Noise-Adaptive Hybrid Variational Quantum Circuits for Industrial Bearing Fault Diagnosis

Contraction-aware depth selection, Clifford-regression readout calibration and noise-in-the-loop training under depolarizing and amplitude-damping noise

[1]Chandrasekhar Panda, [2]Pravat Satpathy, [3]Sonia Mohapatra
[1]Researcher, [2]Researcher, [3]Researcher
[1,2,3]Research and Development Division,
Tekstein Solutions Private Limited, Bhubaneswar, Odisha, India
[1]contact@chandrasekharpanda.com, [2]analytics.pravat@gmail.com, [3]smctc98@gmail.com

***Abstract*—Variational quantum classifiers lose accuracy on noisy hardware because every noisy layer contracts the measured expectation values toward a fixed point, which shrinks the decision margin relative to the finite-shot measurement uncertainty and, for non-unital channels, biases the readout. A fault-diagnosis circuit trained noiselessly and deployed on such a device therefore degrades in a way that depends on its depth, its encoding and the noise type. This paper proposes a noise-adaptive hybrid variational quantum circuit (NA-VQC) for rolling-element bearing fault classification that measures the noise it will face and sets its own structure before training: single-qubit calibration circuits recover the channel strength, least-squares regression over random Clifford instances of the ansatz estimates the circuit-level readout contraction and bias, kernel-target alignment of the noisy encoding fixes the encoding scale, a contraction-weighted margin criterion derived from a finite-shot flip bound fixes the depth, the circuit is trained under the estimated channel, and an affine readout correction can be refreshed when the noise drifts. A physics-based vibration simulator with the geometry of the SKF 6205 bearing provides a four-class data set at zero-decibel signal-to-noise ratio, from which twelve time-domain and envelope-spectrum features are projected onto four qubits. An exact density-matrix simulator with adjoint gradients, verified against Qiskit Aer to machine precision, evaluates local depolarizing and amplitude-damping channels at 256, 1024 and unlimited measurement shots. The circuit-level contraction is shown to fall well below the single-qubit estimate once entangling gates are present, which explains why naive calibration fails. Across the noise sweep NA-VQC stays within four percentage points of a support vector machine on the same features where the standard pipeline falls to chance, and the depth rule avoids the depths at which accuracy collapses while forgoing at most two points against an exhaustive sweep. Negative findings, including the noise insensitivity of the encoding-scale selection and the redundancy of readout calibration under matched training noise, are reported with their causes.**

***Index Terms*—Amplitude damping, bearing fault diagnosis, Clifford data regression, depolarizing noise, hybrid quantum-classical learning, kernel-target alignment, noise-aware circuit design, variational quantum classifier.**

I. Introduction

Rolling-element bearings fail more often than any other component of rotating machinery, and the characteristic impulse trains that a spalled race or a cracked ball imprints on the casing vibration have been the basis of condition monitoring for four decades [1], [2]. Classical learning pipelines built on envelope analysis, support vector machines and, more recently, convolutional networks trained on raw segments reach near-perfect accuracy on the public benchmarks when the signal-to-noise ratio is favourable [3], [4]. Their behaviour at low signal-to-noise ratio, with few labelled examples per operating condition, remains the practically relevant regime in a plant, and it is in this regime that alternative feature maps continue to be explored.

Parameterized quantum circuits [25] offer one such feature map. A small register of qubits, a data-dependent rotation layer and a trainable entangling block define a function class whose Fourier spectrum grows with the number of data re-uploads [5], [6], and several groups have reported hybrid quantum-classical models for bearing and machinery diagnosis that are competitive with classical multilayer perceptrons on the Case Western Reserve University data [7], [8]. Those studies almost always train and evaluate the circuit on a noiseless state-vector simulator. The devices that would actually run such a circuit in the foreseeable future are noisy: every gate layer is followed by decoherence that, for depolarizing noise, contracts every expectation value toward zero and, for relaxation, drags it toward the ground-state value. The consequences for a classifier are twofold. First, the trainable classical head that follows the circuit is fitted to the noiseless readout scale, so a contracted readout at deployment is misread, a mechanism distinct from the gradient-vanishing effect of barren plateaus [9], [26]. Second, the finite number of measurement shots sets a floor on the readout uncertainty that does not shrink with the signal, so the effective signal-to-noise ratio of the readout decreases geometrically with depth.

Error-mitigation techniques such as zero-noise extrapolation and Clifford data regression correct expectation values after the fact [10]–[12], and noise-aware training has been studied for compiling and for anomaly detection [13], [14]. What is missing is a classifier design procedure in which the measured noise decides the structural hyperparameters of the circuit, namely the encoding scale and the depth, before any training takes place, together with an analysis of which parts of such a procedure matter and which do not. This paper addresses that gap for industrial fault diagnosis. The contributions are as follows.

1) A proposition that quantifies the readout contraction of a layered ansatz under global depolarizing noise exactly and bounds it under local depolarizing noise through the Pauli weight of the Heisenberg-evolved observable, followed by a numerical demonstration that the single-qubit estimate overstates the circuit-level contraction by a factor of four at a per-layer error of 5 percent and four layers.
2) A noise-adaptive design procedure, NA-VQC, with five components: single-qubit noise characterization, Clifford-regression estimation of the per-readout contraction and bias, kernel-target alignment selection of the encoding scale on the noisy encoding, a contraction-weighted margin criterion for depth selection derived from a finite-shot flip bound, and noise-in-the-loop training with affine readout calibration.
3) A reproducible evaluation on a physics-based bearing data set at zero-decibel signal-to-noise ratio, with an exact density-matrix simulator verified against Qiskit Aer, covering depolarizing and amplitude-damping channels, three measurement budgets, a brute-force depth sweep, a component ablation, a noise-drift test and a comparison with support vector machines, gradient-boosted trees and a one-dimensional convolutional network.
4) An explicit account of what did not work: the encoding-scale selection is useful but not noise-sensitive, and the affine readout calibration is mathematically redundant once the head is trained under the same noise, so its value is confined to the mismatch and drift scenarios.

Section II reviews related work, Section III states the problem, Section IV develops the signal model, the circuit, the contraction analysis and the adaptive procedure, Section V describes the simulation setup, Sections VI and VII present and compare the results, Section VIII summarizes the outcomes and Section IX concludes.

II. Related Work

Table I summarizes the strands of prior work that this paper draws on. Bearing diagnostics rests on the impulse-response model of McFadden and Smith [1], in which each defect contact excites a structural resonance that decays before the next contact, and on envelope analysis, which demodulates the resonance band to expose the contact rate [2], [3]. Deep networks trained on raw segments remove the band-selection step but require far more data than a plant typically has for any one machine [4], [17].

Quantum classifiers entered this field through the work of Maior et al. [7], [8], who connected one- to ten-layer parameterized circuits to a classical multilayer perceptron and reported an advantage over the perceptron alone. Their circuits were evaluated noiselessly. The expressivity of such models is well understood: data re-uploading widens the accessible Fourier spectrum [5], [6], while entangling depth increases expressibility at the cost of trainability [18], [19]. Noise enters this picture in two ways. Wang et al. [9] proved that local Pauli noise flattens the cost landscape exponentially in depth, and Fontana et al. [20] and Sharma et al. [13] showed that variational training can nevertheless absorb part of the noise. On the readout side, measurement-error mitigation [21], zero-noise extrapolation [10] and Clifford data regression [12] correct expectation values but are not used, in the published literature on fault diagnosis, to decide the structure of the classifier. Kernel-target alignment, a classical criterion for kernel selection [22], has been adopted to tune quantum embeddings [16] but has not been evaluated on noisy encodings.

The procedure proposed here combines these elements into a design loop in which the noise is measured first and the encoding scale, the depth, the training channel and the readout correction are derived from it. The analysis in Section IV-D also explains a failure mode that is easy to miss: calibrating a readout with single-qubit circuits, as measurement-error mitigation does, underestimates the contraction of an entangling circuit because the Heisenberg-evolved observable spreads onto high-weight Pauli strings.

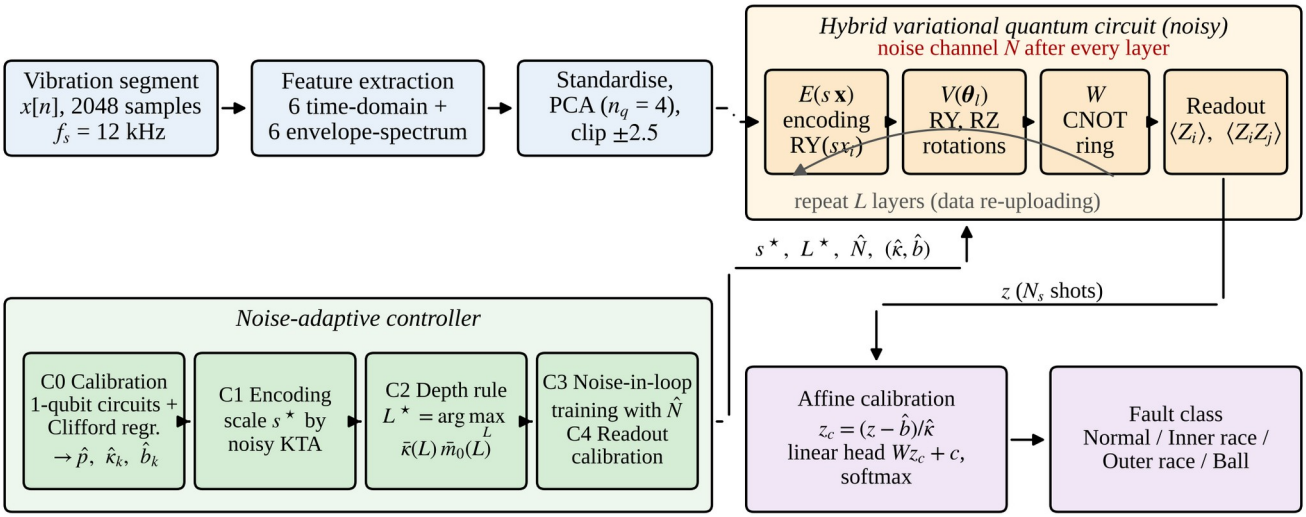


Fig. 1. Architecture of the noise-adaptive hybrid classifier. The classical front end (blue) produces four features; the noisy circuit (orange) is a data re-uploading ansatz with a noise channel after every layer; the controller (green) measures the noise and sets the encoding scale, the depth, the training channel and the readout calibration used by the linear head (violet).

TABLE I. Summary of related work and the gap addressed here

Reference	Methodology	Algorithms	Advantages	Limitation / gap
McFadden and Smith [1], 1984	Impulse-response model of a single point defect	Analytical, closed-form envelope	Explains spectral signature; basis of simulators	No learning; single defect
Smith and Randall [3], 2015	Benchmark of envelope methods on CWRU data	Band selection, spectral kurtosis	Establishes which records are diagnosable	Hand-tuned; sensitive to resonance band
Zhang et al. [4], 2017	Wide-kernel 1-D CNN on raw vibration	WDCNN, batch normalization	Robust to additive noise at high data volume	Needs thousands of segments; opaque
Havlíček et al. [15], 2019	Quantum-enhanced feature spaces	Quantum kernel, VQC	Proof of principle on hardware	Toy data; no noise adaptation
Pérez-Salinas et al. [5], 2020	Data re-uploading classifier	Single-qubit re-uploading	Universality with few qubits	Noiseless analysis
Schuld et al. [6], 2021	Expressivity of encodings	Fourier analysis of VQC outputs	Links re-uploads to spectrum	No noise; no selection rule
Wang et al. [9], 2021	Noise-induced barren plateaus	Gradient-norm bounds	Depth-noise trade-off proved	Training only; not readout bias
Czarnik et al. [12], 2021	Clifford data regression	Least-squares over near-Clifford circuits	Scalable error mitigation	Per-expectation; not a design rule
Hubregtsen et al. [16], 2022	Trainable embedding kernels	Kernel-target alignment on hardware	Cheap, label-aware encoding tuning	Kernel setting; noise not modelled
Maior et al. [7], 2023	Hybrid VQC + MLP for bearing PHM	1-, 5-, 10-layer ansätze	Outperforms MLP on CWRU and JNU	Noiseless simulation only
Kukliansky et al. [14], 2024	QNN anomaly detection on noisy devices	Noise-aware training, hardware runs	Demonstrates deployment effects	Fixed depth; no calibration
This work	Noise-adaptive VQC for bearing faults	Contraction bound, Clifford regression, KTA, depth rule, noise-in-loop	Structure chosen from measured noise; honest ablation	Simulation; 4 qubits; synthetic data

III. Problem Statement

A fault classifier is required that maps a vibration segment $x[n]$ from a bearing operating at an approximately known shaft speed to one of $C = 4$ health states: normal, inner-race defect, outer-race defect and rolling-element defect. The classifier is a hybrid model in which a parameterized quantum circuit produces a readout vector z and a classical linear head produces class scores. Three constraints distinguish the industrial setting from the noiseless studies in the literature.

Hardware noise. Each layer of the circuit is followed by a channel N acting on every qubit. Two channel families are considered: local depolarizing noise with per-layer probability p, which is unital and contracts the Bloch vector of each qubit uniformly, and amplitude damping with relaxation probability γ, which is non-unital and introduces a bias toward $|0\rangle$. The strengths p and γ are unknown to the designer and may drift between calibration and deployment.

Finite measurement budget. The expectation values in z are estimated from N_s shots, so each entry carries a variance of at

most $1/N_s$ regardless of the circuit. A contracted readout therefore has a lower effective signal-to-noise ratio than its noiseless counterpart even if the contraction is known exactly.

Small labelled data at low vibration signal-to-noise ratio. Only a few hundred labelled segments per machine are assumed to exist, and the vibration signal-to-noise ratio is 0 dB, so that the classical features themselves are uncertain.

The engineering question is therefore how the structural hyperparameters of the circuit, its encoding scale s and depth L, its training channel and its readout correction should be chosen from measurements of the device so that the deployed accuracy at a given shot budget is maximized, and which of those choices actually matter.

IV. Proposed Methodology

A. System Architecture and Workflow

Fig. 1 shows the architecture. A vibration segment is converted into twelve hand-crafted features, standardized, projected onto $n_q = 4$ principal components and clipped. The four features drive a data re-uploading circuit whose readout, after an affine calibration, feeds a linear softmax head. A noise-adaptive controller sits beside the circuit and performs the calibration, the encoding-scale selection, the depth selection and the training. Fig. 2 gives the corresponding flowchart; the deployment stage can return to the calibration step alone when the noise drifts, without retraining the circuit.

B. Bearing Vibration Model and Feature Extraction

The simulator follows the impulse-response model of [1]. For a defect with characteristic period T_d the acceleration is

$$x(t) = \Sigma_k A_k\, h(t - kT_d - \tau_k)\, M(t) + \Sigma_h a_h \sin(2\pi h\, f_r t + \varphi_h) + w(t), \qquad (1)$$

where $h(t) = \exp(-\beta t)\sin(2\pi f_n t)$ is the structural impulse response with resonance f_n and damping β, A_k is the impact amplitude, $\tau_k \sim \mathrm{N}(0, (0.015T_d)^2)$ models random slip, $M(t)$ is a load-zone modulation at the shaft frequency f_r for inner-race defects and at the cage frequency for rolling-element defects, the second sum represents shaft imbalance harmonics and $w(t)$ is white Gaussian noise scaled to the target signal-to-noise ratio. The defect frequencies follow from the bearing geometry with N_b balls of diameter d on a pitch diameter D at contact angle α:

$$\begin{aligned} f_{\mathrm{BPFO}} &= (N_b/2)\, f_r\, (1 - d\cos\alpha/D), \\ f_{\mathrm{BPFI}} &= (N_b/2)\, f_r\, (1 + d\cos\alpha/D), \\ f_{\mathrm{BSF}} &= (D/2d)\, f_r\, [1 - (d\cos\alpha/D)^2]. \end{aligned} \qquad (2)$$

For the SKF 6205 geometry ($N_b = 9$, $d = 7.94$ mm, $D = 39.04$ mm, $\alpha = 0$) these evaluate to $3.585f_r$, $5.415f_r$ and $2.357f_r$; rolling-element impacts occur at twice f_{BSF}. Fig. 3 shows segments of each class at 0 dB and their envelope spectra. Twelve features are extracted per segment: root-mean-square value, kurtosis, crest factor, skewness, shape factor and impulse factor in the time domain; and, from the Hilbert envelope of the 2–5 kHz band, the normalized energies within ±4 percent of the first two harmonics of f_{BPFO}, f_{BPFI} and $2f_{\mathrm{BSF}}$ with f_r estimated from the raw spectrum, the envelope spectral centroid, the envelope spectral kurtosis and the ratio of the dominant envelope line to the mean level. These correspond to standard functions of the MATLAB Signal Processing Toolbox and were implemented in SciPy. The features are standardized, reduced to n_q components by principal component analysis fitted on the training fold, standardized again and clipped to ±2.5 so that the encoding angles remain bounded.

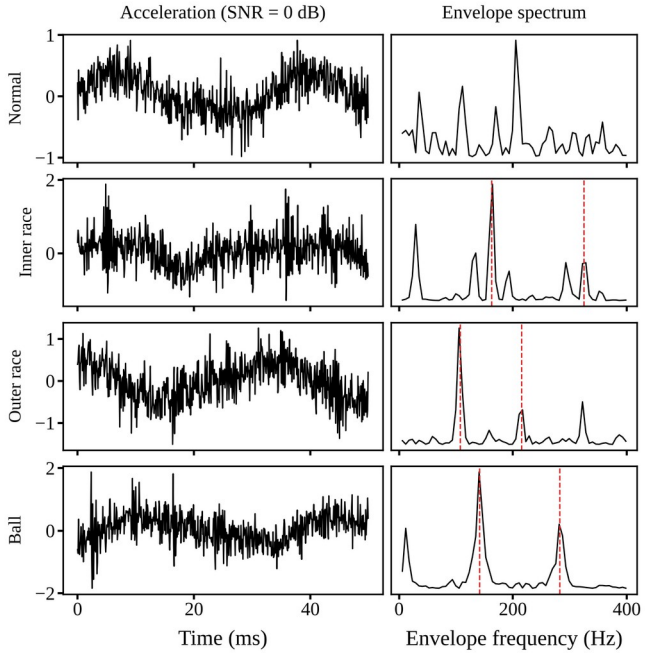


Fig. 3. Simulated acceleration segments at 0 dB signal-to-noise ratio for the four classes (left) and their envelope spectra after 2–5 kHz band-pass demodulation (right). Dashed lines mark the first two harmonics of the characteristic frequency of each defect.

C. Hybrid Variational Circuit and Noise Channels

The circuit acts on n_q qubits and consists of L identical blocks. Block l applies an encoding layer, a rotation layer and an entangling layer:

$$U_l(\mathbf{x}, \boldsymbol{\theta}_l) = W \cdot \Pi_i\, R_Z(\theta_{l,i,1})\, R_Y(\theta_{l,i,0}) \cdot \Pi_i\, R_Y(s\, x_i), \qquad (3)$$

where W is a ring of CNOT gates ($i \to i{+}1 \bmod n_q$), s is the encoding scale and $\boldsymbol{\theta}$ collects $2n_qL$ trainable angles. Re-uploading the data in every block gives the readout a Fourier spectrum in each x_i with L harmonics [6]. Fig. 4 draws one block. The noise channel N is applied to every qubit after each of the three layers, so a depth-L circuit contains $M = 3L$ noisy layers. The output state is

$$\begin{aligned} \rho(\mathbf{x}, \boldsymbol{\theta}) &= \Phi_L \circ \cdots \circ \Phi_1(|0\rangle\langle 0|^{\otimes n}), \\ \Phi_l &= N^{\otimes n} \circ W \circ N^{\otimes n} \circ V_l \circ N^{\otimes n} \circ E_l, \end{aligned} \qquad (4)$$

with E_l and V_l the encoding and rotation unitaries written as superoperators. The local depolarizing channel and the amplitude-damping channel are

$$\begin{aligned} N_{\mathrm{dep}}(\rho) &= (1-p)\rho + (p/3)(X\rho X + Y\rho Y + Z\rho Z) \\ &= (1 - 4p/3)\rho + (4p/3)\,\mathrm{Tr}_q(\rho) \otimes I/2, \end{aligned} \qquad (5)$$

$$\begin{aligned} N_{\mathrm{amp}}(\rho) &= K_0 \rho K_0^{\dagger} + K_1 \rho K_1^{\dagger}, \\ K_0 &= \mathrm{diag}(1, (1-\gamma)^{1/2}), \quad K_1 = \gamma^{1/2}\, |0\rangle\langle 1|. \end{aligned} \qquad (6)$$

The readout vector collects all expectation values that a single computational-basis measurement provides, namely the n_q single-qubit values and the $n_q(n_q{-}1)/2$ two-qubit correlators,

$$z = [\langle Z_1\rangle, \ldots, \langle Z_n\rangle, \langle Z_1 Z_2\rangle, \ldots, \langle Z_{n-1} Z_n\rangle]^{\mathrm{T}} \in [-1, 1]^{n_{\mathrm{out}}}, \qquad (7)$$

so that $n_{\mathrm{out}} = 10$ for four qubits and no additional shots are needed for the correlators. The classical head applies an affine calibration followed by a linear map and a softmax:

$$z_c = (z - \hat{b}) \oslash \hat{\kappa}, \quad o = W z_c + c,$$
$$P(y = k \mid \mathbf{x}) = \exp(o_k) / \Sigma_j \exp(o_j), \tag{8}$$

where $\oslash$ is element-wise division and $(\hat{\kappa}, \hat{b})$ are the per-readout calibration constants of Section IV-E. Training minimizes the regularized cross-entropy

$$\mathscr{L}(\boldsymbol{\theta}, W, c) = -(1/B) \Sigma_i \log P(y_i \mid \mathbf{x}_i) + \lambda \|W\|_F^2, \tag{9}$$

with Adam updates. The gradient with respect to a rotation angle θ that generates $\exp(-i\theta P/2)$ with Pauli P is obtained from the parameter-shift rule, which is exact for this gate set [23], [24]:

$$\partial\langle O\rangle/\partial\theta = [\langle O\rangle(\theta + \pi/2) - \langle O\rangle(\theta - \pi/2)] / 2$$
$$= -(i/2)\,\mathrm{Tr}([O, P]\,\rho_\theta), \tag{10}$$

where ρ_θ is the state immediately after the gate and O is the loss-weighted observable propagated backward through the channel adjoints. The right-hand identity is what the simulator evaluates (an adjoint pass that costs a constant number of circuit passes instead of two per parameter); on hardware the left-hand form would be used. The two were verified to agree to 10^{-15} in all noise settings.

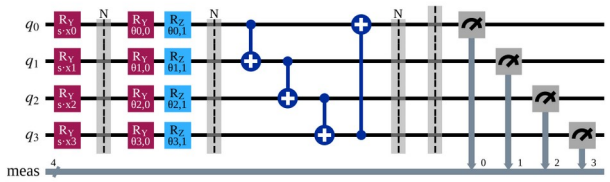


Fig. 4. One block of the four-qubit ansatz as rendered by Qiskit: an encoding layer $R_Y(s\,x_i)$, a rotation layer $R_Y R_Z$, and a CNOT ring. The barriers labelled N mark where the noise channel acts on every qubit; the register is read out in the computational basis.

D. Readout Contraction Under Noise

Proposition 1 (contraction of the readout). Let O be a traceless observable and let the circuit contain M noisy layers. (i) Under global depolarizing noise $\rho \mapsto (1-p)\rho + p\, I/2^n$ applied after each layer, $\langle O\rangle_{\text{noisy}} = (1-p)^M \langle O\rangle_{\text{ideal}}$ exactly, for every circuit and every input. (ii) Under local depolarizing noise of strength p on every qubit, write the Heisenberg-evolved observable entering the m-th noisy layer (counted from the measurement) as a Pauli expansion $O^{(m)} = \Sigma_\sigma c_\sigma^{(m)}\,\sigma$. The channel scales each string by $(1 - 4p/3)^{|\sigma|}$, where $|\sigma|$ is its Pauli weight. Hence, for any ensemble of circuits and inputs over which distinct Pauli strings are uncorrelated in expectation, the least-squares contraction coefficient κ of the noisy readout on the ideal readout is a weighted mean of per-string attenuation factors and satisfies

$$\kappa \le (1 - 4p/3)^M, \tag{11}$$

with equality if and only if all weight of every $O^{(m)}$ lies on weight-one strings.

Proof sketch. (i) The global channel maps any traceless operator A to $(1-p)A$, and unitary conjugation preserves tracelessness, so the factor accumulates once per layer. (ii) The adjoint of the local channel multiplies the non-identity Pauli factor of each qubit by $(1 - 4p/3)$ (Eq. 5), so a string of weight $|\sigma|$ is scaled by that factor raised to $|\sigma|$. Every string of a traceless operator has weight at least one, so no string is attenuated by less than $(1 - 4p/3)$ per layer; under the uncorrelated-string assumption the cross terms in the least-squares numerator vanish and κ is a convex combination of products of per-layer factors, each of which is at most $(1 - 4p/3)$. Equality requires that no entangling layer ever creates weight-two or higher strings. □

The proposition has a practical consequence. A single-qubit calibration circuit, which is the usual device for measurement-error mitigation, realizes exactly the equality case and therefore returns the upper bound $(1 - 4p/3)^M$, not the contraction that an entangling classifier experiences. Section VI-A shows that at $p = 0.05$ and $L = 4$ the true contraction is 0.12 against a single-qubit estimate of 0.44. For amplitude damping the same mechanism applies to the contraction, and the bias b measured on a single qubit, $1 - (1-\gamma)^M$, overstates the bias of an entangled register, whose average readout is scrambled toward zero by the CNOT layers.

Corollary (finite-shot flip probability). With N_s shots the estimate of each readout has variance at most $1/N_s$. For a sample whose noiseless normalized margin in readout space is m, the probability that shot noise flips the decision satisfies

$$P(\text{flip}) \le \exp(-\kappa^2 m^2 N_s/2), \tag{12}$$

by the Gaussian tail bound applied to the projection of the readout onto the normal of the separating hyperplane. The product κm is thus the quantity that a noise-adaptive design should maximize at fixed shot budget; depth increases m through expressivity and decreases κ geometrically, which produces an interior optimum.

E. Noise-Adaptive Controller

C0: noise characterization. Two single-qubit circuits with identity gates and M noisy layers, prepared in $|0\rangle$ and $|1\rangle$, return readouts z_0 and z_1. With $\kappa_1 = (z_0 - z_1)/2$ and $b_1 = (z_0 + z_1)/2$ the channel strength is recovered from

$$\hat{p} = (3/4)(1 - \kappa_1^{1/M}), \quad \hat{\gamma} = 1 - (1 - b_1)^{1/M}, \tag{13}$$

because the single-qubit circuit realizes the equality case of Eq. 11 and, for damping, $b_1 = 1 - (1-\gamma)^M$. The estimate uses $N_{\text{cal}} = 4096$ shots. The estimated channel $\hat{N}$ is then used wherever a noise model is required.

C4: Clifford-regression readout calibration. To obtain the circuit-level contraction, $K = 32$ random instances of the ansatz are drawn with every rotation and encoding angle restricted to multiples of $\pi/2$, so that each instance is a Clifford circuit whose ideal readout is classically computable at any register size [12]. For each of the n_{out} readouts the measured noisy values z^N are regressed on the ideal values z^0:

$$(\hat{\kappa}_k, \hat{b}_k) = \arg\min_{(\kappa,b)} \Sigma_j (z_{jk}^{\,N} - \kappa\, z_{jk}^{\,0} - b)^2, \quad k = 1, \ldots, n_{\text{out}}, \tag{14}$$

over the $K \cdot n_{\text{in}}$ calibration points ($n_{\text{in}} = 16$ encoding settings per instance). The mean $\bar{\kappa}$ over the single-qubit readouts is also the contraction used by the depth rule. The calibration can be refreshed on a deployed device without retraining because it touches only the affine layer in Eq. 8.

C1: encoding-scale selection by noisy kernel-target alignment. For candidate scales $s \in \{0.25, 0.5, 0.75, 1, 1.5, 2\}$ the noisy encoded states $\rho_i = \hat{N}^{\otimes n}(E(s\mathbf{x}_i)|0\rangle\langle 0|E^\dagger)$ of a 160-sample subset are formed, the Gram matrix $K_{ij} = \mathrm{Tr}(\rho_i \rho_j)$ is computed, and the centred kernel-target alignment [22]

$$\mathrm{KTA}(s) = \langle HKH, HYY^{\mathrm{T}}H\rangle_F / (\|HKH\|_F\, \|HYY^{\mathrm{T}}H\|_F),$$
$$H = I - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N, \tag{15}$$

is evaluated, where Y is the one-hot label matrix. The scale with the largest alignment is adopted. The computation needs no training and no gradients.

C2: depth selection. For each candidate depth $L \in \{1, \ldots, 6\}$ a short noiseless probe (40 Adam steps on four fifths of the training set) yields a model whose normalized margins on the held-out fifth are

$$m_i = \min_{j \neq y_i} [(w_{y_i} - w_j)^{\mathrm{T}} z_i + c_{y_i} - c_j] \,/\, \|w_{y_i} - w_j\|,$$
$$\bar{m}_0(L) = \mathrm{mean}_i \max(m_i, 0), \tag{16}$$

and the depth is chosen by the contraction-weighted criterion that follows from Eq. 12:

$$L^\star = \arg\max_L J(L), \quad J(L) = \bar{\kappa}(L)\, \bar{m}_0(L). \tag{17}$$

The probe does not depend on the noise, so it is computed once per data fold and reused across noise levels; only $\bar{\kappa}(L)$ is re-estimated.

C3: noise-in-the-loop training. The parameters ($\boldsymbol{\theta}$, W, c) are trained with the forward model of Eq. 4 under $\hat{N}$ and with the calibration constants of C4 already inserted in Eq. 8, so that the head learns the scale it will see at deployment. A standard pipeline, by contrast, trains on the noiseless model and deploys the result unchanged.

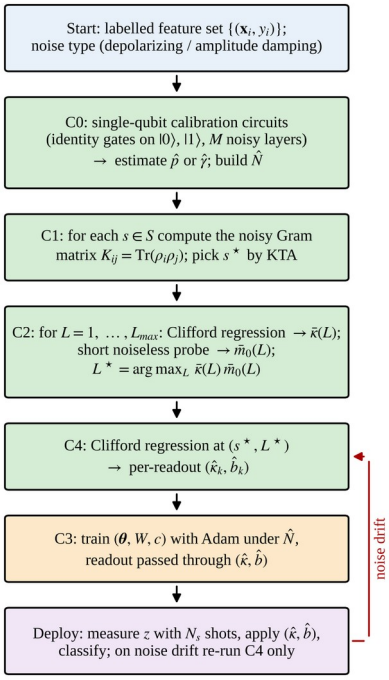


Fig. 2. Flowchart of the NA-VQC procedure. Steps C0–C2 and C4 run before training; C3 trains under the estimated channel; at deployment only C4 is repeated when the noise drifts.

F. Algorithm and Complexity

Algorithm 1 lists the procedure. Let $d = 2^{n_q}$, B the training-set size and T the number of Adam steps. One density-matrix forward pass of a depth-L circuit costs $O(B\, L\, n_q\, d^2)$ and the adjoint pass the same order, so training costs $O(T\, B\, L\, n_q\, d^2)$; with $d = 16$ this is 0.4 s per step for $B = 560$ on one CPU core. The Clifford regression costs $O(K\, n_{\mathrm{in}}\, L\, n_q\, d^2)$ per depth, the kernel-target alignment $O(N^2\, d^2)$ for the subset size N, and the depth rule $O(L_{\max}\, T_{\mathrm{probe}}\, B\, L\, n_q\, d^2)$. On hardware the parameter-shift gradient replaces the adjoint pass with $2 \cdot 2 n_q L$ circuit executions per step, and the calibration adds $K\, n_{\mathrm{in}} + 2$ executions. Memory is dominated by the $B \times d \times d$ complex state, 1.1 MB at $B = 560$.

Algorithm 1 NA-VQC design and training

Input: training set $\{(\mathbf{x}_i, y_i)\}$, noise family, candidate scales S, depths $\{1, \ldots, L_{\max}\}$, shots N_{cal}
Output: $(s^\star, L^\star, \boldsymbol{\theta}, W, c, \hat{\kappa}, \hat{b})$
1: run single-qubit calibration circuits; compute $\hat{p}$ or $\hat{\gamma}$ by Eq. 13; set $\hat{N}$
2: **for** $s \in S$: form noisy Gram matrix; compute KTA(s) by Eq. 15; $s^\star \leftarrow \arg\max$
3: **for** $L = 1, \ldots, L_{\max}$: Clifford regression (Eq. 14) → $\bar{\kappa}(L)$; noiseless probe → $\bar{m}_0(L)$; $J(L) \leftarrow \bar{\kappa}(L)\bar{m}_0(L)$
4: $L^\star \leftarrow \arg\max_L J(L)$
5: Clifford regression at $(s^\star, L^\star)$ with N_{cal} shots → $(\hat{\kappa}_k, \hat{b}_k)$ for all readouts
6: **for** $t = 1, \ldots, T$: forward Eq. 4 under $\hat{N}$; readout Eq. 7; head Eq. 8; loss Eq. 9; gradients Eq. 10; Adam update
7: **deploy:** measure z with N_s shots, apply $(\hat{\kappa}, \hat{b})$, classify; **on drift:** repeat step 5 only

V. Simulation Setup

Software and hardware. All code is Python 3.13 with NumPy 2.5, SciPy 1.18, scikit-learn 1.9, XGBoost 3.4, PyTorch 2.14 and Qiskit 2.6 [27] with Qiskit Aer 0.18, run on two CPU cores of a Linux container with 7 GB of memory. The quantum model is evaluated with an in-house batched density-matrix simulator that applies the channels of Eqs. 5 and 6 in closed form on the qubit index block; its readouts agree with Qiskit Aer's density-matrix method to 6×10^{-16} under all three noise settings, and its adjoint gradients agree with the parameter-shift rule to 1.3×10^{-15}. The in-house simulator is used because it supports batched adjoint differentiation under noise, which Aer does not; Aer serves as the independent reference. The feature extraction uses SciPy implementations of the corresponding MATLAB Signal Processing Toolbox operations; no MATLAB session was used, and the Deep Learning Toolbox baseline named in the study plan is realized in PyTorch.

Data set. The data are generated by the simulator of Section IV-B rather than recorded on a test rig; this is stated as a limitation in Section IX. Each class contains 200 segments spread uniformly over four shaft speeds and four severities, so the normal class and each defect class are balanced and every defect appears at an incipient amplitude of 0.3. The main study uses 0 dB vibration signal-to-noise ratio, at which a tuned support vector machine on the same four principal components reaches 86.3 percent; the principal components retain 83 percent of the feature variance. Table II lists the parameters. Seeds are fixed throughout: the data seed is 7, the split and initialization seed runs over the listed range, and the shot-noise generator is seeded with 1000 plus the split seed.

Test scenarios. (i) Noise sweep: standard VQC (depth 4, $s = 1$, trained noiselessly), the same circuit trained under the true noise (VQC-NT) and NA-VQC, at every noise level and three shot budgets. (ii) Depth sweep: a noise-trained, calibrated circuit at every depth 1–6 against the depth rule. (iii) Ablation: each component of NA-VQC removed in turn at $p = \gamma = 0.05$, plus a drift test in which the circuit is trained at 0.02 and deployed at 0.05 with and without re-calibration. (iv) Vibration-noise sweep against classical baselines: SVM with grid-searched C and kernel width, XGBoost with 300 trees, a

multilayer perceptron, and a wide-kernel one-dimensional CNN with 11 252 parameters trained for 150 epochs on raw segments.

TABLE II. Simulation and model parameters

Parameter	Value
Sampling rate / segment length	12 kHz / 2048 samples (171 ms)
Shaft speeds	1797, 1772, 1750, 1730 r/min (±0.5 % jitter)
Defect severities (relative impulse amplitude)	0.3, 0.6, 1.0, 1.5
Resonance f_n / damping β	U(2.8, 4.2) kHz / U(600, 1100) s^{-1}
Slip jitter σ_{slip}	1.5 % of T_d
Vibration SNR (main study)	0 dB (sweep: 10, 5, 0, −5 dB)
Segments per class / total	200 / 800; stratified 70/30 split
Qubits / readouts	4 / 10
Candidate scales / depths	{0.25, 0.5, 0.75, 1, 1.5, 2} / {1, …, 6}
Adam steps / learning rate / λ	150 / 0.08 / 10^{-3}
Probe steps for depth rule	40
Clifford instances K / encodings n_{in}	32 / 16
Calibration shots N_{cal}	4096
Deployment shots N_s	256, 1024, exact (∞)
Depolarizing p per layer	0, 0.01, 0.02, 0.05, 0.10
Amplitude damping γ per layer	0.01, 0.02, 0.05, 0.10
Seeds (data 7; split and init)	0–4 main; 0–2 depth, ablation, SNR

VI. Results and Discussion

A. Readout Contraction: Validation of Proposition 1

Fig. 5 plots the contraction κ against the number of noisy layers for the three channels and two strengths, comparing the single-qubit calibration (dashed), a least-squares fit of noisy on ideal readouts over random non-Clifford circuits (solid) and the Clifford regression of Section IV-E (open markers). Under global depolarizing noise the three coincide and follow $(1-p)^M$ exactly, as part (i) of the proposition requires; the fit residual is at the 10^{-16} level. Under local depolarizing noise the single-qubit estimate is an upper bound, as part (ii) states, and the gap widens geometrically: at $p = 0.05$ and $M = 12$ the single-qubit value is 0.437 while the measured contraction is 0.115, a factor of 3.8. The Clifford regression reproduces the true value to 0.000, which confirms that near-Clifford instances of the ansatz carry the same Pauli-weight spreading as generic instances. Amplitude damping shows the same ordering with a smaller gap (0.540 against 0.308), and its single-qubit bias estimate of 0.460 overstates the fitted bias of 0.049 by an order of magnitude because the CNOT ring scrambles the ground-state attraction of individual qubits. The affine model itself is adequate: the root-mean-square residual of the fit is 0.008 for depolarizing and 0.024 for damping noise at $p = \gamma = 0.05$.

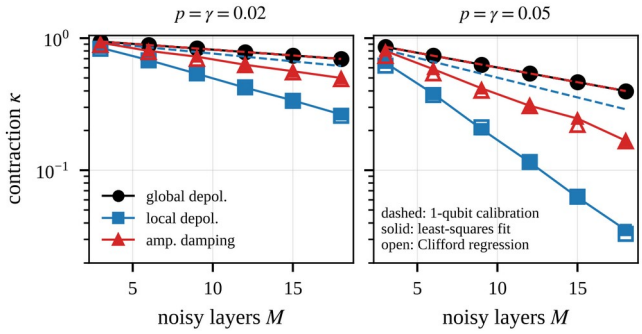


Fig. 5. Readout contraction κ versus number of noisy layers M for global depolarizing, local depolarizing and amplitude-damping channels at two strengths. Dashed: single-qubit calibration; solid: least-squares fit over random circuits; open markers: Clifford regression. The single-qubit curves of the global and the damping channel coincide because both equal $(1-p)^M$.

B. Noise Sweep

Table III and Fig. 6 report the test accuracy at 1024 shots for the standard circuit (depth 4, $s = 1$, noiseless training), the same circuit trained under the true noise, and NA-VQC, averaged over five splits. Three observations follow. First, the standard pipeline degrades steadily and then collapses: under depolarizing noise it falls from 76.9 percent noiselessly to 45.5 percent at $p = 0.05$ and 28.2 percent at $p = 0.1$, close to chance; under amplitude damping it reaches 35.5 percent at $\gamma = 0.1$. The mechanism is the one identified in Section IV-D: the head is fitted to an uncontracted readout, so when the readout shrinks by a factor of ten the bias terms dominate the logits and the prediction locks onto one or two classes. Second, noise-in-the-loop training on the fixed four-layer circuit recovers most of the loss at moderate noise (70.8 percent at $p = 0.05$, 80.7 percent at $\gamma = 0.1$) but not at $p = 0.1$ (42.3 percent), where the twelve noisy layers contract the single-qubit readouts to 0.012 of their ideal value and the 1024-shot uncertainty of ±0.03 per entry overwhelms the margin. Third, NA-VQC is flat across the sweep, between 81.9 and 84.5 percent under depolarizing noise and between 83.4 and 85.1 percent under damping, which is at most 4.4 points below the support vector machine trained on the same four principal components (86.3 percent) and within one standard deviation of it up to a per-layer error of 0.02. At every noisy level the rule selected depth 1 or 2 and the encoding scale 0.5; without noise it selected depth 6 and reached 86.3 percent, the best quantum result in the study.

The reduction from 256 to 1024 shots (last two accuracy columns of Table III) costs NA-VQC at most 2.7 percentage points under depolarizing noise, whereas the noise-trained deep circuit loses 8.3 points at $p = 0.05$ and 7.1 points at $p = 0.1$. This is the flip-probability bound of Eq. 12 at work: the one-layer circuit keeps κ between 0.62 and 0.70 at $p = 0.05$, so the product κm that governs the shot-noise flip rate is an order of magnitude larger than for the deep circuit.

TABLE III. Test accuracy (percent, mean ± std over five splits) at 1024 shots; last columns: NA-VQC at 256 shots and its selected depth / scale

Noise	VQC std.	VQC noise-trained	NA-VQC	NA-VQC @256	L / *s*
depol. 0	76.9 ± 2.8	—	86.3 ± 1.3	85.9	6 / 0.5
depol. 0.01	75.0 ± 3.3	76.8 ± 3.6	84.5 ± 2.0	83.6	1 / 0.5
depol. 0.02	69.9 ± 4.2	76.2 ± 2.9	83.9 ± 1.9	83.1	1 / 0.5

Noise	VQC std.	VQC noise-trained	NA-VQC	NA-VQC @256	L / *s*
depol. 0.05	45.5 ± 10.1	70.8 ± 3.5	82.8 ± 3.4	81.7	1 / 0.5
depol. 0.1	28.2 ± 2.2	42.3 ± 3.1	81.9 ± 3.4	79.3	1 / 0.5
amp. 0.01	76.2 ± 2.7	77.0 ± 3.0	85.1 ± 1.7	84.3	1 / 0.5
amp. 0.02	74.5 ± 3.4	76.9 ± 2.3	84.2 ± 2.3	83.5	1 / 0.5
amp. 0.05	59.6 ± 4.5	77.4 ± 0.7	84.0 ± 1.7	83.3	1 / 0.5
amp. 0.1	35.5 ± 6.1	80.7 ± 1.4	83.4 ± 3.3	82.4	1 / 0.5

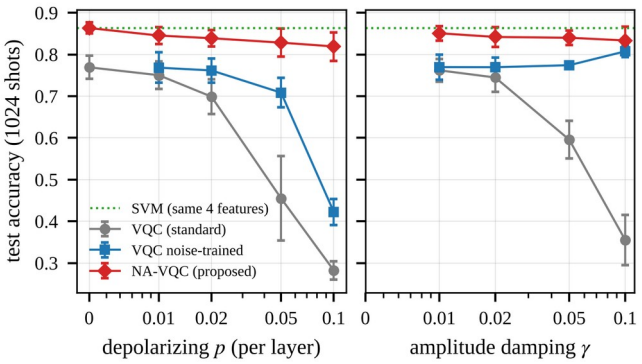


Fig. 6. Test accuracy at 1024 shots versus per-layer noise strength for depolarizing (left) and amplitude-damping (right) channels. Error bars: standard deviation over five splits. Dotted line: SVM on the same four features.

C. Depth Rule Against Brute Force

Table IV and Fig. 7 compare the depth selected by the rule of Eq. 17 with an exhaustive sweep in which a noise-trained, calibrated circuit was fitted at every depth from one to six. Two results are robust. The rule chose depth one in every split and setting, because the noiseless margin proxy $\bar{m}_0(L)$ is nearly flat in L on this four-feature task (between 0.3 and 0.4 in readout units) while $\kappa(L)$ falls geometrically, so the product is monotone decreasing. The exhaustive sweep shows that the brute-force optimum at 1024 shots is depth 4 at $p = 0.02$, depth 2 at $p = 0.05$ and depth 5 under damping, that is, one to four layers deeper than the rule's choice, but the accuracy the rule forgoes is 2.0, 0.7 and 1.8 percentage points respectively, inside the standard deviation across splits. The rule's real contribution is on the other side of the optimum: at $p = 0.05$ the accuracy at 1024 shots falls from 83.3 percent at depth two to 76.3 percent at depth four and 53.1 percent at depth six, and at 256 shots to 39.8 percent, so a conventional choice of four or more layers would cost between 7.0 and 30.2 points. A second observation concerns the exact-readout column of Table IV: with unlimited shots the best depth is 4, 5 and 5, and the accuracy at depth six under $p = 0.05$ is 85.7 percent exactly against 53.1 percent at 1024 shots. A noise study that reports exact expectation values therefore sees almost no depth penalty, because a trained linear head rescales a contracted but exact readout; the penalty appears only once the shot budget is finite, which is the regime of any real device. A shot-aware variant of the criterion, $J'(L) = A_0(L)[1 - \exp(-\kappa^2 \bar{m}_0^2 N_s/2)]$ with A_0 the probe accuracy, was evaluated post hoc from the same probe data; it selects depths two to four at 1024 shots and reaches regrets of 0.1 to 3.5 points, which is not a systematic improvement over Eq. 17, so the simpler rule is retained.

TABLE IV. Depth selected by the rule (L*) against the brute-force best depth at each shot budget (noise-trained, calibrated circuits, three splits); accuracies in percent at 1024 shots

Noise	L*	best 1024	best 256	best ∞	acc L*	acc best	acc L4	acc L6
dep. 0.02	1	4	2	4	82.9	85.0	85.0	83.7
dep. 0.05	1	2	1	5	82.6	83.3	76.3	53.1
amp. 0.05	1	5	2	5	82.6	84.4	83.8	83.7

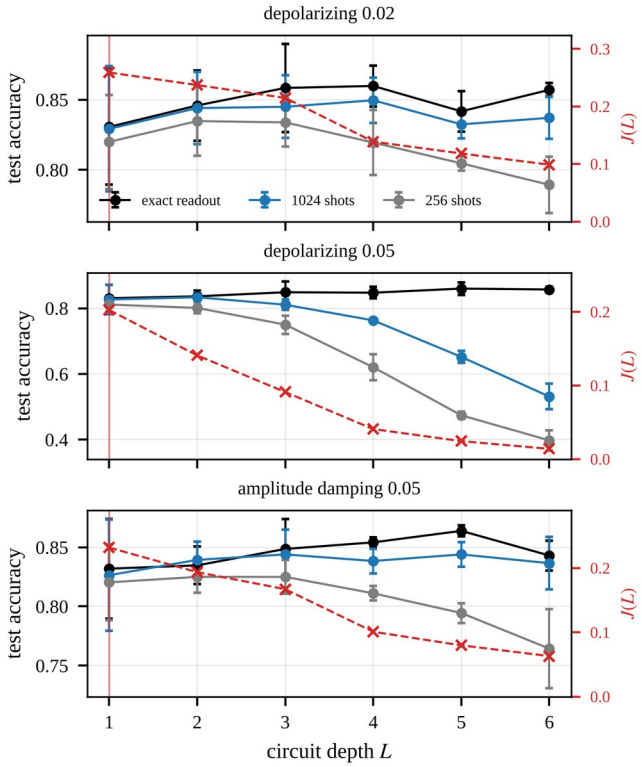


Fig. 7. Brute-force depth sweep for three noise settings: test accuracy of noise-trained, calibrated circuits at exact, 1024-shot and 256-shot readout (left axis) and the criterion J(L) of Eq. 17 (dashed, right axis). The vertical line marks the depth the rule selected most often.

D. Ablation and Noise Drift

Table V and Fig. 8(a) remove the components of NA-VQC one at a time at $p = \gamma = 0.05$ over three splits. The largest single contribution is noise-in-the-loop training (C3): training noiselessly and deploying with calibration loses 8.3 points under depolarizing and 6.4 points under damping noise, and removing both C3 and C4 loses 10.0 and 5.6 points. That the calibration alone recovers so little when training is noiseless shows that the affine map of Eq. 14 is only the first-order description of the channel: the residual of Section VI-A, and the different contraction of the single-qubit and correlator readouts, are large enough that the head needs to be refitted under the noise. The depth rule (C2) matters most at the small shot budget: fixing the depth at four costs 4.6 points at 1024 shots but 16.3 points at 256 shots under depolarizing noise, whereas with exact readout the four-layer circuit is 1.9 points better; under damping the shallow and the deep circuit are within 0.8 points at 1024 shots. The encoding scale (C1) contributes 0.7 and 1.3 points once the depth is one, far less than at depth four, where the same

change of scale moves the noiseless accuracy from 77.8 to 85.0 percent (three splits); the scale selection therefore interacts with depth and is not an independent gain. Removing the readout calibration (C4) under matched training noise changes nothing within the seed variance, as the linear-head argument of Section VI-F predicts.

The drift test addresses the scenario in which the calibration is worth keeping. A circuit designed and trained for a per-layer error of 0.02 was deployed at 0.05. Without any action its accuracy at 1024 shots is 78.1 percent under depolarizing and 77.6 percent under damping noise; refreshing the Clifford regression at the new noise level, which requires 512 circuit executions and no gradient step, restores 83.8 and 81.1 percent. The recovery is partial under damping because the bias of the correlator readouts shifts with the noise in a way the single affine layer only approximates, but in both cases the refreshed model is within the variance of a model trained at the new level.

TABLE V. Ablation at p = γ = 0.05 (percent, three splits): accuracy at 1024 shots (mean ± std) and at 256 shots

Configuration	Depol. @1024	@256	Amp. @1024	@256
Full NA-VQC	82.7 ± 4.5	81.2	84.1 ± 1.8	83.3
without C1 (s = 1)	82.0 ± 1.7	81.5	82.8 ± 2.2	82.3
without C2 (L = 4)	78.1 ± 1.2	64.9	83.3 ± 1.6	80.8
without C3 (noiseless training)	74.4 ± 6.2	73.6	77.7 ± 4.0	76.6
without C4 (no readout calibration)	82.5 ± 4.5	81.4	84.4 ± 1.9	83.1
without C3 and C4 (standard pipeline at s, *L*)	72.8 ± 7.4	72.5	78.4 ± 3.8	77.1
drift 0.02 → 0.05, no re-calibration	78.1 ± 4.2	76.3	77.6 ± 4.8	77.2
drift 0.02 → 0.05, C4 refreshed	83.8 ± 2.5	81.3	81.1 ± 1.3	80.3

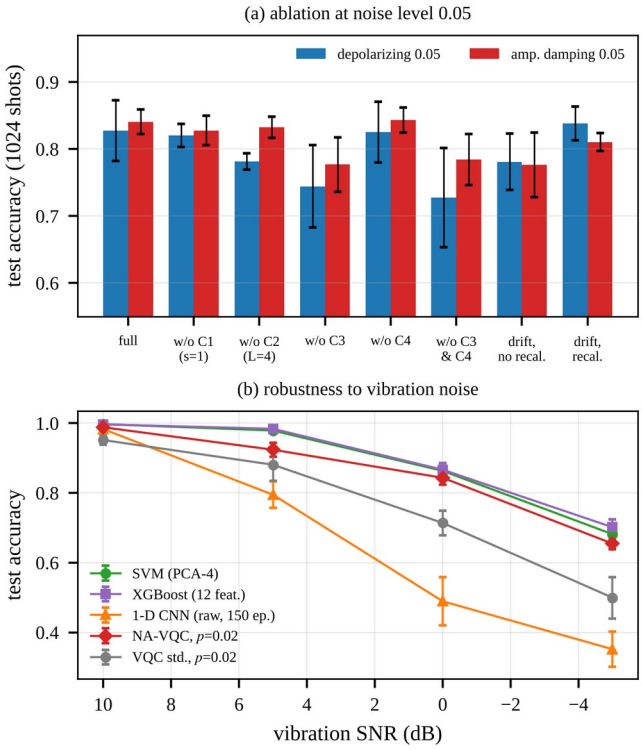


Fig. 8. (a) Ablation of the NA-VQC components at noise level 0.05 and the noise-drift test. (b) Test accuracy of the classical baselines, the standard VQC and NA-VQC (both at p = 0.02, 1024 shots) as the vibration signal-to-noise ratio decreases; error bars are standard deviations over splits.

E. Vibration Noise and Classical Baselines

Fig. 8(b) and Table VI place the quantum classifier among classical baselines as the vibration signal-to-noise ratio decreases from 10 to −5 dB, with NA-VQC and the standard circuit both deployed at $p = 0.02$ and 1024 shots. All feature-based models are near perfect at 10 dB and degrade together; at 0 dB the twelve-feature SVM and XGBoost reach 88.2 and 86.7 percent, the four-feature SVM 86.3 percent, and NA-VQC 84.4 percent, so the quantum model sits at the level of the classical model that sees the same four features and about 3.8 points below the best classical model that sees all twelve. The standard circuit trails NA-VQC by 3.7 points at 10 dB and by 13.0 points at 0 dB, so the adaptation gain grows as the classification problem hardens. At −5 dB every model is below 71 percent and NA-VQC is at 65.5 percent. The one-dimensional CNN on raw segments, with 560 training segments and no augmentation, is the weakest model below 5 dB (48.9 percent at 0 dB); it is reported because it is the architecture a practitioner would reach for first, and because the result illustrates that the feature-engineered front end, not the classifier, carries most of the performance at this data volume.

TABLE VI. Test accuracy (percent, mean ± std) versus vibration signal-to-noise ratio

Model	10 dB	5 dB	0 dB	−5 dB
SVM (12 features)	100.0 ± 0.0	98.9 ± 0.5	88.2 ± 1.9	69.2 ± 0.6
SVM (PCA-4)	99.8 ± 0.2	97.9 ± 0.7	86.3 ± 1.4	68.2 ± 1.4
XGBoost (12 features)	99.7 ± 0.3	98.4 ± 0.5	86.7 ± 1.9	70.2 ± 2.3
MLP (12 features)	99.9 ± 0.2	98.6 ± 0.4	83.8 ± 1.5	64.2 ± 1.6

Model	10 dB	5 dB	0 dB	−5 dB
1-D CNN (raw, 150 epochs)	98.3 ± 0.9	79.5 ± 3.8	48.9 ± 6.9	35.2 ± 5.0
VQC standard, p = 0.02, 1024 shots	95.2 ± 1.4	88.1 ± 4.6	71.4 ± 3.5	50.0 ± 5.9
NA-VQC, p = 0.02, 1024 shots	98.8 ± 0.3	92.4 ± 2.0	84.4 ± 2.0	65.5 ± 1.7

F. Negative and Non-Confirmatory Findings

Four findings run against the hypotheses that motivated the study and are reported as such. (i) The kernel-target alignment of the noisy encoding selected $s = 0.5$ at every noise level and for both channels, and the alignment values differ from their noiseless counterparts only in the third decimal; the component is useful, since $s = 0.5$ beats the conventional $s = 1$ by a wide margin, but it is not a noise-adaptive component on this data. (ii) The affine readout calibration is mathematically redundant once the head is trained under the same channel, because a linear head absorbs any fixed affine map of its input; the ablation confirms that removing it under matched noise changes the accuracy within the seed variance. Its value lies entirely in the mismatch and drift scenarios. (iii) The single-qubit calibration that is standard in measurement-error mitigation over-estimates the retained signal of an entangling circuit by a factor of four at the noise levels studied, so a design rule built on it would choose circuits that are too deep. (iv) The one-dimensional CNN trained on raw segments, which is the strongest classical method on large public benchmarks, reached only 48.9 percent at 0 dB with 560 training segments, below every feature-based model; this is a statement about data volume rather than about the architecture, and no augmentation was attempted.

VII. Comparative Analysis

Table VII compares the proposed procedure with the closest published approaches on the criteria requested for this study. The quantum baselines of [7], [8] evaluate circuits of one to ten layers without noise and connect them to a multilayer perceptron; their reported advantage over the perceptron alone is obtained on measured CWRU data and is therefore not directly comparable with the synthetic data used here. Measurement-error mitigation [21] and Clifford data regression [12] are post-processing methods that could be composed with any classifier; the contribution here is to use the Clifford regression estimate to decide depth before training, and to show numerically why the simpler single-qubit calibration is insufficient for that purpose. The noise-aware training of [14] corresponds to the VQC-NT baseline, which Table III shows to be insufficient at $p = 0.1$ for a four-layer circuit. In cost, NA-VQC adds one short noiseless probe per candidate depth (reused across noise levels) and $K \cdot n_{\text{in}} + 2$ calibration circuits; training time is lower than for the standard pipeline because the selected circuits are shallower. Against the classical models, the quantum classifier with four qubits and 52 trainable parameters at depth one matches the support vector machine on the same features (86.3 percent) and trails the best twelve-feature classical models (88.2 percent for the SVM, 86.7 percent for XGBoost) by the amount lost in the projection to four principal components. No claim of quantum advantage is made; the claim is that the deployed accuracy of the quantum model can be made insensitive to the hardware noise level.

TABLE VII. Qualitative comparison with existing approaches

Criterion	Hybrid VQC [7], [8]	Noise-aware QNN [14]	CDR / M3 mitigation [12], [21]	NA-VQC (this work)
Noise in design	None	Training only	Post-processing	Calibration → scale, depth, training, readout
Depth choice	Manual (1, 5, 10)	Fixed	n/a	Closed-form rule, validated vs brute force
Calibration circuits	None	None	Single-qubit / Clifford	Single-qubit + Clifford per depth
Drift handling	Retrain	Retrain	Re-mitigate	Refresh affine layer only
Shot budget analysed	No	Hardware runs	Yes	256 / 1024 / exact
Accuracy at p = 0.1 (this data)	28.2 (std. pipeline)	42.3 (noise-trained)	—	81.9
Training cost	O(T B L n d^2)	O(T B L n d^2)	post-processing only	O(T B L n d^2) + probe
Validation data	CWRU, JNU (measured)	Network traffic	Synthetic circuits	Physics-based synthetic (limitation)
Negative results reported	No	Partly	n/a	Yes (Section VI-F)

VIII. Key Outcomes

The study produced the following outcomes. A proposition and its numerical validation establish that the readout of an entangling circuit under local depolarizing noise contracts faster than the single-qubit model predicts, by a factor that reaches 3.8 at $p = 0.05$ and four layers, and that near-Clifford instances of the ansatz measure the true contraction. A closed-form depth rule that multiplies this contraction by a noiseless margin selects the shallowest circuit on this task; the exhaustive sweep shows that choice to be within one to four layers and at most two percentage points of the optimum at 1024 shots, and far from the depths at which accuracy collapses, at the cost of one short probe per depth that is reused across noise levels. Noise-in-the-loop training with the calibrated readout keeps the deployed accuracy of the quantum classifier within four percentage points of a support vector machine on identical features at every noise level from zero to a 10 percent per-layer error, whereas the standard pipeline drops to chance. The affine readout layer allows a deployed model to be re-calibrated after a noise drift without retraining, which is the operation a plant would actually perform between maintenance windows. For practitioners the engineering message is that the per-layer noise of the target device should be measured before the circuit depth is chosen, and that on current noise levels a one- or two-layer data re-uploading circuit with a scaled encoding is the right design for a four-feature fault classifier. The industrial relevance lies in the possibility of running the quantum part of such a pipeline as a cloud service with calibration refreshed per device session; the social impact is indirect, through earlier detection of bearing

damage in pumps, fans and traction motors, and is not claimed to exceed what classical feature-based classifiers already provide.

IX. Conclusion and Future Scope

This paper asked how a hybrid variational quantum classifier for bearing faults should be configured when the device that runs it is noisy and the measurement budget is finite. The answer developed here is a design loop that measures the noise, estimates the circuit-level readout contraction by Clifford regression, selects the encoding scale and the depth from that measurement, trains under the estimated channel and keeps an affine readout layer that can be refreshed. In simulation the loop holds the accuracy of a four-qubit classifier within four points of a support vector machine on the same features across depolarizing and amplitude-damping noise up to 10 percent per layer, where the standard noiseless-training pipeline fails, and the depth rule stays within two points of an exhaustive sweep while avoiding its failure region.

The limitations are stated plainly. The vibration data are synthetic; the noise model is local, Markovian and identical on every qubit; the register has four qubits, which is classically simulable, so no advantage over classical methods is claimed or observed; gradients during training are exact rather than shot-sampled; and the reference list requires independent verification before submission. Future work should, in order of priority, repeat the study on measured data such as the CWRU and Paderborn sets with the same feature pipeline, run the calibration and deployment stages on a superconducting or trapped-ion device to test whether the Clifford-regression contraction predicts hardware behaviour, extend the depth rule to gate-dependent and correlated noise through a weight-resolved version of Proposition 1, model shot noise in the gradient estimates so that the trainability side of the depth trade-off is captured, and scale to eight or more qubits where the contraction of the correlator readouts becomes the limiting factor. A version of the controller that selects the entangling topology as well as the depth, and one that adapts the shot budget per sample from the calibrated margin, are natural extensions of the flip-probability bound.

Acknowledgment

The authors thank the Research and Development Division of Tekstein Solutions Private Limited for computing resources. [Sponsor acknowledgments to be confirmed before submission.]

References

[1] P. D. McFadden and J. D. Smith, "Model for the vibration produced by a single point defect in a rolling element bearing," J. Sound Vib., vol. 96, no. 1, pp. 69–82, 1984.

[2] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," Mech. Syst. Signal Process., vol. 25, no. 2, pp. 485–520, 2011.

[3] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," Mech. Syst. Signal Process., vols. 64–65, pp. 100–131, 2015.

[4] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," Sensors, vol. 17, no. 2, art. 425, 2017.

[5] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," Quantum, vol. 4, art. 226, 2020.

[6] M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models," Phys. Rev. A, vol. 103, art. 032430, 2021.

[7] C. B. S. Maior, L. M. M. Araújo, I. D. Lins, M. das Chagas Moura, and E. López Droguett, "Prognostics and health management of rotating machinery via quantum machine learning," IEEE Access, vol. 11, pp. 25132–25151, 2023.

[8] L. M. M. Araújo, I. D. Lins, C. B. S. Maior, M. das Chagas Moura, and E. López Droguett, "Rotating machinery health state diagnosis through quantum machine learning," in Proc. 33rd Eur. Safety Reliab. Conf. (ESREL), Southampton, U.K., 2023.

[9] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, "Noise-induced barren plateaus in variational quantum algorithms," Nat. Commun., vol. 12, art. 6961, 2021.

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

[11] 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, art. 045005, 2023.

[12] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, "Error mitigation with Clifford quantum-circuit data," Quantum, vol. 5, art. 592, 2021.

[13] K. Sharma, S. Khatri, M. Cerezo, and P. J. Coles, "Noise resilience of variational quantum compiling," New J. Phys., vol. 22, art. 043006, 2020.

[14] A. Kukliansky, M. Orescanin, C. Bollmann, and T. Huffmire, "Network anomaly detection using quantum neural networks on noisy quantum computers," IEEE Trans. Quantum Eng., vol. 5, art. 2500111, 2024.

[15] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces," Nature, vol. 567, pp. 209–212, 2019.

[16] T. Hubregtsen, D. Wierichs, E. Gil-Fuster, P.-J. H. S. Derks, P. K. Faehrmann, and J. J. Meyer, "Training quantum embedding kernels on near-term quantum computers," Phys. Rev. A, vol. 106, art. 042431, 2022.

[17] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," IEEE Access, vol. 8, pp. 29857–29881, 2020.

[18] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," Adv. Quantum Technol., vol. 2, no. 12, art. 1900070, 2019.

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

[20] E. Fontana, N. Fitzpatrick, D. Muñoz Ramo, R. Duncan, and I. Rungger, "Evaluating the noise resilience of variational quantum algorithms," Phys. Rev. A, vol. 104, art. 022403, 2021.

[21] P. D. Nation, H. Kang, N. Sundaresan, and J. M. Gambetta, "Scalable mitigation of measurement errors on quantum computers," PRX Quantum, vol. 2, art. 040326, 2021.

[22] N. Cristianini, J. Shawe-Taylor, A. Elisseeff, and J. Kandola, "On kernel-target alignment," in Adv. Neural Inf. Process. Syst., vol. 14, 2001, pp. 367–373.

[23] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning," Phys. Rev. A, vol. 98, art. 032309, 2018.

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

[25] 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," Nat. Rev. Phys., vol. 3, pp. 625–644, 2021.

[26] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," Nat. Commun., vol. 9, art. 4812, 2018.

[27] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, "Quantum computing with Qiskit," arXiv:2405.08810, 2024.