

# Pushing the limits of optical information storage using deep learning

Peter R. Wiecha<sup>1\*</sup>, Aurélie Lecestre<sup>2</sup>, Nicolas Mallet<sup>2</sup> and Guilhem Larrieu<sup>2</sup>

**Diffraction drastically limits the bit density in optical data storage. To increase the storage density, alternative strategies involving supplementary recording dimensions and robust readout schemes must be explored. Here, we propose to encode multiple bits of information in the geometry of subwavelength dielectric nanostructures. A crucial problem in high-density information storage concepts is the robustness of the information readout with respect to fabrication errors and experimental noise. Using a machine-learning-based approach in which the scattering spectra are analysed by an artificial neural network, we achieve quasi-error-free readout of sequences of up to 9 bits, encoded in top-down fabricated silicon nanostructures. We demonstrate that probing few wavelengths instead of the entire spectrum is sufficient for robust information retrieval and that the readout can be further simplified, exploiting the RGB values from microscopy images. Our work paves the way towards high-density optical information storage using planar silicon nanostructures, compatible with mass-production-ready complementary metal-oxide-semiconductor technology.**

Optical information storage promises perennial longevity, high information densities and low energy consumption compared to magnetic storage media<sup>1,2</sup>. The compact disk, the DVD and the Blu-ray disk broadly established optical storage in our society<sup>3,4</sup>. Those media are based on storing a single bit per diffraction-limited area ('zero' or 'one'). Several concepts have been proposed to increase the information density in optical storage. Examples are polarization-sensitive digits<sup>5</sup>, near-field optical recording<sup>6</sup>, the use of fluorescent dyes<sup>7</sup> or three-dimensional (3D) approaches like two-photon point excitation<sup>8</sup>. Yet, all these alternatives suffer from major drawbacks. Either they are hardly superior to commercial planar solutions (polarization-sensitive patterns) or they require very complex storage media (fluorescence) or sophisticated readout schemes (near-field recording, two-photon point excitation). The most promising alternative seemed to be holographic memory, which makes use of the volume of the storage medium, yet there is no commercial product available<sup>9,10</sup>.

Photonic nanostructures are powerful instruments to control light at the nanometre scale<sup>11,12</sup>. Localized surface plasmons (LSPs) in metal nanoparticles<sup>13</sup> or Mie-type resonances in high-index dielectric structures<sup>14</sup> offer high scattering efficiencies, cover the entire visible spectrum and can be tuned by designing appropriate geometric features<sup>15,16</sup>. As a consequence, it has been suggested to encode information in the rich scattering spectra of plasmonic nanostructures, more densely than a single bit<sup>17–21</sup>. The information density might be further increased by addressing layer-wise arranged nanostructures via the focal depth<sup>22</sup> or by light polarization<sup>23</sup>. A key problem of such an approach is the availability and accuracy of readout schemes<sup>18,24</sup>. The main difficulty lies in the fact that different nanostructure geometries can lead to quite similar optical responses, which need to be unambiguously identified during information retrieval<sup>18,25</sup>. Inevitable structural defects and experimental noise further complicate the problem. Drawbacks are also associated with the metallic character of plasmonic nanostructures. One problem is the limited scalability of the production. Another inconvenience with gold (the most common plasmonic material) is the limitation to wavelengths above the interband

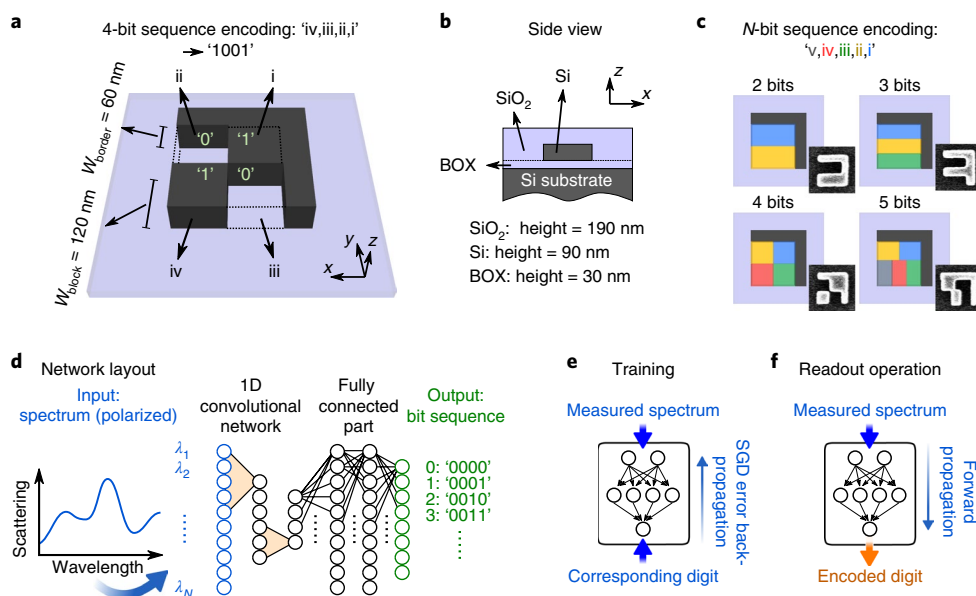
transitions (hence larger than ~520 nm). Shorter wavelengths cannot be used to encode information, which effectively reduces the attainable information density.

To overcome these limitations we have developed a scheme for digital information encoding that is based on silicon nanostructures. Owing to the high refractive index (with low imaginary part) of silicon, silicon nanostructures support low-loss optical resonances that are tunable over the entire visible spectral range<sup>14,16,26</sup>. Silicon also has great technological advantages, in particular its mass-production-ready, high-precision complementary metal-oxide-semiconductor (CMOS)-based processing technology, its low cost and its durability. To reliably retrieve the stored information, we propose a machine-learning-based approach, in which the scattering spectra are analysed by an artificial neural network (ANN). ANNs are computational schemes that can be 'trained' to efficiently solve problems that are hard for classical computer arithmetic<sup>27,28</sup>. ANNs have many applications, for instance in modern smartphones or in medical image interpretation<sup>29,30</sup>. Other than a few recent examples, however, machine learning has rarely been applied to problems in nano-optics. In one work, robust data readout from holographic memory was realized using convolutional ANNs<sup>31</sup>. The potential of ANNs has also been demonstrated in the classification and inverse design of nanoparticles<sup>32–34</sup>. The digital-information retrieval neural network is trained on the measured scattering spectra of several hundred fabricated copies of each nanostructure corresponding to a binary sequence. On all our experimental data sets encoding up to 9 bits of information, effectively going beyond the data density of the Blu-ray disk, the trained ANNs yield a quasi-error-free readout. We demonstrate furthermore that accurate retrieval can be achieved using the scattering intensity at only a small number of discrete wavelengths or even simply the RGB (red–green–blue) colour values from dark-field microscopy images. The latter approach allows, in principle, a massively parallel readout of the stored data.

## Silicon nanostructures for digital information encoding

As illustrated in Fig. 1a,b, we use a planar array-like geometry to encode several bits of information in a single silicon nanostructure.

<sup>1</sup>CEMES, Université de Toulouse, CNRS, Toulouse, France. <sup>2</sup>LAAS, Université de Toulouse, CNRS, INP, Toulouse, France. \*e-mail: [peter.wiecha@cemes.fr](mailto:peter.wiecha@cemes.fr)



**Fig. 1 | Sketch of the nanostructure geometry and the 1D convolutional ANN.** **a**, Illustration of the 4-bit nanostructure geometry. The digital information is encoded in the four silicon blocks (block: '1'; no block: '0'). The structure corresponds to the 4-bit digit '1001' (decimal '9'). The L-shaped sidewall is necessary to distinguish symmetric arrangements via polarized optical spectroscopy. **b**, Side view of the fabricated structure layout. **c**, Sketches of the different geometry models used to experimentally encode sequences of 2, 3, 4 or 5 bits in each nanostructure. In all cases, the height is 90 nm and the L-shaped wall is 60 nm wide. In the 2-bit geometry, the blocks are  $120 \times 240 \text{ nm}^2$ . The 3-bit geometry uses blocks of  $80 \times 240 \text{ nm}^2$ . For the 4-bit encoding, the blocks occupy areas of  $120 \times 120 \text{ nm}^2$ . In the 5-bit geometry, the two upper blocks measure  $150 \times 100 \text{ nm}^2$  and the three lower ones have a size of  $100 \times 150 \text{ nm}^2$ . Insets: selected SEM images show areas of  $450 \times 450 \text{ nm}^2$ . **d**, Sketch of the 1D convolutional neural network used for the classification task. **e**, Training stage of the ANN: measured spectra and corresponding digital information are fed into the network. The error is back-propagated using a variant of the stochastic gradient descent (SGD) algorithm. **f**, The trained network is capable of retrieving the digital information encoded in the structures via their spectra.

If a certain position in the 2D array contains a silicon block, the particular bit is set to '1'; otherwise it is '0'. To distinguish symmetric or rotational arrangements (for example, 4-bit '0010' and '0100'), an L-shaped silicon frame was added, surrounding two sides of the structure. In this way, under linearly polarized illumination each binary number yields a unique spectral response.

We fabricated nanostructures encoding between 2 and 5 bits of information each, as illustrated in Fig. 1c, using electron-beam lithography (EBL) and subsequent dry-etching of commercial silicon-on-insulator substrates with a silicon overlayer with a height of  $H=90 \text{ nm}$ . Subsequent to the etching, the structures were covered by a protective  $\text{SiO}_2$  layer (190 nm height; Fig. 1b). For more details on the fabrication process, see Methods.

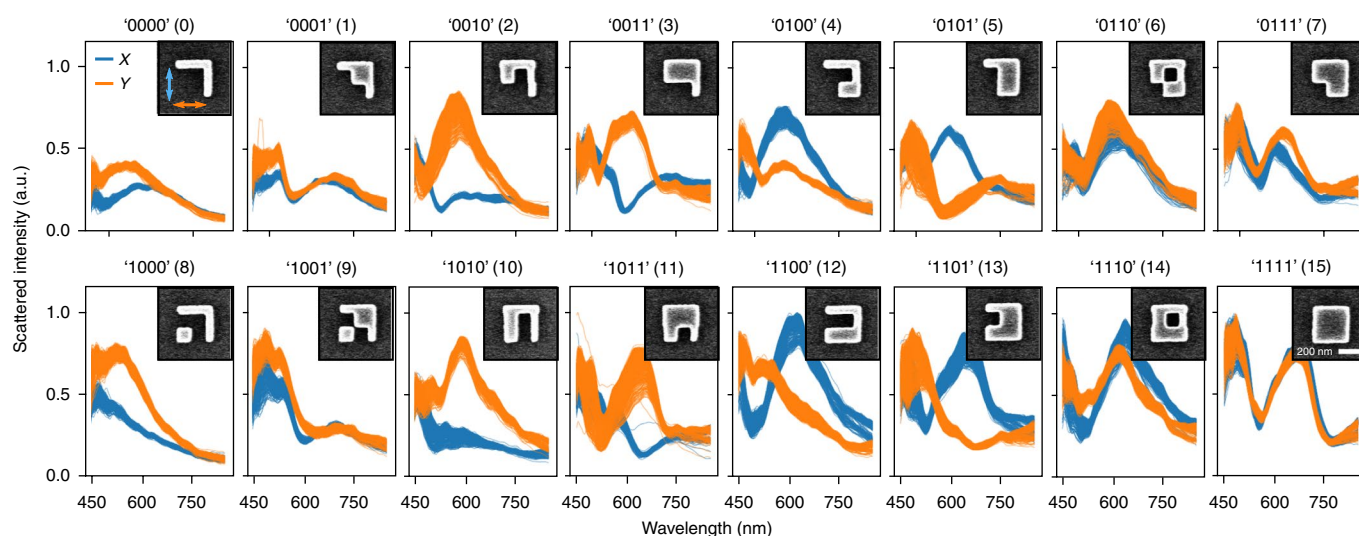
We fabricated  $25 \times 25 = 625$  copies of each geometry. Using an automated set-up with an XY piezo stage, we measured the linearly X- and Y-polarized dark-field spectra of each copy of the structures. All acquired spectra for the 4-bit case together with representative scanning electron microscopy (SEM) images are shown in Fig. 2, superposed and grouped according to the 16 digital numbers. The spectra for the 2-, 3- and 5-bit data sets are provided in Supplementary Figs. 4–11. We note that the spectra of symmetric structures are not completely identical for crossed polarizations (see, for example, '0000'). We attribute this observation to the EBL being horizontally raster-scanned in the fabrication process, leading to small anisotropies ( $\sim 5\text{--}10\%$ ) between horizontal and vertical features (SEM images in Supplementary Figs. 1–3 and 21). This can, in principle, be corrected in an automated fashion during the mask design, as demonstrated in Supplementary Figs. 27 and 28. On the other hand, such asymmetries can even be advantageous for our purpose, because they increase the 'uniqueness' of the individual geometries and their scattering spectra.

### Machine-learning-based digital information readout

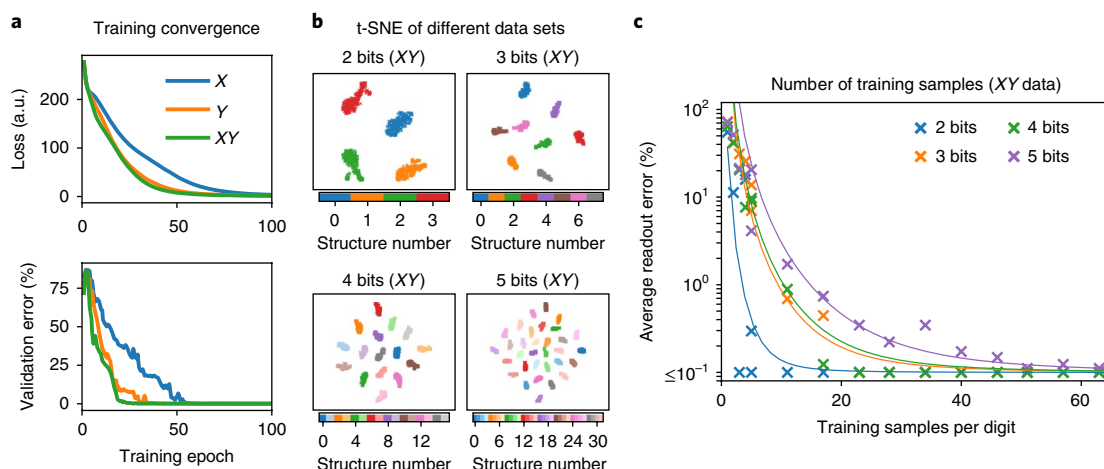
Our goal is to read the information encoded in the geometries of the silicon structures via a far-field optical measurement scheme. The optical scattering spectrum is a promising physical quantity to differentiate between the different structures, in other words to retrieve the bit sequences they represent. Here, we propose a machine-learning approach to the problem. We train an ANN using subsets of the acquired spectra. Subsequently, we evaluate the accuracy of the readout by testing the trained ANN with spectra not used for training.

We use a 1D convolutional neural network (ConvNet), followed by a fully connected network (Fig. 1d), an architecture with particular strength for pattern-recognition tasks<sup>28</sup>. The spectra are fed into the network input layer, which consists of one or two parallel channels, depending on whether a single or both polarization cases are used. At the 'softmax' output layer, each bit sequence is attributed to one neuron. Details on the network and training parameters as well as the preprocessing of the data are provided in the Methods. The network loss and the error rate on the validation set are shown in Fig. 3a for the first 100 epochs of training on the 4-bit data set (note that an 'epoch' is a machine learning term that signifies one full training iteration; in each epoch, the full training data set is randomly shuffled and used in its totality to optimize the network parameters).

The readout scheme is illustrated in Fig. 1f: The scattering spectra of the binary structures are fed into the trained ANN and forward-propagated through the network. The output neuron with the highest activation indicates the encoded bit sequence. In all cases (2–5 bits per structure; X or Y or both (XY) polarizations), the trained ANN yields quasi-error-free readout accuracy. In the cases of 4 and 5 bits, one single spectrum of the test data was incorrectly



**Fig. 2 | Experimental dark-field spectra training data set for 4 bits.** Our data comprise measurements from 625 copies for each of the 16 ‘4-bit’ geometries (a total of  $625 \times 16 \times 2 = 20,000$  acquired spectra). The spectra are superposed above each other. Blue lines show dark-field scattering for X-polarized light; orange lines show Y-polarization. Insets, SEM images of one representative copy of the respective nanostructure; the areas are  $600 \times 600$  nm<sup>2</sup>. The 2-bit data set, not shown, consists of 5,000 scattering spectra, the case of 3 bits comprises 10,000 spectra and the 5-bit set contains 40,000 spectra (shown together with SEM images in Supplementary Figs. 4–11, as well as a separate 4-bit data set and simulations to demonstrate reproducibility, shown in Supplementary Figs. 12–14).

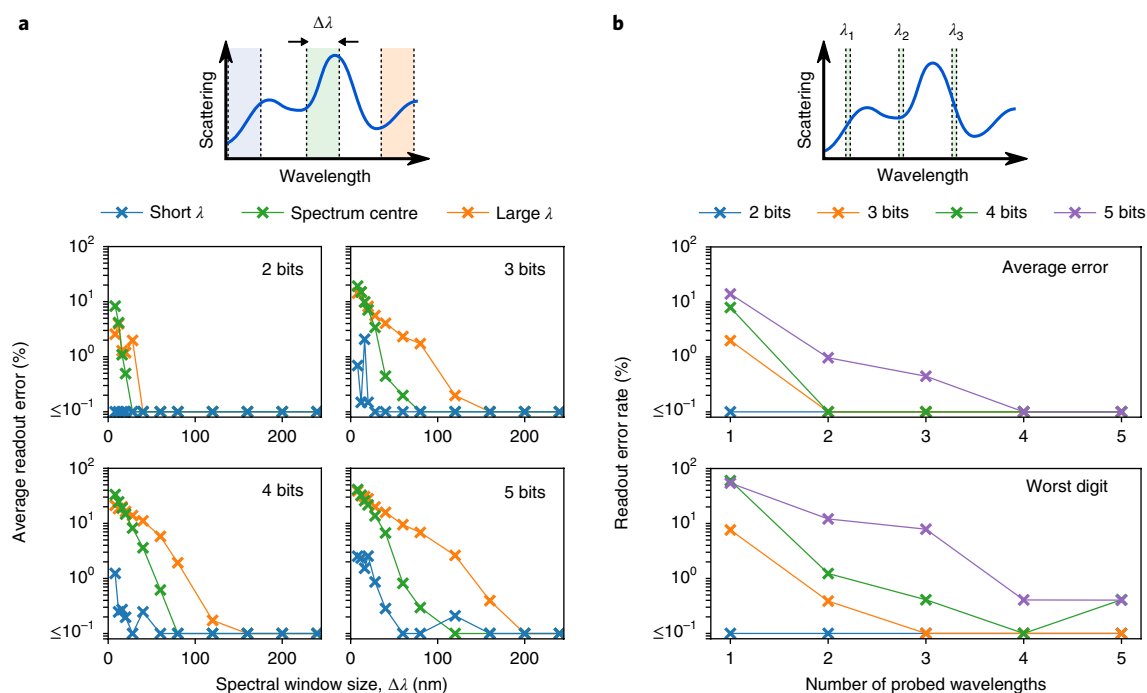


**Fig. 3 | Training convergence and readout accuracy of the ANN trained on the full scattering spectra.** **a**, Convergence of training for the example of 4-bit encoding and with spectra of 300 copies per digit for training. Loss (categorical cross-entropy, top) and validation error rate (bottom) are shown as a function of training epoch for only X (blue) and only Y (orange) polarized spectra and for both polarizations (green) as training data. **b**, t-SNE<sup>35</sup> visualization (see also main text) of the XY polarization training sets for 2–5 bits encoded per structure. **c**, Readout error as a function of the number of spectra per structure used for training of the network. All cases of information density per structure yield quasi-error-free readout using sufficiently large training sets of at most around 60 samples per geometry (5-bit case, purple crosses). As one would expect, the lower bit densities (blue, 2-bit; orange, 3-bit; green, 4-bit) require even fewer training samples for accurate information retrieval. Solid lines are guides to the eye, proportional to  $N^{-3}$ , which qualitatively describe the convergence of readout as function of training set size  $N$ .

interpreted (corresponding to 0.023% and 0.011% for the 4,400 (4-bit) and 8,800 (5-bit) test structures, respectively).

We analysed the data sets using the t-SNE (t-distributed stochastic neighbour embedding) dimensionality reduction to estimate the distinguishability between different geometries, as well as the variance in copies of identical structures<sup>35</sup>. In a t-SNE plot, well separated scatter points correspond to unambiguously differentiable entities in the data set, while nearby and overlapping points correspond to very similar data. The results for the XY data sets are shown in Fig. 3b, and the t-SNE plots for the X and Y data sets are

provided in Supplementary Fig. 15. Each colour corresponds to one type of structure and each dot represents a single measurement. All data sets are characterized by a very good separation of the different spectra in the t-SNE plots, which explains why the ANN can retrieve the binary information with almost no errors. Figure 3c shows the readout error as function of the training samples. As intuitively expected, the required number of training spectra for error-free operation increases with the complexity of the geometrical model. On the other hand, we observed that the main source for dispersion in the measurements is the position of the confocal



**Fig. 4 | Accuracy of network trained on reduced spectral information.** **a**, Training using scattering from a spectral window of reduced width for the 2-, 4- and 5-bit data sets. The scattering intensity was taken from a window either at the short wavelength side (blue lines), in the centre (green) or from the red edge of the spectra (orange). **b**, Average (top plot) and worst-digit (bottom plot) readout accuracy using a network trained on the scattering intensity of X- and Y-polarized light at a discrete number of wavelengths. The explicit positions of  $\lambda_i$  for the different cases are provided in the Methods. t-SNE plots for all reduced spectral information data sets are provided in Supplementary Fig. 15.

hole with respect to the nanostructure. Hence, the spectral dispersion could be easily reduced by optimizing the stability of the acquisition scheme. We also want to emphasize that our conditions are still perfectly sufficient for a quasi-error-free operation in all considered cases.

To assess the amount of optical data required for an accurate readout, we train and test the ANNs using reduced spectral information. Figure 4a shows the error rate as function of the spectral window width  $\Delta\lambda$  used for training and retrieval (on the XY data sets). We evaluated three different positions of the spectral window: at the long wavelength end of the spectra ( $\lambda \leq 850$  nm, orange lines), at short wavelengths ( $\lambda \geq 450$  nm, blue lines) or centred around 650 nm (green lines). While the short and intermediate wavelengths always yield error-free readout for spectral windows as small as  $\leq 100$  nm, we observe that using the red part of the spectra requires a larger spectral window of up to  $\sim 200$  nm for high accuracy. We conclude that the red part of the spectra contains the least amount of information, insufficient to unambiguously distinguish between the binary sequences. This is a direct consequence of the photon energy being inverse proportional to the wavelength. Furthermore, the used geometries have no resonances above  $\sim 750$  nm. Using larger or higher structures could increase the information density in the red by shifting resonances to longer wavelengths. In a second step, we train a fully connected ANN (see Methods) using the scattering intensity only at a low number of discrete wavelengths (Fig. 4b). Probing three wavelengths is always sufficient to obtain a readout accuracy of  $>99\%$ . The error rates on the worst bit sequence (on the order of 10%) can be reduced below 0.5% by probing at four or five wavelengths.

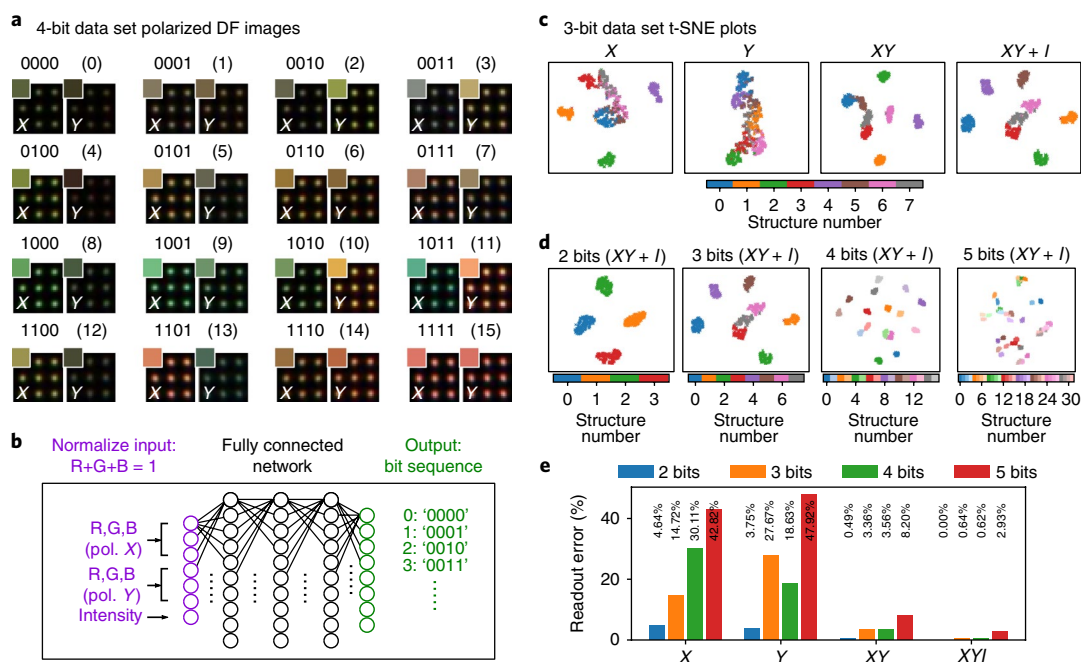
We will study in the following whether the bit sequences can also be recovered using a simpler and faster data acquisition scheme, by using the scattered RGB colour obtained from conventional dark-field microscopy images. Figure 5a shows dark-field images

of  $3 \times 3$  copies for each geometry and both polarizations for the example of the 4-bit geometry. The average RGB colour is shown in the upper left corner of each plot. For training we use the RGB values from the scattered light of each individual structure in the polarization-filtered dark-field image. We normalize the RGB values to  $R + G + B = 1$  and store the scattered intensity separately. We create four training data sets: three sets with only the normalized RGB information (one for X, one for Y and a third combining both polarizations XY), and a fourth set contains the XY data and additionally their intensity values ( $XY + I$ ). We use a fully connected network as depicted in Fig. 5b. Technical details are provided in the Methods.

Figure 5c,d shows t-SNE plots for the different 3-bit data sets and for the XY data of the 2-, 3-, 4- and 5-bit geometries. The partial mixing of scatter points in the t-SNE plots of only X- or only Y-polarized RGB values suggests that these data are not sufficient for accurate identification of the structures. This is confirmed by the insufficient readout performance of the corresponding ANNs, and the error rates of several individual binary numbers are even well above 50% (Fig. 5e and Supplementary Fig. 16).

Using the data sets combining both polarizations, the situation improves drastically. The t-SNE plots (Fig. 5d) now show a clear separation of the different binary numbers. The average error rate drops significantly below 10% in the XY case, and can be reduced further below 1% (2–4 bits) and below 3% (5 bits) when the brightness values are also considered. In the 4-bit XY + I case, for instance, the largest error rate is observed for digits ‘0101’ and ‘1100’, which scatter light with a comparable tone and brightness (Supplementary Fig. 16). The similarity can also be observed in the t-SNE plot in Fig. 5d, where the light green and pink dots (structures ‘5’ and ‘12’) are partly touching. Such an analysis of the t-SNE plots allows us to identify the most problematic digits (for example, ‘6’, ‘8’, ‘29’ and ‘31’ in the 5-bit case). By designing nanostructures with a more significant difference in the scattered colours (for example, using modern





**Fig. 5 | Neural-network based data readout via the RGB colour values.** **a**, Polarization-filtered dark-field colour images of representative  $3 \times 3$  arrays of the 4-bit digit structures. Left, X-polarization; right, Y-polarization. The dark-field images have areas of  $7.5 \times 7.5 \mu\text{m}^2$ . Insets: average RGB colour of the  $3 \times 3$  structures. **b**, Scheme of the fully connected ANN used for the RGB classification task. **c**, t-SNE<sup>35</sup> visualization of the 3-bit training sets for only X, only Y, X + Y or X, Y and scattered intensity  $I$  (from left to right). Only if using simultaneously both polarizations (XY and XY +  $I$ ) do the different bit sequences show a clear separation in the t-SNE plots. **d**, t-SNE plots for the XY +  $I$  cases of the 2-, 3-, 4- and 5-bit training sets (from left to right). **e**, Information retrieval accuracy of the network, trained on the different data sets consisting of only the X-filtered, only the Y-filtered, X + Y-filtered and XY +  $I$ . t-SNE plots for all data sets are provided in Supplementary Fig. 15. In Supplementary Figs. 17 and 18, the activations of the softmax output neurons of the RGB network are shown for the whole validation sets of the 4-bit data sets X and XY +  $I$ .

inverse problem techniques<sup>16,36</sup>), this limitation could be easily overcome and the error rate further decreased.

The RGB colour information allows to simultaneously capture many thousands of structures within a single measurement of a large-area image. In other words, it supports a massively parallel readout of the information (see Methods). Cheap, smartphone-based dark-field microscopy for the RGB readout might become feasible in the near future. Indeed, smartphone-based microscopy is the subject of current research and has undergone tremendous progress in the recent past<sup>37–39</sup>. A further interesting route for improvement concerns specifically designed bright-field colour scatterers<sup>40</sup> for information encoding, to avoid the necessity of a dark-field illumination scheme. Finally, a cheap multi-laser approach, similar to a Blu-ray disk reader with several lasers, could be used to capture the scattering intensity simultaneously at several wavelengths.

### Towards higher information density

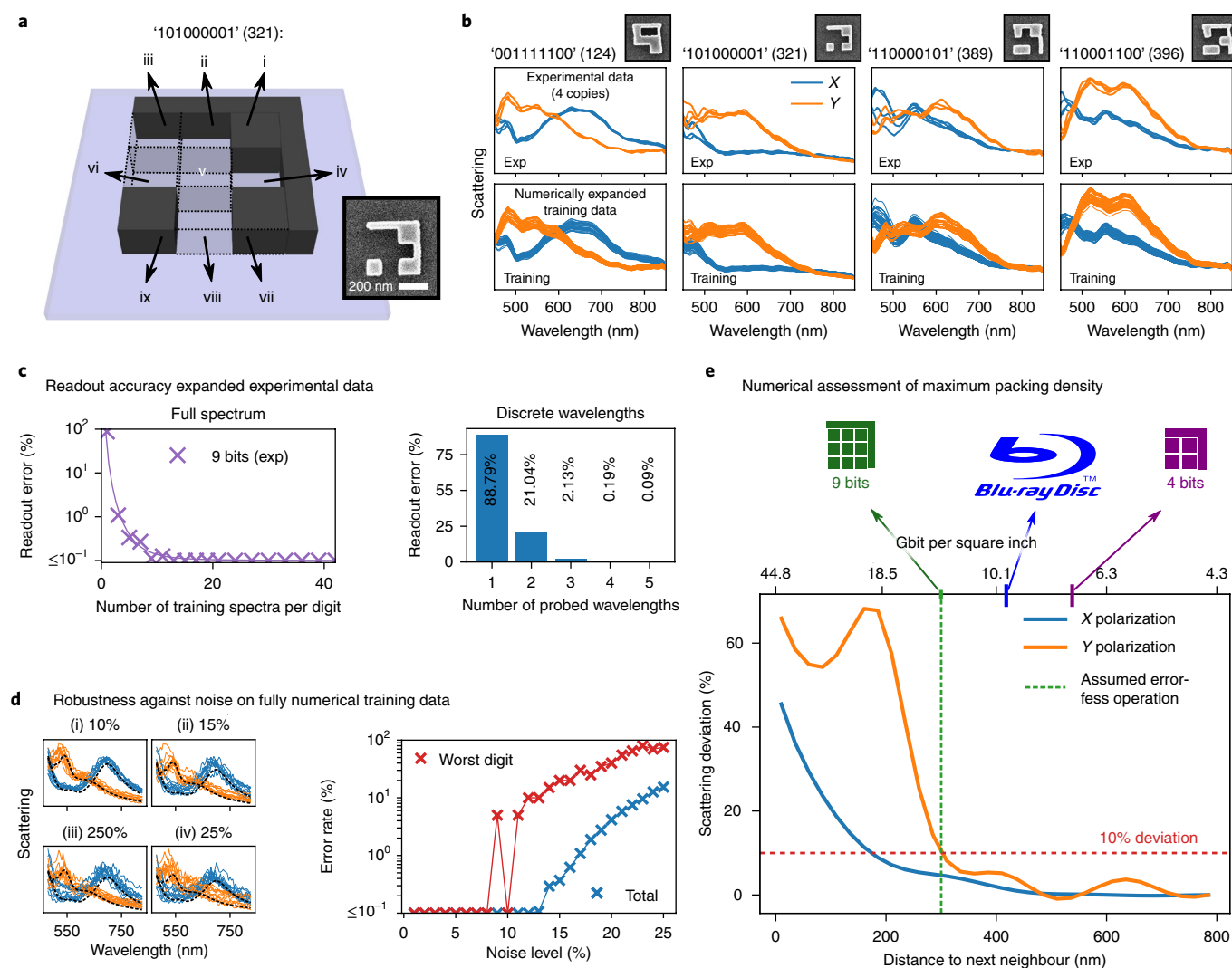
Finally, we want to assess whether the information density of the individual nanostructures can be further increased. As detailed in Fig. 6a, we examine in the following the case of a geometry encoding 9 bits of information, which leads to 512 possible structural arrangements. We note that the ~20% larger size of the 9-bit structures compared to the 2–5-bit geometries leads to a redshift of the resonances. This effectively increases the information content at long wavelengths, which is advantageous for our aim to encode 9 bits per nanostructure.

Having 512 different geometries to distinguish, the main difficulty is now the requirement for a large data set. Unfortunately, while no constraints exist concerning mass-production fabrication, on our scientific equipment, acquiring scattering spectra from

several hundreds of copies per geometry would imply many months of workload. However, there are no general technical constraints, and using dedicated, fully automated set-ups would enable a rapid experimental acquisition of very large data sets. In our case, to obtain a sufficiently large data set we numerically expand a representative set of experimental spectra from four copies of each nanostructure by generating 300 random superpositions of these four spectra. We also introduce a random intensity deviation of up to  $\pm 10\%$ . The results of this numerical data expansion are shown for selected examples in Fig. 6b (see Supplementary Figs. 26 and 29–32 for the whole experimental data set and more examples).

We train networks on 9-bit readout using either the full spectra or the scattering intensity at a discrete number of wavelengths (using simultaneous X and Y polarization). The techniques are identical to those for the above cases of 2–5 bits per nanostructure (see Methods). In the left of Fig. 6c, the readout accuracy using full spectra is shown as a function of training samples per geometry. Quasi-error-free operation is obtained for training sets as small as ~20 spectra per digit. Using a discrete number of probed wavelengths (Fig. 6c, right), very low error rates of ~2% can be obtained by probing only three discrete wavelengths. Using four or five probed wavelengths improves the error far below 1%.

To assess how larger variations and noise influence the readout performance, we use a data set of fully numerical simulations of the 512 geometries (see Methods). These synthetic data allow us to adjust the distortions and noise of the spectra in a quantifiable way, as shown schematically in the left of Fig. 6d. The readout accuracy of an ANN, trained on simulated spectra, is shown on the right of Fig. 6d. Significant error rates start to occur at noise levels around 15%. The average error rate with 20% noise is still at a reasonable value of 4.16%. For the worst digit though, the error rate is



**Fig. 6 | Information encoding with 9 bits per nanostructure.** **a**, Geometry of a silicon nanostructure encoding  $3 \times 3 = 9$  bits (512 possible combinations). Each silicon block occupies an area of  $105 \times 105 \text{ nm}^2$ . The L-shaped sidewall is 45 nm wide, and the height is 90 nm. An SEM image of a fabricated structure is shown in the inset. The shown example represents the decimal number '321'. **b**, Selected examples, illustrating training data generation by a numerical expansion of the experimental spectra. SEM images show areas of  $550 \times 550 \text{ nm}^2$ . We fabricated four copies of each possible 9-bit nanostructure. Via random superposition of the experimental scattering spectra of these four copies, we generated a large set of spectra, allowing us to train and test the performance of the binary information readout ANN (see Supplementary Fig. 23 for more details). **c**, Accuracy of the ANN trained on the experimental 9-bit data using the full spectra (left, the solid line is a guide to the eye) or scattering intensities at a limited number of discrete wavelengths (right). In both cases, X- and Y-polarized data are used simultaneously. **d**, Evaluation of the robustness of 9-bit readout with respect to noise on fully numerical data. We use simulated scattering spectra for all 9-bit geometries with different amounts of random noise (see Supplementary Fig. 33 for details of training data generation). Example spectra for noise levels from (i) 10% to (iv) 25% are shown for the example of structure '001011010' (decimal 90). The plot on the right shows the readout error rate on the numerical data as a function of noise level (training on the full spectra of both polarizations). **e**, Scattering deviation relative to the isolated structure as function of the distance between two digit-encoding structures for focused illumination. For estimation of the feasible information density, 10% deviation from the unperturbed spectrum is assumed to be tolerable (red dashed horizontal line). This leads to an information density (green indicator) around 40% higher than that of a Blu-ray disk (blue indicator). The 4-bit structures yield about 75% of the Blu-ray density.

as high as 35% in this case, which could be improved by designing optically 'more unique' nanostructures for pathological cases. Some examples of false and correct classifications as well as an analysis of the error rate at several noise levels are given in Supplementary Figs. 34 and 35. A comparison of the experimental data with simulations of different noise levels, a t-SNE plot of the experimental 9-bit data set and further analysis on the impact of noise are provided in Supplementary Figs. 22 and 24–26.

We finally compare the information density of the 9-bit nanostructures to the state-of-the-art flat optical storage medium, the 'Blu-ray'

disk. The Blu-ray requires  $67,500 \text{ nm}^2$  per bit, while our approach scales with the structure spacing and uses  $(3 \times 105 + 45 + \text{spacing})^2/9 \text{ nm}^2$  (see Methods, section 'Considerations on the practical implementation of readout schemes and their performance'). In Fig. 6e we analyse the influence of a close neighbour on the scattering from a 9-bit nanostructure. Assuming a tolerable deviation of 10% compared to the isolated nanostructure's spectrum, a spacing of 300 nm would still allow an accurate readout, yielding around 40% higher information density than the Blu-ray (Supplementary Figs. 36 and 37). With a 300 nm spacing, the 4-bit encoding

nanostructures result in 75% of the Blu-ray information density. Using properly designed sets of photonic nanostructures with highly distinguishable optical responses, the accuracy and robustness of the method can be further optimized. Also, by increasing the structure height to redshift the resonances, the area covered by each nanostructure could be reduced. The information per structure could possibly be increased up to 11 or 12 bits (2,048 and 4,096 structure geometries, respectively). However, larger binary sequences seem difficult for an accurate readout, due to the power-scaling by which the number of structure geometries increases ( $2^{N_{\text{bit}}}$ ). A further means to improve the readout accuracy and its robustness might be to include the angle of incidence as a probing parameter. To exploit all three spatial dimensions of the storage medium, multilayer arrangements might be possible<sup>22,23</sup>.

## Conclusions

In summary, we have demonstrated on experimental data that deep neural networks can be effectively trained for the optical retrieval of digital information encoded in the geometry of photonic nanostructures. We have shown, on geometries encoding up to 9 bits per diffraction-limited area, that the optical scattering spectra are more than sufficient for accurate recovery of the encoded data. We have also shown that probing at a few discrete wavelengths or even simply using the RGB colour information obtained from standard dark-field microscopy images provides a precise readout scheme, potentially possible on very simple and cheap equipment. The latter approach would allow a massively parallel retrieval of information. The robustness can easily be improved by tailoring nanostructure geometries with high colour contrast<sup>41</sup>.

Our work paves the way towards high-density planar optical information storage using simple far-field characterization combined with concepts of machine learning. Rewritable storage media might be created around the recently developed technology on catalytic magnesium metasurfaces for dynamic structural colour<sup>42</sup>. Our approach can easily be generalized to other classification tasks in nano-optics, including biological applications and problems in plasmonics.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41565-018-0346-1>.

Received: 11 May 2018; Accepted: 10 December 2018;

Published online: 21 January 2019

## References

- Zhang, J., Gecevičius, M., Beresna, M. & Kazansky, P. G. Seemingly unlimited lifetime data storage in nanostructured glass. *Phys. Rev. Lett.* **112**, 033901 (2014).
- Gu, M., Li, X. & Cao, Y. Optical storage arrays: a perspective for future big data storage. *Light Sci. Appl.* **3**, e177 (2014).
- Satoh, I., Ohara, S., Akahira, N. & Takenaga, M. Key technology for high density rewritable DVD (DVD-RAM). *IEEE Trans. Magn.* **34**, 337–342 (1998).
- Borg, H. J. et al. Phase-change media for high-numerical-aperture and blue-wavelength recording. *Jpn J. Appl. Phys.* **40**, 1592 (2001).
- Zeng, B. J., Ni, R. W., Huang, J. Z., Li, Z. & Miao, X. S. Polarization-based multiple-bit optical data storage. *J. Opt.* **16**, 125402 (2014).
- Tominaga, J., Nakano, T. & Atoda, N. An approach for recording and readout beyond the diffraction limit with an Sb thin film. *Appl. Phys. Lett.* **73**, 2078–2080 (1998).
- Mottaghi, M. D. & Dwyer, C. Thousand-fold increase in optical storage density by polychromatic address multiplexing on self-assembled DNA nanostructures. *Adv. Mater.* **25**, 3593–3598 (2013).
- Strickler, J. H. & Webb, W. W. Three-dimensional optical data storage in refractive media by two-photon point excitation. *Opt. Lett.* **16**, 1780–1782 (1991).
- van Heerden, P. J. Theory of optical information storage in solids. *Appl. Opt.* **2**, 393–400 (1963).
- Psaltis, D. & Burr, G. W. Holographic data storage. *Computer* **31**, 52–60 (1998).
- Girard, C. Near fields in nanostructures. *Rep. Prog. Phys.* **68**, 1883–1933 (2005).
- Novotny, L. & Hecht, B. *Principles of Nano-Optics* (Cambridge Univ. Press, Cambridge, 2006).
- Maier, S. *Plasmonics: Fundamentals and Applications* (Springer, New York, 2010).
- Kuznetsov, A. I. et al. Optically resonant dielectric nanostructures. *Science* **354**, aag2472 (2016).
- Cao, L., Fan, P., Barnard, E. S., Brown, A. M. & Brongersma, M. L. Tuning the color of silicon nanostructures. *Nano Lett.* **10**, 2649–2654 (2010).
- Wiecha, P. R. et al. Evolutionary multi-objective optimization of colour pixels based on dielectric nanoantennas. *Nat. Nanotechnol.* **12**, 163–169 (2017).
- Mansuripur, M. et al. Plasmonic nano-structures for optical data storage. *Opt. Express* **17**, 14001–14014 (2009).
- Chen, W. T. et al. Manipulation of multidimensional plasmonic spectra for information storage. *Appl. Phys. Lett.* **98**, 171106 (2011).
- Cui, Y., Phang, I. Y., Hegde, R. S., Lee, Y. H. & Ling, X. Y. Plasmonic silver nanowire structures for two-dimensional multiple-digit molecular data storage application. *ACS Photon.* **1**, 631–637 (2014).
- El-Rabaa, M. A., Areed, N. F. F. & Obayya, S. S. A. Novel plasmonic data storage based on nematic liquid crystal layers. *J. Lightwave Technol.* **34**, 3726–3732 (2016).
- Zijlstra, P., Chon, J. W. M. & Gu, M. Five-dimensional optical recording mediated by surface plasmons in gold nanorods. *Nature* **459**, 410–413 (2009).
- Taylor, A. B., Kim, J. & Chon, J. W. M. Detuned surface plasmon resonance scattering of gold nanorods for continuous wave multilayered optical recording and readout. *Opt. Express* **20**, 5069–5081 (2012).
- Taylor, A. B., Michaux, P., Mohsin, A. S. M. & Chon, J. W. M. Electron-beam lithography of plasmonic nanorod arrays for multilayered optical storage. *Opt. Express* **22**, 13234–13243 (2014).
- Li, X., Cao, Y., Tian, N., Fu, L. & Gu, M. Multifocal optical nanoscopy for big data recording at 30 TB capacity and gigabits/second data rate. *Optica* **2**, 567–570 (2015).
- Liu, D., Tan, Y., Khoram, E. & Yu, Z. Training deep neural networks for the inverse design of nanophotonic structures. *ACS Photon.* **5**, 1365–1369 (2018).
- Albella, P. et al. Low-loss electric and magnetic field-enhanced spectroscopy with subwavelength silicon dimers. *J. Phys. Chem. C* **117**, 13573–13584 (2013).
- Nielsen, M. A. *Neural Networks and Deep Learning* (Determination Press, 2015); <http://neuralnetworksanddeeplearning.com/>
- Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, Cambridge, 2016).
- Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. Inception-v4, inception-ResNet and the impact of residual connections on learning. Preprint at <https://arxiv.org/abs/1602.07261> (2016).
- Mamoshina, P., Vieira, A., Putin, E. & Zhavoronkov, A. Applications of deep learning in biomedicine. *Mol. Pharm.* **13**, 1445–1454 (2016).
- Shimobaba, T. et al. Convolutional neural network-based data page classification for holographic memory. *Appl. Opt.* **56**, 7327–7330 (2017).
- Jo, Y. et al. Holographic deep learning for rapid optical screening of anthrax spores. *Sci. Adv.* **3**, e1700606 (2017).
- Malkiel, I. et al. Plasmonic nanostructure design and characterization via deep learning. *Light Sci. Appl.* **7**, 60 (2018).
- Peurifoy, J. et al. Nanophotonic particle simulation and inverse design using artificial neural networks. *Sci. Adv.* **4**, eaar4206 (2018).
- van der Maaten, L. & Hinton, G. Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Feichtner, T., Selig, O., Kiunke, M. & Hecht, B. Evolutionary optimization of optical antennas. *Phys. Rev. Lett.* **109**, 127701 (2012).
- Riverson, Y. et al. Deep learning enhanced mobile-phone microscopy. *ACS Photon.* **5**, 2354–2364 (2018).
- Orth, A., Wilson, E. R., Thompson, J. G. & Gibson, B. C. A dual-mode mobile phone microscope using the onboard camera flash and ambient light. *Sci. Rep.* **8**, 3298 (2018).
- Wei, Q. et al. Plasmonics enhanced smartphone fluorescence microscopy. *Sci. Rep.* **7**, 2124 (2017).
- Flauraud, V., Reyes, M., Paniagua-Domínguez, R., Kuznetsov, A. I. & Brugger, J. Silicon nanostructures for bright field full color prints. *ACS Photon.* **4**, 1913–1919 (2017).
- González-Alcalde, A. K. et al. Optimization of all-dielectric structures for color generation. *Appl. Opt.* **57**, 3959–3967 (2018).
- Duan, X., Kamin, S. & Liu, N. Dynamic plasmonic colour display. *Nat. Commun.* **8**, 14606 (2017).

## Acknowledgements

The authors thank A. Arbouet and C. Girard for their advice, for their help and for discussing and proofreading the manuscript, and F. Carcenac for his help with EBL

and automatic SEM images. This work was supported by Programme Investissements d'Avenir under the program ANR-11-IDEX-0002-02, reference ANR-10-LABX-0037-NEXT, by the LAAS-CNRS micro and nanotechnologies platform, a member of the French RENATECH network, and by the computing facility centre CALMIP of the University of Toulouse under grant P12167.

### Author contributions

P.R.W. conceived the idea and designed the research together with G.L. G.L. and A.L. developed the fabrication techniques. A.L. fabricated the nanostructures and performed the electron microscopy with the help of N.M. P.R.W. carried out the optical experiments, did the simulations and the data analysis, and implemented the machine-learning part. P.R.W. wrote the manuscript with contributions from G.L. P.R.W. and G.L. discussed the results and all authors commented on the manuscript at every stage.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41565-018-0346-1>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to P.R.W.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019



## Methods

**Nanofabrication of planar Si structures.** The silicon nanostructures were patterned on a silicon on insulator (SOI) substrate (90 nm active Si layer on 30 nm buried oxide) following a large-scale top-down approach. EBL was used to pattern an 80-nm-thick layer of an inorganic negative-tone resist, hydrogen silsesquioxane (HSQ). After exposure, the HSQ was developed by immersion in 25% tetramethylammonium hydroxide (TMAH) for 1 min (ref. <sup>43</sup>). HSQ patterns were subsequently transferred to the silicon substrate down to the buried oxide by anisotropic reactive ion etching. The structures were then embedded in a 200-nm-thick HSQ layer, deposited by spin coating to perfectly planarize the sample in a nanometrical range<sup>44</sup>. Finally, the HSQ layer ( $\text{SiO}_2\text{H}_2$ ) was converted into a  $\text{SiO}_2$  layer by rapid thermal annealing at 600 °C for 2 min under nitrogen ambience, leading to a final  $\text{SiO}_2$  layer with thickness of 190 nm.

**Confocal dark-field microscopy.** The scattering spectra were acquired on a confocal dark-field microscope (Horiba XploRA). A white lamp was focused on the sample with a  $\times 50$  dark-field objective (NA = 0.5; condenser NA = 0.8–0.95). The backscattered light was filtered by a confocal hole (diameter of 100  $\mu\text{m}$ ) and a polarization filter and dispersed by a grating (300 grooves  $\text{mm}^{-1}$ ) on a Peltier-cooled charge-coupled device (CCD). The acquisition time was  $t_a = 0.2$  s. All spectra were normalized by the spectrum of the lamp. While the spectra used for the results shown in the main text were measured at an acquisition time of  $t_a = 0.2$  s, a second data set was measured with  $t_a = 0.5$  s, leading to similar results (Supplementary Figs. 12–14). We therefore assume that even shorter acquisition times would be sufficient for robust data recognition.

The polarization-filtered dark-field images were taken using the same  $\times 50$  dark-field microscope objective with a colour CCD camera at a resolution of  $1,392 \times 1,040$  and an exposure time of  $t_{\text{exp}} = 0.05$  s. Each structure was perceived as a coloured dot on the microscopy image, covering  $\sim 20$ – $50$  pixels, from which we take the average RGB value.

**Expansion of the 9-bit experimental training data set.** Due to the technical constraints of our measurement set-up, we are not able to acquire, in reasonable time, scattering spectra of several hundreds of all 512 possible 9-bit encoding nanostructures. To nevertheless assess the feasibility of experimental 9-bit readout via our machine-learning-based approach, we fabricate four copies of the 9-bit structures. To obtain the required much larger number of spectra for each geometry, we subsequently created new spectra  $\sigma_{\text{new}}(\lambda)$  from random superpositions of the four experimental spectra  $\sigma_i(\lambda)$  of each nanostructure type. Our condition in this approach is that the random weights  $w_i$  of the four different spectra sum to 1, hence

$$\sigma_{\text{new}}(\lambda) = \sum_{i=1}^4 w_i \sigma_i(\lambda), \quad \text{with} \quad \sum_{i=1}^4 w_i = 1. \quad (1)$$

Finally, we multiply the spectrum by a random coefficient  $C$  between 0.9 and 1.1 to emulate larger intensity fluctuations. In this way we generate 300 semi-experimental spectra with which we train the ANNs for the decoding of the digital information.

**Electrodynamical simulations. Simulated data set.** The 9-bit geometry model used for the fully numerically simulated data set is identical to the experimental structures (Fig. 6a). It consists of  $3 \times 3$  silicon blocks, each  $105 \times 105 \text{ nm}^2$  lateral size and 90 nm height. Two sides are surrounded by an L-shaped block of 45 nm width.

We numerically simulate the scattering spectra under X- and Y-polarized plane wave illumination for all 512 possible geometries using the Green dyadic method (GDM). To assess the robustness against different amounts of perturbation and noise in the spectra, we numerically add noise to the simulated data via a sequence of random modifications. The noise addition steps are illustrated in Supplementary Fig. 33. First we add random noise, apply a random positive offset as well as a scaling factor. Then we multiply the spectra with a linear function of random slope and finally apply a wavelength shift. To do so, in the first place we simulate the spectra on an extended wavelength range. Thanks to this procedure, we are able to adjust the magnitude of the random variations to yield more or less strongly distorted results. Figure 6d shows several randomized spectra for noise magnitudes between 10% (i) and 25% (iv).

**Green dyadic method.** The numerical simulations for the 9-bit structures are performed using the GDM, a frequency-domain approach based on the concept of a generalized propagator<sup>45</sup>. In particular, we use our own implementation in python, ‘pyGDM’<sup>46</sup>.

In the GDM the volume of a nanostructure is discretized with  $N$  cubic meshpoints of edge length  $d$ . A dipolar response is attributed to each of these meshpoints. As detailed, for example, in ref. <sup>47</sup>, this leads to a system of  $3N$  coupled equations:

$$\mathbf{E}_0 = \mathbf{M} \cdot \mathbf{E} \quad (2)$$

which, by inversion of  $\mathbf{M}$ , allows us to relate any incident electric field  $\mathbf{E}_0$  to the induced field  $\mathbf{E}$  inside the particle.  $\mathbf{M}$  is composed of  $3 \times 3$  submatrices:

$$\mathbf{M}_{ij} = \mathbf{I} \delta_{ij} - \alpha_i(\omega) \mathbf{G}(\mathbf{r}_i, \mathbf{r}_j, \omega) \quad (3)$$

where  $\mathbf{I}$  is the Cartesian unitary tensor,  $\delta_{ij}$  the Kronecker delta function and, in c.g.s. units,

$$\alpha_i(\omega) = \frac{\epsilon_i(\omega) - \epsilon_{\text{env}}(\omega)}{4\pi} v_i \quad (4)$$

In equation (4),  $v_i$  is the volume of the cubic discretization cells, and hence  $v_i = d^3$ . We use the dispersion of silicon from ref. <sup>48</sup> for  $\epsilon_i$ . The structures are placed in a homogeneous environment ( $\epsilon_{\text{env}} = 1.45$  for  $\text{SiO}_2$ ) at 30 nm above a silicon substrate.

$\mathbf{G}$  in equation (3) is the Green’s dyad, coupling the cubic meshpoints  $i$  and  $j$ . It is composed of a vacuum term (accounting for the homogeneous environment) and a surface term (accounting for the substrate):

$$\mathbf{G}(\mathbf{r}_i, \mathbf{r}_j, \omega) = \mathbf{G}_0(\mathbf{r}_i, \mathbf{r}_j, \omega) + \mathbf{G}_{\text{surf}}(\mathbf{r}_i, \mathbf{r}_j, \omega) \quad (5)$$

which can be found in the literature<sup>47</sup>. At  $\mathbf{r}_i = \mathbf{r}_j$  the Green’s dyad  $\mathbf{G}_0$  diverges, hence a normalization scheme is applied:

$$\mathbf{G}_0(\mathbf{r}_i, \mathbf{r}_j, \omega) = \mathbf{I} C(\omega) \quad (6)$$

For a cubic mesh we use<sup>47</sup>

$$C(\omega) = -\frac{4\pi}{3} \frac{1}{\epsilon_{\text{env}}(\omega) v_i} \quad (7)$$

We invert equation (2) using lower–upper (‘LU’) decomposition, and the scattering cross-sections can be calculated from the near-field  $\mathbf{E}$  inside the particle<sup>49</sup>.

**Training ANNs for far-field scattering-based classification.** The ANNs were implemented in python using the TensorFlow package<sup>50</sup>.

**Digit retrieval using scattering spectra—1D ConvNet. Preprocessing of scattering spectra.** Prior to the training of the ANN, we preprocess the acquired scattering spectra. After background subtraction and normalization to the spectrum of the white lamp, we apply the following further processing steps on our data. We first apply a median filter with 6 nm kernel size to eliminate spikes from the spectra. Subsequently we apply a smoothing filter based on moving averages with a 15 nm kernel to reduce the noise. We finally apply a downsampling procedure using an order 8 Chebyshev type I filter by which we reduce the number of wavelengths to 99 per spectrum.

**Network architecture.** The scattering-spectra-based network for information retrieval is a 1D convolutional network followed by a fully connected part. The ConvNet consists of three layers using the ‘leaky ReLU’ activation function. The first layer with 64 filters per channel uses a kernel of size 7. The second and third layers both have 32 filters with kernels of size 5 and 3, respectively. Each 1D convolutional layer is followed by a max pooling layer with kernel size 2 as well as by a batch normalization<sup>51</sup>. The fully connected network consists of two layers with 64 and 32 neurons, employing a ‘tanh’ activation. For the ConvNet part, we apply an ‘L2’ regularization, and for the fully connected part a dropout regularization scheme with 80% keep probability. During training the output layer neuron whose index corresponds to the input binary number is set to ‘1’ while the other neuron activations are kept at ‘0’. We train the network as depicted in Fig. 1d using a variant of the stochastic gradient descent algorithm (‘Adam’, for details see ref. <sup>52</sup>) with a batch size of 64. We use the categorical cross-entropy loss, a learning rate of 0.0001 and train the network for 200 epochs. In the case of the 9-bit structures we train the network for 2,000 epochs.

**Discrete wavelengths and RGB digit retrieval—fully connected network. Discrete wavelengths—selection and preprocessing.** The intensity values at specific wavelengths are extracted from the scattering spectra acquired with white-light illumination. We average the intensity over a small window of three neighbouring values to reduce the impact of noise. The following spacings between wavelengths are used:

- First wavelength:  $\lambda_1 = 500 \text{ nm}$
- Two wavelengths:  $\lambda_2 = 630 \text{ nm}$
- Three wavelengths:  $\Delta\lambda = 95 \text{ nm}$
- Four wavelengths:  $\Delta\lambda = 75 \text{ nm}$
- Five wavelengths:  $\Delta\lambda = 60 \text{ nm}$

Because we found that the red part of the spectra contains very little information (see also Fig. 4a), we set the longest wavelength to be not larger than about 740 nm in order to ideally probe the regions of the spectra that contain the most information on the encoded bit sequence.

**RGB—preprocessing of dark-field images.** We automatically process the polarization-filtered dark-field microscopy images by considering all pixels with a brightness of at least three times the background level. Using this procedure, each structure results in a dot of 30–50 pixels in the dark-field images. Our data set is

composed of the average RGB values of the ensemble of pixels corresponding to each nanostructure. We normalize the RGB values to  $R + G + B = 1$  and separately store the scattered intensity (hence the brightness).

**Network architecture.** For the RGB data sets as well as for the readout using the scattering intensity at discrete wavelengths, we use a fully connected network. The scattering intensity at wavelengths  $\lambda_i$  or the R, G, B values (and optionally the intensity) are the inputs to the network, which itself consists of three layers with 'tanh' activation, in the following order: 128, 256 and 64 artificial neurons. We use L2 regularization and dropout with 80% keep probability on the entire ANN. The 16 neurons in the 'softmax' output layer represent the binary numbers encoded in the nanostructures. We train the network on data from 300 samples per geometry, using the original SDG algorithm and the categorical cross-entropy loss function. The batch size is 64. The learning rate is 1.0 with a  $\times 0.96$  decay each 1,000 steps (approximately every 13 epochs).

#### Considerations on the practical implementation of readout schemes

**and their performance.** *Blu-ray bit density for comparison with nanostructure size.* The Blu-ray has a length and width per digit of 150 nm and 130 nm, with 320 nm line spacing, leading to a required area per bit of  $A_{\text{bit}} = (320 + 130) \times 150 \text{ nm}^2 = 67,500 \text{ nm}^2$ . On the other hand, our '9-bit' structures with block edges of 105 nm and an outer wall width of 45 nm require  $(3 \times 105 + 45 + \text{spacing})^2 \text{ nm}^2$  per 9 bits. For a 300 nm spacing, this yields  $48,400 \text{ nm}^2$  per bit.

*Multiwavelength-based readout.* Using multiple lasers in a system similar to a Blu-ray reader, it would be possible to probe the scattering intensity at different wavelengths. As in the case of Blu-ray, the lasers could be scanning a rotating storage medium and acquire the spectral information in a sequential manner (for instance first a blue laser scans the structure, then a green and finally a red laser). The performance of such a system would be limited by the constraints of the available technology and should be comparable to the Blu-ray disk, except that 9 bits of information (instead of a single bit) could be read at every passage of the laser system.

*RGB-based readout.* With our experimental set-up, using a  $\times 50$ ,  $\text{NA} = 0.5$  objective we capture dark-field images covering areas of  $240 \times 180 \mu\text{m}^2$ . Assuming  $700 \times 700 \text{ nm}^2$  per digit including spacing, we could capture around 80,000 structures per image. In the case of five encoded bits per geometry, these nanostructures would encode approximately 0.44 Mbits of information. The Blu-ray disk at its  $1 \times$  data rate yields  $36 \text{ Mbits s}^{-1}$ , so around 100 images per second would be necessary to yield similar performance with our approach. Throughout this study we used an exposure time of  $t_{\text{exp}} = 0.05 \text{ s}$ , which, under otherwise perfect conditions, would correspond to 20 images per second. This is still about five times below Blu-ray performance.

Our set-up, however, is far from ideal for this specific readout task. Using high-NA and low-magnification objectives, the captured area could easily be enlarged significantly. For instance a  $\times 20$ ,  $\text{NA} = 0.75$  microscope objective would

cover a  $2.5 \times 2.5$  times larger area than our microscope, which would lead to 2.75 Mbits per image (again using the 5-bit structures). Also the exposure time could easily be strongly decreased using very bright light sources and optimized CCD cameras. Assuming  $t_{\text{exp}} = 0.01 \text{ s}$  (which is still a very conservative guess), this would then enable a readout rate of  $275 \text{ Mbit s}^{-1}$ .

#### Code availability

The authors declare that all software used to obtain the results of this work is publicly accessible as open-source software: [python](#) including [SciPy](#), [TensorFlow](#), as well as [pyGDM](#)<sup>46</sup>, our own implementation of the GDM. Our scripts can be made accessible from the corresponding author upon reasonable request.

#### Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request. The experimental and simulated scattering data sets are available under <https://doi.org/10.6084/m9.figshare.7326842.v1>.

#### References

43. Guerfi, Y., Carcenac, F. & Larrieu, G. High resolution HSQ nanopillar arrays with low energy electron beam lithography. *Microelectron. Eng.* **110**, 173–176 (2013).
44. Guerfi, Y., Doucet, J. B. & Larrieu, G. Thin-dielectric-layer engineering for 3D nanostructure integration using an innovative planarization approach. *Nanotechnology* **26**, 425302 (2015).
45. Martin, O. J. F., Girard, C. & Dereux, A. Generalized field propagator for electromagnetic scattering and light confinement. *Phys. Rev. Lett.* **74**, 526–529 (1995).
46. Wiecha, P. R. pyGDM—a python toolkit for full-field electro-dynamical simulations and evolutionary optimization of nanostructures. *Comput. Phys. Commun.* **233**, 167–192 (2018).
47. Girard, C., Dujardin, E., Baffou, G. & Quidant, R. Shaping and manipulation of light fields with bottom-up plasmonic structures. *New J. Phys.* **10**, 105016 (2008).
48. Edwards, D. F. In *Handbook of Optical Constants of Solids* (ed. Palik, E. D.) 547–569 (Academic, Burlington, 1997).
49. Draine, B. T. The discrete-dipole approximation and its application to interstellar graphite grains. *Astrophys. J.* **333**, 848–872 (1988).
50. Abadi, M. et al. *TensorFlow: Large-scale Machine Learning on Heterogeneous Distributed Systems*. <https://www.tensorflow.org/> (2015).
51. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Preprint at <https://arxiv.org/abs/1502.03167> (2015).
52. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2014).