



OPEN

## Low-dose shift- and rotation-invariant diffraction recognition imaging

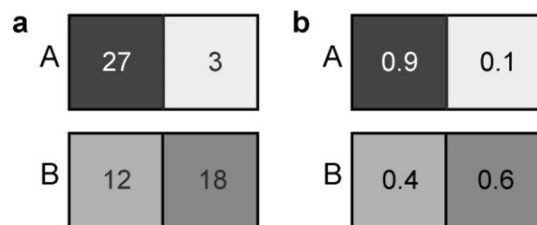
Tatiana Latychevskaia<sup>1,2✉</sup> & Alice Kohli<sup>2</sup>

A low-dose imaging technique which uses recognition rather than recording of a full high-resolution image is proposed. A structural hypothesis is verified by probing the object with only a few particles (photons, electrons). Each scattered particle is detected in the far field and its position on the detector is analysed by applying Bayesian statistics. Already a few detected particles are sufficient to confirm a structural hypothesis at a probability exceeding 95%. As an example, the method is demonstrated as an application in optical character recognition, where a hand-written number is recognized from a set of different written numbers. In other provided examples, the structural hypothesis of a single macromolecule is recognized from a diffraction pattern acquired at an extremely low radiation dose, less than one X-ray photon or electron per Å<sup>2</sup>, thus leaving the macromolecule practically without any radiation damage. The proposed principle of low-dose recognition can be utilized in various applications, ranging from optical character recognition and optical security elements to recognizing a certain protein or its conformation.

One can recognize a familiar object or scene from only a few details without a need for an entire high-resolution image. Thus, a few measurements can be sufficient to extract all the relevant information. The task of recognition finds applications in different areas. In optics, for example, optical character recognition (OCR) allows for identifying the correct symbol from a set of available symbols. Although OCR can be routinely performed by commercially available devices, it is still a challenge to identify a character from a noisy and low-quality image, one solution here is to apply neural networks<sup>1</sup>. In biology, molecular structures can be simulated based on theoretical models in great detail. The purpose of an experiment can thus be replaced from obtaining a full high-resolution image of macromolecule to recognizing a certain molecular structure. The recent breakthrough in solving the protein folding problem using artificial intelligence (AlphaFold) has already provided a large number of models to the three-dimensional (3D) shapes of proteins<sup>2</sup>. At the same time, the experimental imaging of a single protein still remains a challenge due to the radiation damage problem<sup>3</sup>. Most protein structures are being solved using X-ray crystallography and cryo-electron microscopy, and are the result of averaging over tens of thousands of molecules. Presently, there is a strong wish to develop experimental techniques which would allow atomic resolution imaging of truly individual molecules<sup>4–14</sup>. However, radiation damage is the main limiting factor<sup>15</sup>, since a macromolecule is destroyed long before a sufficient number of scattering events is detected for a high-resolution structural analysis<sup>3,4</sup>. Alternative methods for structural determination which minimize the radiation dose have recently been proposed, they allow for the verification of a structural hypothesis with just a few scattering events. In these schemes, a quantum sorter is designed based on the molecular structure hypothesis and is positioned between the macromolecule and the detector<sup>16,17</sup>. Although highly promising, the experimental realization of such methods has been troubled due to several factors, mainly due to the practical limits in nano-fabrication of the structure-defined diffractive elements and the necessity of an atomic-precision alignment of the sorter caused by the high sensitivity of the method to lateral shifts and rotations of either the molecule or the sorter (the latter is illustrated in Fig. S1).

Here, we propose a low-dose diffraction recognition imaging which identifies the correct object from a set of hypotheses by using Bayesian statistics analysis. Two-dimensional (2D) diffraction pattern is replaced with one-dimensional (1D) probability function obtained from azimuthal-averaged diffraction pattern, which solves the two issues in the quantum recognition schemes<sup>16,17</sup>: no sorting element is needed and object can be recognized in any in-plane orientation. The method is independent of the lateral position and in-plane rotation of the object. We demonstrate an application of the method for optical character recognition (OCR)<sup>18</sup>, where a hand-written character is recognized from a set of different written characters. We also provide an example of recognizing the

<sup>1</sup>Paul Scherrer Institute, Forschungsstrasse 111, 5232 Villigen, Switzerland. <sup>2</sup>Physics Department, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. ✉email: tatiana.latychevskaia@psi.ch



**Figure 1.** Two two-pixel images, A and B (a) and their corresponding probability density functions (PDFs) (b).

structural hypothesis of a single macromolecule's orientation from its diffraction pattern. It is shown that already a few detected particles are sufficient to confirm a structural hypothesis at a high probability exceeding 95%.

## Principle

**Low-dose recognition imaging.** For  $M$  object hypotheses,  $M$  corresponding images  $I_i(p)$  are given and the corresponding probability density functions (PDFs) are calculated as:

$$P(p|I_i) = \frac{I_i(p)}{\sum_{p=1}^N I_i(p)}, \quad (1)$$

where  $p$  is the pixel coordinate and  $N$  is the total number of pixels. For a 2D diffraction pattern,  $p$  is running through the  $(q_x, q_y)$  coordinates in the far-field detector plane  $(q_x, q_y)$ . The a priori probabilities of having any of the given hypotheses are equal:

$$P_1 = P_2 = \dots = P_M = \frac{1}{M}. \quad (2)$$

After the first particle is detected in pixel  $p_1$ , the a posteriori probabilities that the corresponding PDF corresponds to the imaged structure are given by

$$P'_i = \frac{P_i \cdot P(p_1|I_i)}{\sum_{j=1}^M [P_j \cdot P(p_1|I_j)]}. \quad (3)$$

The updated probabilities become the input a priori probabilities for the analysis of the next particle detected in pixel  $p_2$ . The probabilities are then updated according to Eq. (3), and so forth. Equation (3) gives  $P'_1 = P_1 = 1$  for  $M=1$ .

The proposed Bayesian recognition principle is illustrated using two two-pixel images, A and B with intensity distributions  $I_A(p)$  and  $I_B(p)$ , shown in Fig. 1a.

According to Eq. (1) the PDF for each image are given by:

$$P(p|A) = \frac{I_A(p)}{\sum_{p=1}^2 I_A(p)}, \quad P(p|B) = \frac{I_B(p)}{\sum_{p=1}^2 I_B(p)},$$

and they are shown in Fig. 1b. The initial probabilities that the detected image corresponds to hypothesis A or B are equal to 0.5:  $P_A = P_B = 0.5$ . After the first particle is detected in pixel  $p$ , according to Eq. (3), the probability that the observed image corresponds to hypothesis A is given by

$$P'_A = \frac{P_A \cdot P(p|A)}{P_A \cdot P(p|A) + P_B \cdot P(p|B)}.$$

When the first particle is detected at pixel  $p = 1$ , the updated probability, according to Eq. (3) is given by:

$$P'_A = \frac{0.5 \cdot P(1|A)}{0.5 \cdot P(1|A) + 0.5 \cdot P(1|B)} = \frac{0.9}{0.9 + 0.4} = 0.692.$$

The updated probability that the detected image corresponds to hypothesis B is given by  $P'_B = 1 - P'_A = 0.308$ .  $P'_A$  and  $P'_B$  are a posteriori probabilities, which are then used as input probabilities for the analysis of the next detected particle. When the second particle is again detected at pixel  $p = 1$ , the updated probability that the observed image corresponds to hypothesis A is given by:

$$P'_A = \frac{0.692 \cdot P(1|A)}{0.692 \cdot P(1|A) + 0.308 \cdot P(1|B)} = \frac{0.692 \cdot 0.9}{0.692 \cdot 0.9 + 0.308 \cdot 0.4} = 0.835,$$

and the probability that the detected image corresponds to hypothesis B is given by  $P'_B = 0.165$ . And so forth, the routine is repeated for each next detected particle. Thus, by counting individual particles arriving at the detector and analysing their position on the detector, the probabilities of the hypotheses are quantitatively evaluated.

**Low-dose shift-invariant diffraction recognition imaging.** In general, an object's image is shifted together with the object. Thus, each image of a shifted object corresponds to a different hypothesis. To create an imaging method which recognizes object independently on its lateral position, a diffraction pattern of the object can be considered instead of its image. A diffraction pattern of an object distribution is obtained by acquiring the squared amplitude of the Fourier transform (FT) of the object distribution. Shifting the object distribution creates an additional phase factor of its complex-valued Fourier transform distribution without changing its amplitude. The resulting diffraction pattern is thus independent on the object shift.

In an optical diffraction experiment, the probing particles (photons, electrons) scattered off an object give rise to a diffraction pattern in the far field. The diffraction pattern, in principle, is a distribution of the probability to detect a particle at a certain point on the detector. A probing particle, after being scattered by the object, changes its state in such a way that the probability of detecting it at a certain point on the detector in the far field is given by the diffraction pattern. A single scattered particle could arrive seemingly at any position at the detector. But when the second, the third and further scattered particles also arrive at the positions that are distributed according to a diffraction pattern corresponding to a certain sample hypothesis, there is a high probability for the structural hypothesis to be correct. In this arrangement, even a few scattered probing particles are sufficient to determine the correct structural hypothesis. While conventional statistics analysis considers the outcomes of detecting scattered particles as being independent of one another, the Bayesian statistical analysis takes into account previous events and updates the probability based on previous outcomes. Quantitatively, the normalized diffraction pattern  $I(q_x, q_y) = I_0(q_x, q_y) / \iint I_0(q_x, q_y) dq_x dq_y$  provides a probability density function (PDF) which gives the probability of detecting the scattered particle at a position  $(q_x, q_y)$  on the detector;  $I_0(q_x, q_y)$  is the recorded diffraction pattern.

The recognition analysis performed on 2D diffraction patterns instead of 2D images of an object does not depend on the in-plane shift of the object. However, it still depends on the in-plane rotation of the object. When an object is rotated in-plane, its 2D diffraction pattern is also rotated by the same degree. Thus, the diffraction patterns of the object and of the rotated object are different and therefore correspond to two different structural hypotheses. The diffraction patterns of the object corresponding to its different in-plane rotations can be used as a set of hypotheses and the in-plane rotation of the object can be determined from them.

**Low-dose shift- and rotation-invariant diffraction recognition imaging.** To make the recognition method independent of the in-plane rotation of the object, a 1D PDF is obtained from azimuthally-averaged 2D diffraction pattern as follows. For each  $q$  value ( $q = \sqrt{q_x^2 + q_y^2}$ ), and each azimuthal angle  $\vartheta$ , the value of diffraction pattern at  $(q, \vartheta)$  is extracted giving  $I_0(q, \vartheta)$ . The azimuthal angle values are run over  $2\pi$  in steps of  $\Delta\vartheta$  and the total sum of all particles detected at a given  $q$  is given by

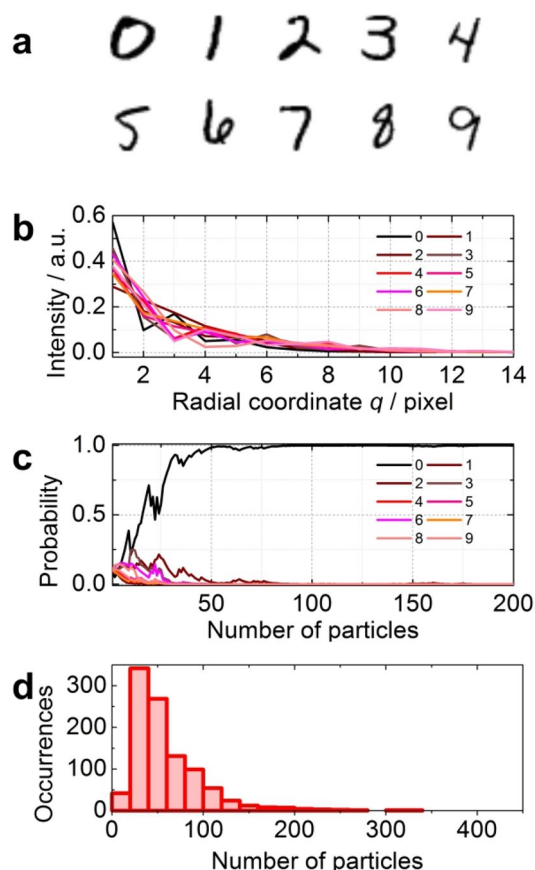
$$I(q) = \frac{2\pi q}{N} \sum_{n=0}^{N-1} I_0(q, n\Delta\vartheta), \quad (4)$$

where  $N = \frac{2\pi}{\Delta\vartheta}$  is the number of steps. The 1D PDF function is then calculated by normalizing  $I(q)$  using Eq. (1). The obtained 1D PDF function is related only to the radial positions of the detected particles. The method is also insensitive to the lateral position of the sample, since a diffraction pattern is insensitive to the lateral (in-plane) shifts of the imaged sample. An example of 1D radial profile evolution as a function of the number of detected particles in a 2D image is provided in Fig. S2. The approach of using 1D radial profile has been previously explored in applying deep-learning techniques for structural recognition in X-ray powder diffraction data<sup>19</sup>.

For a 3D object in a certain orientation, its 2D diffraction pattern is approximately given by the diffraction pattern of the object's 2D projection. An out-of-plane rotation of a 3D object results in a different 2D projection and, as a result, in a different 2D diffraction pattern. A different diffraction pattern, in turn, corresponds to a different hypothesis. Different hypotheses can correspond to different out-of-plane rotations of the same 3D object, or different out-of-plane rotations of different 3D objects. Thus, the recognition process can be realized for: different out-of-plane rotations of the same 3D object, or different out-of-plane rotations of different 3D objects. Once the correct out-of-plane orientation of a 3D object is determined, the in-plane rotation can be found as explained at the end of previous section: by applying recognition of the entire 2D diffraction pattern against diffraction patterns corresponding to different in-plane rotations of the object.

## Results

**Character recognition.** The principle of the low-dose diffraction recognition imaging is demonstrated using an example of 10 images of hand-written numbers, each sampled with  $28 \times 28$  pixels, Fig. 2a. These are the known distribution—hypotheses. For each object, its diffraction pattern was calculated, the azimuthally averaged 1D profile was extracted, and the corresponding 1D PDF was obtained, shown in Fig. 2b. In the numerical experiment, one object was selected, the corresponding diffraction pattern was calculated, and the radial distribution was obtained by azimuthal averaging of the diffraction pattern. To mimic experimental conditions, Gaussian-distributed noise was added in order to reach SNR = 2. The Gaussian-distributed noise was added as follows: at each value of signal  $S$ , an array of Gaussian-distributed noise with standard deviation  $S/\text{SNR}$  was generated using a built in routine (LabView), the noise distribution and its parameters were checked, and the first

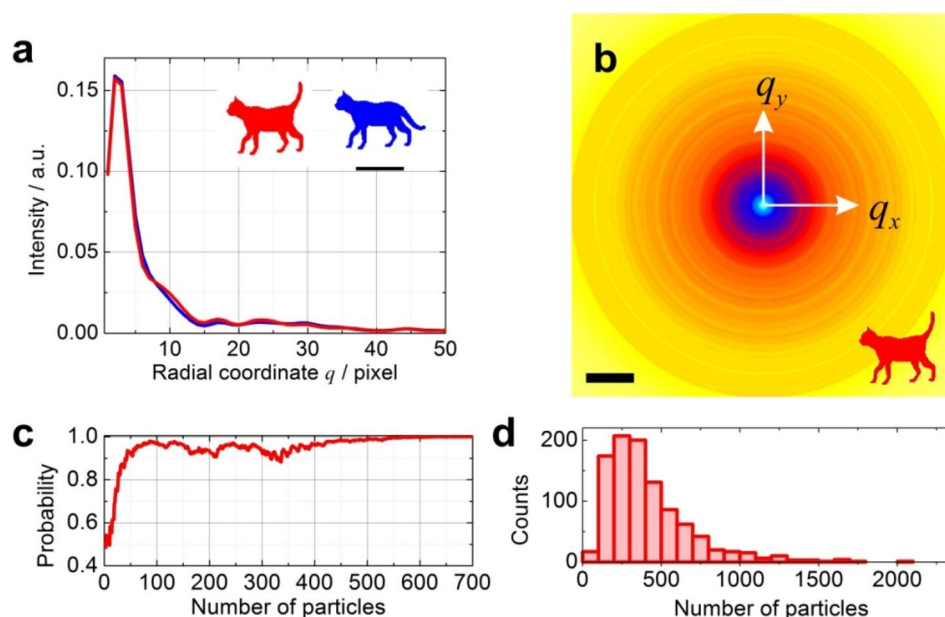


**Figure 2.** Principle of low-dose diffraction recognition imaging. **(a)** Ten samples constituting the then hypotheses. **(b)** 1D PDF profiles corresponding to the diffraction patterns of each sample. **(c)** Evolution of probabilities as a function of number of detected particles. **(d)** Number of particles required to achieve 95% probability that the sample structure hypothesis is correct; results of 1000 experiments.

value from the array was added to the signal value. The PDF was calculated using Eq. (1). The radial positions on the detector of the particle scattered off the probed structure were modelled using Monte Carlo simulations. Each particle scattered off the sample was analysed by Bayesian analysis against the ten noise-free hypotheses using Eq. (3). According to Eq. (2), the initial probability for all hypotheses is 0.1. As the number of the detected particles increased, the probability approached 1 for the correct hypothesis, and 0 for all other hypotheses, Fig. 2c. Numerical experiments showed that based on the results of 1000 numerical experiments, about 40 particles were needed to achieve 95% confidence level of the probability that the hypothesis was correct, Fig. 2d. It was verified that this number was approximately the same for the same images sampled with different number of pixels:  $28 \times 28$  pixels and  $140 \times 140$  pixels. About 40 particles were needed to achieve 95% confidence level of the probability that the hypothesis was correct for each image size (Fig. S3).

The proposed low-dose recognition method allows for the recognition of a 2D object independent of the object's shift or rotation. The method can be applied, for example, in optical character recognition (OCR)<sup>18</sup>, where quality and noise of the images is often a problem for the successful identification of a character. The example demonstrated here shows that character recognition can be realized for noisy and arbitrarily rotated images by acquiring only 100 particles (samples) instead of acquiring full two-dimensional images.

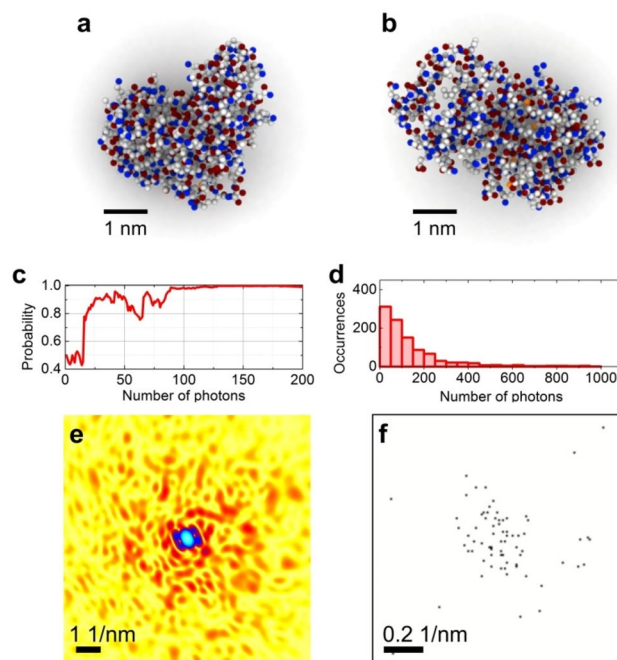
**Conformation recognition.** The example in this section shows how a small change in the sample can be recognized using the proposed technique. Here, a test sample is a cat cartoon in two possible conformations—with its tail up ("up") and with its tail down ("down"), Fig. 3a. 1D PDF profiles obtained from the corresponding azimuthal-averaged diffraction patterns exhibit almost identical distributions, Fig. 3a–b. The probed structure is in the "up" conformation and it is verified against the two conformation hypotheses: "up" and "down". The 1D PDF is calculated from the radial distribution of the azimuthal-averaged diffraction patterns corresponding to the "up" conformation using Eq. (4), with noise added to reach  $\text{SNR}=2$ . The numerical experiments show that approximately 300 particles are required to reach a 95% probability that the structural hypothesis is correct, Fig. 3c–d. Cross-correlation function (CCF) analysis is a conventional approach to quantitatively estimate whether the signal is matching a reference signal. By applying the CCF analysis to the radial distributions updated after each detected particle, we observed that both CCFs exhibit very close values with only about  $10\text{E}-2$  difference (Fig. S4), thus providing a less clear answer to the structural hypothesis.



**Figure 3.** Principle of structure determination demonstrated for a binary sample – a cat cartoon. **(a)** The sample can be found in two conformations: with tail up ("up", shown in red) or tail down ("down", shown in blue). **(a)** 1D PDF profiles corresponding to the diffraction patterns; the region where two signals exhibit difference is shown. The scalebar is 60 a.u. **(b)** Azimuthally-averaged diffraction pattern of "up" conformation. The scalebar is 5 pixels in the Fourier space, with 1 pixel = 1/400 a.u. **(c)** Probability of the right hypothesis ("up") as a function of the number of detected scattered particles. **(d)** Number of particles required to achieve 95% probability that the sample structure hypothesis is correct; results of 1000 experiments are shown.

**Macromolecule orientation recognition from X-ray diffraction pattern.** The low-dose recognition method can be particularly useful for verifying the structure of radiation sensitive biological macromolecules, as for example in the quantum recognition schemes<sup>16,17</sup>. In the next example, a single lysozyme molecule<sup>20,21</sup> is recognized from two possible hypotheses: molecule being in  $xy$  or  $xz$  orientation, Fig. 4a,b. The enzyme was set in the  $xy$ -orientation, Fig. 4a, and probed with X-ray photons of 1 Å wavelength. The diffraction patterns were simulated as described in Methods. The radial positions of the individual particles on the detector were simulated using the Monte Carlo technique, the 1D PDF was created in a similar way as in the previous example, with noise added to reach  $\text{SNR} = 2$ ; the currently reported noise level for diffraction patterns in single particle imaging (SPI) amounts to  $\text{SNR} = 7$ <sup>22</sup>. Approximately 100 elastically scattered photons were needed to reach 95% probability that the structural hypothesis was correct (Fig. 4c,d). This results in 0.11 photons/Å<sup>2</sup> for elastically scattered particles, or an approximate radiation dose of 1.1 photons/Å<sup>2</sup>, considering that only one photon out of ten is elastically scattered<sup>3</sup>. Thus, a structural hypothesis of a single lysozyme molecule can be verified with a radiation dose of one X-ray photon/Å<sup>2</sup>, which is orders of magnitudes less than typical radiation dose in high-resolution X-ray imaging<sup>4–14</sup>. The total number of 100 photons is extremely low. To illustrate this fact, the intensity distribution obtained with 100 photons is shown in Fig. 4e,f. This low-dose diffraction pattern exhibits almost no signal outside of the central region of 0.2 1/nm resolution, the limit which is comparable to the size of molecule itself, 5 nm. Such a low-dose electron diffraction pattern cannot be used for structure determination by an iterative phase retrieval algorithm. In the provided here example a lysozyme molecule was selected to compare the obtained results to the results presented by Neutze et al<sup>3</sup> who proposed "diffract and destroy" experiment and showed that a diffraction pattern of a single lysozyme molecule would exhibit such a low number of counts per pixels at the rim (where the resolution is 2.2 Å) that iterative phase retrieval methods cannot be applied to reconstruct the molecular structure. Here we show that, alternatively, structure of an individual macromolecule can be *recognized* from its low-dose diffraction pattern.

**Macromolecule orientation recognition from electron diffraction pattern.** Similar results were obtained for electron diffraction of an individual macromolecule, as for example EspB, a virulence protein secreted from *Mycobacterium tuberculosis*<sup>23,24</sup>. The diffraction patterns were simulated as described in Methods. Gaussian noise was added to the diffraction pattern so that the SNR at a certain  $q$  value is modelled as  $\text{SNR}(q) = 1/(30q)$ , where  $q$  is in Angstrom; for example  $\text{SNR} = 0.33$  at  $q = 0.1$  1/Å and  $\text{SNR} = 0.11$  at  $q = 0.3$  1/Å. This particular model of  $\text{SNR}(q)$  roughly approximates the experimental observations<sup>25</sup>. The orientation of the protein was validated against the  $xy$  and  $xz$  orientations, Fig. 5a,b. The probed molecule was positioned in the  $xy$ -orientation (Fig. 5a) and approximately 40'000 electrons were needed to reach 95% probability that the structural hypothesis was correct (Fig. 5c,d). For the total probed area of  $8.23\text{E} + 5$  Å<sup>2</sup> this translates into a dose of  $48.6\text{E} - 3$  e/Å<sup>2</sup>, a much smaller dose than the typical  $2\text{--}5$  e/Å<sup>2</sup> required in high-resolution imaging cryo-electron



**Figure 4.** Structure verification of a single lysozyme molecule from a low-dose X-ray diffraction pattern. (a) and (b) structure of the molecule in  $xy$ - and  $xz$ -orientation, respectively. (c) Probability of the right hypothesis ( $xy$ -orientation) as a function of the number of detected photons. (d) Number of photons required to achieve 95% probability that the sample structure hypothesis is correct; results of 1000 experiments are shown. (e) and (f) X-ray diffraction patterns of a lysozyme molecule formed by elastically scattered (e)  $1.24\text{E}+8$  photons/ $\text{\AA}^2$  and (f)  $0.11$  photons/ $\text{\AA}^2$  (f). In (e), the full,  $1024 \times 1024$  pixel diffraction pattern is shown, the maximum is  $8.0\text{E}+7$  counts per pixel (cpp). In (f), only the central  $128 \times 128$  pixel region of  $1024 \times 1024$  diffraction pattern is shown, with the maximum of 1 cpp.

microscopy<sup>26–28</sup>. The electron diffraction pattern produced with a dose of  $48.6\text{E}-3$  e/ $\text{\AA}^2$  exhibits mainly a very low-resolution signal, Figs. 5e–f.

## Discussion

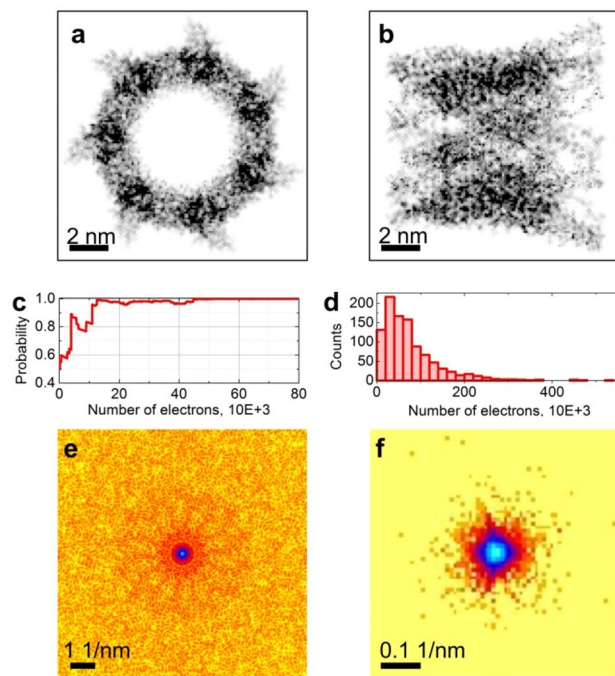
Low-dose diffraction recognition imaging verifies a structural hypothesis by detecting only a few scattered particles. Detecting scattered particles in the far field and analysing only their radial positions on the detector make the method invariant to shift and in-plane rotation of the object. For live low-dose imaging, the scattered particles can be detected and analysed during imaging. For experiments where only a post-experimental low-count image is available, as in the case of a diffraction pattern of a protein, each individual count in a pixel can be considered as a detected particle analysed by the proposed method as explained in the example of two-pixel images.

The proposed method can be adapted for optical image recognition, which finds applications in microscopy, medical imaging, robotic vision, optical remote sensing, and optical security methods. In optical security techniques, image recognition is currently achieved by performing correlation of a reference object and a target object to be recognized, for example by means of holography<sup>29,30</sup>. In the herewith provided example of OCR, a character is recognised using much less samples than when acquiring a full image of the character.

When applied for structural biology, the herewith proposed method allows for the verification of a structural hypothesis by probing the sample with only a few particles (photons, electrons). Previously, Neutze et al<sup>3</sup> described their "diffract and destroy" experiment and showed that a diffraction pattern of a single biological macromolecule (lysozyme) would exhibit such a low number of counts per pixel at the rim of the diffraction pattern. Such a diffraction pattern cannot be used for high-resolution structure reconstruction by iterative phase retrieval methods. With the method proposed here we show that, alternatively, structure of an individual macromolecule can be *recognized* from a diffraction pattern, even when the diffraction pattern is acquired at a very low dose.

High-resolution imaging techniques such as cryo-EM or X-ray SPI<sup>14,31</sup>, conventionally employ a low-resolution model of the molecular structure as reference to perform cross-correlation analysis for alignment of the experimental images into a high-resolution image of the structure. Currently, the EMC method (E for expansion, M for maximization and C for compression), proposed by Loh and Elser, is applied for analysis of SPI data to reconstruct a particle's three-dimensional (3D) diffraction intensity from many photon shot-noise limited two-dimensional measurements<sup>32</sup>.

In the method proposed here, instead of a full image only a few scattered counts are detected, thus demonstrating that the cross-correlation analysis can be replaced by the low-dose recognition method.



**Figure 5.** Structure verification of a single EspB protein by counting single scattered electrons. (a) and (b) the phase distributions of the exit wave are shown calculated for 200 keV electrons, the macromolecule in  $xy$  and  $xz$  orientation, respectively. The phase shift reaches 0.84 in (a) and 0.74 rad in (b). (c) Probability of the right hypothesis ( $xy$ -orientation) as a function of the number of detected electrons. (d) Number of electrons required to achieve 95% probability that the sample structure hypothesis is correct; results of 1000 experiments are shown. (e) and (f) diffraction patterns of a single EspB protein at a radiation dose of (e)  $8.23E+5 \text{ e}/\text{\AA}^2$  (b) and  $48.6E-3 \text{ e}/\text{\AA}^2$  (f). (e) Full,  $1024 \times 1024$  pixel diffraction pattern and (f) only the central  $64 \times 64$  pixel region of  $1024 \times 1024$  diffraction pattern are shown.

Another possible application of the herewith proposed method can be a sequential imaging of macromolecule undergoing conformational or dynamical changes where at each time frame the macromolecule is probed with only a few particles (photons, electrons), so that the conformational changes can be determined without significant damaging of the structure during the entire acquisition time.

We provided several examples of possible applications of the method. In general, the proposed method can be applied for any recognition task where the hypotheses exhibit different PDF distributions. This could include: different orientations of 3D objects, different 2D or 3D objects, different conformations of the same objects, etc.

## Methods

**X-ray diffraction pattern of a single lysozyme molecule.** The diffraction pattern of a single lysozyme molecule were calculated by coherently adding waves scattered off individual atoms:

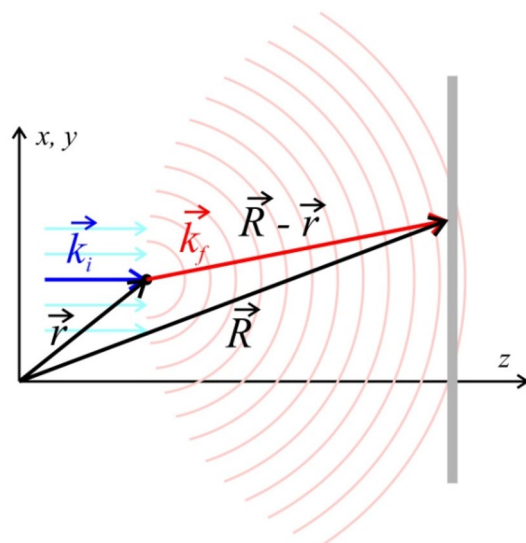
$$I(\vec{R}) = \left| \sum_{j=1}^N f_{i,j}(q) \exp(ikz) \frac{\exp(ik|\vec{R} - \vec{r}_j|)}{|\vec{R} - \vec{r}_j|} \right|^2,$$

where  $f_j(q)$  is the atomic form factor corresponding to  $j$ th chemical element:

$$f_j(q) = \sum_{i=1}^4 a_i \exp \left[ -b_i \left( \frac{q}{4\pi} \right)^2 \right] + c$$

where the parameters  $a_i$ ,  $b_i$ , and  $c$  are provided in the International Tables for Crystallography<sup>33</sup>;  $\vec{r} = (x, y, z)$  is the coordinate of the atom,  $\vec{R} = (X, Y, Z)$  is the coordinate on the detector,  $\vec{q}$  is the momentum transfer  $\vec{q} = \vec{k}_f - \vec{k}_i$ ,  $\vec{k}_i = k(0, 0, 1)$  is the wave vector of the incident wave and  $\vec{k}_f = \frac{k}{|\vec{R} - \vec{r}|} (X - x, Y - y, Z - z)$  is the wave vector of the scattered wave,  $k = \frac{2\pi}{\lambda}$ , Fig. 6. The atomic coordinates of the lysozyme molecule were downloaded from the protein database structure 253L<sup>20,21</sup>.

**Electron diffraction pattern of a single virulence protein.** The electron diffraction pattern of a single virulence protein was calculated using the following multi-slice simulation protocol: (1) The atomic coordinates of the macromolecule were downloaded from the protein database structure 3J83<sup>23,24</sup>. (2) The sequence



**Figure 6.** Coordinates and symbols used in calculation of X-ray scattering.

of atoms was re-arranged in the order of the increasing  $z$ -coordinate, and the atoms were numbered as a1, a2 etc. (3) An incident plane wave with a unit amplitude was assumed,  $u_1(x_1, y_1, z_1) = 1$ . (4) The coordinates of the first atom a1 were read from the text file as  $(x_1^{(1)}, y_1^{(1)}, z_1)$ . (5) The transmission function in the plane at  $z_1$  was calculated as  $t_1(x_1, y_1, z_1) = \exp[i\sigma v_z(x_1, y_1)]$ , where  $\sigma$  is the interaction parameter at 200 keV and  $v_z(x_1, y_1)$  is the projected potential of atom a1, calculated from the tabulated parameters corresponding to the chemical elements as described in reference<sup>34</sup>. (6) The exit wave in the plane  $(x_1, y_1, z_1)$  was calculated as  $u'_1(x_1, y_1, z_1) = u_1(x_1, y_1, z_1)t_1(x_1, y_1, z_1)$ . (7) The  $z$ -coordinate of the next atom a2 was read as  $z_2$ , and the distance  $\Delta z = z_2 - z_1$  was calculated. (8) The wave function  $u'_1(x_1, y_1, z_1)$  was propagated for  $\Delta z$  using the angular spectrum method<sup>35</sup>. The resulting wavefront was  $u_2(x_2, y_2, z_2)$ . (9) The wave function was propagated through the sample, atom by atom, by repeating steps 4 to 8 until the electron wave had propagated through all the atoms. The obtained distribution is the exit wave. (10) The diffraction pattern was calculated as the square of the amplitude of the Fourier transform of the exit wave.

### Data availability

All data are available in the main text or the supplementary information.

Received: 7 February 2022; Accepted: 24 June 2022

Published online: 01 July 2022

### References

- Sabourin, M. & Mitiche, A. Optical character recognition by neural network. *Neural Netw.* **5**, 843–852 (1992).
- Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583 (2021).
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
- Miao, J. W., Charalambous, P., Kirz, J. & Sayre, D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature* **400**, 342–344 (1999).
- Miao, J. W. *et al.* Imaging whole *Escherichia coli* bacteria by using single-particle X-ray diffraction. *PNAS* **100**, 110–112 (2003).
- Shapiro, D. *et al.* Biological imaging by soft X-ray diffraction microscopy. *PNAS* **102**, 15343–15346 (2005).
- Song, C. Y. *et al.* Quantitative imaging of single, unstained viruses with coherent x-rays. *Phys. Rev. Lett.* **101**, 158101 (2008).
- Williams, G. J. *et al.* High-resolution X-ray imaging of plasmodium falciparum-infected red blood cells. *Cytom. Part A* **73A**, 949–957 (2008).
- Huang, X. *et al.* Soft X-ray diffraction microscopy of a frozen hydrated yeast cell. *Phys. Rev. Lett.* **103**, 198101 (2009).
- Nishino, Y., Takahashi, Y., Imamoto, N., Ishikawa, T. & Maeshima, K. Three-dimensional visualization of a human chromosome using coherent X-ray diffraction. *Phys. Rev. Lett.* **102**, 018101 (2009).
- Nelson, J. *et al.* High-resolution X-ray diffraction microscopy of specifically labeled yeast cells. *PNAS* **107**, 7235–7239 (2010).
- Wilke, R. N. *et al.* Hard X-ray imaging of bacterial cells: nano-diffraction andptychographic reconstruction. *Opt. Express* **20**, 19232–19254 (2012).
- Seibert, M. M. *et al.* Single mimivirus particles intercepted and imaged with an x-ray laser. *Nature* **470**, 78–81 (2011).
- Ekeberg, T. *et al.* Three-dimensional reconstruction of the giant mimivirus particle with an X-ray free-electron laser. *Phys. Rev. Lett.* **114**, 098102 (2015).
- Howells, M. R. *et al.* An assessment of the resolution limitation due to radiation-damage in x-ray diffraction microscopy. *J. Electron Spectrosc.* **170**, 4–12 (2009).
- Okamoto, H., Latychevskaia, T. & Fink, H.-W. A quantum mechanical scheme to reduce radiation damage in electron microscopy. *Appl. Phys. Lett.* **88**, 164103 (2006).
- Troiani, F. *et al.* Efficient molecule discrimination in electron microscopy through an optimized orbital angular momentum sorter. *Phys. Rev. A* **102**, 2001.08918 (2020).
- Mori, S., Nishida, H. & Yamada, H. *Optical Character Recognition* (Wiley, New York, 1999).



19. Lee, J.-W., Park, W. B., Lee, J. H., Singh, S. P. & Sohn, K.-S. A deep-learning technique for phase identification in multiphase inorganic compounds using synthetic XRD powder patterns. *Nat. Commun.* **11**, 86 (2020).
20. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. A relationship between protein stability and protein function. *PNAS* **92**, 452–456 (1995).
21. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. PDB ID: 253L. <https://doi.org/10.1073/pnas.1092.1072.1452> (1995).
22. Sobolev, E. *et al.* Megahertz single-particle imaging at the European XFEL. *Commun. Phys.* **3**, 97 (2020).
23. Solomonson, M. *et al.* Structure of EspB from the ESX-1 type VII secretion system and insights into its export mechanism. *Structure* **23**, 571–583 (2015).
24. Solomonson, M. *et al.* PDB ID: 3J83. <https://doi.org/10.1016/j.str.2015.1001.1002> (2015).
25. Palovcak, E., Asarnow, D., Campbell, M. G., Yu, Z. L. & Cheng, Y. F. Enhancing the signal-to-noise ratio and generating contrast for cryo-EM images with convolutional neural networks. *IUCr* **7**, 1142–1150 (2020).
26. Henderson, R. Cryoprotection of protein crystals against radiation-damage in electron and x-ray diffraction. *Proc. R. Soc. Lond. B* **241**, 6–8 (1990).
27. Henderson, R. The potential and limitations of neutrons, electrons and x-rays for atomic-resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
28. Henderson, R. Realizing the potential of electron cryo-microscopy. *Q. Rev. Biophys.* **37**, 3–13 (2004).
29. Poon, T. C. & Kim, T. Optical image recognition of three-dimensional objects. *Appl. Opt.* **38**, 370–381 (1999).
30. Esteve-Taboada, J. J., Garcia, J. & Ferreira, C. Rotation-invariant optical recognition of three-dimensional objects. *Appl. Opt.* **39**, 5998–6005 (2000).
31. Branden, G. *et al.* Coherent diffractive imaging of microtubules using an X-ray laser. *Nat. Commun.* **10**, 2589 (2019).
32. Loh, N. T. D. & Elser, V. Reconstruction algorithm for single-particle diffraction imaging experiments. *Phys. Rev. E* **80**, 026705 (2009).
33. Brown, P. J., Fox, A. G., Maslen, E. N., M. A. O'Keefe, M. A. & Willis, B. T. M. Intensity of diffracted intensities. *Int. Tables Crystallogr. C*, 554–595 (2006).
34. Kirkland, E. J. *Advanced computing in electron microscopy* (Springer, Berlin, 2010).
35. Latychevskaia, T. & Fink, H.-W. Practical algorithms for simulation and reconstruction of digital in-line holograms. *Appl. Opt.* **54**, 2424–2434 (2015).

## Acknowledgements

The research was supported by Swiss National Foundation grant 200021\_197107.

## Author contributions

T.L. conceived the idea, made the simulations and data analysis. T.L. and A.K. discussed the results and wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15486-y>.

**Correspondence** and requests for materials should be addressed to T.L.

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

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022