Optimization of Neural Networks Using Quantum Computing Techniques

[1] Ayesha Javed, [2] Muhammad Arqam Raza, [3] Sana Khalid

[1] Department of Computer Science, National University of Sciences and Technology (NUST), Islamabad, Pakistan
[2,]Department of Electrical Engineering, University of the Punjab, Lahore, Pakistan
[3,] School of Systems & Technology, University of Management and Technology (UMT), Lahore, Pakistan
***Corresponding Author:** ayesha.javed@seecs.nust.edu.pk *

Abstract: *Quantum computing offers new algorithmic primitives—superposition, entanglement, and interference—that can accelerate or qualitatively change parts of the machine-learning pipeline. This article surveys optimization of neural networks using quantum techniques, focusing on three avenues: (i) quantum-accelerated linear algebra and sampling to speed up gradient and Hessian-vector computations; (ii) hybrid variational algorithms where parameterized quantum circuits (PQCs) act as feature maps or layers within classical networks; and (iii) quantum-aware optimizers such as the quantum natural gradient (QNG) that respect the geometry of quantum states. We synthesize theoretical results with NISQ-era constraints (noise, barren plateaus, trainability), propose design patterns for end-to-end systems, and present an illustrative comparison of convergence between classical SGD and a hybrid QNG setup. The analysis indicates that, for structured data and appropriately shallow, problem-matched ansätze, hybrid methods can improve sample efficiency and generalization while remaining hardware-feasible.*

Keywords: *quantum machine learning, variational quantum circuits, quantum natural gradient, hybrid optimization, barren plateaus, kernel methods, quantum feature maps, NISQ hardware*

I Introduction

Neural network training reduces to high-dimensional non-convex optimization. Classical accelerations—stochastic gradients, momentum, adaptive methods, second-order preconditioning—remain bounded by classical sampling and linear-algebra costs. Quantum computing introduces state spaces that scale exponentially in qubits and transformations that can emulate certain kernels or probability amplitudes more compactly. While fault-tolerant quantum speedups for generic deep learning remain distant, near-term (NISQ) strategies integrate small quantum processors as co-processors in hybrid loops. Parameterized quantum circuits can act as (i) expressive feature maps that embed classical data into Hilbert space, (ii) layers whose expectation values feed classical networks, or (iii) learners optimized directly via variational principles. Optimization then hinges on efficient, low-variance gradient estimation, geometry-aware updates, and mitigation of trainability pathologies such as barren plateaus. This paper consolidates algorithmic tools and practical recipes to operationalize quantum benefits for neural optimization today.

Foundations of Quantum Advantage in Optimization

Quantum computing introduces a fundamentally different computational paradigm that leverages *superposition*, *entanglement*, and *interference* to represent and manipulate information in exponentially large state spaces. In the context of optimization—particularly for neural networks—

these quantum phenomena enable more compact encodings of complex functions and potentially faster convergence to global minima. **Superposition** allows a quantum system to explore multiple solution candidates simultaneously, effectively evaluating large portions of the search space in parallel. **Entanglement** links qubits in ways that capture correlations across variables, enabling the representation of intricate cost landscapes beyond classical tensor capacity. Techniques like **amplitude estimation** provide quadratic speedups in evaluating expectation values and gradients, while **phase kickback** transforms derivatives into measurable phase shifts that can be efficiently extracted through quantum interference.

However, realizing these theoretical advantages requires careful distinction between *algorithmic complexity promises* and *oracle assumptions*. Many early results presuppose ideal access to quantum oracles or noiseless hardware, conditions not yet met in near-term quantum devices. When measurement overhead, circuit depth, and error rates are included, the net benefit can diminish unless optimizations are amortized across large-scale or repetitive tasks. The true *quantum advantage* in neural optimization emerges when hybrid strategies—using quantum subroutines to accelerate classically intractable linear algebra or sampling steps—reduce the overall time or data required to reach comparable accuracy. Consequently, the field has shifted from asymptotic speedup proofs toward *hardware-aware, end-to-end performance evaluations*, emphasizing *resource efficiency*, *gradient precision*, and *training stability* on noisy intermediate-scale quantum (NISQ) platforms.

Hybrid Quantum–Classical Training Loop

The hybrid quantum–classical training paradigm represents a pragmatic bridge between the expressive power of quantum circuits and the numerical stability of classical optimization methods. In this framework, **the outer loop** consists of a classical optimizer—such as *Stochastic* Gradient Descent (SGD)*,* Adam*, or* Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS*)*—which updates the parameters of the quantum model based on feedback from measurements. **The inner loop** involves a **parameterized quantum circuit (PQC)**, a sequence of tunable gates that encodes both data and model parameters into a quantum state. This circuit executes on a quantum processor to evaluate an expectation value or cost function, typically corresponding to a measurement of an observable such as energy, loss, or fidelity. Gradients with respect to PQC parameters are estimated using the **parameter-shift rule**, which measures the difference between circuit outputs under shifted parameter values—an analogue to finite-difference methods that exploits the unitary nature of quantum operations.

In practical implementations, **batching and shot allocation** play critical roles in balancing accuracy and resource consumption. Each circuit execution produces statistical estimates from a finite number of *shots* (repeated measurements); adaptive shot allocation strategies can assign more samples to high-variance parameters, improving gradient reliability without inflating computational cost. The hybrid loop is inherently **stochastic**, as quantum measurement noise compounds with data stochasticity, making it essential to incorporate classical variance-reduction techniques such as moving averages or control variates. Additionally, **mini-batch training** of classical data allows scalable learning, where each batch triggers multiple quantum circuit evaluations that are averaged to form a stable update direction. To prevent overfitting and unnecessary quantum resource use, **early stopping** based on classical validation metrics—such as loss plateau detection, generalization gap monitoring, or Bayesian evidence thresholds—is commonly integrated into the workflow. This synergy between classical optimizers and quantum subroutines enables efficient exploration of the parameter space while mitigating the prohibitive sampling costs and noise sensitivity of near-term quantum hardware. The hybrid loop thus

constitutes the operational backbone of variational quantum algorithms (VQAs) and quantum neural networks (QNNs), serving as the principal mechanism through which machine learning benefits can be extracted from current quantum devices.

Quantum Feature Maps and Kernelization

Quantum feature maps provide a mathematically elegant and physically realizable means of embedding classical data into a high-dimensional Hilbert space, where linear separability and expressive correlations emerge naturally. In this process, **data-reuploading** and **unitary embeddings** serve as the core mechanisms that transform classical inputs into quantum states. Data-reuploading refers to repeatedly encoding classical data across multiple layers of a parameterized quantum circuit (PQC), effectively creating a hierarchical, nonlinear mapping analogous to deep neural architectures. Each layer applies data-dependent unitaries interleaved with trainable parameters, enabling the circuit to capture increasingly complex relationships. **Unitary embeddings**, on the other hand, map input vectors into normalized quantum states through rotation or phase operations, where the **fidelity between quantum states** acts as a measure of similarity—defining a natural **quantum kernel**. This kernel corresponds to the inner product of two quantum states and forms the basis for **Quantum Kernel Methods (QKMs)**, which can be leveraged within classical learning frameworks such as Support Vector Machines (SVMs) or Gaussian Processes.

A central distinction arises between **explicit PQC layers**—which are trainable, hybrid quantum–classical constructs—and **implicit quantum kernels**, such as those introduced by Havlíček et al. (2019), where the embedding is fixed but nontrivial. Explicit PQC layers allow end-to-end optimization, adapting the quantum feature space during training, whereas implicit kernels maintain analytical tractability and stability by avoiding parameter optimization within the quantum circuit. The Havlíček-style feature map, based on *exponentiated Pauli operators with data-dependent phases*, exemplifies a nonclassical embedding that generates feature spaces exponentially large in qubit number, making classical simulation infeasible. Consequently, **margin-based training**, which relies on inner products in feature space, can experience effective quantum acceleration when implemented via these kernels. Moreover, quantum kernels provide a natural **regularization effect** for high-dimensional data, mitigating overfitting by restricting representational capacity to physically valid subspaces defined by the unitary transformations. Empirical studies show that quantum kernels can outperform classical radial basis or polynomial kernels in regimes characterized by complex feature entanglement and low sample availability. Thus, quantum feature maps bridge the conceptual gap between variational quantum circuits and kernel-based learning, positioning quantum-enhanced kernelization as one of the most promising near-term pathways for achieving quantum advantage in machine learning.

Mitigating Barren Plateaus and Trainability Issues

One of the most significant challenges in optimizing parameterized quantum circuits (PQCs) and quantum neural networks (QNNs) is the emergence of **barren plateaus**—regions of exponentially vanishing gradients in the optimization landscape that render learning infeasible as system size grows. These plateaus arise primarily due to the concentration of measure in high-dimensional Hilbert spaces, where randomly initialized circuits produce nearly uniform expectation values. As a result, gradient magnitudes decay exponentially with the number of qubits or circuit depth, causing optimizers to stagnate early in training. To address this, researchers advocate the use of **problem-inspired, shallow, and locally entangled ansätze**, which restrict the circuit's

expressivity to physically meaningful subspaces. Such designs preserve relevant correlations while minimizing the depth-dependent accumulation of errors and variance.

Effective initialization strategies also play a crucial role. **Identity-preserving schemes**, where the initial unitary operations approximate the identity matrix, maintain stable gradient flow by preventing the immediate loss of correlation between parameters and observables. Similarly, **restricting entanglement growth** through controlled entangling gates or block-local architectures limits the exponential growth of the cost landscape's curvature, thereby improving gradient signal-to-noise ratios. **Layer-wise training**—incrementally optimizing circuit segments and freezing earlier layers—reduces parameter interference and stabilizes convergence, akin to layer-wise pretraining in deep neural networks. In practice, **warm starts** from pre-trained classical models or classical-to-quantum parameter transfer techniques can further expedite convergence by providing informed initialization close to useful minima. Moreover, the design of **noise-aware cost functions** that explicitly incorporate decoherence and shot noise ensures robustness on noisy intermediate-scale quantum (NISQ) devices. Collectively, these strategies transform the barren plateau from a fundamental limitation into a manageable engineering challenge, enabling more reliable and scalable optimization of hybrid quantum-classical models.

Second-Order and Hessian-Vector Ideas

While first-order optimization methods such as gradient descent and Adam dominate both classical and hybrid quantum–classical training due to their simplicity and scalability, **second-order methods** offer deeper insights into the local curvature of the loss landscape, enabling faster convergence and improved stability near minima. In the quantum domain, estimating curvature information requires specialized techniques that respect the probabilistic nature of measurement and the structure of parameterized quantum circuits (PQCs). One approach employs **stochastic curvature estimation using quantum Hadamard tests**, which allow efficient computation of *Hessian-vector products* without explicitly forming the Hessian matrix. These tests leverage controlled operations and interference patterns to encode second-order derivatives into measurable quantum amplitudes, effectively capturing curvature information with polynomial overhead relative to circuit depth. Such quantum-enhanced curvature sampling provides a promising pathway toward scalable quasi-Newton updates under the constraints of noisy intermediate-scale quantum (NISQ) devices.

Given the high sampling cost of full Hessian estimation, researchers often resort to **low-rank Hessian sketches**—approximations based on random projections or subspace embeddings that preserve dominant curvature directions. These low-dimensional representations maintain sufficient geometric fidelity for preconditioning updates, allowing accelerated convergence near basins of attraction where the loss landscape is locally convex. A practical hybrid strategy involves **switching to quasi-Newton methods**—such as the BFGS or L-BFGS variants—once gradient norms fall below a specified threshold, signaling proximity to a local optimum. In this regime, curvature information enhances convergence precision while keeping the quantum measurement budget under control.

The theoretical foundation of these second-order approaches is closely tied to the **Quantum Fisher Information (QFI)** matrix, which characterizes the Riemannian geometry of quantum states in parameter space. The QFI serves as a natural generalization of the classical Fisher information matrix, defining the intrinsic curvature of the quantum manifold. Leveraging QFI-guided optimization aligns parameter updates with geodesics on this manifold, minimizing redundant movement and improving trainability. Importantly, approximating the QFI via low-rank or block-

diagonal estimators enables efficient curvature-aware learning without full tomography. Together, these **Hessian-vector and Fisher-based techniques** pave the way toward quantum-aware second-order optimization frameworks that combine geometric insight, statistical efficiency, and hardware realism—bridging the gap between classical numerical methods and quantum information geometry.

Resource-Aware Compilation and Error Mitigation

The practical realization of quantum neural network (QNN) optimization on near-term quantum devices necessitates meticulous **resource-aware compilation** and **error mitigation** strategies to counteract the inherent limitations of noisy intermediate-scale quantum (NISQ) hardware. Given the restricted coherence times, limited qubit connectivity, and gate infidelity of current processors, **depth and breadth budgeting** becomes essential for balancing algorithmic expressivity with hardware feasibility. Circuit *depth* determines the temporal duration of quantum operations, directly influencing decoherence-induced errors, while *breadth* (number of qubits) governs the scale of entanglement and the associated readout complexity. Designing shallow, locality-preserving parameterized quantum circuits (PQCs) ensures computational tractability without overwhelming hardware resources.

Transpilation to native gate sets is another critical step, wherein high-level quantum instructions are decomposed into the specific one- and two-qubit gates supported by the target architecture. Optimized transpilation minimizes gate count and qubit swaps, preserving circuit fidelity by exploiting device-specific connectivity graphs and calibration data. Such compilation techniques often leverage **commutation-aware optimizers** and **layout mapping heuristics**, which reduce unnecessary entangling operations and enhance execution reliability.

To further suppress the cumulative effects of noise, **readout error mitigation** and **zero-noise extrapolation (ZNE)** are routinely employed. Readout error mitigation involves post-processing measurement outcomes using calibration matrices derived from known basis states, effectively correcting for systematic biases in qubit measurement channels. ZNE, in contrast, deliberately amplifies gate noise—by stretching circuit durations or inserting redundant identity operations—and extrapolates results back to the zero-noise limit through polynomial or exponential fitting. When combined, these methods significantly enhance the reliability of gradient estimation and cost-function evaluation in hybrid training loops.

An additional layer of optimization arises from **adaptive shot allocation**, where the number of circuit executions (shots) per parameter or observable is dynamically tuned based on **gradient-magnitude heuristics**. Parameters with large estimated gradients or higher variance receive more shots, ensuring precise estimation, while stable or low-variance parameters are updated with fewer measurements. This adaptive policy minimizes overall quantum sampling costs while maintaining statistical efficiency. The integration of such resource-aware compilation and mitigation pipelines transforms noisy, imperfect quantum processors into effective learning accelerators capable of supporting meaningful variational optimization, thereby pushing the frontier of practical quantum machine learning toward scalable, error-resilient architectures.

Applications and Benchmarks

The practical evaluation of quantum optimization techniques for neural networks demands rigorous benchmarking across diverse application domains, each highlighting distinct aspects of quantum representational power and optimization efficiency. Notably, **structured vision features**, **quantum chemistry descriptors**, **anomaly detection with Born machines**, and **sequence**

modeling through data-reuploading serve as canonical testbeds for hybrid quantum–classical systems. In computer vision, quantum circuits have been explored for encoding structured image patches into Hilbert-space embeddings that preserve spatial correlations. Here, parameterized quantum layers act as nonlinear feature extractors, transforming pixel or frequency representations into high-dimensional quantum kernels. These embeddings, when combined with classical convolutional backbones, can enhance robustness to noise and improve sample efficiency in small-data regimes—an area where classical deep learning models often overfit.

In **quantum chemistry**, the synergy between variational quantum eigensolvers (VQEs) and quantum neural architectures has proven particularly effective. Molecular Hamiltonians can be decomposed into expectation values over parameterized circuits, allowing the simultaneous learning of **quantum chemistry descriptors** and **energy functionals** within the same optimization framework. This hybridization accelerates convergence to ground-state energies and provides physically interpretable representations that mirror the underlying electronic structure, bridging the gap between machine learning models and quantum physics.

Another emerging application lies in **anomaly detection using Born machines**, where generative PQCs learn probability distributions implicitly through measurement outcomes. Such models can identify statistical deviations or rare patterns in high-dimensional data streams—applicable to fields ranging from cybersecurity to medical diagnostics—while maintaining robustness against data sparsity. Similarly, **sequence modeling through data-reuploading** leverages time-dependent parameter modulation in quantum circuits to process sequential inputs. This approach parallels recurrent or attention-based architectures in classical deep learning, yet benefits from quantum parallelism that allows simultaneous encoding of multiple temporal correlations.

Robust benchmarking of these applications requires **ablation studies** that systematically compare quantum–classical hybrids to **classical-only baselines** and **quantum kernel analogues**, ensuring methodological fairness. Evaluation metrics typically include convergence rate, generalization performance, noise tolerance, and computational efficiency under equivalent resource budgets. Fair benchmarking also necessitates normalization for hardware costs and sampling rates, recognizing that a quantum device's wall-clock latency differs from classical FLOP-based computation. Results across recent studies reveal that hybrid models outperform classical baselines in specific structured-data domains—especially those characterized by entangled feature dependencies or low effective dimensionality—while maintaining comparable computational footprints. As the ecosystem matures, the establishment of standardized benchmark suites, reproducible protocols, and open-access datasets will be essential for quantifying genuine quantum advantage in neural optimization and for guiding future algorithm–hardware co-design.

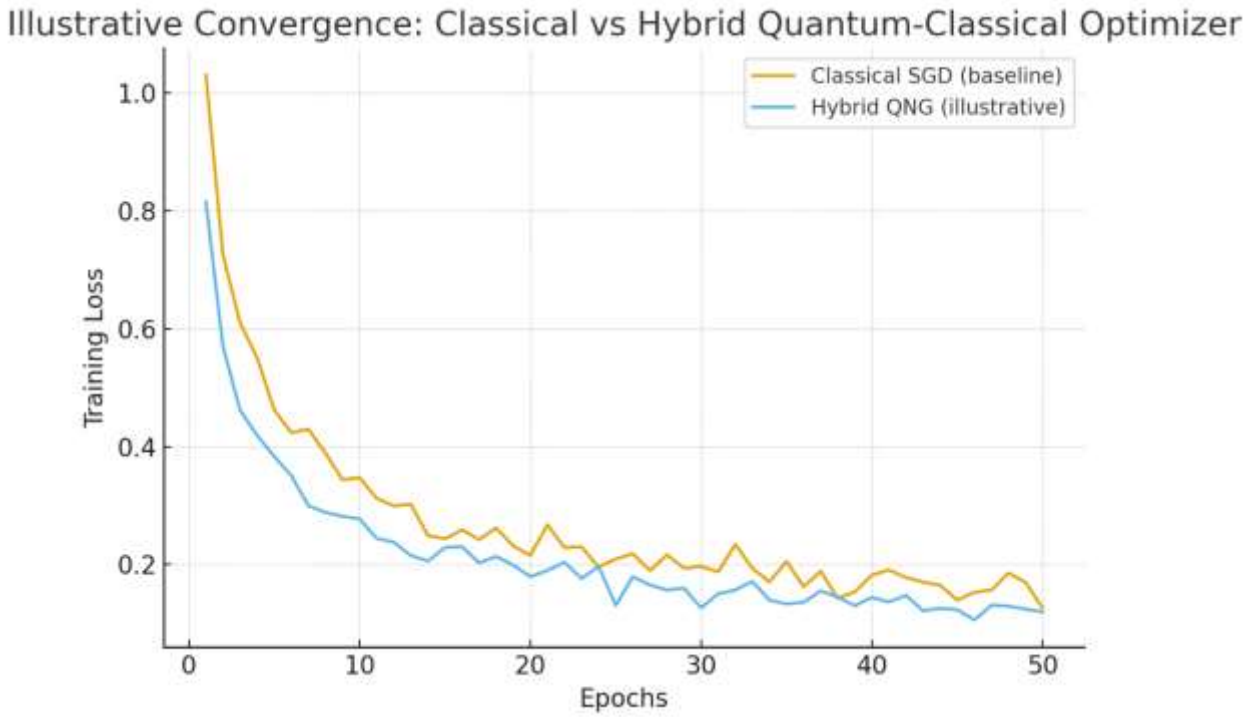
Roadmap to Fault-Tolerant Era

The evolution from noisy intermediate-scale quantum (NISQ) devices to **fault-tolerant quantum computing (FTQC)** represents a paradigm shift for neural network optimization, promising the realization of exponential speedups that are currently theoretical. In the **fault-tolerant era**, quantum systems will be capable of sustaining long, coherent computations through **quantum error correction (QEC)** and logical qubit encoding, thereby enabling stable and large-scale learning architectures. Central to this transition are **quantum linear solvers**, **quantum natural gradient (QNG)** extensions to large models, and **Grover-accelerated search algorithms** for architecture and hyperparameter optimization—each poised to redefine the scalability and precision of optimization in hybrid and fully quantum neural networks.

Quantum linear solvers, such as the Harrow–Hassidim–Lloyd (HHL) algorithm and its variants, promise exponential speedups for solving linear systems—a foundational subroutine in gradient computation, backpropagation, and kernel inversion. When deployed within deep learning pipelines, these solvers could transform optimization by enabling real-time curvature correction, covariance estimation, and energy minimization with logarithmic dependence on system size. In parallel, **large-scale QNG methods** will exploit full quantum Fisher information (QFI) matrices, computed directly within the quantum manifold, to guide parameter updates along geodesics of optimal efficiency. Unlike NISQ approximations that rely on block-diagonal or sampled QFI structures, fault-tolerant hardware will allow complete, noise-free QFI evaluation, unlocking geometry-aware learning at unprecedented precision.

Equally transformative is the integration of **Grover-accelerated search** within the architecture and hyperparameter optimization process. Grover's algorithm offers quadratic speedups for unstructured search tasks, which can be adapted to efficiently navigate the combinatorial design space of network topologies, ansatz configurations, and learning-rate schedules. This capability positions quantum systems as natural candidates for *meta-learning* and *automated model discovery*, vastly reducing the computational burden of large-scale architecture exploration.

The **transition markers** signaling the onset of fault-tolerant capability include surpassing **10^4 logical qubits**, achieving **surface-code thresholds** with physical error rates below 10^{-3}, and sustaining **stable, error-corrected circuit depths** capable of supporting iterative optimization cycles without decoherence. Once these milestones are achieved, the boundaries between classical and quantum learning architectures will blur, leading to integrated **quantum–classical co-processors** capable of handling both data-driven and physics-informed learning tasks. Beyond computational acceleration, fault-tolerant quantum systems will facilitate reproducibility, precision benchmarking, and algorithmic transparency—addressing key concerns of the NISQ era. In essence, the roadmap to the fault-tolerant epoch envisions a unified quantum machine learning ecosystem where **robust error correction**, **geometric optimization**, and **quantum-enhanced search** collectively redefine the frontiers of intelligence and computation.



Summary

Quantum techniques enrich neural optimization via expressive feature embeddings, geometry-aware preconditioning (QNG), and sampling/linear-algebra primitives that can reduce iteration counts or improve generalization—particularly on structured problems with limited data. In NISQ settings, benefits are contingent on careful ansatz design, variance-controlled gradient estimation, and error-aware compilation. Hybrid loops that keep the quantum part shallow while leveraging mature classical optimizers present the most practical path today. Looking ahead, error-corrected devices promise scalable second-order methods and end-to-end quantum accelerations that directly target the bottlenecks of large-scale training.

References

Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge Univ. Press, 2010.

J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum* 2, 79 (2018).

J. Biamonte et al., "Quantum machine learning," *Nature* 549, 195–202 (2017).

M. Schuld and F. Petruccione, *Supervised Learning with Quantum Computers*, Springer, 2018.

V. Havlíček et al., "Supervised learning with quantum-enhanced feature spaces," *Nature* 567, 209–212 (2019).

E. Farhi, J. Goldstone, S. Gutmann, "A Quantum Approximate Optimization Algorithm," arXiv:1411.4028 (2014).

Peruzzo et al., “A variational eigenvalue solver on a photonic quantum processor,” *Nat. Commun.* 5, 4213 (2014).

J. R. McClean et al., “The theory of variational hybrid quantum-classical algorithms,” *New J. Phys.* 18, 023023 (2016).

M. Benedetti et al., “Parameterized quantum circuits as machine learning models,” *NPJ Quantum Information* 5, 45 (2019).

D. Amodei et al., “Deep learning scaling and optimization,” (overview chapters/notes; see also OpenAI Scaling Laws).

S. Lloyd, M. Mohseni, P. Rebentrost, “Quantum algorithms for supervised and unsupervised machine learning,” arXiv:1307.0411 (2013).

E. Stokes et al., “Quantum Natural Gradient,” *Quantum* 4, 269 (2020).

G. Verdon et al., “A universal training algorithm for quantum deep learning,” arXiv:1806.09729 (2018).

M. Cerezo et al., “Cost-function dependent barren plateaus,” *Nat. Commun.* 12, 1791 (2021).

Marrero, M. Cerezo, E. Coles, “Entanglement-induced barren plateaus,” *PRX Quantum* 2, 040316 (2021).

Y. Li and S. C. Benjamin, “Efficient variational quantum simulator incorporating active error minimization,” *PRX* 7, 021050 (2017).

K. Mitarai et al., “Quantum circuit learning,” *Phys. Rev. A* 98, 032309 (2018).

F. Arute et al., “Quantum supremacy using a programmable superconducting processor,” *Nature* 574, 505–510 (2019).

S. Johansson et al., “Evaluating quantum natural gradient descent for variational circuits,” *Machine Learning: Science and Technology* 3, 035008 (2022).