

Quantum state reconstruction from incomplete preparation records

Hikaru Wakaura

hikaruwakaura@gmail.com

QuantScape Inc

Research Article

Keywords: Quantum state reconstruction, Preparation records, Fidelity bounds, Quantum algorithms, Quantum memory, Resource estimation

Posted Date: October 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11430091/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.

Quantum state reconstruction from incomplete preparation records

Hikaru Wakaura

QIRI (Quantum Integrated Research Institute Inc.), 1-16-3, Akasaka,
Minato-ku, Tokyo, 107-0061, Japan.

Corresponding author(s). E-mail(s): h.wakaura@qiri.co.jp;

Abstract

Preparing additional quantum states while retaining a source requires information beyond an arbitrary unknown input. We formulate reconstruction from incomplete classical preparation records as an identity channel on the source and two output preparations. For a short interval of QPE phase states, we derive the exact record-only minimax fidelity and an attaining endpoint superposition, closing the gap between center replay and an endpoint upper bound. An explicit one-ancilla circuit realizes the optimum with a known success probability, while deterministic synthesis provides a finite-circuit alternative. Shor and randomized-simulation examples expose different obstructions from discrete ambiguity and missing random choices. In one randomized-simulation setting, optimal conditional fidelity averaged over records was 0.925, although robust fidelity above five sixths was impossible for eight of ten records. Session calibration and resource-controlled comparisons distinguish average performance from worst-case guarantees. Hardware diagnostics test operational phase models, source waiting and preparation costs. They do not certify intrinsic gate angles, reconstruction fidelity or physical source preservation. The analysis identifies information and operational conditions without changing universal cloning bounds.

Keywords: Quantum state reconstruction, Preparation records, Fidelity bounds, Quantum algorithms, Quantum memory, Resource estimation

1 Introduction

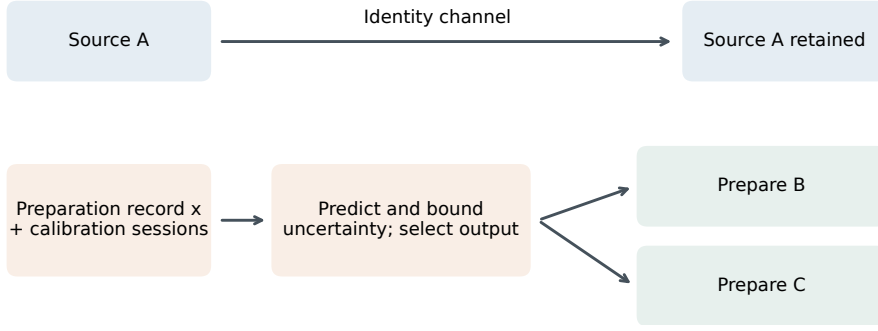
A quantum computation may produce an intermediate state that would be useful in more than one subsequent calculation. Retaining that physical state and preparing

additional instances is a different task from repeatedly measuring it. Without extra information, universal quantum cloning has fixed fidelity limits. For one arbitrary pure input of dimension d , the optimal deterministic symmetric $1 \rightarrow 2$ cloner has single-output fidelity $(d+3)/[2(d+1)]$ [1, 2]. In a preparation device, however, the initial state, circuit description, control settings and random seed may be partially recorded. Such records change the information available for preparing a new state. They must be counted as a resource rather than absorbed into the definition of an unknown input.

For a finite pure-state family containing no orthogonal pairs, exact deterministic cloning with supplementary information is possible only if that information alone can generate the extra copy [3, Theorem 1]. The theorem has an explicit exception when orthogonal pairs are allowed and supplies no general approximate-fidelity bound. State-preserving operations on a known family likewise constrain the information that can be extracted [4]. The reconstruction decision itself also has a substantial history. For a known conditional ensemble of pure states, preparing a maximal-eigenvalue eigenstate of its average density matrix maximizes average overlap [5]. That rule does not establish a worst-case guarantee, solve the squared Uhlmann-fidelity problem for mixed targets, or make the output efficiently preparable. General state-synthesis constructions emphasize the separate cost of preparing a specified vector [6, 7]. These distinctions become decisive when incomplete records are processed by a learned predictor. Variational quantum cloning already optimizes circuit parameters and structure for specified state families, using the quantum input and distinguishing local from global cloning objectives [8, Secs. II and IV, Appendix B.1]. Here the predictor uses external preparation records and never accesses the retained source; that different resource model requires its own baselines and does not establish superiority to a quantum cloner.

Encrypted cloning provides another distinct resource model: correlated quantum signal and key registers permit perfect recovery of one chosen output, while the encoded signals are not simultaneously available as independent unencrypted copies [9]. In contrast, our outputs are prepared from a correlated classical record, and the ideal source channel is the identity. These operational differences preclude ranking the methods by their quoted fidelities alone.

We study the conditions under which incomplete records support two simultaneous additional outputs, with no operation on the source. Our analysis connects record uncertainty to fidelity and distinguishes an achieved lower bound from an impossibility upper bound. We evaluate session-calibrated angle uncertainty and checkpoints of quantum phase estimation (QPE), Shor order finding and qDRIFT Hamiltonian simulation [10–12]. A hardware pilot examines chronological coverage of measured fringe phases, source waiting and two-output preparation cost. A follow-up experiment tests a SPAM-nuisance phase model, balances submitted waiting-arm order and records classical and cloud costs. The central theoretical result is an exact record-only minimax solution for a short QPE phase interval, including an attaining state and an explicit preparation circuit. Algorithm-specific bounds, reproducible numerical checks and hardware diagnostics place that result within broader information and resource conditions. We do not claim a new posterior reconstruction rule, a new conformal



Two additional outputs, independent preparation randomness, success probability 1

Fig. 1 Operational model. The preparation record and separately calibrated uncertainty determine the output preparation. No measurement or gate is applied to source A. Independent preparations produce B and C simultaneously. A record may be correlated with the target because it comes from its preparation device; the blank output registers are not an additional copy of that target.

prediction method or a violation of no-cloning. Throughout, $5/6$ is a chosen single-output target. It coincides with the universal qubit cloning optimum, but is not the universal bound for arbitrary qubit number or for this additional-information task.

2 Results

2.1 A source-preserving channel and three distinct objectives

Let x denote the classical information available both during training and at reconstruction time. The target state ρ_A is compatible with x , but need not be uniquely specified by it. Our operational channel is

$$\Gamma_x(\rho_A) = \rho_A \otimes \sigma_{x,B} \otimes \sigma_{x,C}. \quad (1)$$

For each fixed x , this is a completely positive, trace-preserving map: it tensors the input with two states prepared from independent ancillas and independent classical randomness. Both outputs exist simultaneously, and the success probability is one. The original register is untouched; the same statement holds when it is entangled with a reference. An uninformative record still permits fixed output preparation, but its best worst-case overlap with an arbitrary pure target is only $1/d$ (Supplement S1). Source preservation in Eq. (1) is an ideal-channel property, not a certification of physical memory preservation during a device run. In particular, waiting, crosstalk or leakage may disturb a hardware source even when its intended circuit is the identity.

We use squared Uhlmann fidelity [13, Sec. III, Eq. (4)],

$$F(\rho, \sigma) = \left(\text{Tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right)^2. \quad (2)$$

For a pure target, this is $\langle \psi | \sigma | \psi \rangle$. Conditional average fidelity averages over a specified distribution $p(\psi | x)$. Robust fidelity takes the infimum over the compatible set

S_x . A finite test-set minimum estimates neither of these exactly. The joint-output target is $\rho \otimes \rho$; Eq. (1) gives $F_{BC} = F_B F_C = F_B^2$ for each fixed target. This multiplicativity follows from the product preparation and is not a property of general cloners [1, 14]. Averaging gives $\mathbb{E}[F_B^2]$, not $(\mathbb{E}[F_B])^2$.

For pure targets define $R_x = \mathbb{E}[|\psi\rangle\langle\psi| \mid x]$. The established average optimum is

$$\max_{\sigma} \text{Tr}(R_x \sigma) = \lambda_{\max}(R_x). \quad (3)$$

A corresponding eigenstate attains this value when unrestricted preparation is permitted [5]. By contrast, robust reconstruction is the optimization

$$t_x^* = \max_{\sigma \succeq 0, \text{Tr } \sigma = 1} \inf_{\psi \in S_x} \langle \psi | \sigma | \psi \rangle. \quad (4)$$

For a finite set this is a semidefinite program. Two compatible pure states with overlap magnitude s imply $t_x^* \leq (1 + s)/2$; also $t_x^* \leq \lambda_{\max}(R_x)$ for any supported distribution. These bounds remain valid with preparation restrictions, while attainment can fail under such restrictions. Equation (3) is neither a minimax theorem nor a joint-output optimum. Explicit mixed-target and joint-objective counterexamples are given in Supplement S10.2.

2.2 From calibrated angles to a fidelity guarantee

Consider a known initial pure state and a circuit with known gate axes, order and connectivity, but K uncertain Pauli rotation angles. Write $R_{P_j}(\theta_j) = \exp(-i\theta_j P_j/2)$ with $P_j^2 = I$. If prediction error lies in a circular interval of half-width r_j , replacing the rotations one at a time and using the Fubini–Study triangle inequality [13, Sec. III] gives

$$F_B = F_C \geq L = \cos^2 \left[\min \left\{ \frac{\pi}{2}, \frac{1}{2} \sum_{j=1}^K r_j \right\} \right]. \quad (5)$$

The proof does not require the rotations to commute. It discards cancellations and state-dependent sensitivity, and is therefore sufficient rather than generally necessary. Independent single-qubit R_Y rotations instead have exact box worst-case fidelity $\prod_j \cos^2(r_j/2)$. Perfectly anticorrelated consecutive rotations $R_Y(a)R_Y(-a)$ cancel. Thus the number or fraction of missing angles alone does not determine fidelity.

For a pure target and known terminal depolarization $\mathcal{D}_\epsilon(\sigma) = (1 - \epsilon)\sigma + \epsilon I/d$, the achieved lower bound is $(1 - \epsilon)L + \epsilon/d$. When the square-root argument is between zero and one, a sufficient condition for single-output target $f_\star$ is

$$\sum_j r_j < 2 \arccos \sqrt{\frac{f_\star - \epsilon/d}{1 - \epsilon}}. \quad (6)$$

For $n = 5$, $f_\star = 5/6$ and $\epsilon = 0.025$, a total half-width below 0.7844185327 radians is sufficient under this noise model. Here ϵ is a full-register terminal noise parameter, not a per-gate error probability.

We calibrated these intervals at the preparation-session level. A predictor and residual scales were fixed on training sessions. Each calibration score was the largest

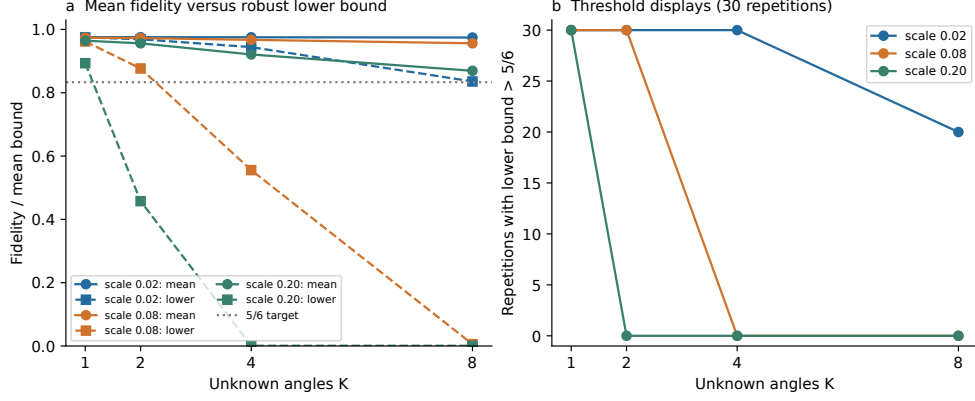


Fig. 2 Session-calibrated reconstruction for five qubits, pure targets and terminal depolarization 0.025. (a) Empirical mean fidelity and the mean interval-based lower bound across thirty repetitions, each with ten test records. Curves show synthetic residual scales, not calibrated interval widths. (b) Number of repetitions displaying a lower bound strictly above $5/6$. All panels use 199 calibration sessions. These counts do not estimate a distribution-free conditional success probability among displayed cases.

normalized circular residual over all ten records and all uncertain angles in its session, enlarged by an independent reference-error bound. A split-conformal order statistic then controls all angles in one new exchangeable session [15]. The guarantee is marginal over calibration and test sessions. It is not a conditional guarantee restricted to records displaying high fidelity, and it does not cover all future sessions simultaneously.

The synthetic experiment used $n \in \{1, 2, 5\}$, $K \in \{1, 2, 4, 8\}$ of sixteen rotations, three residual scales and thirty repetitions per setting. Each repetition contained 120 training records, 1,990 calibration records in 199 sessions and ten records in one new test session. A nested 39-session calibration subset enabled a paired comparison. Across 1,080 cases, the 199-session intervals covered every test angle in 1,037 sessions (96.02%); the 39-session intervals covered 1,033 (95.65%). These are descriptive implementation checks across heterogeneous settings, not estimated conditional coverage for each subgroup.

At $n = 5$, scale 0.02 radians and $K = 8$, the mean output fidelity was 0.974459, while the mean lower bound was 0.835578. Twenty of thirty repetitions displayed a bound above $5/6$, compared with fourteen using 39 calibration sessions. The predictor and test outputs were unchanged; additional calibration altered the bound rather than the achieved fidelity. At scale 0.08 and $K = 8$, mean fidelity remained 0.956221 but the mean lower bound fell to 0.005898. High average performance therefore did not justify a high robust guarantee.

There were no lower-bound violations among interval-covered states. Across 86,400 scoring rows, 120 violations involved 21 unique test records outside their intervals. No main-experiment threshold display was false. A diagnostic shift of 0.25 radians in test angles reduced 199-session coverage to 261/1,080 and produced 296 false threshold-display rows across both calibration sizes. These derived rows are correlated. This stress test exposes the dependence on stable session distributions. These angle records

were synthetic. The new hardware study below instead calibrates operational fringe phases; it does not supply SPAM-independent gate-angle bounds for applying Eq. (5) to the physical checkpoints.

2.3 Quantum phase estimation: amplified phase uncertainty

For a known, preparable eigenstate $|u\rangle$ with $U|u\rangle = e^{2\pi i\phi}|u\rangle$, consider m counting qubits after the first r low-order controlled powers. The counting state is a product of phase states. If a reconstruction uses the central phase and differs by δ cycles, with extra phase error η_j radians at power j , its full-state fidelity is

$$F_r(\delta, \boldsymbol{\eta}) = \prod_{j=0}^{r-1} \cos^2(\pi 2^j \delta + \eta_j/2). \quad (7)$$

For $|\delta| \leq h$, $|\eta_j| \leq e_j$ and $2\pi 2^j h + e_j \leq \pi$, the exact worst value of this center preparation is the product evaluated at $(h, \mathbf{e})$. All factors decrease with their absolute phase error, and that corner simultaneously maximizes each error. A common inverse quantum Fourier transform preserves the full-state fidelity. Equation (7) is not the probability of obtaining a correct phase-estimation bin [10, 16].

With $r = m$, no extra power-specific error and terminal noise 0.025, the center-preparation half-width boundaries are 0.027459606, 0.006786213 and 0.000846794 cycles for $m = 3, 5, 8$, respectively. The full register includes the known one-qubit eigenstate, so $d = 2^{m+1}$. A strict inequality below the boundary gives $F > 5/6$. The required phase precision scales approximately as 2^{-m} in this regime.

The endpoint upper bound can be attained on a short continuous interval. Set $D = 2^r$, $q_k = k - (D-1)/2$ and $|u(t)\rangle = D^{-1/2} \sum_k e^{2\pi i q_k t} |k\rangle$, after removing the known center phase. Let $K_D(t) = \sin(\pi D t) / [D \sin(\pi t)]$, with $K_D(0) = 1$. For $0 \leq h \leq 1/(2D)$, the exact record-only minimax value is

$$t^*(h) = \frac{1 + K_D(2h)}{2}, \quad |v_h\rangle = \frac{|u(h)\rangle + |u(-h)\rangle}{\sqrt{2[1 + K_D(2h)]}}. \quad (7a)$$

The endpoint average density matrix has largest eigenvalue $t^*(h)$, giving an upper bound for every preparation. The overlap of v_h with an interior phase state is a positive weighted sum of cosines, each nonnegative and decreasing with $|t|$ in this interval. Its minimum is therefore attained at the endpoints and equals the upper bound. Supplement S11 gives the continuous-domain proof and the explicit dual certificate. The spectral optimization is established [5]; the additional result is its attainment across the entire compatible interval. Optimizing a measurement on a quantum probe with a prior phase distribution is a different resource and objective [17].

With terminal depolarization, the optimum is $(1 - \epsilon)t^*(h) + \epsilon/d$. For eight counting qubits and $\epsilon = 0.025$, strict fidelity above $5/6$ is attainable if and only if $h < h_*$, where $h_* \simeq 0.000861540576$ cycles solves $(1 - \epsilon)t^*(h_*) + \epsilon/d = 5/6$. The larger intervals are excluded by the same contained pair of boundary states. This closes the gap left by center replay, whose boundary is approximately 0.000846794041 cycles, and increases the admissible half-width by 1.74145% (Fig. 3). This is an increase in tolerated uncertainty, not a relative fidelity gain. Independent power-specific errors are not included in this exact result.

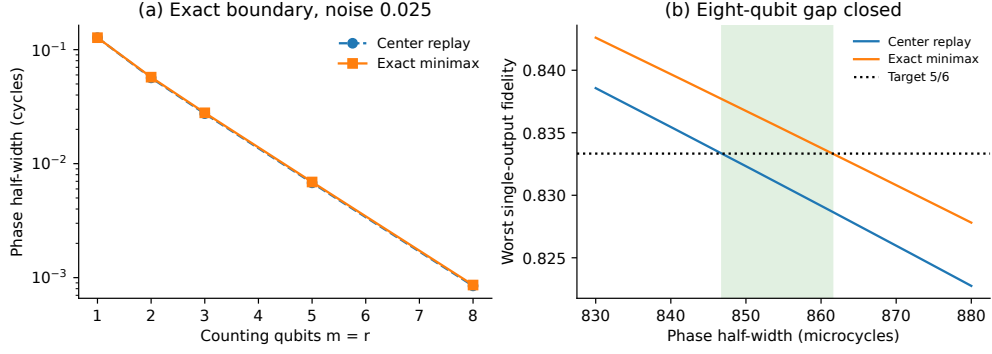


Fig. 3 QPE phase-only reconstruction conditions. (a) Center-replay and exact minimax half-width boundaries for target fidelity $5/6$ at terminal noise 0.025 ; all controlled powers are completed and the known one-qubit eigenstate is included in the noise dimension. (b) For eight counting qubits, the endpoint-superposition preparation closes the previous gap. Strict target excess is attainable below 0.000861540576 cycles and impossible at or above that boundary. The colored strip indicates improvement over center replay. Additional power-specific errors and unknown unpreparable eigenstates are outside this theorem.

The optimizer is a known state and can be synthesized deterministically. Alternatively, one flag qubit, r controlled-phase gates and a flag measurement prepare it with success probability $p = t^*(h) \geq 1/2$ per attempt. Failures consume fresh blank preparations, not the retained source. Two outputs prepared independently have expected total attempt count $2/p$; with at most L attempts per output, joint success probability is $[1 - (1 - p)^L]^2$. At the eight-qubit boundary, $p = 0.854650775$ and this probability is 0.999107550 for $L = 4$. Supplement S11 reports both measurement branches, deterministic and heralded circuit counts, and waiting/reset exclusions. The efficient heralded implementation is not a finite-depth deterministic guarantee.

No phase-only record can reconstruct an unknown, unpreparable eigenstate by this argument. Superpositions of eigenphases require a different joint-state analysis. Acquiring the phase record also has a cost: replay through generic sequential calls to U uses $2^r - 1$ calls to realize the controlled powers, whereas direct phase-state preparation presupposes the phase information.

2.4 Shor checkpoints: discrete ambiguity rather than small angles

At a modular-exponentiation checkpoint, let $Q = 2^m$ and let c_j denote the completed modular multipliers, all units modulo the same known N . The full joint state is

$$|\Psi_c\rangle = Q^{-1/2} \sum_{x=0}^{Q-1} |x\rangle \left| \prod_{j < k} c_j^{x_j} \bmod N \right\rangle. \quad (8)$$

The overlap with a guessed multiplier sequence is the fraction of x values on which the two modular functions agree, and fidelity is its square. If one multiplier differs, fix every bit except that multiplier's control bit. The matching condition takes the form

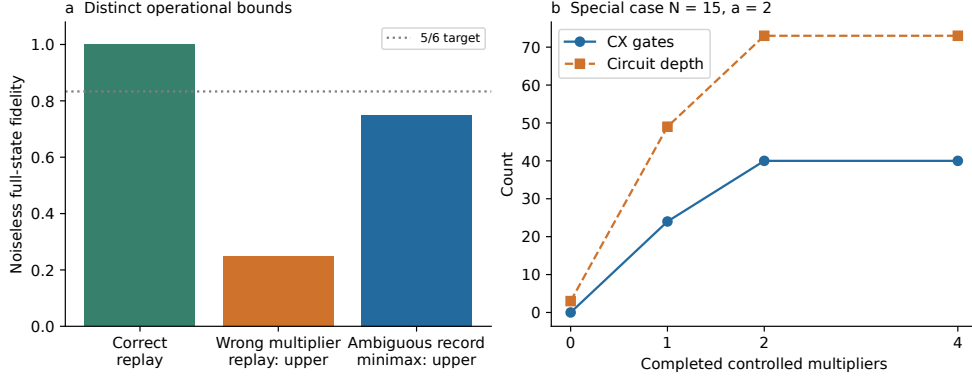


Fig. 4 Shor checkpoint conditions and representative preparation cost. (a) Correct replay attains fidelity one ideally; a replay with a different unit multiplier has fidelity at most $1/4$; true ambiguity between two multiplier sequences imposes an arbitrary-preparation minimax upper bound of $3/4$. The latter two bars are bounds, not measured means. (b) Native controlled-SWAP implementations for the special case $N=15$, $a=2$ use the displayed CX counts and depths. This order-four example is not an estimate of general Shor resource scaling.

$dd_j^{x_j} = 1$ with $d_j \neq 1$. Both $x_j = 0$ and $x_j = 1$ cannot satisfy it. Consequently the matching fraction is at most one half, and wrong-multiplier replay has $F \leq 1/4$, even when several errors partially cancel.

Two genuinely compatible but distinct multiplier sequences therefore have overlap at most one half. Equation (4) then gives $t_x^* \leq 3/4$ for any record-only preparation. This is a stronger obstruction to a $5/6$ guarantee than the fidelity of a particular wrong replay. It applies to the full joint pure state, not automatically to a reduced counting register or a general noisy state.

For ideal Shor circuits the multipliers are $c_j = a^{2^j} \bmod N$. If N , a and the checkpoint are known, repeated squaring regenerates them exactly without a factoring or order oracle [11]. Missing logs then do not represent irreducible uncertainty. Replay has ideal fidelity one but requires the modular arithmetic prefix to be run again. In contrast, unrecorded execution faults may create real uncertainty that knowledge of the intended circuit cannot remove. For the separate model of phase errors on the control bits with correct multipliers, $F = \prod_j \cos^2(\eta_j/2)$; the discrete multiplier obstruction must not be treated as a continuous angle error.

We checked ninety mapping conditions across ten new (N, a) pairs using twelve counting and six work qubits. We also verified twelve eight-qubit controlled-SWAP circuits for $N = 15$, $a = 2$. Noiseless prefixes of zero, one, two and four multipliers used 0, 24, 40 and 40 CX gates and had depths 3, 49, 73 and 73. The eighteen-qubit mapping calculation was not a native-gate synthesis demonstration.

2.5 qDRIFT: high conditional means can coexist with robust impossibility

qDRIFT selects Hamiltonian terms randomly and applies a short rotation for each selected term [12]. Our target is the intermediate state of a particular realized term history, not the exact Hamiltonian-evolution state or the channel averaged over histories. For Pauli terms, write $U_\ell = \exp(-i\tau P_\ell)$ with $\tau = \lambda T/L$ and $\lambda = \sum_\ell |h_\ell|$. When K term choices are missing, simply omitting these steps gives the state-independent bound

$$F_{\text{skip}} \geq \cos^2[\min\{\pi/2, K\tau\}]. \quad (9)$$

For two qubits and terminal depolarization 0.025, $K\tau < 0.4000874755$ is sufficient for $F > 5/6$. At $\tau = 0.05, 0.10, 0.20$, this guarantees up to eight, four and two omitted steps, respectively. At $\tau = 0.1$, $K = 4$, the lower bound is 0.8334071514, only slightly above $5/6$. This narrow ideal-model margin should not be read as robustness to additional unbudgeted hardware error.

We used the two-qubit Hamiltonian

$$H = 0.45XI + 0.35ZI + 0.30IX + 0.25IZ + 0.40XX + 0.25ZZ, \quad (10)$$

with $\lambda = 2$, forty total steps and checkpoints after ten, twenty and forty steps. Ten new initial states and histories were reused across conditions. From 600 visible records we obtained 1,200 noise-conditioned bound classifications and 7,200 realized-target fidelity evaluations. These are not 7,200 independent preparations.

For independent missing term selections, the conditional mean state R_x can be computed exactly by inserting the known random-unitary channel at missing positions [12, Eqs. (5)–(6)]. An independent draw of missing terms prepares R_x ; its pure-target average fidelity is $\text{Tr}(R_x^2)$. Preparing a maximal eigenstate achieves Eq. (3) [5, Eqs. (13)–(16)]. A learning algorithm cannot extract a particular independent hidden draw from unrelated visible draws when their distribution is already known. Better fidelity can nevertheless come from an output adapted to the conditional ensemble rather than from correct classification of every term.

We evaluated exact worst fidelities of the specified methods by enumerating all 6^K candidates for $K \leq 4$. For larger K , we used analytic achievable bounds and pairs drawn from a 256-candidate subset only as upper-bound witnesses. A subset minimum was never used as a lower guarantee over all candidates. For checkpoint twenty, $\tau = 0.1$ and ten missing terms, eight of ten records had robust upper bounds at most $5/6$; two remained unresolved. Nevertheless the record-averaged optimal conditional mean fidelity was 0.924973. Maximal-eigenstate preparation on the ten realized targets had mean 0.899714 and sample minimum 0.776865. After forty steps with twenty missing terms, all ten records excluded a robust $5/6$ excess. These results show why average prediction performance cannot substitute for compatible-set analysis.

2.6 Preparation resources and learned reconstruction

The preceding optimal conditional average assumes unrestricted synthesis of the chosen state. An earlier frozen ablation tested how record-based learning behaves under a fixed preparation budget. It used 100 independent repetitions in six groups, 120

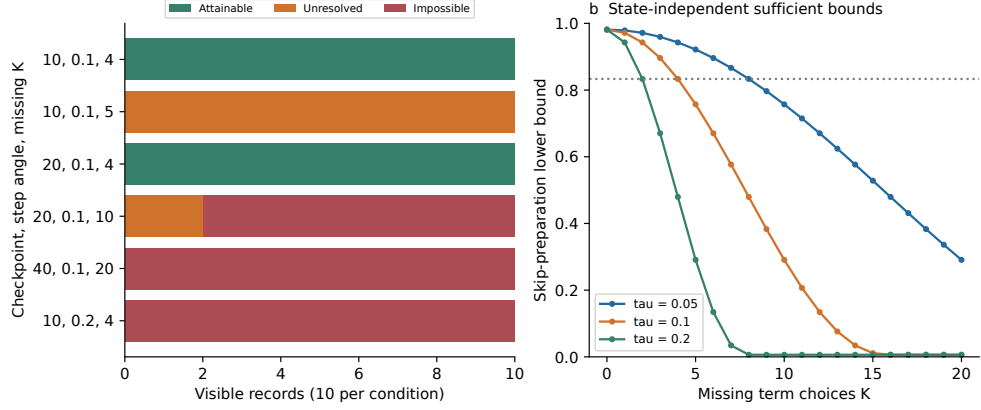


Fig. 5 qDRIFT conditions at terminal depolarization 0.025. Left: attained lower bounds and valid upper witnesses classify ten records per condition. Attainable means an evaluated method guarantees fidelity strictly above $5/6$ for all compatible pure targets; impossible means an upper bound is at most $5/6$; otherwise the case is unresolved. Conditions reuse the ten base preparations. Right: analytic omission bounds, independent of the visible record, for three step angles. Neither an unresolved classification nor a low sufficient bound proves impossibility.

training, thirty validation and ten test records per repetition, and eleven fixed methods. Half of the gates were masked. Across 6,000 unique test inputs, the record-conditioned ten-neighbor mixture improved mean fidelity over a 120-state unconditional mixture by 0.144664 for pure and 0.042677 for mixed targets. Approximate familywise 95% intervals for four prespecified contrasts were $[0.139230, 0.150099]$ and $[0.039661, 0.045694]$. Distribution-free Hoeffding sensitivity intervals included zero; we therefore do not present the approximate intervals as assumption-free evidence.

Under the same operational budget, replacing the local mixture by its maximal eigenstate decreased fidelity by 0.052438 for pure and 0.072521 for mixed targets. Across all methods, 9,655 evaluations exceeded a quantum preparation limit and were scored using the prescribed maximally mixed replacement. No execution failure was excluded. Every five-qubit eigenstate-method case in each of four relevant arms was replaced. This is a limitation of those synthesis choices and budgets, not a fundamental lower bound on state preparation. No pure-target group-method mean exceeded $5/6$. A ridge-regression method also outperformed the neighbor mixture in overall pure-target mean, so the information contrast does not establish a universally best learner. Separation of method choice from test evaluation is essential [18]. Independent calibration further requires a fixed score rule and exchangeable future sessions [15]. Neither safeguard alone proves the other.

Structured angle prediction used the known circuit directly and avoided generic state synthesis. The five-qubit multigate preparation required twelve CX gates per output; a representative native circuit had depth 47, while the all-angle construction bound was 92. The 180 representative two-qubit qDRIFT output circuits were also verified after native synthesis. Their maximal-eigenstate preparations used one CX gate

and depth eleven. This small-system result does not establish efficient generic many-body eigenstate synthesis. Costs in this paragraph are per additional output: preparing both requires two preparations and, if sequential, twice the corresponding work.

2.7 Hardware diagnostics: forward coverage, source waiting and total cost

We tested operational prerequisites in a new pilot on the IBM `ibm_marrakesh` processor. Forty chronological acquisition jobs measured representative QPE, Shor and qDRIFT blocks. Six jobs supplied sixty training records per algorithm, twenty-four jobs calibrated fixed prediction intervals, and ten later jobs supplied one hundred test records per algorithm. Each job contained ten commanded angles per block. A job defines an acquisition session here; distinct jobs do not establish day-to-day independence or exchangeability.

The observable was the fringe phase $\arg(\langle X \rangle + i\langle Y \rangle)$ from independent diagnostic specimens. The QPE probe used a known eigenstate, the Shor probe used a uniform work-register eigenstate for the first $N = 15$, $a = 2$ controlled-multiplier block, and the qDRIFT probe used the ZZ term with a fixed spectator. These probes are distinct from the complete checkpoints used for the memory test. State-preparation and measurement (SPAM) errors and nonunitary effects can bias the fringe phase; we do not identify it with an intrinsic effective gate angle without further calibration.

The calibrated half-widths were 0.5792, 0.9807 and 0.5117 radians, respectively. In the ten later sessions, every independent test-reference interval was contained in its prediction interval in ten QPE, nine Shor and eight qDRIFT sessions (Fig. 6b). The remaining one Shor and two qDRIFT sessions were unresolved because their reference intervals overlapped prediction boundaries; none had a disjoint reference interval. All point-reference estimates were inside their prediction intervals. The distinction matters: point-estimate coverage is weaker than reference-interval containment, and neither observation establishes unconditional coverage of physical gate angles. Under the assumed exchangeability and shot model, the chosen error allocations give a marginal coverage lower bound of 0.95 for the operational phase in one future session of a specified algorithm, not simultaneous validity for all algorithms and future sessions.

A separate job compared immediate source diagnosis, matched waiting without output preparation, and waiting during sequential preparation of B and C. It used ten complete checkpoint cases per algorithm and 512 shots per arm per case. The checkpoints contained three, eight and two source qubits for QPE, Shor and qDRIFT, respectively. Source A’s non-delay gate sequence was identical across each matched triple, and the compiled B/C preparation interval contained only delays on A. Diagnostic inverse circuits followed by measurement yielded the raw all-zero return probability P_A^{ret} , which includes preparation, inverse-circuit and readout errors and is not a directly measured state fidelity.

The mean immediate/matched-idle/active-output return probabilities were 0.7244/0.2918/0.2934 for QPE, 0.1840/0.0389/0.0451 for Shor and 0.7738/0.6285/0.6508 for qDRIFT (Fig. 6a). Waiting arms therefore had lower source-return statistics. Under independent-shot assumptions, conservative simultaneous intervals for the nine fixed-case arm means have half-width 0.0240. The corresponding

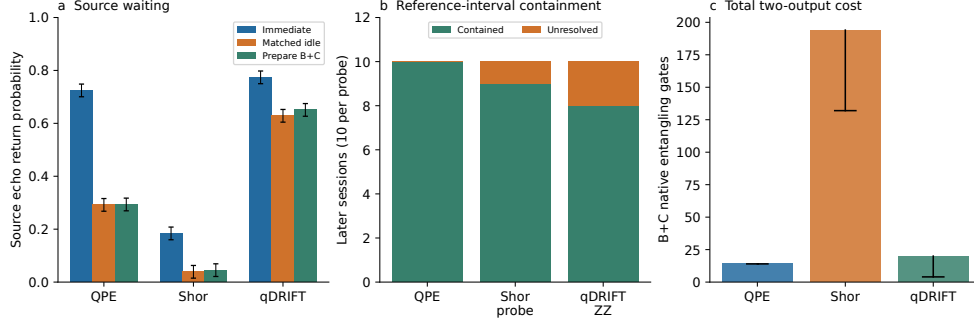


Fig. 6 Hardware diagnostics on `ibm_marrakesh`. (a) Source all-zero echo-return means for ten checkpoint cases per algorithm, with 512 shots per case and arm. Error bars are conservative simultaneous Hoeffding half-widths for the nine fixed-case shot-expectation means; they exclude drift and new-case uncertainty. Return probability is not state fidelity. (b) Number of later acquisition sessions in which all ten test-reference phase intervals were contained or remained unresolved. No disjoint intervals occurred. These are representative block probes, not full-checkpoint effective-angle certificates. (c) Executed two-output native entangling-gate counts: bars show maxima and black spans show the minimum–maximum range across ten cases. All output-preparation costs include both B and C.

active-minus-idle differences, 0.0016, 0.0063 and 0.0223, all had intervals including zero. Thus the results do not resolve an additional effect of B/C activity relative to the matched wait, nor do they isolate a unique physical mechanism. Arm order was not randomized, and the uncertainty intervals do not cover temporal drift or generalization to new states.

Preparing both outputs required fourteen native entangling gates for the QPE examples, 132–194 for Shor and 4–20 for qDRIFT. The sequential B/C preparation intervals were 1.696, 10.976–17.028 and 0.656–2.776 microseconds. Counts refer to the processor’s native entangling gates, not CX-equivalent costs. Logical A+B+C sizes were nine, twenty-four and six qubits; three eight-qubit physical groups were reserved to allow routing. Direct same-shot joint BC return probabilities averaged 0.6928, 0.0451 and 0.8131, respectively, without assuming a product of hardware marginals. Full cost details and confidence conventions appear in Supplementary Information S8.

The provider charged 248 QPU seconds across forty phase jobs and one memory job. These outputs were prepared by known-record prefix replay, rather than by transferring the learned phase correction to other physical layouts. The study therefore demonstrates acquisition, forward evaluation and resource accounting, but does not establish learned improvement of full-state fidelity or an end-to-end advantage over replay. Most importantly, ideal identity-on-source preservation cannot be treated as a hardware-validated property on the basis of these diagnostics.

2.8 Testing the phase model and randomized waiting controls

A follow-up experiment on the same backend collected 293,376 shots across 978 settings in six successful jobs. Three commanded angles, 0.35, 0.75 and 1.15 radians, were measured at eight repetition counts in each probe context. We fitted independent X/Y

offsets, contrasts and reference phases together with a common oscillation magnitude ϕ and damping exponent γ , reserving two repetition counts for prediction checks (Methods; Supplement S9). Under the stated bootstrap diagnostics and multiplicity correction, the single-mode model was rejected at all three QPE and all three Shor settings. It was not rejected at the three qDRIFT-ZZ settings. Nonrejection does not certify the model or separately identify preparation, measurement and gate errors. A two-frequency coherent signal can display an attenuation envelope without following the fitted exponential law (Supplement S10.3). Thus γ is a descriptive parameter of this model, not an identified general nonunitary error rate; the present data do not determine a unique cause of attenuation.

For qDRIFT-ZZ, phase estimates were 0.3823, 0.7625 and 1.1484 radians, with local model-dependent 95% intervals [0.3728, 0.3927], [0.7560, 0.7691] and [1.1426, 1.1542]. These intervals do not provide the circuit-angle error guarantees required by Eq. (5). A native-instruction audit found that Shor’s first routed block differed from subsequent blocks. Absorbing this first block into preparation in a separately documented supplementary analysis did not resolve the model failures. Thus a single fitted phase cannot summarize the tested hardware errors across all three algorithm contexts.

For memory controls, every fixed checkpoint received all six permutations of immediate, matched-idle and active-output arms, once each. Case order was shuffled within blocks. This balances submitted PUB order; actual per-shot execution order was not exposed by the provider. Source inverse-return means and contrasts are reported in Table 1:

Table 1 summarizes the randomized waiting controls.

Table 1: Source inverse-return probabilities and active-minus-idle contrasts; these are not state fidelities

Checkpoint	Immediate	Matched idle	Active B/C	Active minus idle, simultaneous 95% interval
QPE	0.6794	0.2154	0.2178	0.0025 [-0.0242, 0.0292]
Shor	0.1413	0.0408	0.0388	-0.0020 [-0.0287, 0.0248]
qDRIFT	0.7724	0.6029	0.6099	0.0070 [-0.0197, 0.0337]

The intervals cover six prespecified fixed-case contrasts under independent-shot assumptions; they exclude drift and new-state uncertainty. Waiting-arm decrements remained visible, whereas the active-minus-idle contrasts remained unresolved. This is not an equivalence or no-crosstalk result. All entries are noisy inverse-return probabilities, not state fidelities. Direct BC/ABC joint returns, circuit audits and a comparison figure appear in Supplement S9.

Successful acquisition took 220.683 seconds of wall time and 4.258 seconds of client process CPU time. Its charge was 54 QPU seconds; a retained failed attempt

added nine seconds, for a total of 63. Separate local two-output construction, native compilation and serialization required mean CPU times of 0.023497, 0.030736 and 0.015370 seconds per QPE, Shor and qDRIFT case, respectively. These offline replay costs are distinct from the executed microsecond-scale preparation intervals. Provider and client intervals overlap and are not added as disjoint costs. In this implementation, no source was retained across separate IBM jobs: measured cloud turnaround is a workflow cost, not a demonstrated quantum-memory residence time.

3 Discussion

The results identify what must be known to obtain high-fidelity additional states while retaining a quantum source. Classical preparation records alter the available information, and their residual ambiguity determines a conditional reconstruction problem. Circuit structure then translates that ambiguity into different fidelity conditions. Calibrated angle intervals yield a general sufficient bound; QPE admits an exact short-interval minimax solution while magnifying shared phase uncertainty; Shor requires correct discrete modular multipliers; and qDRIFT permits bounded omission of small rotations while retaining a substantial gap between mean and worst-case performance.

The theoretical ingredients have established precedents. The maximal-eigenstate average decision follows earlier reconstruction work, and session coverage is an application of conformal ranking [5, 15]. The particular operational model, derivations for checkpoint uncertainty, computational comparisons, hardware diagnostics and explicit failure cases are the scope of this study. The checkpoint bounds and all experimental values are supported by the derivations and frozen records provided here; the cited algorithm papers establish the checkpoints and sampling laws, not these new numerical results. Supplement S10 distinguishes the relevant operation classes and evidence levels. No claim of first priority is made. None of the results improves a universal unknown-input cloning theorem under its original resource assumptions.

Quantum conformal prediction and conformal fidelity estimation already address uncertainty in quantum predictions and measurement-based fidelity estimates [19, 20]. Certification from incomplete measurement constraints is also established [21]. Our calibrated variables are preparation-record angle errors, which are propagated into a reconstruction bound; this changes the input and operational objective, not the conformal rank theorem. Classical control records and random seeds have also been proposed for checkpoint restoration [22]. The distinction here is the fidelity achievable when the retained records are incomplete. These precedents motivate restricting the theoretical contribution to the stated reconstruction conditions rather than claiming a new calibration or replay principle.

Several limitations matter for practical use. The circuit-angle coverage experiment uses synthetic reference-error records. The new hardware pilot measures operational fringe phases and confirms that reference uncertainty must be retained in forward-session evaluation, but it does not bound intrinsic gate-angle error independently of SPAM. Its same-day, single-backend design does not establish session exchangeability under longer-term drift. The source-return measurements are affected by diagnostic inverses and readout. The initial memory experiment did not randomize order; the

follow-up balanced submitted arm order but did not observe the provider’s per-shot execution sequence. Their decline in waiting arms prevents inferring experimental memory preservation solely from the identity channel in Eq. (1). The single-mode follow-up failed for QPE and Shor, reinforcing the distinction between operational fringe measurements and intrinsic-angle certification. The terminal depolarization model used in the numerical study is not a calibrated model of this processor.

The algorithm calculations include ten base qDRIFT preparations and small native circuits; larger instances may be limited by conditional-state calculation and synthesis cost. Numerical checks use floating-point arithmetic, including independent 70-digit root calculations for the QPE thresholds; continuous QPE optimality follows from the analytic proof, not a sampled grid or interval-arithmetic certificate. Hardware cost measurements include both outputs and their on-device waiting interval. The follow-up additionally records CPU, RPC and scheduling costs, but does not measure source residence throughout those classical and cloud intervals. We have not demonstrated an improvement in downstream factoring probability, phase-estimation success or Hamiltonian-simulation accuracy.

The failed single-mode tests motivate richer channel models with independently justified SPAM assumptions and uniform native repeated blocks. Physical validation would also require memory controls across independent days and a demonstrated controller-and-memory path for the actual classical latency. Beyond the phase-only short-interval solution, sharper bounds are needed for additional power errors and general checkpoint uncertainty. Establishing an algorithmic benefit would additionally require comparing record acquisition, protected source storage and two output preparations against straightforward prefix replay. The present data make those costs explicit while leaving hardware-certified reconstruction fidelity and end-to-end advantage open.

4 Methods

Physical resources and scope. The source is not used for classification, tomography or training at reconstruction time. Initial preparation descriptions, visible circuit parameters and masks are classical inputs. Training targets are generated in separate preparations and do not supply extra copies of a particular test source. For a fixed record, both output registers start in input-independent known states. Random mixture preparation uses fresh independent randomness for B and C. Terminal depolarization is applied independently to each output. Shared random choices would give a different joint channel and require separate global-fidelity evaluation.

Session calibration. A multivariate ordinary least-squares predictor and positive coordinate scales s_j were fitted using twelve training sessions of ten records each. Let y_{ibj} be the independent reference angle in calibration session i , record b , coordinate j , and let its circular error relative to the true angle be bounded by u_{ibj} . With d_{circ} denoting circular distance, the session score was

$$S_i = \max_{b,j} \frac{d_{\text{circ}}(y_{ibj}, \hat{\theta}_j(x_{ib})) + u_{ibj}}{s_j}. \quad (11)$$

For M calibration sessions, the rank was $\ell = \lceil (M + 1)(1 - \alpha) \rceil$ and $q = S_{(\ell)}$, with $q = \infty$ when $\ell > M$. Half-widths were $r_j = \min(\pi, qs_j)$. The synthetic experiment used $\alpha = 0.05$ and hard reference-error bound 0.002 radians. Reference uncertainty was enlarged before taking the rank, not estimated from the test targets. A simultaneous probabilistic metrology failure bound β would weaken the overall guarantee to at least $1 - \alpha - \beta$. A correct linear model is not needed for the rank guarantee, although the numerical generator used linear centers. Session exchangeability and a frozen predictor are needed.

Synthetic evaluation and drift. Thirty repetitions at each combination of three qubit counts, four unknown-angle counts and three residual scales gave 1,080 cases, 2,289,600 total records and 10,800 new test records. Two calibration sizes, pure or isotropically mixed targets and two output-noise values yielded 86,400 scoring rows. Mixed targets used $\rho_p = (1 - p)|\psi\rangle\langle\psi| + pI/d$ with $p = 0.3$, and were evaluated using Eq. (2), not the pure-target decision theorem. The drift diagnostic reused the test records after shifting their true angles and did not create independent additional observations. Predictions were sealed before scoring in this experiment.

Algorithm calculations. QPE boundaries were scalar roots of the analytic product formulas for $m = 3, 5, 8$, selected prefixes and per-power radii 0, 0.02 and 0.1 radians. Shor mapping comparisons used modular-function evaluation, while the twelve small controlled-SWAP implementations were checked by statevector simulation. qDRIFT used three step angles, three checkpoints and missing counts from zero through half the checkpoint, with duplicate conditions removed. Exact conditional mean density matrices used the known term probabilities $|h_\ell|/\lambda$. Hidden realized terms were excluded from output-selection functions; this was a logical data-interface separation, not the all-prediction external seal used in the earlier ablation. No learner was newly trained for the qDRIFT study.

Bound classification and numerical verification. A case was attainable when a valid achieved lower bound exceeded $5/6$, impossible when a valid upper bound was at most $5/6$, and unresolved otherwise. Finite candidate witnesses can establish upper bounds but not exhaustive worst-case lower bounds. A separate implementation re-evaluated the 48 QPE boundaries, ninety Shor mapping conditions, twelve Shor circuits, 600 qDRIFT records, 1,200 classifications and 7,200 fidelity rows. The maximum QPE root residual was 1.62×10^{-10} ; the maximum fidelity difference was 1.13×10^{-8} . All 180 qDRIFT native output circuits agreed with their prescribed states to infidelity below 8.0×10^{-15} . Six algorithm unit checks and ten multigate unit checks passed in their original runs. Independent verification of the calibration study reconstructed 32,400 circuit states and all 86,400 scores. These are separate-code checks within the same project, not an external replication.

Resource-controlled ablation. Fixed arms were compared on paired test inputs with 50% whole-gate masking. One-output limits were forty CX gates, depth 240 and 500 one-qubit gates; the classical execution container had one CPU and 2 GiB memory. Resource overruns returned I/d and remained in every aggregate. The four prespecified contrasts used repetition-level inference with 98.75% individual intervals, providing nominal 95% family coverage by Bonferroni under the stated t approximation. The older selected-method experiment was not pooled with this fixed-arm ablation.

Hardware acquisition and operational phase calibration. The forty phase jobs ran sequentially on 7 October 2026 using Qiskit 2.5.2 and Qiskit IBM Runtime 0.50.0. Each contained ten phases per probe, drawn before acquisition from $[-0.6, 0.6]$ radians, measured in X and Y at repetition factors one and two with 256 shots per setting. Two readout controls per job gave 4,880 settings and 1,249,280 shots overall. Readout controls remained diagnostics; a fitted offset was not subtracted without an uncertainty bound. A least-squares model with features $[1, \sin \theta, \cos \theta]$ fitted the wrapped phase residual from training repetition-one outcomes. Its parameters were frozen before calibration. A session score maximized prediction-to-reference phase distance plus reference half-width over all ten records of that probe. The calibration rank was $\lceil 25(1 - 0.04) \rceil = 24$. Exact Clopper–Pearson intervals [23] used a total reference failure allocation $\beta = 0.01$ over 4,800 planned quadratures. Interval geometry, shot assumptions and the distinction from intrinsic-angle metrology are detailed in Supplementary Information S8. No low-visibility case was excluded.

Hardware memory, uncertainty and resources. The memory job used ninety circuits and 46,080 shots. All gates on A except delays were audited for equality across the three arms; routing was restricted to disjoint register groups. Matched-idle duration equalled the sequential B/C preparation interval. Nine pooled source-arm means each comprised ten cases and 5,120 shots. Hoeffding’s bounded-variable inequality [24] and a union bound give simultaneous 95% half-width $\sqrt{\log(18/0.05)/(2 \cdot 5120)} = 0.0240$, conditional on independent shots; difference intervals use twice this radius. This addresses the average expectation of the fixed cases, not drift or new-state generalization. Counts for BC were read directly from joint bitstrings. Native schedules were reconstructed from saved device gate durations, delays and barriers. A resource-only amendment increased the initial additional-QPU cap from 240 to 260 seconds to complete the fixed plan after observing six-second phase-job charges and an eight-second memory-job charge. Circuit definitions, shot counts, fits, intervals and test cases did not change. A separate 609-item audit checked chronology, frozen hashes, count and confidence calculations, while native-state checks covered 36 probe settings and 270 memory-register circuits.

QPE minimax verification. The new phase-only result was checked against direct overlap vectors, endpoint spectra and compiled state-preparation circuits. A separate implementation used tensor-product states and 70-digit product-form boundary roots. An ideal source-reference Bell pair verified that both successful output preparations leave source-reference correlations unchanged. These local checks submitted no hardware jobs. Supplement S11 states circuit connectivity, success flags, finite retry limits and excluded costs.

Follow-up phase model and randomized controls. Three acquisition rounds each contained 146 phase settings with 512 shots per setting and a separate 180-circuit memory job with 128 shots per circuit. Phase probes used repetitions $m = 0, 1, 2, 3, 4, 6, 8, 12$ and the three fixed positive angles. We fitted $E_a(m) = b_a + (1 - |b_a|)A_a e^{-\gamma m} \cos(m\phi + \xi_a)$ with eight parameters, using $m = 0, 1, 2, 4, 6, 12$ and holding out $m = 3, 8$. Binomial-deviance multistart fits and 999 local parametric bootstrap replicates supplied descriptive intervals and model diagnostics. For the fitted-model bootstrap diagnostics, simulated samples were refitted before recomputing

discrepancies, following the fit–simulate–refit approach [25, Sec. 1.1]. Its asymptotic regularity conditions have not been established for our constrained phase-model fits. Holm adjustment covered 27 training-fit, held-out and between-round homogeneity tests [26]. A Shor warm-up analysis, prescribed by the native-circuit audit before count inspection, used $m - 1$ after absorbing the first routed block into preparation. It retained the same held-out original repetitions. Supplement S9 records the finite-sample and optimizer limitations; neither model is full gate-set tomography, which characterizes preparation, gates and measurement self-consistently subject to gauge freedoms [27, Secs. 2.1.4 and 3.1]. Supplement S10.4 derives the multiplicity rule and distinguishes its conditional guarantee for valid p-values from the approximate fitted-model bootstrap used here.

Follow-up uncertainty, timing and amendments. Each of thirty fixed memory cases received all six arm permutations, with case order shuffled independently in each block; two blocks were assigned to each memory job. Every algorithm-arm mean contained 7,680 shots. Six source contrasts had simultaneous 95% Hoeffding half-width $\sqrt{\log(240)/7680} = 0.0267$ under independent shots [24]. Joint BC and ABC outcomes were counted directly. Monotonic wall and process CPU clocks timed acquisition, computation and RPC segments; provider timestamps separately described queue and run intervals. A confirmed initial timeout returned no counts but charged nine QPU seconds. A documented resource amendment increased job ceilings from 12 seconds to 25 for phase and 15 for memory, retaining the 72-second total charged cap, including failure, and a 20-second reserve. Missing client timing detail from the failed attempt was disclosed. An independent coefficient-based likelihood check detected one numerically improvable QPE fit. Equivalent constrained-coefficient refinement was applied to all data and bootstrap fits; original outputs were preserved, and rejection decisions did not change. No hardware settings, exclusions or test splits changed during this numerical correction.

Use of artificial intelligence tools. OpenAI Codex and associated language-model tools assisted literature organization, analytical exploration, code generation, numerical checking, manuscript drafting and revision. This use extended beyond language editing. Automated checks and separate-code checks within this project are not independent external replication.

5 Data and code availability

The code and supporting data are organized in a versioned reproducibility package (v1.6.1). It contains fixed software dependencies, original code, seeds, numerical reference tables and recorded hardware counts and circuits, with file-level checksums. The package regenerates the QPE minimax calculations, algorithm checkpoint conditions and synthetic angle studies, recomputes ablation statistics from saved scored outputs, and verifies recorded hardware data. It does not retrain all historical ablation models or repeat hardware acquisition and all phase-model bootstrap fits. Supplement S7 maps the results to their underlying records. The package is retained locally; no public repository or archival accession is currently available.

6 Statements and Declarations

Funding. No funding was received for this study.

Competing interests. The author is affiliated with QIRI (Quantum Integrated Research Institute Inc.) and declares no competing interests.

Author contributions. Hikaru Wakaura prepared the manuscript with assistance from the AI tools disclosed in Methods and is its sole author. No other human contributors were involved in writing the manuscript. The author retains responsibility for the scientific content, interpretation and final text.

Supplementary information. The Supplementary Information appended to this preprint contains proofs, study designs, statistical assumptions, hardware diagnostics, resource accounting and evidence provenance. References to Supplement S1–S11 throughout this article identify sections of the appended supplement.

Supplementary Information

S1. Physical operation and the universal-cloning comparison

For a fixed classical record x , let $\sigma_x = \sum_b w_b |\phi_b\rangle\langle\phi_b|$. A physical preparation draws b with probability w_b , initializes a blank register and applies a preparation circuit for $|\phi_b\rangle$. Two independent draws give $\sigma_x \otimes \sigma_x$. The map in main-text Eq. (1) is completely positive and trace preserving because it is an identity channel tensored with preparation channels. For an arbitrary reference R , it acts as

$$(\text{id}_R \otimes \Gamma_x)(\rho_{RA}) = \rho_{RA} \otimes \sigma_x \otimes \sigma_x.$$

Thus every source–reference correlation is preserved in the ideal model. The classical record may itself be correlated with the input through a preparation process. The general input resource is then a classical–quantum state $\sum_x p_x |x\rangle\langle x| \otimes \rho_{A|x}$, not an arbitrary source tensored with a fixed uninformative record. If the record supplies no constraint and the source cannot be interrogated, the best worst-case single-output overlap with arbitrary pure states is $1/d$, attained by I/d . Indeed the minimum overlap for a fixed σ is its least eigenvalue, at most $1/d$.

The usual universal-cloning comparison uses a different operation class. A one-copy cloner can interact with and disturb the input but has no informative preparation record. Its optimal symmetric single-output fidelity is $(d+3)/[2(d+1)]$ [2]. An attaining channel is

$$\Lambda_W(\rho) = \frac{2}{d+1} P_+(\rho \otimes I) P_+, \quad P_+ = (I + S)/2.$$

Complete positivity follows by writing $\rho \mapsto \rho \otimes I$ as a sum of embedding Kraus maps and composing with P_+ . Since $\text{Tr}[P_+(\rho \otimes I)] = (d+1) \text{Tr} \rho/2$, the normalization is trace preserving. Swap symmetry is explicit. Expanding the two projectors and tracing out either output gives

$$\rho_B = \rho_C = \frac{d+2}{2(d+1)} \rho + \frac{I}{2(d+1)}.$$

For a pure target this yields $(d+3)/[2(d+1)]$. The joint fidelity is $2/(d+1)$ because $P_+|\psi\psi\rangle = |\psi\psi\rangle$ [1]. In general this is not the product of the marginal fidelities.

For completeness, a short upper-bound proof for the single-output objective uses the unnormalized Choi convention $J \succeq 0$, $\text{Tr}_{BC} J = I_A$. Let $|\Omega\rangle = \sum_i |ii\rangle$ and $\Phi = |\Omega\rangle\langle\Omega|$. The Haar-averaged symmetric single-output score is $\text{Tr}(JM)$, where

$$M = \frac{2I + \Phi_{AB} \otimes I_C + \Phi_{AC} \otimes I_B}{2d(d+1)}.$$

The two entangled-subspace projectors Φ_{AB}/d and Φ_{AC}/d have cross-overlap singular value $1/d$, so the largest eigenvalue of their sum is $1 + 1/d$. Consequently $\lambda_{\max}(M) = (d+3)/[2d(d+1)]$. Since $\text{Tr} J = d$, the average fidelity is at most $(d+3)/[2(d+1)]$. The worst-case minimum of the two marginal fidelities cannot exceed that average; the covariant symmetric channel attains equality for every pure

input. This proof includes measurements and discarded environments. A circuit-depth constraint is not automatically represented by the Choi constraints.

For $n = 1, 2, 5$, the single-output optima are $5/6$, $7/10$ and $35/66$, and the joint optima are $2/3$, $2/5$ and $2/33$. As $n \rightarrow \infty$, they approach $1/2$ and zero, respectively. Additional input-independent ancillas do not raise these channel-class bounds. In the reconstruction problem, ancillas may enable a chosen preparation within a resource budget but cannot reveal information absent from the record.

S2. Conditional mean, minimax and objective mismatch

For an exactly known pure-state conditional ensemble,

$$\mathbb{E}[F | x] = \text{Tr}(R_x \sigma_x).$$

Diagonalizing R_x shows that this linear functional is at most $\lambda_{\max}(R_x)$ over density matrices, with equality on a maximal eigenvector. This is the earlier accessible-fidelity reconstruction rule [5]. A learned or empirical $\hat{R}_x$ need not be the true conditional mean; maximizing its eigenvalue objective does not prove optimal test fidelity.

For two pure candidates with overlap magnitude s , the nonzero eigenvalues of their equal mixture are $(1 \pm s)/2$. Therefore

$$\inf_{\psi \in \mathcal{S}_x} F(\psi, \sigma) \leq \frac{1}{2}[F(\psi_1, \sigma) + F(\psi_2, \sigma)] \leq (1 + s)/2.$$

For exactly two candidates this bound is attainable by the equal superposition after aligning their relative phase; for a larger set it is only an upper bound. For a finite candidate set, Eq. (4) becomes the SDP

$$\text{maximize } t : \quad \sigma \succeq 0, \quad \text{Tr } \sigma = 1, \quad \text{Tr}(\rho_i \sigma) \geq t \quad \text{for every } i.$$

Its dual minimizes $\lambda_{\max}(\sum_i w_i \rho_i)$ over probability weights w_i . Strong duality follows from a strictly feasible primal point $\sigma = I/d$, $t < 1/d$. The uniform pair bound is one feasible dual choice. Neither this SDP nor the conditional mean calculation imposes native gate count or depth. The current numerical study used these analytic bounds rather than a large-scale SDP sweep.

The product preparation has global pure-target score $(\langle \psi | \sigma | \psi \rangle)^2$, whose expectation depends on second moments. A maximal eigenvector of R_x need not optimize it. For mixed targets, squared Uhlmann fidelity is not linear in σ , so the pure-state eigenvector argument cannot be transferred. These distinctions also separate cloning, whose desired output is $\rho \otimes \rho$, from broadcasting, whose marginal requirements alone are ρ [4, 14]. A commuting family may be broadcast without producing independent copies of each member.

S3. Angle-error bound and coverage proof

Consider one rotation discrepancy δ for a Hermitian involution P . For any normalized pure state,

$$\langle \psi | e^{-i\delta P/2} | \psi \rangle = \cos(\delta/2) - i\langle P \rangle_\psi \sin(\delta/2).$$

Its squared magnitude is at least $\cos^2(\delta/2)$. A circular representative $|\delta| \leq \pi$ therefore produces Fubini–Study distance at most $|\delta|/2$. The Bures angle associated with the squared fidelity in Eq. (2) is a metric and reduces to this angle for pure states [13, Sec. III, p. 062310-4]. Replacing gates successively creates intermediate states whose adjacent distances obey this inequality. All later common unitaries preserve distance. The triangle inequality gives main-text Eq. (5) after capping total distance by $\pi/2$. The circular equivalence is valid for these Pauli rotations up to global phase; a controlled block with physically meaningful relative phase must be modeled with its actual generator instead.

The same reasoning gives a more general Bures-angle condition when known initial-state and non-angle errors are bounded. If the initial mismatch has Bures angle at most γ_0 and validated target/output implementation bounds are η_t, η_o , then

$$F_{\text{actual}} \geq \cos^2 \left[\min\{\pi/2, \gamma_0 + \frac{1}{2} \sum_j r_j + \eta_t + \eta_o\} \right].$$

This is meaningful only when those bounds themselves have been justified. It does not permit setting unknown hardware errors to zero.

To prove session coverage, condition on the fitted training predictor and scales. Let T_i be the session maximum of the normalized *true* circular errors. Assume the calibration sessions and one future session have the same acquisition rule and are exchangeable. With a simultaneous reference-error event, $T_i \leq S_i$ for all calibration sessions by the circular triangle inequality. For $\ell = \lceil (M+1)(1-\alpha) \rceil$, exchangeable ranking gives

$$\Pr\{T_{\text{new}} \leq T_{(\ell)}\} \geq 1 - \alpha,$$

with infinite threshold if $\ell = M+1$ and conservative handling of ties. As $T_{(\ell)} \leq S_{(\ell)}$, the observed threshold is also valid. If $qs_j > \pi$, clipping at π retains the entire angular circle. No independence of records within a session or correctness of the regression model is required. These are applications of standard split-conformal reasoning [15].

If reference validity fails with probability at most β , a union bound gives coverage at least $1 - \alpha - \beta$. This statement requires an unconditional exchangeable-score premise and a simultaneous metrology event; separate per-angle 99% statements do not by themselves give a simultaneous $\beta = 0.01$ event. The input adapter can allocate $\alpha = 0.04$, $\beta = 0.01$ for a 95% lower guarantee. Using $\alpha = 0.05$, $\beta = 0.01$ instead gives 94%.

The guaranteed event is coverage of every record and angle in *one* future session. The event of displaying a fidelity bound above the target while that bound is invalid is contained in noncoverage. Its joint probability is therefore bounded, but the error probability conditional on a display need not be at most α . Multiple future sessions require a separate simultaneous guarantee or error allocation. We do not infer these stronger guarantees from the observed absence of false displays.

S4. Isotropic mixed targets and one-angle conditions

The mixed calculations use $\rho_p = (1-p)|\psi\rangle\langle\psi| + pI/d$ and output $\sigma_q = (1-q)|\phi\rangle\langle\phi| + qI/d$, with $q = p + (1-p)\epsilon$. Define $a_0 = 1-p+p/d$, $a_1 = p/d$, $b_0 = 1-q+q/d$, $b_1 = q/d$ and $t = |\langle\psi|\phi\rangle|^2$. The two-dimensional span of the vectors and its complement give

$$F(\rho_p, \sigma_q) = \left[\sqrt{t(a_0b_0 + a_1b_1) + (1-t)(a_0b_1 + a_1b_0) + 2\sqrt{a_0a_1b_0b_1} + (d-2)\sqrt{a_1b_1}} \right]^2.$$

The coefficient of t inside the radical is $(a_0 - a_1)(b_0 - b_1) \geq 0$. Substituting a valid lower bound for t therefore gives a valid achieved lower bound for this isotropic family. This is not an optimum over arbitrary mixed reconstructions. In particular, a high mixed-state fidelity should be compared to the information-free I/d baseline.

A prior controlled experiment isolated one uncertain rotation:

$$|\psi\rangle = WR_Y(c + \xi)|0\rangle, \quad c = \beta^\top(1, \mathbf{x}), \quad \xi \sim \text{Uniform}[-h, h],$$

where W is a known embedding and preparation circuit. For known c and $0 \leq h < \pi/2$, the center state has optimal mean fidelity $(1 + \sin h/h)/2$ and optimal worst fidelity $(1 + \cos h)/2$. The first follows by averaging the two-dimensional density matrix, and the second is attained at the interval endpoints with equality in the two-candidate upper bound. If the center estimate has error at most b , then the sufficient noisy condition is $h + b$ below the radius in main-text Eq. (6).

At $h = 0.6$, increasing the nested training size from 120 to 480 changed mean fidelity from 0.946281 to 0.946620 at five qubits and terminal noise 0.025. The pointwise 99% bound-display fraction increased from 0.4% to 61.6%; a more conservative simultaneous 95% analysis over 18,000 prespecified points increased from zero to 6.3%. These are different confidence objectives. At $h = 0.8$, even a known center cannot guarantee noisy fidelity above $5/6$ throughout the interval, despite a mean near 0.925. The experiment used 100 repetitions at each of nine widths, 9,000 unique test IDs and 126,000 scoring rows. These are a separate data family from the subsequent multigate session experiment.

S5. Algorithm-specific derivations and units

QPE. Before an inverse Fourier transform, the counting state factors as $\bigotimes_{j < r} (|0\rangle + e^{2\pi i 2^j \phi} |1\rangle) / \sqrt{2}$, with untouched counting qubits remaining in $|+\rangle$. Taking overlaps factor by factor yields main-text Eq. (7). With no extra phase errors,

$$F_r(\delta, 0) = \left| 2^{-r} \sum_{z=0}^{2^r-1} e^{2\pi i z \delta} \right|^2.$$

Within the monotonic box, choosing $\delta = h$ and every $\eta_j = e_j$ attains the worst corner. A common underlying phase error is multiplied by 2^j ; it must not be analyzed as an independent equal-size power error. For independent equal-size power radii e and no common phase error, the condition becomes $\cos^{2^r}(e/2) > [f_\star - \epsilon/d]/(1 - \epsilon)$. These are two different error models.

Shor. Set $d_j = \hat{c}_j c_j^{-1} \bmod N$. If one $d_j \neq 1$, then after fixing other bits, agreement is $Dd_j^{x_j} = 1$. At most one of the two equally weighted choices of x_j satisfies it. Thus every distinct unit-multiplier sequence has overlap at most $1/2$. The bound survives additional differing multipliers and a common inverse Fourier transform. For a method that replays a guessed multiplier sequence, let p_{exact} be the probability of an entirely correct sequence. Its mean noiseless fidelity is at most $p_{\text{exact}} + (1 - p_{\text{exact}})/4$. Hence $p_{\text{exact}} > 7/9$ is necessary, but not sufficient, for mean fidelity above $5/6$. This classification threshold does not apply to arbitrary output states that are not guessed replays.

qDRIFT. Each missing unitary is compared to the identity. Its state-distance change is at most τ for the evaluated $0 < \tau < \pi/2$. Replacing K missing steps gives main-text Eq. (9). An initial Bures angle γ_0 and additional step-parameter uncertainties u_j give total angle budget $K\tau + \sum_j u_j + \gamma_0$. The conventional Pauli gate angle is 2τ , so an angle half-width r_j contributes $u_j = r_j/2$.

At a missing position the conditional mean evolves by $\rho \mapsto \sum_\ell p_\ell U_\ell \rho U_\ell^\dagger$, while visible positions apply their known unitary. Repeating these operations computes the exact conditional mean without enumerating every history. It does not identify the realized hidden trajectory. A saved random seed or term log removes that ambiguity and enables replay, but is an additional resource. The present fidelity is relative to the realized trajectory; it is not qDRIFT’s approximation error relative to $\exp(-iHT)$.

S6. Data families, implementation resources and hardware limits

Table S1 summarizes the numerical study designs.

Table S1: Study designs, sample units and evaluation rows

Study	Independent design units	Unique test records	Evaluation rows	Main role
Fixed-arm ablation v1.2	100 repetitions in six groups	6,000	66,000	Information and synthesis budgets
One-angle conditions	100 repetitions at nine widths	9,000	126,000	Learned center and uncertainty width
Multigate calibration	30 repetitions in 36 settings	10,800	86,400	Session coverage and angle bound
Algorithm qDRIFT	10 base preparations, reused	10 base histories	7,200	Record-specific mean and worst case

Repeated evaluations for noise, mixtures, calibration sizes and methods are not independent samples. The manuscript treats displayed means and counts as finite-design descriptive values except for the explicitly specified ablation contrast intervals. Confidence in the analytic coverage theorem comes from its assumptions and rank proof, not a binomial test on pooled heterogeneous cases.

The ablation and later angle studies have distinct distributions and operational tasks. Their mean fidelities must not be compared as a controlled estimate of progress. All twelve-CX angle preparations use a known five-qubit circuit structure; they do not solve generic five-qubit synthesis. The one-angle known- W implementation used eight CX gates, whereas an example generic state-synthesis route used 122. These are examples of exploiting available structure, not universal synthesis lower bounds.

The qDRIFT native cost checks covered 180 representative two-qubit circuits. Maximal-eigenstate preparation used CX count one, depth eleven and fifteen one-qubit gates. Omission preparations used CX counts 1–21 and depths 13–132; complete replays used CX counts 5–21 and depths 31–132. The mixture output is implemented by sampling hidden terms and replaying the resulting trajectory. Reporting only the classical density-matrix calculation would omit its physical preparation cost.

Prior IBM measurements used the `ibm_marrakesh` backend on 6 October 2026, with one- and two-qubit target families, three jobs and approximately 38 seconds of total recorded QPU time. These were exploratory low-shot tomography experiments, using 64 or 32 shots in their respective settings and reused small synthetic test sets. They were not rerun for this manuscript, did not include five-qubit targets and are not evidence for the session-angle coverage theorem. Tomography consumes measured specimens and does not certify that the same unmeasured physical source remained unchanged. Existing logs contain programmed circuits, counts and device calibration summaries, but lack independent effective-angle references and simultaneous metrology uncertainty bounds. These earlier data remain separate from the forward-session pilot in S8 and follow-up in S9; none establishes hardware-validated high-fidelity reconstruction.

The new pilot records session IDs, runtime-visible commanded phases, reference-fringe estimates, device context and held-out later sessions. A certified intrinsic-angle test would additionally need independently justified SPAM and nonunitary-error bounds for the full checkpoint context. Commanded angles alone cannot substitute for effective-angle measurements. Device drift must be checked against the exchangeability premise. The implemented record adapter validates schema and partition integrity, but cannot validate the truth of a supplied metrological claim.

S7. Reproducibility and claim provenance

Paths below are relative to the research project root. Table S2 identifies the underlying records for each result; the reproducibility package contains the selected inputs and code required by its supported workflows.

Table S2: Provenance of the main-text results

Main-text result	Frozen result or evidence
Calibration counts and fidelities	<code>calibrated-multigate/run-v1/summary.csv</code> and <code>coverage.csv</code>
Coverage theorem and mixed formula	<code>calibrated-multigate/theory-foundation.md</code>
Hardware-record inadequacy	<code>calibrated-multigate/hardware-record-audit.json</code>

Main-text result	Frozen result or evidence
QPE thresholds	<code>algorithm-conditions/run-v1/qpe-boundaries.csv</code>
Shor mapping and circuit checks	<code>shor-multiplier-checks.csv</code> , <code>shor-circuits.csv</code> in the same run
qDRIFT robust counts	<code>algorithm-conditions/run-v1/qdrift-condition-summary.csv</code>
qDRIFT realized fidelities	<code>algorithm-conditions/run-v1/qdrift-actual.csv</code>
qDRIFT native costs	<code>algorithm-conditions/run-v1/qdrift-costs.csv</code>
Separate-code algorithm verification	<code>algorithm-conditions/run-v1/independent-verification.json</code>
Separate-code calibration verification	<code>calibrated-multigate/run-v1/independent-verification.json</code>
One-angle design and numbers	<code>condition-experiment/report.md</code> and <code>run-v1</code>
Fixed-arm contrasts	<code>ablation-evaluation/report.md</code> and <code>run-v1.2</code>
Earlier exploratory QPU evidence	<code>ibm-hardware-report/report.md</code> and its raw-data links
New hardware phase/memory study	<code>hardware-validation/run-v1</code> and <code>hardware-validation/report.md</code>
Hardware split/count verification	<code>hardware-validation/run-v1/record-audit.json</code>
Follow-up phase model, randomized memory and costs	<code>hardware-validation-v2/report.md</code> , <code>verification.json</code> and <code>MANIFEST.json</code>
Native memory timing verification	<code>hardware-validation/run-v1/memory/schedule-audit.json</code>

The reproducibility package fixes software versions and seeds and compares regenerated numerical tables with stored references. Its supported workflows cover the QPE minimax calculation, algorithm conditions, synthetic angle calibration, statistical reanalysis of the fixed ablation scores and verification of recorded hardware data. Historical model and prediction directories omitted from the package are not required for score-based statistical reanalysis. The hardware workflow recomputes count summaries and checks saved circuits, likelihoods, session order and resource accounting; it does not acquire new data or rerun all phase-model bootstrap fits. The initial hardware study comprises forty phase jobs and one memory job (S8); the follow-up comprises six successful jobs and one recorded failed attempt (S9).

S8. Hardware phase records, source memory and total preparation cost

S8.1. Independent phase records and chronological evaluation

We executed forty sequential acquisition jobs on `ibm_marrakesh` on 7 October 2026. Each job contained ten commanded angles for each of three representative blocks: a QPE controlled-phase block with a known eigenstate, the first controlled modular-multiplier block of the $N=15$, $a=2$ Shor example with uniform work-register eigenprobe, and a qDRIFT ZZ rotation with a fixed spectator. Independent X and Y measurements were made at repetition factors one and two, with 256 shots per setting. Two additional readout controls were retained in every job. The forty jobs contained 4,880 measurement settings and 1,249,280 shots.

Six jobs supplied sixty training examples per algorithm. A fixed three-feature linear model predicted the residual operational phase from the commanded angle. Twenty-four subsequent jobs calibrated a session-maximum residual enlarged by the reference-phase confidence radius. The final ten jobs supplied one hundred new test records per algorithm, in ten chronological sessions. Predictors were frozen before calibration, and calibration intervals were frozen before test acquisition. Separate acquisition jobs do not establish exchangeability or independence across days.

The reference phase was the argument of the measured quadrature pair. Exact binomial intervals with a simultaneous Bonferroni failure allocation of 0.01 were transformed to circular phase intervals. Session calibration used $\alpha=0.04$ and rank 24 of 24. Conditional on the statistical assumptions, the combined coverage lower bound is 0.95 for one future session of a specified algorithm; it is not simultaneous across all algorithms or all future sessions. Readout controls and repetition consistency are diagnostics and do not bound every SPAM or nonunitary effect. In particular, the Shor eigenprobe does not certify the full entangled checkpoint’s effective angles or its modular-function correctness.

Table S3 summarizes forward-session interval outcomes.

Table S3: Operational phase intervals and later-session containment

Probe	Calibrated half-width (rad)	Later sessions with every reference interval	Unresolved sessions	Disjoint sessions
		contained		
QPE	0.5792	10/10	0/10	0/10
Shor eigenprobe	0.9807	9/10	1/10	0/10
qDRIFT ZZ	0.5117	8/10	2/10	0/10

All point-reference estimates were inside the prediction intervals. This is weaker than knowing every true phase is covered. The containment criterion requires an entire independent test-reference interval to lie inside the prediction interval; unresolved cases retain finite-shot ambiguity. None of 1,200 repetition checks excluded the doubled-phase model using its allocated confidence intervals, but this lack of rejection is not a proof of

the physical model. The relatively broad radii cannot be transferred without validation to multiple arbitrary gates, different physical layouts or entire algorithmic checkpoints.

S8.2. Source waiting and simultaneous outputs

A separate job evaluated ten complete checkpoint cases per algorithm using immediate, time-matched idle, and active two-output preparation arms. The checkpoints used three qubits for QPE, eight for the N=15 Shor example, and two for a ten-step qDRIFT trajectory. B and C were prepared sequentially on disjoint physical register groups. Native circuits restricted routing between registers. The complete non-delay gate sequence on A was identical in all three arms for each case, and no gate acted on A during the B/C preparation interval. These checks establish the intended circuit structure, not hardware isolation.

After each arm, inverse preparation circuits and computational-basis measurement produced a raw all-zero return statistic. There were 512 shots per circuit, ninety circuits and 46,080 shots. The diagnostic inverses and readout are noisy, so these values are not direct state fidelities.

Table S4 summarizes the initial source-memory controls.

Table S4: Initial hardware source-return probabilities

Checkpoint	A immediate return	A matched-idle return	A active-output return	Idle minus immediate
QPE	0.7244	0.2918	0.2934	-0.4326
Shor	0.1840	0.0389	0.0451	-0.1451
qDRIFT	0.7738	0.6285	0.6508	-0.1453

Under independent-shot assumptions, simultaneous Hoeffding intervals for the nine arm means have half-width 0.0240; the corresponding conservative difference intervals have half-width 0.0480. These intervals concern the average shot expectation of the ten fixed cases and do not include new-state generalization or intersession drift. The active-minus-idle differences were 0.0016, 0.0063 and 0.0223; their conservative intervals all included zero. The experiment shows lower source-return statistics in waiting arms, but does not isolate a unique physical error mechanism. Arm-order randomization and day-to-day repetitions were not performed.

The loss in source-return probability rules out treating ideal identity-on-source preservation as an experimentally established property of this implementation. It does not, by itself, determine the fidelity of an arbitrary unknown source or preservation of its external entanglement.

S8.3. Total two-output preparation cost

Table S5 summarizes two-output preparation resources.

Table S5: Hardware resources and joint return for both prepared outputs

Checkpoint	Logical A+B+C qubits	Native B+C entangling gates	Native B+C one-qubit operations	Sequential B+C duration (microsec- onds)	Direct joint BC return
QPE	9	14	67	1.696	0.6928
Shor	24	132–194	368–558	10.976–17.028	0.0451
qDRIFT	6	4–20	36–122	0.656–2.776	0.8131

Three eight-qubit physical groups were reserved for routing; logical counts therefore do not alone describe all reserved hardware. Native entangling gates are not counted as CX gates. Virtual RZ operations are included in one-qubit operation counts. Joint BC return probabilities were obtained from the same-shot joint counts, without a product approximation. Durations were checked by independently reconstructing schedules from native delays, barriers and the saved device gate-length snapshot; the recorded and reconstructed waits agreed exactly in device time units.

The provider billed 240 QPU seconds for the forty metrology jobs and eight seconds for the memory job, for 248 seconds total. The initial 240-second cap was amended to 260 seconds after observing per-job charges, solely to complete the fixed forty-session plan. No circuit, shot count, model, interval or test case changed. The original cap stop and the resource amendment are retained in the provenance record. The account retained 99 seconds after completion, and no credit was purchased or reset.

Local training-refit and inference CPU timings are saved separately. They are replay benchmarks, not measurements of classical-control latency experienced by a physical source. The memory experiment measures the on-device sequential B/C preparation interval only. An end-to-end advantage over prefix replay is not established: these outputs are themselves known-record prefix replays, and the measurement overhead is a separate resource.

S8.4. Verification and reproducibility

Independent compressed-state simulation checked 36 native probe settings and 270 memory-register circuits, with maximum ideal probability differences below $2.8\text{e-}15$ and $1.6\text{e-}15$ respectively. A separate 609-item audit checked chronological splits, frozen hashes, confidence-tail inversion, interval decisions, bit ordering and joint return counts. All 4,880 metrology count settings had the prescribed shot totals. Raw counts, QPY circuits, job IDs, timing, provider usage and code are available in `hardware-validation/run-v1`; the experiment report and limitations are in `hardware-validation/report.md`.

This experiment supports a hardware diagnostic and resource-accounting result. SPAM-independent effective-angle calibration, conditional fidelity certification for arbitrary algorithm checkpoints, and quantum-reference memory preservation remain unestablished. The corresponding limitations in the main manuscript should therefore be retained.

S8.5. Statistical definitions and scope

For each quadrature, let K be the count of outcome zero in $N = 256$ shots. The model assumes independent stationary Bernoulli observations within a setting. A two-sided Clopper–Pearson interval with error allocation a is $[B_{a/2}^{-1}(K, N - K + 1), B_{1-a/2}^{-1}(K + 1, N - K)]$, with endpoints zero or one at the corresponding count extremes [23]. Here B^{-1} is a beta quantile, and $a = 0.01/(40 \cdot 3 \cdot 10 \cdot 2 \cdot 2)$. The equivalent binomial-tail inversion bounds each tail by $a/2$; a union bound over the 4,800 quadratures therefore gives a simultaneous reference event with probability at least 0.99 under the shot model, without assuming independence across settings.

Probability intervals are mapped to expectation intervals by $c = 2p_0 - 1$. Their X/Y Cartesian product defines a rectangle. Its argument image gives a conservative circular phase interval around the empirical argument; if it includes the origin, phase is unidentified and the radius is π . No programmed angle replaces this measured reference. The measured phase is an operational property of the complete probe and readout procedure. Its interval does not include an unmeasured correction for SPAM or for the difference between that procedure and a microscopic gate generator.

For algorithm a , training fixes $\hat{\phi}_a(\theta) = \text{wrap}[\theta + \mathbf{b}_a^\top(1, \sin \theta, \cos \theta)]$. For calibration session i , define $S_{ia} = \max_b \{d_{\text{circ}}(\hat{\phi}_a(\theta_{ib}), \phi_{ib}^{\text{ref}}) + u_{ib}^{\text{ref}}\}$. The prediction radius is $\min(\pi, S_{(24),a})$. The rank proof in S3 applies to the true operational phase only if the acquisition sessions are exchangeable and the reference intervals have their stated meaning. A metrology failure union bound with $\alpha = 0.04$ and $\beta = 0.01$ gives the stated per-algorithm marginal 0.95 bound. The observed ten-session test counts do not establish these assumptions or a population conditional-coverage rate.

Test containment is declared when the complete independent reference interval lies inside the prediction interval. Disjointness requires no intersection; remaining cases are unresolved. A session is contained only if all ten test intervals are contained; any disjoint record makes the session disjoint. This rule is fixed before scoring and retains uncertainty about reference labels. All point estimates being inside is a different and weaker statistic.

For memory diagnostics, let $Y_{aj\ell} \in \{0, 1\}$ denote return in algorithm-arm group a , fixed case j and shot ℓ . The displayed mean pools ten cases and 512 shots each. Independent bounded observations need not have identical probabilities for Hoeffding’s inequality to give

$$\Pr \left(\left| \hat{P}_a - \frac{1}{5120} \sum_{j,\ell} \mathbb{E}[Y_{aj\ell}] \right| > r \right) \leq 2 \exp(-2 \cdot 5120 r^2).$$

A union bound over nine groups with total error 0.05 yields $r = \sqrt{\log(18/0.05)/(2 \cdot 5120)} = 0.0239753$ [24]. Differences of two groups have conservative half-width $2r = 0.0479506$. Their coverage requires the shot assumptions and concerns fixed-case means, not device drift, new-state variation, randomized causal effects or fidelity. The new hardware data neither justify substituting these returns for the pure-state overlap in Eq. (5) nor establish preservation of source–reference entanglement.

The confidence calculations use the explicit tail-inversion and bounded-variable formulas above and were checked against the raw counts using a separate implementation.

S9. Phase-model validation, randomized controls and workflow costs

S9.1. Measurement model and its falsification

We performed a further six-job experiment on IBM `ibm_marrakesh`, obtaining 293,376 shots across 978 measurement settings. The experiment separates three questions: whether a phase estimate remains meaningful after fitting fixed SPAM nuisance parameters and attenuation, whether source waiting results survive balanced randomized submission order, and how the classical/cloud workflow compares with on-chip preparation time. None of these measurements certifies cloning fidelity for arbitrary unknown inputs.

For three commanded angles (0.35, 0.75, 1.15 rad) in each QPE, Shor and qDRIFT-ZZ probe context, we measured X/Y quadratures at repetition counts 0, 1, 2, 3, 4, 6, 8, 12. Three acquisition rounds provided 1,536 shots per setting. We fitted

$$E_a(m) = b_a + (1 - |b_a|)A_a e^{-\gamma m} \cos(m\phi + \xi_a),$$

using repetitions 0, 1, 2, 4, 6, 12 and reserved 3,8 for prediction checks. The six axis-dependent nuisance parameters describe fixed offsets, contrasts and reference phases; phi and gamma describe a single damped oscillatory mode. We used 999 local parametric bootstrap replicates and Holm adjustment across 27 training-fit, held-out and between-round homogeneity diagnostics. This model does not identify preparation and measurement errors separately. Supplement S10.3 gives an explicit ambiguity in their observable product and a coherent two-frequency counterexample to interpreting attenuation as a single exponential rate. The fitted attenuation parameter is therefore not an identified general measure of channel nonunitarity; no particular hardware mechanism is established by these diagnostics.

The single-mode model was rejected for all three QPE and all three Shor command settings. It was not rejected for the three qDRIFT-ZZ settings. For qDRIFT-ZZ, the fitted phase magnitudes were 0.3823, 0.7625 and 1.1484 rad, with local model-dependent 95% intervals [0.3728,0.3927],[0.7560,0.7691] and [1.1426,1.1542] rad. The corresponding damping exponents were 0.0254, 0.0106 and 0.0145 per block. These intervals are not jointly certified physical-angle errors and cannot be inserted directly into an arbitrary-state reconstruction fidelity bound.

Table S6 summarizes the fitted phase-model diagnostics.

Table S6: Single-mode parameter estimates and model-rejection decisions

Probe	Command (rad)	Fitted magnitude (rad)	Damping exponent	Held-out p	Model rejected
qpe	0.35	0.3570	0.0848	0.001	Yes

Probe	Command (rad)	Fitted magnitude (rad)	Damping exponent	Held-out p	Model rejected
qpe	0.75	0.6899	0.0873	0.001	Yes
qpe	1.15	1.1282	0.0835	0.160	Yes
shor	0.35	0.5918	0.5292	0.001	Yes
shor	0.75	0.7973	0.6822	0.001	Yes
shor	1.15	1.2444	0.6886	0.001	Yes
qdrift_zz	0.35	0.3823	0.0254	0.694	No
qdrift_zz	0.75	0.7625	0.0106	0.136	No
qdrift_zz	1.15	1.1484	0.0145	0.579	No

The displayed p-values are unadjusted Monte Carlo values. Rejection combines all 27 adjusted diagnostics, including training goodness of fit; therefore a held-out p-value alone does not determine the last column. The fitted values for rejected models are descriptive and are not physical angle certificates. Bootstrap intervals are local, model-dependent percentile summaries, not uniform finite-sample coverage guarantees. The initial 199-replicate plan was increased to 999 before inspecting counts because a minimum Monte Carlo p-value of $1/200$ could not resolve the $0.05/27$ Holm threshold. This pre-analysis numerical-resolution amendment did not change quantum resources or splits.

A native-instruction audit, completed before inspecting hardware outcomes, found that Shor’s first routed block differed from subsequent blocks. We retained the original analysis and added a separately documented warm-up model in which the first block became part of preparation and the repeat index was shifted to $m-1$. All three supplementary Shor fits also failed the stated model diagnostics. This result limits the use of a single calibrated angle as a summary of the hardware error. It does not establish that a particular alternative noise mechanism is responsible.

S9.2. Randomized submitted-order waiting controls

For source waiting, each of ten fixed checkpoints per algorithm received all six permutations of the immediate, time-matched idle and active B/C-preparation arms, once each, at 128 shots per setting. Case order was randomized within each block. The provider did not expose the actual per-shot execution order; randomization is therefore established for submitted PUB order. Mean raw source inverse-return probabilities were 0.6794/0.2154/0.2178 for QPE, 0.1413/0.0408/0.0388 for Shor and 0.7724/0.6029/0.6099 for qDRIFT (immediate/idle/active). Active-minus-idle differences were 0.0025, -0.0020 and 0.0070, respectively. All three simultaneous 95% Hoeffding intervals included zero, with halfwidth 0.0267 under independent-shot assumptions. This is an unresolved contrast, not an equivalence or no-crosstalk result. The waiting decrement remained visible in all three algorithms.

Joint BC and ABC returns were counted directly rather than inferred from products. In the active arm, BC/ABC returns were 0.7647/0.1673 (QPE), 0.0427/0.0013 (Shor) and 0.9078/0.5600 (qDRIFT). These observables include inverse-circuit and SPAM errors; they are not state fidelities. Independent schedule reconstruction confirmed

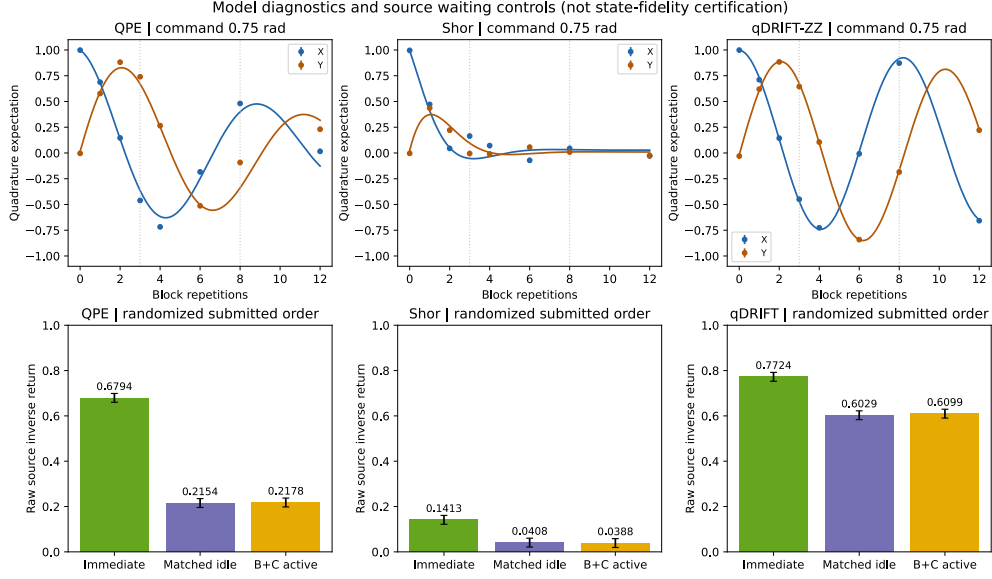


Fig. S1 Follow-up model and memory diagnostics. Top: quadrature expectations at a commanded angle of 0.75 radians, with binomial standard-error bars and fitted single-mode curves. Vertical dotted lines mark held-out repetition counts 3 and 8. The QPE and Shor models are rejected; their curves are descriptive fits. Bottom: source inverse-return means for the balanced randomized submitted orders, with simultaneous nine-group Hoeffding error bars under independent shots. Error bars exclude drift and new-state variation; returns are not state fidelity.

identical source non-delay gate sequences across arms and delay-only source evolution during output preparation.

S9.3. CPU, communication and acquisition resources

The successful six-job acquisition took 220.683 s of wall time and 4.258 s of client process CPU time, with 54 charged QPU seconds. A confirmed initial timeout consumed 9 additional charged seconds and returned no counts; total charge was 63 s. Its provider queue and running intervals were 318.533 s and 35.503 s, respectively. Client RPC/CPU segments for that failed attempt were not flushed and remain missing. The per-job runtime ceiling was raised in a documented recovery amendment, while the overall 72 charged-second cap, including failure, and 20 second reserve were retained.

In a separate local replay, construction, native compilation and serialization of both outputs required mean CPU times 0.023497 s (QPE), 0.030736 s (Shor) and 0.015370 s (qDRIFT) per case. Corresponding wall times were 0.015854, 0.021083 and 0.011220 s. These are timed classical replays, not fresh QPU submissions or online feedback. The actual on-chip B+C preparation intervals remained 1.696 μ s, 10.976–17.028 μ s and 0.656–2.776 μ s, respectively. Native two-qubit counts were 14, 132–194 and 4–20. Record I/O, OLS replay/inference, serialization, submission, polling, retrieval, provider queue/run intervals and offline diagnostic fitting were recorded separately. Provider and client intervals overlap and are not summed as disjoint costs.

Across the final fitting runs, measured phase preparation, offline B/C compilation replay and successful acquisition, recorded process CPU components summed to 52.526 s. Successful-job RPC wall intervals summed to 32.159 s and provider queue intervals to 19.191 s; both overlap the acquisition interval. These are not additive wall-time costs, and missing failed-attempt CPU, earlier discarded analyses and unmeasured development costs preclude a complete project-wide total.

In this implementation, no source was retained across separate IBM jobs. Consequently, cloud turnaround was measured as a workflow cost, not as a physical source-memory residence time. A protocol requiring an already-arrived source to survive classical compilation or a cloud round trip would require a demonstrated memory/controller path or advance preparation from the available classical record. The present results support explicit resource accounting and model-failure diagnosis, but not full SPAM/nonunitarity separation or simultaneous arbitrary-input high-fidelity preservation and replication.

S9.4. Numerical correction and provenance

The original polar-parameter optimizer left one QPE fit numerically improvable. An independent coefficient-based likelihood audit detected this; all data/bootstrap fits were rerun with an equivalent smooth constrained-coefficient refinement. Original outputs and the correction record are retained. The rejection pattern did not change. The independent arithmetic/order/hash audit comprised 1,307 checks, with binomial-deviance disagreement below 2.1×10^{-12} . The optimization checks are not a certificate of global nonconvex optimality.

All raw data, circuit and order definitions, original and corrected analyses, amendments, figures and timing records are provided in hardware-validation-v 2. Previous paper v 1.1 and its underlying experimental records are retained.

S9.5. Confidence calculations and accounting boundaries

For each algorithm and contrast, two arm means contain $N = 7680$ independent bounded Bernoulli observations. The sum of squared coefficient ranges for their difference is $2/N$. Thus Hoeffding’s inequality gives $\Pr(|\hat{\Delta} - \mathbb{E}\hat{\Delta}| > r) \leq 2e^{-Nr^2}$. A union bound over six contrasts sets $r = \sqrt{\log(12/0.05)/7680} = 0.0267137579$. This concerns the shot-expectation average of the fixed cases, not a population of new states or drift-robust causal attribution. Nine arm-mean error bars in Supplementary Fig. 1 instead use $\sqrt{\log(18/0.05)/(2N)}$. Readout controls contain both preparation and measurement errors and were not inverted as an independently calibrated measurement channel.

The primary stationary model pools three rounds. Between-round homogeneity diagnostics compare each of sixteen quadrature settings across three rounds, with 32 asymptotic degrees of freedom. The secondary Shor model uses fourteen settings and 28 degrees of freedom. Its nine training, held-out and homogeneity diagnostics form a separate Holm family. All three secondary settings were rejected; the revised phase magnitudes were 1.6200, 1.7612 and 2.1151 radians. These values illustrate model dependence, not a recovered intrinsic phase. Only ten observations fit eight parameters

in the secondary model. Neither short-timescale homogeneity nor failed rejection would establish exchangeability across days.

Wall-clock categories overlap: submission and retrieval are included in acquisition; provider execution overlaps client polling and waiting; the two fitting processes ran concurrently. CPU components in the cost ledger are separately measured process times, and classical replay is an offline benchmark. Costs of previously acquired training data, missing failed-attempt client intervals, interpreter startup, development and discarded analysis attempts prevent an exact project-wide total. No state remains live across the measured cloud jobs. Preparation of calibration probes consumes independently prepared specimens, not the preserved test source.

The acquisition/source freeze, three design/resource amendments and numerical correction are separately retained. `verification.json` contains 1,307 arithmetic, order and hash checks. A second coefficient-based optimizer reported nonconvergence for one final boundary case, although its deviance improvement was below 2×10^{-5} ; no global nonconvex optimum is certified. Both native simulation and a physically valid two-mode counterexample were used to challenge the model. Reported diagnostics therefore delimit the current inference model rather than certify complete SPAM removal or general channel nonunitarity. Supplement S10.3 gives elementary identifiability examples without attributing these mechanisms to the measured device.

S10. Resource comparisons, counterexamples and evidence scope

S10.1. What is and is not being compared

Table S7 summarizes the resource models of the comparison methods.

Table S7: Operation classes, supported conclusions and comparison limits

Operation or baseline	Information and source access	Supported conclusion	Comparison limit
Universal one-copy cloner	One arbitrary pure source; no informative preparation record; interaction with source allowed	Optimal symmetric single-output fidelity $(d+3)/[2(d+1)]$	Extra records change the resource model [1, 2]
Exact supplementary-information cloning	Finite pure family with no orthogonal pairs; supplementary state indexed by input	An extra copy must be producible from the supplement alone	Exact deterministic theorem, with an orthogonal-pair exception; not an approximate bound [3, Theorem 1 and following remark]

Operation or baseline	Information and source access	Supported conclusion	Comparison limit
Variational quantum cloning	Quantum input, blank receiving registers and optional ancilla; family-dependent trained circuit	Optimizes circuit parameters and structure with distinct local or global costs	Interacts with input; no same-resource superiority is established here [8, Secs. II and IV, Appendix B.1]
Encrypted cloning	Unknown qubit coupled to signals initially entangled with quantum keys; no preparation record required	One chosen signal can be perfectly recovered using all keys	Encoding changes the source; recovery is not two simultaneous independent copies [9, pp. 010801-2–010801-3, Eqs. (1)–(9)]
Record-only reconstruction	Classical record available at execution; no access to retained source	Known pure-ensemble average optimum and record-dependent robust bounds	Synthesis cost and posterior accuracy remain separate; no-record minimax is $1/d$ [5]; Supplement S1–S2
Known-prefix replay	Correct initial-state and circuit information, including recoverable Shor multipliers	Exact ideal replay when the full preparation is available	Two output preparations and record acquisition still cost resources [11, Sec. 3]
qDRIFT completion	Known term distribution and independent missing draws	Conditional mean channel and ensemble-adapted preparation	A new draw is not the lost realized trajectory [12, Eqs. (4)–(6)]

The Shor comparison must first remove ambiguity that can be resolved by classical repeated squaring. The qDRIFT comparison must include known-probability completion, since inference cannot reveal an independent hidden draw from unrelated visible draws. Generic state synthesis and structured replay also have different costs [6, 7]. None of these comparisons supplies a uniform ranking of learners under all budgets or state families. The resource-controlled ablation supports only the stated paired contrasts, including all fallback outputs.

For encrypted cloning, the signal registers alone are maximally mixed. In the published two-signal construction, the combined encoding and selected decoding operation is a SWAP between the input and the selected signal [9, p. 010801-3, Eq. (8)]. Applied to the initial Bell pairs, it leaves the selected signal in the input state and the remaining registers in input-independent Bell pairs. Thus the same encoded instance cannot yield a second simultaneously recovered perfect copy. The key qubits need not be destroyed; it is the correlations needed to recover another copy of that input that are unavailable after this decoding. The original source is not retained in its input state by this construction (see also the End Matter, p. 010801-6). Our ideal channel instead retains the source and prepares two approximate outputs from

classical records. Neither this comparison nor the low local overlap of an encrypted signal establishes a same-resource fidelity advantage.

S10.2. Counterexamples to transferring the average decision rule

The maximal-eigenstate rule concerns an exactly known conditional ensemble of pure targets and unrestricted output preparation [5, Sec. 3, Eqs. (13)–(16)]. For a fixed mixed target $\rho = I/2$, preparing $I/2$ gives squared Uhlmann fidelity one, whereas any pure preparation gives fidelity $1/2$. Thus selecting an eigenvector cannot be a universal mixed-target decision rule.

Even for pure targets, the mean single-output and mean joint-output objectives can disagree. Consider targets $|0\rangle$ and $|+\rangle$ with probabilities $3/5$ and $2/5$. Their mean density matrix has Bloch vector $(2/5, 0, 3/5)$, so its maximal eigenstate has unit Bloch vector $\mathbf{n} = (2, 0, 3)/\sqrt{13}$. For two independently prepared copies of the chosen output state, the average joint score is

$$G(\mathbf{n}) = \frac{3}{5} \left(\frac{1 + n_z}{2} \right)^2 + \frac{2}{5} \left(\frac{1 + n_x}{2} \right)^2.$$

On the unit-circle path with tangent $\mathbf{n}' = (n_z, 0, -n_x)$ at this point, differentiation gives $G' = -3/65 \neq 0$. The single-output maximizer is therefore not even stationary for this joint objective. This example establishes objective mismatch, without asserting a general formula for the joint optimum. Likewise, a learned $\hat{R}_x$ need not equal the true conditional mean R_x , and a native preparation budget may exclude the chosen eigenstate entirely.

S10.3. Why a fringe fit does not separate physical error sources

A simple qubit model already has a preparation/measurement ambiguity. Let the prepared Bloch vector be $r(\cos\theta, \sin\theta, 0)$ and let the measured quadratures have observables vX and vY , where $0 \leq r, v \leq 1$. The expectations are $rv \cos\theta$ and $rv \sin\theta$. Any replacement $r \mapsto cr$, $v \mapsto v/c$ remaining in the physical interval leaves both observations unchanged. Without an independent constraint, contrast alone cannot separate preparation purity from measurement visibility. This is a counterexample to identifiability from these data, not a general gate-set characterization. Gate-set tomography treats preparation, gates and measurements self-consistently; gauge transformations can nevertheless leave all circuit probabilities unchanged [27, Secs. 2.1.4 and 3.1, Eq. (34)]. Our eight-parameter fringe model does not implement that reconstruction and cannot inherit a claim of unique physical-error separation from it.

A coherent two-frequency signal supplies a second caution:

$$\frac{1}{2} \cos[m(\phi + \delta)] + \frac{1}{2} \cos[m(\phi - \delta)] = \cos(m\phi) \cos(m\delta).$$

Controlled unitary evolution with an unobserved two-level register can produce this signal. Its envelope can decrease on a finite measured range even though it is not a stationary exponential. Fitting $e^{-\gamma m}$ over that range therefore does not uniquely identify a physical nonunitary error rate. The example does not identify the cause of the observed device errors, and it does not imply that real preparation, measurement,

leakage or drift errors are absent. A physical angle or channel estimate would require additional experimentally justified assumptions and measurements.

S10.4. Statistical guarantees used in the diagnostics

Binomial reference intervals. Under $K \sim \text{Binomial}(N, p)$, inversion of each one-sided tail test at level $a/2$ gives an interval with noncoverage at most a [23, Sec. 1, pp. 406–407]. The lower endpoint solves $\Pr_{p_L}(K' \geq k) = a/2$ for $k > 0$; the upper solves $\Pr_{p_U}(K' \leq k) = a/2$ for $k < N$, with endpoints zero and one in the extreme cases. The binomial/beta identities give the quantiles in S8.5. Discreteness makes the coverage conservative, not necessarily exactly $1 - a$. Summing the allocated errors over the planned quadratures gives the simultaneous reference event; no independence across settings is required for that union bound. The stationary independent-shot binomial premise within each setting is still needed.

Bounded-shot intervals. A short derivation makes the assumptions behind the named inequality explicit [24, Sec. 2, Theorem 2, Eq. (2.6), and two-sample corollary Eq. (2.7)]. For $Z \in [a, b]$, define $K(t) = \log \mathbb{E} e^{t(Z - \mathbb{E}Z)}$. Under exponential tilting, $K''(t)$ is the variance of a variable still supported on $[a, b]$, hence at most $(b - a)^2/4$. Since $K(0) = K'(0) = 0$, integration gives $K(t) \leq t^2(b - a)^2/8$. Independence allows these cumulant bounds to add. For $W = \sum_i w_i(Z_i - \mathbb{E}Z_i)$, Chernoff optimization on both tails yields

$$\Pr\{|W| \geq r\} \leq 2 \exp \left[- \frac{2r^2}{\sum_i w_i^2 (b_i - a_i)^2} \right].$$

A Bernoulli group mean of N shots has denominator $1/N$; a difference of two disjoint independent N -shot means has denominator $2/N$. Union bounds over nine means or six contrasts then give exactly the radii in S8 and S9. Nonidentical shot probabilities are allowed, but dependence or unmodeled drift can invalidate the stated product-moment argument. The intervals concern the fixed-case shot-expectation averages. Their inclusion of zero is not an equivalence test and does not establish absence of crosstalk.

Step-down multiplicity and bootstrap limits. We use Holm’s sequentially rejective Bonferroni rule [26, Sec. 2, pp. 66–67]. For H hypotheses, sort the p-values increasingly and compare $p_{(j)}$ with $\alpha/(H - j + 1)$, stopping at the first failure. If r null hypotheses are true, the first rejected true null, if any, has threshold at most α/r . A union bound over those true nulls gives familywise error at most α , provided each null p-value is super-uniform. Dependence among valid p-values does not invalidate this argument. The fitted-model bootstrap diagnostics simulate from the fitted model and refit the simulated samples before recomputing discrepancies [25, Sec. 1.1, p. 1097]. The cited analysis supplies asymptotic validity under its smoothness, identifiability and estimator conditions (Sec. 2, pp. 1100–1102); those conditions have not been established for our constrained, nonconvex phase-model fits. The diagnostics here use fitted-model parametric-bootstrap p-values; finite samples, nuisance estimation, boundaries and numerical fitting mean that their super-uniformity is not certified. The step-down rule does not repair those limitations. Local percentile bootstrap intervals

and the 999-replicate Monte Carlo resolution should consequently be interpreted as model-dependent diagnostics, not uniform finite-sample coverage guarantees.

S10.5. Scope of cited results

The cited literature establishes the universal-cloning benchmarks, reconstruction decision rule, algorithm checkpoints, sampling law, synthesis methods and conformal ranking framework. The angle-sum bound, record-specific checkpoint conditions and empirical values reported here are supported by the derivations and data in this study. The encrypted-cloning comparison in S10.1 concerns the published two-signal construction and its specified decoding operation.

The binomial, bounded-variable and multiplicity arguments retain the sampling and validity assumptions stated in S10.4. The bootstrap reference does not establish validity of the p-values for the constrained phase-model fits used here. Numerical agreement and record consistency do not certify hardware state fidelity or general preservation of source-reference correlations. No claim of first priority is made.

S11. Exact minimax reconstruction on a short QPE phase interval

S11.1. Scope and exact value

Let $D = 2^r$, where $r \geq 1$ low-order controlled powers have been completed. The eigenstate and the unused counting factors must be known and preparable. The record specifies a center phase and a half-width h cycles. There is no additional power-specific phase error. After removing the known center and choosing a global phase, the uncertain factor is

$$|u(t)\rangle = D^{-1/2} \sum_{k=0}^{D-1} e^{2\pi i q_k t} |k\rangle, \quad q_k = k - (D-1)/2, \quad |t| \leq h.$$

Write $K_D(t) = \sin(\pi D t) / [D \sin(\pi t)]$, with the removable value $K_D(0) = 1$. For $0 \leq h \leq 1/(2D)$, the exact record-only minimax single-output fidelity is

$$\max_{\sigma \succeq 0, \text{Tr } \sigma = 1} \inf_{|t| \leq h} \langle u(t) | \sigma | u(t) \rangle = \lambda_h = \frac{1 + K_D(2h)}{2}.$$

The maximizing state is

$$|v_h\rangle = \frac{|u(h)\rangle + |u(-h)\rangle}{\sqrt{2[1 + K_D(2h)]}}, \quad \langle k | v_h \rangle = \sqrt{\frac{2}{D[1 + K_D(2h)]}} \cos(2\pi q_k h).$$

The optimization permits arbitrary mixed output states. Known additional factors do not change its value: taking their target matrix element in any full output produces a positive operator of trace at most one on the uncertain factor, to which the same spectral bound applies. Tensoring the optimizer with the known factors attains the bound. A common inverse QFT also leaves the result unchanged.

S11.2. Upper bound and attainment throughout the interval

Assign probability $1/2$ to each endpoint, and let R_h be their average density matrix. For every output σ ,

$$\inf_{|t| \leq h} \langle u(t) | \sigma | u(t) \rangle \leq \text{Tr}(R_h \sigma) \leq \lambda_{\max}(R_h).$$

In the stated interval, $K_D(2h) \geq 0$, so the nonzero eigenvalues of R_h are $[1 \pm K_D(2h)]/2$. Thus $\lambda_h I - R_h \succeq 0$ is an explicit dual upper-bound certificate for the continuous problem. The spectral optimization is the established reconstruction ingredient [5]; the remaining issue is whether its endpoint optimizer also satisfies every interior constraint.

The spectrum is symmetric under $q_k \mapsto -q_k$. Consequently the overlap $\langle u(t) | v_h \rangle$ is a positive normalization factor times

$$A(t) = \sum_k \cos(2\pi q_k h) \cos(2\pi q_k t).$$

For $0 \leq t \leq h \leq 1/(2D)$, each coefficient is nonnegative, and each cosine factor is nonnegative and nonincreasing because $2\pi|q_k|h < \pi/2$. Therefore $A(t)$ is nonnegative and nonincreasing. Evenness handles negative t . Squared overlap is minimized at the endpoints, where it equals λ_h , proving attainment. Equivalently,

$$F_v(t) = \frac{[K_D(h+t) + K_D(h-t)]^2}{2[1 + K_D(2h)]}.$$

This is an analytic continuous-domain argument, independent of the finite numerical grid used for implementation checks. For $0 < h < 1/(2D)$, the top eigenvalue of R_h is nondegenerate, so the optimal density operator is unique. At the boundary the statement guarantees an optimizer without asserting uniqueness.

S11.3. Exact target condition and resources

For the fixed terminal depolarizing model on the full d -dimensional register,

$$F_\epsilon^*(h) = (1 - \epsilon)\lambda_h + \epsilon/d.$$

For a target between $(1 - \epsilon)/2 + \epsilon/d$ and $1 - \epsilon + \epsilon/d$, there is a unique boundary h_* in the short interval. The target is exceeded if and only if $h < h_*$. Strict monotonicity follows by differentiating $K_D(2h) = D^{-1} \sum_k \cos(4\pi q_k h)$: every nonzero-frequency contribution has negative derivative in the open interval. Any wider interval contains the boundary endpoints and therefore also excludes strict target excess. This establishes the threshold even though the complete optimum for wider intervals has not been derived.

For $r = m = 8$, $d = 512$, $\epsilon = 0.025$ and target $5/6$, $h_* \simeq 0.000861540576$ cycles. The center-replay boundary is approximately 0.000846794041 cycles; the optimizer increases the tolerated half-width by 1.74145% . The gain concerns tolerated record uncertainty; it is not a 1.74145% fidelity increase or a universal cloning advantage. For $r = 1$ the optimizing state is the center state, and there is no gain.

A finite deterministic preparation of the known vector v_h exists by ordinary state synthesis. A heralded alternative uses one flag and r data qubits. Initialize all in $|+\rangle$. Apply $P(-2\pi 2^j h)$ to data qubit j , $CP(4\pi 2^j h)$ from the flag to that qubit, and $P[-2\pi(D-1)h]$ to the flag. After a final flag Hadamard, outcome zero prepares v_h up to a global phase, with probability $p = \lambda_h \geq 1/2$. Outcome one prepares the normalized antisymmetric endpoint combination, with probability $1 - p$; at $h = 0$ it never occurs. The record center is restored using known phase gates. The explicit relative flag phase is necessary because the endpoint states were globally phase-aligned in the proof.

This circuit requires a flag connected to every data qubit, the gate set H, P, CP , and flag measurement. Its sequential depth is $O(r)$ and it uses r controlled-phase gates. It prepares fresh known candidates; it never filters or measures the unknown retained source. A failed blank can be reset and retried, with all attempts counted. The probability is independent of the true t . Unlimited retries terminate almost surely but are not a finite worst-case depth implementation. With at most L attempts for each of two outputs, joint success probability is $[1 - (1 - p)^L]^2$, and the conditional output is $|v_h\rangle\langle v_h|^{\otimes 2}$. Expected total attempts without truncation are $2/p$. Waiting for the second output, reset, measurement and classical feedback have costs.

At the eight-qubit threshold, $p = 0.854650775$, expected attempts for two outputs are 2.340137, and at most four attempts per output give joint success probability 0.999107550 in the ideal independent-attempt model. This flag probability is a preparation probability, not an input-dependent probability of correctly identifying the unknown phase.

Table S8: Local circuit synthesis counts at the respective noisy target boundaries, with unrestricted connectivity and basis rz, sx, x, cx. These are counts for the demonstrated implementations, not optimal complexity bounds. Counts exclude measurement, reset, known eigenstate preparation, unused factors, any inverse QFT, routing and physical noise.

Counting qubits r	Flag circuit CX per attempt	Generic deterministic synthesis CX	Flag circuit depth
1	2	0	9
2	4	1	13
3	6	4	17
5	10	26	25
8	16	247	37

The deterministic operation on the retained source and outputs is $\rho_A \mapsto \rho_A \otimes |v_h\rangle\langle v_h|_B \otimes |v_h\rangle\langle v_h|_C$. Both marginals attain the optimum simultaneously. This product construction has global fidelity $F_v(t)^2$; no optimality claim is made for an independently optimized correlated global-output objective. Physical source waiting remains separate from the ideal identity map.

S11.4. Verification and novelty scope

The saved computation checks the trigonometric profile, endpoint minimum, finite-dimensional dual positivity and both flag branches, and verifies compiled circuits

against statevectors. An independent implementation uses tensor products of qubit phase states, endpoint eigendecomposition, 70-digit product-form root calculations, and the saved compiled circuits. A separate source-reference Bell-pair calculation verifies that both heralded outputs can be generated while preserving source-reference correlations in the ideal circuit. The continuous guarantee follows from the proof, not sampling. No new hardware experiment was performed.

This result sharpens the QPE reconstruction conditions under a specified record-only model. It is distinct from optimizing a measurement of a quantum probe with a prior phase distribution [17]. The underlying spectral method and phase-state formula are established. The analysis does not establish priority. Additional power-specific errors, an unknown unpreparable eigenstate, general hardware noise and optimal synthesis under restricted connectivity remain outside the theorem.

References

- [1] Werner, R.F.: Optimal cloning of pure states. *Physical Review A* **58**(3), 1827–1832 (1998) <https://doi.org/10.1103/physreva.58.1827> [arXiv:quant-ph/9804001](https://arxiv.org/abs/quant-ph/9804001)
- [2] Keyl, M., Werner, R.F.: Optimal cloning of pure states, testing single clones. *Journal of Mathematical Physics* **40**(7), 3283–3299 (1999) <https://doi.org/10.1063/1.532887> [arXiv:quant-ph/9807010](https://arxiv.org/abs/quant-ph/9807010)
- [3] Jozsa, R.: A stronger no-cloning theorem (2002). <https://arxiv.org/abs/quant-ph/0204153>
- [4] Koashi, M., Imoto, N.: Operations that do not disturb partially known quantum states. *Physical Review A* **66**(2), 022318 (2002) <https://doi.org/10.1103/physreva.66.022318> [arXiv:quant-ph/0101144](https://arxiv.org/abs/quant-ph/0101144)
- [5] Fuchs, C.A., Sasaki, M.: Squeezing quantum information through a classical channel: Measuring the “Quantumness” of a set of quantum states. *Quantum Information and Computation* **3**(5), 377–404 (2003) <https://doi.org/10.26421/qic3.5-1> [arXiv:quant-ph/0302092](https://arxiv.org/abs/quant-ph/0302092)
- [6] Möttönen, M., Vartiainen, J.J., Bergholm, V., Salomaa, M.M.: Transformation of quantum states using uniformly controlled rotations. *Quantum Information and Computation* **5**(6), 467–473 (2005) <https://doi.org/10.26421/qic5.6-5> [arXiv:quant-ph/0407010](https://arxiv.org/abs/quant-ph/0407010)
- [7] Shende, V.V., Bullock, S.S., Markov, I.L.: Synthesis of quantum-logic circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **25**(6), 1000–1010 (2006) <https://doi.org/10.1109/tcad.2005.855930> [arXiv:quant-ph/0406176](https://arxiv.org/abs/quant-ph/0406176)
- [8] Coyle, B., Doosti, M., Kashefi, E., Kumar, N.: Progress toward practical quantum cryptanalysis by variational quantum cloning. *Physical Review A* **105**, 042604 (2022) <https://doi.org/10.1103/PhysRevA.105.042604>

- [9] Yamaguchi, K., Kempf, A.: Encrypted qubits can be cloned. *Physical Review Letters* **136**, 010801 (2026) <https://doi.org/10.1103/y4y1-1ll6>
- [10] Cleve, R., Ekert, A., Macchiavello, C., Mosca, M.: Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **454**(1969), 339–354 (1998) <https://doi.org/10.1098/rspa.1998.0164> [arXiv:quant-ph/9708016](https://arxiv.org/abs/quant-ph/9708016)
- [11] Shor, P.W.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing* **26**(5), 1484–1509 (1997) <https://doi.org/10.1137/s0097539795293172> [arXiv:quant-ph/9508027](https://arxiv.org/abs/quant-ph/9508027)
- [12] Campbell, E.: Random compiler for fast hamiltonian simulation. *Physical Review Letters* **123**(7), 070503 (2019) <https://doi.org/10.1103/physrevlett.123.070503> [arXiv:1811.08017](https://arxiv.org/abs/1811.08017)
- [13] Gilchrist, A., Langford, N.K., Nielsen, M.A.: Distance measures to compare real and ideal quantum processes. *Physical Review A* **71**, 062310 (2005) <https://doi.org/10.1103/PhysRevA.71.062310>
- [14] Barnum, H., Caves, C.M., Fuchs, C.A., Jozsa, R., Schumacher, B.: Noncommuting mixed states cannot be broadcast. *Physical Review Letters* **76**(15), 2818–2821 (1996) <https://doi.org/10.1103/physrevlett.76.2818> [arXiv:quant-ph/9511010](https://arxiv.org/abs/quant-ph/9511010)
- [15] Angelopoulos, A.N., Bates, S.: Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning* **16**(4), 494–591 (2023) <https://doi.org/10.1561/22000000101> [arXiv:2107.07511](https://arxiv.org/abs/2107.07511)
- [16] Kitaev, A.Y.: Quantum measurements and the Abelian Stabilizer Problem (1995). <https://arxiv.org/abs/quant-ph/9511026>
- [17] Demkowicz-Dobrzański, R.: Optimal phase estimation with arbitrary a priori knowledge. *Physical Review A* **83**, 061802 (2011) <https://doi.org/10.1103/PhysRevA.83.061802>
- [18] Cawley, G.C., Talbot, N.L.C.: On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research* **11**(70), 2079–2107 (2010)
- [19] Park, S., Simeone, O.: Quantum Conformal Prediction for Reliable Uncertainty Quantification in Quantum Machine Learning. Preprint, version 3 (2023). <https://arxiv.org/abs/2304.03398v3>
- [20] Aditi, K., Becker, S.: Uncertainty Quantification for Quantum Fidelity Estimation. Preprint, version 1, 7 October 2026 (2026). <https://arxiv.org/abs/2610.10885>

- [21] Zambrano, L., Farina, D., Pagliaro, E., Taddei, M.M., Acín, A.: Certification of quantum state functions under partial information. *Quantum* **8**, 1442 (2024) <https://doi.org/10.22331/q-2024-08-16-1442>
- [22] Guan, Q., Cao, Q., Lu, X., Niu, S.: Architectural Foundations for Checkpointing and Restoration in Quantum HPC Systems. Preprint, version 1 (2026). <https://arxiv.org/abs/2602.09325v1>
- [23] Clopper, C.J., Pearson, E.S.: The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**(4), 404–413 (1934) <https://doi.org/10.1093/biomet/26.4.404>
- [24] Hoeffding, W.: Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**(301), 13–30 (1963) <https://doi.org/10.1080/01621459.1963.10500830>
- [25] Genest, C., Rémillard, B.: Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* **44**(6), 1096–1127 (2008) <https://doi.org/10.1214/07-AIHP148>
- [26] Holm, S.: A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**(2), 65–70 (1979)
- [27] Nielsen, E., Gamble, J.K., Rudinger, K., Scholten, T., Young, K., Blume-Kohout, R.: Gate set tomography. *Quantum* **5**, 557 (2021) <https://doi.org/10.22331/q-2021-10-05-557>