The advantages of ignorance

M. J. Kewming, 1 S. Shrapnel, 1 A. G. White, 1 and J. Romero 1

1 Centre for Engineered Quantum Systems, School of Mathematics and Physics, University of Queensland, QLD 4072 Australia

(Dated: March 25, 2019)

When we conduct scientific experiments and measure the world around us, we are attempting to fill gaps in our missing knowledge. This absence in knowledge is called ignorance and it too, is worthy of scientific investigation. Ignorance inherent in a system can be distributed and located in many of its constituent parts. Naturally one might ask does ignorance of the whole system imply ignorance of its parts? Our classical intuition tells us yes, if you are ignorant of the whole, you must be ignorant of at least one clearly identifiable part. However, in quantum theory this reasoning is not correct. Experimental proof of this counter-intuitive fact requires controlling and measuring quantum systems of high dimension \((d > 9)\). We provide this experimental evidence using the transverse spatial modes of light, a powerful resource for testing high dimensional quantum phenomenon. Our observations demonstrate the existence of new novel, counter-intuitive, quantum feature of nature.

Scientific inquiry is a process of uncovering our ignorance about the universe. When we do not know the answer to a question, it signals our ignorance of a particular part in our understanding. For example, say you are reading a book and you decide to skip a chapter. This implies that you are ignorant of this chapter’s content. If someone decided to measure your ignorance of the whole book by asking you questions, they would be able to determine the source of your ignorance, namely the chapter you skipped. Books are classical objects and satisfy the statement that ignorance of the whole implies ignorance of the parts. That is, if you don’t know everything about the whole system, you can always point to the part you do not know. However, quantum physics has consistently defied our classical intuition, so it is rightful to question whether this intuition of ignorance holds in the quantum world.

Many foundational principles in modern physics are formulated as mathematical inequalities, most notably Heisenberg’s uncertainty principle and Bell’s theorem. Nowadays, many of these foundational inequalities have been recast in terms of entropy, a more universal measure of a distribution’s randomness [4,5]. Entropy is also used as a a measure of ignorance. Entropic inequalities have found use in a variety of practical settings including non-locality [2], information causality [3], cryptography [4,6], and quantum memories [7].

Vidick and Wehner (VW) formalised the statement ignorance of the whole does not imply ignorance of the parts as a violation of an entropic inequality [8]. This VW-inequality is a non-contextuality inequality: it holds for classical systems, but can be violated by quantum systems. As shown by Kochen and Spekker [9], it is impossible to attribute pre-determined values to an observable property of a quantum system, independently of any compatible observables that may be measured with it—its context. This curious feature of contextuality is known to be a crucial resource in quantum computation [10].

The conceptual significance of this violation can be appreciated by considering a familiar scenario [8]. Imagine a teacher wants to determine a student’s knowledge about a particular textbook. She can test his knowledge by asking him to sit a series of tests. In each test he will be asked one question drawn from any page in the textbook. Each page of the textbook contains a fixed number of possible questions. Before each test, the teacher announces the possible questions—one question drawn from each page—that may appear on the test. The student is told that he can encode the answer to one question in some study notes which he can take into the test. He may use the same encoding strategy for each test, for example, he may always encode the answers from one particular page. During the test, he can only rely on his notes that he prepared because he completely forgets all the answers. To uncover his ignorance, the teacher simply needs to ask a question not included in his notes. Repeating this for many different tests, she can uncover his encoding strategy and the source of his ignorance.

Let us consider the simplest possible case of a two-page textbook. In this example, a teacher probes the knowledge of the student who is required to know the answers \(y_0\) and \(y_1\) to two possible questions from either page. She can ask him a single question about either part \(Y_0\) or \(Y_1\), or about the whole \(Y\). The student’s answers are randomly drawn from the set of possible answers to either page \(Y_0\) and \(Y_1\) where \(y_0\) and \(y_1\in\{0,1,...,d-1\}\) are represented as dits. The combination of answers from both pages is a single dit string \(y=y_0y_1\). Recall that the student can prepare one answer—one dit of information—in his study notes \(P_y\) which are encoded in the register \(E\). The values of the two dits \(y=y_0y_1\) are announced by the teacher, but the student forgets them both before sitting the test and can only use the one dit that is encoded in his study notes. In the test, the teacher asks a single question \(M\) which leads to the conditional probability of successfully guessing \(y\) correctly as \(P(y|M,P_y)\).

Although the test only contains one question, in the long run all questions will be asked, allowing the teacher...
to reconstruct both the student’s knowledge of the whole Y and the student’s knowledge of the parts Y_0 or Y_1. For example, to determine the whole Y, the students sits d^2 tests, each time being asked to guess a different y. The probability of the student guessing the whole book correctly is

\[ p_{\text{guess}}(Y|E) = \max_{\{M\}} \sum_y P_y(y)P(y|M, \mathcal{P}_y) \]  

where \( P_y(y) = 1/d^2 \) (recall \( y_0 \) and \( y_1 \) are selected at random). The maximisation over \( M \) signifies the highest success probability of student for a given test. Therefore, the teacher measures the student’s ignorance using the conditional min-entropy defined as

\[ H_\infty(Y|E) = -\log p_{\text{guess}}(Y|E) \]  

The student’s ignorance of the whole textbook is bounded by the dimension of his study notes \( d \). This occurs because he has one dit of information encoded in his notes, but is required to know 2 dits of information (one dit for either page). Therefore he always lacks one dit of information, hence \( H_\infty(Y|E) = \log d \). In fact, this bound also holds in the quantum case where his notes are replaced by a single qudit. A similar result bounding the maximum retrievable amount of information from quantum systems was first derived by Kraus [12] and later generalised to broader class of Rényi entropies by Maassen and Uffink [13].

Now the teacher wants to determine the student’s ignorance of the parts. Suppose, the student always prepares his notes such that he only encodes the answers from \( Y_1 \), that is \( E=Y_1 \). Because he encodes \( Y_1 \), he guesses all the answers \( y_1 \) with certainty \( P(Y_1|E=Y_1)=1 \). As a consequence, he must guess the answers from \( Y_0 \) randomly \( P(Y_0|E=Y_1)=1/d \). For \( d=2 \), the probability of guessing the parts correctly is \( p_{\text{guess}}(Y_C|E) = 0.75 \) where \( C \) is a binary random variable with outcomes \( c \in \{0,1\} \) and indicates the page he is required to guess. This result holds for any mixed distribution of encoded answers \( Y_0 \) and \( Y_1 \).

Can the teacher point to the source of the student’s ignorance? Classically this is true. Clearly in the above case, \( Y_0 \) is the source. Let us assume that the teacher has access to the random variable \( C \) which is classically correlated with the student’s notes \( E \), such that \( P(c=0|E)=1 \) with certainty. In this case, the teacher always asks him for answers from \( Y_0 \) given he always encodes \( E=Y_1 \). Here, \( C \) points to the source of ignorance. Moreover marginalising over \( C \) does not tamper with the student’s notes \( E \).

This is the intuition behind the VW-inequality. For a given encoding \( E \) of any dimension \( d \), there always exists a random variable \( C \) that can point to the source of ignorance \[ H_\infty(Y_C|E, C) \geq H_\infty(Y|E) - 1 \]  

where \( H_\infty(Y|E) \) is the ignorance of the whole and \( H_\infty(Y_C|E, C) \) is the ignorance of the parts. One can think of this inequality as a constraint on how ignorance of the whole \( H_\infty(Y|E) \) can be split between the two parts \( Y_C \). The minus one accounts for the fact that we have conditioned on \( C \), gaining a single bit of information as a result. Accounting for this makes the inequality robust against the possibility that \( C \) may have assisted the student. We can interpret this inequality as an implication; if you are ignorant of the whole book \( Y \), then you can always find the part \( Y_C \) you are ignorant of. But does this logic hold given our understanding of quantum mechanics?

Let us recast the problem in terms of quantum theory. The student now has access to a quantum set of notes \( \rho_y^E = |\Psi_y\rangle \langle \Psi_y| \) which is a single qudit. The questions become positive-operator valued measurements (POVM) \( \sum_y M_y = 1 \). The random variable \( C \) becomes a classical mixed state \( \sigma_Y^C = q_y |0\rangle \langle 0| + (1-q_y) |1\rangle \langle 1| \) for example a biased coin, where \( q_y \) is a measure of the bias for one outcome. Similarly, the outcome of successfully guessing \( y \) becomes \( P(y|M, \mathcal{P}_y) = \text{tr} (\rho_y^E M_y) \). Using operator notation, we depict a schematic of this conception in Figure 4.

The student can encode the answers \( y_0 \) and \( y_1 \) using two mutually unbiased bases, namely the generalised Pauli operators \( X_d \) and \( Z_d \),

\[ |\Psi_y\rangle = \frac{1}{\sqrt{2}} \left( X_d^{y_0} X_d^{y_1}\left(I + F\right) |0\rangle \right) \]  

where \( F \) is the quantum Fourier transform. In this encoding, the answers from \( Y_0 \) are encoded in the computational basis \( X_d \) and answers from \( Y_1 \) are encoded in the conjugate basis \( Z_d \). These states are referred to as the
Spatial modes are photonic qudits which can be described by orthogonal families of mathematical functions. We use the Laguerre-Gauss (LG) bases in this experiment with each mode fully characterised by two numbers \((l, p)\). The encoded answers can then be represented as a weighted superposition of these bases: 

\[
|\Psi_y\rangle = \sum_i a_i |\psi_i\rangle_{l, p},
\]

where \(a_i\) is a complex number containing the information about the encoded pages. Several example encoded qudits \(\rho^E_y\) are shown in Figure 2.

Large qudit systems can be readily produced and measured using phase masks displayed on liquid-crystal spatial light modulators (SLM’s). We use an 809 nm diode laser attenuated to the single-photon level as our light source. Our experimental apparatus is depicted in Figure 2(a). We use the intensity and phase encoding techniques identified in Ref [25] to test Eq. (2). We prepare single photons in a 4-rail Fock basis qudit using the first SLM where each rail is formed by an orthogonal LG mode \((l, p)\). The superposition of these modes creates a set quantum study notes encoded in the spatial profile of a single photon. The figures show the spatial amplitude overlayed with the phase profile. The \(y_0\) dit is shown in the rows and the \(y_1\) dit is shown in the columns.

FIG. 2. a) In the experiment, the student’s notes for the answer \(y = y_0 y_1\) consists of a single qudit \(\rho^E_y\) encoded in the spatial mode of a single photon. The photon is prepared using a phase mask displayed on a spatial light modulator (SLM). The photon is then propagated to the second SLM through a 4\(f\) lens configuration. The teacher sets the phase mask of the second SLM, corresponding to a question \(M_{y_c}\), where \(c\) indicates the page from which the teacher’s question is derived. The photon is then coupled to a single mode fibre connected to an avalanche photodiode (APD). If the photon is detected at the APD it means the student correctly answered the question. The distribution \(C\) can be correlated with the student’s notes by any classical system (e.g. a computer) b) Each image shows a unique set of answers for \(d=3\) encoded in the spatial profile of a single photon. The teacher tries to point to the student’s ignorance by asking the question \(M_{y_c}\), however the student can answer both questions with equal probability \(p_{\text{guess}}(Y_c|E)=\frac{1}{d}\). For \(d=2\) the student can guess questions from both pages with \(p_{\text{guess}}(Y_c|E,C)\approx 0.853\) gaining an advantage over the classical case. Furthermore, the student’s results are now independent of the system \(\sigma^C_y\), that is \(C\) can no longer point to the source of ignorance. As we increase the Hilbert space dimension for \(d>2\) (the number of possible questions and answers), we find that it becomes impossible to satisfy Eq. (2) using a classically correlated \(\sigma^C_y\). Despite the student being ignorant of at least one page, the teacher cannot uncover which page. In a quantum world ignorance of the whole does not imply ignorance of the parts.

Here we present an experimental violation of Eq. (2) using the transverse spatial mode of photons. The transverse spatial profile of photons is a widely-used and readily accessible qudit basis which has been used for both foundational and technological studies. These include demonstration of the Einstein-Podolsky-Rosen paradox and violations of the Bell inequality in higher dimensions, intra-city quantum cryptography, and free-space communication.

Spatial modes are photonic qudits which can be described by orthogonal families of mathematical functions. We use the Laguerre-Gauss (LG) bases in this experiment with each mode fully characterised by two numbers \((l, p)\). The encoded answers can then be represented as a weighted superposition of these bases: 

\[
|\Psi_y\rangle = \sum_i a_i |\psi_i\rangle_{l, p},
\]

where \(a_i\) is a complex number containing the information about the encoded pages. Several example encoded qudits \(\rho^E_y\) are shown in Figure 2.

Large qudit systems can be readily produced and measured using phase masks displayed on liquid-crystal spatial light modulators (SLM’s). We use an 809 nm diode laser attenuated to the single-photon level as our light source. Our experimental apparatus is depicted in Figure 2(a). We use the intensity and phase encoding techniques identified in Ref [25] to test Eq. (2). We prepare single photons in a 4-rail Fock basis qudit using the first SLM where each rail is formed by an orthogonal LG mode \((l, p)\). The superposition of these modes creates a set quantum study notes encoded in the spatial profile of a single photon. The figures show the spatial amplitude overlayed with the phase profile. The \(y_0\) dit is shown in the rows and the \(y_1\) dit is shown in the columns.

FIG. 2. a) In the experiment, the student’s notes for the answer \(y = y_0 y_1\) consists of a single qudit \(\rho^E_y\) encoded in the spatial mode of a single photon. The photon is prepared using a phase mask displayed on a spatial light modulator (SLM). The photon is then propagated to the second SLM through a 4\(f\) lens configuration. The teacher sets the phase mask of the second SLM, corresponding to a question \(M_{y_c}\), where \(c\) indicates the page from which the teacher’s question is derived. The photon is then coupled to a single mode fibre connected to an avalanche photodiode (APD). If the photon is detected at the APD it means the student correctly answered the question. The distribution \(C\) can be correlated with the student’s notes by any classical system (e.g. a computer) b) Each image shows a unique set of answers for \(d=3\) encoded in the spatial profile of a single photon. The teacher tries to point to the student’s ignorance by asking the question \(M_{y_c}\), however the student can answer both questions with equal probability \(p_{\text{guess}}(Y_c|E,C)\approx 0.853\); gaining an advantage over the classical case. Furthermore, the student’s results are now independent of the system \(\sigma^C_y\), that is \(C\) can no longer point to the source of ignorance. As we increase the Hilbert space dimension for \(d>2\) (the number of possible questions and answers), we find that it becomes impossible to satisfy Eq. (2) using a classically correlated \(\sigma^C_y\). Despite the student being ignorant of at least one page, the teacher cannot uncover which page. In a quantum world ignorance of the whole does not imply ignorance of the parts.

Here we present an experimental violation of Eq. (2) using the transverse spatial mode of photons. The transverse spatial profile of photons is a widely-used and readily accessible qudit basis which has been used for both foundational and technological studies. These include demonstration of the Einstein-Podolsky-Rosen paradox and violations of the Bell inequality in higher dimensions, intra-city quantum cryptography, and free-space communication.

Spatial modes are photonic qudits which can be described by orthogonal families of mathematical functions. We use the Laguerre-Gauss (LG) bases in this experiment with each mode fully characterised by two numbers \((l, p)\). The encoded answers can then be represented as a weighted superposition of these bases: 

\[
|\Psi_y\rangle = \sum_i a_i |\psi_i\rangle_{l, p},
\]

where \(a_i\) is a complex number containing the information about the encoded pages. Several example encoded qudits \(\rho^E_y\) are shown in Figure 2.

Large qudit systems can be readily produced and measured using phase masks displayed on liquid-crystal spatial light modulators (SLM’s). We use an 809 nm diode laser attenuated to the single-photon level as our light source. Our experimental apparatus is depicted in Figure 2(a). We use the intensity and phase encoding techniques identified in Ref [25] to test Eq. (2). We prepare single photons in a 4-rail Fock basis qudit using the first SLM where each rail is formed by an orthogonal LG mode \((l, p)\). The superposition of these modes creates a set quantum study notes encoded in the spatial profile of a single photon. The figures show the spatial amplitude overlayed with the phase profile. The \(y_0\) dit is shown in the rows and the \(y_1\) dit is shown in the columns.
with the theoretical predictions seen in Figure 3(a). Our observations confirm that for $d > 9$, using the encoded states Eq. (3), the RHS of the inequality (ignorance of the whole) becomes greater than the LHS (ignorance of the parts). These results serve as a calibration of the experiment, ensuring that we can create high dimensional qudits accurately.

Violating Eq. (2) requires that the teacher cannot point to the student’s source of ignorance given the binary distribution $\sigma_y$. Here we establish this result by fixing the system at $d=13$, in the region where the VW inequality is violated, and introduce the role of the teacher. She randomly asks the student to answer either $Y_0$ or $Y_1$ according to the probability distribution $\sigma_y$. If we can find a $\sigma_y$ such that our measurement of the parts falls in the shaded green region (signifying statistical overlap with the whole) and crosses into the white area in Figure 3(b), then we have satisfied the VW-inequality. Our measurements show that it is impossible to satisfy the VW-inequality using notes encoded by Eq. (4) (black diamonds) confirming that ignorance of the whole does not imply ignorance of the parts. We can also test the optimal classical case $E=Y_0$ (right) or $E=Y_1$ (left) where we create statistical mixtures of both pages by randomly selecting different classical encodings $E$. The results of this experiment Figure 3(b) show a strong dependence on the teacher’s choice system $\sigma_y$. As a result, we satisfy the VW-inequality when these measurements cross the (green band) into the white region.

Our experiment is a demonstration that given control of dimension, physics at high dimensions can readily be explored. Transverse spatial modes are an excellent candidate when pursuing proof of principle experiments in high dimensions as we have demonstrated. SLMs allow researchers to accurately control the preparation and measurement procedure of high-dimensional states. However, this is limited by the physical size of the beams that can be prepared and measured in the experiment. Other qudit systems do not suffer from such limitations, such as time-bin encoding [26]. It may be possible to combine these qudits with the transverse spatial modes to create a unique architecture for both foundational and technological studies needing high-dimensional quantum information.

Inequalities in physical theories have played a significant role in defining the boundary of the quantum and classical world. Here we provide experimental evidence of another boundary, namely violation of VW-inequality which concludes that ignorance of the whole does not imply ignorance of the parts in quantum systems. A unique component of this inequality is the dependence on dimension. We do not see violation of the inequality until $d > 9$ because we must discount the influence of the teacher’s system $C$ which could aid the student by biasing the tests in their favour. If we continue to increase $d$ the rift between the whole and the parts continues to widen. In the
asymptotic limit $d \to \infty$, the min-entropy of the whole increases as $H_\infty(Y|E) = \log d$ since we cannot fully encode both parts of $y$ into $\rho_y^E$. On the other hand, the ignorance of the parts for both the classical and quantum case approach $H_\infty(Y | E, C) \to 1$. That the quantum case violates the VW-inequality in this asymptotic limit, whereas the classical does not, indicates a significant role of the coherent interference between the two complementary bases $X_d$ and $Z_d$. Information encoded in either basis can be retrieved with equal probability $p_{\text{guess}}(Y | E, C) \approx 50\%$ despite the large number of possible outcomes $d$. This result is known as Bohr complementarity where two conjugate variables can be known individually but not simultaneously. Our result views complementarity from a perspective of ignorance, introducing a novel line of inquiry into the differences between classical and quantum systems.

Acknowledgments. JR is supported by an ARC Discovery Early Career Research Award (DE160100409) and a L’Oreal-UNESCO For Women In Science Fellowship Award. AGW is supported by a University of Queensland Vice-Chancellor Research and Teaching Fellowship. JR would like to thank Thomas Vidick and Stephanie Wehner for initial discussions. We acknowledge Gerard Milburn and Marco Tomamichel for fruitful discussions and suggestions.

* jacq.romero@gmail.com
[1] S. Wehner and A. Winter, New Journal of Physics 12, 025009 (2010).
[2] J. Oppenheim and S. Wehner, Science 330, 1072 (2010).
[3] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, New Journal of Physics 12, 033024 (2010).
[4] I. B. Damgaard, S. Fehr, R. Renner, L. Salvail, and C. Schaffner, arXiv:quant-ph/0612014 (2006).
[5] S. Wehner, C. Schaffner, and B. M. Terhal, Physical Review Letters 100, 220502 (2008).
[6] F. Furrer, T. Gehring, C. Schaffner, C. Pacher, R. Schnabel, and S. Wehner, Nature Communications 9, 1450 (2018).
[7] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, Nature Physics 6, 659 (2010).
[8] T. Vidick and S. Wehner, Physical Review Letters 107, 030402 (2011).
[9] S. Kochen and E. P. Specker, Journal of Mathematics and Mechanics 17, 59 (1967).
[10] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Nature 510, 351 (2014).
[11] R. Konig, R. Renner, and C. Schaffner, IEEE Transactions on Information Theory 55, 4337 (2009).
[12] K. Kraus, Physical Review D 35, 3070 (1987).
[13] H. Maassen and J. B. M. Uffink, Physical Review Letters 60, 1103 (1988).
[14] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information: 10th Anniversary Edition, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).
[15] C. Crpeau, Journal of Modern Optics 41, 2445 (1994).
[16] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane, Physical Review Letters 114 (2015), 10.1103/PhysRevLett.114.170502.
[17] A. Chailhoux, I. Kerenidis, S. Kundu, and J. Sikora, New Journal of Physics 18, 045003 (2016).
[18] A. Forbes, A. Dudley, and M. M. Nielsen, Advances in Optics and Photonics 8, 200 (2016).
[19] J. Leach, B. Jack, J. Romero, A. Jha, A. Yao, S. Franke-Arnold, D. Ireland, R. Boyd, S. Barnett, and M. Padgett, Science 329, 662 (2010).
[20] A. Dada, J. Leach, G. Buller, M. J. Padgett, and E. Andersson, Nature Physics 7 (2011). doi: 10.1038/nphys1996
[21] A. Sit, F. Boukarm, R. Fickler, J. Gagnon-Bischoff, H. Laroque, K. Heshami, D. Elser, C. Peuntinger, K. Günthner, B. Heim, et al., Optica 4, 1006 (2017).
[22] M. Krenn, J. Handsteiner, M. Fink, R. Fickler, R. Ursin, M. Malik, and A. Zeilinger, Proceedings of the National Academy of Sciences 113, 13648 (2016).
[23] G. Molina-Terriza, J. P. Torres, and L. Torner, Nature Physics 3, 305 (2007).
[24] N. K. Langford, R. B. Dalton, M. D. Harvey, J. L. O’Brien, G. J. Pryde, A. Gilchrist, S. D. Bartlett, and A. G. White, Physical Review Letters 93, 053601 (2004).
[25] E. Bolduc, N. Bent, E. Santamato, E. Karimi, and R. W. Boyd, Optics Letters 38, 3546 (2013).
[26] N. T. Islam, C. C. W. Lim, C. Cahall, J. Kim, and D. J. Gauthier, Science Advances 3, e1704191 (2017).
[27] T. Durt, B.-G. Englert, I. Bengtsson, and K. yyczkowski, International Journal of Quantum Information 08, 535 (2010).
[28] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, Reviews of Modern Physics 79, 135 (2007).
SUPPLEMENTARY MATERIAL

Experimental details

We use a 809 nm continuous light source attenuated using neutral density filters until $\sim 10^6$ photons/sec are detected using a Perkins Elmer SPCM-AQR single photon counting module with an average dark count of $\sim 150$ photons/sec. The light source is spatially filtered, expanded and then collimated to ensure a clean single mode profile. We expand the input beam so it is much larger than the encoded beam creating an approximately uniform input profile over encoding phase masks. To generate the encoded state and measure it, we use two Meadowlark 1920 x 1152 analog spatial light modulators. We use the amplitude modulation technique described in Ref [25] on the encoded state to exactly control the spatial transverse modes. The results of this method can be seen in Figure 4 where the phase mask displayed on the SLM. The speckle noise seen in the experimental images are due to the camera and are not present in the actual modes. The encoding phase mask generates a state $|\Psi_y\rangle$.

Given the unitary nature of geometric optics, swapping the input and output fibres of the experiment leads to reversal of the experiment. In this reversed experiment, the measurement SLM generates a state $|\Phi_y\rangle$. Therefore, propagating the spatial profile through a 4f lens configuration from the encoding SLM to the measurement SLM allows us to perform the measurement $\langle \Phi_y | \Psi_y \rangle$. We then couple to a single mode fibre which projects out the zeroth mode yielding the counting measurement $C_y = |\langle \Phi_y | \Psi_y \rangle|^2$ [13, 25]. The probability of successfully detecting a photon that is in $y_0$ can then be measured as

$$P(y_0|E) = \operatorname{tr} \left( \rho_y^E M_{y_0} \right) = \frac{|\langle \Phi_{y_0} | \Psi_y \rangle|^2}{|\langle \Psi_y | \Psi_y \rangle|^2} = \frac{C_{y_0}}{N_y}$$  \hspace{1cm} (4)

where $N_y = |\langle \Psi_y | \Psi_y \rangle|^2$ are the normalised counts for a particular encoding $y$. In the original conception [8], the optimal measurement POVM for the whole is $M_y = |\Psi_y\rangle\langle \Psi_y | / d$ which is determined by a semi-definite program. Measuring experimentally using SLMs is equivalent to

$$P(y|E) = \frac{1}{d} \operatorname{tr} \left( \rho_y^E \rho_y^E \right) = \frac{1}{d} \frac{N_y}{N_y} = \frac{1}{d},$$  \hspace{1cm} (5)

which is not really a measurement. However, it easy to show that for a complete set of orthogonal POVM measurements $\sum_{i=0}^{d-1} M_i = 1$ chosen at random,

$$P(y|E) = \frac{1}{d} \operatorname{tr} \left( \rho_y^E M_{y_0} \right) = \frac{1}{d} \frac{N_y}{N_y} = \frac{1}{d} \sum_{i=0}^{d-1} P(M_i) C_i = \frac{C}{N_y}$$  \hspace{1cm} (6)

where $F$ is the quantum Fourier transform and $X_d$ and $Z_d$ are the generalised Pauli operators that are the generators of the Heisenberg-Weyl group and are defined as

$$X_d|y_0\rangle = |y_0 + 1 \text{ mod } d\rangle \text{ and } Z_d|y_0\rangle = \omega^{y_0}|y_0\rangle$$  \hspace{1cm} (8)

where $\omega = \exp(2\pi i / d)$. The Pauli operators form a canonical conjugate pair and are related by the $Z_d = F^T X_d F$. In quantum mechanics the notion of canonically conjugate quantities is central irrespective of Hilbert space dimension. If the state of a system is such that one canonical variable takes a definite value, then the conjugate must be maximally uncertain. In the original proof of the VW-inequality, the authors assumed $d$ was prime to complete the proof [8]. This assumption is required due to the unanswered questions relating to MUBs for non-prime dimensions [27]. Currently, it is not known how many MUBs exist for composite Hilbert dimensions, but is well known for prime powers and the continuous limit.

The need for this type of encoding is to create a quantum superposition, that is, it is not a deterministic classical mixed state of the $X_d$ and $Z_d$ basis. The inequality

$$P(y|E) = \frac{1}{d} \operatorname{tr} \left( \rho_y^E M_{y_0} \right) = \frac{1}{d} \frac{N_y}{N_y} = \frac{1}{d} \sum_{i=0}^{d-1} P(M_i) C_i = \frac{C}{N_y}$$  \hspace{1cm} (6)

where $F$ is the quantum Fourier transform and $X_d$ and $Z_d$ are the generalised Pauli operators that are the generators of the Heisenberg-Weyl group and are defined as

$$X_d|y_0\rangle = |y_0 + 1 \text{ mod } d\rangle \text{ and } Z_d|y_0\rangle = \omega^{y_0}|y_0\rangle$$  \hspace{1cm} (8)

where $\omega = \exp(2\pi i / d)$. The Pauli operators form a canonical conjugate pair and are related by the $Z_d = F^T X_d F$. In quantum mechanics the notion of canonically conjugate quantities is central irrespective of Hilbert space dimension. If the state of a system is such that one canonical variable takes a definite value, then the conjugate must be maximally uncertain. In the original proof of the VW-inequality, the authors assumed $d$ was prime to complete the proof [8]. This assumption is required due to the unanswered questions relating to MUBs for non-prime dimensions [27]. Currently, it is not known how many MUBs exist for composite Hilbert dimensions, but is well known for prime powers and the continuous limit.

The need for this type of encoding is to create a quantum superposition, that is, it is not a deterministic classical mixed state of the $X_d$ and $Z_d$ basis. The inequality

$$P(y|E) = \frac{1}{d} \operatorname{tr} \left( \rho_y^E M_{y_0} \right) = \frac{1}{d} \frac{N_y}{N_y} = \frac{1}{d} \sum_{i=0}^{d-1} P(M_i) C_i = \frac{C}{N_y}$$  \hspace{1cm} (6)

where $F$ is the quantum Fourier transform and $X_d$ and $Z_d$ are the generalised Pauli operators that are the generators of the Heisenberg-Weyl group and are defined as

$$X_d|y_0\rangle = |y_0 + 1 \text{ mod } d\rangle \text{ and } Z_d|y_0\rangle = \omega^{y_0}|y_0\rangle$$  \hspace{1cm} (8)

where $\omega = \exp(2\pi i / d)$. The Pauli operators form a canonical conjugate pair and are related by the $Z_d = F^T X_d F$. In quantum mechanics the notion of canonically conjugate quantities is central irrespective of Hilbert space dimension. If the state of a system is such that one canonical variable takes a definite value, then the conjugate must be maximally uncertain. In the original proof of the VW-inequality, the authors assumed $d$ was prime to complete the proof [8]. This assumption is required due to the unanswered questions relating to MUBs for non-prime dimensions [27]. Currently, it is not known how many MUBs exist for composite Hilbert dimensions, but is well known for prime powers and the continuous limit.

The need for this type of encoding is to create a quantum superposition, that is, it is not a deterministic classical mixed state of the $X_d$ and $Z_d$ basis. The inequality

$$P(y|E) = \frac{1}{d} \operatorname{tr} \left( \rho_y^E M_{y_0} \right) = \frac{1}{d} \frac{N_y}{N_y} = \frac{1}{d} \sum_{i=0}^{d-1} P(M_i) C_i = \frac{C}{N_y}$$  \hspace{1cm} (6)

where $F$ is the quantum Fourier transform and $X_d$ and $Z_d$ are the generalised Pauli operators that are the generators of the Heisenberg-Weyl group and are defined as

$$X_d|y_0\rangle = |y_0 + 1 \text{ mod } d\rangle \text{ and } Z_d|y_0\rangle = \omega^{y_0}|y_0\rangle$$  \hspace{1cm} (8)

where $\omega = \exp(2\pi i / d)$. The Pauli operators form a canonical conjugate pair and are related by the $Z_d = F^T X_d F$. In quantum mechanics the notion of canonically conjugate quantities is central irrespective of Hilbert space dimension. If the state of a system is such that one canonical variable takes a definite value, then the conjugate must be maximally uncertain. In the original proof of the VW-inequality, the authors assumed $d$ was prime to complete the proof [8]. This assumption is required due to the unanswered questions relating to MUBs for non-prime dimensions [27]. Currently, it is not known how many MUBs exist for composite Hilbert dimensions, but is well known for prime powers and the continuous limit.

The need for this type of encoding is to create a quantum superposition, that is, it is not a deterministic classical mixed state of the $X_d$ and $Z_d$ basis. The inequality

$$P(y|E) = \frac{1}{d} \operatorname{tr} \left( \rho_y^E M_{y_0} \right) = \frac{1}{d} \frac{N_y}{N_y} = \frac{1}{d} \sum_{i=0}^{d-1} P(M_i) C_i = \frac{C}{N_y}$$  \hspace{1cm} (6)

where $F$ is the quantum Fourier transform and $X_d$ and $Z_d$ are the generalised Pauli operators that are the generators of the Heisenberg-Weyl group and are defined as

$$X_d|y_0\rangle = |y_0 + 1 \text{ mod } d\rangle \text{ and } Z_d|y_0\rangle = \omega^{y_0}|y_0\rangle$$  \hspace{1cm} (8)

where $\omega = \exp(2\pi i / d)$. The Pauli operators form a canonical conjugate pair and are related by the $Z_d = F^T X_d F$. In quantum mechanics the notion of canonically conjugate quantities is central irrespective of Hilbert space dimension. If the state of a system is such that one canonical variable takes a definite value, then the conjugate must be maximally uncertain. In the original proof of the VW-inequality, the authors assumed $d$ was prime to complete the proof [8]. This assumption is required due to the unanswered questions relating to MUBs for non-prime dimensions [27]. Currently, it is not known how many MUBs exist for composite Hilbert dimensions, but is well known for prime powers and the continuous limit.
The probability of successfully guessing $y$ is then given by the trace operator $\text{tr}(\rho^E_y M_y)$.

To obtain a measurement of the whole $Y$, we sum over all combinations of $y$ multiplying each probability of the outcome by the probability that $y$ was selected. Hence $p_{\text{guess}}(Y|E) = \max_y \sum_y P_Y(y)\text{tr}(\rho^E_y M_y)$ where $P_Y(y) = 1/d^2$. Similarly, we can measure the parts in the same way e.g. $p_{\text{guess}}(Y_0|E) = \max_y \sum_y P_{Y_0}(y_0)\text{tr}(\rho^E_y M_{y_0})$ where $P_Y(y) = 1/d$.

The min-entropy $H_\infty(Y|E) = -\log p_{\text{guess}}(Y|E)$ can then be computed by taking the log of each guessing probability.

---

**d-rail qudits**

Here we present an overview of the d-rail qudits analysis. Let $d$ represent the number of available modes, then the total Hilbert space is the tensor product of Fock space spanned by the states

$$|n_1, n_2, ..., n_d\rangle \equiv |n_1\rangle \otimes |n_2\rangle \otimes ... \otimes |n_d\rangle .$$

We will further assume that we are working in a subspace in which every state is an eigenstate of the total photon number operator $\hat{N}|\psi\rangle = \hat{N}|\psi\rangle$ where $\hat{N} = \sum_j a_j^\dagger a_j$ where $N$ is a positive integer and $a_j^\dagger, a_j$ are the creation and annihilation operators which satisfy $[a_i, a_j^\dagger] = \delta_{ij}$. A single photon $N = 1$ has a two dimensional Hilbert space and two modes spanned by the Fock states $\{|1, 0\rangle, |0, 1\rangle\}$, often called a dual-rail qubit $\mathbb{PZ}$. Here, we consider the case of a single photon $N = 1$ photon with $d$ modes, hence a Hilbert space of dimension $d$ will be spanned by the Fock states of one photon in each mode.

A reliable single-photon source is required to produce Fock states described above. We will show that a weak single photon coherent state is sufficient for our purposes. In the case of a single mode, a coherent state is defined as

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle ,$$

where $\alpha$ is an arbitrary complex number. Coherent states do not have a fixed photon number but the average is bounded and equal to $|\alpha|^2$. We generate $d$ modes each in a coherent state

$$|\Psi\rangle = |\alpha_1\rangle \otimes |\alpha_2\rangle \otimes ... \otimes |\alpha_d\rangle = |\alpha_1, \alpha_2, ..., \alpha_d\rangle .$$

We can see that this state will be contaminated by undesired multi-photon Fock states. Furthermore linear optics can only transform coherent product states into coherent products states: no entanglement is possible.
It is possible to get around this restriction by introducing an imaginary measurement device that makes a total photon measurement on \( s \) modes without absorbing any photons or mode mixing. We can now ask: what is the conditional state if such a measurement is made on a \( s \)-fold product of coherent states, conditioned on the result of the measurement being \( N \)? We can write this conditional state as

\[
|\Psi : N\rangle = \mathcal{N} \Pi_N |\alpha_1, \alpha_2, \ldots, \alpha_s\rangle,
\]

where we are interested in the single photon case \( N = 1 \). In fact, there is no such non-absorbing measurement device, however we effectively achieve the same result by postponing all measurements to the end and summing the number of photons counted in each mode. We then post-select all the trials where the number is not the chosen integer \( N \). There are several detection problems we must consider including photon loss, non-photon-number-resolving counts and dark counts. Dark counts is the least important for modern APDs and in our experiment, introduce an 0.0001% error.

We can simulate the photon loss using a beam splitter model with transitivity \( \eta \) standing in for the role of non-unit quantum efficiency. Such a filter is a linear optical device that transforms the input annihilation operator \( a_k \rightarrow \sqrt{\eta} a + \sqrt{1 - \eta} b_k \) where \( b_k \) is an auxiliary mode that is initially in the vacuum state.

We can now model the photon loss for a weak coherent state where \( |\alpha|^2 \ll 1 \), expanded to second order that is

\[
|\alpha\rangle \approx e^{-|\alpha|^2/2} \left( |0\rangle + \frac{\alpha^2}{\sqrt{2}} |2\rangle \right)
\]

which we can substitute into Eq. (11) obtaining

\[
|\Psi\rangle \rightarrow |\Psi'\rangle
\]

Making the beam splitter transformations such that

\[
|\Psi'\rangle = \mathcal{N} \left( \sum_{i=1}^{d} \alpha_i a_i^\dagger + \frac{\alpha^2}{\sqrt{2}} \left( \sqrt{1 - \eta} a_i^\dagger b_i^\dagger + \sqrt{1 - \eta} \left( a_i^\dagger b_j^\dagger + a_j^\dagger b_i^\dagger \right) \right) \right) \otimes |0\rangle_{a,i} |0\rangle_{b,i},
\]

Assuming that mean photon number \( |\alpha|^2 \) is evenly distributed across the different modes up to multiplicative factor \( |\beta_j| < 1 \), then we can write \( \alpha_i = \alpha \beta_i \) where \( \sum_{j=1}^{d} |\beta_j|^2 = 1 \) leading to the normalisation constant

\[
\mathcal{N} = \exp \left( - \frac{\sum_{i=1}^{d} |\alpha \beta_i|^2}{2} \right) = e^{-|\alpha|^2/2}
\]

We can interpret \( |\beta_i|^2 \) as the probability amplitude associated with each \( k \) mode. We now want to calculate the conditional state in the case where \( N = 1 \) and trace out the \( b \) mode which we never observe. As a result, we obtain the mixed state

\[
|\Psi' : 1\rangle \langle \Psi' : 1| = p |\Psi\rangle \langle \Psi| + \frac{1 - p}{d} \rho
\]
FIG. 5. The guessing probability is only minutely corrupted by the presence of the contamination states in the single photon limit for a coherent state with $|\alpha|^2 = 0.5$. As we vary the quantum efficiency $\eta$ from 0.8 (dark blue) to 0 (light blue) we observe that the contamination introduce a $\Delta \% < 0.025$ defect in the guessing probability even in the worst case. This can be explained by the fidelity of the contamination state $\tilde{\rho}$ which is almost identical to desired state $\rho$. We again plot it for increasing quantum efficiency $\eta$ using the same colour scheme as the guessing probability. Even when the $\eta = 0$ and the contaminated states are equally as likely as the desired state, the effect on the measurement statistics is negligible.

where $p$ is the probability of observing the desired single photon Fock state $|\Psi\rangle\langle\Psi|$ and $\tilde{\rho}$ are the contamination states. The probability of obtaining the desired state $p$ is found to be

$$p = \left(1 + \frac{|\alpha|^2(1 - \eta)}{4}\left(\sum_{j=1}^{K} |\beta_j|^4 + 1\right)\right)^{-1}.$$  \hspace{1cm} (18)

In this expression, the probability of detecting the correct decreases for increasing mean photon number $|\alpha|^2$ and decreasing quantum efficiency $\eta$ as we would expect. Increasing the number of photon number, increases the likelihood of detecting a contamination state.

We can now compute the effect that non-unit quantum efficiency $\eta$ has on our results using the encoding quantum encoding $|\Psi_y\rangle$ shown for $|\alpha|^2 = 0.5$ in Figure 5. Here we plot the effect that the contamination states have on the absolute difference in guessing probability $\Delta \% = |p_{\text{guess}} - p'_{\text{guess}}|$ as a function of dimension. When $\eta = 1$ then we never detect any contamination states and $\Delta \% = 0$ for all $d$. However, decreasing to $\eta = 0.8$ we see this difference increasing until we reach $\eta = 0$ which maximises the effect of the contamination states. Even so, in this limit the percentage error is $\sim 0.02\%$ which is extremely small.

The perturbation to the guessing probability is small even when the contribution of the contamination states $\tilde{\rho}$ is maximised because they are very similar to the desired state $|\Psi\rangle\langle\Psi|$. This can be seen by plotting the fidelity $F = \text{tr}\left(\sqrt{\tilde{\rho}}\sqrt{\rho}\sqrt{\rho}\sqrt{\tilde{\rho}}\right)$ of the whole state $\tilde{\rho}$ against the desired state $\rho$ seen in Figure 5(right). The contaminated mixed state is very close to the desired state leading to the small perturbation to the guessing probability. Using the same values of $\eta$ as before, we see that the fidelity is always greater than $F > 0.995$. 
