Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Quantum models pay for non-linearity in data encodings

Pavel Sulimov

suli@zhaw.ch

Zurich University of Applied Sciences https://orcid.org/0000-0003-2885-2646

Claude Lehmann

Zurich University of Applied Sciences, University of Zurich

Article

Keywords:

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11614879/v1

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

Additional Declarations: There is **NO** Competing Interest.

Quantum models pay for non-linearity in data encodings

Pavel Sulimov[1,*] and Claude Lehmann[1,2]

[1]Zurich University of Applied Sciences (ZHAW), Winterthur, Switzerland
[2]University of Zurich, Zurich, Switzerland
[*]Correspondence: suli@zhaw.ch

Abstract

Ancillas, mid-circuit measurements, noise, post-selection and classical post-processing are each credited with making quantum machine-learning models non-linear in their input. Here we show that, when the input enters only through Pauli encoding gates, these all draw on one budget: the spectral range of the encodings in each shot. For any procedure with finitely many shots, arbitrary channels, feed-forward and a bounded classical head, the expected output extends to complex inputs $a+ib$ as an entire function bounded by $\prod_i \cosh(E_i b)$, where E_i is the range of shot i, and its Fourier coefficients lie below a Gaussian whose variance is the sum of the E_i^2. Range spent inside one shot buys sharp features at the Heisenberg rate, range spread over shots only at the shot-noise rate, so a hybrid model that encodes once per shot needs quadratically more encodings than a coherent circuit to resolve a jump. What range costs depends on the regime: circuit depth, capped by coherence, when the encoding scale is fixed, and input precision when it can be rescaled. Post-selection escapes the budget only with infinitely many shots. Sixteen experiments on IBM, IQM and IonQ processors, fifteen of them pre-registered, test the predicted rates and costs.

1 Introduction

A parameterised quantum circuit that encodes a scalar x through gates $e^{-ixG/2}$ computes a trigonometric polynomial $f(x) = \sum_\omega c_\omega e^{i\omega x}$ whose frequencies are set by the eigenvalue differences of the generators [1, 2, 3]. Unitary evolution is linear in the state, so non-linearity in x is usually sought elsewhere: in ancillas, mid-circuit measurements with feed-forward [4, 5], dissipation, repeat-until-success circuits [6], reservoirs, or classical post-processing of the measured outcomes. Benchmarks treat these as separate mechanisms [7]. We show that they all draw on the same budget, the spectral range of the data encodings within each shot, under one standing assumption that we state once here and that every result below uses: *the input x enters the model only through its encoding gates.* A classical layer that transforms x before it is encoded, or a choice of circuit that depends on x itself rather than on measured outcomes, lies outside our scope and can add any non-linearity at no quantum cost; so does non-linearity in the quantum state, which requires copies of the state.

The rates we find are those of phase estimation, and several of our corollaries restate known results in the language of data encodings: the polynomial method of query complexity [8], the Cramér–Rao bound with the sequential Heisenberg limit and its standard quantum limit for repetition [9, 10], the approximation rates of Bernstein polynomials [11, 12] and the Chebyshev–Newman theory of rational approximation [13, 14]. What is new is the object that unifies them. For a single shot whose E encodings are interleaved with arbitrary non-covariant operations and feed-forward, where no single generator exists, we bound the total outcome mass of the analytically continued

probabilities by $\cosh(Eb)$; the bound multiplies over adaptively chosen shots, so the expected output of any finite-shot procedure with any bounded classical head on the outcomes extends to an entire function bounded by $\prod_i \cosh(E_i b)$ (Theorem 1; Fig. 1). From this one bound follow the degree (Corollary 1), the slope through the Fisher budget $\sum_i E_i^2$ (Corollary 2), a Gaussian Fourier spectrum with that variance, an exchange rate between depth and shots with its constants (Corollary 3; Supplementary Note 2), a finite-floor exchange rate for post-selection (Theorems 3 and 4) and, through a realisability construction, the statement that every admissible function is reached by a circuit with exactly that many encodings. Input-Lipschitz bounds of the coherent kind were known [15]; we add the incoherent side.

The consequences are concrete. Range spent inside one shot, through depth or through rescaled encodings $R_Y(sx)$, buys sharp features at the Heisenberg rate, whereas the same range spread over shots or over unentangled qubits buys them only at the shot-noise rate, for every kind of sharp feature, not only a jump. A hybrid model in which every shot encodes x once and a classical neural network maps the measured bits to the output therefore needs the square of a coherent circuit's encoding budget to resolve a jump, whatever the network computes. What the range costs, however, depends on the regime. When the scale of each encoding is fixed by the architecture, as in data re-uploading, range costs gate depth and coherence caps it (Theorem 2). When rescaling is free, as on IBM hardware, where $R_Y(16x)$ compiles into the same two physical pulses as $R_Y(x)$, range costs nothing quantum: shots of single rescaled gates resolved a jump at the coherent rate, with an error that fell as $T^{-0.55}$ in the encoding budget T against $T^{-0.23}$ for repeated $R_Y(x)$ shots, with a replication on IQM hardware, much as in entanglement-free Heisenberg-limited phase estimation [16] and geometric encodings [17]. In that regime the precision of the input caps the usable scale: a datum known to precision δ supports a slope of at most $\sqrt{2/\pi}/\delta$ at any budget. The price in the title is paid in range; which resource the range costs, depth, coherence or input precision, is what each regime decides.

Post-selection, the classic exception to the polynomial method (PostBQP = PP [18]; rational degree [14]), escapes the bound only in the limit of infinitely many shots. In query complexity, the post-selected quantum query complexity of a Boolean function equals the least degree of a rational approximation up to a factor of two [14], with no constraint on the success probability. On hardware it is the success floor that sets the shot cost, so we work with a finite floor and show that it fixes the exchange rate between encodings and shots (Theorems 3 and 4); the question of which functions of an outcome probability measurement and post-selection can realise is that of quantum Bernoulli factories [19].

Within this function class, the structure of a circuit decides which coefficients it can reach. Hidden time-reversal and reflection symmetries remove frequencies, and a single upload on m qubits caps the top coefficient at 2^{-m}, which is the Fourier form of the separable limit discussed below. Both properties can be computed before training, and a companion paper [20] develops them and tests them on hardware. Here we test the theory in sixteen experiments on IBM, IQM and IonQ processors, fifteen of them with criteria and a prediction committed to version control before the job (on IBM, the prediction comes from the device's calibrated noise model), and we report the runs that missed alongside those that passed, distinguishing predictions that could have failed from consistency checks (Table 2).

Earlier work computes exact spectra and coefficients for Clifford-plus-rotation circuits [21, 22], bounds the largest frequency from the encoding [2, 23, 24], studies coefficient magnitudes for typical (Haar or 2-design) parameters [25, 26], trainable or exponential encoding scales [27, 17], generalisation in terms of the encoding spectrum [28], the relation between explicit, implicit and re-uploading models [29] and the Fisher information as a capacity measure [30]; classical surrogates of models with concentrated spectra are known [31]. The single-qubit universality that our realisability con-

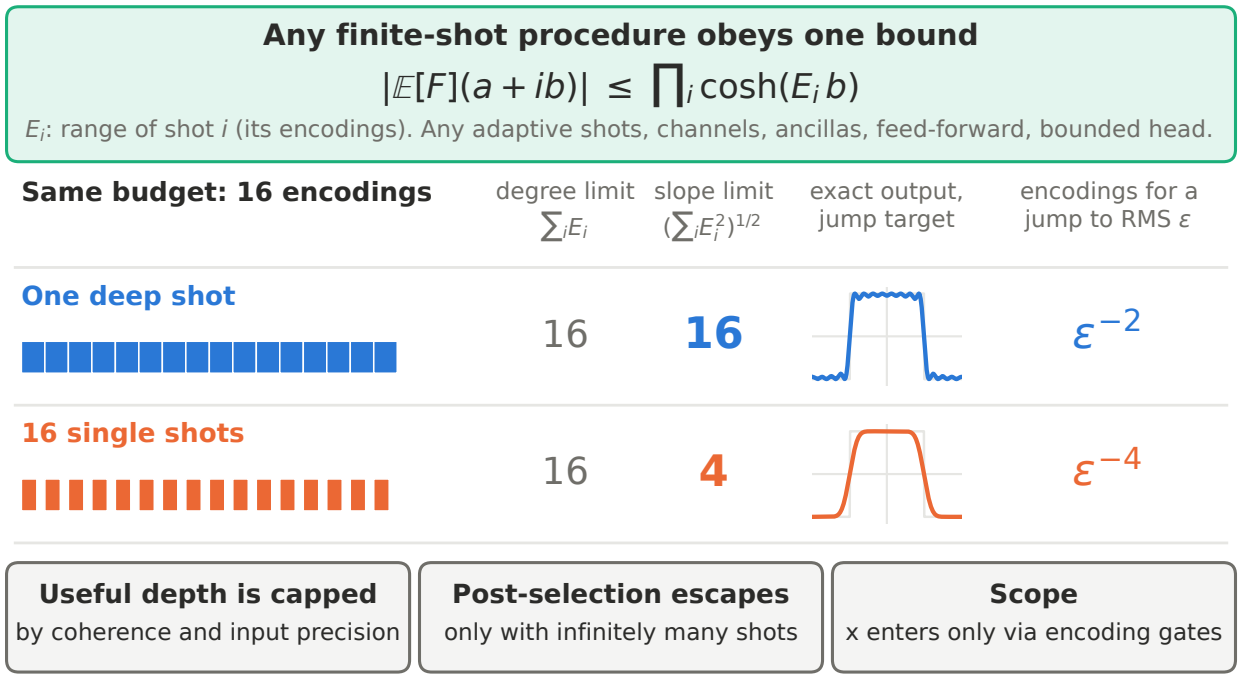


Figure 1: **One budget for non-linearity.** Top, every finite-shot procedure obeys the master bound (Theorem 1), which fixes the degree, the steepness and a Gaussian spectrum. Middle, two ways to spend the same 16 encodings: the degree limit is the same, but one deep shot reaches slope 16 and 16 single shots at most 4, so a jump to RMS error ε costs ε^{-2} encodings in one shot and ε^{-4} by repetition. The output curves are exact: the best bounded polynomial of degree 16 and the best bounded head on 16 single-encoding shots against the jump target (grey). Bottom, coherence and input precision cap the range per shot, post-selection leaves the class only with infinitely many shots, and every result assumes that x enters only through encoding gates.

struction uses was shown in [32], and the ε^{-2} against ε^{-4} contrast is that of amplitude estimation against sampling [33] in a different guise.

2 Results

2.1 Channels do not change the function class

Lemma 1 (Channel closure)**.** *Let x enter only through encoding gates $U_j(x) = e^{-ixG_j/2}$ whose generators have integer eigenvalue half-differences $(\lambda - \lambda')/2$ and spectral range $\Delta_j = \lambda_{\max}(G_j) - \lambda_{\min}(G_j)$. Interleave them with arbitrary x-independent unitaries, CPTP maps and instruments, where measurement outcomes may control later operations, including whether later encodings are applied. Then the outcome-averaged model $f(x) = \mathrm{Tr}[O\rho_{\mathrm{out}}(x)]$ is a trigonometric polynomial with $|\omega| \le \max_{\mathrm{branches}} \sum_j \Delta_j/2$. For Pauli generators ($\Delta_j = 2$) this is $N_{\max}$, the largest number of encodings on any outcome branch.*

We call the largest frequency an encoding can carry, $\Delta_j/2$, the *range* of the encoding (half the spectral range of its generator). It is 1 for a Pauli encoding and s for $R_Y(sx)$, and we count encodings in these units throughout; "spectral range of a shot" below means the sum of the ranges of its encodings. Table 1 confirms the lemma for reset, for mid-circuit measurement with feed-forward

(including an extra upload controlled by the outcome), for repeated measurement and for mixed noise. Two operations leave the class, as they must. Post-selection divides by an x-dependent success probability and makes f rational in e^{ix}, and a coherent over-rotation of the encoding makes the generator spectrum non-integer [34].

Table 1: Channel closure on 3 system qubits and 1 ancilla (two layers of three $R_Y(x)$ encodings, readout $\langle Z\rangle$ on the qubit whose light cone contains every encoding, 20 random circuits per row, 128-point FFT). $N_{\max}$ is the frequency bound of Lemma 1, the number of encodings on the longest branch; it is 8 where an outcome-controlled extra upload adds one encoding per layer. Top frequency: the largest frequency carrying more than 10^{-14} of the peak power. Every mechanism covered by the lemma reaches its bound exactly, with top frequency $N_{\max}$ and power above $N_{\max}$ at the level of floating-point rounding; the two controls that violate its assumptions spread power up to the grid limit (64).

Mechanism	$N_{\max}$	top frequency	power above $N_{\max}$
unitary	6	6	3.5×10^{-31}
ancilla reset between re-uploads	6	6	3.4×10^{-31}
mid-circuit measurement + feed-forward	8	8	5.8×10^{-31}
repeated mid-circuit measurement	6	6	1.5×10^{-30}
depolarizing + amplitude damping + dephasing	6	6	4.4×10^{-31}
post-selection on the ancilla	6	63	1.8×10^{-3}
encoding over-rotation, $\varepsilon = 0.07$	6	64	5.1×10^{-3}

2.2 One bound for every finite-shot procedure

A *procedure* learns about x only by running circuits with Pauli encodings. It may run any number of shots, choose each next circuit and when to stop from all outcomes so far, and output any bounded function F of the full shot record, for example a sample mean passed through a neural network, a ratio estimator for post-selection, a majority vote or the result of a repeat-until-success loop (when a single shot can suffice for a prediction is analysed by Recio-Armengol et al. [35]). Let T be the number of encodings executed along the record, each counted by its range (a rescaled encoding $R_Y(sx)$ with integer s counts as s, Lemma 1), and let E_i be the count within shot i.

Theorem 1 (Master bound)**.** *Let a procedure run shots with spectral ranges $E_1, E_2, \dots$, with any channels, ancillas, mid-circuit measurements and feed-forward inside a shot and any adaptive choice between shots under a bounded stopping rule, and output a head $|F| \le 1$; let $V = \sum_i E_i^2$ (the largest value over records). Then $x \mapsto \mathbb{E}_x[F]$ extends to an entire function with $|\mathbb{E}[F](a+ib)| \le \max_{\text{records}} \prod_i \cosh(E_i b) \le e^{Vb^2/2}$, and every Fourier coefficient obeys $|c_\omega| \le e^{-\omega^2/2V}$. For n incoherent encodings (shots of one-encoding circuits, or single uploads on separable states) $V = n$.*

The proof continues the outcome probabilities to complex $x = a + ib$, shows that the total outcome mass of a shot with E encodings is at most $\cosh(Eb)$, and then shifts the integration contour (Methods). For an unbounded stopping rule the theorem applies to the procedure truncated after k encodings, which differs from the original by at most $\|F\|_\infty \sup_x \Pr_x[T > k]$ (Corollary 1). Bounds on the derivatives of a single coherent circuit's output in terms of its generator norms are known [36]. For one coherent circuit even the bound $|\mathbb{E}[F](a+ib)| \le \cosh(Tb)$ on the *expectation* is classical: every real trigonometric polynomial of degree T bounded by 1 on the real line obeys

it, with equality for $\cos Tx$ (Bernstein's inequality in the strip). That bound does not multiply over shots. A head on N shots is a sum over 2^N records, and the classical inequality controls one polynomial rather than the total outcome mass $\sum_o |p_o|$, which is what Theorem 1 bounds. For one encoding per shot the mass bound reduces to the classical growth of complex Bernstein polynomials, $(|z|+|1-z|)^N = \cosh(b)^N$ at $z = \cos^2((a+ib)/2)$, combined with Hoeffding's lemma $\cosh b \le e^{b^2/2}$. For parallel encodings on a product state it is a concentration property of the generator's spectrum, in the language of modes of asymmetry [37]. To our knowledge, the bound $\cosh(Eb)$ for a shot whose E encodings are interleaved with arbitrary non-covariant operations and feed-forward, where no single generator exists, is new, and so is its product over adaptive shots. The proof takes half a page. The bound is tight for every E: E consecutive encodings with nothing between them have outcome mass exactly $\cosh(Eb)$, and maximising the mass over circuits with an ancilla, a mid-circuit measurement and outcome-controlled extra encodings comes within 10^{-5} of the bound without exceeding it (Methods). It extends to every generator whose eigenvalue half-differences are integers, each encoding counted by its range (Methods). Its interest lies in what the product implies, above all a Gaussian spectrum whose variance V is the Fisher budget of Corollary 2, which makes the theorem a spectral Cramér–Rao bound. One coherent circuit of depth T has $V = T^2$, and inside its degree the bound is close to vacuous (it equals $e^{-1/2}$ at $\omega = T$), as it should be.

2.3 Non-linearity is paid in encodings

Degree, steepness, the exchange rate and the separable limit all follow as corollaries; two more, for curvature and for every kind of sharp feature, are in Supplementary Note 1. Since $\prod_i \cosh(E_i b) \le e^{T|b|}$, the periodic entire function $\mathbb{E}_x[F]$ has exponential type at most T, and shifting the contour removes every Fourier coefficient beyond T:

Corollary 1 (Encoding count)**.** *If $T \le T_{\max}$ on every record, $x \mapsto \mathbb{E}_x[F]$ is a trigonometric polynomial of degree $\le T_{\max}$. In general $\mathbb{E}_x[F]$ lies within $\|F\|_\infty \sup_x \Pr_x[T > k]$ of a polynomial of degree $\le k$, for every k.*

A Cauchy estimate on the circle of radius $V^{-1/2}$ bounds the slope by $\sqrt{eV}$; the Fisher information of the shot record gives the sharp constant:

Corollary 2 (Steepness)**.** $|\frac{d}{dx}\mathbb{E}_x[F]| \le \|F\|_\infty \sqrt{I(x)}$*, where I is the Fisher information of the record, and $I \le \sum_i E_i^2$ for shots with E_i encodings each.*

A single circuit with T encodings can therefore be as steep as T, the Heisenberg scaling of phase estimation, whereas N shots of a depth-E circuit reach only $E\sqrt{N} = \sqrt{TE}$, the standard quantum limit. A bounded function of slope s misses a jump by an RMS error of order $s^{-1/2}$, so resolving a jump costs an error $\propto T^{-1/2}$ with depth and $\propto (TE)^{-1/4}$ with shots. Exact computations confirm each step (Fig. 2; Methods). Three classical heads and an adaptive sequential test stay within the degree bound, and the slope reaches its bound at $N = 1$ only, so for $N > 1$ the Fisher bound is a bound and not an attained value (Fig. 2b; Supplementary Note 2). With post-selection and N shots, the expected ratio estimator is a polynomial of degree at most NE, which approaches the ideal rational output only once $N \gtrsim 1/q_{\min}$ (Fig. 2c). Hybrid models with classical non-linear heads, polynomial readouts, multiple copies and repeat-until-success loops all live inside the same budget. So do classifiers, because the probability of each predicted label is the expectation of a bounded function of the shot record, hence a polynomial of degree at most T whose slope obeys Corollary 2.

Corollary 3 (Exchange rate)**.** *For a jump from -1 to $+1$, every procedure with T encodings, at most $E_{\max}$ per shot, $TE_{\max} \ge 2$, has RMS error $\ge 0.46\,(TE_{\max})^{-1/4}$ in its expected output $\mathbb{E}[F]$ (and*

hence in every single prediction, since mean-squared error is at least squared bias). One coherent circuit reaches RMS error at most $4(T+1)^{-1/2}$ *(Fejér mean) and repeated single-encoding shots at most* $1.88\,T^{-1/4}$ *(majority vote).*

Resolving a jump to RMS ε therefore costs $\Theta(\varepsilon^{-2})$ encodings with depth and $\Theta(\varepsilon^{-4})$ with single-encoding shots, and for these two arms the constants are known to within a factor of 1.5: in exact computations over the class optimum of each arm, error$\times(TE_{\max})^{1/4}$ settles at 0.55 for the best head on one-encoding shots (error $0.55\,T^{-1/4}$) and at 0.70 for the best bounded polynomial of degree T (error $0.70\,T^{-1/2}$), against the lower bound 0.46 (Fig. 2a; the Fejér mean of the proof is a looser $0.90\,T^{-1/2}$). Over every allocation we tested, depth-E Fejér circuits read out by a majority vote, which are not class optima, give constants between 0.55 and 0.96. What matters is the spectral range per shot, $E_{\max}$, whether it comes from depth or from rescaling. A single qubit of ibm_fez shows the two rates directly (Fig. 2a). Optimal bounded step approximants with $T = 2$ to 64 encodings, compiled into circuits with exactly T encodings, resolved a step with error $\propto T^{-0.48}$ (bootstrap 95% interval 0.45–0.50, fitted over $T = 2$–32), while the same encodings spent on repeated shots with a majority vote gave $T^{-0.27}$ (0.24–0.29); the difference of the two exponents is 0.21 (0.18–0.24). The $T = 64$ point is reported but not fitted, because its 128 measurement points sit at the Nyquist limit of a degree-64 polynomial (Supplementary Note 6). At $T = 64$ depth won by a factor of 1.9, even though the deep circuit carried 130 physical pulses, and all pre-registered criteria passed. Adaptivity and mid-circuit noise do not raise the degree either. A one-qubit dynamic circuit on ibm_fez used mid-circuit measurement, reset and feed-forward to choose its second sub-circuit from the first outcome, and stopped after 4 or 6 encodings. Its power above degree 6 stayed at 2.2 times the shot-noise floor (the criterion was below 3 times), and it kept 82% of the ideal low-degree power.

Depth is not the only way to raise $E_{\max}$. On IBM hardware $R_Y(\theta)$ compiles to two fixed $\sqrt{X}$ pulses and virtual Z rotations, so $R_Y(16x)$ costs exactly as much as $R_Y(x)$. On one qubit of ibm_kingston the visibility of $R_Y(sx)$ lay between 0.991 and 0.994 for every s from 1 to 16. We ran shots at scales $1, 2, 4, \ldots, 2^J$ in two quadratures, four per setting, and read them out with a Bayes head fixed in advance. They resolved the jump with error $\propto T^{-0.55}$ (bootstrap 95% interval 0.43–0.67, three budgets $T = 56$, 120, 248), against $T^{-0.23}$ (0.11–0.34) for the same budget spent on $R_Y(x)$ shots; the difference of the exponents is 0.32 (0.15–0.48). At $T = 248$ their error of 0.094 was below 0.116, the bound of Corollary 3 for any procedure with one encoding per shot. Every value lay within 0.015 of the prediction of the calibrated noise model, and all pre-registered criteria passed (an exponent of at least 0.40 with rescaling and at most 0.35 without), with both intervals on the passing side. The square-root penalty therefore comes from repeating one encoding, not from a lack of coherence across shots. As in phase estimation [38, 16, 39] and in exponential encodings [17], geometric scales place almost the whole Fisher budget $V = \sum_i E_i^2$ in the largest scale. On IQM Garnet the rotation angle is a physical pulse parameter rather than a virtual frame change, which makes the test harder. A pre-registered replication on 20 qubits in parallel gave exponents of 0.57 (0.45–0.69) against 0.25 (0.13–0.37, an interval that includes the threshold 0.35) and an error of 0.090 at $T = 248$, below the one-encoding bound; all criteria passed on the point estimates once we had corrected a shot-ordering error in our handling of aggregated counts (Supplementary Note 6).

2.4 Repetition buys sharpness at the square root

The same square root governs every sharp feature, not only jumps. Corollary 1 credits N shots of a depth-E circuit with a literal degree of NE, but their derivatives behave like those of a polynomial of degree $E\sqrt{N}$: the slope of the expected output is at most $E\sqrt{N}$ and its curvature at most $e\,NE^2$, whereas a bounded polynomial of degree T reaches T and T^2 (Supplementary Theorem 1). The Gaussian spectrum of Theorem 1 turns this into a rate for every kind of feature: for a target g

with an α-Hölder singularity, every finite-shot procedure has $\|\mathbb{E}[F] - g\|_2 = \Omega((V \log V)^{-(2\alpha+1)/4})$ (Supplementary Theorem 2), and exact optimisation over all bounded heads on V single-encoding shots attains the exponent to within 0.007 for α from 0 to 1 (Supplementary Note 1). So V incoherent encodings are worth about $\sqrt{V}$ coherent ones whatever the feature. For one encoding per shot these statements are classical: an N-shot head is a Bernstein polynomial with bounded coefficients in the outcome probability, the object of Bernoulli factories [40, 41] and stochastic computing [42], whose approximation rates [11, 12] correspond to depth $\sqrt{N}$. Theorem 1 extends them to circuits with E encodings and arbitrary operations in each shot. For analytic targets the difference disappears, since smooth features need no large derivatives. A feature of width w therefore costs about $1/w$ encodings in a coherent circuit and about $1/w^2$ by repetition. This is where the polynomial method is loose, and it shows that the non-linearity repetition buys inefficiently is sharpness rather than degree.

The resolution limit is visible directly on ibm_fez. For smoothed steps $\tanh(\cos x/w)$ with w from $\frac{1}{2}$ to $\frac{1}{64}$, quadrupling the budget from 8 to 32 encodings shrank the narrowest width resolved to RMS 0.1 by a factor of 6.8 with depth (from 0.14 to 0.021) but only by 2.5 with shots (from 0.26 to 0.10), against same-day predictions of 7 and 2.7 and asymptotic factors of 4 and 2, so these budgets are still pre-asymptotic. At $w = \frac{1}{2}$ both arms stayed below 0.021, and all criteria passed.

The exchange rate is a statement about capacity, not about training. Fitted to its class optimum, a one-qubit circuit with 64 encodings had a third of the test error of the matched equal-scale hybrid model on a sharp step (0.060 against 0.186), and in trained hybrid models the price of the classical head rose from 0.002 to 0.110 in test error as the feature width fell from $\frac{1}{2}$ to $\frac{1}{32}$. Gradient training, however, did not collect the capacity: in five pre-registered attempts, trained circuits with narrow two-qubit layers never beat a fairly read-out hybrid model at matched encodings, a pre-registered task-level sharpness law failed, and a circuit chosen after diagnosing one of those failures beat the hybrid and its theorem ceiling on fresh seeds only when its layout contained the target (Supplementary Note 4). None of the capacity bounds in this paper is a trainability result.

2.5 Only coherent encodings count

The effective degree is one instance of a general rule: encodings count fully only when they act coherently. For one upload of weight w_j on each of m unentangled qubits the outcome mass is at most $\prod_j \cosh(w_j b)$, so for integer weights Theorem 1 holds with $V = \sum_j w_j^2$, and the quantum Fisher information again gives the sharp slope:

Corollary 4 (Copies are shots)**.** *If x enters once on each of m qubits through Pauli encodings with weights w_j and the pre-encoding state is separable, then for every later channel, measurement and head $|F| \le 1$, $|\partial_x \mathbb{E}[F]| \le (\sum_j w_j^2)^{1/2}$. An entangled pre-encoding state reaches $\sum_j |w_j|$.*

This is the separable limit of quantum metrology [9, 43] applied to data encodings. With equal weights, a single upload on unentangled qubits, the most common encoding, is worth only $\sqrt{m}$ coherent encodings in sharpness whatever circuit follows, and it misses a jump by at least $0.46\, m^{-1/4}$ even when its expectation is known exactly. Its degree is m, but with local readout its ideal outcome law is literally that of m shots of one encoding. With geometric weights $w_j = 2^j$ the bound $(\sum_j w_j^2)^{1/2}$ lies within a factor $\sqrt{4/3}$ of the largest weight, so rescaled uploads avoid the penalty without entanglement.

On ibm_fez the separable limit was never exceeded, and an entangled input beat it on IBM and IQM processors. With x encoded once on each of $m \le 24$ qubits of an ibm_fez chain, the separable arms stayed below $\sqrt{m}$, while the same CNOTs placed before the encoding produced a GHZ-type input whose slope exceeded $\sqrt{m}$ at every size, by up to 2.1 times at $m = 12$; GHZ states

with an m-fold phase response have been prepared with trapped ions [44] and Rydberg atoms [45]. A phase-robust replication on IQM Garnet gave slopes of 3.25, 5.61 and 6.40 against $\sqrt{m} = 2$, 2.83 and 4 for $m = 4$, 8 and 16 (the $m = 16$ value was reported but not gated). On IonQ Forte-1 the test was inconclusive: its phase-robust bound was computed after the run, and several blocks returned fitted visibilities above 1 (Supplementary Note 5). The separable arms are consistency checks, since noise channels can only lower a slope (a coherent over-rotation of the encodings could raise it); the entangled arms are predictions that could have failed. Trained entangle-first models did not reach their capacity (Supplementary Note 5), so the corollary bounds capacity rather than what training finds.

2.6 Coherence and input precision cap the range

Theorem 2 (Noise)**.** *(i) If every encoding is followed by dephasing along its axis with coherence factor η, every procedure with T encodings (any control, ancillas and adaptivity) has Fisher information at most cT, $c = \eta^2/(1-\eta^2)$, and misses a jump by RMS $\geq 0.46\,(cT)^{-1/4}$ whatever its allocation (for $cT \geq 2$). (ii) For a general channel whose Kraus span contains the generator, the bound is $c_N T$ with a channel-dependent constant c_N. (iii) Under (i), coherent blocks of $E^* = 1/(2\ln\eta^{-1})$ encodings (for $\eta \geq e^{-1/2}$) repeated T/E^* times have Fisher information $T/(2e\ln\eta^{-1})$, which tends to cT/e as $\eta \to 1$.*

Theorem 2 composes the channel-extension bound of noisy metrology [46, 47, 10, 48] with Corollary 2; in Fourier terms the same dephasing suppresses the component ω by $\eta^{|\omega|}$ [49]. Part (iii) is the optimal Ramsey interrogation time of Huelga et al. [50]. It is a statement about Fisher information, and we do not claim that block strategies attain the RMS bound. Noise thus restores the shot-noise rate for every architecture, with a constant set by the coherence per encoding, and the best circuit spends the coherence of one shot before repeating. When the generator lies outside the Kraus span, error correction with noiseless ancillas restores Heisenberg scaling [51, 52], and whether depth keeps buying non-linearity as models grow then depends on the encoding axis relative to the dominant noise.

The theorem counts noise per encoding gate. A rescaled gate $R_Y(sx)$ carries range s through a single noisy channel, so when the encoding angle is a classical control parameter, gate noise does not cap the spectral range. Noise on the input does. If x carries Gaussian noise of width δ, then $R_Y(sx)$ dephases by $e^{-s^2\delta^2/2}$ and its Fisher information per shot, $s^2 e^{-s^2\delta^2}$, peaks at $s = 1/\delta$. Over N shots the Fisher information is then at most $N/(e\delta^2)$, and every procedure returns to the shot-noise rate beyond $s \approx 1/\delta$. If instead the datum itself is known only to precision δ, identically in every shot, each output is a bounded function convolved with the noise, and its slope is at most $\sqrt{2/\pi}/\delta$ at any budget. Non-linearity bought by rescaling is therefore capped by the precision of the input, just as depth is capped by coherence.

On one qubit of ibm_fez a budget of 64 encodings kept paying for depth under the device's native noise (RMS 0.086 for one circuit of depth 64, against 0.227 for 64 single-encoding shots). Adding a $4\,\mu$s idle after each encoding moved the best allocation to 16 encodings per shot, as predicted, but by less than the pre-registered margin (0.013 against 0.02), because the idle also added a coherent phase that the noise model did not include (Supplementary Note 6). A redesigned run replaced the idle by random Z flips directly after every encoding, which in this run is $R_Z(-x)$ ($p = 0.015$, 32 patterns per point). These flips act along the encoding axis, are virtual on IBM hardware and realise exactly the dephasing of part (i), with $\eta = 0.97$ and $E^* = 16.4$. Among $E \in \{4, 16, 64\}$ the dephased arm did best at $E = 16$ (RMS 0.171, against 0.187 and 0.243), the native arm kept improving up to $E = 64$ (0.090 against 0.138), every value lay within 0.01 of the prediction of the calibrated noise model, and all criteria passed. The run locates E^* only to within a factor of 4,

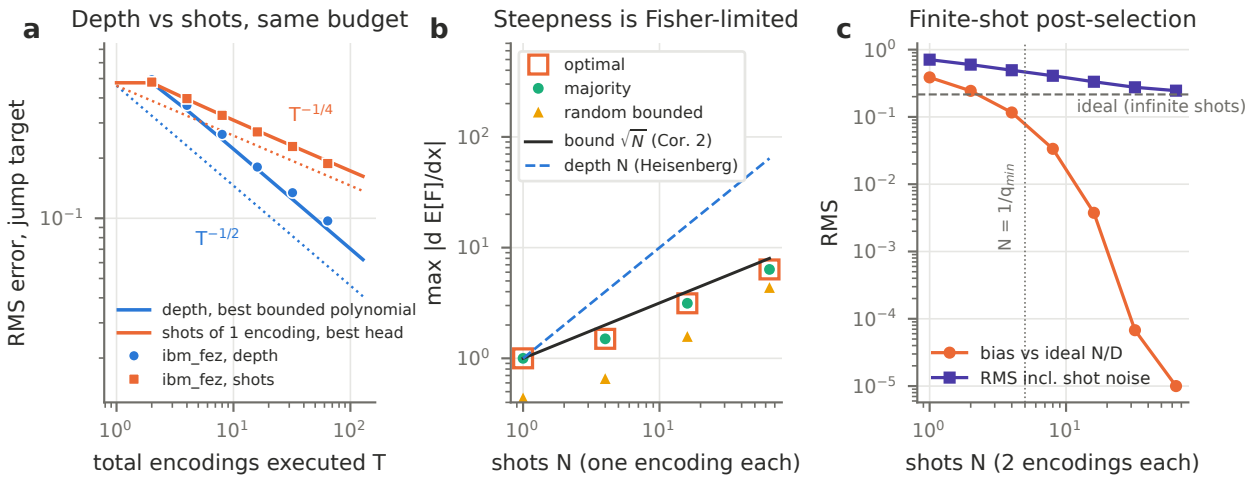


Figure 2: **Non-linearity is paid in encodings.** **a**, at a fixed total of T encodings, error on a jump for the class optimum of each arm: the best polynomial of degree T bounded by 1 (one coherent circuit) and the best bounded head on T single-encoding shots, which for a jump is the majority vote (lines, exact computation); markers, the same two arms measured on one qubit of ibm_fez, where the depth circuits realise the best polynomial bounded by 0.98 (within 6% of the class optimum) and the error is evaluated on the 128 design points; dotted lines: the lower bound $0.46\,(TE_{\max})^{-1/4}$ for each arm. **b**, the steepest slope of the expected output against the number of shots, for three heads, with the Fisher bound $\sqrt{N}$ (Corollary 2) and the depth scaling N; the head that minimises the error on a jump (open squares) is the majority vote at every budget, and the head of largest slope is steeper but stays below $\sqrt{N}$ (Supplementary Note 2). **c**, a post-selected model of the square wave $0.9\,\mathrm{sign}(\sin x)$ ($E = 2$, floor 0.2) estimated from N shots: the bias to the ideal rational output vanishes once $N \gtrsim 1/q_{\min}$, while the error including shot noise stays above the infinite-shot ideal.

and the margin by which $E = 16$ beat $E = 4$ (0.016) is comparable to its shot-noise standard error (about 0.01); the clear result is the reversal at $E = 64$.

2.7 The price of conditioning on outcomes

Conditioning on a measurement outcome is the one operation that leaves the class. A post-selected model with E encodings outputs $f = \mathcal{N}/D$, where $D = q(x)$ is the success probability and $\mathcal{N} = \mathrm{Tr}[(\Pi \otimes O)\rho(x)]$. By Lemma 1 both are trigonometric polynomials of degree at most E, with $|\mathcal{N}| \le D \le 1$. Rational functions can approximate kinks exponentially better than polynomials [13], which is the usual argument for post-selected and repeat-until-success neurons [6], and Mahadev and de Wolf showed that unrestricted post-selection realises exactly the rational degree [14]. How much of that advantage survives depends on the success floor $q = \min_x q(x)$.

Theorem 3 (Post-selection is a polynomial up to ε)**.** *Let $r = (1-\sqrt{q})/(1+\sqrt{q})$. For every $J \ge 0$ there is a trigonometric polynomial p_J of degree $\le (J+1)E$ with $\sup_x |f - p_J| \le 1/T_{J+1}(\frac{1+q}{1-q}) \le 2r^{J+1}$, where T_n is the Chebyshev polynomial.*

A post-selected model therefore reaches precision ε with the expressive power of at most $E\,(1 + \ln(2/\varepsilon)/(2\sqrt{q}))$ unitary encodings. The proof uses the Chebyshev residual polynomial of iterative linear solvers with condition number $1/q$ [53], which also underlies polynomial approximations of

$1/x$ in quantum linear-system algorithms [54, 55] and fixed-point amplitude amplification [56]. A model that post-selects at several stages is covered once its acceptance events are combined into one, whose probability D still has degree at most E. The decay rate r is attained: a three-qubit circuit with a single encoding and floor exactly q outputs $q/(\frac{1+q}{2} + \frac{1-q}{2}\cos x)$, whose Fourier coefficients decay exactly as $\sqrt{q}\, r^{|k|}$. The price is paid in shots. Because D is itself a polynomial of degree E, Bernstein's inequality bounds its curvature, and the mean shot overhead satisfies $\mathbb{E}_x[1/q(x)] \geq \frac{2}{\pi E\sqrt{q}} \arctan \frac{\pi E}{2\sqrt{q}} \approx 1/(E\sqrt{q})$, so a low floor cannot be confined to a single point. Combining the two results gives the main consequence.

Theorem 4 (Post-selection cannot change the approximation rate)**.** *If every trigonometric polynomial within 2ε of a target g has degree at least $K(2\varepsilon)$, every post-selected model with E encodings and error $\leq \varepsilon$ has $\sqrt{q} \leq E\ln(2/\varepsilon)/(2(K(2\varepsilon) - E))$ and hence a mean shot overhead $\gtrsim 2(K(2\varepsilon) - E)/(E^2 \ln(2/\varepsilon))$.*

Post-selection therefore trades encodings for shots at a bounded exchange rate and never turns a polynomial rate into an exponential one. Newman's $e^{-\sqrt{n}}$ rate for $|x|$, realised by post-selection, forces $q \lesssim \mathrm{poly}(E)\, e^{-2\sqrt{E}}$ and a shot overhead $\gtrsim e^{\sqrt{E}}/\mathrm{poly}(E)$. A repeat-until-success neuron whose attempts each succeed with probability at least $1/2$ is worth at most $1 + \ln(2/\varepsilon)/1.76$ unitary encodings per encoding, about 5 at $\varepsilon = 10^{-3}$.

Every admissible pair is realised by a circuit. Any $\mathcal{N}, D$ of degree E with $|\mathcal{N}| \leq D \leq 1$ can be implemented on two qubits with exactly E single-qubit encodings (Methods), by the construction behind quantum signal processing [57, 58] and single-qubit data re-uploading [59, 60]. The same construction on one qubit shows that every bounded real trigonometric polynomial of degree K is a unitary model with K encodings, so Theorem 3 is an operational equivalence once p_J is divided by $1 + \sup_x |f - p_J|$ to make it bounded, which at most doubles its error.

None of 124 trained post-selected models beat the floor of Theorem 3 or the shot-overhead bound. In RMS error over $x \in [-\pi, \pi)$, Theorem 3 gives the floor $\mathrm{floor}(E, q) = \max_J [d_{(J+1)E}(g) - 1/T_{J+1}(\frac{1+q}{1-q})]$, where $d_K(g)$ is the RMS distance from the target g to the trigonometric polynomials of degree K, and no post-selected model beats R_E, the best error of any rational function of degree E, so the error is at least $\max(\mathrm{floor}(E, q), R_E)$. For the square wave $g = 0.9\,\mathrm{sign}(\sin x)$, the target of this section, the optimum stays within a factor of 1.66 of this bound, and below a saturation floor (about 0.2 at $E = 2$) lowering q buys nothing. When encodings, shots and qubits are counted together, post-selection pays off only when depth is the binding constraint (Supplementary Note 3).

We also measured the price on hardware. For the square wave at $E = 2$ and 3 and floors $q \in \{0.5, 0.2, 0.05\}$, we built the optimal post-selected models as two-qubit circuits with exactly E encodings, together with the best unitary models of the same degree, and ran them on ibm_kingston. All pre-registered criteria passed. At fixed encodings post-selection changed the function class: at $E = 2$ the RMS error was 0.231, against 0.420 for the unitary model. Every error stayed above the floor of Theorem 3 at the measured $q_{\min}$, and every mean shot overhead above the Bernstein bound; both hold by construction, because errors and overheads are computed from degree-E fits to the measured probabilities (Methods). At $E = 3$, lowering the floor from 0.5 to 0.05 cut the error from 0.272 to 0.201 while $\mathbb{E}[1/q]$ rose from 1.43 to 3.61. At $E = 2$ the same change bought nothing below $q \approx 0.2$, the predicted saturation. The excess error over the ideal grew as the floor fell (from $+0.011$ to $+0.052$ at $E = 3$), the noise price of dividing by a small success probability (Fig. 3a,b; Supplementary Note 3). Degree-E polynomials fitted to the measured outcome probabilities left residuals at the shot-noise level, a direct hardware check of the closure lemma. A pre-registered scale-up to $E = 4$ on ibm_fez reproduced the pattern. Every post-selected model beat the unitary one (0.197–0.243 against 0.314), lowering the floor from 0.5 to 0.2 helped (0.216 to 0.197), and

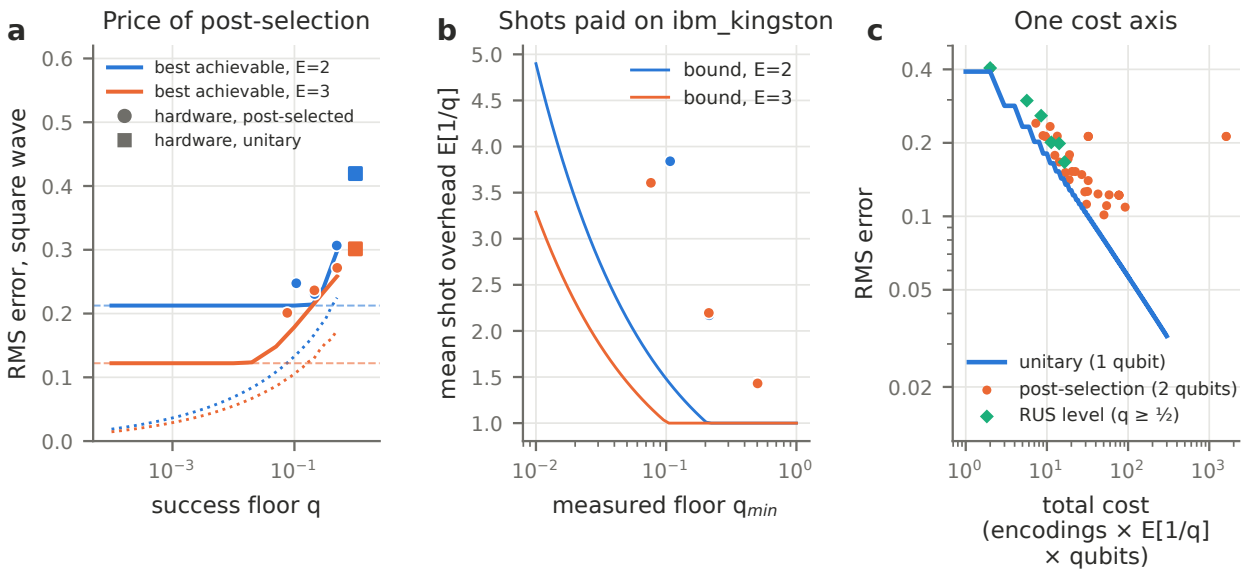


Figure 3: **The price of conditioning.** **a**, RMS error on the square wave $0.9\,\mathrm{sign}(\sin x)$ of the best post-selected model with E encodings against the success floor (solid), the floor of Theorem 3 (dotted), the degree limit R_E (dashed), and hardware runs of the optimal circuits on ibm_kingston (dots; squares: unitary models of the same degree). **b**, measured mean shot overhead against the shot-overhead bound. **c**, error against encodings × shot overhead × qubits, with every unit weighted equally (a convention; Supplementary Note 3); the unitary curve is the L^2 tail of the target beyond degree K, a lower bound on the error of every unitary model with K encodings.

the smallest floor gave the worst post-selected model (0.243 at $q = 0.05$), because the noise price outweighed the gain, as the same-day prediction had foreseen. All criteria passed.

3 Discussion

Non-linearity in a quantum model is a budget rather than a mechanism. Every finite-shot procedure computes a function whose analytic continuation is bounded by $\prod_i \cosh(E_i b)$, where E_i is the spectral range of the encodings in shot i, and the degree, steepness, curvature and spectrum of that function follow from the bound. Spent as spectral range within a shot, through depth or rescaling, the budget buys sharp features at the Heisenberg rate. Spent on repetitions of one encoding, it buys them only at the shot-noise rate. Hybrid models with classical heads are not exempt: they pay in shots, and in trained models we measured that price growing with the sharpness of the target. The budget is a physical cost only through the resource that range consumes in a given regime. When the scale of each encoding is fixed by the architecture, range costs depth and coherence caps it (Theorem 2); only error correction, when the encoding axis lies outside the Kraus span of the noise, keeps depth paying as models grow. When rescaling is free and the input is exact, the quantum model pays nothing for range, and the square-root penalty of repetition is the only price left; when rescaling is free and the input is not exact, its precision caps the range (the input-precision bound). Our hardware sits in the regime of free rescaling with exact inputs, which is why rescaled single gates matched coherent depth there. This is the quantum-metrology view of expressivity: a model can resolve only as much as its measurements learn about x. The same accounting of shots has been used to price the certification of reported Fisher geometries of hybrid models [61]; here it prices

expressivity itself.

The mechanisms usually listed as sources of quantum non-linearity are therefore not independent. Measurements and dissipative gates (reviewed in [6]), repeat-until-success circuits [6, 62], measurement back-action with outcome-dependent gates [5], non-Gaussian gates in continuous-variable networks [63], the tensor-product structure of multi-qubit encodings with entangling gates [64] and quantum reservoirs [65] all draw on the same budget. The steepness and exchange-rate bounds hold for any encoding with a bounded generator, through the quantum Fisher information, and the degree bound holds for Pauli encodings; continuous-variable encodings need an energy constraint (Methods). Apart from conditioning on outcomes, these mechanisms also produce the same class of functions. What separates them is the spectral support and the coefficient magnitudes that a circuit's structure allows, and both can be computed before training [20].

Conditioning on outcomes does change the class, and Theorems 3 and 4 put a price on it. Depth can be exchanged for shots at a rate fixed by the success floor, but the approximation rate cannot be improved. The exchange is visible on hardware, where lower floors bought accuracy at a measured shot price until the degree limit was reached, and added noise along the way.

Our results have limitations. The post-selection bound $\max(\mathrm{floor}(E, q), R_E)$ is tight to within a factor of 1.7 for jumps and loosens smoothly as the target gets smoother, reaching 4.6–5.8 times for a kink at $E = 4$ and up to 6.5 times for a triangle wave at $E \leq 6$, for a reason we identify (Supplementary Note 3). Of the sixteen hardware runs, ten passed every pre-registered criterion at the first scoring. Two passed with a qualification (a hybrid comparison confounded by a one-bit readout, and the IQM multi-scale run, whose first scoring failed two criteria before we corrected the shot-ordering error), one is inconclusive (the IonQ phase-robust run, which met its criteria but returned fitted visibilities above 1), and three missed (a tie within noise, a margin missed by 0.007 and a failed first IonQ run; Table 2 and Supplementary Note 6). Most of the hardware runs test consistency with a bound as well as a prediction that could have failed; Table 2 marks which is which, and the "same-day predictions" on IBM are outputs of the device's calibrated noise model, so agreement with them tests that model as much as the theory. Eleven further runs on selection rules and ceilings, four of them with a miss, are reported in the companion paper [20]. The noise theorem bounds Fisher information, and we have not shown that coherent blocks of E^* encodings attain the RMS bound that follows from it. In training, narrow two-qubit layouts never beat a fairly read-out hybrid model at matched encodings, the pre-registered task-level sharpness law failed, the test of entangling before encoding failed through trainability, and the one trained success used a layout chosen after the diagnosis of a failure (Supplementary Note 4). Supplementary Theorem 1, Corollary 4 and Theorem 2 bound capacity, not trainability. No experiment tests the Gaussian spectrum of Theorem 1 directly; its checks are numerical (Methods). Fitted exponents rest on three to five budgets each and carry bootstrap intervals (Methods). Most of the hardware evidence comes from IBM Heron processors.

4 Methods

Proof of Lemma 1

With $G_j = \sum_\lambda \lambda \Pi_\lambda$, the encoding acts as $\rho \mapsto \sum_{\lambda,\lambda'} e^{-ix(\lambda-\lambda')/2} \Pi_\lambda \rho \Pi_{\lambda'}$, a sum of x-independent linear maps weighted by $e^{-ix\delta}$ with $\delta = (\lambda - \lambda')/2$ an integer, $|\delta| \leq \Delta_j/2$. Averaged over outcomes, the circuit is a sum over branches of compositions of linear maps. Expanding each encoding and using linearity, every term carries $e^{-ix(\delta_1+\cdots+\delta_N)}$ over the N encodings on that branch, and $|\delta_1 + \cdots + \delta_N| \leq \sum_j \Delta_j/2$. (The bound counts ranges, not encodings: $G = 2Z$ has half-differences $0, \pm 2$, so one encoding carries frequency 2.)

Proof of Theorem 1 (master bound)

Purify channels and measurements, so that each record probability is $p_r(x) = \sum_k f_{rk}(x)\overline{f_{rk}(\bar{x})}$, a sum over the purifying register of products of entire amplitudes; the Cauchy–Schwarz steps below run over k as well. For one incoherent encoding on a pure qubit state φ, Cauchy–Schwarz gives $\sum_o |p_o(a+ib)| \le \|e^{-i(a+ib)Y/2}\varphi\|\,\|e^{-i(a-ib)Y/2}\varphi\| = ((\cosh b + y\sinh b)(\cosh b - y\sinh b))^{1/2} \le \cosh b$, with y the Y-expectation of the rotated state. Products multiply the bound, mixtures preserve the coefficient bound by convexity, and adaptive choices are bounded one shot at a time, so $|\mathbb{E}[F](a+ib)| \le \cosh(b)^n \le e^{nb^2/2}$. *Any depth.* For a shot with E encodings interleaved with arbitrary purified operations, write $e^{-i(a+ib)Y/2} = e^{-iaY/2}\cosh(b/2)(I+tY)$ with $t = \tanh(b/2)$. The amplitude operator is $A(b) = \cosh(b/2)^E P(t)$, where $P(t)$ is a product of unitaries, $e^{-iaY/2}$ and $(I+tY)$, and $A(-b) = \cosh(b/2)^E P(-t)$. By Cauchy–Schwarz and the parallelogram law, $\sum_o |p_o| \le \|A(b)\psi\|\,\|A(-b)\psi\| \le \cosh(b/2)^{2E}(\|P_{\text{even}}\psi\|^2 + \|P_{\text{odd}}\psi\|^2)$, where $P_{\text{even/odd}}$ collect even and odd powers of t. Take $b \ge 0$, since the bound is even in b. The degree-k part of P is a sum of $\binom{E}{k}$ terms, each t^k times an operator of norm 1, so $\|P_{\text{even}}\psi\| \le \frac{1}{2}((1+t)^E + (1-t)^E)$ and $\|P_{\text{odd}}\psi\| \le \frac{1}{2}((1+t)^E - (1-t)^E)$. The bracket is therefore at most $\frac{1}{2}((1+t)^{2E} + (1-t)^{2E})$, and since $\cosh(b/2)(1\pm t) = e^{\pm b/2}$ the mass is at most $\cosh(Eb)$. When feed-forward changes how many encodings follow, pad every branch to the largest count E: an x-independent controlled swap sends each surplus encoding to a fresh qubit in the $+1$ eigenstate of Y, where it only multiplies the amplitude by $e^{-ix/2}$ and leaves every record probability unchanged. Every branch then carries E encodings at fixed positions, and the expansion applies as written. Products over shots and adaptive choices are handled as above, so $|\mathbb{E}[F](a+ib)| \le \prod_i \cosh(E_i b) \le e^{Vb^2/2}$. Shifting the contour of $c_\omega = \frac{1}{2\pi}\int \mathbb{E}[F](a)e^{-i\omega a}da$ to $\operatorname{Im} x = -b\operatorname{sgn}\omega$ gives $|c_\omega| \le e^{Vb^2/2 - |\omega| b}$, and $b = |\omega|/V$ gives the claim. The tail beyond K sums to at most $\sqrt{2\pi V}\,e^{-K^2/2V}$. The bound is attained: E consecutive encodings give outcome probabilities $\cos^2(Ex/2)$ and $\sin^2(Ex/2)$, whose continuations have total mass $1 + 2\sinh^2(Eb/2) = \cosh(Eb)$, and independent shots of such circuits attain the product. Checks: random circuits stay below the bound (one qubit with random unitaries between the encodings, 0.92 of it at $E = 12$; an ancilla with dephasing, 0.88 at $E = 6$; branch-dependent encoding counts, at most 0.97), and maximising the mass over every unitary of a two-qubit circuit with a mid-circuit measurement whose outcome adds up to two encodings ($E_{\max} \le 8$) comes within 7×10^{-6} of the bound in all 12 cases and never exceeds it. The worst case over heads meets the coefficient bound at $\omega = 0$. *Other generators.* The bound holds for every generator whose eigenvalue half-differences are integers, with $E_i = \sum_j \Delta_j/2$. Subtracting the midpoint of the spectrum multiplies the amplitudes by a phase that leaves every p_r unchanged, and the shifted eigenvalues lie in $\{-n, -n+2, \dots, n\}$ with $n = \Delta_j/2$. An x-independent isometry that maps each eigenvector of eigenvalue λ to a basis state of Hamming weight $(n-\lambda)/2$ on n qubits (with an ancilla register for degeneracies) then turns the encoding into the n commuting Pauli encodings $e^{-ixZ_i/2}$.

Proof of Corollary 1 (encoding count)

A complete record $\tau = (o_1, \dots, o_n)$ has probability $P_x(\tau) = \prod_i P_x(o_i \mid o_{<i})$; the i-th factor is an outcome probability of the circuit chosen after $o_{<i}$, a trigonometric polynomial of degree at most its number of encodings by Lemma 1. Products add degrees, so $P_x(\tau)$ has degree $\le T(\tau)$ and $\mathbb{E}_x[F] = \sum_\tau F(\tau)P_x(\tau)$ has degree $\le T_{\max}$; classical randomness is a convex combination. For the second part, split the sum into records with $T \le k$ (degree $\le k$) and the rest, whose contribution is at most $\|F\|_\infty \Pr_x[T > k]$. This is the polynomial method of query complexity [8], used for phase estimation with arbitrary interleaved operations by van Dam et al. [66] and by Mande and de Wolf [67], with Fourier degree in place of algebraic degree, classical heads and

classical adaptivity across shots. The first part also follows from Theorem 1 by the contour shift, since $\prod_i \cosh(E_i b) \le e^{T|b|}$.

Proof of Corollary 2 (steepness)

$\frac{d}{dx}\mathbb{E}_x[F] = \mathrm{Cov}_x(F, \partial_x \log P_x(\tau))$, and the score has mean zero, so Cauchy–Schwarz gives $|\frac{d}{dx}\mathbb{E}_x[F]| \le \mathrm{sd}_x(F)\sqrt{I(x)}$. For adaptive records the score is a sum of conditional scores and $I = \sum_i \mathbb{E}[I_i \mid \text{past}]$. Each shot's Fisher information is at most its quantum Fisher information, which for E_i encodings $e^{-ixP/2}$ interleaved with arbitrary x-independent channels is at most E_i^2, the sequential Heisenberg bound [9, 10]. Exact enumeration confirms that the degree stays within the bound for majority, tanh and random-network heads ($N \le 32$), for an adaptive sequential test with random stopping (including the truncation bound) and for the finite-shot post-selection estimator ($N \le 64$), and that the maximal slope is at most $\sqrt{N}$ in all cases, with equality at $N = 1$.

Proof of Corollary 3

Lower bound. Let $|f'| \le s$ and $f(x_0) = a$, $|a| \le 1$, at a jump of g from -1 to $+1$. On the $+1$ side $f(x_0 + u) \le a + su$, so $(1 - f)^2 \ge (1 - a - su)^2$ for $0 \le u \le (1 - a)/s$; on the -1 side $(1 + f)^2 \ge (1 + a - su)^2$ for $0 \le u \le (1 + a)/s$. Integrating, $\int (f - g)^2 \ge [(1 - a)^3 + (1 + a)^3]/(3s) = (2 + 6a^2)/(3s) \ge 2/(3s)$ per jump. Each window extends at most $2/s$ from its jump and the jumps of $\mathrm{sign}(\cos x)$ are π apart, so for $s \ge 4/\pi$ the windows do not overlap and RMS $\ge \sqrt{2/(3\pi s)}$. Corollary 2 allows $s = \sqrt{TE_{\max}}$, which is $\ge 4/\pi$ whenever $TE_{\max} \ge 2$. *Depth.* The Fejér mean $\sigma_T g$ is a degree-T polynomial bounded by 1 (positive kernel); its kernel tail at distance d from a jump is at most $\pi/((T + 1)d)$, so $|\sigma_T g - g| \le \min(2, 2\pi/((T + 1)d))$ and RMS $\le 4(T + 1)^{-1/2}$; by realisability it is a circuit with exactly T encodings. *Shots.* The majority vote of N shots of $R_Y(x)$ errs with probability at most $\exp(-N \cos^2 x/2) \le \exp(-2Nu^2/\pi^2)$ at distance u from a jump (Hoeffding). The error of its expectation is $|\mathbb{E}[F] - g| \le 2\,\Pr[\text{err}]$ (equality for odd N; a tie outputs 0), so $(\mathbb{E}[F] - g)^2 \le 4e^{-4Nu^2/\pi^2}$, and integrating over the two jumps gives RMS $\le (2\sqrt{\pi})^{1/2} N^{-1/4} = 1.88\, N^{-1/4}$ (exact: 0.53–0.56 $N^{-1/4}$ for $N \ge 4$). For the mean-squared error of a single prediction, which also counts its variance, the squared error is at most 4 when the vote errs, which gives $(2\sqrt{2\pi})^{1/2} N^{-1/4} = 2.24\, N^{-1/4}$ (exact: 0.86–1.01 $N^{-1/4}$ for $N \ge 4$). The class-optimal constants quoted in Results are computed in Supplementary Note 2.

Proof of Corollary 4 (copies are shots)

The quantum Fisher information is convex in the state and additive on products. For one qubit under $e^{-ixwG/2}$ it is $4\,\mathrm{Var}(wG/2) \le w^2$, so a separable pre-encoding state gives $F_Q \le \sum_j w_j^2$. Channels, measurements and feed-forward after the encoding cannot increase the Fisher information of the record, and Corollary 2 gives the slope bound. The state $(|\lambda_{\min}\rangle + |\lambda_{\max}\rangle)/\sqrt{2}$ of $H = \sum_j w_j G_j/2$ has $F_Q = (\sum_j |w_j|)^2$, and the observable $|\lambda_{\min}\rangle\langle\lambda_{\max}| + \text{h.c.}$ gives $\cos((\sum_j |w_j|)x)$. The jump bound follows as in Corollary 3. Numerically, 400 random separable-input circuits with ancillas and random heads never exceed $\sqrt{m}$; the best head on m local readouts reaches 0.86–0.92$\sqrt{m}$.

Proof of Theorem 2 (noise)

For T uses of a channel whose Kraus span contains the generator, the channel-extension method bounds the quantum Fisher information of any strategy (parallel with ancillas, or sequential with arbitrary control, which reduces to it [48]) by cT; for dephasing $c = \eta^2/(1 - \eta^2)$ [46, 10]. Corollary 2 turns this into the slope bound. A coherent block of E encodings has contrast η^E and phase Ex,

hence Fisher information $E^2\eta^{2E}$ per shot and $TE\eta^{2E}$ per budget, maximal at $E^* = 1/(2\ln\eta^{-1})$ (when $\eta \geq e^{-1/2}$, else $E^* = 1$) where it equals $T/(2e\ln\eta^{-1})$. Its ratio to cT is $(1-\eta^2)/(2e\eta^2\ln\eta^{-1})$, which tends to $1/e$ as $\eta \to 1$. For general channels violating the Hamiltonian-not-in-Kraus-span condition the constant c_N is an SDP [47, 10].

Proof of Theorem 3

$Q(t) = T_{J+1}(\frac{1+q-2t}{1-q})/T_{J+1}(\frac{1+q}{1-q})$ has $Q(0) = 1$, so $Q(t) = 1 - tP_J(t)$ with $\deg P_J = J$, and $|Q| \leq 1/T_{J+1}(\frac{1+q}{1-q})$ on $[q, 1]$. Put $p_J = \mathcal{N} P_J(D)$. Then $f - p_J = f\,Q(D)$ and $|f| \leq 1$.

Proof of the shot-overhead bound

Let D be a real trigonometric polynomial of degree E with $q \leq D \leq 1$ and $\min D = q$ at x_0. Bernstein's inequality for $D - \frac{1}{2}$ (sup norm at most $\frac{1}{2}$) gives $|D''| \leq E^2/2$, so $D(x) \leq q + \frac{E^2}{4}(x - x_0)^2$. Integrating $1/(q + E^2u^2/4)$ over $u \in [-\pi, \pi]$ and dividing by 2π gives $\mathbb{E}_x[1/D] \geq \frac{2}{\pi E\sqrt{q}} \arctan \frac{\pi E}{2\sqrt{q}}$.

Proof of Theorem 4

Take $J + 1 = \lceil \ln(2/\varepsilon)/\ln(1/r) \rceil$, so $2r^{J+1} \leq \varepsilon$. Theorem 3 gives a polynomial of degree $(J + 1)E$ within ε of f, hence within 2ε of g, so $(J + 1)E \geq K(2\varepsilon)$. With $\ln(1/r) \geq 2\sqrt{q}$ this is the stated bound on $\sqrt{q}$, and the shot-overhead bound converts it into a bound on shots. The degree-limit term follows because dropping the floor only enlarges the class: no post-selected model beats R_E, the best error of any degree-E rational function.

Realisability of every admissible pair

For $|\mathcal{N}| \leq D \leq 1$ of degree E the outcome probabilities $p_{1,2} = (D \pm \mathcal{N})/2$, $p_3 = 1 - D$ are non-negative trigonometric polynomials of degree $\leq E$ summing to one. Fejér–Riesz writes $p_i = |a_i(e^{ix})|^2$, so $v(z) = (a_1, a_2, a_3, 0)$ is a unit polynomial vector in $\mathbb{C}^4$. Its lowest and highest coefficients are orthogonal, so a rank-2 projector P with $Pc_0 = 0$ and $(I - P)c_E = 0$ lowers the degree by one: $w = (I - P + z^{-1}P)v$. After E steps $v(z) = \prod_j (I - P_j + zP_j)w_0$, and each factor is $W_j\,\mathrm{diag}(z, z, 1, 1)\,W_j^\dagger$, i.e. one $R_Z(x)$ on the first qubit between fixed two-qubit unitaries. Measuring both qubits and post-selecting outcomes 1–2 gives $\mathcal{N}/D$ with exactly E encodings (checked to 2×10^{-14} on 24 random instances and used for the hardware runs).

Scope

Every result assumes that x enters only through encoding gates. Mechanisms outside that assumption, and what the bounds say about them: (i) a classical map $x \mapsto \phi(x)$ applied before encoding, trained or not, or a non-integer scale w in $R_Y(wx)$, leaves the integer frequency class; the degree and Gaussian bounds do not apply, and only the Fisher bound survives (below); (ii) a choice of circuit, scale or shot count that depends on x itself, rather than on measured outcomes, is likewise excluded, whereas an outcome-dependent rescaling $R_Y(g(\text{bits})\,x)$ is covered with E the largest range on any branch; (iii) heads with $\|F\|_\infty > 1$, such as a ratio estimator before clipping, scale every bound by $\|F\|_\infty$; (iv) non-linearity in the quantum state (swap tests, kernel estimates) needs copies of the state and is a different question. Corollary 2 uses only the Fisher information of the record, so it holds for every encoding: amplitude encodings (where the degree bound is void, because normalisation already makes the state non-linear in x), trainable frequency scalings $e^{-iw_jxG_j}$ (Fisher information $\leq (\sum_j |w_j|\|G_j\|)^2$ per shot, with $\|G\|$ the spectral range), and Hamiltonian

encodings with non-integer spectra, through the quantum Fisher information of the encoded state. Corollary 1 is the algebraic counterpart for encodings with integer eigenvalue half-differences, and the lower bound of Corollary 3 holds for every encoding. Theorem 1 holds for every generator with integer eigenvalue half-differences, each encoding counted by its range (proof of Theorem 1); a generator with spectrum $\{-2, 0, 2\}$, for example, counts as two encodings. For several inputs the degree bound holds per variable and the slope bound per direction, with the Fisher information matrix. Bounded generators are needed: a continuous-variable displacement encoding has quantum Fisher information $4\,\mathrm{Var}(p)$ and needs an energy constraint. Corollary 4 holds more generally when the state is separable across blocks that each hold one encoded qubit and its own ancillas.

Hardware analysis of the post-selection runs

By the closure lemma the measured outcome probabilities are trigonometric polynomials of degree $\leq E$. We fit each by least squares on the 16 measured points and evaluate $\mathcal{N}/D$ on 4,096 points; the continuous error, $q_{\min}$ and $\mathbb{E}[1/q]$ come from these fits, with multinomial bootstrap errors. The residual of the fits was 0.5–2.0 times its shot-noise expectation, a direct check of the closure lemma on hardware. Because the fitted $\mathcal{N}$ and D have degree E, the floor of Theorem 3 and the shot-overhead bound hold for them by construction (up to small violations of $|\mathcal{N}| \leq D$ in the fits), so those criteria check the analysis rather than the device.

Statistics

No trained post-selected model fell below the floor (124 models, which share two targets and five seeds); since the floor is a theorem, this checks the code rather than estimating a rate. Hardware errors are shot-noise standard errors from the estimator, multinomial bootstraps (post-selection) or block statistics over repeated settings (Open Quantum runs). Every fitted hardware exponent carries a parametric bootstrap 95% interval (4,000 replicates, `exponent_ci.py`): for the exchange-rate run each point estimate is redrawn from its binomial or block law and the bias-corrected RMS and the slope are recomputed; for the multi-scale runs, where only the per-budget RMS and block counts are stored, the per-point squared errors (128 on IBM, 120 on IQM) are resampled as independent, which is approximate. The intervals are: depth 0.48 $[0.45, 0.50]$ against shots 0.27 $[0.24, 0.29]$ (difference $[0.18, 0.24]$); multi-scale 0.55 $[0.43, 0.67]$ against equal 0.23 $[0.11, 0.34]$ on IBM (difference $[0.15, 0.48]$) and 0.57 $[0.45, 0.69]$ against 0.25 $[0.13, 0.37]$ on IQM (difference $[0.15, 0.49]$). The exact-computation exponents in the text are deterministic fits over the stated ranges and carry no sampling error.

Pre-registration and the hardware experiments

Each experiment was frozen in a dated note with thresholds, a pass rule and its interpretation fixed in advance, and committed with the same-day prediction before submission. Supplementary Table 1 lists, for every run, the commit that froze it, its time and the submission time of the job, read from the stored result (`prereg_record.py`). The earliest run reported here (September 2026) is not pre-registered in the version-control sense: its note and prediction file were written before the job but committed the following morning. Every later pre-registration and prediction was committed before the job, with margins from 5 seconds to three hours; the four notes that were touched again after their job record an implementation deviation or a provider-side job failure, and none changed a criterion. On IBM the prediction came from the device's calibrated Aer noise model; on IQM and IonQ, where no calibrated model was available, the prediction was a noiseless or inline expectation. Two learning experiments were re-run under a revised criterion after a miss; every version is on

record and counted (Supplementary Note 7). Deviations and errata are recorded with the data. Of the 16 hardware runs reported here, 10 passed every gating criterion at the first scoring, 2 passed with a qualification, 1 is inconclusive and 3 missed (Table 2); the eleven selection-rule and ceiling runs of the same programme, seven of which passed, are reported in the companion [20]. The handling of aggregated counts on the Open Quantum devices and the qubit layout of the IQM run are described in Supplementary Note 6.

Table 2: Pre-registered hardware experiments reported here (IBM Heron unless noted; eleven further runs are in the companion [20]). Cost as charged: QPU seconds on IBM, Open Quantum credits (cr.) on IQM Garnet and IonQ Forte-1. Type: P, a prediction that could have failed (an exponent, a class change, a slope above a bound, a non-monotonic pattern); C, a consistency check that noise channels can only move in the allowed direction (a degree or slope below a bound) or that holds by construction (the post-selection floors); P+C, both kinds among the criteria. *Note entered version control only after the job.

Experiment	device	QPU s / cr.	criteria	type	outcome
Post-selection, trained models*	kingston	79	4	P+C	3 of 4 (tie vs 5 encodings)
Price of post-selection	kingston	140	7	P+C	pass
Depth-for-shots exchange rate	fez	123	3	P	pass
Resolution limit, depth vs shots	fez	175	4	P	pass
Adaptive dynamic circuit	fez	37	2	P+C	pass
Price of post-selection at $E = 4$	fez	88	5	P+C	pass
Copies vs coherence, 2–24 qubits	fez	112	4	P+C	pass
Noisy allocation law	fez	135	3	P	1 missed by 0.007
Allocation law, exact dephasing	fez	205	3	P	pass
Trained models on hardware	fez	67	3	P	pass; hybrid comparison confounded
Trained models, light-cone-fixed	fez	67	3	P	pass
Rescaled encodings, multi-scale shots	kingston	416	4	P	pass
Copies vs coherence, 4–16 qubits	IonQ Forte-1	25 cr.	2	P	1 failed (phase offsets)
Separable limit, phase-robust	IQM Garnet	12 cr.	1	P	pass
Separable limit, phase-robust	IonQ Forte-1	90 cr.	2	P	inconclusive (post-hoc bound; phase drift)
Rescaled encodings, multi-scale shots	IQM Garnet	194 cr.	4	P	pass after rescoring (first scoring failed 2)

Simulation. Statevector and density-matrix simulation in NumPy and PennyLane 0.44.0. Spectra from FFTs over $2N_{\text{enc}} + 3$ equispaced points (128 points for the closure table, Table 1), averaged over random parameter draws; a frequency is in the support if its mean power exceeds 10^{-12} (in Table 1, if its power exceeds 10^{-14} of the peak).

Hardware. Qiskit 2.5.2, Qiskit Aer 0.17.2 and Qiskit IBM Runtime 0.49.0; EstimatorV2 with TREX, gate twirling and automatic dynamical decoupling off unless stated; the noise model and the device calibration are those of the job date (dates in the result file names; job identifiers in the repository). Thresholds, job identifiers, QPU seconds and raw data are in the repository. Post-selection runs used SamplerV2 with 4,096 shots per point and no readout mitigation; continuous errors were computed from degree-E least-squares fits to the measured outcome probabilities, which the closure lemma justifies.

Use of large language models

Generative AI assistants were used substantively alongside the authors in writing the experiment and scoring code, drafting the pre-registration, searching the literature, and drafting and revising the text. The authors set the research question, made and approved the pre-registered decisions, checked the cited sources, read the final text, and are responsible for its content. Every hardware job was submitted only after an author approved it. All reported results are computed by script from the released artifacts.

Data availability

The pre-registrations, same-day predictions, raw hardware results with job identifiers, and all simulation and training outputs that support the findings of this study are available at `https://drive.google.com/file/d/1j9aALzKzIhe-RdIA6tbdoJ8GF2Osu44o` and will be deposited in Zenodo with a DOI on publication. The version-control history that dates each pre-registration (Supplementary Table 1) is available to the editors and referees on request.

Code availability

The code for every simulation, theory check, learning test and hardware experiment is available at `https://drive.google.com/file/d/1j9aALzKzIhe-RdIA6tbdoJ8GF2Osu44o`. A single command re-scores every hardware experiment reported here against its pre-registered criteria from the stored results, regenerates the figures and builds this manuscript; a full run also repeats the theory checks and the training. The repository's documentation describes both.

Acknowledgements

We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team. Access to IQM Garnet and IonQ Forte-1 was provided through Open Quantum.

Author contributions

P.S. conceived the research programme and the QNLEP benchmark it builds on, derived the master bound, its corollaries and the noise, post-selection and input-precision results, designed, pre-registered and ran the hardware experiments on IBM, IQM and IonQ processors, carried out the simulations and learning tests, built the reproduction pipeline, and wrote the manuscript. C.L. wrote the ancilla-tracing circuit implementation and the IBM Quantum runtime interface on which

the experiments build, co-authored the QNLEP benchmark, and led the pre-submission reviews of the manuscript and the corrections they produced.

Funding

No funding was received for this work.

Competing interests

All authors declare no financial or non-financial competing interests.

References

[1] Gil Vidal, F. J. & Theis, D. O. Input redundancy for parameterized quantum circuits. *Front. Phys.* **8**, 297 (2020).

[2] Schuld, M., Sweke, R. & Meyer, J. J. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Phys. Rev. A* **103**, 032430 (2021).

[3] Gan, B. Y., Leykam, D. & Angelakis, D. G. Fock state-enhanced expressivity of quantum machine learning models. *EPJ Quantum Technol.* **9**, 16 (2022).

[4] Heredge, J., Hill, C. D., Hollenberg, L. C. L. & Sevior, M. Non-unitary quantum machine learning. Preprint at https://arxiv.org/abs/2405.17388 (2024).

[5] Argyle, P., Lakhdar-Hamina, D., Miller, S. H. & Galitski, V. Measurement-induced quantum neural network. Preprint at https://arxiv.org/abs/2603.19200 (2026).

[6] Cao, Y., Guerreschi, G. G. & Aspuru-Guzik, A. Quantum neuron: an elementary building block for machine learning on quantum computers. Preprint at https://arxiv.org/abs/1711.11240 (2017).

[7] Sulimov, P. & Lehmann, C. A reproducible evaluation protocol for quantum non-linearity in variational quantum models. In *Advances in Neural Information Processing Systems 39* (NeurIPS, 2026); `https://openreview.net/forum?id=37XAXg90Dc`.

[8] Beals, R., Buhrman, H., Cleve, R., Mosca, M. & de Wolf, R. Quantum lower bounds by polynomials. *J. ACM* **48**, 778–797 (2001).

[9] Giovannetti, V., Lloyd, S. & Maccone, L. Quantum metrology. *Phys. Rev. Lett.* **96**, 010401 (2006).

[10] Demkowicz-Dobrzański, R., Kołodyński, J. & Guţă, M. The elusive Heisenberg limit in quantum-enhanced metrology. *Nat. Commun.* **3**, 1063 (2012).

[11] Lorentz, G. G. The degree of approximation by polynomials with positive coefficients. *Math. Ann.* **151**, 239–251 (1963).

[12] Holtz, O., Nazarov, F. & Peres, Y. New coins from old, smoothly. *Constr. Approx.* **33**, 331–363 (2011).

[13] Newman, D. J. Rational approximation to $|x|$. *Michigan Math. J.* **11**, 11–14 (1964).

[14] Mahadev, U. & de Wolf, R. Rational approximations and quantum algorithms with postselection. *Quantum Inf. Comput.* **15**, 295–307 (2015).

[15] Berberich, J., Fink, D., Pranjić, D., Tutschku, C. & Holm, C. Training robust and generalizable quantum models. *Phys. Rev. Res.* **6**, 043326 (2024).

[16] Higgins, B. L., Berry, D. W., Bartlett, S. D., Wiseman, H. M. & Pryde, G. J. Entanglement-free Heisenberg-limited phase estimation. *Nature* **450**, 393–396 (2007).

[17] Shin, S., Teo, Y. S. & Jeong, H. Exponential data encoding for quantum supervised learning. *Phys. Rev. A* **107**, 012422 (2023).

[18] Aaronson, S. Quantum computing, postselection, and probabilistic polynomial-time. *Proc. R. Soc. A* **461**, 3473–3482 (2005).

[19] Dale, H., Jennings, D. & Rudolph, T. Provable quantum advantage in randomness processing. *Nat. Commun.* **6**, 8203 (2015).

[20] Sulimov, P. & Lehmann, C. Hidden selection rules and amplitude ceilings in quantum Fourier models. Preprint at https://arxiv.org/abs/2611.01314 (2026).

[21] Nemkov, N. A., Kiktenko, E. O. & Fedorov, A. K. Fourier expansion in variational quantum algorithms. *Phys. Rev. A* **108**, 032406 (2023).

[22] Wiedmann, M., Periyasamy, M. & Scherer, D. D. Fourier analysis of variational quantum circuits for supervised learning. In *Proc. IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2025); preprint at https://arxiv.org/abs/2411.03450.

[23] Holzer, P. & Turkalj, I. Spectral invariance and maximality properties of the frequency spectrum of quantum neural networks. *Quantum Mach. Intell.* **8**, 33 (2026).

[24] Casas, B. & Cervera-Lierta, A. Multidimensional Fourier series with quantum circuits. *Phys. Rev. A* **107**, 062612 (2023).

[25] Mhiri, H. et al. Constrained and vanishing expressivity of quantum Fourier models. *Quantum* **9**, 1847 (2025).

[26] Barthe, A. & Pérez-Salinas, A. Gradients and frequency profiles of quantum re-uploading models. *Quantum* **8**, 1523 (2024).

[27] Jaderberg, B., Gentile, A. A., Ghosh, A., Elfving, V. E., Jones, C., Vodola, D., Manobianco, J. & Weiss, H. Let quantum neural networks choose their own frequencies. *Phys. Rev. A* **109**, 042421 (2024).

[28] Caro, M. C., Gil-Fuster, E., Meyer, J. J., Eisert, J. & Sweke, R. Encoding-dependent generalization bounds for parametrized quantum circuits. *Quantum* **5**, 582 (2021).

[29] Jerbi, S., Fiderer, L. J., Poulsen Nautrup, H., Kübler, J. M., Briegel, H. J. & Dunjko, V. Quantum machine learning beyond kernel methods. *Nat. Commun.* **14**, 517 (2023).

[30] Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A. & Woerner, S. The power of quantum neural networks. *Nat. Comput. Sci.* **1**, 403–409 (2021).

[31] Landman, J., Thabet, S., Dalyac, C., Mhiri, H. & Kashefi, E. Classically approximating variational quantum machine learning with random Fourier features. In *International Conference on Learning Representations (ICLR)* (2023).

[32] Pérez-Salinas, A., López-Núñez, D., García-Sáez, A., Forn-Díaz, P. & Latorre, J. I. One qubit as a universal approximant. *Phys. Rev. A* **104**, 012405 (2021).

[33] Montanaro, A. Quantum speedup of Monte Carlo methods. *Proc. R. Soc. A* **471**, 20150301 (2015).

[34] Franz, M. et al. Out of tune: demystifying noise-effects on quantum Fourier models. Preprint at https://arxiv.org/abs/2506.09527 (2025).

[35] Recio-Armengol, E., Eisert, J. & Meyer, J. J. Single-shot quantum machine learning. *Phys. Rev. A* **111**, 042420 (2025).

[36] Gilyén, A., Arunachalam, S. & Wiebe, N. Optimizing quantum optimization algorithms via faster quantum gradient computation. In *Proc. 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* 1425–1444 (SIAM, 2019).

[37] Marvian, I. & Spekkens, R. W. Modes of asymmetry: the application of harmonic analysis to symmetric quantum dynamics and quantum reference frames. *Phys. Rev. A* **90**, 062110 (2014).

[38] Kitaev, A. Y. Quantum measurements and the Abelian stabilizer problem. Preprint at https://arxiv.org/abs/quant-ph/9511026 (1995).

[39] Berry, D. W. et al. How to perform the most accurate possible phase measurements. *Phys. Rev. A* **80**, 052114 (2009).

[40] Keane, M. S. & O'Brien, G. L. A Bernoulli factory. *ACM Trans. Model. Comput. Simul.* **4**, 213–219 (1994).

[41] Nacu, S. & Peres, Y. Fast simulation of new coins from old. *Ann. Appl. Probab.* **15**, 93–115 (2005).

[42] Qian, W., Riedel, M. D. & Rosenberg, I. Uniform approximation and Bernstein polynomials with coefficients in the unit interval. *Eur. J. Combin.* **32**, 448–463 (2011).

[43] Pezzé, L. & Smerzi, A. Entanglement, nonlinear dynamics, and the Heisenberg limit. *Phys. Rev. Lett.* **102**, 100401 (2009).

[44] Monz, T. et al. 14-qubit entanglement: creation and coherence. *Phys. Rev. Lett.* **106**, 130506 (2011).

[45] Omran, A. et al. Generation and manipulation of Schrödinger cat states in Rydberg atom arrays. *Science* **365**, 570–574 (2019).

[46] Escher, B. M., de Matos Filho, R. L. & Davidovich, L. General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology. *Nat. Phys.* **7**, 406–411 (2011).

[47] Fujiwara, A. & Imai, H. A fibre bundle over manifolds of quantum channels and its application to quantum statistics. *J. Phys. A* **41**, 255304 (2008).

[48] Demkowicz-Dobrzański, R. & Maccone, L. Using entanglement against noise in quantum metrology. *Phys. Rev. Lett.* **113**, 250801 (2014).

[49] Lu, R. et al. A unified frequency principle for quantum and classical machine learning. Preprint at https://arxiv.org/abs/2601.03169 (2026).

[50] Huelga, S. F. et al. Improvement of frequency standards with quantum entanglement. *Phys. Rev. Lett.* **79**, 3865–3868 (1997).

[51] Demkowicz-Dobrzański, R., Czajkowski, J. & Sekatski, P. Adaptive quantum metrology under general Markovian noise. *Phys. Rev. X* **7**, 041009 (2017).

[52] Zhou, S., Zhang, M., Preskill, J. & Jiang, L. Achieving the Heisenberg limit in quantum metrology using quantum error correction. *Nat. Commun.* **9**, 78 (2018).

[53] Saad, Y. *Iterative Methods for Sparse Linear Systems* 2nd edn (SIAM, 2003).

[54] Childs, A. M., Kothari, R. & Somma, R. D. Quantum algorithm for systems of linear equations with exponentially improved dependence on precision. *SIAM J. Comput.* **46**, 1920–1950 (2017).

[55] Gilyén, A., Su, Y., Low, G. H. & Wiebe, N. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proc. 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC)* 193–204 (ACM, 2019).

[56] Yoder, T. J., Low, G. H. & Chuang, I. L. Fixed-point quantum search with an optimal number of queries. *Phys. Rev. Lett.* **113**, 210501 (2014).

[57] Low, G. H. & Chuang, I. L. Optimal Hamiltonian simulation by quantum signal processing. *Phys. Rev. Lett.* **118**, 010501 (2017).

[58] Motlagh, D. & Wiebe, N. Generalized quantum signal processing. *PRX Quantum* **5**, 020368 (2024).

[59] Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E. & Latorre, J. I. Data re-uploading for a universal quantum classifier. *Quantum* **4**, 226 (2020).

[60] Yu, Z., Yao, H., Li, M. & Wang, X. Power and limitations of single-qubit native quantum neural networks. In *Advances in Neural Information Processing Systems 35* 27810–27823 (NeurIPS, 2022).

[61] Sulimov, P. & Lehmann, C. What does it cost to trust a quantum model? Shot-priced certification for hybrid QML. In *NeurIPS 2026 Workshop on Secure and Trustworthy Quantum Machine Learning (SaTQuML)* (non-archival, 2026); `https://openreview.net/forum?id=4wUj6qZBTR`.

[62] Moreira, M. S. et al. Realization of a quantum neural network using repeat-until-success circuits in a superconducting quantum processor. *npj Quantum Inf.* **9**, 118 (2023).

[63] Killoran, N. et al. Continuous-variable quantum neural networks. *Phys. Rev. Res.* **1**, 033063 (2019).

[64] Mitarai, K., Negoro, M., Kitagawa, M. & Fujii, K. Quantum circuit learning. *Phys. Rev. A* **98**, 032309 (2018).

[65] Fujii, K. & Nakajima, K. Harnessing disordered-ensemble quantum dynamics for machine learning. *Phys. Rev. Appl.* **8**, 024030 (2017).

[66] van Dam, W., D'Ariano, G. M., Ekert, A., Macchiavello, C. & Mosca, M. Optimal quantum circuits for general phase estimation. *Phys. Rev. Lett.* **98**, 090501 (2007).

[67] Mande, N. S. & de Wolf, R. Tight bounds for quantum phase estimation and related problems. In *Proc. 31st Annual European Symposium on Algorithms (ESA)* (Schloss Dagstuhl, 2023); preprint at https://arxiv.org/abs/2305.04908.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryinformation.pdf