Scanning Microscopy Vol. 11, 1997 (Pages 391-405) 0891-7035/97$5.00+.25
Scanning Microscopy International, Chicago (AMF O'Hare), IL 60666 USA

PRINCIPLES AND APPLICATIONS OF DEFOCUS-IMAGE MODULATION PROCESSING ELECTRON MICROSCOPY

Yoshizo Takai[1]*, Toshiyuki Ando[1], Takashi Ikuta[2] and Ryuichi Shimizu[1]

[1]Department of Applied Physics, Osaka University, Suita, Osaka
[2]Osaka Electro-Communication University, 18-8 Hatsu-machi, Neyagawa, Osaka, Japan

Abstract

Principles of defocus-image modulation processing (DIMP) are described from a view point of three-dimensional image formation in transmission electron microscopy (TEM). Two types of defocus-image modulation processing electron microscopes (DIMP-EM) have been constructed for realizing spherical aberration-free observation. They are based on changing the objective lens current and modulating the acceleration voltage for defocussing, respectively. The former one has an advantage of practical use that the DIMP-system developed can be easily applied to a commercial type electron microscope without any modifications. The second one enables a real-time image processing to be realized by controlling the irradiation time by the primary electron beam during recording each defocused image. The performance of the two DIMP-EMs are reported and discussed with some new experimental results.

Key Words: High resolution transmission electron microscope, defocus-image modulation processing (DIMP), spherical aberration-free observation, phase electron microscopy, real-time correction of spherical aberration.

*Address for correspondence:
Yoshizo Takai
Department of Applied Physics,
Faculty of Engineering, Osaka University,
2-1 Yamada-oka, Suita, Osaka 565, Japan
Telephone number: 81-6-879-7842
FAX number: 81-6-879-7843
E-mail: takai@ap.eng.osaka-u.ac.jp

Introduction

It is well known that magnetic lenses for electron optical systems inevitably have spherical aberration, which is the main limitation on the resolution of transmission electron microscopes (TEM). The spherical aberration causes blurring in TEM images especially at the areas near a sample edge, an interface or a defect. The blurring severely complicates obtaining structure information directly from the observed images. For obtaining atomic structure information, therefore, the images of a sample observed at different focus conditions have been compared with the corresponding images calculated using a model structure at the same focus conditions. One-by-one comparisons between the observed images and calculated images allow one to conclude whether the model structure is valid or not. However, it takes a long time and, furthermore, a priori knowledge about the structure is required in the through-focussing technique, which restricts wider applications of high resolution electron microscopy to materials science. Since the spherical aberration-free images of phase and amplitude components have much higher potential for determining atomic structures directly from the images, various kinds of approaches have been proposed thus far for the aberration correction or the phase retrieval problem (Schiske, 1968, 1973; Hoppe *et al*., 1973; Saxton, 1978; Hawkes, 1980). The development of a field emission gun, high performances of charge-coupled device (CCD) cameras and computer systems recently make it possible to employ some novel spherical aberration correction methods which are applicable to high resolution images, such as electron holography (Lichte *et al*., 1992), restoration from focus series of images (Ikuta, 1985, 1989; Van Dyck and Op de Beeck, 1990; Coene *et al*., 1992) and tilt series of images (Takai *et al*., 1994a, Kirkland *et al*., 1995).

We have developed the defocus-image modulation processing (DIMP) method for spherical aberration-free observation of phase and amplitude images (Ikuta, 1989; Taniguchi *et al*., 1990a,b, 1992a,b). The method is based on the procedure of multiplying a specific weighting function, which represents the characteristics of the electron optical system, to a series of defocus images and taking a summation of these weighted images to get one final image

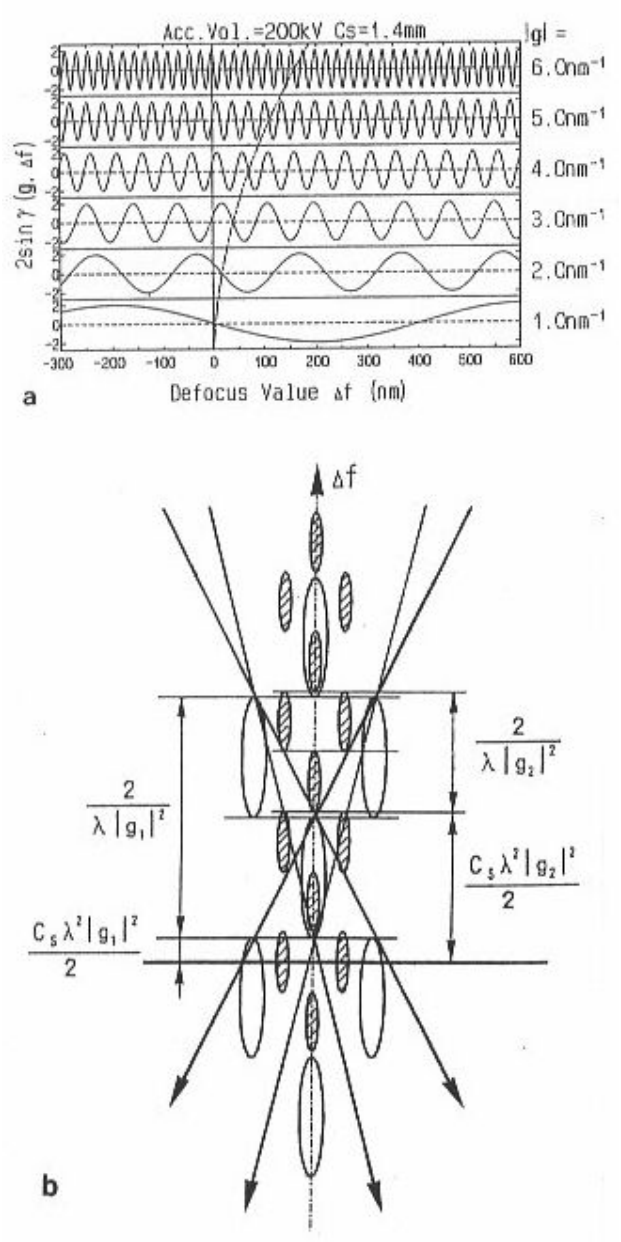
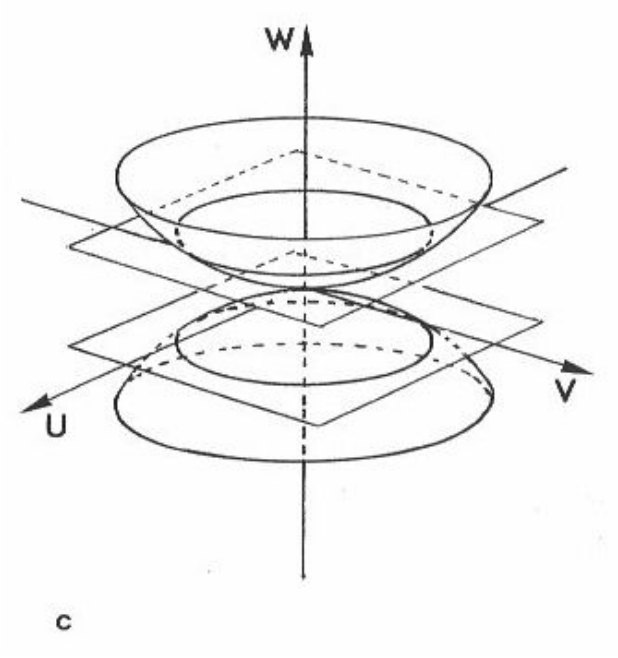


Figure 1. (**a**) Phase contrast transfer functions and (**b**) schematic illustration of Fourier image appearing around Gaussian image plane. Note that the oscillation of contrast transfer function produces periodic contrast change of Fourier images. (**c**) A pair of Ewald spheres formed by linear imaging term and two planes formed by a sinusoidal weighting function in three-dimensional Fourier space.



that provides, in principle, the spherical aberration-free image. In the method, spherical aberration correction and restoration of phase and amplitude images are achieved at the same time without any additional optical devices, such as an electron biprism. Van Dyck *et al.* (1990) have also proposed a method called the focus variation method or the paraboloid method, which is an expanded method of the Schiske method. In the focus variation method, plural images (about 20 images) are used to recover the complex wave function. There are a lot of similarities between the two methods, but also several differences. In the focus variation method, the convolution integral between the Fourier transforms of the through-focus images and the phase correction factors is performed in reciprocal space. According to the mathematical derivations, the convolution integral works as a spatial filter to extract only the information located on a Ewald sphere in three dimensional Fourier space. Therefore, the non-linear image contributions can be effectively reduced by increasing the number of images (Saxton, 1994). In the DIMP method, on the other hand, there is no such filtering effect in the Fourier space. Therefore, the non-linear components may produce some artifacts in the phase and amplitude images when the weakly scattering object approximation is not well satisfied to the sample. However, the convolution integral is done between through focus images and the weighting function in real space, which becomes the largest advantage in the DIMP. Because no Fourier transformation is necessary, which is essentially important for real-time processing. By utilizing the characteristic that all of the processing are done in real space, we have further proposed another DIMP method for achieving faster processing. In the second DIMP, the spherical aberration-free observation is performed within several video frame periods by controlling the irradiation time of primary electron beam during recording each

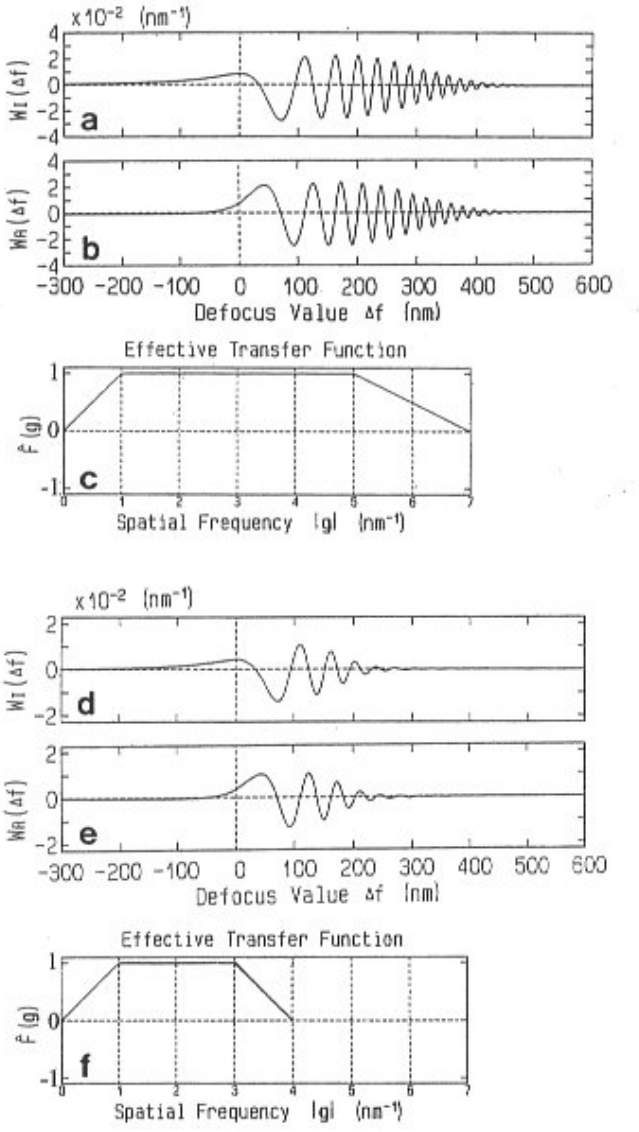


Figure 2. Weighting functions for spherical aberration-free (**a**) phase image and (**b**) amplitude image when an effective contrast transfer function (**c**) was assumed. (**d**), (**e**) are the weighting functions for (**f**) different effective contrast transfer function.

defocused image. The weighted image integration is done in the image detection device itself by the control. Since the second DIMP can keep the entire electron dose to the specimen as small as possible, it is one of the promising methods which are applicable to the spherical aberration-free observation of electron radiation sensitive specimens, such as biological samples.

In the present paper, we attempt to clarify the principles of DIMP using some new experimental results. Two types of DIMP-electron microscopes (EM) constructed for realizing spherical aberration-free observation are summarized with their performance and present restrictions.

Principles of Spherical Aberration Correction by DIMP

Extraction of one specified Fourier component from through-focus images

TEM image contrast of a spatial frequency component $|g_o|$ periodically changes by every focus change of $2/(|g_o|^2\lambda)$, as shown by sinusoidal oscillation of phase contrast transfer functions in Figure 1a. In the figure, phase contrast transfer functions for different spatial frequencies are shown as a function of focus, not a function of spatial frequency as usual. Figure 1a indicates that Fourier images appear periodically with different focus intervals depending on its spatial frequency in a three-dimensional image space, where the focus axis (z-axis) is regarded as the third axis just as shown schematically in Figure 1b. It is also found that the effect of the spherical aberration is given as a phase shift of each simple sinusoidal waveform, which is indicated by the dotted line in Figure 1a. It is known that the narrow band pass filtering can be achieved for one-dimensional temporally altered signals by using the lock-in amplifier under the referenced signal of the same frequency. In the case of the two-dimensional image observation under the coherent illumination, the sinusoidally weighted image-integration over the focus range is compared to the phase sensitive detection of the lock-in amplifier. That is, by integrating the intensity of through-focus images multiplying the contrast transfer function with $|g_o|$ as a weighting function, only the Fourier component $|g_o|$ can be extracted. The effect of the spherical aberration is then corrected by considering the origin shift of the weighting function. Mathematical derivations have been shown in detail by Taniguchi *et al*. (1994). This convolution integral between through-focus images and the weighting function in three-dimensional real space is equivalent to a spatial frequency filtering in three-dimensional reciprocal space, because the Fourier transform of the integral is expressed by a product of Fourier transforms of the three-dimensional image intensity and the sinusoidal weighting function, respectively. The Fourier transform of the image intensity for the linear-imaging component is known to form a pair of Ewald spheres (Taniguchi *et al*., 1991; Van Dyck, 1990). The latter Fourier transform becomes a delta function which works as a spatial frequency filter, enabling to extract one Fourier component in reciprocal space at the two planes of $w= \pm(|g_o|^2\lambda)/2$. Figure 1c shows the filtering process illustrated schematically in three-dimensional reciprocal space, consisting of a w-axis for the focus value in addition to the common u and v axes. Only the Fourier components of the intersection region between

the spheres and the planes are extracted by the integral. The two cross-section rings are then projected to a two-dimensional reciprocal plane by performing an inverse Fourier transformation with respect to the w-axis. The projected ring pattern in the two-dimensional reciprocal plane indicates that one Fourier component was successfully extracted by the image integration of the through-focus images weighted by the sinusoidal function, which will be shown in the experiment.

Weighting functions for spherical aberration correction

The DIMP method extends this extraction processing so as to cover all frequency components. Since the extraction operations for different spatial frequency components are independent from one another, a final weighting function $W_i(\Delta f)$ for extracting all spatial frequency components in the phase image is obtained by the following integral of the phase contrast transfer function (Taniguchi *et al*., 1994):

$$W_i(\Delta f) = \int_0^\infty F(g)\sin(\gamma)d(\lambda|g|^2/2) \qquad (1)$$

where F(g) is the effective transfer function to be realized after the correction. Figure 2a shows the final weighting function for a 200 kV electron microscope with spherical aberration coefficient of 1.4 mm when the effective transfer function shown in Figure 2c is assumed. The weighting function is a unique function for an electron microscope when once an effective transfer function is assumed. For extracting the amplitude component, on the other hand, the amplitude contrast transfer function cos(γ) should be used instead of sin(γ) in Equation (1), leading to another weighting function $W_r(\Delta f)$ in Figure 2b. Figures 2d and 2e show another set of weighting functions that are obtained when the effective transfer function is changed as in Figure 2f. From these figures, the weighting function proves to be a function that oscillates slowly around Scherzer focus and quickly at larger under-focus conditions. When a wider effective contrast transfer function is assumed, the weighting function oscillates in wider focus range and especially shows a quicker oscillation at larger under-focus. Since the achievable highest resolution by the method is restricted by the finest structure present in the images, it does not make sense to assume a very wide effective contrast transfer function in order to get higher resolution beyond the information limit. Smaller spherical aberration coefficient makes the weighting function simpler in a narrower focus range, therefore, a superior lens with small C_s is of course effective for practical use of this technique.

Spherical aberration-free images

Spherical aberration-free images $I_i(\mathbf{r})$ and $I_r(\mathbf{r})$ for the phase component and amplitude component, therefore, can

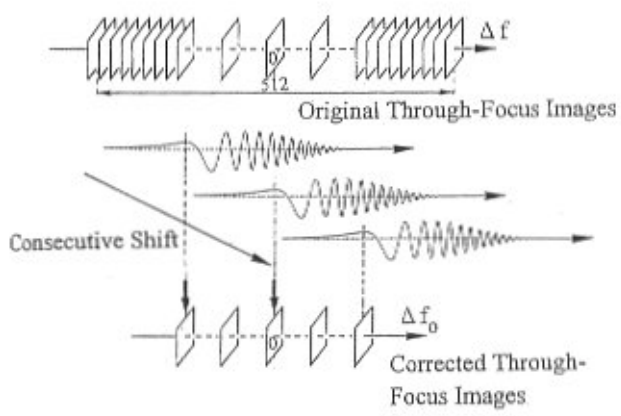


Figure 3. Schematic illustration showing procedures for calculating spherical aberration-free images by the first DIMP.

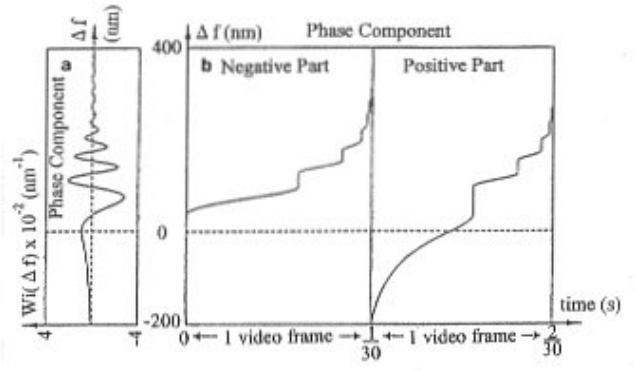


Figure 4. (**a**) Weighting function for the first DIMP and (**b**) the corresponding cumulative weighting function for the second DIMP.

be obtained by integrating through-focus images $I(\mathbf{r}, \Delta f)$ multiplied with the final weighting functions $W_i(\Delta f)$ and $W_r(\Delta f)$, respectively, that is,

$$I_i(\mathrm{r}) = \int_{-\infty}^{\infty} I(\mathrm{r},\Delta f)\, W_i(\Delta f)\, d\Delta f$$
$$I_r(\mathrm{r}) = \int_{-\infty}^{\infty} I(\mathrm{r},\Delta f)\, W_r(\Delta f)\, d\Delta f \qquad (2)$$

t should be noted here that separate imaging of phase and amplitude components is possible by using respective weighting functions, because an orthogonal relation is satisfied between the two weighting functions. Through-

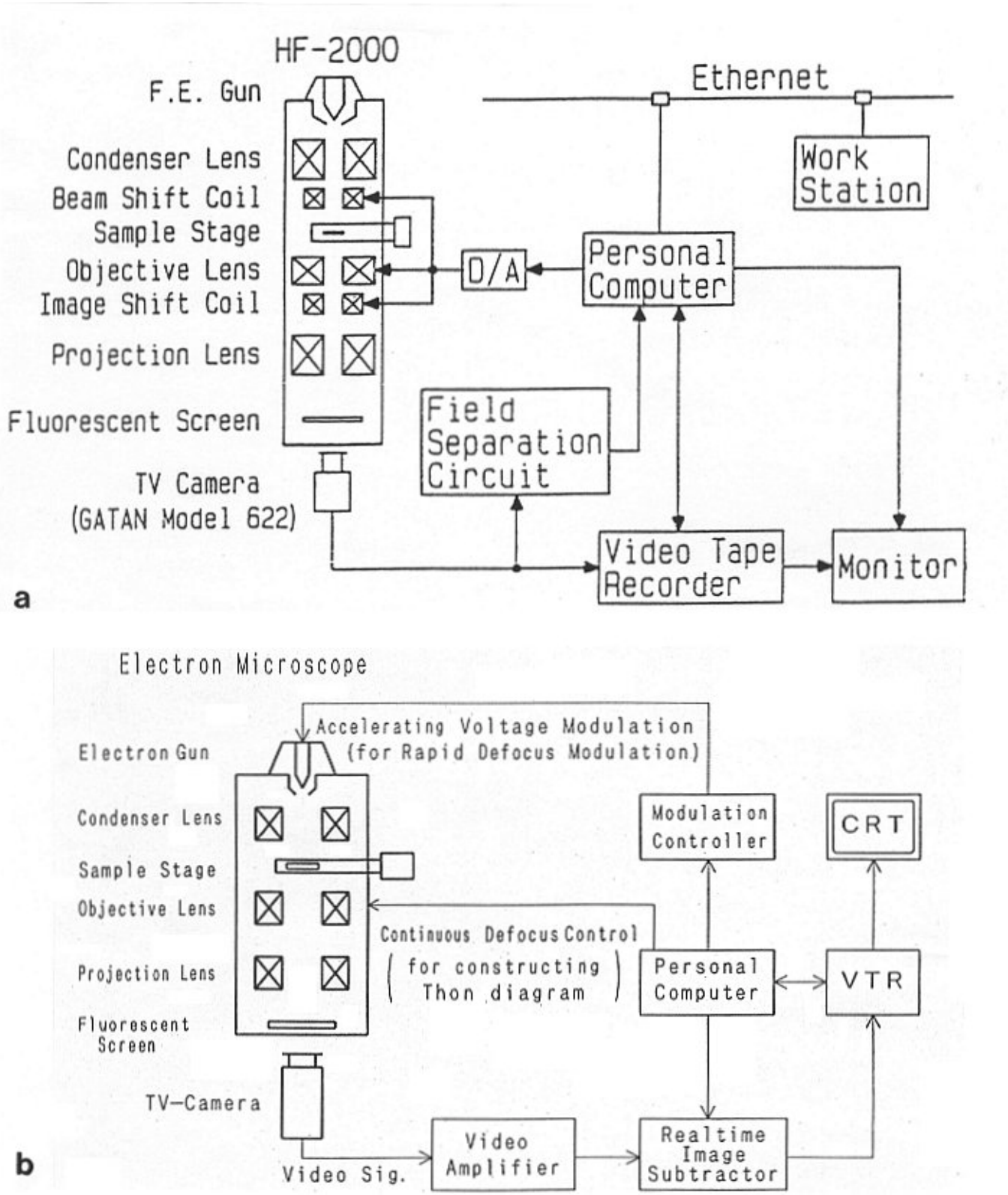


Figure 5. Schematic block diagrams of (**a**) the first DIMP system and (**b**) the second DIMP system.

focus images $I_{i,r}(\mathbf{r}, \Delta fo)$, which are corrected for spherical aberration, can be obtained by using $W_{i,r}(\Delta f+\Delta fo)$ instead of $W_{i,r}(\Delta f)$ in Equation (2), where Δfo is a virtual defocus value for the corrected images. This is because the phase contrast transfer functions in Figure 1 shift their origin positions by a constant value when the focus is changed, independent of the value of spatial frequency. Figure 3 shows the schematic procedure for getting spherical aberration-free images by the first DIMP. By using digitized original through-focus images, the image integration is performed by Equation (2), resulting in spherical aberration-free phase or amplitude image. Through-focus images that are corrected for spherical aberration can be obtained by consecutive shifts of the weighting function as shown in the figure.

Real-time correction of spherical aberration

In the first DIMP, 256 images recorded by a TV camera are generally used for getting the spherical aberration-free images. Therefore, several seconds are necessary at least for recording the images by the TV camera and three hours for digitizing the images and performing the calculation of the integration. Such long exposure time and processing time disturbs us from developing a real-time image processing system. Thus we have proposed the second DIMP, where the processing is performed within a few video frame periods by rapid modulation of focussing. It seems to be one promising way for spherical aberration-free observation of biological specimens, because we can keep the total electron dose to the specimen as small as possible.

The second DIMP-EM uses irradiation-time control of the primary electron beam (Ando *et al*., 1994a,b) for realizing the weighting function in Figure 2. Since a TV camera uses charge-storage-type image detection elements and integrates the image signals over the video frame period (1/30 sec), irradiation-time control enables us to perform the weighted image integration of through-focus images in the detection elements in real-time. Using image signals per unit time, $i(r,\Delta f)$, and the frame time, τ, the phase image in Equation (2) is then rewritten as

$$I_i(\mathrm{r}) = \int_{-\infty}^{\infty} i(\mathrm{r}, \Delta f)[\tau W_i(\Delta f)]\, d\Delta f \qquad (3)$$

Next, consider that one changes the focus within a video frame period, τ. Then, the image signal which is integrated within the video frame period is obtained as,

$$\int_{0}^{\tau} i(\mathrm{r}, \Delta f\ (t))\, dt = \int_{-\infty}^{\infty} i(\mathrm{r}, \Delta f\ (t))\, |\, dt/d\Delta f\, |\, d\Delta f \qquad (4)$$

where the term of $|dt/d\Delta f|$ is defined as the irradiation time of the primary electron beam per unit defocus-value. We control here the irradiation time so as to satisfy the following equation,

$$|dt/d\Delta f| = \tau\, |W_i(\Delta f)| \qquad (5)$$

The weighting function $W_i(\Delta f)$ is a bipolar function having positive and negative values and the irradiation time is always positive. Therefore, the subtracted image between two integrated images for positive and negative values of the weighting function becomes equivalent to the phase image by the first DIMP, that is,

$$I_i(\mathrm{r}) = \int_{-\infty}^{\infty} i(\mathrm{r}, \Delta f\ (t))[\tau\, |\, W_i(\Delta f)\, |\,]_{\text{positive}}\, d\Delta f - \int_{-\infty}^{\infty} i(\mathrm{r}, \Delta f\ (t))[\tau\, |\, W_i(\Delta f)\, |\,]_{\text{negative}}\, d\Delta f \qquad (6)$$

Figure 4a and 4b show a weighting function for the phase component and a cumulative function of the weighting function for the irradiation-time control. The t vs. Δf curve shown in Figure 4b enables us to realize the defocusing as a function of irradiation time, i.e., $\Delta f(t)$. As shown in the figure, longer irradiation times are used to realize larger values of the corresponding weighting function. In order to flip the signs of the different zones of the weighting function, two complete scans are used as shown in Figure 4b, one for weighting the positive images and one for the negative images. Two different video frames that are obtained by the respective scans are then subtracted as formularized in Equation (6) and final processed images are displayed on a CRT monitor in every two video frames, in principle. In this way, one can perform the integrations of Equation (2) in a very short period. Thus, we can operate the electron microscope and observe phase or amplitude images without awareness of image distortion due to the spherical aberration. The most important technical subject is how the rapid and precise modulation of focussing is achieved.

Experimental Apparatus

Figure 5a shows a block diagram of the first DIMP-EM. A field emission electron microscope HF-2000 (Hitachi, Ibaragi, Japan) with a TV camera (Gatan model 622; Gatan, Pleasanton, CA) was used. First, through-focus images with a focus step of 2.28 nm were recorded on a video tape by controlling objective lens current using a personal computer. Second, 256 images were transferred to a HP-9000 work station (Hewlett-Packard, Palo Alto, CA) and then the integration of the digi-tized images was performed according to Equation (2).

The defocus value of each image and the origin of focus were determined using original through-focus images

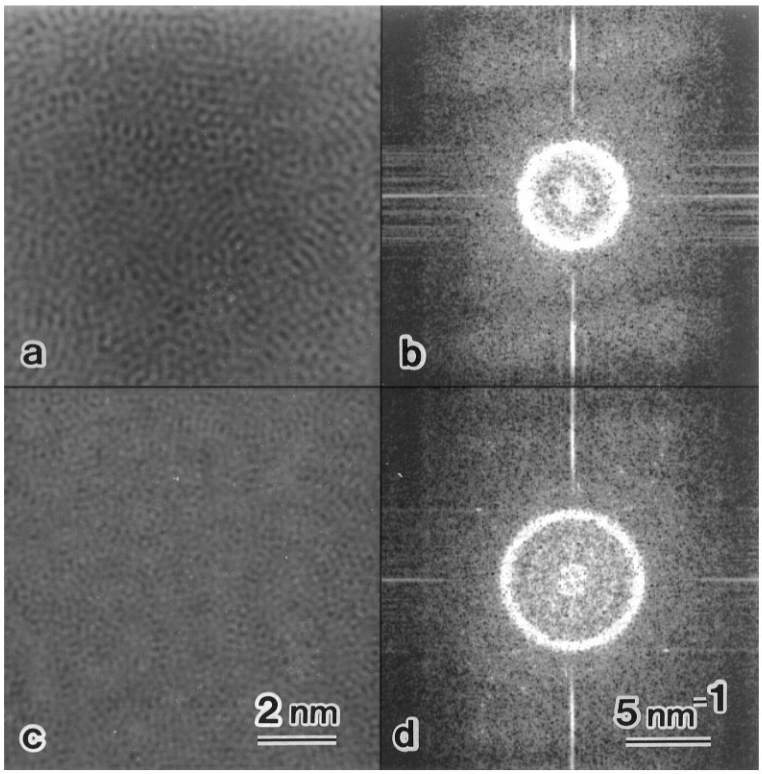


Figure 6. (**a**) Experimental confirmation of extraction of one Fourier component; (**b**) processed image and power spectrum for 3 nm^{-1} and (**c**), (**d**) for 4 nm^{-1}.

of an amorphous carbon thin film. From the experimental point of view, compensation of image drift among the original images is the most important technical factor for obtaining aberration-free images correctly. For this, image shift coils were synchronously controlled while recording images, and three-dimensional analysis of the image data set was performed for fine tuning of image drift with a constant speed (Taniguchi, 1991). Cross-correlation technique was sometimes used as the finest tuning method for irregular variable speed image drift.

Figure 5b shows a block diagram of the second DIMP-EM system. A JEM-200CX (JEOL, Tokyo, Japan) with the same TV camera was used operating at an accelerating voltage of 160 kV. Through-focussing was performed by accelerating voltage modulation instead of objective lens current modulation for rapid scan of focussing within a video frame, because the hysteresis of the objective lens interferes with rapid scan of focussing. The driving signal for the modulation was simply supplied to the feedback circuit of the high-voltage stabilizing unit of the TEM in the present experiment. The processed images were displayed on a CRT at a rate of 4/30 sec, because two video frames are necessary to get respective irradiation time controlled images for positive and negative parts due to the restrictions of the charge storage-type image detection and reading system using the TV camera (Ando *et al*., 1995a).

Experimental Results and Discussions

Extraction of one Fourier component

Figures 6a and 6b show experimental results of the extraction processing for 3 nm^{-1} and 4 nm^{-1} Fourier components from the same series of through focus ima-ges of an amorphous carbon thin film. The phase contrast transfer functions corresponding to g=3 nm^{-1} and 4 nm^{-1} in Figure 1 were used as weighting functions in the convolution integral, respectively. Their power spectra show different sizes of the diffraction ring, which demonstrates that one Fourier component was successfully extracted in each case by the weighted image integration. The effect of spherical aberration has been corrected in the images by considering the origin shift of the phase contrast transfer functions when the convolution integral is performed.

Confirmation of spherical aberration correction by DIMPs

Figures 7a and 7b show the Thon diagram constructed from the original through-focus images and the corrected through-focus images obtained by the first DIMP.

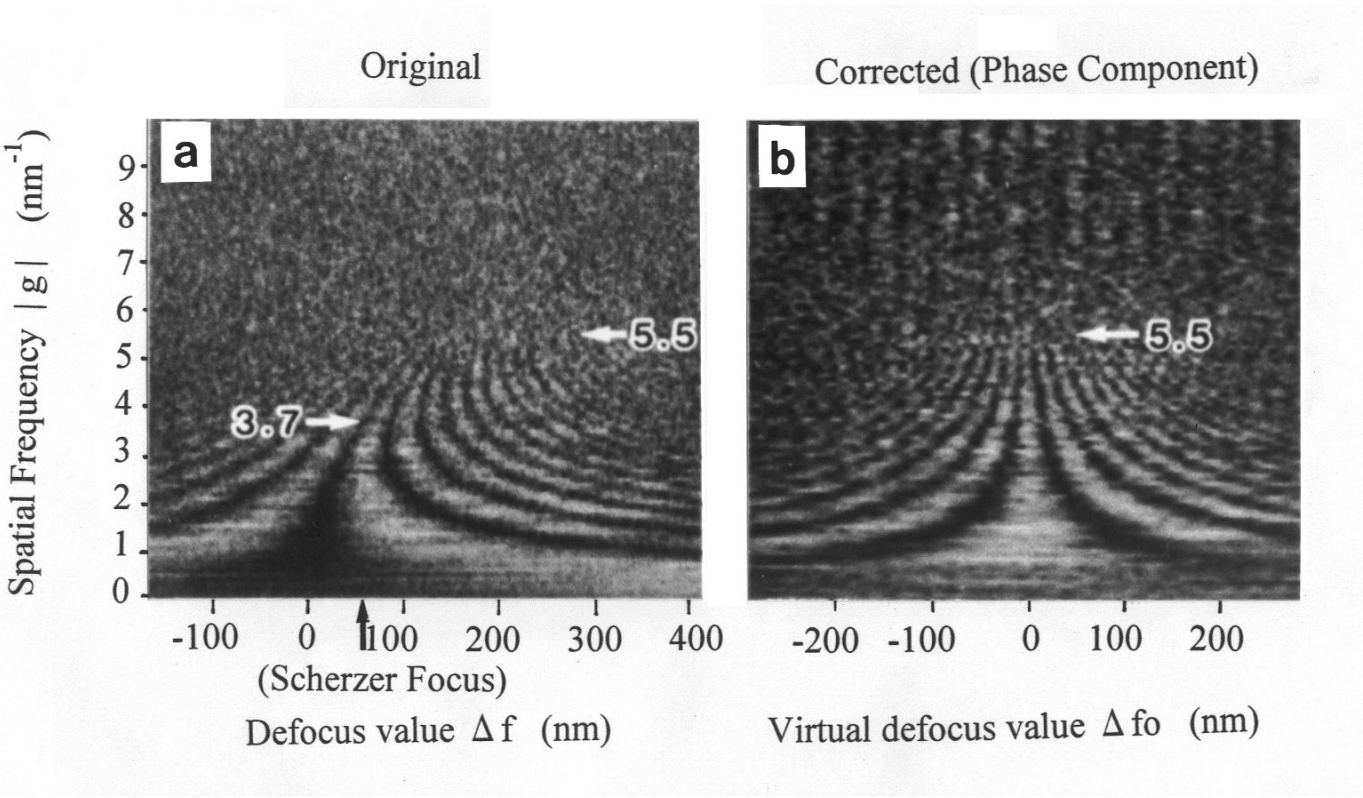


Figure 7. Thon diagrams constructed from (**a**) the original through-focus images and (**b**) corrected images for phase component by the first DIMP. Note that symmetric pattern in (**b**) confirms the correction of the spherical aberration up to 5.5 nm^{-1}.

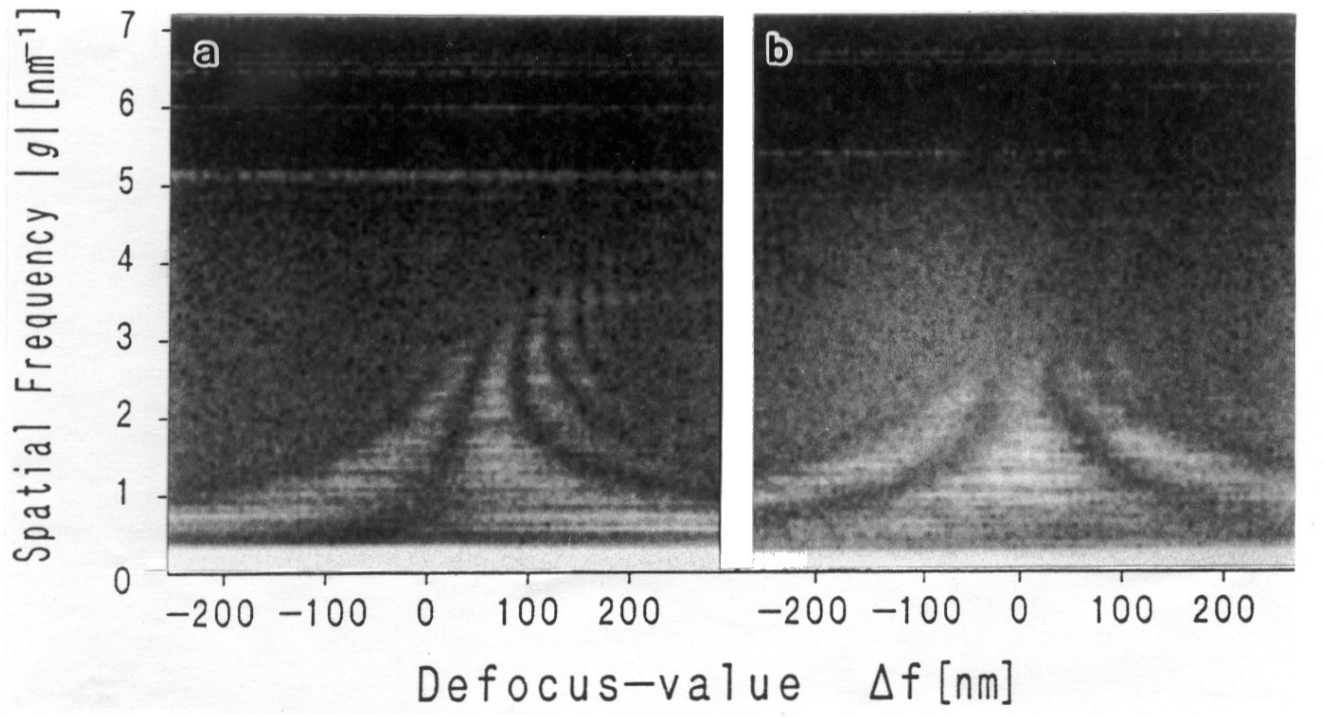


Figure 8. Thon diagrams constructed from (**a**) the original through-focus images and (**b**) the corrected through-focus images obtained in real-time by the second DIMP.

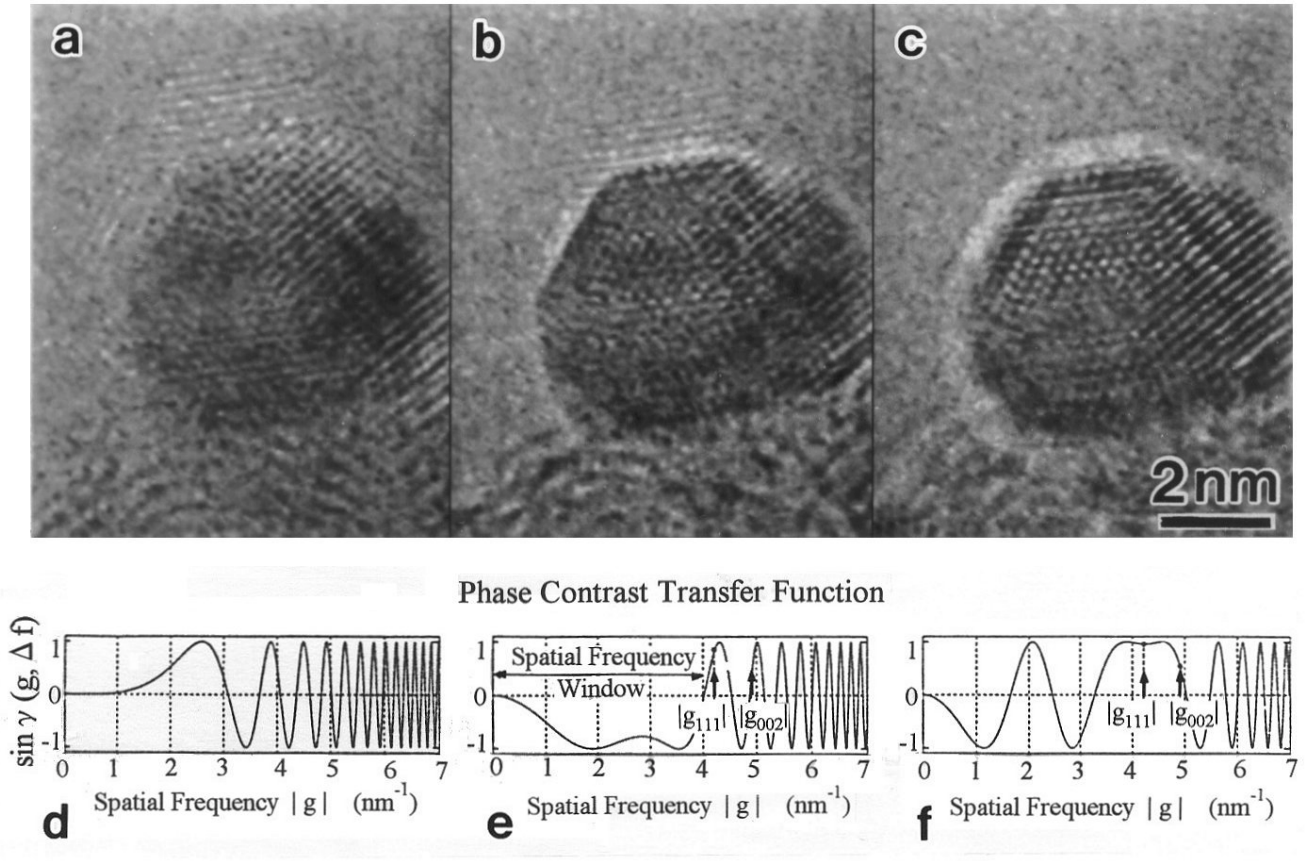


Figure 9. Original through-focus images of a fine gold particle and phase contrast transfer functions. (**a**), (**d**) Δf=0.0 nm; (**b**), (**e**) Δf=1.2 $C_s \lambda$ = 71.1 nm (Scherzer focus) and (**c**), (**f**) Δf = $C_s |g_{111}|^2 \lambda^2$ = 158.8 nm (optimum focus for Au(111)).

In the diagram, the radial intensity distributions of the power spectra of the images are plotted as a function of focus. The diagram in Figure 7b shows a symmetric pattern about the line Δf=0, confirming that the spherical aberration is completely corrected, while an asymmetric pattern is seen in Figure 7a. A straight bright line at Δf=0 in Figure 7b confirms that the phase contrast is observed under in-focus condition. By correcting spherical aberration, the resolution of TEM was successfully improved from Scherzer resolution limit to the information limit, which is the attainable highest resolution.

Figures 8a and 8b show same data sets as those in Figure 7, which were obtained by the accelerating voltage modulation of the second DIMP system. Since the present system has been based on the use of a commercial type TEM without any modification of its basic performance, the application has been restricted to spatial frequencies in the intermediate region (<3.0nm^{-1}) due to the limited response of the acceleration voltage modulation. The resolution will be improved up to the information limit when precise modulation of the accelerating voltage is achieved. We are now developing a modified system where the modulation of the accelerating voltage is achieved by transferring the modulation signals from a function generator through optical pipes to the high voltage power supply directly. Some new results will be reported elsewhere.

The 1st-DIMP performs weighted integration numerically by a computer for the defocused images which were recorded while changing the focus at a constant speed. On the other hand, the 2nd-DIMP performs weighted image integration by the irradiation-time control, where the focus is rapidly changed within a video frame period at a speed which is inversely proportional to the numerical weight. In the second DIMP, the images whose numerical weight is zero are not recorded from the beginning. Therefore, judging from the point of the information theory, the second DIMP is superior to the first DIMP under the condition of the same total electron dose. However, the 2nd-DIMP has a technical difficulty to control the accelerating voltage as quickly and precisely as possible.

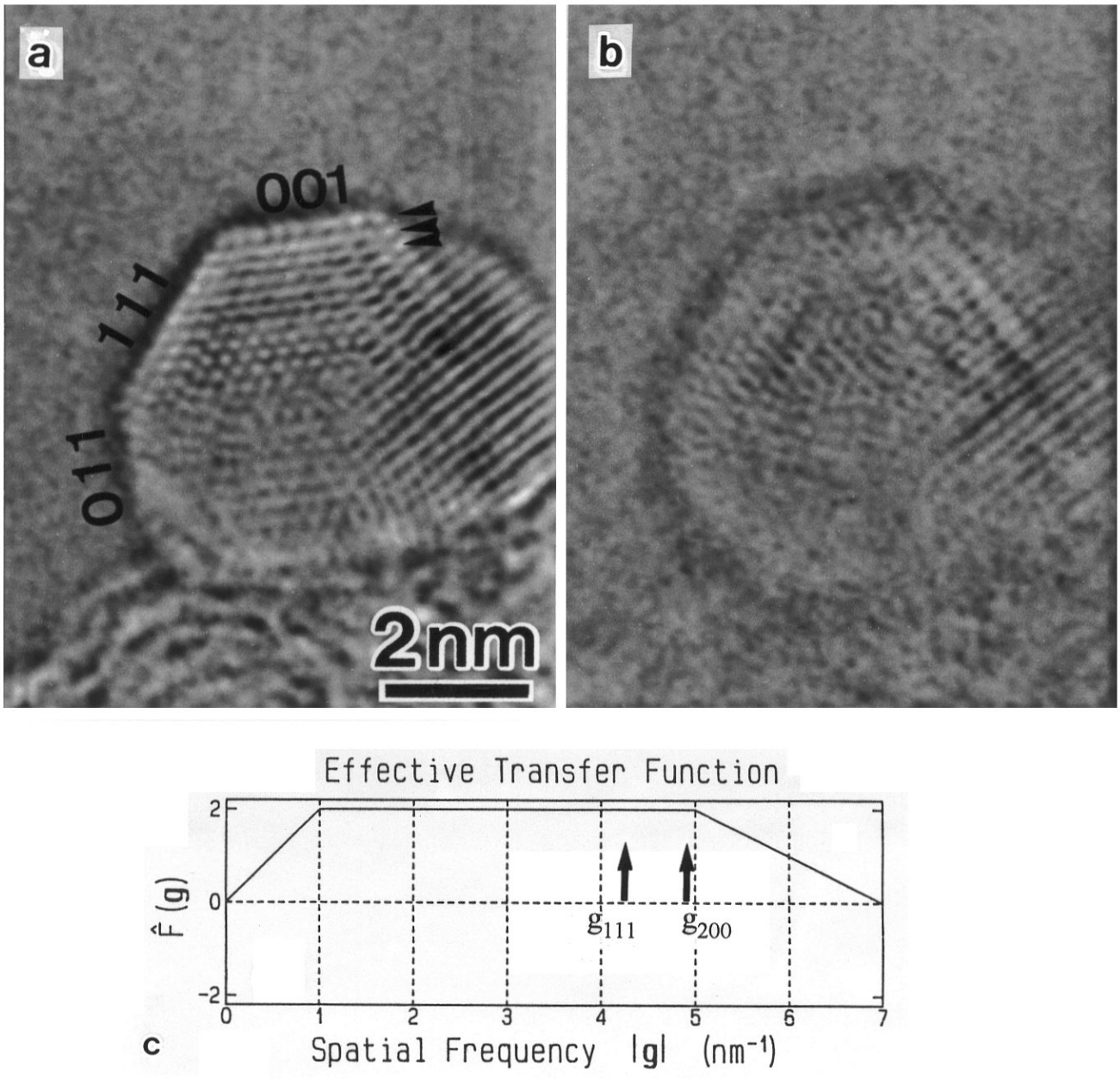


Figure 10. Spherical aberration-free images processed by the first DIMP method for (**a**) the phase component and (**b**) the amplitude component; (**c**) Effective contrast transfer function used in this processing.

Application to the observation of a gold fine particle

A new result of the spherical aberration-free observation of a gold fine particle, which is processed by the first DIMP, is shown in Figures 9 and 10. Figure 9 shows three images from the original through-focus series and respective phase contrast transfer functions. In the in-focus image, as shown in Figure 9a, interference fringes between the primary wave and Au(111) or Au(002) diffracted waves are seen outside the particle. In the image near the Scherzer focus shown in Figure 9b, the shape of the particle can be clearly observed since the main structural information of the particle is distributed over the so-called „spatial frequency window" indicated by an arrow in the corresponding contrast transfer function. The Au(111) and Au(002) lattice planes, however, still exist outside the particle because Au(111) and Au(002)

diffraction spots are not involved in the spatial frequency window. These outside fringes are artifacts caused by the spherical aberration. Figure 9c shows the optimum defocus image for Au(111) lattice planes, Δf=158.8nm. The Au(111) lattice planes are clearly observed inside the particle, however, the shape of the particle becomes blurred compared with Figure 9b. These three images demonstrate how spherical aberration makes it difficult to analyze surface microstructure at atomic resolution.

Figures 10a and 10b show spherical aberration-free images for phase and amplitude components by the first DIMP, respectively. Figure 10c shows an effective transfer function which is used in the present processing. Since the effective contrast transfer function shows a wider „spatial frequency window" than that in Figure 9b, covering $|g_{111}|$ and $|g_{200}|$ without any oscillation, no extra lattice fringes are observed outside the particle in the processed images. The phase image in Figure 10a allows us to identify more precisely not only the shape but also the lattice image of the particle. In the present experiment, since the atomic column of the crystal is not parallel to primary beam, the image is not clear enough to determine atomic structure directly. However, aberration-free observation has a high potential for direct observation of surface structure at atomic resolution. The amplitude image in Figure 10(b) shows that the edge of the particle satisfies the weak phase object approximation relatively well. But the higher contrast inside the particle shows the existence of an amplitude component. In the DIMP, non-linear imaging components give some artifacts in the processed images. In the present images, however, no strong contrast anomalies are observed, even if the sample is not assumed to be an ideal weakly scattering object. We have done some theoretical image calculations of DIMP, where the same weighted image integration was performed in the same way using 256 calculated images (Takai *et al*., 1994b). According to the calculations, the effect of non-linear image component was negligibly small in the processed image contrast when the thickness of gold crystal is less than 4nm. In the thicker sample than 4 nm, of course, some contrast anomaly appears especially at the surface edge area. This technique has now been attracting much attention in the efforts to understand basic mechanism of catalysis properties of gold particles of nanometer sizes (Haruta, 1996).

Conclusion

Two types of DIMP-EM systems were described in terms of their principles and performances. The proposed DIMP techniques have the following distinctive characteristics:

(1) Phase contrast as well as amplitude contrast can be observed at the condition of $C_s = \Delta f = 0$, which means phase electron microscopy is realized by the DIMP methods without using any phase plates. Therefore, the DIMP methods can be regarded as one of the wave front reconstruction techniques analogous to in-line electron holography.

(2) Spherical aberration correction by the first DIMP method enables considerable improvement of resolution from Scherzer resolution limit to information limit.

(3) Through-focus images that are corrected for spherical aberration can be obtained by shifting the origin of the weighting function for the first DIMP and by changing the fundamental focus via objective lens current for the second DIMP method, which is useful to determine how accurately the spherical aberration is corrected.

(4) The whole area recorded by a camera is corrected without using any additional electron optic devices.

(5) All of the processing is done in real space with no Fourier transformation, which makes the method suitable for real-time processing in the second type of DIMP-EM.

At the present stage, the performance of real-time processing is not fully satisfactory due to the technical difficulty to modulate the accelerating voltage with high speed and high precision. But the second method is superior in principle from the image processing point of view, because we integrate the image intensity depending on its importance. Furthermore, the second method of DIMP is a unified technique of image detection and processing and it makes possible real-time image processing. Therefore it should be distinguished from conventional image processing in which image detection and processing are performed separately. This is why we suggest the name „active image processing" for the second type of DIMP.

References

Ando T, Taniguchi Y, Takai Y, Camera Y, Shimizu R (1994a) Active image processing as applied to high resolution electron microscopy [II] Real-time phase-plateless electron phase microscopy by accelerating-voltage modulation. J Electron Microsc **43**: 10-15.

Ando T, Taniguchi Y, Takai Y, Kimura Y, Shimizu R, Ikuta T (1994b) Development of real-time defocus modulation type active image processing (DMAIP) for spherical-aberration-free TEM observation. Ultramicroscopy **54**: 261-267.

Coene W, Janssen G, Op de Beek M, Van Dyck D (1992) Phase retrieval through focus variation for ultra-resolution in field-emission transmission electron microscope. Phys Rev Lett **69**: 3743-3746.

Haruta M, Ueda A, Tsubota S, Torres Sanchez RM (1996) Low-temperature catalytic combustion of methanol and decomposed derivatives over supported gold catalysts. Catal Today, **29**: 443-447.

Hawkes PW (1980) Imgae processing based on the linear theory of image formation. In: Computer Processing of Electron Microscope Images. Hawkes PW (ed). Springer, Berlin. pp 1-33.

Hoppe W, Bussler P, Feltynowski A, Hunsmann N, Hirt A (1973) Some experience with computerized image reconstruction method. In: Image Processing and Computer-Aided Design in Electron Optics. Hawkes PW (ed). Academic Press, London. pp 92-126.

Ikuta T (1985) Active image processing. Appl Opt **24**: 2907-2913.

Ikuta T (1989) Image restoration in coherent imaging system involving spherical aberration. J Electron Microsc **38**: 415-422.

Kirkland AI, Saxton WO, Chau K-L, Tsuno K, Kawasaki M (1995) Super-resolution by aperture synthesis: tilt series reconstruction. Ultramicroscopy **57**: 355-374.

Lichte H, Völkl E, Scheerschmidt K (1992) Electron holography II. First steps of high resolution electron holography into material science. Ultramicroscopy **47**: 231-240.

Saxton WO (1978) Computer techniques for image processing in electron microscopy. In: Advances in Electronics and Electron Physics. Marton L (ed). Academic Press, New York. pp 289-309.

Saxton WO (1994) What is the focus variation? Is it new? Is it direct? Ultramicroscopy **55**: 171-181.

Schiske P (1968) Zur Frage der Bildrekonstruktion durch Fokusreihen (On the question of image reconstruction by through-focus series). Proc 4th Eur Conf Electron Microscopy. Bocciarelli DS (ed). Tipografia Polyglotta Vaticana, Rome. pp 145-146.

Schiske P (1973) Image processing using additional statistical information about the object. In: Image Processing and Computer-Aided Design in Electron Optics. Hawkes PW (ed). Academic Press, London. pp 82-90.

Takai Y, Oba N, Ando T, Ikuta T, Shimizu R (1994a) Spherical aberration-free imaging by hollow-cone illumination processed by the focal-depth extension method. Proc 13th Inst Congr Electron Microsc, Paris Jouffrey B, Colliex C (eds). Les Editions de Physique, Les Ulis, Paris. pp 17-22.

Takai Y, Taniguchi Y, Ikuta T, Shimizu R (1994b) Spherical aberration-free observation of profile images of Au(011) surface by defocus-modulation image processing. Ultramicroscopy **54**: 250-260.

Taniguchi Y, Ikuta T, Endoh H, Shimizu R (1990a) Active image processing as applied to high resolution electron microscopy [I] assessment of misalignment and its correction. J Electron Microsc **39**: 137-144.

Taniguchi Y, Shimizu R, Chaya M, Ikuta T (1990b) Correction of spherical aberration using defocus modulation technique. Proc XIIth Int Congr Electron Microsc. Peachey LD, Williams DB (eds). San Francisco Press, San Francisco. Vol 1, pp 458-459.

Taniguchi Y, Ikuta T, Shimizu R (1991) Assessment of image formation by three-dimensional power spectrum in transmission electron microscopy. J Electron Microsc **40**: 5-10.

Taniguchi Y, Takai Y, Ikuta T, Shimizu R (1992a) Correction of spherical aberration in HREM image using defocus-modulation image processing. J Electron Microsc **41**: 21-29.

Taniguchi Y, Takai Y, Shimizu R, Ikuta T, Isakozawa S, Hashimoto T (1992b) Spherical aberration free observation of TEM images by defocus-modulation image processing. Ultramicroscopy **41**: 323-333.

Taniguchi Y, Ikuta T, Shimizu R (1994) Mathematical background of defocus-modulation image processing. Optik **96**: 129-135.

Van Dyck D, Op de Beek M (1990) New direct methods for phase and structure retrieval by HREM. Proc XIIth Int Congr Electron Microsc. Peachey LD, Williams DB (eds). San Francisco Press, San Francisco. Vol 1, pp 26-27.

Discussion with Reviewers

D. Van Dyck: The theory behind the method is based on linear imaging which is only valid for weak phase objects. Is there any indication for the validity of the method for stronger scattering objects such as crystals in an exact zone orientation?

Authors: The DIMP method is valid only for weak scattering objects (not only weak phase objects but also weak amplitude objects), because the DIMP is based on linear imaging theory. Therefore, it is not generally applicable to crystal specimens in an exact zone orientation. The validity of the method for such stronger scattering objects can be estimated by computer simulations. We have done some computer simulations of the DIMP method using 256 calculated images considering dynamic electron diffraction in the sample and image formation through the lens (Takai *et al*., 1994b). According to the calculations, the Au sample of 4 nm thickness in (011) orientation seems to be valid, because no strong contrast anomaly was detected in the processed images. Strictly speaking, however, quantitative line intensity profile analysis indicates that the effect of non-linear imaging components actually exists and it disturbs the processed image contrast to some extent especially at the surface edge region. For thicker specimen, the DIMP method does not work effectively any more, because the effect of non-linear imaging components become too large to adopt weak scattering object approximation.

D. Van Dyck: What will happen when the object is a crystal

with a square unit cell so that the high resolution images vary periodically with focus (Fourier images). This focus period is $\varepsilon = 2a^2/\lambda$ with a the mesh of the square, i.e., for Au, $a = 0.2$ nm and for $\lambda = 2$ μm (300 keV), giving $\varepsilon = 10$ nm.

Authors: The focus interval in this method is an important factor which is determined by the finest structure to be resolved by the processing. The interval should be smaller than $2g^2/\lambda$, where g is the largest spatial frequency to be resolved, as you pointed out. In the present experiments, we use 256 images taken with 2.28 nm focus interval, which is much smaller than 10 nm. Under such a condition, the DIMP works well even if the image contrast of a crystal varies periodically with focus.

D. Van Dyck: The defocus in the first method (Figure 2a) covers a focal range of more than 500 nm. However, for large focus values the envelope function of beam convergence will deteriorate the information limit ρ_i. For instance, for a beam convergence apex angle α, a measure for the maximum defocus is $\varepsilon < \rho_i/2\alpha$. If $\rho_i = 0.2$ nm and $\alpha = 10^{-3}$ rad, we then have $\varepsilon < 200$ nm.

Authors: Simple averaging of image intensity in a wide focal range surely deteriorates the information limit. However, image integration with bipolar weighting function does not deteriorate the information limit even if images in a wide focal range are used, as proven by the experimental results of the extraction processing in Figure 6. In the experiment, 256 images from -200 nm to 400 nm were used. The reason why we used such long focus range is to restore not only the fine structure but also low spatial frequency components.

D. Van Dyck: Why is the defocus range for method 2 much smaller than for method 1? Is it also related to the response of the drive for voltage modulation?

Authors: Yes, it is. The defocus range in the DIMP is related to the frequency range of the effective contrast transfer function as shown in Figure 2, not to the type of DIMP. In the present experiment of the second DIMP, a narrower range of the effective contrast transfer function had to be assumed due to the limited frequency response of the high voltage modulation by the present driving method, resulting in the smaller focus range in the second DIMP.

D. Van Dyck: What is the accuracy of the method for the small spatial frequencies? From Figure 1b it can be estimated that the selectivity in the reciprocal focus detection is inversely proportional to the focal range, which puts a limit on the accuracy, especially for small spatial frequencies. On the other hand, the highest spatial frequency is the inverse of the distance between the focal images. From this, the spatial frequency range g_{max}/g_{min} (as in Figure 2a) would be roughly equal to $\sqrt{N}$ with N the number of images.

Authors: In order to construct all components in a wide spatial frequency range with high accuracy, a lot of images taken with small defocus intervals in a wide focal range are necessary in the DIMP method. In the present experiments, 256 images with 2.28 nm focus intervals were used for reconstructing the highest spatial frequency component and a wide focus range from -200 nm to 400 nm are adopted for reconstructing the small spatial frequency components. The number of images, $N = 256$, was used for reconstructing phase and amplitude image as precisely as possible.

D. Van Dyck: How do you determine the absolute focus if no amorphous specimen is available?

Authors: It can be rather easily determined within 2-4 nm accuracy by constructing through-focus images that are corrected for spherical aberration. Because the contrast of the amplitude image shows the smallest at in-focus condition at the edge of the crystal, where the thickness should be thin enough to assume a weakly scattering object.

D. Van Dyck: The weighting function in Figure 2 contain both the effect of C_s and Δf. In this way one directly reconstructs the exit wave (but therefore the knowledge of C_s and of absolute focus is required). It is also possible to use a weighting function that only accounts for the defocus relative to a fixed image plane so that the exit wave in that plane will be restored. In a second step one can then correct for C_s and for the absolute defocus onset so as to reconstruct the exit wave of the object.

Authors: As you mentioned, the DIMP need both information of C_s and absolute defocus onset so as to reconstruct the exit wave of the object. They can be easily estimated with enough accuracy by constructing a Thon diagram.

D. Van Dyck: Why is through focussing done by modulating the voltage rather than the objective lens current? Because of speed, hysteresis?

Authors: Through focussing by modulating the objective lens current is limited in speed due to the huge inductance and the magnetic hysteresis of the objective lens. More rapid modulation of focussing is necessary in the second DIMP than in the first DIMP. Therefore we decided to use the high voltage modulation method for the second DIMP, instead of the objective lens current modulation in the first DIMP.

P.W. Hawkes: I am surprised that a gold particle is a suitable test object, as it is not likely to be a weak scatterer. The authors observe that it is indeed not a weak phase object but for the method to work, both phase and amplitude need to be weak. Please comment.

Authors: The validity of the DIMP method for strong

scattering objects can be estimated by computer simulations of the DIMP method. In order to examine whether the DIMP method works well or not for the specimen, we have done some theoretical calculations of the DIMP method using 256 calculated images considering dynamic electron diffraction in the sample and imaging parameters, such as defocus, C_s, beam divergence angle and chromatic defocus spread. According to the contrast calculation, the sample of Au in (011) orientation of 4 nm thickness seems to be valid, because no strong contrast anomaly was detected in the processed images. But, when the thickness becomes larger than 4 nm, of course, some contrast anomaly appeared especially at the sample edge region which seems to be due to the effect of non-linear imaging component.

T. Hanai: The importance of the extraction of Fourier components is not so clear. It seems to prove that the focus step used for the construction of the weighting function is in good agreement with that in the experiment. But I do not understand if the authors claim that the estimated spherical aberration coefficient is also correct. I think that the power spectra shown in Figure 6 can be obtained even if the spherical aberration is miscalculated because the effect of the spherical aberration is a relative phase shift of Fourier components.
Authors: In the extraction processing of one Fourier component, the effect of spherical aberration can be corrected by considering the phase shift shown by the dotted curve in Figure 1a. Therefore, the absolute focus value should be determined for each defocused image. In the present experiment, the absolute value was determined in advance by constructing the Thon diagram as shown in Figure 7.

T. Hanai: I also would like to know the reason why the authors consider that there are no artifacts in the image of the gold particle shown in Figure 10. This technique is based on weak phase approximation, which is not satisfied in this case, as the authors also pointed out. I am afraid that an artifact can be caused by deviation from the approximation.
Authors: Some artifacts may appear in the processed images due to the deviation from the weak scattering object approximation, but they are negligibly small, judging from the observed image contrast. The DIMP method is valid only for weak scattering objects. The validity of the method for a strong scattering object can be estimated by computer simulations of the DIMP method using calculated images considering dynamic electron diffraction in the sample and image formation through the lens (Takai *et al.*, 1994b). According to the calculations, the Au sample of 4 nm thickness in (011) orientation seems to be valid, since no strong contrast anomaly was detected in the processed images. For thicker specimen, the DIMP method does not work effectively any more, because the effect of non-linear imaging components becomes large in the image contrast.

T. Hanai: The spherical aberration coefficients of the used electron microscopes have to be known for construction of the weighting functions. Did you estimate the tolerance in measuring or calculating the aberration coefficient to obtain a resolution beyond the Scherzer limit?
Authors: The spherical aberration coefficient should be estimated within the tolerance of _0.1 mm for achieving 0.2 nm resolution. From the experimental point of view, however, we can confirm whether the estimation of the spherical aberration is accurate or not, by constructing a Thon diagram using through-focus images that are corrected for spherical aberration. Therefore we can improve estimation accuracy by checking the symmetry of the Thon diagram.

T. Hanai: In the off-line processing, the optimum image formed by the uniform transfer function with no defocus can be chosen using the Thon diagram. For real-time DIMP, does the absolute defocus amount always have to be monitored with high accuracy to obtain the in-focus image?
Authors: Yes, it does, but in real-time DIMP, the in-focus condition can be monitored rather easily by checking the contrast minimum of amplitude image while changing the absolute defocus amount via objective lens current. Under the condition, spherical aberration-free phase image can be obtained by changing the weighting function for amplitude component to that for phase component.

T. Hanai: In the second method based on modulation of the accelerating voltage, the image can be accumulated many times instead of taking a focal series in the first method. Is this an essential difference between these two methods from the signal-to-noise ratio point of view?
Authors: There is a more essential difference between the two methods. In the 1st-DIMP, the image integration is performed numerically by a computer for the defocused images which were recorded changing the focus at a constant speed. On the other hand, the 2nd-DIMP performs weighted image integration by the irradiation-time control, where the focus is changed at a speed which is inversely proportional to the numerical weight. In the second DIMP, the images which are important for the aberration correction are recorded at sufficiently long exposure time, on the other hand, the images whose numerical weight is zero are not recorded from the beginning. Therefore, judging from the S/N ratio point of view, the second DIMP is superior to the first DIMP under the condition of the same total electron dose.

P. Kruit: How difficult is it to get the alignment sufficiently rotation-free? Can you give quantitative requirements?

Authors: Additional changes induced by the focus modulation, such as magnification, image rotation, image brightness and image contrast must be as small as possible in the processing. In the present electron microscope, the angle of image rotation was less than 1 degree when the focus was changed in the range of 600 nm. Therefore, the image rotation was not of crucial influence on the processing. The detailed data will be reported later.

P. Kruit: For a large field of view. the method cannot allow magnification charges as you go through focus. How do you avoid those?
Authors: Magnification change must also be negligibly small for the focus modulation, otherwise the processing cannot be performed ideally for a large field of view. The magnitude can be estimated to be less than 0.05%, judging from the focal length of the objective lens and the focus modulation range. The induced magnification change could not be detected in experiments when the focus was changed.