COMPOSITIONAL SPECTRAL REPRESENTATIONS FOR EFFICIENT QUANTUM NEURAL APPROXIMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Universal approximation is now established for several quantum neural-network families, yet it does not determine the circuit cost of representing a structured target. Constructive spectral models compile global Fourier or Chebyshev expansions; products and compositions of simple local factors can make these expansions exponentially long while the generating computation remains small. We introduce stable quantum spectral formula networks (QSFN), which compile the bounded generating formula directly, making circuit resources follow formula complexity even when its global spectrum is exponentially long. Univariate spectral leaves expose signed Pauli-Z expectations, and internal vertices combine them through bounded maps that are affine in each input separately. We prove that these maps are exactly the transformations compatible with the interface, establish universal and quantitative analytic approximation, and give a deterministic circuit compiler with explicit query, gate, depth, normalization, and stability guarantees. For an explicit continuous bounded family, every fixed-error Fourier-sum approximation requires $\Omega(2^d)$ branches. QSFN gives an exact, locally query-optimal $\Theta(d^2)$ construction with unit normalization. Across three experiments, matched-error comparisons reduce entangling depth by up to 10.5-fold, exact semantic controls isolate factorization as the source of compression, and fixed-error tests reproduce the predicted polynomial–versus– exponential scaling. The results establish compositional structure as an independent determinant of quantum approximation cost.

1 INTRODUCTION

Universal approximation is a foundational test of neural-network expressivity: it identifies function classes that can be approximated arbitrarily well (Hornik, 1991). Modern approximation theory goes further by relating error to smoothness, dimension, and model structure (Barron, 1993; Yarotsky, 2017). For quantum learning, these questions also concern the circuit resources needed to achieve a prescribed accuracy (Yu et al., 2024). A constructive spectral route selects a global expansion and compiles each retained term as a circuit branch (Tang et al., 2026). Yet many high-dimensional targets have compact descriptions as products of local factors, low-rank sums, or hierarchical interactions. Multiplication combines local spectral supports, so a short formula can generate a long global spectrum. For example, a product of d univariate factors, each with three Fourier frequencies, has a formula of linear size and an expanded spectrum of 3^d frequencies. An architecture that implements each global term incurs the expanded representation cost even when its coefficients are supplied exactly. This raises a structural question: how can quantum neural networks exploit a compact generating formula while retaining universal approximation?

Our approach expands local functions spectrally while retaining their products, weighted sums, and hierarchical combinations. A stable quantum spectral formula network (QSFN) realizes the resulting formula as a quantum computation tree. Each local circuit represents a bounded function through the signed Pauli-Z expectation of an output qubit. Composition circuits combine independently prepared child outputs, and the root represents the complete formula. We characterize all bounded combinations determined solely by these expectations: they are precisely the maps that are affine in each input separately. This characterization includes multiplication and bounded signed sums and yields a deterministic compiler with explicit resource costs and control of error propagation.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

This construction yields universal approximation for continuous functions on compact domains and quantitative bounds for low-separation-rank analytic functions, connecting local approximation accuracy to formula size and normalization. We further analyze continuous bounded targets for which exact QSFN evaluation uses polynomially many local signal queries, while a global Fourier sum requires exponentially many terms even at fixed error. An analogous separation holds for the native spectral-modulus readout. These results establish formula complexity as a resource measure that captures advantages hidden by global spectral length.

1.1 Related Work

Low-rank tensors and hierarchical networks show how compact compositions can generate large expanded representations (Grasedyck et al., 2013; Mhaskar & Poggio, 2016; Cohen et al., 2016). This motivates *formula complexity*—the number and depth of local factors and composition nodes—as a resource measure alongside spectral length.

Fourier analyses relate quantum function expressivity to the frequencies accessible through data encoding (Schuld et al., 2021), and repeated encoding gives universal approximation even with a single qubit (Pérez-Salinas et al., 2021). Constructive multivariate circuits provide non-asymptotic error and resource bounds (Yu et al., 2024). Recent work establishes universality with qubit counts logarithmic in input dimension (Pérez-Salinas et al., 2025), while SAQNN builds Fourier- or Chebyshev-based approximants with resource bounds tied to the selected terms (Tang et al., 2026). These results link expressive power to circuit resources and highlight the role of a target's representation.

QSP and QSVT implement bounded polynomial transformations with degree-controlled query cost (Low & Chuang, 2017; Gilyén et al., 2019). QKAN composes Chebyshev-based block-encoded layers, modular QSP composes multivariable operator transformations, and EHands performs arithmetic on expectation-encoded values (Ivashkov et al., 2026; Rossi et al., 2025; Balewski et al., 2025).

1.2 Contributions

Our main contributions are:

- We introduce QSFN as a quantum computation tree of local spectral functions and bounded compositions. We characterize the bounded rules supported by independent child output expectations (Theorem 1), derive exact resource recurrences from a deterministic compiler (Theorem 2), and bound each local error's effect on the output (Proposition 4).
- We prove uniform universal approximation and construct approximating formulas of size $O(Rd\log(MAd/\epsilon)/\log\rho)$, with normalization at most eA, for rank-R separated analytic targets (Theorem 3).
- We construct continuous bounded targets whose fixed-error Fourier-sum complexity is $\Omega(2^d)$. Their exact QSFN programs use $\Theta(d^2)$ local signal queries, attain the matching query lower bound, have logarithmic aggregation depth, and retain unit normalization (Theorem 5). Theorem 6 establishes a separate exponential lower bound for the native modulus interface.
- Three experiments link theory to implementation: matched-error comparisons show circuit savings, exact controls isolate factorization, and fixed-error tests support the separation.

2 Structured Approximation Tasks

Approximation task. Let $\mathcal{X} \subset \mathbb{R}^d$ be compact and $f : \mathcal{X} \to \mathbb{R}$ bounded. We seek a finite formula F satisfying $\|f - F\|_{\mathcal{E}} \le \epsilon$ and a quantum circuit that evaluates F. Here $\mathcal{E} = L^\infty$ for uniform approximation and $\mathcal{E} = L^2(\mu)$ for the spectral separations, where μ is a probability measure. The periodic comparisons use the normalized uniform measure on $[-\pi, \pi)^d$.

Structured target functions. We consider bounded continuous functions built from local responses through products, weighted sums, and bounded hierarchical interactions. A representative class

consists of sums of separable products:

$$f(\boldsymbol{x}) = \sum_{r=1}^{R} a_r \prod_{j=1}^{d} g_{rj}(x_j), \qquad |g_{rj}(x_j)| \le 1. \tag{1}$$

Here d is the input dimension, R is the number of products, g_{rj} is the response to coordinate x_j in product r, and a_r weights that product. Local bounds are absorbed into a_r. Examples include products $\prod_j(\alpha_j + \beta_j \cos x_j)$, separable polynomial response surfaces, and low-rank sums of these functions. The same operations also build hierarchies of bounded pairwise interactions. Their compact descriptions can generate many spectral terms when the products are expanded.

Every continuous target on a compact domain admits polynomial approximants of the form equation 1, since monomials are products of one-variable powers. Theorem 3 gives universality and explicit resource bounds for analytic factors.

From local spectra to a preserved formula. For a target of the form equation 1, approximate each g_{rj} in a univariate basis $\{\psi_k\}$ and retain the target's products and weighted sum:

$$g_{rj}(x_j) \approx q_{rj}(x_j) = \sum_{k \in S_{rj}} c_{rjk}\psi_k(x_j), \qquad f(\boldsymbol{x}) \approx F(\boldsymbol{x}) = \sum_{r=1}^{R} a_r \prod_{j=1}^{d} q_{rj}(x_j). \tag{2}$$

Here S_{rj} contains the retained local indices and c_{rjk} are their coefficients. The basis is chosen for the input domain, such as Chebyshev polynomials for normalized coordinates or Fourier functions for phases; each q_{rj} is real-valued. Finite spectral factors can be retained exactly. For general factors, Theorem 3 bounds how their local errors accumulate in F.

Equation 2 specifies a computation tree: its leaves evaluate q_{rj}, product nodes combine factors within each component, and signed-sum nodes combine the products with weights a_r. Hierarchical targets extend this organization with bounded interactions between child values; Figure 1 illustrates one such tree and its circuit. Section 3 realizes these same operations as quantum circuits.

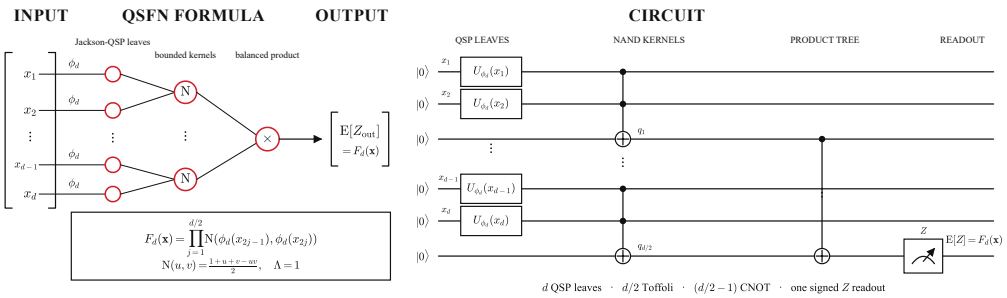


Figure 1: A QSFN formula and circuit, illustrated by the bounded family F_d of Section 3.4. Local ϕ_d leaves feed pairwise N rules and a balanced product tree; the circuit realizes the same composition and reads the root's signed Z expectation.

Expanding Equation 2 selects one index from each local support and yields $\sum_{r=1}^{R}\prod_{j=1}^{d}|S_{rj}|$ written terms before identical basis functions are combined; QSFN retains the local-factor products. Appendix C.6 gives the full expansion and a two-dimensional example in which two three-frequency factors yield nine complex Fourier frequencies.

Domain knowledge or a prescribed dictionary supplies the tree topology, and local and aggregation coefficients are fitted classically before compilation for repeated quantum evaluation. Appendix G gives the fixed-structure learning procedure and capacity bound.

3 Stable Quantum Spectral Formula Networks

A stable quantum spectral formula network (QSFN) realizes the preserved formula as a quantum computation tree. At node v, the subtree computes F_v, represented by the Pauli-Z expectation of

output qubit o_v: $z_v = \langle Z_{o_v} \rangle = F_v/\Lambda_v \in [-1, 1]$, with known positive scale Λ_v. Parent nodes compose these represented values, and the root computes the complete formula F.

For Equation 2, this correspondence can be written at every level. Choose $\Lambda_{rj} > 0$ bounding $\|q_{rj}\|_\infty$. The local-factor outputs and their products satisfy

$$z_{rj} = \frac{q_{rj}(x_j)}{\Lambda_{rj}} = \langle Z_{o_{rj}} \rangle, \quad \Lambda_r = \prod_{j=1}^{d} \Lambda_{rj}, \quad z_r = \prod_{j=1}^{d} z_{rj}, \quad \Lambda_r z_r = \prod_{j=1}^{d} q_{rj}(x_j). \tag{3}$$

Here o_{rj} is the output qubit for q_{rj}, and z_r represents the product of the normalized factors in component r. Multiplying by Λ_r recovers the unnormalized product. Combining the R components with the original weights a_r gives the root output:

$$\Lambda_{\text{root}} = \sum_{r=1}^{R} |a_r| \Lambda_r, \qquad z_{\text{root}} = \sum_{r=1}^{R} \frac{a_r \Lambda_r}{\Lambda_{\text{root}}} z_r, \qquad \Lambda_{\text{root}} z_{\text{root}} = F(\boldsymbol{x}). \tag{4}$$

The normalized weights have absolute sum one, so $|z_{\text{root}}| \le 1$. When every $a_r = 0$, a zero-output circuit suffices. Products and sums with many inputs are arranged as binary trees. These equations describe the expectations represented by the circuit; child outputs remain quantum states throughout coherent composition, with measurement at the root.

The same correspondence extends to a general QSFN tree. A leaf ℓ evaluates a bounded spectral function q_ℓ of coordinate $x_{j(\ell)}$. Each internal node v combines a constant number of children, denoted $\mathrm{ch}(v)$, through a rule G_v whose input and output values lie in $[-1, 1]$. Each non-root node has one parent. The functions computed by these subtrees therefore obey

$$F_\ell = q_\ell(x_{j(\ell)}), \qquad F_v = \Lambda_v G_v\big((F_u/\Lambda_u)_{u \in \mathrm{ch}(v)}\big), \qquad F = F_{\text{root}}. \tag{5}$$

Multiplication and signed sums recover Equations 3–4; the next subsection characterizes the bounded rules supported by these outputs. At the root, the compiled circuit $U_F(\boldsymbol{x})$ gives

$$f(\boldsymbol{x}) \approx F(\boldsymbol{x}) = \Lambda_{\text{root}} \langle Z_{o_{\text{root}}} \rangle. \tag{6}$$

Here $f \approx F$ is the target approximation, and the circuit realizes the selected F exactly. Leaf degrees set signal-query cost, the tree sets aggregation count and depth, and the scales govern measurement cost and error propagation.

3.1 Bounded composition of function values

The products and weighted sums above combine the values represented by child outputs. More generally, a parent rule must depend on these values alone, independently of unobserved details of their quantum states. At node v, a preparation circuit U_v acting on its register represents F_v through output qubit o_v when

$$\mathrm{tr}\big[Z_{o_v} U_v(\boldsymbol{x}) |0\rangle\langle 0| U_v(\boldsymbol{x})^\dagger\big] = z_v(\boldsymbol{x}) = F_v(\boldsymbol{x})/\Lambda_v, \qquad |z_v| \le 1. \tag{7}$$

We call the pair (z_v, Λ_v) its signed-expectation representation. A parent rule is *Z-compatible* if its output depends only on the child Z expectations, so changes to unobserved X and Y coherences leave the represented result unchanged. Sibling subtrees are prepared independently on disjoint registers.

For r such output qubits, the computational-basis probability of bit b_i is $[1 + (-1)^{b_i} z_i]/2$. Assign each bit string $\boldsymbol{b}$ a value $\tau(\boldsymbol{b}) \in [-1, 1]$ and form the diagonal observable $D_\tau = \sum_{\boldsymbol{b}} \tau(\boldsymbol{b}) |\boldsymbol{b}\rangle\langle\boldsymbol{b}|$. Independence makes its expectation

$$G_\tau(\boldsymbol{z}) = \sum_{\boldsymbol{b} \in \{0,1\}^r} \tau(\boldsymbol{b}) \prod_{i=1}^{r} \frac{1 + (-1)^{b_i} z_i}{2}. \tag{8}$$

Each factor is affine in one z_i, so the result is affine in every input separately. Such a function is called *multiaffine*; its corner values determine it throughout the hypercube. This property gives the complete class of bounded scalar combinations available from these child outputs.

Theorem 1 (Exact bounded-composition characterization)**.** *A function $G : [-1,1]^r \to [-1,1]$ is exactly realizable from independent signed-Z outputs through a Z-compatible diagonal rule if and only if it is bounded and multiaffine:*

$$G(\boldsymbol{z}) = \sum_{S \subseteq [r]} \gamma_S \prod_{i \in S} z_i. \tag{9}$$

Equation 8 proves necessity. For sufficiency, set $\tau(\boldsymbol{b}) = G((-1)^{b_1}, \ldots, (-1)^{b_r})$ and use uniqueness of multiaffine interpolation on the hypercube. Each rule admits an output-qubit realization, constructed in Appendix C.1 together with its precision costs. The compiler uses constant-arity rules with specified local realization costs.

Products, normalized signed sums, and the bounded NAND function $\mathsf{N}(u,v) = (1+u+v-uv)/2$ all belong to this class. Every unit-scale rule also satisfies $|\partial G/\partial z_i| \le 1$: changing one child value changes the parent value by at most the same amount. This local property will control errors along the entire tree.

3.2 Structure-preserving circuit compilation

We now realize the leaves and internal nodes of Equation 5, then combine them recursively.

Local spectral leaves. For a normalized coordinate $t \in [-1,1]$, the quantum signal uses an angle θ satisfying $t = \cos\theta$. Given a real degree-p polynomial q bounded by $B > 0$ on $[-1,1]$, quantum signal processing (QSP) realizes $\langle Z_o \rangle = q(t)/B = q(\cos\theta)/B$ using $O(p)$ local signal calls and $O(1)$ work qubits, with leaf scale $\Lambda_\ell = B$. For phase inputs x, the identity $\cos(kx) = T_k(\cos x)$, where T_k is the kth Chebyshev polynomial, connects cosine spectral factors to this polynomial form. This covers the $\cos x$ factor in Appendix C.6 and the local functions used in Section 3.4. Generalized QSP also supports mixed-parity polynomials and bounded Fourier transformations with positive and negative frequencies (Motlagh & Wiebe, 2024). A constant-size Hadamard-test readout converts the real matrix element into a signed Z expectation, giving the local factors in Equation 2 with query cost linear in their degree.

Composition circuits. For independent child states, one CNOT realizes multiplication:

$$\langle Z_b \rangle_{\mathrm{out}} = z_a z_b. \tag{10}$$

Its represented value is $F_a F_b$, with scale $\Lambda_v = \Lambda_a \Lambda_b$. For $F_v = \alpha F_a + \beta F_b$, a selector with probabilities proportional to $|\alpha|\Lambda_a$ and $|\beta|\Lambda_b$ directs the sign-adjusted child value to the output through a controlled swap:

$$z_v = \frac{\alpha \Lambda_a z_a + \beta \Lambda_b z_b}{|\alpha|\Lambda_a + |\beta|\Lambda_b}, \qquad \Lambda_v = |\alpha|\Lambda_a + |\beta|\Lambda_b. \tag{11}$$

When $\alpha = \beta = 0$, use a constant-zero node. These two circuits implement precisely the products and weighted sum in Equations 3–4. For the bounded pairwise interaction N, a Toffoli into $|0\rangle$ gives

$$\langle Z_{\mathrm{out}} \rangle = \mathsf{N}(z_a, z_b). \tag{12}$$

All three rules return a bounded expectation that can enter the next node; Appendix C.2 derives their gate identities.

Resource accounting. One local query applies a single-variable signal circuit or its inverse; for phase input x_j, write it as $W_j(x_j)$, for example $e^{ix_j Z}$. We assume every matrix entry of W_j and its inverse is a Laurent polynomial in e^{ix_j} of degree at most an absolute constant b ($b = 1$ in the example). This *constant-bandwidth* condition bounds the per-call increase in an amplitude's Fourier degree by the constant b and underlies the query lower bound. With T_j calls to W_j or its inverse, $Q_{\mathrm{sig}} = \sum_j T_j$; D_{sig} denotes query depth with at most W_{par} local calls in parallel. Each local call is counted individually; a fully multiplexed oracle over all variables defines a different access model. We also track cross-component aggregation gates G_{agg}, aggregation depth D_{agg}, logical width $Q_{\log}$, and the root scale Λ, which controls measurement cost through the variance of the scaled output.

Compiling the full tree. Prepare each sibling subtree on disjoint registers and apply its parent's local circuit. Repeating this step from the leaves to the root realizes Equation 5; the following theorem gives both its exact output and the resulting resource costs.

Theorem 2 (Structure-preserving tree compiler)**.** *Suppose every leaf has an exact module realization and every internal node has an exact Z-compatible module realization. Then every finite* QSFN *formula compiles exactly to a module of the form equation 7. Operation counts add over the tree; parallel depth is the largest sum of local depths along a root-to-leaf path. At each node, logical width is bounded by the sum of child-register widths and fresh work qubits. Scales follow the product and signed-sum rules in Equations 10–11; other compatible nodes use their specified bounded rule and scale. For local errors η_v, the normalized root error obeys $e_{\mathrm{root}} \le \sum_v I_v \eta_v$, where I_v is the product of coordinate Lipschitz constants along the path from v to the root.*

Appendix C.3 gives the proof by induction over the tree and the resource recurrences; Appendix C.6 develops the two-dimensional circuit example and its higher-dimensional extension.

3.3 Approximation and stability

The compiler evaluates a selected finite formula exactly. We now return to the target f: first establish approximation for every continuous function, then quantify how a short analytic factorization controls degree, formula size, and normalization.

Theorem 3 (Universal and analytic approximation)**.** *Let $K \subset \mathbb{R}^d$ be compact.*

1. *For every $f \in C(K)$ and $\epsilon > 0$, there exists a finite* QSFN *formula F such that $\|f - F\|_{\infty,K} \le \epsilon$.*

2. *On $K = [-1,1]^d$, suppose $f(\boldsymbol{x}) = \sum_{r=1}^{R} a_r \prod_{j=1}^{d} g_{rj}(x_j)$, where $|g_{rj}| \le 1$, every g_{rj} extends analytically to a common Bernstein ellipse E_ρ with norm at most M, and $A = \sum_r |a_r| > 0$. The degree*

$$p \ge \frac{\log(2MAed/((\rho-1)\epsilon))}{\log \rho} \tag{13}$$

suffices for uniform error ϵ. If $S(F)$ charges a degree-p spectral leaf by $p+1$ and every constant-arity internal node by one, the compiled formula has size $O(Rdp) = O(Rd \log(MAd/\epsilon)/\log \rho)$, balanced height $O(\log p + \log d + \log R)$, and normalization $\Lambda \le eA$. The height is the algebraic dependency height; signal-query depth is recorded separately and is $O(p)$ when all univariate leaves run in parallel.

Stone–Weierstrass gives polynomial approximants that the compiler realizes exactly. For analytic factors, local Chebyshev truncations bound both the accumulated approximation error and the root scale (Appendix C.5); the root scale also sets the sampling term in the measured-output guarantee below.

Proposition 4 (End-to-end stability)**.** *Let ϵ_{app} bound $|f - F|$ and ϵ_{imp} bound the deviation of the implemented circuit's mean output from F, including finite-precision and channel errors. From N_{shot} independent root measurements, the estimate $\widehat{F}_{N_{\mathrm{shot}}}$ satisfies, with probability at least $1-\delta$,*

$$|f - \widehat{F}_{N_{\mathrm{shot}}}| \le \epsilon_{\mathrm{app}} + \epsilon_{\mathrm{imp}} + \Lambda\sqrt{2\log(2/\delta)/N_{\mathrm{shot}}}. \tag{14}$$

The compiler weights local errors by their path influences, which are at most one for unit-scale rules. Appendix F derives the implementation-error bound and the synthesis and measurement costs.

3.4 Fixed-Error Resource Separations

The compiler's cost follows the degrees of the local leaves and the size of their composition tree. At fixed error, we compare QSFN's local-query cost with the number of global Fourier terms needed for the same target. The comparator may choose any n integer frequencies and complex coefficients; all L^2 norms use normalized uniform measure on $\mathbb{T}^d$, where $\mathbb{T} = [-\pi, \pi)$.

Fourier-sum comparison. For even d, take the cosine polynomial ϕ_d of Appendix D.1, with $|\phi_d| \le 1$ and degree $p_d = 128d - 3$, and define $F_d(\boldsymbol{x}) = \prod_{j=1}^{d/2} \mathsf{N}(\phi_d(x_{2j-1}), \phi_d(x_{2j}))$. The resulting F_d has squared L^2 norm at least $3/4$, while every global Fourier coefficient has squared magnitude at most 2^{-d} (Appendix D.2).

Theorem 5 (Fourier sums versus formula complexity)**.** *The functions $F_d : \mathbb{T}^d \to [-1, 1]$ defined above form an explicit family of bounded continuous functions for even d. For every fixed $0 < \epsilon < \sqrt{3}/2$, any n-term Fourier sum approximating F_d with L^2 error at most ϵ requires*

$$n \ge (3/4 - \epsilon^2)2^d. \tag{15}$$

In the local signal-oracle model, QSFN *realizes F_d with $Q_{\mathrm{sig}} = \Theta(d^2)$ and $D_{\mathrm{sig}} = \Theta(d)$ using $W_{\mathrm{par}} = \Theta(d)$ local calls in parallel; $G_{\mathrm{agg}} = \Theta(d)$, $D_{\mathrm{agg}} = \Theta(\log d)$, $Q_{\log} = O(d)$, and $\Lambda = 1$.*

The local-query bound is tight for this family (Appendix D.3); the balanced tree also attains the minimum cross-component gate count and depth (Appendix C.4).

Native modulus comparison. The native SAQNN rank-one readout returns $|G_n|$ for a complex n-term Fourier polynomial (Tang et al., 2026). From the same local leaves, define the nonnegative target

$$B_d(\boldsymbol{x}) = \prod_{j=1}^{d} \phi_d(x_j), \qquad H_d^{\mathrm{nat}}(\boldsymbol{x}) = (1 + B_d(\boldsymbol{x}))/2. \tag{16}$$

A product tree computes B_d, and an equal-weight sum with one computes H_d^{nat} at unit scale. Squaring the readout gives $|G_n|^2 = G_n\overline{G_n}$, with at most n^2 pairwise differences of its frequencies; Appendix E bounds how well this readout can capture B_d.

Theorem 6 (Native spectral-modulus separation)**.** *For $d \ge 2$, the functions $H_d^{\mathrm{nat}} : \mathbb{T}^d \to [0, 1]$ in Equation 16 form a bounded continuous family. Let G_n be any complex Fourier polynomial with at most n frequencies, with arbitrary global rescaling absorbed into its coefficients. If $\||G_n| - H_d^{\mathrm{nat}}\|_2 \le 1/16$, then*

$$n \ge \sqrt{11/189}\,(\pi/2)^{d/2}. \tag{17}$$

The corresponding QSFN *construction uses $\Theta(d^2)$ local queries and unit normalization.*

For both target families, preserving local composition gives quadratic quantum query cost alongside exponential Fourier-branch requirements at fixed accuracy.

4 Experimental Results

E1 compares resources at matched accuracy, E2 isolates factorization with exact controls, and E3 tests exponential Fourier branches against polynomial quantum queries at fixed error.

4.1 Protocol

SAQNN-*Reimplementation* independently implements the published formulas and reported optimization settings. SAQNN-*Oracle* receives the true spectral coefficients, and QSFN-*Oracle* compiles the target structure deterministically. Oracle comparisons share a gate set and all-to-all connectivity; compiled resources are medians over five seeds. Appendix H gives the complete settings and result tables.

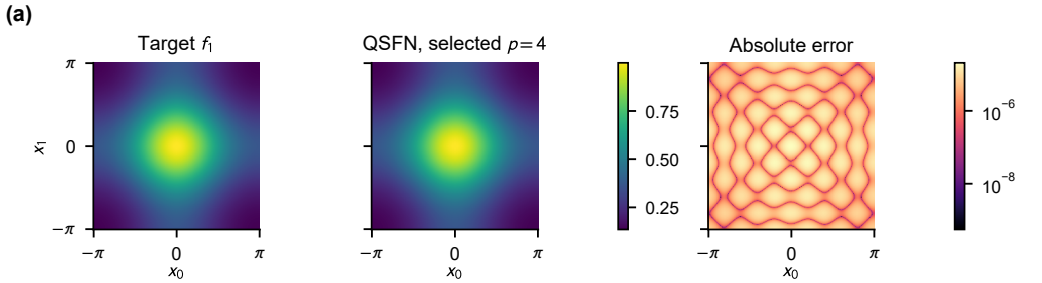
4.2 E1: fidelity and matched resources

We evaluate the three source tasks:

$$f_1(x_0, x_1) = e^{-\sin^2(x_0/2) - \sin^2(x_1/2)}, \tag{18}$$

$$f_2(x_0, x_1) = [1 + \cos(x_0 + x_1 + \pi/6)]/2.5, \tag{19}$$

$$f_3(x_0, x_1) = (x_0 - x_1 + 1)^2/9. \tag{20}$$



(b) Matched comparisons

Task	Model	Grid MSE	Terms	Q_{sig}	Q	G_2	D_2
f_1	SAQNN	1.867×10^{-8}	49	48	7	328	326
	QSFN $p=3$	1.867×10^{-8}	8	10	14	81	31
f_2	SAQNN	5.993×10^{-33}	3	2	3	14	14
	QSFN	0	2	2	5	16	16
f_3	SAQNN	2.627×10^{-33}	6	6	5	206	205
	QSFN	8.943×10^{-34}	6	4	10	33	17

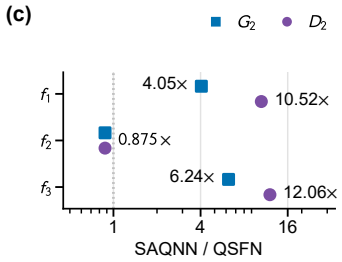


Figure 2: **E1: approximation fidelity and resources.** (a) Target, selected QSFN surface, and absolute error for the product-Fourier task; this is the higher-accuracy $p = 4$ operating point. (b) Dense-grid MSE and oracle resources: terms, signal calls Q_{sig}, logical qubits Q, and median compiled two-qubit gates G_2 and depth D_2. The matched $p = 3$ rows have equal MSE. (c) SAQNN/QSFN gate and depth ratios.

For each of 20 seeds, 200 uniform points are split equally for training and testing; a fixed 256^2 grid supplies the primary error. QSFN represents f_1 as two equal univariate leaves and a product, f_2 with its joint frequency identity, and f_3 by independently preparing two copies of $z = (x_0 - x_1 + 1)/3$ and multiplying them.

At matched $p = 3$ and grid MSE 1.867×10^{-8}, QSFN reduces the representation from 49 to 8 terms, using 4.05-fold fewer two-qubit gates and 10.52-fold lower two-qubit depth. The $p = 4$ operating point reaches grid MSE 4.645×10^{-11}. For f_3, both oracle constructions reach numerical precision; QSFN uses 6.24-fold fewer two-qubit gates and 12.06-fold lower two-qubit depth. These reductions double logical width on both tasks. In contrast, the three-term spectrum of f_2 is already compact: both constructions reach numerical precision, with no QSFN resource advantage. The three tasks thus connect the resource gains to compact product structure in the target functions. On f_1, the 20 trained runs have median grid MSE 1.028×10^{-3} (Figure 8); the matched oracle rows in Figure 2 provide the resource comparison.

4.3 E2: EXACT FACTORIZATION MECHANISM

To isolate structure, we remove training and approximation error and represent

$$P_d(\boldsymbol{x}) = \prod_{j=1}^{d} \frac{1 + \frac{1}{2}\cos x_j}{1 + \frac{1}{2}}. \tag{21}$$

SAQNN-Flat uses all 3^d complex Fourier frequencies of P_d as native branches. QSFN-Expanded organizes the $(3^d + 1)/2$ independent real branches obtained by conjugate pairing into a sum tree, while QSFN-Factorized retains $4d - 1$ formula nodes (Figure 3a).

Across fully transpiled $d \leq 8$, all three realizations match the target to floating-point precision and pass independent statevector checks. At $d = 8$, QSFN-Factorized uses 71 two-qubit gates and depth 11, versus 24,572 for both in SAQNN-Flat. The flat-resource recurrence, validated at every compiled dimension, gives 2.35×10^{13}-fold gate compression at $d = 32$. QSFN-Expanded retains exponential growth while the factorized formula remains linear, isolating preserved factorization as the source of the gain.

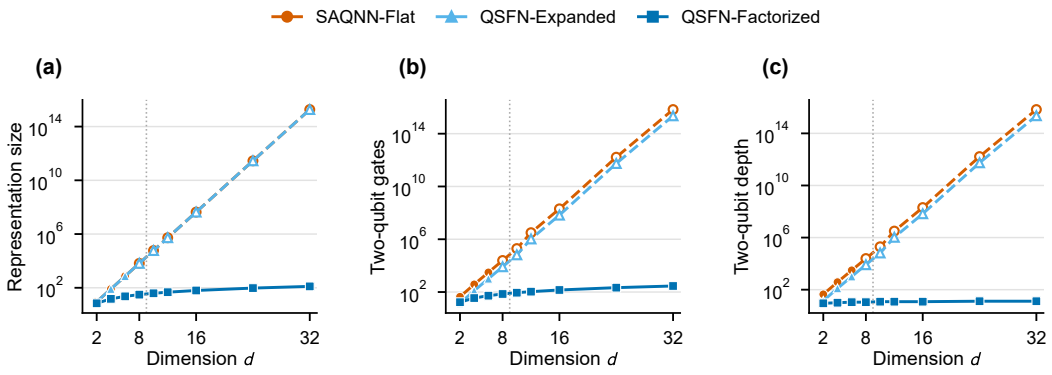


Figure 3: **E2: exact structural separation.** (a) Native complex branches of SAQNN-Flat and formula nodes of both QSFN realizations. (b,c) Two-qubit gates and depth. Flat and expanded circuits are transpiled through $d = 8$; later open points use validated exact recurrences or serial-depth upper bounds. Factorized circuits are transpiled throughout; compression ratios appear in Figure 11.

4.4 E3: FIXED-ERROR AND NATIVE-INTERFACE CERTIFICATES

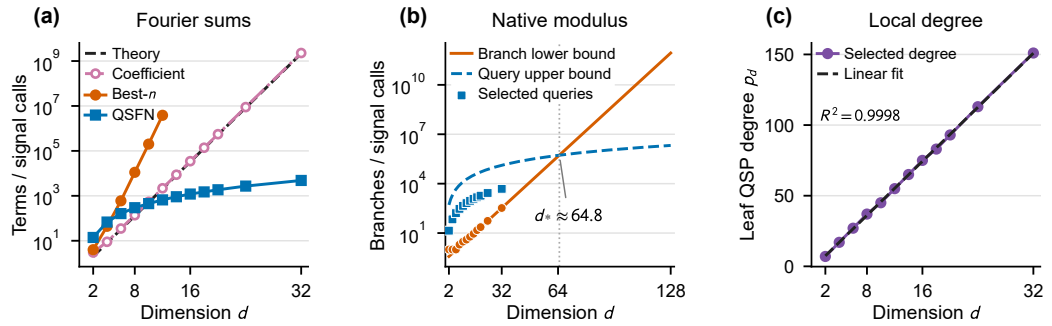


Figure 4: **E3: fixed-error separation.** (a) Theoretical and coefficient-based Fourier-sum lower bounds, exact best-n values where enumeration completes, and QSFN signal calls at error $1/2$. (b) Native-modulus branch lower bound at error $1/16$, the uniform $128d^2$ query upper bound, and selected QSFN query counts. Markers show evaluated dimensions; continuous curves show analytic bounds, with crossover $d \approx 64.8$. (c) Selected QSP degree ($p_d < 128d$) and its linear fit.

At fixed error $\epsilon = 1/2$, Equation 51 certifies $n \geq 2^{d-1}$ for Fourier sums with arbitrary frequencies and coefficients: at $d = 32$, at least 2.147×10^9 branches versus 4,832 QSFN signal calls (Figure 4a). Query and leaf-degree trends are consistent with the proved d^2 and d orders (Figure 4c). For the native target equation 16 at error $1/16$, computed query counts remain below $128d^2$; this upper bound and the branch lower bound of Theorem 6 cross analytically beyond the tested dimensions (Figure 4b and Figure 12 in Appendix H).

5 CONCLUSIONS

QSFN realizes local spectral functions and their composition as quantum circuits, with universal approximation of continuous functions on compact domains and explicit resource and stability guarantees. For specific continuous bounded families, circuits using quadratically many local queries achieve exponential separations from global spectral representations at fixed error, covering both Fourier sums and spectral-modulus readout. Matched-error comparisons and exact factorization controls identify compact composition as the source of compiled resource savings. Beyond function approximation, QSFN can combine expectation-valued features from independently prepared circuits through weighted sums and products in a single final readout for tasks such as quantum state classification and many-body property prediction.

AI USE STATEMENT

Generative AI tools were used to assist with literature search and synthesis, manuscript structure, scientific figure refinement, suggestions for experimental design and architecture, experimental-code review, result interpretation, suggestions for manuscript refinement, and consistency checks. The authors checked the theoretical claims, proofs, experimental protocols, code, numerical results, and cited sources against the archived source assets and primary literature. The authors take responsibility for the final content.

ETHICS STATEMENT

This work studies mathematical representations and simulated quantum circuits and uses no human-subject or personal data. Analytic, simulated, and transpiled evidence is labeled separately throughout.

REPRODUCIBILITY STATEMENT

The appendices state all assumptions and proofs, target functions, hyperparameters, seeds, simulator and compiler settings, metrics, and validation criteria. The anonymous artifact contains the three single-file experiment scripts and archived result packages.

REFERENCES

Jan Balewski, C. Pestano, Mercy G. Amankwah, E. Wes Bethel, Talita Perciano, and Roel Van Beeumen. EHands: Quantum protocol for polynomial computation on real-valued encoded states, 2025. URL https://arxiv.org/abs/2502.15928.

Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. Noncommuting mixed states cannot be broadcast. *Physical Review Letters*, 76:2818–2821, 1996. doi: 10.1103/PhysRevLett.76.2818.

Andrew R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):930–945, 1993. doi: 10.1109/18.256500.

Peter L. Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3:463–482, 2002. URL https://jmlr.org/papers/v3/bartlett02a.html.

Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. doi: 10.1145/502090.502097.

Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002. doi: 10.1090/conm/305/05215.

Nadav Cohen, Or Sharir, and Amnon Shashua. On the expressive power of deep learning: A tensor analysis. In *Proceedings of the 29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 698–728. PMLR, 2016. URL https://proceedings.mlr.press/v49/cohen16.html.

Olivier Delalleau and Yoshua Bengio. Shallow vs. deep sum-product networks. In *Advances in Neural Information Processing Systems*, volume 24, 2011. URL https://proceedings.neurips.cc/paper/2011/hash/8e6b42f1644ecb1327dc03ab345e618b-Abstract.html.

Ronald A. DeVore and George G. Lorentz. *Constructive Approximation*, volume 303 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1993. doi: 10.1007/978-3-662-02888-9.

Yulong Dong, Xiang Meng, K. Birgitta Whaley, and Lin Lin. Efficient phase-factor evaluation in quantum signal processing. *Physical Review A*, 103:042419, 2021. doi: 10.1103/PhysRevA.103.042419.

András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 193–204, 2019. doi: 10.1145/3313276.3316366.

Lukas Gonon, Rodrigo Martínez-Peña, and Juan-Pablo Ortega. Feedback-driven recurrent quantum neural network universality. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=248ysaRatx.

Lars Grasedyck, Daniel Kressner, and Christine Tobler. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36(1):53–78, 2013. doi: 10.1002/gamm.201310004.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2): 251–257, 1991. doi: 10.1016/0893-6080(91)90009-T.

Petr Ivashkov, Po-Wei Huang, Kelvin Koor, Lirandë Pira, and Patrick Rebentrost. QKAN: Quantum kolmogorov-arnold networks with applications in machine learning and multivariate state preparation. *npj Quantum Information*, 12:73, 2026. doi: 10.1038/s41534-026-01202-5.

Ali Javadi-Abhari, Matthew Treinish, Kevin Krsulich, Christopher J. Wood, Jake Lishman, Julien Gacon, Simon Martiel, Paul D. Nation, Lev S. Bishop, Andrew W. Cross, Blake R. Johnson, and Jay M. Gambetta. Quantum computing with Qiskit, 2024. URL https://arxiv.org/abs/2405.08810.

Lorenzo Laneve and Stefan Wolf. On multivariate polynomials achievable with quantum signal processing. *Quantum*, 9:1641, 2025. doi: 10.22331/q-2025-02-20-1641.

Guang Hao Low and Isaac L. Chuang. Optimal hamiltonian simulation by quantum signal processing. *Physical Review Letters*, 118(1):010501, 2017. doi: 10.1103/PhysRevLett.118.010501.

John M. Martyn, Zane M. Rossi, Kevin Z. Cheng, Yuan Liu, and Isaac L. Chuang. Parallel quantum signal processing via polynomial factorization. *Quantum*, 9:1834, 2025. doi: 10.22331/q-2025-08-27-1834.

Hrushikesh N. Mhaskar and Tomaso Poggio. Deep vs. shallow networks: An approximation theory perspective. *Analysis and Applications*, 14(6):829–848, 2016. doi: 10.1142/S0219530516400042.

Danial Motlagh and Nathan Wiebe. Generalized quantum signal processing. *PRX Quantum*, 5(2): 020368, 2024. doi: 10.1103/PRXQuantum.5.020368.

Adrián Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, and José I. Latorre. Data re-uploading for a universal quantum classifier. *Quantum*, 4:226, 2020. doi: 10.22331/q-2020-02-06-226.

Adrián Pérez-Salinas, David López-Núñez, Artur García-Sáez, P. Forn-Díaz, and José I. Latorre. One qubit as a universal approximant. *Physical Review A*, 104(1):012405, 2021. doi: 10.1103/PhysRevA.104.012405.

Adrián Pérez-Salinas, Mahtab Yaghubi Rad, Alice Barthe, and Vedran Dunjko. Universal approximation of continuous functions with minimal quantum circuits. *Physical Review Research*, 7(4): 043282, 2025. doi: 10.1103/t49h-jmty.

Arthur G. Rattew, Po-Wei Huang, Naixu Guo, Lirandë Pira, and Patrick Rebentrost. Accelerating inference for multilayer neural networks with quantum computers. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=QcRto0GjxC.

Neil J. Ross and Peter Selinger. Optimal ancilla-free Clifford+T approximation of z-rotations. *Quantum Information and Computation*, 16(11–12):901–953, 2016. doi: 10.26421/QIC16.11-12-1.

Zane M. Rossi and Isaac L. Chuang. Multivariable quantum signal processing (M-QSP): Prophecies of the two-headed oracle. *Quantum*, 6:811, 2022. doi: 10.22331/q-2022-09-20-811.

Zane M. Rossi, Jack L. Ceroni, and Isaac L. Chuang. Modular quantum signal processing in many variables. *Quantum*, 9:1776, 2025. doi: 10.22331/q-2025-06-18-1776.

Walter Rudin. *Principles of Mathematical Analysis*. McGraw–Hill, 3 edition, 1976.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3):032430, 2021. doi: 10.1103/PhysRevA.103.032430.

Marshall H. Stone. The generalized Weierstrass approximation theorem. *Mathematics Magazine*, 21 (5):237–254, 1948. doi: 10.2307/3029337.

Jialiang Tang, Jialin Zhang, and Xiaoming Sun. SAQNN: Spectral adaptive quantum neural network as a universal approximator. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.09718.

Dmitry Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Networks*, 94: 103–114, 2017. doi: 10.1016/j.neunet.2017.07.002.

Lexing Ying. Stable factorization for phase factors of quantum signal processing. *Quantum*, 6:842, 2022. doi: 10.22331/q-2022-10-20-842.

Zhan Yu, Qiuhao Chen, Yuling Jiao, Yinan Li, Xiliang Lu, Xin Wang, and Jerry Zhijian Yang. Non-asymptotic approximation error bounds of parameterized quantum circuits. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://openreview.net/forum?id=XCkII8nCt3.

A Additional Related-Work Context

Structural and spectral antecedents. Deep sum-product networks show how alternating sums and products can give compact deep representations despite large shallow expansions (Delalleau & Bengio, 2011). Data re-uploading studies the Fourier frequencies accessible through repeated quantum encodings (Pérez-Salinas et al., 2020; Schuld et al., 2021).

Other approximation settings. Feedback-driven recurrent quantum networks establish quantitative universality for temporal state-space systems with input histories and linear readouts (Gonon et al., 2026). This provides a temporal counterpart to the static multivariate approximation studied here.

Circuit synthesis and composition. Phase-factor algorithms turn bounded polynomials into QSP sequences for the local leaves (Dong et al., 2021; Ying, 2022). Multivariable QSP studies joint transformations of operator signals and conditions under which the corresponding polynomials can be implemented (Rossi & Chuang, 2022; Laneve & Wolf, 2025). Parallel QSP factorizes a polynomial across multiple copies to reduce query depth, at the cost of more measurements (Martyn et al., 2025).

Input access and resource accounting. For coherent multilayer inference, the cost changes with access to inputs and weights (Rattew et al., 2026). Section 3.2 therefore reports local signal calls separately from the cost of supplying those signals.

B Mathematical Setting and Proof Roadmap

The universal approximation result covers bounded continuous functions on compact domains. Within this class, Theorem 5 constructs the explicit family F_d and compares it with Fourier sums of arbitrary frequencies and complex coefficients. Theorem 6 uses the nonnegative companion family H_d^{nat} and the native rank-one modulus readout $|G_n|$, with any global scale absorbed into its coefficients. Table 1 summarizes the corresponding settings.

All torus norms use the probability measure $d\mu_d(\boldsymbol{x}) = \prod_{j=1}^{d} dx_j/(2\pi)$ on $\mathbb{T}^d$. For $f \in L^2(\mathbb{T}^d)$, the Fourier convention and Parseval identity are

$$\widehat{f}(\boldsymbol{k}) = \int f(\boldsymbol{x}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} d\mu_d(\boldsymbol{x}), \qquad \|f\|_2^2 = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\widehat{f}(\boldsymbol{k})|^2. \tag{22}$$

For a finite $S \subset \mathbb{Z}^d$, orthogonal projection gives

$$\inf_{\operatorname{supp}(\widehat{g}) \subseteq S} \|f - g\|_2^2 = \|f\|_2^2 - \sum_{\boldsymbol{k} \in S} |\widehat{f}(\boldsymbol{k})|^2. \tag{23}$$

Selecting the n largest coefficient magnitudes gives the best n-term error:

$$\sigma_n(f)_2^2 = \|f\|_2^2 - \sum_{j=1}^{n} |\widehat{f}|_{(j)}^2, \tag{24}$$

where coefficients are ordered by magnitude. The optimal approximation keeps the n largest coefficient magnitudes; its squared error is the sum of the squared magnitudes of the omitted coefficients. The Fourier-sum lower bound follows by constructing a continuous bounded F_d whose squared coefficient magnitudes sum to at least $3/4$, with each at most 2^{-d}.

The query lower bound uses the polynomial method (Beals et al., 2001). By the constant-bandwidth convention in Section 3.2, an amplitude has Laurent degree zero before any query. An input-independent unitary forms linear combinations and leaves this degree unchanged. Multiplying by W_j or W_j^{-1} increases it by at most b. Induction therefore bounds every amplitude after T_j calls by bT_j. An expectation is bilinear in amplitudes and their conjugates, so its Fourier degree is at most $c_q T_j$ with $c_q = 2b$. A nonzero highest frequency p_j thus implies $T_j \geq p_j/c_q = \Omega(p_j)$. We count each local call separately; a unified multiplexed oracle is a different access model.

The logical dependency is

$$\text{Jackson moment} \Rightarrow \text{local norm bound} \Rightarrow \text{Fourier coefficient bounds} \Rightarrow \text{Fourier-sum lower bound},$$

while

$$\begin{aligned} &\text{interface characterization} \Rightarrow \text{tree compiler} \Rightarrow \text{NAND/product construction} \\ &\qquad \Rightarrow \Theta(d^2) \text{ local queries.} \end{aligned}$$

The native-interface theorem reuses the Jackson leaf and adds a squared-modulus difference-set argument.

C Bounded Composition Rules and Compiler Proofs

Figure 5 links the tree compiler to its signed-expectation readout and the CNOT and Toffoli rules used below.

C.1 Proof of the interface characterization

For an interface state $\rho_i = (I + x_i X + y_i Y + z_i Z)/2$, its computational-basis probabilities are $[1 + (-1)^{b_i} z_i]/2$. A diagonal output on the product state therefore equals Equation 8; every factor is affine in one coordinate, proving necessity. Conversely, let G be bounded and multiaffine and assign the diagonal vertex value $\tau(\boldsymbol{b}) = G((-1)^{b_1}, \ldots, (-1)^{b_r})$. The induced G_τ is multiaffine and agrees with G at every vertex. Multiaffine interpolation is unique, so $G_\tau = G$ throughout the cube. Boundedness gives $|\tau(\boldsymbol{b})| \leq 1$.

The Walsh form makes the same result explicit. If $\widehat{\tau}(S) = 2^{-r} \sum_{\boldsymbol{b}} \tau(\boldsymbol{b})(-1)^{\sum_{i \in S} b_i}$, then independence gives

$$G_\tau(\boldsymbol{z}) = \sum_{S \subseteq [r]} \widehat{\tau}(S) \prod_{i \in S} z_i. \tag{25}$$

Furthermore, $\partial_{z_i} G_\tau = \frac{1}{2}[\mathbb{E}(\tau \mid B_i = 0) - \mathbb{E}(\tau \mid B_i = 1)]$, so every coordinate Lipschitz constant is at most one.

Output-qubit realization. Equation 8 defines the desired expectation on the child registers. To export it through a single output qubit, first consider $\tau(\boldsymbol{b}) \in \{-1, +1\}$. Reversibly compute a bit $y_\tau(\boldsymbol{b})$ satisfying $(-1)^{y_\tau(\boldsymbol{b})} = \tau(\boldsymbol{b})$; its Pauli-$Z$ expectation is then exactly $G_\tau(\boldsymbol{z})$. For a general bounded table $\tau(\boldsymbol{b}) \in [-1, 1]$, reversibly compute a δ-accurate value $\widetilde{\tau}(\boldsymbol{b})$ and apply $R_y(\arccos \widetilde{\tau})$ to an output qubit initialized in $|0\rangle$. Its Pauli-Z expectation equals $\widetilde{\tau}(\boldsymbol{b})$ on each basis branch and

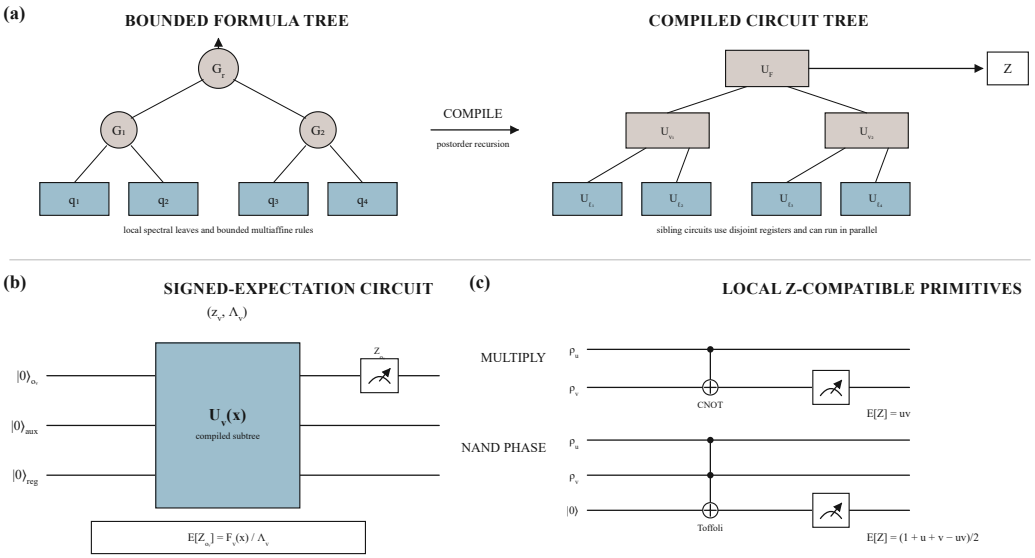


Figure 5: A formula-tree compiler. (a) Local leaves and bounded multiaffine rules compile recursively on disjoint registers. (b) A subtree circuit U_v exposes $z_v = \langle Z_{o_v} \rangle = F_v/\Lambda_v$. (c) CNOT realizes multiplication and Toffoli realizes N for independent child outputs.

hence realizes the corresponding Bernoulli extension. If $C_{\mathrm{rev}}(\tau, \delta)$ and $C_{\mathrm{rot}}(\delta)$ denote the reversible evaluation and controlled-rotation costs, the local realization cost is $O(C_{\mathrm{rev}} + C_{\mathrm{rot}})$. The induced error is a node error in Proposition 4; ideal arithmetic and rotations recover the exact semantic rule used by Theorem 2. A general r-input table has 2^r entries; the constant-arity rules used in the compiler keep this local table size bounded.

C.2 Elementary composition circuits

For multiplication, CNOT conjugation gives

$$\mathrm{CNOT}^{\dagger}_{a\to b}(I \otimes Z_b)\,\mathrm{CNOT}_{a\to b} = Z_a \otimes Z_b. \tag{26}$$

On $\rho_a \otimes \rho_b$, the output is uv. For a signed linear combination $F = \alpha F_a + \beta F_b$, flip signs using $XZX = -Z$, prepare a selector with probabilities proportional to $|\alpha|\Lambda_a$ and $|\beta|\Lambda_b$, and controlled-swap the selected interface into a fixed port. This yields scale $|\alpha|\Lambda_a + |\beta|\Lambda_b$. For the NAND vertex rule $(-1)^{b_1 b_2}$, Equation 8 gives

$$\mathsf{N}(u, v) = 1 - 2\frac{1-u}{2}\frac{1-v}{2} = \frac{1+u+v-uv}{2}. \tag{27}$$

A Toffoli into $|0\rangle$ exposes this value as a Z expectation.

C.3 Exact compilation and sharing

For Theorem 2, the complete resource vector is

$$(Q_{\mathrm{sig}}, D_{\mathrm{sig}}, G_{\mathrm{agg}}, D_{\mathrm{agg}}, Q_{\log}, T, T_D, \Lambda, N_{\mathrm{shot}}), \tag{28}$$

where T and T_D are Clifford+T count and depth after gate-set decomposition. At an internal node v with children u, write c_v^R for the local count cost, d_v^R for the local depth cost, and a_v for the number of fresh work qubits. The resource recurrences are

$$\begin{aligned} R(v) &= \sum_u R(u) + c_v^R, && R \in \{Q_{\mathrm{sig}}, G_{\mathrm{agg}}, T\}, \\ R(v) &= \max_u R(u) + d_v^R, && R \in \{D_{\mathrm{sig}}, D_{\mathrm{agg}}, T_D\}, \\ Q_{\log}(v) &\le \sum_u Q_{\log}(u) + a_v. \end{aligned} \tag{29}$$

For the general sum–product form, let ℓ_{rj} denote the compiled local-leaf cost, such as $p+1$ for a degree-p polynomial leaf. The tree retains one copy of each local description and its composition nodes, giving $S_{\mathrm{form}} = O(\sum_{r,j} \ell_{rj} + Rd)$. Thus R products of d local s-term, degree-$O(s)$ factors can generate Rs^d global terms from an $O(Rds)$ formula. Theorem 3 controls accuracy and scale for analytic factors.

Proof of Theorem 2 proceeds by induction. A height-zero tree is a leaf module with a declared query count, work-qubit count, and scale. For a node of height h, independently execute the compiled children on disjoint registers and apply the node's compatible map. Its semantic rule supplies the desired expectation. Counts add across all children and the local circuit; parallel child blocks contribute their maximum depth; and disjoint live registers add in width. These three observations give every recurrence in Equation 29, including T count and depth after a gate-set decomposition.

The scale recurrence follows from the same semantic induction. Multiplication of child expectations gives $F_a F_b/(\Lambda_a\Lambda_b)$, hence $\Lambda_v = \Lambda_a\Lambda_b$. For $F_v = \alpha F_a + \beta F_b$, the selector probabilities in the preceding subsection give $\Lambda_v = |\alpha|\Lambda_a + |\beta|\Lambda_b$. A unit-scale vertex rule has $\Lambda_v = 1$ when all declared inputs are unit-scale. At root scale Λ, Hoeffding's inequality gives the sufficient measurement count $N_{\mathrm{shot}} \geq 2\Lambda^2\epsilon_{\mathrm{shot}}^{-2}\log(2/\delta)$ for additive accuracy ϵ_{shot} and failure probability δ. This is a root readout cost; coherent internal nodes require no intermediate measurement.

Finally, if $L_{v,u}$ is the coordinate Lipschitz constant on edge $u \to v$, the local implementation error obeys $e_v \leq \eta_v + \sum_u L_{v,u} e_u$. Repeated substitution yields $e_{\mathrm{root}} \leq \sum_v I_v \eta_v$, where I_v is the product of edge constants along the unique path from v to the root.

If a shared node produces a noncommuting state family, no fixed channel can map every state to a bipartite state with both marginals equal to the input (Barnum et al., 1996). An expectation-valued directed acyclic graph must therefore unfold shared nodes, re-estimate and re-encode them, or prove a stronger interface. If a value estimated with error $|\widehat{z} - z|$ feeds paths with influences $I_1, \ldots, I_m$, the root contribution is at most $(\sum_j I_j)|\widehat{z} - z|$; achieving error ϵ by independent shots costs $O((\sum_j I_j)^2\epsilon^{-2}\log(1/\delta))$. Commuting state families also permit broadcasting; the cost of the corresponding channel must be included when using this form of reuse.

C.4 Optimal multiplication aggregation

Initially, m independent inputs occupy distinct information components. A two-qubit gate can merge at most two components, so at least $m-1$ cross- component gates are necessary for one output to depend on every input. In a depth-D circuit of disjoint two-qubit layers, the backward light cone of one output contains at most 2^D inputs. Thus $D \geq \lceil \log_2 m \rceil$. A balanced CNOT tree attains both bounds.

C.5 Universal and analytic approximation

For a compact $K \subset \mathbb{R}^d$, Stone–Weierstrass gives a multivariate polynomial uniformly within ϵ of any $f \in C(K)$ (Stone, 1948; Rudin, 1976). Coordinate leaves, multiplication trees, and signed sums compile this polynomial, proving universal approximation.

For the analytic class on $[-1,1]^d$, $\|f\|_\infty \leq A$; hence the zero formula suffices when $\epsilon \geq A$. Assume $0 < \epsilon < A$. Analytic Chebyshev approximation (DeVore & Lorentz, 1993) gives a degree-p truncation with

$$\|g_{rj} - q_{rj}^{(p)}\|_\infty \leq \delta_p := \frac{2M}{\rho - 1}\rho^{-p}, \qquad \|q_{rj}^{(p)}\|_\infty \leq 1 + \delta_p \tag{30}$$

Equation 13 gives $d\delta_p \leq \epsilon/(eA) < 1$. For a fixed rank index r, insert and subtract one factor at a time:

$$\prod_{j=1}^{d} g_{rj} - \prod_{j=1}^{d} q_{rj}^{(p)} = \sum_{j=1}^{d} \left(\prod_{k<j} q_{rk}^{(p)}\right)(g_{rj} - q_{rj}^{(p)})\left(\prod_{k>j} g_{rk}\right). \tag{31}$$

Taking norms and summing the geometric series gives

$$\left\| \prod_j g_{rj} - \prod_j q_{rj}^{(p)} \right\|_\infty \leq (1+\delta_p)^d - 1 \leq ed\delta_p. \tag{32}$$

The outer ℓ_1 norm then bounds the total error by $Aed\delta_p \leq \epsilon$.

It remains to connect this classical truncation to the quantum leaf interface. Set $B_{rj} = \|q_{rj}^{(p)}\|_\infty \leq 1+\delta_p$. Generalized QSP realizes $q_{rj}^{(p)}/B_{rj}$ with $O(p)$ local queries even when the truncation mixes even and odd degrees (Motlagh & Wiebe, 2024). Thus the product node for term r has scale $\prod_j B_{rj} \leq (1+\delta_p)^d \leq e$, and the signed outer sum gives

$$\Lambda \leq \sum_r |a_r| \prod_j B_{rj} \leq eA. \tag{33}$$

There are Rd leaves, each charged $p+1$ spectral coefficients, at most $R(d-1)$ product nodes, and at most $R-1$ outer-sum nodes. Hence $S(F) = O(Rdp)$ and $Q_{\text{sig}} = O(Rdp)$. Balancing the dimension products and rank sum gives algebraic height $O(\log d + \log R)$, which is within the stated $O(\log p + \log d + \log R)$ bound; parallel leaves have signal-query depth $O(p)$. This proves every claim in Theorem 3.

C.6 A Two-Dimensional Example from Spectral Expansion to Quantum Evaluation

Target and local factors. The example connects the preserved formula in Equation 2 to the circuits in Section 3.2. Consider the product of two identical local factors on $[-\pi, \pi)^2$,

$$P_2(x_1, x_2) = h(x_1)h(x_2), \qquad h(x) = \frac{2}{3} + \frac{1}{3}\cos x. \tag{34}$$

Here $h(x) \in [1/3, 1]$, so $P_2 \in [1/9, 1]$ and both factors have unit scale. This is Equation 1 with $R = 1$, $a_1 = 1$, and $g_{11} = g_{12} = h$. The local expansion is exact: $q_{1j} = g_{1j}$ and $F = P_2$.

Global spectral expansion. Selecting one index $k_j \in S_{rj}$ per coordinate expands Equation 2 as follows. For complex Fourier leaves $\psi_k(x) = e^{ikx}$, the tuple $\boldsymbol{k} = (k_1, \ldots, k_d)$ gives the global basis function $e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$:

$$F(\boldsymbol{x}) = \sum_{r=1}^{R} \sum_{(k_1,\ldots,k_d) \in \prod_j S_{rj}} a_r \left(\prod_{j=1}^{d} c_{rjk_j} \right) \prod_{j=1}^{d} \psi_{k_j}(x_j). \tag{35}$$

For P_2, each factor has the Fourier expansion $h(x) = 2/3 + e^{ix}/6 + e^{-ix}/6$, with frequencies $-1, 0, 1$. Expanding the product gives

$$P_2(\boldsymbol{x}) = \sum_{\boldsymbol{s} \in \{-1,0,1\}^2} c_{\boldsymbol{s}} e^{i\boldsymbol{s}\cdot\boldsymbol{x}}, \tag{36}$$

where $c_{\boldsymbol{s}} = b_{s_1} b_{s_2}$ with $b_0 = 2/3$ and $b_{-1} = b_1 = 1/6$. Thus $c_{(0,0)} = 4/9$, the four axis coefficients are $1/9$, and the four corner coefficients are $1/36$. Conjugate pairing gives the equivalent real expression

$$P_2(\boldsymbol{x}) = \frac{4}{9} + \frac{2}{9}(\cos x_1 + \cos x_2) + \frac{1}{18}\big[\cos(x_1 + x_2) + \cos(x_1 - x_2)\big].$$

There are nine complex frequencies, or five real spectral terms after pairing. The published one-term-per-branch SAQNN construction uses the nine complex terms; its modulus readout returns the same target because P_2 is positive. The preserved formula evaluates the two local factors and multiplies their outputs, retaining the same function and coefficients.

Quantum evaluation of the preserved factors. For each coordinate x_j, prepare independent constant-one and $\cos x_j$ leaves as in Section 3.2. The selector in Equation 11 combines these leaves with weights $2/3$ and $1/3$, giving output qubit o_j with expectation $\langle Z_{o_j} \rangle = 2/3 + (1/3)\cos x_j = h(x_j)$ and scale one. Prepare the two factor circuits on disjoint registers. Applying a CNOT from o_1 to o_2 gives

$$\langle Z_{o_2} \rangle_{\text{out}} = \langle Z_{o_1} Z_{o_2} \rangle_{\text{in}} = h(x_1)h(x_2) = P_2(\boldsymbol{x}), \qquad \Lambda_{\text{root}} = 1.$$

The first equality is the CNOT identity in Equation 10; the second follows from independent preparation of the factor circuits. The selected child outputs remain quantum states through composition, with measurement only at the root. This explicit construction has four leaves, two weighted-sum nodes, and one product node: seven formula nodes in total.

Higher-dimensional scaling and E2. In d dimensions, the same construction uses $2d$ leaves, d weighted-sum nodes, and $d-1$ product nodes:

$$P_d(\boldsymbol{x}) = \prod_{j=1}^{d} h(x_j), \qquad N_{\text{global}} = 3^d, \qquad S_{\text{form}} = 4d - 1. \tag{37}$$

Here N_{global} counts the complex frequencies in the fully expanded product, while S_{form} counts the nodes of the explicit formula above. Conjugate pairing yields $(3^d + 1)/2$ real spectral terms. The equality $h(x) = (1 + \frac{1}{2}\cos x)/(1 + \frac{1}{2})$ identifies this family with the E2 target in Equation 21. E2 compares these representations of the same function and coefficients. Section 3.4 establishes the fixed-error separation using the same local and composition operations on the explicit bounded continuous family F_d.

D Separation Family and Fourier-Sum Lower Bound

Figure 6 previews the two representations of F_d and the fixed-error lower bound proved in this section.

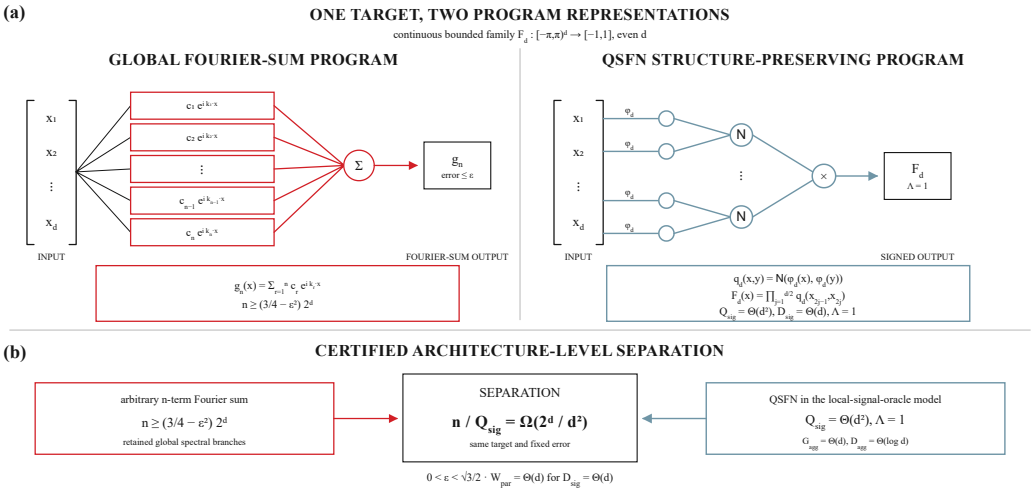


Figure 6: The explicit family F_d: (a) an arbitrary n-term global Fourier sum and the preserved QSFN formula; (b) the fixed-error Fourier-term lower bound and the formula's local signal-query cost. The norm and coefficient estimates in Equation 51 establish the bound.

D.1 The Jackson Leaf

Convolving $r(x) = \operatorname{sgn}(\cos x)$ with the positive normalized Jackson kernel J_n defined below gives a real cosine polynomial $\phi_n = J_n * r$ supported on odd frequencies and satisfying

$$|\phi_n| \le 1, \quad \widehat{\phi}_n(0) = 0, \quad 1 - \|\phi_n\|_2^2 \le 32/n, \quad \deg \phi_n = 2n - 3. \tag{38}$$

The dimension-dependent function in Section 3.4 is $\phi_d = \phi_{n_d}$ with $n_d = 64d$, so its degree is $p_d = 2n_d - 3 = 128d - 3 < 128d$. Its odd cosine expansion is a bounded odd polynomial in $\cos x$, suitable for the QSP construction in Section 3.2. We derive these properties below.

For $n \geq 2$, the Fejér kernel is

$$K_n(t) = \frac{1}{n}\left(\frac{\sin(nt/2)}{\sin(t/2)}\right)^2, \qquad \widehat{K}_n(k) = \max\{1 - |k|/n, 0\}. \tag{39}$$

Parseval gives

$$c_n = \int K_n(t)^2 d\mu(t) = 1 + 2\sum_{k=1}^{n-1}(1 - k/n)^2 = \frac{2n^2 + 1}{3n}. \tag{40}$$

Thus $J_n = K_n^2/c_n$ is nonnegative, normalized, and has degree at most $2n - 2$. We next make the first-moment constant explicit. From the closed form above, $c_n \geq 2n/3$. On $|t| \leq 1/n$, $K_n(t) \leq n$, and hence

$$\begin{aligned} I_1 &:= \int_{|t|\leq 1/n} |t| J_n(t) d\mu(t) \\ &\leq \frac{1}{2\pi}\, 2\int_0^{1/n} t\frac{3n}{2} dt = \frac{3}{4\pi n}. \end{aligned} \tag{41}$$

For $1/n < |t| \leq \pi$, the inequalities $|\sin(nt/2)| \leq 1$ and $\sin(|t|/2) \geq |t|/\pi$ give $K_n(t) \leq \pi^2/(nt^2)$ and $J_n(t) \leq 3\pi^4/(2n^3t^4)$. Therefore

$$\begin{aligned} I_2 &:= \int_{1/n<|t|\leq\pi} |t| J_n(t) d\mu(t) \\ &\leq \frac{3\pi^3}{2n^3}\int_{1/n}^{\pi} t^{-3} dt \leq \frac{3\pi^3}{4n}. \end{aligned} \tag{42}$$

Consequently,

$$\int |t| J_n(t) d\mu(t) = I_1 + I_2 < 24/n < 25/n. \tag{43}$$

The square wave $r = \operatorname{sgn}(\cos x)$ has two jumps per period. For $|t| \leq \pi$, the set on which $r(x)$ and $r(x - t)$ differ has length at most $2|t|$; the pointwise difference there is two. Under normalized Haar measure,

$$\|r - r(\cdot - t)\|_1 \leq 2|t|/\pi. \tag{44}$$

Writing $\phi_n = J_n * r$, positivity and normalization of J_n now give

$$\begin{aligned} \|r - \phi_n\|_1 &\leq \int J_n(t)\|r - r(\cdot - t)\|_1 d\mu(t) \\ &\leq \frac{2}{\pi}\int |t| J_n(t) d\mu(t) \leq \frac{50}{\pi n}. \end{aligned} \tag{45}$$

Since $r^2 = 1$ almost everywhere and $|\phi_n| \leq 1$, $1 - \phi_n^2 \leq 2|r - \phi_n|$. Integration yields

$$1 - \|\phi_n\|_2^2 \leq 2\|r - \phi_n\|_1 \leq 100/(\pi n) < 32/n. \tag{46}$$

Finally, r is even, has zero mean, and satisfies $r(x + \pi) = -r(x)$, so it contains only odd cosine frequencies. At the largest admissible odd frequency, the convolution formula for J_n gives

$$\widehat{J}_n(2n - 3) = \frac{2(2/n)(1/n)}{c_n} = \frac{4}{c_n n^2} > 0, \tag{47}$$

while $\widehat{r}(k) = 2(-1)^{(k-1)/2}/(\pi k) \neq 0$ for odd k. Hence $\deg \phi_n = 2n - 3$. Writing $\cos(kx) = T_k(\cos x)$ makes ϕ_n a bounded odd QSP polynomial and completes the leaf realization.

D.2 Function norms and Fourier coefficient bounds

For even $d = 2m$, the family F_d in Theorem 5 has the explicit form

$$q_d(x, y) = \frac{1 + \phi_d(x) + \phi_d(y) - \phi_d(x)\phi_d(y)}{2}, \qquad F_d(\boldsymbol{x}) = \prod_{j=1}^{m} q_d(x_{2j-1}, x_{2j}). \tag{48}$$

Here $q_d(x, y) = \mathsf{N}(\phi_d(x), \phi_d(y))$, the unit-scale rule in Equation 12. Since ϕ_d is continuous and bounded by one, every $F_d : \mathbb{T}^d \to [-1, 1]$ is a bounded continuous finite trigonometric polynomial.

Write $\rho_d^2 = \|\phi_d\|_2^2$. With $n_d = 64d$, $\rho_d^2 \geq 1 - 1/(2d)$. Since 1, $\phi_d(x)$, $\phi_d(y)$, and $\phi_d(x)\phi_d(y)$ are orthogonal,

$$\|q_d\|_2^2 = \left(\frac{1 + \rho_d^2}{2}\right)^2. \tag{49}$$

Across $m = d/2$ disjoint pairs,

$$\|F_d\|_2^2 = \left(\frac{1 + \rho_d^2}{2}\right)^d \geq (1 - 1/(4d))^d \geq 3/4. \tag{50}$$

If $a_k = \widehat{\phi}_d(k)$, then $a_0 = 0$ and $|a_k| \leq 1$. The coefficients of q_d are $1/2$, $a_k/2$, $a_\ell/2$, and $-a_k a_\ell/2$ in the four zero/nonzero cases, so their magnitudes are at most $1/2$. Tensorization gives $\max_{\boldsymbol{k}} |\widehat{F}_d(\boldsymbol{k})|^2 \leq 2^{-d}$. Together, these give

$$\|F_d\|_2^2 \geq \frac{3}{4}, \qquad \max_{\boldsymbol{k}} |\widehat{F}_d(\boldsymbol{k})|^2 \leq 2^{-d}. \tag{51}$$

Choose any set S of at most n frequencies. Orthogonal projection gives the best approximation on S (Equation 23), and equation 51 shows that these modes capture at most $n2^{-d}$ of the squared norm. Hence every n-term sum g satisfies $\|F_d - g\|_2^2 \geq 3/4 - n2^{-d}$. If its error is at most ϵ, then $n \geq (3/4 - \epsilon^2)2^d$, proving Theorem 5. The set S was arbitrary, so this includes the best n-term sum.

D.3 Constructive and lower-bound resources

Each of the d leaves has nonzero degree $p_d = 2n_d - 3 < 128d$. The leaf convention gives an $O(p_d)$-query construction, while the polynomial-method lemma in Appendix B gives the matching $\Omega(p_d)$ cost. The $d/2$ pairwise NAND rules form one parallel aggregation layer, followed by a balanced product of their outputs. Thus

$$\begin{aligned} Q_{\mathrm{sig}} &= d\,\Theta(p_d) = \Theta(d^2), & D_{\mathrm{sig}} &= \Theta(p_d) = \Theta(d) \quad (W_{\mathrm{par}} = \Theta(d)), \\ G_{\mathrm{agg}} &= d/2 + (d/2 - 1) = d - 1, & D_{\mathrm{agg}} &= 1 + \lceil \log_2(d/2) \rceil, \\ Q_{\log} &= O(d), & \Lambda &= 1. \end{aligned} \tag{52}$$

The width follows because each QSP leaf and each constant-arity aggregator uses $O(1)$ live work qubits. The exact Clifford+T entries additionally depend on the chosen oracle and Toffoli decompositions; Appendix F accounts separately for phase synthesis.

For the matching lower bound, continuity and zero mean give a zero of ϕ_d. Fix the partner of x_i and both variables in every other pair at such zeros. The remaining dependence is $C_i[1 + \phi_d(x_i)]/2$, with $C_i = 2^{-(d/2-1)} \neq 0$, and therefore retains the highest frequency $p_d = \Theta(d)$. The constant-bandwidth query lemma gives $T_i \geq p_d/c_q = \Omega(d)$ for every i, so $Q_{\mathrm{sig}} = \sum_i T_i = \Omega(d^2)$. If only W_{par} local calls can run in parallel,

$$D_{\mathrm{sig}} = \Omega\big(\max\{d, d^2/W_{\mathrm{par}}\}\big), \tag{53}$$

which matches the construction at $W_{\mathrm{par}} = \Theta(d)$.

E Native SAQNN Interface Lower Bound

The target of Theorem 6 is the nonnegative bounded continuous family $H_d^{\mathrm{nat}} : \mathbb{T}^d \to [0, 1]$ in Equation 16, formed from the same ϕ_d as F_d. For its product component B_d, $\|B_d\|_2^2 \geq 9/16$ and

$\max_{\boldsymbol{k}} |\widehat{B}_d(\boldsymbol{k})| \le (2/\pi)^d$. These follow from the product form: each factor has squared norm at least $1 - 1/(2d)$, and Jackson smoothing multiplies each Fourier coefficient of $\mathrm{sgn}(\cos x)$ by a kernel coefficient of magnitude at most one. Hence $\|B_d\|_2^2 \ge (1 - 1/(2d))^d \ge 9/16$ for $d \ge 2$, while each local coefficient has magnitude at most $2/\pi$.

Let $Y = |G_n|$ and $K = Y^2 = G_n\overline{G_n}$. If the support of G_n is S, then $\operatorname{supp} \widehat{K} \subseteq S - S$, so it contains at most n^2 frequencies. Also $K \ge 0$, which implies $|\widehat{K}(\boldsymbol{k})| \le \widehat{K}(\mathbf{0}) = \int K d\mu_d$.

For $H = H_d^{\mathrm{nat}}$ and $\|Y - H\|_2 \le 1/16$, Cauchy–Schwarz gives

$$\|K - H^2\|_1 \le (1/16)(2 + 1/16) = 33/256. \tag{54}$$

Since $H - H^2 = (1 - B_d^2)/4 \ge 0$ and $\|B_d\|_2^2 \ge 9/16$,

$$\|K - H\|_1 \le 33/256 + (1 - 9/16)/4 = 61/256 =: \delta_{\mathrm{nat}}. \tag{55}$$

Using B_d as a witness,

$$\langle B_d, K\rangle \ge \|B_d\|_2^2/2 - \delta_{\mathrm{nat}} \ge 11/256, \qquad \widehat{K}(\mathbf{0}) \le 1/2 + \delta_{\mathrm{nat}} = 189/256. \tag{56}$$

Because each Fourier coefficient of B_d is at most $(2/\pi)^d$,

$$11/256 \le n^2(2/\pi)^d(189/256), \tag{57}$$

which proves Equation 17. QSFN constructs B_d as a signed-Z expectation with a balanced product of d unit-scale leaves. It then applies the selector construction from Appendix C to the constant-one interface and the B_d interface with coefficients $1/2, 1/2$. The exported signed-Z expectation is exactly $H = (1 + B_d)/2$, and its scale is $(1/2)1 + (1/2)1 = 1$. Thus the construction uses the same signed-expectation output interface as the rest of the paper.

F Finite Precision, Sampling, and Noise

F.1 Proof of the end-to-end bound

Let $F^\star = F$ denote the exact output of the selected formula and let F_{imp} denote the mean output of its implemented circuit, including finite-precision and channel errors. To expand Proposition 4, write η_ℓ for normalized leaf errors, δ_j for operator-norm rotation errors, ξ_v for normalized composition errors, and η_g for gate-channel errors in diamond norm. The factors I are the products of coordinate Lipschitz constants along the corresponding paths to the root. With approximation error ϵ_{app} and N_{shot} independent measurements, the full bound holds with probability at least $1 - \delta$:

$$|f - \widehat{F}_{N_{\mathrm{shot}}}| \le \epsilon_{\mathrm{app}} + \Lambda \sum_\ell I_\ell \eta_\ell + 2\Lambda \sum_j I_j \delta_j + \Lambda \sum_v I_v \xi_v \tag{58}$$

$$+ \Lambda \sum_g \eta_g + \Lambda \sqrt{2\log(2/\delta)/N_{\mathrm{shot}}}. \tag{59}$$

The approximation step gives $|f - F^\star| \le \epsilon_{\mathrm{app}}$. For leaf and composition approximations alone, write the resulting output as F_{comp}. The compiler influence inequality gives

$$|F^\star - F_{\mathrm{comp}}| \le \Lambda \sum_\ell I_\ell \eta_\ell + \Lambda \sum_v I_v \xi_v. \tag{60}$$

Replacing rotation j by an operator-norm δ_j approximation changes a unit-norm expectation by at most $2\delta_j$; propagation to the root adds the factor I_j. Channel telescoping similarly bounds the accumulated gate-channel bias by $\Lambda \sum_g \eta_g$. Therefore

$$|F^\star - F_{\mathrm{imp}}| \le \Lambda \sum_\ell I_\ell \eta_\ell + 2\Lambda \sum_j I_j \delta_j + \Lambda \sum_v I_v \xi_v + \Lambda \sum_g \eta_g. \tag{61}$$

For the final Pauli measurement, Hoeffding's inequality gives, with probability at least $1 - \delta$,

$$|F_{\mathrm{imp}} - \widehat{F}_{N_{\mathrm{shot}}}| \le \Lambda \sqrt{2\log(2/\delta)/N_{\mathrm{shot}}}. \tag{62}$$

The triangle inequality applied to these three displays proves Equation 59. Taking the right-hand side of Equation 61 as ϵ_{imp} gives Proposition 4. The leaf, synthesis, composition, and channel budgets are deterministic in this statement. If any is itself certified only with failure probability δ_s, the same bound holds with total failure probability at most $\delta + \sum_s \delta_s$ by the union bound.

F.2 SYNTHESIS

For $U = U_L \cdots U_1$ and approximations satisfying $\|\widetilde{U}_j - U_j\|_\infty \le \delta_j$, a telescoping expansion gives $\|\widetilde{U} - U\|_\infty \le \sum_j \delta_j$. Therefore an observable of norm one has scaled output error at most

$$|F_{\mathrm{syn}} - F_{\mathrm{ideal}}| \le 2\Lambda \sum_j \delta_j. \tag{63}$$

The separation construction has $L_R = \Theta(d^2)$ QSP phase rotations. Uniform precision $\delta_R = \epsilon_{\mathrm{syn}}/(2\Lambda L_R)$ suffices. Using ancilla-free Clifford+T synthesis (Ross & Selinger, 2016), the phase cost is

$$T_{\mathrm{phase}} = O\left[d^2 \left(\log \frac{d^2}{\epsilon_{\mathrm{syn}}} + \log\log \frac{d^2}{\epsilon_{\mathrm{syn}}}\right)\right], \tag{64}$$

in addition to signal-oracle and composition-rule costs. If rotation j has root influence I_j, minimizing the leading logarithmic synthesis cost gives $\delta_j^\star = \epsilon_{\mathrm{syn}}/(2\Lambda L_R I_j)$.

F.3 SHOTS AND COHERENT ESTIMATION

A Pauli sample $Y \in \{-1, 1\}$ has mean F/Λ. For N_{shot} shots,

$$\mathrm{Var}(\Lambda \overline{Y}_{N_{\mathrm{shot}}}) = \frac{\Lambda^2 - F^2}{N_{\mathrm{shot}}} \le \frac{\Lambda^2}{N_{\mathrm{shot}}}. \tag{65}$$

Hoeffding's inequality (Hoeffding, 1963) gives the sufficient condition

$$N_{\mathrm{shot}} \ge 2\Lambda^2 \epsilon_{\mathrm{shot}}^{-2} \log(2/\delta). \tag{66}$$

The dependence is minimax-optimal for independent Bernoulli samples. Coherent amplitude estimation changes the precision dependence to $O(\Lambda/\epsilon)$ controlled queries, with a deeper fault-tolerant execution model (Brassard et al., 2002).

F.4 CHANNEL NOISE AND BLOCKING

If each implemented channel is within diamond distance η_g of its ideal counterpart, channel telescoping yields $|F_{\mathrm{noisy}} - F_{\mathrm{ideal}}| \le \Lambda \sum_g \eta_g$. A Pauli-diagonal channel that contracts the propagated Pauli by κ after each of $m-1$ multiplication gates gives

$$\langle Z_{\mathrm{out}} \rangle_{\mathrm{noisy}} = \kappa^{m-1} \prod_{j=1}^{m} z_j. \tag{67}$$

Debiasing multiplies variance by $\kappa^{-2(m-1)}$. Dividing the product into blocks of size b gives the bound $N_{\mathrm{total}}(b) = O(M^3/[\epsilon^2 \kappa^{2(b-1)}])$, $M = \lceil m/b \rceil$. For $\kappa = 1 - p$, continuous optimization predicts $b^\star \simeq 3/[-2\log(1-p)] \sim 3/(2p)$. This is a hardware limitation and a motivation for hybrid execution; experimental robustness remains a separate evaluation target.

G LEARNING FIXED FORMULA STRUCTURES

At deployment, structure and coefficients are learned classically, the selected formula is compiled deterministically, and the quantum circuit evaluates the result. For a fixed rank-R model

$$f_\Theta(\boldsymbol{x}) = \sum_{r=1}^{R} a_r \prod_{j=1}^{d} \sum_{k=0}^{p} c_{rjk} T_k(x_j), \tag{68}$$

the parameter count is $O(Rdp)$. Holding all dimensions except j fixed turns ridge regression into a convex block problem. With $h_{ir}^{(-j)} = a_r \prod_{\ell \ne j} g_{r\ell}(x_{i\ell})$ and $\Phi_{i,(r,k)}^{(j)} = h_{ir}^{(-j)} T_k(x_{ij})$, its solution is

$$\boldsymbol{c}_j^\star = \left[N^{-1}(\Phi^{(j)})^\top \Phi^{(j)} + \lambda I\right]^{-1} N^{-1}(\Phi^{(j)})^\top \boldsymbol{y}. \tag{69}$$

One sweep costs $O(d(NP_b^2 + P_b^3))$ for $P_b = R(p+1)$. This gives monotone block descent under exact updates. Global optimality for the joint nonconvex problem remains open.

For a dictionary of M unit-scale formula features $|\Phi_j| \le 1$ and outer weights $\|\boldsymbol{a}\|_1 \le A$, the empirical Rademacher complexity satisfies

$$\widehat{\mathfrak{R}}_N \le A\sqrt{2\log(2M)/N} \tag{70}$$

(Bartlett & Mendelson, 2002). The same A bounds the quantum normalization, so it simultaneously controls statistical capacity and independent-shot cost. When structure is unknown, the search objective should penalize empirical risk, formula size, depth, queries, and $\log \Lambda$. Finding a globally minimal formula remains an open combinatorial problem.

H Experimental Details and Additional Results

Table 1 records the information granted and the exact claim supported in each theoretical and experimental comparison.

Table 1: Unified comparison settings for theorems and experiments. Signed and modulus targets, seeded training, and deterministic oracle constructions remain separate.

Scope	Comparator output and matching information	Supported conclusion
Fourier-sum theorem/E3	Best n-term Fourier sum. Any integer frequencies and complex coefficients; F_d; normalized L^2 error $0 < \epsilon < \sqrt{3}/2$; declared local-query accounting for QSFN	Fourier-sum branch lower bound and query scaling
Native theorem/E3	Native SAQNN output $a\|G_n\|$. Any frequencies, coefficients, and scale; H_d^{nat}; normalized L^2 error $1/16$	Projector–modulus branch lower bound
E1	SAQNN-Reimplementation; SAQNN-Oracle. Same source task; resource claims use oracle pairs at matched error and one transpiler configuration	Finite-task fidelity and resources
E2	SAQNN-Flat; QSFN-Expanded. Same exact function and coefficients; no training or truncation; common O0 backend and expanded-branch fusion	Compression from preserved factorization

H.1 Environment and shared compiler settings

All experiments used Python 3.11.8, NumPy 1.26.4, SciPy 1.13.1, pandas 2.2.3, Qiskit 1.2.4 (Javadi-Abhari et al., 2024), Qiskit Aer GPU 0.15.1, and an NVIDIA RTX 4090. QSP experiments used pyqsp 0.2.0. Statevector workloads used the GPU backend. The transpiler basis was $\{\texttt{rz}, \texttt{sx}, \texttt{x}, \texttt{cx}\}$ with all-to-all connectivity. Both optimization levels 0 and 3 were evaluated; the main figures use level 0 to retain construction-level comparability. Five transpiler seeds were used in every experiment.

H.2 E1 protocol and primary results

The source-task reimplementation uses learning rates $0.01, 0.05, 0.05$ for f_1, f_2, f_3, respectively, 80 Adam epochs, and a finite-difference step selected from $\{10^{-2}, 10^{-3}, 10^{-4}\}$ on training-only pilot runs. Each of 20 seeds defines the same 100/100 train/test split for all compared methods. Training summaries use 10,000 paired bootstrap resamples. The dense grid is never used for fitting or degree selection.

The full-domain surfaces in Figure 7 show that the compiled QSFN constructions preserve the geometry of all three source tasks across the evaluation domain, complementing the dense-grid errors in Table 2.

Figure 9 shows the E1 training trajectories and finite-shot scaling checks. Compared with the original paper's reported test MSEs, the independent reimplementation medians are within factors 1.31 and 1.90 for f_1 and f_3; f_2 is lower by a factor of 44.6. The matched resource comparison uses the current independently constructed oracle rows.

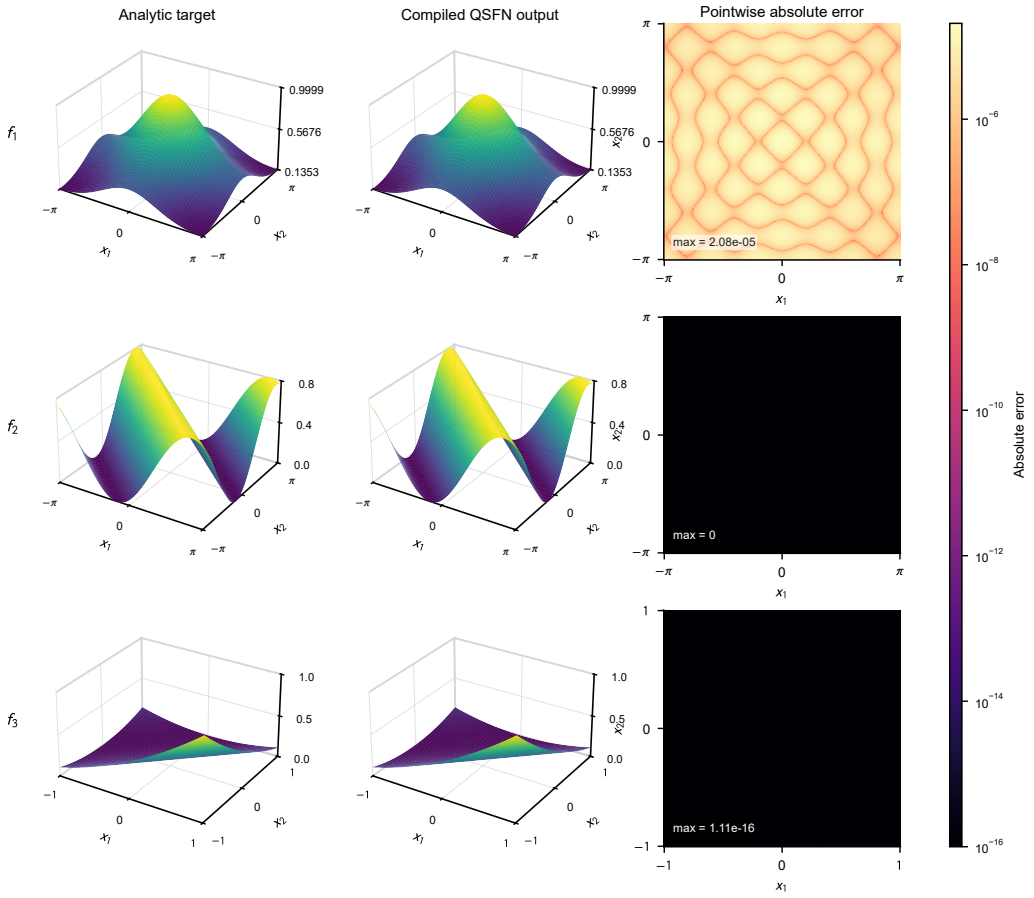


Figure 7: Approximation geometry on the three E1 source tasks. Rows correspond to f_1, f_2, and f_3; columns show the analytic target, the compiled QSFN output, and the pointwise absolute error on the 256×256 evaluation grid. For f_1, the displayed result uses the selected $p = 4$ higher-bandwidth operating point; f_2 and f_3 use the deterministic oracle constructions. Target and output surfaces share the same view and value scale within each row. The 10^{-16} floor is used only to display near-machine-precision errors on a logarithmic color scale.

Table 2: E1 primary dense-grid error and transpiled O0 resources. Deterministic oracle values are repeated across data seeds; trained values are medians over 20 runs. Q denotes logical qubits.

Task	Model	Grid MSE	Terms	Q_{sig}	Q	G_2	D_2
f_1	SAQNN-Reimplementation	1.028×10^{-3}	49	48	7	328	327
	SAQNN-Oracle	1.867×10^{-8}	49	48	7	328	326
	QSFN-Matched-$p3$	1.867×10^{-8}	8	10	14	81	31
	QSFN-Selected-$p4$	4.645×10^{-11}	10	14	14	97	35
f_2	SAQNN-Reimplementation	5.957×10^{-6}	3	2	3	14	14
	SAQNN-Oracle	5.993×10^{-33}	3	2	3	14	14
	QSFN-Oracle	0	2	2	5	16	16
f_3	SAQNN-Reimplementation	9.111×10^{-4}	6	6	5	182	181
	SAQNN-Oracle	2.627×10^{-33}	6	6	5	206	205
	QSFN-Oracle	8.943×10^{-34}	6	4	10	33	17

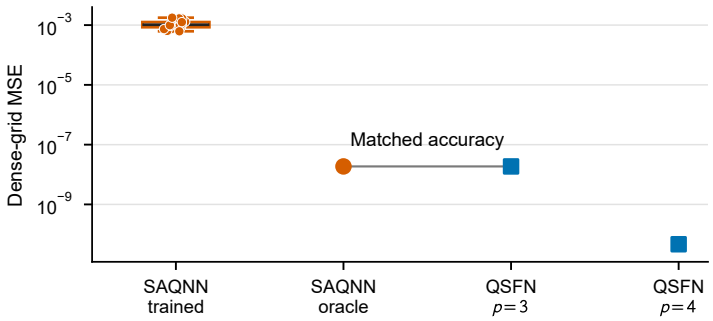


Figure 8: Dense-grid accuracy on f_1. All 20 trained SAQNN runs are shown alongside deterministic oracle constructions. The matched $p = 3$ points coincide; the selected $p = 4$ construction supplies a higher-accuracy operating point.

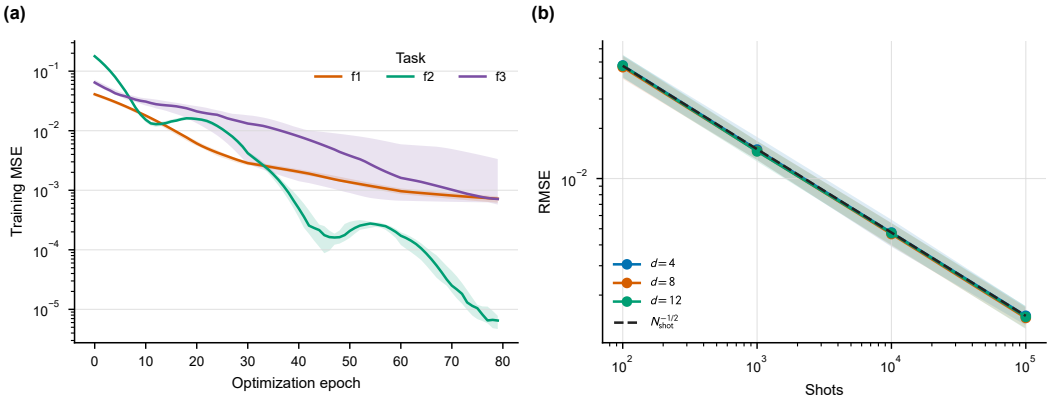


Figure 9: (a) Optimization curves over 20 genuine E1 training seeds. (b) Finite-shot RMSE over 100 repeated estimates for each point follows the $N_{\text{shot}}^{-1/2}$ reference. Bands show empirical intervals.

H.3 E2 PROTOCOL AND SCALING TABLE

Dimensions are $d \in \{2, 4, 6, 8, 10, 12, 16, 24, 32\}$. All three circuits are generated and fully transpiled through $d = 8$. For larger dimensions, SAQNN-Flat and QSFN-Expanded use exact recurrences for the reported non-depth resources, validated on the compiled subset; depth is reported only for explicitly compiled circuits. QSFN-Factorized remains explicitly compiled. Function-equivalence tests use 10,000 Sobol points with formula and circuit tolerances 10^{-12} and 10^{-10}.

The fitted flat-resource slope is $\log_2 3$, and factorized formula size has log–log slope 1. At $d = 8$, representation and two-qubit-gate compression are 211.6 and 346.1; at $d = 32$, the corresponding factors are 1.46×10^{13} and 2.35×10^{13}, using exact flat-resource recurrences (Table 3).

Figure 10 compares exact-contraction and statevector outputs for the E2 implementations.

H.4 E3 PROTOCOL AND CERTIFICATES

The tested dimensions are $\{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 24, 32\}$. Full circuits are transpiled through $d = 8$; higher-dimensional G_2 values use constructive component accounting, and high-dimensional depth is not labeled as measured. The Jackson order is the smallest order passing the prespecified energy condition; actual nonzero degree and phase-sequence calls are reported. QSP validation uses a 100,001-point grid and tolerance 2×10^{-9}. Exact best-n status is assigned only

Table 3: Selected E2 scaling values. Flat high-dimensional gate counts are exact recurrences; factorized values are transpiled measurements. Representation means complex branches for SAQNN-Flat and total formula nodes for the two QSFN rows; QSFN-Expanded contains $(3^d + 1)/2$ real spectral leaves. Dashes mark cases without an explicitly compiled depth.

d	Model	Representation	Q	G_2	D_2	Max. error
4	SAQNN-Flat	81	8	380	380	5.6×10^{-16}
4	QSFN-Expanded	81	7	126	126	5.6×10^{-16}
4	QSFN-Factorized	15	12	35	10	4.4×10^{-16}
8	SAQNN-Flat	6,561	14	24,572	24,572	4.4×10^{-16}
8	QSFN-Expanded	6,561	13	8,190	8,190	4.4×10^{-16}
8	QSFN-Factorized	31	24	71	11	5.6×10^{-16}
32	SAQNN-Flat	1.853×10^{15}	52	6.755×10^{15}	–	–
32	QSFN-Expanded	1.853×10^{15}	51	2.252×10^{15}	–	–
32	QSFN-Factorized	127	96	287	13	–

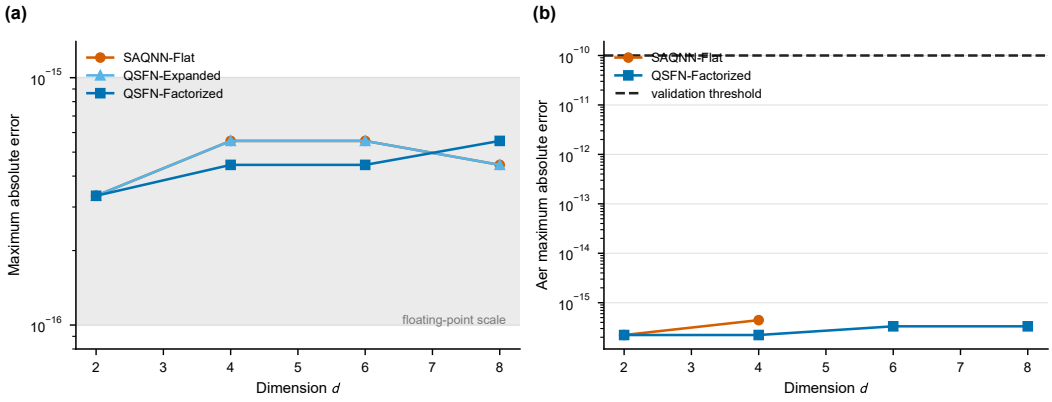


Figure 10: Independent exact-contraction and Aer statevector validation of the E2 implementations. All errors remain below the specified thresholds.

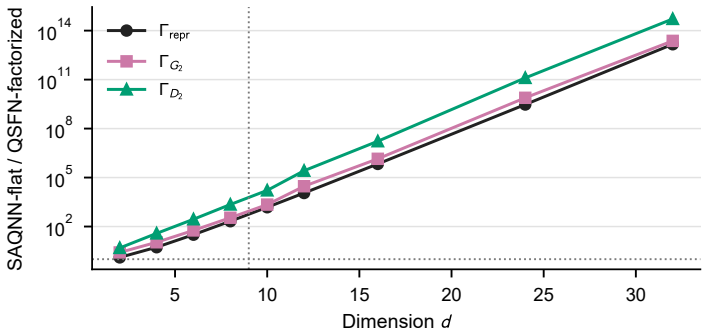


Figure 11: E2 flat-to-factorized compression in representation size, two-qubit gates, and depth. Beyond $d = 8$, representation and gate ratios use validated recurrences; the depth ratio uses the flat serial-depth upper bound and the transpiled factorized depth. The vertical guide separates the compiled flat-circuit range from the recurrence-based range.

when enumeration reaches the target energy; truncated runs retain certified lower bounds. Finite-shot tests use 256 Sobol points, 100 repetitions, and $N \in \{10^2, 10^3, 10^4, 10^5\}$.

The dimension-uniform analytic choice is $n_d = 64d$. For the selected orders, finite-size fits give exponents 2.089 for signal calls and 1.089 for leaf degree, consistent with the quadratic and linear asymptotic orders (Table 4).

Table 4: Selected E3 certificates. Exact $n_{\min}$ is shown only where the enumeration completes. The native column is the integer certified lower bound for error $1/16$.

d	p_d	$\|F_d\|_2^2$	Signed n_{LB}	Exact $n_{\min}$	Q_{sig}	G_2	D_2	Native n_{LB}
4	17	0.7796	8	44	68	293	76	1
8	37	0.7842	128	11,132	296	1,227	156	3
12	55	0.7789	2,048	3,842,736	660	2,705	–	5
16	75	0.7813	32,768	–	1,200	4,887	–	13
24	113	0.7802	8,388,608	–	2,712	10,979	–	74
32	151	0.7797	2,147,483,648	–	4,832	19,503	–	445

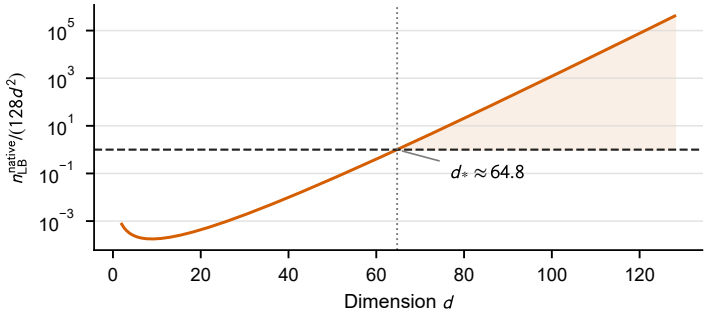


Figure 12: Ratio of the native-modulus branch lower bound to the uniform $128d^2$ QSFN query upper bound at error $1/16$. The analytic bounds cross at $d \approx 64.8$; the shaded region identifies the subsequent separation. These curves extend the proved bounds beyond the evaluated dimensions.

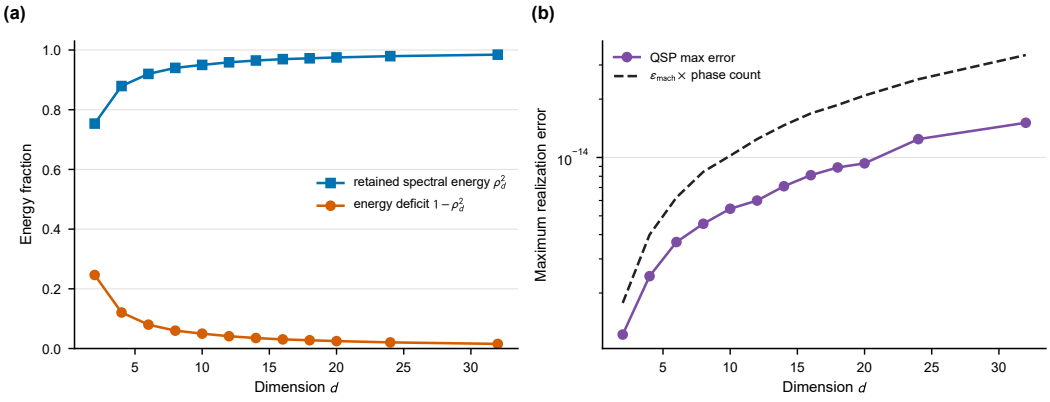


Figure 13: (a) Computed Jackson-leaf energy and energy deficit. (b) Dense-grid QSP semantic residuals and phase counts. The machine-epsilon line is a numerical reference.

The finite-shot fitted slopes are -0.5001, -0.5009, and -0.5008 for $d = 4, 8, 12$; their bootstrap intervals all contain $-1/2$. Figure 13 shows the Jackson-leaf energy and QSP residual checks. The accompanying evaluations verify minimal Jackson order, polynomial boundedness, energy and coefficient invariants, exact/top-K labels, analytic and statevector agreement, constructive resource recurrences, native-interface transfer and correlation margins, and shot scaling.