

Algorithmic Error Mitigation Scheme for Current Quantum Processors

Philippe Suchsland^{1,2,3}, Francesco Tacchino², Mark H. Fischer³, Titus Neupert³, Panagiotis Kl. Barkoutsos², and Ivano Tavernelli²

¹Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland

²IBM Quantum, IBM Research – Zurich, 8803 Rueschlikon, Switzerland

³Department of Physics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

We present a hardware agnostic error mitigation algorithm for near term quantum processors inspired by the classical Lanczos method. This technique can reduce the impact of different sources of noise at the sole cost of an increase in the number of measurements to be performed on the target quantum circuit, without additional experimental overhead. We demonstrate through numerical simulations and experiments on IBM Quantum hardware that the proposed scheme significantly increases the accuracy of cost functions evaluations within the framework of variational quantum algorithms, thus leading to improved ground state calculations for quantum chemistry and physics problems beyond state-of-the-art results.

Introduction. One of the main limitations of available quantum computers is the sensitivity to noise. Applications like quantum chemistry, machine learning and finance require significant precision in the calculation of observables or classical cost functions. However, current hardware noise levels only allow for limited accuracy, thus hindering the achievement of meaningful outputs. While quantum error correction techniques [1–3] could in principle offer a solution, all the proposed schemes demand technological advancements which are still beyond state-of-the-art capabilities. Within a near-term perspective [4], the available algorithmic implementations are therefore limited to shallow circuits applied on easy-to-prepare initial states [5–10]. In most cases, only proof-of-principle experiments have been performed [11–16], whose performances are not easily compared to classical coun-

terparts.

Error mitigation techniques [17] can in principle extend the range of operability of current technology without the need for a fully error corrected quantum device. Recently proposed solutions are based on extrapolation to zero noise [14, 18–22], quasi-probability distribution methods [20, 21], symmetry enforcement [8, 22–25] or stabilizer-like methods [26], and significant improvements have already been demonstrated experimentally [14, 25, 27, 28]. Dedicated protocols have also been developed for mitigating measurement errors [14, 29–33], mostly based on the assumption that the readout noise can be simulated via a classical model where noise acts independently on each qubit.

In this paper, we introduce a technique for quantum error mitigation inspired by the Lanczos algorithm for matrix diagonalization [34, 35]. Our method provides a rigorous way to improve ground state estimates by enhancing the projection of any given solution to the true ground state of the problem Hamiltonian, therefore representing a direct extension in the context of quantum algorithms of the well known classical Krylov space methods. These could for example be employed to increase the quality of the variational ansatz whenever the latter is limited, e.g., by the shallowness of the corresponding quantum circuit or by hardware connectivity constraints [36]. In addition to that, here we go beyond a simple adaptation of classical protocols and we demonstrate a promising parallel application as a systematic quantum error mitigation procedure. The scheme remains fully compatible with the underlying variational principle and can be combined with other error mitigation techniques. Most importantly, our method does not require additional experimental resources (e.g.,

Ivano Tavernelli: ita@zurich.ibm.com

calibration of dedicated pulses) besides an increase in the number of observables to be measured on the quantum register. We demonstrate a practical implementation of our technique on a number of quantum physics and quantum chemistry models using state-of-the-art quantum hardware. The general applicability of the scheme is not restricted to such problems, but can be used for any quantum computing application that requires the calculation of ground state properties. The results indicate that the proposed error mitigation scheme can systematically provide better accuracy for energy estimates without the need for a detailed description of the underlying noise model and without being subject to specific hardware constraints.

Lanczos Algorithm. Variational quantum algorithms are typically formulated in terms of a Hamiltonian H , whose smallest eigenvalue, and possibly ground eigenstate, represent the desired solution. In the variational approach, a quantum circuit with a set of tunable gate parameters $\vec{\theta}$ is used to generate an optimal approximation $|\psi(\vec{\theta}_{\text{opt}})\rangle$ of the target quantum state in the variational subspace, using the expectation value of the Hamiltonian $\langle H \rangle_{\vec{\theta}} = \langle \psi(\vec{\theta}) | H | \psi(\vec{\theta}) \rangle$ as a cost function. Under the effect of gate, thermalization, and readout noise, represented here as a quantum operation \mathcal{E} (see also Appendix A and B and Ref. [37]), the variational estimate of the target ground state energy can be described in terms of a density matrix $\rho = \mathcal{E}(|\psi(\vec{\theta}_{\text{opt}})\rangle)$ as $E = \text{Tr}[\rho H]$.

Following the classical Lanczos algorithm [34], both the ideal $\langle H \rangle_{\vec{\theta}}$ and the noisy energy estimates can be improved by constructing a basis of the order- m Krylov subspace $\mathcal{K}^{(m)}$ through, e.g., a power iteration scheme and then diagonalizing H in $\mathcal{K}^{(m)}$ [34]. As an example, given an arbitrary pure variational state $|\psi(\vec{\theta})\rangle$, any element in $\mathcal{K}^{(2)}$ can be parametrized as $|\Psi_L(a_0, a_1, \vec{\theta})\rangle = (a_0 - a_1 H) |\Psi(\vec{\theta})\rangle / \sqrt{N(a_0, a_1)}$ where $N(a_0, a_1) = \langle \Psi(\vec{\theta}) | (a_0 - a_1 H)^2 | \Psi(\vec{\theta}) \rangle$. An improved upper bound for the true ground state energy E_0 can then be found by minimizing the expectation value of the Hamiltonian over the free parameters a_0 and a_1 . Similarly, for any density matrix ρ representing a noisy variational state, the improved energy estimate E_L can be expressed as (see Appendix C for the general case)

$$E_L = \min_{a_0, a_1 \in \mathbb{R}} \quad (1)$$

Notice that this result coincides with the application of a single Euler step of size $\tau = a_1/a_0$ in imaginary time evolution [38–40]. The estimate in Eq. (1) can be shown to be reliable and at least as good as the initial estimate E even in the presence of noise, as it still obeys the variational principle $E_0 \leq E_L$ while additionally satisfying $E_L \leq E$. Indeed, by defining, $\forall a_0, a_1 \in \mathbb{R}$,

$$\rho_L(a_0, a_1) = \frac{(a_0 - a_1 H) \rho (a_0 - a_1 H)}{\text{Tr}[\rho (a_0 - a_1 H)^2]}, \quad (2)$$

we can rewrite Eq. (1) as $E_L = \min_{a_0, a_1} \text{Tr}[\rho_L(a_0, a_1) H]$. Since $\rho_L(a_0, a_1)$ represents a valid density matrix, i.e., hermiticity, semi-positive-definiteness and normalization are preserved under the application of Eq. (2), the variational bound $E_0 \leq E_L$ must also hold. Moreover, the relation $E_L \leq E$ follows from the observation that the expression minimized in Eq. (1) evaluates to E for $a_0 = 1, a_1 = 0$. With a similar proof, it can be shown that the same ordering $E_0 \leq E_L \leq E$ is respected also for higher order ($m > 2$) versions of the method.

If $\bar{\rho}_L$ is the density matrix achieving the minimum in Eq. (1), an improved estimate for any ground state observable O can be obtained as $\langle O \rangle_L = \text{Tr}[\bar{\rho}_L O]$. The accuracy of such estimate with respect to the true value is expected to grow, compared to the initial noisy result, as the overlap of ρ_L with the true ground state $|\Psi_0\rangle$ will generally be larger than the one of the original ρ . A sufficient condition for this to hold, if \bar{a}_0, \bar{a}_1 are the optimal parameters from Eq. (1), is that

$$(\bar{a}_0 - \bar{a}_1 E_L)^2 / \text{Tr}[\rho(\bar{a}_0 - \bar{a}_1 H)^2] > 1. \quad (3)$$

If $\bar{a}_1 \neq 0$, also $\bar{a}_0/\bar{a}_1 > E_L$ is required.

Intuitively, the application of the Lanczos method modifies the original spectral weights $\rho_{ii} = \langle i | \rho | i \rangle$ on the energy eigenstates $\{|i\rangle\}$, leading to the replacement

$$\rho_{ii} \rightarrow \bar{\rho}_{L,ii} = \frac{\rho_{ii}(\bar{a}_0 - \bar{a}_1 E_i)^2}{\text{Tr}[\rho(\bar{a}_0 - \bar{a}_1 H)^2]}, \quad (4)$$

where E_i is the eigenvalue of H corresponding to eigenvector $|i\rangle$. The minimum condition of Eq. (1) then corresponds, similarly to the parameter shift rule in usual power iteration methods for matrix diagonalization, to a choice of \bar{a}_0, \bar{a}_1 reducing the spectral weight of excited states.

The number of independent operators that need to be measured at order m scales in the

worst case as M^{2m-1} , where M is the number of Pauli terms in the original Hamiltonian. Indeed, expectation values of the form $\langle H^n \rangle$ for $n = 0, \dots, m+1$ appear e.g. in Eq. (1) and its generalization to $m > 2$. However, through grouping of Pauli operators into tensor-product basis sets [12] or by using partial state tomography and state reconstruction techniques [41–43] the actual overhead in the number of measurements required by the scheme can be significantly mitigated. With extensive numerical simulations, we also find that for many relevant quantum chemistry problems encoded on n_q qubits the number of Pauli strings in H^n remains below the theoretical predictions ($\mathcal{O}(M^3)$, i.e., $\mathcal{O}(n_q^{12})$, for $n = 3$) at least up to about $n_q = 20$, see Appendix D for details. Moreover, simple analytical arguments predict a very favorable scaling for local Hamiltonians, like for instance the ones describing spin and lattice models. In the following, we will show practical examples for which even low-order ($m = 2$) results can provide estimates very close to noiseless results.

Implementation of order $m = 2$ error mitigation. For the implementation of the method to lowest order, E_L is expressed in terms of the three expectation values $\langle H^k \rangle = \text{Tr}[\rho H^k]$ with $k = 1, 2, 3$, which can all be efficiently measured in practice on a quantum processor. However, each of such estimates usually comes with an associated standard error σ_{H^k} arising from fluctuations of the readout due to the inherent quantum statistics of measurements as well as to hardware noise. As a matter of fact, when such statistical errors are propagated to the final outcome E_L they can give rise to uncertainties σ_{E_L} which are larger than the standard errors affecting the original noisy estimate $E = \text{Tr}[\rho H]$. The precision is particularly influenced by variations in time of the experimental noise levels and calibration accuracy, and can become critical at ill-conditioned values for which the minimization in Eq. (1) yields a small denominator. Besides increasing the precision of the original $\langle H^k \rangle$ estimates e.g. via an increase in the size of the statistical samples within the same experiment, several strategies can be adopted to enhance the stability of the proposed method. A direct improvement can for example be obtained by repeating the estimation of E_L a number n_{repeat} of times, thus obtaining a set of values $\{e_{L,i}\}$ and of

associated errors $\{\sigma_i\}$ which are then combined in a weighted least square (wls) average $\overline{E}_{L,\text{wls}}$ (see also Appendix E). Alternatives are obtained by replacing Eq. (1) with a cube root estimate $E_{L,\text{cube}} = \sqrt[3]{\langle H^3 \rangle}$, which - under some assumptions - can also enhance the spectral weight corresponding to the exact ground state (see Appendix F), or by a direct empirical choice of the parameters a_0 and a_1 in a region with low standard error yielding E_{L,a_0,a_1} (see Appendix F and G for details). It is also worth mentioning that, in principle, Eq. (1) may become ill-conditioned whenever $|\Psi(\vec{\theta})\rangle$ is already very close to a pure eigenstate of H . However, in all our $m = 2$ experiments we always find that $H|\Psi\rangle$ and $|\Psi\rangle$ can be assumed to be linearly independent. In the more general case, simple numerical schemes can be introduced to detect singularities in the Krylov subspace.

Applications. In all the examples presented below, we successfully use an order $m = 2$ method to correct for hardware noise and to enhance the quality of the ansätze generated by the Variational Quantum Eigensolver (VQE) algorithm [12, 44]. We use the so called $R_y R_z$ variational form (see Appendix H) [12] and optimize the rotation angles $\vec{\theta}$ via the simultaneous perturbation stochastic approximation method (SPSA) [12, 45]. To increase numerical stability and in view of the often rough optimization landscapes under study, in all cases we sample $n_{\text{init}} = 5$ parameter sets $\vec{\theta}$ from uniform distributions on $[0, 2\pi]$ and use the one with the lowest energy expectation value as initial set for the optimization routine with a maximum of n_{steps} SPSA iterations. The optimization procedure is repeated a number n_{VQE} of times, choosing the $\{\vec{\theta}_{\text{opt}}\}$ with the lowest energy as the optimal VQE solution. Error mitigation is then applied to the optimal circuit by taking measurements of the operators H , H^2 and H^3 and employing the methods introduced above. In particular, we obtain the direct solution of Eq. (1) by analytic diagonalization of a 2×2 representation of the Hamiltonian constructed from estimates of $\langle H^k \rangle$, see Appendix C. The outcomes are compared to the original noisy estimate of the ground state energy E .

In the first example, we study the dissociation profile of the Hydrogen molecule H_2 [11, 12] as a function of the internuclear distance d .

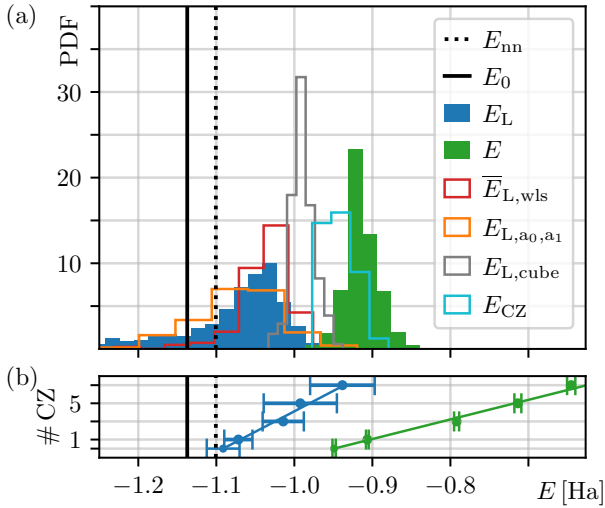


Figure 1: Error mitigation for H_2 encoded on 4 qubits at equilibrium distance with $n_{\text{steps}} = 100$, $n_{\text{VQE}} = 10$, $n_{\text{shots}} = 8192$ run on the *ibmq_ourense* quantum processor. The ground state is approximated with the hardware efficient $R_y R_z$ variational form with one entangling layer. (a) Probability density functions (PDF) for the bare evaluation of $\langle H \rangle_{\vec{\theta}_{\text{opt}}}$ on real hardware (green) and the corresponding mitigated results using Eq. (1) (blue), the cube root method (gray), the wls scheme (red) and the empirical selection of the a_0 and a_1 (orange). E_0 is the exact ground state energy, E_{nn} denotes a numerical evaluation of $\langle H \rangle_{\vec{\theta}_{\text{opt}}}$ using the optimal VQE ansatz, whose parameters were obtained on the real quantum processor. E_{CZ} represents the value of the zero-noise extrapolation carried out on bare energy evaluations with a number of measurements comparable to the ones required for the Lanczos method. (b) Richardson extrapolation towards zero-noise energy estimations. Evaluations of $\langle H \rangle_{\vec{\theta}_{\text{opt}}}$ are obtained by replacing every physical controlled phase (CZ) gate in the variational circuit with an odd (1, 3, 5, 7) number of copies, thus systematically increasing the impact of hardware errors. Markers at $\# \text{CZ} = 0$ represent the outcomes of a linear extrapolation to zero noise. The green curve is constructed with the bare energy evaluations while the data in blue are obtained by applying to each point our proposed error mitigation technique defined in Eq. (1). Combined with Richardson extrapolation, the latter procedure can recover noiseless results (dotted line) almost perfectly. The color code and the meaning of the vertical lines are consistent with the definitions in panel (a).

We compute the Hamiltonian parameters using PySCF [46] in the STO-3G basis set [47] with the restricted Hartree-Fock method [48], leading to a four-qubit encoding. For comparison to previous state-of-the-art work [13], we also simulate H_2 on two qubits exploiting the spin conservation symmetry to reduce the size of the required quantum register [49].

In the second example, we increase the complexity of the simulation by focusing on the H_3 molecule, a prototypical example of a chemical system acquiring a Berry phase under a transformation between obtuse and acute triangular shapes [50–52]. Here, we tune the height h of an originally equilateral molecule at a constant basis side length of 1 \AA : under this deformation, the spectrum exhibits a systematic change due to a level crossing at an arrangement of equidistant nuclei. The Hamiltonian is constructed similarly to the case of H_2 above, but using the unrestricted Hartree-Fock method.

As a last example, we consider the anti-ferromagnetic Heisenberg model for spin $s = 1/2$ on a tetrahedral cell

$$H_{\text{Heis}} = J \sum_{\langle i,j \rangle} \sum_{\alpha=x,y,z} \sigma_{\alpha}^{(i)} \sigma_{\alpha}^{(j)}, \quad (5)$$

where $J > 0$, σ_{α} are the usual Pauli matrices and $\langle i,j \rangle$ denote nearest neighbors. We study the evolution of the ground state energy manifold as a function of the coupling strength of one of the bonds (J') with respect to all the others. At the fully symmetric tetrahedron configuration, achieved for $J = J'$, a level crossing occurs.

Results. In Fig. 1a we report, for the 4-qubit H_2 setup at the equilibrium bond length (0.74 \AA), the distribution of outcomes for the evaluation of the ground state energy on the *ibmq_ourense* quantum processor, accessed via the IBM Quantum Experience using the Qiskit software stack [53]. The histograms are constructed by repeating 1335 times the experimental procedure, namely the evaluation of Hamiltonian expectation values for the optimal VQE parameters $\{\vec{\theta}_{\text{opt}}\}$ and the application of the algorithmic error mitigation method. For each data point, the noisy estimates of $\langle H^k \rangle$ were constructed by using $n_{\text{shots}} = 8192$ runs of the relevant quantum circuit. Compared to the exact result $E_0 \approx -1.137 \text{ Ha}$, obtained with numerical diagonalization, the bare measurement of $E = \langle H \rangle_{\vec{\theta}_{\text{opt}}}$ on the optimized ansatz

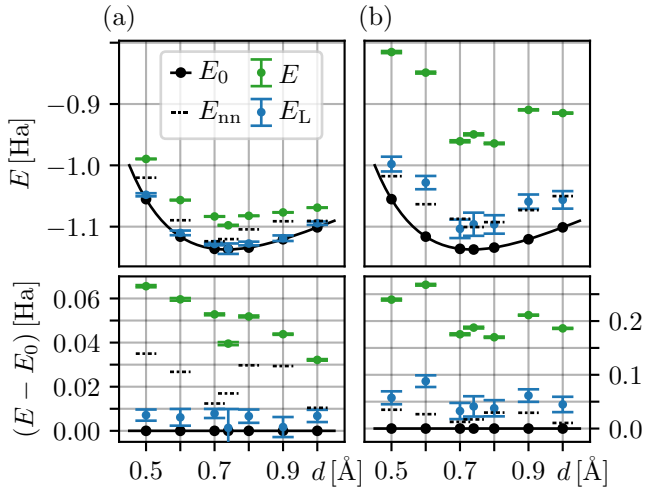


Figure 2: Dissociation profiles for H_2 computed with 2-qubit (a) and 4-qubit (b) on *ibmq_ourense* with $n_{\text{steps}} = 100$, $n_{\text{VQE}} = 10$. Bare results (green) are compared to error-mitigated ones (blue, $n_{\text{shots}} = 8 \cdot 10^5$). Black curves and dots denote the true ground state energies, and the dotted lines reference to the non-noisy energy evaluation obtained using the VQE optimal parameters resulting from (noisy) quantum hardware calculations. In all cases, the $R_y R_z$ ansatz with one entangling layer is used.

(green) yields the lowest accuracy, with an average of about $-0.9145(6)$ Ha. A direct application of the order $m = 2$ error mitigation procedure, according to Eq. (1), significantly shifts the distribution towards the reference value: however, due to the normalization required for the Lanczos-inspired algorithm, the distribution (blue) has a long tail towards small energies and large uncertainty. By increasing n_{shots} one can directly obtain smaller uncertainties σ_{H^k} on the evaluations of $\langle H^k \rangle_{\theta}$ within a single experiment. Despite being computationally expensive (typically $\sigma_{H^k} \propto 1/\sqrt{n_{\text{shots}}}$), this approach leads in our case to an average mitigated result $E_L = -1.084(7)$ Ha with $n_{\text{shots}} = 10^7$. Less demanding alternatives are the application of the cube root method (grey), yielding the smallest improvement on the energy evaluation but the most stable results, the weighted least square over groups of $n_{\text{repeat}} = 5$ repetitions (red), or the empirical selection of the a_0 and a_1 parameters (orange). The latter procedure is carried out by choosing a_0 and a_1 in such a way to keep the average estimated measurement uncertainty below 0.02 Ha, which is far smaller than the investigated energy scales, while still providing an improvement over the bare results. In fact, by tuning the ratio a_0/a_1 one can interpolate be-

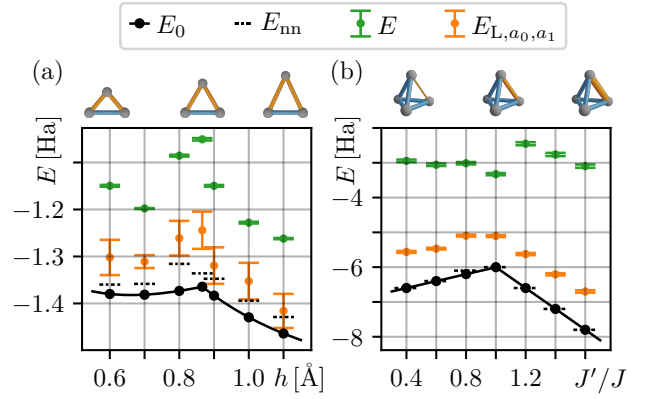


Figure 3: Ground state energies for the triangular H_3 molecule (a) and for the Heisenberg tetrahedron model (b, J' corresponds to the orange bond). Bare results (green) are compared to error-mitigated ones (orange) obtained by fixing a_0 and a_1 to achieve a maximal estimated standard error of 0.04 Ha. Black curves and dots denote the true ground state energies, dotted lines are VQE results optimized numerically and evaluated in the absence of noise. The $R_y R_z$ ansatz with 1 (3) entangling layers is used, with $n_{\text{steps}} = 2000$, $n_{\text{VQE}} = 4$ (10), $n_{\text{shots}} = 8 \cdot 10^5$ (8192), for H_3 (respectively the Heisenberg model).

tween the outcome of Eq. (1) and a more stable result, ultimately converging to the original bare value for $a_0/a_1 \rightarrow \infty$ (see also Appendix G). It is worth mentioning that a reasonable scale for the energy uncertainty, and therefore suitable values for a_0 and a_1 , can be identified through a systematic procedure involving only noisy energy estimates (see Appendix I) without prior knowledge of ideal results.

The performance of the proposed methods in presence of high noise values is assessed in Fig. 1b. Here, for the same configuration of Fig. 1a, we artificially increase the noise levels by replacing every controlled phase (CZ) operation in the entangling block of the variational circuit with an odd (3, 5, 7) number of copies [54, 55]. We observe a dominantly linear dependence on the number of CZ, with an increase of the uncertainties associated to the mitigated results E_L for higher levels of noise. Fig. 1b also shows the application of the Richardson zero-noise extrapolation (ZNE) procedure [14, 18] on both bare energy measurements and Lanczos-mitigated ones, reporting an almost complete cancellation of the hardware noise when both error mitigation strategies are combined.

To compare the Lanczos and the bare ZNE methods on an equal footing, we also take into

account the overall number of measurements required in the two cases. In Fig. 1a, we report an additional histogram, E_{CZ} , representing the distribution of error-mitigated energy reconstructions obtained with the zero-noise extrapolation (ZNE) procedure applied to the bare energies, using a similar amount of resources (i.e., measurements) as needed for the proposed Lanczos-inspired procedure (i.e., E_L results). For the latter, each of the data points contributing to the histogram is obtained by measuring 15 Pauli strings for the expectation value of the Hamiltonian $\langle H \rangle$ and 32 Pauli strings for $\langle H^2 \rangle$ and $\langle H^3 \rangle$. In total, this amounts to 79 Pauli strings measured independently using 8192 shots each. To match as close as possible these overall requirements, we increase the statistics for the energy expectation reconstructions with the ZNE approach. More specifically, for each ZNE realization we measure four times $\langle H \rangle$, one for each choice of $\#CZ = 1, 3, 5, 7$. This consists of a total of $4 \cdot 15 = 60$ Pauli strings. For each Pauli string we double the number of shots used ($2 \cdot 8192$) to estimate the expectation values, such that the overall number of measurements is actually in favour of the ZNE method. The whole procedure is repeated 50 times to construct the corresponding histogram. These results show that the use of more measurements for the ZNE method (keeping fixed the range of $\#CZ$) only slightly reduces the uncertainties, without significantly improving the average energy expectation value (see also Fig. 14 in Appendix N). In fact, we consistently observe that the Lanczos method yields results closer to the true ground state energy, although with typically larger uncertainties. It is also worth stressing that we do not in general consider (as shown in Fig. 1b) the Lanczos and ZNE methods as alternative options, but rather as complementary approaches.

The results for the full dissociation profile of H_2 encoded on 2 and 4 qubits are shown in Fig. 2. The proposed error mitigation scheme enhances the quality of the energy estimates up to a factor of 5. In particular, for the 2-qubit calculation this leads to an accuracy of about 10 mHa, which is competitive with state-of-the-art results obtained, e.g., on dedicated hardware [13, 18] and comparable to the application of quantum subspace expansion (QSE) in Ref. [27]. We notice that for $m = 2$ the Lanczos algorithm re-

quires a smaller subspace size than QSE at the cost of a larger number of measurements (i.e., more Pauli strings). It is also worth pointing out that in some cases our Lanczos-inspired error mitigation provides corrections going beyond a mere noise-cancellation effect in energy evaluations, and which can be interpreted as an improvement on the ansatz itself. This can be seen by comparing the mitigated results with what would be obtained from an ideal noiseless measurement E_{nn} , simulated numerically starting from the variational ansatz learned on the real hardware (dashed lines). Additional quantitative details on both the improvement over noiseless energy estimates, due to the very nature of Krylov-subspace methods, and the fulfillment of Eq. (3), are provided in the Appendices J and K)

Finally, in Fig. 3 we report the outcomes for the H_3 dissociation curve and for the Heisenberg tetrahedron model mapped on 6 and 4 qubits, respectively. Due to the significant increase in complexity and size of the required hardware simulations, for these models the optimal VQE parameters $\vec{\theta}_{opt}$ were computed via numerical simulations, while the final energy evaluations and the corresponding mitigated results were obtained by implementing the optimal circuit on the *ibmq_almaden* (H_3) and *ibmq_ourense* (tetrahedron) quantum processors. At difference with Fig. 2, here we made use of the empirical selection method for a_0 and a_1 in order to reduce the standard error on the mitigated points E_{L,a_0,a_1} (see Appendices for additional data). In both cases, the Lanczos-inspired scheme significantly helps in recovering the most relevant ground state features, i.e., the extremal points of the energy surfaces and level crossings.

Summary. We introduced a hardware agnostic technique to mitigate errors on near term quantum processors and we described its direct application to quantum chemistry and quantum physics problems. The proposed approach significantly improves the calculation of ground state properties by effectively enhancing the spectral weight of noisy variational ansatzes on the respective optimal solutions while ensuring the physical consistency of the outcomes. This method does not require direct control on low-level hardware operations and can in principle be combined with other available noise mitigation protocols. We reported experimental results for energy evalua-

tions closely approaching the exact reference values and demonstrated the robustness of the approach in several test cases with highly non-trivial ground state properties. The computational cost of this error mitigation procedure is ultimately associated to an increase in the number of measurements to be performed and can potentially be reduced by the introduction of more elaborate schemes for the grouping of Pauli terms appearing in the target Hamiltonian. In view of these arguments, we conclude that the proposed method has the potential to become a standard error mitigation scheme for any type of noisy quantum computing platform.

Acknowledgements. Authors acknowledge interesting discussions with Igor O. Sokolov. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (ERC-StG-Neupert-757867-PARATOP). IT acknowledges the financial support from the Swiss National Science Foundation (SNF) through the grant No. 200021-179312. IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. The current list of IBM trademarks is available at <https://www.ibm.com/legal/copytrade>.

References

- [1] Peter W. Shor. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A*, 52:R2493–R2496, Oct 1995. DOI: [10.1103/PhysRevA.52.R2493](https://doi.org/10.1103/PhysRevA.52.R2493).
- [2] A. M. Steane. Error correcting codes in quantum theory. *Phys. Rev. Lett.*, 77:793–797, Jul 1996. DOI: [10.1103/PhysRevLett.77.793](https://doi.org/10.1103/PhysRevLett.77.793).
- [3] A. R. Calderbank and Peter W. Shor. Good quantum error-correcting codes exist. *Phys. Rev. A*, 54:1098–1105, Aug 1996. DOI: [10.1103/PhysRevA.54.1098](https://doi.org/10.1103/PhysRevA.54.1098).
- [4] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018. ISSN 2521-327X. DOI: [10.22331/q-2018-08-06-79](https://doi.org/10.22331/q-2018-08-06-79).
- [5] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):4213, 7 2014. DOI: [10.1038/ncomms5213](https://doi.org/10.1038/ncomms5213).
- [6] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.
- [7] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Phys. Rev. A*, 92:042303, Oct 2015. DOI: [10.1103/PhysRevA.92.042303](https://doi.org/10.1103/PhysRevA.92.042303).
- [8] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, feb 2016. DOI: [10.1088/1367-2630/18/2/023023](https://doi.org/10.1088/1367-2630/18/2/023023).
- [9] Panagiotis Kl Barkoutsos, Jerome F. Gonthier, Igor Sokolov, Nikolaj Moll, Gian Salis, Andreas Fuhrer, Marc Ganzhorn, Daniel J. Egger, Matthias Troyer, Antonio Mezzacapo, Stefan Filipp, and Ivano Tavernelli. Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions. *Phys. Rev. A*, 98(2):022322, 2018. ISSN 24699934. DOI: [10.1103/PhysRevA.98.022322](https://doi.org/10.1103/PhysRevA.98.022322).
- [10] Guglielmo Mazzola, Pauline J. Ollitrault, Panagiotis Kl. Barkoutsos, and Ivano Tavernelli. Nonunitary operations for ground-state calculations in near-term quantum computers. *Phys. Rev. Lett.*, 123:130501, Sep 2019. DOI: [10.1103/PhysRevLett.123.130501](https://doi.org/10.1103/PhysRevLett.123.130501).
- [11] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis. Scalable quantum simulation of molecular energies. *Phys. Rev. X*, 6:031007, Jul 2016. DOI: [10.1103/PhysRevX.6.031007](https://doi.org/10.1103/PhysRevX.6.031007).

- [12] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671): 242–246, Sep 2017. ISSN 1476-4687. DOI: [10.1038/nature23879](https://doi.org/10.1038/nature23879).
- [13] M. Ganzhorn, D.J. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, and S. Filipp. Gate-efficient simulation of molecular eigenstates on a quantum computer. *Phys. Rev. Applied*, 11: 044092, Apr 2019. DOI: [10.1103/PhysRevApplied.11.044092](https://doi.org/10.1103/PhysRevApplied.11.044092).
- [14] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495, 2019. DOI: [10.1038/s41586-019-1040-7](https://doi.org/10.1038/s41586-019-1040-7).
- [15] Pauline J. Ollitrault, Abhinav Kandala, Chun-Fu Chen, Panagiotis Kl. Barkoutsos, Antonio Mezzacapo, Marco Pistoia, Sarah Sheldon, Stefan Woerner, Jay M. Gambetta, and Ivano Tavernelli. Quantum equation of motion for computing molecular excitation energies on a noisy quantum processor. *Phys. Rev. Research*, 2:043140, Oct 2020. DOI: [10.1103/PhysRevResearch.2.043140](https://doi.org/10.1103/PhysRevResearch.2.043140).
- [16] A Chiesa, F Tacchino, M Grossi, P Santini, I Tavernelli, D Gerace, and S Carretta. Quantum hardware simulating four-dimensional inelastic neutron scattering. *Nature Physics*, 15(5):455–459, 2019. DOI: [10.1038/s41567-019-0437-4](https://doi.org/10.1038/s41567-019-0437-4).
- [17] Suguru Endo, Zhenyu Cai, Simon C. Benjamin, and Xiao Yuan. Hybrid quantum-classical algorithms and quantum error mitigation. *Journal of the Physical Society of Japan*, 90(3):032001, 2021. DOI: [10.7566/JPSJ.90.032001](https://doi.org/10.7566/JPSJ.90.032001).
- [18] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski. Cloud quantum computing of an atomic nucleus. *Phys. Rev. Lett.*, 120:210501, May 2018. DOI: [10.1103/PhysRevLett.120.210501](https://doi.org/10.1103/PhysRevLett.120.210501). URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.210501>.
- [19] Ying Li and Simon C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Physical Review X*, 7(2), Jun 2017. ISSN 2160-3308. DOI: [10.1103/physrevx.7.021050](https://doi.org/10.1103/physrevx.7.021050).
- [20] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. *Physical Review Letters*, 119(18), Nov 2017. ISSN 1079-7114. DOI: [10.1103/physrevlett.119.180509](https://doi.org/10.1103/physrevlett.119.180509).
- [21] Suguru Endo, Simon C. Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Phys. Rev. X*, 8:031027, Jul 2018. DOI: [10.1103/PhysRevX.8.031027](https://doi.org/10.1103/PhysRevX.8.031027).
- [22] George S Barron, Bryan T Gard, Orien J Altman, Nicholas J Mayhall, Edwin Barnes, and Sophia E Economou. Preserving symmetries for variational quantum eigensolvers in the presence of noise. *arXiv preprint arXiv:2003.00171*, 2020.
- [23] Jarrod R. McClean, Mollie E. Kimchi-Schwartz, Jonathan Carter, and Wibe A. de Jong. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Physical Review A*, 95(4), Apr 2017. ISSN 2469-9934. DOI: [10.1103/physreva.95.042308](https://doi.org/10.1103/physreva.95.042308).
- [24] Jarrod R. McClean, Zhang Jiang, Nicholas C. Rubin, Ryan Babbush, and Hartmut Neven. Decoding quantum errors with subspace expansions. *Nature Communications*, 11(1), Jan 2020. ISSN 2041-1723. DOI: [10.1038/s41467-020-14341-w](https://doi.org/10.1038/s41467-020-14341-w).
- [25] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O’Brien, and L. DiCarlo. Experimental error mitigation via symmetry verification in a variational quantum eigensolver. *Phys. Rev. A*, 100:010302, Jul 2019. DOI: [10.1103/PhysRevA.100.010302](https://doi.org/10.1103/PhysRevA.100.010302).
- [26] Sam McArdle, Xiao Yuan, and Simon Benjamin. Error-mitigated digital quantum simulation. *Phys. Rev. Lett.*, 122: 180501, May 2019. DOI: [10.1103/PhysRevLett.122.180501](https://doi.org/10.1103/PhysRevLett.122.180501).
- [27] J. I. Colless, V. V. Ramasesh, D. Dahlen,

- M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi. Computation of molecular spectra on a quantum processor with an error-resilient algorithm. *Phys. Rev. X*, 8: 011021, Feb 2018. DOI: [10.1103/PhysRevX.8.011021](https://doi.org/10.1103/PhysRevX.8.011021).
- [28] Chao Song, Jing Cui, H. Wang, J. Hao, H. Feng, and Ying Li. Quantum computation with universal error mitigation on a superconducting quantum processor. *Science Advances*, 5(9), 2019. DOI: [10.1126/sciadv.aaw5686](https://doi.org/10.1126/sciadv.aaw5686).
- [29] Yanzhu Chen, Maziar Farahzad, Shinjae Yoo, and Tzu-Chieh Wei. Detector tomography on ibm quantum computers and mitigation of an imperfect measurement. *Phys. Rev. A*, 100:052315, Nov 2019. DOI: [10.1103/PhysRevA.100.052315](https://doi.org/10.1103/PhysRevA.100.052315).
- [30] J. M. Chow, L. DiCarlo, J. M. Gambetta, A. Nunnenkamp, Lev S. Bishop, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Detecting highly entangled states with a joint qubit readout. *Phys. Rev. A*, 81:062325, Jun 2010. DOI: [10.1103/PhysRevA.81.062325](https://doi.org/10.1103/PhysRevA.81.062325).
- [31] Colm A. Ryan, Blake R. Johnson, Jay M. Gambetta, Jerry M. Chow, Marcus P. da Silva, Oliver E. Dial, and Thomas A. Ohki. Tomography via correlation of noisy measurement records. *Phys. Rev. A*, 91: 022118, Feb 2015. DOI: [10.1103/PhysRevA.91.022118](https://doi.org/10.1103/PhysRevA.91.022118).
- [32] Benjamin Nachman, Miroslav Urbanek, Wibe A de Jong, and Christian W Bauer. Unfolding quantum computer readout noise. *npj Quantum Information*, 6(1):1–7, 2020. DOI: [10.1038/s41534-020-00309-7](https://doi.org/10.1038/s41534-020-00309-7).
- [33] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C. McKay, and Jay M. Gambetta. Mitigating measurement errors in multiqubit experiments. *Phys. Rev. A*, 103: 042605, Apr 2021. DOI: [10.1103/PhysRevA.103.042605](https://doi.org/10.1103/PhysRevA.103.042605).
- [34] Cornelius Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Natl. Bur. Stand. B*, 45:255–282, 1950. DOI: [10.6028/jres.045.026](https://doi.org/10.6028/jres.045.026).
- [35] Mario Motta, Chong Sun, Adrian T. K. Tan, Matthew J. O’Rourke, Erika Ye, Austin J. Minnich, Fernando G. S. L. Brandão, and Garnet Kin-Lic Chan. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nature Physics*, 16(2): 205–210, Nov 2019. ISSN 1745-2481. DOI: [10.1038/s41567-019-0704-4](https://doi.org/10.1038/s41567-019-0704-4).
- [36] Harish J Vallury, Michael A Jones, Charles D Hill, and Lloyd CL Hollenberg. Quantum computed moments correction to variational estimates. *Quantum*, 4:373, 2020. DOI: [10.22331/q-2020-12-15-373](https://doi.org/10.22331/q-2020-12-15-373).
- [37] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: [10.1017/CBO9780511976667](https://doi.org/10.1017/CBO9780511976667).
- [38] G. C. Wick. Properties of bethe-salpeter wave functions. *Phys. Rev.*, 96:1124–1134, Nov 1954. DOI: [10.1103/PhysRev.96.1124](https://doi.org/10.1103/PhysRev.96.1124).
- [39] Wilhelm Magnus. On the exponential solution of differential equations for a linear operator. *Communications on Pure and Applied Mathematics*, 7(4):649–673, 1954. DOI: [10.1002/cpa.3160070404](https://doi.org/10.1002/cpa.3160070404).
- [40] Richard P. Feynman. An operator calculus having applications in quantum electrodynamics. *Phys. Rev.*, 84:108–128, Oct 1951. DOI: [10.1103/PhysRev.84.108](https://doi.org/10.1103/PhysRev.84.108).
- [41] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, Jun 2020. ISSN 1745-2481. DOI: [10.1038/s41567-020-0932-7](https://doi.org/10.1038/s41567-020-0932-7).
- [42] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14 (5):447–450, 2018. DOI: [10.1038/s41567-018-0048-5](https://doi.org/10.1038/s41567-018-0048-5).
- [43] Giacomo Torlai, Guglielmo Mazzola, Giuseppe Carleo, and Antonio Mezzacapo. Precise measurement of quantum observables with neural-network estimators. *Phys. Rev. Research*, 2:022060, Jun 2020. DOI: [10.1103/PhysRevResearch.2.022060](https://doi.org/10.1103/PhysRevResearch.2.022060).
- [44] Nikolaj Moll, Panagiotis Barkoutsos, Lev S Bishop, Jerry M Chow, Andrew Cross, Daniel J Egger, Stefan Filipp, Andreas Fuhrer, Jay M Gambetta, Marc Ganzhorn,

- and et al. Quantum optimization using variational algorithms on near-term quantum devices. *Quantum Science and Technology*, 3(3):030503, Jun 2018. ISSN 2058-9565. DOI: [10.1088/2058-9565/aab822](https://doi.org/10.1088/2058-9565/aab822).
- [45] J. C. Spall. Adaptive stochastic approximation by the simultaneous perturbation method. *IEEE Transactions on Automatic Control*, 45(10):1839–1853, 2000.
- [46] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhen-dong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. Pyscf: the python-based simulations of chemistry framework. *WIREs Computational Molecular Science*, 8(1):e1340, 2018. DOI: <https://doi.org/10.1002/wcms.1340>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1340>.
- [47] W. J. Hehre, R. F. Stewart, and J. A. Pople. Self-consistent molecular-orbital methods. i. use of gaussian expansions of slater-type atomic orbitals. *The Journal of Chemical Physics*, 51(6):2657–2664, 1969. DOI: [10.1063/1.1672392](https://doi.org/10.1063/1.1672392).
- [48] Hideo Fukutome. Unrestricted hartree-fock theory and its applications to molecules and chemical reactions. *International Journal of Quantum Chemistry*, 20(5):955–1065, 1981. DOI: [10.1002/qua.560200502](https://doi.org/10.1002/qua.560200502).
- [49] Sergey Bravyi, Jay M Gambetta, Antonio Mezzacapo, and Kristan Temme. Tapering off qubits to simulate fermionic hamiltonians. *arXiv preprint arXiv:1701.08213*, 2017.
- [50] G. Herzberg and H. C. Longuet-Higgins. Intersection of potential energy surfaces in polyatomic molecules. *Discuss. Faraday Soc.*, 35:77–82, 1963. DOI: [10.1039/DF9633500077](https://doi.org/10.1039/DF9633500077).
- [51] H. von Busch, Vas Dev, H.-A. Eckel, S. Kasahara, J. Wang, W. Demtröder, P. Sebald, and W. Meyer. Unambiguous proof for berry’s phase in the sodium trimer: Analysis of the transition $a^2e'' \leftarrow x^2e'$. *Phys. Rev. Lett.*, 81:4584–4587, Nov 1998. DOI: [10.1103/PhysRevLett.81.4584](https://doi.org/10.1103/PhysRevLett.81.4584).
- [52] Raffaele Resta. Manifestations of berry’s phase in molecules and condensed matter. *Journal of Physics: Condensed Mat-*
- ter*, 12(9):R107–R143, feb 2000. DOI: [10.1088/0953-8984/12/9/201](https://doi.org/10.1088/0953-8984/12/9/201).
- [53] Gadi Aleksandrowicz, Thomas Alexander, Panagiotis Barkoutsos, Luciano Bello, Yael Ben-Haim, David Bucher, Francisco Jose Cabrera-Hernandez, Jorge Carballo-Franquis, Adrian Chen, Chun-Fu Chen, Jerry M. Chow, Antonio D. Corcoles-Gonzales, Abigail J. Cross, Andrew Cross, Juan Cruz-Benito, Chris Culver, Salvador De La Puente Gonzalez, Enrique De La Torre, Delton Ding, Eugene Dumitrescu, Ivan Duran, Pieter Eendebak, Mark Everitt, Ismael Faro Sertage, Albert Frisch, Andreas Fuhrer, Jay Gambetta, Borja Godoy Gago, Juan Gomez-Mosquera, Donny Greenberg, Ikko Hamamura, Vojtech Havlicek, Joe Hellmers, Lukasz Herok, Hiroshi Horii, Shaohan Hu, Takashi Imamichi, Toshinari Itoko, Ali Javadi-Abhari, Naoki Kanazawa, Anton Karazeev, Kevin Krsulich, Peng Liu, Yang Luh, Yunho Maeng, Manoel Marques, Francisco Jose Martin-Fernandez, Douglas T. McClure, David McKay, Srujan Meesala, Antonio Mezzacapo, Nikolaj Moll, Diego Moreda Rodriguez, Giacomo Nannicini, Paul Nation, Pauline J. Ollitrault, Lee James O’Riordan, Hanhee Paik, Jesus Perez, Anna Phan, Marco Pistoia, Viktor Prutyaynov, Max Reuter, Julia Rice, Abden Rodriguez Davila, Raymond Harry Putra Rudy, Mingi Ryu, Ninad Sathaye, Chris Schnabel, Eddie Schoute, Kanav Setia, Yunong Shi, Adenilton Silva, Yukio Siraichi, Seyon Sivarajah, John A. Smolin, Mathias Soeken, Igor Sokolov, Hitomi Takahashi, Ivano Tavernelli, Charles Taylor, Pete Taylour, Kenso Trabing, Matthew Treinish, Wes Turner, Desiree Vogt-Lee, Christophe Vuillot, Jonathan A. Wildstrom, Jessica Wilson, Erick Winston, Christopher Wood, Stephen Wood, Stefan Wörner, Ismail Yunus Akhalwaya, and Christa Zoufal. Qiskit: An Open-source Framework for Quantum Computing, January 2019.
- [54] Francesco Tacchino, Panagiotis Barkoutsos, Chiara Macchiavello, Ivano Tavernelli, Dario Gerace, and Daniele Bajoni. Quantum implementation of an artificial feed-forward neural network. *arXiv preprint*

arXiv:1912.12486, 2019.

- [55] Nikitas Stamatopoulos, Daniel J. Egger, Yue Sun, Christa Zoufal, Raban Iten, Ning Shen, and Stefan Woerner. Option Pricing using Quantum Computers. *Quantum*, 4:291, July 2020. ISSN 2521-327X. DOI: [10.22331/q-2020-07-06-291](https://doi.org/10.22331/q-2020-07-06-291).

- [56] T. O. Lewis and P. L. Odell. A general-

ization of the gauss-markov theorem. *Journal of the American Statistical Association*, 61(316):1063–1066, 1966. ISSN 01621459. DOI: [10.2307/2283200](https://doi.org/10.2307/2283200).

- [57] Gene H Golub and Charles F Van Loan. *Matrix computations*, volume 3. JHU press, 2013.

A Kraus Operator Expression for Errors

The validity of the Lanczos method requires a density matrix description of the state of the quantum computer with noise. This is guaranteed as long as the whole evolution of the state of the quantum computer can be described with Kraus operators. Indeed, in the basic noise model of qiskit, the noise of the quantum computer is divided into gate errors, thermalization errors, and readout errors. While the Kraus operator expression for the gate and thermalization errors are well known and given below, the Kraus operator expression for the readout error is derived in section B. With each application of a gate an error occurs with a probability $p_{\text{depol. error}}$. These errors are modeled via the depolarization channel, described for one qubit as

$$\begin{aligned}\rho_{\text{n,depol.}} &= (1 - \gamma_1)\rho_0 + \gamma_1 \mathbb{1}/2^N = \sum_i K_i \rho_0 K_i^\dagger, \\ K_0 &= \sqrt{1 - 3\gamma_1/4} \mathbb{1}, \\ K_1 &= \sqrt{\gamma_1/4} \sigma_x, \\ K_2 &= \sqrt{\gamma_1/4} \sigma_y, \\ K_3 &= \sqrt{\gamma_1/4} \sigma_z, \\ \gamma_1 &= p_{\text{depol. error}},\end{aligned}\tag{6}$$

returning the density matrix $\rho_{\text{n,depol.}}$ including the gate error given the state of the quantum computer after the application of the respective gate ρ_0 [37].

The thermalization error occurring during a gate duration of t is modelled using two different error sources. These are the phase flip and the generalized amplitude damping channel with respective Kraus operator expressions

$$\begin{aligned}\rho_{\text{n,p.flip}} &= \sum_i K_i \rho_0 K_i^\dagger, \\ K_0 &= \sqrt{1 - p_f/2} \mathbb{1}, K_1 = \sqrt{p_f/2} \sigma_z, \\ p_f &= 1 - e^{-t/\tau_1},\end{aligned}\tag{7}$$

where τ_1 is a decay time to be fitted, and

$$\begin{aligned}\rho_{\text{n,gad}} &= \sum_i K_i \rho_0 K_i^\dagger, \\ K_0 &= \sqrt{p} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1 - \gamma_2} \end{pmatrix}, K_1 = \sqrt{p} \begin{pmatrix} 0 & \sqrt{\gamma_2} \\ 0 & 0 \end{pmatrix}, \\ K_2 &= \sqrt{1 - p} \begin{pmatrix} \sqrt{1 - \gamma_2} & 0 \\ 0 & 1 \end{pmatrix}, K_3 = \sqrt{1 - p} \begin{pmatrix} 0 & 0 \\ \sqrt{\gamma_2} & 0 \end{pmatrix}, \\ p &= \left(e^{\frac{-\omega}{k_b T}} + 1 \right)^{-1}, \gamma_2 = 1 - e^{-t/\tau_2},\end{aligned}\tag{8}$$

where ω is the energy difference between the two qubit levels, T the temperature and τ_2 a decay time. While the phase flip channel describes a decay of the off-diagonal elements of a one-qubit density matrix, with the generalized amplitude damping channel a decay of the diagonal elements towards the Fermi-Dirac distribution is achieved.

For each applied gate these Kraus expressions are applied consecutively on the involved qubits using the noise calibration data of an IBM quantum device to find $\tau_1, \tau_2, p_{\text{depol. error}}, t, T$ and ω .

B Kraus operator expression for readout errors

In this section we explain the simulation of readout errors in qiskit and derive a Kraus operator expression so as to incorporate the readout errors by changing the density matrix.

We consider an operator O in the Pauli basis

$$O = \sum_{\mathbf{i}} o^{\mathbf{i}} \sigma_{i_0} \otimes \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_{N-1}}$$

with an expectation value of

$$\langle O \rangle = \sum_{\mathbf{i}} o^{\mathbf{i}} \langle \sigma_{i_0} \otimes \sigma_{i_1} \cdots \rangle = \sum_{\mathbf{i}, \mathbf{s} = \{s_0, s_1, \dots, s_{N-1}\}} o^{\mathbf{i}} p^{\mathbf{i}}(\mathbf{s}) \underbrace{\prod_j s_j}_{=s}, \quad (9)$$

where $s^j \in \{-1, 1\}$ is the measurement outcome of $\sigma_{i_j} \in \{\sigma_0 = \mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}$ with a measurement probability of $p^{\mathbf{i}}(\mathbf{s})$ for the measurement result $\mathbf{s} \in \{-1, 1\}^{\otimes N}$ for the Pauli string \mathbf{i} . Now, this probability changes with the readout error $p_r^{\mathbf{i}}(\mathbf{s}|\mathbf{s}')$ for a given Pauli string \mathbf{i} , i.e., it encodes the probability to measure \mathbf{s} instead of \mathbf{s}' given that the measurement outcomes of the single qubits were $\mathbf{s}' = \{s'_0, \dots\}$. Now, the measurement result with readout error (r.e.) can be written as

$$\langle O \rangle_{\text{r.e.}} = \sum_{\mathbf{i}, \mathbf{s}} o^{\mathbf{i}} p_{\text{r.e.}}^{\mathbf{i}}(\mathbf{s}), \quad (10)$$

$$p_{\text{r.e.}}^{\mathbf{i}}(\mathbf{s}) = \sum_{\mathbf{s}'} p_r^{\mathbf{i}}(\mathbf{s}|\mathbf{s}') p^{\mathbf{i}}(\mathbf{s}'). \quad (11)$$

In qiskit, $p_r^{\mathbf{i}}(\mathbf{s}|\mathbf{s}')$ is described using the probability for a single qubit flip $p_{r,j}^{i_j}(s_j|s'_j)$ on qubit j , which is an approximation making the probability that a readout error occurs on one qubit independent of the measurement result of all other qubits. Furthermore, we assume that $p_{r,j}^{i_j}(1|-1) = p_{r,j}^{i_j}(-1|1)$ implying $p_{r,j}^{i_j}(1|1) = p_{r,j}^{i_j}(-1|-1)$. If this assumption is not fulfilled, the variational principle is not fulfilled in general anymore (see counterexample below). Intuitively, if $p_{r,j}^{i_j}(1|-1) \neq p_{r,j}^{i_j}(-1|1)$, the operator expression becomes dependent on the measurement result, so that a Kraus operator expression cannot be constructed anymore. This yields

$$p_r^{\mathbf{i}}(\mathbf{s}|\mathbf{s}') = \prod_j p_{r,j}^{i_j}(s_j|s'_j) \quad (12)$$

so that

$$\langle O \rangle_{\text{r.e.}} = \sum_{\mathbf{i}, \mathbf{s}, \mathbf{s}'} o^{\mathbf{i}} \left[\prod_j s_j p_{r,j}^{i_j}(s_j|s'_j) \right] p^{\mathbf{i}}(\mathbf{s}') = \sum_{\mathbf{i}, \mathbf{s}'} o^{\mathbf{i}} \left[\prod_j p_{r,j}^{i_j}(1|s'_j) - p_{r,j}^{i_j}(-1|s'_j) \right] p^{\mathbf{i}}(\mathbf{s}'). \quad (13)$$

This allows us to further simplify

$$p_{r,j}^{i_j}(1|s'_j) - p_{r,j}^{i_j}(-1|s'_j) = s'_j [p_{r,j}^{i_j}(s'_j|s'_j) - p_{r,j}^{i_j}(-s'_j|s'_j)] = s'_j [1 - 2p_{r,j}^{i_j}(-s'_j|s'_j)]. \quad (14)$$

To emphasize the independence of $p_{r,j}^{i_j}(-s'_j|s'_j)$ on s'_j by assumption we write $p_{r,j}^{i_j} \equiv p_{r,j}^{i_j}(-s'_j|s'_j)$. This allows us to write

$$\langle O \rangle_{\text{r.e.}} = \sum_{\mathbf{i}, \mathbf{s}'} o^{\mathbf{i}} \left[\prod_j (1 - 2p_{r,j}^{i_j}) \right] s' p^{\mathbf{i}}(\mathbf{s}') = \sum_{\mathbf{i}} o^{\mathbf{i}} \left\langle \bigotimes_j \sigma_{i_j} (1 - 2p_{r,j}^{i_j}) \right\rangle. \quad (15)$$

To be able to write down Kraus operators, we further assume that $p_{r,j}^x = p_{r,j}^y = p_{r,j}^z$ which we refer to by $p_{r,j}$ in the following. This assumption is reasonable, as the only difference in the measurement of $\sigma_x, \sigma_y, \sigma_z$ is one single qubit rotation, which has a small error, which is neglected here. Furthermore, we have $p_{r,j}^{\mathbb{1}} = 0$ as the identity does not need to be measured. Then, we can write

$$\sigma_{i_j} (1 - 2p_{r,j}^{i_j}) = \epsilon_j(\sigma_{i_j}) \quad (16)$$

with

$$\epsilon_j(\sigma_{i_j}) = \sigma_{i_j} (1 - 2p_{r,j}) + p_{r,j} (\sigma_{i_j} + \sigma_{x,j} \sigma_{i_j} \sigma_{x,j} + \sigma_{y,j} \sigma_{i_j} \sigma_{y,j} + \sigma_{z,j} \sigma_{i_j} \sigma_{z,j}) / 2 \quad (17)$$

$$= \sigma_{i_j} (1 - 2p_{r,j}) + p_{r,j} \text{Tr}(\sigma_{i_j}) \mathbb{1}. \quad (18)$$

Using this expression, it becomes evident, that ϵ_j is a Kraus operator. This simplifies the error calculation to

$$\langle O \rangle_{\text{r.e.}} = \sum_{\mathbf{i}} o^{\mathbf{i}} \left\langle \bigotimes_j \epsilon_j(\sigma_{i_j}) \right\rangle = \sum_{\mathbf{i}} o^{\mathbf{i}} \left\langle \epsilon \left(\bigotimes_j \sigma_{i_j} \right) \right\rangle = \langle \epsilon(O) \rangle, \quad (19)$$

with $\epsilon = \epsilon_0 \circ \epsilon_1 \cdots$ and ϵ_j acts trivially on all sites except for site j .

Finally, we can use that

$$\langle O \rangle_{\text{r.e.}} = \text{Tr}(\rho \epsilon(O)) = \text{Tr}(\epsilon(\rho) O)$$

to describe the readout error by changing the density matrix.

Violation of the Variational Principle for $p_{r,j}^{i_j}(1|1) \neq p_{r,j}^{i_j}(-1|1)$ The variational principle is not in general fulfilled anymore for $p_{r,j}^{i_j}(1|1) \neq p_{r,j}^{i_j}(-1|1)$. For this, consider the example $H = -\sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y - \sigma_z \otimes \sigma_z = -2(\mathcal{S}^2 - 2s^2) = -2(S(S+1) - 2s(s+1))$ with spectrum $\{-1, -1, -1, 3\}$. One eigenstate with eigenvalue -1 is $\Psi_{S=0, S_z=1} = |00\rangle$. Assuming equal readout error probabilities for the two qubits $p_{r,j}^{x,y,z}(s|s) = p(s|s)$ the expectation value of the Hamiltonian becomes

$$\langle 00|H|00\rangle = -1 + 4p(1|0)(1 - p(1|0)) - 2(p(0|1) - p(1|0))^2, \quad (20)$$

which can be below the minimal eigenvalue of -1 for $p(1|0) \neq p(0|1)$ and as shown larger equals -1 for $p(1|1) = p(-1|1)$.

C Relation to the Lanczos Algorithm and Generalization

In the Lanczos algorithm, the ground state is approximated using the ground state of the Hamiltonian expressed in an order m Krylov space $\mathcal{K}^{(m)}$. The matrix expression for the Hamiltonian is real and symmetric, as all matrix entries are real functions of $\{\langle \Psi | H^l | \Psi \rangle \in \mathbb{R}\}_{l=0}^{2m-1}$. Thus, every eigenstate can be chosen to be real. The ground state is the linear combination with real coefficient with minimal eigenvalues of all elements of the Krylov space. Since the minimal eigenvalue is the lowest possible

measurement outcome for all states in $\mathcal{K}^{(2)}$ the minimization in Eq. (1) cannot yield a lower energy estimate. Furthermore, the linear combination of states parametrized with a_0 and a_1 spans $\mathcal{K}^{(2)}$ so that the ground state lies within the space parametrized with a_0 and a_1 . Hence, the minimization in Eq. (1) yields the ground state of the Hamiltonian expressed in $\mathcal{K}^{(2)}$ and Eq. (1) evaluates to its eigenvalue.

This holds also for higher orders m for the generalized expression of Eq. (1) that reads

$$E_{L,k,n,m} = \min_{\mathbf{a} \in \mathbb{R}^m} \sqrt[k]{\frac{\langle \Psi | H^k (\sum_{i=0}^{m-1} a_i H^i)^n | \Psi \rangle}{\langle \Psi | (\sum_{i=0}^{m-1} a_i H^i)^n | \Psi \rangle}}, \quad (21)$$

for which $E_{L,k,n,m} > E_0$ holds with odd k and even n . This follows by generalization of the statements of the main text, this section and Appendix F. For $k = 1, n = 2$ expression (21) is equivalent to the Lanczos method for the smallest eigenvalue in an order- m Krylov space, as can be shown with the same argument as for $\mathcal{K}^{(2)}$. Furthermore, $E_{L,k=1,n,m} \leq E$, as E is the special case of $a_0 = 1, a_{i>0} = 0$.

The direct relationship with the classical Lanczos algorithm immediately suggests a practical way to solve Eq. (1). For the 2-dimensional case ($m = 2$) presented in this work, the minimization is in fact equivalent to the diagonalization of a 2×2 matrix, which can be done analytically. The matrix is constructed by expressing all matrix elements in terms of the measurement results for $\langle H \rangle_{\vec{\theta}}, \langle H^2 \rangle_{\vec{\theta}}$ and $\langle H^3 \rangle_{\vec{\theta}}$. The matrix elements can be obtained by constructing $\mathcal{K}^{(2)} = \{|v_1\rangle = |\Psi(\vec{\theta})\rangle, |v_2\rangle = (H - \alpha_1)|\Psi(\vec{\theta})\rangle/\beta_2\}$ with $\alpha_1 = \langle H \rangle_{\vec{\theta}}, \beta_2 = \sqrt{\langle H^2 \rangle_{\vec{\theta}} - \langle H \rangle_{\vec{\theta}}^2}$. In this basis, the matrix expression for the Hamiltonian reads

$$H_{\mathcal{K}^{(2)}} = \begin{pmatrix} \langle v_1 | H | v_1 \rangle & \langle v_1 | H | v_2 \rangle \\ \langle v_2 | H | v_1 \rangle & \langle v_2 | H | v_2 \rangle \end{pmatrix} = \begin{pmatrix} \alpha_1 & \beta_2 \\ \beta_2 & \alpha_2 \end{pmatrix} \quad (22)$$

where $\alpha_2 = (\langle H^3 \rangle_{\vec{\theta}} - 2\langle H^2 \rangle_{\vec{\theta}}\langle H \rangle_{\vec{\theta}} + \langle H \rangle_{\vec{\theta}}^3)/\beta_2^2$. In higher orders (i.e., $m > 2$), the minimization can in principle be performed in a similar way by constructing and diagonalizing, in a classical post-processing stage, the corresponding matrix $H_{\mathcal{K}^{(m)}}$. Alternatively, one could also achieve the desired result with an iterative minimization procedure whose cost function, parametrized by a_0, a_1 , is defined in Eq. (1) of the main text. A similar approach is also possible at higher orders, when a suitable parametrization to navigate the appropriate Krylov subspace is considered.

Concerning the regularization techniques proposed in the main text and discussed in more details in some Appendix sections below, they can in principle be generalized straightforwardly to higher orders m . This is for instance true for the direct application of the Lanczos method (with the generalized version of Eq. (1) presented in Eq. (21) above) as well as for the closely related weighted least squares averaged version. The same Eq. (21) shows how the empirical tuning of the parameters a_i could scale for higher orders, with $m - 1$ independent parameters to optimize for an order m mitigation. Although it is true that such approach could become impractical for large values of m , a viable solution could be to apply an iterative minimization procedure of a suitably parametrized cost function, which can be constructed from the noisy estimates of powers of the Hamiltonian. This could possibly be used to explore the neighborhood of the formal solution obtained with, e.g., the numerical diagonalization of the reduced Hamiltonian in the Krylov subspace. Finally, the cube root version can be replaced with similar higher order constructions of the form $\sqrt[k]{\langle H^k \rangle_{\vec{\theta}}}$, for odd k .

D Scaling

In general terms and in the absence of more specific assumptions, when two-body interactions are present H^2 and H^3 are, respectively, 8th and 12th order operators, with correspondingly large numbers of Pauli strings to be measured in order to reconstruct average values. However, as we argue below

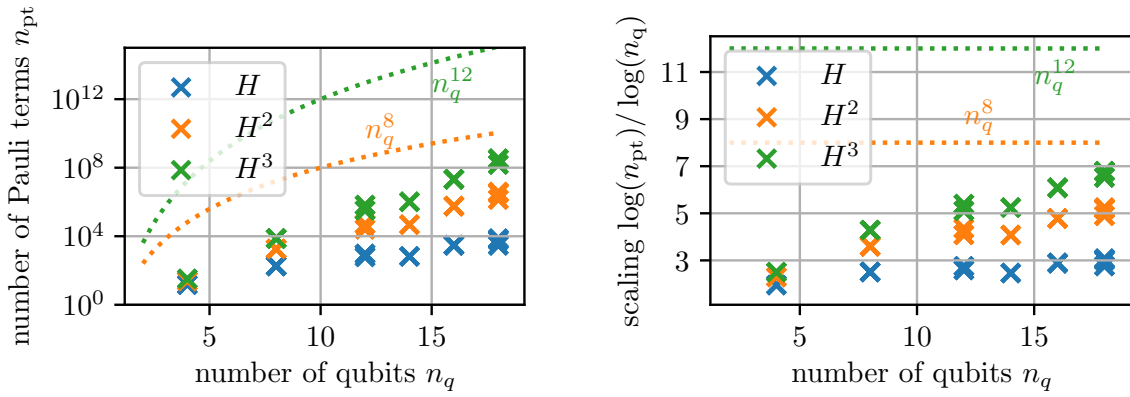


Figure 4: Number of Pauli terms n_{pt} for different real molecules encoded on n_q qubits. In the left panel the values are shown using a logarithmic scale, while in the right panel the exponent y is shown, so that $n_q^y = n_{\text{pt}}$. The terms of the simulated molecules are calculated in the sto3g basis if not stated otherwise: H_2 ($n_q = 4$), H_2 ($n_q = 8$, basis: 631g), H_4 ($n_q = 8$), H_6 ($n_q = 12$), LiH ($n_q = 12$), H_2O ($n_q = 14$), NH_3 ($n_q = 16$), H_8 ($n_q = 16$), H_{10} ($n_q = 18$), NN ($n_q = 18$) and NaH ($n_q = 18$), where the 18-qubit molecules are simulated using a two-qubit reduction scheme for implementation purposes. The Pauli terms for the Hamiltonian H are calculated using PySCF and the H_2 equilibrium interatomic distance of 0.75 \AA and 1.6 \AA for LiH . The geometry of the molecules was chosen to be one dimensional except NH_3 , which was simulated on a tetrahedron. The results for similar number of qubits n_q tend to agree for different molecules.

through an extensive collection of numerical investigations, the reality for many cases of practical interest shows effective results which are much more favorable.

Focusing on quantum chemistry problems, in Fig. 4, we report the number of Pauli terms corresponding to the Hamiltonian operators H^n ($n = 1, 2, 3$) for a collection of 11 different molecules which require up to $n_q = 18$ encoding qubits. These were obtained numerically without resorting to grouping techniques, which might provide a further reduction of approximately an order of magnitude in the number of independent measurement settings required for the reconstruction of average values. We observe, in particular, that the number of terms in H seems to saturate around a power law of the form n_q^3 , making it effectively a 3rd-order operator. This behaviour is typical whenever conservation laws are present, which provide implicit constraints to the number of available physical interactions. Regarding the terms H^2 and H^3 , although the trend is actually still far from achieving a power law behavior with a well established exponent, we can already explicitly observe that the actual number of Pauli strings remains far from the predicted values $\mathcal{O}(n_q^8)$ for H^2 , respectively $\mathcal{O}(n_q^{12})$ for H^3 , and even below the $\mathcal{O}(n_q^6)$ (respectively $\mathcal{O}(n_q^9)$) that could be inferred by the 3rd order nature of the H operators employed. We interpret such results by mentioning two concurrent effects: on one hand, the commutation algebra of the Pauli strings effectively reduces the number of terms appearing in successive products of H . On the other hand, the total number of independent Pauli strings that can be constructed with n_q qubits equals 4^{n_q} , and can be quickly saturated for small and medium sized systems. While we cannot exclude that in the large n_q limit an actual 8th or 12th order scaling is followed, the onset of such regime appears to be still beyond many of the examples of current practical interest in near term applications. The latter represent, in fact, some ideal candidates for our proposed error mitigation strategy.

Much more favourable scaling properties can also be predicted, independently of the size of the problem, for local Hamiltonians: There, the number of terms scales as n_q , so that the Lanczos method scales in the worst case like n_q^3 . This class of Hamiltonians is represented in the main text by the Heisenberg tetrahedron. Local Hamiltonians constitute by themselves an interesting field of research with many connections, e.g., to the theory of magnetism (see also a recent work by Vallury et al. [36], where the authors analyse an application of Krylov methods for local Hamiltonians), and include among others many examples of spin and lattice systems.

Finally, it is worth noticing that all the analysis conducted in the present work remains valid whenever

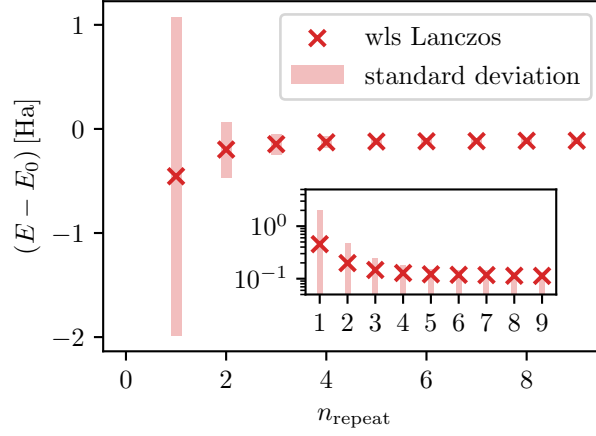


Figure 5: The results of the dependence of the wls method (red) on the number of Lanczos estimates (n_{repeat}) used is shown for a distance of 0.74 \AA for the simulation of H_2 on four qubits. The results are obtained by using the 100 measurements on the quantum computer and choose n_{repeat} of them randomly 1000 times. The mean of the results and the standard deviation of the results is shown for the wls. As a conclusion for $n_{\text{repeat}} \gtrsim 5$ we obtain similar results, validating the use of $n_{\text{repeat}} = 5$ in the main text.

a different observable O is used instead of the Hamiltonian. In such case, only OHO has to be measured, while the variational principle is still fulfilled and the ground state energy estimate will not be worse than the noisy prediction. In fact, H can actually be used to guide the choice of the operator O , which can be as small as the quantum computer is capable of dealing with. However, an analysis of this strategy requires a more detailed study which is left for future work.

E H_2 Saturation of wls

In this section we show the results of taking the weighted least square average (wls) over n_{repeat} measurement outcomes. In this case the error mitigated energy estimate can be obtained via the formula:

$$\bar{E}_{\text{L,wls}} = \frac{\sum_{i=0}^{n_{\text{repeat}}-1} e_{\text{L},i} \sigma_i^{-2}}{\sum_{i=0}^{n_{\text{repeat}}-1} \sigma_i^{-2}}, \quad (23)$$

with standard error $\sigma_{\bar{E}_{\text{L,wls}}}^{-2} = \sum_i \sigma_i^{-2}$. Notice that this solution fulfills the generalized Gauss-Markov Theorem [56], and therefore $\bar{E}_{\text{L,wls}}$ has minimal variance. The results are shown in Fig. 5, where we plotted the dependency of the mean of the wls result and its standard deviation against varying n_{repeat} . As expected, for increasing n_{repeat} the standard deviation of the wls decreases faster than $1/\sqrt{n_{\text{repeat}}}$, which is validating the generalized Gauss-Markov Theorem. Furthermore, we can observe an increase of the mean of the wls result with increasing n_{repeat} . The reason is the small uncertainty for results with larger energy, which therefore have a larger weight. The higher n_{repeat} , the more unlikely are wls results which only contain low energy estimates with larger measurement uncertainty. Hence, the mean of the wls result increases. The mean for $n_{\text{repeat}} = 1$ is below the true ground state energy showing that this value might not be useful and reflecting the necessity of considering the estimated measurement uncertainties.

F Variational Principle for the Cube Root Method

For the cube root method the variational principle holds as well since for any odd k

$$\langle H^k \rangle = \sum_i \alpha_i E_i^k \geq \sum_i \alpha_i E_0^k = E_0^k \quad (24)$$

and $\sqrt[k]{\cdot}$ is monotone restricting the image of $\sqrt[k]{\cdot}$ to the real axis.

The effective enhancement of the ground state spectral weight can be understood by comparing the expectation value

$$\langle H^k \rangle_{\vec{\theta}} = \sum_i (\alpha_i E_i^{k-1}) E_i \quad (25)$$

with

$$\langle H \rangle_{\vec{\theta}}^k = \sum_i (\alpha_i \langle H \rangle_{\vec{\theta}}^{k-1}) E_i, \quad (26)$$

where α_i represents the probability of observing the i -th energy eigenstate on a given variational (noisy) ground state estimate. If we assume that the dominant contributions come from the true ground state ($i = 0$) and other eigenstates for which $|E_i| < |E_0|$, i.e.,

$$\sum_{0, i: |E_i| < |E_0|} \alpha_i \gg \sum_{i: |E_i| \geq |E_0|} \alpha_i, \quad (27)$$

we have that the weight associated to the ground state energy in the expression of Eq. (25) is enhanced by a factor $E_0^{k-1} / \langle H \rangle_{\vec{\theta}}^{k-1} > E_i^{k-1} / \langle H \rangle_{\vec{\theta}}^{k-1}$, which is larger for the ground state, with respect to the quantity in Eq. (26). This assumption is reasonable, for example, in those cases where the ground state energy and the energies of the lowest excited states are negative. Although the cube root method is not *a priori* guaranteed to be helpful in all cases, the fact that it complies with the variational principle makes it straightforward to discard the result $\sqrt[k]{\langle H^k \rangle_{\vec{\theta}}}$ whenever this is larger than the unmitigated value $\langle H \rangle_{\vec{\theta}}$. In such case, one could infer that the (noisy) variational quantum state features a significant occupation of high energy eigenstates.

G Dependence of the Lanczos Energy Estimate on a_0, a_1

In Fig. 6 and 7 the dependence of the energy estimate on the choice of parameters a_0, a_1 is shown. For H_3 even for large number of measurements we could observe an instability of the method. This is related to the fact, that the measurements have been taken over a long time period of six days during which the noise fluctuates significantly. This instability can be detected considering the estimates for the measurement uncertainty. This is shown in Fig. 6, where the measurement uncertainty also increases for too small energy estimates. In all cases considered, the ground state energy was found to be within a likely number of standard deviations of the measurement result so that the measurement uncertainty reflects the accuracy of the method sufficiently well.

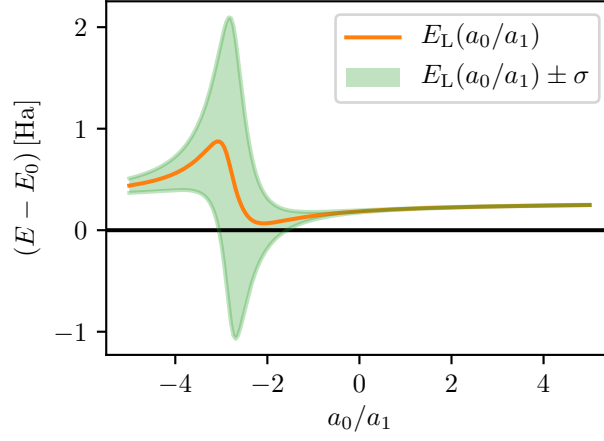


Figure 6: The dependence of the mean of 100 energy estimates using formula (1) for fixed a_0/a_1 with reference to E_0 and the mean of the estimated standard errors of the 100 results are shown for an height of 0.8 \AA for H_3 . In general, we observe that for larger a_0/a_1 the standard deviation decreases, while the energy estimate increases. For $a_0/a_1 \approx E_0$ the standard deviation is the largest as most of the weight of the statevector is projected out, the norm is close to 0. In conclusion, when fixing the parameter a_0/a_1 a value needs to be chosen which provides a good compromise between increased ground state energy estimates and standard errors. For $a_0/a_1 \rightarrow \infty$ the method converges to the bare measurement of $\langle H \rangle_{\vec{\theta}_{\text{opt}}}$. In order to reduce the standard error by a factor of $1/\sqrt{n}$, n estimates can be averaged over.

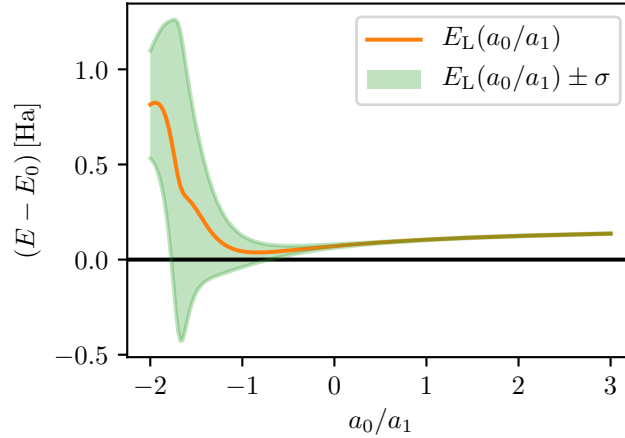


Figure 7: The dependence of the mean of 100 energy estimates using formula (1) for fixed a_0/a_1 with reference to E_0 and mean of the estimated standard errors of the 100 results are shown for an internuclear distance of 0.74 \AA for H_2 on four qubits. In general, we observe that for larger a_0/a_1 the standard deviation decreases, while the energy estimate increases. For $a_0/a_1 \approx E_0$ the standard deviation is the largest as most of the weight of the statevector is projected out, the norm is close to 0. In conclusion, when fixing the parameter a_0/a_1 a value needs to be chosen which provides a good compromise between increased ground state energy estimates and standard errors. For $a_0/a_1 \rightarrow \infty$ the method converges to the bare measurement of $\langle H \rangle_{\vec{\theta}_{\text{opt}}}$. In order to reduce the standard error by a factor of $1/\sqrt{n}$, n estimates can be averaged over.

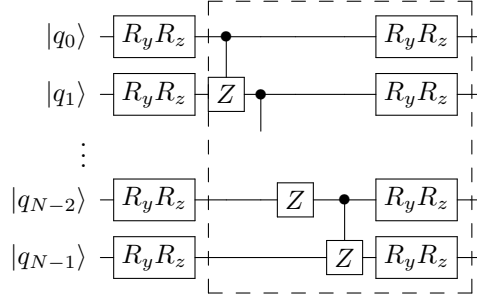


Figure 8: This figure depicts an example of an $R_y R_z$ circuit.

H Variational Form

In this section the used variational form, namely the $R_y R_z$ ansatz, is shown in Fig. 8. It consists of a layer of initial rotations followed by entanglement layers. One such entanglement layer is shown below denoted with the box. This entanglement layer is repeated as often as required to obtain the highest achievable accuracy.

I Empirical selection of the parameters a_0, a_1

In this chapter we want to give a quantitative description for how the bound of $\sigma_{\max} = 40$ mHa on the energy estimates $E_{L,h,a_0/a_1}$, where h is the varied system parameter, is calculated. Such value was chosen by comparing the results for the whole dissociation profiles for different values of a maximal measurement uncertainty σ_{\max} , essentially identifying reasonable energy scales from typical energy differences in the system obtained from noisy estimates. This corresponds to a modification of Eq. (1) in the main text to $E_{L,a_0,a_1} = \min_{a_0,a_1|C} \text{Tr}(\rho_L(a_0, a_1)H)$ with the additional constraint $C : \sigma_{E_{L,a_0,a_1}} \leq \sigma_{\max}$, where $\sigma_{E_{L,a_0,a_1}}$ is the estimated measurement uncertainty on E_{L,a_0,a_1} . Note that reasonable values of σ_{\max} can be found only in the range $[\sigma_E, \sigma_{E_{L,h}}]$. To quantify the choice of σ_{\max} , we evaluate the sum of the energy level distances for consecutive ground state energy estimates

$$\Delta_h E = \sum_h |E_{L,h,a_0/a_1} - E_{L,h+\delta,a_0/a_1}|, \quad (28)$$

which we expect to be the lowest for a more pronounced energy profile (i.e. by requiring adjacent energy estimates to be close and therefore reducing noise induced fluctuations, this strategy highlights peaks or discontinuities in the profile, if they are present, without actually making any assumptions on the existence of such features). This is shown in Fig. 9 (a) for H_3 and (b) for H_2 , for the tetrahedron $\sigma_{E_{L,J'/J}} < \sigma_{\max}$, for any $\sigma_{\max} \lesssim |E_{L,J'/J,a_0/a_1} - E_{L,J'/J+\delta,a_0/a_1}|$ so that $\Delta_{J'/J} E$ is constant. We denote with $\sigma_{E_{L,J'/J}}$ the measurement uncertainty for the energy estimate of Eq. (1) in the main text.

We stress that choosing a σ_{\max} is always legitimate to do, as the results are bounded from below by the true values via the variational principle and from above by the non-mitigated, noisy results. Hence, by tuning the results as a whole an improvement within these boundaries is expected.

In summary, the proposed error mitigation procedure is particularly well suited to refine energy profiles and to bring to light extremal features in the spectrum. Towards this goal, we suggest to choose a common bound on the standard deviation in such a way that the calculated energy differences are the smallest. This does not involve any requirement regarding knowledge of the solution a priori.

J Groundstate overlap improvement

In the main text a sufficient condition for the improvement of the overlap between the true ground state and the ground state estimate is presented. It requires the fulfillment of Eq. (3) of the main text

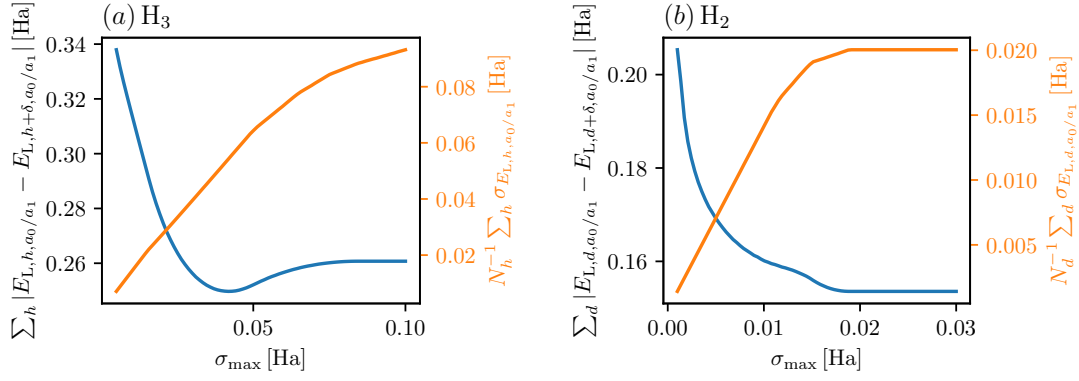


Figure 9: The dependence of the summed level distance of consecutive parameters is shown in (a) for H_3 and (b) for H_2 for different chosen cutoffs σ_{\max} , which is used for all heights h (distances d for H_2). Given the cutoff and the system parameter h , a_0/a_1 is chosen as minimizing the ground state energy estimate yielding $E_{L,h}$ (Eq. (1) main text) if $\sigma_{E_{L,h}} < \sigma_{\max}$ or as the smallest value a_0/a_1 so that $\sigma_{E_{L,h},a_0/a_1} < \sigma_{\max}$. We observe a saturation for large σ_{\max} , since all estimated errors are below the cutoff so that it is not applied. For small cutoffs, the $\Delta_h E$ increases up to the non-error mitigated value. In orange the estimated noise error on $E_{L,h,a_0/a_1}$ averaged over h is shown. Its saturation for large cut offs σ_{\max} reflects the circumstance that the chosen cutoff is larger than the estimated standard error σ_{E_L} .

and $a_0/a_1 > E_L$ for the optimal a_0/a_1 for $a_1 \neq 0$. In Tab. 1 we report the results for the evaluation of such condition for all the data points presented in the main text. We recall that whenever the presented value is larger than one, an improvement of the overlap between the ground state estimate and the true ground state is expected. For each of the values provided, we have that $a_0/a_1 > E_L$ and, as reported in the Table, we evaluate the left hand side of Eq. (3)

$$(\bar{a}_0 - \bar{a}_1 E_L)^2 / \text{Tr} [\rho(\bar{a}_0 - \bar{a}_1 H)^2] \quad (29)$$

to be larger than 1 for all cases. Eq. (3) is therefore fulfilled and the overlap between the ground state and the estimate is improved.

parameter system	values with uncertainties							mean, standard deviation
d [Å]	0.50	0.60	0.70	0.74	0.80	0.90	1.00	
H_2 2-qubit	1.039(4)	1.042(7)	1.033(4)	1.04(2)	1.039(6)	1.05(1)	1.025(6)	1.038, 0.007
d [Å]	0.50	0.60	0.70	0.74	0.80	0.90	1.00	
H_2 4-qubit	1.15(2)	1.15(2)	1.14(3)	1.16(4)	1.14(4)	1.17(3)	1.18(4)	1.16, 0.01
h [Å]	0.60	0.70	0.80	0.87	0.90	1.00	1.10	
H_3	1.11(4)	1.09(2)	1.15(5)	1.15(4)	1.15(5)	1.11(5)	1.17(6)	1.13, 0.03
J'/J	0.60	0.70	0.80	0.87	0.90	1.00	1.10	
Heisenberg	1.341(9)	1.310(9)	1.261(8)	1.216(7)	1.341(9)	1.342(8)	1.324(8)	1.31, 0.05

Table 1: The results for the evaluation of the left side of Eq. 3 are reported for the four considered systems in the main text: H_2 simulated on 2 and 4 qubits, H_3 and the Heisenberg spin chain. The values are given for each parameter configuration used for the respective system with the uncertainty on the last digit due to the measurement uncertainty on H, H^2 and H^3 . In the last column, the average of all the values for each system is given as well as the standard deviation of the values around the given mean.

K Noiseless Results Improvement

In this section, we provide insights about the improvement of the Lanczos estimates over noiseless energy estimates. This effect is a direct consequence of the application of the Lanczos diagonalization

method within Krylov subspaces. Indeed, similarly to a classical diagonalization problem, any noiseless simulation of a variational ansatz essentially corresponds to an approximation of the eigenvector of the target Hamiltonian associated to the minimum eigenvalue. It is therefore not surprising that the quality of such approximation can benefit from the application of the Lanczos method, as this is explicitly designed to increase the spectral overlap with the exact solution. We notice, for example, that this appears to be the main motivation behind some recent work by Vallury et al. [36]. As we discuss in the main text, the aim of our work is, besides remarking the usefulness of this effect for quantum variational algorithms, to demonstrate its very promising parallel application as a systematic error mitigation procedure.

For completeness, we report in Fig. 10 the results obtained from numerical noiseless simulations of the VQE algorithm for the four example models presented in the main text. These are obtained with Qiskit statevector simulator, which works out the linear algebraic calculations without the effect of sampling or hardware noise. It shows that the application of the method described in this work provides an improvement towards the true ground state values, which is purely due to the effective increased spectral overlap provided by the Lanczos procedure.

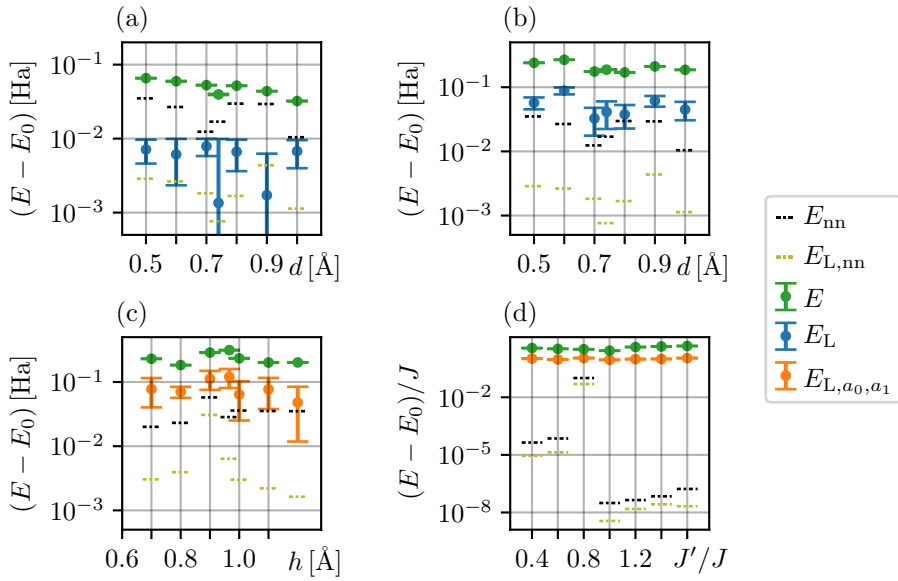


Figure 10: Here we show in addition to the data presented in Fig. 2 and Fig. 3 in the main text the improvement obtained via the application of the Lanczos algorithm on the noiseless results for the four cases discussed in the main text (a) H_2 simulated on two qubits, (b) H_2 simulated on four qubits, (c) H_3 and (d) the Heisenberg model.

L Convergence Rate

In principle, due to the delicate scaling properties of the method already at the lowest $m = 2$ order, we do not expect that higher orders will be used extensively. Hence, we do not provide a detailed numerical study for the convergence of the method as a function of m . Instead we can make use of the Kaniel-Paige convergence theory for density matrices. This yields a convergence $|E_0 - E_L| < \text{constant} \cdot r^n$, where n is the order of the method and r is a convergence rate. This corresponds to what was observed when considering examples. For more information, see e.g. chapter 10.1.5 in Ref. [57] (in combination with the trigonometric definition of the Chebyshev polynomials).

M Effect of shot noise and error propagation

In this section, we provide additional insights about the results for E_L reported in Fig. 1 of the main text, where some data points placed below the true ground state energy appear to contradict the

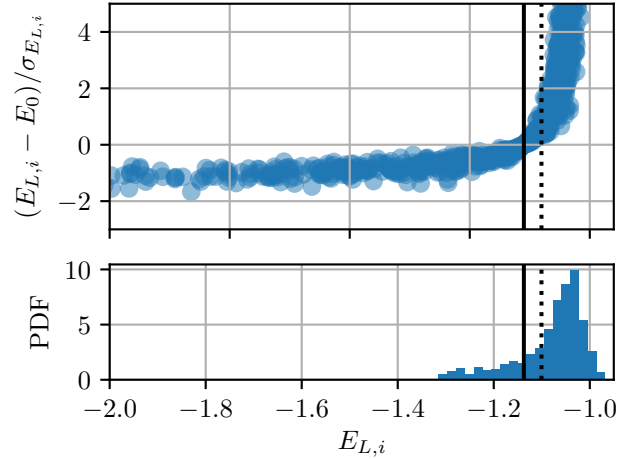


Figure 11: In this figure we show additional properties of the numerical data, which is also presented in Fig. 1 of the main text. For each data point we show on the x axis the energy estimate and, on the y axis, the corresponding deviation from the true ground state value in terms of the measurement uncertainty. In the lower panel the histogram of the data in the upper panel is shown, which is the same as presented in the main text in Fig. 1. Like in the figure in the main text, the solid black line marks E_0 , the dashed black line the noiseless estimate E_{nn} .

validity of the variational principle. We support our discussion with the analysis presented in Fig. 11, where more detailed information is provided particularly regarding the uncertainty associated to the mitigated energy estimates. As it can be seen, the long tails of the energy distribution should be interpreted as a consequence of shot noise and error propagation effects. Indeed, as we highlighted in the main text and in Appendices A and B, all other possible sources of errors act as quantum channels, thus always yielding valid density matrices. However, statistical fluctuations due to the finite size of the samples taken from a quantum hardware can for example lead to apparent violations of the variational bounds whenever the minimization procedure in Eq. (1) of the main text yields an ill-conditioned denominator. Nevertheless, as we report in Fig. 11, such extreme cases are always clearly marked by large intrinsic errors accompanying the energy estimate and can thus be clearly identified and marked as unphysical, in what essentially represents a self-consistence mechanism for the proposed method.

N Additional Figures

This section contains additional results for the experiments presented in the main text. In Fig. 12 the result for the Lanczos method are shown for the parameters a_0, a_1 minimizing the energy expectation value for the dataset also presented in Fig. 3 in the main text. The results for the tetrahedron coincide with the ones given in the main text as the chosen parameters a_0, a_1 minimizing the energy yield an accuracy below $\sigma/J = 0.03$. In Fig. 13 the results for the different Lanczos methods for the VQE results of H_2 generated on four qubits are depicted. Also in this case the results for the parameters a_0, a_1 achieving the exact minimum of Eq. (1) (blue) and the ones yielding a maximum standard error of 0.03 Ha (orange) almost coincide when the experimental estimates of H, H^2, H^3 are obtained with $n_{\text{shots}} = 8 \cdot 10^5$. Finally, Fig. 14 provides additional results for the comparison of the zero-noise extrapolation (ZNE) and the Lanczos error mitigation methods. In particular, we present the effect of increasing the total number of measurements in the ZNE approach.

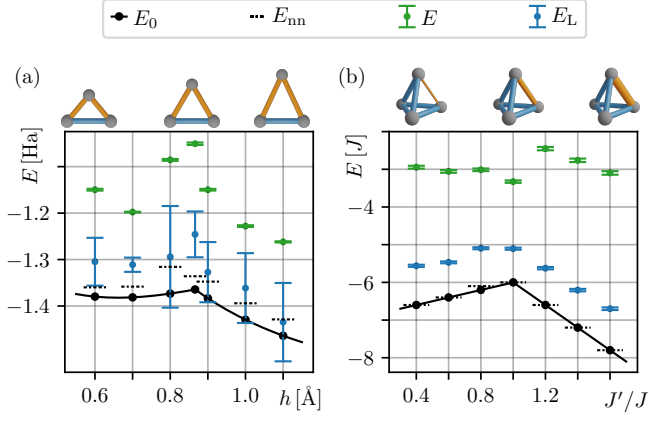


Figure 12: The results for the VQE algorithm for the Heisenberg tetrahedron and H_3 simulated on four qubits with `ibmq_ourense` and on six qubits with `ibmq_almaden` in dependence of the systems parameters are depicted in a) and b), respectively. The colors denote the same results as in Fig. 1 in the main text, blue are the results of the Lanczos algorithm and the results of the bare measurement of the Hamiltonian are shown in green. Black curves and dots denote the true ground state energies, dotted lines are VQE results optimized numerically and evaluated in the absence of noise. The $R_y R_z$ ansatz with 1 (3) entangling layers is used, with $n_{\text{shots}} = 8 \cdot 10^5$ (8192), $n_{\text{steps}} = 2000$, $n_{\text{VQE}} = 4$ (10), for H_3 (respectively the Heisenberg model).

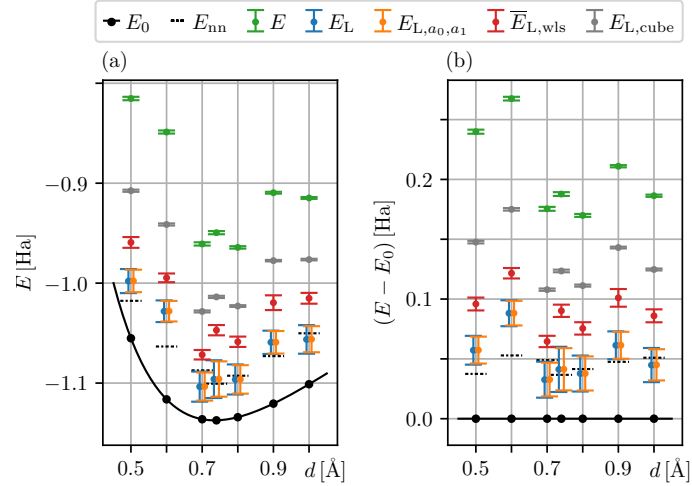


Figure 13: The results for the VQE algorithm for H_2 simulated on four qubits in dependence of the system configuration are depicted in (a) alongside with their accuracy in (b). The colors denote the same results as in Fig. 1 in the main text, blue are the results of the Lanczos algorithm, orange are the results of the fixed parameter method, gray are the results of the cube root method, red are the averaged results of taking the wls over five measurements, while the results of the bare measurement of the Hamiltonian are shown in green. Black denotes the true ground state energy and the dotted line reference to the non-noisy energy result obtained with the circuit optimized on the (noisy) quantum computer.

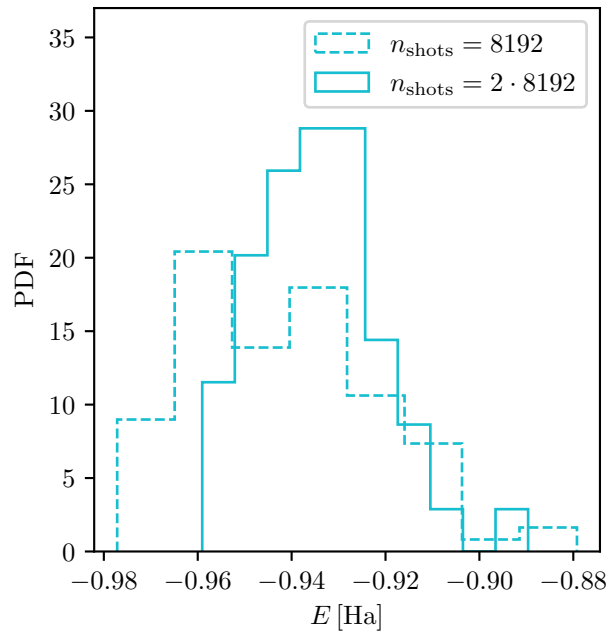


Figure 14: Effect of increasing the number of measurements budget in the zero-noise-extrapolation error mitigation scheme. Shown is the distribution of the final ground state energy estimations obtained with 50 repetitions of the protocol (dashed line). Each value is obtained extrapolating from 4 data points measured using 1, 3, 5 and 7 copies of each CZ gate. The histogram acquired with double the number of shots ($2 \cdot 8192$) is reported with a solid line.