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.

Detection-Limited Eigenvalue Learning for Hybrid Quantum Linear System Solvers

Hikaru Wakaura

hikaruwakaura@gmail.com

QuantScape Inc https://orcid.org/0000-0001-8381-8323

Taiki Tanimae

QIRI (Quantum Integrated Research Institute Inc.)

Research Article

Keywords: Bayesian phase estimation, eigenvalue estimation, hybrid quantum–classical algorithms, power flow, quantum linear system algorithms, quantum phase estimation

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11766188/v1

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

Additional Declarations: The authors declare no competing interests.

Detection-Limited Eigenvalue Learning for Hybrid Quantum Linear System Solvers

Hikaru Wakaura and Taiki Tanimae

***Abstract*—Hybrid quantum linear system solvers of the Harrow–Hassidim–Lloyd (HHL) type learn the eigenvalues that carry weight in the input state and then run a short clock register, but the choice of the learning estimator has not been analysed. We show that the detection of weak spectral components decides this choice and give a counting model: a component of weight w is found with about $(w + w_{\mathrm{fl}})/w^2$ samples, with w_{fl} the weight leaking into its label. Single-ancilla estimators on mixture statistics need about $1/w^2$ samples, and labelling each shot with a clock register (Projective HHL) needs about $1/w$, down to a leakage floor of a few per cent. Recycled sequential HHL keeps one copy of the input, measures it repeatedly with one ancilla and labels the copy by its own Bayesian posterior, without a register; its label error is tuned by a stopping threshold, which is not a proven bound, and measured in simulation. On four systems with up to eight eigenvalues and calibrated ancilla noise it gives a lower median solution infidelity than Projective HHL within the same evolution time in all eight settings, by factors from 2 to 3×10^3, while neither is uniformly better for the compliance. On DC power-flow problems of 9-, 14- and 30-bus systems it is, without dephasing, the only stage tested that resolves the weak component at the smallest eigenvalue of the 9-bus system in most runs; with dephasing none reliably resolves the weak components of the larger systems, and the coherent clock limits every solver.**

***Index Terms*—Bayesian phase estimation, eigenvalue estimation, hybrid quantum–classical algorithms, power flow, quantum linear system algorithms, quantum phase estimation**

I. Introduction

SOLVING linear systems $A\vec{x} = \vec{b}$ underlies most of scientific computing, and the HHL algorithm [1] and its successors [2]–[4] promise to prepare the normalised solution state $|x\rangle$ with a cost that is polylogarithmic in the dimension. The promise comes with well-known conditions on sparsity, condition number κ, state preparation and read-out [4], [5], and with a practical obstacle that dominates every attempt to run the algorithm on present hardware: the eigenvalue estimation stage. In the textbook circuit the eigenvalues are written coherently into a clock register by QPE, and the depth of that register grows exponentially with the number of precision bits, which empirical studies identify as the most severe bottleneck on noisy devices [6].

Hybrid HHL schemes [7]–[10] side-step the problem by learning the eigenvalues $\{\lambda_j\}$ that carry weight in the input, $|b\rangle = \sum_j \beta_j |u_j\rangle$, in a separate classical post-processed experiment, and then running a short clock whose rotation angles are set from the learned values. The learning experiment has been a sampled QFT-type QPE [7]–[9] or, for eigenvalues that are exact binary fractions, an iterative phase estimation [10]. In parallel, the early fault-tolerant literature has produced single-ancilla eigenvalue estimators that reach the Heisenberg limit without eigenstate preparation, including multi-order time-series methods [11], [12], QCELS [13], [14], robust multiple-phase estimation [15], Gaussian-filtered search [16], and measurement-backaction schemes [17]. To our knowledge these continuous-valued estimators have not been compared as learning stages of a linear solver for an input with an unknown number of weighted components, which is the question addressed here. Methods that avoid phase estimation altogether, a variational solver [18] and, for a forecasting task, quantum reservoir computing [19], quantum circuit learning [20] and decoded quantum interferometry [21], are included as further baselines (Sec. IX).

This paper makes three points. The first is a detection law for the learning stage. For a non-eigenstate input a single-ancilla Hadamard test measures the mixture $\sum_j w_j \cos(\lambda_j t + \theta)$; the derivative of the outcome probability with respect to a weight w_j does not depend on w_j, so the Fisher information per shot on each weight is $O(1)$, the relative weight error is $\sim 1/(w_j\sqrt{S})$, and a component is found only when $S \gtrsim w_j^{-2}$. A scheme that labels each shot by a short coarse QPE instead counts the component with Poisson statistics, relative error $\sim 1/\sqrt{w_j S}$ and $S \gtrsim w_j^{-1}$, down to a floor set by clock leakage. The eigenvalue precision of a detected component is a secondary quantity: its local information scales as w_j^2 for the mixture and as w_j with labelling, and at the budgets studied it is the binding constraint only for the smallest eigenvalues under dephasing (Sec. V). Because a missed component at a small eigenvalue carries a large share of the solution, detection decides the solution fidelity. We first benchmark such a register-labelled learning stage, which keeps the single-ancilla refinement but labels each shot by a coarse QPE, against mixture estimators and sampled QPE under dephasing.

The second point is that the floor belongs to the clock register and not to labelling. If one copy of $|b\rangle$ is kept and measured repeatedly with a single ancilla, every measurement operator is a function of A, and the probability of a record is a weighted product over the measurements, as if one eigencomponent had been drawn with probability w_j [12]. This back-action is the basis of the channel framework of [17], whose adaptive scheme reads binary digits, and of solvers built on iterative phase estimation for eigenvalues that are exact binary fractions [10].

We combine it with known Bayesian tools into a learning stage in which each copy is labelled by its own posterior and stopped at a posterior-tail threshold, so that the label has no $1/d^2$ tail, the resolution is set by evolution time rather than

The authors are with QIRI (Quantum Integrated Research Institute Inc.), Tokyo 107-0061, Japan (e-mail: h.wakaura@qiri.co.jp; t.tanimae@qiri.co.jp).
Corresponding author: Hikaru Wakaura (e-mail: h.wakaura@qiri.co.jp).

by a register size, and the number of components need not be known. The measurement settings maximise the rate of expected variance reduction [22], [23], evaluated in closed form from Fourier moments of the posterior, and evolution time can be allocated only to components that are above their precision target. The stopping threshold is not a proven error bound; we measure the label error, including under a mismatched noise model, and we test by an ablation whether the advantage survives a change of the time allocation.

We name the two labelled schemes by the hybrid solver they define. *Projective HHL* learns with the projective estimator, in which the clock outcome of a short QPE labels the shot and collapses the input approximately onto the labelled eigencomponent; this is the register-labelled learning of the hybrid schemes cited above with a refinement measurement added. *Recycled sequential HHL* learns with the sequential estimator, in which one copy of the input is recycled through sequential single-ancilla phase estimation and labelled by its own posterior. In column headings "projective" and "sequential" stand for the two. The names denote the configurations compared here and not new complete solvers: unless a real clock register is stated, the solution fidelity quoted for either is that of an idealised solver that applies the inverse of the learned eigenvalue nearest to each eigencomponent exactly.

The third point is that for a sparse spectral measure the learning stage already answers most questions asked of the solver. That $\langle b|f(A)|b\rangle = \int f\, d\mu_b$ is the spectral theorem, and estimating such integrals from an approximate cumulative distribution function is the route of Lin and Tong [24]; what we add is the atomic case, in which the measure has few atoms and is represented by learned eigenvalues and weights, so that the compliance $\vec{b}^{\mathsf{T}} A^{-1}\vec{b}$, the energy of the solution and its mode weights follow with the per-atom weight error $\sqrt{w_j/S}$ rather than with a uniform error on the distribution function, and a polarisation identity with three further input states extends the construction to linear functionals $\langle v|A^{-1}|b\rangle$. No polynomial transformation and no post-selection are needed, and with the sequential estimator no clock register either. We compare this with two model-free estimators built from the same Hadamard tests, a thresholded time-evolution Krylov solver [25], [26] and the randomised Fourier representation of A^{-1}, and find the regime of few weighted components in which the atomic representation wins. Only observables that are not diagonal in the eigenbasis still need the quantum back end; there we return the learned spectrum to a real clock, uniform or tapered [1], [27]–[29], and to spectrally corrected QSVT inversion [30], and quantify what it buys and what it cannot repair.

Finally we apply all estimators to the DC power-flow problem of standard 9-, 14- and 30-bus test systems [31], a linear system on which HHL-type solvers have been run for three to nine buses [32]–[34], to see which conclusions survive on spectra that were not chosen for the purpose (Sec. VII). A second test is the readout training of a small reservoir computer for three nonlinear time series, a ridge-regression system [35], [36] in which components that hold less than 10^{-3} of the weight carry up to 72% of the solution (Sec. VIII). The result is a design rule for the learning stage and not a claim of advantage: every test system is solved classically at negligible cost, the noise model contains ancilla noise only, and an end-to-end analysis finds no practical advantage of quantum linear solvers for DC power flow [5].

The paper is organised as follows. Section II states the setting and the detection law. Section III defines the three learning stages and Sec. IV benchmarks them. Section V uses the learned measure as a solver for observables, and Sec. VI returns the spectrum to a solver with a real clock register. Section VII applies all estimators to DC power-flow problems and Sec. VIII to the readout training of a reservoir computer, Sec. IX adds baselines that do not use phase estimation, and Sec. X gives the conclusions and the limitations.

II. Setting and the detection law

Let A be Hermitian with eigenpairs $(\lambda_j, |u_j\rangle)$, $0 < \lambda_{\min} \le \lambda_j \le \lambda_{\max} \le 1$, $\kappa = \lambda_{\max}/\lambda_{\min}$, and let $w_j = |\beta_j|^2$ with $\sum_j w_j = 1$ be the weights of the m distinct eigenvalues on the support of $|b\rangle$. The single-ancilla circuit (ancilla $|+\rangle$, controlled e^{-iAt}, phase θ, Hadamard, measurement) returns outcome 0 with probability

$$p(t,\theta) = \tfrac{1}{2} + \tfrac{1}{2}\, c(t) \sum_j w_j \cos(\lambda_j t + \theta), \tag{1}$$

where $c(t) = (1-p_d)e^{-t/T_2}$ is the contrast under depolarising rate p_d and control-qubit dephasing time T_2; we take c to be calibrated. Five statements follow (proofs in Appendix A).

Detection and weight law (mixture). On the simplex a change of w_j is absorbed by the other weights, so $\partial p/\partial w_j = \frac{1}{2}c(\cos\varphi_j - \bar{g})$ with $\varphi_j = \lambda_j t + \theta$ and $\bar{g}$ the cosine of the absorbing components; this does not depend on w_j and is at most c in magnitude, so the Fisher information of one shot on the weight w_j is at most $c^2/[p(1-p)] \le 4c^2/(1-c^2)$, an $O(1)$ number when $c < 1$. For any unbiased estimator of a resolved component, $\sigma(\hat{w}_j) \ge 1/(2c\sqrt{\kappa_p S})$ with $\kappa_p = \overline{[4p(1-p)]^{-1}} \le (1-c^2)^{-1}$, so the relative weight error is $\sim 1/(w_j\sqrt{S})$ and a component is distinguishable from absence only when $w_j\sqrt{S}$ exceeds a threshold z that increases with the number of frequency bins searched: $S \gtrsim z^2/w_j^2$.

Detection and weight law (labelled). If each shot carries a label that identifies the eigencomponent, the number of shots with label j is binomial with mean $w_j S$, so $\sigma(\hat{w}_j) = \sqrt{w_j(1-w_j)/S}$, the relative error is $\sim 1/\sqrt{w_j S}$, and the probability of never seeing the component is $(1-w_j)^S$: $S \gtrsim \ln(1/\delta)/w_j$ suffices. The sampled coherent clock of hybrid HHL is such a labelled scheme, which is why it does not suffer from the mixture law. The two laws differ by the factor w_j in the information, and this is the whole content of the "w_j versus w_j^2" statement: it concerns detection and relative weights, not absolute weights and not, at the budgets considered here, eigenvalues.

A counting model. The two statements can be read in one counting model. Let a labelling procedure assign a sample of component j to the label of j with probability α_j (its acceptance) and let w_{fl} be the total probability that a sample of another component receives that label. The count in the label is binomial with success probability $p_j = \alpha_j w_j + w_{\text{fl}}$. If the

background is not known the component is visible only for $\alpha_j w_j \gtrsim w_{\text{fl}}$, a floor that does not decrease with S; if α_j and w_{fl} are known and the background is subtracted, a signal of z standard deviations needs

$$S \geq z^2 \, \frac{p_j(1-p_j)}{\alpha_j^2 w_j^2} \approx z^2 \, \frac{w_j + w_{\text{fl}}}{w_j^2}, \tag{2}$$

the last form for $\alpha_j \approx 1$ and rare labels. This is a signal-to-noise criterion, not a finite-sample test with stated error rates. For a uniform clock register $w_{\text{fl}} = \sum_{k \neq j} w_k \sum_{y \in Y_j} |a_{ky}|^2$ with a_{ky} the clock amplitudes and Y_j the bins assigned to the label (the peak bin and its two neighbours in the tests below), a few per cent next to a strong component and dependent on the instance, and α_j is the probability of the bins assigned to the component (on the planted-component test below, 0.57 and 0.79 for the peak bin of a five-bit register and 0.88 and 0.93 for the three bins around it). For the sequential estimator of Sec. III, w_{fl} is set by the stopping threshold and is measured in Sec. IV. Mixture statistics have no label; their weight law has the scaling of (2) with $w_{\text{fl}} = O(1)$, which is an analogy and not an identity of the two observation models. In this model the cost of a learning stage is decided by how small it makes w_{fl} and by what a sample costs.

Eigenvalue information (secondary). The Fisher information matrix of one setting with respect to the eigenvalues is $I_{jk} = c^2 t^2 w_j w_k \sin\varphi_j \sin\varphi_k / [4p(1-p)]$, of rank one; averaged over θ and over a time window wider than the inverse gaps it is approximately diagonal with $\bar{I}_{jj} \asymp c^2 \overline{t^2} w_j^2$ (bounds in Appendix A), so a detected component can be refined to $\sigma_j \asymp 1/(w_j T\sqrt{S})$ by the mixture and to $\sigma_j \asymp 1/(\sqrt{w_j} T\sqrt{S})$ with labels, the labelled single-cosine information per shot being $t^2(1-\sqrt{1-c^2})$ after averaging over the phase. With T of a few hundred and $S = 10^4$ these errors are below 10^{-3} for every $w_j \geq 0.03$ and are not what limits the solver.

Error propagation. For the solver output $|\hat{x}\rangle \propto \sum_j (\beta_j/\hat{\lambda}_j)|u_j\rangle$ with $\hat{\lambda}_j = \lambda_j + \delta_j$, write $q_j = (w_j/\lambda_j^2)/\sum_k (w_k/\lambda_k^2)$ for the share of component j in $\|x\|^2$ and $e_j = \delta_j/\lambda_j$. Then

$$1 - F = \sum_j q_j e_j^2 - \Big(\sum_j q_j e_j\Big)^2 + O(\|e\|^3) \;\leq\; \sum_j q_j e_j^2, \tag{3}$$

where the second term is the cancellation of errors parallel to the solution (a common relative error of all eigenvalues leaves the normalised state unchanged); for independent zero-mean errors $\mathbb{E}[1-F] = \sum_j q_j(1-q_j)\mathbb{E}[e_j^2]$. A component that is missed altogether, or frozen at a coarse value, contributes instead $q_j(1-q_j)(r-1)^2/[1+q_j(r^2-1)]$ with $r = \lambda_j/\lambda_{\text{nn}}$, where λ_{nn} is the eigenvalue it is merged with and the other components are exact; this is $\approx q_j(r-1)^2$ to leading order in q_j, $O(q_j)$ rather than $O(e_j^2)$. On the Poisson $N = 8$ system used below the component with $w = 0.032$ at $\lambda = 0.031$ has $q = 0.89$: missing it costs most of the fidelity, while refining it beyond 10^{-2} changes little. A weak neighbour that is merged with it is amplified by the small eigenvalue: the $w = 0.006$ component at $\lambda = 0.121$ ($q = 0.011$) rotated with $\lambda = 0.031$ costs 0.076. Detection, not precision, decides the fidelity, and the two laws above decide detection.

Tracking and unwrapping (heuristic). For a multi-stage mixture estimator that doubles the evolution time between rounds of S_r shots, staying in the correct phase basin requires $\sigma_j T_{r+1} \lesssim 0.4$ and hence a growth factor $f \lesssim c_u w_j \sqrt{S_r}$; components with $w_j \lesssim f/(c_u\sqrt{S_r})$ cannot follow and are frozen (Appendix A). This is a property of that estimator class, not a lower bound, and robust multi-peak methods [15], [16] are not bound by it.

III. Three learning stages

Mixture estimators. We implement two representatives of the single-ancilla family: a multi-level least-squares estimator in the spirit of MM-QCELS [14], which refits all eigenvalues and weights on time grids $t_n = n\tau_\ell$ with $\tau_\ell = \tau_0 2^\ell$, and a Gaussian-filtered search in the spirit of QMEGS [16], which samples $t \sim \mathcal{N}(0, T^2)$ and reads eigenvalues from the peaks of $\overline{\text{Re}[Z_n e^{ixt_n}]}$. Both see (1) and inherit the mixture eigenvalue rate.

Projective HHL: the projective estimator. Each shot runs an n_c-bit QPE with $t_0 = 2\pi/1.25$ whose clock outcome y labels a component (probability w_j) and collapses the register; the same shot then applies the Hadamard test at an adaptive time $t = 0.6/\sigma_j$ and phase $\theta = -\hat{\lambda}_j t - \pi/2$ to the collapsed state and updates a one-dimensional grid posterior for that component (the Wiebe–Granade rule [37], tempered to absorb mislabelling). The number of coarse bits is set by the resolution needed to separate the spectrum and by the coherence budget,

$$n_c = \min\big\{\lceil \log_2(2\lambda_{\max}/\Delta_{\min})\rceil,\ \lfloor \log_2(T_2/t_0)\rfloor\big\}, \tag{4}$$

so that the coherent part is governed by the spectral resolution $\Delta_{\min}$ and not by the target precision ε; the precision is carried by the incoherent refinement, whose maximal time is capped at T_2 ($2T_2$ in Table I; 2000 without dephasing, the window of the mixture baselines). Near-degenerate pairs below the coarse resolution are treated as one cluster. The construction is the explicit-register version of the backaction steering of [17]. A finite clock does not project exactly: outcome y leaves the register in $\propto \sum_j \beta_j a_{jy}|u_j\rangle$, so the refinement sees a mixture with purity $\pi_j = w_j|a_{jy}|^2/\sum_k w_k|a_{ky}|^2$ in the bin of component j, the effective weight in the labelled law is $w_j\pi_j^2$, and a weak component whose bin receives the $1/(\pi^2 d^2)$ tail of a strong one d bins away is lost below a floor $w_j \lesssim \sum_{k\neq j} w_k \sum_{y\in Y_j} |a_{ky}|^2$ that does not decrease with S. This leakage floor, not the shot count, is what limits the labelled scheme, and it is the mechanism behind the two-group results reported below.

Recycled sequential HHL: the sequential estimator. The leakage floor is a property of the finite register, and it is avoided by not using one. A copy of $|b\rangle$ is kept and the single-ancilla circuit of (1) is applied to it repeatedly. If the noise acts on the ancilla only, the operation on the system register for outcome d of a measurement with setting (t,θ) is a mixture of Kraus operators that are functions of A, so for any adaptive rule the probability of a record $d_1 \ldots d_n$ is

$$P(d_1 \ldots d_n) = \sum_j w_j \prod_{i=1}^{n} p_j(d_i; t_i, \theta_i), \tag{5}$$

with $p_j(0;t,\theta) = \frac{1}{2} + \frac{1}{2}c(t)\cos(\lambda_j t + \theta)$ (Appendix A). Without noise this is Eq. (10) of [12], and the back-action behind it is analysed in [17]. Each copy behaves as if one eigencomponent had been drawn with probability w_j and all its outcomes came from λ_j, so single-eigenvalue Bayesian estimation applies to the copy and the latent components of the copies are independent draws from the spectral measure. The label inferred from a finite record is an estimate of that latent component, with the errors quantified below. A digit-by-digit adaptive scheme [17] reproduces the outcome statistics of a register of the same length, tails included; the stopping rule below is designed to avoid them.

The estimator has three parts. (i) *Identification.* Each copy carries a grid posterior on $(0, 1.25)$ with mean μ and standard deviation σ. With the Fourier moments $z_t = \mathbb{E}[e^{i\lambda t}]$ and $q_t = \mathbb{E}[(\lambda-\mu)e^{i\lambda t}]$ of the posterior, a measurement at (t,θ) lowers the posterior variance on average by

$$R(t,\theta) = \frac{c(t)^2}{4}\,\frac{[\mathrm{Re}(e^{i\theta}q_t)]^2}{P_0(1-P_0)}, \quad P_0 = \tfrac{1}{2} + \tfrac{1}{2}c(t)\,\mathrm{Re}(e^{i\theta}z_t), \tag{6}$$

by the law of total variance, where c is the contrast of the likelihood used in the update. We take the pair (t,θ) that maximises R/t, the rate of expected gain [22], [23], over candidate times spaced by $\sqrt{2}$ up to $\min(t_{\mathrm{cap}}, 1.5/\sigma_{\mathrm{id}})$ and sixteen phases. For a Gaussian posterior and a constant contrast the maximum is at $\sigma t = 1/\sqrt{2}$ and at the phase of maximal fringe slope, close to the rule of [37]; with $c(t) \propto e^{-t/T_2}$ it moves to $\sigma^2 t^2 + t/T_2 = \frac{1}{2}$, and for a multimodal posterior it is the time and phase that separate the surviving modes.

The copy stops when $\sigma \le \sigma_{\mathrm{id}}$ and the posterior mass farther than $3\sigma_{\mathrm{id}}$ from μ is at most p_{out}; while only the second condition fails, the setting is $\theta = -\mu t$ with the time that maximises the decay rate of the far mass. The threshold p_{out} is a stopping parameter and not a guarantee: the likelihood is tempered by 0.02 for robustness and the prior of later copies is a plug-in, so the posterior is that of a working model. For an exact Bayesian update with the true prior, stopping at posterior tail mass p_{out} would bound the prior-averaged error rate by p_{out} (Appendix A); for the implemented rule the error rate is measured (Sec. IV). It takes the place of the $1/(\pi^2 d^2)$ tail of the register. The first batch of copies starts from a flat prior; later copies start from the mixture of the components found so far plus a flat mass $u = 0.05$ for undiscovered ones ($u = 0.2$ when the estimator is applied to another input state with the components fixed), which turns identification into a sequential test among known components.

(ii) *Detection.* The per-copy estimates are clustered by peeling: the fullest window of half-width $2\sigma_{\mathrm{id}}$ that contains at least $n_{\min} = 4$ estimates defines a component, the estimates within $4\sigma_{\mathrm{id}}$ of its mean are removed, and the step is repeated. The number of components is not an input, and the weights are the label frequencies.

(iii) *Refinement.* A labelled copy is not discarded. It continues with measurements at the frontier $t = 0.6/\sigma_c$ of the pooled posterior of its component (capped at t_{cap}, at most 16 per copy), and only while that component is above its precision target δ, so that evolution time is spent on the components that need it. The resolution σ_{id} is a choice, by default $\Delta_{\min}/8$ and not finer than $0.6/T_2$; like n_c in (4) it uses the smallest gap, which is assumed known, but it is not tied to a register, and the caps are those of the projective estimator ($t_{\mathrm{cap}} = T_2$, or 2000 without dephasing). Copies are run one after another until a prescribed total evolution time is used up; a copy that would not finish is not counted.

Resource comparison. When the coherent evolution time can grow with the precision, the textbook clock and the two labelled estimators all need a total evolution time of order $1/\varepsilon$; the difference is where it is spent. The clock needs $n_p = \log_2(\lambda_{\max}/\varepsilon)$ coherent ancillas, controlled U^{2^k} and an inverse QFT. The projective estimator needs $n_c + 1$ ancillas, a coarse clock of time $2^{n_c}t_0$ with its own inverse QFT, and $O(w_{\min}^{-1})$ shots for detection followed by an incoherent refinement of depth up to T; its gain is the ancilla count, the $w_{\min}^{-1}$ rather than $w_{\min}^{-2}$ detection cost and the robustness of short repeated circuits, and its price is the coarse register and its leakage. The sequential estimator needs one ancilla, no register and no QFT. Its price is the lifetime of the system register: a copy must stay coherent through all its measurements, here up to about 4×10^4 time units (1×10^4 with a single refinement per copy), against about 2×10^3 for a projective shot, although no single controlled evolution is longer than t_{cap}. A copy costs $c_{\mathrm{id}}/\sigma_{\mathrm{id}}$ with c_{id} given in Table II, against $(2^{n_c}-1)t_0$ for a register label (Sec. IV compares the two). When the single evolution is capped at T, n refinement measurements give $\varepsilon = 1/[c(T)T\sqrt{n}]$, a total time $1/[c(T)^2T\varepsilon^2]$ that grows as ε^{-2}; this is the functional form of the bound $\Theta(1/(q\varepsilon^2))$ for phase estimation with subroutines of q queries [38], but that model discards all qubits between subroutines, whereas the sequential estimator keeps the system register, so no optimality is implied.

IV. Benchmarks of the learning stage

We use four test systems with $\lambda_{\max} \le 1$: a random Hermitian 4×4 matrix with eigenvalue gaps above 0.08, the one-dimensional Dirichlet Poisson matrix for $N = 4$ and $N = 8$ ($\kappa = 9.5$ and 32), and a diagonally dominant symmetric Toeplitz matrix for $N = 8$ whose smallest gap is 0.0034; $|b\rangle$ is a random unit vector in each case, so that weak components (w_j down to 10^{-3}) occur. All measurement statistics are sampled exactly from (1) (and from the corresponding clock amplitudes for QPE, with Gaussian phase kicks of variance $2{\cdot}2^k t_0/T_2$ on each controlled U^{2^k}), with 10^4 shots per learning run. Fidelity is first evaluated for the idealised solver in which eigen-tagging is exact and only the rotation angles come from the learned values, so that the learning stages are compared in isolation.

Fig. 1 and Table I show the result. The mixture estimators are limited by the weak components, as (3) predicts: on Poisson $N = 8$, whose weights are 0.032 and 0.006 at the two smallest eigenvalues ($\lambda = 0.031$ and 0.121) and 0.001 at the fifth, their median fidelity is 0.76–0.88 without dephasing and they reach $F \ge 0.99$ in 1 (MM-QCELS-type) and 13 (QMEGS-type) of the 100 runs; the QMEGS-type runs form

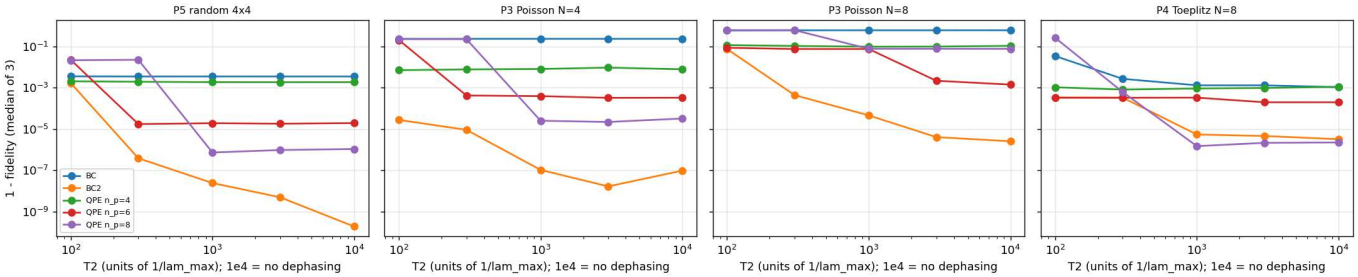


Fig. 1. Solution infidelity $1-F$ of the idealised hybrid solver versus the dephasing time T_2 (units of $1/\lambda_{\max}$; 10^4 denotes no dephasing) for the eigenvalue-learning stages, 10^4 shots, median of three repetitions. "BC" is a mixture estimator (multi-level least squares), "BC2" the projective estimator of (4), "QPE n_p" sampled coherent QPE with n_p bits. The projective estimator stays below 10^{-3} for $T_2 \geq 300$ on all four systems, including points where eight-bit QPE (coherent time 1287) has failed ($1-F \geq 0.02$); at $T_2 = 100$ it reaches 1.7×10^{-3} (random) and 0.075 (Poisson $N = 8$), and on Poisson $N = 8$ at $T_2 = 300$ its distribution over 100 runs has a low group that three runs do not show (Table I). The mixture estimator is limited by the weak components regardless of T_2.

TABLE I

MEDIAN SOLUTION FIDELITY OF THE IDEALISED HYBRID SOLVER (EXACT EIGEN-TAGGING, LEARNED ANGLES) OVER 100 REPETITIONS, 10^4 SHOTS PER RUN, $p_d = 0.02$. INTERQUARTILE RANGES ARE GIVEN WHERE THEY EXCEED 0.01. n_c FROM (4).

Problem	T_2	κ	MM-QCELS-type	QMEGS-type	QPE 6 bits	Projective (n_c)
Random 4×4	∞	2.2	0.995 [0.977, 0.998]	0.979 [0.977, 0.996]	1.0000	1.0000 (5)
Random 4×4	300	2.2	0.998 [0.978, 0.9999]	0.978 [0.977, 0.992]	1.0000	1.0000 (5)
Poisson $N = 4$	∞	9.5	0.941 [0.770, 0.997]	0.889 [0.769, 0.999]	0.9996	1.0000 (4)
Poisson $N = 4$	300	9.5	0.968 [0.813, 0.999]	0.818 [0.769, 0.944]	0.999	1.0000 (4)
Poisson $N = 8$	∞	32	0.880 [0.612, 0.902]	0.757 [0.392, 0.924]	0.999 [0.925, 0.999]	1.0000 (5)
Poisson $N = 8$	300	32	0.844 [0.394, 0.905]	0.647 [0.430, 0.906]	0.925	0.924 [0.924, 0.9997] (5)
Toeplitz $N = 8$	∞	4.0	0.999 [0.976, 0.999]	0.999	0.9998	1.0000 (6)
Toeplitz $N = 8$	300	4.0	0.999 [0.980, 0.999]	0.993 [0.973, 0.998]	0.9997	0.9996 (5)

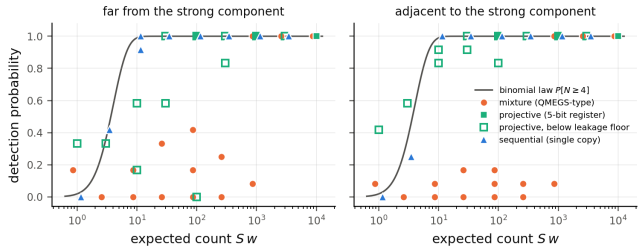


Fig. 2. Detection probability of a planted component of weight $w \in \{0.001, 0.003, 0.01, 0.03, 0.1\}$ (the sequential estimator up to 0.03) in an eight-eigenvalue spectrum with one strong component ($w = 0.5$), versus the expected count Sw for $S \in \{10^3, 10^4, 10^5\}$ samples (twelve runs per point, four for the sequential estimator at $S = 10^5$; markers of the three estimators are offset horizontally for visibility). Left: planted component far from the strong one; right: 1.5 register bins from it. The sequential estimator (triangles) follows the binomial law $P[N \geq 4]$ (line) at both placements and returned no false component in any run. Projective labelling with a five-bit register (squares) is hollow where the planted weight is below the leakage floor w_{fl} of its three bins (0.043 and 0.038), where a flagged peak cannot be attributed to the planted component. The mixture estimator (QMEGS-type, circles) is not governed by Sw: it detects only where $w\sqrt{S} \gtrsim 10$.

several groups there (26 at 0.38–0.40, 21 at 0.92–0.93, 13 at $F \geq 0.99$), and the medians of the first and of the second fifty runs are 0.91 and 0.54, so its median is not a stable summary. The detection runs in Appendix A and the mixture points of Fig. 2 show the mechanism: at 10^4 shots both mixture estimators miss all three components, whose amplitudes are below the per-point noise $1/\sqrt{S_{\mathrm{pt}}} \approx 0.12$ of the time series; at 10^5 shots the QMEGS-type estimator finds the $w = 0.032$ component in every run and reaches $F = 0.92$–0.99, while the MM-QCELS-type estimator still misses it in two of three runs, which is a limitation of that implementation rather than of the data.

Sampled six-bit QPE and the projective estimator both reach a median $F \geq 0.998$ in every case without dephasing, but not with the same reliability: on Poisson $N = 8$ six-bit QPE falls to 0.925 in 32 of the 100 runs, whereas the projective estimator stays above 0.97 in all of them. The two also differ in resources. Under dephasing the difference is in the coherent time (Fig. 1, medians of three runs). Eight-bit QPE (coherent time 1287) falls to $1-F \geq 0.02$ at $T_2 \leq 300$ on the random and Poisson $N = 4$ systems and at $T_2 = 100$ on the Toeplitz system, and six-bit QPE (coherent time 322) at $T_2 = 100$ on the random and Poisson $N = 4$ systems; elsewhere on these three systems both stay below 10^{-3}. On Poisson $N = 8$ both are limited before dephasing matters: in two of the three runs without dephasing eight-bit QPE does not select the $\lambda = 0.121$ component, because side lobes of the strong component at $\lambda = 0.773$ collect more counts, and its median stays at $1-F = 0.076$–0.077 for $T_2 \geq 1000$; six-bit QPE is at 0.075–0.086 for $T_2 \leq 1000$.

The labelled estimator with $n_c = 4$–5 (coherent time 80–160) keeps a median $F \geq 0.9996$ at $T_2 = 300$ on three systems and degrades at $T_2 = 100$ (to 0.998 on the random system and 0.92 on Poisson $N = 8$); on Poisson $N = 8$ at $T_2 = 300$ its runs form two groups, median 0.924 with upper quartile 0.9997: 57 of the 100 runs are below 0.93. In 55 of them ($F = 0.92$–0.93) the dephased coarse clock does not

resolve the $w = 0.006$ component at $\lambda = 0.121$ ($q = 0.011$), and the solver rotates it with the learned value of $\lambda = 0.031$, which costs 0.076 by the exact expression after (3) (0.09 to leading order); in the other two ($F = 0.84$ and 0.86) the $\lambda = 0.031$ component itself is learned at 0.077. None of the 43 runs above 0.93 shows the merger, and the seven low runs of Table IV all do. The 32 low runs of six-bit QPE without dephasing and all 100 at $T_2 = 300$ show the same merger, which is why the two estimators have the same median at $T_2 = 300$. The other low tails of the projective estimator are small: 6 of 100 runs below 0.99 on Poisson $N = 4$ at $T_2 = 300$ (minimum 0.969) and 3 of 100 on Poisson $N = 8$ without dephasing (minimum 0.972). This is the leakage floor of Sec. III, raised by dephasing, and not a shot-count effect.

The refinement time cap does not change this in the range tested (in separate sets of 100 runs that use the same random numbers for every cap between $0.5T_2$ and $3T_2$, the median is 0.924–0.925, the quartiles are 0.924 and 0.999, 60 or 61 runs are below 0.93, and one run changes group; without dephasing the result is unchanged when the refinement is capped at 2000, the window of the mixture baselines), so under dephasing the coherent coarse stage, not the incoherent refinement, is the limiting element.

Table I compares whole estimators at the same number of shots, each with its own time rule, and not at equal evolution time: the refinement of the projective estimator is capped at $2T_2$ (2000 without dephasing), the MM-QCELS-type time grid extends to 295 at $T_2 = 300$ and to 2359 without dephasing, the QMEGS-type estimator draws $|t|$ from a Gaussian of width $T = \min(T_2, 2000)$, so that 32% of its times exceed T, and six-bit QPE has a coherent window of 322; coarse registers have at most six bits. The sequential estimator of Sec. III is compared with the projective one separately, within the same budget of total evolution time, in the next subsection.

A. Recycled sequential HHL

Table II gives the cost and the error of the identification. At $p_{\mathrm{out}} = 10^{-4}$ a copy costs $8.8/\sigma_{\mathrm{id}}$ from a flat prior and $5.0/\sigma_{\mathrm{id}}$ with the learned prior (11.8 and 6.4 at $T_2 = 300$); with the exact spectrum as prior it would cost $4.5/\sigma_{\mathrm{id}}$. The fraction of estimates farther than $3\sigma_{\mathrm{id}}$ from the eigenvalue, the event the stopping rule refers to, does not exceed p_{out} in seven of the eight entries (two of them, at $p_{\mathrm{out}} = 10^{-5}$, have no event, which 24 000 copies cannot distinguish from 10^{-5}) and is 1.7×10^{-4} in the eighth (flat prior, $p_{\mathrm{out}} = 10^{-4}$); no estimate is farther than $4\sigma_{\mathrm{id}}$ for $p_{\mathrm{out}} \le 10^{-4}$, which with 24 000 copies bounds that fraction only at the 10^{-4} level. Lowering p_{out} by a factor of ten adds between 1.0 and 3.0 to c_{id} without dephasing and between 1.3 and 4.5 at $T_2 = 300$, the increments growing as p_{out} falls. The threshold is not robust to a wrong noise model (Table III): when the contrast is over-estimated, by a dephasing time three times too long or a depolarising rate five to ten times too small, the fraction beyond $3\sigma_{\mathrm{id}}$ rises to as much as 9 times p_{out} at $p_{\mathrm{out}} = 10^{-3}$ and 17 times at $p_{\mathrm{out}} = 10^{-4}$, whereas a weak unmodelled dephasing ($T_2 = 1000$ instead of none) leaves it near p_{out}. The degradation is gradual, and it is still far below the $1/(\pi^2 d^2)$ tail of a register, but p_{out} must be read as a design parameter whose effect has to be measured.

Fig. 2 shows the consequence for detection. With planted weights between 10^{-3} and 0.03 the sequential estimator detects the component with the binomial probability of collecting four copies: 0.25–0.42 at $Sw = 3$, 0.92–1 at $Sw = 10$ and 1 above, far from and next to the strong component alike, and it returned no false component in any run up to 10^5 copies. The projective estimator does not find $w = 0.001$ far from the strong component even with 10^5 shots, and next to the strong component its detections below $w \approx 0.04$ are under the leakage floor. A copy costs 3.3–4.2 register shots in this test once the prior is learned (6–8 in the first 10^3 copies); more register shots do not lower the floor.

Table IV compares four labelled estimators within the same budget of total evolution time: the projective and the sequential estimator, each with a refinement measurement on every sample and with the refinement stopped at a precision target. The budget T_{tot} is the median time of the ten projective runs with 10^4 shots, whose individual times are 0.58–$1.03\,T_{\mathrm{tot}}$; the projective estimator with the target uses the largest number of shots that fits (0.86–$1.00\,T_{\mathrm{tot}}$), and the sequential estimator stops at T_{tot}.

With a refinement on every sample, the sequential estimator has the lower median infidelity on all eight system and noise combinations, by factors between 2.3 (random system, no dephasing) and 2.9×10^3 (Poisson $N = 8$, $T_2 = 300$); on Poisson $N = 8$ without dephasing the median falls from 2.8×10^{-6} to 6.8×10^{-8}, and at $T_2 = 300$, where the dephased coarse register does not resolve the $w = 0.006$ component at $\lambda = 0.121$ in seven of these ten runs (55 of 100 in the independent runs of Table I), so that it is rotated with $\lambda = 0.031$ and the projective estimator stays at 0.076, to 2.6×10^{-5}. With the target on both it is lower on all eight as well, by factors of 1.5 to 2.2×10^3. The advantage thus persists under both allocation policies, and giving the projective estimator the target does not remove its failure on Poisson $N = 8$ (2.7×10^{-6} without dephasing; median 0.013 and worst run 0.076 at $T_2 = 300$), which is consistent with its eigenvalues being limited by the purity of the register label and by the unresolved (merged) component. The allocation is not without effect: it changes the medians of both estimators on the other systems, and the four columns compare whole estimators, not the label in isolation. The worst runs of the projective estimator with the target reach 1.3×10^{-3} (Poisson $N = 8$, no dephasing) and 8.7×10^{-3} (Poisson $N = 4$, $T_2 = 300$).

All eight components of Poisson $N = 8$ are found without the number being supplied when there is no dephasing, and seven at $T_2 = 300$. On the Toeplitz system, $\sigma_{\mathrm{id}} = 0.005$ separates the pair at $\lambda = 0.921$ and 0.938 (weights 0.023 and 0.80), which the six-bit register does not resolve, in all ten runs without dephasing and in seven of ten at $T_2 = 300$, and leaves the pair 0.0034 apart merged ($\hat{m} = 7$); resolving that pair needs $\sigma_{\mathrm{id}} = 4 \times 10^{-4}$. Every copy is then identified at that resolution and the same time buys 5.1×10^3 copies instead of 2.1×10^4; all eight components are found and the median infidelity is 1.2×10^{-7} against 1.3×10^{-6} (five repetitions),

TABLE II

IDENTIFICATION COST OF ONE COPY, $c_{\mathrm{id}} = T_{\mathrm{id}}\sigma_{\mathrm{id}}$ (MEAN), AND FRACTION OF ESTIMATES FARTHER THAN $3\sigma_{\mathrm{id}}$ AND $4\sigma_{\mathrm{id}}$ FROM THE EIGENVALUE, VERSUS THE STOPPING THRESHOLD p_{out}, FOR A FLAT PRIOR AND FOR THE PRIOR LEARNED BY A PRECEDING RUN OF THE ESTIMATOR (5×10^6 TIME UNITS; ALL EIGHT COMPONENTS FOUND). "EXACT": THE TRUE SPECTRUM AS PRIOR, FOR REFERENCE. POISSON $N = 8$, $\sigma_{\mathrm{id}} = 0.0112$, 24 000 COPIES PER ENTRY, MATCHED CONTRAST. WITH ZERO EVENTS IN 24 000 COPIES THE ONE-SIDED 95% UPPER LIMIT ON THE FRACTION IS 1.2×10^{-4}.

	flat prior			learned prior			exact	$T_2 = 300$	
p_{out}	c_{id}	$> 3\sigma_{\mathrm{id}}$	$> 4\sigma_{\mathrm{id}}$	c_{id}	$> 3\sigma_{\mathrm{id}}$	$> 4\sigma_{\mathrm{id}}$	c_{id}	c_{id} flat	c_{id} learned
10^{-2}	5.5	2.0e-3	1.0e-3	1.3	3.3e-4	3.3e-4	1.2	7.0	1.5
10^{-3}	6.5	2.1e-4	8.3e-5	2.7	0	0	2.6	8.3	3.3
10^{-4}	8.8	1.7e-4	0	5.0	0	0	4.5	11.8	6.4
10^{-5}	11.6	0	0	8.0	0	0	6.9	16.2	10.0

TABLE III

FRACTION OF ESTIMATES FARTHER THAN $3\sigma_{\mathrm{id}}$ AND $4\sigma_{\mathrm{id}}$ FROM THE EIGENVALUE WHEN THE CONTRAST ASSUMED BY THE ESTIMATOR DIFFERS FROM THE TRUE ONE (ASSUMED $\rightarrow$ TRUE; "MATCHED": NO MISMATCH). FLAT PRIOR, POISSON $N = 8$, 24 000 COPIES PER ENTRY.

	$p_{\mathrm{out}} = 10^{-3}$		$p_{\mathrm{out}} = 10^{-4}$	
contrast	$> 3\sigma_{\mathrm{id}}$	$> 4\sigma_{\mathrm{id}}$	$> 3\sigma_{\mathrm{id}}$	$> 4\sigma_{\mathrm{id}}$
$T_2 = 300$, matched	4.6e-4	4.2e-5	0	0
T_2: 300 $\rightarrow$ 100	8.3e-3	2.6e-3	4.2e-4	4.2e-5
no dephasing, matched	2.1e-4	8.3e-5	1.7e-4	0
p_d: 0.02 $\rightarrow$ 0.10	2.6e-3	1.2e-3	2.9e-4	4.2e-5
p_d: 0.02 $\rightarrow$ 0.20	8.7e-3	4.3e-3	1.7e-3	6.7e-4
T_2: $\infty \rightarrow$ 1000	7.9e-4	1.7e-4	1.3e-4	0

but a tenth of the time runs only 400 copies and finds six. A resolution that is refined only where estimates cluster would avoid this cost and is not implemented.

The estimator fails when the time does not cover both identification and the refinement of the dominant component. With $T_{\mathrm{tot}}/10$ on Poisson $N = 8$ at $T_2 = 300$ only 160 copies are run and the median infidelity is 0.076 (worst 0.61); on the Toeplitz system at $T_2 = 300$ the same fraction gives 8.7×10^{-4} (worst 0.034, four components found). With $\delta = 10^{-4}$ and $T_{\mathrm{tot}}/10$ on Poisson $N = 8$ without dephasing the refinement of the first components exhausts the time and the median is again 0.076; and with $\delta = 10^{-4}$ at $T_2 = 300$ the median is 1.0×10^{-4} and the worst run 0.076. The target must be set from the time available, as is done for observables in Sec. V.

For the weights the picture is different (Table V): neither labelled estimator is uniformly better. A register label is cheaper than a sequential copy, 156 against about 420 time units on Poisson $N = 8$, so that the projective estimator with the target collects 1.3×10^5 shots where the sequential one collects 4.8×10^4 copies, and its compliance error is the smaller there (0.35% against 0.86%). More shots do not, however, systematically lower the projective error: with the target it has 2 to 26 times as many shots as without, and its compliance error falls in three of the eight combinations and does not in five, so that it is not limited by sampling there. The sequential error is lower with the target than without on all eight. Over the eight combinations the smallest error belongs to the sequential estimator with the target in six (two of them near-ties), to the projective estimator in one and to the projective estimator with the target in one; with ten repetitions the ranking should not be over-read. Averaged over a run, a sequential copy costs 1.9 to 3.4 times a register label without dephasing and 4 to 16 times at $T_2 = 300$. In the two main sequential columns of Table IV (every copy; target $\delta = 10^{-3}$) every copy converged and at most 0.3% of the copies were left without a label; with $\delta = 10^{-4}$ up to 0.7% are, with $T_{\mathrm{tot}}/10$ up to 6%, and in the smallest runs of the other settings, with fewer than a hundred copies, up to 23%.

Equation (5) assumes that the system register is disturbed only by the measurements. As a first test we let the copy lose its eigencomponent at a rate $1/T_{\mathrm{sys}}$, redrawing it with probabilities w_j (Poisson $N = 8$, no dephasing, $T_{\mathrm{tot}} = 2 \times 10^6$, $\delta = 10^{-3}$). The median infidelity stays between 2.4×10^{-6} and 5.0×10^{-6} for $T_{\mathrm{sys}} \geq 3\times10^4$ (worst run 1.7×10^{-5}) and rises to 3.1×10^{-5} at $T_{\mathrm{sys}} = 10^4$, where the worst of ten runs is 0.021; no false component appeared, and seven of the eight components were found throughout. A copy lives for up to about 4×10^4 time units in the runs with a target and for 9×10^2 to 1×10^4 with one refinement per copy, against about 2×10^3 for a projective shot.

V. THE LEARNING STAGE AS A SOLVER

The projective estimator returns, for each learned eigenvalue, a label count. Solving the non-negative least-squares problem $\mathrm{counts}(y)/S \approx \sum_j \hat{w}_j P_j(y)$ with the (dephasing-averaged) clock kernel P_j computed from the learned μ_j gives the weights $\hat{w}_j$; reading them directly from the ±1 bins would be biased by the 19% leakage. The pair $\{(\mu_j, \hat{w}_j)\}$ is the spectral measure of A on $|b\rangle$, and for any f that is C^1 on $[\lambda_{\min}, 1]$,

$$\Big|\sum_j \hat{w}_j f(\mu_j) - \langle b|f(A)|b\rangle\Big| \leq \sum_j \big(|\eta_j||f(\lambda_j)| + \hat{w}_j|f'(\xi_j)||\delta_j|\big), \quad (7)$$

with $\eta_j = \hat{w}_j - w_j$. The compliance ($f = 1/\lambda$), the energy ratio and the eigenmode weights of the solution follow without any further quantum circuit. Linear functionals need the cross weights $c_j = \langle v|u_j\rangle\langle u_j|b\rangle$, which the polarisation identity $c_j = \frac{1}{2}(w_j^{(b+v)}\|b + v\|^2 - w_j - w_j^{(v)}) + \frac{i}{2}(w_j^{(b+iv)}\|b + iv\|^2 - w_j - w_j^{(v)})$ obtains from the diagonal weights of four input states, learned with the eigenvalue set fixed by the $|b\rangle$ run. When the refinement has made the eigenvalue errors subdominant, (7) is dominated by the weight errors $\eta_j \sim \sqrt{w_j/S}$ times the conditioning of the kernel, and the shot cost of the observable-first solver follows the sampling law $1/\varepsilon^2$.

TABLE IV

SOLUTION INFIDELITY $1 - F$ OF THE IDEALISED HYBRID SOLVER FOR FOUR LABELLED ESTIMATORS WITHIN THE SAME BUDGET OF TOTAL EVOLUTION TIME T_{tot}, THE MEDIAN TIME OF THE TEN PROJECTIVE RUNS WITH 10^4 SHOTS (UNITS OF $1/\lambda_{\max}$; INDIVIDUAL PROJECTIVE RUNS USE 0.58–$1.03\,T_{\mathrm{tot}}$, THE OTHER THREE ESTIMATORS NEVER MORE THAN T_{tot}); MEDIAN AND WORST OF TEN INDEPENDENT REPETITIONS (NO COMMON RANDOM NUMBERS BETWEEN ESTIMATORS), ENTRIES WRITTEN AS $2.8\mathrm{e}{-6} = 2.8 \times 10^{-6}$. "PROJECTIVE": REGISTER LABEL AND ONE REFINEMENT MEASUREMENT ON EVERY SHOT. "PROJ., TARGET": THE SAME WITH THE REFINEMENT STOPPED ONCE A COMPONENT REACHES THE RELATIVE PRECISION $\delta = 10^{-3}$, WITH THE LARGEST NUMBER OF SHOTS THAT FITS INTO T_{tot}. "SEQ., EVERY COPY": SEQUENTIAL ESTIMATOR WITH ONE REFINEMENT MEASUREMENT ON EVERY COPY. "SEQ., TARGET": REFINEMENT STOPPED AT $\delta = 10^{-3}$ (OR 10^{-4}); $\hat{m}$ IS THE MEDIAN NUMBER OF COMPONENTS FOUND (THE PROJECTIVE ESTIMATOR IS GIVEN m), AND THE LAST COLUMN USES $T_{\mathrm{tot}}/10$. THE SEQUENTIAL ESTIMATOR NEVER EXCEEDS ITS TIME BUDGET. $\sigma_{\mathrm{id}} = \Delta_{\min}/8$, EXCEPT TOEPLITZ, WHERE $\sigma_{\mathrm{id}} = 0.005$, A QUARTER OF A SIX-BIT BIN.

			Projective		Proj., target		Seq., every copy		Seq., target 10^{-3}			10^{-4}	$T_{\mathrm{tot}}/10$
Problem	T_2	T_{tot}	med.	worst	med.	worst	med.	worst	med.	worst	$\hat{m}$	med.	med.
Random 4×4	∞	2.0e7	7.4e-10	5.2e-9	6.3e-7	1.5e-6	3.2e-10	1.7e-9	4.3e-7	9.8e-7	4	3.6e-9	5.8e-7
Random 4×4	300	4.1e6	2.1e-6	1.2e-5	1.4e-6	1.6e-5	3.7e-7	5.2e-6	4.5e-7	8.2e-7	4	7.7e-8	7.8e-7
Poisson $N = 4$	∞	2.0e7	8.1e-8	2.3e-7	4.4e-7	1.4e-6	5.4e-9	3.0e-8	2.3e-7	8.8e-7	4	5.5e-9	2.3e-7
Poisson $N = 4$	300	3.5e6	3.9e-5	2.7e-3	2.1e-5	8.7e-3	2.2e-6	1.2e-5	5.8e-7	1.4e-6	4	2.4e-6	5.2e-5
Poisson $N = 8$	∞	2.0e7	2.8e-6	1.8e-4	2.7e-6	1.3e-3	6.8e-8	8.9e-6	8.8e-8	5.4e-7	8	7.8e-9	2.4e-6
Poisson $N = 8$	300	4.2e6	0.076	0.076	0.013	0.076	2.6e-5	2.9e-4	5.8e-6	1.9e-5	7	1.0e-4	0.076
Toeplitz $N = 8$	∞	2.2e7	4.4e-6	5.7e-6	3.6e-6	3.5e-4	4.9e-7	2.3e-6	1.3e-6	4.7e-6	7	4.8e-7	4.1e-6
Toeplitz $N = 8$	300	4.2e6	3.9e-4	5.8e-4	5.2e-4	1.0e-3	3.2e-6	5.6e-6	3.0e-6	6.6e-6	6	5.6e-6	8.7e-4

TABLE V

RELATIVE ERROR OF THE COMPLIANCE (MEDIAN OF TEN REPETITIONS) AND NUMBER OF SAMPLES FOR THE RUNS OF TABLE IV; THE PROJECTIVE ESTIMATOR USES 10^4 SHOTS.

		compliance error (%)				samples		
Problem	T_2	Proj.	Proj., target	Seq., every	Seq., target	Proj., target	Seq., every	Seq., target
Random 4×4	∞	0.09	0.14	0.15	0.09	128125	8504	50542
Random 4×4	300	0.15	0.21	0.30	0.20	23945	4600	6452
Poisson $N = 4$	∞	0.23	0.18	0.35	0.17	261250	9507	136761
Poisson $N = 4$	300	0.61	0.61	0.63	0.22	34375	7437	8326
Poisson $N = 8$	∞	1.7	0.35	1.9	0.86	125625	8554	48166
Poisson $N = 8$	300	2.9	1.8	2.4	1.7	20351	4644	4312
Toeplitz $N = 8$	∞	0.46	0.55	0.72	0.20	70000	7222	20910
Toeplitz $N = 8$	300	0.67	1.0	0.82	0.51	20624	1955	1718

Two model-free references use exactly the same primitive. The first is time-evolution quantum Krylov linear solving [25], [26]: the states $e^{-iAt_k}|b\rangle$ span a subspace whose overlap and projected matrices are $S_{kl} = g(t_l - t_k)$ and $H_{kl} = h(t_l - t_k)$ with $h(t) = \langle b|Ae^{-iAt}|b\rangle$ (one controlled application of A per Hadamard test), the projected system $Hc = r$ is solved after thresholding the small eigenvalues of S, and the observables follow from c and the same polarised Hadamard tests. Its input data are the mixture functions g and h, so the weights of weak components enter with the mixture detection law, and the thresholded inversion of an ill-conditioned overlap matrix amplifies shot noise; the implementation used here is the plain thresholded variant and does not include the directional-stability constructions of [25], so its numbers bound that variant only. The second is the randomised Fourier representation $A^{-1} = i\int_0^\infty dy\, \mathbb{E}_{z\sim\mathcal{N}(0,1)}[z\, e^{-iyzA}]$, truncated at $y \le Y = \kappa \log(1/\varepsilon)$ and sampled one (y, z) per Hadamard test, which is unbiased up to the truncation but has variance $\sim Y^2/S$ and hence a shot cost $\propto \kappa^2 \log^2(1/\varepsilon)/\varepsilon^2$ [39].

Table VI shows the outcome. With 10^4 shots the spectral-measure solver returns the compliance and the energy to 0.3–2.8% on all four systems without dephasing; under $T_2 = 300$ the Poisson $N = 8$ errors rise to 8–14%, while the other systems stay below 2%; the eigenvalue term of (7) at the smallest eigenvalues, the weight error of the kernel inversion and the unresolved $\lambda = 0.121$ component of Table IV can all contribute there, and their shares were not separated. Linear functionals are harder: the cross weights c_j alternate in sign on the Toeplitz system, on which two pairs of eigenvalues are closer than one six-bit bin (0.414 and 0.417; 0.921 and 0.938, the latter with weights 0.023 and 0.80) and are merged into clusters, and the relative error stays near 48% at both budgets, which is consistent with a cancellation between merged components; on the other three systems 1.0–4.4% is reached with 1.6×10^5 shots. The thresholded Krylov solver, with the same data and shots, is worse by factors of 2 to 280 and does not improve monotonically with shots; the Fourier estimator is worse by factors of 2.5 to 100 and would need of order 10^8 shots at evolution times near 750 for a 1% compliance on Poisson $N = 8$.

A statement that does not depend on the post-processing, within its class of estimators, is the Cramér–Rao bound for a given measurement design: a lower bound on the standard deviation of locally unbiased estimators. For the compliance it must be taken jointly over the weights on the simplex, since errors of different weights can cancel in C (Appendix A). On Poisson $N = 8$ with 10^4 mixture shots and known eigenvalues the bound on the relative standard deviation is 13% for random evolution times of the kind used by the mixture baselines (Gaussian windows from 30 to 2000) and about 6% for the best design found on a grid of settings. These are not bounds on the median errors of Table VI, nor on biased or regularised

TABLE VI

OBSERVABLE-FIRST SOLVER (PROJECTIVE ESTIMATOR, MEDIAN OF SIX REPETITIONS, THE RUNS OF TABLE VII) VERSUS TWO MODEL-FREE ESTIMATORS BUILT FROM THE SAME CONTROLLED-e^{-iAt} HADAMARD TESTS (MEDIAN OF FOUR REPETITIONS): A TIME-EVOLUTION QUANTUM KRYLOV SOLVER (QKS, PROJECTED MATRICES FROM $g(t)$ AND $h(t) = \langle b|Ae^{-iAt}|b\rangle$, THRESHOLDED REGULARISATION, BEST OF A SMALL GRID OF SUBSPACE SIZES) AND THE RANDOMISED FOURIER REPRESENTATION OF A^{-1} (TRUNCATION $Y = \kappa \log 10^3$). RELATIVE ERRORS OF THE COMPLIANCE $C = \langle b|A^{-1}|b\rangle$, THE ENERGY RATIO $E = \langle x|A|x\rangle/\langle x|x\rangle$ AND A LINEAR FUNCTIONAL $L = \langle v|A^{-1}|b\rangle$ WITH A RANDOM UNIT v, FOR TOTAL BUDGETS OF 10^4 AND 1.6×10^5 SHOTS SPLIT OVER THE FOUR INPUT STATES.

Problem	T_2	Spectral: C / E / L		QKS: C / E / L		Fourier: L
		10^4	1.6×10^5	10^4	1.6×10^5	10^4 / 1.6×10^5
Random 4×4	∞	0.31 / 0.36 / 3.0%	0.09 / 0.11 / 1.0%	3.5 / 12 / 51%	1.5 / 2.4 / 15%	168 / 40%
Random 4×4	300	0.54 / 0.94 / 4.6%	0.14 / 0.23 / 2.1%	19 / 27 / 81%	3.3 / 2.7 / 27%	131 / 44%
Poisson $N = 4$	∞	1.7 / 2.6 / 3.4%	0.45 / 0.46 / 1.6%	41 / 61 / 43%	9.2 / 6.6 / 17%	102 / 16%
Poisson $N = 4$	300	0.94 / 1.8 / 6.5%	0.33 / 0.62 / 1.6%	15 / 31 / 38%	8.1 / 18 / 14%	142 / 22%
Poisson $N = 8$	∞	1.4 / 2.8 / 11%	0.84 / 1.1 / 2.5%	29 / 459 / 71%	25 / 301 / 56%	584 / 157%
Poisson $N = 8$	300	8.3 / 14 / 12%	2.9 / 4.9 / 4.4%	38 / 302 / 81%	21 / 177 / 59%	923 / 445%
Toeplitz $N = 8$	∞	1.1 / 1.0 / 48%	0.76 / 1.1 / 48%	57 / 62 / 283%	9.8 / 18 / 95%	451 / 117%
Toeplitz $N = 8$	300	0.66 / 1.6 / 47%	0.81 / 0.47 / 50%	73 / 61 / 256%	3.2 / 8.8 / 142%	728 / 128%

estimators, and an estimate from the approximate cumulative distribution function of [24] may use another design and is not implemented here; they indicate the scale at which mixture data constrain C, which is four to nine times the 1.4% median error of the labelled estimate (Table VI). The Krylov and Fourier errors are above this scale, so they should be read as what the plain variants achieve rather than as the best mixture post-processing.

The comparison fixes the regime of the observable-first solver: few weighted components (m small, $w_{\min}$ not small), where learning the measure costs $O(w_{\min}^{-1})$ samples for detection, a refinement time that grows as $1/\varepsilon$ without a cap on the evolution time and as $1/\varepsilon^2$ with one, and $O(1/\varepsilon^2)$ samples for the weights; when m grows with the dimension the model-free estimators, or the coherent solvers, take over.

With the sequential estimator the weights are label frequencies, $\hat{w}_j = N_j/S$. If the labels were exact, the counts multinomial with fixed S, and the eigenvalue errors unbiased and independent of each other and of the counts, then

$$\operatorname{Var} \hat{C} = \frac{V}{S} + \sum_j \frac{w_j^2}{\lambda_j^2}\, \mathbb{E}[e_j^2], \quad V = \sum_j \frac{w_j}{\lambda_j^2} - C^2, \tag{8}$$

to first order. The implemented estimator shares copies between identification, clustering and refinement, so (8) is a model and not a derived property. The coefficient $\sqrt{V}/C$ of $1/\sqrt{S}$ in the relative error is 0.21, 0.69, 2.0 and 0.63 for the random, Poisson $N = 4$, Poisson $N = 8$ and Toeplitz inputs. In the runs of Table V with the target $\delta = 10^{-3}$ the measured median compliance error is 0.86% with 4.8×10^4 copies on Poisson $N = 8$, where the first term of (8) predicts a median of 0.61%, and 0.09% with 5.1×10^4 copies on the random system, where it predicts 0.06%: the model gives the scale, within a factor of 1.5. For observables we use it to set the refinement target, as a heuristic: each component is refined until its eigenvalue term is a quarter of its $1/\hat{m}$ share of the first term, and no further.

Table VII compares the two labelled estimators on the three observables, with the projective estimator re-run so that error and time come from the same runs. Here the projective estimator refines on every shot of the $|b\rangle$ run, as in Table VI, and the comparison should be read together with Table V, which shows that a target also helps the projective weights. For the compliance and the energy the two estimators are comparable: at the smaller time the sequential estimate of C is the more accurate on three of the eight combinations, at the larger time on seven (0.08–0.59% against 0.09–0.84% without dephasing). The linear functional is where the sequential estimator is more consistently better: at the larger time its error is smaller on all eight combinations, by factors of 1.2 to 19, and at the smaller time on seven of the eight. The largest gain is on the Toeplitz system, from 48% to 2.5% without dephasing and from 50% to 8.4% at $T_2 = 300$. There the sequential estimator separates the pair at 0.921 and 0.938 in most runs (Sec. IV) and leaves the pair 0.0034 apart merged, which suggests that the cancellation error of the projective estimate comes from the former pair; we have not tested this by changing the resolution of the projective estimator alone.

VI. RETURNING THE SPECTRUM TO THE SOLVER

A. A real clock register

The idealised fidelities of Table I assume that each eigencomponent is tagged exactly. With a real n_c-bit clock the QPE amplitude of a component spreads over the bins with the well-known $1/\delta^2$ tails (about 19% outside the two nearest bins), and the inverse QPE returns to the clock state $|0\rangle$ only the component of the rotated amplitude vector along the unrotated one; the remainder leaks. We evaluate this for the ideal circuit, the inverse QPE taken as the exact inverse of the forward circuit including its dephasing phases, so that independent noise in the inverse is not modelled, with the rotation applied only on the ± 1 bins of each learned eigenvalue, the other bins being removed by the ancilla post-selection as in HHL's filter functions; the reported quantity is the fidelity of the joint solution-and-clock state with $|x\rangle|0\rangle$ in the heralded ensemble (Appendix A), a lower bound on the fidelity of the solution register alone.

Table VIII shows that the learned angles improve on the same clock without learning by 0.03–0.30 on the Poisson and Toeplitz systems, except on Poisson $N = 8$ at $T_2 = 300$, where the two agree within the Monte Carlo error, but that the gap to the idealised solver (0.05–0.19) is set by the clock itself. On the random 4×4 system, whose eigenvalues are

TABLE VII

OBSERVABLES FROM THE LEARNED MEASURE WITHIN THE SAME BUDGET OF TOTAL EVOLUTION TIME: PROJECTIVE ESTIMATOR (WEIGHTS FROM THE KERNEL INVERSION) AND SEQUENTIAL ESTIMATOR (WEIGHTS FROM LABEL FREQUENCIES, REFINEMENT TARGET SET FROM THE SENSITIVITY OF C). RELATIVE ERRORS OF C / E / L, MEDIAN OF SIX INDEPENDENT REPETITIONS EACH; THE BUDGET OF THE SEQUENTIAL ESTIMATOR IS THE MEDIAN TIME OF THE SIX PROJECTIVE RUNS WITH 10^4 AND 1.6×10^5 SHOTS, MEASURED ON THOSE RUNS (WHICH INDIVIDUALLY USE 0.78–1.02 OF IT), AND IS NOT EXCEEDED.

		time of 10^4 shots		time of 1.6×10^5 shots	
Problem	T_2	Projective	Sequential	Projective	Sequential
Random 4×4	∞	0.31 / 0.36 / 3.0%	0.23 / 0.40 / 2.3%	0.09 / 0.11 / 1.0%	0.08 / 0.08 / 0.54%
Random 4×4	300	0.54 / 0.94 / 4.6%	0.59 / 0.75 / 4.9%	0.14 / 0.23 / 2.1%	0.15 / 0.17 / 0.60%
Poisson $N = 4$	∞	1.7 / 2.6 / 3.4%	0.35 / 0.54 / 1.7%	0.45 / 0.46 / 1.6%	0.12 / 0.23 / 0.49%
Poisson $N = 4$	300	0.94 / 1.8 / 6.5%	1.0 / 2.7 / 2.4%	0.33 / 0.62 / 1.6%	0.22 / 0.70 / 1.3%
Poisson $N = 8$	∞	1.4 / 2.8 / 11%	2.4 / 2.5 / 7.6%	0.84 / 1.1 / 2.5%	0.59 / 1.2 / 1.4%
Poisson $N = 8$	300	8.3 / 14 / 12%	8.5 / 15 / 12%	2.9 / 4.9 / 4.4%	2.9 / 4.8 / 3.0%
Toeplitz $N = 8$	∞	1.1 / 1.0 / 48%	0.82 / 0.81 / 8.6%	0.76 / 1.1 / 48%	0.32 / 0.57 / 2.5%
Toeplitz $N = 8$	300	0.66 / 1.6 / 47%	2.5 / 4.2 / 24%	0.81 / 0.47 / 50%	0.64 / 1.4 / 8.4%

TABLE VIII

MEDIAN JOINT FIDELITY OF THE SOLUTION-AND-CLOCK STATE WITH $|x\rangle|0\rangle$ FOR A REAL n_c-BIT UNIFORM CLOCK REGISTER AND THE SPECTRUM LEARNED BY THE PROJECTIVE ESTIMATOR (TEN REPETITIONS). "LEARNED ANGLES": ROTATION ONLY ON THE ± 1 BINS AROUND EACH LEARNED EIGENVALUE, OTHER BINS FILTERED BY THE ANCILLA POST-SELECTION. "$+t_0$": CLOCK SCALING CHOSEN TO PLACE THE LEARNED EIGENVALUES ON BINS. "NO LEARNING": ROTATION $1/\lambda_y$ ON EVERY BIN ABOVE $\lambda_{\min}$. "EXACT-SPECTRUM t_0": NO LEARNING, BUT t_0 CHOSEN FROM THE EXACT SPECTRUM (A REFERENCE, NOT AN UPPER BOUND: WITH EIGHT EIGENVALUES ON 32 BINS NO SCALING PLACES THEM ALL). UNDER DEPHASING THE FIDELITY IS THAT OF THE HERALDED ENSEMBLE, WITH ONE PHASE-KICK TRAJECTORY OF THE CLOCK PER SAMPLE ACTING ON ALL EIGENCOMPONENTS (256 SAMPLES; MONTE CARLO STANDARD ERROR OF THE TABULATED MEDIANS BETWEEN 0.001 AND 0.010, AT MOST 0.013 IN A SINGLE REPETITION, STORED WITH EVERY REPETITION); POST-SELECTION PROBABILITIES ARE NOT INCLUDED IN THE FIDELITY.

Problem	T_2	n_c	Idealised	Learned angles	$+t_0$	No learning	Exact-spectrum t_0
Random 4×4	∞	5	1.0000	0.948	0.993	0.964	0.995
Random 4×4	300	5	1.0000	0.898	0.926	0.934	0.968
Poisson $N = 4$	∞	4	1.0000	0.920	0.968	0.623	0.958
Poisson $N = 4$	300	4	1.0000	0.834	0.906	0.621	0.906
Poisson $N = 8$	∞	5	1.0000	0.909	0.914	0.878	0.817
Poisson $N = 8$	300	5	0.924	0.731	0.700	0.737	0.693
Toeplitz $N = 8$	∞	6	1.0000	0.950	0.989	0.902	0.996
Toeplitz $N = 8$	300	5	0.9995	0.871	0.898	0.617	0.889

all large, learning the angles is slightly worse than rotating on every bin, because the bins outside the learned peaks are discarded. With a uniform clock the most effective use of the learned spectrum is to choose the clock scaling t_0 so that the learned eigenvalues sit on bins, the strategy of hybrid HHL++ [8]: this raises the fidelity to 0.993 (random 4×4) and 0.989 (Toeplitz), close to the reference that uses the exact spectrum for the same purpose. It fails for Poisson $N = 8$, whose eight eigenvalues cannot be placed on 32 bins simultaneously (a greedy split needs four separate passes), and under $T_2 = 300$ the coarse clock itself degrades every variant to 0.69–0.74. The conclusion is that improving the learning stage alone does not raise the fidelity of a hybrid solver; the clock design (n_c, t_0, clock state, filtering) dominates.

B. A tapered clock

The leakage that limits Table VIII is that of the uniform clock state. Preparing the clock register in a window state confines the amplitude of each eigencomponent to a main lobe at the price of a wider lobe; the original HHL circuit uses a sine state for this reason [1], cosine and optimal discrete prolate spheroidal windows have been analysed for phase estimation [27]–[29], [40] and for coherent subroutines [41], and the optimal window can be prepared with a matrix-product state of bond dimension four [29]. Table IX applies this known technique to the learned rotation table.

Without dephasing it removes most of the loss: with $NW = 2$ and rotation on ± 2 bins the fidelity is 0.99988 on the random system ($n_c = 5$), 0.99994 on Poisson $N = 4$ ($n_c = 5$), 0.99990 on Poisson $N = 8$ ($n_c = 6$) and 0.99954 on the Toeplitz system ($n_c = 6$), where the uniform clock gives 0.87–0.95 and the optimised t_0 at most 0.993. The two ablation columns show that the window, not the wider rotation, is responsible: widening the rotation of the uniform clock gives 0.91–0.97 in these four cases, while the window with the narrow rotation already gives 0.990–0.9993. The sine state obtains most of the gain, and the success probability of the post-selection changes by at most 11%. The wider lobe must still fit between the eigenvalues: with one bit fewer the fidelity is 0.998 on Poisson $N = 4$ and 0.962–0.987 on Poisson $N = 8$.

At $T_2 = 300$ the gain is small. The best window improves on the uniform clock of the same size by 0.03–0.06 at the register size of Table VIII and by 0.07–0.10 with one more bit, where the uniform clock is lower; it is below the optimised t_0 on Poisson $N = 4$ and, with the larger register, on Poisson $N = 8$, and no clock state exceeds 0.95; dephasing enters as a phase jump on each clock qubit, which spreads amplitude over the register whatever the window [42]. Without dephasing,

TABLE IX

MEDIAN JOINT FIDELITY (AS IN TABLE VIII) WITH AN n_c-BIT CLOCK AND THE SPECTRUM LEARNED BY THE SEQUENTIAL ESTIMATOR (TEN REPETITIONS, $T_{\mathrm{tot}} = 2 \times 10^6$), FOR THE UNIFORM CLOCK STATE, THE SINE STATE OF [1] AND DISCRETE PROLATE SPHEROIDAL (DPSS) WINDOWS WITH TIME-BANDWIDTH PRODUCT NW. THE ROTATION IS APPLIED ON THE BINS WITHIN $\pm h$ OF EACH LEARNED EIGENVALUE; h IS GIVEN WITH EACH CLOCK STATE, AND THE COLUMNS "UNIFORM ± 2" AND "$NW = 2, \pm 1$" SEPARATE THE EFFECT OF THE WINDOW FROM THAT OF THE WIDER ROTATION.

Problem	T_2	n_c	Uniform ± 1	$+t_0$ ± 1	Uniform ± 2	Sine ± 2	$NW = 1.5$, ± 2	$NW = 2$, ± 2	$NW = 2$, ± 1
Random 4×4	∞	5	0.948	0.9928	0.966	0.99980	0.99974	0.99988	0.9993
Random 4×4	∞	6	0.929	0.988	0.966	0.9993	0.99961	0.99998	0.9942
Random 4×4	300	5	0.898	0.930	0.930	0.948	0.947	0.944	0.918
Random 4×4	300	6	0.787	0.840	0.877	0.889	0.885	0.879	0.797
Poisson $N = 4$	∞	4	0.920	0.968	0.927	0.986	0.9981	0.9979	0.9963
Poisson $N = 4$	∞	5	0.908	0.977	0.929	0.9989	0.9991	0.99994	0.9953
Poisson $N = 4$	300	4	0.836	0.906	0.833	0.865	0.867	0.858	0.871
Poisson $N = 4$	300	5	0.785	0.867	0.830	0.859	0.860	0.852	0.823
Poisson $N = 8$	∞	5	0.909	0.914	0.905	0.987	0.980	0.962	0.960
Poisson $N = 8$	∞	6	0.869	0.989	0.907	0.99955	0.9993	0.99990	0.9902
Poisson $N = 8$	300	5	0.749	0.775	0.743	0.784	0.767	0.745	0.759
Poisson $N = 8$	300	6	0.701	0.807	0.745	0.772	0.761	0.755	0.714
Toeplitz $N = 8$	∞	6	0.950	0.9905	0.974	0.9986	0.9992	0.99954	0.9969
Toeplitz $N = 8$	∞	7	0.922	0.9929	0.955	0.9990	0.9992	0.99954	0.9921
Toeplitz $N = 8$	300	5	0.862	0.890	0.906	0.921	0.921	0.917	0.875
Toeplitz $N = 8$	300	6	0.789	0.830	0.861	0.866	0.863	0.855	0.788

then, the learned spectrum and a tapered clock together bring the hybrid solver to an infidelity between 6×10^{-5} and 5×10^{-4}; with dephasing the coherent clock remains the limit.

C. Spectrum-aware inversion

For observables that are not diagonal in the eigenbasis of A a quantum back end is needed. A clock is not the only way to use the learned spectrum there. QSVT inversion [2], [3] applies a polynomial $p(A)$ and its output depends on p only at the eigenvalues, so a polynomial that interpolates $\lambda_{\min}/x$ at the learned nodes can have a far lower degree than the standard $O(\kappa \log(1/\varepsilon))$ approximation on the whole interval; this observation and a practical min-norm correction of a base polynomial were made in [30] for known eigenvalues. Two constraints make the comparison non-trivial: QSVT requires odd parity and $|p| \leq 1$ on $[-1, 1]$, so a naive Lagrange interpolant, which reaches $F = 0.99999$, is unusable because its maximum on $[-1, 1]$ exceeds its value on the spectrum by a factor 3×10^3 (Poisson $N = 8$), which enters the success probability squared.

Table X compares, under both constraints, the min-norm spectral correction at all learned eigenvalues, a linear program that minimises the weighted node residual $\sum_j (\sqrt{w_j}/\mu_j^2)|p(\mu_j) - \lambda_{\min}/\mu_j|$ subject to the bound, and the same program for the uniform interval error. Over the forty learning runs the spectral correction needs the lowest degree. On Poisson $N = 4$ it reaches $F \geq 0.9999$ at degree 7 in 39 of 40 runs, where the interval polynomial needs degree 55 for $F \geq 0.999$ and does not reach 0.9999 up to degree 79; on Poisson $N = 8$ it reaches $F \geq 0.999$ in 37 of 40 runs, in 26 of them at degree 35, where the interval polynomial reaches neither target. It pays in subnormalisation when the degree is very low ($\tau = 4.1$, a factor 17 in success probability, on Poisson $N = 4$), and the tighter target is sensitive to the learned nodes: on Poisson $N = 8$ only 13 of 40 runs reach $F \geq 0.9999$, ten of them at degree 39. The bounded program keeps $\tau \approx 1$ at a higher degree (15–17 on Poisson $N = 4$) and on Poisson $N = 8$ reaches $F \geq 0.999$ in 16 of 40 runs and $F \geq 0.9999$ in one. The two sit at the ends of one degree-versus-success-probability trade-off, and neither is reliable at $F \geq 0.9999$ on the system with the smallest eigenvalue.

The learned nodes carry a second risk: on the Toeplitz system the learning stage misses a component of weight 10^{-3} at $\lambda = 1$, and the node program, which does not control the missed node, reaches $F \geq 0.999$ in only 8 of 40 runs; in the run analysed it saturates at $F = 0.9895$, with 99.8% of the infidelity coming from that component. The spectral correction, whose base polynomial is bounded on the whole interval, reaches $F \geq 0.9999$ at degree 21 in all forty runs ($\tau = 1.30$), and the interval polynomial at degree 39. The dependence on learning errors is bounded by (Appendix A)

$$1 - F \leq \frac{1}{\|\hat{y}\|^2} \sum_j w_j \Big[|r_j| + \Big(d^2 \max_{[-1,1]} |p| + \frac{\lambda_{\min}}{\lambda_j \mu_j} \Big) |\delta_j| \Big]^2, \quad (9)$$

with $\hat{y} = p(A)|b\rangle$, $r_j = p(\mu_j) - \lambda_{\min}/\mu_j$ the node residual and $\mu_j = \lambda_j + \delta_j$ both in $[-1, 1]$, which amplifies eigenvalue errors by the Markov factor d^2; a component without a node enters with $r_j = p(\lambda_j) - \lambda_{\min}/\lambda_j$ uncontrolled. The bound confirms that the admissibility of spectrum-aware inversion is decided by the learning stage, specifically by whether every component is captured, and it also shows that the degrees of Table X presuppose nearly exact nodes: with relative node errors of 10^{-3} the degree-79 node-aware polynomial on Poisson $N = 8$ falls to $F \approx 0.93$, and to 0.54 at 10^{-2}.

VII. APPLICATION TEST: DC POWER FLOW

The DC power-flow problem $B'\theta = P$ gives the bus voltage angles θ of a transmission network from the net real-power injections P (generation minus load) and the reduced bus susceptance matrix B'; line flows are linear functionals of θ. We build it for the 9-, 14- and 30-bus cases distributed with MATPOWER [31], with the injections of each case, and normalise to $A = B'/\lambda_{\max}$ and $|b\rangle = P/\|P\|$. The systems have 8, 13 and 29 unknowns and condition numbers 59, 119

TABLE X

ODD POLYNOMIALS WITH THE QSVT CONSTRAINT $|p| \leq 1$ ON $[-1, 1]$, BUILT ON LEARNED NODES AND EVALUATED ON THE EXACT SPECTRUM, FOR FORTY INDEPENDENT LEARNING RUNS OF THE PROJECTIVE ESTIMATOR (10^4 SHOTS, NO DEPHASING). EACH ENTRY GIVES THE NUMBER OF RUNS THAT REACH THE TARGET FIDELITY AT A DEGREE UP TO 79 AND, AFTER THE COLON, THE DEGREE THAT AT LEAST HALF OF THE FORTY RUNS REACH ("–" IF FEWER THAN HALF REACH THE TARGET), WITH THE MEDIAN SUBNORMALISATION τ OF THE REACHING RUNS IN PARENTHESES. "SC": MIN-NORM SPECTRAL CORRECTION [30] OF A BOUNDED ODD BASE POLYNOMIAL AT ALL LEARNED EIGENVALUES. "NODE LP": LINEAR PROGRAM MINIMISING THE WEIGHTED NODE RESIDUAL UNDER THE BOUND. "INTERVAL": THE SAME PROGRAM FOR THE UNIFORM ERROR ON $[\lambda_{\min}, 1]$, WHICH DOES NOT USE THE LEARNED NODES (MINIMAL DEGREE FOR $F \geq 0.999$ / $F \geq 0.9999$; "–": NOT REACHED UP TO DEGREE 79).

Problem	κ	SC, $F \geq 0.999$	SC, $F \geq 0.9999$	Node LP, $F \geq 0.999$	Node LP, $F \geq 0.9999$	Interval
Poisson $N = 8$	32	37/40: 35 (1.34)	13/40: –	16/40: –	1/40: –	– / –
Poisson $N = 4$	9.5	40/40: 7 (4.10)	39/40: 7 (4.10)	39/40: 15 (1.02)	39/40: 17 (1.00)	55 / –
Toeplitz $N = 8$	4.0	40/40: 21 (1.30)	40/40: 21 (1.30)	8/40: –	1/40: –	13 / 39
Random 4×4	2.2	40/40: 7 (1.01)	40/40: 7 (1.01)	40/40: 7 (1.01)	40/40: 7 (1.01)	13 / 17

TABLE XI

MEDIAN SOLUTION INFIDELITY $1 - F$ OF THE IDEALISED HYBRID SOLVER ON DC POWER-FLOW PROBLEMS OF STANDARD TEST SYSTEMS. ABOVE THE RULE: LEARNING STAGES WITH 10^4 SHOTS EACH AND 100 REPETITIONS, EACH WITH ITS OWN TIME RULE (TOTAL EVOLUTION TIME 4×10^5 TO 5×10^7). BELOW THE RULE: PROJECTIVE HHL WITH 10^4 SHOTS AND RECYCLED SEQUENTIAL HHL (PRECISION TARGET $\delta = 10^{-3}$) WITHIN THE MEDIAN TIME OF THE PROJECTIVE RUNS (2×10^7 WITHOUT DEPHASING, 4×10^6 AT $T_2 = 300$), AT THE RESOLUTION OF THE REGISTER (TEN REPETITIONS) AND AT ITS FINEST RESOLUTION (FIVE REPETITIONS). RUNS WITHOUT A FINITE ESTIMATE ARE EXCLUDED (FOUR OF 3000 ABOVE THE RULE). SEVERAL MEDIANS LIE BETWEEN GROUPS OF RUNS; SEE THE TEXT. "–": NOT RUN (REGISTER LONGER THAN $4T_2$); "SAME": THE TWO RESOLUTIONS COINCIDE.

	9-bus ($\kappa = 59$)		14-bus ($\kappa = 119$)		30-bus ($\kappa = 493$)	
Learning stage	no dephasing	$T_2 = 300$	no dephasing	$T_2 = 300$	no dephasing	$T_2 = 300$
Bayesian clock (mixture)	0.230	0.230	0.055	0.052	0.772	0.772
QMEGS-type (mixture)	0.210	0.197	0.012	0.012	0.692	0.673
MM-QCELS-type (mixture)	0.216	0.214	2.6e-4	0.012	0.688	0.685
QPE 6 bits	0.187	0.186	0.015	0.015	0.668	0.675
QPE 8 bits	0.188	–	3.7e-4	–	0.095	–
QPE 10 bits	0.188	–	1.1e-5	–	0.064	–
Projective HHL	0.187	0.175	0.012	0.012	0.674	0.775
Recycled sequential HHL, register resolution	4.3e-7	0.114	5.6e-5	0.012	0.180	0.689
Recycled sequential HHL, finest resolution	same	same	3.3e-7	0.012	0.066	0.637

and 493; 5, 9 and 20 eigencomponents hold 99% of the weight of the input, so the spectral measures are less sparse than those of Sec. IV, and the smallest gap (0.047, 0.0072, 0.0037) is below the bin of a six-bit register (0.020) on the two larger systems. Time is in units of $1/\|B'\|$. Table XI applies the learning stages of Table I with the same number of shots, and the equal-time protocol of Table IV to the two labelled estimators.

Each system has a weak component that limits the fidelity. As a reference we compute, from the eigendecomposition, the infidelity that results when one component is rotated with the nearest other eigenvalue while all other eigenvalues are exact: 0.19 for the smallest eigenvalue of the 9-bus system ($\lambda = 0.017$, 0.18% of the weight of the input, 27% of the solution), 0.012 for $\lambda = 0.042$ of the 14-bus system (0.33% of the weight) and 0.065 for $\lambda = 0.0069$ of the 30-bus system (0.07%). These are reference levels, not measurements. The learned eigenvalues of every run are stored, and we say that a run assigns a value to a component when the learned value nearest to the component is nearer to it than to any other eigenvalue; a component without an assigned value is rotated with a value that belongs to a neighbour.

The two parts of Table XI follow different protocols. The rows above the rule use 100 repetitions and between 0.10 and 2.4 times the evolution time of the rows below. The rows below use ten repetitions, five at the finest resolution where it differs from the register resolution, and the sequential estimator has the precision target $\delta = 10^{-3}$. Four of the 3000 runs above the rule gave no finite estimate and are excluded from the medians: two MM-QCELS-type fits on the 30-bus system without dephasing did not converge, and two QMEGS-type runs on the same system at $T_2 = 300$ returned a non-finite value. Counting them as failures changes no entry of the table by more than one unit of its last digit.

9-bus system. The medians of the three mixture estimators ($w\sqrt{S} = 0.18$), of the QPE histograms with six, eight and ten bits and of the projective estimator are at the reference level (0.18–0.23), with and without dephasing. Without dephasing none of the 300 QPE-histogram runs and none of the ten projective runs assigns a value to the smallest eigenvalue. The Bayesian clock, the QMEGS-type and the MM-QCELS-type estimator assign one in 12, 25 and 31 of 100 runs, but it is accurate enough for $1 - F < 0.1$ in only 3, 12 and 8 of them. The sequential estimator assigns a value in all ten runs without dephasing (median $1 - F = 4.3 \times 10^{-7}$) and in five of ten at $T_2 = 300$. The median at $T_2 = 300$, 0.11, lies between two groups of runs (4×10^{-4} to 0.04, and 0.19) and is not a value that any run takes.

14-bus system. Without dephasing the medians of the projective and the QMEGS-type estimator are at the reference level 0.012, and that of the six-bit QPE histogram is 0.015. The MM-QCELS-type estimator is below 10^{-3} in 54 of 100

runs (median 2.6×10^{-4}, with a ninth of the time) and assigns a value to the component in 26 of 100 runs; in its other low runs the component is rotated with a learned value that lies nearer to a neighbouring eigenvalue. The QPE histograms with eight bits (3.7×10^{-4}) and ten bits (1.1×10^{-5}, at 2.4 times the time) assign a value in 98 of 100 runs each. The sequential estimator assigns one in all runs (5.6×10^{-5} at the register resolution, 3.3×10^{-7} at its finest). The six-bit register is not the limit of the projective estimator: with a nine-bit clock, which resolves the smallest gap, and twice the time it assigns a value in six of ten runs (median 1.8×10^{-6}, four runs at 0.010–0.012), and the sequential estimator in the same time in all ten (6.3×10^{-8}). At $T_2=300$ every median is at the reference level or above it; the component is assigned a value in four of ten sequential runs, in none of the ten projective runs and in 30 and 23 of 100 QMEGS-type and MM-QCELS-type runs.

30-bus system. Twenty components hold 99% of the weight and no learning stage resolves them at this budget. The medians of the mixture estimators, the six-bit QPE histogram and the projective estimator are 0.67–0.77. In the ten projective runs the three to five smallest eigenvalues, and 18 to 20 of the 29 eigenvalues in all, are without an assigned value; the reference levels for the four and the five smallest are 0.66 and 0.69. The sequential estimator at its finest resolution (0.066, five runs) and the ten-bit QPE histogram at 2.4 times the time (0.064) are at the reference level of the $\lambda=0.0069$ component, which is without an assigned value in all five sequential runs and in 77 of the 100 QPE runs. The median of the sequential estimator at the register resolution, 0.18, comes from three groups of runs (two near 0.07, four at 0.17–0.18, four at 0.61–0.67). At $T_2=300$ every median is above 0.6.

Observables. As a linear functional we take the flow on one line, obtained with the polarisation identity of Sec. V. The line is the one with the largest flow in the classical solution; it is fixed beforehand from that solution, and finding it is not part of the test. We also report the DC quadratic form $C=P^{\mathsf{T}}\theta=\sum_l f_l^2/b_l$ (f_l the flow and b_l the susceptance of line l), multiplied by the base power. For lossless lines, unit voltages, small angles and nominal transformer taps it is the series reactive power consumed by the lines: the angle-dependent part of that consumption agrees with C to within 0.11% on the three systems, and the three off-nominal taps of the 14-bus system add 2.5 MVAr (4.6%) that does not depend on the angles and is not contained in C.

At the larger observable budget (1.6×10^5 projective shots, six repetitions, same time) and without dephasing the error on the 163 MW line of the 9-bus system is 1.8 MW with the sequential estimator and 109 MW with the projective one, and the error of $C=48.7$ MVAr is 0.06 against 1.2 MVAr. On the 148 MW line of the 14-bus system the errors are 10–80 MW over the five settings tested and the sequential estimator is the more accurate in four of them (not at $T_2=300$ with 10^4 shots, 80 against 35 MW); on the 37 MW line of the 30-bus system they are 2–7 MW and the projective estimator is the more accurate in four of five. C is more accurate with the sequential estimator in all fourteen settings of these observable runs; in the fidelity runs of Table XI, where the refinement target is not set from C, the order is reversed on the 14-bus system at $T_2=300$ (error of C 2.2% with the projective and 4.1% with the sequential estimator). The model-free estimators of Table VI (four repetitions) fail on these systems: the Fourier estimate of the flow is off by more than 100% in every setting, and the Krylov estimates by 7–104% for the flow and 10–85% for C.

Real clock register. With a real clock register (joint fidelity as in Table IX, ten repetitions) and no dephasing we compare complete configurations, which does not separate the effect of learning from that of the clock length and the window. Standard HHL with a ten-bit uniform clock and no learning gives 0.985, 0.965 and 0.77 on the 9-, 14- and 30-bus systems. The projectively learned spectrum with a six-bit uniform clock gives 0.66, 0.84 and 0.31, and the sequentially learned spectrum with a seven-bit uniform clock 0.87, 0.98 and 0.62, so that with a uniform clock a learned spectrum on a short register is worse than no learning on a long one except on the 14-bus system with the sequential estimator. The sequentially learned spectrum with a seven-bit DPSS clock gives 0.9999 on the 9- and 14-bus systems; on the 30-bus system, whose spectrum the learning stage does not resolve at this budget, the best tapered clock gives 0.69 and standard HHL with ten bits is the best (0.77). At $T_2=300$ no configuration exceeds 0.73.

Polynomial inversion. Spectrum-aware inversion (Sec. VI-C) does not carry over at this budget: on learned nodes the node-aware and the interval polynomial do not reach $F\geq0.9999$ up to degree 199 on any of the three systems, and the spectrally corrected polynomial reaches it in one of the six combinations of system and learning stage (9-bus, sequential nodes, degree 51, $\tau=1.6$).

More time helps where resolution is the limit. On the 30-bus system the sequential estimator at its finest resolution reaches $1-F=0.065$, 2.0×10^{-5} and 9.6×10^{-7} at one, four and sixteen times the budget of Table XI (five repetitions), with 22, 27 and all 29 components found, whereas the projective estimator stays at 0.67–0.68 with a six-bit clock and at 0.065–0.067 with a ten-bit clock at three and thirteen times the budget. The 118-bus case (117 unknowns, $\kappa=2.9\times10^3$, 72 components for 99% of the weight) is outside the sparse regime: at sixteen times the budget the sequential estimator finds 60 components and reaches $1-F=0.008$, the ten-bit projective estimator reaches 0.09 at thirteen times the budget, and the six-bit one stays at 0.77. All of these systems are solved classically in microseconds. What the test shows is how the estimators behave on weight distributions that were not constructed for the purpose, in which a component with 0.2% (9-bus) or 0.4% (30-bus) of the weight sits at the smallest eigenvalue and carries 27% or 81% of the solution, and where every estimator fails.

VIII. Application test: readout of a reservoir computer

As a second application we train the linear readout of a reservoir computer for nonlinear time series, a least-squares problem of the kind for which quantum linear solvers were

TABLE XII

LEARNING STAGES ON THE RIDGE SYSTEMS OF A RESERVOIR COMPUTER FOR THREE TIME SERIES, ALL WITHIN THE SAME EVOLUTION TIME, TEN TIMES THE MEDIAN TIME OF 10^4 PROJECTIVE SHOTS WITHOUT DEPHASING (2.2×10^8); TEN RUNS EACH. ENTRIES: MEDIAN SOLUTION INFIDELITY $1 - F$ OF THE IDEALISED HYBRID SOLVER AND, IN PARENTHESES, THE NUMBER OF THE TEN RUNS THAT REPRODUCE THE FORECAST OF THE EXACT RIDGE READOUT ($D < 10^{-3}$, SEE THE TEXT). WITHOUT DEPHASING PROJECTIVE HHL HAS 10^5 SHOTS AND EXCEEDS THE TIME BY UP TO 4%. THE FINEST RESOLUTION OF THE SEQUENTIAL ESTIMATOR IS 3×10^{-4} WITHOUT DEPHASING AND 0.002 AT $T_2 = 300$. "–": NOT RUN (REGISTER LONGER THAN $4T_2$).

	NARMA2		Mackey–Glass		Lorenz	
Learning stage	no dephasing	$T_2 = 300$	no dephasing	$T_2 = 300$	no dephasing	$T_2 = 300$
Bayesian clock (mixture)	0.443 (0)	0.448 (0)	0.285 (1)	0.284 (1)	0.781 (0)	0.781 (0)
QMEGS-type (mixture)	0.305 (0)	0.306 (1)	0.068 (4)	0.055 (6)	0.180 (0)	0.305 (0)
MM-QCELS-type (mixture)	0.431 (0)	0.431 (0)	0.059 (7)	0.056 (10)	0.611 (0)	0.370 (1)
QPE 6 bits	0.068 (4)	0.254 (0)	0.390 (5)	0.058 (10)	0.133 (0)	0.363 (0)
QPE 8 bits	0.431 (0)	–	0.052 (10)	–	0.015 (10)	–
QPE 10 bits	0.192 (0)	–	0.055 (10)	–	0.016 (10)	–
Projective HHL	0.431 (0)	0.158 (2)	0.050 (9)	0.049 (6)	0.022 (5)	0.097 (4)
Recycled sequential HHL, $\sigma_{\mathrm{id}} = 0.005$	0.104 (5)	0.104 (4)	1.0e-3 (10)	2.0e-3 (10)	0.016 (10)	0.015 (10)
Recycled sequential HHL, finest resolution	2.7e-3 (10)	3.4e-3 (10)	1.0e-3 (10)	0.055 (10)	1.5e-4 (10)	0.015 (10)

proposed early [35]. An echo state network [36] of 16 tanh nodes with fixed random couplings (spectral radius 0.9) is driven by the input series; after 200 discarded steps, 2000 steps train the readout and the next 1000 test it. With Z the standardised node states and y the centred target, ridge regression solves $(G + \alpha)W = g$ with $G = Z^{\mathsf{T}}Z/2000$ and $g = Z^{\mathsf{T}}y/2000$. We set $\alpha = \lambda_{\max}(G)/99$, so that $A = (G + \alpha)/(\lambda_{\max}(G) + \alpha)$ has $\kappa \leq 100$, and $|b\rangle = g/\|g\|$. The series are NARMA2 driven by uniform noise [43] (target: the output at the same step), the Mackey–Glass system with delay 17 [44] (six time units ahead) and the x coordinate of the Lorenz system [45] (five integration steps of 0.02 ahead).

Forecast error. A learned spectrum is turned into a readout as the idealised hybrid solver does it: every eigencomponent of $|b\rangle$ is divided by the learned eigenvalue nearest to its own, here without normalising the solution, so the forecast measures the learning stage with an ideal back end. We report the normalised mean squared error (NMSE) of the forecasts on the test steps and the mean squared difference D between these forecasts and those of the exact ridge readout, both in units of the variance of the target, and say that a run reproduces the ridge forecast when $D < 10^{-3}$. The exact ridge readout has NMSE 0.120, 0.235 and 0.132 on the three series, against 0.84, 0.67 and 0.26 for repeating the last known value. The bound on κ costs accuracy (without regularisation 0.056, 0.027 and 0.0021), so the reference is the regularised readout and not the best forecast.

Spectra. The spectral measures are sparse in the weight of the input but not in the solution. Three, two and one eigencomponents hold 99% of the weight of $|b\rangle$. Nine, thirteen and thirteen of the sixteen eigenvalues lie below the bin of a six-bit register (0.020), most of them near the regularisation floor; together they carry 3×10^{-4}, 7×10^{-5} and 9×10^{-4} of the weight and 40%, 8% and 72% of the solution. On the Lorenz series a single component at $\lambda = 0.016$ with weight 7.4×10^{-4} holds 51% of the solution.

Protocol. All learning stages get the same budget of evolution time, ten times the median time of 10^4 projective shots with a six-bit register and no dephasing, with and without dephasing and ten runs each (Table XII). The mixture estimators and the QPE histograms get a number of shots with which every run stays within the budget (they use 86% to 100% of it), and the sequential estimator gets the budget itself. Projective HHL gets the shots that keep every run within the budget at $T_2 = 300$ (4.7×10^5 to 4.9×10^5 with a five-bit register); without dephasing it gets 10^5 shots, whose time exceeds the budget by 2% to 4%. The sequential estimator is run at two identification resolutions, $\sigma_{\mathrm{id}} = 0.005$ and the finest one used, 3×10^{-4} without dephasing and $0.002 = 0.6/T_2$ at $T_2 = 300$; eigenvalues at the regularisation floor differ by as little as 10^{-7} and are not meant to be separated. Every run returned a finite estimate.

NARMA2. Only the sequential estimator at its finest resolution reproduces the ridge forecast, in all ten runs with and without dephasing (median $1 - F = 2.7 \times 10^{-3}$ and 3.4×10^{-3}). At $\sigma_{\mathrm{id}} = 0.005$ it does so in five and four runs, the six-bit QPE histogram in four runs without dephasing, and no other stage in more than two. Projective HHL is not limited by the number of shots here: with 10^5 shots its median infidelity is 0.43, as with 10^4, and in no run does it learn a value below 0.3, although thirteen eigenvalues lie there.

Mackey–Glass. Most stages reproduce the forecast: the sequential estimator in all ten runs at both resolutions and noise settings, the eight- and ten-bit QPE histograms without dephasing, the MM-QCELS-type estimator and the six-bit histogram at $T_2 = 300$, and Projective HHL in nine and six runs. A median infidelity near 0.05 is compatible with a reproduced forecast on this series. The infidelity itself is lowest for the sequential estimator at $\sigma_{\mathrm{id}} = 0.005$ (1.0×10^{-3} and 2.0×10^{-3}); at $T_2 = 300$ the finest resolution gives 0.055, so the coarser one is the better choice here.

Lorenz. The sequential estimator reproduces the forecast in all ten runs at both resolutions and noise settings, with $1 - F = 1.5 \times 10^{-4}$ at the finest resolution without dephasing and 0.015 to 0.016 otherwise. The eight- and ten-bit QPE histograms do so without dephasing and are not run at $T_2 = 300$. Projective HHL reproduces it in five and four runs, and no mixture estimator in more than one.

Reading of the table. Three points limit the conclusions. First, the forecast error is far less sensitive than the fidelity: on

TABLE XIII
VARIATIONAL QUANTUM LINEAR SOLVER ON THE SYSTEMS OF TABLES I, XI AND XII AND ON THE 118-BUS SYSTEM: MEDIAN SOLUTION INFIDELITY $1-F$ OVER TEN RUNS FROM RANDOM ANGLES, n QUBITS. "TERMS": PAULI TERMS MEASURED FOR ONE VALUE OF THE COST. "EXACT": EXACT COST AND GRADIENT; "PAIRS": PAIRS OF A CIRCUIT AND A TERM EVALUATED BY THAT RUN (MEDIAN). LAST TWO COLUMNS: STOCHASTIC OPTIMISATION WITHIN 10^4 AND 10^7 CIRCUIT EXECUTIONS. "–": ONE ITERATION DOES NOT FIT. NO DEPHASING IS APPLIED (SEE THE TEXT).

System	n	terms	exact	pairs	10^4	10^7
Random 4×4	2	47	<1e-9	1.3e4	0.690	0.023
Poisson $N=4$	2	9	<1e-9	1.9e3	0.812	0.121
Poisson $N=8$	3	21	<1e-9	1.6e4	0.839	0.865
Toeplitz $N=8$	3	33	<1e-9	1.3e4	0.893	0.064
9-bus	3	63	<1e-9	7.2e4	0.946	0.447
14-bus	4	223	<1e-9	2.6e6	0.958	0.975
30-bus	5	807	<1e-9	1.2e8	0.963	0.966
118-bus	7	8127	0.579	4.8e9	–	0.994
NARMA2 readout	4	256	<1e-9	3.4e6	0.959	0.936
Mackey–Glass readout	4	256	<1e-9	2.0e6	0.959	0.967
Lorenz readout	4	256	<1e-9	1.2e6	0.971	0.952

TABLE XIV
FORECAST NMSE OF BASELINES WITHOUT PHASE ESTIMATION ON THE THREE SERIES OF SEC. VIII (MEDIAN OF TEN RUNS; ONE VALUE WHERE NOTHING IS RANDOM). "RIDGE READOUT": THE EXACT READOUT OF THE ECHO STATE NETWORK. "EXACT": EXACT EXPECTATION VALUES, OR EXACT COST FOR THE VARIATIONAL SOLVER; OTHERWISE THE NUMBER OF CIRCUIT EXECUTIONS. DQI: THE POLYNOMIAL THAT SATISFIES MOST CONSTRAINTS AMONG 10^5 SAMPLES, FOR CONSTRAINT SETS OF r LEVELS.

Method	NARMA2	Mackey–Glass	Lorenz
Ridge readout, exact	0.120	0.235	0.132
Last known value	0.838	0.668	0.263
Variational solver, exact	0.120	0.235	0.132
Variational solver, 10^5	0.968	0.691	0.437
Variational solver, 10^7	0.396	0.567	0.348
Quantum reservoir, exact	0.207	0.233	0.241
Quantum reservoir, 10^5	0.642	0.560	0.433
Circuit learning, exact	0.144	0.097	8.4e-3
Circuit learning, 10^5	1.003	0.954	0.997
DQI, $r=5$, 10^5	1.204	1.639	0.837
DQI, $r=15$, 10^5	1.609	0.655	0.267

the Lorenz series median infidelities of 0.016 and 1.5×10^{-4} of the sequential estimator give NMSE 0.137 and 0.131. Second, which resolution of the sequential estimator is the better depends on the series and the noise; the two were compared after the runs, and no rule that fixes the resolution beforehand was tested. Third, ten runs do not resolve rare failures: six of the 360 runs of the other stages give NMSE above one, the largest above 10^6, and none of the 120 sequential runs does.

Smaller budget. At a tenth of this budget, the protocol of Table XI, the sequential estimator at $\sigma_{\mathrm{id}}=0.005$ reproduces the forecast of the Lorenz series in all ten runs without dephasing and in one at $T_2=300$, and Projective HHL in two and in none. On NARMA2 no stage reproduces it in more than four of ten or 14 of 100 runs.

IX. BASELINES WITHOUT PHASE ESTIMATION

Variational solvers prepare the solution without phase estimation and are the usual proposal for devices without error correction. We run the variational quantum linear solver [18] on the four systems of Table I, on the 9-, 14-, 30- and 118-bus systems of Sec. VII and on the three ridge systems of Sec. VIII, with the dimension padded to a power of two by a unit block. The trial state $|\psi\rangle$ of n qubits is prepared by layers of single-qubit rotations R_Y (R_Y and R_Z on the complex system) separated by a chain of CNOT gates, with one layer more than the number of free amplitudes requires, and the angles minimise the global cost $C=1-|\langle b|A|\psi\rangle|^2/\langle\psi|A^2|\psi\rangle$.

Cost unit and protocols. The unit of cost of this solver is a circuit execution, not evolution time, and none of its circuits contains the controlled evolution to which T_2 refers; only the calibrated read-out contrast $1-p_d$ is applied. We therefore compare by the number of state preparations, which is 10^4 for the learning stages of Tables I and XI. Two protocols are run, ten times each from random angles. In the first the cost and its gradient are exact and the optimiser is L-BFGS with at most 4000 iterations; this is the ceiling of the method, and we count the pairs of a circuit and a Pauli term that it evaluates, which is the number of executions at one shot per pair. In the second the cost is estimated within a fixed number of executions, $\langle b|A|\psi\rangle$ from one Hadamard test per Pauli term of A and $\langle\psi|A^2|\psi\rangle$ from the Pauli terms of A^2, and the angles follow simultaneous-perturbation stochastic approximation [46]; every term gets the largest number of shots that leaves 500 iterations, and at least one. The fidelity of the final state is computed exactly in both protocols.

Results. With an exact cost the median infidelity is below 10^{-9} on ten systems (Table XIII): every run is below 10^{-9} on the nine smaller ones, and on the 30-bus system nine runs are below 10^{-8} and one, which stops at the iteration limit, ends at 1.6×10^{-4}. On the 118-bus system all ten runs stop at the iteration limit, with a median of 0.58. The exact runs on the ten systems evaluate between 1.9×10^3 and 1.2×10^8 pairs (medians), more than the 10^4 state preparations of the learning stages on all of them except Poisson $N=4$. Within 10^4 executions the median infidelity is between 0.69 and 0.97, as for a random state (expected $1-2^{-n}$, 0.75 to 0.97), with single runs between 0.18 and 1.00; on the 118-bus system, whose cost has 8127 terms, one iteration does not fit. Within 10^7 executions the median falls to 0.023–0.45 on the random, the Poisson $N=4$, the Toeplitz and the 9-bus system and stays above 0.8 on the other seven.

Cost and fidelity. On Poisson $N=8$ the runs within 10^7 executions end at a median cost of 0.052 and a median infidelity of 0.865. The cost bounds the infidelity only as $C\geq(1-F)/\kappa^2$ [18], so a small cost does not certify a state near the solution when κ is large. The entries of Table XIII are for one plain implementation: a global cost, no grouping of commuting terms, step sizes of the stochastic optimiser fixed after a small trial on the Poisson $N=4$ system, and ten repetitions. They are not a bound for variational solvers, and the number of pairs depends on the path of the optimiser. The comparison is also between resources of different kinds: a state preparation of a learning stage is followed by controlled evolution of up to t_{cap}, which the variational circuits do not contain.

Forecasting models. For the series of Sec. VIII we add

three methods that produce forecasts without solving the ridge system by phase estimation (Table XIV). The first is a quantum reservoir of four qubits [19]: at every step the first qubit is reset to $\sqrt{1-s}|0\rangle+\sqrt{s}|1\rangle$ with the input s scaled to $[0,1]$, the register evolves for unit time under $H=\sum_{i<j}J_{ij}X_iX_j+\sum_i Z_i$ with J_{ij} uniform in $[-1,1]$, and $\langle Z_i\rangle$ is read at four equally spaced times, which gives sixteen features for the ridge readout of Sec. VIII. Each measured step and time is one circuit, $12\,000$ in all. The second is quantum circuit learning [20] with four qubits: three layers each encode the last four inputs by $R_Y(\arcsin x)R_Z(\arccos x^2)$ and apply a CNOT chain and trainable rotations R_YR_Z; the forecast is $a\langle Z_1\rangle+c$, and the 26 parameters minimise the squared error on the training steps, by L-BFGS with exact expectation values (300 iterations) or by stochastic approximation on minibatches of 100 samples with one shot each.

Decoded quantum interferometry. The third method, decoded quantum interferometry (DQI) [21], is an optimisation algorithm and not a linear solver. We apply it to optimal polynomial intersection, the problem on which it is analysed: a polynomial of degree at most four over $\mathbb{F}_{31}$ maps the level of the most recent input on which the target depends (30 levels) to the level of the target (31 levels), and the constraint set of an input level holds the r output levels around the median training target of that level, with $r=5$ and $r=15$. The code dual to the instance is a Reed–Solomon code of distance six, so the DQI polynomial has degree two. We sample from the exact output distribution of the ideal algorithm with perfect decoding, obtained by enumerating all 31^5 polynomials; its norm and its expected number of satisfied constraints agree with Lemmas 9.1 and 9.2 of [21] to 10^{-9} and 10^{-6}.

Forecasts. With exact expectation values the quantum reservoir gives NMSE 0.21, 0.23 and 0.24, and the trained circuit 0.14, 0.097 and 0.0084. The circuit is better than the regularised ridge readout on the Mackey–Glass and Lorenz series; it is not bound by $\kappa\le 100$, and its exact training evaluates 9×10^6 to 3×10^7 circuits. Within 10^5 executions, the number of state preparations of Projective HHL in Table XII without dephasing, the reservoir has eight shots per circuit and gives 0.64, 0.56 and 0.43, the circuit does not train (0.95 to 1.00), and the readout of the variational solver gives 0.97, 0.69 and 0.44. With 10^4 executions the circuits of the reservoir cannot each be run once.

One DQI sample satisfies 30% ($r=5$) and 64% ($r=15$) of the constraints, against 16% and 48% for a random polynomial, and the best of 10^5 samples satisfies in the median as many as the optimum found by exhaustive search, 63% and 97% to 100%. The forecasts of these polynomials have NMSE between 0.66 and 1.6, and 0.27 on the Lorenz series with $r=15$, the level of repeating the last known value. The limit is the formulation and not the sampling: a polynomial over a finite field has no notion of nearness, so a level whose constraint is not met receives an arbitrary value.

Scope of these baselines. The reservoir couplings, the depth of the trained circuit, the field size and the degree were fixed beforehand and not tuned, each method is one plain implementation with ten runs, and no dephasing is applied to them. With exact expectation values the variational solver reproduces the ridge readout and the trained circuit is better than it on two series, at 10^6 to 3×10^7 evaluations of a circuit or of a circuit and a term. Within 10^5 executions none of the four methods comes near the ridge readout, and only the quantum reservoir on NARMA2 and Mackey–Glass and DQI with $r=15$ on Mackey–Glass are below the error of repeating the last known value.

X. Discussion

Five conclusions follow, each for the four small systems, the noise model and the time budgets studied.

First, for a non-eigenstate input the quantity that decides the learning stage is the detection of weak components. Mixture estimators need $S\gtrsim w_j^{-2}$ shots, a clock register needs $S\gtrsim w_j^{-1}$ down to a floor of a few per cent, and because a component that is missed at a small eigenvalue, or merged with one, dominates the error of the solution, this decides the fidelity; the counting model of (2) summarises the two labelled cases and is an analogy for the mixture case.

Second, the floor is not intrinsic to labelling. Keeping the copy and labelling it by its own posterior, as Recycled sequential HHL does, replaces the register tail by a stopping threshold, removes the register and the QFT from the learning stage, frees the resolution from a register size and the detection from knowledge of the number of components. Within the same time budget, the median time of the register-labelled runs, the median solution infidelity is then lower than with register labels on every system and noise setting, by factors from 2 to 3×10^3 when both estimators refine on every sample and from 1.5 to 2×10^3 when both stop at a precision target. The advantage persists under both allocation policies; the comparison is between whole estimators and does not isolate the label. The price is a system register that must outlive a whole copy, a label error that is tuned through the stopping threshold and measured, not bounded (a matched noise model is necessary for the working model and not sufficient), and a sample that costs two to sixteen register labels.

Third, for a sparse spectral measure the learned atoms answer every $\langle b|f(A)|b\rangle$ question, and through polarisation linear functionals, by classical post-processing. Neither labelled estimator is uniformly better for the compliance and the energy, and more register shots do not systematically lower the projective error; at the larger observable budget the sequential estimator is the more accurate for linear functionals in all eight settings of the four small systems, most clearly where eigenvalues share a register bin; on the 30-bus power-flow system the projective estimator is the more accurate for the line flow in four of five settings.

Fourth, the learned spectrum does not by itself repair a hybrid solver for the remaining observables: with a uniform clock the loss is dominated by leakage, which a tapered clock removes when the register resolves the spectrum and there is no dephasing, and the spectrum is otherwise most valuable for choosing t_0 or, when it is fully captured and its nodes are accurate enough, for spectrally corrected inversion.

Fifth, under dephasing the coherent clock limits the solver for every estimator and clock state tested, while the learning

stage needs no coherent register; its own requirement is the lifetime of the system register over one copy.

The power-flow test of Sec. VII is consistent with the first two conclusions on spectra that were not constructed for the purpose and marks their limit: a component with 0.2% of the weight at the smallest eigenvalue limits the fidelity of the 9-bus solution and, without dephasing, is assigned a learned value in all ten runs of Recycled sequential HHL, in at most 31 of 100 runs of a mixture estimator and in no run of the QPE histograms and of Projective HHL, whereas with dephasing, or with twenty weighted components at the budget of Table XI, no learning stage resolves the spectrum.

The forecasting test of Sec. VIII adds a problem in which the weak components are weaker still. Within ten times the budget of the other tests Recycled sequential HHL at its finest resolution reproduces the forecast of the exact ridge readout in all ten runs on the three series, with and without dephasing; Projective HHL does so in none to nine runs and gains nothing from ten times the shots on NARMA2. The forecast error is far less sensitive than the fidelity, the resolution was compared after the runs, and the baselines without phase estimation of Sec. IX reach the ridge readout only with exact expectation values.

Several limitations should be stated.

Noise model. The noise model is a calibrated contrast on the control qubit and a fixed depolarising rate; gate-level noise on the system register, independent noise in the inverse QPE and miscalibration of the coarse stage were not modelled. For the sequential estimator this matters more than for the others. Equation (5) holds only while the system register evolves under A alone, so Trotter error and system noise that moves population between eigenspaces are outside it; a copy must survive for 9×10^2 to 8×10^4 time units, depending on the allocation and the resolution; and the only system-noise test is the redrawing model of Sec. IV.

Stopping threshold and resolution. The stopping threshold is not an error bound: a matched contrast is necessary for the working model but not sufficient for coverage with the tempered likelihood and the plug-in prior, the measured error fractions are simulation results for one system, and with an over-estimated contrast they exceed the threshold by up to a factor of seventeen. The resolution is uniform, so that one close pair makes every copy expensive, and σ_{id}, like n_c, is set from the smallest gap, which is assumed known.

Prior art. The building blocks are known: the record identity of repeated ancilla measurements [12] and its back-action [17], iterative phase estimation inside a hybrid solver [10], rate-of-gain experiment design [22], [23] and window states for the clock [1], [27]–[29]. The contribution is their combination into a learning stage and its numerical study on small instances; a guarantee on the label error, optimality of the cost and a matched-time advantage beyond the tested settings are not established.

Statistics and baselines. The comparisons between estimators use independent repetitions, not common random numbers, and the five to ten repetitions of the matched-time comparisons do not resolve rare failures: in Table IV the projective result on Poisson $N = 8$ at $T_2 = 300$ has two groups (seven of ten runs near 0.076, three below 4×10^{-4}), "worst" means the worst of the runs made, and only Table I, with 100 repetitions, estimates the frequency of such failures; its reliability does not carry over to the other tables. The mixture baselines are plain implementations: the MM-QCELS-type estimator misses components that a tuned estimator would find (it fails at $w_j\sqrt{S} \approx 10$ where the QMEGS-type estimator succeeds), so the measured gap between mixture and labelled estimators is an upper bound on the gap implied by the detection law.

Test set and repetitions. The test set is four systems with $N \le 8$, three power-flow systems and three ridge systems of one reservoir, with one input vector each, and the repetition counts (100 for Table I, three for Fig. 1, ten for Tables IV, VIII, IX, XII, XIII and XIV, five at the full Toeplitz resolution, six for Table VII and the spectral columns of Table VI, four for its Krylov and Fourier columns, forty learning runs for Table X; in Sec. VII 100 for the mixture and QPE rows of Table XI, ten for its two labelled estimators and five at the finest resolution, six for the observables, five for the scaling runs and four for the model-free estimators) are, except for Table I and the upper rows of Table XI, small for outcomes with two groups; medians are reported with interquartile ranges where they are wide.

Idealisations. The weight estimation of the projective estimator assumes the clock kernel is known, which requires the learned eigenvalues and the dephasing model; a kernel mismatch biases the weights directly. The clock analysis uses the ideal circuit with exact Hamiltonian simulation and does not include block-encoding subnormalisation or QSVT phase-factor synthesis, and the interval reference polynomial is a bounded linear-program solution rather than a Remez construction.

Scaling. For generic inputs the number of learned components grows with the dimension, and the learning cost then removes any asymptotic advantage; the analysis is intended for structured problems with few weighted components, the regime in which hybrid solvers are proposed, and the Krylov baseline used here is the plain thresholded variant rather than the stabilised constructions of [25], [26]. Instance-dependent complexities of this kind have been formalised for fully coherent solvers [47]; the present work gives the corresponding statement for the learning stage and for the observables it already determines.

Appendix A
Proofs and algorithms

Weight information. For a Bernoulli outcome with probability p the Fisher information on a parameter ϑ is $(\partial_\vartheta p)^2/[p(1-p)]$. With $\vartheta = w_j$ (the other weights renormalised), $\partial_{w_j} p = \frac{1}{2}c(\cos\varphi_j - \bar{g})$ with $\bar{g}$ the background, bounded in magnitude by c, so $I_{w_j w_j} \le c^2/[p(1-p)] \le 4c^2/(1-c^2)$ for $c < 1$ (two components with opposite phases and equal weights reach $4c^2$); for a labelled shot the count is binomial and $I_{w_j w_j} = 1/[w_j(1-w_j)]$. In terms of $a_j = \ln w_j$ the two informations are $O(w_j^2)$ and $O(w_j)$, which is the detection law.

Rank-one eigenvalue information. With $\vartheta = \lambda_j$, $\partial_j p = -\frac{1}{2} c w_j t \sin\varphi_j$, which gives the stated outer product of rank one.

Mixture scaling. Averaging the outer product over $\theta \in [0, 2\pi)$ gives $\overline{\sin\varphi_j \sin\varphi_k} = \frac{1}{2}\cos((\lambda_j - \lambda_k)t)$; a further average over a time window wider than the inverse gaps removes the off-diagonal terms by the Riemann–Lebesgue lemma. The lower bound $\bar{I} \succeq \frac{c^2}{2} t^2 W^2 (1-\eta)$ follows from $4p(1-p) \le 1$ with η the finite-window Gram error of the cosines; the upper bound uses $4p(1-p) \ge 1 - c^2$, valid for $c < 1$ (at $c = 1$ a single dominant component makes $\overline{[4p(1-p)]^{-1}}$ diverge while the information stays finite, so the constant must be stated with the contrast). The Cramér–Rao inequality then gives $\mathrm{Var}(\hat{\lambda}_j) \ge [(S\bar{I})^{-1}]_{jj}$; the rates quoted in the text are these local bounds for resolved components, not achievability statements.

Error propagation. With $d = \hat{x} - x$, $1 - F = (\|x\|^2\|d\|^2 - |\langle x|d\rangle|^2)/(\|x\|^2\|\hat{x}\|^2) \le \|d\|^2/\|\hat{x}\|^2$. Expanding $1/\hat{\lambda}_j = (1 - e_j + O(e_j^2))/\lambda_j$ and normalising gives (3); the subtracted term is $|\langle x|d\rangle|^2/\|x\|^4$, the squared mean of e_j under the weights q_j.

Tracking rule and unwrapping. The tracking cost $c/(w_j T^2 S)$ and the freezing cost $w_j \Delta_j^2$ share the factor λ_j^{-4}, so the comparison does not involve λ_j. The likelihood is periodic in λ_j with period $2\pi/T$, so staying in the correct basin across rounds requires $z_\delta \sigma_j T_{r+1} < \pi/2$; inserting the mixture rate $\sigma_j \asymp 1/(w_j T\sqrt{S})$ with per-round shots S_r bounds the growth factor by $f < c_u w_j \sqrt{S_r}$, $c_u = \pi/(2z_\delta)$. The projective estimator used $t = 0.6/\sigma_j$ and a tempering $\epsilon = 0.05$ of the likelihood.

Labelled scaling. If the coarse measurement left the register exactly in $|u_j\rangle$ with probability w_j, (1) would reduce to a single cosine; its phase-averaged information is $\overline{c^2 t^2 \sin^2\varphi/(1 - c^2\cos^2\varphi)} = t^2(1 - \sqrt{1-c^2})$, multiplied by w_j. With a finite clock the register after outcome y is $\propto \sum_k \beta_k a_{ky} |u_k\rangle$ and the single cosine is replaced by a mixture with purity π_j; the information on λ_j is reduced by π_j^2 and the background of the other components enters the likelihood, which the estimator absorbs by tempering.

Equation (5). With the ancilla in $|+\rangle$, a controlled e^{-iAt}, the phase θ and a measurement in the X basis, outcome d applies $K_d = \frac{1}{2}[1 + (-1)^d e^{i(At+\theta)}]$ to the system register; a memoryless noise channel on the ancilla before or after the interaction replaces K_d by a set $\{K_{d,a}\}$ with $K_{d,a} = f_{d,a}(A)$. All these operators are functions of A, hence commute and are diagonal in its eigenbasis, so $P(d_1 \dots d_n) = \langle b| \prod_i \sum_a K_{d_i,a}^\dagger K_{d_i,a} |b\rangle = \sum_j w_j \prod_i \sum_a |f_{d_i,a}(\lambda_j)|^2$, and $\sum_a |f_{0,a}(\lambda)|^2 = \frac{1}{2} + \frac{1}{2} c\cos(\lambda t + \theta)$ for dephasing and depolarisation of the ancilla. Settings that depend on earlier outcomes change nothing, by the chain rule; for degenerate eigenvalues w_j is the weight of the eigenspace. Without noise the identity is Eq. (10) of [12]. It fails when an operation is not a function of A: Trotter error, or system noise that transfers population between eigenspaces. A state-vector propagation of all 32 records of five adaptive measurements on a three-level system agrees with (5) to 10^{-12}, without noise and with a random phase kick on the ancilla.

Equation (6). By the law of total variance the expected posterior variance after a measurement is the prior variance minus $\sum_d P_d(\mu_d - \mu)^2$, with μ_d the posterior mean after outcome d. Since $P_0\mu_0 = \mathbb{E}[\lambda\, p(0|\lambda)]$, $\mu_0 - \mu = \frac{1}{2} c\,\mathrm{Re}(e^{i\theta} q_t)/P_0$ and $\mu_1 - \mu = -\frac{1}{2} c\,\mathrm{Re}(e^{i\theta} q_t)/P_1$, which gives (6); it holds for the likelihood used in the update, so c is 0.98 times the calibrated contrast. For a Gaussian posterior $|q_t| = \sigma^2 t e^{-\sigma^2 t^2/2}$ and q_t is in quadrature with z_t, so $P_0 = \frac{1}{2}$ at $\theta = -\arg q_t$ and $R/t \propto c(t)^2 t e^{-\sigma^2 t^2}$, maximal at $\sigma t = 1/\sqrt{2}$ for constant c and at $\sigma^2 t^2 + t/T_2 = \frac{1}{2}$ for $c \propto e^{-t/T_2}$. For a general posterior $\theta = -\arg q_t$ maximises only the numerator, which is why the phase is taken from a grid.

Stopping rule. Let the eigenvalue of a copy have prior π and let the update be exact. If every stopped record D satisfies $P_\pi(|\lambda - \mu(D)| > r \,|\, D) \le p_{\mathrm{out}}$, then averaging over stopped records gives $P_\pi(|\lambda - \mu| > r \text{ and stopped}) \le p_{\mathrm{out}} P_\pi(\text{stopped})$: the prior-averaged error rate among stopped copies is bounded, for adaptive settings and stopping times alike. This is not coverage for a fixed eigenvalue, and it does not apply as stated to the implementation, whose likelihood is tempered and whose prior for later copies is the plug-in mixture; Tables II and III are the evidence for the implemented rule.

Sequential estimator. Grid spacing $\sigma_{\mathrm{id}}/5$ on $(0, 1.25)$; likelihood tempered by 0.02 in the identification and 0.05 in the pooled refinement; batches of at most 500 copies and at most a quarter of the remaining time, the components and the prior being updated after each batch; a pooled posterior is created at σ_{id} around the cluster mean and re-gridded as it narrows. After re-clustering a cluster keeps a pooled posterior only if the two consist of the same copies (at least 90% of the previously labelled members of the new cluster come from the old one and at least 90% of the old one's members are in the new one); a strong cluster that absorbs a weak one thus keeps its own posterior, a merger of comparable clusters or a split keeps none unless one part holds 90% of the members, and all other clusters start afresh. The rule is tested on constructed mergers and splits in the code; it compares memberships and does not check the location of the inherited posterior, and with a resolution coarser than the default the peeling step can leave the tail of a weak component as a cluster of its own. The weights are the label frequencies among converged copies. The time of a copy that cannot finish within the budget is charged and the copy discarded. Every sequential run of Tables IV and V records its identification and refinement times, the fraction of converged and of unlabelled copies, the counts, the cluster centres and the number of inherited and newly created posteriors.

Compliance bound. For mixture data with known eigenvalues the Fisher matrix of the $m - 1$ free weights is $I = \sum_s \nabla p_s \nabla p_s^{\mathsf{T}} / [p_s(1 - p_s)]$ over the settings s, with $\partial p/\partial w_i = \frac{1}{2} c(\cos\varphi_i - \cos\varphi_m)$, and the bound on $\mathrm{Var}\,\hat{C}$ is $g^{\mathsf{T}} I^{-1} g$ with $g_i = 1/\lambda_i - 1/\lambda_m$; a bound on one weight alone does not bound C, because directions on the simplex exist that change w_1 and leave C unchanged. The design-optimised value minimises $g^{\mathsf{T}} I^{-1} g$ over the distribution of settings on a grid ($t \le 2000$, sixteen phases) by multiplicative updates; it is

a local bound for that grid, not an optimum over all adaptive protocols. All these are bounds on the standard deviation of locally unbiased estimators. With unknown eigenvalues the bound for the random-time designs is the same to two digits for windows of 300 and longer.

Equation (9). $|p(\lambda_j) - \lambda_{\min}/\lambda_j| \le |r_j| + |p(\mu_j) - p(\lambda_j)| + \lambda_{\min}|\delta_j|/(\lambda_j\mu_j)$, with $|p(\mu_j) - p(\lambda_j)| \le \max_{[-1,1]}|p'|\,|\delta_j| \le d^2 \max_{[-1,1]}|p|\,|\delta_j|$ by Markov's inequality for $\lambda_j, \mu_j \in [-1, 1]$; insert into the error-propagation bound for $\hat{y} = p(A)|b\rangle$ against $y = \lambda_{\min}A^{-1}|b\rangle$. The bound is loose: for relative node perturbations between 10^{-4} and 10^{-2} its right-hand side exceeds one on the ill-conditioned systems, so it certifies only the small-perturbation regime.

Equation (7). Mean-value theorem and the triangle inequality; the polarisation identity follows from $|\beta_j + \nu_j|^2 - |\beta_j|^2 - |\nu_j|^2 = 2\,\mathrm{Re}(\nu_j^*\beta_j)$ and $|\beta_j + i\nu_j|^2 - |\beta_j|^2 - |\nu_j|^2 = 2\,\mathrm{Im}(\nu_j^*\beta_j)$. The Fourier estimator follows from $\int dz\, z\, e^{-z^2/2}e^{-iyz\lambda}/\sqrt{2\pi} = -iy\lambda e^{-y^2\lambda^2/2}$ and $\int_0^\infty dy\, y\lambda e^{-y^2\lambda^2/2} = 1/\lambda$.

Detection runs. On Poisson $N = 8$ without dephasing, three runs each: at 10^4 shots the MM-QCELS-type estimator detects (learned eigenvalue within 0.01) the components $\{0.258, 0.426, 0.773, 1.0\}$ or fewer and never $\{0.031, 0.121, 0.605\}$ (weights $0.032, 0.006, 0.001$); the QMEGS-type estimator detects the five largest-weight components and never the same three; the labelled estimator detects all but $w = 0.001$ in every run. At 10^5 shots the QMEGS-type estimator finds $w = 0.032$ in all three runs ($F = 0.92$–0.99), the MM-QCELS-type in one of three, and $w = 0.006$ is found by none of the mixture estimators and by the labelled estimator in two of three. The per-point noise of the mixture time series is $1/\sqrt{S_{\mathrm{pt}}} = 0.12$ at 10^4 and 0.038 at 10^5 shots; the labelled count of the $w = 0.032$ component is 320 ± 18 at 10^4 shots. Fig. 2 repeats the measurement for planted weights and adds the sequential estimator.

Clock fidelity. For component j with clock amplitudes a_{jy} and rotation c_y ($c_y = 1/\hat{\lambda}_y$ on rotated bins, 0 elsewhere), the exact inverse of the forward QPE returns $\bar{c}_j = \sum_y |a_{jy}|^2 c_y$ to the clock state $|0\rangle$ and leaves the rest in orthogonal clock states, so the joint fidelity with $|x\rangle|0\rangle$ is $F = |\sum_j \bar{x}_j^*\beta_j\bar{c}_j|^2 / \sum_j |\beta_j|^2 \sum_y |a_{jy}|^2 c_y^2$, where $\bar{x}$ is the normalised exact solution in the eigenbasis; this is a lower bound on the fidelity of the solution register with the clock traced out. Under dephasing each sample s draws one set of phase kicks on the clock qubits, applied to all eigencomponents, and gives the unnormalised overlap $N_s = |\sum_j \bar{x}_j^*\beta_j\bar{c}_j^{(s)}|^2$ and the success weight $D_s = \sum_j |\beta_j|^2 \sum_y |a_{jy}^{(s)}|^2 c_y^2$; the heralded-ensemble fidelity is $\overline{N_s}/\overline{D_s}$ over 256 samples, with the standard error estimated from eight blocks. For a window state with amplitudes g_k the clock amplitudes are $a_{jy} = M^{-1/2}\sum_k g_k e^{ik(\lambda_j t_0 - 2\pi y/M)}$ with the same phase kicks, and the rotation table assigns to each bin the nearest learned eigenvalue within the main lobe.

Spectral correction. With $p_0 = \sum_\ell c_\ell T_{2\ell+1}$ and residuals $r_k = \lambda_{\min} - \lambda_k p_0(\lambda_k)$ at the K learned eigenvalues, solve $G\alpha = r$ with $G_{ik} = \lambda_i\lambda_k \sum_\ell T_{2\ell+1}(\lambda_i)T_{2\ell+1}(\lambda_k)$ and set $\Delta c_\ell = \sum_k \alpha_k\lambda_k T_{2\ell+1}(\lambda_k)$ [30]; the base polynomial here is the bounded odd interval solution of the linear program.

Code and data availability

The stage scripts, the result files and the scripts that regenerate every table of this paper and check a hand-written list of the numbers quoted in the text against the result files, together with SHA-256 manifests of code and results, are available at https://github.com/deeptell-inc/hhlx-reproduce.

References

[1] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Phys. Rev. Lett.*, vol. 103, p. 150502, 2009.

[2] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, "Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics," *Proc. 51st ACM STOC*, pp. 193–204, 2019.

[3] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, "Grand unification of quantum algorithms," *PRX Quantum*, vol. 2, p. 040203, 2021.

[4] M. E. S. Morales, L. Pira, P. Schleich, K. Koor, P. C. S. Costa, D. An, A. Aspuru-Guzik, L. Lin, P. Rebentrost, and D. W. Berry, "Quantum linear system solvers: A survey of algorithms and applications," *Rev. Mod. Phys.*, vol. 98, p. 025005, 2026.

[5] P. Pareek, A. Jayakumar, C. Coffrin, and S. Misra, "Limitations of fault-tolerant quantum linear system solvers for quantum power flow," *arXiv:2402.08617*, 2024.

[6] M. A. Marfany, A. Sakhnenko, and J. M. Lorenz, "Identifying bottlenecks of NISQ-friendly HHL algorithms," *arXiv:2406.06288*, 2024.

[7] Y. Lee, J. Joo, and S. Lee, "Hybrid quantum linear equation algorithm and its experimental test on IBM quantum experience," *Sci. Rep.*, vol. 9, p. 4778, 2019.

[8] R. Yalovetzky, P. Minssen, D. Herman, and M. Pistoia, "Solving linear systems on quantum hardware with hybrid HHL++," *Sci. Rep.*, vol. 14, p. 20610, 2024.

[9] J. Morgan, E. Ghysels, and H. Mohammadbagherpoor, "An enhanced hybrid HHL algorithm," *arXiv:2404.10103*, 2024.

[10] F. Gao, G. Wu, M. Yang, W. Cui, and F. Shuang, "A hybrid algorithm to solve linear systems of equations with limited qubit resources," *Quantum Inf. Process.*, vol. 21, p. 111, 2022.

[11] A. Dutkiewicz, B. M. Terhal, and T. E. O'Brien, "Heisenberg-limited quantum phase estimation of multiple eigenvalues with few control qubits," *Quantum*, vol. 6, p. 830, 2022.

[12] T. E. O'Brien, B. Tarasinski, and B. M. Terhal, "Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments," *New J. Phys.*, vol. 21, p. 023022, 2019.

[13] Z. Ding and L. Lin, "Even shorter quantum circuit for phase estimation on early fault-tolerant quantum computers with applications to ground-state energy estimation," *PRX Quantum*, vol. 4, p. 020331, 2023.

[14] ——, "Simultaneous estimation of multiple eigenvalues with short-depth quantum circuit on early fault-tolerant quantum computers," *Quantum*, vol. 7, p. 1136, 2023.

[15] H. Li, H. Ni, and L. Ying, "On adaptive low-depth quantum algorithms for robust multiple-phase estimation," *Quantum*, 2023, arXiv:2303.08099.

[16] Z. Ding, H. Li, L. Lin, and H. Ni, "Quantum multiple eigenvalue Gaussian filtered search: an efficient and versatile quantum phase estimation method," *Quantum*, 2024, arXiv:2402.01013.

[17] Y.-D. Jin, S.-Y. Zhang, and W.-L. Ma, "Channel-based framework for phase estimation of multiple eigenvalues," *Phys. Rev. A*, vol. 111, p. 022406, 2025.

[18] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subaşı, L. Cincio, and P. J. Coles, "Variational quantum linear solver," *Quantum*, vol. 7, p. 1188, 2023.

[19] K. Fujii and K. Nakajima, "Harnessing disordered-ensemble quantum dynamics for machine learning," *Phys. Rev. Appl.*, vol. 8, p. 024030, 2017.

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

[21] S. P. Jordan, N. Shutty, M. Wootters, A. Zalcman, A. Schmidhuber, R. King, S. V. Isakov, T. Khattar, and R. Babbush, "Optimization by decoded quantum interferometry," *Nature*, vol. 646, pp. 831–836, 2025.

[22] B. de Neeve, A. V. Lebedev, V. Negnevitsky, and J. P. Home, "Time-adaptive phase estimation," *Phys. Rev. Research*, vol. 7, p. 023070, 2025.

[23] R. D. McMichael, S. Dushenko, and S. M. Blakley, "Sequential Bayesian experiment design for adaptive Ramsey sequence measurements," *J. Appl. Phys.*, vol. 130, p. 144401, 2021.

[24] L. Lin and Y. Tong, "Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers," *PRX Quantum*, vol. 3, p. 010318, 2022.

[25] S. Yu and Y. Zhou, "Quantum Krylov linear solving beyond global conditioning via spectral compression and directional stability," *arXiv:2610.06620*, 2026.

[26] B. Park, K. Jeong, and D. Ahn, "Sample-based quantum linear solver and bounds on Krylov-state sampling," *arXiv:2610.23138*, 2026.

[27] G. Rendon, T. Izubuchi, and Y. Kikuchi, "Effects of cosine tapering window on quantum phase estimation," *Phys. Rev. D*, vol. 106, p. 034503, 2022.

[28] D. Patel, S. J. S. Tan, Y. Subaşı, and A. T. Sornborger, "Optimal coherent quantum phase estimation via tapering," *PRX Quantum*, vol. 7, p. 020302, 2026.

[29] K. Kristjuhan and D. W. Berry, "Quantum phase estimation with optimal confidence interval using three control qubits," *Quantum*, vol. 10, p. 2193, 2026.

[30] K. Suresh, "Spectrally corrected polynomial approximation for quantum singular value transformation," *arXiv:2603.03998*, 2026.

[31] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, 2011.

[32] R. Eskandarpour, K. Ghosh, A. Khodaei, L. Zhang, A. Paaso, and S. Bahramirad, "Quantum computing solution of DC power flow," *arXiv:2010.02442*, 2020.

[33] R. Eskandarpour, K. Ghosh, A. Khodaei, and A. Paaso, "Experimental quantum computing to solve network DC power flow problem," *arXiv:2106.12032*, 2021.

[34] F. Gao, G. Wu, S. Guo, W. Dai, and F. Shuang, "Solving DC power flow problems using quantum and hybrid algorithms," *arXiv:2201.04848*, 2022.

[35] N. Wiebe, D. Braun, and S. Lloyd, "Quantum algorithm for data fitting," *Phys. Rev. Lett.*, vol. 109, p. 050505, 2012.

[36] H. Jaeger and H. Haas, "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, pp. 78–80, 2004.

[37] N. Wiebe and C. Granade, "Efficient Bayesian phase estimation," *Phys. Rev. Lett.*, vol. 117, p. 010503, 2016.

[38] A. Ambainis, A. Gilyén, and M. Kokainis, "Tight bounds for hybrid quantum-classical query algorithms," *arXiv:2610.06803*, 2026.

[39] S. Hariprakash, R. Van Beeumen, K. Klymko, and D. Camps, "The practicality of randomized quantum linear systems solvers," *arXiv:2510.13766*, 2025.

[40] E. Pelofske and S. Eidenbenz, "Effects of Trotter error, digitization error, and initial state overlap on tapered quantum phase estimation for minimum eigenvalue computation," *arXiv:2609.06249*, 2026.

[41] H. Apel, C. L. Cortes, J. Lemieux, and M. Steudtner, "Reducing quantum resources for observable estimation with window-assisted coherent QPE," *arXiv:2508.06677*, 2025.

[42] M. O. Gordeichuk and O. M. Kiselev, "Measurement-dependent reduction of open dynamics in the quantum phase estimation algorithm: the visible subspace and the damped Fejér kernel," *arXiv:2610.35332*, 2026.

[43] A. F. Atiya and A. G. Parlos, "New results on recurrent network training: unifying the algorithms and accelerating convergence," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 697–709, 2000.

[44] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, pp. 287–289, 1977.

[45] E. N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.*, vol. 20, no. 2, pp. 130–141, 1963.

[46] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Autom. Control*, vol. 37, no. 3, pp. 332–341, 1992.

[47] A. M. Dalzell, J. Li, and Y. Su, "Faster quantum linear system solver beyond the condition number," *arXiv:2607.07691*, 2026.