Addressing the Current Challenges of Quantum Machine Learning through Multi-Chip Ensembles

Junghoon Justin Park
Interdisciplinary Program in Artificial Intelligence
Seoul National University
Gwanakgu, Seoul, 08826
utopie9090@snu.ac.kr

Jiook Cha
Department of Psychology
Department of Brain and Cognitive Sciences
Interdisciplinary Program in Artificial Intelligence
Seoul National University
Gwanakgu, Seoul, 08826
connectome@snu.ac.kr

Samuel Yen-Chi Chen *
Wells Fargo
New York, USA
yen-chi.chen@wellsfargo.com

Huan-Hsin Tseng
Computational Science Initiative
Brookhaven National Laboratory
New York, USA
htseng@bnl.gov

Shinjae Yoo
Computational Science Initiative
Brookhaven National Laboratory
New York, USA
sjyoo,@bnl.gov

Abstract

Quantum Machine Learning (QML) holds significant promise for solving computational challenges across diverse domains. However, its practical deployment is constrained by the limitations of noisy intermediate-scale quantum (NISQ) devices, including noise, limited scalability, and trainability issues in variational quantum circuits (VQCs). We introduce the multi-chip ensemble VQC framework, which partitions high-dimensional computations across smaller quantum chips to enhance scalability, trainability, and noise resilience. We show that this approach mitigates barren plateaus, reduces quantum error bias and variance, and maintains robust generalization through controlled entanglement. Designed to align with current and emerging quantum hardware, the framework demonstrates strong potential for enabling scalable QML on near-term devices, as validated by experiments on standard benchmark datasets (MNIST, FashionMNIST, CIFAR-10) and real world dataset (PhysioNet EEG).

1 Introduction

The past decade has witnessed significant advancements in quantum computing hardware, algorithms, and applications [1–5]. Among these, Quantum Machine Learning (QML) has emerged as a promising

---

*The views expressed in this article are those of the authors and do not represent the views of Wells Fargo. This article is for informational purposes only. Nothing contained in this article should be construed as investment advice. Wells Fargo makes no express or implied warranties and expressly disclaims all legal, tax, and accounting implications related to this article.

approach to harness quantum computing for real-world challenges, leveraging properties such as superposition and entanglement to achieve computational advantages over classical methods [6, 7]. Applications range from quantum simulation in chemistry [8, 9] and materials science [4, 10] to healthcare [11, 12], sensing [13–15], and high-energy physics [16, 17].

Despite these prospects, QML faces critical challenges due to the limitations of noisy intermediate-scale quantum (NISQ) devices [18, 19]. These devices, characterized by noise, limited coherence, and sparse qubit connectivity, pose significant barriers to the scalability and trainability of QML models, particularly variational quantum circuits (VQCs) [20]. The emergence of barren plateaus [21], where gradients vanish in the optimization landscape, further exacerbates these challenges.

To overcome these NISQ-era barriers, we introduce multi-chip ensemble VQCs—a modular architecture that splits the input across several small quantum chips and aggregates their measurements classically. This partitioning scales to high-dimensional data without enlarging any single chip; tames barren plateaus by capping inter-chip entanglement; boosts generalization through implicit regularization; dampens hardware noise because each subcircuit is shorter and errors average out. Instead of dissecting every possible design choice (e.g., circuit depth, quantum gate set, partition heuristics), we demonstrate theoretically and empirically that these advantages stem from the framework's ability to control quantum entanglement. While direct experiments on real quantum hardware was infeasible, we validated the multi-chip ensemble framework under realistic depolarizing and amplitude-damping noise, mirroring today's NISQ conditions. The scheme is fully compatible with existing hardware and fits the multi-chip roadmaps of IBM, IonQ, and Rigetti [22–24].

2 Background

2.1 Status-Quo: Single-Chip VQCs

In VQCs, the input data $\boldsymbol{x}$ is first encoded into a quantum state $\rho(\boldsymbol{x})$. A parameterized unitary operator $U(\boldsymbol{\theta})$ then acts on this state, where $\boldsymbol{\theta}$ represents the tunable parameters of the quantum circuit. The evolution of the quantum state is given as $U(\boldsymbol{\theta})\rho(\boldsymbol{x})U^\dagger(\boldsymbol{\theta})$. Measurements are then performed on the $U(\boldsymbol{\theta})\rho(\boldsymbol{x})U^\dagger(\boldsymbol{\theta})$ to produce classical outputs.

The output is represented by the function:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \mathrm{Tr}[HU(\boldsymbol{\theta})\rho(\boldsymbol{x})U^\dagger(\boldsymbol{\theta})], \tag{1}$$

where $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ represents the output of the quantum model, derived from the expected value of measurements $\mathrm{Tr}[HU(\boldsymbol{\theta})\rho(\boldsymbol{x})U^\dagger(\boldsymbol{\theta})]$ performed by the Hermitian operator H. This expected value reflects the average outcomes based on the probability distribution of measurement results.

During training, the parameters $\boldsymbol{\theta}$ are adjusted to optimize the performance of the quantum circuit $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ by minimizing the loss function $\mathcal{L}(\boldsymbol{x}, y; \boldsymbol{\theta})$ over the training dataset $D = \{(\boldsymbol{x}_i, y_i)\}_i$. For example, the loss function for a regression task may be $\mathcal{L}(\boldsymbol{x}, y; \boldsymbol{\theta}) = \sum_i \|f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i\|^2$. The value of the loss $\mathcal{L}(\boldsymbol{x}, y; \boldsymbol{\theta})$ (or of its gradient) are estimated on a quantum circuit are then fed into a classical optimizer, which attempts to solve the optimization task $\arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{x}, y; \boldsymbol{\theta})$.

VQCs leverage the parameter-shift rule to compute analytical gradients by evaluating circuits at shifted parameter values [25]. This enables seamless integration with classical components through backpropagation, supporting end-to-end training of hybrid quantum-classical models.

2.2 Limitations of VQC

VQCs are hybrid quantum-classical algorithms designed to utilize the computational power of NISQ devices. They parameterize quantum circuits to solve optimization problems, where a classical optimizer iteratively updates parameters to minimize a cost function. By leveraging quantum hardware for state preparation and measurement, and classical resources for optimization, VQCs bridge the gap between NISQ limitations and real-world computational demands. This hybrid approach has enabled applications in quantum chemistry, combinatorial optimization, and quantum machine learning, making VQCs a cornerstone of near-term quantum computing research [26].

Despite their potential, VQCs face significant challenges due to the limitations of current quantum hardware and algorithmic scalability [18]. While theoretically capable of achieving quantum advantage, these limitations hinder their ability to solve high-dimensional and complex problems:

Scalability   Current NISQ devices are restricted to tens or low hundreds of qubits, with limited coherence times and sparse qubit connectivity [18]. While it is possible to process high-dimensional data with limited number of qubits using amplitude encoding, it requires exponential circuit depth [27–29]. As quantum noise grows exponentially with circuit depth [30], large-scale quantum circuits are often infeasible. These constraints prevent VQCs from processing high-dimensional datasets or representing rich quantum states essential for tasks like classification and regression.

Trainability   Barren plateaus—regions in the optimization landscape where gradients vanish—are a significant obstacle to VQC optimization [20, 21, 31]. Gradients often vanish exponentially with the number of qubits, making optimization difficult for many problems [32]. Noise and random parameter initialization exacerbate these challenges, creating uninformative and noisy loss landscapes that hinder effective parameter updates [33]. These factors reduce the generalizability of VQCs on complex datasets and limit their utility in quantum machine learning.

Noise Resilience   Noise, inherent to NISQ hardware, affects all stages of VQC execution, from state preparation to measurements [26]. Noise arises from interactions between qubits and their environment, introducing errors in quantum gates, measurements, and state preparation. This limits circuit depth and computational accuracy [19]. Accumulated noise reduces circuit fidelity exponentially with depth, limiting the expressivity of VQCs and exacerbating barren plateau issues [33]. These combined effects restrict VQCs to shallow circuits, narrowing the range of problems they can address.

In summary, while VQCs represent a critical step toward quantum advantage in the NISQ era, their practical utility is constrained by challenges in scalability, trainability, and noise resilience. Addressing these limitations is crucial to translating their theoretical promise into impactful applications in artificial intelligence and beyond.

2.3   Comparison with Related Works

2.3.1   Distributed QML Approaches

The inherent limitations of NISQ devices have spurred approaches to distribute quantum computations across multiple smaller processing units. While distributed quantum computing has been studied extensively [34–37] and applied to variational algorithms like variational quantum eigensolvers [38, 39] and quantum approximate optimization algorithms [40], its application to QML remains nascent [41, 42]. Frameworks such as QUDIO [43] demonstrate that distributed approaches can accelerate convergence and reduce circuit depth on current hardware. However, most distributed QML research focuses on empirical scaling or data partitioning without addressing the theoretical foundations of critical QML challenges [41–46].

2.3.2   Limitations of existing Distributed QML strategies

Current distributed QML approaches face four principal limitations. First, circuit cutting techniques [44, 46] fragment large circuits into smaller segments but incur exponential sampling overhead, undermining their practical scalability. Second, communication-based methods [47] that integrate mid-circuit classical information exchange suffer from latency and coherence disruptions. Third, feature-based partitioning approaches [45] divide input data across multiple circuits with classical output aggregation, but lack theoretical foundations regarding trainability, generalization, and noise resilience. Fourth, quantum federated learning methods [48], while promising for privacy preservation, remain primarily heuristic without rigorous analysis of fundamental QML challenges.

These limitations coalesce around three critical gaps: (1) insufficient theoretical analysis of barren plateaus, quantum bias-variance trade-offs, and noise resilience; (2) empirical evaluations restricted to small-scale, noiseless simulations with limited insight into real hardware performance; and (3) minimal consideration of compatibility with emerging modular quantum hardware architectures being developed by industry. These gaps highlight the need for theoretically grounded, hardware-compatible distributed QML frameworks that address fundamental quantum learning challenges.

2.3.3 Novel contribution of the present work

We propose a novel multi-chip ensemble VQC framework that addresses the limitations above both theoretically and empirically. Our contributions are as follows:

Theoretical Rigor   We provide formal analysis linking entanglement to trainability and generalization. Specifically, we prove that restricting inter-chip entanglement increases gradient variance (mitigating barren plateaus) and improves generalization by regulating model complexity via the quantum bias–variance trade-off (Appendix B, C, D, E).

Noise Resilience without Overhead   Unlike traditional error mitigation techniques that reduce error bias at the cost of increased variance [49], our method reduces both simultaneously through architectural design. This is analytically proven (Appendix F) and empirically validated under depolarizing and amplitude damping noise.

Scalability and Hardware Compatibility   : The proposed framework processes high-dimensional data using ensembles of shallow circuits, enabling scalability without increasing qubit count per chip. Our architecture aligns with modular quantum hardware roadmaps (e.g., IBM, IonQ, Rigetti), making it forward-compatible with near-term devices [22–24].

Unified Resolution of QML Challenges   To the best of our knowledge, this is the first distributed QML approach to simultaneously and systematically address scalability, trainability (barren plateaus), generalizability, and noise resilience, supported by both theoretical guarantees and large-scale empirical results.

In summary, our work fills a critical gap in the field by advancing distributed QML from engineering patchwork to a principled and scalable learning framework. It is both practically implementable on near-term hardware and theoretically grounded to tackle foundational QML limitations.

3   Multi-Chip Ensemble VQCs

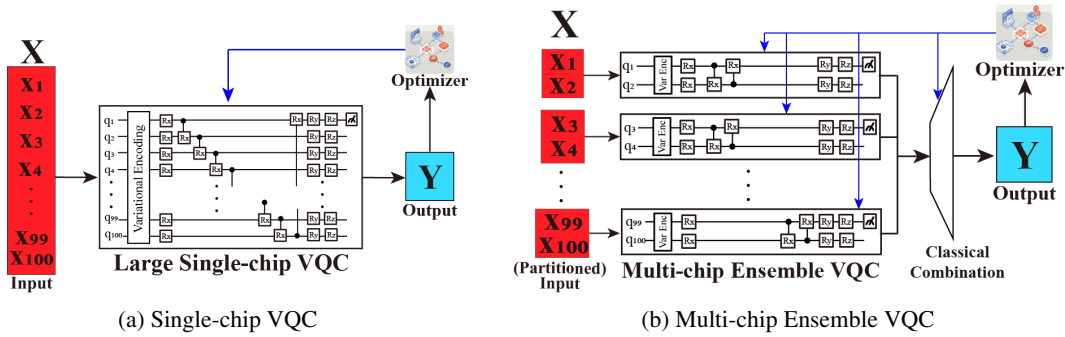(a) Single-chip VQC   (b) Multi-chip Ensemble VQC

Figure 1: Comparison between Single-chip vs Multi-chip Ensemble VQCs

3.1   Multi-Chip Ensemble Framework Overview

Multi-chip ensemble VQC introduces a novel architecture that combines k disjoint quantum chips with each chip comprised of small l-qubits quantum subcircuits to create a n-qubit large quantum circuit ($n = k \times l$) (Figure 1b). Here, each quantum chip has individual subcircuit $U_i(\boldsymbol{\theta}_i)$, acting on l-qubits. Crucially, there are no gates connecting different chips–so the total action is:

$$U_{MC}(\boldsymbol{\theta}) = \bigotimes_{i=1}^{k} U_i(\boldsymbol{\theta}_i). \tag{2}$$

This indicates that there are no cross-chip entanglement in the multi-chip ensemble VQCs.

To perform computations on multiple quantum chips, the input data space $\mathbb{R}^n$ is partitioned into k smaller subspaces, where $\mathbb{R}^n = \mathbb{R}^\ell \times \cdots \times \mathbb{R}^\ell$ (k times). Each input $\boldsymbol{x} \in \mathbb{R}^n$ is split into k

concatenated subvectors $\boldsymbol{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k]$, where $\boldsymbol{x}_i \in \mathbb{R}^\ell$ corresponds to the input for the i-th quantum chip.

Each subvector $\boldsymbol{x}_i \in \mathbb{R}^\ell$ is then processed by an independent quantum circuit $U_i(\boldsymbol{\theta}_i)$ on a separate quantum chip, where each circuit operates on ℓ-qubits. The quantum state preparation for each subcircuit follows $\rho_i(\boldsymbol{x}_i) = V_i(\boldsymbol{x}_i) \left(^{\otimes \ell}|0\rangle\langle 0|^{\otimes \ell}\right) V_i^\dagger(\boldsymbol{x}_i)$, where $V_i(\boldsymbol{x}_i)$ represents the data encoding unitary for the i-th subcircuit.

The measurement outputs from the k subcircuits are classically combined to produce the final output. Specifically, each subcircuit measures an observable H_i to produce an expectation value:

$$f_{\boldsymbol{\theta}_i}(\boldsymbol{x}_i) = \text{Tr}[H_i U_i(\boldsymbol{\theta}_i) \rho_i(\boldsymbol{x}_i) U_i^\dagger(\theta_i)]. \tag{3}$$

Here, H_i can vary across subcircuits or be identical, depending on the application. For example, in image classifications, H_i could correspond to observables encoding class probabilities. The circuit outputs $f_{\boldsymbol{\theta}_1}(\boldsymbol{x}_1), ..., f_{\boldsymbol{\theta}_k}(\boldsymbol{x}_k)$ are then combined through a classical function $g : \mathbb{R}^k \to \mathbb{R}^m$ to produce a final output:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = g(f_{\boldsymbol{\theta}_1}(\boldsymbol{x}_1), ..., f_{\boldsymbol{\theta}_k}(\boldsymbol{x}_k)). \tag{4}$$

The choice of combination function g depends on the specific learning task. For instance, it could be a weighted sum for regression tasks or a more complex nonlinear function implemented via a shallow neural network for classification tasks.

The training procedure for multi-chip ensemble VQC maintains the hybrid quantum-classical nature of traditional VQCs while incorporating the distributed architecture. The parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k\}$ are optimized jointly to minimize the overall loss function:

$$\mathcal{L}(\boldsymbol{x}, y; \boldsymbol{\theta}) = \mathcal{L}_{ensemble}(f_{\boldsymbol{\theta}}(\boldsymbol{x}), y), \tag{5}$$

where $\mathcal{L}$ denote a task-dependent loss function (e.g., mean squared error for regression or cross-entropy for classification), $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ the multi-chip ensemble VQC's prediction, and y the target outcome. Notably, the gradients for each subcircuit can be computed independently and in parallel, enabling efficient training even as the number of subcircuits k increases. This parallelization of both inference and training represents a significant advantage over single-chip approaches, particularly for high-dimensional data processing.

3.2 Compatibility with Current and Near-Future Quantum Hardware

The multi-chip ensemble VQC framework aligns naturally with both current NISQ devices and emerging modular quantum architectures [5, 18, 19, 50]. By distributing computations across multiple smaller chips without requiring inter-chip quantum communication, our approach works within existing hardware constraints while anticipating future developments.

Current NISQ hardware faces fundamental limitations in qubit count, coherence time, and connectivity [18, 51]. Our approach addresses these challenges by confining operations to smaller, high-coherence regions and using classical aggregation instead of noisy inter-chip quantum gates. This design enables processing of high-dimensional data even with modest qubit counts per chip.

This architecture maps directly to emerging industry roadmaps, including IBM's quantum interconnect strategy [22], Rigetti's modular superconducting systems [24], and IonQ's reconfigurable multicore architecture [23]. For detailed hardware compatibility analysis, see Appendix A.

3.3 Entanglement in Multi-Chip Ensembles

Quantum entanglement—the non-classical correlation between quantum subsystems—fundamentally shapes the computational capabilities and limitations of QML models. When unitary operators $U(\boldsymbol{\theta})$ act on encoded quantum states $\rho(\boldsymbol{x})$, the resulting state $U(\boldsymbol{\theta})\rho(\boldsymbol{x})U^\dagger(\boldsymbol{\theta})$ can exhibit entanglement patterns that fundamentally determine the model's expressibility, trainability, and generalizability.

A key feature of our multi-chip ensemble architecture is its controlled entanglement structure: quantum connections exist only among qubits within each chip, with no entanglement between qubits across different chips. This design choice leads to quantifiably lower global entanglement compared to equivalent single-chip circuits (see Appendix B for formal proofs).

5

While this reduction in entanglement restricts the accessible portion of the Hilbert space [52–54], it simultaneously addresses two critical challenges in quantum machine learning. First, it mitigates barren plateaus—regions where gradients vanish exponentially with system size [55–57]—by preventing the global entanglement patterns that trigger this phenomenon. Second, it reduces overfitting by limiting the model's capacity to represent excessively complex functions [58], effectively serving as an implicit regularization mechanism.

Our approach demonstrates a fundamental principle: optimal quantum model performance requires calibrating entanglement appropriately rather than maximizing it. The multi-chip ensemble architecture provides a systematic framework for achieving this balance, as confirmed by our experimental results across diverse datasets.

4 Advantages of Multi-Chip Ensembles

Here, we present the advantages of multi-chip ensemble VQCs over the conventional single-chip VQCs. We show that given a fixed number of total qubits, applying multi-chip ensemble approach to single-chip VQCs can improve scalability, trainability, generalizability, and noise resilience of the model.

4.1 Improved Scalability

The scalability of multi-chip ensemble VQCs offers significant advantages over conventional single-chip VQCs, particularly for processing high-dimensional data in machine learning applications. Single-chip VQCs face scalability limitations, as variational encoding requires n qubits to process n-dimensional data. This often necessitates classical dimension reduction techniques, such as principal component analysis, autoencoders, or learnable neural networks, which can introduce information loss and degrade model performance (Figure 3a). While amplitude encoding theoretically enables encoding 2^n-dimensional data with n qubits by leveraging quantum superposition, its practical use is limited by deep circuit requirements and complex state preparation, making it infeasible for current NISQ hardware [27–29].

In contrast, multi-chip ensemble VQCs distribute the computational load across k independent quantum chips, each processing $l = n/k$ dimensions of the input data (Figure 3c). This architecture eliminates the need for mandatory dimension reduction by scaling horizontally through additional chips rather than requiring larger single chips. For instance, a small quantum circuit with 10 physical qubits can be combined 200 times to conceptually create an ensemble model with $2000 = 10 \times 200$ qubits, enabling 2000-dimensional data processing using only 10 physical qubits. Additionally, each subcircuit can specialize in specific feature subspaces, capturing nuanced patterns that might be lost with global dimension reduction.

The multi-chip design also enables advanced parallel data processing strategies. Correlated features can be grouped and processed on the same chip, while independent features are distributed across different chips. This natural partitioning of the feature space aligns well with many real-world datasets, where features often exhibit clustered correlations.

4.2 Improved Trainability

A fundamental challenge in QML is the barren plateau phenomenon—regions in the optimization landscape where gradients vanish exponentially with system size [21, 31]. These plateaus emerge when circuits achieve volume-law entanglement, causing quantum states to become so entangled that parameter perturbations produce minimal output changes [55, 57].

4.2.1 Gradient Variance Enhancement

Our multi-chip ensemble architecture directly addresses this challenge by constraining the dimension of entangled subspaces. By limiting entanglement to within-chip boundaries, we prevent the global entanglement patterns that trigger barren plateaus. Our theoretical analysis in Appendix C demonstrates that for a fixed total qubit count n, partitioning into k chips of size $l = n/k$ significantly increases gradient variance compared to a fully-entangled single-chip implementation. Our experimental results in Section 5.3 confirm this relationship, showing increased gradient variance with higher chip counts.

4.2.2 Escaping the Classical Simulability Dilemma

Recent approaches to avoid barren plateaus [59–65] often restrict circuits to polynomial subspaces that, while trainable, become classically simulable and thus reduce quantum computational advantage [31, 66]. This creates a fundamental dilemma: circuits that avoid barren plateaus are often classically simulable, while those that maintain quantum advantage suffer from vanishing gradients.

Our multi-chip ensemble architecture provides a pathway to resolve this dilemma through a careful balance of local complexity and global structure. By scaling chip size l with system size n, we create a framework where each l-qubit subcircuit maintains sufficient complexity to resist efficient classical simulation. Simultaneously, the overall n-qubit system avoids the global 2-design randomization conditions that trigger barren plateaus due to the absence of cross-chip entangling gates. This dual property enables trainable quantum models that maintain potential quantum advantage by operating outside known classically simulable polynomial subspaces while simultaneously mitigating barren plateaus. Formal proofs of these properties are provided in Appendix D.

4.3 Improved Generalizability

Generalization—a model's ability to perform well on unseen data—represents a critical challenge in machine learning that manifests uniquely in quantum systems. Our multi-chip ensemble approach provides a principled framework for optimizing generalization through controlled entanglement.

Quantum Bias-Variance Trade-off In classical machine learning, the bias-variance trade-off balances model complexity against overfitting risk [67]. This fundamental principle extends to quantum systems, where increased circuit complexity reduces bias (underfitting) but potentially increases variance (overfitting) [68, 69]. Quantum models with appropriately calibrated model complexity can achieve exceptional generalization, even with limited training samples [70, 71], but excessive complexity leads to deteriorating test performance. This relationship creates a critical need for methods that can navigate this trade-off effectively in quantum circuits.

Entanglement as a Complexity Regulator Our theoretical analysis, detailed in Appendix E, establishes that quantum entanglement directly modulates this bias-variance trade-off. We demonstrate that higher entanglement levels (γ_k) expand the representable function class, reducing bias in the learning process. However, this expanded expressibility simultaneously increases the risk of overfitting, captured by a complexity penalty $\Omega(\gamma_k)$ in our generalization bounds. The optimal generalization performance occurs at an intermediate entanglement level that balances these competing factors. This relationship is formalized as:

$$\text{gen}(\theta) \leq \min_{k, \theta \in \Theta_k} \left\{ \underbrace{R(\theta) - R_S(\theta)}_{\text{bias}} + \underbrace{\epsilon_k + \Omega(\gamma_k)}_{\text{variance}} \right\} \tag{6}$$

This formulation demonstrates that entanglement serves as a fundamental regulator of the quantum bias-variance trade-off, with important implications for quantum model design.

Multi-Chip Ensemble as Implicit Regularization Our multi-chip ensemble architecture implements this theoretical insight by confining entanglement within chip boundaries. By partitioning n qubits into k independent chips, we reduce global entanglement levels proportionally to the number of chips ($\gamma_k \propto 1/k$). This design creates an implicit regularization mechanism: each chip maintains sufficient internal expressibility to capture relevant data patterns, while the absence of cross-chip entanglement prevents the variance surge associated with global entanglement. The architecture naturally positions the model near the optimal point on the bias-variance curve without requiring additional techniques like entanglement dropout [58].

4.4 Improved Noise Resilience

Noise in quantum hardware induces bias and errors that severely impact VQC performance [49]. Traditional approaches face significant limitations: quantum error correction [72–74] requires substantial qubit overhead (approximately 1,000 physical qubits per logical qubit [75]), while quantum error mitigation [49, 76–78] techniques like zero-noise extrapolation (ZNE) [79, 80] introduce a fundamental bias-variance trade-off, reducing error bias at the cost of increased variance [49, 77].

Our multi-chip ensemble framework inherently reduces both bias and variance of quantum errors simultaneously without additional resources. For a single-chip circuit with n-qubits and N_g noisy gates, depolarizing-channel analysis shows that bias grows exponentially as $\exp(nN_g\varepsilon)$, while variance scales inversely with circuit runs as $1/N_{\text{cir}}$. When partitioning this workload across k independent $l(=n/k)$-qubit chips, each chip contributes a significantly smaller bias term $\exp\left(\frac{n}{k}N_g\varepsilon\right)$. Summing across chips yields a total bias of $k\exp\left(\frac{n}{k}N_g\varepsilon\right)$, which remains exponentially smaller than single-chip bias for any $k > 1$. Furthermore, because chip noise patterns are uncorrelated, their classical averaging reduces variance to $1/(kN_{\text{cir}})$.

This dual-benefit noise reduction contrasts sharply with traditional error mitigation techniques that typically improve one error component at the expense of the other. Our approach requires no additional quantum resources beyond running the already partitioned subcircuits, providing a hardware-compatible route to robust QML on noisy devices. Mathematical proofs in Appendix F formalize these advantages, which our experimental results confirm across various datasets and noise conditions.

5 Experiments

(a) Model Performance
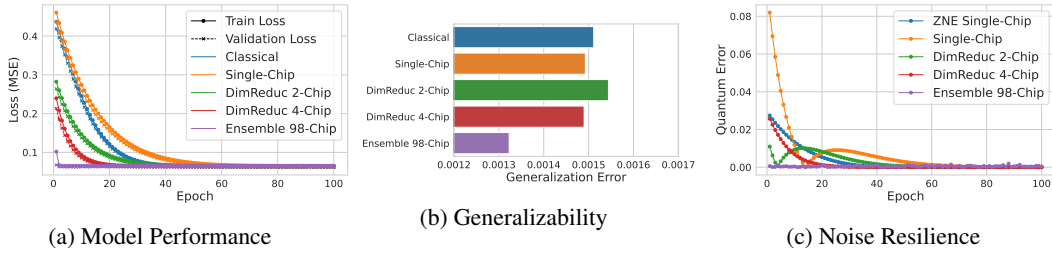(b) Generalizability
(c) Noise Resilience

Figure 2: Experimental Results on MNIST. DimReduc 2-Chip and 4-Chip indicate multi-chip ensemble VQC models with classical dimension reduction. Ensemble 98-Chip denotes multi-chip ensemble VQC model without classical dimension reduction.

We designed our experiments to address two fundamental questions: (1) Do multi-chip ensemble VQCs demonstrate enhanced performance, improved generalizability, and greater noise resilience compared to single-chip VQCs? (2) Can multi-chip ensembles effectively process high-dimensional data without classical dimension reduction? We also conducted additional experiments to validate our approach on datasets beyond standard benchmarks.

5.1 Experimental Design

We implemented three model configurations for comparative analysis: (a) single-chip VQC with classical dimension reduction (Figure 3a), (b) multi-chip ensemble VQC with classical dimension reduction (Figure 3b), and (c) multi-chip ensemble VQC without classical dimension reduction (Figure 3c). All models were built within a quantum-classical hybrid autoencoder framework, ensuring that differences in performance stem solely from their quantum structures rather than architectural variations. Comparing models (a) and (b) addresses question (1), while comparing models (a) and (c) examines question (2).

For standard benchmark evaluation, we used MNIST [81], FashionMNIST [82], and CIFAR-10 [83] datasets. After flattening, input dimensions were 784 (MNIST, FashionMNIST) and 3072 (CIFAR-10). Each model used 8 qubits for MNIST and FashionMNIST, and 12 qubits for CIFAR-10. The multi-chip ensemble without dimension reduction (model c) distributed inputs across $98(= 784/8)$ chips for MNIST/FashionMNIST and $256(= 3072/12)$ chips for CIFAR-10. We also implemented a classical autoencoder as a baseline.

To demonstrate broader applicability, we applied our multi-chip ensemble approach to quantum convolutional neural networks (QCNN) [4] trained on PhysioNet EEG time-series data [84, 85], which has 3264-dimensional spatio-temporal features. The multi-chip ensemble QCNN processed this data using 272 chips with 12 qubits each, without requiring classical dimension reduction.

8

Because run-time on today's cloud quantum processors is scarce and queue times prohibit the thousands of circuit executions required for gradient-based training, we emulate NISQ conditions with calibrated depolarizing and amplitude-damping noise; this yields the same error profiles reported in current hardware data sheets while allowing controlled, repeatable comparisons. Due to space constraints, we present MNIST results in the main text, with detailed experimental design and additional results available in Appendices G and H.

5.2 Performance & Scalability

As shown in Figure 2a, the multi-chip ensemble VQC without classical dimension reduction (Ensemble 98-Chip) achieved the best performance, converging to the optimal loss value in fewer than 10 epochs. This is significantly faster than both the single-chip VQC and the classical baseline, demonstrating the multi-chip ensemble VQC's ability to effectively learn high-dimensional data without classical dimension reduction.

Additionally, multi-chip ensemble VQC models with classical dimension reduction (DimReduc 2-Chip, 4-Chip) outperformed the single-chip VQC. Increasing the number of chips from 2 to 4 further improved performance, indicating that multi-chip ensemble VQCs outperform single-chip VQCs under comparable conditions. Similar trends were observed in the FashionMNIST and CIFAR-10 datasets, as detailed in Figures 4a and 5a (Appendix H).

Results on the PhysioNet EEG task confirm the trend: the 272-chip QCNN outperforms both a single-chip QCNN and a matched classical CNN, attaining higher balanced accuracy while showing a much smaller train–validation gap, indicating reduced overfitting.

5.3 Trainability

Using the entangling capability measure [52], we quantified the quantum entanglement levels in single-chip and multi-chip ensemble VQCs with classical dimension reduction. For a total of 8 qubits, increasing the number of chips (i.e., dividing a large VQC into smaller subcircuits) reduced the overall entanglement while increasing the variance of gradients (Table **??**).

Since barren plateaus are characterized by exponentially small gradient variance, the higher gradient variance observed in multi-chip ensemble VQCs suggests a reduced risk of barren plateaus. These results, consistent with our theoretical analysis in Appendix C, demonstrate that the multi-chip ensemble approach can mitigate the risk of barren plateaus, thereby enhancing trainability.

5.4 Generalizability

Generalization error, defined as the difference between test loss and final training loss [69, 70], was used to evaluate model generalizability. Lower generalization error indicates better generalizability. As depicted in Figure 2b and further detailed in Appendix H (Figures 4b and 5b), multi-chip ensemble VQC models with classical dimension reduction consistently exhibited smaller generalization error compared to single-chip VQC models. These findings demonstrate that the multi-chip ensemble approach enhances the generalizability of VQC models.

5.5 Noise Resilience

We evaluated noise resilience by measuring quantum error—the absolute difference between validation losses of noisy and noiseless circuits. Lower quantum error indicates higher resilience. Our noise model incorporated both depolarizing noise (uniform random errors across qubits) and amplitude damping noise (energy dissipation effects), reflecting common noise profiles in current quantum hardware [86]. We compared our approach against ZNE, a widely used error mitigation technique for near-term quantum devices [79, 80].

As shown in Figure 2c, multi-chip ensemble VQCs consistently achieved lower quantum errors than both unmitigated single-chip VQCs and those with ZNE error mitigation. This advantage persisted across all datasets (see Appendix H, Figures 4c and 5c). Notably, while ZNE reduces error bias at the cost of increased variance—requiring more circuit runs to stabilize results—our multi-chip ensemble approach simultaneously reduces both error components without this trade-off. This enables robust

performance from the first iteration, demonstrating superior noise resilience without the limitations or overhead of traditional error mitigation techniques.

6 Discussion

The multi-chip ensemble VQC framework effectively addresses key limitations of conventional single-chip quantum machine learning models. By distributing computations across smaller, independent quantum chips and leveraging classical postprocessing, this approach overcomes scalability constraints, reduces noise accumulation, and enhances both generalizability and noise resilience. Experimental results demonstrate that multi-chip ensembles outperform single-chip VQCs in processing high-dimensional data, even without classical dimension reduction, and achieve superior performance, lower generalization loss, and reduced quantum errors.

Multi-chip ensemble VQCs' ability to directly process high-dimensional data without dimension reduction preserves data fidelity. Machine learning tasks in areas such as computer vision, natural language processing, and genomics frequently involve high-dimensional data, where complex feature interactions are critical for model performance. By maintaining these relationships, multi-chip ensemble VQCs can capture more sophisticated patterns compared to single-chip VQCs, which often sacrifice information to meet hardware limitations.

Future research on multi-chip ensemble VQCs could explore the design of heterogeneous subcircuits for individual chips. Tailoring subcircuit architectures or assigning specific feature subspaces to different chips may further optimize performance and enhance the framework's adaptability to diverse tasks. Additionally, investigating advanced strategies for inter-chip communication and integration with emerging multi-chip quantum hardware will be critical for scaling the approach to even larger datasets and more complex problems.

The multi-chip ensemble VQC represents a forward-compatible and practical solution for advancing quantum machine learning, paving the way for robust, scalable applications in the NISQ era and beyond.

References

[1] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, "Validating quantum computers using randomized model circuits," *Phys. Rev. A*, vol. 100, p. 032328, Sep 2019.

[2] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.

[3] R. Acharya, D. A. Abanin, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, *et al.*, "Quantum error correction below the surface code threshold," *Nature*, 2024.

[4] I. Cong, S. Choi, and M. D. Lukin, "Quantum convolutional neural networks," *Nature Physics*, vol. 15, no. 12, pp. 1273–1278, 2019.

[5] Y. Gujju, A. Matsuo, and R. Raymond, "Quantum machine learning on near-term quantum devices: Current state of supervised and unsupervised techniques for real-world applications," *Phys. Rev. Appl.*, vol. 21, p. 067001, Jun 2024.

[6] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

[7] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, "Training deep quantum neural networks," *Nature Communications*, vol. 11, no. 1, p. 808, 2020.

[8] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor," *Nature Communications*, vol. 5, no. 1, p. 4213, 2014.

[9] C. Cîrstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, "Variational fast forwarding for quantum simulation beyond the coherence time," *npj Quantum Information*, vol. 6, no. 1, p. 82, 2020.

[10] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering," *Science*, vol. 361, no. 6400, pp. 360–365, 2018.

[11] S. Mensa, E. Sahin, F. Tacchino, P. Kl Barkoutsos, and I. Tavernelli, "Quantum machine learning framework for virtual screening in drug discovery: a prospective quantum advantage," *Machine Learning: Science and Technology*, vol. 4, no. 1, p. 015023, 2023.

[12] U. Ullah, A. G. O. Jurado, I. D. Gonzalez, and B. Garcia-Zapirain, "A fully connected quantum convolutional neural network for classifying ischemic cardiopathy," *IEEE Access*, vol. 10, pp. 134592–134605, 2022.

[13] J. L. Beckey, M. Cerezo, A. Sone, and P. J. Coles, "Variational quantum algorithm for estimating the quantum fisher information," *Phys. Rev. Res.*, vol. 4, p. 013083, Feb 2022.

[14] J. J. Meyer, J. Borregaard, and J. Eisert, "A variational toolbox for quantum multi-parameter estimation," *npj Quantum Information*, vol. 7, no. 1, p. 89, 2021.

[15] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O'Brien, J. G. Rarity, A. Laing, and M. G. Thompson, "Experimental quantum hamiltonian learning," *Nature Physics*, vol. 13, no. 6, pp. 551–555, 2017.

[16] S. Y.-C. Chen, T.-C. Wei, C. Zhang, H. Yu, and S. Yoo, "Quantum convolutional neural networks for high energy physics data analysis," *Phys. Rev. Res.*, vol. 4, p. 013231, Mar 2022.

[17] A. Di Meglio, K. Jansen, I. Tavernelli, C. Alexandrou, S. Arunachalam, C. W. Bauer, *et al.*, "Quantum computing for high-energy physics: State of the art and challenges," *PRX Quantum*, vol. 5, no. 3, p. 037001, 2024.

[18] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, *et al.*, "Noisy intermediate-scale quantum algorithms," *Rev. Mod. Phys.*, vol. 94, p. 015004, Feb 2022.

[19] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018.

[20] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, "Challenges and opportunities in quantum machine learning," *Nature Computational Science*, vol. 2, no. 9, pp. 567–576, 2022.

[21] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature Communications*, vol. 9, no. 1, p. 4812, 2018.

[22] IBM Quantum, "IBM Quantum Development & Innovation Roadmap," report, October 2024.

[23] J. Kim, "Reconfigurable multicore quantum architecture." `https://ionq.com/resources/reconfigurable-multicore-quantum-architecture`, January 08 2025. Date Accessed: 2025-01-22.

[24] M. Field, A. Q. Chen, B. Scharmann, E. A. Sete, F. Oruc, K. Vu, V. Kosenko, J. Y. Mutus, S. Poletto, and A. Bestwick, "Modular superconducting-qubit architecture with a multichip tunable coupler," *Phys. Rev. Appl.*, vol. 21, p. 054063, May 2024.

[25] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, "Evaluating analytic gradients on quantum hardware," *Phys. Rev. A*, vol. 99, p. 032331, Mar 2019.

[26] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, "Variational quantum algorithms," *Nature Reviews Physics*, vol. 3, no. 9, pp. 625–644, 2021.

[27] M. Plesch and i. c. v. Brukner, "Quantum-state preparation with universal gate decompositions," *Phys. Rev. A*, vol. 83, p. 032302, Mar 2011.

[28] X. Sun, G. Tian, S. Yang, P. Yuan, and S. Zhang, "Asymptotically optimal circuit depth for quantum state preparation and general unitary synthesis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 10, pp. 3301–3314, 2023.

[29] X.-M. Zhang, M.-H. Yung, and X. Yuan, "Low-depth quantum state preparation," *Phys. Rev. Res.*, vol. 3, p. 043200, Dec 2021.

[30] G. De Palma, M. Marvian, C. Rouzé, and D. S. França, "Limitations of variational quantum algorithms: A quantum optimal transport approach," *PRX Quantum*, vol. 4, p. 010309, Jan 2023.

[31] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, "Barren plateaus in variational quantum computing," *Nature Reviews Physics*, vol. 7, no. 4, pp. 174–189, 2025.

[32] L. Bittel and M. Kliesch, "Training variational quantum algorithms is NP-hard," *Phys. Rev. Lett.*, vol. 127, p. 120502, Sep 2021.

[33] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, "Noise-induced barren plateaus in variational quantum algorithms," *Nature Communications*, vol. 12, no. 1, p. 6961, 2021.

[34] H. J. Kimble, "The quantum internet," *Nature*, vol. 453, no. 7198, pp. 1023–1030, 2008.

[35] C. Monroe, R. Raussendorf, A. Ruthven, K. R. Brown, P. Maunz, L.-M. Duan, and J. Kim, "Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects," *Phys. Rev. A*, vol. 89, p. 022317, Feb 2014.

[36] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, "Simulating large quantum circuits on a small quantum computer," *Phys. Rev. Lett.*, vol. 125, p. 150504, Oct 2020.

[37] D. Barral, F. J. Cardama, G. Díaz-Camacho, D. Faílde, I. F. Llovo, M. Mussa-Juane, J. Vázquez-Pérez, J. Villasuso, C. Piñeiro, N. Costas, J. C. Pichel, T. F. Pena, and A. Gómez, "Review of distributed quantum computing: From single qpu to high performance quantum computing," *Computer Science Review*, vol. 57, p. 100747, 2025.

[38] I. Khait, E. Tham, D. Segal, and A. Brodutch, "Variational quantum eigensolvers in the era of distributed quantum computers," *Phys. Rev. A*, vol. 108, p. L050401, Nov 2023.

[39] Y. Zhang, L. Cincio, C. F. A. Negre, P. Czarnik, P. J. Coles, P. M. Anisimov, S. M. Mniszewski, S. Tretiak, and P. A. Dub, "Variational quantum eigensolver with reduced circuit complexity," *npj Quantum Information*, vol. 8, no. 1, p. 96, 2022.

[40] K.-C. Chen, X. Xu, F. Burt, C.-Y. Liu, S. Yu, and K. K. Leung, "Noise-aware distributed quantum approximate optimization algorithm on near-term quantum hardware," in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 02, pp. 144–149, 2024.

[41] J. Wu, T. Hu, and Q. Li, "Distributed quantum machine learning: Federated and model-parallel approaches," *IEEE Internet Computing*, vol. 28, no. 2, pp. 65–72, 2024.

[42] L. Pira and C. Ferrie, "An invitation to distributed quantum neural networks," *Quantum Machine Intelligence*, vol. 5, no. 2, p. 23, 2023.

[43] Y. Du, Y. Qian, X. Wu, and D. Tao, "A distributed learning scheme for variational quantum algorithms," *IEEE Transactions on Quantum Engineering*, vol. 3, pp. 1–16, 2022.

[44] S. C. Marshall, C. Gyurik, and V. Dunjko, "High Dimensional Quantum Machine Learning With Small Quantum Computers," *Quantum*, vol. 7, p. 1078, Aug. 2023.

[45] Y. Kawase, "Distributed quantum neural networks via partitioned features encoding," *Quantum Machine Intelligence*, vol. 6, no. 1, p. 15, 2024.

[46] A. Marchisio, E. Sychiuco, M. Kashif, and M. Shafique, "Cutting is all you need: Execution of large-scale quantum neural networks on limited-qubit devices," 2024.

[47] K. Hwang, H.-T. Lim, Y.-S. Kim, D. K. Park, and Y. Kim, "Distributed quantum machine learning via classical communication," *Quantum Science and Technology*, vol. 10, p. 015059, dec 2024.

[48] S. Y.-C. Chen and S. Yoo, "Federated quantum machine learning," *Entropy*, vol. 23, no. 4, 2021.

[49] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, "Quantum error mitigation," *Rev. Mod. Phys.*, vol. 95, p. 045005, Dec 2023.

[50] X. Wang, Y. Du, Y. Luo, and D. Tao, "Towards understanding the power of quantum kernels in the NISQ era," *Quantum*, vol. 5, p. 531, Aug. 2021.

[51] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, "The complexity of NISQ," *Nature Communications*, vol. 14, no. 1, p. 6001, 2023.

[52] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900070, 2019.

[53] M. Ballarin, S. Mangini, S. Montangero, C. Macchiavello, and R. Mengoni, "Entanglement entropy production in Quantum Neural Networks," *Quantum*, vol. 7, p. 1023, May 2023.

[54] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, "The power of quantum neural networks," *Nature Computational Science*, vol. 1, no. 6, pp. 403–409, 2021.

[55] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, "Entanglement-induced barren plateaus," *PRX Quantum*, vol. 2, p. 040316, Oct 2021.

[56] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, "Connecting ansatz expressibility to gradient magnitudes and barren plateaus," *PRX Quantum*, vol. 3, p. 010313, Jan 2022.

[57] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin, "Entanglement devised barren plateau mitigation," *Physical Review Research*, vol. 3, no. 3, p. 033090, 2021.

[58] M. Kobayashi, K. Nakaji, and N. Yamamoto, "Overfitting in quantum machine learning and entangling dropout," *Quantum Machine Intelligence*, vol. 4, no. 2, p. 30, 2022.

[59] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, "Absence of barren plateaus in quantum convolutional neural networks," *Phys. Rev. X*, vol. 11, p. 041011, Oct 2021.

[60] K. Zhang, L. Liu, M.-H. Hsieh, and D. Tao, "Escaping from the barren plateau via gaussian initializations in deep variational quantum circuits," in *Advances in Neural Information Processing Systems*, vol. 35, pp. 18612–18627, 2022.

[61] C.-Y. Park and N. Killoran, "Hamiltonian variational ansatz without barren plateaus," *Quantum*, vol. 8, p. 1239, Feb. 2024.

[62] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, "Diagnosing Barren Plateaus with Tools from Quantum Optimal Control," *Quantum*, vol. 6, p. 824, Sept. 2022.

[63] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. Ortiz Marrero, M. Larocca, and M. Cerezo, "A lie algebraic theory of barren plateaus for deep parameterized quantum circuits," *Nature Communications*, vol. 15, no. 1, p. 7172, 2024.

[64] Q. T. Nguyen, L. Schatzki, P. Braccia, M. Ragone, P. J. Coles, F. Sauvage, M. Larocca, and M. Cerezo, "Theory for equivariant quantum neural networks," *PRX Quantum*, vol. 5, p. 020328, May 2024.

[65] A. Skolik, M. Cattelan, S. Yarkoni, T. Bäck, and V. Dunjko, "Equivariant quantum circuits for learning on weighted graphs," *npj Quantum Information*, vol. 9, no. 1, p. 47, 2023.

[66] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, "Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing," 2024.

[67] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA: MIT Press, 2 ed., 2018.

[68] L. Banchi, J. Pereira, and S. Pirandola, "Generalization in quantum machine learning: A quantum information standpoint," *PRX Quantum*, vol. 2, p. 040321, Nov 2021.

[69] M. C. Caro, E. Gil-Fuster, J. J. Meyer, J. Eisert, and R. Sweke, "Encoding-dependent generalization bounds for parametrized quantum circuits," *Quantum*, vol. 5, p. 582, Nov. 2021.

[70] M. C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, and P. J. Coles, "Generalization in quantum machine learning from few training data," *Nature Communications*, vol. 13, no. 1, p. 4919, 2022.

[71] H. Zhao and D.-L. Deng, "Entanglement-induced provable and robust quantum learning advantages," 2024.

[72] B. M. Terhal, "Quantum error correction for quantum memories," *Rev. Mod. Phys.*, vol. 87, pp. 307–346, Apr 2015.

[73] P. W. Shor, "Scheme for reducing decoherence in quantum computer memory," *Phys. Rev. A*, vol. 52, pp. R2493–R2496, Oct 1995.

[74] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, "High-threshold and low-overhead fault-tolerant quantum memory," *Nature*, vol. 627, no. 8005, pp. 778–782, 2024.

[75] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: Towards practical large-scale quantum computation," *Phys. Rev. A*, vol. 86, p. 032324, Sep 2012.

[76] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," *Nature*, vol. 567, no. 7749, pp. 491–495, 2019.

[77] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, "Hybrid quantum-classical algorithms and quantum error mitigation," *Journal of the Physical Society of Japan*, vol. 90, no. 3, p. 032001, 2021.

[78] S. Endo, S. C. Benjamin, and Y. Li, "Practical quantum error mitigation for near-future applications," *Phys. Rev. X*, vol. 8, p. 031027, Jul 2018.

[79] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits," *Phys. Rev. Lett.*, vol. 119, p. 180509, Nov 2017.

[80] Y. Li and S. C. Benjamin, "Efficient variational quantum simulator incorporating active error minimization," *Phys. Rev. X*, vol. 7, p. 021050, Jun 2017.

[81] Y. LeCun, C. Cortes, and C. Burges, "Mnist handwritten digit database," *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, vol. 2, 2010.

[82] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," 2017.

[83] A. Krizhevsky, G. Hinton, *et al.*, "Learning multiple layers of features from tiny images," 2009.

[84] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[85] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

[86] J. Sun, X. Yuan, T. Tsunoda, V. Vedral, S. C. Benjamin, and S. Endo, "Mitigating realistic noise in practical noisy intermediate-scale quantum devices," *Phys. Rev. Appl.*, vol. 15, p. 034026, Mar 2021.

[87] A. Galicia, B. Ramon, E. Solano, and M. Sanz, "Enhanced connectivity of quantum hardware with digital-analog control," *Phys. Rev. Res.*, vol. 2, p. 033103, Jul 2020.

[88] Y. Nam, J.-S. Chen, N. C. Pisenti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, *et al.*, "Ground-state energy estimation of the water molecule on a trapped-ion quantum computer," *npj Quantum Information*, vol. 6, no. 1, p. 33, 2020.

[89] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, "Circuit-centric quantum classifiers," *Physical Review A*, vol. 101, no. 3, p. 032308, 2020.

[90] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, pp. 242–246, 2017.

[91] D. A. Meyer and N. R. Wallach, "Global entanglement in multiparticle systems," *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4273–4278, 2002.

[92] G. K. Brennen, "An observable measure of entanglement for pure states of multi-qubit systems," *Quantum Info. Comput.*, vol. 3, p. 619–626, Nov. 2003.

[93] A. W. Harrow and R. A. Low, "Random quantum circuits are approximate 2-designs," *Communications in Mathematical Physics*, vol. 291, no. 1, pp. 257–302, 2009.

[94] P. Hayden, D. W. Leung, and A. Winter, "Aspects of generic entanglement," *Communications in Mathematical Physics*, vol. 265, no. 1, pp. 95–117, 2006.

[95] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, "Local random quantum circuits are approximate polynomial-designs," *Communications in Mathematical Physics*, vol. 346, no. 2, pp. 397–434, 2016.

[96] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," *Nature Communications*, vol. 12, no. 1, p. 1791, 2021.

[97] M. T. West, J. Heredge, M. Sevior, and M. Usman, "Provably trainable rotationally equivariant quantum machine learning," *PRX Quantum*, vol. 5, p. 030320, Jul 2024.

[98] Y. Wang, B. Qi, C. Ferrie, and D. Dong, "Trainability enhancement of parameterized quantum circuits via reduced-domain parameter initialization," *Phys. Rev. Appl.*, vol. 22, p. 054005, Nov 2024.

[99] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, "Exploiting symmetry in variational quantum machine learning," *PRX Quantum*, vol. 4, p. 010328, Mar 2023.

[100] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, *et al.*, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," 2022.

A Compatibility of Multi-Chip Ensemble VQC with Current and Near-Future Quantum Hardware

A.1 Current Quantum Hardware

The constraints of NISQ hardware, including limited scalability, noise, decoherence, and sparse qubit connectivity, necessitate novel algorithmic approaches tailored to these limitations [5, 18, 19, 50]. The proposed multi-chip ensemble VQC addresses these challenges by distributing quantum computations across multiple smaller chips, with classical postprocessing of measurement outputs to derive the final result. This modular design aligns with the constraints of current NISQ devices, offering a practical and scalable framework for hybrid quantum-classical computation.

The multi-chip ensemble VQC divides a large quantum circuit into k smaller subcircuits, each executed on a separate chip containing ℓ qubits ($n = k \times \ell$, where n is the total number of qubits). Unlike monolithic circuits on single chips, which suffer from increasing noise and decoherence with greater circuit depth [18, 51], this modular approach confines quantum operations to smaller, high-coherence regions, thereby mitigating noise accumulation. Each chip operates independently, and outputs are combined classically, ensuring robustness even in the presence of inter-chip noise or limited quantum connectivity [19]. By minimizing reliance on inter-chip quantum communication, the multi-chip ensemble VQC is particularly suited to NISQ devices, where such interactions are noisy or not fully realized.

This modular design directly addresses scalability limitations in NISQ hardware, particularly the restricted number of physical qubits. Current NISQ devices struggle to scale beyond a few dozen qubits due to fabrication and yield constraints [18, 51]. The multi-chip ensemble VQC overcomes these limitations by employing small quantum circuits multiple times and combining their measurement results classically. This enables processing of high-dimensional data even on devices with

relatively few physical qubits. By distributing computation across smaller chips, the framework allows independent fabrication and optimization of each chip, reducing complexity while increasing the yield of high-quality qubits. Additionally, the use of classical postprocessing leverages the computational power of classical hardware, extending the reach of quantum computation without overburdening quantum components.

Noise and decoherence, pervasive challenges in NISQ devices, are naturally mitigated within the multi-chip ensemble framework. By limiting the depth of quantum operations on each chip, qubits are less exposed to prolonged noise. The classical aggregation step also enhances resilience, as noise in individual chip outputs can be statistically absorbed or corrected in the final result. Techniques such as zero-noise extrapolation and probabilistic error cancellation [49, 79] can further improve the reliability of individual chip outputs, reinforcing the overall robustness of the framework.

Sparse qubit connectivity, another significant constraint of NISQ devices [18, 19, 87], is effectively managed by the multi-chip ensemble VQC. Each chip operates independently, and intra-chip connectivity suffices for implementing required quantum operations within subcircuits. The absence of strong inter-chip connectivity does not hinder functionality, as the classical processing step eliminates the need for high-fidelity inter-chip quantum gates. This compatibility ensures the framework's applicability to current superconducting [3, 74], ion-trap [88], and other quantum platforms, where connectivity is often constrained by physical qubit arrangements and fabrication limitations.

A.2 Near-Future Quantum Hardware

Quantum computing architectures have traditionally relied on single-chip designs where all qubits and associated control and readout circuitry are fabricated on a single monolithic chip. While effective for small-scale quantum processors, this approach faces severe scalability limitations due to increasing fabrication complexity, crosstalk, and reduced yield as the number of qubits grows. Multi-chip quantum computing, by contrast, addresses these limitations by modularizing the quantum system, distributing qubits and control elements across multiple chips. This design is inherently compatible with next generation quantum computing architecture explored in Rigetti's multi-chip tunable coupler [24], IBM Quantum Development & Innovation Roadmap [22], and IonQ's Reconfigurable Multicore Quantum Architecture (RMQA) [23].

Rigetti's modular superconducting qubit architectures using tunable couplers for high-fidelity communication between chips [24]. The multi-chip ensemble VQC aligns with this architecture by limiting inter-chip quantum operations, thereby minimizing the reliance on tunable couplers. Instead, each chip processes independent subcircuits, and classical postprocessing aggregates results. The modular design of the hardware supports the algorithm's scalability by enabling fabrication and optimization of smaller, high-quality chips, perfectly suited for independent computation within the VQC framework.

IBM's roadmap emphasizes modularity through quantum interconnects, such as m-couplers and l-couplers, designed to link multiple chips with minimal coherence loss [22] . The multi-chip ensemble VQC is fully compatible with this vision, as its classical aggregation of measurement results reduces the need for frequent inter-chip entanglement or communication. Additionally, IBM's development of error correction and circuit knitting technologies directly supports the scalable deployment of the algorithm by enabling more reliable execution of distributed quantum circuits across multiple chips.

IonQ's RMQA takes a reconfigurable multicore approach, where each core acts as an independent processor with photonic interconnects for high-fidelity communication [23]. The multi-chip ensemble VQC aligns seamlessly with this architecture by leveraging the modularity of RMQA. Each core can independently execute a subcircuit of the VQC, while the classical postprocessing step efficiently combines results. The flexibility of RMQA ensures that the algorithm can scale while mitigating noise and connectivity limitations inherent to trapped-ion systems.

In summary, the multi-chip ensemble VQC is inherently compatible with these cutting-edge or near-future multi-chip quantum architectures. By distributing quantum computations across smaller subcircuits and relying on classical postprocessing, it leverages the strengths of modular designs while minimizing inter-chip communication, offering a scalable solution for quantum optimization and machine learning tasks. As multi-chip systems continue to mature, the proposed multi-chip ensemble VQC framework represents a forward-compatible approach for leveraging the current and future quantum hardware.

B Quantum Entanglement of Multi-Chip Ensemble VQC

As the unitary operators $U(\boldsymbol{\theta})$ act on the encoded quantum state $\rho(x)$, the resulting state $U(\boldsymbol{\theta})\rho(\boldsymbol{x})U^{\dagger}(\boldsymbol{\theta})$ can exhibit quantum entanglement, manifesting as correlations between different parts of the quantum system. We define the entanglement as:

$$Ent(\rho; \boldsymbol{\theta}) = Ent\big[U(\boldsymbol{\theta})\rho(\boldsymbol{x})U^{\dagger}(\boldsymbol{\theta})\big], \tag{7}$$

where $Ent[\cdot]$ represents a measurement function of quantum entanglement.

One way to measure quantum entanglement is to obtain the entangling capability. The entangling capabiliy of a variational quantum circuit quantifies the circuit's proficiency in effectively delineating the solution space of the machine learning task and capturing non-trivial correlations within the quantum dataset [89,90]. The entangling capability can be obtained by sampling the circuit parameters and calculating the sample average of the Meyer-Wallach measure [91] for the resulting states [52]. More precisely, we take the estimate of the entangling capability to be

$$Ent = \frac{1}{|S|} \sum_{\theta_i \in S} Q(|\psi_{\theta_i}\rangle), \tag{8}$$

where $S = \{\theta_i\}$ is the set of sampled circuit parameter vectors and Q is the Meyer-Wallach measure. This measure with n-qubits is defined as

$$Q(|\psi\rangle) = \frac{2}{n} \sum_{j=1}^{n} (1 - Tr(\rho_j^2)), \tag{9}$$

where $\rho_j = Tr_{\backslash j}(|\psi\rangle\langle\psi|)$ is the reduced density matrix of the j-th qubit [92]. An entangling capability score of 0 indicates that the quantum cicrcuit exclusively generates product states, whereas a score of 1 denote that the circuit consistently produces highly entangled states.

To compare the level of quantum entanglement in single-chip and multi-chip ensemble VQCs, we consider two maximally entangled VQC models with equal number of total qubits (n) and identical ansatz design. The multi-chip ensemble VQC has k quantum chips, each composed of l-qubit subcircuits to form a large n-qubit circuit $(n = k \times l)$.

We first show that the Meyer-Wallach measure of the multi-chip ensemble VQC Q_{MC} is the average of the measures Q_c of the individual chips. Let $|\psi_{MC}\rangle$ be the n-qubit state generated by the multi-chip ensemble. By design, this state has a product structure across chips:

$$|\psi_{MC}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_k\rangle \tag{10}$$

where $|\psi_c\rangle$ is the l-qubit state generated by the c-th chip's VQC (with the same ansatz design but acting on l-qubits). The Meyer-Wallach measure for the *entire* n-qubit multi-chip ensemble system is:

$$Q_{MC}(|\psi_{MC}\rangle) = \frac{2}{n} \sum_{j=1}^{n} (1 - Tr(\rho_{j,MC}^2)) \tag{11}$$

where $\rho_{j,MC}^2 = Tr_{\backslash j}(|\psi_{MC}\rangle\langle\psi_{MC}|)$.

Let $Q_c(|\psi_c\rangle)$ be the Meyer-Wallach measure for the c-th chip's l-qubit state:

$$Q_c(|\psi_c\rangle) = \frac{2}{l} \sum_{j' \in c} (1 - Tr(\rho_{j',c}^2)), \tag{12}$$

where the sum is over the local indices j' within chip c, and $\rho_{j',c} = Tr_{\backslash j'}(|\psi_c\rangle\langle\psi_c|)$.

As shown previously, because $|\psi_{MC}\rangle$ is a product state across quantum chips, the reduced state of a qubit j located in chip c is determined only by the sate $|\psi_c\rangle$. That is, $\rho_{j,MC} = \rho_{j'c}$. Substituting this into the expression for Q_{MC}:

$$Q_{MC}(|\psi_{MC}\rangle) = \frac{2}{n} \sum_{c=1}^{k} \sum_{j' \in c} \big(1 - Tr(\rho_{j',MC}^2)\big). \tag{13}$$

We can rewrite this by multiplying and dividing by l:

$$Q_{MC}(|\psi_{MC}\rangle) = \frac{2}{n} \sum_{c=1}^{k} \frac{l}{2} \left(\frac{2}{l} \sum_{j' \in c} \left(1 - Tr(\rho_{j',MC}^2) \right) \right) = \frac{2}{n} \sum_{c=1}^{k} \frac{l}{2} Q_c(|\psi_c\rangle) \tag{14}$$

Since $n = k \times l$:

$$Q_{MC}(|\psi_{MC}\rangle) = \frac{2}{kl} \sum_{c=1}^{k} \frac{l}{2} Q_c(|\psi_c\rangle) = \frac{1}{k} \sum_{c=1}^{k} Q_c(|\psi_c\rangle). \tag{15}$$

Thus, Q_{MC} is the average of the Meyer-Wallach measures of the individual chips.

Next, we show that the Meyer-Wallach measure of single-chip VQC Q_{SC} is greater or equal to that of multi-chip ensemble VQC Q_{MC}. The core of the argument relies on comparing the potential purity $Tr(\rho_j^2)$ for a qubit j in the single-chip VQC versus the multi-chip ensemble VQC. In the multi-chip ensemble VQC, $\rho_{j,MC}$ only reflects entanglement with the $l-1$ other qubits in the same chip. Let j be in chip c, then $\rho_{j,MC} = \rho_{j',c}$. In the single-chip VQC, $\rho_{j,SC}$ reflects entanglement with all $n-1$ other qubits. Entanglement of a qubit j with additional qubits (those outside its quantum chip partition in the single-chip VQC) can only decrease its purity (or keep it the same if there is no entanglement with them). Therefore, $Tr(\rho_{j,SC}^2) \leq Tr(\rho_{j,MC}^2)$. This implies $(1 - Tr(\rho_{j,SC}^2)) \geq (1 - Tr(\rho_{j,MC}^2))$. Summing over all n qubits $j = 1, ..., n$:

$$\sum_{j=1}^{n} (1 - Tr(\rho_{j,SC}^2)) \geq \sum_{j=1}^{n} (1 - Tr(\rho_{j,MC}^2)) \tag{16}$$

Multiplying by $\frac{2}{n}$:

$$\frac{2}{n} \sum_{j=1}^{n} (1 - Tr(\rho_{j,SC}^2)) \geq \frac{2}{n} \sum_{j=1}^{n} (1 - Tr(\rho_{j,MC}^2)) \tag{17}$$

This gives the result

$$Q_{SC}(|\psi_{SC}\rangle) \geq Q_{MC}(|\psi_{MC}\rangle). \tag{18}$$

Finally, we consider the entangling capability of single-chip VQC and that of multi-chip ensemble VQC. The entangling capability defined in Equation B is the average of Meyer-Wallach measure over the parameter distributions. We assume the paameter distributions $\mathcal{P}_{SC}$ and $\mathcal{P}_l$ are chosen appropriately for the respective circuits (e.g., uniform random angles for rotation gates). If we average the inequality $Q_{SC} \geq Q_{MC}$ over the corresponding parameter distributions, the inequality is preserved:

$$\mathbb{E}_{\theta_{SC} \sim \mathcal{P}_{SC}} \left[Q_{SC}(|\psi_{SC}(\boldsymbol{\theta}_{SC})\rangle) \right] \geq \mathbb{E}_{\theta_{MC} \sim \mathcal{P}_{MC}} \left[Q_{MC}(|\psi_{MC}(\boldsymbol{\theta}_{MC})\rangle) \right], \tag{19}$$

where $\mathcal{P}_{MC}$ represents the joint distribution of independent draws from $\mathcal{P}_l$ for each quantum chip's parameters. This directly translates to:

$$Ent_{SC} \geq Ent_{MC}. \tag{20}$$

Therefore, multi-chip ensemble VQCs generally have higher levels of quantum entanglement than status-quo single-chip VQCs.

C   Relationship between Entanglement and Trainability

While high entanglement and the consequential high expressibility improve the quantum circuit's capacity to represent complex quantum states, they can also flatten the loss landscape, leading to barren plateaus [55–57]. Barren plateaus are regions in the loss landscape where the gradients of the loss function become exponentially small, making optimization challenging [21]. Given the loss function of a n-qubit VQC $\mathcal{L}(\boldsymbol{x}, y; \boldsymbol{\theta})$, a barren plateau occurs when the variance of partial derivatives $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}$ vanishes exponentially in n [31, 66]:

$$Var\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \right) \in \mathcal{O}(\frac{1}{2^n}). \tag{21}$$

As entanglement scales with the number of entangling gates (volume-law entanglement), the quantum states become so entangled that small changes in parameters have minimal impact on the output, causing gradients to vanish [55]: $Var\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right) \to 0$. This results in a flat loss landscape, where the model struggles to make progress during training.

As discussed by Patti et al. [57], the relationship between bipartite quantum entanglement and variance of gradients is:

$$Var\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right) \propto \frac{1}{2^S}, \tag{22}$$

where $S(\rho_\alpha) = -Tr[\rho_\alpha\log_2\rho_\alpha]$ denotes bipartite entanglement entropy and $\rho_\alpha = Tr_\alpha[\rho]$ is the reduced density matrix.

However, many variational quantum circuits (VQCs) generate *multipartite* entanglement across all qubits, and in that context a common global measure is the Meyer–Wallach Q or its average over circuit parameters—i.e., the entangling capability (Ent). Intuitively:

- Large bipartite entanglement in any cut shrinks gradient variance, leading to barren plateaus.
- Even more strongly, large multipartite entanglement across the entire n-qubit system indicates a circuit behaves almost *like the Haar distribution*—and that randomization also exacerbates barren plateaus.

Thus, one can replace bipartite entropy $S(\rho_\alpha)$ with a multipartite measure (such as Meyer–Wallach Q), and show a similar scaling that high entangling capability implies low gradient variance—that is, a higher risk of barren plateaus.

It is well known [52, 56, 93–95] that if the typical quantum state $|\psi(\boldsymbol{\theta})\rangle$ has near-maximal multipartite entanglement ($Ent \approx 1$), then the circuit distribution over unitaries $U(\boldsymbol{\theta})$ is typically close to being an approximate 2-design on $(\mathbb{C}^2)^{\otimes n}$. In other words, high entangling capability leads to Haar-like randomization of the entire n-qubit system. This is significant because random circuits that emulate Haar (or a 2-design) are precisely the ones known to exhibit barren plateaus for broad classes of cost functions.

Now we show that excessive multipartite entanglement (measured as entangling capability) can induce barren plateaus. As shown by McClean et al. [21], if $U(\boldsymbol{\theta})$ forms a global 2-design—or even approximates one—then for large n, the variance of partial derivatives $\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}$ vanishes as $O(\frac{1}{2^n})$. This indicates that 2-design behavior leads to exponentially vanishing gradient variance (i.e., barren plateaus).

When a circuit typically yields globally entangled states (i.e., *any* partition of the qubits is strongly entangled, and specifically the *one-qubit* marginals are thoroughly mixed), it implies that the circuit distribution matches the Haar measure up to *second moments*—the hallmark of a 2-design [93]. Equivalently, high entangling capability $Ent \approx 1$ means typical states $|\psi(\boldsymbol{\theta})\rangle$ are close to typical Haar-random states.

Given that 2-design behavior leads to exponentially vanishing gradient variance and excessive Ent means 2-design-like behavior, we obtain:

$$Var\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right) \propto \frac{1}{2^{Ent}}. \tag{23}$$

Therefore, the bigger the global (multipartite) entangling capability of the VQC, the more likely it is to randomize completely, forcing the training landscape into a barren plateau [21, 55, 57].

D Reducing Risk of Barren Plateaus Without Classical Simulability in Multi-Chip Ensembles

When the loss landscape of a VQC exhibit barren plateaus, exponential resources are required for training, prohibiting the successful scaling of the quantum circuit. Hence, identifying architectures and training strategies that provably do not lead to barren plateaus has become a highly active area of research. Examples of such strategies include shallow circuits with local measurements [59, 96], dynamics with small Lie algebras [62, 63, 97], identity initializations [60, 61, 98], and embedding symmetries into the circuit's architecture [64, 65, 99].

However, loss landscapes which provably do not exhibit barren plateaus can be simulated using a classical algorithm that runs in polynomial time [66]. Importantly, this simulation does not require VQCs implemented on a quantum device nor hybrid quantum-classical optimization loops. These arguments can be understood as a form of dequantization of the information processing capabilities of VQCs in barren plateau-free landscapes.

Here, we show that *multi-chip ensembles approach can reduce the risk of barren plateaus without making the VQC classically simulable*. While our method may not guarantee provable barren plateau-free VQCs, it can certainly reduce the risk of such phenomena while maintaining sufficient complexity to avoid classical simulability.

A circuit family is said to be in a *classically identifiable polynomial subspace* (hence *classically simulable*) if for all states or measurement outcomes generated by the circuit, the cost of exact (or approximate) classical simulation scales *polynomially* in n [66]. In practice:

- If each quantum chip (i.e., subcircuit) of the multi-chip ensemble VQC is constant size $l = \text{const}$, then the overall quantum state is a tensor product of $\mathcal{O}(n)$ small states, each with dimension $2^l = \text{const}$. Storing or simulating each sub-state requires $\mathcal{O}(1)$ resources, repeated n/l times, so total $\mathcal{O}(n)$. Hence, a multi-chip ensemble VQC with fixed l is trivially classically simulable.

- If l grows substantially with n—e.g. $l = \mathcal{O}(n)$—the dimension becomes exponentially large, so simulating each subcircuit can be exponentially costly. This is not guaranteed to be a known polynomial subspace.

We now show that sets l to grow with n—thus each chip alone can be large/hard—yet the factorized structure can stop full randomization across all n qubits, preventing or mitigating a global barren plateau.

D.1 Classical Hardness

Let $k = n/l$ where l is proportional to n. Then k does not grow exponentially with n. Each subcircuit of the multi-chip ensemble VQC:

- Acts on l-qubits with $l \in \mathcal{O}(n)$;
- Potentially deep or universal enough within each l-sized subcircuit that simulating it *classically* may cost $\mathcal{O}(2^l)$, i.e., exponential in l, which is exponential in n.

Hence, a naive classical simulation of the total quantum state in multi-chip ensemble VQC is:

$$\rho_{MC} = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_k, \tag{24}$$

where each ρ_j is l-qubit and $l \sim n$. Storing or simulating ρ_j in full amplitude format costs $\mathcal{O}(2^l)$ per subcircuit, thus overall $\mathcal{O}(k2^l) \approx 2^l$. This is exponential in n. Therefore, this multi-chip ensemble architecture is *not in a known polynomial subspace*.

D.2 Reduced risk of Barren Plateaus

2-design arguments typically require global randomization across all n-qubits. Since multi-chip ensemble factorizes the unitary into k disjoint subcircuits, the entire distribution cannot be a global 2-design. This is particularly due to the following reasons:

No Global Entanglement By design, multi-chip ensemble VQC has no global entanglement. The total circuit of the multi-chip ensemble is defined as $U_{MC}(\boldsymbol{\theta}) = \bigotimes_{j=1}^{k} U_j(\boldsymbol{\theta}_j)$, which means that there is no inter-chip quantum connections between individual subcircuits. Thus, randomizing within each subcircuit does not generate global (inter-chip) entanglement.

Global Loss Functions A number of research [21, 55–57] has shown that deep random circuits produce barren plateaus precisely by approximating a global 2-design on $(\mathbb{C}^2)^{\otimes n}$. But multi-chip ensemble VQC is strictly subspace-limited: it can only produce states in a factorized manifold.

Gradient Variance Argument Let $\mathcal{L}(\boldsymbol{x}, y; \boldsymbol{\theta})$ be the loss function of the multi-chip ensemble VQC. The partial derivative w.r.t. θ_i in subcircuit j is confined to that specific quantum chip. Because there is *no global mixing*, the randomization effect that typically drives the variance to $\mathcal{O}(\frac{1}{2^n})$ does not apply to the *full n*-qubits. Each chip might see variance $\mathcal{O}(\frac{1}{2^l})$ for its local subset, but not $\mathcal{O}(\frac{1}{2^n})$.

- If $l \ll n$, one may get $\mathcal{O}(\frac{1}{2^l})$, which is still a big improvement over $\mathcal{O}(\frac{1}{2^n})$.
- If l grows with n, one wants to ensure the circuit is not fully randomizing even that chip. One can keep a moderate depth or partial connectivity within each chip so as to avoid approximate 2-design in each l-qubit subcircuit as well.

Because the circuit never forms a global 2-design, typical cost gradients do not vanish as $\mathcal{O}(\frac{1}{2^n})$. Instead, one can get

$$Var\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}\right) \in \Omega(\frac{1}{2^l}). \tag{25}$$

And since $l \leq n$, that scaling is strictly better than the dreaded $\frac{1}{2^n}$. Therefore, a multi-chip ensemble VQC with moderate or carefully chosen subcircuit depth can mitigate the exponential gradient decay that plagues fully entangling global random circuits.

D.3   Summary

We can design multi-chip ensemble VQCs which are not in a polynomial subspace because with $l \in \mathcal{O}(n)$, each chip is an l-qubit subcircuit that is not trivially simulable. The product of these l-qubit states is still dimension $2^{kl} = 2^n$, requiring exponential overhead in n if one attempts naive amplitude simulation.

By design, there are no inter-chip entangling gates in multi-chip ensemble VQCs. This means that there is no global randomization, which leads to no approximation of global 2-design. The gradient variance w.r.t. parameters in each block is not suppressed by a factor $\mathcal{O}(\frac{1}{2^n})$ but only by something related to subcircuit size l. Given that $l = n/k$, the circuit might randomize an $\frac{n}{k}$-subset at best. Even then, one can maintain partial structure within each quantum chip to avoid a local 2-design. Thus, the typical scaling of gradient variance is $\mathcal{O}(\frac{1}{2^l})$, which is not necessarily $\mathcal{O}(\frac{1}{2^n})$. Indeed, one can design the subcircuit so it, too, fails to become fully random (or at least not a 2-design).

Hence we have shown that there exist multi-chip ensemble architectures (for instance, with $l \propto n$ and moderate subcircuit depth that fails to approximate a 2-design) that simultaneously:

- Generate states requiring exponential overhead to simulate classically (i.e. not in a known polynomial subspace).
- Avoid or mitigate a barren plateau, because no single set of gates randomizes the entire n qubits.

E   Relationship between Entanglement and Generalizability

To understand how entanglement affects the generalizability of quantum circuits, we consider the generalization error [70]. The goal of a quantum circuit model is to minimize the expected loss over the distribution $\mathcal{P}$ of data $(\boldsymbol{x}, \boldsymbol{y})$, represented as:

$$R(\boldsymbol{\theta}) = \mathbb{E}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mathcal{P}}\left[\mathcal{L}(f_{\boldsymbol{\theta}}(\boldsymbol{x}), y)\right], \tag{26}$$

where $\mathcal{L}$ is a loss function (e.g. mean-squared loss, cross-entropy, etc.). As the distribution $\mathcal{P}$ is unknown, the expected loss $R(\boldsymbol{\theta})$ is typically estimated from a finite training set $S = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^N$, leading to the training loss:

$$R_S(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}\left(f_{\boldsymbol{\theta}}(\boldsymbol{x}), y_i\right). \tag{27}$$

A trained model $\hat{\boldsymbol{\theta}}(S)$ is chosen to minimize or nearly minimize R_S. The generalization error is the difference between the expected loss and the training loss:

$$gen(\boldsymbol{\theta}) = R(\hat{\boldsymbol{\theta}}) - R_S(\hat{\boldsymbol{\theta}}), \tag{28}$$

which quantifies how well the model generalizes to unseen data.

Classically, under mild conditions on $\mathcal{L}$, one can decompose the expected test error into bias and variance terms—reflecting how (on average) a learned hypothesis $\hat{\boldsymbol{\theta}}(S)$ differs from the optimal function [67]. Formally, one writes:

$$\mathbb{E}_S[(f_{\hat{\boldsymbol{\theta}}(S)}(x) - y)^2] = \underbrace{\left(\mathbb{E}_S[f_{\hat{\boldsymbol{\theta}}(S)}(x)] - f^*(x)\right)^2}_{\text{bias}^2} + \underbrace{\mathbb{E}_S\left[\left(f_{\hat{\boldsymbol{\theta}}(S)}(x) - \mathbb{E}_S[f_{\hat{\boldsymbol{\theta}}(S)}(x)]\right)^2\right]}_{\text{variance}} \quad (29)$$

A model with high capacity typically has lower bias but higher variance. In QML, circuit capacity can be tied to Rademacher complexity [68] or expressibility [69]. Next, we incorporate quantum entanglement into this narrative.

We first partition the space of all possible $\boldsymbol{\theta}$ into sets Θ_k such that for each $\theta \in \Theta_k$, the average entanglement satisfies

$$Ent(\theta) \leq \gamma_k, \quad (30)$$

where Ent is the entangling capability and γ_k is some increasing sequence $\gamma_1 < \gamma_2 < \cdots \gamma_M \leq 1$. Then we construct a nested hierarchy of circuit families:

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_M = \{f_\theta : \boldsymbol{\theta} \in \Theta\} \quad (31)$$

where each $\mathcal{F}_k$ is is realized by a set of parameters that produce an upper bound on the average entanglement:

$$Ent(\mathcal{F}_k) \leq \gamma_k. \quad (32)$$

In other words, at level k, we only allow those parameter choices (and gate structures) that keep the average global entanglement less or equal to γ_k. Constraining Ent effectively limits the circuit's capacity to produce large-scale entanglement across the entire data set.

We next define a complexity penalty $\Omega(\gamma_k)$ that increases with γ_k. Intuitively, if the circuit can reach larger entangling capability, it has a larger capacity and thus is penalized for potentially overfitting. The simplest approach is to let $\Omega(\gamma_k) = \alpha\gamma_k$ for some $\alpha > 0$. Using the complexity penalty, we can define the penalized training loss:

$$\tilde{R}_S(\theta, k) = R_S(\theta) + \Omega(\gamma_k), \Omega' \geq 0. \quad (33)$$

By design, picking a higher γ_k means letting the circuit produce more entanglement on average, which we penalize with a larger $\Omega(\gamma_k)$.

Using covering-number or uniform convergence arguments (which is standard in statistical learning), one obtains:

$$\Pr_S\left[\sup_{\theta \in \Theta_k} |R(\theta) - R_S(\theta)| \leq \epsilon_k(\gamma_k, N)\right] \geq 1 - \delta_k, \quad (34)$$

for some for some error probability δ_k. Typically ϵ_k is an increasing function of the capacity of Θ_k; here that capacity is correlated with γ_k.

We unify these bounds across $\mathcal{F}_1, ..., \mathcal{F}_M$, obtaining the following statement:

$$\Pr_S\left[\forall k, \forall \theta \in \Theta_k : |R(\theta) - R_S(\theta)| \leq \epsilon_k\right] \geq 1 - \sum_k \delta_k. \quad (35)$$

The chosen solution $\hat{\theta} \in \Theta_{\hat{k}}$ satisfies

$$R_S(\hat{\theta}) + \Omega(\gamma_{\hat{k}}) \leq R_S(\theta) + \Omega(\gamma_k) \quad \forall k, \theta \in \Theta_k. \quad (36)$$

Combining with the uniform convergence guarantee $|R(\theta) - R_S(\theta)| \leq \epsilon_k$:

$$R(\hat{\theta}) \leq R_S(\hat{\theta}) + \epsilon_{\hat{k}} \leq \left[R(\theta) + \epsilon_{\hat{k}}\right] + \Omega(\gamma_k) - \Omega(\gamma_{\hat{k}}), \quad \forall k, \theta \in \Theta_k. \quad (37)$$

After arrangement, this becomes:

$$R(\hat{\theta}) - R_S(\hat{\theta}) \leq \left[R(\theta) - R_S(\theta)\right] + \left[\epsilon_k + \Omega(\gamma_k)\right] - \left[\epsilon_{\hat{k}} + \Omega(\gamma_{\hat{k}})\right]. \quad (38)$$

Minimizing the right-hand-side over k and $\theta \in \Theta_k$ yields:

$$gen(\theta) = R(\hat{\theta}) - R_S(\hat{\theta}) \leq \min_{k, \theta \in \Theta_k} \left\{ \underbrace{\left(R(\theta) - R_S(\theta)\right)}_{\text{bias}} + \underbrace{\epsilon_k + \Omega(\gamma_k)}_{\text{variance}} \right\}. \quad (39)$$

21

The term $R(\theta) - R_S(\theta)$ can be viewed as a form of bias, because in a limited sub-family, the best circuit might not perfectly fit all data. The $\epsilon_k + \Omega(\gamma_k)$ stands in for a variance/complexity penalty that increases with γ_k.

Therefore, increasing the allowed entangling capability γ_k reduces bias (the circuit can represent a more complex function to match data) but increases variance risk (ϵ_k or $\Omega(\gamma_k)$). This is the quantum analog of the bias–variance trade-off based on quantum entanglement.

F Comparison between Quantum Errors in Multi-Chip Ensemble VQCs and Single-Chip VQCs

Given an input quantum state $\rho_{in}(x)$ from input data x, the ideal output quantum state ρ_{out}^{ideal} is expressed as

$$\rho_{out}^{ideal} = \mathcal{U}_{N_g} \circ \mathcal{U}_{N_g-1} \cdots \circ \mathcal{U}_1(\rho_{in}(x)), \tag{40}$$

where $\mathcal{U}(\cdot)$ represents ideal noiseless quantum channels corresponding to unitary gates U, i.e., $\mathcal{U}(\cdot) \equiv U(\cdot)U^\dagger$. The number of gates is denoted as N_g. For noisy output states, we have:

$$\rho_{out} = \mathcal{E}_{N_g} \circ \mathcal{U}_{N_g} \cdots \circ \mathcal{E}_1 \circ \mathcal{U}_1(\rho_{in}(x)), \tag{41}$$

where $\mathcal{E} \circ \mathcal{U}$ denotes noisy quantum gate, with $\mathcal{E}$ as the noise channel. Following prior work [77], we assume Markovian noise where noise channels $\mathcal{E}$ are independent.

The goal of quantum error mitigation is to estimate the expectation value of $\mathrm{Tr}[H\rho_{out}^{ideal}]$ for a given Hermitian observable H. Using the quantum circuit outputs, an estimator $\hat{H}$ for $\mathrm{Tr}[H\rho_{out}^{ideal}]$ can be constructed. The mean square error (MSE) of $\hat{H}$, quantifying its deviation from the true value, is given by:

$$MSE[\hat{H}] = \mathbb{E}[(\hat{H} - \mathrm{Tr}[H\rho_{out}^{ideal}])^2]. \tag{42}$$

Reducing $MSE[\hat{H}]$ is the primary goal of quantum error mitigation [49, 77]. After N_{cir} runs of the noisy quantum circuit, the noisy sample mean $\bar{H}_{\rho_{out}}$ estimates $\mathrm{Tr}[H\rho_{out}]$, with its MSE expressed as:

$$MSE[\bar{H}_{\rho_{out}}] = (\mathrm{Tr}[H\rho_{out}] - \mathrm{Tr}[H\rho_{out}^{ideal}])^2 + \frac{(\mathrm{Tr}[H^2\rho_{out}] - \mathrm{Tr}[H\rho_{out}]^2)}{N_{cir}}, \tag{43}$$

where the first term is the bias and the second term is the variance of quantum errors.

F.1 Bias of Quantum Errors $(\mathbf{Tr}[H\rho_{out}] - \mathbf{Tr}[H\rho_{out}^{ideal}])^2$

For single-chip VQCs, the noisy output state involves n-qubits affected by N_g noise channels: $\rho_{out} = \mathcal{E}_{N_g} \circ \mathcal{U}_{N_g} \cdots \circ \mathcal{E}_1 \circ \mathcal{U}_1(\rho_{in}(x))$. Each noise channel $\mathcal{E}_i$ is modeled as $\mathcal{E}_i(\rho) = (1-\epsilon)\rho + \epsilon\mathcal{D}i(\rho)$, where ϵ is the error rate and $\mathcal{D}i$ represents an error operation. Errors compound multiplicatively across n qubits and N_g gates, yielding:

$$(\mathrm{Tr}[H\rho_{out}] - \mathrm{Tr}[H\rho_{out}^{ideal}])^2 \propto (1+\epsilon)^{nN_g} \approx \exp(nN_g\epsilon). \tag{44}$$

For multi-chip ensemble VQCs, each chip processes $l = n/k$ qubits, producing noisy output states ρ_{out}^i. The bias for each chip scales similarly:

$$(\mathrm{Tr}[H_i\rho_{out}^i] - \mathrm{Tr}[H_i\rho_{out}^{i,ideal}])^2 \propto \exp(lN_g\epsilon) = \exp\left(\frac{n}{k}N_g\epsilon\right). \tag{45}$$

When combining outputs from k chips, individual biases add linearly:

$$\sum_{i=1}^{k} \exp\left(\frac{n}{k}N_g\epsilon\right) = k\exp\left(\frac{n}{k}N_g\epsilon\right). \tag{46}$$

Thus, while the bias for single-chip VQCs scales as $\exp(nN_g\epsilon)$, the bias for multi-chip ensemble VQCs scales as $k\exp\left(\frac{n}{k}N_g\epsilon\right)$. Since $\exp(n) > k\exp\left(\frac{n}{k}\right)$ for $k > 1$, multi-chip ensemble VQCs have significantly reduced bias of quantum errors.

F.2   Variance of Quantum Errors $\frac{(\mathbf{Tr}[H^2\rho_{out}] - \mathbf{Tr}[H\rho_{out}]^2)}{N_{cir}}$

In a single-chip VQC, variance of quantum error arises from random noise (e.g., gate errors, decoherence, finite sampling) that accumulates across all n-qubits. The variance decreases with the number of circuit runs (i.e., shots of the quantum circuit), scaling as $\frac{1}{N_{cir}}$.

For multi-chip ensemble VQCs, variance of quantum error is confined to each chip because there are no entangling gates between subsystems. Each chip contributes independently to the final measurement output. Combining the outputs of k-chips averages out some of the variance, reducing the overall variance compared to the single-chip VQC. The total variance of a multi-chip ensemble VQC decreases as $\frac{1}{kN_{cir}}$ due to averaging over k independent chips.

Since $\frac{1}{N_{cir}} > \frac{1}{kN_{cir}}$ for $k > 1$, multi-chip ensemble VQCs achieve lower variance of quantum errors compared to single-chip VQCs.

F.3   Summary

Multi-chip ensemble VQCs reduce both bias and variance of quantum errors compared to single-chip VQCs. This dual reduction is achieved without a bias-variance trade-off, providing enhanced noise resilience. By leveraging circuit decomposition, multi-chip ensemble VQCs offer a "free lunch" where both bias and variance are improved, ensuring robust and scalable quantum computation.

G   Experimental Model Design

We implemented three quantum-classical autoencoder models: (1) single-chip VQC (Figure 3a); (2) multi-chip ensemble VQC with classical dimension reduction (Figure 3b); and (3) multi-chip ensemble VQC without dimension reduction (Figure 3c). These models integrate classical and quantum components for efficient processing of high-dimensional data, ensuring scalability and robust learning. A classical autoencoder corresponding to the single-chip VQC autoencoder was also used to establish a baseline.

All experiments utilized the Pennylane library [100], integrated with PyTorch for seamless quantum-classical hybrid computations. The experiments were conducted on a Linux server (Kernel 5.14) with 128 CPU cores, 256 threads (x86-64 architecture), 503.14 GB RAM, and an NVIDIA A100-PCIE GPU with 40 GB memory. The software environment included Python 3.11.7, PyTorch 2.5.0+cu121, and CUDA 12.1.

G.1   Single-Chip VQC with Classical Dimension Reduction

The single-chip quantum autoencoder comprises a classical encoder, a VQC, and a classical decoder.

The classical encoder reduces the high-dimensional input data (*input_dim*) into a lower-dimensional quantum-compatible representation (n_{qubits}) using a fully connected layer: *input_dim* $\rightarrow n_{qubits}$.

The VQC then processes the encoded features using a single quantum circuit with n_{qubits} and a depth of d. The circuit includes parameterized RX, RY, RZ, and CRX gates, enabling flexible state evolution and entanglement. A single measurement (Pauli-Z expectation) is taken from the first qubit, yielding one output.

Finally, the classical decoder reconstructs the original input from the single measurement using a fully connected layer: $1 \rightarrow$ *input_dim*.

G.2   Multi-Chip Ensemble VQC with Classical Dimension Reduction

The multi-chip ensemble quantum autoencoder extends the single-chip design by distributing computations across n_{chips} independent VQCs.

The classical encoder maps the high-dimensional input data (*input_dim*) into n_{qubits}, where n_{qubits} is the total number of qubits across all chips: *input_dim* $\rightarrow n_{qubits}$. The encoded features are shuffled to ensure that each chip processes a representative subset of the input data.

23

(a) Single-chip

(b) Multi-chip Ensemble (reduced)
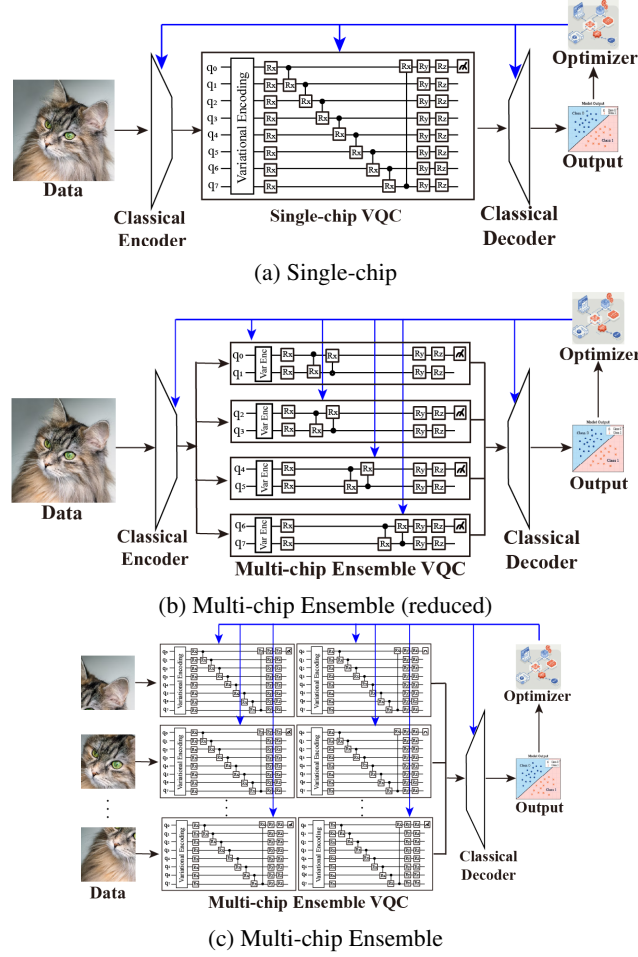

(c) Multi-chip Ensemble

Figure 3: Processing High Dimensional Data: Single-chip vs Multi-chip Ensemble VQCs

The quantum component is composed of n_{chips} VQCs, each operating on a disjoint subset of the n_{chips}. Each chip processes n_{qubits}/n_{chips} independently. For each chip:

- Parameterized RX, RY, RZ gates and CRX gates are applied to encode data and introduce intra-chip entanglement.
- A single measurement (Pauli-Z expectation) is taken from the first qubit of each chip, resulting in n_{chips} outputs.

Finally, the classical decoder aggregates the n_{chips} quantum measurements and reconstructs the original input using a fully connected layer: $n_{chips} \rightarrow input\_dim$.

G.3 Multi-Chip Ensemble without Classical Dimension Reduction

The quantum component consists of n_{chips} independent VQCs, each operating on n_{qubits} qubits with a circuit depth of d. Input data is divided into n_{chips} disjoint subsets, ensuring that each chip processes a representative portion of the input. To achieve this, the input is first split and then shuffled across the chips. This mechanism mimics classical ensemble learning techniques, such as feature bagging, to promote diverse and robust learning across the ensemble.

G.3.1 Quantum Circuits

Each VQC consists of parameterized RX, RY, and RZ gates for single-qubit rotations and CRX gates to introduce intra-chip entanglement. Entanglement is confined to qubits within a chip, preventing

Table 1: Experimental Model Design: Proof-of-Concept

(a) Classical & Single-Chip Autoencoder

Model Features	Classical	Single-Chip Quantum (with dimension reduction)
Optimizer	Adam	Adam
Learning Rate	0.001	0.001
Classical Encoder	1 Linear layer	1 Linear layer
Classical Decoder	1 Linear layer	1 Linear layer
Variational Encoding	-	RY
Number of Layers	2 Linear layers	2 VQC layers
Layer Structure	Linear(32,32)	RX, RY, RZ, CRX

(b) Multi-Chip Ensembles Quantum Autoencoder

Model Features	With dimension reduction	Without dimension reduction
Optimizer	Adam	Adam
Learning Rate	0.001	0.001
Classical Encoder	1 Linear layer	–
Classical Decoder	1 Linear layer	1 Linear layer
Variational Encoding	RY	RY
Number of Layers	2 VQC layers	2 VQC layers
Layer Structure	RX, RY, RZ, CRX	RX, RY, RZ, CRX

cross-chip quantum connections. At the end of the quantum circuit, a single measurement (Pauli-Z expectation value) is performed on the first qubit of each chip, yielding n_{chips} measurements in total. These outputs serve as the quantum component's contribution to the overall model.

G.3.2 Classical Components

In this multi-chip ensemble VQC model, there is no classical encoding layer for dimension reduction. Instead, the input data is simply partitioned into n_{chips} subsets and shuffled to ensure representative feature allocation to each quantum circuit.

The classical decoder aggregates the outputs of all VQCs (n_{chips}) using a fully connected layer to reconstruct the original input. The decoder maps the quantum measurements to the input dimension ($n_{chips} \rightarrow input\_dim$).

G.3.3 Scalability and Reproducibility

The modular design of multi-chip ensemble VQCs enables efficient scaling of the quantum component by increasing n_{chips}, while each chip processes a fixed number of qubits, ensuring compatibility with NISQ devices. The quantum backend leverages PennyLane's lightning.qubit simulator, and the entire system is integrated with PyTorch for seamless hybrid computations. The parameter-shift rule allows for end-to-end training via backpropagation, ensuring compatibility with standard optimization workflows.

This design balances quantum and classical resources, demonstrating the practicality of multi-chip ensemble VQCs for high-dimensional machine learning tasks. By explicitly defining the model components and data flow, we ensure reproducibility for future studies and extensions.

G.4 Classical Autoencoder

To establish a baseline for comparison with the quantum autoencoder models, we implemented a classical autoencoder that mirrors the overall structure of the single-chip VQC quantum autoencoder. This design ensures a fair comparison by aligning the architecture of the classical and quantum models, particularly in terms of the encoder, circuit (hidden layers), and decoder components.

The classical encoder maps the high-dimensional input (*input_dim*) to a lower-dimensional latent space with size n_{qubits}, matching the input to the quantum circuit in the single-chip VQC model.

The intermediate processing stage comprises fully connected layers that simulate the operations of the VQC. The number of hidden layers matches the circuit depth d of the single-chip VQC. The first layer maps $n_{qubits} \rightarrow 32$ and the final layer maps $32 \rightarrow n_{chips}$, corresponding to the number of measurements in the single-chip quantum autoencoder. This processing stage is implemented using a sequence of fully connected layers.

Finally, the latent features are reconstructed into the original input using a linear decoder: $n_{chips} \rightarrow$ *input_dim*. This mirrors the output structure of the quantum autoencoder.

G.5   PhysioNet EEG Dataset

We use the motor-imagery subset of the PhysioNet EEG [84] corpus recorded with the BCI2000 system [85]. The release comprises 1522 one- and two-minute runs from 109 subjects; we select the left- versus right-hand imagery runs sampled at 16 Hz from 64 scalp channels. Each trial is reshaped to a 64×51 matrix and flattened to a 3264-dimensional vector. The learning task is binary classification of the imagined hand movement.

G.6   QCNN Architecture for PhysioNet EEG

Table 2: Experimental Model Design: QCNN

Model Features	Classical	Single-Chip (with dimension reduction)	Multi-Chip Ensemble (with dimension reduction)
Optimizer	Adam	Adam	Adam
Learning Rate	0.001	0.001	0.001
Classical Preprocessing	1 Linear layer	1 Linear layer	-
Classical Postprocessing	1 Linear layer	1 Linear layer	1 Linear layer
Variational Encoding	-	RY	RY
Convolutional Layers	2 Linear layers	2 VQC layers	2 VQC layers
Pooling Layers	2 Linear layers	2 VQC layers	2 VQC layers
Layer Structure	Linear(32,32)	U3, IsingZZ, IsingYY, IsingXX	U3, IsingZZ, IsingYY, IsingXX

We implement a quantum convolutional neural network (QCNN) [4] and adapt it to our multi-chip setting and the high-dimensional PhysioNet EEG input ($62 \times 52 = 3264$ features).

The single-chip baseline uses $n = 8$ qubits, number of convolutional and pooling layers $d = 2$, and a fully connected layer that maps the raw feature vector to one rotation angle per qubit; these angles are loaded with R_Y-type variational encoding. Each convolutional layer applies per neighboring wire pair the sequence $\text{U3} \rightarrow \text{IsingZZ} \rightarrow \text{IsingYY} \rightarrow \text{IsingXX} \rightarrow \text{U3}$, giving $18dn$ trainable parameters, followed by conditional pooling that measures every second qubit, applies a three-parameter U3 to its partner, and thus halves the wire list. Including pooling and measurement, the circuit contains $18dn + 3d(n/2) + n$ quantum parameters and $d_{\text{in}}n + n$ classical parameters from the input layer and bias.

To remove classical dimension reduction in the multi-chip ensemble variant, we split the 3264 features evenly over $k = 272$ chips with $l = 12$ qubits each; every chip executes an independent copy of the QCNN block and the scalar outputs are averaged before the final softmax loss.

Retaining the U3+Ising kernel preserves the expressibility of the original QCNN while maintaining differentiability via parameter-shift rules, conditional pooling logarithmically reduces qubit count and mitigates barren plateaus, and the partition-then-average strategy allows the full EEG sequence to be processed quantum-natively, directly testing the scalability claim of our multi-chip framework.

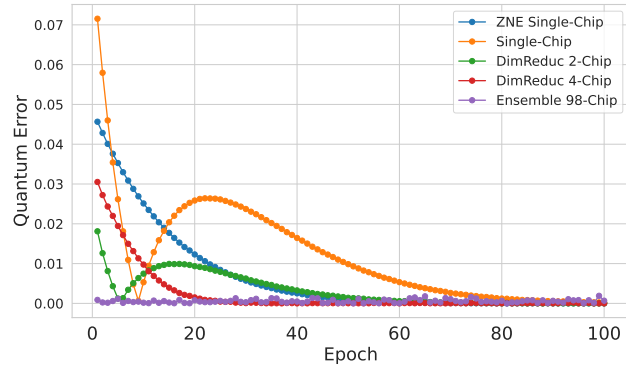## H   Experimental Results: FashionMNIST, CIFAR-10, PhysioNet EEG

We present the experimental results of FashionMNIST, CIFAR-10, and PhysioNet EEG datasets.
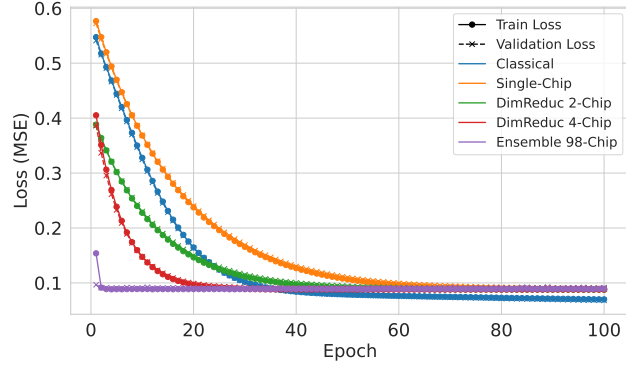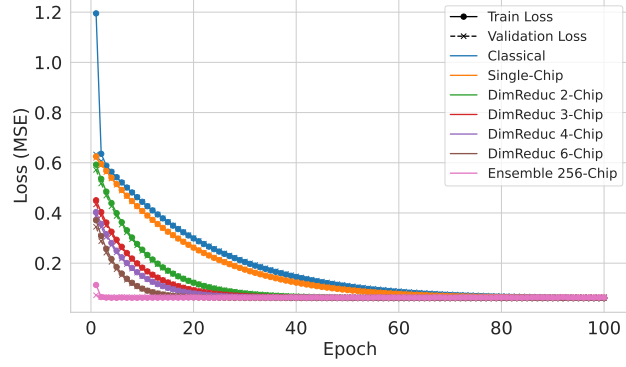
(a) Model Performance
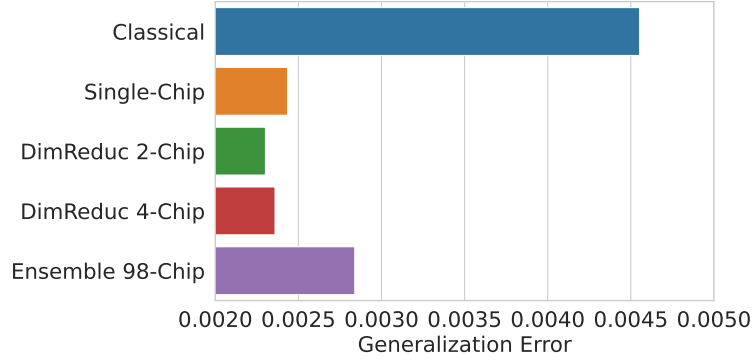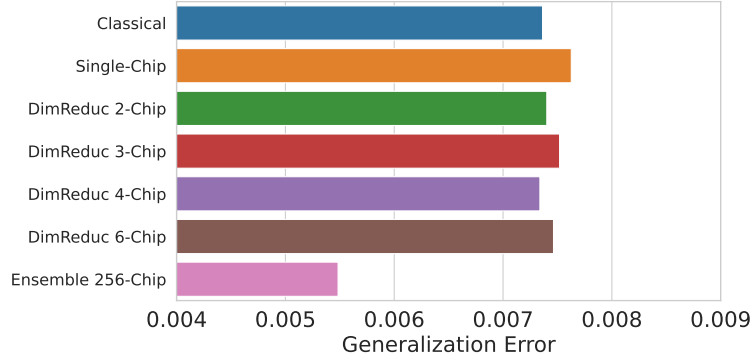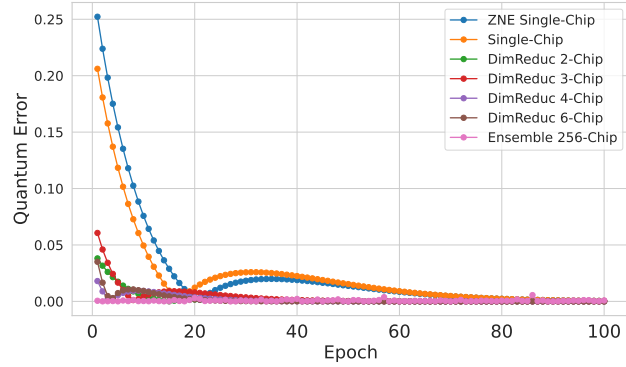

(b) Generalizability

(c) Noise Resilience

Figure 4: Experimental Results on FashionMNIST. DimReduc 2-Chip and 4-Chip indicate multi-chip ensemble VQC models with classical dimension reduction. Ensemble 98-Chip denotes multi-chip ensemble VQC model without classical dimension reduction.

(a) Model Performance

(b) Generalizability

(c) Noise Resilience

Figure 5: Experimental Results on CIFAR-10. DimReduc 2-Chip, 3-Chip, 4-Chip, and 6-Chip indicate multi-chip ensemble VQC models with classical dimension reduction. Ensemble 256-Chip denotes multi-chip ensemble VQC model without classical dimension reduction.
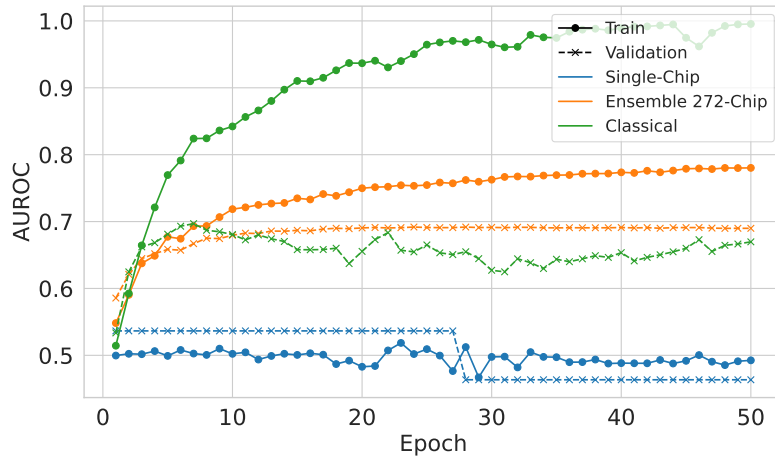
Figure 6: Model Performance

Figure 7: Experimental Results on PhysioNet EEG dataset. Ensemble 272-Chip denotes multi-chip ensemble VQC model without classical dimension reduction.