



VarSaw: Application-tailored Measurement Error Mitigation for Variational Quantum Algorithms

Siddharth Dangwal*
University of Chicago
USA

Kaitlin N. Smith
Infleqtion
USA

Gokul Subramanian Ravi*
University of Michigan
USA

Jonathan M. Baker
The University of Texas at Austin
USA

Poulami Das
The University of Texas at Austin
USA

Frederic T. Chong
University of Chicago
USA

ABSTRACT

For potential quantum advantage, Variational Quantum Algorithms (VQAs) need high accuracy beyond the capability of today's NISQ devices, and thus will benefit from error mitigation. In this work we are interested in mitigating measurement errors which occur during qubit measurements after circuit execution and tend to be the most error-prone operations, especially detrimental to VQAs. Prior work, JigSaw, has shown that measuring only small subsets of circuit qubits at a time and collecting results across all such 'subset' circuits can reduce measurement errors. Then, running the entire ('global') original circuit and extracting the qubit-qubit measurement correlations can be used in conjunction with the subsets to construct a high-fidelity output distribution of the original circuit. Unfortunately, the execution cost of JigSaw scales polynomially in the number of qubits in the circuit, and when compounded by the number of circuits and iterations in VQAs, the resulting execution cost quickly turns insurmountable.

To combat this, we propose VarSaw, which improves JigSaw in an application-tailored manner, by identifying considerable redundancy in the JigSaw approach for VQAs: spatial redundancy across subsets from different VQA circuits and temporal redundancy across globals from different VQA iterations. VarSaw then eliminates these forms of redundancy by *commuting* the subset circuits and *selectively executing* the global circuits, reducing computational cost (in terms of the number of circuits executed) over naive JigSaw for VQA by 25x on average and up to 1000x, for the same VQA accuracy. Further, it can recover, on average, 45% of the infidelity from measurement errors in the noisy VQA baseline. Finally, it improves fidelity by 55%, on average, over JigSaw for a fixed computational budget. VarSaw can be accessed here: <https://github.com/siddharthdangwal/VarSaw>

*Both authors contributed equally. Correspondence: siddharthdangwal@uchicago.edu, gsravi@umich.edu

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASPLOS '23, March 25–29, 2023, Vancouver, BC, Canada

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0394-2/23/03...\$15.00

<https://doi.org/10.1145/3623278.3624764>

CCS CONCEPTS

- Computer systems organization → Quantum computing.

KEYWORDS

Quantum Computing, Variational Quantum Algorithms, Error Mitigation, Measurement Errors, Noisy Intermediate-Scale Quantum, Variational Quantum Eigensolver

ACM Reference Format:

Siddharth Dangwal*, Gokul Subramanian Ravi*, Poulami Das, Kaitlin N. Smith, Jonathan M. Baker, and Frederic T. Chong, 2023. VarSaw: Application-tailored Measurement Error Mitigation for Variational Quantum Algorithms. In *28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 4 (ASPLOS '23), March 25–29, 2023, Vancouver, BC, Canada*. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3623278.3624764>

1 INTRODUCTION

Quantum computers leverage superposition, interference, and entanglement to give them potentially significant computing advantage in chemistry [28], optimization [38], machine learning [10] and other domains of critical interest.

In near-term quantum computing, called Noisy Intermediate-Scale Quantum (NISQ), we expect to work with machines that comprise 100-1000s of imperfect qubits [43].

NISQ devices suffer from high error rates in the form of state preparation and measurement (SPAM) errors, gate errors, qubit decoherence, crosstalk, etc. One of the most promising quantum tasks for near-term quantum advantage in the NISQ era are variational quantum algorithms (VQAs). They have wide application in approximation [38], chemistry [41] etc, and are usually designed as minimization problems. VQAs are hybrid quantum-classical algorithms iteratively run a parameterized quantum circuit (called an *ansatz*) on the quantum machine. The QC parameters are optimized each iteration by a classical tuner/optimizer to try and approach the global minimum of the variational objective function. The classical optimizer attempts to adjust the QC to the noise characteristics of the quantum device, and hence, *in theory*, gives VQAs the potential for quantum advantage even on noisy machines. But *in practice*, noise has prevented current quantum computers from surpassing the capabilities of classical computers in almost all applications, including VQAs. Considering the potential for VQAs on NISQ machines [20, 41], it is imperative to explore techniques to maximize their quality of execution on today's machines.

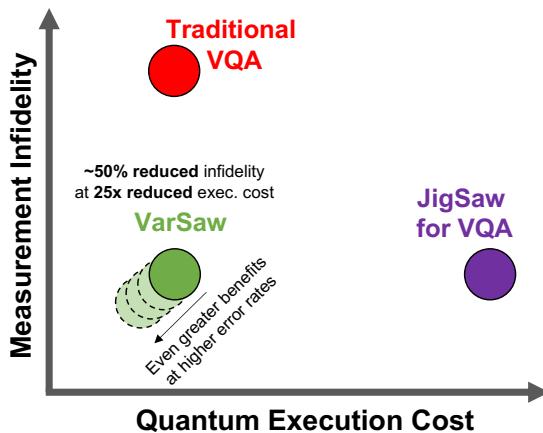


Figure 1: Traditional VQA is significantly impacted by measurement error. JigSaw combats measurement error but introduces high execution cost which is particularly harmful for VQA. VarSaw minimally achieves JigSaw’s measurement error mitigation at an execution cost similar to traditional VQA. When measurement error is very high, its benefits in terms of both fidelity and computational cost is even greater.

Measurement errors are often the most dominant source of error on current superconducting quantum computers, with average error rates ranging as high as 2-7% [58]. Furthermore, measurement operations suffer from measurement crosstalk [18], which means performing several measurement operations concurrently can substantially increase the error rate of each measurement operation by an order of magnitude [47]. Thus, larger programs are especially susceptible to measurement errors (crosstalk and otherwise) due to the higher number of measurement operations. This can be particularly detrimental to variational quantum algorithms which a) use low depth circuits and are thus often dominated by measurement errors, and b) can require very high accuracy for real-world applications such as molecular chemistry [44]. In this work, we improve the fidelity of VQAs by specifically targeting measurement errors.

Recent work, JigSaw [18], proposed a measurement error mitigation scheme for general quantum circuits. The key insight in this work is to divide the target circuit execution into two components. The first component is the idea of measurement subsetting (the ‘subset’ runs), wherein only a subset of the target circuit’s qubits are measured, for all such subsets that span the entire circuit. Measurement subsetting encounters fewer errors due to a) appropriately mapping the target logical qubits to be measured onto the physical qubits with highest measurement fidelity, and b) reduced measurement crosstalk, since fewer qubits are measured together. But, subsets alone are insufficient to construct the output of the target circuit, since they lack information about the global relations across all the qubits. The second component involves running the target circuit as is (the ‘global’ run), and then using its outcome to construct a high fidelity, relatively measurement-error-free output from the subset runs. To do so, a Bayesian Reconstruction algorithm

is employed, which enables a particular probability distribution (obtained from the subsets) to be appropriately modified using other information (obtained from the Global). Bayesian Reconstruction is inspired by Bayesian updating in statistics, whereby a prior probability estimate is updated using additional information [26].

While this prior work achieves substantial benefits, these benefits come at non-trivial computational cost. For example, VQE for a relatively small 6-qubit CH_4 molecule, when run with JigSaw does not converge to an accurate solution in over 100 hours of VQE simulation. This is because the additional subset runs scale polynomially (linear to quadratic) in the number of qubits in the circuit. While this cost might be tolerable for single-run circuits (the primary focus of their work), the costs are too substantial for variational algorithms, for three reasons: a VQA is an iterative task, requiring 1000s of iterations to navigate the noisy quantum device landscape to find the optimal solutions b for real world applications like molecular chemistry, the Hamiltonian is made up of 1000s of Pauli terms/strings, which result in a considerable number of circuits to be executed per VQA iteration; and c target applications (and thus the quantum circuits) can require 100s of qubits for any potential quantum advantage. Without the additional layer of measurement error mitigation, VQA is already computationally demanding. If used naively with the simplest subsetting, the additional measurement error mitigation layer adds yet another factor of computational requirements by running the same VQA instances multiple times. These costs are illustratively quantified and discussed further in Section 3 and Fig.8.

In this work, we seek to combat the effect of measurement errors in variational algorithms at a reasonable cost. To do so, we build application-aware optimizations to the JigSaw approach and propose **VarSaw: Application-tailored Measurement Error Mitigation for Variational Quantum Algorithms**.

VarSaw’s primary goal is to reduce computational costs by identifying two forms of redundancy: a *Spatial Redundancy in Subsets*: The generation of measurement Subsets for each Pauli string/circuit of the problem Hamiltonian, leads to significant redundancy (i.e., Subsets can repeat or commute), and b *Temporal Redundancy in Globals*: The incremental gradient based iterative approach used in VQA optimization leads to significant redundancy among the proximate ‘Global’ runs that JigSaw uses for Bayesian reconstruction (i.e., nearby Globals create nearly the same distributions). Both these forms of redundancy go unnoticed in the prior application-agnostic circuit-focused methodology.

Primarily, VarSaw reduces the quantum resources required to exploit measurement error mitigation by taking an end-to-end, application tailored, holistic approach. In doing so, it is also able to achieve VQA accuracy improvements. The same CH_4 example from earlier, when run with VarSaw, is able to converge to an accurate solution in under 10 hours of VQE simulation. An overview of VarSaw is provided in Fig.2 and its caption.

VarSaw key insights, contributions and results:

- (1) VarSaw improves JigSaw [18], specifically in the context of VQAs, by identifying *Spatial Redundancy in JigSaw Subsets* (across VQA Pauli strings) and *Temporal Redundancy in JigSaw Globals* (across VQA iterations).

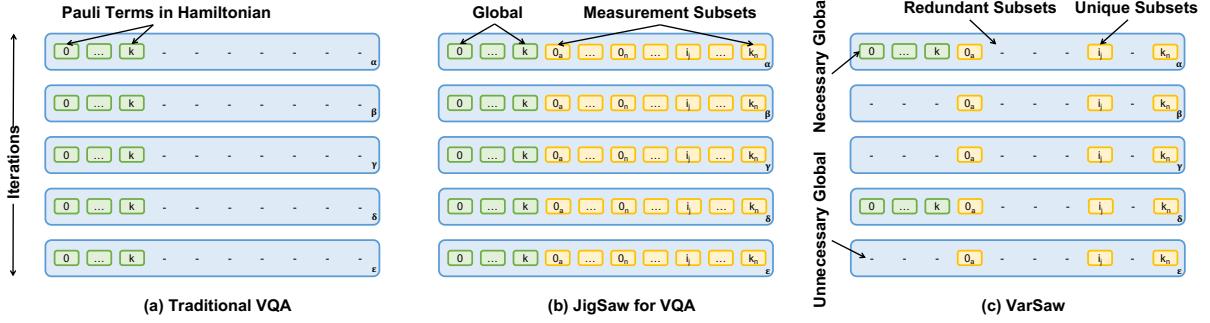


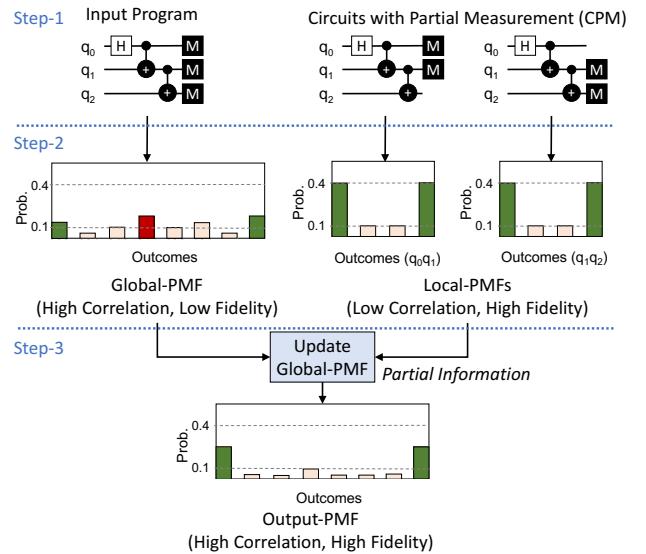
Figure 2: (a) Traditional VQA task run over multiple iterations. Each iteration executes multiple circuits corresponding to the different Pauli Strings in the problem Hamiltonian. (b) For every circuit in the original task, JigSaw runs multiple circuits - these are the Global circuits (same as original) and the Measurement Subsets (which scale linearly in the number of circuit qubits). These are run every iteration. (c) VarSaw optimizes JigSaw in a VQA-cognizant manner. It exploits spatial redundancy (repetitions and commutativity) to reduce the number of Subsets executed every iteration. Further, it exploits temporal redundancy (similarity in adjacent iterations' distributions) to reduce the number of iterations on which the Globals are executed.

- (2) To combat Spatial Redundancy, VarSaw proposes *Committing of Pauli String Subsets*: which takes a Hamiltonian-aware approach to commute and eliminate unnecessary subset circuits generated across all Pauli strings of the problem Hamiltonian.
- (3) To combat Temporal Redundancy, VarSaw proposes *Selective Global Executions*: which takes a VQA tuning-aware approach to only perform Global executions occasionally and increase/decrease their execution via a feedback-based approach.
- (4) In all, VarSaw reduces computational cost over naive JigSaw for VQA by 25x on average and up to 1000x (in terms of number of circuits executed) for the same target accuracy.
- (5) Further, it can recover, on average, 45% of the infidelity from measurement errors in the noisy VQA baseline.
- (6) Additionally, it improves fidelity by 55%, on average, over JigSaw for a fixed computational budget.
- (7) Importantly, this work showcases the overwhelming benefits from tailoring state-of-the-art optimizations in a domain-specific manner. Rejigging JigSaw from the ground up in a VQA-cognizant manner significantly benefits measurement error mitigation for VQAs, in terms of both computational cost as well as fidelity.

2 BACKGROUND

2.1 Noisy Intermediate Scale Quantum Computing

Today's quantum devices are noisy [43]. The sensitivity to external noise channels and imperfections in control and readout circuitry limit qubits from retaining their state indefinitely and causes the error rate of quantum operations to be very high. For example, the average error rate of a two-qubit gate is about 1% on existing hardware from IBM and Google, whereas that of measurement operations is about 4% [8], limiting the fidelity of quantum programs on NISQ devices. Consequently, software mitigation of hardware



errors on NISQ devices is an active area of research. Multiple forms of error mitigation strategies have been proposed to reduce the impact of different forms of errors. These include, but are not limited to, noise aware compilation [39, 56]; correcting measurement errors [14, 54]; crosstalk-aware scheduling [19, 40]; decoherence mitigation [11, 17, 27, 30, 42, 51], spin-echo correction [23], efficient gate scheduling [49]; etc. In addition, some of these can be used in conjunction to achieve better fidelity [45].

2.2 Measurement Errors

During computation, the quantum system exists as a linear superposition of many states. To obtain the final outcome, qubits

are measured to obtain a classical bitstring. Unfortunately, these operations cannot be implemented perfectly and are a dominant source of error on most existing quantum systems [8]. Program execution is unsuccessful and produces an incorrect outcome even if one of the measurements fails. The problem worsens with the program size because the probability of successfully measuring all the qubits simultaneously diminishes rapidly with the number of measurements. Thus, measurement errors constrain the size of the largest program (in terms of number of qubits) that can be run on NISQ devices.

Fast and accurate qubit measurements at scale remains an open problem. Measurement operations are imperfect due to several reasons. First, they require coordination between many sophisticated analog and digital instruments that operate across thermal domains (from 20 milli-Kelvins to room temperatures) [31] increasing susceptibility of these operations to thermal noise. Second, crosstalk arises from the unwanted couplings between the measurement apparatus when multiple qubits are simultaneously measured. For example, simultaneous qubit measurements are 1.26x times more likely to encounter errors on average compared to isolated measurements on Google Sycamore [8]. Last, these operations incur long latencies (about 800ns on Google devices and 4-5 microseconds on IBM's devices [7, 18]) during which qubits may lose their information due to decoherence. It is non-trivial to completely eradicate measurement errors at the device-level [31]. Measurement errors manifest as bit-flips and several recent studies specifically focus on minimizing their impact at the application-level [9, 13, 18, 25, 32, 55].

2.3 JigSaw for Measurement Error Mitigation

JigSaw [18] is a recent measurement error-mitigation technique whose overview is shown in Figure 3 and is comprised of three key steps. First, it takes a program as an input and prepares multiple subset circuits or 'Circuits with Partial Measurements' (CPM). Each CPM performs the exact same quantum operations as the input program but measures only a unique subset of qubits. For example, the two CPM in Figure 3 measures qubit subsets $[q_0, q_1]$ and $[q_1, q_2]$ respectively, for a 3-qubit GHZ program. Second, JigSaw executes the input program and the CPM on the NISQ device. The input program produces a Global-PMF (here PMF stands for Probability Mass Function) that captures the global correlation of the program over all the qubits but has lower fidelity. On the other hand, each CPM produces a partial or marginal Local-PMF over only the measured subset of qubits with high fidelity owing to fewer measurement operations. The third and final step of JigSaw leverages the partial information from the Local-PMFs to adjust the probabilities of each outcome in the Global-PMF using Bayesian reconstruction and produces a high-fidelity Output-PMF for the program. By default, JigSaw uses CPM that scales linear in the number of qubits so that high-fidelity Local-PMFs are obtained for each program qubit. In general, for a circuit with n qubits and a subset of size m , there exist $\binom{n}{m}$ CPMs.

As an example, for a 3 qubit circuit, if one of the subset circuits targeted the first two qubits, then the subset circuit produces a Local PMF corresponding to '00', '01', '10', '11'. The probability of each of these bitstrings is used to create the 3-qubit Output-PMF. For instance, the probability of '00' from the Local PMF is used to

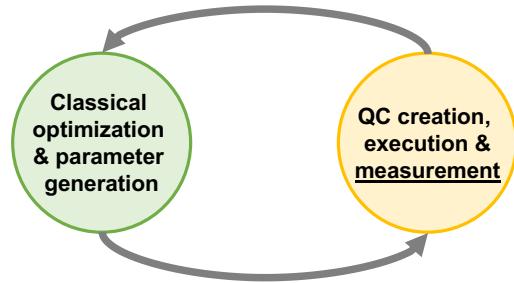


Figure 4: VQA: a hybrid algorithm that alternates between classical optimization and quantum execution.

create the probability of '000' and '001' in the Output PMF. This is done by weighing the probability of '00' in the Local PMF with the relative probabilities of '000' and '001' from the Global PMF. The subset circuit suffers from lower measurement errors since it can be mapped to the good qubits, plus it is more sparse and results in lower measurement crosstalk. On the other hand, Bayesian reconstruction is a statistical method and is hence imperfect and introduces some approximation errors.

A detailed depiction of the reconstruction is illustrated in Fig. 6 of the Jigsaw paper [18]. Jigsaw performs evaluation up to 18 qubits and shows that even for larger circuits, the benefits from avoiding measurement errors far outweigh errors from Bayesian reconstruction. Our own results are in agreement.

2.4 Variational Quantum Algorithms

Variational algorithms expect to have innate error resilience due to hybrid alternation with a noise-robust classical optimizer [36, 41]. An overview of this process is illustrated in Fig. 4. There are multiple applications in the VQA domain such as the Quantum Approximate Optimization Algorithm (QAOA) [20] and the Variational Quantum Eigensolver (VQE) [41]. Our applications in this work target VQE, but VarSaw is applicable to all VQA problems - more on this in Section 7.3. VQE in itself has wide applications in molecular chemistry, condensed matter physics, quantum Ising optimization problems, and a variety of quantum mechanics many-body problems, etc. An important application of VQE is the ground state energy estimation of a molecule, a task that is exponentially difficult in general for a classical computer [22]. Estimating the molecular ground state has important applications in chemistry, such as determining reaction rates and molecular geometry. The quantum circuit used in each iteration of VQE (and VQA in general) is termed an ansatz which describes the range of valid physical systems that can be explored and thus determines the optimization surface. Traditionally, the ansatz is parameterized by 1-qubit rotation gates. A good ansatz provides a balance between a simple representation, efficient use of available native hardware gates, and sufficient sensitivity with the input parameters. The classical tuner/optimizer [5, 34] variationally updates (often via a gradient based approach) the parameterized circuit until the measured objective converges to a minimum. An example 3-qubit ansatz (i.e., the parameterized circuit which is tuned in VQA) is shown in Fig. 5 - measurements have to be performed

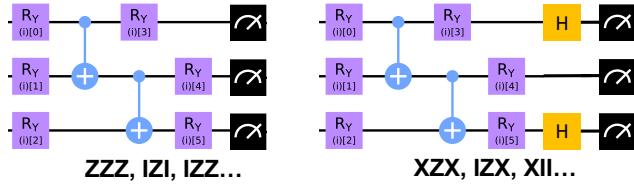


Figure 5: Ansatz circuit measured on different Pauli bases (left: ‘ZZZ’, right: ‘XZX’). Qubit commutativity allows for a set of Pauli strings to be measured on a single basis.

Workload	Ref. Energy	Noisy VQE	VQE+JigSaw (subset size 2)
LiH	1.72	6.11	2.73
H ₂ O	-109.86	-97.77	-107.71
H ₂	10.46	17.00	12.51
CH ₄	-28.55	-21.22	-26.63

Table 1: For a given circuit, JigSaw is effective in mitigating measurement errors, improving the circuit’s energy estimate.

on this ansatz every iteration, which is key to the motivation for this work (more on this later).

Estimating the VQE global optimum with high accuracy has proven challenging in the NISQ era even with sophisticated optimizers, a well-chosen ansatz, and error mitigation [9, 12, 16, 45, 46, 53, 57, 59]. Considering the significant disparity between NISQ VQA accuracy and real world requirements, it is imperative to further aid VQA to the best extent possible.

2.5 JigSaw for a VQA circuit

As in the case of other circuits, JigSaw successfully mitigates measurement errors at the circuit-level for VQAs. In Table 1, we show 4 VQE instances from molecular chemistry. In these instances, the VQE ansatz is parameterized with optimal parameters (known from ideal simulation) and executed in noisy simulation. Clearly, the noisy VQE instance performs considerably worse than the ideal reference simulation. We employ JigSaw on top of this noisy VQE instance and show that JigSaw successfully manages to mitigate measurement errors well, recovering more than 70% of the optimal energy. This clearly shows that measurement error plays a significant role in VQA and JigSaw, at a circuit level, is able to achieve considerable benefits. However, the circuit cost for performing JigSaw is high and the circuit-level benefits can be lost at the application-level, as is discussed next.

3 VARSAW: MOTIVATION AND PROPOSAL

3.1 Hamiltonian, Paulis, Measurement Commutation

The VQA problem is represented as a Hamiltonian and is a linear combination of multiple Pauli terms. The lowest eigenvalue of the Hamiltonian corresponds to the system’s ground state energy [36].

- (1) $\mathbf{H}_{\text{Base}} = ZZIZ + ZIZX + ZZII + IIIZ + ZXXZ + XZIZ + ZXIZ + IXZZ + XIZZ + XXIX$ [10 terms]
- (2) $\mathbf{C}_{\text{Comm}} = ZZIZ + ZIZX + ZXXZ + XZIZ + IXZZ + XIZZ + XXIX$ [7 circuits]
- (3) $\mathbf{C}_{\text{JigSaw}} = ZZ-- + -ZI- + --IZ + ZI-- + -IZ- + --ZX + ZX-- + -XX- + --XZ + XZ-- + -ZI- + --IZ + IX-- + -XZ- + --ZZ + XI-- + -IZ- + --ZZ + XX-- + -XI- + --IX$ [21 circuits]
- (4) $\mathbf{C}_{\text{VarSaw}} = ZZ-- + --ZX + ZX-- + -XX- + -XZ + XZ-- + -XZ- + --ZZ + XX--$ [9 circuits]

Figure 6: (1) 4-qubit VQA Hamiltonian with 10 Pauli terms. (2) Qubit commutativity (red terms in 1) allows for all 10 terms in the Hamiltonian to be measured with 7 circuits/terms. (3) JigSaw uses a 2-qubit sliding window to create measurement Subsets, thus resulting in (4-1)*7=21 circuits. (4) VarSaw identifies redundant / commuting terms in the Jigsaw Subsets (red terms in 3), to reduce the required executions to only 9 circuits.

Every iteration, the VQA objective function calculates the expectation value of this Hamiltonian. This objective function is derived from ansatz measurements over different bases.

An example VQE Hamiltonian with 10 terms is shown in Eq.1 in Fig.6. This Hamiltonian represents a 4-qubit system, so each term is 4-wide. While only 10 terms are shown for this trivial example, this number grows quickly for larger Hamiltonians (targeting more complex molecules, for instance). For example, the Cr₂ molecule, which is considered a benchmark for quantum advantage in molecular chemistry minimally requires around 50 qubits and has nearly 50,000 Pauli terms [44]. In general, the number of Pauli strings scales as $O(N^4)$ [21]. Different Pauli strings are measured by measuring the ansatz in different bases - this is shown in Fig.5 for a 3-qubit system. Different bases correspond to adding appropriate gates at the end of the ansatz. Not all Pauli strings require different circuits - this is discussed next.

Prior work has shown that the number of circuits executed can be smaller than the total number of Hamiltonian Pauli strings, because strings which correspond to commuting observables can be measured with a single circuit. The simplest form of this commutation based reduction is shown in Eq.2 in Fig.6 - here the number of terms to be measured reduces from 10 to 7 in Eq.1 since the terms in red trivially commute with one or more terms in black. A more detailed example of commutation is shown in Fig.7. Here, different Pauli strings (only those made up of ‘I’, ‘X’ and ‘Z’) for a 3-qubit system are shown and the arrows indicate commuting relations. ‘III’ (blue) commutes with all of the other Paulis (blue arrows) and therefore can use all other Paulis as its commuting parent. Similarly, ‘IIZ’ has 8 other commuting Pauli parents (green), ‘IZZ’ has 2

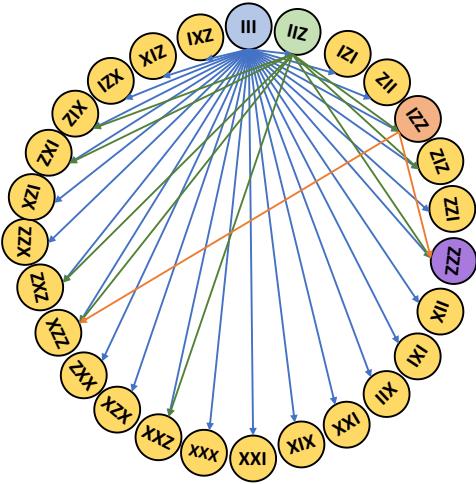


Figure 7: Qubit commutativity graph for 27 3-qubit Pauli strings ($X'/Z'/T'$). Arrows are drawn from Paulis that can be commutatively measured by others, to the Paulis that can measure them. Arrows are shown for ‘III’ (26), ‘IIZ’ (8) and ‘IZZ’ (2) and ‘ZZZ’ (0).

(orange) and ‘ZZZ’ has none. If the Hamiltonian has the terms ‘IZZ’ and ‘ZZZ’, then a circuit measured in the all-Z basis is sufficient for both of these terms. Extending from the above, in Fig.5 the circuit to the left is used to measure commuting Paulis such as ‘ZZZ’, ‘IZI’, ‘IZZ’ etc. Other circuits or Paulis can be inferred similarly.

More sophisticated forms of commutation are possible, which can further reduce the number of terms [21, 24, 48], but these forms can: a) non-trivially increase circuit depth and b) can suffer exponential cost to construct. While VarSaw benefits will potentially be applicable on top of these as well, we limit our scope to the trivial qubit commutations as shown in the examples.

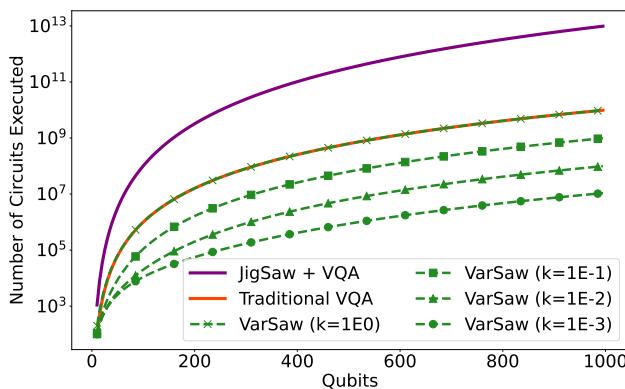


Figure 8: JigSaw incurs high circuit cost for VQA, with executed circuits every iteration roughly scaling as $O(Q^5)$ in the number of qubits, whereas the baseline scales as $O(Q^4)$. VarSaw scales as $O(Q^2)$ to $O(Q^4)$ while also mitigating errors equal to or better than JigSaw.

3.2 Spatial Redundancy in JigSaw Subsets

Eq.2 in Fig.6 shows that there are 7 Pauli strings post-commutation in C_{Comm} . JigSaw generates Subsets for each of these strings. The optimal version of JigSaw constructs 2-qubit measurement Subsets using a sliding window, thus 3 subsets are created for each Pauli string, resulting in a total of 21 terms. This is shown in Eq.3 as C_{JigSaw} . Now note that there is substantial redundancy among these terms due to repetitions and commutations - these are shown in red. For example, ‘-ZI-’, which requires only the 2nd qubit from the left to be measured, commutes with ‘ZZ-’. Other terms can be similarly inferred. By taking a Hamiltonian-level approach to parsing through the Subsets, VarSaw is able to identify these redundancies, and prunes the number of terms required to only 9. This is shown in Eq.4 as C_{VarSaw} .

While the VarSaw term reduction is around 2.3x and the number is equal to C_{Comm} in this example, more terms or qubits in the Hamiltonian will lead to even greater redundancy, and thus even greater reduction benefits with VarSaw. This is intuitive from Fig.7. Terms with more ‘T’s have a larger commuting family (i.e., more arrows), thus they are more likely to be combined with other terms. More generally, the number of terms H_{Base} scale as $O(Q^4)$ [21] where Q is the number of qubits (since each qubit can be ‘T’ / ‘Z’ / ‘X’ / ‘Y’), whereas the number of terms in C_{VarSaw} scales as $O(Q^{1-2})$ (Sliding window or $\binom{Q}{2}$).

This scaling benefit is illustrated in Fig.8. The graph shows increasing number of circuits executed (on a log scale) against the number of qubits in the problem Hamiltonian *per iteration* of VQA. The number of circuits for traditional VQA scales roughly as $O(P)$, with roughly $P = 0.01 * Q^4$, whereas for JigSaw it scales as $O(P + P * Q^{1-2})$, where the first term is the number of global executions and the second is the subsets, with P as the number of Pauli terms in the Hamiltonian. On the other hand, the number of terms for VarSaw scales as $O(k * P + Q^{1-2})$, where the first term is the number of global executions and the second is the subsets. While ‘k’ (0 to 1) and the number of global executions is discussed next, it is evident that VarSaw’s scaling is at worst similar to traditional VQA (the line with $k=1$ overlaps Traditional VQA). Importantly, VarSaw is at least $O(Q)$ lower quantum computational cost than JigSaw, with at least the same measurement error mitigation benefit.

3.3 Temporal Redundancy in JigSaw Globals

Next, we discuss temporal redundancy in JigSaw Globals. Global executions are performed so as to generate correlations among the different qubits of the target circuit, and these correlations are then used, along with the measurement subsets, to construct the measurement error mitigated probability distribution (via Bayesian Reconstruction). There are three characteristics of interest associated with the Globals. First, they suffer from measurement error, since the entire circuit is being executed. Second, the Bayesian reconstruction itself is associated with some statistical errors. Third, the Globals are a function of the ansatz parameters of the particular iteration.

Our hypothesis for temporal redundancy stems from the following: On the one hand, generating new Globals every iteration will cause lower error due to: a) faithfully following the ansatz parameters, and b) no propagation of Bayesian Reconstruction error

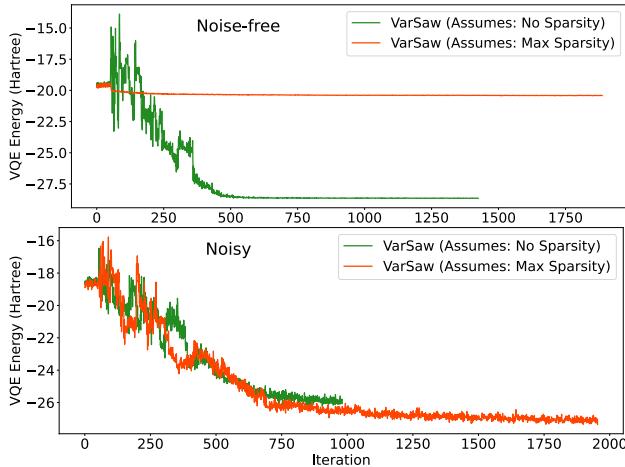


Figure 9: In the absence of noise, increasing VarSaw temporal sparsity in Globals is detrimental to VQA. But in practical settings with considerable noise, temporal sparsity can achieve fidelity similar or better than the alternative, while also achieving considerable reduction in execution cost.

across iterations. On the other hand, generating new Globals every iteration will cause higher error due to even more measurement errors influencing the correlation and reconstruction. Thus, as long as the effect of ansatz parameter variation (among proximate iterations) and Bayesian Reconstruction error is minimal compared to the additional measurement errors, it is redundant to perform Global executions every iteration (and potentially beneficial to not do so). Instead, they can be performed in a more sparse manner, thus further saving computational cost and, in fact, also improving VQA fidelity.

To analyze the above hypothesis, we construct experiments to compare two extreme scenarios of VarSaw - one performing Globals every iteration ('No-Sparsity'), and the other performing Globals only once, right at the beginning of VQA ('Max-Sparsity'). Both are run for a fixed wall-clock tuning time. These experiments are performed on a 6-qubit CH_4 molecule Hamiltonian with two versions studied in simulation, and are shown in Fig.9. The first version is a noise-free experiment with no measurement errors (top). The second version employs noise mimicking IBM's quantum machine IBMQ Mumbai (bottom). More details of the experimental methodology is discussed later in Section 5.

In the noise-free experiment, Varsaw with Max-Sparsity performs very poorly compared to the No-Sparsity. In the absence of measurement error and any form of significant perturbation, the overwhelming influence of the fixed one-time global execution causes the Max-Sparsity scenario to quickly get stuck at a local minima, whereas the No-Sparsity scenario is faithful to the true distribution and can reach the ideal value. In the noisy experiment, though, it can be observed that the Max-Sparsity scenario is, in fact, able to marginally outperform the No-Sparsity scenario. As discussed above, the lower impact of parameter variation and Reconstruction error compared to measurement error, is clearly evident. Interestingly, the presence of random perturbation, allows

the No-Sparsity scenario to jump out of any early local-minima despite the heavy influence of the fixed single Global execution. Note that it is not always the case that the Max-Sparsity scenario will outperform (fidelity-wise) the No-Sparsity scenario, but clearly this is sufficient empirical evidence for the potential similarity in their behaviors. Also, it is observable in both experiments that the Max-Sparsity scenario is able to perform more iterations than No-Sparsity scenario (for the same time / computational budget). This is intuitive, since Max-Sparsity saves on computational resource / time, since it does not run Globals every iteration. The iteration ratio difference between the top and bottom experiments occurs due to the non-determinism of the tuner, apart from other external variations.

Perhaps non-intuitively, high sparsity can enable computational cost reduction even over the baseline. This is because the number of circuits in each global execution or traditional VQA iteration scales as $O(Q^4)$, whereas the number of subset circuits scales as $O(Q^{1-2})$. The reduced computational cost for different sparsity ratios, as a function of number of qubits, is shown in Fig.8. Clearly the number of circuits executed is lowered even compared to Traditional VQA as the sparsity of Global executions increases.

It is clear that exploiting temporal redundancy is highly beneficial. In terms of computational cost, VarSaw clearly benefits over prior work and also potentially over the baseline. Further, in terms of fidelity, it can potentially achieve even higher fidelity (lower VQA expectation) in comparison to the non-sparse scenarios and over the baseline.

4 VARSAW DESIGN

In this Section, we discuss the two design components of VarSaw to exploit and eliminate the spatial and temporal redundancy produced by JigSaw when targeting VQAs.

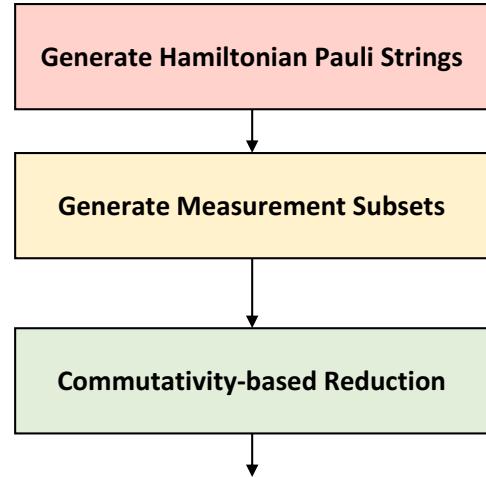


Figure 10: VarSaw performs subsetting across all Paulis and aggregates them *before* commutativity based reduction.

4.1 Commuting of Pauli String Subsets

The high spatial redundancy in JigSaw’s Pauli subsets that repeat and/or commute with each other was motivated in Section 3.2. The reason for this is that JigSaw is agnostic to the application and is focused on measurement-subsetting only at the circuit level. VarSaw eliminates this redundancy by integrating measurement-subsetting into the end-end VQA framework. It performs measurement subsetting immediately after the generation of Hamiltonian Pauli strings and before commutativity-based reduction. Once the subsets are generated, commuting terms can then be identified and eliminated, with higher redundancy and thus higher benefits, as was described in Section 3.2. This is shown to the right of Fig.10.

The commutativity based reduction is performed similar to prior work. For example, [28, 50] reduce commuting terms on an ad hoc basis, via inspection of the Hamiltonian terms. More systematic methods are used in OpenFermion [37] and Rigetti PyQuil [15] libraries.

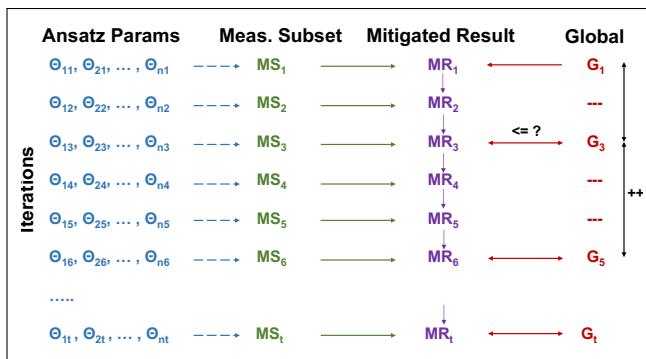


Figure 11: Design to exploit temporal redundancy by selectively executing the Globals. Measurement subsets are generated every iteration for every new instance of ansatz parameters, whereas the Globals are done only every k iterations where k is dynamically optimized through comparisons against the mitigated results which are derived explicitly from the current iteration’s Subsets and implicitly from the most recent Global.

4.2 Selective Execution of Globals

Next, we discuss the VarSaw design to tackle temporal redundancy among the Global executions which, by default, are performed every iteration in JigSaw. The redundancy exists because, in the presence of non-insignificant noise, the Global execution's influence on the Bayesian reconstruction is less effected by fine granularity ansatz parameter changes and any propagated Bayesian error. As an additional benefit, by not executing new Globals, there are no new measurement errors introduced. The benefits of reduced measurement errors often overrides the potential loss incurred from eliminating certain Global executions. Thus, it is sufficient to perform Global Executions sparsely, but how sparse would be dependent on the VQA gradients and the machine noise.

VarSaw designs a simple dynamic optimization scheme to increase or decrease the sparsity of Global Executions depending on

Molecule	Qubits	Pauli terms	Temporal?
H_2	4	15	Y
LiH	6	118	Y
LiH	8	193	Y
H_2O	6	62	Y
H_2O	8	193	Y
H_2O	12	670	N
CH_4	6	94	Y
CH_4	8	241	Y
H_6	10	919	N
BeH_2	12	670	N
N_2	12	660	N
C_2H_4	20	10510	N
Cr_2	34	32699	N

Table 2: Molecules for VQE, for VarSaw evaluation, with Qubits and Pauli terms in the Hamiltonian. ‘Temporal?’ indicates if evaluation of Temporal Redundancy is performed.

its actual impact. At the beginning, the Globals are set to execute on every k^{th} VQA iteration i.e., these are actual full executions and not just the measurement subsets. On every non k^{th} iteration, the Globals are not executed, only the Subsets are run. The Mitigated Result of this iteration is then generated from the previous iteration's Mitigated Result and the current iteration's Subsets. The measurement Mitigated Result for the k^{th} iteration is then computed/verified with the Global and the Subsets – i.e., we compute the result in both possible ways (a) Executing both Globals and Subsets (b) Executing only the Subsets and using the previous iteration's Mitigated Result. If the value of the result in case (a) is less than that in case (b), then the value of k is decreased. If not, it is increased. This is a simple hill-climbing method, in which the granularity of Globals is increased/decreased by verifying the quality of the mitigated result.

This is illustrated in Fig.11. In the figure, each row represents a VQA iteration, with specific ansatz parameters chosen by the tuner. The measurement subsets are designed and executed, as discussed in Section 4.1. On iteration 1, a Global execution is performed, and the initial sparsity window is 2 cycles (i.e., the Global execution is set to perform every alternate cycle). The mitigated result MR_1 is then generated from MS_1 and G_1 . On iteration 2, there is no G_2 , only MS_2 . Instead MR_2 is generated from MS_2 and MR_1 . On iteration 3, MR_3 is still generated from MR_2 and MS_3 . But G_3 is also run and the result obtained from G_3 and MS_3 is compared against MR_3 . If MR_3 has similar or lower energy than the result computed using G_3 and MS_3 , that means that the error from fine-granularity ansatz parameter change and propagated Bayesian error is no worse than additional measurement errors. Thus, MR_3 is kept as is, and Global executions are made more sparse. In the figure, the next Global execution is set to the iteration 5 - doubling the sparsity. If MR_i is greater than the result obtained using G_i and MS_i , then MR_i is instead regenerated with G_i and the sparsity is lowered.

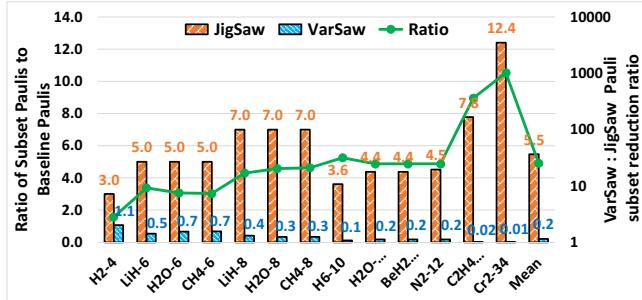


Figure 12: Pauli term reduction in Measurement Subsets achieved by VarSaw compared to JigSaw. Orange columns / left axis: VarSaw and JigSaw subsets expressed relative to the number of Pauli terms in the baseline. Green line / right secondary axis: reduction in subsets for VarSaw rel. to JigSaw.

5 METHODOLOGY

5.1 Applications

We limit ourselves to one VQA domain, the VQE, which was introduced in Section 2.4. Our evaluations encompass 9 different molecules, the details for which are provided in Table 2. Some molecules are run with more than a single molecular configuration. Systems with more molecular orbitals frozen require less qubits. The ideal noiseless ground state energy of a molecule is the same across different specifications.

Note that only a subset of the applications are used for full VarSaw evaluation of temporal+spatial redundancy elimination. This is because the larger molecules/configurations have too many terms/qubits to effectively simulate VQA (especially since we need to run 100s of VQA iterations). For these, we only show the spatial benefits of VarSaw. Applications where we demonstrate spatial+temporal redundancy elimination are restricted to systems of up to 8 qubits.

We use the hardware efficient SU2 [3] ansatz which are of low depth and therefore suited to NISQ devices. The ansatz is constructed for the ‘full’ entanglement, which means that entanglement (i.e. CX gates) is allowed between all qubit pairs. 2 blocks of repetition are used in the ansatz, where each block is an additional layer of CX gates and tunable 1-qubit gate parameters. We use the SPSA [5] and ImFil [34] classical tuners across all our evaluations. We find a subset size of two qubits to be optimal to our work, fairly in agreement with JigSaw (evaluation in Appendix.A)

5.2 Infrastructure and Optimization Process

We integrate VarSaw as part of the Qiskit [6] framework. The tackling of spatial redundancy is implemented as a pre-processing step before actual VQA execution - this was illustrated in Fig.10. Qiskit interfaces with the PySCF library [52] in the process of constructing Hamiltonians from molecular specifications. JigSaw is then performed after Hamiltonian generation. This is followed by the commutativity step to reduce the number of terms.

Varsaw’s tackling of temporal redundancy of Jigsaw Globals is an on-the-fly optimization that can be integrated with any VQA classical optimizer. We implement it in Python and use it with the

Qiskit VQA framework, through which it interacts with quantum execution.

Due to the large number of VQA Hamiltonian Pauli terms, as well as the number of VQA iterations, we primarily evaluate VarSaw on noisy simulation modeled on IBM Quantum Devices (except in Section 6.5 which shows real machine analysis). We use the noise model of IBMQ Mumbai (27 qubits) – machine details can be found on the IBM Quantum Systems page [4]. While Qiskit Runtime [2] offers support to run iterative applications on the actual quantum hardware, it is very limited in its support for on-the-fly optimizations, hence unsuited to evaluate VarSaw. Running one iteration at a time on the quantum hardware without Runtime (i.e., using our own classical computer between every quantum execution) is possible, but impractical for most long running applications since machine availability is very sparse and wait-times between iterations are enormous (hours to days). Where applicable, results are averaged over up to 10 different trials which run the VQA optimizer with different random seeds.

5.3 Evaluation Comparisons and Metrics

Comparisons:

Baseline: Traditional VQA without measurement error mitigation, but using Pauli string commutation.

JigSaw: JigSaw applied on top of the Baseline, with each VQA iteration employing 2-qubit measurement Subsets using a sliding window, along with Global executions.

VarSaw: The proposed approach which optimizes JigSaw (for VQA) for spatial and temporal redundancy.

Metrics:

Cost: Quantum computational cost, in terms of the number of quantum circuits that have to be executed on the quantum device (Lower is better).

Energy: Molecular VQE ground state energy estimates using the standard Hartree Energy metric (Lower is better). This is the equivalent of fidelity for VQA.

6 EVALUATION

6.1 Computational cost reduction in Subsets

Fig.12 shows the Pauli term reduction in Measurement Subsets achieved by VarSaw in comparison to JigSaw. Both VarSaw and JigSaw subsets are expressed relative to the number of Pauli terms in the baseline in the orange columns / left axis. The number of Pauli terms/circuits in JigSaw increases loosely in proportion to the number of qubits. The worst case increase would be $(Q-1)*P$ where Q is the number of qubits and P is the number of baseline Pauli strings. The number is often lower than this because measurement subsets which have are all T’s are already weeded out i.e, those terms will require no measurements. But clearly, the large increase in the number of terms is evident. Our largest molecule, a 34-qubit Cr_2 molecule, sees a 12x increase in the number of terms, while an average increase of 5.5x is observed for molecular systems ranging from 4-qubits to 34-qubits.

On the other hand, the Pauli term Measurement Subsets in VarSaw is substantially lower due to the elimination of redundant repetitions and commuting terms. On average, the number of subsets is

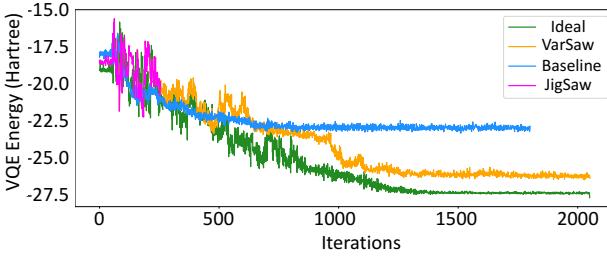


Figure 13: VQE accuracy improvements by VarSaw compared to the noisy baseline and noise-free ideal for CH_4

only 0.2x of the total Pauli terms. Thus, the maximum per-iteration computational cost increase for VarSaw is, on average, only 20% greater than the baseline, while achieving the benefits of measurement error mitigation. Note that the cost-per-iteration incurred by VarSaw is usually much less than the baseline since for most iterations, the Globals are not executed at all. This is discussed later in Section 6.3.

Also notable is that the number of Subsets, relative to the baseline Pauli terms, decreases for larger molecules. This is because, as the number of baseline Pauli terms increases, there is a considerably larger fraction of the Subsets which repeat or commute. This is especially promising for the scalability of VarSaw’s benefits. For example, Cr_2 only requires a 1% increase in the worst-case computational cost.

On the green line graph / right secondary axis, we show the reduction in measurement subsets for VarSaw compared to JigSaw. Note that the axis is on a log scale. Term reduction is greater than 1000x for the largest molecule, while the average reduction is around 25x. This clearly highlights the benefits of VarSaw term reduction compared to JigSaw, and in general is indicative of the potential of application-aware improvements to state-of-the-art quantum optimizations.

Overall, VarSaw achieves a 25x lower runtime on average compared to JigSaw, and as much as 1000x lower runtime for our largest problems. Further, VarSaw has a 10x lower runtime on average compared to the baseline.

6.2 Single-application CH_4 analysis

Fig.13 shows the VQE energy plot for VarSaw, compared to JigSaw, the noisy baseline, and the noise-free ideal, for the CH_4 molecule. These results are shown for a fixed circuit budget, i.e., all 4 scenarios run the same number of circuits. VarSaw is able to achieve the highest accuracy of VQE energy estimates, comparable to the Ideal. The baseline is affected by measurement errors which is expected. It also runs fewer iterations than VarSaw. This is because VarSaw has a lower cost per iteration due to minimal number of global iterations. This is discussed further in Section 6.3. On the other hand, while JigSaw also performs measurement error mitigation, it is only able to run a fraction of the iterations due to its high circuit cost per iteration. Thus, it achieves a lower accuracy than the baseline in this fixed budget scenario. VarSaw to JigSaw accuracy comparison, under a fixed circuit budget, is discussed further in Section 6.4.

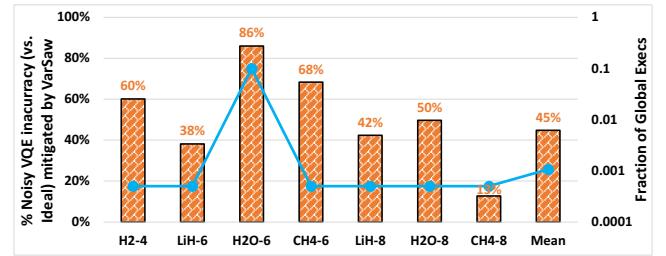


Figure 14: VQE accuracy improvements by VarSaw compared to the baseline over 2000 iterations. Secondary axis shows the optimal fraction of global executions required for VarSaw.

While other forms of error continue to persist (and is orthogonal to this work), it is evident that the subsetting technique is able to eliminate a considerable fraction of the measurement error, leading to significant accuracy improvements. Further, VarSaw is able to do so at low computational cost, meaning that it is able to optimize further, to better accuracy, under a fixed computational budget.

6.3 VQE Accuracy Improvements through Error Mitigation and Global execution fraction

Next, we show the improvements in VQE accuracy achieved by VarSaw compared to the baseline over 2000 iterations of VQA tuning in Fig.14 (orange columns / left axis). The improvements are achieved thanks to the measurement error mitigation adopted from the JigSaw approach. VQE accuracy improvement ranges from 13% to 86% with a mean improvement of 45%. Note that the results shown here are only for the smaller benchmarks (up to 8 qubits) since the larger ones are not amenable to noisy simulation. Clearly, VarSaw achieves considerable improvement to VQE accuracy while only resulting in minimal increase to the worst case per-iteration computational cost.

The blue line / right axis shows the optimal fraction of global executions required for VarSaw. Clearly the fraction is very low, meaning that performing only a few Globals is most beneficial. This is so for two reasons. First, performing fewer Globals means that new measurement errors are not being introduced into the tuning process. Second, fewer Global executions means that the computational resources (i.e., quantum circuits) can be utilized towards running more subset iterations - allowing for further VQA optimizations, albeit with imperfect correlation. VarSaw performs optimally at a Global granularity of around 1 in 100 iterations, which means that the total number of circuits executed, in comparison to the baseline, is also more than 10x lower. This can, of course, change with different noise characteristics, problem complexity etc, but VarSaw will always be computationally competitive with the baseline while achieving significant measurement error mitigation benefits. We evaluate VarSaw’s benefits over a noise sweep in Appendix.B.

6.4 VQE Benefits over JigSaw for fixed circuit budget

While VarSaw employs the same error mitigation as JigSaw, it is much more computationally efficient. Thus, under a fixed circuit execution budget, the VQE accuracy improvements achieved by

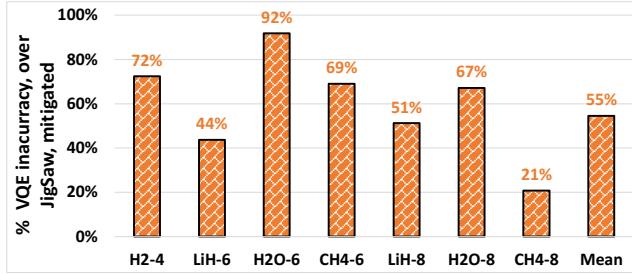


Figure 15: VQE Accuracy Improvements over JigSaw for fixed circuit budget

VarSaw is considerably greater than that of JigSaw. This is because VarSaw can execute orders of magnitude greater number of VQA iterations for the same circuit budget. This is important both in terms of quantum device availability in the cloud as well as monetary cost which scales proportionally to the circuits executed. Fig.15 shows the improvements in VQE accuracy achieved by VarSaw compared to JigSaw for a fixed circuit budget that allows VarSaw to complete nearly 2000 iterations, but JigSaw can only complete a few 100s. Thus, VQE accuracy improvement for VarSaw over JigSaw ranges from 21% to 92% with a mean improvement of 55%. These benefits will increase as we scale to larger molecules since the computational cost difference between JigSaw and VarSaw increases with system size as discussed earlier.

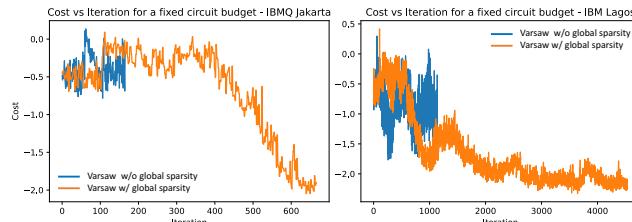


Figure 16: VarSaw on real IBM Q devices - Lagos and Jakarta

6.5 Real device experiments

The experiments so far have focused on noisy simulated quantum devices. In this section, we discuss smaller VQA experiments run on real quantum devices (IBM's Lagos and Jakarta) to highlight the benefits of VarSaw's temporal optimization by selectively executing the Globals. This is shown in Fig.16. To keep circuit overheads to a minimum (for the runs to be amenable to real device execution), we run VQE on a Transverse Field Ising Model (TFIM) Hamiltonian with 5 qubits and only 3 Pauli terms.

It is clearly evident that the VarSaw temporal optimization that employs the Global selective execution / sparsity, considerably reduces the circuit overhead per VQA iteration while still being faithful to the overall Hamiltonian objective. Therefore, sparse VarSaw is able to complete a significantly greater number of meaningful VQA iterations (nearly 4x) compared to VarSaw with full Global execution, and achieves 1.5x-3x VQA objective improvements on real devices. These benefits are expected to be greater for more

Workload	Full	Linear	Circular	Asymmetric
CH_4	44.42	65.87	23.26	31.24
H_2O	32.09	96.49	63.86	54.51
LiH	26.90	85.09	37.86	65.64

Table 3: % VQE Inaccuracy Mitigated by VarSaw with Global Selective Execution, over VarSaw without Global Selective Execution, for Different Ansatz Types

Workload	$p = 1$	$p = 2$	$p = 4$	$p = 8$
CH_4	29.98	26.73	21.17	5.29
H_2O	-1.46	31.83	27.88	7.97
LiH	58.67	53.59	26.04	7.79

Table 4: % VQE Inaccuracy Mitigated by VarSaw with Global Selective Execution, over VarSaw without Global Selective Execution, for Different Ansatz Depths

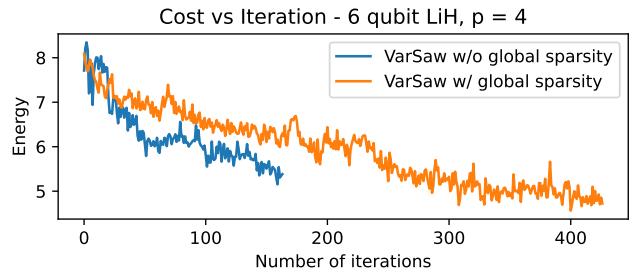


Figure 17: VarSaw w/ and w/o global sparsity for $p = 4$

complex VQA problems which will have even greater impact of measurement fidelity and will thus gain from the VarSaw method.

Finally, note that for the VarSaw spatial optimization, the benefit is orthogonal to running on a real device since the primary benefit is circuit overhead reduction which is the same for the real device or simulation.

6.6 VarSaw benefits on different ansatz structures

Next, we investigate the behavior of VarSaw with Global Selective Execution (i.e., sparsity) against VarSaw without Global Selective Execution for a variety of ansatz types and ansatz depths. We perform our simulation based analysis on 6-qubit CH_4 , 6-qubit H_2O , and 6-qubit LiH . The results of our experiments are summarized in Tables 3 and 4. Table 3 shows that VarSaw with Selective Execution outperforms VarSaw without Selective Execution for all molecules and ansatz types. For varying depths (Table 4), we see that Global sparsity helps in all cases but one, where it performs marginally worse compared to the no-sparsity case.

We also observe some interesting instances where the rate of convergence of VarSaw with selective execution is poorer than the rate of convergence of VarSaw without selective execution. However, selective execution reduces the per-iteration overhead to such a great extent that the program ends up performing many

more meaningful iterations compared to VarSaw with no sparsity, and converges to a lower energy value. A typical plot is shown in Fig. 17. This is expected since a deeper circuit has a greater number of parameters, and the effect of computing the global distribution using stale parameters is more pronounced - the VarSaw temporal optimization benefit is still seen to have considerable benefit.

6.7 Isolated effect of each VarSaw optimization

VarSaw spatial optimization vs Baseline, for circuit overhead reduction: VarSaw's spatial optimization as a circuit reduction scheme does not have any direct impact on the baseline since these subset circuits are not a part of the baseline.

VarSaw spatial optimization vs JigSaw, for circuit overhead reduction: Fig.12 showed the impact of VarSaw's spatial optimization in terms of subset circuit reduction over JigSaw. If we add the constant global component to both (assuming no temporal optimization), then VarSaw is 5x better on average and 12x better for the largest systems.

VarSaw temporal optimization vs Baseline, for circuit reduction: Fig.14 (secondary axis) shows the benefit of temporal optimization in terms of reduced circuits compared to the baseline. VarSaw only has to execute the Globals around once every 100 iterations. Of course, VarSaw executes the local circuits instead. If we do not use the spatial optimization, then VarSaw would still be much worse than the baseline because the number of local circuits is roughly 6x the baseline circuits. But if we did use temporal optimization as well, then VarSaw is 10x lower circuits than the baseline.

VarSaw temporal optimization vs JigSaw, for circuit reduction: Similar to above, referring to Fig.14, VarSaw would execute global circuits only roughly 1% of the time that JigSaw does. But if spatial optimization is not performed, this benefit is less useful. If spatial optimization is also performed, then VarSaw is 25x lower circuits than JigSaw on average and more than 1000x lower in the largest systems.

VarSaw vs JigSaw, for VQA accuracy: If there was no runtime or circuit budget, the VarSaw optimizations do not have a direct impact on accuracy over JigSaw because VarSaw is focused on reducing circuit overhead compared to JigSaw while employing the same measurement error mitigation. If there was a fixed budget, then VarSaw's spatial optimization would achieve better accuracy than JigSaw. With both spatial and temporal optimization, this accuracy improvement is 55% relative to JigSaw's improvements. If we only had temporal optimization, the benefit would be negligible (as explained earlier, temporal optimization is only beneficial post spatial optimization). The accuracy benefit for spatial optimization alone is around 35%.

VarSaw vs baseline, for VQA accuracy: Overall we see a 45% absolute improvement over the baseline. Without a circuit/runtime budget, this accuracy improvement is independent of the spatial optimization and is unaffected by the temporal optimization, which are primarily circuit overhead reduction techniques.

6.8 Integration with other mitigation techniques

VarSaw can be combined with IBM's matrix-based complete measurement error mitigation (MBM) [1] for higher fidelity than either

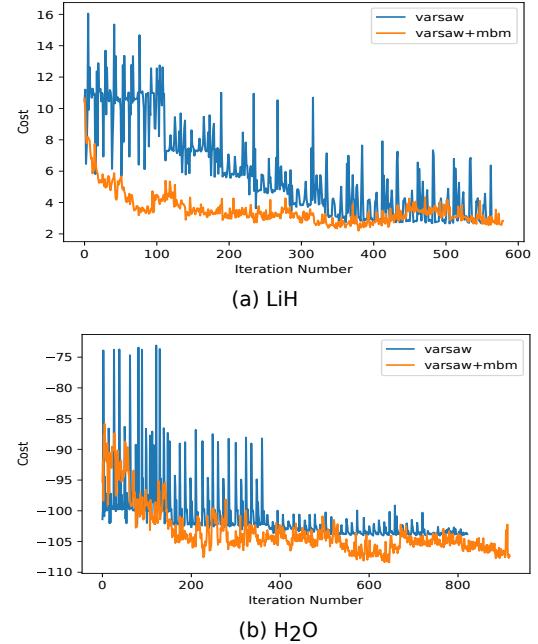


Figure 18: Training VQE with VarSaw and VarSaw+IBM's Measurement Error Mitigation for LiH and H_2O

scheme as standalone [18]. VQE results are shown in Fig.18 for two molecules. From our evaluations we see that VarSaw+MBM improved the VQE estimate by around 10% for H_2O and negligibly, but with less noise, for LiH . A tailored approach to deploying multiple measurement error mitigation techniques is worth pursuing in future work.

7 DISCUSSION

7.1 Why VarSaw works

The benefit from VarSaw's spatial optimization is fairly straightforward. Eliminating the spatial redundancy in JigSaw subsets is a significant win with no trade-offs. The benefit from VarSaw's temporal optimization is more interesting to consider. Clearly this is a trade-off. On the one hand, reducing global executions means that the 'weights' used to combine the VarSaw subsets together are stale. On the other hand, the global executions themselves have inaccuracy due to measurement errors (among other errors), so fresh global weights does not necessarily imply better overall VQE results.

Therefore, in the presence of non-trivial measurement noise, reducing the impact of the error-prone global runs is advantageous for VQA. It should also be noted that the VQA ansatz is inherently flexible. The rotational gates of the ansatz are tuned by the optimizer to produce a global state that minimizes the overall objective function. With VarSaw, it is intuitive the optimizer could converge towards the optimal state that minimizes the problem objective along an optimization path which is somewhat less dependent on global weights. This flexibility is offered by VQA and is one of the reasons why there are so many ansatz choices to solve a target VQA problem. Also important is that the reduced global executions

allows VQE to run many more iterations in the VarSaw scenario, so the VQA algorithm is able to search the problem space much faster even if the weighting mechanism is imperfect.

If measurement noise is negligible, then VarSaw's temporal optimization will not be beneficial. In such a scenario, we cannot expect VarSaw's temporal optimization to outperform the baseline approach. Any additional accuracy in VQA objective estimation that is enabled by the perfect global weights will be lost with VarSaw's stale weights - so this is VarSaw's worst case scenario (which is shown in the top of Fig.9). However, measurement errors are very significant, so we do not expect this to happen in practice. Additionally, it is possible that device calibration information could be used to influence the use of VarSaw's temporal optimization (or VarSaw as a whole). If some qubits have near-zero measurement errors, then VarSaw, or measurement error mitigation in general, is not required for these qubits.

7.2 Generalizing VarSaw

As a quantum application, variational algorithms clearly have the potential for a meaningful shot at a near-term quantum advantage. Thus, we need to enable them towards this goal by building solutions that are as tailored to them as possible. In this work, we are showing that general-purpose error mitigation techniques, when implemented in a domain-specific manner for VQAs, can provide significant gains beyond the capability of the original technique. Even more significant is the fact that the original technique (i.e., Jigsaw), when targeting this domain of VQA, is often detrimental at the application level due to its high computational cost (see Fig.13 for example), even though it has clear benefits at the circuit level (see Table.1). We show how its application-level benefits can be transformed by taking a domain-specific approach to provide substantial gains. While this work has focused on the intersection of VQA and measurement error mitigation, other optimization techniques, such as circuit synthesis and pulse-level control, can be tailored specifically to VQA. Similarly, measurement error mitigation can be tailored to other quantum applications like QFT and stabilizers.

7.3 VarSaw-specific extensions

While we have focused on VQE, VarSaw is useful for all VQA problems. The temporal optimization will produce benefits for all VQAs, but more benefits will be observed for problems that have Pauli terms spread across different measurement bases because this would increase the 'Global' cost. The spatial optimization is beneficial to those VQAs which have Pauli terms spread across different measurement bases (as this would present an opportunity for redundant 'Subsets' across the different Pauli terms) but not detrimental otherwise. These applications include, apart from VQE, time-evolving Hamiltonian simulations that encompass a broad range of algorithms such as the Ising model, Heisenberg model, XY model etc [33]. Quantitatively evaluating VarSaw on such applications would be useful future work. There is potential to employ measurement error mitigation only in specific phases of VQA and to only specific terms in the Hamiltonian - i.e., only employ mitigation where it matters most. This trade-off of cost vs accuracy is a suitable immediate extension to VarSaw. Next, in this work,

we have focused on VQA with a hardware-efficient ansatz [28]. A hardware-efficient ansatz can be limited in its capabilities because it is application-agnostic. In future work, we will look to explore these biases, as well as evaluate the benefits of VarSaw for other ansatz.

7.4 Related Work

Prior work VAQEM [45], dynamically tailors existing error mitigation techniques to the actual, dynamic noisy execution characteristics of VQAs on a target quantum machine. While similar in spirit, VAQEM incorporates parameters of different error mitigation techniques into the variational tuning harness - this is specifically suited to mitigation techniques that have tunable parameters and are inherently quantum optimizations (like dynamical decoupling [42].) On the other hand, our work specifically focuses on measurement error mitigation, and optimizes it in ways that is orthogonal to the VAQEM approach. Software error mitigation techniques, especially focused on measurement error mitigation have also been proposed in [14, 32, 54]. Measurement error mitigation techniques tailored to VQAs have been proposed in [9]. Measurement error can also be mitigated by circuit level advancements. For example. [35] has proposed a variety of techniques for machine-learning assisted qubit readout. Previous work has been done to improve the performance of VQAs using error mitigation techniques. Ref.[29] uses Zero Noise Extrapolation (ZNE) to increase the accuracy of VQAs on real quantum hardware. Ref.[60] shows that error mitigation can improve the trainability of VQAs and overcome the effects of noise-induced barren plateaus.

8 CONCLUSION

In this work, we seek to combat the effect of measurement errors in variational algorithms at reasonable cost. To do so, we build application-aware optimizations to the general-purpose JigSaw approach for measurement-error mitigation. VarSaw improves JigSaw [18], specifically in the context of VQAs, by identifying *Spatial Redundancy in JigSaw Subsets* (across VQA Pauli strings) and *Temporal Redundancy in JigSaw Globals* (across VQA iterations) and combating these in novel ways. In all, VarSaw reduces computational cost over naive JigSaw for VQA substantially while achieving significant improvements in VQA fidelity over both the unmitigated baseline as well as over JigSaw under budget constraints. Importantly, this work showcases the overwhelming benefits from tailoring state-of-the-art optimizations in a domain-specific manner. Designing JigSaw from the ground up in a VQA-cognizant manner, enables significant benefits.

ACKNOWLEDGMENTS

This work is funded in part by EPiQC, an NSF Expedition in Computing, under award CCF-1730449; in part by STAQ under award NSF Phy-1818914; in part by NSF award 2110860; in part by the US Department of Energy Office of Advanced Scientific Computing Research, Accelerated Research for Quantum Computing Program; and in part by the NSF Quantum Leap Challenge Institute for Hybrid Quantum Architectures and Networks (NSF Award 2016136) and in part based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science

Research Centers. FTC is Chief Scientist for Quantum Software at ColdQuanta and an advisor to Quantum Circuits, Inc.

REFERENCES

- [1] [n. d.]. IBM. 2010. Measurement Error Mitigation. <https://qiskit.org/textbook/ch-quantum-hardware/measurement-error-mitigation.html>. Accessed: 2020-7-1.
- [2] [n. d.]. IBM Quantum delivers 120x speedup of quantum workloads with Qiskit Runtime. <https://research.ibm.com/blog/120x-quantum-speedup>. Accessed: 2021-07-30.
- [3] [n. d.]. IBM Quantum SU2 ansatz. <https://qiskit.org/documentation/stubs/qiskit.circuit.library.EfficientSU2.html>. Accessed: 2021-12-01.
- [4] [n. d.]. IBM Quantum Systems. <https://quantum-computing.ibm.com/services/systems-all>. Accessed: 2021-12-01.
- [5] [n. d.]. SPSA: Simultaneous Perturbation Stochastic Approximation method. <https://www.jhuapl.edu/spsa/>. Accessed: 2021-30-07.
- [6] Héctor Abraham, AduOffei, Rochisha Agarwal, Ismail Yunus Akhalwaya, Gadi Aleksandrowicz, Thomas Alexander, Matthew Amy, Eli Arbel, Arijit02, Abraham Asfaw, Artur Avkhadiev, Carlos Azaustre, AzizNgoueya, Abhik Banerjee, Aman Bansal, Panagiotsis Barkoutsos, George Barron, George S. Barron, Luciano Bello, Yael Ben-Haim, Daniel Bevenius, Arjun Bhobe, Lev S. Bishop, Carsten Blank, Sorin Bolos, Samuel Bosch, Brandon, Sergey Bravyi, Bryce-Fuller, David Bucher, Artemiy Burov, Fran Cabrera, Padraig Calpin, Lauren Capelluto, Jorge Carballo, Ginés Carrascal, Adrian Chen, Chun-Fu Chen, Edward Chen, Jielun (Chris) Chen, Richard Chen, Jerry M. Chow, Spencer Churchill, Christian Claus, Christian Clauss, Romilly Cocking, Filipe Correa, Abigail J. Cross, Andrew W. Cross, Simon Cross, Juan Cruz-Benito, Chris Culver, Antonio D. Córcoles-Gonzales, Sean Dague, Tareq El Dandachi, Marcus Daniels, Matthieu Dartiailh, DavideFr, Abdón Rodríguez Davila, Anton Dekusar, Delton Ding, Jun Doi, Eric Drechsler, Drew, Eugene Dumitrescu, Karel Dumon, Ivan Duran, Kareem EL-Saffy, Eric Eastman, Grant Eberle, Pieter Eendebak, Daniel Egger, Mark Everitt, Paco Martín Fernández, Axel Hernández Ferrera, Romain Fouilland, FranckChevallier, Albert Frisch, Andreas Fuhrer, Bryce Fuller, MELVIN GEORGE, Julien Gacon, Borja Godoy Gago, Claudio Gambella, Jay M. Gambetta, Adhisha Gammampila, Luis Garcia, Tanya Garg, Shelly Garion, Austin Gilliam, Aditya Giridharan, Juan Gomez-Mosquera, Salvador de la Puente González, Jesse Gorzinski, Ian Gould, Donny Greenberg, Dmitry Grinko, Wen Guan, John A. Gunnels, Mikael Haglund, Isabel Haide, Ilko Hamamura, Omar Costa Hamido, Frank Harkins, Vojtech Havlicek, Joe Hellmers, Lukasz Herok, Stefan Hillmich, Hiroshi Horii, Connor Howington, Shaohan Hu, Wei Hu, Junye Huang, Rolf Huisman, Haruki Imai, Takashi Imamichi, Kazuaki Ishizaki, Raban Iten, Toshinari Itoko, JamesSeaward, Ali Javadi, Ali Javadi-Abhari, Jessica, Madhav Jivrajani, Kiran Johns, Scott Johnstun, Jonathan-Shoemaker, Vismai K, Tal Kachmann, Naoki Kanazawa, Kang-Bae, Anton Karazeev, Paul Kassebaum, Josh Kelso, Spencer King, Knabberjoe, Yuri Kobayashi, Arseny Kovyrshin, Rajiv Krishnakumar, Vivek Krishnan, Kevin Krsulich, Prasad Kunkar, Gawel Kus, Ryan LaRose, Enrique Lacal, Raphaël Lambert, John Lapeyre, Joe Latone, Scott Lawrence, Christina Lee, Gushu Li, Dennis Liu, Peng Liu, Yunhe Maeng, Kahan Majmudar, Aleksei Malyshev, Joshua Manela, Jakub Marecek, Manoel Marques, Dmitrii Maslov, Dolph Mathews, Atsushi Matsu, Douglas T. McClure, Cameron McGarry, David McKay, Dan McPherson, Srujan Meesala, Thomas Metcalfe, Martin Mevissen, Andrew Meyer, Antonio Mezzacapo, Rohit Midha, Zlatko Minev, Abby Mitchell, Nikolaj Moll, Jhon Montanez, Michael Duane Mooring, Renier Morales, Niall Moran, Mario Motta, MRF, Prakash Murali, Jan Müggenburg, David Nadlinger, Ken Nakanishi, Giacomo Nannicini, Paul Nation, Edwin Navarro, Yehuda Naveh, Scott Wyman Neagle, Patrick Neuweiler, Johan Nicander, Pradeep Niroula, Hassi Norlen, NuoWenLei, Lee James O'Riordan, Oluwatobi Ogunbayo, Pauline Ollitrault, Raul Otaolea, Steven Oud, Dan Padilha, Hanheo Paik, Soham Pal, Yuchen Pang, Simone Perriello, Anna Phan, Francesco Piro, Marco Pistoia, Christophe Piveteau, Pierre Poreau, Alejandro Pozas-ikErstjens, Viktor Prutyanov, Daniel Puzzuoli, Jesús Pérez, Quintii, Rafey Iqbal Rahman, Arun Raja, Nipun Ramagiri, Anirudh Rao, Rudy Raymond, Rafael Martin-Cuevas Redondo, Max Reuter, Julia Rice, Marcello La Rocca, Diego M. Rodríguez, RohithKarur, Max Rosmannek, Minki Ryu, Tharmashastha SAPV, SamFerracin, Martin Sandberg, Hirmay Sandesara, Ritvik Sapra, Hayk Sargsyan, Aniruddha Sarkar, Ninad Sathaye, Bruno Schmitt, Chris Schnabel, Zachary Schoenfeld, Travis L. Scholten, Eddie Schoute, Joachim Schwarm, Ismael Faro Sertage, Kanav Setia, Nathan Shammah, Yunong Shi, Adenilton Silva, Andrea Simonetto, Nick Singstock, Yukio Siraichi, Iskandar Situdikov, Seyon Sivarajah, Magnus Berg Sletfjording, John A. Smolin, Mathias Soeken, Igor Olegovich Sokolov, Igor Sokolov, SooluThomas, Starfish, Dominik Steenken, Matt Stypulkoski, Shaojun Sun, Kevin J. Sung, Hitomi Takahashi, Tanvesh Takawale, Ivana Tavernelli, Charles Taylor, Pete Taylor, Soolu Thomas, Mathieu Tillet, Maddy Tod, Miroslav Tomaskin, Enrique de la Torre, Kenso Trabing, Matthew Treinish, TrishaPe, Davendra Tulsi, Wes Turner, Yotam Vaknin, Carmen Recio Valcarce, Francois Varchon, Almudena Carrera Vazquez, Victor Villar, Desirée Vogt-Lee, Christophe Vuillot, James Weaver, Johannes Weidenfeller, Rafał Wieczorek, Jonathan A. Wildstrom, Erick Winston, Jack J. Woehr, Stefan Woerner, Ryan Woo, Christopher J. Wood, Ryan Wood, Stephen Wood, Steve Wood, James Wootton, Daniyar Yeralin, David Yonge-Mallo, Richard Young, Jessie Yu, Christopher Zachow, Laura Zdanski, Helena Zhang, Christa Zoufal, Zoufalc, a kapila, a matsuo, bcamorrison, brandhsn, nick bronn, chlorophyll zz, dekel.meiron, dekelmeiron, dekool, dime10, drholmie, dtrenev, ehchen, elfrorampeador, faisaldebouni, fanizamarco, gabrieleagl, gadial, galeinston, georgios ts, gruu, hhorii, hykavitha, jagunther, jliu45, jscott2, kanejess, klinvill, krtuk2966, kurarr, lerongil, ma5x, merav aharoni, michelle4654, ordmoj, sagar pahwa, rmo- yard, saswati qiskit, scottkelso, sethmerkel, strickroman, sumitpuri, tigerjack, toural, tsura crisaldo, vvilpas, welien, willhbang, yang.luh, yotamvakninibm, and Mantas Čepulkovskis. 2019. Qiskit: An Open-source Framework for Quantum Computing. <https://doi.org/10.5281/zenodo.2562110>
- [7] Google Quantum AI. 2021. Exponential suppression of bit or phase errors with cyclic error correction. *Nature* 595, 7867 (2021), 383.
- [8] Google Quantum AI. Accessed: June 19, 2021. Quantum Computer Datasheet. <https://quantumai.google/hardware/datasheet/weber.pdf>.
- [9] George S Barron and Christopher J Wood. 2020. Measurement error mitigation for variational quantum algorithms. *arXiv preprint arXiv:2010.08520* (2020).
- [10] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. 2017. Quantum machine learning. *Nature* 549, 7671 (2017), 195–202.
- [11] M J Biercuk, A C Doherty, and H Uys. 2011. Dynamical decoupling sequence construction as a filter-design problem. *Journal of Physics B: Atomic, Molecular and Optical Physics* 44, 15 (Jul 2011), 154002. <https://doi.org/10.1088/0953-4075/44/15/154002>
- [12] Ludmila Botelho, Adam Glos, Akash Kundu, Jaroslaw Adam Miszczak, Özlem Salehi, and Zoltán Zimbrás. 2021. Error mitigation for variational quantum algorithms through mid-circuit measurements. *arXiv:2108.10927* [quant-ph]
- [13] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C Mckay, and Jay M Gambetta. 2020. Mitigating measurement errors in multi-qubit experiments. *arXiv preprint arXiv:2006.14044* (2020).
- [14] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C Mckay, and Jay M Gambetta. 2021. Mitigating measurement errors in multiqubit experiments. *Physical Review A* 103, 4 (2021), 042605.
- [15] Rigetti Computing. 2019. Pyquil documentation. URL <http://pyquil.readthedocs.io/en/latest> (2019), 64–65.
- [16] Piotr Czarnik, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio. 2020. Error mitigation with Clifford quantum-circuit data. *arXiv:2005.10189* [quant-ph]
- [17] Poulam Das, Swapnil Tannu, Siddharth Dangwal, and Moinuddin Qureshi. 2021. Adapt: Mitigating idling errors in qubits via adaptive dynamical decoupling. In *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture*. 950–962.
- [18] Poulam Das, Swapnil Tannu, and Moinuddin Qureshi. 2021. JigSaw: Boosting Fidelity of NISQ Programs via Measurement Subsetting. In *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture* (Virtual Event, Greece) (MICRO '21). Association for Computing Machinery, New York, NY, USA, 937–949. <https://doi.org/10.1145/3466752.3480044>
- [19] Yongshan Ding, Pranav Gokhale, Sophia Fuhui Lin, Richard Rines, Thomas Propson, and Frederic T Chong. 2020. Systematic Crosstalk Mitigation for Superconducting Qubits via Frequency-Aware Compilation. *arXiv preprint arXiv:2008.09503* (2020).
- [20] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. 2014. A Quantum Approximate Optimization Algorithm. *arXiv:1411.4028* [quant-ph]
- [21] Pranav Gokhale, Olivia Anguli, Yongshan Ding, Kaiwen Gui, Teague Tomesh, Martin Suchara, Margaret Martonosi, and Frederic T Chong. 2019. Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families. *arXiv preprint arXiv:1907.13623* (2019).
- [22] Pranav Gokhale, Yongshan Ding, Thomas Propson, Christopher Winkler, Nelson Leung, Yunong Shi, David I. Schuster, Henry Hoffmann, and Frederic T. Chong. 2019. Partial Compilation of Variational Algorithms for Noisy Intermediate-Scale Quantum Machines. *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture* (Oct 2019). <https://doi.org/10.1145/3352460.3358313>
- [23] E. L. Hahn. 1950. Spin Echoes. *Phys. Rev.* 80 (Nov 1950), 580–594. Issue 4. <https://doi.org/10.1103/PhysRev.80.580>
- [24] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P Lanyon, Peter Love, Ryan Babbush, et al. 2018. Quantum chemistry calculations on a trapped-ion quantum simulator. *Physical Review X* 8, 3 (2018), 031022.
- [25] IBM. 2010. Measurement Error Mitigation. <https://qiskit.org/textbook/ch-quantum-hardware/measurement-error-mitigation.html>. [Accessed July-2020].
- [26] James Joyce. 2003. Bayes' theorem. (2003).
- [27] Petar Jurcevic, Ali Javadi-Abhari, Lev S Bishop, Isaac Lauer, Daniela Borgorin, Markus Brink, Lauren Capelluto, Oktay Gunluk, Toshinari Itoko, Naoki Kanazawa, et al. 2021. Demonstration of quantum volume 64 on a superconducting quantum computing system. *Quantum Science and Technology* (2021).
- [28] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. 2017. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* 549, 7671 (2017), 242–246.

[29] Abhinav Kandala, Kristan Temme, Antonio D Córcoles, Antonio Mezzacapo, Jerry M Chow, and Jay M Gambetta. 2019. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* 567, 7749 (2019), 491–495.

[30] Kaveh Khodjasteh and Daniel A. Lidar. 2007. Performance of deterministic dynamical decoupling schemes: Concatenated and periodic pulse sequences. *Physical Review A* 75, 6 (Jun 2007). <https://doi.org/10.1103/physreva.75.062310>

[31] Philip Krantz, Morten Kjaergaard, Fei Yan, Terry P Orlando, Simon Gustavsson, and William D Oliver. 2019. A quantum engineer’s guide to superconducting qubits. *Applied Physics Reviews* 6, 2 (2019), 021318.

[32] Hyekjea Kwon and Joonwoo Bae. 2020. A hybrid quantum-classical approach to mitigating measurement errors. *arXiv preprint arXiv:2003.12314* (2020).

[33] Lingling Lao and Dan E. Browne. 2021. 2QAN: A quantum compiler for 2-local qubit Hamiltonian simulation algorithms. <https://doi.org/10.48550/ARXIV.2108.02099>

[34] Wim Lavrijen, Ana Tudor, Julianne Müller, Costin Iancu, and Wibe de Jong. 2020. Classical Optimizers for Noisy Intermediate-Scale Quantum Devices. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*. 267–277. <https://doi.org/10.1109/QCE49297.2020.00041>

[35] Benjamin Lienhard. 2021. *Machine Learning Assisted Superconducting Qubit Readout*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[36] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. 2016. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics* 18, 2 (2016), 023023.

[37] Jarrod R McClean, Nicholas C Rubin, Kevin J Sung, Ian D Kivlichan, Xavier Bonet-Monroig, Yudong Cao, Chengyu Dai, E Schuyler Fried, Craig Gidney, Brendan Gimby, et al. 2020. OpenFermion: the electronic structure package for quantum computers. *Quantum Science and Technology* 5, 3 (2020), 034014.

[38] Nikolaj Moll, Panagiotis Barkoutsos, Lev S Bishop, Jerry M Chow, Andrew Cross, Daniel J Egger, Stefan Filipp, Andreas Fuhrer, Jay M Gambetta, Marc Ganzhorn, et al. 2018. Quantum optimization using variational algorithms on near-term quantum devices. *Quantum Science and Technology* 3, 3 (2018), 030503.

[39] Prakash Murali, Jonathan M Baker, Ali Javadi-Abhari, Frederic T Chong, and Margaret Martonosi. 2019. Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers. In *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*. 1015–1029.

[40] Prakash Murali, David C McKay, Margaret Martonosi, and Ali Javadi-Abhari. 2020. Software mitigation of crosstalk on noisy intermediate-scale quantum computers. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*. 1001–1016.

[41] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O’Brien. 2014. A variational eigenvalue solver on a photonic quantum processor. *Nature communications* 5 (2014), 4213.

[42] Bibek Pokharel, Namit Anand, Benjamin Fortman, and Daniel A. Lidar. 2018. Demonstration of Fidelity Improvement Using Dynamical Decoupling with Superconducting Qubits. *Physical Review Letters* 121, 22 (Nov 2018). <https://doi.org/10.1103/physrevlett.121.220502>

[43] John Preskill. 2018. Quantum Computing in the NISQ era and beyond. *Quantum* 2 (2018), 79.

[44] Gokul Subramanian Ravi, Pranav Gokhale, Yi Ding, William M. Kirby, Kaitlin N. Smith, Jonathan M. Baker, Peter J. Love, Henry Hoffmann, Kenneth R. Brown, and Frederic T. Chong. 2022. CAFQA: Clifford Ansatz For Quantum Accuracy. <https://doi.org/10.48550/ARXIV.2202.12924>

[45] Gokul Subramanian Ravi, Kaitlin N. Smith, Pranav Gokhale, Andrea Mari, Nathan Earnest, Ali Javadi-Abhari, and Frederic T. Chong. 2021. VAQEM: A Variational Approach to Quantum Error Mitigation. *arXiv:2112.05821* [quant-ph]

[46] Elliott Rosenberg, Paul Ginsparg, and Peter L. McMahon. 2021. Experimental error mitigation using linear rescaling for variational quantum eigensolving with up to 20 qubits. *Quantum Science and Technology* (Nov 2021). <https://doi.org/10.1088/2058-9565/ac3b37>

[47] Mohan Sarovar, Timothy Proctor, Kenneth Rudinger, Kevin Young, Erik Nielsen, and Robin Blume-Kohout. 2020. Detecting crosstalk errors in quantum information processors. *Quantum* 4 (2020), 321.

[48] Yunong Shi, Nelson Leung, Pranav Gokhale, Zane Rossi, David I. Schuster, Henry Hoffmann, and Frederic T. Chong. 2019. Optimized Compilation of Aggregated Instructions for Realistic Quantum Computers. *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems* (Apr 2019). <https://doi.org/10.1145/3297858.3304018>

[49] Kaitlin N Smith, Gokul Subramanian Ravi, Prakash Murali, Jonathan M Baker, Nathan Earnest, Ali Javadi-Abhari, and Frederic T Chong. 2021. Error Mitigation in Quantum Computers through Instruction Scheduling. *arXiv preprint arXiv:2105.01760* (2021).

[50] Robert S Smith, Michael J Curtis, and William J Zeng. 2016. A practical quantum instruction set architecture (2016). *arXiv preprint arXiv:1608.03355* 29 (2016).

[51] Alexandre M Souza, Gonzalo A Alvarez, and Dieter Suter. 2012. Robust dynamical decoupling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370, 1976 (2012), 4748–4769.

[52] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. 2017. The Python-based Simulations of Chemistry Framework (PySCF). *arXiv:1701.08223* [physics.chem-ph]

[53] Ryuji Takagi, Suguru Endo, Shintaro Minagawa, and Mile Gu. 2021. Fundamental limits of quantum error mitigation. *arXiv:2109.04457* [quant-ph]

[54] Swamit S Tannu and Moinuddin K Qureshi. 2019. Mitigating measurement errors in quantum computers by exploiting state-dependent bias. In *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*. 279–290.

[55] Swamit S Tannu and Moinuddin K Qureshi. 2019. Mitigating Measurement Errors in Quantum Computers by Exploiting State-Dependent Bias. In *MICRO*. 279–290.

[56] Swamit S Tannu and Moinuddin K Qureshi. 2019. Not all qubits are created equal: a case for variability-aware policies for NISQ-era quantum computers. In *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*. 987–999.

[57] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H. Booth, and Jonathan Tennyson. 2021. The Variational Quantum Eigensolver: a review of methods and best practices. *arXiv:2111.05176* [quant-ph]

[58] Benjamin Villalonga, Dmitry Lyakh, Sergio Boixo, Hartmut Neven, Travis S Humble, Rupak Biswas, Eleanor G Rieffel, Alan Ho, and Salvatore Mandrà. 2020. Establishing the quantum supremacy frontier with a 281 Pflop/s simulation. *Quantum Science and Technology* 5, 3 (apr 2020), 034003. <https://doi.org/10.1088/2058-9565/ab7eeb>

[59] Samson Wang, Piotr Czarnik, Andrew Arrasmith, M. Cerezo, Lukasz Cincio, and Patrick J. Coles. 2021. Can Error Mitigation Improve Trainability of Noisy Variational Quantum Algorithms? *arXiv:2109.01051* [quant-ph]

[60] Samson Wang, Piotr Czarnik, Andrew Arrasmith, Marco Cerezo, Lukasz Cincio, and Patrick J Coles. 2021. Can error mitigation improve trainability of noisy variational quantum algorithms? *arXiv preprint arXiv:2109.01051* (2021).

A EVALUATING DIFFERENT SUBSET SIZES

In the JigSaw framework, (i.e., when targeting a single a single circuit), fixing the subset size naively to 2 may not be optimal. The optimal size would have to be chosen as a trade-off between measurement errors, circuit overheads, reconstruction errors etc. In general for JigSaw, smaller subsets means lower measurement error (until some point of saturation) but suffers from higher overheads.

On the other hand, with VarSaw, the smaller subsets are always beneficial—both in terms of accuracy as well as in terms of number of circuits. There are two aspects here—the first (similar to JigSaw) is that smaller subsets have potentially greater potential for measurement error mitigation (discussed prior). The second aspect, and unique to VarSaw, is that smaller subsets actually produce the least number of total circuits for execution. This is because smaller subsets enable considerably larger amounts of commutativity based circuit reduction. Thus, VarSaw with 2-qubit subsets is attractive both in terms of maximum error mitigation as well as in terms of lowest number of circuits. Thus, it is the clear choice for our chosen applications. We expect this trend to broadly hold, except for some variation depending on application and machine characteristics.

In Fig.19, we run a single instance of the VQE circuit with ideal parameters in noisy simulation and perform VarSaw’s measurement error mitigation on top of this. We plot VarSaw’s fidelity improvement over the noisy baseline for 6 qubit LiH, H₂O, and CH₄. We show results for 4 different subset sizes, all of which produce substantial benefits over the baseline. For all 3 benchmarks, the accuracy benefits did not vary significantly with subset size – in LiH, the smallest subset performs around 10% better than other choices. In others, the different subset sizes are all attractive in terms of accuracy improvement. But the important aspect to note is the window size of 2 achieves a considerably smaller number of total circuits (this number increases with subset size). Thus, smaller subsets are more attractive for VarSaw.

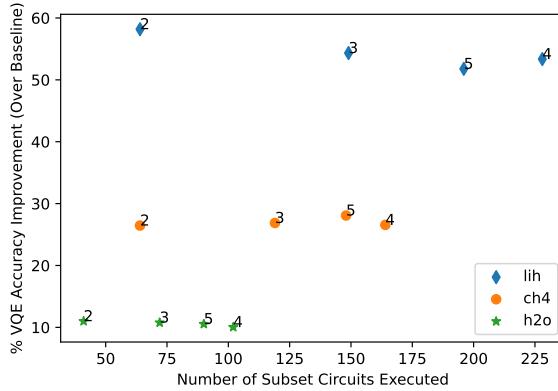


Figure 19: For one iteration of VQE at optimal parameters, VarSaw is applied on the noisy baseline. While accuracy improvements from subsetting are high, the accuracy variance among small subset sizes is low. But the circuit overhead is substantially lower for smallest circuit subsets.

B PERFORMANCE OF SPARSIFICATION AT VARIED NOISE

Finally, we evaluate the VQE accuracy benefits from VarSaw's temporal optimization (i.e., adding sparsity) at different noise levels, to showcase that this optimization is beneficial over a wide noise

range. To do so, we scale the device noise model and use it to run the VQE baseline, VarSaw with No Sparsity and VarSaw with Max Sparsity. Table 5 shows the results for H₂O-6 molecule. We see that as the noise scales, VarSaw at Max Sparsity always achieves benefits over the baseline and the benefits are always similar (and sometimes better) compared to the No Sparsity scenario. Thus, not only is the VQE result highly attractive when temporal optimization is employed, it also significantly reduces computational cost. And this trend holds across a variety of noise. Note that when there is no noise at all, then Max Sparsity can perform poorly - this is expected and was discussed in Fig.9 and Section 3.3.

Noise Scale	Baseline	VarSaw (No Sparsity)	VarSaw (Max Sparsity)
5	-89.33	-91.29	-91.36
3	-90.86	-94.51	-93.32
1	-98.28	-101.73	-102.73
0.8	-99.45	-102.95	-101.08
0.5	-101.62	-104.30	-103.23
0.1	-104.15	-104.46	-104.69
0.05	-104.57	-104.75	-104.77

Table 5: VarSaw with Max Sparsity and No Sparsity compared to the baseline for different noise levels.

Received 22 October 2022; revised 2 March 2023; accepted 27 April 2023