

ARTICLE OPEN



Unfolding quantum computer readout noise

Benjamin Nachman¹✉, Miroslav Urbanek², Wibe A. de Jong² and Christian W. Bauer¹

In the current era of noisy intermediate-scale quantum computers, noisy qubits can result in biased results for early quantum algorithm applications. This is a significant challenge for interpreting results from quantum computer simulations for quantum chemistry, nuclear physics, high energy physics (HEP), and other emerging scientific applications. An important class of qubit errors are readout errors. The most basic method to correct readout errors is matrix inversion, using a response matrix built from simple operations to probe the rate of transitions from known initial quantum states to readout outcomes. One challenge with inverting matrices with large off-diagonal components is that the results are sensitive to statistical fluctuations. This challenge is familiar to HEP, where prior-independent regularized matrix inversion techniques (“unfolding”) have been developed for years to correct for acceptance and detector effects, when performing differential cross section measurements. We study one such method, known as iterative Bayesian unfolding, as a potential tool for correcting readout errors from universal gate-based quantum computers. This method is shown to avoid pathologies from commonly used matrix inversion and least squares methods.

npj Quantum Information (2020)6:84; <https://doi.org/10.1038/s41534-020-00309-7>

INTRODUCTION

While quantum algorithms are promising techniques for a variety of scientific and industrial applications, current challenges limit their immediate applicability. One significant limitation is the rate of errors and decoherence in noisy intermediate-scale quantum (NISQ) computers¹. Mitigating errors is hard in general because quantum bits (“qubits”) cannot be copied^{2–4}. An important family of errors are readout errors. They typically arise from two sources: (1) measurement times are significant in comparison to decoherence times, and thus a qubit in the $|1\rangle$ state can decay to the $|0\rangle$ state during a measurement, and (2) probability distributions of measured physical quantities that correspond to the $|0\rangle$ and $|1\rangle$ states have overlapping support, and there is a small probability of measuring the opposite value. The goal of this paper is to investigate methods for correcting these readout errors. This is complementary to efforts for gate error corrections. One strategy for mitigating such errors is to build in error correcting components into quantum circuits. Quantum error correction^{5–9} is a significant challenge because qubits cannot be cloned^{2–4}. This generates a significant overhead in the additional number of qubits and gates required to detect or correct errors. Partial error detection/correction has been demonstrated for simple quantum circuits^{10–19}, but complete error correction is infeasible for current qubit counts and moderately deep circuits. As a result, many studies with NISQ devices use the alternative zero noise extrapolation strategy, whereby circuit noise is systematically increased and then extrapolated to zero^{20–25}. Ultimately, both gate and readout errors must be corrected for a complete measurement and a combination of strategies may be required.

Correcting measured histograms for the effects of a detector has a rich history in image processing, astronomy, high energy physics (HEP), and beyond. In the latter, the histograms represent binned differential cross sections and the correction is called unfolding. Many unfolding algorithms have been proposed and are currently in use by experimental high energy physicists (see, e.g., refs. ^{26–28} for reviews). One of the goals of this paper is to introduce these methods and connect them with current

algorithms used by the quantum information science (QIS) community.

Quantum readout error correction can be represented as histogram unfolding, where each bin corresponds to one of the possible $2^{n_{\text{qubit}}}$ configurations, where n_{qubit} is the number of qubits (Fig. 1). Correcting readout noise is a classical problem (though there has been a proposal to do it with quantum annealing²⁹), but relies on calibrations or simulations from quantum hardware. Even though discussions of readout errors appear in many papers (see, e.g., refs. ^{24,30–32}), we are not aware of any dedicated study comparing unfolding methods for QIS applications. Furthermore, current QIS methods have pathologies that can be avoided with techniques from HEP. In particular, the most popular quantum simulators `pyQuil`³³ (by Rigetti), `Cirq`^{34,35} (by Google), and `XACC`^{36–38} implement a version of matrix inversion, and the other popular simulator `Qiskit` by IBM^{39,40} uses a least squares method (see also refs. ^{41,42}) that is the same as the matrix inversion solution when the latter is nonnegative. Challenges with methods based on matrix inversion will be discussed in more detail below.

RESULTS

Unfolded quantum readout

For the results presented here, we simulate a quantum computer using `qiskit-terra` 0.9.0, `qiskit-aer` 0.2.3, and `qiskit-ignis` 0.1.1³⁹. Note that multi-qubit readout errors are not supported in all versions of `qiskit-aer` and `qiskit-ignis`. This work uses a custom measure function to implement the response matrices. We choose a Gaussian distribution as the true distribution, as this is ubiquitous in quantum mechanics as the ground state of the harmonic oscillator. An alternative distribution that is mostly zero with a small number of spikes is presented in Supplementary Fig. 5. This system has been recently studied in the context quantum field theory as a benchmark $0 + 1$ -dimensional noninteracting scalar field theory^{43–50}. In practice, all of the qubits of a system would be entangled to achieve the

¹Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. ²Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. ✉email: bnachman@lbl.gov

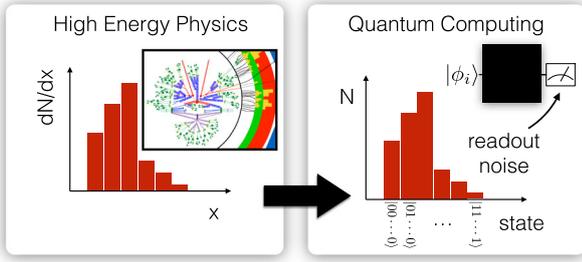


Fig. 1 A schematic diagram illustrating the connection between binned differential cross section measurements in high energy physics (left) and interpreting the output of repeated measurements from quantum computers (right). Given this connection, techniques used to mitigate detector effects for measurements in high energy physics can be studied for readout error corrections in quantum computing.

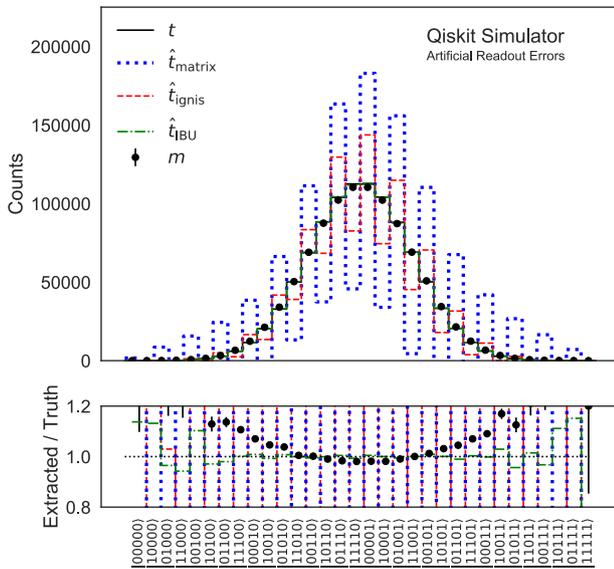


Fig. 2 The measurement of a Gaussian distribution (ground state of harmonic oscillator) using the pathological response matrix. One million total shots are used both to sample from m and to construct R . The IBU method uses ten iterations.

harmonic oscillator wave function. However, this is unnecessary for studying readout errors, which act at the ensemble level. The Gaussian is mapped to qubits using the following map:

$$t(b) \propto \exp\left[-\frac{1}{2\sigma} (b - 2^{n_{\text{qubit}}-1})^2\right], \quad (1)$$

where b is the binary representation of a computational basis state, i.e., $|00000\rangle \mapsto 0$ and $|00011\rangle \mapsto 3$. For the results shown below, $\sigma = 3.5$.

As a first test, a pathological response matrix (see the “Methods” section) is used to illustrate a failure mode of the matrix inversion and *ignis* approaches. Figure 2 compares \hat{t}_{matrix} , \hat{t}_{ignis} , and \hat{t}_{IBU} for a four-qubit Gaussian distribution. As the matrix inversion result is already nonnegative, the \hat{t}_{ignis} result is nearly identical to \hat{t}_{matrix} . Both of these approaches show large oscillations. In contrast, the iterative Bayesian unfolding (IBU) method with ten iterations is nearly the same as the truth distribution. This result is largely insensitive to the choice of iteration number, though it approaches the matrix inversion result for more than about a thousand iterations. This is because the IBU method converges to

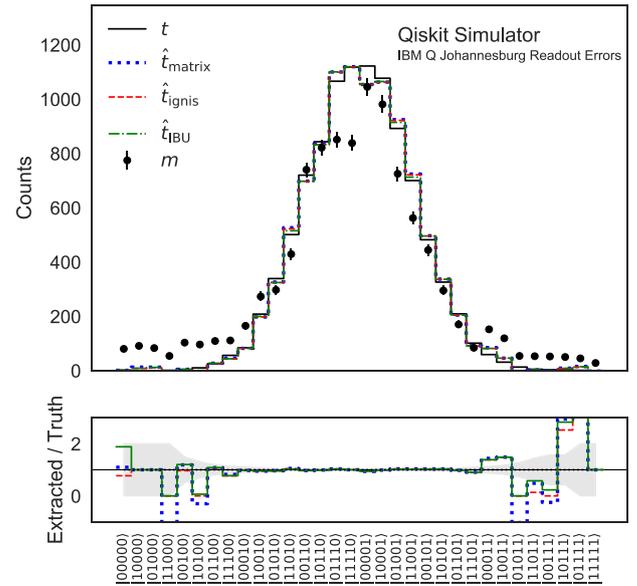


Fig. 3 The measurement of a Gaussian distribution using the response matrix from the IBM Q Johannesburg machine. One million total shots are used to construct R and 10^4 are used for t and m . The IBU method uses 100 iterations. The significant deviations on the far left and right of the distributions are due in part to large statistical fluctuations, where the counts are low. The uncertainty band in the ratio is the statistical uncertainty on t .

a maximum likelihood estimate⁵¹, which in this case aligns with the matrix inversion result. If the matrix inversion would have negative entries, then it would differ from the asymptotic limit of the IBU method, which is always nonnegative.

While Fig. 2 is illustrative for exposing a potential failure mode of matrix inversion and the *ignis* method, it is also useful to consider a realistic response matrix. Figure 3 uses the same Gaussian example as above, but for an IBM Q Johannesburg response matrix (see the “Methods” section). Even though the migrations are large, the pattern is such that all three methods qualitatively reproduce the truth distribution. The dip in the middle of the distribution is due to the mapping between the Gaussian and qubit states: all of the state to the left of the center have a zero as the last qubit, while the ones on the right have a one as the last qubit. The asymmetry of $0 \rightarrow 1$ and $1 \rightarrow 0$ induces the asymmetry in the measured spectrum. For example, it is much more likely to migrate from any state to $|00000\rangle$ than to $|11111\rangle$.

The three unfolding approaches are quantitatively compared in Fig. 4. The same setup as Fig. 2 is used, repeated over 1000 pseudo-experiments. For each of the 1000 pseudo-experiments, a discretized Gaussian state is prepared (same as Fig. 2) with no gate errors and it is measured 10^4 times. The bin counts without readout errors are the “true” values. These values include the irreducible Poisson noise that the unfolding methods are not expected to correct. The measured distribution with readout errors is unfolded using the three methods, and the resulting counts are the “predicted” values. Averaging over many pseudo-experiments, the spread in the predictions is a couple of percent smaller for \hat{t}_{IBU} compared to \hat{t}_{ignis} , and both of these are $\sim 10\%$ more precise than \hat{t}_{matrix} . The slight bias of \hat{t}_{ignis} and \hat{t}_{IBU} to the right results from the fact that they are nonnegative. Similarly, the sharp peak at zero results from \hat{t}_{ignis} and \hat{t}_{IBU} values that are nearly zero when $t_i \sim 0$. In contrast, zero is not special for matrix inversion so there is no particular feature in the center of its distribution.

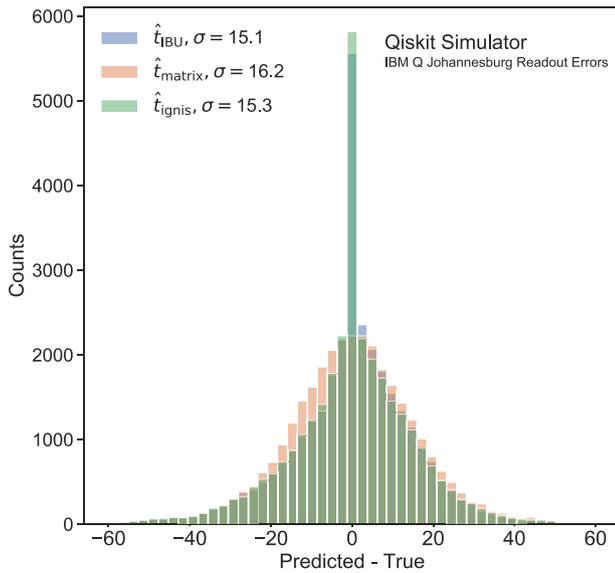


Fig. 4 The distribution of the difference between true and predicted counts from a Gaussian distribution using the response matrix from the IBM Q Johannesburg machine. The simulation is repeated 1000 times. For each of the 1000 pseudo-experiments, a discretized Gaussian state is prepared with no gate errors and it is measured 10^4 times. The bin counts without readout errors are the “true” values. The measured distribution with readout errors is unfolded and the resulting counts are the “predicted” values. Each of the 2^5 states over all 1000 pseudo-experiments contribute one entry to the above histogram. The standard deviations of the distributions are given in the legend. The IBU method uses 100 iterations.

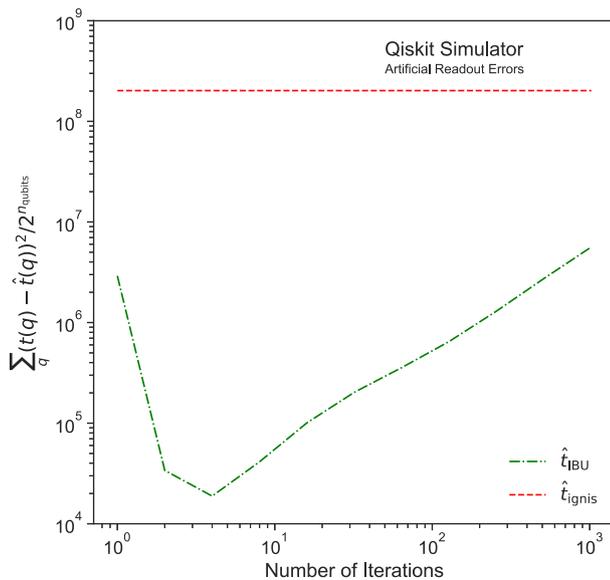


Fig. 5 The difference between \hat{t}_{IBU} and t as a function of the number of iterations for the example presented in Fig. 2. By definition, the `ignis` method does not depend on the number of iterations.

Regularization and uncertainties

One feature of any regularization method is the choice of regularization parameters. For the IBU method, these parameters are the prior and number of iterations. Figure 5 shows the average bias from statistical and systematic sources in the measurement

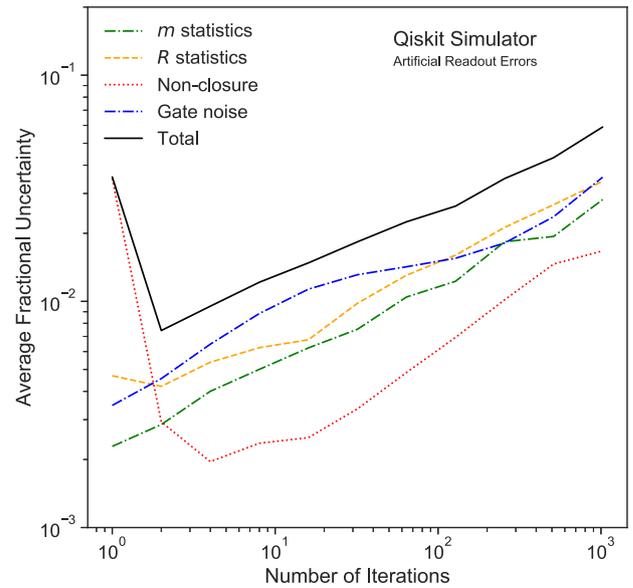


Fig. 6 Sources of uncertainty for \hat{t}_{IBU} as a function of the number of iterations for the example presented in Fig. 2. Each uncertainty is averaged over all states. The total uncertainty is the sum in quadrature of all the individual sources of uncertainty, except gate noise (which is not used in the measurement simulation, but would be present in practice).

from the Gaussian example shown in Fig. 2, as a function of the number of iterations. With a growing number of iterations, \hat{t}_{IBU} approaches the oscillatory \hat{t}_{ignis} . The optimal number of iterations from the point of view of the bias is three. However, the number of iterations cannot be chosen based on the actual bias because the true answer is not known a priori. In HEP, the number of iterations is often chosen before unblinding the data by minimizing the total expected uncertainty. In general, there are three sources of uncertainty: statistical uncertainty on m , statistical and systematic uncertainties on R , and non-closure uncertainties from the unfolding method. Formulae for the statistical uncertainty on m are presented in ref. ⁵² and can also be estimated by bootstrapping⁵³ the measured counts. Similarly, the statistical uncertainty on R can be estimated by bootstrapping and then repeating the unfolding for each calibration pseudo-dataset. The sources of statistical uncertainty are shown as dot-dashed and dashed lines in Fig. 6. Adding more iterations enhances statistical fluctuations and so these sources of uncertainty increase monotonically, with the number of iterations.

The systematic uncertainty on R and the method non-closure uncertainty are not unique and require careful consideration. In HEP applications, R is usually determined from simulation, so the systematic uncertainties are simulation variations that try to capture potential sources of mis-modeling. These simulation variations are often estimated from auxiliary measurements with data. There are additional uncertainties related to the modeling of background processes, as well as related to events that fall into or out of the measurement volume. These are not relevant for QIS. In the QIS context, R is determined directly from the data, so the only uncertainty is on the impurity of the calibration circuits. In particular, the calibration circuits are constructed from a series of single-qubit X gates. Due to gate imperfections and thermal noise, there is a chance that the application of an X gate will have a different effect on the state than intended. In principle, one can try to correct for such potential sources of bias by an extrapolation method. In such methods, the noise is increased in a controlled fashion and then extrapolated to zero noise^{20–24}. This method may have a residual bias and the uncertainty on the method

would then become the systematic uncertainty on R . A likely conservative alternative to this approach is to modify R by adding in gate noise, and taking the difference between the nominal result and one with additional gate noise, simulated using the `thermal_relaxation_error` functionality of `qiskit`. This is the choice made in Fig. 6, where the gate noise is shown as a dot-dashed line. In this particular example, the systematic uncertainty on R increases monotonically with the number of iterations, just like the sources of statistical uncertainty.

The non-closure uncertainty is used to estimate the potential bias from the unfolding method. One possibility is to compare multiple unfolding methods and take the spread in predictions as an uncertainty. Another method advocated in ref. 54 and widely used in HEP is to perform a data-driven reweighting. The idea is to reweight the t^0 so that when folded with R , the induced m^0 is close to the measurement m . Then, this reweighted m^0 is unfolded with the nominal response matrix and compared with the reweighted t^0 . The difference between these two is an estimate of the non-closure uncertainty. The reweighting function is not unique, but should be chosen so that the reweighted t^0 is a reasonable prior for the data. For Fig. 6, the reweighting is performed using the nominal unfolded result itself. In practice, this can be performed in a way that is blinded from the actual values of \hat{t}_{IBU} so that the experimenter is not biased, when choosing the number of iterations.

Altogether, the sources of uncertainty presented in Fig. 6 show that the optimal choice for the number of iterations is 2. In fact, the difference in the uncertainty between two and three iterations is <1% and so consistent with the results from Fig. 5. Similar plots for the measurement in Fig. 3 can be found in Supplementary Figs. 3 and 4.

DISCUSSION

This work has introduced a suite of readout error correction algorithms developed in HEP for binned differential cross section measurements. These unfolding techniques are well-suited for quantum computer readout errors, which are naturally binned and without acceptance effects (counts are not lost or gained during readout). In particular, the iterative Bayesian method has been described in detail, and shown to be robust to a failure mode of the matrix inversion and `ignis` techniques. When readout errors are sufficiently small, all the methods perform well, with a preference for the `ignis` and Bayesian methods that produce nonnegative results. The `ignis` method is a special case of the `TUnfold` algorithm, where the latter uses the covariance matrix to improve precision and incorporates regularization to be robust to the failure modes of matrix inversion. It may be desirable to augment the `ignis` method with these features or provide the iterative method as an alternative approach. In either case, Fig. 3 showed that even with a realistic response matrix, readout error corrections can be significant and must be accounted for in any measurement on near-term hardware.

An important challenge facing any readout error correction method is the exponential resources required to construct the full R matrix (see the “Methods” section). While R must be constructed only once per hardware setup and operating condition, it could become prohibitive when the number of qubits is large. On hardware with few connections between qubits, per-qubit transition probabilities may be sufficient for accurate results. When that is not the case, one may be able to achieve the desired precision with polynomially many measurements. These ideas are left to future studies.

Another challenge is optimizing the unfolding regularization. The numerical examples presented in the “Results” section considered various sources of uncertainty, and studied how they depend on the number of iterations in the IBU method. The full measurement is performed for all $2^{n_{\text{qubit}}}$ states and the studies in

the “Results” section collapsed the uncertainty across all bins into a single number by averaging across bins. This way of choosing a regularization parameter is common in HEP, but is not a unique approach. Ultimately, a single number is required for optimization, but it may be that other metrics are more important for specific applications, such as the uncertainty for a particular expectation value, or the maximum or most probable uncertainty across states. Such requirements are application specific, but should be carefully considered prior to the measurement.

With active research and development across a variety of application domains, there are many promising applications of quantum algorithms in both science and industry on NISQ hardware. Readout errors are an important source of noise that can be corrected to improve measurement fidelity. HEP experimentalists have been studying readout error correction techniques for many years under the term unfolding. These tools are now available to the QIS community and will render the correction procedure more robust to resolution effects, in order to enable near-term breakthroughs.

METHODS

The unfolding challenge

Let t be a vector that represents the true bin counts before the distortions from detector effects (HEP) or readout noise (QIS). The corresponding measured bin counts are denoted by m . These vectors are related by a response matrix R as $m = Rt$, where $R_{ij} = \text{Pr}(m = i | t = j)$. In HEP, the matrix R is usually estimated from detailed detector simulations, while in QIS, R is constructed from measurements of computational basis states. The response matrix construction is discussed in more detail below.

The most naive unfolding procedure would be to simply invert the matrix R : $\hat{t}_{\text{matrix}} = R^{-1}m$. However, simple matrix inversion has many known issues. Two main problems are that \hat{t}_{matrix} can have unphysical entries, and that statistical uncertainties in R can be amplified and can result in oscillatory behavior. For example, consider the case

$$R = \begin{pmatrix} 1 - \epsilon & \epsilon \\ \epsilon & 1 - \epsilon \end{pmatrix}, \quad (2)$$

where $0 < \epsilon < 1/2$. Then, $\text{Var}(\hat{t}_{\text{matrix}}) \propto 1/\det(R) = 1/(1 - 2\epsilon) \rightarrow \infty$ as $\epsilon \rightarrow 1/2$. As a generalization of this example to more bins (from ref. 55), consider a response matrix with a symmetric probability of migrating one bin up or down,

$$R = \begin{pmatrix} 1 - \epsilon & \epsilon & 0 & \cdots \\ \epsilon & 1 - 2\epsilon & \epsilon & \cdots \\ 0 & \epsilon & 1 - 2\epsilon & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (3)$$

Unfolding with the above response matrix and $\epsilon = 25\%$ is presented in Fig. 7. The true bin counts have the Gaussian distribution with a mean of zero and a standard deviation of 3. The values are discretized with 21 uniform unit width bins spanning the interval $[-10, 10]$. The leftmost bin correspond to the first index in m and t . The indices are monotonically incremented with increasing bin number. The first and last bins include underflow and overflow values, respectively. Due to the symmetry of the migrations, the true and measured distributions are nearly the same. For this reason, the measured spectrum happens to align with the true distribution; the optimal unfolding procedure should be the identity. The significant off-diagonal components result in an oscillatory behavior and the statistical uncertainties are also amplified by the limited size of the simulation dataset used to derive the response matrix.

Even though it is simple, matrix inversion is an important benchmark for comparing with the methods described below because it is widely used in quantum computing, as introduced at the end of the “Introduction” section. Many of these implementations are even simpler than full matrix inversion by assuming that errors are uncorrelated across qubits. In some cases, additional mitigation strategies are used to make the matrix inversion corrections smaller. This is demonstrated in ref. 35 (see also ref. 56), where a symmetrization step is first applied with X gates prior to matrix inversion. A minimal modification to matrix inversion is to ensure that the counts are nonnegative. One such approach is to use a least

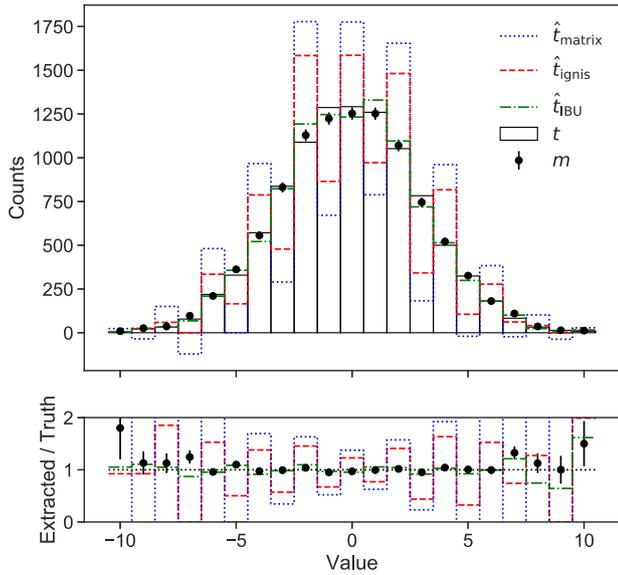


Fig. 7 A comparison of unfolding techniques for a Gaussian example and the R matrix from Eq. (3). The symbols t and m denote the true and measured probability mass functions, respectively. Simple matrix inversion is represented by \hat{t}_{matrix} . The ignis and IBU methods are represented by \hat{t}_{ignis} and \hat{t}_{IBU} , respectively. For this example, $\|m\|_1 = 10^4$ and R is assumed to be known exactly. The IBU method uses ten iterations and a uniform prior (other iterations choices are studied in Supplementary Fig. 1). The simulation used in this plot is based on standard Python functions and does not use a quantum computer simulator (see instead Fig. 2).

squares method^{40–42} that will be described in more detail in the next section.

Unfolding methods

The fact that matrix inversion can result in unphysical outcomes ($\hat{t}_i < 0$ or $\hat{t}_i > \|t\|_1$, where $\|x\|_1 = \sum |x_i|$ is the L^1 norm) is often unacceptable. One solution is to find a vector that is as close to \hat{t} as possible, but with physical entries. This is the method implemented in `qiskit-ignis`^{39,40}, a widely used quantum computation software package. The `ignis` method solves the optimization problem

$$\hat{t}_{\text{ignis}} = \arg \min_{t': \|t'\|_1 = \|m\|_1, t' > 0} \|m - Rt'\|^2. \quad (4)$$

The same method was also recently studied by other authors as well, see, e.g., refs. ^{41,42}. Note that the norm invariance is also satisfied for simple matrix inversion: $\|\hat{t}_{\text{matrix}}\|_1 = \|m\|_1$ by construction because $\|Rx\|_1 = \|x\|_1$ for all x so in particular for $y = R^{-1}m$, $\|Ry\|_1 = \|y\|_1$ implies that $\|m\|_1 = \|R^{-1}m\|_1$. This means that $\hat{t}_{\text{ignis}} = \hat{t}_{\text{matrix}}$ whenever the latter is nonnegative, and so \hat{t}_{ignis} inherits some of the pathologies of \hat{t}_{matrix} .

Three commonly used unfolding methods in HEP are IBU⁵² (also known as Richardson–Lucy deconvolution^{57,58}), singular value decomposition (SVD) unfolding⁵⁹, and TUnfold⁶⁰. There are other less widely used methods, such as fully Bayesian unfolding⁶¹ and others^{62–66}. Even though IBU calls for the repeated application of Bayes’ theorem, this is not a Bayesian method as there is no prior/posterior over possible distributions, only a prior for initialization. TUnfold is similar to Eq. (4), but imposes further regularization requirements in order to avoid pathologies from matrix inversion. The SVD approach applies some regularization directly on R before applying matrix inversion. The focus of this paper will be on the widely used IBU method, which avoids fitting and matrix inversion altogether with an iterative approach. Reference ⁵² also suggested that one could combine the iterative method with smoothing from a fit, in order to suppress amplified statistical fluctuations.

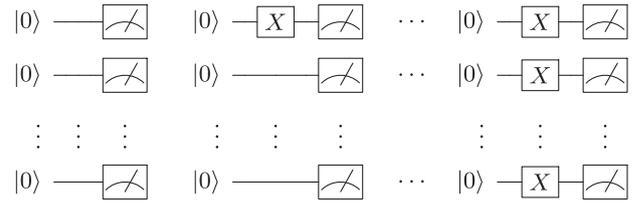


Fig. 8 The set of 2^n qubit calibration circuits. The calibration circuits are simple combinations of initialization followed by measurement with all possible combinations of single X gates in between.

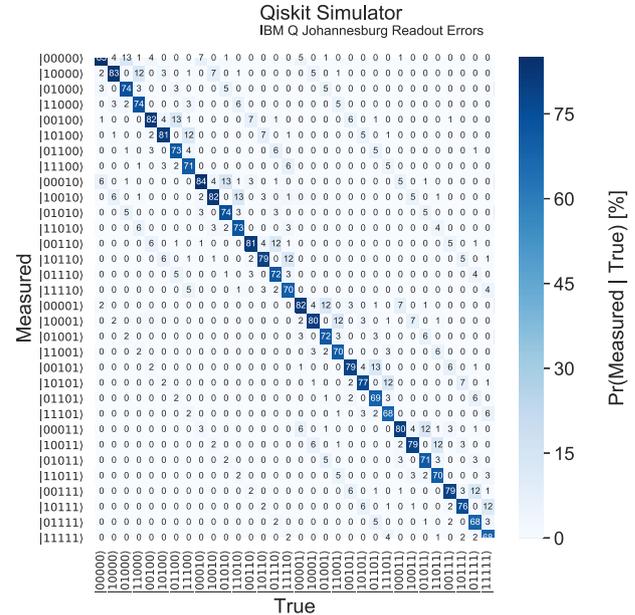


Fig. 9 An example R matrix from five qubits of the IBM Q Johannesburg machine using 8192 shots for each of the 2^5 possible states. Note that this matrix depends on the calibration quality, so while it is representative, it is not precisely the version valid for all measurements made on this hardware.

Given a prior truth spectrum $t_i^0 = \Pr(\text{truth is } i)$, the IBU technique proceeds according to the equation

$$\begin{aligned} t_i^{n+1} &= \sum_j \Pr(\text{truth is } i | \text{measure } j) \times m_j \\ &= \sum_j \sum_k \frac{R_{jk} t_k^n}{\sum_k R_{jk} t_k^n} \times m_j, \end{aligned} \quad (5)$$

where n is the iteration number and one iterates a total of N times. The advantage of Eq. (5) over simple matrix inversion is that the result is a probability (nonnegative and unit measure) when $m \geq 0$. In HEP applications, m can have negative entries resulting from background subtraction. In this case, the unfolded result can also have negative entries. The parameters t_i^0 and N must be specified ahead of time. A common choice for t_i^0 is the uniform distribution. The number of iterations needed to converge depends on the desired precision, how close t_i^0 is to the final distribution, and the importance of off-diagonal components in R . In practice, it may be desirable to choose a relatively small N prior to convergence to regularize the result. Typically, $\lesssim \mathcal{O}(10)$ iterations are needed.

In addition to the \hat{t}_{matrix} and \hat{t}_{ignis} approaches discussed in the previous section, Fig. 7 shows the IBU result with $N = 10$. Unlike the \hat{t}_{matrix} and \hat{t}_{ignis} results, the \hat{t}_{IBU} does not suffer from rapid oscillations and like \hat{t}_{ignis} , is nonnegative. Analogous results for quantum computer simulations are presented in the “Results” section.

Constructing the response matrix

In practice, the R matrix is not known exactly, and must be measured for each quantum computer and set of operating conditions. One way to

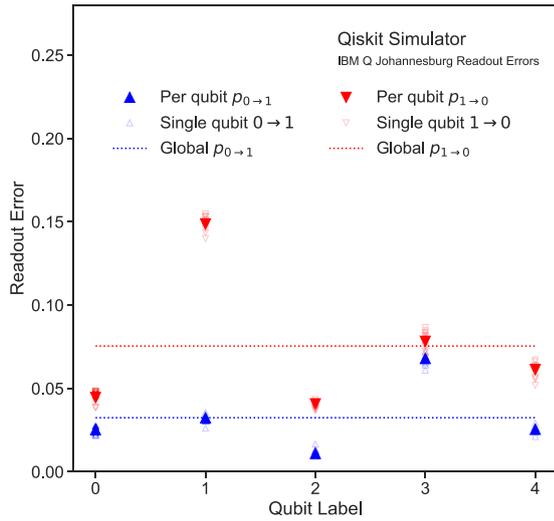


Fig. 10 Test of universality in the response matrix R . Horizontal dotted lines show the result of a global fit to $p_{0 \rightarrow 1}$ and $p_{1 \rightarrow 0}$, assuming an equal rate of readout errors for each qubit. Big filled markers represent a fit with independent values for each qubit. Small semitransparent open markers show the $2^{n_{\text{qubit}}-1}$ transition probabilities for each qubit, when the other $n_{\text{qubit}} - 1$ qubits are unchanged between truth and measured states.

measure R is to construct a set of $2^{n_{\text{qubit}}}$ calibration circuits as shown in Fig. 8. Simple X gates are used to prepare all of the possible qubit configurations and then they are immediately measured.

Figure 9 shows the R matrix from five qubits of the IBM Q Johannesburg machine. As expected, there is a significant diagonal component that corresponds to cases, when the measured and true states are the same. However, there are significant off-diagonal components, which are larger toward the right when more configurations start in the one state. The diagonal stripes with transition probabilities of $\sim 5\text{--}7\%$ are the result of the same qubit flipping from $0 \leftrightarrow 1$. This matrix is hardware dependent and its elements can change over time due to calibration drift. Machines with higher connectivity have been observed to have more readout noise.

Figure 10 explores the universality of qubit migrations across R . If every qubit was identical, and there were no effects from the orientation and connectivity of the computer, then one may expect that R can actually be described by just two numbers $p_{0 \rightarrow 1}$ and $p_{1 \rightarrow 0}$, the probability for a zero to be measured as a one and vice versa. These two numbers are extracted from R by performing the fit in Eq. (6):

$$\min_{p_{0 \rightarrow 1}, p_{1 \rightarrow 0}} \sum_{ij} \left| R_{ij} - p_{0 \rightarrow 1}^a (1 - p_{0 \rightarrow 1})^{\alpha'} p_{1 \rightarrow 0}^{\beta} (1 - p_{1 \rightarrow 0})^{\beta'} \right|^2, \quad (6)$$

where a is the number of qubits corresponding to the response matrix entry R_{ij} that flipped from 0 to 1, α' is the number that remained as a 0 and β, β' are the corresponding entries for a one state; $a + \alpha' + \beta + \beta' = n_{\text{qubit}}$. For example, if the R_{ij} entry corresponds to the state $|01101\rangle$ migrating to $|01010\rangle$, then $a = 1$, $\alpha' = 1$, $\beta = 2$, and $\beta' = 1$.

A global fit to these parameters for the Johannesburg machine results in $p_{0 \rightarrow 1} \approx 3.2\%$ and $p_{1 \rightarrow 0} \approx 7.5\%$. In reality, the qubits are not identical and so one may expect that $p_{0 \rightarrow 1}$ and $p_{1 \rightarrow 0}$ depend on the qubit. A fit to n_{qubit} values for $p_{0 \rightarrow 1}$ and $p_{1 \rightarrow 0}$ (Eq. (6), but fitting ten parameters instead of two) are shown as filled triangles in Fig. 10. While these values cluster around the universal values (dotted lines), the spread is a relative 50% for $p_{0 \rightarrow 1}$ and 60% for $p_{1 \rightarrow 0}$. Furthermore, the transition probabilities can depend on the values of neighboring qubits. The open markers in Fig. 10 show the transition probabilities for each qubit with the other $n_{\text{qubit}} - 1$ qubits held in a fixed state. In other words, the $2^{n_{\text{qubit}}-1}$ open triangles for each qubit show

$$\Pr(|q_0, \dots, q_i, \dots, q_{n_{\text{qubit}}}\rangle \rightarrow |q_0, \dots, q', \dots, q_{n_{\text{qubit}}}\rangle), \quad (7)$$

where $(q_i, q') \in \{(0, 1), (1, 0)\}$ with the other qubits q_j held in a fixed configuration. The spread in these values is smaller than the variation in the solid markers, which indicates that per-qubit readout errors are likely sufficient to capture most of the salient features of the response matrix. This spread has a contribution from connection effects, but also from Poisson

fluctuations as each measurement is statistically independent. However, this is hardware dependent and higher connectivity computers may depend more on the state of neighboring qubits.

Constructing the entire response matrix requires exponential resources in the number of qubits. While the measurement of R only needs to be performed once per quantum computer per operational condition, this can be untenable when $n_{\text{qubit}} \gg 1$. The tests above indicate that a significant fraction of the $2^{n_{\text{qubit}}}$ calibration circuits may be required for a precise measurement of R . Sub-exponential approaches may be possible and will be studied in future work.

DATA AVAILABILITY

The data shown in this paper are available upon request from the corresponding author.

CODE AVAILABILITY

The code for the work presented here can be found at <https://github.com/bnachman/QISunfolding>.

Received: 15 October 2019; Accepted: 11 August 2020;

Published online: 25 September 2020

REFERENCES

- Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
- Park, J. L. The concept of transition in quantum mechanics. *Found. Phys.* **1**, 23–33 (1970).
- Wootters, W. K. & Zurek, W. H. A single quantum cannot be cloned. *Nature* **299**, 802–803 (1982).
- Dieks, D. Communication by EPR devices. *Phys. Lett. A* **92**, 271–272 (1982).
- Gottesman, D. An introduction to quantum error correction and fault-tolerant quantum computation. Preprint at <https://arxiv.org/abs/0904.2557> (2009).
- Devitt, S. J., Munro, W. J. & Nemoto, K. Quantum error correction for beginners. *Rep. Prog. Phys.* **76**, 076001 (2013).
- Terhal, B. M. Quantum error correction for quantum memories. *Rev. Mod. Phys.* **87**, 307–346 (2015).
- Lidar, D. A. & Brun, T. A. *Quantum Error Correction* (Cambridge University Press, 2013).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th edn (Cambridge University Press, New York, NY, 2011).
- Miroslav Urbaneck, B. N. & de Jong, W. A. Quantum error detection improves accuracy of chemical calculations on a quantum computer. *Phys. Rev. A* **102**, 022427 (2020).
- Wootton, J. R. & Loss, D. Repetition code of 15 qubits. *Phys. Rev. A* **97**, 052313 (2018).
- Barends, R. et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500–503 (2014).
- Kelly, J. et al. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature* **519**, 66–69 (2015).
- Linke, N. M. et al. Fault-tolerant quantum error detection. *Sci. Adv.* **3**, e1701074 (2017).
- Takita, M., Cross, A. W., Córcoles, A. D., Chow, J. M. & Gambetta, J. M. Experimental demonstration of fault-tolerant state preparation with superconducting qubits. *Phys. Rev. Lett.* **119**, 180501 (2017).
- Roffe, J., Headley, D., Chancellor, N., Horsman, D. & Kendon, V. Protecting quantum memories using coherent parity check codes. *Quantum Sci. Technol.* **3**, 035010 (2018).
- Vuillot, C. Is error detection helpful on IBM 5Q chips? *Quantum Inf. Comput.* **18**, 0949–0964 (2018).
- Willsch, D., Willsch, M., Jin, F., De Raedt, H. & Michielsen, K. Testing quantum fault tolerance on small systems. *Phys. Rev. A* **98**, 052348 (2018).
- Harper, R. & Flammia, S. T. Fault-tolerant logical gates in the IBM Quantum Experience. *Phys. Rev. Lett.* **122**, 080504 (2019).
- Kandala, A. et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019).
- Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050 (2017).
- Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).

23. Endo, S., Benjamin, S. C. & Li, Y. Practical quantum error mitigation for near-future applications. *Phys. Rev. X* **8**, 031027 (2018).
24. Dumitrescu, E. F. et al. Cloud quantum computing of an atomic nucleus. *Phys. Rev. Lett.* **120**, 210501 (2018).
25. He, A., Jong, W. A. d., Nachman, B. & Bauer, C. Resource efficient zero noise extrapolation with identity insertions. *Phys. Rev. A* **102**, 012426 (2020).
26. Cowan, G. A survey of unfolding methods for particle physics. *Conf. Proc. C0203181*, 248 (2002).
27. Blobel, V. Unfolding methods in particle physics. In *PHYSTAT 2011 Proceedings*, 240–251 (CERN, Geneva, 2011).
28. Blobel, V. In *Data Analysis in High Energy Physics*, Ch. 6, 187–225 (Wiley, Hoboken, 2013).
29. Cormier, K., Di Sipio, R. & Wittek, P. Unfolding as quantum annealing. *JHEP* **11**, 128 (2019).
30. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
31. Klco, N. & Savage, M. J. Minimally-entangled state preparation of localized wavefunctions on quantum computers. *Phys. Rev. A* **102**, 012612 (2020).
32. Yeter-Aydeniz, K. et al. Scalar quantum field theories as a benchmark for near-term quantum computers. *Phys. Rev. A* **99**, 032306 (2019).
33. Rigetti Forest Software Development Kit. *Source Code for pyquil.noise*. <http://docs.rigetti.com/en/stable/noise.html> (2020).
34. The Cirq Contributors. *Cirq, a Python Framework for Creating, Editing, and Invoking Noisy Intermediate Scale Quantum (NISQ) circuits*. <https://github.com/quantumlib/Cirq> (2020).
35. Arute, F. et al. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. Preprint at <https://arxiv.org/abs/2004.04197> (2020).
36. McCaskey, A. J., Lyakh, D. I., Dumitrescu, E. F., Powers, S. S. & Humble, T. S. Xacc: A system-level software infrastructure for heterogeneous quantum-classical computing. Preprint at <https://arxiv.org/abs/1911.02452> (2019).
37. The XACC Contributors. *XACC Documentation*. <https://xacc.readthedocs.io/en/latest/> (2020).
38. McCaskey, A. J. et al. Quantum chemistry as a benchmark for near-term quantum computers. *npj Quantum Inf.* **5**, 99 (2019).
39. IBM Research. *Qiskit*. <https://qiskit.org> (2019).
40. IBM Research. *Qiskit Ignis*. <https://qiskit.org/ignis> (2019).
41. Yanzhu Chen, S. Y., Farahzad, M. & Wei, T.-C. Detector tomography on ibm 5-qubit quantum computers and mitigation of imperfect measurement. *Phys. Rev. A* **100**, 052315 (2019).
42. Maciejewski, Filip B., Zimborás, Z. & Oszmaniec, M. Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum* **4**, 257 (2020).
43. Jordan, S. P., Krovi, H., Lee, K. S. M. & Preskill, J. BQP-completeness of scattering in scalar quantum field theory. *Quantum* **2**, 44 (2018).
44. Jordan, S. P., Lee, K. S. M. & Preskill, J. Quantum computation of scattering in scalar quantum field theories. *Quant. Inf. Comput.* **14**, 1014–1080 (2014).
45. Jordan, S. P., Lee, K. S. M. & Preskill, J. Quantum algorithms for quantum field theories. *Science* **336**, 1130–1133 (2012).
46. Jordan, S. P., Lee, K. S. M. & Preskill, J. Quantum algorithms for fermionic quantum field theories. Preprint at <https://arxiv.org/abs/1404.7115> (2014).
47. Somma, R. D. Quantum simulations of one dimensional quantum systems. *Quantum Info. Comput.* **16**, 1125?1168 (2016).
48. Macridin, A., Spentzouris, P., Amundson, J. & Harnik, R. Electron-phonon systems on a universal quantum computer. *Phys. Rev. Lett.* **121**, 110504 (2018).
49. Macridin, A., Spentzouris, P., Amundson, J. & Harnik, R. Digital quantum computation of fermion-boson interacting systems. *Phys. Rev. A* **98**, 042312 (2018).
50. Klco, N. & Savage, M. J. Digitization of scalar fields for quantum computing. *Phys. Rev. A* **99**, 052335 (2019).
51. Shepp, L. A. & Vardi, Y. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging* **1**, 113–122 (1982).
52. D'Agostini, G. A Multidimensional unfolding method based on Bayes' theorem. *Nucl. Instrum. Methods Phys. Res. A* **362**, 487–498 (1995).
53. Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1–26 (1979).
54. Malaescu, B. An iterative, dynamically stabilized method of data unfolding. Preprint at <https://arxiv.org/abs/0907.3791> (2009).
55. Blobel, V. Unfolding methods in high-energy physics experiments. In *Proceedings, CERN School of Computing DESY-84-118* 88–127 (CERN School of Computing, Aigua Blava, 1984).
56. Tannu, S. S. & Qureshi, M. K. 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*, 279–290 (Association for Computing Machinery, New York, NY, 2019).
57. Lucy, L. B. An iterative technique for the rectification of observed distributions. *Astron. J.* **79**, 745 (1974).
58. Richardson, W. H. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.* **62**, 55–59 (1972).
59. Hocker, A. & Kartvelishvili, V. SVD approach to data unfolding. *Nucl. Instrum. Methods Phys. Res. A* **372**, 469–481 (1996).
60. Schmitt, S. TUnfold: an algorithm for correcting migration effects in high energy physics. *J. Instrum.* **7**, T10003 (2012).
61. Choudalakis, G. Fully Bayesian unfolding. Preprint at <https://arxiv.org/abs/1201.4612> (2012).
62. Gagunashvili, N. D. Machine learning approach to inverse problem and unfolding procedure. Preprint at <https://arxiv.org/abs/1004.2006> (2010).
63. Glazov, A. Machine learning as an instrument for data unfolding. Preprint at <https://arxiv.org/abs/1712.01814> (2017).
64. Datta, K., Kar, D. & Roy, D. Unfolding with generative adversarial networks. Preprint at <https://arxiv.org/abs/1806.00433> (2018).
65. Zech, G. & Aslan, B. Binning-free unfolding based on Monte Carlo Migration. In *PHYSTAT 2003 Proceedings*, Vol. C030908, TUGT001 (SLAC, Stanford, CA, 2003).
66. Lindemann, L. & Zech, G. Unfolding by weighting Monte Carlo events. *Nucl. Instrum. Methods Phys. Res. A* **354**, 516–521 (1995).

ACKNOWLEDGEMENTS

This work is supported by the U.S. Department of Energy, Office of Science under contract DE-AC02-05CH11231. In particular, support comes from Quantum Information Science Enabled Discovery (QuantISED) for High Energy Physics (KA2401032) and the Office of Advanced Scientific Computing Research (ASCR) through the Quantum Algorithms Team program. We acknowledge access to quantum chips and simulators through the IBM Quantum Experience and Q Hub Network through resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. B.N. would also like to thank Michael Geller for spotting a typo, Jesse Thaler for stimulating discussions about unfolding, and the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611.

AUTHOR CONTRIBUTIONS

B.N. conceived the project idea, wrote the code, performed the numerical analysis, and wrote the manuscript. W.A.d.J. organized the measurements on IBM Q. All authors discussed the results and revised the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41534-020-00309-7>.

Correspondence and requests for materials should be addressed to B.N.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.