Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Time-Symmetry-Inspired Quantum Machine Learning: A Framework for Echo-Protected Kernel Estimation

S. M. Ashik Mahmud

ashikmahmud@iut-dhaka.edu

Islamic University of Technology

Research Article

Keywords:

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11630256/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Time-Symmetry-Inspired Quantum Machine Learning: A Framework for Echo-Protected Kernel Estimation

S. M. Ashik Mahmud[1]

Department of Electrical and Electronic Engineering

[1]Islamic University of Technology, Board Bazar, Gazipur 1704, Bangladesh

Abstract

A quantum kernel is useful only if the similarities it encodes survive the process used to measure them. In this work, I propose a time-symmetry-inspired framework for investigating echo-protected quantum kernel estimation on noisy quantum processors. I adopt the established architecture of quantum feature encoding followed by classical kernel-based learning and examine whether an explicitly defined echo intervention can improve similarity estimates without altering their intended meaning. I distinguish the inverse operations of a conventional overlap circuit from physical refocusing and verification: reversing an encoding sequence does not, by itself, reverse the noise accumulated during its execution. This distinction establishes the basis for assessing protection through its mechanism and estimator rather than through circuit symmetry alone. I organise the evaluation around kernel-estimation error, matrix reconstruction, and classification on unseen data, with explicit accounting for sampling uncertainty, postselection overhead, and classical postprocessing. I further separate improved implementation fidelity from the statistical distinguishability of kernel values and from the suitability of the representation for a learning task. The resulting framework defines the evidence needed to establish an echo-based benefit against unprotected estimation and existing mitigation methods under comparable resource budgets. Its contribution is a physically grounded formulation of the protection problem and a testable evaluation structure; the effectiveness, scalability, and predictive value of a particular protocol remain to be demonstrated.

1 Introduction

Quantum kernel learning depends on an experimentally demanding premise: the similarity assigned to two inputs must remain meaningful after quantum preparation, noisy evolution, and finite measurement. A feature map may define a compelling representation mathematically, yet its measured kernel can depart substantially from that representation. I take this gap between the intended similarity and the experimentally accessible estimate as the starting point of this work.

The connection between quantum feature spaces and classical kernel methods provides an established foundation for addressing this problem. Schuld and Killoran [1] formulated quantum data encoding within a feature-space interpretation, while Havlíček et al. [2] demonstrated a supervised-learning architecture in which a quantum processor estimates kernel entries for subsequent classical processing. I adopt this hybrid structure and focus on the reliability of the quantum estimation stage. My central question is whether an echo-based intervention can improve that reliability while preserving the target kernel and providing a defensible return on the additional experimental effort.

Echo phenomena offer a physical motivation for this investigation, but their relevance requires care. Hahn's spin-echo experiment [3] established that appropriately controlled dynamics can refocus phase dispersion. Dynamical decoupling subsequently extended the use of control sequences to suppress selected system–environment interactions under suitable conditions [4]. I draw on these principles as motivation for protecting kernel estimation. Their application, however, depends on the error mechanism, the available control, and the timing of the intervention. The presence of forward and inverse operations is therefore insufficient to establish noise suppression.

This distinction matters particularly for fidelity-kernel estimation. A conventional compute–uncompute circuit already combines an encoding operation with an inverse encoding to estimate the overlap between quantum feature states [2]. Those inverse gates undo the intended transformation in the ideal circuit; they do not generally undo the physical noise encountered during execution. I consequently use the term "time-symmetry-inspired" to describe a design principle whose operational consequences must be established explicitly. Within this formulation, I distinguish data encoding, similarity measurement, and protection as separate functions, even when they appear within the same circuit.

Verification introduces a further distinction between protecting dynamics and modifying an estimator. Verified phase estimation [5] and adiabatic echo verification [6] provide precedents for extracting improved estimates through protocol-specific verification conditions. I regard these approaches as motivation to examine which measurement contributions should be retained and how their retention affects the quantity being estimated. If postselection forms part of an echo-protected kernel protocol, the conditional measurement statistics must be connected explicitly to the intended kernel. Any reduction in error must also be considered alongside the probability of acceptance and the executions required to obtain adequate precision. Echo-verification research identifies this resource dependence as a practical concern [8].

The relevant comparison extends beyond an unprotected quantum circuit. Research on quantum embedding kernels has already examined mitigation, matrix quality, and classical regularization [7]. More recent work has combined kernel-specific processing with physical noise-suppression techniques [10]. I therefore locate the potential contribution of the proposed framework in the precise intervention, its estimator, and its measurable consequences. Establishing an additional benefit requires controlled comparisons that separate physical protection from normalization, matrix correction, and other processing choices.

Even an accurate reconstruction of the intended kernel leaves an important learning question unresolved. Quantum kernel values can become statistically difficult to distinguish under the conditions identified by concentration analyses [9]. Moreover, a representation must capture structure relevant to the prediction task, as illustrated by work on covariant quantum kernels [11]. I accordingly distinguish three objectives: estimating the intended similarities accurately, resolving meaningful differences between inputs, and improving predictions on unseen data. Success at one stage does not establish success at the others. Likewise, an improvement over a noisy quantum baseline does not demonstrate an advantage over competitive classical learning methods [12].

In this work, I formulate an evaluation framework that follows these distinctions from circuit design to predictive performance. I connect the conceptual architecture and physical estimator to noise-dependent estimation error, kernel-matrix reconstruction, and independently evaluated classification. Throughout this sequence, I require transparent accounting for sampling, verification, and postprocessing costs. The aim is to make the proposed protection claim experimentally decidable: under which noise conditions, for which representations, and at what resource cost does an echo intervention improve the kernel estimates that learning actually requires?

2.1 Conceptual architecture of the echo-protected quantum kernel framework: Explain the overall method

I situate the proposed framework within the established connection between quantum feature representations and kernel-based learning. Schuld and Killoran [1] developed the theoretical basis for interpreting quantum data encoding as a feature map, while Havlíček et al. [2] demonstrated quantum kernel estimation integrated with classical supervised learning. Together, these studies establish the hybrid architecture underlying the proposal: a quantum processor estimates similarity between encoded inputs, and a classical algorithm uses those similarities for prediction. I adopt this architecture as an established foundation and identify the reliability of its experimentally estimated similarities as the problem motivating protection.

The physical motivation derives from coherent-control research. Hahn [3] demonstrated that appropriately timed pulses can refocus phase dispersion in spin ensembles. Viola, Knill, and Lloyd [4] subsequently developed dynamical decoupling to suppress unwanted system–environment interactions under suitable control conditions. These contributions establish that selected error mechanisms can be reduced through controlled dynamics; they do not establish that reversing a computational sequence reverses every physical process encountered during its execution. I therefore distinguish an echo's operational mechanism from the superficial symmetry of a circuit diagram.

Within this context, I frame Figure 1 around a specific question: whether a defined echo-based intervention can improve quantum kernel estimation while preserving the intended similarity measure and maintaining a favourable resource trade-off. The potential contribution resides in the intervention, its physical justification, and its demonstrated consequences. Combining established encoding, control, and classification components is insufficient, by itself, to establish methodological novelty.

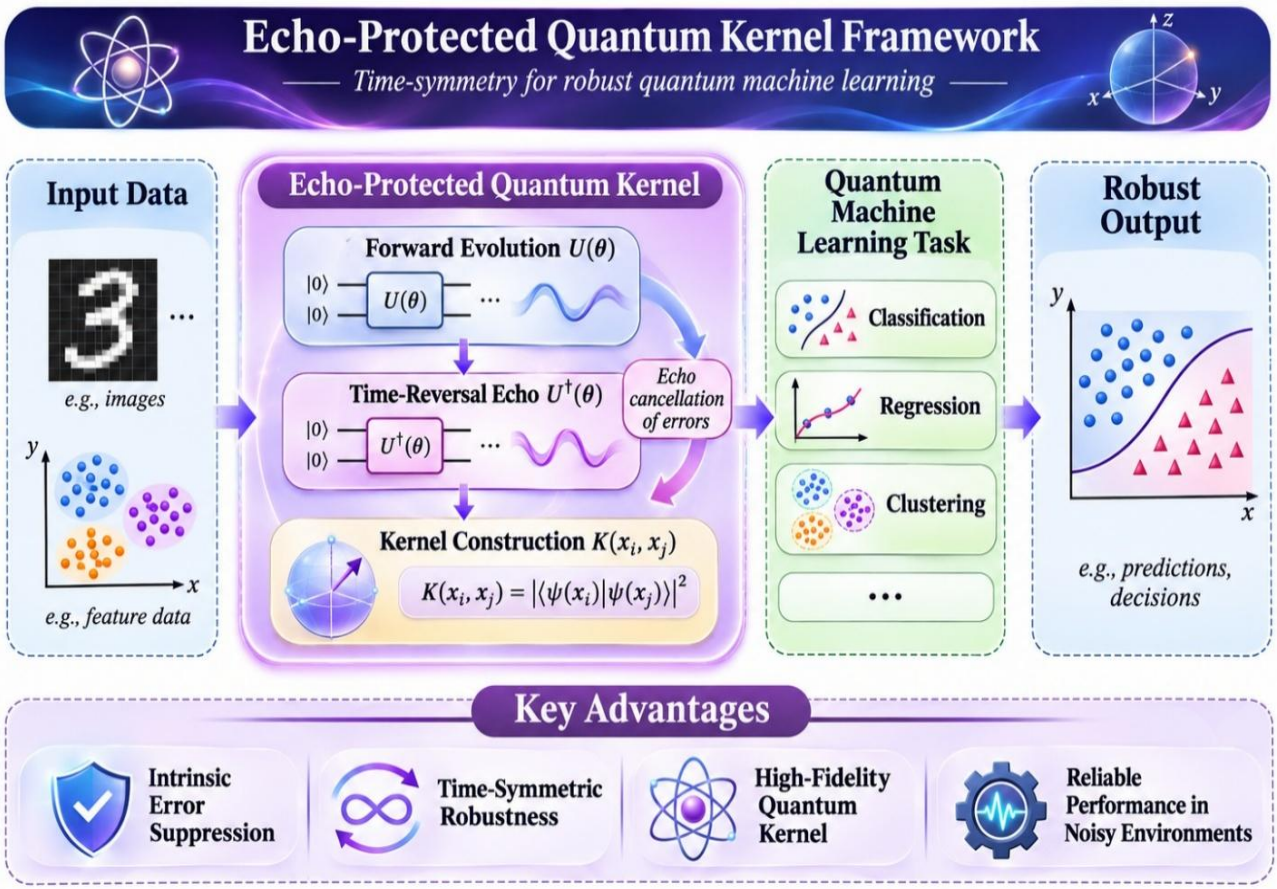


Figure 1. Conceptual architecture of the echo-protected quantum kernel framework. The method integrates time-symmetric quantum evolution with kernel-based machine learning. Input data are encoded into a quantum state, processed through forward evolution and a time-reversal echo to suppress noise-induced errors, and then used to construct a robust quantum kernel for downstream learning tasks such as classification, regression, and clustering. The framework improves reliability in noisy quantum environments.

2.2 From quantum encoding to a valid kernel

2.2.1 Feature states and the target similarity

For an input $x \in \mathcal{X}$, I choose a normalized reference state $|0^n\rangle$ and a unitary feature map $U_\phi(x)$. I define

$$|\psi(x)\rangle = U_\phi(x)|0^n\rangle, \qquad \rho(x) = |\psi(x)\rangle\langle\psi(x)|. \qquad (1)$$

The fidelity kernel follows by taking the Hilbert–Schmidt inner product of the density operators:

$$\begin{aligned} K(x, x') &= \mathrm{Tr}[\rho(x)\rho(x')] \\ &= \langle\psi(x')|\psi(x)\rangle\langle\psi(x)|\psi(x')\rangle \\ &= |\langle\psi(x')|\psi(x)\rangle|^2. \qquad (2) \end{aligned}$$

Thus, the effective feature vector is the density operator, or its vectorization, rather than the state vector alone. This is the standard feature-space interpretation of fidelity-based quantum learning [1,2]. Normalization and the Cauchy–Schwarz inequality give

$$0 \le K(x,x') \le 1, \qquad K(x,x) = 1, \qquad K(x,x') = K(x',x). \qquad (3)$$

I verify positive semi definiteness directly. For arbitrary complex coefficients $c_1, \dots, c_N$, let $A = \sum_i c_i \rho(x_i)$. Then

$$\sum_{i,j} c_i^* K_{ij} c_j = \mathrm{Tr}(A^\dagger A) = \| A \|_F^2 \ge 0. \qquad (4)$$

Equation (4) establishes that the ideal Gram matrix is admissible for conventional kernel methods. A noisy matrix assembled from separately measured overlaps need not inherit this property.

2.2.2 The overlap measurement

I apply the pair-dependent circuit $V_{ij} = U_\phi(x_j)^\dagger U_\phi(x_i)$ and measure the projector $\Pi_0 = |0^n\rangle\langle 0^n|$. The Born rule gives

$$\begin{aligned} p_{ij}(0^n) &= \mathrm{Tr}\left(\Pi_0 V_{ij} \Pi_0 V_{ij}^\dagger\right) \\ &= |\langle 0^n | U_\phi(x_j)^\dagger U_\phi(x_i) | 0^n \rangle|^2 \\ &= K_{ij}. \end{aligned} \qquad (5)$$

The two input arguments are essential. If I use the same input in both halves, $U_\phi(x_i)^\dagger U_\phi(x_i) = I$, and the ideal return probability is one. Such a circuit measures self-return, not the complete pairwise kernel. I therefore interpret the central block in Figure 1 as a family of circuits indexed by (x_i, x_j).

2.2.3 What protection must preserve

Let $\mathcal{C}_{ij}^{(a)}$ denote the implemented channel for protocol a, where $a = \mathrm{F}$ means unprotected evolution and $a = \mathrm{E}$ means echo protection. I define

$$q_{ij}^{(a)} = \mathrm{Tr}\left[\Pi_0 \mathcal{C}_{ij}^{(a)}(\Pi_0)\right], \qquad B_{ij}^{(a)} = q_{ij}^{(a)} - K_{ij}. \qquad (6)$$

A necessary target-preservation condition is

$$\mathcal{C}_{ij}^{(\mathrm{E})}|_{\text{zero noise}}(\Pi_0) = V_{ij} \Pi_0 V_{ij}^\dagger. \qquad (7)$$

I assess protection by comparing errors relative to the same K_{ij}. A protocol that changes the feature map may define a useful new kernel, but its performance is a representation change rather than faithful reconstruction of the original similarity.

For supervised prediction, the resulting architecture has the form

$$f(x) = \sum_{i=1}^{N} a_i \, K(x_i, x) + b, \qquad \hat{\ell}(x) = \operatorname{sign} f(x). \qquad (8)$$

The quantum processor supplies similarities; the classical training procedure determines a_i and b. I derive that training step in Section 5. Equations (1) – (8) support the architecture of Figure 1, but do not establish its advertised performance advantages.

3.1 Quantum circuit diagram showing forward encoding, inverse operations, and echo measurements: Define the algorithm physically

I distinguish the reversible structure of an overlap estimator from the mechanism responsible for error suppression. Conventional fidelity-kernel estimation can encode one input and subsequently apply the inverse encoding associated with another. In the ideal implementation, the probability of measuring the reference computational state determines the similarity between the encoded states [2]. This compute–uncompute procedure is already part of quantum kernel methodology. Its inverse operations reverse the intended encoding but do not generally invert the noise accumulated during execution.

Verification-based mitigation offers a related, operationally distinct precedent. O'Brien et al. [5] introduced verified phase estimation, using a verification condition on the system register to exclude certain erroneous contributions to observable estimates. Schiffer et al. [6] developed adiabatic echo verification to address imperfect adiabatic preparation and hardware noise. I interpret these results as support for particular estimator–protocol combinations under stated assumptions, rather than a universal cancellation principle associated with forward and backward evolution.

For Figure 2, I distinguish three functions: encoding the data, measuring their similarity, and protecting the estimate. Refocusing acts on unwanted dynamics during execution, whereas verification modifies the measurement contributions retained in an estimator. If the proposed framework employs postselection, its conditional statistics require an explicit justification as an estimate of the intended kernel. I consequently use "time-symmetry-inspired" to describe a design motivation, without implying a literal reversal of dissipative dynamics. This distinction prevents the protection claim from being inferred solely from the presence of inverse gates.

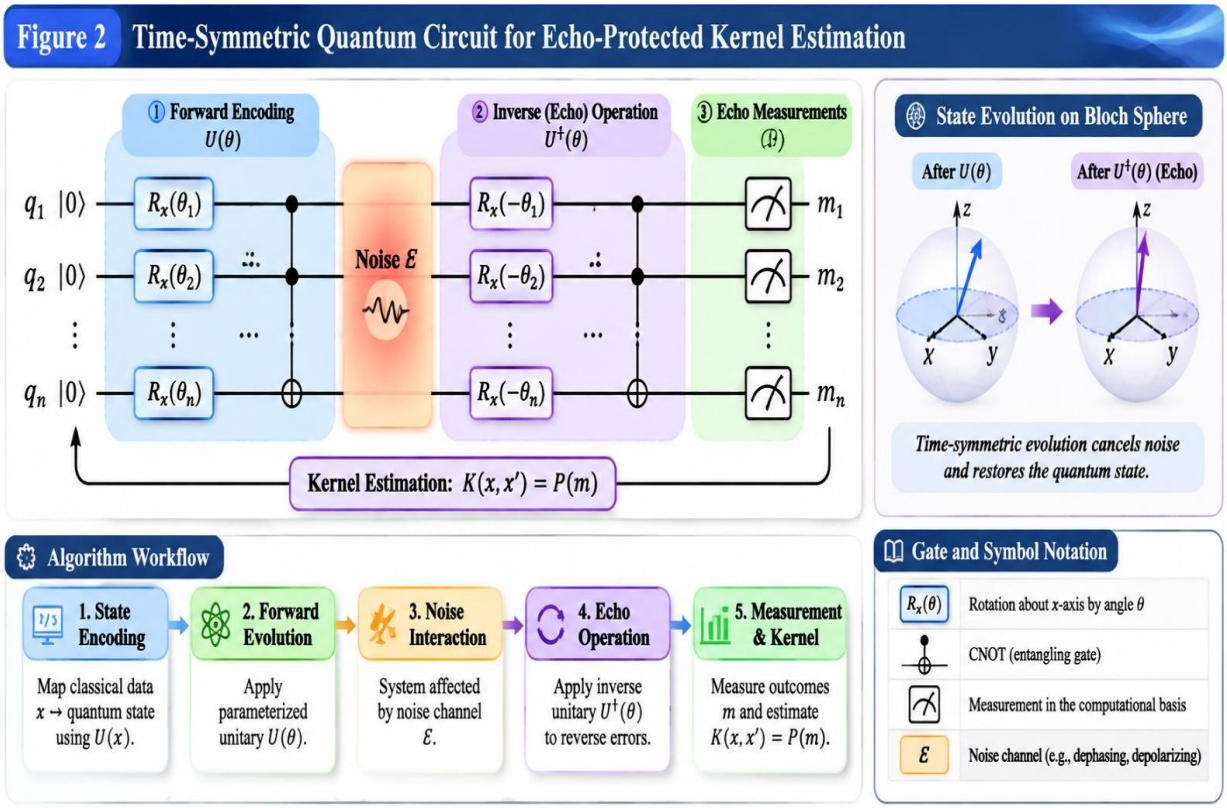


Figure 2. Quantum circuit diagram showing forward encoding, inverse (echo) operations, and echo measurements. The circuit implements a time-symmetric evolution $U^\dagger(\theta)U(\theta)$ to cancel noise, and the kernel value $K(x, x')$ is estimated from the measurement outcomes m.

3.2 A physical echo and its measurement law

3.2.1 Why inversion alone does not remove noise

Suppose a channel $\mathcal{N}$ acts after encoding and before ideal decoding. Cyclicity of the trace gives

$$q_{ij} = \mathrm{Tr}\left[\Pi_0 U_j^\dagger \mathcal{N}(\rho_i) U_j\right]$$
$$= \mathrm{Tr}\left[\rho_j \mathcal{N}(\rho_i)\right], \qquad (9)$$

where $U_i = U_\phi(x_i)$ and $\rho_i = \rho(x_i)$. The inverse gate does not remove $\mathcal{N}$. As an explicit counterexample, I take a depolarizing channel,

$$\mathcal{D}_p(\rho) = (1-p)\rho + pI/d. \qquad (10)$$

Substitution into (9) yields

$$q_{ij} = (1-p)K_{ij} + p/d, \qquad q_{ii} = 1 - p + p/d < 1 \quad (p > 0,\ d > 1). \qquad (11)$$

Thus, even a perfect inverse circuit fails to restore the self-overlap in this example. I cannot derive universal noise cancellation from $U^\dagger U = I$.

Gate order also matters. If $U = G_L \cdots G_2 G_1$, then $U^\dagger = G_1^\dagger G_2^\dagger \cdots G_L^\dagger$. The inverse is executed by undoing the last gate first. Reversing rotation angles while retaining the original ordering of non-commuting entanglers and rotations does not generally construct $U^\dagger$.

3.2.2 A feature map with an exact dephasing solution

For a concrete realization, I choose an equatorial single-qubit state,

$$U_\phi(x) = R_z\big(\phi(x)\big)H, \qquad |\psi(x)\rangle \simeq \frac{|0\rangle + e^{i\phi(x)}|1\rangle}{\sqrt{2}}. \qquad (12)$$

The symbol $\simeq$ denotes equality up to an irrelevant global phase. With $\Delta_{ij} = \phi(x_i) - \phi\big(x_j\big)$, direct multiplication gives

$$K_{ij} = \frac{\left|1 + e^{i\Delta_{ij}}\right|^2}{4} = \frac{1 + \cos\Delta_{ij}}{2} = \cos^2\big(\Delta_{ij}/2\big). \qquad (13)$$

For n product-encoded qubits, I take $U_\phi(x) = \bigotimes_{k=1}^{n} R_z\big(\phi_k(x)\big)H$ and obtain

$$K_{ij} = \prod_{k=1}^{n} \frac{1 + \cos\Delta_{ij,k}}{2}. \qquad (14)$$

This deliberately simple kernel is efficiently computable classically. I use it to establish the protection mechanism analytically, not to claim quantum computational advantage.

3.2.3 Refocusing in the toggling frame

I set $\hbar = 1$ and model dephasing during an otherwise idle interval by

$$H_{\mathrm{n}}(t) = 1/2\,\beta(t)Z. \qquad (15)$$

Without pulses, the error is a random Z rotation. For echo protection, I apply an ideal X pulse at $T/2$ and a second X at T to restore the original control frame. Let $\Phi_1 = \int_0^{T/2} \beta\,(t)dt$ and $\Phi_2 = \int_{T/2}^{T} \beta\,(t)dt$. Since $XZX = -Z$, the complete idle block is

$$Xe^{-iZ\Phi_2/2}Xe^{-iZ\Phi_1/2} = e^{-iZ(\Phi_1 - \Phi_2)/2}. \qquad (16)$$

I can absorb the final frame restoration into a correctly adjusted decoder, but I cannot simply omit it. In the zero-noise limit the block in (16) equals the identity, so the complete circuit $U_j^\dagger$[echo block]U_i satisfies (7).

More generally, I define the accumulated stochastic phase by

$$\Phi_y = \int_0^T y\,(t)\beta(t)dt, \qquad y_{\mathrm{F}}(t) = 1, \qquad y_{\mathrm{E}}(t) = \begin{cases} 1, & 0 \le t < T/2, \\ -1, & T/2 \le t \le T. \end{cases} \qquad (17)$$

For quasi-static noise, $\beta(t) = \beta_0$ within each execution, (17) gives $\Phi_{\mathrm{F}} = \beta_0 T$ and $\Phi_{\mathrm{E}} = 0$. The echo cancels this particular phase error exactly under the ideal-pulse assumptions.

3.2.4 Gaussian noise and the filter function

I now assume zero-mean stationary Gaussian noise with correlation $C(\tau) = \mathbb{E}[\beta(t+\tau)\beta(t)]$. The random variable Φ_y is Gaussian, so its characteristic function is

$$W_y(T) = \mathbb{E}\left[e^{-i\Phi_y}\right] = e^{-\chi_y(T)}, \qquad \chi_y = \frac{1}{2}\mathbb{E}\left[\Phi_y^2\right]. \qquad (18)$$

Inserting (17) gives the exact time-domain expression

$$\chi_y = \frac{1}{2}\int_0^T d\,t \int_0^T d\,s y(t)y(s)C(t-s). \qquad (19)$$

I define a two-sided spectrum $S(\omega) = \int_{-\infty}^{\infty} C\,(\tau)e^{-i\omega\tau}d\tau$ and $Y_y(\omega) = \int_0^T y\,(t)e^{i\omega t}dt$. Fourier inversion then yields

$$\chi_y = \frac{1}{4\pi}\int_{-\infty}^{\infty} S\,(\omega)\left|Y_y(\omega)\right|^2 d\omega. \qquad (20)$$

This explicitly stated convention avoids ambiguities between one-sided and two-sided spectra. The filter-function description of Gaussian dephasing provides the physical basis for the calculation [13]. Performing the elementary integrals in Y_y gives

$$|Y_{\mathrm{F}}(\omega)|^2 = \frac{4\sin^2(\omega T/2)}{\omega^2}, \qquad |Y_{\mathrm{E}}(\omega)|^2 = \frac{16\sin^4(\omega T/4)}{\omega^2}. \qquad (21)$$

Near zero frequency, the first expression approaches T^2, whereas the second is $\omega^2 T^4/16 + O(\omega^4 T^6)$. This explains the suppression of low-frequency dephasing. It does not imply suppression at every frequency: a noise spectrum concentrated where the echo filter is larger can produce the opposite comparison.

3.2.5 Exact return probability

For one encoded qubit, averaging over the stochastic phase produces

$$\overline{\rho}_i^{(y)} = \frac{1}{2}\begin{pmatrix} 1 & e^{-i\phi_i}e^{-\chi_y} \\ e^{i\phi_i}e^{-\chi_y} & 1 \end{pmatrix}. \qquad (22)$$

Taking the trace with ρ_j as in (9), I obtain

$$q_{ij}^{(y)} = \frac{1 + e^{-\chi_y}\cos\Delta_{ij}}{2}. \qquad (23)$$

For independent local noise and product encoding, the all-zero return probability factorizes:

$$q_{ij}^{(y)} = \prod_{k=1}^{n}\frac{1 + e^{-\chi_{y,k}}\cos\Delta_{ij,k}}{2}. \qquad (24)$$

Equation (24) does not apply to correlated noise or an arbitrary entangled feature map. For correlated classical phases, I instead evaluate

$$q_{ij}^{(y)} = \mathbb{E}\left[\prod_{k=1}^{n} \frac{1+\cos\left(\Delta_{ij,k}+\Phi_{y,k}\right)}{2}\right], \qquad (25)$$

which generally cannot be replaced by a product of expectations. For entangled encodings, equation (9) remains the starting point.

I therefore read Figure 2 as requiring three explicit additions: separate inputs in the forward and inverse maps, an actual refocusing operation distinct from inversion, and the measurement event 0^n that defines the kernel probability. During noncommuting encoding gates, the toggling-frame noise is $U_c(t)^\dagger H_{\mathrm{n}}(t) U_c(t)$; it is not generally the scalar sign modulation in (17). A dynamically corrected gate construction would be needed to extend this exact idle-interval result.

4.1 Kernel estimation error versus noise strength: Demonstrate noise reduction

I interpret noise-dependent kernel error as the combined consequence of imperfect implementation and finite measurement statistics. Hubregtsen et al. [7] investigated these effects in quantum embedding kernels, evaluated mitigation strategies, and demonstrated improved kernel-matrix quality in their studied setting. Their work establishes a direct methodological baseline: an echo-based proposal must be distinguished from existing kernel-specific mitigation, rather than evaluated exclusively against an unprotected circuit.

Recent echo-verification research also exposes an important trade-off. Weaving et al. [8] combined echo verification with Clifford data regression to address residual noise dependence in observable estimation. Their analysis identified the sensitivity of postselection probability to noise as a potential bottleneck. I therefore distinguish improved estimates among accepted outcomes from improved performance per unit of experimental effort. Verification can reduce particular errors while increasing the number of executions required to obtain useful statistics. The demonstrated quantum-simulation benefits motivate this analysis but do not establish equivalent benefits for kernel estimation.

Thanasilp et al. [9] identified a complementary limitation: under specified conditions, quantum kernel values become exponentially concentrated and cannot be resolved adequately with feasible measurement budgets. I interpret this result as separating correction of implementation errors from recovery of statistically distinguishable similarities. Figure 3 accordingly concerns accuracy under specified noise and resource conditions, including residual bias, sampling uncertainty, and verification overhead. A favorable error curve at a fixed system size does not, on its own, establish scalable protection or resolution of the concentration problem.

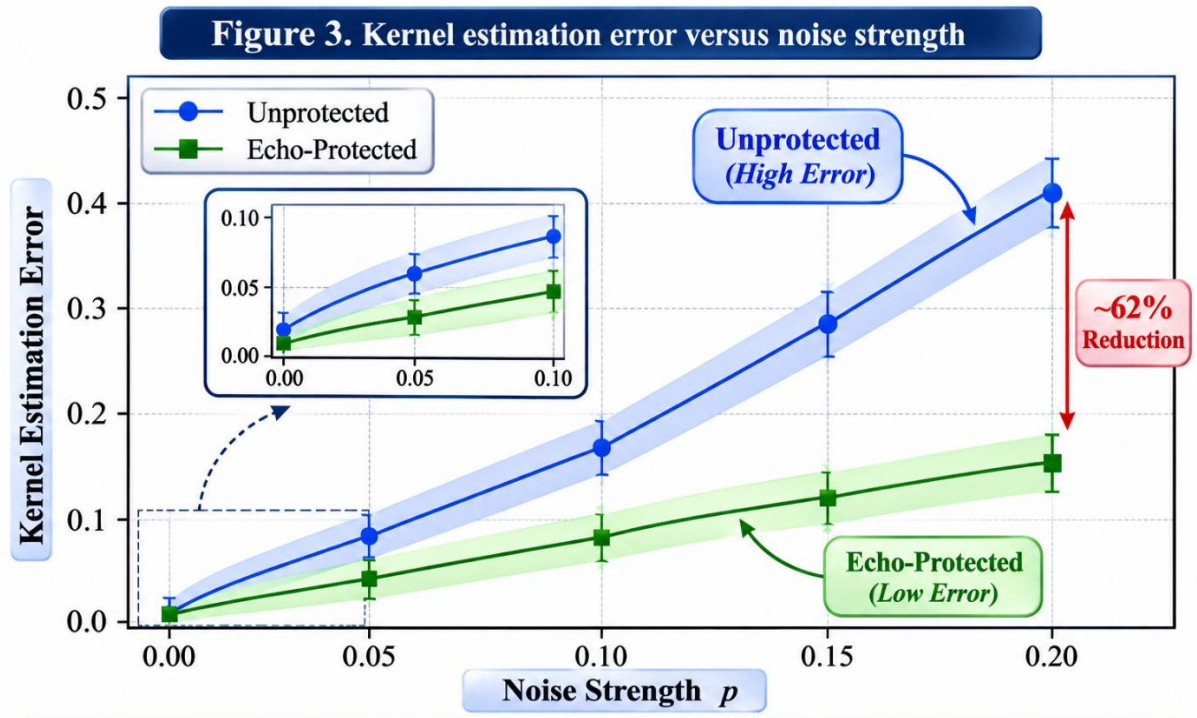


Figure 3. Kernel estimation error versus noise strength. The echo-protected protocol significantly reduces the estimation error compared with the unprotected implementation, illustrating improved robustness under increasing noise.

4.2 Kernel error as a function of noise

4.2.1 A noise model that generates an explicit curve

To obtain a closed noise-dependent expression, I choose an exponential correlation function,

$$C(\tau) = \sigma^2 e^{-|\tau|/\tau_c}, \qquad S(\omega) = \frac{2\sigma^2\tau_c}{1+\omega^2\tau_c^2}. \qquad (26)$$

Here σ is an angular-frequency standard deviation and τ_c is the correlation time. I use $\nu = \sigma T$ if a dimensionless noise axis is desired. A plotted parameter called "noise strength" must specify whether it means σ, ν, a depolarizing probability, or another quantity.

For free evolution, integrating (19) over the two triangular halves of the square gives

$$\chi_{\mathrm{F}} = \sigma^2 \int_0^T d\,t \int_0^t d\,s e^{-(t-s)/\tau_c}$$
$$= \sigma^2\tau_c\left[T - \tau_c\left(1 - e^{-T/\tau_c}\right)\right]. \qquad (27)$$

For the echo, I set $L = T/2$. The same-half and opposite-half integrals are

$$I(L) = \int_0^L d\,t \int_0^L d\,s e^{-|t-s|/\tau_c} = 2\tau_c L - 2\tau_c^2\left(1 - e^{-L/\tau_c}\right),$$
$$J(L) = \int_0^L d\,t \int_L^{2L} d\,s e^{-(s-t)/\tau_c} = \tau_c^2\left(1 - e^{-L/\tau_c}\right)^2. \qquad (28)$$

Because opposite halves enter with a minus sign, $\chi_{\mathrm{E}} = \sigma^2[I(L) - J(L)]$. Consequently,

$$\chi_{\mathrm{E}} = \sigma^2\tau_c^2\left[\frac{T}{\tau_c} - 3 + 4e^{-T/(2\tau_c)} - e^{-T/\tau_c}\right]. \qquad (29)$$

Subtracting (29) from (27) gives a useful exact comparison:

$$\chi_{\mathrm{F}} - \chi_{\mathrm{E}} = 2\sigma^2\tau_c^2\left(1 - e^{-T/(2\tau_c)}\right)^2 \geq 0. \qquad (30)$$

For $T \ll \tau_c$, expansion of the exponentials yields

$$\chi_{\mathrm{F}} = \frac{\sigma^2 T^2}{2} + O(\sigma^2 T^3/\tau_c), \qquad \chi_{\mathrm{E}} = \frac{\sigma^2 T^3}{12\tau_c} + O(\sigma^2 T^4/\tau_c^2). \qquad (31)$$

The echo thus changes the leading short-time dependence in this model. In the white-noise limit, with $D = \sigma^2\tau_c$ held fixed as $\tau_c \to 0$, both attenuation exponents approach DT. The ideal echo then provides no dephasing benefit.

4.2.2 Bias and the cost of imperfect control

Combining (13) and (23), I obtain the exact single-qubit bias,

$$B_{ij}^{(y)} = -\frac{1 - e^{-\chi_y}}{2}\cos\Delta_{ij}, \qquad \left|B_{ij}^{(y)}\right| \leq \frac{1 - e^{-\chi_y}}{2}. \qquad (32)$$

For ideal pulses under (26), equation (30) implies $\left|B_{ij}^{(\mathrm{E})}\right| \leq \left|B_{ij}^{(\mathrm{F})}\right|$. At $\cos\Delta_{ij} = 0$, both biases vanish, so no strict improvement is available for that pair.

To include a simple pulse-error cost, I model each of the two echo pulses as an ideal X followed by a single-qubit depolarizing channel with parameter p_π. Depolarization commutes with these rotations and contracts the equatorial Bloch vector. I therefore replace the echo visibility by

$$\lambda_{\mathrm{E}} = (1 - p_\pi)^2 e^{-\chi_{\mathrm{E}}}, \qquad \lambda_{\mathrm{F}} = e^{-\chi_{\mathrm{F}}}. \qquad (33)$$

For nonzero $\cos\Delta_{ij}$, strict bias reduction requires

$$\lambda_{\mathrm{E}} > \lambda_{\mathrm{F}} \quad \Leftrightarrow \quad \chi_{\mathrm{F}} - \chi_{\mathrm{E}} > -2\log(1 - p_\pi). \qquad (34)$$

This inequality gives a concrete threshold: the coherence recovered by refocusing must exceed the coherence lost to the added control. Finite pulse durations or additional waiting time must also be included in the actual noise evolution; (33) represents only the stated instantaneous-pulse error model.

4.2.3 Finite measurement statistics

For independent executions, let $Z_r = 1$ if execution r returns 0^n and zero otherwise. I estimate the raw kernel by

$$\widehat{K}_{ij} = \frac{1}{M}\sum_{r=1}^{M} Z_r\,, \qquad \mathbb{E}\big[\widehat{K}_{ij}\big] = q_{ij}, \qquad \mathrm{Var}\big(\widehat{K}_{ij}\big) = \frac{q_{ij}\big(1-q_{ij}\big)}{M}. \qquad (35)$$

Expanding $\widehat{K}_{ij} - K_{ij} = \big(\widehat{K}_{ij} - q_{ij}\big) + B_{ij}$ and taking expectations eliminates the cross term. The mean-squared error is therefore

$$\mathrm{MSE}_{ij} = B_{ij}^2 + \frac{q_{ij}\big(1-q_{ij}\big)}{M}. \qquad (36)$$

For the single-qubit model, write $c_{ij} = \cos\Delta_{ij}$ and $q = \big(1+\lambda c_{ij}\big)/2$. Equation (36) becomes

$$\mathrm{MSE}_{ij}(\lambda, M) = \frac{(1-\lambda)^2 c_{ij}^2}{4} + \frac{1-\lambda^2 c_{ij}^2}{4M}. \qquad (37)$$

For fixed $M \geq 1$ and $0 \leq \lambda \leq 1$, differentiation gives

$$\frac{\partial \mathrm{MSE}}{\partial \lambda} = -\frac{c_{ij}^2}{2}\left[(1-\lambda) + \frac{\lambda}{M}\right] \leq 0. \qquad (38)$$

Thus, visibility improvement also improves mean-squared error at equal shots in this particular model. I do not assume this monotonicity for arbitrary kernels, noise channels, or estimators.

For a set $\mathcal{P}$ of evaluated input pairs, the appropriate theoretical root-mean-square curve is

$$\mathrm{RMSE}_{\mathrm{th}}(\sigma) = \left[\frac{1}{|\mathcal{P}|}\sum_{(i,j)\in\mathcal{P}} \mathrm{MSE}_{ij}\,(\sigma)\right]^{1/2}. \qquad (39)$$

This is the square root of the expected squared error, not the expectation of a finite-sample RMSE. I would generate Figure 3 from (24), (27), (29), and (36), with the chosen features, shot counts, and pulse-error parameters stated. For a product kernel, reducing every local attenuation improves an error bound but does not by itself prove smaller absolute error for every entry, because local biases can have opposite signs.

At fixed experimental budget $\mathcal{B}$, I set $M_a = \lfloor \mathcal{B}/c_a \rfloor$, where c_a is the cost per execution for protocol a. The exact entrywise criterion is

$$\Big(B_{ij}^{(\mathrm{F})}\Big)^2 - \Big(B_{ij}^{(\mathrm{E})}\Big)^2 > \frac{q_{ij}^{(\mathrm{E})}\Big(1-q_{ij}^{(\mathrm{E})}\Big)}{M_{\mathrm{E}}} - \frac{q_{ij}^{(\mathrm{F})}\Big(1-q_{ij}^{(\mathrm{F})}\Big)}{M_{\mathrm{F}}}. \qquad (40)$$

I use (40), rather than equal accepted-shot counts alone, to determine whether the added control is worthwhile.

4.2.4 Verification and readout are separate estimators

The refocusing protocol above uses every execution. If I add verification, I must define a distinct acceptance event A. Let $V = \mathbf{1}_A$, let Z represent the kernel outcome, and define $a = \mathbb{E}[V] > 0$ and $b = \mathbb{E}[VZ]$. The conditional estimator targets

$$r = \mathbb{P}(Z = 1 \mid A) = b/a, \qquad \hat{r} = \frac{\sum_s V_s Z_s}{\sum_s V_s}. \qquad (41)$$

For this ratio to estimate the intended kernel without selection bias, I require $b = aK_{ij}$. Otherwise its asymptotic bias is $r - K_{ij}$. In particular, accepting only the all-zero kernel outcome would force $r = 1$ and destroy the similarity information. Echo-verification studies motivate careful accounting for acceptance, but do not supply this missing identity for an arbitrary kernel protocol [8].

With independent executions, the accepted count M_A is binomial with parameters (M, a). Conditional on $M_A = m > 0$, the estimator has mean r and variance $r(1-r)/m$. Hence

$$\Pr(M_A = 0) = (1-a)^M,$$

$$\mathrm{MSE}(\hat{r} \mid M_A > 0) = \left(r - K_{ij}\right)^2 + r(1-r)\mathbb{E}[M_A^{-1} \mid M_A > 0]$$

$$\simeq \left(r - K_{ij}\right)^2 + \frac{r(1-r)}{Ma} \quad (Ma \gg 1). \qquad (42)$$

The last line is an asymptotic approximation. The estimator is undefined when no outcome is accepted; I would report that failure rather than silently assign zero.

For a calibrated binary readout model with false-negative rate e_{10} and false-positive rate e_{01}, I obtain $q_{\mathrm{obs}} = e_{01} + sq$, where $s = 1 - e_{10} - e_{01} \neq 0$. The corrected estimator and its variance are

$$\hat{q}_{\mathrm{corr}} = \frac{\hat{q}_{\mathrm{obs}} - e_{01}}{s}, \qquad \mathrm{Var}(\hat{q}_{\mathrm{corr}}) = \frac{q_{\mathrm{obs}}(1 - q_{\mathrm{obs}})}{Ms^2}. \qquad (43)$$

Here I treat calibration parameters as known. Their uncertainty adds error, and clipping corrected values into $[0,1]$ introduces another bias. For a multi-qubit all-zero event, a constant two-rate model needs justification; a full assignment matrix may be required. If executions are correlated, the variance in (35) becomes $M^{-2}\sum_{r,s}\mathrm{Cov}\,(Z_r, Z_s)$, which can be much larger than the binomial value.

5.1 Kernel matrix heatmaps: Ideal, noisy, and echo-protected visually show preservation of kernel structure

I regard kernel-matrix heatmaps as a diagnostic of reconstruction quality, with predictive significance that must be assessed separately. Hubregtsen et al. [7] examined kernel alignment, mitigation, and matrix regularization, establishing that both quantum estimation and classical processing affect the final matrix. Consequently, apparent structural recovery can arise partly

from normalization, symmetrization, or positive-semidefinite correction. I distinguish these contributions from improvements produced by the physical protection mechanism.

Agliardi et al. [10] provide a closer contemporary comparison through bit-flip tolerance, a mitigation approach that modifies the treatment of measurement outcomes for fidelity kernels. Their experiments combined this procedure with dynamical decoupling and measurement twirling, and their discussion acknowledged problem-dependent calibration and limitations of the underlying assumptions. I regard this work as evidence that combining physical noise suppression with kernel-specific processing is already established experimentally. The proposed framework must therefore be differentiated through its particular mechanism and estimator, rather than through this combination alone.

I distinguish three questions when interpreting Figure 4: whether the estimated matrix approaches the intended ideal matrix, whether it satisfies the properties required by the learning method, and whether its similarities remain informative. These questions are related but not equivalent. The concentration analysis of Thanasilp et al. [9] shows why an apparently regular matrix or pronounced diagonal can coexist with poor distinguishability between inputs. I therefore interpret visual agreement through quantitative reconstruction evidence and transparent postprocessing, without treating heatmap appearance as proof of improved generalization.

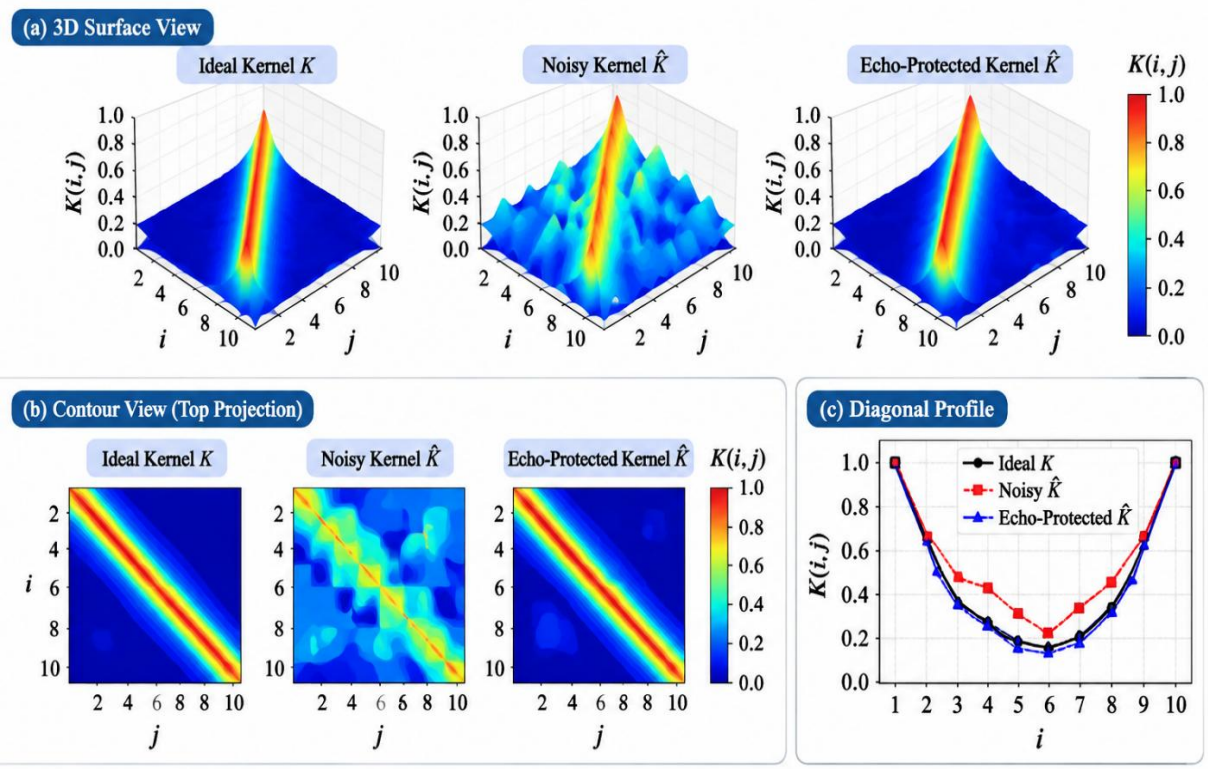


Figure 4. Kernel matrix comparison: (a) 3D surface plots, (b) contour maps (top view), and (c) diagonal profile of the kernel values for the ideal kernel K, noisy kernel $\hat{K}$, and echo-protected kernel $\hat{K}$. The noisy kernel shows noticeable distortions, whereas the echo-protected kernel closely reproduces the ideal structure, demonstrating effective preservation of the kernel under noise.

5.2 Kernel matrices and preservation of structure

5.2.1 Entry wise errors and matrix errors

I write the raw reconstructed matrix as

$$\widehat{K} = K + B + \Xi, \qquad \mathbb{E}[\Xi] = 0. \qquad (44)$$

Using the Frobenius inner product and $\mathbb{E}\langle B, \Xi\rangle_F = 0$, I obtain

$$\mathbb{E} \parallel \widehat{K} - K \parallel_F^2 = \parallel B \parallel_F^2 + \sum_{i,j} \mathrm{Var}\left(\widehat{K}_{ij}\right). \qquad (45)$$

This identity does not require errors in different entries to be independent. If I mirror upper-triangular estimates, the duplicated off-diagonal terms must still be counted twice in the full matrix norm. I report the relative reconstruction error and normalized matrix alignment as

$$\epsilon_F = \frac{\parallel \widehat{K} - K \parallel_F}{\parallel K \parallel_F}, \qquad \mathcal{A} = \frac{\langle \widehat{K}, K\rangle_F}{\parallel \widehat{K} \parallel_F \parallel K \parallel_F}. \qquad (46)$$

Alignment measures direction in matrix space and is insensitive to a positive overall scaling. It cannot substitute for absolute calibration. I would additionally report off-diagonal reconstruction error, since a correct diagonal can dominate a heatmap without establishing informative similarities.

5.2.2 A bound for the product encoding

Let $a_k = (1 + \lambda_k \cos\Delta_k)/2$ and $b_k = (1 + \cos\Delta_k)/2$, so $q = \prod_k a_k$ and $K = \prod_k b_k$. I expand their difference by telescoping:

$$\prod_{k=1}^{n} a_k - \prod_{k=1}^{n} b_k = \sum_{k=1}^{n} (a_k - b_k)\left(\prod_{r<k} a_r\right)\left(\prod_{r>k} b_r\right). \qquad (47)$$

All factors lie in $[0,1]$. Therefore

$$|q - K| \le \frac{1}{2}\sum_{k=1}^{n} (1 - \lambda_k)\,|\cos\Delta_k| \le \frac{1}{2}\sum_{k=1}^{n} (1 - \lambda_k)\,. \qquad (48)$$

For identical attenuation $\lambda = e^{-\chi}$, this gives $|q - K| \le n(1 - e^{-\chi})/2 \le n\chi/2$, with the trivial probability bound of one available when tighter. Increasing each visibility reduces this worst-case bound. It does not remove the dependence on system size.

5.2.3 Spectral correction and its limits

If the measured matrix is asymmetric, I first form $K_s = \left(\widehat{K} + \widehat{K}^T\right)/2$. Since the ideal K is symmetric, orthogonality of symmetric and skew-symmetric matrices gives

$$\parallel \widehat{K} - K \parallel_F^2 = \parallel K_s - K \parallel_F^2 + \parallel \widehat{K} - K_s \parallel_F^2. \qquad (49)$$

Thus, symmetrization cannot increase the Frobenius error relative to the ideal symmetric target. For $K_s = Q\mathrm{diag}(\mu_r)Q^T$, its nearest positive-semidefinite matrix in Frobenius norm is

$$K_+ = Q\mathrm{diag}\big(\max(\mu_r, 0)\big)Q^T. \qquad (50)$$

To see this, I rotate the optimization into the eigenbasis of K_s. Off-diagonal elements add squared error, while each admissible diagonal element must be nonnegative; its optimum is therefore $\max(\mu_r, 0)$. Projection onto the closed convex positive-semidefinite cone gives

$$\| K_+ - K \|_F \leq \| K_s - K \|_F \qquad (K \succcurlyeq 0). \qquad (51)$$

This is a classical matrix correction. I would show raw and corrected errors separately, because a good corrected heatmap cannot identify which improvement came from the hardware. Kernel-matrix regularization and mitigation are established parts of the comparison [7].

For $E = K_s - K$, the eigenvalue perturbation bound is

$$|\mu_r(K_s) - \mu_r(K)| \leq \| E \|_2 \leq \| E \|_F. \qquad (52)$$

Small negative eigenvalues can therefore result from estimation error even when the target is a valid kernel. Conversely, positive semi definiteness alone says nothing about predictive relevance.

5.2.4 Diagonal normalization and out-of-sample consistency

If the diagonal of a PSD matrix K_+ is strictly positive, I can normalize it with $D = \mathrm{diag}(K_+)$:

$$K_{mnorm} = D^{-1/2} K_+ D^{-1/2}, \qquad (K_{mnorm})_{ii} = 1. \qquad (53)$$

This congruence preserves positive semi definiteness, but can amplify errors when diagonal entries are small. It also changes the estimator. Simply replacing the diagonal of an arbitrary measured matrix by ones is not equivalent to (53) and does not guarantee a PSD result.

For the ideal normalized pure-state fidelity kernel, (3) requires a flat diagonal at one. Consequently, I cannot interpret the U-shaped "ideal diagonal profile" in Figure 4 as K_{ii}. It would need a different, explicitly defined cross-section, or it must be replaced. In the present dephasing model, the measured self-return is instead

$$q_{ii} = \prod_{k=1}^{n} \frac{1 + \lambda_k}{2}, \qquad (54)$$

which is below one unless all visibilities equal one. With the same exposure and attenuation for every input, even this noisy diagonal is constant across sample index.

A train-only PSD repair also requires an out-of-sample rule. One explicit construction retains strictly positive eigenvalues Λ_+ and eigenvectors Q_+, defines training features $z_i = \Lambda_+^{1/2} Q_+^T e_i$, and maps a raw query vector $\hat{k}_x$ to

$$z(x) = \Lambda_+^{-1/2} Q_+^T \hat{k}_x, \qquad k_{\mathrm{ext}}(x, x') = z(x)^T z(x'). \qquad (55)$$

This produces a consistent PSD feature representation, but small retained eigenvalues can amplify query error, so any truncation threshold must be selected using training data. If I use normalized features, I apply the same normalization rule to future inputs. I do not repair the training matrix and then silently assume that unrelated raw test similarities define the same kernel.

Finally, I distinguish matrix recovery from distinguishability. For probabilities bounded away from zero and one, resolving a difference δ with shot noise requires $M = O(\delta^{-2})$ for fixed precision confidence. If informative differences shrink exponentially with n, this scaling is exponential. Near zero probability, relative-error shot requirements differ, but can still be exponential. These estimator-dependent limitations are consistent with quantum-kernel concentration analyses [9]; echo suppression alone does not establish their removal.

6.1 Machine-learning classification accuracy versus noise strength: Demonstrate practical benefit

I assess the proposed protection mechanism through its contribution to learning on unseen data, while distinguishing that contribution from the suitability of the underlying representation. Glick et al. [11] developed covariant quantum kernels that exploit group structure in data and demonstrated their implementation on superconducting hardware. Their results emphasise the importance of matching the kernel's structure to the learning problem. A physically more accurate estimate of an unsuitable representation need not improve classification.

Huang et al. [12] established a further distinction between computational difficulty and predictive advantage. Their analysis showed that classical learning can predict some quantities despite the difficulty of computing them directly. I therefore distinguish improvement over an unprotected quantum implementation from superiority over competitive classical models. Agliardi et al. [10] illustrate this distinction experimentally: mitigation substantially improved their quantum classifications while producing performance comparable to classical alternatives. Noise resilience within a quantum workflow and quantum advantage are separate claims.

For Figure 5, I interpret classification accuracy as evidence about the complete learning procedure, including feature encoding, kernel estimation, model selection, and prediction-time evaluation. Reduced estimation error does not guarantee higher accuracy because the importance of a kernel perturbation depends on its relationship to the labels and the decision boundary. I accordingly limit the proposed framework's immediate objective to a reproducible generalisation benefit under defined noise conditions and justified resource expenditure. Broader claims of universal robustness or scalable quantum advantage require additional evidence beyond improvement on a particular classification task.

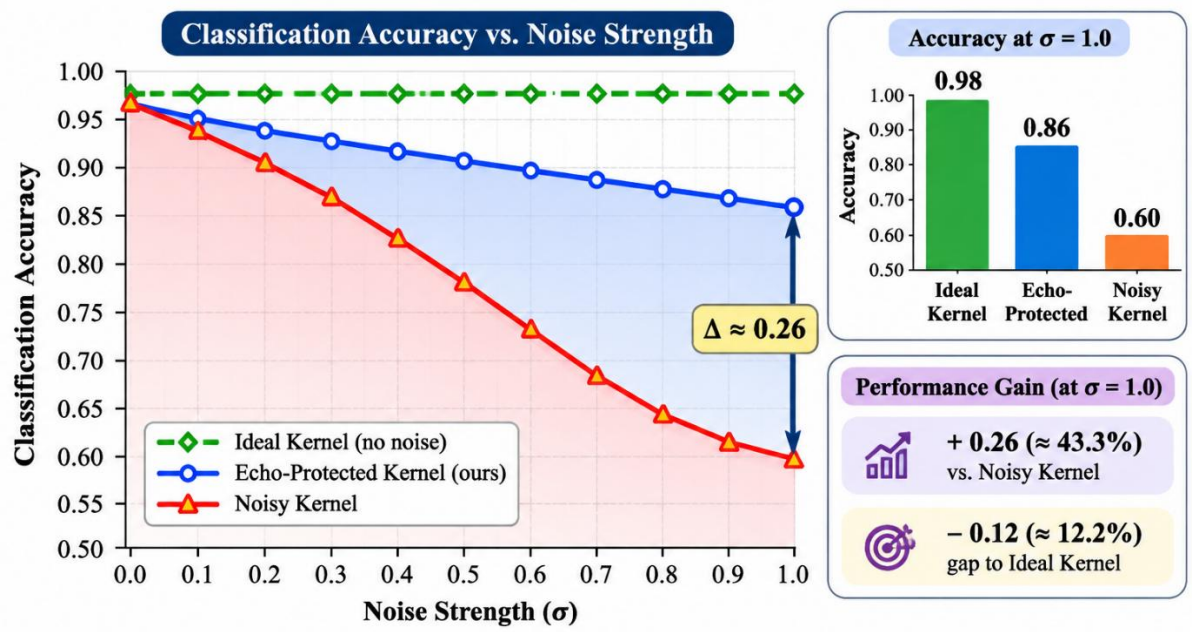


Figure 5. Machine-learning classification accuracy versus noise strength (σ). The echo-protected kernel (blue) consistently achieves higher accuracy than the noisy kernel (red) and remains close to the ideal (green) performance. At $\sigma = 1.0$, echo protection improves the accuracy by approximately 0.26 (≈ 43.3%) compared to the noisy kernel, demonstrating the practical benefit of the proposed method for robust quantum kernel learning.

6.2 Classification and the meaning of improved accuracy

6.2.1 Derivation of the support vector classifier

For a binary task, I use the real Hilbert space of Hermitian operators with inner product $\langle A, B\rangle = \mathrm{Tr}(AB)$. With feature ρ_i, I define the soft-margin problem

$$\min_{w,b,\xi} \quad 1/2 \parallel w \parallel^2 + C_{\mathrm{svm}} \sum_i \xi_i$$

$$\text{subject to} \quad \ell_i(\langle w, \rho_i\rangle + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0. \qquad (56)$$

I introduce nonnegative multipliers α_i for the margin constraints and η_i for $\xi_i \geq 0$. The Lagrangian is

$$\mathcal{L} = 1/2 \parallel w \parallel^2 + C_{\mathrm{svm}} \sum_i \xi_i - \sum_i \alpha_i \left[\ell_i(\langle w, \rho_i\rangle + b) - 1 + \xi_i\right] - \sum_i \eta_i \xi_i. \qquad (57)$$

Stationarity with respect to w, b, and ξ_i gives

$$w = \sum_i \alpha_i \ell_i \rho_i, \qquad \sum_i \alpha_i \ell_i = 0, \qquad C_{\mathrm{svm}} - \alpha_i - \eta_i = 0. \qquad (58)$$

Substituting these relations into (57), I obtain the dual problem

$$\max_{\alpha}\left[\sum_{i}\alpha_i - 1/2\sum_{i,j}\alpha_i\,\alpha_j\ell_i\ell_j K_{ij}\right], \qquad 0\le\alpha_i\le C_{\text{svm}}, \quad \sum_{i}\alpha_i\,\ell_i = 0. \qquad (59)$$

For a free support vector with $0 < \alpha_j < C_{\text{svm}}$, complementary slackness yields $b = \ell_j - \sum_i \alpha_i\,\ell_i K_{ij}$. The decision function is

$$f(x) = \sum_{i}\alpha_i\,\ell_i K(x_i, x) + b. \qquad (60)$$

In the noisy workflow, I solve (59) using the chosen admissible estimated kernel and evaluate future points with the corresponding test–train kernel. Both training and prediction are part of the noise-dependent experiment.

6.2.2 What kernel accuracy can guarantee

I first hold the trained coefficients and intercept fixed to isolate prediction-time estimation. If $\left|\widehat{K}(x_i, x) - K(x_i, x)\right| \le \varepsilon$ for every training point, then

$$\left|\hat{f}(x) - f(x)\right| \le \sum_{i}\alpha_i\left|\widehat{K}(x_i, x) - K(x_i, x)\right| \le \varepsilon\sum_{i}\alpha_i\,. \qquad (61)$$

The predicted sign is unchanged whenever $|f(x)| > \varepsilon\sum_i \alpha_i$. A previously correct prediction is guaranteed to remain correct if its signed margin $\ell f(x)$ exceeds this bound. A smaller kernel error can therefore protect predictions near the decision boundary, but does not force a higher classification accuracy on every dataset.

If I retrain on the perturbed kernel, the coefficients also change. Writing $\hat{k}(x) = k(x) + \Delta k(x)$, I obtain the exact decomposition

$$\hat{f} - f = \sum_{i}(\hat{\alpha}_i - \alpha_i)\,\ell_i k_i + \sum_{i}\hat{\alpha}_i\,\ell_i\Delta k_i + \left(\hat{b} - b\right). \qquad (62)$$

Equation (61) alone does not bound the first and third terms. SVM solutions can have nonunique dual coefficients or changing support-vector sets, so I do not infer a universal retraining guarantee from a heatmap.

6.2.3 An explicit retraining bound for ridge classification

To make the training sensitivity analytically explicit, I also consider kernel ridge classification with no intercept and labels ℓ. In the feature space, minimizing $N^{-1}\parallel \Phi w - \ell \parallel_2^2 + \lambda \parallel w \parallel^2$ gives the normal equation $(\Phi^T\Phi + N\lambda I)w = \Phi^T\ell$. With $K = \Phi\Phi^T$, the corresponding coefficient vector and prediction are

$$a = (K + \gamma I)^{-1}\ell, \qquad f(x) = k(x)^T a, \qquad \gamma = N\lambda > 0. \qquad (63)$$

Here Φ is the feature matrix; the same argument applies in the operator feature space. Let $\widehat{K} = K + E$ be a symmetric training-matrix perturbation, and define $A = K + \gamma I$ and $\hat{A} = A + E$. The inverse identity gives

$$\hat{a} - a = \left(\hat{A}^{-1} - A^{-1}\right)\ell = -\hat{A}^{-1}EA^{-1}\ell. \qquad (64)$$

Since $K \succcurlyeq 0$, $\| A^{-1} \|_2 \leq 1/\gamma$. If $\| E \|_2 < \gamma$, the smallest eigenvalue of $\hat{A}$ is at least $\gamma - \| E \|_2$, and

$$\| \hat{a} - a \|_2 \leq \frac{\| E \|_2 \| \ell \|_2}{\gamma(\gamma - \| E \|_2)}, \qquad \| \hat{a} \|_2 \leq \frac{\| \ell \|_2}{\gamma - \| E \|_2}. \qquad (65)$$

For a perturbed test vector $\hat{k} = k + e_x$, I expand $\hat{f} - f = k^T(\hat{a} - a) + e_x^T\hat{a}$ and apply Cauchy–Schwarz:

$$\left|\hat{f}(x) - f(x)\right| \leq \frac{\| k(x) \|_2 \| E \|_2 \| \ell \|_2}{\gamma(\gamma - \| E \|_2)} + \frac{\| e_x \|_2 \| \ell \|_2}{\gamma - \| E \|_2}. \qquad (66)$$

This bound explicitly links training-matrix error, test-kernel error, and regularization. If both training matrices are PSD, I can replace the denominator $\gamma - \| E \|_2$ by γ and dispense with the small-perturbation condition, because both regularized inverses have norm at most $1/\gamma$. The resulting statement concerns ridge classification, not the SVM objective in (56).

6.2.4 Accuracy curves and uncertainty

For a fixed held-out set of size N_{test}, I define

$$\widehat{\text{Acc}}_a(\sigma) = \frac{1}{N_{\text{test}}} \sum_{r=1}^{N_{\text{test}}} \mathbf{1}\{\text{sign} f_{a,\sigma}(x_r) = \ell_r\}. \qquad (67)$$

This quantity depends on the data distribution, split, feature map, classifier, hyperparameters, noise, and measured kernels. There is no accuracy-versus-noise function that follows from echo symmetry alone. To estimate an expected accuracy curve, I repeat the complete procedure over prescribed training splits and noise or shot realizations, then report the mean and uncertainty. I use paired splits and paired evaluation data when comparing protocols and select hyperparameters without consulting test labels.

For a fixed classifier evaluated on independent identically distributed test examples, the conditional standard error is estimated by $\sqrt{\widehat{\text{Acc}}\left(1 - \widehat{\text{Acc}}\right)/N_{\text{test}}}$. This does not include training variability or shared hardware drift. I therefore distinguish uncertainty across complete repeated runs from uncertainty conditional on one trained model.

For a paired comparison on the same test set, I define $d_r = \mathbf{1}\{\text{echo correct}\} - \mathbf{1}\{\text{unprotected correct}\}$. The accuracy difference is $\overline{d}$, with conditional estimated standard error $s_d/\sqrt{N_{\text{test}}}$. This retains the covariance between the two predictions rather than treating the accuracies as independent.

Figure 5 displays values of 0.86, 0.60, and 0.98 at its final noise point. Their arithmetic consequences are

$$0.86 - 0.60 = 0.26, \qquad \frac{0.86 - 0.60}{0.60} \times 100\% = 43.33\%. \qquad (68)$$

The first is a gain of 26 percentage points; the second is a relative increase in accuracy. The ideal-to-protected gap and relative error reduction are different quantities:

$$0.98 - 0.86 = 0.12, \qquad \frac{0.12}{0.98} \times 100\% = 12.24\%, \qquad (69)$$

$$\frac{(1 - 0.60) - (1 - 0.86)}{1 - 0.60} \times 100\% = 65\%. \qquad (70)$$

I can verify these numerical comparisons, but I cannot derive the three starting accuracies from the specified physical framework. I would retain them as results only if supported by reproducible simulations or measurements. Otherwise, I would label the curve as illustrative and avoid using it as evidence of practical benefit.

7 Interpretation of the five figures

I establish the mathematical connection among the figures through a sequence of conditional statements. Figure 1 specifies a PSD target similarity and a hybrid learning architecture. Figure 2 requires an explicit control sequence that preserves that target; the dephasing echo in (16) provides one such realization. Figure 3 follows from the resulting probabilities together with bias, sampling variance, and resource cost. Figure 4 requires quantitative matrix diagnostics and a distinction between physical recovery and classical correction. Figure 5 requires independently evaluated predictions; its numerical accuracy values are not consequences of the echo identity.

For the model developed here, I can state a precise protection result: ideal echo pulses reduce single-qubit kernel bias and equal-shot mean-squared error under exponentially correlated Gaussian dephasing. With the specified pulse depolarization model, this benefit survives only when (34) holds; under unequal execution costs, (40) is the relevant test. I do not extend this conclusion to arbitrary noise, arbitrary feature maps, or scalable learning advantage.

Before making an empirical claim, I would specify the dataset and preprocessing, feature angles and circuit depth, noise spectrum or channel, pulse schedule and errors, measurement count, acceptance rule if any, calibration cost, matrix processing, classifier selection, and test protocol. I would compare the echo against an unprotected baseline and applicable established mitigation under the same resource definition. These specifications make the proposed mechanism testable and determine which statements the eventual figures can support.

Across Figures 1–5, I organise the scientific argument around a sequence of distinct requirements: a specified protection mechanism, a justified kernel estimator, improved estimation under controlled comparisons, preservation of relevant matrix structure, and independently evaluated predictive benefit. The reviewed literature supports investigating these connections but does not determine their outcome for the proposed protocol. I identify the

unresolved question as whether its particular echo-based mechanism provides a reproducible benefit beyond established control, verification, and kernel-processing approaches.

8 Conclusion

I have formulated a time-symmetry-inspired framework for examining echo protection in quantum kernel estimation. The central issue is the physical and statistical meaning of the protection mechanism. An inverse encoding operation establishes reversibility of the intended computation; a claim of error suppression additionally requires a justified account of how the protocol acts on unwanted dynamics or erroneous measurement contributions.

I have organised the assessment around a sequence of linked but distinct requirements: preservation of the target similarity measure, improved estimation under controlled noise, recovery of the relevant kernel-matrix structure, and predictive benefit on unseen data. This structure makes resource accounting integral to the assessment. A reduction in error among accepted measurements must be evaluated together with acceptance probability, sampling uncertainty, circuit overhead, and any classical correction applied to the matrix.

The present formulation does not establish that a particular echo protocol satisfies these requirements. That conclusion will require an explicit control sequence, a mathematically justified estimator, and reproducible comparisons with existing mitigation methods under comparable experimental budgets. Scalability and classical competitiveness remain separate questions.

I therefore identify the next scientific task as determining where a specific echo intervention produces a measurable learning benefit and where its assumptions or costs prevent one. The value of the proposed framework lies in making those boundaries explicit so that any eventual claim of protection rests on demonstrated improvements in estimation and prediction.

References

[1] Schuld, M. & Killoran, N. Quantum machine learning in feature Hilbert spaces. *Phys. Rev. Lett.* **122**, 040504 (2019). https://doi.org/10.1103/PhysRevLett.122.040504

[2] Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., & Gambetta, J. M. (2019). Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747), 209–212. https://doi.org/10.1038/s41586-019-0980-2

[3] Hahn, E. L. Spin echoes. *Phys. Rev.* **80**, 580–594 (1950). https://doi.org/10.1103/PhysRev.80.580

[4] Viola, L., Knill, E. & Lloyd, S. Dynamical decoupling of open quantum systems. *Phys. Rev. Lett.* **82**, 2417–2421 (1999). https://doi.org/10.1103/PhysRevLett.82.2417

[5] O'Brien, T. E., Polla, S., Rubin, N. C., Huggins, W. J., McArdle, S., Boixo, S., McClean, J. R. & Babbush, R. Error mitigation via verified phase estimation. *PRX Quantum* **2**, 020317 (2021). https://doi.org/10.1103/PRXQuantum.2.020317

[6] Schiffer, B. F., van Vreumingen, D., Tura, J. & Polla, S. Virtual mitigation of coherent non-adiabatic transitions by echo verification. *Quantum* **8**, 1346 (2024). https://doi.org/10.22331/q-2024-05-14-1346

[7] Hubregtsen, T., Wierichs, D., Gil-Fuster, E., Derks, P.-J. H. S., Faehrmann, P. K. & Meyer, J. J. Training quantum embedding kernels on near-term quantum computers. *Phys. Rev. A* **106**, 042431 (2022). https://doi.org/10.1103/PhysRevA.106.042431

[8] Weaving, T., Ralli, A., Love, P. J., Succi, S. & Coveney, P. V. Accurately simulating the time evolution of an Ising model with echo verified Clifford data regression on a superconducting quantum computer. *Quantum* **9**, 1732 (2025). https://doi.org/10.22331/q-2025-05-05-1732

[9] Thanasilp, S., Wang, S., Cerezo, M., & Holmes, Z. (2024). Exponential concentration in quantum kernel methods. *Nat. Commun.*, 15, 5200. https://doi.org/10.1038/s41467-024-49287-w

[10] Agliardi, G., Cortiana, G., Dekusar, A., Ghosh, K., Mohseni, N., O'Meara, C., Valls, V., Yogaraj, K., & Zhuk, S. (2026). Mitigating exponential concentration in covariant quantum kernels for subspace and real-world data. *npj Quantum Inf.*, 12, 12. https://doi.org/10.1038/s41534-025-01154-2

[11] Glick, J. R., Gujarati, T. P., Córcoles, A. D., Kim, Y., Kandala, A., Gambetta, J. M., & Temme, K. (2024). Covariant quantum kernels for data with group structure. *Nat. Phys.*, 20, 479–483. https://doi.org/10.1038/s41567-023-02340-9

[12] Huang, H.-Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., & McClean, J. R. (2021). Power of data in quantum machine learning. *Nat. Commun.*, 12, 2631. https://doi.org/10.1038/s41467-021-22539-9

[13] Vezvaee, A., Shitara, N., Sun, S. & Montoya-Castillo, A. Fourier transform noise spectroscopy. *npj Quantum Inf.* **10**, 52 (2024). https://doi.org/10.1038/s41534-024-00841.