



## UvA-DARE (Digital Academic Repository)

### Benchmarking noisy intermediate scale quantum error mitigation strategies for ground state preparation of the HCl molecule

Weaving, T.; Ralli, A.; Kirby, W.M.; Love, P.J.; Succi, S.; Coveney, P.V.

**DOI**

[10.1103/PHYSREVRESEARCH.5.043054](https://doi.org/10.1103/PHYSREVRESEARCH.5.043054)

**Publication date**

2023

**Document Version**

Final published version

**Published in**

Physical Review Research

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Weaving, T., Ralli, A., Kirby, W. M., Love, P. J., Succi, S., & Coveney, P. V. (2023). Benchmarking noisy intermediate scale quantum error mitigation strategies for ground state preparation of the HCl molecule. *Physical Review Research*, 5(4), Article 043054. <https://doi.org/10.1103/PHYSREVRESEARCH.5.043054>

**General rights**



It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*

## Benchmarking noisy intermediate scale quantum error mitigation strategies for ground state preparation of the HCl molecule

Tim Weaving <sup>1</sup>, Alexis Ralli,<sup>1</sup> William M. Kirby,<sup>2</sup> Peter J. Love,<sup>2,3</sup> Sauro Succi <sup>4,5,6</sup> and Peter V. Coveney<sup>1,7,8</sup>

<sup>1</sup>Centre for Computational Science, Department of Chemistry, University College London, WC1H 0AJ London, United Kingdom

<sup>2</sup>Department of Physics and Astronomy, Tufts University, Medford, Massachusetts 02155, USA

<sup>3</sup>Computational Science Initiative, Brookhaven National Laboratory, Upton, New York 11973, USA

<sup>4</sup>Center for Life Nano-Neuro Science @ La Sapienza, Italian Institute of Technology, 00161 Roma, Italy

<sup>5</sup>Department of Mechanical Engineering, University College London, WC1E 7JE London, United Kingdom

<sup>6</sup>Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

<sup>7</sup>Advanced Research Computing Centre, University College London, WC1H 0AJ London, United Kingdom

<sup>8</sup>Informatics Institute, University of Amsterdam, Amsterdam, 1098 XH, Netherlands



(Received 22 March 2023; accepted 14 September 2023; published 17 October 2023)

Due to numerous limitations including restrictive qubit topologies, short coherence times, and prohibitively high noise floors, few quantum chemistry experiments performed on existing noisy intermediate-scale quantum hardware have achieved the high bar of chemical precision, namely energy errors to within 1.6 mHa of full configuration interaction. To have any hope of doing so, we must layer contemporary resource reduction techniques with best-in-class error mitigation methods; in particular, we combine the techniques of qubit tapering and the contextual subspace variational quantum eigensolver with several error mitigation strategies comprised of measurement-error mitigation, symmetry verification, zero-noise extrapolation, and dual-state purification. We benchmark these strategies across a suite of eight 27-qubit IBM Falcon series quantum processors, taking preparation of the HCl molecule's ground state as our testbed.

DOI: [10.1103/PhysRevResearch.5.043054](https://doi.org/10.1103/PhysRevResearch.5.043054)

### I. INTRODUCTION

We find ourselves in the era of noisy intermediate-scale quantum (NISQ) computation, which is characterized by various obstacles including restrictive qubit topologies, short coherence times, and imperfect quantum gates; these factors compound to limit what is achievable using existing or near-term quantum devices. The development of quantum error mitigation (QEM) techniques has therefore been a necessary pursuit, aiming to extract usable data from the raw output of NISQ machines.

A plethora of techniques have been proposed that exploit various properties of noise in NISQ devices. Some evaluate collections of Clifford circuits (which are classically efficient to simulate) either to mitigate against measurement errors [1,2] or for learning-based approaches that aim to characterize the noise model [3–5]. Others average over the effect of noise by recompiling circuits at random [6–8] or make predictions informed by the character of the noise and its behavior under amplification [6–12]. There are also techniques that use problem-specific properties (e.g., known symmetries) to identify and discard invalid outcomes [13–15] and purification-based approaches that promote some pure component of the noisy states prepared in hardware [16–20]. We refer the reader to the works of Endo *et al.* [21] and Cai *et al.*

[22] for a comprehensive review of the QEM literature, and to Resch and Karpuzcu [23] for an exposition of noise sources in quantum computation.

While QEM has permitted some degree of success in obtaining usable results from NISQ computers, a number of works have cautioned that QEM may be restricted by some fundamentals limits [24–26]. With this in mind, it is not clear whether “quantum advantage” will be feasible using QEM alone, and we may still require partially error-corrected machines for this to be realized in practice.

In this work, we place an emphasis on *scalable* quantum error mitigation techniques for the NISQ era. As such, we benchmark the following:

- (i) Measurement-error mitigation (MEM)—IV B,
- (ii) non- $\mathbb{Z}_2$  symmetry verification (SV)—IV C,
- (iii) zero-noise extrapolation (ZNE)—IV D,
- (iv) dual-state purification (DSP)—IV E,
- (v) tomography purification (TP) applied to DSP,

including every possible combination given by the compatibility matrix in Fig. 1. For a fixed shot budget, we intend to identify which combined strategy is most effective in mitigating errors, executed across a suite of IBM quantum hardware.

The problem we take as a testbed for this QEM benchmark is preparation of the HCl molecule ground state, with the ultimate goal of measuring the corresponding energy to chemical precision (errors within 1.6 mHa of full configuration interaction). Of the numerous quantum chemistry experiments performed on NISQ hardware to date [10,27–48], only a select few have achieved this threshold; of those that have, most consist of hydrogen chains of varying size.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

	MEM	SV	ZNE	DSP	TP
MEM	✓	✓	✓	✓	✗
SV	✓	✓	✓	✗	✗
ZNE	✓	✓	✓	✓	✗
DSP	✓	✗	✓	✓	✓
TP	✗	✗	✗	✓	✓

FIG. 1. Compatibility matrix of the error mitigation techniques investigated in this work. Note that tomography purification is compatible with each of these techniques in principle, however it is not in general scalable due to the exponential number of bases one must measure to reconstruct the density matrix via full state tomography. We claim compatibility with dual-state purification since we need only apply it to a single ancilla qubit.

## II. THE HARDWARE

The IBM Quantum hardware is equipped with the universal gate set  $\{\text{CNOT}, R_z, X, \sqrt{X}\}$  and, at the time of writing, eight 27-qubit Falcon series quantum processors were available to us. From the point of view of gate errors and coherence, these devices are the most reliable available through IBM Quantum at present, with the greatest quantum volumes (QV) [23,49]; in Table I we provide a snapshot of the hardware specification at the point of execution of our Qiskit Runtime programs.

One way we may assess the quality of these devices is to evaluate quantum state fidelities for increasing numbers of qubits. Namely, we shall prepare the  $N$ -qubit Greenberger-

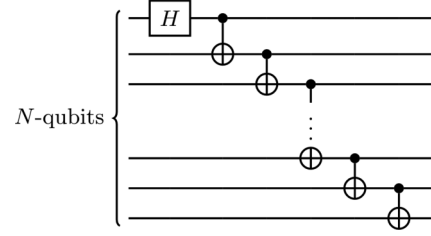


FIG. 2. The  $N$ -qubit GHZ circuit, consisting of one Hadamard and  $N - 1$  CNOT gates.

Horne-Zeilinger (GHZ) state

$$|\psi_N\rangle = (|0\rangle^{\otimes N} + |1\rangle^{\otimes N})/\sqrt{2} \quad (1)$$

via the circuit given in Fig. 2 and determine the fidelity

$$f(N) = \left| \langle \psi_N^{\text{true}} | \psi_N^{\text{noisy}} \rangle \right|^2 = \frac{1}{2}(\sqrt{p_0} + \sqrt{p_1})^2, \quad (2)$$

where  $\psi_N^{\text{noisy}}$  is the noisy state prepared on the hardware, and  $p_0, p_1$  are the probabilities with which we obtain the all  $\mathbf{0}$  or  $\mathbf{1}$  state, respectively.

In Fig. 3 we observe a decay in fidelity as more qubits are included in the GHZ state preparation procedure, with a sharp drop to near-zero fidelity at  $N = 22$ . This is due to the longest connected path of qubits being of length 21, given the chip topology of Fig. 4; beyond this point, we

TABLE I. Breakdown of quantum hardware specification restricted to the chosen qubit cluster at the point of executing the Qiskit Runtime programs. We provide the Quantum Volume (QV), chosen 5-qubit cluster, T1/T2 times and gate duration/error for entangling (CNOT), local ( $R_z, X, \sqrt{X}$ ), and readout operations.

	QV	Chosen 5q Cluster	Coherence		Gate Specification		
			Type	Time [ $\mu\text{S}$ ]	Type	Time [nS]	Error $\times 10^3$
ibmq_montreal	128	{0, 1, 2, 3, 4}	T1:	140.92 $\pm$ 16.77	Entangling:	471.11 $\pm$ 78.69	7.85 $\pm$ 1.06
			T2:	82.16 $\pm$ 39.10	Local:	35.56 $\pm$ 0.00	0.22 $\pm$ 0.03
					Readout:	5201.78 $\pm$ 0.00	14.08 $\pm$ 2.52
ibmq_kolkata	128	{16, 19, 20, 22, 25}	T1:	150.92 $\pm$ 16.80	Entangling:	348.44 $\pm$ 177.77	5.14 $\pm$ 0.65
			T2:	135.59 $\pm$ 66.36	Local:	35.56 $\pm$ 0.00	0.17 $\pm$ 0.04
					Readout:	675.56 $\pm$ 0.00	10.68 $\pm$ 1.99
ibmq_mumbai	128	{0, 1, 2, 3, 4}	T1:	129.80 $\pm$ 28.12	Entangling:	556.44 $\pm$ 136.77	8.63 $\pm$ 2.20
			T2:	104.20 $\pm$ 69.62	Local:	35.56 $\pm$ 0.00	0.31 $\pm$ 0.17
					Readout:	3552.00 $\pm$ 0.00	18.24 $\pm$ 0.91
ibm_hanoi	64	{0, 1, 2, 4, 7}	T1:	135.29 $\pm$ 54.61	Entangling:	270.67 $\pm$ 59.21	6.99 $\pm$ 2.65
			T2:	174.89 $\pm$ 83.31	Local:	32.00 $\pm$ 0.00	0.21 $\pm$ 0.10
					Readout:	817.78 $\pm$ 0.00	8.94 $\pm$ 1.59
ibm_cairo	64	{8, 11, 13, 14, 16}	T1:	95.59 $\pm$ 37.85	Entangling:	462.22 $\pm$ 296.50	9.17 $\pm$ 4.30
			T2:	92.76 $\pm$ 69.45	Local:	24.89 $\pm$ 0.00	0.22 $\pm$ 0.06
					Readout:	732.44 $\pm$ 0.00	20.86 $\pm$ 11.74
ibm_auckland	64	{8, 11, 13, 14, 16}	T1:	162.99 $\pm$ 73.58	Entangling:	376.89 $\pm$ 28.61	6.28 $\pm$ 1.22
			T2:	123.67 $\pm$ 72.80	Local:	35.56 $\pm$ 0.00	0.23 $\pm$ 0.03
					Readout:	757.33 $\pm$ 0.00	8.30 $\pm$ 1.55
ibmq_toronto	32	{5, 8, 9, 11, 14}	T1:	113.96 $\pm$ 6.53	Entangling:	382.22 $\pm$ 61.35	7.96 $\pm$ 0.83
			T2:	171.38 $\pm$ 19.39	Local:	35.56 $\pm$ 0.00	0.28 $\pm$ 0.06
					Readout:	5962.67 $\pm$ 0.00	12.04 $\pm$ 4.30
ibm_geneva	32	{5, 8, 9, 11, 14}	T1:	311.27 $\pm$ 101.19	Entangling:	586.67 $\pm$ 110.22	3.97 $\pm$ 0.31
			T2:	300.56 $\pm$ 127.31	Local:	38.40 $\pm$ 5.69	0.20 $\pm$ 0.17
					Readout:	1600.00 $\pm$ 0.00	30.26 $\pm$ 13.08

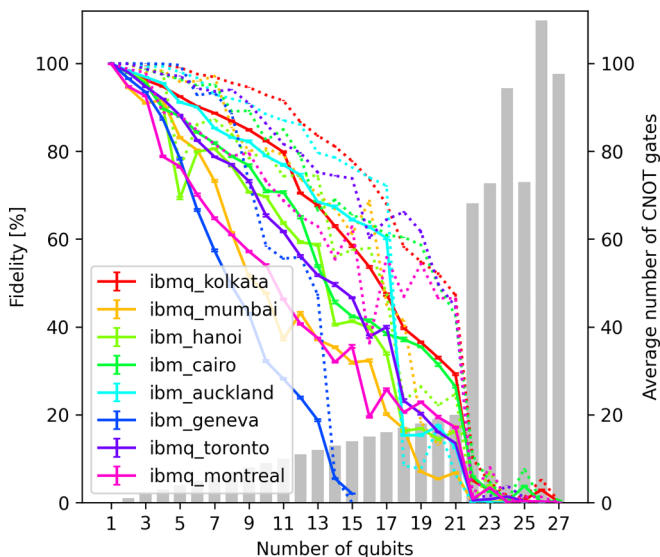


FIG. 3. Decay in quantum state fidelity against the number of qubits for GHZ preparation; dotted lines indicate the measurement-error mitigated result. We could not utilize more than 15-qubits on *ibmq\_geneva* due to a faulty qubit. The gray bars indicate the average number of CNOT gates required to prepare the relevant  $N$ -qubit GHZ state, with differences arising between chips due to the transpiler.

incur expensive SWAP operations that rapidly consume the remaining fidelity, indicated by the dramatic jump in the number of CNOT gates from 22-qubit onwards. We also include

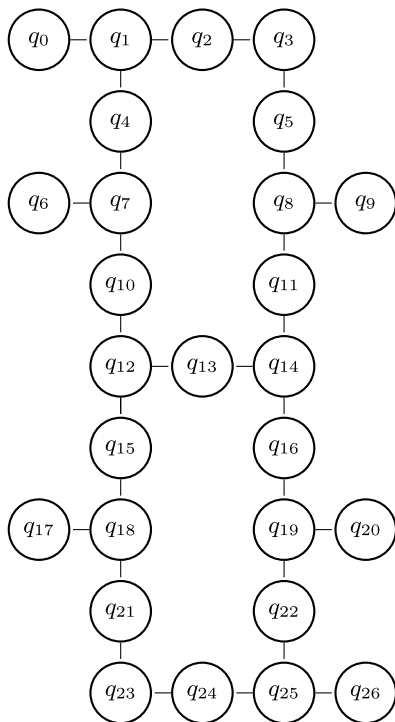


FIG. 4. The IBM *Falcon* series 27-qubit chip “heavy-hex” topology. For our quantum simulations, we identified optimal qubit clusters by assigning scores based on gate and readout errors. For DSP we require 5-qubit clusters of the form given in Fig. 11 to facilitate every readout configuration; the specific clusters used are detailed in Table I.

the effect of measurement-error mitigation on the fidelity, and note that we are able to recover approximately 10–20% fidelity in most cases.

### III. QUBIT REDUCTION TECHNIQUES

Taken in the minimal STO-3G basis, the full HCl problem consists of 20 qubits, and therefore direct treatment is not yet feasible on current quantum computers. For the hardware to accommodate our problem, we layered the qubit reduction techniques of tapering [50,51] (Sec. III A) and contextual subspace [52–54] (Sec. III B) to yield a dramatically condensed 3-qubit Hamiltonian

$$H = \sum_i h_i P_i, \tag{3}$$

where we provide the explicit coefficients  $h_i \in \mathbb{R}$  and Pauli terms  $P_i = q_0^{(i)} \otimes q_1^{(i)} \otimes q_2^{(i)}$  in Table II. The exact ground state energy of this Hamiltonian lies within 0.837 mHa of the full configuration interaction (FCI) energy (−455.157 Ha, calculated using PySCF [55]); this is nearly half what is generally considered chemical precision (1.6 mHa), although we stress that, due to the minimal basis set used here, one should not expect agreement with experimentally obtained energy values. Subtracting the relatively large identity term leaves a target energy of −2.066 Ha; with respect to chemical precision, this represents a challenging 0.077% error ratio that we aim to capture via QEM.

Due to incompatibility with some of the error mitigation techniques investigated here, we do not implement any measurement reduction strategies such as (qubit-wise) commuting decompositions or unitary partitioning [56,57]. Instead, each Hamiltonian term is treated independently so there is zero covariance between expectation value estimates, and the overall variance is therefore obtained as

$$\text{var}(H) = \sum_i h_i^2 \text{var}(P_i); \tag{4}$$

the statistical analysis is conducted with a bootstrapping of the raw quantum measurement data.

#### A. Qubit tapering

Tapering allows one to map Hamiltonian  $\mathbb{Z}_2$  symmetries onto distinct qubits and consequently project over them, thus reducing the effective dimension of the problem. This works by identifying an independent set of Pauli operators  $\mathcal{S} \subset \mathcal{P}_N$  such that  $[S, T] = 0 \forall S \in \mathcal{S}, T \in \mathcal{T}$ , which we refer to as *symmetry generators* and can be identified efficiently using the *Symmer* Python package [58]. Assuming the elements of  $\mathcal{S}$  commute among themselves (if not, select the largest commuting subset within), one may perform a Clifford rotation mapping each symmetry to a distinct qubit position and consequently project onto the corresponding stabilizer subspace; under this procedure, it is possible to remove  $|\mathcal{S}|$  qubits from the Hamiltonian while remaining isospectral.

Since this is a fermionic system, we are guaranteed a reduction of at least two qubits arising from the preservation of spin-up/-down parities; under the Jordan-Wigner mapping [59] these manifest as  $S_{\text{up/down}} = Z^{\otimes \mathcal{I}_{\text{up/down}}}$ , where the sets

TABLE II. The 3-qubit contextual subspace HCl Hamiltonian, terms ordered by coefficient magnitude, we take as a testbed for the error mitigation strategies investigated in this work.

Index	$q_0$	$q_1$	$q_2$	Coefficient	Index	$q_0$	$q_1$	$q_2$	Coefficient
0	I	I	I	-453.090742	17	Y	Y	X	0.035219
1	I	Z	Z	0.846721	18	I	I	X	-0.015458
2	Z	I	Z	0.846721	19	I	Z	X	0.015458
3	I	Z	I	0.620754	20	Z	I	X	0.015458
4	Z	I	I	0.620754	21	Z	Z	X	-0.015458
5	I	I	Z	0.393828	22	I	X	X	-0.009644
6	Z	Z	I	0.258369	23	I	Y	Y	-0.009644
7	Z	Z	Z	0.238049	24	Z	X	X	0.009644
8	X	Z	I	-0.061959	25	Z	Y	Y	0.009644
9	Z	X	I	0.061959	26	X	I	X	0.009644
10	Z	X	Z	-0.061959	27	X	Z	X	-0.009644
11	X	Z	Z	0.061959	28	Y	I	Y	0.009644
12	Y	Y	I	-0.055599	29	Y	Z	Y	-0.009644
13	Y	Y	Z	0.055599	30	I	X	I	0.004504
14	X	X	X	-0.035219	31	I	X	Z	-0.004504
15	X	Y	Y	-0.035219	32	X	I	I	-0.004504
16	Y	X	Y	-0.035219	33	X	I	Z	0.004504

$\mathcal{I}_{\text{up}}, \mathcal{I}_{\text{down}}$  index qubit positions encoding up ( $\alpha$ ), down ( $\beta$ ) electron spin orbitals, respectively. These spin-parity operators are still  $\mathbb{Z}_2$  symmetries (i.e., single-Pauli terms) under the Bravyi-Kitaev mapping [60], however their closed form is less convenient since individual qubits do not represent distinct spin-orbitals. For our particular formulation of the 20-qubit HCl system with even (odd) indices encoding spin-up (-down) electrons, we have

$$\begin{aligned} S_{\text{up}} &= ZIZIZIZIZIZIZIZIZI, \\ S_{\text{down}} &= IZIZIZIZIZIZIZIZI. \end{aligned} \quad (5)$$

We also identified two additional  $\mathbb{Z}_2$  symmetries

$$\begin{aligned} S_{\sigma_h} &= IIIIIIZZIIIIIZZII, \\ S_{C_2} &= IIIIZZZZIIIIZZZZII, \end{aligned} \quad (6)$$

that arise from the Abelian subgroup  $C_{2v}$  of the non-Abelian point group  $C_{\infty v}$  (to which all heteronuclear diatomic molecules belong) generated by reflections along the molecular plane ( $\sigma_h$  symmetry) and rotations through an angle of  $180^\circ$  ( $C_2$  symmetry). In all, with the symmetry generating set  $\mathcal{S} = \{S_{\sigma_h}, S_{C_2}, S_{\text{up}}, S_{\text{down}}\}$ , qubit tapering permits a reduction of 20 to 16 qubits while exactly preserving the energy spectrum.

### B. Contextual subspace

Whereas tapering exploits physical symmetries of the Hamiltonian to remove redundant qubits, it is possible to achieve further reductions by imposing pseudosymmetries on the system. This is the contextual subspace approach [52–54] in which we partition the Hamiltonian into noncontextual and contextual components; the former may be mapped onto a classical optimization problem, whereas the latter yields quantum corrections obtained via some eigenvalue-finding algorithm (VQE, QPE etc.). The qubit reduction is effected by enforcing noncontextual symmetries on the contextual Hamil-

tonian, thus ensuring any quantum corrections are consistent with the noncontextual ground state configuration.

The choice of which noncontextual symmetries to enforce is highly nontrivial. Here, we select stabilizers that preserve commutativity with the most dominant coupled-cluster amplitudes, thus maximizing variational flexibility in the contextual subspace. Using this heuristic, we are able to project onto a 3-qubit contextual subspace that permits chemical precision. In Table II we provide explicit details of the corresponding Hamiltonian, whose ground state energy has absolute error 0.837 mHa with respect to the FCI energy.

This dramatic reduction in qubit resource is likely due to CCSD being near exact (we obtained an error of  $3.403 \times 10^{-8}$  Ha with respect to FCI, five orders of magnitude below chemical precision), as there are just two unoccupied spin-orbitals in the minimal STO-3G basis set and therefore excitations above doubles are not possible.

## IV. ERROR MITIGATION

In this section, we review the technical aspects of each quantum error mitigation (QEM) technique investigated through our benchmark, the results of which are discussed later in Sec. V E.

### A. Estimators

The language we shall use to describe our QEM techniques is that of *estimators*. Suppose that we are interested in some observable  $O$  (a Hermitian operator, i.e.,  $O^\dagger = O$ ) and have access to a general quantum state  $\rho$ ; we wish to estimate the quantity  $\text{Tr}(\rho O)$ , but may only probe the state via some finite sample of quantum measurements  $\mathcal{M} = \{m_i\}_{i=1}^M$ , where  $m_i \in \mathbb{Z}_2^N$ . The way in which we collect and subsequently combine our sample to approximate the desired observable property defines an estimator  $\mathcal{E} : \mathcal{M} \rightarrow \mathbb{R}$ ; the goal of QEM is to construct effective estimators that are capable of suppressing errors and extracting some usable data from the noise.

For example, we may define a naive estimator for the expectation value of a Pauli operator  $P \in \mathcal{P}_N$ . Given a pure quantum state  $|\psi\rangle$ , we may sample from the quantum device in a compatible basis (i.e., one that commutes with  $P$ ) and obtain eigenstates  $|m_i\rangle$  such that  $P|m_i\rangle = m_i|m_i\rangle$ , where  $m_i = \pm 1$  to estimate the expectation value  $\langle P \rangle_\psi := \langle \psi | P | \psi \rangle$ . The raw estimator is

$$\mathcal{E}_{\text{RAW}}^P(\mathcal{M}) = \frac{1}{M} \sum_{i=1}^M m_i \rightarrow \langle P \rangle_\psi \quad (M \rightarrow \infty). \quad (7)$$

Since any Hermitian operator may be decomposed as  $O = \sum_P o_P P$  with  $o_P \in \mathbb{R}$ , this allows us to extend our estimator to the full observable by linearity,

$$\mathcal{E}_{\text{RAW}} = \sum_P o_P \mathcal{E}_{\text{RAW}}^P, \quad (8)$$

which shall form a baseline for our QEM benchmark.

We shall use the following metrics to assess the efficacy of QEM techniques:

$$\begin{aligned} \text{var}(\mathcal{E}) &= \mathbb{E}(\mathcal{E}^2) - \mathbb{E}(\mathcal{E})^2, \\ \text{bias}(\mathcal{E}) &= \mathbb{E}(\mathcal{E} - \langle O \rangle_\psi) \end{aligned} \quad (9)$$

and the related quantity

$$\begin{aligned} \text{MSE}(\mathcal{E}) &= \mathbb{E}((\mathcal{E} - \langle O \rangle_\psi)^2) \\ &= \text{var}(\mathcal{E}) + \text{bias}(\mathcal{E})^2, \end{aligned} \quad (10)$$

or mean-squared error. Taking  $O = H$  and  $|\psi\rangle$  the ground state of  $H$ , our objective is to approximate  $\mathbb{E}(\mathcal{E}) \approx \langle \psi | H | \psi \rangle = \langle H \rangle_\psi = E_{\text{FCI}}$ . The goal of QEM is to reduce bias as far as possible [ideally within the threshold of chemical precision, i.e.,  $|\text{bias}(\mathcal{E}_{\text{QEM}})| < 1.6$  mHa] while aiming not to amplify variance severely.

Although it would be preferable to run multiple instances of each quantum simulation to evaluate  $\mathbb{E}(\mathcal{E})$ , this is not feasible given the length of time taken to produce each energy estimate. Instead, we rely on the statistical tool of *bootstrapping*, introduced in further detail in Appendix, whereby we generate resampled data from the empirical measurement outcomes.

### B. Measurement-error mitigation

Measurement-error mitigation (MEM) aims to characterize the errors incurred during the readout phase of a quantum experiment [1]; it treats the state preparation itself as a black box and does not consider errors that occur prior to measurement.

A naive, non-scalable approach to MEM is to prepare-and-measure each of the  $2^N$  basis states individually; given some  $|b_i\rangle$  with  $\mathbf{b}_i \in \mathbb{Z}_2^N$  we perform measurements to obtain a noisy distribution of binary outcomes  $|\mu_{\text{noisy}}^{(i)}\rangle = \sum_j p_{i,j} |\mathbf{b}_j\rangle$ , where  $p_{i,j} = \langle \mathbf{b}_i | A | \mathbf{b}_j \rangle$  denotes the probability of preparing the state  $|\mathbf{b}_i\rangle$  and measuring  $|\mathbf{b}_j\rangle$ . The doubly stochastic matrix  $A = \sum_{i,j} p_{i,j} |\mathbf{b}_j\rangle \langle \mathbf{b}_i|$  is referred to as the assignment (or transition) matrix and lies at the core of this technique.

Now, suppose we wish to implement a circuit with noiseless measurement output  $|\mu_{\text{ideal}}\rangle = \sum_i m_i |\mathbf{b}_i\rangle$ ; since

$A|\mathbf{b}_i\rangle = |\mu_{\text{noisy}}^{(i)}\rangle$ , then by linearity we have

$$A|\mu_{\text{ideal}}\rangle = \sum_i m_i |\mu_{\text{noisy}}^{(i)}\rangle =: |\mu_{\text{noisy}}\rangle. \quad (11)$$

More realistically, what we will actually have access to is  $|\mu_{\text{noisy}}\rangle$ , the output from some quantum experiment. Therefore, by inverting the assignment matrix we obtain a measurement-error mitigated distribution  $|\mu_{\text{ideal}}\rangle = A^{-1}|\mu_{\text{noisy}}\rangle$ .

In its current form, it will not be possible to construct the assignment matrix for large numbers of qubits. The “tensoring” approach of Nation *et al.* [2] is designed to assess the *qubit-wise* measurement assignment error, namely evaluating the probability  $p_k$  that qubit  $k$  is erroneously flipped  $|0\rangle \rightleftharpoons |1\rangle$ . The single-qubit assignment matrix for this process is

$$A^{(k)} = \begin{matrix} & |0\rangle & |1\rangle \\ \begin{matrix} |0\rangle \\ |1\rangle \end{matrix} & \begin{bmatrix} 1 - p_k & p_k \\ p_k & 1 - p_k \end{bmatrix} \end{matrix} \quad (12)$$

and we subsequently reconstruct the full  $N$ -qubit assignment error probability by taking products over the relevant single-qubit transitions,

$$A_{i,j} \approx \prod_{k=0}^{N-1} A_{(\mathbf{b}_i)_k, (\mathbf{b}_j)_k}^{(k)}. \quad (13)$$

This expression makes some strong assumptions on the character of the readout errors, in particular that they are predominantly uncorrelated. On the IBM Quantum hardware, Nation *et al.* found this to be a reasonable assumption (using *ibmq\_kolkata*), with little difference observed between this tensoring approach versus a complete measurement calibration until inducing correlations by increasing the readout pulse amplitudes from their optimized values [2].

The expression of  $A$  in terms of single-qubit readout errors (13) requires just  $2N$  quantum experiments to be carried out, versus  $2^N$  in a complete measurement calibration. Furthermore, its form is particularly convenient as it is amenable to matrix-free iterative linear algebra techniques [61]. The Python package *mithree* developed through the work of Nation *et al.* is available in Qiskit; we utilized this for our QEM benchmark, and it is the only technique presented here that we did not implement ourselves. In Fig. 5 we present the measurement distribution pre- and post-MEM for a 21-qubit GHZ preparation procedure on *ibmq\_kolkata*, recalling from Fig. 3 that we observed an increase from 29.3% to 47.5% in GHZ state fidelity. The effect of  $T_1$  relaxation is also visible in this plot, whereby the  $|0\rangle$  state occurs with considerably greater probability than  $|1\rangle$  since the former is energetically favorable.

### C. Symmetry verification

An inexpensive method of error mitigation is to take known symmetries of the Hamiltonian (usually those of the  $\mathbb{Z}_2$  variety, i.e., Pauli operators that commute termwise across the Hamiltonian) and enforce stabilizer constraints on the measured binary strings resulting from a quantum experiment; we shall refer to this as *symmetry verification* (SV) [13–15]. On the other hand, in Sec. III we described how those same

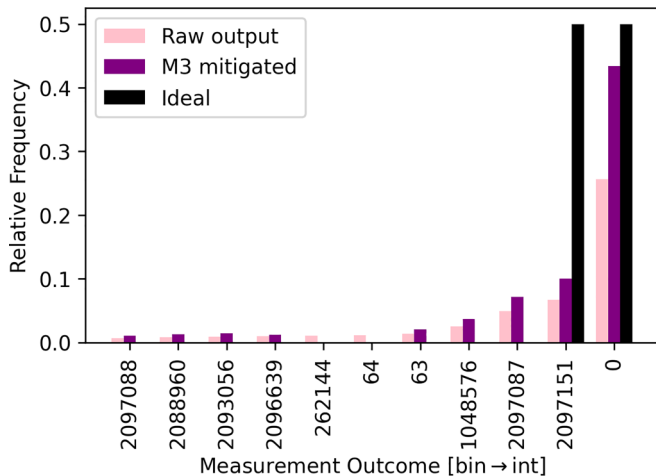


FIG. 5. Comparing raw and MEM measurement distributions against the ideal output for 21-qubit GHZ preparation on *ibmq\_kolkata* (the greatest number of qubits possible without SWAP operations) with  $2^{15}$  circuit shots. Only outcomes exceeding a frequency of  $10^{-2}$  are plotted here; this contributes 46.8% and 71.4% of the raw and MEM distributions, respectively.  $T_1$  relaxation results in a reduced frequency of  $|1\rangle$  measurement outcomes compared with  $|0\rangle$  whereas they should be observed with equal probability 50%.

symmetries may instead be utilized for the purposes of qubit reduction, which allowed us to dramatically reduce the dimension of our Hamiltonian. In doing so, we may no longer use  $\mathbb{Z}_2$  symmetries to postselect allowed measurement outcomes as the reduced Hamiltonian has been abstracted from them. However, there still exist symmetries of a more general nature that need not commute with each term individually, but do so with respect to the full Hamiltonian. Examples in the setting of electronic structure are the (Jordan-Wigner encoded) particle and spin quantum number operators

$$S_N = \frac{1}{2} \left( N \times I^{\otimes N} - \sum_{i=1}^N Z_i \right), \quad S_z = \frac{1}{4} \sum_{i=1}^N (-1)^i Z_i; \quad (14)$$

note how the latter differs from the up/down spin-parity operators of (5). These are not  $\mathbb{Z}_2$  symmetries as they do not commute with individual terms in the Hamiltonian and are therefore nontrivial in the contextual subspace; the projection procedure respects commutation and therefore we may use the reduced operators

$$S_N = 17 \times III - IIZ - \frac{1}{2}(IZI + IZZ + ZII + ZIZ), \quad (15)$$

$$S_z = \frac{1}{4}(IZI + IZZ - ZII - ZIZ)$$

for error mitigation in our HCl 3-qubit contextual subspace— as an exercise, we suggest the reader confirms that these operators do indeed commute with the Hamiltonian described by the terms in Table II. An interesting feature of this reduced  $S_N$  operator is the identity term that was not present in the original formulation of the number operator in (14); the coefficient indicates the number of particles that have been effectively enforced in the contextual subspace, in this case 17 out of the 18 available electrons. In the qubit tapering setting of Sec. III A we were able to enforce  $\mathbb{Z}_2$  symmetry perfectly, thus removing the corresponding degrees of free-

dom from the resulting Hamiltonian; in the non- $\mathbb{Z}_2$  setting we have only treated the symmetry approximately, given that the corresponding symmetry operator (15) is nontrivial. In the instance that a non- $\mathbb{Z}_2$  symmetry operator is projected onto the identity under the contextual subspace procedure, this indicates that the symmetry is fixed in the resulting subspace and therefore need not be accounted for in the construction of an ansatz circuit. The rotations involved in the projection procedure abstract the reduced system from the underlying physical system, however this observation suggests there may be some natural interpretation of the contextual subspace method, which would be an interesting pursuit for further research.

An important point is that we may only mitigate errors of terms that commute with the number and spin operators, which, in this case, means only the diagonal ones; this may still yield significant improvements in error since these terms have the greatest coefficient magnitude, and errors here will be amplified proportionally.

Given an ensemble of measurements  $\{\mathbf{b}\}$ , we discard any binary strings  $\mathbf{b} \in \mathbb{Z}_2^N$  that do not respect the number and spin symmetries; given that we know the number of particles  $n$  in the system and the allowed spin values  $\{s_0, \dots, s_{M-1}\}$ , where  $s_i = s - i$  for quantum number  $s$  (multiplicity  $M = 2s + 1$ ), we require that  $S_N|\mathbf{b}\rangle = n|\mathbf{b}\rangle$  and  $S_z|\mathbf{b}\rangle = s_i|\mathbf{b}\rangle$  for some  $i \in \{0, \dots, M - 1\}$ . Our HCl problem is in a singlet configuration, hence the only allowable spin value is  $s = 0$  and thus valid quantum measurements are those in the kernel of  $S_z$ .

This QEM technique requires no additional coherent overhead and only minor postprocessing, yet we observe respectable error suppression from enforcing number and spin symmetries on the diagonal Hamiltonian terms, as seen in Table IV. We intend to investigate the use of non-Abelian point group symmetries (see Sec. III A) for the purposes of error mitigation in future work, although it is not immediately clear whether this will be possible.

#### D. Zero-noise extrapolation

The technique of *zero-noise extrapolation* (ZNE), also referred to in the literature as *Richardson extrapolation*, operates on the principle that one may methodically amplify noise present in our quantum measurement output, obtaining a collection of increasingly noisy energy estimates before extrapolating the data and inferring the experimentally un-touchable point of “zero noise” [6–12]. There are many methods of amplifying noise in our quantum circuits: some do so continuously by stretching gates temporally, requiring pulse-level control over the hardware, whereas others employ discrete approaches that either insert identity blocks of increasing complexity (e.g., unitary folding) or replace the target gate with a product over its roots.

It is the latter method we employ here. Given a quantum circuit  $U$ , some constituent native gate  $G$ , and a noise parameter  $\lambda \in \mathbb{N}$ , we shall replace each instance of  $G$  in-circuit with the equivalent operation  $\prod_{i=1}^{\lambda} \sqrt[i]{G}$  to yield a noise-amplified circuit  $U_\lambda$ . One may note that  $\lambda = 1$  corresponds with the unmodified circuit, whereas we intend to infer a value for  $\lambda = 0$  by evaluating expectation values  $E_\lambda = \langle \psi_{\text{ref}} | U_\lambda^\dagger P U_\lambda | \psi_{\text{ref}} \rangle$  at integer values  $\lambda \in \{1, 2, 3, \dots\}$  and extrapolating.

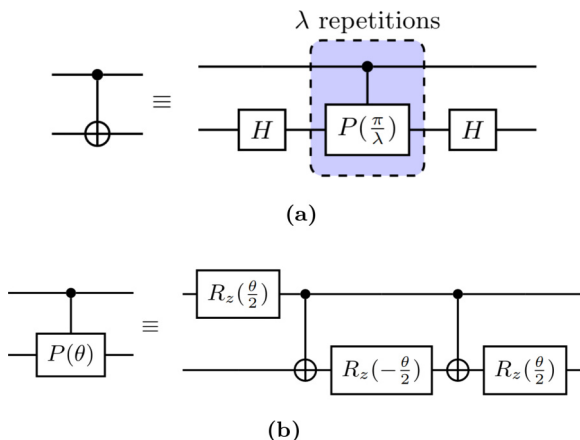


FIG. 6. Noise amplification method used for zero-noise extrapolation. Given a noise amplification factor  $\lambda \in \mathbb{N}$ , each CNOT is replaced by  $2\lambda$  CNOTs,  $3\lambda$  single-qubit Z-rotations, and two Hadamard gates. (a) Each CNOT gate may be replaced in-circuit with a product over its roots, namely  $\sqrt[\lambda]{\text{CNOT}} = H_1 \text{CPhase}(\frac{\pi}{\lambda}) H_1$ . (b) Since the IBM hardware takes the CNOT as its native entangling gate, the CPhase decomposition of (a) is transpiled back in terms of CNOTs at the point of execution.

In particular, we shall take  $G = \text{CNOT}$  since this is the dominant source of error by an order of magnitude, as seen in Table I. To decompose CNOT into its roots, we define the two-qubit gate

$$\begin{aligned} \text{CPhase}(\theta) &= \frac{1}{2}[(1 + Z) \otimes I + (1 - Z) \otimes P(\theta)] \\ &= \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & P(\theta) \end{bmatrix}, \end{aligned} \tag{16}$$

where  $P(\theta) := e^{i\theta/2} R_z(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix}$  and note that  $\text{CNOT} = H_1 \times \text{CPhase}(\pi) \times H_1$ . In other words, the Hadamard gates applied on the target qubit diagonalize the CNOT gate and thus

$$\sqrt[\lambda]{\text{CNOT}} = H_1 \times \text{CPhase}\left(\frac{\pi}{\lambda}\right) \times H_1. \tag{17}$$

The CNOT root-product decomposition is given as a circuit in Fig. 6(a). When it comes down to implementation of zero-noise extrapolation on a quantum computer, one must be mindful of which gates are native to said device and should avoid circuit optimization routines since these may result in an unpredictable scaling of noise. For example, as stated in Sec. II, the CNOT is in fact the native entangling gate on IBM Quantum systems; therefore, CPhase operations will be transpiled back in terms of CNOT and  $R_z$  gates at the point of execution, the decomposition of which is given in Fig. 6(b). Such considerations can wreak havoc on zero-noise extrapolation if not controlled carefully.

For our specific implementation of ZNE, we shall assume that the individual noise amplified estimates have been obtained via an estimator  $\mathcal{E}_\lambda$  so that  $E_\lambda = \mathbb{E}(\mathcal{E}_\lambda)$ , which might have previously had some other QEM strategy applied. We shall then evaluate estimates for  $\lambda \in \{1, 2, 3, 4\}$  before performing weighted least-squares (WLS) regression

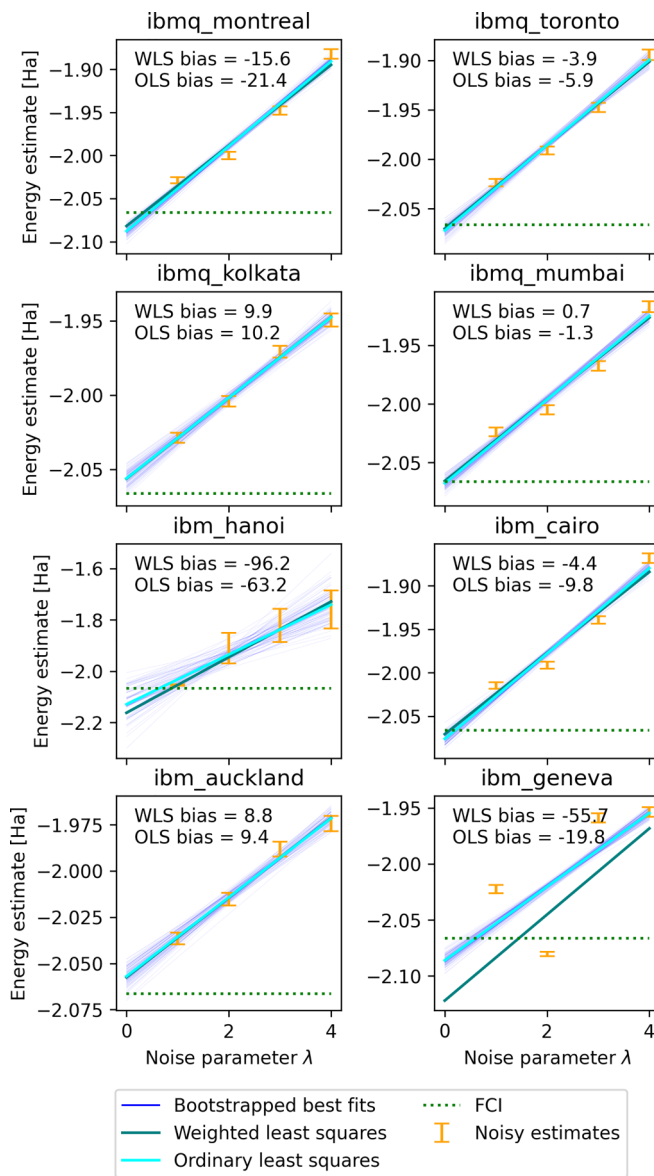


FIG. 7. Zero-noise extrapolation of our HCl problem, comparing weighted and ordinary least squares in addition to possible bootstrapped fits. Each of the noisy estimates has had measurement-error mitigation and symmetry verification applied.

with weights  $w_\lambda = 1/\text{var}(\mathcal{E}_\lambda)$  to infer a “zero-noise” estimate  $\mathcal{E}_{\text{ZNE}} = E_0$ . This penalizes highly varying points in the extrapolation; in Fig. 7 we compare WLS against ordinary least squares (OLS) and a bootstrapped collection of possible ZNE curves. We note that such a regression approach allows us to quantify the success of our extrapolation via the coefficient of determination, or  $R^2$  value, expressed as a ratio of residual and total sum of squares [62]. WLS yields the smallest bias in all but two cases: *ibmq\_hanoi* and *ibmq\_geneva*. In the former we have a low-variance, low-bias point at  $\lambda = 1$  that is pinning the extrapolation while the noisier estimates vary dramatically, whereas the latter exhibits a problematic low-variance, negatively biased point at  $\lambda = 2$  that is causing the extrapolation to fail.



### E. Dual-state purification

Purification-based error mitigation techniques operate on the basis that in quantum computation we are often interested in preparing some pure state  $|\psi_0\rangle = U|\mathbf{0}\rangle$ , corresponding with a channel

$$\rho = \mathcal{U}(|\mathbf{0}\rangle\langle\mathbf{0}|) = U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger = |\psi_0\rangle\langle\psi_0|, \quad (18)$$

whereas in reality what is actually prepared on the noisy quantum hardware is some mixed state

$$\rho = \sum_{i=0}^{2^N-1} \lambda_i \rho_i, \quad (19)$$

where  $\rho_i = |\psi_i\rangle\langle\psi_i|$  and we assume  $\lambda_i > \lambda_j$  for  $i < j$ . The central observation that purification-based methods exploit is

$$\rho^M / \text{Tr}(\rho^M) \rightarrow \rho_0 \quad (M \rightarrow \infty), \quad (20)$$

and the convergence is exponentially fast. This is precisely the formulation of *virtual distillation* [16], in which one prepares  $M$  copies of the mixed state  $\rho$  over disjoint quantum registers and induces their product via application of a cyclic shift operator. However, this permutation circuit is expensive and not feasible for near-term applications; the error mitigation technique we investigate here—*dual state purification* (DSP), also referred to in the literature as *echo verification* [22]—is closely related but may be implemented at significantly reduced cost. While the technique was first presented in the context of quantum phase estimation (QPE) [17], it was subsequently extended to the NISQ era [18,20]. The idea behind this method is that one prepares some quantum state, performs an intermediary readout, and subsequently uncomputes the circuit before postselecting on zero measurement outcomes; this bears some resemblance to second-order virtual distillation ( $M = 2$ ) but with the state

$$(\rho\bar{\rho} + \bar{\rho}\rho) / 2\text{Tr}(\rho\bar{\rho}) \quad (21)$$

as opposed to the form given in (20), with the dual state given as  $\bar{\rho} = \mathcal{V}(|\mathbf{0}\rangle\langle\mathbf{0}|) := U^\dagger|\mathbf{0}\rangle\langle\mathbf{0}|U$  [20].

We now describe explicitly the steps one must follow to implement DSP. The setting is that of a Pauli operator  $P \in \mathcal{P}_N$  whose expectation value we wish to evaluate with respect to an  $N$ -qubit state  $|\psi\rangle = U|\mathbf{0}\rangle$ . Denoting by  $\mathcal{I}$  the set of nonidentity qubit indices, we may identify a change-of-basis operator  $B$  such that  $BPB^\dagger = Z_{\mathcal{I}} = \otimes_{i \in \mathcal{I}} Z_i$  defined as

$$B_i = \begin{cases} I, & P_i \in \{I, Z\}, \\ H, & P_i = X, \\ HS, & P_i = Y. \end{cases} \quad (22)$$

Now, we note the effect of applying a CNOT gate controlled on a qubit position  $i \in \mathcal{I}$  to an ancilla register. With the expression

$$\text{CNOT} = \frac{1}{2}[(I+Z) \otimes I + (I-Z) \otimes X] \quad (23)$$

we observe

$$\begin{aligned} \text{CNOT}_{i,a}(|\psi\rangle \otimes |0\rangle_a) &= \frac{1}{2}(|\psi\rangle \otimes |0\rangle_a + Z_i|\psi\rangle \otimes |0\rangle_a \\ &\quad + |\psi\rangle \otimes |1\rangle_a - Z_i|\psi\rangle \otimes |1\rangle_a) \\ &= \frac{1}{\sqrt{2}}(|\psi\rangle \otimes |+\rangle_a + Z_i|\psi\rangle \otimes |-\rangle_a). \end{aligned} \quad (24)$$

Finally, as demonstrated by Huo and Li [20], we may uncompute the circuit  $U$  that prepares  $|\psi\rangle$  and postselect on measurement outcomes  $|\mathbf{0}\rangle$ , occurring with probability  $p_0$ , to drive the ancilla register into the state  $\frac{1}{\sqrt{2p_0}}(|+\rangle_a + \langle\psi|Z_i|\psi\rangle|-\rangle_a)$ .

We now describe the full process of computing the expectation value  $\langle P \rangle_\psi$ . First of all, the circuit is initialized in the state

$$|\psi_0\rangle = |\mathbf{0}\rangle \otimes |0\rangle_a \quad (25)$$

before applying the unitary  $U$  and basis transformation  $B$  supported on some qubit subset  $\mathcal{I}$ :

$$\begin{aligned} |\psi_1\rangle &= (BU \otimes I)|\psi_0\rangle \\ &= BU|\mathbf{0}\rangle \otimes |0\rangle. \end{aligned} \quad (26)$$

We now compute and store the parity of qubits  $\mathcal{I}$  on the ancilla register:

$$\begin{aligned} |\psi_2\rangle &= \prod_{i \in \mathcal{I}} \text{CNOT}_{i,a} |\psi_1\rangle \\ &= \frac{1}{\sqrt{2}}(BU|\mathbf{0}\rangle \otimes |+\rangle_a + Z_{\mathcal{I}}BU|\mathbf{0}\rangle \otimes |-\rangle_a). \end{aligned} \quad (27)$$

Inverting the change-of-basis and unitary circuit, we obtain

$$\begin{aligned} |\psi_3\rangle &= (U^\dagger B^\dagger \otimes I)|\psi_2\rangle \\ &= \frac{1}{\sqrt{2}}(|\mathbf{0}\rangle \otimes |+\rangle_a + U^\dagger \underbrace{B^\dagger Z_{\mathcal{I}} B}_{=P} U|\mathbf{0}\rangle \otimes |-\rangle_a) \\ &= \frac{1}{\sqrt{2}}(|\mathbf{0}\rangle \otimes |+\rangle_a + U^\dagger P U|\mathbf{0}\rangle \otimes |-\rangle_a). \end{aligned} \quad (28)$$

Finally, we perform a projective measurement onto the  $\mathbf{0}$  outcome, effected by the projection operator  $P_0 = |\mathbf{0}\rangle\langle\mathbf{0}|$  with probability  $p_0 = \langle\psi|P_0|\psi\rangle$ :

$$\begin{aligned} |\psi_4\rangle &= \frac{1}{\sqrt{p_0}}(P_0 \otimes I)|\psi_3\rangle \\ &= \frac{1}{\sqrt{2p_0}}(|\mathbf{0}\rangle \otimes |+\rangle_a + |\mathbf{0}\rangle \underbrace{\langle\mathbf{0}|U^\dagger P U|\mathbf{0}\rangle}_{=(P)_\psi} \otimes |-\rangle_a) \\ &= |\mathbf{0}\rangle \otimes \underbrace{\frac{1}{\sqrt{2p_0}}(|+\rangle_a + (P)_\psi|-\rangle_a)}_{:=|\phi\rangle_a}; \end{aligned} \quad (29)$$

in practice, this projective measurement is realized by postselecting on zero measurement outcomes.

In effect, we have induced a virtual calculation of the desired expectation value on the ancilla qubit. The quantity  $\langle P \rangle_\psi$  may be extracted by performing measurements of the ancilla state  $|\phi\rangle$  in the  $X$  and  $Z$  bases, as we shall demonstrate now.

Using the normalization condition for  $|\phi\rangle$ , we infer that

$$p_0 = \frac{1 + \langle P \rangle_\psi^2}{2}, \quad (30)$$

which we note is at least  $\frac{1}{2}$ , meaning we should in principle retain at worst 50% of the samples taken from the quantum hardware. Inspecting (29), sampling from the  $|\phi\rangle$  state in the

X-basis yields  $|+\rangle$  and  $|-\rangle$  with probabilities  $p_0^X$  and  $p_1^X$ , respectively, from which we obtain the estimator

$$\mathcal{E}^X = p_0^X - p_1^X \approx \frac{1 - \langle P \rangle_\psi^2}{2p_0} = \frac{1 - \langle P \rangle_\psi^2}{1 + \langle P \rangle_\psi^2} \quad (31)$$

for the ancilla expectation value  $\langle X \rangle_\phi$ .

We may also express  $|\phi\rangle$  in the Z-basis

$$|\phi\rangle = \frac{1}{2\sqrt{p_0}}[(1 + \langle P \rangle_\psi)|0\rangle_a + (1 - \langle P \rangle_\psi)|1\rangle_a]; \quad (32)$$

sampling from this state, we obtain  $|0\rangle$  and  $|1\rangle$  with probabilities  $p_0^Z$  and  $p_1^Z$ , respectively. From this we may derive an estimator

$$\begin{aligned} \mathcal{E}^Z &= p_0^Z - p_1^Z \\ &\approx \frac{1}{4p_0}[(1 + \langle P \rangle_\psi)^2 - (1 - \langle P \rangle_\psi)^2] \\ &= \frac{\langle P \rangle_\psi}{p_0} \\ &= \frac{2\langle P \rangle_\psi}{1 + \langle P \rangle_\psi^2} \end{aligned} \quad (33)$$

for the ancilla expectation value  $\langle Z \rangle_\phi$ .

Finally, by combining (31) and (33) we may reconstruct an error mitigated estimator for the desired quantity  $\langle P \rangle_\psi$ :

$$\mathcal{E}_{\text{DSP}} = \frac{\mathcal{E}^Z}{1 + \mathcal{E}^X}. \quad (34)$$

One may actually reconstruct  $\langle P \rangle_\psi$  using *only* the Z-basis measurements by noting  $\langle X \rangle_\phi^2 + \langle Z \rangle_\phi^2 \equiv 1$  and therefore

$$\langle P \rangle_\psi = \frac{\langle Z \rangle_\phi}{1 + \sqrt{1 - \langle Z \rangle_\phi^2}}, \quad (35)$$

which one may arrive at by forming a quadratic equation from (33) and solving. Doing the same for the X-basis measurements yields

$$\langle P \rangle_\psi = \pm \sqrt{\frac{1 - \langle X \rangle_\phi}{1 + \langle X \rangle_\phi}}, \quad (36)$$

however it is not possible to determine the correct sign using these measurements alone; supplementary Z-basis measurements would be required to indicate the sign here.

One might also note that, expressing  $|\phi\rangle$  in the Y-basis

$$\begin{aligned} |\phi\rangle &= \frac{1}{2\sqrt{2p_0}}\{[(1 + \langle P \rangle_\psi) - i(1 - \langle P \rangle_\psi)]|+\rangle_a \\ &\quad + [(1 + \langle P \rangle_\psi) + i(1 - \langle P \rangle_\psi)]|-\rangle_a\}, \end{aligned} \quad (37)$$

we must have  $\langle Y \rangle_\phi = p_0^Y - p_1^Y = 0$ ; this was also noted by Huo and Li [20] and we might be able to exploit this fact for additional error mitigation in future work.

In Fig. 8 we present the DSP circuit. The only errors that are not suppressed through this process are those occurring in the readout phase, since errors may propagate through to the ancilla register and are not canceled during the subsequent uncomputation. However, there is one additional trick we may employ here; if the circuit is error-free, then the state of the

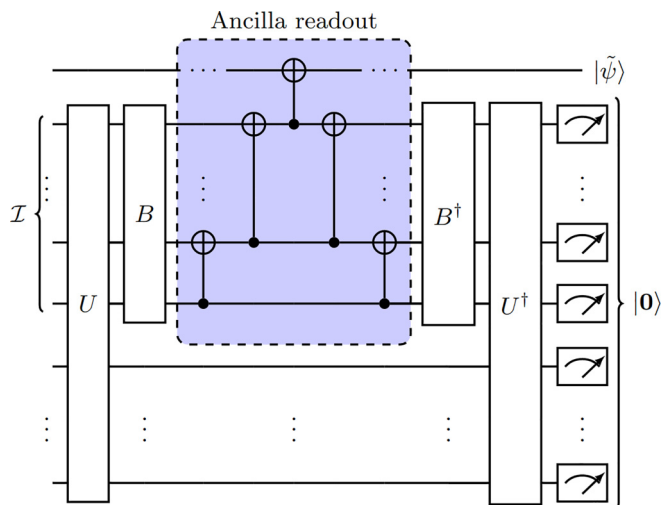


FIG. 8. Schematic of the dual-state purification protocol with readout to a single ancilla qubit, where  $\mathcal{I}$  indicates the nonidentity qubit positions of the Hamiltonian term being measured. In reality, one must construct the readout subcircuit with careful consideration of the chip topology to avoid excessive SWAP usage; for the 3-qubit HCl problem we report the optimal readout blocks in Fig. 12.

ancilla qubit is necessarily pure. In practice, the ancilla will be described by a mixed state

$$\rho = (1 - \epsilon)|\varphi_0\rangle\langle\varphi_0| + \epsilon|\varphi_1\rangle\langle\varphi_1|, \quad (38)$$

where  $\epsilon$  is the infidelity, which we may characterize fully via state tomography. Measuring the ancilla in the X, Y, Z bases, we may reconstruct  $\rho = \frac{1}{2}(I + \gamma_X X + \gamma_Y Y + \gamma_Z Z)$ , where  $\gamma_P = \text{Tr}(P\rho)$ , and identifying the largest eigenvalue with corresponding eigenvector  $|\varphi_0\rangle$  we take this as an approximation to the pure state  $|\phi\rangle$  obtained in the noiseless setting. Huo and Li [20] found this additional state tomography procedure to be essential in obtaining accurate results from dual-state purification.

Furthermore, from (31) we note that  $\langle X \rangle_\phi \geq 0$ , but in practice it is possible for a negative value to appear from quantum experiments. In fact, the depolarizing noise can be sufficiently high such that the corresponding eigenvalue of  $\rho$  dominates, resulting in spurious expectation values that can violate this non-negativity constraint maximally. This is a considerable problem when one considers the form (34), since this can result in division by zero, yielding a potentially infinite expectation value estimate for  $\langle P \rangle_\psi$ . We combat this by always choosing the eigenvalue with positive  $\langle X \rangle_\phi$ , even in the case when it does not hold the greatest weight. We observed this in particular for terms necessitating expensive SWAP operations; for our HCl circuit, this meant only terms of the form  $ZZI$  under change-of-basis (see Sec. VB for details), since this results in a closed loop of three CNOT gates which is not directly expressible on any IBM system (the heavy-hex topology of Fig. 4 does not contain cycles of three connected qubits). We also observed instability of the tomography purification method when  $\langle Z \rangle_\phi \approx 0$  whereby the error can be increased through this procedure. Therefore, we opted only to run this additional step when the raw expectation value exceeded some threshold near zero, taking the standard DSP result otherwise.

A potential modification for future work would be to flip the initial state of the system register  $|0\rangle \rightarrow |1\rangle$  via a layer of  $X$  gates and postselect on  $\mathbf{1}$  measurement outcomes. While this should theoretically be no different from initializing with  $|0\rangle$ , the effect of  $T_1$  relaxation is for qubits to decay into the energetically favorable  $|0\rangle$  state (as was observed in Fig. 5), resulting in the erroneous postselection of invalid measurements. By flipping the initial state, we should expect to retain fewer measurements in the postselected data, but the probability of these corresponding with successful circuit runs should be improved.

## V. GROUND STATE PREPARATION

Before proceeding to the quantum error mitigation benchmark, there are a few additional considerations to resolve. First, one must identify a suitable ansatz circuit that is sufficiently expressible to realize the desired ground state. Second, we discuss the mapping of our circuits onto physical qubits, in particular for dual-state purification since one should be mindful of the added qubit connectivity constraints arising from parity computation stored on the ancilla qubit. Third, despite not implementing any shot reduction methods in this work, we still wish to distribute the shot budget in an informed manner, preferably tailored to each device; this is the final point of discussion before moving onto the results of our benchmark.

We stress that, while eigenvalue estimation is a fundamental component of the VQE algorithm, we are not running the optimization itself on hardware, only the ground state preparation for means of benchmarking the QEM techniques presented in Sec. IV. There are additional considerations to take into account when running full VQE on quantum hardware. For example, the algorithm itself exhibits a level of robustness against noise as the optimizer can effectively combat hardware error due to the variational principle. The caveat here is that, since it is not in general possible to measure the full Hamiltonian simultaneously, the optimization may become nonvariational due to state preparation discrepancies when measuring different subsets of Pauli operators.

### A. Ansatz construction

Initially, we tested the noncontextual projection ansatz [53] derived from the 316-term CCSD operator. The projection into the 3-qubit contextual subspace yields a 6-term excitation pool from which we identify four operators via qubit-ADAPT-VQE that permit chemical precision. Despite this dramatic reduction in circuit depth from the full UCCSD ansatz, the resulting noncontextual projection ansatz consists of 12 CNOT gates which we found to be prohibitive in achieving chemical precision.

To remedy this, we abandon chemical intuition in the name of hardware efficiency. It is already known that an arbitrary 3-qubit quantum state may be prepared on quantum hardware using at most four CNOT gates [63]. In fact, we found that only two CNOT gates are sufficient in constructing a 3-qubit ansatz circuit that is sufficiently expressible for our electronic structure problem, presented in Fig. 9. In Fig. 10 we present the outcome of a noiseless VQE simulation over this ansatz to illustrate its expressibility.

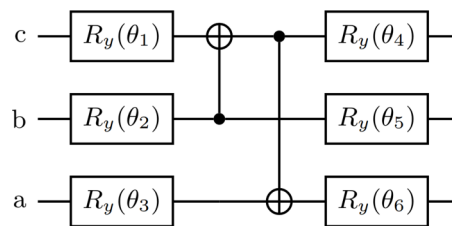


FIG. 9. Hardware efficient HCl 3-qubit contextual subspace ansatz; the  $Y$ -rotation gates are decomposed into native gates as  $R_y = \sqrt{X}R_z\sqrt{X}$ . The optimal parametrization obtained from the state vector simulation in Fig. 10 is  $\theta_1 = -0.064\,926\,67$ ,  $\theta_2 = 2.898\,361\,52$ ,  $\theta_3 = 0.263\,738\,07$ ,  $\theta_4 = -0.067\,090\,62$ ,  $\theta_5 = 0.010\,068\,33$ ,  $\theta_6 = -0.265\,850\,46$ .

### B. Ancilla readout mapping for DSP

The main bottleneck for dual-state purification is the ancilla readout step. Given the limited topology of the available quantum systems (Fig. 4) and the structure of our ansatz (Fig. 9), it is not possible to realize every 3-qubit Pauli  $Z$  measurement basis ( $IIZ, IZI, IZZ, ZII, ZIZ, ZZI, ZZZ$ ) without the aid of SWAP operations since at least one basis will always result in a closed loop of three CNOTs, which cannot be directly implemented on the hardware. We identified an optimal readout mapping that ensures that just one measurement basis requires a SWAP operation by selecting a cluster of five qubits of the form in Fig. 11 and implementing the readout as per Fig. 12.

### C. Shot budget distribution

To ensure a fair comparison, we define a fixed shot budget  $B$  up front and distribute according to the particular combined error mitigation strategy. The optimal shot distribution is in proportion with  $v_P := |h_P| \sqrt{\text{var}(P)}$ , where  $\text{var}(P) = 1 - \langle P \rangle_\psi^2$  [64]; however, the state-dependency means this may only be evaluated in-circuit. Therefore, we allocate 0.1% ( $b = 0.001$ ) of the overall budget to determine a rough estimate of the variance for each Hamiltonian term in order to rebalance the shot distribution accordingly; after this

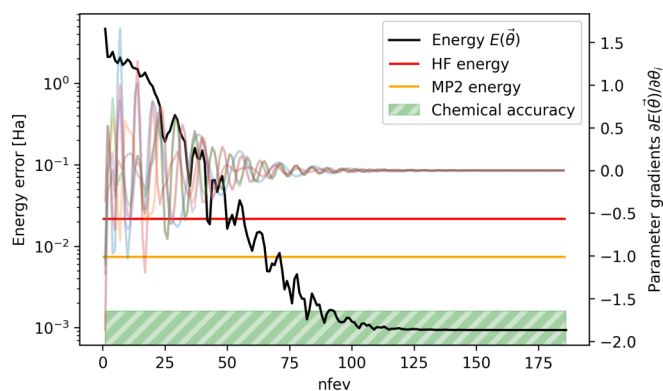


FIG. 10. Noiseless 3-qubit CS-VQE simulation of the HCl molecule over the hardware efficient ansatz presented in Fig. 9. The classical optimizer used is adaptive moment (Adam) estimation with gradients calculated using the parameter shift rule; we see that the ansatz is sufficiently expressible to achieve chemical precision.

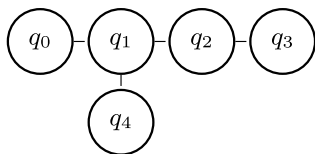


FIG. 11. The five-qubit cluster we require for dual-state purification in order to facilitate readout in every possible measurement basis. There are 18 such clusters on the 27-qubit Falcon chip (see Fig. 4) and we selected the optimal one with respect to gate and readout errors.

preliminary step, we are left with  $B' = (1 - b)B$  remaining shots. For example, defining  $V = \sum_p v_p$  we allow the following:

- (i) ZNE:  $\frac{B'v_P}{\Lambda V}$  circuit shots for each Pauli term  $P$  per noise amplification factor, where  $\Lambda$  is the number of noisy estimates desired for the energy extrapolation procedure.
- (ii) DSP:  $\frac{B'v_P}{2V}$  circuit shots for each Pauli term  $P$ , where the factor of  $\frac{1}{2}$  comes from performing both  $X$  and  $Z$  measurements over the ancilla qubit.
- (iii) DSP + ZNE:  $\frac{B'v_P}{2\Lambda V}$  circuit shots for each Pauli term  $P$  per noise amplification factor.

Since the shot budget is fixed, layering multiple error mitigation techniques may result in increased variance since fewer shots might be allocated to individual point estimates. It is the goal of this work to practically evaluate this tradeoff between absolute error and uncertainty in the energy estimate, which has been noted in numerous studies [22,24].

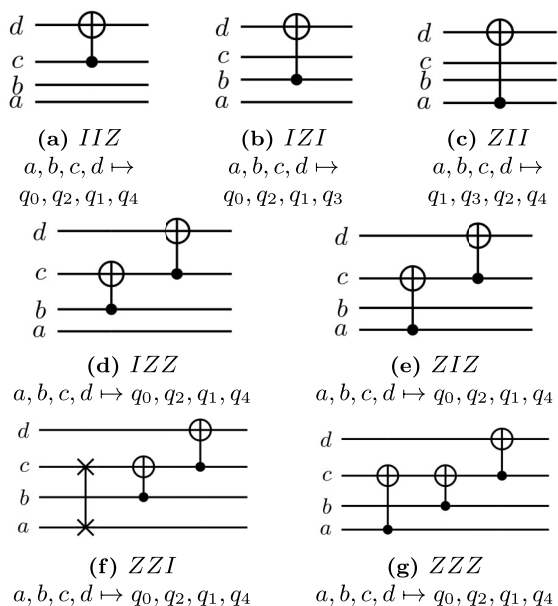


FIG. 12. Ancilla readout mappings given qubit clusters of the form in Fig. 11. Given the qubit topology of Fig. 4 and the form of our ansatz in Fig. 9 (where qubits  $a, b, c$  are the same as above, with  $d$  the ancilla qubit), we may not entangle qubits  $a$  and  $b$  since it would result in a closed loop of three CNOT operations that is not expressible on the available quantum systems. We avoid this situation by introducing a single SWAP operation (represented in-circuit as  $\times-\times$ ) for Hamiltonian terms of the form  $ZZI$  as in (f).

### D. Methods

To construct the molecular Hamiltonian for HCl (bond length 1.341 Å), we first performed a restricted Hartree-Fock calculation in *PySCF* [55] in the STO-3G basis. *OpenFermion* was then used to build the second-quantized fermionic molecular Hamiltonian [65] and was mapped onto Pauli operators via the Jordan-Wigner transformation [59]. This was then converted into the *Symmer* [58] operator representation to leverage the included tapering and contextual subspace functionality, which facilitated a reduction to 3-qubits while incurring a ground state energy error of just 0.837 mHa in the resulting contextual subspace Hamiltonian with respect to full configuration interaction (FCI); Sec. III discusses this in further detail.

We used *Qiskit* [66] for the construction of our hardware efficient ansatz circuit, and the state preparation jobs required for each quantum error mitigation (QEM) strategy were composed as *Qiskit Runtime* programs. These were submitted to the *IBM Quantum* service and allowed us to retrieve all the necessary quantum circuit samples in the shortest amount of time possible to mitigate against noise drift.

The *mtree* [2] package was utilized to perform measurement-error mitigation (MEM, Sec. IV B) whereas we wrote bespoke implementations for all the other QEM techniques introduced in Sec. IV, namely symmetry verification (SV, Sec. IV C), zero-noise extrapolation (ZNE, Sec. IV D), and dual-state purification (DSP, Sec. IV E) with or without tomography purification (TP). The linear regression functionality of *statsmodels* [67] was utilized for the purposes of ZNE, and the relevant postprocessing required for each QEM technique was parallelized with *multiprocessing* to permit a greater number of resamples to be extracted in our bootstrapping procedures (see Appendix for details).

We also provide all the Hamiltonian data, runtime program scripts, quantum experiment data, and postprocessing functions to aid the reader in reproducing the results of this paper, accessible via GitHub [68].

### E. Results

In Table IV we report the results of benchmarking our suite of error mitigation strategies for the 3-qubit HCl problem across every 27-qubit system currently available to us through IBM Quantum with a shot budget of  $B = 10^6$ ; the order in which each QEM technique (MEM, SV, ZNE, DSP, TP) appears in the combined strategy identifier indicates the order in which each method is being applied. Table III presents the average error suppression in relation to the raw estimate, calculated as

$$\left(1 - \left| \frac{\text{bias}(\mathcal{E}_{\text{QEM}})}{\text{bias}(\mathcal{E}_{\text{RAW}})} \right| \right) \times 100\%, \quad (39)$$

and change in standard deviation  $\sigma$  across our suite of systems excluding *ibm\_hanoi* and *ibm\_geneva*, due to these systems performing suboptimally (resulting in a failure of ZNE in Fig. 7). When  $\text{bias}(\mathcal{E}_{\text{QEM}})$  is near zero, the error suppression will approach 100%, whereas values close to 0% indicate little (or no) improvement over the raw estimator; negative values of error suppression correspond with instances whereby the

TABLE III. Average error suppression and change in standard deviation under each error mitigation strategy evaluated across all 27-qubit Falcon IBM Quantum devices (excluding *ibm\_hanoi* and *ibm\_geneva*, which did not perform well, as seen in the ZNE plots of Fig. 7 and Table IV). Ordered by decreasing mean error suppression.

	Error Suppression (%)			Change in Std. Dev.		
	Mean	Best	Worst	Mean	Best	Worst
MEM+SV+ZNE	94.327	99.392	88.101	3.680	1.078	7.207
DSP+TP	93.253	99.713	80.793	2.543	1.583	3.833
MEM+DSP+TP	92.661	98.601	75.508	2.113	0.789	3.472
MEM+ZNE	87.094	97.877	69.185	7.069	1.202	25.063
MEM+SV	82.678	96.643	67.108	0.638	0.519	0.758
SV+ZNE	79.799	94.938	52.882	4.270	0.385	8.594
MEM	76.505	96.704	65.358	0.669	0.517	0.762
SV	63.577	80.992	33.191	0.748	0.645	0.887
MEM+DSP+TP+ZNE	59.767	99.758	-98.987	6.738	3.104	8.853
DSP+TP+ZNE	34.012	95.721	-107.805	7.462	5.593	9.523
ZNE	33.384	52.230	18.303	6.699	0.642	24.689
MEM+DSP+ZNE	-10.180	93.343	-238.874	6.779	3.715	8.601
MEM+DSP	-18.002	75.430	-330.030	2.224	1.080	3.440
DSP+ZNE	-68.019	1.854	-298.966	7.235	5.453	8.989
DSP	-76.967	29.366	-393.687	2.620	1.896	3.726

QEM strategy has had a detrimental effect on the energy estimate, a highly unfavorable situation.

The shot budget yields a raw standard deviation of  $2 < \sigma < 3$  mHa, quantified via a bootstrapping procedure (discussed in Appendix). In Fig. 13 we plot the bootstrap distributions for a selection of the best performing QEM strategies to illustrate the tradeoff between estimator bias and variance in practice, serving as a valuable comparison with previous theoretical analyses [22].

We observed that application of the MEM and SV techniques served to consistently lower both the estimator bias and standard deviation, which can be attributed to these approaches rectifying readout errors. Used in combination, the MEM + SV strategy permitted a respectable reduction in bias while also suppressing deviations with very little classical overhead.

Unlike MEM and SV, the ZNE and DSP techniques necessitate modification to the quantum circuits themselves: the former, a decomposition of each CNOT gate into procedurally more complex circuit blocks, and the latter requiring a prepare-readout-invert structure with a supplementary ancilla qubit. This amounts to an increase in circuit depth for both techniques that must be taken into consideration when using them. For example, DSP doubles the circuit depth and therefore one must ensure the primitive ansatz is executable within half the coherence time of the device (minus the time required for measurement). On the other hand, the increase in circuit depth arising from ZNE limits the permitted noise amplification factors, although it is not recommended to go to large factors regardless. Both of these methods can be seen to inflate the standard deviation.

By itself, DSP performs very poorly (indeed, the worst four strategies were all DSP-based), but when used in combination with TP we are permitted dramatic reductions in bias which exceed all other QEM strategies in the benchmark. The dependence on tomography purification for the ancilla

qubit was also observed in Huo and Li [20] and is essential to obtain good results from dual-state purification. We stress that, although state tomography is not scalable in general, here it is applied to a single qubit and hence does not contribute a significant cost in the number of measurements required.

We found mixed success with ZNE-based strategies depending on which other QEM techniques were deployed in combination. Applied on top of MEM and SV, we observed a significant improvement in error suppression, for *ibm\_hanoi*, and *ibm\_geneva* where extrapolation failed (Fig. 7), albeit at a significant increase in standard deviation. On the other hand, performing noise amplification on the ancilla qubit for the purposes of DSP produced disappointing results. These observations might be attributed to coherent errors causing unpredictable noise amplification behavior; this could have been improved by including probabilistic error cancellation [6], converting coherent error into incoherent error that may be extrapolated more confidently.

## VI. CONCLUSION

In this work, we compared various quantum error-mitigation strategies for estimating the ground state energy of the HCl molecule on NISQ hardware. Indicated by the benchmark results of Sec. VE, we identified three hybrid strategies with the strongest performance:

(i) *Dual-state purification with tomography purification (DSP + TP)*. This yields compelling error suppression (93.253% on average) although at an increase in standard deviation (2.543 times the raw value on average); given a generous shot budget and sufficient qubit connectivity, this strategy should produce reliably accurate results. Implementing dual-state purification requires heavy modification to the ansatz resulting in doubled circuit depth, although the errors incurred here are suppressed. Further layering measurement-error mitigation produces a similar suppression in error,

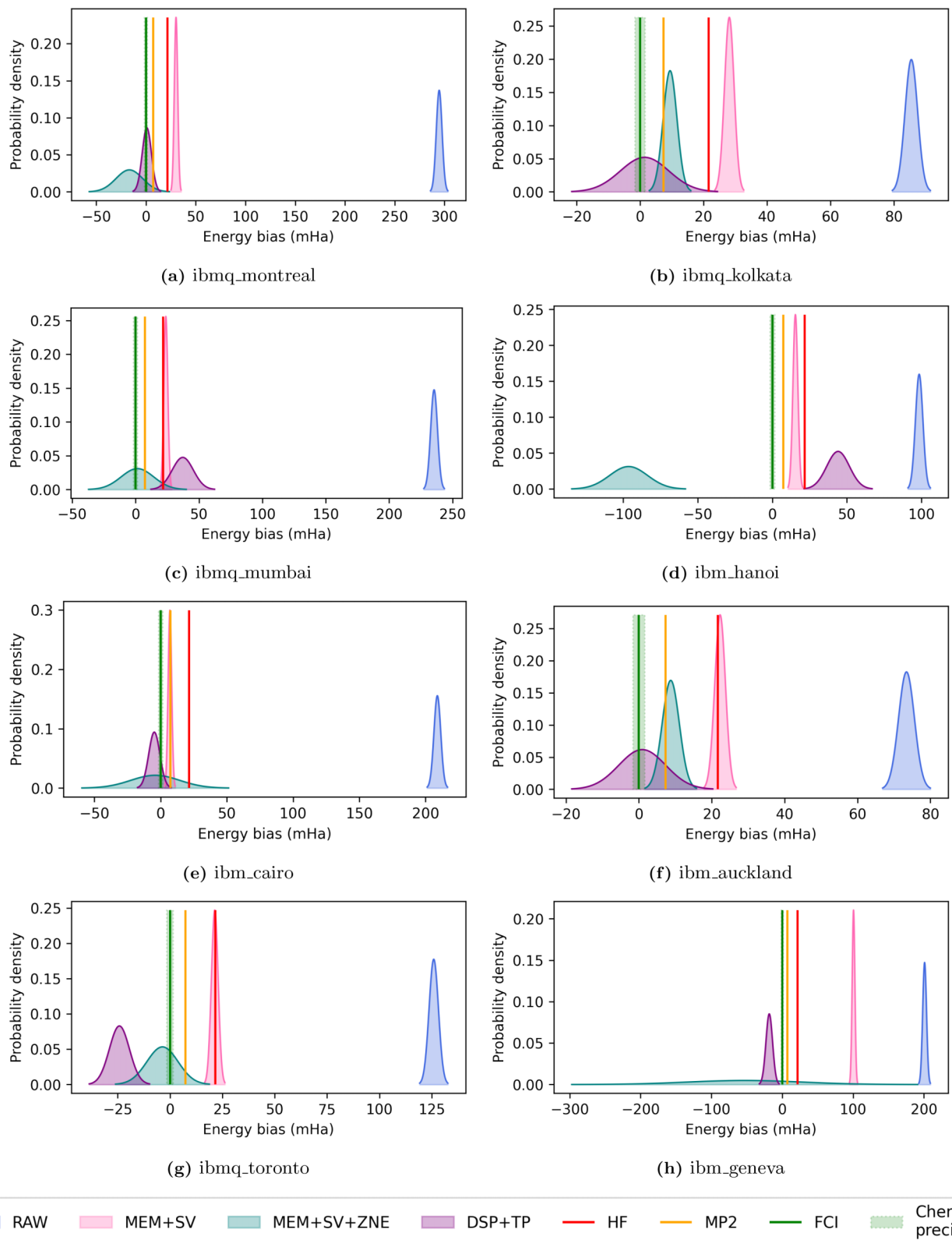


FIG. 13. Bootstrapped distributions for the best three QEM strategies identified through our benchmark. The mean energy of each distribution corresponds with the estimator bias. Note the failure of ZNE on *ibmq\_hanoi* and *ibmq\_geneva*, which is explained in Sec. IV D and illustrates the sensitivity of this technique to erroneous fluctuations in the noise amplification.

although the increase in standard deviation is slightly less (2.113 times the raw value on average).

(ii) *Measurement-error mitigation with symmetry verification (MEM + SV)*. This comes with very low overhead (top of respectable error suppression (82.678% on average) on top of a reduction in standard deviation (0.638 times the raw value on average). Furthermore, there is no required modification to the ansatz circuit since both techniques operate solely on the binary measurement output. We recommend this strategy for restrictive shot budgets or where qubit topology does not permit the readout block needed for dual-state purification.

(iii) *Zero-noise extrapolation on top of measurement error mitigation with symmetry verification (MEM + SV + ZNE)*. This is sensitive to many factors, but used carefully it can yield excellent results (94.327% average error suppression when we exclude the cases where extrapolation failed in Fig. 7). There are many approaches to implementing ZNE, even extending to the pulse-level. On superconducting devices this might be preferable since it offers fine control over noise amplification. ZNE produced the largest inflation in standard deviation (3.680 times the raw value on average) and therefore a significantly greater shot budget would be necessary, due to error propagation in the extrapolation and since we evaluate several noise factors per expectation value.

As indicated by Table III, each of these strategies achieved an average error suppression exceeding 80% across the suite of 27-qubit IBM Quantum chips. Given the level of noise present on these devices, reflected in the raw energy estimates, the high bar of chemical precision would necessitate

a suppression of 98.783%. This was obtained for three out of eight instances of DSP + TP (on the highest QV=128 systems *ibmq\_montreal* and *ibmq\_kolkata*, plus the QV=64 system *ibmq\_auckland*, with further device specifications given in Table I) and a single instance of MEM + SV + ZNE (on the QV=128 system *ibmq\_mumbai*), bearing in mind the standard deviation exceeds the chemically precise region, and an increased shot budget would be necessary to counteract this.

From the empirical results presented in this work, it is clear that we must rely heavily on methods of quantum error mitigation if we are to obtain usable results from NISQ hardware. Through our benchmark on the IBM Quantum 27-qubit Falcon processors, we have demonstrated the most effective combined strategies, which we intend to take forward in our future simulation work.

### ACKNOWLEDGMENTS

T.W. and A.R. acknowledge support from the Unitary Fund and the Engineering and Physical Sciences Research Council (EP/S021582/1 and EP/L015242/1, respectively). T.W. also acknowledges support from CBKSciCon Ltd., Atos, Intel, and Zapata. W.M.K. and P.J.L. acknowledge support by the NSF STAQ project (PHY-1818914). W.M.K. acknowledges support from the National Science Foundation, Grant No. DGE-1842474. S.S. wishes to acknowledge financial support from the National Centre for HPC, Big Data and Quantum Computing (Spoke 10, CN00000013). P.V.C. is grateful for funding from the European Commission for VECMA (800925) and EPSRC for SEAVEA (EP/W007711/1). Access

TABLE IV. Comparison of estimator bias and standard deviation  $\sigma$  (given in mHa) for various error mitigation strategies performed across a suite of IBM Quantum 27-qubit Falcon devices with a shot budget  $B = 10^6$ . The columns are ordered left-to-right by decreasing average error suppression (detailed in Table III), with the exception of the raw estimate, which is provided for reference. Note the following *classical* quantum chemistry error benchmarks: HF, 21.621 mHa; MP2, 7.360 mHa; Chemical Precision, 1.6 mHa.

	RAW	MEM+SV+ZNE	DSP+TP	MEM+DSP+TP	MEM+ZNE	MEM+SV	SV+ZNE	MEM	SV	MEM+DSP+TP+ZNE	DSP+TP+ZNE	ZNE	MEM+DSP+ZNE	MEM+DSP	DSP+ZNE	DSP
<i>ibmq_montreal</i>	<b>bias</b> 294.7	<b>16.7</b>	<b>0.8</b>	<b>6.9</b>	<b>19.0</b>	<b>30.2</b>	<b>14.9</b>	<b>85.9</b>	<b>59.2</b>	<b>61.1</b>	<b>143.2</b>	<b>240.7</b>	<b>65.6</b>	<b>72.4</b>	<b>301.9</b>	<b>208.1</b>
	$\sigma$ 2.9	13.4	4.6	3.7	16.5	1.7	14.0	2.0	1.9	9.0	18.1	13.5	10.8	4.2	16.8	5.8
<i>ibmq_kolkata</i>	<b>bias</b> 85.5	<b>9.4</b>	<b>1.4</b>	<b>3.3</b>	<b>1.8</b>	<b>28.1</b>	<b>40.3</b>	<b>28.2</b>	<b>57.1</b>	<b>8.8</b>	<b>10.6</b>	<b>65.5</b>	<b>93.0</b>	<b>37.3</b>	<b>144.5</b>	<b>88.7</b>
	$\sigma$ 2.0	2.2	7.7	6.9	7.4	1.5	1.9	1.5	1.8	17.5	18.2	4.7	17.2	6.9	18.0	7.4
<i>ibmq_mumbai</i>	<b>bias</b> 235.4	<b>1.4</b>	<b>37.2</b>	<b>3.3</b>	<b>36.0</b>	<b>23.8</b>	<b>18.6</b>	<b>42.9</b>	<b>44.7</b>	<b>0.6</b>	<b>54.5</b>	<b>118.1</b>	<b>146.7</b>	<b>165.1</b>	<b>332.4</b>	<b>349.8</b>
	$\sigma$ 2.7	12.9	8.4	6.9	12.9	1.6	15.5	1.7	1.7	23.9	25.7	12.7	23.2	6.7	23.8	7.7
<i>ibm_auckland</i>	<b>bias</b> 73.4	<b>8.7</b>	<b>0.9</b>	<b>7.1</b>	<b>5.8</b>	<b>22.3</b>	<b>29.1</b>	<b>25.4</b>	<b>38.1</b>	<b>146.1</b>	<b>152.6</b>	<b>54.2</b>	<b>248.8</b>	<b>315.8</b>	<b>293.0</b>	<b>362.5</b>
	$\sigma$ 2.2	2.4	6.5	5.5	2.6	1.5	0.8	1.5	1.7	15.7	15.8	1.4	16.4	6.0	16.5	6.6
<i>ibm_cairo</i>	<b>bias</b> 208.9	<b>4.1</b>	<b>4.8</b>	<b>4.7</b>	<b>64.4</b>	<b>7.0</b>	<b>26.5</b>	<b>6.9</b>	<b>64.0</b>	<b>19.5</b>	<b>8.9</b>	<b>99.8</b>	<b>255.2</b>	<b>200.4</b>	<b>205.4</b>	<b>242.8</b>
	$\sigma$ 2.6	18.5	4.2	5.3	64.2	1.3	22.0	1.3	1.9	20.8	18.2	63.3	19.6	5.5	17.5	4.9
<i>ibm_hanoi</i>	<b>bias</b> 98.5	<b>96.4</b>	<b>44.1</b>	<b>31.1</b>	<b>157.9</b>	<b>15.5</b>	<b>64.5</b>	<b>20.8</b>	<b>30.8</b>	<b>12.0</b>	<b>26.6</b>	<b>110.6</b>	<b>59.6</b>	<b>90.2</b>	<b>84.7</b>	<b>171.9</b>
	$\sigma$ 2.5	12.7	7.6	7.3	23.8	1.6	12.2	1.7	1.9	21.2	20.1	23.4	21.5	7.2	21.0	7.2
<i>ibmq_toronto</i>	<b>bias</b> 125.9	<b>3.6</b>	<b>24.2</b>	<b>30.8</b>	<b>18.8</b>	<b>21.4</b>	<b>11.0</b>	<b>28.6</b>	<b>37.9</b>	<b>2.3</b>	<b>125.6</b>	<b>87.7</b>	<b>8.4</b>	<b>55.0</b>	<b>123.6</b>	<b>162.4</b>
	$\sigma$ 2.2	7.5	4.8	1.8	4.5	1.6	11.5	1.7	1.8	9.9	12.6	7.2	10.3	2.4	12.2	5.0
<i>ibm_geneva</i>	<b>bias</b> 200.9	<b>53.0</b>	<b>18.4</b>	<b>3.6</b>	<b>109.5</b>	<b>100.6</b>	<b>65.9</b>	<b>100.6</b>	<b>153.7</b>	<b>27.8</b>	<b>37.9</b>	<b>107.2</b>	<b>358.7</b>	<b>28.0</b>	<b>471.2</b>	<b>232.3</b>
	$\sigma$ 2.7	81.4	4.7	7.8	148.4	1.9	24.3	1.9	2.3	14.6	12.3	60.4	15.2	7.8	16.1	5.5

to the IBM Quantum Computers was obtained through the IBM Quantum Hub at CERN with which the Italian Institute of Technology (IIT) is affiliated. We would also like to thank George Ralli, Andrew Tranter at Quantinuum, William Simon and Oliver Maupin at Tufts University, Marco Maronese at IIT, and Michele Grossi and Oriol Kiss at CERN for valuable discussions.

### APPENDIX: BOOTSTRAPPING

To evaluate the uncertainty in our energy estimates, we rely on the statistical technique of bootstrapping [69]. Ideally, one would perform quantum experiments many times to probe the “true” population, but from a practical standpoint this is not feasible due to the length of time required for each energy estimate (in our case  $\approx 30$  min for a shot budget of  $B = 10^6$ ). Instead, we perform the experiment just once and generate resampled measurement data from the empirical distribution. This technique is widespread in statistics and makes the statistical analysis very convenient, not least as we may assume normality under the central limit theorem, which we verified using the *normaltest* function in *SciPy* [70] that implements the D’Agostino-Pearson test [71].

Suppose we perform an  $n$ -shot quantum experiment and obtain a collection of binary measurement outcomes  $M = \{m_1, \dots, m_n\}$ , where  $m_i \in \mathbb{Z}_2^N$ . Our various QEM strategies combine these measurements in some way to yield an energy estimate  $\mathcal{E}(M)$ , but we would like to say something about the uncertainty in each estimator *without* having

to perform further experiments. The bootstrapping approach involves resampling from the empirical measurement distribution  $M$ , namely sampling elements  $m'_i \in M$  with replacement to form a new set of  $n$  measurements  $M'$ . We perform this process as many times as possible given the available compute resource, say  $R \in \mathbb{N}$  repetitions, to approximate

$$\text{var}(\mathcal{E}) \approx \frac{1}{R^2} \sum_{r=1}^R \sum_{s>r}^R (\mathcal{E}(M'_r) - \mathcal{E}(M'_s))^2; \quad (\text{A1})$$

this is how we obtained the variances in Table IV.

One might question whether bootstrapping is well-motivated here. *A priori*, one has no reason to expect acceptable agreement with the true population parameters, hence we ran 225 instances of our quantum experiment applied just to the diagonal terms of the Hamiltonian (given in Table II), necessitating only computational basis measurements. We performed 10 000 circuit shots in each experiment, for a combined total of  $2.25 \times 10^6$  point samples before assessing the quality of the bootstrapped distributions against the overall sample. The 225 quantum experiments provide a *target* standard deviation  $\sigma$ , indicated by the vertical line in Fig. 15, and we compare with this the *bootstrap* standard deviations obtained per experiment.

In Fig. 14 we plot the result of our bootstrapping test and see reasonable agreement with the true energy distribution obtained from the NISQ hardware; the standard deviations all coincide with the experimentally obtained value to  $O(10^{-3})$  (on the order of chemical precision), as indicated in Fig. 15, and therefore we employ bootstrapping with confidence.

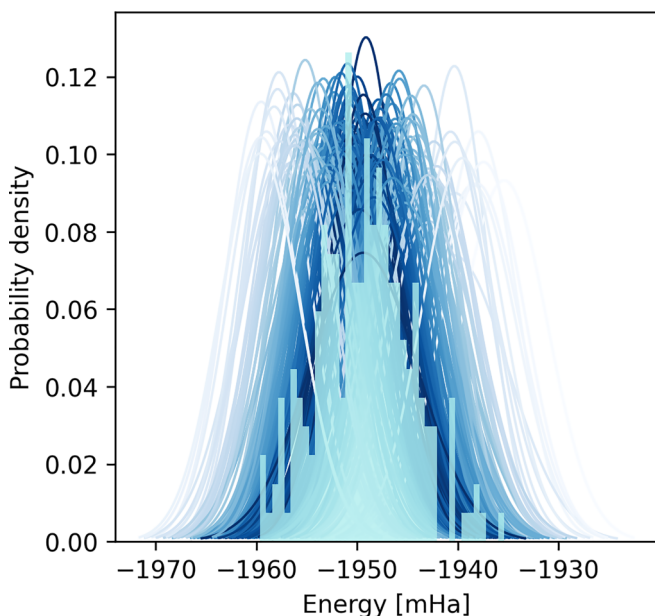


FIG. 14. The true distribution of energy estimates obtained from 225 quantum experiments on *ibmq\_kolkata*, each consisting of 10 000 circuit shots. Overlaid are the bootstrapped distributions for individual measurement sets to understand the relation between bootstrapping and the true population; the color gradient indicates how far a given sample lies from the true (empirical) mean.

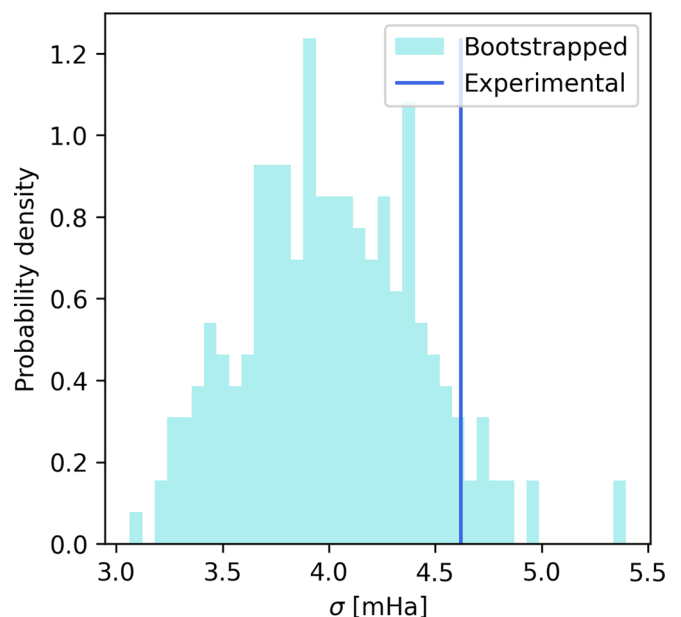


FIG. 15. Distribution of bootstrapped standard deviations  $\sigma$  vs the experimentally obtained value on *ibmq\_kolkata*. We observe good agreement, with the bootstrapped values correct up to  $O(10^{-3})$ .



- [1] S. Bravyi, S. Sheldon, A. Kandala, D. C. McKay, and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, *Phys. Rev. A* **103**, 042605 (2021).
- [2] P. D. Nation, H. Kang, N. Sundaresan, and J. M. Gambetta, Scalable mitigation of measurement errors on quantum computers, *PRX Quantum* **2**, 040326 (2021).
- [3] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, Learning-based quantum error mitigation, *PRX Quantum* **2**, 040330 (2021).
- [4] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with clifford quantum-circuit data, *Quantum* **5**, 592 (2021).
- [5] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Unified approach to data-driven quantum error mitigation, *Phys. Rev. Res.* **3**, 033098 (2021).
- [6] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [7] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, *Phys. Rev. X* **8**, 031027 (2018).
- [8] A. Mari, N. Shammah, and W. J. Zeng, Extending quantum probabilistic error cancellation by noise scaling, *Phys. Rev. A* **104**, 052607 (2021).
- [9] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, *Phys. Rev. X* **7**, 021050 (2017).
- [10] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, *Nature (London)* **567**, 491 (2019).
- [11] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, Digital zero noise extrapolation for quantum error mitigation, in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Piscataway, NJ, 2020), pp. 306–316.
- [12] A. He, B. Nachman, W. A. de Jong, and C. W. Bauer, Zero-noise extrapolation for quantum-gate error mitigation with identity insertions, *Phys. Rev. A* **102**, 012426 (2020).
- [13] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O’Brien, Low-cost error mitigation by symmetry verification, *Phys. Rev. A* **98**, 062339 (2018).
- [14] S. McArdle, X. Yuan, and S. Benjamin, Error-mitigated digital quantum simulation, *Phys. Rev. Lett.* **122**, 180501 (2019).
- [15] Z. Cai, Quantum error mitigation using symmetry expansion, *Quantum* **5**, 548 (2021).
- [16] W. J. Huggins, S. McArdle, T. E. O’Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, *Phys. Rev. X* **11**, 041036 (2021).
- [17] T. E. O’Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean, and R. Babbush, Error mitigation via verified phase estimation, *PRX Quantum* **2**, 020317 (2021).
- [18] Z. Cai, Resource-efficient purification-based quantum error mitigation, [arXiv:2107.07279](https://arxiv.org/abs/2107.07279).
- [19] P. Czarnik, A. Arrasmith, L. Cincio, and P. J. Coles, Qubit-efficient exponential suppression of errors, [arXiv:2102.06056](https://arxiv.org/abs/2102.06056).
- [20] M. Huo and Y. Li, Dual-state purification for practical quantum error mitigation, *Phys. Rev. A* **105**, 022427 (2022).
- [21] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, *J. Phys. Soc. Jpn.* **90**, 032001 (2021).
- [22] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O’Brien, Quantum error mitigation, [arXiv:2210.00921](https://arxiv.org/abs/2210.00921).
- [23] S. Resch and U. R. Karpuzcu, Benchmarking quantum computers and the impact of quantum noise, *ACM Comput. Surv.* **54**, 1 (2022).
- [24] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, *npj Quantum Inf.* **8**, 1 (2022).
- [25] R. Takagi, H. Tajima, and M. Gu, Universal sample lower bounds for quantum error mitigation, [arXiv:2208.09178](https://arxiv.org/abs/2208.09178).
- [26] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, [arXiv:2210.11505](https://arxiv.org/abs/2210.11505).
- [27] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 4213 (2014).
- [28] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure, *Phys. Rev. A* **95**, 020501(R) (2017).
- [29] 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, Aspuru-A. Guzik, and J. M. Martinis, Scalable quantum simulation of molecular energies, *Phys. Rev. X* **6**, 031007 (2016).
- [30] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, D. P. Tew, X. Zhou, J. L. O’Brien, and M. G. Thompson, Witnessing eigenstates for quantum simulation of Hamiltonian spectra, *Sci. Adv.* **4**, 1 (2018).
- [31] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature (London)* **549**, 242 (2017).
- [32] 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 (2018).
- [33] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush *et al.*, Quantum chemistry calculations on a trapped-ion quantum simulator, *Phys. Rev. X* **8**, 031022 (2018).
- [34] Y. Nam, J.-S. Chen, N. C. Pinti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf, *et al.*, Ground-state energy estimation of the water molecule on a trapped-ion quantum computer, *npj Quantum Inf.* **6**, 33 (2020).

- [35] A. J. McCaskey, Z. P. Parks, J. Jakowski, S. V. Moore, T. D. Morris, T. S. Humble, and R. C. Pooser, Quantum chemistry as a benchmark for near-term quantum computers, *npj Quantum Inf.* **5**, 99 (2019).
- [36] S. E. Smart and D. A. Mazziotti, Quantum-classical hybrid algorithm using an error-mitigating  $N$ -representability condition to compute the Mott metal-insulator transition, *Phys. Rev. A* **100**, 022517 (2019).
- [37] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell *et al.*, Hartree-Fock on a superconducting qubit quantum computer, *Science* **369**, 1084 (2020).
- [38] Q. Gao, G. O. Jones, M. Motta, M. Sugawara, H. C. Watanabe, T. Kobayashi, E. Watanabe, Y.-y. Ohnishi, H. Nakamura, and N. Yamamoto, Applications of quantum computing for investigations of electronic transitions in phenylsulfonyl-carbazole TADF emitters, *npj Comput. Mater.* **7**, 70 (2021).
- [39] Y. Kawashima, E. Lloyd, M. P. Coons, Y. Nam, S. Matsuura, A. J. Garza, S. Johri, L. Huntington, V. Senicourt, A. O. Maksymov, J. H. V. Nguyen, J. Kim, N. Alidoust, A. Zaribafiyani, and T. Yamazaki, Optimizing electronic structure simulations on a trapped-ion quantum computer using problem decomposition, *Commun. Phys.* **4**, 245 (2021).
- [40] J. E. Rice, T. P. Gujarati, M. Motta, T. Y. Takeshita, E. Lee, J. A. Latone, and J. M. Garcia, Quantum computation of dominant products in lithium-sulfur batteries, *J. Chem. Phys.* **154**, 134115 (2021).
- [41] A. Eddins, M. Motta, T. P. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield, and S. Sheldon, Doubling the size of quantum simulators by entanglement forging, *PRX Quantum* **3**, 010309 (2022).
- [42] M. Motta, G. O. Jones, J. E. Rice, T. P. Gujarati, R. Sakuma, I. Liepuoniute, J. M. Garcia, and Y. Ohnishi, Quantum chemistry simulation of ground- and excited-state properties of the sulfonium cation on a superconducting quantum processor, *Chem. Sci.* **14**, 2915 (2023).
- [43] K. Yamamoto, D. Z. Manrique, I. T. Khan, H. Sawada, and D. M. Ramo, Quantum hardware calculations of periodic systems with partition-measurement symmetry verification: Simplified models of hydrogen chain and iron crystals, *Phys. Rev. Res.* **4**, 033110 (2022).
- [44] J. J. M. Kirsopp, C. Di Paola, D. Z. Manrique, M. Krompiec, G. Greene-Diniz, W. Guba, A. Meyer, D. Wolf, M. Strahm, and D. Muñoz Ramo, Quantum computational quantification of protein-ligand interactions, *Int. J. Quantum Chem.* **122**, e26975 (2022).
- [45] I. T. Khan, M. Tudorovskaya, J. J. M. Kirsopp, D. M. Ramo, P. W. Warrier, D. K. Papanastasiou, and R. Singh, Chemically aware unitary coupled cluster with ab initio calculations on an ion trap quantum computer: A refrigerant chemicals' application, *J. Chem. Phys.* **158**, 214114 (2023).
- [46] T. E. O'Brien *et al.*, Purification-based quantum error mitigation of pair-correlated electron simulations, [arXiv:2210.10799](https://arxiv.org/abs/2210.10799).
- [47] L. Zhao, J. Goings, K. Wright, J. Nguyen, J. Kim, S. Johri, K. Shin, W. Kyoung, J. I. Fuks, J.-K. K. Rhee, and Y. M. Rhee, Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers, [arXiv:2212.02482](https://arxiv.org/abs/2212.02482).
- [48] O. Kiss, M. Grossi, P. Lougovski, F. Sanchez, S. Vallecorsa, and T. Papenbrock, Quantum computing of the  ${}^6\text{Li}$  nucleus via ordered unitary coupled clusters, *Phys. Rev. C* **106**, 034325 (2022).
- [49] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, *Phys. Rev. A* **100**, 032328 (2019).
- [50] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, Tapering off qubits to simulate fermionic Hamiltonians, [arXiv:1701.08213](https://arxiv.org/abs/1701.08213).
- [51] K. Setia, R. Chen, J. E. Rice, A. Mezzacapo, M. Pistoia, and J. D. Whitfield, Reducing qubit requirements for quantum simulations using molecular point group symmetries, *J. Chem. Theor. Comput.* **16**, 6091 (2020).
- [52] W. M. Kirby, A. Tranter, and P. J. Love, Contextual subspace variational quantum eigensolver, *Quantum* **5**, 456 (2021).
- [53] T. Weaving, A. Ralli, W. M. Kirby, A. Tranter, P. J. Love, and P. V. Coveney, A stabilizer framework for the contextual subspace variational quantum eigensolver and the noncontextual projection ansatz, *J. Chem. Theor. Comput.* **19**, 808 (2023).
- [54] A. Ralli, T. Weaving, A. Tranter, W. M. Kirby, P. J. Love, and P. V. Coveney, Unitary partitioning and the contextual subspace variational quantum eigensolver, *Phys. Rev. Res.* **5**, 013095 (2023).
- [55] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, Pyscf: the python-based simulations of chemistry framework, *WIREs Comput. Mol. Sci.* **8**, e1340 (2018).
- [56] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method, *J. Chem. Theor. Comput.* **16**, 190 (2020).
- [57] A. Ralli, P. J. Love, A. Tranter, and P. V. Coveney, Implementation of measurement reduction for the variational quantum eigensolver, *Phys. Rev. Res.* **3**, 033195 (2021).
- [58] A. Ralli and T. Weaving, symmer, <https://github.com/UCL-CCS/symmer> (2022).
- [59] P. Jordan and E. P. Wigner, Über das paulische äquivalenzverbot, *Z. Phys. A* **47**, 631 (1928).
- [60] S. B. Bravyi and A. Y. Kitaev, Fermionic quantum computation, *Ann. Phys.* **298**, 210 (2002).
- [61] Y. Saad, *Iterative Methods for Sparse Linear Systems* (SIAM, Philadelphia, 2003).
- [62] N. R. Draper and H. Smith, *Applied Regression Analysis* (Wiley, 1998), Vol. 326.
- [63] M. Žnidarič, O. Giraud, and B. Georgeot, Optimal number of controlled-not gates to generate a three-qubit state, *Phys. Rev. A* **77**, 032320 (2008).
- [64] N. C. Rubin, R. Babbush, and J. McClean, Application of fermionic marginal constraints to hybrid quantum algorithms, *New J. Phys.* **20**, 053020 (2018).
- [65] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby *et al.*, Openfermion: The electronic structure package for quantum computers, *Quantum Sci. Technol.* **5**, 034014 (2020).

- [66] M. S. Anis *et al.*, Qiskit: An open-source framework for quantum computing (2021).
- [67] S. Seabold and J. Perktold, statsmodels: Econometric and statistical modeling with python, in *9th Python in Science Conference* (2010).
- [68] T. Weaving, quantum-error-mitigation, <https://github.com/TimWeaving/quantum-error-mitigation> (2023).
- [69] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (CRC, Boca Raton, FL, 1994).
- [70] P. Virtanen *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in python, *Nat. Methods* **17**, 261 (2020).
- [71] R. D'Agostino and E. S. Pearson, Tests for Departure from Normality. Empirical Results for the Distributions of  $b_2$  and  $\sqrt{b_1}$ , *Biometrika* **60**, 613 (1973).