Tensor network compressed sensing with unsupervised machine learning

Shi-Ju Ran ●,[1,*] Zheng-Zhi Sun ●,[2] Shao-Ming Fei,[3,4] Gang Su,[2,5] and Maciej Lewenstein[6,7]

[1]*Department of Physics, Capital Normal University, Beijing 100048, China*
[2]*School of Physical Sciences, University of Chinese Academy of Sciences, P. O. Box 4588, Beijing 100049, China*
[3]*School of Mathematical Sciences, Capital Normal University, Beijing 100048, China*
[4]*Max-Planck-Institute for Mathematics in the Sciences, 04103, Leipzig, Germany*
[5]*Kavli Institute for Theoretical Sciences, and CAS Center for Excellence in Topological Quantum Computation,*
University of Chinese Academy of Sciences, Beijing 100190, China
[6]*ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[7]*ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain*

We propose the tensor-network compressed sensing (TNCS) by incorporating the ideas of compressed sensing, tensor network (TN), and machine learning. The primary idea is to compress and communicate the real-life information through the generative TN state and by making projective measurements in a designed way. First, the state $|\Psi\rangle$ is obtained by the unsupervised learning of TN, and then the data to be communicated are encoded in the separable state with the minimal distance to the projected state $|\Phi\rangle$, where $|\Phi\rangle$ can be acquired by partially projecting $|\Psi\rangle$. A protocol analogous to the compressed sensing assisted by neural-network machine learning is thus suggested, where the projections are designed to rapidly minimize the uncertainty of information in $|\Phi\rangle$. To characterize the efficiency of TNCS, we propose a quantity named as q sparsity to describe the sparsity of quantum states, which is analogous to the sparsity of the signals required in the standard compressed sensing. The need of the q sparsity in TNCS is essentially due to the fact that the TN states obey the area law of entanglement entropy. The tests on the real-life data (handwritten digits and fashion images) show that the TNCS has competitive efficiency and accuracy.

I. INTRODUCTION

Hybridizing the ideas and techniques in information theories and quantum physics has given birth to significant and fruitful achievements. On one hand, quantum physics may enhance the communications of classical information. Taking dense/superdense coding protocols [1–7] as examples, the idea is to use previously shared entangled state between a sender and the receiver(s) to send more classical information than is possible without the resource of entanglement. On the other hand, classical techniques can assist quantum approaches. One example is to use compressed sensing [8] (see also the book in Ref. [9]) to improve quantum state tomography [10–13].

We here consider to combine the ideas of compressed sensing [8], quantum communication [14], and unsupervised tensor network (TN) machine learning [15]. Compressed sensing is a powerful scheme for classical data compression by sampling, which is particularly useful when the samplings of the data are difficult or expensive. For instance in the

magnetic resonance imaging, compressed sensing can largely compress the required samplings, thus significantly improve the efficiency [16].

In quantum communication, measurements are also expensive, since quantum states are difficult to prepare and each measurement will collapse or disturb the state to some extent. Consequently, the quantum communications of real-life data [e.g., images of $O(10^3)$ bits or more], even including the corresponding simulations of the quantum processes on classical computers, are extremely challenging. For most of the existing quantum schemes, it is difficult to generalize them to implement realistic tasks, by using many-qubit (or many-body) states instead of the states with a small number of qubits. This is also partially because the Hilbert space grows exponentially with the number of qubits, and it is extremely challenging to use classical numeric simulations to guide the quantum schemes with many qubits. Meanwhile, the quantum technologies are advanced rapidly, and the number of qubits that can be entangled in a well-controlled way increases fast (see for instance several recent progresses in Refs. [17–19]). Therefore, developing algorithms and protocols to handle real-life problems using many-qubit states is important and urgent.

Recently, TN [20–24] is rapidly developed into a powerful quantum-inspired computational tool for machine learning, which brings new possibilities and wide perspectives to process real-life data, such as images and texts, in the quantum processes based on many-qubit (or many-body) states

---

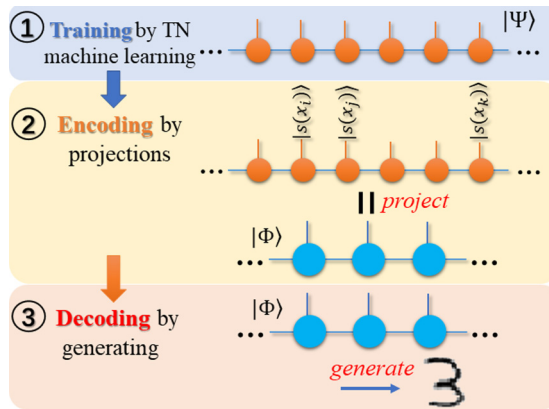*Corresponding author: sjran@cnu.edu.cn

FIG. 1. Illustration of the main steps of TNCS: (1) train the Born machine $|\Psi\rangle$ representing the probability distribution of the data that Alice considers to send, (2) encode the specific piece of information to be sent by projecting $|\Psi\rangle$, (3) decode the information as a generative process by the projected Born machine.

[15,25–32]. High efficiencies have been demonstrated at least for the classical simulations of these quantum processes. Meanwhile, the underlying relations between TN and the models executable on quantum platforms (e.g., as quantum circuits [31,33] or through quantum channels [34]) are revealed. These achievements allow and motivate to explore novel quantum schemes that previously could not be efficiently simulated even by classical simulations, which will provide valuable results for the future investigations on the genuine quantum hardware [31].

In this work, the tensor-network compressed sensing (TNCS) is proposed. The main idea is to encode and communicate the information by implementing designed projections on the state $|\Psi\rangle$ (also called Born machine [35]) that is trained by the unsupervised TN machine learning algorithm [15]. To explain TNCS, let us consider the following scenario. Alice wants to send a piece of classical information $\{x\}$, e.g., an image of hand-written digit "3" consisting of $O(10^2)$ pixels, to Bob. She intends to send only a small number of pixels (or features in the terminology of machine learning) denoted by $\{x^{\text{sent}}\}$ to Bob by classical communication which might be unsafe or even public. The rest of the information $\{x^{\text{rest}}\}$ (with $\{x\} = \{x^{\text{sent}}\} \cup \{x^{\text{rest}}\}$) will be encoded in the Born machine $|\Psi\rangle$. To recover $\{x^{\text{rest}}\}$, Bob projects $|\Psi\rangle$ that is previously provided by Alice in the way determined by $\{x^{\text{sent}}\}$. After the projections, $|\Psi\rangle$ will be projected to another entangled state denoted as $|\Phi\rangle$, and by design, $\{x^{\text{rest}}\}$ will be encoded in the separable state that has the minimal distance to $|\Phi\rangle$. Therefore, Bob can reliably recover $\{x^{\text{rest}}\}$ by implementing projections on $|\Phi\rangle$, or generate $\{x^{\text{rest}}\}$ by sampling on $|\Phi\rangle$ in the terms of TN machine learning. A flowchart of TNCS is given in Fig. 1.

There remain two key questions: how to construct $|\Psi\rangle$ and how to design the projections on it, so that the information to be sent can be optimally encoded in $|\Phi\rangle$ in the above way. Our proposal is the following. First, Alice trains $|\Psi\rangle$ by the unsupervised TN machine learning algorithm [15], so that $|\Psi\rangle$ represents the probability distribution of a huge amount of information that Alice considers to send. $|\Psi\rangle$ is called a Born

machine since the probability of each piece of information is the square of the corresponding coefficient in $|\Psi\rangle$ [35]. Then to send a specific piece of information, she chooses to send Bob the pixels, with which the uncertainty of the rest of the pixels in the probability distribution will be minimized. The full information $\{x\}$ is efficiently compressed to (or in other words, can be accurately reconstructed from) a small part of the image $\{x^{\text{rest}}\}$ and the Born machine $|\Psi\rangle$.

We test our TNCS with the datasets of hand-written digits and fashion images (namely MNIST [36] and fashion-MNIST [37]) by classical simulations. Any image in the training or testing sets can be reconstructed reliably and efficiently. The efficiency is indicated by the compression ratio $r = N_f/N \simeq 10\%$, where $N = \#\{x\}$ denotes the total number of features in $\{x\}$ and $N_f = \#\{x^{\text{[sent]}}\}$ denotes the number of the sent features. In other words, the information Bob accesses is about 10 times the information that Alice needs to send through the classical channels. Most of the information is encoded in the Born machine (quantum state). Similar to the compressed sensing, randomly choosing $\{x^{\text{[sent]}}\}$ already leads to small compression ratios. Better performance is reached by choosing $\{x^{\text{[sent]}}\}$ with a sampling protocol based on the entanglement of $|\Psi\rangle$ and by implementing post-selections to access $\{x^{\text{[rest]}}\}$. Finally, q sparsity to characterize the sparsity of quantum state is proposed. For TNCS, q sparsity characterizes how fast the Shannon entropy of the probability distribution will decrease by projecting the Born machine $|\Psi\rangle$ and how efficient the compressed sampling can be via $|\Psi\rangle$. An empirical equation to estimate the required number of pixels for reliable reconstructions is given.

Our proposal shares a similar spirit with the recent works that combine the standard compressed sensing with the classical probabilistic models. For instance, the autoencoders parameterized by deep neural networks are used to design an efficient Markov sampling process to significantly reduce the compression ratio [38,39]. Here, we employ a matrix product state (MPS) [40] that possesses simple and shallow architecture to encode the information and to design the sampling strategy based on the entanglement. While this work will focus on demonstrating TNCS by classical simulations, our scheme could in principle be implemented on quantum platforms. The TNCS would pave a way to processing real-life data through quantum many-body states. The possible advantages of the TNCS in efficiency, security, etc. when implemented on quantum platforms are to be discussed in future.

This paper is organized as follows. In Sec. II, we explain the basic concepts and theories of TNCS. In Sec. III, TNCS is improved by introducing a quantum-inspired protocol and post-selections. In Sec. IV, we introduce q sparsity to characterize the sparsity of quantum states based on entanglement and projections and in this work to qualitatively characterize the efficiency of TNCS. At last, we present a summary and the perspective on quantum communications in Sec. V.

II. TENSOR-NETWORK COMPRESSED SENSING

In this section, we will mainly explain TNCS as a TN scheme that efficiently deals with quantum many-body states (optimizations, projections, etc.) by classical simulations.

Note that TNCS can in principle be generalized to quantum platforms. Some discussions on the quantum nature of TNCS can be found in Appendices D, E, and F from the perspectives of entanglement, security, and efficiency.

Suppose Alice wants to send Bob an image of a handwritten digit "3" by TNCS (Fig. 1). She first trains the state $|\Psi\rangle$ as the generative model for the training set of many "3" images in MNIST. This can be done with the unsupervised TN machine learning algorithm [15]. The idea is to first map the images to quantum states. For example, the nth pixel ($0 \leqslant x_n \leqslant 1$) is mapped to a state of a qubit as

$$x_n \rightarrow |s(x_n)\rangle = \cos(x_n\pi/2)|0\rangle + \sin(x_n\pi/2)|1\rangle, \quad (1)$$

with $|0\rangle$ and $|1\rangle$ the two eigenstates of the Pauli matrix $\hat{\sigma}^z$. In this way, one image with pixels $\{x\} = (x_1, x_2, \ldots)$ is mapped to a separable state $|\psi\rangle = \prod_n |s(x_n)\rangle$. Then the Born machine $|\Psi\rangle$ is optimized to capture the probability distribution of the training set, by minimizing the distance (negative-log likelihood) to the probability distribution of the images in the training set. See more details in Appendix A. Here, we take $|\Psi\rangle$ in the form of MPS [40]. Note that TNCS is a general scheme, where one may also choose other TN forms to represent $|\Psi\rangle$, such as tree TN or MERA [26,31,32], or simply a quantum state without a specific entanglement structure.

In the sense of machine learning, though we only use the "3" images in the training set to optimize $|\Psi\rangle$, it is expected that $|\Psi\rangle$ approximately gives the probability distribution of any "3" images. In other words, $|\Psi\rangle$ learns the probability distribution of the "3" images from a finite (training) set but can generalize to generate and/or recognize arbitrary "3" images that $|\Psi\rangle$ might have never learned. The ability of a machine-learning model to process the information beyond the training set is known as the generalization power (see, e.g., Ref. [41]). As shown in the previous works [15,25,26,28–32], TN models (including MPS) possess remarkable generalization power that is competitive to neural networks. Notably, TN models surpass neural networks as they possess high interpretability and allow us to implement quantum processes.

As $|\Psi\rangle$ gives the probability distribution of the "3" images in the training set and beyond (due to its generalization power), it is then possible to use $|\Psi\rangle$ to communicate any "3" images. As a direct advantage, Alice can train $|\Psi\rangle$ without knowing the specific "3" image that will be sent to Bob. In other words, different "3" images can be communicated with the same state $|\Psi\rangle$, as long as $|\Psi\rangle$ can "recognize" (in the sense of machine learning) it as an image of "3" (see Appendix B for more discussions).

In the communication, Alice sends Bob only a small part of this image $\{x^{\text{sent}}\}$ and $|\Psi\rangle$; then Bob projects $|\Psi\rangle$ according to $\{x^{\text{sent}}\}$ as

$$|\Phi\rangle = \prod_{x_n \in \{x^{[\text{sent}]}\}} \langle s(x_n)|\Psi\rangle/C, \quad (2)$$

with C a constant to normalize $|\Phi\rangle$. $\{x^{\text{sent}}\}$ should be selected so that #$\{x^{\text{sent}}\}$ is small and meanwhile Bob can accurately reconstruct the rest of the pixels $\{x^{\text{rest}}\}$ from $|\Phi\rangle$. Note that in the classical sense, Bob just needs the data of $|\Psi\rangle$ (e.g., the tensors in the MPS) to simulate the projections. When considering TNCS as a quantum protocol, many copies of $|\Psi\rangle$

will be needed to obtain at least one expected projected state. Some relevant discussions can be found in Appendix F.

The selection of $\{x^{\text{sent}}\}$ is analog to the sampling process of compressed sensing [9]. In a standard compressed sensing scheme, one may randomly choose a certain number of pixels from the image as $\{x^{\text{sent}}\}$. The sampling can be compressed since the randomly selected pixels approximately lead to averagely distributed frequencies, and meanwhile an image in the frequency space is normally sparse. These are known as incoherence and sparsity, which are the two conditions for the compressed sensing to efficiently work.

For the TNCS, one may also randomly choose $\{x^{\text{sent}}\}$ from $\{x\}$, which is denoted as random ordering (RO). Each projection by $|s(x_n)\rangle$ in Eq. (2) projects the state towards the separable state $|\psi\rangle = \prod_n |s(x_n)\rangle$. With sufficient data in $\{x^{\text{sent}}\}$, $|\Phi\rangle$ will eventually be projected to such a state, where $\prod_n |s(x_n)\rangle$ ($x_n \in \{x^{\text{rest}}\}$) is the separable state that has the minimal distance to $|\Phi\rangle$ among all separable states. Therefore, Bob can access $\{x^{\text{rest}}\}$ by simply projecting on $|\Phi\rangle$. Until now, we can already see that the data can be efficiently compressed by the TNCS if $\prod_{x_n \in \{x^{\text{rest}}\}} |\langle \Phi|s(x_n)\rangle| \rightarrow 1$ with #$\{x^{\text{sent}}\} \ll$ #$\{x\}$. In other words, Bob will accurately have #$\{x^{\text{sent}}\}$ when $|\Phi\rangle$ is mapped to a state that is sufficiently close to the separable state corresponding to #$\{x^{\text{rest}}\}$. Later in Sec. IV, we will give more discussions about the efficiency from the sparsity of quantum states.

For extracting $\{x^{\text{rest}}\}$, let us consider that Bob samples only once from $|\Phi\rangle$ for each pixel, which is dubbed as the one-shot measurement. When preparing $|\Phi\rangle$ as a quantum state (instead of the classical data as a TN), the one-shot measurement corresponds to measuring only on one copy of $|\Phi\rangle$. To generate $\{x^{\text{rest}}\}$, he measures the qubits in the basis of the Pauli matrix $\hat{\sigma}^z$. The probability $P(x_n)$ of the nth pixel $x_n = 0$ or 1 is determined by $\hat{\rho}_n$ as $P(x_n) = \langle x_n|\hat{\rho}_n|x_n\rangle$, where $\hat{\rho}_n$ is the reduced density matrix with respect to the nth qubit

$$\hat{\rho}_n = \text{Tr}_{/n}|\Phi\rangle\langle\Phi|, \quad (3)$$

with $\text{Tr}_{/n}$ the trace over all degrees of freedom except for the nth qubit. Note $\sum_{x_n} P(x_n) = \text{Tr}\hat{\rho}_n = 1$ due to the normalization of $|\Phi\rangle$. Considering TNCS as a quantum protocol, the one-shot measurement will be quite cheap and feasible in experiments as it requires only one copy of the state. One drawback is that only black-or-white pixels ($x_n = 0$ or 1) will be generated, instead of the grayscale ones. From the perspective of machine learning on classical computers, such a way of obtaining $\{x^{\text{rest}}\}$ is in fact to generate $\{x^{\text{rest}}\}$ by the Born machine $|\Phi\rangle$ [15], which involves the sampling processes based on the probability distributions $P(x_n)$.

We test the TNCS with RO and the one-shot measurement on MNIST and fashion-MNIST datasets, which consists of the real-life images of handwritten digits and Zalando's articles, respectively. Each dataset contains 10 classes of images and in total has 60 000 training images and 10 000 testing images. Each image contains $28 \times 28 = 784$ grayscale pixels. In Figs. 2(a) and 2(b), we show the accuracy of TNCS with different compression ratios $r = N_f/N$ (green solid and the purple dashed lines). The accuracy is characterized by the average peak signal-to-noise ratio (PSNR), which (say between the original data $\{x\}$ and the reconstructed data $\{y\}$) is
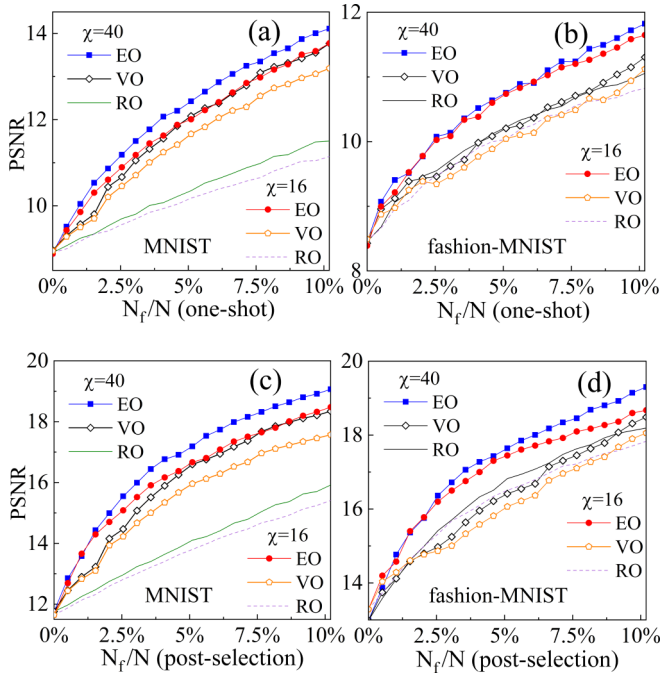
FIG. 2. Average peak signal-to-noise ratio (PSNR) of the constructed images in the testing dataset of the handwriting digits "3" in MNIST and the dresses in fashion-MNIST. The images are generated from $|\Phi\rangle$ in the one-shot way [(a) and (b)] or with the post-selection [(c) and (d)]. The dimension of the MPS is taken as $\chi = 16$ or 40. The number of known pixels for reconstruction ranges from about $N_f/N = 0\%$ to 10%. We try three orderings, which are entanglement ordering (EO), variance ordering (VO), and random ordering (RO). See the details in the text.

defined as

$$\mathrm{PSNR}(\{x\}, \{y\}) = 10 \log_{10} \frac{784}{\sum_n (x_n - y_n)^2}. \quad (4)$$

We average the PSNR by the results of reconstructing all the images in the testing set, which the Born machine did not learn in the training process. We take the bond dimensions of the MPS $\chi = 16$ and 40. Generally, the PSNR increases with r and χ as expected, and TNCS works well by simply sampling a small number of $\{x^{\mathrm{sent}}\}$ randomly from $\{x\}$ and implementing one-shot measurement on $|\Psi\rangle$.

III. IMPROVING EFFICIENCY WITH ENTANGLEMENT-ORDERED SAMPLING PROTOCOL AND POST-SELECTIONS

While RO works well for TNCS, in the following, we will propose to improve the performance (i.e., of higher PSNR and higher efficiency with smaller compression ratio) by incorporating with a quantum-inspired sampling protocol based on entanglement and the post-selections of measurements.

Regarding the sampling, the results will change if Alice selects differently the $\{x^{\mathrm{sent}}\}$. Other than RO, a natural selection way dubbed as variance ordering (VO) is to select the pixels according to the variance. The variance of the nth pixel

is calculated from the training set as

$$V_n = \sum_i \left[x_{i,n} - \left(\sum_j x_{j,n}/K \right) \right]^2 / K. \quad (5)$$

where $x_{i,n}$ is the nth pixel in the ith image of the training set and K is the number of the training images. By choosing $\{x^{\mathrm{sent}}\}$ as the pixels with the highest variance, the PSRN is obviously improved [see the black diamonds and orange pentagons in Figs. 2(a) and 2(b)].

A more reasoned way is to select based on the entanglement of $|\Psi\rangle$, so that $\{x^{\mathrm{sent}}\}$ will minimize the uncertainty of $\{x^{[\mathrm{rest}]}\}$ from the probability distribution given by the Born machine. Knowing $\{x^{[\mathrm{sent}]}\}$, the (conditional) probability distribution of $\{x^{[\mathrm{rest}]}\}$ satisfies

$$P(\{x^{[\mathrm{rest}]}\}|\{x^{[\mathrm{sent}]}\}) = | \prod_{x_n \in \{x^{[\mathrm{rest}]}\}} \langle s(x_n)|\Phi\rangle|^2, \quad (6)$$

where $|s(x_n)\rangle$ stands for the state associated with the nth pixel x_n [see Eq. (1)], and $|\Phi\rangle$ satisfies Eq. (2). The task is to find the N_f pixels $\{x^{[\mathrm{sent}]}\}$ that minimize the Shannon entropy

$$S^{\mathrm{Shan}} = - \sum_{\{x^{[\mathrm{rest}]}\}} P(\{x^{[\mathrm{rest}]}\}|\{x^{[\mathrm{sent}]}\}) \ln P(\{x^{[\mathrm{rest}]}\}|\{x^{[\mathrm{sent}]}\}). \quad (7)$$

Aiming at this task, let us begin with a simpler question: Which pixel should be sent if Alice sends only one pixel? This can be determined by the single-site entanglement entropy (SEE) that (say for the nth qubit) is defined as

$$S_n^{\mathrm{ent}} = -\mathrm{Tr}\hat{\rho}_n \ln \hat{\rho}_n. \quad (8)$$

S_n^{ent} quantifies the information of the rest of the system that will be gained if one has the information of the nth qubit. Such a quantity has been utilized to safely reduce the number of pixels for efficient supervised TN machine learning [42]. With S_n^{ent}, Alice can choose the $\tilde{n}$th pixel with $\tilde{n} = \arg\max_n S_n^{\mathrm{ent}}$, so that Bob will gain as much information as possible from one sent pixel.

Based on the above scheme, we propose the following Markov sampling strategy to select $\{x^{[\mathrm{sent}]}\}$, dubbed as entanglement-ordered sampling protocol (EOSP).

(1) With an N-qubit state $|\Psi(N)\rangle$ (initialized as $|\Psi\rangle$), calculate the SEE S_n^{ent} of all qubits and find the qubit that has the maximal S_n^{ent}, i.e., $\tilde{n} = \arg\max_n S_n^{\mathrm{ent}}$.

(2) From the reduced density matrix of the $\tilde{n}$th qubit, $\hat{\rho}_{\tilde{n}}$, calculate its dominant eigenstate $|s_{\tilde{n}}\rangle$.

(3) Projecting the $\tilde{n}$th qubit of $|\Psi(N)\rangle$, a $(N-1)$-qubit state is obtained as $|\Psi(N-1)\rangle = \langle s_{\tilde{n}}|\Psi(N)\rangle/C$, with C a constant to normalize $|\Psi(N-1)\rangle$.

(4) If N_f qubits have been projected, record the positions of these qubits and transfer the pixels at these positions of the image to Bob. Note we have $|\Psi(N-N_f)\rangle = |\Phi\rangle$ [Eq. (2)]. Otherwise, go back to Step 1 with $|\Psi(N-1)\rangle$.

In short, EOSP selects the pixels in the order of entanglement (EO). A simple example that helps to understand the EOSP is provided in Appendix C. We shall emphasize that the order of projections in the EOSP is solely determined by the entanglement properties of the generative MPS and does not depend on the specific images to be sent.

To obtain better accuracy for reconstructing grayscale images (note the images in the MNIST dataset are grayscale), we generate the pixels $\{x^{[\text{rest}]}\}$ by locating the separable state with maximal probability, i.e.,

$$\{x^{[\text{rest}]}\} = \arg\max_{\{x\}} \left| \prod_n \langle s(x_n)|\Phi\rangle \right|^2, \tag{9}$$

where the product $\prod_n$ goes through $\{x^{[\text{rest}]}\}$. It means that each projective basis $|s(x_n)\rangle$ is the dominant eigenstate of the corresponding single-site reduced density matrix of $|\Phi\rangle$ [Eq. (3)]. We dub such a generation way as "post-selection," considering that post-selections [43] will be needed to reach the maximum in Eq. (9) when the generative MPS is prepared as a quantum state. Note for the classical simulations of TNCS, the post-selection scheme will not increase the cost too much compared with the one-shot scheme, as Bob can sample as many times as he wants from one $|\Psi\rangle$, or he may simply calculate the reduced density matrices and their dominant eigenvectors by classical computers from the MPS. Figures 2(c) and 2(d) show the results with post-selections. One can see that the PSNR's for all three selection ways (EO, VO, and RO) are significantly improved. With EO and post-selections, the image can be accurately and efficiently communicated to Bob (with PSNR $\simeq 20$) for $r \simeq 10\%$.

Combined with the classical compressed sensing, it was proposed to utilize the classical probabilistic models to significantly reduce the compression ratio [38,39]. For instance, Ref. [38] investigated the compressed sensing assisted by autoencoders on MNIST and calculated the mean-square error (MSE) defined as $\text{MSE} = \frac{\sum_n (x_n - y_n)^2}{784} = 10^{-\text{PSNR}/10}$. The autoencoder is formed by two parts: a four-layer fully-connected deep neural network as the so-called recognizer to encode the data to some latent variables and a neural network with the same architecture as the generator to decode the latent variables. The autoencoder is trained and tested on all images of the ten classes in the dataset simultaneously. The MSE per image ranges approximately as $\text{MSE} \simeq 0.06 \sim 0.01$ when taking the compressed ratio as $r \simeq 1\% \sim 10\%$ for all the samples in the testing dataset. For TNCS, we train the MPS by the images in a single class of the train set, in order to control the computational complexity and focus on the performances of different reconstruction ways given a well-trained generative MPS. We obtain $\text{MSE} \simeq 0.047 \sim 0.016$ for the samples in each class of the testing set. More numerical results are provided in the Supplemental Material [44].

IV. Q SPARSITY

A prerequisite for the conventional compressed sensing to work efficiently is the sparsity of the signals. For processing images, it is known that the signals are usually not sparse in the real space. Therefore, transformation (such as discrete cosine/wavelet transformation) is implemented to transform to another space in which the signals are sparse.

In TNCS, the prerequisite is that the quantum probability distribution, i.e., the state $|\Psi\rangle$, should be "sparse." Here, the sparsity is gained in a completely different way from the standard compressed sensing, which is by mapping the data to the higher-dimensional quantum Hilbert space. This is analog
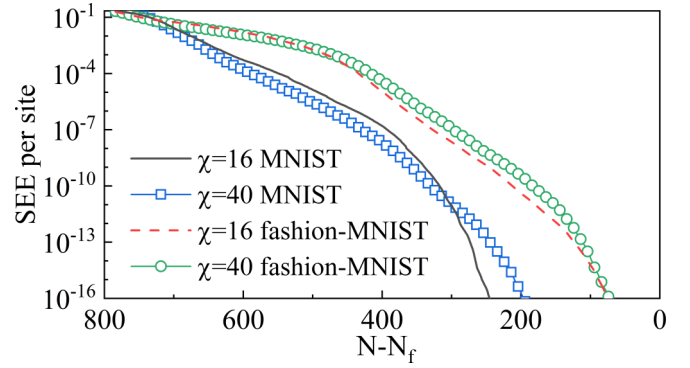

FIG. 3. The single-site entanglement entropy (SEE) [Eq. (8)] per site of $|\Phi\rangle$ in the entanglement-ordered sampling protocol (EOSP) versus the number of the unprojected qubits $N - N_f$. The more steeply the SEE per site decays, the faster the information of a quantum state can be gained by projections.

to the support vector machines [45] by mapping to a higher-dimensional space where the data can be better classified. In the unsupervised TN machine learning algorithm, each pixel x is mapped to the state of a qubit [Eq. (1)], then one image is mapped to the direct product state of N qubits with N the number of pixels. Such a vector is defined in a (2^N)-dimensional space $\mathcal{H}$. The MPS $|\Psi\rangle$ describes the joint probability distribution of the "vectorized" images in $\mathcal{H}$. Essentially, one still deals with the data in the real space. However, the probability distribution becomes sparse in this higher-dimensional space, since it can be well captured by an MPS. An MPS is sparse because such a representation can only reach a small corner of $\mathcal{H}$ that satisfies the so-called one-dimensional area law of entanglement entropy [46,47].

However, it is not easy to characterize the sparsity of a many-qubit state (including MPS) by simply treating it as a vector, as its dimension is exponentially large. We here propose to use EOSP to do so. In each step of EOSP, the qubit with the maximal SEE is projected. The entanglement of the state $|\Phi\rangle$ formed by the unprojected qubits decreases after each projection. Figure 3 shows the SEE per site $\bar{S}(\tilde{n}) = \sum_n S_n^{\text{ent}}(\tilde{n})/\tilde{n}$ of $|\Phi(\tilde{n})\rangle$ [see Eq. (8)] with a different number of unprojected qubits $\tilde{n} = N - N_f$. One can see that $\bar{S}(\tilde{n})$ decays rapidly as $\tilde{n}$ decreases, meaning the unprojected qubits are almost in a separable state for small $\tilde{n}$. For $\bar{S}(\tilde{n}) = 0$, no information will be gained by knowing the unprojected pixels. It means all information is contained in the projected pixels, and there is no uncertainty for the rest pixels, when $\bar{S}(\tilde{n})$ becomes zero.

From the implication of $\bar{S}(\tilde{n})$ discussed above, we define q sparsity to qualitatively describe the sparsity of a quantum state (including MPS) as

$$\mathcal{S}^q = \prod_{\tilde{n}=1}^N d^{\frac{S(\tilde{n})}{\ln d} - 1}, \tag{10}$$

with d the dimension of one vectorized pixel. In this work, we focus on qubits with $d = 2$. Q sparsity characterizes how fast the information of a quantum state can be extracted (or how fast the uncertainty of the rest can be reduced) by projections. Thus it characterizes the sparsity of quantum probability

distributions in an essentially different way from the sparsity of (classical) probability distributions, since quantum processes (e.g., measurements) are involved.

For any given state, the q sparsity satisfies

$$0 \leqslant \mathcal{S}^q \leqslant 1. \tag{11}$$

Obviously, we have $\mathcal{S}^q = 0$ for separable states. Considering the maximally-entangled state [48], we have $\mathcal{S}^q = 1$ since $\bar{S}(\tilde{n}) = \ln 2$ for any $\tilde{n}$. For what is in the middle, we take the N-qubit GHZ state as an example. We have $\bar{S}(N) = \ln 2$ originally and $\bar{S}(\tilde{n} \neq N) = 0$ after one projection. Therefore, we have $\mathcal{S}^q = 2^{-N+1}$. For the conventional k sparsity in comparison, we have $\mathcal{S}^k = 2/2^N = 2^{-N+1} = \mathcal{S}^q$ since it only has two nonzero coefficients in the 2^N-component vector. For the generative MPS's, we numerically have $\mathcal{S}^q = 2^{-768.6}$ and $2^{-765.6}$ with $\chi = 16$ for MNIST and fashion-MNIST, respectively, and $\mathcal{S}^q = 2^{-770.0}$ and $2^{-767.5}$ with $\chi = 40$.

For TNCS, $\mathcal{S}^q$ characterizes the efficiency, i.e., the compression ratio. The smaller $\mathcal{S}^q$ is, the faster $\bar{S}(\tilde{n})$ decays in general with the projections, and the less $\{x^{[\text{sent}]}\}$ Bob will require to accurately reconstruct the full information by TNCS. Therefore, analog to the conventional compressed sensing, TNCS requires the quantum probability distribution to be sparse in the higher-dimensional Hilbert space, i.e., $N + \log_2 \mathcal{S}^q \ll N$. Based on our results on MNIST and fashion-MNIST (Fig. 2), we have $N_f \simeq 10\% N$ (with $N = 784$) to reach PSNR $\simeq 20$ using EOSP and

$$N_f \simeq c(N + \log_2 \mathcal{S}^q), \tag{12}$$

with $c \simeq 6$ [the precision of c is taken up to $O(1)$ considering the fluctuations of different classes and datasets]. For other orderings (RO and VO), c will be larger (meaning that more pixels are required) to reach a comparable PSNR than the one that EO achieves.

V. SUMMARY AND PERSPECTIVE

In this work, we propose a novel compressed sensing approach by combining the ideas of compressed sensing, quantum communication, and unsupervised TN machine learning. The key step is to train the state $|\Psi\rangle$ (a Born machine) by the unsupervised TN machine learning algorithm, so that the targeted piece of information can be encoded in the separable state with the minimal distance to $|\Phi\rangle$ that is obtained by implementing projections on $|\Psi\rangle$ in a designed way. The q sparsity is proposed as a property of any quantum states and is used to estimate the efficiency of TNCS. We apply TNCS to the real-life datasets (handwritten digits and fashion images).

While the TNCS permits us to compress and communicate data with competitive efficiency and accuracy as a classical method, the scheme can be potentially generalized to quantum states on quantum setups. One may build and optimize the Born machine on quantum platforms to compress and transmit information based on quantum theories. In particular, one could use TN to design the quantum circuit as the Born machine [31] or directly realize the Born machine as a many-qubit state. The main difficulties are to implement TN models (or the many-qubit states) on quantum devices and also the measurements of entanglement, which are not yet feasible with nowadays technologies. But it is hopeful that quantum

devices will eventually surpass the classical counterparts in the future, with which the relevant obstacles would be tackled. Our work provides possibilities for processing real-life data by secure quantum communications (see more discussions in Appendix E).

APPENDIX A: UNSUPERVISED TENSOR-NETWORK MACHINE LEARNING ALGORITHM

In the generative TN machine learning algorithm proposed in Ref. [15], each image is mapped to a product state of N qubits as

$$|\phi_i\rangle = \prod_n |s(x_{i,n})\rangle, \tag{A1}$$

with $|s(x_{i,n})\rangle = \cos(x_{i,n}\pi/2)|0\rangle + \sin(x_{i,n}\pi/2)|1\rangle$ and N the total number of pixels in one image. Here, $x_{i,n}$ is the nth pixel (gray with $0 \leqslant x_{i,n} \leqslant 1$) of the ith image. The coefficients in the quantum state $|\Psi\rangle$ are optimized to minimize the negative log-likelihood (NLL) defined as

$$f = \ln |\langle \Psi | \Psi \rangle|^2 - \frac{\sum_i \ln |\langle \Psi | \phi_i \rangle|^2}{N}. \tag{A2}$$

The summation $\sum_i$ is over all the training images. NLL characterizes the resemblance between two probability distributions.

In this work, we choose the TN to be matrix product state (MPS). The coefficients of $|\Psi\rangle$ are in a special form satisfying

$$|\Psi\rangle = \sum_{\{a\}} \prod_n \sum_{s_n=0,1} A^{[n]}_{s_n a_n, a_{n+1}} |s_n\rangle. \tag{A3}$$

$A^{[n]}$ represents a tensor that corresponds to the nth pixel. The indexes $\{a\}$ are known as virtual bonds of the MPS; their dimensions are bounded by $\dim(a_n) \leqslant \chi$, with χ called virtual bond dimension. MPS is an efficient representation of

quantum-many-body states where the total number of parameter scales linearly with N as $\sim 2N\chi^2$. Note that the dimension of the Hilbert space actually scales exponentially as $\sim 2^N$. The tensors in the MPS are updated alternatively by the gradient method as $A^{[n]} \leftarrow A^{[n]} - \tau \partial f / \partial A^{[n]}$, with τ the gradient step; see Ref. [25] or [15] for more details.

After converging, $|\Psi\rangle$ gives the joint probability of the pixels. The probability for any image $\{x\}$ in $|\Psi\rangle$ is given as

$$P(\{x\}) = \left| \prod_n \langle s(x_n)|\Psi\rangle \right|^2. \tag{A4}$$

Note the probability is the square of the corresponding coefficient, thus such a TN state is also called the Born machine [35].

APPENDIX B: AMBIGUOUS CORRELATIONS OF INFORMATION IN TNCS

Another immediate question about TNCS is how to determine the samples (denoted by $\mathbb{A}$) for training the Born machine $|\Psi\rangle$, and what are the relations to the information (denoted by $\mathbb{B}$) that can be transferred or reconstructed through $|\Psi\rangle$. Obviously, we have $\mathbb{A} \subseteq \mathbb{B}$. The size of the complementary set $\mathbb{C} = \mathbb{B} - \mathbb{A}$ characterizes the generalization power of the Born machine.

Evidently, $\mathbb{C}$ has to be "ambiguously" correlated to $\mathbb{A}$ somehow. Let us consider an extreme situation, where all training samples in $\mathbb{A}$ are formed by uncorrelated random numbers. The trained state $|\Psi\rangle$ is an entangled state. However, such a state obviously cannot be used to effectively transfer a random image as no correlations exist between the random image and the state.

In this work, we choose $\mathbb{A}$ and $\mathbb{B}$ as the training and testing images of the same dataset, respectively. For instance, $\mathbb{A}$ and $\mathbb{B}$ are handwritten digits "3" or images of dresses. Although the "microscopic information" (pixels) of all the images in $\mathbb{A}$ and $\mathbb{B}$ are different from each other, a human being can recognize the "macroscopic information" of each image as a digit "3" (or a dress) without any problem. This suggests that $\mathbb{A}$ and $\mathbb{B}$ (thus $\mathbb{A}$ and $\mathbb{C}$) must be correlated somehow. In other words, we here ensure the existence of the "ambiguous" correlations between $\mathbb{A}$ and $\mathbb{B}$ by the "macroscopic" information.

With the TN machine learning, we can define the "ambiguous" correlation in a relatively more rigorous way: $\mathbb{A}$ and $\mathbb{B}$ are "ambiguously" correlated if the Born machine trained by $\mathbb{A}$ can accurately recognize the data in $\mathbb{B}$. For instance, one may train two Born machines by the "3" and "4" images in the training set, respectively, and construct a classifier that accurately recognizes "3" and "4" images [49]. To classify an image in $\mathbb{B}$ or the testing set, one compares the probability of have this image in the two Born machines, and classification is given by finding the largest probability.

The above recognition scheme can give us much useful information. For instance, the Born machine trained by the "3" images can be used to implement the TNCS for an image "3" written by the reader, as long as it can be recognized by the Born machine. Obviously, the TNCS cannot be implemented by the Born machine of "3" if the reader writes a "4." How

to more rigorously characterize and quantify such ambiguous correlations is an important issue to TNCS. One direction is to develop more universal classifiers for pattern recognition (not limited to digits or some certain kind of data). This will also be helpful to further understand and model the recognition process.

APPENDIX C: A SIMPLE EXAMPLE TO UNDERSTAND ENTANGLEMENT-ORDERED SAMPLING PROTOCOL

To explain why the entanglement-ordered sampling protocol (EOSP) works, let us consider the following four-qubit state as an example,

$$\begin{aligned} |\Psi\rangle &= \left(\frac{\sqrt{2}}{2}|01\rangle + \frac{\sqrt{2}}{2}|10\rangle \right) \otimes \left(\frac{1}{2}|01\rangle + \frac{\sqrt{3}}{2}|10\rangle \right) \\ &= \frac{\sqrt{2}}{4}|0101\rangle + \frac{\sqrt{6}}{4}|0110\rangle + \frac{\sqrt{2}}{4}|1001\rangle + \frac{\sqrt{6}}{4}|1010\rangle. \end{aligned} \tag{C1}$$

Such a state can describe a dataset of four images (0, 1, 0, 1), (0, 1, 1, 0), (1, 0, 0, 1), and (1, 0, 1, 0), with the probability $P = 1/8, 3/8, 1/8$, and $3/8$, respectively.

If Alice wants to send two pixels and encode the remaining two in the state, the pixel that Alice should first choose is obviously the first (or the second) pixel. Since the first two qubits are in the maximally entangled state, one of the pixels can be determined by knowing the other pixel. The second pixel Alice chooses should be the third or the fourth one. These two qubits are entangled (but not maximally), thus knowing one of them will gain certain (but not the full) information of the other. In all, Alice should send the first (or second) and the third (or the fourth) pixels to Bob.

The EOSP gives the same answer. The SEE of $|\psi\rangle$ satisfies $S_1^{\text{ent}} = S_2^{\text{ent}} = \ln 2 \simeq 0.693$, and $S_3^{\text{ent}} = S_4^{\text{ent}} = -\frac{1}{4}\ln\frac{1}{4} - \frac{3}{4}\ln\frac{3}{4} \simeq 0.562$. In step 1 of the EOSP, Alice chooses the first or the second pixel. The reduced density matrices satisfy $\hat{\rho}_1 = \hat{\rho}_2 = I/2$, with I the 2×2 identity. Therefore, Alice decides to measure the first qubit by $|0\rangle\langle 0|$ or $|1\rangle\langle 1|$. In either case, the resulting three-qubit state will be $|\Psi(3)\rangle = |x\rangle \otimes (\frac{1}{2}|01\rangle + \frac{\sqrt{3}}{2}|10\rangle)$ with $x = 0$ or 1. In the second iteration, Alice has $\tilde{S}_2^{\text{ent}} = 0$ and $S_3^{\text{ent}} = S_4^{\text{ent}} \simeq 0.562$, thus she decides to send the third (or forth) pixel. In comparison, Alice will choose to send the first and second pixels according to the variance, which is not a good idea since Bob will not be able to gain any information about the third and fourth pixels. Again, we would like to emphasize that this example is to help understand EOSP; it is too simple to draw any general conclusions about the advantages/disadvantages of quantum methods over classical ones.

APPENDIX D: QUANTUM NATURE IN TNCS

With $N_f = \#\{x^{\text{sent}}\} = 0$, Bob will randomly generate an image according to the probability distribution give by $|\Psi\rangle$. If the post-selections are used, the result will approach the separable state that has the minimal distance to $|\Psi\rangle$. This separable state gives the image that has the maximal probability in the probability distribution. We dub such an image
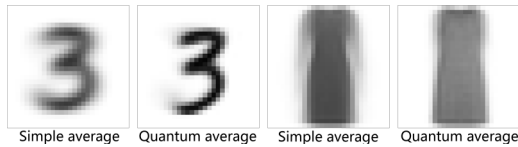
FIG. 4. The images by taking simple average of each pixel and by taking the quantum average (generated by MPS with no known pixel).

from no known pixel as the quantum average. One $|\Psi\rangle$ gives one unique quantum average (we assume that all $\hat{\rho}_n$'s have nondegenerated eigenvalues). As shown in Fig. 4, the quantum average is different from the simple average $\bar{x}_n = \sum_i x_{i,n}/K$ where no correlations are considered. Correlations (and entanglement) are considered in the quantum average when calculating the reduced density matrix.

Figure 5 shows which pixels are selected in EO and VO with different values of N_f. To illustrate the orders, we mark a pixel redder than those pixels that are behind this pixel in the order. Both EO and VO manage to capture the general shapes. Particularly, the "checkerboard" pattern appears in EO with relatively large N_f. This brings higher efficiency for the following reason. Since each two nearest-neighbor pixels should possess a strong correlation, the corresponding qubits are expected in a highly entangled state. It means that one only needs to know the information of one qubit (pixel) to access the information of the other qubit (pixel). Taking the maximally entangled two-qubit state $|01\rangle + |10\rangle$ as an example, if one knows that the first qubit is in the state $|0\rangle$ (or $|1\rangle$), meaning that the first pixel $x_1 = 0$ (or $x_1 = 1$), one will know that the second qubit is in the state $|1\rangle$ (or $|0\rangle$), meaning that the second pixel $x_2 = 1$ (or $x_2 = 0$). In this case, one only needs to send the information of one of the pixels, and the rest will be obtained from the state.

Intuitively, both the quantum entanglement and the (classical) variance measure the amount of the carried information. For instance, considering a pixel (labeled as n) that is always black in all the training images, such a pixel obviously carries no information, and we have $S_n = V_n = 0$. On the other hand, if a pixel changes dramatically with the training images, not necessarily but normally, this pixel may contain more
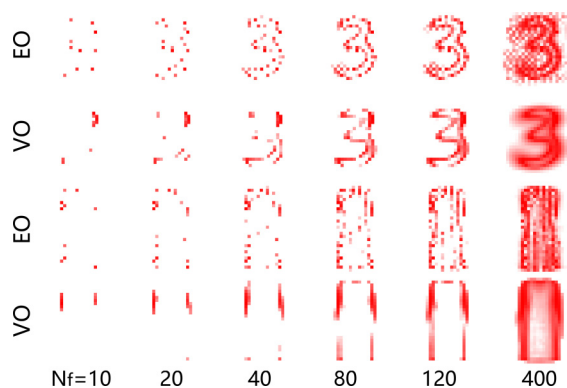

FIG. 5. Which N_f pixels are selected in EO and VO. To illustrate the orders by color, we mark a pixel redder than those behind this pixel in the order.

information, and we will have large S_n and V_n. One essential difference is that S_n and V_n are properties from the quantum state and the classical data, respectively. In our case, the quantum quantity (EO) outperforms the classical one (VO), providing evidence of the quantum advantage in the TNCS.

However, we cannot state here any general quantum advantages over classical information with these two specific methods. As we stated before, EO considers certain nonlocal properties while VO is purely local. Nevertheless, TNCS indeed provides a new path to investigate quantum advantages over classical information techniques. Several important and interesting questions are to be investigated, such as how to define new (classical or quantum) quantities that better suppress the compression ratio and/or increase the accuracy. Possible choices include the (classical) co-variance of the training data, the (quantum) correlation functions from $|\Psi\rangle$, and the multipartite entanglement. The performance of both quantum and classical methods for selecting $\{x^{\text{sent}}\}$ need to be pushed to their limits to discuss more clearly about the possible quantum advantages.

APPENDIX E: SECURITY OF TNCS WITH QUANTUM MANY-BODY STATES

In the scenario depicted above, TNCS can be potentially used to securely send information via quantum many-body states. The information is secure under the assumption that $|\Psi\rangle$ cannot be intercepted or replaced without letting Alice or Bob know, and those without $|\Psi\rangle$ cannot reconstruct the full information solely from $N_f \ll N$ pixels.

With a subset of the pixels of the image sent by Alice, Bob will need many copies of the state (the generative model) to project the corresponding qubits onto a product state according to the values of the sent pixels. An attacker Eve may just intercept the entire communication to get the information intended for Bob. Eve may also be able to hide her presence by sending Bob the same pixels she intercepted and certain new many-body states, e.g., the product states, which encode the image she received. As we argue below, this interception attack of Eve can be detected if the sent states are highly entangled.

Let us remind the readers that security in the "standard" quantum cryptographic protocols relies on shared nonlocality in the Ekert's scheme [50] or on the entanglement of the effective shared two-party density matrix (cf. [51,52]) in the prepare-and-measure schemes like the Bennett-Brassar one [53]. Since in the present situation, Alice sends to Bob many copies of a quantum many-body state, she may restrict herself to sending a (moderately or highly) entangled state. Our scheme belongs to such a case according to the entanglement properties of the generative models shown in Fig. 3. Moreover, for Bob, who has the access to many copies of this state, he can measure and detect this entanglement using many entanglement witnesses [54], for instance the ones aimed at measuring the many-body Bell correlations with pair correlators (cf. [55–59]). The witness operators for these protocols might be even announced publicly.

If Eve intercepts the communication, she needs to send to Bob the states having the same entanglement properties: not too weak, but also not too strong. Thus she has to measure

what she needs to send on several, say initial copies, and replace them by something completely unrelated. Bob by performing the same measurement on these unrelated copies should be able to detect interception. Of course, this way of assuring security is not easy or cheap, but in a sense it consists of using necessary methods to characterize the sent state, which might be useful and even necessary for its application for other quantum protocols (simulation, computation, metrology, etc.).

In this scheme, there is a possible leak of the information from $\{x^{[\text{sent}]}\}$ that is sent by classical channels. Regarding this issue, there are many ways to enhance the security classically. For example, Alice can introduce a one-to-one (reversible) deterministic map $\{y^{[\text{sent}]}\} = F(\{x^{[\text{sent}]}\}; \{x^{[\text{rest}]}\})$ to encrypt $\{x^{[\text{sent}]}\}$. Without F, the $\{x^{[\text{sent}]}\}$, which might be unsafe, could contain critical information (see for example Fig. 5, which are almost meaningful images for $N_f > 40$). The purpose of F is to avoid containing any meaningful information in $\{x^{[\text{sent}]}\}$.

Such an F-encrypted TNCS will contain the following steps: (1) Alice designs the function F, and trains $|\Psi\rangle$ by the images formed by $\{x^{[\text{rest}]}\}$ and $\{y^{[\text{sent}]}\}$; (2) Alice sends $|\Psi\rangle$ to Bob; (3) For the information to be sent, Alice sends $\{y^{[\text{sent}]}\} = F(\{x^{[\text{sent}]}\}; \{x^{[\text{rest}]}\})$ and the function F to Bob through classical channels that may not be safe; (4) Bob obtains $\{x^{[\text{rest}]}\}$ by $|\Psi\rangle$ and $\{y^{[\text{sent}]}\}$ (same to the standard TNCS), and obtains $\{x^{[\text{sent}]}\}$ by $\{y^{[\text{sent}]}\}$, $\{x^{[\text{rest}]}\}$, and the inverse of F. Then Bob will have the full information $\{x^{[\text{sent}]}\} + \{x^{[\text{rest}]}\}$. The information will be safe since those without $|\Psi\rangle$ cannot have $\{x^{[\text{rest}]}\}$, thus cannot obtain $\{x^{[\text{rest}]}\}$) even if they have F and $\{y^{[\text{sent}]}\}$.

Since the information to be sent is not restricted to the data that train $|\Psi\rangle$, Alice can provide previously the copies of $|\Psi\rangle$ to multiple parties and send any piece of "ambiguously" correlated information to each party anytime afterwards. Different pieces of information can be sent via the copies of the same state.

Meanwhile, Alice does not allow other parties to access the coefficients of $|\Psi\rangle$, to guarantee herself as the only provider of the state. One potential risk is that Alice provides too many copies of $|\Psi\rangle$ to others, with which the coefficients of $|\Psi\rangle$ can be cracked by, e.g., quantum state tomography [60]. In our case, this risk should be low since N is large. Specifically, as the state is designed in the form of MPS, Alice should avoid sending out polynomially many states (or more) [61].

In the scenario discussed above, Alice sends a small part of the classical information $\{x^{[\text{sent}]}\}$ and the whole state $|\Psi\rangle$ to Bob. Bob then generates the missing information $\{x^{[\text{rest}]}\}$ from $|\Psi\rangle$ and $\{x^{[\text{sent}]}\}$. In this scenario, one does not need to stabilize the entanglement between the qubits that are far separated.

This process can be replaced by a more standard quantum communication scheme. First, Alice trains and prepares $|\Psi\rangle$. Then she sends the qubits corresponding to $\{x^{[\text{rest}]}\}$ to Bob and keeps those corresponding to $\{x^{[\text{sent}]}\}$ to herself. Note that these qubits of $\{x^{[\text{sent}]}\}$ and $\{x^{[\text{rest}]}\}$ form the whole entangled state $|\Psi\rangle$. To send the information, Alice projects her qubits according to $\{x^{[\text{sent}]}\}$. Afterwards, Bob generates the $\{x^{[\text{rest}]}\}$ from his qubits.

In this scenario, Alice only gives a part of the qubits in $|\Psi\rangle$ to Bob or other receivers and does not need to transfer the information of $\{x^{[\text{sent}]}\}$ through a classical channel. It avoids the risks in communicating $\{x^{[\text{sent}]}\}$ classically. The disadvantage is that the qubits with Alice and the receivers need to be kept entangled in certain distance until Alice implements measurements on her qubits. The discussions about security given above also apply to this scheme.

APPENDIX F: EFFICIENCY OF TNCS WITH QUANTUM MANY-BODY STATES

It is not obvious whether the TNCS with quantum many-body states could exhibit any advantages over the one using classical computers from the perspective of efficiency. Let us try to give some preliminary discussions by considering the total number of qubits in the copies that Bob needs to reconstruct the information by measurements and the number of bits to send the state in the form of MPS classically. For the quantum-state scheme, we consider that Bob uses one-shot measurement for generating $\{x^{\text{rest}}\}$ for simplicity. Therefore, the number of the copies of the state $|\Psi\rangle$ that Bob needs is determined by how efficiently $|\Psi\rangle$ can be projected to the targeted state $|\Phi\rangle$ [Eq. (2)].

The probability of projecting $|\Psi\rangle$ to $|\Phi\rangle$ can be estimated as

$$P = 1 - \langle\tilde{\Phi}|\tilde{\Phi}\rangle, \tag{F1}$$

with $|\tilde{\Phi}\rangle = \prod_{x_n \in \{x^{[\text{sent}]}\}} \langle s(x_n)|\Psi\rangle$ the measured "state" without normalization. If Bob has M copies of $|\Psi\rangle$, the probability of obtaining at least one $|\Phi\rangle$ satisfies

$$\bar{P}(M) = 1 - (1 - P)^M. \tag{F2}$$

To further simplify our discussion on estimating P, let us consider the following "ideal" case. We assume that the images are binary, meaning each pixel can only take the value 0 (black) or 1 (white). Then one can see that the product states $\{|\phi_i\rangle\}$ [Eq. (A1)] are orthonormal to each other. We also suppose that $|\Psi\rangle$ that gives the minimum of the loss function [Eq. (A2)] can be reached. In this case, we consider to send any images in the training set and have $|\langle\Psi|\phi_i\rangle|^2 = 1/N_i$ with N_i the total number of the training images. In other words, the probabilities of the images are equally distributed at the minimal point. Due to the orthonormal properties of $\{|\phi_i\rangle\}$, we have $\langle\tilde{\Phi}|\tilde{\Phi}\rangle = |\prod_{x_n \in \{x^{[\text{sent}]}\}} \langle s(x_n)|\Psi\rangle|^2 \simeq 1/N_i$ when the $\{x^{[\text{sent}]}\}$ of any two images are not identical. In other words, $|\Psi\rangle$ will be projected to any one of $\{|\phi_i\rangle\}$ with an equal probability. We have $P = 1 - 1/N_i$ and $\bar{P}(M) = 1 - (1 - 1/N_i)^M$. Figure 6 demonstrates $\bar{P}(M)$ versus N_i and M. With $\bar{P}(M)$, one can estimate the number of needed copies (M) to obtain at least one expected projection with a given probability. The total number of needed qubits will be $N_{\text{qubit}} = MN$ with N the number of pixels in one image, which is the number of qubits in one copy. The bits to send an MPS classically is estimated as $N_{\text{bit}}^{\text{MPS}} = 2kN\chi^2$ with $k = 64$ if one takes the double-float precision for each element of the tensors in the MPS. Compared with $N_{\text{bit}}^{\text{MPS}}$ with $\chi = 40$ as an example, we have $M \sim O(10^5)$ so that $N_{\text{qubit}} \simeq N_{\text{bit}}^{\text{MPS}}$. With this number of copies and supposing $N_i \sim O(10^3)$, the probability of having at least one expected projected state is high.

We shall stress that the case with the fashion-MNIST dataset (and most other real-life cases) is much more complicated than the ideal one discussed above. The images are grayscale (or colorful), where the number of values each
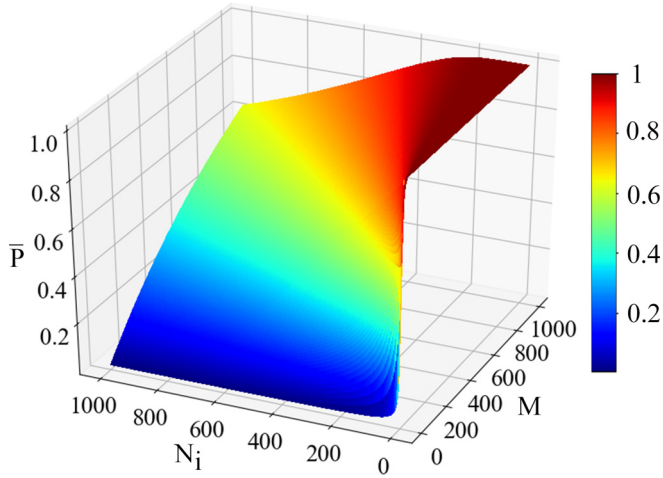
FIG. 6. The estimated probability $\bar{P} = 1 - (1 - 1/N_i)^M$ of having at least one expected projected state $|\Phi\rangle$ [Eq. (2)] given M copies of the state $|\Psi\rangle$. N_i is the number of images. See the details in Appendix F.

pixel can take is much larger than 2. The images to be sent are normally not restrained to the training set. Consequently,

$|\langle\Psi|\phi_i\rangle|^2$ will probably suffer from large fluctuations for, e.g., different images (indexed by i) and different datasets. Meanwhile, it is unclear if any tricks (e.g., state tomography) can be used to reduce the number of needed copies. Therefore, systematic investigations and thorough simulations are needed to draw any conclusions, which we leave for our future study.

APPENDIX G: IMAGES GENERATED BY TNCS

Figure 7 provides some reconstructed images with different numbers N_f of $\{x^{\mathrm{sent}}\}$ to give some intuitive impressions of how the accuracy increases with N_f. The pixels $\{x^{\mathrm{sent}}\}$ are picked in three different orders (EO, RO, and VO). Take the reconstruction of a dress image as an example (last three rows in Fig. 7). The quantum average ($N_f = 0$) is quite different from the image to be sent. With only $N_f \simeq 5$ known pixels picked by EO, the sleeves emerge. In contrast, the sleeves appear until 50 pixels are known if they are picked randomly. For the VO, the sleeves also emerge with 5 pixels but in a bad shape. The shape of the sleeves is reconstructed with $N_f \simeq 20$ in VO to a similar quality as $N_f \simeq 5$ in EO. The length of the sleeves is corrected with $N_f \simeq 50$ for EO and $N_f \simeq 110$ for RO and VO.
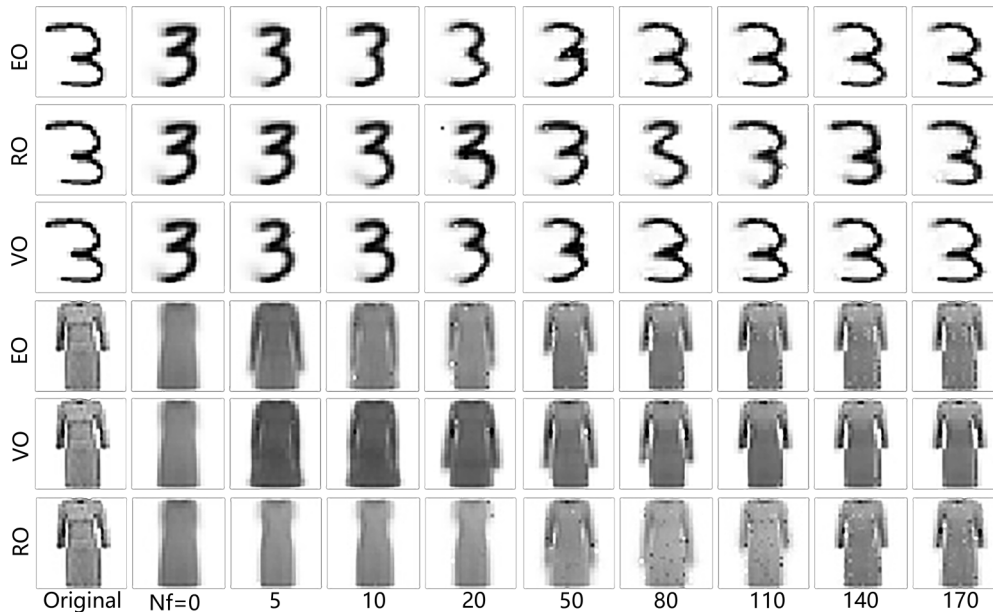

FIG. 7. Examples of the original and generated images in MNIST and fashion-MNIST in the entanglement order (EO), random order (RO), and variance order (VO). The number of known features N_f varies from 0 to 170, while the total number of features in an image is 784. We take the bond dimension of the generative MPS as $\chi = 40$.

[1] C. H. Bennett and S. J. Wiesner, Communication via One- and Two-Particle Operators on Einstein-Podolsky-Rosen States, Phys. Rev. Lett. **69**, 2881 (1992).

[2] X. S. Liu, G. L. Long, D. M. Tong, and F. Li, General scheme for superdense coding between multiparties, Phys. Rev. A **65**, 022304 (2002).

[3] H.-j. Lee, D. Ahn, and S. W. Hwang, Dense coding in entangled states, Phys. Rev. A **66**, 024304 (2002).

[4] D. Bruß, G. M. D'Ariano, M. Lewenstein, C. Macchiavello, A. Sen(De), and U. Sen, Distributed Quantum Dense Coding, Phys. Rev. Lett. **93**, 210501 (2004).

[5] A. Harrow, P. Hayden, and D. Leung, Superdense Coding of Quantum States, Phys. Rev. Lett. **92**, 187901 (2004).

[6] C. Wang, F.-G. Deng, Y.-S. Li, X.-S. Liu, and G. L. Long, Quantum secure direct communication with high-dimension quantum superdense coding, Phys. Rev. A **71**, 044305 (2005).

[7] A. K. Pati, P. Parashar, and P. Agrawal, Probabilistic superdense coding, Phys. Rev. A **72**, 012329 (2005).

[8] D. L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory **52**, 1289 (2006).

[9] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications* (Cambridge University Press, New York, 2012).

[10] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum State Tomography via Compressed Sensing, Phys. Rev. Lett. **105**, 150401 (2010).

[11] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, New J. Phys. **14**, 095022 (2012).

[12] A. Smith, C. A. Riofrío, B. E. Anderson, H. Sosa-Martinez, I. H. Deutsch, and P. S. Jessen, Quantum state tomography by continuous measurement and compressed sensing, Phys. Rev. A **87**, 030102(R) (2013).

[13] A. Kalev, R. L. Kosut, and I. H. Deutsch, Quantum tomography protocols with positivity are compressed sensing protocols, npj Quantum Inf. **1**, 15018 (2015).

[14] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, 2002).

[15] Z.-Y. Han, J. Wang, H. Fan, L. Wang, and P. Zhang, Unsupervised Generative Modeling using Matrix Product States, Phys. Rev. X **8**, 031012 (2018).

[16] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, Compressed sensing MRI, IEEE Signal Process. Mag. **25**, 72 (2008).

[17] C. Song, K. Xu, H. Li, Y.-R. Zhang, X. Zhang, W. Liu, Q. Guo, Z. Wang, W. Ren, J. Hao *et al.*, Generation of multicomponent atomic schrödinger cat states of up to 20 qubits, Science **365**, 574 (2019).

[18] A. Omran, H. Levine, A. Keesling, G. Semeghini, T. T. Wang, S. Ebadi, H. Bernien, A. S. Zibrov, H. Pichler, S. Choi, J. Cui, M. Rossignolo, P. Rembold, S. Montangero, T. Calarco, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Generation and manipulation of schrödinger cat states in rydberg atom arrays, Science **365**, 570 (2019).

[19] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[20] F. Verstraete, V. Murg, and J. Ignacio Cirac, Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems, Adv. Phys. **57**, 143 (2008).

[21] J. Ignacio Cirac and F. Verstraete, Renormalization and tensor product states in spin chains and lattices, J. Phys. A: Math. Theor. **42**, 504004 (2009).

[22] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, Ann. Phys. **349**, 117 (2014).

[23] J. Haegeman and F. Verstraete, Diagonalizing transfer matrices and matrix product operators: A medley of exact and computational methods, Annu. Rev. Condens. Matter Phys. **8**, 355 (2017).

[24] S.-J. Ran, E. Tirrito, C. Peng, X. Chen, L. Tagliacozzo, G. Su, and M. Lewenstein, *Tensor Network Contractions* (Springer, Cham, 2020).

[25] E. Stoudenmire and D. J. Schwab, Supervised learning with tensor networks, in *Advances in Neural Information Processing Systems 29*, edited by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., 2016), pp. 4799–4807.

[26] D. Liu, S.-J. Ran, P. Wittek, C. Peng, R. B. García, G. Su, and M. Lewenstein, Machine learning by unitary tensor network of hierarchical tree structure, New J. Phys. **21**, 073059 (2019).

[27] V. Pestun and Y. Vlassopoulos, Tensor network language model, arXiv:1710.10248.

[28] C. Guo, Z. Jie, W. Lu, and D. Poletti, Matrix product operators for sequence-to-sequence learning, Phys. Rev. E **98**, 042114 (2018).

[29] I. Glasser, N. Pancotti, and J. I. Cirac, From probabilistic graphical models to generalized tensor networks for supervised learning, IEEE Access **8**, 68169 (2020).

[30] E. M. Stoudenmire, Learning relevant features of data with multi-scale tensor networks, Quantum Sci. Technol. **3**, 034003 (2018).

[31] W. Huggins, P. Patil, B. Mitchell, K. B. Whaley, and E. M. Stoudenmire, Towards quantum machine learning with tensor networks, Quantum Sci. Technol. **4**, 024001 (2019).

[32] S. Cheng, L. Wang, T. Xiang, and P. Zhang, Tree tensor networks for generative modeling, Phys. Rev. B **99**, 155131 (2019).

[33] T. H. Johnson, J. D. Biamonte, S. R. Clark, and D. Jaksch, Solving search problems by strongly simulating quantum circuits, Sci. Rep. **3**, 1235 (2013).

[34] V. Giovannetti, S. Montangero, and R. Fazio, Quantum Multiscale Entanglement Renormalization Ansatz Channels, Phys. Rev. Lett. **101**, 180503 (2008).

[35] S. Cheng, J. Chen, and L. Wang, Information perspective to probabilistic modeling: Boltzmann machines versus born machines, Entropy **20**, 583 (2018).

[36] L. Deng, The mnist database of handwritten digit images for machine learning research [best of the web], IEEE Signal Process. Mag. **29**, 141 (2012).

[37] H. Xiao, K. Rasul, and R. Vollgraf, Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms, arXiv:1708.07747.

[38] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, Compressed sensing using generative models, in *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70 (ICML17, 2017), pp. 537–546.

[39] A. Grover and S. Ermon, Uncertainty autoencoders: Learning compressed representations via variational information maximization, The 22nd International Conference on Artificial Intelligence and Statistics, Proc. Machine Learning Res. **89**, 2514 (2019).

[40] D. Pérez-García, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, Quantum Inf. Comput. **7**, 401 (2007).

[41] F. J. Pineda, Generalization of Back-Propagation to Recurrent Neural Networks, Phys. Rev. Lett. **59**, 2229 (1987).

[42] Y. Liu, X. Zhang, M. Lewenstein, and S.-J. Ran, Entanglement-guided architectures of machine learning by quantum tensor network, arXiv:1803.09111.

[43] Y. Aharonov, D. Z. Albert, and L. Vaidman, How the Result of a Measurement of a Component of the Spin of a Spin-1/2 Particle can Turn Out to be 100, Phys. Rev. Lett. **60**, 1351 (1988).

[44] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.2.033293 for more data and the images reconstructed by TNCS.

[45] C. Cortes and V. Vapnik, Support-vector networks, Mach. Learn. **20**, 273 (1995).

[46] J. Eisert, M. Cramer, and M. B. Plenio, Colloquium: Area laws for the entanglement entropy, Rev. Mod. Phys. **82**, 277 (2010).

[47] A. J. Ferris, Area law and real-space renormalization, Phys. Rev. B **87**, 125139 (2013).

[48] N. Gisin and H. Bechmann-Pasquinucci, Bell inequality, bell states and maximally entangled states for n qubits, Phys. Lett. A **246**, 1 (1998).

[49] Z.-Z. Sun, C. Peng, D. Liu, S.-J. Ran, and G. Su, Generative tensor network classification model for supervised machine learning, Phys. Rev. B **101**, 075135 (2020).

[50] A. K. Ekert, Quantum Cryptography Based on Bell's Theorem, Phys. Rev. Lett. **67**, 661 (1991).

[51] M. Curty, M. Lewenstein, and N. Lütkenhaus, Entanglement as a Precondition for Secure Quantum Key Distribution, Phys. Rev. Lett. **92**, 217903 (2004).

[52] M. Curty, O. Gühne, M. Lewenstein, and N. Lütkenhaus, Detecting two-party quantum correlations in quantum-key-distribution protocols, Phys. Rev. A **71**, 022306 (2005).

[53] C. H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, Theor. Comput. Sci. **560**, 7 (2014), theoretical Aspects of Quantum Cryptography - celebrating 30 years of BB84.

[54] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[55] J. Tura, R. Augusiak, A. B. Sainz, T. Vértesi, M. Lewenstein, and A. Acín, Detecting nonlocality in many-body quantum states, Science **344**, 1256 (2014).

[56] F. Baccari, J. Tura, M. Fadel, A. Aloy, J.-D. Bancal, N. Sangouard, M. Lewenstein, A. Acín, and R. Augusiak, Bell correlation depth in many-body systems, Phys. Rev. A **100**, 022121 (2019).

[57] J. Tura, A. Aloy, F. Baccari, A. Acín, M. Lewenstein, and R. Augusiak, Optimization of device-independent witnesses of entanglement depth from two-body correlators, Phys. Rev. A **100**, 032307 (2019).

[58] A. Aloy, J. Tura, F. Baccari, A. Acín, M. Lewenstein, and R. Augusiak, Device-Independent Witnesses of Entanglement Depth from Two-Body Correlators, Phys. Rev. Lett. **123**, 100507 (2019).

[59] A. Piga, A. Aloy, M. Lewenstein, and I. Frérot, Bell Correlations at Ising Quantum Critical Points, Phys. Rev. Lett. **123**, 170604 (2019).

[60] K. Vogel and H. Risken, Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase, Phys. Rev. A **40**, 2847 (1989).

[61] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, Nat. Commun. **1**, 149 (2010).