

Measurement Error Mitigation in Quantum Computers Through Classical Bit-Flip Correction

Lena Funcke¹, Tobias Hartung², Karl Jansen³, Stefan Kühn⁴, Paolo Stornati^{3,5}, and Xiaoyang Wang⁶

¹Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON N2L 2Y5, Canada

²Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom

³NIC, DESY Zeuthen, Platanenallee 6, 15738 Zeuthen, Germany

⁴Computation-based Science and Technology Research Center, The Cyprus Institute, 20 Kavafi Street, 2121 Nicosia, Cyprus

⁵Institut für Physik, Humboldt-Universität zu Berlin, Zum Großen Windkanal 6, D-12489 Berlin, Germany

⁶School of Physics, Peking University, 5 Yiheyuan Rd, Haidian District, Beijing 100871, China

(Dated: July 7, 2020)

We develop a classical bit-flip correction method to mitigate measurement errors on quantum computers. This method can be applied to any operator, any number of qubits, and any realistic bit-flip probability. We first demonstrate the successful performance of this method by correcting the noisy measurements of the ground-state energies of the longitudinal and transversal Ising models. We then generalize our results to arbitrary operators and test our method both numerically and experimentally on IBM quantum hardware. As a result, our correction method reduces the measurement error on the quantum hardware by up to one order of magnitude.

1 Introduction

Quantum computers have the potential to outperform classical computers in a variety of tasks ranging from combinatorial optimization over cryptography to machine learning. In particular, the prospect of being able to efficiently simulate quantum systems makes them a promising tool for solving quantum many-body problems in physics and chemistry. Despite recent progress, a large scale, fault tolerant digital quantum computer is still not available, and current intermediate scale devices suffer from a considerable level of noise. Although this limits the depth of the circuits that can be executed faithfully, these Noisy Intermediate-Scale Quantum (NISQ) devices [1] are already able to exceed the capabilities of clas-

sical computers in certain cases [2].

In the context of quantum many-body systems, a promising approach for exploiting the power of NISQ devices is variational quantum simulation (VQS), a class of hybrid quantum-classical algorithms for solving optimization problems [3, 4]. These make use of a feedback loop between a classical computer and a quantum coprocessor; the latter is used to efficiently evaluate the cost function for a given set of variational parameters, which are optimized on a classical computer based on the measurement outcome obtained from the quantum coprocessor. In particular, it has been experimentally demonstrated that VQS allows for finding both the ground state and low-lying excitations of systems relevant for condensed matter and particle physics as well as quantum chemistry [5–14].

NISQ devices are susceptible to errors, which can only be partially mitigated using error correction procedures (see, e.g., Refs. [6, 15–32]). The final qubit measurement is among the most error-prone operations on NISQ devices, with error rates ranging from 8% to 30% for current hardware [32]. These errors arise from bit flips, i.e., from erroneously recording an outcome as 0 given it was actually 1, and vice versa. Even though such errors can be mitigated for a small number of qubits, this method is exponentially costly with respect to the number of qubits [33].

The goal of this paper is to mitigate these type of measurement errors for any local operator, any number of qubits, and any bit-flip probability. We develop an efficient mitigation method that relies on cancellations of different erroneous measurement outcomes. This cancellation re-

sults from relative minus signs stemming from the default measurement basis of current hardware, $Z = \text{diag}(1, -1)$. The only input requirement for this approach is the knowledge of the different bit-flip probabilities during readout for each qubit. The simplicity of the method relies on the fact that measurement bit flips are uncorrelated between the qubits for multi-qubit measurements, at least in good approximation (see, e.g., Ref. [34]).

Our paper is organized as follows. In Sec. 2, we demonstrate the performance of our mitigation method by correcting the noisy energy histograms for the longitudinal and transversal Ising models. For simplicity, we assume all bit-flip probabilities to be equal. In Sec. 3, we generalize our method to different bit-flip probabilities and arbitrary operators. We now correct each bit flip directly at the measurement step, which allows us to mitigate the measurement errors of any expectation value of any operator. In Sec. 4, we demonstrate the experimental applicability of our method on IBM quantum hardware. In Sec. 5, we summarize and discuss our results.

2 Mitigation of measurement errors for energy histograms

Throughout this article, we focus on classical bit-flip errors and neglect any other sources of error, such as gate errors and decoherence. Thus, we assume that the quantum device prepares a pure state $|\psi\rangle$ for N qubits, which we measure in the computational basis

$$|\psi\rangle = \sum_{i=0}^{2^N-1} c_{i+1} |i\rangle. \quad (1)$$

Here, $|i\rangle$ is a shorthand notation for the computational-basis state corresponding to a bit string for the binary representation of i (e.g., for $N = 4$ the state $|5\rangle$ corresponds to $|0101\rangle$). A perfect, noise-free projective measurement would thus yield the bit string i with probability $|c_{i+1}|^2$, however, bit flips during readout can lead to erroneously recording $j \neq i$ instead. Throughout this article, we make the assumption that each bit flips independently of the others, which is a good approximation on current quantum hardware (see, e.g., Ref. [34]).

In this section, we assume for simplicity that all bit-flip probabilities are equal, $p(|0\rangle \rightarrow |1\rangle) =$

$p(|1\rangle \rightarrow |0\rangle)$, for all qubits. Using this assumption, we introduce our mitigation method and demonstrate its performance by correcting the noisy energy histograms of the longitudinal and transversal Ising models. We will turn to the more general case in Sec. 3, where we will discuss different bit-flip probabilities, arbitrary operators, and arbitrary (pure or mixed) states.

2.1 Binomial distribution of measurements

To start with, let us focus on a diagonal Hamiltonian \mathcal{H} . To evaluate the corresponding energy, $E = \langle\psi|\mathcal{H}|\psi\rangle$, on a quantum device, we have to (i) run the quantum circuit preparing the state $|\psi\rangle$, (ii) projectively measure the energy in the (computational) basis, and (iii) record the outcomes. We repeat this procedure a number of times and refer to the repetitions as the number of shots, s , throughout the rest of the article. Performing s shots, we record k correct results but $s - k$ incorrect results with $E^n \neq E$, where the superscript n denotes a noisy outcome. For simplicity, we assume that (i) a wrong measurement originates from a *single* bit flip with probability p and (ii) each bit flip yields the *same* deviation from E . We will see later that these assumptions will need to be modified in the presence of multi-qubit interactions, for example, for the longitudinal and transversal Ising models.

The probability of getting k correct measurement results is given by the probability mass function

$$f(k, s, 1 - p) = \binom{s}{k} (1 - p)^k p^{s-k}, \quad (2)$$

where p is the probability of *incorrectly* measuring the energy. The resulting noisy energy histograms can be described in terms of the number k of correct measurements,

$$\begin{aligned} E^n(k) &= E + (s - k)\Delta E^n \\ &= \begin{cases} E & \text{for } k = s, \\ E + s\Delta E^n & \text{for } k = 0, \end{cases} \end{aligned} \quad (3)$$

where ΔE^n is the deviation from E per bit flip. In terms of the bit-flip probability p , the resulting noisy expectation \mathbb{E} of the measured energy E^n reads

$$\begin{aligned} \mathbb{E}E^n &= E + sp\Delta E^n \\ &= \begin{cases} E & \text{for } p = 0, \\ E + s\Delta E^n & \text{for } p = 1. \end{cases} \end{aligned} \quad (4)$$

We note that “expectation” here means the expectation with respect to the probability p , which should not be confused with the quantum mechanical expectation value of the Hamiltonian, $\langle \psi | \mathcal{H} | \psi \rangle = E$. Thus, the expectation $\mathbb{E}\mathcal{H}^n$ is the expected value (as an operator to be measured subject to bit flips, see also Sec. 3) for the noisy Hamiltonian \mathcal{H}^n , while $\mathbb{E} \langle \psi | \mathcal{H}^n | \psi \rangle = \mathbb{E}E^n$ is the expected value for the noisy (quantum mechanical) expectation value $\langle \psi | \mathcal{H}^n | \psi \rangle = E^n$.

For a large number of shots s , the noisy energy histograms can be described by a normal distribution with mean energy $\mathbb{E}E^n$ given by Eq. (4). The only free parameter of this measurement noise model is ΔE^n , since s and p are known input parameters.

2.2 Mean energy vanishes for $p = 0.5$

The first step towards eliminating the free parameter ΔE^n is to study the dependence of this parameter on the bit-flip probability p , for example for $p = 0.5$. Let us consider the noise-free Hamiltonian \mathcal{H} acting on the state $|\psi\rangle = c_1|0\rangle + c_2|1\rangle$ and yielding the energy

$$\langle \psi | \mathcal{H} | \psi \rangle = (c_1^* \langle 0 | + c_2^* \langle 1 |) \mathcal{H} (c_1 | 0 \rangle + c_2 | 1 \rangle) = E. \quad (5)$$

The noisy measurement of this energy on the quantum hardware is performed along the basis $Z = \text{diag}(1, -1)$. We note that this noisy measurement yields $\mathbb{E}E^n = 0$ for $p = 0.5$, due to the opposite signs of the terms resulting from bit-flipping the terms in Eq. (5).

Let us demonstrate this for a simple example, $\mathcal{H}_X = X = HZH$, where H is the Hadamard gate, and study the possible outcomes of the energy measurements in the single-qubit case:

- The absence of any bit flip gives the true energy of the noise-free Hamiltonian:

$$\begin{aligned} \langle \psi | \mathcal{H}_X | \psi \rangle &= \langle \psi | HZH | \psi \rangle \\ &= [c_1^*(\langle 0 | + \langle 1 |) + c_2^*(\langle 0 | - \langle 1 |)] \\ &\quad Z [c_1(|0\rangle + |1\rangle) + c_2(|0\rangle - |1\rangle)] \\ &= |c_1 + c_2|^2 - |c_1 - c_2|^2 = E. \end{aligned} \quad (6)$$

- The bit flip $|0\rangle \rightarrow |1\rangle$, $|1\rangle \rightarrow |1\rangle$ changes one sign: $\langle \psi | \mathcal{H}_X^n | \psi \rangle = -|c_1 + c_2|^2 - |c_1 - c_2|^2$.
- The bit flip $|0\rangle \rightarrow |0\rangle$, $|1\rangle \rightarrow |0\rangle$ changes the other sign: $\langle \psi | \mathcal{H}_X^n | \psi \rangle = |c_1 + c_2|^2 + |c_1 - c_2|^2$.

- The bit flip $|0\rangle \rightarrow |1\rangle$, $|1\rangle \rightarrow |0\rangle$ changes both signs: $\langle \psi | \mathcal{H}_X^n | \psi \rangle = -|c_1 + c_2|^2 + |c_1 - c_2|^2$ and thus yields the outcome $-E$.

For $p = 0.5$, each of these four possible outcomes has the same probability $p^2 = 0.25$, and thus cancellation yields $\mathbb{E}E^n = 0$. This result holds true for any operator and any number of qubits, as we will show later in Eq. (45).

2.3 ΔE^n for non-interacting Hamiltonians

The next step towards eliminating the free parameter ΔE^n is to examine the four possible measurement outcomes from the previous section for *any* bit-flip probability p . We observe that the second and third outcomes have *opposite* signs and *equal* probability and thus cancel, given the above assumption that $p(|0\rangle \rightarrow |1\rangle) = p(|1\rangle \rightarrow |0\rangle)$. For N qubits, one can similarly show that among the 4^N possible measurement outcomes, *all* outcomes cancel apart from the ones corresponding to no bit flip and all bit flips. This justifies our previous assumption that we either measure a correct energy with probability $1-p$ or an incorrect energy with probability p . Crucially, the latter probability is *not* given by p^{2N} as one might expect at first sight. Thus, each incorrect measurement yields the same deviation from the correct energy of $-2E$ with the same probability p . This can be seen by evaluating the probabilities of the four different outcomes above:

- The absence of any bit flip, $|0\rangle \xrightarrow{1-p} |0\rangle$, $|1\rangle \xrightarrow{1-p} |1\rangle$, gives $\langle \psi | \mathcal{H}_X | \psi \rangle = E$ with probability $(1-p)^2$.
- The “mixed” bit flips $|0\rangle \xrightarrow{p} |1\rangle$, $|1\rangle \xrightarrow{1-p} |1\rangle$ and $|0\rangle \xrightarrow{1-p} |0\rangle$, $|1\rangle \xrightarrow{p} |0\rangle$ give $\langle \psi | \mathcal{H}_X^n | \psi \rangle = 0$ with a combined probability of $2p(1-p)$.
- The “total” bit flip $|0\rangle \xrightarrow{p} |1\rangle$, $|1\rangle \xrightarrow{p} |0\rangle$ gives $\langle \psi | \mathcal{H}_X^n | \psi \rangle = -E$ with probability p^2 .

This yields the simple relation for the mean energy

$$\begin{aligned} \mathbb{E}E^n &= (1-p)^2 E + 2p(1-p)0 + p^2(-E) \\ &= E - 2pE = (1-2p)E. \end{aligned} \quad (7)$$

Combining Eqs. (4) and (7) we find for the parameter ΔE^n

$$\mathbb{E}E^n = E + sp\Delta E^n \quad \leftrightarrow \quad \Delta E^n = -\frac{2E}{s}, \quad (8)$$

where ΔE^n is normalized by the number of shots s , i.e., the number of evaluations of the energy (5) required to produce the energy histogram. For $p = 1$, the first three possible measurement outcomes have zero probability, independently of any cancellations, and only the last outcome with $\langle \psi | \mathcal{H}_X^n | \psi \rangle = -E$ contributes.

As we will discuss in the next subsection, Eq. (8) *only* applies to non-interacting Hamiltonians, i.e., without any multi-qubit interaction terms. For example, for the Hamiltonians $\mathcal{H}_X = h \sum_{i=1}^N X_i$ or $\mathcal{H}_Z = h \sum_{i=1}^N Z_i$ with the ground-state energy $E_0 = -Nh$, we would get $\Delta E_0^n = 2Nh/s$ when measuring the ground-state energy. Thus, after measuring the noisy expectation value of any (trivial) non-interacting Hamiltonian on a quantum computer, Eq. (8) allows us to predict the corresponding true energy.

2.4 ΔE^n for interacting Hamiltonians

For two-qubit interaction terms in the Hamiltonian, e.g., for $\mathcal{H}_{ZZ} = J \sum_{i=1}^N Z_i Z_{i+1}$, our previous considerations need to be modified in two ways: first, we observe that the *one-qubit* bit flips from the previous subsection give the same contribution to the mean energy as before, but now with a probability of $2p(1-p)$ instead of p^2 . This is because the one-qubit “total” bit flips yield $\langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle = -E$. Here, “one-qubit “total” bit flip” means that one of the two qubits experiences a bit flip during readout ($|0\rangle \rightarrow |1\rangle$, $|1\rangle \rightarrow |0\rangle$), while the other qubit has no bit flip ($|0\rangle \rightarrow |0\rangle$, $|1\rangle \rightarrow |1\rangle$). Second, the mean energy receives small $\mathcal{O}(p^2)$ corrections since the parameter ΔE^n becomes p -dependent for the interacting case. These $\mathcal{O}(p^2)$ corrections come from the *two-qubit* bit flips and have the *opposite* sign of the $\mathcal{O}(p)$ terms, because the two minus signs from the measurement bases Z_1 and Z_2 cancel. Indeed, the two-qubit “total” bit flips, i.e., $|0\rangle \rightarrow |1\rangle$ and $|1\rangle \rightarrow |0\rangle$ for both qubits, yield $\langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle = E$ with probability p^2 .

Let us demonstrate the latter for the simple two-qubit Hamiltonian $\mathcal{H}_{11} = 1_1 1_2 = Z_1 Z_1 Z_2 Z_2$, which gives

$$\begin{aligned} \langle \psi | \mathcal{H}_{11} | \psi \rangle &= \langle \psi | Z_1 Z_1 Z_2 Z_2 | \psi \rangle \\ &= [c_1^* \langle 00 | + c_2^* \langle 01 | - c_3^* \langle 10 | - c_4^* \langle 11 |] Z_1 Z_2 \\ &\quad [c_1 | 00 \rangle - c_2 | 01 \rangle + c_3 | 10 \rangle - c_4 | 11 \rangle] \\ &= |c_1|^2 + |c_2|^2 + |c_3|^2 + |c_4|^2 = E \end{aligned} \quad (9)$$

without any bit flip. For two-qubit bit flips with $p = 1$, we obtain the same result and thus recover the true energy E ,

$$\begin{aligned} \langle \psi | \mathcal{H}_{11}^n | \psi \rangle &= \langle \psi | Z_1^n Z_1^n Z_2^n Z_2^n | \psi \rangle \\ &= [c_1^* \langle 11 | + c_2^* \langle 10 | - c_3^* \langle 01 | - c_4^* \langle 00 |] Z_1 Z_2 \\ &\quad [c_1 | 11 \rangle - c_2 | 10 \rangle + c_3 | 01 \rangle - c_4 | 00 \rangle] \\ &= |c_1|^2 + |c_2|^2 + |c_3|^2 + |c_4|^2 = E, \end{aligned} \quad (10)$$

since the minus signs from the Z -matrices cancel.

The contributions from “mixed” bit flips, such as all basis states $|b_1 b_0\rangle$ flipping to $|11\rangle$, cancel for any p due to opposite signs and equal probabilities, just as in the non-interacting case. Therefore, the “total” two-qubit bit flips as discussed in Eq. (10) have a probability of p^2 instead of p^{4N} . The same is given for the \mathcal{H}_{ZZ} Hamiltonian, which yields for the total mean energy

$$\begin{aligned} \mathbb{E}E^n &= (1-p)^2 E + 2p(1-p)(-E) + p^2 E \\ &= E - 4pE + 4p^2 E = (1-2p)^2 E. \end{aligned} \quad (11)$$

Thus, the parameter ΔE^n now has two contributions,

$$\mathbb{E}E^n = E + sp\Delta E^n \Leftrightarrow \Delta E^n = -\frac{4E}{s}(1-p). \quad (12)$$

Equations (11) and (12) imply that the two-qubit interacting Hamiltonian yields the correct energy $\mathbb{E}E^n = E$ for *both* $p = 0$ and $p = 1$, in contrast to the non-interacting case where $p = 1$ gave $\mathbb{E}E^n = -E$ (see Eq. (7)). Moreover, $\mathbb{E}E^n = 0$ is still given for $p = 0.5$.

2.5 Prediction for the longitudinal Ising model

We now apply our previous results to predict the noisy ground-state energy histograms of the longitudinal Ising (LI) model with periodic boundary conditions. The Hamiltonian of this model reads

$$\mathcal{H}_{\text{LI}} = J \sum_{i=1}^N Z_i Z_{i+1} + h \sum_{i=1}^N Z_i, \quad (13)$$

where we assume $J < 0$ and $h > 0$ and we identify $N + 1$ with 1. The true ground-state energy of the model is

$$E_0 = E_0^J + E_0^h = NJ - Nh, \quad (14)$$

which is the sum of the individual ground-state energies for $h = 0$ and $J = 0$, which we call E_0^J and E_0^h , respectively.

Based on Eqs. (7) and (11), the mean energy of the noisy ground-state energy histograms reads

$$\begin{aligned}\mathbb{E}E_0^n &= E_0 + sp\Delta E_0^n = (1-p)^2(E_0^J + E_0^h) \\ &\quad + 2p(1-p)(-E_0^J) + p^2(E_0^J - E_0^h) \quad (15) \\ &= (1-2p)E_0^h + (1-2p)^2E_0^J.\end{aligned}$$

Thus, the parameter ΔE_0^n now has three contributions,

$$\Delta E_0^n = -\frac{1}{s} \left(2E_0^h + 4E_0^J - 4pE_0^J \right). \quad (16)$$

For $|J| = |h|$ and $p = 1$, we obtain $\mathbb{E}E_0^n = 0$ in Eq. (15). This is because the two terms in the Hamiltonian (13) contribute equally to E_0 and cancel for $p = 1$ due to their opposite signs, $\mathbb{E}E_0^{J,n} = E_0^J$ and $\mathbb{E}E_0^{h,n} = -E_0^h$.

Our method allows us to predict the variance of the noisy energy histograms as well, as we will explain in detail in Sec. 3.5 and Appendix A. Based on these results, Fig. 1 shows the energy histograms for the ground state of \mathcal{H}_{LI} with different choices of the parameters N , h , s , and p , where we measure the ground state 2048 times for each parameter combination. The noise model, with the mean energy from Eq. (15) and the variance from Eq. (72), agrees with the data for all the parameters. Indeed, our prediction (solid orange line in Fig. 1) perfectly matches the fitted data of the histogram (dashed black line). This allows to retrieve the true ground state energy E_0 (dashed green line) using Eq. (15).

2.6 Prediction for the transversal Ising model

Next, we apply our results to the ground-state energy of the transversal Ising (TI) model with the Hamiltonian

$$\mathcal{H}_{\text{TI}} = J \sum_{i=1}^N Z_i Z_{i+1} + h \sum_{i=1}^N X_i, \quad (17)$$

where we again assume $J < 0$ and $h > 0$ and periodic boundary conditions. The true ground-state energy can be derived as [35–38]

$$\begin{aligned}E_0 &= -\frac{1}{2} \sum_k \gamma (\alpha^2 + 4\beta^2) \\ &= -\frac{1}{2} \sum_k \gamma \left[4h^2 + 4J^2 - 8Jh \cos\left(\frac{2\pi k}{N}\right) \right], \quad (18)\end{aligned}$$

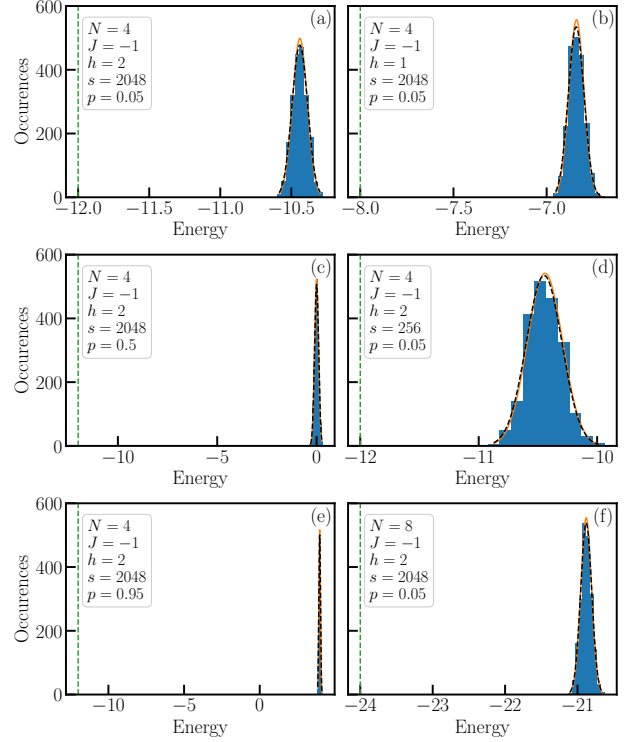


Figure 1: Energy histograms for the longitudinal Ising model. The vertical dashed green line indicates the true ground-state energy, the solid orange line the prediction from Eqs. (15) and (72), and the dashed black line a fit to the data. The left column corresponds to $N = 4$, $J = -1$, $h = 2$, $s = 2048$ with (a) $p = 0.05$, (c) $p = 0.50$, and (e) $p = 0.95$. The right column shows varied N , h , and s : (b) $h = 1$, (d) $s = 256$, and (f) $N = 8$.

where the sum runs from $k = -\left(\frac{N-1}{2}\right)$ to $\left(\frac{N-1}{2}\right)$ and the constants α , β , and γ are defined as

$$\begin{aligned}\alpha &= 2h - 2J \cos\left(\frac{2\pi k}{N}\right), \\ \beta &= J \sin\left(\frac{2\pi k}{N}\right), \\ \gamma &= \frac{\text{sign}(\alpha)}{\alpha} \sqrt{\frac{\alpha^2}{\alpha^2 + 4\beta^2}}.\end{aligned} \quad (19)$$

Just as in Eq. (15), the mean energy of the noisy ground-state energy histograms receives three different contributions,

$$\mathbb{E}E_0^n = (1-p)^2 E_1 + 2p(1-p) E_2 + p^2 E_3. \quad (20)$$

The probabilities of the three different terms in Eqs. (15) and (20) are the same because they are determined by the number of interacting qubits in the different terms of the respective Hamiltonian. However, the measurement outcomes E_i in Eq. (20) deviate from the ones in Eq. (15) because E_0 in Eq. (18) is not simply the sum of

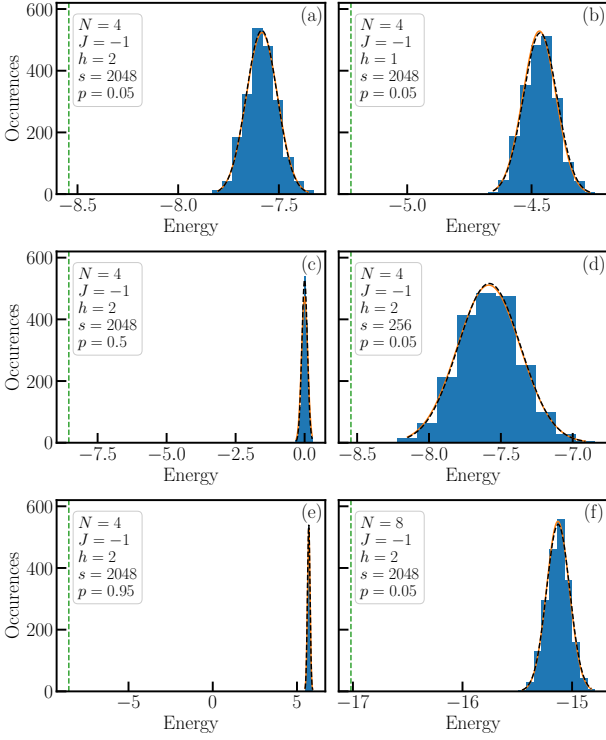


Figure 2: Energy histograms for the transversal Ising model. The vertical dashed green line indicates the true ground-state energy, the solid orange line the prediction and the dashed black line a fit to the data. The left column corresponds to $N = 4$, $J = -1$, $h = 2$, $s = 2048$ with (a) $p = 0.05$, (c) $p = 0.50$, and (e) $p = 0.95$. The right column shows varied N , h , and s : (b) $h = 1$, (d) $s = 256$, and (f) $N = 8$.

the J - and h -dependent ground-state energies E_0^J and E_0^h as in Eq. (14).

The different measurement outcomes E_i in Eq. (20) can be derived in the following way. First, we know that $E_1 = E_0$. Second, we know that $\mathbb{E}E^n$ vanishes for $|J| = |h|$ and $p = 1$ because the two terms in the Hamiltonian (17) contribute equally to E_0 and thus cancel for $p = 1$. This cancellation happens due to opposite signs of the non-interacting and interacting energy contributions in case of a total bit flip, as discussed above. In particular, any mixed terms, such as the mixed Jh -term in Eq. (18), vanish for $p = 1$, as also discussed above. This fixes E_3 . Third, we know that $\mathbb{E}E^n(p = 0.5) = 0$, so we can find E_2 by solving Eq. (20) for $p = 0.5$ and inserting the known expressions for E_1 and E_3 . In total, we obtain

$$\begin{aligned} E_1 &= E_0^J + E_0^h, \\ E_2 &= -E_0^J, \\ E_3 &= E_0^J - E_0^h, \end{aligned} \quad (21)$$

just as in Eq. (15), but with E_0^J and E_0^h given by

$$\begin{aligned} E_0^J &= -\frac{1}{2} \sum_k \gamma \left[4J^2 - 4Jh \cos\left(\frac{2\pi k}{N}\right) \right], \\ E_0^h &= -\frac{1}{2} \sum_k \gamma \left[4h^2 - 4Jh \cos\left(\frac{2\pi k}{N}\right) \right]. \end{aligned} \quad (22)$$

Thus, the mean energy in Eq. (20) can be brought into a similar form as the true ground-state energy in Eq. (18),

$$\begin{aligned} \mathbb{E}E_0^n &= (1 - 2p)E_0^h + (1 - 2p)^2 E_0^J \\ &= -\frac{1}{2} \sum_k \gamma \left[(1 - 2p)4h^2 + (1 - 2p)^2 4J^2 \right. \\ &\quad \left. - (1 - 3p + 2p^2)8Jh \cos\left(\frac{2\pi k}{N}\right) \right]. \end{aligned} \quad (23)$$

The resulting parameter ΔE_0^n is identical to the one for the longitudinal Ising model in Eq. (16) but with E_0^J and E_0^h given by Eq. (22). We note that for the longitudinal Ising model, ΔE_0^n rises strictly linearly with N . For the transversal Ising model, the sum over k yields N contributions to each E_i in Eq. (21), which are equal for E_3 but differ for E_1 and E_2 due to the N -dependence of the cosine. Thus, $\Delta E_0^n(N)$ only becomes approximately linear for large N , where these small differences average out.

In Fig. 2, we plot the energy histograms for the ground state of \mathcal{H}_{TI} with different N , h , s , and p , where we again measure the ground state 2048 times for each parameter combination. As before, the noise model with the mean energy from Eq. (23) and the variance from Eq. (72) agrees with the data for any choice of parameters we study. Note that the variance is larger compared to the longitudinal case in Fig. 1, because the measurement Z -basis is not an eigenbasis of the X_i operator. Thus, the histograms are wider for the transversal case.

3 Mitigation of measurement errors for arbitrary operators

In this section, we generalize our previous results to *arbitrary* operators acting on N different qubits $q = 1, \dots, N$ with $p(|0\rangle \rightarrow |1\rangle) \neq p(|1\rangle \rightarrow |0\rangle)$. This generalization is greatly aided by a change in point of view. Whereas previously, we treated the bit-flip error as part of the measurement process, i.e., we projectively measured the

state $|\psi\rangle$ onto a basis bit string and randomly flipped the bits of this bit string, we now consider the bit flip as part of the operator. In other words, the measurement process no longer includes the bit flips and instead we consider *random* operators to be measured. While this point of view is conceptually very different, we will demonstrate that these random operators yield a distribution of measurements that precisely coincides with the distribution of measurements for a non-random operator subject to bit flips.

Our analysis will be split into four parts. First, we will consider a single Z operator acting on a single qubit, while allowing for different flip probabilities, $p(|0\rangle \rightarrow |1\rangle) \neq p(|1\rangle \rightarrow |0\rangle)$, in Sec. 3.1. In particular, we will compute the operator's expectation as a random operator subject to classical bit flips during measurement. This computation will be the stepping stone to subsequently construct the expectations for noisy measurements of $Z_N \otimes \cdots \otimes Z_1$ operators with $N > 1$ in Sec. 3.2. This construction is inductive with respect to N and will allow us to construct a classical bit-flip correction procedure for the noisy measurement of $Z_N \otimes \cdots \otimes Z_1$. It is important to note that the classical bit-flip correction procedure can be pre-processed (replacing the operator to be measured) as well as post-processed (measuring the necessary information first and then extracting the bit-flip corrected expectation values from the measured data).

In Sec. 3.3, we will consider the special case of equal bit-flip probabilities for all qubits, to compare the results directly to Sec. 2. In Sec. 3.4, we will generalize the classical bit-flip correction procedure to arbitrary operators that are measured from bit-string distributions of the state $|\psi\rangle$. We note that Sec. 3.4 denotes a change in measurement paradigm compared to the previous sections, which affects the variance of the histogram means. We will discuss the different measurement paradigms in detail in Sec. 3.5 and return to the transversal Ising model for an explicit illustration. The derivation of the corresponding variances is provided in Appendix A.

3.1 Measurement of a single Z operator

3.1.1 Prediction for the noisy expectation value

For $N = 1$, the noise-free operator Z_q gets replaced by the random noisy operator Z_q^n , which

can take the values

- Z_q with probability $(1 - p_{q,0})(1 - p_{q,1})$,
- $-\mathbb{1}_q$ with probability $p_{q,0}(1 - p_{q,1})$,
- $\mathbb{1}_q$ with probability $(1 - p_{q,0})p_{q,1}$,
- or $-Z_q$ with probability $p_{q,0}p_{q,1}$.

Here, $p_{q,b}$ is the probability of flipping the qubit q given that it is in the state $b = |0\rangle$ or $|1\rangle$. For example, $p_{3,0}$ is the probability of flipping $|0\rangle \rightarrow |1\rangle$ for qubit 3.

Then, we obtain the noisy expectation $\mathbb{E}Z_q^n$ for the random operator Z_q^n ,

$$\mathbb{E}Z_q^n = (1 - p_{q,0} - p_{q,1})Z_q + (p_{q,1} - p_{q,0})\mathbb{1}_q, \quad (24)$$

which reduces to Eq. (7) for $p_{q,0} = p_{q,1} =: p$. As before, “expectation” here means the expectation with respect to the bit-flip probabilities, which should not be confused with the quantum mechanical expectation value $\langle\psi|O|\psi\rangle$ of the operator O . The expectation $\mathbb{E}O^n$ is the expected value (as an operator) for the noisy operator O^n , while $\mathbb{E}\langle\psi|O^n|\psi\rangle$ is the expected value for the noisy (quantum mechanical) expectation value $\langle\psi|O^n|\psi\rangle$ of the operator O^n .

3.1.2 Density matrix description and visualization of measurement noise

For the single-qubit case it is instructive to express our results in terms of density matrices. Starting from an arbitrary single-qubit density operator

$$\rho = (\mathbb{1} + \vec{r} \cdot \vec{\sigma})/2, \quad (25)$$

where \vec{r} is a real vector with $\|\vec{r}\| \leq 1$ and $\vec{\sigma}$ is the vector containing the Pauli matrices, any quantum channel acting on the state ρ is as an affine linear map

$$\vec{r} \mapsto \vec{r}' = M\vec{r} + \vec{c}, \quad (26)$$

where M is a 3×3 real matrix and \vec{c} is a constant real vector [39]. In particular, a noise-free projective measurement in the computational basis corresponds to a unital channel with $M = \text{diag}(0, 0, 1)$ and $\vec{c} = 0$. For an arbitrary pure

single-qubit state, $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, with density operator

$$\rho = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{pmatrix}, \quad (27)$$

such a projective measurement yields the classical mixture $\rho_c = \text{diag}(|\alpha|^2, |\beta|^2)$.

In case of a noisy measurement, the bit flips change the classical state that one obtains after

the measurement. As discussed above, (i) with probability $(1 - p_0)(1 - p_1)$ we obtain the original state, (ii) with probability $p_0(1 - p_1)$ the $|0\rangle$ flips to a $|1\rangle$, (iii) with probability $(1 - p_0)p_1$ the $|1\rangle$ flips to a $|0\rangle$, and (iv) with probability p_0p_1 both measurement outcomes flip. The resulting classical state can be expressed as a convex linear combination of the different outcomes

$$\begin{aligned} \rho_c^n &= \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix} (1 - p_0)(1 - p_1) + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} p_0(1 - p_1) + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} p_1(1 - p_0) + \begin{pmatrix} |\beta|^2 & 0 \\ 0 & |\alpha|^2 \end{pmatrix} p_1 p_0 \\ &= \begin{pmatrix} (1 - p_0 - p_1)|\alpha|^2 + p_1 & 0 \\ 0 & (1 - p_0 - p_1)|\beta|^2 + p_0 \end{pmatrix}. \end{aligned} \quad (28)$$

The expectation value of the Z_q operator then reads

$$\begin{aligned} \langle Z_q^n \rangle &= \text{Tr}(\rho_c^n Z_q) \\ &= (1 - p_0 - p_1)(|\alpha|^2 - |\beta|^2) + p_1 - p_0, \end{aligned} \quad (29)$$

which is equivalent to computing the quantum expectation value of Eq. (24), $\text{Tr}(\rho \mathbb{E} Z_q^n)$.

Moreover, we see that Eq. (28) arises from the original density operator ρ by applying the quantum channel

$$M^n = \begin{pmatrix} 0 & & \\ & 0 & \\ & & 1 - p_0 - p_1 \end{pmatrix}, \quad \vec{c}^n = \begin{pmatrix} 0 \\ 0 \\ p_1 - p_0 \end{pmatrix}. \quad (30)$$

From the equation above, it is apparent that the channel is no longer unital. For $p_0 = p_1$ all quantum states ρ in the equatorial plane of the Bloch sphere, corresponding to $r_z = 0$, are unaffected. The closer the state is to the polar region of the sphere, the more pronounced is the effect of the measurement errors. Compared to the classical state ρ_c obtained from a noise-free projective measurement, the Bloch vector corresponding to ρ_c^n is shortened because of M^n , and translated along the z axis by \vec{c}^n (see Fig. 3). Moreover, for $p_0 + p_1 = 1$ the channel maps any state to the same point inside the Bloch sphere. As a result, our mitigation method is not applicable to the

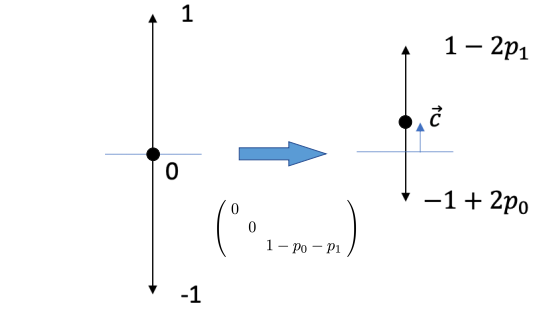


Figure 3: Left panel: Possible range of Bloch vectors of the classical states ρ_c obtained from a noise-free projective measurement in the computational basis. Right panel: Deformed range of Bloch vectors corresponding to the classical state ρ_c^n resulting from a measurement in the presence of readout noise.

special case of $p_0 + p_1 = 1$, which will be further discussed in the next section.

3.2 Measurement of $Z_N \otimes \dots \otimes Z_1$ operators

Going beyond $N = 1$, we can now compute the noisy expectations for arbitrary operators $Z_N \otimes \dots \otimes Z_1$ with $N > 1$. For this, we assume that the expectations of the individual operators can be measured *independently* of each other. In this case, the noisy expectation of the tensor product $Z_N^n \otimes \dots \otimes Z_1^n$ equals the tensor product of the

individual noisy expectations,

$$\mathbb{E}(Z_N^n \otimes \cdots \otimes Z_1^n) = \mathbb{E}Z_N^n \otimes \cdots \otimes \mathbb{E}Z_1^n. \quad (31)$$

Equation (31) can be proven by considering two different noisy operators O_1^n and O_2^n acting on different qubits, and defining their conditional expectations $\mathbb{E}^{O_1^n} O_1^n =: \Omega_1$ and $\mathbb{E}^{O_2^n} O_2^n =: \Omega_2$. The term “conditional” here means that the expectations are only taken with respect to the qubits the operators are acting on, leaving the other qubits untouched. Now, if we assume O_1^n to take the values χ_α with probabilities p_α , for example $O_1^n = Z_q^n$ could take $\chi_\alpha \in \{Z_q, -\mathbb{1}_q, \mathbb{1}_q, -Z_q\}$ as above, then we observe

$$\begin{aligned} \mathbb{E}(O_1^n \otimes O_2^n) &= \sum_{\alpha} p_{\alpha} \mathbb{E}^{O_2^n} (\chi_{\alpha} \otimes O_2^n) \\ &= \sum_{\alpha} p_{\alpha} \chi_{\alpha} \otimes \Omega_2 = \Omega_1 \otimes \Omega_2, \end{aligned} \quad (32)$$

which directly yields Eq. (31).

Our final goal is to *reconstruct* the noise-free quantum mechanical expectation value $\langle \psi | O | \psi \rangle$ of an arbitrary operator $O = O_N \otimes \cdots \otimes O_1 \in \{\mathbb{1}, Z\}^{\otimes N}$ from its noisy measurement. To this end, we need to find a matrix ω^{-1} that multiplies the noisy expectations $\mathbb{E} \langle \psi | O^n | \psi \rangle$ and yields the noise-free expectation values $\langle \psi | O | \psi \rangle$, as we will achieve in Eq. (41).

For this, we first express the noisy expectation of $Z_N^n \otimes \cdots \otimes Z_1^n$ in Eq. (31) in terms of the noise-free operators $O_N \otimes \cdots \otimes O_1$. Using Eq. (24), we find

$$\begin{aligned} \mathbb{E}(Z_N^n \otimes \cdots \otimes Z_1^n) \\ = \sum_{O \in \{\mathbb{1}, Z\}^{\otimes N}} \gamma(O_N) O_N \otimes \cdots \otimes \gamma(O_1) O_1, \end{aligned} \quad (33)$$

where the coefficients γ in front of the noise-free operators are defined as

$$\gamma(O_q) := \begin{cases} 1 - p_{q,0} - p_{q,1} & \text{for } O_q = Z_q, \\ p_{q,1} - p_{q,0} & \text{for } O_q = \mathbb{1}_q. \end{cases} \quad (34)$$

In order to construct the value of $\mathbb{E}(Z_Q^n \otimes \cdots \otimes Z_1^n)$ in Eq. (33) inductively, it is advantageous to choose the “lexicographic order” \preceq for both the noise-free operators $O \in \{\mathbb{1}, Z\}^{\otimes N}$ and the noisy operators $O^n \in \{\mathbb{1}, Z\}^{\otimes N}$,

$$\begin{aligned} \mathbb{1}_3 \otimes \mathbb{1}_2 \otimes \mathbb{1}_1 &\preceq \mathbb{1}_3 \otimes \mathbb{1}_2 \otimes Z_1 \\ &\preceq \mathbb{1}_3 \otimes Z_2 \otimes \mathbb{1}_1 \preceq \mathbb{1}_3 \otimes Z_2 \otimes Z_1 \\ &\preceq Z_3 \otimes \mathbb{1}_2 \otimes \mathbb{1}_1 \preceq Z_3 \otimes \mathbb{1}_2 \otimes Z_1 \\ &\preceq Z_3 \otimes Z_2 \otimes \mathbb{1}_1 \preceq Z_3 \otimes Z_2 \otimes Z_1 \preceq \dots \end{aligned} \quad (35)$$

This choice implies $O_N \otimes \cdots \otimes O_1 \preceq Z_N \otimes \cdots \otimes Z_1$ and will later ensure that the matrix ω in Eq. (41) is a lower triangular matrix, which is invertible as long as none of its diagonal entries vanish. To determine the matrix ω , we need to generalize Eq. (33) to arbitrary noisy operators,

$$\begin{aligned} \mathbb{E}(O_N^n \otimes \cdots \otimes O_1^n) \\ = \sum_{O \in \{\mathbb{1}, Z\}^{\otimes N}} \Gamma(O_N | O_N^n) O_N \otimes \cdots \otimes \Gamma(O_1 | O_1^n) O_1, \end{aligned} \quad (36)$$

where the coefficients Γ in front of the noise-free operators are now defined as

$$\Gamma(O_q | O_q^n) = \begin{cases} \gamma(O_q) & \text{for } O_q^n = Z_q^n \\ 1 & \text{for } O_q = \mathbb{1}_q \wedge O_q^n = \mathbb{1}_q^n \\ 0 & \text{for } O_q = Z_q \wedge O_q^n = \mathbb{1}_q^n. \end{cases} \quad (37)$$

Using this definition, we can now define the matrix ω as

$$\begin{aligned} \omega(O | O^n) &:= \prod_{q=1}^N \Gamma(O_q | O_q^n), \\ \omega &:= (\omega(O | O^n))_{O^n, O \in \{\mathbb{1}, Z\}^{\otimes N}}. \end{aligned} \quad (38)$$

It is important to note that $O^n \prec O$ implies $\omega(O | O^n) = 0$. In other words, ω is a lower triangular matrix and therefore invertible as long as none of its diagonal entries vanish. The diagonal entries are $\prod_{q=1}^N \Gamma(O_q | O_q^n)$ and thus can only vanish if one of the $\gamma(Z_q)$ vanishes, i.e., ω is invertible as long as $\forall q : p_{q,0} + p_{q,1} \neq 1$. If that is the case, then we obtain the bit-flip corrected operators

$$(O)_{O \in \{\mathbb{1}, Z\}^{\otimes N}} = \omega^{-1} (\mathbb{E} O^n)_{O^n \in \{\mathbb{1}, Z\}^{\otimes N}}. \quad (39)$$

In particular, for $O = Z_2 \otimes Z_1$, we obtain

$$\begin{aligned} Z_2 \otimes Z_1 &= \frac{1}{\gamma(Z_2)\gamma(Z_1)} \mathbb{E}(Z_2^n \otimes Z_1^n) \\ &\quad - \frac{\gamma(\mathbb{1}_1)}{\gamma(Z_2)\gamma(Z_1)} \mathbb{E}(Z_2^n) \otimes \mathbb{1}_1 \\ &\quad - \frac{\gamma(\mathbb{1}_2)}{\gamma(Z_2)\gamma(Z_1)} \mathbb{1}_2 \otimes \mathbb{E}(Z_1^n) \\ &\quad + \frac{\gamma(\mathbb{1}_2)\gamma(\mathbb{1}_1)}{\gamma(Z_2)\gamma(Z_1)} \mathbb{1}_2 \otimes \mathbb{1}_1. \end{aligned} \quad (40)$$

We can now evaluate Eq. (39) on an arbitrary state $|\psi\rangle$ to find the bit-flip corrected expectation

values. For all $O \in \{\mathbb{1}, Z\}^{\otimes N}$, we find

$$\langle \psi | O | \psi \rangle = \sum_{O^n \in \{\mathbb{1}, Z\}^{\otimes N}} \omega_{O, O^n}^{-1} \mathbb{E} \langle \psi | O^n | \psi \rangle. \quad (41)$$

In Fig. 4, we show the relative error for the bit-flip corrected expectation value of $\langle \psi | Z_N^n \otimes \cdots \otimes Z_1^n | \psi \rangle$, as retrieved from histogram data using Eq. (41), compared to the bit-flip free expectation value $\langle \psi | Z_N \otimes \cdots \otimes Z_1 | \psi \rangle$:

$$\frac{|\langle \psi | Z_N^n \otimes \cdots \otimes Z_1^n | \psi \rangle - \langle \psi | Z_N \otimes \cdots \otimes Z_1 | \psi \rangle|}{|\langle \psi | Z_N \otimes \cdots \otimes Z_1 | \psi \rangle|}. \quad (42)$$

We also plot the standard deviation of this relative error, alternatively to plotting the error bars. Figure 4 also contains a fit $y(s) = Cs^{-\alpha}$ of the relative error in Eq. (42), where s is again the number of shots, i.e., the number of $\langle \psi | Z_N^n \otimes \cdots \otimes Z_1^n | \psi \rangle$ evaluations to produce the histogram. In particular, the fit indicates Monte-Carlo type convergence $\alpha \approx 1/2$ for $N \in \{1, 2, 3, 4\}$. Figure 4 has been generated using 4096 random states $|\psi\rangle$ satisfying $|\langle \psi | Z_N \otimes \cdots \otimes Z_1 | \psi \rangle| \geq .25$ to avoid dividing by small numbers when computing relative errors. For each $|\psi\rangle$ we randomly chose the bit-flip probabilities $p_{q,b}$ uniformly in $(.05, .25)$.

3.3 Measurement of $Z_N \otimes \cdots \otimes Z_1$ operators assuming equal bit-flip probabilities

To compare the results of the previous two subsections with the results obtained in Sec. 2, we now set all bit-flip probabilities $p_{q,b} = p$ to be equal. For the case $N = 1$, the expectation $\mathbb{E}Z_q^n$ in Eq. (24) reduces to

$$\mathbb{E}Z_q^n = (1 - 2p)Z_q, \quad (43)$$

in agreement with Eq. (7). For $N > 1$, the expectation in Eq. (31) reduces to

$$\mathbb{E}(Z_N^n \otimes \cdots \otimes Z_1^n) = (1 - 2p)^N Z_N \otimes \cdots \otimes Z_1, \quad (44)$$

which yields Eq. (11) for $N = 2$. This implies that the matrix ω in Eq. (38) becomes diagonal with

$$\mathbb{E}(O_N^n \otimes \cdots \otimes O_1^n) = (1 - 2p)^{\#Z(O)} \times O_N \otimes \cdots \otimes O_1, \quad (45)$$

where $\#Z(O)$ is the number of terms $O_q = Z_q$ in the tensor product $O = O_N \otimes \cdots \otimes O_1$. In

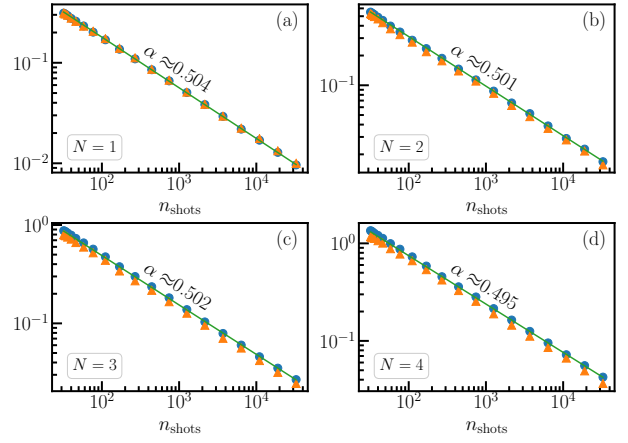


Figure 4: Relative errors (blue dots) and standard deviations (orange triangles) for the bit-flip corrected expectation values of $\langle \psi | Z_N^n \otimes \cdots \otimes Z_1^n | \psi \rangle$, as retrieved from histogram data using Eq. (41), compared to the “true” bit-flip free expectation values of $\langle \psi | Z_N \otimes \cdots \otimes Z_1 | \psi \rangle$, see Eq. (42). Shown are the four different operators Z_1 (a), $Z_2 \otimes Z_1$ (b), $Z_3 \otimes Z_2 \otimes Z_1$ (c), and $Z_4 \otimes Z_3 \otimes Z_2 \otimes Z_1$ (d). The average relative errors are fitted with a power law, $y(s) \propto s^{-\alpha}$ (green lines), the slopes obtained are indicated in the different panels. The standard deviations of the relative errors are extracted from 4096 random states $|\psi\rangle$ and random bit-flip probabilities $p_{q,b}$.

particular, ω is invertible as long as $p \neq 1/2$. We again observe in Eqs. (44) and (45) that the noisy expectations of arbitrary operators can be related to the true operators in a surprisingly simple way, which requires no knowledge of the quantum hardware apart from the different bit-flip probabilities of the qubits.

3.4 Measurement of general operators \mathcal{H} from bit-string distributions of $|\psi\rangle$

3.4.1 Prediction for the noisy expectation value

Our analysis of the bit-flip error above assumed that we measure general operators \mathcal{H} by expressing them as linear combinations of operators U^*OU with $O \in \{\mathbb{1}, Z\}^{\otimes N}$ on an N -qubit machine, and by measuring each O independently (U being the transformation into the Z basis). For example, if we are interested in measuring $\mathcal{H}_{ZZ} = J \sum_{i=1}^N Z_i Z_{i+1}$ with $N = 3$ qubits, then we generate independent histograms for $\langle \psi | \mathbb{1}_3 \otimes Z_2 \otimes Z_1 | \psi \rangle$, $\langle \psi | Z_3 \otimes Z_2 \otimes \mathbb{1}_1 | \psi \rangle$, and $\langle \psi | Z_3 \otimes \mathbb{1}_2 \otimes Z_1 | \psi \rangle$, extract their expectation values, and recover $\langle \psi | \mathcal{H} | \psi \rangle$ accordingly.

Alternatively, as we will discuss in the following, we can measure the *distribution* of $|\psi\rangle$ and

obtain a histogram in terms of the computational basis $\{|j\rangle; j \in \mathbb{N}_{0, < 2^N}\}$. Hence, if the probability of measuring $|j\rangle$ is p_j , then we can recover $\langle\psi|\mathcal{H}|\psi\rangle$ from $\sum_j p_j \langle j|\mathcal{H}|j\rangle$. In this case, the measurements of $\langle\psi|U^*OU|\psi\rangle$ comprising $\langle\psi|\mathcal{H}|\psi\rangle$ are no longer independent. This has an impact on the variance of measurement histograms, as we will discuss in Sec. 3.5. However, it has no impact on the expectation subject to bit flips, since linearity of the expectation value implies

$$\begin{aligned}\mathbb{E}\langle\psi|\mathcal{H}^n|\psi\rangle &= \mathbb{E}\langle\psi|\sum_{\alpha}\lambda_{\alpha}U_{\alpha}^*O_{\alpha}^nU_{\alpha}|\psi\rangle \\ &= \langle\psi|\sum_{\alpha}\lambda_{\alpha}U_{\alpha}^*(\mathbb{E}O_{\alpha}^n)U_{\alpha}|\psi\rangle,\end{aligned}\quad (46)$$

which is precisely the expression we would obtain from summing the independently measured operators O_{α}^n .

3.4.2 Prediction for the bit-flip corrected operator

In order to correct for bit flips in this setting, we need to keep in mind that the general case requires measurements of all operators $O \preceq O_{\alpha}$ (with respect to the lexicographic order \preceq on $\{\mathbb{1}, Z\}^{\otimes N}$) for all operators O_{α} in $\mathcal{H} = \sum_{\alpha}\lambda_{\alpha}U_{\alpha}^*O_{\alpha}U_{\alpha}$. Hence, the histogram for $\langle\psi|\mathcal{H}^n|\psi\rangle$ does not contain sufficient information. However, we can use the classical bit-flip correction method as discussed above to find coefficients $\omega_{\alpha,O}$ such that

$$O_{\alpha} = \sum_{O \preceq O_{\alpha}} \omega_{\alpha,O} \mathbb{E}O^n \quad (47)$$

holds. Inserting this into \mathcal{H} , we can express \mathcal{H} as

$$\mathcal{H} = \sum_{\alpha}\lambda_{\alpha}U_{\alpha}^* \sum_{O \preceq O_{\alpha}} \omega_{\alpha,O} \mathbb{E}O^n U_{\alpha}. \quad (48)$$

In other words, we can replace the operator \mathcal{H} by the bit-flip corrected noisy operator

$$\mathcal{H}_{\text{bfc}}^n := \sum_{\alpha}\lambda_{\alpha}U_{\alpha}^* \sum_{O \preceq O_{\alpha}} \omega_{\alpha,O} O^n U_{\alpha} \quad (49)$$

and obtain

$$\mathbb{E}\langle\psi|\mathcal{H}_{\text{bfc}}^n|\psi\rangle = \langle\psi|\mathcal{H}|\psi\rangle. \quad (50)$$

3.4.3 Prediction for equal bit-flip probabilities

To compare our results to Secs. 2 and 3.3, let us assume that the bit-flip probabilities $p_{q,b}$ satisfy $p_{q,0} = p_{q,1} = p_q$, i.e., there is no difference

between $p(|0\rangle \rightarrow |1\rangle)$ and $p(|1\rangle \rightarrow |0\rangle)$ for each qubit, but this value might depend on the individual qubit. Then we obtain $\omega_{\alpha,O} = 0$ unless $O = O_{\alpha} = O_{\alpha,N} \otimes \cdots \otimes O_{\alpha,1}$, for which we find

$$\omega_{\alpha,O_{\alpha}} =: \omega_{\alpha} = \prod_q \frac{1}{(1 - 2p_q)}, \quad (51)$$

where q ranges over all qubits satisfying $O_{\alpha,q} = Z_q$. For $p_{q,b} = p$, this result agrees with Eqs. (7), (11), and (44).

Thus, the bit-flip corrected noisy operator

$$\mathcal{H}_{\text{bfc}}^n := \sum_{\alpha}\lambda_{\alpha}\omega_{\alpha}U_{\alpha}^*O_{\alpha}^nU_{\alpha} \quad (52)$$

has the same Pauli-sum structure as the original operator \mathcal{H} , changing only the coefficients. This is completely analogous to the independent measurement case. In both cases, if we have $p_{q,0} = p_{q,1}$, then we can correct for bit flips without additional cost to the quantum device.

3.5 Impact of measurement choices

In general, we will extract the quantum mechanical expectation of an operator by running the circuit preparing $|\psi\rangle$ a number of times followed by a projective measurement in the computational basis. As before, we refer to these repetitions as the number of shots, s . Of course, these shots are still subject to statistical fluctuations. Hence, if we generate N_{hist} histograms with s shots each, we can generate a histogram from the means extracted from each histogram. This will yield results as in Fig. 1 and Fig. 2. Using bit-flip corrected operators as in Eq. (49), we can shift the expected mean to coincide with the quantum mechanical expectation of the operator we wish to measure. However, the variance of histogram means is then highly dependent on the measurement paradigm.

For illustration, let us consider the transversal Ising model $\mathcal{H}_{\text{TI}} = J \sum_{j=1}^N Z_j Z_{j+1} + h \sum_{j=1}^N X_j$, which we will measure on the ground state $|\psi\rangle$. The first step is to compute the bit-flip corrected noisy Hamiltonian $\mathcal{H}_{\text{TI,bfc}}^n$. For simplicity, we will assume all bit-flip probabilities $p_{q,b}$ to coincide with some value p . This yields

$$\mathcal{H}_{\text{TI,bfc}}^n = J_p \sum_{j=1}^N Z_j^n Z_{j+1}^n + h_p \sum_{j=1}^N X_j^n \quad (53)$$

with $J_p := J(1 - 2p)^{-2}$ and $h_p := h(1 - 2p)^{-1}$. Of course, this process changes the variances. In particular, since Fig. 1 and Fig. 2 show histograms without the bit-flip correction, the prediction of variances in Fig. 1 and Fig. 2 uses J and h instead of J_p and h_p .

At this point, we need to decide upon the precise way of measuring the Hamiltonian. Essentially, we have a spectrum of possibilities which contains three interesting cases:

- *Method 1:* measure each $Z_j^n Z_{j+1}^n$ and X_j^n in Eq. (53) independently
- *Method 2:* measure the entire Hamiltonian $\mathcal{H}_{\text{TI,bfc}}^n$ in Eq. (53) from distributions of $|\psi\rangle$ measurements
- *Method 3:* measure $\mathcal{H}_{ZZ}^n := J_p \sum_{j=1}^N Z_j^n Z_{j+1}^n$ and $\mathcal{H}_X^n := h_p \sum_{j=1}^N X_j^n$ independently from distributions of $|\psi\rangle$ measurements

Methods 1 and 2 are the two extremes discussed in Secs. 3.1–3.3 and Sec. 3.4, respectively, whereas Method 3 is a reasonable compromise. In fact, Method 3 is precisely the method we used for Fig. 1 and Fig. 2. Method 3 is also an example that is closely related to implementations of quantum algorithms which are optimized for the number of calls to the quantum device, i.e., implementations in which only parts of an operator can be measured simultaneously and both Methods 1 and 2 are unfeasible.

The variance of histogram means has two contributions: bit-flip variance and quantum mechanical variance. These contributions for each of the three methods are shown in Fig. 5. The derivation of these variances can be found in Appendix A; in particular, Fig. 5 shows Eq. (74), Eq. (76), and Eq. (79). To remove the dependence on the number of shots per histogram, all variances are multiplied by the number of shots s , i.e., all values in Fig. 5 correspond to the normalization $s = 1$.

It is interesting to note that not only the full variance varies in magnitude but also the relative contribution from bit flips and quantum mechanics is vastly different between the three methods.

If we compare the two extremes – Method 1 and Method 2 – we notice that for Method 1 the bit-flip induced variance is small compared to the quantum mechanical variance, whereas

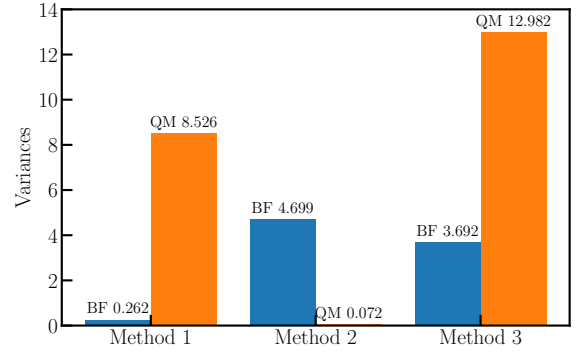


Figure 5: Contributions to the variance of histogram means for the bit-flip corrected TI Hamiltonian in Eq. (53) evaluated on the ground state of the “true” TI Hamiltonian in Eq. (17). The different bars correspond to the bit-flip (BF, blue) and quantum mechanical (QM, orange) variance contributions for the three different measurement methods. We used the parameters $N = 4$, $J = -1$, $h = 2$, and $p_{q,b} = p = 0.05$. All values are normalized by setting $s = 1$.

for Method 2 the situation is reversed. Generically, this pattern is to be expected. Method 1 is likely to produce a much smaller bit-flip contribution since all summands are measured independently. Meanwhile, measuring with Method 2 introduces $O(4^N)$ covariance terms, which vanish in Method 1 due to independent measurements of summands. Moreover, concerning Method 2, we note that the quantum mechanical variance vanishes upon evaluation on an eigenstate of the operator. In Fig. 5, we evaluated the bit-flip corrected transversal Ising Hamiltonian $\mathcal{H}_{\text{TI,bfc}}^n = J_p \sum_{j=1}^N Z_j^n Z_{j+1}^n + h_p \sum_{j=1}^N X_j^n$ with equal bit-flip probabilities $p_{q,b} = p = .05$ on the ground state of the “true” transversal Ising Hamiltonian $\mathcal{H}_{\text{TI}} = J \sum_{j=1}^N Z_j Z_{j+1} + h \sum_{j=1}^N X_j$. For small values of p , we can interpret the bit-flip correction as a small perturbation to the original operator. Hence, the ground state of \mathcal{H}_{TI} is close to an eigenstate of $\mathcal{H}_{\text{TI,bfc}}^n$ and thus the quantum mechanical contribution to the variance is small.

For intermediate methods, such as Method 3, it is generally difficult to predict the different contributions to the variance using similar arguments as above. Depending on the practical limitation of any given implementation, it will be imperative to balance the different contributions to the variance with the number of quantum device calls. For example, for the transversal Ising model, fewer quantum device calls per evaluation of the Hamiltonian introduce more covari-

ance terms. In turn, this requires more quantum device calls to obtain the necessary statistical power if we aim to extract a histogram mean with a required level of precision. Thus, this balancing act is highly problem specific. However, considering Method 3 for the transversal Ising model, it clearly shows that great care has to be taken when constructing an intermediate method if the aim is to reduce the overall variance on a given budget of quantum device calls.

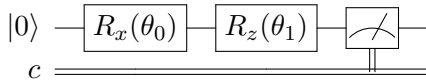
4 Experimental results

In order to demonstrate the experimental applicability of our measurement error mitigation method, we generate data on IBM quantum hardware using the Qiskit software development kit (SDK) [40]. To assess the performance of our correction procedure, we first simulate the quantum hardware classically using the noise models for the different backends provided by Qiskit, before we proceed to the actual hardware.

4.1 Single-qubit case

To begin with, let us focus on the simplest case of a single qubit. In a first step, we determine the bit-flip probabilities p_0 (p_1) of the qubit, which can be easily obtained by preparing the state $|0\rangle$ and recording the number of 1 outcomes (by preparing the state $|1\rangle$ through applying a single X gate to the initial $|0\rangle$ state and recording the number of 0 outcomes). In order to account for statistical fluctuations, we repeat this procedure several times and average over the flip probabilities obtained for each run (see Appendix B.1 for details).

After obtaining the flip probabilities, we measure $\langle\psi|Z|\psi\rangle$ for a randomly chosen $|\psi\rangle$. Starting from the initial state $|0\rangle$, we can prepare any state on the Bloch sphere by first applying a rotation gate around the x -axis followed by a rotation around the z -axis. Hence, we choose the circuit



in our experiments, where the angles θ_0 , θ_1 are both drawn uniformly from the interval $[0, 2\pi]$. Our measurement outcomes allow us to determine the noisy expectation value of Z , $\mathbb{E}(Z^n)$. Subsequently, we can apply our correction procedure using Eq. (24). To acquire statistics for

$\mathbb{E}(Z^n)$, we repeat the process for 1050 randomly chosen $|\psi\rangle$ and monitor the mean and the standard deviation of the absolute error

$$|\langle\psi|Z^n|\psi\rangle_{\text{measured}} - \langle\psi|Z|\psi\rangle_{\text{exact}}| \quad (54)$$

for both the noisy expectation value and the corrected expectation value. Moreover, each individual measurement for fixed values of θ_0 and θ_1 requires running the circuit multiple times to get the probability distribution of basis states in $|\psi\rangle$. Thus we also explore the dependence of our results on the number of shots s .

4.1.1 Classical simulation of quantum hardware

To benchmark the performance of our correction procedure, we first simulate `ibmq_london` [41] and `ibmq_burlington` [42] classically. The Qiskit SDK provides a noise model for each of the respective chips comprising various sources of error, including readout errors during the measurement process, which can be switched on and off individually. To begin with, we simulate the quantum hardware incorporating the measurement errors only, subsequently we use the full noise model to see the effect of the various other errors. Our results for the mean and the standard deviation of the absolute error as a function of s are shown in Fig. 6.

Focusing on the case with readout error only in Fig. 6(a) and 6(b), we see that correcting our results according to Eq. (24) clearly reduces the mean and the standard deviation of the absolute error in both cases. Without correction, the mean (standard deviation) of the absolute error converges to a value around 8×10^{-1} (9×10^{-1}), and increasing s beyond 1024 does not significantly improve the results. In particular, this stagnation already happens for values of s below the maximum one possible on real hardware, hence showing that the measurement error severely limits the precision that can be achieved. On the contrary, the corrected results show a significant improvement and a power-law decay of these quantities with s . In particular, in the ideal, completely noise-free case, performing a projective measurement on $|\psi\rangle$ is nothing but sampling from a probability distribution, thus one would expect the mean error to decay as $\propto s^{-1/2}$. To check for that behavior, we can fit the same functional form as in Sec. 3.2 to our data, the resulting exponents are shown in

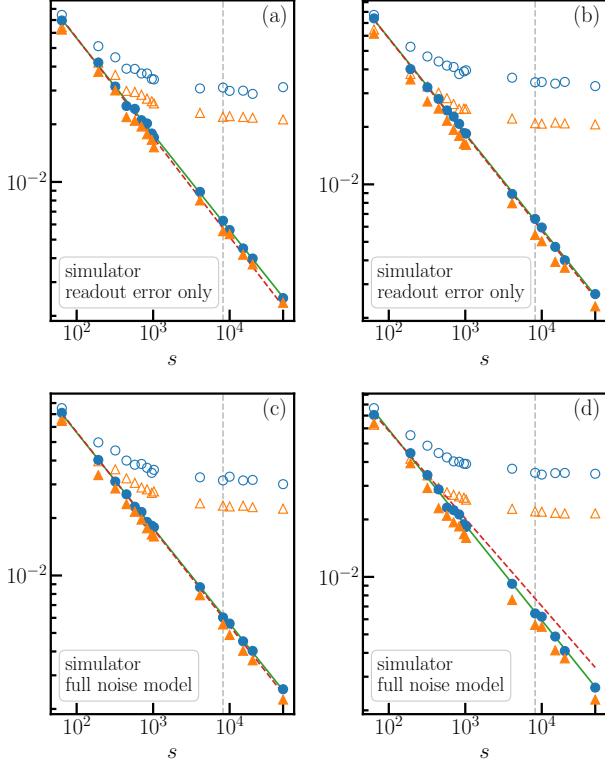


Figure 6: Mean value (blue dots) and standard deviation (orange triangles) of the absolute error Eq. (54) after applying the correction procedure (filled symbols) and without it (open symbols) as a function of the number of shots s . The different panels correspond to the data obtained by classically simulating the quantum hardware `ibmq_london` (left column) and `ibmq_burlington` (right column) including measurement noise only (upper row) and using the full hardware noise model (lower row). The solid green line corresponds to a power law fit to all our data points for the mean absolute error, the red dashed line to fit including the lowest four number of shots. The vertical gray dashed line indicates the maximum number of shots, 8192, that can be executed on the actual hardware.

Tab. 1. Indeed, we recover $\alpha = 1/2$, thus demonstrating that our correction procedure essentially allows us to recover the noise-free case.

Taking into account the full noise model in our simulations, which contains for instance gate errors and decoherence, we obtain the results in Fig. 6(c) and 6(d). Compared to the case with measurement errors only, the picture is very similar, which shows that the dominant error contribution for the single-qubit case is coming from the readout procedure. The mean and the standard deviation of the absolute error without any correction only approach marginally higher values than previously. Again, we observe a sig-

readout error only	first 4 points	full range
<code>ibmq_london</code>	0.519	0.501
<code>ibmq_burlington</code>	0.503	0.499
full noise model	first 4 points	full range
<code>ibmq_london</code>	0.508	0.500
<code>ibmq_burlington</code>	0.459	0.503

Table 1: Exponents α obtained from fitting the power law $Cs^{-\alpha}$ to our data for the mean absolute error in Fig. 6 after applying the correction.

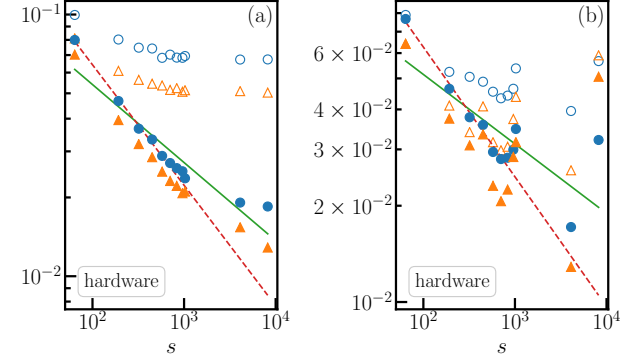


Figure 7: Mean value (blue dots) and standard deviation (orange triangles) of the absolute error Eq. (54) after applying the correction procedure (filled symbols) and without it (open symbols) as a function of the number of shots s . The solid green line corresponds to a power law fit to all our data points for the mean absolute error, the red dashed line to fit including the lowest four number of shots. Different panels correspond to data obtained on quantum hardware `ibmq_london` (a) and `ibmq_burlington` (b).

nificant reduction of the mean and the standard deviation of the absolute error after applying the correction procedure, and a power law decay with s . Fitting a power law to our data yields once more exponents around $1/2$ (see Tab. 1).

4.1.2 Quantum hardware

Our experiments can be readily carried out on quantum hardware, and we repeat the same simulations on `ibmq_london` and `ibmq_burlington`. The only difference with respect to the classical simulation is that s on those two devices is limited to a maximum number of 8192. Fig. 7 shows our results obtained on real devices.

Comparing our results for the chip `ibmq_london` in Fig. 7(a) to the classical simulation of the quantum hardware in Fig. 6(a) and Fig. 6(c), we observe qualitative agree-

ment for $s \leq 1024$. Compared to the classical simulation of the quantum device, the mean value and the standard deviation of the absolute error are in general larger on the hardware. Correcting for the readout errors yields again a significant improvement and reduces the mean and the standard deviation of absolute error considerably. As before, we can fit our data to a power law. While for a small number of shots $s < 500$ we observe again an exponent of $1/2$, for a larger number of measurements the curve for the corrected result starts to flatten out and the exponent obtained for fitting the entire range is considerably smaller than $1/2$ (see Tab. 2 for details). Since increasing s should decrease the inherent statistical fluctuations of the projective measurements, and measurement errors can be dealt with our scheme, this might be an indication that in addition to readout errors also other sources of noise play a significant role. Their effects cannot be corrected with our procedure and thus dominate from a certain point on.

Looking at the results for `ibmq_burlington` in Fig. 7(b) and comparing them with the classical simulation of the quantum hardware in Fig. 6(b) and Fig. 6(d), we see that the discrepancies in this case are more severe and the data is less consistent. Applying our mitigation to the data again yields an improvement, which is less pronounced than in the case of `ibmq_london`. For a small number of shots, the mean of the absolute error after correction shows again roughly a power law decay. The exponent obtained from a fit to our data in that range is smaller compared to the one from our data from `ibmq_london` (see Tab. 2 for details). From $s = 1024$ on, the uncorrected data is already less consistent. Making use of our mitigation scheme still yields an improvement, however, the corrected results scatter similarly to the original ones and do not follow the same power law as for a small number of shots, as a fit to our data reveals. This suggests that noise other than the one resulting from the readout has a considerable contribution.

4.2 Two-qubit case

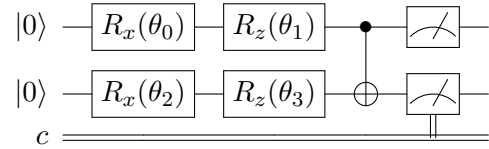
Since our correction procedure is not limited to the single-qubit setup, we can straightforwardly apply it to multiple qubits. To assess the performance for that case, we repeat the same pro-

chip name	first 4 points	full range
<code>ibmq_london</code>	0.460	0.298
<code>ibmq_burlington</code>	0.405	0.217

Table 2: Exponents α obtained from fitting the function $Cs^{-\alpha}$ to our data for the mean absolute error in Fig. 7 after applying the correction.

cedure we did previously but now for a circuit encompassing two qubits. Since we assume the flip-probabilities $p_{q,b}$ (with $q = 1, 2$, $b = 0, 1$) of the qubits to be independent of each other, we apply the same procedure that we used to obtain the flip probabilities in the single-qubit case, but this time for each qubit individually.

Subsequently we prepare a two-qubit state using the following circuit



where the angles $\theta_0, \dots, \theta_3$ are again random numbers drawn uniformly from $[0, 2\pi]$, and the final CNOT gate allows for creating entanglement between the two qubits. Analogous to the single-qubit case, we first simulate the quantum hardware classically before we eventually carry out our experiments on a real quantum device. In both cases we measure the noisy expectation value of $Z_2 \otimes Z_1$, $\mathbb{E}(Z_2^n \otimes Z_1^n)$, and apply Eq. (41) to correct for noise caused by readout errors. Again, we repeat the procedure for 1050 randomly chosen sets of angles and compute the mean and the standard deviation of the absolute error

$$|\langle \psi | Z_2^n \otimes Z_1^n | \psi \rangle_{\text{measured}} - \langle \psi | Z_2 \otimes Z_1 | \psi \rangle_{\text{exact}}| \quad (55)$$

as a function of the number of shots, s , with and without applying the mitigation scheme.

4.2.1 Classical simulation of quantum hardware

As for the single-qubit case, we use the Qiskit SDK to classically simulate the chips `imq_london` and `ibmq_burlington` first with readout error only and subsequently using the full noise model. Figure 8 shows our results for both cases.

Looking at Fig. 8(a) and Fig. 8(b), we see that the two-qubit case with just measurement error

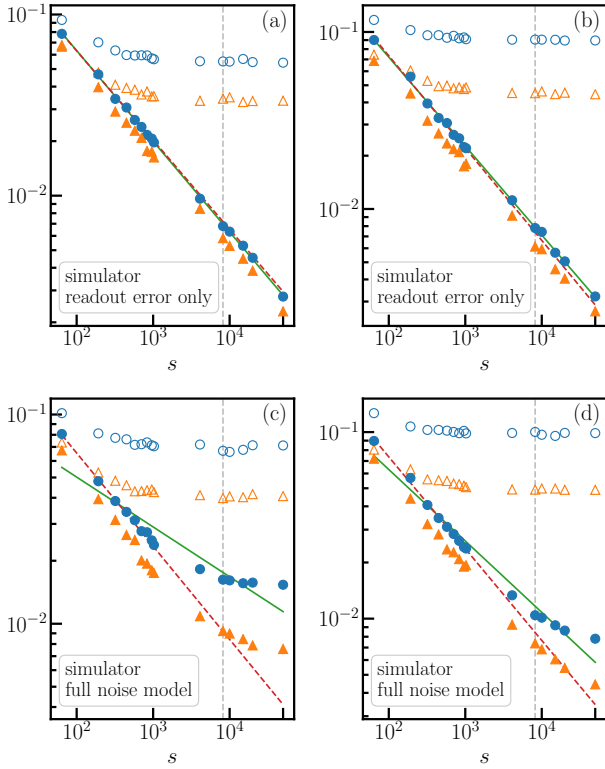


Figure 8: Mean value (blue dots) and standard deviation (orange triangles) of the absolute error Eq. (55) after applying the correction procedure (filled symbols) and without it (open symbols) as a function of the number of shots s . The different panels correspond to the data obtained by classically simulating the quantum hardware `ibmq_london` (left column) and `ibmq_burlington` (right column) including measurement noise only (upper row) and using the full hardware noise model (lower row). The solid green line corresponds to a power law fit to all our data points for the mean absolute error, the red dashed line to fit including the lowest four number of shots. The vertical gray dashed line indicates the maximum number of shots that can be executed on the actual hardware.

behaves like the single-qubit case. Without applying any correction, the mean and the standard deviation of the absolute error initially decrease with increasing s , before eventually converging to fixed values which are slightly higher than for the single-qubit case (compare Fig. 6(a) with Fig. 8(a) and Fig. 6(b) with Fig. 8(b)). Applying the correction procedure, we can significantly decrease the values and observe again a power-law decay with an exponent of $1/2$ over the entire range of s we study, as a fit to our corrected data reveals (see also Tab. 3).

Repeating the same simulations, but this time with the full noise model, yields the results in Fig. 8(c) and Fig. 8(d). Comparing this to the

readout error only	first 4 points	full range
<code>ibmq_london</code>	0.492	0.501
<code>ibmq_burlington</code>	0.522	0.503
full noise model	first 4 points	full range
<code>ibmq_london</code>	0.446	0.238
<code>ibmq_burlington</code>	0.492	0.383

Table 3: Exponents α obtained from fitting the power law $Cs^{-\alpha}$ to our data for the mean absolute error in Fig. 8.

case with readout error only, we see a more pronounced effect than in the single-qubit case. Applying the correction reduces the mean and the standard deviation of the absolute error still considerably, nevertheless one can observe that data after correction converges to a fixed value with increasing s . In particular, the power law decay with $\alpha = 1/2$ is only present for a small number of shots. Considering the entire range of s we study, the classical simulation of `ibmq_london` predicts that the data is not very well compatible with a power law. In contrast, our simulation data for `ibmq_burlington` is still reasonably well described by a power law, however with an exponent of 0.38 and thus considerably smaller than $1/2$ (see Tab. 3 for details). Most notably, a comparison between the results for classically simulating two qubits using the full noise model to the single-qubit case in Fig. 6(c) and Figs. 6(d), we see that noise has a substantially larger effect in the two-qubit case. This can be partially explained by the CNOT gate in the circuit, as the error rates for two-qubit gates are in general much larger than for single-qubit rotations.

4.2.2 Hardware

For the two-qubit case, we can carry out the simulations on real quantum hardware as well. Using again `ibmq_london` and `ibmq_burlington` we obtain the data shown in Fig. 9.

Our results for `ibmq_london` in Fig. 9(a) shows qualitative agreement with the classical simulation. Once more, we see that the mean and the standard deviation of the absolute error obtained on the hardware converge to higher values than the ones obtained from the simulation (compare Fig. 8(c) and Fig. 9(a)). Correcting our data according to Eq. (41), the mean of the absolute error and its standard deviation are sig-

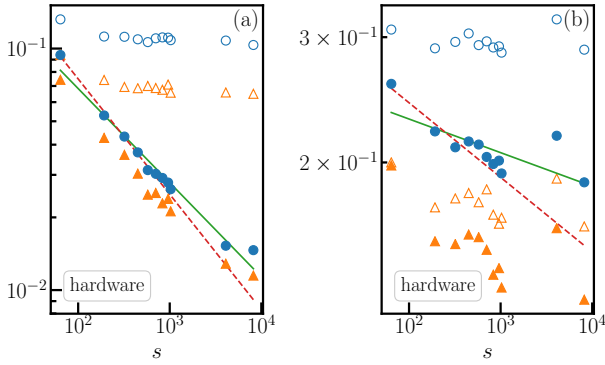


Figure 9: Mean value (blue dots) and standard deviation (orange triangles) of the absolute error Eq. (55) after applying the correction procedure (filled symbols) and without it (open symbols) as a function of the number of shots s . The solid green line corresponds to a power law fit to all our data points for the mean absolute error, the red dashed line to fit including the lowest four number of shots. Different panels correspond to data obtained on quantum hardware `ibmq_london` (a) and `ibmq_burlington` (b).

nificantly reduced. Comparing the reduction to the single-qubit case in Fig. 7, we observe that for the two-qubit case, the improvement is even larger. In particular, for our largest number of shots $s = 8192$, the mean and the standard deviation of the absolute error are reduced by approximately one order of magnitude. The corrected data is again well described by a power law. Fitting the first 4 data points, we obtain an exponent of 0.48. Using the entire range of s for the fit, the exponent only decreases moderately to 0.39 (see also Tab. 4), thus showing that the readout error has still a significant contribution to the overall error.

Turning to our results for `ibmq_burlington` in Fig. 9(b), we see that the data for this chip is significantly worse. For one, the mean value (standard deviation) of the absolute error without applying any correction procedure is roughly a factor 3 (2) larger than the one obtained on `ibmq_london`. Applying the correction procedure still yields an improvement, however, this time it is a lot smaller than for `ibmq_london`, as a comparison between Fig. 9(a) and Fig. 9(b) shows. While for a small number of shots, the mean value of the absolute error after correction still shows a power law decay, albeit with an exponent a lot smaller than $1/2$, for a large number of shots this trend stops, as fits to our data reveal (see also Tab. 4). This behavior is giving an

chip name	first 4 points	full range
<code>ibmq_london</code>	0.478	0.390
<code>ibmq_burlington</code>	0.105	0.047

Table 4: Exponents α obtained from fitting the power law $Cs^{-\alpha}$ to our data for the mean absolute error in Fig. 9.

indication that for `ibmq_burlington`, the readout error is not the dominant one, but rather other errors have a significant contribution which cannot be corrected for using our scheme.

5 Conclusions

In this paper, we proposed a classical bit-flip correction method to mitigate measurement errors on noisy quantum computers. This method relies on cancellations of different erroneous measurement outcomes and requires the knowledge of the different bit-flip probabilities during readout for each qubit. We tested the performance of this method by correcting the noisy energy histograms of the longitudinal and transversal Ising models. Moreover, we demonstrated that the method can be applied to any operator, any number of qubits, and any realistic bit-flip probability. For the single-qubit case, we also provided a density matrix description and a visualization scheme of the measurement noise. Finally, we tested our method both numerically and experimentally for the IBM quantum devices `ibmq_london` and `ibmq_burlington` for a single qubit and two qubits. We observe that our method is able to improve the data significantly for both cases and to reduce the error by up to one order of magnitude.

Our method of replacing noisy operators with random operators that model the noise behavior, as we presented in Sec. 3, is generally applicable to arbitrary observables and could also be applied to other error sources as well. In the general case, this does incur a certain amount of overhead cost since each tensor product of n single-qubit Z operators is replaced by up to 2^n operators. Hence, in the worst case scenario, the computational cost is exponential. However, in many applications this will not be the case. For example, if we consider a generic operator acting on N qubits, it will already be a linear combination of all 4^N N -qubit Pauli matrices (with

N -qubit Pauli matrices we mean a tensor product of N 2×2 Pauli matrices, which include the 2×2 identity). The replacement method will then only change the coefficients but not incur any overhead on the quantum device.

For most physically relevant cases, the overhead will only be moderate in the number of qubits. If we consider local Hamiltonians, then each N -qubit Pauli matrix in the linear combination comprising the Hamiltonian has only very few terms that are not the 2×2 identity. For example, the longitudinal and transversal Ising models only contain tensor products of up to two Z -terms. This always incurs at most three additional terms (one of which is the identity) and many of them coincide; e.g., both $Z_2 Z_1$ and $Z_3 Z_2$ generate the additional Z_2 term. Hence, we require measurements for at most N additional operators. In general, if the Hamiltonian contains at most n -fold tensor products of non-identity Pauli matrices, then we need to measure at most $\binom{N}{n} 2^n$ additional terms. For local Hamiltonians, n is fixed and thus the cost is polynomial in the number of qubits N .

Moreover, this cost will often be further reduced due to already existing lower-order terms. For example, if we consider the longitudinal Ising model $\mathcal{H}_{\text{LI}} = J \sum_{i=1}^N Z_i Z_{i+1} + h \sum_{i=1}^N Z_i$, then each $Z_i Z_{i+1}$ generates the additional terms Z_i , Z_{i+1} , and $\mathbf{1}$. With the exception of $\mathbf{1}$, they are already part of the Hamiltonian and only the coefficients change. Furthermore, the additional $\mathbf{1}$ simply results in a constant and needs no measurement. Thus, the longitudinal Ising model incurs no overhead cost on the quantum device.

In addition to the moderate overhead cost, another advantage of our mitigation scheme is that it can be readily integrated into hybrid quantum-classical algorithms, as for example the Quantum Approximate Optimization Algorithm [43] and VQS. After initially measuring the flip probabilities, one can simply correct the values obtained for the cost function from the quantum device before passing them on to a classical algorithm for optimizing the variational parameters. Moreover, our method is completely platform independent and lends itself not only to superconducting qubits, but also to other architectures such as trapped ions. As long as the readout errors are constant to a certain degree and not excessively large, they can be reliably corrected for with our

procedure. These advantages make our mitigation method promising for various applications on NISQ devices and beyond.

Acknowledgments

The authors thank Xu Feng, Tom Weber, and Philipp Stratmann for valuable discussions and comments. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade. S.K. acknowledges financial support from the Cyprus Research and Innovation Foundation under project "Future-proofing Scientific Applications for the Supercomputers of Tomorrow (FAST)", contract no. COMPLEMENTARY/0916/0048. P.S. thanks the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS) for funding. We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team.

A Prediction for the variances of noisy expectation values

In this appendix, we compute the variances of the different noisy expectation values presented in Secs. 2 and 3. We follow the structure of Sec. 3: we first discuss a single Z operator in Sec. A.1, followed by the general case of $Z_N \otimes \dots \otimes Z_1$ operators in Sec. A.2. We then simplify our results to the case of equal bit-flip probabilities in Sec. A.3. Finally, we discuss the case of measuring general operators from bit-string distributions of $|\psi\rangle$ in Sec. A.4. In this section, we will also discuss different measurement paradigms and their impact on the variance of means extracted from histogram data. We will eventually return to the transversal Ising model for an explicit illustration.

A.1 Measurement of single Z operator

For computing the variance $\mathbb{V}Z_q^n$ of the noisy expectation in Eq. (24),

$$\begin{aligned}\mathbb{V}Z_q^n &= \mathbb{E}(Z_q^n \otimes Z_q^n) - (\mathbb{E}Z_q^n)^2 \\ &= \Phi'_{Z_q^n}(0) \otimes \Phi'_{Z_q^n}(0) - \Phi''_{Z_q^n}(0),\end{aligned}\quad (56)$$

we need to evaluate the derivatives $\Phi'_q(0) = i\mathbb{E}Z_q^n$ and $\Phi''_{Z_q^n}(0) = -\mathbb{E}(Z_q^n)^2$ of the characteristic function

$$\Phi_{Z_q^n}(t) := \mathbb{E} \exp \left[i \operatorname{Tr} \left(t^* Z_q^n \right) \right]. \quad (57)$$

This yields

$$\begin{aligned}\Phi'_{Z_q^n}(0) &= i(1 - p_{q,0} - p_{q,1})Z_q + (p_{q,1} - p_{q,0})\mathbb{1}_q \\ \Phi''_{Z_q^n}(0) &= -(1 - p_{q,0} - p_{q,1} + 2p_{q,0}p_{q,1})Z_q \otimes Z_q \\ &\quad - (p_{q,0} - p_{q,1} + 2p_{q,0}p_{q,1})\mathbb{1}_q \otimes \mathbb{1}_q.\end{aligned}\quad (58)$$

Thus, the variance operator in Eq. (56) reads

$$\begin{aligned}\mathbb{V}Z_q^n &= [(p_{q,0} + p_{q,1})(1 - p_{q,0} - p_{q,1}) + 2p_{q,0}p_{q,1}] \\ &\quad \times Z_q \otimes Z_q \\ &\quad - (1 - p_{q,0} - p_{q,1})(p_{q,1} - p_{q,0})Z_q \otimes \mathbb{1}_q \\ &\quad - (1 - p_{q,0} - p_{q,1})(p_{q,1} - p_{q,0})\mathbb{1}_q \otimes Z_q \\ &\quad + (p_{q,0} + p_{q,1} - p_{q,0}^2 - p_{q,1}^2)\mathbb{1}_q \otimes \mathbb{1}_q.\end{aligned}\quad (59)$$

A.2 Measurement of $Z_N \otimes \cdots \otimes Z_1$ operators

We now generalize the variance for $N = 1$ in Eq. (59) to operators acting on multiple qubits, i.e., $N > 1$. According to Eq. (32), operators O_1^n and O_2^n acting on different qubits are uncorrelated, i.e., the covariance vanishes,

$$\begin{aligned}\operatorname{Cov}_{\otimes}(O_1^n, O_2^n) &:= \mathbb{E}(O_1^n \otimes O_2^n) - \mathbb{E}(O_1^n) \otimes \mathbb{E}(O_2^n) \\ &= 0.\end{aligned}\quad (60)$$

Hence, we obtain the variance operator

$$\begin{aligned}\mathbb{V}(Z_N^n \otimes \cdots \otimes Z_1^n) &= \mathbb{E}(Z_N^n \otimes \cdots \otimes Z_1^n \otimes Z_N^n \otimes \cdots \otimes Z_1^n) \\ &\quad - \mathbb{E}(Z_N^n \otimes \cdots \otimes Z_1^n) \otimes \mathbb{E}(Z_N^n \otimes \cdots \otimes Z_1^n) \\ &= U^* \left(\mathbb{E}(Z_N^n \otimes Z_N^n) \otimes \cdots \otimes \mathbb{E}(Z_1^n \otimes Z_1^n) \right. \\ &\quad \left. - \bigotimes_{q=1}^N (\mathbb{E}Z_q^n \otimes \mathbb{E}Z_q^n) \right) U \\ &= U^* \left((\mathbb{V}Z_N^n + \mathbb{E}Z_N^n \otimes \mathbb{E}Z_N^n) \otimes \cdots \right. \\ &\quad \left. \otimes (\mathbb{V}Z_1^n + \mathbb{E}Z_1^n \otimes \mathbb{E}Z_1^n) \right. \\ &\quad \left. - \bigotimes_{q=1}^N (\mathbb{E}Z_q^n \otimes \mathbb{E}Z_q^n) \right) U,\end{aligned}\quad (61)$$

where the unitary operation U re-orders the tensor products from $|\psi_N\rangle \otimes \cdots \otimes |\psi_1\rangle \otimes |\psi_N\rangle \otimes \cdots \otimes |\psi_1\rangle$ to $(|\psi_N\rangle \otimes |\psi_N\rangle) \otimes \cdots \otimes (|\psi_1\rangle \otimes |\psi_1\rangle)$. That is, for two qubits the re-ordering maps the basis state $|b_3b_2b_1b_0\rangle$ to $|b_3b_1b_2b_0\rangle$, and for three qubits the re-ordering maps $|b_5b_4b_3b_2b_1b_0\rangle$ to $|b_5b_2b_4b_1b_3b_0\rangle$, etc.

A.3 Measurement of $Z_N \otimes \cdots \otimes Z_1$ operators assuming equal bit-flip probabilities

For $N = 1$, the variance in Eq. (61) reduces to

$$\mathbb{V}Z_q^n = 2p(1-p)(Z_q \otimes Z_q + \mathbb{1}_q \otimes \mathbb{1}_q). \quad (62)$$

For $N = 2$, the re-ordering of the tensor product $|\psi\rangle \otimes |\psi\rangle$ in Eq. (61) becomes important, which yields

$$\begin{aligned}\mathbb{V}(\langle\psi| Z_2^n \otimes Z_1^n |\psi\rangle) &= (\langle\psi| \otimes \langle\psi|) U^* (\mathbb{V}Z_2^n \otimes \mathbb{V}Z_1^n) U (|\psi\rangle \otimes |\psi\rangle) \\ &\quad + (\langle\psi| \otimes \langle\psi|) U^* (\mathbb{V}Z_2^n \otimes \mathbb{E}Z_1^n \otimes \mathbb{E}Z_1^n) U (|\psi\rangle \otimes |\psi\rangle) \\ &\quad + (\langle\psi| \otimes \langle\psi|) U^* (\mathbb{E}Z_2^n \otimes \mathbb{E}Z_2^n \otimes \mathbb{V}Z_1^n) U (|\psi\rangle \otimes |\psi\rangle).\end{aligned}\quad (63)$$

For arbitrary N , we can evaluate the variance operator in Eq. (61) for the ground state, $|\psi\rangle =$

$|0 \dots 0\rangle$, and obtain the expression

$$\begin{aligned}
& \mathbb{V}(\langle \psi | Z_N^n \otimes \dots \otimes Z_1^n | \psi \rangle) \\
&= \prod_{q=1}^N \langle \psi | \mathbb{V} Z_q^n + \mathbb{E} Z_q^n \otimes \mathbb{E} Z_q^n | \psi \rangle \\
&- \prod_{q=1}^N \langle \psi | \mathbb{E} Z_q^n \otimes \mathbb{E} Z_q^n | \psi \rangle \\
&= \prod_{q=1}^N (4p(1-p) + (1-2p)^2) - (1-2p)^{2N} \\
&= 1 - (1-2p)^{2N}.
\end{aligned} \tag{64}$$

This surprisingly simple result can be verified directly by noting that the measurement of $\langle 0 \dots 0 | Z_N^n \otimes \dots \otimes Z_1^n | 0 \dots 0 \rangle$ yields the values $+1$ with probability p_1 and -1 with probability p_{-1} . Thus, we conclude

$$\begin{aligned}
& \mathbb{V}(\langle 0 \dots 0 | Z_N^n \otimes \dots \otimes Z_1^n | 0 \dots 0 \rangle) \\
&= \mathbb{E}(\langle 0 \dots 0 | Z_N^n \otimes \dots \otimes Z_1^n | 0 \dots 0 \rangle^2) \\
&- \mathbb{E}(\langle 0 \dots 0 | Z_N^n \otimes \dots \otimes Z_1^n | 0 \dots 0 \rangle)^2 \\
&= p_1 + p_{-1}(-1)^2 \\
&- (1-2p)^{2N} \langle 0 \dots 0 | Z_N^n \otimes \dots \otimes Z_1^n | 0 \dots 0 \rangle^2 \\
&= 1 - (1-2p)^{2N}.
\end{aligned} \tag{65}$$

A.4 Measurement of general operators \mathcal{H} from bit-string distributions of $|\psi\rangle$

A.4.1 Prediction for the variance of operators

While measuring the entire Hamiltonian simultaneously makes no difference for the measured mean value, the variance on the other hand is affected by this change in measurement paradigm. If we consider \mathcal{H}_{ZZ} with $N = 2$ and $J = 1$, i.e., $\mathcal{H}_{ZZ} = Z_2 Z_1 + Z_1 Z_2$, then we would formally compute $\langle \psi | Z_2 \otimes Z_1 | \psi \rangle$ twice independently using the approach considered so far, whereas the expectation from the bit-string distribution of $|\psi\rangle$ directly extracts $2 \langle \psi | Z_2 \otimes Z_1 | \psi \rangle$. Thus the variance using independent histograms for each summand is given by

$$\begin{aligned}
& \mathbb{V}_{\text{ind}} \langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle = \\
&= \mathbb{V} \langle \psi | Z_2^n \otimes Z_1^n | \psi \rangle + \mathbb{V} \langle \psi | Z_2^n \otimes Z_1^n | \psi \rangle \tag{66} \\
&= 2 \mathbb{V} \langle \psi | Z_2^n \otimes Z_1^n | \psi \rangle
\end{aligned}$$

whereas the variance using the bit-string distribution of $|\psi\rangle$ is

$$\begin{aligned}
\mathbb{V}_{\text{bsd}} \langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle &= \mathbb{V}(2 \langle \psi | Z_2^n \otimes Z_1^n | \psi \rangle) \\
&= 4 \mathbb{V} \langle \psi | Z_2^n \otimes Z_1^n | \psi \rangle \tag{67} \\
&= 2 \mathbb{V}_{\text{ind}} \langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle.
\end{aligned}$$

In general, if $\mathcal{H}^n = \sum_{\alpha} \lambda_{\alpha} U_{\alpha}^* O_{\alpha}^n U_{\alpha}$, we are still able to predict the variance $\mathbb{V}_{\text{bsd}} \langle \psi | \mathcal{H}^n | \psi \rangle$ using the same method as above albeit the covariance terms no longer vanish (each O_{α}^n is a tensor product $O_{\alpha,N}^n \otimes \dots \otimes O_{\alpha,1}^n$). For $O_{\alpha,q}^n = Z_q^n$, $O_{\alpha,q}^n$ takes one of the possible values $\{Z_q, -\mathbb{1}_q, \mathbb{1}_q, -Z_q\}$, as in Sec. 3.1. For $O_{\alpha,q}^n = \mathbb{1}_q^n$, $O_{\alpha,q}^n$ always takes the value $\mathbb{1}_q$. Using these replacements for all summands in \mathcal{H}^n , we obtain that \mathcal{H}^n takes finitely many (up to 2^N) values \mathcal{H}_{α} with probability p_{α} . Hence, the characteristic function $\Phi_{\mathcal{H}^n}$ is given by

$$\begin{aligned}
\Phi_{\mathcal{H}^n}(t) &:= \mathbb{E} \exp(i \text{Tr}(t^* \mathcal{H}^n)) \\
&= \sum_{\alpha} p_{\alpha} \exp(i \text{Tr}(t^* \mathcal{H}_{\alpha})).
\end{aligned} \tag{68}$$

As such, we can directly conclude

$$\Phi'_{\mathcal{H}^n}(0) = \sum_{\alpha} p_{\alpha} i \mathcal{H}_{\alpha} = i \mathbb{E} \mathcal{H}^n, \tag{69}$$

$$\Phi''_{\mathcal{H}^n}(0) = - \sum_{\alpha} p_{\alpha} \mathcal{H}_{\alpha} \otimes \mathcal{H}_{\alpha}, \tag{70}$$

and find the variance operator

$$\begin{aligned}
\mathbb{V}_{\text{bsd}} \mathcal{H}^n &= \Phi'_{\mathcal{H}^n}(0) \otimes \Phi'_{\mathcal{H}^n}(0) - \Phi''_{\mathcal{H}^n}(0) \\
&= \left(\sum_{\alpha} p_{\alpha} \mathcal{H}_{\alpha} \otimes \mathcal{H}_{\alpha} \right) - (\mathbb{E} \mathcal{H}^n) \otimes (\mathbb{E} \mathcal{H}^n) \\
&= \left(\sum_{\alpha} p_{\alpha} \mathcal{H}_{\alpha} \otimes \mathcal{H}_{\alpha} \right) - \left(\sum_{\alpha, \beta} p_{\alpha} p_{\beta} \mathcal{H}_{\alpha} \otimes \mathcal{H}_{\beta} \right).
\end{aligned} \tag{71}$$

Similarly, we can measure the operator \mathcal{H}^n on the state $|\psi\rangle$ and obtain the variance

$$\begin{aligned}
\mathbb{V}_{\text{bsd}} \langle \psi | \mathcal{H}^n | \psi \rangle &= \left(\sum_{\alpha} p_{\alpha} \langle \psi | \mathcal{H}_{\alpha} | \psi \rangle^2 \right) \\
&- \left(\sum_{\alpha, \beta} p_{\alpha} p_{\beta} \langle \psi | \mathcal{H}_{\alpha} | \psi \rangle \langle \psi | \mathcal{H}_{\beta} | \psi \rangle \right).
\end{aligned} \tag{72}$$

A.4.2 Prediction for the variance of histogram means

Lastly, we can combine the bit-flip induced variances with quantum mechanically induced variances to obtain the full variances observed in

measuring histogram means. In particular, we will construct the variances for the three methods discussed in Sec. 3.5. There we measured the bit-flip corrected transversal Ising Hamiltonian $\mathcal{H}_{\text{TI,bfc}}^n = J_p \sum_{j=1}^N Z_j^n Z_{j+1}^n + h_p \sum_{j=1}^N X_j^n$ in Eq. (53) subject to bit flips on the ground state of the “true” transversal Ising Hamiltonian $\mathcal{H}_{\text{TI}} = J \sum_{j=1}^N Z_j Z_{j+1} + h \sum_{j=1}^N X_j$. For simplicity, we assumed that all bit-flip probabilities $p_{q,b}$ equal p . The three methods are

- *Method 1:* measure each $Z_j^n Z_{j+1}^n$ and X_j^n in Eq. (53) independently
- *Method 2:* measure the entire Hamiltonian $\mathcal{H}_{\text{TI,bfc}}^n$ in Eq. (53) from distributions of $|\psi\rangle$ measurements
- *Method 3:* measure $\mathcal{H}_{ZZ}^n := J_p \sum_{j=1}^N Z_j^n Z_{j+1}^n$ and $\mathcal{H}_X^n := h_p \sum_{j=1}^N X_j^n$ independently from distributions of $|\psi\rangle$ measurements

Method 1: Since each $Z_j^n Z_{j+1}^n$ and X_j^n is measured independently, the bit-flip contributions $\mathbb{V}_{\text{bf}} \langle \psi | Z_j^n Z_{j+1}^n | \psi \rangle$ and $\mathbb{V}_{\text{bf}} \langle \psi | X_j^n | \psi \rangle$ to the variance can be directly obtained from Eq. (61) keeping in mind that $X_j^n = H_j Z_j^n H_j$ where H_j is the Hadamard gate on qubit j . But since $|\psi\rangle$, in general, will not be an eigenstate of all $Z_j Z_{j+1}$ and X_j simultaneously, we also have a contribution from the quantum mechanical variances $\mathbb{V}_{\text{QM}} \langle \psi | Z_j Z_{j+1} | \psi \rangle = 1 - \langle \psi | Z_j Z_{j+1} | \psi \rangle^2$ and $\mathbb{V}_{\text{QM}} \langle \psi | X_j | \psi \rangle = 1 - \langle \psi | X_j | \psi \rangle^2$. We therefore obtain the variance of histogram means

$$\begin{aligned} \mathbb{V}_{\text{M1}} \langle \psi | \mathcal{H}_{\text{TI,bfc}}^n | \psi \rangle &= \frac{J_p^2}{s} \sum_{j=1}^N \mathbb{V}_{\text{bf}} \langle \psi | Z_j^n Z_{j+1}^n | \psi \rangle \\ &+ \frac{J_p^2}{s} \sum_{j=1}^N \mathbb{V}_{\text{QM}} \langle \psi | Z_j Z_{j+1} | \psi \rangle \\ &+ \frac{h_p^2}{s} \sum_{j=1}^N \mathbb{V}_{\text{bf}} \langle \psi | X_j^n | \psi \rangle \\ &+ \frac{h_p^2}{s} \sum_{j=1}^N \mathbb{V}_{\text{QM}} \langle \psi | X_j | \psi \rangle. \end{aligned} \quad (73)$$

In particular, if the state $|\psi\rangle$ is translationally invariant, such as the ground state of \mathcal{H}_{TI} , then

this further simplifies to

$$\begin{aligned} \mathbb{V}_{\text{M1}} \mathcal{H}_{\text{TI,bfc}}^n &= \frac{J_p^2 N}{s} \mathbb{V}_{\text{bf}} \langle \psi | Z_j^n Z_{j+1}^n | \psi \rangle \\ &+ \frac{J_p^2 N}{s} \mathbb{V}_{\text{QM}} \langle \psi | Z_j Z_{j+1} | \psi \rangle \\ &+ \frac{h_p^2 N}{s} \mathbb{V}_{\text{bf}} \langle \psi | X_j^n | \psi \rangle \\ &+ \frac{h_p^2 N}{s} \mathbb{V}_{\text{QM}} \langle \psi | X_j | \psi \rangle \end{aligned} \quad (74)$$

for any choice of j .

Method 2: In this case, the bit-flip contribution $\mathbb{V}_{\text{bf}} \langle \psi | \mathcal{H}_{\text{TI,bfc}}^n | \psi \rangle$ is given by Eq. (72) and the quantum mechanical variance is given by

$$\begin{aligned} \mathbb{V}_{\text{QM}} \langle \psi | \mathcal{H}_{\text{TI,bfc}}^n | \psi \rangle &= \langle \psi | (\mathcal{H}_{\text{TI,bfc}}^n)^2 | \psi \rangle \\ &- \langle \psi | \mathcal{H}_{\text{TI,bfc}}^n | \psi \rangle^2. \end{aligned} \quad (75)$$

Hence, the variance of histogram means is

$$\begin{aligned} \mathbb{V}_{\text{M2}} \langle \psi | \mathcal{H}_{\text{TI,bfc}}^n | \psi \rangle &= \frac{1}{s} \mathbb{V}_{\text{bf}} \langle \psi | \mathcal{H}_{\text{TI,bfc}}^n | \psi \rangle \\ &+ \frac{1}{s} \mathbb{V}_{\text{QM}} \langle \psi | \mathcal{H}_{\text{TI,bfc}}^n | \psi \rangle. \end{aligned} \quad (76)$$

While this expression appears simpler than its counterpart for Method 1, it is also important to note that $O(4^N)$ terms are required to compute $\mathbb{V}_{\text{M2}} \langle \psi | \mathcal{H}_{\text{TI,bfc}}^n | \psi \rangle$ whereas the number of terms required to compute $\mathbb{V}_{\text{M1}} \mathcal{H}_{\text{TI,bfc}}^n$ is only $O(N)$ and can even be reduced to $O(1)$ for translationally invariant states $|\psi\rangle$.

Method 3: Being a combination of Method 1 and Method 2, the variance can be constructed combining the results from Method 1 and 2. The bit-flip contributions $\mathbb{V}_{\text{bf}} \langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle$ and $\mathbb{V}_{\text{bf}} \langle \psi | \mathcal{H}_X^n | \psi \rangle$ follow from Eq. (72) again. Furthermore, the quantum mechanical variances contribute as

$$\begin{aligned} \mathbb{V}_{\text{QM}} \langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle &= \langle \psi | (\mathcal{H}_{ZZ}^n)^2 | \psi \rangle \\ &- \langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle^2 \end{aligned} \quad (77)$$

and

$$\mathbb{V}_{\text{QM}} \langle \psi | \mathcal{H}_X^n | \psi \rangle = \langle \psi | (\mathcal{H}_X^n)^2 | \psi \rangle - \langle \psi | \mathcal{H}_X^n | \psi \rangle^2. \quad (78)$$

The variance of histogram means is thus

$$\begin{aligned} \mathbb{V}_{M3} \langle \psi | \mathcal{H}_{\text{TI,bfc}}^n | \psi \rangle &= \frac{1}{s} \mathbb{V}_{\text{bf}} \langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle \\ &+ \frac{1}{s} \mathbb{V}_{\text{bf}} \langle \psi | \mathcal{H}_X^n | \psi \rangle \\ &+ \frac{1}{s} \mathbb{V}_{\text{QM}} \langle \psi | \mathcal{H}_{ZZ}^n | \psi \rangle \\ &+ \frac{1}{s} \mathbb{V}_{\text{QM}} \langle \psi | \mathcal{H}_X^n | \psi \rangle. \end{aligned} \quad (79)$$

B Technical details of the simulations

Here we briefly summarize the details on how to determine the flip probabilities, the simulations, and data evaluation procedure for the results shown in Sec. 4.

B.1 Calibration of the flip probabilities

Although the Qiskit SDK [40] provides values for the flip probabilities for the different qubits on the different chips, we choose to calibrate $p_{q,0}$ and $p_{q,1}$ ourselves. To obtain $p_{q,0}$, we simply measure the initial state using $s_{\text{calibration}}$ shots and record the number of 1 outcomes. Similarly, we determine $p_{q,1}$ by first applying an X-gate to the qubit q , thus preparing the state $|1\rangle$ and measure the resulting state again $s_{\text{calibration}}$ times and record the number of 0 outcomes. For all data shown in the main text we use $s_{\text{calibration}} = 8192$ which is the maximum number of repetitions possible on the real quantum hardware. Moreover, to acquire some statistics how the obtained values for the flip probabilities fluctuate, we repeat this procedure multiple times. Subsequently, we average all the data obtained for $p_{q,b}$. The resulting flip probabilities are the ones used for correcting the data in Sec. 4.

B.1.1 Single-qubit case

In Fig. 10 and Fig. 11, we show the flip probabilities we obtained for `ibmq_london` and `ibmq_burlington`. Looking at the data resulting from simulating `ibmq_london` classically with readout noise only in Fig. 10(a), we observe that the flip probabilities our calibration procedure yields scatter around the value provided by the noise model. Using the full noise model does not change the picture a lot, only the values for $p_{0,1}$ scatter slightly more around the value of the noise model, as Fig. 10(b) reveals. The data generated

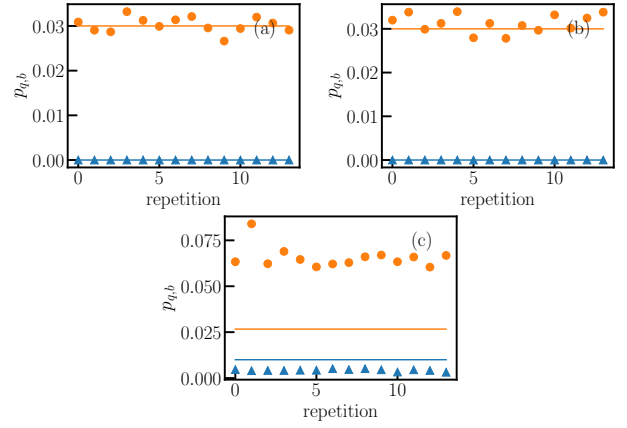


Figure 10: Flip probabilities $p_{0,0}$ (blue triangles) and $p_{0,1}$ (orange dots) for the single qubit case measured with the calibration procedure as a function of the repetition for (a) classically simulating `ibmq_london` with readout errors only, (b) the full noise model, and (c) data obtained on the quantum hardware. The solid lines correspond to the data provided by the noise model.

on the actual `ibmq_london` quantum hardware in Fig. 10(c) does not agree very well with the values of the noise model. Even the values for $p_{0,0}$, which does not involve a single gate, is in general lower than the value provided by the noise model. In contrast, $p_{0,1}$ exceeds the value of the noise model. Despite the fact that the values for the experimentally obtained flip probabilities deviate from the noise model, they only fluctuate moderately and we can extract a reasonable flip probability by averaging over all repetitions. Comparing the different panels of Fig. 10 closely, one can also observe that the values for the flip probabilities provided by the noise model in panel (c) differ slightly from those in panel (a) and (b). The reason for that is that the data in the noise model is updated every day and our classical simulations as well as our simulations on real quantum hardware were not carried out the same day.

The corresponding results for `ibmq_burlington` are shown in Fig. 11. Again, the classical simulation of the chip using the noise model produces as expected flip probabilities in agreement with the values provided. Looking at the data from the real chip in Fig. 11(c), we see that these fluctuate over a wide range between different repetitions. Thus, in this case the flip probabilities cannot be extracted as reliably as for `ibmq_london`. Since our correction procedure relies on being able to estimate the flip probabilities precisely, this partially explains why the improvement in

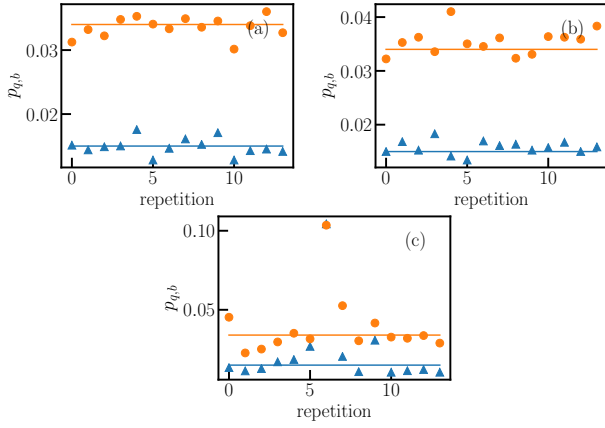


Figure 11: Flip probabilities $p_{0,0}$ (blue triangles) and $p_{0,1}$ (orange dots) for the single qubit case measured with the calibration procedure as a function of the repetition for (a) classically simulating `ibmq_burlington` with readout error only, (b) the full noise model, and (c) data obtained on the hardware. The solid lines correspond to the data provided by the noise model.

Sec. 4.1 after applying the correction to our data for `ibmq_burlington` is smaller.

B.1.2 Two-qubit case

Analogously to the single-qubit case, Fig. 12 and Fig. 13 show the data for extracting the flip probabilities for `ibmq_london` and `ibmq_burlington` obtained in our two-qubit simulations.

The results for the two-qubit case on `ibmq_london` in Fig. 12 show a fairly similar behavior to the single-qubit case. The classical simulations in panels (a) and (b) yield as expected good agreement with the values provided in the noise model. In contrast, the data obtained on the real quantum device (Fig. 12(c)) does not agree with the data in the noise model, in particular for $p_{0,1}$ and $p_{1,0}$. Nevertheless the experimental data is fairly consistent and allows us to reliably determine the flip probabilities for `ibmq_london`. Again, we see that the theoretical values differ between the panels in the upper row and the lower row differ noticeably. This is once more due to the fact that the hardware data was taken on a different day than the simulator data and the noise model has been updated in between.

The flip probabilities obtained from classical simulating `ibmq_burlington` in in Fig. 13(a) and Fig. 13(b) show a similar picture than the previous cases and agree well with the values pro-

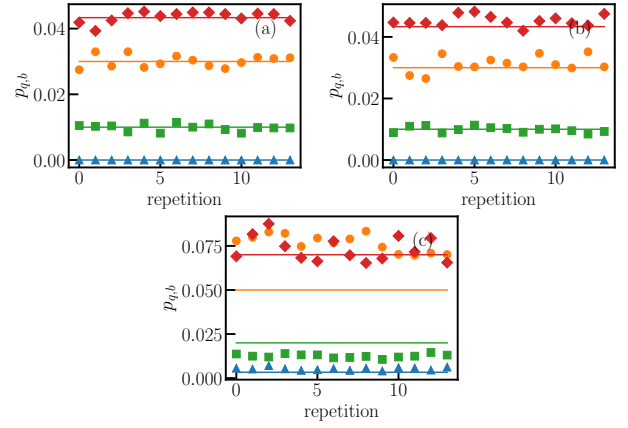


Figure 12: Flip probabilities $p_{0,0}$ (blue triangles), $p_{0,1}$ (orange dots), $p_{1,0}$ (green squares), $p_{1,1}$ (red diamonds) for the two-qubit case measured with the calibration procedure as a function of the repetition for (a) classically simulating `ibmq_london` with readout error only, (b) the full noise model, and (c) data obtained on the hardware. The solid lines correspond to the data provided by the noise model.

vided in the noise model. On the contrary, the data from the real quantum device does again not agree very well with the values provided in the noise model. Moreover, the values for $p_{0,0}$ and $p_{0,1}$ show large fluctuations. In this case as well the theoretical values for the flip probabilities differ between the simulator data and the hardware data. Most noticeably the theoretical value for $p_{1,1}$ almost doubled during the time span between carrying out the classical simulations and the experiments on quantum hardware.

B.2 Technical details for generating the experimental data on quantum hardware

Each of the data points in Fig. 6 to Fig. 9 is obtained by preparing 1050 random wave functions $|\psi\rangle$ using the circuits shown in Sec. 4. While running the 1050 circuits is unproblematic for classical simulations, as of completion of this paper one can only submit 75 circuits per job to real quantum hardware. Thus, we divide them into 14 chunks of 75 circuits. This procedure is repeated for every value of s . Since we have to run a considerable amount of jobs which might take some time depending on how busy the queue of the device is, we insert a job running the circuits for determining the flip probabilities before every chunk. This way we can monitor the flip probabilities over the duration of the run and detect potential outliers.

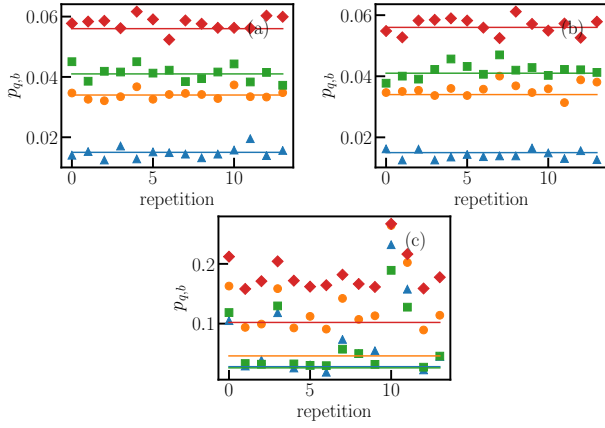


Figure 13: Flip probabilities $p_{0,0}$ (blue triangles), $p_{0,1}$ (orange dots), $p_{1,0}$ (green squares), $p_{1,1}$ (red diamonds) for the two-qubit case measured with the calibration procedure as a function of the repetition for (a) classically simulating `ibmq_burlington` with readout error only, (b) the full noise model, and (c) data obtained on the hardware. The solid lines correspond to the data provided by the noise model.

Moreover, before running our circuits, we use the transpiler to optimize them for the hardware we intend to use. To ensure we have the same mapping between logical and physical qubits in every case, we inspect the transpiler results obtained for the circuits used to extract the bit-flip probabilities and to prepare the random wave function $|\psi\rangle$. For all the data reported in the main text we checked that the mapping between logical and physical qubits is indeed the same and the flip probabilities we extract correspond to the qubits we use for generating our random wave functions.

References

- [1] J. Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, 2018. DOI: [10.22331/q-2018-08-06-79](https://doi.org/10.22331/q-2018-08-06-79).
- [2] F. Arute et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505, 2019. DOI: [10.1038/s41586-019-1666-5](https://doi.org/10.1038/s41586-019-1666-5).
- [3] A. Peruzzo, J. McClean, P. Shadbolt, M. Yung, X. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nat. Comm.*, 5:4213, 2014. DOI: [10.1038/ncomms5213](https://doi.org/10.1038/ncomms5213).
- [4] J. R. McClean, J. Romero, R. Babbush,

- and A. Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.*, 18(2):023023, 2016. DOI: [10.1088/1367-2630/18/2/023023](https://doi.org/10.1088/1367-2630/18/2/023023).
- [5] P. J. J. O’Malley et al. Scalable quantum simulation of molecular energies. *Phys. Rev. X*, 6:031007, 2016. DOI: [10.1103/PhysRevX.6.031007](https://doi.org/10.1103/PhysRevX.6.031007).
- [6] A. Kandala, A. Mezzacapo, K. Temme, et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nat.*, 549:242, 2017. DOI: [10.1038/nature23879](https://doi.org/10.1038/nature23879).
- [7] 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, 2017. DOI: [10.1103/PhysRevA.95.020501](https://doi.org/10.1103/PhysRevA.95.020501).
- [8] 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. DOI: [10.1103/PhysRevX.8.011021](https://doi.org/10.1103/PhysRevX.8.011021).
- [9] 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, 2018. DOI: [10.1103/PhysRevLett.120.210501](https://doi.org/10.1103/PhysRevLett.120.210501).
- [10] C. Hempel et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Phys. Rev. X*, 8:031022, 2018. DOI: [10.1103/PhysRevX.8.031022](https://doi.org/10.1103/PhysRevX.8.031022).
- [11] 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, 2019. DOI: [10.1103/PhysRevApplied.11.044092](https://doi.org/10.1103/PhysRevApplied.11.044092).
- [12] C. Kokail, C. Maier, R. van Bijnen, et al. Self-verifying variational quantum simulation of lattice models. *Nat.*, 569:355, 2019. DOI: [10.1038/s41586-019-1177-4](https://doi.org/10.1038/s41586-019-1177-4).
- [13] T. Hartung and K. Jansen. Zeta-regularized vacuum expectation values. *J.*

- Math. Phys.*, 60(9):093504, 2019. DOI: [10.1063/1.5085866](https://doi.org/10.1063/1.5085866).
- [14] K. Jansen and T. Hartung. Zeta-regularized vacuum expectation values from quantum computing simulations. *Proc. Sci. LATTICE2019*, 363:153, 2020. DOI: [10.22323/1.363.0153](https://doi.org/10.22323/1.363.0153).
 - [15] Y. Li and S. C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X*, 7:021050, 2017. DOI: [10.1103/PhysRevX.7.021050](https://doi.org/10.1103/PhysRevX.7.021050).
 - [16] K. Temme, S. Bravyi, and J. M. Gambetta. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.*, 119:180509, 2017. DOI: [10.1103/PhysRevLett.119.180509](https://doi.org/10.1103/PhysRevLett.119.180509).
 - [17] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A*, 95:042308, 2017. DOI: [10.1103/PhysRevA.95.042308](https://doi.org/10.1103/PhysRevA.95.042308).
 - [18] 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. DOI: [10.1103/PhysRevA.98.062339](https://doi.org/10.1103/PhysRevA.98.062339).
 - [19] S. Endo, S. C. Benjamin, and Y. Li. Practical quantum error mitigation for near-future applications. *Phys. Rev. X*, 8:031027, 2018. DOI: [10.1103/PhysRevX.8.03102](https://doi.org/10.1103/PhysRevX.8.03102).
 - [20] S. McArdle, X. Yuan, and S. Benjamin. Error-mitigated digital quantum simulation. *Phys. Rev. Lett.*, 122:180501, 2019. DOI: [10.1103/PhysRevLett.122.180501](https://doi.org/10.1103/PhysRevLett.122.180501).
 - [21] S. Endo, Q. Zhao, Y. Li, S. Benjamin, and X. Yuan. Mitigating algorithmic errors in a hamiltonian simulation. *Phys. Rev. A*, 99:012334, Jan 2019. DOI: [10.1103/PhysRevA.99.012334](https://doi.org/10.1103/PhysRevA.99.012334).
 - [22] A. Kandala, K. Temme, A. D. Córcoles, et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nat.*, 567:491, 2019. DOI: [10.1038/s41586-019-1040-7](https://doi.org/10.1038/s41586-019-1040-7).
 - [23] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven. Decoding quantum errors with subspace expansions. *Nat. Comm.*, 11(1), 2020. DOI: [10.1038/s41467-020-14341-w](https://doi.org/10.1038/s41467-020-14341-w).
 - [24] M. Otten and S. K. Gray. Recovering noise-free quantum observables. *Phys. Rev. A*, 99:012338, 2019. DOI: [10.1103/PhysRevA.99.012338](https://doi.org/10.1103/PhysRevA.99.012338).
 - [25] M. Otten and S. K. Gray. Accounting for errors in quantum algorithms via individual error reduction. *npj Quantum Inf.*, 5: 11, 2019. DOI: [10.1038/s41534-019-0125-3](https://doi.org/10.1038/s41534-019-0125-3).
 - [26] R. Sagastizabal et al. Experimental error mitigation via symmetry verification in a variational quantum eigensolver. *Phys. Rev. A*, 100:010302, 2019. DOI: [10.1103/PhysRevA.100.010302](https://doi.org/10.1103/PhysRevA.100.010302).
 - [27] M. Urbanek, B. Nachman, and W. A. de Jong. Quantum error detection improves accuracy of chemical calculations on a quantum computer. *arXiv:1910.00129*, 2019. URL <https://arxiv.org/abs/1910.00129>.
 - [28] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley. Efficient quantum measurement of pauli operators in the presence of finite sampling error. *arXiv:1908.06942*, 2019. URL <https://arxiv.org/abs/1908.06942>.
 - [29] B. Chungheon, O. Tomohiro, T. Seigo, and C. Byung-Soo. Density matrix simulation of quantum error correction codes for near-term quantum devices. *Quantum Sci. Technol.*, 5(1):015002, 2019. DOI: [10.1088/2058-9565/ab5887](https://doi.org/10.1088/2058-9565/ab5887).
 - [30] A. D. Córcoles, E. Magesan, S. J. Srinivasan, A. W. Cross, M. Steffen, J. M. Gambetta, and J. M. Chow. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nat. Commun.*, 6(1):6979, 2015. DOI: [10.1038/ncomms7979](https://doi.org/10.1038/ncomms7979).
 - [31] S. Sheldon, E. Magesan, J. M. Chow, and Jay M. Gambetta. Procedure for systematically tuning up cross-talk in the cross-resonance gate. *Phys. Rev. A*, 93(6):060302, 2016. DOI: [10.1103/PhysRevA.93.060302](https://doi.org/10.1103/PhysRevA.93.060302).
 - [32] S. S. Tannu and M. K. Qureshi. Mitigating measurement errors in quantum computers by exploiting state-dependent bias. In *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO ’52, page 279, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450369381. DOI: [10.1145/3352460.3358265](https://doi.org/10.1145/3352460.3358265).

- [33] A. Asfaw et al. Learn quantum computation using qiskit, 2020. URL <http://community.qiskit.org/textbook>. accessed on 2020/07/03.
- [34] Qiskit Aer API documentation and source code. accessed on 2020/07/03.
- [35] E. Lieb, T. Schultz, and D. Mattis. Two soluble models of an antiferromagnetic chain. *Ann. Phys.*, 16(3):407, 1961. DOI: [10.1016/0003-4916\(61\)90115-4](https://doi.org/10.1016/0003-4916(61)90115-4).
- [36] A. Kitaev. Unpaired majorana fermions in quantum wires. *Phys. Usp.*, 44(10S):131, 2001. DOI: [10.1070/1063-7869/44/10s/s29](https://doi.org/10.1070/1063-7869/44/10s/s29).
- [37] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev. Entanglement in quantum critical phenomena. *Phys. Rev. Lett.*, 90:227902, 2003. DOI: [10.1103/PhysRevLett.90.227902](https://doi.org/10.1103/PhysRevLett.90.227902).
- [38] J. I. Latorre, E. Rico, and G. Vidal. Ground state entanglement in quantum spin chains. *Quantum Info. Comput.*, 4(1):48, 2004. URL <http://dl.acm.org/citation.cfm?id=2011572.2011576>.
- [39] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.
- [40] H. Abraham et al. Qiskit: An open-source framework for quantum computing. *Zendo*, 2019. DOI: [10.5281/zenodo.2562110](https://doi.org/10.5281/zenodo.2562110).
- [41] *ibmq_london* v1.0.2, IBM Quantum team. URL <https://quantum-computing.ibm.com>. Retrieved from <https://quantum-computing.ibm.com> (2020).
- [42] *ibmq_burlington* v1.0.2, IBM Quantum team. URL <https://quantum-computing.ibm.com>. Retrieved from <https://quantum-computing.ibm.com> (2020).
- [43] E. Farhi, J. Goldstone, and S. Gutmann. A quantum approximate optimization algorithm. *arXiv:1411.4028*, 2014. URL <https://arxiv.org/abs/1411.4028>.