

Umberto Borso
22-944-326

Quantum Advantage in Machine Learning: Investigating Entanglement and Quantum Feature Maps in Neural Networks

Semester Thesis

Master's Degree Program in Quantum Engineering
Scalable Parallel Computing Laboratory (SPCL) Group
Swiss Federal Institute of Technology (ETH) Zurich

Supervision

Dr. M. Besta
Prof. Dr. T. Hoefler
Prof. Dr. R. Renner

June 23, 2023

Abstract

This work explores different design choices of quantum neural network (QNN) models, focusing on the effect of quantum feature maps and variational circuits on model performance. Their impact on QNN models in terms of capacity and trainability is systematically evaluated using the Fisher information spectrum and the effective dimension, as proposed by Abbas et al. [2].

Furthermore, the Meyer-Wallach measure, as suggested by Sim et al. [31], is employed to quantify the entangling capability of the analysed quantum feature maps and variational circuits. Considering that entanglement is thought to play a pivotal role in achieving quantum advantage [14], it is of the utmost interest to systematically explore this aspect. This investigation aims at understanding the interplay between entanglement descriptors and QNN performance metrics, thereby laying the groundwork for a theoretical framework essential for characterising these key components in quantum machine learning.

Among the analysed Pauli feature maps, this study identifies parameter choices $k = 2$, $P_0 = Z$ and $P_{0,1} = ZZ$ as a superior design, exhibiting a high effective dimension and entangling capability, thus affording a potential quantum advantage [14]. As for variational circuits, particle-preserving designs, employing circular and scalar entanglement patterns demonstrate excellent model performance, and are expected to be well suited for devices with restricted nearest neighbor connectivity [34].

The findings highlight the crucial role of entanglement in feature map design for effective quantum encoding, while the performance of variational circuits relies on their efficient transformation of the quantum state using parameterised gate operations. These results indicate the necessity of prioritising entangling capability in the design of quantum feature maps to achieve optimal QNN models.

Contents

Abstract	ii
1 Introduction	2
2 Quantum Neural Networks	4
2.1 General QNN Structure	4
2.2 Quantum Feature Maps	5
2.3 Variational Circuits	7
3 Evaluation Metrics and Strategy	10
3.1 Fisher Information	10
3.2 Trainability: Fisher Information Spectrum	11
3.3 Capacity: Effective Dimension	11
3.4 Entangling Capability: Meyer-Wallach Measure	13
4 Evaluation Methodology	15
4.1 Calculating Fisher Information and Effective Dimension	15
4.2 Measuring the Entangling Capability	17
5 Evaluation	19
5.1 Quantum Feature Maps Analysis	19
5.1.1 Trainability	19
5.1.2 Capacity	20
5.1.3 Entangling Capability	22
5.1.4 Discussion	23
5.2 Variational Circuits Analysis	25
5.2.1 Trainability	25
5.2.2 Capacity	26
5.2.3 Entangling Capability	27
5.2.4 Discussion	28
6 Conclusion and Outlook	30
A Appendix	32
A.1 Feature Maps Circuit Diagrams	32

1 Introduction

Quantum machine learning (QML) is a research field that merges advancements in quantum computing hardware and algorithms with concepts and challenges from artificial intelligence. It aims to leverage the unique properties of quantum devices, such as superposition and entanglement, to achieve potential advantages in computation [3].

Within QML, quantum neural networks (QNNs), which are the quantum analogs of classical neural networks (CNNs), are parameterised quantum circuits with adjustable gate operations. Two principal components comprise a QNN: a quantum feature map and a variational circuit [29].

The quantum feature map is responsible for state preparation, translating classical data into quantum states. It is designed to optimize a model’s performance for a specific task and remains constant during training [30]. The variational circuit is subsequently applied, and its parameters are optimised during the training process to minimise a given loss function [10, 29]. The output is derived by executing a measurement on the quantum circuit and implementing a classical post-processing function.

However, designing parameterised circuits presents a significant challenge, considering both their structural architecture and parameterisation. Their structure and design dictate the circuit’s expressive power [9], ability to navigate the Hilbert space, and encode probability distributions more efficiently than alternative models. Highly expressive circuits, though, can lead to vanishing gradients and barren plateaus in the loss landscape [20], hindering optimal parameterisation.

This study aims to evaluate the impact of quantum feature maps and variational circuits on the performance of quantum neural networks (QNNs), a topic that has not been extensively explored in the literature [1]. Various designs are examined to assess their effects on QNN model trainability and capacity, utilising the Fisher information spectrum and effective dimension as evaluation metrics, as proposed by Abbas et al. [2].

The investigation also focuses on understanding the role of entanglement in model trainability and the detection of higher-order correlations in data. The entangling capability of different quantum feature maps and variational circuits is assessed using the Meyer-Wallach measure [31]. This methodology facilitates the examination of the relationship between entanglement descriptors and QNN performance metrics, providing a theoretical framework for the characterisation and comparison of quantum feature maps and variational circuits in the field of quantum machine learning.

The structure of a QNN is outlined in Section 2.1, while the quantum feature maps and variational circuits analysed in this study are introduced in Sections 2.2 and 2.3, respectively. Evaluation metrics are defined in Section 3. The connection between the Fisher information spectrum and the appearance of barren plateaus in the loss landscape, along with an analysis of model trainability, is discussed in Section 3.2. The effective dimension, a capacity metric, is defined in Section 3.3. The application of the Meyer-Wallach measure as an entanglement metric is described in Section 3.4. The numerical methodology used to examine different designs of feature maps and variational circuits is outlined in Section 4, and the results are presented and analysed in Section 5. The study concludes in Section 6 with a summary of the findings and potential avenues for future research.

2 Quantum Neural Networks

In this chapter, the general structure of a quantum neural network model will be outlined, including its principal components. The various quantum feature maps and variational circuits analysed in this study will be introduced, along with their key characteristics and the reasons for their selection.

2.1 General QNN Structure

A typical quantum neural network structure is depicted in Figure 2.1 and it is composed of two key components: a quantum feature map and a variational circuit [29].

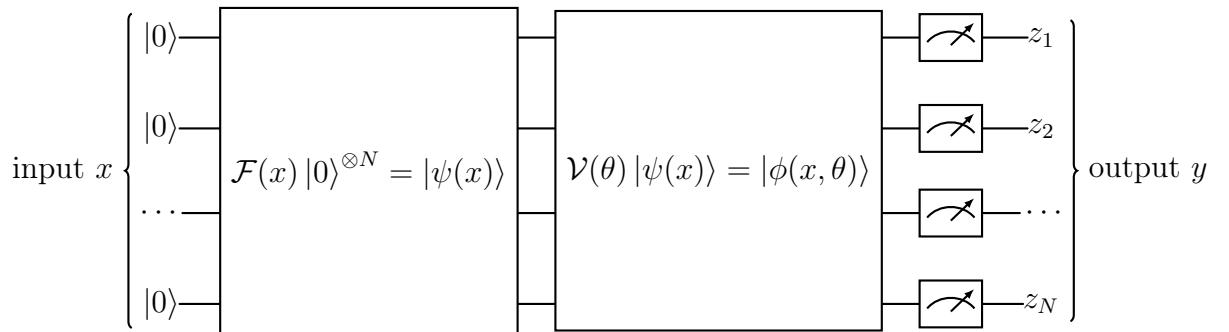


Figure 2.1: Circuit diagram showcasing the typical structure of a quantum neural network model. The classical input data x is encoded through the feature map $\mathcal{F}(x)$ and the resulting quantum state is fed to the variational circuit $\mathcal{V}(\theta)$. During the training phase the parameters θ are adjusted to minimise a loss function. The output of the model y is obtained by measuring all the qubits in the Z-basis and applying a classical-post processing function to the measurement outcome.

The quantum feature map $\mathcal{F}(x)$ is tasked with state preparation, i.e. encoding the classical input data x into a quantum state $|\psi(x)\rangle$. For a QNN with N qubits, the input state to the quantum feature map is simply $|0\rangle^{\otimes N}$, where all N qubits are in the zero basis state. The quantum feature map then applies a series of gate operations which are adjusted accordingly to the classical input data x . Therefore the output state $\mathcal{F}(x)|0\rangle^{\otimes N} = |\psi(x)\rangle$ depends directly on the classical input data x , meaning that for each classical data sample, a specific quantum state is generated [29].

The second component is the variational circuit $\mathcal{V}(\theta)$ which is applied subsequent to the feature map. This circuit, often referred to as a parameterised quantum circuit or ansatz circuit, has adjustable parameters θ that are iteratively optimised during

the training process to minimise a specified loss function. The input state to the variational circuit is $|\psi(x)\rangle$, and the output state is $\mathcal{V}(\theta)|\psi(x)\rangle = |\phi(x, \theta)\rangle$ [29].

Finally, after the application of the variational circuit, all the qubits are measured in the Z-basis and the output y is obtained by applying a classical post-processing function to the measurement outcomes $\{z_1, \dots, z_N\}$. In this work we focus on binary classification, the output labels $y = \{-1, +1\}$ are extracted using the following parity function:

$$y = \begin{cases} -1 & \text{if } z = z_1, \dots, z_N \text{ is even} \\ +1 & \text{if } z = z_1, \dots, z_N \text{ is odd} \end{cases} \quad (2.1)$$

Meaning that if in the binary output string $z = z_1, \dots, z_N$ the digit 1 occurs an even number of times, the classification label is -1 , otherwise the classification label is $+1$. This approach is inspired by [2].

2.2 Quantum Feature Maps

The quantum feature maps examined in this study fall under the category of Pauli feature maps. This class encompasses a diverse range of quantum circuits designed to encode classical data into qubit states through a series of parameterised gates, and their structure is inspired by [14]. The number of qubits in these circuits corresponds to the dimension of the input data vector. In this analysis, we focus on 4-qubit feature maps to ensure comparability with the original paper that introduced the concept of effective dimension for QNNs [2]. Additionally, due to the exponential scaling of the quantum state space with the number of qubits, limiting the analysis to 4 qubits helps to manage simulation and evaluation times effectively.

This particular category of feature maps has demonstrated good performance in classification tasks despite its relatively shallow depth. Moreover, it was selected based on the premise that to achieve a quantum advantage over classical approaches, it is necessary to employ circuits that are challenging to simulate using classical methods, as discussed in [14]. Specifically, the authors of this paper argue that evaluating inner products generated from this class of feature maps presents computational difficulties in a classical setting.

Pauli feature maps encode the classical data vector $\vec{x} \in \mathbb{R}^N$ into a quantum state generated by applying the unitary $\mathcal{U}_\Phi(\vec{x}) = U_{\Phi(\vec{x})} H^{\otimes N}$ to N qubits initialised in the zero state $|0\rangle^{\otimes N}$. Here $H^{\otimes N}$ represents the action of Hadamard gates on all

N qubits, to create a uniform superposition of all basis states. The operator $U_{\Phi(\vec{x})}$ is given by [14]

$$U_{\Phi(\vec{x})} = \exp \left(i \sum_{S \subseteq [N]} \phi_S(\vec{x}) \prod_{i \in S} P_i \right) \quad (2.2)$$

where $P = \{I, X, Y, Z\}$ is the collection of single qubit Pauli matrices, $S \subseteq [N]$ describes the connectivity of the qubits $S \in \{\binom{N}{k} \text{ combinations}, k = 1, \dots, N\}$, and $\phi_S(\vec{x})$ represents a transformation of the classical data

$$\phi_S(\vec{x}) = \begin{cases} x_0 & \text{if } k = 1 \\ \prod_{j \in S} (\pi - x_j) & \text{otherwise} \end{cases} \quad (2.3)$$

An example of the binomial combinations in set S for the case $k = 2$ and $N = 4$ qubits corresponds to the 6 ways to choose 2 elements from $\{1, 2, 3, 4\}$ which are $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$, and $\{3, 4\}$. Figure 2.2 presents an illustrative example of a Pauli feature map for the case $k = 2$, $P_0 = Z$ and $P_{0,1} = ZZ$.

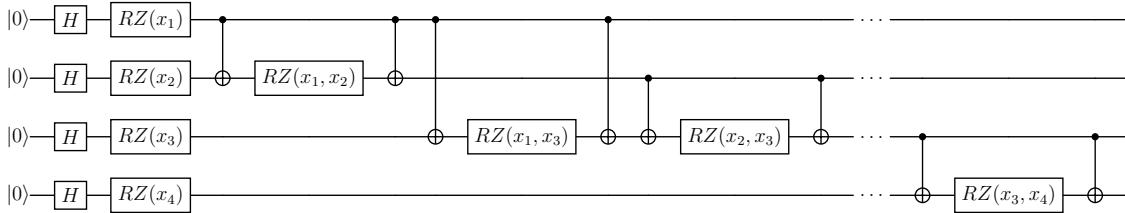


Figure 2.2: Pauli feature map FM3 for the case $k = 2$, $P_0 = Z$ and $P_{0,1} = ZZ$. The quantum circuit consists of a layer of Hadamard gates generating a uniform superposition of basis states, followed by a layer of RZ rotations. Additionally, the feature map includes a layer of entangling RZZ gates.

Table 2.1 displays the values of the parameters k and P_i used to generate the feature maps analysed in this work. For a detailed list of the circuit diagrams corresponding to these feature maps see Appendix A.

Feature map ID	k	P_0	$P_{0,1}$	$P_{0,1,2}$
FM1	1	Z		
FM2	2		ZZ	
FM3	2	Z	ZZ	
FM4	2	Z	XX	
FM5	2	Y	ZX	
FM6	3	Z		ZZZ
FM7	3	Z	YY	XXX

Table 2.1: The table presents the feature maps analysed in the study, identified by their circuit ID. Each feature map incorporates gates operating on a maximum of k qubits concurrently. The parameter P_0 denotes the type of single-qubit interactions included in each feature map, while $P_{0,1}$ and $P_{0,1,2}$ indicate the type of 2-qubit and 3-qubit interactions, respectively.

2.3 Variational Circuits

The variational circuits examined in this study are presented in Figure 2.3, and they include two categories of quantum gates: single-qubit gates, which enforce state superposition, such as Hadamard gates, and two-qubit gates, which enable the control of entanglement between qubits. These gates can be further classified into parameterised gate operations, dependent on parameters θ which are optimised during training, and fixed gate operations, which exhibit no parameter dependency and remain constant during training [29]. To achieve similar number of training parameters, the different variational circuits are repeated for a specific number of layers. In particular, shorter circuits comprising fewer gate operations are repeated for more layers, while longer circuits are repeated for fewer layers.

The selection of variational circuits includes a diverse range of entanglement patterns and parametrised gates. These circuits were inspired by previous studies, demonstrating their relevance and utility in various applications. For instance, circuit VC1, initially proposed in [22], serves as a simple circuit capable of approximating unitary 2-designs, this circuit was further explored in [20] to investigate the phenomenon of barren plateaus. Circuits VC2, VC3, VC5, VC6, and VC9 represent alternative variants of particle-preserving circuits, extensively utilised in quantum chemistry [5, 12], and differ by their entanglement pattern. Circuits VC4 and VC8 generate quantum states with real amplitudes and find applications in both variational quantum eigensolvers (VQEs) [37] and quantum neural networks [2]. Lastly, circuit VC7 represents a hardware-efficient circuit suitable for preparing trial wave

functions in variational quantum algorithms [11].

Table 2.2 provides a concise overview of these circuits, presenting essential features such as the entanglement pattern employed, the number of parameters, and the types of gates utilised.

Circuit ID	Parameters	Entanglement Pattern	Gates
VC1	40	pairwise	RY, RX, CZ
VC2	40	circular	RZ, RXX, RYY
VC3	40	full	RZ, RXX, RYY
VC4	40	full	RY, CX
VC5	39	pairwise	RZ, RXX, RYY
VC6	39	linear	RZ, RXX, RYY
VC7	40	circular	RY, Y, CX
VC8	40	linear	RY, CX
VC9	40	scalar	RZ, RXX, RYY

Table 2.2: The table presents the variational circuits analysed in the study, identified by a circuit ID. It includes information such as the number of parameters θ , the entanglement pattern, and the type of gate utilised for each variational circuit.

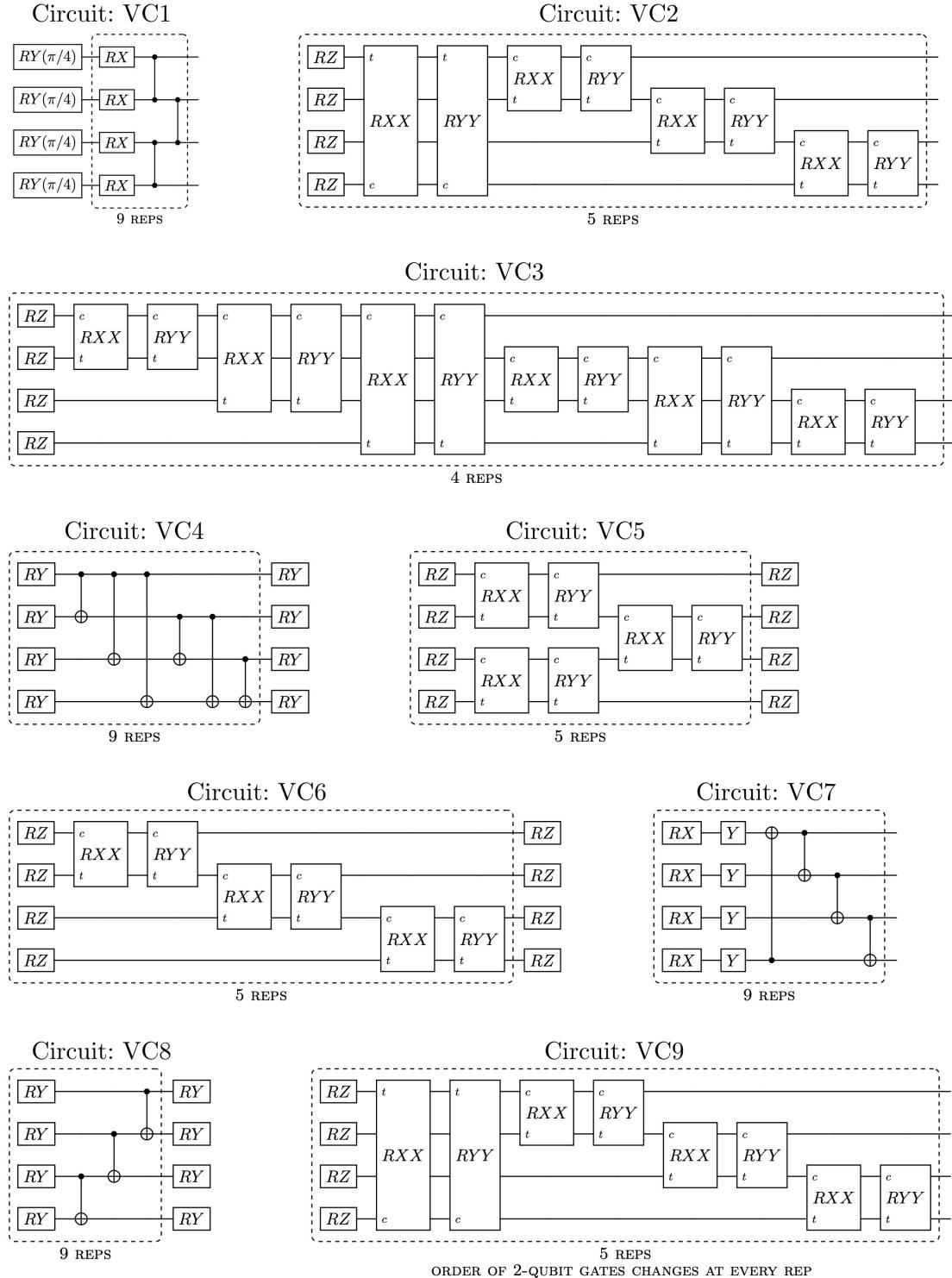


Figure 2.3: Circuit diagrams of the examined variational circuits. The dashed lines indicate gates applied for multiple iterations.

3 Evaluation Metrics and Strategy

The aim of this study is to investigate how different quantum feature maps and variational circuits impact the performance of QNN models in terms of trainability and model capacity. Additionally, it is of great interest to understand the relationship between a circuit’s ability to generate entangled states and the overall model performance.

To assess trainability and capacity, we adopt an evaluation strategy inspired by Abbas et al. [2]. Specifically, we examine the Fisher information spectrum associated with various feature maps and variational circuits to compare their effects on model trainability. The effective dimension is then used as a metric to quantify model capacity. Furthermore, the Meyer-Wallach measure is employed to estimate the entangling capability of different model structures, by following the methodology introduced by Sim et al. [31].

This chapter formally introduces the above mentioned metrics, and justifies their selection in the context of this analysis.

3.1 Fisher Information

In the context of machine learning, wherein neural networks are interpreted as statistical models, the Fisher Information plays a crucial role in understanding the model’s learning characteristics [13]. It provides a measure of the sensitivity of a model’s output distribution to changes in its parameters [32]. Formally, for a model with parameters θ , we describe the relationship of (x, y) pairs with the joint distribution $p_\theta(x, y) = p_\theta(y|x)p(x)$, where $x \in \mathcal{X} \subset \mathbb{R}^N$, $y \in \mathcal{Y}$ and $\theta \in \Theta \subset [-1, +1]^d$. The distribution $p(x)$ is a prior input distribution and $p_\theta(y|x)$ describes the input output relation for a fixed value of the parameter θ [17]. The Fisher Information matrix $F(\theta)$ is defined as the expectation of the outer product of the gradient of the log-likelihood with respect to the parameters:

$$F(\theta) = \mathbb{E}_{x,y \sim p_\theta(x,y)} [\nabla_\theta \log p_\theta(y | x) \nabla_\theta \log p_\theta(y | x)^T] \quad (3.1)$$

Where the expectation is taken over the data pairs (x, y) sampled from the joint distribution $p_\theta(x, y)$ [17].

Given that the true data distribution is often unknown in practical scenarios, the Fisher Information matrix can be approximated using the empirical Fisher Information. This approximation replaces the expectation over the true data distribution

with an average over a dataset of n samples, $\{x_i\}_{i=1}^n$, drawn from the data distribution [17]:

$$\tilde{F}(\theta) = \frac{1}{n} \sum_i \nabla_{\theta} \log p_{\theta}(y_i | x_i) \nabla_{\theta} \log p_{\theta}(y_i | x_i)^T \quad (3.2)$$

The empirical Fisher Information thus provides a practical means of estimating the Fisher Information when the true data distribution is inaccessible, and it is commonly used in the training of neural networks to guide the update of the model parameters [4].

3.2 Trainability: Fisher Information Spectrum

In a classical setting, the eigenvalue spectrum of the Fisher information provides valuable insights into the optimisation landscape of a machine learning model. The eigenvalues' magnitudes reflect the curvature of the model landscape for a given parameterisation. When there is a significant concentration of eigenvalues close to zero, the optimisation landscape tends to be mostly flat, posing challenges for training parameters using gradient-based methods [16].

This behaviour is also observed for the Hessian matrix of a model, which measures the second-order derivatives of the model's loss function with respect to its parameters, and is closely related to the Fisher information [18]. A well-conditioned Fisher information, with non-zero eigenvalues, indicates a landscape that is conducive to efficient learning. In such cases, gradients and updates during training lead to substantial progress. However, the occurrence of barren plateaus, flat regions in the loss landscape where gradients become vanishingly small, can lead to inefficient learning [16].

As shown in a previous study [2], the spectrum of the Fisher Information matrix in the quantum regime can also be associated with barren plateaus. In line with the approach proposed by Abbas et al., the trainability of QNN models constructed with various feature maps and variational circuits will be evaluated by analysing the Fisher information spectrum.

3.3 Capacity: Effective Dimension

The effective dimension of a statistical model provides an insightful measure of the model's complexity, specifically in terms of the number of parameters that are ef-

fectively being learnt from the data. The objective of the effective dimension is to approximate the size occupied by a model within the model space, which encompasses all conceivable functions within a given model class [6]. In this context, the Fisher Information matrix assumes the role of a metric, aiding in the estimation process. The work by Berezniuk et al. [6] provides a useful definition of the effective dimension. It defines the effective dimension of a statistical model \mathcal{M}_θ with parameters θ living in a d -dimensional parameter space Θ , taking into account a dataset consisting of more than one ($n > 1$) data samples. The effective dimension is formally defined as follows:

$$\dim_n(\mathcal{M}_\theta) := 2 \frac{\log \left(\frac{1}{V_\Theta} \int_\Theta \sqrt{\det \left(\text{Id}_d + \frac{n}{2\pi} \hat{F}(\theta) \right)} d\theta \right)}{\log \frac{n}{2\pi}} \quad (3.3)$$

where $V_\Theta := \int_\Theta d\theta$ is the volume of the parameter space Θ , and $\hat{F}(\theta)$ is the normalised Fisher information matrix:

$$\hat{F}_{ij} := d \frac{V_\Theta}{\int_\Theta \text{tr } F(\theta) d\theta} F_{ij} \quad (3.4)$$

These two normalisation factors ensure that the effective dimension $\dim_n(\mathcal{M}_\theta)$ is scaling invariant in θ .

In contrast to other definitions of effective dimension, such as those based on the Gardner phasespace approach [24] or the VC dimension [35], which are more applicable in scenarios where the number of data samples n approaches infinity, the provided definition takes into account the practical limitation of having only a finite-size dataset. This unique approach recognizes that real-world experiments rely on finite-sized data and aims to provide a more accurate estimation of the dimension of the true model space.

In essence, the size of the available data determines the resolution at which the model is observed, thus influencing its dimension. This acknowledgment highlights the importance of considering the data's finite size when estimating the effective dimension, as it provides a better understanding of the true model space observed in experimental settings.

Moreover, in the work conducted by Amira Abbas et al. [2], an additional parameter γ is incorporated into the definition of the effective dimension. This inclusion enables them to establish an upper bound on the generalisation error, based on the effective dimension. The generalisation error bound serves as a common metric to

quantify the performance of a machine learning model on unseen data. Thus, the effective dimension demonstrates its relevance as a capacity measure by providing a generalisation bound, validating its utility in assessing model performance.

In this study, the capacity of different models is compared using the effective dimension metric. According to Equation 3.3, it can be observed that the effective dimension reaches its maximum value more rapidly when the Fisher information spectrum is evenly distributed on average [2]. Consequently, QNN models with higher capacity, indicating superior performance, will have a higher effective dimension for a given input dataset size.

3.4 Entangling Capability: Meyer-Wallach Measure

An essential objective of this study is to investigate the role of entanglement in quantum neural networks and, specifically, how the entangling capability of a model relates to its capacity and trainability. Within the domain of variational hybrid quantum algorithms, generating highly entangled states using low-depth circuits offers potential advantages. These include the efficient representation of solution spaces for tasks such as ground state preparation or data classification, as well as capturing non-trivial correlations within quantum data [28, 15]. However, it has also been observed that excessive entanglement between visible and hidden units in a quantum neural network can impede the learning process [19]. Consequently, conducting a more comprehensive analysis of how entanglement influences the performance of quantum models would prove valuable in identifying optimal model structures.

The Meyer-Wallach measure serves a useful tool in this analysis, as it allows to evaluate the entangling capability of a parametrised quantum circuit, meaning its ability to generate entangled states [21]. This measure was selected due to its computational simplicity and its demonstrated effectiveness in real-world, non-simulated scenarios, without the need for full quantum state tomography [8].

The Meyer-Wallach measure, quantifying the entanglement of a quantum state $|\psi\rangle$, can be defined as:

$$Q(|\psi\rangle) = 2 \left(1 - \frac{1}{n} \sum_{k=0}^{n-1} \text{Tr} [\rho_k^2] \right) \quad (3.5)$$

where ρ_k is the reduced density operator of qubit k , and $\text{Tr} [\rho_k^2]$ can be inter-

preted as the purity of the state of qubit k [8]. This expression provides a clear interpretation of multi-particle entanglement, representing it as the average entanglement between each qubit and the remaining system. The measure Q exhibits a linear relationship with the average purity of individual qubits, and it can be verified that it fulfills the following properties: (i) $0 \leq Q(|\psi\rangle) \leq 1$. Specifically $Q(|\psi\rangle) = 0$ if and only if $|\psi\rangle$ is a product state, and $Q(|\psi\rangle) = 1$ for certain entangled states, such as GHZ states, (ii) $Q(|\psi\rangle)$ is invariant under unitaries U_j acting on single qubits [8, 31]. The Meyer-Wallach measure has proven to be a valuable tool in various quantum information applications, such as characterising entangled states in quantum error correcting codes [21] and quantum phase transitions [33]. It has also been successfully employed to track the evolution of entanglement in Grover's algorithm [21]. Due to its effectiveness in quantifying the ability of a parameterised quantum circuit to generate entangled states, the Meyer-Wallach measure is well-suited for assessing the entangling capability of quantum feature maps and variational circuits.

4 Evaluation Methodology

This study focuses on the examination of the feature maps and variational circuits introduced in Section 2.2 and Section 2.3, respectively. The goal is to conduct an in-depth investigation into their effects on the trainability, capacity, and entangling capability of QNN models.

As previously outlined, the Fisher information spectrum is used to analyse model trainability, the effective dimension as an estimate of model capacity and the Meyer-Wallach measure to assess the entangling capability of different model structures.

To ensure a comprehensive evaluation of the feature maps, multiple quantum models are generated by combining each feature map with the same variational circuit, specifically VC8. This approach isolates the impact of the feature map as the primary factor contributing to performance variations, with VC8 chosen for its simplicity which allows for an efficient computation of evaluation metrics.

Likewise, the performance assessment of QNNs derived from different variational circuits involves pairing each circuit with the same feature map, FM3. As discussed later in Section 5.1, FM3 is recognised as the most versatile feature map among those examined, showcasing favorable capacity, trainability, and entangling capability. Additionally, the relatively shallow depth of FM3 enables efficient simulations within reasonable time frames.

To ensure a fair comparison across models, several common characteristics are maintained. These include a comparable number of trainable parameters, either $d_\theta = 40$ or 39 , consistent input dimensions of $d_{input} = 4$, and output dimensions of $d_{output} = 2$. This allows for a more balanced evaluation through the effective dimension metric, which is dependent on the number of parameters when the available training data samples are limited [6].

4.1 Calculating Fisher Information and Effective Dimension

In order to evaluate the effective dimension and estimate the Fisher information spectrum of the QNN models, a Monte Carlo sampling approach is utilised, as suggested in the study by Abbas et al. [2]. This approach involves iteratively generating random parameters for the QNN models, allowing for the exploration of the entire parameter space and input space. At each iteration, the empirical Fisher information matrix is estimated, and the estimates from all iterations are averaged to obtain a reliable approximation of the true value. The resulting Fisher information matrix is then normalised and used to compute the eigenvalue spectrum and the effective dimension, providing valuable insights into the model's capacity

and trainability.

The process involves generating 100 sets of input data samples and 100 sets of parameter samples. The input data samples, denoted as \mathbf{x}_i , are drawn from a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. These samples are then encoded into the quantum statespace using the selected quantum feature maps. The parameters, represented as $\boldsymbol{\theta}_i$, are uniformly sampled from the parameter space $\Theta = [-1, 1]^{d_\theta}$.

Next, a specific pair consisting of a feature map and variational circuit is selected for analysis. For each specific set of input samples, the classical data \mathbf{x}_i is encoded using the chosen quantum feature map, transforming the initial quantum state $|0\rangle^{\otimes N}$ into the output state $|\psi(\mathbf{x}_i)\rangle$. The variational circuit is then parametrised with a specific set of parameter samples $\boldsymbol{\theta}_i$ and applied to the output state $|\psi(\mathbf{x}_i)\rangle$, resulting in the final state $|\phi(\mathbf{x}_i, \boldsymbol{\theta}_i)\rangle$. Z-basis measurements are performed on all qubits, and the resulting classical bitstring undergoes a parity post-processing function to generate the final classification outcome. This entire process, which corresponds to the model forward pass, is repeated for each combination of input samples and parameter samples, yielding the model output vector.

In the following step, the backward pass is executed to obtain the model gradient vector. To calculate the individual gradients of the entire quantum circuit with respect to the gate parameters $\boldsymbol{\theta}_i$, the parameter shift rule is applied [27]. This rule involves running the original quantum circuit twice while shifting a single gate parameter to determine the corresponding gradient component.

The average Jacobian matrix is computed using the model output and gradient vectors, providing an empirical estimate of the Fisher information matrix using Equation 3.2. The Fisher information is subsequently normalised using Equation 3.4.

The eigenvalue distribution of the normalised Fisher information is then calculated and plotted as a histogram for each QNN model under consideration. These distributions are compared in the analysis of model trainability.

Finally, utilising the normalised Fisher information matrix, the effective dimension of the model is calculated for 11 different values of the dataset size $n \in [10^3, 10^6]$ using Equation 3.3. The resulting effective dimensions of the generated QNNs are compared in the analysis of model capacity.

It is important to acknowledge that the metrics employed to assess the different QNNs - the effective dimension as an indicator of capacity and the Fisher information spectrum revealing insights into model trainability - are influenced by the number

of data samples and parameter samples utilised during the Monte Carlo sampling process. For this study, the number of data samples and parameter samples has been set to 100. A sensitivity analysis conducted in [2] confirms that this sample size is adequate to yield stable and reliable results in calculating the effective dimension for a model with a specific number of trainable parameters $d_\theta = 40$.

4.2 Measuring the Entangling Capability

In order to examine the impact of entanglement on QNNs and its relation to model performance, the entangling capability of the considered feature maps and variational circuits is evaluated using the Meyer-Wallach measure. This measure allows for an assessment of their ability to generate entangled states, which is then correlated with key performance metrics such as effective dimension and Fisher information spectrum.

Since quantum feature maps and variational circuits are parametrised circuits, it is necessary to estimate their ability to generate quantum states over the entire parameter space. To do so a Monte Carlo sampling approach is employed, following the methodology outlined by Sim et al. [31]. This involves generating 5000 sets of random input samples \mathbf{x}_i and parameter samples $\boldsymbol{\theta}_i$ as described above.

The feature maps are initially prepared in the zero basis state $|0\rangle^{\otimes N}$, where $N = 4$ represents the number of qubits. These maps are responsible for encoding the classical input samples \mathbf{x}_i into the corresponding quantum states $|\psi(\mathbf{x}_i)\rangle$.

On the other hand, the variational circuits, parameterised by the samples $\boldsymbol{\theta}_i$, generate output states $|\phi(\cdot, \boldsymbol{\theta}_i)\rangle$. The entangling capability of these circuits is assessed under three distinct input scenarios:

1. The variational circuits are initialised in the zero basis state $|0\rangle^{\otimes N}$.
2. The input states to the variational circuits are generated by randomly parameterising the feature map FM5 and using its output states as the input.
3. The input states to the variational circuits are generated by randomly parameterising the feature map FM3 and using its output states as the input.

Approaches 2 and 3 are considered meaningful as they reflects the typical quantum machine learning scenario where the input to a variational circuit corresponds to the output of an encoding feature map. Furthermore the third approach is particularly relevant for this analysis since the trainability and capacity analyses were

performed using FM3 in the studied QNN models. Therefore, the choice of using FM3 as the input state aligns with the context of the work.

Finally, the entanglement of the generated output states is evaluated using the Meyer-Wallach measure, and averaged over the inputs and parameter sets. The resulting entangling capabilities of the feature maps \mathcal{F} and of the variational circuits \mathcal{V} are therefore defined as:

$$EC_{\mathcal{F}} = \frac{1}{|S_x|} \sum_{\boldsymbol{x}_i \in S_x} Q(|\psi(\boldsymbol{x}_i)\rangle) \quad (4.1)$$

$$EC_{\mathcal{V}} = \frac{1}{|S_\theta|} \sum_{\boldsymbol{\theta}_i \in S_\theta} Q(|\phi(\cdot, \boldsymbol{\theta}_i)\rangle) \quad (4.2)$$

where $S_x = \{\boldsymbol{x}_i\}$ is the collection of sets of randomly sampled parameters fed to the feature maps, and $S_\theta = \{\boldsymbol{\theta}_i\}$ is the collection of sets of randomly sampled parameters fed to the variational circuits [31]. By employing this measure, a quantum circuit that exclusively generates product states will yield an entangling capability score of 0. Conversely, a circuit that consistently produces highly entangled states will be associated with a score approaching 1.

5 Evaluation

In this chapter, the results of the numerical experiments are presented, focusing on the analysis and comparison of various feature maps and variational circuits using the previously defined metrics.

To evaluate model trainability, the Fisher information spectra of the QNN models are plotted and compared. A concentration of eigenvalues near zero, with few large eigenvalues, indicates flat regions in the model landscape, which implies poor trainability.

For assessing model capacity, the estimated effective dimensions are also plotted. Specifically, the effective dimension calculated for the largest dataset size considered, $n = 10^6$, is used to compare the models' performance in terms of capacity.

Additionally, the average Meyer-Wallach measure is employed to estimate and compare the entangling capability of different feature maps and variational circuits.

5.1 Quantum Feature Maps Analysis

This section is dedicated to the analysis of the quantum feature maps introduced in Section 2.2. The primary goal is to evaluate the impact of the various structures of these feature maps on the capacity and trainability of the resulting quantum model, as well as their ability to generate entangled states.

5.1.1 Trainability

Figure 5.1 presents the normalised Fisher information spectrum for the QNNs generated using different feature maps, providing insight into their impact on the trainability of the resulting models. Each model's Fisher information matrix is computed, and the distribution of the eigenvalues is presented as a histogram.

As previously discussed in Section 3.2, a Fisher information spectrum characterised by a high density of small eigenvalues and a scarcity of large eigenvalues indicates the presence of barren plateaus within the parameter space. These barren plateaus pose challenges to parameter optimisation using gradient descent techniques, thereby affecting the trainability of the model.

The visual analysis of the Fisher information spectrum reveals distinct characteristics for various feature maps. FM1 and FM2 show a considerable concentration of eigenvalues near zero, an undesirable property indicative of lower model trainability. Conversely, FM3, FM4, FM5, FM6 and FM7 exhibit a more uniform distribution of Fisher information eigenvalues, suggesting that these feature maps can contribute

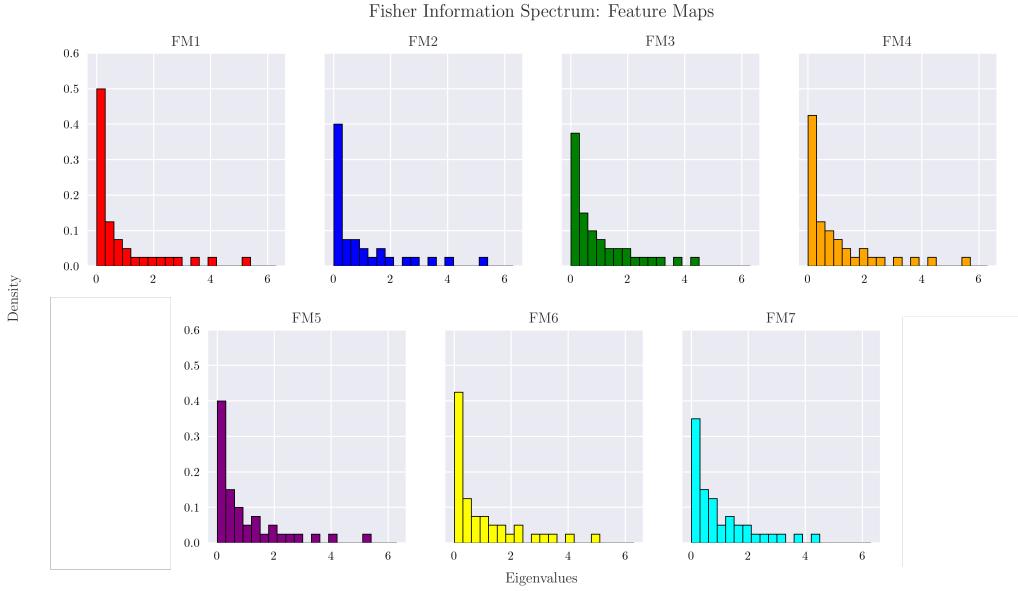


Figure 5.1: Fisher information spectrum of the analysed quantum feature maps. Every feature map described in table 2.1 is combined with the variational circuit VC8 to generate a unique quantum neural network model. The Fisher information matrix is computed using Monte Carlo sampling, and the average eigenvalue distribution is plotted for each model.

to improved training performance in resulting models.

5.1.2 Capacity

Figure 5.2 displays the normalised effective dimensions corresponding to different choices of feature maps. The normalised effective dimension of each model, given by $\dim_n(\mathcal{M}_\theta)/d_\theta$, where d_θ represents the number of θ parameters in the model, is plotted against the input dataset size n .

Feature maps with higher effective dimensions indicate models with greater capacity as explained in Section 3.3. This relationship arises from the close connection between the effective dimension and the Fisher information spectrum. A well-performing model, characterised by a more evenly distributed spectrum of eigenvalues, leads to a normalised effective dimension that converges faster toward 1.

Consistent with the measured Fisher information spectra, it can be observed that FM1 and FM2 exhibit the lowest effective dimensions across all ranges of finite data considered, aligning with expectations. In particular they achieve maximum effective dimension values of 0.814 and 0.670 respectively. Table 2.1 reveals that these feature maps solely employ one type of gate. For instance, FM1 exclusively utilizes Z

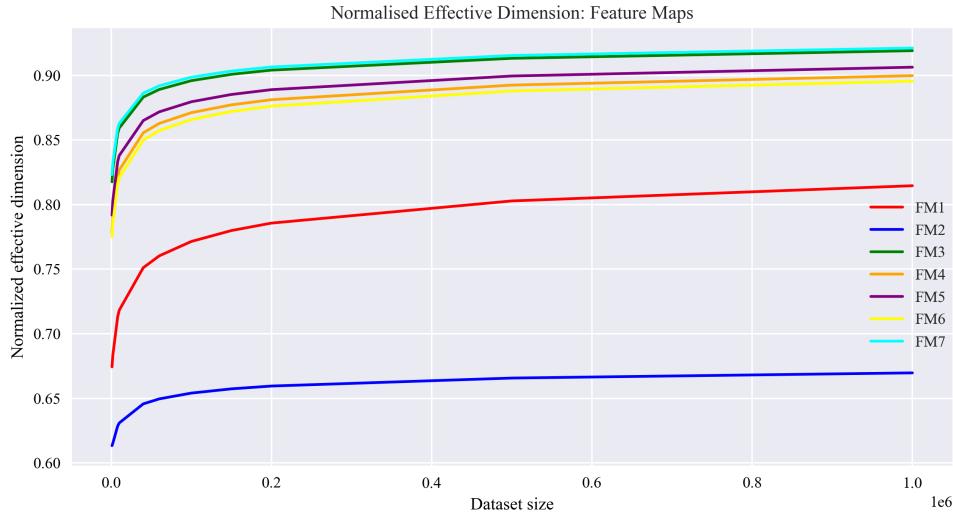


Figure 5.2: Normalised effective dimension of different quantum feature maps. Every feature map described in table 2.1 is combined with the variational circuit VC8 to generate a unique quantum neural network model. The normalised effective dimension of each of these models is calculated and plotted to facilitate comparison.

type rotations, which severely restricts the circuit's ability to generate entangled states, as clarified later in the entangling capability analysis, Section 5.1.3. Similarly, FM2 solely employs ZZ type rotations, which are two-qubit gates allowing for the encoding of classical information only through joint interactions of two classical data entries. The absence of single-qubit gates prevents the independent encoding of individual data entries, resulting in diminished model capacity.

FM6 ranks as the third-worst performing feature map, achieving a maximum effective dimension of 0.895. Although this feature map incorporates both one-qubit and three-qubit gates, the absence of two-qubit gates appears to significantly impact its performance, particularly in terms of effective dimension. Comparatively, FM3, FM4, and FM5 replace three-qubit gates with two-qubit gates and achieve higher effective dimensions (0.919, 0.899 and 0.906 respectively) for the same dataset size compared to FM6. Among the various gate combinations within these three feature maps, the best performing configuration is FM3, which is solely employing Z-type interactions at both single-qubit and two-qubit levels.

Finally, the overall best performing feature map is FM7, which achieves an effective dimension of 0.921. This feature map encompasses all previously mentioned interactions - one-qubit, two-qubit, and three-qubit gates - and the comprehensive inclusion of these gates allows for a more robust encoding of classical data into the resulting quantum states, resulting in a model with higher capacity.

5.1.3 Entangling Capability

Figure 5.3 illustrates the entangling capabilities of the examined feature maps, estimated through Monte Carlo sampling. An entangling capability of 0 implies the inability to generate entangled states, while a value close to 1 indicates the capacity to produce highly entangled states.

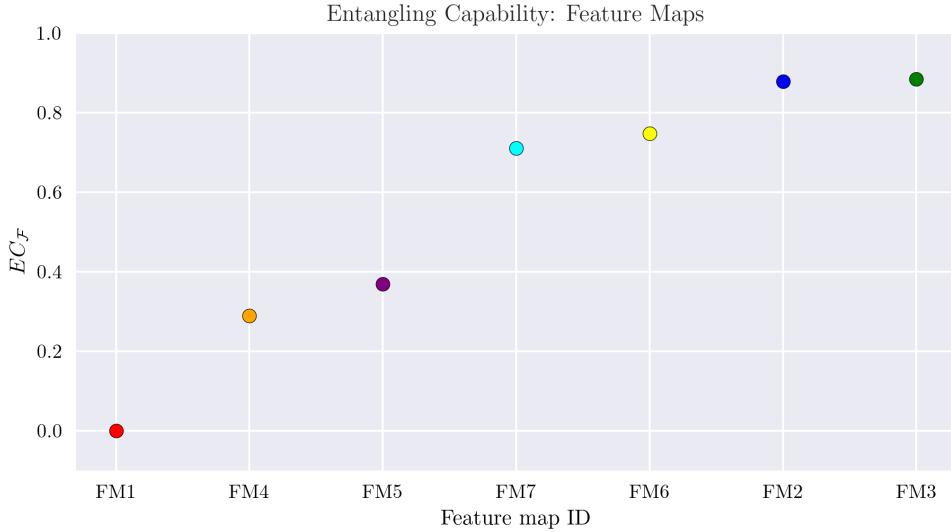


Figure 5.3: Entangling capability of different quantum feature maps initialised in state $|0\rangle^{\otimes N}$.

FM1, composed exclusively of single-qubit gates, demonstrates an entangling capability of 0.000. This outcome aligns with expectations, as FM1 exclusively utilizes single-qubit Z rotations without any inclusion of two-qubit interactions, thus hindering the generation of entanglement. It is well-established that the presence of multi-qubit gates is necessary to introduce entanglement in a quantum circuit [7].

In contrast, FM2, exclusively employing two-qubit ZZ rotations, demonstrates the ability to generate highly entangled states, with an entangling capability of 0.878.

The highest entangling capability is observed in FM3 (0.884), which employs Z-type interactions at both the single-qubit and two-qubit levels.

Similarly, FM4 and FM5 also include both single and two-qubit gates, but with different combinations of Pauli matrices. These feature maps exhibit lower entangling capabilities compared to FM3, indicating that the most effective combination of gates for generating highly entangled states involves Z and ZZ interactions.

Lastly, FM6 and FM7, incorporating up to three-qubit interactions, demonstrate entangling capabilities of 0.748 and 0.710, respectively, which are still lower than

FM2 and FM3.

5.1.4 Discussion

This analysis yields two important conclusions. Firstly, it is evident that both single-qubit and two-qubit gates are crucial for constructing meaningful quantum feature maps. Single-qubit gates enable the independent encoding of individual data entries, while two-qubit gates introduce fundamental entanglement interactions between data pairs, which are necessary to create feature maps that are hard to simulate classically and can lead to quantum advantage [14].

Feature map FM1, which solely consists of single-qubit gates, is unable to generate entangled states, making it easy to simulate classically. As a result, it offers no quantum advantage and exhibits a reduced effective dimension, consistent with findings in the literature [2, 14].

On the other hand, the presence of multi-qubit gates alone is insufficient for effective encoding schemes. FM2, which only includes two-qubit gates, demonstrates high entangling capability yet performs poorly in terms of model capacity.

When combining single and two-qubit gates, it is possible to generate feature maps with high capacity and entangling capability. In this context, Z type interactions are observed to be the optimal choice for both categories of gates. FM3 serves as an exemplary feature map in terms of achieving these desirable characteristics. It is worth noting that Z type gates are commonly utilised in experimental demonstrations, particularly in superconducting architectures. Variational quantum algorithms have often employed layers of controlled- Z rotations in various experimental studies [23, 25].

While three-qubit gates can facilitate the encoding of higher ordered correlations in the data, as proved by the capacity performance of FM7, their practical implementation in quantum computers can pose challenges. In particular, when only standard single and two-qubit gate sets are available, the compilation of these three-qubit gates results in additional gate overhead and significantly increases the circuit depth [36, 38]. Moreover, the increase in model capacity achieved by incorporating three-qubit gates is limited, as demonstrated by the almost identical effective dimensions of FM3 and FM7.

Secondly, it is observed that a higher entangling capability, which implies a more challenging classical simulation of a circuit, can assist in selecting a feature map with better model capacity. However, the ability to generate entangled states alone is not sufficient for meaningful classical data encoding in the quantum feature

space. In the case of Pauli feature maps with $k = 2$, which include both single and two-qubit interactions (FM3, FM4, and FM5), the entangling capability increases with the effective dimension. Nevertheless, as mentioned earlier, entanglement alone does not guarantee meaningful classical data encoding, as exemplified by FM2.

Considering these key findings and acknowledging the higher expected circuit depth for feature maps involving three-qubit gates, it can be concluded that the more versatile feature map among those analysed in this study is FM3. This feature map exclusively incorporates single and two-qubit gates, resulting in a relatively small circuit depth. Moreover, FM3 exhibits a higher effective dimension compared to most other feature maps, ultimately enabling the generation of highly entangled states, which may be necessary for specific applications.

5.2 Variational Circuits Analysis

This section focuses on the analysis of the variational circuits presented in Figure 2.3. The main objective is to evaluate how the different structures of these variational circuits impact the capacity and trainability of the resulting quantum model, as well as their ability to generate entangled states.

5.2.1 Trainability

To investigate the impact of different variational circuits on the trainability of QNNs, the Fisher information spectrum is plotted (Figure 5.4) and analysed.

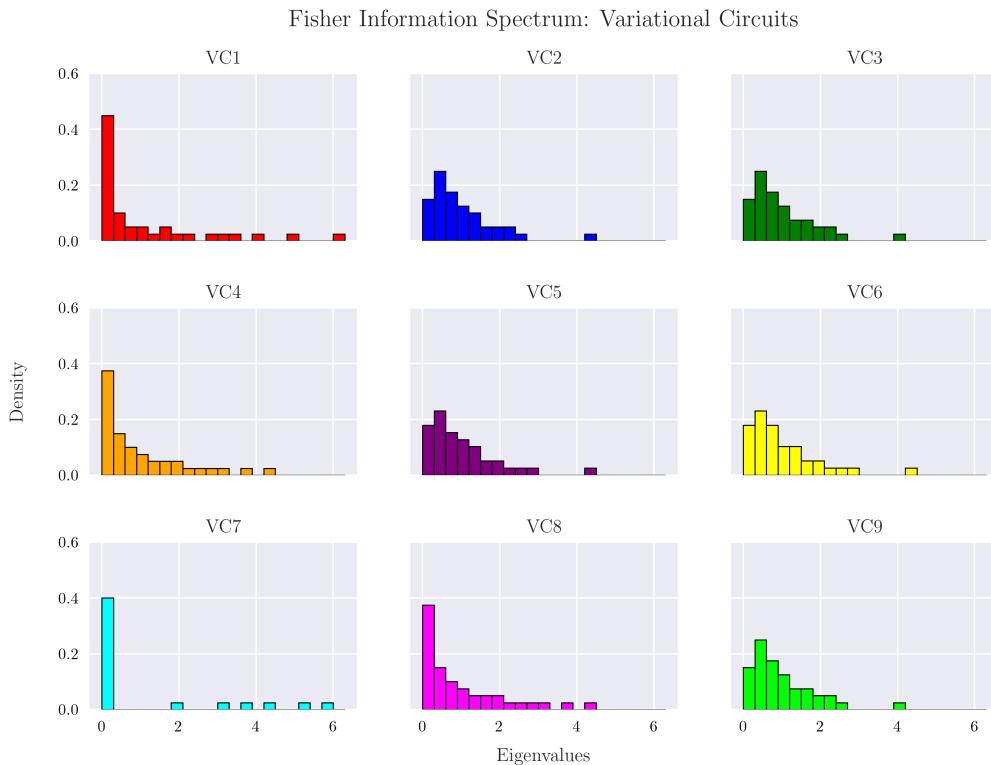


Figure 5.4: Fisher information spectrum of the analysed variational circuits. Every variational circuit described in table 2.2 is combined with the feature map FM3 to generate a unique quantum neural network model. The Fisher information matrix is computed using 100 times using different parameter samples, and the average eigenvalue distribution is plotted for each model.

The Fisher information spectrum of VC1 and VC7 exhibits a concentration of eigenvalues around zero, with only a few large eigenvalues, indicating the presence of barren plateaus, which can lead to vanishing gradients in the parameter space and suboptimal training. Therefore, VC1 and VC7 are expected to have reduced model

performance. On the other hand, VC4 and VC8 demonstrate a more evenly distributed spectrum of eigenvalues, suggesting better trainability. Additionally, VC2, VC3, VC5, VC6, and VC9 exhibit similar distributions of eigenvalues with a smaller concentration near zero. These circuits are expected to generate improved training landscapes for parameter optimisation, leading to high performance in machine learning tasks. This behavior is further reflected in the capacity analysis conducted in the following section, as well as in the observed effective dimensions.

The similarity in the eigenvalue spectra observed between VC4 and VC8 can be attributed to their common characteristic as variations of real amplitude circuits. In contrast, VC2, VC3, VC5, VC6, and VC9 represent different variations of particle-preserving circuits. Although these circuits utilize the same type of gates, they exhibit distinct entanglement patterns, as shown in Table 2.2. The similarity in eigenvalue distributions suggests that the specific entanglement patterns, generated by different combinations of two-qubit gates, do not significantly impact the resulting parameter landscape of the quantum neural network.

5.2.2 Capacity

In this section, the capacity of the analysed quantum neural networks is compared in terms of their effective dimension. The normalised effective dimension is calculated as previously described, and the results are presented in Figure 5.5.

It is observed that VC1 and VC7 exhibit the lowest effective dimensions across the entire range of dataset sizes considered, indicating a reduced model capacity.

Among the variational circuits, the particle-preserving circuits demonstrate the best performance in terms of effective dimension. Specifically, VC2, VC3, and VC9 exhibit nearly identical trends and achieve the highest effective dimensions (0.966, 0.967 and 0.965), while VC5 and VC6 show slightly lower capacities (0.926 and 0.927 respectively). Examining the structural differences between these circuits, as elucidated in Table 2.2, it is evident that the *full*, *circular*, and *scalar* entanglement patterns lead to superior performance. The *scalar* pattern is very similar to the *circular* pattern, differing only in the order of two-qubit gates in successive repetitions of the circuit. In contrast, the *pairwise* and *linear* entanglement patterns result in lower effective dimensions.

Following the particle-preserving circuits, the real-amplitude circuits rank immediately after. VC4, with a *full* entanglement pattern, and VC8, with a *linear* entanglement pattern, exhibit very similar effective dimensions of 0.917 and 0.919 respectively, indicating comparable levels of capacity.

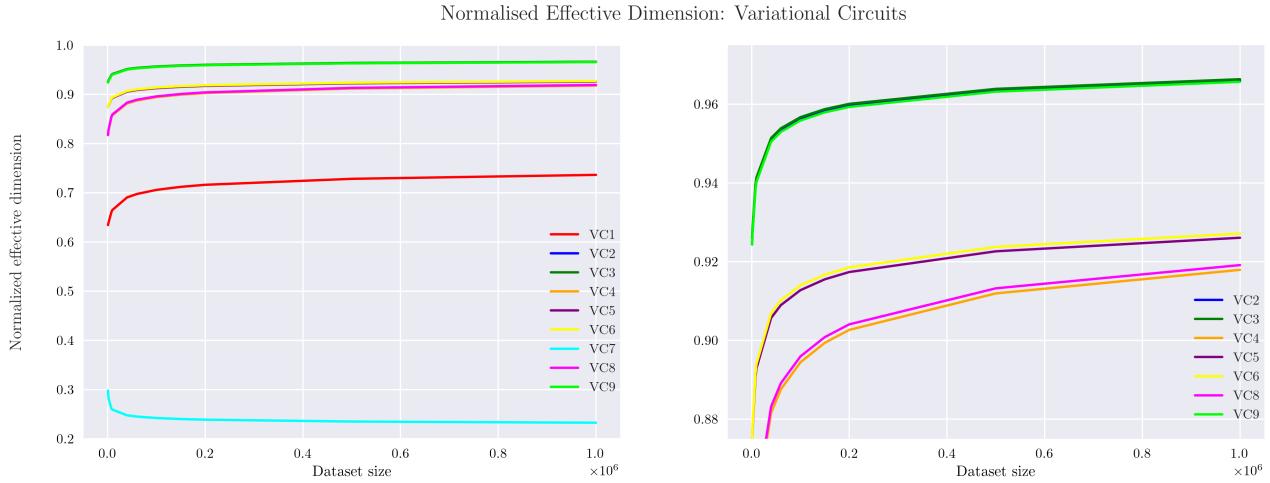


Figure 5.5: Normalised effective dimension of different variational circuits. Each variational circuit, as described in Table 2.2, is paired with the FM3 feature map to generate a unique quantum neural network model. (*left*) The normalised effective dimension is then calculated and plotted for each model, allowing for easy comparison. (*right*) Additionally, a zoomed plot is provided to better visualize the performance of the top-performing variational circuits.

These results align with the previous analysis of the Fisher information spectra, as expected.

5.2.3 Entangling Capability

Figure 5.6 illustrates the entanglement capability of the analysed variational circuits. As described in Section 4.2 three different input scenarios are considered.

When considering the first scenario with input state $|0\rangle^{\otimes N}$, VC7 exhibits the highest entangling capability with an estimated value of 0.592. This is followed by VC8, VC1, and VC5 with entangling capabilities of 0.394, 0.393, and 0.330, respectively. Notably, the particle-preserving circuits (VC2, VC3, VC5, VC6, and VC9) have an entangling capability of 0.000 when applied to the zero basis state.

In the second input scenario, where the variational circuits are applied to the output of feature map FM5, all circuits show improved entangling capabilities surpassing 0.511. VC7 remains the best at generating entangled states, while the particle-preserving circuits demonstrate a significant improvement in performance, reaching the level of VC1.

In the third input scenario, where the variational circuits are applied to the output of feature map FM3, all the circuits demonstrate greater entangling capabilities in the range of 0.851 to 0.832. This behavior can be attributed to the fact that

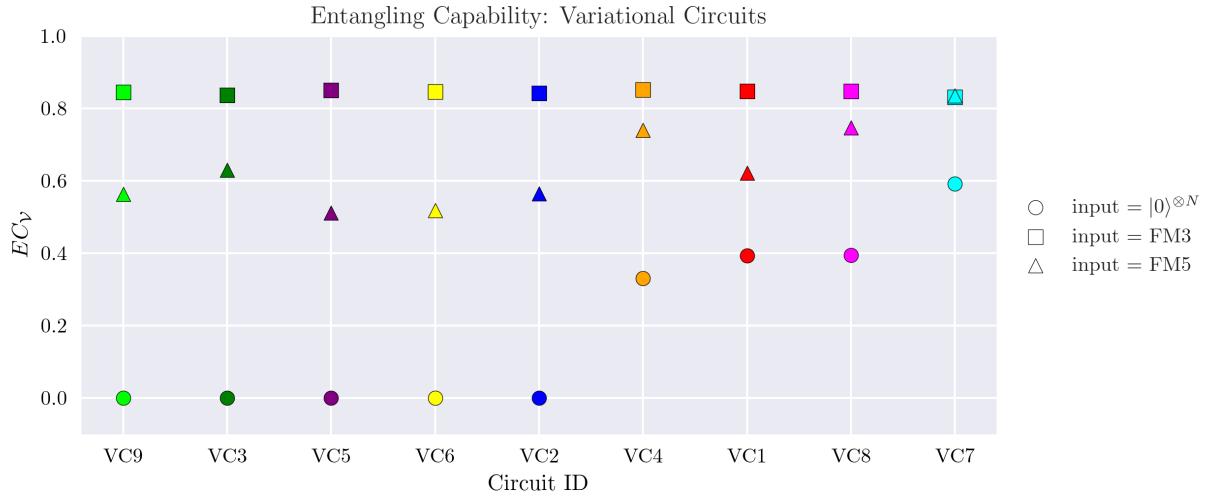


Figure 5.6: Entangling capability of the variational circuits described in Table 2.2. The entangling capability is evaluated and plotted for three different input scenarios: when the initial state is $|0\rangle^{\otimes N}$ (dots), when it corresponds to the output of a randomly parameterised FM5 feature map (triangles) and FM3 feature map (squares).

FM3 inherently possesses a higher entangling capability (0.884) compared to FM5 (0.710), as seen in Figure 5.3.

This analysis reveals that when considering the overall entangling capability of QNN models, the factor mostly affecting the ability to generate entangle states is the quantum feature map. It demonstrates that variational circuits with initially no entangling capability, such as the particle-preserving circuits, can generate quantum models with entangling capabilities above 0.800 when paired with highly-entangling feature maps like FM3.

5.2.4 Discussion

Based on the numerical analysis conducted in this study, it can be concluded that the particle-preserving circuits exhibit the best performance in terms capacity and trainability of QNN models. These circuits demonstrate an evenly distributed Fisher information spectrum, which helps avoid barren plateaus in the loss landscape and indicates optimal trainability. Specifically, the entanglement patterns *full*, *circular*, and *scalar* are found to be the most effective based on the effective dimension analysis.

Considering the prevalent use of superconducting circuit technology in quantum computing hardware, which typically features nearest neighbor connectivity

between qubits, implementing the *full* entanglement pattern with "all-to-all" interactions between qubits may be challenging in the short term. This pattern would require additional swap operations during the execution of the quantum model [34]. Consequently, considering the current state of quantum computing technology, the more versatile circuits identified are VC2 and VC9, which employ *circular* and *scalar* entanglement patterns that are better suited for implementation on devices limited to nearest neighbor connectivity.

Following the particle-preserving circuits, the real-amplitude circuits (VC4 and VC8) demonstrate strong model performance, making them a viable alternative in scenarios where a limited set of gates can be efficiently implemented. These circuits solely consist of single-qubit parametrised gates, specifically *RZ* rotations, while the other gates are fixed Hadamards and CNOTs. Despite their simplicity, these circuits are able to achieve competitive results and provide effective parameterisation options.

Interestingly, the analysis reveals that the entangling capability of the variational circuits is not directly correlated with their performance in terms of model trainability and capacity. This discrepancy arises from the fact that variational circuits are typically applied to the output states of an encoding feature map, rather than the zero basis state $|0\rangle^{\otimes N}$.

Even if a variational circuit lacks the ability to generate entangled states when initialised in the zero basis state, it can still achieve good model performance by operating on the output states of the encoding feature map, which can be designed to generate highly entangled states.

Overall, the analysis demonstrates that the quantum feature map primarily influences the entangling capability of QNN models. It emphasizes that variational circuits with initially no entangling capability, such as the particle-preserving circuits, can generate quantum models with entangling capabilities above 0.800 when paired with highly entangling feature maps like FM3.

Therefore, it can be concluded that entanglement plays a crucial role in feature map design to create effective encodings that fully leverage the quantum feature space and provide quantum advantage by generating hard-to-simulate feature maps [14]. On the other hand, when it comes to variational circuits, their effectiveness lies in their ability to efficiently transform the quantum state generated by the feature map using parameterised gate operations. Hence, entangling capability should be primarily considered in feature map design rather than variational circuits.

6 Conclusion and Outlook

This work aims to investigate the impact of quantum feature maps and variational circuits on the performance of quantum neural network (QNN) models. This area has not been well-explored, especially in terms of the entangling capability of feature maps and variational circuits [1]. To address this, we compare different designs and analyze their effects on QNN model trainability and capacity, using the Fisher information spectrum and effective dimension as metrics, as proposed by Abbas et al. [2].

To evaluate the entangling capability of various quantum feature maps and variational circuits, we utilize the Meyer-Wallach measure as a metric, as suggested by Sim et al. [31]. This approach enables to investigate the relationship between entanglement descriptors and QNN model performance metrics. It provides a theoretical framework for characterising and comparing quantum feature maps and variational circuits in the context of quantum machine learning.

Among the analysed feature maps, FM3 stands out as the most versatile design. This Pauli feature map, generated for the case $k = 2$, employs single-qubit gates ($P_0 = Z$) and two-qubit gates ($P_{0,1} = ZZ$). FM3 exhibits a higher effective dimension compared to most other feature maps, enabling the generation of highly entangled states that are difficult to simulate classically and can achieve quantum advantage.

Regarding variational circuits, the particle-preserving circuits, particularly VC2 and VC9, demonstrate the best model performance. These circuits employ *circular* and *scalar* entanglement patterns, which are better suited for implementation on devices with limited nearest neighbor connectivity.

The most versatile QNN structure, showing the best performance in terms of model capacity and trainability, is achieved by combining FM3 and VC2. For a dataset size of $n = 10^6$, this combination achieves a normalised effective dimension of 0.966.

In conclusion, entanglement is found to play a crucial role in feature map design, allowing for effective encodings that leverage the quantum feature space and may provide quantum advantage as described in the literature [14]. On the other hand, the effectiveness of variational circuits lies in their ability to efficiently transform the quantum state generated by the feature map using parameterised gate operations. Therefore, when designing QNN models, it is important to primarily consider entangling capability in feature map design rather than in variational circuits.

Future work may involve developing a unified metric that combines descriptors to evaluate both entangling capability and model capacity of different feature maps.

Additionally, exploring noisy simulations can provide insights into QNN performance in more realistic settings. This study provides a solid framework for evaluating and approaching these challenges, allowing for the comparison of different circuit designs based on metrics such as model capacity and entangling capability. It can serve as a useful starting point for designing new quantum feature maps and variational circuits and benchmarking existing solutions.

A Appendix

A.1 Feature Maps Circuit Diagrams

Figures A.1 and A.2 show the circuits diagrams for the analysed feature maps. Note that three-qubit gates have been decomposed into single and two-qubit gates.

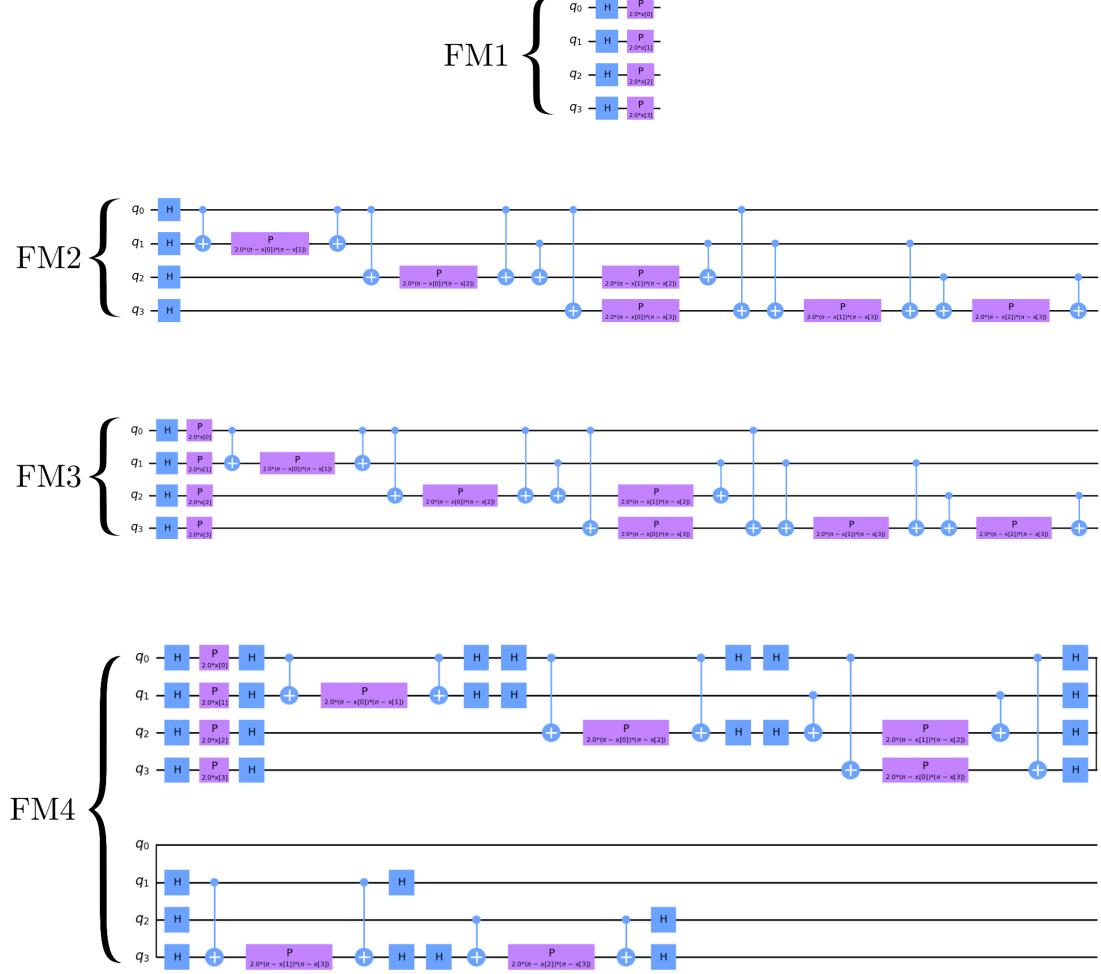


Figure A.1: Circuit diagrams for feature maps FM1, FM2, FM3 and FM4, plotted using the Qiskit python library [26].

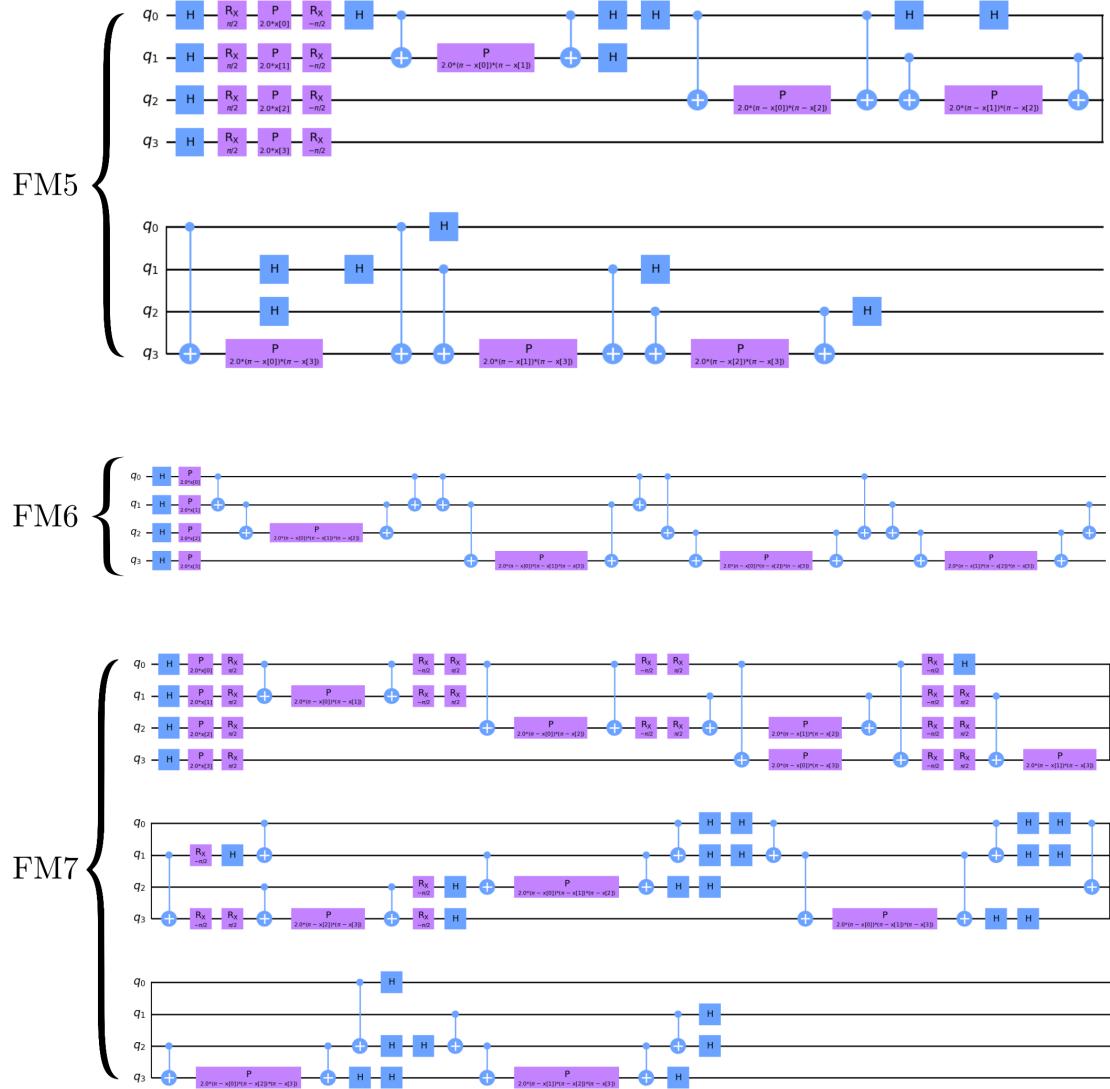


Figure A.2: Circuit diagrams for feature maps FM5, FM6 and FM7, plotted using the Qiskit python library [26].

Bibliography

- [1] Scott Aaronson. Read the fine print. *Nature Physics*, 11(4):291–293, 2015.
- [2] Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, jun 2021.
- [3] Sergio Altares-López, Angela Ribeiro, and Juan José García-Ripoll. Automatic design of quantum feature maps. *Quantum Science and Technology*, 6(4):045015, aug 2021.
- [4] Shun-ichi Amari. Natural Gradient Works Efficiently in Learning. *Neural Computation*, 10(2):251–276, 02 1998.
- [5] Juan Miguel Arrazola, Olivia Di Matteo, Nicolás Quesada, Soran Jahangiri, Alain Delgado, and Nathan Killoran. Universal quantum circuits for quantum chemistry. *Quantum*, 6:742, jun 2022.
- [6] Oksana Bereznuk, Alessio Figalli, Raffaele Ghigliazza, and Kharen Musaelian. A scale-dependent notion of effective dimension, 2020.
- [7] János A. Bergou, Mark Hillery, and Mark Saffman. *Entanglement*, pages 31–63. Springer International Publishing, Cham, 2021.
- [8] Gavin K. Brennen. An observable measure of entanglement for pure states of multi-qubit systems, 2003.
- [9] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, and Dacheng Tao. Expressive power of parametrized quantum circuits. *Phys. Rev. Res.*, 2:033125, Jul 2020.
- [10] Vedran Dunjko and Hans J Briegel. Machine learning amp; artificial intelligence in the quantum domain: a review of recent progress. *Reports on Progress in Physics*, 81(7):074001, jun 2018.
- [11] Lena Funcke, Tobias Hartung, Karl Jansen, Stefan Kähn, and Paolo Stornati. Dimensional expressivity analysis of parametric quantum circuits. *Quantum*, 5:422, mar 2021.
- [12] Bryan T. Gard, Linghua Zhu, George S. Barron, Nicholas J. Mayhall, Sophia E. Economou, and Edwin Barnes. Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. *npj Quantum Information*, 6(1), jan 2020.
- [13] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [14] Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, mar 2019.
- [15] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.

- [16] Ryo Karakida, Shotaro Akaho, and Shun-ichi Amari. Universal statistics of fisher information in deep neural networks: Mean field approach. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1032–1041. PMLR, 16–18 Apr 2019.
- [17] Frederik Kunstner, Lukas Balles, and Philipp Hennig. Limitations of the empirical fisher approximation for natural gradient descent, 2020.
- [18] Zhibin Liao, Tom Drummond, Ian Reid, and Gustavo Carneiro. Approximate fisher information matrix to characterize the training of deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):15–26, 2020.
- [19] Carlos Ortiz Marrero, MÁjria KieferovÁj, and Nathan Wiebe. Entanglement induced barren plateaus, 2021.
- [20] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, 2018.
- [21] David A. Meyer and Nolan R. Wallach. Global entanglement in multiparticle systems. *Journal of Mathematical Physics*, 43(9):4273–4278, sep 2002.
- [22] Yoshifumi Nakata, Christoph Hirche, Ciara Morgan, and Andreas Winter. Unitary 2-designs from random ix/i- and iz/i-diagonal unitaries. *Journal of Mathematical Physics*, 58(5):052203, may 2017.
- [23] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis. Scalable quantum simulation of molecular energies. *Phys. Rev. X*, 6:031007, Jul 2016.
- [24] Manfred Opper. Learning and generalization in a two-layer neural network: The role of the vapnik-chervonvenskis dimension. *Phys. Rev. Lett.*, 72:2113–2116, Mar 1994.
- [25] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. Schuyler Fried, S. Hong, P. Karalekas, C. B. Osborn, A. Papageorge, E. C. Peterson, G. Prawiroatmodjo, N. Rubin, Colm A. Ryan, D. Scarabelli, M. Scheer, E. A. Sete, P. Sivarajah, Robert S. Smith, A. Staley, N. Tezak, W. J. Zeng, A. Hudson, Blake R. Johnson, M. Reagor, M. P. da Silva, and C. Rigetti. Unsupervised machine learning on a hybrid quantum computer, 2017.
- [26] Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023.
- [27] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A*, 99:032331, Mar 2019.

- [28] Maria Schuld, Alex Bocharov, Krysta M. Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Physical Review A*, 101(3), mar 2020.
- [29] Maria Schuld and Francesco Petruccione. *Variational Circuits as Machine Learning Models*, pages 177–215. Springer International Publishing, Cham, 2021.
- [30] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3), mar 2021.
- [31] Sukin Sim, Peter D. Johnson, and Alá n Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies*, 2(12):1900070, oct 2019.
- [32] Alexander Soen and Ke Sun. On the variance of the fisher information for deep learning, 2021.
- [33] Rolando Somma, Gerardo Ortiz, Howard Barnum, Emanuel Knill, and Lorenza Viola. Nature and measure of entanglement in quantum phase transitions. *Phys. Rev. A*, 70:042311, Oct 2004.
- [34] Roberto Stassi, Mauro Cirio, and Franco Nori. Scalable quantum computer with superconducting circuits in the ultrastrong coupling regime. *npj Quantum Information*, 6(1):67, 2020.
- [35] Vladimir Vapnik, Esther Levin, and Yann Le Cun. Measuring the VC-Dimension of a Learning Machine. *Neural Computation*, 6(5):851–876, 09 1994.
- [36] Christopher W. Warren, Jorge Fernández-Pendás, Shahnawaz Ahmed, Tahereh Abad, Andreas Bengtsson, Janka Biznárová, Kamanasish Debnath, Xiu Gu, Christian Križan, Amr Osman, Anita Fadavi Roudsari, Per Delsing, Göran Johansson, Anton Frisk Kockum, Giovanna Tancredi, and Jonas Bylander. Extensive characterization and implementation of a family of three-qubit gates at the coherence limit. *npj Quantum Information*, 9(1):44, 2023.
- [37] Daniel Yoffe, Amir Natan, and Adi Makmal. A qubit-efficient variational selected configuration-interaction method, 2023.
- [38] Nengkun Yu, Runyao Duan, and Mingsheng Ying. Five two-qubit gates are necessary for implementing the toffoli gate. *Phys. Rev. A*, 88:010304, Jul 2013.

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):

First name(s):

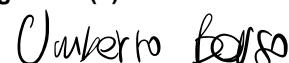
With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Signature(s)



For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.