EPJ Quantum Technology
a SpringerOpen Journal

RESEARCH   **Open Access**

On the learnability of quantum state fidelity

Norhan Elsayed Amer[1*], Walid Gomaa[1,2], Keiji Kimura[3], Kazunori Ueda[3] and Ahmed El-Mahdy[1,2]

*Correspondence:
norhan.elsayed@ejust.edu.eg
[1]Department of Computer Science
and Engineering, Egypt-Japan
University for Science and
Technology, Alexandria, Egypt
Full list of author information is
available at the end of the article

Abstract

Current quantum processing technology is generally noisy with a limited number of qubits, stressing the importance of quantum state fidelity estimation. The complexity of this problem is mainly due to not only accounting for single gates and readout errors but also for interactions among which. Existing methods generally rely on either reconstructing the given circuit state, ideal state, and computing the distance of which; or forcing the system to be on a specific state. Both rely on conducting circuit measurements, in which computational efficiency is traded off with obtained fidelity details, requiring an exponential number of experiments for full information. This paper poses the question: Is the mapping between a given quantum circuit and its state fidelity learnable? If learnable, this would be a step towards an alternative approach that relies on machine learning, providing much more efficient computation. To answer this question, we propose three deep learning models for 1-, 3-, and 5-qubit circuits and experiment on the following real-quantum processors: ibmq_armonk (1-qubit), ibmq_lima (5-qubit) and ibmq_quito (5-qubit) backends, respectively. Our models achieved a mean correlation factor of 0.74, 0.67 and 0.66 for 1-, 3-, and 5-qubit random circuits, respectively, with the exponential state tomography method. Additionally, our 5-qubit model outperforms simple baseline state fidelity estimation method on three quantum benchmarks. Our method, trained on random circuits only, achieved a mean correlation factor of 0.968 while the baseline method achieved 0.738. Furthermore, we investigate the effect of dynamic noise on state fidelity estimation. The correlation factor substantially improved to 0.82 and 0.74 for the 3- and 5-qubit models, respectively. The results show that machine learning is promising for predicting state fidelity from circuit representation and this work may be considered a step towards efficient end-to-end learning.
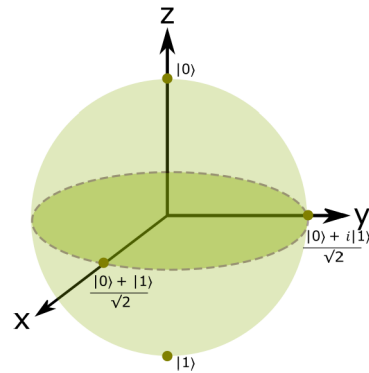
Keywords: Quantum computing; State tomography; Convolutional neural network; Classical optimization; Noise characterization; Quantum circuit; Fidelity estimation

1 Introduction

With the advancement of technology in the quantum computing field, we are entering the *Noisy Intermediate-Scale Quantum (NISQ)* era [1]. NISQ refers to the limited number of available qubits ranging from 50 to a few hundred that can potentially perform more complex tasks. In addition, the term "Noisy" refers to the noise inherent in qubits that needs error correction schemes. However, these schemes are constrained by the limited number of available qubits, hindering continuous quantum error correction. Noise in the NISQ era occurs due to multiple reasons such as crosstalk between gates executed at the same time,

Springer

Figure 1 Bloch Sphere

decoherence errors due to undesirable interactions with the environment, gates errors, and readout errors. Moreover, mapping of qubits according to layout constraints and the complexity of the given quantum circuit contribute further to noise.

Characterising the noise is therefore essential in guiding circuit optimisations, and quantifying circuit reliability. The term 'state fidelity' is well-known in characterising such noise [2]. It is generally defined as the probability of the noisy quantum circuit state to be equal to the ideal one. Unfortunately, there is no accurate analytical noise model for estimating state fidelity. Thus, state of the art relies on exhaustive experimentation to reconstruct the noisy state. This process is called 'state tomography' [3]. To illustrate, consider a one qubit state. The state can be represented as a point onto the Bloch sphere (Fig. 1). To reconstruct the state, it suffices to measure the x, y, z coordinates by generating three circuits, each projecting the quantum state into a corresponding direction. For a two-qubit circuit, 3×3 circuits are generated each projecting to a combination of the axes, e.g. xx, xy, $\cdots$. This results into 3^n different circuits for an n-qubit circuit, severely limiting scalability of the state tomography approach.

While there are many methods in the literature to compute state fidelity (as detailed in Sect. 1.1), the state of the art mainly relies on state tomography. Other fast methods exist such as random benchmarking [4, 5] and simple multiplicative models [6, 7]; however, they either provide partial noise characterisation or suffer from low accuracy.

As the state of the art requires generating an extensive number of circuits (as well as their executions), it might be interesting to consider an alternate approach that merely relies on one circuit with no measurements. In particular, we consider the question: Is the mapping between a given quantum circuit and its state fidelity learnable? This is a necessary step towards a time efficient, end-to-end fidelity estimation process.

To answer the above question, we need to provide evidence that machine learning results in accurate estimates of state fidelity for complex circuit models and quantum backends. This is challenging given the limited availability of quantum resources to the research community. Therefore, we use a simple multiplicative noise model as a baseline in evaluating the machine learning results. Moreover, we consider random circuits as well as quantum circuits for solving real-world problems to decrease bias and assess the generality of the approach.

Another issue is the time required for generating the training dataset. A training record consists of a representation of an input circuit together with its state fidelity. However, computing the state fidelity is an exponential process as mentioned before (for both the

circuit and the ideal one). Nevertheless, our question focuses on the learning aspect as a starting point towards the more general utility of the approach.

Due to the limited availability of open-access quantum backends, we consider three quantum backends provided by IBM (ibmq_armonk (1-qubit), ibmq_lima (5-qubit) and ibmq_quito (5-qubit)) for generating random circuits consisting of 1-, 3-, and 5-qubit. These small number of qubits can still perform complex tasks, representing 2^n states.

For each of the considered machines, we utilise the typical convolutional neural network architecture for estimating state fidelity. In addition, we utilize an embedding layer to provide a better dense representation of the circuit encoding.

Our results has reconfirmed the dynamic variation of noise across the day, which significantly affects the state fidelity of the same circuit when run at different times across the day [8]. Generally, we have observed two noise types: static and dynamic noise. Static noise refers to noise that occur to specific circuit's characteristic in addition to stable noise levels that can vary slightly in the quantum backend. Dynamic noise refers to system instability that are circuit independent while occurring in certain time-frame, resulting in system failure and low circuit fidelity.

The neural networks resulted in a mean correlation factor of 0.74, 0.67 and 0.66 for 1-, 3-, and 5-qubit random circuits, respectively, with the exponential state tomography method. When filtering out the dynamic noise, the mean correlation factors substantially improved to 0.82 and 0.74 for the 3-and 5-qubit models, respectively.

Additionally, for our most complex 5-qubit model, it outperformed a simple baseline state fidelity estimation method on three quantum benchmarks; our method achieved a mean correlation factor of 0.968 while the baseline method achieved 0.738, while trained on the random circuits only.

In summary, this paper has the following contributions:

1 Propose a 1D convolutional neural network model for state fidelity estimation for single qubit circuits, and extend the proposed model to three and five qubits circuits, exploiting transfer learning to test the generalization of our 1D convolutional models.

2 Usage of circuit representation instead of tomographic measurements which reduces the input space complexity from exponential to polynomial for practical quantum circuits.

3 Demonstrate the learnability potential of the neural network models when compared with baseline state fidelity estimation on both random and real-world quantum benchmarking circuits.

1.1 Related work

In this section, we review state of the art research in quantum state reconstruction using techniques with varying complexities. These techniques can be divided into employing optimization techniques, exploiting properties in certain entangled quantum states, and utilizing machine learning techniques. Other simple fidelity estimation techniques were proposed such as exploiting randomized benchmarking and utilizing gates and readout errors.

Several researchers transformed the problem into an optimization problem [9, 10] relying on maximum likelihood (MLE) estimation method for quantum state reconstruction. However, the MLE method is resource intensive. Gross et al. [11] and Flammia et al. [12]

utilized compressed sensing techniques on low-rank states. They randomly choose a subset of $O(rn^2 2^n)$ Pauli measurements, where r is the rank of the density matrix, and n represents the number of qubits. They then use convex optimization to estimate the density matrix. Bolduc et al. [13] proposed three methods for state reconstruction using different variations of projected gradient descent algorithm. Their method achieved a time complexity of $O(N2^{2n})$, where N is the number of projectors. Shang et al. [14] accelerated the convergence of the MLE method using projected gradients. Their method achieved a time complexity of $O(k_r^{n+1})$, where k_r represents the number of probability-operator measurement. Qi et al. [15] proposed a linear regression estimation method for state reconstruction in which the unknown parameters are estimated by the least squares method. Their results show that the method is faster than the MLE. However, their method is still exponential requiring $O(2^{4n})$. This method was enhanced by other works [16, 17]. Hou et al. [16] reduced the number of Pauli bases needed achieving a complexity of $O(8^n)$. Qi et al. [17] used an adaptive recursive method. Specifically, they recursively estimate the state based on the available measurements. This estimate is then updated with new measurements. Their method accelerated their previous estimation method [15] achieving $O(2^{3n})$. Ferrie [18] introduced self guided quantum tomography. Their method utilizes stochastic optimization techniques to estimate quantum state using 2 measurements per iteration on 2-qubit. The time complexity of their method is $O(2^n)$. However, the convergence of their technique is not scalable with the number of qubits. This technique was then extended to more qudits [19] and to mixed states [20]. Other researchers relied on Bayesian inference for state reconstruction [21–23].

Other studies have exploited properties in certain entangled types of quantum states [24, 25]. Tiurev and Sørensen [25] introduced a method for fidelity estimation on cluster state class, which refers to the entanglement of multi-qubit. Their method relies on the use of stabilizers that require number of measurements that increases linearly with the number of qubits. However, their work was limited to a special case which is the cluster state.

Several researchers have focused on exploiting machine learning methods. Qian and Shuqi [26] relied on dense neural network to reconstruct density matrices from tomographic measurements. The hyperparameters of their network and the number of training samples differ according to the number of qubits that range up to 5-qubit. Their simulated training dataset consists of ideal measurements, whereas the testing dataset contains Gaussian noise. Their results show that their method is comparable with the maximum likelihood technique. However, their work contains multiple limitations. One limitation is the use of simulated instead of real qauntum backend. Another limitation is that the input to their network will always be exponential as their network depends on tomographic measurements totalling 6^n. Other limitation is that the reconstruction requires exponential complexity achieving $O(2^{3n})$. Lohani et al. [27] proposed a 2D convolutional neural network to reconstruct an unknown state using tomographic measurements totalling 6^n where n refers to number of qubits. They collected 35,500 simulated random states for up to 4-qubit. They also explored the reconstruction of states for circuits with limited number of shots. Their results show that their proposed method is comparable with the maximum likelihood technique. However, one important issue with their technique is that their method depend on tomographic measurements as an input, which means that the inference time of a given circuit will be exponential since it will be transformed to

6^n measurements. Zhang et al. [28] reduced the input space by feeding a machine learning method a subset of the 3^n Pauli operators. Specifically, they proposed a dense neural network consisting of 4 layers that classifies the fidelity output as 1 of the 122 available fidelity intervals. They used separate models for qubits ranging from 2 to 7 on both general quantum states and classes of entangled states like bell, cluster, dicke, and Greenberger–Horne–Zeilinger state. For each model, they choose k Pauli operators that maximize expectation for a state. Their results show that fidelity classification accuracy is proportional to the number of Pauli operators. However, one limitation to their method is the need to calculate all 6^n measurements to select the maximum k operators. Cha et al. [29] introduced attention based tomography, in which they utilize the transformer model used in natural language processing applications. Their work is concerned with Greenberger–Horne–Zeilinger state, a class of entangled states. Their method consists of an embedding layer followed by 6 stacked transformer layers which are then fully connected to a dense layer. The input to their model is the positive operator valued measurements and the output is the reconstructed density matrix. Their results are comparable to maximum likelihood estimation methods. However, they were limited to a certain entangled state class.

Other researchers relied on generative neural network models. Torlai et al. [30] and Carrasquilla et al. [31] exploited restricted Boltzmann machines (RBMs). Ahmed et al. [32] employed conditional generative adversarial networks (CGAN), in which they rely on a generator to reconstruct a density matrix estimate from measurement operators and statistics. The discriminator is then used to distinguish between the reconstructed density matrix and the statistical one. However, this method requires exponential measurements as an input to the CGAN.

Due to the complexities of the aforementioned studies, Murali et al. [6] and Nishio et al. [7] relied on simple method for approximating fidelity calculation that scales linearly with the total number of gates. Specifically, they rely on daily calibration errors provided by IBM's quantum backends for single and two qubits gates errors in addition to readout errors. Thus, the *estimated success probability* (ESP) of the circuit is based on error rates in single (ϵ_s) and two qubits (ϵ_e) gates along with readout errors (ϵ_r), shown in Eq. (1) where $G1$, $G2$, and N represent the number of single qubit gates, edges used by two qubits gates, and total number of qubits, respectively. This equation calculates the success rate of the circuit after subtracting gates' errors and qubits' readout errors. However, this method has multiple limitations. One limitation is that this method considers separate error forms and does not account for non-negligible errors like crosstalk. Another limitation is due to the product form equation of this method such that if one gate has an error of 1 then the whole circuit will have 0 success rate.

$$ESP = \prod_{s=1}^{G1}(1 - \epsilon_s) \times \prod_{e=1}^{G2}(1 - \epsilon_e) \times \prod_{r=1}^{N}(1 - \epsilon_r). \tag{1}$$

Liu and Zhou [5] estimated the reliability of circuits through the *probability of successful trials* (PST) metric which is the ratio between successful trials and total number of trials. Specifically, they utilize randomized benchmarking to generate random circuits using Clifford gates group such that each circuit is appended by its equivalent inverse gate at the end of the circuit so that ideally the quantum state will return to the original input state.

They then extract characteristics from the circuits like connectivity map, number of single and two qubit gates to estimate PST using two reliability estimation models: polynomial fitting and shallow neural network. Their results show that their method is comparable with Qiskit noise model but outperforms the ESP method. Their neural network model achieved a correlation factor of 0.9 while the Qiskit noise and the ESP methods achieved correlation factors of 0.89, and 0.78, respectively, with the observed PST values on quantum benchmarks. However, one limitation of this method is that it will infer the same PST value for different circuits if they have the same depth and number of gates.

The rest of this paper is organized as follows: Sect. 2 details the dataset collection process, describes the architecture of the proposed models, and explains the experimental design. Section 3 analyses and provides insights about the performance of our models in comparison with the baseline method on both random circuits and real-world quantum benchmarks. Finally, Sect. 4 concludes this paper and provides future work directions.

2 Models and methods

2.1 State fidelity estimation

Generally, quantum states can be divided into pure and mixed states. Pure states refer to the ideal condition of the quantum system. A pure quantum state can be described using the state vector notation, shown in Eq. (2) where $\alpha_{i \in [0,2^n-1]}$ is the amplitude and phase of the basis state i, or the density matrix notation, shown in Eq. (3).

$$|\psi_{\text{pure}}\rangle = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ . \\ . \\ \alpha_{2^n-1} \end{bmatrix}, \tag{2}$$

$$\rho = |\psi_{\text{pure}}\rangle\langle\psi_{\text{pure}}|. \tag{3}$$

Mixed states refer to the mixture of probabilities of multiple pure states that can only be represented using the density matrix, shown in Eq. (4) where p_i is the probability of pure state $|\psi_i\rangle$.

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{4}$$

To measure the closeness between any two quantum states, we estimate the *state fidelity*, shown in Eq. (5) where ρ_1 and ρ_2 are the density matrices for pure or mixed states and Tr is the trace. In this paper, we are concerned with the reconstruction of a pure quantum state $|\psi_{\text{pure}}\rangle$. Thus the state fidelity equation reduces to Eq. (6) where ρ_1 reduces to $|\psi_{\text{pure}}\rangle\langle\psi_{\text{pure}}|$ and ρ_2 represents the reconstructed density matrix for the mixed state outcome. Specifically, Eq. (6) measures the probability of the noisy mixed quantum state ρ_2 being equal to the ideal one $|\psi_{\text{pure}}\rangle$.

$$F(\rho_1, \rho_2) = \left[\text{Tr}\left(\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\right)\right]^2, \tag{5}$$

$$F(\rho_1, \rho_2) = \langle\psi_{\text{pure}}|\rho_2|\psi_{\text{pure}}\rangle, \quad \text{where } \rho_1 = |\psi_{\text{pure}}\rangle\langle\psi_{\text{pure}}|. \tag{6}$$

For dataset collection, we maximize the likelihood function, shown in Eq. (7) where $|y_j\rangle\langle y_j|$ are the projective measurements repeated with frequencies f_j, to reconstruct the noisy quantum state ρ_2. This is equivalent to the least-squares minimization under the assumption of Gaussian measurement noise [10]. The least-squares minimization method is fitted under the constraint that the reconstructed density matrix has a unit trace and is positive semi-definite.

$$L(\rho_2) = \prod_j \langle y_j|\rho_2|y_j\rangle^{f_j}. \tag{7}$$

2.2 Dataset collection

We generated random circuits consisting of different combinations of basis gates for 1, 3, and 5-qubit circuits using IBM's Qiskit software [33]. These basis gates include single qubit gates (identity, rotation Z, square-root NOT, and NOT) and a two qubit gate (controlled NOT). The identity gate (ID) is used for no operation, that is, a qubit is left as it is. The rotation Z gate (RZ) is a phase parameterized gate that rotates a qubit state around the z-axis with a specified angle. The square-root NOT gate (SX) transforms a qubit into a superposition state with different phase. The NOT gate (X) flips a qubit state. The controlled NOT gate (CX) is a two qubit gate consisting of a control qubit that flips the target qubit if the control qubit is in state $|1\rangle$. For each RZ gate, the angle was randomly generated. Each IBM device supports certain universal gate set referred to as "basis or physical gates". Therefore, gates in a quantum circuit must be composed using the basis gates before running on the IBM backend. For each circuit, we use Qiskit software to provide the corresponding state vector, generate state tomography projections, and estimate the state fidelity distance between the state tomography of the circuit and its state vector. We set the optimization level for all circuits to be 0, which means no optimization, in order to prevent any variations and accurately measure the state preparation and measurement errors on a real quantum device for the given high-level circuit. The number of shots for all circuits was fixed to 8190 shots.

For single qubit dataset, we generated different combinations of single qubit basis gates (ID, RZ, SX, and X) on ibmq_armonk machine of 5 stages depth threshold. We collected a total of 1024 circuits considering all possible combinations.

For three and five qubits datasets, we generated different combinations of both single and two qubits basis gates (ID, RZ, SX, X, and CX). For the three qubits dataset, we collected on ibmq_lima machine a total of 8,295 unique random circuits. The quantum volume for ibmq_lima machine is 8, specifying that the maximum square circuit size that can run successfully on this backend is 3×3. Thus, we fixed the depth of the collected 3-qubit circuits to 3. For the five qubits dataset, we collected on ibmq_quito machine a total of 5,429 unique random circuits, which is considered a small subset of the design space. For consistency with the 3-qubit dataset, we fixed the depth of the 5-qubit circuits to 3, which is close to the quantum volume of ibmq_quito with value 16. These datasets were then divided into 80% training and 20% testing without any overlap. Additionally, we generated 500 test circuits on ibmq_lima (3-qubit) and ibmq_quito (5-qubit) with 4 stages depth threshold, to test the generality of our model.

Figure 2 Single qubit convolutional model
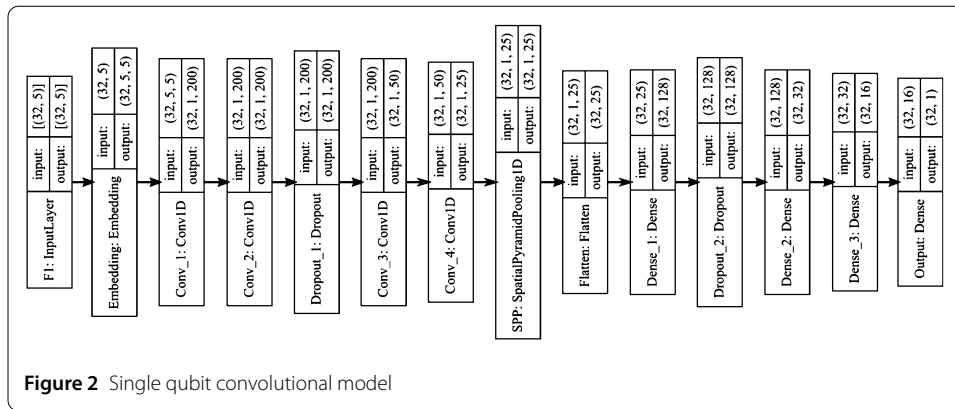
2.3 Proposed models

In this section, we proposed three deep learning models that have similar architectures, but differ in few hyperparameters, the number of layers used, and the input shapes. Additionally, the dataset used for each model is collected on different IBM backend (ibmq_armonk, ibmq_lima, and ibmq_quito). Each proposed architecture consists of an embedding layer, 1D convolution layers, a spatial pyramid pooling layer, and dense layers. The embedding layer provides a better dense encoding representation that learns relationships among gates in the circuit, unlike the use of integer encoding directly. The 1D convolution layers learn and extract unique patterns among gates from the embedding representation of the circuit. The spatial pyramid pooling layer is used to allow variable-sized circuit input while having a fixed-sized output for the dense layers. The dense layers are then finally used to estimate the state fidelity of a circuit.

2.3.1 1-*qubit convolutional model*

The input to this model is the integer encoding of gates in each circuit. Specifically, we map each gate to a unique integer value (ID: 1, RZ: 2, SX: 3, and X:4). The input is a vector of length 5, corresponding to the maximum depth of each circuit. The model, shown in Fig. 2, consists of an embedding layer that encodes the input into a dense representation resulting in an output dimension of length 5 for each element in the input. This representation is fed into 4 1D convolutional layers to extract unique features about the circuit consisting of 200, 200, 50, and 25 filters with kernel sizes of 5, 1, 1, and 1, respectively. The output is then followed by a spatial pyramid pooling layer consisting of 1 bin to allow variable-sized input. The output of the pooling layer is flattened to vector of length 25 which is then fully connected to 3 dense layers consisting of 128, 32, 16 neurons. This is finally connected to the output dense layer to estimate state fidelity. There are 2 dropout layers with 40% and 50% dropout rates. The activation function for all layers is the ReLU function except the output layer which has a sigmoid activation function as the expected fidelity value has a range from 0 to 1.

2.3.2 3-*qubit and* 5-*qubit convolutional models*

The input to these models consists of 2 matrices: gates and edges matrices. The gates matrix is the integer encoding as described in the single qubit model with the addition of the encoding of the two qubits gate (CX:5). The edges matrix is the encoding for the availability of CX gate and its direction. The availability of a CX gate is encoded to value 6

Figure 3 ibmq_lima and ibmq_quito backend layout
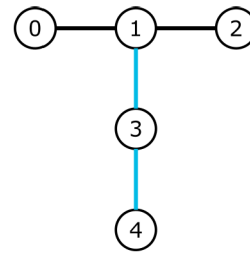
and 8 according to the connection direction of the control and target qubit. The absence of a CX gate is encoded to value 7.

Each of the 3-qubit circuits consists of 3-qubit and 3 stages depth threshold, thus the gates matrix has a shape of 3×3, which is flattened to a vector of length 9. There is a total of 2 edges, highlighted in blue in Fig. 3, for the 3-qubit circuits consisting of 3 stages, consequently the resulting edges matrix has a shape of 2×3 which is flattened to a vector of length 6. The flattened vectors of both matrices are fed to our model, shown in Fig. 4, which are first concatenated, then fed to an embedding layer to encode a dense representation. This model consists of 6 1D convolution layers consisting of 100, 50, 50, 50, 50, and 25 filters with alternating kernel sizes of (3, and 1) for each 2 consecutive convolution layers. This is then followed by a spatial pyramid pooling layer of 2 bins resulting in 2 different pooling layers with sizes 1, and 3. The resulting fixed length output has a length of 9. These sizes were empirically found to produce better results. The output of the pooling layer is then flattened to a vector of length 100 and fully connected to 2 dense layers of sizes 64 and 32, respectively. This is then connected to the output dense layer to finally output the state fidelity value. There are 2 dropout layers in this network, consisting of dropout rates of 40%. The activation function for all layers is the ReLU function except the output layer which has a sigmoid function.

Similar to the 3-qubit circuits, the gates matrix for 5-qubit circuits has a shape of 5×3, which is flattened to a vector of length 15. There is a total of 4 edges in the layout of the five qubits device, shown in Fig. 3, hence the resulting edge matrix has a shape of 4×3 which is flattened to a vector of length 12. Figure 5 represents the gates and edges encoding of a 5-qubit random circuit. The flattened vectors of both matrices are fed to our model, shown in Fig. 6, which are first concatenated to a length of 27 then fed to an embedding layer to encode a dense representation. This model consists of 6 1D convolution layers consisting of 100, 100, 50, 25, 50, and 25 filters with kernel sizes of 3, 1, 1, 1, 3, and 3, respectively. This is then followed by spatial pyramid pooling of 3 bins of sizes 1, 3, and 15 resulting in a fixed length output of 19. These sizes were empirically found to produce better results. The output of the pooling layer is then flattened to a vector of length 475 and fully connected to 2 dense layers of sizes 64 and 32, respectively. This is then connected to output dense layer to finally produce the state fidelity value. There are 3 dropout layers in this network, consisting of dropout rates of 30%, 30% and 50%. In the spatial pyramid pooling, we empirically found that using 3 bins of sizes 1, 3, and 15 produced better results. The resulting fixed length output has a length of 19. The activation function for all layers is the ReLU function except the output layer which has a sigmoid function.
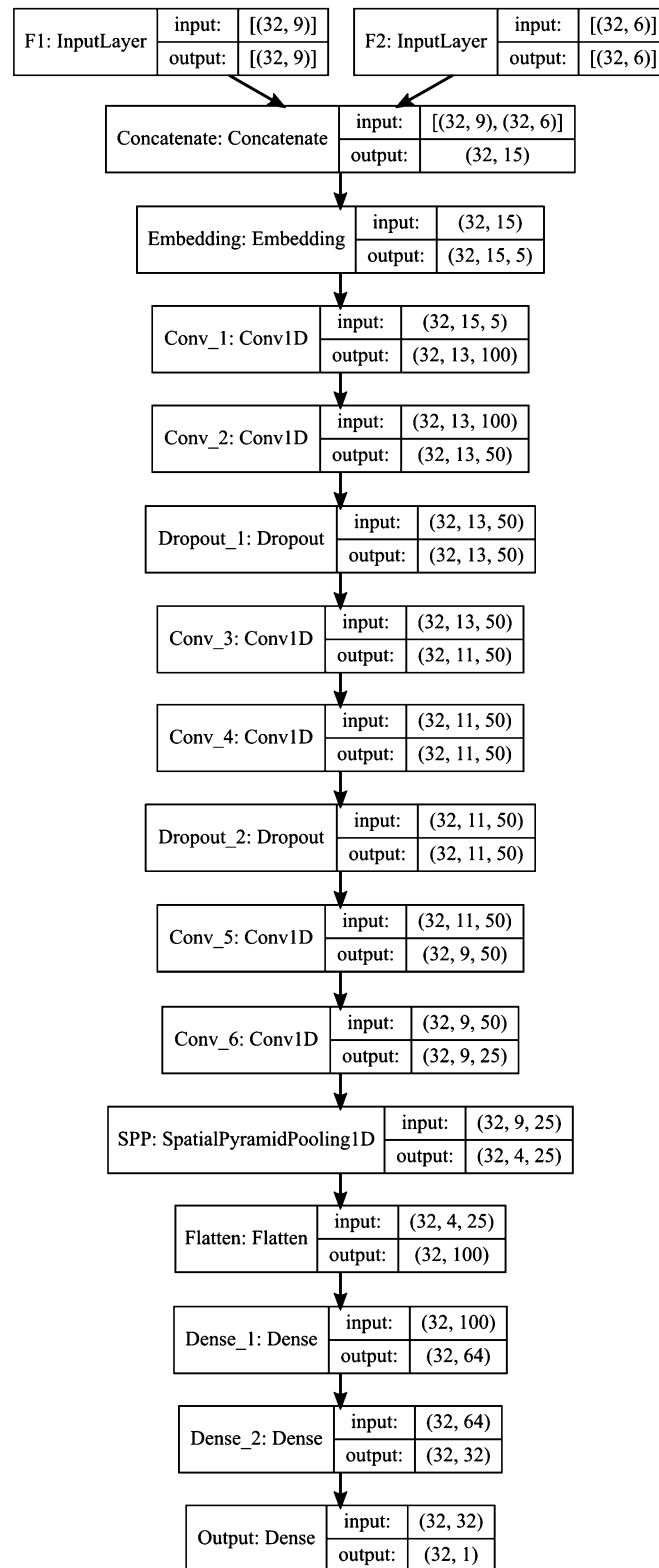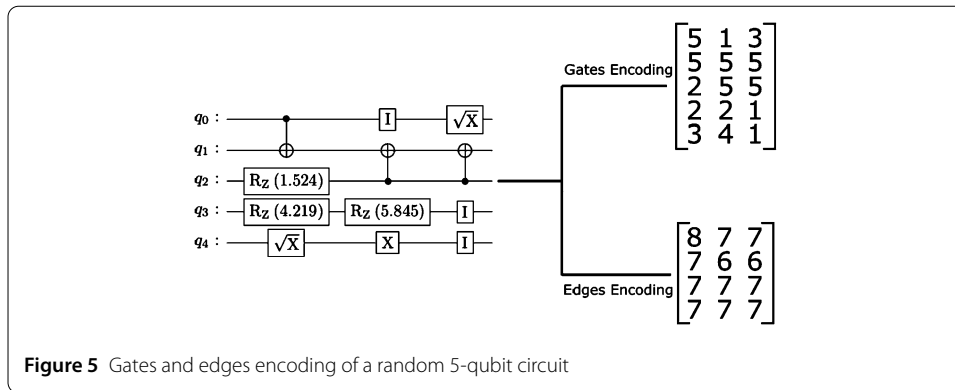
Figure 4 Three qubits convolutional model

Figure 5 Gates and edges encoding of a random 5-qubit circuit

2.4 Experimental design

The data collection process was run on IBM Quantum lab, including the state tomography process to generate the ground truth fidelities. First, we generate random circuits consisting of basis gates defined on IBM quantum backends. For ibmq_lima (5-qubit) backend, we fix three qubits (1, 3, and 4) in the topology layout, highlighted in blue in Fig. 3, so that the logical 3-qubit of the circuit will be mapped to these physical qubits in the same order. Similarly, we fix the logical to physical mapping of the 5-qubit circuits to be in the same order on ibmq_quito backend (for example logical 0 will be mapped to physical 0). Additionally, we constrain the controlled NOT (CX) basis gate according to the layout of the backend devices, shown in Fig. 3, where we allow the CX gate to only be applied if the 2-qubit are directly connected (e.g. qubits 1 and 3). The goal of these constraints is to learn the underlying hardware noise and avoid the addition of gates resulting from the swap operation. For each generated circuit, we estimate the ideal outcome state vector, shown in Eq. (2).

For the state tomography estimation, we utilize IBM's Qiskit software to generate the 3^n Pauli tomographic measurements for each circuit. After that we used IBM's tomography fitter to reconstruct the density matrix for the mixed state outcome from the tomographic measurements by the least-squares minimization method.

Then, we calculate the state fidelity distance, shown in Eq. (6) between the expected pure state vector $|\psi_{\text{pure}}\rangle$ and the reconstructed mixed state represented by the density matrix ρ_2.

The input to our models is the integer encoding circuit matrices instead of the tomographic measurements. Our representation relies on the circuits depths which can increase polynomially with the number of qubits for practical quantum circuits; if the depth of quantum circuits increases exponentially, then the quantum advantage over classical methods will be invalidated and will be impractical to run on current available backends that deal with decoherence noise as the results will be unreliable. For practical quantum circuits, our representation reduced the complexity of the input space from exponential to polynomial assuming arbitrary number of stages or circuit depth. In this paper, since we fixed the circuits' depth, the complexity of the input space is linear as it only linearly scales with the number of qubits. The output of our regressor models is the corresponding state fidelity of the circuit which indicates *the reliability of executing the circuit on a noisy hardware*.

Our 3 models were then implemented using Keras [34] as a python API of tensorflow [35]. We used Glorot uniform as a kernel initializer for convolution and dense layers. We
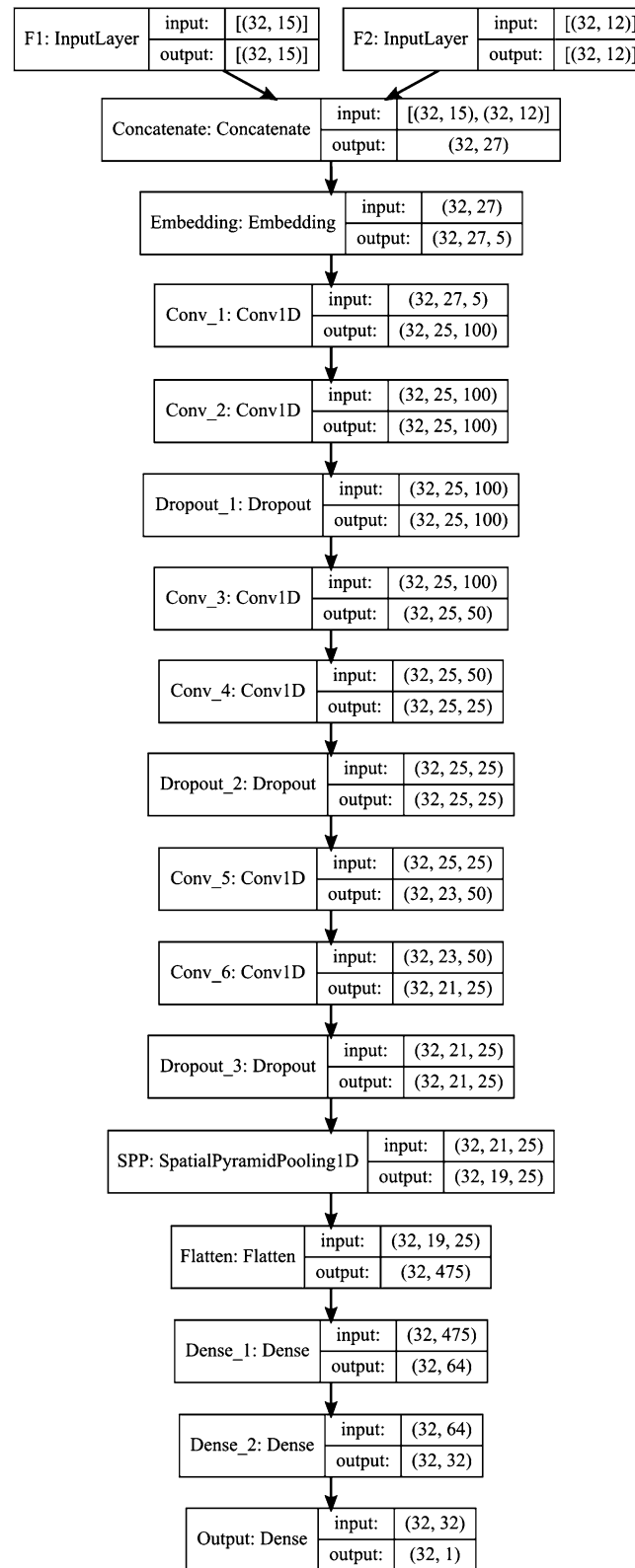
Figure 6 Five qubits convolutional model

chose to use an embedding layer because if we just relied on integer encoding of the circuit, it will imply an ordinal relationship among gates which is not the case. Also, if we used one hot encoding, it would result in a sparse input representation with no relationship among gates. Additionally, the use of an embedding layer in each model was found to provide a substantial improvements in the results, emphasizing its role in modeling complex relationships among gates in a circuit. Since convolutional layers allow more generalization in extracting features than dense layers, they are used for the unsupervised layers concerning feature extraction followed by a few dense layers for supervised classification or regression. This is demonstrated by the well-known deep neural networks as AlexNet [36] and VGGNet [37] networks which have multiple consecutive convolutional layers to demonstrate hierarchical representation of learning features. These convolutional layers are usually followed by a few dense layers for supervised classification. Therefore, we used multiple convolutional layers to extract and learn patterns from the embedding representation of a circuit which is followed by a few dense layers to output the predicted state fidelity. The 3 models were trained on a GPU (NVidia GeForce GTX 1060). We basically had 2 phases: the hyperparameter tuning phase and the experimental analysis phase. In the first phase, we performed manual hyperparameter tuning to the neural network and evaluated the model on a separate data sample (validation set). This manual tuning includes exploring different numbers of 1D convolutional layers (ranging from 2 to 8), varying numbers of dense layers (ranging from 1 to 6), using different optimizers (Adam, AdaDelta, and RMSProp) along with different learning rates (in the range from 0.0001 to 0.009). We also explored different kernel sizes (ranging from 2 to 8), different filter sizes (ranging from 10 to 200), different activation functions (ReLU, and sigmoid), and different batch sizes (16, 32, and 64). After we settled on the network architecture, we no longer tuned the 1-, 3- and 5-qubit models and proceeded with the second phase. In the experimental analysis phase, we repeated the experiment 10 times, for the 1-, 3- and 5-qubit models, with different random splits of training and testing to calculate the mean and standard deviation for the correlation factor, mean absolute error, and root mean square error to ensure that the model is stable and unbiased to certain split. This experiment was done after settling on the 1-, 3- and 5-qubit models' hyperparameters so that there is no data leakage. The three models were optimized using Adam optimizer with batch size set to 32. For single qubits models, we set the learning rate to 0.001, respectively. For the three and five qubits models before noise removal, we set the learning rate to 0.0009 and 0.0001, respectively. After noise removal, we slightly increased the learning rate for the 3 and 5-qubit models to 0.001 and 0.0003, respectively. Additionally, we used an activity regularizer before the output layer of 0.001 and 0.005 for single and five qubits models, respectively. The optimizer is then used to minimize the mean absolute error shown in Eq. (8), where S represents number of samples, Pred stands for predicted, and GT stands for ground truth.

$$\text{MAE} = \frac{1}{S} \sum_{i=0}^{S-1} \left| F_{\text{Pred}}^i - F_{\text{GT}}^i \right|. \tag{8}$$

Since available literature methods require exponential time complexity, we consider the method used by [6, 7] as the baseline method. Additionally, the ground truth fidelity we are considering already requires exponential complexity as the reconstruction process using maximum likelihood estimation relies on the tomographic measurements, serving as both ground truth and comparison with literature review.

3 Results and discussion

3.1 Results

Fidelity estimation is a regression task in which our goal is to estimate the state fidelity of a circuit *through its representation only without undergoing exponential tomographic measurements*, thus reducing the input space complexity. For a 5-qubit circuit, our 2 matrices representation would require a total of 27 elements for depth-3 that could increase polynomially with the number of qubits. Thus, our representation for an efficient quantum circuit consisting of n-qubit and N gates and having depth-l will be $O(\text{poly}(n))$. However, the exponential tomographic measurement would require a total of $7,776$ measurements outcomes given that there is a total of 3^n possible Pauli combinations that lead to 6^n possible outcomes as each measurement operator has 2 possible outcomes, where n is the number of qubits.

To quantify the performance of our models, we measure the correlation factor, the mean absolute error (MAE), and the root mean square error (RMSE) between the predicted and the ground truth fidelities shown in equations (8), (9), and (10) respectively, where Pred stands for predicted, GT stands for ground truth, Cov represents the covariance, σ is the standard deviation, and S is the number of samples. We also calculate the percentage error (*%Error_i*), shown in equation (11), between predicted and ground truth state fidelity for circuit i. Due to the small variation in the state fidelity range, the correlation factor illustrates better the relationship between the predicted and ground truth fidelity than the use of mean absolute error and root mean square error.
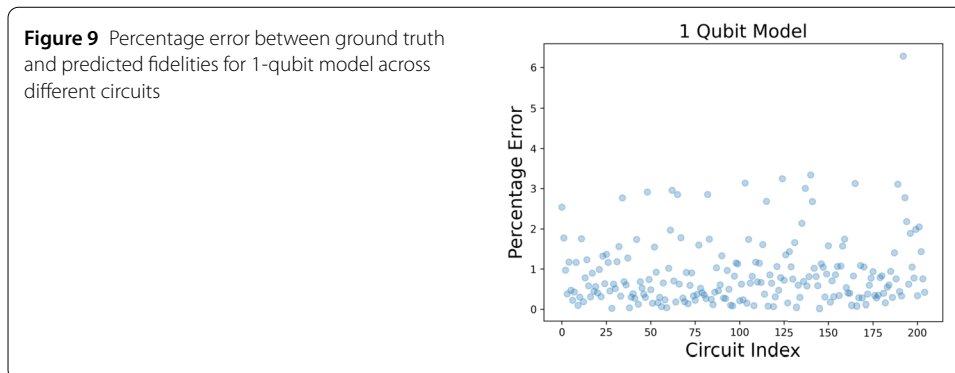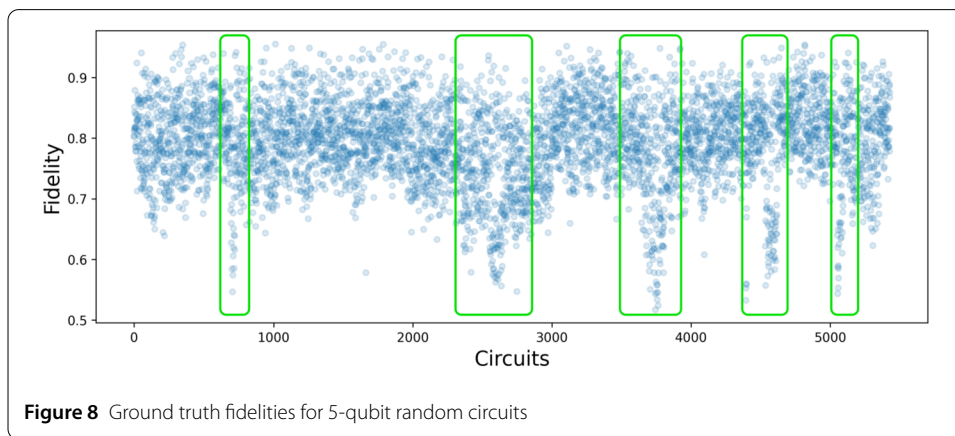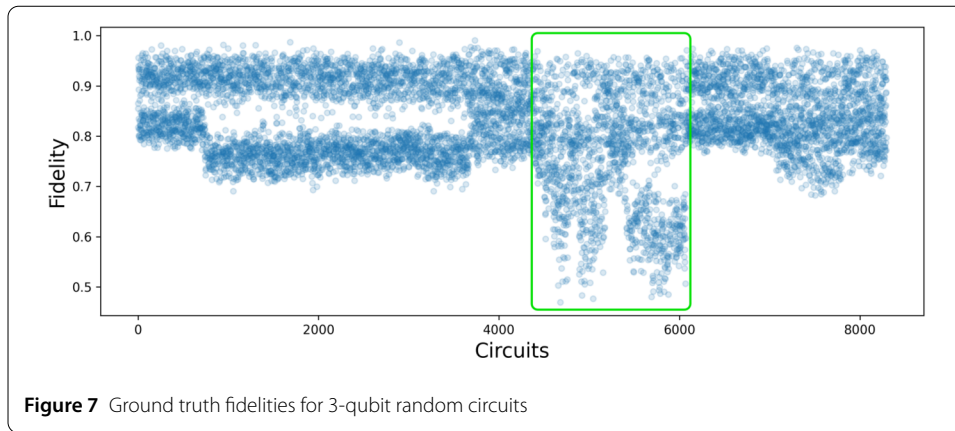
$$r(F_{\text{Pred}}, F_{\text{GT}}) = \frac{\text{Cov}(F_{\text{Pred}}, F_{\text{GT}})}{\sigma_{F_{\text{Pred}}} \times \sigma_{F_{\text{GT}}}}, \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{S} \sum_{i=0}^{S-1} \left(F_{\text{GT}}^i - F_{\text{Pred}}^i\right)^2} \tag{10}$$

$$\%\text{Error}_i = \frac{|F_{\text{GT}}^i - F_{\text{Pred}}^i|}{F_{\text{GT}}^i} * 100. \tag{11}$$

Additionally, we analyze the dynamic effect of noise on both 3-qubit and 5-qubit circuits' state fidelities by removing noisy regions, shown in Figs. 7 and 8, respectively. These figures represent the ground truth state fidelities of random circuits running on IBM backends sequentially in time. The bounded rectangular regions in both figures, highlighted in green, represent noise clustered regions that are circuit independent as they occur in limited regions and do not spread across the whole dataset. We eliminate these regions as they represent instability periods of backend devices, resulting in erroneous fidelities. In this paper, we consider both scenarios: before and after the removal of noise.

For the single qubit dataset, Fig. 9 represents the percentage error between ground truth and predicted state fidelities across different circuits of our 1D convolutional model on our test set running on ibmq_armonk backend. The results show that the percentage error across different circuits has a mean of $0.89\% \pm 0.84\%$ reflecting the high prediction accuracy of our model with the ground truth fidelities. We achieved a mean correlation factor of 0.7447 ± 0.0374, shown in Fig. 10, a mean absolute error of 0.0094 ± 0.008, and a root mean square error of 0.0129 ± 0.001. The results on this small-scale model shows that there is a potential relationship between transformations applied to qubits and its state fidelity on a specific hardware device. These results motivated us to extend this model to

Figure 7 Ground truth fidelities for 3-qubit random circuits

Figure 8 Ground truth fidelities for 5-qubit random circuits

Figure 9 Percentage error between ground truth and predicted fidelities for 1-qubit model across different circuits

3 and 5-qubit circuits, as the single qubit model provided insight about the feasibility of this method for this simplistic case.

For the three qubit dataset, Figs. 11 and 12 show the percentage error between ground truth and predicted state fidelities across different circuits of the test set for 3-qubit circuits running on ibmq_lima backend before and after noise removal. The results show that the percentage error before noise removal across different circuits has a mean of $5.65\% \pm 8.54\%$, while after noise removal the percentage error is further reduced achieving a mean of $3.23\% \pm 3.04\%$. The high standard deviation before noise removal demonstrates high dynamic noise rates available in ibmq_lima. This is also illustrated in Fig. 12

Figure 10 Predicted and ground truth comparison plot for 1-qubit model of depth-5 stages ($r = 0.74$)
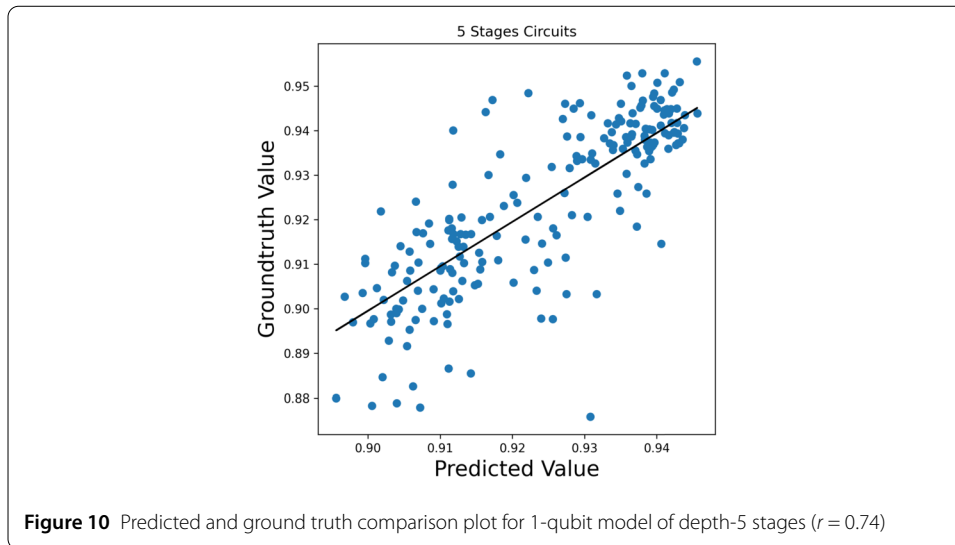
where percentage error is more compact with almost no outliers as in Fig. 11. Similarly, for the five qubits dataset, Figs. 16 and 17 represent the percentage error between ground truth and predicted fidelities across different circuits on ibmq_quito backend. The results show that the percentage error across different circuits before noise removal has a mean of $5.06\% \pm 5.09\%$, while the percentage error after noise removal has a mean of $3.92\% \pm 3.41\%$. This reduction in both the mean and the standard deviation is also depicted in Fig. 17 where few outliers exist than in Fig. 16.

Table 1 shows a comparison between the results before and after noise removal for the 3 and 5-qubit models. The results show significant improvements over the consideration of the dataset without noise removal, and emphasize the effect of stable backend devices to achieve high reliability. For the 3-qubit model, the correlation factor after noise removal was 0.828 ± 0.0129, shown in Fig. 14, while the one before noise removal was 0.6745 ± 0.0152, shown in Fig. 13. Similarly, for the 5-qubit model, the correlation factor after noise removal was 0.745 ± 0.0119, shown in Fig. 19, while the one before noise removal was 0.6637 ± 0.0121, shown in Fig. 18. Also, the MAE and RMSE after noise removal slightly decreased than before noise removal for both models.

For the rest of the paper, we will work with 3- and 5-qubit models without noise removal, assuming worst-case scenario. We then employed transfer learning to test our model on circuits with increased depth of 4 stages while trained on circuits with depth-3 for both 3- and 5-qubit models. For the 3-qubit models, we achieved a mean correlation factor of 0.3471 ± 0.0448, shown in Fig. 15, a mean absolute error of 0.0536 ± 0.0034, and a root mean square error of 0.0694 ± 0.0038. For the 5-qubit models, we achieved a mean correlation factor of 0.3582 ± 0.0147, shown in Fig. 20, a mean absolute error of 0.0821 ± 0.0036, and a root mean square error of 0.1013 ± 0.0036. These results show that even though our models were trained on certain depth, they can still generalize to higher depth. This opens the direction for further research to fine-tune these models to higher depth.

Additionally, we compare our 5-qubit model with a simple baseline method (ESP), shown in Eq. (1), on 250 additionally randomly generated circuits on ibmq_quito. Although the ESP baseline method achieved a mean absolute error (MAE) of 0.04 and a root mean square error (RMSE) of 0.06 with ground truth fidelities, it achieved a low cor-

Figure 11 Percentage error between ground truth and predicted fidelities for 3-qubit model across different circuits
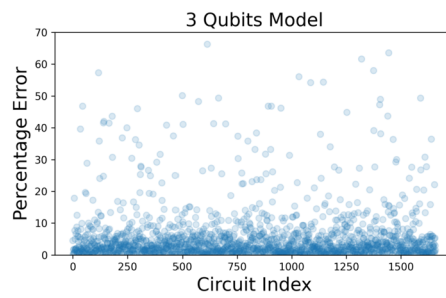
Figure 12 Percentage error between ground truth and predicted fidelities for 3-qubit model across different circuits after noise removal from dataset
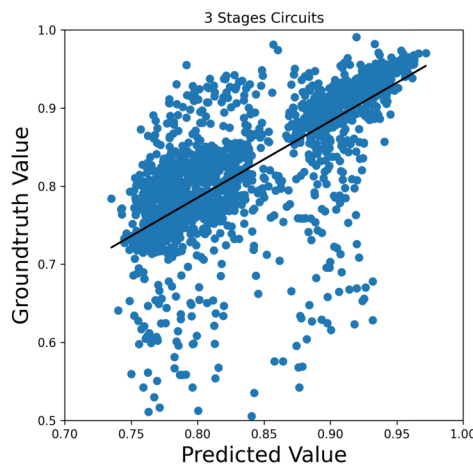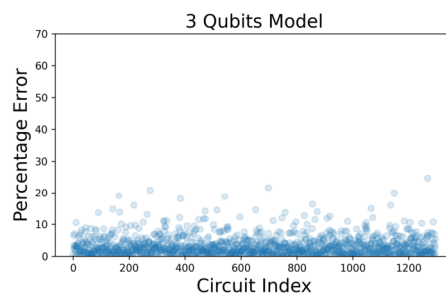
Figure 13 Predicted and ground truth comparison plot for 3-qubit model on depth-3 stages ($r = 0.67$)

relation factor of 0.10, shown in Fig. 21. Our deep learning method achieved a MAE of 0.03, a RMSE of 0.04, and a correlation factor of 0.687. Thus, the approximate fidelity calculation method, given its simplicity, cannot be used for state tomography due to its poor performance.

3.2 Quantum benchmarks

In this section, we demonstrate the effectiveness of our 5-qubit model trained on random circuits on specific quantum benchmarks. Specifically, we verify our proposed model on three real-world quantum algorithms and protocols (Bernstein–Vazirani, Simon, and Superdense coding). Bernstein–Vazirani algorithm [38] is concerned with finding a secret string by querying an oracle one time only, providing a polynomial speedup over its clas-
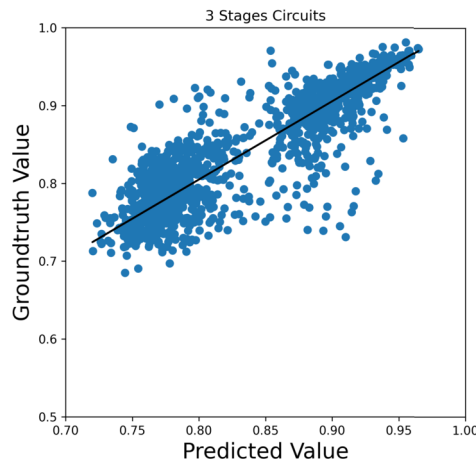
Figure 14 Predicted and ground truth comparison plot for 3-qubit model on depth-3 stages after noise removal from dataset ($r = 0.82$)
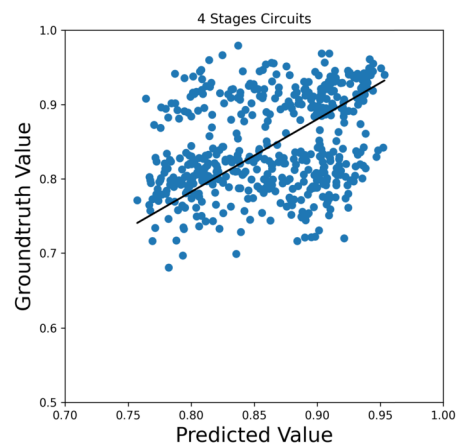
Figure 15 Predicted and ground truth comparison plot for 3-qubit model on depth-4 stages ($r = 0.34$)
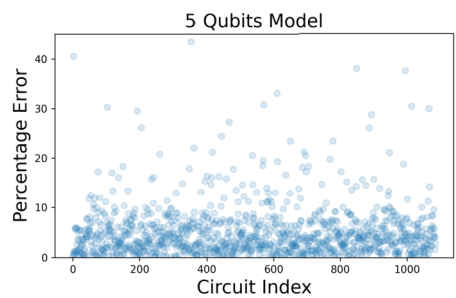

Figure 16 Percentage error between ground truth and predicted fidelities for 5-qubit model across different circuits

sical counterpart. Simon's algorithm [39] solves the problem of finding a periodic string s such that when querying an oracle both input x and $x \oplus s$ will have the same output. This algorithm provides an exponential speedup over the classical one. Superdense coding [40] is the process of sending 2 classical bits between two distant parties across a quantum communication channel. The sender encodes a single qubit such that when decoded, the receiver is able to reveal the classical bits.

Figure 17 Percentage error between ground truth and predicted fidelities for 5-qubit model across different circuits after noise removal from dataset
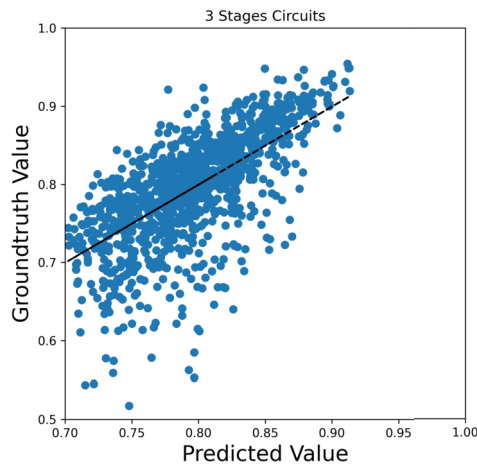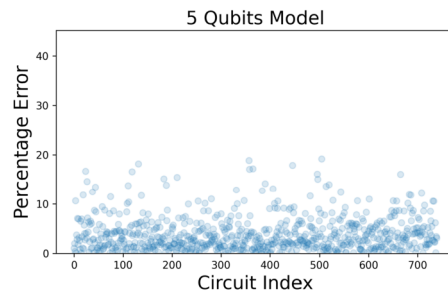

Figure 18 Predicted and ground truth comparison plot for 5-qubit model on depth-3 stages ($r = 0.66$)
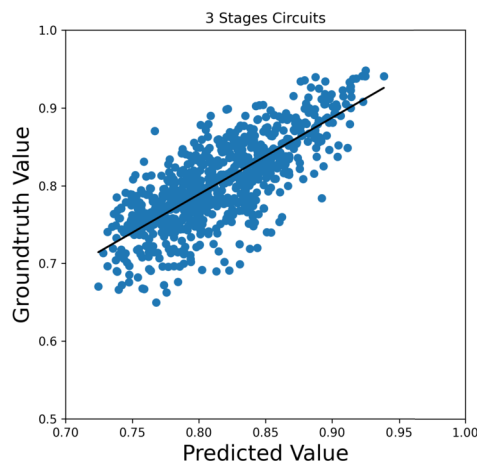

Figure 19 Predicted and ground truth comparison plot for 5-qubit model of depth-3 stages after noise removal from dataset ($r = 0.74$)

For Bernstein–Vazirani and Simon's algorithms, we use a simplified version of these algorithms such that the circuits conform with our constraints. Specifically, we remove stages that contain only RZ gates, shown in Fig. 22, as IBM Qiskit implements this gate virtually [41] requiring zero time and error. Thus, these circuits fit to our constraints with-
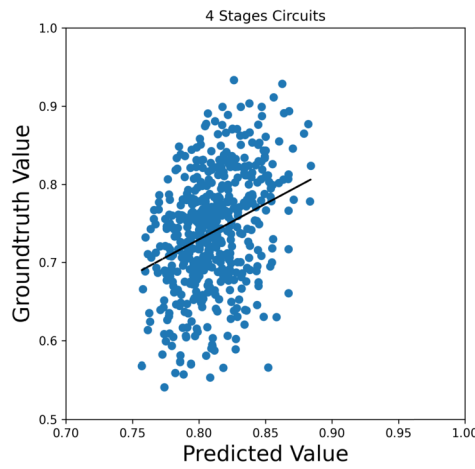
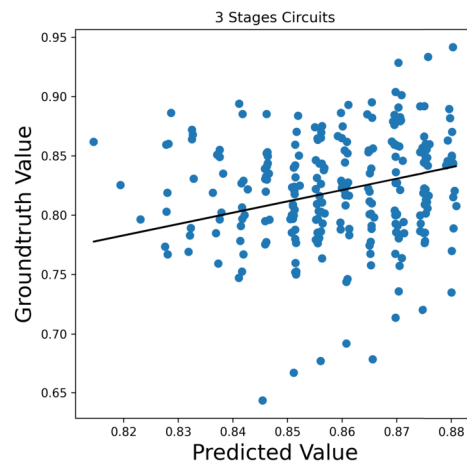**Figure 20** Predicted and ground truth comparison plot for 5-qubit model on depth-4 stages ($r = 0.35$)

Figure 21 Approximate Fidelity and ground truth comparison plot for 5-qubit circuits on depth-3 stages ($r = 0.1$)

Table 1 Mean results (± standard deviation) before and after noise removal between ground truth state tomography and our models

Model	Before Noise Removal			After Noise Removal		
	Correlation factor	MAE	RMSE	Correlation factor	MAE	RMSE
3-qubit model	0.6745	0.0414	0.0655	0.828	0.029	0.0401
	±0.0152	±0.0012	±0.0020	±0.0129	±0.0009	±0.0014
5-qubit model	0.6637	0.0388	0.0519	0.745	0.0307	0.0389
	±0.0121	±0.0008	±0.0010	±0.0119	±0.0010	±0.0011

out sacrificing any gate error. For the superdense coding protocol, we only use the optimized versions of the circuits as they already fulfill our constraints. For comparison, we implement the baseline fidelity estimation, shown in Eq. (1), to the simplified versions of the three benchmarks.
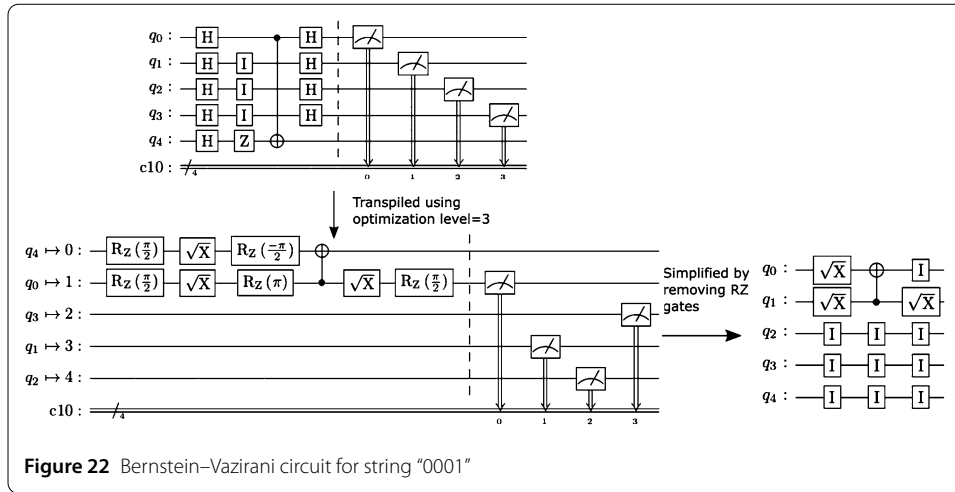
Figure 22 Bernstein–Vazirani circuit for string "0001"

Table 2 Comparison between ground truth state tomography (optimized and simplified circuits), predicted and baseline state fidelities using our model

Benchmark	Output	GT Optimized Fidelities	GT Simplified Fidelities	Predicted Fidelities	Baseline (ESP) Fidelities
Bernestein–Vazirani 4-bit	"0001"	0.84	0.83	0.85	0.84
Simon 2-bit	"00"	0.78	0.76	0.81	0.86
	"01"	0.84	0.81	0.82	0.90
	"10"	0.83	0.80	0.82	0.89
Superdense coding	"01"	0.89	–	0.89	0.93
	"10"	0.89	–	0.88	0.97
	"11"	0.84	–	0.85	0.93

Table 2 represents a comparison between ground truth state tomography (optimized and simplified circuits) with the predicted and baseline fidelities of our model across benchmarks. The discrepancy between optimized and simplified ground truth fidelities is due to the readout errors which are inversely proportional to the number of ones or hamming weight in a measured state according to [42]. To demonstrate the advantage of the predicted fidelities of our model over the estimated fidelities of the baseline method, we computed the correlation factor with the ground truth fidelities of the simplified versions of Bernstein–Vazirani and Simon and the optimized version of superdense coding. For the three benchmarks, our method achieved a correlation factor of 0.968 while the baseline method achieved a correlation factor of 0.738 with the ground truth fidelities.

4 Conclusion

In this paper, we proposed three deep learning models to estimate the state fidelities for 1-, 3-, and 5-qubit circuits. In particular, we exploited an embedding layer to encode a dense representation of a circuit instead of tomographic measurements that scale exponentially with the number of qubits, reducing the input space for practical quantum circuits from exponential to polynomial complexity. Specifically, we aimed at modeling the relationship between gates and circuits' state fidelities. Experimental results show that there is a high correlation between the predicted and ground truth state fidelities achieving a correlation factor of 0.74, 0.67, and 0.66 for single, three, and five-qubits circuits, respectively. To test the generality of the three and five qubits circuits, we increased the circuits' depth to 4 while training our model only on depth-3 circuits. We achieved a mean correlation factor

of 0.34 and 0.35 on three and five qubits circuits, respectively. We also investigated the dynamic effect of noise on the circuits' state fidelities. We achieved correlation factor of 0.828 and 0.745 for three and five-qubits circuits after removing noisy regions from our dataset. Additionally, we compared our 5-qubit model and the baseline state fidelity estimation method with the exponential state tomography method on three quantum benchmarks. Our method achieved a correlation factor of 0.96 with the exponential state tomography method while the baseline method achieved 0.73. Thus, our results demonstrate the learnability of quantum state fidelity which is considered a step towards an efficient end-to-end learning. Machine learning methods typically require few data points to train a neural network model. To generate a data point which is composed of a circuit and its corresponding ground truth state fidelity, we need to utilize state tomography (which is of exponential complexity on the number of qubits). As future work, we plan to reduce this complexity to allow scalability to a higher number of qubits. Additionally, we plan to train our model for specific quantum algorithms like quantum approximate optimization algorithm (QAOA). We also plan to explore stage based fidelity estimation using long short-term memory neural network.

Declarations

Competing interests
The authors declare no competing interests.

Author contributions
AE formulated the problem of utilizing machine learning for fidelity estimation. NE performed the data collection and code implementation. WG analyzed and interpreted the results. KK and KU contributed to problem positioning and presentation. All authors read and approved the final manuscript.

Author details
[1]Department of Computer Science and Engineering, Egypt-Japan University for Science and Technology, Alexandria, Egypt. [2]Faculty of Engineering, Alexandria University, Alexandria, Egypt. [3]Waseda University, Tokyo, Japan.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Preskill J. Quantum computing in the NISQ era and beyond. Quantum. 2018;2:79.
2. Jozsa R. Fidelity for mixed quantum states. J Mod Opt. 1994;41(12):2315–23.
3. Altepeter JB, Jeffrey ER, Kwiat PG. Photonic state tomography. Adv At Mol Opt Phys. 2005;52:105–59.
4. Emerson J, Alicki R, Zyczkowski K. Scalable noise estimation with random unitary operators. J Opt B, Quantum Semiclass Opt. 2005;7:347–52. https://doi.org/10.1088/1464-4266/7/10/021. arXiv:quant-ph/0503243.
5. Liu J, Zhou H. Reliability modeling of nisq-era quantum computers. In: 2020 IEEE international symposium on workload characterization (IISWC). New York: IEEE Press; 2020. p. 94–105.
6. Murali P, Baker JM, Javadi-Abhari A, Chong FT, Martonosi M. Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers. In: Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems. 2019. p. 1015–29.
7. Nishio S, Pan Y, Satoh T, Amano H, Meter RV. Extracting success from ibm's 20-qubit machines using error-aware compilation. ACM J Emerg Technol Comput Syst. 2020;16(3):1–25.

8. Wilson E, Singh S, Mueller F. Just-in-time quantum circuit transpilation reduces noise. In: 2020 IEEE international conference on quantum computing and engineering (QCE). New York: IEEE Press; 2020. p. 345–55.

9. Banaszek K, D'ariano G, Paris M, Sacchi M. Maximum-likelihood estimation of the density matrix. Phys Rev A. 1999;61(1):010304.

10. Smolin JA, Gambetta JM, Smith G. Efficient method for computing the maximum-likelihood quantum state from measurements with additive Gaussian noise. Phys Rev Lett. 2012;108(7):070502.

11. Gross D, Liu Y-K, Flammia ST, Becker S, Eisert J. Quantum state tomography via compressed sensing. Phys Rev Lett. 2010;105(15):150401.

12. Flammia ST, Gross D, Liu Y-K, Eisert J. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. New J Phys. 2012;14(9):095022.

13. Bolduc E, Knee GC, Gauger EM, Leach J. Projected gradient descent algorithms for quantum state tomography. npj Quantum Inf. 2017;3(1):1–9.

14. Shang J, Zhang Z, Ng HK. Superfast maximum-likelihood reconstruction for quantum tomography. Phys Rev A. 2017;95(6):062336.

15. Qi B, Hou Z, Li L, Dong D, Xiang G, Guo G. Quantum state tomography via linear regression estimation. Sci Rep. 2013;3(1):1–6.

16. Hou Z, Zhong H-S, Tian Y, Dong D, Qi B, Li L, Wang Y, Nori F, Xiang G-Y, Li C-F et al. Full reconstruction of a 14-qubit state within four hours. New J Phys. 2016;18(8):083036.

17. Qi B, Hou Z, Wang Y, Dong D, Zhong H-S, Li L, Xiang G-Y, Wiseman HM, Li C-F, Guo G-C. Adaptive quantum state tomography via linear regression estimation: theory and two-qubit experiment. npj Quantum Inf. 2017;3(1):1–7.

18. Ferrie C. Self-guided quantum tomography. Phys Rev Lett. 2014;113:190404. https://doi.org/10.1103/PhysRevLett.113.190404.

19. Rambach M, Qaryan M, Kewming M, Ferrie C, White AG, Romero J. Robust and efficient high-dimensional quantum state tomography. Phys Rev Lett. 2021;126(10):100402.

20. Farooq A, Ullah MA, Ramadhani S, Shin H, et al. Self-guided quantum state learning for mixed states. 2021. arXiv preprint. arXiv:2106.06166.

21. Blume-Kohout R. Optimal, reliable estimation of quantum states. New J Phys. 2010;12(4):043034.

22. Granade C, Combes J, Cory D. Practical Bayesian tomography. New J Phys. 2016;18(3):033024.

23. Lukens JM, Law KJ, Jasra A, Lougovski P. A practical and efficient approach for Bayesian quantum state estimation. New J Phys. 2020;22(6):063038.

24. Gühne O, Lu C-Y, Gao W-B, Pan J-W. Toolbox for entanglement detection and fidelity estimation. Phys Rev A. 2007;76(3):030305.

25. Tiurev K, Sørensen AS. Fidelity measurement of a multiqubit cluster state with minimal effort. 2021. arXiv preprint. arXiv:2107.10386.

26. Xu Q, Xu S. Neural network state estimation for full quantum state tomography. 2018. arXiv preprint. arXiv:1811.06654.

27. Lohani S, Searles TA, Kirby BT, Glasser RT. On the experimental feasibility of quantum state reconstruction via machine learning. 2020. arXiv preprint. arXiv:2012.09432.

28. Zhang X, Luo M, Wen Z, Feng Q, Pang S, Luo W, Zhou X. Direct fidelity estimation of quantum states using machine learning. 2021. arXiv preprint. arXiv:2102.02369.

29. Cha P, Ginsparg P, Wu F, Carrasquilla J, McMahon PL, Kim E-A. Attention-based quantum tomography. 2020. arXiv preprint. arXiv:2006.12469.

30. Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R, Carleo G. Neural-network quantum state tomography. Nat Phys. 2018;14(5):447–50.

31. Carrasquilla J, Torlai G, Melko RG, Aolita L. Reconstructing quantum states with generative models. Nat Mach Intell. 2019;1(3):155–61.

32. Ahmed S, Muñoz CS, Nori F, Kockum AF. Quantum state tomography with conditional generative adversarial networks. 2020. arXiv preprint. arXiv:2008.03240.

33. Anis MS et al. Qiskit: an open-source framework for quantum computing. 2021. https://doi.org/10.5281/zenodo.2573505.

34. Chollet F, et al., Keras. 2015. https://keras.io.

35. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. TensorFlow: large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org. https://www.tensorflow.org/.

36. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25.

37. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2015.

38. Bernstein E, Vazirani U. Quantum complexity theory. SIAM J Comput. 1997;26(5):1411–73.

39. Simon DR. On the power of quantum computation. SIAM J Comput. 1997;26(5):1474–83.

40. Bennett CH, Wiesner SJ. Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states. Phys Rev Lett. 1992;69:2881–4. https://doi.org/10.1103/PhysRevLett.69.2881.

41. McKay DC, Wood CJ, Sheldon S, Chow JM, Gambetta JM. Efficient z gates for quantum computing. Phys Rev A. 2017;96(2):022330.

42. Tannu SS, Qureshi MK. Mitigating measurement errors in quantum computers by exploiting state-dependent bias. In: Proceedings of the 52nd annual IEEE/ACM international symposium on microarchitecture. 2019. p. 279–90.