Supporting Information: Limitations of Quantum Hardware for Molecular Energy Estimation Using VQE

Abel Carreras,[*,†] Román Orús,[†,‡,¶] and David Casanova[‡,¶]

†*Multiverse Computing, Donostia, 20014, Euskadi, Spain*

‡*Donostia International Physics Center (DIPC), 20018 Donostia, Euskadi, Spain*

¶*IKERBASQUE, Basque Foundation for Science, 48009 Bilbao, Euskadi, Spain*

E-mail: abelcarreras83@gmail.com

Contents

S1 Effective Hamiltonian

To construct the effective Hamiltonian, we apply the frozen core approximation. In this approach, the molecular spin-orbitals are divided into three sets:

– **Core spin-orbitals**: low-energy spin-orbitals that are treated as an effective potential in the Hamiltonian;

– **Active spin-orbitals**: which are treated explicitly in the Hamiltonian;

– **Truncated spin-orbitals**: corresponding to high-energy spin-orbitals that are not included in the Hamiltonian.

In this work we used a spin-orbital basis, where spin-up (α) and spin-down (β) components are interleaved. In this basis the effective Hamiltonian is defined as

$$\hat{\mathcal{H}}_{\text{eff}} = \sum_{pq}^{\text{act}} \tilde{h}_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs}^{\text{act}} g_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s + \hat{V}_{\text{eff}}, \tag{S.1}$$

where $\tilde{h}_{pq}$ are the effective one-electron integrals, defined as

$$\tilde{h}_{pq} = h_{pq} + \sum_{i}^{\text{core}} \left(g_{iqpi} - g_{iipq} \right), \tag{S.2}$$

with h_{pq} denoting the one-electron integrals from the original Hamiltonian; and the effective potential term $\hat{V}_{\text{eff}}$, accounting for the core spin-orbitals, is given by

$$\hat{V}_{\text{eff}} = \sum_{i}^{\text{core}} h_{ii} + \frac{1}{2} \sum_{ij}^{\text{core}} \left(g_{jiij} - g_{jjii} \right). \tag{S.3}$$

Note that indices i, j run over the core spin-orbitals, while p, q, r, s run over the active spin-orbitals.

S2 Compression of electronic Hamiltonians

In this section, we analyze the impact of Hamiltonian compression on the ground state energy. Additionally, we examine how the size of the Hamiltonian evolves as the number of active space orbitals increases, considering different compression thresholds. For this study, we construct the Hamiltonian of benzene molecule for increasing active space sizes, ranging from 2 to 12 molecular orbitals, and apply different compression thresholds (1e-2, 1e-3, and 1e-4). The results are calculated using canonical orbitals (Figures S3 and S3), and using natural orbitals S1 and S2.

Focusing in natural orbitals, Figure S1 (left) shows the number of Hamiltonian terms as a function of the active space size. As expected, the number of terms increases with the active space size. For small active spaces, there is almost no difference between the three compression thresholds. However, beyond 8 orbitals, a significant difference emerges between 1e-2 and the other two (1e-3 and 1e-4). In this regime, the number of terms is notably reduced for 1e-2, while for 1e-3 and 1e-4, the difference remains marginal.

Figure S1 (right) illustrates the impact of compression from an energy perspective by plotting the ground state (GS) energy as a function of the number of Hamiltonian terms. The results indicate that a compressed Hamiltonian provides a better GS energy-to-terms ratio compared to the uncompressed Hamiltonian. This trend holds for all compression thresholds except 1e-2, where, around 1000 terms, the ratio deteriorates, leading to a GS energy higher than that of the uncompressed Hamiltonian for the same number of terms.

In Figure S2, we analyze the evolution of the GS energy error (defined as the difference between the GS energy of the compressed and uncompressed Hamiltonians) with respect to both the active space size (left) and the number of Hamiltonian terms (right). Even for a compression threshold of 1e-2, the GS energy error remains within chemical accuracy up to an active space of 6 orbitals. For 1e-3 compression, the GS energy remains below chemical accuracy up to 12 molecular orbitals. Notably, the difference between 1e-3 and 1e-4 is minimal.

Examining the energy error as a function of the number of Hamiltonian terms, we observe that a 1e-3 compression is highly efficient up to 6 active orbitals, yielding a Hamiltonian with negligible loss relative to the uncompressed one. Beyond this point, the energy error increases exponentially with both the number of active orbitals and the number of Hamiltonian terms.

These results suggest that as the active space size increases—and thus the accuracy of the GS energy improves—small terms become more relevant, limiting the extent to which compression can be applied. However, for active spaces of up to 6 orbitals, a compression threshold of 1e-2 appears sufficient to maintain the GS energy error below chemical accuracy.

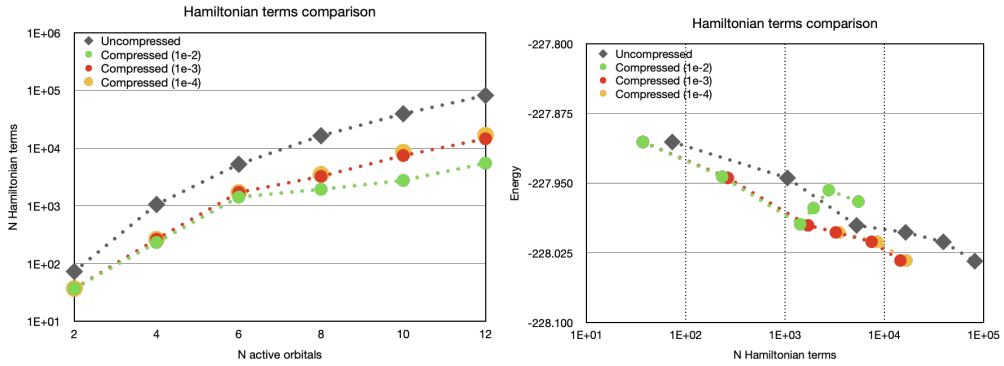

Figure S1: Analysis of the impact of Hamiltonian compression on the number of Hamiltonian terms and the ground state energy. On the left, we show the number of Hamiltonian terms as a function of the active space size for different Hamiltonian with three compression thresholds (1e-2, 1e-3, and 1e-4), compared to the uncompressed Hamiltonian. On the right, we plot the ground state (GS) energy as a function of the number of Hamiltonian terms. The GS energies have been computed using the FCI method for three different compression thresholds (1e-3, 1e-4, and 1e-5). The Hamiltonian is expressed in the basis of natural Orbitals.

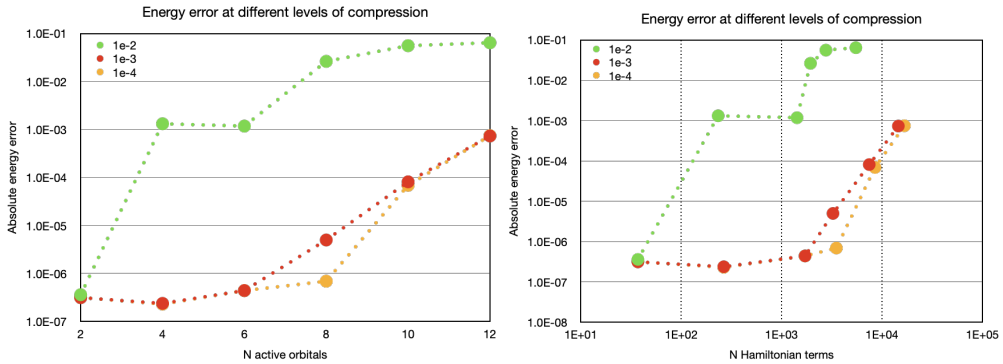

Figure S2: Analysis of the energy error with respect to different Hamiltonian compression thresholds (1e-2, 1e-3, and 1e-4). On the left, we show the ground state energy error as a function of the active space size. On the right, we present the GS energy error as a function of the number of Hamiltonian terms. The ground state energy energies have been computed using the FCI method. The Hamiltonian is expressed in the basis of natural Orbitals.

Examining the results obtained using canonical orbitals (Figures S3 and S4), we observe that the use of canonical orbitals is detrimental compared to the use of natural orbitals. The energy-to-number-of-terms ratio presented in Figure S3 (right) is significantly worse than that observed with natural orbitals, where compression provides little to no benefit at any level. Furthermore, we observe that the GS energy error with respect to the uncompressed Hamiltonian (Figure S4) is substantially higher when using canonical orbitals, frequently exceeding chemical accuracy.

This suggests that the use of natural orbitals leads to Hamiltonians where the most important interactions are concentrated in a smaller number of terms, making the compression procedure more effective. However, it is important to note that as higher precision is required, the contribution of small Hamiltonian terms becomes more significant, and their impact on the ground state energy may no longer be negligible.

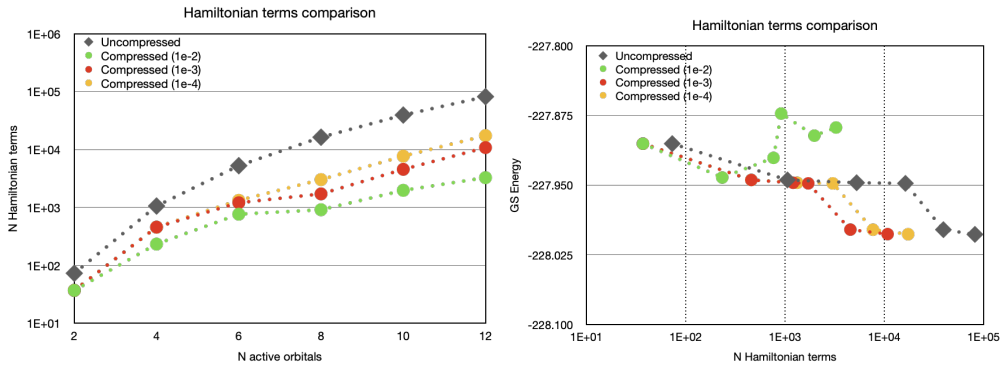

Figure S3: Analysis of the impact of Hamiltonian compression on the number of Hamiltonian terms and the ground state energy. On the left, we show the number of Hamiltonian terms as a function of the active space size for different Hamiltonian with three compression thresholds (1e-2, 1e-3, and 1e-4), compared to the uncompressed Hamiltonian. On the right, we plot the ground state (GS) energy as a function of the number of Hamiltonian terms. The GS energies have been computed using the FCI method for three different compression thresholds (1e-3, 1e-4, and 1e-5). The Hamiltonian is expressed in the basis of canonical orbitals.

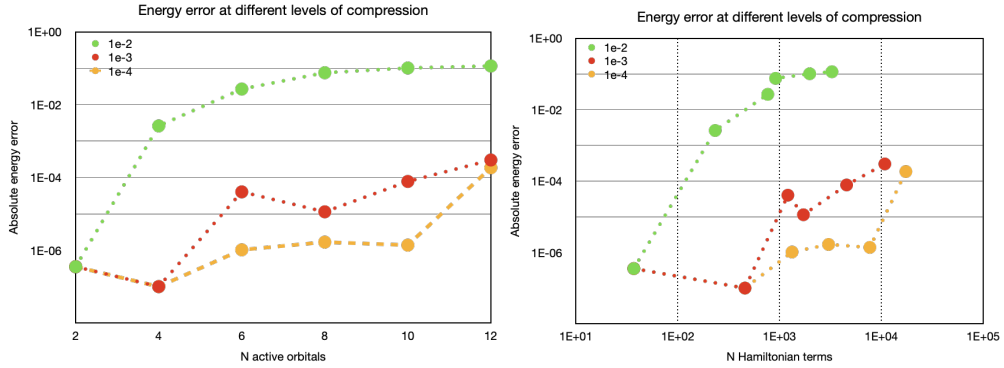

Figure S4: Analysis of the energy error with respect to different Hamiltonian compression thresholds (1e-2, 1e-3, and 1e-4). On the left, we show the ground state energy error as a function of the active space size. On the right, we present the GS energy error as a function of the number of Hamiltonian terms. The ground state energy energies have been computed using FCI method with a bond dimension of 200. The Hamiltonian is expressed in the basis of canonical Orbitals.

To evaluate the impact of Hamiltonian compression in the fermionic-operator basis, we performed a series of exact simulations of benzene molecule with a active space of 4 orbitals using our qubit-ADAPT implementation with a trotterized ansatz analogous to the circuits deployed in the quantum-hardware experiments, and tested multiple compression thresholds. The results, shown in Figure S5, indicate that a compression threshold of (1e-2) yields a

converged energy with an error of approximately (1e-3) Ha relative to the exact ground state, whereas a threshold of (1e-3) reproduces the uncompressed Hamiltonian nearly exactly. The effect of the (1e-2) compression is negligible for the first four selected operators and remains very small up to the first seven; beyond this point, deviations become more noticeable. We do not include smaller compression thresholds, as the results for (1e-2) and the uncompressed Hamiltonian are essentially indistinguishable.

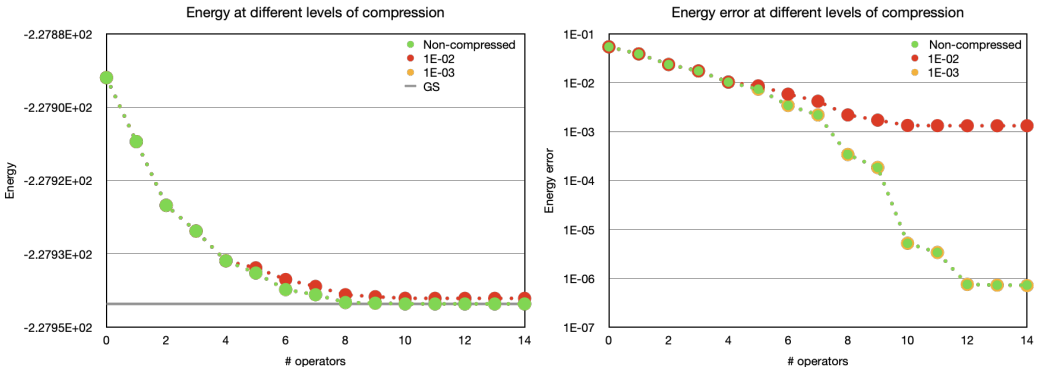

Figure S5: Qubit-ADAPT exact simulation of benzene (4 active orbitals) using a hamiltonian with different compression thresholds (1e-2, 1e-3) in the basis of natural orbitals and computed using a trotterized ansatz (1 step). On the left, we show the energy at each ADAPT-VQE step as a function of the number of operators, on the right the energy error with respect to the exact ground state.

S3 Fermion vs. Qubit operator pools

The quantum circuits generated by fermionic ADAPT-VQE are constructed by mapping fermionic operators to qubit operators and implemented on quantum hardware using a first-order Trotter approximation. We note that this Trotterization can potentially break some symmetries.

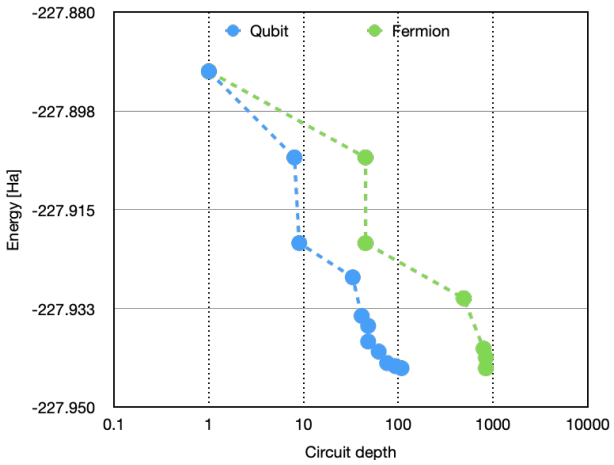

Figure S6: Ground state energy (in Hartrees) as a function of the number of iterations for the simulation of the benzene molecule with fermion (green) and qubit (blue) ADAPT-VQE algorithm.

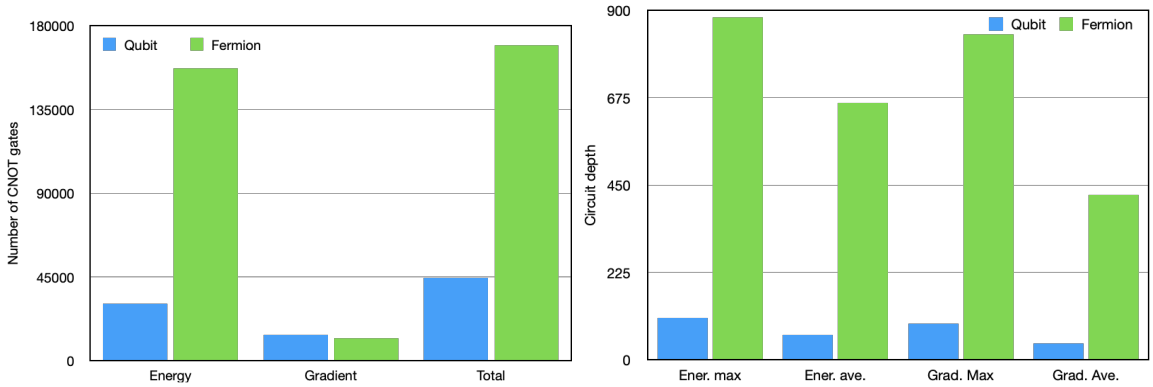

Figure S7: Comparison of the computational cost of VQE algorithms using fermion vs qubit operators for benzene molecule. On the left there is a break down of the total (accumulated) number of CNOTs used for all circuits in the full simulation. We separate this number in energy evaluations and gradient evaluations. On the right there is an analysis of the circuit depths of the circuits used to evaulate the energy and the gradients. We show the maxium circuit depth and the average circuit depth.

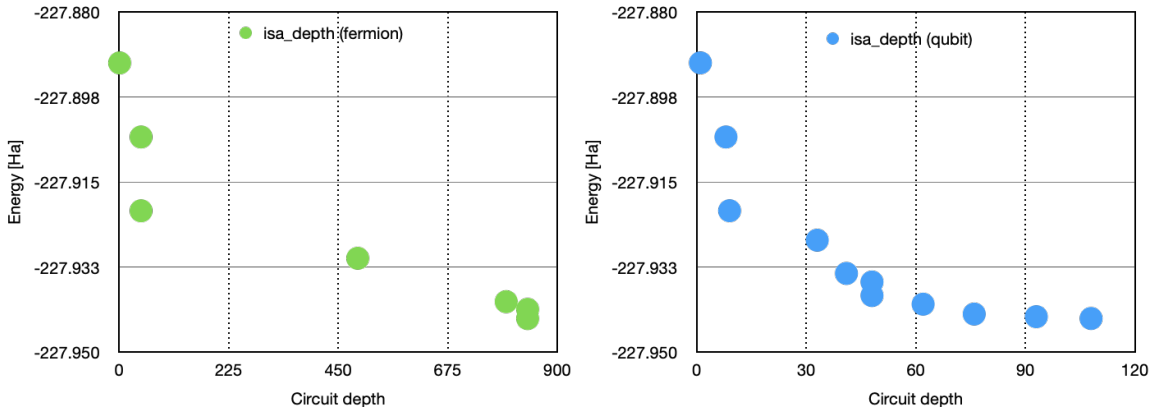

Figure S8: Comparison of the circuit depth (of the circuit used to evaluate the energy) vs energy obtained from an ADAPT-VQE simulation (exact). On the left fermion-ADAPT-VQE, on the right qubit-ADAPT.

S4 Classical optimizer

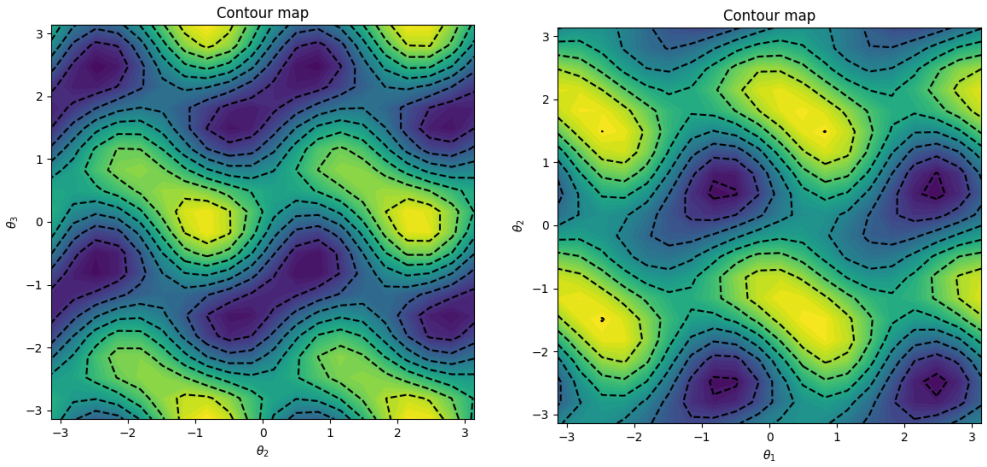

Figure S9: Energy surface generated by the scan over the coefficients of a two operator qubit-ADAPT ansatz of H_4 molecule. Left figure shows the scan along coefficients 1 and 2 and right figure is the scan along coefficients 2 and 3

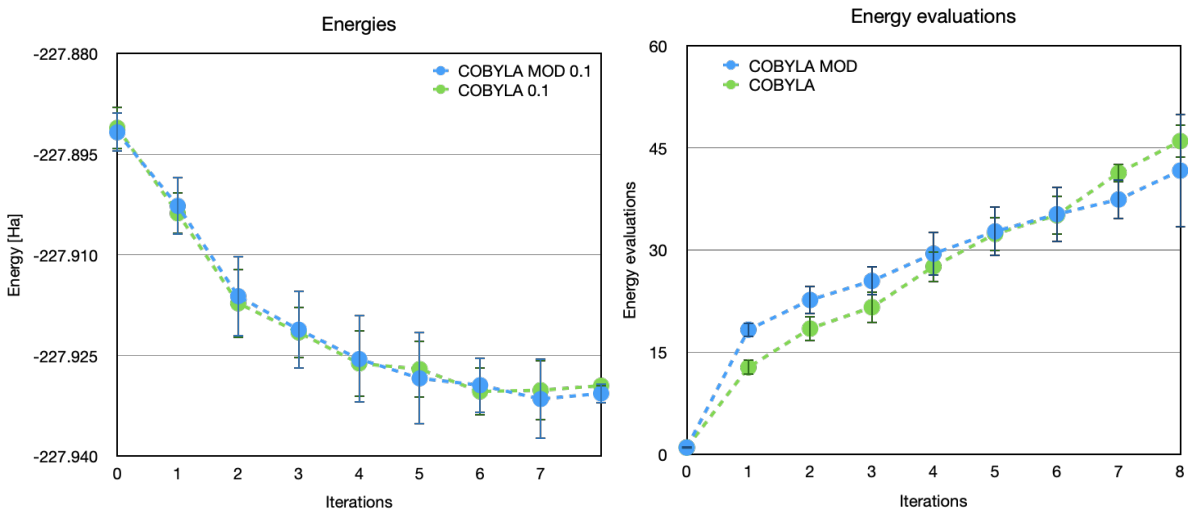

Figure S10: (Left) Ground state energy and (right) energy evaluations of benzene molecule along the qubit-ADAPT iterations using COBYLA (green) and mod-COBYLA (blue). Each point corresponds to the average of 30 simulations using 1000 shots. Default tolerance in standard COBYLA is 10^{-3} Hartrees.

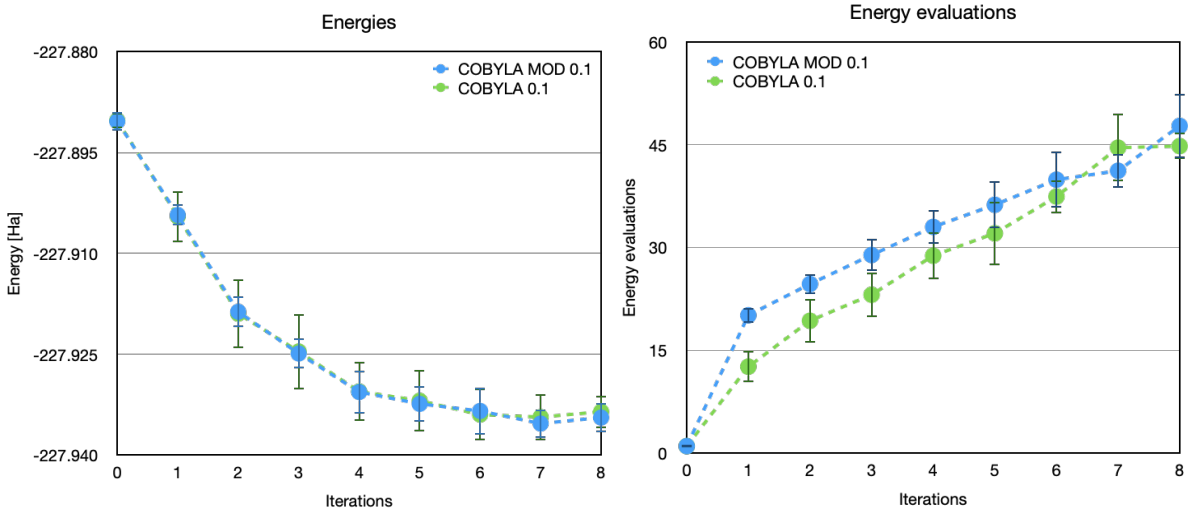

Figure S11: (Left) Ground state energy and (right) energy evaluations per iteration of benzene molecule along the qubit-ADAPT iterations using COBYLA (green) and mod-COBYLA (blue). Each point corresponds to the average of 30 simulations using 5000 shots. Default tolerance in standard COBYLA is 10^{-3} Hartrees.

To estimate the cost of the Qubit-ADAPT algorithm in the quantum hardware we have analyzed the number of measurements that will be necessary per ADAPT-VQE iteration. For this we account for the number of qubit operators to be measures for the evaluation of the energy (89 qubit operators) and the gradients (46 qubit operators). This is summarized

in Figure S12, where we plot the number of energy measurements per iteration using both COBYLA and mod-COBYLA and the number of measurements needed for the gradients per iteration.

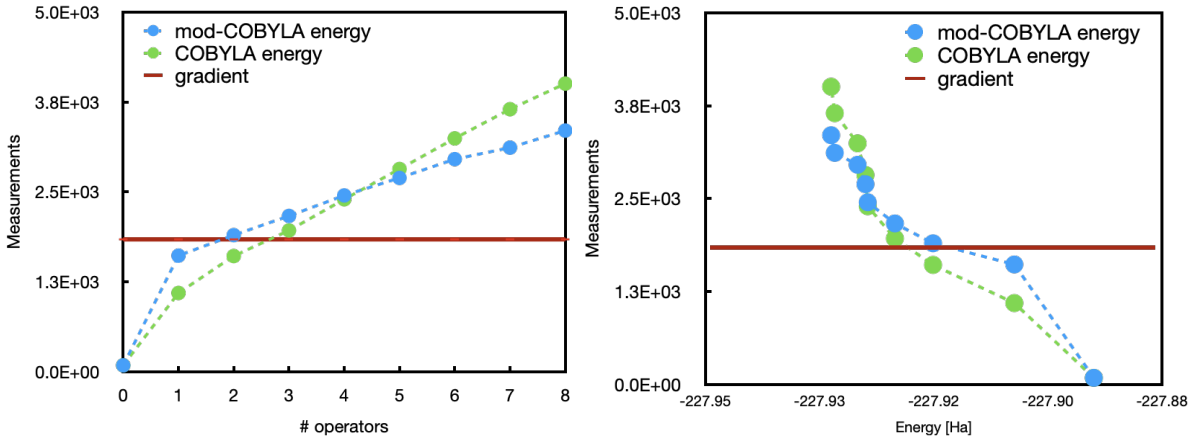

Figure S12: (Left) measurements per iteration of a simulation of the benzene molecule with qubit-ADAPT using COBYLA and mod-COBYLA. (Right) measurements for the same simulation as a function of the energy at each iteration. Qiskit Aer simulator with 500 shots per measurement where used.

As shown, the computational cost of the energy evaluation per iteration increases approximately linearly with the number of operators, whereas the gradient evaluation exhibits a constant cost across iterations, and is determined by the product of the number of shots (500), the number of qubit operators in the pool (40), and the number of qubit terms appearing in the Hamiltonian commutators (46). This leads to a total of 920000 measurements per iteration. For small iteration counts, the gradient evaluation dominates the total computational effort; however, as the number of operators grows and the simulation advances, its relative contribution becomes smaller, consistent with the expected scaling behaviour.

S4.1 Stretched H_4 model

To test the mod-COBLYLA with highly correlated system we have analyzed the linear H_4 molecule with bond distance of 2 Å computed with the STO-3G basis set. For this example we have also used Qiskit Aer simulator to compute the energy with 1000 shots. In figure S13

we show a comparison of a the average of 30 qubit-ADAPT simulations using both COBYLA and mod-COBYLA. In the left plot we show the energy as a function of each iteration step and on the right we plot the number of energy evaluations. For reference we include two exact simulation computed using Fermion operators and qubit operators.

In this example we observe a significant improvement in the convergence of the energy using mod-COBYLA with respect to COBYLA, however they converge to a similar final energy value. Also we observe that for each iteration of the ADAPT-VQE algorithm the number of energy evaluations required for the optimization is lower using mod-COBYLA. Looking at the error bars associates to the values presented in these figures we also observe that dispersion obtained with mod-COBYLA is smaller dispersion than ussing COBYLA. This indicates that the results are more reliable using mod-COBYLA. This can be observed in figure S14 where we compare the energy distribution of the final converged energy using both COBYLA and mod-COBYLA. In this figure we observe that the average energy using COBYLA is slighly smaller but the deviation of the energy distribution is much larger than mod-COBYLA.

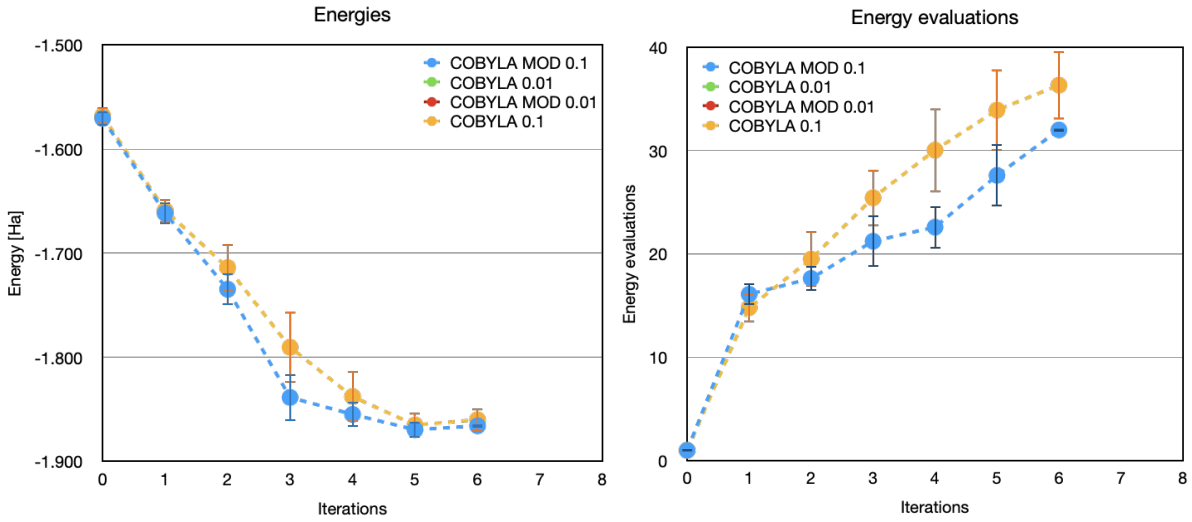

Figure S13: Molecule H_4., Comparison of the optimization using COBYLA and mod-COBYLA using a value of 0.1 for rhobeg. Number of shots is 1000. Default tolerance is 1e-3 Ha.

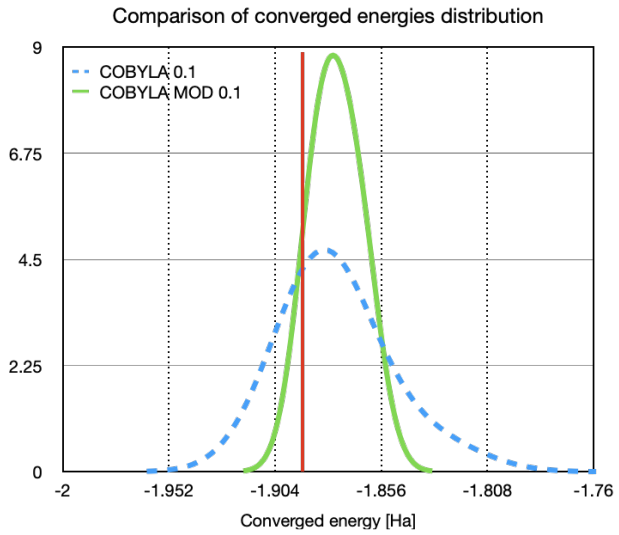

Figure S14: Molecule H_4. Distribution of the energy of the converged ground state. Comparison of the optimization using COBYLA and mod-COBYLA. Number of shots is 1000. Default tolerance is 1e-3 Ha.

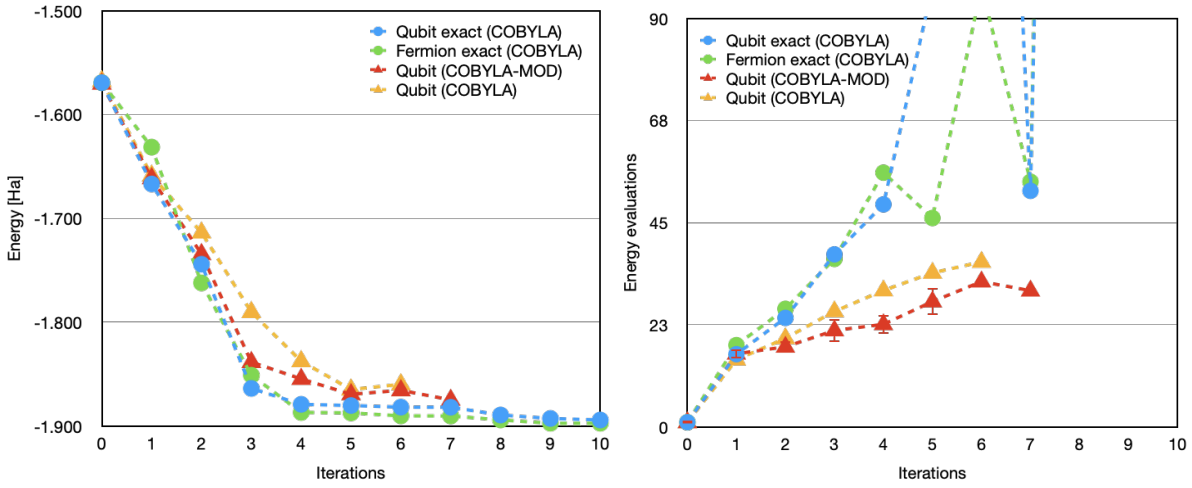

Figure S15: Molecule H_4. Comparison of the optimization using COBYLA and mod-COBYLA. Fermion vs qubit operators using a value of 0.1 for rhobeg. Number of shots is 1000. Energy as a function of the iterations (left) and evaluations of the energy (right)

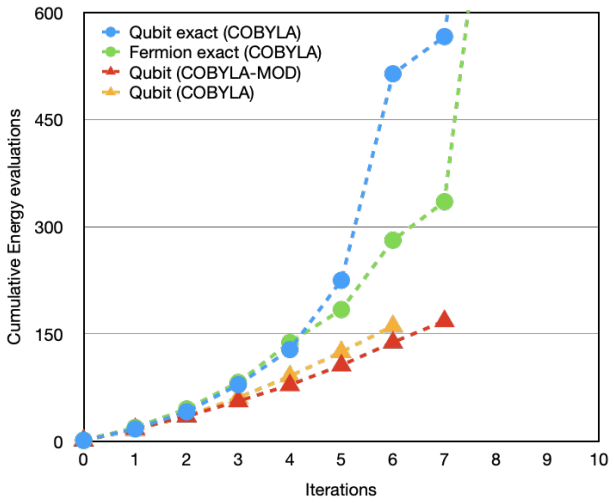

Figure S16: Molecule H_4. Comparison of the optimization using COBYLA and mod-COBYLA. Fermion vs qubit operators using a value of 0.1 for rhobeg. Number of shots is 1000

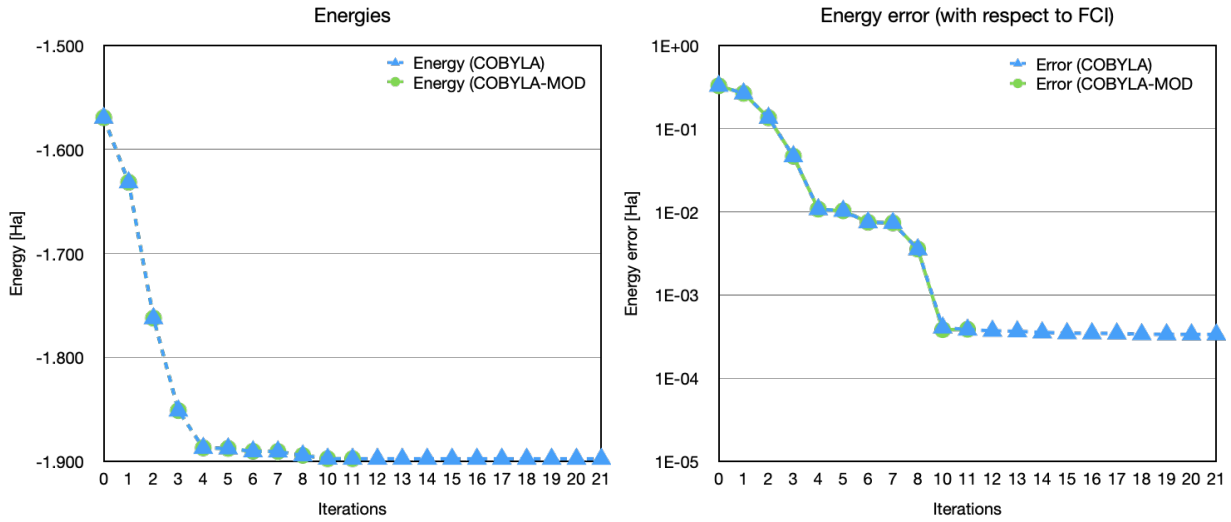

Figure S17: Molecule H_4. Comparison of the optimization using COBYLA and mod-COBYLA. Exact energy evaluation vs ADAPT iterations (left) and energy error with respect to FCI (right)

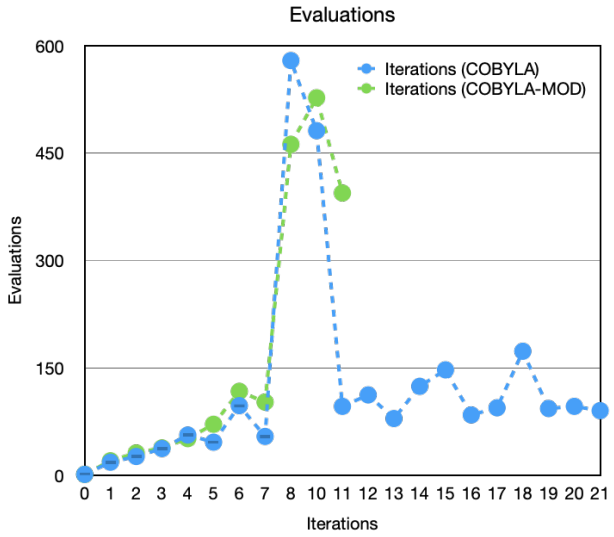

Figure S18: Molecule H_4. Comparison of the optimization using COBYLA and mod-COBYLA. Energy evaluations vs ADAPT iterations

We conducted a targeted ablation study to assess the performance of the mod-COBYLA algorithm relative to the original COBYLA implementation in the scipy library. To this end, we performed a series of Qubit-ADAPT calculations on the H_4 molecule using Qiskit Aer simulations, systematically varying the tolerance parameter of mod-COBYLA. For each parameter set, we carried out 30 independent repetitions to account for statistical noise. In mod-COBYLA, the tolerance is set to the standard error of the energy measurement. In this analysis, we multiply this baseline tolerance by a scaling parameter ("tolerance factor") to examine the stability and suitability of the criterion. We also carried out the same calculations with and without the pre-optimization scan of the initial guess.

The results, shown in Figures S19–S20, indicate that enabling the pre-optimization scan increases the number of energy evaluations required for convergence, yet in general improves the convergence the final energy estimate and reduces its standard error. Varying the tolerance factor reveals the expected trade-off: larger tolerances lead to less accurate converged energies but fewer energy evaluations, whereas smaller tolerances improve accuracy at the expense of increased computational cost. Overall, we find that setting the tolerance equal to the standard error of the energy measurement, as done in mod-COBYLA, offers a balanced

compromise between accuracy and efficiency.

Comparing these results with the original COBYLA algorithm (green line), we observe that mod-COBYLA systematically yields more accurate energies, as evidenced by the lower standard errors. For a tolerance factor equal to 1, the converged energies are nearly identical to those of the original COBYLA, but require consistently fewer energy evaluations. This further supports that our chosen tolerance criterion provides a favourable balance between accuracy and computational cost.

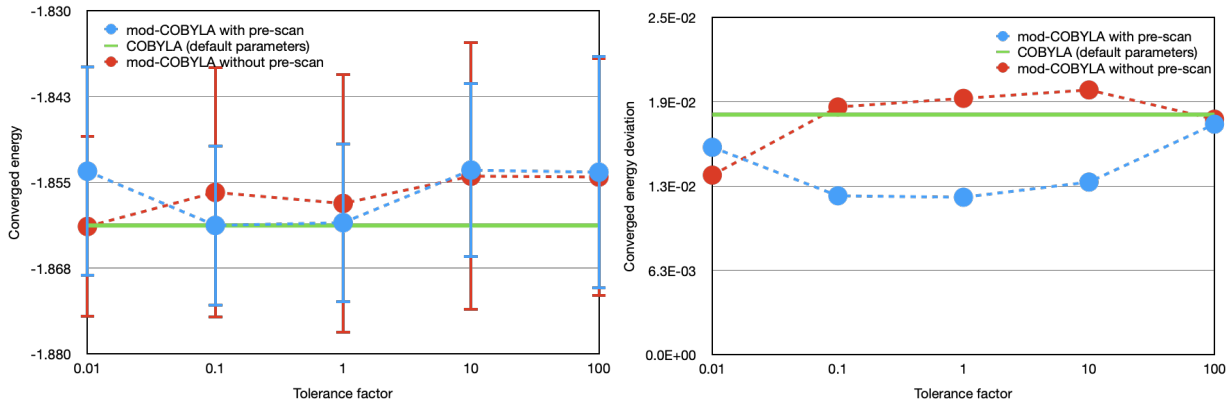

Figure S19: Comparison of optimization performance using the standard COBYLA algorithm and the modified mod-COBYLA variant for the H_4 molecule. The panels show the converged Qubit-ADAPT energy (left) and the corresponding standard error (right) as a function of the tolerance factor. The green line represents the original COBYLA implementation in scipy. Blue symbols correspond to mod-COBYLA calculations including the pre-optimization scan of the initial guess, whereas red symbols indicate mod-COBYLA without the scan. For each tolerance factor, 30 independent repetitions were performed to capture statistical fluctuations. The tolerance in mod-COBYLA is defined as the standard error of the energy measurement, multiplied here by the indicated tolerance factor.

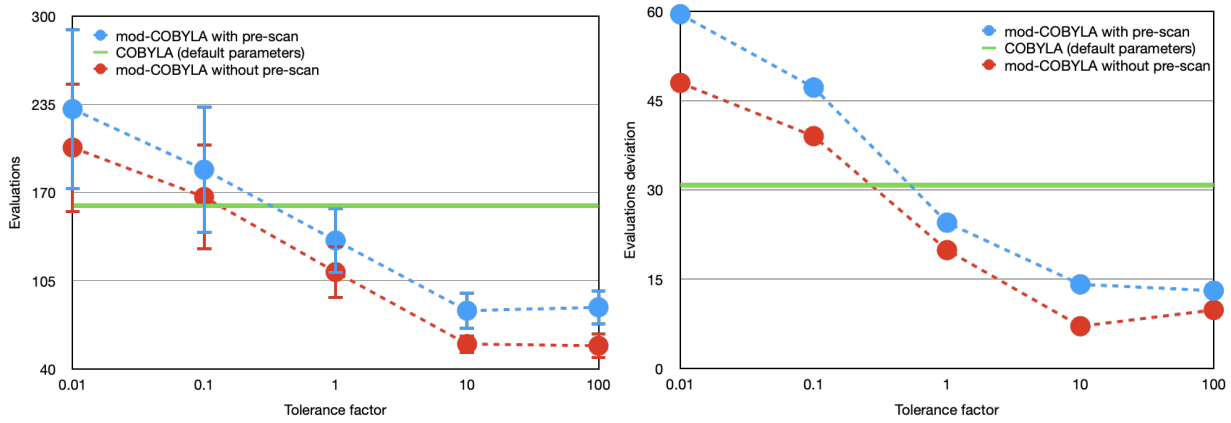

Figure S20: Comparison of optimization performance using the standard COBYLA algorithm and the modified mod-COBYLA variant for the H_4 molecule. The panels show the number of energy evaluations required to converge the Qubit-ADAPT simulation (left) and the corresponding standard error (right) as a function of the tolerance factor. The green line represents the original COBYLA implementation in scipy. Blue symbols correspond to mod-COBYLA calculations including the pre-optimization scan of the initial guess, whereas red symbols indicate mod-COBYLA without the scan. For each tolerance factor, 30 independent repetitions were performed to capture statistical fluctuations. The tolerance in mod-COBYLA is defined as the standard error of the energy measurement, multiplied here by the indicated tolerance factor.

S5 Circuit optimization

Table S1: Circuit depth of the ansatz used at each iteration step of the Qubit-ADAPT using the two different CNOT orientations and the one generated by our algorithm (Optimal). For each version we show also the depth of the transpiled circuit for `IBM torino` quantum computer (ISA depth) using the cascade pattern.

	Standard			Reverse			Optimal		
# op.	depth	CNOT	ISA	depth	CNOT	ISA	depth	CNOT	ISA
0	1	0	1	1	0	1	1	0	1
1	10	6	22	8	6	24	8	6	24
2	10	12	82	8	12	74	8	12	80
3	19	18	126	17	18	132	17	18	132
4	35	32	173	33	32	181	33	32	181
5	42	38	207	39	38	215	39	38	215
6	44	44	197	41	44	215	41	44	215
7	57	54	242	54	54	253	54	54	253

Table S2: Circuit depth of the ansatz used at each iteration step of the Fermion-ADAPT using the two different CNOT orientations and the one generated by our algorithm (Optimal). For each version we show also the depth of the transpiled circuit for `IBM torino` quantum computer (ISA depth) using the cascade pattern.

	Standard			Reverse			Optimal		
# op.	depth	CNOT	ISA	depth	CNOT	ISA	depth	CNOT	ISA
0	1	0	1	1	0	1	1	0	1
1	73	48	110	68	48	64	67	46	64
2	73	96	232	68	96	149	67	92	149
3	696	576	1244	683	576	1059	616	514	1060
4	1110	896	1949	1090	896	1631	979	794	1628
5	1182	944	2072	1157	944	1719	1045	840	1726
6	1254	992	2074	1157	992	1720	1045	886	1718
7	1408	1040	2183	1224	1040	2293	1111	932	1790

S6 Job Size Limitations in IBM Quantum Computers

To estimate the computational limitations of IBM quantum computers, we submitted a series of jobs using the `Estimator` primitive from the IBM Qiskit Runtime. These jobs varied in both the number of two-qubit gates and the number of Pauli observables measured.

The data shown in the plot below reflects the limitations observed at the time of writing this manuscript. A job was considered not allowed if the server returned an error at submission time, indicating that the circuit exceeded current resource or policy constraints.

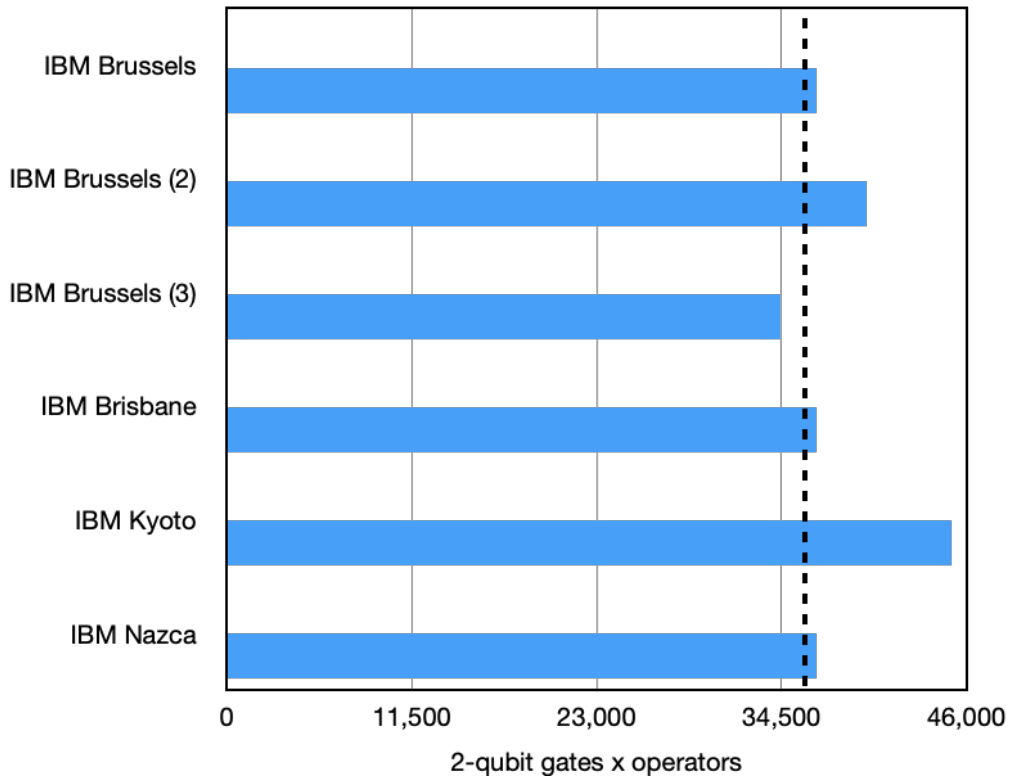

Figure S21: Maximum job size accepted by various IBM quantum computers, measured as the product of the number of two-qubit gates and the number of measured Pauli operators in the observable. IBM Brussels was tested 3 times

S7 Quantum hardware experiments

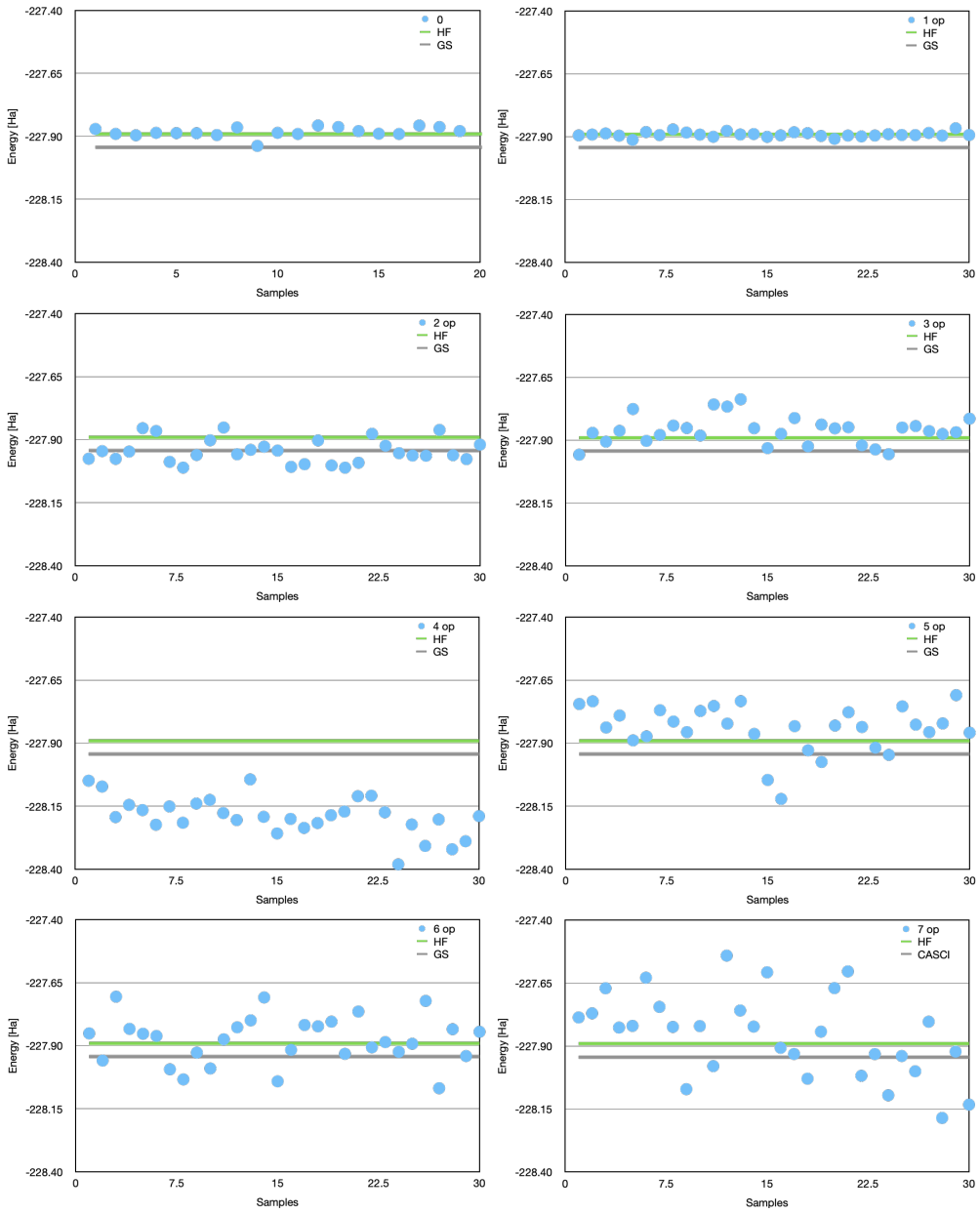

Figure S22: Energy evaluations (in Ha) with Qubit-ADAPT ansätze with 0-7 qubit operators measured in the `IBM Torino` computer using 9000 shots per sample, CASCADE pattern and error mitigation. The coefficients of the ansatz are obtained from an exact calculation. Horizontal lines indicate the exact HF (green) and FCI with 4 active orbitals and 4 active electrons (gray). The energy difference between HF and exact GS (correlation energy) is 53.9 mHa.

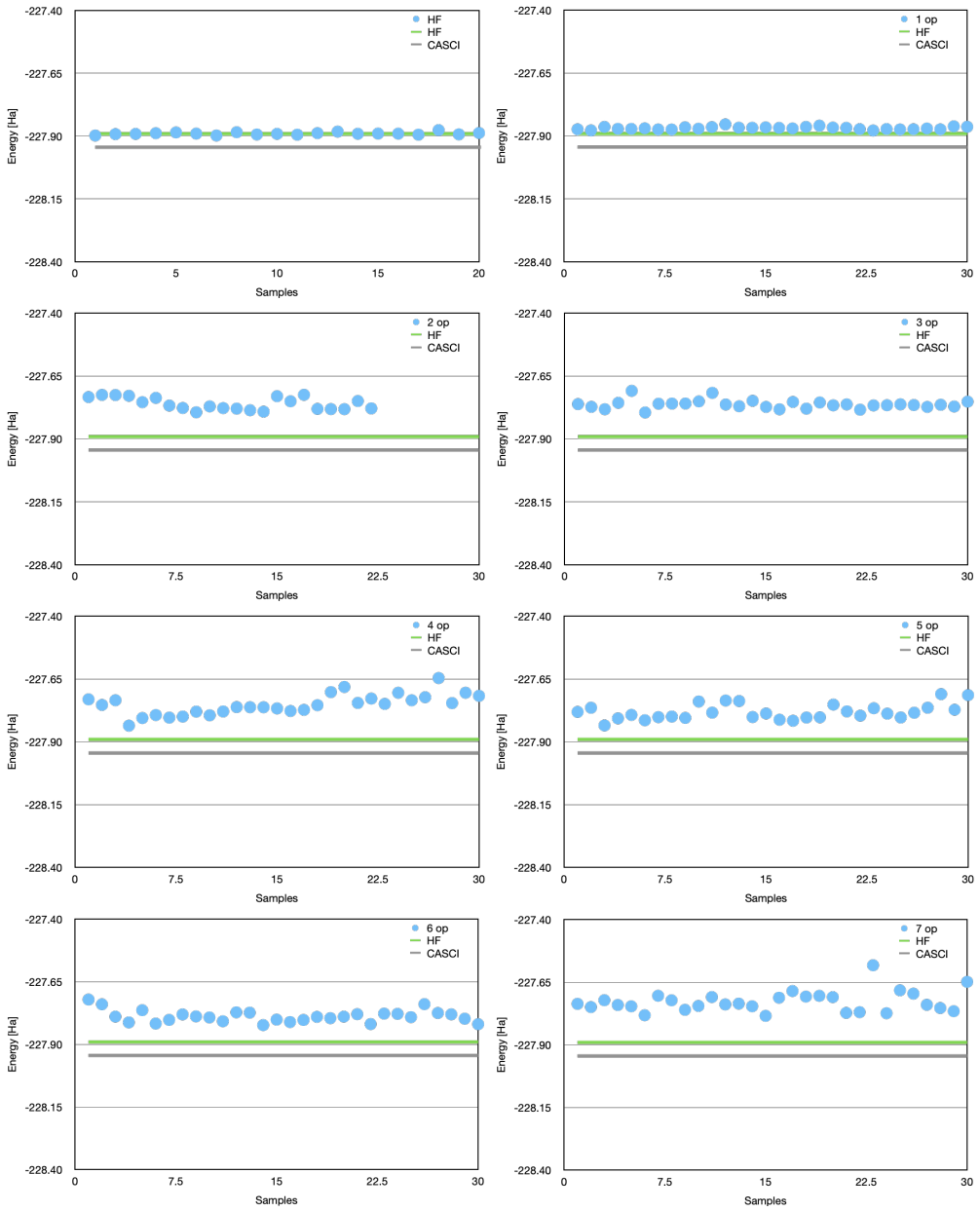

Figure S23: Energy evaluations (in Ha) with Qubit-ADAPT ansätze with 0-7 qubit operators measured in the `IBM Torino` computer using 9000 shots per sample, FAN pattern and error mitigation. The coefficients of the ansatz are obtained from an exact calculation. Horizontal lines indicate the exact HF (green) and FCI with 4 active orbitals and 4 active electrons (gray). The energy difference between HF and exact GS (correlation energy) is 53.9 mHa.

Table S3: Standard energy errors and residual biases obtained on the `IBM torino` quantum machine using the CASCADE pattern, with and without error mitigation. The ISA-compiled circuit depth at each ansatz size (number of operators) is additionally provided.

		mitigated		non-mitigated	
# op	**ISA depth**	**std. err.**	**residual bias**	**std. err.**	**residual bias**
0	1	0.009640	0.007750	0.004314	0.065547
1	24	0.009786	0.014616	0.008208	0.130618
2	80	0.048527	0.024430	0.011697	0.286382
3	132	0.053708	0.065366	0.016977	0.349918
4	181	0.072687	0.255181	0.022630	0.391574
5	215	0.092860	0.092992	0.018493	0.421805
6	215	0.095753	0.060726	0.020422	0.525207
7	253	0.167928	0.083850	0.018801	0.541124

In Tables S3 and S4, we report the results obtained from the quantum-hardware calculations. These tables compare the outcomes of the mitigated and non-mitigated calculations. As observed, the use of error mitigation reduces the residual bias, at the expense of increasing the standard error relative to the non-mitigated results. This behaviour is expected and appears in both circuit patterns. As discussed in the main text, the reduction in residual bias is more pronounced for the FAN pattern, making it more effective than the CASCADE pattern. This improvement can be attributed to the lower ISA circuit depth achieved in the FAN topology, as shown in the tables.

Table S4: Standard energy errors and residual biases obtained on the `IBM torino` quantum machine using the FAN pattern, with and without error mitigation. The circuit depth of the ISA-compiled circuits at each iteration (ISA depth) is also reported.

		mitigated		non-mitigated	
# op	**ISA depth**	**std. err.**	**residual bias**	**std. err.**	**residual bias**
0	1	0.005037	0.000086	0.004314	0.065547
1	19	0.005471	0.036794	0.004525	0.099224
2	70	0.024646	0.160981	0.014175	0.278944
3	124	0.017337	0.164561	0.010135	0.306543
4	155	0.040834	0.182286	0.009426	0.370087
5	184	0.030711	0.153979	0.013950	0.442582
6	195	0.024272	0.150725	0.014881	0.347148
7	213	0.042336	0.212549	0.008964	0.345339

S8 Error-mitigation on IBM Hardware

All hardware results on IBM-Torino using error mitigation were obtained using the Qiskit Runtime Estimator primitive with **resilience level = 2**. At this level, the runtime automatically enables a Zero Noise Extrapolation (ZNE) procedure combined with measurement error mitigation.

ZNE was performed using the built-in noise-scaling method based on gate folding, with scaling factors 1,3,5. The extrapolation to the zero-noise limit employed an exponential fit of the measured expectation values with linear fallback when variance is large.

Pauli-based TREX (Twirled Readout Error eXtinction) is applied with a per-job calibration cadence, meaning that the calibration circuits required for the readout-mitigation model are freshly executed for each submitted job rather than reused across runs. TREX relies on twirled measurements, implemented by randomly replacing each measurement gate with a three-step sequence consisting of (i) application of a Pauli-X gate, (ii) measurement, and (iii) classical bit-flip post-processing.

The twirling randomization mode is set to automatic (IBM Runtime default), which selects random twirling operations according to the backend configuration and the active

mitigation pipeline, as specified in the IBM Runtime twirling options documentation.

S9 Thermal relaxation noise model

In Figure S24, we present an additional view of the data from the simulations using the noise models described in the main text. This representation highlights more clearly that, under the noise model, ansätze with fewer operators are less affected by decoherence, while larger ansätze systematically yield higher energies. As the coherence factor increases, the energies obtained for all ansatz sizes converge toward their exact values.

In Figure S24 (right), we show the number of ansätze with residual biases smaller than 1 mHa. This plot reveals that, for coherence factors of approximately 100 or higher, all ansätze fall below this threshold.

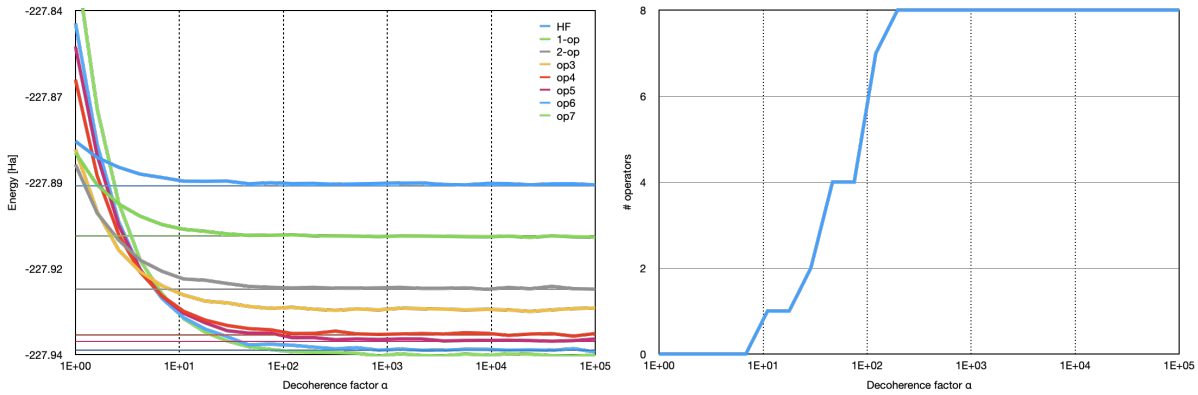

Figure S24: Simulation of the benzene molecule using qubit-ADAPT with a thermal relaxation noise model in Qiskit Aer. (Left) Energy of the ansatz with different numbers of operators as a function of the coherence factor; horizontal lines indicate the exact energies for each ansatz size. (Right) Number of operators exhibiting residual biases larger than 1 mHa as a function of the coherence factor.

S9.1 Instruction execution times

The instruction times used in our Qubit-ADAPT simulations with the thermal relaxation noise model are summarized in Table S5, corresponding to:

- Parameterized single-qubit rotation gates

  - U1: phase shift gate.
  - U2: single-qubit rotation with two parameters.
  - U3: general single-qubit unitary rotation with three parameters.

- CX: CNOT gate.
- Reset: reset operation initializing a qubit back to the $|0\rangle$ state.
- Measure: the measurement operation collapses a qubit's state into either $|0\rangle$ or $|1\rangle$ and records the result.

Table S5: Instruction execution times (in ns) used in the thermal relaxation noise model.

instruction	time
U1	0
U2	50
U3	100
CX	300
Reset	1000
Measure	1000