Possibility of single biomolecule imaging with coherent amplification of weak scattering x-ray photons

Tsumoru Shintake
RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan
(Received 23 April 2008; revised manuscript received 21 August 2008; published 3 October 2008)

The number of photons produced by coherent x-ray scattering from a single biomolecule is very small because of its extremely small elastic-scattering cross section and low damage threshold. Even with a high x-ray flux of 3×10^{12} photons per 100-nm-diameter spot and an ultrashort pulse of 10 fs driven by a future x-ray free electron laser (x-ray FEL), it has been predicted that only a few 100 photons will be produced from the scattering of a single lysozyme molecule. In observations of scattered x rays on a detector, the transfer of energy from wave to matter is accompanied by the quantization of the photon energy. Unfortunately, x rays have a high photon energy of 12 keV at wavelengths of 1 Å, which is required for atomic resolution imaging. Therefore, the number of photoionization events is small, which limits the resolution of imaging of a single biomolecule. In this paper, I propose a method: instead of directly observing the photons scattered from the sample, we amplify the scattered waves by superimposing an intense coherent reference pump wave on it and record the resulting interference pattern on a planar x-ray detector. Using a nanosized gold particle as a reference pump wave source, we can collect 10^4–10^5 photons in single shot imaging where the signal from a single biomolecule is amplified and recorded as two-dimensional diffraction intensity data. An iterative phase retrieval technique can be used to recover the phase information and reconstruct the image of the single biomolecule and the gold particle at the same time. In order to precisely reconstruct a faint image of the single biomolecule in Angstrom resolution, whose intensity is much lower than that of the bright gold particle, I propose a technique that combines iterative phase retrieval on the reference pump wave and the digital Fourier transform holography on the sample. By using a large number of holography data, the three-dimensional electron density map can be assembled.

DOI: 10.1103/PhysRevE.78.041906 PACS number(s): 87.85.Ng, 42.30.Rx, 07.85.Tt, 42.40.Ht

I. INTRODUCTION

Protein crystallography has advanced through the use of high-quality x-ray beams generated by synchrotron light sources [1] and with the development of high-brightness x-ray beams produced by an undulator device placed in low-emittance electron storage rings [2]. The increase in the brightness of x-ray beams has made it possible to increase the flux incident on small protein crystals. Currently, crystals with a diameter of 100 μm are routinely analyzed, and the 10-μm-diameter crystals are under investigation. It is comparatively easier to produce high-quality small crystals than larger crystals with a lesser number of defects.

However, crystallization of membrane proteins is extremely challenging because they require detergents to maintain them in isolation or solution; unfortunately, such detergents often interfere with the crystallization process. Membrane proteins form a large component of genomes and include several proteins of great physiological importance, such as ion channels and receptors [3].

Recently, the practical applications of single-pass free-electron lasers (FELs) [4,5] at short wavelengths have increased as a result of long-term research and development of important devices and technologies, such as low-emittance electron sources, accurate undulator devices, and high-precision electron-beam control techniques. Currently, three large x-ray FELs emitting 1 Å x-ray laser beams are under development and will be available for use in the next few years [6–8].

Since x-ray FELs produce extremely high peak power with a short pulse duration, the potential use of these devices to determine the structure of a single biomolecule or individual cells from scattered photons in one shot has been examined [9]. Through a computer simulation, which considers in detail the damage associated with photoabsorption and Auger electron emission and their various effects, it has been demonstrated that coherently scattered photons from a single biomolecule can be collected using femtosecond pulses before atoms start drifting due to Coulomb forces. However, the number of available photons is quite limited. A previous report has predicted that a maximum of 682 photons can be scattered from a single lysozyme molecule. In the calculation, however, unrealistic parameters have been assumed: 5×10^{13} photons per pulse in a 100-nm-diameter spot with a 1-fs pulse duration at a wavelength of 1 Å. If we refer to the design parameters of the x-ray FELs currently under development, the feasible parameters appear to be 3×10^{12} photons or less per pulse in a 100-nm-diameter spot with a 10-fs pulse duration; with these parameters, only 100 photons would be scattered into scattering angle for 2–30 Å resolution. This is an extremely small number, which makes it impossible to carry out single-shot imaging.

In this paper, initially, I discuss the double slit thought experiment in order to show signal amplification phenomena associated with the interference effect. Then, I propose a method in which the scattered photons are amplified by superimposing a reference pump wave and the resulting interference fringes are recorded on a planar x-ray detector; similar to the technique followed in holography. On the basis of the four-slit thought experiment, I discuss the signal contents in the interference fringes. To recover an image, I discuss the

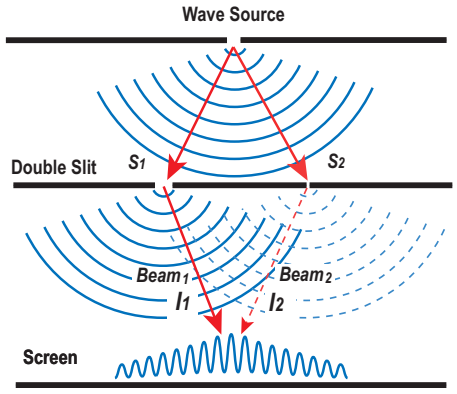

FIG. 1. (Color online) Modified double-slit experiment.

interactive phase retrieval technique and a technique that combines iterative phase retrieval on low-pass filtered data and digital Fourier transform holography.

II. DOUBLE-SLIT THROUGH EXPERIMENT

Figure 1 shows a modified version of Young's double-slit experiment. The width of slit S_2 is considerably less than that of S_1 such that the intensity (and photon flux) of light passing through S_2 is considerably lower than that passing through S_1. The wave functions of the two beams arriving at the screen are expressed as follows:

$$\psi_1 = A_1 \exp i(\omega t - kz + \phi_1), \tag{1}$$

$$\psi_2 = A_2 \exp i(\omega t - kz + \phi_2). \tag{2}$$

These amplitudes are normalized such that the conjugate product $\psi^*\psi = |\psi|^2$ is equal to the photon flux of each beam. Therefore, the probability of finding a photon in the beams is proportional to $I_1 = \psi_1^*\psi_1$ for beam −1 and $I_2 = \psi_2^*\psi_2$ for beam −2.

The two waves interfere and the resulting interference fringes, which are observed on the screen, can be represented by the linear summation of the two wave functions, $\psi = \psi_1 + \psi_2$. The probability of finding a photon on the screen is given by

$$\begin{aligned} I = |\psi|^2 &= \psi^*\psi = (\psi_1{}^* + \psi_2{}^*)(\psi_1 + \psi_2) \\ &= |\psi_1|^2 + 2|\psi_1||\psi_2|\cos(\phi_1 - \phi_2) + |\psi_2|^2 \\ &= I_1 + 2\sqrt{I_1 I_2}\cos(\phi_1 - \phi_2) + I_2. \end{aligned} \tag{3}$$

In the x-ray diffraction experiment, ψ_1 represents the reference pump wave and ψ_2 represents the diffracted wave from the single biomolecule. The second term in Eq. (3) represents the interference fringes, which contain information about the second slit, i.e., its width, intensity of the light beam passing through it, and its distance from the first slit. If there are multiple slits, there will be multiple frequency components in Eq. (3), which can be analyzed by Fourier transformation.

Next, we discuss information that can be retrieved from fringe contrast. From the maximum and minimum amplitudes given in Eq. (3), we determine the modulation depth as follows:

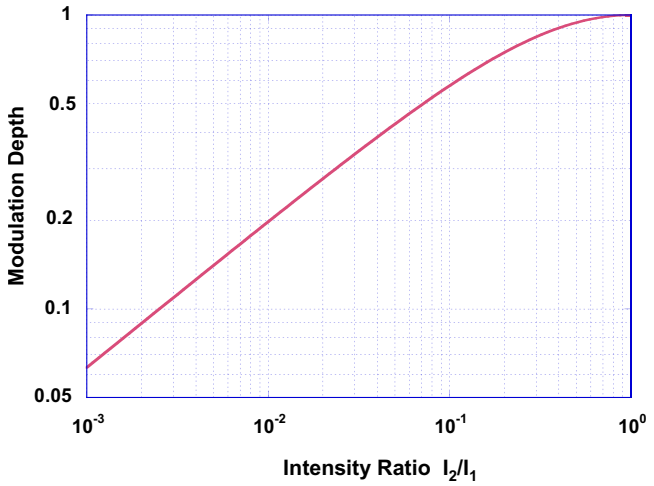

FIG. 2. (Color online) Modulation depth of interference fringe from the modified double-slit experiment.

$$M = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\sqrt{I_2/I_1}}{1 + I_2/I_1}. \tag{4}$$

Figure 2 shows the modulation depth (or visibility) as a function of the intensity ratio I_2/I_1. The modulation depth has a maximum value of 1 when the intensities are equal, i.e., $I_2 = I_1$. By decreasing I_2, the modulation depth decreases, although the rate of this decrease is slower than a first-order dependence, which indicates that the modulation depth or fringe contrast is visible even for very small values of I_2.

This phenomenon has been widely used to detect a weak signal in a radio frequency communication system and in an optical laser system; it is called the heterodyne detection method or homodyne detection method when the frequencies of the two beams are identical. The high intensity beam is a local oscillator in the case of radio frequency applications or a reference pump beam in the optical laser system. In x-ray crystallography, there exists a very important technique that is based on this phenomena; it is called the isomorphous replacement method [10]. One or a few heavy atoms attached to a large protein molecule provide sufficient change intensities of the diffraction spots, from which the phase of Bragg's diffraction spots can be determined. The effect of intensity change was estimated by Crick and Magdoff [11], and they derived the root mean square relation, which is basically the same as Eq. (4).

In this paper, I propose to place a heavy nanosized atom near a single biomolecule, so that diffraction will be amplified and its phase can be determined. If we count the number of photons directly, we will observe only a few 100 photons scattered from the single biomolecule. We need at least ten times more photons for biomolecular imaging. According to the basic concept of wave-particle duality in quantum mechanics, all matter exhibits both wavelike and particlelike behavior [12,13]. We can choose not to directly observe the x-ray photons scattered from the single biomolecule and allow them to propagate in free space as a wave and overlap with coherent and more intense waves from S_1. Since the resulting wave has a considerably large amplitude, a large

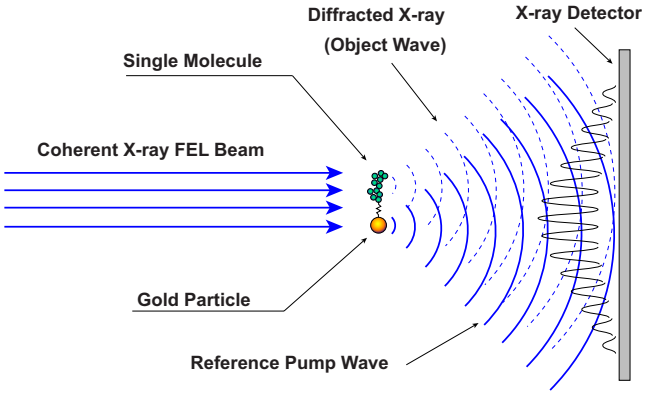

FIG. 3. (Color online) Holographic recording for single-molecule imaging.

number of photoionization events occur in the detector; thus, we can obtain more precise information about the biomolecule, which is recorded in the amplitude modulation of the interference fringes.

III. PROPOSED SETUP FOR HOLOGRAPHIC RECORDING

As shown in Fig. 3, we place a single biomolecule along with a small nanosized gold particle, which is used as the scattering source for the reference pump wave, in the path of an intense x-ray beam generated by an x-ray FEL. As discussed later, the size of the scattering source must be small for better separation of frequency components in the x-ray scattering data and the pump wave intensity must be high. Hence, we choose a heavy atom with a high-Z material since the form factor of atomic scattering scales as Z^2. The gold particle and single biomolecule correspond to the wide and narrow slits, respectively, in the double-slit experiment described previously. The scattered waves from these two objects interfere and form an interference pattern on the planar detector. There are various components in the interference pattern; among them, the simplest pattern is a periodic sine wave pattern associated with two-wave interference, whose width is given by

$$p = \frac{L}{D}\lambda, \tag{5}$$

where L is the distance between the objects and the detector; D, the distance between the single biomolecule and the gold particle; and λ, the x-ray wavelength. We assume that the planar detector (100 mm $\times$ 100 mm) with a resolution of 256×256 pixels is placed at a distance of 100 mm from the sample (the resolution at the rim corresponds to 2 Å for a wavelength of 1 Å). The interference fringe width should be 1 mm or more for the pattern to be clearly visible on the detector. From Eq. (5), we find that the distance between the two objects must be 10 nm or less, which is less than the coherent length of the x-ray beam emitted from the FEL.

Here, we consider a single lysozyme molecule as an example, which has a mass of 14.4 kDa. It is one of the standard enzymes, with a size of approximately $45 \times 30 \times 30$ Å. It mainly consists of carbon, oxygen, and nitrogen atoms. In order to observe its structure at the spatial resolution of around 2 Å, the number of Fourier components required to represent the two-dimensional (2D) projection image of this structure is roughly 50×50, where an over sampling factor of 2–3 is assumed. There must be at least 10 photon events for each Fourier component; thus, the total number of photons becomes $50 \times 50 \times 10 = 2.5 \times 10^4$. This is a fairly rough estimate for this small sample case. When the size of the sample becomes larger and more complex, a larger number of pixels and x-ray photons are required.

However, a previous study [9] has predicted the number of available photons to be approximately 100 after scattering from a single lysozyme molecule, which is considerably less than the number of photons required for imaging. We need 100–1000 times more flux in one shot.

Now, we roughly estimate the required number of atoms in the gold particle. The elastic scattering cross section is nearly proportional to the square of the atomic number Z. The scattering power decreases with the scattering angle and decreases rapidly for a low-Z material [14]; we neglect this effect for simplicity. The number of scattering photons under the same flux becomes

$$\frac{n_{\mathrm{p.Au}}}{n_{\mathrm{p.C}}} = \frac{N_{\mathrm{Au}}\sigma_{\mathrm{Au}}}{N_{\mathrm{C}}\sigma_{\mathrm{C}}} = \frac{N_{\mathrm{Au}}Z_{\mathrm{Au}}^2}{N_{\mathrm{C}}Z_{\mathrm{C}}^2}. \tag{6}$$

We approximate the lysozyme molecule as a 1000-atom carbon cluster, denoted by the subscript C. Using $Z_{\mathrm{C}}=6$, $N_{\mathrm{C}}=1000$, $Z_{\mathrm{Au}}=79$, and the same number of Au atoms as those in the lysozyme, i.e., $N_{\mathrm{Au}}=1000$, we find that the intensity ratio $n_{\mathrm{p.Au}}/n_{\mathrm{p.C}} \approx 200$. A gold particle in the metallic phase with a diameter of less than 3 nm contains 1000 atoms. Using the latest advanced gold-labeling technology [15], we can bind a nanosized gold particle to various types of biomolecules such as proteins, lipids, or adenosine triphosphate (ATP). Products such as Nanogold are manufactured using a well-established technology and are commercially available with diameters from 1.4 to 40 nm. The linking arm of the gold particle can be specifically made to react with thiols; thus, the location of the link to the biomolecule is well defined.

When an intense x-ray beam with a flux of 3×10^{12} photons per pulse in a 100-nm-diameter spot is incident on the gold particle linked to a single lysozyme molecule, as shown in Fig. 3, we obtain 2×10^4 photons into scattering angle for 2–30 Å resolution on the x-ray detector. In order to provide uniform illumination of the reference pump wave within the planar x-ray detector, a noncrystal structure will be suitable for the gold particle. Intensive research will be required to optimize its metallic structure (amorphous gold, quasicrystal gold, glassy alloy, etc.), but for simplicity, we assume a perfect gold particle with a random atomic distribution.

IV. FOUR-SLIT THROUGH EXPERIMENT

The experimental setup shown in Fig. 3 is identical to that used for Fourier transform holography or lensless holographic imaging [16,17]; only the reference wave used in these experiments is different.

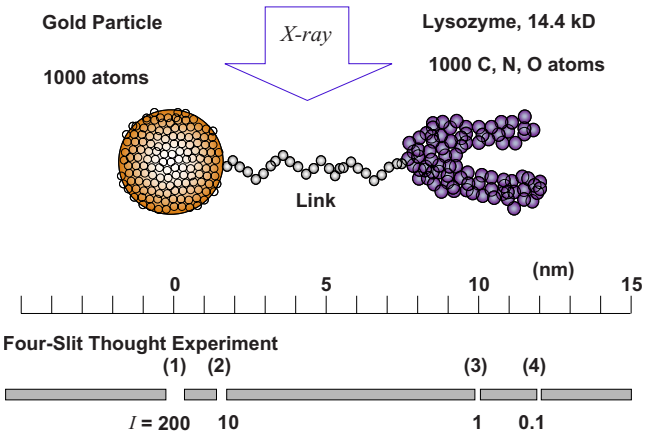

FIG. 4. (Color online) Conceptual diagram of a single lysozyme molecule linked to a gold particle (diameter of atoms is not drawn to scale). The gold particle produces 200 times more coherent x-ray scattering than the single lysozyme molecule. The bar at the bottom of the figure represents a four-slit thought experiment.

Fourier transform holography uses spherical waves as reference waves to record the phase of an object wave. The intensity of the reference wave is chosen such that it is comparable to that of the object wave to obtain the best contrast. To obtain better image quality, the size of the reference wave source should be considerably smaller than the object. The image-recovery process in Fourier transform holography is simple; basically, an inverse Fourier transform performed on the obtained intensity data provides the real image. This technique is well established, and Fourier transform holography has been successfully applied to the observation of magnetic nanostructures in thin films using soft x rays [18].

In the method, the intensity of the reference wave is considerably higher; as a result, the reference wave amplifies the weak signal and also provides a phase reference. However, the diameter of the gold particle is comparable to that of the single biomolecule and it is not a point source. Hence, the scattered x rays cannot be simple spherical waves. In fact, they form a speckle pattern, whose phase and amplitude are not known. Therefore, simple holographic image processing cannot be applied in this case.

To solve this problem, I have devised a thought experiment with each slit, in the double-slit experiment, being subdivided into two slits, i.e., there are four slits in total as shown at the bottom of Fig. 4. The gold particle, the internal structure or an imperfection of the gold particle, the sample, and the internal structure of the sample function as the first, second, third, and fourth slits, respectively. This is the simplified model of a gold-labeled molecule. When coherent light is incident, the probability of finding a photon on the screen is given by

$$I = |\psi|^2 = \psi^* \psi = \sum_{i=1}^{4} \psi_i^* \sum_{j=1}^{4} \psi_j = I_1 + 2\sqrt{I_1 I_2}\cos\Delta\phi_{12} + 2\sqrt{I_1 I_3}\cos\Delta\phi_{13} + 2\sqrt{I_1 I_4}\cos\Delta\phi_{14} + I_2 + 2\sqrt{I_2 I_3}\cos\Delta\phi_{23} + 2\sqrt{I_2 I_4}\cos\Delta\phi_{24} + I_3 + 2\sqrt{I_3 I_4}\cos\Delta\phi_{34} + I_4, \quad (7)$$

where $\Delta\phi_{ij} = \phi_j - \phi_i$ is the phase difference given by

$$\Delta\phi_{ij} = \phi_j - \phi_i = \frac{2\pi d_{ij}}{\lambda}\sin(2\theta). \quad (8)$$

Here, I_i represents the flux from the ith slit, 2θ is the scattering angle (as defined in crystallography), and d_{ij} is the distance between the i and jth slits. The slit locations and flux ratio suitable for our example case shown in Fig. 4 are as follows: $d_{12}=1.5$ nm, $d_{23}=8.5$ nm, $d_{34}=2$ nm, $I_1=200$, $I_2=10$, $I_3=1$, $I_4=0.1$, and $\lambda=1$ Å. Figure 5(a) shows the flux density distribution estimated using Eq. (7). The distribution is considerably complicated because the fringes are formed by the interference of four waves.

In Fig. 5(c), the curve at the bottom indicates the scattered wave from the biomolecule when it is directly observed without using the reference pump wave. It is a very weak signal with a relative intensity of approximately 1. In practice, the signal is quantized by the photon energy, leading to a loss of detailed information. The dashed curve (magnified 10 times) also shows an interference pattern, which represents the internal structure of the biomolecule; our aim is to study this pattern. By superposition of the reference pump wave, the signal wave is amplified, and the resulting interference pattern is recorded. In order to demonstrate the amplification effect clearly, the reference pump wave is assumed to be perfect with $\psi_2=0$. In Fig. 5(c), the curve at the top shows the amplified signal, which is recorded by the amplitude modulation of the interference pattern. From Eq. (7), we find that

$$I = I_1 + 2\sqrt{I_1 I_3}\cos\Delta\phi_{13} + 2\sqrt{I_1 I_4}\cos\Delta\phi_{14} + I_3 + 2\sqrt{I_3 I_4}\cos\Delta\phi_{34} + I_4. \quad (9)$$

In our experiment, the reference wave has a considerably higher intensity than that from the sample; $I_1 \gg I_3 > I_4$. Hence, we can eliminate the last three terms (direct wave) and Eq. (9) becomes

$$I \approx I_1 + 2\sqrt{I_1 I_3}\cos\Delta\phi_{13} + 2\sqrt{I_1 I_4}\cos\Delta\phi_{14}. \quad (10)$$

These two amplified signals have different frequencies, which results in the formation of a beat wave at $\Delta\phi_{14} - \Delta\phi_{13} = \Delta\phi_{34}$, whose frequency is equal to that of the interference signal of the direct sample signal. As discussed later, the sample signal is amplified and recorded as a beat wave. The signal gain is given by

$$G = \frac{\text{modulation in amplified signal}}{\text{modulation in direct signal}} = \frac{2(\sqrt{I_1 I_3} + \sqrt{I_1 I_4}) - 2(\sqrt{I_1 I_3} - \sqrt{I_1 I_4})}{4\sqrt{I_3 I_4}} = \frac{\sqrt{I_1 I_4}}{\sqrt{I_3 I_4}} = \sqrt{I_1/I_3}. \quad (11)$$

In this example, $G=\sqrt{200}=14.1$.

The signal to noise ratio S/N is significantly improved as follows. We assume that noise is caused by statistical fluctuations associated with the quantization process at photoionization. The improvement in S/N can be estimated as

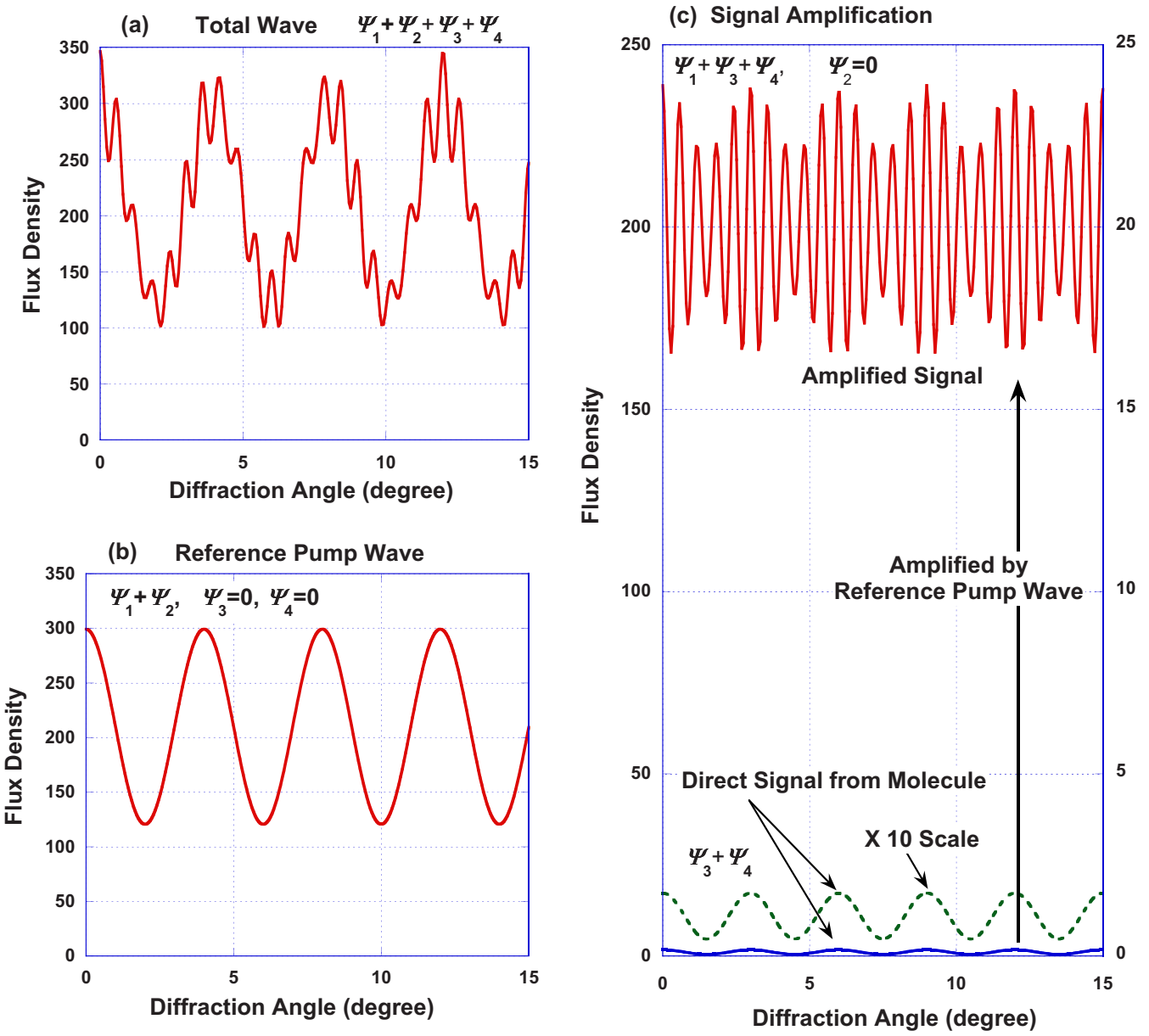

FIG. 5. (Color online) Four-slit through experiment

$$\frac{S/N|_{\text{holography}}}{S/N|_{\text{direct}}} = \frac{\sqrt{I_1 I_4}/(\sqrt{I_1}/I_1)}{\sqrt{I_3 I_4}/(\sqrt{I_3}/I_3)} = \frac{I_1}{I_3} = 200. \tag{12}$$

Therefore, a more accurate image can be obtained with this method. It should be noted that we do not detect the photons directly after they are scattered from the biomolecule; we detect them by their interference with the superimposed reference pump wave, which yields an amplified signal, resulting in considerably high power and lower statistical deviations associated with the quantization process.

Figure 5(b) shows the reference pump wave, which has an interference term of $\psi_1+\psi_2$. In practice, it forms a speckle pattern due to the random distribution of the gold atoms. The problem is that the amplified signal is influenced by the randomness of the speckle. In order to employ Fourier transform holography, we must determine the phase and amplitude of the reference pump wave.

In Eq. (7), under the condition $I_1 > I_2 \gg I_3 > I_4$, we can eliminate direct waves from the sample as follows:

$$I = I_1 + 2\sqrt{I_1 I_2}\cos\Delta\phi_{12} + 2\sqrt{I_1 I_3}\cos\Delta\phi_{13} + 2\sqrt{I_1 I_4}\cos\Delta\phi_{14} + I_2 + 2\sqrt{I_2 I_3}\cos\Delta\phi_{23} + 2\sqrt{I_2 I_4}\cos\Delta\phi_{24}. \tag{13}$$

In this equation, the information from the sample is contained in the interference patterns as $\psi_1\psi_3$, $\psi_1\psi_4$, $\psi_2\psi_3$, and $\psi_2\psi_4$.

In order to measure the spatial frequency components of the interference pattern, we define the angular frequency as follows:

$$\phi'_{ij} = \frac{\partial \Delta \phi_{ij}}{\partial \sin(2\theta)} \approx \frac{2\pi d_{ij}}{\lambda}. \tag{14}$$

The angular frequency is proportional to the distance d_{ij} or the diameter of the particle. If we multiply the angular frequency with the distance between the object and the planar detector, we obtain the spatial frequency on the detector. The maximum angular frequency of a speckle from the gold particle is determined by interference of two atoms at the left edge and right edge, i.e., $\phi'_{\mathrm{sp.max}} \approx 2\pi \times 2d_{12}/\lambda = 2\pi D_{\mathrm{Au}}/\lambda$, where D_{Au} is the gold particle diameter. If the gap between the linked structure and the gold particle is designed to be larger than the gold diameter, i.e., $d_{23} > D_{\mathrm{Au}}$, the angular frequency of the interference fringes of the signal wave exceeds the speckle angular frequency, i.e., $\phi'_{13}, \phi'_{14}, \phi'_{23}, \phi'_{24} \geqslant \phi'_{\mathrm{sp.max}}$. Therefore, if we use a low-pass spatial filter, we may be able to separate the speckle pattern from the high frequency components which contain information of sample structure.

V. ITERATIVE PHASE RETRIEVAL AND IMAGE RECONSTRUCTION

In coherent x-ray diffraction imaging or lensless x-ray diffraction microscopy, x-ray diffractions from all electrons in the sample create an interference pattern that is recorded as 2D diffraction intensity data, which usually forms a fairly complicated and random image similar to a speckle pattern. Recently, advancements in digital signal processing technology have made it possible to reconstruct the object image from such diffraction data. The method used for this purpose is called iterative phase retrieval or the oversampling method [19–23].

As described in the previous section, the signal from the single biomolecule is amplified on the intense reference pump wave and recorded as an interference pattern. By applying the iterative phase retrieval technique, we can recover the phase of the diffraction data, which includes the interference pattern between two objects (gold particle and single molecule), and we can reconstruct the images of the single molecule linked with the gold particle as one object.

The advantages and disadvantages of this method are discussed below.

(1) The benefit of using an intense reference pump wave is that the S/N ratio can be significantly improved, as shown in Eq. (12), which results in better image quality. We can use a high intensity reference pump wave, until it reaches the limit of the dynamic range in the x-ray detector.

(2) The ultimate resolution of diffraction imaging is defined by the maximum diffraction angle at the edge of the planar detector: $d \approx \lambda/2 \sin \theta_{\max}$. For example, using x rays with a wavelength of 1 Å and the maximum diffraction angle of $2\theta_{\max} = 30°$, the resolution limit is obtained as 2 Å. However, in the case of single molecular imaging, the number of available photons is quite limited; as a result the practical resolution reduces significantly due to background noise and large statistical fluctuations of the number of photons in each pixel. By introducing an intense pump wave, the S/N ratio can be significantly improved, as shown in Eq. (12), which results in a high practical resolution.

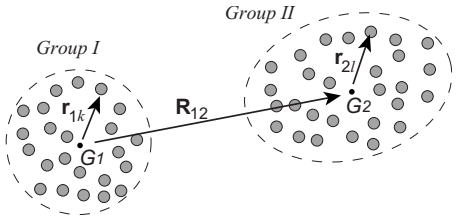

FIG. 6. Two group models of atoms. G_1 and G_2 are the mean centers of electron charge in each group.

(3) When the gold particle is introduced, the total size of the sample increases, which causes the maximum spatial frequency in the diffraction data to increase. To capture this image perfectly, we must use a finer pixel size in the planar detector, which means lower number of photons in each pixel. In the example shown in Fig. 4, the full size of the sample is about 4 times larger than that of the single lysozyme; therefore, we have to use $4 \times 4 = 16$ times more pixels, and the number of photons in each pixel becomes 16 times lower. This is a drawback of this method; however, it can be offset by a significant increase in the gain of S/N ratio, i.e., 200 times in the present example.

(4) In the case of a large molecule such as a single protein molecule or single viral capsid, the diffraction pattern becomes much more complicated and extends to a higher spatial frequency, which makes the iterative phase retrieval process slower and potentially unreliable.

(5) Even if the most advanced phase retrieval algorithm is used, there are a few percent level of discrepancies in the retrieved images [24]. Therefore, it will be difficult to study in detail the structure of a single biomolecule in the retrieved image, whose intensity is at least 100 times lower than that of the bright gold particle.

In order to solve this problem, I propose a scheme that combines iterative phase retrieval on low-pass filtered data and digital Fourier transform holography.

VI. HOLOGRAPHIC IMAGE RECONSTRUCTION

Here, we consider two groups of atoms, group I and group II, as shown in Fig. 6. Group I has a higher scattering power than group II. The resultant diffracted wave is given as follows:

$$\begin{aligned}\psi &= \sum_{k=1}^{N_1} \psi_{1k} + \sum_{l=1}^{N_2} \psi_{2l} = \sum_{k=1}^{N_1} f_k \exp[2\pi i \mathbf{r}_{1k} \cdot \mathbf{S}] \\ &\quad + \sum_{l=1}^{N_2} f_l \exp[2\pi i (\mathbf{R}_{12} + \mathbf{r}_{2l}) \cdot \mathbf{S}] \\ &= \sum_{k=1}^{N_1} f_k \exp[2\pi i \mathbf{r}_{1k} \cdot \mathbf{S}] \\ &\quad + \exp[2\pi i \mathbf{R}_{12} \cdot \mathbf{S}] \sum_{l=1}^{N_2} f_l \exp[2\pi i \mathbf{r}_{2l} \cdot \mathbf{S}] = R + O,\end{aligned} \tag{15}$$

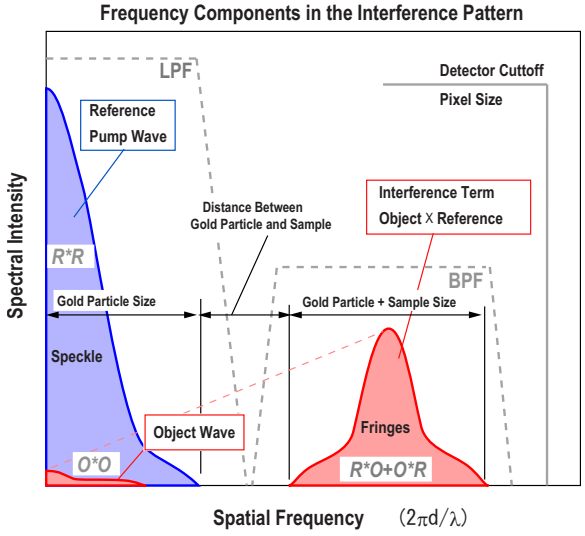

FIG. 7. (Color online) Spatial frequency components in the diffraction intensity pattern of Eq. (16).

where f is the atomic scattering factor and $\mathbf{S}$ is the scattering vector given by $|\mathbf{S}|=2\sin\theta/\lambda$. The origin of each group, G_1 and G_2, is defined as the center of the charge density. $\mathbf{r}_{1k}$ and $\mathbf{r}_{2l}$ represent the particle distance from the origins, and $\mathbf{R}_{12}$ is the distance between the origins. R and O denote the reference and object waves; here, we follow the same notation as that used in holography. The intensity of the diffracted wave becomes

$$\psi^*\psi = (R+O)^*(R+O) = R^*R + R^*O + O^*R + O^*O \approx R^*R + R^*O + O^*R. \quad (16)$$

We eliminate the last term because the contribution of the direct signal O^*O is considerably smaller as compared to other signals. As given in Eq. (15), the diffracted wave from group II has the phase term $\exp[2\pi i\mathbf{R}_{12}\cdot\mathbf{S}]$, which has considerably high spatial frequency components.

Figure 7 shows the spatial frequency components of Eq. (16). If there is a gap between the reference wave and the object wave, we can separate these signals using frequency filters. The condition for this is

$$g_{\text{link}} > D_{\text{Au}}, \quad (17)$$

where g_{link} is the gap between the right end of the gold particle and the left end of the sample. In the example shown in Fig. 4, the gap g_{link} must be longer than 3 nm.

If we use a band-pass filter (BPF) in Fig. 7, we can obtain the interference term R^*O+O^*R. However, it contains unknown phase terms R^* and R. To find the phase, we apply the iterative phase-retrieval method.

Initially, we use a low-pass filter (LPF in Fig. 7) and eliminate the high frequency components. If we select the turn over frequency of the LPF filter at the gap, we obtain the approximated reference wave as follows:

$$\psi^*\psi|_{filtered} \simeq \sum_{k=1}^{N_1}\psi_{1k}^*\sum_{k=1}^{N_1}\psi_{1k} = R^*R. \quad (18)$$

Then, we apply the iterative phase-retrieval method to R^*R data. A reference pump wave contains a large number of photons diffracted from the gold particle, and to some extent, it is an isolated object whose spatial frequency is much lower than the maximum sampling frequency as shown in Fig. 7. Therefore, the oversampling ratio increases significantly, and the iterative process will always converge and provide accurate solutions [24]. The error associated with the digital filter is discussed in the next section.

Once we determine the reference wave, the image recovery process is simple for a single biomolecule. From Eq. (16),

$$\frac{\psi^*\psi - R^*R}{R^*} = O + RO^*/R^*, \quad (19)$$

where the second term is a conjugate image, which will create a ghost image opposite to the gold particle and does not deteriorate the real image [26]. The subtraction $\psi^*\psi - R^*R$ plays an important roll in holographic image reconstruction, which erases the intense reference wave and the residual becomes only the interference term. Therefore, the inverse Fourier transform of Eq. (20) produces an accurate image of the single molecule. On the other hand, if the iterative phase retrieval technique is applied to both objects as one image, even a small phase error will cause image leakage from the bright gold particle, which will contaminate the single molecule image. The operator $1/R^*$ on the left-hand side of the equation applies amplitude normalization and phase correction to the diffraction pattern for a nonuniform reference wave distribution (correction of the speckle pattern formed from the gold particle).

To reconstruct the electron density map of the sample, we apply the inverse Fourier transform as follows [25]:

$$\rho(\mathbf{r}) = \int (\psi^*\psi - R^*R)/R^* \exp[-2\pi i\mathbf{r}\cdot\mathbf{S}]\cdot d\mathbf{S} \quad (20)$$

where $\rho(\mathbf{r})$ includes the atomic scattering factor f. The approximated 2D projection of the object is obtained as follows:

$$\rho(x,y) = \int_{S_y}\int_{S_x}(\psi^*\psi - R^*R)/R^* \times\exp[-2\pi i(xS_x + yS_y)]dS_x dS_y \quad (21)$$

We may use the BPF (Fig. 7) to select the object wave and reject the background noise [27,28] and parasitic diffractions coming from the upstream of the x-ray beam line.

In actual experiments, a large number of biomolecules with gold labeling must be injected into a beam line formed by a spraying technique and illuminated by an x-ray FEL. The orientation of the biomolecules will be random. The experiment is carried out as follows (see Fig. 8):

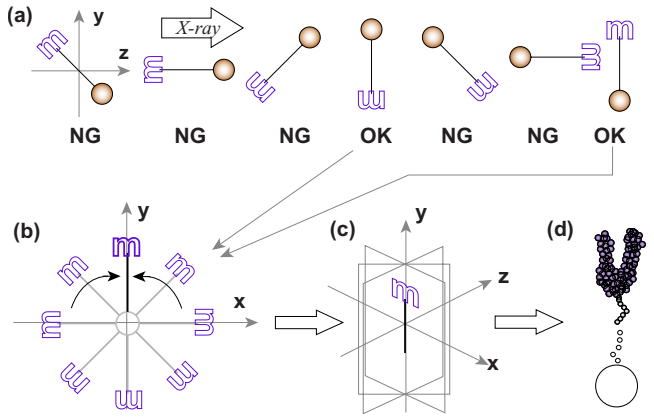

FIG. 8. (Color online) Creation of 3D structure from randomly oriented molecules. (a) Select diffraction images in which the gold particle and biomolecule lie in the *xy* plane, (b) rotate image on the *xy* plane to align on the *y* axis, (c) assemble many images according to the azimuth angle, (d) and create the 3D structure.

(1) Image selection: By applying the fast Fourier transform (FFT) to 2D data obtained from the planar detector, we obtain a peak corresponding to $\mathbf{R}_{12}\cdot\mathbf{S}=1$, from which we determine the axis vector $\mathbf{R}_{12}$. We must select the optimum diffraction images from a large number of images of randomly oriented molecules. The best image will have the largest $\mathbf{R}_{12}$ vector, whose length must be the same as that of the designed link structure of the gold particle, and the highest contrast. Such an image will be obtained when the gold label and biomolecule lie in a plane normal to the beam axis. A few degrees of angle error can be corrected in step (5) by tilting the image numerically.

(2) Retrieval of reference pump wave: The LPF filter (2D smoothing) is applied to the scattering x-ray data to eliminate the object waves; then, the iterative phase retrieval is carried out. Using the filtered scattering data as the amplitude and the retrieved phase, we compute the reference pump wave.

(3) Acquisition of 2D projection image of single molecule: A numerical operation is used to extract the object wave by Eq. (19), and the inverse FFT of Eq. (21) is applied to obtain the approximated 2D projected image.

(4) Rotation of axis: The retrieved images is rotated on *xy* plane and aligned to *y* axis. To accurately find the correct azimuthal rotation angle, an additional smaller gold label can be used as an angle marker.

(5) 3D structure reconstruction: By combining the electron densities given by Eq. (20) according to the azimuth angle, we obtain a 3D electron density map. Averaging over many data sets will improve resolution.

The hologram obtained by Eq. (19) contains depth information, i.e., it can represent 3D images as common holograms. We may tilt a single molecular image numerically [28], while the maximum rotation angle is restricted within a fraction of the solid angle of planar detector acceptance. However, this feature will enable us to easily assemble 3D structures from a large number of 2D projection images by rotation and interpolation.

VII. DISCUSSIONS

Once we resolve the reference wave from the diffraction data, we can apply the holographic imaging technique, which simplifies the assembly of the 3D structure, as discussed in the previous section. Here, I discuss an error associated with the digital filter used to obtain the reference wave. The low pass filter (LPF) is not perfect; thus, some signal of R^*O+O^*R will leak out and mix with R^*R, or the LPF will filter a portion of the high frequency part of R^*R. We may choose the cutoff frequency of LPF to be much closer to the interference signal R^*O+O^*R or within it, so that we do not loose the high frequency edge of the reference wave, which contains geometrical information of the nonspherical surface structure of the gold particle. In the iterative phase retrieval technique, it is known that iteration becomes unstable for a symmetric structure. Therefore, it is better to use a nonspherical structure, whose edge should not be rounded by the LPF. In Fig. 4, the gold particle is drawn as a spherical ball, but in practice, nanosized gold particles usually take various nonspherical geometries.

The leakage signal of R^*O+O^*R will mix with the reference wave and cause image degradation. By processing the numerical operation of Eq. (19), the left-hand side becomes

$$\psi^*\psi-\{R^*R+\delta(R^*O+O^*R)\}=(1-\delta)(R^*O+O^*R), \tag{22}$$

where $\delta(R^*O+O^*R)$ represents the leakage signal into the reference wave. Equation (22) indicates that the signal leakage of LPF causes a loss of the holographic image by a factor of $1-\delta$, which is not negligible at the lowest frequency part of R^*O+O^*R signal. It corresponds to the left end of the sample nearest to the gold particle. Therefore, the left end part of the single molecule becomes lower brightness in the reconstructed image. The denominator $1/R^*$ is also influenced by the signal leakage, while it does not cause substantial error on the reconstructed image because the phase is dominated by intense reference pump wave. For more detail discussion, it will be necessary to carry out careful study using numerical simulation on practical models.

It should be noted that if the gold particle has a crystal structure, the reference wave will be concentrated onto Bragg's diffraction spots. At other locations, the intensity becomes almost zero; thus, the numerical operation of $1/R^*$ on the left-hand side of Eq. (19) becomes unstable and introduces a substantial error. Therefore, we need to find the optimum structure of the gold particle, i.e., a random atom distribution, to provide uniform illumination. Intensive studies will be required to produce such a gold particle, and carry out numerical simulations that carefully take into account these details, and optimize the numerical procedures.

VIII. CONCLUSIONS

In this paper, I have discussed the potential of a scheme for determining the atomic structure of a single biomolecule using an intense coherent x-ray beam from an x-ray FEL. A gold particle is linked to biomolecular functions as a scattering source for a reference pump wave, which amplifies the

weak diffracted wave, and large flux x-ray photons are incident on the detector; the scattered photons carry sufficient structural information that can be recorded by a holography technique. To reconstruct an image from the x-ray diffraction data, I have developed a technique that combines iterative phase-retrieval and digital Fourier transform holography. This technique can be used to reconstruct the image of a single biomolecule. By using a combination of advanced gold labeling technology and intense x-ray pulses from x-ray FELs, we can study the structures of various proteins with a resolution close to few angstroms, without crystallization. Detailed studies will be required to optimize the gold particle structure (amorphous, quasicrystaline, glassy alloy, etc.), and determine the damage caused to the gold particle by radiation. A proof of principle experiment will be required when the x-ray FELs emitting 1 Å x-ray laser beam are ready.

[1] Z. Dauter, Acta Crystallogr., Sect. D: Biol. Crystallogr. **D62**, 1 (2006).

[2] C. Nave, Acta Crystallogr., Sect. D: Biol. Crystallogr. **D55**, 1663 (1999).

[3] K. Lundstrom, Cell. Mol. Life Sci. **63**, 2597 (2006).

[4] R. Bonifacio, C. Pellegrini, and L. M. Narducci, Opt. Commun. **50**, 373 (1984).

[5] W. Ackermann *et al.*, Nat. Photonics **1**, 336 (2007).

[6] J. Arthur *et al.*, Linac Coherent Light Source (LCLS) Conceptual Design Report No. SLAC-R593, Stanford, 2002 (unpublished).

[7] SCSS X-FEL Conceptual Design report, RIKEN Harima Institute, Hyogo, Japan, 2005, edited by T. Tanaka and T. Shintake (unpublished).

[8] XFEL: The European X-ray Free-electron Laser Technical Design Report No. DESY 2006-097, DESY, Hamburg, 2006, edited by M. Altarelli *et al.* (unpublished).

[9] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, Nature (London) **406**, 752 (2000).

[10] J. Drenth, *Principles of Protein X-ray Crystallography*, 2nd ed. (Springer, Berlin, 1994), Chap. 7.4, p. 145.

[11] F. H. C. Crick and Beatrice S. Magdoff, Acta Crystallogr. **9**, 901 (1956).

[12] C. Jönsson, Am. J. Phys. **42**, 4 (1974).

[13] A. Tonomura, J. Endo, T. Matsuda, T. Kawasaki, and H. Ezawa, Am. J. Phys. **57**, 117 (1989).

[14] J. Als-Nielsen and D. McMorrow, *Element of Modern X-ray Physics*, 1st ed. (Wiley, New York, 2001), Vol. 1, Chap. 4.2, p. 112.

[15] J. F. Hainfeld and R. D. Powell, J. Histochem. Cytochem. **48**, 471 (2000).

[16] J. T. Winthrop and C. R. Worthington, Phys. Lett. **15**, 124 (1965).

[17] W. S. Haddad, D. Cullen, J. C. Solem, J. W. Longworth, A. McPherson, K. Boyer, and C. K. Rhodes, Appl. Opt. **31**, 4973 (1992).

[18] S. Eisebitt, J. Luning, W. F. Schlotter, M. Lorgen, O. Hellwig, W. Eberhardt, and J. Stohr, Nature (London) **432**, 885 (2004).

[19] D. Sayre, Acta Crystallogr. **5**, 843 (1952).

[20] R. J. Fienup, Appl. Opt. **21**, 2758 (1982).

[21] D. Sayre, H. N. Chapman, and J. Miao, Acta Crystallogr., Sect. A: Found. Crystallogr. **A54**, 232 (1998).

[22] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, Nature (London) **400**, 342 (1999).

[23] H. N. Chapman *et al.*, Nat. Phys. **2**, 839 (2006).

[24] C. Song, D. Ramunno-Johnson, Y. Nishino, Y. Kohmura, T. Ishikawa, C. C. Chen, T. K. Lee, and J. Miao, Phys. Rev. B **75**, 012102 (2007).

[25] J. Drenth, *Principles of Protein X-ray Crystallography*, [10], Chap. 4.10, p. 93.

[26] P. Hariharan, *Basics of Holography*, 1st ed. (Cambridge University Press, Cambridge, 2002), Chap. 1.5, p. 11.

[27] F. Charrière, J. Kühn, T. Colomb, F. Monfort, E. Cuche, Y. Emery, K. Weible, P. Marquet, and Ch. Depeursinge, Appl. Opt. **45**, 829 (2006).

[28] U. Schnars and Werner Jueptner, *Digital Holography*, 1st ed. (Springer, Berlin, 2005), Chap. 3.3, p. 56.