

Quantum Kolmogorov--Arnold networks are classically reproducible at their attained accuracy, and a discrete-logarithm encoding need not be: a measurement-based map of where quantum advantage can live

Hikaru Wakaura

hikaruwakaura@gmail.com

QuantScape Inc

Research Article

Keywords: quantum Kolmogorov--Arnold networks (QKANs) , tensor network (TN) , IQP kernel

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10200068/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Quantum Kolmogorov–Arnold networks are classically reproducible at their attained accuracy, and a discrete-logarithm encoding need not be: a measurement-based map of where quantum advantage can live

Hikaru Wakaura

QIRI (Quantum Integrated Research Institute Inc.), Tokyo 107-0061, Japan
June 22, 2026

Quantum machine-learning models are useful only if classical computers cannot cheaply imitate them. We test this directly for quantum Kolmogorov–Arnold networks (QKANs)—neural networks whose tunable “activation functions” are small quantum circuits—by asking whether each model can be contracted as an efficient tensor network (TN). Our task-aware yardstick is the bond dimension a classical surrogate needs to copy the model’s output at the accuracy the model itself reaches. We find a clear ladder. The amplitude-encoded Chebyshev QKAN uses only logarithmically many qubits and is exactly classical. The variational QKAN, built from all-to-all controlled rotations, looks hard at random settings but generalizes only weakly once trained (held-out $R^2 \leq 0.27$) and is copied at that attained accuracy by a low-bond-dimension surrogate ($n \leq 10$); at fixed strict tolerance the required bond dimension grows near-exponentially. Depolarizing noise erases it (purity collapse) near $p \approx 0.03$ – 0.05 at the sizes we reach. Every generic attempt to force hardness—a global measurement or an interaction-based (IQP) feature map—instead destroys trainability through exponentially small gradients (a “barren plateau”). The escape is a discrete-logarithm feature map: its kernel variance is flat from 4 to 14 qubits while a generic IQP kernel concentrates exponentially, and a support-vector machine with it learns a classically hard concept to $\geq 99\%$ accuracy while generic and classical kernels stay at chance; wired into the KAN architecture, it fits discrete-log targets to $R^2 \approx 1$. Hardness placed in number theory stays classically hard, shot-efficient, and learnable. The work provides a measurement-based reproducibility criterion and a design rule: seek advantage in structured, provably hard encodings rather

than in deep entanglement.

1 Introduction

Quantum computers promise speed-ups, but only for tasks a classical computer cannot already do efficiently. In quantum machine learning this caveat is sharp: many quantum models can be reproduced by classical algorithms, erasing any advantage [1, 4]. The decisive question for any proposed quantum model is therefore not “does it run on a quantum device?” but “can a classical computer imitate it cheaply?”

The cleanest classical imitator is the *tensor network* (TN)—a way of writing a quantum circuit as a graph of small tensors that are multiplied together (“contracted”) [5]. Any quantum circuit is a tensor network; the cost of contracting it classically is set by one number, the *bond dimension* χ , which counts how much quantum entanglement crosses a cut through the circuit. When entanglement obeys an “area law” the bond dimension stays small ($\chi = \text{poly}$) and a matrix-product state (MPS) simulates the system efficiently [6]; when it reaches a “volume law” the bond dimension explodes ($\chi \sim 2^{n/2}$ for n qubits) and no efficient contraction is known. This same dial—how fast χ grows—explains both the classical reproduction of quantum-supremacy experiments [7] and the collapse of noisy random-circuit sampling into the classically easy regime [8, 9].

Quantum Kolmogorov–Arnold networks (QKANs) are a fast-growing model family in which each learnable edge function of a (classical) Kolmogorov–Arnold network [10] is realised by a quantum subroutine [11, 12]; several variants are proposed as routes to quantum advantage. However, whether any concrete QKAN actually resists efficient tensor-network simulation—and whether such resistance can coexist with the ability to train the model—has not been established.

In this work we address the question, for two published QKAN constructions, by *measuring* their tensor-network simulability rather than assuming it.

Hikaru Wakaura: h.wakaura@deeptell.jp

We introduce a task-aware bond-dimension criterion and apply it along a sequence of increasingly entangling QKANs. We find that the Chebyshev amplitude QKAN is exactly classical; that the variational QKAN is volume-law at random parameters, learns only weakly, and is reproducible at its attained accuracy while hardening near-exponentially at fixed accuracy; that depolarizing noise erases it by $p \approx 0.05$; and that every generic route to hardness collides with a barren plateau [13] or kernel concentration [14]. We then show numerically that a single structured construction—a discrete-logarithm feature map [17]—threads the needle, remaining classically hard while staying shot-efficiently estimable. The conceptual ingredients are established: tensor-network dequantization of variational models, including the bound that a model trained on M_t data lies in an MPS space of bond dimension $\chi \leq M_t$ [2]; the trainability–dequantization relation [3]; classical surrogates [1]; area-law simulability [6]; barren plateaus [13]; kernel concentration [14]; and the discrete-logarithm speed-up [17]. Building on these, our contribution is to (i) apply a task-aware, accuracy-matched bond-dimension diagnostic to *concrete published QKAN constructions*; (ii) attribute the hardness to its source—encoding versus ansatz—identifying a distinct low-entanglement, construction-hard category (the discrete-logarithm map); and (iii) exhibit, for this family, the empirical hardness–trainability map together with a shot-efficient learnability separation. All technical detail needed to reproduce every number is given in the Supplemental Material (SM) [27]; data and code paths are recorded in a separate provenance file [28].

2 Models and methods

We define every abbreviation at first use: TN (tensor network), MPS (matrix-product state), MPO (matrix-product operator), OSEE (operator-space entanglement entropy), IQP (instantaneous quantum polynomial-time), SVM (support-vector machine), BP (barren plateau), RMS (root-mean-square), SPSA (simultaneous-perturbation stochastic approximation). Single-qubit gates are $R_y(\theta) = \exp(-i\theta Y/2)$ and $R_z(\theta) = \exp(-i\theta Z/2)$ with Pauli matrices X, Y, Z ; a controlled- R_y (written CR_y) applies R_y to a target qubit conditioned on a control qubit being $|1\rangle$.

2.1 The two QKAN constructions

Chebyshev amplitude QKAN (CCQKAN). Each edge activation is the Chebyshev expansion $a(x) = \sum_{k=0}^d c_k T_k(x)$, with T_k the order- k Chebyshev polynomial of the first kind [11]. In the “merged” encoding a classical vector $\mathbf{v}$ of dimension $n_{\text{in}}(d+1)$ is amplitude-encoded into $q = \lceil \log_2 n_{\text{in}}(d+1) \rceil$ qubits,

giving the state $|\mathbf{v}\rangle = \sum_j v_j |j\rangle / \|\mathbf{v}\|$, and the model reads out the overlap with the uniform state $|u\rangle = 2^{-q/2} \sum_j |j\rangle$. Because

$$\langle u | \mathbf{v} \rangle \|\mathbf{v}\| 2^{q/2} = \sum_j v_j, \quad (1)$$

the “quantum” output is exactly a classical weighted sum (full derivation in SM [27], Sec. S2). The decisive structural fact is $q = O(\log n_{\text{in}} d)$.

Variational QKAN (VQKAN). The data $\mathbf{x} \in [-1, 1]^n$ is angle-encoded, then L reuploading layers are applied. We state the circuit in the same order as the source program [28]:

1. Encoding: for each qubit j , apply $R_y(\arccos x_j)$.
2. For each layer $t = 1, \dots, L$:
 - (a) measure the intermediate magnetizations $\langle Z_j \rangle$;
 - (b) for each qubit j , apply $R_y(\pi \phi_{t,j}^R(s_j))$, where $s_j = \frac{1}{2}(\langle Z_j \rangle + 1)$ and $\phi_{t,j}^R$ is a learnable B-spline edge function (the Kolmogorov–Arnold activation);
 - (c) recompute $\langle Z_j \rangle$; for every ordered pair $j < \ell$ apply the all-to-all entangler $\text{CR}_y(\pi \phi_{t,j\ell}^C(s_j))$.
3. Readout: a Pauli observable H (below).

The intermediate $\langle Z_j \rangle$ are read exactly from the state in simulation; on hardware they are estimated by sampling. The state is propagated coherently—the measurements only set classical control angles—so entanglement accumulates. We reimplement VQKAN generically in n (SM [27], Sec. S3) so the architecture can be scaled past the $n=4$ of the original code [28], whose own focus is the model under noise; our noise study (R3) connects to that setting while posing a different, simulability-oriented question.

2.2 Loss function, target, and training

For regression the prediction is $p_{\boldsymbol{\theta}}(\mathbf{x}) = \langle H \rangle / N_p$ with the 2-local readout

$$H = \sum_{k=1}^{n/2} Z_{2k-1} Z_{2k}, \quad N_p = \frac{n}{2}, \quad (2)$$

and parameters $\boldsymbol{\theta} = \{\phi^R, \phi^C\}$ (the spline control points). The training loss is the mean-squared error over a dataset $\{\mathbf{x}_m\}_{m=1}^M$,

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{M} \sum_{m=1}^M (p_{\boldsymbol{\theta}}(\mathbf{x}_m) - f(\mathbf{x}_m))^2, \quad (3)$$

with the smooth target

$$f(\mathbf{x}) = \tanh\left(3\left[\frac{2}{n} \sum_{k=1}^{n/2} \sin(x_{2k-1}^2 + x_{2k}^2) - \frac{1}{2}\right]\right). \quad (4)$$

We use the *mean* over pairs so that $\text{Var}(f)$ stays $O(1)$ as n grows; the sum-based variant in the original program [28] has $\text{Var}(f) \rightarrow 0$ (e.g. $0.106 \rightarrow 0.019$ for $n=6 \rightarrow 12$), which we found makes a constant predictor hard to beat and drives *apparent* dequantization through under-fitting; we therefore adopt Eq. (4) and report the coefficient of determination R^2 against the constant (mean) predictor. We minimize Eq. (3) with SPSSA [22] (gradient-free, two loss evaluations per step) using three random restarts and keeping the best (settings in SM [27], Sec. S6). The discrete-logarithm kernel model of Sec. 3 is instead trained by a convex SVM [23], which has no spurious local minima and no barren plateau.

2.3 Tensor-network simulability diagnostics

For a state $|\psi\rangle$ on n qubits we cut the chain in half and compute the bipartite (von Neumann) entanglement entropy $S(n/2)$ from the Schmidt spectrum (singular values σ_i of the reshaped amplitude matrix), $S = -\sum_i \sigma_i^2 \log_2 \sigma_i^2$; an area law has $S = O(1)$, a volume law $S \propto n$. The bond dimension that captures a state to fixed fidelity scales as $\chi \sim e^S$ (a flat Schmidt spectrum of rank χ has $S = \log_2 \chi$), so area law $\Rightarrow$ polynomial χ and volume law $\Rightarrow$ exponential χ ; for a single arbitrary D -dimensional amplitude vector the Schmidt rank across any cut is at most $\sqrt{D}$. The bond dimension needed by a truncated MPS to reproduce the readout within tolerance ϵ is

$$\chi_{\text{need}}(\epsilon) = \min \left\{ \chi : |\langle H \rangle_{\text{MPS}_\chi} - \langle H \rangle| < \epsilon \right\}, \quad (5)$$

with $\langle H \rangle_{\text{MPS}_\chi}$ evaluated after compressing $|\psi\rangle$ to bond dimension χ (algorithm in SM [27], Sec. S4). The *task-aware* choice sets $\epsilon = \sqrt{\mathcal{L}}$, the RMS error the trained model attains: a model that cannot beat accuracy ϵ does not require a bond dimension that only matters at finer tolerance. This criterion is the methodological core of the paper.

For noise we evolve the density matrix ρ under a per-gate, per-qubit depolarizing channel of strength p and compute the OSEE—the bond entropy of the MPO representation of ρ , obtained as the entanglement entropy of the vectorized ρ across the cut (SM [27], Sec. S5). Trainability is quantified by the standard BP diagnostic [13]: the variance, over random parameters θ , of a single partial derivative $\partial \langle H \rangle / \partial \theta_i$ (observation protocol in SM [27], Sec. S6). All diagnostics use a self-contained NumPy statevector/density-matrix engine [24, 25]; figures use Matplotlib [26].

2.4 Scope and validity of the numerical model

Three caveats bound our claims and we state them up front. (i) *Proxy model*. The trainability and trained- χ results use a generic- n reimplement of VQKAN that retains the all-to-all CR_y entangler—the feature

governing hardness—but simplifies the edge functions to depend on a single control qubit and uses $L=2$, $G=4$; the original code is fixed at $n=4$. Because the simplification *reduces* entangling power, it can only *under-estimate* the hardness of the full model, so our dequantization finding (R2) is conservative in direction; exact-model and multi-seed confirmation is left for future work. (ii) *What χ_{need} certifies*. We evaluate Eq. (5) by explicit statevector truncation, which is exponential in memory; a bounded χ_{need} at $n \leq 12$ therefore certifies *bounded bond dimension at the tested sizes*, not an asymptotic polynomial-time classical algorithm. (iii) *Average tolerance*. The task-aware $\epsilon = \sqrt{\mathcal{L}}$ is the dataset-averaged RMS error; it bounds the surrogate’s mean readout error, not the worst-case per-input error, and we also report the strict $\epsilon = 10^{-2}$ for contrast. We do not provide a classical KAN baseline matching the model’s accuracy, so “dequantizable” here means “reproducible by a low-bond-dimension surrogate for the tested task,” not a proven absence of advantage in general.

3 Results

Throughout, statements inferred from our measurements cite the relevant figure; statements inferred from prior work cite the reference. Unless noted, VQKAN runs use $L=2$ reuploading layers and B-splines with 4 control points (SM [27]).

(R1) The Chebyshev amplitude QKAN is exactly classical. By Eq. (1) the model output is a classical sum, and the merged encoding uses $q = O(\log n_{\text{in}} d)$ qubits, so its entire Hilbert space has dimension $2^q = O(n_{\text{in}} d)$. The point is not the state-preparation cost but that the function the model computes is the classical sum of Eq. (1): it is evaluated in time $O(n_{\text{in}} d)$ on a classical computer, with bond dimension $\chi \leq \sqrt{n_{\text{in}} d}$ (SM [27], Sec. S2), so the quantum circuit is unnecessary. The parallel encoding uses many qubits but in an unentangled product of single-edge registers. There is no entanglement resource and hence no regime of quantum advantage; this is consistent with the construction’s own stated scope [11].

(R2) Random VQKAN is volume-law; trained VQKAN dequantizes. At random parameters the all-to-all entangler drives the entanglement toward a volume law: the normalized half-chain entropy $S(n/2)/(n/2)$ rises from 0.50 at $n=4$ to 0.88 at $n=14$ [Fig. 1, red]—growing with system size, though not yet saturated to unity at the sizes reached—and even the 2-local observable then needs the full $\chi = 2^{n/2}$. This is the random-circuit expectation and not the model’s operating point. We train VQKAN on Eqs. (3)–(4). On the training set the models beat the constant predictor ($R^2 = 0.62 \pm 0.08, 0.36 \pm 0.12$,

0.30 ± 0.05 at $n = 6, 8, 10$; four seeds), but this is an *in-sample* control: with 24 training points and 168–440 parameters, held-out evaluation on 200 fresh inputs gives *negative* R^2 —the models memorize. Retraining on 200 points restores weak but genuine generalization (held-out $R^2 = +0.14/+0.27$ at $n=6$, $+0.12/+0.03$ at $n=8$; two seeds), with χ_{need} at the attained accuracy still small (1.7–4.5). The honest statement is therefore modest: this VQKAN is a *weak learner* on this target, and at the weak accuracy it attains it is reproducible at low bond dimension. (We stress this control: with the original sum-based target the models scored $R^2 < 0$ for $n \geq 8$, so their small χ_{need} was an artifact of under-fitting; that pitfall is why we adopt Eq. (4).) For these genuinely trained models the task-aware bond dimension [Eq. (5)] stays small relative to $\chi_{\text{max}} = 2^{n/2}$: the ratio is 0.29 ± 0.06 , 0.24 ± 0.03 , 0.14 ± 0.02 [Fig. 1, green], falling with n . We do not claim a precise scaling, only that χ_{need} does *not* track χ_{max} (no exponential growth). A fixed-tolerance control disentangles this trend from the criterion itself—the task-aware tolerance $\epsilon = \sqrt{\mathcal{L}}$ loosens as the fit degrades with n : re-evaluating the same trained models at n -independent tolerances gives $\chi_{\text{need}}/\chi_{\text{max}} = 0.40, 0.38, 0.35$ at $\epsilon=0.1$ but $0.90, 0.85, 0.81$ at the strict $\epsilon=10^{-2}$ (SM [27], Sec. S6). In *raw* terms the fixed-tolerance bond dimension grows near-exponentially, $\chi_{\text{need}} \approx 3.2 \rightarrow 6.1 \rightarrow 11.2$ at $\epsilon=0.1$ ($\sim 2^{0.45n}$, close to the $2^{0.5n}$ of χ_{max}); only the *ratio* falls. The decrease at task-aware tolerance is therefore largely criterion-driven (the tolerance loosens as the fit degrades), and the defensible statement is correspondingly precise: the model is reproduced by a low-bond-dimension surrogate *at the accuracy it actually attains*, while at any fixed accuracy the state hardens near-exponentially with n . We also note a normalization subtlety in Eq. (5): $\epsilon = \sqrt{\mathcal{L}}$ is defined on the normalized prediction scale but applied to the unnormalized readout $\langle H \rangle$, i.e. N_p -fold stricter than self-consistency requires; under the self-consistent tolerance the task-aware χ_{need} drops to 1–2, so the task-aware criterion at these weak accuracies carries little information beyond the fixed- ϵ results, which we regard as primary. The state stays volume-law, yet a model that demonstrably fits a smooth 2-local target is reproduced by a surrogate of bond dimension $\chi \sim 2$ –5. Scope: this concerns the studied 2-local readout and smooth target [Eq. (4)] at $n \leq 10$; targets with broad Fourier support or genuine n -body (parity-like) correlations are expected to need higher χ_{need} and are not covered. We report χ_{need} averaged over probe inputs; the per-probe maximum (SM [27], Sec. S4) shows no heavy tail here. Finally, we compare directly with the sample-count bound $\chi \leq M_t$ of Ref. [2] by sweeping the training-set size at $n=8$: $\chi_{\text{need}} = 8.0, 5.3, 3.2, 3.7$ at $M_t = 6, 12, 24, 48$ ($R^2 = 0.92$ – 0.34). The two quantities are not trivially ordered—at $M_t=6$ the state-

level χ_{need} *exceeds* M_t , because the bound constrains the function-space MPS of the hypothesis rather than the prepared state—while for $M_t \geq 12$ the task-aware value lies a factor 2–5 *below* $\min(M_t, \chi_{\text{max}})$. The diagnostic therefore carries information independent of, and in the practically relevant regime sharper than, the sample-count bound.

(R3) Noise restores classicality. Under depolarizing noise the OSEE of ρ collapses: at $n=7$ it falls from 4.5 at $p=0$ (4.92 ± 0.23 over five seeds) to 1.16 ± 0.13 by $p=0.05$ and to near 0 by $p=0.1$ [Fig. 2]. Two honest qualifications temper the interpretation. First, the purity collapses immediately— $\text{Tr} \rho^2 = 0.062$ already at $p=0.02$, within an order of magnitude of the fully mixed value 2^{-n} —so the classicalization at these sizes is driven by near-complete depolarization rather than by a subtle simulability transition, and the analogy to noisy-random-circuit sampling [8, 9] should be read as motivation, not correspondence. Second, because every gate carries noise, the total noise grows as $O(Ln^2)$ and the collapse necessarily drifts to smaller p with size (relative OSEE at $p=0.05$: 52%, 38%, 21%, 18%, 3.5% for $n = 5, 6, 7, 8, 10$); we therefore report a size-dependent crossover, not a threshold, and a gate-count-normalized analysis is future work. The behavior parallels the established classical-simulability crossover of noisy random circuits [8, 9]; we claim only the same qualitative mechanism, not identical exponents.

(R4) Forcing hardness triggers a barren plateau. (R1)–(R2) identify the two reasons VQKAN is easy: a low-weight readout and an effectively low-rank target. The natural fixes are a global readout and a hard feature map. Figure 3 sweeps four configurations. Only the shipped angle+local model keeps a flat gradient variance (~ 0.1 , trainable)—and it is the easy one. A global parity readout $H_{\text{glob}} = \prod_j Z_j$ leaves the state (and hence its entanglement) unchanged but still collapses the gradients; an IQP feature map [16] raises the entanglement further [Fig. 3, right]; both drive $\text{Var}_\theta[\partial\langle H \rangle/\partial\theta]$ down geometrically [Fig. 3, left]; a coarse log-linear fit gives $\text{Var} \sim 2^{-0.7n}$ for the global readout, the signature of a barren plateau. The diagnostic perturbs one deterministically chosen active parameter (a last-layer entangling control point; SM [27], Sec. S6), so the slope is indicative rather than a precise exponent. An IQP encoding followed by a *shallow local* trainable layer—a variational stand-in for a kernel method, not the genuine kernel support-vector machine studied in (R5)—also concentrates [Fig. 3]; we attribute this to the highly entangling IQP map producing states whose local observables concentrate, consistent with—though we do not independently measure the Haar-ness via the frame potential—the analysis of Ref. [14]. Our

measurements thus reproduce, for this model family, the analytic link between absence of barren plateaus and classical simulability [15] and the exponential concentration of expressive quantum kernels [14].

(R5) A discrete-logarithm feature map threads the needle. The escape is to place hardness in structure, not in generic entanglement. We implement the Liu–Arunachalam–Temme discrete-logarithm feature map [17]. For a prime p , a generator g of $\mathbb{Z}_p^*$, and a power-of-two window $M = 2^s$ (our concentration experiment uses $M \approx (p-1)/4$; the learning experiments use $M = (p-1)/2$ rounded down—SM [27]), the feature state on $n = \lceil \log_2 p \rceil$ qubits is

$$|\phi(x)\rangle = \frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} |xg^i \bmod p\rangle, \quad (6)$$

the output of a modular-exponentiation circuit $|i\rangle|0\rangle \mapsto |i\rangle|xg^i \bmod p\rangle$ with the index register discarded (SM [27], Sec. S7). Writing $S_x = \{xg^i \bmod p\}$ and using orthonormality of computational basis states, the kernel is

$$k(x, x') = |\langle \phi(x) | \phi(x') \rangle|^2 = \left(\frac{|S_x \cap S_{x'}|}{M} \right)^2. \quad (7)$$

Since $\log_g(xg^i) = \log_g x + i$, the set S_x is an interval of length M in the discrete-logarithm coordinate, so $|S_x \cap S_{x'}| = \max(0, M - d)$ for the circular distance d between $\log_g x$ and $\log_g x'$: a smooth triangle kernel (full derivation in SM [27], Sec. S7). The non-concentration relies on the window choice $M \sim (p-1)/2$; a very small M/p would shrink typical overlaps and could itself concentrate, so the favorable behavior is a property of this regime, not of the encoding for all M . Evaluating it classically requires $\log_g$, the discrete-logarithm problem, believed hard [17]. Measured off-diagonal kernel variance rises from 0.046(4) at $n=4$ and saturates at ≈ 0.072 for $n \geq 8$ (five seeds), in agreement with the closed form $\text{Var}[k] = \frac{1}{10} - \frac{1}{36} \approx 0.0722$ of the triangle kernel in the discrete-log coordinate—an *analytic* structural guarantee of non-concentration (the flatness carries no quantum content by itself; the quantum content is that *evaluating* the kernel classically requires $\log_g$), whereas the generic IQP kernel concentrates: its variance decays as 4^{-n} (from $4.1(4) \times 10^{-3}$ at $n=4$ to $7.0(3) \times 10^{-8}$ at $n=12$, five seeds) and its mean entry as 2^{-n} —the latter setting the $\sim 2^n$ -shot read-out cost [Fig. 4]. Non-concentration means the kernel is $O(1/\epsilon^2)$ -shot estimable and trainable by a convex SVM with no barren plateau, while remaining classically hard—the one route *in our study* that reconciles hardness with trainability. It is not unique in the literature: symmetry-structured (covariant) kernels have been shown analytically to evade exponential concentration [21], and our discrete-logarithm window is best

read as a numerically explicit member of that broader structured, non-concentrating class. This result is at the kernel level; result (R7) below embeds the same map into the Kolmogorov–Arnold architecture itself.

(R6) The discrete-logarithm kernel is a demonstrated learner. Non-concentration is necessary but not sufficient for advantage, so we test learning directly. We define the Liu–Arunachalam–Temme concept class—a half-space $y(x) = \text{sgn}[(p-1)/2 - \log_g x]$ in the discrete-logarithm coordinate, classically hard to evaluate—and train a support-vector machine with each of three precomputed kernels (SM [27], Sec. S8). The discrete-logarithm kernel reaches held-out test accuracy ≥ 0.99 at $n = 6, 8, 10, 12$, essentially n -independent, whereas the generic IQP kernel (0.53–0.62) and a classical radial-basis-function baseline on the *raw* inputs (0.44–0.57) remain at chance [Fig. 5]. An important fairness caveat scopes this result: at these sizes the discrete logarithm is classically trivial, and a classical RBF *given the discrete-log coordinate* matches the quantum kernel (0.97–1.00 in our check). The separation shown is therefore a numerical illustration of the Liu–Arunachalam–Temme construction—hardness enters only asymptotically, through the classical cost of computing $\log_g$ —not an empirical demonstration of hardness at the tested sizes. The structured kernel thus both avoids concentration and generalizes on a classically hard concept, while the generic quantum and classical baselines do neither—turning the variance signature of (R5) into an explicit learnability separation. This separation survives finite measurement statistics: under a hardware-shot model that replaces each kernel entry by a binomial estimate from S shots (followed by the standard symmetrize-and-project-to-PSD post-processing of measured Gram matrices), the discrete-logarithm SVM already reaches 0.99 test accuracy at $S=30$ ($n=8$; 1.00 at $n=10$; SM [27], Sec. S8), confirming the kernel is shot-efficient. The IQP kernel, whose entries are $\sim 2^{-n}$, would instead need $\sim 2^n$ shots merely to resolve them from zero and stays at chance.

(R7) A discrete-logarithm QKAN. Finally, we wire the hard encoding into the Kolmogorov–Arnold architecture itself. The model computes quantum features $F_{d,j} = k(x_d, c_{d,j})$ —orbit-window overlaps of coordinate x_d with $J=4$ fixed random centers at three window sizes $M = (p-1)/2, (p-1)/8, (p-1)/32$ (each entry a modular-exponentiation-orbit overlap, quantum-evaluable and $O(1)$ -valued, hence shot-efficient)—and feeds them to an additive KAN layer $f(\mathbf{x}) = \sum_i \phi_i(F_i)$ whose learnable piecewise-linear edge functions are fit by convex ridge least squares: trainable by construction, with no barren plateau. On a smooth target defined on the discrete-logarithm coordinates (pseudorandom in the raw inputs; SM [27],

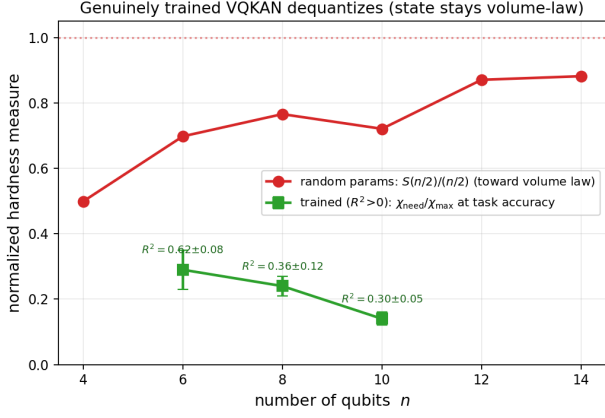


Figure 1: (fig_dequant.png) Random VQKAN is volume-law (red: normalized half-chain entropy $S(n/2)/(n/2) \rightarrow 1$), but the trained model dequantizes (green: the task-aware bond dimension $\chi_{\text{need}}/\chi_{\text{max}}$, $\chi_{\text{max}} = 2^{n/2}$, falls with the qubit number n). Trained points are genuine fits (R^2 against the constant predictor annotated); error bars and R^2 are mean $\pm$ std over four seeds. Data: result (R2).

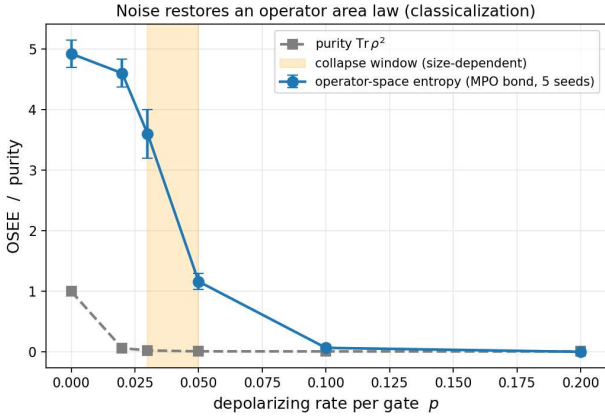


Figure 2: (fig_noise.png) Depolarizing noise of rate p restores an operator-space area law: the MPO bond entropy (OSEE) and purity $\text{Tr} \rho^2$ collapse across $p^* \approx 0.03\text{--}0.05$ ($n = 7$). Data: result (R3).

Sec. S8), this DL-KAN reaches held-out test $R^2 = 1.00, 1.00, 0.99$ over three seeds at $n=8$, while the *identical* architecture on IQP features or on classical (cos, sin) features fails (test $R^2 \leq 0$). A necessity control tempers the claim: plain linear ridge regression on the same discrete-logarithm features already reaches test $R^2 = 0.96$, so the separation is carried by the *feature map*, with the KAN layer contributing only the final nonlinear residual. (R7) therefore demonstrates *compatibility* of the hard encoding with a Kolmogorov–Arnold readout—hardness in the encoding, learnability in a convex readout—not necessity of the KAN structure itself.

4 Discussion

We have measured, rather than assumed, the tensor-network simulability of concrete quantum Kolmogorov–Arnold networks, using a task-aware bond-dimension criterion [Eq. (5)] that asks how much entanglement a model’s *output* actually requires at the accuracy the model reaches (a criterion in the spirit of accuracy-matched dequantization [20]). Our coverage is two published constructions—the amplitude-Chebyshev and reuploading-variational families; other proposals (the QSVT/block-encoding QKAN of Ref. [12], enhanced and adaptive VQKAN variants) remain to be analyzed. Our results must also be reconciled with two adjacent findings. Optimized QAOA circuits were shown to end in $\chi=2$ matrix-product states while an entanglement barrier persists at intermediate times [19]—consistent with our output-level statement, and a reminder that low output- χ does not by itself certify an efficient classical *simulation path*. Conversely, Ref. [18] found that trained variational classifiers with high test accuracy are underpinned by *higher* entanglement, with truncation sharply degrading accuracy—the opposite regime to ours. We conjecture the difference lies in readout weight, target smoothness, and attained accuracy (our 2-local readout, smooth target, and weak learner versus their global-feature, high-accuracy classifiers); mapping the boundary between these regimes with our diagnostic is a concrete next step.

The findings form a chain. The Chebyshev amplitude QKAN is exactly classical (R1)—logarithmically many qubits, no entanglement resource. The variational QKAN is volume-law at random parameters but dequantizes once trained (R2): χ_{need} stays small while χ_{max} grows exponentially, in contrast to the naive expectation that an all-to-all entangler confers hardness—a concrete numerical instance, for a published QKAN, of the general bound that an M_t -sample-trained model lies in an MPS space of bond dimension $\chi \leq M_t$ [2]. And every generic attempt to restore hardness—global readout, deeper entangler, IQP encoding (R4)—collides with an exponentially concentrated gradient or kernel, reproducing for this family the principle that trainable variational models tend to be classically simulable [15, 14]. It is useful to separate the two axes explicitly: of the four cells of the trainable $\times$ hard table, our configurations populate only “trainable and easy” (angle+local) and “hard and untrainable” (global readout, IQP); the desired “trainable and hard” cell is empty for every generic route and is reached only by the structured kernel of (R5). Noise only sharpens the conclusion, classicalizing the model above $p^* \approx 0.03\text{--}0.05$ (R3).

Placed against the broader program of quantum advantage in machine learning, these results delineate where advantage can live: not in generic expressivity or entanglement, but in problem structure

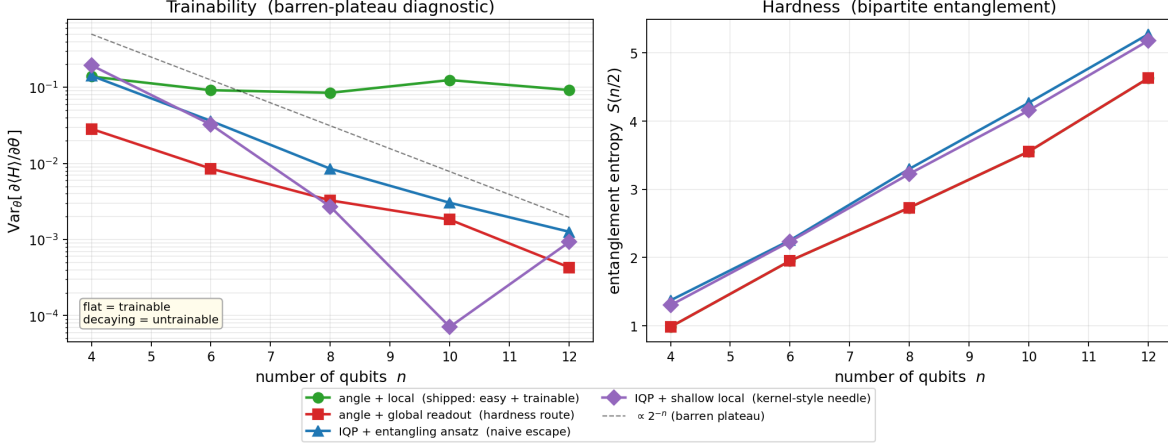


Figure 3: (fig.tradeoff.png) Hardness $\Leftrightarrow$ trainability tradeoff. Left: gradient variance $\text{Var}_\theta[\partial\langle H\rangle/\partial\theta]$ (barren-plateau diagnostic) versus qubit number n ; only angle+local stays flat. Right: bipartite entanglement $S(n/2)$. Every route that raises hardness drives the gradient variance down exponentially. Each point: variance over 300 random parameter samples at fixed input. Data: result (R4).

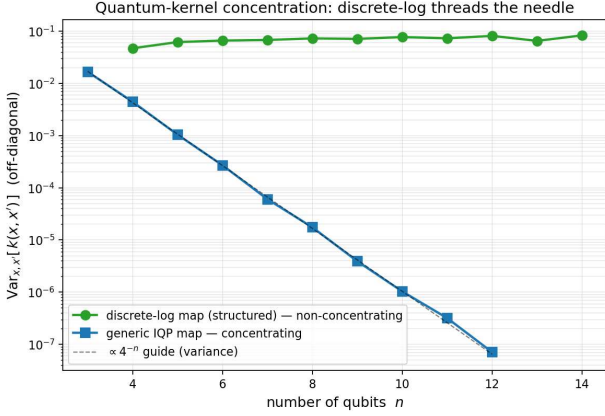


Figure 4: (fig_kernel.png) Quantum-kernel concentration. The discrete-logarithm feature map (green) has n -independent off-diagonal kernel variance $\text{Var}_{x,x'}[k(x,x')]$, whereas the generic IQP map (blue) concentrates (variance $\sim 4^{-n}$, mean entry $\sim 2^{-n}$). Each point: 400 random off-diagonal pairs; the discrete-logarithm value is 0.07 ± 0.01 over five seeds. Data: result (R5).

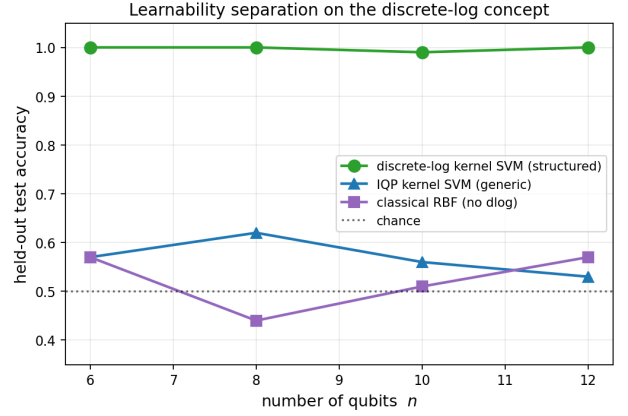


Figure 5: (fig_learning.png) Learnability separation on the discrete-log concept. A support-vector machine with the discrete-logarithm kernel (green) reaches ≥ 0.99 held-out test accuracy, n -independent, while the generic IQP kernel (blue) and a classical radial-basis-function baseline (purple) stay at chance (dotted). Train/test sizes $\leq (p-1)/2$ each. Data: result (R6).

with a proven classical-hardness core. The discrete-logarithm feature map (R5) is the concrete witness—non-concentrating, hence shot-efficient and trainable, yet classically hard [17]. We offer this as a numerical verification and design exemplar built on the established speed-up of Ref. [17], not as new hardness theory. The two sources of tensor-network hardness can moreover be separated by ablating the ansatz versus the encoding (SM [27], Sec. S9): the all-to-all entangler and the IQP encoding each reach near-maximal entanglement on their own, so the full-state entropy cannot tell them apart, but the discrete-logarithm feature states are only *moderately* entangled ($S \approx 2.5\text{--}3.0$ versus the near-maximal $n/2$ of the

other two). Their intractability is thus not high bond dimension but the classical cost of *constructing* the kernel from the input—a distinct, low-entanglement, construction-hard category, which is exactly why it escapes the barren plateau that entanglement-derived hardness incurs. The practical message is both a warning and a design rule: within the scope we tested, a QKAN that is comfortably trainable on near-term hardware is, by that very fact, likely classically reproducible, and our results suggest that demonstrations of advantage are better sought in structured, provably hard encodings than in deep entangling ansätze. This design rule is an inference from one negative family and one positive example, not a theorem; we do not

claim it holds for every QKAN task.

Limitations. Our study is numerical and small-scale ($n \leq 14$), with no asymptotic proof. The trained- χ result (R2) controls for fit quality (four seeds, constant-predictor R^2) but uses a simplified proxy ansatz (Sec. 2.4) and $n \leq 10$ (with restarts); at $n=12$ only one of three seeds passes the fit-quality control at our optimizer budget (its ratio, 0.09, is consistent with the trend; SM [27]), so iso-accuracy training at larger n remains to be done. The per-figure system sizes are heterogeneous, set by the exponential cost of exact statevector and density-matrix simulation rather than by design. The learnability results (R6, R7) use one concept class / one smooth target family and integer-bit input encodings; we include a hardware-shot (finite- S) model, but broader datasets, deeper DL-KANs (beyond one additive layer), and a full device-noise model are natural next steps. Two concrete directions follow: extending the DL-KAN of (R7) to multi-layer compositions while preserving non-concentration, and a finite-size-scaling analysis of $p^*(n)$ against the noisy-random-circuit transition [8].

Code and data availability. All diagnostics, the generic- n VQKAN reimplementation, the training, tradeoff, and discrete-logarithm experiments, and the figure generators are available from the author; their locations are listed in the provenance file [28]. Full reproduction instructions, per-figure commands, parameters, and random seeds are tabulated in the SM [27] (Sec. S9).

References

- [1] F. J. Schreiber, J. Eisert, and J. J. Meyer, *Classical surrogates for quantum learning models*, Phys. Rev. Lett. **131**, 100803 (2023).
- [2] S. Shin, Y. S. Teo, and H. Jeong, *Dequantizing quantum machine learning models using tensor networks*, Phys. Rev. Research **6**, 023218 (2024).
- [3] E. Gil-Fuster, C. Gyurik, A. Pérez-Salinas, and V. Dunjko, *On the relation between trainability and dequantization of variational quantum learning models*, ICLR (2025); arXiv:2406.07072.
- [4] M. Cerezo *et al.*, *Variational quantum algorithms*, Nat. Rev. Phys. **3**, 625 (2021).
- [5] R. Orús, *A practical introduction to tensor networks*, Ann. Phys. **349**, 117 (2014).
- [6] G. Vidal, *Efficient classical simulation of slightly entangled quantum computations*, Phys. Rev. Lett. **91**, 147902 (2003).
- [7] F. Pan, K. Chen, and P. Zhang, *Solving the sampling problem of the Sycamore quantum circuits*, Phys. Rev. Lett. **129**, 090502 (2022).
- [8] D. Aharonov *et al.*, *A polynomial-time classical algorithm for noisy random circuit sampling*, in Proc. 55th ACM STOC (2023), p. 945.
- [9] K. Noh, L. Jiang, and B. Fefferman, *Efficient classical simulation of noisy random quantum circuits in one dimension*, Quantum **4**, 318 (2020).
- [10] Z. Liu *et al.*, *KAN: Kolmogorov–Arnold networks*, arXiv:2404.19756 (2024).
- [11] H. Wakaura, *qkan: Quantum Kolmogorov–Arnold networks—a unified framework, library, and benchmark* (2025).
- [12] P. Ivashkov *et al.*, *QKAN: Quantum Kolmogorov–Arnold networks*, arXiv:2410.04435 (2024).
- [13] J. R. McClean *et al.*, *Barren plateaus in quantum neural network training landscapes*, Nat. Commun. **9**, 4812 (2018).
- [14] S. Thanasilp, S. Wang, M. Cerezo, and Z. Holmes, *Exponential concentration in quantum kernel methods*, Nat. Commun. **15**, 5200 (2024).
- [15] M. Cerezo *et al.*, *Does provable absence of barren plateaus imply classical simulability?*, arXiv:2312.09121 (2023).
- [16] V. Havlíček *et al.*, *Supervised learning with quantum-enhanced feature spaces*, Nature **567**, 209 (2019).
- [17] Y. Liu, S. Arunachalam, and K. Temme, *A rigorous and robust quantum speed-up in supervised machine learning*, Nat. Phys. **17**, 1013 (2021).
- [18] A. C. Nakhl, T. Quella, and M. Usman, *Calibrating the role of entanglement in variational quantum circuits*, Phys. Rev. A **109**, 032413 (2024).
- [19] M. Dupont, N. Didier, M. J. Hodson, J. E. Moore, and M. J. Reagor, *Entanglement perspective on the quantum approximate optimization algorithm*, Phys. Rev. A **106**, 022423 (2022).
- [20] R. Sweke, E. Recio *et al.*, *Potential and limitations of random Fourier features for dequantizing quantum machine learning*, Quantum **9**, 1640 (2025).
- [21] L. Henderson *et al.*, *Quantum advantage without exponential concentration: trainable kernels for symmetry-structured data*, arXiv:2509.14337 (2025).
- [22] J. C. Spall, *Multivariate stochastic approximation using a simultaneous perturbation gradient approximation*, IEEE Trans. Autom. Control **37**, 332 (1992).
- [23] C. Cortes and V. Vapnik, *Support-vector networks*, Mach. Learn. **20**, 273 (1995).

- [24] C. R. Harris *et al.*, *Array programming with NumPy*, *Nature* **585**, 357 (2020).
- [25] P. Virtanen *et al.*, *SciPy 1.0: fundamental algorithms for scientific computing in Python*, *Nat. Methods* **17**, 261 (2020).
- [26] J. D. Hunter, *Matplotlib: a 2D graphics environment*, *Comput. Sci. Eng.* **9**, 90 (2007).
- [27] See Supplemental Material (file `SI.QKAN-trained-dequantization-and-discrete-log-escape.tnqa-session_2026-06-22`) for the engine, derivations, algorithms, protocols, and reproduction settings.
- [28] Provenance file `PROVENANCE_data-and-source-paths_2026-06-22.md` (parent data and source paths), kept separate from the manuscripts.