

Cost-effective scalable quantum error mitigation for tiled Ansätze

Oskar Graulund Lentz Rasmussen,^{*,†} Erik Kjellgren,[†] Peter Reinholdt,[†] Stephan

P. A. Sauer,[‡] Sonia Coriani,[¶] Karl Michael Ziems,[§] and Jacob Kongsted[†]

[†]*Department of Physics, Chemistry and Pharmacy, University of Southern Denmark,
Campusvej 55, DK-5230 Odense M, Denmark*

[‡]*Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100
Copenhagen Ø, Denmark*

[¶]*Department of Chemistry, Technical University of Denmark, Kemitorvet Building 207,
DK-2800 Kongens Lyngby, Denmark*

[§]*School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK*

E-mail: oskr@sdu.dk

Abstract

We introduce a cost-effective quantum error mitigation technique that builds on the recent Ansatz-based gate and readout error mitigation method (M0). The technique, tiled M0, leverages the unique structure of tiled Ansätze (e.g., tUPS, QNP, hardware-efficient circuits) to apply a locality approximation to M0 that results in an exponential reduction in the QPU cost of the noise characterization. We validate the technique for molecular ground state energy calculations with the tUPS Ansatz on LiH, H₂, H₂O, butadiene, and benzene (4 – 12 qubits), demonstrating little to no loss in accuracy compared to M0 in noisy simulations. We also show the performance of the technique in quantum experiments, highlighting its potential use in near-term applications.

1 Introduction

Quantum computing is thought to offer solutions to problems in quantum chemistry that are intractable classically. Given the severity of noise on current noisy intermediate-scale quantum (NISQ) computers,¹ however, the touted benefits cannot yet be fully realized. Nonetheless, through the use of quantum error mitigation, which seeks not to correct but to mitigate errors through classical data processing methods, it is possible to perform more accurate computational tasks with existing devices. Thus, even though hardware improvements (i.e., larger, faster, and less error-prone quantum computers) together with quantum error correction promise to unlock reliable quantum computing in the future,^{2,3} error mitigation will continue to play an important role in the current NISQ era and also beyond.⁴

Many error mitigation techniques, including the one we propose herein, require some characterization and knowledge of the noise on the quantum computer, which is then used to remove errors from measurement results through different schemes. In general, error mitigation techniques increase the sampling cost needed to estimate an observable with a desired accuracy and certainty. The sampling cost has been shown to increase exponentially with the circuit depth and gate error rates, regardless of the specifics of the error mitigation protocol.^{5,6} This makes error mitigation useful only if the noise is at a manageable level or if the circuit is sufficiently shallow. A number of different error mitigation techniques have been developed for quantum computing such as zero-noise extrapolation (ZNE),⁷⁻⁹ probabilistic error cancellation (PEC),^{7,10} readout error mitigation (REM),¹¹ Clifford data regression (CDR),¹²⁻¹⁵ tensor-network error mitigation (TEM)¹⁶ and Ansatz-based gate and readout error mitigation (M0).¹⁷ An introduction to the field of quantum error mitigation, as well as a review of some of the aforementioned techniques, is given in Ref. 18.

The main idea behind our proposed error mitigation technique is to utilize the layered structure of tiled Ansätze such as the tiled unitary product state (tUPS) Ansatz¹⁹ to characterize gate and readout noise efficiently. We focus on the tUPS Ansatz, but in principle the technique also works for other tiled Ansätze such as QNP²⁰ and hardware-efficient Ansätze

with tiled structures.^{21–26} Our technique builds on the recently proposed M0 error mitigation method by Ziem, Kjellgren et al.¹⁷ where the authors extend the REM technique by not only measuring the qubits during the noise characterization step, but by also including the Ansatz with all variational parameters set to 0 (an identity operation on a noiseless device) with the intent of encoding gate noise in addition to readout noise in the assignment matrix. The M0 technique has been used in hardware experiments^{17,27} and simulated noise studies,²⁸ but it requires the construction of a single fully correlated assignment matrix, which imposes an exponential QPU overhead with respect to the number of qubits, thereby limiting its application to small systems.

In our proposed technique, which we refer to as *tiled M0*, we combine M0 with a locality approximation that exploits the repeating structure of tiled Ansätze. The approximation is similar to the tensor product approach suggested for REM in Ref. 11, and it allows us to reduce the exponential scaling in the QPU processing time required for the noise characterization to a constant independent of the system size. The reduction is achieved by approximating the full-system assignment matrix through a number of smaller matrices, one for each tile in the Ansatz. Crucially, in tiled Ansätze, where an increase of layers leads to convergence to the exact solution,^{19,29} tiled M0 also works independently of the layer depth. The downside to the locality approximation in tiled M0 is that it partially neglects correlated noise, but we do not find this to be a significant issue (see Sec. 4).

We show the performance of tiled M0 for ground state energy calculations with the tUPS Ansatz on different molecules in quantum experiments using IBM quantum computers and in noisy simulations. We look at LiH and H₂ with (2, 2) active spaces and up to four layers in the tUPS Ansatz, butadiene and H₂O with (4, 4) active spaces and one layer in the Ansatz, and benzene with a (6, 6) active space and one layer in the Ansatz.

2 Background

2.1 Tiled unitary product state Ansatz

The tUPS Ansatz introduced by Burton in Ref. 19 is a commonly employed tiled Ansatz. It is built from a repeating structure of layers, and layers themselves consist of smaller repeating units called tiles (see Fig. 1 for an illustration). A tile is a set of gates that cause electronic transitions between four spin orbitals using three variational parameters, which ensures that any state can be reached in the subset.^{19,30} Each layer consists of two columns of tiles (one if the number of qubits is less than 6). The tiles in the second column of a layer in tUPS are offset compared to the tiles in the first column, which is necessary to ensure that any electronic configuration can be reached, given enough layers.²⁹ Generally, the more layers that are included in the Ansatz, the greater the expressivity. In tiled M0, we use the fact that there are only ever four qubits associated with a tile, and tiles are exactly the same across layers, meaning that qubits and gates stay the same for a given tile. Following Ref. 19, the unitary operator for a tile can be expressed in terms of the fermionic excitation operators $\hat{\kappa}_{pq}^{(1)}$ and $\hat{\kappa}_{pq}^{(2)}$ as

$$\hat{U}_{pq}^{(m)} = \exp\left(\theta_{pq,1}^{(m)} \hat{\kappa}_{pq}^{(1)}\right) \exp\left(\theta_{pq,2}^{(m)} \hat{\kappa}_{pq}^{(2)}\right) \exp\left(\theta_{pq,3}^{(m)} \hat{\kappa}_{pq}^{(1)}\right), \quad (1)$$

where p and q are spatial orbital indices and m refers to the layer number (see Fig. S1 in the supporting information for a quantum circuit diagram). The tUPS Ansatz for an n -qubit system is then

$$\hat{U}_{\text{tUPS}}(\theta) = \prod_{m=1}^L \left(\prod_{p=1}^R \hat{U}_{2p+1,2p}^{(m)} \prod_{p=1}^S \hat{U}_{2p,2p-1}^{(m)} \right), \quad (2)$$

where L is the number of layers, and $S = \text{floor}(\frac{n}{4})$ and $R = \text{floor}(\frac{n-2}{4})$ are the number of tiles in the first and second column, respectively. The tile numbering t_i as shown in Fig. 1 relates to the tUPS Ansatz numbering in terms of p and q as follows (with one-based indexing for

p and q):

$$p = \begin{cases} 2i - 1, & \text{if } i \leq S \quad (\text{first columns}); \\ 2(i - S), & \text{otherwise} \quad (\text{second columns}). \end{cases} \quad (3)$$

$$q = p + 1 \quad (4)$$

These in turn relate to the qubit number on which a tile starts by $q_{\text{num}} = 2(p - 1)$. For clarity and ease of presentation, in the following we will only refer to the tile numbering, t_i , and refer to the above for connecting a specific tile to its unitary and circuit implementation.

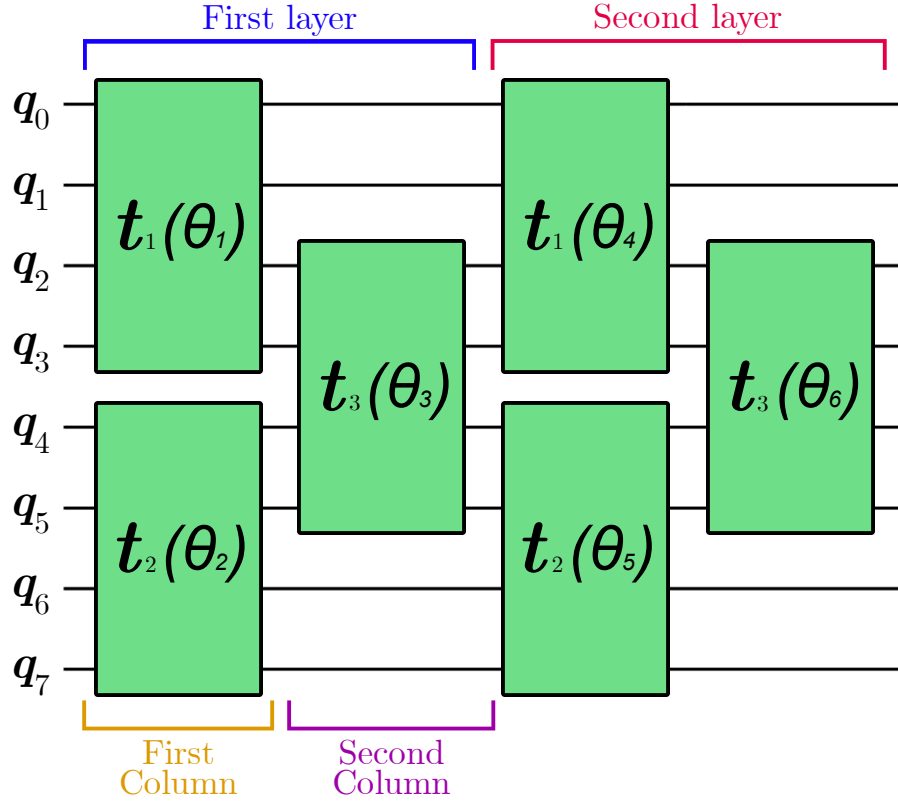


Figure 1: The tUPS Ansatz with two layers for an 8-qubit system. Each tile (labelled t_i) consists of gates that correspond to single and double electronic excitations. Three variational parameters are associated with every tile. The input qubit register is initialized to the desired reference state.

2.2 Readout error mitigation

Our proposed error mitigation technique (*vide infra*) is based on readout error mitigation (REM). REM was introduced by Bravyi et al.¹¹ to correct readout noise under the assumption that the noise can be modelled by a stochastic matrix, \mathbf{A}^{REM} , called an assignment or confusion matrix. The entry $(\mathbf{A}^{\text{REM}})_{yx}$, which is referred to as a transition probability, gives the probability of measuring the bit string $|y\rangle$ when the system was prepared in the basis state $|x\rangle$:

$$(\mathbf{A}^{\text{REM}})_{yx} = \Pr(|y\rangle \mid |x\rangle). \quad (5)$$

To construct the full assignment matrix, every possible bitstring $|x\rangle \in \{|b_i^{\text{REM}}\rangle\}$ needs to be prepared and measured. These are defined as

$$|b_i^{\text{REM}}\rangle = \hat{X}_i |0\rangle^{\otimes n} \quad (6)$$

where \hat{X}_i is an operator with the X -gates necessary to create any bit string $|b_i\rangle$. When the state $|x\rangle \in \{|b_i^{\text{REM}}\rangle\}$ has been prepared according to Eq. (6), the entry $(\mathbf{A}^{\text{REM}})_{yx}$ is estimated from a finite number of K measurement results, $|s_1\rangle, \dots, |s_K\rangle$, as

$$(\mathbf{A}^{\text{REM}})_{yx} \approx K^{-1} \sum_{i=1}^K \langle s_i | y \rangle. \quad (7)$$

The idea proposed with the M0 technique¹⁷ is to include not only X -gates when the basis states are prepared, but the full Ansatz \hat{U} with all variational parameters set to 0, i.e.

$$|b_i^{\text{M0}}\rangle = \hat{U}(0) \hat{X}_i |0\rangle^{\otimes n}, \quad (8)$$

and use this to construct the M0 assignment matrix analog to \mathbf{A}^{REM} , \mathbf{A}^{M0} , with equations analogous to Eqs. (5) and (7) for the entries. For Ansätze such as tUPS, $\hat{U}(0) = \hat{I}$ in

the absence of hardware noise, and the inclusion of $\hat{U}(0)$ has no effect. On noisy devices, however, the errors associated with the gates in \hat{U} will influence the transition probabilities, thereby encoding gate noise in the assignment matrix. In the remainder of the text, we refer to assignment matrices \mathbf{A} constructed from a general basis state preparation circuit and omit qualifiers such as REM and M0 unless there is need for a distinction.

2.3 Expectation value estimation

The error characterized in \mathbf{A} is applied to a subsequent noisy simulation or quantum experiment as follows: Given a set of N noisy measurement results, $|r_1\rangle, \dots, |r_N\rangle$, the error mitigated outcomes are obtained by multiplying with the inverse of the assignment matrix, i.e. $\mathbf{A}^{-1}|r_1\rangle, \dots, \mathbf{A}^{-1}|r_N\rangle$. In reality, defining the raw (unmitigated) probability vector $|p_{\text{raw}}\rangle = N^{-1} \sum_{i=1}^N |r_i\rangle$, we can write the error mitigated results as $|p_{\text{mitigated}}\rangle = \mathbf{A}^{-1}|p_{\text{raw}}\rangle$. Thus, an estimate of the expectation value of an operator \hat{O} which is diagonal in the computational basis is obtained as

$$\bar{O} = \sum_{x \in \{0,1\}^n} O_x \langle x | \mathbf{A}^{-1} | p_{\text{raw}} \rangle \quad (9)$$

where n is the number of qubits and O_x is the eigenvalue corresponding to the computational basis state $|x\rangle$.

2.4 Tiled error mitigation

The major drawback of REM and M0 is the exponential scaling in the QPU processing time needed for the noise characterization which arises from the scaling of the bitstring space (see Eqs. (6) and (8)). To alleviate this problem, we combine M0 with the concept of noise locality from the REM tensor product model.¹¹ In the tensor product model, noise is assumed to act locally on individual qubits and \mathbf{A} can be written as a tensor product of the 2×2 assignment

matrices for all qubits, i.e.

$$\mathbf{A} = \mathbf{A}_{q_{n-1}} \otimes \cdots \otimes \mathbf{A}_{q_0} \quad (10)$$

where \mathbf{A}_{q_i} is the assignment matrix for qubit i . The approximation results in an exponential saving in the QPU cost needed to construct \mathbf{A} . In tiled M0, we assume noise locality for qubits that share a common tile in the tUPS Ansatz, and the crux of our technique is that we construct a 16×16 assignment matrix for every tile in the first layer of the Ansatz. For ease of explanation, assume in the following that there is no readout but only gate noise (we deal with readout noise afterwards).

Let t_i refer to tile number i in the tUPS Ansatz (see Fig. 1 for the ordering) and let \mathbf{A}_{t_i} be the corresponding tile assignment matrix (we explain what quantum circuits we execute to determine \mathbf{A}_{t_i} momentarily). We assume that tile assignment matrices are the same across layers, i.e. $\mathbf{A}_{t_i} = \mathbf{A}_{t_{(i+mT)}}$ where T is the number of tiles in the first layer of the Ansatz and $m \in \mathbb{N}$. We then approximate the full n -qubit assignment matrix for $n \geq 4$ and n even as

$$\mathbf{A} = \begin{cases} \left[(\mathbf{I}^{\otimes 2} \otimes \mathbf{A}_{t_T} \otimes \cdots \otimes \mathbf{A}_{t_{S+1}} \otimes \mathbf{I}^{\otimes 2}) (\mathbf{A}_{t_S} \otimes \cdots \otimes \mathbf{A}_{t_1}) \right]^L, & \text{if } 4 \mid n ; \\ \left[(\mathbf{A}_{t_T} \otimes \cdots \otimes \mathbf{A}_{t_{S+1}} \otimes \mathbf{I}^{\otimes 2}) (\mathbf{I}^{\otimes 2} \otimes \mathbf{A}_{t_S} \otimes \cdots \otimes \mathbf{A}_{t_1}) \right]^L, & \text{otherwise;} \end{cases} \quad (11)$$

where $S = \text{floor}(\frac{n}{4})$ is the number of tiles in the first column of the Ansatz as defined previously and $T = \text{floor}(\frac{n}{2}) - 1$. T scales linearly with the number of qubits, so we only need to construct a linearly scaling number of tile assignment matrices to approximate \mathbf{A} .

The assignment matrix for a given tile is determined by including the Ansatz gates that belong to that tile (with all parameters set to 0). Since each tile has four qubits, we need to prepare only 16 different computational basis states on the QPU for every tile to determine the transition probabilities (with a sufficient amount of shots for each). Moreover, since tiles that lie in the same column in the tUPS Ansatz have no overlapping qubits, we can prepare and measure the 16 computational basis states for each of those tiles in parallel. For the example in Fig. 1, this means that we determine \mathbf{A}_{t_1} and \mathbf{A}_{t_2} from the same probability

vectors. Since there are two columns in a layer (for systems comprised of 6 or more qubits), and since we assume that gate noise is the same across layers, we need to execute only 32 different quantum circuits to construct all tile assignment matrices for arbitrary layers and depth. We now explain how readout noise fits in this picture.

2.5 Preventing the overcorrection of readout noise

When we determine the transition probabilities as described in the previous section, the tile assignment matrices will encode not only gate noise but also readout noise. If at any point we have two or more tile assignment matrices in Eq. (11) that share any qubits, we will correct the readout noise for those qubits more than once. This is expected to introduce a bias in the error-mitigated results, which will become worse the more tiles and layers we have. To fix this problem, we assume that any assignment matrix can be written as the product of a readout part and a gate part

$$\mathbf{A}_{t_i} = \mathbf{A}_{t_i}^{\text{REM}} \mathbf{A}_{t_i}^{\text{gate}} . \quad (12)$$

This means that we assume that gate and readout noise are completely independent of each other. Since $\mathbf{A}_{t_i}^{\text{REM}}$ can be determined separately for every tile using the traditional REM approach, we can remove the readout part from \mathbf{A}_{t_i} , thereby obtaining the pure gate noise

$$\mathbf{A}_{t_i}^{\text{gate}} = (\mathbf{A}_{t_i}^{\text{REM}})^{-1} \mathbf{A}_{t_i} . \quad (13)$$

This yields the full-system gate error assignment matrix similar to Eq. (11) as

$$\mathbf{A}^{\text{gate}} = \begin{cases} \left[(\mathbf{I}^{\otimes 2} \otimes \mathbf{A}_{t_T}^{\text{gate}} \otimes \dots \otimes \mathbf{A}_{t_{S+1}}^{\text{gate}} \otimes \mathbf{I}^{\otimes 2}) (\mathbf{A}_{t_S}^{\text{gate}} \otimes \dots \otimes \mathbf{A}_{t_1}^{\text{gate}}) \right]^L, & \text{if } 4 \mid n ; \\ \left[(\mathbf{A}_{t_T}^{\text{gate}} \otimes \dots \otimes \mathbf{A}_{t_{S+1}}^{\text{gate}} \otimes \mathbf{I}^{\otimes 2}) (\mathbf{I}^{\otimes 2} \otimes \mathbf{A}_{t_S}^{\text{gate}} \otimes \dots \otimes \mathbf{A}_{t_1}^{\text{gate}}) \right]^L, & \text{otherwise.} \end{cases} \quad (14)$$

This matrix encodes, ideally, all the gate noise of the Ansatz but none of the readout noise. Note that we need to execute 64 different quantum circuits in total to construct the gate tile assignment matrices — a number that is independent of the system size (for systems comprised of 6 or more qubits). The full-system readout assignment matrix \mathbf{A}^{REM} can be approximated as

$$\mathbf{A}^{\text{REM}} = \begin{cases} \mathbf{A}_{t_S}^{\text{REM}} \otimes \dots \otimes \mathbf{A}_{t_1}^{\text{REM}}, & \text{if } 4 \mid n ; \\ \mathbf{A}_{q_{n-1}, q_{n-2}}^{\text{REM}} \otimes \mathbf{A}_{t_S}^{\text{REM}} \otimes \dots \otimes \mathbf{A}_{t_1}^{\text{REM}}, & \text{otherwise;} \end{cases} \quad (15)$$

where $\mathbf{A}_{q_{n-1}, q_{n-2}}^{\text{REM}}$ is the 4×4 two-qubit readout assignment matrix for the last two qubits. Thus, the readout error characterization uses the first column REM assignment matrices $\mathbf{A}_{t_i}^{\text{REM}}$ that we construct to determine $\mathbf{A}_{t_i}^{\text{gate}}$ anyways, and therefore comes at no additional cost in the case of $4 \mid n$ or with only 4 additional circuits to measure otherwise. This means that the complete noise characterization costs at most $(64 + 4) \times K$ quantum circuit measurements, where K is the shot count. We choose $K \approx 15,000$ in our work, independent of the system size. More details follow in Sec. 3.

2.6 Overhead screening

For the tensor product model, Bravyi et al. show in Ref. 11 that the number of shots needed to estimate the expectation value of an observable with a desired accuracy and certainty scales exponentially — compared to the noiseless case — with the sum of the off-diagonal entries over all \mathbf{A}_{q_i} in Eq. (10). The authors define the noise strength

$$\gamma = \sum_{i=0}^{n-1} \max \left[(\mathbf{A}_{q_i})_{10}, (\mathbf{A}_{q_i})_{01} \right] \quad (16)$$

and show that the sampling cost (i.e., shot count) scales with $e^{4\gamma}$. This is referred to as the error mitigation overhead. Since the assignment matrices \mathbf{A}_{q_i} can be approximated at very

little QPU cost (we use $2 \times 15,000$ shots independent of the system size), they can be used as a cheap way to estimate the overhead for a given circuit and determine the feasibility of performing a calculation. Note that we determine the matrices \mathbf{A}_{q_i} with the inclusion of the full tUPS circuit, $\hat{U}_{\text{tUPS}}(0)$.

We observe on hardware that γ can vary considerably from day to day and with the quality of the qubits that one’s circuit is mapped to on the device. Most of the calculations we show in this work were done after approximating all \mathbf{A}_{q_i} matrices and determining γ through Eq. (16). In most cases, we deemed the overhead to be at a reasonable level and proceeded with the calculation, but in a few cases the noise was found to be unacceptably high and we postponed the calculation.

3 Computational Details

We performed ground state energy calculations with the tUPS Ansatz on equilibrium geometries of: 1) LiH and H₂ with (2, 2) active spaces and up to four layers in the tUPS Ansatz; 2) H₂O and butadiene with (4, 4) active spaces and one layer in the Ansatz and 3) benzene with a (6, 6) active space and one layer in the Ansatz.

The fermionic Hamiltonians and quantum circuits for all molecular systems were obtained with the Python library SlowQuant³¹ using the STO-3G basis set.³² Qubit Hamiltonians were obtained from the fermionic Hamiltonians with the Jordan-Wigner transformation using the Qiskit Nature library.^{33,34} Reference energies and optimal parameters for the tUPS Ansätze were calculated classically with SlowQuant’s ideal state vector simulator. PySCF³⁵ was used for the initial Hartree-Fock calculations. Qiskit and Qiskit Aer were utilized for all device noise simulations, and all quantum experiments were performed on IBM quantum computers through the IBM Quantum Platform, Qiskit, and the Qiskit IBM Runtime library.³³ For the hardware experiments, the `ibm_fez` and `ibm_marrakesh` backends were used. The device noise simulations adopted noise models imported from those same backends and included

transpilation (with a fixed seed and respecting the connectivity map of the backend). Perfect-pairing was used for the input states for all energy calculations.¹⁹

We executed all quantum circuits needed for the construction of any assignment matrix with a shot count that determines the transition probabilities with an accuracy of at least 0.01 with a probability of at least 90%. The shot count can be determined from Hoeffding’s inequality, which provides an upper bound on the probability that a sample mean value, \bar{x} , differs more than δ from the corresponding population mean value, μ ,

$$P(|\bar{x} - \mu| \geq \delta) \leq 2 \exp \left(- \frac{2K^2\delta^2}{\sum_i^K (b_i - a_i)^2} \right) \quad (17)$$

where \bar{x} is the mean of K measurement results x_1, \dots, x_k , and a_i and b_i are lower and upper bounds for the measurement result x_i , i.e. $a_i \leq x_i \leq b_i$. For the transition probabilities, $a_i = 0$ and $b_i = 1$ for all i , and the necessary shot count for $\delta = 0.01$ and an upper bound of 10% is 14,979 shots for each computational basis state. For the butadiene, H₂O and benzene (4,4) and (6,6) systems, we used $64 \times 14,979 = 958,656$ shots in total to construct all tile assignment matrices in Eqs. (14) and (15). That corresponds to approximately 5 minutes of QPU time being spent characterizing the noise (on `ibm_fez` and `ibm_marrakesh`). For the H₂ and LiH (2,2) systems, we used 479,328 shots in total to construct the matrices. Note that those numbers exclude the 30,000 shots used for overhead screening, which takes about 10 seconds of QPU processing time.

For measuring the electronic Hamiltonian, we employed a heuristic minimum clique cover algorithm to group the corresponding Pauli strings of the Hamiltonian based on qubit-wise commutativity (QWC) using the largest-first approach.^{36,37} Again, we used Hoeffding’s inequality (see Eq. (17)) to determine the number of shots that were required for each clique to estimate the energy with a desired accuracy and certainty. This necessitated an estimate of the error mitigation overhead that we obtained through the procedure described in Sec. 2.6. We repeated the energy calculations up to five times for each molecular system to get some

crude insight into the underlying statistics.

The approach of using Hoeffding’s inequality is approximate, but it allows us to allocate more shots to cliques of Pauli strings that have large absolute weights. To briefly illustrate how the weights play in, consider an operator $\hat{O} = w_1\hat{O}_1 + w_2\hat{O}_2$ where \hat{O}_1 and \hat{O}_2 are Pauli strings that commute qubit-wise, and w_1 and w_2 are the corresponding weights. Assume that \hat{O}_1 and \hat{O}_2 are diagonal in the computational basis, and that we want to determine the expectation value of \hat{O} for an arbitrary state — for simplicity, without doing error mitigation. Each measurement result will yield one of the following values for \hat{O} : $w_1 + w_2$, $w_1 - w_2$, $-w_1 + w_2$, or $-w_1 - w_2$. The lower and upper bounds for the measurement results are $-(|w_1| + |w_2|)$ and $|w_1| + |w_2|$, respectively, so the denominator in the exponential in Eq. (17) is $2K(|w_1| + |w_2|)$. In general, the denominator will be $2K \sum_j |w_j|$, where the sum is over all Pauli strings in the clique. The larger the sum, the more shots will be needed to estimate the expectation value with a fixed accuracy and certainty.

The number of shots that we used to estimate the expectation values and to characterize the noise (including the shots needed to construct all tile assignment matrices but excluding the 30,000 shots used for overhead screening) for all molecular systems is shown in Table 1 along with information on the number of Pauli strings in the qubit Hamiltonians, the number of QWC cliques, the L1 norms of the Pauli string weights (excluding the weight of the all-identity Pauli string), the number of times the calculations were repeated, and the QPU processing times. The shot counts that are listed are for a single calculation, and they were the same for quantum experiments and noisy simulations on the same molecule. We also calculated the energies with normal M0 on the simulator. The number of shots that we used to construct the M0 assignment matrix for an n -qubit system was $14,979 \times 2^n$, and those shots are not included in the table. Note that the shot counts that are shown for H_2 and LiH are the total combined shots that were used for all the layer calculations, while the shot counts for the other molecules are for experiments with just one layer. The shot counts for each layer calculation for H_2 and LiH can be found in Table S1 in the supporting

Table 1: Details about the qubit Hamiltonians for the tested molecular systems, the number of shots used for the quantum experiments and noisy simulations (excluding the 30,000 shots used for overhead screening and excluding the shots used for normal M0 on the simulator), the number of times the experiments were repeated, and the QPU processing times.

Molecule	LiH	H ₂	Butadiene	H ₂ O	Benzene
Number of qubits	4	4	8	8	12
Pauli strings	27	15	185	361	555
Cliques	9	5	46	73	132
L1 norm	1.47	1.89	5.06	9.44	9.15
Shots, total*	1,944,302 [†]	2,920,557 [†]	3,426,225	15,749,269	26,162,607
Repeats [‡]	5, 5	5, 5	5, 5	3, 5	3, 5
QPU time, total*	~ 10 min	~ 15 min	~ 20 min	~ 1.3 hrs	~ 2.1 hrs
QPU time, tiled M0	~ 2.5 min	~ 2.5 min	~ 5 min	~ 5 min	~ 5 min

* Per calculation (including both noise characterization and expectation value estimation, but excluding overhead screening).

[†] The shot counts for H₂ and LiH are the total combined shots that were used for all the layer calculations from 1 to 4 layers.

[‡] The number of times the energy calculation was repeated in the quantum experiment and simulation, respectively.

information. The repeated energy calculations were all carried out on separate days for the quantum experiments, and new assignment matrices were constructed each time. New matrices were also constructed for each noisy simulation experiment.

4 Results and discussion

We show the performance of tiled M0 for ground state energy calculations with the tUPS Ansatz on the aforementioned molecular systems with results from both quantum hardware experiments and noisy simulations. We compare our results to exact energies calculated with an ideal state vector simulator. In the noisy simulations, we also compare to the exponentially more expensive M0 technique which allows us to quantify the correlation that we miss due to the tiled locality approximation. The graphs that are presented in the text show average deviations over all repeated quantum experiments and simulations. Results

from the individual calculations are included in the supporting information, together with calculations that have been excluded from the main text based on excessive noise.

4.1 LiH and H₂

The results for LiH and H₂ with (2, 2) active spaces are shown in Fig. 2. The number of CZ gates in the circuits was $39 \times L$, where L is the layer count.

From the LiH quantum experiments, we see that tiled M0 is successful in reducing the energy deviations for all the calculations compared to the unmitigated results. We get a one order of magnitude reduction in the deviation at 1, 2, and 3 layers and slightly less at 4 layers. The best accuracy we achieve with error mitigation is 8.7 mH at 1 layer, and the deviations increase with the number of layers for both the mitigated and unmitigated energies. A comparison of the quantum experiments and simulations for LiH shows that we get more accurate results in the simulations for both the unmitigated and tiled M0 energies. Crucially, tiled M0 achieves impressive one and two orders of magnitude reductions in the energy deviations. The best result is for the 3 layer calculation, where the tiled M0 energy deviates only -0.5 mH from the exact value. Interestingly, the tiled M0 energies on the simulator do not increase consistently with the number of layers the same way they do for the quantum experiments, highlighting the differences between noise models and real hardware.

The tests on H₂ in Fig. 2b show another example of tiled M0 reliably reducing the error for calculations on a 4 qubit system. The number of CZ gates was the same as for LiH. In both the quantum experiments and noisy simulations, we observe approximately the same improvements as for LiH, getting close to chemical accuracy in the simulations. Importantly, tiled M0 performs qualitatively the same as normal M0 in the simulations on LiH and H₂ with only minor quantitative differences. Note that tiled M0 and M0 are the same for a 4 qubit system when the number of layers is 1. The results differ slightly in our tests because separate assignment matrices were constructed for the tiled M0 and regular M0 tests.

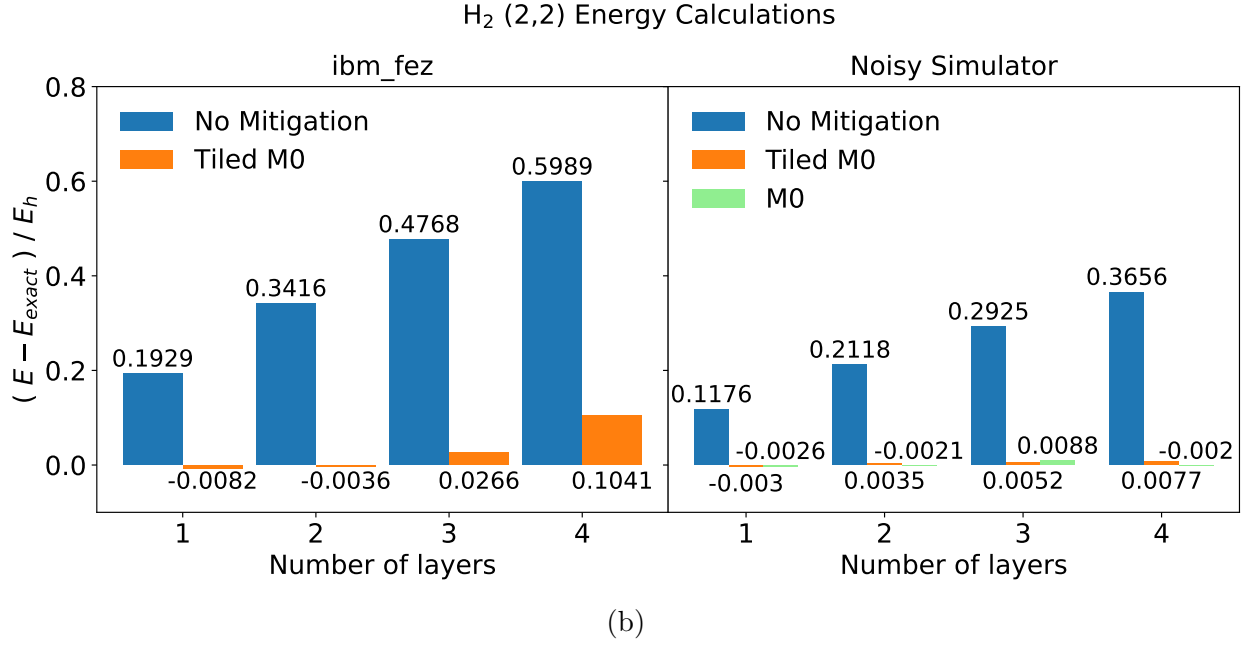
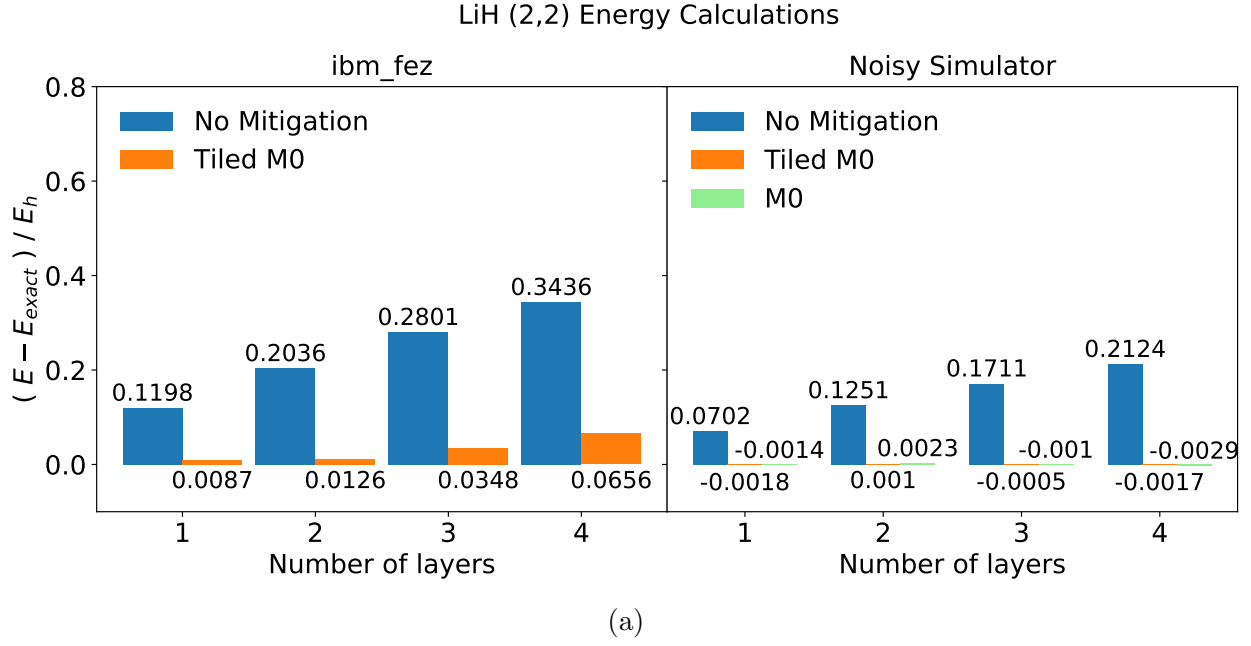


Figure 2: Average results from energy calculations with the tUPS Ansatz on LiH and H₂. We refer to Table 1 for the number of shots used. Each quantum experiment and noisy simulation was repeated 5 times. The noise model used for the simulations was imported from `ibm_fez`. Note that for both LiH and H₂, a single layer is sufficient to reach the ground state energy, but the addition of more layers serves to introduce more noise, allowing for the viability of tiled M0 to be tested.

4.2 Butadiene and H₂O

Fig. 3 shows results from energy calculations on butadiene and H₂O with $(4, 4)$ active spaces and with 151 – 157 CZ gates in the one-layer tUPS Ansatz (the exact number depends on the qubit mapping). Again, tiled M0 is successful in reducing the energy deviations by approximately one order of magnitude, e.g. for butadiene from 454 mH to -28.7 mH and from 262 mH to -18.9 mH in the quantum experiments and noisy simulations, respectively. For the tests on water, better error mitigation is obtained with tiled M0 on the simulator, where we see a reduction from 618 mH to 48.7 mH compared to a reduction from 1028 mH to 263 mH in the quantum experiment. The quantum experiment shows only an improvement of a factor of 4, which we tentatively suggest is due to noise drift and noise fluctuations — a point that we discuss further in Sec. 4.4. The fact that tiled M0 performs very similarly to M0 in the simulations once again justifies the tiled locality approximation.

4.3 Benzene

Finally, we test tiled M0 on a large system for current NISQ device capabilities: benzene with a $(6, 6)$ active space, corresponding to 12 qubits, and with 287 – 299 CZ gates in the one-layer tUPS circuit. The results from energy calculations are shown in Fig. 4. Here, we truly see a large difference between the quantum experiments and noisy simulations. In the quantum experiments, we succeed in bringing down the energy deviation by a factor of 2 with tiled M0 while we obtain a better than one order of magnitude improvement on the simulator with both tiled M0 and M0. We note that the noise levels varied significantly between the three different quantum experiments, as seen in Fig. S10 in the supporting information.

While the hardware results are not as impressive (see Sec. 4.4 for a discussion), the noisy simulation results show the potential capabilities of the technique. We also stress that despite the large system size, the noise characterization takes only 5 minutes on the QPU or about 4% of the total QPU processing time, making it practically free. This is strikingly different from recent quantum experiments using M0^{17,27} where the noise characterization

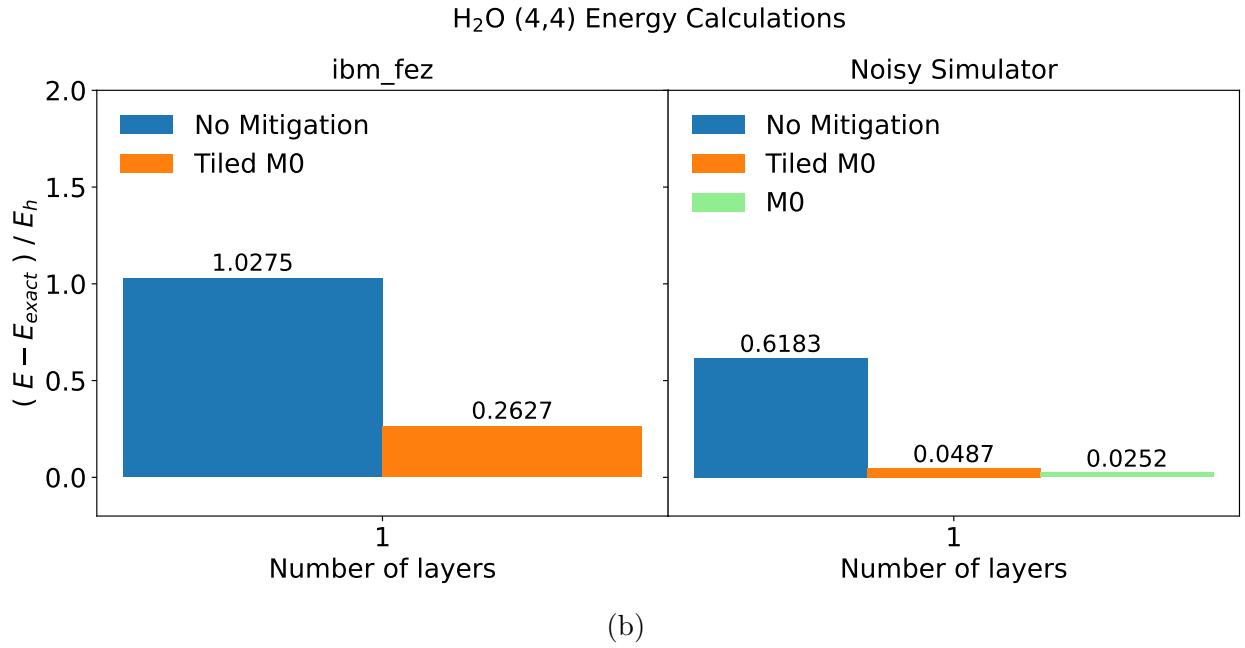
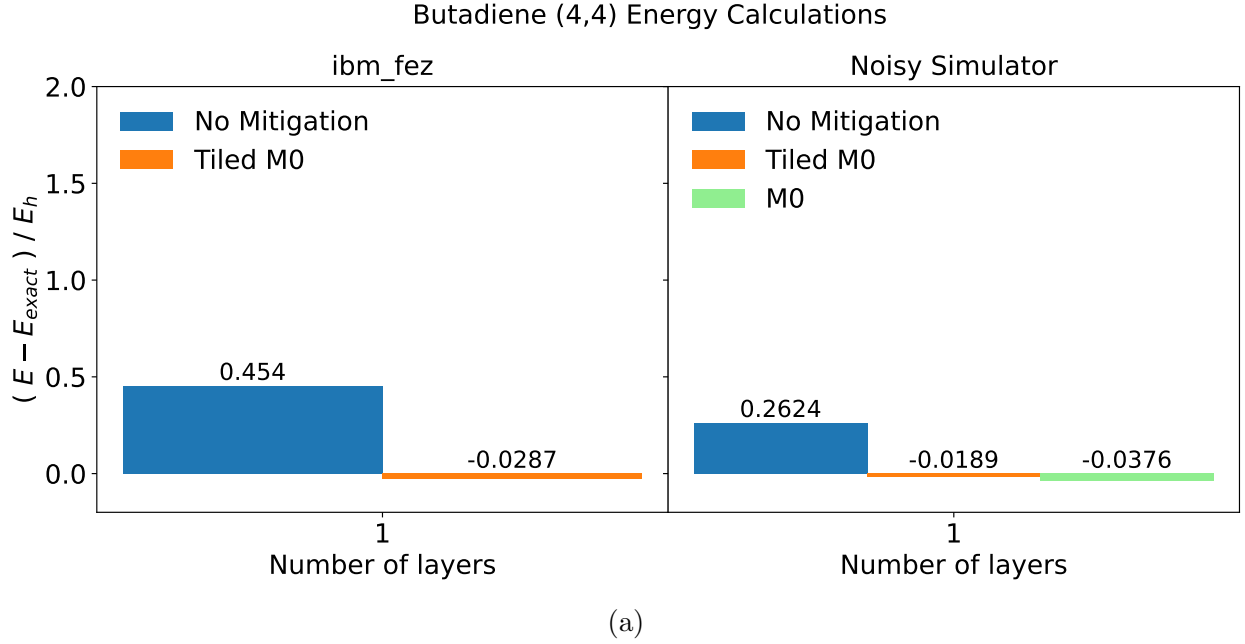


Figure 3: Average results from energy calculations with the tUPS Ansatz on butadiene and H₂O. We refer to Table 1 for the number of shots used. Each quantum experiment and noisy simulation was repeated 5 times for butadiene. For H₂O, the quantum experiments and noisy simulations were repeated 3 and 5 times, respectively. The noise model used for the simulations was imported from `ibm_fez`.

took up a substantial amount of time.

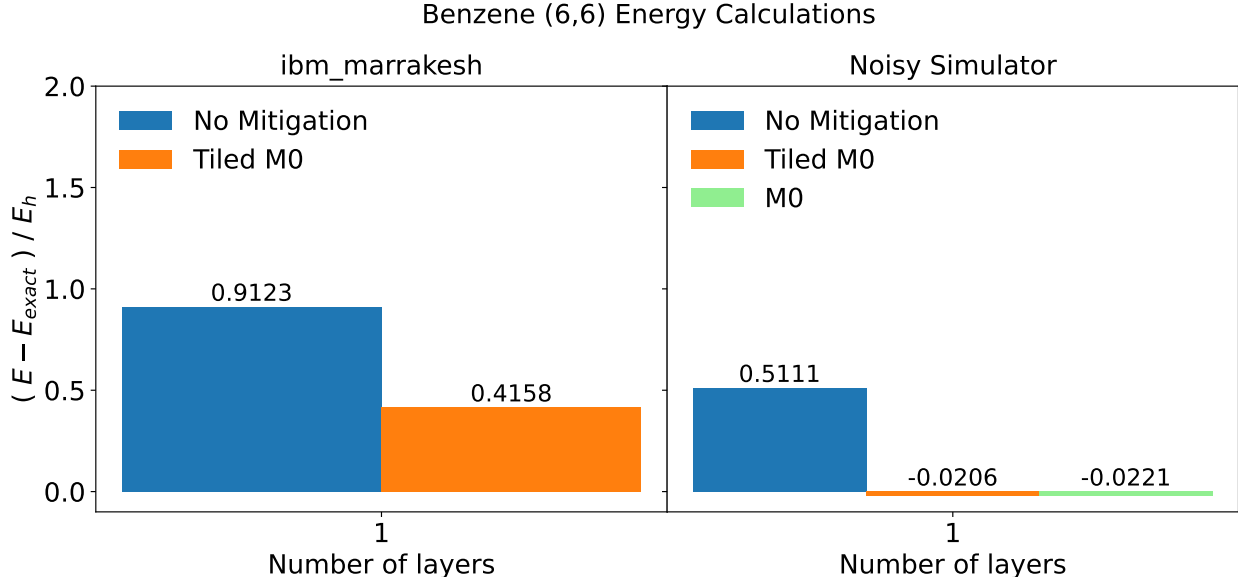


Figure 4: Average results from energy calculations with the tUPS Ansatz on benzene. We refer to Table 1 for the number of shots used. Each quantum experiment and noisy simulation was repeated 3 and 5 times, respectively. The noise model used for the simulations was imported from `ibm_fez`.

4.4 Noise fluctuations

The differences between the quantum experiment and noisy simulation results for H_2O and benzene are significantly larger compared to LiH , H_2 , and butadiene. For both H_2O and benzene, the energy deviation with tiled M0 is one order of magnitude smaller on the noisy simulator compared to the energy deviation in the quantum experiment. Based on the consistent results that we achieve across the repeated quantum experiments and simulations, as seen in the supporting information, the problem is likely not due to undersampling and statistical noise. We suspect that the poor results in the quantum experiments are, in part, due to noise drift and noise fluctuations. The QPU processing times for the H_2O and benzene experiments were approximately 1 hour and 15 minutes and 2 hours and 5 minutes, respectively, for each repeated calculation. The noise was characterized at the start of the experiments, and if the noise drifts as we execute the quantum circuits needed to calculate

the energies, this will introduce an error. Since the calculations on LiH, H₂, and butadiene all took at most 20 minutes of QPU time, the noise profile was unlikely to change as much over the course of the experiments.

To get an idea of how noise drift might have influenced our results, we did two tests that each lasted 1 hour and 10 minutes on `ibm.fez`, where we repeatedly executed the 1 layer tUPS circuit for H₂O with a (4,4) active space. We ran 150 jobs in immediate succession with 100,000 shots each. In Fig. S12 in the supporting information, a plot of a fixed element of the raw probability vector (no error mitigation) from each job result is seen as a function of the time when the job finished executing. A simulation test is also shown for reference. The results from both tests on `ibm.fez` indicate that the noise is moderately stable over the 1 hour and 10 minute durations, but with a period of about 10 minutes for each test where severe fluctuations occur. We tentatively suggest that fluctuations like those, together with small drifts in the noise, are partially responsible for the poor results obtained for H₂O and benzene in the quantum experiments.

5 Summary

We have proposed an error mitigation technique, tiled M0, that characterizes and corrects gate and readout noise for tiled Ansätze. Our technique is similar to readout error mitigation, but it differs in two main ways: 1) parts of the quantum chemical Ansatz are included when constructing the assignment matrices, an idea adopted from the recently proposed M0 technique, and 2) special to tiled M0, a locality approximation based on the structure of tiled Ansätze is introduced to exponentially reduce the scaling in the QPU cost of the noise characterization compared to M0. The cost of the noise characterization is constant and independent of the qubit number and circuit depth for systems of 6 qubits and more, and in this work, we spent a maximum of 5 minutes worth of QPU time characterizing the noise for a given energy calculation.

In quantum experiments and noisy simulations, we have demonstrated the ability of tiled M0 to reliably improve ground state energy calculations with the tUPS Ansatz for molecular systems varying in size between 4 and 12 qubits with up to 299 CZ gates in the Ansatz. An order of magnitude reduction in the energy error is seen in most tests, and we reach chemical accuracy for the two- and three-layer noisy simulations on LiH. In the quantum experiments, tiled M0 performed well for the LiH, H₂, and butadiene systems with an order of magnitude reduction in the energy error, but its impact was limited for H₂O and benzene, only reducing the error by a factor of 2 – 4. Based on preliminary findings, we suspect that this is at least in part due to noise drift and noise fluctuations, which are exacerbated by long QPU processing times. These effects were not present in the corresponding noisy simulations, where tiled M0 performed consistently well and at a level similar to M0.

In summary, the proposed error mitigation technique provides a simple, cost-effective, and scalable way of mitigating errors for tiled Ansätze with a noise characterization cost that is independent of the number of qubits and circuit depth. We are especially encouraged by the noisy simulation results, where orders of magnitude improvements over the unmitigated energies were obtained in all our tests. We expect tiled M0 to find use on improved quantum hardware where the noise is lower and less prone to drift, and in near-term quantum computing applications. In the future, we will further investigate the role of drift and noise fluctuations and look at ways of potentially mitigating their effect.

Acknowledgement

We acknowledge the financial support of the Novo Nordisk Foundation for the focused research project *Hybrid Quantum Chemistry on Hybrid Quantum Computers* (HQC)², grant number NNFSA220080996. We further thank The National Quantum Algorithm Academy (NQAA) under Danish e-infrastructure Consortium (DeiC) for financial support in form of a PhD stipend to OGLR.

References

- (1) Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79.
- (2) Shor, P. W. Fault-tolerant quantum computation. Proceedings of 37th Conference on Foundations of Computer Science. 1996; pp 56–65.
- (3) Aharonov, D.; Ben-Or, M. Fault-tolerant quantum computation with constant error. Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing. 1997; pp 176–188.
- (4) Zimborás, Z.; Koczor, B.; Holmes, Z.; Borrelli, E.-M.; Gilyén, A.; Huang, H.-Y.; Cai, Z.; Acín, A.; Aolita, L.; Banchi, L.; Brandão, F. G. S. L.; Cavalcanti, D.; Cubitt, T.; Filippov, S. N.; García-Pérez, G.; Goold, J.; Kálmán, O.; Kyoseva, E.; Rossi, M. A. C.; Sokolov, B.; Tavernelli, I.; Maniscalco, S. Myths around quantum computation before full fault tolerance: What no-go theorems rule out and what they don't. 2025; <https://arxiv.org/abs/2501.05694>.
- (5) Takagi, R.; Tajima, H.; Gu, M. Universal Sampling Lower Bounds for Quantum Error Mitigation. *Phys. Rev. Lett.* **2023**, *131*, 210602.
- (6) Quek, Y.; Stilck França, D.; Khatri, S.; Meyer, J. J.; Eisert, J. Exponentially tighter bounds on limitations of quantum error mitigation. *Nature Physics* **2024**, *20*, 1648–1658.
- (7) Temme, K.; Bravyi, S.; Gambetta, J. M. Error Mitigation for Short-Depth Quantum Circuits. *Phys. Rev. Lett.* **2017**, *119*, 180509.
- (8) Li, Y.; Benjamin, S. C. Efficient Variational Quantum Simulator Incorporating Active Error Minimization. *Phys. Rev. X* **2017**, *7*, 021050.
- (9) Giurgica-Tiron, T.; Hindy, Y.; LaRose, R.; Mari, A.; Zeng, W. J. Digital zero noise

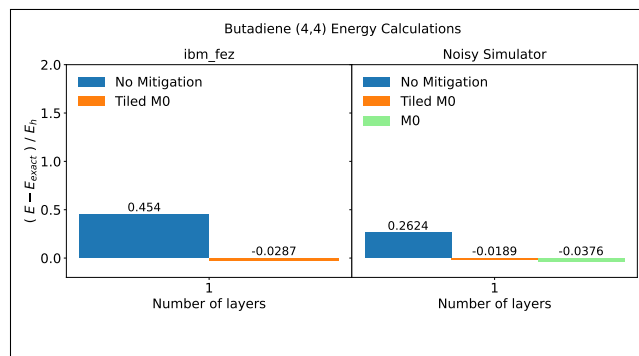
- extrapolation for quantum error mitigation. 2020 IEEE International Conference on Quantum Computing and Engineering (QCE). 2020; pp 306–316.
- (10) Endo, S.; Benjamin, S. C.; Li, Y. Practical Quantum Error Mitigation for Near-Future Applications. *Phys. Rev. X* **2018**, *8*, 031027.
 - (11) Bravyi, S.; Sheldon, S.; Kandala, A.; McKay, D. C.; Gambetta, J. M. Mitigating measurement errors in multiqubit experiments. *Phys. Rev. A* **2021**, *103*, 042605.
 - (12) Czarnik, P.; Arrasmith, A.; Coles, P. J.; Cincio, L. Error mitigation with Clifford quantum-circuit data. *Quantum* **2021**, *5*, 592.
 - (13) Strikis, A.; Qin, D.; Chen, Y.; Benjamin, S. C.; Li, Y. Learning-Based Quantum Error Mitigation. *PRX Quantum* **2021**, *2*, 040330.
 - (14) Czarnik, P.; McKerns, M.; Sornborger, A. T.; Cincio, L. Improving the efficiency of learning-based error mitigation. *Quantum* **2025**, *9*, 1727.
 - (15) Zhao, Z.; Kjellgren, E. R.; Coriani, S.; Kongsted, J.; Sauer, S. P. A.; Ziem, K. M. Quantum error mitigation using energy sampling and extrapolation enhanced Clifford data regression. 2025; <https://arxiv.org/abs/2511.03556>.
 - (16) Filippov, S.; Leahy, M.; Rossi, M. A. C.; García-Pérez, G. Scalable tensor-network error mitigation for near-term quantum computing. 2023; <https://arxiv.org/abs/2307.11740>.
 - (17) Ziem, K. M.; Kjellgren, E. R.; Sauer, S. P. A.; Kongsted, J.; Coriani, S. Understanding and mitigating noise in molecular quantum linear response for spectroscopic properties on quantum computers. *Chem. Sci.* **2025**, *16*, 4456–4468.
 - (18) Cai, Z.; Babbush, R.; Benjamin, S. C.; Endo, S.; Huggins, W. J.; Li, Y.; McClean, J. R.; O’Brien, T. E. Quantum error mitigation. *Rev. Mod. Phys.* **2023**, *95*, 045005.

- (19) Burton, H. G. A. Accurate and gate-efficient quantum Ansätze for electronic states without adaptive optimization. *Phys. Rev. Res.* **2024**, *6*, 023300.
- (20) Anselmetti, G.-L. R.; Wierichs, D.; Gogolin, C.; Parrish, R. M. Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems. *New J. Phys.* **2021**, *23*, 113010.
- (21) Berthussen, N. F.; Trevisan, T. V.; Iadecola, T.; Orth, P. P. Quantum dynamics simulations beyond the coherence time on noisy intermediate-scale quantum hardware by variational Trotter compression. *Phys. Rev. Res.* **2022**, *4*, 023097.
- (22) Sun, R.-Y.; Shirakawa, T.; Yunoki, S. Efficient variational quantum circuit structure for correlated topological phases. *Phys. Rev. B* **2023**, *108*, 075127.
- (23) Tepaske, M. S. J.; Luitz, D. J.; Hahn, D. Optimal compression of constrained quantum time evolution. *Phys. Rev. B* **2024**, *109*, 205134.
- (24) Ayeni, B. M. Efficient particle-conserving symmetric quantum circuits. *Phys. Rev. A* **2025**, *111*, 022629.
- (25) Miháliková, I.; Krejčí, M.; Friák, M. The impact of quantum circuit architecture and hyperparameters on variational quantum algorithms exemplified in the electronic structure of the GaAs crystal. *Sci. Rep.* **2025**, *15*, 15746.
- (26) Uvarov, A. V.; Kardashin, A. S.; Biamonte, J. D. Machine learning phase transitions with a quantum processor. *Phys. Rev. A* **2020**, *102*.
- (27) Jensen, P. W. K.; Hedemark, G. S.; Ziems, K. M.; Kjellgren, E. R.; Reinholdt, P.; Knecht, S.; Coriani, S.; Kongsted, J.; Sauer, S. P. A. Hyperfine coupling constants on quantum computers: Performance, errors, and future prospects. *J. Chem. Theory Comput.* **2025**, *21*, 7878–7889.

- (28) Reinholdt, P.; Kjellgren, E.; Ziemis, K. M.; Coriani, S.; Sauer, S. P. A.; Kongsted, J. Self-consistent Quantum Linear Response with a Polarizable Embedding Environment. *J. Phys. Chem. A* **2025**, *129*, 1504–1515.
- (29) Evangelista, F. A.; Chan, G. K.; Scuseria, G. E. Exact parameterization of fermionic wave functions via unitary coupled cluster theory. *J. Chem. Phys.* **2019**, *151*, 244112.
- (30) Burton, H. G.; Marti-Dafcik, D.; Tew, D. P.; Wales, D. J. Exact electronic states with shallow quantum circuits from global optimisation. *npj Quantum Information* **2023**, *9*, 75.
- (31) Kjellgren, E.; Ziemis, K. SlowQuant. *GitHub*. <https://github.com/erikkjellgren/SlowQuant/tree/master> **2025**,
- (32) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent molecular-orbital methods. I. use of Gaussian expansions of Slater-type atomic orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
- (33) Javadi-Abhari, A.; Treinish, M.; Krsulich, K.; Wood, C. J.; Lishman, J.; Gacon, J.; Martiel, S.; Nation, P. D.; Bishop, L. S.; Cross, A. W.; Johnson, B. R.; Gambetta, J. M. Quantum computing with Qiskit. 2024; <https://arxiv.org/abs/2405.08810>.
- (34) The Qiskit Nature developers and contributors. Qiskit Nature 0.6.0. 2023; <https://doi.org/10.5281/zenodo.7828768>.
- (35) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1340.
- (36) Verteletskyi, V.; Yen, T.-C.; Izmaylov, A. F. Measurement optimization in the varia-

- tional quantum eigensolver using a minimum clique cover. *J. Chem. Phys.* **2020**, *152*, 124114.
- (37) Gokhale, P.; Angiuli, O.; Ding, Y.; Gui, K.; Tomesh, T.; Suchara, M.; Martonosi, M.; Chong, F. T. $O(N^3)$ measurement cost for variational quantum eigensolver on molecular Hamiltonians. *IEEE Trans. Quantum Eng.* **2020**, *1*, 1–24.

TOC Graphic



Supporting Information:

Cost-effective scalable quantum error mitigation for tiled Ansätze

Oskar Graulund Lentz Rasmussen,^{*,†} Erik Kjellgren,[†] Peter Reinholdt,[†] Stephan
P. A. Sauer,[‡] Sonia Coriani,[¶] Karl Michael Ziems,[§] and Jacob Kongsted[†]

[†]*Department of Physics, Chemistry and Pharmacy, University of Southern Denmark,
Campusvej 55, DK-5230 Odense M, Denmark*

[‡]*Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100
Copenhagen Ø, Denmark*

[¶]*Department of Chemistry, Technical University of Denmark, Kemitorvet Building 207,
DK-2800 Kongens Lyngby, Denmark*

[§]*School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK*

E-mail: oskr@sdu.dk

1 Quantum circuit for a tile in the tUPS Ansatz

Fig. S1 shows how a single tile in the tUPS Ansatz might look in terms of common single and two qubit quantum gates. The variational parameters are θ_1, θ_2 and θ_3 in the illustration.

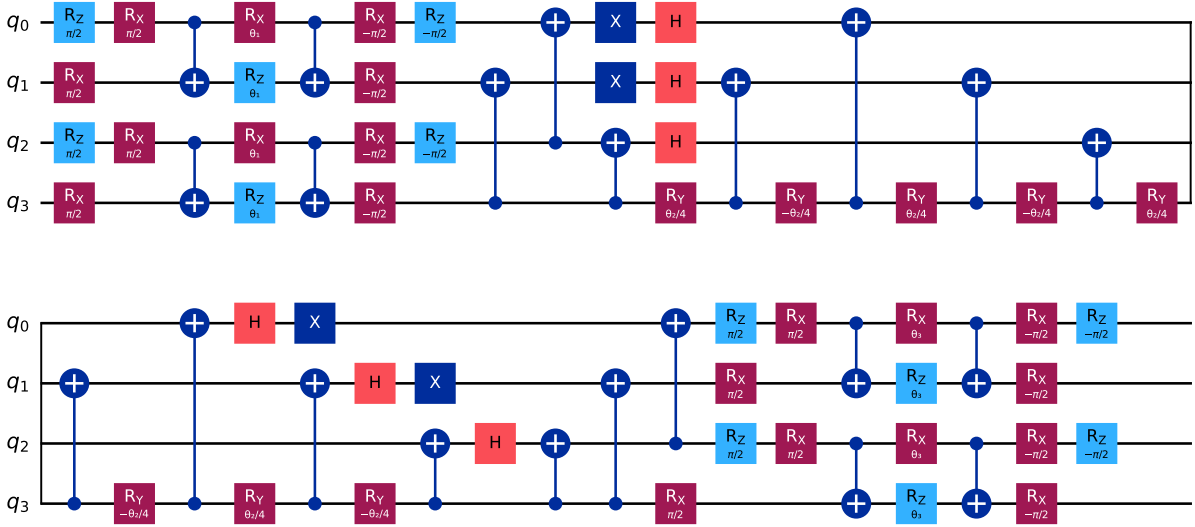


Figure S1: The quantum circuit for a tile in the tUPS Ansatz with variational parameters θ_1, θ_2 and θ_3 .

Table S1: The number of shots used for the tUPS energy calculations on LiH and H_2 at different numbers of layers. The shot counts listed here do not include the shots that were used for noise characterization, but only those used for expectation value estimations.

Molecule	LiH	H_2
One layer	117,696	196,179
Two layers	221,964	369,856
Three layers	366,452	610,631
Four layers	758,861	1,264,562

2 Energy Calculations

The graphs in this section show results from the individual energy calculation experiments that we performed. The graphs in the main text show averages over all experiments. The shot counts that were used are given in Table 1 in the main text. The shot counts that were used for the individual layer calculations for LiH and H_2 are given in Table S1. In Table S2, we show the exact (noiseless) electronic energies that we compare the noisy and error mitigated energies with.

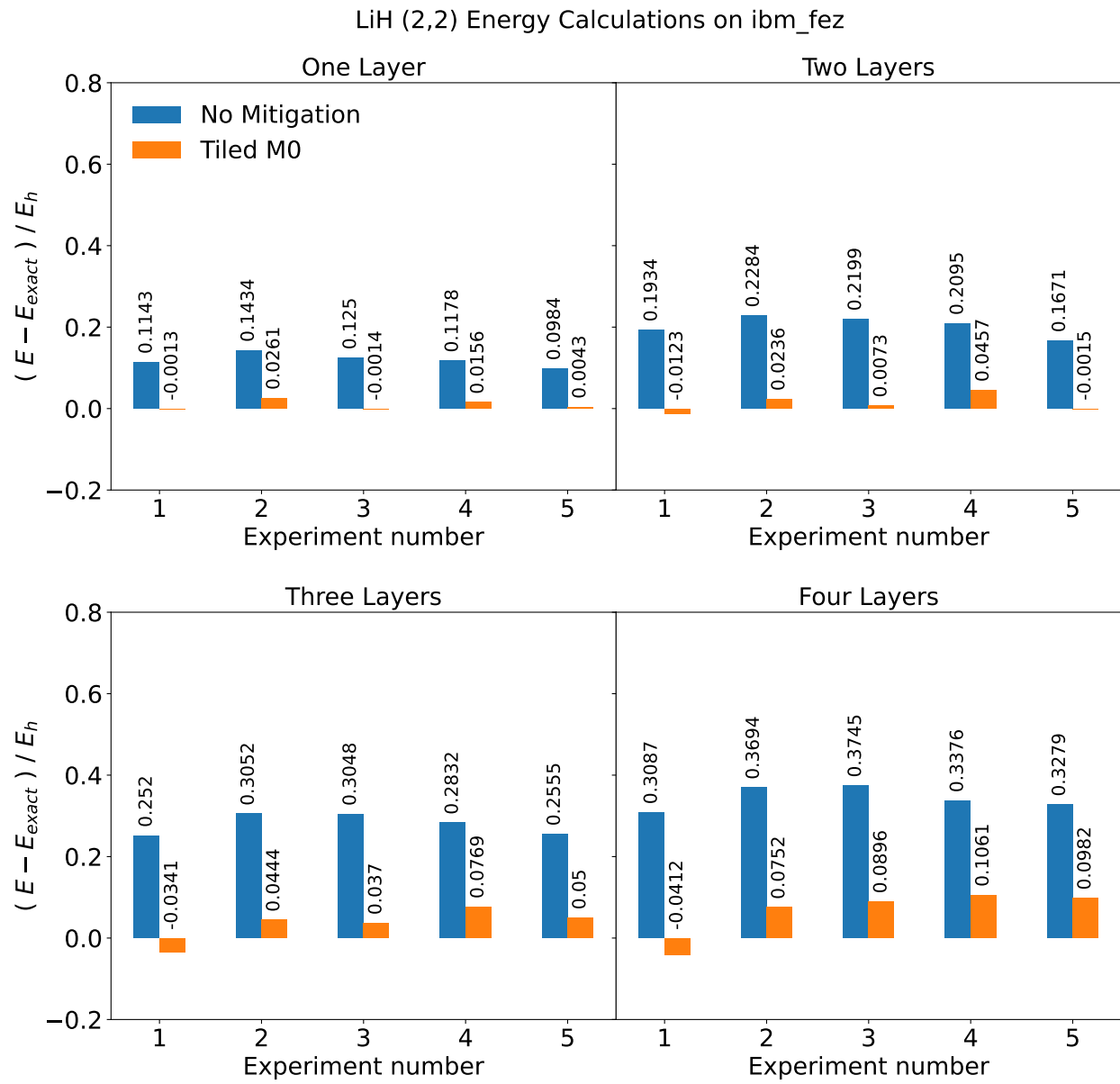


Figure S2: Results from LiH (2,2) energy calculation experiments on `ibm.fez` at different numbers of layers in the tUPS Ansatz.

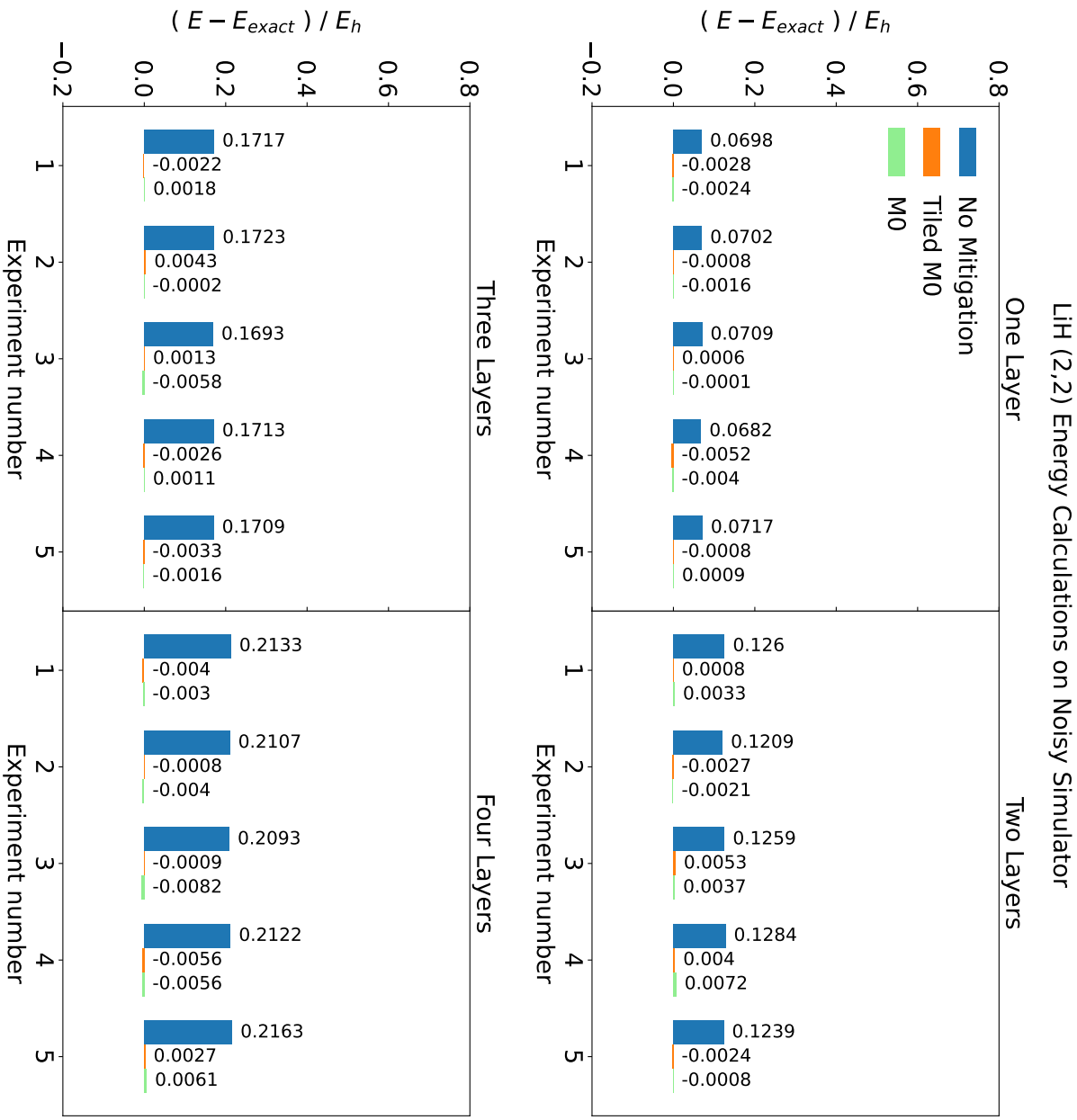


Figure S3: Results from LiH (2,2) energy calculation experiments on a noisy simulator at different numbers of layers in the tUPS Ansatz. The noise model was imported from the `ibm_fez` backend.

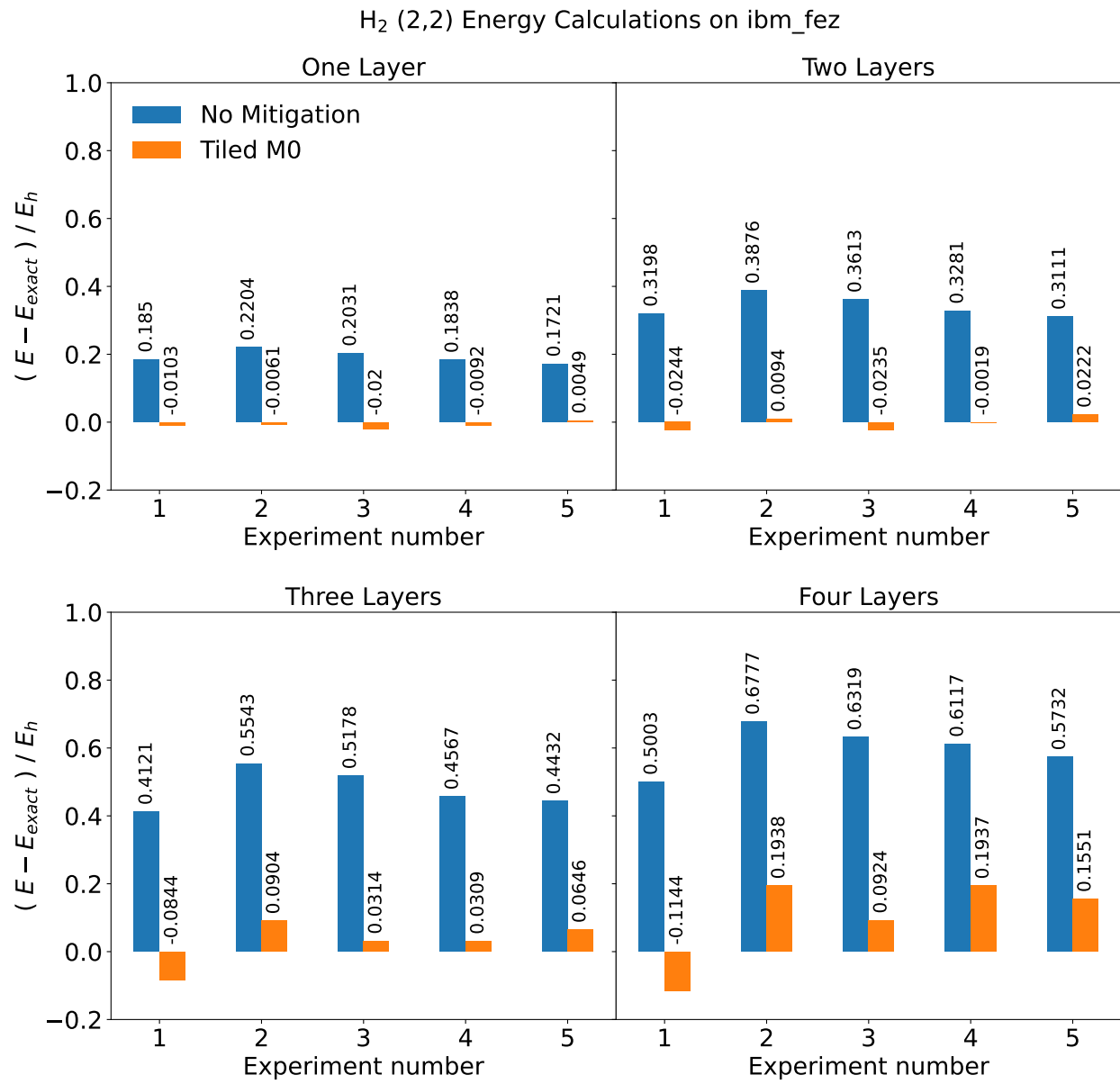


Figure S4: Results from H₂ (2,2) energy calculation experiments on `ibm_fez` at different numbers of layers in the tUPS Ansatz.

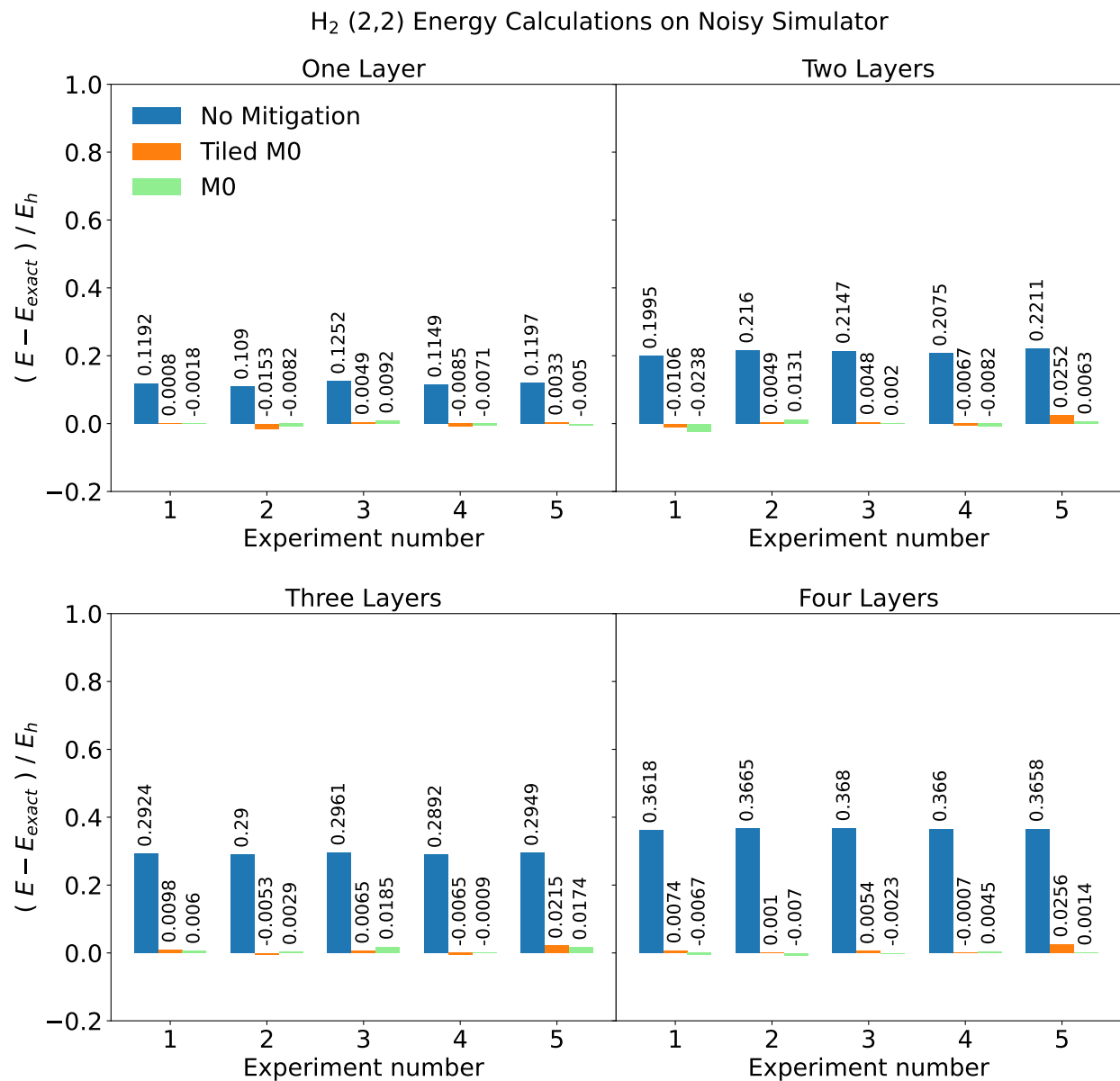


Figure S5: Results from H_2 (2,2) energy calculation experiments on a noisy simulator at different numbers of layers in the tUPS Ansatz. The noise model was imported from the `ibm_fez` backend.

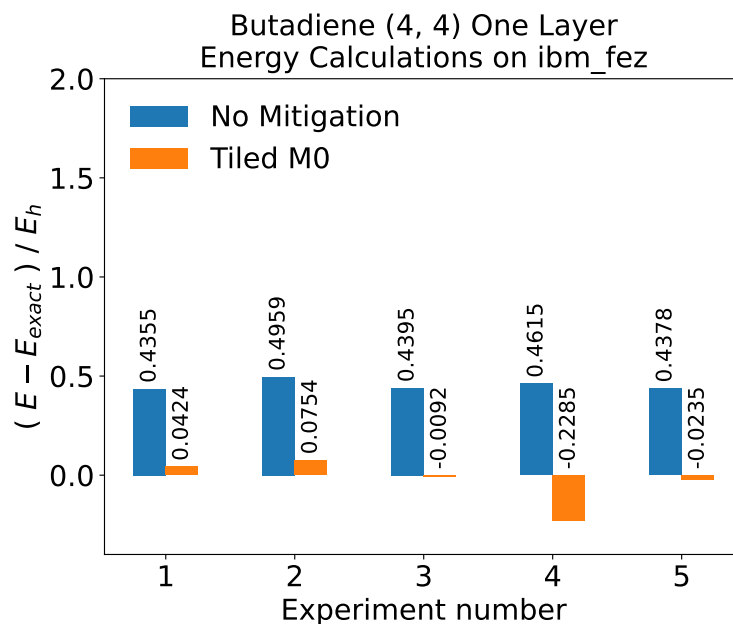


Figure S6: Results from butadiene (4,4) energy calculation experiments on `ibm_fez` at one layer in the tUPS Ansatz.

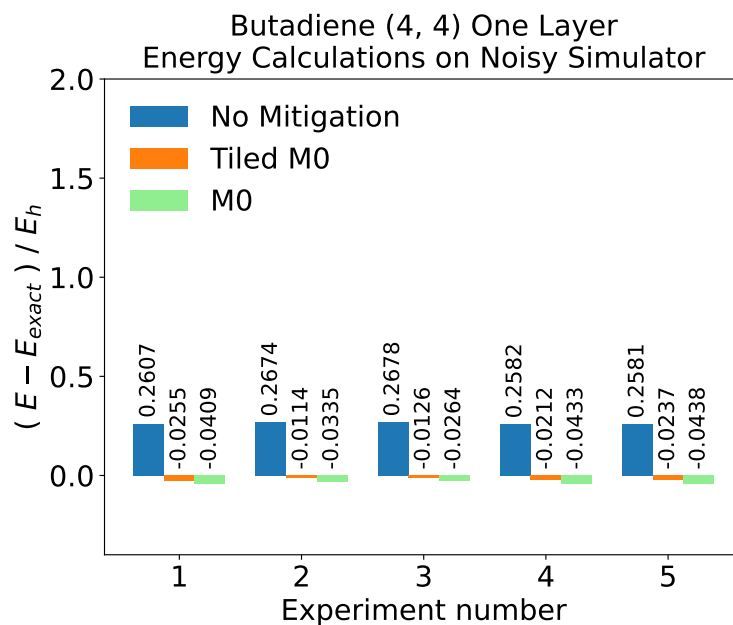


Figure S7: Results from butadiene (4,4) energy calculation experiments on a noisy simulator at one layer in the tUPS Ansatz. The noise model was imported from the `ibm_fez` backend.

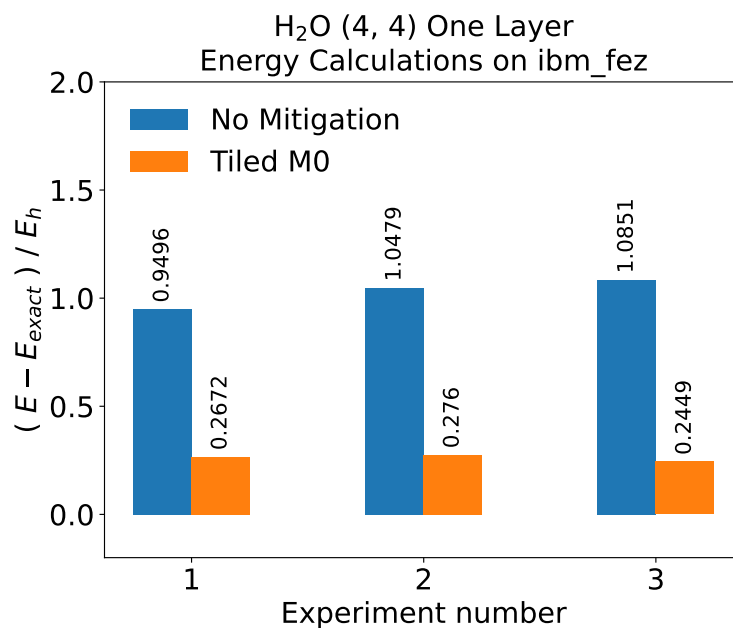


Figure S8: Results from H₂O (4,4) energy calculation experiments on `ibm_fez` at one layer in the tUPS Ansatz.

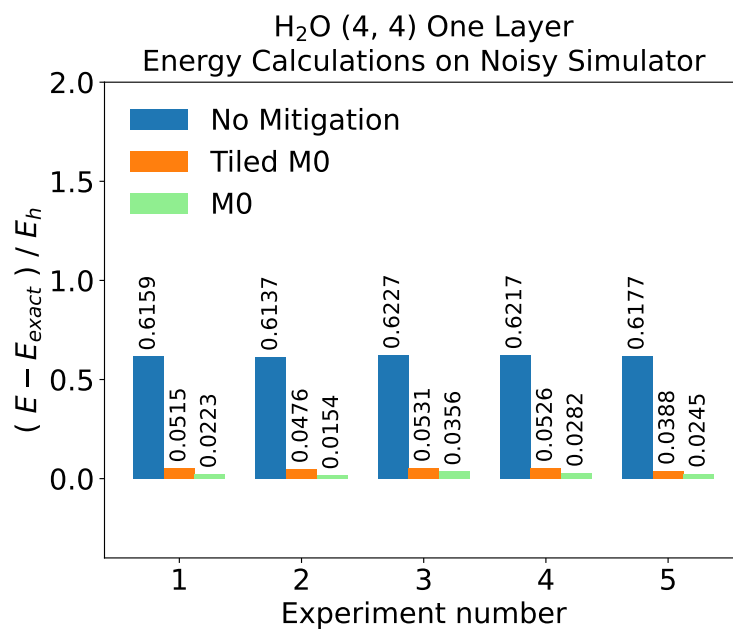


Figure S9: Results from H₂O (4,4) energy calculation experiments on a noisy simulator at one layer in the tUPS Ansatz. The noise model was imported from the `ibm_fez` backend.

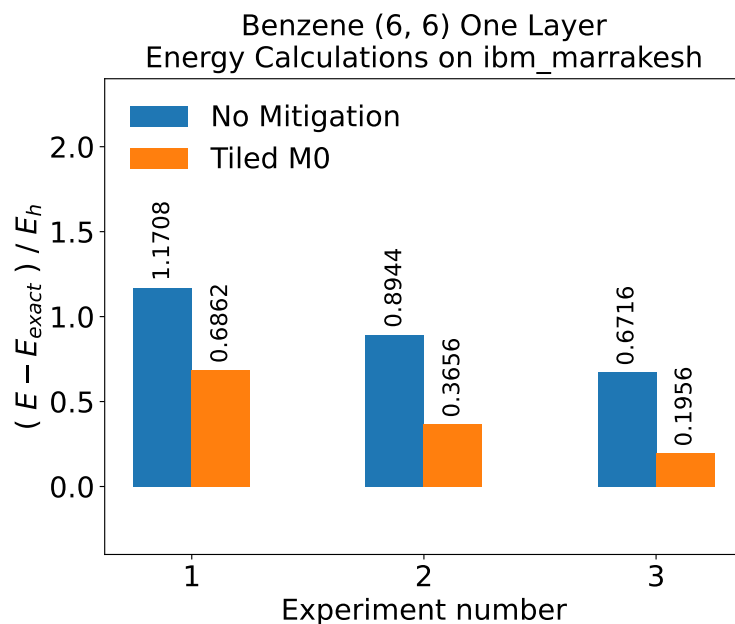


Figure S10: Results from benzene (6,6) energy calculation experiments on `ibm_marrakesh` at one layer in the tUPS Ansatz.

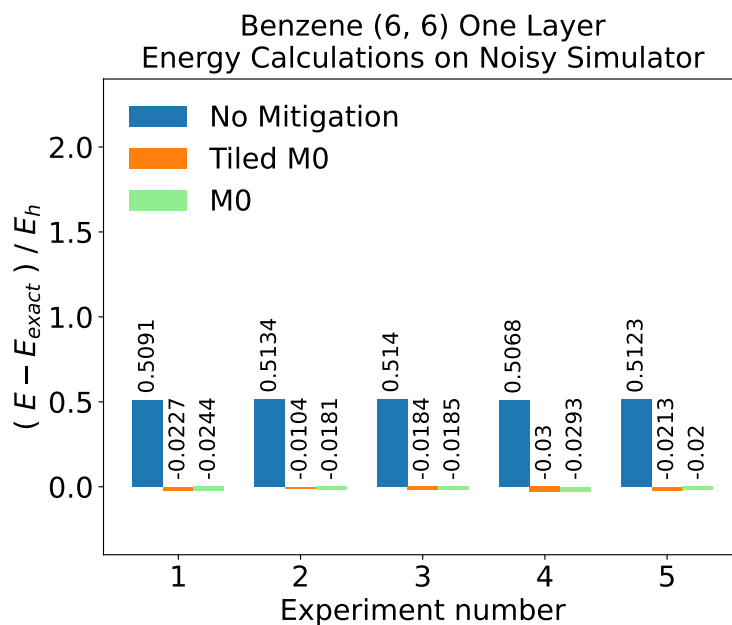


Figure S11: Results from benzene (6,6) energy calculation experiments on a noisy simulator at one layer in the tUPS Ansatz. The noise model was imported from the `ibm_fez` backend.

Table S2: The exact (noiseless) electronic energies for the molecular systems that we compare the noisy and error mitigated energies with. By exact we mean with respect to the given level of theory, the number of layers in the tUPS Ansatz (one), the chosen initial state (perfect-pairing), and the specific variational parameters that we obtained from the classical optimization routine. Note that the energies for LiH and H₂ were the same for all numbers of layers. The energies were calculated with an ideal state vector simulator.

Molecule	LiH	H ₂	Butadiene	H ₂ O	Benzene
E_{exact} / E_h	-8.87325	-1.85239	-257.17168	-84.00123	-432.43028

3 Noise drift and noise fluctuations

Fig. S12 shows how a fixed element of the raw probability vector (no error mitigation) resulting from executions of the one-layer tUPS circuit for H₂O (4,4) can change over time on a quantum computer and noisy simulator. The quantum computer tests were done on IBM’s `ibm_fez` backend by executing 150 tUPS circuits in immediate succession as separate jobs with 100,000 shots each. The same simulation test with a noise model imported from `ibm_fez` is shown for reference.

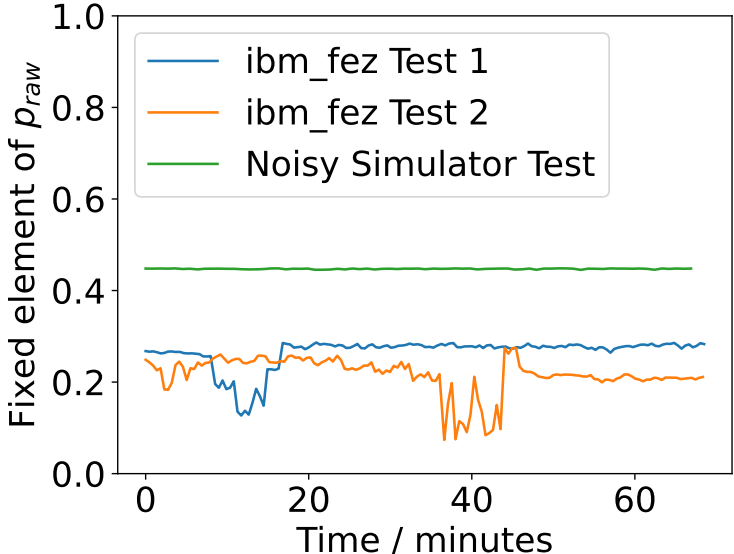


Figure S12: Plots of a fixed element of the probability vector resulting from repeated executions of the one-layer tUPS circuit (with optimal, non-zero parameters) for H₂O (4,4) as a function of time. The time is relative to when the corresponding test began. Each probability vector was obtained with 100,000 shots. Both hardware tests were done on `ibm_fez`. The exact (noiseless) value of the fixed element is approximately 0.75.

Table S3: Results from quantum experiments on `ibm_fez` where the noise was too severe for the tiled M0 technique to succeed. The energies were all calculated with one layer in the tUPS Ansatz.

$(E - E_{exact}) / E_h$ with:	Tiled M0	No Mitigation
Butadiene experiment	250.65	0.62004
Butadiene experiment	-3.18234	0.66804
H ₂ experiment	0.282205	0.564870

4 Tiled M0 under excessive noise

The levels of hardware noise on a given backend can vary with time. If the noise is excessive, our proposed error mitigation technique will fail. In Table S3, we show results from calculations that we did on `ibm_fez` where the noise was too large for our technique to succeed. This is from before we started doing overhead screening. The setups were the exact same in terms of shot counts etc. as for all the other experiments. The energies shown are from calculations with one layer in the tUPS Ansatz. Note that the energies obtained with no error mitigation are also worse compared to the energies from the experiments that we include in the main text, highlighting the stronger levels of noise that were present during the calculations.