RAIRO Operations Research
Will be set by the publisher

STEANE CODE ANALYSIS BY RANDOMIZED BENCHMARKING *

Anderson de Souza Barbosa[1], Franklin de Lima Marquezino[1] and Renato Portugal[2]

Abstract. Quantum error correction codes (QECC) play a fundamental role in protecting the information processed in today's noisy quantum computers. To build good error correction schemes, it is essential to understand how noise affects the behavior of these codes. In this research paper, we analyze Steane code, a 7-qubit QECC, using a randomized benchmarking (RB) protocol. With RB protocols, we can partially characterize the quality of implementation of a set of quantum gates. We show a scenario where Steane code with one logical qubit is advantageous compared to the situation with no quantum code. We obtained our results using a quantum simulator with custom noise models considering different numbers of noisy qubits.

Keywords: Quantum error correction, Steane code, performance evaluation, randomized benchmarking

Mathematics Subject Classification. 68Q12, 81P68

1. Introduction

Quantum computers should solve certain problems exponentially faster than classical computers with the best available algorithms. A well-known example of such a problem is integer factorization, which could be solved efficiently on

January, 2023.

* *This research has been financially supported by the Brazilian agencies FAPERJ, CAPES and CNPq.*

[1] Systems Engineering and Computer Science Program, Federal University of Rio de Janeiro, Av. Horácio Macedo 2030, Rio de Janeiro, RJ 21941-914, Brazil

[2] National Laboratory of Scientific Computing, Av. Getúlio Vargas 333, Petrópolis, RJ 25651-075, Brazil

This provisional PDF is the accepted version. The article should be cited as: RAIRO: RO, doi: 10.1051/ro/2023030

a quantum computer using Shor's algorithm [1]. However, the noise present in current quantum devices partially destroys the quantum information. Quantum algorithms such as Shor's usually require many qubits and deep circuits, which generate even more noise. Therefore, we still cannot take advantage of the full potential of quantum computers for solving real-world problems.

Quantum error correction and detection codes are a way to attempt to get around this problem. They seek to protect information by creating redundancies, that is, encoding one qubit into several ones, thereby being able to detect and correct a limited set of errors, which depends on the code. It is important to completely characterize the quantum noise in order to build good error correction schemes. That is possible through quantum process tomography. However, this method is not scalable, which makes it unfeasible for large systems [2]. An alternative is to partially characterize the noise acting in a quantum system and, for that, Magesan *et al.* [3] proposed a scalable randomized benchmarking (RB) protocol.

This protocol is suitable for characterizing noise at the level of physical qubits, but to analyze the performance of a quantum code, a method that evaluates it at the level of encoded logical qubits is the ideal solution. In that regard, Combes *et al.* [4] proposed a logical randomized benchmarking protocol that, as the name implies, is capable of evaluating an error correction implementation at the logical level. In this work, we use this protocol to evaluate the performance of Steane code, an important quantum error correction code with characteristics that facilitate the implementation.

In addition to the logical RB, several other variations of the original protocol were proposed. Morvan *et al.* [5] extended the RB to make it able to analyze qutrits. Brown and Eastin [6] examined RB protocols based on subgroups of the Clifford group that are not unitary 2-designs. Magesan *et al.* [7] proposed a new protocol called interleaved randomized benchmarking to analyze individual quantum gates. Proctor *et al.* [8] proposed a *direct* RB protocol that avoids compiling n-qubit Clifford gates, for large n, into large circuits composed of the native gates of the device to be characterized.

This work was inspired by the experiments performed by Harper and Flammia [9] with a quantum error detection code. They analyzed a 4-qubit code with distance 2 that encodes two qubits with a variation of the RB protocol called Real Randomized Benchmarking [10], which can be used with any gate set that is an orthogonal 2-design. They observed an improvement in the fidelity of the gates from 94.2%, when no code is used, to 99.4%, with the use of the 4-qubit code.

Our results show that, under certain circumstances, it is worthwhile to use Steane's code, as we obtain a higher fidelity than in the case with no code. Although the conditions under which the experiments were performed are not those found in the normal use of quantum devices, the results indicate that in these cases we would be able to obtain similar behaviors.

The rest of this paper is organized as follows. In Section 2, we present the background of this work. In Section 3, we explain our experiments. In Section 4, we present our final results. Finally, in Section 5, we summarize our conclusions.

2. Background

In this section we give some details about the code we used, Steane code, explain how to perform the randomized benchmarking protocol, and show how the quantum channels we considered act on a quantum state.

2.1. Steane code

Steane code is a quantum error correction code introduced by Steane [11] that encodes one logical qubit into seven physical ones and has distance 3, which means it can correct arbitrary errors up to 1 qubit. Because of these characteristics, we say it is a $[[7, 1, 3]]$ code. Furthermore, Steane code is a stabilizer code, that is, its code space can be generated by a subgroup of the Pauli group, which we call a stabilizer. Table 1 shows all six generators of Steane code, which consist of tensor products of the Pauli X and Z gates and the Identity gate (I).

Table 2 shows the main encoded Clifford gates in Steane code. We can see that all qubits can be implemented in a bitwise fashion and, as all Clifford gates can be implemented as a product of the gates shown in Table 2, all of them can be implemented in this fashion. We call this a transversal implementation. That is a feature of a class of quantum codes called CSS (Calderbank-Shor-Steane), to which the Steane code belongs. This class was discovered by Calderbank and Shor [12] and Steane [13] and is particularly useful due to its easy implementation.

Table 1. Steane code generators

Name	Operator
g_1	$I \otimes I \otimes I \otimes X \otimes X \otimes X \otimes X$
g_2	$I \otimes X \otimes X \otimes I \otimes I \otimes X \otimes X$
g_3	$X \otimes I \otimes X \otimes I \otimes X \otimes I \otimes X$
g_4	$I \otimes I \otimes I \otimes Z \otimes Z \otimes Z \otimes Z$
g_5	$I \otimes Z \otimes Z \otimes I \otimes I \otimes Z \otimes Z$
g_6	$Z \otimes I \otimes Z \otimes I \otimes Z \otimes I \otimes Z$

Table 2. Encoding of the Pauli X, Y and Z gates, the Hadamard gate (H), and the Phase gate (S), which are some of the main Clifford gates, in Steane code

Gate	Encoding
X	$X \otimes X \otimes X \otimes X \otimes X \otimes X \otimes X$
Y	$Y \otimes Y \otimes Y \otimes Y \otimes Y \otimes Y \otimes Y$
Z	$Z \otimes Z \otimes Z \otimes Z \otimes Z \otimes Z \otimes Z$
H	$H \otimes H \otimes H \otimes H \otimes H \otimes H \otimes H$
S	$ZS \otimes ZS \otimes ZS \otimes ZS \otimes ZS \otimes ZS \otimes ZS$

2.2. Randomized benchmarking

Randomized benchmarking is a protocol for partially characterizing the quality of the implementation of a set of quantum gates. That is done by estimating the *fidelity* between the identity channel and the average noise $\mathcal{E}$ acting on the set, which is given by

$$F(\mathcal{E}) = \int \mathrm{d}\psi \left\langle\psi\right| \mathcal{E}(\left|\psi\right\rangle\left\langle\psi\right|) \left|\psi\right\rangle, \tag{1}$$

where $\mathrm{d}\psi$ is the Haar measure on the space of quantum states $\left|\psi\right\rangle$. The fidelity is a distance measure, that is, it is a way of quantifying how similar two quantum states are, that is why it is important in the context of error correction. The fidelity between a pure state $\left|\psi\right\rangle$ and a mixed state ρ is given by

$$F(\left|\psi\right\rangle, \rho) = \sqrt{\left\langle\psi\right| \rho \left|\psi\right\rangle}. \tag{2}$$

The RB protocol was proposed by Magesan *et al.* [2, 3] and is scalable in the number of qubits of the system. It consists of the following steps:

(1) Choose a sequence length m and prepare a state ρ.
(2) Choose uniformly at random a sequence of m Clifford gates, apply them to the circuit and apply the inversion gate, which, in an ideal situation, inverts all the sequence and makes the circuit equivalent to the identity.
(3) Perform the measurement to verify if the register returned to the initial state ρ.

After the development of this RB protocol, several modifications were proposed to adapt it to different situations and needs. One of these modifications was the logical randomized benchmarking proposed by [4], which is suitable for benchmarking quantum gates encoded in QECCs. The logical RB protocol can be performed through the following steps:

(1) Choose an initial state ρ and a sequence length m.
(2) Choose uniformly at random a sequence of m Clifford gates

$$s_m = \{U_1, U_2, \ldots, U_m\}.$$

(3) Perform the following steps T times:
    (a) Encode ρ using Steane code.
    (b) For each $U \in \{U_1, U_2, \ldots, U_m\}$:
        (i) Apply the encoded version of gate U.
        (ii) Apply the error correction procedure.
    (c) Apply the gate $U^{-1} = \prod_{U \in s_m} U^\dagger$ encoded.
    (d) Apply the error correction procedure.
    (e) Apply the decoding procedure.
    (f) Measure ρ with E, an operator of a POVM.
(4) Calculate the percentage F_m of executions that returned to the initial state for m gates.

POVM (Positive Operator-Valued Measure) is a formalism for the analysis of quantum measurements that is suitable for applications where we are mainly interested in measurement probabilities associated with each operator. In the logical RB protocol, the POVM element E represents the situation in which the qubit returns to the initial state, whose probability we want to estimate. The value F_m obtained with the protocol above corresponds to the probability that the initial state ρ be recovered at the end of the circuit for a given sequence length ρ. To estimate the average of this probability for all possible sequences, it is necessary to perform the steps above for different sequences of m gates. In addition, the whole procedure must also be done for different values of m so that we obtain several points that associate a sequence length with a probability, known as survival probability. The curve obtained from these points is known as the fidelity decay curve, which can be fitted to the model

$$F_L(m) = Ap^m + B, \tag{3}$$

where A and B are constants that absorb state preparation and measurement errors and p is given by

$$p = \frac{dF(\Lambda_L) - 1}{d - 1}, \tag{4}$$

where Λ_L is the average noise acting on the code gate set and $d = 2^n$, where n is the number of logical qubits. That means that by finding the value of p through the fit, we can determine the average fidelity $F(\Lambda_L)$.

2.3. Quantum channels

Below we present the three quantum channels that we used to build the noise models: the depolarizing channel, the amplitude damping, and the phase damping, which are some of the most important. Let ρ be a mixed quantum state.

- Depolarizing channel: The action of the n-qubit depolarizing channel is given by

$$\mathcal{E}_D(\rho) = \frac{pI_n}{2^n} + (1-p)\rho \tag{5}$$

  where p is the parameter of the channel and I_n is the $n \times n$ identity matrix.
- Amplitude damping: The action of the 1-qubit amplitude damping channel is given by

$$\mathcal{E}_A(\rho) = A_0\rho A_0^\dagger + A_1\rho A_1^\dagger \tag{6}$$

  where the Kraus operators are given by

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{bmatrix} \tag{7}$$

  and

$$A_1 = \begin{bmatrix} 0 & \sqrt{\lambda} \\ 0 & 0 \end{bmatrix}, \tag{8}$$

where λ is the parameter of the channel. The Kraus operators of the 2-qubit amplitude damping are given by [14]

$$A_{ij} = A_i \otimes A_j,\ i,j = 0,1. \tag{9}$$

- Phase damping: The action of the 1-qubit phase damping is given by

$$\mathcal{E}_P(\rho) = P_0 \rho P_0^\dagger + P_1 \rho P_1^\dagger \tag{10}$$

where the Kraus operators are given by

$$P_0 = A_0 \tag{11}$$

and

$$P_1 = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{\lambda} \end{bmatrix}, \tag{12}$$

where λ is the parameter of the channel. The Kraus operators of the 2-qubit phase damping are given by

$$P_{ij} = P_i \otimes P_j,\ i,j = 0,1. \tag{13}$$

3. Methodology

The experiments from this work were performed in the simulators *ibmq_qasm_simulator*, available on IBM Quantum, and *qasm_simulator*, which can be run locally through Qiskit [15]. We decided to use simulators instead of real quantum devices because some non-unitary instructions necessary to implement Steane code, such as mid-circuit measurements, returning a qubit to the state $|0\rangle$ (*reset*) and applying gates conditioned to the value of the classical register (*if*), were not available on IBM Quantum computers. The first two instructions became available later, but the *if* remains unavailable.

To simulate noise, we created noise models based on three of the most important quantum channels, the depolarizing channel, the amplitude damping, and the phase damping. The first one is characterized by a parameter p and the others by a parameter λ. Our goal was to compare the performance of experiments using one physical qubit and using one logical qubit encoded in Steane code, which corresponds to 13 physical qubits, seven from the code block and six ancilla ones. For that, we created noise models for each quantum channel acting on different numbers of qubits from the code block, from 2 to 7, and we sought to find a parameter value (threshold), for each model, in which the fidelity decay curves were approximately equal for encoded and non-encoded experiments. To each model were added 1-qubit and 2-qubit errors from the respective quantum channel. The 1-qubit errors were associated with the gates $U1$, $U2$, $U3$ and I, and the 2-qubit errors with the $CNOT$ gate. It is important to mention that the circuits to be run were initially built with the gates X, Y, Z, H, S, $S^\dagger$ and $CNOT$, but later were converted into equivalent circuits with the gates $U1$, $U2$, $U3$ and $CNOT$, that were the standard gates on IBM Quantum devices. We did not add noise on ancilla qubits because the threshold would be so low that it would require a number of gates in the thousands to generate a useful curve and, as a result, an amount of memory much higher than the 8 GB available on IBM Quantum.

We performed experiments with initial states $|0\rangle$ and $|1\rangle$ and with 30 different values for m starting at 2 and with equal intervals. The lower the threshold found, the greater the interval necessary to generate a good curve. For each value of m, we considered the average of the values obtained for $|0\rangle$ and $|1\rangle$. Each circuit was run 1,024 times to estimate the probability of returning to the initial state, which means $T = 1024$ in the logical RB protocol.

The use of Steane code consists of three procedures: encoding, correction, and decoding. The encoding is done at the beginning of the circuit, starting from the original state with which we want to work. Figure 1 shows the circuit that we used to encode the initial state in Steane code, which includes only the seven qubits from the code block and performs no action on the ancilla. The correction procedure is performed after each encoded gate is applied and consists of two steps: syndrome measurement and recovery. The syndrome measurement is performed by measuring each generator of the Steane code. To measure each generator, one ancilla qubit is needed, and that is why we need six. The combination of the results of all the measurements is the error syndrome. Let β_i be the result of the measurement of the generator g_i and $\{E_j\}$ be a set of correctable errors for the Steane code. We perform the recovery by applying E_j such that $E_j g_i E_j^\dagger = \beta_i g_i$ for all i. Finally, the decoding converts the logical state encoded in Steane code into the physical state with the original number of qubits and is performed by inverting the circuit from Figure 1. These three procedures are steps in the logical RB protocol presented above.

4. Results

We created a different noise model for each quantum channel and each number of noisy qubits from 2 to 7, that is, we performed experiments for 18 different models. We found the threshold for each of these models. Figure 2 shows the thresholds obtained for each case. The graph is in logarithmic scale for better visualization. The channel that has the highest thresholds was the phase damping channel, which shows it is the one that causes less damage to the information for the same parameter value. The other channels have similar thresholds, but those from the depolarizing channel were slightly lower.

We can notice that, in this scale, the points form curves that are close to lines, which indicates that a function of the form $f(x) = ax^b$ can be fitted to them. The fitted function for the phase damping was $f(x) = 1.99x^{-2.32}$, for the amplitude damping was $f(x) = 0.73x^{-2.30}$ and for the depolarizing channel was $f(x) = 0.75x^{-2.59}$. The lines shown in Figure 2 represent the fitted functions. Although our experiments were limited to simulating noise on the code block, the

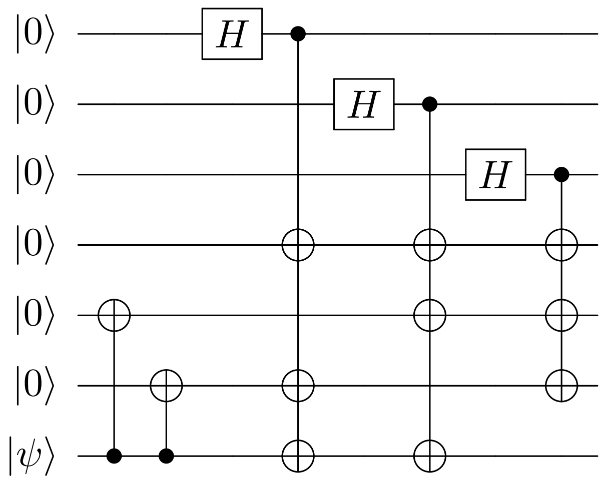

FIGURE 1. Circuit to encode quantum state $|\psi\rangle$ in Steane code

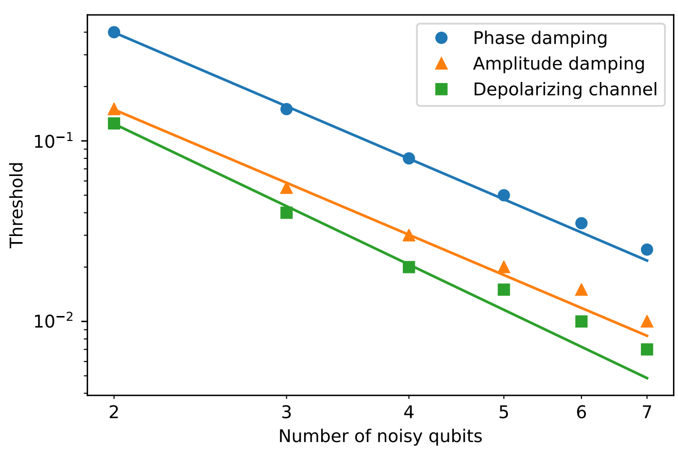

FIGURE 2. Threshold as a function of the number of noisy qubits for each quantum channel

curves we obtained in Figure 2 show that we can also expect to find thresholds when simulating noise on the ancilla, which would represent a more realistic scenario where the use of Steane code is advantageous.

Table 3 shows the main average fidelities found. They refer to the encoded and non-encoded experiments in which the models, for each channel, were generated using the threshold for seven noisy qubits. In the case of encoded experiments, column Configuration indicates the number of noisy qubits from the code block, and only the cases from 5 to 7 qubits are shown. We can see that the average fidelities from the encoded experiments with seven noisy qubits and the non-encoded ones are almost the same, with a maximum difference of 3×10^{-4}. That indicates that, with lower parameter values, we would obtain better results when using Steane code than when no quantum code is used. Furthermore, the fidelities show a scenario, albeit restricted, that clearly favors Steane code, which is the one in which there are five noisy qubits with noise that can be modeled using the thresholds corresponding to seven noisy qubits, for each channel.

Table 3. Average fidelities found. Column Configuration indicates if the fidelity is a result of non-encoded experiments or the number of noisy qubits in the cases where the Steane code was used.

Channel	Parameter	Configuration	Average fidelity
Depolarizing channel	$p = 0.007$	5 qubits	0.9983
		6 qubits	0.9977
		7 qubits	0.9969
		non-encoded	0.9967
Amplitude damping	$\lambda = 0.01$	5 qubits	0.9985
		6 qubits	0.9977
		7 qubits	0.9971
		non-encoded	0.9968
Phase damping	$\lambda = 0.025$	5 qubits	0.9981
		6 qubits	0.9971
		7 qubits	0.9961
		non-encoded	0.9960

5. Conclusion

Our objective in this work was to find a scenario in which using a quantum error correction code was beneficial. We chose Steane code for our experiments because it is an influential code that is also easy to implement. To evaluate the code and compare it to the situation without error correction, we used the fidelity metric, which can be estimated using a randomized benchmarking protocol. Furthermore, we chose to run our experiments in quantum simulators with custom noise models due to the limitations of the quantum devices available in IBM Quantum.

We found the threshold for all 18 noise models created. In a scenario with the same number of noisy qubits with noise modeled by one of the three quantum channels considered, if the parameter value is less than the corresponding threshold, it would be advantageous to use the Steane code. The quantum channel that generated the lowest thresholds was the depolarizing channel, which means that it is the one that damages the information the most. On the other hand, phase damping was the channel that generated the highest thresholds. In addition, we found that on a logarithmic scale the curve of the threshold as a function of the number of noisy qubits is approximately a line, indicating the possibility of finding thresholds in cases where ancilla qubits are also noisy. However, it is noteworthy that this curve probably has a different behavior when considering ancilla qubits, because in this block the structure of the circuit is different.

References

[1] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proceedings 35th annual symposium on foundations of computer science*, pp. 124–134, Ieee, 1994.

[2] E. Magesan, J. M. Gambetta, and J. Emerson, "Characterizing quantum gates via randomized benchmarking," *Physical Review A*, vol. 85, no. 4, p. 042311, 2012.

[3] E. Magesan, J. M. Gambetta, and J. Emerson, "Scalable and robust randomized benchmarking of quantum processes," *Physical Review Letters*, vol. 106, no. 18, p. 180504, 2011.

[4] J. Combes, C. Granade, C. Ferrie, and S. T. Flammia, "Logical randomized benchmarking," *arXiv preprint arXiv:1702.03688*, 2017.

[5] A. Morvan, V. Ramasesh, M. Blok, J. Kreikebaum, K. O'Brien, L. Chen, B. Mitchell, R. Naik, D. Santiago, and I. Siddiqi, "Qutrit randomized benchmarking," *Physical Review Letters*, vol. 126, no. 21, p. 210504, 2021.

[6] W. G. Brown and B. Eastin, "Randomized benchmarking with restricted gate sets," *Physical Review A*, vol. 97, no. 6, p. 062323, 2018.

[7] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. Da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, *et al.*, "Efficient measurement of quantum gate error by interleaved randomized benchmarking," *Physical Review Letters*, vol. 109, no. 8, p. 080505, 2012.

[8] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, "Direct randomized benchmarking for multiqubit devices," *Physical Review Letters*, vol. 123, no. 3, p. 030503, 2019.

[9] R. Harper and S. T. Flammia, "Fault-tolerant logical gates in the ibm quantum experience," *Physical Review Letters*, vol. 122, no. 8, p. 080504, 2019.

[10] A. Hashagen, S. Flammia, D. Gross, and J. Wallman, "Real randomized benchmarking," *Quantum*, vol. 2, p. 85, 2018.

[11] A. M. Steane, "Error correcting codes in quantum theory," *Physical Review Letters*, vol. 77, no. 5, p. 793, 1996.

[12] A. R. Calderbank and P. W. Shor, "Good quantum error-correcting codes exist," *Physical Review A*, vol. 54, no. 2, p. 1098, 1996.

[13] A. Steane, "Multiple-particle interference and quantum error correction," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 452, no. 1954, pp. 2551–2577, 1996.

[14] Y.-n. Guo, Q.-l. Tian, K. Zeng, and Z.-d. Li, "Quantum coherence of two-qubit over quantum channels with memory," *Quantum Information Processing*, vol. 16, no. 12, pp. 1–18, 2017.

[15] G. Aleksandrowicz, T. Alexander, P. Barkoutsos, *et al.*, "Qiskit: An Open-source Framework for Quantum Computing," 2019.