

Randomized and Diverse Input State Generation for Quantum Program Testing

Maryse Ernzer

University of Luxembourg
Luxembourg, Luxembourg
maryse.ernzer@uni.lu

Seung Yeob Shin

University of Luxembourg
Luxembourg, Luxembourg
seungyeob.shin@uni.lu

Fabrizio Pastore

University of Luxembourg
Luxembourg, Luxembourg
fabrizio.pastore@uni.lu

Domenico Bianculli

University of Luxembourg
Luxembourg, Luxembourg
domenico.bianculli@uni.lu

Abstract—With the accelerating development of quantum technologies and their growing computational potential, quantum systems are being adapted for simulations and other critical tasks across diverse domains, making the reliability of the corresponding quantum software an essential concern. Although recent efforts have started to incorporate quantum-specific properties such as magnitude, phase, and entanglement under the form of input-coverage criteria into software testing, the unique structure of the quantum state space demands for more comprehensive testing. In particular, the notion of complete state-space exploration has so far received little attention.

To address this gap, we propose a framework for evaluating test circuit generators with respect to their coverage of the quantum state space. Our contribution is threefold: we develop a set of diversity scores that capture both local and global indicators of the extent to which the state space is explored; we propose a test circuit generator that produces test input states via a Brick-Circuit (BC) construction designed to approximate ideal random states using hardware-compatible gates; we compare the proposed construction with existing generators based on their ability to generate uniformly distributed random test input states.

Our extended diversity scores quantify the local correlations and global spread of magnitude, phase and entanglement. Using these scores, we evaluate the expressibility, defined as the capability to span the quantum state space uniformly, and entangling capabilities of existing generators relative to the BC generator. Our results show that the hardware-compatible BC generator achieves higher expressibility and entanglement performance at shallower depths than existing circuit generators.

I. INTRODUCTION

With the continued progress of quantum technologies, software development evolves in parallel with its hardware counterpart. Given the novelty of the field, quantum software is particularly susceptible to programming mistakes, underscoring the need for systematic approaches to reliability. Additionally, the inherent characteristics of quantum systems, such as probabilistic outcomes, sensitivity to measurement, and the exponentially large state space, impose challenges on the testing methods that go beyond those of classical systems and demand for adapted testing techniques [1].

Recent approaches aim to address these challenges by exploiting quantum-specific properties such as superpositions, entanglement and reversibility [2]–[4]. These works indicate that incorporation of quantum-specific properties improves fault detection capabilities, especially by uncovering quantum-specific faults [5]–[8].

To further enhance fault detection and overall testing effectiveness in a systematic manner, test case diversity has been investigated. Early works introduced coverage criteria and test input optimization methods, along with systematic frameworks, search-based techniques incorporating quantum features and fuzzing-based state-space exploration for fault detection [9]–[13]. These studies also highlight the value of quantum-aware input coverage, and demonstrate the effectiveness of including state superpositions and entanglement to ensure test thoroughness. More recent approaches explicitly aim to quantify and increase the diversity of generated test inputs, addressing the challenge of achieving effective coverage of the quantum input state space and of generating test input states that exhibit genuinely quantum properties, which typically require nontrivial quantum circuit constructions [14]. In the latter, the local state properties (e.g., magnitude and phase) and non-local correlation properties (e.g., entanglement), are directly considered and formulated as diversity scores quantifying the property distributions across bipartite qubit sectors. These scores are used to measure the input coverage achieved by certain test circuit generators. The test generators follow different patterns, including simple concatenation of random gates, sets of random rotation combined with entangling gates and circuits based on the inverse quantum Fourier transforms.

While the methods described above address some manifestations of quantum states such as phase, superposition, and entanglement, a key aspect of quantum mechanics—namely the complex structure of the underlying state space—has so far been mostly overlooked in quantum software testing (QST). Indeed, for multi-qubit systems, the dimensionality of the state space has to be considered in the diversity assessment, since generating test input states with a wide distribution of magnitude and phase does not prevent the state from being confined to a small subset of the entire (Hilbert) state volume. To assess expressibility, defined by the ability to span the entire state space and generate entanglement, more informative indicators than magnitude and phase have to be considered. For example, recent works have established a direct link between maximizing state-space exploration and the uniformity of the generated states under the form of Haar-distributed ideal random states [15]–[17] which can be approximated via circuits implementing unitary-2 designs [18]–[21].

We propose to advance existing QST approaches along these lines, by integrating state-of-the-art techniques developed within the quantum information processing community [15], [16]. Specifically, to ensure that test input states span a substantial part of the state-space volume, we propose to extend diversity scores such that they capture not only local features but also multi-qubit correlations, in order to establish a comprehensive expressibility profile of test circuit generators, which produce test input states. These scores include a general estimation of state participation and distinguish between structured subsystems and global properties, addressing the need for more fine-grained, informative coverage measures in QST. We verify their capability in capturing the targeted properties by applying them to well-defined reference quantum states.

Moreover, to guarantee high coverage through an unbiased exploration of the quantum state space, we implement an efficient test circuit generator based on the Brick-Circuit (BC) construction [18], [19], [21], a quantum circuit pattern designed to approximate the Haar-random distribution using only nearest-neighbor 2-qubit entangling gates, ensuring feasibility on the current quantum processors. We compare the degree of randomness of the circuit matrices generated via the hardware-compatible BC construction to three different construction methods illustrated in the state of the art [14], namely the Random, UCNOT and iQFT circuit designs. We evaluate the generated randomness through the statistical proximity of the resulting unitary ensembles to ideal Haar-distributed samples. Our results show that our BC generator saturates the randomness threshold with fewer gates than previously used test circuit generators. We also observe that circuits constructed by combining only a few randomly drawn gates exhibit insufficient randomness.

Finally, we compare the proposed BC construction with existing test circuit generators, by analyzing the diversity of the test input states produced by their respective circuit designs, as quantified by our set of diversity scores, thereby addressing current limitations in evaluating the coverage and expressibility of quantum circuit generators. Therefore, we apply the extended diversity scores to quantify the performance of the circuit generators in expressibility and entangling capabilities. To this end, we consider qubit-sector occupations, local phase and amplitude correlations, effective state participation, as well as single-qubit and delocalized entanglement. We find that the BC construction consistently produces test input states that reach the highest diversity scores, closest to the expected values of ideal random states.

In summary, the main contributions of this paper are:

- We extend diversity scores to capture essential quantum state features.
- We investigate the BC construction, known for its efficient random state generation, as a test circuit generator for diverse test input states.
- We evaluate different test circuit generators based on their expressibility and entangling capabilities.

Our results pave the way towards effective and comprehensive test circuit generation for QST. In particular, they combine

a circuit generator that produces test input states which efficiently approximate Haar-random states with diversity scores that capture expressibility and entangling capabilities, enabling a more complete exploration of the admissible quantum state space during testing.

The rest of the paper is organized as follows: Section II provides some background information; Section III presents the extended diversity scores and the proposed test circuit generator. In Section IV, we evaluate and compare our developed diversity scores and test circuit generator with existing techniques. In Section V we review related work. Section VI draws conclusions and outlines future work.

II. BACKGROUND

A. Quantum States

The general state of a n -qubit system belongs to the complex vector space $\mathbb{C}^d$ of dimension $d = 2^n$ growing exponentially in the qubit number n . Let $\{|j\rangle\}_{j=0}^{2^n-1} = \{|j_0 j_1 \dots j_{n-1}\rangle \mid j_k \in \{0, 1\}\}$ be the orthonormal computational basis where each state is identified by its binary expansion $j = \sum_{k=0}^{n-1} j_k 2^{n-1-k}$; an arbitrary pure state can be described in terms of its *state vector* $|\psi\rangle = \sum_{j=0}^{2^n-1} \alpha_j |j\rangle$ with the *complex amplitudes* $\alpha_j = r_j e^{i\phi_j}$ associated with the state $|j\rangle$ which can be decomposed in terms of *magnitude* $r_j = |\alpha_j|$ and *phase* $\phi_j \in [0, 2\pi)$. The state is necessarily normalized by $\sum_{j=0}^{2^n-1} |\alpha_j|^2 = 1$, where the terms $|\alpha_j|^2$ express the probability of observing the basis state $|j\rangle$ upon measurement in the computational basis.

As a consequence of maintaining multi-qubit phase correlations within the complex state structure of quantum systems, non-classical correlation properties such as *quantum entanglement* emerge. These result in a non-separability of multi-qubit states, i.e., the combined state of subsystems A and B cannot be written as the separable product of its constituents $|\psi\rangle_{AB} = |\psi_A\rangle \otimes |\psi_B\rangle$.

Illustrative examples representing two classes of strongly entangled states are given by the Greenberg-Horne-Zeilinger (GHZ) state incorporating cat-state entanglement and the W state implementing delocalized single-excitation entanglement [22], [23], defined as follows:

$$\psi_{\text{GHZ}} = \frac{1}{\sqrt{2}} (|0\rangle^{\otimes n} + |1\rangle^{\otimes n}), \quad (1)$$

$$\psi_{\text{W}} = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} |0 \dots 0 1_k 0 \dots 0\rangle. \quad (2)$$

B. Entanglement measures

Since quantum entanglement is a primary resource for quantum computation, various measures have been developed to quantify the entanglement strength. Such measures employ different approaches depending on the aspects they aim to capture and the system configuration; they have to satisfy monotonicity under Local Operations and Classical Communication (LOCC), return 0 if and only if the considered states are separable, and 1 for maximally-entangled states.

We consider two measures that evaluate the strength of the entanglement via the presence of entropy in the subsystems. The *von Neumann entropy* captures the average strength of the entanglement in the system in terms of the density matrix $\rho = |\psi\rangle\langle\psi|$ of state $|\psi\rangle$ [24], [25], and is defined as

$$S(\rho) = -\text{Tr}(\rho \log \rho). \quad (3)$$

An alternative entanglement measure consists in the *Rényi-2 entropy*, which quantifies the entanglement between a subsystem A and its complement [25], [26]. It is defined as

$$S_2(A) = -\log \text{Tr}(\rho_A^2), \quad (4)$$

where ρ_A is the reduced density matrix of the considered subsystem A. Unlike the von Neumann entropy, this measure does not require diagonalization and the knowledge of the full eigenvalue spectrum and is based solely on easily accessible measured quantities. As a result, it is particularly attractive for its numerical and experimental practicality [27]–[29]. Through its sensitivity to the chosen partition, the Rényi-2 entropy provides detailed insights into the extent and delocalization of entanglement in multipartite subsystems; moreover, since it does not rely on averaging, it is especially sensitive to extreme features in the entanglement spectrum.

Considering the global spread of entanglement across all qubits, a widely used measure consists of the *Meyer-Wallach (MW) measure* [30], [31]. It returns the average value of the entanglement strength between each qubit and the rest of the system; for a pure state $|\psi\rangle$ it is defined as

$$Q(|\psi\rangle) = 2\left(1 - \frac{1}{n} \sum_{k=1}^n \text{Tr}(\rho_k^2)\right), \quad (5)$$

for a n -qubit system and ρ_k the reduced density matrix of the k -th qubit. It is especially interesting for its computational efficiency, scalability and its ability to indicate the onset and local saturation of entanglement [32]. The average value of the MW measure has been used to estimate the entangling capability of quantum circuits [15].

C. Unitary Description

The general closed evolution of quantum states is reversible and can be expressed by a *unitary matrix* U , such that the resulting state can be written as $|\psi\rangle = U|\psi_{\text{in}}\rangle$ in terms of the initial state $|\psi_{\text{in}}\rangle$. Within the scope of quantum computing, this means that the combined effect of all gates acting on a target state can be summarized via the matrix product of the individual unitary representation acting on an initial state, which in the default case is given by $|\psi_{\text{in}}\rangle = |0\rangle^{\otimes n}$.

The *unitary circuit* is the matrix obtained by multiplication of the applied gate sequence. Uniform sampling of the quantum state space therefore corresponds to sampling uniformly from the space of unitary matrices, i.e., drawing a random matrix. A uniform distribution over all unitaries must be invariant under all unitary transformations. This property uniquely characterizes the *Haar measure* μ , defined by its left- and right-invariance [33]. For every measurable subset

$S \subseteq U(N)$ and every matrix $M \in U(N)$ the probability measure μ must satisfy $\mu(SM) = \mu(MS) = \mu(S)$ with normalization $\mu(U(N)) = 1$. Hence, the Haar measure assigns equal weight to all regions, which can be used to derive randomized states as explained in the following paragraph. Furthermore, the significant properties of Haar-random states that we consider, such as amplitude and phase correlations as well as entropy, are defined by at most second order statistical moments, and thus are captured by a unitary-2 design approximation [18], [20], [21].

D. Randomized states

For an n -qubit system, we sample a Haar matrix $U_{\text{Haar}} \sim \mu_n$ where μ_n denotes the Haar measure on $U(2^n)$ and define differently *structured random states*, such as generic Haar-random states or randomized maximally-entangled bipartite states

$$|\psi_{\text{hr}}\rangle = U_{\text{Haar}} |0\rangle^{\otimes n}, \quad (6)$$

$$|\psi_{\text{me}}\rangle = (I_A \otimes U_B) \frac{1}{\sqrt{2^{n_A}}} \sum_{a=0}^{2^{n_A}-1} |a\rangle_A \otimes |a\rangle_B, \quad (7)$$

with the identity matrix I_A of the subsystem A of size n_A , and a Haar matrix $U_B \sim \mu_{n_B}$ applied to subsystem B of size $n_B = n - n_A$. Further interesting classes of random states consist of uniform flat-amplitude states and random product states defined as

$$|\psi_{\text{ua}}\rangle = \frac{1}{\sqrt{2^n}} \sum_{j=0}^{2^n-1} e^{i\phi_j} |j\rangle, \quad (8)$$

$$|\psi_{\text{p}}\rangle = \bigotimes_{k=1}^n \left(\cos \frac{\theta_k}{2} |0\rangle + e^{i\phi_k} \sin \frac{\theta_k}{2} |1\rangle \right), \quad (9)$$

with random angles $\theta_k \in [0, \pi]$, $\phi_k \in [0, 2\pi]$.

E. Expressibility

Expressibility of quantum circuits refers to their ability to represent any possible state of the underlying complex Hilbert space. Since this corresponds exactly to the capability of generating a uniform distribution of states, expressibility is directly related to the generation of Haar-random states. Together with the entangling capability, expressibility has been identified as one of the two key circuit descriptors that provide a quantitative characterization of parametrized quantum circuits [15]. In the context of test-case generation, expressibility therefore captures the diversity of test input states with respect to the entire accessible state space, and has been directly linked to proximity to Haar-random states and the presence of entanglement [15], [34], [35].

III. DIVERSITY SCORES AND TEST GENERATION

This section presents the six diversity scores that we are going to use for characterizing test input states with respect to their magnitude and phase. Building on measures developed in statistical mechanics and chaos theory, we transfer and adapt a set of diversity scores to quantify the expressibility of test input

states [14], [33], [36]–[39]. Furthermore, we describe the test circuit generator that we adapted from studies in theoretical and experimental quantum physics [18], [19], [21], [40], [41], to achieve our goal of efficiently producing test input states.

A. Diversity Scores

To quantify the diversity of a state generated by a test circuit, we characterize the exploration of the quantum state space by a set of scores addressing the magnitude and phase of the individual state, analogous and complementary to the established entanglement measures described above.

We start by considering the magnitude and phase scores originally introduced in QuraTest [14], which capture how evenly the magnitude and phase contributions are distributed between the ensembles of states forming two sectors defined by a given qubit being in state 0 or 1 with all other qubit entries kept fixed. Although this provides insight into the balance of magnitude and phase via one-to-one correlations between the two sectors, it neglects the more complex aspects of all-to-all correlations within the amplitude and phase distribution.

To establish a comprehensive set of scores that describes the expressibility of a given state, we reformulate and formalize the original QuraTest scores in a unified framework, which we further complement with correlation and global spread scores. We express the scores via the amplitudes α_j of the state vector $|\psi\rangle$ resulting from applying the circuit unitary U to the default initial state $|0\rangle^{\otimes n}$. For each qubit k , we define a binary partition of the state space into two qubit sectors corresponding to the ensembles of computational basis states for which the k -th qubit is fixed to 0 or 1, respectively. Following the convention of notation introduced in Sec. II-A, we define the sectors $S_0^k = \text{span}\{|j\rangle : j_k = 0\}$ and $S_1^k = \text{span}\{|j\rangle : j_k = 1\}$, where “span” indicates the linear span operator. The vectors in opposite sectors are separated by the index value $p_k = 2^{n-1-k}$. Furthermore, besides the general support $\mathcal{I} = \{j : |\alpha_j| \neq 0\}$ with dimension $N = |\mathcal{I}|$, for each qubit k , we define the support for pairs of states with non-zero amplitude mass $\mathcal{I}_p^k = \{j : j_k = 0, |\alpha_j| \neq 0, |\alpha_{j+p_k}| \neq 0\}$ resulting in the number of pairs of contributing basis states $N_p^k = |\mathcal{I}_p^k|$. For states either being 0 or 1 we define the supports $\mathcal{I}_0^k = \{j : j_k = 0, |\alpha_j| \neq 0\}$ and $\mathcal{I}_1^k = \{j : j_k = 1, |\alpha_j| \neq 0\}$ with dimensions $N_0^k = |\mathcal{I}_0^k|$ and $N_1^k = |\mathcal{I}_1^k|$.

1) *Magnitude measures:* For a given qubit k in a n -qubit system, we define the **Magnitude Score** as the difference between the total probability weight between the sectors S_0^k and S_1^k at a index distance p_k , such that

$$\text{MS}_k = \frac{1}{2} \left(1 + \sum_{j \in S_0^k} |\alpha_{j+p_k}|^2 - |\alpha_j|^2 \right). \quad (10)$$

The normalization of the score is directly enforced by the sum of probabilities resulting in 1. To ensure a consistent comparison between scores, we apply a shift for it to cover the interval $[0,1]$. This score measures the population balance between the qubit sectors S_0^k and S_1^k , expected to converge to $\text{MS}_k \sim 0.5$ for uniformly distributed magnitudes. In contrast,

if all the probability mass is either concentrated in the state $|0\rangle^{\otimes n}$ or $|1\rangle^{\otimes n}$, MS_k will return the values 0 or 1.

To get a more differentiated assessment, we expand the one-to-one correlation measure described above to a more general all-to-all correlations of magnitudes. Therefore, we sum the pairwise product of magnitudes over both sectors S_0^k and S_1^k , such that the **Magnitude Correlation Score** can be written as

$$\text{MCS}_k = \frac{1}{2^{n-2}} \sum_{i \in S_0^k} \sum_{j \in S_1^k} |\alpha_i| |\alpha_j|. \quad (11)$$

This allows us to evaluate the distribution of significant amplitude mass carried across sectors, and provides a measure of global amplitude overlap between qubit sectors. Since the amplitude of Haar-random states follow the Porter-Thomas distribution, we expect MCS_k to approach $\pi/4 \sim 0.78$ for random states [35], whereas for states that contain only mass in one of the two sectors, we expect MCS_k to return 0, and for uniform-amplitude states we expect it to return 1.

To measure the global localization determined by the complete probability weight distribution, we rely on the effective dimensionality. This is a fourth order-participation measure, corresponding to the amount of basis states effectively participating in a given system state. Therefore, we define the **Magnitude Dimensionality Score** for the complete n -qubit system as

$$\text{MDS} = \frac{1}{D} \frac{1}{\sum_{j=0}^{2^n-1} |\alpha_j|^4}, \quad (12)$$

normalized in terms of the system’s dimensionality $D = 2^n$. This score provides a direct measure of the amplitude delocalization within the system. For a Haar-random state we expect MDS to converge towards $1/2$, meaning that, on average, half of all possible states participate significantly in a given state. We note here that the dimensionality score is expected to return 1 for uniform flat-amplitude states, and $1/D$ for states with only a single basis state contribution.

In summary, the described magnitude scores constitute three complementary measures corresponding to the first, second, and fourth moments (exponents) of the magnitude, capturing different aspects of a state’s amplitude distribution.

2) *Phase measures:* Analogously to the magnitude, we consider the phase distribution through a treatment of the pairwise angular phase comparison, relying on the relevant contributions of the k -th qubit across sectors S_0^k and S_1^k at the index distance p_k respectively. We can formulate the **Phase Score** as

$$\text{PS}_k = \frac{1}{N_p^k} \sum_{j \in \mathcal{I}_p^k} |\varphi_j - \varphi_{j+p_k}| \quad (13)$$

with the phase at the i -th index $\varphi_j = \arg(\alpha_j)$, and the number of contributing basis state pairs N_p^k . In practice, we compute the phase difference modulo 2π such that this score measures a circular distance. With this expression, we measure the one-to-one phase distance between paired qubit states differing only in qubit k . For a perfectly random state we expect PS_k to

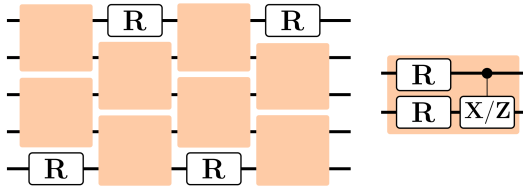


Fig. 1: Circuit diagram of BC construction pattern applied on alternating layers (left), and brick composition (right) in terms of the single-qubit random rotation $R = R_Z(\phi)R_X(\theta)R_Z(\omega)$ where ϕ, ω are drawn randomly from the interval $[0, 2\pi)$, $\cos \theta$ from $[-1, 1]$, and 2-qubit controlled X or Z gates.

be balanced and peaked at 0.5, for states where every paired amplitude has a relative phase of 0, we expect it to return 0, and for a relative phase of π , we expect it to return 1.

In order to generalize phase difference between sectors to an all-to-all correlation measure, we consider the angular difference between all states across the sectors S_0^k and S_1^k , and define the **Phase Correlation Score** as

$$\text{PCS}_k = 1 - \left| \frac{1}{N_0^k N_1^k} \sum_{i \in \mathcal{I}_0^k} \sum_{j \in \mathcal{I}_1^k} \cos(\varphi_i - \varphi_j) \right|. \quad (14)$$

This score provides a measure for cross-sector phase decorrelation. We solely consider the phases of the elements that have a non-zero amplitude mass, to avoid contributions from empty entries. For a Haar-random state we expect PCS_k to converge towards 1, indicating perfect phase independence, whereas for a state with perfectly aligned phases we expect it to return 0.

To evaluate the global spread of the phase and verify its uniformity, we consider the circular **Phase Variance Score** defined as

$$\text{PVS} = 1 - \left| \frac{1}{N} \sum_{j \in \mathcal{I}} e^{i\varphi_j} \right|, \quad (15)$$

with the phase term $e^{i\varphi_j} = \alpha_j / |\alpha_j|$. This score measures the global phase alignment of contributing basis states with non-zero amplitudes. For Haar-random states we expect the PVS to approach 1 up to a factor $1/\sqrt{D}$, and for phase-coherent states we expect it to tend towards 0.

In summary, the three phase scores we propose capture the phase alignment at different granularities, between cross-sector pairs, general cross-sector states or considering the states globally. Overall, the ensemble of presented scores characterizes state expressibility, including structural aspects, such as amplitude correlations and delocalization as well as phase decorrelation and variance.

B. Brick-Circuit (BC) construction for random test input state generation

The diversity scores introduced above, together with the entanglement measures presented in sec. II-B quantify quantum properties such as magnitude, phase, and entanglement. Together they form a comprehensive set to characterize the

overall diversity present in quantum test input states. Furthermore, it has been shown that the expressibility (ability to explore the entire admissible state space, cf. II-E) of a generated test state is directly related to its proximity to Haar randomness [15]–[17]. Accordingly, we expect test circuit generators whose output states closely approximate Haar-random states to be particularly effective for expressive test input state generation, guiding our choice of the appropriate test circuit generator for QST.

Since generating true Haar-random matrices using quantum circuits is computationally expensive, efficient constructions of approximate random unitaries have been investigated in the literature. Ideas for using a two-step construction with parallel gates [18], [20], employing only hardware-compatible nearest-neighbor-gates [19], [21], as well as applying short circuit templates on alternating layers [21], [34], [41], [42] have been put forward.

In our work, we propose combining parallel local random circuits on alternating layers with unitary Haar-random single-qubit rotations into the Brick-Circuit (BC) construction shown in Fig. 1, aiming to implement an efficient circuit design to generate approximate Haar-random states. Indeed, this particular circuit construction follows the general designs that have been studied within the fields of random matrices and shown to efficiently approximate unitary 2-designs [16], [18]–[21], [40], [42], [43], even at very shallow circuit depths [41]. Our BC construction implements alternating layers of spherically uniform single-qubit Haar rotations and 2-qubit entangling gates, i.e., controlled X or Z gates, such that the circuit diagram takes the shape of a brick wall as shown in Fig. 1.

Note that our BC construction aligns with an established experimental construction scheme based on applying random rotation to each qubit individually, and entangling gates implementing controlled gates between adjacent qubits. The employed gate set is compatible with the nearest-neighbour connectivity often found in hardware platforms, allowing efficient implementation of 2-qubit gates without additional swap operations, making it particularly suitable for hardware characterization. Moreover, it is theoretically estimated to approximate the ideal Haar-random state with a depth scaling at most polynomial in the gate number [20], [40], [41], [43].

Owing to its expected advantageous scaling of quantum property generation with a low cost of resources, we identify the BC construction as a promising candidate for test circuit generation.

IV. EVALUATION

To assess the validity of our set of diversity scores for characterizing test input states, and the effectiveness of our BC construction for generating the latter, we formulate the following research questions (RQs):

- **RQ1:** To what extent do the proposed diversity scores capture the relevant state properties needed to quantify expressibility and entanglement strength of test input states?

- **RQ2:** How do various circuit generators fare in terms of circuit depth to approximate Haar-random unitaries?
- **RQ3:** How do different circuit generators fare in terms of the diversity of the generated test input states?

In **RQ1** we aim to experimentally verify the capability of the theoretically established diversity scores to capture the structures in which the magnitude, phase and entanglement are present in diverse, randomized and characteristic quantum states.

RQ2 aims to gauge the capability of our BC generator, in comparison with existing test circuit generators, to approximate Haar-random unitaries, since proximity to a Haar-random unitary is a commonly used criterion for evaluating randomness, uniformity and expressibility.

Finally, **RQ3** assesses the diversity of the test input states in terms of our established scores, where the proposed set of diversity scores enables a fine-grained, per-score analysis compared to the proximity analysis to Haar-random unitaries conducted for RQ2. The goal is to get a detailed assessment of the quantitative presence of quantum properties and hence the diversity in the test input states produced by different circuit generators.

A. RQ1: Diversity Scores for evaluating state properties

Methodology: To answer RQ1, we analyzed the discrimination capabilities of the proposed scores with respect to the presence or lack of specific properties and structures in quantum states. To this end, we compared the distribution of score values between randomized reference states and characteristic entangled states. Specifically, we evaluated the scores on four characteristic types of randomized n -qubit states. Therefore, we generated 5000 instances of 6-qubit states for each type, implementing Haar-random (HR), maximally-entangled bipartite (ME), uniform flat-amplitude (UA) and product (P) states, as defined in Eqs. 6-9. For the ME state, we chose an equal partition that divides the system evenly into two subsystems A and B . Additionally, we chose two well-defined entangled states, namely the GHZ and W states, as defined in Eqs. 1-2 to ensure a complete exploration of the score's output range.

For the entanglement scores, we consider the Meyer-Wallach measure Q (cf. Eq. 5), the von Neumann entropy S (cf. Eq. 3), and the Rényi-2 entropy S_2 as defined in Eq. 4, where the latter is evaluated on two different subsystems. We chose a balanced bipartition of the global state on which we evaluate S_2 by clustering the considered qubits into blocks. By $S_2(A)$ we denote the entropy that is evaluated on a block fully contained within subsystem A of the fixed balanced bipartition $A|B$, whereas $S_2(AB)$ is evaluated on a block that includes the partition between the subsystems A and B . To ensure a fair comparison, S_2 is normalized by the size of the block under consideration, and thus returns the Rényi-2 entropy per qubit in the block.

Results: The resulting distributions of the evaluated scores are shown in Fig. 2, grouped by the three addressed properties,

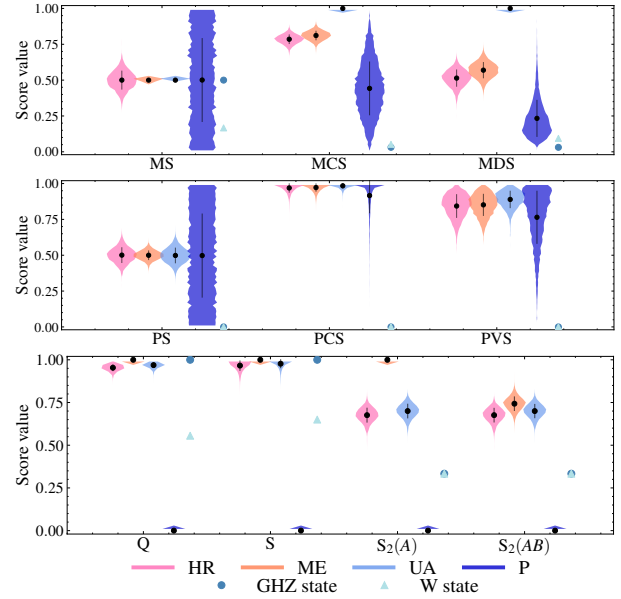


Fig. 2: Scores addressing the magnitude (top), phase (middle) and entanglement (bottom) evaluated on the GHZ and W state, as well as on 5000 instances of Haar-random (HR), maximally-entangled (ME), uniform-amplitude (UA) and product (P) states for 6 qubits.

namely magnitude (top), phase (middle) and entanglement (bottom), evaluated for qubit $k = 0$.

As shown in Fig. 2 (top), for **MS**, we observe that the HR, ME, P and GHZ states are centered around 0.5, indicating an equilibrated balance between qubit sectors, as expected from states that are symmetric in each qubit's $|0\rangle$ and $|1\rangle$ state. By construction, the W state has a larger amplitude mass in the $|0\rangle$ states, which is confirmed by its lower MS value (MS = 0.17). The widest distribution is achieved by the P state, due to its random instantiation of each individual qubit. In this case, each qubit is sampled independently and uniformly, according to the Haar measure on single-qubit states, resulting in a uniformly distributed probability as captured by the sector-wise score MS. Considering **MCS** and **MDS**, the UA states achieve the maximal score, since by design they include contributions from all possible basis states. In contrast, the W and GHZ states with very few basis state contributions result in a minimal value of the score. In terms of amplitude distribution, the ME states show a high similarity to the HR states. The retrieved average expected value of the MCS for the HR states results in 0.78 and its MDS value agrees with the expected 0.5. The P states show a moderate MCS and MDS, as expected by the factorization of their probability mass.

Fig. 2 (middle) shows that the distributions obtained for the **PS** closely resemble those observed for the **MS**, with all randomized states lying around the mean value indicating balanced phases, and the P states spanning the entire range from 0 to 1. For the **PCS** and **PVS** all of the randomized states generate strongly decorrelated phases and a large phase

spread. In contrast, the W and GHZ states correctly result in low phase score values since their phases are completely aligned by design. The pronounced tail for the P states reflects their more limited ability to generate strong phase-scrambling.

Fig. 2 (bottom) confirms that all entanglement scores correctly assign zero entanglement to the P states, as required for any valid entanglement measures. For both the **Q** and **S** scores, HR, ME, UA and the GHZ states exhibit distributions that are strongly peaked near their maximal values, confirming that on average each qubit is strongly entangled with the remainder of the system. The W state, by contrast, shows only moderate entanglement according to Q and S (~ 0.6), consistent with its delocalization of a single excitation over many qubits and the resulting weak single-qubit entanglement.

A qualitatively different behavior emerges when considering the Rényi-2 entropy S_2 , which probes the scaling of entanglement with subsystem size. For the ME states, we observe a pronounced reduction in entanglement when evaluating blocks that cross the state construction bipartition $A|B$, as captured by $S_2(AB)$. Indeed, while $S_2(A)$ reaches the maximal value of 1.0, $S_2(AB)$ returns a mean value of 0.74. This behavior is expected, since by design volume-law entanglement is realized only within block A . For both GHZ and W states, S_2 yields a value of $\frac{1}{3}$ independent of the chosen block, consistent with the presence of one global bit of entanglement shared across the 3-qubit subsystems. In contrast, HR and UA states achieve similarly high values for Q and S, but moderately reduced values for $S_2(A)$ and $S_2(AB)$, consistent with the range of entanglement which is shorter than that of the ME states. The independence of the entanglement strength on the block choice for HR and UA states confirms an isotropic entanglement distribution, while the S_2 values, significantly exceeding those of GHZ and W state, indicate a favorable extensive entanglement scaling in the randomized ensembles.

These results confirm that the diversity scores we introduced, addressing magnitude, phase, and entanglement, are capable of differentiating between the structural characteristics of the states considered. Indeed, the selected entanglement measures recover the expected entanglement signatures of well-understood reference states and reliably distinguish structured entanglement from that arising in randomized ensembles.

Answer to RQ1. The set of the proposed diversity scores capture well the different properties present in the quantum reference states we considered, resulting in good agreement with the underlying theory.

B. RQ2: Randomness Analysis

Methodology: **RQ2** addresses the difficulty of generating a true Haar-random unitary using a finite set of discrete gates without encountering exponentially increasing circuit depth with the system size. To this end, we implement the BC construction as shown in Fig. 1, choosing the controlled Z gate as the entangling gate and compare the amount of gates required to generate circuit matrices that are Haar-random

TABLE I: Required gate numbers to reach the randomness criterion for different test circuit generators and system sizes.

Number of Qubits	BC	RC	UCNOT	iQFT
3	10	15	16	18
5	17	30	29	36
7	24	50	42	58

beyond statistical distinguishability with three existing test circuit generators.

We employ the statistical test of maximum mean discrepancy (MMD) using the normalized Hilbert-Schmidt inner-product kernel on generated circuit matrices to evaluate their proximity to Haar-random matrices. The MMD hypothesis test is used to compare distances between the embeddings of two distributions given by their unitary matrices [44]. We attribute a p-value to conclude whether the difference between the realization of the distributions is statistically insignificant and adopt a threshold value of $p > 0.01$.

In practice, ideal Haar unitary matrices are obtained via the standard QR decomposition method [45]. We generate a set of these ideal random matrices, as well as circuit matrices that are produced by our BC generator. In addition, our experiments include three existing test circuit generators, i.e., Random Circuit (RC), UCNOT and iQFT, proposed by QuraTest [14], a state-of-the-art approach in quantum software testing.

We employed the PennyLane framework [46] to generate quantum circuits and extract their circuit matrix representation. To vary the size of the considered system we generate circuit matrices for 3, 5 and 7 qubits.

Results: Fig. 3 shows how circuit depth influences the randomness of circuit matrices generated by the different circuit generators for three different system sizes, namely 3, 5 and 7 qubits. In the case for 5 qubits, circuit unitaries generated using the **BC** construction already satisfy our randomness criterion at a single-layer depth, corresponding to circuits with 17 gates. In contrast, the **RC** generator requires approximately 30 gates to meet our randomness threshold, and 40 to reach a comparable approximation to a Haar-random unitary than the BC generator. The onset of a plateau observed for the p-values beyond a certain gate number is consistent with the expected saturation of expressibility [15]. The **UCNOT** generator reaches the randomness threshold with one single construction layer, requiring 29 gates. Since increasing the number of layers does not modify its p-value we only show the result for its shallowest configuration. Finally, the **iQFT** generator has a fixed circuit depth determined by its algorithmic structure, which corresponds to 36 gates for a 5-qubit system. The resulting unitaries meet the randomness criterion.

In general, larger qubit systems require a higher gate count to satisfy the randomness criterion. We observe that the scaling advantage of the BC construction scheme compared to the other circuit construction methods is reinforced with larger system sizes, since the required gate count grows more slowly with the system's qubit number, as shown in Table I.

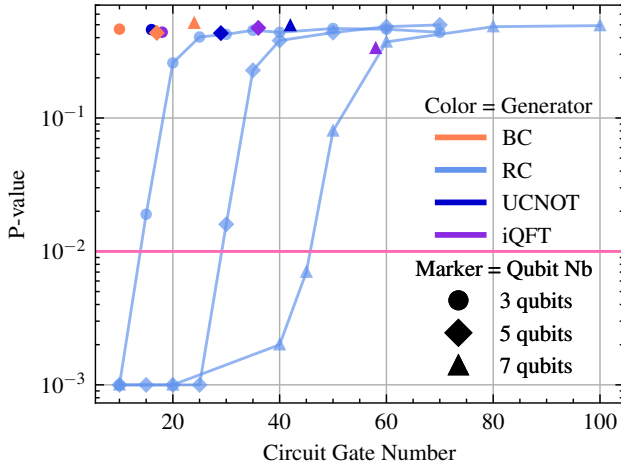


Fig. 3: P-value comparison of RC, BC, UCNOT and iQFT circuit generators for 3, 5 and 7 qubits and varying circuit depth.

Answer to RQ2. The BC generator fulfills the randomness criterion with the least amount of gates, followed by the UCNOT, iQFT generators, and finally the RC generator.

C. RQ3: Diversity Scores for expressibility and entangling capabilities

Methodology: To address **RQ3**, we evaluate test input states generated by the BC construction and three existing test-case generators using our established scores. Specifically, we apply the generated circuit matrices onto the default initial state in order to generate test input states, and assess the diversity of the resulting states in terms of expressibility and entangling capability.

Therefore, we generated 2000 instances of circuit matrices for each of the four generators (BC, RC, UCNOT, iQFT) for a 5-qubit system. We explored the capability of the BC and RC generators to allow for a variable gate number to study the diversity of the generated test input states with increasing circuit depth. Additionally, we computed 2000 instances of Haar-random states, representing a uniform state distribution on the whole multi-qubit state space and thus acting as a baseline for expressibility. We evaluated each of these with the proposed scores targeting magnitude, phase, and entanglement distributions. We used the Wasserstein metric to quantify the statistical distance between the distributions obtained from evaluating the diversity scores on Haar-random states and on test input states constructed by the different generators.

Results: First, we analyze the varying distribution of magnitude and phase in the generated test circuits with increasing circuit depth. In Fig. 4, we show the magnitude and phase scores MS and PS (top), the magnitude and phase correlation scores MCS and PCS (middle), and entanglement scores S and $S_2(A_3)$ (bottom) for the RC and BC construction with increasing gate numbers.

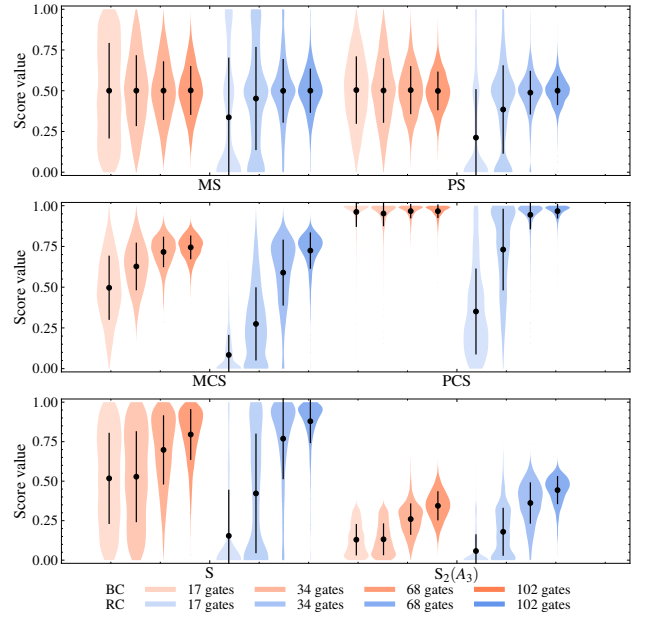


Fig. 4: Diversity Scores for increasing gate numbers. Magnitude (MS) and phase (PS) scores (top), magnitude (MCS) and phase (PCS) correlation scores (middle) and entanglement entropy (S) and Rényi-2 entropy ($S_2(A_3)$) (bottom).

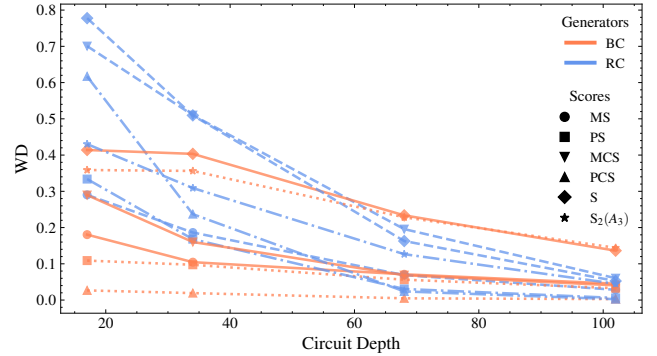


Fig. 5: Wasserstein Distance (WD) of the diversity scores evaluated on the test input states generated by BC and RC constructions with varying circuit depth with respect to Haar-random reference scores.

We observe a clear variation in the shape of the distribution with increasing circuit depth for the considered generators and scores. In the case of the BC construction, the mean values of the scores **MS** and **PS** remain constant around the balanced value 0.5, however, their distributions become more narrowly concentrated, in agreement with the expectations for Haar-random states. For the RC generator, the change accompanied by increasing gate number is more drastic. In this case also the mean values, so the average generated magnitude (**MS** from 0.34 to 0.5) and phase (**PS** from 0.21 to 0.5) weights increase significantly, indicating broadened magnitude and phase diversity. A very similar picture is obtained for the correlation scores, where for **MCS** both generators produce

increasingly correlated values with the gate number, whereas for **PCS**, the BC produces maximally phase decorrelated state already with shallow circuits, whereas RC requires a significant number of gates (~ 100) to fulfill this condition. For the entanglement scores, **S** and $S_2(A_3)$, both generators approach the maximal entanglement score, with increasingly narrow distributions as circuit depth grows. We even observe that for deep circuits, the states produced by the unstructured RC generator outperform the values obtained for the structured BC construction. Furthermore, both generators achieve a higher mean value for **S**, namely 0.8 and 0.88 for BC and RC respectively, than for $S_2(A_3)$, which is limited to 0.34 and 0.44 for BC and RC. This illustrates again (cf. previous section) that it is significantly more difficult to generate genuinely multipartite, delocalized entanglement compared to single-qubit entanglement, especially for structured generators.

Compared to the results of RQ2, we note that our diversity scores provide a more detailed tool to assess the expressibility of the generated test circuits, rather than solely considering their distinguishability in terms of randomness. Indeed, we observe that the generated circuits are becoming more diverse (achieving higher scores, which lie closer to the Haar-random value) even for gate counts above 17 for the BC construction and 30 for RC that exceeded the randomness threshold set in the previous section.

By increasing the number of allowed gates, especially for the RC construction, the distributions of the considered scores progressively approach that obtained from the Haar-random states, as illustrated by the statistical distance plotted in Fig. 5. Smaller distances indicate a closer match to the Haar-random ensemble and therefore higher expressibility of the corresponding circuit generator. Furthermore, we observe that a gate depth of 17 and below is insufficient for the RC construction to reproduce the state properties that we expect for random circuits.

The complete overview of the score distributions obtained for states originating from different generators is shown in Fig. 6. To allow for a fair comparison between generators, the RC and BC constructions are evaluated on circuit depths that are similar to the fixed gate generators UCNOT and iQFT (2 layers for the BC, corresponding to 34 gates for the RC).

For **MS**, all test states are symmetric around $MS \sim 0.5$, and thus show balanced two-sector states. For the **MCS** and **MDS**, the BC achieves the highest scores (closest to Haar-random predictions), followed by the UCNOT, iQFT and RC construction. As for the **PS**, all generators except the RC are balanced around 0.5 and are capable of generating states with completely decorrelated phases **PCS** ~ 1 , whereas all generators achieve a uniform phase spread captured by **PVS** close to its maximal value. For the entropy **S** the BC construction achieves the highest mean score, closest to the Haar-random distribution; all other generators result in very similar mean values, whereas, with the more narrow distribution, the iQFT generator is slightly advantageous in its reliability to generate entanglement compared to RC and UCNOT. In the case of the Meyer-Wallach measure **Q**, for which the mean

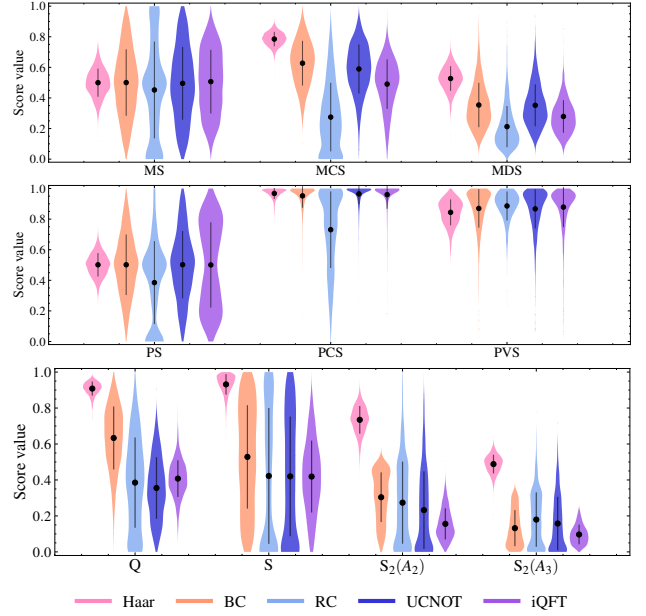


Fig. 6: Diversity scores for magnitude (top), phase (middle) and entanglement (bottom) evaluated on test input states produced by BC (2 layers), RC (34 gates), UCNOT and iQFT generators, compared to Haar-random states for 5 qubits.

value $\bar{Q}$ can be directly related to the entangling capability [15], the score results in the average value $\bar{Q} = 0.63$ for the BC construction, 0.41 for iQFT, 0.39 for RC and 0.36 for UCNOT, in comparison to the Haar-random state which achieves $\bar{Q} = 0.9$. In fact, the computed WDs of the **Q** score with respect to Haar-random states result in 0.27 for the two layer BC construction, 0.52 for the RC generator, 0.55 for UCNOT and 0.50 for iQFT. For the Rényi-2 entropy S_2 , for the 2-qubit block, the BC scores the highest, followed by RC, UCNOT and iQFT generators. For the larger 3-qubit block, the RC slightly outperforms the other circuit generators, followed by the very similar distribution generated by UCNOT, BC and iQFT. The larger values obtained for **S** in comparison to S_2 , and especially their reduction with increasing block size, indicate that although single-qubit entanglement is strong, the entanglement is limited to a short range and localized. We note that the qualitative features of the distributions captured by **S** are also present in the computationally advantageous **Q** measure, which is thus preferable for practical applications.

Answer to RQ3. The BC construction achieves the highest diversity scores approaching the baseline value we obtain for Haar-random states in all three aspects, namely magnitude, phase and entanglement, followed by the UCNOT, iQFT and RC generators.

D. Threats to Validity and Limitations

To reduce the threats of random variation and insufficient statistical power, we employ at least 2000 samples of each

generator, and used a common statistical test, such as the maximum mean discrepancy, and the Wasserstein distance between the score distribution of generated test input states and the Haar-random reference to analyze the results. However, due to the inherent variability of the randomness involved, some results concerning the state scores may still be affected by statistical outliers and finite-size effects.

Our diversity scores address quantities that are not direct physical observables, and as such they are not uniquely defined. To mitigate this threat, we selected three different measures to capture different aspects of each property and apply them consistently across all evaluations.

The proposed diversity scores might face scalability limitations beyond classically tractable system sizes, as their evaluation involves increasing computational costs with growing system dimensions. To estimate the scalability of the required depth of the BC to meet the randomness criteria, the system’s underlying connectivity has to be considered, following the sub-polynomial constructions boundaries established for local quantum circuits that have been proven in all circuit architectures [41].

V. RELATED WORK

In the context of quantum software testing, test case generation requires quantum input states to be produced by a quantum circuit. Prior work, most notably QuraTest [14] has investigated the use of quantum-inspired measures to assess the diversity of test inputs generated by different circuit constructions. In this setting, diversity is quantified by three scores probing the magnitude, phase and entanglement across a bipartition, effectively capturing average cross-sector correlations associated with individual qubit being in state $|0\rangle$ or $|1\rangle$. In contrast, our diversity scores provide a more comprehensive characterization by explicitly targeting state-space exploration, rather than limiting the analysis to a single-cut correlation measure. Furthermore, these scores are used to evaluate the performance in input and output state coverage of three test circuit generators, consisting of a Random gate combination, a UCNOT pattern and an IQFT circuit. In contrast to our BC generator, none of these existing test circuit generators is specifically optimized for circuit depth efficiency or for producing strongly entangled, highly expressive states. To the best of our knowledge, despite its highly desirable properties for generating diverse and comprehensive test inputs, this work constitutes the first systematic investigation of the BC construction as a test circuit generator for QST.

Research on quantum random circuits spans a wide range of applications, including complex quantum many-body and back-hole physics [47], [48], device benchmarking [49]–[51], variational quantum algorithms [15], [16], [52], quantum machine learning [53], [54], and near-term demonstrations of quantum advantage [34], [55], [56]. Since implementing ideal Haar-random unitaries using discrete gates is computationally expensive, practical approaches focus on efficiently approximating Haar randomness up to an error ϵ , formalized through the framework of unitary t -designs [19], [21]. Accordingly,

a central challenge resides in the practical construction of pseudo-random circuits that approximate Haar properties with minimal resources. To this end, numerous efficient schemes for approximate unitary design have been investigated for their convergence to unitary t -designs at shallow circuit depth [18]–[21], [34], [40]–[43], [52]. In contrast, the BC generator introduced in this work instantiates the established parallel local random-circuit paradigm with a specific aim at efficient QST, combining low gate counts with nearest-neighbor entangling gates, making it well-suited for implementation on real-world hardware [55]–[57].

Random-circuit designs are especially relevant for randomized device benchmarking, where it is used to estimate process fidelities and characterize noise in quantum processors [49]–[51] by generating constructible families of states that approximate uniform sampling over the full Hilbert space. Related ideas also appear in variational quantum algorithms, where expressibility serves as a key metric characterizing how well a given ansatz can cover the Hilbert space [15]–[17]. Building on these prior studies, our work leverages the established link between expressibility and entangling capabilities of parametrized quantum circuits, as well as their proximity to Haar-distributed ideal random states to assess the diversity of test input states using the proposed diversity scores. This plays a central role in QST, since the Hilbert space coincides with the set of all admissible input states and its extensive exploration forms the basis of any comprehensive testing scheme.

VI. CONCLUSIONS

In this work, we have proposed a set of diversity scores to guide test circuit generation for comprehensive quantum software testing, and paired them with our BC construction method for efficient random test circuit generation. We have compared the expressibility and entangling capabilities of differently structured test circuit generators and observed that, from the considered implementations, our BC construction reaches the highest scores, with the most shallow gate depth.

As part of future work, we plan to refine the analysis of the circuit generator’s required gate depth by distinguishing between single and two-qubit gates. We also plan to study the impact of an expressible and entangling test circuit generator on the efficacy of the fault finding process, for instance, leveraging fault injection. As part of future applications, we see potential in improving the effective generation and deployment of random circuits, the test-case guidance for comprehensive and resource-efficient QST, the efficiency of holistic benchmarking strategies, and the fine-grained assessment of overall quantum processor performance.

DATA AVAILABILITY STATEMENT

A replication package containing the code and the experimental data is provided online [58].

REFERENCES

- [1] M. Pezzè, S. Abrahão, B. Penzenstadler, D. Poshvanyk, A. Roychoudhury, and T. Yue, “A 2030 roadmap for software engineering,” *ACM Transactions on Software Engineering and Methodology*, vol. 34, no. 5, pp. 1–55, 2025.
- [2] S. Honarvar, M. R. Mousavi, and R. Nagarajan, “Property-based testing of quantum programs in q#,” in *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops*, 2020, pp. 430–435.
- [3] R. Abreu, J. P. Fernandes, L. Llana, and G. Tavares, “Metamorphic testing of oracle quantum programs,” in *Proceedings of the 3rd International Workshop on Quantum Software Engineering*, 2022, pp. 16–23.
- [4] G. Pontolillo, M. R. Mousavi, and M. Grzesiuk, “Qcheck: A property-based testing framework for quantum programs in qiskit,” *arXiv preprint arXiv:2503.22641*, 2025.
- [5] A. Miranskyy, J. Campos, A. Mjeda, L. Zhang, and I. G. R. de Guzmán, “On the feasibility of quantum unit testing,” *arXiv preprint arXiv:2507.17235*, 2025.
- [6] Y. Huang and M. Martonosi, “Statistical assertions for validating patterns and finding bugs in quantum programs,” in *Proceedings of the 46th International Symposium on Computer Architecture*, 2019, pp. 541–553.
- [7] G. Li, L. Zhou, N. Yu, Y. Ding, M. Ying, and Y. Xie, “Projection-based runtime assertions for testing and debugging quantum programs,” *Proceedings of the ACM on Programming Languages*, vol. 4, no. OOPSLA, pp. 1–29, 2020.
- [8] N. H. Oldfield, C. Laaber, T. Yue, and S. Ali, “Faster and better quantum software testing through specification reduction and projective measurements,” *ACM Transactions on Software Engineering and Methodology*, vol. 34, no. 7, pp. 1–39, 2025.
- [9] S. Ali, P. Arcaini, X. Wang, and T. Yue, “Assessing the effectiveness of input and output coverage criteria for testing quantum programs,” in *2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST)*. IEEE, 2021, pp. 13–23.
- [10] X. Wang, P. Arcaini, T. Yue, and S. Ali, “Quito: a coverage-guided test generator for quantum programs,” in *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 2021, pp. 1237–1241.
- [11] —, “Generating failing test suites for quantum programs with search,” in *International symposium on search based software engineering*. Springer, 2021, pp. 9–25.
- [12] J. Wang, F. Ma, and Y. Jiang, “Poster: Fuzz testing of quantum program,” in *2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST)*. IEEE, 2021, pp. 466–469.
- [13] P. Long and J. Zhao, “Testing multi-subroutine quantum programs: From unit testing to integration testing,” *ACM Transactions on Software Engineering and Methodology*, vol. 33, no. 6, pp. 1–61, 2024.
- [14] J. Ye, S. Xia, F. Zhang, P. Arcaini, L. Ma, J. Zhao, and F. Ishikawa, “Quratest: Integrating quantum specific features in quantum program testing,” in *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 2023, pp. 1149–1161.
- [15] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, “Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms,” *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900070, 2019.
- [16] K. Nakaji and N. Yamamoto, “Expressibility of the alternating layered ansatz for quantum computation,” *Quantum*, vol. 5, p. 434, 2021.
- [17] P. C. Azado, G. I. Correr, A. Drinko, I. Medina, A. Canabarro, and D. O. Soares-Pinto, “Expressibility, entangling power, and quantum average causal effect for causally indefinite circuits,” *Physical Review A*, vol. 111, no. 4, p. 042620, 2025.
- [18] J. Emerson, Y. S. Weinstein, M. Saraceno, S. Lloyd, and D. G. Cory, “Pseudo-random unitary operators for quantum information processing,” *science*, vol. 302, no. 5653, pp. 2098–2100, 2003.
- [19] A. W. Harrow and R. A. Low, “Random quantum circuits are approximate 2-designs,” *Communications in Mathematical Physics*, vol. 291, no. 1, pp. 257–302, 2009.
- [20] C. Dankert, R. Cleve, J. Emerson, and E. Livine, “Exact and approximate unitary 2-designs and their application to fidelity estimation,” *Physical Review A—Atomic, Molecular, and Optical Physics*, vol. 80, no. 1, p. 012304, 2009.
- [21] F. G. Brandao, A. W. Harrow, and M. Horodecki, “Local random quantum circuits are approximate polynomial-designs,” *Communications in Mathematical Physics*, vol. 346, no. 2, pp. 397–434, 2016.
- [22] H. J. Briegel and R. Raussendorf, “Persistent entanglement in arrays of interacting particles,” *Physical Review Letters*, vol. 86, no. 5, p. 910, 2001.
- [23] W. Dür, G. Vidal, and J. I. Cirac, “Three qubits can be entangled in two inequivalent ways,” *Physical Review A*, vol. 62, no. 6, p. 062314, 2000.
- [24] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, “Mixed-state entanglement and quantum error correction,” *Physical Review A*, vol. 54, no. 5, p. 3824, 1996.
- [25] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge university press, 2010.
- [26] P. Calabrese and J. Cardy, “Entanglement entropy and quantum field theory,” *Journal of statistical mechanics: theory and experiment*, vol. 2004, no. 06, p. P06002, 2004.
- [27] M. B. Hastings, I. González, A. B. Kallin, and R. G. Melko, “Measuring renyi entanglement entropy in quantum monte carlo simulations,” *Physical review letters*, vol. 104, no. 15, p. 157201, 2010.
- [28] R. Islam, R. Ma, P. M. Preiss, M. Eric Tai, A. Lukin, M. Rispoli, and M. Greiner, “Measuring entanglement entropy in a quantum many-body system,” *Nature*, vol. 528, no. 7580, pp. 77–83, 2015.
- [29] J. Zhao, B.-B. Chen, Y.-C. Wang, Z. Yan, M. Cheng, and Z. Y. Meng, “Measuring renyi entanglement entropy with high efficiency and precision in quantum monte carlo simulations,” *npj Quantum Materials*, vol. 7, no. 1, p. 69, 2022.
- [30] D. A. Meyer and N. R. Wallach, “Global entanglement in multiparticle systems,” *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4273–4278, 2002.
- [31] G. K. Brennen, “An observable measure of entanglement for pure states of multi-qubit systems,” *arXiv preprint quant-ph/0305094*, 2003.
- [32] T. R. de Oliveira, G. Rigolin, M. C. de Oliveira, and E. Miranda, “Multipartite entanglement signature of quantum phase transitions,” *Physical review letters*, vol. 97, no. 17, p. 170401, 2006.
- [33] K. Zyczkowski and M. Kus, “Random unitary matrices,” *Journal of Physics A: Mathematical and General*, vol. 27, no. 12, p. 4235, Jun. 1994. [Online]. Available: <https://doi.org/10.1088/0305-4470/27/12/028>
- [34] M. Liu, J. Liu, Y. Alexeev, and L. Jiang, “Estimating the randomness of quantum circuit ensembles up to 50 qubits,” *npj Quantum Information*, vol. 8, no. 1, p. 137, 2022.
- [35] A. A. Mele, “Introduction to haar measure tools in quantum information: A beginner’s tutorial,” *Quantum*, vol. 8, p. 1340, 2024.
- [36] N. I. Fisher, *Statistical analysis of circular data*. cambridge university press, 1995.
- [37] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009.
- [38] Y. Liu, P. Sierant, P. Stornati, M. Lewenstein, and M. Płodzień, “Quantum algorithms for inverse participation ratio estimation in multiqubit and multiqubit systems,” *Physical Review A*, vol. 111, no. 5, p. 052614, 2025.
- [39] P. W. Claeys and G. De Tomasi, “Fock-space delocalization and the emergence of the porter-thomas distribution from dual-unitary dynamics,” *Physical Review Letters*, vol. 134, no. 5, p. 050405, 2025.
- [40] J. Haferkamp, “Random quantum circuits are approximate unitary t -designs in depth $O(nt^{5+o(1)})$,” *Quantum*, vol. 6, p. 795, 2022.
- [41] T. Schuster, J. Haferkamp, and H.-Y. Huang, “Random unitaries in extremely low depth,” *Science*, vol. 389, no. 6755, pp. 92–96, 2025.
- [42] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, “Cost function dependent barren plateaus in shallow parametrized quantum circuits,” *Nature communications*, vol. 12, no. 1, p. 1791, 2021.
- [43] A. W. Harrow and S. Mehraban, “Approximate unitary t -designs by short random quantum circuits using nearest-neighbor and long-range gates,” *Communications in Mathematical Physics*, vol. 401, no. 2, pp. 1531–1626, 2023.
- [44] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, “Integrating structured biological data by kernel maximum mean discrepancy,” *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, 2006.
- [45] F. Mezzadri, “How to generate random matrices from the classical compact groups,” *Notices of the American Mathematical Society*, vol. 54, no. 5, pp. 592–604, 2007.
- [46] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, M. S. Alam, G. Alonso-Linaje, B. AkashNarayanan, A. Asadi *et al.*, “Pennylane: Automatic differentiation of hybrid quantum-classical computations,” *arXiv preprint arXiv:1811.04968*, 2018.
- [47] A. Nahum, S. Vijay, and J. Haah, “Operator spreading in random unitary circuits,” *Physical Review X*, vol. 8, no. 2, p. 021014, 2018.

- [48] P. Hayden and J. Preskill, "Black holes as mirrors: quantum information in random subsystems," *Journal of high energy physics*, vol. 2007, no. 09, pp. 120–120, 2007.
- [49] J. Emerson, R. Alicki, and K. Życzkowski, "Scalable noise estimation with random unitary operators," *Journal of Optics B: Quantum and Semiclassical Optics*, vol. 7, no. 10, pp. S347–S352, 2005.
- [50] 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.
- [51] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, "Scalable randomised benchmarking of non-clifford gates," *npj Quantum Information*, vol. 2, no. 1, p. 16012, 2016.
- [52] Y. Nakata, D. Zhao, T. Okuda, E. Bannai, Y. Suzuki, S. Tamiya, K. Heya, Z. Yan, K. Zuo, S. Tamate *et al.*, "Quantum circuits for exact unitary t-designs and applications to higher-order randomized benchmarking," *PRX Quantum*, vol. 2, no. 3, p. 030339, 2021.
- [53] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature communications*, vol. 9, no. 1, p. 4812, 2018.
- [54] J. Liu, K. Najafi, K. Sharma, F. Tacchino, L. Jiang, and A. Mezzacapo, "Analytic theory for the dynamics of wide quantum neural networks," *Physical Review Letters*, vol. 130, no. 15, p. 150601, 2023.
- [55] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, "Characterizing quantum supremacy in near-term devices," *Nature Physics*, vol. 14, no. 6, pp. 595–600, 2018.
- [56] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, "Quantum supremacy using a programmable superconducting processor," *nature*, vol. 574, no. 7779, pp. 505–510, 2019.
- [57] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, "Quantum entanglement growth under random unitary dynamics," *Physical Review X*, vol. 7, no. 3, p. 031016, 2017.
- [58] M. Ernzer, S. Y. Shin, F. Pastore, and D. Bianculli, "[replication package] Randomized and Diverse Input State Generation for Quantum Program Testing," 2026. [Online]. Available: <https://doi.org/10.6084/m9.figshare.32105962>