Electron microscopy in semiconductor inspection

Koji Nakamae (Osaka University)

Abstract

Currently, semiconductor devices are manufactured in a technology node of several nanometers. Electron microscopy is mainly used in semiconductor inspection in manufacturing stages since accelerated electrons have wavelengths of nanometers or less, and a high spatial resolution can be expected. Among various electron microscopes since the scanning electron microscope can observe the sample as it is without processing the sample, the scanning electron microscope-based inspection instrument is mainly used at each stage of manufacturing the semiconductor device. The paper presents a review of SEM-based electron microscopy in semiconductor inspection. First, an overview of electron microscopy is described to understand the electron-sample interaction, the characteristics of electrons emitted from an irradiated specimen, charging, noise, and so on. Next, application areas such as mask inspection are introduced. Finally, future challenges are discussed.

Keywords: scanning electron microscope, semiconductor inspection, critical dimension, defect detection, defect classification, image processing

1. Introduction

Semiconductor device researchers and manufacturers have continued to develop new device structures such as 3D structures, new materials, manufacturing methods and manufacturing instruments to follow the well-known Moore Law (dimensional scaling). As a result, the integration density of semiconductor devices continues to increase according to Moore's Law. In 2019, extreme UV lithography was introduced for mass production of logic-integrated circuits on a 7 nm technology node. At each manufacturing stage of devices with a size of several nanometers, an inspection method with an accuracy of nanometer or less is required for inspecting whether the device is manufactured as designed at each manufacturing stage.

Electron microscopy irradiates a specimen with an accelerated electron beam and obtains information about the specimen from various emitted signals. The wavelength of the accelerating electrons used is related to the spatial resolution of the obtained information. For example, an electron accelerated at 100 V (energy 100 eV) has a wavelength of approximately 0.12 nm. Thus, wavelengths below nanometers (nm) can be easily achieved. For this reason, electron microscopes are mainly used to inspect each manufacturing stage of semiconductor devices.

There are several types of electron microscopes, including scanning electron microscopes (SEMs), transmission electron microscopes, scanning transmission electron microscopes, and electron holography. Of these, SEMs are microscopes that can observe specimens as they are (without processing for observation) at each stage of manufacturing. For this reason, SEM-based instrument is mainly used for inspection at each stage. Semiconductor devices are manufactured by repeating a number of processes such as film formation, lithography, impurity diffusion, and electrode formation. Thus, the following inspections are carried out by using SEM-based instruments: 1) critical dimension (CD) measurement to check whether a line with the designed width was formed

correctly without fluctuation, 2) overlay measurement to check whether the next pattern was formed in the correct position on the previously formed pattern, 3) defect and foreign particle detection, review, and automatic classification, and 4) three dimensional (3D) measurements of deep hole for 3D integration of semiconductor devices, etc.

By using a scanning electron microscope, the observed image of the specimen can be obtained in a digital representation. As a result, powerful digital image processing can be applied. In addition, advanced digital image processing and analysis techniques such as recent deep learning have been used.

The paper presents a review of SEM-based electron microscopy in semiconductor inspection. This review is organised as follows: Section 2 gives an overview of electron microscopy. To understand the physical constraints, it is important to understand the electron-sample interaction and scanning electron microscopy. Section 3 describes its application to semiconductor inspection. Section 4 describes future challenges. In addition, Section 5 presents the conclusions.

2. Outline of electron microscopy

To understand why SEM is used for semiconductor inspection, it is necessary to understand the interaction between incident electrons and specimens, contrast-forming factors, image forming methods, etc. This section gives an overview of these. It also outlines the simulations used in model-based inspection methods. These understandings will help us know considerations in the SEM-based instruments in Section 3. In addition to these, this chapter introduces recent related studies.

2.1 Electron-specimen interactions

Incident electrons with energies interact with specimen atoms and electrons in various ways and consume their energy. Many of the resulting phenomena can be used for semiconductor inspection. This section describes the characteristics of these interactions.

2.1.1 Electron scattering and diffusion

The interaction between the incident electron and the specimen atom occurs through various electron scattering mechanisms. As a result, the direction and energy of the incident electrons change. Energy transfer to the specimen also occurs, resulting in some form of secondary radiation emission [1]. Two types of scattering, elastic scattering and inelastic scattering, are well recognised. Elastic scattering re-orients electron orbitals, but the energy remains constant. Inelastic scattering gives some of the kinetic energy to the specimen atom and reduces the energy of the electrons.

The scattering process is expressed using a cross section, which means the probability that the process will occur. In some cases, instead of the cross section, the mean distance that electrons pass through the specimen for the process to occur, that is, the mean free path λ, is used.

When incident electrons with energies (hereinafter primary electrons (PEs)) pass near the nucleus, the electrons have a chance to be elastically scattered by the Coulomb field of the nucleus. For this elastic scattering the Rutherford cross section equation is well known. In this equation, the angular scattering cross section decreases with the

scattering angle θ in the form of sin-4 (θ / 2) [2]. The scattering angle θ is the angle measured from the incident direction. In this model formula, it diverges in the forward direction ($\theta \rightarrow 0$), resulting in non-physical results. This condition corresponds to the case where the electron passes far away from the nucleus. In reality, electrons are shielded by surrounding atoms. Therefore, the Coulomb force between the passing electrons and the nucleus is strongly suppressed at a distance. The result is a finite scattering cross section. This modified model is called the screened Rutherford cross section. It is well used as a reliable elastic scattering model for keV electrons [3]. However, if the scattering angle is large, such as $\theta \rightarrow \pi$, another correction is required. Here, the electrons are attracted to the strong positive potential of the nucleus and acquire kinetic energy. As a result, the electron velocity is faster than the original incident velocity. This effect is noticeable when low-energy electrons enter a material with a high atomic number. The correction, called the Mott cross section, is based on the Dirac equation of electrons with partial wave expansion of the atomic potential [4], [5].

The inelastic scattering process results from the interaction of PEs with atomic electrons. Various types of inelastic scattering occur depending on whether the electrons in the specimen are excited alone or collectively. Part of the energy transferred to the specimen is used to emit secondary electrons, Auger electrons, X-rays, etc. from the specimen, and these emitted quanta carry information useful for inspection. Thus, the inelastic scattering reduces the energies of the PEs.

If a particular process involves the characteristic energy of excitation, measuring the energy spectrum of the electrons emitted from the specimen can reveal the characteristic energy loss. As a result, it helps identify the process.

The secondary electrons (SEs), that is, the electrons emitted from the specimen by the PEs, are generated as a result of energy transfer from the primary electrons to the conduction band electrons of the specimen constituent atoms. The energy transferred to conduction band electrons is approximately 1 to 50 eV. Due to the low energy of secondary electrons, their range within a solid is approximately 5 nm. Considering this range, the interpretation of secondary electron signals can be tricky, as the surface of the specimen is oxidised or contaminated even in vacuum.

Most SEs are low-energy (less than 50 eV) electrons, but collisions occur between the PE and atomic electrons and large energy transfers can occur in rare cases. The SEs generated as a result are called fast SEs. Fast SEs can be generated with energies up to that of the PE.[1] This electron has a wide range due to its high energy. Murata et al. (1981) [6] demonstrated the importance of fast SEs in the definition of spatial resolution in the exposure of electron beam resists. Joy et al. (1982) [7] studied the effects of fast SEs on the spatial resolution of the analysis using an analytical electron microscope.

Primary electrons can interact with tightly bound inner-shell electrons and emit electrons from atoms. The minimum energy required to emit an electron from an atom is called critical excitation (ionisation) energy. Bethe (1930) [8] has expressed the cross section of the inner shell ionisation.

Also, interactions with many electrons in the specimen atom are known.

When PE passes through the Coulomb field of the atom, it has a chance to be decelerated and to decrease its energy. The reduced energy is used as the photon emission of electromagnetic radiation. This "bremsstrahlung" forms a continuous distribution of photon energy from zero to incident energy. This continuous X-ray is the main component in the background of the X-ray spectrum from the specimen.

As the PEs pass through the specimen, the Coulomb field of the PEs can perturb the electrons of the specimen

solid over a relatively long distance. The vibration caused by this perturbation is called plasmon. Ferrel (1956) [9] expressed the cross section of the plasmon excitation. In plasmon scattering, PEs lose energy in quantised units. Excited plasmons have been reported to have a wavelength of approximately 100 interatomic distances (Hirsch et al., 1965) [10].

If the specimen is transparent to PEs, the transmitted electrons can be recorded by a detector situated below the specimen support. This is called Transmitted Electron (TE) mode. Atoms in the crystal are thermally vibrating around their lattice positions. PEs passing through the crystal can interact with these crystal vibrations to create or annihilate phonons (Hirsch et al., 1965) [10]. Although the energy change due to this scattering is insignificant (~ 0.02 eV), it deflects the orbit of PE. As a result, when observing Bragg diffraction spots, the effect of thermal vibration appears as the diffusion background between the diffraction spots in TE mode.

As mentioned above, the interaction of PEs with atoms occurs through various mechanisms of electron scattering, which can result in energy transfer to the specimen. The result is some form of secondary radiation (Figure 1). In semiconductors, a few thousands of electron-hole pairs are created per incident PE. In the depletion layer of a p-n junction, the electric field separates the charge carries and minority carriers can reach the junction by diffusion. This results in a charge-collection current or electron-beam-induced current (EBIC). Also, the ultraviolet or visible light (cathodoluminescence: CL) can be emitted by the PE irradiation. These secondary radiations such as EBIC and CL are used in analytical modes to measure semiconductor and device parameters. Below, SEs and BSEs that are mainly used in the semiconductor inspection are described.

Figure 2 shows the energy spectrum of the electrons emitted as a secondary radiation. Bongeleru et al. [12] summarised the experimental results of the measurement range (penetration depth) in units of g cm^{-2} when PE with energy E_{PE} in the range of 0.5 to 20 keV is incident on various elements (see Figure 6 in Reference [12]). The range in units of centimetres is obtained by dividing by the density in units of g cm^{-3}. For example, the range when 1 keV PE is incident on Si is approximately 9 μg/cm^2. Since the density of Si is 2.33 g/cm^3, the range in metres is approximately 40 nm.

2.1.2 Emission of backscattered and secondary electrons

Backscattered electrons (BSEs) and secondary electrons (SEs) are commonly used signals for inspection images. The BSE signal (backscatter coefficient η) and SE signal (secondary electron yield δ) depend on the inclination of the specimen surface, the specimen material, the energy of the PEs, and so on. Understanding these characteristics is essential for interpreting the specimen information contained in the image.

Electrons with emitted energies greater than 50 eV are conventionally called BSEs, and electrons with less than 50 eV are called SEs. Taking the current of backscattered electrons as I_{BSE}, η is given by the following equation.

$$\eta = I_{BSE} / I_P ,$$

where I_P denotes the incident PE current. As well, taking the current of secondary electrons as I_{SE}, δ is given by the following equation.

$$\delta = I_{SE} / I_P .$$

A total yield σ is given by the following equation.

$$\sigma = \eta + \delta = (I_{BSE} + I_{SE}) / I_P \ .$$

When σ> 1, more electrons leave the specimen. If the specimen is an insulator, the specimen will be positively charged, whereas σ <1 means it will be negatively charged.

Inspection uses the contrast of the available signal in the inspection area. When the detection signals at inspection points A and B are I_A and I_B, respectively, the following equation defines contrast C.

$$C = (I_A - I_B) / (I_A + I_B)$$

Backscattered electrons (BSEs)

It is known that the BSE signals have five types of contrast depending on the generation factors [13].

(1) Thickness contrast, film thickness estimation

Niedrig [13] summarized the experimental results on the backscattering coefficient η_∞ of bulk materials vs atomic number Z at normal incidence (see Figure 6 in Reference [13]). It is seen from his result that the backscattering coefficient η_∞ of bulk material monotonically increases with increasing atomic number. Therefore, the thin film (top layer) on bulk substrate affect the backscattering coefficient $\eta_{S\infty}$ of the clean substrate in the case of the atomic numbers of the top thin film (Z_f) and of the substrate (Z_S) are different. As the top layer thickness increases, the total backscattering coefficient η_{total} of such a specimen varies between $\eta_{S\infty}$ and $\eta_{f\infty}$ for very thick top layers. For top layers of different thicknesses, corresponding thickness contrast can, therefore, be observed in the SEM. The thickness of the top layer on the bulk substrate can be estimated by measuring the backscattering coefficient.

Bongeleru et al (1993) [12] reported experimental η_∞s for different elements and for different PE energies (see Figure 7 in Reference [12]). From their results, it can be seen that η_∞ has little dependence on PE energy in the range of approximately 5 keV to 200 keV. However, at lower PE energies, η_∞ exhibits different characteristics. For specimen atoms with an atomic number Z of 30 or less, η_∞ increases as the PE energy decreases. In contrast, it decreases for specimen atoms with Z of 30 or more. These characteristics should be taken into account during the inspection. [13]).

(2) Z-contrast, material contrast

Bongeleru et al (1993) [12] summarized the average energies of emitted BSEs from a bulk specimen of a single element from C to Pb (see Figure 11 in Reference [12]). It can be seen that the average energy of BSE is within from $0.45E_{PE}$ to less than $0.8E_{PE}$ in the range of 0.5 keV to 370 keV of PE energy E_{PE}. It is also seen that the average energy increases with atomic number Z. However, it is known that the energy of the maximum value of the energy distribution of BSE is higher than the average energy when the atomic number is high. Also, the BSE signal with $E < 0.5\ E_{PE}$ does not change much with Z [12]. These electrons reduce the contrast of the material. Considering these characteristics using an energy filter for the BSE signal can enhance the material contrast.

(3) Angular distribution contrast

The angular distribution of BSE radiation depends on the PE energy, the incident angle of the PE, the elements (atomic number) that make up the specimen, the specimen structure, etc.

For thick specimens, multiple scatters or diffuses are superimposed on the elastic backscatter distribution. The

angular distribution of BSE is the cosine angular distribution on the surface of bulk materials with high atomic numbers (for example, see Figure 14(a) in Reference [13]). However, for specimens with normal incidence and thin specimen of low atomic numbers, the angular distribution of the backscatter becomes flat because most BSEs are due to a single elastic scattering.

When the PEs are incident at an angle, the peak of BSE angular distribution is in the direction of light reflection (for example, see Figure 14(b) in Reference [13]). In addition, the backscattering coefficient increases as the incident angle α increases. Therefore, when PE is vertically incident on surfaces with different tilt angles, tilt contrast (topographic contrast) occurs due to the different backscattering coefficients.

For obliquely incident PE, the directions of the maximum peak of the backscatter distribution of high atomic number (Au in the figure) and low atomic number (Al in the figure) are the same. Also, the magnitude of the peak-backscattering coefficient $d\eta / d\Omega$ is approximately the same. See Figure 14(b) in Reference [13]. If the detector is in this direction, the material contrast will be reduced, but the topographic contrast will remain. In the case of normal incidence, the material contrast is high.

In the case of low voltage SEM (PEs with energy less than 3 or 5 keV), the backscattering coefficient shows slightly different characteristics. Bongeleru et al. (1993)[12] reported a Monte Carlo simulation results of the angular distribution of BSE emissions when vertically incident on Al, Cu, Ag, and Au. 1 keV, 3 keV, and 5 keV are used as PE energies. See Figure 12 in Reference [12]. For Al, all energies show a distribution close to the cosine angle distribution. However, as atomic number Z increases, decreasing ζ emits more electrons than $\cos\zeta$ where ζ is the angle between exit direction and surface normal. Bongeleru et al. (1993)[12] showed the measured values of η (Z, ϕ) for 10 keV and 1 keV electrons (see Figure 10 in Reference [12]). ϕ is the tilt angle. ϕ=0 means the normal incident. The η (ϕ) curve shows a lower than expected increase with increasing ϕ, which is more pronounced at 1 keV than at 10 keV.

When using BSE for inspection, it is necessary to understand these characteristics.

(4) Low-loss electrons

The electrons with low energy loss have mainly suffered a single large-angle scattering event and have an information depth of about 10 nm or less. They can be separated from those electrons with higher energy losses (and hence with a greater information depth) by energy filtering. This "low-loss image" is more sensitive to shallow surface topography than is the secondary electron image.

(5) Diffraction effects, orientation contrast

Crystalline specimens produce a BSE signal that depends on the angle of incidence of PE. This is caused by the diffraction of PE and/or scattered electrons on the lattice plane of the crystal. This allows differently oriented grains in polycrystalline material to be imaged by the crystal orientation or channeling contrast. In order to better understand the observed contrast, it is necessary to analyze the electron wave field (Bloch wave) generated in the crystal according to the diffraction theory. Inspection techniques using this contrast include electron channeling patterns (ECP), electron backscatter diffraction (EBSD), and electron channeling contrast imaging (ECCI). In ECP, the BSE intensity at a point of interest is recorded as a function of beam-specimen angle to know crystal local orientation. In

EBSD, the PE is scanned across the sample surface and at each grid point over a region of interest, the cones of diffracted electrons are detected by a 2D electron detector to form an electron backscatter pattern. From detected data, the local crystal structure and crystal orientation at the surface of a specimen can be determined. ECCI is the imaging of channeling contrast over a region of interest at the pre-determined specimen tilt condition.

Secondary Electrons (SEs)

The SE signal that reaches the detector contains four different types of signals depending on where it generated. These signals are shown in Figure 3 as SE (1), SE (2), SE (3) and SE (4).

The SE (1) signal is SEs produced by PE and emitted only near the irradiation point of PE. For this reason, it is important to detect only this SE (1) signal when high spatial resolution is required, but in reality, it is difficult. Also, this signal is sensitive to surface conditions. In standard SEM, a contaminated layer can be formed on the surface, which reduces the contrast caused by the specimen material. Where, the SE (1) emission coefficient only by PE is described as δ_{PE}.

The SE (2) signal is SEs produced by BSE at the surface of the specimen. The signal reduces the spatial resolution due to the lateral spread of BSE (RE in Figure 3) within the specimen. Figure 4 shows the schematic spatial distribution of the signal intensities of SE (1) and SE (2). This signal shows a higher material contrast than SE (1) because the difference in BSE η with respect to the specimen material can be greater than that in SE δ. In addition, the contrast of this material is largely independent of the thin contaminated layer. The emission ratio of SE (2) by BSE will be described as $\eta\delta_{BSE}$.

The SE (3) signal is a signal generated when the emitted BSE collides with the wall surface of an objective lens or the like. It can amplify the SE (2) signal, but it reduces the spatial resolution as well as the SE (2). The emission ratio of SE (3) released on the wall surface by BSE will be described as $\alpha\eta\delta_{walls}$. α is a constant that depends on the detector configuration and takes a number between 0 and 1.

The SE (4) is a signal generated when PEs collide with a wall surface, etc. when passing through an electro-optical column, and is called floating SE. This signal can also enter the detector and becomes noise.

SE signals also have several types of contrast [14], since SE yield δ depends on the inclination of the specimen surface, the specimen material, the energy of the primary electrons, and so on.

(1) Material contrast

Bongeleru et al. (1993)[12] reported the measured SE yield δs where PE in the range of from 0.3 keV to 5 keV or less is irradiated to the bulk specimen (C, Si, Ag, Au) from the vertical direction (see Figure 14 in Reference [12]). From their results, it can be seen that δ also depends on the energy of PE, and that δ differs depending on the specimen element. Thus, on uncontaminated surfaces, surface materials with different δs create contrast. For example, if you want to observe material A and material B with the clearest possible contrast, it is well performed to observe with the PE of E^m_{PE} that gives the maximum δ of material A (δ^m_A), or observe with the PE of E^m_{PE} that gives δ^m_B of material B. Figure 5 shows the definitions of E^m_{PE}, δ^m, E^I_{PE}, and E^{II}_{PE}. The maximum E^m_{PE} and SE yields of δ^m for various atomic numbers Z are shown in Table 1 of Reference [15]. Note that in a standard vacuum system, the contaminated layer can be created by PE and BSE, so the δ of SE can change over time. As the thickness of the contaminated layer increases, the contrast disappears.

(2) Topography contrast

Bongeleru et al (1993)[12] reported also the measured SE yield δ with respect to the inclination angle of the specimen surface where PE is used in the range of from 0.5 keV to 10 keV and specimens are Al, Cu, and Au (see Figure 17 in Reference[12]). It is seen that the SE yield δ increases with increasing incident angle θ measured relative to the surface normal when θ <80 °, according to the following equation.

$$\delta(\theta) = \delta_0(\cos\theta)^{-n}; \quad \delta_0 = \delta(\theta = 0°).$$

A value of n = 1 is valid for materials with a value of approximately Z = 30. For lighter elements, n increases to about 1.3, and for heavy elements, n decreases to approximately 0.8. The increase in δ with the increase in θ is due to the small escape depth of SE. The longer the PE penetrates within the SE escape depth, the higher the yield. As θ increases, δ and E^m_{PE} increase as shown in Figure 6.

Topography contrast is caused by the dependence of δ on the angle of incidence. Specimen details with different slopes θ and (θ + Δθ) with respect to the PE beam are imaged with different brightness.

(3) Electron channeling and crystallographic contrast

As described in BSE, when the specimen is a single crystal, the penetration depth of PE into the specimen changes depending on the crystal orientation. As a result, the contrast called an electronic channeling pattern (ECP) is generated because the amount of emitted BSE or SE changes.

(4) Voltage type I contrast

The yield of SE depends on the energy of PE that irradiated the specimen surface. If the specimen has a region with a potential different from the earth potential, the incident energy of PE at the location changes by that potential, and the yield of SE also changes. The contrast due to this factor is called Voltage type I contrast. For example, if the specimen contains an insulator, it will be charged when the insulator is irradiated with PE with energy other than E^I_{PE} or E^{II}_{PE} shown in Figure 5, and the potential will be non-zero at that location. The contrast created in this way is used for inspecting electrical networks (contactless test).

(5) Voltage type II contrast

The energy distribution of the emitted SE is determined by the specimen material and remains the same as long as the specimen is not contaminated. When the specimen surface potential changes, the SE energy distribution also shifts by the change in potential. Contrast can be created by arranging a retarding electrode, etc. in front of the specimen to filter the energy and detecting only the electrons that have passed through the retarding electrode with the applied potential. This contrast is called "Voltage type II contrast". The surface potential can be measured by recording the energy distribution of the emitted SE with an energy analyzer. The voltage resolution of the analyzer depends on the signal-to-noise ratio, the bandwidth of the detection system, and so on. Detection of a voltage difference of approximately 1 mV has been reported [16], [17].

(6) Voltage type III contrast

This type of contrast does not occur if there is a detector that can detect all SEs emitted from the specimen surface. In reality, only the emitted SEs that enter a certain region reach the detector. If there are regions with different potentials on the specimen surface, an electric field will be generated on the specimen surface, deflecting the trajectories of the SEs with low energy. As a result, the number of SEs that reaches the detector changes, resulting in contrast. The contrast due to the deflection of SEs in the electric field on this surface is called "Voltage type III contrast". Nakamae et al. (1981) [18] have reported the voltage measurement accuracy when this type of contrast was used in three types of detectors: a conventional SEM detector system, a retarding field energy analysis system, and a high-resolution energy analysis system.

(7) Contrast at the surface of ferroelectrics.

The domain structure of a ferroelectric substance is composed of spontaneous polarization with different orientations. When such a ferroelectric specimen is irradiated with PEs, the number of emitted SEs changes due to the interaction between PE and polarized atoms inside the specimen, or the emitted SEs is affected by the electric field that locally generated over the surface due to the spontaneous polarisation charges in the specimen. The latter is the same as the generation of "Voltage type III contrast". As a result, contrast appears.

(8) Magnetic type contrast

If the specimen contains a structure that generates a magnetic field, the effect of the generated magnetic field causes contrast. For example, a specimen is assumed to be a part of an integrated circuit that contains a metal wire that carries a signal. A magnetic field is generated when an electric current is passed through this metal wiring. The internal magnetic field can deflect the PE and BSE trajectories in the specimen. In addition, the magnetic field leaking onto the specimen surface can deflect the SEs and prevent them from reaching the detector.

The contrast generated by these factors is called "Magnetic type contrast".

2.1.3 Charging

If the number of incident PEs is greater than the number of emitted electrons (SE + BSE), the specimen will accumulate a negative charge, and if it is less, a positive charge will be accumulated. This charging changes the amount of SE signal collected, deflects the PE trajectory, and even creates an electric field that can damage the specimen. If the specimen is conductive and connected to the ground, the following relationship holds between the PE current I_{PE}, SE current, and specimen current I_{SC} flowing between the specimen and ground.

$$I_{PE} = \delta I_{PE} + \eta I_{PE} + I_{SC}$$

If the specimen is not connected to ground, or if it is not conductive (insulator), the I_{SC} will be zero. As a result, the amount of charge ΔQ shown by the following equation is accumulated in the specimen per second.

$$\Delta Q = I_{PE}(1-(\delta+\eta)) \quad per\ \mathrm{sec}ond$$

At PE energies higher than EIIPE shown in Figure 5, the total electron yield ($\sigma = \delta + \eta$) is less than 1 and negative charges are accumulated. These negative charges have the effect of repelling some SEs that return to the surface. In addition, PE electrons (accelerated from an electron source with a negative potential) reduce the energy of PE

electrons due to the negative charge of the specimen. This will be closer to EIIPE, so the SE emission ratio δ will be higher. In addition, the electric field strength between the specimen surface and the SE detector will increase. As a result, the SE signal in the negatively charged area increases and becomes brightly on the display.

At PE energies between E^{I}_{PE} and E^{II}_{PE} in Figure 5, σ is greater than 1. Many electrons emit from the specimen, leaving a positive charge. In the positively charged region, part of the emitted SEs is attracted to the positively charged region of the specimen, resulting in a diminished SE signal from that region and a darker appearance. If the PE energy is just at E^{I}_{PE} or E^{II}_{PE}, no net charge will occur even if the specimen is not conductive. However, this condition will not be met once a contaminated layer is formed on the surface. E^{I}_{PE} and E^{II}_{PE} are typically 50 - 150 eV and 0.5 - 3 keV, respectively. Joy and Joy (1996) [15] reported typical E^{II}_{PE} values for materials used in the manufacture of semiconductor devices as shown in Table 1. For example, as to resist on oxide, E^{II}_{PE} = 0.9 keV. For SiO2(quartz), E^{II}_{PE} = 3 keV. The fact that E^{II}_{PE} is in the range of several keV for technically important materials such as semiconductors, polymers and ceramics is an important factor in the development of SEM-based inspection technology.

In order to suppress the influence of charging and inspect a non-conductive specimen, the entire specimen may be coated with a conductive material. Table 11.3 in Reference [19] shows the thermal conductivity and electrical resistance of several elements used in surface coatings. In some cases, a thin (nm) layer of conductive metal is deposited (coated) on the surface of the specimen to suppress the generation of a local electric field on the surface. Table 11.4 in Reference [19] shows the heat conduction and electrical resistivity of some elements and typical specimen groups. If you want to avoid using such antistatic measures and to inspect, one of the following can be used: 1) measure with as few PEs as possible, 2) use BSE with higher energy for detection, and 3) use the PEs of energy corresponding to E^{I}_{PE} or E^{II}_{PE} shown in Figure 5. In the case of an SEM-based instrument, the specific method of 1) is to reduce the current of the PE beam, to change the scanning speed from low speed to TV scan speed and to perform signal processing of the obtained images.

Recently, Takemasa et al (2018) [20] have reported a charge control method. This corresponds to method 1) described in the paragraph above. The proposed control has two scanning schemes: fast scan and special scan. The fast scan method scans faster than traditional scans. Two times and four times standard TV scan speed are available. The special scanning method is designed to control the charging effect in all directions. The details of this scanning method are not explained. They reported that applying the special scanning method to the EUV resist hole pattern was effective in suppressing errors caused by charging.

2.1.4 Simulation

Model-based inspection requires modeling what signal is obtained when the specimen is irradiated with PE. Simulation is used for this purpose. The simulation performs a calculation of the interaction between PE and the specimen. There are two different approaches to the calculation of electron-specimen interactions: wave nature and particle nature. In the case of crystalline specimens and atomic size inspections, it is necessary to take into account the wave nature of electrons. The incident PE wave interacts with the potential field of the aggregate of charged particles in the specimen. Due to this interaction, inelastic scattering changes the amplitude and phase of the electron wave. This phenomenon is calculated by the wave-nature approach.

The particle-nature approach is currently commonly used. Monte Carlo simulation is to perform discrete particle processing of the interaction between electrons and specimen using the scattering cross section and mean free path described in Section 2.1.1. The following is a brief overview of this simulation [1].

In the Monte Carlo simulation, the interaction between PE and the specimen is calculated step by step. One element is shown in Figure 7. The PEs that have reached the position P_N interact (elastic scattering, etc.), and as a result, the direction of the PE is changed by an angle θ with respect to the direction from P_{N-1} to P_N, and is rotated by the azimuth angle γ. Then PE travels linearly by the distance S and reach the next position P_{N+1}. For the scattering angle θ, only elastic scattering is considered in this overview. Since the elastic scattering angle θ can be any value from 0 ° to 180 °, the following equation determines the scattering angle θ by using the linear random number R (0 <R <1).

$$\cos\theta = 1 - \left[2 p_s R / (1 + p_s - R)\right]$$
$$p_s = 3.4 \times 10^{-3} Z^{2/3} / E_{PE}(keV)$$

The azimuth scattering angle γ in Figure 7 can take any value in the range of 0 ° to 360 °. Using another random number R, the angle γ is determined from the following equation.

$$\gamma = R \times 360°$$

Since the travel length S in one step of PE in Figure 7 is distributed around the mean value of the mean free path λ described in 2.1.1, S is determined by the following equation. In this case as well, a random number R is generated.

$$S = -\lambda \log_e R$$

From the scattering angle, azimuth, and travel length obtained in this way, the x, y, and z coordinates of the next position P_{N+1} can be calculated.

When determining the scattering angle θ, inelastic scattering is ignored. However, as PE travels a unit distance, the energy loss due to inelastic scattering is assumed to occur according to the following equation.

$$\Delta E = S \times (dE / dS)$$

This energy loss dE/dS is given by Bethe's continuous energy loss function.

Since the energy of the electron and the distance it travels are continuously known, individual inelastic scattering processes, such as inner-shell ionization or fast secondary electron production, can be calculated.

Furthermore, the electron drift and diffusion in the specimen under the influence of electric and magnetic fields can be considered in the simulation.

The Monte Carlo simulation can take into account complex boundary conditions such as specimen topography and material composition (including interfaces). Since the specimen surface positions of SEs and BSEs emitted from the specimen, their emission directions, and their possessed energies can be known using the simulation, the electrons (signals) reaching the detector can be predicted. In order to know the number of electrons actually detected, it is necessary to calculate the trajectories of the emitted electrons considering the electromagnetic field formed between the specimen and the detector.

Many simulations must be repeated to obtain significant results from the Monte Carlo simulation. For example, if the number of particles detected follows a Poisson distribution, it is necessary to consider $\sqrt{N}$ fluctuations (deviations) when detecting an average of N particles. SNR = N / $\sqrt{N}$ = $\sqrt{N}$. If you want to get a signal with SNR = 10, then N =

100. In this way, the simulations are repeated many times depending on the accuracy required for the inspection. Therefore, a long calculation time is required.

Recently, L. van Kessel et al. (2019) [21] have reported that it is not enough for the inspection of several nm nodes to consider the physical mechanisms of elastic scattering and inelastic scattering in the above simulation. They have shown that surface effects (such as surface plasmons) need to be considered in the simulation so that the simulation time becomes longer.

A simulator that can simulate the charging phenomena is required. The charging phenomenon shows a dynamically changing contrast. Arat el al. (2019) [22] have introduced two physical mechanisms into the simulation to perform dynamic charging accurately in the Monte Carlo simulation. The first is the introduction of a first-principles scattering model into the interaction mechanism between PE and specimen material. The second is the introduction of models that are dynamically coupled to the charge distribution, such as local field calculations in the specimen dielectric and surface, electron tracking in the field, and charge redistribution in the specimen material. The results show that it is possible to reproduce the dynamically changing contrast during charging. However, the calculation time required for the simulation is considerably long. Lee et al. (2018) [23] have developed a simulator with a flexible model for simulating the dynamic charging. This simulator uses the Langevin equation to simulate electron scattering in a solid so that the calculation time is reduced. Their results show that the calculation time is reduced by approximately 1/10 compared to the conventional Monte Carlo method.

Simulation helps to obtain physical quantities that are difficult to obtain experimentally, and is indispensable for studying measurement methods, that is, model-based measurement.

2.2 Scanning electron microscope (SEM)

A scanning electron microscope scans a fine beam of electrons onto the surface of a specimen to detect the selected signal emitted from the specimen. The detected signal is sent to the display device synchronized with the beam on the specimen via a video amplifier. The brightness of any point on the display device depends on the intensity of the signal from the corresponding point on the specimen. In this way, the image of the specimen is displayed on the device.

2.2.1 Spatial resolution or primary electron beam focusing

How clear and sharp the SEM image looks depends on both the PE beam current (which mainly determines the signal-to-noise ratio) and the probe size of the PE beam (which mainly determines the spatial resolution). The PE probe size d in the specimen can be estimated as follows (Goldstein et al., 1992 [24]).

$$d = \left[\frac{4I_{PE}}{\beta\pi^2\alpha^2} + \left(\frac{1}{2} C_S \alpha^3 \right)^2 + \left(\frac{0.61\lambda}{\alpha} \right)^2 + \left(C_C \alpha \frac{\Delta E}{E_{PE}} \right)^2 + \textit{defocus terms etc.} \right]^{1/2}$$

where λ is the electron wavelength, Cs and Cc are the spherical aberration and chromatic aberration coefficients of the objective lens, α is the half angle of the beam convergence in the specimen, E_{PE} is the primary electron beam energy and ΔE is the energy spread of PE, and β is the brightness of the electron source.

The above equation gives an estimate of the probe size, but not the distribution of PE in the cross section of the

beam itself. Joy and Joy (1996) [15] used a numerical ray tracing program to examine the distribution of PE at focused positions. The results are shown in Figure 8. The landing positions of 250 PEs are displayed. The energy of the PE beam emitted from the field emission gun source is (a) 30 keV, (b) 5 keV, and (c) 1 keV, respectively. At 30 keV, PEs are densely clustered around the optical axis (centre point). As a result of this numerical integration, it was shown that the profile of the current distribution (charge distribution) in the probe has a Gaussian distribution. At 5 keV, PEs are more diffused to the periphery, and fewer electrons are gathered on or near the optical axis. When it goes down to 1keV, it becomes difficult to determine the position of the optical axis. This decrease in probe formation ability is due to the effect of chromatic aberration that causes electrons to move away from the axis (centre).

Zach and Haider (1995) [25] have developed a high-resolution SEM with a quadrupole / octupole corrector dedicated to low beam energy. This corrector can correct spherical and chromatic aberrations. These aberration corrections theoretically allow a probe size of 1 nm with a beam energy of 0.5 - 1 keV. They used an experimental instrument to demonstrate less than 2 nm at 1 keV. As factors that did not reach the theoretical probe size, they cited high-frequency oscillations of electrostatic corrector power supplies, weak magnetic shields, and some low-frequency mechanical vibrations.

Recently, Cheng et al (2019) [26] have reported an SEM platform with an automatic aberration correction system. This system automatically corrects chromatic and spherical aberrations. The performance of the inspection systems using this system is evaluated with E_{PE} energy down to 100 eV. Their calculated probe sizes reduce by more than 0.4 nm in the E_{PE} range of 100 eV to 2000 eV compared to the case without aberration correction. The effect of aberration correction increases as the E_{PE} decreases. When the E_{PE} is 100 eV, the calculated spot size is reduced to less than 45% of the spot size without correction. Improved image sharpness has been experimentally demonstrated with Au particle specimens. Image sharpness was assessed by DR magnitude [27] and improved from 3.76 nm to 2.04 nm with an aberration correction at 100 eV E_{PE}. This result was close to the simulation. They stated that the difference is due to electromagnetic noise, shot noise, and mechanical vibration of the experimental platform.

In SEM, the brightness of any point on the display depends on the intensity of the selected signal (SE, BSE, etc.) from the corresponding point on the specimen. Due to the interaction of electron beam with the specimen, the apparent beam diameter (spatial resolution) will not be the same as the final probe size d. The apparent beam diameter increases by many factors such as specimen topography, specimen charging, specimen heating, vibration, and image capture process [28]. Thus, the spatial resolution deteriorates.

2.2.2 Detectors

Not all signals (SE, BSE, etc.) emitted from the specimen reach the detector. Some of the emitted signals collide with the walls of the objective lens and of the optical column. In addition, the electromagnetic field in the space between the specimen and the detector affects the signal components. In this way, the structure of the detector system including space from the specimen to the detector affects the detected signals. The Everhart-Thornley detector (ETD) for SE detection has been used for a long time. The detector pulls the SE through a high-voltage cage that causes the accelerated SEs to collide with subsequent scintillators. The resulting photons are amplified by a photomultiplier tube, converted into electrical signals, and sent to the display device. Solid silicon detectors can be used to detect BSEs. For more information on these detectors, see Wells's book [29], Reimer's book [11], or Seiler's review paper

[14]. Here, the detector systems reported in several recent papers are described.

Figure 8 shows the detector system used in the review SEM [30]. This system consists of three detectors, as shown in Figure 8(a). The upper detector is mainly used to detect SEs emitted from the specimen. This three-detector system makes the topography contrast clear, especially pattern edges (see Figure 8(c)). The detectors placed on the left and right are mainly used to detect BSEs emitted from the specimen. The shadows in the images obtained from these detectors indicate the direction of the inclination of the specimen surface (see Figure 8(c)). In addition, since BSEs are mainly detected, it is possible to obtain contrast (Z contrast) due to the difference in specimen material. In addition, the detectors in the specimen chamber can be arranged and the electromagnetic field in the chamber can be adjusted so that SEs and BSEs can be selectively detected. Figure 8(b) is a layout schematic of the two detectors on the sides. In this way, the left and right detectors are located at 45 ° to the chip layout coordinate system. With this detector configuration, both horizontal and vertical circuit patterns can be observed in good contrast with respect to the chip layout. Figure 8(c) shows an example of the images acquired by these three detectors.

Recently, Konvalina et al (2019) [31] have reported an in-lens detection system with bandpass filter characteristics. You can see Figures by entering the following URL (open access). https://www.mdpi.com/1996-1944/12/14/2307. Figure 1 in Reference [31] shows this detection system. The specimen is placed in a strong magnetic field like an immersion objective. The SEs emitted from the specimen are pulled up by a high magnetic field while drawing an upward spiral orbit in the lens column. The pulled SEs are deflected in the direction of ETD by the potentials applied to the reflecting electrode and the push electrode shown in the figure, and are detected. They investigated the bandpass filter characteristics by extensively simulating SEs and BSEs with a precise 3D model of this detector. As a result, it was seen that this detector has an energy window of approximately 3 eV and thus functions as a bandpass filter for SEs. This window can be shifted on the energy axis, showing that the energy selection of the detected SEs is possible. This in-lens detection eliminates the detection of SE (3) in Figure 3 and detects only the true SEs (SE (1) and SE (2)), thereby improving the signal-to-noise ratio of the SE signal. Note that this detector is not be suitable for inspecting specimens that generate high magnetic fields.

Suri et al. (2020) [32] have developed an SE detector that has a bandpass filter function. This detector is called the Bessel Box (BB) electron energy analyzer. You can see Figures (but small) by entering the following URL (not open access). https://www.sciencedirect.com/science/article/pii/S0368204818301853. Figure 4 in Reference [32] shows a schematic diagram of the BB detector (energy analyzer). The BB detector consists of a central cylindrical electrode and the two end cap electrodes (Input Electrode and Output Electrode in the figure). Controlling the potentials on these electrodes allows the BB detector to act as a bandpass filter. The name "Bessel Box" comes from the fact that the fields in the cylindrical structure depend on the modified mathematical Bessel function. The input electrode is grounded and the same negative potential is applied to both the output electrode and the central cylindrical BB electrode. As a result, as shown in Figure 4(a), SEs with a certain energy are focused in the hole of the output electrode. SEs with energies lower than this passing energy are repelled as shown in Figure 4(b). For SEs with high energy, as shown in Figure 4(c), they are terminated at the output electrode without focusing. Thus, only SEs with energy within a narrow band can pass through the hole in the output electrode and are detected (Figure 4(a)). Figure 7 in Reference [32] shows the experimental setup of the BB detector. The BB detector and the electron gun are orthogonal to each other. The specimen is mounted at 45 ° to the axis of the BB detector and the axis of the electron

gun. The BB detector is coupled with a count mode channel electron multiplier tube. They have demonstrated a compact BB detector (12 mm^3 volume). This detector with a calculated energy resolution of 0.3% (for example, half-width 30 meV when the peak path energy is 10 eV) is experimentally verified. Thus, detecting only SEs with certain energies can not only change the contrast of the image, but also display details that cannot be seen otherwise.

For a dedicated detector for detecting BSEs, see Niedel's Review Paper (1982) [33]. Here we introduce the recently reported BSE detector systems. Lin et al. (2018) [34] reported a detector consisting of silicon p-n diodes in a multi-annular configuration to detect BSEs. You can see Figures by entering the following URL (open access). https://www.mdpi.com/1424-8220/18/9/3093. Figure 1(a) [33] shows the BSE semiconductor detector used in commercial SEMs that consist of four sectors. The BSE signal incident on the four sectors becomes independent signals and is used for later signal processing (addition and subtraction). Lin et al. (2018) [34] integrated the multi-annular detectors with a lateral p-n junction and aluminum grid structure to increase the sensitivity of Z contrast in BSE images and to improve surface topography contrast. Figure 1(c) [33] shows their BSE detector that was designed and manufactured to detect BSEs in the range of 5 keV to 8 keV. Experimental results show that BSEs in the range of 0.5 keV to 1 keV can also be detected. They claim that the signals from the individual p-n junctions of the detector can be added or subtracted to some extent to distinguish between topography and Z contrast.

Rau et al. (2019) [35] have reported the optimal design of the BSE detector to more reliably distinguish between topographic contrast and Z contrast in BSE images. It can also be used to obtain images of various depths within the specimen volume. They developed a BSE detector consisting of four inner Si plates and four outer Si plates (see Figure 5 in Reference [35]). Si plates consist of a planar p-n junction. The inner plate is tilted at a 30 ° angle in the plane of the detector, and the outer plate is tilted at a 60 ° angle. The BSE-collected solid angle Ω1 of each internal detector is approximately equal to the solid angle Ω2 of the external detector, considering the decrease in the number of BSEs as a function of the BSE emission angle θ. As a result of the experiment, they reported that the contrast of the image is improved by approximately 1.5 times.

In-lens BSE detectors have also been developed. RadliCka et al. (2018) [36] reported an in-lens detector with energy filtering that enables the detection of low-loss energy backscattered electrons (energy close to E_{PE}). You can see Figures (but small) by entering the following URL (not open access): https://www.sciencedirect.com/science/article/abs/pii/S0304399117304187. Figure 2 in Reference [36] shows the design of the BSE detector. BSE is pulled up by the leakage magnetic field of the objective lens near the specimen and enters the objective lens. By design, BSEs with an emission angle of 40 ° or less enter the objective lens. Inside of the objective lens, (a) shielding electrode, (b) first focus electrode, (c) second focus electrode, (d) filtering grid, and (e) BSE detector are installed. Only BSEs that have passed through the filtering grid (d) can reach the detector. In the figure the trajectories of approximately 2 keV BSEs emitted from the specimen irradiated with E_{PE} 2 keV (landing energy in figure) are shown where a potential of -2 kV is applied to the filter grid (d). Note that the grid planes are not equipotential planes, so even BSEs with energies less than 2 keV can reach the detector. In this way, the BSE energy to be detected is selected according to the potential applied to the filter grid. BSE orbital simulations have shown that the energy window goes from 100 eV to 10 eV with a reasonable collection efficiency.

2.2.3 Noise

Whether or not the signal at each irradiation point (one pixel) obtained by the SEM-based inspection device is a valid signal depends on the amount of noise contained in the signal. The signal-to-noise ratio (SNR) is typically used for this evaluation. SNR is defined as the ratio of the average signal to the fluctuation (noise) of that signal. If the SNR is above a certain criterion, the detected signal is judged to be valid. Assuming that the signal of one pixel consists of n electrons on average and the fluctuation of the detected electrons follows the Poisson distribution, the noise is $\sqrt{N}$ and SNR = $\sqrt{N}$. If a signal with an SNR of 5 or higher is desirable, it is seen that an average of 25 or more electrons should be detected.

In the SEM-based inspection instrument, the specimen is irradiated with PEs emitted from an electron gun, and as a result, SEs or BSEs are emitted, a part of which reaches the detector, is amplified, and is finally transmitted to the display device and becomes a signal. It is important to know which part of such a signaling path determines the SNR of the final signal. Figure 10 shows a signal transmission path using a standard ETD detector [29]. The vertical axis in Figure 10(b) is the number of electrons in one pixel in this transmission path (described as signal quanta in the figure). The horizontal axis shows the location of the transmission path shown in Figure 10(a). For example, "1" is the number of PEs irradiated to the specimen that becomes a one-pixel signal.

In the case of Figure 10, the SNR of the final signal is mainly determined by the position 2, because the number of electrons is the smallest at the position 2 of the transmission path [29]. Such a position is called the noise bottleneck because it determines the final SNR. For SE detection, the noise bottleneck lies between the specimen and detector. Timischl et al. (2012) [37] have calculated SNRtot (k) at each stage of the transmission path by using the realistic expectations of the commercially available SEM JSM-6610LV. "k" indicates the location of the transmission path. You can see Figures by entering the following URL (Wiley Online Library): https://onlinelibrary.wiley.com/doi/full/10.1002/sca.20282. Figure 5 in Reference [37] shows the result. It is seen that SNR on the second stage is close to the final SNR. This result is consistent with that in Figure 10 [29].

Sakakibara et al. (2019) [38] developed a method for measuring fluctuations in the number of SEs emitted from a specimen. The noise contained in the final SE signal can be minimised by detecting only the SEs (SE (1) and SE (2)) emitted from the specimen. An experiment was performed in which a Si wafer specimen was irradiated with 500 eV PE, and the fluctuation of the emitted SE was measured. As a result, it was 1.9 times the value obtained assuming the Poisson process. The value in this experiment agreed with the value in the Monte Carlo simulation. Therefore, they reported that the noise contained in the signal can be predicted by simulation.

Recently, Bunday (2020) [39] has considered how noise distribution changes spatially by simulating various specimen structures that will allow us to understand the noise dependencies on material and local geometry. His proposed hypothesis is that when the interaction volume is in positions interacting with multiple materials or features, the noise tends to approach a Gaussian distribution (see Figure 10 in Reference [39]). This consideration is needed when adding noise to analytically generated images.

2.2.4 Image contrast (imaging modes) and signal processing

Recent SEMs use a digital scanning scheme to capture an image of the selected signal. This image allows us to observe the various contrasts described in Section 2.1.2. The raster scan scheme is well used. In this scheme, after acquiring the signal for one line (the position of PE is on the right end), PE is returned to the left end and is shifted downward by one pixel. By repeating this operation, an image is acquired. The time required for this swing back is also taken into consideration. The

signal amount of one pixel depends on the time that PE stays in one pixel. The acquired signal amount of one pixel is converted into a digital signal of several bits. In this way, a digital SEM image is obtained. As a result, powerful digital image processing can be applied to the acquired images.

Lenthe et al. (2018) [40] have identified two problems in the digital raster scan scheme. One is the shift errors in the vertical direction between horizontal line scan. The other is scan stabilisation errors that occur due to scanning parameters and microscope hardware response time. They reduced the shift errors by incorporating a high resolution, high speed IO controller and using a snake scan scheme. You can see Figures (but small) by entering the following URL (not open access): https://www.sciencedirect.com/science/article/abs/pii/S0304399118300305. Figure 2 in Reference [40] shows the raster and snake beam scanning schemes. As to the second, they reported that scan stabilisation error can be corrected using time-resolved fast sampling detector.

Papavieros and Constantoudis (2017) [41] have reported an issue of image digitisation for measureing the line edge roughness. They have focused on the discretization of edge point coordinates in the direction across the edge/line axis of line and space patterns, i.e. the discretization of edge point ordinates. They analysed the effect using synthesised SEM images so that the critical role of the ratio of the roughness metrics of the edge to the pixel size is identified. Its mitigation method was discussed. Without mitigation, the roughness measurement deteriorates with the large amount of measurement noise due to digitisation. This deterioration biases the measurement and blurs the true roughness with measurement noise, overestimating the true roughness.

To accurately inspect nanometer-sized specimens with SEM-based instruments, it is necessary to suppress the effects of noise. The process of integrating images from the same sample can reduce noise, but due to the increased number of PEs irradiating the sample, the inelastic scattering events that break chemical bonds can damage the specimen. Alternatively, in the low electron dose, noise begins to dominate and the edges of the image pattern become obscured. Therefore, the maximum allowable electron dose inherent in the specimen limits the signal-to-noise ratio (SNR) and spatial resolution. In the following, recent studies on digital image processing for denoising are described.

Noise reduction techniques using digital image processing can be divided into two main categories: the classical approach and the deep learning approach. The former approach includes filtering techniques in real space and in transformed space (e.g., spatial frequency space using Fourier transforms) [42]. Filtering in real space takes advantage of the high correlation of adjacent pixels to filter the value of each pixel from the values of the adjacent pixels (spatial mask). It can be further divided into linear (such as averaging) filters and non-linear filters (such as median). As in the case of real space, the filtering method in the transformed space (frequency space) can be thought of as a frequency space mask. These filtering methods have a smoothing effect on the image. For more information, see Roy's review paper [42]. Real-space filtering techniques or transform-space filtering techniques reduce noise, which can bias the edge position measurements. Therefore, for a bias-free measurement, it is desirable to perform edge detection from the measured image without using image filtering.

Mack et al. (2019) [43] have reported a method for measuring edge position without bias. Their approach is to fit the analytical line scan model (ALM) [44] to the actual measurement data to determine the edge positions. ALM considers the physics of electron scattering and secondary electron generation. By adapting the ALM to a rigorous Monte Carlo simulation, the parameters contained in the ALM are determined. In addition, it is possible to calibrate ALM parameters for specific actual inspection data. Through the calibrated ALM, the left and right edge positions

that optimally fit the model to the detected data can be obtained without using image filtering. The detected edges include the roughness of the actual line pattern and the effects of noise from the image. The noise of the image is statistically removed from the power spectral density of the obtained edge roughness to obtain the true unbiased roughness.

Several noise reduction methods using deep learning (DL) approaches have also been reported. Giannatou et al. (2019) [45] have used a Convolutional Neural Network (CNN) based on the DnCNN [46] architecture for image denoising. Instead of outputting a noise-free image directly, it is designed to predict noise at each pixel of the observed image and output an estimated residual image. Then, the estimated zero noise images can be obtained by subtracting the estimated residual image from the high noise image. The DnCNN model is trained with a composite SEM image set with known Gauss and Poisson noise. Midoh and Nakamae (2019) [47] have used the pix2pix DL architecture [48]. The results of the simulated noisy SEM image show that stable LER measurements are possible regardless of the noise level (within consideration). In the actual SEM image, the line width roughness was reduced to approximately 1/4. Liangjiang et al. (2020) [49] proposed an unsupervised machine learning-based image quality enhancement framework that uses only noisy SEM images for training. It is based on the fact that the statistical expectations of noisy observations are the same as that of the clean signal [50]. The framework has been tested with both simulated and real SEM images. The results showed the effectiveness of the proposed framework.

3 Application to semiconductor inspection

3.1 Mask inspection

Most advanced manufacturing facilities are adopting extreme ultraviolet (EUV) lithography as technology nodes shrink to 7 nm and below. This size scaling further reduces the killer defect size on both the wafer and mask. SEM based inspection systems have the advantage of high spatial resolution of images. However, it has the disadvantage of low throughput, but it can be mitigated to some extent by using a projection electron microscope (PEM) with a wider illumination beam or by using a multi-beam SEM type inspection system.

Iida et al. (2016) [51] have investigated the defect detectability of EUV masks by comparing PEM and SEM inspection systems. See Figure 1 (a schematic diagram of the PEM inspection system) in Reference [51]. In PEM, PEs covering a large area illuminates the mask surface, and the emitted SEs reach the detector via imaging electron optics. When 16 nm size defects were inspected, compared with both PEM and SEM for a line & space patterns of half-pitch 64 nm, it was found that both were within the allowable range of less than 2 nm line edge roughness (3σ). They claimed that the throughput of PEM inspection systems corresponds to that of multi-beam SEM-based inspection systems with 200 to 1850 beams.

Ma et al. (2018) [52] have developed a multi-beam inspector (MBI) to meet the industry's needs for EUV lithography. See Figure 4 (a schematic diagram of a multi-beam electro-optical system) in Reference [52]. The optical system consists of an aperture array that divides one beam source into multiple beamlets, a double condenser lens that avoids image rotation, and an electro-optical module manufactured with MEMS technology that controls each beamlet individually. In addition, a Wien filter is attached to the primary beam module to bend the SEs emitted from the specimen towards the detector. Then SEs from different beamlets are projected onto the detector array. They have validated the operation of MBI with 9 beamlets.

The first comprehensive assessment of the EUV mask infrastructure was presented at the 2015 BACUS

Photomask Conference. There, the Actinic Pattern Mask Inspection (APMI) was highlighted as an infrastructure gap. Since then, great progress has been made in all areas, especially the development of pellicle and the deployment of APMI for EUV mask production at Intel's mask shop. The wavelength of actinic light is 13.5 nm. Electron beam mask inspection (EBMI) was developed as a temporary measure before APMI became available. EBMI can provide the required spatial resolution, but is less sensitive to defects that do not have strong Z-contrast or topography profiles such as defects in multi-layer (ML) layers, and have low throughput. Recently, Liang et al. (2020) [53] have reported that APMI can detect defects that cannot be inspected by DUV optical inspection instruments or EBMI. For the first time, they showed that APMI enables though-pellicle inspection of patterns on pelliclized EUV masks. With the advent of APMI, EBMI will play a limited role.

3.2 Detection, review, and automatic classification of defects and particles

Device manufacturing technology is continually advancing through the introduction of new processes and materials to reduce the size of circuit patterns in semiconductor devices. Besides, the defect size, which has a great influence on the operation of the device, is getting smaller and smaller. Therefore, rapid detection and analysis of defects in the manufacturing process and appropriate countermeasures are more important for rapid yield improvement and stable manufacturing process (Figure 11).

In-line inspection systems typically consist of optical wafer inspection instruments and scanning electron microscope (SEM) -based review instruments, which are deployed at semiconductor wafer manufacturing sites for process monitoring. Optical wafer inspection instruments are used to quickly detect defects and particles on wafers. Such instruments capture images of the wafer surface and compare them with images of adjacent chips (chips are formed in an array on the wafer). The pixel size of these images is on the order of tens to 100 nm, which is insufficient to determine the presence of defects. Detected defects are reviewed and analysed using SEM-based review instruments with nanometer-order resolution. Defect review/analysis includes manual review of defect appearance, elemental composition analysis, and defect classification based on size and type. Next, the review/analysis results (defect root cause analysis) are used to statistically analyse the tendency and frequency of defect occurrence to determine the priority of countermeasures.

SEM-based review instruments that automatically collect defect images with a high throughput and classify them with a high degree of precision are required in order to the efficiency of the defect review/analysis process.

The automatic defect classification (ADC) is a function that automatically classifies a defect region in image into a predetermined defect class based on the appearance of the defect when it is recognized as having a defect. Figure 12 shows the general flow of ADC processing. The image to be inspected and the reference image of the defect-free area are input to the ADC. First, the ADC recognises the defect area and circuit pattern area from these two images. For the defect areas, various defect characteristics are extracted and described as feature vectors. The feature vector includes appearance features (defect size, length, etc.) and the positional relationship between the circuit pattern and the defect. Then the defect regions are classified into one of several predetermined classes using their feature vectors. The classes are defined using defect type, defect criticality, and a combination of both indicators. The defect criticality is calculated by the positional relationship between the circuit pattern and the defects.

Nakagaki et al. (2009) [30] reported a defect area recognition technique using the detector system in Figure 9,

where three types of SEM images are effectively utilized. As mentioned in the 2.2.2 section, the three-detector system allows selective detection of SEs and BSEs emitted from the specimen. In the proposed method based on the comparison of the inspection image and the reference image, the difference images between the three types of images are calculated, and the ratio of mixing the three difference images is determined according to the surface roughness analysis result. The final difference image is obtained using this mixing ratio, and the defect area is obtained from the threshold procedure. Shading information is used for surface roughness analysis, and the roughness characteristics have two indexes: defect height level and unevenness. As a result of applying the proposed method to 200 images selected from nine defect classes contained in two processed wafers, they have verified a recognition accuracy of 98.9%.

Nakagaki et al. (2010) [54] also reported on the recognition of the circuit pattern area using the detector system shown in Figure 9. Their method takes advantage of the fact that the image of the circuit pattern depends on the pattern structure, pattern material, and pattern layout. The method consists of three parts: pattern density evaluation, segmentation-based pattern recognition, and threshold-based pattern recognition. These three algorithms are used adaptively depending on the specimen inspection image. Experimental results on five wafers (including 421 circuit pattern images) showed that the proposed method can automatically recognise circuit patterns with an accuracy of 99.8%.

As the technology nodes shrink below 7nm, killer defect sizes on both wafers and masks become even smaller, making it difficult for SEM-based inspection systems to automatically capture clear images of different types of defects. Harada et al. (2017) [55] have proposed an auto-focus algorithm using multiple band-pass filters. They proposed an index "focus measure" that indicates the focussing state of the PE beam and a method for designing a bandpass filter. To achieve an optimal specific frequency response for various shape defect images, multiple band-pass filters are introduced. In the proposed method, two series of focus measures are calculated using multiple band-pass filters independently, and the optimum focussing position (conditions) is automatically selected according to the reliability of the series of focus measures. Also, the signal-to-noise ratio of an image for acceptable auto-focus precision is determined through simulation images. Experiment showed that the success rate of auto focus was improved from 79.4% to 95.6%.

As critical defect sizes become smaller and smaller, highly sensitive optical wafer inspection instruments used in the first phase of defect detection mistakenly detect process fluctuations as defects and generate large amounts of "nuisance" information. Figure 13 shows the process flow from the optical wafer inspection instrument to the ADC. The coordinate information obtained by the optical wafer inspection instrument contains a position error of approximately $\pm$ 1 μm. Therefore, if a high magnification (e.g. 300 000) image is taken after moving the stage, there may be no defects in the image (about 0.45 μm). To work around this issue, a low magnification image (e.g. 50 000; 2.7 μm field of view) is obtained centered on the defect location detected by the optical inspection instrument. The defect location is determined by comparing the image to be inspected with the image obtained on the adjacent chip (see Figure 12). Next, a high-magnification image (observation image in Figure 13) is acquired at the determined defect location.

In order to automatically discriminate between defects and nuisance information, Harada et al (2019) [56] have proposed a true defect detection method through 2 steps. In the first step, multiple reference images are used to decrease the number of defect candidates since true defects are detected even compared to any reference image. In the second step,

the inspection image is divided into patch images, and the presence or absence of defects is determined by comparing each patch with patch image models without defects. Defect-free patch image models are created and stored in advance from comparison results with a large number of reference images. Experimental results showed a nuisance image discrimination rate (true negative rate) of 84.4% and a defect detection rate (true positive rate) of 93.3%, which are higher than those of the one-class support vector machine.

Recently introduced EUV lithography technology can cause unexpected defects. Therefore, a faster inspection is desired to inspect the entire wafer surface. Inspection methods based on image comparison include the Die-to-Database (D2DB) scheme as well as the Die-to-Die (D2D) scheme used above. Unlike D2D inspection, D2DB inspection does not require a reference image because it compares the inspection image with the design layout, the high-speed inspection is possible. However, in conventional D2DB inspection, the circuit pattern is not produced as designed in the manufacturing process, and the deformation is detected as a defect. Ouchi et al. (2020) [57] proposed a deep learning-based D2DB inspection that can distinguish between deformations due to manufacturing processes and defects. Their model is a CNN model including several encoder and decoder layers and predicts the mean and standard deviation images as the luminosity distributions from an input design layout. The predicted mean and standard deviation images can specify the luminosity distributions on the inspection image. The model was trained with pairs of an inspection image and a design layout. Thus, the trainable D2DB learns acceptable deformation on a circuit as the luminosity distributions, which are defined as the pixel-wise probability density distribution on inspection images. The trainable D2DB directly judges whether each pixel on an inspection image belongs to a defect or does not. Therefore, the trainable D2DB can inspect low-resolution inspection images and detect extremely small defects consisting of several pixels. Results show that it successfully detected all defects without mis-detection at 1, 2, 3, and 4 nm/pixel resolutions. When using the inspection image of 4 nm/pixel resolution, high-speed inspection whose speed is 1.25 hour/mm^2 can be realised. This inspection is more than ten times as fast as conventional e-beam inspection. Because their proposal requires only normal images, the model can be trained without defect images, which are difficult to obtain with enough variety.

The ADC extracts various types of defect features (defect size, length, height, texture, etc.) from the defect image and stores them as feature vectors for classification. The ADC then performs a classification process from the feature vector of each defect. Rule-based and learning-based methods have been proposed for the classification process. The former classifies defects according to pre-determined rules. This method has the advantage of not having to collect samples in advance, but has the disadvantage of being inflexible in changing classification criteria. The latter classifies defects based on the statistical properties of the feature vector and refers to training data created from pre-collected specimen defects for each class. This method is relatively compliant with changes in classification criteria, but is unstable in situations where it is difficult to collect large amounts of specimen data, especially in the case of killer defects. Several ADC machine learning (ML) methods are also being considered. However, ML algorithms tend to generate models that prefer defect classes with a majority sample size. Another challenge for ADCs is the lack of labelled data. Annotating data (labelled data) is laborious and requires special expertise to correctly identify defects. This problem is known as an "unbalanced dataset". Among various solutions for the imbalance, the Synthetic Minority Oversampling Technique (SMOTE) [58] has shown its strength in ML when dealing with imbalanced datasets. The SMOTE use a combination of over-sampling the minority (abnormal) classes and under-sampling the majority (normal) class. In over-sampling, the minority classes involve synthetic minority class examples. Lee et al. (2020) [59] have performed experiments to test the performance of ML methods such as Random

Forest, AdaBoost, XGBoost, and SVM on two imbalanced datasets with and without SMOTE. Results show that the ML algorithms combined with SMOTE using 25%-75% of the total training data achieved higher average class-wise accuracy than the same classifiers trained with 100% data but without SMOTE. They claimed that SMOTE can improve ADC performance for imbalanced datasets. It also reduces the number of ground truth samples required for training, making it easier to apply the ADC if there are not enough human resources to label the ground truth data.

3.3 Critical dimension (CD) measurements

3.3.1 Line edge roughness (LER)

As semiconductor device scaling continues with the introduction of three-dimensional (3-D) structures, device performance and reliability are becoming increasingly sensitive to the size and shape of local parts of circuit functionality. In addition, the number of such local features on the wafer reaches 1 trillion per device layer, and their fluctuations have a significant impact on manufacturing yield. Therefore, controlling process variability is as important as process-centric control. For pattern inspection, it is necessary to monitor variations such as line edge roughness (LER) and local critical dimension uniformity (LCDU) as well as the average size. Causes and effects of LER range from short-period (intra-feature) to medium/long-period (between inter-feature) fluctuations and cause fluctuations and degradation in the device performance and reliability. Various attempts have been made to properly evaluate the LER and improve the material and patterning process. Most LER comes from the probabilistic behaviour of processes and materials. A real semiconductor device contains many of the smallest size features of the same design, and as the 2D size of such features approaches the fluctuation period, size fluctuation LCDUs are observed in those features.

Fukuda et al. (2018) [60] have reported general guidelines and practical solutions for LER measurement using CD-SEM. First, they discussed LER measurement from the four aspects: (1) edge detection algorithm, (2) power-spectral-density (PSD) prediction algorithm, (3) sampling strategy, and (4) noise mitigation method. Then, they extended some of the above results to LCDU measurement, focussing on noise impact on measurement bias and repeatability. General guidelines and practical solutions for LER measurement are summarised as follows. 1) Advanced algorithms such as the wave-matching method detect pattern edges from low SNR (small frame number) image pictures and reduce the influence of SEM noise while minimising specimen damage. 2) Weak filtering (2 to 4 pixels along the edge depending on pixel size) suppresses SEM noise and aliasing by minimising artefacts in observed LER. However, it is difficult to visualise residual SEM noise and special care is needed in predicting apparent noise floor level. 3) The multitaper method is effective for suppressing the sampling noise within an individual line-edge but still requires multiple lines (or image pictures) for covering line-line variation. 4) Over-sampling and weak high-cut filtering are effective for suppressing aliasing. 5) Two types of SEM noise (LER bias) mitigation methods are effective, “apparent noise floor” subtraction the method and LER-noise decomposition method.

Extreme ultraviolet (EUV) lithography was introduced for the high-volume manufacturing of state-of-the-art semiconductor devices in 2019. One of the issues for the CD metrology of an EUV resist pattern is electron irradiation damage to the resist that is known as “shrinkage”. Since the ratio of shrinkage to CD increases in EUV lithography compared with that in immersion argon fluoride (iArF) lithography, the low-shrink metrology with a high spatial resolution needed in EUV lithography is required [61]. Reducing the irradiation energy of a PE in a SEM is effective

to reduce the resist shrinkage. However, reducing the energy of PEs degrades spatial resolution. The use of low-energy SEM with aberration corrector has been proposed [26], but the aberration corrector has complex power control issues and will not provide a solution to the metrology of the EUV resist in the short term. Suzuki et al. (2018) [5] have shown the possibility of reducing resist shrinkage for the thinner EUV resist by using high-energy PEs. Bizen et al. (2019) [62] have reported that the CD metrology of an EUV resist pattern with low shrinkage and high resolution can be realised using PEs with a high energy of 4000 eV. The LER obtained for those with an energy of 4000 eV was approximately the same as that for those with energies of 200 eV and 800 eV. The electron irradiation damage for an under layer and the amount of shrinkage depending on the pattern size could cause issues.

Pu et al. (2019) [63] have reported accurate quantification of the amount of line-edge roughness (LER) by means of power spectral density (PSD) analysis as well as other features such as correlation length, roughness exponent, and unbiased standard deviation. The observed PSD is affected not only by the inherent roughness of the measured features but also by external factors such as SEM shot noise, SEM instrument stage movement, and SEM beam profile. They suggested a method to correct the PSD for the effects of the SEM beam profile to improve the accuracy of the LER measurement. In the simulation, such a process is easy to perform, but in practice it can be difficult to have enough information about the SEM beam profile.

Kessel et al (2020) [64] have reported the influence of sidewall roughness (SWR) on LER measured with the top-down CD-SEM. They investigated by measuring the LER from a simulated inspection image with a known SWR. Their results show that the popular method of PSD analysis works very well if the structures have no roughness in the vertical direction (line direction). The main factor limiting the measurement of true pattern PSD is the PE beam profile. The beam profile blurs the inspection image and suppresses the high spatial frequency components of the PSD. If the structures are isolated and rough in the vertical direction, the CD-SEM observes the outermost edge over the full height of the structure. For dense lines and spaces, the situation is complicated since SEM is less sensitive to the lower layers.

3.3.2 High-Aspect Ratio (HAR) contact holes

As mentioned earlier, device structures are changing from planar structures to complex three-dimensional structures such as 3D-NAND. One of the features of 3D-NAND is the high aspect ratio (HAR) channel hole. The aspect ratio is defined as the ratio of depth to top diameter of the channel hole. This aspect ratio has increased to over 50: 1. One of the challenges in manufacturing high-aspect-ratio holes is to etch holes (or trenches) from top to bottom without distortion. Mass production requires uniform etching of over 1 trillion holes in all wafers. Therefore, in-line measurements to measure the typical features of HAR holes, such as lower CD uniformity and upper and lower centre overlays (OVL), are essential for yield improvement. Recently, there has been a strong demand for in-line measurements of cross-sectional or 3D profiles of HAR holes with complex shapes such as bending and twisting. Although many techniques have been considered, inspection techniques that use BSEs in SEM-based instrument are widely accepted in HAR hole inspection because of the linear trajectory of high-energy BSE and the increase in yield with atomic number.

As a small amount of residue may remain after the etching process at the bottom of the HAR hole, Han (2018) [65] has investigated the BSE inspection conditions for this residue by using simulation. The presence of residue can cause serious disruption to the device operation. He showed that high-energy BSE mode improved the detection sensitivity of 2 nm thick

residues by more than 10-fold. Also, he proposed to use an energy window in the energy range of PE energy to approximately 20 eV lower than PE energy for inspection.

Nishihata et al. (2019) [66] have reported a technique for measuring HAR hole depth using a high-voltage SEM. BSE generated by high voltage SEM includes BSE with high energy. This high voltage BSE is detected because its high energy allows it to penetrate the material around the HAR hole, even if it occurs at the bottom of the HAR hole. They analytically modelled the relationship between the transmitted BSE strength, the hole depth, and the hole diameter. The two unknown parameters contained in the analysis formula are determined by adapting two or more of the bottom BSE signal intensity, hole depth, and surface density. Based on this model, they proposed a depth measurement method that uses only the BSE image on the top surface. Channel hole measurement errors in 3D flash memory devices with a depth of a few microns were assessed using 30 keV PE. According to the evaluation results, the error range was 62 nm and the measurement reproducibility was ± 18 nm.

Sun et al (2019) [67] have also investigated the top down BSE imaging of the HAR holes by using high voltage SEM. They also modelled the relationship between BSE intensity and the PE irradiation depth location in the hole into a mathematical formula. One unknown parameter is the attenuation coefficient μ that means the probability of a single electron being scattered within a distance of dh. The value of coefficient μ is obtained by adapting the simulated BSE signal intensity into the formula. They showed that the BSE signal intensity decreases in the same trend for HAR holes with different depths. Also, the higher is the PE energy, the smaller the value of μ. In addition, they showed a weak dependence of μ on the depth of the hole by using high-energy PE, such as 45 keV. See Figure 4 in Reference [67]. Thus, holes with different depths can be measured using such high energy PEs and the same value of μ. Based on the model, they proposed a 3D profile measurement algorithm and experimentally verified the feasibility of the proposed algorithm, where several kinds of HAR holes with different positive taper angles and different bowing profiles were fabricated.

With a large number of high aspect ratio (HAR) hole formations, slight deviations in etching conditions can result in the formation of diagonally curved holes. See Figure 1 in Reference [68]. For feedback to etching and other processes, it is important to measure the top CD, the bottom CD, and the slope and bend of hole. However, if the pattern is greatly bent, the bottom CD cannot be measured correctly because part of the bottom is blocked by the top. Tu et al. (2020) [68] have developed and demonstrated an automatic electron beam tilting function to overcome this problem. They used three detector systems that consisted of top BSE, upper SE, and lower BSE detectors. See Figure 2 in Reference [68]. BSEs are captured with the lower detector under the objective lens and the upper detector above the lens. The SEs are captured with the middle detector above the objective lens. The procedure of automatic electron beam tilting is as follows. The primary electron beam is irradiated at different angles of incidence, and the beam tilt correction coefficients in the X and Y directions are obtained, respectively. Next, the tilt angle of the primary electron beam is adjusted to be parallel to the tilt hole pattern. Therefore, the best angle for detecting the bottom signal at high PE energy is obtained. In this way, the exact bottom CD and the angle and orientation of the bent holes can be automatically measured. Thus, quick feedback to the etching process can be realised.

Sun et al. (2020) [69] have reported 3D profiling of HAR holes using a deep learning approach. They simulated many BSE line profiles for learning. The simulation considers the effects of electron beam aperture angle and aberrations. Next, they trained a deep learning model with these simulation data. Its learning objective is to enable mapping from the BSE line profile to the HAR hole cross-section profile. Two one-dimensional neural network architectures: CNN and Multiscale

CNN (MS-CNN) are trained and different loss functions are investigated to optimise the network. They said that the cross-sectional profile predicted by the MS-CNN model is closer to true compared to the previously proposed exponential model [67].

3.3.3 3D surface reconstruction

The SE and BSE signals contain a signal component (topography contrast) that reflects the shape of the specimen, as described in Section 2.1.2. The three-dimensional structure reconstruction methods of the specimen surface using this property have been studied for a long time. Tafti et al. (2015) [70] review a 3D surface reconstruction system from SEM images. They categorised 3D surface reconstruction algorithms into three classes: single view, multiview, and hybrid. The single-view approach uses a 2D image set from a single viewpoint. The multi-view approach uses 3D computer vision algorithms. This method uses a stereo pair taken with the specimen tilted. The corresponding feature points are searched in the image pair and the 3D points are reconstructed. The reconstruction algorithm called the hybrid method takes advantage of both the single-view and multi-view approaches. In addition, there are methods such as focused ion beam-scanning electron microscope (FIB-SEM) and electron tomography (ET), which require the operation of gradually thinning the specimen or dividing the specimen into slices. This operation results in specimen destruction.

Villarrubia et al. (2016) [71] have evaluated three commercial stereo photogrammetric software packages for 3D reconstruction using simulated image pairs. The basic shape of the specimen for simulation is a line pattern with a trapezoidal cross-section with rounded upper two corners. In addition, a surface roughness with an arbitrary power spectral density is added on the surface of this basic shape. Three software packages are evaluated using simulation SEM image pairs of this specimen. As a result of comparing the 3D model reconstructed with software with the known specimen shape, two of the three software packages reconstructed the 3D specimen within an error of 1 nm. They state that all packages have shown some difficulty in reconstructing the surface roughness details.

Valade et al. (2019) [72] have introduced a PE beam tilting mechanism for reconstructing a 3D shape from multiple SEM images. The tilt (deflection) of the PE beam is performed through the magnetic field generated by the coil at the outlet of the column just above the specimen. Their experimental results are shown in Figure 19 of Reference [72]. The results are consistent with the traditional 3D measurement methods. Their experiment showed a three-sigma standard deviation of 10 nm on the estimated height for heights ranging from 50 nm to more than 200 nm.

Neumann et al (2020) [73] have carried out 3D tomography with FIB-SEM to analyze the 3D structure of the device. The specimen is milled with FIB and the surface appeared is observed with SEM. After that, further milling is performed to obtain an SEM image. Repeating this procedure will result in a series of 2D images. A series of 2D images can be rendered into a 3D volume from the information on the amount of each milling (thickness scraped). Thus, the 3D information about the structure at each position in the 3D volume is obtained.

3.4 Overlay

Semiconductor devices are manufactured by stacking multiple layers. In each layer, the patterns of materials required for the function of the device are arranged. The pattern of each of these layers is properly aligned and connected as

designed to the lower and upper layers for the final device to function properly. For example, the contacts, signal wires, and all transistors are aligned to properly function. Overlay control, which directs such placement over the underlying layer, plays an essential role during manufacturing. If there is any misalignment, it can cause short circuits and connection failures [74]. Overlay control is even more demanding for technology nodes of a few nm. Scanning electron microscope-based overlay (SEM-OL) inspection has long been studied as an overlay measurement solution with higher measurement precision/accuracy.

Inoue and Hasumi (2019) [75] have reviewed SEM-OL metrology applied in CD-SEM and high-voltage SEM (HV-SEM). They summarised mechanisms and factors for overlay measurement in Figure 1 in Reference [75] schematically.

Abramovitz et al. (2020) [76] have proposed a flow of SEM overlay verification using inline FIB that was performed in a full-scale wafer and a short time to results. In the flow, first the wafer overlay was measured using the EB instrument with high energy PE. Then a single step milling was performed on the precise location of interest with the inline FIB in a full wafer scale. Finally, the wafer was returned to the EB instrument to measure overlay with low energy PE.

SEM-OL measurements are expected to become a complementary or alternative technology to traditional optical measurements for overlay control.

3.5 Design for manufacturing (DFM)

Design for Manufacturing (DFM) is the creation of rules (called design rules) that take into account manufacturing capability so that devices can be manufactured as designed. For the advanced nm technology node, DFM rules are more complicated. As a result, semiconductor foundries analyse the importance of each part of new design layout before starting production. Babu et al. (2020) [77] have proposed a layout risk assessment flow. The flow queries the existing database to ascertain if the new design contains risky patterns. The database stores information about all design patterns, design weaknesses, and defects. The defect data are obtained through review SEM and so on. This database is utilised for machine learning algorithm development to analyse new designs. For example, a supervised ML model using both design (width, space, etc.) and defect (size, type, etc.) data in the database can predict the risk patterns included in a new design before manufacturing. The defect information detected by SEM in this way is used in DFM. In addition, it can be used to prioritise inspections of devices with new layouts.

4. Future challenges

As described in Section 3.1, the technology node has been shrinking to 7nm and below through the extreme ultraviolet (EUV) lithography. Thus, the killer defect size become even smaller on both wafer and mask. This size is close to the diameter of the hydrogen atom approximately 0.1 nm. For this reason, expectations for atomic resolution electron microscopes are increasing.

Under these circumstances, the requirements for SEM-based inspection instruments are sufficient spatial resolution (sub nanometer) in image, sufficient SNR in detected signal, tight control over the landing position of the PE beam, and non-invasive to the specimen as possible.

Regarding the spatial resolution, the SEM with aberration correction system have attained 2.04 nm in the image

sharpness through aberration correction at 100 eV E_{PE}, while the calculated spot size is about 0.8 nm [26]. Though the apparent beam size (spatial resolution) would not be identical to the final probe size due to electron beam interaction effects, the difference is considered resulting mainly from electromagnetic noise, shot noise and mechanical vibration of the experimental platform. Thus, steady research is needed to eliminate or mitigate these factors.

Regarding noise, according to Sakakibara et al (2019) [38] in Section 2.2.3, noise in SEM-image can be minimized by directly detecting SEs (SE (1) and SE (2)) emitted from the specimen. This means that efforts should be made to detect only the signals that are really needed. Thus, in order to arrange the detector properly, and to achieve the optimized measurement conditions and parameters for sufficient SNR, the use of simulation and modeling is indispensable.

It is not easy to irradiate an electron beam at the desired inspection point, because there are the specimen stage drift (mechanical vibrations, acoustical noise) and electromagnetic disturbances. Specimen stage drift may be mitigated by matching between images acquired by using fast scans in Section 2.1.3. For the landing error of the PE beam due to the local charging of the specimen, Ura' paper (1981) [78] could be used as a reference.

When PEs with high energy are used for SEM-based inspection, high-energy electrons interacting with the specimen can damage sensitive devices [79]. Therefore, low energy PE inspection (less than 2.5 keV) is believed to eliminate or at least minimise damage to such devices. Devices in several nm technology nodes are expected to be more sensitive to the energy provided by the PE, so care should be taken even when using low energy PE. Non-invasive (non-destructive) inspection should be performed using low PE energy and low dose amount where electron beam damage is not a problem so that the chemical bonds of the specimen are not broken. Therefore, it is necessary to understand the electrical properties of the specimen device structure at each stage of manufacturing before inspection. Physics based simulations can be useful for this understanding.

Furthermore, there are the following challenges.

Electron beam induced contamination can be a serious problem, since it changes the specimen itself, and the number, orbits, and energies of electrons leaving the specimen. Due to contamination, spatial resolution can deteriorate and reproducible quantitative measurements are difficult. Low-energy plasma cleaning and chemical cleaning methods are available to eliminate such contamination, but care should be taken when inspecting nm-technology node devices. Physics based simulations can also be useful for this affect.

In the SEM-based image inspection, the whole specimen area is scanned to acquire the image, though the areas actually needed for measurements are small. Thus, an intelligent scanning scheme is needed. For example, a scanning method that utilises layout information, such as the D2DB in Section 3.2, will be useful. It is better to use the problematic layout area information such as the design layout risk assessment in Section 3.5.

As the feature size of the device approaches a few atoms, applying SEM-based inspection methods requires attention to both electron microscope physics and inspection method physics in details. Simulations used for model bases should consider physical phenomena that were previously negligible. However, the simulation time will increase. Then, it is necessary to reduce the simulation time. One approach is to use a computer that allows parallel computing. Another approach is to adopt analytical calculation schemes in simulations such as Lee's paper (2018) [23]. In addition, as the size of the feature approaches the size of a few atoms, a simulation based on the wave nature of the incident PE becomes necessary, which is a probability-based approach.

As described in the previous sections, inspection researches using deep learning are increasing. This trend continues to increase. In the near future, there may be inspection systems with artificial intelligence that have learned human expert's knowledge and experiences in this field.

New inspection systems continue to be devised, making full use of knowledges of physics and statistics.

5. Conclusions

This paper presents a review of SEM-based electron microscopy in semiconductor inspection. The physics of image formation, electron-specimen interaction and useful contrast in the detected signal in SEM are described along with inspection applications.

Although not mentioned in the text, it is necessary to evaluate whether the SEM-based electron microscope instrument can provide data that can be used for process control. For this purpose, the ratio of measurement accuracy to process fluctuation tolerance is used. The measurement accuracy is obtained from the short-term repeatability and long-term reproducibility of the inspection device. Therefore, the long-term stability of the device becomes an important issue. Steady research is required.

Furthermore, the number of inspection items is increasing due to the miniaturisation, three-dimensionalization, and structural complexity of semiconductor devices. For this reason, hybrid methods of electron microscopy and other inspection methods are being studied, and the amount of measurement data tends to increase. Signal processing challenges will also increase.