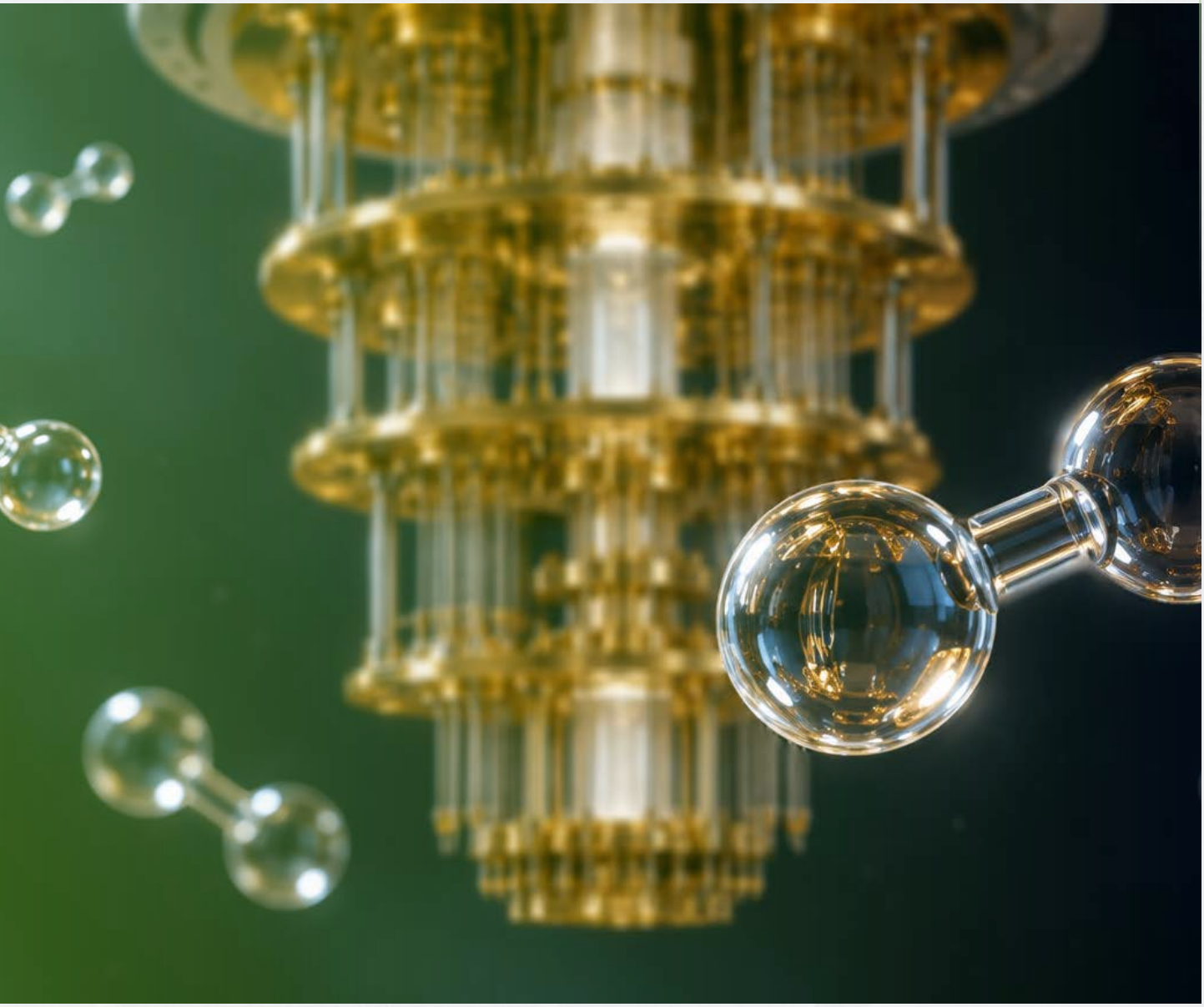
COMPAL



Benchmarking the Variational Quantum Eigensolver for H_2 on IQM Quantum Hardware

A Cross-Platform Reproducibility Study Following a Standardized VQE Benchmarking Methodology

November 2026

EXECUTIVE SUMMARY

This white paper reports a Compal benchmarking study of the Variational Quantum Eigensolver (VQE) applied to the ground-state energy of the hydrogen molecule (H_2) on IQM's superconducting quantum processors (Garnet / Emerald), accessed through NVIDIA's CUDA-Q platform. The study follows the standardized, transparent benchmarking methodology introduced by Larrucea, Oliv, and Lorenz for reference VQE calculations of H_2 on IBM Quantum hardware [1], enabling a direct, apples-to-apples comparison across two different superconducting-qubit vendors.

H_2 is encoded using three fermion-to-qubit mapping strategies of increasing circuit simplification — a 4-qubit Jordan-Wigner (JW) mapping, a 2-qubit parity mapping with two-qubit reduction (PF), and a fully symmetry-tapered 1-qubit parity mapping (PT). Across all three encodings, we report real-hardware VQE energies, deviation from the exact reference energy, and a noise-floor analysis that isolates hardware error from optimizer error by evaluating the circuit at the exact ground-state angle.

Key findings:

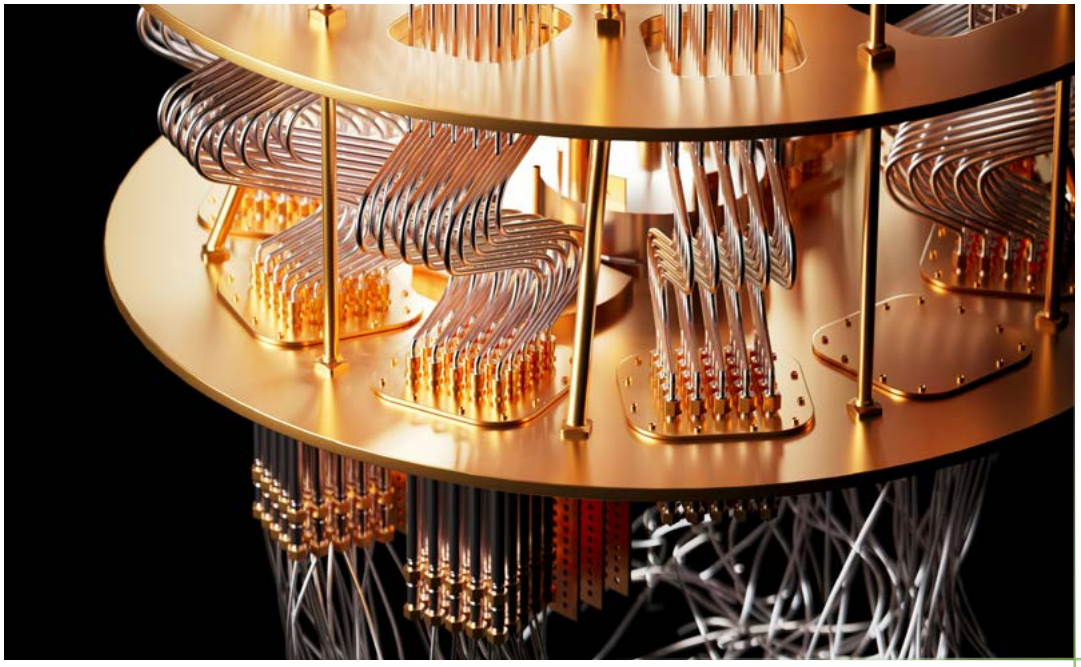
- Circuit simplification is the dominant lever for hardware accuracy on IQM hardware, exactly as reported for IBM hardware in [1]: across the representative experiments, the 1-qubit PT mapping achieved errors ranging from single-digit to low-tens of mHa against the exact reference, whereas the 2-qubit PF mapping produced errors in the tens-of-mHa range. The unreduced 4-qubit JW mapping, by contrast, produced a 641 mHa error—effectively yielding no chemically meaningful signal.

- A noise-floor analysis at the fixed exact ground-state angle (isolating hardware noise from optimizer error) shows a stable, systematic upward bias on IQM hardware — never below the reference energy, with low run-to-run variance — indicating a stable hardware bias rather than random sampling noise. This dataset does not isolate which hardware mechanism (gate error vs. readout error) drives the bias; that would require further characterization.

- State-preparation convention ("frame A" vs. "frame B" encodings of the same tapered Hamiltonian) changed the 1-qubit result by roughly 2.5–3× in accuracy, pointing to a low-cost, hardware-informed optimization: favoring the basis state closer to the physical $|0\rangle$ to exploit the amplitude asymmetry of IQM's readout error.

- Overall, the results on IQM hardware support the central finding of the IBM reference study [1] — that tapered mappings, rather than hardware brand, are a more consistent factor in near-term VQE accuracy, while also showing differences in the residual noise floor across hardware vendors.

These results establish a transparent, reproducible reference dataset for VQE on IQM hardware, directly comparable to the open IBM dataset in [1], and lay the groundwork for continued Compal quantum-chemistry benchmarking on IQM hardware.



AUTHORS & CONTRIBUTORS

Yi Chou, Compal Electronics Inc.
Kuo-Wei Huang, Compal Electronics Inc.
Yiching Chen, Compal Electronics Inc.

COMPAL

Compal Electronics Inc.
VQE circuit design, CUDA-Q implementation,
IQM hardware execution, benchmarking analysis.

1. Introduction

1.1 The Challenge

The Variational Quantum Eigensolver is one of the most promising near-term algorithms for quantum chemistry, but its practical accuracy on real Noisy Intermediate-Scale Quantum (NISQ) hardware is governed by a dense set of workflow choices — qubit mapping, ansatz, shot count, backend calibration, and error-mitigation settings — whose combined effect is difficult to predict from theory alone. Larrucea, Oliv, and Lorenz [1] addressed this gap for IBM Quantum hardware, publishing a hardware-validated, tutorial-aligned reference dataset for H_2 that quantifies how these choices trade off against accuracy, runtime, and cost, and found that circuit simplification through tapered qubit mappings is the most consistent driver of accuracy gains.

As Compal's AI team builds out a quantum-chemistry pipeline on IQM hardware — spanning VQE for small molecules today and for larger active spaces going forward — an open question is whether these IBM-hardware findings generalize to a different superconducting architecture with different native gates, connectivity, and calibration. Answering this question tells Compal whether its workflow can be benchmarked consistently on a second hardware platform, and gives practitioners a like-for-like reference across hardware platforms.

1.2 What This Study Does

Using the same H_2 benchmark system, the same accuracy metric (deviation from the exact ground-state energy, in mHa), we replicate and extend the reference-VQE methodology on IQM Garnet / Emerald hardware via CUDA-Q. The result is a directly comparable, cross-vendor reference dataset that:

- Validates whether the accuracy ordering established on IBM hardware (fewer qubits / gates → higher accuracy) holds on IQM hardware.
- Decomposes the residual hardware error into an optimizer-independent noise floor, isolating the contribution of gate and readout error from that of the classical optimization loop.
- Documents practical, low-cost mitigations — shot budget selection and basis-state convention — that measurably improve accuracy without additional error-mitigation overhead.

2. Technology Overview & Methodology

2.1 Benchmark System

The benchmark system is the electronic ground state of H_2 in a minimal basis, matching [1]. The molecular Hamiltonian is constructed via PySCF / OpenFermion and mapped to qubits using three encodings of increasing circuit simplification:

Mapping	Qubits	Description
JW	4	Jordan-Wigner mapping, no symmetry reduction
PF	2	Parity mapping with two-qubit reduction (particle-number / spin symmetry)
PT	1	Parity mapping, fully symmetry-tapered to a single qubit

Table 1. Fermion-to-qubit mappings evaluated in this study.

2.2 Hardware & Stack

- **QPUs**: IQM Garnet and IQM Emerald superconducting processors, accessed via IQM Resonance.
- **Compilation / execution**: NVIDIA CUDA-Q and cudaq-solvers.
- **Ansatz**: Qiskit's unitary coupled-cluster (UCC) ansatz, matching the construction used in [1]. After qubit tapering, the UCC excitation generator reduces to a small number of parameterized single-qubit rotations — a single RY rotation for the fully-tapered 1-qubit (PT) encoding — rather than being a generic, problem-agnostic template.
- **Classical optimizer**: COBYLA for shot-based hardware runs; exact diagonalization used for the reference energy.
- The 1-qubit tapered Hamiltonian was evaluated under two basis-state conventions — "frame A" (Qiskit ParityMapper pipeline) and "frame B" (OpenFermion SCBK / manual tapering) — which are mathematically equivalent but assign the computational basis states differently, allowing us to test sensitivity to state-preparation convention on real hardware.

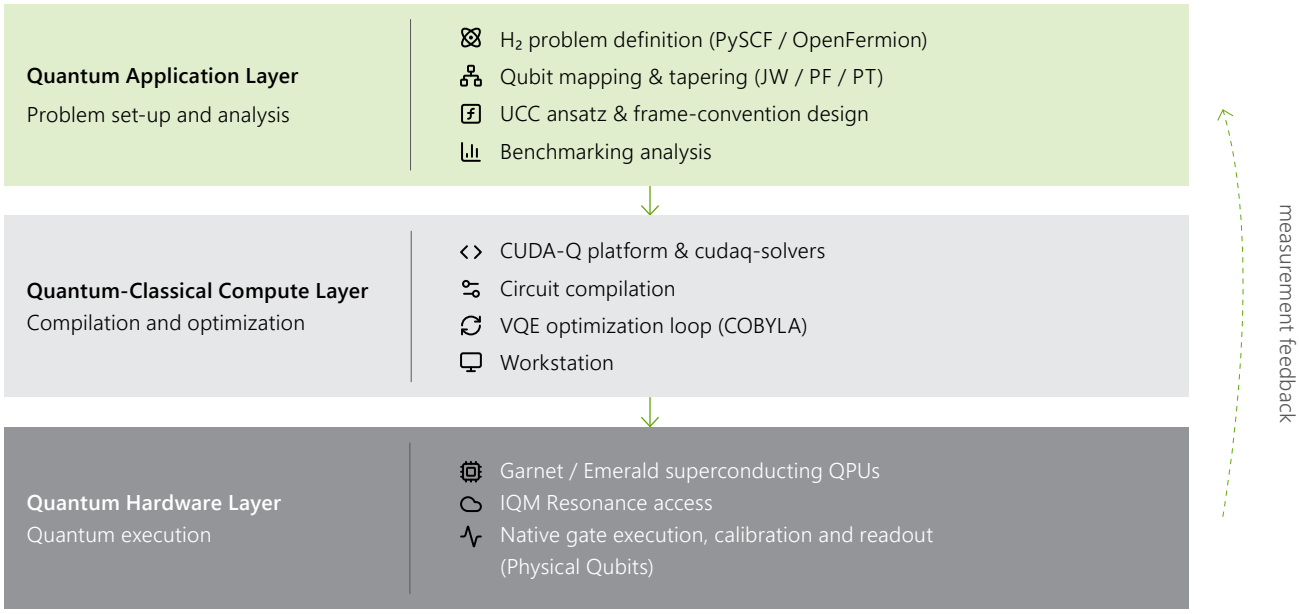


Figure 1. Tools used in this study: problem set-up and analysis, compilation and optimization (NVIDIA CUDA-Q), and quantum execution (IQM Garnet / Emerald via IQM Resonance).

2.3 Comparability with the IBM Reference Study

To keep the comparison meaningful, this study mirrors [1] in benchmark system, mapping choices, ansatz choices (Qiskit's UCC construction) and accuracy metric (|E − E_ref| in mHa). Where [1] models circuit reliability from IBM's reported per-gate error rates, we apply the same functional form and the same published constants as a shared theoretical baseline, then measure how each platform's real hardware performs against it.

[1] estimates, from IBM's own gate-count and calibration data, an error probability of roughly one-in-three for the unreduced 4-qubit (JW) circuit versus roughly one-in-27 for the 2-qubit (PF) circuit — motivating the same circuit-simplification strategy evaluated here on IQM hardware.

3. Real-Hardware Results on IQM

3.1 1-Qubit (PT Mapping) — Tapered Reference

Reference energy E_ref (PT mapping, fully tapered, 1-qubit) = −1.137306 Ha.

Frame	Shots	E (Ha)	|E − E_ref| (mHa)	Measured state
Frame A	16,384	−1.13289344	4.41	0.9948|0⟩ + 0.1018|1⟩
Frame A	1,024	−1.132626	4.68	0.9962|0⟩ + 0.0874|1⟩
Frame B	16,384	−1.12277628	14.53	−0.1001|0⟩ + 0.9950|1⟩
Frame B	1,024	−1.12591159	11.39	−0.0988|0⟩ + 0.9951|1⟩

Table 2. IQM real-hardware VQE results, 1-qubit tapered (PT) H_2 Hamiltonian.

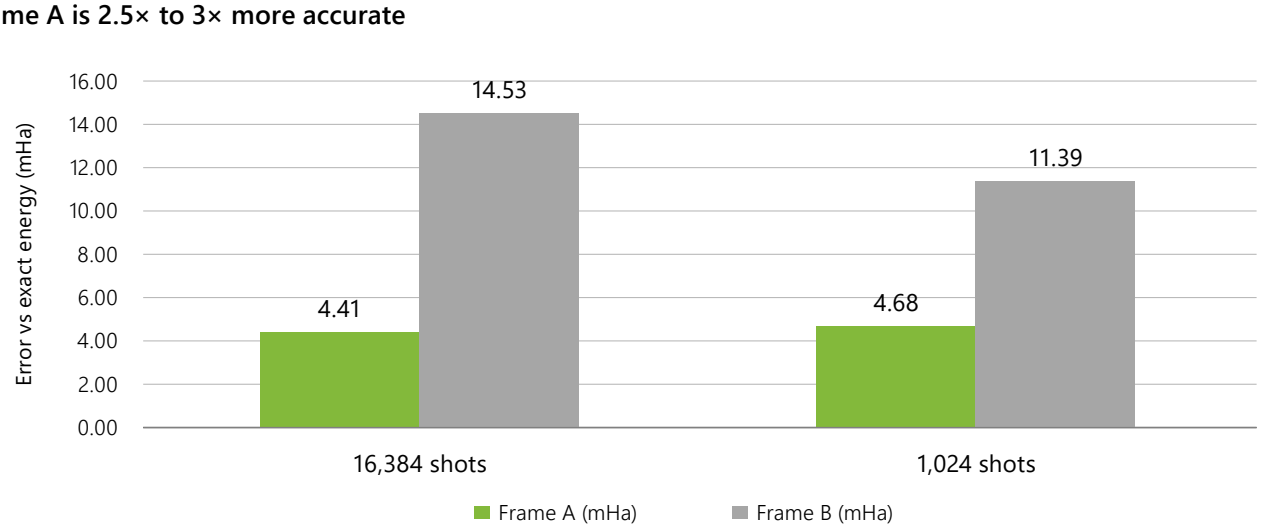


Figure 2. Frame A vs frame B, 1-qubit (PT) mapping on IQM hardware: error at 16,384 and 1,024 shots. Data from Table 2.

Consistent with the shot-count trade-off reported in [1] (where 1,024 nominal shots is identified as a practical compromise between stability and cost), we likewise find only modest accuracy loss when reducing from 16,384 to 1,024 shots (≤ 0.3 mHa for frame A) against a 16× reduction in circuit executions.

Frame A consistently outperforms frame B by roughly 2.5–3× in accuracy. Both encode the same physical Hamiltonian, so the gap is attributable to which computational basis state carries the dominant amplitude of the ground state: frame A places it on |0⟩, while frame B places it on |1⟩. This is consistent with superconducting-hardware readout being asymmetric — T1 decay during dispersive readout biases measurements toward |0⟩ — so a state-preparation convention that aligns the dominant amplitude with |0⟩ is an essentially free accuracy gain, worth adopting as a standing convention for tapered VQE encodings on this class of hardware.

3.2 Comparison with IBM Reference Results

We additionally executed the same 1-qubit tapered circuit and frame conventions on IBM hardware (backend ibm_fez), using the same Eerr = |E − E_exact| accuracy metric as [1] (there presented as Eerr × 10, in atomic units). This run yielded Eerr × 10 = 0.0651 (frame A) and 0.227 (frame B) — compared to 0.044 (frame A) and 0.145 (frame B) on the same scale from our IQM run (Table 2). Both platforms show the same frame A / frame B ordering, and the IQM result is somewhat more accurate in this single head-to-head comparison; given that IBM's own study documents substantial backend-to-backend and calibration-drift variability for nominally identical configurations, this single-run difference should be read as a data point rather than a general accuracy ranking between vendors. This caution is corroborated by [1]'s own repeated-run data: for the PT mapping at 1,024 shots without mitigation, Eerr × 10 on ibm_fez alone spans 0.095–0.267 across three repeated runs, and up to two orders of magnitude (0.00549–0.6671) across backends and repeats more broadly — variability comparable in size to, or larger than, the single-run gap observed here between IQM and IBM.

3.3 2-Qubit (PF Mapping)

Reference energy E_ref (PF mapping, 2-qubit) = −1.137306 Ha.

Shots	E (Ha)	|E − E_ref| (mHa)
4,096	−1.08143908	55.87
1,024	−1.07436804	62.94

Table 3. IQM real-hardware VQE results, 2-qubit (PF) H_2 Hamiltonian.

3.4 Noise Floor at the Ground-State Angle (2-Qubit)

To separate hardware noise from optimizer error, we evaluate the circuit at the exact ground-state angle (θ* = −0.1118) rather than the VQE-optimized angle, over 3 repeats of 16,384 shots each.

Run	E (Ha)	|E − E_ref| (mHa)
1	−1.072858	64.45
2	−1.076327	60.98
3	−1.075223	62.08

Table 4. Noise-floor measurements, 2-qubit encoding, at the exact ground-state angle.

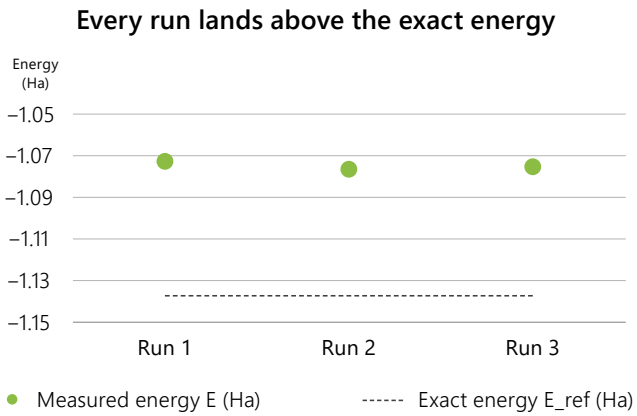


Figure 3. Noise floor, 2-qubit encoding: measured energy of the three runs at the exact ground-state angle (θ* = −0.1118) compared with the exact energy E_ref (dashed line). Data from Table 4.

Reference points from [1]: noisy Aer simulator VQE baseline (Table 4) in [1] ≈ 76.9 mHa; real quantum hardware (Figure 7 in [1]) ≈ 110 mHa (ibm_aachen), an approximate reading from the bar chart. IQM's measured noise floor at this encoding, ≈ 62.5 mHa, is of the same order of magnitude as these VQE-based figures.

- All 3 runs land above E_ref — never below — consistent with a systematic bias rather than random sampling noise.
- The small spread across repeats (σ ≈ 1.8 mHa) indicates the bias is stable: more shots would sharpen the measured value but would not close the gap to E_ref.
- Because θ* is fixed and not VQE-optimized, this residual is attributable entirely to hardware noise (gate and readout error), not optimizer behavior.

3.5 4-Qubit (JW Mapping) — Unreduced Encoding

Reference energy E_ref (JW mapping, 4-qubit) = −1.137306 Ha.

Shots	E (Ha)	|E − E_ref| (mHa)
1,024	−0.4963246287	640.98

Table 5. IQM real-hardware VQE result, unreduced 4-qubit (JW) H_2 Hamiltonian.

At 641 mHa, the 4-qubit VQE result carries no usable chemical signal — an error roughly 400× the 1.6 mHa chemical-accuracy threshold. This is the direct hardware manifestation of the ≈31.5% circuit error probability [1] estimates for this gate count (138 single-qubit and 48 two-qubit gates): close to one in three circuit executions is expected to be corrupted by at least one gate error, sufficient to destroy the VQE signal entirely without any qubit tapering.

3.6 Noise Floor at the Ground-State Angle (4-Qubit)

Same protocol as Section 3.4 (θ* = −0.11176851, 3 repeats × 16,384 shots), applied to the 4-qubit encoding.

Run	E (Ha)	|E − E_ref| (mHa)
1	−0.470133	667.17
2	−0.492856	644.45
3	−0.487541	649.76

Table 6. Noise-floor measurements, 4-qubit encoding, at the exact ground-state angle.

Reference points from [1]: noisy Aer simulator VQE baseline (Table 4) in [1] ≈ 554.7 mHa; real quantum hardware (Figure 7 in [1]) ≈ 400 mHa (ibm_aachen), an approximate reading from the bar chart. IQM's measured noise floor here, ≈ 654 mHa, is of the same order of magnitude as these VQE-based figures.

As with the 2-qubit case, all three runs land systematically above E_ref, with a comparably tight spread, again consistent with a stable hardware bias rather than statistical noise.

4. Cross-Platform Comparison

The central qualitative finding of [1] — that circuit simplification through tapered mappings is the most consistent lever for VQE accuracy on NISQ hardware — holds on IQM hardware as well. Across both platforms, the accuracy ordering PT (1-qubit) > PF (2-qubit) ≫ JW (4-qubit) is reproduced.

Our own hardware-executed VQE results (Section 3) are of the same order of magnitude as the corresponding figures in [1] at both the 2-qubit and 4-qubit encodings.

A second, more actionable finding is that the 1-qubit "frame A vs. frame B" gap (4.4–14.5 mHa vs. the same physical Hamiltonian) is a zero-hardware-cost optimization: choosing the basis-state convention that places the dominant ground-state amplitude on $|0\rangle$ recovers most of the accuracy gap attributable to asymmetric qubit readout. This complements the heavier error-mitigation techniques evaluated in [1] (e.g., resilience-level correction) and is worth adopting as a standing convention in any tapered-mapping VQE workflow, on either platform.

Finally, the systematic (never-below-reference), low-variance character of the noise floor on IQM hardware — observed at both the 2-qubit and 4-qubit encodings — indicates the residual error is dominated by a stable hardware bias rather than shot noise. This dataset does not isolate whether gate error or readout error drives the bias. A stable, systematic bias is still favorable for error mitigation: unlike a purely statistical noise floor, it is in principle correctable.

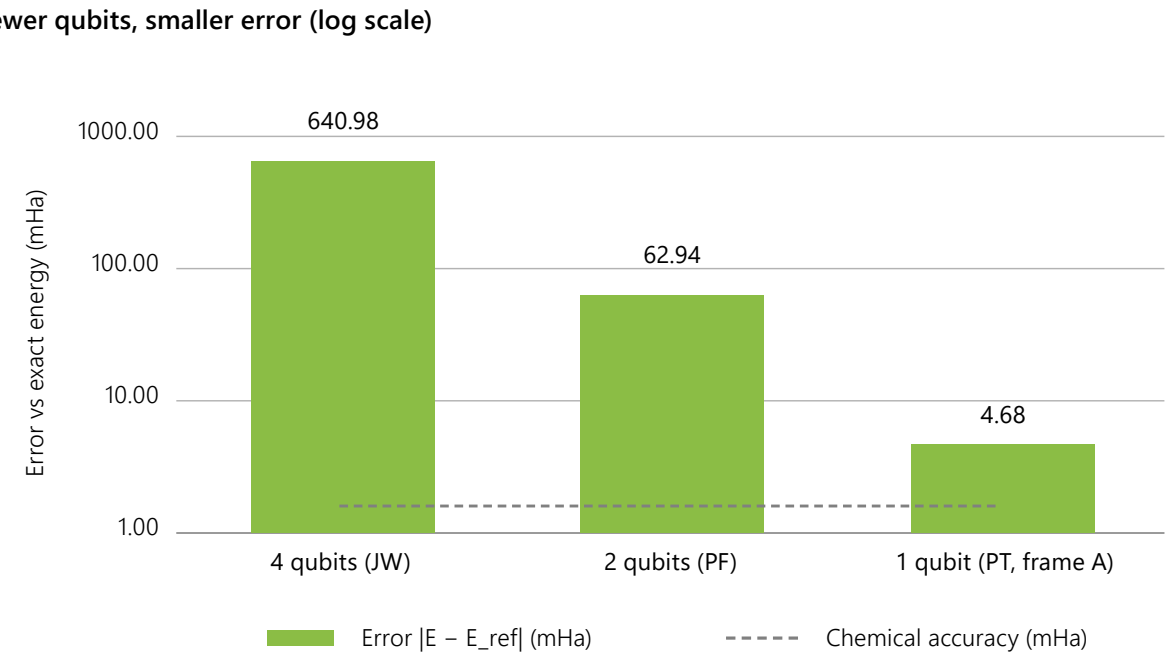


Figure 4. VQE error per mapping on IQM hardware at 1,024 shots (log scale). Dashed line: 1.6 mHa chemical-accuracy threshold. Data from Tables 2, 3 and 5.

5. Impact

- Shows that Compal's VQE workflow can be run and benchmarked on IQM hardware, using a methodology directly comparable to the published IBM reference dataset [1].
- Confirms that qubit-mapping / tapering strategy — not hardware vendor — is the dominant lever for near-term VQE accuracy, a portable and immediately actionable finding for any team building on either platform.
- Delivers a low-cost, hardware-informed convention (basis-state alignment with $|0\rangle$) that improves 1-qubit VQE accuracy roughly 2.5–3× with no additional shots or error mitigation.
- Provides a transparent, reproducible reference dataset for H_2 from Compal's runs on IQM hardware, positioned as a companion to the open IBM dataset in [1] for cross-platform benchmarking by the wider quantum-chemistry community.

6. Future Roadmap

Compal's next steps:

- Extend the standardized benchmarking methodology beyond H_2 to larger active spaces (BeH_2, H_2O).
- Evaluate error-mitigation techniques in Compal's workflow, analogous to IBM's resilience levels, including confusion-matrix-based readout correction and dynamical decoupling to suppress spectator-qubit crosstalk.
- Publish Compal's own shot-count-vs-accuracy-vs-cost trade-off curves from its runs on IQM hardware, mirroring the cost transparency goal of [1].

ACKNOWLEDGMENTS

The authors thank IQM Quantum Computers for access to the Garnet and Emerald superconducting quantum processors used in this study.

REFERENCES

[1] Larrucea, Julen, Marita Oliv, and Jeanette Lorenz. "Accuracy-Cost Trade-offs for Reference VQE Calculations of H_2 on IBM Quantum Hardware." arXiv preprint arXiv:2604.11478 (2026).

IMAGES

Cover illustration created with the assistance of AI. It does not depict an actual quantum computer.

COMPAL

FROM VISION TO REALITY

Innovation Empowered