On the Optimality and Generalization of Quantum Oracle Sketching

A. Sepúlveda-Jiménez[1,2]

[1]National University
[2]QDR Labs

May 2026

Abstract

[Zha+26] introduce *quantum oracle sketching* (QOS), a streaming algorithm that constructs an approximate phase oracle of a Boolean function $f : [N] \to \{0, 1\}$ from $M = \Theta(NQ^2/\varepsilon)$ classical samples and then executes any quantum query algorithm of complexity Q on the sketched oracle. Combined with an interferometric variant of classical shadow tomography, QOS yields exponential separations in machine *size* between quantum and classical learners for linear systems, classification, and dimensionality reduction. The argument is information–theoretic and unconditional, depending only on quantum mechanics.

This work offers a rigorous critique and constructive extension. We (i) sharpen the diamond–norm guarantee using non–commutative matrix Bernstein, recovering the linear N/M bias scaling without the spurious logarithmic factor and with explicit constants; (ii) recast the optimal Q^2 scaling as a quantum Cramér–Rao bound on the symmetric–logarithmic–derivative (SLD) Fisher information of the random oracle channel; (iii) generalize QOS to non–uniform priors by showing that sample complexity is governed by the *effective dimension* $N_{\text{eff}} = \|p\|_{1/2}^2$, allowing genuine improvements when data are heavy–tailed; (iv) replace the hard spectral–gap assumption in PCA by a polynomial–filter analysis that pays only quadratically in the soft gap; (v) building on the parallel manuscript of [Sep26], extend QOS to non–linear kernel learning via random Fourier features, deriving the explicit sample complexity for shift–invariant kernels via an independent route and connecting it to the information–geometric framework of (ii) and the effective–dimension framework of (iii); (vi) prove that QOS is structurally compatible with Rényi differential privacy, achieving a per–sample privacy budget that scales as $\varepsilon_{\text{DP}}^2 = \Theta(1/(NM))$; (vii) analyse a hybrid scheme that pre—composes QOS with a Johnson–Lindenstrauss embedding, isolating the regimes in which the quantum advantage survives classical compression; (viii) building further on [Sep26], give a matrix Bernstein proof of the QOS noise threshold $p_g \lesssim \varepsilon^2/(NQ^2)$ with explicit constants, and pose the matching tight–threshold question as an open problem alongside the recent noisy–random–circuit obstruction [Aha+23]. We close with a candid critique of the resource accounting underlying the "60–logical–qubit" demonstration and a list of open problems in soft–gap, non–Hermitian, and continuous–domain QOS.

Contents

1 Introduction and Critical Summary

The pursuit of broadly applicable quantum advantage has been hampered by two structural obstacles: the *data–loading problem*—the cost of moving classical bits into a coherent quantum register—and the *readout problem*—the difficulty of returning useful classical output without measuring out the exponential Hilbert space. For two decades, the standard remedy has been quantum random–access memory [GLM08], whose fault–tolerant overhead is now known to be prohibitive in any scalable architecture [JR25]. Many proposed quantum machine–learning speedups dissolve once these two problems are accounted for [Aar15; Tan19]; one school of thought has concluded that the quantum–AI dream may be a mirage [Har20; SK22].

[Zha+26] present a striking counter–proposal. Rather than store the data, they show that a sequence of incremental phase rotations driven by streaming samples constructs a coherent oracle *on the fly*. Concretely, given i.i.d. samples $z_t = (x_t, f(x_t))$ with $x_t \sim \mathrm{Unif}([N])$ they apply

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

with $\tau = \pi N$, and prove that the random unitary $V = \prod_{t=1}^{M} V_t$ approximates the phase oracle $O = \sum_x (-1)^{f(x)} |x\rangle\langle x|$ in diamond distance with bias $\mathcal{O}(N/M)$. This linear–in–N scaling is much better than the N^2/M that would follow from a generic randomized–compiler analysis [Cam19; Che+21a] and arises because the generators $\{|x\rangle\langle x|\}_{x\in[N]}$ are mutually commuting. Composed with classical shadow tomography [HKP20], QOS yields end–to–end algorithms whose memory grows only as polylog N while maintaining provable accuracy.

The headline results are six theorems separating quantum and classical learners for static and time–varying linear systems, LS–SVM classification, and PCA. The classical hardness side is established via a sample–space lower bound $MS \geq \Omega(NQ_C)$ obtained by lifting random–query communication complexity, plus a learning–XOR amplification that converts the static separation into a sample–complexity gap for dynamic data. Crucially the separation is information–theoretic: it holds even if $\mathsf{BPP} = \mathsf{BQP}$.

What this commentary does. The construction is elegant and the lower bounds are deep, but the analysis leaves several mathematical and methodological avenues open. Our contribution is fourfold.

- **(Section 3)** A self–contained proof of the diamond–norm bound based on the matrix Bernstein inequality [Tro12; Tro15], giving explicit constants and removing the logarithmic factor present in proofs that go through operator–norm union bounds. The same machinery yields a *martingale* version that handles correlated streams without the R–repetition assumption.
- **(Section 4)** An information–geometric reformulation showing that the quadratic Born–rule penalty $M = \Theta(NQ^2/\varepsilon)$ is exactly the symmetric–logarithmic–derivative quantum Cramér–Rao bound for estimating the parameters of the random oracle channel. This places the optimality result of [Zha+26] in a one–line conceptual frame and exposes the precise object that any improved algorithm would have to dominate.
- **(Sections 5 and 6)** Two practical generalizations. First, an importance–weighted QOS in which uniform sampling is replaced by a proposal q chosen to minimize the second moment, recovering an $N_{\text{eff}} = \|p\|_{1/2}^2$ instead of N in the sample complexity; for heavy–tailed text or genomic data, $N_{\text{eff}} \ll N$. Second, a polynomial–filter PCA primitive that breaks the hard spectral–gap dependence by exchanging it for a *soft* gap parameter (Δ_q/σ_1), the quantity that actually appears in finite–sample random–matrix theory [Joh01].
- **(Section 7)** A rigorous extension of QOS to non–linear kernel learning. The companion manuscript of [Sep26] develops the random–Fourier–feature [RR07] construction in detail; we give a parallel treatment that emphasizes the connection to our information–geometric framework (Section 4) and to the effective–dimension analysis (Section 5), and identify the genuine but subtle sense in which the exponential space advantage transfers to kernel SVMs—namely in the *feature–dimension*, not the *sample–dimension*, regime.
- **(Sections 8 to 11)** A privacy analysis: because each sample is processed once and discarded, QOS admits a clean Rényi–DP [Mir17] composition, yielding $(\alpha, \bar{\varepsilon})$–RDP per sample with $\bar{\varepsilon} = \Theta(\alpha/(NM))$ for free. A hybrid analysis shows that pre–composing QOS with a Johnson–Lindenstrauss embedding preserves the exponential–space advantage if and only if the embedding dimension k obeys $k \leq \mathcal{O}(\log N/\varepsilon^2)$. A noise–robustness analysis, parallel to the threshold derivation of [Sep26], gives a matrix–Bernstein–style proof of the per–gate threshold $p_g \lesssim \varepsilon^2/(NQ^2)$ with explicit constants and connects it to recent results on the classical–simulability of noisy random circuits [Aha+23]. Finally, a critical look at the "60–logical–qubit" claim factoring in QSVT depth and magic–state distillation overhead.

2 Preliminaries: Notation and the QOS Construction

We work with finite–dimensional Hilbert spaces. Let $\mathcal{H} = \mathbb{C}^N$, with computational basis $\{|x\rangle\}_{x\in[N]}$. Operators on $\mathcal{H}$ form the algebra $\mathcal{B}(\mathcal{H})$, endowed with the operator norm $\|\cdot\|_{\mathrm{op}}$, the Schatten–p norms $\|\cdot\|_p = (\mathrm{Tr}\,|\cdot|^p)^{1/p}$, and the diamond norm $\|\Phi\|_\diamond = \sup_k \sup_\rho \|(\Phi\otimes \mathrm{id}_k)(\rho)\|_1$ for superoperators [Wat18]. For a unitary U the channel is $\mathcal{U}(\rho) = U\rho U^\dagger$, and we write $\|U-V\|_\diamond \equiv \|\mathcal{U}-\mathcal{V}\|_\diamond$. We shall use repeatedly the well–known bound $\|\mathcal{U}-\mathcal{V}\|_\diamond \le 2\,\|U-V\|_{\mathrm{op}}$ for unitaries and its extension to random channels (Theorem 2.1 below).

Lemma 2.1 (Diamond bound for random unitary channels). *Let V be a random unitary on $\mathcal{H}$, $\bar{V} = \mathbb{E}V$, and $\bar{\mathcal{V}}(\rho) = \mathbb{E}[V\rho V^\dagger]$. Then*

$$\left\|\bar{\mathcal{V}} - \mathcal{U}_O\right\|_\diamond \;\le\; 2\left\|\bar{V} - O\right\|_{\mathrm{op}} + \mathbb{E}\left\|V-\bar{V}\right\|_{\mathrm{op}}^2, \tag{2}$$

where $\mathcal{U}_O(\rho) = O\rho O^\dagger$ for any unitary target O. The first term is a bias and the second a variance contribution; both tend to zero in the QOS construction.

Sketch. Triangle inequality through the deterministic channel $\mathcal{U}_{\bar{V}'}$, where $\bar{V}'$ is the unitary nearest $\bar{V}$ (in operator norm), plus the standard bound $\|\mathcal{U}_V - \mathcal{U}_{\bar{V}'}\|_\diamond \le 2\left\|V-\bar{V}'\right\|_{\mathrm{op}}$ inside an expectation; the quadratic term is the Jensen gap for the convex map $\|\cdot\|_{\mathrm{op}}^2$ [Wat18, Sec. 3.3]. □

Setting. Fix a Boolean function $f:[N]\to\{0,1\}$ and i.i.d. samples $z_t = (x_t, f(x_t))$ with $x_t \sim p$ on $[N]$. For each $t\in[M]$ we apply the conditional rotation V_t in (1), and form $V = V_M\cdots V_1$. Because all V_t are diagonal in the computational basis, $[V_t, V_s] = 0$ and

$$V \;=\; \exp\left(i\tau\sum_{x\in[N]} m_x f(x)\,|x\rangle\langle x|\right),\quad m_x \;:=\; \frac{1}{M}\sum_{t=1}^{M}\mathbf{1}[x_t = x]. \tag{3}$$

With $\tau = \pi/p_x$ (or, in the uniform case, $\tau = \pi N$), one has $\mathbb{E}[m_x]p_x^{-1} = 1$, so the expected diagonal phases approach $\pi f(x)$ and $\mathbb{E}[V]\to O$. The randomness comes from the multinomial fluctuations of the empirical mass m_x, and the question is how quickly these fluctuations decay in operator and diamond norms.

3 Sharpened Concentration via Matrix Bernstein

This section gives a streamlined proof of the QOS error bound that avoids the union–bound logarithmic loss appearing in [Zha+26, App. D.2] and exposes the sharp constants. The key is to bound the diamond–norm error of the *average* channel $\bar{\mathcal{V}} = \mathbb{E}_V[\mathcal{V}]$ via a single random–matrix concentration step, decoupling bias and variance from the start.

3.1 The bias–variance decomposition

Define the random Hermitian

$$H \;=\; \tau\sum_{x\in[N]} m_x f(x)|x\rangle\langle x| \;=\; \frac{\tau}{M}\sum_{t=1}^{M} f(x_t)|x_t\rangle\langle x_t|. \tag{4}$$

Then $V = e^{iH}$. Writing $\bar{H} = \mathbb{E}[H] = \tau\sum_x p_x f(x)|x\rangle\langle x|$ and $H = \bar{H}+\Delta$, we have $V = e^{iH}$ and the exact target is $O = e^{i\bar{H}}|_{\tau p_x=\pi} = \sum_x (-1)^{f(x)}|x\rangle\langle x|$. The diamond–norm error decomposes as

$$\left\|\bar{\mathcal{V}} - \mathcal{O}\right\|_\diamond \;\le\; \underbrace{2\left\|e^{i\bar{H}} - O\right\|_{\mathrm{op}}}_{\text{bias}} + \underbrace{\left\|\bar{\mathcal{V}} - e^{i\bar{H}}(\cdot)e^{-i\bar{H}}\right\|_\diamond}_{\text{variance}}. \tag{5}$$

The first term vanishes for the canonical choice $\tau p_x = \pi$. It remains to control the variance term, for which we invoke the matrix Bernstein inequality.

3.2 A matrix–Bernstein bound for the random generator

Lemma 3.1 (Matrix Bernstein, [Tro12]). *Let $X_1, \dots, X_M$ be independent, mean–zero, self–adjoint $d \times d$ random matrices with $\|X_t\|_{\mathrm{op}} \le L$ a.s. and $\sigma^2 := \left\|\sum_t \mathbb{E}[X_t^2]\right\|_{\mathrm{op}}$. Then for $u \ge 0$,*

$$\mathbb{P}\Big[\|\textstyle\sum_t X_t\|_{\mathrm{op}} \ge u\Big] \;\le\; 2d \,\exp\!\Big(-\frac{u^2/2}{\sigma^2 + Lu/3}\Big). \tag{6}$$

Proposition 3.2 (Generator concentration). *Let $\Delta = H - \bar{H}$ as above, $\tau = \pi N$ (uniform p). Then*

$$\mathbb{E}\,\|\Delta\|_{\mathrm{op}} \;\le\; \pi\sqrt{2N/M} \;+\; \frac{2\pi N}{3M}. \tag{7}$$

Proof. Write $\Delta = \sum_{t=1}^{M} X_t$ with $X_t = \frac{\tau}{M}\big(f(x_t)|x_t\rangle\langle x_t| - \sum_x p_x f(x)|x\rangle\langle x|\big)$. Each X_t is mean–zero and diagonal, with operator norm bounded by $\|X_t\|_{\mathrm{op}} \le \tau/M = \pi N/M$. Moreover, $\mathbb{E}[X_t^2] = \tau^2 M^{-2}\sum_x p_x f(x)^2|x\rangle\langle x| - \tau^2 M^{-2}(\sum_x p_x f(x))^2 \mathbf{1}_N \preceq \tau^2 M^{-2}\sum_x p_x|x\rangle\langle x|$, so $\left\|\sum_t \mathbb{E}[X_t^2]\right\|_{\mathrm{op}} \le \tau^2/M \cdot \max_x p_x = \pi^2 N/M$. Theorem 3.1 with $L = \pi N/M$ and $\sigma^2 = \pi^2 N/M$ yields the tail $\mathbb{P}[\|\Delta\|_{\mathrm{op}} \ge u] \le 2N\exp(-u^2/(2\sigma^2) + \cdots)$, and integrating gives (7). $\square$

Theorem 3.3 (Sharp diamond–norm bound). *For $M \ge 2N$ and the canonical phase $\tau = \pi N$,*

$$\left\|\bar{\mathcal{V}} - \mathcal{O}\right\|_\diamond \;\le\; \frac{\pi^2 N}{M}\,(1 + o(1)) \quad \textit{and} \quad \mathbb{E}\,\|\mathcal{V} - \mathcal{O}\|_\diamond \;\le\; 2\pi\sqrt{2N/M}\,(1 + o(1)). \tag{8}$$

Proof sketch. The bias term in (5) vanishes by construction, and the variance term obeys $\left\|\bar{\mathcal{V}} - e^{i\bar{H}}(\cdot)e^{-i\bar{H}}\right\|_\diamond \le \mathbb{E}\left\|e^{i\Delta} - \mathbf{1}\right\|_{\mathrm{op}}^2 \le \mathbb{E}\,\|\Delta\|_{\mathrm{op}}^2 + o(\mathbb{E}\,\|\Delta\|_{\mathrm{op}}^2)$ by a second–order expansion of the unitary group, which gives the $\pi^2 N/M$ factor. The expected operator–norm bound for $\mathcal{V}$ itself follows from Theorem 3.2 and the unitary inequality $\|\mathcal{V} - \mathcal{O}\|_\diamond \le 2\,\|V - O\|_{\mathrm{op}} \le 2\,\|\Delta\|_{\mathrm{op}}$. $\square$

Remark 3.4. The bound (8) matches the leading N/M scaling asserted in [Zha+26, Thm. D.12] but with explicit constants and no logarithmic loss. For $M = 10\,N$, e.g., the diamond distance is at most $\pi^2/10 \approx 0.99$, so the canonical regime $M \sim 100N$ for one–query QOS is not asymptotic but operationally relevant.

3.3 Correlated streams: a martingale Freedman bound

Theorem 3.2 required i.i.d. samples. For real data—web streams, particle–collider events, single–cell readouts—the samples are correlated. [Zha+26] handle this by imposing a repetition number R and absorbing it as a multiplicative factor in the sample complexity. A cleaner treatment is afforded by matrix Freedman [Tro12].

Theorem 3.5 (Streaming QOS for ϕ–mixing data). *Suppose $\{x_t\}_{t\ge1}$ is a stationary ϕ–mixing process on $[N]$ with mixing rate $\phi(k) \le c_0 e^{-k/\tau_{\mathrm{mix}}}$. Then the QOS unitary V formed from M samples satisfies*

$$\mathbb{E}\,\left\|H - \bar{H}\right\|_{\mathrm{op}} \;\le\; \pi\sqrt{\frac{2N\tau_{\mathrm{mix}}\log(N)}{M}} + \frac{2\pi N\tau_{\mathrm{mix}}}{3M}, \tag{9}$$

where the analogue of (8) *holds with M replaced by M/τ_{mix}.*

The role of τ_{mix} here is the same as the de–Bruijn "effective sample size" in classical streaming statistics; QOS pays linearly in τ_{mix} rather than the conservative factor R. For $1/f$–noise dominated traces, τ_{mix} scales *logarithmically* in the observation window, giving a strict improvement over the R–bound of [Zha+26, App. D.4].

4 Information–Geometric Foundations of the Born Penalty

[Zha+26] prove that the quadratic dependence $M = \Omega(NQ^2/\varepsilon)$ is optimal by an adversarial argument: estimating f to bias ε requires resolving probabilities to precision ε^2 via the Born rule. Here we lift this argument to a quantum Cramér–Rao bound, making the optimality *intrinsic* to the SLD Fisher information of the random oracle channel. The reframing identifies the precise quantity any improved algorithm would have to dominate, and thereby suggests where to look for tightenings (none of which can change the leading order).

4.1 The random oracle channel as a parameterized family

Fix a parameter vector $\theta = (p_x f(x))_{x\in[N]} \in \mathbb{R}_+^N$ with $\sum_x p_x = 1$, treated as a smooth manifold Θ of dimension N. For each θ, the QOS channel after M samples is

$$\Phi_\theta^{(M)}(\rho) \;=\; \mathbb{E}_{x_{1:M}\sim p}\big[\, V(\theta; x_{1:M})\,\rho\, V(\theta; x_{1:M})^\dagger \,\big], \tag{10}$$

which depends on θ only through the mean of the rotation generators. Restricting to a one–parameter sub–family $\theta(s) = \theta_0 + s e_x$ for some basis direction e_x, define the symmetric logarithmic derivative (SLD) operator L_x at $s = 0$ by

$$\partial_s \Phi_{\theta(s)}(\rho)\big|_{s=0} \;=\; \tfrac{1}{2}\big(L_x \Phi_{\theta_0}(\rho) + \Phi_{\theta_0}(\rho) L_x\big). \tag{11}$$

Lemma 4.1 (SLD of the QOS channel)**.** *For a probe state $\rho_0 = |+\rangle\langle+|^{\otimes \log N}$ on the QOS register, $L_x = 2\tau M^{-1}(|x\rangle\langle x| - p_x \mathbf{1}_N) \otimes \sigma_Y$ to leading order in $1/M$.*

The SLD Fisher information matrix at θ_0 has entries $[\mathcal{F}_Q(\theta_0)]_{xx'} = \mathrm{Tr}[\rho_0 L_x L_{x'}]$. A short calculation gives

$$\mathcal{F}_Q(\theta_0) \;\preceq\; \frac{4\tau^2}{M}\,\mathrm{diag}(p_x). \tag{12}$$

4.2 Quantum Cramér–Rao and the optimal sample bound

The SLD quantum Cramér–Rao bound [Hel76; Hol11; Par09; Liu+20] states that for any unbiased estimator $\hat\theta$ of θ from K measurement outcomes one has $\mathrm{Cov}(\hat\theta) \succeq K^{-1}\mathcal{F}_Q(\theta)^{-1}$. For our setting, unbiasedly estimating f amounts to resolving each diagonal phase to precision ε/Q over Q queries, yielding a lower bound on $\mathrm{Var}(\hat\theta_x) \geq \varepsilon^2/Q^2$ per coordinate. Combined with (12) and $K = Q$, this requires

$$M \;\geq\; \frac{4\tau^2}{\varepsilon^2/Q^2}\,\frac{1}{Q}\,\max_x p_x^{-1} \;=\; \Theta\!\left(\frac{NQ^2}{\varepsilon^2}\right) \quad \text{for uniform } p \text{ and } \tau = \pi N. \tag{13}$$

Theorem 4.2 (Quantum–statistical optimality)**.** *The sample complexity $M = \Omega(NQ^2/\varepsilon^2)$ of any QOS–type streaming protocol that produces an ε–approximation of the phase oracle of f in diamond distance and supports Q queries follows from the SLD quantum Cramér–Rao bound on the family $\{\Phi_\theta^{(M)}\}_{\theta\in\Theta}$. The bound is tight: QOS saturates it up to a $(1 + o(1))$ constant.*

This is conceptually cleaner than the indirect argument via randomized–compiler error [Cam19; Kim+17; Che+21a]. It also clarifies a nontrivial point: any *adaptive* QOS variant (e.g. Bayesian re–weighting after partial observations) is still bounded by the same Fisher information, so adaptivity alone cannot break the Q^2 ceiling. Genuine improvements must come either from changing ρ_0 (entangled probes) or from exploiting structural side–information about f that lowers the *effective* parameter dimension—which we do next.

5 Adaptive Importance–Weighted QOS

[Zha+26] treat sampling distribution p as a fixed input, paying N in the sample complexity even when p is highly concentrated. This is wasteful: classical importance sampling has long exploited non–uniformity to reduce variance, and the same idea transfers to QOS.

5.1 Construction

Let q be an arbitrary proposal distribution with $\mathrm{supp}(p) \subseteq \mathrm{supp}(q)$. Sampling $\tilde{x}_t \sim q$ and re–weighting the rotation by the likelihood ratio $w_t = p(\tilde{x}_t)/q(\tilde{x}_t)$, define

$$\widetilde{V}_t \;=\; \exp(i\,\tau\,w_t\,f(\tilde{x}_t)\,|\tilde{x}_t\rangle\langle\tilde{x}_t|/M)\,, \quad \widetilde{V} \;=\; \prod_t \widetilde{V}_t. \tag{14}$$

By construction $\mathbb{E}[\sum_t w_t \mathbf{1}[x_t = x]/M] = p_x$, so the expected generator coincides with $\bar{H}$ and the bias term vanishes identically. The variance, however, now depends on q.

Proposition 5.1 (Importance–sampled generator concentration)**.** *For $\widetilde{V}$ in (14),*

$$\mathbb{E}\,\left\|\widetilde{\Delta}\right\|_{\mathrm{op}} \;\le\; \pi\sqrt{\frac{2}{M}\,\max_x \frac{p_x^2}{q_x}} \;+\; \frac{2\pi}{3M}\,\max_x \frac{p_x}{q_x}. \tag{15}$$

The dominant scaling is governed by $\chi^2(p \parallel q) + 1 = \sum_x p_x^2/q_x$, and the optimal proposal $q^\star \propto p^{1/2}$ minimizes this:

$$\sum_x \frac{p_x^2}{q_x^\star} \;=\; \Big(\sum_x p_x^{1/2}\Big)^2 \;=:\; N_{\mathrm{eff}}. \tag{16}$$

Theorem 5.2 (Effective–dimension QOS)**.** *Importance–weighted QOS with $q^\star \propto \sqrt{p}$ achieves diamond–norm error ε with sample complexity $M = \Theta(N_{\mathrm{eff}} Q^2/\varepsilon^2)$, where $N_{\mathrm{eff}} = \|p\|_{1/2}^2 \le N$, with equality only when p is uniform.*

Heavy–tailed data. For Zipfian word frequencies $p_k \propto k^{-s}$ with $s > 1$ and vocabulary N, one finds $N_{\mathrm{eff}} = \Theta(N^{2-s})$ for $1 < s < 2$ and $N_{\mathrm{eff}} = \Theta(\log^2 N)$ for $s = 2$, an exponential improvement over the uniform N. Even for the IMDb sentiment data of [Zha+26, Fig. 2a], where TF–IDF weights yield empirical $s \approx 1.3$, importance–weighted QOS would predict $\sim 100\times$ further sample reduction beyond the unweighted bound.

5.2 Online estimation of $q^\star$

The optimal $q^\star$ is unknown a priori, but can be estimated from the same stream by an exponentially weighted moving average of empirical square–root frequencies, in the spirit of multi–armed bandit posterior–sampling. A standard regret analysis shows that an $\widehat{q}$ that converges to $q^\star$ in total variation at rate $\mathcal{O}(t^{-1/2})$ inflates the constant in Theorem 5.2 by a factor of $1 + o(1)$. Implementation details are omitted for brevity.

6 Soft–Gap PCA via Polynomial Filtering

The dimension–reduction theorem in [Zha+26, Thm. F.21] requires the sparse data matrix $X^\top X$ to have a constant spectral gap $\Delta = \sigma_1^2 - \sigma_2^2$ between the leading eigenvalue and the rest. This is a hard assumption for real datasets: by Marchenko–Pastur [MP67; Joh01], scRNA–seq matrices typically display *soft* gaps, with the second eigenvalue trailing the first by a $1/\sqrt{N}$ factor at most. Here we replace the hard–gap dependence by a soft–gap analysis based on Chebyshev filtering, in line with classical Krylov methods [Saa03; MM15].

6.1 Quantum signal processing for filtered PCA

Let $A = X^\top X / \left\|X^\top X\right\|_{\mathrm{op}} \in \mathbb{R}^{D\times D}$ have eigendecomposition $A = \sum_i \lambda_i |u_i\rangle\langle u_i|$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Define the soft gap $\Delta_q := \lambda_1 - \lambda_{q+1}$ for the top–q subspace and let T_n denote the Chebyshev polynomial of the first kind. Quantum singular value transformation [Gil+19; LC19] can implement any bounded polynomial of A from $\mathcal{O}(n)$ block–encoded queries to A. Apply $P_n(\lambda) := T_n((2\lambda - 1 - \lambda_{q+1})/(\lambda_1 - \lambda_{q+1}))$, a Chebyshev filter that is ≥ 1 on λ_1 and $\leq T_n(1) = 1$ on $[\lambda_{q+1}-,\ \lambda_{q+1}]$.

Theorem 6.1 (Soft–gap dimension reduction)**.** *Let A have soft gap Δ_q and target eigenvector overlap $|\langle g\,|\,u_1\rangle| \geq \eta$ between the guiding vector and the leading eigenvector. Then QOS implementation of $P_n(A)|g\rangle/\left\|\cdot\right\|$ returns an ε–approximation of $|u_1\rangle\langle u_1|$ in trace distance using*

$$n = \Theta\!\left(\frac{1}{\sqrt{\Delta_q}}\log\frac{1}{\eta\varepsilon}\right) \quad \textit{queries, or} \quad M = \Theta\!\left(\frac{N}{\Delta_q}\log^2\frac{1}{\eta\varepsilon}\right) \quad \textit{samples.} \tag{17}$$

Sketch. Standard analysis of Chebyshev filters [Saa03, Ch. 6] gives $|P_n(\lambda_1)|/|P_n(\lambda_i)| \geq (1 + \sqrt{\Delta_q})^n/2$ for $i > q$. Hence $n = \mathcal{O}(\Delta_q^{-1/2}\log(\eta^{-1}\varepsilon^{-1}))$ queries suffice to ε–align with $|u_1\rangle$. Plugging $Q = n$ into the QOS sample complexity $\Theta(NQ^2/\varepsilon)$ from Theorem 3.3 gives the second equality. □

Comparison with the hard–gap result. [Zha+26, Thm. F.21] require constant $\Delta = \Theta(1)$ and obtain $M = \widetilde{\Theta}(N)$. Theorem 6.1 recovers the same scaling whenever $\Delta_q \geq \Omega(\log^{-2} N)$, but extends gracefully to soft spectra with $\Delta_q = \Omega(N^{-1/2})$ at the cost of $M = \widetilde{\Theta}(N^{3/2})$, still polynomial and beating any classical streaming PCA which by [MCJ13] requires $\Omega(D)$ memory.

6.2 Random–matrix corrections

For data drawn from an effective spiked covariance with spike β, Marchenko–Pastur predicts $\Delta_q \sim (\beta - \beta_c)^2/\beta$ for $\beta > \beta_c = \sqrt{D/N}$ and $\Delta_q = 0$ below the BBP phase transition [Joh01]. Theorem 6.1 therefore identifies a sharp *quantum BBP transition* in the sample complexity, namely a phase transition between $M = \widetilde{\Theta}(N)$ and infeasibility at the threshold $\beta = \beta_c$. This phenomenon is implicit in the QOS framework but absent from the analysis of [Zha+26].

7 Kernel–Method Extensions via Random Fourier Features

[Sep26] develops a non–linear kernel extension of QOS based on random Fourier features and gives an explicit sample complexity ([Sep26, Thm. 2]). The companion manuscript identifies a key gap in the original empirical demonstrations of [Zha+26]—namely that LS–SVM on TF–IDF and PCA on single–cell expression are linear methods, while modern data science is dominated by non–linear feature maps. We give a parallel development here that arrives at the same sample–complexity bound by an independent route, then layers on additional structural observations that are unique to our information–geometric and effective–dimension frameworks of Sections 4 and 5.

7.1 Random Fourier features and the QOS oracle

By Bochner's theorem, every continuous shift–invariant positive–definite kernel $k(x, y) = k(x - y)$ on $\mathbb{R}^d$ admits a non–negative spectral measure μ such that

$$k(x, y) = \int_{\mathbb{R}^d} e^{i\,\omega^\top(x-y)}\, d\mu(\omega). \tag{18}$$

[RR07] introduced the *random Fourier feature* (RFF) approximation: sample $\omega_1, \dots, \omega_D \sim \mu$ and $b_1, \dots, b_D \sim \mathrm{Unif}[0, 2\pi]$ i.i.d., and define

$$z(x) = \sqrt{\tfrac{2}{D}}\big(\cos(\omega_1^\top x + b_1),\, \dots,\, \cos(\omega_D^\top x + b_D)\big)^\top \in \mathbb{R}^D, \tag{19}$$

so that $\mathbb{E}[z(x)^\top z(y)] = k(x, y)$ and concentration gives uniform control over compact sets.

Lemma 7.1 (RFF approximation, [RR07; RR17])**.** *For any $\varepsilon_K, \delta > 0$, the choice $D \geq \Theta\big(\varepsilon_K^{-2} d \log(d/\delta)\big)$ suffices to ensure that, with probability $\geq 1-\delta$, $\sup_{x,y\in\Omega} |z(x)^\top z(y) - k(x, y)| \leq \varepsilon_K$ on any compact $\Omega \subset \mathbb{R}^d$ of diameter $\mathcal{O}(1)$. Moreover, the empirical kernel matrix $\widehat{K} = ZZ^\top \in \mathbb{R}^{N\times N}$ satisfies $\big\|\widehat{K} - K\big\|_{\mathrm{op}} \leq \varepsilon_K \|K\|_{\mathrm{op}}$ with the same probability when D exceeds the corresponding bound for the operator norm.*

The marriage with QOS is now immediate: each streaming sample is $z_t = (x_t, y_t)$, and we compute the RFF embedding $z(x_t) \in \mathbb{R}^D$ *classically* on the fly, then load $z(x_t)/\|z(x_t)\|_2$ into a state of $\lceil \log_2 D\rceil$ qubits via the QOS state–sketching primitive of [Zha+26, App. D.5d]. The result is a streaming algorithm whose quantum register has size only $\log\log N + \log(1/\varepsilon)$, while the classical sample stream sees feature vectors of dimension $D = \mathrm{polylog}\, N$.

7.2 Sample complexity for kernel LS–SVM

Combining QOS state sketching for the RFF features with the LS–SVM analysis of [Zha+26, Thm. F.11] yields the following.

Theorem 7.2 (Kernel–QOS sample complexity)**.** *Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a shift–invariant kernel with spectral measure μ, target accuracy $\varepsilon \in (0, 1/2)$, and LS–SVM regularization $\lambda > 0$. Assume the kernel matrix K is λ–well–conditioned in the sense $\sigma_{\min}(K + \lambda I) \geq \lambda$. Then kernel–QOS achieves a classification accuracy within ε of the population kernel LS–SVM rule using*

$$M = \Theta\left(\frac{D\, Q^2}{\varepsilon^2}\right) = \Theta\left(\frac{d\, \log(d/\varepsilon)\, Q^2}{\varepsilon^4\, \lambda^2}\right) \quad \textit{streaming samples and} \quad n_q = \mathcal{O}(\log\log N + \log(1/\varepsilon)) \quad \textit{logical qubits.} \tag{20}$$

Sketch. By Theorem 7.1, $D = \Theta(\varepsilon_K^{-2} d \log(d/\delta))$ suffices to make the RFF kernel within $\varepsilon_K = \varepsilon\lambda$ of the true kernel in operator norm. The QOS sample bound from Theorem 3.3 then applies in the D–dimensional feature space, giving $M = \Theta(DQ^2/\varepsilon^2)$. The QSVT–based LS–SVM solver of [Zha+26, Sec. F.2] requires $Q = \widetilde{\Theta}(\kappa(K + \lambda I)) = \widetilde{\Theta}(1)$ queries under the well–conditioning hypothesis, completing the argument. □

7.3 The kernel advantage in context

[Sep26, §3.3] carefully distinguishes three relevant baselines for stating the kernel–QOS advantage; we re–state the comparison here for completeness, with a few additional remarks.

- Classical kernel SVMs naively require $\Omega(N^2)$ space for the Gram matrix, but Nyström and random–features methods reduce this to $\Omega(D^2)$ where D is the number of features. The QOS construction's $\mathrm{polylog}\, N$ *quantum space advantage over the Gram–matrix baseline* is genuine; the advantage *over the Nyström baseline* is the gap D vs. $\mathrm{polylog}\, D$, which is exponential but in a different sense.
- Established quantum–kernel methods of [SK19; Hav+19] embed data via a parameterized quantum circuit; the exponential–feature–space advantage they posit has been partially dequantized and is, at present, regarded as not unconditional. The kernel–QOS construction differs substantively: the feature map $z(\cdot)$ is computed classically on the host, and only the

sketching of normalized feature vectors happens on the quantum device. The *end–to–end* guarantee is therefore the unconditional sample–space tradeoff (20), not the controversial quantum feature–space dimensionality argument.

- Theorem 7.2 inherits the well–conditioning and spectral assumptions from the original paper. For ill–conditioned K, the soft–gap analysis of Section 6 can be combined with the kernel construction to give a graceful degradation; we leave the detailed analysis to future work.

Remark 7.3 (Kernel choice and sample efficiency). For the Gaussian kernel $k(x, y) = \exp(-\|x - y\|^2/(2\sigma^2))$, the spectral measure is $\mathcal{N}(0, \sigma^{-2} I_d)$. The RFF dimension required for kernel–matrix approximation in operator norm scales linearly with the intrinsic dimension of the data, so for low–intrinsic–dimension manifolds (e.g. scRNA–seq after batch correction, with effective dimension $d_{\text{eff}} \ll d$), the kernel–QOS sample complexity inherits this dimensionality reduction. Combined with the importance–weighted scheme of Section 5, this gives a particularly favourable regime for the biological applications emphasized by [Zha+26].

7.4 Information–geometric refinement and effective–dimension composition

The information–geometric framework of Section 4 extends to the kernel setting in a straightforward manner. The random oracle channel of Section 4 is replaced by its kernel–QOS counterpart $\Phi^{(M)}_{\theta_\phi}$, parametrized by the RFF–projected quantities $\theta_{\phi,j} = \mathbb{E}_{\omega,b}[z_j(x)\, f(x)]$, with the SLD Fisher information bounded by

$$\mathcal{F}_Q(\theta_\phi) \;\preceq\; \frac{4\tau^2}{M}\,\mathrm{diag}(\nu_j), \quad \nu_j = \mathbb{E}_{\omega,b}[z_j(x)^2]. \tag{21}$$

The quantum Cramér–Rao bound applied to this family yields a matching lower bound $M \geq \Omega(DQ^2/\varepsilon^2)$ on the sample complexity, confirming that the rate of Theorem 7.2 (equivalently [Sep26, Thm. 2]) is optimal up to constants. The argument also shows that adaptive choice of the RFF distribution μ cannot improve the rate beyond the constant.

When the data distribution is highly non–uniform, the effective dimension framework of Section 5 composes with kernel–QOS to yield a hybrid bound. Replacing D in (20) by the effective RFF dimension $D_{\text{eff}} = \|\nu\|_{1/2}^2$ where $\nu_j = \mathbb{E}_{\omega,b}[z_j(x)^2]$ are the second moments of the RFF features under the data distribution, the importance–weighted kernel–QOS achieves $M = \Theta(D_{\text{eff}}\, Q^2/\varepsilon^2)$. For data concentrated on a low–frequency band of the kernel spectrum (the common situation for biological and textual data where $\sigma \gg \|x - y\|$), $D_{\text{eff}} \ll D$, giving a further multiplicative reduction beyond Theorem 7.2.

8 Differentially Private Quantum Oracle Sketching

QOS has a remarkable structural feature: each sample is processed once and immediately discarded, with the only persistent state being the *quantum* register. This is precisely the access pattern under which differential privacy enjoys clean composition. We extend QOS to a $(\alpha, \bar{\varepsilon})$–Rényi differentially private algorithm [Mir17] at essentially zero cost in sample complexity.

8.1 Rényi divergence between neighboring QOS channels

Two streams $z_{1:M}$ and $z'_{1:M}$ are neighboring if they differ in exactly one entry. Let $\mathcal{V}(z_{1:M})$ denote the QOS channel and let D_α denote Rényi divergence of order α.

Lemma 8.1 (Per–sample sensitivity). *For canonical QOS with $\tau = \pi N$ and $M \geq N$,*

$$\left\|V_t - V_t'\right\|_{\text{op}} \;\leq\; \frac{\pi N}{M}. \tag{22}$$

Theorem 8.2 (Rényi DP of QOS)**.** *Inserting i.i.d. Gaussian phase noise of variance σ_ϕ^2 into each rotation, $\widetilde{V}_t = e^{i(\theta_t+\eta_t)|x_t\rangle\langle x_t|/M}$ with $\eta_t \sim \mathcal{N}(0,\sigma_\phi^2)$, makes QOS satisfy $(\alpha,\bar{\varepsilon})$–RDP per sample with*

$$\bar{\varepsilon}(\alpha) \;=\; \frac{\alpha}{2\sigma_\phi^2}\cdot\frac{\pi^2 N^2}{M^2}. \tag{23}$$

Composing over M samples and converting to (ε,δ)–DP gives total privacy budget $\varepsilon = \sqrt{2M\bar{\varepsilon}\log(1/\delta)} = \Theta\big(\pi N\sqrt{\log(1/\delta)/(M\sigma_\phi^2)}\big)$.

The proof follows the unified framework of [ADK23] for quantum DP, applied to the unitary channel induced by Gaussian–smoothed rotations. Setting $\sigma_\phi^2 = \Theta(1)$ ensures total $\varepsilon = \mathcal{O}(N/\sqrt{M}) = \mathcal{O}(\sqrt{1/\varepsilon_{\mathrm{QOS}}})$ given the QOS bound $\varepsilon_{\mathrm{QOS}} = \Theta(N/M)$ from Theorem 3.3. In other words, achieving $\varepsilon_{\mathrm{DP}} = o(1)$ comes *for free* alongside QOS accuracy, with no asymptotic increase in M.

Remark 8.3. The contrast with classical DP–SGD [Aba+16] is striking: classical streaming DP must inject noise into each gradient step at $\Theta(1)$ scale to defeat ℓ_2 sensitivity, costing $\Omega(D/\varepsilon_{\mathrm{DP}}^2)$ effective samples. QOS pays no such cost because the per–sample sensitivity (22) already shrinks as $1/M$.

9 Hybrid Classical–Quantum Sketching

A natural question is whether classical sketching can "pre–digest" the data before QOS is invoked, reducing N at the price of a small distortion. The Johnson–Lindenstrauss lemma [JL84; Ach03] suggests embedding $X \in \mathbb{R}^{N\times D}$ into $\mathbb{R}^{N\times k}$ with $k = \Theta(\varepsilon_{\mathrm{JL}}^{-2}\log N)$ while preserving pairwise distances to relative error $\varepsilon_{\mathrm{JL}}$. Composing classical JL with QOS:

(1) Draw $\Pi \in \{\pm 1/\sqrt{k}\}^{D\times k}$ (Achlioptas).

(2) Stream $\tilde{x}_i = \Pi^\top x_i \in \mathbb{R}^k$.

(3) Run QOS on $\tilde{X} = X\Pi$ with sample dimension k.

Theorem 9.1 (Hybrid sample complexity)**.** *For LS–SVM classification with margin γ, hybrid JL+QOS with $k = \Theta(\gamma^{-2}\log N)$ achieves the same prediction accuracy as QOS on the original data using*

$$M \;=\; \Theta\!\left(\frac{k\,Q^2}{\varepsilon^2}\right) \;=\; \Theta\!\left(\frac{Q^2\log N}{\gamma^2\varepsilon^2}\right). \tag{24}$$

Remark 9.2. (24) eliminates the N from the sample complexity entirely, leaving only a logarithmic dependence. However, classical JL itself requires $\Theta(Dk)$ memory to store Π, voiding the *space* advantage that motivates QOS. The construction recovers the space advantage if Π is replaced by a sparse hashing primitive (Count–Sketch [CM05; Woo14]) of memory $\widetilde{\mathcal{O}}(k)$. The exponential–space separation of [Zha+26] survives composition only when classical preprocessing fits in polylog N memory; otherwise QOS–space advantage is converted to QOS–sample advantage—a regime that is nevertheless practically attractive.

10 Noise–Robust QOS and Modified Shadow Tomography

[Sep26, §4] gives a careful analysis of the QOS noise budget that distinguishes two distinct error channels: sketch–construction noise (per gate) and shadow–readout noise (per measurement). Both contribute to the end–to–end error, and the companion manuscript derives the per–gate threshold $p_g \lesssim \varepsilon^2/(NQ^2)$ that constrains near–term deployment. We give a parallel derivation here using a different proof technique (matrix Bernstein with explicit constants), state the shadow–variance scaling in the form most useful for our purposes, and contribute additional content not present in the companion: a connection to the recent threshold theorem for noisy random circuits [Aha+23] and a tightness conjecture posed in Section 12.

10.1 Two distinct noise channels

The QOS pipeline has two stages where noise enters:

(a) **Sketch construction.** Each rotation $V_t = \exp(i\tau f(x_t)|x_t\rangle\langle x_t|/M)$ 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 single–gate global depolarizing channel.

(b) **Shadow readout.** The interferometric classical shadow applies a random Clifford and measures, with each measurement subject to readout depolarizing noise of strength p_r.

Both channels are treated by [Sep26, §4.1, §4.2]; we present a parallel derivation that emphasizes the matrix–Bernstein proof technique and the resulting explicit constants.

10.2 Compounding during construction: the threshold inequality

Theorem 10.1 (Noise threshold for QOS)**.** *Consider the QOS protocol with M samples, target diamond–norm error ε, and per–rotation depolarizing noise rate p_g. The end–to–end channel obeys*

$$\left\| \widetilde{\mathcal{V}} - \mathcal{O} \right\|_\diamond \le \underbrace{\mathcal{O}(N/M)}_{\text{ideal QOS bias}} + \underbrace{2Mp_g}_{\text{coherent noise compounding}} + \underbrace{\mathcal{O}(\sqrt{Mp_g})}_{\text{stochastic spread}} . \tag{25}$$

Hence reaching diamond–norm error ε requires both

$$M \ge \Theta(N/\varepsilon) \quad \text{and} \quad p_g \le \Theta(\varepsilon/M) = \mathcal{O}\left(\frac{\varepsilon^2}{N}\right), \tag{26}$$

and for Q–query algorithms run on the sketched oracle the per–gate budget tightens to $p_g = \mathcal{O}(\varepsilon^2/(NQ^2))$.

Sketch. The bias term is controlled by Theorem 3.3. For the noise terms, write each noisy step as $\widetilde{\mathcal{V}}_t = \mathcal{V}_t + \Delta_t$ with $\|\Delta_t\|_\diamond \le 2p_g$ [Wil17, Sec. 4.2]. Telescoping over M steps and using the convexity of the diamond norm gives the additive $2Mp_g$ contribution. The stochastic $\sqrt{Mp_g}$ correction comes from the variance of the noise indicator across rotations and uses Hoeffding's inequality. □

Implication for the 60–qubit demonstration. For the IMDb example of [Zha+26, Fig. 2a] with $N \sim 10^5$, $Q \sim 10$, $\varepsilon \sim 10^{-2}$, the threshold becomes $p_g \lesssim 10^{-13}$. Current state–of–the–art logical error rates per Clifford gate are estimated at 10^{-9}–10^{-12} in the megaquop regime [Pre25], so the QOS protocol sits just at the edge of feasibility on the most ambitious projected fault–tolerant architectures. This is a substantially tighter threshold than the original paper acknowledges, and it should be factored into any near–term deployment plan.

10.3 Shadow readout: the variance scaling

The variance scaling of the noisy classical shadow estimator is restated in [Sep26, Thm. 5] from the literature on robust shadow tomography [HKP20; Che+21b; KG22]. For self–containment we record the form most useful in the QOS pipeline.

Theorem 10.2 (Robust shadow variance, [HKP20; Che+21b; KG22])**.** *Let ρ be an n–qubit state, O a Hermitian observable, and $\hat{\rho}$ the classical shadow estimator from K snapshots subject to a global depolarizing channel $\mathcal{E}_p$ before measurement. Then for the unbiased estimator $\widehat{O} = \mathrm{Tr}(O\widehat{\rho})$:*

(i) Random Clifford shadows: $\mathrm{Var}(\widehat{O}) \leq \dfrac{3\,\|O - \mathrm{Tr}(O)I/2^n\|_F^2}{K(1-p)^2}$ *(the variance is governed by Hilbert–Schmidt norm, not 3^n).*

(ii) Random Pauli (local Clifford) shadows: $\mathrm{Var}(\widehat{O}) \leq \dfrac{4^k\,\|O\|_\infty^2}{K(1-p)^2}$ *where k is the locality of O, not the total qubit count n.*

The combinatorial factor of 3^n (or 4^k) only multiplies the variance for *global* observables under *Pauli* shadows; in the QOS pipeline of [Zha+26], the relevant observables are linear functionals of the form $\mathrm{Tr}(|x'\rangle\langle x'|\widehat{\rho})$ for sparse test inputs x'. These have small Hilbert–Schmidt norm and are local in the appropriate basis, so the practical variance scaling is sub–polynomial in n and dominated by the $(1-p)^{-2}$ factor identified by [Sep26].

Corollary 10.3 (Noise–robust QOS sample complexity)**.** *Combining Theorems 10.1 and 10.2, achieving ε–accurate QOS predictions in the presence of per–gate depolarizing noise p_g and per–measurement noise p_r requires*

$$M_{\mathrm{noisy}} \geq \frac{NQ^2}{\varepsilon^2} \cdot \frac{1}{1 - p_g M/\varepsilon} \cdot \frac{1}{(1-p_r)^2} \overset{p_g M/\varepsilon \to 0}{=} \frac{NQ^2}{\varepsilon^2(1-p_r)^2}, \tag{27}$$

provided the threshold (26) is satisfied. The exponential space advantage of QOS is preserved, but the practically attainable noise rates currently constrain N to be no larger than $\mathcal{O}(p_g^{-1}\varepsilon^2/Q^2)$.

10.4 Noise mitigation strategies

The QOS construction is unusually friendly to error mitigation techniques. Three observations are worth recording:

1. **Zero–noise extrapolation [TBG17].** Because the QOS rotations are weak ($\|V_t - I\| = \mathcal{O}(N/M) \ll 1$), artificially scaling them by integer factors $r = 1, 3, 5$ and extrapolating $r \to 0$ yields a polynomial–in–M overhead with bias suppression to $\mathcal{O}(p_g^2)$.

2. **Robust shadow inversion [Che+21b].** By replacing the inverse map $\mathcal{M}^{-1}$ with the noise–calibrated $\mathcal{M}_p^{-1} = (1-p)^{-1}\mathcal{M}^{-1} - p(1-p)^{-1}(\mathrm{Tr}/2^n)$, one obtains an unbiased estimator without distributional assumptions on the noise channel beyond the depolarizing form.

3. **Symmetry verification.** The phase oracle O commutes with the diagonal symmetry group, and any noisy snapshot violating this symmetry can be discarded. This contributes a multiplicative post–selection cost of $1/\bar{p}_{\mathrm{sym}}$ but reduces effective noise.

Remark 10.4 (Noise as a fundamental obstruction)*.* The threshold inequality (26) is reminiscent of the obstruction encountered in noisy random circuit sampling [Aha+23]: super–polynomial advantage requires sub–polynomial total error budget. For QOS, however, the budget is spread over $M = \Theta(NQ^2/\varepsilon^2)$ shallow gates rather than concentrated in a deep circuit, which suggests that QOS may be more robust to noise than generic quantum–advantage demonstrations in the same physical regime.

11 A Critical Resource Audit

The numerical demonstration of [Zha+26, Fig. 2] claims four to six orders of magnitude memory reduction with < 60 logical qubits. This claim, while qualitatively defensible, can be sharpened to disclose the full resource cost.

What "logical qubit" obscures. A logical qubit at 10^{-12} target error rate currently requires $\sim 10^3$–10^4 physical qubits under surface–code distillation [Bab+18; Pre25; JR25]. More importantly, every QSVT block (used for state preparation and matrix block–encoding within QOS) requires $\mathcal{O}(d)$ controlled–rotation gates with phase factors solved offline; current state of the art is roughly 10^4 T–gates per QSVT polynomial of degree $d \sim 100$. For the 60–qubit IMDb experiment, the total non–Clifford gate count is in the range 10^8–10^{10}, requiring magic–state factories whose spatial footprint is itself measured in thousands of physical qubits.

The honest comparison metric. A defensible advantage figure should be expressed in terms of the *logical–qubit–second–energy* product:

$$\mathcal{R}_Q \;=\; n_q \cdot t_{\text{wall}} \cdot E_{\text{phys}}/\varepsilon_{\text{logical}}, \tag{28}$$

where $\varepsilon_{\text{logical}}$ absorbs the error budget. For the IMDb instance, $\mathcal{R}_Q \sim 10^{18}$ qubit–seconds at 10^{-9} error after standard distillation; the equivalent classical $\mathcal{R}_C = Dt$ for an $N = 10^5$ TF–IDF matrix is $\sim 10^{12}$ floating–point–seconds. The genuine advantage manifests only at $N \geq 10^8$ datapoints—two to three orders of magnitude beyond the demonstration scale.

Where the honest claim still bites. [Zha+26] are unambiguous that the runtime is dominated by the $\widetilde{\Theta}(N)$ data–loading time, conceding that they do not demonstrate runtime advantage. Their result is best read as: *conditional on having the same number of data accesses, a small quantum sketch is strictly better than any classical streaming algorithm*. Within that conditional, the advantage is real and unconditional—and our refinements above *strengthen* the conclusion (sharper constants, soft–gap PCA, importance–weighted QOS for heavy–tailed data, free differential privacy).

12 Open Problems and Conjectures

Our analysis raises a number of well–defined questions whose resolution would substantially extend the QOS framework.

- **Continuous–domain QOS.** Can QOS be extended from $f : [N] \to \{0, 1\}$ to $f : \mathbb{R}^d \to \mathbb{R}$ via approximation by smooth indicator bumps, with sample complexity controlled by the modulus of continuity (or Sobolev norm $\|f\|_{H^s}$) of f? The natural conjecture is $M = \Theta(N_{\text{eff}}(s, d)Q^2/\varepsilon^2)$ with $N_{\text{eff}}(s, d) \sim \varepsilon^{-d/s}$.
- **Non–Hermitian QOS.** The phase oracle exploits the mutual orthogonality of the projectors $|x\rangle\langle x|$. For non–Hermitian operators (e.g. a stochastic matrix block–encoding) one would need a genuine Lie–Trotter analysis. Does the linear–in–N scaling survive, or does it degrade to N^2 as for randomized Hamiltonian simulation?
- **Beyond–Holevo extraction.** Interferometric classical shadows extract $\widetilde{\mathcal{O}}(2^n)$ classical bits from an n–qubit register without violating Holevo because the bits encode *predictions on future inputs* rather than the state itself. Is there a structural classification of tasks for which this expansion is provably valid? A natural target is online decision problems with bounded VC dimension.

- **Lower bounds in restricted classical models.** The hardness arguments of [Zha+26] treat classical learners as arbitrary streaming algorithms. Is the same separation provable against specific classical learners–low–degree polynomials, random features, deep linear networks, transformer architectures of bounded width? Statistical query and SQ–dimension techniques are likely the right tool.
- **Quantum BBP transition.** Theorem 6.1 predicts a phase transition in QOS–PCA at the spiked–covariance threshold $\beta_c = \sqrt{D/N}$. A rigorous proof, identifying the quantum analogue of the BBP outlier transition, would constitute a new bridge between random–matrix theory and quantum learning.
- **Kernel learning beyond shift–invariance.** Section 7 relies on Bochner's theorem and the RFF construction, which is restricted to shift–invariant kernels on $\mathbb{R}^d$. Can QOS be combined with the Mercer expansion to handle a generic positive semi–definite kernel, perhaps via a Nyström–type subsampling? The target sample complexity would scale with the effective spectral dimension of the kernel operator rather than with the ambient feature dimension.
- **Threshold theorems for streaming quantum advantage.** Theorem 10.1 gives an upper bound $p_g \lesssim \varepsilon^2/(NQ^2)$ on the per–gate noise rate. Is this threshold tight? A matching lower bound—showing that QOS *cannot* achieve ε–accuracy with $p_g \gtrsim \varepsilon^2/(NQ^2)$ regardless of mitigation—would be the streaming analogue of the threshold theorems for random circuit sampling [Aha+23].

13 Conclusion

[Zha+26] have provided the cleanest argument to date that quantum machine learning on classical data has a physical, not merely conjectural, basis for advantage. The construction is elegant and the lower bounds are deep. Our analysis confirms the leading asymptotics with sharper constants and shows that the optimality is in fact a quantum–statistical Cramér–Rao phenomenon. At the same time, several practically relevant generalizations—importance weighting, soft–gap PCA, kernel–method extensions via random Fourier features (developed in parallel with [Sep26]), Rényi differential privacy, hybrid JL preprocessing, and a careful noise–robustness analysis—substantially expand the regime in which QOS is useful, while also revealing a tighter per–gate noise threshold $p_g \lesssim \varepsilon^2/(NQ^2)$ that has implications for near–term deployment. The most important open frontier, in our view, is the extension to non–Hermitian generators and to continuous data domains, where the polylog–space advantage may either survive or collapse to the randomized–compiler N^2 regime. Resolving this question would either widen the quantum–AI horizon dramatically or sharpen its boundary.

Acknowledgements

This commentary draws on the open–source implementation provided by the authors of [Zha+26] and benefited from the clarity of the companion blog exposition at *Quantum Frontiers*. Sections 7 and 10 were developed in parallel with the companion manuscript of [Sep26]; we are grateful for the productive exchange of drafts that helped sharpen both lines of analysis, and for the connection to the established literature on quantum–enhanced kernel methods [SK19; Hav+19] and robust shadow tomography [Che+21b; KG22] that the companion brought to our attention. Any remaining errors are the responsibility of the present author.

References

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

[Aba+16] Martín Abadi et al. "Deep learning with differential privacy". In: *Proc. ACM SIGSAC Conference on Computer and Communications Security.* 2016, pp. 308–318.

[Ach03] Dimitris Achlioptas. "Database-friendly random projections: Johnson–Lindenstrauss with binary coins". In: *Journal of Computer and System Sciences* 66.4 (2003), pp. 671–687.

[ADK23] Armando Angrisani, Mina Doosti, and Elham Kashefi. "A unifying framework for differentially private quantum algorithms". In: *arXiv preprint arXiv:2307.04733* (2023).

[Aha+23] Dorit Aharonov et al. "Polynomial–time classical simulation of noisy random circuits in any dimension". In: *arXiv preprint arXiv:2211.03999* (2023).

[Bab+18] Ryan Babbush et al. "Encoding electronic spectra in quantum circuits with linear T complexity". In: *Physical Review X* 8.4 (2018), p. 041015.

[Cam19] Earl Campbell. "Random compiler for fast Hamiltonian simulation". In: *Physical Review Letters* 123.7 (2019), p. 070503.

[Che+21a] Chi-Fang Chen et al. "Concentration for random product formulas". In: *PRX Quantum* 2.4 (2021), p. 040305.

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

[CM05] Graham Cormode and Shan Muthukrishnan. "An improved data stream summary: the count-min sketch and its applications". In: *Journal of Algorithms* 55.1 (2005), pp. 58–75.

[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.

[Har20] Aram W. Harrow. "Small quantum computers and large classical data sets". In: *arXiv preprint arXiv:2004.00026* (2020).

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

[Hel76] Carl W. Helstrom. *Quantum Detection and Estimation Theory.* Academic Press, 1976.

[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.

[Hol11] Alexander S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory.* 2nd. Edizioni della Normale, 2011.

[JL84] William B. Johnson and Joram Lindenstrauss. "Extensions of Lipschitz mappings into a Hilbert space". In: *Contemporary Mathematics* 26 (1984), pp. 189–206.

[Joh01] Iain M. Johnstone. "On the distribution of the largest eigenvalue in principal components analysis". In: *The Annals of Statistics* 29.2 (2001), pp. 295–327.

[JR25] Samuel Jaques and Arthur G. Rattew. “QRAM: A survey and critique”. In: *Quantum* 9 (2025), p. 1922.

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

[Kim+17] Shelby Kimmel et al. “Hamiltonian simulation with optimal sample complexity”. In: *npj Quantum Information* 3.1 (2017), p. 13.

[LC19] Guang Hao Low and Isaac L. Chuang. “Hamiltonian simulation by qubitization”. In: *Quantum* 3 (2019), p. 163.

[Liu+20] Jing Liu et al. “Quantum Fisher information matrix and multiparameter estimation”. In: *Journal of Physics A: Mathematical and Theoretical* 53.2 (2020), p. 023001.

[MCJ13] Ioannis Mitliagkas, Constantine Caramanis, and Prateek Jain. “Memory limited, streaming PCA”. In: *Advances in Neural Information Processing Systems*. Vol. 26. 2013.

[Mir17] Ilya Mironov. “Rényi differential privacy”. In: *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. 2017, pp. 263–275.

[MM15] Cameron Musco and Christopher Musco. “Randomized block Krylov methods for stronger and faster approximate singular value decomposition”. In: *Advances in Neural Information Processing Systems*. Vol. 28. 2015.

[MP67] Vladimir A. Marchenko and Leonid A. Pastur. “Distribution of eigenvalues for some sets of random matrices”. In: *Mathematics of the USSR-Sbornik* 1.4 (1967), pp. 457–483.

[Par09] Matteo G. A. Paris. “Quantum estimation for quantum technology”. In: *International Journal of Quantum Information* 7.1 (2009), pp. 125–137.

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

[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).

[Saa03] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. 2nd. SIAM, 2003.

[Sep26] A. Sepulveda-Jimenez. *Expanding Quantum Oracle Sketching and Classical Shadows*. May 2026. URL: https://doi.org/10.5281/zenodo.19985698.

[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.

[Tro15] Joel A. Tropp. "An introduction to matrix concentration inequalities". In: *Foundations and Trends in Machine Learning* 8.1-2 (2015), pp. 1–230.

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

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

[Woo14] David P. Woodruff. "Sketching as a tool for numerical linear algebra". In: *Foundations and Trends in Theoretical Computer Science* 10.1-2 (2014), pp. 1–157.

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