

Pushing the limits of optical information storage using deep learning

Peter Wiecha, Aurélie Lecestre, Nicolas Mallet, Guilhem Larrieu

▶ **To cite this version:**

Peter Wiecha, Aurélie Lecestre, Nicolas Mallet, Guilhem Larrieu. Pushing the limits of optical information storage using deep learning. Nature Nanotechnology, 2019, 14 (3), pp.237-244. 10.1038/s41565-018-0346-1 . hal-02382045

HAL Id: hal-02382045
https://laas.hal.science/hal-02382045

Submitted on 19 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pushing the limits of optical information storage using deep learning

Peter R. Wiecha,[1,*] Aurélie Lecestre,[2] Nicolas Mallet,[2] and Guilhem Larrieu[2]

[1] *CEMES, Université de Toulouse, CNRS, Toulouse, France*
[2] *LAAS, Université de Toulouse, CNRS, INP, Toulouse, France*

Diffraction drastically limits the bit density in optical data storage. To increase the storage density, alternative strategies involving supplementary recording dimensions and robust read-out 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 analyzed by an artificial neural network, we achieve quasi error free read-out of sequences of up to 9 bit, 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 CMOS technology.

Optical information storage promises perennial longevity, high information densities and low energy consumption compared to magnetic storage media.[1,2] The compact disc, the DVD and the Blu-ray disc broadly established optical storage in our society.[3,4] 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,[5] near-field optical recording,[6] the use of fluorescent dyes[7] or three-dimensional approaches like two-photon point-excitation[8]. 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 read-out 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 with no commercial product available.[9,10]

Photonic nanostructures are powerful instruments to control light at the nanometer scale.[11,12] Localized surface plasmons (LSP) in metal nanoparticles[13] or Mie-type resonances in high-index dielectric structures[14] offer high scattering efficiencies, cover the entire visible spectrum and can be tuned by designing appropriate geometric features.[15,16] In consequence, the idea has been raised to encode information in the rich scattering spectra of plasmonic nanostructures, denser than a single bit.[17–21] The information density might be further increased by addressing layer-wise arranged nanostructures via the focal depth[22] or by the light polarization[23]. A key problem of such approach is the availability and accuracy of read-out schemes.[18,24] 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 the information retrieval.[18,25] 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 $\approx 520\,\mathrm{nm}$. Shorter wavelengths cannot be used to encode information, which effectively reduces the attainable information density.

To overcome these limitations, we develop a scheme for digital information encoding, based on silicon nanostructures. Owing to its high refractive index with low imaginary part, silicon nanostructures support low-loss optical resonances, tunable over the entire visible spectral range.[14,16,26] In addition, silicon has great technological advantages, first of all the mass-production ready, high-precision complementary metal-oxide-semiconductor (CMOS) based processing technology, its low cost and durability. To reliably retrieve the stored information, we propose a machine learning (ML) based approach, in which the scattering spectra are analyzed by an artificial neural network (ANN). ANNs are computational schemes that can be "trained" to efficiently solve problems, hard for classical computer arithmetics.[27,28] ANNs have many applications for instance in modern smartphones or in medical image interpretation.[29,30] Besides a few recent examples, ML is being scarcely applied on problems in nano-optics. In one work, robust data read-out from holographic memory was realized using convolutional ANNs.[31] The potential of ANNs has been demonstrated also in classification and inverse design of nanoparticles.[32–34] We train the digital-information retrieval neural network on the measured scattering spectra of several hundred fabricated copies of each nanostructure corresponding to a binary sequence. On all our experimental datasets encoding up to 9 bit of information, effectively going beyond the data density of the Blu-ray disc, the trained ANNs yield a quasi

* e-mail : peter.wiecha@cemes.fr

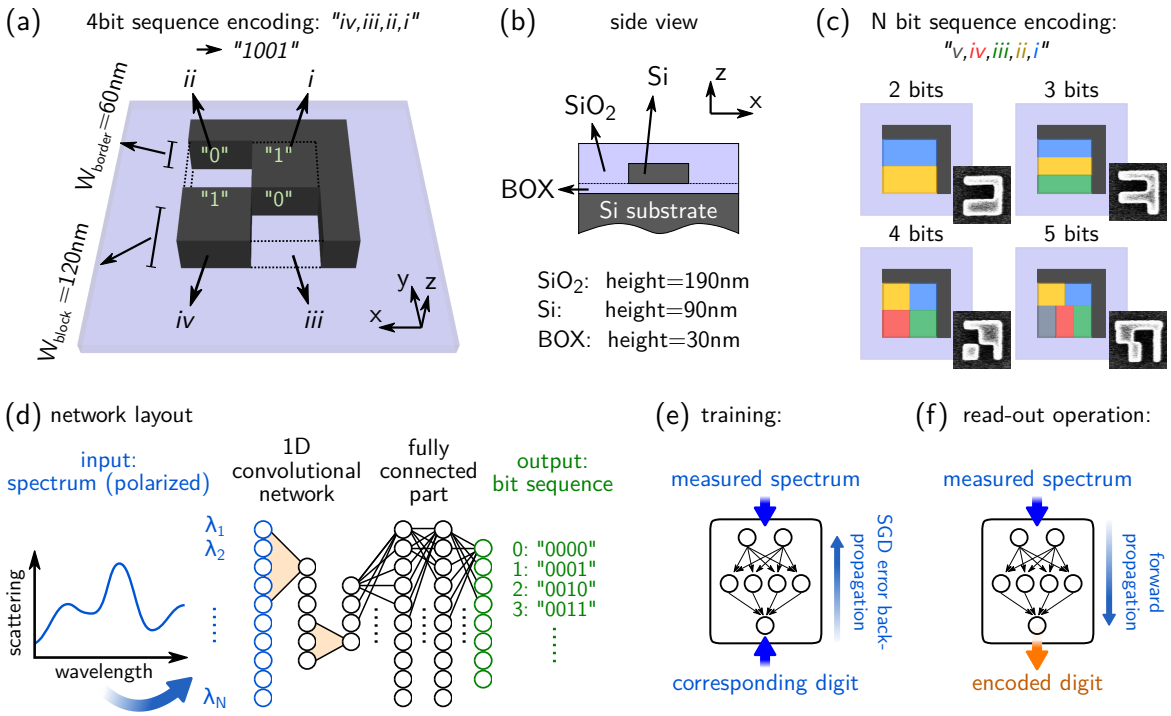


FIG. 1. **Sketch of the nanostructure geometry and the 1D convolutional artificial neural network.** (a) Illustration of the "4 bit" nanostructure geometry. The digital information is encoded in the four silicon blocks (block: "1", no block: "0"). The structure corresponds to the 4 bit digit "1001" (decimal "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 $120 \times 240\,\mathrm{nm}^2$ large. The 3 bit geometry uses blocks of $80 \times 240\,\mathrm{nm}^2$. For the 4 bit encoding, the blocks occupy areas of $120 \times 120\,\mathrm{nm}^2$. In the 5 bit geometry, the two upper blocks measure $150 \times 100\,\mathrm{nm}^2$ and the three lower ones have a size of $100 \times 150\,\mathrm{nm}^2$. Insets of selected SEM images show areas of $450 \times 450\,\mathrm{nm}^2$. (d) sketch of the 1D convolutional neural network used for the classification task. (e) training stage of the artificial neural network: measured spectra and corresponding digital information are fed into the network. The error is back-propagated using a variant of the stochastic gradient descent algorithm. (f) the trained network is capable to retrieve the digital information encoded in the structures via their spectra.

error-free read-out. We demonstrate furthermore that an 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) color values from darkfield microscopy images. The latter approach allows in principle a massively parallel read-out of the stored data.

I. SILICON NANOSTRUCTURES FOR DIGITAL INFORMATION ENCODING

As illustrated in figure 1a-b, we use a planar array-like geometry to encode several bits of information in a single silicon nanostructure. If a certain position in the 2D array contains a silicon block, the according bit is set to "1", otherwise it is "0". In order to distinguish symmetric or rotational arrangements (for example 4 bit "0010" and "0100"), an L-shaped silicon frame is added, surrounding two sides of the structure. In this way, under linearly polarized illumination each binary number yields a unique spectral response.

We fabricate nanostructures encoding between 2 and 5 bits of information each, as illustrated in figure 1c, using electron-beam lithography (ebeam) and subsequent dry-etching of commercial silicon-on-insulator substrates with a silicon overlayer of $H = 90\,\mathrm{nm}$ height. Subsequent to the etching, the structures are covered by a protective SiO_2 layer of 190 nm height (see figure 1b). For more details on the fabrication process, see the Methods section.

We fabricate $25 \times 25 = 625$ copies of each geometry. Using an automated setup with an XY piezo stage, we measure the linearly X and Y polarized darkfield 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 figure 2, superposed and grouped according to the 16 digital numbers. The spectra for the 2, 3 and 5 bit datasets can be found in the supporting informations (SI), Figs. S4-S11. We note that the spectra of symmetric structures

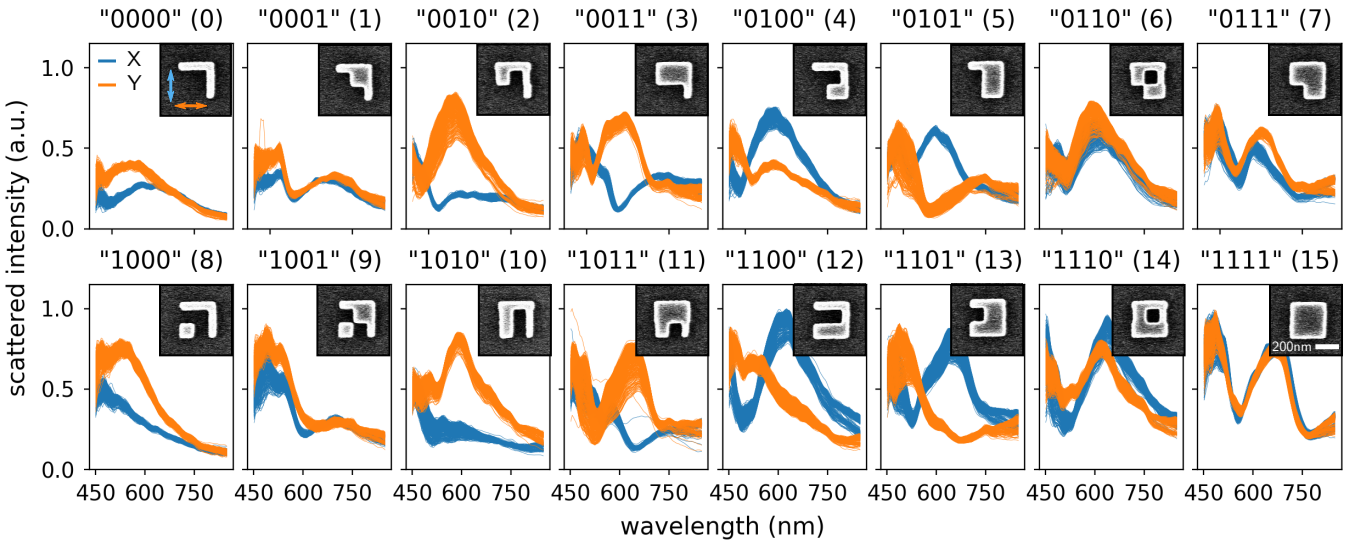


FIG. 2. **Experimental darkfield spectra training data-set for 4 bits.** Our data comprises measurements from 625 copies for each of the 16 "4 bit" geometries (this makes a total of $625 \times 16 \times 2 = 20000$ acquired spectra). The spectra are superposed above each other. Blue lines: darkfield scattering for X-polarized light, orange lines: Y-polarization. Insets show SEM images of one representative copy of the respective nanostructure, the areas are $600 \times 600\,\mathrm{nm}^2$ large, the scalebar in "1111" is 200 nm. The not shown 2 bit dataset consists of 5000, the case of 3 bits comprises 10000 and the 5 bit set contains 40000 scattering spectra (all shown together with SEM images in the supporting informations, (Figs. S4-S11), where also a separate 4 bit dataset and simulations are shown in order to demonstrate the reproducibility (Figs. S12-S14).

are not completely identical for crossed polarizations (see *e.g.* "0000"). We attribute this observation to the ebeam being horizontally rasterscanned in the fabrication process, leading to small anisotropies (about $5-10\%$) between horizontal and vertical features (see SEM images in SI Figs. S1-S3 and S21). This can in principal be corrected in an automated fashion during the mask design, as demonstrated in the SI, Figs. S27-S28. On the other hand, such asymmetries can be even advantageous for our purpose, because they increase the "uniqueness" of the individual geometries and their scattering spectra.

II. MACHINE LEARNING BASED DIGITAL INFORMATION READ-OUT

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 sub-sets of the acquired spectra. Subsequently, we evaluate the accuracy of the read-out by testing the trained ANN with spectra not used for training.

We use a one-dimensional convolutional neural network (ConvNet), followed by a fully connected network, as depicted in figure 1d – an architecture with particular strength at pattern-recognition tasks.[28] The spectra are fed in 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 on the preprocessing of the data can be found in the Methods section. The network loss and the error rate on the validation set are shown in figure 3a for the first 100 epochs[36] of training on the 4 bit dataset.

The read-out 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 read-out accuracy. In the cases of 4 and 5 bits, one single spectrum of the test-data was incorrectly interpreted (corresponding to 0.023% and 0.011% for the 4400 (4 bit), respectively 8800 (5 bit) test-structures).

We analyzed the datasets using the "t-SNE" dimensionality reduction, in order to estimate the distinguishability between different geometries, as well as the variance in copies of identical structures.[35] In a t-SNE plot, well separated scatter points correspond to unambiguously differentiable entities in the dataset, while nearby and overlapping points correspond to very similar data. The results for the "XY" datasets are shown in figure 3b, the t-SNE plots for the "X" and "Y" datasets can be found in the SI, Fig. S15. Each color corresponds to one type of structure,

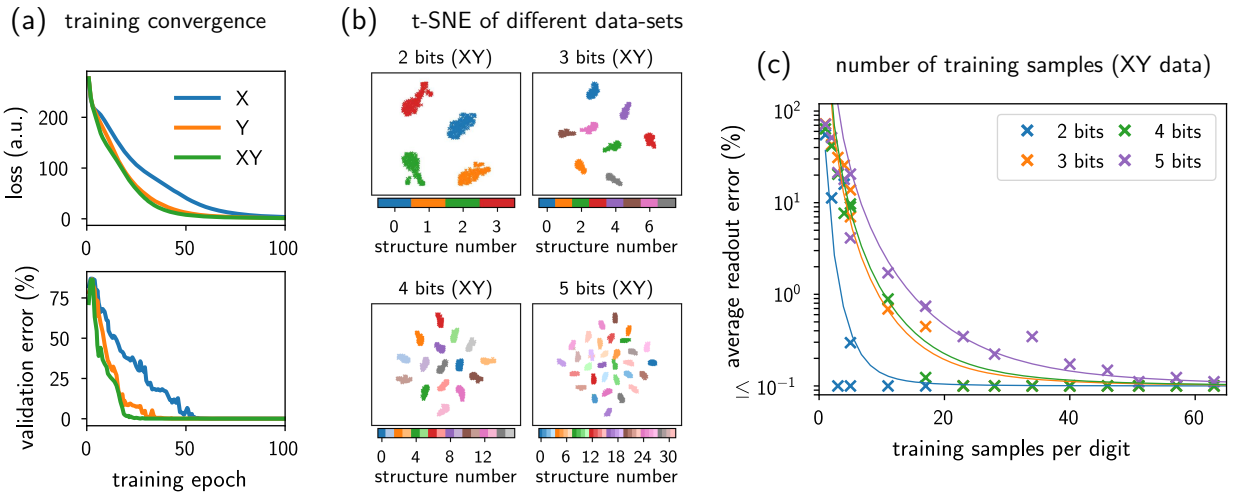


FIG. 3. **Training convergence and read-out accuracy of the artificial neural network trained on the full scattering spectra.** (a) convergence of the training at the example of 4 bit encoding and spectra of 300 copies per digit for training. Loss (categorical cross-entropy, top) and validation error rate (bottom) as function of the training epoch for only X (blue), only Y polarized spectra (orange) and both polarizations (green) as training data. (b) t-SNE[35] visualization (see also text) of the "XY" polarization training sets for 2-5 bits encoded per structure. (c) read-out error as function of the number of spectra per structure used for the training of the network. All cases of information density per structure yield quasi error-free read-out 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 less training samples for an accurate information retrieval. Solid lines are guides to the eye, proportional to N^{-3}, which qualitatively describe the convergence of read-out as function of training set size N.

each dot represents a single measurement. All datasets 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 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 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" datasets). We evaluated three different positions of the spectral window: Either at the long wavelength end of the spectra ($\lambda \leq 850\,\mathrm{nm}$, orange lines), at short wavelengths ($\lambda \geq 450\,\mathrm{nm}$, blue lines), or centered around 650 nm (green lines). While the short and intermediate wavelengths always yield error free readout for spectral windows as small as $\lesssim 100\,\mathrm{nm}$, we observe that using the red part of the spectra requires a larger spectral window of up to $\approx 200\,\mathrm{nm}$ 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 $\approx 750\,\mathrm{nm}$. 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 also Methods) using the scattering intensity only at a low number of discrete wavelengths, as shown in figure 4b. Probing three wavelengths is always sufficient to obtain a readout accuracy of $> 99\,\%$. The error rates on the worst bit-sequence in the order of $10\,\%$ can be reduced below $0.5\,\%$ by probing at 4 or 5 wavelengths.

We will study in the following if the bit sequences can be recovered also using a simpler and faster data acquisition scheme, namely the scattered RGB color obtained from conventional darkfield microscopy images. Figure 5a shows darkfield images of 3×3 copies for each geometry and both polarizations at the example of the 4 bit geometry. The average RGB color 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 darkfield image. We normalize the RGB values to $R + G + B = 1$ and store the scattered intensity separately. We create four training datasets: Three sets with only the normalized RGB information. One for X, one for Y polarized data and a

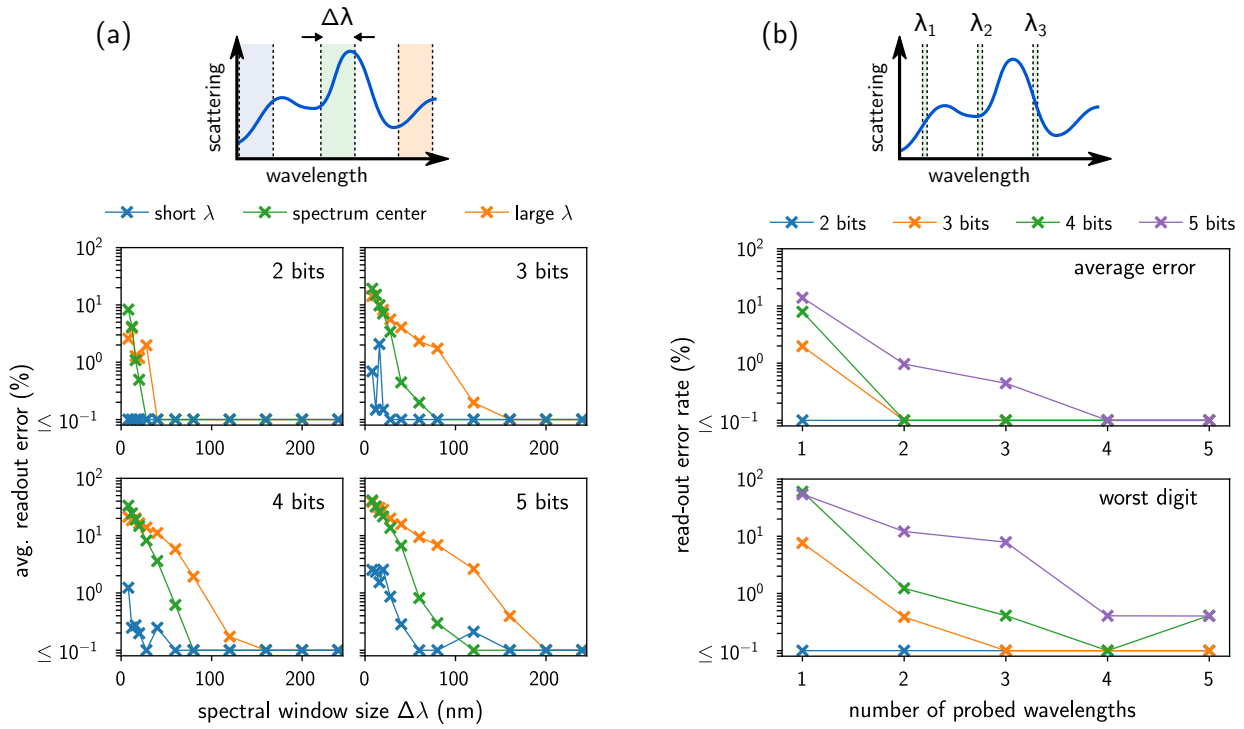


FIG. 4. **Accuracy of network trained on reduced spectral information.** (a) Training using scattering from spectral window of reduced width for the 2, 3, 4 and 5 bit datasets (left top to right bottom plot). The scattering intensity was taken from a window either at the short wavelength side (blue lines), in the center (green lines) or from the red edge of the spectra. (b) Average (top plot) and worst-digit (bottom pot) read-out 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 λ_i for the different cases are given in the Methods section. t-SNE plots for all reduced spectral information datasets can be found in the SI, Figs. S15.

third dataset combining both polarizations ("XY"). The forth 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 given in the Methods section.

Figures 5c and d respectively show "t-SNE" plots for the different 3 bit datasets and for the "XY" data of the 2, 3, 4 and 5 bit geometries. The partial mixing of scatterpoints 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, the error rates of several individual binary numbers are even well above 50% (see Figure 5e and SI, Fig. S16).

Using the datasets combining both polarizations, the situation drastically improves. 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 in a resembling tone and brightness (see also SI Figs S16). The similarity can also be observed in the t-SNE plot (figure 5d), where the light green and pink dots (structures "5" and "12") are partly touching. Such an analysis of the t-SNE plots allows to identify the most problematic digits (e.g. "6", "8", "29" and "31" in the 5 bit case). By designing nanostructures with a more significant difference in the scattered colors (e.g. using modern inverse problem techniques[16,37]), this limitation could be easily overcome and the error rate further decreased.

The RGB color 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 read-out of the information (see also Methods). Cheap, smartphone-based darkfield microscopy for the RGB read-out might become feasible in the near future. Indeed, smartphone-based microscopy is subject of current research and has undergone tremendous progress in the recent past.[38–40] A further interesting route for improvement are specifically designed bright-field color scatterers[41] for information encoding, in order to avoid the necessity of a darkfield illumination scheme. Finally, a cheap multi-laser approach, similar to a Blu-ray disc reader with several lasers could be used to capture the scattering intensity simultaneously at several wavelengths.

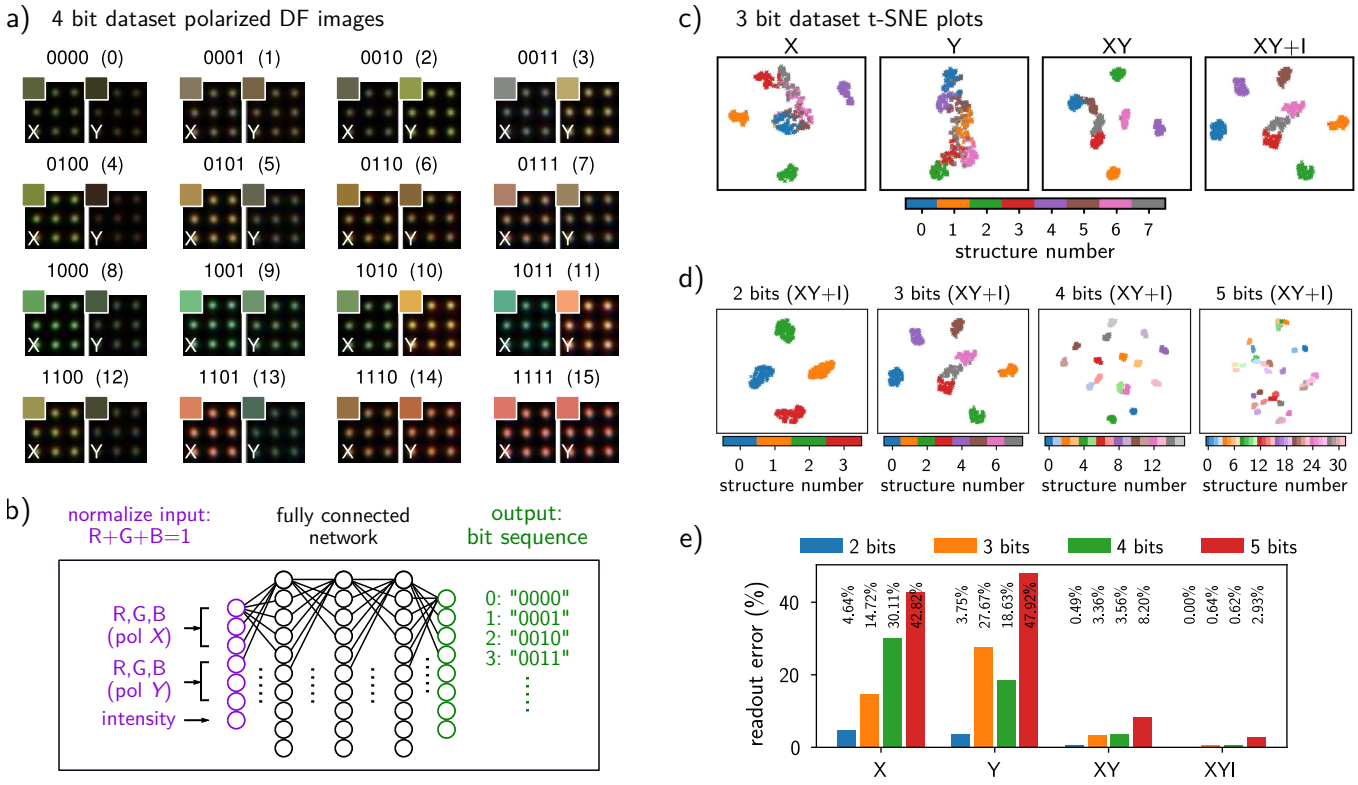


FIG. 5. **Neural-network based data read-out via the RGB color values.** (a) polarization filtered dark-field color images of representative 3×3 arrays of the "4 bit" digit structures. Left: X-polarization, right: Y-polarization. The darkfield images show areas of $7.5 \times 7.5\,\mu m^2$. The insets show the average RGB color of the 3×3 structures. (b) scheme of the fully connected artificial neural network used for the RGB classification task. (c) t-SNE[35] visualization of the "3 bit" training sets for only X, only Y, X plus Y or X, Y and scattered intensity I (from left to right). Only if using simultaneously both polarizations ("XY" and "XY+I"), 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 + the scattered intensities I. t-SNE plots for all data-sets can be found in the SI, Fig. S15. In Figs. S17-S18 of the SI, the activations of the softmax output neurons of the "RGB" network are shown for the whole validation sets of the 4 bit datasets "X" and "XY+I".

III. TOWARDS HIGHER INFORMATION DENSITY

Finally, we want to assess if the information density of the individual nanostructures can be further increased. As detailed in figure 6a, we therefore 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 $\approx 20\%$ larger size of the 9 bit structures compared to the 2-5 bit geometries leads to a red-shift 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 of a large dataset. 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. But there are no general technical constraints and using dedicated, fully automated setups would enable a rapid experimental acquisition of very large datasets. In our case, in order to obtain a sufficiently large dataset, we numerically expand a representative set of experimental spectra from 4 copies of each nanostructure by generating 300 random superpositions of these 4 spectra. We additionally introduce a random intensity deviation of up to $\pm 10\%$. The result of this numerical data expansion is shown at selected examples in figure 6b (see SI, Figs. S26 and S29-S32 for the whole experimental dataset 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 simultaneously X and Y polarization). The techniques are identical to the above cases of 2-5 bits per nanostructure (see also Methods). In the left of figure 6c, the readout accuracy using full spectra is shown as a function of training samples per geometry.

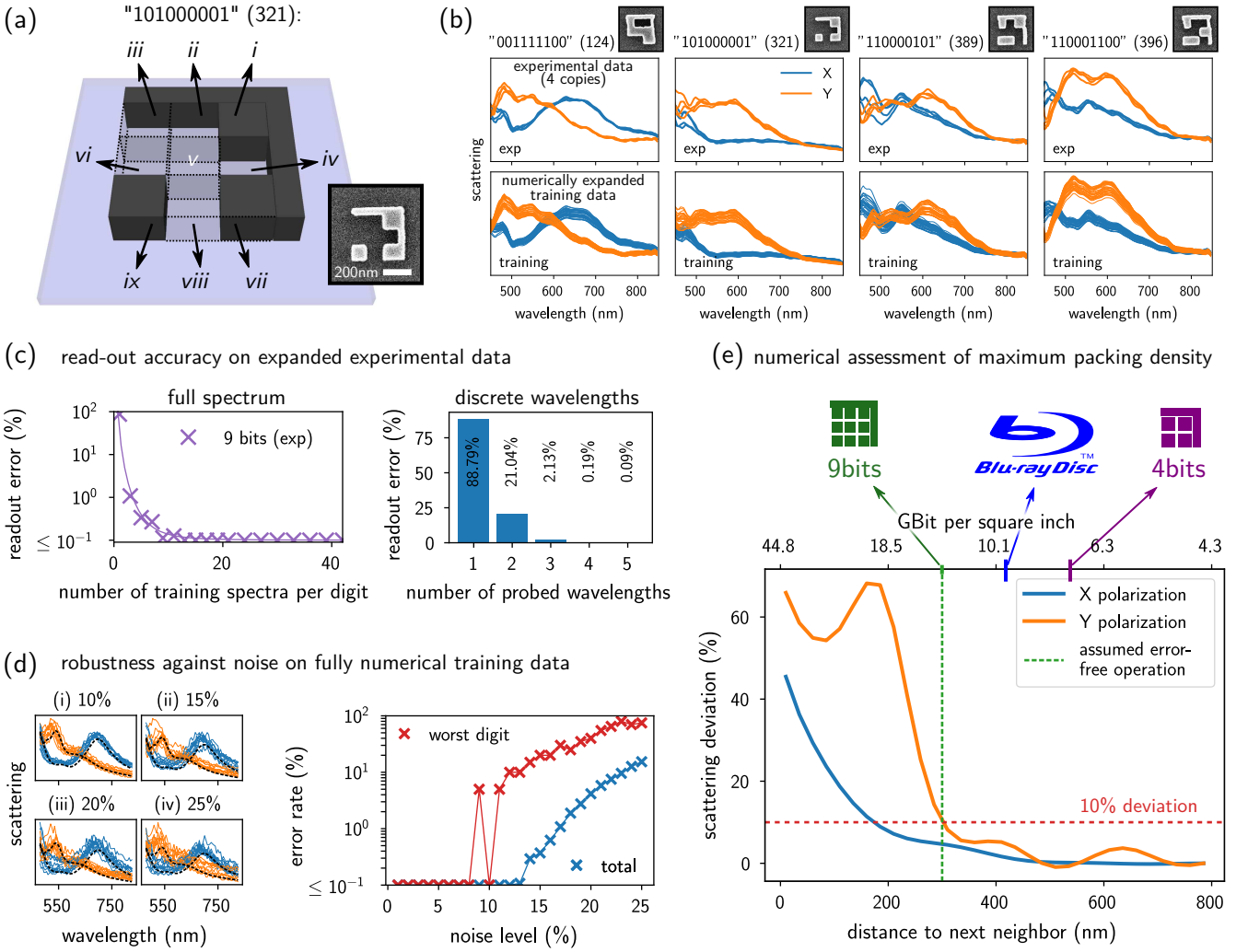


FIG. 6. **9 bit per nanostructure information encoding.** (a) geometry of a silicon nanostructure encoding $3 \times 3 = 9$ bits (512 possible combinations). Each silicon block occupies an area of $105 \times 105\,\mathrm{nm}^2$. The L-shaped sidewall is 45 nm wide. The height is 90 nm. An SEM image of a fabricated structure is given in the inset, where the scalebar is 200 nm. The shown example represents the decimal number "321". (b) selected examples, illustrating the training data generation by a numerical expansion of the experimental spectra. SEM images show areas of $550 \times 550\,\mathrm{nm}^2$. We fabricated 4 copies of each possible 9 bit nanostructure. Via random superposition of the experimental scattering spectra of these 4 copies, we generate a large set of spectra, allowing us to train and test the performance of the binary information readout ANN (see also SI Fig. S23 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 is used simultaneously. (d) evaluation of the robustness of 9 bit read-out 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 SI Fig. S33 for details on the training data generation). Example spectra for noise levels from (i) 10% to (iv) 25% are shown at the example of structure "001011010" (decimal 90). The plot on the right shows the readout error rate on the numerical data as 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 the estimation of the feasible information density, 10% deviation from the unperturbed spectrum are assumed to be tolerable (red dashed horizontal line). This leads to an information density (green indicator) around 40% higher than the Blu-ray disc (blue indicator). The 4 bit structures yield about 75% of the Blu-ray density.

Quasi error-free operation is obtained for training-sets as small as about 20 spectra per digit. Using a discrete number of probed wavelengths (Fig. 6c, right), very low error rates of $\approx 2\%$ can be obtained by probing only three discrete wavelengths. Using 4 or 5 probed wavelengths improves the error far below 1%.

In order to assess how larger variations and noise influence the readout performance, we use a dataset of fully

numerical simulations of the 512 geometries (see Methods). This synthetic data allows to adjust the distortions and noise of the spectra in a quantifiable way, which is shown schematically in the left of figure 6d. The read-out 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 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 the SI Figs. S34-S35. A comparison of the experimental data with simulations of different noise levels, a t-SNE plot of the experimental 9 bit dataset and further analysis on the impact of noise can be found in SI Fig. S22 and S24-S26.

We finally compare the information density of the 9 bit nanostructures to the state-of-the art flat optical storage medium, the "Blu-ray disc". The Blu-ray requires $67500\,\mathrm{nm}^2$ per bit, while our approach scales with the structure-spacing and uses $(3\cdot 105+45+\mathrm{spacing})^2/9\,\mathrm{nm}^2$ (see Methods, section V F). In figure 6e we analyze the influence of a close neighbor 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 read-out, yielding around 40% higher information density than the Blu-ray (see also SI Fig. S36-S37). 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 red-shift 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 (2048, respectively 4096 structure geometries). Larger binary sequences seem however difficult for an accurate read-out, due to the power-scaling by which the number of structure geometries increases ($2^{N_{\mathrm{bit}}}$). A further means to improve the read-out accuracy and its robustness might be to include the angle of incidence as probing parameter. In order to exploit all three spatial dimensions of the storage medium, multi-layer arrangements might be possible.[22,23]

IV. CONCLUSIONS

In summary, we 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 demonstrated on geometries encoding up to 9 bits per diffraction limited area, that the optical scattering spectra are more than sufficient for an accurate recovery of the encoded data. We showed that probing at a few discrete wavelengths or even simply using the RGB color information obtained from standard darkfield microscopy images, is a precise read-out scheme, potentially possible on very simple and cheap equipment. The latter approach would allow a massively parallel retrieval of the information. The robustness can easily be improved by tailoring nanostructure geometries with high color-contrast.[42]

Our work paves the way towards high density planar optical information storage using simple far-field characterization combined with concepts of machine learning. Re-writeable storage media might be created around the recently developed technology on catalytic magnesium metasurfaces for dynamic structural color.[43] Our approach can be easily generalized to other classification tasks in nano-optics, including biological applications and problems in plasmonics.

ACKNOWLEDGMENTS

We gratefully thank Arnaud Arbouet and Christian Girard for their inspiring advise, their valuable help and for discussing and proof-reading the manuscript, and Franck Carcenac for his help on electron beam lithography 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 center 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. and N.M. fabricated the nanostructures and performed the electron microscopy. P.R.W. carried out the optical experiments, did the simulations, the data analysis and implemented the machine learning part. P.R.W. wrote the manuscript with contributions from G.L. All authors discussed the results and commented on the manuscript at every stage.

ADDITIONAL INFORMATION

Supplementary information is available in the online version of the paper. Reprints and permission information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.R.W.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

[1] Zhang, J., Gecevičius, M., Beresna, M. & Kazansky, P. G. Seemingly Unlimited Lifetime Data Storage in Nanostructured Glass. *Physical Review Letters* **112**, 033901 (2014).

[2] Gu, M., Li, X. & Cao, Y. Optical storage arrays: A perspective for future big data storage. *Light: Science & Applications* **3**, e177 (2014).

[3] Satoh, I., Ohara, S., Akahira, N. & Takenaga, M. Key technology for high density rewritable DVD (DVD-RAM). *IEEE Transactions on Magnetics* **34**, 337–342 (1998).

[4] Borg, H. J. *et al.* Phase-Change Media for High-Numerical-Aperture and Blue-Wavelength Recording. *Japanese Journal of Applied Physics* **40**, 1592 (2001).

[5] Zeng, B. J., Ni, R. W., Huang, J. Z., Li, Z. & Miao, X. S. Polarization-based multiple-bit optical data storage. *Journal of Optics* **16**, 125402 (2014).

[6] Tominaga, J., Nakano, T. & Atoda, N. An approach for recording and readout beyond the diffraction limit with an Sb thin film. *Applied Physics Letters* **73**, 2078–2080 (1998).

[7] Mottaghi, M. D. & Dwyer, C. Thousand-Fold Increase in Optical Storage Density by Polychromatic Address Multiplexing on Self-Assembled DNA Nanostructures. *Advanced Materials* **25**, 3593–3598 (2013).

[8] Strickler, J. H. & Webb, W. W. Three-dimensional optical data storage in refractive media by two-photon point excitation. *Optics Letters* **16**, 1780–1782 (1991).

[9] van Heerden, P. J. Theory of Optical Information Storage in Solids. *Applied Optics* **2**, 393–400 (1963).

[10] Psaltis, D. & Burr, G. W. Holographic data storage. *Computer* **31**, 52–60 (1998).

[11] Girard, C. Near fields in nanostructures. *Reports on Progress in Physics* **68**, 1883–1933 (2005).

[12] Novotny, L. & Hecht, B. *Principles of Nano-Optics* (Cambridge University Press, Cambridge ; New York, 2006).

[13] Maier, S. *Plasmonics: Fundamentals and Applications* (Springer US, 2010).

[14] Kuznetsov, A. I., Miroshnichenko, A. E., Brongersma, M. L., Kivshar, Y. S. & Luk'yanchuk, B. Optically resonant dielectric nanostructures. *Science* **354** (2016).

[15] Cao, L., Fan, P., Barnard, E. S., Brown, A. M. & Brongersma, M. L. Tuning the Color of Silicon Nanostructures. *Nano Letters* **10**, 2649–2654 (2010).

[16] Wiecha, P. R. *et al.* Evolutionary multi-objective optimization of colour pixels based on dielectric nanoantennas. *Nature Nanotechnology* **12**, 163–169 (2017).

[17] Mansuripur, M. *et al.* Plasmonic nano-structures for optical data storage. *Optics Express* **17**, 14001–14014 (2009).

[18] Chen, W. T. *et al.* Manipulation of multidimensional plasmonic spectra for information storage. *Applied Physics Letters* **98**, 171106 (2011).

[19] 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 Photonics* **1**, 631–637 (2014).

[20] El-Rabiaey, M. A., Areed, N. F. F. & Obayya, S. S. A. Novel Plasmonic Data Storage Based on Nematic Liquid Crystal Layers. *Journal of Lightwave Technology* **34**, 3726–3732 (2016).

[21] Zijlstra, P., Chon, J. W. M. & Gu, M. Five-dimensional optical recording mediated by surface plasmons in gold nanorods. *Nature* **459**, 410–413 (2009).

[22] 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. *Optics Express* **20**, 5069–5081 (2012).

[23] Taylor, A. B., Michaux, P., Mohsin, A. S. M. & Chon, J. W. M. Electron-beam lithography of plasmonic nanorod arrays for multilayered optical storage. *Optics Express* **22**, 13234–13243 (2014).

[24] 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).

[25] Liu, D., Tan, Y., Khoram, E. & Yu, Z. Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures. *ACS Photonics* **5**, 1365–1369 (2018).

[26] Albella, P. *et al.* Low-Loss Electric and Magnetic Field-Enhanced Spectroscopy with Subwavelength Silicon Dimers. *The Journal of Physical Chemistry C* **117**, 13573–13584 (2013).

[27] Nielsen, M. A. *Neural Networks and Deep Learning* (Determination Press, 2015).

[28] Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).

[29] Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv:1602.07261 [cs]* (2016). 1602.07261.

[30] Mamoshina, P., Vieira, A., Putin, E. & Zhavoronkov, A. Applications of Deep Learning in Biomedicine. *Molecular Pharmaceutics* **13**, 1445–1454 (2016).

[31] Shimobaba, T. *et al.* Convolutional neural network-based data page classification for holographic memory. *Applied Optics* **56**, 7327–7330 (2017).

[32] Jo, Y. *et al.* Holographic deep learning for rapid optical screening of anthrax spores. *Science Advances* **3**, e1700606 (2017).

[33] Malkiel, I. *et al.* Plasmonic nanostructure design and characterization via Deep Learning. *Light: Science & Applications* **7**, 60 (2018).

[34] Peurifoy, J. *et al.* Nanophotonic particle simulation and inverse design using artificial neural networks. *Science Advances* **4**, eaar4206 (2018).

[35] van der Maaten, L. & Hinton, G. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

[36] An "epoch" is a ML term, signifying one full training iteration. In each epoch, the full training dataset is randomly shuffled and used in its totality to optimize the network parameters.

[37] Feichtner, T., Selig, O., Kiunke, M. & Hecht, B. Evolutionary Optimization of Optical Antennas. *Physical Review Letters* **109**, 127701 (2012).

[38] Rivenson, Y. *et al.* Deep Learning Enhanced Mobile-Phone Microscopy. *ACS Photonics* (2018).
[39] 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. *Scientific Reports* **8**, 3298 (2018).
[40] Wei, Q. *et al.* Plasmonics Enhanced Smartphone Fluorescence Microscopy. *Scientific Reports* **7**, 2124 (2017).
[41] Flauraud, V., Reyes, M., Paniagua-Domínguez, R., Kuznetsov, A. I. & Brugger, J. Silicon Nanostructures for Bright Field Full Color Prints. *ACS Photonics* (2017).
[42] González-Alcalde, A. K. *et al.* Optimization of all-dielectric structures for color generation. *Applied Optics* **57**, 3959–3967 (2018).
[43] Duan, X., Kamin, S. & Liu, N. Dynamic plasmonic colour display. *Nature Communications* **8**, 14606 (2017).
[44] Guerfi, Y., Carcenac, F. & Larrieu, G. High resolution HSQ nanopillar arrays with low energy electron beam lithography. *Microelectronic Engineering* **110**, 173–176 (2013).
[45] Guerfi, Y., Doucet, J. B. & Larrieu, G. Thin-dielectric-layer engineering for 3D nanostructure integration using an innovative planarization approach. *Nanotechnology* **26**, 425302 (2015).
[46] Martin, O. J. F., Girard, C. & Dereux, A. Generalized Field Propagator for Electromagnetic Scattering and Light Confinement. *Physical Review Letters* **74**, 526–529 (1995).
[47] Wiecha, P. R. pyGDM—A python toolkit for full-field electro-dynamical simulations and evolutionary optimization of nanostructures. *Computer Physics Communications* **233**, 167–192 (2018).
[48] Girard, C., Dujardin, E., Baffou, G. & Quidant, R. Shaping and manipulation of light fields with bottom-up plasmonic structures. *New Journal of Physics* **10**, 105016 (2008).
[49] Edwards, D. F. Silicon (Si)*. In Palik, E. D. (ed.) *Handbook of Optical Constants of Solids*, 547 – 569 (Academic Press, Burlington, 1997).
[50] Draine, B. T. The Discrete-Dipole Approximation and its Application to Interstellar Graphite Grains. *Astrophysical Journal* **333**, 848–872 (1988).
[51] Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. *https://www.tensorflow.org/* (2015).
[52] Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167 [cs]* (2015). 1502.03167.
[53] Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]* (2014). 1412.6980.

V. METHODS

A. Nanofabrication of planar Si structures

The silicon nanostructures were patterned on silicon on insulator (SOI) substrate (90 nm active Si layer on 30 nm buried oxide) following a large scale top-down approach. Electron beam lithography was used to pattern 80 nm thick layer of an inorganic negative-tone resist, namely hydrogen silsesquioxane (HSQ). After exposure, HSQ was developed by immersion in 25% tetramethylammonium hydroxide (TMAH) for one minute.[44] HSQ patterns were subsequently transferred to the silicon substrate down to the buried oxide by anisotropic reactive ion etching (RIE). Then, the structures were embedded in a 200 nm thick HSQ layer, deposited by spin coating in order to perfectly planarize the sample in a nanometrical range.[45] Finally, the HSQ layer (SiO_xH_y) was converted into a SiO_x layer by rapid thermal annealing at 600°C / 2 min under nitrogen ambiance, leading to a final SiO_x thickness layer of 190 nm.

B. Confocal darkfield microscopy

The scattering spectra were acquired out on a confocal dark-field microscope (Horiba XploRA). A white lamp was focused on the sample by a ×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 µm) and a polarization filter and dispersed by a grating (300 grooves mm^{-1}) on a Peltier-cooled 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 dataset was measured with $t_a = 0.5$ s, leading to similar results (see SI, Figs. S12-S14). We therefore assume that even shorter acquisition times would be sufficient for a robust data recognition.

The polarization filtered dark-field images were taken using the same ×50 darkfield microscope objective, with a color CCD camera at a resolution of 1392×1040 and an exposure time of $t_{\text{exp}} = 0.05$ s. Each structure is perceived as a colored dot on the microscopy image, covering $\approx 20 - 50$ pixels, from which we take the average RGB value.

C. Expansion of 9 bit experimental training data set

Due to technical constraints of our measurement setup, we are not able to acquire in reasonable time, scattering spectra of several hundreds of all 512 possible 9 bit encoding nanostructures. In order to nevertheless assess the feasibility of experimental 9 bit readout via our machine learning based approach, we fabricate 4 copies of the 9 bit structures. To obtain the required much larger amount of spectra of each geometry, we subsequently create new spectra $\sigma_{\text{new}}(\lambda)$ from random superpositions of the 4 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 up to one, hence

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

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 artificial neural networks for the decoding of the digital information.

D. Electrodynamical simulations

Simulated dataset

The 9 bit geometry model used for the fully numerically simulated dataset is identical to the experimental structures (see figure 6a). It consists of 3×3 silicon blocks of each $105 \times 105\,\mathrm{nm}^2$ 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). In order to assess the robustness against different amount 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 the supporting informations, figure S33. 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. In order to do so, in the first place we simulated the spectra on an extended wavelength range. Thanks to this procedure, we are capable of adjusting 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 Green Dyadic Method (GDM), a frequency-domain approach based on the concept of a generalized propagator.[46] In particular, we use an own implementation in python, "pyGDM".[47]

In the GDM the volume of a nanostructure is discretized with N cubic meshpoints of edge length d. To each of these mesh-points, a dipolar response is attributed. As detailed e.g. in reference[48], this leads to a system of $3N$ coupled equations

$$\mathbf{E}_0 = \mathbf{M} \cdot \mathbf{E}, \tag{2}$$

which, by inversion of $\mathbf{M}$, allows to relate any incident electric field $\mathbf{E}_0$ to the induced field $\mathbf{E}$ inside the particle. $\mathbf{M}$ is composed of 3×3 sub-matrices

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

$\mathbf{I}$ is the Cartesian unitary tensor, δ_{ij} the Kronecker delta function and, in cgs units,

$$\alpha_i(\omega) = \frac{\epsilon_i(\omega) - \epsilon_{\mathrm{env}}(\omega)}{4\pi} v_i\,. \tag{4}$$

In the latter equation, v_i is the volume of the cubic discretization cells hence $v_i = d^3$. We use the dispersion of silicon from Palik[49] for ϵ_i. The structures are placed in a homogeneous environment ($\epsilon_{\mathrm{env}} = 1.45$ for SiO_2) at 30 nm above a silicon substrate.

$\mathbf{G}$ in Eq. (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):

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

which can be found in literature.[48] At $\mathbf{r}_i = \mathbf{r}_j$ the Green's Dyad $\mathbf{G}_0$ diverges, hence a normalization scheme is applied:

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

For a cubic mesh we use[48]

$$C(\omega) = -\frac{4\pi}{3}\frac{1}{\epsilon_{\mathrm{env}}(\omega) v_i}\,. \tag{7}$$

We invert equation (2) using LU-decomposition, the scattering cross-sections can be calculated from the near-field $\mathbf{E}$ inside the particle.[50]

E. Training artificial neural networks for far-field scattering based classification

The artificial neural networks are implemented in python using the tensorflow package.[51]

Digit retrieval using scattering spectra – 1D ConvNet

Preprocessing of scattering spectra Prior to the training of the ANN, we pre-process 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 one-dimensional 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.[52] 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, 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. 53) 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 2000 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 3 neighboring values to reduce the impact of noise. Following spacings between wavelengths are used

- first wavelength: $\lambda_1 = 500\,\mathrm{nm}$
- 2 wavelengths: $\lambda_2 = 630\,\mathrm{nm}$
- 3 wavelengths: $\Delta\lambda = 95\,\mathrm{nm}$
- 4 wavelengths: $\Delta\lambda = 75\,\mathrm{nm}$
- 5 wavelengths: $\Delta\lambda = 60\,\mathrm{nm}$

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

RGB: Preprocessing of darkfield images We automatically process the polarization filtered darkfield microscopy images by considering all pixels with a brightness of at least 3× the background level. Using this procedure, each structure results in a dot of 30-50 pixels in the darkfield images. Our dataset 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 datasets as well as for the read-out using the scattering intensity at discrete wavelengths, we use a fully connected network. The scattering intensity at the wavelengths λ_i or the R, G, B values (and optionally the intensity) are the input to the network, which itself consist 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 1000 steps ($\approx$ every 13 epochs).

F. Considerations on the practical implementation of readout schemes and their performance

Blu-ray bit density for comparison with the nanostructure size

The Blu-ray has a length and width per digit of 150 nm, respectively 130 nm, with 320 nm line spacing, leading to a required area per bit of $A_{1\mathrm{bit}} = (320+130)\cdot 150\,\mathrm{nm}^2 = 67500\,\mathrm{nm}^2$. On the other hand, our "9 bit" structures with block edges of 105 nm and an outer wall width of 45 nm require $(3\cdot 105+45+\mathrm{spacing})^2\,\mathrm{nm}^2$ per 9 bits. For 300 nm spacing, this yields $48400\,\mathrm{nm}^2$ per bit.

Multi-wavelength 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 at last a red laser). The performance of such a system would be basically limited by the constraints of the available technology and should be comparable to the Blu-ray disc, except that 9 bit of information (instead of a single bit) could be read at every passage of the laser system.

RGB based readout

With our experimental setup, using a ×50, NA 0.5 objective, we capture darkfield images covering areas of $240 \times 180\,\mu\mathrm{m}^2$. Assuming $700 \times 700\,\mathrm{nm}^2$ per digit including spacing, we could capture around 80,000 structures per image. In the case of 5 encoded bit per geometry, these nanostructures would encode approximately 0.44 Mbit of information. The Blu-ray disc at its 1× datarate yields 36 Mbit/s, 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 $t_{\mathrm{exp}} = 0.05\,\mathrm{s}$, which, under otherwise perfect conditions, would correspond to 20 images per second. This is still about 5 times below Blu-ray performance.

Our setup however is far from ideal for this specific readout task. Using high-NA and low magnification objectives, the captured area could easily be significantly enlarged. For instance a $\times 20$, NA 0.75 microscope objective would cover a 2.5×2.5 times larger area than our microscope, which would lead to 2.75 Mbit 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 $t_{\mathrm{exp}} = 0.01$s (which is still a very conservative guess), this would then enable a readout rate of 275MBit/s.

G. 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 datasets are available under https://doi.org/10.6084/m9.figshare.7326842.v1.

H. Code availability

The authors declare that all software used to obtain the results of this work are publicly accessible as open-source software: python including scipy, tensorflow, as well as pyGDM[47], our own implementation of the Green Dyadic Method. Our scripts can be made accessible from the corresponding author upon reasonable request.