

Experimental Simultaneous Learning of Multiple Non-Classical Correlations

Mu Yang,^{1,2} Chang-liang Ren,³ Yue-chi Ma,^{4,5} Ya Xiao,⁶ Xiang-Jun Ye,^{1,2} Lu-Lu Song,^{5,7}
Jin-Shi Xu,^{1,2,*} Man-Hong Yung,^{5,7,8,†} Chuan-Feng Li,^{1,2,‡} and Guang-Can Guo^{1,2}

¹CAS Key Laboratory of Quantum Information, University of Science
and Technology of China, Hefei 230026, People's Republic of China

²CAS Center For Excellence in Quantum Information and Quantum Physics,
University of Science and Technology of China, Hefei 230026, People's Republic of China

³Center for Nanofabrication and System Integration,
Chongqing Institute of Green and Intelligent Technology, 400714,
Chinese Academy of Sciences, Peoples Republic of China

⁴Center for Quantum Information, Institute for Interdisciplinary Information Sciences,
Tsinghua University, Beijing 100084, People's Republic of China

⁵Shenzhen Institute for Quantum Science and Engineering and Department of Physics,
Southern University of Science and Technology, Shenzhen 518055, China

⁶Department of Physics, Ocean University of China, Qingdao 266100, People's Republic of China

⁷Shenzhen Key Laboratory of Quantum Science and Engineering,
Southern University of Science and Technology, Shenzhen, 518055, China

⁸Central Research Institute, Huawei Technologies, Shenzhen, 518129, China

(Dated: November 19, 2018)

Non-classical correlations can be regarded as resources for quantum information processing. However, the classification problem of non-classical correlations for quantum states remains a challenge, even for finite-size systems. Although there exist a set of criteria for determining individual non-classical correlations, a unified framework that is capable of simultaneously classifying multiple correlations is still missing. In this work, we experimentally explored the possibility of applying machine-learning methods for simultaneously identifying non-classical correlations such as entanglement, non-locality, and quantum steering, using the same set of features. In particular, we constructed quantum-state classifiers from three different machine-learning approaches, including neural network, support vector machine, and decision tree. We found that for a family of quantum states, all three approaches can achieve high accuracy for the classification problem, but the run-times are very different. Overall, we found that support vector machine can significantly reduce the computing time with a mild loss of accuracy, compared with the neural network. Finally, we further demonstrated the scalability of this approach by analyzing the learnability of the quantum-state classifiers from labels generated by semi-definite programming.

Introduction.—In 1935, Einstein, Podolsky, and Rosen (EPR) [1] questioned the completeness of quantum mechanics (QM), as the theory seems to allow “spooky action at a distance” (known as EPR paradox). In quantum information science, much efforts have been devoted to achieve a deeper understanding of EPR’s paradox in terms of non-classical correlations, such as quantum entanglement [2], EPR steering [3], and Bell non-locality [4]. The various relationships among different non-classical correlations not only shape the foundation of the quantum theory, but also find themselves many interesting applications, ranging from quantum key distribution [5–9], communication complexity [10, 11], cloning of correlations [12, 13], quantum metrology [14], quantum state merging [15, 16], remote state preparation [17], and random number generation [18] etc.

The question is, how may one characterize the non-classical correlation for any given quantum state? To tackle such a problem, there are several challenges. (i) Even though various mathematical criteria and inequalities constraining non-classical correlations are known, for general multipartite states, the classification of entanglement, EPR steering, or Bell non-locality are generally

computationally-hard problems [19–21]. (ii) Many existing methods require the information of the whole density matrix; experimentally, a full quantum state tomography would be required. (iii) Each type of non-classical correlation has a different set of criteria; it is not known whether one can, simultaneously, classify all the non-classical correlations, based on the *same* set of measurements or observables.

On the other hand, machine learning represents a branch of artificial intelligence, aiming at producing a predictive function or a computer program based on a set of training data [22–25]. Beyond industrial applications, machine learning has created a profound impact on quantum information science, leading to a new research field called quantum machine learning, where many progresses have been achieved, including Hamiltonian learning [26], automated quantum experiments search [27], identification of phase transition [28], topological phase of matter [29] etc.

Recently, binary classification of quantum correlations have been achieved using the tools of machine learning, such as determination of nonlocality [30], and separability [31–33]. Our goal here is to take one step further

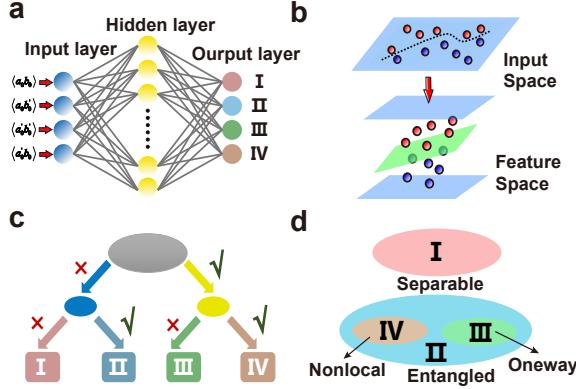


FIG. 1. Quantum correlation is divided into 4 categories by three typical ML models. **a.** The structures of artificial neural network (ANN). **b.** The support vector machine (SVM). **c.** The decision tree (DT). **d.** The Venn diagram of these four categories, in which I, II, III and IV represent the separable state, entangled state, one-way steerable state and Bell nonlocal state.

and explore if machine learning can be applied, in both theoretical and experimental setting, to *simultaneously* characterize multiple non-classical correlations with only partial information about the given quantum state.

In addition to binary classifiers, we experimentally constructed a *multi-label state classifier* for characterizing different classes of non-classical correlations through machine learning. Specifically, we compared three different multi-label state classifiers using three different machine-learning methods (see Fig. 1), including artificial neural network (ANN), support vector machine (SVM) and decision tree (DT), where each classifier only takes partial information for each member in a family of quantum states. The label on the state is determined by using the positive partial transpose (PPT) criteria [34], steering radius in two settings [35] and Clauser-Horne-Shimony-Holt (CHSH) inequalities [36], which are applied only to the training set but not the testing set. For classifying non-classical correlations, we found that the three models have different characteristics: (i) neural network provides the highest classification accuracy but consumes more computing resources; (ii) decision tree can save computing resources but has the lowest accuracy; (iii) Support vector machine can make use of fewer computing resources while maintaining high accuracy.

Labeling the training states.—For the purpose of demonstration, we shall focus on a family of quantum states, for which we can label unambiguously the class of non-classical correlations for each member. For feature extraction, we use partial information (four observables) of the quantum states, instead of the whole density matrix,

$$\langle a_0 b_0 \rangle, \langle a_0 b'_0 \rangle, \langle a'_0 b_0 \rangle, \langle a'_0 b'_0 \rangle, \quad (1)$$

to be the input, where $a_0 = \sigma_z$, $a'_0 = \sigma_x$, $b_0 = (\sigma_z - \sigma_x)/\sqrt{2}$, $b'_0 = (\sigma_z + \sigma_x)/\sqrt{2}$.

Specifically, the family of quantum states under investigation is of the following form, which has been applied for demonstrating one-way steering [37]:

$$\rho_{AB}(p, \theta) = p|\psi_\theta\rangle\langle\psi_\theta| + (1-p)I_A/2 \otimes \rho_B^\theta, \quad (2)$$

where $|\psi_\theta\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$, and $\rho_B^\theta = \text{tr}_A(|\psi_\theta\rangle\langle\psi_\theta|)$. The members in the two-parameter family are characterized by the combination of the two parameters $\theta \in (0, 2\pi)$, $p \in (0, 1)$.

The non-classical correlations for the testing state $\rho_{AB}(p, \theta)$ can be determined by the following rules: (i) for separability, e.g. through PPT criterion, one can show that whenever $p < 1/3$, the states are separable. Otherwise, they are entangled. (ii) for one-way steering, within the range, $1/\sqrt{2} < p < 1/\sqrt{1 + \sin^2 2\theta}$, the states are one-way steerable [38]. (iii) for non-locality, whenever $p > 1/\sqrt{1 + \sin^2 2\theta}$, the states become non-local [39].

However, in the experimental setting, the values of p and θ are not given; we just extracted a limited number of observables from the state. Each set of the features are associated with a label in the set I, II, III, IV, shown as FIG. 1d.

Experimental setting.—The experimental setup is shown in FIG. 2. Photon pairs entangled in the polarization basis $\{H, V\}$ are created through a type-II spontaneous parametric down conversion in a 20 mm-long periodical KTiOPO₄ (PPKTP) crystal, which is located on the Sagnac interferometer [40] and pumped by a 404 nm continuous-wave diode laser. In order to generate the entangled state, $\cos\theta|HH\rangle + \sin\theta|VV\rangle$, a half-wave plate (HWP1) is used to control the parameter θ . One of the two photons is sent to an unbalanced interferometer (UI) while the other photon is sent to Bob directly.

In the UI, the photon is separated into two paths by a beam splitter (BS). The state in one of the paths remains unchanged. Two sufficiently long calcite crystals (CCs) with HWP4 set to be 22.5° between them are placed on the other path to completely destroy the coherence between different components. Two variable filters (VFs) are used to manipulate the relative amplitude between these two paths. Combining these two paths into one, arbitrary two-qubit states $\rho_{AB}(p, \theta)$ can be prepared.

A quarter-wave plate (QWP), a HWP, and a polarization beam splitters (PBS) on both sides are used for quantum state tomography and four characteristic observables (features) measurement (partial information measurement). The labels can be deduced from the reconstructed density matrix $\rho_{AB}(p, \theta)$, by using PPT criteria, steering radius in two settings and CHSH inequalities. We implemented two independent experiments to obtain a training set and a test set respectively, which are used to train and test the machine learning models.

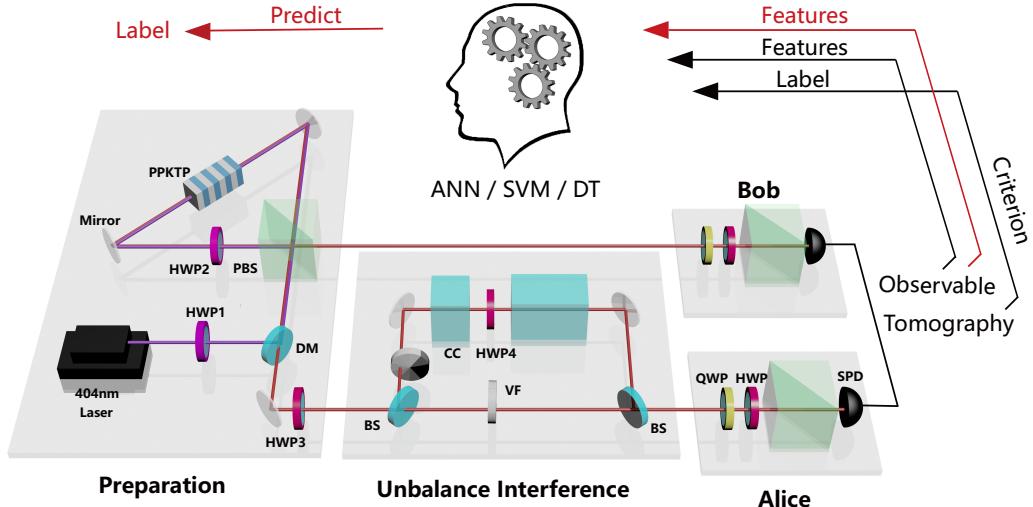


FIG. 2. Experimental setup. A pair of polarization-entangled photons are generated in the preparation stage by pumping a type-II PPKTP crystal located in a Sagnac interferometer with an ultraviolet laser at 404 nm. The polarization state of the pump light is rotated by a half-wave plate (HWP₁). The dual-wavelength HWP₂ set to be 45° in the Sagnac interferometer is used to exchange the polarizations of the pump light and down-converted photons. The HWP₃ set to be 45° is used to change the form of the entangled state. One of the photons passes through an unbalanced interferometer and is sent to Alice. HWP₄ is set at 22.5° and two sufficiently long calcite crystals (CCs) with the length of the last one being two times larger than the first one to completely destroy the coherence in different components. Two variable filters (VFs) are used to control the relative amplitudes between these two arms. The other photon is sent directly to Bob. Quarter-wave plates (QWPs), HWPs and polarization beam splitters (PBSs) on both sides of Alice and Bob are used for observables and tomography measurement. The photons are detected by single-photon detectors (SPDs), and the signals are sent for coincidence. The training and testing procedures are marked with black and red colors respectively. In the training case, criterion of quantum correlations are determined from tomographic data and the results are sent to the ML models together with the four characteristic features obtained from partial information measurement. In the testing case, only the four measured characteristic features are used to predict the labels.

The fidelities between the located states and experimental states are higher than 99%, where the experimentally constructed density matrix ρ_{exp} is mapped to a physical density matrix through minimizing the value of $Tr[\rho_{AB}(p, \theta) - \rho_{exp}]$. To train the input data, we prepared 445 states as the training set. For the trained models, we experimentally prepared a different set of 455 states for testing, which are shown in FIG. 3b-d in the (p, θ) space.

Experimental results.—The four characteristic features of the testing states are then sent to different ML models. First, we consider applying the three ML models to perform *binary* classification (YES/NO labels). We found that the accuracy can reach more than 97%, which is comparable to that of a recent related experiment [33]. For example, when determining whether a state is separable or not, the accuracy of the ML models can reach over 98%. When applied to one-way steering and Bell non-locality, the prediction accuracies can be over 97% and 99%, respectively. These results are shown in FIG. 3a.

Next, we consider applying the ML models to simultaneously label all four classes of states. The patterns of

the states trained through ANN, SVM and DT are shown in FIG. 3b, c and d. The accuracies are 93.93%, 93.25% and 91.68%, respectively. States in red and blue colors are labeled to be separable and entangled, in green color are labeled to be one-way steerable and in yellow color are labeled to be Bell non-local, respectively.

Gray lines in FIG. 3b-d represent the theoretical boundaries determined by the value of p and θ . We can see that the predicted errors mainly occur at the boundaries. Since the calculation of steering radius is extremely sensitive when θ is close to 0, $\pi/2$, π and $3\pi/2$, we omit these states in the test of the ability of the learning networks to reduce the errors induced by the physical criteria. A more detailed discussion is shown in SM [41].

Data analysis.—The accuracy of the ML methods depends on the settings of several parameters, which are further investigated (see FIG. 4). The variations of the loss function and accuracy as a function of Epochs (stage) for the ANN is shown in FIG. 4a. The implementation of ANN is based on the TensorFlow API, where the loss function is defined by the categorical cross-entropy. The training process is to minimize the loss function by

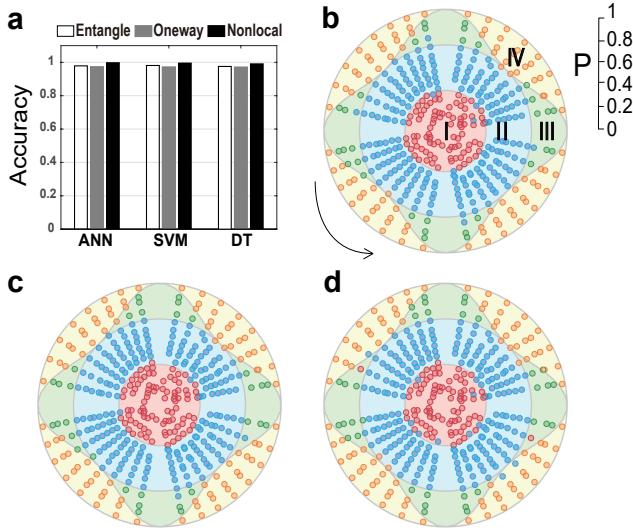


FIG. 3. a. The accuracies of ANN, SVM and DT, when distinguish whether a state is entangled, one-way steerable, or Bell nonlocal. The test results with ANN, SVM and DT are shown in b., c. and d. with the accuracies of 93.93%, 93.25% and 91.68%, respectively. The parameter θ varies from 0 to 2π while the visibility p changes from 0 to 1. The space is divided into four parts by the gray theoretical boundary, in which states in I, II, III and IV are theoretical predicted to be separable, entangled, one-way steerable and Bell nonlocal, respectively.

RMSprop [42]. The accuracy is define as the proportion of the number of samples that are correctly labeled by ML models. From the FIG. 4a, we found that the loss function decreases rapidly with the training accuracy growing quickly. An epoch is one complete presentation of the data set to be learned to a learning machine. In our work, we set max Epochs to 100. The value of loss function is 0.1500 and the accuracy is 93.55%. The change of Epochs will cause slight disturbance to prediction accuracy. The value of 93.93% is achieved when Epochs=100.

We further analyzed the impact of penalty factor of SVM and DT's maximum tree depth on the prediction accuracy, as shown in FIG. 4b. In the initial stage, with the increase of tree depth, the prediction ability becomes stronger for DT. When reaching the optimal depth 4, the DT prediction accuracy is 91.68%. Larger depth will cause overfitting. The initial properties of SVM are similar to that of DT, when penalty factor is greater than 25, the prediction accuracy will converge to 93.25%.

Scalability.—For systems with a higher dimension, e.g. a pair of qutrits, there is no known efficient way of labeling the training set. In this case, one may rely on numerical means, for example, semi-definite programming [43] (SDP) for labeling. In the following, we shall demonstrate how our ML model can also be applied to learn the labels generated by SDP, using ANN on a pair qutrits.

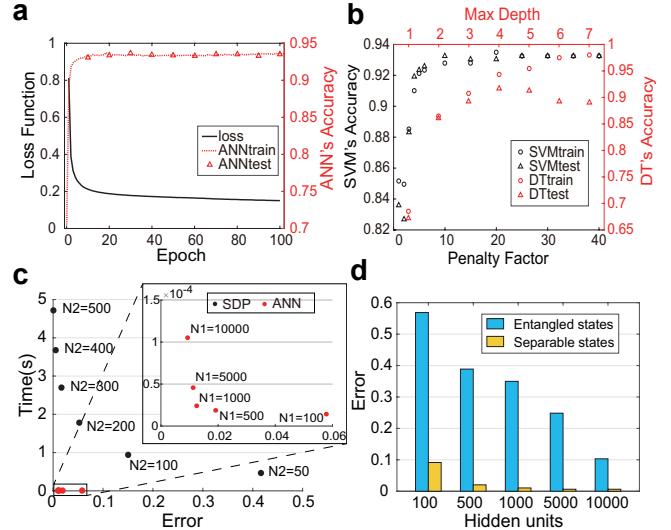


FIG. 4. The comparison between three ML methods. a. The performance of ANN versus the Epochs. The black curves represent value of the loss function and the red dotted line represents the training accuracy. Red triangles show the tested accuracy in different Epochs. b. The performance of DT and SVM versus max tree's depth and error penalty factor with the data maked in black and red colors, respectively. Circles and triangles represent the training and testing results respectively. c. The relationship among labeling time, error and iterations (N2) of SDP (black); The relationship among labeling time, error and hidden units (N1) of ANN (red). d. The variation of error along with the number of hidden units for distinguishing entangled/separable states.

In the following, we assume that SDP with 500 iterations for each data point (state) would be sufficient for separating the entangled states for a random set of states; For 100000 pairs of qubits, we found that the accuracy of SDP 500 can reach nearly 100% compared with the PPT criterion. Furthermore, we studied the error and running time in classifying 768280 (90% training + 10% testing) general 2-qutrit states labeled by SDP, which are equally divided between the entangle states and separable states. The result is shown in Fig. 4c (red points), where N1 is the number of hidden units. For comparison, we also plot the running time and error relative to SDP 500 for SDP with different iterations N2 as black points in Fig. 4c. These results show that our ML model can efficiently learn SDP 500 with a small error.

We also analyze the main error sources of ANN. Fig.4d illustrates the relationship between prediction error and hidden units for different types of 2-qutrit states (8000 entangle states + 12936 separable states). With the increase of hidden units, the entanglement/separable states error decrease. For the same number of hidden units, the ANN prediction error of entangled states is much larger than that of separable states. Therefore, the prediction error of neural network is mainly due to the inaccurate

prediction of entangled states.

Conclusion.—In this work, we experimentally explored the possibility of the classification problem of multiple quantum correlations by comparing three ML approaches, namely ANN, SVM, and DT. It is shown that, all three methods can be experimentally trained to efficiently learn and classify quantum states, without state tomography. Comparing these three ML methods, we concluded that the predictive power is ordered as follows: ANN \gtrapprox SVM > DT. However, the required training time also follows the same order ANN(43.7s) > SVM(1.2s) \gtrapprox DT(1.1s). Therefore, even though ANN can achieve the best performance, it also requires the longest training time. On the other hand, DT can be applied for classification problems with less suitable for classification without considerable precision for saving computation resources. To improve the accuracy of prediction, we can use ANN with more computation resources. While SVM is alternative when we care about the trade off between the computational resources and prediction accuracy.

The performances of predicting quantum correlation using only two characteristic features are further discussed in SM [41]. Although all the three ML methods are theoretically shown to be of high predicted accuracies, they are experimentally shown to be of lower predicted accuracies compared to the case with four characteristic features. It implies that the four features used here are of higher noise immunity. Furthermore, we discuss the influence of the data source of the training set [41]. It suggests that an effective classifier need to be trained with experimental states instead of the simulation produced by computer, illustrating the necessity of the experimental verification.

Acknowledgement.—This work was supported by the National Key Research and Development Program of China (Grant No. 2016YFA0302700), the National Natural Science Foundation of China (Grants No. 61725504, 61327901, 11774335 and 11821404), the Key Research Program of Frontier Sciences, Chinese Academy of Sciences (CAS) (Grant No. QYZDY-SSW-SLH003), Anhui Initiative in Quantum Information Technologies (AHY060300 and AHY020100), the Fundamental Research Funds for the Central Universities (Grant No. WK2470000020 and WK2470000026). C.-L. R. was supported by National key research and development program (No. 2017YFA0305200), the Youth Innovation Promotion Association (CAS) (Grant No. 2015317), the National Natural Science Foundation of China (Grant No. 11605205), the Natural Science Foundation of Chongqing (Grants No. cstc2015jcyjA00021 and No. cstc2018jcyjA2509), the Entrepreneurship and Innovation Support Program for Chongqing Overseas Returnees (Grants No. cx2017134 and No. cx2018040). M.-H. Y. was supported by the National Natural Science Foundation of China (11875160), the Guangdong Innovative and Entrepreneurial Research Team

Program (No. 2016ZT06D348), Natural Science Foundation of Guangdong Province (2017B030308003), and the Science, Technology and Innovation Commission of Shenzhen Municipality (JCYJ20170412152620376, JCYJ20170817105046702, ZDSYS201703031659262).

M. Y. and C.-L. R. contributed equally to this work.

* jsxu@ustc.edu.cn

† yung@sustc.edu.cn

‡ cfli@ustc.edu.cn

- [1] A. Einstein, B. Podolsky and N. Rosen, Phys. Rev. **47**, 777 (1935).
- [2] R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
- [3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
- [4] D. Cavalcanti and P. Skrzypczyk, Rep. Prog. Phys. **80**, 024001 (2017).
- [5] A. K. Ekert, Quantum cryptography based on Bells theorem, Phys. Rev. Lett. **67**, 661 (1991).
- [6] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, Rev. Mod. Phys. **74**, 145 (2002).
- [7] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).
- [8] C. Branciard, E. G. Cavalcanti, S. P. Walborn, V. Scarani, and H. M. Wiseman, Phys. Rev. A **85**, 010301 (2012).
- [9] S. Kocsis, M. J.W. Hall, A. J. Bennet, and G. J. Pryde, Nat. comm. **6**, 5586 (2015).
- [10] Č. Brukner, M. Žukowski, J. W. Pan, and A. Zeilinger, Phys. Rev. Lett. **92**, 127901 (2004).
- [11] M. Piani and J. Watrous, Phys. Rev. Lett. **114**, 060404 (2015).
- [12] M. Piani, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **100**, 090502 (2008).
- [13] S. Luo and W. Sun, Phys. Rev. A **82**, 012338 (2010).
- [14] K. Modi, H. Cable, M. Williamson, and V. Vedral, Phys. Rev. X **1**, 021022 (2011).
- [15] M. Horodecki, J. Oppenheim, and A. Winter, Nature (London) **436**, 673 (2005).
- [16] D. Cavalcanti, L. Aolita, S. Boixo, K. Modi, M. Piani, and A. Winter, Phys. Rev. A **83**, 032324 (2011).
- [17] B. Dakić, Y. O. Lipp, X. Ma, M. Ringbauer, S. Kropatschek, S. Barz, T. Paterek, V. Vedral, A. Zeilinger, Č. Brukner and P. Walther, Nat. Phys. **8**, 666 (2012).
- [18] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning and C. Monroe, Nature (London) **464**, 1021 (2010).
- [19] L. Gurvits, J. Comput. Syst. Sci. **69**, 448 (2004).
- [20] Y. C. Huang, New J. Phys. **16**, 033027 (2014).
- [21] J. Batle, C. R. Ooi, S. Abdalla, and A. Bagdasaryan, Quantum Inf. Process. **15**, 2649 (2016).
- [22] T. Mitchell, *Machine Learning* 1st edn (New York: McGraw-Hill) (1997).
- [23] L. Devroye, L. Györfi and G. Lugosi, *A Probabilistic Theory of Pattern Recognition* 1st edn (Berlin: Springer) (1996).

- [24] V. Vapnik, *Statistical Learning Theory* (New York: Wiley) (1998).
- [25] J. H. Friedman, T. Hastie and R. Tibshirani, *Elements of Statistical Learning: Data Mining, Inference and Prediction* (Berlin: Springer) (2003).
- [26] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, Phys. Rev. Lett. **112**, 190501 (2014).
- [27] M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz, and A. Zeilinger, Phys. Rev. Lett. **116**, 090405 (2016).
- [28] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, Nat. Phys. **12**, 469 (2016).
- [29] Y. Zhang and E. A. Kim, Phys. Rev. Lett. **118**, 216401 (2017).
- [30] D. L. Deng, X. Li and S. DasSarma, Phys. Rev. X **7**, 021021 (2017).
- [31] S. R. Lu, S. L. Huang, K. R. Li, J. Li, J. X. Chen, D. W. Lu, Z. F. Ji, Y. Shen, D. L. Zhou and B. Zeng, Phys. Rev. A **98**, 012315 (2018)
- [32] Y. C. Ma and M. H. Yung, Npj Quantum Inform. **4**, 34 (2018).
- [33] J. Gao, L. F. Qiao, Z. Q. Jiao, Y. C. Ma, C. Q. Hu, R. J. Ren, A. L. Yang, H. Tang, M. H. Yung and X. M. Jin, Phys. Rev. Lett. **120**, 240501, (2018).
- [34] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **283**, 1 (2001).
- [35] K. Sun, X. J. Ye, J. S. Xu, X. Y. Xu, J. S. Tang, Y. C. Wu, J. L. Chen, C. F. Li, and G. C. Guo, Phys. Rev. Lett. **116**, 160404 (2016)
- [36] J. F. Clauser, M. A. Horne, A. Shimony, R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969)
- [37] Y. Xiao, X. J. Ye, K. Sun, J. S. Xu, C. F. Li, and G. C. Guo, Phys. Rev. Lett. **118**, 140404 (2017)
- [38] J. Bowles, F. Hirsch, M. T. Quintino, and N. Brunner, Phys. Rev. A **93**, 022121 (2016).
- [39] H. Y. Su, C. L. Ren, J. L. Chen, F. L. Zhang, C. F. Wu, Z. P. Xu, M. L. Gu, and S. Vinjanampathy and L. C. Kwek, Phys. Rev. A **93**, 022110 (2016)
- [40] A. Fedrizzi, T. Herbst, A. Poppe, T. Jennewein, and A. Zeilinger, Opt. Express **15**, 15377 (2007).
- [41] see Supplemental Material for more details.
- [42] T. Tielemans, G. Hinton, COURSERA: Neural networks for machine learning. **4**, 26 (2012)
- [43] A. C. Doherty, P. A. Parrilo, F. M. Spedalieri, Phys. Rev. Lett. **88**, 187904 (2002).

Supplemental Material of “Experimental Simultaneous Learning of Multiple Non-Classical Correlations”

Mu Yang,^{1,2} Chang-liang Ren,³ Yue-chi Ma,^{4,5} Ya Xiao,⁶ Xiang-Jun Ye,^{1,2} Lu-Lu Song,^{5,7}
Jin-Shi Xu,^{1,2,*} Man-Hong Yung,^{5,7,8,†} Chuan-Feng Li,^{1,2,‡} and Guang-Can Guo^{1,2}

¹*CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, People’s Republic of China*

²*CAS Center For Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, People’s Republic of China*

³*Center for Nanofabrication and System Integration, Chongqing Institute of Green and Intelligent Technology, 400714, Chinese Academy of Sciences, Peoples Republic of China*

⁴*Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, People’s Republic of China*

⁵*Shenzhen Institute for Quantum Science and Engineering and Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China*

⁶*Department of Physics, Ocean University of China, Qingdao 266100, People’s Republic of China*

⁷*Shenzhen Key Laboratory of Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China*

⁸*Central Research Institute, Huawei Technologies, Shenzhen, 518129, China*

In this supplements, we provide and discuss more experimental results. In Sec. (I), we discuss the performance of predicting quantum correlation using less information. In Sec. (II), we show the performance of predicting experimental states by a well-trained multi-label classifier using simulated states. It efficiently proves the importance of using experimental states as training data but not the simulated state produced by computer. In Sec. (III), predicting results including states with θ close to 0, $\pi/2$, π and $3\pi/2$. was further discussed. In Sec. (IV), we introduce Semidefinite programs in detail.

I. THE PERFORMANCE OF PREDICTING QUANTUM CORRELATION USING LESS INFORMATION

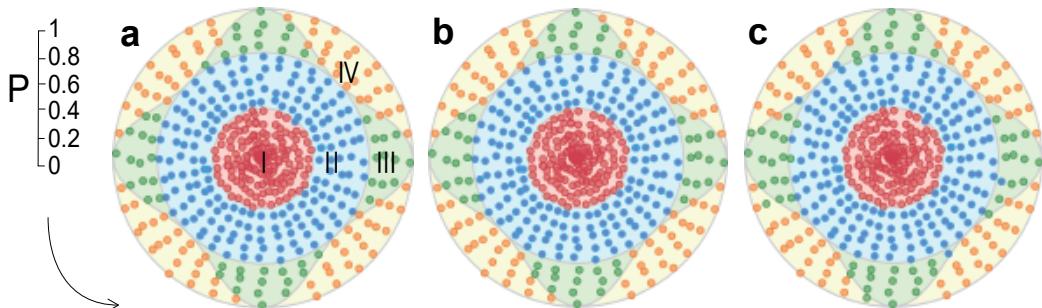


FIG. 1. Theoretical results for predicting states with less information. The test results with ANN, SVM and DT are shown in **a.**, **b.** and **c.** with the accuracies of 99.60%, 99.00% and 97.41%, respectively. States in red and blue colors are determined to be separable and entangled, in green color are determined to be one-way steerable and in yellow color are determined to be Bell nonlocal, respectively. The parameter θ varies from 0 to 2π while the visibility p changes from 0 to 1. The space is divided into four parts by the gray theoretical boundary, in which states in I, II, III and IV are theoretical predicted to be separable, entangled, one-way steerable and Bell nonlocal, respectively.

Here, we use only two of the mean value of observables,

$$\langle a_0 b'_0 \rangle, \langle a'_0 b_0 \rangle, \quad (1)$$

as the characteristic features to explore the performance of predicting quantum correlation using less information, where $a_0 = \sigma_z$, $a'_0 = \sigma_x$, $b_0 = (\sigma_z - \sigma_x)/\sqrt{2}$, $b'_0 = (\sigma_z + \sigma_x)/\sqrt{2}$. We generated 1004 different quantum states by

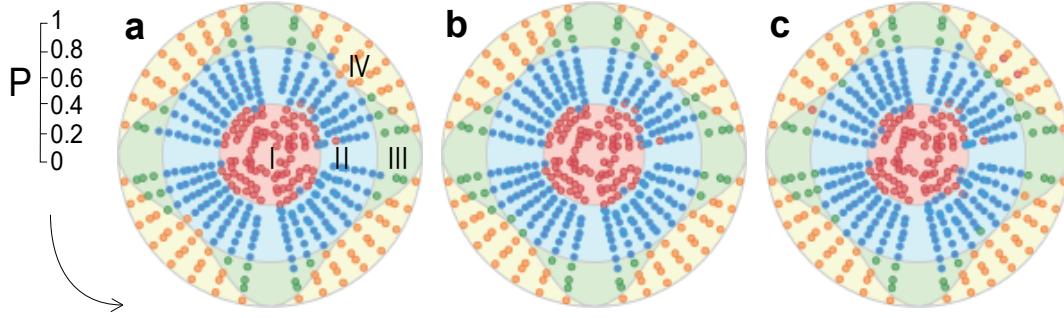


FIG. 2. Experimental results for predicting states with less information. The test results with ANN, SVM and DT are shown in **a.**, **b.** and **c.** with the accuracies of 92.35%, 92.58% and 88.00%, respectively. States in red and blue colors are determined to be separable and entangled, in green color are determined to be one-way steerable and in yellow color are determined to be Bell nonlocal, respectively. The parameter θ varies from 0 to 2π while the visibility p changes from 0 to 1. The space is divided into four parts by the gray theoretical boundary, in which states in I, II, III and IV are theoretically predicted to be separable, entangled, one-way steerable and Bell nonlocal, respectively.

computer with the form of

$$\rho_{AB}(p, \theta) = p|\psi(\theta)\rangle\langle\psi(\theta)| + (1-p)I_A/2 \otimes \rho_B^\theta. \quad (2)$$

Taking half of them as training set, where these states are labeled by different quantum criteria. Then we can construct three multiple quantum correlation classifiers by different machine learning (ML) methods. The rest half part is used as the test set, in which the distribution, is shown in FIG. 1a in the space of the parameters p and θ . Compared with the result determined by quantum correlation criteria from quantum state tomography data, the predicted accuracies of ANN, SVM and DT, are 99.60%, 99.00% and 97.41% respectively, as illustrated in FIG. 1b-d. This result shows that the machine learning models can work efficiently with only two features in theory.

However, when we use the experimental data to construct the multiple quantum correlation classifier by using only two features, the predictions are not as good as we expected, which are shown in FIG. 2. The accuracies are 92.35%, 92.58% and 88.00% of ANN, SVM and DT, which are less than that of ML models with four features. Therefore, using $\langle a_0 b_0 \rangle$, $\langle a_0 b'_0 \rangle$, $\langle a'_0 b_0 \rangle$, $\langle a'_0 b'_0 \rangle$ as input features is of higher noise immunity.

II. THE PERFORMANCE OF PREDICTING EXPERIMENTAL STATES BY A WELL-TRAINED MULTIPLE QUANTUM CORRELATION CLASSIFIER USING SIMULATED STATES

Here we want to clarify an interesting and important point: can we use a well-trained multiple quantum correlation classifier which is trained by using simulated states to efficiently predict the real experimental data? One may expect that these well-trained multiple quantum correlation classifier can efficiently avoid experimental noise during the training process, and it can also achieve a very high accuracy in the cross validation process. Hence, it should be very effective. But such intuition is incorrect. We generated 1004 different quantum states by computer and constructed three well-trained multiple quantum correlation classifiers with different ML methods. Then taking such quantum correlation classifiers to predict the real experimental data (FIG. 3a). The accuracies of ANN, SVM and DT are 78.43%, 73.48% and 83.37%, which are shown in FIG. 3b-d, respectively. The prediction error on the boundary is very large, and none of a one-way steerable state is predicted. The theoretically trained ML models are failures for experimental states. It illustrates that the sources of the training data and the test data heavily affect the efficiency of the classifiers. This also explains the necessity of our experimental demonstration.

III. PREDICTING RESULTS INCLUDING STATES WITH θ CLOSE TO 0, $\pi/2$, π AND $3\pi/2$.

When the parameter θ is close to 0, $\pi/2$, π , $3\pi/2$, the steering ellipsoid will become very small. So the criterion of steering radius will be extremely unstable. The result is shown in FIG. 4a, in which the one-way steerable states (green

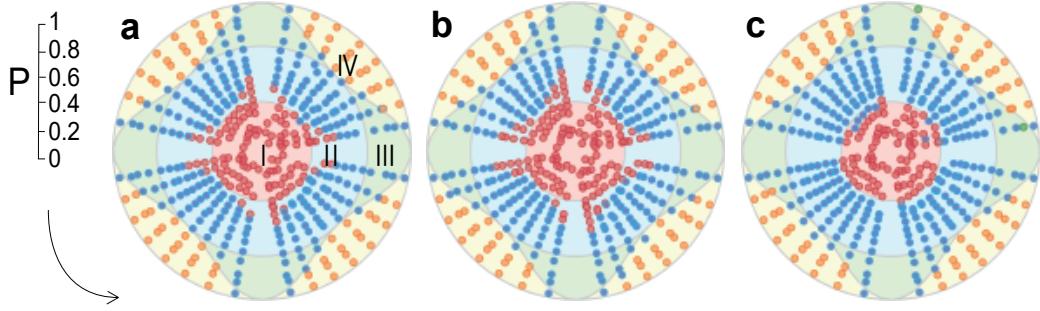


FIG. 3. The results of machine-learning models trained by numerical states. States in red and blue colors are determined to be separable and entangled, in green color are determined to be one-way steerable and in yellow color are determined to be Bell nonlocal, respectively. The test results with ANN, SVM and DT are shown in **a.**, **b.** and **c.** with the accuracies are 78.43%, 73.48% and 83.37%, respectively. The parameter θ varies from 0 to 2π while the visibility p changes from 0 to 1. The space is divided into four parts by the gray theoretical boundary, in which states in I, II, III and IV are theoretical predicted to be separable, entangled, one-way steerable and Bell nonlocal, respectively.

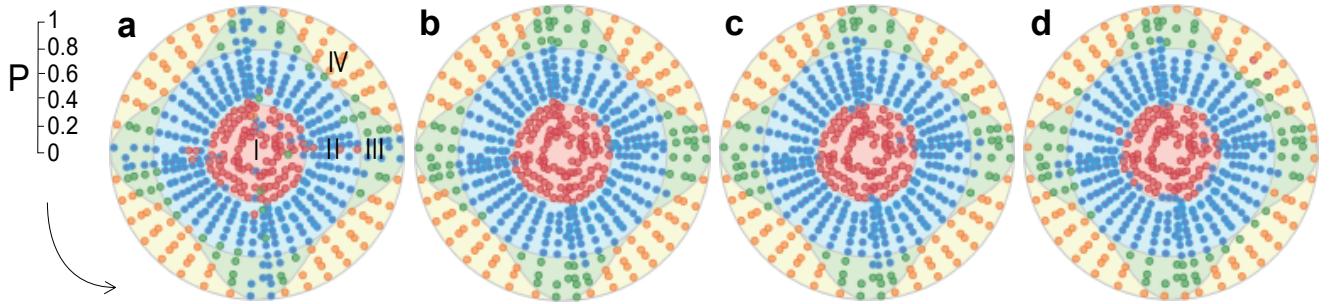


FIG. 4. The results including states with θ close to 0, $\pi/2$, π , $3\pi/2$. The parameter θ varies form 0 to 2π while the visibility p changes from 0 to 1. The space is divided into 4 parts by the gray boundary, in which I, II, III and IV represent the separable state, entangled state, one-way steerable state, and Bell nonlocal state, respectively. **a.** The quantum states are labeled by different quantum correlation criteria with tomography data. States in red and blue colors are determined to be separable and entangled, in green color are determined to be one-way steerable and in yellow color are determined to be Bell nonlocal, respectively. The quantum states are predicted by the ANN (**b.**), SVM (**c.**) and DT (**d.**).

color) determined from (p, θ) are found not to be one-way steerable (blue color) determined from the reconstructed density matrixes. Therefore ,we do not take into the consideration of the states when θ close to 0, $\pi/2$, π , $3\pi/2$.

In our work, we use the states without θ closed to 0, $\pi/2$, π , $3\pi/2$ as the training set and use the trained ML models to predict the states in whole p - θ space. It is to noted that our ML models correctly predicts the quantum states which are difficult to judge by quantum correlation criteria. The result is shown in FIG. 4b-d with the methods of ANN, SVM and DT.

IV. SEMIDEFINITE PROGRAMS FOR DISTINGUISHING SEPARABLE AND ENTANGLED STATES

Background

One of the most popular criteria for “witnessing” entangled states is based on taking partial transposition, introduced by Peres[1]. For any density matrix ρ , with the matrix elements defined by $\rho_{ik,jl} = \langle i| \otimes \langle k | \rho | j \rangle \otimes | l \rangle$, its partial transpose is defined as $\rho_{ik,jl}^{T_A} = \rho_{jk,il}$. Thus if the states ρ is separable, it must have a positive partial transpose. Furthermore, any state for which ρ^{T_A} is not positive semidefinite is necessarily entangled. Horodecki et al proved that PPT (positive partial transpose) is both necessary and sufficient for identifying separability for a pair of qubits

$\mathcal{H}_2 \otimes \mathcal{H}_2$ and qubit-qutrit systems $\mathcal{H}_2 \otimes \mathcal{H}_3$ [2]. However, for higher dimensional systems, there exist entangled states that cannot be identified by the PPT criterion [3].

Alternatively, due to the approach known as semidefinite program(SDP)[5] for convex optimization problem, one may theoretically and numerically [4] construct a separability criteria based on the properties of separable states.

A separable state ρ can be written as a convex combination of product states $\rho = \sum p_i |\psi_i\rangle\langle\psi_i| \otimes |\phi_i\rangle\langle\phi_i|$, where $|\psi_i\rangle$ and $|\phi_i\rangle$ are on the spaces \mathcal{H}_A and \mathcal{H}_B . Let $\tilde{\rho}$ defined on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_A$, is given by

$$\tilde{\rho} = \sum p_i |\psi_i\rangle\langle\psi_i| \otimes |\phi_i\rangle\langle\phi_i| \otimes |\psi_i\rangle\langle\psi_i| \quad (3)$$

Thus, the state $\tilde{\rho}$ satisfies following properties. First, the state $\tilde{\rho}$ is an extension of ρ (the partial trace over the third party is equal to ρ). Second, the state $\tilde{\rho}$ is symmetric under interchanging the two copies of \mathcal{H}_A . Finally, the extension $\tilde{\rho}$ is a tripartite separable state.

Thus, if a state ρ on $\mathcal{H}_A \otimes \mathcal{H}_B$ is separable, there exists an extension $\tilde{\rho}$ on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_A$ such that $\tilde{\rho}^{T_A} \geq 0$ and $\tilde{\rho}^{T_B} \geq 0$. We can also generalize the criterion to an arbitrary copies of \mathcal{H}_A and \mathcal{H}_B . That is, if a state ρ on $\mathcal{H}_A \otimes \mathcal{H}_B$ is separable, there is an extension $\tilde{\rho} = \sum p_i |\psi_i\rangle\langle\psi_i|^{\otimes k} \otimes |\phi_i\rangle\langle\phi_i|^{\otimes l}$ on the symmetric subspace $\mathcal{H}_A^k \otimes \mathcal{H}_B^l$ such that $\tilde{\rho}$ has a positive partial transpose for all partitions.

Thus, given an state ρ , the usual PPT criterion is first used to test. If the test fails, the state is entangled; otherwise, the state can be entangled or separable. For the latter case, we look for an extension $\tilde{\rho}$ of ρ to three parties such that it satisfies all possible partial transpose, which can be solved through SDP introduced in the next. If extension $\tilde{\rho}$ can not be found, we say that the state is entangled; otherwise, we can look for an entension for four parties.

Semidefinite programs for separability problem

The above problem can been solved through SDP, which is basically expressed as

$$\begin{aligned} &\text{minimize} && c^T \mathbf{x}, \\ &\text{subject to} && F(\mathbf{x}) \geq 0, \end{aligned} \quad (4)$$

where c is a given vector, $\mathbf{x} = (x_1, \dots, x_m)$ and $F(\mathbf{x}) = F_0 + \sum_i F_i x_i$, for some fixed n -by- n Hermitian matrices F_j . The inequality in the second line of (2) means that the matrix $F(\mathbf{x})$ is semidefinite. For a particular case when $c = 0$, the problem reduces to whether it is possible to find \mathbf{x} such that $F(\mathbf{x})$ is semidefinite. This turns into a feasibility problem and lightens a separability criteria.

We only consider the problem of searching for an extension of ρ to three parties here. Let $\{\sigma_i^A\}_{i=1,\dots,d_A^2}$ and $\{\sigma_j^B\}_{j=1,\dots,d_B^2}$ be bases for the space of Hermitian matrices that operate on \mathcal{H}_A and \mathcal{H}_B respectively, such that they satisfy

$$\text{Tr}(\sigma_i^X \sigma_j^X) = \alpha \delta_{ij} \quad \text{and} \quad \text{Tr}(\sigma_i^X) = \alpha \delta_{i1} \quad (5)$$

where X stands for A or B , and α is some constant. Thus we can write $\rho = \sum_{ij} \rho_{ij} \sigma_i^A \sigma_j^B$, with $\rho_{ij} = \alpha^{-2} \text{Tr}[\rho \sigma_i^A \sigma_j^B]$. The extension $\tilde{\rho}$ can be written in a similar way

$$\tilde{\rho} = \sum_{ij, i < k} \tilde{\rho}_{kji} \{ \sigma_i^A \otimes \sigma_j^B \otimes \sigma_k^A + \sigma_k^A \otimes \sigma_j^B \otimes \sigma_i^A \} + \sum_{kj} \{ \sigma_k^A \otimes \sigma_j^B \otimes \sigma_k^A \} \quad (6)$$

We also need to satisfy that the trace of $\tilde{\rho}$ for the third party is ρ , that is $\text{Tr}_C[\tilde{\rho}] = \rho$, which leads to $\tilde{\rho}_{ij1} = \rho_{ij}$. The remaining components of $\tilde{\rho}$ will be variables in our SDP. Thus we require that the extension $\tilde{\rho}$ and its partial transpose must be semidefinite. For example, if we want the condition $\tilde{\rho} \geq 0$ to be transformed into $F(\mathbf{x}) = F_0 + \sum_i F_i x_i \geq 0$, we can define

$$\begin{aligned} F_0 &= \sum_j \rho_{1j} \sigma_1^A \otimes \sigma_j^B \otimes \sigma_1^A + \sum_{i=2, j=1} \rho_{ij} \{ \sigma_i^A \otimes \sigma_j^B \otimes \sigma_1^A + \sigma_1^A \otimes \sigma_j^B \otimes \sigma_i^A \}, \\ F_{iji} &= \sigma_i^A \otimes \sigma_j^B \otimes \sigma_i^A, \quad i \geq 2, \\ F_{ijk} &= (\sigma_i^A \otimes \sigma_j^B \otimes \sigma_k^A + \sigma_k^A \otimes \sigma_j^B \otimes \sigma_i^A), \quad k > i \geq 2. \end{aligned} \quad (7)$$

The coefficients $\tilde{\rho}_{ijk}$ ($k \neq 1, k \geq i$) play the role of variable \mathbf{x} . Moreover, positivity of the partial transpose leads to two more inequalities, $\tilde{\rho}^{T_A} \geq 0$ and $\tilde{\rho}^{T_B} \geq 0$.

Therefore, the whole problem, whether a stat ρ is separable or not, turns into finding $\tilde{\rho}_{ijk}(k \neq 1, k \geq i)$ with $G = \tilde{\rho} \otimes \tilde{\rho}^{T_A} \otimes \tilde{\rho}^{T_B} \geq 0$. If such $\tilde{\rho}_{ijk}(k \neq 1, k \geq i)$ don't exist, the state ρ is entangled; If we cannot find such $\tilde{\rho}_{ijk}(k \neq 1, k \geq i)$, the state ρ may be separable or entangled.

Numerical SDP solvers have been introduced in detail in [5]. Given the symmetric subspace $\mathcal{H}_A^k \otimes \mathcal{H}_B^l$, the iteration number for checking this criteria is no more than $O(d_A^{13k/2} d_B^{13l/2})$. Specifically, in our work, we apply SDP in the space $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_A$.

* jsxu@ustc.edu.cn

† yung@sustc.edu.cn

‡ cfli@ustc.edu.cn

- [1] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).
- [2] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).
- [3] P. Horodecki, Phys. Lett. A 232, 333 (1997).
- [4] Doherty A C, Parrilo P A, Spedalieri F M. Physical Review Letters, **18**, 88 (2002).
- [5] L. Vandenbergh and S. Boyd, SIAM Rev. 38, 49 (1996).