

Implementation of Gate Set Tomography on Quantum Hardware

Henrique Guimarães Silvério
henrique.silverio@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2019

Abstract

Finding and understanding errors in quantum devices is instrumental to their development. Among error characterization techniques, Gate Set Tomography (GST) stands as the most comprehensive protocol. In this work, we implement a working GST protocol to characterize single-qubit operations of selected IBM Q's quantum devices.

By benchmarking GST with simulated data, we find the optimal solutions to be degenerate, with the initial state and readout estimations changing correlatedly. To fix this, we propose a protocol modification which assumes a perfect initial state. The modified GST protocol produces accurate and consistent estimations, with precision limited mainly by sampling noise.

In the application of GST to IBM Q devices, we observe that the average single-qubit gate quality is close to the limit of GST's precision. The same no longer holds for the readout errors, whose larger values agree with IBM's calibrations. We also characterize a qubit's readout error over time and find that it often changes in the order of 1% – 2% per hour.

Keywords: Quantum computing, GST, qubit characterization, IBM Q

1. Introduction

In recent years, the field of quantum computation has benefited from substantial interest and investment, which has been translated into the development of the first widely available prototypes of quantum computers, with up to 20 qubits. However, these improvements are not yet enough to take full advantage of most of the quantum algorithms under development since the 1990s, in large part due to technological challenges. Further improving current quantum computers relies heavily on understanding the nature and origin of existing complications, to which extent an accurate method of characterization is essential. The main focus of this work will be the study, implementation and application to real devices¹ of state of the art characterization techniques.

1.1. The need for qubit characterization

In the development of quantum devices, testing the hardware's performance is always an essential part of the process. In a first iteration, this can be done through the computation of quantum algorithms whose outcomes are known; the magni-

tude of the deviations that appear then speak for the overall quality of the device. However, a minimal amount of information can be extracted using this naive approach. Thus, a more systematic approach to hardware testing is needed, more specifically one that tries to describe the individual elements of the device to predict its behaviour on all possible scenarios.

Qubit characterization provides rigorous tools for error diagnostics and assessment of the hardware's performance. The objective is to find out the sources of errors so that they can be mitigated, or to reliably set an upper bound on the error of an arbitrary algorithm's outcome. The different ways to approach qubit characterization can be roughly sorted by the amount of information they intend to extract from the procedure. Among the more popular methods, Randomized Benchmarking (RB) [7] and Gate Set Tomography (GST) [2] stand at opposite ends of this spectrum.

RB is a process in which gate sequences of different lengths are applied and the outcome compared with the predicted value. By doing this, one attains a measurement of the average process fidelity, which can then be used to extract the average gate fidelity [6]. Additionally, this process

¹Namely the IBM Quantum Experience (IBM Q) quantum processors.

can be extended to multi-qubit operations with only polynomial scaling with the number of qubits [10]. In contrast with other techniques, RB is simple, robust and gives reliable average gate fidelities for multi-qubit processes that are useful in ascertaining the overall quality of the gates in a quantum device. However, it can only provide the average fidelity of specific gates, which makes it unsatisfactory as a standalone characterization technique.

On the other hand, GST tries to achieve a complete characterization of the system. With GST, the goal is to achieve a description of the hardware that correctly predicts the outcome of any circuit. This feature makes GST much more informative than RB, but has the serious drawbacks of being costly in resources and unpractical for multiple qubits.

In this work, the method of choice for the characterization of IBM's quantum processors was GST, for several reasons: First and foremost, because we were interested in achieving a description of the system that was as detailed as possible. Secondly, due to the fact that RB has already been extensively applied to IBM's systems, while GST results remain, to our knowledge, either undone or unpublished. Lastly, because the IBM Q is remotely accessible to the scientific community and has thus been subjected to extensive testing, there are multiple reports of its poor performance when compared with simulation results [5]. These reports made clear that understanding what is wrong and trying to improve the current devices is a pressing issue, for which purpose GST is a far more useful tool than RB.

1.2. Objectives

The principal focus of this work was to implement a working GST protocol capable of fully characterizing single-qubit operations of present-day quantum devices and testing its performance in both simulated and real experimental data. Nonetheless, rather than debugging IBM Q's systems in particular, our main intention was to explore GST's general capabilities, to uncover potential limitations and to establish the extent of its applicability and usefulness in experimental settings.

2. The principles of GST

2.1. Quantum Circuit Model

The computation model used describes quantum algorithms as circuits, in which a sequence of quantum gates (i.e. operators) act on qubits (i.e. quantum states). Any quantum computation can then be divided into three stages: *state preparation*, *conditional manipulation* and *state readout*. As an example, figure 1 depicts a typical quantum circuit, subdivided into these three stages.

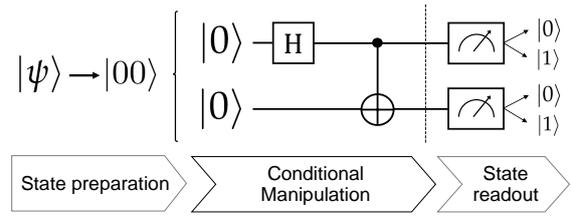


Figure 1: Diagram representing the stages of a quantum computation, with the example of a circuit that prepares the maximally entangled state $(|00\rangle + |11\rangle)/\sqrt{2}$. The state preparation represents the transformation of an arbitrary quantum state onto the $|00\rangle$ state and is usually omitted in quantum circuits. The circuit itself shows the implementation of a Hadamard gate on the first qubit, which creates an equal superposition state on that qubit, followed by a CNOT gate that flips the state of the second qubit conditionally on the state of the first. In the end, each qubit is individually measured in the computational basis, returning one of two possible basis states.

2.2. Operations with Density Matrices

In this work, we adopt the density matrix formalism, which represents a system as an ensemble of quantum states with probability p_i , $\{p_i, |\psi_i\rangle\}$. It is defined as an operator

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i| \quad (1)$$

A density matrix, ρ , represents a physical system if $\text{Tr}[\rho] = 1$, $\rho = \rho^\dagger$ and ρ is semi-definite positive (i.e. $\rho \succeq 0$) [14]. To describe changes to a system, we use quantum operations, which are defined as maps acting on the system's density matrix, $\rho \rightarrow \Lambda(\rho)$. Maps that describe physical processes must meet two basic requirements [6]:

1. Trace preservation: If $\text{Tr}[\rho] = 1$, then $\text{Tr}[\Lambda(\rho)] = 1$
2. Complete Positivity: For a composite system $\rho \succeq 0$, then $(\Lambda(\cdot) \otimes I)\rho \succeq 0$, where Λ only acts on part of ρ .

For obeying the conditions above, a physical map is often referred to as a CPTP (completely positive trace-preserving) map. In fact, any physical map can be conveniently written in its *Kraus form* [6]:

$$\Lambda(\rho) = \sum_{i=1}^N K_i \rho K_i^\dagger \quad (2)$$

where $N \leq d^2$, with $d = 2^n$ being the Hilbert space dimension of the n -qubit space. The K_i are the Kraus operators and they need only fulfill the completeness condition, $\sum_{i=1}^N K_i^\dagger K_i = I$.

2.3. Superoperator Formalism

Despite its generality, the quantum operations formalism lacks the ease of manipulation the Dirac

notation offers the state vector formalism. We circumvent this issue with the superoperator formalism, which represents density operators ρ (of dimension $d \times d$) in the Hilbert space of dimension d as vectors $|\rho\rangle\rangle$ in the *Hilbert-Schmidt* space of dimension d^2 . In turn, quantum operations now consist of $d^2 \times d^2$ matrices in this space. We define the Hilbert-Schmidt inner product of two density operators A and B as

$$\langle\langle A|B\rangle\rangle = \text{Tr}[A^\dagger B] \quad (3)$$

This formalism also supports the representation of some observable, E , since

$$\langle\langle E\rangle\rangle = \text{Tr}[E\rho] = \langle\langle E|\rho\rangle\rangle \quad (4)$$

Therefore, a generic observable E is represented as $\langle\langle E|\rceil$ in the superoperator formalism.

Choice of basis

Although superoperators can be written in any basis of \mathcal{H}_n , a particularly convenient choice is the normalized Pauli basis. This is a non-standard normalization in which the basis elements belong to the set

$$\sigma_k \in \left\{ I/\sqrt{2}, X/\sqrt{2}, Y/\sqrt{2}, Z/\sqrt{2} \right\}^{\otimes n} \quad (5)$$

with $\{I, X, Y, Z\}$ being the standard Pauli matrices. Let us denote the basis elements in matrix form as σ_k and their superoperator counterparts as $|k\rangle\rangle$. In this basis, a general density operator A is decomposed as

$$A = \sum_k a_k \sigma_k \quad (6)$$

where $a_k = \text{Tr}[\sigma_k A] = \langle\langle k|A\rangle\rangle$ are the coefficients of the vector $|A\rangle\rangle$. With the usage of the normalized Pauli basis, one ensures that the basis states in the superoperator formalism obey the completeness relation $\sum_k |k\rangle\rangle\langle\langle k| = \sum_k \sigma_k \text{Tr}[\sigma_k \cdot] = I$.

Pauli Transfer Matrix Representation

The Pauli Transfer Matrix (PTM) representation allows us to define quantum operations as matrices in the superoperator formalism. For an arbitrary quantum operation Λ , there will be an associated PTM R_Λ such that

$$R_\Lambda = \sum_{jk} |j\rangle\rangle\langle\langle j|R_\Lambda|k\rangle\rangle\langle\langle k| \quad (7)$$

$$\langle\langle j|R_\Lambda|k\rangle\rangle = \text{Tr}[\sigma_j \Lambda(\sigma_k)] = (R_\Lambda)_{jk} \quad (8)$$

Later on, we represent a PTM, R_Λ , solely by the quantum operation, Λ , that it describes.

These definitions lead naturally to (see Section 2.2 of Ref.[6]):

$$|\Lambda(\rho)\rangle\rangle = R_\Lambda|\rho\rangle\rangle \quad (9)$$

$$R_{\Lambda_1 \circ \Lambda_2} = R_{\Lambda_1} R_{\Lambda_2} \quad (10)$$

Thus, we see that the superoperator formalism permits a quantum operation acting on a density matrix to be represented as a PTM acting on a state vector. A sequence of quantum operations can also be represented as a simple multiplication of the respective PTMs.

Physicality Constraints

A physical quantum operation maps a density matrix ρ to another density matrix $\Lambda(\rho)$. This implies that $\Lambda(\rho)$ can be decomposed in the normalized Pauli basis with real coefficients in the interval $[-1/\sqrt{d}, 1/\sqrt{d}]$. Considering the PTM's parameters definition in eq. (8), one concludes that all entries must fulfil

$$(R_\Lambda)_{ij} \in [-1, 1] \quad (11)$$

Furthermore, a quantum operation needs to be a CPTP map in order to be a representation of a physical system. Therefore, we must constrain our PTM's parameters in a way that ensures the represented operation is both trace-preserving and completely positive.

Trace preservation: Taking into account that the Pauli matrices are traceless (i.e. $\text{Tr}[\sigma_i] = \sqrt{d}\delta_{0i}$, since σ_0 is the normalized identity matrix), a trace-preserving map must verify $\text{Tr}[\Lambda(\sigma_i)] = \sqrt{d}\delta_{0i}$, which in turn implies that $(R_\Lambda)_{0i} = \text{Tr}[\sigma_0 \Lambda(\sigma_i)] = \delta_{0i}$.

Complete Positivity: Ensuring a given PTM represents a CP quantum operation is, unfortunately, not so straightforward. To impose this condition, we rely on the fact that the Choi-Jamiolkowski matrix, ρ_Λ , associated with a completely-positive map is positive semi-definite and takes the form [12]:

$$\rho_\Lambda = \frac{1}{d} \sum_{ij} (R_\Lambda)_{ij} \sigma_j^T \otimes \sigma_i \quad (12)$$

Therefore, one needs to impose $\rho_\Lambda \succeq 0$, meaning ρ_Λ is an hermitian matrix with non-negative eigenvalues.

2.4. Qubit Characterization

We shall see that Gate Set Tomography (GST) [2] is the only reliable method of characterization in the presence of state preparation and measurement (SPAM) errors. However, in order to understand GST, one should start with its predecessors: Quantum State Tomography (QST) [8] and Quantum Process Tomography (QPT) [4].

QST consists of determining the complete density matrix of an unknown quantum state. Like every other characterization procedure, QST requires access to a large ensemble of identical states — obtained, for instance, through repetition of the same experiment. Given this ensemble, the expectation value of different observables (usually the Pauli matrices) can be determined by repeating the same measurement and averaging over the results.

QPT is similar to QST, though the goal is now to characterize some unknown quantum process instead of a quantum state. QPT requires repetitive access to the process it wishes to describe; this process can be regarded as a *quantum black box*, to which one can feed different inputs (ρ_{in}) and read the corresponding outputs (ρ_{out}). The end goal is then to construct the quantum operation that best maps each ρ_{in} to the measured ρ_{out} , which in turn requires QST to be performed on every output.

By default, QPT does not take into account possible errors in the preparation of the input states, nor in the measurement of the corresponding output; the protocol assumes that the input and output states are known exactly, while this is usually not the case. In fact, most of the states are prepared using gates also included within the process QPT is supposed to characterize [2]. In this way, QPT makes an incorrect assumption about the system that creates a self-consistency issue, ultimately tainting the obtained results.

GST offers a solution to this problem by including the faulty gates in the estimation itself. In contrast with QPT, GST needs to perform, simultaneously, an estimation of an entire set of gates and states

$$\{|\rho\rangle\rangle, \langle\langle E|\mathcal{G}\} \quad (13)$$

where $|\rho\rangle\rangle$ is the initial state, $\langle\langle E|$ is a two-outcome measurement and $\mathcal{G} = \{G_k\}$ is the *gate set*, which contains the quantum operations performed by the system.

The first step in performing GST is the choice of the gate set, \mathcal{G} , to be characterized, which needs to permit the system's preparation and measurement in an informationally complete basis [3]. This can be achieved solely through individual gates in \mathcal{G} , or through a composition of these. In either case, we define a *state preparation and measurement* (SPAM) gate set, \mathcal{F} , whose gates are used to obtain all the starting states $|\rho_i\rangle\rangle = F_i|\rho\rangle\rangle$ and measurement projectors $\langle\langle E_j| = \langle\langle E|F_j$. The other requirement is that both \mathcal{G} and \mathcal{F} include the "null" gate (represented as $\{\}$) which is, by definition, an operation that does nothing for no time. The standard example of a minimal gate (i.e. one with as few gates as possible) for a single-qubit system set

is then

$$\mathcal{G} = \{G_0, G_1, G_2\} = \{\{\}, X_{\frac{\pi}{2}}, Y_{\frac{\pi}{2}}\} \quad (14)$$

where $X_{\frac{\pi}{2}}$ and $Y_{\frac{\pi}{2}}$ represent a $\frac{\pi}{2}$ rotation around the x or y axis, respectively. The above stated requisites are verified if one can write all the required SPAM gates using gates within \mathcal{G} . In this case, one has

$$\mathcal{F} = \{F_1, F_2, F_3, F_4\} = \{\{\}, X_{\frac{\pi}{2}}, Y_{\frac{\pi}{2}}, X_{\frac{\pi}{2}}X_{\frac{\pi}{2}}\} \quad (15)$$

With \mathcal{G} and \mathcal{F} defined, one then conducts all the experiments and organizes their results in a tensor of the form

$$p_{ijk} = \langle\langle E|F_iG_kF_j|\rho\rangle\rangle \quad (16)$$

where $G_k \in \mathcal{G}$ and $F_i, F_j \in \mathcal{F}$. Considering for instance that $E = |0\rangle\langle 0|$, each experiment should be run N times and count the number of trials a $|0\rangle$ state was measured — here represented as $[\#|0\rangle]_{ijk}$; with this, the experimental probability can then be computed as $p_{ijk} = \frac{[\#|0\rangle]_{ijk}}{N}$. As a consequence of the sampling being finite, each expectation value will have an accompanying sampling error, $\sigma_{ijk} = \sqrt{p_{ijk}(1-p_{ijk})/N}$. In possession of all the experimental data, one then moves on to the gate set estimation phase.

2.4.1 Linear Inversion Gate Set Tomography (LGST)

Having the tensor of experimental points, one can start by reorganizing the information it contains. Introducing the completeness relation in the Pauli basis, one reaches

$$p_{ijk} = \sum_{rs} \langle\langle E|F_i|r\rangle\rangle \langle\langle r|G_k|s\rangle\rangle \langle\langle s|F_j|\rho\rangle\rangle \quad (17)$$

$$= \sum_{rs} A_{ir}(G_k)_{rs} B_{sj} \quad (18)$$

$$= (AG_kB)_{ij} \quad (19)$$

With this construction, we then define

$$\tilde{G}_k = AG_kB \quad (20)$$

$$g = \tilde{G}_0 = AB \quad (21)$$

$$|\tilde{\rho}\rangle\rangle = A|\rho\rangle\rangle = \sum_i |i\rangle\rangle \langle\langle E|F_i|\rho\rangle\rangle \quad (22)$$

$$\langle\langle \tilde{E}| = \langle\langle E|B = \sum_j \langle\langle E|F_j|\rho\rangle\rangle \langle\langle j| \quad (23)$$

Note that we can estimate g — and thus isolate the product of matrices A and B — because we imposed $G_0 = \{\}$ when choosing the gate set. Also,

noting that $|\hat{\rho}\rangle\rangle$ and $\langle\langle\hat{E}|$ are component-wise identical, the following relations can be established with the acquired data:

$$(\tilde{G}_k)_{ij} = \langle\langle E|F_i G_k F_j|\rho\rangle\rangle \quad (24)$$

$$g_{ij} = \langle\langle E|F_i F_j|\rho\rangle\rangle \quad (25)$$

$$|\hat{\rho}\rangle\rangle_i = \langle\langle E|F_i|\rho\rangle\rangle = \langle\langle\tilde{E}|_i \quad (26)$$

One then simply inverts g to get $g^{-1} = B^{-1}A^{-1}$ and achieve the following estimations

$$\hat{G}_k = g^{-1}\tilde{G}_k = B^{-1}G_k B \quad (27)$$

$$|\hat{\rho}\rangle\rangle = g^{-1}|\hat{\rho}\rangle\rangle = B^{-1}|\rho\rangle\rangle \quad (28)$$

$$\langle\langle\hat{E}| = \langle\langle\tilde{E}| = \langle\langle E|B \quad (29)$$

This displays the gauge freedom in terms of B present in the results, which with the available experimental data, we cannot measure. Moreover, take into account that $B_{sj} = \text{Tr}[\sigma_s F_j(\rho)]$ cannot be the identity, as this would require $F_j(\rho) = \sigma_s$, which is not a CPTP map.

One can, nonetheless, try to devise some methods of partially circumventing this gauge freedom problem. The most obvious one is to assume that the initial state and the SPAM gates are perfect, thus enabling the explicit calculation of B , which we call the *exact gauge*. A more refined approach is to try to find the gauge that results in the closest possible gate set to the ideal case [6], which we call the *optimized gauge*. None of these result in an accurate estimation, but they produce estimations that serve as sufficiently good starting points for the next step, Maximum Likelihood Estimation (MLE).

2.4.2 Maximum Likelihood Estimation (MLE)

MLE surpasses both shortcomings of LGST: for one, it does not suffer from LGST's gauge freedom problem; secondly, it enforces the physicality constraints that LGST cannot. The MLE procedure employed is reasonably standard: one finds the adequate parameters that yield expected results compatible with the experimental data. This is achieved through the minimization of a cost function, subjected to the physicality constraints. To start, one constructs the probability estimators

$$\hat{p}_{ijk}(\vec{t}) = \langle\langle\hat{E}(\vec{t})|\hat{F}_i(\vec{t})\hat{G}_k(\vec{t})\hat{F}_j(\vec{t})|\hat{\rho}(\vec{t})\rangle\rangle \quad (30)$$

where each quantity with a "hat" represents an estimation, as a function of the free parameters in \vec{t} . Then, one minimizes the difference between the experimental points and the estimators through the minimization of a cost function. In this work, we found a standard least-squares cost function of the

form

$$LS(\vec{t}) = \sum_{ijk} (p_{ijk} - \hat{p}_{ijk}(\vec{t}))^2 \quad (31)$$

produced the best results. The minimization itself should use a global optimization technique, as the cost function has multiple local minima. In this work, we use the Differential Evolution (DE) optimization method [15, 16], which also allows us to enforce the physicality conditions

$$(R_\Lambda)_{ij} \in [-1, 1], \quad \forall \Lambda \in \mathcal{G} \quad (32)$$

$$\rho \succeq 0 \quad (33)$$

$$E \succeq 0 \quad (34)$$

$$I - E \succeq 0 \quad (35)$$

$$\rho_\Lambda = \frac{1}{d} \sum_{ij} (R_\Lambda)_{ij} \sigma_j^T \otimes \sigma_i \succeq 0, \quad \forall \Lambda \in \mathcal{G} \quad (36)$$

with relative ease.

Once the cost function minimum is reached, we use the fact that the gradient at that point is null to propagate the sampling errors and thus determine the uncertainties of the free parameters in \vec{t} . This amounts to calculating a matrix $D = -NM^{-1}$, whose entries are element-wise defined as

$$D_{mj} = \frac{\partial t_j}{\partial p_m} \quad (37)$$

$$M_{ji} = \frac{\partial^2}{\partial t_j \partial t_i} C(t_k, p_l) \quad (38)$$

$$N_{mi} = \frac{\partial^2}{\partial p_m \partial t_i} C(t_k, p_l) \quad (39)$$

where $C(t_k, p_l)$ is the cost function, t_k are the free parameters and p_l the flattened data tensor p_{ijk} , with the index l running through all the entries. N and M are calculated numerically to obtain D , from which the uncertainties are calculated as

$$\vec{\varepsilon} = \sqrt{(D^T)^{\circ 2} \vec{\sigma}^{\circ 2}} \quad (40)$$

where $\vec{\varepsilon}$ and $\vec{\sigma}$ are the errors in \vec{t} and \vec{p} written as column vectors, and the symbol in the exponents stands for the *Hadamard power*, which exponentiates each entry of a vector or matrix element-wise.

2.5. State of the art GST

The above-given description of GST represents a primitive version of the method, which its creators have since improved [3]. The state of the art GST implementation surpasses its predecessor in accuracy, precision, speed and capability. Such improvements are achieved through a modified data-fitting protocol, in which new data points are included iteratively. These points come from increasingly long sequences of gates, specially constructed to amplify existing errors in the gates.

Despite its superior performance, this process requires a severe increase in the size of the data set. In our case, we are restricted by the number of experiments that IBM Q's devices can run, which impedes the acquisition of all the required data points. For this reason, we have to resort to a more primitive version of GST that, on the flip side, can work with more modest data sets.

2.6. Error quantification

The error of an operation quantifies its distance from its ideal form. In this work, we use the diamond norm

$$\|A - B\|_{\diamond} = \frac{1}{2} \max_{\rho} \text{Tr} \left| (A \otimes I)[\rho] - (B \otimes I)[\rho] \right| \quad (41)$$

which measures the distance between the output states of two processes, A and B , maximized over all the possible input states. The distance between output states is quantified through the trace distance [14],

$$d(\rho, \sigma) = \frac{1}{2} \text{Tr} |\rho - \sigma| \quad (42)$$

where $|A| \equiv \sqrt{A^\dagger A}$ is the positive square root of $A^\dagger A$.

As an alternative to the diamond norm, we also resort to the widely used average gate fidelity, defined as [9, 13, 11]:

$$\overline{\mathcal{F}}_{A,B} = \frac{\text{Tr}[R_A^T R_B] + d}{d(d+1)} \quad (43)$$

where d is the system's dimension ($d = 2$ in our single-qubit case). While the diamond norm is an upper-bound on the gate error, the average gate fidelity is a measure of average gate quality, obtained via another distance measure, the fidelity [14]. Furthermore, what is commonly referred to as average gate error is simply $1 - \overline{\mathcal{F}}_{A,B}$.

3. Benchmarking GST

Before applying it to real data, we tested all aspects of GST's estimation procedure using simulated faulty data. Our set of meaningful quantities was $\{|\rho\rangle\rangle, \langle\langle E|, X_{\frac{\pi}{2}}, Y_{\frac{\pi}{2}}, I\}$, which consists of the minimal gate set and the identity gate. We took these quantities' ideal forms and introduced slight deviations consistent with reported physical errors in IBM Q's devices. This faulty gate set mimicked a real device, with which all the required experiments were performed to construct a data tensor like that of eq. (16). We also added sampling noise by simulating the finite sampling process that takes place in real devices. In this way, we obtained a set of data points while knowing the gates that generated it, and could thus gauge GST's ability to characterize a device correctly.

For LGST, we found, unsurprisingly, that the estimations did not obey the physicality conditions. Nonetheless, we compared the closeness to the exact results of estimations that used the exact and optimized gauges and found the exact gauge to be preferable when the initial state is expected to be perfect.

Proceeding to MLE, we first realized that, when accounting for identical sequences in p_{ijk} , the number of data points is smaller than the number of free fitting parameters. To correct this flaw, we extended the SPAM gate set to

$$\mathcal{F}' = \left\{ \{ \}, X_{\frac{\pi}{2}}, Y_{\frac{\pi}{2}}, X_{\frac{\pi}{2}} X_{\frac{\pi}{2}}, X_{\frac{\pi}{2}} X_{\frac{\pi}{2}} X_{\frac{\pi}{2}}, Y_{\frac{\pi}{2}} Y_{\frac{\pi}{2}} Y_{\frac{\pi}{2}} \right\} \quad (44)$$

With the resulting larger data set, we performed a series of MLE trials on the data without sampling noise to assess the method's accuracy and consistency. Having found large deviations in $\hat{\rho}$ and \hat{E} among trials, we looked for correlations between the deviations in these quantities using the Pearson correlation coefficient [17] as a quantifier. This analysis produced the results in figure 2, which shows the deviations in \hat{E} , $\hat{\rho}$ and \hat{I} are perfectly correlated. At the same time, none of these are correlated to deviations in the cost function value at convergence, Δ_{Cost} , which led us to conclude that they can vary in a way that compensates one another and does not influence the cost function value.

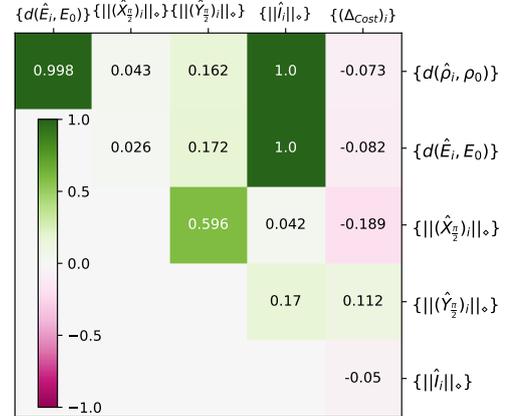


Figure 2: Pearson correlation coefficients between the figures of merit of a series of MLE trials, without sampling noise. Δ_{Cost} represents the difference between the cost function value at the end of the minimization and the cost function value for the exact solution. All values are obtained through comparison of the estimated quantities with the exact ones used in the creation of the simulated points — symbolically explicit when the trace distance is used (ρ_0 and E_0) or implicit in the diamond norm notation and in Δ_{Cost} .

The fact that \hat{E} , $\hat{\rho}$ and \hat{I} can vary concordantly and without affecting the cost function value meant that they can assume incorrect values. To eliminate these fluctuations, we fixed $\hat{\rho}$ and repeated

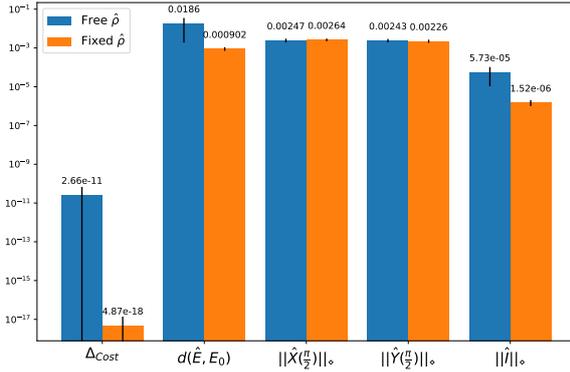


Figure 3: MLE performance with free and fixed initial states, without sampling noise. Presented here are the mean and standard deviation of the labeled quantities, extracted from a sample of ten estimations.

the MLE trials. The results in figure 3 show how this alteration improves the accuracy and consistency of the affected quantities, which further supports our hypothesis.

When MLE was applied to data with sampling noise, we found that our methods only worked systematically for around 25000 samples per data point — for inferior sample sizes, the sampling error was too large and caused the optimization algorithm to misconverge. Additionally, the cost function value at convergence was always lower than the one the exact result would yield, which implied that the algorithm managed to reach an estimation that described the noisy data better than the exact result.

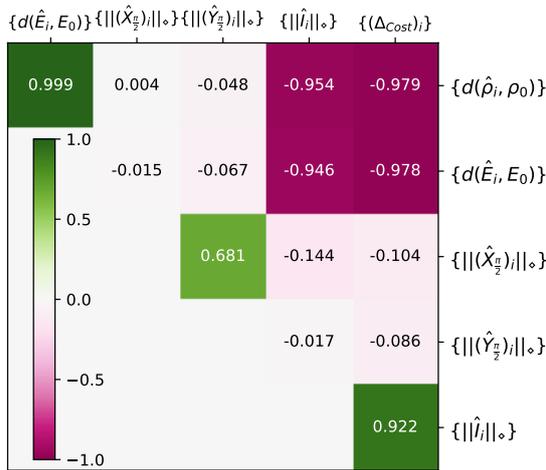


Figure 4: Pearson correlation coefficients between the figures of merit of a series of MLE trials, with sampling noise.

Overall, we observed that the results with sampling noise resembled those without it, apart from an expectable decrease in quality. The correlated behaviour between the deviations in $\hat{\rho}$ and \hat{E} was still present, though this time they were perfectly

anti-correlated with the deviations in \hat{I} and the cost function value at convergence, as depicted in figure 4. This behaviour meant that $\hat{\rho}$ and \hat{E} now deviated from their ideal values to accommodate the discrepancies induced by the sampling noise and thus reach lower cost function values. Also opposing previously observed behaviour is the fact that the identity gate now changed to accommodate the sampling noise discrepancies whenever \hat{E} and $\hat{\rho}$ did not, though it did so less successfully considering how it was positively correlated with Δ_{Cost} . As before, fixing $\hat{\rho}$ eliminated the correlations and improved the accuracy and consistency of the involved quantities.

4. Application to real devices

Armed with a better understanding of GST, we employed it to characterize IBM Q's Melbourne and Yorktown quantum processors. To do so, we gathered the necessary data through a series of quantum circuits, submitted remotely to run on IBM Q's devices via the IBM Q Experience website [1]. We used the same gate set as that of section 3, with which 92 distinct experiments of the form of eq. (16) were constructed. Each experiment was run consecutively three times in order to gather a large enough sample size. In total, 276 quantum circuits were run at the maximum number of shots (8192), taking around 40 minutes to collect the necessary data for a single qubit's characterization. This process was repeated in close succession for every qubit in each device so that the experimental conditions changed as little as possible.

Based on the observations of section 3, we fixed the initial state estimator in the analysis of real devices so that an adequate estimation of the readout error could be reached. This meant that we had to assume the qubits always started in their ground state instead of characterizing it.

4.1. Device characterizations

We performed complete characterizations of all selected devices' qubits. Although the quality of the characterizations varied among them, we found that even the best characterizations, of which the $X_{\frac{\pi}{2}}$ -gate estimation in figure 5 is an example, have errors which, though small, still encompass the respective ideal value. This suggests that we are near the threshold of single-qubit gate quality for which this GST implementation can still detect deviations between the ideal and the real cases.

Nonetheless, this GST implementation still proved itself to be rather useful in detecting errors in current IBM Q devices; this was especially true for readout errors, which were commonly at least one order of magnitude higher than regular gate errors. As a distinct example, figure 6 shows a GST estimation of the measurement operator with the

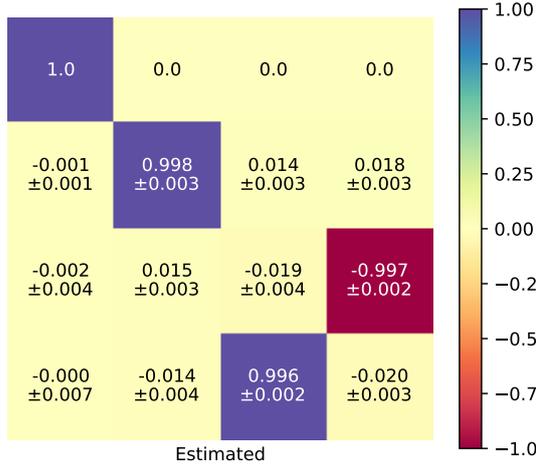


Figure 5: $X_{\pi/2}$ -gate's PTM estimation, for IBM Q Melbourne's 11th qubit. In an ideal case, the purple entries would have value "1", the red entry "-1" and the rest "0". The matrix's entries, r_{ij} , are represented in the standard order, with $i = 0, 1, 2, 3$ indexing the matrix line and $j = 0, 1, 2, 3$ the column.

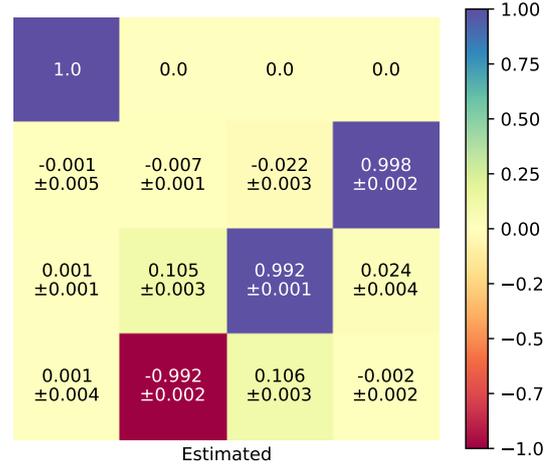


Figure 7: $Y_{\pi/2}$ -gate's PTM estimation, for IBM Q Yorktown's 3th qubit. In an ideal case, the purple entries would have value "1", the red entry "-1" and the rest "0".

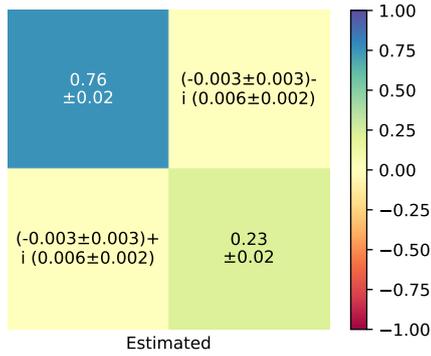


Figure 6: GST characterization of IBM Q Melbourne's qubit 7's measurement operator. In its ideal form, $E = |0\rangle\langle 0|$. The matrix's entries, r_{ij} , are represented in the standard order, with $i = 0, 1$ indexing the matrix line and $j = 0, 1$ the column.

4.2. Comparison with IBM's calibrations

Having a complete qubit characterization, one can also compute the error figures used by IBM to describe their devices. In this way, we compared our results with IBM's calibration data, which provided values for the single-qubit gate and readout errors.

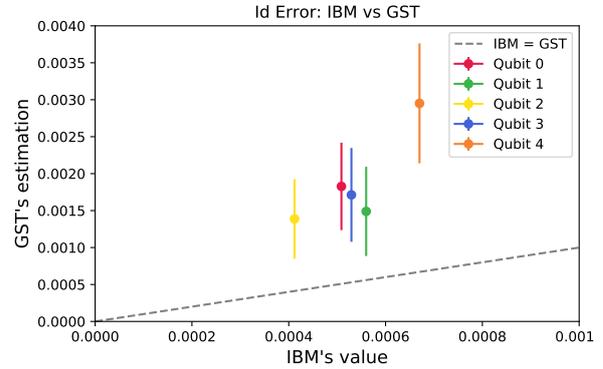


Figure 8: IBM vs GST: Identity gate errors in IBM Q Yorktown. The dashed grey line represents where the points should lie if GST's estimations and IBM's values were identical.

most substantial error out of all the characterized qubits, where the deviations from the ideal case are unequivocal even though the uncertainties are one order of magnitude above those of figure 5.

Another feature particular to GST is its ability to determine the actual values of the "ideally zero" parameters of a PTM, in which sometimes significant deviations are concentrated. This is especially relevant because the widely used gate fidelity, from eq. (43), only looks at the overlap between the two PTM's, which means that PTM values that would be zero in the perfect case do not affect the gate error. Take for instance the $Y_{\pi/2}$ -gate estimation in figure 7: although the "non-zero" parameters of the PTM are close to their ideal values, there are significant deviations in the "ideally zero" parameters.

For the gate errors, we generally witnessed that GST's precision tends to be too low for the order of magnitude of the errors we are dealing with; even for the points with smaller errors we did not see any trend, with the GST's gate error estimation lying either below, above or in agreement with those of IBM, depending on the qubit. The exception to the rule was the error of the identity gate in IBM Q Yorktown's qubits, which we compare with IBM's value in figure 8, and where we see that IBM's gate error values show a tendency to be underestimated.

As for the readout errors, exemplified in figure 9, we noticed a much more consistent match between the GST's estimations and IBM's values. For IBM Q

Melbourne’s qubits, we confirmed that most of the qubits had readout errors in the 2.5% – 7.5% range, with qubit 7 having a significantly higher readout error than the rest — something we had already remarked through figure 6, which exhibits a readout error that is even larger than the one provided by IBM.

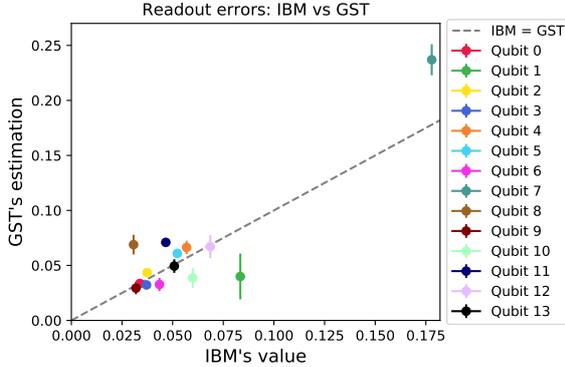


Figure 9: IBM vs GST: Readout errors in IBM Q Melbourne. The dashed grey line represents where the points should lie if GST’s estimations and IBM’s values were identical.

4.3. Device performance over time

Using GST, we chose to study IBM Q Melbourne’s 7th qubit’s behaviour over time, mainly to probe the evolution of its readout error.

The results, taken over around 48 hours, are presented in figure 10, where we observe how erratic the qubit’s readout is over time and how far apart its error can be from the values provided in IBM’s calibrations (in red). Although we cannot discern any specific pattern to these oscillations, these results show that the system’s readout error repeatedly changes in a rate of 1 – 2% per hour, which is especially worrying when we consider that our data collection takes a total of 40 minutes.

These fluctuations have far-reaching implications, in the sense that they compromise the accuracy of the whole estimations; this becomes particularly detrimental when considering its influence on the accuracy of smaller quantities like the single-qubit gate errors. In fact, what we observe in this case is that the gate error estimations end up having error bars that are too large to observe whether some type of fluctuations are happening to a significant extent — and so we refrain from showing those points. We stress, however, that in this particular case this is an unavoidable consequence that would arise in the results of even the most sophisticated GST implementations, simply because we are dealing with measurements that drift significantly throughout the data acquisition, which is something the physical model does not contemplate.

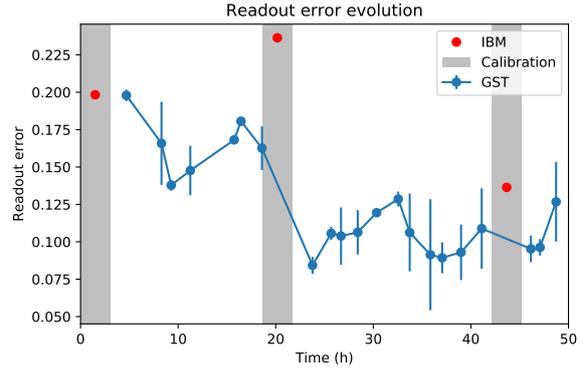


Figure 10: Readout error of IBM Q Melbourne’s qubit 7 over time. The grey boxes highlight when the calibrations were performed, while the red points are the readout error’s reported by IBM at the time of the calibrations.

5. Conclusions, findings and achievements

In this work, we set out to implement a fully functioning qubit characterization tool based on the GST framework. This process involved not only the study and subsequent programming of the whole procedure, but also a thorough examination of its capabilities and shortcomings. In this regard, we found that the method can produce reliable estimations, but only under a specific set of potentially demanding conditions.

Experimentally, we found that GST demands several experiments to be performed (in our case, we needed 92 distinct experiments) and each experiment to be repeated a large enough number of times (in our case, around 25000) to mitigate sampling noise. This results, depending on the device, on considerably long data acquisition times (in our case, for the IBM Q devices, of around 40 minutes), which leaves room for the system to drift significantly throughout the data collection phase. If this is the case, then GST’s capabilities are compromised by its inability to contemplate non-markovian behaviour.

When it comes to the estimation procedure, we discovered that the global minima of the MLE cost functions were degenerate. This degeneracy stemmed mainly from correlated deviations in the initial state and readout estimations that compensated one another and thus produced no net change in the cost function value. This was found to be a property endemic to the probability estimator of eq. (30), and thus present in all cost functions that try to match estimations to experimental points. Since this behaviour is particularly detrimental to the accuracy of the correlated quantities, if a useful estimation of one is desired, it is necessary to fix the other, thus forcing a new assumption about the system to be made.

We did not attempt to replicate the state of the art GST implementation because of its prohibitive

demands in data collection. Instead, we based our work on the more primitive implementation of ref. [6], upon which we made some modifications and improvements. These included the incorporation of an error estimation procedure and the adoption of differential evolution as a simple and efficient algorithm for global optimization. Moreover, we adapted the original method to account for entries in the data tensor that corresponded to identical experiments and fixed the existence of more free parameters than experimental points by increasing the number of different experiments without introducing extra gates.

By applying our GST implementation to the IBM Q devices, we noticed that not all results were of equal precision, with some displaying larger uncertainties than others as a virtue of their cost function landscape at the convergence point. Usually, we found that the average single-qubit gate quality of IBM's devices is close to the limit of GST's capacity to consistently and accurately detect deviations from their ideal form. This was not the case for the readout errors, whose estimated values have adequate uncertainty values, stemming from the fact that readout errors are usually one to two orders of magnitude higher than single-qubit gate errors.

Nonetheless, we found some instances in which GST's estimations conflicted with IBM's characterization values. Many hypotheses can be formulated to explain this, including ones that put into question either their or our characterization procedures. On the other hand, by characterizing a qubit with a particularly lousy readout over time, we found recurrent fluctuations in the readout error, often in the order of 1% – 2% per hour. Although it is difficult to say whether such drifts are present in the other qubits, smaller fluctuations would be sufficient to explain the discrepancy between our and IBM's characterizations.

References

- [1] Active quantum devices & simulators - IBM Q Experience.
- [2] R. Blume-Kohout, et al. Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit. 10 2013.
- [3] R. Blume-Kohout, et al. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nature Communications*, 8:14485, 2 2017.
- [4] I. L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. 10 1996.
- [5] P. J. Coles, et al. Quantum Algorithm Implementations for Beginners. 4 2018.
- [6] D. Greenbaum. Introduction to Quantum Gate Set Tomography. 9 2015.
- [7] E. Knill, et al. Randomized benchmarking of quantum gates. *Physical Review A*, 77(1):012307, 1 2008.
- [8] D. Leibfried, et al. Experimental Determination of the Motional Quantum State of a Trapped Atom. Technical report, 1996.
- [9] E. Magesan, R. Blume-Kohout, and J. Emerson. Gate fidelity fluctuations and quantum process invariants. *Physical Review A*, 84(1):012309, 7 2011.
- [10] E. Magesan, J. M. Gambetta, and J. Emerson. Scalable and Robust Randomized Benchmarking of Quantum Processes. *Physical Review Letters*, 106(18):180504, 5 2011.
- [11] E. Magesan, J. M. Gambetta, and J. Emerson. Characterizing quantum gates via randomized benchmarking. *Physical Review A*, 85(4):042311, 4 2012.
- [12] S. T. Merkel, et al. Self-consistent quantum process tomography. *Physical Review A*, 87(6):062119, 6 2013.
- [13] M. A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303(4):249–252, 10 2002.
- [14] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [15] K. Price, R. M. Storn, and J. A. Lampinen. *Differential Evolution: A Practical Approach to Global Optimization (Natural Computing Series)*. Springer-Verlag, Berlin, Heidelberg, 2005.
- [16] R. Storn and K. Price. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.
- [17] G. Upton and I. Cook. *A Dictionary of Statistics*. Oxford University Press, 2008.