

The Shot Gap: Bloch-Sphere Geometry and the Hidden Measurement Budget of Single-Qubit Quantum Classifiers

Alfredo Merlet

alfredo.merlet@gmail.com

Universidad del Alba <https://orcid.org/0009-0007-8558-164X>

Research Article

Keywords: Bloch sphere, Helstrom bound, quantum machine learning, data re-uploading, shot budget, evaluation of AI systems

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11435727/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.

Abstract

Single-qubit data re-uploading classifiers map every input to a point on the Bloch sphere and are routinely evaluated by thresholding an expectation value estimated from many repetitions of the circuit (shots). A single measurement, however, can deliver at most the Helstrom accuracy of the two class-conditional mixtures, which for one qubit reduces to the length of the difference between the class-mean Bloch vectors. We call the difference between thresholded accuracy and single-shot accuracy the shot gap and measure it in a computational experiment with known ground truth: four synthetic tasks with exact Bayes accuracy, five circuit depths, three training losses, two label-noise levels and ten seeds (1,200 runs), simulated exactly and validated against a state-vector simulation, finite-difference gradients and a brute-force search over measurements. The Helstrom bound held in every run. For circuits of depth two or more, models whose thresholded accuracy averaged 0.857 delivered 0.696 per single measurement, a mean shot gap of 0.161; at depth six, models with 0.93–0.98 thresholded accuracy on the linear, circle and XOR tasks delivered 0.67–0.86 per shot. The embedding, not the choice of measurement, accounted for 96.3% of the single-shot shortfall. Training with the single-shot log-likelihood narrowed the shot gap relative to a smoothed hinge loss in 89% of paired configurations (median reduction 0.085) at a median cost of 0.011 in thresholded accuracy, and eleven shots recovered 74% of the gap against 41% under the hinge loss. Two preregistered hypotheses were not supported: the gap did not grow with boundary nonlinearity, and training loss predicted it better than class-mean Bloch-vector length. Accuracy reports for quantum classifiers should state the number of shots per prediction, as they state training data and compute.

1. Introduction

A single qubit is the smallest object on which a quantum classifier can be built, and it is also the most transparent. Every pure state of one qubit is a unit vector on the Bloch sphere, every single-qubit gate is a rotation of that vector, and every projective measurement asks on which side of a plane through the origin the vector lies (Nielsen & Chuang, 2010). Data re-uploading classifiers exploit this geometry by interleaving data-dependent rotations with trainable ones, so that a single qubit can approximate rich decision boundaries (Pérez-Salinas et al., 2020). Subsequent work has shown that such models can match classical baselines on small datasets (Peña Tapia et al., 2023) and has proposed purely classical training of single-qubit models (Aminpour et al., 2026).

How these classifiers are scored matters as much as how they are built. In the standard protocol the circuit is run many times for each input, the expectation value of the measured observable is estimated, and the predicted label is the sign of that estimate. The reported accuracy is therefore the accuracy of a thresholded expectation value, which presupposes that the input can be re-encoded and measured as many times as needed. A different question is what the classifier delivers from a single preparation and a single measurement. Quantum state discrimination answers that question exactly: the probability of a correct guess from one copy of an unknown state drawn from one of two mixtures is bounded by the

Helstrom limit (Helstrom, 1969), and for a single qubit that limit depends only on the class-mean Bloch vectors.

This article measures the distance between those two notions of accuracy, which we call the shot gap, in a setting where every reference quantity is known. Synthetic tasks with symmetric label noise give an exact Bayes accuracy; an exact Bloch-vector simulator gives exact accuracies at any number of shots; and the Helstrom bound gives an exact ceiling for single-shot performance that is independent of the measurement actually trained. Three contributions follow. First, a closed-form decomposition of the single-shot shortfall of a single-qubit classifier into an embedding loss and a measurement loss, both non-negative, with an exact mapping between clean and noisy labels. Second, a factorial experiment of 1,200 training runs that quantifies the shot gap across tasks, depths, losses and label noise. Third, a test of four preregistered hypotheses, two of which were not supported, reported as such.

The motivation is partly methodological and partly a matter of evaluation practice. In the evaluation of artificial-intelligence systems more broadly, a reported score is interpretable only together with the resources consumed to produce it, such as training data, compute or the number of attempts allowed at inference. Shots are an inference-time resource of exactly this kind. The analogy is developed in the Discussion and is labelled there as an analogy; the results themselves concern single-qubit classifiers only.

2. Related work

The literature relevant to this study was mapped with five Scopus queries on 19 October 2026: the Bloch sphere combined with machine learning or classifiers (41 records); data re-uploading or single-qubit classifiers (116); quantum classifiers combined with Helstrom, state discrimination or trace distance (38); quantum classifiers combined with shot noise, number of shots or measurement budget (99); and the conjunction of the last two themes (0). The zero was checked by applying the logic of the conjunctive query to every record of the other exports: no record satisfied it, and the third and fourth exports shared no records. The two literatures closest to this article therefore exist side by side without intersecting.

On the discrimination side, Zhang and Zhuang (2022) showed that the classification error of extensive variational circuits on quantum data decays with depth until it saturates at the Helstrom limit, provided that the outputs are post-processed by maximum likelihood. Lee et al. (2024) trained a variational circuit to implement the minimum-error measurement and used it as a supervised classifier. Du et al. (2023) related the power of quantum neural networks on multiclass problems to properties of the embedded states, and Mordacci et al. (2025) trained encodings to increase the trace distance between classes. Jerbi et al. (2023) placed re-uploading models in a broader landscape of quantum learning models beyond kernel methods. None of these studies compares single-shot performance with the thresholded accuracy that is usually reported for classical-data classifiers.

On the measurement-budget side, shot noise has been studied mainly as a cost or an obstacle during training. Phalak and Ghosh (2023) reduced the number of shots needed to train quantum machine-

learning architectures, Kim et al. (2025) adapted the shot allocation dynamically in variational algorithms, Scriva et al. (2024) quantified how shot noise inflates the resources needed by variational optimisation, Turab et al. (2026) analysed convergence with finite-shot oracles, and Etar et al. (2025) documented the degradation of quantum neural networks trained under increasing shot noise. The closest single-qubit study is that of Tsili et al. (2025), who trained a single-qubit classifier using a single measurement per input; their concern is the sampling used during training rather than the accuracy available per prediction after training. The present article connects the two literatures by using the Helstrom bound as the reference for what a single measurement can deliver from a trained embedding, and by expressing the usual thresholded accuracy as the limit of a finite-shot majority vote.

3. Theory

3.1 A single-qubit classifier on the Bloch sphere

Let $x \in \mathbb{R}^2$ be an input with label $y \in \{0, 1\}$, and write $s = +1$ for class 0 and $s = -1$ for class 1. A depth- L data re-uploading circuit applies L layers $U(\varphi_L) = R_z(\varphi_{L1}) R_y(\varphi_{L2}) R_z(\varphi_{L3})$ to $|0\rangle$, with $\varphi_L = \theta_L + (w_{L1} x_1, w_{L2} x_2, 0)$, following Pérez-Salinas et al. (2020). Each layer has five trainable parameters. Because the output state is pure, it is fully described by its Bloch vector $r(x) = (\langle X \rangle, \langle Y \rangle, \langle Z \rangle)$, with $\|r(x)\| = 1$, and each rotation acts on $r(x)$ as an element of $SO(3)$. The circuit is measured in the Z basis, so a single shot returns outcome 0 with probability $p_0(x) = (1 + z(x))/2$, where $z(x)$ is the third component of $r(x)$. Since the last layer is trainable, measuring Z after it is equivalent to measuring along a trainable axis.

3.2 Single-shot accuracy and the Helstrom bound

If the classifier is used once per input and outputs the measured outcome as its label, its accuracy is $A_1 = E[(1 + s \cdot z(x))/2]$. Writing π_c for the class priors and $\bar{r}_c$ for the mean Bloch vector of class c , this equals

$$A_1 = \frac{1}{2} + \frac{1}{2} (\pi_0 \bar{z}_0 - \pi_1 \bar{z}_1).$$

The two class-conditional states seen by the measurement are the mixtures $\rho_c = (I + \bar{r}_c \cdot \sigma)/2$. By Helstrom's theorem, no two-outcome measurement distinguishes them with probability greater than $\frac{1}{2} + \frac{1}{2} \|\pi_0 \rho_0 - \pi_1 \rho_1\|_1$. For one qubit, the operator $\pi_0 \rho_0 - \pi_1 \rho_1$ has eigenvalues $((\pi_0 - \pi_1) \pm \|v\|)/2$ with $v = \pi_0 \bar{r}_0 - \pi_1 \bar{r}_1$, which gives the closed form

$$A_H = \frac{1}{2} + \frac{1}{2} \max(|\pi_0 - \pi_1|, \|\pi_0 \bar{r}_0 - \pi_1 \bar{r}_1\|),$$

and, for balanced classes, $A_H = \frac{1}{2} + \|\bar{r}_0 - \bar{r}_1\|/4$. Proposition 1 follows: for every trained classifier, $A_1 \leq A_H$, because the trained measurement realises the component v_z while the bound uses the full length $\|v\|$. A_H is a property of the embedding alone. It shrinks whenever the states of one class spread over the sphere, because spreading shortens the class-mean vector even when each individual state lies on the surface.

3.3 Thresholded and finite-shot accuracy

The conventional score thresholds the expectation value: $A_\infty = E[1\{s \cdot z(x) > 0\}]$. It corresponds to an unlimited number of shots per prediction. With S shots and a majority vote (S odd), the probability of a correct label for input x is the upper tail of a binomial distribution with S trials and success probability $(1 + s \cdot z(x))/2$, and A_S is its expectation over inputs. A_S equals A_1 at $S = 1$ and converges to A_∞ as S grows, at a speed governed by the distribution of the margins $|z(x)|$: states near the poles are classified correctly almost every time, states near the equator almost at random. We define the shot gap as $A_\infty - A_1$. The ordering $A_1 \leq A_\infty$ is not guaranteed in general, because a model that is wrong with small margins on some region can be worse when thresholded than when sampled. In the experiment, $A_\infty < A_1$ occurred in 126 of the 1,200 runs; their thresholded accuracy had a median of 0.497, so they were models that had not learned the task (83 of them on the checkerboard and 73 at depth 1).

3.4 Exact decomposition and label noise

All accuracies above are computed against the noise-free labelling rule, for which the Bayes accuracy is 1. The single-shot shortfall then decomposes exactly as

$$1 - A_1 = (1 - A_H) + (A_H - A_1),$$

where the first term is the embedding loss and the second the measurement loss; both are non-negative by Proposition 1. When labels are flipped with probability η independently of x , every population accuracy A measured against the noisy labels equals $\eta + (1 - 2\eta)A$, the Bayes accuracy becomes $1 - \eta$, and the Helstrom bound transforms in the same way, because the noisy class-weighted difference of mean Bloch vectors is $(1 - 2\eta)$ times the clean one (Proposition 2). Label noise therefore affects the results only through training, and all figures and tables report clean-label accuracies, which can be converted exactly to any η .

4. Methods

4.1 Tasks, model and training

Inputs were drawn uniformly from the square $[-1, 1]^2$. Four deterministic rules each assign exactly half of the square to class 0: a linear boundary ($x_1 > x_2$), a circle ($x_1^2 + x_2^2 < 2/\pi$), XOR ($x_1 x_2 > 0$) and a 4×4 checkerboard. Each training set had 400 points; each task had a fixed test set of 20,000 points. Training labels were flipped with probability $\eta \in \{0, 0.1\}$.

Three training losses were compared, all functions of the signed margin $m(x) = s \cdot z(x)$. The fidelity cost of Pérez-Salinas et al. (2020) minimises the mean of $((1 - m)/2)^2$, which is equivalent to the squared error between $\langle Z \rangle$ and the label. The single-shot log-likelihood minimises the mean of $-\log((1 + m)/2)$, smoothed by $\delta = 10^{-6}$; it is the negative log-probability that one measurement returns the correct label. The smoothed hinge minimises the mean of $\text{softplus}(k(\mu - m))/k$ with $k = 20$ and margin $\mu = 0.1$; it

rewards only the sign of $\langle Z \rangle$ beyond a small margin and is the natural loss for the thresholded score. Parameters θ were initialised uniformly in $(-\pi, \pi)$ and weights w from a standard normal distribution, and were trained with Adam (Kingma & Ba, 2015) for 400 full-batch steps at learning rate 0.05. Gradients were computed by automatic differentiation of the exact Bloch-vector simulator.

4.2 Design and metrics

The design crossed 4 tasks $\times$ 5 depths ($L = 1, 2, 3, 4, 6$) $\times$ 3 losses $\times$ 2 noise levels $\times$ 10 seeds, for 1,200 runs. Within each combination of task, depth, noise level and seed, the three losses shared the same training data and the same initial parameters, so that losses can be compared in paired form. For every trained model the test set yielded A_1, A_H, A_∞, A_S for $S \in \{3, 5, 11, 21, 51, 101, 1001\}$, the lengths of the two class-mean Bloch vectors, the mean absolute margin and the final training loss.

4.3 Hypotheses

Four hypotheses were specified before the factorial run was launched; a single-seed pilot of nine configurations on the circle task preceded them. H1 is a check of Proposition 1 and is reported as validation, not as a finding: $A_1 \leq A_H$ in every run. H2: the shot gap increases with the nonlinearity of the decision boundary, ranked linear $<$ circle $<$ XOR $<$ checkerboard. H3: the mean length of the class-mean Bloch vectors predicts the shot gap better than the final training loss, within each loss function. H4: the single-shot log-likelihood yields a smaller shot gap than the other two losses, at some cost in thresholded accuracy. H2 and H3 were assessed with Spearman correlations and H4 with paired Wilcoxon signed-rank tests over the 400 matched configurations.

4.4 Validation

Three checks preceded the factorial run. Automatic gradients agreed with central finite differences for all three losses at depths 1, 3 and 6, with a maximum relative discrepancy of 4.5×10^{-8} . The Bloch-rotation simulator agreed with an explicit simulation of 2×2 unitaries acting on state vectors to within 7.8×10^{-16} . The closed-form Helstrom bound agreed with the eigenvalue formula and with a brute-force search over 260,000 measurement directions to six decimals (0.533290 in all three).

5. Results

5.1 The Helstrom bound held in every run (H1)

No run had $A_1 > A_H$; the largest value of $A_1 - A_H$ was -4.4×10^{-7} . The measurement loss $A_H - A_1$ was small: its median was 0.0024 and its mean 0.012. Across all runs, the embedding loss $1 - A_H$ accounted for 96.3% of the total single-shot shortfall. Training therefore found an almost Helstrom-optimal measurement for whatever embedding it produced, and the limit on single-shot accuracy came from the geometry of the embedded states on the Bloch sphere, not from the choice of measurement (Fig. 1).

5.2 The size of the shot gap

The shot gap was large and widespread. For depths of two or more, thresholded accuracy averaged 0.857 and single-shot accuracy 0.696, a mean gap of 0.161, and 79.2% of these runs had a gap above 0.10. The gap grew with depth from 0.044 at $L = 1$, where most models barely learned, to 0.168 at $L = 6$ (Fig. 2). Label noise in training increased it from 0.123 to 0.153 on average. Table 1 reports the depth-6 results: models scoring 0.93–0.98 when thresholded on the linear, circle and XOR tasks delivered 0.67–0.86 from a single measurement.

Table 1

Clean-label accuracies at depth 6, by task and training loss (means over 20 runs: two noise levels $\times$ ten seeds; standard deviation of the gap in parentheses). A_1 : one shot; A_H : Helstrom bound of the trained embedding; A_{11} and A_{101} : majority vote of 11 and 101 shots; A_∞ : threshold on $\langle Z \rangle$.

Task · loss	A_1	A_H	A_{11}	A_{101}	A_∞	Shot gap (SD)
Linear · Fidelity	0.862	0.864	0.954	0.974	0.978	0.116 (0.020)
Linear · Single-shot LL	0.861	0.865	0.946	0.965	0.969	0.108 (0.028)
Linear · Smoothed hinge	0.717	0.775	0.851	0.954	0.980	0.263 (0.097)
Circle · Fidelity	0.776	0.779	0.893	0.933	0.940	0.164 (0.026)
Circle · Single-shot LL	0.780	0.784	0.887	0.921	0.926	0.146 (0.023)
Circle · Smoothed hinge	0.670	0.685	0.810	0.921	0.963	0.293 (0.048)
XOR · Fidelity	0.821	0.823	0.920	0.947	0.952	0.130 (0.017)
XOR · Single-shot LL	0.819	0.822	0.912	0.936	0.940	0.121 (0.019)
XOR · Smoothed hinge	0.699	0.714	0.830	0.927	0.962	0.263 (0.075)
Checkerboard · Fidelity	0.688	0.690	0.786	0.823	0.830	0.142 (0.023)
Checkerboard · Single-shot LL	0.682	0.685	0.766	0.792	0.796	0.114 (0.019)
Checkerboard · Smoothed hinge	0.560	0.565	0.620	0.682	0.715	0.155 (0.047)

5.3 The gap did not grow with boundary nonlinearity (H2 not supported)

H2 predicted a positive association between boundary nonlinearity and the shot gap. The observed Spearman correlation was negative ($\rho = -0.42$, $p < 10^{-50}$), and it was negative at every depth, weakening to $\rho = -0.09$ at $L = 6$. The checkerboard, the most nonlinear task, had the smallest gap (0.046 on average) because most models learned it poorly: a model that is near chance when thresholded has little thresholded accuracy to lose. H2 is therefore not supported. An exploratory analysis, not preregistered, restricted to models with $A_\infty > 0.6$ and expressed the gap as a share of above-chance thresholded accuracy, $(A_\infty - A_1)/(A_\infty - \frac{1}{2})$. That relative gap showed a weak positive association with nonlinearity

($\rho = 0.17$), which suggests that the preregistered test was dominated by learnability; it is reported for completeness and does not rescue H2.

5.4 Training loss predicted the gap better than Bloch-vector length (H3 not supported)

Within each loss function, the mean length of the class-mean Bloch vectors was associated with the shot gap in the expected direction, longer vectors going with smaller gaps ($\rho = -0.46$ for the fidelity cost, -0.26 for the single-shot log-likelihood and -0.24 for the smoothed hinge). The final training loss, however, was more strongly associated with the gap in all three cases ($\rho = -0.58$, -0.61 and -0.61). H3 is not supported. The class-mean length summarises the geometry relevant to a single shot, while the gap also depends on how far thresholded accuracy rises, which the training loss tracks more closely.

5.5 The single-shot log-likelihood narrowed the gap (H4 supported)

In paired comparisons, the single-shot log-likelihood produced a smaller shot gap than the smoothed hinge in 89.2% of the 400 matched configurations (median difference -0.085 ; Wilcoxon $p < 10^{-59}$) and a smaller gap than the fidelity cost in 78.0% (median -0.008 ; $p < 10^{-31}$). The cost in thresholded accuracy was small: a median of 0.011 relative to the hinge and 0.008 relative to the fidelity cost (both $p < 10^{-16}$). Single-shot accuracy was indistinguishable from that obtained with the fidelity cost (median difference 0.000; $p = 0.39$) and higher than with the hinge (median 0.070; $p < 10^{-59}$). H4 is supported (Fig. 3). The hinge is the most consequential case: it optimises exactly the quantity that is usually reported, and it produced the best thresholded accuracies at depth 6 on three tasks and the largest gaps of all, between 0.26 and 0.29 on the linear, circle and XOR tasks.

5.6 How many shots close the gap

Figure 4 reports the median share of the shot gap recovered by majority voting, for runs with a gap above 0.02. Under the fidelity cost and the single-shot log-likelihood, three shots recovered about 35% of the gap, eleven shots 71–74%, 101 shots 96% and 1,001 shots more than 99%. Under the smoothed hinge the same budgets recovered 14%, 41%, 85% and 98%. A model trained for the thresholded score thus needs roughly an order of magnitude more shots per prediction to approach the accuracy it reports.

6. Discussion

The central result is simple to state. A single-qubit classifier evaluated by thresholding $\langle Z \rangle$ reports the accuracy of an idealised procedure with unlimited repetitions. What it can deliver per measurement is bounded by a quantity that can be read from the Bloch sphere, the separation between the class-mean Bloch vectors, and in this experiment that bound sat on average 16 points below the reported score among models of depth two or more. Because the trained measurement was almost always Helstrom-optimal, no better choice of measurement basis could have closed the difference. The gap is a property

of the embedding, and it is created by losses that reward the sign of $\langle Z \rangle$ rather than the polarisation of the state.

Three practical consequences follow. First, the Helstrom bound of a trained single-qubit classifier is cheap to audit. It requires only the class-mean expectation values of X, Y and Z, which can be estimated by measuring in three bases, and it requires no access to the circuit. Second, accuracy should be reported together with the number of shots per prediction, and preferably as a curve such as Fig. 4, because the same model can be described truthfully as 0.98 accurate or 0.72 accurate depending on a budget that is rarely stated. Third, when single-shot or few-shot use is intended, the training loss should be chosen accordingly: the single-shot log-likelihood reduced the gap substantially at a cost of about one point of thresholded accuracy.

By analogy, and only as an analogy, the finding resembles a familiar problem in the evaluation of artificial-intelligence systems, where scores obtained with many attempts or with heavy inference-time computation are compared with scores obtained from a single attempt without stating the difference. Shots are the inference-time compute of a quantum classifier. Evaluation practice that already asks for the training data and compute behind a model could ask, for quantum models, for the shot budget behind each reported accuracy. This recommendation does not depend on any claim about quantum advantage, and the experiment does not test any such claim.

6.1 Limitations

The study is deliberately narrow. It uses one qubit, one ansatz, two-dimensional synthetic inputs drawn from a uniform distribution, and exact noiseless simulation; hardware noise would shorten Bloch vectors further and would probably enlarge the gap, but this was not tested. The Helstrom bound in Section 3 is specific to a single measurement of a single copy; collective measurements on several copies follow different bounds and were not considered. The test of H2 was confounded by learnability, as the checkerboard was rarely learned at the depths studied. The results on loss functions concern three losses and one optimiser with a fixed budget of 400 steps. Finally, multi-qubit classifiers with entangling gates do not admit the same Bloch-sphere picture, although the decomposition into embedding and measurement losses carries over with the Helstrom bound computed from the class-conditional density matrices.

7. Conclusion

Single-qubit classifiers are scored on an expectation value, but used, if they are used one measurement at a time, on a single outcome. The difference between the two, the shot gap, averaged 0.16 accuracy points across 960 trained models of depth two or more, was explained almost entirely by the geometry of the embedded states on the Bloch sphere, and was largest precisely for the loss that optimises the conventional score. The Helstrom bound of the class-mean Bloch vectors gives a measurement-

independent ceiling that any reader can check, and the shot budget per prediction should accompany every reported accuracy.

Declarations

Competing interests.

The author declares no competing interests.

Ethics approval.

Not applicable. The study uses only synthetic data and involves no human participants, animals or personal data.

Use of AI tools.

The author used Claude (Anthropic) as an assistant for code development, analysis scripting and language editing. The author reviewed all code, results and text and takes full responsibility for the content of the article.

Funding.

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author contributions.

A.A.M.E. conceived the study, designed the experiment, carried out the analysis and wrote the manuscript.

Data availability.

All data are generated by code from fixed seeds. The per-run results, the figures and the scripts that reproduce them are openly available on Zenodo under a CC BY 4.0 licence at <https://doi.org/10.5281/zenodo.23481776>.

Code availability.

The exact simulator, task generators, training and analysis scripts, the finite-difference gradient check and the physics validation scripts are included in the same Zenodo record (<https://doi.org/10.5281/zenodo.23481776>).

References

1. Aminpour S, Banad Y, Sharif S (2026) Learning with a single qubit: Classical training of quantum-inspired classifiers for structured tasks. *Proceedings of SPIE*, 13919, 1391900. <https://doi.org/10.1117/12.3081062>
2. Du Y, Yang Y, Tao D, Hsieh M-H (2023) Problem-dependent power of quantum neural networks on multiclass classification. *Phys Rev Lett* 131(14):140601. <https://doi.org/10.1103/PhysRevLett.131.140601>
3. Etar A, Tripathi S, Soni J, Upadhyay H, Perez-Pons A (2025) Study of effects of shot noise on quantum machine learning models. In *Proceedings of the 2025 Conference on AI, Science, Engineering, and Technology (AIxSET)* (pp. 111–118). IEEE. <https://doi.org/10.1109/AIxSET65682.2025.00023>
4. Helstrom CW (1969) Quantum detection and estimation theory. *J Stat Phys* 1(2):231–252. <https://doi.org/10.1007/BF01007479>
5. Jerbi S, Fiderer LJ, Nautrup P, Kübler H, Briegel JM, H. J., Dunjko V (2023) Quantum machine learning beyond kernel methods. *Nat Commun* 14:517. <https://doi.org/10.1038/s41467-023-36159-y>
6. Kim Y, Jang E, Kim H, Choi S, Lee C, Kim D, Kyoung W, Shin K, Ro WW (2025) Distribution-adaptive dynamic shot optimization for variational quantum algorithms. *Phys Rev Res* 7(4):043253. <https://doi.org/10.1103/vwhk-22b8>
7. Kingma DP, Ba J (2015) Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*. <https://arxiv.org/abs/1412.6980>
8. Lee D, Baek K, Huh J, Park DK (2024) Variational quantum state discriminator for supervised machine learning. *Quantum Sci Technol* 9(1):015017. <https://doi.org/10.1088/2058-9565/ad0a05>
9. Mordacci M, Pandey M, Santini P, Amoretti M (2025) Triplet loss based quantum encoding for class separability. *CEUR Workshop Proceedings*, 4153, 115–128
10. Nielsen MA, Chuang IL (2010) *Quantum computation and quantum information* (10th anniversary ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511976667>
11. Peña Tapia E, Scarpa G, Pozas-Kerstjens A (2023) A didactic approach to quantum machine learning with a single qubit. *Phys Scr* 98(5):054001. <https://doi.org/10.1088/1402-4896/acc5b8>
12. Pérez-Salinas A, Cervera-Lierta A, Gil-Fuster E, Latorre JI (2020) Data re-uploading for a universal quantum classifier. *Quantum* 4:226. <https://doi.org/10.22331/q-2020-02-06-226>
13. Phalak K, Ghosh S (2023) Shot optimization in quantum machine learning architectures to accelerate training. *IEEE Access* 11:41514–41523. <https://doi.org/10.1109/ACCESS.2023.3270419>

14. Scriva G, Astrakhantsev N, Pilati S, Mazzola G (2024) Challenges of variational quantum optimization with measurement shot noise. Phys Rev A 109(3):032408.
<https://doi.org/10.1103/PhysRevA.109.032408>
15. Tsili A, Maragkopoulos G, Mandilara A, Syvridis D (2025) Learning the N-input parity function with a single-qubit and single-measurement sampling. Electronics 14(5):901.
<https://doi.org/10.3390/electronics14050901>
16. Turab A, Latif J, Zaidi SA, Hussain A (2026) NOISE-VQA: Convergence and complexity analysis of variational quantum algorithms with finite-shot and biased oracles. J Comput Appl Math 488:117857. <https://doi.org/10.1016/j.cam.2026.117857>
17. Zhang B, Zhuang Q (2022) Fast decay of classification error in variational quantum circuits. Quantum Sci Technol 7(3):035017. <https://doi.org/10.1088/2058-9565/ac70f5>

Figures

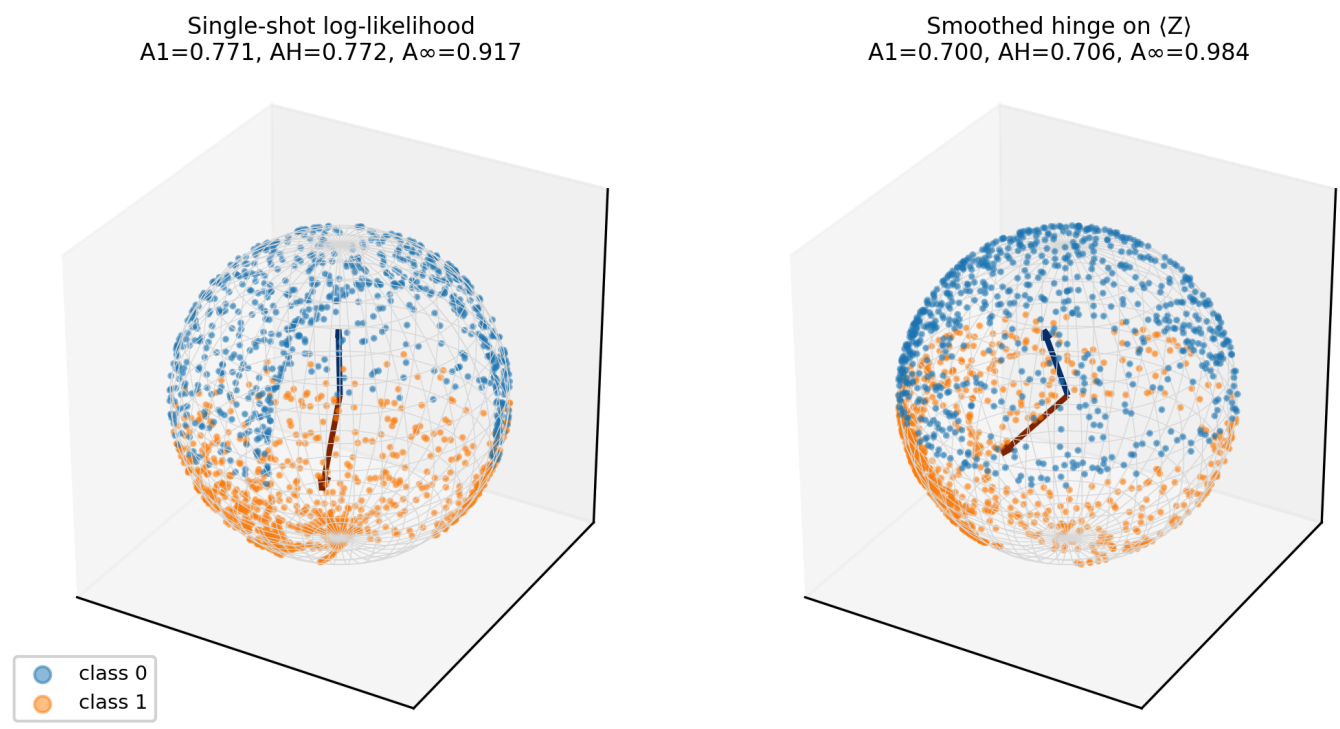


Figure 1

Final Bloch vectors of two trained depth-6 classifiers on the circle task (no label noise, seed 0), for 1,500 test inputs coloured by class. Arrows show the class-mean Bloch vectors. The single-shot log-likelihood (left) pushes the classes towards opposite poles; the smoothed hinge (right) separates them by the sign of $\langle Z \rangle$ but leaves many states near the equator, which shortens the class-mean vectors and lowers the Helstrom bound even though thresholded accuracy is higher.

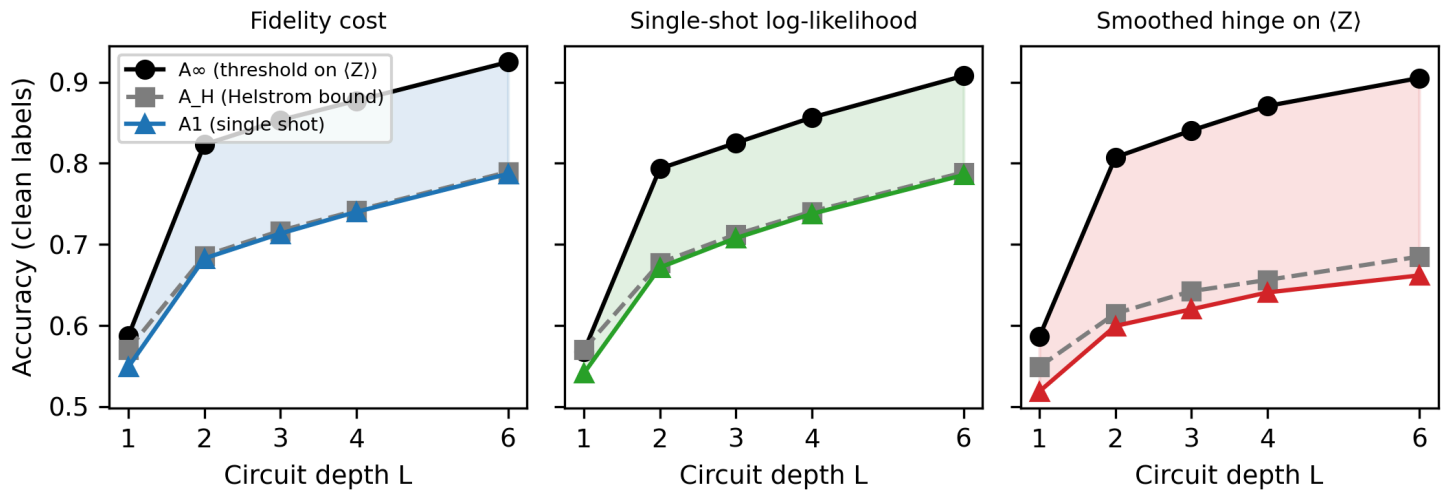


Figure 2

Mean thresholded accuracy (A_∞), Helstrom bound (A_H) and single-shot accuracy (A_1) by circuit depth for each training loss, averaged over tasks, noise levels and seeds (80 runs per point). The shaded band is the shot gap. A_H and A_1 almost coincide, so nearly all of the gap lies between the Helstrom bound and the thresholded score.

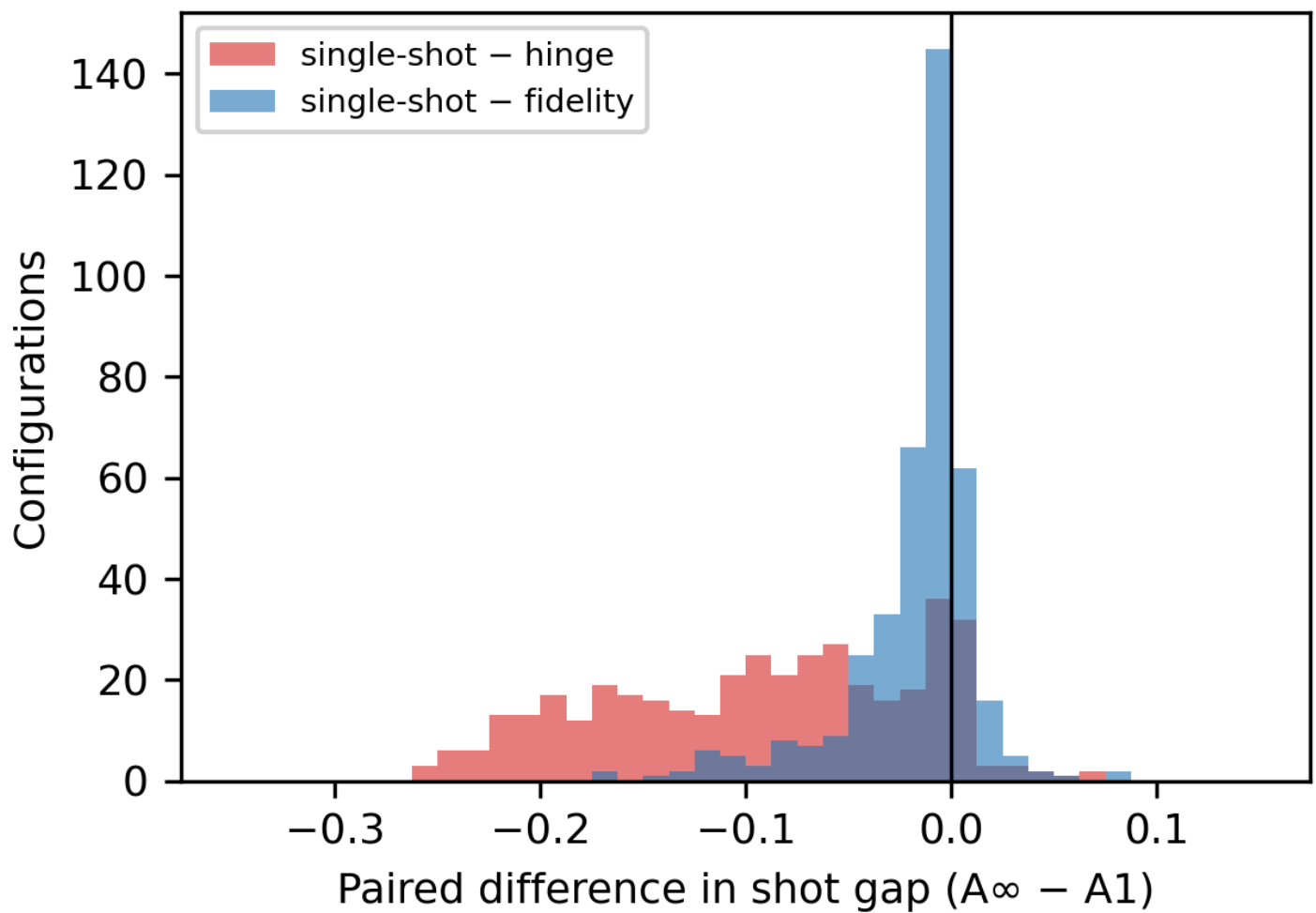


Figure 3

Paired differences in the shot gap between the single-shot log-likelihood and each of the other losses, over 400 configurations sharing data and initialisation. Negative values mean a smaller gap under the single-shot log-likelihood.

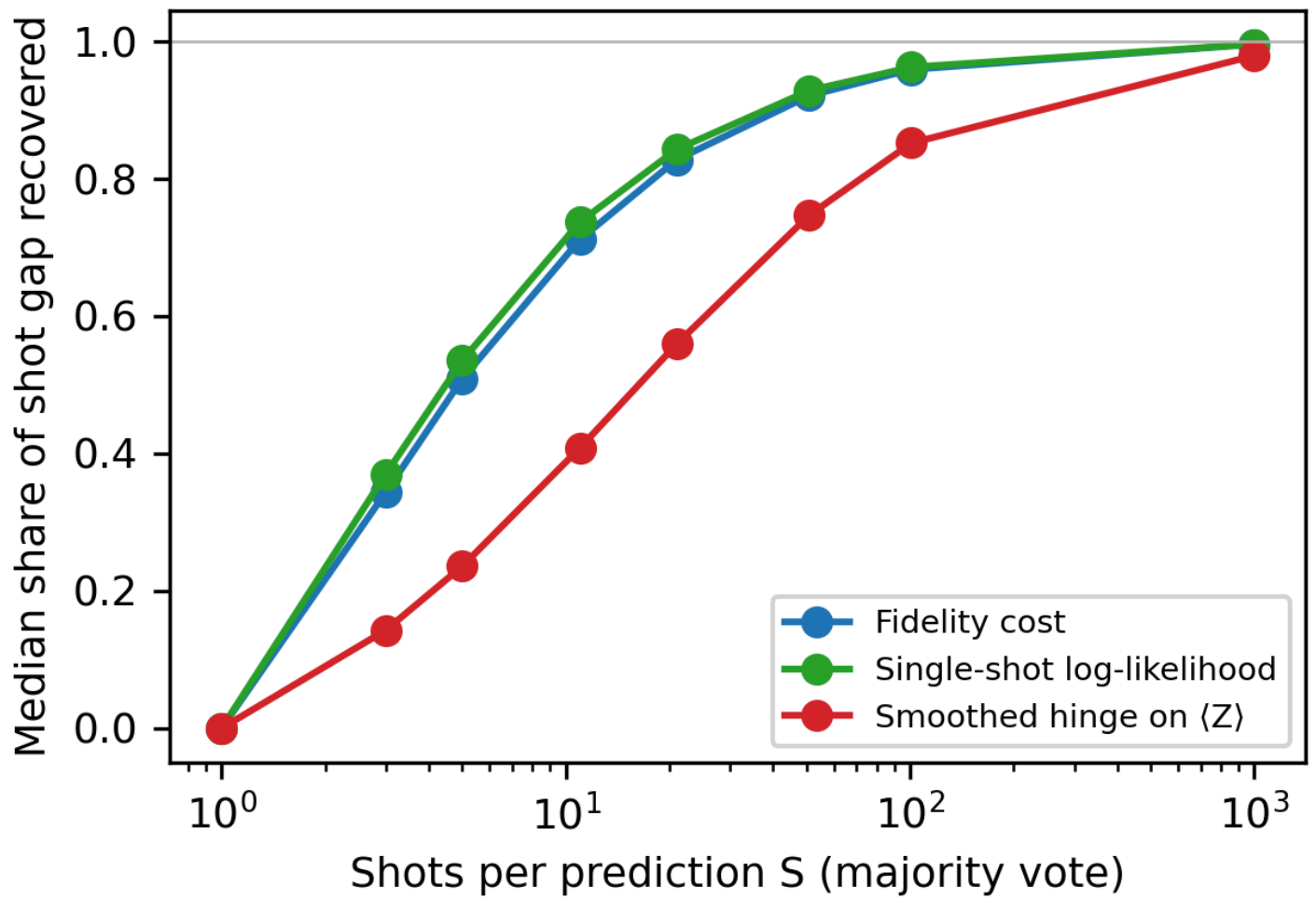


Figure 4

Median share of the shot gap, $(A_S - A_1)/(A_\infty - A_1)$, recovered by a majority vote of S shots, by training loss (runs with a gap above 0.02).