Random projection using random quantum circuits

Keerthi Kumaran,[1] Manas Sajjan,[2] Sangchul Oh,[2,3] and Sabre Kais[1,2,4,*]
[1]*Department of Physics and Astronomy, Purdue University, West Lafayette, Indiana 47907, USA*
[2]*Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, USA*
[3]*Department of Physics, Southern Illinois University, Carbondale, Illinois 62901-4401, USA*
[4]*Department of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, USA*

(Received 19 August 2023; accepted 29 November 2023; published 3 January 2024)

The random sampling task performed by Google's Sycamore processor gave us a glimpse of the "quantum supremacy era." This has definitely shed some light on the power of random quantum circuits in this abstract task of sampling outputs from the (pseudo)random circuits. In this paper, we explore a practical near-term use of local random quantum circuits in dimensional reduction of large low-rank data sets. We make use of the well-studied dimensionality reduction technique called the random projection method. This method has been extensively used in various applications such as image processing, logistic regression, entropy computation of low-rank matrices, etc. We prove that the matrix representations of local random quantum circuits with sufficiently shorter depths [$\sim O(n)$] serve as good candidates for random projection. We demonstrate numerically that their projection abilities are not far off from the computationally expensive classical principal components analysis on MNIST and CIFAR-100 image datasets. We also benchmark the performance of quantum random projection against the commonly used classical random projection in the tasks of dimensionality reduction of image data sets and computing von Neumann entropies of large low-rank density matrices. And finally, using variational quantum singular value decomposition, we demonstrate a near-term implementation of extracting the singular vectors with dominant singular values after quantum random projecting a large low-rank matrix to lower dimensions. All such numerical experiments unequivocally demonstrate the ability of local random circuits to randomize a large Hilbert space at sufficiently shorter depths with robust retention of properties of large data sets in reduced dimensions.

DOI: 10.1103/PhysRevResearch.6.013010

I. INTRODUCTION

Many problems in machine learning and data science involve the dimensional reduction of large data sets with low ranks [1] (e.g., image processing). Dimensional reduction as a preprocessing step reduces computational complexity in the later stages of processing. Principal component analysis (PCA) [2], reliant on singular value decomposition (SVD), is one such method to reduce the dimension of data sets by retaining only the singular vectors with dominant singular values. There are quantum circuit implementations for PCA (and SVD) [3–6] and for related applications [7], some of which are near-term (noisy intermediate-scale quantum technologies era [8]) algorithms [6].

Techniques like PCA (and SVD) involve a complexity of $O(N^3)$, where N is the size (or the dimension) of data vectors. An alternative to such computationally expensive dimensional reduction methods is the random projection method [9–11]. In the random projection method, we multiply the data sets with certain random matrices and project them to a lower-dimensional subspace. Recent years have witnessed fruitful usage of an especially thoughtful variant of such random projections which are known to preserve the distance between any two vectors in the data set (say $\vec{x_1}$ and $\vec{x_2}$) in the projected subspace up to an error that scales as $O(\sqrt{\frac{\log(N)}{k}})$, where N is the original dimension and k is the reduced dimension of each data vector. This choice is motivated by the Johnson-Lindenstrauss (JL) lemma [12] introduced at the end of the last century. Since this paper will exclusively use such transformations to validate all the key results, we hereafter refer to such candidates as good random projectors. Such projection techniques are beneficial to myriad applications because the preservation of distances between data vectors ensures that their distinctiveness is uncompromised, thereby rendering them usable for discriminative tasks such as classification schemes like logistic regression [13].

Classically, this is advantageous compared to other methods like PCA because the random matrix used for projection is independent of the data set considered. The time complexity involved in the random projection arises from a matrix multiplication complexity $O(N^{2.37})$ [14] followed by the usual SVD complexity of $O(N^2 \text{poly}\log(N))$, making the resulting scheme cheaper than PCA (or SVD). It must be emphasized that a further reduction in the time complexity to $O(N\text{poly}\log(\text{N}))$ can be afforded using the fast

*kais@purdue.edu

Johnson-Lindenstrauss transforms [15]. Several candidates have been studied in classical random projection, including Haar random matrices, Gaussian random matrices, etc. But the memory complexity of storing such matrices can be potentially huge (proportional to N^2 times the precision of each matrix entry). This has engendered the introduction of several competing candidates with better memory complexity (containing sparse matrices with random integer entries) and multiplication complexity. The latter category is mainly considered in practical applications today [10] and will also be used to compare the results of the quantum variants in this paper.

Classically random projections performed by using projectors sampled from Haar random unitaries suffer from the innate problem of storage due to their exceptionally high memory usage. Even in the quantum setting, implementing Haar random unitaries requires exponential resources, as shown in some counting arguments [16]. As a result, it is natural to consider unitary t-designs which only match the Haar measure up to tth moments. Quantum implementation of such t-designs, as has been studied in this paper, is efficient owing to the fact that local random quantum circuits approach approximate unitary t-designs [17–19] in sufficiently shorter [$O(\log(N)t^{10.5}$] depths [20,21] (Here, we have assumed that the number of qubits n required to encode a data vector or a wave vector of size N is $\sim \log(N)$). It was shown recently that even shorter depths suffice [22]. The primary workhorse of this paper will be based on such quantum circuits which as we eventually show not only perform better in accuracy than standard more commonly used classical variants but also require a lesser number of single-qubit random rotation gates $O(\text{poly}(\log(N)))$ for implementation.

The flow of the paper is as follows. In Sec. II, we begin with an introduction to the JL lemma and how it makes the random projection method effective. This is followed by a brief introduction to the Haar measure and approximate Haar unitaries generated from local random quantum circuits. Then, we explicitly prove that the local random quantum circuits which are exact unitary 2-designs can satisfy the JL lemma with the same high probability as Haar random matrices, thereby making them good random projectors. We then extend the results to approximate unitary 2-designs and discuss the bounds on depths to achieve a certain error threshold in the JL lemma and derive a slightly different probability of the satisfaction of the latter. We would note that the quantum memory required to store a 2-design or approximate 2-design is $O(\text{poly}(\log N))$ where N is the size of the data vector. It is worth noting that it was previously shown in Ref. [23] that approximate unitary t-designs with $t = O(k)$ can be used to satisfy the JL lemma, thus corroborating our assertions that even they are good candidates for random projection. The exponentially low limit obtained in Ref. [23] is better than the limit derived in this paper only for system sizes $N \geqslant O(10^4)$. For $N \sim O(10^3)$, limits derived in this paper for unitary 2-designs are tighter.

For numerical quantification of the key assertions, we first use the MNIST (Modified National Institute of Standards and Technology) and CIFAR(Canadian Institute for Advanced Research)-100 image datasets [24,25] and show that the quantum random projection preserves distances postprojection not far off from the computationally expensive algorithms like PCA (along the lines similar to Ref. [11]) and is similar to the classical random projection. This task does not require one to know the singular values or the singular vectors explicitly. We compare the performance of quantum random projection with the commonly used classical random projection technique. Instead of benchmarking it against the Haar random matrices generated classically, we make use of classical random projectors whose storage and multiplications are efficient. To this end, we use a subsampled randomized Hadamard transform (SRHT) [15] for different sizes of data sets (1024 and 2048, corresponding to 10 and 11 qubits, respectively). As a second instance, we look at a task that requires us to calculate the singular values of large low-rank data matrices and the singular vectors associated with them. In this regard, we perform the computation of entropies of low-rank density matrices by randomly projecting them to reduced subspace (along the lines of Refs. [26,27]) to get the dominant singular values postprojection. We also demonstrate that one can construct the simplest quantum random projector by performing quantum random projection and extracting the dominant singular values using the variational quantum singular value decomposition (VQSVD) [6]. Here, random projection to a lower dimension allows us to optimize using a lower-dimensional variational *Ansatz* at one end. The combined effect of the variational nature of the algorithm and the fact that unitary t-designs are short depth establishes good testing grounds for the implementation of this demonstration in near-term devices [8]. These demonstrations highlight the ability of local random circuits to efficiently randomize a large Hilbert space (and hence require exponentially fewer parameters to create a random projector) and serve as good random projectors for dimensionality reduction.

II. THEORETICAL BACKGROUND

A. Random projection

The random projection method is a computationally efficient technique for dimensionality reduction and is useful in many problems in data science, signal processing, machine learning, etc. (see, for example, Refs. [10,11]). The reason behind the effectiveness of the method stems from the Johnson-Lindenstrauss lemma [12].

Lemma 1. For any $0 < \epsilon < 1$ and $N \in \mathbb{Z}_+$, let us also consider $k \in \mathbb{Z}_+$ such that

$$k \sim O\left(\frac{\log(N)}{\epsilon^2}\right). \tag{1}$$

Then, for any set of vectors $S = \{\vec{x}_i\}_{i=1}^N$ with $\vec{x}_i \in \mathbb{R}^d$, $\exists\ f : \mathbb{R}^d \to \mathbb{R}^k$ such that $\forall\ \vec{x}_i, \vec{x}_j \in S$,

$$(1-\epsilon)|\vec{x}_1 - \vec{x}_2|_2 \leqslant |f(\vec{x}_1) - f(\vec{x}_2)|_2 \leqslant (1+\epsilon)|\vec{x}_1 - \vec{x}_2|_2, \tag{2}$$

where $||_2$ refers to the l_2 norm.

Proof. See lemma in Ref. [12]. ■

Definition 1: Random projections vs good random projections

Multiplication with Gaussian or Haar random matrices along with a scaling factor followed by projection to a

reduced subspace is one function that obeys Eq. (2) [28,29]. Essentially, it follows from the fact that the expected value of Euclidean distance post–random projection is equal to the Euclidean distance in the original subspace. And the distances post–random projection are not distorted beyond an ϵ factor with high probability because the variance of the distances post–random projection is sufficiently low.

From now on, we consider random projections that satisfy the JL lemma in Eq. (2) to be good random projections. In this regard, the JL lemma says that any set of N points in a high-dimensional Euclidean space (say, $\mathbb{R}^N$) can be embedded into a lower number of dimensions [say, $k = O(\epsilon^{-2}\log N)$] by a random projection, preserving all the pairwise distances to within a multiplicative factor of $1 \pm \epsilon$. This is also equivalent to preserving all the pairwise inner products (or angles). Formally,

$$(1-\epsilon)|\vec{x}_1 - \vec{x}_2|_2 \leqslant |\Pi\cdot\vec{x}_1 - \Pi\cdot\vec{x}_2|_2 \leqslant (1+\epsilon)|\vec{x}_1 - \vec{x}_2|_2, \tag{3}$$

where $||_2$ refers to the l_2 norm and $\vec{x}_i \in \mathbb{R}^N \ \forall\, i$ and Π denote the random projection matrix of size $k \times N$ (or $N \times k$ in which case the random matrix multiplies the data vectors from the right) which obeys Eq. (3) and will be called good random projectors from now onwards.

Several other candidates which satisfy the JL lemma have been considered for random projection in various applications. These random matrices include the SRHT and input sparsity transform (IST) [10,15,30,31]. These random projectors are database friendly because, unlike Gaussian or Haar random matrices whose storage memory cost is proportional to the number of matrix entries and precision, these could be retrieved by matrices that are sparse and have whole number entries.

For benchmarking the quantum random projection in our analysis later, we will be using the SRHT [32] to compare the performances of random projection using random quantum circuits. We picked the SRHT because we want to compare different random matrices that could be efficiently stored and multiplied. In a classical setting, that would be the SRHT, and in a quantum setting, it would be the 2-designs that act as quantum random projectors. We construct a SRHT random projector as in Algorithm 1.

Algorithm 1 Construct a classical random projector (CRP) $\Pi_{CRP}^{N\times k}$ which satisfies JL lemma [see Eq. (3)].

INPUT: integers $N = 2^n$, k with $k \sim \log(N/\epsilon^2)$
1. Assign a diagonal matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$ whose elements are independent random signs $1, -1$.
2. Assign a matrix $H \in \mathbb{R}^{N\times N}$ to be the normalized Walsh-Hadamard matrix.
3. Assign a matrix $S \in \mathbb{R}^{N\times k}$ by randomly sampling k columns from the $N \times N$ identity matrix.
4. The subsampled randomized Hadamard transform matrix (which is our CRP) is obtained as $\Pi_{CRP} = \sqrt{\frac{N}{k}} D \cdot H \cdot S$.

B. Approximate unitary t-designs

In the next section, we will show that the random matrices sampled uniformly from the Haar measure satisfy the JL lemma. Though the exact replication of Haar random unitaries is not possible as a quantum circuit because of the fact that they require exponential resources [16], we will show that to satisfy the JL lemma, exact or approximate t-designs [19], which match the Haar measure only until second moment, would suffice. We will introduce the definitions related to the approximate t-designs in this section and provide theorems on approximate t-designs (or 2-designs) satisfying the JL lemma.

1. Definition 1: Moment operator

The tth moment of a superoperator defined with respect to a probability distribution $\mathcal{U}(N)$ defined on the unitary group $\mathbb{U}(N)$ is defined as

$$\Phi_{\mathcal{U}(N)}^{(t)}(\cdot) = \int_{U\sim\mathcal{U}(N)} U^{\otimes t}(\cdot)(U^\dagger)^{\otimes t} d\nu(U), \tag{4}$$

where $d\nu(U)$ is the volume element of the probability distribution $\mathcal{U}(N)$.

2. Definition 2: Exact unitary t-design

Let us define $\Delta_{\mathcal{U}(N)}^{(t)}(\cdot)$ [31,33,34] as

$$\Delta_{\mathcal{U}(N)}^{(t)}(\cdot) := \int_{U\sim\mathcal{U}} U^{\otimes t}(\cdot)(U^\dagger)^{\otimes t} d\nu(U) - \int_{U'\sim\mu_H} U'^{\otimes t}(\cdot)(U'^\dagger)^{\otimes t} d\nu(U'), \tag{5}$$

where μ_H refers to the uniform distribution over the Haar measure. Unitaries like U sampled from a distribution $\mathcal{U}(N)$ are said to form a t-design if and only if $\Delta^{(t)}(X) = \mathbb{0}\ \forall\, X = f(U) \in \mathbb{U}(N)$. This essentially means that the $\mathcal{U}(N)$ mimics the Haar measure up to the tth moment.

3. Definition 3: α approximate unitary t-designs

The unitary group $\mathbb{U}(N)$ is said to form an α approximate unitary t-design if and only if

$$\left|\left|\Delta_{\mathcal{U}(N)}^{(t)}\right|\right|_\diamond \leqslant \alpha/N^t, \tag{6}$$

where $||\cdot||_\diamond$ refers to the diamond norm (see, for example, Ref. [22]). Though the α approximate unitary design definition here involves the diamond norm, formulations using other norms exist [35], and the theorems in the following section generalize for those formulations as well. Local random quantum circuits with depths $O(\log(N)(\log(N) + \log(1/\alpha)))$ become α approximate 2-designs [22].

III. RANDOM QUANTUM CIRCUITS AS RANDOM PROJECTORS

In this section, we show that local random quantum circuits which are approximate unitary 2-designs (or exact unitary 2-designs) are suitable candidates for the random projection (and will be called quantum projectors from now on). We show that quantum projectors satisfy the Johnson-Lindenstrauss lemma so that their random projection is an l_2 subspace embedding with a very high probability of having a

very low error. And if one were to compute specific quantities like entropy, one should quantify whether such random matrices produce projected singular values that are closer to the true singular values with higher probability. In a later section, we will discuss how the projection can be done on real quantum computers and how the reduced dimensional vectors and their singular values can be read out from near-term quantum computers. In the following theorems, let us denote the Haar measure distribution as μ_H and the distribution corresponding to α approximate $t = 2$ design as $\mu_{2,\alpha}$. Proofs of the theorems can be found in Appendix A.

Theorem 1. Let $U \in \mathbb{U}(N)$ be sampled uniformly from the Haar measure (μ_H) and let $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^N$. Then the matrix $\Pi^{k\times N}$ obtained by considering any k rows of U followed by multiplication with $\sqrt{\frac{N}{k}}$ satisfies

$$(1-\epsilon)|\vec{x}_1 - \vec{x}_2|_2 \leqslant |\Pi \cdot (\vec{x}_1 - \vec{x}_2)|_2 \leqslant (1+\epsilon)|\vec{x}_1 - \vec{x}_2|_2 \quad (7)$$

with probability greater than $(1 - \frac{N-k}{4kN\epsilon^2})$ with $\epsilon \in R_{\geqslant 0}$ being the error threshold.

Proof. See Appendix A. ■

Theorem 2. Let $U \in \mathbb{U}(N)$ be sampled uniformly from the α approximate unitary 2-design $(\mu_{2,\alpha})$ and let $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^N$. Then the matrix $\Pi^{k\times N}$ obtained by considering any k rows of U followed by multiplication with $\sqrt{\frac{N}{k}}$ satisfies

$$(1-\epsilon)|\vec{x}_1 - \vec{x}_2|_2 \leqslant |\Pi \cdot (\vec{x}_1 - \vec{x}_2)|_2 \leqslant (1+\epsilon)|\vec{x}_1 - \vec{x}_2|_2 \quad (8)$$

with a probability greater than $1 - (\frac{N-k}{4kN\epsilon^2} + \frac{\alpha}{4\epsilon^2 k})$ with $\epsilon \in R_{\geqslant 0}$ being the error threshold.

Proof. See Appendix A. ■

It is worth mentioning that upper bounds on the distortion have been obtained before with exponential scaling [23], namely, $2^4 \exp(-2^{-4}\epsilon k)$ for Haar measure and $2^{10}\exp(-2^{-10}\epsilon^2 k)$ for approximate t-designs. These limits are better than the limits obtained here only for $(N, k) > 10^4$. For the cases that are to be explored in the paper, our limits are tighter than the exponential limits.

For the plots in the experiments section of the paper, we use the *Ansatz* used in Ref. [36] which is assumed to be an exact 2-design *Ansatz* beyond a certain depth (Fig. 1). The main text contains the depths at which the *Ansatz* matches the exact 2-design limit (∼150). There are many candidate local random circuit architectures which are α approximate 2-designs [22]. Instead of studying the projection abilities of different local random circuits architecture, Appendix E contains some experiments where we look at a less expensive *Ansatz* and hence is in an approximate unitary 2-design regime by choosing a lower depth (∼50) (analogous to Ref. [34]) of the same circuit in Fig. 1.

IV. EXPERIMENTS ON QUANTUM RANDOM PROJECTORS

In this section, we consider two different experiments to benchmark the performance of quantum random projection discussed in the previous section against the SRHT projection which will be labeled as classical random projection in the plots. This should be looked at as a comparison of random projectors that can be stored and applied efficiently in terms of memory and time complexity in classical vs quantum settings. Since quantum random projectors approximate the Haar measure, their projection abilities are expected to be better than the SRHT projectors because the latter is less random compared to the Haar measure. However, in certain applications, it is known that they both converge to similar performance when the size of the data set tends to infinity [29]. We see in Appendix D that their performances start becoming closer when we increase the size of the data matrices, and vectors from 1024 to 2048 (corresponding to 10 and 11 qubits, respectively).

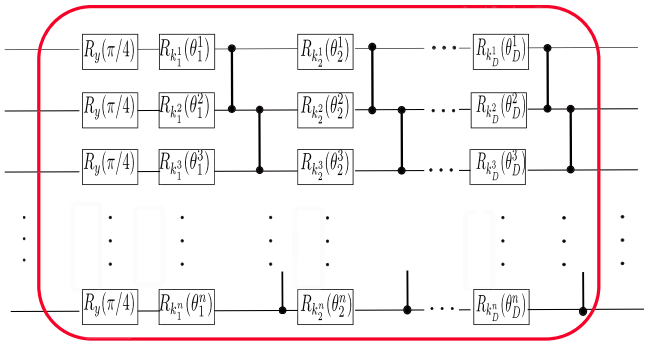

FIG. 1. The local random quantum circuit used in preparing a quantum random projector is the *Ansatz* that was used in Ref. [36] and is known to converge to an exact 2-design limit of the variance of the local cost function beyond a certain depth. The circuit contains a layer of $R_y(\pi/4)$ rotations (often used to make all the directions symmetric in a variational training procedure; we do not necessarily need to have this component). This is followed by alternating random single-qubit rotations and ladders of CPHASE operations repeated D times. For $n = 10, 11$ (the dimensions studied in this paper), the circuit reaches the exact 2-design limit (variance limit) at $D \geqslant 150$.

We initially consider the task that does not require us to know the singular values and is concerned with only dimensionality reduction. In this regard, we reduce the dimensions of the MNIST [24] and CIFAR-100 [25] image data sets and benchmark the performance of quantum random projection against classical random projection. We also compare it with the computationally expensive PCA, which is supposed to give the exact projection to the dominant singular vectors of the data sets and cannot be outperformed beyond a certain rank.

In the second task, we calculate the von Neumann entropy of low-rank density matrices (along the lines of Ref. [26]) which requires us to know the singular values after random projection in addition to the dimensionality reduction. We compare the performance of quantum random projection (QRP) vs classical random projection (CRP) for this task over different ranks (r) of the density matrices.

In this section, we pick the local random quantum circuit from Fig. 1 and we assume that we can make arbitrary projection operators with any number of basis vectors, i.e., if the P_k operator projects to the first k basis state $(|p_1\rangle, |p_2\rangle, \ldots, |p_k\rangle)$ in any basis. Then,

$$P_k = \sum_{i=1}^{k} |p_i\rangle\langle p_i|, \quad (9)$$

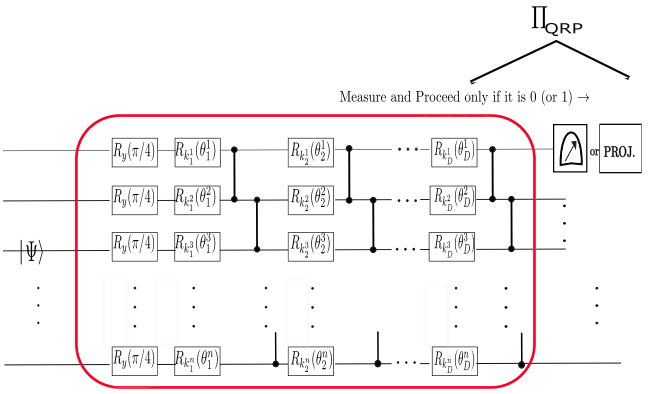

FIG. 2. The schematic of performing the quantum random projection. The data vector has to be encoded into the circuit through one of the existing encoding schemes (see main text). This is followed by the local random quantum circuit and partial measurements (the number of qubits measured depends on how low the final reduced dimensions are) or an arbitrary projection operator. For partial measurements, the algorithm proceeds only if the measurement results in qubits in only 0 (or only 1). This is equivalent to reducing the data set's size by 1/2, 1/4, 1/8, and so on, depending on how many qubits are measured.

where we do not have any restriction on what basis we pick and what values k can take. In a later section, we discuss the simplest projection operator one can construct by measuring one or more qubits and restricting to particular outputs (0 or 1) in those qubits, as shown in Fig. 2. It is worth noting that this scheme has a structure similar to the quantum autoencoders [37] but the circuit here is data agnostic.

A. Dimensionality reduction of image data sets

In this section, we benchmark the performance of QRP against CRP in the task of dimension reduction of subsets of two different image data sets, MNIST and CIFAR-100. We also plot the performance of the computationally expensive PCA which is supposed to capture all the nonzero singular valued singular vectors. When the reduced dimension is greater than the rank of the system, PCA could never be outperformed.

MNIST contains 28×28 grayscale images. The matrix representations of the images were boosted to 32×32 so that they can be reshaped into 1024×1 normalized vectors by adding zeros. (Note that this is not a common quantum encoding scheme. We use QRP on the normalized data vectors for a direct comparison with CRP.) We have to do this preprocessing step because the projectors that we consider (both CRP and QRP) are of the form $2^n \times k$ and hence take only 2^n-dimensional vectors as the input.

CIFAR-100, in addition to being composed of 28×28 images, also contains colored images and had to be converted to 32×32 grayscale so that they can be fed as input to our projectors. But unlike MNIST, which contains handwritten integers from 0 to 9, the CIFAR-100 data set contains images belonging to 100 different classes including airplanes, automobiles, birds, cats, trucks, etc. As a result, CIFAR-100 is expected to have more features in its data sets and hence

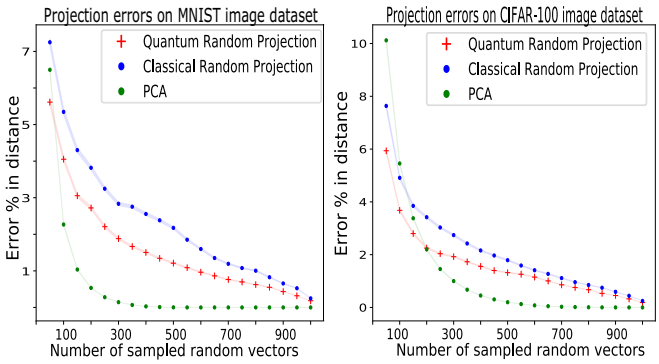

FIG. 3. The mean percentage errors in the distance between 10 000 different random pairs of data vectors in the MNIST and CIFAR-100 data sets. The envelopes represent their 95% confidence intervals. We see that PCA outperforms the random projection methods beyond a certain rank. Among the random projection methods, though there is not much difference between the classical random projection (CRP) and quantum random projection (QRP), we observe that the latter performs slightly better.

greater rank compared to MNIST if we consider subsets from each of these data sets.

To perform the comparison between CRP and QRP, we took 1000 images from each of these data sets. And in each of these subsets, we reshaped the images into 1024×1 normalized vectors and performed random projection to lower dimensions (x axis of Fig. 3). Then, we randomly sampled two data vectors and compared the error percentage in their l_2 norm (Euclidean distance) between them in the original space and the reduced dimensional space obtained after random projection. This procedure is repeated 10 000 times and the mean error percentages and their 95% confidence intervals for different reduced dimensions are reported in the plots in Fig. 3.

The random projections are performed by multiplying the vectors with random matrices (see Algorithms 1 and 2)

Algorithm 2 Constructing quantum random projector (QRP) $\Pi_{QRP}^{N\times k}$ which satisfies the JL lemma (Theorems 1 and 2).

INPUT: integers $N = 2^n$, k with $k \sim \log(N/\epsilon^2)$
1. Choose $\alpha \sim O(\epsilon^2 k)$.
2. Construct a local random quantum circuit with depth $\sim O(n(n + \ln(1/\alpha)))$ which is an α approximate 2-design or pick an exact unitary 2-design *Ansatz*.
3. Append a projection operator P_k acting after the local random quantum circuit to form $\Pi^{k\times N}$.

$$\vec{\tilde{x}} = \Pi_{CRP} \cdot \vec{x}, \tag{10}$$

$$|\tilde{x}\rangle = P_k \cdot U_{QRP} \cdot |x\rangle, \tag{11}$$

where Π_{CRP} is the SRHT projector and U_{QRP}, P_k are the sampled from the matrix representation of the local random quantum circuit and projectors used. And PCA projection is obtained by first computing singular value decomposition on the data set and projecting them to the subspace of dominant singular vectors:

$$\vec{\tilde{x}} = \Pi_{PCA} \cdot \vec{x}. \tag{12}$$

Figure 3 shows that the PCA outperforms the random projection methods beyond a certain rank. This is because, beyond the rank of the data set considered, PCA projects exactly to the subspace with nonzero singular values. Despite that, we see that random projection methods which are not computationally extensive (because they do not compute the subspace with nonzero singular values) perform to the same extent and even better than PCA at lower reduced dimensions. This dominance in performance at lower reduced dimensions is visible in larger-dimensional data sets (see Appendix D). We also see that PCA takes more reduced dimensional vectors to catch up with the random projection algorithms in the case of CIFAR-100 because it has comparatively more rank (loosely because it has more features) than the MNIST data set. These data vectors dimensionally reduced via quantum random projection could be used in quantum machine learning applications such as training an image recognition or classification model (see, for example, Ref. [38]).

Within the random projection methods, quantum random projection performs slightly better than the classical random projection mainly because Haar random matrices are more random and have tighter JL lemma bounds than the classical random projector. The performance of quantum random projectors which are away from the exact 2-design limit has been analyzed in Appendix E by looking at shorter depths ($\sim$50) of Fig. 1 and hence a lesser expressive *Ansatz*.

The discussion in this section assumed the existence of an exact amplitude encoding scheme for the data vectors. This would require impractical depths of $O(2^n)$ unless the data vectors are genuinely quantum, e.g., ground states of a family of local Hamiltonians. However, for general data vectors like image data vectors, we do not necessarily need exact encoding. Preserving the distinctness of image data vectors $(\vec{x}, \vec{y})$ to a good enough accuracy enables us to use them for many image processing applications, such as recognition and classification. In this regard, there has been substantial work on approximate amplitude encoding. These schemes encompass approximately encoding data vectors whose amplitudes are all positive [39], real [40], and even complex data vectors [41] using shallow parametrized quantum circuits.

With the plots in Fig. 3, we showed that for exactly encoded data vectors $(\vec{x}, \vec{y})$ and quantum random projected vectors $(\vec{\tilde{x}}, \vec{\tilde{y}})$

$$||\vec{x} - \vec{y}| - |\vec{\tilde{x}} - \vec{\tilde{y}}|| \leqslant \delta \tag{13}$$

on average for pairs of images in the data set used. Here δ is a very small fraction of $|\vec{x} - \vec{y}|$. A good approximate amplitude encoding scheme is bound to preserve this distance with minimal error, since it preserves the distinctness of the samples as well as (calling $\vec{x}_{\text{app}}, \vec{y}_{\text{app}}$ approximate encoded vectors)

$$||\vec{x} - \vec{y}| - |\vec{x}_{\text{app}} - \vec{y}_{\text{app}}|| \leqslant \Delta \tag{14}$$

on average. Here Δ is a small fraction of $|\vec{x} - \vec{y}|$.

With Eqs. (13) and (14), it is clear that, even with the approximate amplitude encoding, quantum random projection would preserve the distinctness of samples (up to a perturbation of $\delta + \Delta$) and be useful for image processing applications. The exact value of Δ depends on the efficiency of the approximate encoding used.

The other alternative to circumvent the impractical depths of the exact data encoding issue is by adopting different encoding schemes. One can start by reducing the resolution of the images (equivalent to reducing the pixels), which results in reduced classical image data vector dimension (to, say, $m < 2^n$), and using any other existing data encoding schemes that use qubits greater than m but with polynomial depths (for example, Refs. [42,43]).

If $\Phi(\cdot)$ is the encoding function that takes the original data vector and encodes it as a data vector of dimension 2^d, then, to check how well the distinctness is preserved, experiments need to be run on the d qubits with a quantum random projector corresponding to d qubits. Mathematically, we need to check how low the following values are (on average) for two data vectors $\vec{x}, \vec{y}$ from the original data set:

$$||\Phi(\vec{x}) - \Phi(\vec{y})| - |\Phi(\vec{x})_r - \Phi(\vec{y})_r||, \tag{15}$$

where $\Phi(\vec{x})_r$, $\Phi(\vec{y})_r$ are reduced randomly projected encoded vectors.

In this work, we confined ourselves to experiments involving an exact encoding scheme despite impractical depths because the preservation of distance for the exact encoding scheme implies the same for the approximate encoding schemes as described earlier. Checking the preservation of distance for other encoding schemes would require knowing the exact form of encoding in Eq. (15).

Just like its classical counterpart, we can also reconstruct the images back to the original size after the random projection. For classical methods (for a data vector $\vec{x}$ and its reduced data vector $\vec{\tilde{x}}$),

$$\vec{x}_{\text{recons}} = \Pi_{PCA}^T \cdot \vec{\tilde{x}}, \tag{16}$$

$$\vec{x}_{\text{recons}} = \Pi_{CRP}^T \cdot \vec{\tilde{x}}. \tag{17}$$

For the quantum case, we need to put in the extra qubits or the subspace to which we projected to get back to the original size and then apply the inverse of the unitary circuit used for projection. For example, if we had projected to the subspace where one of the qubits is in the $|0\rangle$ state, we boost the size back to the original size by having a new qubit at $|0\rangle$ and add the inverse unitary circuit $(U_{QRP}^\dagger)$ on this new system. For a general projector, $|\tilde{x}\rangle \to |\tilde{x}\rangle \otimes |\tilde{p}\rangle$ where tensor product with the $|\tilde{p}\rangle$ ensures that we get back the original size of the data (image). And the reconstruction is done as follows (for the data set $|x\rangle$):

$$|x\rangle_{\text{recons}} = U_{QRP}^\dagger \cdot |\tilde{x}\rangle. \tag{18}$$

This is similar to the reconstruction done in Ref. [37]. These reconstructions work on the premise that the product $\Pi^\dagger \Pi \sim \mathbb{I}$ of the original data dimension. It is trivial to see that this holds for the PCA projector. It turns out that this also holds for the random projectors. This is because, in a larger dimensional space, finding almost orthogonal vectors becomes more common; this was studied in Ref. [44] and was used in the discussion of Ref. [11]. Figure 4 shows how one would reconstruct an image from the MNIST data set after dimensionally reducing a subset of the MNIST images.

With the reconstructed quantum data vectors, there are processing applications which have computational advantage over classical processing. For example, the complexity of the

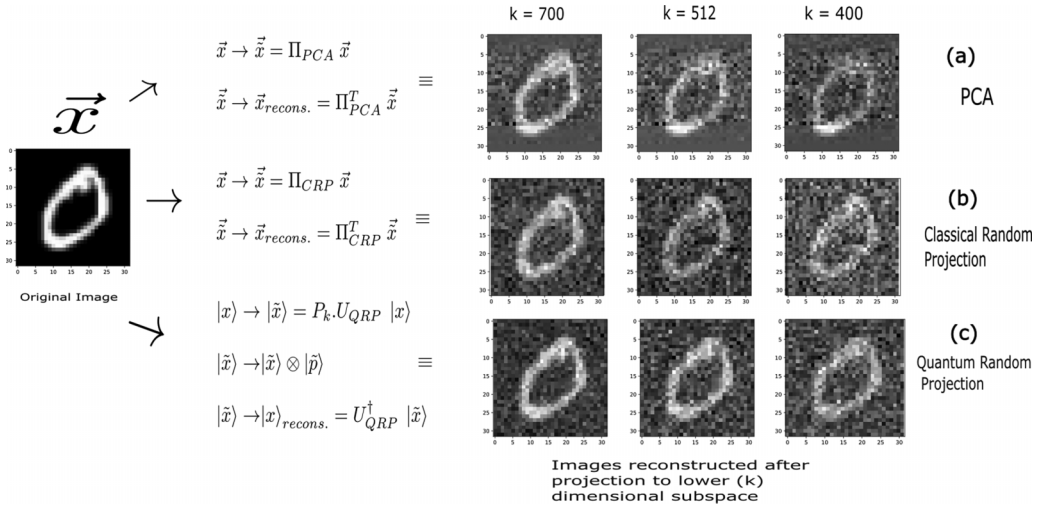

FIG. 4. The schematic of steps involved in the dimensionality reduction and image reconstruction of image data sets using (a) PCA, (b) CRP, and (c) QRP. The figure shows the reconstructed images for various reduced dimensions. Though projectors with dimensions 700 and 400 are not straightforward to construct, a reduced dimension of 512 represents projection by measuring one of the qubits (so the size drops from 1024 to 512) and processing further only if it is 0 or 1. The figure illustrates the reconstruction of one of the data vectors from the MNIST data subset with which we are experimenting; a quantitative description of how the reconstruction performs compared to classical methods is discussed in Appendix B along with a description of the construction of projection operators.

quantum edge detection algorithm [45] is polynomial and does not require exponential resources if we have an image encoded either exactly or approximately. The measurement outputs of the edge detection algorithm contain information about the edges. To get the outputs of this experiment, one can also adopt the classical shadows [46,47] approach to get the probabilities of all the bit strings with less number of measurements than full-state tomography.

B. Entropy estimation of low-rank density matrices

In this section, we compare the performance of the quantum random projectors against the classical random projector, SRHT, on a task that requires one to obtain the approximate singular values of the dominant singular vectors of a data matrix after the dimensionality reduction. Unlike the previous task, this task concerns reducing the dimensions of a large data matrix instead of individual data vectors by random projection. After the dimensionality reduction, we check how well the system captures the properties of the data set by computing the error percentage in a particular property of the data matrix which requires the knowledge of all its singular values.

Specifically, we will consider randomly generated semipositive-definite density matrices with random singular vectors but their singular values follow a certain profile. We then compute their entropy after quantum random projection and check their accuracy (along the lines of Ref. [26]). The exact singular values profile of these density matrices depends on the nature of the system. In this experiment, we consider singular values which are linearly decaying and exponentially decaying until the rank of the system and zero afterward. These profiles could be motivated through the existence of physical systems with such profiles. A thermal ensemble of a simple harmonic oscillator mode of frequency ν with N internal degrees of freedom has an exponentially decaying profile. Here, singular values of ρ will be proportional to 1, $e^{-h\nu}$, $e^{-2h\nu}$, $e^{-3h\nu}$, and so on. And, it is known that a maximal second-order Rényi entropy ensemble of a system with a simple harmonic oscillator mode of frequency ν and N internal degrees of freedom follows a linearly decaying singular value profile for its density matrix [48]. The main text contains the plots related to the linearly decaying singular profile and Figure in Appendix C contains the plots related to the exponential decay profile.

Here is the procedure to perform random projection given a semipositive-definite matrix M of dimension $N \times N$:

(i) Project the original density matrix of size $N \times N$ to a lower dimension $N \times k$ using Π_{CRP} and Π_{QRP}.

(ii) Perform SVD (classical) or quantum SVD (QSVD) on the lower-dimensional matrix to get the singular vectors with singular values $\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \ldots, \tilde{p}_k$ which are approximations to $p_1, p_2, \ldots, p_k$.

(iii) Then we obtain an approximation to entropy (S) using $\tilde{S} = \sum \tilde{p}_i \ln \frac{1}{\tilde{p}_i}$.

The accuracy in the approximated entropy is bounded in Theorem 3.

Theorem 3. For a random matrix Π satisfying the Johnson-Lindenstrauss (JL) lemma with a distortion $\epsilon\sqrt{\delta}$, where $\epsilon \leqslant 1/6$ and $\delta \leqslant 1/2$, the difference in the von Neumann entropy of a density matrix ρ computed using the random projection with Π (denoted as $\tilde{S}(\rho)$) and the true entropy ($S(\rho)$) can be bounded as follows:

$$|\tilde{S}(\rho) - S(\rho)| \leqslant \sqrt{3\epsilon} S(\rho) + \sqrt{\frac{9}{2}}\epsilon \tag{19}$$

with probability at least $(1 - \delta)$.

Proof. See Appendix A. ■

Figure 5 shows the error percentage in the computed entropy after random projection for density matrices of size 1024×1024 with linearly decaying singular values until a certain rank ($r = 10, 50, 100, 40$ in the plot) and zero

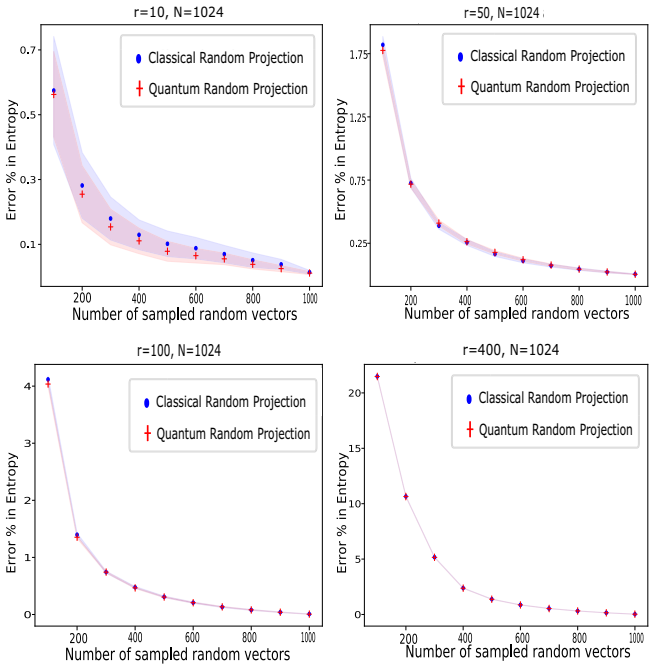

FIG. 5. The plots in this figure show the accuracies of quantum random projection and classical random projection in the entropy computation of randomly generated density matrices of size $N = 1024$ and ranks $r = 10, 50, 100, 400$ with a linearly decaying singular value profile. The envelopes represent their 90% confidence intervals by running the experiments over 100 randomly generated density matrices. The accuracies improve with decrease in the ranks as expected.

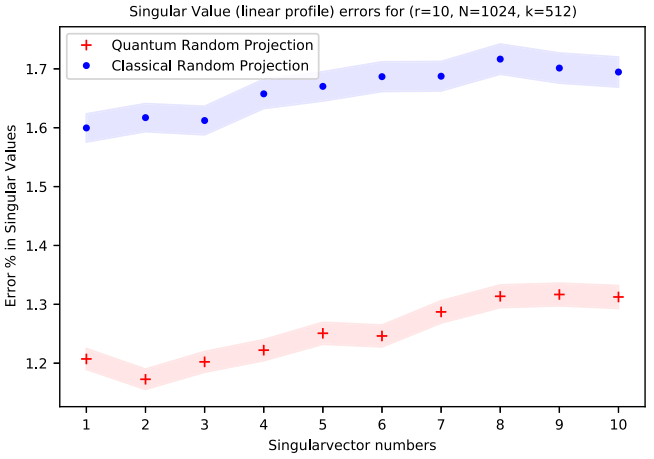

FIG. 6. The plot shows the accuracy with which the quantum random projector and the classical random projector pick the singular values of the density matrix for $r = 10$ when reducing the system size by half. The envelope represents 95% confidence intervals by running the experiments over 10 000 randomly generated density matrices.

afterward. The x axis represents the different reduced dimensions (k). The accuracies are better for low ranks as expected and get worse for larger ranks. We observe that the quantum random projector and the classical random projector perform to similar extents (if not a better quantum performance than classical performance for density matrices with very low rank) in this task. This matches the trends reported in Ref. [26] of similar performance for various other classical random projection matrices like Gaussian, SRHT, and IST. We also show the accuracies with which the quantum and classical random projectors capture the singular values of the system for the rank $r = 10$ when the system's size has been reduced by half in Fig. 6. Here, we see that the quantum random projectors perform better than their classical counterpart mainly because Haar random matrices that the random circuits try to approximate are more random than any classical random projectors that could be stored with similar or comparable complexity. The Appendix contains a discussion regarding how the accuracy improves when we increase the size of the original data sets from 1024×1024 to 2048×2048.

We discuss the same plots for density matrices with exponentially decaying singular value profile until a certain rank in Appendix C. We observe there that increasing the rank does not change the singular value profile as much and hence the accuracy with which the random projection algorithms work remains pretty much constant. The Appendix E contains the error plots for the accuracy in individual singular values for the case $r = 10$ obtained using lesser expressive random circuits (depth $\sim$50).

V. HOW TO PROJECT IN A REAL QUANTUM COMPUTER

For the quantum random projection to work, in addition to sampling a unitary from the exact (or approximate 2-designs), we also need to have a circuit component for projector operators. In one of the previous sections, we considered arbitrary projection operators which might not be able to be efficiently implemented in a quantum computer with polynomial resources. However, we can look at the simplest projection operations that one can use for the quantum random projection. In Fig. 2, we looked at the simplest projection operation, which is measuring some of the qubits and proceeding only if the qubits are in a certain state ($|0\rangle$ or $|1\rangle$). This is equivalent to projecting it to the subspace where those qubits take that specific value. For example, when you have a circuit of ten qubits, measuring one of the qubits and proceeding only when that qubit is in $|0\rangle$ means a reduction in the data vector (or ket) dimension from 1024 to 512.

However, the projection through measurement discussed above is different compared to the classical projection because a quantum measurement (wave-function collapse) automatically takes care of the normalization factor and the extra $\sqrt{\frac{N}{k}}$ is not needed. Since the Hilbert space we consider here is large and the ket entries are randomized, the normalization that happens because of the wave function is the same as the prefactor we would get in a classical random projection.

To demonstrate the quantum random projection with a simple projection operation, we will consider projection operators of the form $\frac{1}{2}(1+\sigma_z^i)$ which is a projection operator to the space where the ith qubit is at the $|0\rangle$ state. To demonstrate this, we perform such a quantum random projection for a large data matrix with a linearly decaying singular value profile of size 1024×1024 and rank $r = 5$ by reducing the

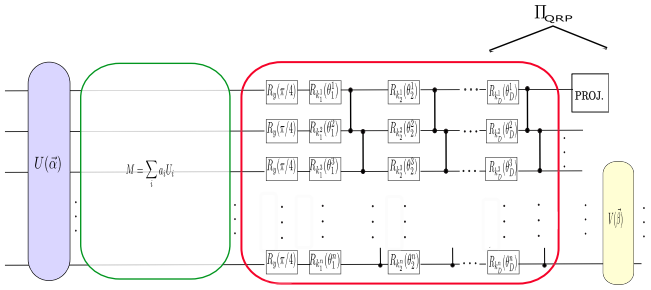

FIG. 7. The figure shows the schematic of the variational quantum SVD post–quantum random projection to lower dimensions. The data matrix M needs to be loaded using a set of unitary gates with techniques like importance sampling (see related discussion in the Appendix of Ref. [6]). Similar to the setup in Fig. 2, we perform projection by measuring a few qubits at the top. This is followed by a training procedure to obtain the dominant singular vectors and their singular values. The singular vectors on the right end belong to the lower-dimensional space and hence require a lower-dimensional *Ansatz*.

data vectors to sizes 512, 256, and 128 by projecting out one, two, and three qubits, respectively. Then, we retrieve the dominant singular vectors by performing a VQSVD [6]. But since the data matrix has been dimensionally reduced, the *Ansatz* we use for finding the right singular vectors is also of reduced size (Fig. 7). The details regarding the implementation of VQSVD, and the *Ansatz* type used, can be found in Appendix F.

Figure 8 shows the accuracy with which we were able to retrieve the singular values after quantum random projection for individual singular vectors. This demonstrates how one can perform quantum random projection in near-term devices as the VQSVD algorithm used to retrieve the dominant vectors is a near-term algorithm. The accuracy with which the singular values have been retrieved depends on the expressivity of the *Ansatz* and whether or not it falls into a barren plateau

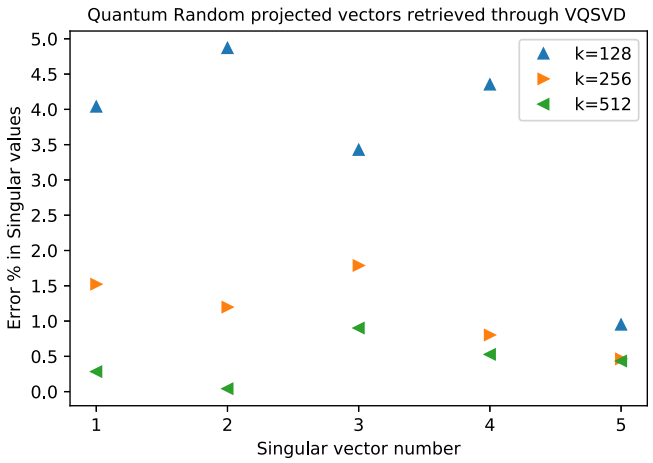

FIG. 8. The figure shows the errors in the singular values obtained by reconstructing the dominant singular vectors post–quantum random projecting a randomly generated data matrix of rank $r = 5$ using variational quantum SVD for various reduced dimensions $k = 512, 256, 128$.

during the training procedure. We have not discussed the most accurate retrieval of the singular vectors, as that is beyond the scope of this paper. There are many strategies to avoid falling into the barren plateau and improve the convergence rate [49–51]. We used the identity block strategy [51] to avoid barren plateaus (more details on that are in Appendix F).

VI. CONCLUSION

In this work, we explored a practically useful application of local random quantum circuits in the task of dimensional reduction of large low-rank data sets. The main essence of the applicability of the local random circuits in this task is their ability to anticoncentrate rapidly at linear or sublinear depths [52,53]. This makes them a good random projector to lower dimensions, meaning they preserve the distinctness of different dominant data vectors in a large data set after dimensional reduction.

The theorems discussed in the paper show that just like the Haar random matrices which are good random projectors, their approximate quantum implementations, the exact and approximate t-designs are also good random projectors. The rapid anticoncentration of Hilbert space at linear depths means that the number of random parameters (the random rotation parameters) required to create and reproduce a random projector is logarithmic in the size of the data sets. Such efficiency in the storage complexity of classically generated Haar random matrices or in any classical random projector is not possible. We then benchmarked its performance against the commonly used classical random projector, SRHT. The quantum random projectors performed slightly better than this classical candidate because they are trying to approximate Haar random matrices which are more random than the classical candidate. We then demonstrated these comparisons for various tasks such as image compression, reconstruction, retrieving the singular values of the dominant singular vectors post–dimension reduction, etc.

Though the initial discussion assumed arbitrary projection operators to arbitrary subspaces, we showed the simplest projection operators and projection subspaces exist. We demonstrated this simplest quantum random projection and retrieved the dominant singular vectors post–quantum random projection via VQSVD [6]. This shows the applicability of such quantum random projections and their retrievals in near-term devices.

Dimensionality reduction facilitated by random projections as discussed in this work can also precede kernel-based variants of PCA wherein eigenvalue decomposition of the Gram matrix associated with the higher-dimensional embedding (often called kernel) is sought [54], especially if said Gram matrix is low rank. Beyond the precincts of classical data, such a technique can act as an effective precursor to improve the efficiency of simulation even on quantum data as has been studied in recent work [55]. The essential crux of the idea is heavily rooted in PCA but applied to quantum data wherein repeated Schmidt decomposition of the states and vectorized form of arbitrary operators are performed followed by subsequent removal of singular vectors associated with nondominant singular values akin to PCA. The techniques explored in this work involving good random projections can

be used in conjunction prior to the application of such a protocol to contract the effective space of the states and/or operators involved. Owing to the demonstrated near-term applicability, similar reduction can also be afforded as a pre-processing step in a host of quantum algorithms manipulating quantum data [56] on noisy hardware. Such protocols are of active interest to the scientific community due to their profound physicochemical applications ranging from exotic condensed-matter physical systems like Rydberg excitonic arrays [57], modeling higher-dimensional spin-graphical architectures in quantum gravity [58], and in learning theory of neural networks [59], constructing unknown Hamiltonians through time-series analysis [60,61], tomographic estimation of quantum states [62,63], in the electronic structure of molecules and periodic materials [64], quantum preparation of low energy states of desired symmetry [64,65], or even order-disorder transitions in conventional Ising spin glass using quantum annealers [66] and quantum variants of the Sherrington-Kirkpatrick model [67], to name a few.

We did an extensive comparison using a deep (∼150) exact 2-design *Ansatz* and deferred the discussion about circuits away from the exact 2-design limit to Appendix E. This is because there exist various random circuit architectures which anticoncentrate just like the exact 2-design *Ansatz* and hence could be good candidates for random projection. This could be a good starting point for future study. Also, it is worth studying and constructing quantum random projectors suited for specific applications and data sets (for example, the data sets in health care [68,69]). It has to be noted that the results derived in the main text assumed noiseless quantum gates and measurements. Similar theorems need to be understood for real quantum computers where different noise sources are unavoidable. This leads to a possible future study to understand the extent to which the theorems in the main text are valid on real quantum computers by performing statistical analysis on the bitstrings from the output of real quantum computers (see Refs. [70,71]).

The classical and the quantum random projection matrix (and the rotation parameters used to generate the circuit) used for the comparisons, the data matrix used to generate Fig. 8, along with the code for generating the plots in this paper will be made available upon reasonable request. The simulation for the retrieval of dominant singular vectors through VQSVD was done in the Paddle quantum framework [72].

ACKNOWLEDGMENTS

We would like to acknowledge funding from the Office of Science through the Quantum Science Center (QSC), a National Quantum Information Science Research Center, the U.S. Department of Energy (DOE) (Office of Basic Energy Sciences), under Award No. DE-SC0019215, and the National Science Foundation under Award No. 1955907.

APPENDIX A: PROOFS OF THEOREMS IN THE MAIN TEXT

Theorem 4 (Theorem 1 in the main text). Let $U \in \mathbb{U}(N)$ be sampled uniformly from the Haar measure (μ_H) and let $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^N$. Then the matrix $\Pi^{k\times N}$ obtained by considering any k rows of U followed by multiplication with $\sqrt{\frac{N}{k}}$ satisfies

$$(1-\epsilon)|\vec{x}_1-\vec{x}_2|_2 \leqslant |\Pi\cdot(\vec{x}_1-\vec{x}_2)|_2 \leqslant (1+\epsilon)|\vec{x}_1-\vec{x}_2|_2 \tag{A1}$$

with probability greater than $(1-\frac{N-k}{4kN\epsilon^2})$ with $\epsilon \in R_{\geqslant 0}$ being the error threshold.

Proof. Let P_k be the projector operator to any k basis states of the Hilbert space. Without loss of generality, let us consider $\vec{x}_1-\vec{x}_2=\vec{v}$ to be of unit norm. The components of such $\vec{x}_1-\vec{x}_2$ will be $v_1, v_2, v_3, \ldots, v_N$ such that $\Sigma_i|v_i|^2=1$. Then,

$$|\Pi\cdot\vec{v}|_2^2 = \frac{N}{k}\sum_{i=1}^{k}|y_i|^2, \tag{A2}$$

where $y_1, y_2, \ldots, y_k$ are k components of the vector $U\cdot(\vec{x}_1-\vec{x}_2)$ depending on the P_k's projection subspace. Then,

$$E_{U\sim\mu_H}[|y_i|^2] = E_{U\sim\mu_H}[|y_j|^2] \tag{A3}$$

$\forall i, j$ because we know that the Haar measure satisfies the translational invariance

$$\int_{U\sim\mu_H} f(U)d\mu(U) = \int_{U\sim\mu_H} f(VU)d\mu(U) \tag{A4}$$

when V is an element belonging to the $U(N)$ group. By choosing V to be the matrix that switches ith and jth components of an N-dimensional complex vector, we get the required relation, Eq. (A3). Now using the fact that $\vec{v}$ is unit norm, we get

$$E_{U\sim\mu_H}\left[|\Pi\cdot\vec{v}|_2^2\right] = 1. \tag{A5}$$

Now, to show that sampling U from the Haar measure produces a low distortion (ϵ) with a very high probability, we should look at the variance of the projected norm. We should essentially show that $P_{U\sim\mu_H}((1-2\epsilon)| \leqslant |\Pi\cdot\vec{v}|_2^2 \leqslant (1+2\epsilon))$ with a very high probability. Consider variance Var_{μ_H}

$$\mathrm{Var}_{\mu_H} = E_{U\sim\mu_H}\left[\left(\sum_{i=1}^{k}|y_i|^2\right)^2\right] - E^2_{U\sim\mu_H}\left(\sum_{i=1}^{k}|y_i|^2\right) \tag{A6}$$

$$= E_{U\sim\mu_H}\left[\sum_{i=1}^{k}|y_i|^4\right] + E_{U\sim\mu_H}\left[\sum_{i\neq j}^{k}|y_i|^2|y_j|^2\right] - \frac{k^2}{N^2}. \tag{A7}$$

We know that the Haar measure and quantum circuits which are 2-designs (with $O(\log(N)$ depth) anticoncentrate [52,53]. Specifically for Haar measure at large N, we have

$$\sum_{i}^{N} E_{U\sim\mu_H}[|y_i|^4] = \frac{2}{N}. \tag{A8}$$

Using $(\sum_{i=1}^{N}(|y_i|^2))^2=1$, we get

$$E_{U\sim\mu_H}\left[\sum_{i\neq j}^{N}(|y_i|^2|y_j|^2)\right] = \frac{N-2}{N}. \tag{A9}$$

Then, similar to our previous arguments, we can show that

$$E_{U\sim\mu_H}[(|y_i|^4)] = E_{U\sim\mu_H}[(|y_j|^4)] \tag{A10}$$

$\forall\, i, j$ and

$$E_{U\sim\mu_H}[(|y_i|^2|y_j|^2)] = E_{U\sim\mu_H}[(|y_k|^2|y_l|^2)] \tag{A11}$$

$\forall(i, j)$ and k, l. Equations (A10) and (A11) can be proved by using V in Eq. (A4) to be an $i \leftrightarrow j$ basis switch and $(i, j) \leftrightarrow (k, l)$ basis switch. Equations (A10) and (A11) allow us to write Eq. (A7) as

$$\mathrm{Var}_{U\sim\mu_H} = \frac{2k}{N^2} + \frac{N-2}{N}\frac{k^2-k}{N^2-N} - \frac{k^2}{N^2} = \frac{(N-k)k}{N^2(N-1)}. \tag{A12}$$

The variance for the $|\Pi\cdot\vec{v}|_2^2$ would be $\frac{N-k}{k(N-1)}$ which scales as $O(\frac{N-k}{Nk})$ for large N, k. Using the derived variance and the deviation from the mean to be 2ϵ gives (using Chebyshev's inequality)

$$P_{U\sim\mu_H}[||\Pi\cdot\vec{v}|_2 - 1| \leqslant \epsilon] \geqslant \left(1 - \frac{N-k}{4kN\epsilon^2}\right). \tag{A13}$$

■

Theorem 5 (Theorem 2 in main text). Let $U \in \mathbb{U}(N)$ be sampled uniformly from the α approximate unitary 2-design $(\mu_{2,\alpha})$ and let $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^N$. Then the matrix $\Pi^{k\times N}$ obtained by considering any k rows of U followed by multiplication with $\sqrt{\frac{N}{k}}$ satisfies

$$(1-\epsilon)|\vec{x}_1 - \vec{x}_2|_2 \leqslant |\Pi\cdot(\vec{x}_1 - \vec{x}_2)|_2 \leqslant (1+\epsilon)|\vec{x}_1 - \vec{x}_2|_2 \tag{A14}$$

with a probability greater than $1 - (\frac{N-k}{4kN\epsilon^2} + \frac{\alpha}{4\epsilon^2 k})$ with $\epsilon \in R_{\geqslant 0}$ being the error threshold.

Proof. Define a function

$$M(U) = \left(|v\rangle\langle v| - \frac{1}{N}\right)^{\otimes 2}. \tag{A15}$$

Let $\mathrm{Var}_{\mu_{2,\alpha}}$ be defined as follows:

$$\mathrm{Var}_{\mu_{2,\alpha}} = E_{U\sim\mu_{2,\alpha}}\left[\left(\sum_{i=1}^{k}|y_i|^2\right)^2\right] - E^2_{U\sim\mu_{2,\alpha}}\left(\sum_{i=1}^{k}|y_i|^2\right) \tag{A16}$$

$$= E_{U\sim\mu_{2,\alpha}}\left[\sum_{i=1}^{k}|y_i|^4\right] + E_{U\sim\mu_{2,\alpha}}\left[\sum_{i\neq j}^{k}|y_i|^2|y_j|^2\right] - \frac{k^2}{N^2}. \tag{A17}$$

In terms of $M(U)$ we can write Eqs. (A7) and (A17) as

$$\mathrm{Tr}\left(E_{U\sim\mu_{2,\alpha}}\left[P_k^{\otimes 2}U^{\otimes 2}(M(U))U^{\otimes 2}P_k^{\otimes 2}\right]\right) = \mathrm{Var}_{\mu_{2,\alpha}}, \tag{A18}$$

$$\mathrm{Tr}\left(E_{U\sim\mu_H}\left[P_k^{\otimes 2}U^{\otimes 2}(M(U))U^{\otimes 2}P_k^{\otimes 2}\right]\right) = \mathrm{Var}_{\mu_H}. \tag{A19}$$

Then,

$$|\mathrm{Var}_{\mu_{2,\alpha}} - \mathrm{Var}_{\mu_H}| = \left|\mathrm{Tr}\left(E_{U\sim\mu_{2,\alpha}}\left[P_k^{\otimes 2}U^{\otimes 2}(M(U))U^{\otimes 2}P_k^{\otimes 2}\right] - E_{U\sim\mu_H}\left[P_k^{\otimes 2}U^{\otimes 2}(M(U))U^{\otimes 2}P_k^{\otimes 2}\right]\right)\right| \tag{A20}$$

$$\leqslant \left|P_k^{\otimes 2}\right|_2 |E_{U\sim\mu_{2,\alpha}}[U^{\otimes 2}(M(U))U^{\otimes 2}] - E_{U\sim\mu_H}[U^{\otimes 2}(M(U))U^{\otimes 2}]|$$

$$\leqslant k\frac{\alpha}{N^2}, \tag{A21}$$

where we used the monomial definition of approximate unitary 2-designs and its equivalence (see Ref. [35]) to the diamond norm definition used in the main text. Using Theorem 4, we get

$$\mathrm{Var}_{\mu_{2,\alpha}} \geqslant \frac{(N-k)k}{N^2(N-1)} - k\frac{\alpha}{N^2}. \tag{A22}$$

The variance for the $|\Pi\cdot\vec{v}|_2^2$ would be greater than $\frac{N-k}{k(N-1)} - \frac{\alpha}{k}$ which scales as $O(\frac{N-k}{Nk}) - \frac{\alpha}{k}$ for large N, k. Using the derived variance and the deviation from the mean to be 2ϵ gives (using Chebyshev's inequality)

$$P_{U\sim\mu_H}[||\Pi\cdot\vec{v}|_2 - 1| \leqslant \epsilon] \geqslant 1 - \left(\frac{N-k}{4kN\epsilon^2} + \frac{\alpha}{4\epsilon^2 k}\right). \tag{A23}$$

■

Theorem 6 (Theorem 3 in main text). For a random matrix Π satisfying the Johnson-Lindenstrauss (JL) lemma with a distortion $\epsilon\sqrt{\delta}$, where $\epsilon \leqslant 1/6$ and $\delta \leqslant 1/2$, the difference in the von Neumann entropy of a density matrix ρ computed using the random projection with Π (denoted as $\tilde{S}(\rho)$) and the true entropy $(S(\rho))$ can be bounded as follows:

$$|\tilde{S}(\rho) - S(\rho)| \leqslant \sqrt{3\epsilon}S(\rho) + \sqrt{\frac{9}{2}}\epsilon \tag{A24}$$

with probability at least $(1-\delta)$.

To prove the theorem we will need to use Theorems 10 and 13 of Ref. [26]. According to Theorem 13, let **DAD** be a symmetric positive-definite matrix such that **D** is a diagonal matrix and $A_{ii} = 1$ for all i. Also, let **DED** be a perturbation matrix such that $|E|_2 \leqslant \lambda_{\min}(A)$. Now, let λ_i be the ith eigenvalue of **DAD** and let λ_i' be the eigenvalue of $\mathbf{D}(\mathbf{A}+\mathbf{E})\mathbf{D}$. Then, $\forall i$,

$$|\lambda_i - \lambda_i'| \leqslant \frac{|E|_2}{\lambda_{\min}(A)}. \tag{A25}$$

Now, consider a distribution $\mathcal{D}$ on matrices $\Pi \in \mathbb{R}^{k\times N}$ (or $\mathbb{R}^{N\times k}$) satisfying

$$E_{\Pi\sim\mathcal{D}}\left|\left|\Pi x\right|_2^2 - 1\right|^2 \leqslant \epsilon^2\delta. \tag{A26}$$

Then, according to Theorem 13 of Ref. [73], for any orthonormal matrix O with N rows,

$$\mathrm{Pr}_{\Pi\sim D}[|O^T\Pi^T\Pi O - I|_2 \geqslant 3\epsilon] \leqslant \delta. \tag{A27}$$

Having these two results, we can derive the bounds on the accuracy of entropy computed post–random projection. In this regard, let us look at a general semipositive-definite density matrix ρ which could be written as $\rho = W\Sigma_p W^T$, where W has orthonormal columns and Σ_p is a diagonal matrix containing the singular values of ρ. We note that the eigenvalues of $\rho\rho^T = W\Sigma_p^2 W^T$ are equal to the eigenvalues of the diagonal matrix Σ_p^2. Similarly, the eigenvalues of $W\Sigma_p W^T\Pi\Pi^T W\Sigma_p W^T$ are equal to the eigenvalues of $\Sigma_p W^T\Pi\Pi^T W\Sigma_p$. Now, using Eq. (A25), we can compare the

eigenvalues of the matrices

$$\Sigma_p I_k \Sigma_p \quad \text{and} \quad \Sigma_p W^T \Pi \Pi^T W \Sigma_p.$$

Using $E = W^T \Pi \Pi^T W - I_k$ in Eq. (A27), we get that

$$|E|_2 \leqslant 3\epsilon \leqslant 1 \tag{A28}$$

with a probability greater than $1 - \delta$.

The eigenvalues of $\Sigma_p I_k \Sigma_p$ are equal to p_i^2 for $i = 1, \ldots, k$ and the eigenvalues of the $\Sigma_p W^T \Pi \Pi^T W \Sigma_p$ are equal to $\tilde{p}_i{}^2$, where $\tilde{p}_i$ are the singular values of $\Sigma_p W^T \Pi$. These are exactly equal to the singular values of $W \Sigma_p W^T \Pi = \rho \Pi$. This along with Eqs. (A28) and (A25) leads us to conclude

$$\left| p_i^2 - \tilde{p}_i{}^2 \right| \leqslant 3\epsilon p_i^2. \tag{A29}$$

This result guarantees that the singular values of ρ (from the main text) are captured with a perturbative factor of 3ϵ. Using this, we can find bounds to the error in the entropy computed after the random projection. We start with the upper bound:

$$\begin{aligned}
\Sigma_{i=1}^k \tilde{p}_i \ln\left(\frac{1}{\tilde{p}_i}\right) &\leqslant \Sigma_{i=1}^k (1+3\epsilon)^{1/2} p_i \ln\left(\frac{1}{(1-3\epsilon)^{1/2} p_i}\right) \\
&\leqslant (1+3\epsilon)^{1/2} S(\rho) + \frac{\sqrt{1+3\epsilon}}{2} \ln\left(\frac{1}{1-3\epsilon}\right) \\
&\leqslant (1+3\epsilon)^{1/2} S(\rho) + \frac{\sqrt{1+3\epsilon}}{2} \ln(1+6\epsilon) \\
&\leqslant (1+\sqrt{3\epsilon}) S(\rho) + \sqrt{\frac{9}{2}}\epsilon.
\end{aligned}$$

In the second-to-last inequality, we used $1/(1-3\epsilon) \leqslant (1+6\epsilon)$ for any $\epsilon \leqslant 1/6$, and in the last inequality we used $\ln(1+6\epsilon) \leqslant 6\epsilon$ for $0 \leqslant \epsilon \leqslant 1/6$. Similarly, we can find the lower bound to be

$$\Sigma_{i=1}^k \tilde{p}_i \ln\left(\frac{1}{\tilde{p}_i}\right) \geqslant (1-\sqrt{3\epsilon}) S(\rho) - \frac{3}{2}\epsilon.$$

Combining both bounds we get the bound for the error in entropy obtained after random projection $\tilde{S}$ with respect to true entropy S:

$$|\tilde{S}(\rho) - S(\rho)| \leqslant \sqrt{3\epsilon} S(\rho) + \sqrt{\frac{9}{2}}\epsilon. \tag{A30}$$

APPENDIX B: IMAGE RECONSTRUCTION POST-QUANTUM RANDOM PROJECTION

Figure 4 was a demonstration of the reconstruction of one image vector of the data set. Since the dimensionality reduction is for the data set as a whole, here we attach the plot of average $|\vec{x} - \vec{x}_{\text{recons}}|$ for all the image vectors $\vec{x}$ (the subset of MNIST containing 1000 images) and their reconstructed vectors $\vec{x}_{\text{recons}}$ in Fig. 9 for various reduced dimensions along with their 95% confidence intervals. The average norm of the difference between any two unit norm vectors in 1024 dimensional space is $\sqrt{2} \sim 1.414$. This sets a standard to compare the average norms plotted in Fig. 9.

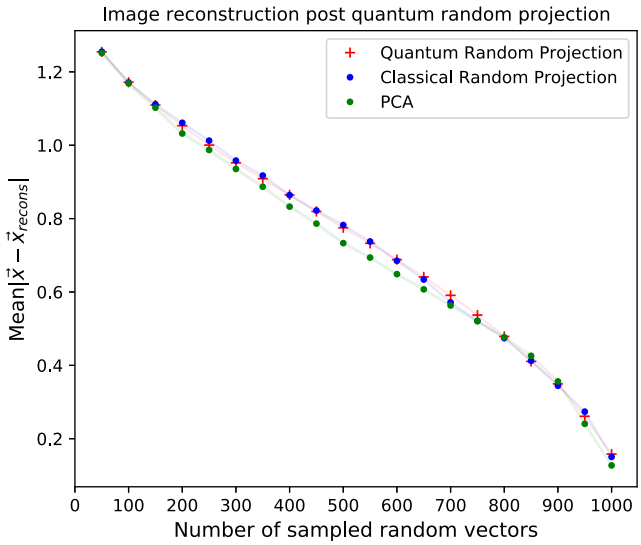

FIG. 9. The average norms of difference between image data vectors and the image vectors reconstructed from PCA, classical random projection, and quantum random projection schemes for varying reduced dimensions.

As mentioned in the experiments section, we assume that we can make arbitrary projection operators with any number of basis vectors, i.e., if the P_k operator projects to the first k basis state ($|p_1\rangle, |p_2\rangle, \ldots, |p_k\rangle$) in any basis. Then,

$$P_k = \sum_{i=1}^{k} |p_i\rangle\langle p_i|, \tag{B1}$$

where we do not have any restriction on what basis we pick and what values k can take. Quantum projectors are easy to construct when k takes values 128, 256, and 512 (corresponding to measuring one, two, and three qubits, respectively). Though it is impractical to construct such quantum projection operators for $k = 400$ and 700, we used those values only to compare the performance of quantum random projection against classical random projection which does not have any restrictions on the values k can take.

For the specific numbers mentioned in the figure, we used the first 400 and 700 components of the wave vector after the random quantum circuit (U) and added extra basis states with zero amplitudes to make them 1024 dimensional. Then, the wave vectors are reconstructed by the action of the inverse of U, i.e., $U^\dagger$. We understand that this process is more insightful when the projection operators are just one-, two-, or three-qubit measurements to specific states ($|0\rangle$ or $|1\rangle$). There, if the quantum random projection is made such that one of the qubits is projected to, say, $|0\rangle$ state, then reconstruction is done by appending the $U^\dagger$ circuit to the reduced quantum state plus measured qubit in $|0\rangle$. (This is equivalent to adding zeros to the basis elements where the measured qubit is in state $|1\rangle$ just like how we boosted dimensions from $k = 400$ or 700 to 1024.)

APPENDIX C: PERFORMANCE OF QUANTUM RANDOM PROJECTION ON EXPONENTIAL DECAY PROFILE

Figure 10 contains analogous plots discussed in the main text for the density matrices with exponential decay profile.

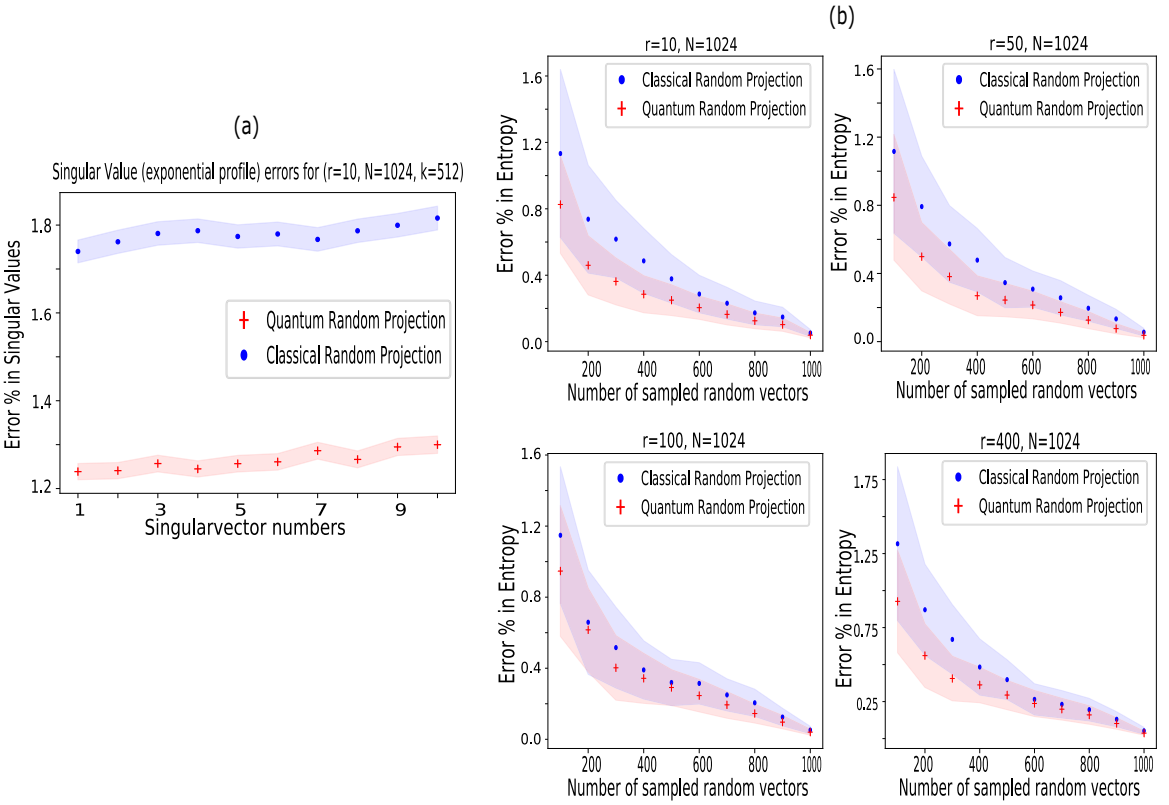

FIG. 10. Analogous plots discussed in the main text for the density matrices with exponential decay profile. (a) The accuracy with which the quantum random projector and the classical random projector pick the singular values of the density matrix for $r = 10$ when reducing the system size by half. The envelope represents 95% confidence intervals by running the experiments over 10 000 randomly generated density matrices. The percent error observed for the exponential decay profile is similar to the values we obtained for the linear decay profile. (b) The accuracies of quantum random projection and classical random projection in the entropy computation of randomly generated density matrices of size $N = 1024$ and ranks $r = 10$, 50, 100, and 400 with exponentially decaying profile. The envelopes represent their 90% confidence intervals by running the experiments over 100 randomly generated density matrices. The accuracies remain constant because the singular values do not vary much while varying the ranks owing to the rapid drop in successive singular values in an exponential decay profile.

APPENDIX D: PERFORMANCE OF QUANTUM RANDOM PROJECTION ON LARGER DATASETS

Figure 11 contains analogous plots discussed in the main text for system size = 2048 ($n = 11$).

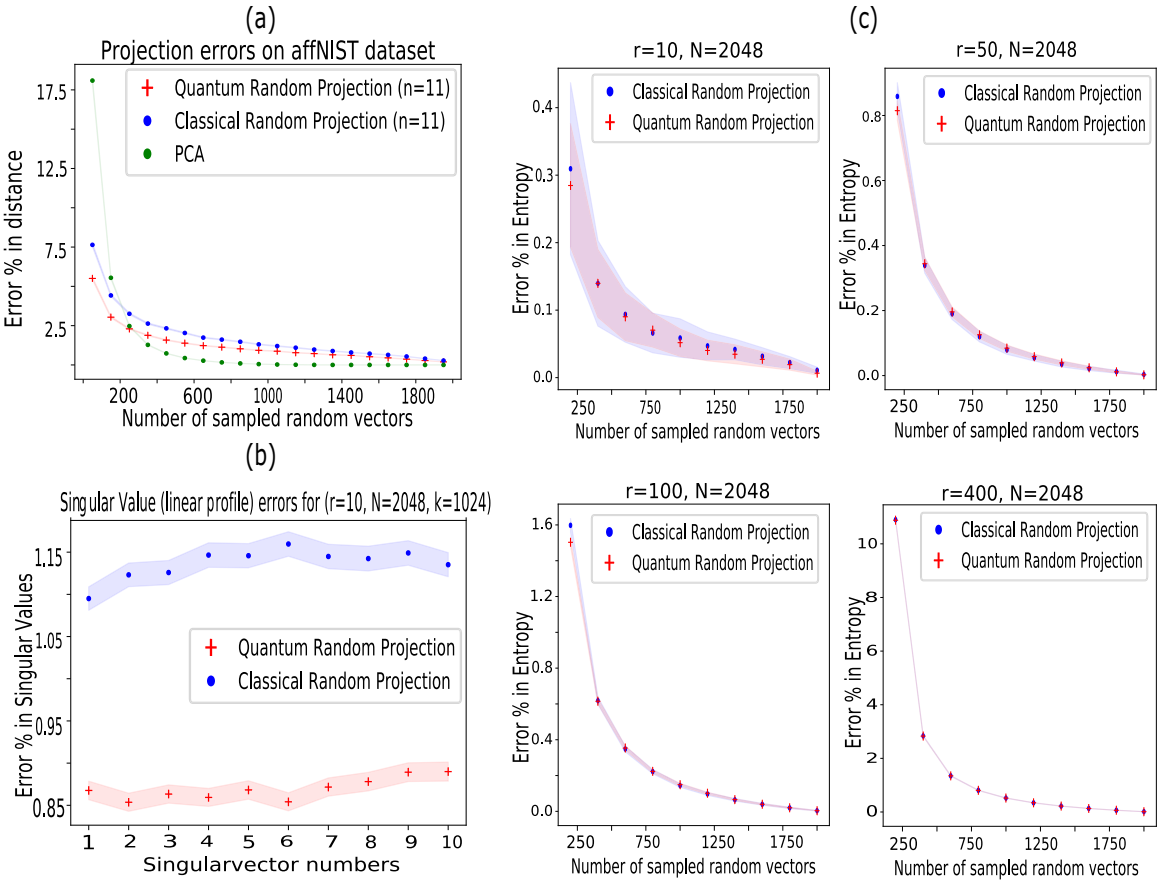

FIG. 11. Analogous plots discussed in the main text for system size = 2048 ($n = 11$). The percent error observed for $n = 11$ is better than the values obtained for the $n = 10$ case. This is because of improved efficiency in random projection for larger data sets as the error bounds derived in the theorems in our main text become tighter. (a) The mean percentage errors in the distance between 10 000 different random pairs of data vectors in the affNIST image data set which are 40×40 images and was boosted to 2048×1 data vectors for our experiment. Here, we can clearly notice that, at lower reduced dimensions, the random projection methods perform better than even PCA. Everywhere, the quantum random projection performs slightly better than the classical random projection. (b) The accuracy with which the quantum random projector and the classical random projector pick the singular values of the density matrix for $r = 10$ when reducing the system size by half. The envelope represents 95% confidence intervals by running the experiments over 10 000 randomly generated density matrices. (c) The accuracies of quantum random projection and classical random projection in the entropy computation of randomly generated density matrices of size $N = 2048$ and ranks $r = 10, 50, 100$, and 400 with linearly decaying singular value profile. The envelopes represent their 90% confidence intervals by running the experiments over 100 randomly generated density matrices. The accuracies improve with a decrease in the rank of the system just like the $n = 10$ case. However, the accuracies here are better than the $n = 10$ case due to the improved efficiency of random projection for larger data sets.

APPENDIX E: PERFORMANCE OF QUANTUM RANDOM PROJECTION CONSTRUCTED USING LESSER EXPRESSIVE LOCAL RANDOM QUANTUM CIRCUITS

Figure 12 contains analogous plots discussed in the main text for a lesser expressive local random circuit.

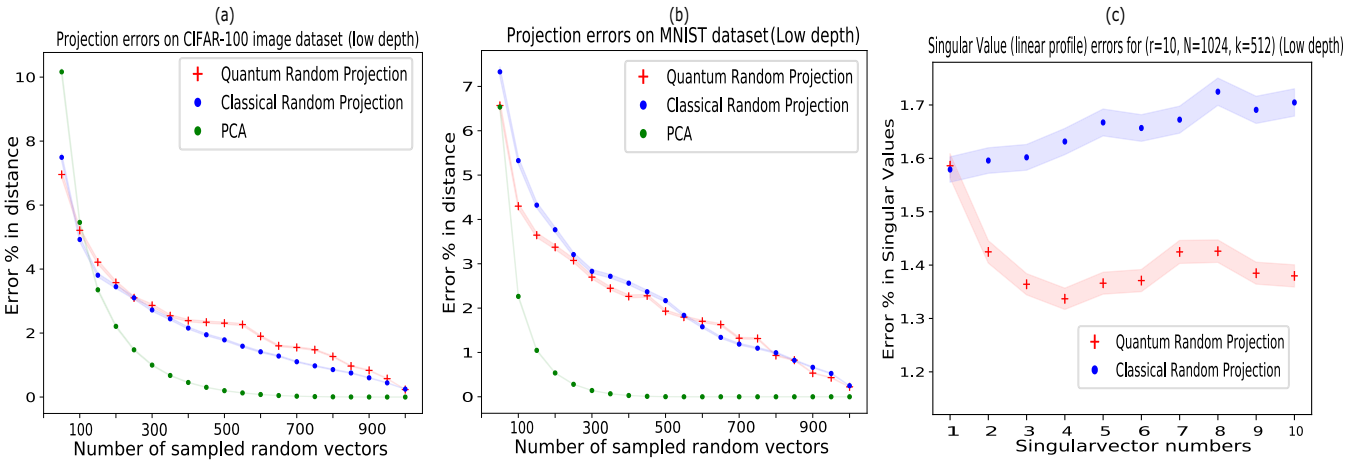

FIG. 12. Analogous plots discussed in the main text but the quantum random projector used here has been obtained from a local random quantum circuit far from the exact 2-design limit. This is our proxy for approximate 2-design. More specifically, we picked a depth of 50 in the quantum circuit shown in Fig. 1. Mean percentage errors in the distance between 10 000 different random pairs of data vectors in the (a) MNIST image data set and (b) CIFAR-100 image data set which are 28×28 images and were boosted to 1024×1 data vectors for our experiment. The performance of this quantum random projector is not as good as the exact 2-design quantum random projector. This is because the error bounds derived in the theorems for approximate unitary 2-designs are not as tight as an exact 2-design quantum random projector. (c) The accuracy with which the approximate unitary 2-design quantum random projector and the classical random projector pick the singular values of the density matrix for $r = 10$ when reducing the system size by half. The envelope represents 95% confidence intervals by running the experiments over 10 000 randomly generated density matrices. The drop in performance of the quantum random projector here could be attributed to the error bounds in the JL lemma not being as tight as the exact 2-design quantum random projector.

APPENDIX F: DETAILS REGARDING THE VQSVD PERFORMED TO RETRIEVE THE DOMINANT SINGULAR VECTORS

The variational quantum singular value decomposition was used to retrieve the dominant singular vectors of a randomly generated data matrix with rank $r = 5$ and singular values following a linearly decaying profile. The matrix on which one has to perform quantum random projection and quantum SVD needs to be loaded as a sum of unitaries or Pauli strings (unit depth). This process has exponential complexity for an exact representation of the matrix. But with importance sampling (as mentioned in Ref. [6]) one only uses a subset of Pauli strings which approximates the matrix to sufficient accuracy. However, for our computation, just like the exact encoding scheme, we used the exact matrix. The demonstration of accurately retrieving the singular vectors implies the same when the importance sampling creates the matrix with high accuracy.

Since the system size we used for this variational algorithm is large, we had to use the block initialization strategy discussed in Ref. [51] with two identity blocks in our training *Ansatz* to avoid the barren plateau issue [36]. Each block used in our *Ansatz* has a hardware efficient *Ansatz* circuit [74] of depth 25 followed by its inverse circuit (this part will remain an inverse circuit only at the beginning of the training; during training, the parameters of these two parts update independently) to start the training procedure close to identity and hence avoid the barren plateau. Figure 13 shows the rate of convergence of VQSVD in reconstructing the individual singular vectors after quantum random projection. The accuracies with which we retrieved the singular values and the dominant singular vectors could also be improved with other *Ansätze* which avoid barren plateau.

The projection operators in the figure are just measurements of certain qubits and making sure they are in a certain state ($|0\rangle$) or operators of the form $\frac{1}{2}(1+\sigma_z^i)$ which project to state $|0\rangle$ of qubit i.

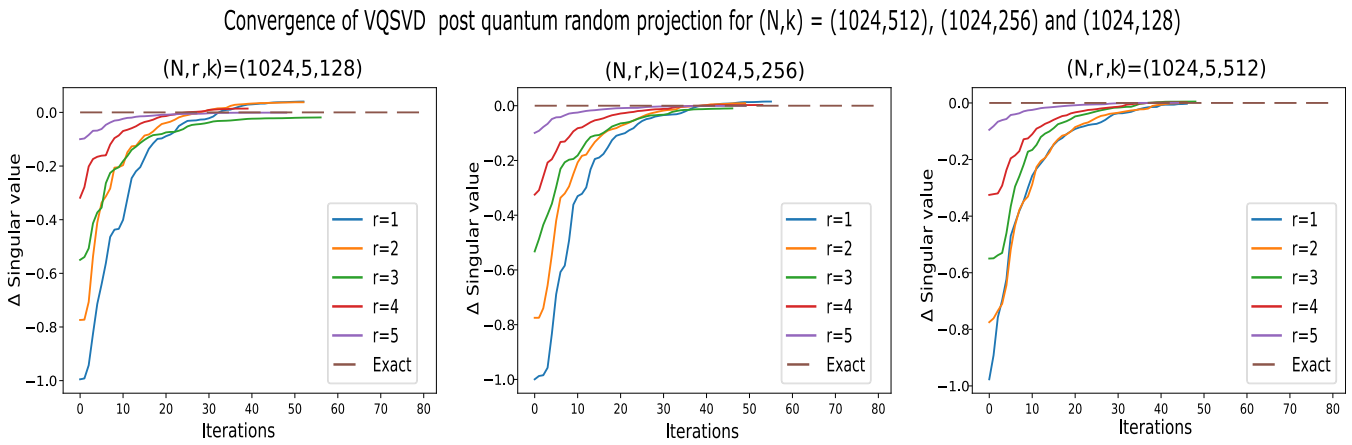

FIG. 13. The convergence of singular values obtained using the variational quantum singular value decomposition algorithm to their true values post–quantum random projection (2-design quantum random projector). We observe that sometimes the converged singular value is greater than the true value; this is because of the distortion of singular values created by the quantum random projection.

[1] S. Velliangiri, S. Alagumuthukrishnan, and S. I. Thankumar Joseph, A review of dimensionality reduction techniques for efficient computation, Procedia Comput. Sci. **165**, 104 (2019).

[2] N. Salem and S. Hussein, Data dimensional reduction and principal components analysis, Procedia Comput. Sci. **163**, 292 (2019).

[3] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, Nat. Phys. **10**, 631 (2014).

[4] P. Rebentrost, A. Steffens, I. Marvian, and S. Lloyd, Quantum singular-value decomposition of nonsparse low-rank matrices, Phys. Rev. A **97**, 012327 (2018).

[5] I. Kerenidis and A. Prakash, Quantum recommendation systems, *Information Technology Convergence and Services* (2016).

[6] X. Wang, Z. Song, and Y. Wang, Variational quantum singular value decomposition, Quantum **5**, 483 (2021).

[7] L. Wossnig, Z. Zhao, and A. Prakash, Quantum linear system algorithm for dense matrices, Phys. Rev. Lett. **120**, 050502 (2018).

[8] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[9] P. Drineas and M. W. Mahoney, RandNLA: Randomized numerical linear algebra, Commun. ACM **59**, 80 (2016).

[10] D. Achlioptas, Database-friendly random projections: Johnson-Lindenstrauss with binary coins, J. Comput. Syst. Sci. **66**, 671 (2003).

[11] E. Bingham and H. Mannila, Random projection in dimensionality reduction: applications to image and text data, *Knowledge Discovery and Data Mining* (2001).

[12] W. B. Johnson and J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, Contemporary mathematics **26**, 189 (1984).

[13] S. Paul, C. Boutsidis, M. Magdon-Ismail, and P. Drineas, Random projections for linear support vector machines, ACM Trans. Knowlegge Discovery Data **8**, 1 (2012).

[14] J. Alman and V. V. Williams, A refined laser method and faster matrix multiplication, in *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)* (SIAM, 2021), pp. 522–539.

[15] N. Ailon and B. Chazelle, The fast Johnson-Lindenstrauss transform and approximate nearest neighbors, SIAM J. Comput. **39**, 302 (2009).

[16] E. Knill, Approximation by quantum circuits, arXiv:quant-ph/9508006.

[17] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[18] C. Dankert, Efficient simulation of random quantum states and operators, arXiv:quant-ph/0512217.

[19] A. Ambainis and J. Emerson, Quantum t-designs: t-wise independence in the quantum world, in *Proceedings of the 22nd Annual IEEE Conference on Computational Complexity (CCC'07)* (IEEE, New York, 2007), pp. 129–140.

[20] A. W. Harrow and R. A. Low, Random quantum circuits are approximate 2-designs, Commun. Math. Phys. **291**, 257 (2009).

[21] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial designs, Commun. Math. Phys. **346**, 397 (2016).

[22] J. Haferkamp, Random quantum circuits are approximate unitary t-designs in depth $O(nt^{5+o(1)})$, Quantum **6**, 795 (2022).

[23] P. Sen, An efficient superpostional quantum Johnson–Lindenstrauss lemma via unitary t-designs, Quantum Info. Proc. **20**, 312 (2021).

[24] L. Deng, The MNIST database of handwritten digit images for machine learning research, IEEE Signal Process. Mag. **29**, 141 (2012).

[25] A. Krizhevsky, Learning multiple layers of features from tiny images, Technical report, 2009 (unpublished).

[26] E.-M. Kontopoulou, G.-P. Dexter, W. Szpankowski, A. Grama, and P. Drineas, Randomized linear algebra approaches to estimate the von Neumann entropy of density matrices, IEEE Trans. Info. Theory **66**, 5003 (2020).

[27] Y. Dong, T. Gong, S. Yu, H. Chen, and C. Li, Robust and fast measure of information via low-rank representation, *AAAI Conference on Artificial Intelligence* (2022).

[28] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, Johnson-Lindenstrauss lemma, linear and nonlinear random projections, random Fourier features, and random kitchen sinks: Tutorial and survey, arXiv:2108.04172.

[29] J. Lacotte and M. Pilanci, Optimal randomized first-order methods for least-squares problems, *International Conference on Machine Learning* (2020).

[30] P. Drineas, M. Mahoney, S. Muthukrishnan, and T. Sarlós, Faster least squares approximation, Numer. Math. **117**, 219 (2011).

[31] J. Nelson and H. L. Nguyen, Sparsity lower bounds for dimensionality reducing maps, in *Proceedings of the 45th ACM Symposium on Theory of Computing (STOC), Palo Alto, CA* (ACM Press, New York, 2013).

[32] J. A. Tropp, Improved analysis of the subsampled randomized Hadamard transform, Adv. Adapt. Data Anal. **3**, 115 (2011).

[33] K. Nakaji and N. Yamamoto, Expressibility of the alternating layered ansatz for quantum computation, Quantum **5**, 434 (2021).

[34] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX Quantum **3**, 010313 (2022).

[35] R. A. Low, Pseudo-randomness and learning in quantum computation, arXiv:1006.5227.

[36] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[37] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum autoencoders for efficient compression of quantum data, Quantum Sci. Technol. **2**, 045001 (2017).

[38] Y. Lü, Q. Gao, J. Lü, M. Ogorzałek, and J. Zheng, A quantum convolutional neural network for image classification, in *Proceedings of the 2021 40th Chinese Control Conference (CCC), Shanghai, China* (IEEE, New York, 2021), pp. 6329–6334.

[39] C. Zoufal, A. Lucchi, and S. Woerner, Quantum generative adversarial networks for learning and loading random distributions, npj Quantum Inf. **5**, 103 (2019).

[40] K. Nakaji, S. Uno, Y. Suzuki, R. Raymond, T. Onodera, T. Tanaka, H. Tezuka, N. Mitsuda, and N. Yamamoto, Approximate amplitude encoding in shallow parameterized quantum circuits and its application to financial market indicators, Phys. Rev. Res. **4**, 023136 (2022).

[41] N. Mitsuda, K. Nakaji, Y. Suzuki, T. Tanaka, R. Raymond, H. Tezuka, T. Onodera, and N. Yamamoto, Approximate complex amplitude encoding algorithm and its application to data classification problems, arXiv:2211.13039.

[42] P. Q. Le, F. Dong, and K. Hirota, A flexible representation of quantum images for polynomial preparation, image compression, and processing operations, Quantum Inf. Process. **10**, 63 (2011).

[43] Y. Zhang, K. Lu, Y. Gao, and M. Wang, NEQR: A novel enhanced quantum representation of digital images, Quantum Inf. Process. **12**, 2833 (2013).

[44] R. Hecht-Nielsen, Context vectors: General purpose approximate meaning representations self-organized from raw data, in *Computational Intelligence: Imitating Life* (IEEE Press, Piscataway, NJ, 1994), pp. 43–56.

[45] X.-W. Yao, H. Wang, Z. Liao, M.-C. Chen, J. Pan, J. Li, K. Zhang, X. Lin, Z. Wang, Z. Luo, W. Zheng, J. Li, M. Zhao, X. Peng, and D. Suter, Quantum image processing and its application to edge detection: Theory and experiment, Phys. Rev. X **7**, 031041 (2017).

[46] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[47] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of Pauli observables by derandomization, Phys. Rev. Lett. **127**, 030503 (2021).

[48] G. Giudice, A. Çakan, J. I. Cirac, and M. C. Bañuls, Rényi free energy and variational approximations to thermal states, Phys. Rev. B **103**, 205128 (2021).

[49] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[50] M. Cerezo, K. Sharma, A. Arrasmith, and P. J. Coles, Variational quantum state eigensolver, npj Quantum Inf. **8**, 113 (2022).

[51] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, Quantum **3**, 214 (2019).

[52] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert, Anticoncentration theorems for schemes showing a quantum speedup, Quantum **2**, 65 (2018).

[53] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Random quantum circuits anticoncentrate in log depth, PRX Quantum **3**, 010333 (2022).

[54] Q. Wang, Kernel principal component analysis and its applications in face recognition and active shape models, arXiv:1207.3538.

[55] A. Daskin, R. Gupta, and S. Kais, Dimension reduction and redundancy removal through successive Schmidt decompositions, Appl. Sci. **13**, 3172 (2023).

[56] M. Sajjan, J. Li, R. Selvarajan, S. H. Sureshbabu, S. S. Kale, R. Gupta, V. Singh, and S. Kais, Quantum machine learning for chemistry and physics, Chem. Soc. Rev. **51**, 6475 (2022).

[57] M. Sajjan, H. Alaeian, and S. Kais, Magnetic phases of spatially modulated spin-1 chains in Rydberg excitons: Classical and quantum simulations, J. Chem. Phys. **157**, 224111 (2022).

[58] R. Borissov, S. Major, and L. Smolin, The geometry of quantum spin networks, Class. Quantum Grav. **13**, 3183 (1996).

[59] M. Sajjan, V. Singh, R. Selvarajan, and S. Kais, Imaginary components of out-of-time-order correlator and information scrambling for navigating the learning landscape of a quantum machine learning model, Phys. Rev. Res. **5**, 013146 (2023).

[60] R. Gupta, R. Selvarajan, M. Sajjan, R. D. Levine, and S. Kais, Hamiltonian learning from time dynamics using variational algorithms, J. Phys. Chem. A **127**, 3246 (2023).

[61] W. Yu, J. Sun, Z. Han, and X. Yuan, Robust and efficient Hamiltonian learning, Quantum **7**, 1045 (2023).

[62] R. Gupta, M. Sajjan, R. D. Levine, and S. Kais, Variational approach to quantum state tomography based on maximal entropy formalism, Phys. Chem. Chem. Phys. **24**, 28870 (2022).

[63] R. Gupta, R. Xia, R. D. Levine, and S. Kais, Maximal entropy approach for quantum state tomography, PRX Quantum **2**, 010318 (2021).

[64] M. Sajjan, S. H. Sureshbabu, and S. Kais, Quantum machine-learning for eigenstate filtration in two-dimensional materials, J. Am. Chem. Soc. **143**, 18426 (2021).

[65] R. Selvarajan, M. Sajjan, and S. Kais, Variational quantum circuits to prepare low energy symmetry states, Symmetry **14**, 457 (2022).

[66] R. Yaacoby, N. Schaar, L. Kellerhals, O. Raz, D. Hermelin, and R. Pugatch, Comparison between a quantum annealer and a classical approximation algorithm for computing the ground state of an Ising spin glass, Phys. Rev. E **105**, 035305 (2022).

[67] P. M. Schindler, T. Guaita, T. Shi, E. Demler, and J. I. Cirac, Variational ansatz for the ground state of the quantum Sherrington-Kirkpatrick model, Phys. Rev. Lett. **129**, 220401 (2022).

[68] S. Mirniaharikandehei, M. Heidari, G. Danala, S. Lakshmivarahan, and B. Zheng, Applying a random projection algorithm to optimize machine learning model for predicting peritoneal metastasis in gastric cancer patients using CT images, Comput. Methods Programs Biomed. **200**, 105937 (2021).

[69] M. Heidari, S. Lakshmivarahan, S. Mirniaharikandehei, G. Danala, S. K. R. Maryada, H. Liu, and B. Zheng, Applying a random projection algorithm to optimize machine learning model for breast lesion classification, IEEE Trans. Biomed. Eng. **68**, 2764 (2020).

[70] S. Oh and S. Kais, Comparison of quantum advantage experiments using random circuit sampling, Phys. Rev. A **107**, 022610 (2023).

[71] S. Oh and S. Kais, Statistical analysis on random quantum circuit sampling by Sycamore and Zuchongzhi quantum processors, Phys. Rev. A **106**, 032433 (2022).

[72] Paddle Quantum (2020), https://github.com/PaddlePaddle/Quantum.

[73] D. P. Woodruff, Sketching as a tool for numerical linear algebra, Found. Trends Theor. Comput. Sci. **10**, 1 (2014).

[74] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).