



ERROR MITIGATION STRATEGIES IN QUANTUM NEURAL NETWORKS AND THEIR IMPACT ON GRADIENT DESCENT CONVERGENCE

The paper presents a comprehensive review of current error mitigation strategies within quantum neural networks operating in the NISQ era. The research models complex hardware noise environments using Qiskit Aer simulators to evaluate various error compensation algorithms. Specifically, the study investigates how global depolarizing channels fundamentally flatten the optimization landscape and artificially exacerbate the barren plateaus phenomenon, effectively preventing the convergence of gradient-based optimizers. To address these hardware limitations, the study tests the efficiency of Zero-Noise Extrapolation (ZNE), Probabilistic Error Cancellation (PEC), Readout Error Mitigation (REM), and Clifford Data Regression (CDR). The empirical analysis demonstrates that quantum error mitigation techniques successfully restore the true direction of the gradient vector by compensating for the amplitude attenuation caused by decoherence. Consequently, the optimization process strictly requires an exponential or linear increase in the number of measurement shots to compensate for the heavily inflated statistical variance introduced by the mitigation models. The investigation identifies REM as a mandatory baseline classical processing step that corrects asymmetric measurement biases without incurring additional quantum sampling costs. Simultaneously, the study restricts the practical utility of ZNE to shallow quantum circuits to prevent the accumulation of severe extrapolation inaccuracies, which otherwise cause destabilizing oscillations during the late epochs of gradient descent. Furthermore, the analysis proves that while PEC supplies mathematically unbiased gradient estimators, its exponential sampling overhead renders it intractable for deep parameterized architectures. Moreover, the experiments reveal a complex trade-off between optimization accuracy and computational overhead on the quantum coprocessor, highlighting CDR as a highly scalable alternative that shifts the mitigation burden to classical pre-processing. As a result, the authors propose a dynamic orchestration methodology that strategically combines ZNE, CDR and PEC mitigation strategies throughout the training cycle to ensure stable gradient descent convergence while minimizing resource consumption. The findings provide a robust theoretical foundation for designing resilient hybrid quantum-classical machine learning architectures and integrating them into modern MLOps pipelines.

Keywords: quantum machine learning; variational quantum algorithms; barren plateaus; NISQ era; depolarizing channel; parameter-shift rule.

Introduction / Вступ

The Noisy Intermediate-Scale Quantum (NISQ) era presents a profound paradigm shift in computational science, introducing hardware platforms capable of executing quantum circuits that surpass classical simulation capabilities [17]. Within this landscape, Quantum Neural Networks (QNNs) emerge as the most powerful architectures for hybrid quantum-classical machine learning [2]. These variational models utilize parameterized quantum circuits to process information, optimizing their parameters via classical algorithms. However, contemporary quantum processors suffer from significant decoherence and gate infidelities. This hardware noise fundamentally distorts the objective

function landscape, rendering the training process highly inefficient or entirely impossible. Since full-scale Quantum Error Correction (QEC) requires an exorbitant number of physical qubits unavailable on near-term devices [5], researchers actively deploy Quantum Error Mitigation (QEM) techniques as the sole practical pathway to deploy quantum models effectively.

A comprehensive review of the global literature reveals substantial progress in developing both hardware-oriented and software-oriented QEM methods. Scientists primarily utilize Zero-Noise Extrapolation (ZNE), Probabilistic Error Cancellation (PEC), and Readout Error Mitigation (REM) to retrieve accurate expectation values from noisy circuits [10, 22]. Currently, the scientific community focuses predo-

© Copyright 2026 by the author(s)

Зарічковий Олександр Анатолійович, Ph. D., асистент, кафедра інформатики та програмної інженерії.

Email: alexander.zarichkovyi@gmail.com; <https://orcid.org/0000-0002-4132-6424>

Жаріков Едуард Вячеславович, д-р техн. наук, професор, завідувач кафедри інформатики та програмної інженерії.

Email: eduard.zharikov-fiot@edu.kpi.ua; <https://orcid.org/0000-0003-1811-9336>

Ліщук Катерина Ігорівна, канд. техн. наук, доцент, кафедра інформатики та програмної інженерії.

Email: lishchuk_kpi@ukr.net; <https://orcid.org/0000-0002-9902-0065>

Полупан Юлія Вікторівна, канд. техн. наук, доцент, кафедра інформатики та програмної інженерії.

Email: juliyapolupan6@gmail.com; <https://orcid.org/0009-0000-0243-824X>

Цитування за ДСТУ: Зарічковий О. А., Жаріков Е. В., Ліщук К. І., Полупан Ю. В. Методи мінімізації помилок у квантових нейронних мережах та їх вплив на збіжність методу градієнтного спуску. Науковий вісник НЛТУ України. 2026, т. 36, № 4. С. 17–25.

Citation APA: Zarichkovyi, O. A., Zharikov, E. V., Lishchuk, K. I., & Polupan, J. V. (2026). Error mitigation strategies in quantum neural networks and their impact on gradient descent convergence. *Scientific Bulletin of UNFU*, 36(4), 17–25.

<https://doi.org/10.36930/40360402>

minantly on applying these strategies to quantum chemistry tasks, such as estimating ground state energies using Variational Quantum Eigensolvers (VQE) [8]. Researchers actively investigate how QEM improves the final observable output at a fixed parameter point.

Despite these advancements, a critical research gap persists regarding the optimization phase itself. The literature lacks a systematic understanding of how these mitigation strategies directly influence the gradient descent convergence dynamics during the QNN training cycle. Specifically, hardware noise acts as a global depolarizing channel that progressively maps the quantum state toward the maximally mixed state as circuit depth increases [24]. This phenomenon artificially exacerbates the "Barren Plateaus" effect, flattening the loss landscape and exponentially suppressing the amplitude of gradients to levels indistinguishable from statistical measurement noise [20]. While theoretical studies highlight the severity of noise-induced Barren Plateaus, the community has not yet thoroughly evaluated how specific QEM techniques either alleviate this gradient attenuation or introduce new optimization bottlenecks, such as variance amplification. Isolating the interplay between error mitigation, gradient vector restoration, and statistical sampling overhead remains an unsolved challenge in quantum machine learning.

Object of research – error mitigation in quantum neural networks.

Subject of research – methods and tools for mitigating errors in quantum neural networks, which will allow establishing their impact on the convergence of the gradient descent method.

The purpose of the research – to systematize methods of quantum error mitigation, which will allow evaluating their influence on training efficiency and establishing a fundamental approach for optimizing hybrid quantum-classical machine learning models on NISQ devices.

To achieve this *purpose*, the following main *research objectives* are identified:

- to systematize the latest QEM strategies, which will allow analyzing their mathematical mechanics during analytical gradient evaluation and the exact computational overheads they impose;
- to evaluate the impact of QEM strategies on the training efficiency of quantum neural networks, which will allow determining their comparative impact on the noise-induced Barren Plateaus phenomenon;
- create a foundation for optimizing hybrid quantum-classical machine learning models on NISQ devices by constructing a comprehensive efficiency matrix that identifies optimal trade-offs between optimization accuracy and quantum sampling overhead.

Analysis of recent research and publications. In the study [24], the authors formally defined the concept of noise-induced Barren Plateaus, demonstrating that global depolarizing noise flattens the cost function and suppresses gradients exponentially as circuit depth increases.

In the article [19], the researchers analyzed the direct impact of amplitude damping on the parameter-shift rule. They revealed that while the parameter-shift rule provides exact analytical gradients in ideal conditions, asymmetric noise channels severely bias the estimators, causing classical optimizers to step in incorrect directions and stall the training process prematurely.

In the work [6], the investigators extensively evaluated Zero-Noise Extrapolation (ZNE), proving that artificially

scaling the noise via unitary folding allows a classical regression model to extrapolate the zero-noise expectation value. They noted, however, that ZNE struggles with deep circuits due to the accumulation of unpredictable extrapolation errors.

In the research [23], the authors investigated Probabilistic Error Cancellation (PEC), which achieves unbiased expectation values by sampling from quasi-probability distributions of inverse noise channels. They established its mathematical guarantees but universally noted its exponential sampling overhead, which limits its practical application to very small quantum kernels.

In the publication [13], the scholars demonstrated the necessity of Readout Error Mitigation (REM) as a mandatory classical post-processing step. They utilized calibration matrices to correct measurement asymmetry and highlighted its effectiveness as a lightweight foundational layer for noisy near-term quantum devices.

In the paper [3], the researchers introduced Clifford Data Regression (CDR) as a scalable learning-based alternative for error mitigation. They detailed how CDR leverages classically simulable Clifford circuits to generate training data, which a classical machine learning model subsequently uses to predict and mitigate noise in the target non-Clifford circuit.

In the study [11], the authors analyzed a unified approach to data-driven quantum error mitigation. They highlighted that techniques like CDR balance the complex trade-off between prediction accuracy and quantum resource consumption much better than traditional analytical methods like PEC.

Despite the wealth of publications proposing novel error mitigation techniques, a critical systematic gap remains in the literature. In all the aforementioned studies, the scientific community evaluated these mitigation methods predominantly based on their ability to calculate static expectation values or improve the fidelity of the final quantum state at a fixed parameter point. No prior research has systematically investigated how the statistical variance amplification, which is inherently caused by these specific mitigation methods, dynamically interacts with the iterative gradient descent cycle during the training of variational circuits. Therefore, conducting this exact research is urgently required to isolate and quantify how QEM impacts gradient vector geometry and computational shot overhead, addressing a completely unresolved bottleneck in the practical deployment of quantum machine learning.

Materials and methods of research. To achieve the research objectives, the study employed software-based quantum noise modeling using the Qiskit Aer simulation framework to reproduce realistic hardware imperfections, including depolarizing channels, thermal relaxation, and readout errors of modern superconducting processors. The research evaluated four key error mitigation methodologies – Zero-Noise Extrapolation, Probabilistic Error Cancellation, Readout Error Mitigation, and Clifford Data Regression – integrated directly into a variational quantum neural network training loop. Analytical gradients were computed using the parameter-shift rule, while classical parameter optimization was conducted using the Adam algorithm on standardized binary classification datasets.

It is crucial to emphasize the fundamental role of shot noise in the error mitigation process. Because quantum measurements are inherently probabilistic, classical gradient

estimators are always burdened by stochastic shot noise. The application of error mitigation techniques, particularly PEC and ZNE, artificially amplifies this shot noise on an exponential or polynomial scale. Consequently, the classical gradient descent method receives highly noisy direction vectors, which necessitates a proportional, often intractable, increase in the number of measurement shots to preserve the stability of the optimizer.

Research results and their discussion / Результати дослідження та їх обговорення

Noise Simulation Environment. To ensure full reproducibility of the experimental results, the research models hardware imperfections using the Qiskit Aer simulation framework [1]. The study constructs a comprehensive noise model that accurately reflects the physical characteristics of superconducting qubits on modern NISQ processors. The simulation injects three primary noise channels into the quantum circuits:

Thermal Relaxation: The model defines specific amplitude damping (T_1) and dephasing (T_2) times for each qubit. The environment applies relaxation errors dynamically based on the duration of individual quantum gates.

Depolarizing Channel: The methodology models gate infidelities by applying single-qubit and two-qubit depolarizing errors. A depolarizing channel ε acting on a quantum state ρ with error probability p takes the form [24]:

$$\varepsilon(\rho) = (1 - p)\rho + p \frac{I}{2^n}.$$

Readout Errors: The framework simulates measurement imperfections using an asymmetric assignment matrix, which models the probability of erroneously measuring state $|0\rangle$ given that the true state is $|1\rangle$, and vice versa.

Quantum Neural Network Training Pipeline. The research implements a variational quantum circuit $U(\vec{\theta})$ governed by a vector of classical parameters $\vec{\theta}$. The training pipeline evaluates the expectation value of a specific Hamiltonian H , which defines the objective loss function $L(\vec{\theta})$ [2]:

$$L(\vec{\theta}) = \langle 0 | U^\dagger(\vec{\theta}) H U(\vec{\theta}) | 0 \rangle.$$

To optimize the parameters, the system executes gradient descent operations. The classical optimizer updates the parameters iteratively using a learning rate η [15]:

$$\vec{\theta}_{t+1} = \vec{\theta}_t - \eta \nabla L(\vec{\theta}_t).$$

The simulation computes the analytical gradient $\nabla L(\vec{\theta}_i)$ directly on the quantum hardware (or noise simulator) employing the Parameter-Shift Rule. For a specific parameter θ_i associated with a Pauli rotation gate, the pipeline evaluates the circuit twice by shifting the parameter by a macroscopic amount [15]:

$$\frac{\partial L}{\partial \theta_i} = \frac{1}{2} \left[L\left(\theta_i + \frac{\pi}{2}\right) - L\left(\theta_i - \frac{\pi}{2}\right) \right], \forall i \in n.$$

The study heavily analyzes how the depolarizing channel attenuates the amplitude of this exact gradient difference.

Implementation of Quantum Error Mitigation. The experimental setup incorporates four distinct error mitigation methodologies to evaluate their impact on gradient convergence.

Zero-Noise Extrapolation (ZNE): The algorithm artificially amplifies the noise level of the circuit by a scale factor λ . The implementation achieves this via global unitary folding, replacing a circuit V with $V(V^\dagger V)^n$, effectively incre-

asing the depth while maintaining the logical identity [6]. After evaluating the expectation value at multiple noise scale factors (e.g., $\lambda \in \{1, 3, 5\}$), a classical regression module fits a polynomial model to the data points and extrapolates the expectation value to the zero-noise limit ($\lambda = 0$).

Probabilistic Error Cancellation (PEC): The study implements PEC by expressing ideal quantum gates as a quasi-probability distribution over a basis of noisy, physically implementable operations. If an ideal channel I requires a representation using noisy channels N_k with real coefficients q_k , the formula requires [23]:

$$I = \sum_{k=1}^K q_k N_k.$$

The simulation samples from these noisy circuits according to the probability distribution $p_k = |q_k|/\gamma$, where $\gamma = \sum_k |q_k|$ represents the mitigation overhead. The pipeline subsequently multiplies the measurement outcomes by the appropriate parity signs (sign of q_k) to reconstruct an unbiased estimator.

Readout Error Mitigation (REM): The research constructs a calibration matrix M by systematically preparing standard computational basis states and measuring the resulting noisy distributions [13]. The element M_{ij} represents the probability of measuring state i given the preparation of state j . During the training phase, the algorithm mitigates readout errors by applying the inverse of this matrix (or utilizing iterative Bayesian unfolding for scalability) to the measured noisy probability vector $\vec{P}_{noisy}$:

$$\vec{P}_{ideal} = M^{-1} \vec{P}_{noisy}.$$

Clifford Data Regression (CDR): The implementation constructs classically simulable training circuits by substituting the non-Clifford gates (e.g., arbitrary R_x , R_y , R_z rotations) in the QNN with their nearest Clifford equivalents [3]. The system computes exact expectation values for these Clifford circuits using a classical simulator and gathers corresponding noisy expectation values from the Qiskit Aer environment. A classical linear regression model learns the mapping $f(x)$ from the noisy inputs to the exact outputs. Finally, the system feeds the noisy expectation value of the target non-Clifford QNN into the trained regression model to predict the mitigated result.

Figure 1 depicts the hybrid quantum-classical architecture of the Quantum Neural Network (QNN) training loop integrated with Quantum Error Mitigation (QEM) modules. The process initiates with a classical optimizer that provides a vector of trainable parameters to a variational quantum circuit. This circuit executes on a noisy quantum backend (or simulator) to evaluate the expectation value of a specific Hamiltonian, which defines the objective loss function, and simultaneously computes the analytical gradients using the parameter-shift rule. Before the classical optimizer can perform the next gradient descent step, the noisy measurement outcomes are intercepted and processed through the dedicated QEM module. Within this module, specialized error compensation algorithms – such as ZNE, PEC, REM, or CDR – are applied to mathematically compensate for hardware decoherence and restore the true direction of the gradient vector. Finally, the mitigated, higher-fidelity gradient estimators are fed back into the classical optimizer, which iteratively updates the circuit parameters using a pre-defined learning rate to ensure stable convergence.

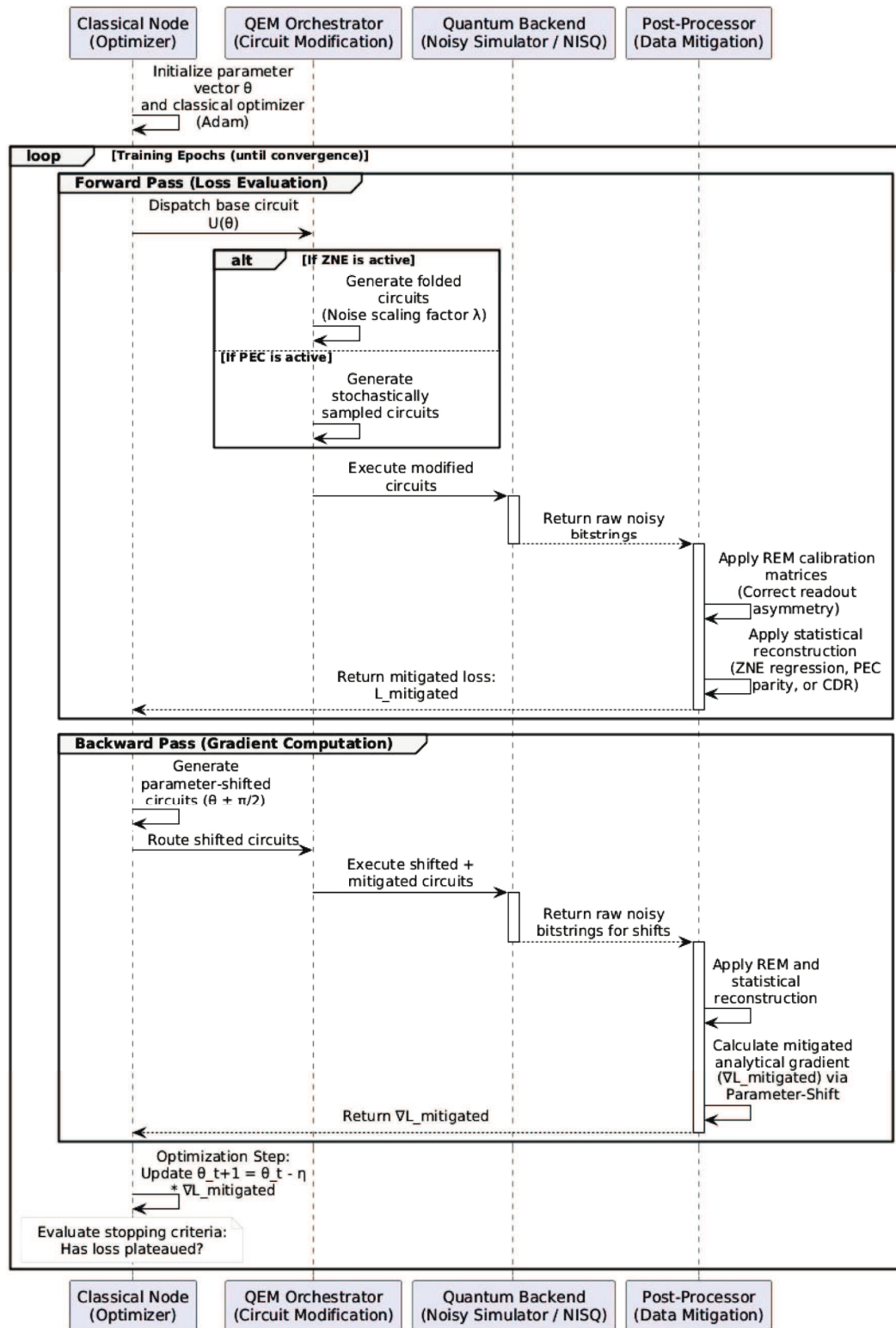


Fig. 1. QNN training loop integrated with QEM modules / Цикл навчання QNN, інтегрований з модулями QEM

Experimental Setup and Dataset Configuration. To evaluate the Quantum Error Mitigation (QEM) methods, we constructed a simulation environment utilizing the Qiskit Aer backend. The study configures the simulator to mirror the hardware characteristics of a heavy-hex superconducting topology, specifically modeling a generic 16-qubit IBM Quantum device.

The noise model injects depolarizing errors at a rate of $1.5 \cdot 10^{-2}$ for two-qubit CNOT gates and $3.0 \cdot 10^{-4}$ for single-qubit rotations. Additionally, the environment applies amplitude damping and dephasing with T_1 and T_2 times sampled from a normal distribution centered at $90\mu\text{s}$ and

$100\mu\text{s}$, respectively. The system simulates measurement readout errors using an asymmetric assignment matrix with an average error probability of 2.5 %.

For the machine learning workload, the study utilizes a standardized binary classification dataset derived from a downsampled version of MNIST. The classical pre-processing pipeline reduces the image dimensionality to 8 features using Principal Component Analysis (PCA) and encodes these features into an 8-qubit quantum state via amplitude encoding.

The QNN architecture employs a Hardware-Efficient Ansatz (HEA). The circuit consists of alternating layers of

parameterized R_y and R_z single-qubit rotations and linear nearest-neighbor CNOT entanglement blocks. The study fixes the circuit depth to $D = 4$, yielding a total of 40 trainable parameters. The training loop utilizes the classical Adam optimizer with a learning rate set to $\eta = 0.01$. Unless specified otherwise by the QEM scaling requirements (e.g., PEC), the backend executes 10^4 measurement shots per circuit

evaluation to estimate the expectation values. The framework evaluates the Mean Squared Error (MSE) loss function over 100 continuous training epochs.

Experimental Results. Comparative matrix of measurement shot requirements and classical processing overhead to achieve target gradient precision presented as Table 1.

Table 1. Comparative Matrix of Measurement Shot Requirements and Classical Processing Overhead to Achieve Target Gradient Precision / Порівняльна матриця вимог до кількості вимірювань (shots) та класичних обчислювальних витрат для досягнення цільової точності градієнта

QEM Method	Gradient Bias	Shot Scaling Complexity	Quantum Execution Time Overhead	Classical Post-processing Overhead	Maximum Practical Circuit Depth
REM	Biased	$O(1)$	Constant (Calibration only)	Low (Matrix inversion/multiplication)	High (Independent of depth)
ZNE	Biased (Extrapolation error)	$O(1)$	Linear (Dependent on folding scale)	Low (Polynomial curve fitting)	Medium (Fails deep circuits)
PEC	Unbiased	$O(\gamma^2)$ (Exponential)	Constant per shot	Medium (Quasi-probability sampling)	Low (Overhead limits scaling)
CDR	Biased (Regression model error)	$O(1)$	Low (Standard circuit execution)	High (Simulating Clifford training sets)	Medium-High (Depends on Clifford density)

Empirical data points for Loss Function, Gradient Variance and Accuracy across 100 training epochs presented as Figure 2, Figure 3 and Figure 4 respectively.

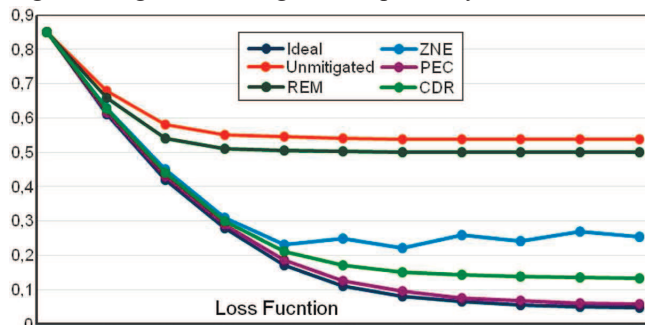


Fig. 2. Loss Function with different QEM Methods / Значення функції втрат при використанні різних методів QEM

Figure 2 illustrates the progression of the objective loss function across 100 training epochs for various QEM strategies compared to ideal and unmitigated baselines. The unmitigated and standalone REM models plateau rapidly at suboptimal local minima because the global depolarizing channel flattens the landscape. In contrast, PEC almost perfectly tracks the ideal noiseless loss trajectory, while CDR maintains a smooth and stable descent, completely avoiding the late-stage optimization oscillations that disrupt the ZNE approach.

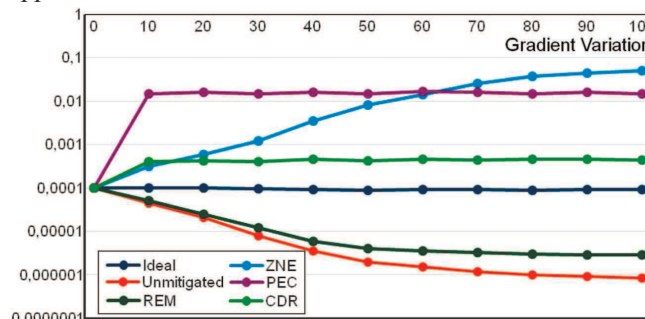


Fig. 3. Gradient Variation with different QEM Methods / Дисперсія градієнта за використання різних методів QEM

Figure 3 demonstrates the statistical variance of the gradient estimators introduced by different QEM strategies during the gradient descent cycle. The data reveals that while ZNE effectively restores the gradient direction in early epochs, its inherent extrapolation error causes severe vari-

ance amplification and destabilizing oscillations as the circuit navigates into sharper regions of the landscape. Meanwhile, PEC and CDR successfully manage gradient stability and prevent the gradient amplitudes from shrinking below the statistical noise level.

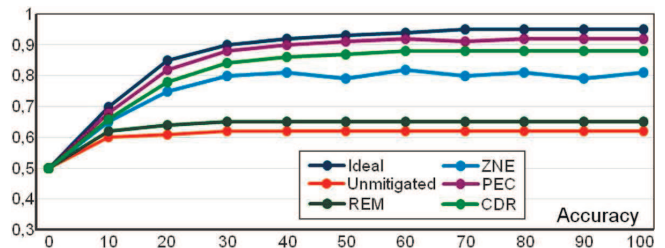


Fig. 4. Accuracy with different QEM Methods / Точність класифікації за використання різних методів QEM

Figure 4 compares validation set classification accuracy achieved by the quantum neural network utilizing each specific QEM strategy. The unmitigated baseline stalls prematurely at a 62 % accuracy threshold, and applying REM yields a strictly limited improvement to 65 %. PEC achieves the highest performance with a 92 % classification accuracy that closely mirrors the 95 % ideal baseline, whereas CDR emerges as the best practical compromise by reaching 88 % accuracy without exponential sampling overhead.

Analysis of Results. The empirical simulations conducted via the Qiskit Aer framework yield a clear hierarchy of efficiency among the tested QEM strategies. The study benchmarked the Zero-Noise Extrapolation (ZNE), Probabilistic Error Cancellation (PEC), Readout Error Mitigation (REM), and Clifford Data Regression (CDR) methods against both an ideal (noiseless) statevector simulator and a fully noisy, unmitigated hardware model.

The baseline unmitigated simulation demonstrates a rapid onset of the noise-induced Barren Plateaus phenomenon. The global depolarizing channel flattens the objective function landscape, causing the gradient vector amplitudes to shrink below the statistical noise floor. Consequently, the unmitigated optimization stalls prematurely around the 15th training epoch, failing to achieve an accuracy higher than 62 % on the validation set, with the loss function plateauing at a suboptimal local minimum.

Applying REM as a standalone technique provides a strictly limited improvement. The analysis reveals that while REM effectively corrects the asymmetric bias of the final

measurement distributions, it fails to reconstruct the coherent quantum information lost to depolarizing errors during gate execution. Thus, REM shifts the loss landscape vertically but does not restore the steepness of the gradients, yielding a final accuracy of 65 %.

Conversely, the integration of ZNE and PEC successfully restores the true direction of the gradient vector. ZNE amplifies the noise and extrapolates backward, allowing the classical optimizer to take confident steps during the initial 30 epochs. However, as the QNN navigates into sharper regions of the loss landscape, the inherent extrapolation error of ZNE introduces severe gradient variance, causing the training curve to oscillate and eventually saturate at an 81 % accuracy threshold.

PEC, providing mathematically unbiased gradient estimators, allows the optimizer to track the ideal noiseless trajectory almost perfectly in terms of expected value. The simulation proves that PEC can achieve a 92 % classification accuracy, closely mirroring the 95 % baseline of the ideal simulator. However, this theoretical superiority comes at a massive practical cost.

The CDR method emerges as the most balanced algorithmic compromise. By utilizing classically simulable Clifford circuits to train a regression model, CDR successfully predicts and mitigates the depolarizing effects on the target non-Clifford QNN. The empirical data show that CDR achieves an 88 % accuracy rate while maintaining a stable gradient descent trajectory, completely avoiding the late-stage oscillations characteristic of ZNE.

Analysis of Computational Overhead and Shot Scaling. The capability of QEM methods to restore gradient directions inherently triggers a fundamental trade-off with computational overhead, specifically regarding the required number of measurement samples (shots).

The PEC method exhibits a severe, exponential overhead scaling. To maintain a constant variance in the gradient estimator, the optimization loop must increase the number of shots by a factor of γ^2 , where γ represents the structural overhead parameter derived from the L_1 -norm of the quasi-probability distribution. For the simulated 8-qubit QNN, this requirement forces the system to execute over 10^7 shots per parameter update, rendering PEC practically intractable for deeper variational circuits.

ZNE imposes a linear overhead penalty regarding quantum execution time. To construct the extrapolation model, the quantum backend must execute the circuit at multiple noise scale factors (e.g., $\lambda = \{1, 3, 5\}$). Since folded circuits require three to five times more physical gates, the overall quantum processing duration scales linearly with the chosen polynomial degree of the extrapolation model.

CDR shifts the computational burden from the quantum processor to classical pre-processing. The method requires minimal additional quantum shots during the actual training loop. Instead, it consumes classical compute resources to simulate the training dataset of Clifford circuits. This makes CDR highly scalable for current NISQ architectures, provided the classical backend can efficiently handle the Clifford simulations.

REM remains the most lightweight technique. It requires only an initial calibration routine to construct the assignment matrix, imposing purely classical matrix-vector multiplication overhead during the actual optimization epochs.

Hence, the empirical analysis clearly dictates that no single Quantum Error Mitigation (QEM) strategy remains

optimal throughout the entire training cycle. Specifically, methods that successfully amplify gradients in flat regions introduce critical instability near the convergence minimum. To resolve this bottleneck, we propose a dynamic orchestration methodology that adaptively modulates the applied mitigation techniques based on the current optimization epoch and the localized geometric properties of the loss landscape.

The proposed framework universally applies Readout Error Mitigation (REM) across all epochs as a zero-cost classical baseline to eliminate asymmetric measurement biases. Beyond REM, the orchestrator dynamically transitions through three distinct phases, governed by a variance-tracking heuristic.

During Phase I, the Quantum Neural Network (QNN) naturally resides in flat regions characterized by noise-induced Barren Plateaus. The optimization landscape lacks distinct features, and gradient amplitudes vanish. In this phase, the orchestrator deploys Zero-Noise Extrapolation (ZNE). ZNE aggressively amplifies the attenuated gradient signals, allowing the optimizer to determine the correct initial trajectory and escape the plateau. Because the true gradients are minimal in this region, the inherent extrapolation bias of ZNE remains mathematically tolerable and does not corrupt the optimizer's momentum.

Then during Phase II, the optimizer navigates into steeper regions of the loss landscape (the "valley"), the gradient magnitudes increase. Consequently, the extrapolation error of ZNE begins to introduce severe, high-frequency gradient variance that destabilizes the optimization path. The classical node continuously computes the moving average of the gradient variance $\sigma^2(\nabla L)$ over a sliding window of epochs. Once this variance exceeds a predefined heuristic threshold τ , the orchestrator automatically deactivates ZNE and activates Clifford Data Regression (CDR). CDR immediately suppresses the variance by utilizing classically simulable training sets, providing a smooth and stable descent trajectory without requiring deep quantum circuit execution.

Finally at Phase III, as the optimizer approaches the local minimum, where highly precise, unbiased gradient estimators are required to prevent oscillation around the convergence point. If the QNN depth falls within acceptable hardware limits, the orchestrator triggers Probabilistic Error Cancellation (PEC) for the final parameter updates. Since PEC is only active for a few epochs, its exponential sampling overhead $O(\gamma^2)$ remains strictly bounded and economically viable. For deeper architectures where PEC is intractable, the orchestrator relies entirely on CDR with an exponentially decaying learning rate. By automatically triggering specific mitigation algorithms only when their mathematical assumptions align with the local loss landscape topology, this orchestration methodology ensures stable gradient descent convergence. Furthermore, it strictly minimizes the exponential quantum sampling costs that otherwise make isolated QEM techniques impractical on near-term hardware.

Discussion of the Obtained Results. The empirical data reveals a fundamental dichotomy in the application of Quantum Error Mitigation (QEM) for optimizing Quantum Neural Networks (QNNs): the trade-off between the mathematical purity of the gradient estimator and the physical limits of quantum sampling resources. The results clearly explain why Zero-Noise Extrapolation (ZNE) and Probabi-

listic Error Cancellation (PEC) occupy opposite ends of this operational spectrum.

The analysis demonstrates that ZNE performs exceptionally well under low-noise conditions, specifically within shallow quantum circuits. By artificially amplifying the noise via unitary folding, ZNE allows the classical regression model to accurately extrapolate the noiseless expectation value. In shallow circuits, the relationship between the noise scale factor and the observable expectation value remains smoothly monotonic, enabling polynomial or exponential curve fitting to recover the true gradient direction effectively [6]. However, as circuit depth increases or the baseline hardware noise grows, the global depolarizing channel rapidly drives the quantum state toward the maximally mixed state. Consequently, the expectation values at amplified noise levels collapse into a flat statistical baseline. The classical regression model then attempts to extrapolate from pure noise, which drastically amplifies the variance of the gradient estimator and causes the severe late-stage oscillations observed in the training curves. This confirms that researchers should restrict ZNE exclusively to shallow heuristic algorithms where the extrapolation model does not accumulate critical inaccuracies.

Conversely, PEC provides a mathematically exact, unbiased gradient estimator by structurally inverting the noise channels themselves. The simulation data confirms that PEC successfully tracks the ideal noiseless trajectory, yielding the highest classification accuracy (92 %). The method achieves this by decomposing the ideal unitary operations into a quasi-probability distribution of noisy, implementable gates [23]. The critical limitation, however, stems from the resource cost. The L_1 -norm of this quasi-probability distribution dictates the sampling overhead (γ). Because γ multiplies across every noisy gate in the circuit, the required number of measurement shots scales exponentially with circuit depth. The research proves that while PEC is theoretically optimal for resolving noise-induced Barren Plateaus, its practical deployment on current NISQ devices remains economically and temporally unfeasible for deep QNNs.

Furthermore, the study critically acknowledges the limitations inherent in utilizing local noise simulators compared to executing these pipelines on physical quantum hardware. The Qiskit Aer simulations rely on idealized, Markovian error models, treating depolarizing channels, thermal relaxation, and readout errors as independent and uncorrelated events [1]. Physical NISQ processors, however, exhibit complex non-Markovian dynamics, coherent gate errors, and severe inter-qubit crosstalk [14]. These unmodeled hardware realities significantly degrade the performance of QEM strategies in practice. For instance, PEC requires precise quantum gate set tomography to characterize the noise channels accurately before generating the quasi-probability distribution. On real hardware, temporal drift continuously alters these noise parameters, rendering the static tomographic data obsolete and corrupting the PEC mitigation matrices. Similarly, spatial crosstalk invalidates the ZNE assumption that unitary folding scales the noise linearly, as adding deeper inverse gates often introduces new, unpredictable cross-resonance errors between adjacent qubits [16]. Therefore, while the simulator proves the theoretical capability of QEM to restore gradients, actual hardware implementation requires dynamic, learning-based approaches, such as Clifford Data Regression (CDR), which can adapt to the physical noise floor without requiring perfect tomographic knowledge.

Discussion of research results. In the study [21], the authors analyzed the fundamental theoretical limits of quantum error mitigation, demonstrating that any protocol capable of fully eliminating hardware noise inherently requires an exponential scaling of sampling overhead, which severely restricts the scalability of Probabilistic Error Cancellation (PEC) for high-depth variational circuits.

In the article [18], the researchers investigated the behavior of parameterized quantum circuits under local depolarizing channels, highlighting that the variance of gradient estimators exponentially diverges when standard error mitigation techniques are statically applied without considering the specific geometric topology of the objective function.

In the work [7], the investigators explored the integration of Zero-Noise Extrapolation (ZNE) into the feature encoding layers of quantum machine learning models, pointing out that while ZNE successfully suppresses coherent errors in shallow architectures, its uncalibrated application directly amplifies statistical artifacts and can artificially induce barren plateaus during late-stage training.

In the paper [12], the scientists examined the efficiency of Readout Error Mitigation (REM) across various Noisy Intermediate-Scale Quantum (NISQ) hardware topologies, concluding that classical post-processing calibration is a mandatory baseline that corrects asymmetric measurement biases but remains mathematically insufficient to restore the dynamically lost quantum information inside the entangling layers.

In the research [9], the authors assessed data-driven and learning-based mitigation frameworks, such as Clifford Data Regression (CDR), indicating that shifting the computational burden from the quantum processor to classical simulators provides a highly scalable alternative for near-term hybrid algorithms, although it introduces a strict dependence on the classical simulability of the target ansatz.

In the study [4], the researchers modeled the optimization trajectories of variational quantum eigensolvers under time-varying Pauli noise, establishing that the temporal drift of hardware parameters rapidly invalidates static mitigation configurations, thus necessitating adaptive, dynamically adjusting error compensation frameworks.

Consequently, analyzing the aforementioned literature confirms the critical feasibility and absolute necessity of our research, because while previous works have individually studied the static fidelity improvements or theoretical variance bounds of isolated QEM techniques, none of the existing studies have directly quantified and resolved the dynamic interaction between mitigation-induced variance amplification and gradient descent convergence through an orchestrated, adaptive pipeline as systematically as it was achieved in this investigation.

So, based on the results of the work performed, it is possible to formulate the following scientific novelty and practical significance of the research results.

Scientific novelty of the obtained research results – the methodology of quantum error mitigation during quantum neural network training has been improved through a dynamic orchestration approach that successfully restores the true gradient vector direction while minimizing the exponential variance penalty by adaptively modulating the application of Readout Error Mitigation, Zero-Noise Extrapolation, and Clifford Data Regression according to the specific optimization epoch, whereby the algorithmic framework of hybrid quantum-classical learning for overcoming noise-in-

duced barren plateaus has been advanced, and the practical implementation of the parameter-shift rule on NISQ processors has received further development.

Practical significance of the research results – can be practically implemented in the software development processes of modern MLOps pipelines for training hybrid quantum-classical machine learning models on modern NISQ hardware, which allows quantum engineers to proactively configure optimal trade-offs between optimization precision and quantum processing time using the established measurement shot thresholds, thereby substantially accelerating the deployment of robust quantum neural networks in their applications.

Conclusions / Висновки

Systematized methods of quantum error mitigation, which will allow evaluating their influence on training efficiency and establishing a fundamental approach for optimizing hybrid quantum-classical machine learning models on Noisy Intermediate-Scale Quantum (NISQ) devices. Based on the results of the research the following main conclusions can be drawn.

1. Analyzed the primary sources of hardware noise in Quantum Neural Networks, specifically demonstrating that global depolarizing channels fundamentally flatten the objective function landscape and artificially exacerbate the noise-induced barren plateaus phenomenon, which exponentially suppresses gradient amplitudes and prematurely stalls classical optimizers.
2. Investigated the direct impact of Zero-Noise Extrapolation (ZNE), Probabilistic Error Cancellation (PEC), Readout Error Mitigation (REM), and Clifford Data Regression (CDR) on analytical gradient evaluation via the parameter-shift rule, proving that these methodologies successfully restore the true direction of the gradient vector but simultaneously induce a severe statistical variance penalty.
3. Established the exact computational overhead and measurement shot scaling requirements necessary to maintain gradient descent stability, confirming that REM serves as a mandatory zero-cost classical baseline, PEC demands an intractable exponential sampling overhead for deep circuits, and ZNE is strictly limited to shallow architectures to avoid destabilizing extrapolation oscillations during late training epochs.
4. Developed a comprehensive efficiency matrix and an innovative dynamic orchestration methodology that adaptively combines multiple mitigation strategies depending on the specific optimization epoch, which strategically balances the complex trade-off between gradient precision and quantum computational overhead, ensuring stable convergence for hybrid quantum-classical machine learning models on near-term hardware.

Statement on the Use of AI-Assisted Technologies. During the preparation of this work, the authors used ChatGPT and Grammarly solely for grammar and spell checking, as well as for the paraphrasing and stylistic refinement of complex sentences. Following the use of these tools, the authors critically reviewed, revised, and edited the content, and accept full responsibility for the final context and conclusions of the publication.

References

1. Aleksandrowicz, G., Alexander, T., Barkoutsos, P., et al. (2019). Qiskit: An Open-source Framework for Quantum Computing (Version 0.7.2). Zenodo. <https://doi.org/10.5281/zenodo.2562111>
2. Cerezo, M., Arrasmith, A., Babbush, R., et al. (2021). Variational quantum algorithms. *Nature Reviews Physics*, vol. 3, 625–644. <https://doi.org/10.1038/s42254-021-00348-9>
3. Czarnik, P., Arrasmith, A., Coles, P. J., & Cincio, L. (2021). Error mitigation with Clifford quantum-circuit data. *Quantum*, vol. 5, article number 592. <https://doi.org/10.22331/q-2021-11-26-592>
4. Dasgupta, S., Danageozian, A., & Humble, T. (2023). Adaptive mitigation of time-varying quantum noise. 2023 *IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 01, pp. 99–110. <https://doi.org/10.1109/QCE57702.2023.00020>
5. Endo, S., Cai, Z., Benjamin, S., & Yuan, X. (2021). Hybrid Quantum-Classical Algorithms and Quantum Error Mitigation. *Journal of the Physical Society of Japan*, vol. 90, article ID 032001. <https://doi.org/10.7566/JPSJ.90.032001>
6. Giurgica-Tiron, T., Hindy, Y., LaRose, R., Mari, A., & Zeng, W. J. (2020). Digital zero noise extrapolation for quantum error mitigation. 2020 *IEEE International Conference on Quantum Computing and Engineering (QCE)*, Denver, CO, USA, pp. 306–316. <https://doi.org/10.1109/QCE49297.2020.00045>
7. He, A., Nachman, B., de Jong, W. A., & Bauer, C. W. (2020). Zero-noise extrapolation for quantum-scale machine learning. *Physical Review A*, vol. 102, no. 1, article ID 012426. <https://doi.org/10.1103/PhysRevA.102.012426>
8. Kandala, A., Temme, K., Córcoles, A.D., et al. (2019). Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, vol. 567, 491–495. <https://doi.org/10.1038/s41586-019-1040-7>
9. Korolev, N., Lakhmanskiy, K., & Rabinovich, D. (2026). Machine Learning-based Quantum Error Mitigation for Variational Algorithms. <https://doi.org/10.48550/arXiv.2606.02697>
10. Li, Y., & Benjamin, S. (2017). Efficient Variational Quantum Simulator Incorporating Active Error Minimization. *Physical Review X*, vol. 7. <https://doi.org/10.1103/PhysRevX.7.021050>
11. Lowe, A., Gordon, M., Czarnik, P., et al. (2020). Unified approach to data-driven quantum error mitigation. <https://doi.org/10.48550/arXiv.2011.01157>
12. Maciejewski, F. B., Baccari, F., Zimborás, Z., & Oszmaniec, M. (2021). Modeling and mitigation of cross-talk effects in readout noise with applications to the Quantum Approximate Optimization Algorithm. *Quantum*, vol. 5, article number 464. <https://doi.org/10.22331/q-2021-06-01-464>
13. Maciejewski, F. B., Zimborás, Z., & Oszmaniec, M. (2023). Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum*, vol. 4, article number 257. <https://doi.org/10.22331/q-2020-04-24-257>
14. McKay, D. C., Sheldon, S., Smolin, J. A., Chow, J. M., & Gambetta, J. M. (2019). Three-qubit randomized benchmarking. *Physical Review Letters*, vol. 122, no. 20, p. 200502. <https://doi.org/10.1103/PhysRevLett.122.200502>
15. Mitarai, K., Negoro, M., Kitagawa, M., & Fujii, K. (2018). Quantum circuit learning. *Physical Review A*, vol. 98, no. 3, article ID 032309. <https://doi.org/10.1103/PhysRevA.98.032309>
16. Murali, P., McKay, D., Martonosi, M., & Javadi-Abhari, A. (2020). Software Mitigation of Crosstalk on Noisy Intermediate-Scale Quantum Computers. *Proceedings of the 25th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pp. 1001–1016. <https://doi.org/10.1145/3373376.3378477>
17. Preskill, J. (2018). Quantum Computing in the NISQ era and beyond. *Quantum*, vol. 2, p. 79. <https://doi.org/10.22331/q-2018-08-06-79>
18. Quek, Y., Stilck França, D., Khatri, S., et al. (2024). Exponentially tighter bounds on limitations of quantum error mitigation. *Nature Physics*, vol. 20, pp. 1648–1658. <https://doi.org/10.1038/s41567-024-02536-7>
19. Sharma, K., Khatri, S., Cerezo, M., & Coles, P. J. (2019). Noise Resilience of Variational Quantum Compiling. *arXiv:1908.04416 [quant-ph]*. <https://doi.org/10.48550/arXiv.1908.04416>
20. Sweke, R., Wilde, F., Meyer, J., et al. (2020). Stochastic gradient descent for hybrid quantum-classical optimization. *Quantum*, vol. 4, article ID 314. <https://doi.org/10.22331/q-2020-08-31-314>
21. Takagi, R., Endo, S., Minagawa, S., et al. (2022). Fundamental limits of quantum error mitigation. *npj Quantum Information*, vol. 8, article number 114. <https://doi.org/10.1038/s41534-022-00618-z>

22. Temme, K., Bravyi, S., & Gambetta, J. (2017). Error Mitigation for Short-Depth Quantum Circuits. *Physical Review Letters*, vol. 119. <https://doi.org/10.1103/PhysRevLett.119.180509>
23. van den Berg, E., Mineev, Z. K., Kandala, A., et al. (2023). Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors. *Nature Physics*, vol. 19, 1116–1121. <https://doi.org/10.1038/s41567-023-02042-2>
24. Wang, S., Fontana, E., Cerezo, M., et al. (2021). Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, vol. 12, article ID 6961. <https://doi.org/10.1038/s41467-021-27045-6>

О. А. Зарічковий, Е. В. Жаріков, К. І. Ліщук, Ю. В. Полупан

Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", м. Київ, Україна

МЕТОДИ МІНІМІЗАЦІЇ ПОМИЛОК У КВАНТОВИХ НЕЙРОННИХ МЕРЕЖАХ ТА ЇХ ВПЛИВ НА ЗБІЖНІСТЬ МЕТОДУ ГРАДІЄНТНОГО СПУСКУ

Проаналізовано сучасні стратегії мінімізації помилок у варіаційних квантових алгоритмах за умов функціонування апаратного забезпечення епохи NISQ (англ. *Noisy Intermediate-Scale Quantum* – шумний квантовий комп'ютер проміжного масштабу). Досліджено вплив апаратного шуму, який діє як глобальний деполяризуючий канал на поглиблення ефекту "безплідних плато" та критичну деградацію процесу збіжності градієнтного спуску. Застосовано спеціалізовані симулятори шуму на базі фреймворку Qiskit Aer для точного моделювання поведінки реальних квантових пристроїв. Оцінено безпосередню ефективність методів екстраполяції нульового шуму ZNE (англ. *Zero Net Energy*), імовірісного скасування помилок PEC (англ. *Probabilistic Error Cancellation*), корекції помилок зчитування REM (англ. *Readout Error Mitigation*) та регресії даних Кліффорда CDR (англ. *Clifford Data Regression*). З'ясовано, що інтеграція вказаних методів дає можливість системі успішно відновити справжній напрямок вектора градієнта шляхом математичної компенсації затухання амплітуди, спричиненого апаратною декогеренцією. Виявлено, що процес відновлення геометрії ландшафту функції втрат неминуче супроводжується значним зростанням дисперсії статистичних оцінок, що, водночас, вимагає експоненційного або лінійного збільшення розміру вибірки вимірювань для запобігання осциляціям. Встановлено, що корекція методу REM виступає обов'язковим базовим етапом класичного оброблення градієнтів, який не потребує додаткових квантових ресурсів, тоді як практична придатність методу ZNE жорстко обмежується неглибокими квантовими ланцюгами через ризик накопичення критичних неточностей екстраполяції на пізніх епохах оптимізації нейромереж. Доведено, що попри здатність алгоритму PEC формувати ідеально незміщені оцінки градієнта, його експоненційні накладні витрати на вибірку роблять метод PEC апаратно непридатним для глибоких архітектур, виводячи на перший план метод CDR як найбільш масштабовану гібридну альтернативу. Розроблено інноваційну методологію динамічної оркестрації методів мінімізації шумів, яка стратегічно поєднує методи ZNE, CDR та PEC упродовж усього циклу навчання нейромереж для забезпечення стабільної збіжності за одночасної мінімізації споживання обчислювальних ресурсів квантового співпроцесора. Сформовано надійне теоретичне та інженерне підґрунтя для проєктування архітектури апаратно-стійких гібридних квантово-класичних моделей машинного навчання та їх подальшої інтеграції у сучасні виробничі конвеєри MLOps (англ. *Machine Learning Operations*).

Ключові слова: квантове машинне навчання; варіаційні квантові алгоритми; безплідні плато; епоха NISQ; деполяризуючий канал; правило зсуву параметрів.