Expanding Quantum Oracle Sketching and Classical Shadows

A. Sepulveda–Jimenez[1]

[1]QDR Labs, National University

May 2026

Abstract

The recent breakthrough by [Zha+26] demonstrates a provable exponential quantum advantage in processing massive classical data using polylogarithmic quantum space, primarily through the innovations of *quantum oracle sketching* (QOS) and *interferometric classical shadows*. This brief manuscript reviews that framework, and as also discussed in the *Quantum Frontiers* forum [Qua26], and proposes two mathematical extensions: (i) a non–linear kernel–QOS protocol based on random Fourier features, with explicit sample complexity, and (ii) a rigorous noise analysis that distinguishes errors during sketch construction from errors during shadow readout. We close with practical implications for hybrid machine learning pipelines.

1 Introduction

The intersection of quantum computing and classical machine learning has long sought a provable, exponential advantage that does not rely on loading the entirety of massive classical datasets into quantum memory—a known bottleneck of Quantum Random–Access Memory [GLM08]. Several proposed quantum–machine–learning speedups have been partially or wholly dequantized once the data–loading cost is properly accounted for [Aar15; Tan19], which has fueled scepticism about the practical reach of the field [SK22].

[Zha+26] present a definitive resolution of this tension by proving that a quantum computer with a polylogarithmic number of qubits (e.g., fewer than 60 logical qubits in the worked examples) can perform large–scale classification and dimension reduction on massive classical data. The mechanism is twofold: *quantum oracle sketching*, which constructs an approximate phase oracle of a Boolean function from streaming samples; and *interferometric classical shadows*, a variant of shadow tomography [HKP20] compatible with the QOS oracle. Together they bypass the classical I/O bottleneck [Qua26].

This manuscript analyses the strengths and limits of the framework with a view towards real–world machine-learning deployment, where pipelines are typically prototyped in R, Python, or Julia and run on imperfect hardware. Two issues stand out. First, the empirical demonstrations in [Zha+26]—LS–SVM on TF–IDF features and PCA on single–cell RNA expression—are linear methods, while modern data science is dominated by non–linear feature maps. Second, the analysis in the original paper is in the ideal–gate model, while present and near–future hardware operates with non–negligible noise. We address both gaps below.

2 Quantum Oracle Sketching: A Brief Review

[Zha+26] consider a Boolean function $f : [N] \to \{0,1\}$ and i.i.d. samples $z_t = (x_t, f(x_t))$ with x_t drawn from a distribution p on $[N]$. For each sample they apply the conditional phase rotation

$$V_t \;=\; \exp\Big(\, i\,\tau\, f(x_t)\, |x_t\rangle\langle x_t| \,/\, M \,\Big), \tag{1}$$

with $\tau = \pi N$ in the uniform–prior case, and form the streaming unitary $V = V_M \cdots V_1$. Because all V_t commute,

$$V \;=\; \exp\Big(i\tau \sum_{x\in[N]} m_x f(x)|x\rangle\langle x|\Big), \quad m_x = \frac{1}{M}\sum_{t=1}^{M} \mathbf{1}[x_t = x], \tag{2}$$

and the empirical mass m_x concentrates around p_x at rate $1/\sqrt{M}$. Standard matrix concentration arguments [Tro12] give the diamond–norm bound

$$\|\mathbb{E}[\,\mathcal{V}\,] - \mathcal{O}\|_\diamond \;=\; \mathcal{O}(N/M), \tag{3}$$

where $\mathcal{O}$ is the ideal phase–oracle channel and $\mathcal{V}(\rho) = V\rho V^\dagger$. Subsequent quantum singular value transformation [Gil+19] converts the sketched oracle into the algorithmic primitives required for LS–SVM classification, linear–system solving, and PCA, with classical shadows returning useful predictions to the host.

2.1 Why linearity matters in the original framework

The classical shadow $\hat{\rho}_X$ produced by interferometric tomography satisfies $\mathbb{E}[\mathrm{Tr}(O\hat{\rho}_X)] = \mathrm{Tr}(O\rho_X)$ for any linear observable O. This is sufficient when the downstream task asks linear questions of the data—a polynomial of bounded degree, the top eigenvector of a covariance matrix, the inner product $\langle x', \hat{w}\rangle$ for a linear classifier $\hat{w}$. However, many modern data–science workflows operate on non–linear manifolds: sentiment classification with kernel SVMs, manifold learning in single–cell genomics, financial regime detection with mixture models. A practitioner attempting to apply QOS to such tasks needs a principled extension to non–linear feature maps, which is the purpose of Section 3.

3 Mathematical Expansion I: Non–Linear Kernel Sketching

3.1 Random Fourier features as a feature–map oracle

We extend QOS to non–linear kernel learning by composing it with the random Fourier feature (RFF) construction of [RR07]. Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a continuous, shift–invariant, positive–definite kernel, $k(x,y) = k(x-y)$, and let μ denote its spectral measure. Bochner's theorem gives the integral representation

$$k(x,y) \;=\; \int_{\mathbb{R}^d} e^{i\,\omega^\top(x-y)}\, d\mu(\omega). \tag{4}$$

Drawing $\omega_1,\dots,\omega_D \sim \mu$ and $b_1,\dots,b_D \sim \mathrm{Unif}[0,2\pi]$ independently, define the random feature vector

$$z(x) \;=\; \sqrt{\tfrac{2}{D}}\Big(\cos(\omega_1^\top x + b_1),\,\dots,\,\cos(\omega_D^\top x + b_D)\Big)^\top \in \mathbb{R}^D, \tag{5}$$

satisfying $\mathbb{E}[z(x)^\top z(y)] \;=\; k(x,y)$. Each streaming sample $x_t \in \mathbb{R}^d$ is mapped *classically* on the host to $z(x_t) \in \mathbb{R}^D$, then encoded as a quantum state on $\lceil \log_2 D\rceil$ qubits via amplitude encoding,

$$|\psi_{z(x_t)}\rangle \;=\; \frac{1}{\|z(x_t)\|_2}\sum_{j=1}^{D} z_j(x_t)\,|j\rangle, \tag{6}$$

using the QOS state–preparation primitive of [Zha+26, App. D.5]. This gives the kernel–QOS density matrix $\rho_\phi = M^{-1}\sum_{i=1}^{M}|\psi_{z(x_i)}\rangle\langle\psi_{z(x_i)}|$.

3.2 Sample complexity and qubit budget

The accuracy of the classical RFF approximation is well–understood:

Lemma 1 ([RR07; RR17])**.** *For accuracy $\varepsilon_K \in (0,1)$, failure probability $\delta > 0$, and any compact $\Omega \subset \mathbb{R}^d$ of diameter $\mathcal{O}(1)$,*

$$D \;\geq\; C\,\frac{d}{\varepsilon_K^2}\log\frac{d}{\delta\varepsilon_K} \tag{7}$$

suffices to guarantee $\sup_{x,y\in\Omega}\left|z(x)^\top z(y) - k(x,y)\right| \leq \varepsilon_K$ with probability $\geq 1-\delta$, where C depends only on the moment of the spectral measure. The empirical kernel matrix $\widehat{K} = ZZ^\top$ then satisfies $\left\|\widehat{K} - K\right\|_{\mathrm{op}} \leq \varepsilon_K\,\|K\|_{\mathrm{op}}$.

Combining this with the QOS sample bound (3) yields the following.

Theorem 2 (Kernel–QOS sample complexity)**.** *Let k be a continuous shift–invariant kernel and consider an LS–SVM classifier trained on the kernel matrix $K + \lambda I$ with $\sigma_{\min}(K+\lambda I) \geq \lambda > 0$. For target accuracy $\varepsilon \in (0, \frac{1}{2})$, kernel–QOS achieves classification accuracy within ε of the population kernel rule using*

$$M \;=\; \Theta\!\left(\frac{D\,Q^2}{\varepsilon^2}\right) \;=\; \Theta\!\left(\frac{d\,\log(d/\varepsilon)\,Q^2}{\varepsilon^4\,\lambda^2}\right) \tag{8}$$

streaming samples and $n_q = \mathcal{O}(\log\log N + \log(1/\varepsilon))$ logical qubits, where $Q = \widetilde{\Theta}(\kappa(K+\lambda I))$ is the QSVT query complexity of the LS–SVM solver.

Proof sketch. Theorem 1 with $\varepsilon_K = \varepsilon\lambda$ ensures the RFF kernel matrix is within ε of the true kernel in operator norm. The kernel–QOS oracle is then a phase oracle on the D–dimensional RFF space; (3) applied with $N \mapsto D$ gives $M = \Theta(DQ^2/\varepsilon^2)$. The QSVT–based LS–SVM solver of [Zha+26, §F.2] requires $Q = \widetilde{\Theta}(\kappa)$ queries, and the qubit count includes $\lceil \log_2 D \rceil$ for the feature register plus $\mathcal{O}(\log(1/\varepsilon))$ ancillas. □

3.3 Critical interpretation of the kernel advantage

A crucial point that was glossed over in our earlier draft: kernel SVMs traditionally operate on an $N \times N$ Gram matrix, and the "space advantage" of QOS for kernel learning must be stated against this baseline carefully.

- Compared to a *naïve* Gram–matrix solver requiring $\Omega(N^2)$ memory, kernel–QOS gives an unconditional exponential advantage: $\mathcal{O}(\operatorname{polylog} N)$ qubits suffice.
- Compared to a *Nyström* [WS01] or RFF solver that already runs in $\mathcal{O}(D^2)$ classical memory, kernel–QOS still gives an exponential advantage in the *feature–dimension* regime, since the quantum register holds $\log D$ qubits versus D classical floats.
- Compared to quantum–feature–map kernels of [SK19; Hav+19], kernel–QOS is *conceptually distinct*: the feature map $z(\cdot)$ is computed classically, and only the streaming sketch and the QSVT–based solver run on the quantum device. This avoids the dequantization concerns that have shadowed quantum–feature–map proposals and gives an end–to–end unconditional sample/space tradeoff.

Remark 3 (Effective dimensionality on real data)*.* For data approximately concentrated on a low–dimensional manifold of intrinsic dimension $d_{\text{eff}} \ll d$, the RFF dimension required for operator–norm kernel approximation reduces accordingly, giving favourable constants for the biological and textual benchmarks in [Zha+26].

4 Mathematical Expansion II: Rigorous Noise Bounds

For near–term deployment we must account for noise. The QOS pipeline has two distinct error channels, both of which contribute to the end–to–end error:

(a) **Sketch–construction noise:** each rotation V_t in (1) is implemented by a noisy gate $\widetilde{\mathcal{V}}_t = \mathcal{E}_{p_g} \circ \mathcal{V}_t$, where $\mathcal{E}_{p_g}(\rho) = (1-p_g)\rho + p_g\, I/2^n$ is a depolarizing channel of strength p_g.

(b) **Shadow–readout noise:** the random Clifford measurement that produces classical snapshots is itself subject to depolarizing noise of strength p_r.

Our earlier draft addressed only (b). A complete analysis must combine both, and the dominant constraint typically comes from (a).

4.1 Sketch–construction threshold (corrects §3.2 of the earlier draft)

Theorem 4 (Per–gate noise threshold for QOS)**.** *Let $\widetilde{V} = \widetilde{V}_M \cdots \widetilde{V}_1$ denote the noisy QOS unitary with per–gate depolarizing rate p_g, and let $\mathcal{O}$ be the ideal phase–oracle channel. Then*

$$\left\| \mathbb{E}[\,\widetilde{\mathcal{V}}\,] - \mathcal{O} \right\|_\diamond \;\leq\; \underbrace{\mathcal{O}(N/M)}_{\textit{ideal QOS bias}} \;+\; \underbrace{2\,M\,p_g}_{\textit{noise compounding}} \;+\; \underbrace{\mathcal{O}(\sqrt{M\,p_g})}_{\textit{stochastic spread}}\,. \tag{9}$$

Reaching diamond–norm error ε therefore requires

$$M \;\geq\; \Theta\!\left(\frac{N}{\varepsilon}\right) \quad \textit{and} \quad p_g \;\leq\; \Theta\!\left(\frac{\varepsilon}{M}\right) \;=\; \mathcal{O}\!\left(\frac{\varepsilon^2}{N}\right). \tag{10}$$

For Q–query downstream algorithms the per–gate budget tightens further to $p_g \leq \mathcal{O}\big(\varepsilon^2/(NQ^2)\big)$.

Proof. The bias term follows from (3). Each noisy step admits the diamond–norm bound $\left\|\widetilde{\mathcal{V}}_t - \mathcal{V}_t\right\|_\diamond \leq 2p_g$ [Wil17, §4.2]. Concatenating M such channels and applying the triangle inequality gives the additive $2Mp_g$ term; the $\sqrt{Mp_g}$ correction is a Hoeffding bound on the variance of noise indicators across rotations. □

Practical consequence. For the IMDb sentiment example of [Zha+26, Fig. 2a] ($N \approx 10^5$, $Q \approx 10$, target $\varepsilon \approx 10^{-2}$), the threshold becomes $p_g \lesssim 10^{-13}$. Projected logical error rates in the megaquop fault–tolerant regime are 10^{-9} to 10^{-12} [Pre25], so QOS sits at the upper edge of near–term feasibility. This is a substantially tighter constraint than our earlier draft acknowledged.

4.2 Shadow–readout variance: correcting the 3^n claim

Our earlier note stated that the variance of the noisy classical shadow scales as $\mathrm{Var}(\hat{\rho}) \propto 3^n/(1-p)^2$, and that 3^n "grows modestly" since $n = \mathcal{O}(\mathrm{polylog}\, N)$. This last statement is mathematically incorrect: $3^{\mathrm{polylog}\, N}$ is super–polynomial in N, so a literal 3^n factor would defeat the QOS advantage. The correct picture, restated below, is more favourable.

Theorem 5 (Robust shadow variance, restated from [HKP20; Che+21; KG22])**.** *Let ρ be an n–qubit state and O a Hermitian observable, and let $\hat{\rho}$ be the classical shadow estimator from K snapshots under depolarizing readout noise of strength p_r. Define $\widehat{O} = \mathrm{Tr}(O\hat{\rho})$. Then*

(i) ***Random Clifford shadows:***

$$\mathrm{Var}(\widehat{O}) \;\leq\; \frac{3\,\|\,O - \mathrm{Tr}(O)I/2^n\,\|_F^2}{K\,(1-p_r)^2}; \tag{11}$$

the variance is governed by the (traceless) Hilbert–Schmidt norm of the observable, not by 3^n.

(ii) ***Random Pauli (local Clifford) shadows:***

$$\mathrm{Var}(\widehat{O}) \;\leq\; \frac{4^k \, \|O\|_{\mathrm{op}}^2}{K\,(1-p_r)^2}, \tag{12}$$

where k is the locality of O in the Pauli basis, not *the total qubit count.*

The exponential factor 4^k in (ii) only appears for *Pauli* shadows applied to global observables. In the QOS pipeline of [Zha+26], the relevant observables are linear functionals such as $\mathrm{Tr}(|x'\rangle\langle x'|\hat{\rho}_\phi)$ for sparse test inputs x', which have small Hilbert–Schmidt norm and modest locality. Hence the practical variance scaling is sub–polynomial in n, and the dominant cost is the $(1-p_r)^{-2}$ inflation factor we identified in our earlier draft. The qualitative conclusion—that sample complexity inflates quadratically in the noise rate—remains correct, as we now state precisely.

Corollary 6 (Combined noise–robust sample complexity)**.** *Under the threshold* (10) *on per–gate noise p_g and assuming readout depolarizing noise p_r, robust shadow inversion [Che+21] yields an unbiased estimator with sample requirement*

$$M_{\mathrm{noisy}} \;=\; \frac{N\,Q^2}{\varepsilon^2\,(1-p_r)^2}\,\big(1+o(1)\big), \tag{13}$$

provided the per–gate threshold is satisfied. The exponential–space advantage of QOS is preserved.

4.3 Practical mitigation strategies

The QOS construction is unusually friendly to error mitigation. Three techniques deserve mention:

- **Robust shadow inversion** [Che+21]: replace the standard inverse $\mathcal{M}^{-1}$ by the noise–calibrated map $\mathcal{M}_p^{-1} = (1-p)^{-1}\mathcal{M}^{-1} - p(1-p)^{-1}\big(\mathrm{Tr}\,/2^n\big)$, yielding an unbiased estimator without distributional assumptions beyond the depolarizing form.
- **Zero–noise extrapolation** [TBG17]: because each QOS rotation is weak ($\|V_t - I\| = \mathcal{O}(N/M)$), integer scaling of the rotation angle by factors $r \in \{1,3,5\}$ and Richardson extrapolation suppresses bias to $\mathcal{O}(p_g^2)$ at polynomial overhead.
- **Symmetry verification:** the phase oracle commutes with the diagonal symmetry group, so noisy snapshots violating this symmetry can be discarded as a post–selection step.

5 Implications for Machine Learning

Moving these algorithms from theory to production requires seamless integration with existing data–science ecosystems. R, Python, and Julia remain the standard for data manipulation, statistics, and scientific computing respectively. The QOS protocol fits naturally into a hybrid pipeline:

1. **Sample.** A classical node (R `data.table`, Python `polars`, Julia `DataFrames.jl`) randomly samples the data stream; for non–linear tasks the same node computes the RFF embedding $z(x_t)$ on the fly.
2. **Stream to QPU.** Each $z(x_t)$ is dispatched to the QPU through an Anthropic–style channel API, where a single rotation V_t updates the streaming sketch.

3. **Shadow readout.** After the sketch is constructed, randomized Clifford measurements produce classical shadow snapshots, returned to the host as a stream of bitstrings.

4. **Downstream classical analysis.** Predictions $\mathrm{Tr}(O\hat{\rho})$ are computed in R/Python/Julia for non–linear classification, manifold visualization, or downstream statistical inference.

This hybrid feedback loop is the true "unleashing" of quantum AI for production machine learning workflows. Crucially, the data scientist need not understand quantum mechanics to use the pipeline: the QPU acts as an accelerated linear–algebra coprocessor on streaming data, with the sample/space tradeoffs of Theorem 2 and Theorem 6 as the relevant performance contract.

6 Conclusion

The framework provided by [Zha+26] is a landmark achievement, confirming that exponential quantum advantages on classical data are accessible without QRAM. By expanding this methodology to include non–linear kernel approximations (Section 3) and establishing rigorous error bounds for both sketch–construction and shadow–readout noise (Section 4), we lay the groundwork for hybrid quantum–classical statistical learning pipelines.

This brief manuscript claims the exponential quantum advantage in space extends to non–linear kernel learning under standard well–conditioning hypotheses; the sample–complexity inflation from depolarizing noise is quadratic in $(1-p_r)^{-1}$ as previously argued, but the per–gate construction threshold $p_g \lesssim \varepsilon^2/(NQ^2)$ is the truly binding constraint for near–term hardware. Future work will focus on two directions: integrating these robust quantum oracle sketches into open–source statistical learning libraries (`tidymodels` in R, `scikit-learn` in Python, and `MLJ.jl` in Julia), and extending the kernel–QOS framework beyond shift–invariant kernels via Mercer expansions and Nyström–type subsampling [WS01].

Acknowledgements

The author thanks [Zha+26] for the paper which is the impetus of this brief work and the clarifying discussion in [Qua26] .

References

[Aar15] Scott Aaronson. "Read the fine print". In: *Nature Physics* 11.4 (2015), pp. 291–293.

[Che+21] Senrui Chen et al. "Robust shadow estimation". In: *PRX Quantum* 2.3 (2021), p. 030348.

[Gil+19] András Gilyén et al. "Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics". In: *Proc. 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019, pp. 193–204.

[GLM08] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum random access memory". In: *Physical Review Letters* 100.16 (2008), p. 160501.

[Hav+19] Vojtěch Havlíček et al. "Supervised learning with quantum-enhanced feature spaces". In: *Nature* 567.7747 (2019), pp. 209–212.

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". In: *Nature Physics* 16.10 (2020), pp. 1050–1057.

[KG22] Dax Enshan Koh and Sabee Grewal. "Classical shadows with noise". In: *Quantum* 6 (2022), p. 776.

[Pre25] John Preskill. "Beyond NISQ: The megaquop machine". In: *ACM Transactions on Quantum Computing* 6.3 (2025), pp. 1–7.

[Qua26] Quantum Frontiers. *Unleashing the advantage of quantum AI.* https://quantumfrontiers.com/2026/04/09/unleashing-the-advantage-of-quantum-ai/. Accessed: 2026–05–01. 2026.

[RR07] Ali Rahimi and Benjamin Recht. "Random features for large-scale kernel machines". In: *Advances in Neural Information Processing Systems* 20 (2007).

[RR17] Alessandro Rudi and Lorenzo Rosasco. "Generalization properties of learning with random features". In: *Advances in Neural Information Processing Systems* 30 (2017).

[SK19] Maria Schuld and Nathan Killoran. "Quantum machine learning in feature Hilbert spaces". In: *Physical Review Letters* 122.4 (2019), p. 040504.

[SK22] Maria Schuld and Nathan Killoran. "Is quantum advantage the right goal for quantum machine learning?" In: *PRX Quantum* 3.3 (2022), p. 030101.

[Tan19] Ewin Tang. "A quantum-inspired classical algorithm for recommendation systems". In: *Proc. 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019, pp. 217–228.

[TBG17] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. "Error mitigation for short-depth quantum circuits". In: *Physical Review Letters* 119.18 (2017), p. 180509.

[Tro12] Joel A. Tropp. "User-friendly tail bounds for sums of random matrices". In: *Foundations of Computational Mathematics* 12.4 (2012), pp. 389–434.

[Wil17] Mark M. Wilde. *Quantum Information Theory*. 2nd. Cambridge University Press, 2017.

[WS01] Christopher K. I. Williams and Matthias Seeger. "Using the Nyström method to speed up kernel machines". In: *Advances in Neural Information Processing Systems* 13 (2001).

[Zha+26] Haimeng Zhao et al. "Exponential quantum advantage in processing massive classical data". In: *arXiv preprint arXiv:2604.07639* (2026).