IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 November 2025, accepted 22 November 2025, date of publication 3 December 2025,
date of current version 15 December 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3639914

APPLIED RESEARCH

A Teleportation Protocol Variant for Single-QPU Benchmarking

CRISTIAN MÁRQUEZ[1], DANIEL SIERRA-SOSA[2], (Member, IEEE), AND KELLY GARCÉS[1]
[1]Department of Systems and Computing Engineering, Universidad de los Andes, Bogotá 111711, Colombia
[2]Department of Computer Science, The Catholic University of America, Washington, DC 20064, USA

Corresponding author: Daniel Sierra-Sosa (sierrasosa@cua.edu)

ABSTRACT With the advent of more quantum computing applications, it is imperative to provide developers with a set of metrics to determine how much trust they can place in the execution of their programs. In this study, we propose a variation of the quantum teleportation to systematically evaluate the performance of quantum circuits and quantify their execution reliability. Our protocol consists of three parts: the payload, composed of entangled qubits with random gates applied to each wire; the bridge, which corresponds to the teleportation-inspired protocol; and an inverse payload operation, designed to return the circuit to its initial state. We validated the algorithm by elaborating its mathematical framework, followed by a comparative analysis with simulation results using Qiskit AerSimulator and ultimately through an execution on the IBM quantum and Rigetti services. A key aspect of our methodology was to set the circuit transpilation to its lowest optimization level (level 0). We recognize that even without aggressive optimization, randomly generated gates can form self-canceling sequences. As a result, our analysis correlated the pre-runtime metrics of the final, minimally-transpiled circuit (such as its actual depth and gate count) with post-runtime results (such as success rate). This approach reveals significant correlations and offers developers quantitative metrics to assess both the reliability of quantum circuit execution and scalability limitations in their applications.

INDEX TERMS Quantum computing, qubits, quantum frameworks, benchmarking, quantum quality metrics.

I. INTRODUCTION

In recent years, companies such as Google, AWS, IonQ, IBM, and Microsoft have led the race to build a practical quantum processing unit (QPU). Up to this day, a number of chips have been released, including Willow [1], Majorana 1 [2], Ocelot [3] and Heron [4], taking quantum utility one step closer to practical realization. A common characteristic of these advanced chips is the enhancement in error correction techniques. In particular, the chips developed by Google and Amazon employ randomized benchmarking (RB) techniques to estimate average error rates. The focus on error correction is fundamental to fully exploiting the potential of logical qubits, which represent an advanced abstraction level wherein information is shielded from noise by being encoded across multiple physical qubits.

The accelerated advancement in hardware technology presents a significant challenge for software developers who use this technology. With powerful QPUs becoming more accessible, the focus of the community shifts from the feasibility of hardware construction to the reliability of circuit execution, a challenge that defines the core of our research.

To ensure reliable results, developers currently rely on a combination of logical qubits and benchmarking protocols, each with specific limitations from a circuit-reliability prediction standpoint. For instance, logical qubit abstraction aims to mitigate errors through foundational error-correction codes [5], but its hardware-dependent implementation makes it difficult for a developer to anticipate how their unique circuit structure will interact with the error correction layer. Similarly, hardware-centric benchmarks like Quantum Volume [6] or IBM's "layer of fidelity" [7] provide a single holistic score based on random circuits, which reveals little about how a specific structured algorithm will behave. Even Randomized Benchmarking (RB), while

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato.

excellent for calculating an average gate error rate [8], [9], is explicitly designed to eliminate sequence-specific errors that are critical to the reliability of a particular application. This failure to deliver circuit-specific predictive insights makes these existing approaches unsuitable for improving the understandability of circuits, nor can they serve as a foundational framework for the development of reliability models for user-defined circuits.

This disconnect has led to calls within the research community for more application-motivated benchmarks that can bridge the gap between general hardware capability and specific circuit reliability [10], revealing a knowledge gap in the quantum computing domain. Addressing this gap requires moving beyond a purely hardware-focused evaluation. The task involves a technology built on multiple layers of abstraction, from inaccessible physical qubits to the error-corrected logical qubit layer, where tests are performed [11], [12].

To bridge this gap between general hardware capability and specific circuit reliability, this paper introduces a comprehensive approach for circuit evaluation. We propose a complete benchmarking strategy that includes a methodology for generating controlled random circuits to systematically validate the correctness of the circuit. In our experimentation, we establish a crucial correlation between pre-runtime metrics (such as number of gates and two-qubit gate count) and post-runtime outcomes (e.g., success rate), providing a predictive link between a circuit's design and its execution fidelity.

The heart of our approach is a variation of the Teleportation Protocol (vTP). The protocol is named because it utilizes the same initial entanglement structure as the canonical teleportation protocol. But, instead of using the measurement operations and the required classical communication, we propose the use of fixed local unitary operations to reconstruct the state of the *payload*, i.e., the set of qubits and gate operations that constitute the quantum state being benchmarked. This modification transforms the teleportation protocol from a communication utility into a testing environment that can be used on a single QPU today while being designed for future distributed quantum services.

This framework allows us to investigate the fundamental aspects of circuit reliability. Specifically, our work is driven by the following research question.

RQ1 How do the pre-runtime structural characteristics of quantum circuits (such as payload size, circuit depth, gate count, and gate types) relate to post-runtime outcomes, including execution success rates, execution time, cost, and quantum information degradation?

The remainder of this paper is organized as follows. Section II presents a review of the classification of benchmarking techniques, including our formulated benchmarking approach, and describes both pre-runtime and post-runtime metrics. Sections II-D and III outline the mathematical formulation and conduct simulations for both the teleportation protocol and a variant. Section IV presents our observations and results. Finally, Section V summarizes our findings.

II. BACKGROUND

This section begins by presenting pre-runtime and post-runtime metrics, grouping relevant reported metrics according to the software quality attributes described in the ISO/IEC 25010 standard [13]. Although this standard was created for classical computing, the literature argues that various QC constructs align with such quality descriptors [14], [15], [16], [17], given that quantum algorithms represent an alternative paradigm for addressing computational challenges. It is also important to consider constraints at the business level. While quality attributes from standards such as ISO/IEC 25010 primarily focus on technical and user-centric software quality, factors such as risk management, business success, budget, and financial costs are not explicitly part of quality descriptors. In this study, we recognize the significance of business constraints for the viability of a software project and assert that they must be considered when formulating solutions that incorporate quantum computing.

The section proceeds with a review of established benchmarking classifications, followed by a presentation of our own benchmarking strategy used to run the experiment. Finally, a description of the teleportation protocol is provided to serve as the foundation for the algorithm discussed in the next section.

A. PRE-RUNTIME AND POST-RUNTIME METRICS

In order to understand the complexities associated with designing a benchmark, it is fundamental to first analyze the procedure adopted by developers in QC, which can be generally categorized into two overarching steps.

In the first step, developers prepare quantum states by applying a combination of quantum gates to a set of qubits. This process is analogous to the creation of classical computer pseudocodes. Once the quantum algorithm is ready, it is coded using a library such as Qiskit, Q#, or PennyLane, which is equivalent to using a programming language like Prolog or a hardware description language like VHDL or Verilog. The result of this step is known as a quantum circuit and serves as the basis for the next step.

In the second step, the circuit is run on a QPU and measurements are collected. It is important to note that before reaching the runtime service, the circuit is transpiled and optimized to match the target quantum chip. The process is equivalent to optimizing high-level programming languages before converting them into low-level instructions.

Alternatively, as an intermediate step, developers may run the circuit on a quantum simulator. However, the number of qubits that can be simulated using classical hardware is limited; for example, simulating 30 qubits requires 16 GB of RAM [18], and even with parallel processing, simulation allows one to go only up to about 240 qubits [19].

In accordance with this workflow, which spans from the development of the circuit to its execution on a QPU or

simulator, it is possible to differentiate between pre-runtime and post-runtime metrics.

1) PRE-RUNTIME METRICS

The pre-runtime metrics are computed after coding the algorithm and before executing the quantum circuit.

The literature highlights that, in addition to some metrics unique to the quantum domain, metrics commonly used in classical computing are also applicable to quantum computing, such as the speed of operations, the number of processing units, and the probability that errors occur in the computation [20]. A number of metrics identified in the literature align with our definition of pre-runtime metrics and, in general, can be categorized under the dimensions of Maintainability.

Maintainability metrics. This category addresses the practical, long-term aspects of developing QC algorithms. Maintainability metrics help quantify the effort and resources required to understand, modify, and manage a quantum program throughout its lifecycle.

Maintainability metrics include a variety of related structural, operational and complexity metrics that quantify the effort to understand and manage a quantum program [20], [21]. Key structural metrics include circuit width (CW), which is the total number of qubits, and circuit depth (CD), representing the longest path from input to output. An example of complexity metrics are the number of circuit gates (CCG), conditional instructions (CI) to quantify the classical control flow, and quantum cyclomatic complexity (QCC) to measure structural complexity. Finally, some of the operational metrics include tracking the number of measurement operations (MO), initial and reset operations (IRO), and any required auxiliary qubits (AQ).

2) POST-RUNTIME METRICS

The post-runtime metrics are computed from the QPU results, which in some cases involves creating a correlation with the pre-runtime metrics. The literature highlights a variety of metrics that are consistent with our post-runtime definition and, in general, these metrics primarily measure the actual Performance Efficiency and Reliability of the quantum system.

a: PERFORMANCE EFFICIENCY METRICS

This characteristic represents the performance of a system in relation to the resources consumed under specified conditions [13]. In the post-runtime QC context, these metrics evaluate the overall effectiveness of an algorithm by measuring the quality and speed of the results.

Performance Efficiency metrics include evaluations of task-solving efficiency, such as quantum volume and total quantum factor [22]. The quality of the results is assessed through precision, which examines the degree to which repeated measurements are related, and accuracy, which determines the degree to which the results are related to the true value [23]. Furthermore, crucial operational aspects are evaluated by measuring decoherence, representing information lost to environmental factors [24], and the response times between a request and its result [25]. An overview of these performance metrics is available in [26] and [27].

b: RELIABILITY METRICS

This characteristic is defined as the degree to which a system performs specified functions under specified conditions for a specified period of time [13]. In the quantum context, this classification is centered on the final results of a algorithm execution, specifically addressing either its successful or failure conclusion.

Reliability metrics primarily evaluate the success rate [28] and the fidelity of the circuit [8]. The most common protocol for calculating them is Randomized Benchmarking (RB), which, in summary, creates random circuits, applies a mirror circuit, and expects to obtain the identity matrix after measurement [29].

3) PHYSICAL HARDWARE CHARACTERISTICS

In contrast to pre- and post-runtime metrics, hardware characteristics constitute a set of metrics designed to measure the inherent physical properties of quantum hardware. These metrics, which are governed by hardware calibrations, technologies, and configurations rather than software-level controls, required the establishment of a distinct category for their presentation within this study.

Hardware quality metrics include physical measurements such as T1 (energy relaxation, ''the time needed for a qubit to move from the excited state $|1\rangle$ to the ground state $|0\rangle$'' [30]) and T2 (dephasing, ''the elapsed time before a qubit's resonance frequency becomes unidentified'' [30], [31]). In addition, it is possible to measure the error due to noise characterization in quantum gates [32].

B. BENCHMARKING REVIEW

Malhotra et al. [33] identify several benchmarking variants used to evaluate the performance and reliability of QPUs. The principal variants are outlined as follows.

1) RANDOMIZED BENCHMARKING (RB)

RB is a protocol used to estimate the average error rates of quantum circuits by applying sequences of randomly chosen gates (typically from the Clifford group) and observing the decay in success rate, error propagation, and/or gate fidelity as the size of the generated random circuit increases [29]. The authors highlight that RB is robust to ''state preparation and measurement'' (SPAM) errors and scalable to larger systems [33].

In this research, we used elements of RB since it aligns with our definitions of pre-runtime and post-runtime. More precisely, RB offers a structured framework for the analysis of quantum circuits by employing pre-runtime metrics, including gate count, circuit depth, and Clifford group composition, aimed at facilitating the investigation of their

correlation with post-runtime metrics, such as execution time or success and error rate [33].

2) APPLICATION-ORIENTED BENCHMARKING

This approach focuses on running known quantum algorithms to assess QPU performance and utility [33]. The idea behind this benchmarking method is to compare measurements with theoretical results.

In this investigation, we chose not to employ application-oriented benchmarking. Although this methodology offers valuable insights via pre-runtime metrics derived from established algorithms, our study advocates the use of random circuits, as their characteristics are better suited for creating a general-purpose transfer protocol.

3) VOLUMETRIC BENCHMARKING

In this method, a QPU is mapped using random quantum circuits with varying widths (number of qubits) and depths (number of gate layers) [33]. Due to the nature of volumetric benchmarking, which is intended to evaluate hardware, no elements of this group are included in this study.

4) QUANTUM TOMOGRAPHY

Quantum tomography is a set of techniques that are used to characterize quantum systems by reconstructing their quantum states or operations [29]. There are three types of post-runtime metrics in this group: state tomography, which reconstructs a quantum state's density matrix; process tomography, which characterizes quantum operations; and gate set tomography, which describes a set of quantum gates in a self-consistent way [33].

These benchmarking methods provide detailed information on quantum states and operations. However, they are resource-intensive and require a large number of measurements, which makes them less scalable for large quantum systems [29]. Additionally, they are not directly controllable at the software layer, as they require specific hardware configurations and measurement protocols [33]. Due to these factors, our study does not consider variants of this category.

5) CYCLE BENCHMARKING

Cycle benchmarking is a protocol that repeatedly applies quantum gates to amplify and characterize errors in a quantum system [33]. This approach allows for systematic analysis of errors in quantum circuits. The method is particularly useful for understanding how errors affect circuit success rates, as it provides a direct measurement of error rates by repeatedly applying the same gates to a qubit [29].

Considering the nature of understanding system errors, our study integrates cycle benchmarking methodologies. Specifically, we include gates operations on each qubit included in the payload of the designed protocol to determine the success rate for each experiment. See Section Variation of the Teleportation Protocol (vTP) for more details about the protocol.

6) MISCELLANEOUS APPROACHES

These approaches combine components from previously described benchmark variants. Several protocols fit this group, including hybrid benchmarking, which combines classical and quantum components, matchgate benchmarking, which focuses on specific gate sets, and random circuit sampling, which incorporates elements of RB with application-specific metrics [34].

In the present study, we propose a benchmarking method that fits within this category by integrating cycle benchmarking for error characterization and RB for success rate.

C. OUR BENCHMARKING DESIGN

This research introduces a benchmarking design based on a variant of the teleportation protocol, intended to evaluate the holistic reliability of an entire circuit rather than the error of individual gates. To achieve this, our methodology focuses on identifying the most effective circuit descriptors by correlating a selected subset of pre- and post-runtime metrics. The majority of these pre-runtime metrics are programmatically gathered from the Qiskit `QuantumCircuit` class, allowing for a systematic and reproducible approach.

1) PRE-RUNTIME METRICS

The pre-runtime metrics selected for this study focus on circuit complexity and configuration attributes. Although a comprehensive list and formal definitions of these metrics are detailed in the literature [20], [21], our investigation focuses on a specific subset. For clarity, we define Circuit Depth (CD) as the longest path of sequential gates, Circuit Width (CW) as the total number of qubits, and Circuit Size as the total number of gates. To understand the composition in more detail, we also track the Gate Count Operations, which provides a specific count for each type of gate used. Finally, we track a configuration metric unique to our experimental design, the payload size, which refers to the number of qubits that comprise the message in our vTP protocol and are initially entangled through CNOT operations.

2) POST-RUNTIME METRICS

The post-runtime metric monitored in this study is the **Success Rate**, defined as the proportion of quantum circuit executions that produce the desired measurement. In the context of our experiment, where the circuit is designed to return to the all-zero state, the success rate is calculated as:

$$\text{Success Rate} = \frac{\text{Number of measurements with all-zero outcome}}{\text{Total number of shots}} \quad (1)$$

A rate of 1.0 indicates that every execution produced the correct all-zero state, while a rate of 0.0 indicates complete failure. In addition to this, we also track the **Error Rate**, which is the complement of the success rate ($1-\textit{successrate}$) and quantifies the proportion of executions that resulted in an incorrect final state.

3) BENCHMARKING STRATEGY

The primary objective of this experiment is to identify pre-runtime metrics that exhibit a correlation with the success rate. To this end, our strategy integrates elements of randomization and cycle benchmarking to assess the success rate of transmitting a quantum state across the circuit. This is accomplished by employing RB within the payload to generate messages, thereby repeatedly applying quantum gates to intentionally amplify errors. The success rate is finally determined by executing the conjugate operation prior to measurement with the aim of obtaining a result of zero from the measured qubit.

The experimentation phase was structured into three distinct tests: (1) a stress test involving the use of N random gates to serve as the message; (2) a strategy to maintain a specified circuit depth; and (3) an evaluation of how success rates scale with an increasing number of qubits. Details about each experiment are presented in Section IV.

D. THE TELEPORTATION PROTOCOL

The teleportation protocol (TP) is a well-defined procedure —introduced by Bennett et al. [35] in 1993— that typically involves three qubits: one holding the message to be teleported (Alice's message qubit, M) and two others forming an entangled pair. The protocol assumes that Alice and Bob are physically separated; Alice possesses one qubit from the entangled pair (qubit A), and Bob holds the other (qubit B). In addition, a classical communication channel between them is mandatory.

TP proceeds as follows: Initially, Alice performs operations to entangle her message qubit (M) and her own qubit (A), which was previously entangled with Bob's qubit (B). She then measures these two qubits (M and A). This measurement projects their combined quantum states onto a definite classical outcome, which yields two classical bits of information, but, in doing so, irreversibly alters the original states of M and A. Following this, Alice communicates these classical bits to Bob. Based on the information received, Bob applies a specific set of corrective unitary operations to his qubit (B), and as a result, he recreates the original quantum state of Alice's message (M) on his qubit.

For a formal mathematical treatment of the TP, please refer to the appendix. Of particular importance is Equation 19, which forms the basis of the protocol variant discussed in this investigation. The sequence of operations is visually summarized in the quantum circuit diagram in Figure 1, which provides an intuitive overview of the entire process.

It is important to note that the efficacy of the protocol is dependent on a classical communication channel. Specifically, Alice is required to transmit her measurement outcome to Bob, enabling him to execute the corresponding corrective operation. This classical communication requirement ensures that no information is transmitted faster than the speed of light, thereby maintaining consistency with the principles of relativistic causality.

To close this section, it is pertinent to acknowledge the complexities inherent in executing the teleportation protocol in practical scenarios. In addition to the challenges related to the loss incurred during the separation of entangled qubits, there must also be an active classical communication channel available. However, existing literature has documented the successful transmission of a single photon across a distance exceeding 14,000 km utilizing a ground-to-satellite quantum teleportation system [36].

III. VARIATION OF THE TELEPORTATION PROTOCOL (vTP)

Running the protocol on a single qubit involves additional complexity to ensure that the chip utilizes physically separated qubits. Typically, targeting two distant qubits on the same chip is achievable by performing multiple swap operations before protocol preparation, although this configuration might introduce additional errors.

Therefore, we modified the protocol so that, instead of transporting quantum information between two distant qubits, we could test the quantum state while being subject only to errors from the required gates of the algorithm and those introduced by the hybrid benchmark in the payload.

In this paper, we propose an algorithm for testing quantum states in a quantum circuit. Inspired by TP, our algorithm propagates the state through the circuit instead of transmitting quantum information between distant qubits. This method enables us to analyze random operations using a hybrid benchmark framework.

As with TP, vTP can be described in three main phases: first, the preparation of the quantum state (the message qubit M) to be processed; second, the application of the core protocol gates (Bell State setup); and third, the state reconstruction or transformation. vTP notably differs from standard TP in the last phase: it employs CNOT and CZ gates instead of measurement operations and does not require a classical communication channel for state reconstruction.

The vTP circuit diagram is presented in Figure 2 where the barriers demarcate the key stages. The diagram shows payload preparation, followed by the standard TP gates, then the unique vTP operations. The final part illustrates the validation sequence and the measurement operations used to determine the execution success.

To initiate the theoretical examination of vTP, we start by using the Bell state given in Equation 19 for qubits M and A in terms of the computational basis.

The standard TP involves Alice executing the measure operation on her qubits M and A, followed by sending the results to Bob via the classical communication channel in order to reconstruct her quantum state.

However, at this point, vTP and TP diverge. Instead of conditional operations and measurements, vTP unconditionally applies two gates: first, a **CX** gate between qubits A and B (CX_{AB}, with A as control and B as target), followed by a **CZ** gate between qubits M and B (CZ_{MB}, with M as control and B as target).

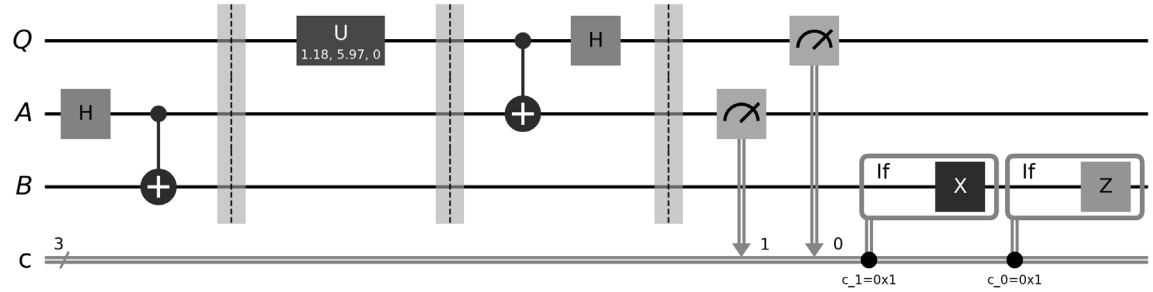


FIGURE 1. A visual representation of TP's quantum circuit. Notice how the last two gates have a conditional operation based on the observation result.

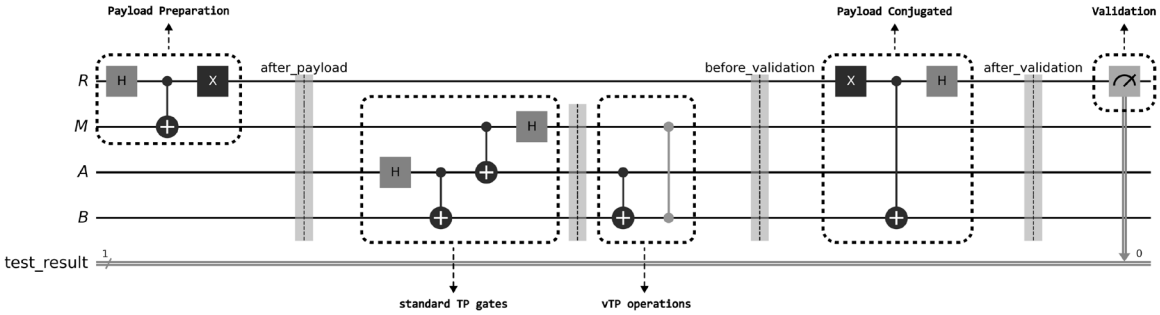


FIGURE 2. The quantum circuit for vTP.

To observe the effect of this new gate sequence on the state $|\Psi_{MAB}\rangle$ of Equation 19, we apply these operations sequentially.

A. EFFECT OF CX_{AB}

Because quantum gates act linearly on superpositions, we can evaluate the action of CX_{AB} on the state in equation 19 term by term. Recall that CX_{AB} flips the target qubit B if the control qubit A is in state $|1\rangle$. Applying it to each computational basis component yields the following.

$$\begin{aligned}
&CX_{AB}\,|00\rangle_{MA}\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B\\
&\quad=|00\rangle_{MA}\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B,\\
&CX_{AB}\,|01\rangle_{MA}\,(\alpha\,|1\rangle+\beta\,|0\rangle)_B\\
&\quad=|01\rangle_{MA}\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B,\\
&CX_{AB}\,|10\rangle_{MA}\,(\alpha\,|0\rangle-\beta\,|1\rangle)_B\\
&\quad=|10\rangle_{MA}\,(\alpha\,|0\rangle-\beta\,|1\rangle)_B,\\
&CX_{AB}\,|11\rangle_{MA}\,(\alpha\,|1\rangle-\beta\,|0\rangle)_B\\
&\quad=|11\rangle_{MA}\,(\alpha\,|0\rangle-\beta\,|1\rangle)_B.
\end{aligned}\tag{2}$$

Combining these, the state after CX_{AB}, denoted $|\Psi'_{MAB}\rangle$, becomes:

$$\begin{aligned}
|\Psi'_{MAB}\rangle=\frac{1}{2}[&|00\rangle_{MA}\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B\\
&+|01\rangle_{MA}\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B\\
&+|10\rangle_{MA}\,(\alpha\,|0\rangle-\beta\,|1\rangle)_B\\
&+|11\rangle_{MA}\,(\alpha\,|0\rangle-\beta\,|1\rangle)_B]
\end{aligned}\tag{3}$$

This state can be factored as follows.

$$\begin{aligned}
|\Psi'_{MAB}\rangle=|+\rangle_A\otimes\frac{1}{\sqrt{2}}[&|0\rangle_M\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B\\
&+|1\rangle_M\,(\alpha\,|0\rangle-\beta\,|1\rangle)_B]
\end{aligned}\tag{4}$$

where $|+\rangle_A=\frac{1}{\sqrt{2}}(|0\rangle_A+|1\rangle_A)$.

B. EFFECT OF CZ_{MB}

Next, we apply CZ_{MB} to the state $|\Psi'_{MAB}\rangle$ from Equation 4. In this case, the $|+\rangle_A$ factor is unaffected.

We focus on the MB part:

$$\begin{aligned}
&CZ_{MB}(\frac{1}{\sqrt{2}}[|0\rangle_M\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B\\
&\quad+|1\rangle_M\,(\alpha\,|0\rangle-\beta\,|1\rangle)_B])\\
&=\frac{1}{\sqrt{2}}[|0\rangle_M\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B+|1\rangle_M\,Z_B(\alpha\,|0\rangle-\beta\,|1\rangle)_B]\\
&=\frac{1}{\sqrt{2}}[|0\rangle_M\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B+|1\rangle_M\,(\alpha\,|0\rangle+\beta\,|1\rangle)_B]\\
&=\frac{1}{\sqrt{2}}(|0\rangle_M+|1\rangle_M)\otimes(\alpha\,|0\rangle+\beta\,|1\rangle)_B\\
&=|+\rangle_M\otimes(\alpha\,|0\rangle+\beta\,|1\rangle)_B
\end{aligned}\tag{5}$$

Therefore, the total final state after applying $CZ_{MB}CX_{AB}$ is:

$$|\Psi''_{MAB}\rangle=|+\rangle_A\otimes|+\rangle_M\otimes(\alpha\,|0\rangle+\beta\,|1\rangle)_B\tag{6}$$

From equation 6, we observe that the quantum state of Bob's qubit (B) is $(\alpha\,|0\rangle + \beta\,|1\rangle)_B$, which is identical to the original state of Alice's message qubit M (as described in equation 13). Furthermore, qubits M and A are now in the $|+\rangle_M$ and $|+\rangle_A$ states, respectively, and all three qubits (M, A, B) are disentangled. This confirms that vTP successfully transfers the initial state of M to B without relying on measurement operations, a classical communication channel, or conditional gate logic, making it suitable for hybrid benchmarks, including those on a single QPU.

It is important to acknowledge a key limitation when translating the preceding theoretical analysis to real hardware. Although the derivations assume ideal gate operations, current NISQ hardware such as IBM Sherbrooke and Rigetti Ankaa-3 introduces errors at each gate that accumulate throughout circuit execution. On these platforms, single-qubit gate error probabilities typically range from 0.01% to 0.1%, while two-qubit gate error probabilities range from 0.1% to 1% [32], [37].

Given these error rates, vTP's reliance on multiple CNOT operations becomes significant. The protocol requires three CNOT gates in the core bridge structure plus one additional CNOT per payload qubit for entanglement. As a result, the cumulative error is expected to grow with the size of the payload, causing the success rate to deviate from the ideal predictions. Nevertheless, the theoretical framework remains essential, since it establishes the correct baseline behavior against which hardware-induced deviations can be measured and understood.

IV. EXPERIMENT

This section outlines the experimental strategy used to address the research question RQ1 presented in Section I, following the methodology detailed in Section II-C. Our approach proceeds in two global phases. We first establish the baseline for the vTP algorithm by detailing its mathematical formulation and comparing the resulting quantum state with the results derived from running the circuit on a noiseless Qiskit Aer simulator. Having established the foundation, we advance to the core experiment, which is structured in three sequential stages aiming to identify correlations between pre-runtime metrics such as gate count, circuit depth, and payload size, with the success rate.

An important aspect for this entire experimental strategy is the restriction of all randomized operations to gates within the Clifford group. The rationale for this restriction is based on the well-established and classifiable mathematical structure of the group [38]. A primary advantage is their inherent efficient simulation on classical hardware; unlike universal quantum circuits, Clifford-based circuits can be simulated on classical computers in polynomial time, which allows for direct corroboration of experimental outputs against a noiseless, theoretical framework. In addition, these gates are integral to quantum error correction, underscoring the importance of their characterization in the advancement of fault-tolerant systems. Because of this, the use of gates within the Clifford group provides a controlled and manageable environment for the development of robust benchmarking methodologies.

A. EXPERIMENTAL HYPOTHESES

Based on our preliminary review of quantum runtime services, we formulate the following hypotheses:

- **H1:** The number of single qubit gates will have a minimal impact on the final success rate.
- **H2:** There exists a negative correlation between the depth of the circuit and the success rate.
- **H3:** The payload size emerges as a significant factor correlated with the success rate.

B. STAGE 1: PREPARATION

Initially, the Qiskit Aer simulator was employed to assess the alignment between the vTP's theoretical forecasts and the simulation results. Figure 2 illustrates the fundamental circuit, while comprehensive details can be found in III.

The experimental setup was designed to execute jobs consisting of 1024 shots each. Upon aggregation of the results, a 100% success rate was observed, which is expected since the simulator did not include any noise model. The primary objective was to measure qubit R after the validation sequence, with an anticipated result of $|0\rangle$. In particular, all the jobs yielded a count of {'0': 1024}.

To compare the results of the simulator with the theoretical analysis, a four-qubit system ($|RMAB\rangle$) should be considered at key stages of circuit execution, as detailed below.

- The quantum state to be teleported, represented by the `'after_payload'` barrier in Figure 1), corresponds to the state vector:

$$|\psi_{\text{ap}}\rangle = \frac{\sqrt{2}i}{2}(|0001\rangle - |0010\rangle) \tag{7}$$

  The state can be written as $|0_R 0_M\rangle \otimes \frac{i}{\sqrt{2}}(|0_A 1_B\rangle - |1_A 0_B\rangle)$.
- After applying the vTP protocol, at the barrier *"before validation"*, the state vector becomes:

$$\begin{aligned}|\psi_{\text{bv}}\rangle = \frac{\sqrt{2}i}{4}(&|0001\rangle + |0011\rangle + |0101\rangle \\ &+ |0111\rangle - |1000\rangle - |1010\rangle \\ &- |1100\rangle - |1110\rangle)\end{aligned} \tag{8}$$

- Finally, after the validation sequence, at *"after_validation"* barrier, the system's state vector is:

$$|\psi_{\text{av}}\rangle = \frac{1}{2}(|0000\rangle + |0010\rangle + |0100\rangle + |0110\rangle) \tag{9}$$

Note that the state can be written as $|0_R\rangle \otimes \frac{1}{\sqrt{2}}(|0_M\rangle + |1_M\rangle) \otimes \frac{1}{\sqrt{2}}(|0_A\rangle + |1_A\rangle) \otimes |0_B\rangle$, which simplifies to $|0_R\rangle\,|+_M\rangle\,|+_A\rangle\,|0_B\rangle$.

The final state $|\psi_{\text{av}}\rangle$ shows that qubit R is in the state $|0\rangle$. For this reason, measurements of qubit R invariably result in '0', thereby confirming the successful implementation and

verification of the vTP. This outcome is consistent with the 100% success rate observed in the simulation.

1) COMPARING SIMULATION WITH MATH ANALYSIS

The simulator data mirror the step-by-step state evolution displayed in Section III (Variation of the Teleportation Protocol) (vTP). In that section, the conceptual state is written for the three logical qubits (M, A, B), whereas the circuit register is ordered (R, M, A, B). Dropping the leftmost bit of every four-bit computational basis label gives us a state that must be compared with Equations 6–19:

- **After payload.** Removing R from the simulated vector $\frac{\sqrt{2}i}{2}(|0001\rangle - |0010\rangle)$ produces $\frac{i}{\sqrt{2}}(|001\rangle - |010\rangle)_{MAB}$, which matches the anti-symmetric Bell component obtained immediately after the payload gate in the theoretical result.
- **Before validation.** Tracing R in $|\psi_{\text{bv}}\rangle$ reproduces the superposition of Eq. 3, the state predicted after the sequence CX_{AB} followed by CZ_{MB}.
- **After validation.** Discarding R from $|\psi_{\text{av}}\rangle$ yields $|+\rangle_M|+\rangle_A|0\rangle_B$, exactly Eq. 6. Hence, Bob's qubit B holds the original message state, while M and A are measured in the $|+\rangle$ state and are disentangled from B, as predicted by theory.

Based on these observations, it is evident that the practical implementation aligns accurately with the predictions of the analytical model.

C. STAGE 2: N RANDOM GATES EXPERIMENT

Having confirmed the correctness of the vTP algorithm, we proceeded to validate the first hypothesis (IV-A). This stage involved a stress test targeting the circuit's reliability on real quantum services (ibm_sherbrooke and Rigetti Ankaa-3) through IBM, qBraid and AWS quantum services. The experiment was designed to progressively insert a number of single qubit operations into the payload, with the number of gates ranging from 0 to 20, 000 and the payload size spanning from 1 to 5 qubits.

To manage the extensive parameter space, jobs were executed in batches, each covering distinct ranges of random gates. For IBM hardware, the selected gate operations ranges were: (200, 205), (500, 505), (1000, 1005), (1500, 1505), (2000, 2005), (3000, 3005), (5000, 5005), (10000, 10005), (20000, 20005). However, for the Rigetti Ankaa-3 machine, the experiment was limited to 20k gates due to provider restrictions, therefore, on that hardware, the maximum range was limited to (7000, 7005). The motivation for these ranges is to perform identical tasks with nearby gate counts while enabling the distinction between ranges.

In addition to determining the appropriate gate count ranges, a crucial aspect of the experimental design was selecting the gate set for random operations. We initially considered including unitary gates not restricted to the Clifford group, but this approach presented three significant challenges: primarily, the conjugate operations necessary for validation were not natively supported by the real hardware. Secondly, determining the conjugate of circuits using gates beyond the Clifford group may not yield straightforward Pauli operations, which creates challenges for implementation on real hardware. Finally, the presence of quantum states beyond the scope of Pauli operations could result in the introduction of additional errors during measurement, as elaborated in Section II.

Considering the issues previously mentioned, we restricted the framework's random component to generate gates exclusively from the Clifford group before conducting the second experiment.

After adapting the code base to exclusively apply random Clifford gates to the payload of the circuit, the experimentation phase continued. Tables 1 and 2 summarize the success rate statistics for vTP after executing circuits with different gate count ranges in IBM Sherbrooke and Rigetti Ankaa-3, respectively. Figure 3 illustrates that while overall success rates remain low and highly variable, a declining trend is observed in gate ranges of ≤ 70.

The American Statistical Association (ASA) defines the p-value as the probability, under a specified statistical model that assumes no correlation, of observing the given data [39].

TABLE 1. Success rate statistics for vTP on IBM Sherbrooke across different gate count ranges.

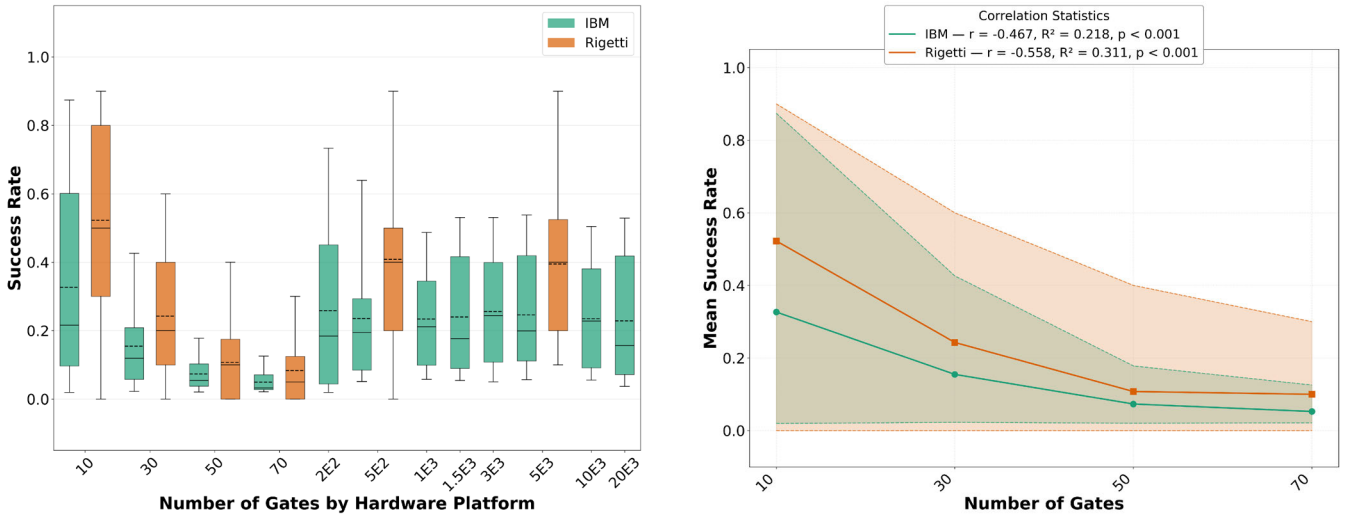
Gate Count Range	Success Rate Range (%)	Mean (%)
10	2.0–87.4	32.7
30	2.3–42.6	15.5
50	2.1–17.8	7.3
70	2.1–12.5	5.0
200	2.0–73.3	25.9
500	5.1–64.0	23.5
1,000	5.8–48.7	23.4
1,500	5.4–53.1	24.0
3,000	5.1–53.1	25.6
5,000	5.7–53.8	24.6
10,000	5.6–50.4	23.5
20,000	3.7–52.9	22.9

TABLE 2. Success rate statistics for vTP on Rigetti Ankaa-3 across different gate count ranges.

Gate Count Range	Success Rate Range (%)	Mean (%)
10	0.0–90.0	52.2
30	0.0–60.0	24.3
50	0.0–40.0	10.8
70	0.0–30.0	8.3
500	0.0–90.0	40.8
5,000	10.0–90.0	39.5

TABLE 3. Correlation statistics between number of gates and success rate for gate counts $\leq$ 70.

Statistic	IBM Sherbrooke	Rigetti Ankaa-3
Pearson r	-0.467	-0.558
Spearman ρ	-0.554	-0.568
R^2	0.218	0.311
Slope	-0.007	-0.009
p-value	6.04×10^{-17}	5.09×10^{-16}



(a) All gate count ranges

(b) Filtered view (gates ≤ 70)

FIGURE 3. Success rate analysis with respect to number of gates. (a) Box plot comparison across all gate count ranges. (b) Filtered correlation view focusing on lower gate counts (≤ 70).

In this context, the calculated p-values ($p < 10^{-15}$) for both platforms reported in Table 3 strongly suggest a mismatch between the observed data and the null hypothesis of zero correlation, indicating that there is a correlation between the gate count and the success rate.

Although the p-values confirm a statistically significant correlation, this overall trend deserves a deeper look. Figure 4 presents a segmented analysis that separates the data by payload size. This analysis reveals that the payload size is the dominant factor that influences the success rate and, at the same time, makes clear that the gate count has only a modest effect (10 - 20%) in the success rate within each category. For instance, on IBM, circuits with 1-qubit payloads maintain relatively high success rates (∼50-60%) regardless of gate count, while 5-qubit payloads consistently exhibit low success rates (∼5%) across all experiments.

The experimental results at this stage partially support our initial hypothesis (IV-A), though the observed correlation is stronger than expected. Although we hypothesized a minimal impact of gate count, the data reveal a moderate negative correlation between gate count and success rate, as summarized in Table 3.

However, the moderate correlation with gate count is overshadowed by a more significant factor. The second key conclusion of this experiment is a strong inverse correlation between payload size and success rate, establishing it as the primary limiting factor. This trend is consistent across hardware platforms, indicating that although the gate count impact is quantifiable, it is less significant compared to the influence of payload size.

To close this section, we concluded that the moderate gate count correlation observed is attributed to the cumulative effect of imperfect single-qubit gate operations, although these operations generally display high fidelity in today's quantum hardware [40]. Our framework determined that for vTP, the use of fewer than 20 sequential single-qubit gates in the payload (40 throughout the complete circuit) is safe to achieve a mean success rate of at least 30%. Data supporting this section can be accessed at https://github.com/xthecapx/qc_experiment/tree/main/experiment/ieee_analysis.

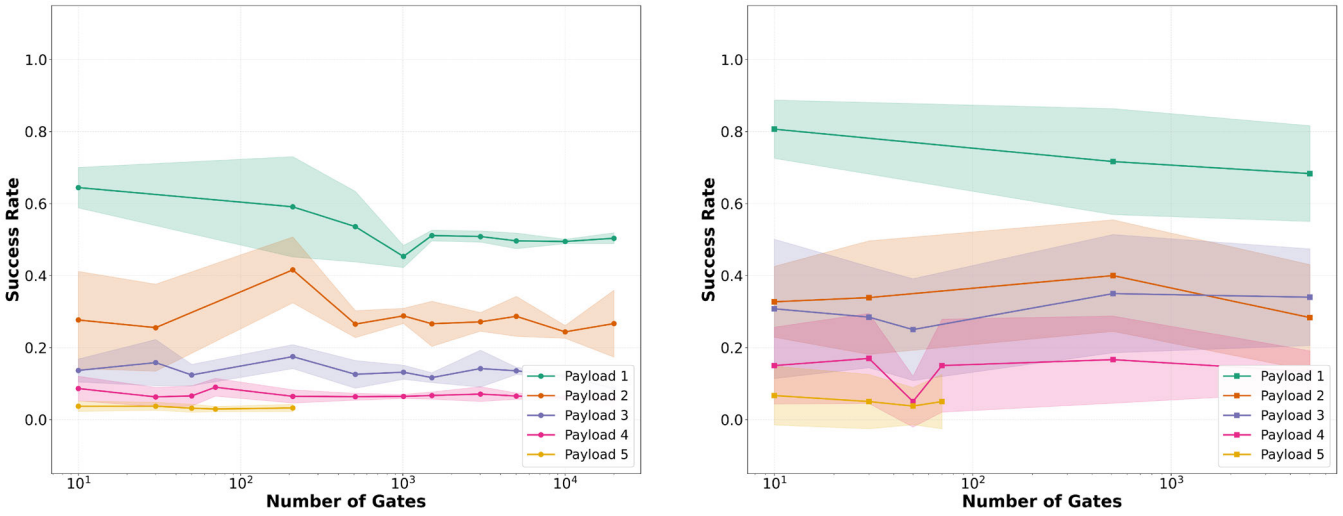
D. STAGE 3: CIRCUIT DEPTH EXPERIMENT

To investigate the impact of circuit depth, our approach was to generate circuits with specific target depths by systematically varying the `payload_size` and `num_gates`. For this purpose, we developed a function that takes a target depth and returns a list of (`payload_size`, `num_gates`) pairs predicted to produce that depth. The function relies on a simple linear model (Equation 10), which is sufficient to generate the varied circuit structures needed for this correlation analysis.

The circuit depth helper function uses a simple linear model, as defined in 10, to generate the list of pairs. Since this research aims to determine the correlations with the success rates, rather than developing an optimal model for generating circuits with specific depths, the simple model serves the purpose.

$$\text{base\_depth} = 13 + 2 \times (\text{payload\_size} - 1) \quad (10)$$

The `base_depth` formula was derived empirically through the process of experimenting with vTP. The 13 value in the formula represents the minimum circuit depth observed for a single-qubit payload (`payload_size` $=$ 1) with a baseline X-gate operation. The linear term ($2 \times$ (`payload_size` -1)) corresponds to the additional circuit depth required for each additional qubit in the payload, where approximately two depth units are added per qubit due to



(a) IBM Sherbrooke (b) Rigetti Ankaa-3

FIGURE 4. Success rate versus gate count segmented by payload size for IBM Sherbrooke and Rigetti Ankaa-3. Each line represents a different payload size, with shaded regions indicating standard deviation. The x-axis is in Log-scale to enabling visualization across the wide gate range tested.

TABLE 4. Success rate statistics for vTP on IBM Sherbrooke across different circuit depth ranges.

Circuit Depth Range	Success Rate Range (%)	Mean (%)
5–9	58.5–87.4	66.6
10–14	6.6–86.5	40.5
15–19	2.0–68.2	13.7
20–24	2.3–69.2	22.3
25–29	2.7–68.3	23.1
30–34	2.1–69.1	24.7
35–39	2.1–69.2	22.2
40–44	2.5–69.3	23.7

TABLE 5. Success rate statistics for vTP on Rigetti Ankaa-3 across different circuit depth ranges.

Circuit Depth Range	Success Rate Range (%)	Mean (%)
10–14	70.0–100.0	84.0
15–19	10.0–90.0	51.3
20–24	0.0–100.0	33.5
25–29	0.0–90.0	30.4
30–34	0.0–100.0	31.0
35–39	0.0–90.0	33.0
40–44	0.0–90.0	31.5
45–49	0.0–100.0	32.3

entanglement operations and random gate operations. This empirical model serves as a simple predictor for circuit depth estimation in preparing experiments with different payload configurations.

Following this, the `target_depth` is defined by taking into account the mean number of supplementary random gates for each payload qubit:

$$\text{target\_depth} = \text{base\_depth} + 2 \times \left(\frac{\text{num\_gates}}{\text{payload\_size}} \right) - 2 \tag{11}$$

TABLE 6. Correlation statistics between circuit depth and success rate at lower depths.

Statistic	IBM Sherbrooke	Rigetti Ankaa-3
Pearson r	-0.714	-0.538
Spearman ρ	-0.761	-0.532
R^2	0.510	0.289
Slope	-0.053	-0.049
p-value	6.04×10^{-28}	9.50×10^{-6}

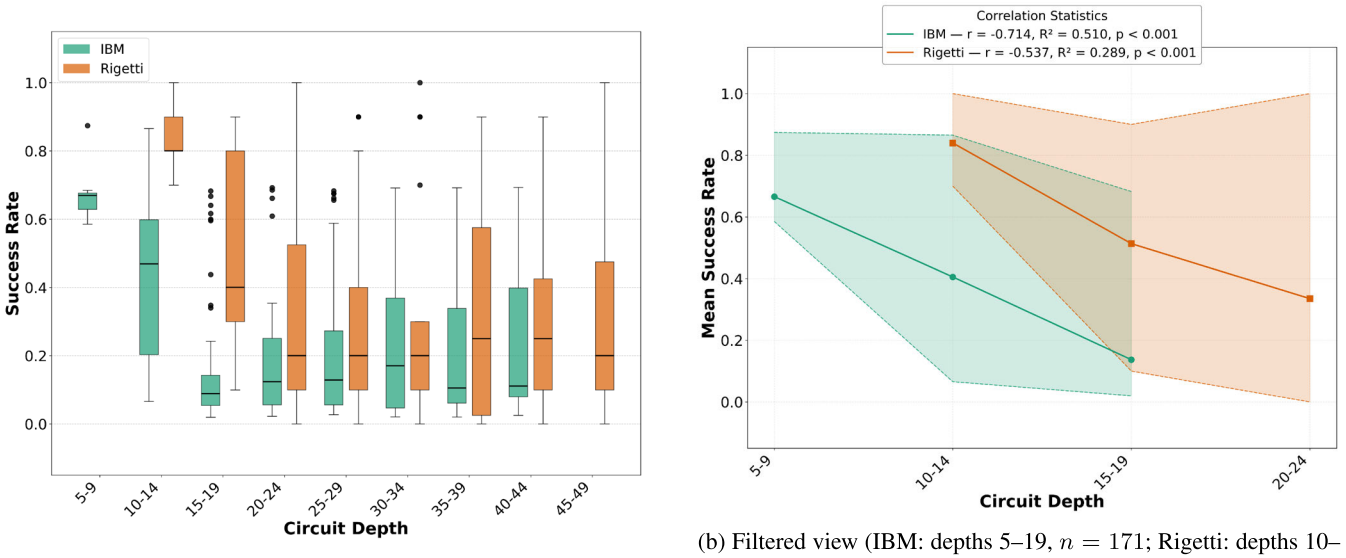
TABLE 7. Correlation statistics between payload size and success rate for payload sizes 1-5.

Statistic	IBM Sherbrooke	Rigetti Ankaa-3
Pearson r	-0.877	-0.831
Spearman ρ	-0.950	-0.843
R^2	0.770	0.690
Slope	-0.144	-0.174
p-value	3.13×10^{-156}	7.53×10^{-59}

Within equation 11, the relation between the number of gates and the payload size serves to estimate the mean depth introduced by the random Clifford gates applied to the payload. The subtraction of 2 accounts for instances where random gates are deployed, replacing the default X-gate (nominal depth of 1, but effectively 2 when the validation sequence is considered) on each payload qubit whose depth contribution was included as part of `base_depth`. To determine number_gates based on predetermined `target_depth` and `payload_size`, Equation 11 is adjusted accordingly.

$$\text{num\_gates} = \frac{(\text{target\_depth} - \text{base\_depth} + 2) \times \text{payload\_size}}{2} \tag{12}$$

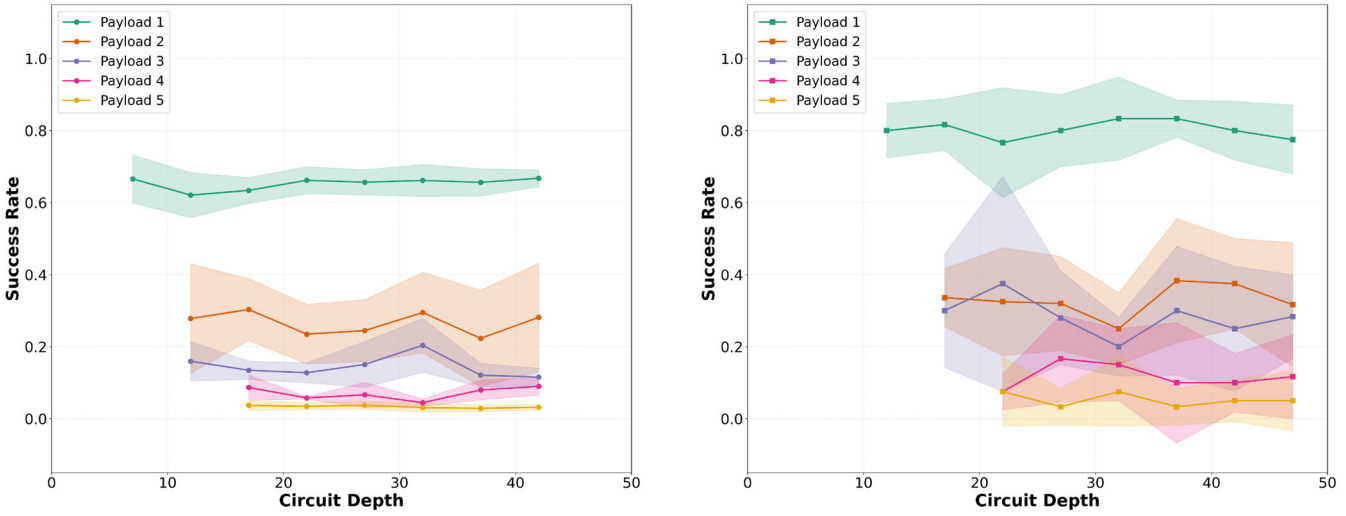
Observe that equation 12 is capable of generating floating-point numbers; consequently, within the code, these values are cast to integers, which results in the truncation of the decimal component without the application of any rounding



(a) Complete dataset across all depth ranges

(b) Filtered view (IBM: depths 5–19, $n = 171$; Rigetti: depths 10–24, $n = 60$)

FIGURE 5. Success rate correlation analysis for circuit depth experiment. (a) Box plot comparison showing all depth ranges. (b) Filtered envelope plot.



(a) IBM Sherbrooke

(b) Rigetti Ankaa-3

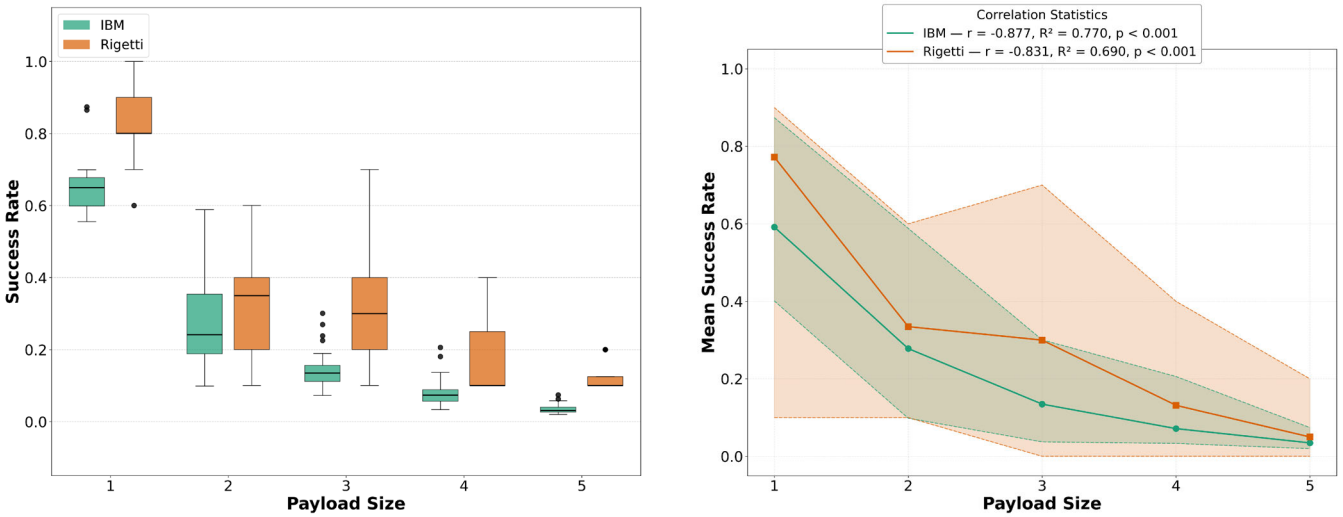
FIGURE 6. Success rate versus circuit depth segmented by payload size for IBM Sherbrooke and Rigetti Ankaa-3. Each line represents a different payload size, with shaded regions indicating standard deviation.

operations. Furthermore, it is important to acknowledge that `base_depth` may potentially exceed `target_depth`; therefore, the code includes a restriction to incorporate only positive values.

It is worth noting that the valid configurations were chosen exclusively when this calculation resulted in a non-negative integer for `num_gates`. This methodology enabled the systematic generation of randomized circuits with consistent target depths. However, it is important to note that the final circuit depth on the quantum hardware varied from our target because of optimizations during the transpilation phase. This discrepancy is a known challenge in quantum programming, where the hardware's compiler can significantly alter a circuit's structure to adapt it to the processor's topology and native gate set.

Tables 4 and 5 summarize the success rate statistics for vTP after executing circuits with different circuit depth ranges in IBM Sherbrooke and Rigetti Ankaa-3, respectively. The data reveal a correlation between the depth of the circuit and the success rate before the cumulative error dominates the results, with the success rates of IBM Sherbrooke dropping from 66. 6% (depth 5-7) to 13. 7% (depth 15-19) and on Rigetti Ankaa-3 dropping from 84. 0% (depth 10-14) to 33. 5% (depth 20-24).

Figure 5 visualizes this inverse relationship through two complementary visualizations, consisting of a boxplot



(a) Success rate distribution by payload size.

(b) Filtered success rate correlation vs payload size.

FIGURE 7. Success rate analysis as a function of payload size for IBM Sherbrooke and Rigetti Ankaa-3. (a) Box plots comparing success rate distributions by payload size across both platforms. (b) Filtered correlation plot with mean success rate envelope, demonstrating a strong negative correlation between payload size and success rate.

TABLE 8. IBM Sherbrooke execution summary.

Stage	Jobs	Shots
1. Hardware baseline	485	1,986,443
2. Gate count	485	1,986,443
3. Circuit depth	291	1,191,936
4. Payload size	485	1,986,443

TABLE 9. Rigetti Ankaa-3 execution summary.

Stage	Jobs	Shots
1. Hardware baseline	225	2,260
2. Gate count	225	2,260
3. Circuit depth	187	1,870
4. Payload size	225	2,260

(Subfigure 5a) showing the complete dataset and a filtered view (Subfigure 5b) focusing on the lower depth ranges where the correlation is most evident. The trend visualized in the figures is statistically validated in Table 6, which provides a comprehensive correlation analysis. The negative Pearson and Spearman correlation coefficients confirm a moderate inverse monotonic relationship, with IBM showing a stronger correlation ($r = -0.714$) than Rigetti ($r = -0.538$). The corresponding p-values ($< 10^{-5}$ for both platforms) indicate that this relationship is highly significant, supporting the interpretation that deeper circuits degrade the success rate before payload size effects become dominant.

To isolate the impact of the depth of the circuit from other variables, Figure 6 presents a segmented analysis based on the size of the payload. This view reinforces the conclusion of Stage IV-C where the payload size remains the dominant factor that influences the success rate. This trend is clearly visible on the IBM platform, where circuits with 1-qubit payloads maintain relatively high success rates (∼65%) regardless of circuit depth, while 5-qubit payloads consistently exhibit low success rates (∼3%) across all experiments.

The results of stage 3 support our second hypothesis (IV-A), confirming a moderate negative correlation between the depth of a quantum circuit and its success rate. This behavior is expected because deeper circuits require more computational steps and, as a result, are more susceptible to time-dependent errors that degrade reliability. This finding is consistent with the results documented by Pérez-Antón et al [41].

This analysis identifies circuit depth as a significant factor in the vTP's performance, second only to payload size. From a practical point of view, our data suggest that to maintain a mean success rate above 40%, the depth of the vTP circuit should be kept below 15 in IBM Sherbrooke and below 25 in Rigetti Ankaa-3. Exceeding these depths leads to a decline in performance, reinforcing that while circuit depth has a clear impact, it is the error introduced by multiqubit operations tied to payload size that remains as the bottleneck.

E. STAGE 4: PAYLOAD SIZE EXPERIMENT

In the final experimental stage, we investigate the impact of increasing the payload size. A critical aspect of this stage is that each additional qubit added to the payload must be entangled with the existing message qubits to form a cohesive quantum state. In our experimental setup, this is achieved by adding a Controlled-NOT gate for each new payload qubit. As a result, this stage is designed not only to study the effect of adding more qubits but, more importantly, to analyze the impact of the corresponding linear increase in high-error two-qubit entangling operations. Also note that for each payload qubit, three stochastic unitary gates from the Clifford group were also incorporated.

TABLE 10. Bob's corrective Pauli operations based on Alice's measurement outcome (communicated classically) to recover the initial quantum message $|\psi\rangle$.

Initial State ($\|\psi\rangle$)	Alice's Measurement	Bob's State (Before Correction)	Operation by Bob	Bob's Final State
$\alpha\|0\rangle + \beta\|1\rangle$	$\|00\rangle_{MA}$	$\alpha\|0\rangle + \beta\|1\rangle$	None (I)	$\alpha\|0\rangle + \beta\|1\rangle$
$\alpha\|0\rangle + \beta\|1\rangle$	$\|01\rangle_{MA}$	$\alpha\|1\rangle + \beta\|0\rangle$	X	$\alpha\|0\rangle + \beta\|1\rangle$
$\alpha\|0\rangle + \beta\|1\rangle$	$\|10\rangle_{MA}$	$\alpha\|0\rangle - \beta\|1\rangle$	Z	$\alpha\|0\rangle + \beta\|1\rangle$
$\alpha\|0\rangle + \beta\|1\rangle$	$\|11\rangle_{MA}$	$\alpha\|1\rangle - \beta\|0\rangle$	XZ	$\alpha\|0\rangle + \beta\|1\rangle$

Figure 7 presents the success rate analysis for Stage 4 using two complementary visualizations. The boxplot in Figure 7a shows a negative trend in success rates as the payload size increases. The filtered correlation analysis in Figure 7b quantifies this relationship, showing a strong negative correlation with payload size for both platforms.

Table 7 presents the correlation statistics for payload size versus success rate. Both platforms exhibit strong negative correlations, IBM showing a Pearson correlation coefficient of -0.877 ($R^2 = 0.770$) and Rigetti showing -0.831 ($R^2 = 0.690$). As explained previously in Stage 2, the p-values ($p < 0.001$) indicate that these correlations are highly statistically significant, providing strong evidence that the payload size is the dominant factor affecting the vTP success rates.

The results of Stage 4 confirm our third hypothesis IV-A. There is a strong negative correlation between payload size and success rate, which establishes payload size as the most significant factor that influences circuit reliability. The p-values for payload size (3.13×10^{-156} for IBM and 7.53×10^{-59} for Rigetti) provide the strongest statistical evidence among all stages, being substantially more extreme than those observed for gate count (Stage 2: 6.04×10^{-17} and 5.09×10^{-16}) and circuit depth (Stage 3: 6.04×10^{-28} and 9.50×10^{-6}). This analysis confirms that the degradation arises not just from including new qubits in the circuit but from the necessary inclusion of lower-fidelity CNOT operations required to entangle the message qubits within the payload.

F. FINAL CONSIDERATIONS

In the last section of our experiment, we present an analysis of the core component used by vTP. Implementing vTP without a payload requires two Hadamard gates, three CNOT gates, one swap operation, and a measurement. Single-qubit operations minimally affect the cumulative error; however, the three CNOT operations significantly affect the circuit success rate. Additionally, read-out error must be considered as it may further distort the output quantum state. As a result, vTP stacks up moderate error in the NISQ era, which explains why useful data are obtained only from payloads of less than three qubits.

It is also important to note that success rates exhibit considerable variance in the results of the experiments. For example, circuits with identical gate counts yield success rates ranging from near 0% to more than 50%. This variability reflects the stochastic nature of both random circuit generation and noise processes inherent to NISQ devices. Although the correlations identified in this study are statistically significant, the wide dispersion in success rates means that they should be interpreted as indicative trends rather than deterministic predictors of circuit outcomes.

Tables 8 and 9 compile the workload breakdown for each experimental stage. Stages 1, 2, and 4 share the same data set, while stage 3 uses a filtered subset limited to circuit depths ≤ 50. The shot counts are reconstructed from measurement count dictionaries returned by each platform.

V. CONCLUSION

This article addresses the critical gap between hardware-level benchmarks and circuit-specific reliability by introducing a new framework. Combining a benchmarking strategy with a variation of the quantum teleportation protocol, our approach enables the identification of correlations between a circuit's pre-runtime characteristics and its post-runtime outcomes.

Beyond experimentation, the research was supported by theoretical analysis and simulations on classical hardware, providing a robust foundation for evaluating the algorithm execution performed on IBM Sherbrooke and Rigetti Ankaa-3 quantum devices.

Our findings suggest that during the NISQ era, quantum algorithm design should prioritize minimizing entangled qubit operations and circuit depth, as these factors negatively impact success. At the same time, we determine that single gate operations are not a concern if limited to fewer than 40 operations.

The major contribution of this work is a framework that systematically measures how pre-runtime metrics influence execution success on real quantum hardware. Specifically, it provides a method to quantify the influence that different circuit components have on overall reliability, offering developers practical, predictive insights into algorithm performance that go beyond traditional hardware-centric benchmarks.

FUTURE WORK

The current study establishes a foundation for future research into a broader range of pre-runtime metrics capable of characterizing quantum circuits. A key objective will be to utilize

these metrics to develop a robust, reduced-dimensionality ''footprint'' for quantum circuits. Such a footprint, enhanced by machine learning and larger datasets, could enable more precise forecasting of post-runtime metrics, thereby providing developers with valuable preliminary insights before QPU execution.

A practical objective is the development of a comprehensive toolkit to assist developers and researchers in capturing pre-runtime metrics and correlating them with post-runtime outcomes derived from various QPUs. Subsequent work will involve rigorously testing an expanded set of pre-runtime metrics to identify correlations with additional post-runtime characteristics, including execution time and computational cost.

Additionally, the vTP algorithm itself is designed with future extensibility in mind. As distributed quantum systems become available, the benchmarking framework and correlations established in this work can be directly adapted. Researchers and developers will be able to transition from the vTP to the canonical TP by replacing the fixed gate sequence with the conditional sequence of gates and the required measurements described by the protocol. This makes the methodology a future-proof strategy for benchmarking not only single QPUs today but also the quantum distributed system of tomorrow.

Finally, our long-term vision is to use the data gathered through this approach to develop artificial intelligence models. As a significant volume of data is acquired, future research will focus on creating machine learning models to forecast the behavior of quantum circuits, especially those whose complexity prevents them from being simulated using classical computational methods.

APPENDIX
MATHEMATICAL DESCRIPTION OF TP

The teleportation protocol proceeds through a series of well-defined quantum operations. The process begins with a formal definition of the quantum state pertaining to Alice's message qubit as:

$$|\psi\rangle_M = \alpha\,|0\rangle + \beta\,|1\rangle \tag{13}$$

Initially, Alice's qubit (A) and Bob's qubit (B) are prepared in the $|0\rangle$ state. To create the necessary entanglement between Alice and Bob, a Hadamard gate is applied to Alice's qubit followed by a CNOT gate with Alice's qubit as control and Bob's as target:

$$\begin{aligned}&\mathrm{CNOT}_{AB}H_A\,|0\rangle_B\,|0\rangle_A\\&= \mathrm{CNOT}_{AB}\,|0\rangle_B\,\frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A)\\&= \mathrm{CNOT}_{AB}\frac{1}{\sqrt{2}}(|00\rangle_{AB} + |01\rangle_{AB})\\&= \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB}) = \left|\Phi^+\right\rangle_{AB}\end{aligned} \tag{14}$$

At this point, the message qubit M is independent of the entangled pair, so the combined three-qubit state can be written as:

$$\begin{aligned}&|\psi\rangle_M \otimes \left|\Phi^+\right\rangle_{AB}\\&= \frac{1}{\sqrt{2}}(\alpha\,|0\rangle_M + \beta\,|1\rangle_M)(|00\rangle_{AB} + |11\rangle_{AB})\end{aligned} \tag{15}$$

Equation 15 denotes the initial state of the three qubits before Alice performs a measurement. Notice that the M qubit remains independent of the entangled pair AB, as they have not yet interacted. This is the starting point for the Alice teleportation operation.

Expanding equation 15 on the computational basis with qubits ordered as M, A, B:

$$|\psi_{MAB}\rangle = \frac{1}{\sqrt{2}}(\alpha\,|000\rangle + \alpha\,|011\rangle + \beta\,|100\rangle + \beta\,|111\rangle) \tag{16}$$

Alice now performs the measurement operation by applying a CNOT gate with M as control and A as target, followed by a Hadamard gate on M. The CNOT operation yields the following.

$$\begin{aligned}&\mathrm{CNOT}_{MA}\,|\psi_{MAB}\rangle\\&= \frac{1}{\sqrt{2}}(\alpha\,|000\rangle + \alpha\,|011\rangle + \beta\,|110\rangle + \beta\,|101\rangle)\end{aligned} \tag{17}$$

Applying the Hadamard gate to qubit M:

$$\begin{aligned}&H_M(\mathrm{CNOT}_{MA}\,|\psi_{MAB}\rangle)\\&= \frac{1}{2}[\alpha(|0\rangle + |1\rangle)\,|00\rangle + \alpha(|0\rangle + |1\rangle)\,|11\rangle\\&\quad + \beta(|0\rangle - |1\rangle)\,|10\rangle + \beta(|0\rangle - |1\rangle)\,|01\rangle]\end{aligned} \tag{18}$$

Expanding and regrouping terms according to the measurement outcomes of qubits M and A:

$$\begin{aligned}|\psi_{MAB}\rangle = \frac{1}{2}[&|00\rangle_{MA}\,(\alpha\,|0\rangle_B + \beta\,|1\rangle_B)\\&+ |01\rangle_{MA}\,(\alpha\,|1\rangle_B + \beta\,|0\rangle_B)\\&+ |10\rangle_{MA}\,(\alpha\,|0\rangle_B - \beta\,|1\rangle_B)\\&+ |11\rangle_{MA}\,(\alpha\,|1\rangle_B - \beta\,|0\rangle_B)]\end{aligned} \tag{19}$$

This expression reveals that when Alice measures qubits M and A, she obtains one of four equally probable outcomes (each with probability $\frac{1}{4}$), and each measurement outcome projects Bob's qubit into a corresponding state. The relationship between Alice's measurement results and Bob's resulting states is summarized in Table 10.

It is worth noting that Equation 19 can be equivalently represented within the Bell basis as:

$$\begin{aligned}|\psi_{MAB}\rangle = \frac{1}{2}[&\left|\Phi^+\right\rangle_{MA}\,(\alpha\,|0\rangle + \beta\,|1\rangle)_B\\&+ \left|\Phi^-\right\rangle_{MA}\,(\alpha\,|0\rangle - \beta\,|1\rangle)_B\\&+ \left|\Psi^+\right\rangle_{MA}\,(\alpha\,|1\rangle + \beta\,|0\rangle)_B\\&+ \left|\Psi^-\right\rangle_{MA}\,(\alpha\,|1\rangle - \beta\,|0\rangle)_B]\end{aligned} \tag{20}$$

where the Bell states are defined as:

$$\left|\Phi^+\right\rangle_{MA} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)_{MA}$$

$$\begin{aligned}
\left|\Phi^{-}\right\rangle_{MA} &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)_{MA} \\
\left|\Psi^{+}\right\rangle_{MA} &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)_{MA} \\
\left|\Psi^{-}\right\rangle_{MA} &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)_{MA}
\end{aligned} \tag{21}$$

It is possible to calculate the inverse transformation from computational basis to Bell basis and update the equation 19.

$$\begin{aligned}
|00\rangle_{MA} &= \frac{1}{\sqrt{2}}(\left|\Phi^{+}\right\rangle + \left|\Phi^{-}\right\rangle)_{MA} \\
|11\rangle_{MA} &= \frac{1}{\sqrt{2}}(\left|\Phi^{+}\right\rangle - \left|\Phi^{-}\right\rangle)_{MA} \\
|01\rangle_{MA} &= \frac{1}{\sqrt{2}}(\left|\Psi^{+}\right\rangle + \left|\Psi^{-}\right\rangle)_{MA} \\
|10\rangle_{MA} &= \frac{1}{\sqrt{2}}(\left|\Psi^{+}\right\rangle - \left|\Psi^{-}\right\rangle)_{MA}
\end{aligned} \tag{22}$$

Equation 20 is the Bell base representation, which highlights the essential framework of the TP.

AI TOOLS AND SERVICES ACKNOWLEDGMENT

The authors would like to thank the use of AI tools and services for manuscript preparation and research assistance. Writefull.com supported language, grammar, and writing improvements; Elicit.com assisted in literature discovery; and Cursor helped with code development, including plotting scripts and quantum circuit understanding. These tools enhanced the authors' capabilities, while the core research, design, data interpretation, and conclusions are the authors' intellectual work.

REFERENCES

[1] R. Acharya et al., "Quantum error correction below the surface code threshold," *Nature*, vol. 638, no. 8052, pp. 920–926, Feb. 2024.
[2] D. Aasen et al., "Roadmap to fault tolerant quantum computation using topological qubit arrays," 2025, *arXiv:2502.12252*.
[3] H. Putterman et al., "Hardware-efficient quantum error correction via concatenated bosonic qubits," *Nature*, vol. 638, no. 8052, pp. 927–934, Feb. 2025.
[4] M. AbuGhanem, "IBM quantum computers: Evolution, performance, and future directions," *J. Supercomput.*, vol. 81, no. 5, p. 687, Apr. 2025.
[5] R. Acharya et al., "Suppressing quantum errors by scaling a surface code logical qubit," *Nature*, vol. 614, no. 7949, pp. 676–681, Feb. 2023.
[6] T. Proctor, K. Young, A. D. Baczewski, and R. Blume-Kohout, "Benchmarking quantum computers," *Nature Rev. Phys.*, vol. 7, no. 2, pp. 105–118, Jan. 2025.
[7] D. C. McKay, I. Hincks, E. J. Pritchett, M. Carroll, L. C. G. Govia, and S. T. Merkel, "Benchmarking quantum processor performance at scale," 2023, *arXiv:2311.05933*.
[8] J. Emerson, R. Alicki, and K. Zyczkowski, "Scalable noise estimation with random unitary operators," *J. Opt. B, Quantum Semiclass. Opt.*, vol. 7, no. 10, pp. S347–S352, Oct. 2005.
[9] E. Magesan, J. M. Gambetta, and J. Emerson, "Scalable and robust randomized benchmarking of quantum processes," *Phys. Rev. Lett.*, vol. 106, no. 18, May 2011, Art. no. 180504.
[10] D. Mills, S. Sivarajah, T. L. Scholten, and R. Duncan, "Application-motivated, holistic benchmarking of a full quantum computing stack," *Quantum*, vol. 5, p. 415, Mar. 2021, doi: 10.22331/q-2021-03-22-415.
[11] E. T. Campbell, B. M. Terhal, and C. Vuillot, "Roads towards fault-tolerant universal quantum computation," *Nature*, vol. 549, no. 7671, pp. 172–179, Sep. 2017.
[12] Y. Tomita and K. M. Svore, "Low-distance surface codes under realistic quantum noise," *Phys. Rev. A, Gen. Phys.*, vol. 90, no. 6, Dec. 2014, Art. no. 062320.
[13] *Systems and Software Engineering-systems and Software Quality Requirements and Evaluation (square)-system and Software Quality Models*, Standard ISO/IEC 25010, 2023.
[14] A. B. Carrillo, P. R. Mateo, and M. R. Monje, "Metrics to evalute fuctional quality: A sistematic review," in *Proc. 7th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2012, pp. 1–6.
[15] A. Acuaviva, D. Aguirre, R. Pena, and M. Sanz, "Benchmarking quantum computers: Towards a standard performance evaluation approach," 2024, *arXiv:2407.10941*.
[16] A. D. Muñoz, M. R. Monje, and M. G. P. Velthuis, "Towards a set of metrics for hybrid (quantum/classical) systems maintainability," *JUCS-J. Universal Comput. Sci.*, vol. 30, no. 1, pp. 25–48, Jan. 2024.
[17] C. Márquez, J. H. O. Silva, and K. Garcés, "Quality concerns when integrating quantum and classical computing: A systematic mapping study," in *Proc. Congresso Ibero-Americano em Engenharia de Softw. (CIbSE)*, May 2025, pp. 380–381. [Online]. Available: https://sol.sbc.org.br/index.php/cibse/article/view/35330
[18] M. Kordzanganeh, M. Buchberger, B. Kyriacou, M. Povolotskii, W. Fischer, A. Kurkin, W. Somogyi, A. Sagingalieva, M. Pflitsch, and A. Melnikov, "Benchmarking simulated and physical quantum processing units using quantum and hybrid algorithms," *Adv. Quantum Technol.*, vol. 6, no. 8, Aug. 2023, Art. no. 2300043.
[19] Y. Zhou, E. M. Stoudenmire, and X. Waintal, "What limits the simulation of quantum computers?" *Phys. Rev. X*, vol. 10, no. 4, Nov. 2020, Art. no. 041038.
[20] D. Lall, A. Agarwal, W. Zhang, L. Lindoy, T. Lindström, S. Webster, S. Hall, N. Chancellor, P. Wallden, R. Garcia-Patron, E. Kashefi, V. Kendon, J. Pritchard, A. Rossi, A. Datta, T. Kapourniotis, K. Georgopoulos, and I. Rungger, "A review and collection of metrics and benchmarks for quantum computers: Definitions, methodologies and software," 2025, *arXiv:2502.06717*.
[21] J. A. Cruz-Lemus, M. Rodríguez, R. Barba-Rojas, and M. Piattini, *Quantum Software Quality Metrics*. Cham, Switzerland: Springer, 2024, pp. 125–142, doi: 10.1007/978-3-031-64136-7_6.
[22] M. Salm, J. Barzen, F. Leymann, and B. Weder, "About a criterion of successfully executing a circuit in the NISQ era: What wd << 1$varepsilon_{eff}$ really means," in *Proc. 1st ACM SIGSOFT Int. Workshop Architectures Paradigms Eng. Quantum Softw.*, Nov. 2020, pp. 10–13. [Online]. Available: https://dl.acm.org/doi/10.1145/3412451.3428498
[23] Y. I. Bogdanov, "Quantum measurements and high-precision control of quantum states," in *Proc. Int. Conf. Micro- Nano-Electronics*, vol. 12157, V. F. Lukichev and K. V. Rudenko, Eds., SPIE/International Society for Optics and Photonics, Jan 2022, Paper 121571V. [Online]. Available: https://doi.org/10.1117/12.2624317
[24] R. Saini, R. Sewada, and A. Arora, "Current challenge and limitations in quantum computation," *J. Nonlinear Anal. Optim.*, vol. 13, no. 1, pp. 26–35, 2023.
[25] E. Moguel, J. Rojo, D. Valencia, J. Berrocal, J. Garcia-Alonso, and J. M. Murillo, "Quantum service-oriented computing: Current landscape and challenges," *Softw. Quality J.*, vol. 30, no. 4, pp. 983–1002, Dec. 2022. [Online]. Available: https://link.springer.com/article/10.1007/s11219-022-09589-y
[26] J. Wang, G. Guo, and Z. Shan, "SoK: Benchmarking the performance of a quantum computer," *Entropy*, vol. 24, no. 10, p. 1467, Oct. 2022.
[27] A. Wack, H. Paik, A. Javadi-Abhari, P. Jurcevic, I. Faro, J. M. Gambetta, and B. R. Johnson, "Quality, speed, and scale: Three key attributes to measure the performance of near-term quantum computers," 2021, *arXiv:2110.14108*.
[28] M. Bandic, P. le Henaff, A. Ovide, P. Escofet, S. Ben Rached, S. Rodrigo, H. van Someren, S. Abadal, E. Alarcon, C. G. Almudever, and S. Feld, "Profiling quantum circuits for their efficient execution on single- and multi-core architectures," 2024, *arXiv:2407.12640*.
[29] 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, Gen. Phys.*, vol. 77, no. 1, Jan. 2008, Art. no. 012307.
[30] R. Youssef, "Measuring and Simulating T1 and T2 for Qubits," Fermi National Accelerator Lab., Batavia, IL, USA, Tech. Rep., 2020.
[31] M. Alam, A. Ash-Saki, and S. Ghosh, "Addressing temporal variations in qubit quality metrics for parameterized quantum circuits," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design (ISLPED)*, Jul. 2019, pp. 1–6.

[32] V. Tripathi, D. Kowsari, K. Saurav, H. Zhang, E. M. Levenson-Falk, and D. A. Lidar, "Benchmarking quantum gates and circuits," *Chem. Rev.*, vol. 125, no. 12, pp. 5745–5775, Jun. 2025.
[33] P. Malhotra, A. Kumar, and S. Garhwal, "A systematic review of quantum benchmarking," *Int. J. Theor. Phys.*, vol. 63, no. 11, pp. 1–34, Oct. 2024.
[34] T. Chasseur, D. M. Reich, C. P. Koch, and F. K. Wilhelm, "Hybrid benchmarking of arbitrary quantum gates," *Phys. Rev. A, Gen. Phys.*, vol. 95, no. 6, Jun. 2017, Art. no. 062335.
[35] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Phys. Rev. Lett.*, vol. 70, no. 13, pp. 1895–1899, Mar. 1993.
[36] J. Ren et al., "Ground-to-satellite quantum teleportation," *Nature*, vol. 549, no. 7670, pp. 70–73, Sep. 2017.
[37] F. Arute et al., "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.
[38] D. Grier and L. Schaeffer, "The classification of Clifford gates over qubits," *Quantum*, vol. 6, p. 734, Jun. 2022.
[39] R. L. Wasserstein and N. A. Lazar, "The asa statement on p-values: Context, process, and purpose," *Amer. Statistician*, vol. 70, no. 2, pp. 129–133, 2016.
[40] A. Shukla, M. Sisodia, and A. Pathak, "Complete characterization of the directly implementable quantum gates used in the IBM quantum processors," *Phys. Lett. A*, vol. 384, no. 18, Jun. 2020, Art. no. 126387.
[41] R. Pérez-Antón, A. Corbi, J. I. L. Sánchez, and D. Burgos, "Reliability of IBM's public quantum computers," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 7, no. 1, pp. 66–78, Jan. 2021.

CRISTIAN MÁRQUEZ received the bachelor's degree in electronic engineering and the master's degree in engineering from the Universidad Nacional de Colombia, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree in engineering, doing research on quantum computing. He is an experienced Software Engineer with over a decade's worth of experience developing and leading teams to develop and deploy web applications services using JavaScript and Python. In 2022, he resumed his research career by co-authoring a book titled *Modelos Matemáticos Para la Gestión Curricular*. His research interests include distributed systems, software architectures, and quantum technologies.

DANIEL SIERRA-SOSA (Member, IEEE) is currently an Assistant Professor with the Department of Computer Science, The Catholic University of America. He is also an Active Researcher in the fields of quantum computing, machine learning, healthcare data processing, image processing, and data analytics. He has been involved in multiple research activities, including industry contracts, collaborations with public health entities, and other academic partners; in addition to his involvement in academic research proposals. Prior to joining Catholic University, he participated in a collaborative industry initiative with a healthcare company, working on projects that included mobile application development, virtual reality, medical imaging, and predictive analytics. He has led various projects and has taken significant responsibilities in mentoring graduate and undergraduate students. He is the co-author and the lead author of several manuscripts published in recognized journals. He is also a Qiskit Advocate and a Certified Instructor in quantum computing, data science, and artificial intelligence.

KELLY GARCÉS received the Ph.D. degree from the Université de Nantes, in September 2010. She is currently an Associate Professor with the Department of Systems and Computing Engineering, Universidad de los Andes, Bogotá, Colombia. Prior to this, she was a Research and Development Engineer and a Software Engineer with Netfective Technology. In 2011, she was a Postdoctoral Fellow with the INRIA Laboratory. She has been participating in research and development projects (proprietary or open source), since 2005, financed with private and public funds. Her research interests include software architecture, evolution, and maintenance of complex software systems (e.g., legacy applications, the IoT systems, quantum systems, and microservice-based applications). Some of her lectures are: software architecture and design, the software for Internet of Things (IoT), software modeling, software modernization, and model driven engineering.