The Chemistry Performance Envelope

Circuit resources, energy prediction and conditional processor budgets

Pally Sai Tilak
Atomic Cubit, Hyderabad, India · tillaksai@gmail.com

Abstract

Estimating the cost of variational quantum chemistry requires the molecular reference, the compiled circuit and the measurement procedure to be considered together. We examine these quantities using small active-space benchmarks, local quantum-processor result records and a reproducible six-qubit circuit. The circuit contains 1,433 controlled-Z gates and has an ideal energy 3.954 mHa above the supplied Hamiltonian's global minimum. Its recorded processor estimate is 998.220 mHa above that reference and 16.645 mHa from the Hamiltonian's trace average. A model with the archived fitted exponent predicts 929.236 mHa, while a retrospective unit-exponent variant predicts 998.376 mHa without an additive floor. Their different residuals and sensitivity to calibration inputs show why numerical agreement must be assessed against a baseline and the model-selection procedure. In a separate 21-point hydrogen campaign, mean absolute deviations are 9.734, 4.873 and 3.153 mHa for raw, readout-mitigated and combined readout/zero-noise-mitigated estimates. One point is below the 1.6 mHa target; the available repeat intervals straddle it. We formulate a conditional energy budget that separates ideal-state error, noise bias and sampling uncertainty. An illustrative BeH_2 inversion gives an effective gate-error limit of 8.110×10^{-6} under its stated assumptions. The results provide reproducible circuit and energy comparisons and a framework for designing prospective tests of chemistry-specific resource budgets.

Keywords: variational quantum eigensolver; quantum chemistry; circuit resources; noise modeling; error mitigation; reproducibility.

1. Introduction

The variational quantum eigensolver (VQE) estimates molecular energies through a sequence of state preparations and measurements [1]. Qubit count determines the size of the encoded problem, but it does not specify the number of noisy operations, the precision of the estimator or the accuracy of the chosen active space. Two calculations with the same qubit count can therefore have very different resource requirements.

Circuit-based benchmarking addresses part of this problem by assessing performance beyond register size [2]. For chemistry, simulations link tolerable gate noise to the number of entangling gates [3], while circuit-error-sum methods use calibration differences between mappings to support error extrapolation [4]. These approaches motivate a circuit-specific budget. Such a budget also needs a chemical reference: a small energy error is useful only when its reference, sector, geometry and active-space approximation are specified.

We use the term *chemistry performance envelope* for the set of circuit and noise parameters satisfying an explicitly stated energy-budget model. It is a conditional design calculation. Its boundary depends on the ideal trial state, the noise response and the measurement procedure; experimental forecasts require separate validation. This distinction is especially important near the trace average of a Hamiltonian, where substantially different noise models can produce similar energy estimates.

The study brings together three types of evidence. First, compiled active-space circuits describe the trade between gate count and ideal energy. Second, a supplied six-qubit circuit and observable permit direct numerical replay. Third, a hydrogen campaign permits a comparison of mitigation arms and a partial assessment of repeatability. The analysis retains their different references and sources rather than treating all rows as interchangeable observations.

The main results are the circuit and energy replay in Section 6, the full-cohort mitigation comparison, and the conditional inversion in Section 8. The accompanying Supplementary Information [12] gives input definitions, complete numerical tables and record provenance. All computational systems considered here are small enough for classical reference calculations. The molecular labels identify active-space computational targets; they do not establish biological activity or a quantum advantage.

---

2. Hamiltonians, references and circuit conventions

2.1 Molecular operators and qubit mappings

In a finite spin-orbital basis, an electronic Hamiltonian has the form

$$H = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} g_{pqrs} a_p^\dagger a_q^\dagger a_r a_s + E_{\mathrm{nuc}}. \tag{1}$$

The integral convention must match the operator ordering in Eq. (1). A fermion-to-qubit mapping produces a sum of Pauli strings,

$$H = \sum_j c_j P_j, \qquad c_0 = 2^{-n} \operatorname{Tr}(H). \tag{2}$$

Here c_0 is the coefficient of the identity string, including any constant energy shift retained in the supplied operator. For an untapered spin-orbital encoding, $n = 2N_{\mathrm{orb}}$. Symmetry reduction can reduce the register size [5], so the active-space label alone does not determine n.

With $|0\rangle$ denoting an empty orbital and $|1\rangle$ an occupied orbital, the Jordan-Wigner convention is

$$a_p = \left(\prod_{k<p} Z_k \right) \frac{X_p + i Y_p}{2}. \tag{3}$$

The parity string enforces fermionic anticommutation. A two-electron Hamiltonian contains up to $O(N_{\mathrm{orb}}^4)$ terms before combination or factorization. This count is not the cost of exact diagonalization, and it does not mean that every Hamiltonian term must appear as a separate ansatz generator.

2.2 Energy references and the variational bound

Let E_0 be the global minimum eigenvalue of the supplied Hermitian operator. Every normalized state and every physical density matrix satisfy

$$\langle\psi|H|\psi\rangle \geq E_0, \qquad \mathrm{Tr}(\rho H) \geq E_0. \tag{4}$$

Noise does not remove this full-space bound. A comparison to the minimum in a particular particle-number or spin sector additionally requires the state to remain in that sector. A finite-shot estimate or a mitigated estimator can cross a variational bound because it is not necessarily the exact expectation of a physical density matrix.

Energy agreement also has limited implications for the state. For a nondegenerate ground state separated by a known gap $\Delta > 0$, the excited-state population is bounded above by $(E - E_0)/\Delta$. Without an appropriate gap and reference, an energy residual does not certify wavefunction fidelity or the accuracy of other observables.

We write the ideal circuit deficit as $\delta = E_{\mathrm{ideal}} - E_0$ and the ideal-to-trace contrast as $D = c_0 - E_{\mathrm{ideal}}$. The separate trace-to-ground contrast is $C = c_0 - E_0 = D + \delta$. Neither D nor C is generally the spectral width $E_{\max} - E_{\min}$. The supplied six-qubit operator is the computational reference for Section 6; its original molecular geometry and sector assignment are not independently reconstructed by this replay.

2.3 Compiled resources and noise coordinates

G denotes a two-qubit gate count for a specified compiled circuit. A reproducible count requires the gate basis, compiler settings, layout and coupling map. Counts from a compilation-model ladder and counts from an executed instruction-set circuit describe different objects. The numerical examples identify which object is used.

We define the phenomenological coordinate $\chi = G\varepsilon$. For inhomogeneous errors, an alternative is $\chi_{\mathrm{w}} = \sum_e G_e \varepsilon_e$, with G_e the number of uses of edge e. Calibration dates and the aggregation rule must accompany either value. Current calibration is appropriate for a new forecast; replay of an earlier calculation uses its archived inputs.

A reported error rate is not automatically a depolarizing probability. For a d-dimensional depolarizing channel, $p_{\mathrm{dep}} = d\, r_{\mathrm{avg}}/(d-1)$ when r_{avg} is average gate infidelity, whereas $p_{\mathrm{dep}} = d^2 r_{\mathrm{pro}}/(d^2-1)$ for process infidelity. For two qubits these factors are $4/3$ and $16/15$. A fitted coefficient may already absorb such a conversion. We therefore specify the convention used by each calculation and do not apply an additional conversion to the archived replay.

---

3. Measurement, optimization and noise budgets

3.1 Sampling an energy

For independently measured Pauli terms, let $v_j = 1 - \langle P_j\rangle^2$ and omit the identity term from sampling. Under variance-optimal allocation, the shot budget for a target standard error s_E is

$$M = \frac{\left(\sum_{j\neq 0} |c_j|\sqrt{v_j}\right)^2}{s_E^2} \leq \frac{\left(\sum_{j\neq 0} |c_j|\right)^2}{s_E^2}. \tag{5}$$

The first expression depends on the state; the upper bound uses only coefficients. They should be distinguished when reporting measurement cost. Allocating shots during optimization may also require estimating the variances, so the optimal expression is not automatically an executable schedule.

Commuting measurements can reduce the required sampling [6]. For a jointly measured group, the standard deviation of its weighted sum is at most the sum of the individual weighted standard deviations. Under ideal variance-optimal allocation, grouping therefore cannot increase the minimum shot count. The comparison does not include the extra basis-change gates or their noise, and it does not establish a wall-clock saving on hardware.

3.2 Optimization diagnostics

Small gradients, local stationary points and limited ansatz expressibility are different explanations for poor optimization. Barren-plateau results concern gradient concentration under specified assumptions on the circuit and cost [7]. They are properties of the landscape and can remain relevant with exact statevector evaluation; removing shot noise does not prove their absence. Finite precision and optimizer stopping tolerances still matter.

Multiple starts help reveal dependence on initialization, but finitely many optimization runs cannot certify the global minimum over an ansatz manifold. A spread of final energies is consistent with distinct basins or incomplete convergence [8]. Distinguishing these possibilities requires gradient and curvature diagnostics or a more complete search.

Adaptive constructions select operators using state-dependent energy derivatives [9,10]. For an anti-Hermitian generator A_k, the derivative at zero parameter is

$$\left.\frac{\partial E}{\partial \theta_k}\right|_{0} = \langle \psi | [H, A_k] | \psi \rangle. \tag{6}$$

This criterion reduces unnecessary circuit growth, but selecting an operator requires evaluating a pool of gradients. Parameter count, compiled gate count and selection-measurement cost therefore need separate accounting. A zero first derivative also does not prove that an operator can never lower the energy at a finite parameter value.

3.3 A conditional noise response

Under global depolarizing interpolation, a state becomes $\rho_q = q\rho_{\text{ideal}} + (1-q)I/2^n$. Its energy satisfies

$$E_q - E_0 = \delta + (1-q)D. \tag{7}$$

An exponential response $q = \exp(-a\chi)$ gives a useful phenomenological curve. Here q is a mixing weight, not generally the state fidelity. Local stochastic errors, coherent errors, leakage, relaxation and mitigation need not follow this curve. The single energy expectation in Eq. (7) also does not determine the full density matrix.

When $D > 0$, an additive accounting model can allocate a nonnegative allowance b to remaining bias and an allowance u to sampling uncertainty:

$$\delta + \left(1 - e^{-akG\varepsilon}\right)D + b + u \le \tau. \tag{8}$$

The coefficient k specifies the conversion from the stated error coordinate, and τ is the target tolerance. This is a design assumption with explicit allowances. The physical noise channels and the coverage of u must be checked before Eq. (8) is used as a probabilistic guarantee.

---

4. Classical circuit benchmarks

4.1 Restart sensitivity

An archived exact-statevector benchmark optimized an eight-qubit H2O active-space circuit with R_y layers and a linear CNOT entangler. The restart sets contain 32 initializations at depth 2 and 16 at depth 4, drawn from $U(-\pi, \pi)$. Table 1 reports the stored summary statistics; errors are relative to the benchmark reference.

Table 1. Restart summaries for the eight-qubit H_2O benchmark. Energy errors and standard deviations are in mHa.

Depth	Starts	Minimum	Median	Maximum	Std. dev.	Below 1.6
2	32	5.298	7.482	398.681	68.090	0
4	16	4.991	7.464	8.096	0.645	0

The depth-4 distribution is narrower, and its best observed error is 4.991 mHa. None of the 48 starts reaches the 1.6 mHa threshold. These observations describe the tested optimization procedure. They do not provide an expressibility bound for the entire family, nor imply that further optimization cannot improve the result.

The stored mean particle number at depth 2 is 3.9688 rather than exactly four. The maximum within-state particle-number variance is approximately 1.24×10^{-3}. A small variance alone does not locate a state in the intended sector: a state concentrated in a different sector also has small variance. Per-run sector probabilities are needed before every restart can be treated as a chemically comparable state. The aggregate record does not supply all those probabilities.

4.2 Routed ansatz resources

Table 2 uses one compiled-resource record consistently. It compares UCCSD and ADAPT circuits with their stored ideal errors. The reduced two-qubit examples have equal gate counts, while the larger examples have different counts. Equality of counts does not establish equality of circuits or noise response.

Table 2. Routed gate counts and stored ideal errors (mHa) from one resource record.

Target	n	HF error	UCCSD G	Ideal error	ADAPT G	Ideal error
H_2, compact	2	20.31	2	0.0000	2	0.0000
H_2, stretched	2	87.28	2	0.0000	2	0.0000
LiH CAS(2,2)	2	0.26	2	0.0000	2	0.0143
LiH CAS(2,3)	6	1.05	124	0.0000	69	0.2281
LiH CAS(4,4)	8	1.20	800	0.0006	72	0.3698
BeH_2 CAS(4,4)	8	5.89	349	0.0007	176	0.0687
H_2O CAS(4,4)	8	7.42	361	0.0005	176	1.4580

For the four larger targets, the UCCSD-to-ADAPT count ratios are 1.80, 11.11, 1.98 and 2.05 in table order. The values apply to these compiled circuits. They do not imply a universal hardware threshold at six qubits. LiH's Hartree-Fock error in these active spaces is already below 1.6 mHa, which also changes the significance of meeting that threshold.

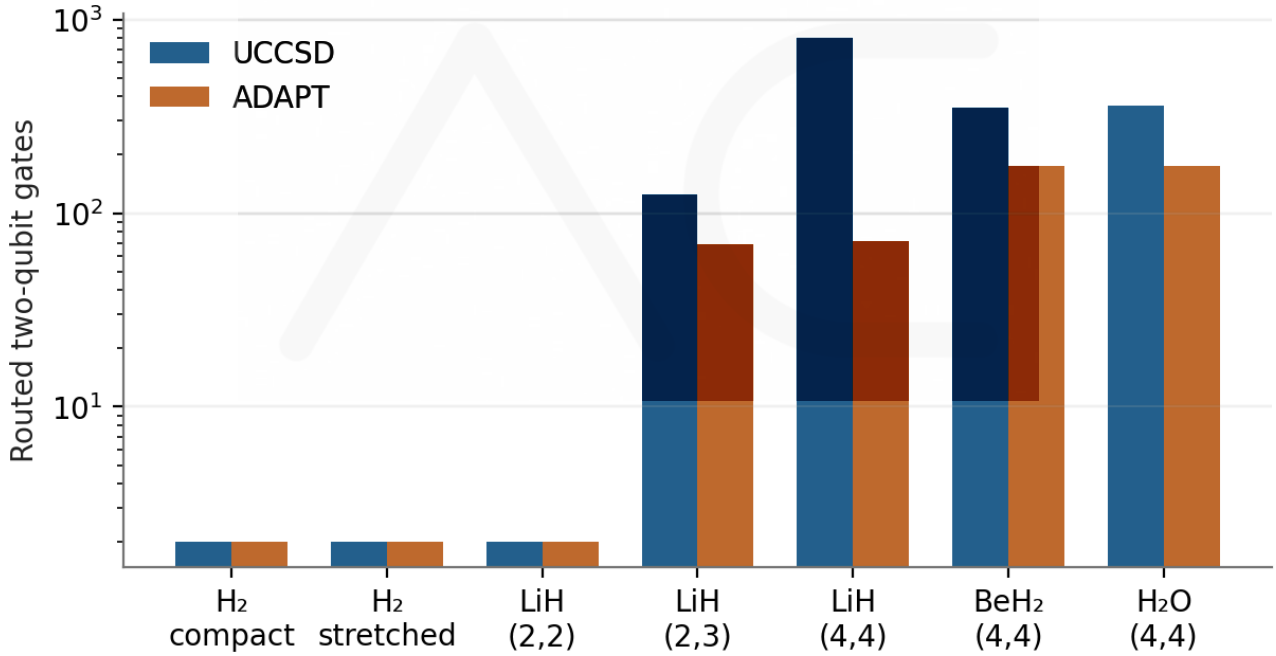

Figure 1. Two-qubit counts from the resource record used in Table 2. The comparison is between compiled UCCSD and ADAPT circuits. No experimental feasibility boundary is inferred from this plot. Plotting code was prepared with assistance from OpenAI Codex (GPT-6), using the numerical records described in the text.

Some hardware-efficient entries in the same source record have large ideal errors and initialization diagnostics that do not establish converged, sector-correct states. They are excluded from the resource-performance comparison. Checking particle number, the initial gradient and the final energy is necessary before interpreting an optimizer's success flag as evidence of a useful variational solution.

5. Empirical energy models

5.1 Fitted response and additive residual

The empirical model examined in the numerical replay is

$$\widehat{\Delta E} = \Phi + \left(1 - e^{-\alpha \chi}\right) C, \tag{9}$$

where $C = c_0 - E_0$ and Φ is an additive floor. Equation (9) is the archived prediction convention. It approximates the ideal deficit as zero, whereas Eq. (7) retains it explicitly. The two forms should not be substituted for each other when reproducing the numerical comparison.

A stored simulator fit reports $a = 5.05$ with a bootstrap interval [4.821, 5.614] for six two-qubit points, and $a = 0.558$ with interval [0.529, 0.593] for three six-qubit points. The corresponding floors are 19.9 and 7.4 mHa. These are summaries of those fits; the intervals do not establish transferability to another molecule, mapping or processor. The prediction record used in Section 6 specifies $a = 0.58$ and $\Phi = 7.41$ mHa.

A residual obtained by subtracting a gate model from an observed energy deviation is an *implied* floor. It can include gate-model error, readout effects and other bias. Its presence alone does not isolate readout as the cause. An additive floor is likewise not a general theorem about strongly noisy circuits: measurement assignment errors can shift an estimated energy on either side of the trace average.

5.2 Calibration and implementation sensitivity

Gate errors enter Eq. (9) through both their values and their aggregation rule. Section 6 compares a device-level archived error coordinate with a use-weighted edge coordinate. The change tests sensitivity; it does not recalibrate the exponent or provide an alternative prospectively validated model.

Independent numerical implementations help distinguish software errors from modeling assumptions. Agreement for the same specified channel checks the computation; it does not validate that channel as a description of a processor. A physical readout model must act in the measurement basis, which can differ across Pauli terms.

A hardware test intended to distinguish gate count from qubit count should vary the compiled gate exposure at fixed width, observable and geometry, and should record duration. In the small fragment cohort, width and gate count change together. Such a cohort cannot isolate their individual effects. Noise-amplification experiments additionally require checking the compiled fold factors; a requested fold factor is not evidence that the submitted circuit attained it.

6. Numerical replay and processor results

6.1 Six-qubit circuit and observable

The supplied reduced circuit contains 1,433 CZ, 2,258 SX, 1,548 RZ and 90 X gates. Its dependency depth is 3,823. Direct NumPy evolution and a separate Qiskit statevector calculation agree within 2.0×10^{-15} in the maximum complex-amplitude difference, including the supplied global phase. This verifies the local reduced representation. It does not authenticate the original full-width submission or its provider linkage.

The supplied observable has 325 Pauli terms. Its global minimum is -2.525385852047 Ha; the circuit expectation is -2.521431970975 Ha. Thus the bound circuit's ideal deficit is 3.953881 mHa. This quantity concerns one circuit at its supplied parameters, not the best reachable state of the ansatz.

A regenerated operator aligns with the supplied one after qubit relabeling but differs in coefficients. Their difference has spectral norm 0.267543 mHa, which bounds expectation differences for the same state in a common basis. All headline values in this section use the supplied operator, so the reference remains consistent.

6.2 Energy comparison

The local result record gives -1.527165893224 Ha. Against the supplied operator this is a deviation of 998.219959 mHa. The trace average is -1.510520464724 Ha, giving $C = 1014.865387$ mHa and a distance of 16.645428 mHa from the result to the trace average. The energy lies close to the trace baseline; this one observable does not establish that the state is maximally mixed or contains no molecular information.

The archived error coordinate is $\varepsilon = 0.002874954624$, giving $\chi = 4.119809976$. Table 3 evaluates Eq. (9) at this coordinate and compares its variants with the trace baseline. A signed residual is the observed deviation minus the predicted deviation.

Table 3. Six-qubit energy comparison at the archived error coordinate.

Model or baseline	Prediction (mHa)	Residual (mHa)
Fitted a = 0.58, with floor	929.236135	+68.983824
Fitted a = 0.58, no floor	921.826135	+76.393824
Retrospective a = 1, with floor	1005.786259	-7.566300
Retrospective a = 1, no floor	998.376259	-0.156300
Trace baseline	1014.865387	-16.645428

The fitted model with floor has an absolute residual 4.14 times that of the trace baseline. The unit-exponent variants are retrospective comparisons. The no-floor variant is numerically close at the archived error coordinate, but its residual becomes +44.639 mHa when the use-weighted edge coordinate $\chi_{\mathrm{w}} = 2.806986495$ is substituted. This sensitivity is part of the model assessment.

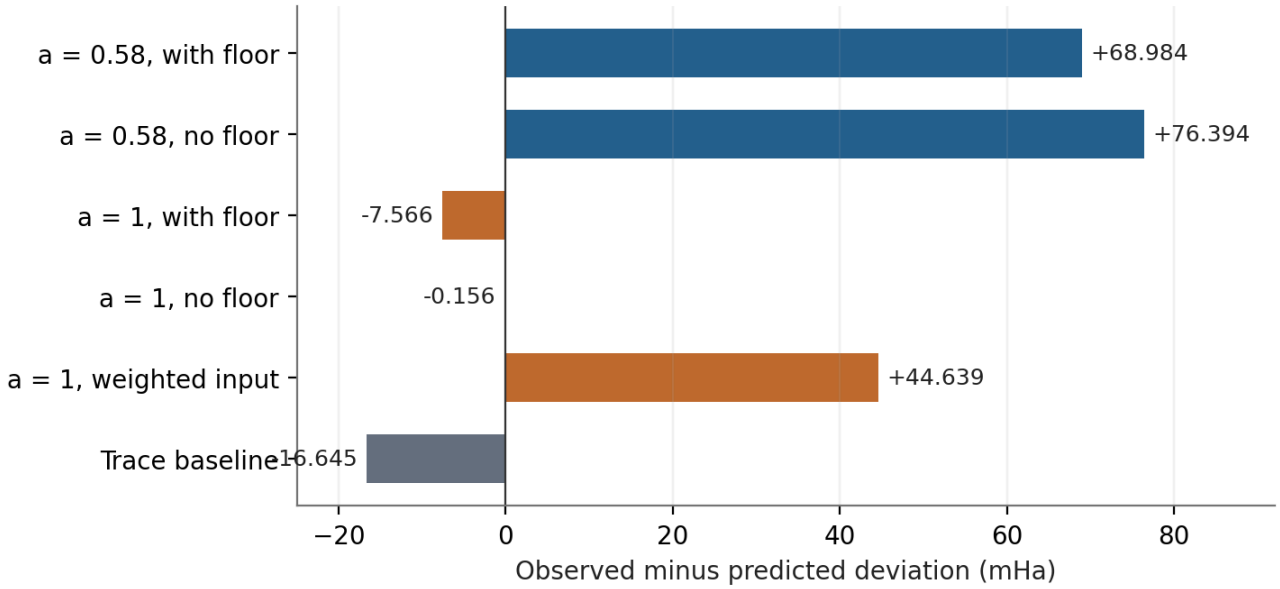

Figure 2. Signed energy residuals for the six-qubit comparison. The unit-exponent variants are retrospective; the weighted variant changes the error aggregation without refitting the exponent. Zero denotes equality of the two scalar energies, not validation of the underlying noise model. Plotting code was prepared with assistance from OpenAI Codex (GPT-6), using the numerical records described in the text.

The result record supplies uncertainty fields of 8.333732 and 8.480969 mHa. Dividing the unit-exponent-plus-floor residual by them gives -0.907913 and -0.892150. These are sampling-normalized distances. They omit uncertainty in the model, calibration and reference and are not calibrated significance tests.

The available seven-entry prediction file does not include this 1,433-gate target. Local records place the model file before the first located retrieval, but the job had already completed. Prior access to the outcome is not established. We therefore treat Table 3 as a numerical replay and retrospective comparison rather than a blind prospective validation. Supplementary Section S2 gives the chronology and exact inputs.

6.3 Hydrogen mitigation campaign

The hydrogen campaign contains seven geometries in each of three arms: raw estimation, twirled readout error extinction (TREX), and TREX with zero-noise extrapolation (ZNE). Recorded electronic energies are combined with their geometry-specific constant shifts before comparison with the stored ideal reference. Table 4 summarizes all 21 point estimates.

Table 4. Hydrogen campaign point estimates. Deviations are in mHa.

Arm	Points	Mean absolute deviation	Range	Below 1.6 mHa
Raw	7	9.733930	5.744439 to 14.562669	0
TREX	7	4.873019	2.298373 to 9.638705	0
TREX + ZNE	7	3.153040	1.465640 to 5.314871	1

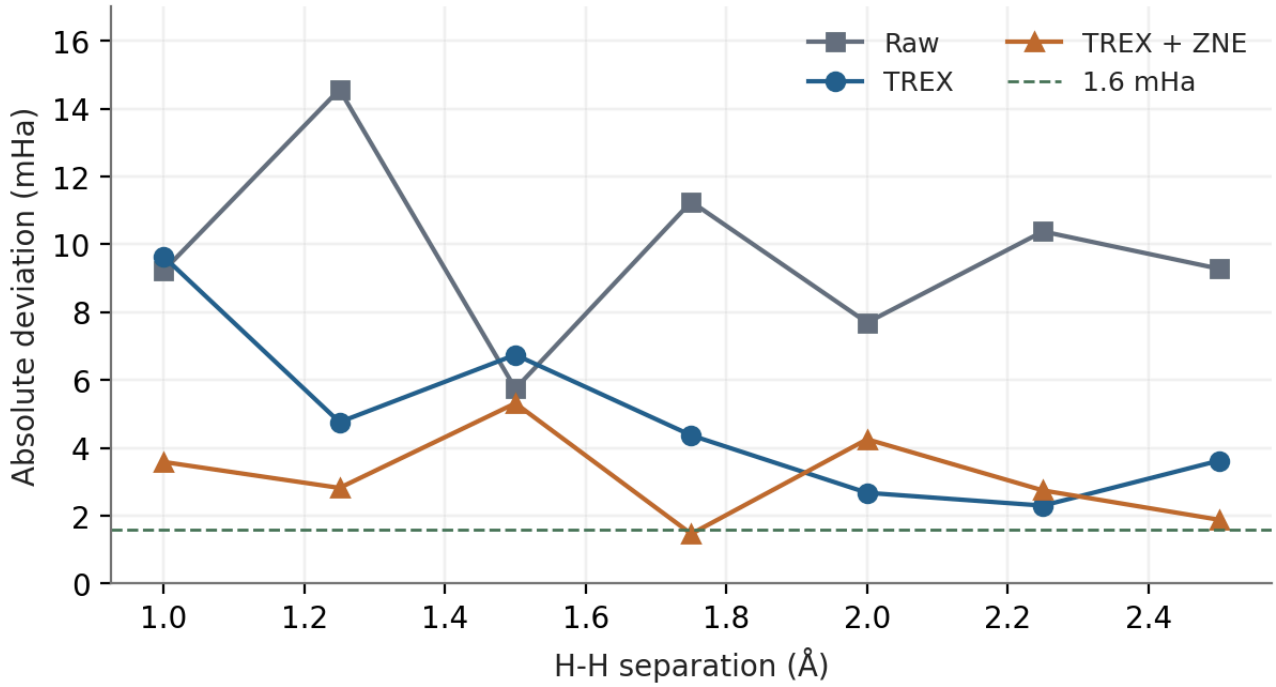

Figure 3. Absolute deviations for all seven geometries in each campaign arm. The dashed line marks 1.6 mHa. The curves join point estimates for readability and do not represent repeat-qualified confidence bands. Plotting code was prepared with assistance from OpenAI Codex (GPT-6), using the numerical records described in the text.

The mean deviation decreases across the three arms. At $R = 1.75$ Å, the combined-mitigation point is 1.465640 mHa, below the target. The campaign manifest predicts 1.319 mHa for this point; a generic prediction file using different inputs predicts 1.768 mHa. The relevant input version must therefore accompany any assessment of forecasting performance.

Across the full cohort, the manifest's threshold classifications are correct for 16 of 21 points. Always predicting a deviation above 1.6 mHa is correct for 20 of 21. Of six predicted below-threshold outcomes, one occurs. The point result is useful as a measurement, but the cohort does not establish superior classification or a reduction in required quantum resources.

The reduction in mean deviation is descriptive. Readout calibration, circuit multiplication, shots and classical processing contribute to mitigation cost [11]. Equal science-circuit shot counts would not by themselves establish zero overhead or improved time to solution.

6.4 Repeat measurements

The stage-3 plan specifies six executions at each of two geometries and evaluates a two-sided 95% Student-t interval for mean run-level absolute deviation. The rule requires the upper endpoint to be below 1.6 mHa to establish the target; an interval straddling the target is inconclusive. Three values per geometry are available in the supplied partial record.

Table 5. Partial repeat analysis under the recorded threshold rule.

R (Å)	Available / planned	Mean (mHa)	95% t interval (mHa)	Outcome
1.75	3 / 6	0.844148	[0.018292, 1.670004]	Inconclusive
2.00	3 / 6	1.732675	[-2.419499, 5.884850]	Inconclusive

Both intervals straddle 1.6 mHa. The first mean is below the target, but the available interval does not yet meet the rule. The second interval is also inconclusive under the stated rule. Its negative lower endpoint is an artifact of a symmetric interval approximation for nonnegative observations, not a negative physical absolute error.

These are partial-data calculations with two degrees of freedom. They are not completed six-run endpoints or an authorized sequential stopping analysis. The source supplies deviation magnitudes, not all signed energies and original provider payloads. The interpretation of the intervals also depends on independence and sampling assumptions. Supplementary Section S4 provides the values and the distinction between point accuracy, mean absolute deviation and signed bias.

6.5 Computational preparation and verification

OpenAI Codex (GPT-6) assisted with preparation of numerical replay and analysis scripts, arithmetic checks, statistical summaries and plotting code from the supplied local records. The preparation was directed toward checking the manuscript against those records and separating retrospective comparisons from prospective predictions. Computational checks included a NumPy statevector replay, a separate Qiskit calculation for the same circuit, and regeneration of the reported table arithmetic. Agreement between implementations checks the specified computation; it does not independently authenticate the source records or validate the noise model. The reproducibility procedure and its evidence limits are given in Supplementary Section S6.

---

7. A resource-screening procedure

A useful screening calculation begins by fixing the scientific target. Specify the geometry, orbital basis, active space, symmetry sector and reference. Compare the Hartree-Fock error with the desired tolerance: when Hartree-Fock already meets the target, a successful VQE estimate is not evidence of an improved chemical answer, although it may still be useful for device characterization.

Next, evaluate the ideal compiled circuit and quantify its deficit. Count its two-qubit operations after routing and specify the associated calibration data. The ideal deficit, a gate-bias estimate and the uncertainty allowance then enter Eq. (8). If the allocated terms exceed the tolerance, revise the proposed circuit or experimental design before assigning the result a chemistry-accuracy interpretation.

The trace baseline should be computed alongside the model whenever the operator is available. A model that only predicts an energy close to this baseline must demonstrate useful discrimination against it on held-out targets. A single arbitrary threshold such as $\chi \geq 1$ is not a universal stopping rule: at $a = 1$, $\chi = 1$ leaves an interpolation weight of $e^{-1} \approx 0.37$.

For the supplied six-qubit circuit, the ideal deficit is already 3.954 mHa. Under Eq. (7) with $D > 0$, further depolarizing mixing increases the deficit, so this particular state cannot satisfy a 1.6 mHa budget within that model. This does not rule out a different optimized state, a different circuit, or a different estimator. It identifies the first term that a proposed calculation needs to address.

Classical preprocessing is inexpensive for the small supplied operators used here. Dense diagonalization nevertheless scales exponentially with the number of encoded qubits. A larger-system forecast would need an independently justified reference or bound, together with its uncertainty; constructing an $O(N_{\mathrm{orb}}^4)$ term list does not make exact diagonalization polynomial.

8. Conditional processor budgets and co-design

8.1 Inverting an energy allowance

For Eq. (8), define the remaining gate-bias allowance $B = \tau - \delta - b - u$. If $D > 0$ and $0 < B < D$, the assumed response yields

$$\varepsilon \leq \varepsilon_\star = \frac{-\ln(1 - B/D)}{akG}. \tag{10}$$

If $B < 0$, the other allocated terms already exceed the tolerance. If $B = 0$, the assumed positive gate-bias contribution must vanish. If $B \geq D$, this particular depolarizing-energy term imposes no finite upper bound. Other physical errors can still matter.

Table 6 uses trace-to-ground contrast inputs C and defines $D = C - \delta$ before applying Eq. (10). The LiH, BeH2 and H2O reference summaries identify their stored contrasts as trace-to-ground differences; the H2 value is retained as an illustrative input. We set $a = k = 1$ and $b = u = 0$. These are conditional model-coordinate limits, not measured processor specifications. Their use in chip design requires compatible molecular references, a calibrated response and explicit allowances for measurement error.

Table 6. Conditional inversion with D = C - δ. Energies are in mHa; a = k = 1 and b = u = 0.

Target	G	δ	C	D	ε*	0.00083 / ε*
H_2 stretched (2q)	2	0.000	341.000	341.000	2.351562e-03	0.353
LiH CAS(2,3) (6q)	69	0.230	599.000	598.770	3.319776e-05	25.002
BeH_2 CAS(4,4) (8q)	176	0.069	1073.494	1073.425	8.109625e-06	102.348
H_2O CAS(4,4) (8q)	176	1.458	1858.000	1856.542	4.345978e-07	1909.812

For BeH2, $G = 176$, $C = 1073.493899$ mHa and $\delta = 0.069$ mHa give $D = 1073.424899$ mHa. The gate-bias allowance is 1.531 mHa, yielding $\varepsilon_\star = 8.109625 \times 10^{-6}$ and a coordinate ratio of 102.348. These calculations use the source's rounded ideal deficits as specified inputs; their displayed precision does not imply a measured hardware tolerance.

The H2O allowance is only 0.142 mHa because its ideal deficit is 1.458 mHa. This makes its conditional ratio much larger. A single improvement factor therefore cannot be transported across molecules. Changing a from 1 to 0.58 increases a limit by $1/0.58$ only if all other assumptions are held fixed.

8.2 Accounting for combined improvements

Device selection, layout, gate basis and ansatz selection can each change a circuit's error exposure. Their gains must be evaluated on a common baseline. A requirement computed with ADAPT's gate count already includes that ansatz choice; a requirement using a selected error coordinate already includes that selection. Multiplying either gain a second time overstates the improvement.

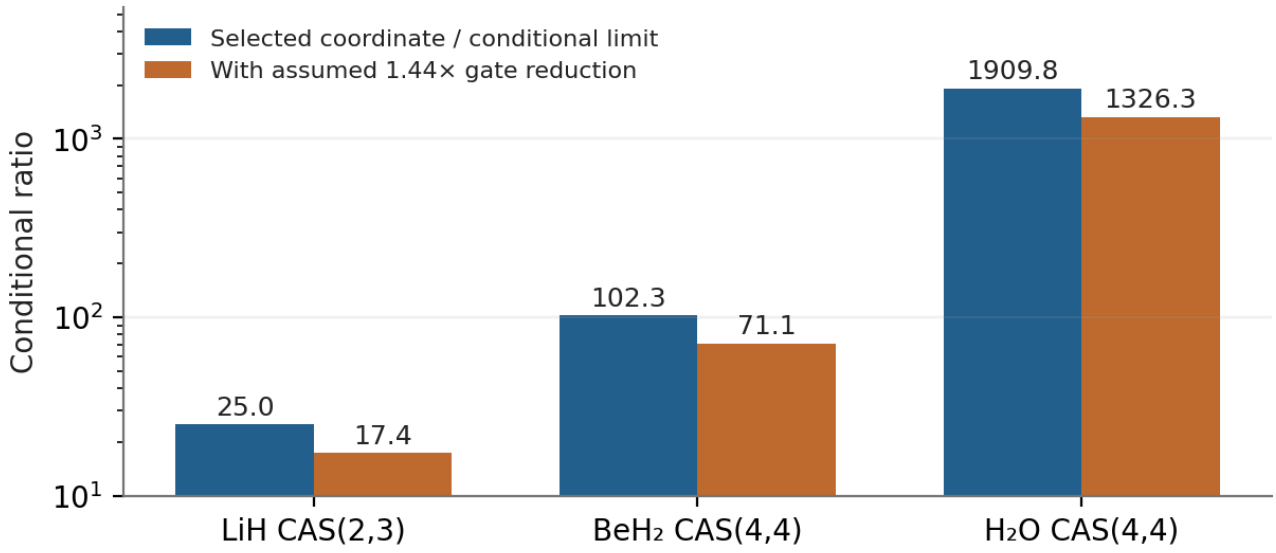

Figure 4. Ratios of the selected coordinate 0.00083 to the conditional limits in Table 6, and the corresponding ratios after an assumed 1.44-fold gate-count reduction. Both series are model calculations. They do not represent independent hardware demonstrations or a common measured shortfall. Plotting code was prepared with assistance from OpenAI Codex (GPT-6), using the numerical records described in the text.

An assumed 1.44-fold reduction in routed gate count changes the ratios to 17.362 for LiH, 71.075 for BeH_2 and 1326.258 for H_2O. These values share inputs and assumptions and cannot be counted as independent validations. A favorable edge percentile also need not form a connected usable layout, and the best gate, ansatz and routing gains may not coexist.

A practical co-design comparison should compile each complete candidate implementation and score its total error budget and cost. Parallel execution can improve throughput, but does not automatically improve the accuracy of an individual energy estimate. Neither the conditional table nor the calibration snapshot establishes a universal limit on a hardware family.

9. Optimization and mitigation cost

Gate count measures one part of a VQE calculation. An end-to-end cost record should also include optimizer evaluations, adaptive-pool measurements, shots per observable group, calibration jobs, mitigation circuits and classical processing. A short circuit that needs many evaluations can cost more than a longer circuit that converges reliably.

The supplied restart summaries report 793.1 s for 32 depth-2 runs and 859.4 s for 16 depth-4 runs. These are archived execution times for those workloads, not a controlled comparison of per-answer hardware cost. Machine configuration, software versions and workload accounting are needed before extrapolating them.

For mitigation, the campaign establishes changes in point-estimate deviations. It does not yet establish an optimum in accuracy per unit cost. A prospective comparison would use common scientific endpoints, complete repeat sets and a cost record for each arm. Fold integrity and measurement-calibration overhead should be checked before attributing changes to ZNE or readout mitigation alone.

10. Relation to prior work and experimental design

Adaptive VQE methods show how state-dependent operator selection can reduce circuit resources [9,10]. The compiled examples here quantify that trade for a small set of active spaces. They also show why a gain in parameter count should not be used as a substitute for a routed-gate or measurement-cost comparison.

Noise-tolerance calculations for chemistry provide a precedent for conditional inversions [3]. Their numerical thresholds depend on the tested algorithms and noise models. Circuit-error-sum extrapolation provides a related way to use spatially inhomogeneous calibrations [4]. The present replay makes the input convention and trace-baseline comparison explicit; it does not establish that one scalar exposure variable captures every physical error channel.

The next discriminating experiment should vary exposure at fixed width and geometry, record circuit duration, and use a target-specific prediction fixed before execution. Intermediate exposures can be more informative than a strongly saturated energy. The exponent, floor, error aggregation, estimator settings, accuracy endpoint and acceptance rule should all be specified in advance.

A separate validation cohort is needed to estimate predictive performance. It should include the full target set, simple baselines and uncertainty from calibration and model fitting. The campaign's 16/21 classification score against a 20/21 constant baseline illustrates why selected favorable points are insufficient for this comparison.

11. Scope and limitations

The numerical replay verifies supplied local files and their arithmetic. It does not independently authenticate provider execution, reconstruct all original molecular inputs or establish outcome blindness. The original full-width input and its linkage to the reduced export remain separate provenance questions.

The circuit benchmarks cover small active spaces and selected optimization procedures. Finite restarts do not bound the full ansatz family. Exact-statevector simulation omits sampling and physical noise, and the aggregate symmetry statistics do not replace per-state sector checks.

The noise response is phenomenological. Duration, coherent errors, leakage, calibration drift, estimator covariance and several measurement effects are not captured by a single fitted exponent. Simulator fits and retrospective scalar agreement do not establish transferability to new circuits.

The repeat analysis uses partial records. Its intervals depend on unverified independence and distributional assumptions, while the conditional processor table leaves no explicit allowance for readout or sampling error. These restrictions define the use of the current results and the measurements needed to extend them.

12. Conclusion

This study connects compiled circuit resources with explicit energy references and reproducible model comparisons. The six-qubit replay establishes the supplied circuit's 1,433 CZ gates, 3,823 dependency depth and 3.954 mHa ideal deficit. Its processor-result record lies 16.645 mHa from the trace average. The fitted-exponent model gives a larger residual than that baseline; a retrospective unit-exponent comparison fits closely at one calibration coordinate and changes materially when the coordinate is reweighted.

The hydrogen campaign shows a reduction in mean absolute deviation from 9.734 to 3.153 mHa across the raw and combined-mitigation arms. One point falls below 1.6 mHa. The available repeat intervals leave the planned accuracy criterion unresolved. Reporting the whole cohort and its baseline makes the achieved point result and the remaining validation work explicit.

The conditional inversion provides a way to allocate a molecular accuracy target across ideal-state error, noise bias and uncertainty. Its value is in making assumptions and tradeoffs calculable. Establishing predictive processor requirements now calls for complete records, target-specific prospective tests and validation against simple baselines across multiple configurations.

Funding

This research was financially supported by Atomic Cubit.

Competing interests

Pally Sai Tilak is the founder of Atomic Cubit, which is developing a four-qubit quantum processor and intends to apply this research to processor design.

Acknowledgments

OpenAI Codex (GPT-6) assisted with substantive revision of the scientific explanations, interpretation and organization of the manuscript, under the author’s direction to prepare the supplied research for publication. Its computational role and the checks performed during preparation are described in Section 6.5.

Data availability

The Supplemental Material [12] includes a research-data archive with the reduced circuit, supplied Pauli operator, selected result records, calibration inputs, campaign data, benchmark summaries, conditional-budget inputs and offline reproduction code. A provenance file identifies copied records and field-level extracts; a SHA-256 manifest identifies the released files. These materials reproduce the circuit replay and reported numerical transformations. Complete original hardware input objects and primary outputs for every summarized benchmark are not included, and offline reproduction does not independently authenticate provider execution or historical timing.

References

[1] A. Peruzzo et al., “A variational eigenvalue solver on a photonic quantum processor,” *Nature Communications* **5**, 4213 (2014). doi:10.1038/ncomms5213.

[2] T. Proctor, K. Rudinger, K. Young, E. Nielsen and R. Blume-Kohout, “Measuring the capabilities of quantum computers,” *Nature Physics* **18**, 75-79 (2022). doi:10.1038/s41567-021-01409-7.

[3] K. Dalton et al., “Quantifying the effect of gate errors on variational quantum eigensolvers for quantum chemistry,” *npj Quantum Information* **10**, 18 (2024). doi:10.1038/s41534-024-00808-x.

[4] A. Uvarov et al., “Mitigating quantum gate errors for variational eigensolvers using hardware-inspired zero-noise extrapolation,” *Physical Review A* **110**, 012404 (2024). doi:10.1103/PhysRevA.110.012404.

[5] S. Bravyi, J. M. Gambetta, A. Mezzacapo and K. Temme, “Tapering off qubits to simulate fermionic Hamiltonians,” arXiv:1701.08213 (2017). arXiv.

[6] W. J. Huggins et al., “Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers,” *npj Quantum Information* **7**, 23 (2021). doi:10.1038/s41534-020-00341-7.

[7] 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.

[8] E. R. Anschuetz and B. T. Kiani, “Quantum variational algorithms are swamped with traps,” *Nature Communications* **13**, 7760 (2022). doi:10.1038/s41467-022-35364-5.

[9] H. R. Grimsley, S. E. Economou, E. Barnes and N. J. Mayhall, “An adaptive variational algorithm for exact molecular simulations on a quantum computer,” *Nature Communications* **10**, 3007 (2019). doi:10.1038/s41467-019-10988-2.

[10] H. L. Tang et al., “qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor,” *PRX Quantum* **2**, 020310 (2021). doi:10.1103/PRXQuantum.2.020310.

[11] K. Temme, S. Bravyi and J. M. Gambetta, “Error mitigation for short-depth quantum circuits,” *Physical Review Letters* **119**, 180509 (2017). doi:10.1103/PhysRevLett.119.180509.

[12] See Supplemental Material at [URL will be inserted by publisher] for circuit and Hamiltonian conventions, complete numerical tables, calibration inputs, record provenance, repeat-estimate calculations and reproducibility procedures.