ENCRYPTABILITY AS A COORDINATE CHOICE: DEPTH-ONE HOMOMORPHIC FEDERATED LEARNING OF QUANTUM NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Encrypted training relies on keeping server-side updates low-degree. This constraint traditionally excludes models whose weights inhabit a compact Lie group (notably variational quantum circuits, where every trainable weight is an $\mathrm{SU}(2)$ rotation). Expressed in Euler angles or discrete alphabets, these updates appear transcendental, historically demanding prohibitive costs: one client–server round per gate, or upwards of $25{,}000$ operations per weight. This penalty is strictly an artefact of coordinates. In the unit-quaternion (spin) chart, group composition is exactly bilinear (degree two, with coefficients in $\{-1, 0, +1\}$). Consequently, encrypted rotation updates cost one multiplicative level and federated averaging costs zero in *any* levelled homomorphic scheme, completely eliminating bootstrapping. This implementation-independent algebraic property is confirmed across two cryptographic backends, introducing only 0.0 and -2.0×10^{-12} rad of aggregation error. Leveraging this reduction yields a non-interactive protocol for encrypted federated training of hybrid quantum–classical networks. It includes correctness proofs for aggregation and sign handling, plus a compilation lemma proving parameterised entanglers add only constant-factor overhead without altering the depth class. Empirically, a paired five-seed study confirms zero measurable utility tax ($\Delta = +9 \times 10^{-6}$ MSE, $p = 0.92$), and a noise-budget ablation falsifies the hypothesis that encryption noise regularises. These convergence trends replicate across datasets and scale to 20 clients. Finally, hardware validation on a 156-qubit processor achieves 0.9918 fidelity against a 0.99957 unencrypted control.

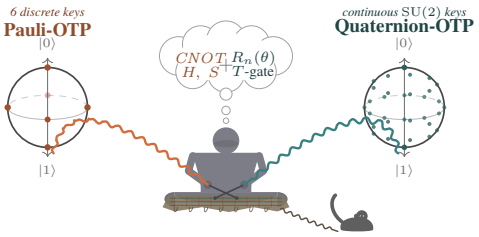


Figure 1: **The hybrid encryption scheme as a half-finished blanket.** An exact, constant-depth compilation between disparate algebraic bases (both necessary for universal quantum computation) is proposed to encrypt federated learning. It resolves the continuous-rotation bottleneck without invoking prohibitive bootstrapping. The two yarns represent two encryption schemes for $\mathrm{SU}(2)$ used in compilation. The Pauli One-Time-Pad (orange), using discrete keys at the six Bloch lattice points, handles Clifford gates ($CNOT, H, S$) via a single XOR update but stalls on continuous rotations. Conversely, the Quaternion-OTP (teal), using continuous keys over the Bloch sphere, evaluates arbitrary rotations $R_n(\theta)$ via degree-2 polynomial arithmetic (Proposition 1) but cannot factorise across $CNOT$. By weaving both tracks, the protocol completes the circuit: continuous rotations are encrypted by quaternions and Cliffords by brief Pauli-track detours. The cat plays no formal role.

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.

1 INTRODUCTION

Encrypted federated learning provides a natural way to train shared models whilst protecting client privacy, and efficient implementations are rapidly emerging. Homomorphic FedAvg is practical because linear weighted averaging avoids multiplicative depth (Zhang et al., 2020); encrypted inference succeeds because low-degree polynomial activations eliminate the need for bootstrapping (Gilad-Bachrach et al., 2016) As a general principle, the reach of homomorphic encryption in ML is set by one question: *is the server-side parameter update a low-degree polynomial in the parameters?* When this is the case, encryption largely amounts to bookkeeping; when it is not, general-purpose methods introduce cumbersome computational overhead.

Undermining generality, that question has been answered architecture by architecture only, and the field now treats low-degree representability as a property of *layer types*: linear yes, ReLU no-but-approximable, attention expensive. That premise becomes harder to sustain for a growing class of models because their parameters are not free vectors in $\mathbb{R}^d$ at all. They are elements of a compact Lie group: orthogonal and unitary recurrent networks (Arjovsky et al., 2016; Lezcano-Casado & Martínez-Rubio, 2019), rotation-equivariant architectures (Thomas et al., 2018), quaternion-valued networks (Parcollet et al., 2020) and – where the group structure is a consequence of the underlying physics – variational quantum circuits. Their every trainable weight is a rotation in $\mathrm{SU}(2)$ (Benedetti et al., 2019; Cerezo et al., 2021).

The final complication for group-valued setting is that the degree is coordinate-dependent: the same update may be transcendental in one chart and bilinear in another. An architecture-first analysis misses this distinction, asking about efficiency in the wrong coordinate system. For example, the Pauli one-time pad (Broadbent & Jeffery, 2015), the standard encryption for quantum states, is closed under the discrete Clifford group but not under continuous rotations: applying $R(\theta)$ to a padded state leaves a phase error conditioned on the secret key.

Both accepted repairs for this phase error are expensive. Measurement-based blind delegation (Broadbent et al., 2009; Fitzsimons, 2016) restores security at one client–server round *per gate*, $O(T)$ in circuit depth T; Solovay–Kitaev compilation (Dawson & Nielsen, 2005) restores it by replacing each continuous weight with a discrete-alphabet sequence, typically above $25{,}000$ gates per rotation (Ma & Li, 2022) where each non-Clifford element demands an active key repair.

Ultimately, these cryptographic penalties translate directly into bottlenecks. An entire model family is excluded from the encrypted federated stack the rest of ML already uses: privacy-preserving quantum federated learning is built either on plaintext updates, which gradient-inversion attacks make indefensible (Zhu et al., 2019; Geiping et al., 2020), or on interactive delegation, which does not compose with classical secure aggregation (Bonawitz et al., 2017). The remaining remedies force trade-offs across vulnerable dimensions – utility for differential privacy in small-data regimes (Abadi et al., 2016; Dwork & Roth, 2014), communication for multi-party computation where round-trips bind (Kairouz et al., 2021). Moreover, the overheads are large enough that the protocols are never run end-to-end, leaving a literature of cost models without measurements. Ultimately, the most profound cost is conceptual: so long as encryptability is framed purely as an architectural constraint, the field remains blind to alternative solutions.

This paper deems the solution to be algebraic, as in Fig. 1. $\mathrm{SU}(2)$ is isomorphic to the unit quaternions, and under that identification group composition is quaternion multiplication, which is *bilinear*: composition is a polynomial map of total degree exactly two on $\mathbb{R}^4$ with coefficients in $\{-1, 0, +1\}$ (Proposition 1). This shift in coordinate system perfectly satisfies the low-degree requirement for the sought efficient homomorphic encryption. Three consequences form the backbone of this work: an encrypted rotation update consumes one multiplicative level; encrypted weighted averaging (the FedAvg primitive) consumes zero; and critically, neither needs bootstrapping. By bypassing the need for thousands of discrete gates or active key repairs, the encrypted quantum training round is *shallower* in multiplicative depth than the encrypted classical MLP round it is usually compared against. Two observations make this more than a mathematical trick. The proposed encryption is scheme-agnostic, since degree two is a property of the algebra rather than of CKKS (Cheon et al., 2017) (verified by running the same aggregation on two independent backends in §5.2). Also, the encryption generalises, since the same bilinearity holds for rotors in the even Clifford subalgebra, meaning any $\mathrm{Spin}(n)$-parameterised layer inherits the bound (Corollary 1).

Contributions.

- **An algebraic depth ledger for group-valued parameters** (§4): Since $\mathrm{SU}(2)$ composition is degree-2 in spin coordinates, an encrypted rotation update consumes just one multiplicative level, and encrypted FedAvg consumes zero in *any* levelled scheme (also in $\mathrm{Spin}(n)$ rotors).
- **A practical protocol and empirical evaluation** (§4.2–§5): Communication overhead is reduced to a single exchange per round (down from 37–73). Furthermore, a paired experimental design with a matched unencrypted hardware control demonstrates no measurable utility tax, falsifying confounding regularisation effects that a single-seed run would obscure. This yields the first known public implementation of continuous-rotation quantum homomorphic encryption, validated across multiple datasets and scaled to 20 clients.
- **Correctness of the encrypted federated loop** (§4.3–§4.4): Aggregation correctness is established via a mask-placement lemma that separates shared from fresh randomness, alongside the derivation of the exact analytic discrepancy, $-\frac{1}{24}\sum_k w_k(\theta_k - \bar{\theta})^3 + O(\delta^5)$, between quaternion and Euclidean averaging (validated empirically to $\geqslant 0.9991$).
- **Constant-overhead entangler compilation** (§4.5): Parameterised two-qubit gates compile exactly onto CNOTs and single-qubit rotations, preserving the multiplicative depth class. This is verified end-to-end at 8.3×10^{-3} rad against a 1.25×10^{-2} rad shot-noise floor.

2 SETTING: LEAKAGE

To support the algebraic reduction detailed in §4, the relevant quantum gates necessary for universal computation are discussed (more in the primer in Appendix A). A universal gateset splits into two categories: the discrete Clifford group (the normaliser of the Pauli group, containing gates such as H, S, and CNOT) and continuous non-Clifford operations. A hybrid quantum–classical model interleaves a parameterised quantum circuit with small classical layers, where the trainable quantum parameters are exclusively these continuous, non-Clifford single-qubit rotation angles (elements of $\mathrm{SU}(2)$). Gradients for these weights are evaluated via the parameter-shift rule (Mitarai et al., 2018; Schuld et al., 2019) or simulator backpropagation. While Clifford gates are readily managed by standard quantum encryption, continuous generic rotations are not, creating the algebraic bottleneck that the quaternion encryption resolves.

Federated training of such a model has *two* distinct exposure points. **(i) Delegated execution:** quantum hardware is remote, so a client must submit its circuit, and the rotation angles – the weights – cross the trust boundary before any aggregation. **(ii) Federated aggregation:** the server that averages updates sees them, and gradient-inversion attacks reconstruct training data from exactly such transmissions (Zhu et al., 2019; Geiping et al., 2020). Both are targeted.

An honest-but-curious server and peers and an active network adversary on authenticated channels are assumed. The cloud QPU sees only quaternion-masked states; the server sees ciphertexts, message sizes and the public weights n_k/n, holding an evaluation-only context in which decryption raises an error; peers see the decrypted global aggregate and nothing further by protocol (Table 10, Appendix J). Confidentiality of the classical layer rests on the hardness of (Ring-)LWE (Regev, 2009; Lyubashevsky et al., 2010); the quantum layer's rests on physics – averaged over the pad key, the delegated state is maximally mixed, so the guarantee is information-theoretic and immune to unbounded compute, including a future large quantum computer (Shor, 1997).

3 RELATED WORK

Low-degree representations for encrypted ML. The design pattern is well established classically: CryptoNets replaces activations with low-degree polynomials to enable levelled schemes (Gilad-Bachrach et al., 2016), homomorphic FedAvg exploits the linearity of averaging (Zhang et al., 2020; Cheon et al., 2017), secure aggregation relies on masking (Bonawitz et al., 2017; Shamir, 1979), and differential privacy trades utility for security (Abadi et al., 2016; Dwork & Roth, 2014). The pattern is applied here one level deeper (to the *parameter group* rather than the nonlinearity), exploiting a group law that is intrinsically low-degree and requires no approximation.

Group-parameterised models. Unitary and orthogonal recurrent networks (Arjovsky et al., 2016; Lezcano-Casado & Martínez-Rubio, 2019), rotation-equivariant architectures (Thomas et al., 2018), and quaternion networks (Parcollet et al., 2020) constrain weights to a compact group for optimisation or inductive-bias reasons. Corollary 1 demonstrates this is incidentally an encryption-friendly choice, since any $\mathrm{Spin}(n)$-parameterised layer inherits the depth-one bound (previously unexplored).

Quantum homomorphic encryption and quantum FL. Blind quantum computing provides information-theoretic delegation at $O(T)$ interaction (Broadbent et al., 2009; Fitzsimons, 2016). Pauli-OTP is non-interactive on Cliffords but stalls on continuous rotations (Broadbent & Jeffery, 2015), whilst Mahadev's encrypted-CNOT enables a fully classical client under LWE (Mahadev, 2023). Closest to the present approach, Ma & Li (2022) introduce the quaternion representation for single-client delegated computation. The federated setting, however, shifts the requirement from isolated delegation to a continuous training loop. This is resolved here by formalising the depth ledger for that loop, establishing the correctness and geometry of aggregating group-valued updates, detailing the entangler compilation, and providing a hardware-validated implementation. Conversely, existing quantum FL frameworks either exchange plaintext updates (Chen & Yoo, 2021; Chehimi & Saad, 2022; Ballester et al., 2025) or rely on interactive delegation that cannot compose with classical secure aggregation (Li & Deng, 2025; Dutta et al., 2024); robustness to malicious clients is treated separately (El Maouaki et al., 2025).

4 Encryptability is a coordinate choice

4.1 The degree-2 reduction and the depth ledger

To efficiently encrypt the continuous non-Clifford $\mathrm{SU}(2)$ rotations identified in §2, a coordinate representation q evaluating as a low-degree polynomial is needed. Let $\mathbb{H}_1 = \{q \in \mathbb{R}^4 : \|q\|_2 = 1\}$ be the unit quaternions with the Hamilton product

$$(q \otimes q')_i \;=\; \sum_{j,k} M^{(i)}_{jk}\, q_j\, q'_k, \qquad M^{(i)}_{jk} \in \{-1, 0, +1\}, \quad i, j, k \in \{0, 1, 2, 3\}, \tag{1}$$

and let $\Phi : \mathbb{H}_1 \to \mathrm{SU}(2)$ be the standard double cover, so that a rotation $U(\theta, \hat{u}) = \cos(\theta/2)I - i\sin(\theta/2)(u_x X + u_y Y + u_z Z)$ is encoded by

$$q(\theta, \hat{u}) = \big(\cos\tfrac{\theta}{2},\, -u_x \sin\tfrac{\theta}{2},\, -u_y \sin\tfrac{\theta}{2},\, -u_z \sin\tfrac{\theta}{2}\big), \qquad \|q\|_2 = 1. \tag{2}$$

Equation 1 is the whole argument: the group law is *bilinear*, hence of total degree two, hence within reach of a levelled homomorphic scheme at a single multiplication.

Proposition 1 (Depth ledger for encrypted rotation update)**.** *Let $\mathcal{E}$ be any levelled homomorphic scheme supporting ciphertext addition at depth 0, plaintext–ciphertext multiplication at depth 1, and $\mathbb{R}$-linear combination with plaintext scalars at depth 0. Let $q \in \mathbb{H}_1$ encode a rotation $\Phi(q)$ and let $V \in \mathrm{SU}(2)$ have quaternion representation $r \in \mathbb{H}_1$. Then* (i) *$q' = q \otimes r$, representing $\Phi(q)\Phi(r)$, is a polynomial of total degree two in (q, r) with coefficients in $\{-1, 0, +1\}$;* (ii) *for plaintext r and encrypted q, $\mathrm{Enc}(q')$ is computable from $\mathrm{Enc}(q)$ and r by plaintext–ciphertext multiplications and constant-coefficient additions alone, consuming exactly one multiplicative level; and* (iii) *the weighted mean $\bar{q} = \sum_{k=1}^{K} \frac{n_k}{n} q^{(k)}$ of K ciphertexts is computable at depth 0, with the renormalisation $\bar{q}/\|\bar{q}\|$ performed client-side in plaintext after decryption and hence also at depth 0. Consequently a federated round of one server-side rotation update per encrypted key followed by one weighted aggregation fits a single multiplicative level, and no bootstrapping is required in the inner loop.*

Appendix B gives the proof and Corollary 2 the counts on unbounded round. Two remarks elucidate the value of these statement to a broader non-quantum community.

Corollary 1 (Spin generalisation)**.** *The spin group $\mathrm{Spin}(n)$ is the double cover of the n-dimensional rotation group $\mathrm{SO}(n)$, generalising the relationship between unit quaternions and 3D rotations. Let $\mathrm{Spin}(n)$ act by rotors in the even subalgebra $\mathrm{Cl}^{[0]}(n)$ (Lounesto, 2001). Because the geometric product is bilinear, composition of rotors is a degree-2 polynomial map on the 2^{n-1} even-graded coordinates, so every statement of Proposition 1 holds verbatim for $\mathrm{Spin}(n)$-parameterised layers, with $\mathbb{H}_1 \cong \mathrm{Spin}(3)$ being the exact case $n = 3$.*

The per-round accounting is as follows. The rotation update consumes one multiplicative level and requires no interaction; weighted FedAvg and the Clifford key update incur zero multiplicative cost. Meanwhile, every branch, sign, renormalisation, and recovery step remains entirely free by running client-side in plaintext. Because clients train locally in plaintext and encrypt only the resulting angles, the per-round homomorphic work reduces to a single depth-0 aggregation, *whatever the ansatz contains*. Furthermore, parameter-shift gradients scale only the local plaintext quantum work, leaving the encrypted workload entirely unaffected.

4.2 Protocol

Figure 2 shows one federated round; the numbered walk-through below is the definition, and every symbol in the figure is defined here rather than in its caption.

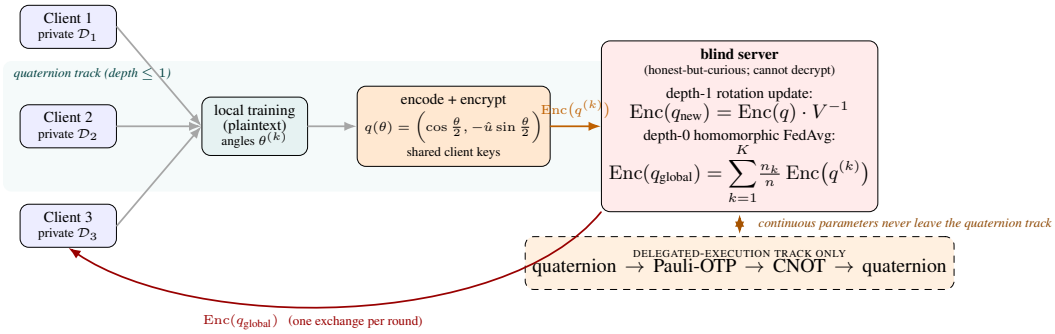


Figure 2: One federated round. Clients train in the clear, encode each trainable angle as a unit quaternion (Eq. 2) and encrypt its four components under keys shared among clients. The server applies plaintext rotations at one multiplicative level and the weighted FedAvg sum at zero, never decrypting; clients decrypt, renormalise and recover angles by $\mathrm{atan2}$. The Pauli-OTP detour is entered only on the delegated execution track (Lemma 2).

Protocol (one federated round t).

1. **Local training** (client, plaintext). Train on private $\mathcal{D}_k$ to obtain angles $\theta^{(k)}$; circuit evaluation is local, the delegated track of §4.5 is used only when offloading to a remote QPU.
2. **Fix branch and sign, encode, encrypt** (client, depth 0). Wrap to $(-\pi, \pi]$, align the hemisphere representative with the previous global quaternion (§4.4), form $q\big(\theta^{(k)}\big)$ by Eq. 2, encrypt its four components under the clients' shared keys, upload.
3. **Aggregate** (server, depth 0). Compute $\mathrm{Enc}(q_{\text{global}}) = \sum_k \frac{n_k}{n} \mathrm{Enc}\big(q^{(k)}\big)$ by ciphertext addition and plaintext scaling, optionally applying a plaintext rotation V^{-1} at depth 1. The server holds an evaluation-only context and cannot decrypt.
4. **Decrypt and recover** (client, plaintext, depth 0). Renormalise, recover $\hat{\theta} = 2\,\mathrm{atan2}(-t_1, t_0)$, re-wrap, begin round $t{+}1$.

Exactly one sequential exchange occurs per round, independent of circuit depth, ansatz, and the number of parameter-shift evaluations used locally.

The compatibility of the two tracks is a closure argument. Pauli-OTP is closed under Clifford conjugation but not under continuous rotation, which is precisely the classical obstruction; the quaternion track removes this problem because composition is degree 2; the two meet only at frame conversions, a constant number per entangler, never sharing arithmetic on the payload; and FedAvg, being linear, commutes with the homomorphic layer at zero depth. Appendices E and F give the conversion machinery and the full construction.

4.3 Correctness of the encrypted aggregation

The correctness of the reduction can be established by considering aggregation and mask-sharing.

Aggregation. The object summed by the server is the *unmasked* update quaternion, encrypted under homomorphic-layer keys shared by the clients: client k encodes each trainable angle by Eq. 2 and encrypts the four components, the server forms $\sum_k n_k \mathrm{Enc}\big(q^{(k)}\big)$ using ciphertext additions and plaintext-integer weightings alone – no ciphertext–ciphertext multiplication, no decryption – and clients decrypt, divide by n, renormalise and recover $\hat{\theta} = 2\,\mathrm{atan2}(-t_1, t_0)$. Correctness is inherited from linearity of the homomorphic layer exactly as in classical HE-FedAvg (Zhang et al., 2020). Quaternion one-time-pad keys are a distinct object living only on the QPU delegation channel; they are never averaged (Appendix C).

Mask sharing. The following divides the labour, while Appendix C gives the proof and the corresponding statement for the delegation channel.

Lemma 1 (Mask placement)**.** *Let $L_u(q) = u \otimes q$ be left multiplication by a unit quaternion u. (i) A* common *left mask commutes with weighted averaging, $\sum_k w_k L_u(q_k) = L_u\big(\sum_k w_k q_k\big)$, so it is stripped after decryption at zero depth. (ii) A* quantum-channel *mask must instead be per-client and freshly sampled: a shared unitary U preserves overlaps, $\mathrm{Tr}[(U\rho_i U^\dagger)(U\rho_j U^\dagger)] = \mathrm{Tr}(\rho_i\rho_j)$, leaking the pairwise geometry of client states, whereas independent Pauli keys twirl any pair to $\frac{I}{d} \otimes \frac{I}{d}$.*

4.4 Well-definedness: branch, sign, and which mean FedAvg computes

Quaternions double-cover $\mathrm{SU}(2)$, meaning q and $-q$ denote the same rotation, and angles wrap at $\pm\pi$, so any protocol averaging these representatives must specify both the chosen representative and the target mean. Both are resolved client-side in plaintext, incurring zero depth and zero communication. Specifically, angles are wrapped to the principal branch, $\theta \leftarrow \theta - 2\pi\lfloor(\theta+\pi)/2\pi\rfloor$, and the hemisphere representative satisfying $\langle q_k, \bar{q}_{\text{prev}}\rangle \geqslant 0$ is selected. The subsequent recovery via $\mathrm{atan2}$ is sign-exact, and renormalisation is scale-invariant. This leaves one discrepancy: the renormalised quaternion mean constitutes the chordal mean on the sphere (Markley et al., 2007), whereas standard FedAvg averages angles directly.

Given a fixed axis, weights $w_k \geqslant 0$ summing to one, a standard mean $\bar{\theta} = \sum_k w_k\theta_k$, and deviations $\delta_k = \theta_k - \bar{\theta}$, the angle $\hat{\theta}$ recovered from the renormalised weighted quaternion mean satisfies the following relation. Provided all representatives share a hemisphere,

$$\hat{\theta} - \bar{\theta} \;=\; -\frac{1}{24}\sum_k w_k \delta_k^3 \;+\; O(\delta^5) \tag{3}$$

(Proposition 2, proved in Appendix D). This bound fails by exactly π for configurations straddling the branch cut, a scenario excluded by construction via the aforementioned wrapping step.

The factor $\frac{1}{24}$ rather than the naive $\frac{1}{6}$ comes from the encoding: the protocol averages quaternion components at $\theta/2$. This formulation yields a discrepancy that can be computed *a priori* directly from observed client disagreement, and it identifies near-antipodal disagreement as the singular configuration in which quaternion FedAvg and standard angle FedAvg diverge (measured in §5.4).

4.5 Entanglers cost's dependence on gates and independence of depth

A reasonable objection to any rotation-only reduction is that modern ansatzes depend on *parameterised* entanglement. It dissolves under compilation.

Lemma 2 (Constant-overhead entangler compilation)**.** *Every controlled single-qubit rotation decomposes exactly into at most two CNOTs plus single-qubit rotations (Barenco et al., 1995), and every two-qubit unitary into at most three (Vatan & Williams, 2004); in particular $R_{ZZ}(\theta) = \mathrm{CNOT}\cdot(I \otimes R_Z(\theta))\cdot\mathrm{CNOT}$. The continuous parameter therefore always lands on a single-qubit rotation, costing one level by Proposition 1(ii), while the CNOTs are fixed Clifford elements whose key update is the depth-0 XOR relabelling $(a_1, b_1, a_2, b_2) \mapsto (a_1,\, b_1 \oplus b_2,\, a_1 \oplus a_2,\, b_2)$, and commutation through the pad, $R_Z(\theta)X^a = X^a R_Z((-1)^a\theta)$, makes only the* sign *of the angle key-conditioned. Parameterised entanglers thus multiply gate count by a constant factor and leave the multiplicative-depth class unchanged. (Proof in Appendix C.)*

Much like lowering convolutions to matrix multiplication, the operation compiles exactly onto existing kernels at constant overhead, avoiding the need for custom extensions. §5.5 validates this empirically.

5 EXPERIMENTS

5.1 SETUP AND CLAIMS' SCOPE

Mechanism-level claims – the depth ledger, non-interactivity, backend independence, exact entangler compilation, the aggregation-geometry bound – are properties of the algebra and the protocol and are expected to generalise. **Run-level** claims – every MSE value – characterise a small hybrid model in one configuration, and are reported as feasibility evidence with error bars. The task is California Housing regression (Pace & Barry, 1997), with Wine Quality (Cortez et al., 2009) as an independent replication; the model is a six-qubit, six-layer variational circuit (36 trainable angles, a CNOT ring after each rotation layer) feeding a small classical head, evaluated by exact statevector simulation with parameter-shift gradients (Qiskit contributors, 2023; Mitarai et al., 2018). Encryption uses the LWE-based fixed-point multi-key scheme of Ma & Li (2022) at 12-bit precision unless stated, cross-checked against CKKS over OpenFHE (Al Badawi et al., 2022); hardware runs use `ibm_fez` (IBMQuantum, 2024). Configurations, seeds and job provenance are in Appendix G.

5.2 THE DEPTH LEDGER'S EXACTNESS AND UNIVERSALITY

Backend	Single client	Post-FedAvg	Overhead
Multi-key (12-bit)	1.56×10^{-4}	1.56×10^{-4}	$\mathbf{0.0}$
OpenFHE CKKS	4.76×10^{-12}	2.75×10^{-12}	$\mathbf{-2.0 \times 10^{-12}}$

Table 1: Depth-0 aggregation exactness (float64 decode, 3 clients). Overhead isolates single-client vs. post-FedAvg error.

Arm	Final MSE	Paired Δ vs. Plain (95% CI)
Plaintext	1.3436 ± 0.72	—
Encrypted (12-bit)	1.3436 ± 0.72	$+9.0 \times 10^{-6}$ $[-2.3, +2.5] \times 10^{-4}$
Encrypted (8-bit)	1.3442 ± 0.72	$+6.0 \times 10^{-4}$ $[-2.6, +3.8] \times 10^{-3}$

Table 2: Utility under encryption (5 paired seeds, 3 clients). Maximum round-0 loss gap is exactly 0.00 across all seeds.

Proposition 1(iii) predicts that weighted aggregation is not merely cheap but *error-free relative to per-client encoding*, in any levelled scheme. This prediction is evaluated via a single measurement – three clients, 36 angles each: encode, aggregate homomorphically, decrypt, compare – executed on two cryptographically distinct backends, decoding in float64 so no cast floor masks the result. Aggregation overhead is 0.0,rad on the fixed-point multi-key stack (single-client and post-FedAvg angle MAE both 1.5632×10^{-4}) and -2.0×10^{-12},rad on OpenFHE CKKS, with the latter representing independent rounding errors averaging down rather than a systematic effect (Table 1). The two stacks differ by eight orders of magnitude in their inherent encoding error yet agree exactly on the algebraic depth-1 and depth-0 properties. Finally, the same run demonstrates that the noise budget functions as a dial – each precision bit roughly halves the angle error, spanning a $60\times$ range (Table 6) – establishing the 12-bit operating point at 1.52×10^{-4} rad as a deliberate set-point.

5.3 UTILITY AS A FUNCTION OF ENCRYPTION AND NOISE

The practical question is simpler: what effect does switching encryption on have on the loss? Answering this requires a paired design, as the between-seed spread of the final MSE is ≈ 0.72 – three to five orders of magnitude larger than any plausible encryption effect. Data loaders are reseeded and rebuilt immediately before each arm, isolating the encryption round trip as the *only* variable. In experiments, arms starting identically must yield equal round-0 loss, and the maximum round-0 gap across five seeds is exactly 0. The result is parity – plaintext 1.3436 ± 0.7228 against encrypted 1.3436 ± 0.7229, with a mean paired difference of $+9 \times 10^{-6}$ MSE (95% CI $[-2.3, +2.5] \times 10^{-4}$, $p = 0.923$ paired t, and $p = 0.625$ Wilcoxon; Table 2).

A single-seed comparison in this configuration would have shown the encrypted arm ahead, and the available explanation – that bounded approximation noise acts as a stochastic regulariser on a rugged variational landscape (McClean et al., 2018) – is superficially compelling. However, it makes a falsifiable prediction: increased noise should yield greater improvements. To test this, the encrypted arm was re-run at 8-bit precision (a $16\times$ larger quantisation step, identical seeds). The noisier arm offers no improvement ($\Delta = +0.0006$ MSE, $p = 0.636$). Thus, encryption incurs no utility tax.

5.4 THE AGGREGATION-GEOMETRY BOUND

Proposition 2 is evaluated by scattering five clients' 36 angles around a base vector at dispersions $\sigma \in \{0.05, 0.2, 0.5, 1.0\}$ rad and comparing the realised chordal-versus-Euclidean difference through

the secure stack against the closed-form prediction. The cubic term tracks the measured bias with correlations of 1.0000, 1.0000, 0.9999 and 0.9991. At the operating dispersion, the 2.50×10^{-6} rad bias closely matches the predicted 2.49×10^{-6} and sits $47\times$ below the concurrent encryption noise floor, ensuring the choice of mean alters no reported figure (Figure 3c). An adversarial double-cover configuration is also constructed (client groups straddling the branch cut, $\approx 2\pi$ apart), where the naive arithmetic mean errs by 3.1416 rad against 0.0000 for the exact chordal mean. This defines the strict boundary of the bound, but it never arises in trained runs, as the operations in §4.4 structurally exclude it at zero depth and zero communication by re-wrapping into $(-\pi, \pi]$ every round.

5.5 Parameterised entanglers

Lemma 2 predicts $R_{ZZ}(\theta)$ traverses the encrypted pipeline as two fixed CNOTs and one rotation, adding no rotation-track depth. This exact circuit was executed end-to-end for 20 angles in $[0, \pi]$ at 4096 shots each. The mean recovered-angle error is 8.3×10^{-3} rad (maximum 2.1×10^{-2}), evaluated against the estimator's intrinsic floor: a Z-basis estimate from N shots yields an expected pure-noise MAE of 1.25×10^{-2} rad ($\sigma = 1.6 \times 10^{-2}$ rad). With the measured mean strictly *below* this floor and the worst trial reaching only 1.3σ, the encrypted entangler introduces no resolvable angle error. Furthermore, every trial consumed exactly two CNOTs independently of θ, and the second qubit returned to $|0\rangle$ with probability 1.000—a direct verification of the XOR key relabelling, as any update failure would leave residual entanglement.

Per trial, encryption requires 6 ms and decryption 51 ms, whilst homomorphic CNOT key updates consume ≈ 1.5 h per gate. Since this update is algorithmically $O(1)$ and angle-independent (two zero-depth XOR operations on pad bits), this disparity stems entirely from implementation overhead, not complexity. Boolean gates are currently evaluated within a single-threaded fixed-point reference chosen for auditability; migrating to a Boolean-native TFHE scheme (Chillotti et al., 2020) resolves this. Crucially, this cost vanishes in the federated setting, where local plaintext training reduces per-round homomorphic work to a single depth-0 aggregation; it applies solely to delegated encrypted execution (Appendix G.4). The run thus proves exactness and constant overhead, but not throughput.

5.6 Hardware

An encryption layer contributing more than the device's native gate error would cap achievable accuracy regardless of training budget. The encrypted round trip on `ibm_fez` attains state fidelity $F = 0.9918$. To *isolate* the encryption's contribution,a matched unencrypted control was run – 100 angles under the same seeded sampling scheme, 4096 shots each, same device, calibration snapshot alongside – giving $F = 0.99957$ (s.d. 0.00107). The 0.0078 difference is comparable to the device's median readout error at the time (0.0083), i.e. ordinary device variation, while the encryption arithmetic is separately bounded at 1.55×10^{-4} rad by the homomorphic-layer-only round trip. The system is hardware-limited, not encryption-limited.

5.7 Independence from the dataset and client count

The evaluation is extended in two directions (five seeds per point; tables in Appendix G.6). Scaling from 3 to 20 clients increases the final MSE slightly: from 1.3035 to 1.3763 on California Housing, and from 1.4214 to 1.5246 on Wine Quality. This degradation is mild and monotonic, showing no aggregation blow-up despite a $6.7\times$ jump in participants. Furthermore, training out to 50 rounds improves the mean on both datasets ($1.2285 \to 1.0831$; $1.3148 \to 1.1275$) and narrows the seed spread ($0.54 \to 0.37$; $0.20 \to 0.09$). This behaviour confirms standard federated convergence, entirely unhindered by the encryption layer. Finally, although Wine Quality differs in feature count, scale, and target, it reproduces every qualitative trend. This includes a precision–bandwidth relation that is numerically identical across both datasets, confirming it as a property of the protocol itself.

5.8 Memory cost

Non-interactivity is bought with bandwidth, and the presentation states where the trade stops being beneficial. Per client-round the protocol moves 22.67 MB up and the same down in *one* sequential exchange; interactive baselines move about 0.55 MB per client-round – roughly $41\times$ less traffic over a 50-round schedule – but need 37 to 73 sequential exchanges per federated round, $1{,}850$ to $3{,}650$ over the schedule, each bound by network latency. On one axis (Table 9, Appendix H.2): at 50 ms round-trip latency the interactive floor is 1.9–3.7 s per round before any payload moves, while this

protocol's 45.3 MB take 0.36 s at 1 Gbps, 3.6 s at 100 Mbps and 36 s at 10 Mbps. The crossover sits near 100 Mbps at wide-area latency: above it, the protocol is beneficial; below it or on a low-latency metro link, the interactive baselines outweigh the benefit. The cross-institution regime this work targets sits above the crossover; edge deployment does not. Precision is a second lever: traffic falls from $6{,}485$ MB at 12 bits to $4{,}375$ MB at 8 bits with final loss statistically unchanged (§5.3). On compute, protocol accounting gives a $10\times$ to $30\times$ per-rotation reduction against Solovay–Kitaev as ϵ tightens from 10^{-2} to 10^{-6} (Appendix H.3) – a gate-count estimate, not a wall-clock measurement.

6 Discussion and limitations

Mechanism-level claims rest on proofs and configuration-insensitive measurements: the depth ledger holds on two independent backends, the aggregation bound spans two orders of magnitude of dispersion, and the entangler compilation is exact. Run-level claims provide feasibility evidence at a deliberately small scale, rigorously designed for falsifiability – pairing is verified, and the most flattering interpretation of the empirical data was explicitly tested and rejected (§5.3).

The protocol converts a latency problem into a throughput problem, offering a favourable trade above roughly 100 Mbps at wide-area latency (§5.8). Ciphertext compression represents the natural next step. This poses an equity concern alongside performance, given that environments requiring federated privacy are often the least well-connected. Separately, the encrypted-CNOT path costs ≈ 1.5 h per gate in the current reference implementation (an implementation constraint that currently bounds delegated encrypted execution).

The composed guarantee is information-theoretic at the quantum-state level and computational under (Ring-)LWE on the classical layer, leaving the system bounded by the weaker link. The reported runs utilise a reference lattice profile chosen strictly for structural verification and carry no formal security guarantees (Appendix J). Decryption capability is shared across clients – the standard shared-key assumption, with threshold key generation (Mouchet et al., 2021) as a drop-in hardening – whilst the honest-but-curious model precludes anomaly detection, leaving silent poisoning unaddressed. Generalising to larger client populations, further datasets, and alternative ansatz families remains future work, and hardware validation required access to a 156-qubit processor (Appendix L).

7 Conclusion

Encrypted training is usually framed as an architectural challenge, but for a growing class of models, it is a coordinate problem. In the spin chart, the composition law of $\mathrm{SU}(2)$ (and any $\mathrm{Spin}(n)$) is bilinear. Therefore, encrypted rotation updates cost one multiplicative level, federated averaging costs zero, and bootstrapping never enters the loop. The exportable principle is clear: when parameters reside on a group, the group law dictates encryption feasibility.

Reproducibility statement

Every number reported in this paper is produced by a script released upon acceptance or upon request. Appendix G records, per experiment, the claim under test, the entry point, the exact command line, the seeds, the cryptographic stack and precision, the hardware, and the job identifiers; Appendix J records the lattice parameter profiles and states which runs used which. Proposition 1, Proposition 2 and Lemma 2 are proved in Appendices B, D and C respectively. The two backend paths, the two aggregation compositions used by the training harness and by the protocol, and the float64 decode setting used to avoid a float32 cast floor in the exactness measurements are documented in Appendix G so that any reported figure can be traced to the code path that produced it. Appendix M discloses the use of LLMs.

Ethics statement

This work strengthens federated confidentiality without introducing new surveillance or data extraction capabilities. Three structural limitations remain. Firstly, server blindness precludes update auditing, creating tension with model-inspection regimes and exposing the system to silent client poisoning. Secondly, physical validation requires institutional quantum and HPC access, risking the concentration of benefits among already well-resourced actors. Finally, the bandwidth profile disadvantages the exact low-connectivity participants for whom federated privacy guarantees are most valuable. These constraints and dual-use/compliance implications are detailed in Appendix L.

REFERENCES

M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 308–318, 2016. doi: 10.1145/2976749.2978318. URL https://doi.org/10.1145/2976749.2978318.

A. Al Badawi, J. Bates, F. Bergamaschi, D. B. Cousins, S. Erabelli, N. Genise, S. Halevi, H. Hunt, A. Kim, Y. Lee, et al. OpenFHE: Open-source fully homomorphic encryption library. In *Proceedings of the 10th Workshop on Encrypted Computing and Applied Homomorphic Cryptography (WAHC)*, pp. 53–63, 2022. doi: 10.1145/3560827.3563379. URL https://doi.org/10.1145/3560827.3563379.

M. Arjovsky, A. Shah, and Y. Bengio. Unitary evolution recurrent neural networks. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1120–1128, 2016. URL https://proceedings.mlr.press/v48/arjovsky16.html.

R. Ballester, J. Cerquides, and L. Artiles. Quantum federated learning: a comprehensive literature review of foundations, challenges, and future directions. *Quantum Machine Intelligence*, 7(2):73, 2025. doi: 10.1007/s42484-025-00292-2. URL https://doi.org/10.1007/s42484-025-00292-2.

A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter. Elementary gates for quantum computation. *Physical Review A*, 52 (5):3457–3467, 1995. doi: 10.1103/PhysRevA.52.3457. URL https://doi.org/10.1103/PhysRevA.52.3457.

M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, 2019. doi: 10.1088/2058-9565/ab4eb5. URL https://doi.org/10.1088/2058-9565/ab4eb5.

D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, K. H. Li, T. Parcollet, P. P. B. de Gusmão, and N. D. Lane. Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390*, 2020. URL https://arxiv.org/abs/2007.14390.

K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 1175–1191, 2017. doi: 10.1145/3133956.3133982. URL https://doi.org/10.1145/3133956.3133982.

A. Broadbent and S. Jeffery. Quantum homomorphic encryption for circuits of low T-gate complexity. In R. Gennaro and M. Robshaw (eds.), *Advances in Cryptology – CRYPTO 2015*, pp. 609–629, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg. ISBN 978-3-662-48000-7. URL https://doi.org/10.1007/978-3-662-48000-7_30.

A. Broadbent, J. Fitzsimons, and E. Kashefi. Universal blind quantum computation. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pp. 517–526, 2009. doi: 10.1109/FOCS.2009.36. URL https://doi.org/10.1109/FOCS.2009.36.

M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3:625–644, 2021. doi: 10.1038/s42254-021-00348-9. URL https://doi.org/10.1038/s42254-021-00348-9.

M. Chehimi and W. Saad. Quantum federated learning with quantum data. In *ICASSP 2022 – IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8617–8621, 2022. doi: 10.1109/ICASSP43922.2022.9746622. URL https://doi.org/10.1109/ICASSP43922.2022.9746622.

S. Y.-C. Chen and S. Yoo. Federated quantum machine learning. *Entropy*, 23(4):460, 2021. doi: 10.3390/e23040460. URL https://doi.org/10.3390/e23040460.

J. H. Cheon, A. Kim, M. Kim, and Y. Song. Homomorphic encryption for arithmetic of approximate numbers. In *Advances in Cryptology – ASIACRYPT 2017*, volume 10624 of *Lecture Notes in Computer Science*, pp. 409–437. Springer, Cham, 2017. doi: 10.1007/978-3-319-70694-8_15. URL https://doi.org/10.1007/978-3-319-70694-8_15.

I. Chillotti, N. Gama, M. Georgieva, and M. Izabachène. TFHE: Fast fully homomorphic encryption over the torus. *Journal of Cryptology*, 33(1):34–91, 2020. doi: 10.1007/s00145-019-09319-x. URL https://doi.org/10.1007/s00145-019-09319-x.

P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547–553, 2009. doi: 10.1016/j.dss.2009.05.016. URL https://doi.org/10.1016/j.dss.2009.05.016.

C. M. Dawson and M. A. Nielsen. The Solovay-Kitaev algorithm. *arXiv preprint quant-ph/0505030*, 2005. URL https://arxiv.org/abs/quant-ph/0505030.

W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644–654, 1976. doi: 10.1109/TIT.1976.1055638. URL https://doi.org/10.1109/TIT.1976.1055638.

S. Dutta, P. P. Karanth, P. M. Xavier, N. Innan, S. Ben Yahia, M. Shafique, I. L. de Freitas, and D. E. Bernal Neira. Federated learning with quantum computing and fully homomorphic encryption: a novel computing paradigm shift in privacy-preserving ML. *arXiv preprint arXiv:2409.11430v3*, 2024. URL https://arxiv.org/abs/2409.11430.

C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014. doi: 10.1561/0400000042. URL https://doi.org/10.1561/0400000042.

W. El Maouaki, N. Innan, A. Marchisio, T. Said, M. Shafique, and M. Bennai. Robqfl: Robust quantum federated learning in adversarial environment. In *2025 IEEE International Conference on Quantum Artificial Intelligence (QAI)*, pp. 128–134, 2025. doi: 10.1109/QAI63978.2025.00027. URL https://doi.org/10.1109/QAI63978.2025.00027.

J. F. Fitzsimons. Private quantum computation: An introduction to blind quantum computing and related protocols, 2016. URL https://arxiv.org/abs/1611.10107.

J. Geiping, H. Bauermeister, H. Dröge, and M. Moeller. Inverting gradients – how easy is it to break privacy in federated learning? In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 16937–16947, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/c4ede56bbd98819ae6112b20ac6bf145-Paper.pdf.

C. Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pp. 169–178, 2009. doi: 10.1145/1536414.1536440. URL https://doi.org/10.1145/1536414.1536440.

R. Gilad-Bachrach, N. Dowlin, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing. CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, volume 48 of *Proceedings of Machine Learning Research*, pp. 201–210, 2016. URL https://proceedings.mlr.press/v48/gilad-bachrach16.html.

IBM. Ibm quantum products, 2026. URL https://www.ibm.com/quantum/products. Accessed: 2026.

IBMQuantum. IBM Quantum, 2024. URL https://quantum.ibm.com/.

P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, et al. Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1–2):1–210, 2021. doi: 10.1561/2200000083. URL https://doi.org/10.1561/2200000083.

M. Lezcano-Casado and D. Martínez-Rubio. Cheap orthogonal constraints in neural networks: A simple parametrization of the orthogonal and unitary group. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3794–3803, 2019. URL https://proceedings.mlr.press/v97/lezcano-casado19a.html.

W. Li and D.-L. Deng. Quantum delegated and federated learning via quantum homomorphic encryption. *Research Directions: Quantum Technologies*, 3:e3, 2025. doi: 10.1017/qut.2025.2. URL https://doi.org/10.1017/qut.2025.2. arXiv:2409.19359.

P. Lounesto. *Clifford Algebras and Spinors*, volume 286 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 2nd edition, 2001. doi: 10.1017/CBO9780511526022. URL https://doi.org/10.1017/CBO9780511526022.

V. Lyubashevsky, C. Peikert, and O. Regev. On ideal lattices and learning with errors over rings. In *Advances in Cryptology – EUROCRYPT 2010*, pp. 1–23. Springer, 2010. doi: 10.1007/978-3-642-13190-5_1. URL https://doi.org/10.1007/978-3-642-13190-5_1.

G. Ma and H. Li. Quantum fully homomorphic encryption by integrating pauli one-time pad with quaternions. *Quantum*, 6:866, 2022. doi: 10.22331/q-2022-12-01-866. URL https://doi.org/10.22331/q-2022-12-01-866.

U. Mahadev. Classical homomorphic encryption for quantum circuits, 2023. URL https://arxiv.org/abs/1708.02130.

F. L. Markley, Y. Cheng, J. L. Crassidis, and Y. Oshman. Averaging quaternions. *Journal of Guidance, Control, and Dynamics*, 30(4):1193–1197, 2007. doi: 10.2514/1.28949. URL https://doi.org/10.2514/1.28949.

J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812, 2018. doi: 10.1038/s41467-018-07090-4. URL https://doi.org/10.1038/s41467-018-07090-4.

B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282. PMLR, 2017. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. Quantum circuit learning. *Physical Review A*, 98 (3):032309, 2018. doi: 10.1103/PhysRevA.98.032309. URL https://doi.org/10.1103/PhysRevA.98.032309.

C. Mouchet, J. Troncoso-Pastoriza, J.-P. Bossuat, and J.-P. Hubaux. Multiparty homomorphic encryption from ring-learning-with-errors. *Proceedings on Privacy Enhancing Technologies*, 2021 (4):291–311, 2021. doi: 10.2478/popets-2021-0071. URL https://doi.org/10.2478/popets-2021-0071.

M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniversary edition, 2010. doi: 10.1017/CBO9780511976667. URL https://doi.org/10.1017/CBO9780511976667.

R. K. Pace and R. Barry. Sparse spatial autoregressions. *Statistics & Probability Letters*, 33(3): 291–297, 1997. doi: 10.1016/S0167-7152(96)00140-X. URL https://doi.org/10.1016/S0167-7152(96)00140-X.

T. Parcollet, M. Morchid, and G. Linarès. A survey of quaternion neural networks. *Artificial Intelligence Review*, 53:2957–2982, 2020. doi: 10.1007/s10462-019-09752-1. URL https://doi.org/10.1007/s10462-019-09752-1.

Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023. URL https://doi.org/10.5281/zenodo.2562110.

O. Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM*, 56(6):1–40, 2009. doi: 10.1145/1568318.1568324. URL https://doi.org/10.1145/1568318.1568324.

M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, 2019. doi: 10.1103/PhysRevA.99.032331. URL https://doi.org/10.1103/PhysRevA.99.032331.

A. Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979. doi: 10.1145/359168.359176. URL https://doi.org/10.1145/359168.359176.

P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997. doi: 10.1137/S0097539795293172. URL https://doi.org/10.1137/S0097539795293172.

N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219*, 2018. URL https://arxiv.org/abs/1802.08219.

F. Vatan and C. Williams. Optimal quantum circuits for general two-qubit gates. *Physical Review A*, 69(3):032315, 2004. doi: 10.1103/PhysRevA.69.032315. URL https://doi.org/10.1103/PhysRevA.69.032315.

C. Zhang, S. Li, J. Xia, W. Wang, F. Yan, and Y. Liu. BatchCrypt: Efficient homomorphic encryption for cross-silo federated learning. In *2020 USENIX Annual Technical Conference (USENIX ATC)*, pp. 493–506, 2020. URL https://www.usenix.org/conference/atc20/presentation/zhang-chengliang.

L. Zhu, Z. Liu, and S. Han. Deep leakage from gradients. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/60a6c4002cc7b29142def8871531281a-Paper.pdf.

APPENDIX OVERVIEW

A quantum and federated-learning preliminaries; **B** proof of the degree-2 depth ledger and its $\mathrm{Spin}(n)$ generalisation; **C** aggregation correctness, the mask lemma, and the entangler compilation proof; **D** aggregation geometry, with the proof of the cubic bound; **E** protocol details for Pauli-OTP and the quaternion track; **F** the detailed construction of encrypted QNN training; **G** experimental specifications, per experiment; **H** cost models; **I** scale tests; **J** implementation and security profiles; **K** extended discussion and limitations; **L** broader impact; **M** use of large language models.

A QUANTUM AND FEDERATED-LEARNING PRELIMINARIES

This appendix collects the minimum needed to follow §4. Readers fluent in federated learning can skip §A.1; readers new to quantum gates should read §A.2. Standard references are Nielsen & Chuang (2010) for the quantum material and Kairouz et al. (2021) for federated learning.

A.1 FEDERATED LEARNING

K clients hold private datasets $\mathcal{D}_k$ of size n_k, with $n = \sum_k n_k$. In each round every client initialises from the global parameters, trains locally, and returns an update; the server forms $\bar{\vartheta} = \sum_k \frac{n_k}{n} \vartheta^{(k)}$ (McMahan et al., 2017). The privacy failure that motivates encryption is that $\vartheta^{(k)}$ is informative about $\mathcal{D}_k$: gradient-inversion attacks recover training examples from transmitted updates alone (Zhu et al., 2019; Geiping et al., 2020). Homomorphic aggregation removes the server from the trust boundary by making the sum computable on ciphertexts (Gentry, 2009; Zhang et al., 2020).

A.2 QUANTUM PRELIMINARIES

A qubit state is a unit vector in $\mathbb{C}^2$, written $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$; n qubits live in $(\mathbb{C}^2)^{\otimes n}$, a 2^n-dimensional space that does not factorise into independent classical parts, which is what entanglement means operationally. Gates are unitary matrices. The Pauli group is generated by X, Y, Z; the Clifford group is the normaliser of the Pauli group, containing H, S and CNOT. Single-qubit rotations $R_{\hat{u}}(\theta) = \exp(-i\theta\,\hat{u}\cdot\vec{\sigma}/2)$ are the trainable weights of a variational circuit; they are not Clifford for generic θ, which is the entire source of difficulty in §1. Measurement in the Z basis returns a bit with probability given by the squared amplitudes, so all reported quantities are estimated from finite shot counts and carry sampling error $\propto 1/\sqrt{N}$.

A variational quantum circuit alternates rotation layers with a fixed entangling pattern and is trained by gradient descent, with gradients from the parameter-shift rule (Mitarai et al., 2018; Schuld et al., 2019) or from simulator backpropagation. Optimisation pathologies specific to this model class – barren plateaus (McClean et al., 2018) – are the reason a purely quantum configuration underperforms the hybrid design (Appendix H.4).

B THE DEPTH LEDGER: PROOFS

Proof of Proposition 1. (i) By Eq. 1 each component of $q \otimes r$ is a sum of terms $\pm q_j r_k$, each of total degree two, with coefficients $M^{(i)}_{jk} \in \{-1, 0, +1\}$ read off the Hamilton table; no other monomials occur. *(ii)* Substituting plaintext r_k into Eq. 1 leaves a $\mathbb{R}$-linear form in the encrypted q_j whose coefficients are plaintext, so $\mathrm{Enc}(q')$ is obtained by plaintext–ciphertext multiplications and additions, consuming one multiplicative level by assumption on $\mathcal{E}$. *(iii)* $\sum_k \frac{n_k}{n} \mathrm{Enc}\big(q^{(k)}\big)$ is an $\mathbb{R}$-linear combination of ciphertexts with plaintext scalars, which $\mathcal{E}$ supports at depth 0; renormalisation is applied to the decrypted plaintext and is therefore outside the homomorphic layer entirely. Composing (ii) and (iii) gives the round bound. □

Corollary 2 (Unbounded round count)**.** *A parameterisation of $\mathcal{E}$ supporting multiplicative depth $\geqslant 1$ suffices for an unbounded number of federated rounds, each consisting of one server-side plaintext rotation applied to every encrypted key followed by one weighted aggregation, provided clients re-encrypt between rounds. No bootstrapping is required in the inner loop.*

Proof. Each round consumes at most one level by Proposition 1, and re-encryption at the client resets the level budget at zero homomorphic cost because clients already decrypt the aggregate in step 5 of the protocol. □

Proof of Corollary 1. Rotors $R \in \mathrm{Spin}(n) \subset \mathrm{Cl}^{[0]}(n)$ compose by the geometric product, which is bilinear on the algebra: writing $R = \sum_A R_A e_A$ over the even-graded basis $\{e_A\}$, the product RR' has components $\sum_{A,B} c^C_{AB} R_A R'_B$ with structure constants $c^C_{AB} \in \{-1, 0, +1\}$ fixed by the algebra (Lounesto, 2001). This is Eq. 1 with a larger index set, and the argument of Proposition 1 applies unchanged. For $n = 3$, the even subalgebra is $\mathbb{H}$ and the statement reduces to the quaternion case. □

Remark 1 (Why not Euler angles)**.** *The obstruction the reduction removes is worth naming precisely. In Euler coordinates the composition law involves* $\arctan$, $\arccos$ *and products of trigonometric functions of encrypted arguments. A levelled scheme reaches these only through polynomial approximation, at a multiplicative depth that grows with the required accuracy and typically forces bootstrapping inside the training loop (Cheon et al., 2017). In the discrete-alphabet route, the same composition is exact but requires* $\Theta(\log^c(1/\epsilon))$ *gates per rotation (Dawson & Nielsen, 2005), in practice above* $25{,}000$ *(Ma & Li, 2022), each non-Clifford element demanding an active key repair. Degree two is therefore not a marginal improvement over these routes; it is a different complexity class for the same computation.*

C Correctness of Aggregation and Entangler Compilation

C.1 What the server computes, and why it is correct

Let client k hold trained angles $\theta_g^{(k)}$ for gate g, encoded as $q_g^{(k)} = q(\theta_g^{(k)})$ by Eq. 2 and encrypted component-wise under keys shared by the clients. The server computes

$$\mathrm{Enc}\big(q_g^{\mathrm{glob}}\big) \;=\; \sum_{k=1}^{K} n_k \cdot \mathrm{Enc}\left(q_g^{(k)}\right), \tag{4}$$

using ciphertext addition and plaintext-integer weighting only. Decryption correctness of $\mathcal{E}$ gives $\mathrm{Dec}\left(\mathrm{Enc}\big(q_g^{\mathrm{glob}}\big)\right) = \sum_k n_k q_g^{(k)} + \varepsilon$ with $\|\varepsilon\|$ bounded by the scheme's encoding error, measured in §5.2; clients divide by n, renormalise and recover $\hat{\theta}_g = 2\,\mathrm{atan2}(-t_1, t_0)$. No ciphertext–ciphertext multiplication occurs, so no relinearisation or rescaling is needed and the level budget is untouched. The server holds an evaluation-only context: the aggregation object exposes addition, scaling and serialisation, and decryption raises an error.

C.2 Proof of the mask lemma

Proof of Lemma 1. (i) Quaternion multiplication is $\mathbb{R}$-bilinear, so for fixed u the map L_u is $\mathbb{R}$-linear and commutes with any finite $\mathbb{R}$-linear combination: $\sum_k w_k(u \otimes q_k) = u \otimes (\sum_k w_k q_k)$. Hence a mask shared by all clients survives averaging intact and is removed by one plaintext multiplication by $u^{-1} = \bar{u}$ after decryption. *(ii)* For the delegation channel, let ρ_i, ρ_j be two clients' states. Conjugation by a shared unitary is an isometry of the Hilbert–Schmidt inner product, $\mathrm{Tr}[(U\rho_i U^\dagger)(U\rho_j U^\dagger)] = \mathrm{Tr}(\rho_i \rho_j)$, so a common mask hides each state individually but preserves all pairwise overlaps, which is a non-trivial function of the clients' private parameters. With independent uniform Pauli keys, the twirl $\frac{1}{4}\sum_{a,b}(X^a Z^b \otimes I)(\rho_i \otimes \rho_j)(X^a Z^b \otimes I)^\dagger$ applied to each factor sends the pair to $\frac{I}{d} \otimes \frac{I}{d}$, so no function of the pair survives. □

The practical reading: masks on the *aggregation* channel may be common (and in the reported protocol are simply absent, since the aggregated object is the unmasked update quaternion under shared keys), whereas masks on the *delegation* channel must be per-client and fresh.

C.3 Proof of the entangler compilation lemma

Proof of Lemma 2. Exact decompositions of controlled rotations into at most two CNOTs plus single-qubit gates are given by Barenco et al. (1995), and of a general two-qubit unitary into at most three CNOTs by the Cartan decomposition argument of Vatan & Williams (2004); both are exact identities, not approximations. Applying either places the continuous parameter on a single-qubit rotation, which by Proposition 1(ii) costs one multiplicative level. The CNOTs are Clifford, so conjugation maps the Pauli frame to another Pauli frame; for $X^{a_1}Z^{b_1} \otimes X^{a_2}Z^{b_2}$ the induced key map is $(a_1, b_1, a_2, b_2) \mapsto (a_1,\ b_1 \oplus b_2,\ a_1 \oplus a_2,\ b_2)$, a fixed XOR relabelling of pad bits carrying no multiplicative depth and no dependence on θ. Finally $R_Z(\theta)X^a = X^a R_Z((-1)^a\theta)$, so passing the rotation through the pad conditions only the sign of the angle on the key. Hence the number of encrypted operations per entangler is a constant independent of θ and of the ansatz, and the rotation-track depth is unchanged. □

Remark 2 (Consequence for expressivity)**.** *Because the decompositions are exact, no expressivity is lost: any parameterised entangler an unrestricted ansatz could use is available here at a fixed gate-count factor. The depth class – and therefore the absence of bootstrapping – depends only on the number of* sequential *rotations. In the federated loop it cannot grow at all, because clients train locally and encrypt only the resulting angles.*

D Aggregation Geometry

Proposition 2 (Aggregation geometry)**.** *Fix an axis $\hat{u}$ and weights $w_k \geqslant 0$, $\sum_k w_k = 1$, and let $\bar{\theta} = \sum_k w_k\theta_k$ with deviations $\delta_k = \theta_k - \bar{\theta}$. Let $\hat{\theta}$ be the angle recovered from the renormalised weighted quaternion mean. Then*

$$\hat{\theta} - \bar{\theta} \;=\; -\frac{1}{24}\sum_k w_k\delta_k^3 \;+\; O(\delta^5),$$

provided all representatives lie in a common hemisphere. The bound fails, by exactly π, for configurations straddling the branch cut, which the wrapping step above excludes by construction.

Proof of Proposition 2. Fix the axis and write $\varphi_k = \theta_k/2$, $\bar{\varphi} = \bar{\theta}/2$ and $\delta_k = \theta_k - \bar{\theta}$, so $\varphi_k = \bar{\varphi} + \delta_k/2$ and $\sum_k w_k\delta_k = 0$. The weighted quaternion mean has components

$$t_0 = \sum_k w_k \cos\varphi_k = \cos\bar{\varphi}\, m_c - \sin\bar{\varphi}\, m_s, \quad -t_1 = \sum_k w_k \sin\varphi_k = \sin\bar{\varphi}\, m_c + \cos\bar{\varphi}\, m_s, \quad (5)$$

with $m_c = \sum_k w_k \cos(\delta_k/2)$ and $m_s = \sum_k w_k \sin(\delta_k/2)$. Therefore $\mathrm{atan2}(-t_1, t_0) = \bar{\varphi} + \psi$ with $\tan\psi = m_s/m_c$. Expanding, $m_s = \frac{1}{2}\sum_k w_k\delta_k - \frac{1}{48}\sum_k w_k\delta_k^3 + O(\delta^5) = -\frac{1}{48}\sum_k w_k\delta_k^3 + O(\delta^5)$ using $\sum_k w_k\delta_k = 0$, and $m_c = 1 - \frac{1}{8}\sum_k w_k\delta_k^2 + O(\delta^4)$. Hence $\psi = -\frac{1}{48}\sum_k w_k\delta_k^3 + O(\delta^5)$ and $\hat{\theta} = 2(\bar{\varphi} + \psi) = \bar{\theta} - \frac{1}{24}\sum_k w_k\delta_k^3 + O(\delta^5)$, which is Eq. 3. Renormalisation does not affect the recovered angle because atan2 is scale-invariant. The expansion requires the representatives to share a hemisphere; if two groups straddle the branch cut, $m_c \to 0$ and the recovered angle jumps by π. □

The factor $\frac{1}{24}$, rather than the $\frac{1}{6}$ of a naive full-angle expansion, is the visible trace of the half-angle encoding: the protocol averages quaternion components at $\theta/2$.

Measured validation. Table 3 reports the dispersion sweep of §5.4: five clients, 36 angles each, five repetitions per dispersion, through the secure stack at 12-bit precision. "Exact bias" is the chordal-versus-arithmetic difference computed on plaintext angles, isolating geometry from encryption noise; "prediction" is Eq. 3; "floor" is the encryption noise of the same runs.

Table 3: Aggregation-geometry sweep. The cubic bound tracks the realised bias across two orders of magnitude of client dispersion, and the bias is below the encryption noise floor throughout.

σ (rad)	Exact bias (rad)	Prediction (rad)	Correlation	Encryption floor (rad)
0.05	2.4956×10^{-6}	2.4949×10^{-6}	1.0000	1.1705×10^{-4}
0.20	1.4067×10^{-4}	1.4014×10^{-4}	1.0000	2.0381×10^{-4}
0.50	2.1721×10^{-3}	2.1126×10^{-3}	0.9999	2.1764×10^{-3}
1.00	1.7974×10^{-2}	1.6236×10^{-2}	0.9991	1.7996×10^{-2}

The branch-cut case, stated precisely. With two client groups placed at $\bar{\theta} \pm (\pi - \epsilon)$, $\epsilon = 0.05$, the naive arithmetic mean of angles errs by 3.1416 rad against the true circular mean, and the exact chordal mean by 0.0000 rad. Two compositions appear in our codebase and it is worth separating them explicitly: the *protocol* path aggregates encrypted quaternions and recovers the angle by $\mathrm{atan2}$, hence computes the chordal mean; the *scale-test harness* composes a per-client encrypted round trip with masked secure aggregation of the recovered angles, hence computes the Euclidean mean and inherits the $\pm\pi$ failure in this constructed case (3.1414 rad). Neither is exercised adversarially in any trained run, because the decode step re-wraps into $(-\pi, \pi]$ every round; the principal-branch and hemisphere-alignment steps of §4.4 close the case structurally, in plaintext, at zero depth and zero communication. Since the two means differ by 2.50×10^{-6} rad at the operating dispersion – $47\times$ below the encryption noise floor – no reported number depends on which is used.

E Quantum Homomorphic Encryption Details

E.1 Pauli One-Time Pad

Pauli-OTP (Broadbent & Jeffery, 2015) encrypts $|\psi\rangle$ as $|\psi'\rangle = X^a Z^b |\psi\rangle$ with $a, b \in \{0,1\}^n$ and $X^a Z^b = (X^{a_1} Z^{b_1}) \otimes \cdots \otimes (X^{a_n} Z^{b_n})$. Keys are sampled fresh per use and shared with the server only in encrypted form. Averaged over the key the ciphertext is maximally mixed, so the scheme is information-theoretically secure.

E.1.1 Clifford gates

For Clifford C and Pauli P, $CPC^\dagger \in \mathcal{P}$, so

$$C|\psi'\rangle = CX^a Z^b |\psi\rangle = (CX^a Z^b C^\dagger)(C|\psi\rangle) = X^{a'} Z^{b'} (C|\psi\rangle), \tag{6}$$

and only the classical keys change. The server updates the encrypted keys purely classically; no client interaction is required.

E.1.2 Non-Clifford gates

For the T gate,

$$T|\psi'\rangle = T(X^a Z^b |\psi\rangle) = X^a Z^{a\oplus b} P^a (T|\psi\rangle), \tag{7}$$

so the residual error P^a is conditioned on a secret bit. The accepted repairs are deferred correction, $O(N_T^2)$ in the number of T gates (Broadbent & Jeffery, 2015), or auxiliary magic states with resource cost exponential in T-depth (Broadbent et al., 2009). For arbitrary continuous rotations the Solovay–Kitaev route requires above $25{,}000$ gates per rotation (Dawson & Nielsen, 2005; Ma & Li, 2022). This is the obstruction that §4.1 removes by changing coordinates rather than by paying either price.

E.1.3 Encrypted CNOT

Mahadev (2023) constructs an encrypted-CNOT primitive from trapdoor claw-free functions under LWE, allowing a fully classical client: the server prepares a superposition over function inputs, entangles it with the target and measures the output, so the conditional flip is controlled by an encrypted bit without decryption.

E.2 QUATERNION TRACK

A quaternion $q = t_1 + t_2\mathbf{i} + t_3\mathbf{j} + t_4\mathbf{k}$ with $\|t\|_2 = 1$ maps to a single-qubit unitary by

$$U(t) = t_1 I_2 + t_2(iX) + t_3(iZ) + t_4(iY), \tag{8}$$

so the encrypted object is a constrained real 4-vector, which is why an approximate-arithmetic scheme over reals is the natural backend. On the delegation channel the key is a unit quaternion and the masked state $U(t)\,|\psi\rangle$ is a uniformly random point on the Bloch sphere. When the server applies a rotation V the key updates as $U(t_{\text{new}}) = U(t)V^{-1}$, a degree-2 homomorphic update (Proposition 1).

Two-qubit gates couple the keys,

$$\mathrm{CNOT}\big(U(t_1) \otimes U(t_2)\big) = \big(U(t_1') \otimes U(t_2')\big)\mathrm{CNOT}(|\psi_1\psi_2\rangle), \tag{9}$$

preventing independent per-qubit updates. Following Ma & Li (2022) the state is converted to the Pauli frame immediately before an entangling gate, where the update is the XOR relabelling of Lemma 2, and converted back afterwards. The conversion decomposes the encrypted angle into Euler form and approximates each angle bit-wise, applying conditional rotations under the encrypted-CNOT primitive; its bit-serial character is its cost, and it is amortised across the rotation layers it enables. The frame conversions are a constant number per entangler and are levelled, requiring no periodic carry bootstrapping (Appendix J).

F ENCRYPTED QNN TRAINING: DETAILED CONSTRUCTION

F.1 CLIENT SIDE

A trainable rotation $U(\theta, \hat{u}) = \cos(\theta/2)I - i\sin(\theta/2)(u_x X + u_y Y + u_z Z)$ is encoded by Eq. 2. In the federated path the client encrypts the four components of the *unmasked* update quaternion under the clients' shared homomorphic keys, after principal-branch wrapping and hemisphere alignment (§4.4). Quaternion one-time-pad keys are a separate object used only on the delegation channel, per client and per use, in accordance with Lemma 1.

F.2 SERVER SIDE

The server applies plaintext rotations by right multiplication, $\mathrm{Enc}(q_{\text{new}}) = \mathrm{Enc}(q) \cdot V^{-1}$ at one multiplicative level, and forms the weighted sum of Eq. 4 at depth 0. It never decrypts and never multiplies ciphertext by ciphertext.

F.3 ENTANGLING GATES

Entangling gates are handled by the short representation switch of §E.2, yielding the division of labour that makes the construction work: continuous parameters ride the quaternion track, where composition is degree 2; Clifford structure rides the Pauli track, where updates are XOR relabellings; and the two meet only at frame conversions of constant count.

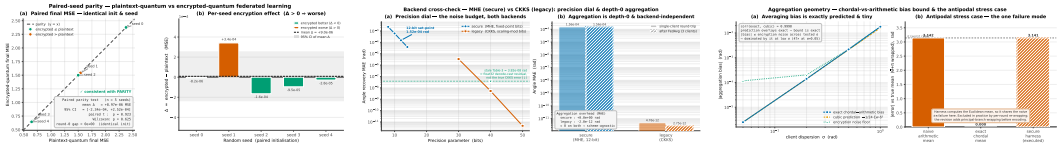
G EXPERIMENTAL SPECIFICATIONS

For each experiment we record: the claim under test, the procedure and entry point, the explicit assumptions, and the observed result. Unless stated otherwise, runs use the secure fixed-point multi-key stack at 12-bit quaternion precision, a six-qubit six-layer ansatz with a CNOT ring after each rotation layer (36 trainable angles), and exact statevector simulation with parameter-shift gradients. Cluster jobs were executed on NVIDIA T4, A40 and A100 nodes; job identifiers are given per experiment so that logs can be matched to results.

Two code paths, named once. Two aggregation compositions exist in the release and are used for different purposes. The *protocol path* aggregates encrypted quaternions and recovers angles by $\mathrm{atan2}$ (chordal mean); the *scale-test harness* composes a per-client encrypted round trip with masked secure aggregation of recovered angles using pairwise-cancelling masks and Shamir sharing (Shamir, 1979;

Diffie & Hellman, 1976) for dropout tolerance (Euclidean mean). The server decrypts in neither. Depth-0 aggregation claims refer to the protocol path and are measured in Appendix G.2; federated learning curves come from the harness; Appendix D quantifies the difference between the two as 2.50×10^{-6} rad at the operating dispersion.

Decode precision. Exactness measurements decode in float64. This matters: a float32 decode path pins the residual near 10^{-7} relative precision and therefore cannot represent the true CKKS aggregation error, which is four orders of magnitude smaller. All values in Table 1 are float64 decodes.



(a) Paired parity: encrypted vs. plaintext, five seeds.

(b) Precision dial and depth-0 aggregation, two backends.

(c) Aggregation bias against the cubic bound of Eq. 3.

Figure 3: The three load-bearing measurements. (a) With initialisation and batch order held identical, encrypted and plaintext arms are indistinguishable. (b) Weighted FedAvg adds no error beyond per-client encoding on either backend. (c) The realised aggregation bias is predicted by Eq. 3 and sits below the encryption noise floor throughout.

G.1 E1: Paired-seed parity

Claim. With initialisation and data order held identical, encrypted and plaintext federated training reach statistically indistinguishable final loss.

Procedure. For each seed in $\{0, 1, 2, 3, 4\}$, reconfigure reproducibility and rebuild the data loaders *immediately before each arm*, then run the plaintext and encrypted arms in turn; five federated rounds, three clients, secure stack, 12-bit. Report per-seed final validation MSE, a paired t-test, a Wilcoxon signed-rank test and a 95% confidence interval.

Assumptions and controls. Pairing is verified rather than assumed: arms that begin identically must have equal round-0 loss. The maximum round-0 gap across seeds is 0.00×10^{0}. Exact noiseless simulation is used deliberately, since injecting device noise would confound the quantity being isolated.

Result. Plaintext 1.3436 ± 0.7228, encrypted 1.3436 ± 0.7229; per-seed differences $\{-8.2 \times 10^{-6}, +3.4 \times 10^{-4}, -1.7 \times 10^{-4}, -9.5 \times 10^{-5}, -2.8 \times 10^{-5}\}$; mean $+9.0 \times 10^{-6}$, 95% CI $[-2.3 \times 10^{-4}, +2.5 \times 10^{-4}]$; $t = 0.103$, $p = 0.923$; Wilcoxon $W = 5.0$, $p = 0.625$. Job `265819`. The between-seed spread of final MSE is ≈ 0.72 (individual finals range 0.64 to 2.38), which is why single-run comparisons in this regime are uninformative about the encryption effect and why the paired design is necessary rather than merely tidy.

Table 4: Utility under encryption. Five paired seeds, five federated rounds, three clients; final validation MSE (mean $\pm$ s.d.). Pairing verified: maximum round-0 gap $= 0.00 \times 10^{0}$ across all seeds. Rows 1–2 isolate the encryption effect; rows 2–3 the noise-budget effect.

Arm	Final MSE	Paired Δ (95% CI)	Paired t / Wilcoxon
Plaintext	1.3436 ± 0.7228	–	–
Encrypted, 12-bit	1.3436 ± 0.7229	$+9.0 \times 10^{-6}$ $[-2.3, +2.5] \times 10^{-4}$	p=0.923 / p=0.625
Encrypted, 8-bit (16× noise)	1.3442 ± 0.7230	$+6.0 \times 10^{-4}$ $[-2.6, +3.8] \times 10^{-3}$	p=0.636 / p=1.0

G.2 E2: BACKEND CROSS-CHECK AND THE NOISE BUDGET

Claim. Weighted FedAvg over encrypted quaternions adds no error beyond per-client encoding, on cryptographically distinct backends; and the encryption noise floor is a tunable dial.

Procedure. Encrypt three clients' 36-angle vectors, aggregate homomorphically, decrypt, and measure single-client versus post-aggregation angle MAE; the difference is the aggregation overhead. Repeat on the secure multi-key stack and on OpenFHE CKKS (ring 2^{16}, depth 40, 50-bit scaling, `FLEXIBLEAUTO`, `HYBRID` key switching). Sweep precision (`qotp_bits` $\in \{8, 10, 12, 14\}$; CKKS scaling modulus $\in \{30, 40, 50\}$), 50 repetitions per point, seed 42, float64 decode.

Assumptions. Both stacks are genuine cryptographic engines rather than mocks; the CKKS arm performs real context generation and is memory-heavy.

Result. Table 1. Aggregation overhead 0.0 (secure) and -2.0089×10^{-12} rad (CKKS); the negative sign is independent rounding errors averaging down. Job `265860`.

Table 5: Depth-0 aggregation is exact on both backends. The difference between the two middle columns is the overhead Proposition 1(iii) predicts to be zero. Float64 decode, three clients, 36 angles, radians; precision sweep in Table 6.

Backend	Single client	Post-FedAvg	Aggregation overhead
Fixed-point multi-key (12-bit)	1.5632×10^{-4}	1.5632×10^{-4}	$\mathbf{0.0}$
OpenFHE CKKS (50-bit modulus)	4.7573×10^{-12}	2.7484×10^{-12}	$\mathbf{-2.0 \times 10^{-12}}$

Table 6: Encryption noise budget as a function of the precision knob, float64 decode, 50 repetitions per point. Angle MAE in radians.

Multi-key, precision bits	8	10	12	14
Angle MAE (rad)	2.31×10^{-3}	6.11×10^{-4}	1.52×10^{-4}	3.91×10^{-5}
CKKS, scaling modulus	30	40	50	–
Angle MAE (rad)	3.39×10^{-6}	5.27×10^{-9}	4.58×10^{-12}	–

G.3 E3: AGGREGATION-GEOMETRY STRESS TEST

Claim. The chordal-versus-Euclidean aggregation bias is predicted by Eq. 3, is negligible in the operating regime, and fails only in a constructed branch-cut configuration that the protocol excludes.

Procedure. Five clients, 36 angles, dispersions $\sigma \in \{0.05, 0.2, 0.5, 1.0\}$ rad, five repetitions each, through the secure round-trip and aggregation primitives; plus a near-antipodal case with two groups at $\bar{\theta} \pm (\pi - 0.05)$. Seed 42.

Result. Table 3; antipodal case 3.1416 rad (arithmetic mean), 0.0000 rad (chordal mean), 3.1414 rad (harness). The run is deterministic: an independent re-execution on different hardware reproduces every reported digit.

G.4 E4: PARAMETERISED ENTANGLER THROUGH THE ENCRYPTED PIPELINE

Claim. $R_{ZZ}(\theta)$ compiles exactly onto the supported gate set and traverses the encrypted pipeline within the shot-noise floor, at constant gate overhead.

Procedure. Execute $R_{ZZ}(\theta) = \mathrm{CNOT} \cdot (I \otimes R_Z(\theta)) \cdot \mathrm{CNOT}$ end to end through the encrypted stack – encrypted state preparation, encrypted CNOT, encrypted rotation, encrypted CNOT, decrypt and run – for 20 angles spanning $[0, \pi]$ at 4096 shots, recovering θ from the measured distribution. Seed 42, 12-bit precision.

Result. Mean recovered-angle error 8.33×10^{-3} rad, maximum 2.06×10^{-2} rad, against a sampling standard deviation of 1.56×10^{-2} rad and an expected mean absolute error of 1.25×10^{-2} rad under pure sampling noise; worst trial at 1.3 standard deviations. Exactly two CNOTs per trial, independent of θ; disentangling sanity check (second qubit returns to $|0\rangle$) passes with probability 1.000 in all 20 trials. Timing: encryption 6.7 ms, decryption 63.6 ms, encrypted evaluation $\approx$ 3.06 h per trial, i.e. $\approx$ 1.5 h per encrypted CNOT.

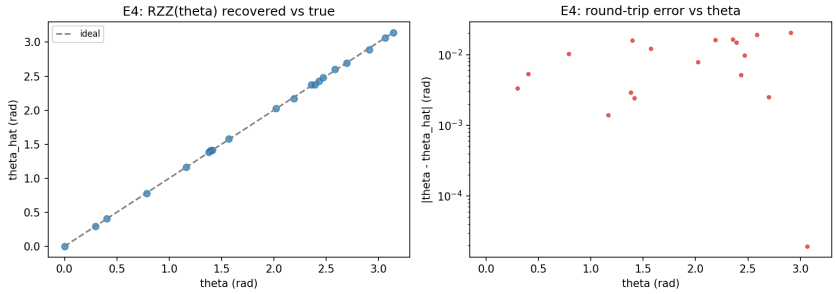


Figure 4: E4: recovered versus true θ for $R_{ZZ}(\theta)$ through the encrypted pipeline, with the shot-noise band. The encrypted entangler adds no angle error resolvable at 4096 shots.

G.5 E5: Noise-budget falsification

Claim. Encryption noise does not act as a beneficial regulariser.

Procedure. Re-run the encrypted arm at 8-bit precision – a $16\times$ larger quantisation step – using E1's per-seed methodology so that each 8-bit run starts from identical weights and data order to the corresponding 12-bit run. The 12-bit values are read from E1's result file rather than re-measured, so the comparison is genuinely paired.

Result. 8-bit 1.3442 ± 0.7230 against 12-bit 1.3436 ± 0.7229; per-seed differences -0.0003, $+0.0048$, -0.0016, -0.0012, $+0.0013$; mean $+0.0006$, 95% CI $[-0.0026, +0.0038]$; $t = 0.511$, $p = 0.636$; Wilcoxon $W = 7.0$, $p = 1.0$. The prediction implied by the regulariser reading – more noise, more benefit – does not hold. Job `266019`.

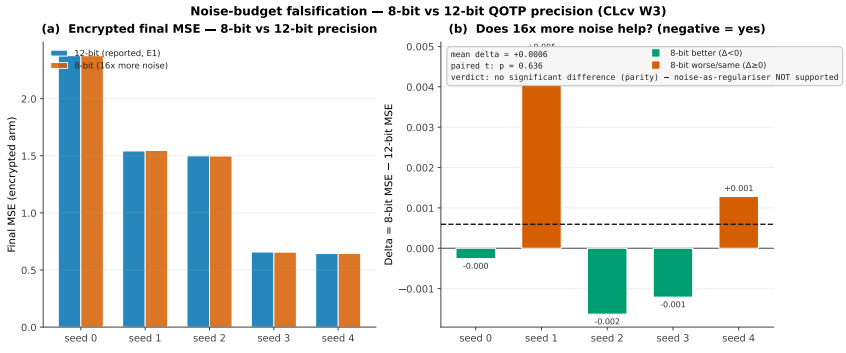


Figure 5: E5: 8-bit versus 12-bit encrypted arms, paired by seed. A $16\times$ increase in quantisation noise produces no detectable change in final loss.

G.6 E6: Cross-dataset scale study

Claim. Client scaling, horizon and precision–bandwidth behaviour are not artefacts of one dataset or of the originally tested client range.

Procedure. Three sub-sweeps – client scaling $K \in \{3, 5, 10, 20\}$ at 10 rounds; horizon $\in \{20, 35, 50\}$ rounds; precision $\in \{8, 10, 12, 14\}$ bits – run identically on California Housing (Pace & Barry, 1997) and Wine Quality (Cortez et al., 2009), five seeds per point, secure stack, 12-bit unless swept.

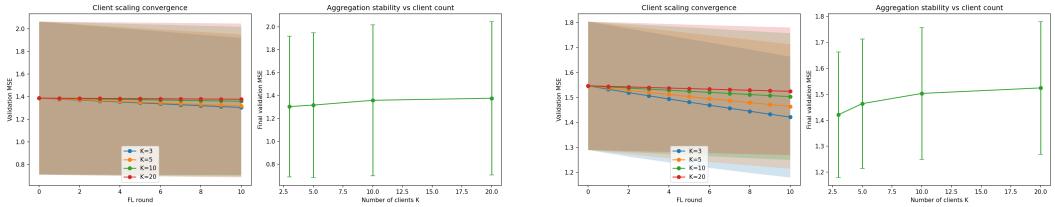
Result. Table 7 and Table 8. Growth with client count is mild and monotone (5.6% and 7.3% for a 6.7× increase in participants); longer horizons improve both the mean and the seed spread; the precision–bandwidth relation is identical across datasets, as a protocol-level property should be. Jobs `265935`, `266315` (California, split across two submissions after an unrelated cluster filesystem failure; configuration parity between the two is recorded in the result metadata) and `265936` (Wine, single job).

Table 7: Cross-dataset scale study, final validation MSE (mean ± s.d. over five seeds), secure stack at 12-bit precision. Client scaling at 10 federated rounds; horizon sweep at three clients.

	Clients K				**Rounds**		
Dataset	3	5	10	20	20	35	50
California Housing	1.3035	1.3166	1.3583	1.3763	1.2285	1.1395	1.0831
	±0.614	±0.633	±0.659	±0.670	±0.541	±0.444	±0.374
Wine Quality	1.4214	1.4641	1.5034	1.5246	1.3148	1.1963	1.1275
	±0.242	±0.249	±0.254	±0.255	±0.205	±0.141	±0.095

Table 8: Precision, ciphertext size, reconstruction error and total protocol traffic over a 50-round, three-client schedule. Traffic and ciphertext size are dataset-independent by construction and reproduce identically on both datasets. Angle MAE here is measured on 256 probe angles; the independent float64 measurement of Table 1 gives 1.52×10^{-4} at 12 bits, consistent within the probe-set difference.

Precision bits	8	10	12	14
Ciphertext size	100.0 KB	125.0 KB	150.0 KB	175.0 KB
Angle MAE (rad)	2.600×10^{-3}	6.585×10^{-4}	1.676×10^{-4}	4.051×10^{-5}
Total traffic	4,375 MB	5,430 MB	6,485 MB	7,539 MB



(a) California Housing, $K \in \{3, 5, 10, 20\}$.

(b) Wine Quality, $K \in \{3, 5, 10, 20\}$.

Figure 6: E6: client scaling replicates on an unrelated dataset, with no aggregation blow-up at 20 clients.

G.7 Hardware experiments

Claim. The encrypt → rotate → decrypt round trip preserves single-qubit circuit behaviour on physical hardware, and the encryption layer's own contribution is below the device error floor.

Procedure. Sample rotation angles; for each trial run encrypted state preparation, an encrypted $R_x(\theta)$, and decrypt-and-run; compute circuit-level fidelity from measured Z-basis counts against the analytic reference; separately measure a homomorphic-layer-only round trip to isolate arithmetic noise. The matched unencrypted control repeats the identical seeded sampling scheme and shot budget (100 angles, 4096 shots) on the same device with a calibration snapshot recorded alongside.

Assumptions. The Z-basis estimator is a lower bound on state agreement. Simulator and hardware tracks use different shot budgets and are reported separately.

Result. Encrypted $F = 0.9918$ on `ibm_fez`; matched unencrypted control $F = 0.99957$ (s.d. 0.00107, min 0.99312); simulator $F = 0.9989$; homomorphic-layer-only angle MAE 1.55×10^{-4} rad. The difference between encrypted and control, 0.0078, is comparable to the device's median readout error at the time (0.0083; two-qubit gate error 0.0029).

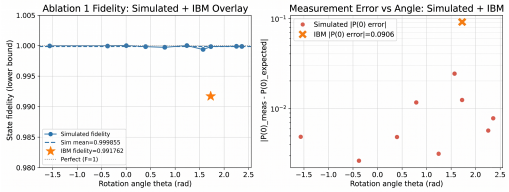


Figure 7: Round-trip state fidelity for $R_x(\theta)\,|0\rangle$ under encrypt $\rightarrow$ rotate $\rightarrow$ decrypt, simulator and `ibm_fez`.

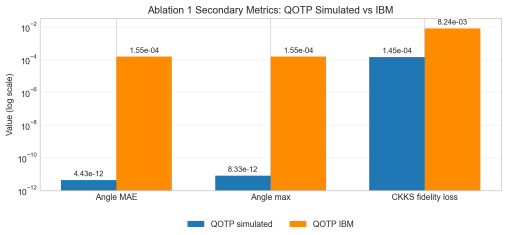


Figure 8: Homomorphic-layer angle-recovery error for the same round trip; an order of magnitude below the device gate-error floor.

G.8 Supporting ablations

Ablation 2: single-client utility preservation. Round-tripping a trained client's angles through encryption and re-evaluating leaves validation MSE unchanged in the reported run – the single-client special case of E1.

Ablation 3: multi-round error accumulation and runtime. Repeated round trips across local training rounds show no observable angle drift; the encryption stage dominates per-round runtime (≈ 167 s against ≈ 76 s of local training).

Ablation 4: homomorphic aggregation error. Superseded, and made precise, by E2: see Table 1.

Ablation 5: encryption scope versus utility. Three matched conditions (plaintext, quaternion-only, full homomorphic) under one federated scaffold. Single-seed differences between conditions fall well inside the between-seed spread measured in E1 and are therefore not interpretable as treatment effects; the paired design of E1 supersedes this comparison.

Ablation 6: precision sweep. Error decreases sharply with precision and runtime increases; the two-backend version is E2.

Ablation 7: secure-refresh frequency. Sweeping refresh interval $k \in \{1, 2, 5, 10\}$ against a no-refresh baseline, the no-refresh configuration is best in the current implementation (MAE $= 1.24 \times 10^{-4}$ rad against 2.7–4.3 rad for periodic cadences), which is why the protocol aggregates in a single pass. This characterises the present secure-runtime implementation, not all homomorphic backends.

H Cost Models

H.1 Hardware profile

Calibration timings for the deployed workflow: $T_{\text{rot,qotp}} \approx 8856.5$ ms against $T_{\text{rot,pauli}} \approx 11844.3$ ms; $T_{\text{to,qotp}} \approx 5487.5$ ms against $T_{\text{to,pauli}} \approx 9221.1$ ms; analytic crossover $p^\star \approx 0.970$. Descriptive profiling rather than a controlled sweep.

H.2 Interaction versus bandwidth

Protocol accounting over 50 federated rounds and three clients gives 50 sequential exchanges in total for this protocol against $1{,}850$ (Pauli-OTP) and $3{,}650$ (blind delegation), at 6.33 GB total traffic against ≈ 156 MB. Table 9 converts both into wall-clock terms. Fixed ciphertext size and message accounting are assumed; totals are model-based rather than network traces.

Table 9: Where non-interactivity pays. Left: time to move this protocol's 45.3 MB per federated round. Right: the latency floor of interactive baselines at 37–73 sequential exchanges, before any payload moves.

Link	This protocol (payload time)	Interactive baselines (latency floor)
1 Gbps, 50 ms RTT	0.36 s	1.9–3.7 s
100 Mbps, 50 ms RTT	3.6 s	1.9–3.7 s
10 Mbps, 50 ms RTT	36 s	1.9–3.7 s
any link, 1 ms RTT	as above	0.04–0.07 s

H.3 Solovay–Kitaev comparison

Per-rotation compute reduction against Solovay–Kitaev decomposition, from gate counts at matched target precision: $10\times$ at $\epsilon = 10^{-2}$, $20\times$ at 10^{-4}, $30\times$ at 10^{-6}. This is an accounting estimate over gate counts, not an end-to-end wall-clock measurement, and should not be read as one.

H.4 Pure-quantum diagnostic

A four-qubit data-reuploading circuit with only minimal classical output calibration (Ballester et al., 2025) fails to converge under the same encrypted federated pipeline: validation MSE rises monotonically from 1.24 to 1.56 across five rounds. Together with Ablation 7 this makes the hybrid co-design an architectural necessity rather than a convenience: classical components carry the gradient signal through landscapes a small variational circuit cannot navigate (McClean et al., 2018), and single-pass aggregation avoids mid-training key refreshes that current secure runtimes handle poorly.

I Scale Tests

The scale suite (client scaling I, horizon I, bandwidth I) is reported at full breadth in E6; this section records the original single-dataset sweeps for completeness.

Multi-seed variance. Five seeds, ten rounds, three clients. This sweep runs the two arms sequentially over shared loaders, so the arms differ in initialisation and batch order as well as in encryption; the resulting apparent gap is initialisation variance and is superseded by the paired design of E1, which measures the encryption effect at $+9 \times 10^{-6}$ MSE against a between-seed spread of ≈ 0.72.

Client scaling. Final MSE grows mildly and monotonically with participant count; tail standard deviation *improves* at larger K ($0.0178 \to 0.0061$), i.e. late-round behaviour is smoother with more clients. Extended to $K = 20$ on two datasets in Table 7.

Extended horizon. Final MSE decreases monotonically with horizon and cross-seed spread contracts, on both datasets (Table 7).

Bandwidth sensitivity. Monotonic trade of angle error against traffic (Table 8). Communication totals are protocol byte accounting, not network traces.

J Implementation and Security Profiles

Cryptographic backends. The executed stack implements the GSW-style LWE constructions of Ma & Li (2022, Schemes 5.1 and 5.2) over fixed-point bit-level arithmetic on the quaternion

components $t = (t_1, t_2, t_3, t_4)$ and Euler angles (α, β, γ), combined with pairwise Diffie–Hellman masking (Diffie & Hellman, 1976) and information-theoretic mask cancellation for the secure-aggregation harness. A parallel CKKS path over OpenFHE (Al Badawi et al., 2022) (ring dimension 2^{16}, multiplicative depth 40, 50-bit scaling modulus, `FLEXIBLEAUTO`, `HYBRID` key switching) is used for the backend cross-check of E2. Security labelling should therefore track the path: the executed stack is LWE-based, the CKKS instantiation RLWE-based (Regev, 2009; Lyubashevsky et al., 2010).

Table 10: Threat model. Decryption capability is shared across clients and withheld from the server (the aggregation object exposes only add, scale and serialise). Threshold key generation is the drop-in hardening for the peer row; see Appendix J.

Party	**Observes**	**Does not obtain**
Cloud QPU	Quaternion-masked states	The underlying state. Masking is information-theoretic with per-client, per-use keys removed by their owner.
Aggregation server	Ciphertexts, message sizes, client weights n_k/n	Any plaintext. Decryption raises an error by construction.
Peer client	The decrypted global aggregate	Nothing further *by protocol*; under the shared-key model a peer holding another's ciphertext could decrypt it, the standard shared-key FHE-FedAvg assumption.
Network adversary	Authenticated ciphertext traffic	Plaintexts; undetected tampering.

Parameter profiles, and what the reported runs are entitled to claim. The default profile is a structural-verification setting ($n = 1$, $q = 256$, $m = 8$, $\beta_{\text{init}} = 4$), which satisfies the noise bound $\beta \geqslant 2\sqrt{n}$ of Ma & Li (2022) and is chosen for auditability, not for security level. Production profiles ($n = 64$, $q = 2^{32}$, $\beta_{\text{init}} = 16$; and $n = 256$, $\beta_{\text{init}} = 32$) are available. *No security property is claimed for the reported runs*: they establish protocol structure, exactness and depth, and the security argument rests on the scheme, not on the profile used for verification.

Angle extraction and the entangler path. Euler-angle extraction from $\mathrm{Enc}(t)$ uses CORDIC vectoring in fixed point, fully oblivious (zero scalar decryptions), with no singularity at the origin and precision controlled by $\max(12, \min(24, \text{bits} + 2))$ iterations. The encrypted-rotation pipeline is a levelled circuit whose level budget is $3kL + 1$ for k precision bits and L circuit levels (37 levels at the default $k = 12$, $L = 1$); level promotion refreshes ciphertexts as needed and no periodic carry bootstrapping is required. The Boolean sub-layer is where the ≈ 1.5 h per encrypted CNOT of §5.5 is spent; a Boolean-native backend (Chillotti et al., 2020) is the natural remedy.

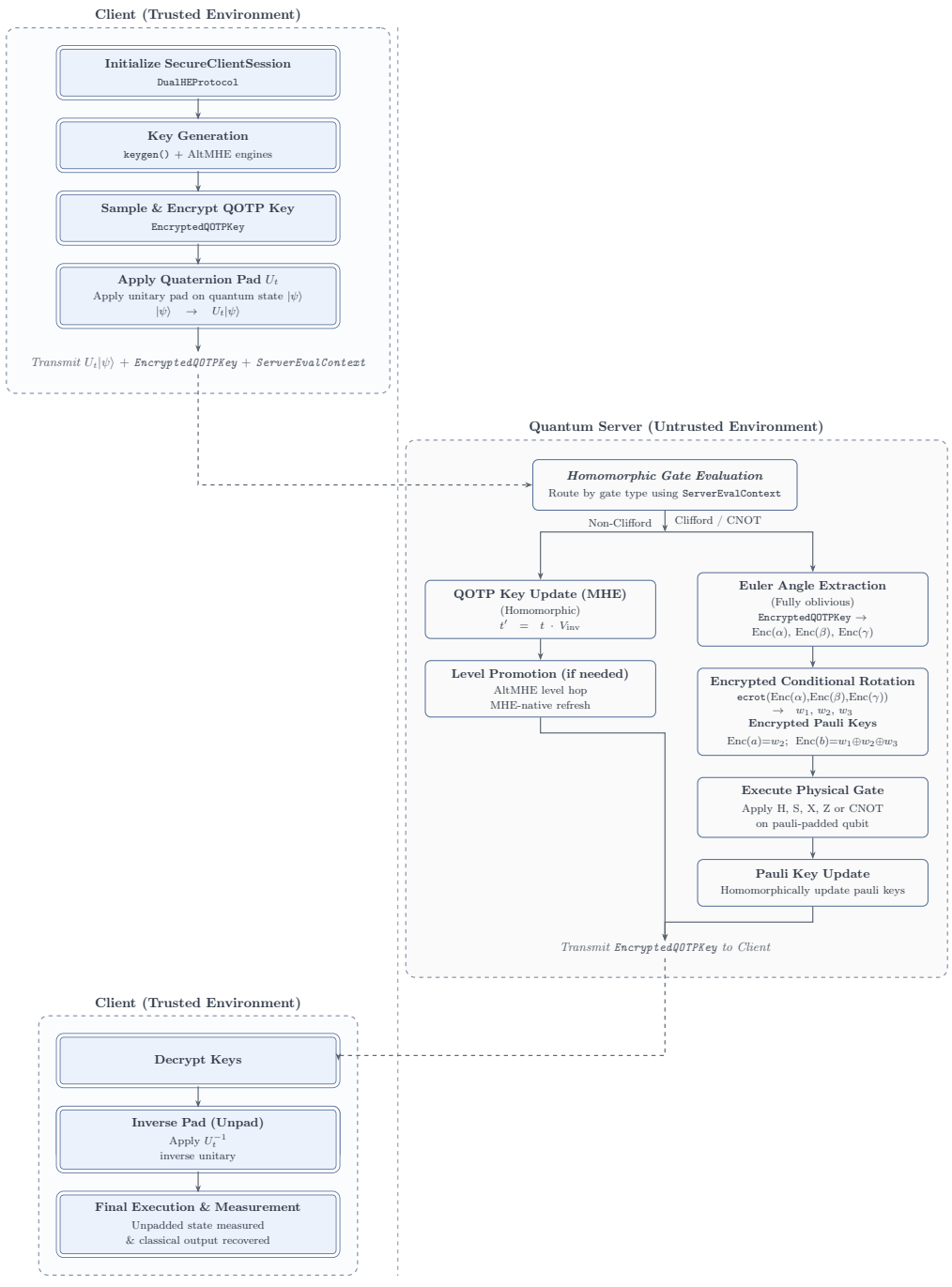


Figure 9: End-to-end implementation flow for the quaternion track.

K Extended Discussion and Limitations

K.1 Bandwidth is the new bottleneck

The dominant cost moves from sequential round-trips to per-round bandwidth: 6.33 GB over a 50-round run, driven by 22.67 MB ciphertexts per client per round, with single-round encryption ($\approx$ 167 s) exceeding local plaintext training ($\approx$ 76 s). The crossover analysis of §5.8 makes the

boundary explicit rather than rhetorical: the trade pays above roughly 100 Mbps at wide-area latency and does not pay below it or on low-latency links. Precision recovers about a third of the traffic at no measured utility cost (§5.3), and ciphertext compression is the obvious next lever.

K.2 Security, in two layers

The quantum layer provides information-theoretic security at the state level; the classical layer provides computational security under (Ring-)LWE. The system-level guarantee is bounded by the weaker link, so a future polynomial-time attack on the lattice assumption would compromise aggregation even though quantum ciphertexts remain opaque; practitioners should size parameters to their threat model. Decryption capability is shared across clients, which is the standard shared-key homomorphic-FedAvg assumption and the one place where a peer is more powerful than the protocol's spirit suggests; threshold key generation (Mouchet et al., 2021) removes this at the cost of an additional key-setup phase.

K.3 Robustness against malicious clients

A blind server cannot inspect updates, so poisoned updates are undetectable by construction – the standard Byzantine pathology, sharpened by encryption. Robust aggregation compatible with the homomorphic layer, along the lines of El Maouaki et al. (2025), is the natural next component. Until then the protocol suits institutionally vetted federations where client identity is verified.

K.4 Implementation scope

The core protocol path is implemented and validated end to end; full application-layer integration into a production federated stack (Beutel et al., 2020) is incomplete, and several planned cost-model studies (evaluation time versus CNOT fraction, depth scaling at fixed CNOT fraction, gate ordering) were not executed. Hardware validation is constrained by quantum-cloud budgets: at prevailing runtime pricing (IBM, 2026) multi-seed hardware sweeps at the scale conventional in classical ML are structurally out of reach for most groups. We record this as a limitation of the evidence, and separately as an open question for the community about whether empirical norms calibrated for classical compute are the right ones to apply to hardware-constrained research.

L Broader Impact

Privacy. The construction strengthens confidentiality in federated training: even an adversary who later breaks the classical layer obtains no transcript of the delegated quantum computation, which existed only as transient mixed states. Organisations should nonetheless evaluate whether the residual lattice-based classical-layer security matches their threat model; the classical scheme is replaceable without touching the reduction.

Auditability. Blindness prevents compliance auditing as well as surveillance. In jurisdictions requiring model inspection or data-access rights, the protocol's opacity may conflict with existing obligations, and selective-disclosure mechanisms compatible with the security model are a worthwhile line of work.

Adversarial robustness. Under the honest-but-curious model a minority of malicious clients could degrade the global model with no audit signal. Deployment in adversarial environments should wait for homomorphic-compatible robust aggregation.

Access and equity. Physical validation required institutional quantum-cloud and HPC access. As the approach matures, benefits risk accruing to already well-resourced actors. The bandwidth profile compounds this: community clinics, mobile devices and low-connectivity nodes – exactly the settings where federated privacy guarantees are most socially valuable – are least able to absorb the cost. Ciphertext compression and lightweight backends are equity concerns as much as performance ones.

Honest framing of readiness. The hybrid design is a necessity, not a convenience: the pure-quantum diagnostic degrades monotonically, and periodic key refresh in the current runtime is catastrophic relative to single-pass aggregation. Readers should not extrapolate the reported numbers beyond the configuration validated here.

M USE OF LARGE LANGUAGE MODELS

Large language models assisted with code development, figure generation, and polishing the manuscript's language. AI was also used for literature retrieval and discovery (e.g., to initially discover papers and place our claims in the broader context of the field). Finally, AI tools helped in generating the synthetic datasets used for the small-scale empirical evaluations, such as the synthetically scattered client rotation angles ($\sigma \in \{0.05, 0.2, 0.5, 1.0\}$ rad) used in the aggregation-geometry stress tests, and the seeded random angles generated for the hardware evaluations. All experimental design, final analysis, and scientific claims remain the authors' own work.