \Pi_0\}$, and label the outcome "0" for the channel Φ_0 and "1" otherwise. The goal of Alice therefore is to find the input state ρ^{in} and the POVM elements $\Pi = \{\Pi_0, \Pi_1\}$ such that they maximize the probability p_{00} of getting the measurement outcome 0 if Bob applied the channel Φ_0 and the probability p_{11} of getting the outcome 1 if the applied channel was Φ_1. These probabilities are

$$\text{p}_{00} = \tfrac{1}{2}\text{Tr}(\Pi_0\Phi_0[\rho^{\text{in}}]),$$
$$\text{p}_{11} = \tfrac{1}{2}\text{Tr}(\Pi_1\Phi_1[\rho^{\text{in}}]),$$

where the factors $\frac{1}{2}$ are essentially the prior probabilities of application of the channel Φ_0 or Φ_1 since Bob chooses them randomly and equiprobably. As $\text{p}_{00} + \text{p}_{11}$ forms the total probability of successful discrimination, then the task of Alice is to maximize it. This can formally described as the following optimization problem:

$$\begin{aligned}\text{p}_{\text{s}} &\equiv \max\{\text{p}_{00} + \text{p}_{11}\} \\ &= \frac{1}{2}\max_{\rho^{\text{in}},\Pi}\{\text{Tr}(\Pi_0\Phi_0[\rho^{\text{in}}]) + \text{Tr}(\Pi_1\Phi_1[\rho^{\text{in}}])\}. \end{aligned} \tag{1}$$

Alongside with the probability of erroneous discrimination $\text{p}_{\text{e}} = \text{p}_{01} + \text{p}_{10}$, it obviously sums to unity, $\text{p}_{\text{s}} + \text{p}_{\text{e}} = 1$.

To be able to solve the problem (1) for Alice, there is the training stage in the game. During this stage, we assume that for each pair (ρ^{in}, Π) picked by Alice, Bob provides as many copies of the state $\rho^{\text{out}} = \Phi_y[\rho^{\text{in}}]$ as needed, without changing the label $y \in \{0, 1\}$. Moreover, for each output state, Alice is informed about the channel label y. This is equivalent to the assumption that Alice can measure the output state ρ^{out} arbitrarily many times, i.e., it is possible to compute the probabilities in (1) exactly. This is opposed to the situation at the game's active stage, when only one measurement is allowed, and the channel label y is kept in secret. Clearly, the described game scheme is analogous to binary classification task as realized in the context of supervised machine learning.

In this particular setting, the probability of successful quantum channel discrimination that can be achieved is upper bounded by [55]

$$\text{p}_{\diamond} = \tfrac{1}{2} - \tfrac{1}{4}||\Phi_0 - \Phi_1||_{\diamond}, \tag{2}$$

where

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

is the so-called diamond norm with Φ being a channel that maps density operators on a Hilbert space $\mathcal{H}$, $\mathbb{1}$ the identity map on $\mathcal{H}_E$, ρ a density operator on $\mathcal{H} \otimes \mathcal{H}_E$, and $||A||_1 = \text{Tr}\sqrt{A^\dagger A}$.

A. Discrimination strategies

So far we have described a channel guessing game when a channel Φ_y is applied only once. If, however, Alice is allowed to pass a chosen probe state ρ^{in} through Bob's channel Φ_y a finite and fixed number of times p, provided that for each time the channel label y remains the same, then Alice can adjust the discrimination strategy by asking Bob to apply the channel Φ_y in a specific way. One of such ways is applying the channel p times in parallel, so that the channel acts simultaneously on the separate subsystems of a composite input state ρ^{in}. Another approach is to apply the channel sequentially such that it acts on a single subsystem of a potentially composite state ρ^{in}. These two discrimination methods are widely discussed in literature and are called the parallel and sequential strategies, respectively. In what follows, we formally describe these strategies assuming that the channels Φ_y are qubit-to-qubit mappings.

1. Parallel strategy

Consider the previously discussed quantum channel discrimination game and suppose that now Alice can ask Bob to apply the channel Φ_y a fixed number of times p. In general words then, the parallel channel discrimination strategy implies that the input state of Alice ρ^{in} is at least p qubit, and Bob acts by the channels Φ_y on each of the p qubits separately and simultaneously. Alice is also allowed to have the input state of more than p qubits, i.e., to add $r \geqslant 0$ auxiliary qubits to it, as it might potentially help in solving quantum channel discrimination task in case of an entangled input state [14,56,57]. The resultant $(p + r)$-qubit state is then measured with a POVM Π. This discrimination strategy is schematically shown in Fig. 1.

Let us describe the parallel strategy more formally. Suppose that the channel Φ_y can be applied p times. Then, first, Alice prepares a $(p + r)$-qubit state ρ^{in}_{PR}, with P and R specifying the registers of p and r qubits, respectively. The qubits of P are then sent through the channels of Bob in parallel, while the register R remains unaffected. Formally, the output state is

$$\rho^{\text{out}} = \left(\Phi_y^{\otimes p} \otimes \mathbb{1}^{\otimes r}\right)\left[\rho^{\text{in}}_{PR}\right], \tag{3}$$

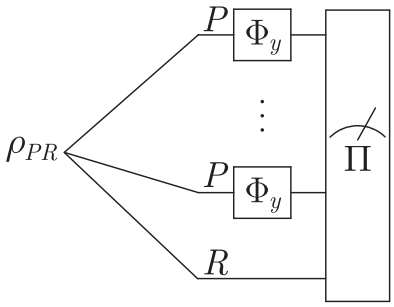

FIG. 1. A schematic of the parallel strategy. Provided with the opportunity of p applications of Bob's channel Φ_y, Alice prepares a composite state ρ_{PR}^{in} of p qubits of the register P and r qubits of the register R. The qubits of P are sent through the channels Φ_y, while the qubits of R remain untouched. In the scheme, the lines coming from ρ_{PR}^{in} indicate the subsystems of the corresponding registers. At the end, all $p+r$ qubits are measured with the POVM Π.

where $\Phi_y^{\otimes p}$ acts only on the subsystem P and $\mathbb{1}^{\otimes r}$ is the identity on the subsystem R. Alice then measures the output state ρ^{out} with $\Pi = \{\Pi_0, \Pi_1\}$. Therefore, in the parallel strategy, according to Eq. (1) one has to maximize

$$\mathrm{p}_{\mathrm{s}}^{\text{par}} = \frac{1}{2}\max_{\rho_{PR}^{\text{in}},\Pi}\big\{\mathrm{Tr}\big(\Pi_0\big(\Phi_0^{\otimes p}\otimes\mathbb{1}^{\otimes r}\big)\big[\rho_{PR}^{\text{in}}\big]\big) + \mathrm{Tr}\big(\Pi_1\big(\Phi_1^{\otimes p}\otimes\mathbb{1}^{\otimes r}\big)\big[\rho_{PR}^{\text{in}}\big]\big)\big\}, \tag{4}$$

where the optimization over r, specifying the number of qubits in the register R, is implicitly included. As was mentioned earlier, introducing this auxiliary register allows to have the entanglement between the qubits of the registers P and R, which may lead to more efficient channel discrimination. For the described strategy, the probability of successful channel discrimination is yielded by

$$\mathrm{p}_{\diamond}^{\text{par}}(p) = \tfrac{1}{2} - \tfrac{1}{4}\big\|\Phi_0^{\otimes p} - \Phi_1^{\otimes p}\big\|_{\diamond} \tag{5}$$

for p parallel applications of Bob's channel.

2. *Sequential strategy*

Now suppose that Alice is again allowed to apply the channel Φ_y a fixed number of times p. This time, in the sequential discrimination strategy, Alice's input state ρ^{in} can be single qubit, and this qubit can be passed through a channel p times in a row. But, after each application of the channel Φ_y, Bob sends the corresponding output state back to Alice, who is allowed to modify it before sending back to Bob again. After the pth-channel application, the resultant state is measured with the POVM Π. Like in the parallel strategy, Alice can add r auxiliary qubits to have a $(1+r)$-qubit entangled input state. The sequential channel discrimination strategy is schematically shown in Fig. 2.

Formally, this strategy can be described as follows. First, Alice prepares the input state ρ_{PR}^{in} which consists of $(1+r)$ qubits: the subsystem P of one qubit is acted by the channel Φ_y, whereas the subsystem R of r qubits remains unaffected. Suppose Alice is allowed to apply the channel p times. Alice then sends the input state ρ_{PR}^{in} to Bob and receives back the state $\tilde{\rho}_{PR} = (\Phi_y\otimes\mathbb{1}^{\otimes r})[\rho_{PR}^{\text{in}}]$. After that, Alice applies a quantum channel ε_1 on the state $\tilde{\rho}_{PR}$ which gives $\rho_1 = \varepsilon_1[\tilde{\rho}_{PR}]$.

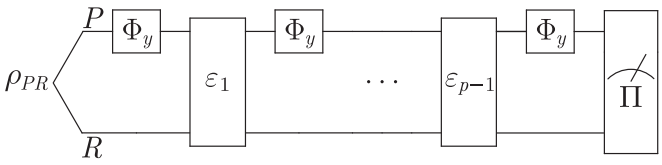

FIG. 2. A schematic of the sequential strategy. Provided with the opportunity of p applications of Bob's channel Φ_y, Alice prepares the composite state ρ_{PR}^{in} of one qubit of the register P and r qubits of the register R. The qubit of P is sent through the channels Φ_y, while the qubits of R remain untouched. After the jth application of the channel Φ_y, Alice modifies the whole state by applying the channel $\mathcal{E}_j$. In the scheme, the lines coming from ρ_{PR}^{in} indicate the subsystems of the corresponding registers. At the end, all $1+r$ qubits are measured with the POVM Π.

This procedure is repeated $(p-1)$ times until Alice has the state $\rho_{p-1} = \varepsilon_{p-1}[\rho_{p-2}]$, and at the end Alice passes the subsystem P through the channel Φ_y the last, pth time, and gets $\rho^{\text{out}} = (\Phi_y\otimes\mathbb{1}^{\otimes r})[\rho_{p-1}]$. More formally, the whole process can be described as

$$\rho^{\text{out}} = \mathcal{C}(\Phi_y, \mathcal{E})\big[\rho_{PR}^{\text{in}}\big], \tag{6}$$

where $\mathcal{E} = \{\varepsilon_j\}_{j=1}^{p-1}$ and

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

with the channel composition operation $(B\circ A)[\rho] \equiv B[A[\rho]]$. The output state ρ^{out} is then measured using the POVM $\Pi = \{\Pi_0, \Pi_1\}$. In this strategy, the optimization problem (1) becomes

$$\mathrm{p}_{\mathrm{s}}^{\text{seq}} = \frac{1}{2}\max_{\rho_{PR}^{\text{in}},\Pi,\mathcal{E}}\big\{\mathrm{Tr}\big(\Pi_0\,\mathcal{C}(\Phi_0,\mathcal{E})\big[\rho_{PR}^{\text{in}}\big]\big) + \mathrm{Tr}\big(\Pi_1\,\mathcal{C}(\Phi_1,\mathcal{E})\big[\rho_{PR}^{\text{in}}\big]\big)\big\}, \tag{7}$$

where in addition to the input state ρ_{PR}^{in} and measurement Π, Alice has to optimize over the channels $\mathcal{E}$ as well.

As a rule, the sequential strategy, incarnating the idea of quantum comb [5,58], provides better discrimination results compared to the parallel strategy [22,23]. Meanwhile, it is worth mentioning a more general discrimination strategy that is based on the so-called indefinite casual order of channel application [15]. It was shown that for p^{par}, p^{seq}, and p^{ico} being the upper bounds for successful discrimination probabilities of the parallel, sequential, and indefinite casual order strategies, respectively, there exists a pair of target channels Φ_0 and Φ_1 satisfying

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

Although the indefinite casual order strategy gives advantage over the parallel and sequential ones, we focus on the latter two in what follows.

B. Variational circuit formulation

Let us embed the parallel and sequential discrimination strategies into the framework of variational quantum circuits. That is, we replace all transformations of quantum states by

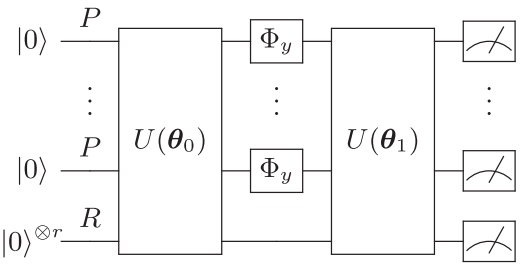

FIG. 3. A variational quantum circuit implementing the parallel channel discrimination strategy from Fig. 1. The initial state ρ_{PR}^{in} is prepared from $|0\rangle_P^{\otimes p} \otimes |0\rangle_R^{\otimes r}$ via the unitary transformation $\mathcal{U}(\boldsymbol{\theta_0})$, while the unitary $\mathcal{U}(\boldsymbol{\theta_1})$ is used to rotate the measurement basis. If the number of allowed applications of the channel Φ_y is p and the number of ancillary qubits is r, the technique requires $(p + r)$ qubits. Note that r could be set to zero.

parametrized unitary operators. Having done that, we accordingly reformulate the optimization problems for $\text{p}_\text{s}^{\text{par}}$ and $\text{p}_\text{s}^{\text{seq}}$ defined in (4) and (7), respectively. For the case of a single-channel application ($p = 1$), a similar approach of embedding was applied for discriminating various qubit-to-qubit channels [59].

1. Parallel strategy

The parallel channel discrimination strategy (4) embedded into the framework of variational circuits is depicted in Fig. 3. In this circuit, the probe state ρ_{PR}^{in} is prepared 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 unitary operator parametrized by a set of real numbers $\boldsymbol{\theta}_0$. The register P of p qubits for ρ_{PR}^{in} is acted then by the p parallel applications of the channel Φ_y, as it is done in (3). After that, applied is the unitary $U(\boldsymbol{\theta}_1)$, which can be viewed as a rotation of the measurement basis. The output state then becomes

$$\rho^{\text{out}}(\boldsymbol{\theta}, \Phi_y, p, r) = \mathcal{U}(\boldsymbol{\theta_1}) \circ \mathcal{F}(\Phi_y, p, r) \circ \mathcal{U}(\boldsymbol{\theta_0})[\rho_0(p, r)], \tag{8}$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}_0 \cup \boldsymbol{\theta}_1$ and $\mathcal{F}(\Phi_y, p, r) = (\Phi_y^{\otimes p} \otimes \mathbb{1}^{\otimes r})$. Clearly, $d = 2^{p+r}$ states $|j\rangle = \{|i_1 i_2 \ldots i_{p+r}\rangle\}$ with the entries $i_n = \{0, 1\}$ is a span of the Hilbert space of the registers P and R. In our numerical simulations, we split the computational basis in two parts of $d/2$ basis vectors each and associate the measurement outcomes with the projectors

$$\Pi_0 = \sum_{j=1}^{d/2} |j\rangle\langle j|, \quad \Pi_1 = \sum_{j=d/2+1}^{d} |j\rangle\langle j|. \tag{9}$$

As the result, we come up with the optimization problem that can be addressed using a hybrid quantum-classical setup:

$$\text{p}_\text{s}^{\text{par}} = \tfrac{1}{2} \max_{\boldsymbol{\theta}, r} \{\text{Tr}[\Pi_0\, \rho^{\text{out}}(\boldsymbol{\theta}, \Phi_0, p, r)] + \text{Tr}[\Pi_1\, \rho^{\text{out}}(\boldsymbol{\theta}, \Phi_1, p, r)]\}. \tag{10}$$

Thus, (10) together with the circuit in Fig. 3 define the framework for numerically testing the parallel discrimination strategy.

2. Sequential strategy

Similarly to the parallel channel discrimination strategy, in Fig. 4 we depict the sequential strategy (7) formulated in the framework of variational circuits. In this variational approach, Alice's channels $\mathcal{E}$ can be applied to the input state ρ_{PR}^{in} via the Stinespring representation [60]. Particularly, $\varepsilon[\rho_{PR}]$ can be implemented by adding an ancillary register E in the state ρ_E of e qubits and performing a unitary evolution U of the joint state $\rho_{PR} \otimes \rho_E$, followed by tracing out the register E. That is, we have

$$\varepsilon[\rho_{PR}] = \text{Tr}_E[U(\rho_{PR} \otimes \rho_E)U^\dagger]. \tag{11}$$

For this transformation to be general, the register E should contain e qubits twice as the size of the registers P and R together [61]. In our setting, this is equal to $e = 2(1 + r)$ with one qubit in P and r qubits in R.

However, in the followup analysis, we reduce the transformation (11) to the one shown in Fig. 4. In this approach, we incorporate the register E into R. Furthermore, instead of the channels $\mathcal{E} = \{\varepsilon_j\}_{j=1}^{p-1}$ we use parametrized unitaries $\{U(\boldsymbol{\theta}_j)\}_{j=1}^{p-1}$, alongside with the operator $U(\boldsymbol{\theta}_0)$ which prepares the initial state and the operator $U(\boldsymbol{\theta}_p)$ which rotates the measurement basis, as done in (8). Analogous to (6), the output state is

$$\rho^{\text{out}}(\boldsymbol{\theta}, \Phi_y, p, r) = \mathcal{C}(\boldsymbol{\theta}, \Phi_y, p, r)[\rho_0(r)], \tag{12}$$

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

$$\mathcal{C}(\boldsymbol{\theta}, \Phi_y, p, r) = \mathcal{U}(\boldsymbol{\theta}_p) \circ (\Phi_y \otimes \mathbb{1}^{\otimes r}) \circ \mathcal{U}(\boldsymbol{\theta}_{p-1}) \circ \cdots \circ (\Phi_y \otimes \mathbb{1}^{\otimes r}) \circ \mathcal{U}(\boldsymbol{\theta}_1) \circ (\Phi_y \otimes \mathbb{1}^{\otimes r}) \circ \mathcal{U}(\boldsymbol{\theta}_0),$$

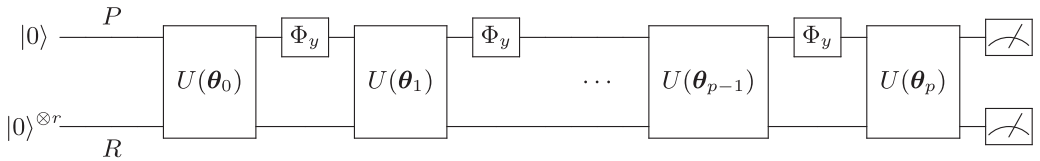

FIG. 4. A variational quantum circuit implementing the sequential channel discrimination strategy from Fig. 2. Alice's channels $\mathcal{E} = \{\varepsilon_j\}_{j=1}^{p-1}$ are replaced by the parametrized unitaries $\{U(\boldsymbol{\theta}_j)\}_{j=1}^{p-1}$, where p is the number of allowed applications 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.

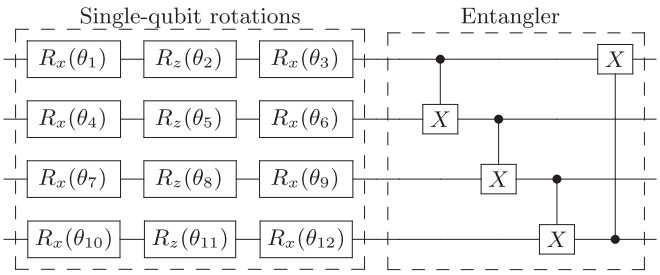

FIG. 5. A layer of the hardware-efficient *Ansatz* with 4 input qubits and 12 variational parameters. Here, $R_\sigma(\theta) = e^{-\imath\theta\sigma}$ with $\sigma \in \{X, Y, Z\}$ specifying the Pauli operator and $\theta \in [0, 2\pi)$ being the optimization parameters.

with $\boldsymbol{\theta} = \bigcup_{k=0}^{p} \boldsymbol{\theta}_k$. The channels $\mathcal{U}$ should be interpreted as in (8). Note that there is also $U(\boldsymbol{\theta}_0)$ which is used to prepare the input state $\rho_{PR}^{\text{in}} = \mathcal{U}(\boldsymbol{\theta}_0)[\rho_0(r)]$. Technically, in line with Fig. 4, this can be viewed as a channel ε_0 which maps the single-qubit state $|0\rangle\langle 0|_P$ from the Hilbert space $\mathcal{H}_P$ to a $(1 + r)$-qubit state in $\mathcal{H}_P \otimes \mathcal{H}_R$ as

$$\rho_{PR}^{\text{in}} = \varepsilon_0[|0\rangle\langle 0|_P] = E_0|0\rangle\langle 0|_P E_0^\dagger + E_1|0\rangle\langle 0|_P E_1^\dagger, \tag{13}$$

where the Kraus operators are $E_j = U(\boldsymbol{\theta}_0)(|j\rangle_P \otimes |0\rangle_R)\langle j|_P$. The rest of the transformations are unitary and performed in the extended Hilbert space $\mathcal{H}_P \otimes \mathcal{H}_R$.

The output state (12) is then measured using the POVM elements (9) which leads to the optimization problem

$$\begin{aligned} \mathrm{p}_{\mathrm{s}}^{\text{seq}} = \frac{1}{2} \max_{\boldsymbol{\theta}, r} \{ &\mathrm{Tr}[\Pi_0\, \rho^{\text{out}}(\boldsymbol{\theta}, \Phi_0, p, r)] \\ &+ \mathrm{Tr}[\Pi_1\, \rho^{\text{out}}(\boldsymbol{\theta}, \Phi_1, p, r)] \}. \end{aligned} \tag{14}$$

As one can notice, the expressions for $\mathrm{p}_{\mathrm{s}}^{\text{par}}$ in (10) and $\mathrm{p}_{\mathrm{s}}^{\text{seq}}$ in (14) do not differ much except for the structure and the genesis of the output state ρ^{out}. In variational quantum computing, one may consider these expressions as objective functions, maximization of which leads to training of the corresponding circuits.

C. Numerical experiments

We herein describe the results of our numerical simulations demonstrating the capability of variational quantum channel discrimination. Although the number of repetitions p and the number of qubits r in the register R enter Eqs. (10) and (14), we do not optimize with respect to these parameters. Instead, we fix $p = 1, 2$ and vary the amount of ancillary qubits r upon optimizing the *Ansatz* parameters $\boldsymbol{\theta}$. The parametrized unitary operators $U(\boldsymbol{\theta}_k)$ are implemented in terms of the hardware-efficient *Ansatz* [62] whose four-qubit structure is shown in Fig. 5. This circuit is composed of several layers constituted by the universal single-qubit rotations and an entanglement block. In this *Ansatz*, the number of variational parameters s is polynomial in the total number of qubits $q = p + r$ and the number of layers l, $s = 3ql$ for $q > 2$. In what follows, we test the variational approach to discriminate the depolarizing channels and entanglement-breaking channels, mapping two-qubit states into one-qubit states.

1. *Entanglement-breaking channel discrimination*

We start our numerical analysis with the variational circuit approach to discriminating the entanglement-breaking channels

$$\Phi_0[\rho] = \sum_{j=1}^{5} A_j \rho A_j^\dagger, \quad \Phi_1[\rho] = \sum_{j=1}^{5} B_j \rho B_j^\dagger, \tag{15}$$

described by the Kraus operators

$$\begin{aligned} &A_1 = |0\rangle\langle 00|, \quad A_2 = |0\rangle\langle 01|, \quad A_3 = |0\rangle\langle 10|, \\ &A_4 = \frac{1}{\sqrt{2}}|0\rangle\langle 11|, \quad A_5 = \frac{1}{\sqrt{2}}|1\rangle\langle 11|, \end{aligned} \tag{16}$$

$$\begin{aligned} &B_1 = |+\rangle\langle 00|, \quad B_2 = |+\rangle\langle 01|, \quad B_3 = |1\rangle\langle 1+|, \\ &B_4 = \frac{1}{\sqrt{2}}|0\rangle\langle 1-|, \quad B_5 = \frac{1}{\sqrt{2}}|1\rangle\langle 1-| \end{aligned} \tag{17}$$

with $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. In this particular scenario, the parallel discrimination strategy was demonstrated to never reach unity of the successful discrimination probability (5), i.e., $\mathrm{p}_{\diamond}^{\text{par}}(p) < 1$ for any finite number of channel applications p [12]. At the same time, the sequential strategy with only $p = 2$ repetitions allows one to distinguish the channels perfectly, and the input state does not need to be entangled. To estimate $\mathrm{p}_{\diamond}^{\text{par}}$ defined in (5), one needs to calculate the diamond norm which can be done via semidefinite programming [63]. Using the CVXPY package [64,65], we calculated the probability $\mathrm{p}_{\diamond}^{\text{par}}(p)$ for $p = 1, 2$; our results reveal that $\mathrm{p}_{\diamond}^{\text{par}} \approx 0.9268$ for $p = 1$ and $\mathrm{p}_{\diamond}^{\text{par}} \approx 0.9771$ for $p = 2$.

We proceed by training the variational circuits depicted in Figs. 3 and 4. Recall that the circuits are trained by maximizing (10) and (14), respectively. For this task, we made use of the L-BFGS-B [66] optimization method. The explicit quantum circuits to be trained to discriminate the entanglement-breaking channels with $p = 2$ repetitions are shown in Fig. 6 for both the parallel and sequential strategies. In the parallel strategy, for $p = 2$ the success probability $\mathrm{p}_{\mathrm{s}}^{\text{par}} \approx \mathrm{p}_{\diamond}^{\text{par}} \approx 0.9771$ was achieved with the use of the hardware-efficient *Ansatz* of $l = 5$ layers representing $U(\boldsymbol{\theta}_0)$ and $U(\boldsymbol{\theta}_1)$. Note that to get this probability no ancillary qubits are needed, i.e., $r = 0$. In the sequential strategy, even a one-layer, $l = 1$, hardware-efficient *Ansatz* parametrizing the operators $U(\boldsymbol{\theta}_0)$, $U(\boldsymbol{\theta}_1)$, and $U(\boldsymbol{\theta}_2)$ provides the success probability $\mathrm{p}_{\mathrm{s}}^{\text{seq}} \approx 1$ for $p = 2$. Note that for $p = 1$, the circuits for both parallel and sequential discrimination look alike, and using a one-layer hardware-efficient *Ansatz* results in $\mathrm{p}_{\mathrm{s}} \approx \mathrm{p}_{\diamond} \approx 0.9268$.

It, however, might be excessive to use circuits with that many parameters, especially in the case of the sequential strategy. A successful discrimination, as implemented in the seminal work [12], did not use anything but the channels Φ_y and a single measurement at all. Meanwhile, even with these (over)parametrized circuits one is capable of identifying optimal input states and measurement bases such that $\mathrm{p}_{\mathrm{s}}^{\text{seq}} \approx 1$. It is thus reasonable to expect that the variational approach could be useful for channels Φ_y without any prior knowledge of them.

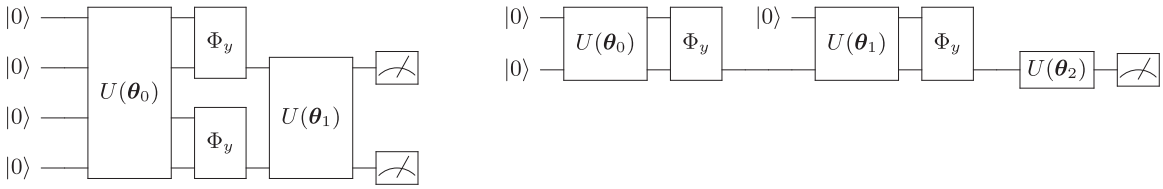

FIG. 6. The explicit variational quantum circuits trained for discriminating the entanglement-breaking channels (15) with $p = 2$ channel applications. On the left shown is the circuit for the parallel strategy and on the right is for the sequential strategy. Note that the channels to be discriminated map two-qubit states into one-qubit states. Therefore, for the sequential strategy, after the first application of the channel Φ_y, one has to add an extra qubit in some state (in our case, $|0\rangle$).

2. *Depolarizing channel discrimination*

Our study is continued with training the variational quantum circuits for discriminating depolarizing channels as given by

$$\Phi(\alpha)[\rho] = (1-\alpha)\rho + \frac{\alpha}{3}(\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z), \quad (18)$$

where the coefficient α determines the depolarization factor, while σ_i with $i = x, y, z$ stands for the Pauli operators. We consider a pair of channels with $0 \leqslant \alpha_0 \neq \alpha_1 \leqslant 1$. In our numerical simulations, we fixed the number of channel applications to $p = 2$; the number of qubits in the ancillary register R is set to $r = 3$ for the parallel strategy and $r = 4$ for the sequential strategy, so that the total number of qubits was $q = 5$ for both cases (see Figs. 3 and 4).

In Fig. 7 shown are the probabilities (10) and (14) for both parallel and sequential strategy, respectively, as evaluated for a pair $(\alpha_0, \alpha_1 = \alpha_0 + 0.1)$ starting from $\alpha_0 = 0$. To speed up the calculations, one might use random *Ansatz* parameters for the initial guess at $(\alpha_0 = 0.0,\ \alpha_1 = 0.1)$; whereas for each next pair up to $(\alpha_0 = 0.4,\ \alpha_1 = 0.5)$ the optimal parameters are taken as obtained at an earlier step. The same strategy can be successfully adopted for a set of parameters starting from $(\alpha_0 = 0.9,\ \alpha_1 = 1.0)$ with random initialization $\boldsymbol{\theta}$ down to $(\alpha_0 = 0.5,\ \alpha_1 = 0.6)$.

We have also explored how the achieved probabilities $\mathrm{p_s}$ depend on l, the number of layers of hardware-efficient *Ansatz* in variational circuits. Clearly, the sequential strategy gives better results and smaller variance with fewer number of layers. In this strategy, $l = 14$ layers turns out to be enough for achieving the success probability $\mathrm{p}_\diamond^{\mathrm{par}}$ for all pairs (α_0, α_1). In the parallel strategy, this result cannot be reproduced no matter how big the value l is (we tested for up to $l = 30$ layers). One may also notice that despite the same diamond distances (5), it is harder to achieve higher $\mathrm{p}_\diamond^{\mathrm{par}}$ for depolarization factor pairs on the right to $\alpha = 0.5$. As an instance, for the pairs $(\alpha_0 = 0.0,\ \alpha_1 = 0.1)$ and $(\alpha_0 = 0.9,\ \alpha_1 = 1.0)$ the theoretical probabilities are $\mathrm{p}_\diamond^{\mathrm{par}} = 0.595$, but the obtained probabilities $\mathrm{p_s}$ are lower for the second pair of depolarization factors. We conclude that there might be some dependence on the trace product between the states passed through the channels, $\mathrm{Tr}(\rho_0\rho_1)$ with $\rho_y \equiv \Phi(\alpha_y)[\rho]$ (see Appendix A for details).

III. BINARY QUANTUM CLASSIFIER

Previously, we discussed how one can use variational quantum computing framework to reformulate and solve the problem of channel discrimination. With a proper postprocessing of measurements, variational quantum circuits can also serve as a means to solve classification tasks. For

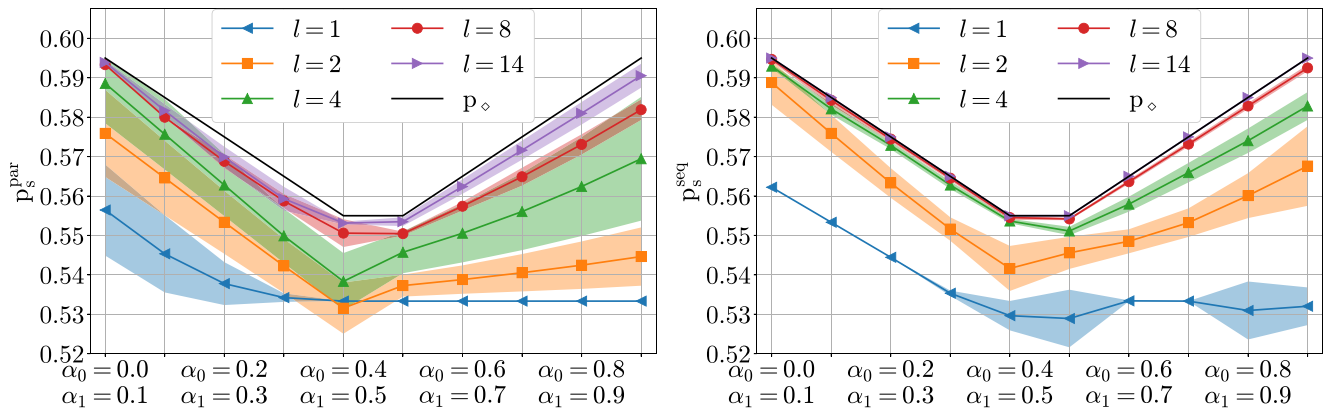

FIG. 7. Probability of successful discrimination $\mathrm{p_s}$ between a pair of depolarizing channels with α_0 and α_1 achieved with the parallel (left panel) and sequential (right panel) strategies. With different colors shown are the results for specific number of *Ansatz* layers l. Marks connected by solid lines stand for the average probabilities obtained after 10 independent runs, and shaded areas are to show the standard deviation. The black solid line indicates the maximum achievable probability for the parallel strategy $\mathrm{p}_\diamond^{\mathrm{par}}$ for $p = 2$ channel evaluations; for the sequential strategy, this line overlaps with the purple curve corresponding to $l = 14$.

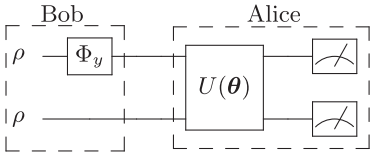

FIG. 8. A schematic of the variational classifier of quantum channels. Bob prepares the states $\Phi_y[\rho]\otimes\rho$ for Alice who then applies the unitary $U(\boldsymbol{\theta})$ and measures the resultant state in the computational basis. The measurement results are further used to compute the prediction value p as yielded by Eq. (20). To train the circuit, Alice minimizes the square distances (21) between the predictions p and the true labels y.

example, one might think of classifying phases of matter [42,67,68]. That is, having a variational classifier trained on labeled data points (quantum states of different phases), one may predict unknown labels of given states. In this section, we solve a similar problem, limiting ourselves to binary classification of the depolarizing channel (18) with two different depolarization factors α_0 and α_1.

Similarly to the variational channel discrimination discussed in Sec. II, the problem of building a variational classifier of quantum channels can be described in the form of a game between Alice and Bob. In this game, Alice wants to train a variational circuit $U(\boldsymbol{\theta})$ such that given an output state of a channel Φ_y from Bob, after postprocessing the results of measurements, this circuit allows to predict the label y of the channel applied. There are several peculiarities in the game we consider. First, Bob sends to Alice not only the output state $\rho_y=\Phi_y[\rho]$, but the original state ρ as well, i.e., from Bob Alice receives the states $\rho_y\otimes\rho$. Second, Alice does not control the original state ρ: it is prepared by Bob, and it is random and mixed. Third, the postprocessing of measurement results assumes that Alice is allowed to perform arbitrarily many measurements or, equivalently, Bob is assumed to give an arbitrary number of copies of the channel's output state and the corresponding original states. This game is schematically shown in Fig. 8.

Let us describe the game more formally. At the beginning, Bob picks two values $0\leqslant\alpha_0\neq\alpha_1\leqslant 1$. Bob then selects a label $y\in\{0,1\}$, creates two copies of a randomly generated (in general, mixed) qubit state ρ, and passes one of these states through a quantum channel giving $\rho_y=\Phi(\alpha_y)[\rho]$. The other copy of the state remains unaffected. Bob sends the state $\rho_y\otimes\rho$ to Alice who feeds this as an input for the variational circuit $U(\boldsymbol{\theta})$, and the resultant state is determined by

$$\rho(\alpha_y,\boldsymbol{\theta})=U(\boldsymbol{\theta})(\rho_y\otimes\rho)U^\dagger(\boldsymbol{\theta}). \tag{19}$$

Alice measures the observable $\sigma_z\otimes\sigma_z$ for calculating

$$p(\boldsymbol{\theta})=\tfrac{1}{2}(1+\mathrm{Tr}[\rho(\alpha_y,\boldsymbol{\theta})\,(\sigma_z\otimes\sigma_z)]), \tag{20}$$

that is used to quantify the prediction of the label y. The task of Alice is to train the circuit $U(\boldsymbol{\theta})$ such that given a pair $\rho_y\otimes\rho$ one is able to predict the label y of the depolarizing channel based on (20), i.e., for some $0<b<1$ Alice returns $y=0$ if $p\leqslant b$, and $y=1$ otherwise.

To train the circuit, Bob provides Alice with the training set $\{\rho_{y_j}^j\otimes\rho^j,y_j\}_{j=1}^{N_{\text{train}}}$ where $y_j\in\{0,1\}$ are true labels and the superscript in ρ^j indicates that each state is different since generated randomly. Then, Alice feeds each pair into the circuit $U(\boldsymbol{\theta})$ and calculates the predictions $p_j(\boldsymbol{\theta})$ as defined by (20). Having obtained $\{p_j(\boldsymbol{\theta}),y_j\}_{j=1}^{N_{\text{train}}}$, Alice makes use of the least-squares loss function to determine the optimal circuit parameters $\boldsymbol{\theta}$:

$$f(\boldsymbol{\theta})=\sum_{j=1}^{N_{\text{train}}}[y_j-p_j(\boldsymbol{\theta})]^2. \tag{21}$$

Knowing the true labels y_j and having sorted predictions p_j, Alice can determine the parameter b that separates two classes. Calculating can be based on maximizing the accuracy during training. The search of b is carried out iteratively. At each step t, Alice takes the prediction p_t and groups all the elements that are less than or equal to b to belong to the first class $(p_1,p_2,\ldots,p_t)\in\{\text{“0”}\}$ and to the second class $(p_{t+1},p_{t+2},\ldots,p_{N_{\text{train}}})\in\{\text{“1”}\}$ otherwise. As a result, b equals to the prediction value p_t, the division by which gives the best accuracy during the training.

Having found the optimal circuit parameters $\boldsymbol{\theta}^{\text{opt}}$ and the border value b, Alice must test the obtained classifier. To do that, Alice receives from Bob a set $\{\rho_{y_j}^j\otimes\rho^j\}_{j=1}^{N_{\text{test}}}$, which does not contain the true labels y_j. As during the training, Alice feeds each state $\rho_{y_j}^j\otimes\rho^j$ from the test set to the circuit $U(\boldsymbol{\theta}^{\text{opt}})$, computes the corresponding prediction values p_j in (20), and assigns the label $y_j=0$ if $p_j\leqslant b$ or $y_j=1$ if $p_j\geqslant b$.

To test the described approach of building a variational channel classifier, we performed numerical experiments of distinguishing the depolarizing channels (18) with different depolarization factors α_0 and α_1. To represent the variational circuit $U(\boldsymbol{\theta})$, we considered the three *Ansätze*:

$$U_1(\boldsymbol{\theta})=CR_y^{12}(\theta_7)[R_x(\theta_3)R_z(\theta_2)R_x(\theta_1)]\otimes[R_x(\theta_6)R_z(\theta_5)R_x(\theta_4)], \tag{22}$$

$$U_2(\boldsymbol{\theta})=[R_z(\theta_2)R_x(\theta_1)]\otimes[R_z(\theta_4)R_x(\theta_3)], \tag{23}$$

$$U_3(\boldsymbol{\theta})=R_z(\theta_2)R_x(\theta_1). \tag{24}$$

The first *Ansatz* $U_1(\boldsymbol{\theta})$, where $CR_y^{12}(\theta)$ represents a controlled Y rotation with the first control and the second target qubit, is up to the two-qubit gate essentially a hardware-efficient *Ansatz* of a single layer shown in Fig. 5. The second *Ansatz* $U_2(\boldsymbol{\theta})$ is a truncated realization of $U_1(\boldsymbol{\theta})$ with no entangling gate present. The classifiers built on the *Ansätze* $U_1(\boldsymbol{\theta})$ and $U_2(\boldsymbol{\theta})$ are two qubit, and they are trained on the pairs $\rho_{y_j}\otimes\rho^j$, as described in the beginning of the section. The third *Ansatz* $U_3(\boldsymbol{\theta})$ is of single-qubit structure, and in this case the classifier is trained on the states ρ_{y_j} that passed through a channel, without feeding in the original states ρ^j.

The results of our numerical experiments with the classifiers based on the *Ansätze* (22)–(24) are shown in Fig. 9. These results are obtained after training the classifiers on sets of the size $N_{\text{train}}=1000$ and tested on sets of the same size, $N_{\text{test}}=1000$. A close inspection of the plots suggests that even the U_3-based classifier that takes only the states ρ_{y_j} for training is capable of discriminating the quantum channels

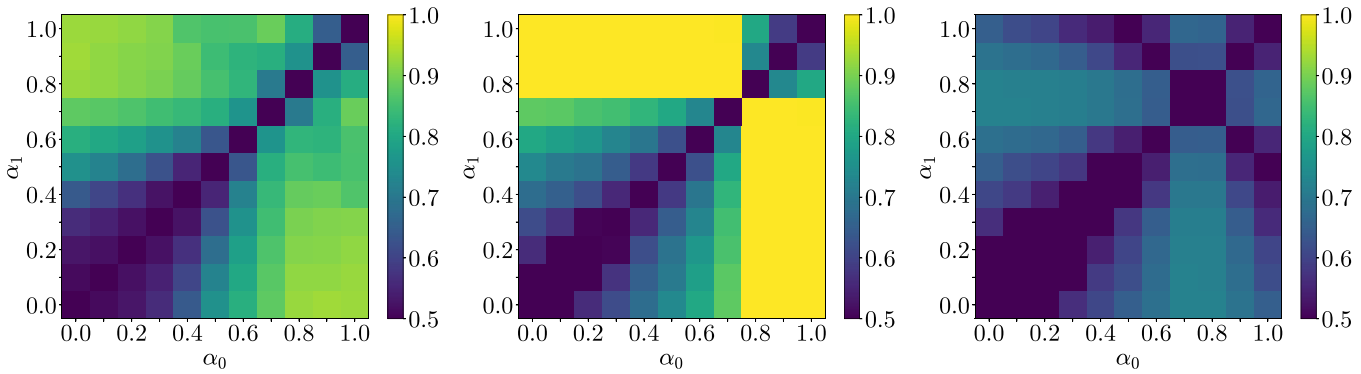

FIG. 9. The accuracy of quantum channel discrimination as obtained on the test set for depolarizing channels with α_0 and α_1. The left, center, and right panels show the accuracy of U_1-, U_2-, and U_3-based classifiers defined in (22)–(24), respectively. The training and test sizes are $N_{\text{train}} = N_{\text{test}} = 1000$.

with some accuracy. Among U_1 and U_2 classifiers which are trained on pairs $\rho_{y_j} \otimes \rho^j$, the better accuracy is achieved by the one that uses a simpler though less expressive *Ansatz* with no two-qubit gates. In our numerical simulations, as shown in Fig. 9 this U_2-based classifier unveils excellent accuracy in discriminating the channels with $\alpha_0 \lesssim 0.75 \lesssim \alpha_1$. Interestingly, the U_3-based classifier yields the highest degree of discrimination accuracy for the depolarization factors $\alpha = 0.7$ or 0.8, i.e., near 0.75. This agrees with the fact that extremum of the function

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

is reached at $\alpha = 0.75 - \epsilon/2$ for any $\rho \neq \mathbb{1}/2$, provided $0 \leqslant (\alpha + \epsilon) \leqslant 1$. This can be established by solving the equation $\partial_\alpha \text{Tr}(\rho_\alpha \rho_{\alpha+\epsilon}) = 0$ for α.

IV. KERNEL-BASED CLASSIFIER

An alternative way for discriminating quantum channels using a quantum processor can be traced back to the kernel methods which can be formulated as follows. Suppose we have a set of states $\mathcal{X} = \{\rho_i\}$. The kernel is essentially a function $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ that guarantees the Gram matrix $\mathcal{K}_{ij} = \mathcal{K}(\rho_i, \rho_j)$ to be positive semidefinite [37]. In particular, the trace of the product of density operators,

$$\mathcal{K}(\rho_i, \rho_j) = \text{Tr}(\rho_i \rho_j), \tag{25}$$

mentioned in previous sections, does possess such properties. The kernel-based classification methods are built on the so-called representer theorem [69]. One can think of supervised machine learning based on the support vector machine where the so-called kernel trick is widely utilized [70,71]. In this method, given a training set $\{\rho_j, y_j\}_{j=1}^{N_{\text{train}}}$ with labels $y_j \in \{-1, +1\}$, the cost function for maximization is

$$f(\boldsymbol{\theta}) = \sum_{i=1}^{N_{\text{train}}} \theta_i - \frac{1}{2} \sum_{i,j=1}^{N_{\text{train}}} \theta_i \theta_j \, \mathcal{K}(\rho_i, \rho_j) \, y_i y_j \tag{26}$$

with respect to $\boldsymbol{\theta} = \{\theta_j\}_{j=1}^{N_{\text{train}}}$, on condition that $\sum_{i=1}^{N_{\text{train}}} \theta_i y_i = 0$ and $\theta_i \geqslant 0$ [72]. Having found the optimal parameters $\boldsymbol{\theta}^{\text{opt}} = \arg\max_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$, one returns the labels based on the prediction function

$$p(\rho) = \sum_{i=1}^{N_{\text{train}}} \theta_i^{\text{opt}} y_i \, \mathcal{K}(\rho_i, \rho) + b, \tag{27}$$

where the bias b is defined by

$$b = \sum_{i=1}^{N_{\text{train}}} \theta_i^{\text{opt}} y_i \, \mathcal{K}(\rho_i, \rho_m) - y_m$$

for any m such that $\theta_m^{\text{opt}} > 0$. In binary classification, the class ones assigns to a given ρ is determined by $y = \text{sgn}[p(\rho)]$.

To formulate the problem of channel discrimination based on quantum kernel estimation, we again consider the game between Alice and Bob. Again, in this game Alice tries to discriminate two depolarizing channels of the form (18). However, this time the problem for Alice is harder: the channels will be associated not with fixed values of the depolarizing factors, but with ranges of it.

Let us formalize the game. First, Alice chooses an input state ρ^{in} and sends N_{train} copies of it to Bob. Then, Bob selects two intervals $\overline{\alpha}_{-1}$ and $\overline{\alpha}_{+1}$ such that $\overline{\alpha}_y \subset [0, 1]$. After that, Bob tosses a fair coin and attributes heads to $y = -1$ and tails to $y = +1$. Finally, Bob picks up a random $\alpha_y \in \overline{\alpha}_y$ and applies the depolarizing channel to one of Alice's states, $\rho_y = \Phi(\alpha_y)[\rho^{\text{in}}]$. As was mentioned, the class labels $y = \pm 1$ are attributed not to the specific values of the depolarization factor α, but to the fixed intervals of it. Having done that for all the states, Bob sends the training set $\{\rho_{y_j}^j, y_j\}_{j=1}^{N_{\text{train}}}$ to Alice who trains the classifier by maximizing the function (26). Here in $\rho_{y_j}^j$, the subscript $y_j \in \{\pm 1\}$ tells the interval the depolarization factor α is taken from, while the superscript j highlights that this factor is in general different for different input states ρ^{in} (recall that Bob picks $\alpha_y \in \overline{\alpha}_y$ randomly). Note that Bob does not tell Alice the intervals $\overline{\alpha}_y$ or the depolarizing factors α_y that were chosen, Alice knows only their true labels $y \in \{-1, +1\}$. To test the classifier, Alice sends N_{test} copies of the state ρ^{in} to Bob, who sends back the test set $\{\rho_{y_j}^j = \Phi(\alpha_{y_j}^j)[\rho^{\text{in}}]\}_{j=1}^{N_{\text{test}}}$ prepared similarly to the training one. For each state $\rho_j \equiv \rho_{y_j}^j$ of the test set, Alice calculates the prediction $p(\rho_j)$ as specified by (27) and assigns to this prediction the class label $y_j = \text{sgn}[p(\rho_j)]$. In practice,

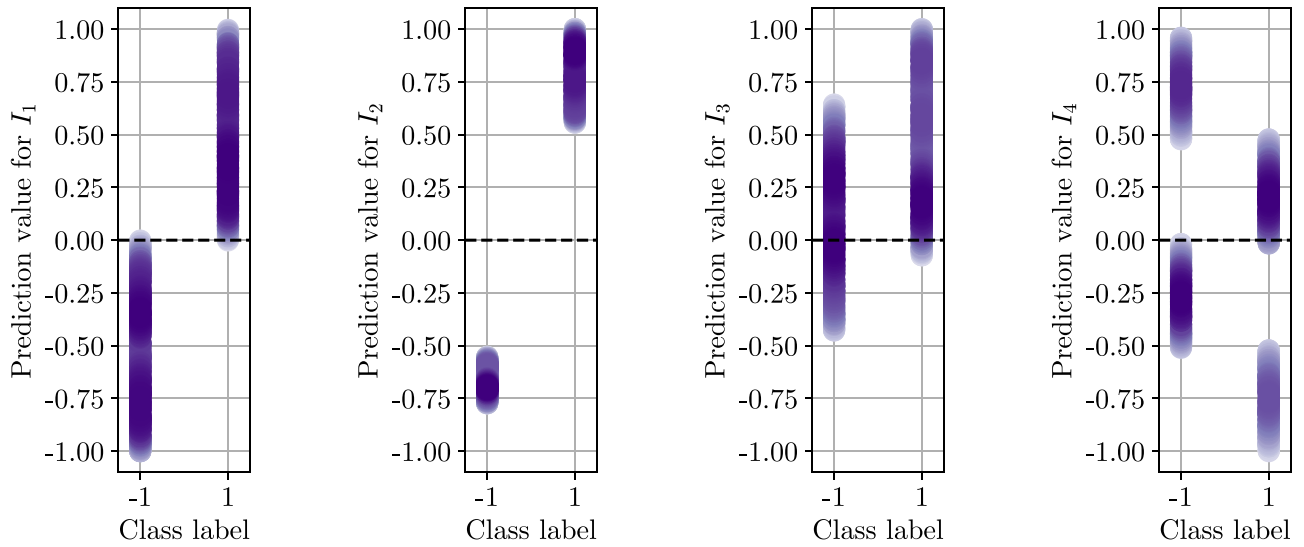

FIG. 10. The accuracy of quantum classifiers trained to discriminate depolarizing channels corresponding to the intervals of the depolarization factor I_1, I_2, I_3, and I_4 and defined by Eqs. (28)–(31), respectively. The corresponding classification accuracies are $A_1 = 0.999$, $A_2 = 1.0$, $A_3 = 0.67$, and $A_4 = 0.519$. The input state is $\rho^{\text{in}} = |+\rangle\langle+|$, and the sizes of the training and test sets are $N_{\text{train}} = 100$ and $N_{\text{test}} = 1000$. The vertical axis shows the normalized prediction value determined by (27). The color intensity features the density of data points.

Alice could estimate the kernel $\mathcal{K}(\rho_i, \rho_j) = \text{Tr}(\rho_i\rho_j)$ via the so-called controlled-SWAP test routine [73].

To test the kernel-based approach of classification, we performed numerical experiments by training such a classifier on sets of the size $N_{\text{train}} = 100$ and testing it on sets of the size $N_{\text{test}} = 1000$. In what follows, we consider four classifier instances trained for discriminating the channels with the following pairs of intervals of the depolarization factors α:

$$I_1 = \{\overline{\alpha}_{-1} = [0.0, 0.5),\ \overline{\alpha}_{+1} = [0.5, 1.0]\}, \tag{28}$$

$$I_2 = \{\overline{\alpha}_{-1} = [0.1, 0.2],\ \overline{\alpha}_{+1} = [0.7, 0.9]\}, \tag{29}$$

$$I_3 = \{\overline{\alpha}_{-1} = [0.0, 0.75],\ \overline{\alpha}_{+1} = [0.25, 1.0]\}, \tag{30}$$

$$I_4 = \{\overline{\alpha}_{-1} = [0.0, 0.25) \cup [0.5, 0.75),\ \overline{\alpha}_{+1} = [0.25, 0.5) \cup [0.75, 1.0]\}. \tag{31}$$

In I_1, the classifier is trained to recognize if a given state ρ_α is taken from $\alpha < 0.5$ or $\alpha \geqslant 0.5$. In I_2, the classes are chosen to comprise subsets of [0, 1] which do not overlap. The intervals I_3 are selected to test the performance of the classifier for intersecting regions. I_4 divides [0,1] into four parts such that the first and the third parts belong to the class $y = -1$ and the second and the fourth parts are marked by $y = +1$. The input state was set to $\rho^{\text{in}} = |+\rangle\langle+|$, and the cost function (26) was maximized using the SLSQP method which supports bounds and constraints [74].

The results of our numerical simulations as presented in Fig. 10 reveal that the classifier trained to discriminate the channels from I_1 and I_2 provides excellent accuracy. Moreover, the higher accuracy is achieved in case the regions $\overline{\alpha}_{-1}, \overline{\alpha}_{+1}$ are separated. In contrast, we expect *a priori* low accuracy for I_3. For example, the states ρ_{α_j} from $[0.25, 0.75] = \overline{\alpha}_{-1} \cap \overline{\alpha}_{+1}$ could happen to have different labels y for the same depolarization factor α. In this case, it turns out that the training performance is substantially dependent on the initial assignments of the parameters $\boldsymbol{\theta}$. Remarkably, in case of I_4 the regions $\overline{\alpha}_{-1}$ and $\overline{\alpha}_{+1}$ do not intersect which translates to the fact that the states ρ_j are expected to be classifiable, and yet the classifier fails. To remedy this issue, Alice could train the classifier on n copies of the output states $\rho_j^{\otimes n}$. This allows one to modify the kernel accordingly:

$$\mathcal{K}(\rho_i, \rho_j) = \text{Tr}\left(\rho_i^{\otimes n}\rho_j^{\otimes n}\right) = [\text{Tr}(\rho_i\rho_j)]^n. \tag{32}$$

The numerical results with this kernel are elaborated on in Appendix B.

V. DISCUSSION

To summarize, we discuss the approaches we used to solve the quantum channel discrimination problem. First, we did put the task into the framework of variational quantum computing paradigm. Namely, we stated the optimization problem of Eq. (1) in terms of optimizing the parameters of an *Ansatz* circuit [see (10) and (14)]. Potentially, this gives an opportunity for discriminating channels using noisy intermediate-scale quantum (NISQ) multiqubit systems [75]. In the context of variational quantum computing, we stressed out that the sequential strategy (7) is superior to the parallel strategy (4) which is in line with the previous studies. The sequential strategy with $p = 2$ channel applications allows one to perfectly discriminate the entanglement-breaking channels (15). In case of depolarizing channel (18), the sequential strategy still performs better (see Fig. 7), although the the total number of qubits to be used is the same for both methods.

Being reformulated in terms of the variational quantum computing, the parallel strategy with p-channel applications requires a quantum computer of p primary and r ancillary qubits, so that the total amount of qubits is $Q_{\text{par}} = p + r$. On the other hand, in the sequential strategy, $Q_{\text{seq}} = 1 + r$ qubits have to be provided, revealing thus no dependence on p.

Despite the advantage in the number of qubits, with growing p and r, the sequential strategy might be worse in terms of the number of optimization parameters $C = |\boldsymbol{\theta}|$. Indeed, if every unitary $U(\boldsymbol{\theta}_k)$ is a hardware-efficient *Ansatz* of l layers, it necessitates $C_{\text{seq}} = 3l(1+r)(p+1) \sim O(pr)$ parameters to optimize over. In contrast, in the parallel strategy, $C_{\text{par}} = 2 \times 3l(p+r) \sim O(p+r)$. That is, by choosing a strategy one trades quantum resources Q for classical resources C, and vice versa. Recall that for $p = 2$ our observation suggests that for the same total amount of qubits, the sequential strategy outperforms the parallel strategy.

In this work, we also addressed the quantum channel discrimination problem solved using a variational-circuit-based quantum classifier. It was mentioned that the best performance is achieved when the classifier is trained on the pairs of the original state ρ and its copy $\Phi[\rho]$ which passed through a channel. Inspired by the approach to quantum channel discrimination as realized with the use of parallel and sequential strategies, we attempted to train the classifier on the pairs of the state $\Phi[\rho]$ and the state $|0\rangle\langle 0|^{\otimes r}$, so that the variational circuit is a $(r+1)$-layered hardware-efficient *Ansatz*. Furthermore, we performed the training on the r copies of the state $\Phi[\rho]^{\otimes r}$. However, none of these two training ways results in a good performance. Remarkably, training such a quantum classifier with a simpler and less expressive *Ansatz* is advantageous. In principle, the circuit of seven parameters U_1 defined in (22) is capable of preparing any pure two-qubit state, showing yet worse performance compared to the one trained with the *Ansatz* U_2 in (23) containing no entangling gates. We tried to add a CX gate to this circuit, which increases its expressive power without introducing any optimization parameters. Still, the *Ansatz* gives a lower performance, which suggests that this fact cannot be attributed to overparametrization. Despite the assumption that we are given a pair of ρ and $\Phi[\rho]$ states, and we may, in principle, perform arbitrary number of measurements, this quantum-machine-learning approach is very powerful. First, the original input states may be random and even mixed. Second, one needs a circuit-based quantum computer of two qubits only and no entangling gates.

Kernel-based methods for quantum channel discrimination were also studied in this work. We deliberately considered a more complex task with the channels being specified by the intervals $\overline{\alpha}_y$ of the depolarization factor α and not of its fixed values α_y. The reason for this is that with fixed input states ρ^{in} and the two depolarizing channels with $\alpha_{\pm 1}$, it would be enough to have a training set of only $N_{\text{train}} = 2$ states, $\{\Phi(\alpha_{-1})[\rho^{\text{in}}], \Phi(\alpha_{+1})[\rho^{\text{in}}]\}$. Special attention should be paid to the case of discriminating the depolarizing channels corresponding to the intervals I_4 in (31), which divides the line $[0, 1]$ into four parts with assigning the class $y = -1$ to the first and third parts and the class $y = +1$ to the second and fourth parts. In principle, these classes are expected to be separable, and yet our classifier fails to do that. In Appendix B, we show this issue can be relaxed by modifying the kernel as $\mathcal{K}(\rho_i, \rho_j) = \text{Tr}(\rho_i^{\otimes n}\rho_j^{\otimes n}) = [\text{Tr}(\rho_i\rho_j)]^n$ for $n \in \mathbb{N}$, which is similar to the classical kernel $\mathcal{K}(x_i, x_j) = (|\boldsymbol{x}_i \cdot \boldsymbol{x}_j|^2)^n$ for $\boldsymbol{x} \in \mathbb{R}^d$ [76]. Moreover, this simple modification of the kernel allows one to use random and mixed input states ρ^{in} instead of the fixed $\rho^{\text{in}} = |+\rangle\langle +|$.

It should be stressed that machine-learning tasks based on quantum kernel estimation are classical quantum. That is, one first maps classical data points $\boldsymbol{x} \in \mathbb{R}^d$ into pure quantum states $|\boldsymbol{x}\rangle$ of a Hilbert space, for which the density operators are $\rho = |\boldsymbol{x}\rangle\langle\boldsymbol{x}|$ and the kernel reduces to $\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = |\langle \boldsymbol{x}_i|\boldsymbol{x}_j\rangle|^2$. Such transformation is called a feature map, and in its simplest form it is specified by

$$\boldsymbol{x} = \{x_i\}_{i=1}^d \longrightarrow |\boldsymbol{x}\rangle = \bigotimes_{i=1}^{d} [\cos(x_i/2)|0\rangle + \sin(x_i/2)|1\rangle].$$

In principle, one can suggest more efficient mapping scheme, meanwhile the necessity of encoding $\boldsymbol{x} \to |\boldsymbol{x}\rangle$ is considered as an important shortcoming of classical-quantum machine learning. However, in the task of quantum channel discrimination, the data points ρ are quantum and do not need to be encoded, although these quantum states are in general mixed.

Interestingly, the function $\mathcal{K}(\rho_i, \rho_j) = \text{Tr}(\rho_i\rho_j)$ seems to play an important role in the other considered approaches of channel discrimination. That is, for the approach of variational computing embedding, we found that the less expressive the *Ansatz* is (i.e., the fewer layers l it has) the more it correlates with the trace of the product (see Appendix A). In addition, for the variational quantum classifier, we observed that with a proper *Ansatz* the classification is perfect for the depolarization factors (α_0, α_1) such that $\alpha_0 \lesssim 0.75 \lesssim \alpha_1$. For the output states of the depolarizing channel $\rho_\alpha = \Phi(\alpha)[\rho]$, the point $\alpha = 0.75$ is the extremum of the function $\mathcal{K}(\rho_\alpha) = \text{Tr}\rho_\alpha^2$ with $\forall\ \rho \neq \mathbb{1}/2$ [or, more generally, the minimum of $\mathcal{K}(\rho_\alpha, \rho_{\alpha+\epsilon}) = \text{Tr}(\rho_\alpha\rho_{\alpha+\epsilon})$ is at $\alpha = 0.75 - \epsilon/2$] (see Appendix A for details). By this we suggest that while solving the quantum channel discrimination problem, one must pay attention not only to the diamond-norm distance between the target channels, but also to the trace of the product of their output states.

All the approaches considered in our study have their pros and cons, as well as different assumptions. The first approach, variational circuit embedding assumes that we are given a number of channel applications p and an arbitrary number of measurements in the training stage, but in the active stage only one measurement is allowed. On the other hand, the second approach, the variational quantum classifier assumes only a single-channel application, but requires to be trained on the pairs of the output and original states $\rho_\alpha \otimes \rho$, and also needs many measurements for estimating expectation values. But, at the same time the states ρ can be random and even mixed. The third approach which is based on quantum kernel estimation also requires many measurements for computing the kernel, but allows to discriminate parameter-dependent channels for different ranges of parameters that belong to different classes. As pointed out in Appendix B, this technique of quantum channel discrimination could be improved by training on n copies of input states, which is equivalent to raising the kernel to the power of n.

The data and code that support the findings of this study are available from A.S.K. upon reasonable request.

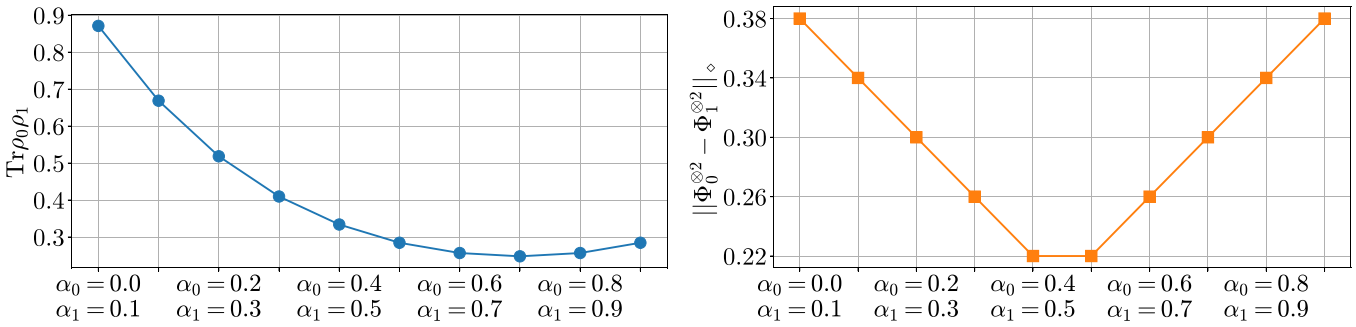

FIG. 11. Trace of the product (left) and the diamond distance (right) for the parallel discrimination strategy versus depolarization factors (α_0, α_1) for $p = 2$ channel applications. Herein, $\Phi_y \equiv \Phi(\alpha_y)$ is the depolarizing channel (18), and $\rho_y = \Phi(\alpha_y)[\rho]$.

ACKNOWLEDGMENTS

This work was supported in the framework of the Roadmap for Quantum computing (Contracts No. 868-1.3-15/15-2021 and No. P2163/11148). D.Y. acknowledges the support from the Russian Science Foundation Project No. 22-11-00074.

APPENDIX A: DEPENDENCE ON THE TRACE OF THE PRODUCT

While solving the quantum channel discrimination problem in the variational quantum computing framework [see (10) and (14) in the main text], one may expect that the performance significantly depends on the number of *Ansatz* layers l and on the diamond-norm distance between the channels which determines $\mathrm{p}_\diamond^{\mathrm{par}}$ as given in (5). Considering the depolarizing channels Φ_0 and Φ_1 in the form of (18) with the depolarization factors α_0 and α_1, it can be seen that the diamond distance $||\Phi_0^{\otimes p} - \Phi_1^{\otimes p}||_\diamond$ for $p = 2$ channel applications is symmetric (see Fig. 11). At the same time, in our numerical simulations, we observe that it is harder to achieve theoretical success probability $\mathrm{p}_\diamond^{\mathrm{par}}$ for the pairs of factors (α_0, α_1) which are on the right side to $\alpha = 0.5$ (see Fig. 7). Interestingly, this seems to correlate with the trace of the product of density operators $\mathrm{Tr}(\rho_0\rho_1)$, where $\rho_y = \Phi(\alpha_y)[\rho]$, $\forall\ \rho \neq \mathbb{1}/2$ (see Fig. 11). Moreover, it appears that the more *Ansatz* layers l one uses to maximize the success probability $\mathrm{p_s}$, the less the convergence properties depend on the trace of the product. In Fig. 12, this can be observed upon close inspection of the Pearson's coefficients.

APPENDIX B: MODIFIED KERNEL

In machine learning, the kernel is a complex- or real-valued function $\mathcal{K}(\rho_i, \rho_j)$ that produces a positive-semidefinite matrix $\mathcal{K}_{ij} = \mathcal{K}(\rho_i, \rho_j)$. Among various kernels considered in the domain of classical machine learning, the simplest one is $\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (|\boldsymbol{x}_i \cdot \boldsymbol{x}_j|^2)^n$ for some $n \in \mathbb{N}$ and classical data $\boldsymbol{x} \in \mathbb{R}^d$. When the data are quantum, i.e., ρ is a density operator on $\mathbb{C}^d$, a similar kernel is $\mathcal{K}(\rho_i, \rho_j) = \mathrm{Tr}(\rho_i^{\otimes n}\rho_j^{\otimes n}) = [\mathrm{Tr}(\rho_i\rho_j)]^n$ given in (32) in the main text.

In our numerical simulations, we can see that the performance of the kernel-based classifier may depend on n, the number of copies of the channel output states used to train the classifier. That is, in Fig. 10 one can notice that for the intervals

$$I_4 = \{\overline{\alpha}_{-1} = [0.0, 0.25) \cup [0.5, 0.75),$$
$$\overline{\alpha}_{+1} = [0.25, 0.5) \cup [0.75, 1.0]\}$$

mentioned in (31), the classifier trained on the single-copy states $(n = 1)$ fails to separate the classes. In Fig. 13, we

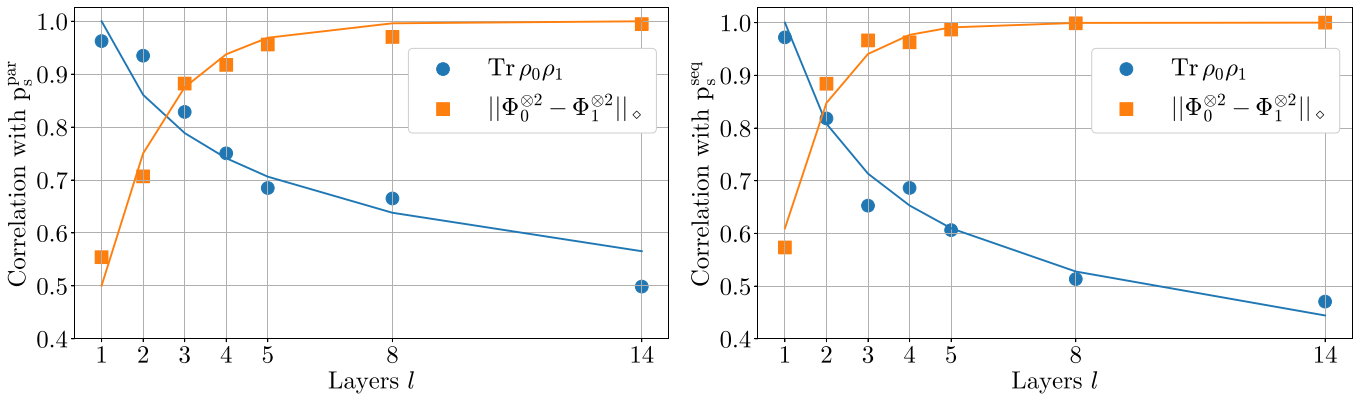

FIG. 12. The Pearson correlation coefficients between the trace of the product, diamond distance, and the average successful discrimination probabilities. On the left are the results for the parallel strategy, and on the right are for the sequential strategy. Each data point is obtained by averaging out five independent runs. By solid lines, data points are fitted by the functions of l, the number of layers of the hardware-efficient *Ansatz*. We used $f(l) = l^{-1/a}$ for fitting the trace of the product, and $g(l) = 1 - e^{-bl}$ for the diamond distance.

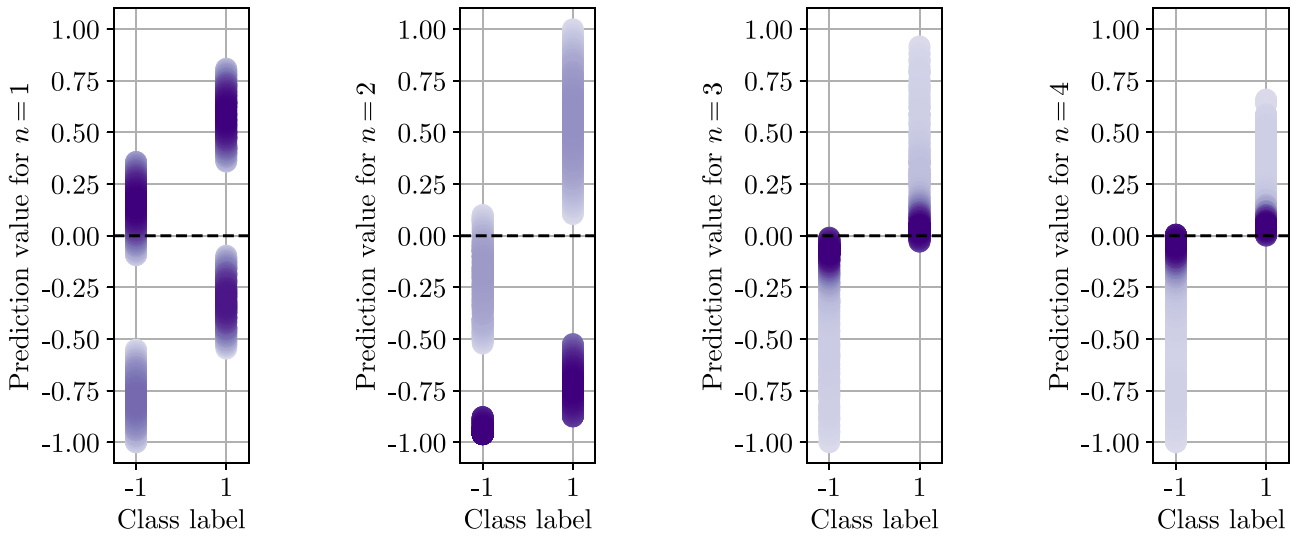

FIG. 13. The accuracy of the kernel-based classifier obtained on the test set for the intervals I_4 defined in (31). From left to right given are the accuracy for the classifiers based on the kernel (32) for different $n \in \{1, 2, 3, 4\}$. The corresponding classification accuracies are $A_1 = 0.522, A_2 = 0.708, A_3 = 0.941$, and $A_4 = 0.981$. The input state is $\rho^{\text{in}} = |+\rangle\langle+|$, and the sizes of the training and test sets are $N_{\text{train}} = 100$ and $N_{\text{test}} = 1000$. The vertical axis shows the normalized prediction value defined in (27). The color intensity features the density of data points.

show the results of classification for I_4 obtained with different numbers of state copies n. Apparently, for achieving the best classification accuracy in this case one should use $n = 4$ copies of the channel output states for training. In this case, one modifies the kernel $\mathcal{K}(\rho_i, \rho_j) = \text{Tr}(\rho_i \rho_j)$ such that it is just raised to the power of $n = 4$.

As mentioned in the main text, we also discovered that this modification of the kernel makes the classifier more powerful in terms of the allowed input states. That is, instead of $\rho^{\text{in}} = |+\rangle\langle+|$, the input state can be random and mixed, as for the variational quantum classifier we tested in our study. In Fig. 14, we show classification accuracy for the intervals

$$I_1 = \{\overline{\alpha}_{-1} = [0.0, 0.5),\ \overline{\alpha}_{+1} = [0.5, 1.0]\}$$

and random mixed input states $(\rho^{\text{in}})^{\otimes n}$ with $n \in \{1, 2, 3, 4\}$. As can be seen, with $n = 1$ the classifier fails to predict the labels when the input states are random, compared to the case when they are fixed [recall that the intervals I_1 are discussed in (28) and tested for the classifier with fixed input $\rho^{\text{in}} = |+\rangle\langle+|$, see Fig. 10].

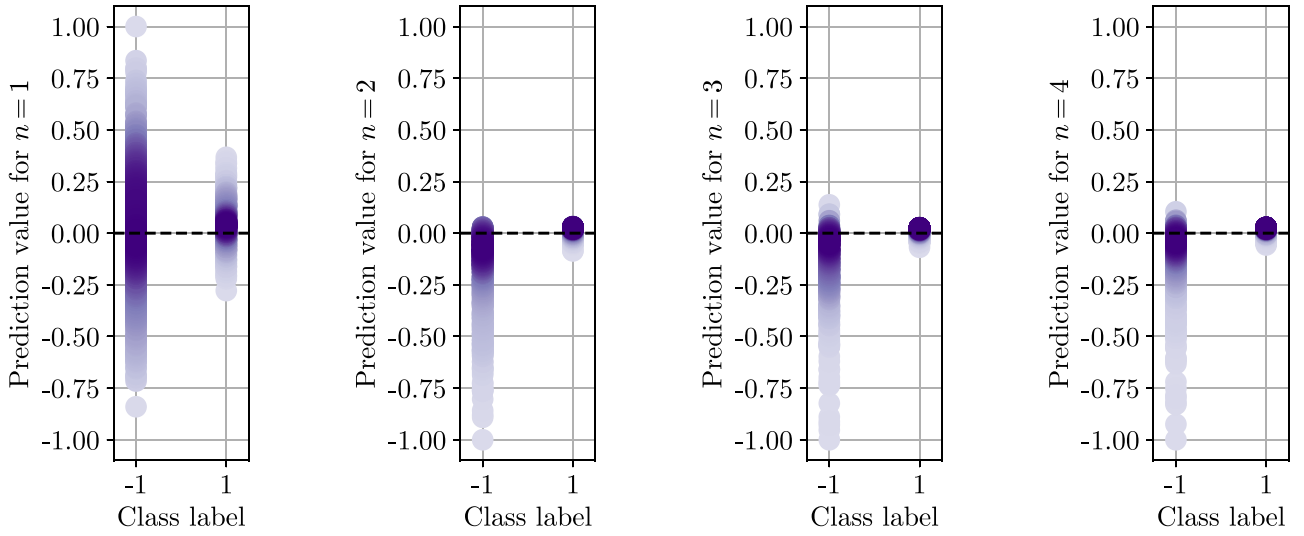

FIG. 14. The accuracy of the kernel-based classifier obtained on the test set for the intervals I_1 defined in (28). From left to right given are the accuracy for the classifiers based on the kernel (32) for different $n \in \{1, 2, 3, 4\}$. The corresponding classification accuracies are $A_1 = 0.595, A_2 = 0.847, A_3 = 0.869$, and $A_4 = 0.858$. The input states ρ^{in} are random and mixed, and the sizes of the training and test sets are $N_{\text{train}} = 100$ and $N_{\text{test}} = 1000$. The vertical axis shows the normalized prediction value defined in (27). The color intensity features the density of data points.

[1] A. Y. Kitaev, Russian Math. Surveys **52**, 1191 (1997).
[2] G. Wang and M. Ying, Phys. Rev. A **73**, 042301 (2006).
[3] R. Duan, Y. Feng, and M. Ying, Phys. Rev. Lett. **103**, 210501 (2009).
[4] A. Holevo, *Quantum Systems, Channels, Information: A Mathematical Introduction* (de Gruyter & Co, Berlin, 2012)
[5] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, Nat. Photonics **12**, 724 (2018).
[6] S. Lloyd, Science **321**, 1463 (2008).
[7] S. Barzanjeh, S. Guha, C. Weedbrook, D. Vitali, J. H. Shapiro, and S. Pirandola, Phys. Rev. Lett. **114**, 080503 (2015).
[8] G. Ortolano, E. Losero, S. Pirandola, M. Genovese, and I. Ruo-Berchera, Sci. Adv. **7**, eabc7796 (2021).
[9] H. E. Brandt, Am. J. Phys. **67**, 434 (1999).
[10] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).
[11] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, npj Quantum Inf. **6**, 20 (2020).
[12] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, Phys. Rev. A **81**, 032339 (2010).
[13] M. M. Wilde, in *2020 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, NJ, 2020), pp. 1915–1920.
[14] J. ur Rehman, A. Farooq, Y. Jeong, and H. Shin, Quantum Inf. Proc. **17**, 271 (2018).
[15] J. Bavaresco, M. Murao, and M. T. Quintino, Phys. Rev. Lett. **127**, 200504 (2021).
[16] T. Cooney, M. Mosonyi, and M. M. Wilde, Commun. Math. Phys. **344**, 797 (2016).
[17] V. Katariya and M. M. Wilde, Quantum Inf. Proc. **20**, 78 (2021).
[18] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, npj Quantum Inf. **5**50 (2019).
[19] J. L. Pereira and S. Pirandola, Phys. Rev. A **103**, 022610 (2021).
[20] Q. Zhuang and S. Pirandola, Phys. Rev. Lett. **125**, 080505 (2020).
[21] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, Lett. Math. Phys. **110**, 2277 (2020).
[22] V. Katariya and M. M. Wilde, Phys. Rev. A **104**, 052406 (2021).
[23] F. Salek, M. Hayashi, and A. Winter, arXiv:2011.06569.
[24] C. Harney and S. Pirandola, npj Quantum Inf. **7**, 153 (2021).
[25] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New J. Phys. **18**, 023023 (2016).
[26] J. Li, X. Yang, X. Peng, and C.-P. Sun, Phys. Rev. Lett. **118**, 150503 (2017).
[27] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone *et al.*, Sci. Adv. **4**, eaap9646 (2018).
[28] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, Nat. Commun. **5**, 4213 (2014).
[29] J. Biamonte, Phys. Rev. A **103**, L030401 (2021).
[30] M. Cerezo *et al.*, Nat. Rev. Phys. **3**, 625 (2021).
[31] R. LaRose, A. Tikku, É. O'Neel-Judy, L. Cincio, and P. J. Coles, npj Quantum Inf. **5**, 57 (2019).
[32] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, and D. Jaksch, Phys. Rev. A **101**, 010301(R) (2020).
[33] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature (London) **549**, 195 (2017).
[34] W. Huggins, P. Patil, B. Mitchell, K. B. Whaley, and E. M. Stoudenmire, Quantum Sci. Technol. **4**, 024001 (2019).
[35] M. Schuld, I. Sinayskiy, and F. Petruccione, Contemp. Phys. **56**, 172 (2015).
[36] M. Schuld and F. Petruccione, *Machine Learning with Quantum Computers* (Springer, Berlin, 2021).
[37] M. Schuld and N. Killoran, Phys. Rev. Lett. **122**, 040504 (2019).
[38] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Phys. Rev. A **98**, 032309 (2018).
[39] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Phys. Rev. A **101**, 032308 (2020).
[40] S. Y.-C. Chen, C.-H. H. Yang, J. Qi, P.-Y. Chen, X. Ma, and H.-S. Goan, IEEE Access **8**, 141007 (2020).
[41] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, arXiv:2001.03622.
[42] A. V. Uvarov, A. S. Kardashin, and J. D. Biamonte, Phys. Rev. A **102**, 012415 (2020).
[43] A. Uvarov, J. D. Biamonte, and D. Yudin, Phys. Rev. B **102**, 075104 (2020).
[44] A. Kardashin, A. Uvarov, D. Yudin, and J. Biamonte, Phys. Rev. A **102**, 052610 (2020).
[45] A. Kardashin, A. Pervishko, J. Biamonte, and D. Yudin, Phys. Rev. A **104**, L020402 (2021).
[46] M. Lazzarin, D. E. Galli, and E. Prati, Phys. Lett. A **434**, 128056 (2022).
[47] A. Patterson, H. Chen, L. Wossnig, S. Severini, D. Browne, and I. Rungger, Phys. Rev. Research **3**, 013063 (2021).
[48] H. Chen, L. Wossnig, S. Severini, H. Neven, and M. Mohseni, Quantum Machine Intelligence **3**, 1 (2021).
[49] R. Mengoni and A. Di Pierro, Quantum Machine Intelligence **1**, 65 (2019).
[50] S. Yu, F. Albarrán-Arriagada, J. C. Retamal, Y.-T. Wang, W. Liu, Z.-J. Ke, Y. Meng, Z.-P. Li, J.-S. Tang, E. Solano *et al.*, Adv. Quantum Technol. **2**, 1800074 (2019).
[51] M. Horodecki, P. W. Shor, and M. B. Ruskai, Rev. Math. Phys. **15**, 629 (2003).
[52] M. B. Ruskai, Rev. Math. Phys. **15**, 643 (2003).
[53] A. Shaham, A. Halevy, L. Dovrat, E. Megidish, and H. S. Eisenberg, Sci. Rep. **5**, 10796 (2015).
[54] A. Shaham and H. S. Eisenberg, Opt. Lett. **37**, 2643 (2012).
[55] G. Benenti and G. Strini, J. Phys. B: At., Mol. Opt. Phys. **43**, 215508 (2010).
[56] D. Puzzuoli and J. Watrous, *Annales Henri Poincaré* (Springer, Berlin, 2017), Vol. 18, pp. 1153–1184.
[57] W. Matthews, M. Piani, and J. Watrous, Phys. Rev. A **82**, 032302 (2010).
[58] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 060401 (2008).
[59] P.-H. Qiu, X.-G. Chen, and Y.-W. Shi, IEEE Access **7**, 50214 (2019).
[60] J. Biamonte, arXiv:1912.10049.
[61] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).
[62] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature (London) **549**, 242 (2017).
[63] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).
[64] A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd, J. Control Decision **5**, 42 (2018).

[65] S. Diamond and S. Boyd, J. Machine Learning Res. **17**, 1 (2016).
[66] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, SIAM J. Sci. Comput. **16**, 1190 (1995).
[67] A. A. Shirinyan, V. K. Kozin, J. Hellsvik, M. Pereiro, O. Eriksson, and D. Yudin, Phys. Rev. B **99**, 041108(R) (2019).
[68] A. Berezutskii, M. Beketov, D. Yudin, Z. Zimborás, and J. D. Biamonte, J. Phys. Complex. **1**, 03LT01 (2020).
[69] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, MA, 2018).
[70] I. Steinwart and A. Christmann, *Support Vector Machines* (Springer, New York, 2008).
[71] A. Rahimi and B. Recht, Adv. Neural Inf. Proc. Syst. **20**, 1 (2007).
[72] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature (London) **567**, 209 (2019).
[73] H. Kobayashi, K. Matsumoto, and T. Yamakami, in *International Symposium on Algorithms and Computation* (Springer, Berlin, 2003), pp. 189–198.
[74] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, Nat. Methods **17**, 261 (2020).
[75] J. Preskill, Quantum **2**, 79 (2018).
[76] M. Schuld, arXiv:2101.11020.