Experimental Realization of a Quantum Autoencoder: The Compression of Qutrits via Machine Learning

Alex Pepper, Nora Tischler,[*] and Geoff J. Pryde[†]

Centre for Quantum Dynamics, Griffith University, Brisbane, QLD 4111, Australia

With quantum resources a precious commodity, their efficient use is highly desirable. Quantum autoencoders have been proposed as a way to reduce quantum memory requirements. Generally, an autoencoder is a device that uses machine learning to compress inputs, that is, to represent the input data in a lower-dimensional space. Here, we experimentally realize a quantum autoencoder, which learns how to compress quantum data using a classical optimization routine. We demonstrate that when the inherent structure of the dataset allows lossless compression, our autoencoder reduces qutrits to qubits with low error levels. We also show that the device is able to perform with minimal prior information about the quantum data or physical system and is robust to perturbations during its optimization routine.

Introduction.—Quantum technologies promise to provide us with advantages over their classical counterparts in a variety of tasks, including faster computation [1], secure communication [2], and increased measurement precision [3,4]. However, they depend on quantum resources, for example, quantum coherence, which can be challenging to produce, control, and preserve effectively. As such, quantum resources are precious, and devices that allow us to minimize the use of these resources are valuable. One such device is the quantum autoencoder [5–8].

An autoencoder uses machine learning to represent data in a lower-dimensional space, as illustrated in Fig. 1(a). Autoencoders for classical data form one of the core architectures in machine learning, and offer a range of tools for image processing and other applications [9–11]. Models of autoencoders that compress *quantum* data were recently proposed and theoretically studied in Refs. [5–8,12].

Quantum data compression can benefit applications such as quantum simulation [5,14] and the communication and distributed computation between nodes in a quantum network [7,8], by reducing requirements on quantum memory [5,15], quantum communication channels [8], and the size of quantum gates [7,13]. Reversible, and therefore lossless, compression is possible if a set of quantum states does not span the full Hilbert space in which they are initially encoded. Some methods of compressing quantum data have been proposed and demonstrated previously [15,16]. In those methods, the compression was based on specific assumptions about properties of the quantum states, for example, that a set of qubits be separable. The main advantage of autoencoders is that they do not rely on assumptions of this kind; instead of exploiting a fixed structure of the data, they are capable of learning the structure based on a training dataset. This ability to learn makes them versatile.

Here, we propose and experimentally realize a simple quantum autoencoder scheme. Our device works on a similar principle to the theoretical model of Ref. [5], but with some notable differences. Whereas the proposal of Ref. [5] concerns the compression of some number of qubits to fewer qubits, we pursue a dimensionality reduction in a qudit representation [see Fig. 1(b)]. Consequently, the desired outputs of our device, as well as the definition and evaluation of the cost function, differ from the original model, as described later on. Generally, our scheme supports
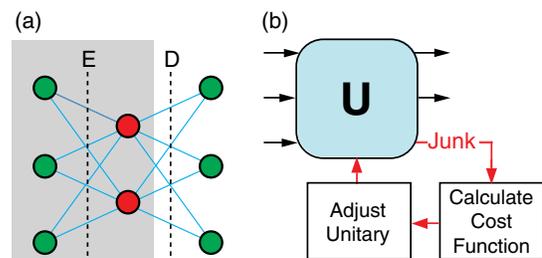

FIG. 1. (a) The concept of an autoencoder. Through an encoding process (*E*), autoencoders represent data in a lower-dimensional space; if the compression is lossless, the original inputs can be perfectly recovered through a decoding process (*D*). (b) The scheme of our qudit-based autoencoder, equivalent to the gray shaded section in (a), for the case of a compression of qutrits to qubits. A unitary transformation *U* characterized by a set of parameters converts between three input modes and three output modes. The iterative training of the parameters aims to minimize the occupation probability of the third output mode, the "junk" mode, across a set of training input states. Lossless compression is achieved when the junk mode is unoccupied. Given that the encoding step is performed with a unitary transformation, the decoding step would simply be the inverse of the unitary, using a vacuum state in the third input mode.
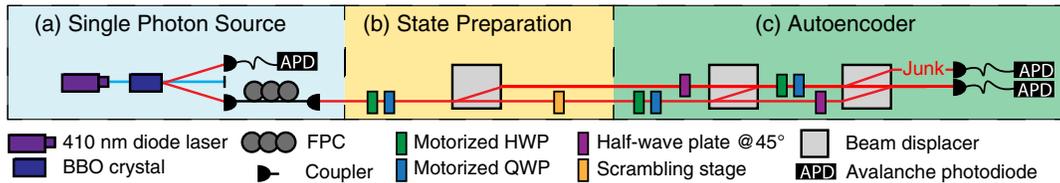
FIG. 2. Experimental layout. (a) Single photon source. We use a 410 nm continuous-wave diode laser to pump a beta barium borate (BBO) crystal to produce photon pairs via type-I spontaneous parametric down-conversion. One photon is collected and detected to herald the second photon; this second photon is fiber coupled and passes through a fiber polarization controller (FPC). (b) State preparation. First, a qubit is encoded in the polarization state of the photon via a motorized half-wave plate (HWP) and quarter-wave plate (QWP). Then the photon passes through a polarizing beam displacer. In the lower spatial mode, a random but fixed birefringent element (a randomly oriented wave plate for a different wavelength) scrambles the polarization, mapping vertical polarization to some superposition of horizontal and vertical. The top spatial mode has matching optical elements (not shown) set to the optic axis, in order to match the optical path length to within the photon's coherence length. By changing the motorized wave plates, sets of qutrits are created in polarization and path modes in such a way that they are compressible into qubits. (c) Autoencoder. A 3×3 unitary transformation with four free parameters is implemented via a series of wave plates, which perform 2×2 unitaries, and beam displacers, which realize mode permutations. The training of the unitary transformation is performed through the rotation of the motorized half- and quarter-wave plates, using a gradient descent routine. The cost function is evaluated based on the measured photon occupation probability of the upper output mode, which contains only one polarization.

a compression of qudits to qunits, where $d > n$ [17]. We demonstrate it experimentally for the case of $d = 3$, $n = 2$, with a photonic device that reduces qutrits to qubits. One potential use of this type of compression lies in applications of quantum communication, for instance between nodes in a quantum network. In quantum communication, photons are a natural choice as quantum information carriers due to their high speed and strong coherence properties. Moreover, the output qubit from our device is encoded in the polarization degree of freedom, which is easily manipulated and can be transmitted over long distances [18].

The combination of machine learning and quantum information processing is a growing research area, which aims to either draw on classical machine learning techniques to aid quantum information tasks or utilize quantum information processing to speed up classical machine learning calculations [19,20]. The quantum autoencoder belongs to the former category. Complementary to recent demonstrations of several classification tasks [21–23], Hamiltonian learning [24], and the reconstruction of quantum states [25], the present work experimentally establishes the compression of quantum data as another use of quantum machine learning.

In this work, we apply the quantum autoencoder to the task of lossless compression, for which we use families of qutrit states that are compressible to qubits. The device can be trained based on a few examples of the family of qutrit states, and can subsequently be tested with other qutrits from the family. Our autoencoder uses a classical machine learning algorithm, gradient descent, to optimize a unitary transformation for compressing the quantum states [17]. The automated experimental unitary optimization circumvents the need to obtain a classical description of the training states, externally design the appropriate unitary, and carry out a full characterization of the optical elements. As an additional benefit, the device is also robust to

disturbances during its optimization routine, as discussed later. These advantages of optimizing directly based on experimental data are akin to previous observations in other systems [23,26].

Experimental scheme.—Our autoencoder consists of a 3×3 unitary transformation with four free parameters, along with a feedback mechanism that is based on measurements of one of the output modes. The setup of our photonic implementation is depicted in Fig. 2. The input qutrits are encoded as superpositions over three optical modes: one spatial mode supporting two polarization modes, and another spatial mode with a fixed polarization. The transformation is implemented as a sequence of 2×2 unitaries, each realized by a set of half- and quarter-wave plates, while mode permutations are realized by beam displacers and half-wave plates set to 45° [27–29]. Such an implementation using joint polarization and spatial modes provides high stability, because all the spatial modes enter and exit the beam displacers through the same facets, and it allows a simple way of controlling the unitary transformation.

Our quantum states are encoded in single photons. In principle, the training process could also be performed with coherent states instead of single photons. However, in applications of the quantum autoencoder within quantum technologies, the training states are more likely to be available in the form of single photons. Given a set of training states, the device performs a unitary transformation with the aim to map all of the training qutrits onto qubits. The desired mapping is achieved when the third output mode, which we will refer to as the "junk" mode, is unused, that is, when the photon cannot be found in that mode [see Fig. 1(b)]. In that case, the mode can be discarded without any effect, leaving the outputs in a qubit space. By contrast, whenever the compression to the qubit output is imperfect, there is a nonzero probability of finding
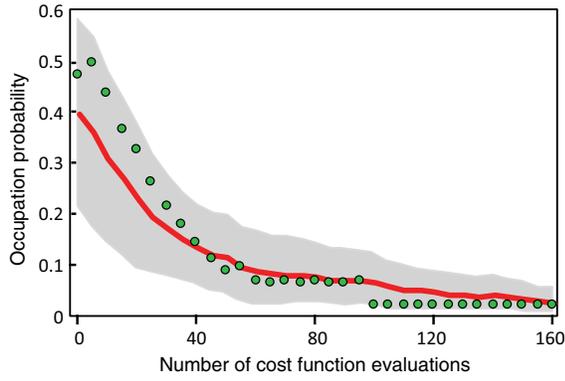
FIG. 3. Training starting from different initializations of the unitary. Here, we used two fixed, randomly selected training states to perform the optimization routine. The cost function, which is the occupation probability of the junk mode averaged over the two training states, is plotted against the number of cost function evaluations. The optimization was carried out 20 times, each time starting with a different randomly initialized unitary. To save training time, a training run was terminated once the average occupation probability reached a threshold of 0.02, and the average occupation probability for that training run was thereafter set to the last measured value. The red line shows the mean values of the 20 training runs, with the gray shaded area indicating ±1 standard deviation of the results. We observe an average occupation probability of 0.03 ± 0.03 after 160 cost function evaluations. The green circles illustrate a sample trajectory. The uncertainties of individual occupation probability measurements, calculated based on Poissonian counting statistics, are smaller than the markers used.

photons in the junk output, and this can be interpreted as a measure of error. The occupation probability of the junk mode P_j quantifies the compression performance twofold: (i) the success probability of the encoding process is $1 - P_j$, and (ii) the fidelity between the given input state and the output after an encoding-decoding sequence is likewise $1 - P_j$ [17]. Therefore, the goal of the training process is to minimize the occupation probability of the junk mode across the training states, ideally to zero. In our experiment, we define the cost function as the average junk mode occupation probability over the different training states. We probe the junk output, calculate the cost function, and use a classical gradient descent optimization routine to adjust the unitary [17].

In order to prepare our input states, polarization qubits are created with a single photon source paired with a half- and quarter-wave plate. We then map these states onto qutrits by using a beam displacer and scrambling the polarization of one spatial output mode. Although the resulting qutrits occupy all three modes, by construction, they lie within a subspace that allows them to be compressed to qubits. We obtain the training states, as well as further states to test the performance of the autoencoder, by varying the half- and quarter-wave plate settings in the state preparation.
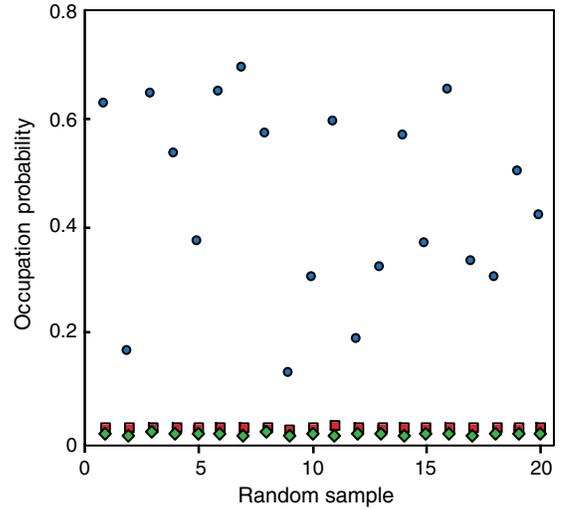

FIG. 4. The effect of the number of training states. The optimization routine was run using one, two, and three training states taken from a compressible subspace. After each training process, 20 random states from the same subspace were prepared and sent through the device to test the compression performance of the unitary. The occupation probability of the junk mode after using one training state (blue circles) was 0.4 ± 0.3, with two training states (red squares) it was 0.03 ± 0.02, and with three training states (green diamonds) it was 0.02 ± 0.02. The uncertainties of individual data points, calculated based on Poissonian counting statistics, are smaller than the markers used.

Results.—The training process is illustrated in Fig. 3, which shows how the device adapts to the training states over time. The training process was repeated 20 times, each time with a different random initialization of the unitary. We observe that the device was able to achieve a high-quality performance, reducing the occupation probability of the junk output mode to 0.03 ± 0.03. The training was performed with a random but fixed set of two training states and no prior calibration of any wave plates, other than the two half-wave plates fixed at 45°. This demonstrates our device performing in general conditions, with no prior information needed about the unitary transformation or quantum states.

The quality of the unitary found by the machine learning process depends on the number of different training states used, as shown in Fig. 4. We explored this relationship by training the unitary with different sized sets of qutrits from a compressible subspace, terminating the training at 200 cost function evaluations, and then testing the device with new, random qutrits from the same subspace. Using a single training state, a low occupation probability of the junk mode was obtained for the training state, but randomized test states were unable to pass through with a similarly low probability of exiting the junk mode. This is unsurprising because a single state does not span the qubit subspace. Using two training states, we saw a consistently high level of performance for all of the test states. Repeating this
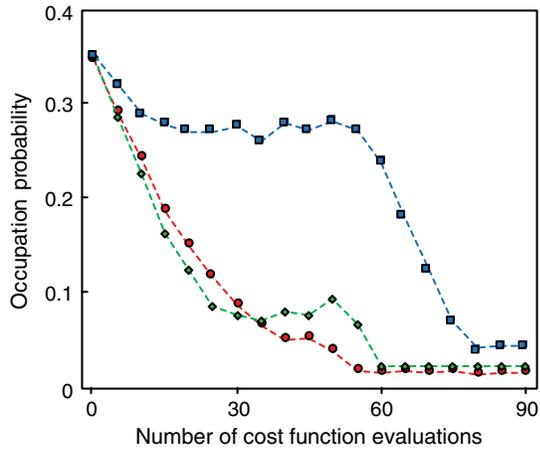
FIG. 5. Test of a channel drift during training. A set of two training states was used to optimize the device under normal conditions, providing a control dataset (red circles) of the occupation probability of the junk mode, averaged over the two training states, versus the number of cost function evaluations. For the green and blue datasets, the same unitary initialization was used, but the necessary encoding was perturbed. This was done by keeping the initial qubits in the state preparation fixed but rotating the wave plate of the scrambling stage by 4° per five cost function evaluations. The datasets of blue squares and green diamonds correspond to different rotation directions. The connecting lines in the plot serve merely as a guide to the eye. The uncertainties of individual occupation probability measurements, calculated based on Poissonian counting statistics, are smaller than the markers used.

process with three training states, we found that the unitary achieved a slightly lower occupation probability of the junk mode across the test states. However, the total training time increased by approximately 40% compared to the run time for two training states (about 90 min), with only minor benefits to the performance. This is why we used two training states in the other parts of the experiment. Note, however, that the run times in our experiment were limited by wave plate rotations, rather than the calculations of the algorithm, and could be made much faster by using fast switching with electro-optic devices. The choice of feedback routine also affects the training time and compression performance. We chose a gradient descent algorithm as a simple way of achieving a reliable device performance, but alternatives, for instance genetic algorithms, could also be considered.

Finally, we investigated the robustness of the device by introducing a channel drift during the optimization routine, as shown in Fig. 5. The aim was to emulate a drift in environmental conditions that affect the required encoding. We performed this by systematically rotating the polarization scrambling wave plate in the state preparation by 4° per five cost function evaluations. Despite an increase in the number of required cost function evaluations, the autoencoder was able to achieve an average occupation probability below 0.05. This shows that it is capable of adapting to changes in

environmental conditions. The learned transformations are provided in the Supplemental Material [17], Sec. D.

Discussion.—In this work, we have proposed and experimentally demonstrated a new, simple scheme for a quantum autoencoder. We have shown that our device is able to compress qutrits to qubits by exploiting the underlying structure of the dataset. The autoencoder is an example of a hybrid quantum-classical machine that optimizes quantum information processing via classical machine learning based on training data. This unsupervised learning provides the valuable capability of adjusting to different datasets, which is absent in alternative methods of compressing quantum data. Indeed, the choice of qubit subspace in our experiment was arbitrary, and was even subjected to a drift during the optimization.

Comparing our scheme with the proposal of Ref. [5], we use a different cost function. Our cost function is defined in terms of the occupation probability of the junk mode, which requires a simple measurement of the mode. By contrast, the cost function of Ref. [5] is defined in terms of the overlap between a fixed reference state and part of the output of the encoding unitary, and is estimated via a SWAP test. Although these definitions and measurements appear quite different, they are based on the same principle that no information should be contained in the discarded parts of the state.

Of the encoding-decoding sequence illustrated in Fig. 1(a), we have implemented the encoding step. Since this encoding is based on a unitary transformation, the decoding step would simply be the inverse of that transformation [17]. Furthermore, although our experiment was designed for qutrits, the same design can be extended to compress higher-dimensional qudits. In that case, the number of input modes and the size of the unitary would be increased, with a polynomial scaling in the number of parameters to be optimized [17]. Depending on the desired output dimensionality, one or more output modes could be monitored. The setup we use can be extended beyond qutrits because of the good stability of beam displacer interferometers and polarization-based two-mode unitary transformations with wave plates [27,28]. However, an even more promising approach for large unitaries lies in integrated photonics, where a reconfigurable unitary transformation in six dimensions has been demonstrated [31]. Indeed, a recent optical implementation of machine learning for a classical application, vowel recognition, has already shown the feasibility of training unitaries in a photonic waveguide architecture [32].

To assess the performance of our device, we have focused on lossless compression, for which we prepared families of qutrits that are in principle perfectly compressible into qubits. The opportunity for lossless compression can arise in certain situations, for example, when a physical symmetry restricts quantum states to a subspace of the full Hilbert space [5], or when a signal originates from a lower-dimensional encoding, but is spread over a larger Hilbert

space through an imperfect transmission channel. However, compression can also be of interest in cases where it is impossible to compress the input data completely faithfully. For instance, it might be worth sacrificing the ability to recover the exact input states for the sake of reducing quantum memory requirements. Exploring such irreversible compression with the autoencoder presents an interesting future research direction.

---

[*]n.tischler@griffith.edu.au

[†]g.pryde@griffith.edu.au

[1] L. K. Grover, in *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing* (ACM, New York, USA, 1996), pp. 212–219.

[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, UK, 2010).

[3] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photonics **5,** 222 (2011).

[4] S. Slussarenko, M. M. Weston, H. M. Chrzanowski, L. K. Shalm, V. B. Verma, S. W. Nam, and G. J. Pryde, Nat. Photonics **11,** 700 (2017).

[5] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum Sci. Technol. **2,** 045001 (2017).

[6] K. H. Wan, O. Dahlsten, H. Kristjánsson, R. Gardner, and M. S. Kim, npj Quantum Inf. **3,** 36 (2017).

[7] L. Lamata, U. Alvarez-Rodriguez, J. D. Martín-Guerrero, M. Sanz, and E. Solano, Quantum Sci. Technol. **4,** 014007 (2019).

[8] G. R. Steinbrecher, J. P. Olson, D. Englund, and J. Carolan, arXiv:1808.10047.

[9] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, Neurocomputing **187,** 27 (2016).

[10] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, Neurocomputing **234,** 11 (2017).

[11] A. Menshawy, *Deep Learning by Example: A Hands-on Guide to Implementing Advanced Machine Learning Algorithms and Neural Networks* (Packt Publishing, Birmingham, UK, 2018).

[12] Recently, we became aware of another, parallel experimental implementation of a quantum autoencoder, reported in Ref. [13].

[13] Y. Ding, L. Lamata, M. Sanz, X. Chen, and E. Solano, arXiv:1807.10643.

[14] J. Thompson, A. J. P. Garner, J. R. Mahoney, J. P. Crutchfield, V. Vedral, and M. Gu, Phys. Rev. X **8,** 031013 (2018).

[15] L. A. Rozema, D. H. Mahler, A. Hayat, P. S. Turner, and A. M. Steinberg, Phys. Rev. Lett. **113,** 160504 (2014).

[16] L. Bartůšková, A. Černoch, R. Filip, J. Fiurášek, J. Soubusta, and M. Dušek, Phys. Rev. A **74,** 022325 (2006).

[17] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.122.060501 for further details.

[18] M. Krenn, M. Malik, T. Scheidl, R. Ursin, and A. Zeilinger, in *Optics in Our Time*, edited by M. D. Al-Amri, M. El-Gomati, and M. S. Zubairy (Springer International Publishing, Cham, 2016), pp. 455–482.

[19] V. Dunjko and H. J. Briegel, Rep. Prog. Phys. **81,** 074001 (2018).

[20] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature (London) **549,** 195 (2017).

[21] Z. Li, X. Liu, N. Xu, and J. Du, Phys. Rev. Lett. **114,** 140504 (2015).

[22] X.-D. Cai, D. Wu, Z.-E. Su, M.-C. Chen, X.-L. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, Phys. Rev. Lett. **114,** 110504 (2015).

[23] 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).

[24] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O'Brien, J. G. Rarity, A. Laing, and M. G. Thompson, Nat. Phys. **13,** 551 (2017).

[25] S. Yu, F. Albarrán-Arriagada, J. C. Retamal, Y.-T. Wang, W. Liu, Z.-J. Ke, Y. Meng, Z.-P. Li, J.-S. Tang, E. Solano, L. Lamata, C.-F. Li, and G.-C. Guo, arXiv:1808.09241.

[26] F. Frank, T. Unden, J. Zoller, R. S. Said, T. Calarco, S. Montangero, B. Naydenov, and F. Jelezko, npj Quantum Inf. **3,** 48 (2017).

[27] A. Laing, Quantum photonic circuits and high dimensional systems, Ph.D. thesis, University of Bristol, 2011.

[28] E. Martín-López, Photonic quantum information processing with the complex Hadamard operator, Ph. D. thesis, University of Bristol, 2014.

[29] An implementation of an arbitrary unitary transformation as a sequence of 2×2 unitaries is always possible, as established in Refs. [27,28,30]. A general unitary would additionally require a half- and quarter-wave plate in the lower spatial mode of Fig. 2, as well as single-mode phase elements, before the detectors. However, these elements are unnecessary for our purposes, since they do not affect our cost function.

[30] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Phys. Rev. Lett. **73,** 58 (1994).

[31] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, G. D. Marshall, M. G. Thompson, J. C. F. Matthews, T. Hashimoto, J. L. O'Brien, and A. Laing, Science **349,** 711 (2015).

[32] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, Nat. Photonics **11,** 441 (2017).