
Position: Quantum Kernel Machines Should Move Beyond Scalar-Valued Kernels to Realize Their Potential

Anonymous Authors¹

Abstract

Quantum kernels are reproducing kernel functions built using quantum-mechanical principles and have emerged as a centerpiece of quantum machine learning. The initial enthusiasm for quantum kernel machines has been tempered by recent studies suggesting that quantum kernels could not offer significant computational or statistical advantages when learning from classical data. However, most of the research in this area has been devoted to scalar-valued kernels in standard classification or regression settings for which classical kernel methods are efficient and effective, leaving very little room for improvement with quantum kernels. In this position paper, we argue that progress in this field requires moving beyond scalar-valued kernels toward more expressive kernel frameworks. Scalar-valued kernels lack the degrees of freedom necessary to fully exploit intrinsically quantum resources such as entanglement and are not rich enough to deal with complex learning tasks where classical learning methods struggle. Building on recent advances in operator-valued kernel learning and C^* -algebraic kernel representations, we propose a roadmap for designing quantum kernels capable of leveraging entanglement and non-commutative structures to tackle complex structured prediction problems. To support this viewpoint, we present an initial proof-of-concept illustrating how quantum entangled operator-valued kernel formulations can reveal structural dependencies that remain difficult to access for scalar-valued kernel methods. This shift in focus could open a pathway toward a new generation of quantum kernel machines and a more faithful exploration of their potential advantages.

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

1. Introduction

Quantum machine learning (QML) is an emerging field of research at the intersection between machine learning and quantum computing with the goal of using quantum computing paradigms and technologies to improve the speed and performance of learning algorithms (see Figure 1) (Biamonte et al., 2017; Du et al., 2025). Since the seminal works by Havlíček et al. (2019) and Schuld and Killoran (2019), quantum kernel machines have generated a lot of enthusiasm in this field, especially for exploring the applications of noisy intermediate-scale quantum computers to machine learning (Mengoni and Di Pierro, 2019; Blank et al., 2020; Wang et al., 2021; Heyraud et al., 2022). This enthusiasm has been tempered by recent studies that have suggested that quantum kernels could not offer speed-ups when learning on classical data (Huang et al., 2021; Kübler et al., 2021; Jerbi et al., 2023). However, these works have dealt with standard classification and regression tasks for which classical kernel methods are efficient and effective, leaving very little room for improvement with quantum kernels. We strongly believe that we have to focus on more complicated ML tasks where classical kernel methods face clear limitations in order to reveal the full potential of quantum kernels. This is the lens through which we will explore quantum kernel machines, relying on recent advances in classical ML. We present here our point of view on this ongoing field of research and argue that **quantum kernel research should move beyond the restrictive setting of scalar-valued kernels and shift its attention toward more expressive kernel frameworks to realize the full potential of quantum kernels.**

Operator-valued kernels (OVKs) generalize standard kernel functions and offer the possibility of tackling various ML problems ranging from multitask learning to multi-view learning and differential equations modeling (Michelli and Pontil, 2005; Alvarez et al., 2012; Evgeniou et al., 2005; Kadri et al., 2016; Lim et al., 2015; Huusari et al., 2018; Stepaniants, 2023). In this paper, we advocate for the adoption of operator-valued kernels for QML. The operator-valued kernel framework is flexible and gives rise to more expressive quantum feature spaces. We also discuss C^* -algebra-valued kernels as a generalization of OVKs. This could open up a new way to apply mathematical theories

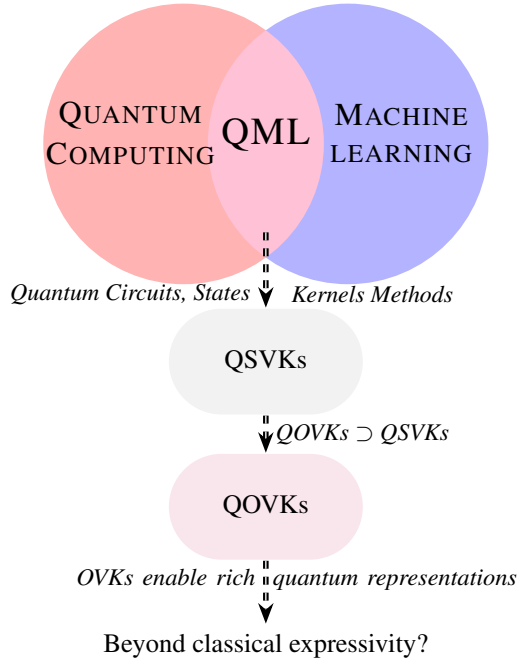


Figure 1. Quantum machine learning (QML) is a recent field of research at the intersection between quantum computing (QC) and machine learning (ML). The interaction is two-sided: quantum-enhanced ML (from QC to ML), and ML-based quantum computing (from ML to QC). Most of the interest has concentrated on the use of quantum computing paradigms to improve machine learning algorithms. Quantum scalar-valued kernels (QSVKs) have generated a great deal of interest in the field of QML due to their natural alignment with the kernel trick and their compatibility with hybrid quantum-classical architectures. However, recent findings suggest that their expressive power may be limited in classical data regimes. Quantum operator-valued kernels (QOVKs) offer a more general and expressive framework, potentially unlocking richer hypothesis spaces that are inaccessible to classical or quantum SVKs. QOVKs generalize QSVKs and provide opportunities to go beyond classical expressivity.

to QML for more complicated tasks such as structured prediction. Structured output learning is the task of learning a mapping between objects of different nature that each can be characterized by complex data structures such as curves, trees and graphs (Tsochantaridis et al., 2005; Geurts et al., 2006; Kadri et al., 2013; Brouard et al., 2016). Quantum structured prediction should extend the application scope of quantum kernels. To substantiate our viewpoint, we outline a possible quantum implementation of OVKs and provide a proof-of-concept demonstration on quantum channel estimation, highlighting how operator-valued formulations can outperform traditional scalar-valued kernels.

Notation The so-called ‘bra-ket’ notation is used to describe the state of a quantum system. A column vector ψ is represented as ‘ket’ $|\psi\rangle$. The conjugate transpose (Hermitian transpose) of a ket, a row vector, is denoted by ‘bra’

$\langle\psi| := |\psi\rangle^\dagger$, where \dagger denotes the conjugate transpose. The inner product of two vectors $|\psi_1\rangle$ and $|\psi_2\rangle$ can be written in bra-ket notation as $\langle\psi_1|\psi_2\rangle$, and their tensor product, can be expressed as $|\psi_1\rangle|\psi_2\rangle$, i.e., $|\psi_1\rangle \otimes |\psi_2\rangle$.

2. Kernels from Classical to Quantum

Machine learning is the branch of artificial intelligence that seeks to develop computer systems which detect patterns in data in order to improve their performance automatically with experience (Jordan and Mitchell, 2015). The widespread development of acquisition tools together with increasing storage capacities has had us witness an explosion in the amount of available data, as well as the urge to develop methods to handle them properly. In this context, it is crucial to design new large-scale machine learning systems that are able to deal with big data.

Quantum machine learning is a relatively recent field of research (Biamonte et al., 2017; Ciliberto et al., 2018; Dunjko and Briegel, 2018). This research field is largely driven by the desire to develop artificial intelligence that uses quantum technologies to improve the speed and performance of learning algorithms. There is also interest in investigating the use of machine learning for tackling quantum computing and quantum information problems. The field is evolving rapidly, but many open questions remain and should be addressed to better understand how quantum computers may outperform classical computers on machine learning tasks.

This paper identifies potential effective interactions between the fields of quantum computing and kernel machines and lays the ground towards a deeper understanding of what kernel-based learning looks like in a quantum world.

2.1. Classical Kernel Methods

The research work described in this paper belongs to a large class of learning algorithms, the so-called kernel methods. Since they were introduced by Boser et al. (1992) as a way to construct a nonlinear extension of Support Vector Machines, these methods become very popular. Kernel methods exploit training data through implicit definition of a similarity between data points that can be expressed as a dot product in a feature space, namely reproducing kernel Hilbert space (RKHS). The idea is to transform the data via a feature map ϕ into a suitable feature space in which linear learning algorithms could be applied. The inner product between features can be computed using the kernel function; this is the well-known kernel trick, i.e., $k(x, y) = \langle\phi(x), \phi(y)\rangle$ (see Figure 2). It should be pointed out that the notion of kernels as dot products in Hilbert spaces was first brought to the field of machine learning by Aizerman et al. (1964), while the theoretical foundation of reproducing kernels and their Hilbert spaces dates back to at least Aronszajn (1950). Kernel meth-

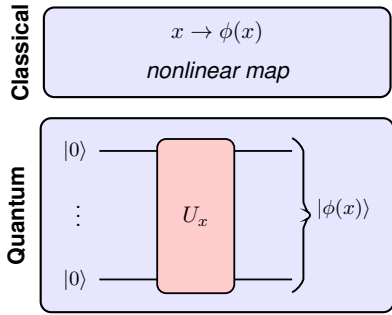


Figure 2. Kernel feature map. (top) In the classical setting, a data point x is mapped to a high-dimensional space via an implicit nonlinear feature map ϕ . The mapping ϕ may be unknown but the inner product between two data points, x and y , mapped by ϕ is equal to the kernel function evaluated at x and y , i.e., $\langle \phi(x), \phi(y) \rangle = k(x, y)$. (bottom) In the quantum setting, the feature map is known explicitly. Encoding a data point x into a quantum state $|\phi(x)\rangle$ using a unitary matrix (quantum gate) U_x defines a quantum feature map. A quantum kernel is then defined as the fidelity between two data-encoding quantum feature states, i.e., $k(x, y) = |\langle \phi(x) | \phi(y) \rangle|^2$.

ods became a mature field able to address many problems in machine learning and statistical data analysis (Hofmann et al., 2008). The tremendous achievements in the field show that these methods provide elegant and powerful learning algorithms for analyzing nonlinear features and processing complex data structures, and offer a comprehensive suite of mathematically well-founded nonparametric modeling techniques for a wide range of learning problems. One major limitation of kernel methods is their high computational cost when the number of training examples is large. This motivates the study of the impact of quantum computation in their computational capabilities. This is of importance since it can give rise to novel and effective strategies to scale up kernel methods for large-scale problems.

2.2. Quantum Kernels

A major difference between quantum computing and its classical counterpart is that information is carried by qubits (Nielsen and Chuang, 2010; De Wolf, 2019). Unlike bits which have only two possible states, 0 and 1, qubits can exist in those and any combination of them. More formally, a qubit is an element of a 2-dimensional Hilbert space \mathcal{H} , a complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product. An arbitrary qubit $|\psi\rangle$ may be written as $a_0 |\psi_0\rangle + a_1 |\psi_1\rangle$, where $|\psi_0\rangle$ and $|\psi_1\rangle$ form a complete orthonormal basis of \mathcal{H} . Multiple qubit states can be obtained from the tensor product of qubit states. A quantum machine learning algorithm needs data in the form of quantum states. Therefore, classical data should be first encoded into quantum states, i.e., the transformation of a classical

data x to a quantum state $|\phi(x)\rangle$. Most of the interest in quantum kernels comes from the observation that encoding classical data into a quantum computer defines an explicit feature representation of the data (see Figure 2). Moreover, all operations that can be performed on quantum feature states are linear. This draws a parallel with classical kernel machines (Havlíček et al., 2019; Schuld and Killoran, 2019). Using the kernel trick, a quantum kernel is then defined as the fidelity between two data-encoding feature states, i.e., $k(x, y) = |\langle \phi(x) | \phi(y) \rangle|^2$. Different data-encoding strategies and their quantum kernels have been proposed in the literature (see, e.g., Schuld and Killoran 2019).

2.3. The Good, the Bad and the Ugly

Quantum kernels have generated a lot of interest in the field of QML (Blank et al., 2020; Coyle et al., 2020; Kusumoto et al., 2021; Glick et al., 2024). The analogy between quantum data encoding and kernel feature representation provides a conceptual framework for understanding and analyzing quantum machine learning algorithms. This sheds light on the synergies between kernel machines and quantum computing and leads to more interaction between the two fields, whether theoretical or practical. Also, quantum kernels provide a scheme with which to realize hybrid quantum-classical learning. A quantum computer can be used to create feature representations and compute quantum kernels which are then fed into classical learning algorithms (Liu et al., 2021b). This makes them suitable for the noisy intermediate-scale quantum (NISQ) era (Preskill, 2018), where quantum computation has to be performed with limited quantum resources.

From another point of view, this analogy is too narrow to support a quantum advantage for machine learning. Some recent studies have argued that supervised quantum machine learning models are kernel methods (Schuld, 2021) or showed that random Fourier features, a widely known method for kernel function approximation, are able to classically approximate variational quantum machine learning (Landman et al., 2023). This leads to the question of whether quantum advantage is the right goal for quantum machine learning (Schuld and Killoran, 2022). Moreover, most of quantum data encoding strategies result in kernel functions that are already known and/or efficiently computable by a classical computer. Whether there are interesting kernel functions that can be computed via quantum states and are classically intractable is still an open question.

More problematic is the generalization ability of quantum ML methods based on kernel functions. Recent studies analyzed generalization error bounds for learning with quantum kernels and the results appear to be negative (Huang et al., 2021; Kübler et al., 2021). The expressive power of quantum models may hinder generalization. Finding suitable

quantum kernels is not easy because the kernel evaluation might require exponentially many measurements. In other words, when using a large number of qubits, the kernel matrix (i.e., the matrix obtained by evaluating the kernel function on all pairs of data points) gets close to the identity matrix, resulting in overfitting and poor generalization performance (Suzuki et al., 2022).

2.4. To Be or Not to Be

Most of the previous studies have focused on supervised learning of scalar-valued functions in the context of standard classification or regression. Classical kernel machines in this setting are well-established models and have been extensively studied in the last three decades. Efficient kernel approximations with randomization techniques have been proposed to reduce their computation and storage requirements while performing quite well in various applications (Liu et al., 2021a). Moreover, deep learning, which finds its root in the field of neural networks, has enabled tremendous progress for learning on various types of datasets, such as image, language or audio datasets, and achieved impressive performance on classification and regression tasks (LeCun et al., 2015). All these do not leave much space for improvement with quantum kernels in the context of standard supervised learning.

In this paper, we propose moving beyond current quantum kernel models toward operator-valued kernels, enabling a shift from simple kernel similarities to a new generation of quantum kernel machines capable of tackling challenging learning tasks. To realize this shift, we advocate for a research agenda centered on three pillars: (i) the development of quantum operator-valued kernels (QOVKs) that leverage entanglement as a computational resource, (ii) the adoption of C^* -algebraic frameworks to exploit non-commutative data structures, and (iii) a strategic focus on quantum structured prediction. This roadmap integrates advances in statistical kernel theory and structured output learning with quantum computing and information.

3. Moving Beyond Scalar-Valued to Operator-Valued Quantum Kernels

Classical OVKs Operator-valued kernels (OVKs) appropriately generalize the well-known notion of reproducing kernels and provide a means for extending the theory of reproducing kernel Hilbert spaces from scalar- to vector-valued functions. They were introduced as a machine learning tool by Michelli and Pontil (2005) and have since been investigated for use in various machine learning tasks, including multitask learning (Evgeniou et al., 2005), functional or operator regression (Kadri et al., 2016), multiview learning (Huusari et al., 2018), and PDE (partial differential equations) learning (Stepaniants, 2023). Despite this

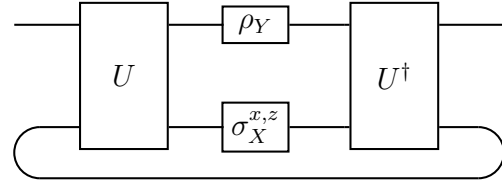


Figure 3. Diagram representation of the quantum operator-valued kernel (1). Input data are embedded into a feature matrix $\sigma_X^{x,z}$. We dilate the input system by tensorizing $\sigma_X^{x,z}$ with an output density matrix ρ_Y to form a larger composite system that enables interactions between input and outputs. A (unitary) interaction is then applied via U , resulting in the evolution of the composite input-output system. By tracing out the input system, we obtain the final state on outputs, which defines the value of the kernel function ($K(x, z)$ is a matrix acting on outputs).

progress, the current status of the field of operator-valued kernels suggests further explorations to shed fresh light on old questions, frame new ones and potentially offer new alternatives to existing kernel machines. For more details on (classical) operator-valued kernels and their associated reproducing Hilbert spaces, see Appendix A.

From classical to quantum OVKs A major limitation of operator-valued kernels is their high computational expense. In contrast to the scalar-valued case, the kernel matrix associated to a reproducing operator-valued kernel is a block matrix of dimension $np \times np$, where n is the number of data samples and p the dimension of the output space. Manipulating and inverting matrices of this size become particularly problematic when dealing with large n and p . Moreover, questions on how to design operator-valued kernels, what sort of interactions they should learn and quantify and how they should learn them from data are still open. Quantum-based kernel machines represent a promising approach for addressing these challenges. The kernel matrix plays a central role in solving kernel learning problems. For many ML tasks, computing the solution involves matrix-vector operations and solving optimization problems. Quantum linear system solvers (Morales et al., 2024) and quantum optimization algorithms (Abbas et al., 2024) hold potential for addressing these tasks more efficiently than their classical counterparts. The kernel matrix in the operator-valued setting can be significantly larger than its scalar-valued counterpart. This suggests that quantum algorithms could have a more substantial impact in the operator-valued setting, although the implementation of quantum solvers on current quantum devices still faces many challenges.

Quantum operator-valued kernels (QOVKs) have not been investigated yet. Building on the framework introduced by Huusari and Kadri (2021), we extend the notion of operator-valued kernels to the quantum domain, introducing the concept of *entangled quantum operator-valued kernel* (QOVK).

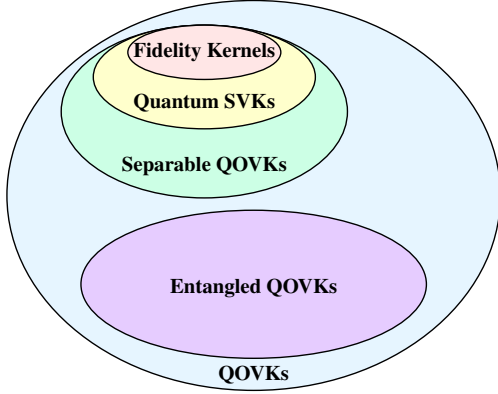


Figure 4. Illustration of inclusions among quantum kernel classes considered in this paper. Entangled quantum operator-valued kernels (QOVKs) are not separable, i.e., dependencies between input and output variables cannot be considered separately. The class of separable QOVKs coincides with the class of quantum scalar-valued kernels (QSVKs) when the output dimension is equal to one. If in addition the input feature matrix takes the form of a product of two pure density matrices, the separable QOVKs class becomes the fidelity kernel class.

Definition 3.1. (Entangled QOVK)

An entangled quantum operator-valued kernel $K : \mathbb{C}^d \times \mathbb{C}^d \rightarrow \mathbb{C}^{p \times p}$ is defined, $\forall x, z \in \mathbb{C}^d, p > 1$, as

$$K(x, z) = \text{Tr}_X \left[U_{YX} (\rho_Y \otimes \sigma_X^{x,z}) U_{YX}^\dagger \right], \quad (1)$$

where $U_{YX} \in \mathbb{C}^{pm \times pm}$ is a non-separable unitary matrix (i.e., a unitary matrix that cannot be written as $U_{YX} = A_Y \otimes B_X$, with $A \in \mathbb{C}^{p \times p}$ and $B \in \mathbb{C}^{m \times m}$), p is the dimension of output data, and m is the dimension of input features. In addition, $\rho_Y \in \mathbb{C}^{p \times p}$ is a density matrix on the output space, $\sigma_X^{x,z} \in \mathbb{C}^{m \times m}$ is a feature matrix extracted from inputs x and z , and Tr_X is the partial trace on X .

A diagram representation of entangled quantum operator-valued kernel is given in Figure 3.

Remark 3.2. It is worth noting that when U_{YX} is separable and equals to $I \otimes B_X$ the kernel $K(x, z)$ in (1) simplifies to a *separable quantum operator-valued kernel* computed using the scalar-valued quantum kernel $\text{Tr}(\sigma_X^{x,z})$, i.e., $K(x, z) = \text{Tr}(\sigma_X^{x,z}) \rho_Y$ (recall that B_X is a unitary matrix).

Remark 3.3. When the output dimension p is equal to one, the class of separable QOVK coincides with the class of *quantum scalar-valued kernels*. Moreover, if $\sigma_X^{x,z} = \rho_X^x \rho_X^z$, where ρ_X^x and ρ_X^z are pure density matrices (i.e., $\rho_X^x = |\phi(x)\rangle \langle \phi(x)|$ and $\rho_X^z = |\phi(z)\rangle \langle \phi(z)|$), we recover the class of *fidelity kernels* illustrated in Figure 2.

An illustration of inclusions among the quantum kernel classes discussed above is provided in Figure 4. It is easy to see that scalar-valued kernels can be recovered from operator-valued kernels by considering a separable kernel

built using a scalar-valued quantum kernel on inputs and a density matrix on outputs, i.e., $K(x, z) = k(x, z) \rho_Y$. This results in a kernel matrix G of the form $g \otimes \rho_Y$, where g is the scalar-valued kernel matrix. If we restrict ourselves to this class of kernels, QOVKs will suffer from the same limitations as QSVKs. The crucial question is whether alternative classes of OVKs might provide better mechanisms for addressing these limitations.

Are QOVKs a ‘magic bullet’ for quantum kernel machines? At present, there is no definitive answer. One might reasonably suspect that QOVKs inherit some of the limitations of quantum scalar-valued kernels, which could raise doubts about their usefulness. However, as operator-valued kernels strictly generalize scalar-valued ones, they offer additional structure and degrees of freedom whose impact remains largely unexplored. Determining whether, and to what extent, QOVKs overcome the limitations of QSVKs therefore calls for a systematic investigation.

The scalar-valued kernel framework may be too restrictive to effectively demonstrate the potential advantage of quantum kernel machines over classical kernel methods. In other words, the SVK framework lacks sufficient degrees of freedom to effectively address the challenges and exploit the unique capabilities of quantum kernels. QSVKs are a special case of QOVKs which correspond to simple, separable QOVKs. Entanglement can be a key resource for achieving quantum advantage (Jozsa and Linden, 2003; Wang et al., 2024). The framework of operator-valued kernels provides a means to investigate entanglement and incorporate it into the kernel learning process. This allows us to explore richer and more complex functions that could be learned more efficiently using quantum computation. Moreover, a key advantage of operator-valued kernels is their ability to naturally integrate input data from multiple modalities and targets from multiple tasks. This significantly improves and expands the application potentials of quantum kernels.

Entangled QOVKs, unlike separable kernels, do not have a Kronecker product structure and can be constructed using quantum correlations. This may open the door for the design of quantum kernels which can be implemented quantumly much more efficiently than classically. Entanglement can play a role in speeding-up quantum computation (Jozsa and Linden, 2003) and OVKs offer a framework for identifying how entanglement may contribute to achieving quantum advantage in kernel-based learning. On the other hand, operator-valued kernels naturally incorporate more data structure than scalar-valued kernels. Adding structure could improve generalization performance and is a well known technique for mitigating overfitting when enhancing expressivity. Entanglement can also have an impact on the number of measurements (Nakhl et al., 2024), which could offer new possibilities for the generalization of quantum kernels.

4. Call to Action

To enable the transition from scalar-valued to more expressive operator-valued quantum kernel machines, we propose the following actionable steps.

Action 1: Quantum implementation of OVKS Extending quantum data-encoding schemes and providing quantum circuit implementations of OVKS is of great importance to be able to characterize how well these kernels fit into a quantum computer. Quantum states can be represented by density operators, which are positive semi-definite, self-adjoint operators with unit trace. Identifying and exploring synergies between density operator formalism and OVKS is an interesting path to investigate. Furthermore, quantum superposition, a fundamental concept in QC, is the means by which quantum algorithms like Grover’s search can outperform classical ones (Grover, 1996). The objective is also to design quantum algorithms based on superposition for learning with OVKS in order to provide non-trivial improvements in terms of not only their computational complexity but also their statistical efficiency (Roget et al., 2022).

Action 2: Quantum entangled OVKS Some classes of operator-valued kernels have been proposed in the literature, with separable kernels being one of the most widely used for learning vector-valued functions due to their simplicity and computational efficiency. These kernels are formulated as a product between a kernel function for the input space alone, and a matrix that encodes the interactions among the outputs. However, there are limitations in using separable kernels. They use only one output matrix and one input kernel function and thus cannot capture different kinds of dependencies and correlations, and assume a strong repetitive structure in the operator-valued kernel matrix that models input and output interactions. As discussed above, *entangled* QOVKS go beyond separable kernels and offer new opportunities for quantum kernel design. This class should be better investigated to shed light on its potential in finding correlations that cannot be described by classical statistics.

Action 3: A C^* -algebraic detour C^* -algebras provide a unified framework for an operational formulation of classical and quantum mechanics (Bru and de Siqueira Pedra, 2023). Reproducing kernel Hilbert C^* -module (RKHM) is a generalization of reproducing kernel Hilbert space (RKHS) by means of C^* -algebra (Hashimoto et al., 2021). Recently, Hashimoto et al. (2023b;a) have paved the way for supervised learning in RKHMs. This provides a new twist to the state-of-the-art kernel-based learning algorithms and the development of a novel kind of reproducing kernels. Advantages of RKHM over RKHS are that we can make use of: i) the C^* -algebra characterizing the RKHM to construct rich feature representations and explore a larger function space (Hashimoto et al., 2023b), and ii) the properties of C^* -algebras such as operator norm and spectral truncation

to achieve better generalization and design kernels that control local and global dependencies of output data on input data (Hashimoto et al., 2023a; 2024).

From the perspective of the connection with quantum, C^* -algebra has rich notions related to quantum mechanics. For example, we can represent quantum gates and density operators as elements of a C^* -algebra. Thus, we can obtain them as outputs of the kernel machines with RKHMs. While the application of C^* -algebra to QML is a promising way to design quantum kernels, we need further investigations to relate theory with practice. It would be interesting to : i) Investigate connections between C^* -algebra-valued kernels and quantum information with a particular attention to learning in RKHM quantum systems (Gebhart et al., 2023); ii) Study the impact of quantum computing on the computational complexity of learning in RKHM.

It is worth noting that solving learning problems with such kernels involves tackling optimization problems over noncommutative groups. Noncommutative optimization (Burgdorf et al., 2016; Bürgisser et al., 2019) is a promising direction to explore within this framework since it makes connections with both noncommutative kernels (Belinschi et al., 2023; Hashimoto et al., 2024) and quantum information (Gribling et al., 2018). It will be interesting to investigate whether noncommutative optimization techniques can improve learning with quantum OVKS.

Action 4: Application to quantum structured prediction Operator-valued kernels hold promise to expand the application realm of quantum kernels. In many practical problems such as network inference (Lim et al., 2015) and graph prediction (Shi et al., 2020), we are faced with the task of learning a mapping between objects of different nature that each can be characterized by complex data structures (Bakir et al., 2007). Therefore, designing algorithms that are sensitive enough to detect structural dependencies among these complex data is of great importance. While classical learning algorithms can be easily extended to complex inputs, more refined and sophisticated algorithms are needed to handle complex outputs. In this case, several mathematical and methodological difficulties arise and these difficulties increase with the complexity of the output space. The task of structured output learning is much more complicated than binary/multiclass classification or scalar-valued regression, which leaves room for improvement.

One difficulty encountered when working with structured data is that usual Euclidean methodology cannot be applied in this case. Reproducing kernels provide an elegant way to overcome this problem. Defining a suitable kernel on the structured data allows to encapsulate the structural information in a kernel function and transform the problem to a Euclidean space. Kernel-based approaches for structured output learning can be found in the literature (Tsochantaridis

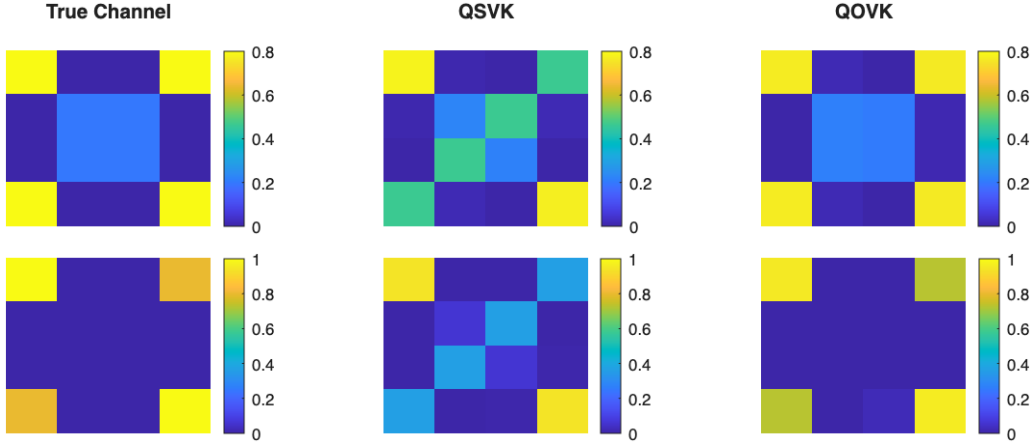


Figure 5. Bit-flip channel (first row) and dephasing channel (second row) estimation results with $a = 0.2$, $n = 100$ and $\alpha = 0.1$. Left plot is the Choi matrix of the correct channel from which data was created. Middle plot is the Choi matrix of the quantum channel recovered by QSVK (separable). Right plot is the Choi matrix of the quantum channel recovered by QOVK (entangled).

et al., 2005; Geurts et al., 2006; Kadri et al., 2013; Lim et al., 2015; Brouard et al., 2016; El Ahmad et al., 2024). These methods generally require an exhaustive pre-image computation (Honeine and Richard, 2011). Of special interest is supervised learning when input and output data are graph-structured. This is a complicated task that appears in various practical applications such as graph link prediction (Zhang and Chen, 2018). Graph kernels have received a lot of attention in the field of machine learning (Borgwardt et al., 2020). A few attempts have been made to introduce quantum kernels on graphs (Bai et al., 2015; Tang and Yan, 2022; Albrecht et al., 2023; Bai et al., 2024; Thabet et al., 2024). However, further investigations are needed to improve our understanding of: i) How to build quantum-based structured kernels and what advantages they offer compared to classical kernels? ii) How to solve efficiently the pre-image problem for quantum structured prediction?

5. Support for OVks in QML

In this section, we provide initial support for our proposed shift towards quantum OVks.

5.1. Learning with SPD Matrices: Application to Quantum Channel Estimation

As a preliminary experimental proof-of-concept, we explore how moving from quantum scalar-valued to operator-valued kernels can enhance symmetric positive-definite (SPD) matrix learning. SPD matrices are widely used in machine learning as they naturally encode covariance structures that capture relationships between variables (Minh and Murino, 2017). Here, we focus on the classical estimation of quantum channels, a task that can be formulated as a SPD matrix-valued regression problem. The objective of quantum channel estimation is to determine the dynamical transformation,

known as a quantum channel, that governs the evolution of a given quantum system (Fujiwara, 2001).

In our experiment, we consider two prototypical single-qubit noise models: the bit-flip channel and the dephasing channel. Each channel depends on a parameter $a \in [0, 1]$, which specifies the probability of a bit-flip or a phase-flip. These channels provide simple yet instructive examples of quantum processes and are widely used to characterize noise in quantum devices. Details of the channels and the complete experimental setting are provided in Appendix B. The code used to generate the results reported in this work will be publicly available in a public Github repository.

We compare kernel learning performance using QSVK and QOVK. The QSVK is the fidelity kernel, defined as $k(\sigma_i, \sigma_j) = \text{Tr}(\sigma_i \sigma_j)$, while QOVK is defined by (10), with $\sigma_X^{i,j} = \sigma_i \sigma_j$. The QOVK formulation is flexible and gives rise to more expressive kernel feature spaces. The choice of U_{YX} is crucial: if it is the identity, QOVK reduces to QSVK. To incorporate entanglement, we use a non-separable U_{YX} , specifically the composition of a CNOT and a SWAP gate, enabling quantum correlations between input and output subsystems. Both models were trained via kernel ridge regression, with the regularization parameter selected by cross-validation.

Figure 5 illustrates an example of the true channels alongside the channels recovered using kernel ridge regression with SVks and entangled OVks. The OVk successfully captures the underlying structure, whereas the SVk struggles to do so. By treating the input and output spaces within a unified, entangled Hilbert space, QOVks offer a more natural framework. They can encode matrix structure directly and model correlations between input and output spaces. Entangled QOVks go further by using non-separable unitary operations to represent richer input–output dependencies

385 than separable or scalar-valued kernels can capture. Fur-
 386 ther results on quantum channel recovery with QOVKs and
 387 QSVKs are reported in Appendix B.

5.2. Quantum Implementation

389 An important question is whether an OVK could be im-
 390 plemented on a quantum computer. Here we present to
 391 our knowledge the first attempt to design a quantum circuit
 392 for OVKs. We build upon previous work on quantum im-
 393 plementation of scalar-valued kernels based on the swap
 394 test (Buhrman et al., 2001; Blank et al., 2020; Di Marcantonio
 395 et al., 2023). See Appendix C for more details.

398 We now present a quantum circuit for computing the OVK
 399 defined in (1). We consider the case where ρ_Y is a pure
 400 state (i.e., $\rho_Y = |\phi\rangle_Y \langle\phi|_Y$). We generalize the quantum
 401 fidelity kernel to the operator-valued setting by introducing
 402 the quantum circuit depicted in Figure 6. See Appendix D
 403 for a detailed description of the circuit implementation. En-
 404 tanglement between inputs and outputs is encoded via the
 405 matrix U acting on registers X and Y . The scalar-valued
 406 kernel can be computed via measurements of the ancillary
 407 qubit, and the separable quantum operator-valued kernel
 408 can be recovered if the operation U is separable, i.e., U
 409 cannot be written as a tensor product of unitaries acting
 410 independently on each subsystem.

412 The input feature matrix $\sigma_X^{x,z}$ produced by the quantum
 413 circuit is a density matrix, as it is obtained via a partial
 414 measurement and is therefore Hermitian. In contrast, the general
 415 definition of an entangled QOVK does not require $\sigma_X^{x,z}$
 416 to be Hermitian. For instance, in quantum channel estimation
 417 experiments, the QOVK is constructed from products of
 418 density matrices (i.e., $\sigma_X^{x,z} = \sigma_x \sigma_z$), which are generally
 419 non-Hermitian. The quantum circuit described here pro-
 420 vides an initial implementation of a QOVK, but it represents
 421 only a subset of the possibilities enabled by the framework.
 422 Future research should also explore alternative circuit de-
 423 signs that can efficiently realize more general forms of $\sigma_X^{x,z}$.
 424 Other strategies could include computing the vectorization
 425 of the QOVK, which reformulates the kernel in terms of
 426 pure states rather than mixed density matrices. By encoding
 427 the kernel in terms of pure-state representations, vectoriza-
 428 tion may enable more flexible computations and expand the
 429 range of quantum features that can be incorporated, although
 430 it may require additional quantum resources. Continued re-
 431 search in this area is essential for implementing QOVKs
 432 that address complex ML tasks and explore the full potential
 433 of quantum kernels.

6. Alternative view

435 Can research on scalar-valued kernels lead to new develop-
 436 ments in the field of quantum kernels? Understanding the

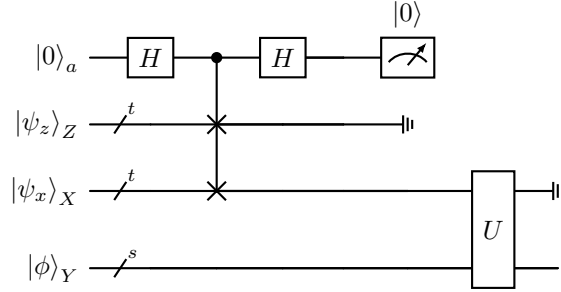


Figure 6. A quantum circuit for preparing a quantum state corresponding to the value of an operator-valued kernel of the form (1). The application of the partial trace on the register Z by measuring the ancillary qubit in the state $|0\rangle$ produces a valid input feature matrix $\sigma^{x,z}$ for the entangled kernel.

generalization abilities of quantum scalar-valued kernels could be a good alternative to explore the full power of quantum kernels. It is worth investigating generalization of quantum kernel machines in both noiseless and noisy settings (Canatar et al., 2023; Wang et al., 2021; Heyraud et al., 2022). Most of the research in this topic did not incorporate recent findings on generalization of classical learning. The study of the generalization of quantum kernel machines should take into account new phenomena of modern machine learning, such as double descent and benign overfitting (Belkin et al., 2019; Bartlett et al., 2020). Only very few studies have recently appeared in the literature that attempt to look at generalization in overparameterized quantum machine learning models (Larocca et al., 2023; Peters and Schuld, 2023; Gil-Fuster et al., 2024; Thanasilp et al., 2024; Kempkes et al., 2025; Tomasi et al., 2025). Another interesting research direction is to learn the quantum feature map of quantum kernels via quantum neural networks (a.k.a quantum parametrized circuits) (Lloyd et al., 2020; Gentinetta et al., 2023; Hubregtzen et al., 2022; Rodriguez-Grasa et al., 2024; Incudini et al., 2024; Glick et al., 2024; Lei et al., 2024). This should give rise to new quantum kernel features adapted to the task at hand. Such research avenues may also inform the design and analysis of quantum operator-valued kernels.

7. Conclusion

This position paper highlights quantum operator-valued kernels as a promising extension of scalar-valued quantum kernels, with the potential to exploit entanglement for richer and more expressive quantum learning models. We outline key challenges and a roadmap for advancing operator-valued kernels within QML, aiming to inspire further exploration of this framework and to foster dialogue across the machine learning and quantum computing communities.

References

- Amira Abbas, Andris Ambainis, Brandon Augustino, Andreas Bärtzchi, Harry Buhrman, Carleton Coffrin, Giorgio Cortiana, Vedran Dunjko, Daniel J Egger, Bruce G Elmegreen, et al. Challenges and opportunities in quantum optimization. *Nature Reviews Physics*, pages 1–18, 2024.
- A. Aizerman, E. M. Braverman, and L. I. Rozoner. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.
- Boris Albrecht, Constantin Dalyac, Lucas Leclerc, Luis Ortiz-Gutiérrez, Slimane Thabet, Mauro D’Arcangelo, Julia RK Cline, Vincent E Elfving, Lucas Lassablière, Henrique Silvério, et al. Quantum feature maps for graph machine learning on a neutral atom quantum processor. *Physical Review A*, 107(4):042615, 2023.
- Mauricio A Alvarez, Lorenzo Rosasco, Neil D Lawrence, et al. Kernels for vector-valued functions: A review. *Foundations and Trends® in Machine Learning*, 4(3): 195–266, 2012.
- N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.
- Lu Bai, Luca Rossi, Andrea Torsello, and Edwin R Hancock. A quantum jensen–shannon graph kernel for unattributed graphs. *Pattern Recognition*, 48(2):344–355, 2015.
- Lu Bai, Lixin Cui, Ming Li, Yue Wang, and Edwin Hancock. Qbmk: Quantum-based matching kernels for unattributed graphs. In *International Conference on Machine Learning*, 2024.
- G. H. Bakir, T. Hofmann, B. Schölkopf, A. J. Smola, B. Taskar, and S. V. N. Vishwanathan. *Predicting structured data*. MIT press, 2007.
- Peter L Bartlett, Philip M Long, Gábor Lugosi, and Alexander Tsigler. Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences*, 117(48): 30063–30070, 2020.
- Serban Belinschi, Victor Magron, and Victor Vinnikov. Non-commutative christoffel-darboux kernels. *Transactions of the American Mathematical Society*, 376(01):181–230, 2023.
- Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.
- Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.
- Carsten Blank, Daniel K Park, June-Koo Kevin Rhee, and Francesco Petruccione. Quantum classifier with tailored quantum kernel. *npj Quantum Information*, 6(1):41, 2020.
- Karsten Borgwardt, Elisabetta Ghisu, Felipe Llinares-López, Leslie O’Bray, Bastian Rieck, et al. Graph kernels: State-of-the-art and future challenges. *Foundations and Trends® in Machine Learning*, 13(5-6):531–712, 2020.
- B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Annual Workshop on Computational Learning Theory*, pages 144–152. ACM, June 1992.
- Céline Brouard, Marie Szafranski, and Florence d’Alché Buc. Input output kernel regression: Supervised and semi-supervised structured output prediction with operator-valued kernels. *Journal of Machine Learning Research*, 17:np, 2016.
- Jean-Bernard Bru and Walter Alberto de Siqueira Pedra. *C*-Algebras and Mathematical Foundations of Quantum Statistical Mechanics: An Introduction*. Springer Nature, 2023.
- Harry Buhrman, Richard Cleve, John Watrous, and Ronald De Wolf. Quantum fingerprinting. *Physical review letters*, 87(16):167902, 2001.
- Sabine Burgdorf, Igor Klep, Janez Povh, et al. *Optimization of polynomials in non-commuting variables*, volume 2. Springer, 2016.
- Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Oliveira, Michael Walter, and Avi Wigderson. Towards a theory of non-commutative optimization: Geodesic 1st and 2nd order methods for moment maps and polytopes. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 845–861, 2019.
- Abdulkadir Canatar, Evan Peters, Cengiz Pehlevan, Stefan M. Wild, and Ruslan Shaydulin. Bandwidth enables generalization in quantum kernel models. *Transactions on Machine Learning Research*, 2023.
- Andrea Caponnetto, Charles A Micchelli, Massimiliano Pontil, and Yiming Ying. Universal multi-task kernels. *The Journal of Machine Learning Research*, 9:1615–1646, 2008.
- Claudio Carmeli, Ernesto De Vito, Alessandro Toigo, and Veronica Umanitá. Vector valued reproducing kernel Hilbert spaces and universality. *Analysis and Applications*, 8(01):19–61, 2010.

- 495 Carlo Ciliberto, Mark Herbster, Alessandro Davide Ialongo,
496 Massimiliano Pontil, Andrea Rocchetto, Simone Severini,
497 and Leonard Wossnig. Quantum machine learning: a
498 classical perspective. *Proceedings of the Royal Society A:
499 Mathematical, Physical and Engineering Sciences*, 474
500 (2209):20170551, 2018.
- 501 Brian Coyle, Daniel Mills, Vincent Danos, and Elham
502 Kashefi. The born supremacy: quantum advantage and
503 training of an ising born machine. *npj Quantum Informa-
504 tion*, 6(1):60, 2020.
- 505 Ronald De Wolf. Quantum computing: Lecture notes. *arXiv
506 preprint arXiv:1907.09415*, 2019.
- 507 Francesco Di Marcantonio, Massimiliano Incudini, Davide
508 Tezza, and Michele Grossi. Quantum advantage seeker
509 with kernels (quask): a software framework to speed up
510 the research in quantum machine learning. *Quantum
511 Machine Intelligence*, 5(1):20, 2023.
- 512 Yuxuan Du, Xinbiao Wang, Naixu Guo, Zhan Yu, Yang
513 Qian, Kaining Zhang, Min-Hsiu Hsieh, Patrick Reben-
514 trost, and Dacheng Tao. *A Gentle Introduction to Quan-
515 tum Machine Learning*. Springer Nature, 2025.
- 516 Vedran Dunjko and Hans J Briegel. Machine learning &
517 artificial intelligence in the quantum domain: a review of
518 recent progress. *Reports on Progress in Physics*, 81(7):
519 074001, 2018.
- 520 Tamim El Ahmad, Junjie Yang, Pierre Laforgue, and Flo-
521 rence d’Alché Buc. Deep sketched output kernel regres-
522 sion for structured prediction. In *Joint European Confer-
523 ence on Machine Learning and Knowledge Discovery in
524 Databases*, pages 93–110, 2024.
- 525 Theodoros Evgeniou, Charles A Micchelli, Massimiliano
526 Pontil, and John Shawe-Taylor. Learning multiple tasks
527 with kernel methods. *Journal of machine learning re-
528 search*, 6(4), 2005.
- 529 Akio Fujiwara. Quantum channel identification problem.
530 *Physical Review A*, 63(4):042304, 2001.
- 531 Valentin Gebhart, Raffaele Santagati, Antonio Andrea Gen-
532 tile, Erik M Gauger, David Craig, Natalia Ares, Leonardo
533 Banchi, Florian Marquardt, Luca Pezzè, and Cristian Bon-
534 ato. Learning quantum systems. *Nature Reviews Physics*,
535 5(3):141–156, 2023.
- 536 Gian Gentinetta, David Sutter, Christa Zoufal, Bryce Fuller,
537 and Stefan Woerner. Quantum kernel alignment with
538 stochastic gradient descent. In *2023 IEEE International
539 Conference on Quantum Computing and Engineering
540 (QCE)*, volume 1, pages 256–262. IEEE, 2023.
- 541 Pierre Geurts, Louis Wehenkel, and Florence d’Alché Buc.
542 Kernelizing the output of tree-based methods. In *Pro-
543 ceedings of the 23rd international conference on Machine
544 learning*, pages 345–352, 2006.
- 545 Elies Gil-Fuster, Jens Eisert, and Carlos Bravo-Prieto. Un-
546 derstanding quantum machine learning also requires re-
547 thinking generalization. *Nature Communications*, 15(1):
548 2277, 2024.
- 549 Jennifer R Glick, Tanvi P Gujarati, Antonio D Corcoles,
550 Youngseok Kim, Abhinav Kandala, Jay M Gambetta, and
551 Kristan Temme. Covariant quantum kernels for data with
552 group structure. *Nature Physics*, 20(3):479–483, 2024.
- 553 Sander Gribling, David de Laat, and Monique Laurent.
554 Bounds on entanglement dimensions and quantum graph
555 parameters via noncommutative polynomial optimization.
556 *Mathematical programming*, 170:5–42, 2018.
- 557 Lov K Grover. A fast quantum mechanical algorithm for
558 database search. In *Proceedings of the twenty-eighth
559 annual ACM symposium on Theory of computing*, pages
560 212–219, 1996.
- 561 Yuka Hashimoto, Isao Ishikawa, Masahiro Ikeda, Fuyuta
562 Komura, Takeshi Katsura, and Yoshinobu Kawahara. Re-
563 producing kernel Hilbert C^* -module and kernel mean
564 embeddings. *The Journal of Machine Learning Research*,
565 22(1):12292–12347, 2021.
- 566 Yuka Hashimoto, Masahiro Ikeda, and Hachem Kadri.
567 Deep learning with kernels through rkhn and the perron-
568 frobenius operator. *Advances in Neural Information Pro-
569 cessing Systems*, 36:50677–50696, 2023a.
- 570 Yuka Hashimoto, Masahiro Ikeda, and Hachem Kadri.
571 Learning in RKHM: a C^* -algebraic twist for kernel ma-
572 chines. In *International Conference on Artificial Intelli-
573 gence and Statistics*, pages 692–708, 2023b.
- 574 Yuka Hashimoto, Ayoub Hafid, Masahiro Ikeda, and
575 Hachem Kadri. Spectral truncation kernels: Noncommu-
576 tativity in C^* -algebraic kernel machines. *arXiv preprint
577 arXiv:2405.17823*, 2024.
- 578 Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme,
579 Aram W Harrow, Abhinav Kandala, Jerry M Chow, and
580 Jay M Gambetta. Supervised learning with quantum-
581 enhanced feature spaces. *Nature*, 567(7747):209–212,
582 2019.
- 583 Valentin Heyraud, Zejian Li, Zakari Denis, Alexandre
584 Le Boité, and Cristiano Ciuti. Noisy quantum kernel
585 machines. *Physical Review A*, 106(5):052421, 2022.
- 586 Thomas Hofmann, Bernhard Schölkopf, and Alexander J
587 Smola. Kernel methods in machine learning. *The Annals
588 of Statistics*, 36(3):1171–1220, 2008.

- 550 Paul Honeine and Cedric Richard. Preimage problem in
551 kernel-based machine learning. *IEEE Signal Processing*
552 *Magazine*, 28(2):77–88, 2011.
- 553 Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni,
554 Ryan Babbush, Sergio Boixo, Hartmut Neven, and
555 Jarrod R. McClean. Power of data in quantum
556 machine learning. *Nature Communications*, 12
557 (1), 05 2021. ISSN 2041-1723. doi: 10.1038/
558 s41467-021-22539-9. URL [http://dx.doi.org/
559 10.1038/s41467-021-22539-9](http://dx.doi.org/10.1038/s41467-021-22539-9).
- 560 Thomas Hubregtsen, David Wierichs, Elies Gil-Fuster,
561 Peter-Jan HS Derks, Paul K Faehrmann, and Jo-
562 hannes Jakob Meyer. Training quantum embedding ker-
563 nels on near-term quantum computers. *Physical Review*
564 *A*, 106(4):042431, 2022.
- 565 Riikka Huusari and Hachem Kadri. Entangled kernels - be-
566 yond separability. *Journal of Machine Learning Research*,
567 22(24):1–40, 2021.
- 568 Riikka Huusari, Hachem Kadri, and Cécile Capponi. Multi-
569 view metric learning in vector-valued kernel spaces. In
570 *International Conference on Artificial Intelligence and*
571 *Statistics*, pages 415–424, 2018.
- 572 Massimiliano Incudini, Daniele Lizzio Bosco, Francesco
573 Martini, Michele Grossi, Giuseppe Serra, and Alessandra
574 Di Pierro. Automatic and effective discovery of quan-
575 tum kernels. *IEEE Transactions on Emerging Topics in*
576 *Computational Intelligence*, 2024.
- 577 Sofiene Jerbi, Lukas J Fiderer, Hendrik Poulsen Nautrup,
578 Jonas M Kübler, Hans J Briegel, and Vedran Dunjko.
579 Quantum machine learning beyond kernel methods. *Nature*
580 *Communications*, 14(1):517, 2023.
- 581 Michael I Jordan and Tom M Mitchell. Machine learning:
582 Trends, perspectives, and prospects. *Science*, 349(6245):
583 255–260, 2015.
- 584 Richard Jozsa and Noah Linden. On the role of entangle-
585 ment in quantum-computational speed-up. *Proceedings*
586 *of the Royal Society of London. Series A: Mathematical,*
587 *Physical and Engineering Sciences*, 459(2036):2011–
588 2032, 2003.
- 589 Hachem Kadri, Mohammad Ghavamzadeh, and Philippe
590 Preux. A generalized kernel approach to structured out-
591 put learning. In *International Conference on Machine*
592 *Learning*, pages 471–479, 2013.
- 593 Hachem Kadri, Emmanuel Duflos, Philippe Preux, Stéphane
594 Canu, Alain Rakotomamonjy, and Julien Audiffren.
595 Operator-valued kernels for learning from functional re-
596 sponse data. *Journal of Machine Learning Research*, 16:
597 1–54, 2016.
- 598 Marie Kempkes, Aroosa Ijaz, Elies Gil-Fuster, Carlos Bravo-
599 Prieto, Jakob Spiegelberg, Evert van Nieuwenburg, and
600 Vedran Dunjko. Double descent in quantum machine
601 learning. *arXiv preprint arXiv:2501.10077*, 2025.
- 602 Jonas Kübler, Simon Buchholz, and Bernhard Schölkopf.
603 The inductive bias of quantum kernels. *Advances in Neu-
604 ral Information Processing Systems*, 34:12661–12673,
2021.
- Takeru Kusumoto, Kosuke Mitarai, Keisuke Fujii, Masahiro
Kitagawa, and Makoto Negoro. Experimental quantum
kernel trick with nuclear spins in a solid. *npj Quantum
Information*, 7(1):94, 2021.
- Jonas Landman, Slimane Thabet, Constantin Dalyac, Hela
Mhiri, and Elham Kashefi. Classically approxim-
ating variational quantum machine learning with random
fourier features. In *International Conference on Learning
Representations*, 2023.
- Martin Larocca, Nathan Ju, Diego García-Martín, Patrick J
Coles, and Marco Cerezo. Theory of overparametriza-
tion in quantum neural networks. *Nature Computational
Science*, 3(6):542–551, 2023.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep
learning. *nature*, 521(7553):436–444, 2015.
- Cong Lei, Yuxuan Du, Peng Mi, Jun Yu, and Tongliang Liu.
Neural auto-designer for enhanced quantum kernels. In
International Conference on Learning Representations,
2024.
- Néhémy Lim, Florence d’Alché Buc, Cédric Auliac, and
George Michailidis. Operator-valued kernel-based vector
autoregressive models for network inference. *Machine
learning*, 99:489–513, 2015.
- Fanghui Liu, Xiaolin Huang, Yudong Chen, and Johan AK
Suykens. Random features for kernel approximation: A
survey on algorithms, theory, and beyond. *IEEE Transac-
tions on Pattern Analysis and Machine Intelligence*, 44
(10):7128–7148, 2021a.
- Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme.
A rigorous and robust quantum speed-up in supervised
machine learning. *Nature Physics*, 17(9):1013–1017,
2021b.
- Seth Lloyd, Maria Schuld, Aroosa Ijaz, Josh Izaac, and
Nathan Killoran. Quantum embeddings for machine
learning. *arXiv preprint arXiv:2001.03622*, 2020.
- Riccardo Mengoni and Alessandra Di Pierro. Kernel meth-
ods in quantum machine learning. *Quantum Machine
Intelligence*, 1(3-4):65–71, 2019.

- 605 Charles A Micchelli and Massimiliano Pontil. On learning
606 vector-valued functions. *Neural computation*, 17(1):177–
607 204, 2005.
- 608 Hà Quang Minh and Vittorio Murino. *Covariances in com-
609 puter vision and machine learning*. Morgan & Claypool
610 Publishers, 2017.
- 611 Mauro ES Morales, Lirandë Pira, Philipp Schleich, Kelvin
612 Koor, Pedro Costa, Dong An, Alán Aspuru-Guzik, Lin
613 Lin, Patrick Rebentrost, and Dominic W Berry. Quan-
614 tum linear system solvers: A survey of algorithms and
615 applications. *arXiv preprint arXiv:2411.02522*, 2024.
- 616 Azar C Nakhli, Thomas Quella, and Muhammad Usman.
617 Calibrating the role of entanglement in variational quan-
618 tum circuits. *Physical Review A*, 109(3):032413, 2024.
- 619 Michael A Nielsen and Isaac L Chuang. *Quantum compu-
620 tation and quantum information*. Cambridge university
621 press, 2010.
- 622 Evan Peters and Maria Schuld. Generalization despite over-
623 fitting in quantum machine learning models. *Quantum*, 7:
624 1210, 2023.
- 625 John Preskill. Quantum computing in the NISQ era and
626 beyond. *Quantum*, 2:79, 2018.
- 627 Pablo Rodriguez-Grasa, Yue Ban, and Mikel Sanz. Neural
628 quantum kernels: training quantum kernels with quantum
629 neural networks. *arXiv preprint arXiv:2401.04642*, 2024.
- 630 Mathieu Roget, Giuseppe Di Molfetta, and Hachem Kadri.
631 Quantum perceptron revisited: Computational-statistical
632 tradeoffs. In *Uncertainty in Artificial Intelligence*, pages
633 1697–1706, 2022.
- 634 Maria Schuld. Supervised quantum machine learning mod-
635 els are kernel methods. *arXiv preprint arXiv:2101.11020*,
636 2021.
- 637 Maria Schuld and Nathan Killoran. Quantum machine learn-
638 ing in feature Hilbert spaces. *Physical review letters*, 122
639 (4):040504, 2019.
- 640 Maria Schuld and Nathan Killoran. Is quantum advantage
641 the right goal for quantum machine learning? *Prx Quan-
642 tum*, 3(3):030101, 2022.
- 643 Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione.
644 An introduction to quantum machine learning. *Contem-
645 porary Physics*, 56(2):172–185, 2015.
- 646 L. Schwartz. Sous-espaces hilbertiens d’espaces vectoriels
647 topologiques et noyaux associés (noyaux reproduisants).
648 *Journal d’Analyse Mathématique*, 13:115–256, 1964.
- 649 Chence Shi, Minkai Xu, Hongyu Guo, Ming Zhang, and
650 Jian Tang. A graph to graphs framework for retrosynthe-
651 sis prediction. In *International conference on machine
652 learning*, pages 8818–8827, 2020.
- 653 George Stepaniants. Learning partial differential equations
654 in reproducing kernel Hilbert spaces. *Journal of Machine
655 Learning Research*, 24(86):1–72, 2023.
- 656 Yudai Suzuki, Hideaki Kawaguchi, and Naoki Yamamoto.
657 Quantum fisher kernel for mitigating the vanishing simi-
658 larity issue. *arXiv preprint arXiv:2210.16581*, 2022.
- 659 Yehui Tang and Junchi Yan. Graphqntk: Quantum neu-
ral tangent kernel for graph data. *Advances in Neural
Information Processing Systems*, 35:6104–6118, 2022.
- Slimane Thabet, Mehdi Djellabi, Igor Sokolov, Sachin Kas-
ture, Louis-Paul Henry, and Loïc Henriët. Quantum posi-
tional encodings for graph neural networks. In *Internat-
ional Conference on Machine Learning*, 2024.
- Supanut Thanasilp, Samson Wang, M Cerezo, and Zoë
Holmes. Exponential concentration in quantum kernel
methods. *Nature Communications*, 15(1):5200, 2024.
- Joachim Tomasi, Sandrine Anthoine, and Hachem Kadri.
Benign overfitting with quantum kernels. *arXiv preprint
arXiv:2503.17020*, 2025.
- Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hof-
mann, Yasemin Altun, and Yoram Singer. Large margin
methods for structured and interdependent output vari-
ables. *Journal of machine learning research*, 6(9), 2005.
- Xinbiao Wang, Yuxuan Du, Yong Luo, and Dacheng Tao.
Towards understanding the power of quantum kernels in
the NISQ era. *Quantum*, 5:531, 2021.
- Xinbiao Wang, Yuxuan Du, Zhuozhuo Tu, Yong Luo, Xiao
Yuan, and Dacheng Tao. Transition role of entangled data
in quantum machine learning. *Nature Communications*,
15(1):3716, 2024.
- John Watrous. *The theory of quantum information*. Cam-
bridge university press, 2018.
- Muhan Zhang and Yixin Chen. Link prediction based on
graph neural networks. *Advances in neural information
processing systems*, 31, 2018.

A. Operator-Valued Kernels and Vector-Valued RKHSs

Consider the supervised learning problem where the goal is to learn a function $f : \mathcal{X} \rightarrow \mathcal{Y}$ given a training set $\{(x_i, y_i)\}_{i=1}^n$, where x_i is in some space \mathcal{X} and y is in a Hilbert space \mathcal{Y} . The space \mathcal{Y} can be finite or infinite-dimensional. For example, in multitask learning where the objective is to solve simultaneously p learning problems, the output space \mathcal{Y} can be \mathbb{R}^p . In functional regression, output data are curves represented by functions and the output space \mathcal{Y} can be the space L^2 of square integrable functions. Learning the function f in such situations is more challenging than finding a scalar-valued function such is the case for standard classification or regression. The framework of scalar-valued kernels is not rich enough to learn nonlinear vector-valued functions that maps complex input data to complex outputs. Operator-valued kernels provide an elegant solution to this problem. The kernel in this case is a function that takes two input data points and outputs an operator rather than a scalar as usual, i.e., $K(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{L}(\mathcal{Y})$, where $\mathcal{L}(\mathcal{Y})$ is the space of bounded operators from \mathcal{Y} to itself. The operator allows to encode prior information about the outputs, and then take into account the output structure. More formally,

Definition A.1. (psd operator-valued kernel)

A $\mathcal{L}(\mathcal{Y})$ -valued kernel K on $\mathcal{X} \times \mathcal{X}$ is a function $K(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{L}(\mathcal{Y})$; it is positive semi-definite (psd) if:

- i. $K(\mathbf{x}, \mathbf{z}) = K(\mathbf{z}, \mathbf{x})^*$, where superscript $*$ denotes the adjoint operator,
- ii. and, for every $n \in \mathbb{N}$ and all $\{(\mathbf{x}_i, \mathbf{y}_i)_{i=1}^n\} \in \mathcal{X} \times \mathcal{Y}$,

$$\sum_{i,j} \langle \mathbf{y}_i, K(\mathbf{x}_i, \mathbf{x}_j) \mathbf{y}_j \rangle_{\mathcal{Y}} \geq 0.$$

Definition A.2. (vector-valued RKHS)

A Hilbert space \mathcal{H} of functions from \mathcal{X} to \mathcal{Y} is called a reproducing kernel Hilbert space if there is a positive semi-definite $\mathcal{L}(\mathcal{Y})$ -valued kernel K on $\mathcal{X} \times \mathcal{X}$ such that:

- i. $\mathbf{z} \mapsto K(\mathbf{x}, \mathbf{z})\mathbf{y}$ belongs to \mathcal{H} , $\forall \mathbf{z}, \mathbf{x} \in \mathcal{X}$, $\mathbf{y} \in \mathcal{Y}$,
- ii. $\forall f \in \mathcal{H}, x \in \mathcal{X}$, $\mathbf{y} \in \mathcal{Y}$,

$$\langle f, K(\mathbf{x}, \cdot) \mathbf{y} \rangle_{\mathcal{H}} = \langle f(\mathbf{x}), \mathbf{y} \rangle_{\mathcal{Y}} \text{ (reproducing property).}$$

A key point for learning with kernels is the ability to express functions in terms of a kernel providing the way to evaluate a function at a given point. This is possible because there exists a bijection relationship between a large class of kernels and associated reproducing kernel spaces which satisfy a regularity property. Bijection between scalar-valued kernels and RKHS was first established by Aronszajn (1950, Part I, Sections 3 and 4). Then Schwartz (1964, Chapter 5) shows that this is a particular case of a more general situation. This bijection in the case of operator-valued kernels is still valid.

Theorem A.3. (bijection between vector-valued RKHS and positive semi-definite operator-valued kernel)

An $\mathcal{L}(\mathcal{Y})$ -valued kernel K on $\mathcal{X} \times \mathcal{X}$ is the reproducing kernel of some Hilbert space \mathcal{H} , if and only if it is positive semi-definite.

For further reading on operator-valued kernels and their associated RKHSs, see, e.g., (Caponnetto et al., 2008; Carmeli et al., 2010; Alvarez et al., 2012; Kadri et al., 2016).

B. Additional Experimental Details

B.1. Experimental Setting

We consider a quantum system described by a Hilbert space \mathcal{F} , with the associated quantum channel represented as $\Gamma : \mathcal{S}(\mathcal{F}) \rightarrow \mathcal{S}(\mathcal{F})$, where $\mathcal{S}(\mathcal{F})$ denotes the set of density operators over \mathcal{F} . The quantum channel Γ is a trace-preserving completely positive map (Watrous, 2018, Chap. 2). Quantum channel estimation can be viewed as a structured prediction problem in which both the inputs and outputs are quantum states. In our experiment, each quantum state is classically encoded as a density matrix, which is a Hermitian, positive semidefinite matrix with unit trace. Leveraging the structural properties of density matrices during learning is essential for accurately reconstructing the quantum channel from observed

Table 1. Experimental results on quantum channel estimation with quantum scalar-valued and quantum operator-valued kernel ridge regression. The recovery error measure is $\|\text{Channel}_{\text{true}} - \text{Channel}_{\text{learned}}\|_F$.

Channel	QSVK	QOVK
Bit-flip	0.569 ± 0.304	0.101 ± 0.035
Dephasing	0.525 ± 0.297	0.096 ± 0.031

data. We apply kernel-based learning algorithms to this setting, considering both quantum scalar-valued kernels (Canatar et al., 2023) and quantum operator-valued kernels as defined in (1).

We focus on learning general completely positive maps, so that the channel takes as input a SPD matrix of size $d \times d$ and outputs a SPD matrix of size $p \times p$. Now the Choi matrix representation of this channel is a matrix of size $dp \times dp$. For certain types of channels (especially ones modeling physical systems) the input and output matrices are of the same size. In our experiment we consider two prototypical single-qubit noise models: the bit-flip channel and the dephasing channel. These channels represent simple yet instructive examples of quantum processes and are widely used for characterizing noise in quantum devices. The bit-flip channel models quantum noise where a qubit is flipped from $|0\rangle$ to $|1\rangle$ and vice versa with probability a . It is defined, for any density matrix σ , as

$$\mathcal{E}_{\text{bf}}(\sigma) := (1 - a)\sigma + a X \sigma X,$$

where $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the Pauli- X operator and $a \in [0, 1]$ is the bit-flip probability. The dephasing channel is a very simple decoherence process that takes a density matrix σ to

$$\mathcal{E}_{\text{deph}}(\sigma) := (1 - a)\sigma + a \text{diag}(\sigma),$$

where $\text{diag}(\sigma)$ denotes the matrix comprising only the diagonal elements of σ , and $a \in [0, 1]$ measures the extent to which the off-diagonal elements, usually called coherences, are reduced in magnitude. The true channel to be estimated is fixed with a given a and represented in its Choi matrix form Φ , which serves as the ground truth in our evaluation. We generate n training examples by sampling random density matrices $\sigma_i, i = 1, \dots, n$, as inputs and passing them through the channel to produce outputs ρ_i . To simulate noisy situations, we added a small perturbation: each output was replaced by a convex combination of the channel output and a random density matrix with mixing parameter $\alpha \in [0, 1]$, i.e.,

$$\rho_i = (1 - \alpha)\mathcal{E}_{\text{bf/deph}}(\sigma_i) + \alpha \tau_i.$$

This provides a more challenging estimation problem and tests the models’ robustness to deviations from ideal data. Random density matrices are generated using QETLAB¹.

B.2. Quantum Channel Estimation Results

The recovery errors, defined as $\|\text{Channel}_{\text{true}} - \text{Channel}_{\text{learned}}\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm, are reported in Table 1, averaged over 10 repetitions of the experiment. These results report the mean performance for channel estimation, with parameter a ranging from 0 to 1 with a step size of 0.1. The OVK regression achieves significantly better performance than SVK regression.

C. Circuit Implementation of Quantum SVKs via Swap Test

The swap test is a quantum algorithm that estimates the fidelity between two quantum states $|\psi_x\rangle$ and $|\psi_z\rangle$, i.e., $F(\psi_x, \psi_z) := |\langle \psi_x | \psi_z \rangle|^2$. Let $|\psi_x\rangle := U_x |0_t\rangle$ be an encoding of a data point $x \in \mathcal{X}$ into a quantum state of t qubits obtained by applying a parametrized unitary operation U_x to the initial state $|0_t\rangle := |0\rangle^{\otimes t}$. Using the density matrix formalism, the encoding corresponds to mapping an input data x to a pure (i.e., rank-one) density matrix $\rho_x := U_x |0_t\rangle \langle 0_t| U_x^\dagger = |\psi_x\rangle \langle \psi_x|$. The fidelity kernel is the function $k(x, z) := \text{Tr}[\rho_x \rho_z] = |\langle \psi_x | \psi_z \rangle|^2$. The quantum circuit implementing the computation of this kernel using a swap test is given in Figure 7.

¹<http://www.qetlab.com>.

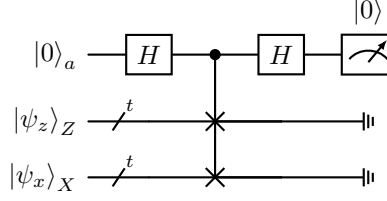


Figure 7. Quantum circuit for computing the fidelity kernel using a swap test. Measuring the ancillary qubit provides the fidelity between the two quantum states $|\psi_x\rangle$ and $|\psi_z\rangle$. The probability to measure the ancillary qubit in the state $|0\rangle$ is given by $\mathbb{P}(|0\rangle_a) = \frac{1}{2} + \frac{1}{2} |\langle \psi_x | \psi_z \rangle|^2$ (Schuld et al., 2015).

D. Circuit Implementation of Quantum OVks via Swap Test

We provide here the details of the computation of the quantum circuit implementing the QOVK given in (1) (see Figure 8). The circuit uses Hadamard and controlled-swap gates. Recall that the Hadamard gate H maps an basis state $|i\rangle$, $i \in \{0, 1\}$ to the equal-weight superposition of basis states, i.e., $H|i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (-1)^i|1\rangle)$. The controlled-swap gate on two states $|\psi_x\rangle_X$ and $|\psi_z\rangle_Z$, controlled on the single qubit $|i\rangle_a$ is defined as

$$CSWAP_{aZX} |i\rangle_a |\psi_z\rangle_Z |\psi_x\rangle_X = \begin{cases} |i\rangle_a |\psi_z\rangle_Z |\psi_x\rangle_X & \text{if } i = 0, \\ |i\rangle_a |\psi_x\rangle_Z |\psi_z\rangle_X & \text{if } i = 1. \end{cases}$$

The initial state of the circuit is

$$|\Psi_0\rangle = |0\rangle_a |\psi_z\rangle_Z |\psi_x\rangle_X |\phi\rangle_Y. \quad (2)$$

Applying a Hadamard gate to the ancilla qubit of $|\Psi_0\rangle$ leads to:

$$|\Psi_1\rangle = (H_a \otimes I_{ZXY}) |\Psi_0\rangle = \frac{1}{\sqrt{2}} (|0\rangle_a + |1\rangle_a) |\psi_z\rangle_Z |\psi_x\rangle_X |\phi\rangle_Y. \quad (3)$$

$|\Psi_2\rangle$ is obtained by applying the controlled-swap gate to $|\Psi_1\rangle$:

$$\begin{aligned} |\Psi_2\rangle &= CSWAP_{aZX} |\Psi_1\rangle = CSWAP_{aZX} \left(\frac{|0\rangle_a + |1\rangle_a}{\sqrt{2}} \right) |\psi_z\rangle_Z |\psi_x\rangle_X |\phi\rangle_Y \\ &= \frac{1}{\sqrt{2}} (|0\rangle_a |\psi_z\rangle_Z |\psi_x\rangle_X + |1\rangle_a |\psi_x\rangle_Z |\psi_z\rangle_X) |\phi\rangle_Y. \end{aligned} \quad (4)$$

A Hadamard gate is then applied to the ancillary qubit of $|\Psi_2\rangle$ giving

$$\begin{aligned} |\Psi_3\rangle &= (H_a \otimes I_{ZXY}) |\Psi_2\rangle = (H_a \otimes I_{ZXY}) \frac{1}{\sqrt{2}} (|0\rangle_a |\psi_z\rangle_Z |\psi_x\rangle_X + |1\rangle_a |\psi_x\rangle_Z |\psi_z\rangle_X) |\phi\rangle_Y \\ &= \frac{1}{2} [(|0\rangle_a + |1\rangle_a) |\psi_z\rangle_Z |\psi_x\rangle_X + (|0\rangle_a - |1\rangle_a) |\psi_x\rangle_Z |\psi_z\rangle_X] |\phi\rangle_Y \\ &= \frac{1}{2} [|0\rangle_a (|\psi_z\rangle_Z |\psi_x\rangle_X + |\psi_x\rangle_Z |\psi_z\rangle_X) + |1\rangle_a (|\psi_z\rangle_Z |\psi_x\rangle_X - |\psi_x\rangle_Z |\psi_z\rangle_X)] |\phi\rangle_Y. \end{aligned} \quad (5)$$

This pure state $|\Psi_3\rangle$ can be represented by its density matrix $\rho^{\Psi_3} = |\Psi_3\rangle \langle \Psi_3|$. Defining $\rho_x = |\psi_x\rangle \langle \psi_x|$, $\rho_z = |\psi_z\rangle \langle \psi_z|$ and omitting register subscripts for readability, ρ^{Ψ_3} writes as:

$$\rho^{\Psi_3} = \sigma^{\Psi_3} \otimes |\phi\rangle \langle \phi|, \quad (6)$$

where

$$\begin{aligned} \sigma^{\Psi_3} &= \frac{1}{4} \left[|0\rangle \langle 0| \otimes (\rho_z \otimes \rho_x + \rho_x \otimes \rho_z + |\psi_z\rangle \langle \psi_x| \otimes |\psi_x\rangle \langle \psi_z| + |\psi_x\rangle \langle \psi_z| \otimes |\psi_z\rangle \langle \psi_x|) \right. \\ &\quad + |0\rangle \langle 1| \otimes (\rho_z \otimes \rho_x - \rho_x \otimes \rho_z - |\psi_z\rangle \langle \psi_x| \otimes |\psi_x\rangle \langle \psi_z| + |\psi_x\rangle \langle \psi_z| \otimes |\psi_z\rangle \langle \psi_x|) \\ &\quad + |1\rangle \langle 0| \otimes (\rho_z \otimes \rho_x - \rho_x \otimes \rho_z + |\psi_z\rangle \langle \psi_x| \otimes |\psi_x\rangle \langle \psi_z| - |\psi_x\rangle \langle \psi_z| \otimes |\psi_z\rangle \langle \psi_x|) \\ &\quad \left. + |1\rangle \langle 1| \otimes (\rho_z \otimes \rho_x + \rho_x \otimes \rho_z - |\psi_z\rangle \langle \psi_x| \otimes |\psi_x\rangle \langle \psi_z| - |\psi_x\rangle \langle \psi_z| \otimes |\psi_z\rangle \langle \psi_x|) \right]. \end{aligned}$$

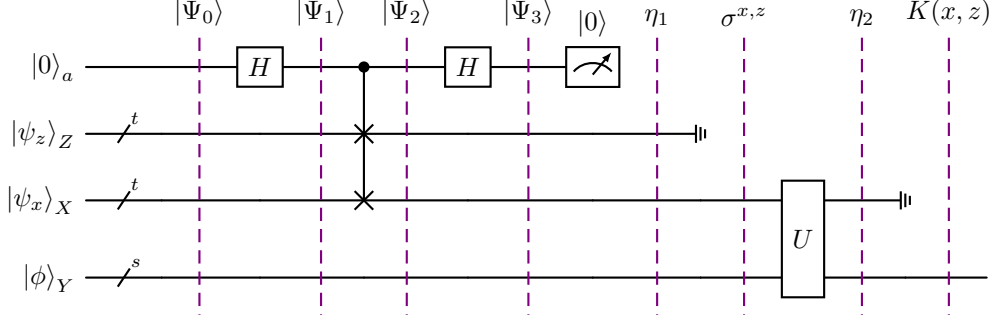


Figure 8. A quantum circuit implementation of the operator-valued kernel given in (1).

The state η_1 is then obtained by measuring the ancilla qubit in the state $|0\rangle$, i.e.,

$$\begin{aligned} \eta_1 &= \frac{\text{Tr}_a[\rho^{\Psi_3}(|0\rangle\langle 0|_a \otimes I_{ZXY})]}{\text{Tr}[\rho^{\Psi_3}(|0\rangle\langle 0|_a \otimes I_{ZXY})]} \\ &= \frac{\rho_z \otimes \rho_x + \rho_x \otimes \rho_z + |\psi_z\rangle\langle\psi_z| \otimes |\psi_x\rangle\langle\psi_x| + |\psi_x\rangle\langle\psi_x| \otimes |\psi_z\rangle\langle\psi_z|}{2(1 + |\langle\psi_z|\psi_x\rangle|^2)} \otimes |\phi\rangle\langle\phi|. \end{aligned} \quad (7)$$

The input feature matrix $\sigma_X^{x,z}$ is obtained in the register X after measuring the subsystem Z by partial tracing η_1 :

$$\sigma_X^{x,z} = \frac{\rho_x + \rho_z + \langle\psi_x|\psi_z\rangle|\psi_x\rangle\langle\psi_z| + \langle\psi_z|\psi_x\rangle|\psi_z\rangle\langle\psi_x|}{2(1 + |\langle\psi_z|\psi_x\rangle|^2)}. \quad (8)$$

Once we have prepared the feature matrix (8), we can proceed with the evaluation of the operator-valued kernel. First, entanglement between X and Y is introduced by a non-separable unitary U , giving the state η_2 :

$$\eta_2 = U_{YX}(|\phi\rangle\langle\phi|_Y \otimes \sigma_X^{x,z})U_{YX}^\dagger. \quad (9)$$

The evaluation of the operator-valued kernel is then obtained by measuring the register X of η_2 :

$$K(x, z) = \text{Tr}_X[U_{YX}(|\phi\rangle\langle\phi|_Y \otimes \sigma_X^{x,z})U_{YX}^\dagger]. \quad (10)$$