

Posterior Algebra of Binary Quantum Measurements: From Shor's Orthogonal-Measurement Conjecture to Information–Error Pareto Frontiers

Jinbo Wang,^{1,*} Qihang Wang,^{1,*} and Kun Chen^{2,†}

¹*School of Mathematical Sciences, Peking University, Beijing 100871, China*

²*Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

We introduce a posterior algebra that characterizes when an affine information bound is attainable and classifies all optimal finite measurements in that regime. Binary ensembles yield a one-generator proof of Shor's orthogonal-measurement statement and a certified receiver. For binary qubits, a posterior moment formulation reveals three-outcome advantages between projective information and minimum-error endpoints. A decoder-derived posterior loss guides readout optimization in a synthetic repetition code, where recovery can improve without improving local information–error scores.

Introduction—For two known quantum states, a measurement can minimize the probability of misidentifying the state or maximize mutual information between the preparation label and the outcome. Both objectives admit an optimal projective measurement, but their optimal bases can differ. We study combined objectives and the use of measurements for error correction.

The maximum mutual information over all measurements is called the accessible information. Generalized measurements can outperform projective measurements for larger signal alphabets [1, 2]. Related early work compared local nonorthogonal and joint detection strategies [3]. For binary ensembles, projective sufficiency is the statement associated with Shor's orthogonal-measurement conjecture. Early results treated pure signals and information bounds [4, 5]; Keil proved the qubit case [6]. The general measured-divergence theorem of Fang, Fawzi, and Fawzi implies the finite-dimensional binary result [7]. Explicit binary reductions and convex formulations are also given by Jo and Dall'Arno [8, 9], including a spectral replacement for each finite POVM.[10]

In this letter, we characterize information-optimal measurements using operator analogues of posterior-to-prior probability ratios. The posterior is the probability of a preparation label conditioned on an outcome, not the post-measurement quantum state. Optimizing the operators gives an affine upper bound U_{aff} on accessible information I_{acc} . For any finite ensemble of full-rank states, the algebra $\mathcal{A}_\star$ generated by the optimized operators satisfies

$$\boxed{I_{\text{acc}} = U_{\text{aff}} \iff \mathcal{A}_\star \text{ is abelian.}} \quad (1)$$

Thus a measurement attains the bound if and only if the optimized operators commute. Their common eigenspaces determine all optimal finite measurements, even when the signal states themselves do not commute. For two signal states, the optimized operators are functions of a single Hermitian matrix. Measuring in its eigenbasis attains accessible information. The construction extends to singular states and gives a computable bound on information loss at each optimization step. For full-rank binary signals, it also limits the weight that a nearly optimal measurement

assigns outside each outcome's dominant spectral block.

We also examine how the optimal measurement changes when the objective includes discrimination error or final recovery. For mixed signals the information-optimal and minimum-error measurements generally differ [11, 12], and information and guessing probability need not order binary measurements alike [13], even though minimum error has a projective solution [14]. For two full-rank qubit states, a weighted information–error objective has an optimal measurement with at most three outcomes. Some ensembles require all three over an open interval of weights (Fig. 1). In a separate repetition-code model, optimizing recovery can reduce the final error while lowering local information at unchanged discrimination error.

Posterior algebra—Let $\mathcal{E} = \{(p_i, \rho_i)\}_{i=1}^m$ have positive priors and faithful (positive-definite) states on a finite-dimensional Hilbert space, and put $\tau = \sum_i p_i \rho_i$. A POVM $\{M_y\}$ consists of positive effects summing to I ; a PVM consists of orthogonal projectors. Write $q(y|i) = \text{Tr } \rho_i M_y$ and $q_y = \text{Tr } \tau M_y$. Zero effects are omitted. Then $I(\mathcal{E}, M) = \sum_{iy} p_i q(y|i) \log[q(y|i)/q_y]$, with $I_{\text{acc}} = \sup_M I(\mathcal{E}, M)$. Logarithms are natural unless bits are specified. A finite optimizer exists [15].

Choose minimal real affine coordinates $\rho_i = \tau + \sum_a s_{ia} A_a$, with $\sum_i p_i s_{ia} = 0$. For a tuple of Hermitian matrices $\mathbf{H} = (H_1, \dots, H_r)$, define

$$G_i(\mathbf{H}) = I + \sum_a s_{ia} H_a \succ 0, \\ \Phi(\mathbf{H}) = \sum_i p_i \text{Tr } \rho_i \log G_i(\mathbf{H}), \quad U_{\text{aff}} = \max_{\mathbf{H}} \Phi. \quad (2)$$

The feasible domain is bounded, and the objective tends to minus infinity at its boundary. Strong concavity then gives a unique interior maximizer $\mathbf{H}_\star$. The algebra $\mathcal{A}_\star = C^*(G_1^\star, \dots, G_m^\star)$ is independent of the chosen affine coordinates. Here C^* means the unital operator algebra generated by the indicated Hermitian matrices. The optimized operators need not diagonalize the signal states. The normalization $\sum_i p_i G_i = I$ bounds each feasible operator by $p_i^{-1} I$.

To bound a measurement, set $h_{ay} = \text{Tr}(A_a M_y)/q_y$

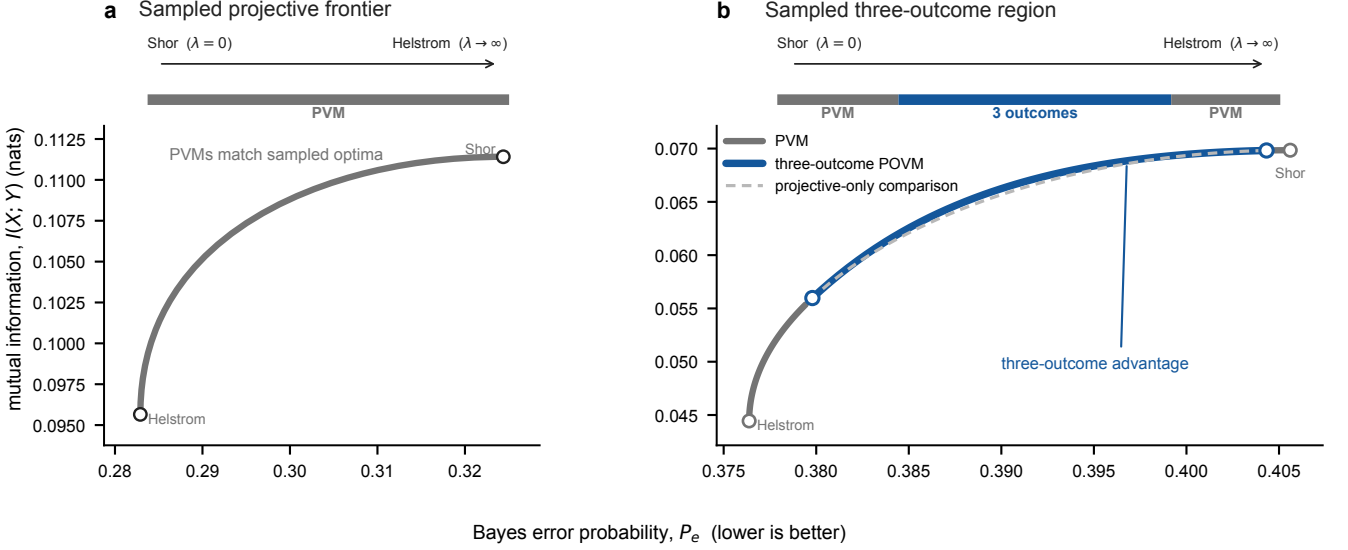


FIG. 1. Information–error trade-offs for two binary-qubit ensembles. Gray shows the sampled projective frontier; blue shows a numerically resolved advantage of three-outcome measurements. (a) Projective measurements match the optimum at the sampled weights. (b) Three outcomes improve the intermediate trade-off between minimum error (Helstrom) and maximum information (Shor). The finite grids test only the sampled weights; the analytic existence result is proved separately.

and $H_{a,M} = \sum_y h_{ay} M_y$. Bayes' rule identifies $g_{iy} = q(y|i)/q_y = \Pr(i|y)/p_i$, and $G_i(\mathbf{H}_M) = \sum_y g_{iy} M_y$. Operator Jensen therefore gives $I(\mathcal{E}, M) \leq \Phi(\mathbf{H}_M) \leq U_{\text{aff}} \leq \chi$, where χ is the Holevo quantity [16–20].

To prove Eq. (1), first suppose the optimized operators commute, with joint spectral PVM $P_\star = \{P_z\}$ and $G_i^\star P_z = g_{iz} P_z$. Direct expansion yields

$$I(\mathcal{E}, P_\star) - \Phi(\mathbf{H}_\star) = \sum_z \text{Tr}(\tau P_z) D(\alpha_{.z} \| \beta_{.z}) \geq 0, \quad (3)$$

where $\alpha_{iz} = p_i \text{Tr}(\rho_i P_z) / \text{Tr}(\tau P_z)$ and $\beta_{iz} = p_i g_{iz}$. The upper bound forces equality, so the spectral measurement is optimal. The measured and predicted posterior probability vectors therefore coincide.

Conversely, a finite POVM attaining U_{aff} must saturate Jensen's inequality. Uniqueness of the maximizer and the equality condition for operator compression imply $g_{iy} M_y^{1/2} = M_y^{1/2} G_i^\star$. Each effect lies in a common eigenspace; completeness forces commutativity. Stationarity gives $P_z \rho_i P_z = g_{iz} P_z \tau P_z$. Thus all optimal finite POVMs, and only those, satisfy

$$M_y = P_z M_y P_z \quad (y \in Y_z), \quad \sum_{y \in Y_z} M_y = P_z. \quad (4)$$

Resolving outcomes within a block adds no label information. The outcome distributions are interconvertible by label-independent classical postprocessing, or Blackwell-equivalent [21, 22]. The classification requires attainment of the affine bound, a stronger condition than projective optimality for a general ensemble. Detailed equality arguments appear in Appendix B.

Binary ensembles and Shor's conjecture—For priors $p, q = 1 - p > 0$, put $A = \rho_0 - \rho_1$. Then $\rho_0 = \tau + qA$ and $\rho_1 = \tau - pA$, so

$$G_0^\star = I + qH_\star, \quad G_1^\star = I - pH_\star. \quad (5)$$

Both optimized operators are functions of the same Hermitian matrix and therefore commute, whether or not the signal states do. A measurement in its eigenbasis attains accessible information.

The conclusion extends to singular states. On $\text{supp } \tau$, maximize the closed objective on $-q^{-1}I \leq H \leq p^{-1}I$, using zero for kernel contributions outside the signal support and $-\infty$ when a signal has support in the kernel. This upper-semicontinuous, strictly concave objective has a unique maximizer; the lower-semicontinuous form of Eq. (3) proves spectral attainment. Boundary Jensen equality gives $M_y^{1/2}(H_\star - h_y I)\rho_i^{1/2} = 0$ for both signals. Their supports span the effective space, so every effect lies in one spectral block. Blockwise stationarity, including zero-posterior endpoints, proves Eq. (4); boundary details are in Appendix D. Consequently every finite-dimensional binary ensemble admits a rank-one projective information optimum, and every optimal finite POVM, compressed to $\text{supp } \tau$, is a label-independent refinement of its canonical posterior PVM. Degenerate blocks allow different bases; unoccupied ambient subspaces have no statistical effect. Within each canonical block, the posterior probabilities are identical; the eigenvalues of the signal states need not be. Refining the measurement inside a block adds outcomes but no information about the label. This statement concerns the fixed ensemble and the information objective.

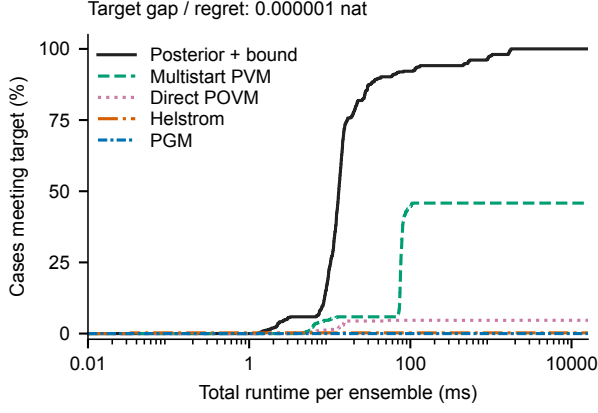


FIG. 2. Accuracy–runtime profile of 408 fixed-budget stress cases. The posterior method uses its analytic gap evaluated in float64; comparators use empirical regret against its PVM. Fractions count test cases.

Near-optimal measurements—For faithful binary states, let $\epsilon_M = I_{\text{acc}} - I(\mathcal{E}, M)$ and $a_i = p_i^3 \lambda_{\min}(\rho_i)$. Quantitative Jensen and strong concavity give

$$\sum_{iy} a_i \|g_{iy} M_y^{1/2} - M_y^{1/2} G_i^*\|_F^2 \leq 4\epsilon_M. \quad (6)$$

Here $\|\cdot\|_F$ denotes the Frobenius norm. If Δ_{post} is the minimum weighted separation of distinct posterior eigenvalue vectors, the total effect weight outside each outcome’s dominant block obeys $L(M) \leq 8\epsilon_M / \Delta_{\text{post}}^2$. Explicitly, with $z(y)$ maximizing $\text{Tr}(P_z M_y)$,

$$\begin{aligned} \Delta_{\text{post}}^2 &= \min_{z \neq z'} \sum_i a_i (g_{iz} - g_{iz'})^2, \\ L(M) &= \sum_y \sum_{z \neq z(y)} \text{Tr}(P_z M_y). \end{aligned} \quad (7)$$

For a fixed full-rank ensemble with separated posterior eigenvalues, a small information loss therefore limits the weight outside the assigned blocks. The bound deteriorates as the states become singular or the blocks merge; for a single block the weight outside it is zero. Definitions and finite-iterate bounds are in Appendix C.

Constructing a receiver with an information bound—For two signal states, we optimize one Hermitian matrix rather than all POVM effects. At a feasible H , diagonalization gives a PVM P_H and the interval

$$L = I(\mathcal{E}, P_H) \leq I_{\text{acc}} \leq U = \Phi(H) + \frac{\text{Tr}[R(H)^2]}{2C}, \quad (8)$$

where $R(H) = pq[D \log(G_0)[\rho_0] - D \log(G_1)[\rho_1]]$ and $C = p^2 q^2 [p \lambda_{\min}(\rho_0) + q \lambda_{\min}(\rho_1)]$. Strong concavity gives the upper bound, while the information obtained by the spectral measurement gives the lower bound. Their difference measures how far the current measurement can be from optimal.

Algorithm 1 in Appendix E gives the explicit updates and stopping conditions. At each step, an ascent direction and backtracking preserve positivity and increase $\Phi(H)$. Diagonalizing the updated matrix gives the PVM and its information interval. The bound is independent of the update rule; the benchmark uses inverse BFGS.

Logarithms are evaluated only at positive trial matrices. The interval and stopping flag are valid in exact arithmetic; floating-point certification requires controlled rounding and a terminating line search. Appendix E details matrix-function evaluation, inverse-BFGS proposals, and the benchmark protocol.

The 408-case benchmark spans dimensions 2, 4, 8, 16, spectral floors down to 10^{-6} , priors up to 0.99, and nearly coincident signals (Fig. 2). Under the declared budgets, the posterior receiver reports a gap of at most 10^{-6} nat on all cases, versus 45.8% for multistart PVM and 4.7% for direct POVM ascent. The comparison uses fixed iteration budgets. Readout requires a basis rotation and an orthogonal measurement; costs and accuracy depend on the device. Appendix E gives the full profiles and budgets.

Information-error Pareto frontiers—We next optimize information and discrimination error together for two distinct full-rank qubit states with fixed priors. The Helstrom measurement minimizes error, while the posterior-spectral measurement maximizes information. Their optimal bases need not coincide.

Let $\eta_y = P(X = 0|y)$ be the posterior probability of signal 0 after outcome y , and let $q_y = P(y)$. Information and error are averages of different functions of this same probability:

$$\begin{aligned} I(\mathcal{E}, M) &= h(p) - \sum_y q_y h(\eta_y), \\ P_e(M) &= \sum_y q_y \min(\eta_y, 1 - \eta_y). \end{aligned} \quad (9)$$

Here $h(t) = -t \log t - (1 - t) \log(1 - t)$ is binary entropy; the posterior-information form is also used in Ref. [12]. Error averages the probability of a wrong guess; information measures the reduction of uncertainty about the preparation label. Both quantities concern the classical measurement outcomes. The trade-off does not describe disturbance of the quantum state after measurement.

The Pareto frontier collects attainable performance pairs for which neither score can improve without worsening the other. We probe these trade-offs by maximizing $F_\lambda = I - \lambda P_e$, with $\lambda \geq 0$. Zero weight selects maximum information. As $\lambda \rightarrow \infty$, optimal performance approaches minimum error, with information breaking ties. Write $V_{\text{POVM}}(\lambda)$ and $V_{\text{PVM}}(\lambda)$ for the best weighted scores over generalized and projective measurements, respectively.

Our central result is that intermediate optima can require a genuinely nonprojective measurement, even though neither endpoint requires one. For every weight,

some optimal receiver has at most three outcomes. Moreover, a strict advantage,

$$\Delta_\lambda = V_{\text{POVM}}(\lambda) - V_{\text{PVM}}(\lambda) > 0, \quad (10)$$

occurs precisely when the smallest outcome count among weighted-score optimizers is three. Every two-effect qubit POVM is a classical postprocessing of a PVM, so two outcomes cannot realize this advantage. The three-outcome attainment statement concerns each weighted optimum, not every point of a flat exposed performance face.

The three-outcome bound follows by rewriting the measurement problem as an optimization over probability distributions with three fixed moments. Write the normalized posterior operator and average state in a common Bloch frame, $Q = \tau^{-1/2} p \rho_0 \tau^{-1/2} = cI + d\sigma_z$ and $\tau = (I + r_x \sigma_x + r_z \sigma_z)/2$, with $r_x \geq 0$. A rank-one measurement corresponds to a pure-state decomposition of τ , using the normalized-effect transformation of Ref. [12]; explicit forward and inverse maps are given in Appendix G. Setting $s(z) = \sqrt{1 - z^2}$, the weighted optimum is $h(p) - \lambda/2$ plus

$$\max_{\mu} \left\{ \int g_\lambda(z) d\mu : \int z d\mu = r_z, \quad \int s(z) d\mu \geq r_x \right\}, \quad (11)$$

where μ is a probability measure on $[-1, 1]$ and $g_\lambda(z) = -h(c + dz) + \lambda|c + dz - 1/2|$. Strict convexity makes the last constraint saturate at an optimum. An extreme maximizing distribution has at most three support points, since only the moments $1, z, s(z)$ are fixed. To prove that three outcomes are necessary, we construct an upper bound that is attained at exactly three points. When the mean lies inside the triangle formed by those points, all three must have positive weight. The corresponding measurement has nonorthogonal effects, and every outcome contributes to information and discrimination; none is discarded or assigned a rejection cost. This distinguishes the mechanism from the three-outcome measurements that interpolate between minimum-error and unambiguous discrimination by admitting a fixed rate of inconclusive results [23–25], and from the pure-state setting, where relative convexity or concavity with respect to the Bhattacharyya parameter selects the Helstrom projection or, for suitable priors, the unambiguous three-outcome measurement within the admissible objective classes [26]. Here both endpoint objectives are projective, and the signals are mixed.

Nor can random switching between projective bases reproduce a strict gap. Choose a basis b independently of the signal with probability w_b , and retain its label in the record. Then $F_\lambda = \sum_b w_b F_\lambda(P_b) \leq V_{\text{PVM}}(\lambda)$. Forgetting the basis label cannot improve either score. Thus the advantage is over all such randomized projective strategies, not just over the Helstrom and Shor measurements separately.

Figure 1 shows two numerical examples. In panel (a), projective measurements match the optimum at the sam-

pled weights. In panel (b), a three-outcome measurement improves the intermediate trade-off. Our construction establishes gaps persisting over open intervals of weights. The examples do not prove that every ensemble has a single connected nonprojective region; that topology remains conjectural. When a common measurement optimizes both endpoints, the frontier instead collapses to one performance point. We prove this collapse for every weight when the posterior spectrum is centered, $p^2 \det \rho_0 = (1 - p)^2 \det \rho_1$. The same condition is the explicitly solvable case in which Fuchs found the information-optimal qubit basis to coincide with the Helstrom basis [11], and for equal priors it is the equal-Bloch-radius coincidence reported in Ref. [12].

At fixed $Q = cI + d\sigma_z$, some faithful average state admits a strict weighted-score gap if and only if $c - d < 1/2 < c + d$ and $c \neq 1/2$. The construction yields an open set of such states. The full moment argument, global certificates, and proof of this existence criterion are given in Appendix G. The operator Q determines the possible posterior probabilities; it is distinct from the optimized matrix $H_\star$ used to construct the information-optimal measurement. We next consider a different objective: the error remaining after decoding.

Choosing the measurement for error correction—Soft-information decoding uses outcome probabilities rather than only hard binary decisions [27, 28]. Here we also optimize the measurement that produces those outcomes. Consider a synthetic distance-three repetition code with stochastic Pauli- X faults, fresh binary qubit pointers, and no ideal terminal syndrome. Let S be one syndrome bit, C the other measurement records available to the decoder, and E the complete error vector. Holding other receivers fixed and assuming the chosen pointer depends only on S , with $0 < P(S = 0) < 1$, a Pauli- X recovery maximizing entanglement fidelity gives the recovery risk $R_{\text{rec}} = \sum_y q_y r(t_y)$, where $t_y = P(S = 0|y)$ and

$$r(t) = \sum_c \left[\sum_e b_{ec}(t) - \max_e b_{ec}(t) \right], \quad b_{ec}(t) = tP(e, c|S = 0) + (1 - t)P(e, c|S = 1). \quad (12)$$

For a logical-reference Bell input, the stated X -only error group has trivial intersection with the Z -type stabilizer group. A wrong complete error vector therefore gives zero fidelity in this recovery class; choosing the most probable vector for each full record yields the loss above. The concave, piecewise-linear loss r depends on the code, noise model, and other measurements. With these fixed, we minimize its average over outcomes. It generally differs from entropy or a single binary decision loss, so the strict-convexity Pareto proof does not apply directly.

The other records are averaged over at the design stage; all measurements can be fixed before the experiment. Changing another receiver changes the joint distribution

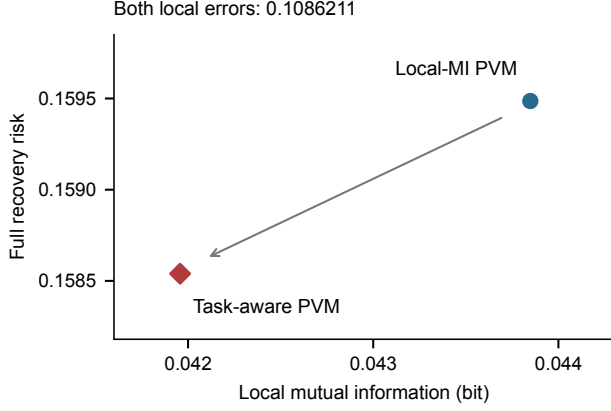


FIG. 3. A task-aware PVM lowers recovery risk while decreasing local information; local error is unchanged. Asymmetric 60° pointers, three rounds, fault probability 0.02, final-site replacement. Risks are deterministic model enumerations, without replicate error bars.

and r , so alternating site updates need not find the globally optimal schedule.

Figure 3 exhibits a 0.59357% risk reduction despite lower local information and unchanged error. Local mutual information and discrimination error therefore do not fully predict the final recovery error. The decoder also uses correlations with the other records; this does not violate data processing or establish Blackwell dominance. Across the 36 tested cases, optimizing recovery improved on locally information-optimal PVMs in 30 cases, mainly by changing the measurement basis. The protocols, generalized-measurement comparisons, and sensitivity to parameter changes are given in Appendix N, with full results in Appendix P. These calculations concern the synthetic model, rather than device performance or a fault-tolerance threshold.

Conclusion—For binary ensembles, the optimized posterior algebra identifies a canonical projective measurement and classifies all information-optimal finite measurements, including singular signals. For full-rank binary signals, it also gives a computable information bound and controls deviations of nearly optimal measurements from separated spectral blocks. Combining information and discrimination error can instead require three outcomes. In the repetition-code model, optimizing final recovery can favor a measurement with less local information and unchanged local error. The useful measurement therefore depends on both the signals and the decision made from its outcomes.

The recovery analysis uses its own loss function and does not extend the affine exactness theorem to arbitrary losses. A device-level comparison would need to include readout backaction, leakage, implementation costs, and conditional quantum states. Questions remain about the

outcome count in higher dimensions, the connectivity of the nonprojective qubit regions, and stability near singular states or merging spectral blocks.

Acknowledgments—We acknowledge Gewu Intelligence Lab for providing the collaborative research environment in which this project was developed. K.C. acknowledges support from the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDB1680102.

Disclosure of automated assistance—OpenAI GPT-5.6 Sol, OpenAI GPT-6 Astra, OpenAI GPT-6.1 Sol, Anthropic Claude Fable 5, and Anthropic Claude Fable 5.1 were used in developing and preparing this manuscript. The AI-assisted workflow developed the posterior algebra and optimal-measurement arguments; checked derivations, equality conditions, singular limits, stability bounds, and receiver certificates; examined information–error trade-offs and decoder-dependent recovery; reviewed numerical protocols and results; checked references; and prepared the manuscript and figures. The authors defined the research objectives, reviewed the AI-assisted derivations and citations, and revised the manuscript.

Data availability—The analytical derivations, numerical protocols, and reported results are presented in this article and its appendices. The accompanying manuscript source contains the figure assets used here. Complete benchmark inputs and executable reproduction programs are not included in that source package.

* These authors contributed equally to this work.

† Contact author: chenkun@itp.ac.cn

- [1] M. Sasaki, S. M. Barnett, R. Jozsa, M. Osaki, and O. Hirota, Accessible information and optimal strategies for real symmetrical quantum sources, *Phys. Rev. A* **59**, 3325 (1999).
- [2] P. W. Shor, On the number of elements needed in a POVM attaining the accessible information (2000), arXiv:quant-ph/0009077.
- [3] A. Peres and W. K. Wootters, Optimal detection of quantum information, *Phys. Rev. Lett.* **66**, 1119 (1991).
- [4] L. B. Levitin, Optimal quantum measurements for two pure and mixed states, in *Quantum Communications and Measurement*, edited by V. P. Belavkin, O. Hirota, and R. L. Hudson (Plenum Press, New York, 1995) pp. 439–448.
- [5] C. A. Fuchs and C. M. Caves, Ensemble-dependent bounds for accessible information in quantum mechanics, *Phys. Rev. Lett.* **73**, 3047 (1994).
- [6] A. Keil, Proof of the orthogonal measurement conjecture for qubit states, *Int. J. Quantum Inf.* **22**, 2440008 (2024), arXiv:0809.0232 [quant-ph].
- [7] K. Fang, H. Fawzi, and O. Fawzi, Uhlmann’s theorem for measured divergences, *IEEE Trans. Inf. Theory* **72**, 1751 (2026), arXiv:2502.07745.
- [8] S. Jo, Shor’s conjecture is true: Projective measurements suffice for binary accessible information (2026),

- arXiv:2609.19003.
- [9] M. Dall’Arno, Proof of Shor’s conjecture on the accessible information of quantum dichotomies (2026), arXiv:2609.20600.
 - [10] An earlier version of the present work, containing the binary proof for arbitrary finite-dimensional ensembles including singular states, was publicly deposited on 13 August 2026 (doi:10.5281/zenodo.21911604), before the public appearance of Refs. [8, 9].
 - [11] C. A. Fuchs, *Distinguishability and Accessible Information in Quantum Theory*, Ph.D. thesis, University of New Mexico, Albuquerque (1995), arXiv:quant-ph/9601020.
 - [12] Á. Varga, P. Adam, and J. A. Bergou, Maximum information measurement for qubit states, *Sci. Rep.* **14**, 11888 (2024).
 - [13] K. D. A. Thai and M. Dall’Arno, On Shor’s conjecture on the accessible information of quantum dichotomies, *IEEE Trans. Inf. Theory* **72**, 7641 (2026), arXiv:2512.11233.
 - [14] C. W. Helstrom, *Quantum Detection and Estimation Theory*, Mathematics in Science and Engineering, Vol. 123 (Academic Press, New York, 1976).
 - [15] E. B. Davies, Information and quantum measurement, *IEEE Trans. Inf. Theory* **24**, 596 (1978).
 - [16] A. S. Holevo, Bounds for the quantity of information transmitted by a quantum communication channel, *Probl. Inf. Transm.* **9**, 177 (1973), English translation of Problemy Peredachi Informatsii 9, 3–11 (1973).
 - [17] F. Hansen and G. K. Pedersen, Jensen’s operator inequality, *Bull. Lond. Math. Soc.* **35**, 553 (2003).
 - [18] K. Matsumoto, On single-copy maximization of measured f -divergence between a given pair of quantum states (2014), arXiv:1412.3676 [quant-ph].
 - [19] F. Hiai, *Quantum f -Divergences in von Neumann Algebras: Reversibility of Quantum Operations*, Mathematical Physics Studies (Springer, Singapore, 2021).
 - [20] M. Berta, O. Fawzi, and M. Tomamichel, On variational expressions for quantum relative entropies, *Lett. Math. Phys.* **107**, 2239 (2017).
 - [21] Y. Kuramochi, Minimal sufficient positive-operator valued measure on a separable Hilbert space, *J. Math. Phys.* **56**, 102205 (2015), arXiv:1506.07288.
 - [22] Y. Kuramochi, Minimal sufficient statistical experiments on von Neumann algebras, *J. Math. Phys.* **58**, 062203 (2017), arXiv:1701.03394.
 - [23] A. Chefles and S. M. Barnett, Strategies for discriminating between non-orthogonal quantum states, *J. Mod. Opt.* **45**, 1295 (1998).
 - [24] J. Fiurásek and M. Ježek, Optimal discrimination of mixed quantum states involving inconclusive results, *Phys. Rev. A* **67**, 012321 (2003), arXiv:quant-ph/0208126.
 - [25] E. Bagan, R. Muñoz-Tapia, G. A. Olivares-Rentería, and J. A. Bergou, Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes, *Phys. Rev. A* **86**, 040303 (2012), arXiv:1206.4145.
 - [26] I. Katz, A. Samorodnitsky, and Y. Kochman, Optimal discrimination between two pure states and Dolinar-type coherent-state detection, *IEEE Trans. Inf. Theory* **70**, 2701 (2024), arXiv:2311.02366.
 - [27] C. A. Pattison, M. E. Beverland, M. P. da Silva, and N. Delfosse, Improved quantum error correction using soft information (2021), arXiv:2107.13589.
 - [28] H. Ali, J. Marques, O. Crawford, J. Majaniemi, M. Serraperralta, D. Byfield, B. Varbanov, B. M. Terhal, L. Di Carlo, and E. T. Campbell, Reducing the error rate of a superconducting logical qubit using analog readout information, *Phys. Rev. Applied* **22**, 044031 (2024).
 - [29] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Iterative procedure for computing accessible information in quantum communication, *Phys. Rev. A* **71**, 054303 (2005).
 - [30] Y. C. Eldar and G. D. Forney, Jr., On quantum detection and the square-root measurement, *IEEE Trans. Inf. Theory* **47**, 858 (2001), arXiv:quant-ph/0005132.

Appendix A: Affine geometry and the information upper bound— Throughout this section, $p_i > 0$ and every ρ_i is positive definite. Let $\mathbf{V} = \text{span}_{\mathbb{R}}\{\rho_i - \tau\}$ and choose a minimal representation

$$\rho_i = \tau + \sum_{a=1}^r s_{ia} A_a, \quad \sum_i p_i s_{ia} = 0, \quad G_i(\mathbf{H}) = I + \sum_a s_{ia} H_a \succ 0. \quad (\text{A1})$$

The feasible domain is nonempty and convex. Since $\sum_i p_i G_i(\mathbf{H}) = I$, every feasible posterior satisfies $G_i \preceq p_i^{-1} I$. Full column rank of the coordinate matrix bounds $\mathbf{H}$. If a feasible sequence approaches the boundary, some G_j acquires an eigenvalue tending to zero. Faithfulness of ρ_j implies $\text{Tr } \rho_j \log G_j \rightarrow -\infty$, while all other terms remain bounded above. The maximum is therefore attained in the interior.

For $X \succ 0$, $R \succ 0$, and Hermitian K , operator concavity of the logarithm and its integral representation give

$$\frac{\lambda_{\min}(R)}{\|X\|_{\infty}^2} \|K\|_F^2 \leq -D^2 \text{Tr } R \log X[K, K] \leq \frac{\|R\|_{\infty}}{\lambda_{\min}(X)^2} \|K\|_F^2. \quad (\text{A2})$$

The Hessian follows by integrating $2 \text{Tr}[R(X + uI)^{-1} K(X + uI)^{-1} K(X + uI)^{-1}]$. The integrand is a squared Frobenius norm; bounding R by its extreme eigenvalues and then the resolvents by those of X gives both constants. For a tuple direction $\mathbf{K}$, put $K_i = \sum_a s_{ia} K_a$. Since $G_i \preceq p_i^{-1} I$, Eq. (A2) yields

$$-D^2 \Phi(\mathbf{H})[\mathbf{K}, \mathbf{K}] \geq \sum_i p_i^3 \lambda_{\min}(\rho_i) \|K_i\|_F^2. \quad (\text{A3})$$

Minimality of the coordinates makes the right side a norm, so Φ is strongly concave and its interior optimizer is unique. Its stationarity conditions are

$$\begin{aligned} \sum_i p_i s_{ia} \mathcal{L}_{G_i^*}(\rho_i) &= 0, \\ \mathcal{L}_F(X) &= D \log F[X] = \int_0^\infty (F + uI)^{-1} X (F + uI)^{-1} du. \end{aligned} \quad (\text{A4})$$

For a finite POVM without zero effects, set

$$\begin{aligned} c_y &= \text{Tr } \tau M_y, & h_{ay} &= \frac{\text{Tr } A_a M_y}{c_y}, \\ H_{a,M} &= \sum_y h_{ay} M_y, & g_{iy} &= \frac{\text{Tr } \rho_i M_y}{c_y}. \end{aligned} \quad (\text{A5})$$

Then

$$G_i(\mathbf{H}_M) = \sum_y g_{iy} M_y. \quad (\text{A6})$$

The map $X \mapsto \log X$ is operator concave, hence

$$\log G_i(\mathbf{H}_M) \succeq \sum_y (\log g_{iy}) M_y. \quad (\text{A7})$$

Tracing against $p_i \rho_i$, summing over i , and using the definition of g_{iy} proves $I(\mathcal{E}, M) \leq \Phi(\mathbf{H}_M) \leq U_{\text{aff}}$.

For completeness, introduce the unrestricted value

$$U_{\text{unres}} = \max_{F_i \succ 0, \sum_i p_i F_i = I} \sum_i p_i \text{Tr } \rho_i \log F_i. \quad (\text{A8})$$

Affine feasibility gives $U_{\text{aff}} \leq U_{\text{unres}}$. For an unrestricted feasible family, set $q_i = \text{Tr } \tau F_i$ and $Q_i = F_i/q_i$. Then $\text{Tr } \tau Q_i = 1$, and the variational formula for measured relative entropy implies

$$\text{Tr } \rho_i \log F_i \leq D_{\mathbb{M}}(\rho_i \| \tau) + \log q_i. \quad (\text{A9})$$

Moreover, $\sum_i p_i q_i = 1$, so scalar Jensen gives $\sum_i p_i \log q_i \leq 0$. Consequently,

$$I_{\text{acc}} \leq U_{\text{aff}} \leq U_{\text{unres}} \leq \sum_i p_i D_{\mathbb{M}}(\rho_i \| \tau) \leq \chi(\mathcal{E}). \quad (\text{A10})$$

The variational formula is Lemma 1 and Theorem 2 of Ref. [20]; see also the measured-divergence treatments [18–20]. This last bound is not needed to prove posterior-algebra exactness. It shows that U_{aff} is an upper bound on accessible information that is no larger than the Holevo quantity.

Appendix B: Posterior exactness and all optimal measurements— We first show directly that a finite POVM attains the maximum information. Refine every effect into rank-one effects and write each as dwP , where P is a rank-one projector and the nonnegative weights sum to one. Completeness is the moment constraint $\int P d\mu(P) = I/d$. The information is the continuous linear functional

$$d \int \left[\sum_i p_i \operatorname{Tr}(\rho_i P) \log \frac{\operatorname{Tr}(\rho_i P)}{\operatorname{Tr}(\tau P)} \right] d\mu(P).$$

The rank-one projector space is compact, and the feasible probability measures form a compact convex set. Hence a maximum is attained at an extreme measure. Such a measure has at most d^2 atoms: more disjoint positive-mass support sets would give a linear dependence among their Hermitian matrix moments, and a bounded signed density perturbation would preserve the moments in both directions. The trace constraint also preserves total mass, contradicting extremality. Refinement cannot reduce mutual information, so this supplies a finite optimizer without an additional outcome-count assumption.

We first record equality in compression Jensen. Let $V : \mathcal{H} \rightarrow \mathcal{K}$ be an isometry, $T \succ 0$, and $G = V^*TV$. In the decomposition $\operatorname{ran} V \oplus (\operatorname{ran} V)^\perp$, write

$$T = \begin{pmatrix} G & B \\ B^* & C \end{pmatrix}. \quad (\text{B1})$$

The integral representation of the logarithm and the Schur complement yield

$$\log G - V^*(\log T)V = \int_0^\infty \left[(G + uI - B(C + uI)^{-1}B^*)^{-1} - (G + uI)^{-1} \right] du \succeq 0. \quad (\text{B2})$$

Equality holds if and only if $B = 0$, equivalently $TV = VG$. For the POVM dilation

$$Vx = \bigoplus_y M_y^{1/2}x, \quad T_i = \bigoplus_y g_{iy}I, \quad (\text{B3})$$

this gives

$$g_{iy}M_y^{1/2} = M_y^{1/2}G_i \quad \text{for every } i, y. \quad (\text{B4})$$

Suppose first that the optimized posteriors commute. Let $P_\star = \{P_z\}_z$ be their minimal joint-spectral PVM and write $G_i^*P_z = g_{iz}P_z$. Define

$$\alpha_{iz} = \frac{p_i \operatorname{Tr}(\rho_i P_z)}{\operatorname{Tr}(\tau P_z)}, \quad \beta_{iz} = p_i g_{iz}. \quad (\text{B5})$$

Both are probability vectors for each z , and a direct expansion gives

$$I(\mathcal{E}, P_\star) - \Phi(\mathbf{H}_\star) = \sum_z \operatorname{Tr}(\tau P_z) D(\alpha_{\cdot z} \| \beta_{\cdot z}) \geq 0. \quad (\text{B6})$$

Since $I(\mathcal{E}, P_\star) \leq I_{\text{acc}} \leq U_{\text{aff}} = \Phi(\mathbf{H}_\star)$, equality holds throughout and $P_\star$ is optimal.

Conversely, assume $I_{\text{acc}} = U_{\text{aff}}$. Finite-dimensional attainment supplies an optimal finite POVM M . Equality in

$$I(\mathcal{E}, M) \leq \Phi(\mathbf{H}_M) \leq \Phi(\mathbf{H}_\star) \quad (\text{B7})$$

forces $\mathbf{H}_M = \mathbf{H}_\star$ by uniqueness. The positive difference in Jensen's inequality has zero trace against each positive-definite state, so the difference itself is the zero operator. Equation (B4) holds with $G_i = G_i^*$. The support of every nonzero effect lies in a common eigenspace of all G_i^* ; POVM completeness makes these supports span $\mathcal{H}$. Hence $\mathcal{A}_\star$ is abelian.

It remains to classify the optimizers. Compress Eq. (A4) to a joint spectral block P_z and put

$$T_z = P_z \tau P_z, \quad B_{az} = P_z A_a P_z, \quad C_z = \sum_i \frac{p_i}{g_{iz}} s_i s_i^\top. \quad (\text{B8})$$

Minimal affine coordinates make each H_a^* a real linear combination of the $G_i^* - I$. It is therefore scalar on every minimal joint-spectral block, which defines the coefficients h_{az} below. The matrix C_z is positive definite. If $h_z = (h_{1z}, \dots, h_{rz})^\top$, normalization gives $\sum_i (p_i/g_{iz})s_i + C_z h_z = 0$, while the compressed stationarity equations give $C_z(B_z - h_z T_z) = 0$. Therefore

$$B_{az} = h_{az} T_z, \quad P_z \rho_i P_z = g_{iz} P_z \tau P_z. \quad (\text{B9})$$

Equation (B4) shows that every optimal effect is supported in one block. Conversely, any block refinement satisfies

$$\text{Tr } \rho_i M_y = g_{iz} \text{Tr } \tau M_y, \quad y \in Y_z. \quad (\text{B10})$$

Thus the conditional refinement channel

$$R(y|z) = \frac{\text{Tr}(\tau M_y)}{\text{Tr}(\tau P_z)} \quad (\text{B11})$$

is independent of the label i , while $y \mapsto z$ is deterministic. The induced classical experiments are Blackwell-equivalent [21, 22].

Appendix C: Robust posterior rigidity— We next bound the deviation from the equality condition for a measurement whose information is close to optimal. If $0 \prec T \preceq \beta I$, the block representation used in Eq. (B2) and the resolvent identity imply

$$\text{Tr}[\log G - V^*(\log T)V] \geq \frac{1}{2\beta^2} \|TV - VG\|_F^2. \quad (\text{C1})$$

Indeed, writing $X_u = G + uI$ and $S_u = B(C + uI)^{-1}B^*$, the trace of the integrand is bounded below by $(\beta + u)^{-3} \|B\|_F^2$. Integration gives the stated constant and $\|B\|_F = \|TV - VG\|_F$. If $\rho \succeq \mu I$, the same positive difference gives the corresponding lower bound after tracing against ρ .

Define the intrinsic posterior norm

$$\|K\|_{\text{post}}^2 = \sum_i a_i \|K_i\|_F^2, \quad a_i = p_i^3 \lambda_{\min}(\rho_i). \quad (\text{C2})$$

Equation (A3) says that Φ is one-strongly concave in this norm. Split the information gap as

$$\epsilon_M = U_{\text{aff}} - I(\mathcal{E}, M) = \epsilon_{\text{opt}} + \epsilon_J, \quad \begin{cases} \epsilon_{\text{opt}} = \Phi(\mathbf{H}_\star) - \Phi(\mathbf{H}_M), \\ \epsilon_J = \Phi(\mathbf{H}_M) - I(\mathcal{E}, M). \end{cases} \quad (\text{C3})$$

Strong concavity gives $\sum_i a_i \|G_i(\mathbf{H}_M) - G_i^*\|_F^2 \leq 2\epsilon_{\text{opt}}$. For the POVM dilation, $T_i \preceq p_i^{-1}I$ because $g_{iy} = q(y|i)/q(y) \leq p_i^{-1}$. Applying Eq. (C1) and then tracing against $p_i \rho_i$ gives the corresponding residual relative to $G_i(\mathbf{H}_M)$ bounded by $2\epsilon_J$. Minkowski's inequality in the weighted direct sum then yields

$$R_\star(M) = \sum_{i,y} a_i \|g_{iy} M_y^{1/2} - M_y^{1/2} G_i^*\|_F^2 \leq 4\epsilon_M. \quad (\text{C4})$$

No commutativity assumption enters this estimate.

When the optimized posterior algebra is abelian, write $G_i^* = \sum_z g_{iz} P_z$ and define

$$\Delta_{\text{post}}^2 = \min_{z \neq z'} \sum_i a_i (g_{iz} - g_{iz'})^2. \quad (\text{C5})$$

Write $g_z = (g_{1z}, \dots, g_{mz})$, $g_y = (g_{1y}, \dots, g_{my})$, and $\|v\|_a^2 = \sum_i a_i v_i^2$. For an outcome y , expansion over the spectral blocks gives

$$\sum_i a_i \|M_y^{1/2} G_i^* - g_{iy} M_y^{1/2}\|_F^2 = \sum_z \|g_z - g_y\|_a^2 \text{Tr } P_z M_y. \quad (\text{C6})$$

Lemma 1 (Weighted separation)— Let $u_1, \dots, u_K$ obey $\|u_z - u_{z'}\|_a^2 \geq \Delta^2$ for $z \neq z'$, and let $w_z \geq 0$, $W = \sum_z w_z$. For every u ,

$$\sum_z w_z \|u_z - u\|_a^2 \geq \frac{\Delta^2}{2} (W - \max_z w_z). \quad (\text{C7})$$

Proof— If $W = 0$, both sides vanish. For $W > 0$, the left side is minimized at $\bar{u} = W^{-1} \sum_z w_z u_z$. The weighted variance identity gives

$$\sum_z w_z \|u_z - \bar{u}\|_a^2 = \frac{1}{2W} \sum_{z, z'} w_z w_{z'} \|u_z - u_{z'}\|_a^2. \quad (\text{C8})$$

The off-diagonal terms are at least Δ^2 , and $\sum_z w_z^2 \leq W \max_z w_z$, proving Eq. (C7). $\blacksquare$

Assign y to a block $z(y)$ maximizing $\text{Tr } P_z M_y$. Apply Lemma 1 with $w_z = \text{Tr } P_z M_y$, sum over outcomes, and use Eq. (C4). This proves the wrong-block mass bound

$$L(M) = \sum_y \sum_{z \neq z(y)} \text{Tr } P_z M_y \leq \frac{8\epsilon_M}{\Delta_{\text{post}}^2}. \quad (\text{C9})$$

The explicit effects

$$B_y = P_{z(y)} M_y P_{z(y)}, \quad D_z = P_z - \sum_{y: z(y)=z} B_y \quad (\text{C10})$$

form an exact block refinement $\widetilde{M} = \{B_y\}_y \cup \{D_z\}_z$, with zero effects omitted. For experiments on the fixed signal ensemble, define the directed deficiency from N to M as $\inf_K \max_i \text{TV}(q_M(\cdot|i), K q_N(\cdot|i))$, where K ranges over stochastic maps; δ_{LR} is the maximum of the two directed deficiencies. Trace-norm Cauchy–Schwarz gives a symmetric finite-experiment deficiency bound

$$\delta_{\text{LR}}(M, \widetilde{M}) \leq \min\{1, L(M) + \sqrt{L(M)[d - L(M)]}\}. \quad (\text{C11})$$

For completeness, put $P = P_{z(y)}$, $Q = I - P$, and $L_y = \text{Tr } Q M_y Q$. Then

$$\|M_y - B_y\|_1 \leq L_y + 2\sqrt{(\text{Tr } P M_y P) L_y}. \quad (\text{C12})$$

Summing and using Cauchy–Schwarz once more gives

$$\sum_y \|M_y - B_y\|_1 \leq L(M) + 2\sqrt{L(M)[d - L(M)]}. \quad (\text{C13})$$

To simulate M from $\widetilde{M}$, report y after outcome B_y and distribute each D_z according to any fixed probability distribution over all original outcome labels. This remains defined even if no y was assigned to block z . In the reverse direction, map outcome y of M to the correspondingly named outcome of $\widetilde{M}$ and never report a D_z label. For every signal state, the two resulting total-variation errors are bounded by one half of the last display plus $L(M)/2$. Taking the larger directed deficiency proves Eq. (C11) (and the trivial upper bound is one). Thus the outcome distributions of a nearly optimal measurement can be approximated by those of a refinement of the canonical posterior PVM, after classical postprocessing in either direction. The bound concerns these distributions for the fixed signal ensemble; it does not give operator-by-operator or diamond-norm closeness of the measurements.

At a feasible numerical tuple $\widehat{\mathbf{H}}$, let $R_a = \nabla_a \Phi(\widehat{\mathbf{H}})$ and

$$C = \sum_i a_i s_i s_i^\top, \quad r_g^2 = \sum_{a,b} (C^{-1})_{ab} \text{Tr}(R_a R_b). \quad (\text{C14})$$

This is the squared dual norm of the gradient for the weighted norm defined in Eq. (C2). One-strong concavity implies, for $\mathbf{K} = \mathbf{H}_\star - \widehat{\mathbf{H}}$,

$$\Phi(\mathbf{H}_\star) - \Phi(\widehat{\mathbf{H}}) \leq \langle \nabla \Phi(\widehat{\mathbf{H}}), \mathbf{K} \rangle - \frac{1}{2} \|\mathbf{K}\|_{\text{post}}^2. \quad (\text{C15})$$

Dual Cauchy–Schwarz and maximization of the scalar quadratic give the first bound below. For the second, strong monotonicity of $-\nabla \Phi$ and stationarity at $\mathbf{H}_\star$ imply $\|\mathbf{K}\|_{\text{post}}^2 \leq \langle \nabla \Phi(\widehat{\mathbf{H}}), \mathbf{K} \rangle \leq r_g \|\mathbf{K}\|_{\text{post}}$:

$$U_{\text{aff}} \leq \Phi(\widehat{\mathbf{H}}) + \frac{1}{2} r_g^2, \quad \sum_i a_i \|G_i(\widehat{\mathbf{H}}) - G_i^\star\|_F^2 \leq r_g^2. \quad (\text{C16})$$

The first inequality is the binary stopping certificate used in the main text. A formally computer-assisted certificate additionally requires outward-rounded matrix functions, eigenvalue bounds, and mutual-information evaluation; the reported benchmark uses float64 arithmetic.

Appendix D: Binary construction and singular closure— If one prior vanishes, or if the two states coincide, every measurement has zero mutual information and any rank-one PVM is optimal. We henceforth take $0 < p, q < 1$ and distinct states.

For a faithful binary ensemble $\mathcal{E} = \{(p, \rho_0), (q, \rho_1)\}$, define

$$\begin{aligned} \Phi_{\text{bin}}(H) = & p \operatorname{Tr} \rho_0 \log(I + qH) \\ & + q \operatorname{Tr} \rho_1 \log(I - pH), \end{aligned} \quad (\text{D1})$$

with $I + qH \succ 0$ and $I - pH \succ 0$, and choose

$$\rho_0 = \tau + q(\rho_0 - \rho_1), \quad \rho_1 = \tau - p(\rho_0 - \rho_1). \quad (\text{D2})$$

The affine program is the single-variable maximization of Eq. (D1). Its optimized posteriors are

$$G_0^* = I + qH_*, \quad G_1^* = I - pH_*, \quad (\text{D3})$$

so they are functions of one generator and commute. Their common spectral PVM is the spectral PVM of H_* , and the posterior-algebra theorem proves that it attains I_{acc} . Refining degenerate spectral projectors to rank one cannot decrease mutual information and therefore preserves optimality.

We now allow singular signal states and classify all optimal finite POVMs. Restrict to $\mathcal{H}_0 = P_0 \mathcal{H}$, where $P_0 = \operatorname{supp} \tau$. Since $p, q > 0$, $\operatorname{supp} \rho_i \subseteq \mathcal{H}_0$, and τ is faithful on $\mathcal{H}_0$. On the compact operator interval

$$\overline{\mathcal{D}} = \{H : -q^{-1}I \preceq H \preceq p^{-1}I\}, \quad (\text{D4})$$

set $G_0(H) = I + qH$ and $G_1(H) = I - pH$. For $G \succeq 0$, use the extended convention

$$\operatorname{Tr} \rho \log G = \begin{cases} \operatorname{Tr}_{\operatorname{supp} G} \rho \log(G|_{\operatorname{supp} G}), & \operatorname{supp} \rho \subseteq \operatorname{supp} G, \\ -\infty, & \text{otherwise.} \end{cases} \quad (\text{D5})$$

The resulting $\overline{\Phi}_{\text{bin}}$ is concave and upper semicontinuous, so it attains its maximum on $\overline{\mathcal{D}}$. The value at $H = 0$ is finite.

The maximizer is unique. Indeed, suppose $H \neq K$ were maximizers and examine an interior point of their segment. Common kernel blocks of $G_i(H)$ and $G_i(K)$ carry no weight in ρ_i , by Eq. (D5); positivity also forces the segment direction to have no row or column joining such a block. On the remaining support, the integral representation of the logarithm gives, for $D = H - K$,

$$-\frac{d^2}{dt^2} \operatorname{Tr} \rho_i \log G_i(K + tD) = 2s_i^2 \int_0^\infty \operatorname{Tr} \rho_i R_{i,u} D R_{i,u} D R_{i,u} du \geq 0, \quad (\text{D6})$$

where $s_0 = q$, $s_1 = -p$, and $R_{i,u} = (G_i + uI)^{-1}$. The integrand equals $\|R_{i,u}^{1/2} D R_{i,u} \rho_i^{1/2}\|_F^2$. Equality for all u implies $D \rho_i^{1/2} = 0$ by taking $u \rightarrow \infty$. Equality in the concavity of the sum would therefore make D annihilate both signal supports. Their ranges span $\operatorname{supp} \tau = \mathcal{H}_0$, so $D = 0$, a contradiction. Denote the unique maximizer by H_* .

Every finite POVM on $\mathcal{H}_0$ defines a feasible point of the closed optimization problem. After deleting zero effects, put

$$c_y = \operatorname{Tr} \tau M_y > 0, \quad h_y = \frac{\operatorname{Tr}(\rho_0 - \rho_1) M_y}{c_y}, \quad g_{0y} = 1 + qh_y, \quad g_{1y} = 1 - ph_y, \quad (\text{D7})$$

and $H_M = \sum_y h_y M_y$. Then $G_i(H_M) = \sum_y g_{iy} M_y \succeq 0$. Applying compression Jensen to $T_{i,\epsilon} = \bigoplus_y (g_{iy} + \epsilon)I$ and taking $\epsilon \downarrow 0$ yields

$$I(\mathcal{E}, M) \leq \overline{\Phi}_{\text{bin}}(H_M) \leq \overline{\Phi}_{\text{bin}}(H_*). \quad (\text{D8})$$

Terms with $g_{iy} = 0$ contribute zero because then $\operatorname{Tr} \rho_i M_y = c_y g_{iy} = 0$. These identities also show $\operatorname{supp} \rho_i \subseteq \operatorname{supp} G_i(H_M)$: if $x \in \ker G_i(H_M)$, positivity of the summands gives $M_y x = 0$ when $g_{iy} > 0$, whereas $g_{iy} = 0$ implies $\rho_i^{1/2} M_y = 0$; summing over y yields $\rho_i^{1/2} x = 0$.

Let $P_* = \{P_z\}_z$ be the spectral PVM of H_* , with eigenvalues h_z and $g_{0z} = 1 + qh_z$, $g_{1z} = 1 - ph_z$. Finiteness of the objective implies $P_z \rho_i P_z = 0$ whenever $g_{iz} = 0$. Thus the gap identity

$$I(\mathcal{E}, P_*) - \overline{\Phi}_{\text{bin}}(H_*) = \sum_z \operatorname{Tr}(\tau P_z) D(\alpha_{\cdot z} \| \beta_{\cdot z}) \geq 0 \quad (\text{D9})$$

remains valid with $\alpha_{iz} = p_i \text{Tr}(\rho_i P_z) / \text{Tr}(\tau P_z)$, $\beta_{iz} = p_i g_{iz}$, and the usual zero conventions. Combining Eqs. (D8) and (D9) shows

$$I(\mathcal{E}, P_\star) = I_{\text{acc}} = \max_{H \in \overline{\mathcal{D}}} \overline{\Phi}_{\text{bin}}(H). \quad (\text{D10})$$

It remains to characterize equality for an arbitrary optimal finite POVM. We use the following boundary form of compression-Jensen equality. Let $T \succeq 0$, $G = V^*TV$, and $\rho \succeq 0$, with the extended logarithmic traces finite. If the limiting weighted Jensen gap vanishes, then

$$(TV - VG)\rho^{1/2} = 0. \quad (\text{D11})$$

To see this, replace T by $T + \epsilon I$ and use the block decomposition in Eq. (B2). If B is its off-diagonal block, then B is independent of ϵ , and the corresponding positive resolvent integrand $K_\epsilon(u)$ satisfies, uniformly for $0 \leq \epsilon \leq 1$,

$$u^3 K_\epsilon(u) \longrightarrow BB^* \quad (u \rightarrow \infty). \quad (\text{D12})$$

Consequently, if $B^* \rho^{1/2} \neq 0$, then on some fixed interval $u \in [R, 2R]$ the weighted integrand has a strictly positive lower bound, independent of sufficiently small ϵ . This contradicts a vanishing limiting Jensen gap. Hence $B^* \rho^{1/2} = 0$, which is precisely Eq. (D11).

If M is optimal, both inequalities in Eq. (D8) are equalities and uniqueness gives $H_M = H_\star$. Applying Eq. (D11) to each signal and outcome gives

$$M_y^{1/2}(H_\star - h_y I) \rho_0^{1/2} = 0, \quad M_y^{1/2}(H_\star - h_y I) \rho_1^{1/2} = 0. \quad (\text{D13})$$

The two signal ranges span $\mathcal{H}_0$, hence $M_y^{1/2}(H_\star - h_y I) = 0$. Consequently every nonzero M_y is supported in one spectral block P_z .

Finally, the signal states are proportional after compression to each block:

$$P_z \rho_i P_z = g_{iz} P_z \tau P_z. \quad (\text{D14})$$

For $-q^{-1} < h_z < p^{-1}$ this follows by varying H arbitrarily inside P_z and using stationarity; since $G_i^*|_{P_z} = g_{iz} P_z$, the resulting condition is $P_z \rho_0 P_z / g_{0z} = P_z \rho_1 P_z / g_{1z}$. At an endpoint, one g_{iz} vanishes, and Eq. (D5) gives the same identity directly. Therefore the optimal finite POVMs on $\mathcal{H}_0$ are exactly the block refinements

$$M_y = P_z M_y P_z \quad (y \in Y_z), \quad \sum_{y \in Y_z} M_y = P_z. \quad (\text{D15})$$

Their refinement channel is label-independent, so every one is Blackwell-equivalent to $P_\star$. Rank-one refinement proves Shor's orthogonal-measurement statement. On the ambient Hilbert space, every POVM has the same signal statistics as its compression $P_0 M_y P_0$; all positive extensions compatible with POVM completeness are statistically invisible beyond that compression.

The binary variational formulations in Refs. [8, 9] are related to Eq. (D1) by $T = p(I + qH)$, after matching signal labels. Indeed, $I - T = q(I - pH)$, and for an embedded measurement $T_M = \sum_y \Pr(0|y) M_y = p(I + qH_M)$. These affine changes preserve the spectral projectors. Applying the gap identity to H_M , rather than only to $H_\star$, gives $I(\mathcal{E}, M) \leq \overline{\Phi}_{\text{bin}}(H_M) \leq I(\mathcal{E}, P_{H_M})$. Thus the pointwise spectral replacement and binary variational value agree; the equality analysis above additionally classifies every finite optimizer, including at singular boundaries.

For comparison, we derive projective sufficiency from the theorem of Fang, Fawzi, and Fawzi. Their finite-dimensional projective-reduction theorem states the following. Let f be convex and lower semicontinuous with $(0, \infty) \subset \text{dom } f$. If an interval J admits a one-to-one parametrization $\psi : J \rightarrow \text{dom } f^*$ for which ψ is operator concave and $f^* \circ \psi$ is operator convex, then the measured f -divergence optimized over POVMs equals the value optimized over PVMs [7, Theorem 2].

If all signals in a collinear ensemble coincide, every measurement has zero mutual information. Otherwise choose two extreme signal states ρ_0 and ρ_1 and write $\rho_i = (1 - \lambda_i)\rho_0 + \lambda_i\rho_1$, where $0 \leq \lambda_i \leq 1$, some $\lambda_i = 0$, some $\lambda_i = 1$, and $\bar{\lambda} = \sum_i p_i \lambda_i \in (0, 1)$. The mutual information of a measurement is the measured f -divergence of the two endpoint outcome distributions for

$$f(t) = \sum_i p_i [(1 - \lambda_i) + \lambda_i t] \log \frac{(1 - \lambda_i) + \lambda_i t}{(1 - \bar{\lambda}) + \bar{\lambda} t}. \quad (\text{D16})$$

Here the displayed formula is for $t > 0$; set $f(0) = \lim_{t \downarrow 0} f(t)$ and $f(t) = +\infty$ for $t < 0$ when taking the convex conjugate. Set

$$h = \frac{t-1}{(1-\bar{\lambda}) + \bar{\lambda}t}, \quad \psi(h) = \sum_i p_i \lambda_i \log[1 + (\lambda_i - \bar{\lambda})h]. \quad (\text{D17})$$

On $J = (-1/(1-\bar{\lambda}), 1/\bar{\lambda})$, direct differentiation and Fenchel conjugacy give

$$f'(t) = \psi(h), \quad (\text{D18})$$

$$f^*(\psi(h)) = - \sum_i p_i (1 - \lambda_i) \log[1 + (\lambda_i - \bar{\lambda})h]. \quad (\text{D19})$$

The first expression is operator concave and the second is operator convex. Furthermore,

$$\psi'(h) = (1 - \bar{\lambda}h) \sum_i \frac{p_i (\lambda_i - \bar{\lambda})^2}{1 + (\lambda_i - \bar{\lambda})h} > 0. \quad (\text{D20})$$

The Möbius map sends $t \in (0, \infty)$ bijectively onto J . Since $\psi(h) = f'(t)$ and $\psi' > 0$, one-dimensional Fenchel duality identifies its range with $\text{dom } f^*$. Indeed, the signal with $\lambda_i = 1$ gives $f'(0+) = -\infty$, while the signal with $\lambda_i = 0$ makes the opposite limiting slope inadmissible because $f(t) - tf'(\infty) = -w_0 \log t + O(1) \rightarrow -\infty$, where $w_0 = \sum_{i:\lambda_i=0} p_i > 0$. Thus the opposite endpoint has infinite conjugate value and no point of the conjugate domain is omitted. With the lower-semicontinuous conventions at ratios zero and infinity, all hypotheses of Fang–Fawzi–Fawzi’s theorem hold. Therefore the POVM and PVM suprema agree; rank-one refinement and compactness give attainment. Thus projective sufficiency follows from that earlier theorem. The present construction also identifies the canonical posterior algebra, classifies all finite optimal measurements, and gives the stability and finite-iterate bounds above.

Appendix E: Posterior-spectral receiver and numerical benchmarks— For a faithful binary ensemble $\mathcal{E} = \{(p, \rho_0), (q, \rho_1)\}$, $q = 1 - p$, the affine coordinates are $s_0 = q$ and $s_1 = -p$, and the posterior program is Eq. (D1). Its unique optimizer generates both posterior operators. The spectral PVM of $H_\star$ is therefore information-optimal. At a feasible iterate $\hat{H}$, intrinsic strong concavity gives

$$I_{\text{acc}} \leq U_{\text{aff}} \leq \Phi(\hat{H}) + \frac{1}{2} r_g^2, \quad (\text{E1})$$

where r_g is the gradient’s dual norm for the weighted norm defined in Appendix C. The difference between this upper bound and the information obtained by the spectral PVM gives the stopping criterion used below.

Algorithm 1: Posterior-spectral receiver

Input: faithful states ρ_0, ρ_1 ; prior $0 < p < 1$; target gap $\varepsilon > 0$; iteration limit N ; initial step $\alpha_0 > 0$ and backtracking parameters $0 < \beta, \eta < 1$.

Output: a rank-one PVM P , information interval $[L, U]$, and a flag indicating whether $U - L \leq \varepsilon$.

1. Set $q \leftarrow 1 - p$, $H \leftarrow 0$, and compute the curvature constant C defined below Eq. (8).
 2. **For** $k = 0, \dots, N$ **do**
 3. Form $G_0 = I + qH$, $G_1 = I - pH$; evaluate $\Phi(H)$ and $R(H) = \nabla \Phi(H)$.
 4. Diagonalize H and refine any degenerate eigenspaces to obtain a rank-one PVM P_H .
 5. Set $L \leftarrow I(\mathcal{E}, P_H)$ and $U \leftarrow \Phi(H) + \text{Tr}[R(H)^2]/(2C)$.
 6. **If** $0 \leq U - L \leq \varepsilon$, **return** $(P_H, [L, U], \text{target met})$.
 7. **If** $k = N$, **return** $(P_H, [L, U], \text{target not met})$.
 8. Choose a Hermitian ascent direction D with $\text{Tr}[R(H)D] > 0$ (default $D = R(H)$); set $\alpha \leftarrow \alpha_0$.
 9. **While** either $I + q(H + \alpha D) \not\geq 0$ or $I - p(H + \alpha D) \not\geq 0$, or the feasible trial fails $\Phi(H + \alpha D) \geq \Phi(H) + \eta \alpha \text{Tr}[R(H)D]$, set $\alpha \leftarrow \beta \alpha$.
 10. Update $H \leftarrow H + \alpha D$.
 11. **End for**
-

Implementation of the posterior-spectral receiver: For the binary coordinates $s_0 = q$ and $s_1 = -p$, the intrinsic curvature matrix is the scalar

$$C = p^2 q^2 [p \lambda_{\min}(\rho_0) + q \lambda_{\min}(\rho_1)] > 0, \quad (\text{E2})$$

and the gradient of Eq. (D1) is

$$R(H) = pq [\mathcal{L}_{I+qH}(\rho_0) - \mathcal{L}_{I-pH}(\rho_1)]. \quad (\text{E3})$$

Intrinsic strong concavity gives $U_{\text{aff}} - \Phi(H) \leq \text{Tr}[R(H)^2]/(2C)$. A backtracking line search preserves $I + qH \succ 0$ and $I - pH \succ 0$; the implementation uses inverse BFGS, although the certificate is independent of the proposal method. Matrix logarithms and Fréchet derivatives are evaluated in an eigenbasis using stable divided differences. A formally computer-assisted certificate replaces floating-point values by outward-rounded enclosures. The implemented certificate and benchmark concern faithful inputs. Applying the numerical method to depolarized singular inputs would require an additional continuity bound; the exact singular theorem above is independent of this numerical extension.

Algorithms and benchmark protocol: Generative-AI assistance with analytical checks and review of the numerical protocols and results is described in the Acknowledgments.

We use a fixed-seed suite containing 408 stress instances: 24 cases at every dimension or conditioning point. It spans dimensions $d = 2, 4, 8, 16$, signal-state spectral floors down to 10^{-6} , priors up to 0.99, and nearly coincident state pairs $\rho_1 = (1 - \delta)\rho_0 + \delta\sigma$ with δ down to 10^{-2} .

The compared methods are as follows.

1. *Posterior-spectral receiver*: feasibility-preserving inverse BFGS applied to Eq. (D1), followed by spectral decomposition and Eq. (E1).
2. *Direct POVM ascent*: a multistart, SOMIM-style mutual-information ascent over d^2 POVM effects, using multiplicative positive updates and POVM renormalization [29]. Starts are described as uniform, Helstrom, and PGM in the benchmark protocol. An unperturbed identity-effect start is stationary for gradient ascent; the start construction and symmetry-breaking convention affect this baseline's performance.
3. *Multistart PVM*: BFGS on $U(d)$ in local exponential coordinates, initialized by identity, Helstrom, and Haar-random bases.
4. *Helstrom and PGM*: the minimum-error spectral measurement and the square-root measurement, respectively [14, 30]. The mixed-state PGM effects are $M_i = p_i \tau^{-1/2} \rho_i \tau^{-1/2}$ on $\text{supp } \tau$, where the inverse is restricted to that support. They are positive and sum to its identity. This is a feasible baseline, without an assertion of information or error optimality.

The baseline implementations optimize the same mutual information but have no independent global upper bound. In the stress suite the optimization budgets are deliberately fixed: at most 300 posterior iterations, 100 accepted direct- POVM iterations per deterministic start, and 30 PVM-BFGS iterations per start. The information loss in this benchmark therefore measures convergence within the specified iteration budgets. All methods optimize the same mathematical objective.

Solver profiles: Figure 4 shows the full profiles for all 408 stress instances. Panel (a) is an accuracy profile: for every target regret ϵ it reports the fraction of instances meeting the target. Panel (b) is a time-to-accuracy profile at $\epsilon = 10^{-6}$ nat. The posterior-spectral receiver reaches the target on all 408 cases, with median successful time 12.9 ms. Multistart PVM and direct POVM ascent reach the target on 45.8% and 4.7% of cases; Helstrom and PGM reach it on 0.25% and 0%. The black curve uses the affine upper-minus-PVM gap, whereas the remaining curves use empirical regret relative to the posterior-spectral PVM.

Numerical certificate values are evaluated in float64 arithmetic from the analytic bounds. Runtime comparisons use the implementations and budgets specified above; interval enclosures would additionally require outward-rounded arithmetic.

Appendix F: Scope, prior work, and limitations— The existence of a projective optimum for finite collinear ensembles is also a specialization of the measured- f projective-reduction theorem of Fang, Fawzi, and Fawzi [7]. The posterior-algebra proof also identifies a canonical measurement, classifies all optimal finite measurements, bounds deviations from the spectral blocks, and gives information bounds at finite optimization steps. Classical minimal sufficiency and postprocessing equivalence motivate the Blackwell interpretation [21, 22].

The stability constants for full-rank states deteriorate as an eigenvalue approaches zero. The closed-domain binary theorem nevertheless extends the canonical posterior PVM and complete finite-optimizer classification to arbitrary

singular states after support compression. A uniform ambient Frobenius-rigidity theorem does not follow: varying-support pure states require a visibility-weighted seminorm. The benchmark tests convergence at fixed iteration budgets after reducing the optimization to one Hermitian matrix. It does not determine polynomial bit complexity or asymptotic scaling exponents.

Appendix G: Scope and notation— We give self-contained proofs of the scalar formulation, the minimal-outcome criterion, the centered common optimum, and the constructive spectral existence theorem. The proofs do not assume that weights with a nonprojective advantage form one connected interval. Throughout, $\lambda \geq 0$, $0 < p < 1$, the two qubit states are distinct and positive definite, and entropy $h(q) = -q \log q - (1 - q) \log(1 - q)$ uses natural logarithms. Set $A = p\rho_0$, $B = (1 - p)\rho_1$, $\tau = A + B$ and $Q = \tau^{-1/2}A\tau^{-1/2}$. In the posterior normal form,

$$Q = cI + d\sigma_z, \quad d > 0, \quad 0 < c - d < c + d < 1, \quad \tau = \frac{1}{2}(I + r_x\sigma_x + r_z\sigma_z), \quad r_x \geq 0, \quad r_x^2 + r_z^2 < 1. \quad (\text{G1})$$

Conversely, every such Q, τ defines a faithful ensemble by

$$p = \text{tr}(\tau Q), \quad \rho_0 = \frac{\tau^{1/2}Q\tau^{1/2}}{p}, \quad \rho_1 = \frac{\tau^{1/2}(I - Q)\tau^{1/2}}{1 - p}. \quad (\text{G2})$$

The normalized posterior operator Q is distinct from the optimized matrix $H_\star$ used above. The representation below describes all rank-one measurements; it does not imply that the eigenbasis of Q maximizes information.

Appendix H: Exact scalar problem and saturation— The following normalized-effect transformation is given in Ref. [12, Eqs. (13)–(19)]. We use it to formulate the weighted information–error objective and establish its outcome-count and spectral existence criteria. For each rank-one effect write

$$\tau^{1/2}M_y\tau^{1/2} = q_y|\psi_y\rangle\langle\psi_y|. \quad (\text{H1})$$

Then q_y is the outcome probability, its posterior is $\eta_y = \langle\psi_y|Q|\psi_y\rangle$, and completeness is equivalent to $\sum_y q_y|\psi_y\rangle\langle\psi_y| = \tau$. The inverse is $M_y = q_y\tau^{-1/2}|\psi_y\rangle\langle\psi_y|\tau^{-1/2}$. Here $q_y = \text{Tr}(\tau M_y) > 0$ for each nonzero effect. Conversely, the displayed inverse maps every pure-state decomposition of τ to positive rank-one effects summing to I ; taking traces gives $\sum_y q_y = 1$. The posterior formula follows from $\text{Tr}(AM_y) = q_y\langle\psi_y|Q|\psi_y\rangle$.

Define $s(z) = \sqrt{1 - z^2}$ and $g_\lambda(z) = -h(c + dz) + \lambda|c + dz - 1/2|$. The maximum of the weighted objective equals $h(p) - \lambda/2$ plus

$$\max_{\mu \in \mathcal{P}([-1, 1])} \left\{ \int g_\lambda d\mu : \int z d\mu = r_z, \quad \int s d\mu \geq r_x \right\}. \quad (\text{H2})$$

Necessity follows from $n_x \leq s(n_z)$ for every pure Bloch vector. For sufficiency let $S = \int s d\mu$. If $S > 0$, split the atom at each z between $(s(z), 0, z)$ and $(-s(z), 0, z)$ in the common relative proportions $(1 + r_x/S)/2$ and $(1 - r_x/S)/2$. The resulting average is $(r_x, 0, r_z)$ and the score is unchanged. If $S = 0$, then $r_x = 0$ and the two poles suffice. The inverse correspondence gives a POVM. Spectral refinement of a general effect cannot reduce the score, since $-h(\eta) - \lambda \min(\eta, 1 - \eta)$ is convex and the unrefined posterior is the probability-weighted mean of its refinements. Thus no general POVM is excluded by the scalar optimization.

The capacity inequality in Eq. (H2) saturates at every optimum. Let ν be the two-pole distribution with mean r_z . Strict convexity of g_λ gives $\int g_\lambda d\nu > \int g_\lambda d\mu$ whenever μ has interior mass. If $S > r_x$, replace μ by

$$\tilde{\mu} = \frac{r_x}{S}\mu + \left(1 - \frac{r_x}{S}\right)\nu. \quad (\text{H3})$$

Its mean is unchanged, its capacity is r_x , and its score strictly increases. At saturation, equality in $\mathbb{E}n_x \leq \mathbb{E}s$ requires $n_x = s(z)$ at every occupied atom. All optimal rank-one atoms therefore lie on the positive real semicircle, including its poles.

Appendix I: Outcome count and global contact certificates— The saturated scalar problem has three linear moments, $1, z, s(z)$. An extreme feasible measure has at most three atoms: otherwise four disjoint positive-mass support sets yield a linear dependence among their three moment vectors, allowing a nonzero signed perturbation in both directions while preserving positivity and all moments. Compactness supplies an extreme maximizer of the continuous linear objective. Two rank-one qubit effects summing to I are orthogonal projections. Their pure-state decomposition images form a chord through (r_x, r_z) and need not be antipodal. Moreover, every two-effect POVM $\{E, I - E\}$ is a stochastic postprocessing of the eigenbasis measurement of E . Such postprocessing cannot increase information or decrease Bayes error. Hence the best two-outcome value equals the best PVM value.

For $\Delta_\lambda = V_{\text{POVM}}(\lambda) - V_{\text{PVM}}(\lambda)$ this gives

$$\Delta_\lambda > 0 \iff \text{the optimal minimal number of outcomes is three.} \quad (\text{I1})$$

This does not mean that every optimizer has exactly three labeled effects: redundant splitting is possible. A three-outcome measurement attains the maximum of each weighted objective. This statement does not impose the same outcome bound on every performance pair along a flat segment of the frontier.

A useful dual of Eq. (H2) is

$$\min_{\alpha, \beta \in \mathbb{R}, \kappa \geq 0} \{ \alpha + \beta r_z - \kappa r_x : \alpha + \beta z - \kappa s(z) \geq g_\lambda(z) \ \forall z \}. \quad (\text{I2})$$

The point mass at r_z has strictly feasible capacity, since $s(r_z) > r_x$. Equivalently a supporting-plane construction for the compact concave roof gives strong duality. A feasible upper bound touching the objective at exactly three noncollinear points proves a strict advantage whenever the average state lies inside their triangle. The bound is attained by assigning positive weight to all three points. A distribution on any two of them cannot reproduce an interior mean. Compactness of the PVM family then makes its shortfall strictly positive.

Appendix J: Centered spectra and a common endpoint optimum— Suppose $c = 1/2$. Set $a = 2d \in (0, 1)$ and

$$H_a(u) = h\left(\frac{1 + a\sqrt{1-u^2}}{2}\right), \quad \phi_\lambda(u) = -H_a(u) + \lambda d\sqrt{1-u^2}. \quad (\text{J1})$$

Then $g_\lambda(z) = \phi_\lambda(s(z))$. For $0 < u < 1$, with $t = \sqrt{1-u^2}$,

$$H'_a(u) = \frac{au}{t} \operatorname{atanh}(at) > 0, \quad H''_a(u) = \frac{a}{t^3} \left[\operatorname{atanh}(at) - \frac{at(1-t^2)}{1-a^2t^2} \right] > 0. \quad (\text{J2})$$

The last sign follows from $\operatorname{atanh}(at) > at \geq at(1-t^2)/(1-a^2t^2)$. Thus ϕ_λ is strictly concave and decreasing, so

$$\mathbb{E}g_\lambda(z) \leq \phi_\lambda(\mathbb{E}s(z)) \leq \phi_\lambda(r_x). \quad (\text{J3})$$

Let $z_0 = \sqrt{1-r_x^2}$. The two states $(r_x, 0, \pm z_0)$ with probabilities $(1 \pm r_x/z_0)/2$ attain equality for every λ . Both probabilities are positive, and the inverse correspondence gives a PVM. Its posterior values are $1/2 \pm dz_0$; consequently

$$P_e^{\min} = \frac{1}{2} - dz_0, \quad I_{\max} = h(\frac{1}{2} + dr_z) - h(\frac{1}{2} + dz_0). \quad (\text{J4})$$

This one performance point dominates all others.

For completeness, the invariant centered condition is

$$\operatorname{tr}[(A+B)^{-1}A] = 1 \iff \det A = \det B \iff p^2 \det \rho_0 = (1-p)^2 \det \rho_1. \quad (\text{J5})$$

To verify it, write $\det(tA+B) = t^2 \det A + tC + \det B$ and differentiate at $t = 1$. The trace is $(2 \det A + C)/(\det A + C + \det B)$. This characterizes centered spectra, not all possible common-optimum ensembles; commuting ensembles provide other common-optimum examples. In Bloch-vector form the centered condition reads $p^2(1-r_0^2) = (1-p)^2(1-r_1^2)$, where r_i is the Bloch length of ρ_i . It is the fourth explicitly solvable case, Eq. (3.255), of Fuchs's variational analysis of the information-optimal qubit basis, where that basis was observed to coincide with the Helstrom basis [11]; for equal priors it reduces to equal Bloch radii, the coincidence condition reported in Ref. [12]. A common endpoint optimum also optimizes every nonnegative weighted sum of the two endpoint scores; the argument above constructs this measurement explicitly.

Appendix K: Constructive spectral existence theorem— For a fixed posterior spectrum $0 < c-d < c+d < 1$, there exist a faithful average state τ and a weight $\lambda > 0$ with $V_{\text{POVM}}(\lambda) > V_{\text{PVM}}(\lambda)$ if and only if $c-d < 1/2 < c+d$ and $c \neq 1/2$. When these conditions hold, the advantage occurs for an open set of average states and persists over an open interval of weights for each such state. This is an existence criterion at fixed Q , not a claim that every compatible τ exhibits an advantage.

Fix Q with $0 < c-d < 1/2 < c+d < 1$ and $c \neq 1/2$. We construct a slope $\lambda_0 > 0$ and an open set of average states with a strict three-outcome advantage. Reflect labels and the posterior axis if necessary so $c < 1/2$, and put $k = (1/2 - c)/d \in (0, 1)$. Define

$$R(z) = \frac{d^2 s(z)^3}{(c+dz)(1-c-dz)}. \quad (\text{K1})$$

It has a unique strict maximum K_* at some $m < k$. Indeed, with $\eta = c + dz$, $a = c - d$ and $b = c + d$,

$$\frac{d^2 \log R}{d\eta^2} = -\frac{3}{2} [(\eta - a)^{-2} + (b - \eta)^{-2}] + \eta^{-2} + (1 - \eta)^{-2} < 0. \quad (\text{K2})$$

The sign follows termwise from $a > 0$, $b < 1$. Also R vanishes at both ends and $(\log R)'(k) = -3k/(1 - k^2) < 0$, where this last derivative is in z .

For $0 < \kappa < K_*$ write

$$F_\kappa(z) = h(c + dz) - \kappa s(z), \quad F_\kappa''(z) = \frac{\kappa - R(z)}{s(z)^3}. \quad (\text{K3})$$

If $u < v$ are the two roots of $R = \kappa$, then F_κ is convex on the two outer intervals and concave between them, with derivatives tending to $-\infty$ and $+\infty$ at the respective endpoints. For each $t \in (F_\kappa'(v), F_\kappa'(u))$ there are unique outer stationary points $z_1(t) < u$ and $z_2(t) > v$ of $F_\kappa - tz$. Their depth difference

$$L(t) = [F_\kappa(z_2) - tz_2] - [F_\kappa(z_1) - tz_1], \quad L'(t) = z_1 - z_2 < 0 \quad (\text{K4})$$

is positive at the lower slope endpoint and negative at the upper one. These signs follow by integrating $F_\kappa' - t$ between the limiting outer stationary points. Hence there is one slope for which the two minima have equal value. The corresponding common tangent $\ell(z) = \ell_0 + t_\kappa z$ lies below F_κ on the whole interval and touches it only at z_1, z_2 .

As $\kappa \uparrow K_*$ both contacts approach m . At K_* the function is strictly convex, its second derivative vanishing only at m ; a limiting support cannot have two separated contacts. Since $z_1 < u < v < z_2$ and $u, v \rightarrow m$, the common limit is m . Choose κ close enough that $z_2 < k$ and put $D_0 = F_\kappa - \ell$. Then $D_0 \geq 0$, with exactly those two zeros, and $D_0'' > 0$ on $[k, 1)$. The function

$$T(C) = D_0(C) - (C - k)D_0'(C) \quad (\text{K5})$$

has $T(k) > 0$, $T'(C) = -(C - k)D_0''(C) < 0$, and $T(C) \rightarrow -\infty$ as $C \rightarrow 1$. The last limit uses finite $D_0(1)$ and $D_0'(1-) = +\infty$. There is a unique root $z_3 \in (k, 1)$, and $D_0'(z_3) > 0$ because $D_0(z_3) > 0$. Set $\lambda_0 = D_0'(z_3)/(2d) > 0$ and

$$D(z) = \begin{cases} D_0(z), & z \leq k, \\ D_0(z) - 2\lambda_0 d(z - k), & z \geq k. \end{cases} \quad (\text{K6})$$

The right piece is strictly convex and has a double zero at z_3 ; the left piece has exactly the zeros z_1, z_2 . Thus $D \geq 0$ globally, with exactly three zeros. For $\alpha = \lambda_0 dk - \ell_0$ and $\beta = -\lambda_0 d - t_\kappa$,

$$D(z) = \alpha + \beta z - \kappa s(z) - g_{\lambda_0}(z). \quad (\text{K7})$$

This is a global dual certificate. Since $n_x \leq s(z)$ and $\kappa > 0$, it also bounds all complex pure-state atoms, with equality only at $(s(z_i), 0, z_i)$ for $i = 1, 2, 3$.

Choose any average-state point in the open interior of their triangle. Its barycentric weights are strictly positive, and Eq. (G2) gives faithful signal states. The contact criterion proves a strictly positive gap at λ_0 . The optimal POVM and PVM values are both Lipschitz in λ , because $0 \leq P_e \leq \min(p, 1 - p)$. The gap therefore persists on an open interval. The triangle gives an open set in the normal-form half disk; rotating about the posterior axis gives an open set of average states in the full Bloch ball.

For the converse, centered spectra have the common optimum just proved. If the spectrum lies entirely on one side of $1/2$, every posterior does also, so every measurement has error $\min(p, 1 - p)$. The established binary-qubit information theorem [6] supplies a PVM maximizing every support. This completes the spectral “if and only if” assertion, including spectra that merely touch $1/2$.

Appendix L: Numerical illustrations and open extensions— Figure 1 illustrates two sampled binary-qubit scenarios. For panel (b), the Bloch-frame signals are $\rho_0 = (I + u_0 \sigma_z)/2$ and $\rho_1 = [I + u_1(\cos \theta \sigma_z + \sin \theta \sigma_x)]/2$, with $p = 0.4072397056684426$, $u_0 = 0.9170949735598445$, $u_1 = 0.5130827427120952$, and $\theta = 0.7195921216129726$ rad. The scan uses 321 support weights, 1440 posterior-circle angles, and an 8192-point PVM mesh, with gap threshold 10^{-7} nat. Panel (a) is a 241-point numerical illustration. Both panels illustrate sampled behavior; the analytic existence proof is independent of them. The analytic results above establish the moment reduction, outcome-count criterion, and existence of an open set with a three-outcome advantage. For general ensembles, the number and connectivity of weight intervals with a nonprojective advantage remain open. Extending the Pareto analysis to nonfaithful states is another direction.

Appendix M: Noise model and recovery objective— The error-correction test uses a distance- d repetition code subject only to independent stochastic Pauli- X faults between syndrome rounds. A binary syndrome S is encoded in a fresh qubit pointer with conditional states ρ_0, ρ_1 ; C denotes all other records supplied to the decoder, and $E \in \{0, 1\}^d$ is the complete physical error vector. Receivers at different sites are classical-conditionally independent once the hidden history is fixed. The primary screen has $d = 3$, one to three rounds, and no ideal terminal syndrome. Starting with $E^{(0)} = 0$, round t first draws independent $B_j^{(t)} \sim \text{Bernoulli}(f)$ and sets $E^{(t)} = E^{(t-1)} \oplus B^{(t)}$. The adjacent-check syndromes are $S_j^{(t)} = E_j^{(t)} \oplus E_{j+1}^{(t)}$, for $j = 1, \dots, d-1$. Each check emits a fresh pointer with outcome law $\Pr(y_j^{(t)} | E^{(t)}) = \text{Tr}(\rho_{S_j^{(t)}} M_{j, y_j^{(t)}}^{(t)})$. Checks are read in increasing j , with no intervening data faults; no additional fault interval follows the last round. The target vector is $E = E^{(R)}$, and the decoder retains the full time-ordered record. Here f denotes the per-round fault probability, written p in the numerical protocol below, and is distinct from the signal prior.

For a measurement $M = \{M_y\}$, the joint probability needed by the decoder is

$$P(E = e, C = c, Y = y) = \text{Tr}(A_{ec} M_y), \quad A_{ec} = \sum_s P(e, c, s) \rho_s. \quad (\text{M1})$$

The decoder chooses a record-dependent Pauli- X correction. In the declared model, the repetition-code stabilizers are Z type, so their intersection with the X -only error group is the identity. Acting on a logical-reference Bell state, recovery therefore has unit entanglement fidelity exactly when the chosen correction equals the complete error vector. Maximizing separately for each record gives

$$F_{\text{rec}}(M) = \sum_{c, y} \max_e \text{Tr}(A_{ec} M_y), \quad R_{\text{rec}}(M) = 1 - F_{\text{rec}}(M). \quad (\text{M2})$$

This identity relies on the X -only error set and the specified recovery class. It is not used as a circuit-level identity for general stabilizer codes, coherent faults, leakage, or noisy syndrome extraction.

Conditioning only on the pointer outcome gives $q_y = P(y)$ and $t_y = P(S = 0 | y)$, and Bayes' rule rewrites Eq. (M2) as an average of a posterior loss:

$$\begin{aligned} R_{\text{rec}}(M) &= \sum_y q_y r(t_y), \\ r(t) &= \sum_c \left[\sum_e b_{ec}(t) - \max_e b_{ec}(t) \right], \\ b_{ec}(t) &= tP(e, c | 0) + (1-t)P(e, c | 1). \end{aligned} \quad (\text{M3})$$

The function r is concave and piecewise linear because it is a sum of an affine function and the negative of a pointwise maximum of affine functions. Thus recovery error is also an average of a function of the posterior probability. That function is determined by the decoder and the other measurements, rather than by entropy or a single binary decision.

The factorization used here requires that the chosen pointer's state depends only on S , with the remaining receiver settings fixed. It gives a design-time loss averaged over C , not a receiver that conditions its setting on future observations. If a different receiver site is subsequently changed, the joint distribution and its induced loss must be recomputed. Optimizing each coordinate separately does not certify a global schedule optimum. The classical posterior t_y is the common statistical variable; this derivation does not identify Q with the information-optimized generator $H_\star$ or extend the affine exactness theorem to every loss.

Appendix N: Numerical protocol and results— Four pointer families were used. The asymmetric controls are $\rho_0 = (I + 0.8Z)/2$ and $\rho_1 = [I + 0.6(\cos \theta Z + \sin \theta X)]/2$ for $\theta = 60^\circ, 120^\circ$. The symmetric controls are

$$\rho_s = \frac{1}{2} \left[I + 0.99 \left(kZ + (-1)^s \sqrt{1 - k^2} X \right) \right], \quad k \in \{0.5, 0.8\}. \quad (\text{N1})$$

For each family, the primary screen used $R = 1, 2, 3$ and fault probabilities $p = 0.005, 0.02, 0.05$. One receiver site at a time was replaced while all other sites remained at their round-dependent local mutual-information PVMs. Every round has two receiver sites. The 36 cases contain 144 one-site rows, and each case reports the lowest risk over its tested single-site replacements. Improvement counts use this casewise minimum and the 10^{-7} absolute-risk criterion shown in the table. PVM directions were searched on the real Bloch great circle and locally refined; generalized-POVM candidates were then compared at the same fixed environment. The reported values use float64 arithmetic. Saved schedules were reevaluated as described below; the finite searches do not certify continuum optima.

TABLE I. Primary distance-three, no-terminal transfer screen. “Endpoint” is the better of the local mutual-information and Helstrom-spectral PVMs.

comparison	declared result
task PVM improves on local MI by $> 10^{-7}$	30/36
task PVM improves on the better endpoint by $> 10^{-7}$	18/36
largest relative risk reduction from local MI	3.8768245%
largest absolute risk reduction from local MI	0.007196346862
largest resolved same-site PVM-to-POVM increment	0.000286710021

The largest relative case is the symmetric $k = 0.5$, $R = 3$, $p = 0.02$ instance: $R_{\text{MI}} = 0.053161858761$ and $R_{\text{task}} = 0.051100866781$. The example showing why the unconditional local pair (I, P_e) does not fully predict recovery performance is the asymmetric 60° , $R = 3$, $p = 0.02$ case at the final site. There

$$\begin{aligned} I_{\text{MI}} &= 0.04384833735 \text{ bit}, & I_{\text{task}} &= 0.04195720460 \text{ bit}, \\ P_e^{\text{MI}} &= P_e^{\text{task}} = 0.10862110515, & R_{\text{rec}} &: 0.15948611110 \longrightarrow 0.15853944866. \end{aligned} \quad (\text{N2})$$

The recovery-optimized receiver therefore has worse local information and the same local error, yet reduces the final recovery risk by 0.593570%. The decoder also uses the time-ordered records, which the two local numbers do not capture. This does not violate data processing or establish a Blackwell ordering of the measurements.

Fixed-site comparison and numerical gap: For a fixed environment, let a be a decoder action assigning an error estimate $a(c)$ to each context record, and define the positive cost matrix $W_a = \sum_c [\sum_e A_{ec} - A_{a(c),c}]$. Then $R_{\text{rec}}(M) = \sum_y \min_a \text{Tr}(W_a M_y)$. For a two-outcome PVM, optimizing its two decoder actions gives

$$R_{\text{PVM}}^{\min} = \min_{a,b} \frac{\text{Tr}(W_a + W_b) - \|W_a - W_b\|_1}{2}. \quad (\text{N3})$$

This is the binary minimum-cost spectral construction for each action pair. Action pairs include $a = b$, so constant decisions are represented. The real pointer states make all W_a real, so real spectral projectors suffice even when complex qubit PVMs are allowed. Grouping effects by decoder action gives the general-POVM primal problem $\min \sum_a \text{Tr}(W_a E_a)$ with $E_a \succeq 0$, $\sum_a E_a = I$; its dual is $\max \text{Tr} \Gamma$ with $\Gamma \preceq W_a$ for every action. The reported fixed-site gap is the feasible receiver risk minus this dual lower bound, evaluated in float64 after a uniform feasibility shift of Γ . The same-site increment in the table compares that receiver with Eq. (N3), rather than a joint-schedule optimum. The displayed residuals are numerical evaluations of these bounds; outward rounding would be required for rigorous floating-point enclosures.

Joint schedules, context, and perturbation diagnostics: Panel (a) of Fig. 5 reports a separate joint optimization of the measurement schedule, with eight distance-three cases and four distance-five follow-up cases at $p = 0.02$. Coordinate restarts and, for the distance-three audit, differential evolution produce feasible fixed-PVM schedules. Their best-found reductions relative to the local-MI schedule range from 0.0311% to 8.4040% at distance three; the distance-five follow-up reaches 10.9412%. These are the best schedules found in finite searches. Their protocol and optimality guarantees differ from those of the 30/36 single-site comparison, so the two sets of results are reported separately.

The comparison of available measurement records in panel (c) uses the symmetric $k = 0.5$, distance-three, two-round model. A fixed task-aware receiver uses the full decoder context only through the averaged design loss. A causal policy may instead switch the measurement using earlier records, assuming ideal zero-latency, zero-cost switching. The all-context oracle may additionally condition on later records and is noncausal except at the final site. For each saved case the risk decomposition

$$R_{\text{MI}} - R_{\text{oracle}} = (R_{\text{MI}} - R_{\text{fixed}}) + (R_{\text{fixed}} - R_{\text{causal}}) + (R_{\text{causal}} - R_{\text{oracle}}) \quad (\text{N4})$$

is a telescoping allocation of the reported risk differences; its rounding residual is below 10^{-12} . Causal settings may depend only on records preceding the target site, whereas oracle settings may depend on all other records. This distinguishes design-time averaging from granting additional information at measurement time.

Panel (d) compares four symmetric distance-three cases over a six-member uncertainty set including the nominal model. Around $(f, v, k) = (0.02, 0.99, k_0)$, the other five members change only one parameter: $f \mapsto 0.9f$ or $1.1f$, $v \mapsto 0.95v$, and $k \mapsto 0.95k$ or $1.05k$, with $k_0 = 0.5$ or 0.8 . The pointer form is Eq. (N1) with 0.99 replaced by v . A nominal task-PVM schedule can incur worst regret of about 1.15×10^{-2} in a $k = 0.8$, three-round case, and freezing the decoder can make performance worse. The finite-set minimax search uses seeds 424242 and 8675309, with differential

evolution limited to 500 iterations, population factor 20, and absolute and relative tolerances 10^{-10} . Its candidates improve worst regret against the fixed nominal-MI baseline in all four cases, but against separately retuned MI baselines only the $k = 0.8$ cases improve. The comparison tests robustness at these six parameter choices. It does not determine the response to continuous calibration drift.

Appendix O: Numerical consistency and model scope— The transfer-test calculations use exact finite hidden-history enumeration within the specified stochastic model. The numerical implementation passed 24 unit tests; saved fixed-site gaps were at most 2.1×10^{-9} , and reevaluation of saved schedules reproduced the risks within absolute tolerance 10^{-11} . Fixed-site certificates are evaluated in float64 arithmetic, while joint-schedule and finite-set minimax searches provide feasible candidate schedules.

The recovery results apply to the synthetic repetition-code model and the specified receiver and decoder families. Circuit-level noise, continuous calibration drift, code thresholds, and physical readout backaction are separate questions for extending this model.

Appendix P: Complete numerical diagnostics— Figures 4 and 5 show the complete numerical diagnostics for the receiver benchmarks and the decoder comparisons, respectively.

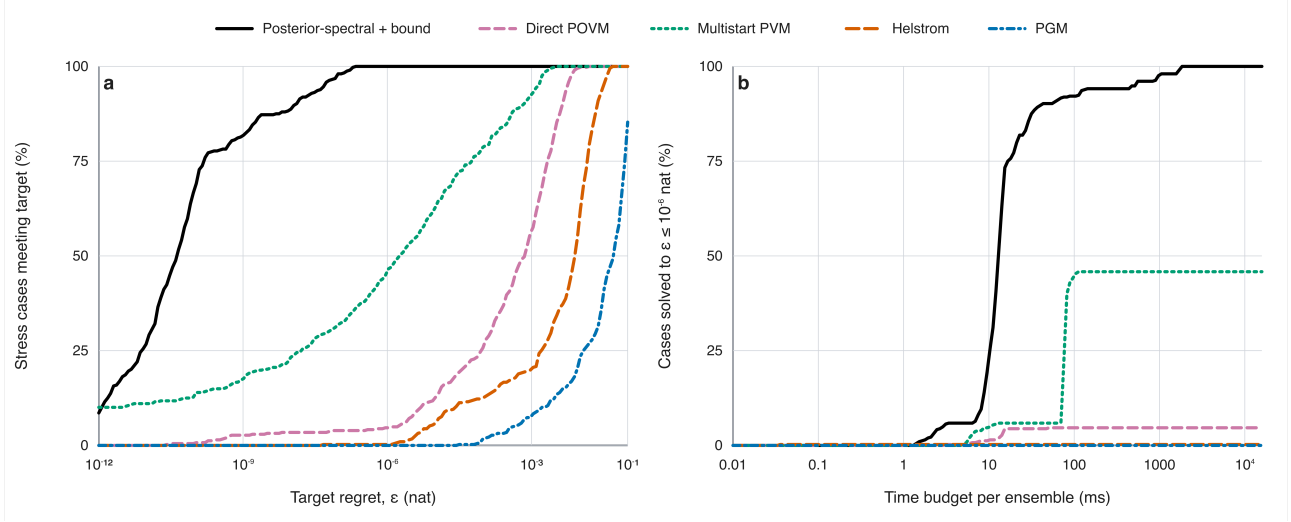


FIG. 4. Accuracy and runtime profiles for the 408-case suite. The posterior curve evaluates its analytic information bound in float64; other curves show information loss relative to its PVM. The comparisons use the fixed implementations and iteration budgets described in Appendix E.

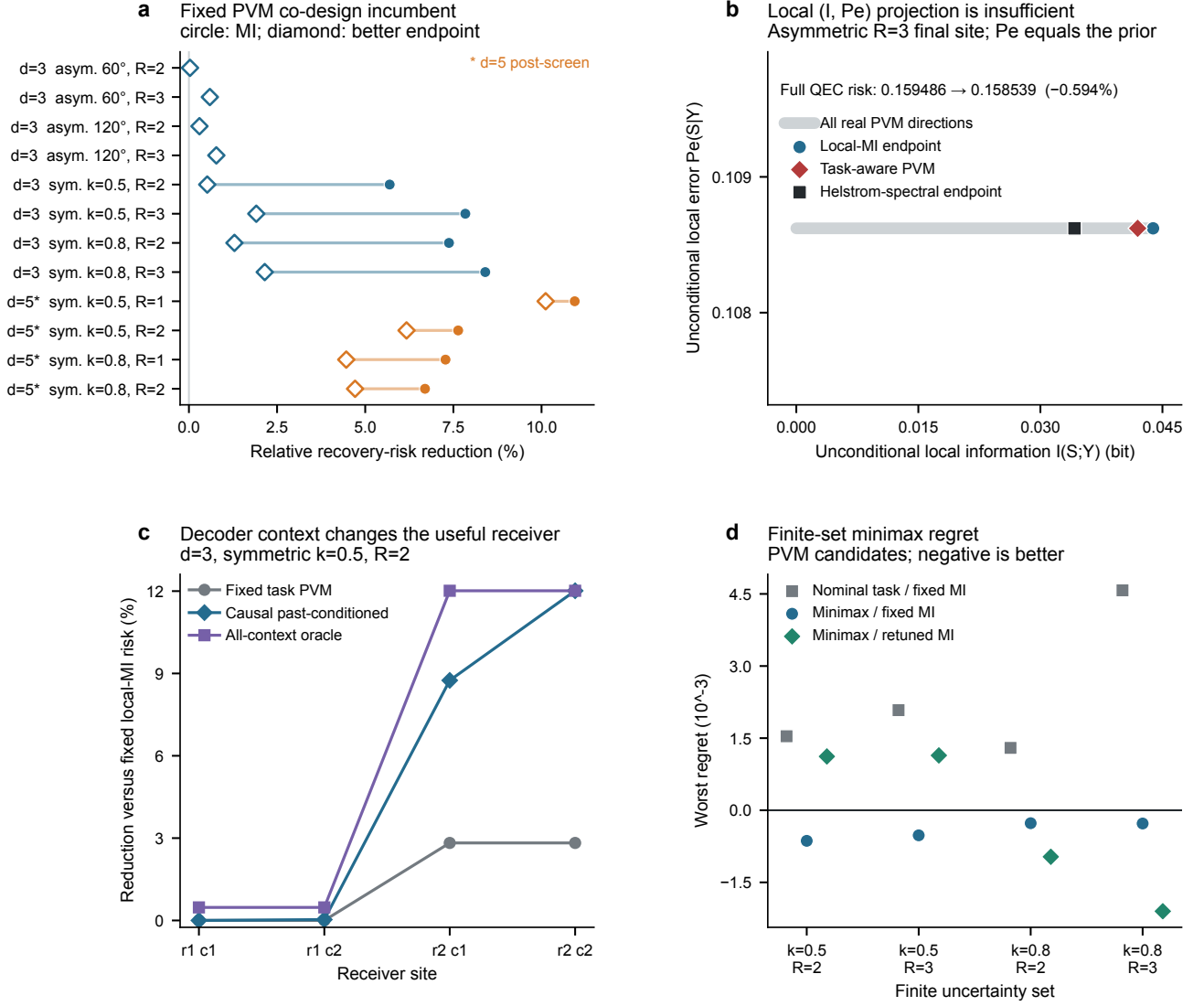


FIG. 5. Full decoder diagnostics: (a) best-found fixed-PVM joint schedules; starred distance-five rows are follow-ups; (b) lower local information and unchanged local error accompany improved recovery; (c) fixed receivers, zero-latency causal switching, and a generally noncausal all-context oracle; (d) worst regret over the six-member uncertainty set defined in Appendix N. The plotted schedules are the best found in finite searches, and the robustness comparison uses only the specified parameter set. Risks are deterministic within the stated stochastic model.