Single-Qubit Cross Platform Comparison of Quantum Computing Hardware

Adrien Suau, Jon Nelson, Marc Vuffray, Andrey Y. Lokhov, Lukasz Cincio, Carleton Coffrin

Single-Qubit Cross Platform Comparison of Quantum Computing Hardware

Adrien Suau,[1, 2, 3] Jon Nelson,[4] Marc Vuffray,[1] Andrey Y. Lokhov,[1] Lukasz Cincio,[1] and Carleton Coffrin[4]

[1] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[2] *CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse, France*
[3] *LIRMM, University of Montpellier, 161 rue Ada, 34095 Montpellier, France*
[4] *Advanced Network Science Initiative, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

As a variety of quantum computing models and platforms become available, methods for assessing and comparing the performance of these devices are of increasing interest and importance. Despite being built of the same fundamental computational unit, radically different approaches have emerged for characterizing the performance of qubits in gate-based and quantum annealing computers, limiting and complicating consistent cross-platform comparisons. To fill this gap, this work proposes a single-qubit protocol (Q-RBPN) for measuring some basic performance characteristics of individual qubits in both models of quantum computation. The proposed protocol scales to large quantum computers with thousands of qubits and provides insights into the distribution of qubit properties within a particular hardware device and across families of devices. The efficacy of the Q-RBPN protocol is demonstrated through the analysis of more than 300 gate-based qubits spanning eighteen machines and 2000 annealing-based qubits from one machine, revealing some unexpected differences in qubit performance. Overall, the proposed Q-RBPN protocol provides a new platform-agnostic tool for assessing the performance of a wide range of emerging quantum computing devices.

I. INTRODUCTION

In the current era of Noisy Intermediate-Scale Quantum (NISQ) devices [1], a wide variety of technologies are being developed that leverage quantum mechanics to conduct, and hopefully accelerate, computations [2–5]. Given the noisy nature of these emerging technologies, measuring and tracking the fidelity of quantum hardware platforms is essential to understanding the limitations of these devices and quantifying progress as these platforms continue to improve. Measuring the performance of gate-based quantum computers (QC) has been studied extensively through the topics of quantum characterization, verification, and validation (QCVV) [6]. The scope of QCVV is broad and ranges from testing individual quantum operations (e.g., error rates of one- and two-qubit gates [7]), verifying small circuits (e.g., Randomized Benchmarking [8, 9], Gate Set Tomography [10]), to full system-level protocols (e.g., quantum volume estimation [11], random quantum circuits [12]). Over the years these QCVV tools have become an invaluable foundation for benchmarking and measuring progress of quantum processors [13], culminating with a quantum supremacy demonstration in 2019 [14].

Interestingly, this large body of QCVV work cannot usually be applied to the assessment of quantum annealing (QA) computers, such as the quantum devices developed by D-Wave Systems [4, 15]. The fundamental challenge in conducting characterization, verification, and validation of quantum annealing devices (QAVV) is that available hardware platforms only allow measuring the state of the system in a fixed basis (the so-called computational z-basis) and at the completion of a specified annealing protocol. Consequently, the QA user can only observe a fairly limited projection of the quantum state that occurs during the hardware's computation. Despite these challenges, recent work has proposed the QASA protocol for performing single-qubit fidelity assessment of QA hardware platforms [16].

The notable differences in the computational models for QC and QA present significant challenges for conducting consistent comparisons of these hardware platforms. Such efforts have focused almost exclusively on system-level benchmarks for applications like optimization and sampling featuring with 10s to 100s of qubits [17–20]. Conducting consistent comparisons of this type are notoriously difficult as these applications require the execution of complex quantum programs and often feature a large number of implementation parameters that impact performance measures, leaving significant room for debate around the fairness of these comparisons. A key benefit in conducting QCVV/QAVV on small quantum systems (e.g., 1-3 qubits) is a reduction in protocol complexity that can reduce debate and increase confidence on the benchmark's results.

The core observation of this work is that the single-qubit QASA protocol, that was originally developed for tracking the performance of QA qubits [16], can be adapted for execution on the more general QC hardware platforms. To that end, this work proposes the Qubit Response, Bias, Positive saturation, Negative saturation (Q-RBPN) protocol to measure four basic properties of the qubits in quantum hardware platforms. In the broader context of QCVV, the Q-RBPN protocol is limited in the results it yields: however, its principal advantage is in providing a comparison of qubit performance that is agnostic to the underlying computational model. To the best of our knowledge, this work provides the first side-by-side comparison of single-qubit performance across the QC (IBM-Q) and QA (D-Wave Systems) hardware platforms.

This work begins with a theoretical discussion of the computational tasks that underpins the Q-RBPN protocol and how it can be executed in different computing

platforms in Section II. This explanation is followed by an analysis of more than 300 QC qubits and 2000 QA qubits in Section III, with some discussion of the disintct features observed in different platforms. Section IV concludes the paper with a discussion of the value of universal qubit performance metrics and the broader implications of the results produced by the Q-RBPN protocol in this work.

II. SINGLE-QUBIT PROGRAMMING MODELS

At first glance, the computational task used in this work is peculiar. It is inspired by sampling from Gibbs distributions, a task that both QC and QA devices were not intended to solve. However, as this work demonstrates, this unconventional task has the advantage of being realizable in both computational models in practice. In the current NISQ era, we observe that this sampling task provides a useful measure of the consistency of qubit behaviors and output realizations in both QC and QA hardware platforms.

At the most abstract level, this work leverages a single input parameter (i.e., h^{in}) to control the relative amount of up/down measurements (i.e., $\sigma \in \{-1, +1\}$) that are observed in the output of the quantum device. The hardware's performance is determined by how consistently it responds to different values of h^{in} within a range of -1.0 to 1.0. Given the restrictions imposed by current QA hardware platforms this work only considers measuring spin projections in a fixed computational basis (denoted as the z-basis).

Drawing inspiration from [16], this work will focus on fitting Gibbs distributions to the outcomes of a single-qubit quantum programs, i.e.,

$$P(\sigma) \propto \exp\left(\beta h^{\mathrm{in}} \sigma\right).,\tag{1}$$

where σ is an observed spin state, h^{in} is an input parameter, and β is a parameter corresponding to a thermal equilibrium at an effective temperature $1/\beta$. In QA, the β parameter is determined by the details of the annealing protocol [16], whereas in QC it can be selected as part of the protocol. The probability distribution over any single binary variable $\sigma \in \{-1, +1\}$ can be fully characterized by a single parameter h^{eff}, coined the effective field, in the following manner:

$$\mathbb{P}\left(\sigma = \pm 1\right) = \frac{\exp\left(h^{\mathrm{eff}} \sigma\right)}{2 \cosh h^{\mathrm{eff}}}.\tag{2}$$

The value of h^{eff} depends on the experiment's input parameters and is, in particular, a function of the input field h^{in}. In the case of a classical magnet placed into a persistent external magnetic field h in a thermal equilibrium at temperature $1/\beta$, one will observe a linear relationship between the output and input fields of the form $h^{\mathrm{eff}} = \beta h^{\mathrm{in}}$. This linear mapping is called a *classical* Gibbs distribution for a single spin; however, as noted in

[16], the noisy nature of quantum hardware yields deviations from this linear mapping revealing imperfections and limitations of the hardware. The remainder of this section details how the QC and QA platforms can be programmed to conduct single-qubit Gibbs sampling and how those observations are converted into h^{eff} values.

A. Theoretical Models

1. Gate-based Quantum Computing

Our goal is to develop a quantum circuit that will include a free input parameter h^{in} and whose output will follow the Gibbs distribution from Eq. (1). To this end, we transform a single qubit from its initialized state $|0\rangle$ to a general state on the Bloch sphere,

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle,\tag{3}$$

and perform a measurement along the computational basis. The mapping is realized using the 1-qubit gates $R_y(\theta) = e^{-i\theta Y}$ and $R_z(\phi) = e^{-i\phi Z}$, where X, Y, Z denote the Pauli matrices. In order to obtain the linear relationship between the output h^{eff} and input h^{in} fields, we prescribe the following mapping between the input parameter h^{in} and the angle θ,

$$\theta = \cos^{-1}\left(\tanh\left(\beta h^{\mathrm{in}}\right)\right).\tag{4}$$

This value of θ depends on β and h^{in}, and links the observed output field h^{eff} and the input field h^{in} with the ideal linear relationship $h^{\mathrm{eff}} = \beta h^{\mathrm{in}}$. This property means that, in an error-less QC model, one can compute an output field that is equal to βh^{in} from the data collected on QC device.

To see that the aforementioned linear relationship is indeed the theoretically expected one, observe that the expected value of a spin prepared in the state $|\psi\rangle$ from Eq. (3) and measured along the computational Z basis is,

$$\langle\psi|Z|\psi\rangle = \cos\left(\theta\right).\tag{5}$$

Substituting the expression of θ from Eq. (4) in Eq. (5), gives

$$\langle\psi|Z|\psi\rangle = \tanh\left(\beta h^{\mathrm{in}}\right).\tag{6}$$

Now, notice that the expectation of the measured output distribution from Eq. (2) is,

$$\mathbb{E}\left[\sigma\right] = \tanh\left(h^{\mathrm{eff}}\right).\tag{7}$$

Therefore, we see that if the single spin system works ideally, one will obtain the following linear relationship between h^{in} and h^{eff},

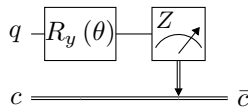$$h^{\mathrm{eff}} = \beta h^{\mathrm{in}}.\tag{8}$$

FIG. 1: Generic 1-qubit quantum program used by the Q-RBPN protocol on QC hardware. The rotation angle θ is related to an h^{in} value according to Eq. (4).

This linear relationship is, in theory, invariant with respect of the value of ϕ. In Appendix A, we show that, despite potential qubit imperfections, this invariance remained true in practice as well. Based on this observation, the results of this work use the special case where $\phi = 0$, which only requires R_y rotations. The quantum circuit that is ultimately utilized by the Q-RBPN protocol on QC hardware is depicted in Fig. 1.

2. Quantum Annealing

In contrast to QC, which can encode arbitrary quantum Hamiltonians, current quantum annealing platforms specialize in modeling the Hamiltonian of the transverse field Ising model, i.e.,

$$H_{\text{T-Ising}} = \Gamma \sum_i X_i + \sum_{i,j} J_{ij} Z_i Z_j + \sum_i h_i Z_i, \quad (9)$$

where h, J encode the local fields and interaction strengths of a classical Ising model and Γ controls the strength of a global transverse field on that model. This specialized Hamiltonian is of practical interest as the Ising model component can readily encode challenging computational problems arising in the study of magnetic materials, machine learning, and optimization [21–24] while the transverse field component yields complex entanglement structures that can be challenging to simulate [25, 26].

The quantum annealing protocol strives to find the low-energy assignments to a user-specified Ising problem by conducting an interpolation process of $H_{\text{T-Ising}}$ as follows:

$$H_{\text{T-Ising}}(s) =$$
$$A(s) \sum_i X_i + B(s) \left(\sum_{i,j} J_{ij} Z_i Z_j + \sum_i h_i Z_i \right). \quad (10)$$

The interpolation process starts with $s = 0$ and ends with $s = 1$. The two interpolation functions $A(s)$ and $B(s)$ are designed such that $A(0) \gg B(0)$ and $A(1) \ll B(1)$, that is, starting with a Hamiltonian dominated by the transverse field and slowly transitioning to a Hamiltonian dominated by an Ising model. The outcome of this quantum annealing process is specified by a binary variable σ_i that takes a value $+1$ or -1 and corresponds to

the observation of the spin projection in the computational basis Z.

In an idealized setting, and when the annealing interpolation transition process is sufficiently slow, the quantum annealing is referred to as adiabatic quantum computation. The adiabatic theorem states that if the interpolation is sufficiently slow and the quantum system is isolated, the proposed QA protocol will always find the ground state of the H_{Ising} problem [25, 27]. However, existing QA hardware platforms are open-quantum systems [15, 28–30] yielding outputs that are similar to thermal equilibrium distributions at an inverse effective temperature of $\beta \approx 10$ [16, 31].

Focusing on the single-qubit context considered by this work, the QA Hamiltonian (10) reduces to,

$$H_{\text{T-Ising}}(s) = A(s) X + B(s) h^{\text{in}} Z. \quad (11)$$

Notice that the expectation of the measured output of this QA program can be used to compute h^{eff} similarly to the QC case. Despite the simplicity of this model, it has been observed that the imperfections of real-world QA platforms make it a useful tool for assessing the performance of individual qubits in practice [16].

B. Illustration on Typical Qubits

To make this single-qubit evaluation procedure concrete, Fig. 2a provides an example of performing the complete protocol on representative qubits from the IBM-Q computer `ibm_lagos` and the D-Wave 2000Q QA computer `DW_2000Q_LANL`. In this illustration, a variety of input fields (i.e., h^{in}) ranging from -1 to $+1$ are executed on the hardware and the h^{eff} values are computed. Noting that the QA protocol of `DW_2000Q_LANL` yields $\beta \approx 10$ [16, 31], in order to have comparable $h^{\text{in}}/h^{\text{eff}}$ relationship between the two models of computation, we fix $\beta = 10$ in Eq. (4) of the QC model as well.

Recall from the discussion in Section II that we expect a linear relationship between the values of $h^{\text{in}}/h^{\text{eff}}$. Reviewing the results presented in Fig. 2a largely confirms the expected linear relationship; however, one can observe notable deviations from the expected model including h^{eff} saturation effects and minor deviations in the linear response. Two observations are of particular note: (1) the QA hardware realizes a maximal h^{eff} magnitude around 5, which is about twice as large as the maximum magnitude realized by the QC hardware (≈ 2.5); and (2) there is a notable asymmetry in the maximum magnitude of the negative (≈ -2.0) and positive (≈ 2.5) h^{eff} values realized by the QC hardware. The root source of these deviations and the differences between QC and QA platforms present interesting questions; however, we will leave those investigations to future work and begin by proposing metrics that provide a coarse summary of a qubit's characteristics in this $h^{\text{in}}/h^{\text{eff}}$ procedure. As illustrated in Fig. 2a, the 4 metrics we propose are: *response / bias*, the slope and the offset extracted from

fitting an affine model in the linear operating region of the qubit; and *positive / negative saturation*, the maximum and minimum observed h^{eff} values capturing the hardware's output saturation points. To fit the affine model, a subset of the data is selected $h^{\text{in}} \in [-0.1, 0.1]$, where the response behaves linearly in both hardware platforms. In general, this range may need to be revised based on the performance of a specific hardware device to yield and accurate fit of the data.

It is important to briefly remark on the data requirements for an accurate estimation of h^{eff}, especially for large h^{in} values. For each value of h^{in}, one collects M samples to extract a conditional expectation $\widehat{\mathbb{E}}\left[\sigma \mid h^{\text{in}}\right]$, which corresponds to an empirical effective field $h^{\text{eff}} = \tanh^{-1} \widehat{\mathbb{E}}\left[\sigma \mid h^{\text{in}}\right]$. For a particular value of M, this estimator is subject to an accuracy limit caused by finite sampling. For large values of $|h^{\text{eff}}|$, the probability of observing a qubit misaligned with the effective field decreases exponentially with the field's intensity; see Eq. (2). Therefore, if $h^{\text{eff}} = 5$ one only expects to see one misaligned spin configuration in every $22,000$ observations, requiring millions of samples to have a confident estimation of h^{eff}. Hence, it is necessary to adjust these data collection requirements to be consistent with the hardware's performance. This finite sampling accuracy challenge is addressed in this work by setting M to a level that provides tight confidence intervals around the estimation of h^{eff} for the particular QC and QA hardware that were considered; this resulted in $M = 8 \times 10^3$ and $M = 5 \times 10^6$ for each hardware platform respectively. To that end, 99.7% confidence intervals (i.e., 3 sigma) for the experimental data collected by this work are shown in Fig. 2a.

C. Limitations

It worth briefly mentioning some of the limitations of this protocol and metrics proposed in this work. It is widely known that current qubits suffer from notable amounts of state preparation and measurement errors (SPAM) [1] yielding spin flips on the order of 1 in 100 measurements [32], which limits the h^{eff} that can be observed on that hardware. QA hardware is known to suffer from a wide range of *integrated control errors* (ICE) [15], which include: background susceptibility; flux noise; Digital-to-Analog Conversion quantization; Input/Ouput system effects; and variable scale across qubits. The differences in how these hardware errors impact qubit performance across these platforms presents a notable challenge for developing hardware agnostic performance metrics. Although the metrics proposed in this work do not capture these specific known issues with qubit performance, they have the advantage of not requiring any assumptions about the particular computing model or underlying qubit implementation. Developing approaches for capturing the signatures of a wider variety of qubit errors presents an opportunity for expanding the Q-RBPN protocol in future work.
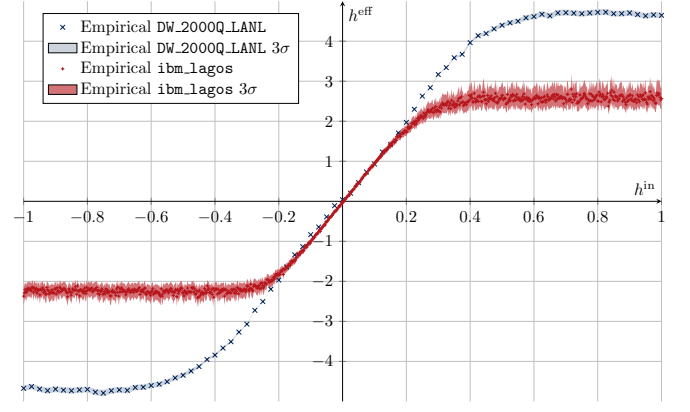
(a) Data obtained after applying the protocol described in Section II A on two representative qubits, QC qubit (6) of `ibm_lagos` and QA qubit (305) of `DW_2000Q_LANL`.
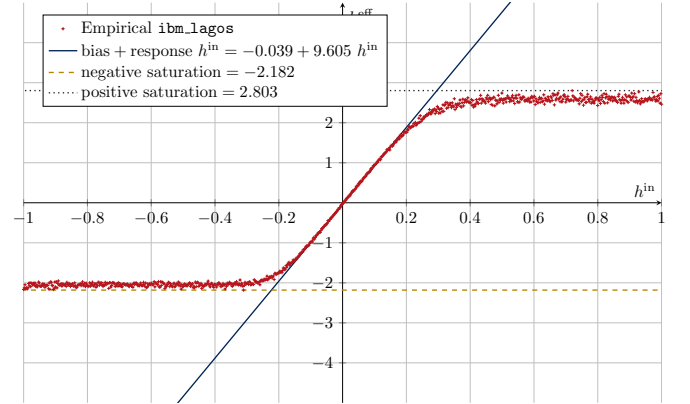
(b) A visual representation of the 4 features of interest to this work. The linear function f is obtained by performing a linear regression on the h^{eff} data in the interval $h^{\text{in}} \in [-0.1, 0.1]$. The g (respectively h) function is obtained by computing the minimum and maximum h^{eff} values over the whole range $h^{\text{in}} \in [-1, 1]$.

FIG. 2: Comparison of the response values recovered from the Q-RBPN protocol on IBM-Q QC and D-Wave QA hardware. 900 and 81 evenly spaced h^{in} values are shown for QC and QA respectively.

III. FULL-CHIP PARAMETER COMPARISONS

The ultimate goal of this work is to develop scalable qubit performance metrics that can be used across both QC and QA computing platforms. To that end, we observe that the single-qubit protocol discussed in the previous section can be executed in parallel for every qubit in QC and QA hardware devices. Consequently, we propose the Qubit Response, Bias, Positive saturation, Negative saturation (Q-RBPN) protocol that runs the $h^{\text{in}}/h^{\text{eff}}$ analysis in parallel characterizes the distribution of response, bias, positive/negative saturation, across an entire quantum computing hardware device. This protocol enables the comparison of a large numbers of qubits spanning a range of quantum hardware to highlight the

consistency of qubits across systems, not just individual qubit comparisons.

To explore the potential of the Q-RBPN protocol, this work analyzes 2300 qubits spanning 18 QC platforms in IBM's Q-Hub (311 qubits in total) and a D-Wave 2000Q Quantum Annealer located at Los Alamos National Laboratory (2031 qubits), known as DW_2000Q_LANL. A summary of all of the quantum computers considered in this work and their Q-RBPN results is available in Table I.

On the IBM-Q hardware, the following data collection settings were used: the number of shots performed is 8192, and the 1-qubit circuits have all been compiled with a custom transpiler pass that guarantees that the result of the compilation will always include 5 hardware-native gates, which covers the most general case for a single qubit. This custom transpilation pass is needed to avoid certain special-cases where the 5 gate decomposition can be simplified to 3 gates or even a single R_z gate, which would distort the results obtained by the protocol as executing fewer gates results in lower error rates. See Appendix B for additional details around the 5 gate decomposition used in this work.

On the D-Wave hardware the following data collection settings were used, *flux drift compensation* is disabled, which prevents automatic corrections to input fields based on a calibration procedure that is run a few times each hour; the *num reads* is set to 10000, specifying the number of identical executions performed for a single programming cycle of the chip; and the *annealing time* is set to 1 μs. For each h^{in} input value, the h^{eff} is estimated with 5×10^6 identical executions to ensure high accuracy of the h^{eff} estimation. The following sections discuss the results of running this Q-RBPN protocol on all 2300 qubits considered revealing some insights into the variations between these two quantum computing platforms.

A.   Parameter Distributions

Response: The response parameter recovered by the Q-RBPN protocol is arguably the most important value as it reflects the hardware accuracy in responding to a given input value and highlights the h^{in} parameter range where the hardware will exhibit the best performance. Figure 3 provides a summary of the response values recovered from the QA (top) and QC (bottom) hardware platforms. Both platforms provide a range of values for the response with mean and standard deviations of 10.03 $\pm$ 0.26 (QA) and 9.04 $\pm$ 0.88 (QC). These specific mean values are not of particular note as they can be adjusted in both platforms by changing how the quantum program is executed, see Eq. (4) and [16]. However, the variance value is interesting because it reflects the variability of the response value across a cohort of qubits that ideally would have identical response functions. These recovered variance values suggest that the QC qubits exhibit around twice the variability in the response than the QA
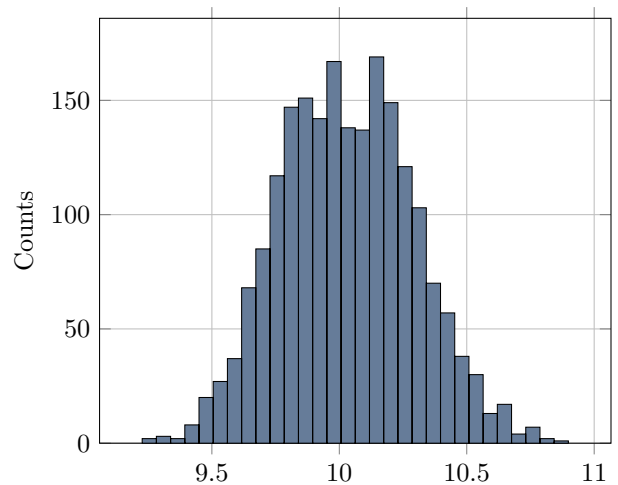
(a) Histogram of fitted response values on D-Wave ($n = 2031$).
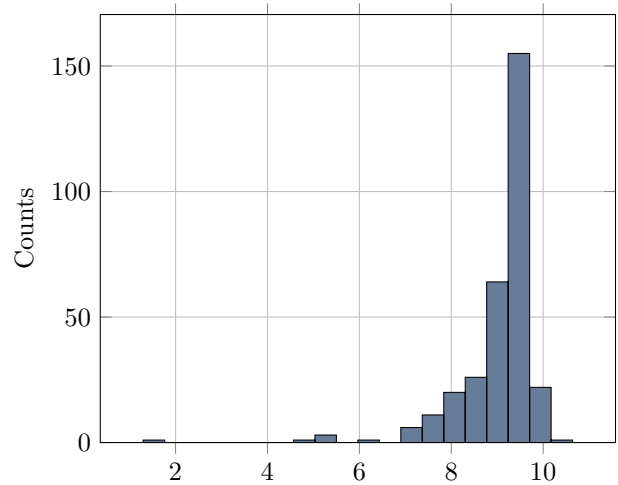
(b) Histogram of fitted response values on IBM-Q ($n = 311$).

FIG. 3: Comparison of the response values recovered from the Q-RBPN protocol on D-Wave and IBM-Q hardware.

qubits, on average. Given that averages over multiple QC chips may be misleading, we suggest reviewing the chip-by-chip response variance values in Table I for a more nuanced comparison where some QC hardware is able to match the QA response variability and some have notably higher values.

Bias: The bias parameter recovered by the Q-RBPN protocol is arguably the least important value as can easily be corrected by prepossessing the input data. However, its magnitude and variability do provide a measure of the calibration accuracy of the qubits. Figure 4 provides a summary of the bias values recovered from the QA (top) and QC (bottom) hardware platforms. Both platforms provide a fairly narrow range of bias values with mean and standard deviations of 0.02 $\pm$ 0.07 (QA) and
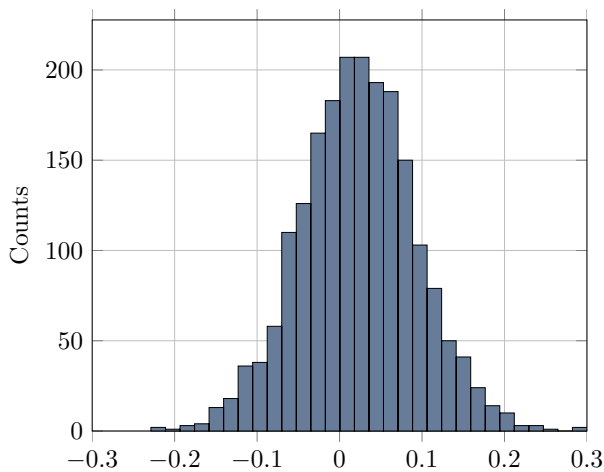
(a) Histogram of the computed biases for all the qubits on the D-Wave chip ($n = 2031$).
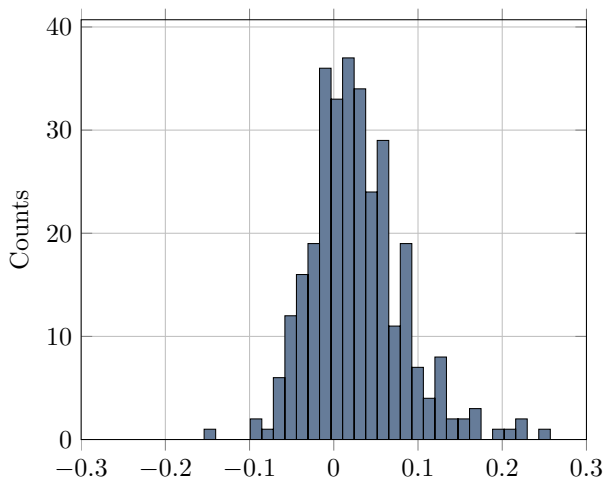
(b) Histogram of the computed biases for all the qubits on all the IBM-Q chips ($n = 311$).

FIG. 4: Comparison of the bias distribution between the quantum-annealing model of computation with D-Wave chip and the gate-based model of computation with IBM-Q hardware. For consistency both plot's axis range of -0.3 to 0.3, which excludes three outliers from the D-Wave dataset.

0.03 ± 0.06 (QC). Although the two histograms appear to be distinct, computation of the mean and standard deviation reveals that their performance is remarkably similar. This level of similarity is notable given the desperate implementations of these two types of quantum devices. One possible explanation is that both platforms are built on super conducting loops operating a milli-Kelvin temperatures; possibly this underlying technology is the limiting the calibration accuracy of these qubits. Executing the Q-RBPN protocol on a quantum hardware platform that uses a different foundational technology (e.g., trapped ions) could provide additional insights into this hypothesis and is left for future work.

Positive / Negative Saturation: The saturation values recovered by the Q-RBPN protocol provide useful information on the range of input parameters that can be realized in practice. For example, notice in Fig. 2a that there is no effective difference in h^{eff} for the range of $|h^{\text{in}}| \in [0.6, 1.0]$, suggesting a programmable range of $h^{\text{in}} \in [-0.4, 0.4]$ would be more suitable if the user would like to the hardware respond consistently to different h^{in} values.

Figure 5 provides a summary of the saturation values recovered from the QA (top) and QC (bottom) hardware platforms. This parameter highlights the biggest differences in both of these hardware platforms. A first observation is that QA hardware realizes notably larger saturation values (≈ 5.0) than the QC hardware (≈ 2.5). This result suggests a larger programmable range for the QA hardware as compared to the QC hardware. Note that using the protocol presented in Section II to realize large values of h^{eff} requires the QC hardware to resolve very fine angles (i.e., to distinguish values of $h^{\text{eff}} \geq 5$ requires an angle resolution below 0.01 radians), presenting a significant test of the hardware's control system. A second observation is that the variation of saturation points of the QA hardware (≈ 0.15) tend to be more consistent that the QC hardware (≈ 0.30). Finally, these results highlight that the asymmetry of the saturation values in the QC hardware that are shown in Fig. 2a are a fairly consistent feature of the IBM-Q hardware. We suspect this asymmetry is related to operational challenges that encourage the hardware to prefer one state over another; however, a comprehensive test of this hypothesis is left for future work. For application developers, knowledge of this QC hardware asymmetry is useful as it can inspired different treatment of negative and positive h^{in} values. Overall, these saturation metrics provide a clear distinction between the qubits in the QA and QC hardware considered here and highlights one feature for possible improvement in QC hardware.

IV.   DISCUSSION & CONCLUSION

In the era of NISQ devices, QCVV plays a valuable role in measuring and tracking the fidelity of quantum hardware platforms. However, QCVV protocols that can measure and compare the performance of both QC and QA hardware have been limited. In this work, we proposed the Q-RBPN protocol to fill this gap and provide a simple and scalable single-qubit performance comparison across different quantum hardware platforms. We hope that the four metrics proposed in this work – response, bias, positive/negative saturation – will provide valuable tools for tracking the progress of a wide range of quantum hardware platforms over time, irrespective of the computational model under consideration.

In addition to tracking qubit performance, the metrics proposed in this work also have interesting implications for specific applications of quantum computing. In

(a) Histogram of the negative and positive saturation values on D-Wave ($n = 2031$).

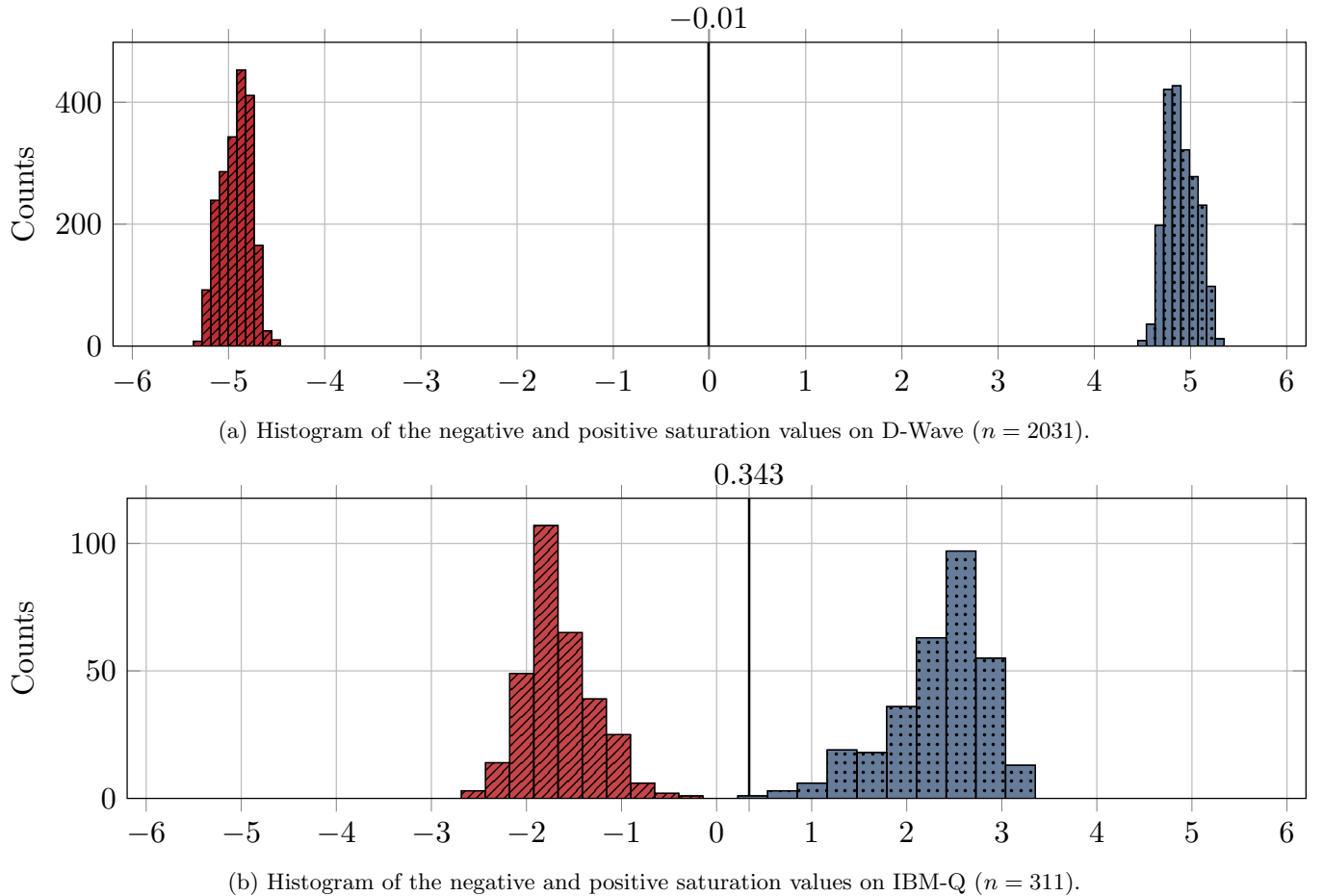(b) Histogram of the negative and positive saturation values on IBM-Q ($n = 311$).

FIG. 5: Comparison of the distributions of negative and positive saturation for D-Wave hardware (above) and IBM-Q hardware (below). The vertical black line on each plot corresponds to the mean over all the collected data for the chip considered. The D-Wave mean of -0.01 indicates that the negative/positive saturation values are distributed in a nearly symmetric fashion. The 0.343 mean value for IBM-Q hardware indicates that the negative/positive saturation distributions are biased tower positive values.

the context of quantum accelerated Boltzmann sampling, the saturation analysis in this work provides insights into what types of distribution parameters can be accurately represented on a given hardware device. If we consider calibration of variational quantum algorithms with hardware in the loop, these metrics provide bounds on the parameter values where the hardware has a good response to the input functions. In the context of quantum accelerated optimization, the programmer often has freedom to rescale the problem parameters. The saturation values provided by Q-RBPN indicate the optimal problem scaling on a given hardware platform, which can increase the accuracy of cross-platform optimization performance comparisons.

Finally, this work has revealed some key directions for future investigation. The first direction is to execute the Q-RBPN protocol on a wider range of quantum hardware platforms [2]. Ion trap based QC [5] and emerging QA [33] hardware platforms are natural next steps to under-

stand how the proposed metrics vary across additional hardware realizations. A second direction is to better understand what are the underlying mechanisms that result in non-ideal behavior of qubits in the Q-RBPN protocol. Such insights could improve the interpretability of the Q-RBPN metrics, introduce additional measures of qubit performance and potentially point to fundamental limits on the performance of the underlying hardware.

V. ACKNOWLEDGEMENTS

Chip Name	Qubits	Response	Bias	Negative S.	Positive S.
DW_2000Q_LANL	2031	10.03 ± 0.26	0.02 ± 0.07	-4.92 ± 0.15	4.90 ± 0.16
all ibmq backends	311	9.04 ± 0.88	0.03 ± 0.06	-1.65 ± 0.37	2.33 ± 0.53
ibmq_brooklyn	65	9.21 ± 0.57	0.02 ± 0.05	-1.68 ± 0.29	2.38 ± 0.32
ibmq_manhattan	65	8.95 ± 0.86	0.03 ± 0.06	-1.60 ± 0.40	2.24 ± 0.56
ibmq_montreal	27	9.23 ± 0.57	0.02 ± 0.06	-1.80 ± 0.33	2.47 ± 0.48
ibmq_mumbai	27	8.96 ± 0.60	0.01 ± 0.03	-1.61 ± 0.25	2.06 ± 0.47
ibmq_sydney	27	8.73 ± 1.58	0.04 ± 0.05	-1.54 ± 0.36	2.25 ± 0.68
ibmq_toronto	27	8.80 ± 1.21	0.04 ± 0.05	-1.67 ± 0.48	2.35 ± 0.72
ibmq_guadalupe	16	9.42 ± 0.16	0.03 ± 0.05	-1.75 ± 0.16	2.75 ± 0.29
ibmq_casablanca	7	9.38 ± 0.28	0.04 ± 0.03	-1.73 ± 0.23	2.71 ± 0.22
ibmq_jakarta	7	9.19 ± 0.31	0.03 ± 0.06	-1.65 ± 0.26	2.32 ± 0.27
ibm_lagos	7	9.79 ± 0.41	-0.02 ± 0.03	-2.32 ± 0.36	2.50 ± 0.23
ibmq_belem	5	9.09 ± 0.35	0.10 ± 0.06	-1.43 ± 0.22	2.39 ± 0.31
ibmq_bogota	5	9.08 ± 0.68	0.05 ± 0.04	-1.58 ± 0.39	2.53 ± 0.33
ibmq_lima	5	9.08 ± 0.51	0.10 ± 0.08	-1.45 ± 0.38	2.78 ± 0.18
ibmq_manila	5	9.38 ± 0.06	0.00 ± 0.02	-1.68 ± 0.06	2.35 ± 0.29
ibmq_quito	5	9.05 ± 0.46	0.08 ± 0.03	-1.38 ± 0.22	2.48 ± 0.44
ibmq_santiago	5	8.95 ± 0.50	0.05 ± 0.02	-1.57 ± 0.33	2.15 ± 0.50
ibmqx2	5	7.31 ± 1.03	-0.04 ± 0.08	-1.16 ± 0.24	1.30 ± 0.44
ibmq_armonk	1	8.23 ± 0.00	0.01 ± 0.00	-1.28 ± 0.00	1.41 ± 0.00

TABLE I: A summary of the Q-RBPN metrics broken down by specific hardware devices including one D-Wave quantum annealing hardware and eighteen IBM-Q quantum computing hardware. The statistics of each of the Q-RBPN metrics are presented in the format [mean] $\pm$ [standard deviation].

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

[2] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, "Quantum computers," Nature **464**, 45–53 (2010).

[3] John Clarke and Frank K. Wilhelm, "Superconducting quantum bits," Nature **453**, 1031–1042 (2008).

[4] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose, "Quantum annealing with manufactured spins," Nature **473**, 194–198 (2011).

[5] S. Debnath, N. M. Linke, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, "Demonstration of a small programmable quantum computer with atomic qubits," Nature **536**, 63–66 (2016).

[6] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi, "Quantum certification and benchmarking," Nature Reviews Physics **2**, 382–390 (2020).

[7] K. Wright, K. M. Beck, S. Debnath, J. M. Amini, Y. Nam, N. Grzesiak, J.-S. Chen, N. C. Pisenti, M. Chmielewski, C. Collins, K. M. Hudek, J. Mizrahi, J. D. Wong-Campos, S. Allen, J. Apisdorf, P. Solomon, M. Williams, A. M. Ducore, A. Blinov, S. M. Kreikemeier, V. Chaplin, M. Keesan, C. Monroe, and J. Kim, "Benchmarking an 11-qubit quantum computer," Nature Communications **10**, 5464 (2019).

[8] Easwar Magesan, J. M. Gambetta, and Joseph Emerson, "Scalable and robust randomized benchmarking of quantum processes," Phys. Rev. Lett. **106**, 180504 (2011).

[9] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, "Randomized benchmarking of quantum gates," Phys. Rev. A **77**, 012307 (2008).

[10] Erik Nielsen, John King Gamble, Kenneth Rudinger, Travis Scholten, Kevin Young, and Robin Blume-Kohout, "Gate set tomography," (2020), arXiv:2009.07301.

[11] Andrew W. Cross, Lev S. Bishop, Sarah Sheldon, Paul D. Nation, and Jay M. Gambetta, "Validating quantum computers using randomized model circuits," Phys. Rev. A **100**, 032328 (2019).

[12] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J. Bremner, John M. Martinis, and Hartmut Neven, "Characterizing quantum supremacy in near-term devices,"

Nature Physics **14**, 595–600 (2018).

[13] Jay Gambetta and Sarah Sheldon, "Cramming more power into a quantum device," Published online at https://www.ibm.com/blogs/research/2019/03/power-quantum-device/ (2019), accessed: 03/28/2021.

[14] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis, "Quantum supremacy using a programmable superconducting processor," Nature **574**, 505–510 (2019).

[15] D-Wave Systems Inc., "D-wave system documentation," Published online at https://docs.dwavesys.com/docs/latest/ (2020), accessed: 03/17/2021.

[16] Jon Nelson, Marc Vuffray, Andrey Y. Lokhov, and Carleton Coffrin, "Single-qubit fidelity assessment of quantum annealing hardware," IEEE Transactions on Quantum Engineering , 1–1 (2021).

[17] D-Wave Systems, "A head-to-head comparison of d-wave and rigetti qpus," (2018), 14-1025A-D.

[18] Madita Willsch, Dennis Willsch, Fengping Jin, Hans De Raedt, and Kristel Michielsen, "Benchmarking the quantum approximate optimization algorithm," Quantum Information Processing **19**, 197 (2020).

[19] Hayato Ushijima-Mwesigwa, Ruslan Shaydulin, Christian F. A. Negre, Susan M. Mniszewski, Yuri Alexeev, and Ilya Safro, "Multilevel combinatorial optimization across quantum architectures," ACM Transactions on Quantum Computing **2** (2021), 10.1145/3425607.

[20] Elijah Pelofske, John Golden, Andreas Bärtschi, Daniel O'Malley, and Stephan Eidenbenz, "Sampling on nisq devices: "who's the fairest one of all?"," (2021), arXiv:2107.06468.

[21] John J Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the national academy of sciences **79**, 2554–2558 (1982).

[22] Dileep Kumar Panjwani and Glenn Healey, "Markov random field models for unsupervised segmentation of textured color images," IEEE Transactions on pattern analysis and machine intelligence **17**, 939–954 (1995).

[23] Andrey Y Lokhov, Marc Vuffray, Sidhant Misra, and Michael Chertkov, "Optimal structure and parameter learning of ising models," Science advances **4**, e1700791 (2018).

[24] Gary Kochenberger, Jin-Kao Hao, Fred Glover, Mark Lewis, Zhipeng Lü, Haibo Wang, and Yang Wang, "The unconstrained binary quadratic programming problem: a survey," Journal of Combinatorial Optimization **28**, 58–81 (2014).

[25] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser, "Quantum computation by adiabatic evolution," (2018), arXiv:quant-ph/0001106.

[26] Sergey Bravyi, David P. Divincenzo, Roberto Oliveira, and Barbara M. Terhal, "The complexity of stoquastic local hamiltonian problems," Quantum Info. Comput. **8**, 361–385 (2008).

[27] Tadashi Kadowaki and Hidetoshi Nishimori, "Quantum annealing in the transverse ising model," Phys. Rev. E **58**, 5355–5363 (1998).

[28] Tameem Albash and Daniel A. Lidar, "Decoherence in adiabatic quantum computation," Phys. Rev. A **91**, 062320 (2015).

[29] Sergio Boixo, Vadim N. Smelyanskiy, Alireza Shabani, Sergei V. Isakov, Mark Dykman, Vasil S. Denchev, Mohammad H. Amin, Anatoly Yu Smirnov, Masoud Mohseni, and Hartmut Neven, "Computational multi-qubit tunnelling in programmable quantum annealers," Nature Communications **7**, 10327 (2016).

[30] Anatoly Yu Smirnov and Mohammad H Amin, "Theory of open quantum dynamics with hybrid noise," New Journal of Physics **20**, 103037 (2018).

[31] Marc Vuffray, Carleton Coffrin, Yaroslav A Kharkov, and Andrey Y Lokhov, "Programmable quantum annealers as noisy gibbs samplers," arXiv preprint arXiv:2012.08827 (2020).

[32] IBM Quantum Team, "IBM Quantum Systems panel," https://quantum-computing.ibm.com/services?services=systems (2021), last accessed: 2021-08-09.

[33] Jeffrey A. Grover, James I. Basham, Alexander Marakov, Steven M. Disseler, Robert T. Hinkey, Moe Khalil, Zachary A. Stegen, Thomas Chamberlin, Wade DeGottardi, David J. Clarke, James R. Medford, Joel D. Strand, Micah J. A. Stoutimore, Sergey Novikov, David G. Ferguson, Daniel Lidar, Kenneth M. Zick, and Anthony J. Przybysz, "Fast, lifetime-preserving readout for high-coherence quantum annealers," PRX Quantum **1**, 020314 (2020).

[34] Andrew W Cross, Lev S Bishop, John A Smolin, and Jay M Gambetta, "Open quantum assembly language," arXiv preprint arXiv:1707.03429 (2017).

[35] Lukasz Cincio, Kenneth Rudinger, Mohan Sarovar, and Patrick J. Coles, "Machine learning of noise-resilient quantum circuits," PRX Quantum **2**, 010324 (2021).
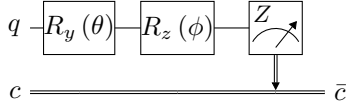
FIG. 6: A more general 1-qubit quantum program used to conduct the Q-RBPN protocol. The rotation angle ϕ is a free input parameter and the rotation angle θ is obtained from specified h^{in} values via Eq. (4).
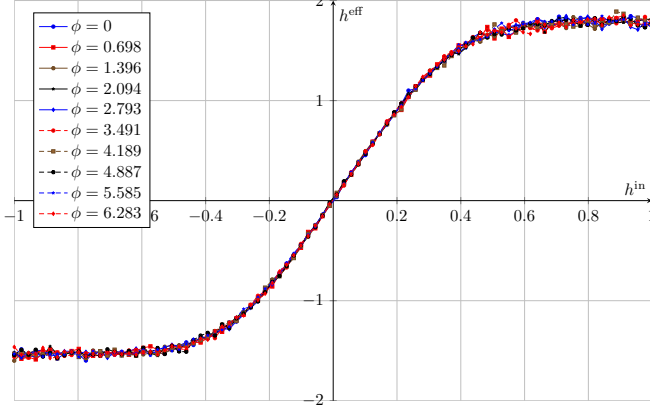

FIG. 7: Result with multiple different values for ϕ zoomed. Data gathered on qubit 0 of `ibm_lagos`. Plot obtained with a value of $\beta = 10$.

Appendix A: Alternative Quantum Gate Programs

The quantum circuit presented in Fig. 1 is suitable for implementing the Q-RBPN protocol developed in this work. However, this is not the only program that could have been used in QC context. In particular, this 1-parameter circuit is only able to explore states that are in the X-Z plane of the Bloch sphere. It is reasonable to postulate that more sophisticated QC programs may yield different results in the Q-RBPN protocol. One option for considering a more general QC program is to perform the same 1-parameter experiment but in another plane, for example, in the Y-Z plane. This process is accomplished by adding an $R_z(\phi)$ rotation that will rotate the plane in which the experiment is performed. For the Y-Z plane, $\phi = \frac{\pi}{2}$. The resulting quantum circuit is depicted in Fig. 6.

To explore the potential of performing the experiments in a different plane, we repeat the Q-RBPN protocol for a single qubit with 10 evenly spaced values of $\phi \in [0, 2\pi]$ and the results are presented in Fig. 7. These results are what is expected theoretically as an $R_z(\phi)$ gate applied just before the qubit measurement has no effect on the expected outcome of the measurement. In particular,

the $R_x(\theta)$ rotation is simply an $R_y(\theta)$ rotation followed by a $R_z\left(\frac{\pi}{2}\right)$ rotation. Based on the consistency of these results we conclude that the simpler quantum circuit presented in Fig. 1 is suitable for implementing the Q-RBPN protocol developed in this work.
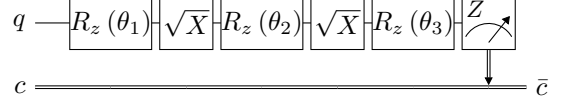

FIG. 8: The 1-qubit quantum program in the IBM-Q native gate set that was used in this work to perform any single qubit operation with a standardized program size.

Appendix B: Quantum Gate Program Consistency

Given the NISQ nature of available QC quantum hardware, the specific details of the quantum programs can have a notable impact on performance [34, 35]. When performing benchmarking and extracting metrics from several different quantum circuits, consistency is paramount; i.e., one does not want a few quantum circuits being treated differently than the others. In particular, it is important in the Q-RBPN protocol that all of the quantum circuits used to perform an experiment have the same number and type of quantum gates. This configuration is important because each gate introduces some amount of error into the computation [32], and, hence, executing smaller circuits changes overall error model. Running identical circuits reduces the potential biases that might emerge when different equivalent circuits are executed.

As this work is only concered with benchmarking individual qubits, we observed that there exists a 5 gate program (using IBM-Q native gates) that can implement any 1-qubit operation as illustrated in Fig. 8. This 5 gate program tends to also be the smallest program that can implement a given single-qubit operation; however, some care needs to be taken with special-cases that arise; for example, when the 1-qubit quantum gate can be compiled to a single $R_z(\theta)$ gate. In such cases, for consistency, we want the quantum circuit to still contain the 5 gates from Fig. 8, even though 1 gate would be enough. This result is achieved by implementing a specific transpilation pass for IBM's qiskit transpiler, to ensure that all 5-gates are generated for any desired single-qubit operation.