
Article:
Banchi, Leonardo, Zhuang, Quntao and Pirandola, Stefano orcid.org/0000-0001-6165-5615 (2020) Quantum-enhanced barcode decoding and pattern recognition. Physical Review Applied. 064026. ISSN 2331-7019

https://doi.org/10.1103/PhysRevApplied.14.064026

Quantum-enhanced barcode decoding and pattern recognition

Leonardo Banchi,[1,2] Quntao Zhuang,[3,4] and Stefano Pirandola[5]

[1] *Department of Physics and Astronomy, University of Florence, via G. Sansone 1, I-50019 Sesto Fiorentino (FI), Italy*
[2] *INFN Sezione di Firenze, via G. Sansone 1, I-50019, Sesto Fiorentino (FI), Italy*
[3] *Department of Electrical and Computer Engineering, University of Arizona, Tucson, Arizona 85721, USA*
[4] *James C. Wyant College of Optical Sciences, University of Arizona, Tucson, Arizona 85721, USA*
[5] *Department of Computer Science, University of York, York YO10 5GH, UK*

(Dated: October 9, 2020)

Quantum hypothesis testing is one of the most fundamental problems in quantum information theory, with crucial implications in areas like quantum sensing, where it has been used to prove quantum advantage in a series of binary photonic protocols, e.g., for target detection or memory cell readout. In this work, we generalize this theoretical model to the multi-partite setting of barcode decoding and pattern recognition. We start by defining a digital image as an array or grid of pixels, each pixel corresponding to an ensemble of quantum channels. Specializing each pixel to a black and white alphabet, we naturally define an optical model of barcode. In this scenario, we show that the use of quantum entangled sources, combined with suitable measurements and data processing, greatly outperforms classical coherent-state strategies for the tasks of barcode data decoding and classification of black and white patterns. Moreover, introducing relevant bounds, we show that the problem of pattern recognition is significantly simpler than barcode decoding, as long as the minimum Hamming distance between images from different classes is large enough. Finally, we theoretically demonstrate the advantage of using quantum sensors for pattern recognition with the nearest neighbor classifier, a supervised learning algorithm, and numerically verify this prediction for handwritten digit classification.

I. INTRODUCTION

Quantum information science has undergone remarkable advances in the recent years, progressing in all its sub-fields of computing [1, 2], communication [3] and sensing [4, 5], both with discrete- and continuous-variable systems [6–8]. In this wide scenario, quantum sensing is arguably one of the most mature areas for near-term technological deployment. Its theoretical and experimental developments have been strongly based on quantum metrology [9, 10] and quantum hypothesis testing [11–14]. Especially, the latter approach has allowed to show a quantum advantage over classical strategies in tasks of target detection [15, 16] and data readout [17], modelled as binary problems of quantum channel discrimination.

The discrimination of quantum channels is an incredibly rich area of investigation [18, 19], with unexplored consequences but also non-trivial difficulties. It represents a double optimization problem where both input states and output measurements need to be varied. Furthermore, in the bosonic setting, it has to be formulated as an energy-constrained problem, where the mean number of input photons is limited to some finite, small, value. In such a scenario, the central question is that of showing quantum advantage: Can truly-quantum states, e.g., entangled, lead to an advantage over classical, i.e. coherent, states? Addressing this question in the multi-ary case is difficult, since the theory is missing powerful tools that are instead available for the binary case.

In this work, we take a step forward by developing the theory of quantum hypothesis testing for the multi-ary setting of barcode decoding and pattern classification. We start from a general model of digital image, where each pixel is described by an ensemble of quantum channels defined over a finite alphabet. We specialize to the case where the single-pixel alphabet is binary, so that there are 2^n possible hypothesis or configurations for an n-pixel barcode. We then show how the use of quantum sources of light, based on entangled states, can clearly outperform classical strategies based on coherent states for the readout of the barcode configuration, i.e., the retrieval of its data. In particular, we derive an analytical condition for the maximum number of pixels or the minimum number of probings such that quantum advantage is obtained. This result holds not only for a uniform distribution of the possible configurations, but also when data is stored by the position of k white pixels among a grid of otherwise black pixels.

Besides data readout or barcode decoding, we consider the general problem of pattern recognition, where the task is to classify an image, e.g. a handwritten digit, without necessarily reconstructing it pixel by pixel. Here the image distribution is not uniform and generally unknown, and optimal classification has to be approximated via a collection of correctly classified examples, following supervised learning strategies [20]. We consider the ultimate limits of this procedure, where we may optimize over the optical circuit, measurements and subsequent classical post-processing algorithms. Introducing relevant bounds, we theoretically prove that this problem is significantly simpler than that of barcode decoding, as long as the minimum Hamming distance between images from different classes is large enough. Moreover, we

show how clear quantum advantage can be obtained as a function of the number of training data. Finally, we consider a simplified scheme for recognizing black and white patterns, such as digital images of handwritten digits, by means local measurements followed by a classical nearest neighbor classifier. More specifically, we apply this classifier to the measurement outcomes that are obtained by either using entangled states or coherent states at the input of the grid of pixels. We are able to show a clear quantum advantage which holds even when we employ sub-optimal photon counting measurements for the quantum case, which are particularly relevant for near-term experiments. The advantage becomes particularly evident at relatively small energies where a total of a few hundred of photons are irradiated over each pixel.

The paper is organized as follows. In Section II we discuss the problem of barcode decoding and show the possible advantage in using quantum detectors with entangled input states. In Section III we discuss the related problem of pattern recognition, showing a similar advantage. Discussions are drawn in Section IV.

II. BARCODE DECODING

A. Quantum mechanical model of a digital image

A basic imaging system irradiates light over an array of pixels which can be read in transmission or in reflection. From the ratios between input and output intensities, the system generates a corresponding array of grey-levels that constitutes a monochromatic image. In a quantum mechanical setting, each pixel can therefore be modeled as a bosonic lossy channel $\mathcal{E}_i$ whose transmissivity depends on the grey-level i. This lossy channel can be probed by an input state (with some limited energy) and a corresponding output measurement, generally described by positive-operator valued measure (POVM). Finally, the outcome is processed by a decision test that identifies the channel and, therefore, the grey-level i.

Let us formalize the problem in more mathematical detail, which can be seen as a multi-ary and multi-pixel generalization of the basic model of quantum reading [17]. Assume that each pixel is described by a channel ensemble $\{\mathcal{E}_i\}$ spanned by the label $0 \leq i \leq C-1$, where C is the discrete number of grey-levels that can be assumed by the pixel. Let us define an *image* over n pixels as a sequence $\boldsymbol{i} := i_0, \cdots, i_{n-1}$, together with an associated probability distribution $\pi_{\boldsymbol{i}}$, which is simply $\pi_{\boldsymbol{i}} = C^{-n}$ in the uniform case. The global channel describing the entire array of n pixels is the tensor product $\mathcal{E}_{\boldsymbol{i}}^n := \mathcal{E}_{i_0} \otimes \cdots \otimes \mathcal{E}_{i_{n-1}}$. Thus, an image can equivalently be represented by an ensemble of multi-channels $\{\pi_{\boldsymbol{i}}, \mathcal{E}_{\boldsymbol{i}}^n\}$.

In order to read the image, let us assume that we have a generic $2n$-mode state $\tilde{\rho}$ at the input: n signal modes are sent through the pixels, while n idler modes are used to help the measurement. At the output, there is an ensemble of possible states $\{\pi_{\boldsymbol{i}}, \rho_{\boldsymbol{i}}\}$ where $\rho_{\boldsymbol{i}} := \mathcal{I}^n \otimes \mathcal{E}_{\boldsymbol{i}}^n(\tilde{\rho})$. In the case of a classical transmitter, the signal modes are prepared in coherent states while the idler modes are in vacuum states. In the case of a quantum transmitter, signal and idler modes are entangled pairwise. In particular, each signal-idler pair is described by a two-mode squeezed vacuum (TMSV) state.

In general, we probe the image $\boldsymbol{i}$ with identical inputs for M times, leading to the overall input state $\tilde{\rho}^{\otimes M}$ and corresponding output $\rho_{\boldsymbol{i}}^{\otimes M} := [\mathcal{I}^n \otimes \mathcal{E}_{\boldsymbol{i}}^n(\tilde{\rho})]^{\otimes M}$. We measure this output with a collective POVM, where each measurement operator $\Pi_{\boldsymbol{i}'}$ represents the decision that the image is $\boldsymbol{i}'$. Because input states are energy-constrained, there will be readout errors described by the conditional probabilities

$$p_{\text{read}}(\boldsymbol{i}'|\boldsymbol{i}) = \text{Tr}\left(\Pi_{\boldsymbol{i}'} \rho_{\boldsymbol{i}}^{\otimes M}\right) \ . \tag{1}$$

By including the priors $\{\pi_i\}$, we may therefore define the success probability or, equivalently, the error probability

$$p_{\text{succ}} := \sum_{\boldsymbol{i}} \pi_{\boldsymbol{i}} p_{\text{read}}(\boldsymbol{i}|\boldsymbol{i}), \quad p_{\text{err}} = 1 - p_{\text{succ}}. \tag{2}$$

Using Refs. [21, 22] and the multiplicativity of the fidelity over tensor products, one finds that the minimum error probability (optimizd over POVMs) satisfies

$$\frac{1}{2} \sum_{\boldsymbol{i} \neq \boldsymbol{j}} \pi_{\boldsymbol{i}} \pi_{\boldsymbol{j}} F_{\boldsymbol{i}:\boldsymbol{j}}^{2M} \leq p_{\text{err}} \leq \sum_{\boldsymbol{i} \neq \boldsymbol{j}} \sqrt{\pi_{\boldsymbol{i}} \pi_{\boldsymbol{j}}} F_{\boldsymbol{i}:\boldsymbol{j}}^{M}, \tag{3}$$

where

$$F_{\boldsymbol{i}:\boldsymbol{j}} := F(\rho_{\boldsymbol{i}}, \rho_{\boldsymbol{j}}) = \|\sqrt{\rho_{\boldsymbol{i}}}\sqrt{\rho_{\boldsymbol{j}}}\|_1 = \text{Tr}\sqrt{\sqrt{\rho_{\boldsymbol{i}}}\rho_{\boldsymbol{j}}\sqrt{\rho_{\boldsymbol{i}}}} \tag{4}$$

is the fidelity between two generic single-probing multi-pixel output states, $\rho_{\boldsymbol{i}}$ and $\rho_{\boldsymbol{j}}$. The inequalities in Eq. (3) bound the performances of a pretty good measurement [23–25] and have no explicit dependence on the dimension of the Hilbert space, so that they hold for bosonic states as long as these states are energy-constrained. Below, we build on these inequalities to derive our bounds for decoding barcodes.

B. Barcode discrimination

An important case of the general problem discussed in the previous section is barcode decoding, whose schematic setup is shown in Fig. 1. A *barcode* is either a one-dimensional (1D) or two-dimensional (2D) grid of pixels with two possible colors, black (B) or white (W). With a slight abuse of jargon, we call *pixels* the elements of the 1D or 2D grid that defines a barcode. In the 1D case (Fig. 1a), a pixel is a black or white vertical bar, while in the 2D case (Fig. 1b) a pixel is an elementary square. It is worth noting that many of the conclusions drawn for barcode decoding can be extended to more general images. Indeed, a higher number of grey-levels

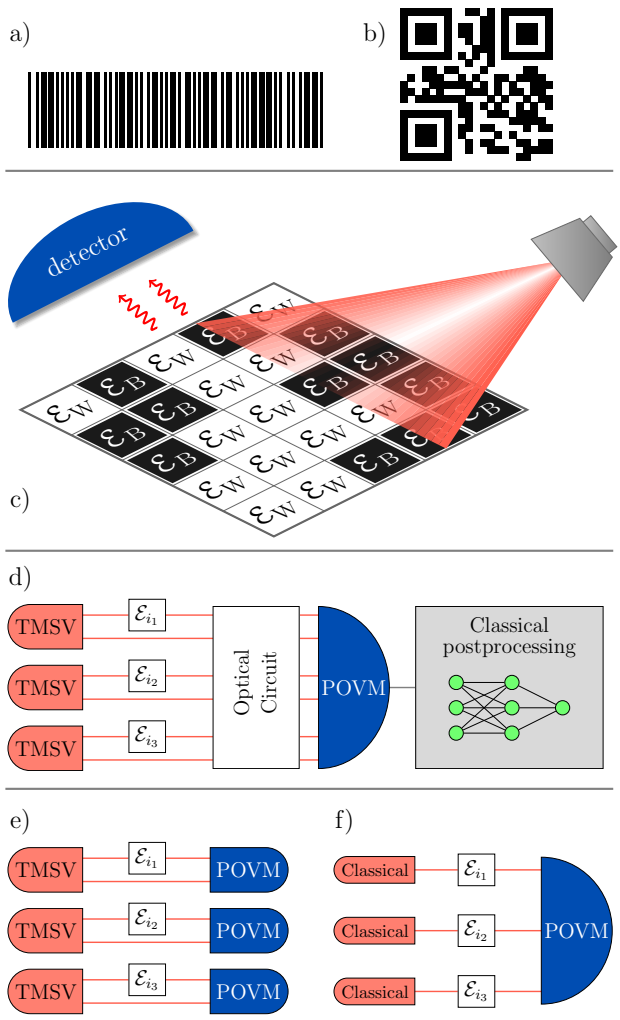

FIG. 1. Barcode decoding. Examples of 1D (a) and 2D (b) barcodes. (c) Schematic physical setup for decoding a 2D barcode with $n = 5 \times 5$ pixels. A source shines light on the grid of pixels, each modeled by a quantum channel, which is either $\mathcal{E}_B$ or $\mathcal{E}_W$ depending on the pixel grey-level, black (B) or white (W). The reflected/scattered light is collected by a detector, which aims at recognizing *all* pixel values depending on the detected photons. In a quantum setup, the signal modes shined over the pixels are entangled with idler modes that are directly sent to the detector for a joint measurement. (d) General theoretical setup scheme for optimal barcode discrimination using entangled TMSV states, where the mixing optical circuit, the measurements and following classical post-processing must be optimized. (e,f) Special setup to claim quantum advantage: we compare independent entangled TMSV states on each pixel followed by independent *local* measurements (e) with classical coherent sources followed by global measurements. Quantum advantage is claimed whenever (e) beats (f).

$C > 2$ can always be formally mapped into a barcode. For instance, $C = 256$ corresponds to an 8-bit grey scale and each bit can be represented as a binary variable with two possible configurations (B or W, by convention). As such, images with $C > 2$ can be mapped into a "barcode" image with a higher number of pixels.

The general problem of barcode discrimination can be depicted as in Fig. 1c and 1d. According to our notation, each pixel of a barcode has two possible grey-levels $i \in \{B, W\}$ and therefore corresponds to two possible quantum channels $\mathcal{E}_B$ and $\mathcal{E}_W$. For barcode decoding, we assume that the pixels are independently probed, so that the input state takes the tensor-product form $\tilde{\rho} = \rho_0^{\otimes n}$. Note that this assumption does not reduce the generality of our treatment. In fact, for the quantum source this leads to one of the best possible choices (tensor product of TMSV states). For the classical source, we know that independent and identical coherent states are able to saturate the lower bounds for general mixtures of multimode coherent states [19]. As for detection, the general scheme to correctly distinguish the various configurations consists in choosing a mixing optical circuit, followed by measurements and classical post-processing algorithms as in Fig. 1d. From an operational point of view, a suboptimal solution can be found for this problem by restricting to a cascade of beam splitters and phase shifter with tunable parameters, followed by independent measurements (e.g. homodyne or photodetection), similar to that of Ref. [18]; while for the classical post-processing we may employ statistical classification algorithms commonly employed in machine learning applications, e.g. based on neural networks [20]. The suboptimal solution is then numerically investigated by minimizing the parameters of the optical and neural networks in order to minimize $p_{\rm err}$. When photodetection measurements are employed, analytic gradients can be computed following Ref. [26] to speed-up the optimization algorithm. In this paper however we focus on the most general case and study the fundamental limits of barcode decoding and patter recognition, introducing different theoretical limits that any possible scheme must satisfy. Indeed, the physical optical circuit and measurements, and also the classical post-processing algorithm in Fig. 1d, can all be reabsorbed into an abstract POVM that must be optimized.

We start by considering the case where a barcode with n pixels is prepared in one of all possible 2^n patterns, each with equal prior. Then, we will consider the case where the patterns are restricted to specific configurations, where k white pixels are randomly positioned within a grid of otherwise black pixels.

Starting from the input state $\tilde{\rho} = \rho_0^{\otimes n}$, the possible states at the output of the barcode $\rho_{\boldsymbol{i}} = \mathcal{I}^n \otimes \mathcal{E}_{i_1} \otimes \cdots \otimes \mathcal{E}_{i_n}(\tilde{\rho})$, take the product form

$$\rho_{\boldsymbol{i}} = \bigotimes_{k=1}^{n} \rho_{i_k}. \tag{5}$$

Correspondingly, the fidelities can be simplified as

$$F(\rho_{\boldsymbol{i}}, \rho_{\boldsymbol{j}}) = F(\rho_W, \rho_B)^{\text{hamming}(\boldsymbol{i},\boldsymbol{j})}, \tag{6}$$

where $\rho_i = \mathcal{I} \otimes \mathcal{E}_i(\rho_0)$ for $i \in \{B, W\}$ and $\text{hamming}(\boldsymbol{i}, \boldsymbol{j})$ is the Hamming distance between the two binary images

$\boldsymbol{i}$ and $\boldsymbol{j}$, namely the number of pixels in which the two images differ. Using the properties of the Hamming distance, in Appendix A we show that, for uniform *a priory* probabilities, the M-probing bounds (3) become

$$\frac{(F_{\max}^{2M}+1)^n-1}{2^{n+1}} \leq p_{\text{err}} \leq 1-\left(1-\frac{F_{\max}^{M}}{2}\right)^n, \qquad (7)$$

where $F_{\max}$ is the fidelity between any two (different) images with minimum Hamming distance [cf. Eq. (6)].

The minimum Hamming distance is achieved when the two images differ by a single pixel. Thus, we get $F_{\max} = F(\rho_B, \rho_W)$, namely the maximum fidelity between any two images is given by the fidelity between the states describing the grey-levels of a single pixel. Using the Bernoulli's inequality, we then simplify Eq. (7) as

$$\frac{n}{2^{n+1}}F_{\max}^{2M} \leq p_{\text{err}} \leq \frac{n}{2}F_{\max}^{M}\ . \qquad (8)$$

In Appendix B, we show that the upper bound can be achieved using *local* measurements, namely where each pixel is measured independently from the others and $\Pi_{\boldsymbol{i}} = \bigotimes_{j=0}^{n-1}\Pi_{i_j}$ in Eq. (1), though each pixel and its respective idler may be measured together (see Fig. 1d). Once we restrict to local operations, the optimum is achieved by independent Helstrom measurements [27] and the upper bound in Eq. (8) follows from Fuchs–van de Graaf inequalities [28]. A sub-optimal local measurement is obtained by combining the signal and idler via a beam splitter followed by independent measurements [17]. Moreover, in the supplementary material we also discuss different inequalities on p_{err} based on the multiple quantum Chernoff bound [29, 30].

Two interesting observations can be made from the bounds (8). First, the upper bound for the error probability becomes small whenever $nF^M(\rho_W, \rho_B) \ll 1$. This implies that, although the set of images (namely barcode configurations) grows exponentially with the number of pixels as 2^n, the required fidelities to accurately distinguish all configurations should decrease polynomially with $1/n$. In particular, $M = \mathcal{O}(\log n)$ copies are needed for correct discrimination. The second observation is that, due the factor 2^{-n}, the lower bound in Eq. (8) decreases exponentially with n. As we show in Appendix B, this factor disappears from the lower bound when local measurements are employed. It is known that, in general, optimum mixed state discrimination requires a joint measurement [31, 32], yet in our setting optimal global measurements may in principle exponentially reduce the probability of error. Nonetheless, it is currently an open question to verify whether and exponentially decreasing error is achievable with optimal quantum measurements. In the next section we will claim quantum advantage whenever the upper bound on p_{err} obtained with entangled states and local measurements is smaller than the lower bound on p_{err} obtained with classical states and possibly global measurements, as schematically shown in Figs. 1d) 1e). Therefore, if the lower bound in (7) is loose, the regimes for quantum advantage are larger.

In the previous bounds we considered a uniform distribution of black an white pixels in the barcode. We may also consider a different encoding with a fixed number of white pixels, generalizing the results of Ref. [19]. The task is then to find the position of k white pixels in a barcode with n bars. The number of possible configurations is $\binom{n}{k} \approx 2^{nH(k/n)}$ where H is the binary entropy function and the approximation holds when both n and k are large. Therefore, in that regime, the configuration space grows exponentially with n, as in the uniform case discussed above. In the asymptotic regime we obtain the following bounds

$$\frac{k(n-k)}{2^{nH(k/n)+1}}F_{\max}^{4M} \lesssim p_{\text{err}}^{k-\text{whites}} \lesssim k(n-k)F_{\max}^{2M}, \qquad (9)$$

while the exact expressions for finite M, n and k are discussed in Appendix C.

C. Quantum enhancement

We now discuss the regime where we get an advantage from using entangled photons as input. We compare the case where each input ρ_0 is a TMSV state $|\Phi_{N_S}\rangle$ with N_S average photons and the case where the input is a coherent state with the same number of signal photons $\left|\sqrt{N_S}\right\rangle \otimes |0\rangle$ (where the vacuum state means that no idler is used). Note that one can replace the vacuum idler with an arbitrary state, such as a strong local oscillator in a coherent state, however that will not give a better performance when the optimum measurement is considered. Assuming M probings of the barcode, we have a total of $N_{\text{tot}} = MN_S$ mean photons irradiated over each pixel. According to the analysis from the previous section, provable quantum advantage can be achieved whenever the upper bound from Ineqs. (8), obtained with TMSV input states, is less than the lower bound obtained with coherent state inputs. Since the upper bound in (8) is obtained with local measurements, what we call "provable advantage" means that possibly non-optimal local measurement strategies with entangled inputs beat any strategy with coherent states, even when the latter is enhanced by complex global measurements. Provable quantum advantage may be more difficult for larger n, given the exponentially decreasing factor in the lower bound of Eq. (8), but here we show that it can be achieved for every number of pixels n with suitably large number of probings M.

Using the formula for the fidelity between two generally-mixed Gaussian states [33, 34], for TMSV states at the input, we compute (see Appendix D)

$$F_q(\rho_W, \rho_B)^M = \left(\frac{1}{1+N_S\Delta_q}\right)^M \geq e^{-MN_S\Delta_{\text{q}}}, \qquad (10)$$

where the index q stands for *quantum* and

$$\Delta_{\text{q}} = 1-\sqrt{(1-\eta_W)(1-\eta_B)}-\sqrt{\eta_W\eta_B}. \qquad (11)$$

For a coherent-state input, we instead have

$$F_{\rm c}(\rho_W,\rho_B)^M = e^{-MN_S\Delta_{\rm c}}, \quad \Delta_{\rm c} = \frac{(\sqrt{\eta_B}-\sqrt{\eta_W})^2}{2}. \tag{12}$$

By comparing Eqs. (10) and (12) we see that, for fixed M, the fidelity between coherent states displays an exponential decay as a function of N_S, while for quantum states we see a polynomial decay in N_S. Nonetheless, for large M and small N_S, the inequality in Eq. (10) becomes tight and, since $\Delta_{\rm q} \geq \Delta_{\rm c}$, in that limit we find that quantum light always provides an advantage for discrimination, irrespective of the values of η_W and η_B. The limits of small N_S and large M are widely employed to show quantum advantage and can be realized experimentally with little imperfections [35]. Therefore, from now on we will focus on such limits, $M \to \infty$ and $N_S \to 0$, while keeping fixed the total mean number of photons MN_S irradiated over each pixel.

To properly demonstrate the advantage, we need to show that the upper bound on the probability of error using quantum light is smaller than the lower bound on the probability of error using coherent states. From Ineqs. (8), we see that this happens when $F_{\rm c}^{2M} \geq 2^n F_{\rm q}^M$. Setting $n = \nu M N_S$, the previous inequality implies that quantum advantage is obtained for

$$\nu \leq \nu_{\rm th} = \frac{\Delta_{\rm q} - 2\Delta_{\rm c}}{\log 2}, \tag{13}$$

which is a barcode multi-pixel generalization of the "threshold energy" theorem proven in the context of single-cell quantum reading [17].

According to Eq. (13), whenever the number n of pixels is smaller than a certain threshold, entangled light always provides an advantage in the discrimination of barcode configurations (barcode decoding) with respect to the best classical strategy with the same signal energy, even when the latter uses possibly complex global measurements. The behaviour of $\nu_{\rm th}$ as a function of η_W and η_B is numerically shown in Fig. 2.

Quantum advantage can also be proven when we consider a prior distribution for the barcode configurations that is non-uniform, more precisely for the case where the number k of white pixels is fixed. Using Ineqs. (9), we find that there is a provable quantum advantage when $F_{\rm c}^{4M} \geq 2^{nH(k/n)+1} F_{\rm q}^{2M}$, namely when $nH(k/n)+1 \leq 2\nu_{\rm th} M N_S$. Therefore, as in the previous case, quantum advantage may be observed when the number of pixels is sufficiently small or the number of probes M is sufficiently large, as long as $\nu_{\rm th} \geq 0$.

It is currently an open question to prove whether or not the lower bound in (8) can be achieved when classical light is employed. Nonetheless, our analysis shows that even assuming that such bound can be achieved with classical inputs, a strategy based on entangled light and the much simpler local measurements can beat any approach based on coherent states. On the other hand, if only local measurements can be performed, then the factor 2^{-n} in the lower bound (8) disappears (see Appendix B). This corresponds to the case $\nu = 0$. Therefore, in that case, whenever $\nu_{\rm th} > 0$, namely when $\eta_W > 1 - \eta_B$, quantum light provides an advantage for decoding uniformly-distributed barcodes, irrespective of n.

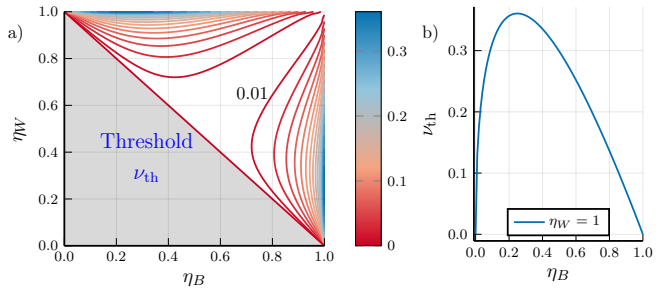

FIG. 2. Regimes of provable quantum advantage. a) Threshold value from Eq. (13) as a function of η_W and η_B. The threshold $\nu_{\rm th}$ is negative in the filled gray area and positive for $\eta_W > 1 - \eta_B$. Contours are from 0.01 in steps of 0.02. b) Threshold $\nu_{\rm th}$ for $\eta_W = 1$. Whenever $n \leq MN_S\nu_{\rm th}$, or similarly $M \geq \frac{n}{N_s\nu_{\rm th}}$, entangled light beats classical strategies based on coherent states.

III. PATTERN RECOGNITION

A. Statistical pattern classification

We now focus on the problem of pattern recognition. Consider the problem of recognizing handwritten digits as shown in Fig. 3a, whose images have been adapted from the MNIST dataset [36]. Each image depicts a single handwritten digit and the task is to extract from the image the corresponding number 0-9. From an algorithmic perspective, this task is more complex than the mere decision of whether a pixel is black or white but, from a physical point of view, this problem is actually simpler as errors are tolerated. Indeed, a human is able to instantly recognize all the numbers in Fig. 3a even when some of the pixels are randomly flipped. Therefore, for reliable pattern recognition, it is not necessary to perfectly reconstruct the entire image. Compared to the barcode configurations of Fig. 1, where each pixel provides important information, here the goal is to recognize a global property that is robust against individual pixel errors, which means that entirely different strategies are possible.

In statistical learning theory [37], different learning tasks, such as image classification, can be modeled using probabilities. We consider the abstract space of all possible images and define the probability $\pi_{\boldsymbol{i}}$ of getting the image $\boldsymbol{i}$ – this is unknown and generally not uniform. Image classification is a rule that attaches a certain label c, or class, to a given image $\boldsymbol{i}$. If this rule is deterministic, then it can be modeled via a function $c = f(\boldsymbol{i})$ but, more generally, the strategy is stochastic: given a certain

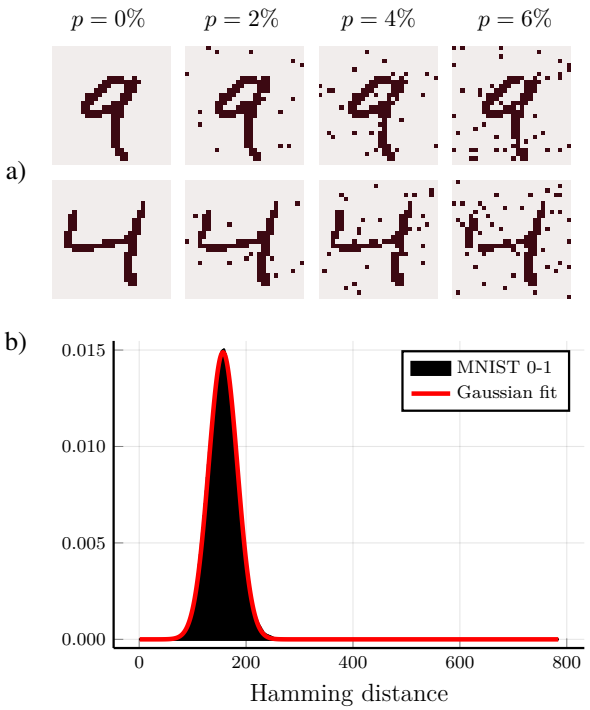

FIG. 3. Pattern recognition. (a) Images from the MNIST dataset without pixel recognition error ($p = 0\%$) and with pixel error probabilities $p = 2\%, 4\%, 6\%$, where each pixel is randomly flipped with probability p. (b) Probability $P_{cc'}(h)$ that one image from class c has Hamming distance h, with $0 \leq h \leq 28\times28$, from another image from class c'. The empirical histogram is evaluated for images from the MNIST dataset that correspond to digits 0 and 1. The Gaussian fit has mean $\mu_{01} \simeq 157$ and standard deviation $\sigma \simeq 27$. Different digits show a similar behaviour with $110 \lesssim \mu_{cc'} \lesssim 167$, where the minimum is achieved between 1 and 7.

image, the rule predicts different possible classes c with a probability distribution $P(c|\boldsymbol{i})$. Let us consider a pair $(c, \boldsymbol{i})$ and assume that, given our data, we have built a classifier $\tilde{c}(\boldsymbol{i})$ that assigns a certain class $\tilde{c}(\boldsymbol{i})$ to the image $\boldsymbol{i}$. The error in our classification can be described by a loss matrix with elements $L_{c\tilde{c}}$ that models the error of misclassification. The common choice is the 0-1 loss with $L_{cc} = 0$ and $L_{c\tilde{c}} = 1$ for $\tilde{c} \neq c$. By using the conditioning $P(c, \boldsymbol{i}) = P(c|\boldsymbol{i})\pi_{\boldsymbol{i}} = P(\boldsymbol{i}|c)P(c)$, the expected classification error can be written as

$$E = \mathop{\mathbb{E}}_{(c,\boldsymbol{i})\sim P(c,\boldsymbol{i})}[L_{c,\tilde{c}(\boldsymbol{i})}] = 1 - \sum_{\boldsymbol{i}} P(\tilde{c}(\boldsymbol{i})|\boldsymbol{i})\,\pi_{\boldsymbol{i}} \ . \qquad (14)$$

For known $P(c, \boldsymbol{i})$ the optimal classifier is then the one minimizing the expected classification error, $\tilde{c}_{\rm B}(\boldsymbol{i}) = \mathrm{argmax}_c\, P(c|\boldsymbol{i})$, which is called *Bayes classifier*, while the resulting error from (14), is called *Bayes rate*. The Bayes rate represents the theoretical minimum error that can be expected with the optimal classifier.

We now study the error of pattern classification when images are noisy, for instance due to an imperfect detection. The setup is the same of Fig. 1, where light, either quantum or classical, is used to illuminate the pattern (e.g. a handwritten digit as in Fig. 3a) and, from the detected output, the task is to find the correct class (e.g. a number between 0-9). For this purpose, we introduce the minimum error as a generalization of Eqs. (14) and (1)

$$E^{\rm Q} := \min_{\{\Pi_c\}} \sum_{c\neq\tilde{c}} \sum_{\boldsymbol{i}} \mathrm{Tr}\big[\Pi_{\tilde{c}} \rho_{\boldsymbol{i}}^{\otimes M}\big] P(c, \boldsymbol{i}) \ , \qquad (15)$$

where the operators $\{\Pi_c\}$ define a POVM whose measurement outcome c predicts the class of the image $\boldsymbol{i}$ encoded into the quantum state $\rho_{\boldsymbol{i}}$. For a two-class decision problem, the optimal POVM can be explicitly found by Helstrom theorem [27]. When the number of classes is larger, a "pretty good" approximation to the optimal POVM can be obtained with pretty good measurements, so that we may derive bounds similar to (3), generalizing [21, 22, 38] (see Appendix E for details).

$$B[F(\rho_B, \rho_W)^{2M}] \leq E^{\rm Q} \leq 2KB[F(\rho_B, \rho_W)^{M}], \qquad (16)$$

where K is such that K^{-1} is the minimum non-zero value of $\sqrt{P(c, \boldsymbol{i})P(c', \boldsymbol{i}')}$, which is independent of M, ρ_W and ρ_B, and we have defined

$$B[F] := \frac{1}{2} \sum_{c\neq c'} P(c)P(c') \sum_{h=1}^{n} P_{cc'}(h) F^h, \qquad (17)$$

where $P_{cc'}(h)$ is the probability that two images from different classes c and c' have Hamming distance h. For large M, the term with minimum Hamming distance dominates and we may write

$$B[F^M] \propto F^{M h_{\rm min}} \ , \qquad (18)$$

where $h_{\rm min}$ is the minimum Hamming distance between two images from different classes. The Ineqs. (16) and the expansion (18) represent the most important results of this section, generalizing Ineqs. (8) and (9) to the problem of pattern recognition. By comparing those bounds, we find that quantum-enhanced pattern recognition is significantly simpler than barcode discrimination when $h_{\rm min} > 1$, as the error decreases with the faster rate (18).

The error $E^{\rm Q}$ is a quantum generalization of Bayes rate, and quantifies the theoretical optimal performance of the classification rule. However, alike the Bayes rate, it is difficult to compute since the distribution $P(c, \boldsymbol{i})$ is typically unknown, and no closed-form solutions to (15) exist beyond the two-class case. To solve these issues, in the next section we propose a supervised learning approach where an optimal classification measurement is estimated from a collection of correctly classified data.

B. Supervised quantum pattern recognition

In data driven approaches the task is to approximate the optimal classifier via a collection of already classified examples $(c_k^{\mathcal{T}}, \boldsymbol{i}_k^{\mathcal{T}})$. The set $\mathcal{T} = \{(c_k^{\mathcal{T}}, \boldsymbol{i}_k^{\mathcal{T}})$ for $k =$

$1,\dots,T\}$ is called *training set* and T is its cardinality. In the framework of statistical learning theory, we can treat the elements of this set as *samples* from the abstract and unknown joint probability distribution $P(c,\boldsymbol{i})$ introduced above. Then, in the limit of large T we may approximate the averages with respect to $P(c,\boldsymbol{i})$ with *empirical* averages over the training set. This allows us to explicitly compute the classification error (15) and the theoretical bounds (16). Therefore we define an empirical learning method, also called "training", as an optimization of the POVM $\{\Pi_c\}$ to correctly classify, as much as possible, the known samples from the training set $\mathcal{T}$

$$\text{training}: \quad \min_{\{\Pi_c\}} \frac{1}{T}\sum_{k=1}^{T}\sum_{c\neq c_k^{\mathcal{T}}} \mathrm{Tr}\left[\Pi_c \rho_{\boldsymbol{i}_k^{\mathcal{T}}}^{\otimes M}\right] =: E_{\mathcal{T}}^{\mathrm{Q}}. \tag{19}$$

From an operational point of view, a suboptimal solution to optimal detection $\{\Pi_c\}$ can be found for instance as shown Fig. 1d and discussed in section II, by optimizing over the available optical circuit, measurement schemes and classical post-processing. Here on the other hand we study the ultimate theoretical limits that any classification task must satisfy, studying the minimum training error $E_{\mathcal{T}}^{\mathrm{Q}}$ via bounds like (16), while the ability to classify unseen data will be discussed in the next section. Indeed, upper and lower bounds on $E_{\mathcal{T}}^{\mathrm{Q}}$ can be obtained (see Appendix E) as an average fidelity between states $\rho_{\boldsymbol{i}_k^{\mathcal{T}}}$ and $\rho_{\boldsymbol{i}_{k'}^{\mathcal{T}}}$ whose images from the training set have different classes, $c_k^{\mathcal{T}} \neq c_{k'}^{\mathcal{T}}$. Thanks to Eq. (6) we finally get

$$B_{\mathcal{T}}[F(\rho_B,\rho_W)^{2M}] \leq E_{\mathcal{T}}^{\mathrm{Q}} \leq 2T B_{\mathcal{T}}[F(\rho_B,\rho_W)^{M}], \tag{20}$$

where we have defined

$$B_{\mathcal{T}}[F] = \sum_{k,k':c_k^{\mathcal{T}}\neq c_{k'}^{\mathcal{T}}} \frac{F^{\mathrm{hamming}(\boldsymbol{i}_k^{\mathcal{T}},\boldsymbol{i}_{k'}^{\mathcal{T}})}}{2T^2}. \tag{21}$$

It is simple to show that $B_{\mathcal{T}}[F]$ is a particular case of $B[F]$ from Eq. (17) in which averages over the abstract distribution are substituted with averages over the empirical distribution. As such, we may rewrite $B_{\mathcal{T}}$ as in Eq. (17) and obtain the large-M scaling (18).

As a relevant example, we consider the problem of handwritten digit classification with the MNIST dataset [36]. The MNIST dataset is composed of a training set of 60000 images and corresponding classes, and a testing set of 10000 images and corresponding classes. Each original image is in grey scale and has $n = 28\times28$ pixels. For simplicity we first map each pixel to either black or white, depending on the closest grey-level. In this way, every image can be seen as a 2D barcode. For the MNIST dataset we see from Fig. 3b) that the probability $P_{cc'}(h)$ that two images from different classes have Hamming distance h resembles a Gaussian distribution with mean $\mu_{cc'}$ and standard deviation $\sigma_{cc'}$, and minimum non-zero value $h_{cc'}^{\min}$. Using this approximation, we find in Appendix E analytical approximations for $B_{\mathcal{T}}[F]$, recovering the scaling (18), where $h_{\min} = \min_{c\neq c'} h_{cc'}^{\min}$. For the MNIST dataset, we find $h_{\min} = 25$. Therefore, from (20) we may get an error that decays as $E_{\mathcal{T}}^{\mathrm{Q}} \approx F(\rho_B,\rho_W)^{\alpha M h_{\min}}$, independently on the number of pixels n and with $1 \leq \alpha \leq 2$. Moreover, thanks to Ineqs. (20) we may define a guaranteed quantum advantage when the upper bound obtained with entangled states is smaller than the lower bound obtained with classical data, namely when $2TF_q^{Mh_{\min}} \leq F_c^{2Mh_{\min}}$. Since the training set is normally very large, we may set $2T = 2^{\nu M h_{\min} N_S}$ for some ν and the above inequality becomes equivalent to (13), in the limit $M \to \infty$ and $N_S \to 0$. Therefore, we may repeat the same analysis of Sec. II C: whenever $\nu_{\mathrm{th}} > 0$ (see Fig. 2), quantum advantage can be proven for training sets whose dimension is bounded as $2T \leq 2^{\nu_{\mathrm{th}} M N_S h_{\min}}$. In other terms, setting $N_{\mathrm{tot}} = MN_S$ we find a simple relation between the number of photons to show quantum advantage and the dimension of the training set as

$$N_{\mathrm{tot}} \geq \frac{\log_2(2T)}{\nu_{\mathrm{th}} h_{\min}} \simeq 0.65\,\nu_{\mathrm{th}}^{-1}\ . \tag{22}$$

In the above expression the first inequality holds in general, while the approximated numerical value is for the MNIST dataset, where $h_{\min} = 25$ and $T = 6\times10^4$.

To conclude this section we note that unlike (7), the upper bound in (20) is achieved with global measurements, so a strategy like the one in Fig. 1f may be needed to achieve such classification accuracy. Bounds with local measurement errors are discussed in the next section, where each pixel is detected independently.

C. Independent on-pixel measurements

In the previous section we have studied the ultimate physical limits for pattern recognition by optimizing over all the elements of the optical apparatus, namely the optical circuit, the measurements and the classical post-processing routines (Fig. 1c). Together these can all be described as an abstract global POVM, as in Eq. (15). Here we consider a simplified setup, similar to that of Fig. 1c but without the optical circuit and with local measurements $\Pi_{\boldsymbol{i}} = \prod_{j=1}^{N} \Pi_{i_j}$. Here a noisy image is reconstructed first, and then a classical algorithm is used to classify it. As before, we call $\boldsymbol{i}$ the real physical configuration of the n pixels, each either black or white $i_j = \{B, W\}$, and $\tilde{\boldsymbol{i}}$ the binary variables corresponding to the reconstructed image, read by the sensors. Using M copies to perform the detection, all possible reconstructed images can appear with probability $p_{\mathrm{read}}(\tilde{\boldsymbol{i}}|\boldsymbol{i})$ as in Eq. (1). Considering also the classical classification routine, the local setup consists in choosing a non-optimal POVM in Eqs. (15) or (19) as

$$\Pi_c = \sum_{\boldsymbol{i}} A(c|\tilde{\boldsymbol{i}}) \prod_{j=1}^{n} \Pi_{\tilde{i}_j}\ , \tag{23}$$

where $A(c|\boldsymbol{i})$ is any reliable (possibly non-linear) machine learning algorithm that can classify the reconstructed images. The above equation defines a POVM as long as $\sum_c A(c|\boldsymbol{i}) = 1$ for all $\boldsymbol{i}$, which is an obvious requirement since every image must be in at least one class.

The classical algorithm must be noise resilient, because some pixels might not be properly reconstructed, see e.g. Fig. 3a. Noise naturally occurs in readouts that are made in reflection where the light is diffused back to the receiver. Classification in the presence of different forms of noise has a large literature in machine learning [39]. Here, we assume that our training set is composed of noiseless images that are correctly classified, namely that $c_k^{\mathcal{T}}$ is the true class of $i_k^{\mathcal{T}}$. Although not explicitly discussed here, it is possible to extend our analysis to noisy training sets via the method of importance reweighting [40–42].

As for the classical algorithm in Eq. (23), there are different strategies to define a classifier given the training set, all with different performances and ranges of applicability [37]. Here we focus on the nearest neighbor classifier [43], defined as

$$\tilde{c}^{\mathcal{T}}_{\rm NN}(\boldsymbol{i}) = c^{\mathcal{T}}_{k_{\rm min}} \ , \qquad k_{\rm min} = \underset{k}{\rm argmin}\, D(\boldsymbol{i}, \boldsymbol{i}_k^{\mathcal{T}}) \ , \qquad (24)$$

where $D(\boldsymbol{i}, \boldsymbol{i}')$ is a suitable distance between two images. In other terms, classification of an unknown image $\boldsymbol{i}$ is done by selecting the class $c^{\mathcal{T}}_{k_{\rm min}}$ of the image from the training set that is closest to $\boldsymbol{i}$, according to distance D. The corresponding algorithm in (23) is $A(c|\boldsymbol{i}) = \delta_{c,\tilde{c}^{\mathcal{T}}_{\rm NN}(\boldsymbol{i})}$. More advanced neural-network based algorithms will be considered in another paper [44].

In spite of being very simple, the nearest neighbor classifier has many desirable features. Indeed, under mild conditions, it has been proven [43] that, for $T \to \infty$, the classification error using the nearest neighbor classifier is at most twice the Bayes rate, irrespective of the number of classes. More details are shown in the Supplemementary material, where we study the performance of this classifier for finite T, i.e., for finite training sets. Another feature is the ability to choose the most appropriate distance D. Here we choose the Hamming distance, which allows us to exploit many results from previous sections.

In this section we consider quantum sources and sensors, but classical algorithms for nearest neighbor classification. Quantum computers can perform nearest neighbor classification quicker than any classical counterpart [45], but how to mix those quantum algorithms with optical detection schemes is still an open problem.

Inserting Eq. (23) into (15) and employing the nearest neighbor classifier we get

$$E^{\rm NN}_{\mathcal{T}} := \sum_{\boldsymbol{i},\tilde{\boldsymbol{i}}} \sum_{c\neq\tilde{c}^{\mathcal{T}}_{\rm NN}(\tilde{\boldsymbol{i}})} P(c,\boldsymbol{i}) \prod_{j=1}^{n} {\rm Tr}\left[\Pi_{\tilde{i}_j} \rho_{i_j}^{\otimes M}\right], \qquad (25)$$

To understand this error, suppose that the pixel error probability p is independent on whether the pixel is black or white. In this case, the probability that the reconstructed image $\tilde{\boldsymbol{i}}$ differs from the true image $\boldsymbol{i}$ in $k = {\rm hamming}(\boldsymbol{i}, \tilde{\boldsymbol{i}})$ pixels is a binomial distribution $\propto p^k(1-p)^{n-k}$, with mean np. Thanks to the analysis shown in Fig. 3b we know that, on average, as long as the number of wrongly detected pixels is smaller than the typical separation in Hamming distance between different classes, the nearest neighbour classifier should provide the correct result. For the transformed MNIST dataset, $n = 28{\times}28$ and the typical number of flips between different classes is ≈ 160 (see Fig. 3b), so a pixel error probability up to $p \simeq 160/784 \simeq 20\%$ should be tolerated by the algorithm.

In Fig. 4 we study the robustness of the nearest neighbor classifier via a numerical analysis with the trans-

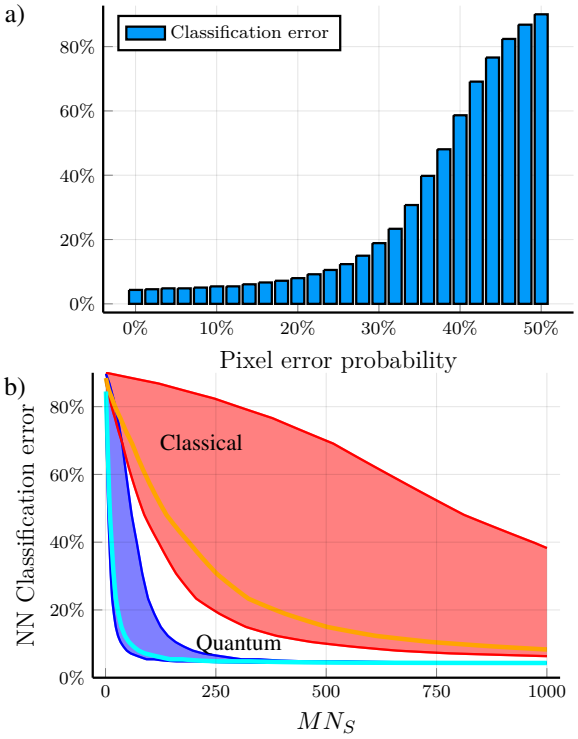

FIG. 4. Pattern recognition with independent on-pixel measurements. (a) Classification error, namely empirical probability of recognizing the wrong digit using the nearest neighbor classifier with Hamming distance, as a function of the pixel error probability. (b) Classification error when each pixel is probed using either coherent inputs or entangled TMSV states. The plot is generated by combining the error coming from the single pixel error probability (see Appendix B) with the classification error from (a). We focus on the limit $M \to \infty$, $N_S \to 0$ while keeping fixed the total mean number of photons MN_S irradiated over each pixel. The colored areas represent the region between the upper and lower bounds assuming quantum (blue) or classical (red) sources combined with optimal local measurements. These bounds depend on the quantum and classical fidelities from Eqs. (10) and (12). The cyan and orange lines represents the performance with quantum light (cyan) or classical coherent states (orange) using (non-optimal) photodetection measurements. We set $\eta_B = 0.9$ and $\eta_W = 0.95$.

formed MNIST dataset, where each image is transformed into a 2D barcode as described in Sec. III B. We use such transformed training set to build a nearest neighbor classifier, and then estimate the error (25) as an average over the testing set, namely as $N_{\rm wrong}/10000$ where $N_{\rm wrong}$ is the number of times that in the 10000 entries of the testing set, the predicted digit is different from the true one. Since the images from the testing set are samples from the abstract and unknown probability $P(c,\boldsymbol{i})$ in the limit of infinitely large testing sets such estimate converges to $E_{\mathcal{T}}^{\rm NN}$ from Eq. (25). Moreover, since the images from the testing set are different from the ones in the training set, this error contains two terms: an error due to imperfect detection and an generalization error, since we are classifying previously unseen data.

In Fig. 4a we study the classification error as a function of the probability p of wrong pixel detection. As we see, even for noiseless images, namely when $p = 0$, the classification error is still non-zero, as the nearest neighbor classifier may provide wrong outcomes. Nonetheless, as predicted, Fig. 4a shows that the nearest neighbor classifier is remarkably robust against relatively high pixel error probabilities p.

In Fig. 4b we combine the bound on the pixel error probability (see Eq. (B5), for a single pixel $n = 1$) with the theoretical curve that predicts the classification error from the pixel error probability in Fig. 4a. The bounds on the pixel error probabilities are obtained from the fidelities, Eqs. (10) and (12), which consider either coherent states or entangled TMSV states with the same average number of photons MN_S. The results from Fig. 4b show that the classification error when we use quantum light is lower than the corresponding classical value. These results are basesd on the assumption that the detector performs the optimal Helstrom measurement, which may be complex to implement experimentally. Therefore, in Fig. 4b, we also consider the simpler photodetection measurement, where the POVM in Eq. (1) is a projection onto the Fock basis. The resulting pixel error probabilities with both coherent states and TMSV inputs are studied in the Supplementary Material, adapting the analysis from Ref. [46]. We see that even for this non-optimal measurement, entangled inputs always provide an advantage against purely classical coherent states for all possible values of MN_S. This advantage can be experimentally observed via a setup like that of Ref. [46].

IV. DISCUSSION

In this work, we have investigated multi-pixel problems of quantum channel discrimination, namely the identification of barcode configurations (equivalent to readout of the stored data) and the classification of black and white patterns, e.g. given by noisy digital images of handwritten digits. In both cases, we have shown that the use of quantum light based on entangled states clearly outperforms classical strategies based on coherent states.

For both quantum-enhanced barcode decoding and pattern recognition, we have analytically studied, via bounds, the physical limits to the classification error that we may get by optimizing over all optical elements, measurements and classical post-processing. This allows us to to derive explicit analytical conditions for the quantum advantage to hold. Moreover, the analysis of our bounds allows us to rigorously prove that quantum-enhanced pattern recognition can vastly reduce the classification error with respect to the mere independent measurement of each pixel.

Nonetheless, being easier from the experimental point of view, we have also considered a simplified setup where all pixels are probed independently and, for the problem of pattern recognition, we found that photon counting measurements are sufficient to show quantum advantage, paving the way for an experimental demonstration with state of the art quantum technology.

ACKNOWLEDGMENTS

L.B. acknowledges support by the program "Rita Levi Montalcini" for young researchers. Q.Z. acknowledges support from the Defense Advanced Research Projects Agency (DARPA) under Young Faculty Award (YFA) Grant No. N660012014029 and Craig M. Berge Dean's Faculty Fellowship of University of Arizona. S.P. acknowledges funding from EU Horizon 2020 Research and Innovation Action under grant agreement No. 862644 (Quantum Readout Techniques and Technologies, QUARTET).

Appendix A: Bounds on barcode discrimination error

Consider a barcode with n pixels or bars, where each bar can assume two possible values (black or white) as in Fig. 1. There are in total 2^n possible configurations and the discriminator must be able to correctly identify each code. Using uniform a priori probabilities $\pi_{\boldsymbol{i}} = 2^{-n}$ in (3) we get

$$p_{\rm err} \geq 2^{-(2n+1)} \sum_{\boldsymbol{i}\neq\boldsymbol{j}} F^{2M}(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{j}}), \tag{A1}$$

$$p_{\rm err} \leq 2^{-n} \sum_{\boldsymbol{i}\neq\boldsymbol{j}} F^{M}(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{j}}). \tag{A2}$$

For some specific choices of the input, e.g. for coherent states or TMSV states, the states take the product form (5) and the fidelities can be written as (6), where $\text{hamming}(\boldsymbol{i},\boldsymbol{j})$ is the Hamming distance between the two images. In oder to find computable bounds for $p_{\rm err}$ we study then the following quantity

$$D_n(f) = \frac{1}{2^n} \sum_{\boldsymbol{i}\neq\boldsymbol{j}} f^{\text{hamming}(\boldsymbol{i},\boldsymbol{j})}. \tag{A3}$$

We find a closed form for $D_n(f)$ by recursion. We may write $\boldsymbol{i} = \{i_1, \boldsymbol{i'}\}$ where $\boldsymbol{i'}$ has $n-1$ components and similarly for $\boldsymbol{j}$. Then in (A3) we may separately consider all four possible values of i_1 and j_1, noting that when $i_1 \neq j_1$ we may also have $\boldsymbol{i'} = \boldsymbol{j'}$. Therefore we get

$$D_n = \frac{1}{2^n} \sum_{\boldsymbol{i} \neq \boldsymbol{j}} f^{\text{hamming}(i_1,j_1)} f^{\text{hamming}(\boldsymbol{i'},\boldsymbol{j'})} \tag{A4}$$

$$= \frac{D_{n-1} + D_{n-1} + 2f(D_{n-1} + 2^{n-1}/2^{n-1})}{2} \tag{A5}$$

$$= (1+f)D_{n-1} + f \ , \tag{A6}$$

where each term in the sum corresponds to $(i_1, j_1) = \{(0,0),(1,1),(1,0),(0,1)\}$. The solution to the above recursion is

$$D_n(f) = (f+1)^n - 1 \ , \tag{A7}$$

from which we get the following inequalities

$$p_{\text{err}} \geq \frac{[F^{2M}(\rho_W, \rho_B) + 1]^n - 1}{2^{n+1}}, \tag{A8}$$

$$p_{\text{err}} \leq [F^M(\rho_W, \rho_B) + 1]^n - 1 \leq$$
$$\leq e^{nF^M(\rho_W,\rho_B)} - 1, \tag{A9}$$

where in the last line we use the inequality $(1+x)^r \leq e^{rx}$. A small error $p_{\text{err}} \leq \epsilon$ is then obtained if we use M copies with

$$M \geq \frac{-\log \frac{\log(1+\epsilon)}{n}}{-\log F}. \tag{A10}$$

Finally we study the asymptotic performance for $M \to \infty$. From the definition of the state (5) we may write

$$F_{\max} = F(\rho_W, \rho_B)^{\min_{i \neq j} \text{hamming}(\boldsymbol{i},\boldsymbol{j})} = F(\rho_W, \rho_B), \tag{A11}$$

and taking the limit in Eqs.(A8)-(A9) we get

$$\frac{n}{2^{n+1}} F^{2M} \lesssim p_{\text{err}} \lesssim n F^M, \tag{A12}$$

which is valid for large M.

Appendix B: Local measurements

When local measurements are employed together with product states as in (5), then the success probability can be simplified as the probability of independently discriminating each pixel, namely

$$p_{\text{succ}}^{\text{local}} = (p_{\text{succ}}^{\text{pixel}})^n, \tag{B1}$$

where $p_{\text{succ}}^{\text{local}}$ is the success probability of discriminating each image (2) using local measurements only and $p_{\text{succ}}^{\text{pixel}}$ is the success probability of detecting the grey-level of a single pixel. For equal *a priory* probabilities, the success probability can be written as $p_{\text{succ}}^{\text{pixel}} = (\text{Tr}[\Pi_B \rho_B] + \text{Tr}[\Pi_W \rho_W])/2 = \frac{1}{2} + \frac{1}{4}\text{Tr}[(\Pi_B - \Pi_W)(\rho_B - \rho_W)]$ and, thanks to the Helstrom theorem [27, 47], minimum error is obtained when Π_B and Π_W are, respectively, the projectors onto the positive and negative subspace of $\rho_B - \rho_W$. Therefore, with optimal Helstrom measurements the success probability is given by $p_{\text{succ}}^{\text{pixel}} = 1/2 + 1/4\|\rho_W - \rho_B\|_1$. Using the Fuchs–van de Graaf inequalities [28],

$$1 - F(\rho,\sigma) \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho,\sigma)^2}, \tag{B2}$$

and the fact that $F_{\max} = F(\rho_B, \rho_W)$ we find

$$1 - \frac{F_{\max}}{2} \leq p_{\text{succ}}^{\text{pixel}} \leq \frac{1 + \sqrt{1 - F_{\max}^2}}{2}, \tag{B3}$$

$$\frac{1 - \sqrt{1 - F_{\max}^2}}{2} \leq p_{\text{err}}^{\text{pixel}} \leq \frac{F_{\max}}{2} \ . \tag{B4}$$

From the above inequalities, we may get a bound on the error probability $p_{\text{err}}^{\text{local}} = 1 - p_{\text{succ}}^{\text{local}}$ using local measurements and M copies. The result is

$$1 - \frac{(1 + \sqrt{1 - F_{\max}^{2M}})^n}{2^n} \leq p_{\text{err}}^{\text{local}} \leq 1 - \left(1 - \frac{F_{\max}^M}{2}\right)^n. \tag{B5}$$

Performing the asymptotic analysis, for large M we find

$$\frac{n}{4} F^{2M} \lesssim p_{\text{err}} \lesssim \frac{n}{2} F^M, \tag{B6}$$

which results in the same decaying rate of Eq. (A12). Therefore, for small n (and large M), the use of global measurements does not increase our ability to distinguish the product states (5). However, for large n the factor 2^{-n-1} in (A12) shows that non-local quantum measurements might provide an important advantage.

We now show that the upper bound from Eq. (B5) is tighter than that of Eq. (A9), namely that

$$(F_{\max}^M + 1)^n - 1 \geq 1 - \left(1 - \frac{F_{\max}^M}{2}\right)^n \ . \tag{B7}$$

Indeed, for any $0 \leq f \leq 1$ we find

$$(f+1)^n + (1 - f/2)^n = \tag{B8}$$
$$= 2 + \sum_{k=1}^{n} \binom{n}{k} f^k [1 + (-1/2)^k] \geq 2 \ ,$$

given that all terms in the sum are positive. Substituting $f = F_{\max}^M$ in the above equation we get (B7). The upper bound from (B5) is always tighter than (A9) and we may combine the two bounds (A8) and (B5) as in Eq. (7). Finally, using the Bernoulli's inequality, $(1+x)^n \geq 1 + nx$ when $x \geq -1$, we find Eq. (8), which is valid for any n and M. The above bounds coincide with what we get from the asymptotic analysis, Eqs. (A12) and (B6), showing that the Bernoulli's inequality is tight in that regime.

Appendix C: Non-uniform a priori probabilities

Here we consider a different problem where the set of images or configurations is restricted to have only k white pixels. We may also call this problem k channel position finding, k-CPF, being a generalization of the idea of CPF introduced in Ref. [19]. Namely, we know that there will be exactly k target channels (white pixels), but we do not know their positions. There are n choose k configurations in this set. The upper and lower bounds in Eqs. (A1) and (A2) can be solved as follows.

Similar to Eq. (A3), we can define

$$D_n^k(f) = \frac{1}{\binom{n}{k}} \sum_{\boldsymbol{i}\neq\boldsymbol{j},k-CPF} f^{\text{hamming}(\boldsymbol{i},\boldsymbol{j})}, \tag{C1}$$

where the summation is over configurations $\boldsymbol{i} \neq \boldsymbol{j}$ with k target channels, and we can express Eqs. (A1) and (A2) as

$$p_{\text{err}}^{k-\text{CPF}} \geq \frac{D_n^k[F(\rho_B,\rho_W)^{2M}]}{2\binom{n}{k}}, \tag{C2}$$

$$p_{\text{err}}^{k-\text{CPF}} \leq D_n^k[F(\rho_B,\rho_W)^{M}]. \tag{C3}$$

Our task reduces to solving the summation in Eq. (C1).

To begin with, we consider the simple case of $k = 2$, as the case of $k = 1$ is solved in Ref. [19]. For the case of 2-CPF, we will have two kinds of terms in the summation of Eq. (C1), one with Hamming distance 4 (corresponding to patterns with no target channels overlapping), and one with Hamming distance 2 (corresponding to patterns with one target channel overlapping), so that

$$\sum_{\boldsymbol{i}\neq\boldsymbol{j},2-CPF} f^{\text{hamming}(\boldsymbol{i},\boldsymbol{j})} = \binom{n}{4}\binom{4}{2} f^4 + 6\binom{n}{3} f^2. \tag{C4}$$

As further example, consider 2-CPF with 4 pixels. There are 6 configurations $1100, 1010, 1001, 0110, 0101, 0011$. In total we have $6 \times 5 = 30$ terms to sum up. There are 6 terms with f^4: $(1100, 0011), (1010, 0101), (1001, 0110)$ and their reverse. There are 24 terms with f^2, for example $(1100, 1010)$.

In general, for k-CPF with n pixels, we have $\binom{n}{k}$ configurations of patterns. Therefore, Eq. (C1) is a summation of $\binom{n}{k}\left(\binom{n}{k}-1\right)$ terms. There are terms with Hamming distance of from 2 to $2k$. By counting the number of terms with an identical hamming distance, we can compute

$$\sum_{\boldsymbol{i}\neq\boldsymbol{j},k-CPF} f^{\text{hamming}(\boldsymbol{i},\boldsymbol{j})} = \sum_{t=k+1}^{2k} \binom{n}{t}\binom{k}{2k-t}\binom{t}{k} f^{2(t-k)}$$
$$= \frac{(1+k)\binom{n}{k+1}}{n-k}\left({}_2F_1\left(-k, k-n, 1, f^2\right) - 1\right), \tag{C5}$$

where ${}_2F_1$ is the standard hypergeometric function.

Therefore we can solve the function D_n^k as

$$D_n^k(f) = {}_2F_1\left(-k, k-n, 1, f^2\right) - 1, \tag{C6}$$

and then evaluate the bounds through Ineqs. (C2) and (C3).

For large M we may expand D_n^k as

$$D_n^k = k(n-k)f^2 + O(f^4)\ , \tag{C7}$$

and for large n and k we may write $\binom{n}{k} \approx 2^{nH(k/n)}$, where $H(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ is the binary entropy function. Employing these approximations in Eqs. (C2)-(C3) we find Eq. (9).

Appendix D: Quantum and classical fidelities

In the case of a quantum source, the input state is a two-mode Gaussian state, described by the quadrature operators $Q = (x_1, p_1, x_2, q_2)$. A TMSV state Φ_{N_S} has zero first moments and covariance matrix (CM) $V_{jk} = \langle\{Q_j, Q_k\}\rangle/2$ given by

$$V_{\text{input}} = \frac{1}{2}\begin{pmatrix} \mu I & \mu' Z \\ \mu' Z & \mu I \end{pmatrix}, \quad \begin{matrix}\mu = 2N_S + 1 \\ \mu' = \sqrt{\mu^2 - 1},\end{matrix} \tag{D1}$$

where I is the 2x2 identity matrix and Z is the Pauli Z operator. In the above expression the variance of the vacuum noise is $1/2$, and N_S is the mean number of thermal photons in each mode. If we now apply $\mathcal{I}\otimes\mathcal{E}_\eta$ to the TMSV state, we get a Gaussian output state with CM

$$V(\eta) = \frac{1}{2}\begin{pmatrix} \mu(1) I & \sqrt{\eta}\mu' Z \\ \sqrt{\eta}\mu' Z & \mu(\eta) I \end{pmatrix} \tag{D2}$$

where

$$\mu(\eta) := \eta\mu + (1-\eta). \tag{D3}$$

The fidelity $F(\rho_W, \rho_B)$ can be computed using the formulae from Refs. [33, 34], and depends on the invariants

$$\Delta = \det(V(\eta_B) + V(\eta_W)) = \tag{D4}$$
$$= \left(N_S^2\left(\sqrt{\eta_W} - \sqrt{\eta_B}\right)^2 - 2N_S\left(\sqrt{\eta_W\eta_B} - 1\right) + 1\right)^2,$$
$$\Gamma = 2^4 \det(\Omega V(\eta_B)\Omega V(\eta_W) - I\otimes I/4) \tag{D5}$$
$$= \Delta + 4(\eta_W - 1)(\eta_B - 1)N_S^2\left(N_S\left(1 - \sqrt{\eta_W\eta_B}\right) + 1\right)^2,$$
$$\Lambda = 2^4 \det(V(\eta_B) + i\Omega/2)\det(V(\eta_W) + i\Omega/2) = 0, \tag{D6}$$

where $\Omega = (iY)^{\oplus 2}$ and Y is the Pauli Y operator. In terms of the above quantities the fidelity can be written as

$$F(\rho_W, \rho_B) = \sqrt{\frac{1}{\sqrt{\Gamma} + \sqrt{\Lambda} - \sqrt{(\sqrt{\Gamma} + \sqrt{\Lambda})^2 - \Delta}}}$$
$$= \frac{1}{\sqrt{\frac{\sqrt{\Gamma}+\sqrt{\Lambda}+\sqrt{\Delta}}{2}} - \sqrt{\frac{\sqrt{\Gamma}+\sqrt{\Lambda}-\sqrt{\Delta}}{2}}}, \tag{D7}$$

where in the second line we simplify the first expression, derived in [33, 34], by explicitly evaluating the square root. Inserting the values from Eqs. (D4)-(D6) we get the result shown in Eq. (10). The tightness of the inequality in Eq. (10) can be proven by setting $N_{\rm tot} = MN_S$ and using $e^x = \lim_{M\to\infty}(1+x/M)^M$

$$\lim_{M\to\infty}\left(\frac{1}{1+N_{\rm tot}\Delta_Q/M}\right)^M = e^{-N_{\rm tot}\Delta_q} \ . \tag{D8}$$

Now let us compute the output fidelity corresponding to an input coherent state $D(\sqrt{N_s})\,|0\rangle$, where D is the displacement operator. The action of the channel with transmissivity η is described by the Heisenberg evolution

$$a^\dagger \to \sqrt{\eta}a^\dagger + \sqrt{1-\eta}b^\dagger, \tag{D9}$$

where b describes the photons in the (vacuum) bath. After the application of the channel we get a new coherent state $D(\sqrt{\eta N_S})\,|0\rangle$, so the fidelity is

$$\begin{aligned} F_{\rm cl} &= |\,\langle 0|\,D(\sqrt{\eta_B N_S})^\dagger D(\sqrt{\eta_W N_S})\,|0\rangle\,| = \\ &= \left| e^{-\frac{N_S\eta_B}{2}} e^{-\frac{N_S\eta_W}{2}} \sum_n \frac{(N_S\sqrt{\eta_B\eta_W})^n}{n!}\right| = \\ &= e^{-\frac{N_s}{2}(\sqrt{\eta_B}-\sqrt{\eta_W})^2}, \end{aligned} \tag{D10}$$

which is the result shown in Eq. (12).

Appendix E: Quantum pattern recognition with pretty-good measurements

A pretty good solution to the optimization problem (15) can be obtained with pretty good measurements [23–25], which are defined by $\Pi_c^{\rm PGM} = \rho_{\rm tot}^{-1/2}P(c)\rho_c^M\rho_{\rm tot}^{-1/2}$, where $P(c,\boldsymbol{i}) = P(\boldsymbol{i}|c)P(c)$ via the Bayes rule, $\rho_c^M = \sum_{\boldsymbol{i}} P(\boldsymbol{i}|c)\rho_{\boldsymbol{i}}^{\otimes M}$ and $\rho_{\rm tot} = \sum_c P(c)\rho_c^M = \sum_{c,\boldsymbol{i}} P(c,\boldsymbol{i})\rho_{\boldsymbol{i}}^{\otimes M}$. Using these measurements, we may find upper and lower bounds [21, 22]

$$E^{\rm Q} \geq \frac{1}{2}\sum_{c\neq c'} P(c)P(c')F(\rho_c^M,\rho_{c'}^M)^2, \tag{E1}$$

$$E^{\rm Q} \leq \sum_{c\neq c'} \sqrt{P(c)P(c')}F(\rho_c^M,\rho_{c'}^M), \tag{E2}$$

similar to (3). The lower bound can be simplified thanks to the concavity of F^2 [48], so as to get

$$E^{\rm Q} \geq \frac{1}{2}\sum_{c\neq c'}\sum_{\boldsymbol{i},\boldsymbol{i'}} P(c,\boldsymbol{i})P(c',\boldsymbol{i'})F(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{i'}})^{2M}, \tag{E3}$$

while the same simplification cannot be obtained for the upper bound. As a consequence, the upper bound cannot be directly computed, as ρ_c^M is not a quantum Gassian state. In order to introduce a computable upper bound we prove the following lemma, generalizing methods from [21, 38].

Lemma: Let us fix a function f. Then using pretty good measurements we can derive the following bound for the probability of error

$$p_{\rm E} = \sum_{x,y:f(x)\neq f(y)} p_x\,{\rm Tr}[\Pi_y\rho_x] \tag{E4}$$

$$\leq \sum_{x,y:f(x)\neq f(y)} \sqrt{p_x p_y}F(\rho_x,\rho_y)\ . \tag{E5}$$

Proof: Let us write $\sigma_x = p_x\rho_x = \sum_k \lambda_{xk}\,|\psi_{xk}\rangle\langle\psi_{xk}|$, where λ_{xk} and $|\psi_{xk}\rangle$ form the eigen-decomposition of σ_x. We also set $\Pi_y = \sigma^{-1/2}\sigma_y\sigma^{-1/2}$, $\sigma = \sum_x \sigma_x$, and define the Gram matrix $G_{kl}^{xy} = \sqrt{\lambda_{xk}\lambda_{yl}}\,\langle\psi_{xk}|\,\psi_{yl}\rangle$. Then the following identities hold (see [38] for a proof):

$$\|\sqrt{G}^{xy}\|_2^2 = p_x\,{\rm Tr}[\Pi_y\rho_x]\ , \tag{E6}$$

$$\|G^{xy}\|_1 = \sqrt{p_x p_y}F(\rho_x,\rho_y)\ , \tag{E7}$$

where $\|A\|_2^2 = {\rm Tr}\big[A^\dagger A\big]$, $\|A\|_1 = {\rm Tr}\,\sqrt{A^\dagger A}$ and G^{xy} is a matrix with elements G_{kl}^{xy}. Moreover, $\|\sqrt{G}^{xy}\|_2^2 \leq \|G^{xy}\|_1$ from lemma 4 in [21]. Therefore,

$$p_{\rm E} = \sum_{x,y:f(x)\neq f(y)} \|\sqrt{G}^{xy}\|_2^2 \leq \sum_{x,y:f(x)\neq f(y)} \|G^{xy}\|_1\ ,$$

which produces (E5) via (E7). □

Thanks to the above lemma, we may now derive a simpler upper bound than (E2). Indeed, we may rewrite (15) as in (E4) by setting $f(c,\boldsymbol{i}) = c$, calling $\rho_{c,\boldsymbol{i}} := \rho_{\boldsymbol{i}}$ and $\Pi_{c'} := \sum_{\boldsymbol{i'}} \Pi_{c',\boldsymbol{i'}}$ and defining multi-indices $x = (c,\boldsymbol{i})$ and $y = (c',\boldsymbol{i'})$. Then from the above lemma we get

$$E^{\rm Q} \leq \sum_{c\neq c'}\sum_{\boldsymbol{i},\boldsymbol{i'}} \sqrt{P(c,\boldsymbol{i})P(c',\boldsymbol{i'})}F(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{i'}})^M. \tag{E8}$$

Now we show how to get Ineqs. (16). Indeed, via (E3) and (6) we may write $E^{\rm Q} \geq B[F(\rho_B,\rho_W)^{2M}]$ where

$$B[F] = \frac{1}{2}\sum_{c\neq c'} P(c)P(c')B_{cc'}[F], \tag{E9}$$

$$B_{cc'}[F] = \sum_h P_{cc'}(h)F^h, \tag{E10}$$

$$P_{cc'}(h) = \sum_{\boldsymbol{i},\boldsymbol{i'}} P(\boldsymbol{i}|c)P(\boldsymbol{i'}|c')\delta_{h,{\rm hamming}(\boldsymbol{i},\boldsymbol{i'})}. \tag{E11}$$

Similarly, we may rewrite the upper bound (E8) as

$$\begin{aligned} E^{\rm Q} &\leq K\sum_{c\neq c'}\sum_{\boldsymbol{i},\boldsymbol{i'}} P(c,\boldsymbol{i})P(c',\boldsymbol{i'})F(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{i'}})^M = \\ &= KB[F(\rho_B,\rho_W)^M], \end{aligned} \tag{E12}$$

where K^{-1} is the minimum non-zero value of $\sqrt{P(c,\boldsymbol{i})P(c',\boldsymbol{i'})}$. Therefore, as long as $P_{cc'}(h)$ is zero for $h < h_{cc'}^{\rm min}$ in the limit of large M we get

$$CF(\rho_B,\rho_W)^{2Mh_{\rm min}} \lessapprox E^{\rm Q} \lessapprox 2KCF(\rho_B,\rho_W)^{Mh_{\rm min}},$$

up to a constant C.

We now focus on the bounds in Eq. (20), which follow by approximating $P(c, \boldsymbol{i})$ via the training distribution

$$P(c, \boldsymbol{i}) \simeq \frac{1}{T} \sum_{k=1}^{T} \delta_{c,c_k^{\mathcal{T}}} \delta_{\boldsymbol{i},\boldsymbol{i}_k^{\mathcal{T}}}, \tag{E13}$$

and noting that, in (E8), we may use $\sqrt{\delta_{ij} + \delta_{kl}} \leq \delta_{ij} + \delta_{kl}$. Alternatively, we may employ (E12) and note that $K = T$ for the distribution (E13).

As for the bounds (20), let us study the function defined in Eq. (21), which can be rewritten as

$$B_{\mathcal{T}}[F] = \sum_{c \neq c'} \frac{T_c T_{c'}}{2T^2} B_{cc'}^{\mathcal{T}}[F] \tag{E14}$$

with $T_c = \sum_k \delta_{c,c_k^{\mathcal{T}}}$, $T = \sum_c T_c$ and

$$B_{cc'}^{\mathcal{T}}[F] = \sum_h P_{cc'}^{\mathcal{T}}(h) F^h, \tag{E15}$$
$$P_{cc'}^{\mathcal{T}}(h) = \frac{1}{T_c T_{c'}} \sum_{k,k'} \delta_{c,c_k^{\mathcal{T}}} \delta_{c',c_{k'}^{\mathcal{T}}} \delta_{h,\mathrm{hamming}(\boldsymbol{i}_k^{\mathcal{T}},\boldsymbol{i}_{k'}^{\mathcal{T}})}.$$

The probability $P_{01}^{\mathcal{T}}(h)$ is numerically studied in Fig. 3b for the MNIST dataset, where we see that $P_{01}^{\mathcal{T}}(h)$ closely matches a normal distribution. Therefore, approximating $P_{cc'}^{\mathcal{T}}(h)$ as a normal distribution with mean $\mu_{cc'}$ and standard deviation $\sigma_{cc'}$ we may write

$$\begin{aligned} B_{cc'}^{\mathcal{T}}[F] &\simeq \int_{h_{cc'}^0}^{\infty} p_{\mu_{cc'},\sigma_{cc'}}(h) F^h \\ &= \frac{1}{2} F^{\mu_{cc'}} e^{\frac{1}{2}\sigma_{cc'}^2 \log^2(F)} \left(\mathrm{erf}\left(\frac{w_{cc'}}{\sqrt{2}\sigma_{cc'}} \right) + 1 \right), \end{aligned} \tag{E16}$$

where $p_{\mu,\sigma}(h)$ is the probability density function of a normal distribution with mean μ and standard deviation σ, $w_{cc'} := \sigma_{cc'}^2 \log(F) + \mu_{cc'} - h_{cc'}^0$ and $h_{cc'}^0$ is the minimum value of Hamming distance in the empirical distribution. When $\mu \gg \sigma$ we find $B_{cc'}^{\mathcal{T}} \simeq F^{\mu_{cc'}}$. On the other hand, for $F \to 0$ we find

$$B_{cc'}^{\mathcal{T}}[F] \overset{F\to 0}{\simeq} \frac{\sigma F^{h_{cc'}^0} e^{-\frac{(h_{cc'}^0 - \mu_{cc'})^2}{2\sigma_{cc'}^2}}}{\sqrt{2\pi}\,(-w_{cc'})}. \tag{E17}$$

This leads to the asymptotic decay rate $B_{cc'}[F^M] \propto F^{M h_{cc'}^0}$ for large M, up to constants and logarithmic corrections. Using the latter asymptotic decay expression in Eq. (E15), we find that, for large M, only the classes with smallest minimum Hamming distance $h_{cc'}^0$ survive. Moreover, we assume that the ratio T_c/T does not scale with M, e.g. for digit reconstruction we assume that the number of 4s and 9s are basically constant if we increase the number of images. As such we may write

$$\lim_{M\to\infty} \frac{\log B[F^M]}{M} = h_{\min} \log F, \tag{E18}$$

where $h_{\min} = \min_{c\neq c'} h_{cc'}^0$, and from the bounds (20)

$$-h_{\min} \log F \leq -\lim_{M\to\infty} \frac{\log E_{\mathcal{T}}^{\mathrm{Q}}}{M} \leq -2h_{\min} \log F. \tag{E19}$$

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

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

[3] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. Pereira, M. Razavi, J. S. Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, Advances in Quantum Cryptography, arXiv:1906.01645 (2019).

[4] C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum sensing, Rev. Mod. Phys. **89**, 035002 (2017).

[5] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, Advances in photonic quantum sensing, Nat. Photon. **12**, 724 (2018).

[6] S. L. Braunstein and P. van Loock, Quantum Information with Continuous Variables, Rev. Mod. Phys. **77**, 513 (2005).

[7] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods* (CRC Press, 2017).

[8] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, Rev. Mod. Phys. **84**, 621 (2012).

[9] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, Phys. Rev. Lett. **72**, 3439 (1994).

[10] M. G. A. Paris, Quantum estimation for quantum technology, International Journal of Quantum Information **07**, 125 (2009).

[11] C. Helstrom, *Quantum Detection and Estimation Theory*, Mathematics in Science and Engineering: a series of monographs and textbooks (Academic Press, 1976).

[12] S. M. Barnett and S. Croke, Quantum state discrimination, Adv. Opt. Photon. **1**, 238 (2009).

[13] A. Chefles, Quantum state discrimination, Contemp. Phys. **41**, 401 (2000).

[14] J. A. Bergou, Discrimination of quantum states, Journal of Modern Optics **57**, 160 (2010).

[15] S. Lloyd, Enhanced sensitivity of photodetection via quantum illumination, Science **321**, 1463 (2008).

[16] S.-H. Tan, B. I. Erkmen, V. Giovannetti, S. Guha,

S. Lloyd, L. Maccone, S. Pirandola, and J. H. Shapiro, Quantum illumination with gaussian states, Phys. Rev. Lett. **101**, 253601 (2008).
[17] S. Pirandola, Quantum reading of a classical digital memory, Phys. Rev. Lett. **106**, 090504 (2011).
[18] Q. Zhuang and Z. Zhang, Physical-layer supervised learning assisted by an entangled sensor network, Phys. Rev. X **9**, 041023 (2019).
[19] Q. Zhuang and S. Pirandola, Entanglement-enhanced testing of multiple quantum hypotheses, Communications Physics **3**, 103 (2020).
[20] K. P. Murphy, *Machine learning: a probabilistic perspective* (MIT press, 2012).
[21] H. Barnum and E. Knill, Reversing quantum dynamics with near-optimal quantum and classical fidelity, J. Math. Phys. **43**, 2097 (2002).
[22] A. Montanaro, A lower bound on the probability of error in quantum state discrimination,, 2008 IEEE Information Theory Workshop Porto , 378 (2008).
[23] A. S. Holevo, Theor. Probab. and Appl. **23**, 411 (1978).
[24] P. Hausladen and W. K. Wootters, A pretty good measurement for distinguishing quantum states, Journal of Modern Optics **41**, 2385 (1994).
[25] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996).
[26] L. Banchi, N. Quesada, and J. M. Arrazola, Training gaussian boson sampling distributions, Phys. Rev. A **102**, 012417 (2020).
[27] C. W. Helstrom, Quantum detection and estimation theory, Journal of Statistical Physics **1**, 231 (1969).
[28] C. A. Fuchs and J. Van De Graaf, Cryptographic distinguishability measures for quantum-mechanical states, IEEE Transactions on Information Theory **45**, 1216 (1999).
[29] J. Bae and L.-C. Kwek, J. Phys A: Math Theor. **48**, 083001 (2015).
[30] M. Nussbaum, A. Szkoła, *et al.*, An asymptotic error bound for testing multiple quantum hypotheses, The Annals of Statistics **39**, 3211 (2011).
[31] J. Calsamiglia, J. I. de Vicente, R. Muñoz Tapia, and E. Bagan, Local discrimination of mixed states, Phys. Rev. Lett. **105**, 080504 (2010).
[32] S. Bandyopadhyay, More nonlocality with less purity, Phys. Rev. Lett. **106**, 210402 (2011).
[33] L. Banchi, S. L. Braunstein, and S. Pirandola, Phys. Rev. Lett. **115**, 260501 (2015).
[34] P. Marian and T. A. Marian, Uhlmann fidelity between two-mode gaussian states, Physical Review A **86**, 022340 (2012).
[35] Z. Zhang, M. Tengner, T. Zhong, F. N. Wong, and J. H. Shapiro, Entanglement's benefit survives an entanglement-breaking channel, Physical review letters **111**, 010501 (2013).
[36] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE **86**, 2278 (1998).
[37] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction* (Springer Science & Business Media, 2009).
[38] A. Montanaro, Pretty simple bounds on quantum state discrimination, arXiv preprint arXiv:1908.08312 (2019).
[39] D. Angluin and P. Laird, Learning from noisy examples, Machine Learning **2**, 343 (1988).
[40] T. Liu and D. Tao, Classification with noisy labels by importance reweighting, IEEE Transactions on pattern analysis and machine intelligence **38**, 447 (2015).
[41] J. A. Aslam and S. E. Decatur, On the sample complexity of noise-tolerant learning, Information Processing Letters **57**, 189 (1996).
[42] N. Manwani and P. Sastry, Noise tolerance under risk minimization, IEEE transactions on cybernetics **43**, 1146 (2013).
[43] T. Cover and P. Hart, Nearest neighbor pattern classification, IEEE transactions on information theory **13**, 21 (1967).
[44] C. Harney *et al.*, Ultimate limits of thermal imaging, (in preparation) (2020).
[45] N. Wiebe, A. Kapoor, and K. M. Svore, Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning, Quantum Information & Computation **15**, 316 (2015).
[46] G. Ortolano, E. Losero, I. R. Berchera, S. Pirandola, and M. Genovese, Experimental quantum reading with photon counting, arXiv:2004.10211 (2020).
[47] I. Bengtsson and K. Życzkowski, *Geometry of quantum states: an introduction to quantum entanglement* (Cambridge university press, 2017).
[48] R. Jozsa, Fidelity for mixed quantum states, Journal of modern optics **41**, 2315 (1994).
[49] S. Pirandola and S. Lloyd, Phys. Rev. A **78**, 012331 (2008).

SUPPLEMENTARY MATERIALS

Multiple Chernoff bound

Here we show that, in the uniform case, we may also employ the multiple quantum Chernoff bound [29, 30]. Combining this with known inequalities [49], we find

$$\frac{F_{\max}^{2M}}{2^M} \lesssim p_{\rm err} \lesssim F_{\max}^{\frac{M}{3}}, \tag{S1}$$

where $F_{\max} := \max_{\boldsymbol{i}\neq\boldsymbol{j}} F_{\boldsymbol{i},\boldsymbol{j}}$, and the approximations are valid in the large-M limit. The above inequality becomes exact for $M \to \infty$, while (3) is valid for any M.

Following [29, 30] for i.i.d. hypotheses, we may also consider the following quantum Chernoff bound

$$\frac{1}{3}\xi_{CB} \leq \lim_{M\to\infty} -\frac{1}{M}\log p_{\rm err} \leq \xi_{CB}, \tag{S2}$$

where

$$\xi_{CB} = \min_{\boldsymbol{i}\neq\boldsymbol{j}}[-\log(Q(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{j}}))] = -\log Q_{\max}, \tag{S3}$$

and

$$Q(\rho_0,\rho_1) := \inf_s \mathrm{Tr}(\rho_0^s\rho_1^{1-s}) \leq \mathrm{Tr}\sqrt{\rho_0}\sqrt{\rho_1}, \tag{S4}$$

$$Q_{\max} = \max_{\boldsymbol{i}\neq\boldsymbol{j}} Q(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{j}}) \leq \max_{\boldsymbol{i}\neq\boldsymbol{j}} \mathrm{Tr}\big[\sqrt{\rho_{\boldsymbol{i}}}\sqrt{\rho_{\boldsymbol{j}}}\big], \tag{S5}$$

which can be computed for any Gaussian state [49]. Using known inequalities [28, 49]

$$1-\sqrt{1-F(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{j}})^2} \leq Q(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{j}}) \leq F(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{j}}) \ , \tag{S6}$$

we may write

$$1-\sqrt{1-F_{\max}^2} \leq Q_{\max} \leq F_{\max} \ , \tag{S7}$$

where

$$F_{\max} = \max_{\boldsymbol{i}\neq\boldsymbol{j}} F(\rho_{\boldsymbol{i}},\rho_{\boldsymbol{j}}) \ . \tag{S8}$$

From the above we finally get the following inequalities

$$-\log F_{\max} \leq \xi_{CB} \leq -\log\Big(1-\sqrt{1-F_{\max}^2}\Big) \ , \tag{S9}$$

$$\begin{aligned} -\frac{1}{3}\log F_{\max} \leq \lim_{M\to\infty} -\frac{1}{M}\log p_{\rm err} \leq \\ \leq -\log\Big(1-\sqrt{1-F_{\max}^2}\Big) \leq \\ \leq -\log\frac{F_{\max}^2}{2}, \end{aligned} \tag{S10}$$

where in the last inequality we use $1-\sqrt{1-x} \geq x/2$ for $0 \leq x \leq 1$, which is tight for $x \simeq 0$. In the large-M limit we get the inequalities (S1). Nonetheless, the exact computation of Q via (S4) may provide a tighter inequality via (S2).

Moreover, from the definition of the states (5), and from (S5) we get

$$Q(\tilde{\rho}_i,\tilde{\rho}_j) = Q(\rho_W,\rho_B)^{\mathrm{hamming}(\boldsymbol{i},\boldsymbol{j})}, \tag{S11}$$

and accordingly, since $Q \leq 1$, we find

$$Q_{\max} = Q(\rho_W,\rho_B)^{\min_{i\neq j}\mathrm{hamming}(\boldsymbol{i},\boldsymbol{j})} = Q(\rho_W,\rho_B). \tag{S12}$$

Together with (S2), this gives computable upper (and possibly lower) bounds on the error rate. Similarly, we may write

$$F_{\max} = F(\rho_W,\rho_B)^{\min_{i\neq j}\mathrm{hamming}(\boldsymbol{i},\boldsymbol{j})} = F(\rho_W,\rho_B), \tag{S13}$$

so we may study the quantum Chernoff bound via (S9) and (S10).

An interesting comparison is between the asymptotic performance given by Eq. (S10), and the explicit results that we may get from from Eq. (A12). Indeed, for large-M we get

$$-\log F \leq -\lim_{M\to\infty}\frac{\log p_{\rm err}}{M} \leq -2\log F, \tag{S14}$$

which is independent on n, as it is Eq. (S10).

Photodetection

We study the performance of barcode decoding and pattern recognition using photodetection. Photodetection is a local measurement, where each pixel is probed independently. As such, this strategy cannot achieve the lower bound in Eq. (8), but it is nonetheless interesting because it can be implemented experimentally with current technology. According to Ref. [46], the pixel error probability for a coherent-state input is given by

$$p_{\rm err,cl}^{\rm pixel} = \frac{1}{2}\left[1-\frac{\gamma(\eta_B)-\gamma(\eta_W)}{\lfloor n_{\rm th}\rfloor!}\right] \ , \tag{S15}$$

where $n_{\rm th} = N_S(\eta_W-\eta_B)/\log(\eta_W/\eta_B)$, while $\gamma(\eta) := \Gamma(\lfloor n_{\rm th}+1\rfloor, N_S\eta)$ and $\Gamma(x,y)$ is the incomplete Gamma function.

For a TMSV with mean photon number N_S we may write $|\mathrm{TMSV}\rangle = \sum_n c_{N_s}(n)\,|n,n\rangle$ with $c_N = \frac{N_S^n}{(1+N_S)^{n+1}}$. After the first mode of the TMSV state is transmitted through a channel with transmissivity η, the resulting photon number distribution is

$$P(n_1,n_2|\eta) = \begin{cases} P_0(n_2)B(n_2,n_1|\eta) & \text{when } n_2 \geq n_1, \\ 0 & \text{when } n_2 < n_1, \end{cases} \tag{S16}$$

where $B(n_2,n_1|\eta) = \binom{n_2}{n_1}\eta^{n_1}(1-\eta)^{n_2-n_1}$ and $P_0(n_2) = c_{N_S}^2(n_2)$ is the initial distribution. The success probability of distinguishing the two channels with transmissivity

η_B and η_W is then

$$p^{\text{pixel}}_{\text{succ,TMSV}} = \frac{1}{2}\sum_{n_2=0}^{\infty}\sum_{n_1=0}^{n_2^{\text{th}}(n_1)} P(n_1,n_2|\eta_B) + \\ + \frac{1}{2}\sum_{n_2=0}^{\infty}\sum_{n_1=n_2^{\text{th}}(n_1)}^{\infty} P(n_1,n_2|\eta_W), \tag{S17}$$

where $n_2^{\text{th}}(n_2) = cn_2$ and

$$c = \left(\frac{\log(\eta_W/\eta_B)}{\log[(1-\eta_W)(1-\eta_B)]} + 1\right)^{-1} . \tag{S18}$$

For M probings, one may substitute the initial distribution in Eq. (S16) with the Poisson distribution $P_0(n_2) = e^{-\lambda}\lambda^{n_2}/n_2!$ with parameter $\lambda \approx MN_S$ [46].

For large average photon number, namely for large λ, the sum in Eq. (S17) is hard to evaluate. Nonetheless, we may approximate the Poisson distribution $P_0(n_2)$ with a normal distribution $n_2 \sim \mathcal{N}(\lambda,\lambda)$, where $\mathcal{N}(\mu,\sigma^2)$ is the normal distribution with mean μ and variance σ^2. Similarly, we may approximate the binomial distribution in (S16) as $n_1 \sim \mathcal{N}(n_2\eta, n_2\eta(1-\eta))$. Introducing the quantity

$$g(\lambda,\eta) = \int_0^{\infty} dn_2 \int_0^{cn_2} dn_1 f_{\lambda,\lambda}(n_2) f_{n_2\eta,n_2\eta(1-\eta)}(n_1) , \tag{S19}$$

where $f_{\mu,\sigma^2}(x)$ is the probability density function of $\mathcal{N}(\mu,\sigma^2)$, and approximating the sum with an integral in (S17), we then get

$$p^{\text{pixel}}_{\text{succ,TMSV}} \simeq \frac{1}{2}\left[g(MN_S,\eta_B) + 1 - g(MN_S,\eta_W)\right] , \tag{S20}$$

and accordingly

$$p^{\text{pixel}}_{\text{err,TMSV}} \simeq \frac{1}{2} - \frac{1}{2}\left[g(MN_S,\eta_B) - g(MN_S,\eta_W)\right] . \tag{S21}$$

Nearest neighbor classifier

Let us review a few important aspects of the nearest neighbor classifier. Using the nearest neighbor classifier (24) with a training set composed of T classified images, the expected classification error for a new image $\boldsymbol{i}$ with true class c is then

$$E_{\text{NN}}(T) = \mathop{\mathbb{E}}_{(c,\boldsymbol{i}),\mathcal{T}}[L_{c,\tilde{c}_{\text{NN}}(\boldsymbol{i})}] , \tag{S22}$$

where T is the size of the training set. Under mild conditions, it has been proven [43] that in the limit $T\to\infty$, the expected classification error $E_{\text{NN}} = \lim_{T\to\infty} E_{\text{NN}}(T)$ satisfies

$$E_{\text{B}} \le E_{\text{NN}} \le E_{\text{B}}\left(2 - \frac{K}{K-1}E_{\text{B}}\right) \le 2E_{\text{B}} , \tag{S23}$$

where E_{B} is the Bayes rate and K is the number of classes. Therefore, the expected error from the nearest neighbor classifier is at most twice the Bayes rate, irrespective of the number of classes.

To study the error for finite T in (S22) we first note that even the training set is made by samples, so the predicted nearest neighbour classifier $\tilde{c}$ is a random variable that depends on the training set $\mathcal{T}$. (Note that we have removed the subscript NN from $\tilde{c}$ to simplify the notation.) Therefore, we may rewrite Eq. (S22) as

$$E_{\text{NN}}(T) = \mathop{\mathbb{E}}_{(c,\boldsymbol{i}),\tilde{c},\mathcal{T}}[L_{c,\tilde{c}}P(\tilde{c}|\boldsymbol{i},\mathcal{T})] \\ = \mathop{\mathbb{E}}_{\boldsymbol{i},\mathcal{T}}\left[\sum_{c\neq\tilde{c}} P(c|\boldsymbol{i})P(\tilde{c}|\boldsymbol{i},\mathcal{T})\right] . \tag{S24}$$

We know from Eq. (24) that $\tilde{c}$ is equal to one element of the training set. Let us call $\tilde{\boldsymbol{i}}$ all possible images that have class $\tilde{c}$. Then by the law of total probability

$$P(\tilde{c}|\boldsymbol{i},\mathcal{T}) = \sum_{\tilde{\boldsymbol{i}}} P(\tilde{c}|\tilde{\boldsymbol{i}})P_{\text{NN}}(\tilde{\boldsymbol{i}}|\boldsymbol{i},\mathcal{T}), \tag{S25}$$

where P_{NN} is the probability that, given an image $\boldsymbol{i}$, in the training set $\mathcal{T}$ the closest image to $\boldsymbol{i}$ is $\tilde{\boldsymbol{i}}$.

In the next sections we then show that $P_{\text{NN}}(\tilde{\boldsymbol{i}}|\boldsymbol{i},\mathcal{T}) \to \delta_{\boldsymbol{i},\tilde{\boldsymbol{i}}}$ in the limit $T\to\infty$. The explanation of this limit is straightforward: since the set of images is finite, in the limit of large training sets the probability of finding the image $\boldsymbol{i}$ inside the training set approaches 1 and, accordingly, $P_{\text{NN}}(\tilde{\boldsymbol{i}}|\boldsymbol{i}) \to \delta_{\boldsymbol{i},\tilde{\boldsymbol{i}}}$. Finally, since the number of classes is smaller than the dimension of the set of images, we may choose the mapping $P(c|\boldsymbol{i})$ to be deterministic and unique, e.g. $P(c|\boldsymbol{i}) = \delta_{c,f(\boldsymbol{i})}$, where f is a function that assigns a class c to i. With these assumptions we find

$$\lim_{T\to\infty} E_{\text{NN}}(T) = 0 . \tag{S26}$$

The above result is indeed consistent with (S23). In fact, for finite dimensional spaces, each image has a unique class associated, so that the Bayes rate is zero.

Nearest neighbor classifier with reading error

Here we generalize the error (S24) to the case where the input is a reconstructed image $\boldsymbol{i}'$, which corresponds to the original image $\boldsymbol{i}$ up to an error probability $p_{\text{read}}(\boldsymbol{i}'|\boldsymbol{i})$ as in (1). When $p_{\text{read}}(\boldsymbol{i}'|\boldsymbol{i}) = \delta_{\boldsymbol{i}',\boldsymbol{i}}$ the reconstruction is perfect, and the definition (S24) applies. On the other hand, when $\boldsymbol{i}' \neq \boldsymbol{i}$, depending on the noise levels, classification algorithms may output the wrong class $\tilde{c}$. Here we focus on the nearest neighbor classifier and generalize the expected classification error (S24) as

$$E_{\text{NN}} = \sum_{\boldsymbol{i}'} \mathop{\mathbb{E}}_{(c,\boldsymbol{i}),\tilde{c},\mathcal{T}}[L_{c,\tilde{c}}P(\tilde{c}|\boldsymbol{i}',\mathcal{T})p_{\text{read}}(\boldsymbol{i}'|\boldsymbol{i})] \tag{S27}$$
$$= 1 - \sum_{c}\sum_{\boldsymbol{i}'} \mathop{\mathbb{E}}_{\boldsymbol{i},\mathcal{T}}[P(c|\boldsymbol{i})P(c|\boldsymbol{i}',\mathcal{T})p_{\text{read}}(\boldsymbol{i}'|\boldsymbol{i})] .$$

In the above equation, we assume that $(c, \boldsymbol{i})$ are sampled from the unknown distribution $P(c, \boldsymbol{i})$ that describes all possible images and their *true* label. As in the previous section, we assume that the training set is made by samples from $P(c, \boldsymbol{i})$. Then, Eq. (S27) can be interpreted as follows: we sample a new physical image $\boldsymbol{i}$ and the corresponding true label c, which is unknown to us. This physical image is measured by a sensor and we get a reconstructed image $\boldsymbol{i}'$ with probability (1). Then we apply the nearest neighbor classifier to find the closest image $\tilde{\boldsymbol{i}}$ to $\boldsymbol{i}'$ from the training set, and get its corresponding label $\tilde{c}$. A *loss* matrix $L_{c,\tilde{c}}$ is the used to weigh the error when $\tilde{c} \neq c$.

Thanks to the above definition, the optimal measurement in (1) is then the one minimizing the expected classification error

$$E^*_{\rm NN} = \min_{\{\Pi_{i'}\}} E_{\rm NN}. \tag{S28}$$

Employing (S25) and taking expectation values with respect to the training set we find

$$E_{\rm NN} = 1 - \sum_c \sum_{\boldsymbol{i}',\tilde{\boldsymbol{i}}} \mathop{\mathbb{E}}_{\boldsymbol{i}} \left[P(c|\boldsymbol{i}) P(c|\tilde{\boldsymbol{i}}) P_{\rm NN}(\tilde{\boldsymbol{i}}|\boldsymbol{i}') p_{\rm read}(\boldsymbol{i}'|\boldsymbol{i}) \right] .$$

If a given image belongs to a single class, as it is the case with the nearest neighbor classifier, we may write $P(c|\boldsymbol{i}) = \delta_{c,f(\boldsymbol{i})}$ for a certain classifier function f, and accordingly

$$\sum_c P(c|\boldsymbol{i}) P(c|\boldsymbol{j}) = \delta_{f(\boldsymbol{i}),f(\boldsymbol{j})} . \tag{S29}$$

so that we may write

$$E_{\rm NN} = 1 - \sum_{\boldsymbol{i},\boldsymbol{i}',\tilde{\boldsymbol{i}}} \Delta(\boldsymbol{i},\tilde{\boldsymbol{i}}) P_{\rm NN}(\tilde{\boldsymbol{i}}|\boldsymbol{i}', \mathcal{T}) p_{\rm read}(\boldsymbol{i}'|\boldsymbol{i}) \pi_{\boldsymbol{i}} , \tag{S30}$$

For the nearest neighbor classifier, f is the function that associates to an image from either the training or testing set the corresponding class. For instance, in Eq. (S30) the image $\boldsymbol{i}$ is from the testing set, say with class c, and the image $\tilde{\boldsymbol{i}}$ is the nearest neighbor to the noisy reconstructed image, and accordingly $\tilde{\boldsymbol{i}}$ is from the training set. Calling $\tilde{c}$ the corresponding class from the training set, $\Delta(\boldsymbol{i},\tilde{\boldsymbol{i}}) = \delta_{c,\tilde{c}}$. Moreover, in the previous section we have seen that, when T is much bigger than the dimension of the image space ($T \gg 2^N$ in this case), we may write $P_{\rm NN}(\boldsymbol{j}|\boldsymbol{i}) = \delta_{\boldsymbol{i},\boldsymbol{j}}$ and we finally get

$$E_{\rm NN} \overset{T\gg 2^N}{\approx} \mathop{\mathbb{E}}_{\boldsymbol{i}} \left[\sum_{\boldsymbol{i}'} (1 - \delta_{f(\boldsymbol{i}),f(\boldsymbol{i}')}) p_{\rm read}(\boldsymbol{i}'|\boldsymbol{i}) \right] \leq p_{\rm err} . \tag{S31}$$

Therefore, in the limit $T \to \infty$ the classification error is smaller than the error in the state discrimination (2).

Finite sample-error in the nearest neighbor classifier

Here we explictly show that $P_{\rm NN}(\tilde{\boldsymbol{i}}|\boldsymbol{i}, \mathcal{T}) \to \delta_{\boldsymbol{i},\tilde{\boldsymbol{i}}}$ in the limit $T \to \infty$. To prove this, we study the statistics of $d_k = \text{dist}(\boldsymbol{i}^{\mathcal{T}}_k, \boldsymbol{i})$ and define $d_{\rm min} = \min_k d_k$, where we employ the Hamming distance. If the samples $\boldsymbol{i}_k$ are independent and identically distributed we get $P_T(d_{\rm min}{\leq}D) \equiv \mathbb{E}_{\mathcal{T}} P(d_{\rm min}{\leq}D|\boldsymbol{i}, \mathcal{T})$ and

$$\begin{aligned} P_T(d_{\rm min}{\leq}D|\boldsymbol{i}, \mathcal{T}) &= 1 - P_T(d_1{>}D, \ldots, d_T{>}D|\boldsymbol{i}) \\ &= 1 - (1 - P(d{\leq}D|\boldsymbol{i}))^T , \end{aligned} \tag{S32}$$

where

$$P(d{\leq}D|\boldsymbol{i}) = \sum_{k=1}^{D} P(\text{dist}(\boldsymbol{i}', \boldsymbol{i}){=}k|\boldsymbol{i}) \tag{S33}$$

is the probability that a new image $\boldsymbol{i}'$ has at most distance D from the given image $\boldsymbol{i}$. From Eq. (S32) we get

$$\begin{aligned} P_T(d_{\rm min}{=}0|\boldsymbol{i}) \equiv P_T(d_{\rm min}{\leq}0|\boldsymbol{i}) = \\ = (1 - (1 - P(d{=}0|\boldsymbol{i}))^T) , \end{aligned} \tag{S34}$$

and for any $D > 0$ we write

$$P_T(d_{\rm min}{=}D|\boldsymbol{i}) \equiv P(d_{\rm min}{\leq}D|\boldsymbol{i}) - P(d_{\rm min}{\leq}D{-}1|\boldsymbol{i}) . \tag{S35}$$

From the above equation and from (S32), it is clear that, as long as $P(d{=}0|\boldsymbol{i}) > 0$, for $T \to \infty$ we find

$$P_T(d_{\rm min}{=}0|\boldsymbol{i}) \to 1, \qquad P_T(d_{\rm min}{>}0|\boldsymbol{i}) \to 0. \tag{S36}$$

In particular, if the set of images is finite and of dimension N, then $P(d{=}0|\boldsymbol{i}) \approx N^{-1}$ and a training set with $T \gg N$ is required to approximate the limit (S36). For any $\boldsymbol{i}$, the solution of $d_{\rm min} = 0$ is unique and, therefore, we have that $\boldsymbol{i}^{\mathcal{T}}_k = \boldsymbol{i}$ for some k.

We may get the above result more explicitly by defining an order between images

$$\boldsymbol{j} \preceq_{\boldsymbol{i}} \boldsymbol{k} \qquad \Leftrightarrow \qquad \text{dist}(\boldsymbol{i},\boldsymbol{j}) \leq \text{dist}(\boldsymbol{i},\boldsymbol{k}) , \tag{S37}$$

so that

$$P(\boldsymbol{j} \preceq_{\boldsymbol{i}} \boldsymbol{k}|\boldsymbol{i}) = \sum_{\text{dist}(\boldsymbol{i},\boldsymbol{j}) \leq \text{dist}(\boldsymbol{i},\boldsymbol{k})} p_{\boldsymbol{j}} , \tag{S38}$$

$$P(\boldsymbol{j} \prec_{\boldsymbol{i}} \boldsymbol{k}|\boldsymbol{i}) = \sum_{\text{dist}(\boldsymbol{i},\boldsymbol{j}) < \text{dist}(\boldsymbol{i},\boldsymbol{k})} p_{\boldsymbol{j}} . \tag{S39}$$

With the above definitions, by calling $P_{\rm NN}(\tilde{\boldsymbol{i}}|\boldsymbol{i}) = \mathbb{E}_{\mathcal{T}} P_{\rm NN}(\tilde{\boldsymbol{i}}|\boldsymbol{i}, \mathcal{T})$, we get

$$P_{\rm NN}(\tilde{\boldsymbol{i}}|\boldsymbol{i}) := P_T(\boldsymbol{j} \preceq_{\boldsymbol{i}} \tilde{\boldsymbol{i}}|\boldsymbol{i}) - P_T(\boldsymbol{j} \prec_{\boldsymbol{i}} \tilde{\boldsymbol{i}}|\boldsymbol{i}) , \tag{S40}$$

$$P_T(\boldsymbol{j} \preceq_{\boldsymbol{i}} \tilde{\boldsymbol{i}}|\boldsymbol{i}) = 1 - (1 - P(\boldsymbol{j} \preceq_{\boldsymbol{i}} \tilde{\boldsymbol{i}}|\boldsymbol{i}))^T , \tag{S41}$$

and, in particular, for $\tilde{\boldsymbol{i}} = \boldsymbol{i}$ we find

$$P_{\rm NN}(\boldsymbol{i}|\boldsymbol{i}) := 1 - (1 - p_{\boldsymbol{i}})^T . \tag{S42}$$