Supplementary Information

The Chemistry Performance Envelope

Pally Sai Tilak · Atomic Cubit · tillaksai@gmail.com

S1. Data organization and scope

This supplement specifies the records and conventions used for the main article. The numerical package groups plotting code, numerical inputs and selected source records. Its manifest reports full SHA-256 hashes for current file identity. A matching hash does not establish when a file first existed, whether an outcome was previously inspected, or whether a provider authenticated the file.

The source material comprises reduced circuit and observable exports, local processor-result records, campaign metadata, prediction parameters, calibration snapshots and benchmark summaries. The numerical checks use these supplied files without accessing a provider account. The main article's “processor result” is therefore an energy from a local result record. Original full-width execution inputs and independent provider authentication remain outside this offline calculation.

The general workflow is: identify one operator and energy-shift convention; build its matrix; evaluate the supplied circuit; compute the reference and trace average; transform each recorded energy using that same convention; and evaluate each explicitly named model variant. The table and figure data are generated from these quantities rather than entered independently.

Released file	Role
`circuit.qasm3`, `observable.json`	Reduced circuit and exact supplied Pauli terms
`result_record.json`, `edge_calibrations.json`	Local result metadata and used-edge calibration values
`prediction_inputs.json`, `fit_summary.json`	Archived prediction inputs and stored fit summaries
`campaign_meta.json`, `campaign_energies.json`, `campaign_scores.json`	Full 21-point campaign, references and prediction comparisons
`stage2_repeats.json`, `replay_reference.json`	Repeat magnitudes and numerical reference values
`resource_summary.json`, `restart_summary.json`	Compiled-resource and restart summaries
`conditional_budgets.json`, `contrast_reference.json`	Explicit energy conventions and model-budget inputs
`comparison_operator.json`	Later reconstructed comparison operator
`PROVENANCE.json`, `MANIFEST.json`	Selection descriptions, source hashes and released-file hashes

The archive contains a documented selection of records and fields, rather than the full internal project. All campaign outcomes are retained. It excludes unrelated candidate metadata and internal editorial records. Source hashes identify the originals from which extracts were made; they do not establish independent provider authentication. The circuit and Pauli operator support an exact replay of the supplied mathematical instance, not reconstruction of its original molecular preparation.

S2. Six-qubit representation and energy chain

S2.1 Circuit and observable

The supplied reduced OpenQASM circuit has six qubits and an explicit global phase of 2.356194490194321 radians. Its gate totals are 1,433 CZ, 2,258 SX, 1,548 RZ and 90 X. The reported depth of 3,823 is a dependency depth with unit gate levels, not an elapsed duration.

Pauli labels use the little-endian convention, with qubit zero represented at the right of a label. The observable has 325 distinct Pauli terms. Direct state evolution including global phase agrees with the separate Qiskit calculation to approximately 2.0×10^{-15} in maximum amplitude difference. This tests the supplied reduced representation. It does not reconstruct the original full-width submitted object.

The reduced circuit's state norm is 0.999999999999991. Its energy and the operator's global minimum are computed from the same supplied Pauli coefficients. A separate regenerated operator aligns under permutation (0, 1, 5, 4, 3, 2). The norm of the difference between the aligned operators is 0.267542736 mHa. That number is a deterministic bound on an expectation difference for the same state, not a stochastic uncertainty estimate.

Table S1. One consistent energy chain from the supplied observable.

Quantity	Value
Ideal circuit energy (Ha)	-2.521431970975
Supplied global minimum (Ha)	-2.525385852047
Ideal deficit (mHa)	3.953881071
Result-record energy (Ha)	-1.527165893224
Identity coefficient c_0 (Ha)	-1.510520464724
Trace-to-ground contrast C (mHa)	1014.865387322
Result deviation (mHa)	998.219958823
Result distance to trace (mHa)	16.645428499

The primary local result record identifies job `d97piu52su3c739idlvg`. It describes execution on 9 July 2026. A local model-file record is dated 13 August, and the first located result retrieval is later on 13 August. The seven targets listed in that prediction file have gate counts 2, 2, 2, 120, 2, 2 and 359. The 1,433-gate circuit is absent. This source chain does not establish a target-specific pre-registration or that the already-completed result was unseen when the model was chosen.

S2.2 Model branches and uncertainty fields

For the comparison, use $G = 1433$, $\varepsilon = 0.0028749546239971446$, $C = 1014.8653873223901$ mHa and the observed deviation 998.219958822959 mHa. The fitted-exponent branch uses $a = 0.58$ and floor 7.41 mHa. The unit-exponent branches set $a = 1$ retrospectively. The no-floor comparisons set the additive term to zero without changing any other archived input.

Table S2. Full branch comparison; energy quantities are in mHa.

Model or baseline	Prediction	Residual	Residual / s_1	Residual / s_2
Fitted a = 0.58, with floor	929.236135	+68.983824	+8.277663	+8.133956
Fitted a = 0.58, no floor	921.826135	+76.393824	+9.166820	+9.007676
Retrospective a = 1, with floor	1005.786259	-7.566300	-0.907913	-0.892150
Retrospective a = 1, no floor	998.376259	-0.156300	-0.018755	-0.018430
Retrospective a = 1, weighted input	953.581022	+44.638937	+5.356416	+5.263424
Trace baseline	1014.865387	-16.645428	-1.997356	-1.962680

Here $s_1 = 8.333732007749159$ mHa and $s_2 = 8.480968874790474$ mHa are the result record's `stds` and `ensemble_standard_error`. The calculation does not independently establish the server's covariance model, estimator coverage or the uncertainty of the prediction. Ratios to s_1 and s_2 should therefore not be interpreted as p-values or significance levels.

S2.3 Edge-weighted sensitivity

Under the supplied physical-qubit map, the CZ uses are 210 on edge (123,136), 351 on (136,143), 128 on (140,141), 358 on (141,142), and 386 on (142,143). They sum to 1,433. Weighting the local edge-calibration values by these uses gives $\chi_w = 2.8069864949638674$.

Holding $a = 1$ fixed and omitting the floor gives a predicted deviation 953.5810216436198 mHa and a residual 44.6389371793392 mHa. This replaces one aggregation rule with another while leaving the model coefficient unchanged. It is a sensitivity calculation; it is not a new fit or an independently validated forecast.

S3. Full hydrogen campaign

The source contains seven two-qubit hydrogen operators at geometries from 1.0 to 2.5 Å in 0.25 Å increments. Each operator is represented by the Pauli labels II, IZ, ZI, ZZ and XX. Electronic estimates receive the geometry-specific constant shift in the metadata before comparison with the stored ideal energy.

For arm A and geometry R, the scored deviation is

$$d_{A,R} = 1000 \left| E_{A,R}^{\text{electronic}} + s_R - E_R^{\text{ideal}} \right| \quad \text{mHa.} \tag{S1}$$

The ideal references and the supplied exact references differ by less than 6.5×10^{-5} mHa over these rows. The score below uses the ideal-reference convention of the campaign so that it reproduces the manifest evaluation.

Table S3. Complete campaign deviations and combined-mitigation predictions (mHa).

R (Å)	Raw	TREX	TREX + ZNE	Manifest ZNE prediction
1.00	9.223538	9.638705	3.585764	1.837
1.25	14.562669	4.759269	2.820491	1.562
1.50	5.744439	6.743811	5.314871	1.401
1.75	11.255255	4.370593	1.465640	1.319
2.00	7.692602	2.676681	4.254396	1.288
2.25	10.378175	2.298373	2.746838	1.284
2.50	9.280831	3.623698	1.883281	1.293

All 21 point deviations are included. Means over the seven geometries are 9.733930 mHa (raw), 4.873019 mHa (TREX) and 3.153040 mHa (TREX+ZNE). The last arm has a below-threshold point at $R = 1.75$ Å. These geometry averages are descriptive; the geometries are not interchangeable independent repeats of one target.

The manifest classification is obtained by comparing each manifest prediction and observation with 1.6 mHa. It has one true positive, five false positives, fifteen true negatives and zero false negatives. Thus its accuracy is 16/21, while the always-above-threshold rule obtains 20/21. Classification accuracy on this imbalanced cohort should be reported with these counts.

Local campaign records associate the raw arm with `d9vhqid0vrcc73bp8qdg`, TREX with `d9vhqiob1g9c73a8l7d0`, and TREX+ZNE with `d9vhqj50vrcc73bp8qf0`. Several geometries belong to the same job; 21 point estimates must not be described as 21 independently authenticated jobs.

At $R = 1.75$ Å, the manifest prediction is 1.319 mHa and the generic campaign prediction is 1.768 mHa. Local logs print the manifest prediction before the recorded result for this separate campaign. Such ordering is supporting local chronology; it does not independently authenticate execution or establish blindness for the six-qubit result.

S4. Repeated estimates and the scoring rule

S4.1 Stage 2 summaries

The available stage-2 record contains rounded repeat magnitudes. Table S4 recomputes a two-sided 95% Student-t interval from the three values at each geometry, using the sample standard deviation and two degrees of freedom. Small differences from stored interval endpoints are expected because the available input values are rounded.

Table S4. Three-repeat stage-2 summaries, recalculated from rounded inputs.

R (Å)	Repeat magnitudes (mHa)	Mean	95% t interval (mHa)
1.50	0.468, 6.167, 5.819	4.151333	[-3.784499, 12.087166]
1.75	0.64, 6.771, 0.199	2.536667	[-6.589220, 11.662553]
2.00	3.87, 1.933, 2.229	2.677333	[0.085312, 5.269355]
2.25	5.244, 1.054, 5.426	3.908000	[-2.236045, 10.052045]
2.50	2.833, 4.805, 2.986	3.541333	[0.816140, 6.266527]

Each interval straddles 1.6 mHa. A label indicating that an acceptance test was not met should not be confused with a demonstrated lower confidence bound above the target. Combining stages also changes the sample and requires a rule specified for that combination.

S4.2 Stage 3 partial data

The planned stage-3 endpoint uses six executions per geometry at 32,768 shots per manifest entry. Three magnitudes are available at each geometry:

Geometry (Å)	Run-level absolute deviations (mHa)
1.75	0.5451755429433103; 1.2021664139327815; 0.7851021253066914
2.00	3.380716700046138; 0.03871279681089135; 1.7785959615703728

For the available $n = 3$ values we evaluate

$$\bar{d} \pm t_{0.975,n-1}\frac{s_d}{\sqrt{n}}, \qquad s_d^2 = \frac{1}{n-1}\sum_i (d_i - \bar{d})^2. \tag{S2}$$

Table S4b. 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

The endpoint concerns the mean of absolute deviations. It does not measure signed bias or guarantee the accuracy of a future individual execution. Recovering signed energies and complete result payloads would allow additional reference and covariance checks. Local linkage inferred from job ordering should be labelled as an inference, not treated as an authenticated per-value payload.

A Student-t interval assumes an appropriate independent sampling model and is especially sensitive to that assumption at $n = 3$. These partial intervals do not authorize early stopping. Completing the planned observations and applying the declared endpoint are separate from recalculating an interval on the currently available subset.

S5. Conditional inversion and resource records

S5.1 Input conventions

The budget uses $D = C - \delta$, where $C = c_0 - E_0$ is the trace-to-ground contrast and $\delta = E_{\mathrm{ideal}} - E_0$. Thus

$$\varepsilon_\star = -\frac{\ln[1 - (1.6 - \delta)/(C - \delta)]}{G}, \qquad a = k = 1, \quad b = u = 0. \tag{S3}$$

All energies are in mHa. Table 6 of the main article gives the four selected examples; `conditional_budgets.json` provides full-precision inputs and derived values. `contrast_reference.json` preserves the LiH, BeH_2 and H_2O reference summary establishing the trace-to-ground convention. The source co and exact energies are printed at lower precision than its stored contrast, so their rounded difference is not used to replace the latter. The H_2 contrast is an illustrative stored value, not a separately regenerated operator.

The BeH_2 input is C = 1073.493899 mHa and δ = 0.069 mHa; D is therefore 1073.424899 mHa. The source's rounded ideal deficits are treated as specified inputs. The coordinate 0.00083 is a selected archived comparison value, not a current market-wide gate-error benchmark. Conversion between different gate-error metrics requires a separate calibration.

S5.2 Co-design arithmetic

The diagnostic ratio before routing change is $r = 0.00083/\varepsilon_\star$. If a common factor 1.44 reduces the gate count while leaving every other input unchanged, the ratio becomes $r/1.44$. The resulting values for LiH, BeH_2 and H_2O are 17.362284, 71.074669 and 1326.258233. These rows reuse a noise coordinate and routing assumption, and their budgets differ substantially. They are neither independent tests nor measurements of a common processor requirement.

A candidate layout must support the full circuit with its selected edges. The selected gate-error value, routing factor and ansatz savings cannot be assumed to combine merely because each appears favorable in a separate analysis. Device, gate and edge selection already represented in the chosen error coordinate must not be multiplied into the improvement a second time.

S5.3 Compiled-resource convention

The main article's Table 2 draws its UCCSD counts 124, 800, 349 and 361 from a single resource record. A separate pinned-configuration summary contains 120 for LiH and 359 for H_2O. These records describe distinct compilation outputs. Mixing them within a ratio would make the baseline ambiguous. The released resource summary and fit summary retain both conventions with their roles identified.

The resource record's hardware-efficient entries include ideal-error and sector diagnostics that do not establish converged chemical states. Their low gate counts alone do not support a superior accuracy claim. The UCCSD and ADAPT entries retained in the main comparison use their own recorded ideal errors.

S6. Reproducibility procedure

The research-data archive contains selected numerical inputs, documented extracts, offline replay scripts and a released-file hash manifest. Run `python reproduce_energy.py` for the six-qubit statevector and operator replay, then `python verify_results.py` for model residuals, campaign deviations, repeat intervals, resource-summary checks and conditional budgets. Run `python plot_figures.py` to regenerate the four figures. The README specifies dependencies and each check's scope.

The circuit replay starts from the released gates and Pauli coefficients. Other checks reproduce arithmetic from recorded point estimates or summaries; they do not rerun missing optimization jobs or establish original hardware execution. The archived fit confidence intervals remain stored summaries, not newly fitted intervals. Original full-width circuit inputs, complete repeat-result payloads and per-state sector probabilities are outside the released source set. These boundaries limit what can be independently reproduced.