Dual-Lens Electron Holography for Junction Profiling and Strain Mapping of Semiconductor Devices

Y.Y. Wang,[1]* A. Domenicucci,[1] and J. Bruley[2]

[1]IBM Microelectronic Division, Zip 40E, Hudson Valley Research Park, 2070 Route 52, Hopewell Junction, New York 12533

[2]IBM T. J. Watson Research Center, 1101 Kitchawan Road, Route 134, Yorktown Heights, New York 10598

*wangyy@us.ibm.com

Introduction

Electron microscopes have been used extensively to look at structure at the nanometer scale. Most of the information obtained from electron microscopes is amplitude information. Yet, the phase information of electron microscopy, which can be obtained from off-axis electron holography, provides unique information on electronic structure and structural changes in a wide variety of materials.

For the semiconductor industry, junction profiling and strain mapping in Si at high spatial resolution provide information that is critical for further scaling of semiconductor devices. Because of the complex process conditions involved to control the junction position relative to the gate position, determination of junction position at high spatial resolution can help to reduce the cost and cycle time of development. Bright-field holography can measure the phase change of electrons traversing the materials, which is directly related to the mean inner potential of Si, indicating the junction position at the nanometer scale. Furthermore, in recent years stressors have been incorporated into devices to change the semiconductor lattice constant in the channel region and thereby enhance hole and electron mobility. Like the junction definition, the extra processing steps involved to add strain in a device have increased development and manufacturing costs. One way to minimize development cycle time is to monitor, at a nanometer scale, changes in channel deformation resulting from process changes. In 2008, Hytch et al. reported that dark-field holography can provide a promising path to nanometer scale strain mapping [1]. Cooper et al. reported using dark-field holography to measure strain related to different process conditions [2, 3].

The requirements of electron holography to inspect the current generation of semiconductor devices are: (1) a fringe width (fringe overlap) in the range of about 100 to 800 nm for an adequate field of view (FOV), (2) fringe spacing between 0.5 and 10 nm for meaningful spatial resolution, (3) visibility of the fringe contrast (10–30%) for useful signal-to-noise ratio, and (4) adjustability of both the FOV and the fringe spacing relative to the sample. In previous papers and patent disclosures, we reported that we had developed a dual-lens electron holography method on a JEOL instrument to meet the above requirements, and later we implemented the same method on FEI instruments to provide a similar operational range for electron holography [4–6]. This dual-lens operation allows electron holography to be performed from low to high magnification and provides the FOV and fringe spacing necessary for two-dimensional (2D) junction profiling and strain measurements for devices with various sizes.

In this article, we describe the electron optics for this method. We also describe several examples of junction profiling and strain mapping to show how to use dual-lens electron holography to resolve semiconductor device issues at high spatial resolution.

Theory of Electron Holography

The electron passing through a specimen contains amplitude information as well as phase information. Regular transmission electron microscopy can only measure the amplitude information. However, electron holography can be used to measure both amplitude and phase information of the electron wave function passing through the sample. The electron holography is achieved by interfering two electron beams through a biprism: one electron beam passing through the region of interest and other electron beam passing through reference region, as shown in Figure 1 and Figure 2 for bright-field electron holography and dark-field electron holography, respectively.

In electron holography, three parameters are critical: fringe overlap for FOV, fringe spacing for spatial resolution, and fringe contrast for signal-to noise ratio. The fringe spacing, σ_i, and fringe overlap, W_i, at the image plane is determined by the biprism voltage and have a limited range because of the requirement of fringe contrast value, μ, for data processing [4–12]. Their relationship to the object (or specimen), σ_o and W_o, can be written as:

$$\sigma_o = \frac{\sigma_i}{M_o} \quad (1)$$

and

$$W_o = \frac{W_i}{M_o}, \quad (2)$$

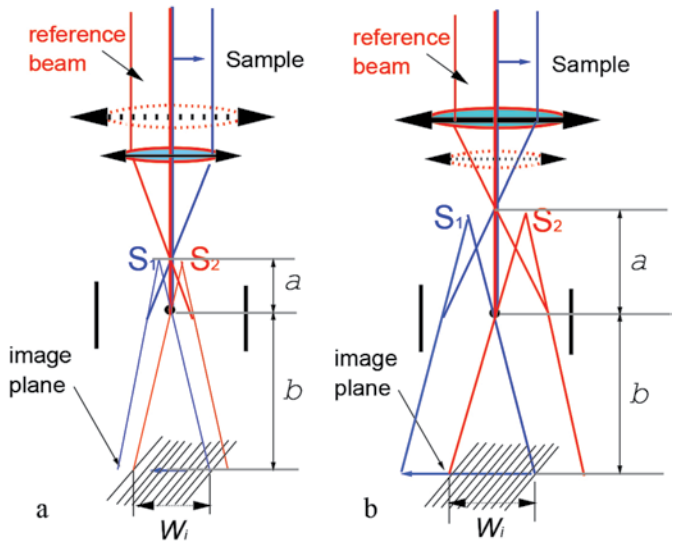

Figure 1: Electron hologram formation using a single lens in a dual-lens imaging system. The interference pattern with fringe spacing, σ_i, and fringe width, W_i, is formed by two virtual sources, S_1 and S_2. (a) Low-magnification single-lens holography where only the second lens is turned on. (b) High-magnification single-lens holography, where only the first lens is turned on.

where M_o is the magnification of the imaging lens(es). The equation shows that magnificaton is a dominant factor for the variation of fringe spacing and fringe overlap relative to the object.

Bright-Field Electron Holography for Junction Profiling

The intensity of the wave function for bright-field electron holography after the biprism can be written as:

$$I = A_0^2 + A^2(\vec{r}) + 2\mu A(\vec{r}) A_0 \cos[2\pi q_c r + \varphi(\vec{r}) + \theta] \quad (3)$$

where $q_c = 1/\sigma_o$, which denotes the frequency of the interference fringes, μ is the fringe contrast, and θ is the constant phase shift. The phase map, which is related to the mean inner potential, and the amplitude map, which is related to the specimen thickness and material scattering factor, can be extracted through a fast Fourier transformation (FFT) of the hologram and an inverse FFT of an appropriately sized side band [13, 14].

Dark-Field Electron Holography for Strain Mapping

The dark-field holography method was first developed by Hytch et al [1]. A detailed method of strain mapping by electron holography was described by Beche et al. [15]. A theoretical description of dark-field holography in relationship with strain measurement can be found in Hytch et al.'s 1998 paper [16] and Rouviere and Sarigiannnidou's 2005 paper [17]. The principle of dark-field holography is to use a biprism to overlap a strained region of the device with an unstrained region in Si, as shown in Figure 2a (off-axis dark-field holography) or in Figure 2b (on-axis dark-field holography). The intensity of a dark-field hologram can be written as following:

$$I = A_{obj}^2 + A_{ref}^2 + 2\mu A_{obj} A_{ref} \cos[2\pi q_c r + 2\pi(\vec{g}_{ref} - \vec{g}_{obj}) \bullet \vec{r} + \varphi_{ref} - \varphi_{obj}] \quad (4)$$

where $\vec{g}_{obj}$ and $\vec{g}_{ref}$ are g vectors from the region of the interest (object) and the reference, respectively. The phase of the hologram is:

$$\varphi = 2\pi(\vec{g}_{ref} - \vec{g}_{obj}) \bullet \vec{r} + \varphi_{ref} - \varphi_{obj} \quad (5)$$

The lattice constant change to the first order can be approximated as:

$$\varepsilon_{ii} = \frac{d_{obj}^{(i)} - d_{ref}^{(i)}}{d_{ref}^{(i)}} \approx \frac{g_{ref}^{(i)} - g_{obj}^{(i)}}{g_{ref}^{(i)}} \approx \frac{1}{2\pi g_{ref}^{(i)}} \frac{\partial \varphi^{(i)}}{\partial r_i} \quad (6)$$

where $i = x, y$, $g = 1/d$, and d is the lattice spacing. By selecting a specific diffracted beam with an objective aperture to obtain a dark-field hologram, the difference in lattice spacing between strained and unstrained Si can be measured.

Dual-Lens Electron Holography

Prior to the dual-lens operation, electron holography had limited application because of its fixed FOV and fringe spacing relative to the sample, with a microscope operated in a single-objective lens mode. This is because in a single-lens mode, the magnification is usually fixed and Equations (1) and (2) show that the magnification is the dominant factor to determine the fringe spacing and fringe width relative to the sample.

Figure 1a shows the low-magnification operation mode with second objective lens excited, whereas Figure 1b shows high-magnification operation mode with the first objective lens excited. The fringe spacing of low-magnification mode (Figure 1a) is too large for the current generation of devices, whereas the FOV of high-magnification mode (Figure 1b) is too small for semiconductor devices. These two single-lens operations cannot cover the range of the current generation devices. Because of this problem, we developed the dual-lens electron holography method to obtain wide fringe spacing and FOV to cover the device dimensions of several generations [4–6]. However, the application of dual-lens method is not limited to only semiconductor device characterization.

Figure 3 illustrates the electron optical ray diagram of a dual-lens system. The voltage (or current) of the first objective lens (OL) is set so that the position of the object is behind the first focal point f_{1OL} but before the lens, thereby forming a virtual image of the object. The second objective lens (OM) is used to project a real image of the OL's virtual image onto the intermediate image plane beyond the biprism. As shown in Figure 3, when the focal point of the OL gets closer to the sample, the virtual image moves further away from the sample and is magnified. In order to refocus the virtual image onto the same image plane, the focal length of the OM is increased (by decreasing the OM strength) to compensate for the movement of the virtual image position. Thereby, the magnification of the sample at the image plane, M_o, can be adjusted by varying the focal length of the first objective lens (OL) to give a virtual image of variable size, which is refocused to the intermediate image plane by the second objective lens (OM).

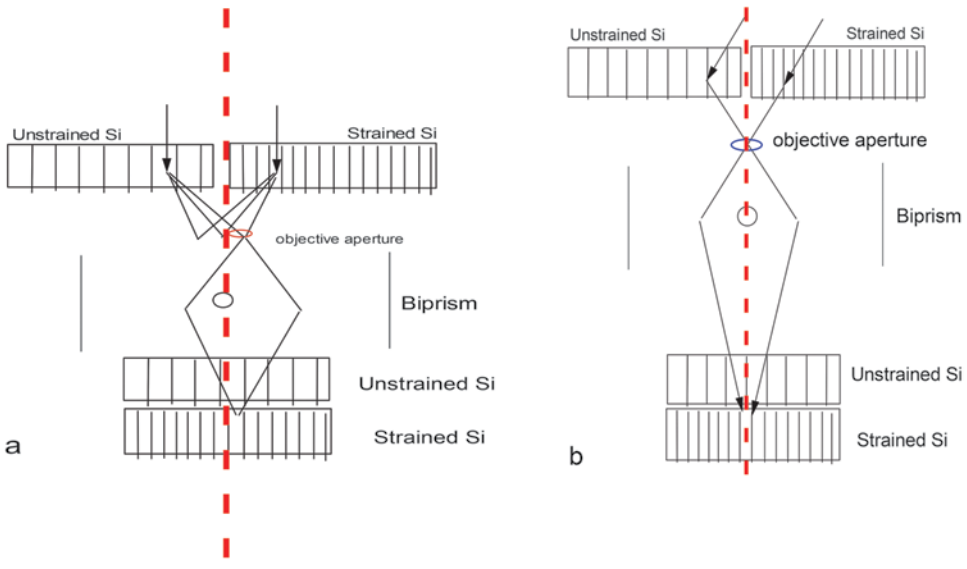

Figure 2: (a) Off-axis dark-field electron holography. A diffracted beam is selected by a moving objective aperture to the diffracted beam location. (b) On-axis dark-field electron holography. A diffracted beam is selected by tilting the illumination to bring the diffracted beam to the optical axis.

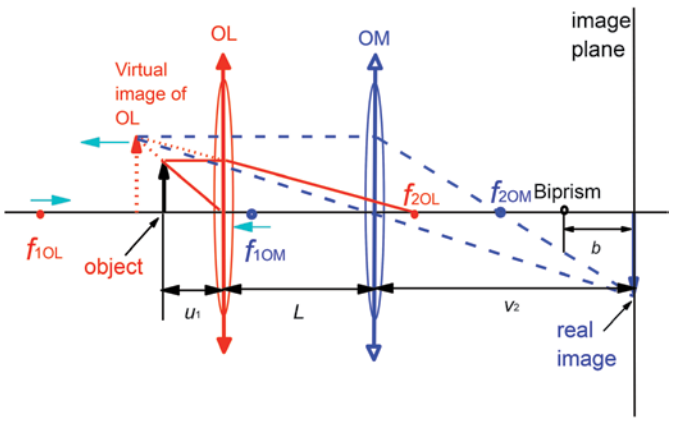

Figure 3: Optical ray diagram of dual-lens operation. A magnified virtual image of OL (first objective lens) is shown as a dashed image and a real image is projected by OM (second objective lens) onto the imaging plane. As the focal point gets closer to the object by increasing the OL strength, a virtual image is magnified and moves away from OL. To refocus the virtual image back to the imaging plane, the strength of the second lens is reduced to increase the focal length of the second lens, as shown in the movement of f_{1OM} by the nearby arrow.

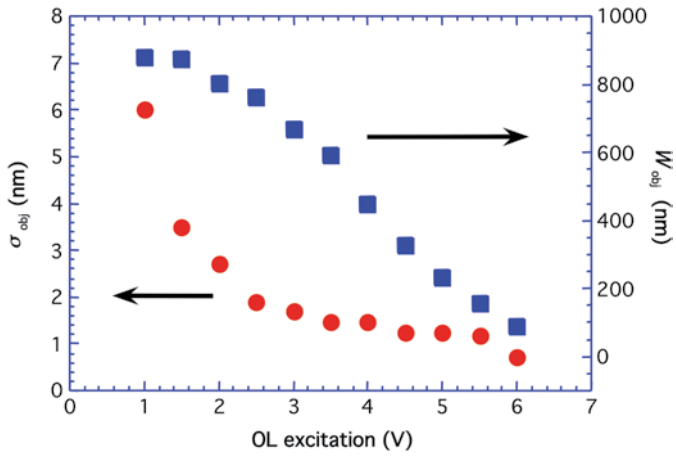

Figure 4: Fringe spacing σ_{obj} (circle) and fringe width W_{obj} (square) as a function of objective lens excitation in the unit of voltage applied to the lens with a constant biprism voltage of 20 V for microscope A.

Figure 4 summarizes experimental results for fringe width and fringe spacing relative to the objective lens excitation on microscope A. The graphical results show that the range of values for FOV and fringe spacing necessary for semiconductor characterization are achieved. The fringe spacing decreases from 6 nm to 0.5 nm; while the FOV changes from 800 nm to 100 nm, when the first objective lens current increases with a constant biprism voltage. The second objective lens is adjusted to refocus the image back to the same imaging plane as shown in Figure 3. With the constant biprism voltage, the contrast varies slightly through the operational ranges [4].

We obtained similar curves on other instruments with dual-lens configurations. Figure 5 shows the fringe spacing and fringe overlap decrease with increasing objective lens excitation for instrument B (Figure 5a) and instrument C (Figure 5b) [6]. The rate of fringe spacing decreases with the objective current on instrument B and C being slightly different from the one obtained on instrument A, but the trends are similar.

In all, the data show that the dual-lens mode leads to a large operational space for electron holography without sacrificing fringe contrast, which determines the signal-to-noise ratio in the amplitude and phase (potential) maps derived from the electron holograms.

Junction Profiling by Dual-Lens Electron Holography

Example 1: Holograms from devices of different sizes. To demonstrate the flexibility of the dual-lens method, we obtained electron holograms of pFET devices with poly silicon gate lengths of 220 nm and 70 nm on bulk Si and on silicon on insulator (SOI) substrates, respectively. Two holograms were obtained using a biprism voltage of 20 V: one is the hologram with the sample and the other is the hologram without the sample (known as a reference hologram). Phase and amplitude maps were obtained from these two holograms through FFTs and inverse FFTs. A potential map can be extracted from the phase map.

Contour potential maps for the two devices are shown in Figure 6. Figure 6a is for the device with 220 nm gate length and Figure 6b is a device with 70 nm gate length. The intensity of the maps near the sides of the gates (side wall spacers and $CoSi_2$) changes sign abruptly due to phase wrapping. As shown in this example, both σ_o and W_o can be scaled by simply varying the OL excitation without losing fringe contrast. This allows the characterization of different size devices with similar signal sensitivity. In these two cases, device scale changes about 3×, and the electron holography imaging scale also changes 3× to show the flexibility of dual-lens operation.

In high-performance CMOS logic devices, low-dose ion implants are placed between the channel and the low-resistance (high-dose) source/drain regions. In Figure 6b, the upper curved portion of the P-N junction on either side of the device channel represents the result of such a low-dose implant, which is clearly visible because of the high spatial resolution in the hologram from which the potential map was reconstructed.

Example 2: P-N junction mapping. Another example is the junction mapping on a narrow Si diffusion region (shown in Figure 7a). At certain process conditions, a high-leakage current is observed from the contact region to the substrate in this narrow Si diffusion region. In order to understand the leakage path, a junction map was acquired by electron holography, (Figure 7b). Similar P-N junctions are observed in the potential maps for two process conditions (Figure 7b and Figure 7c). Near the edge of the Si diffusion region (Figure 7b), the junction is slightly curved upward toward $CoSi_2$. The slight displacement of the junction near the edge of the insulator could be due to either implant shadowing or boron loss at the Si to SiO_2 interface. Without high spatial resolution, it would be difficult to detect this small shift of the junction near the edge of Si diffusion region. Figure 7c shows another process condition where shadowing of the implant occurs along the side of the Si diffusion region. In this picture, the junction at the edge of the Si diffusion region is curved more upward toward $CoSi_2$. Its junction position is closer to

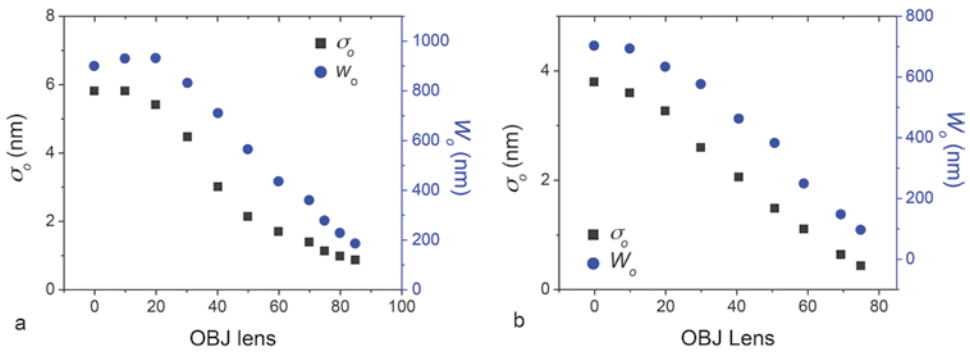

Figure 5: (a) Fringe spacing and fringe width relative to the object as a function of objective lens excitation current in the unit of percentage from the full excitation of the lens on instrument B with the biprism voltage of 100 V. (b) Fringe spacing and fringe width relative to the object versus objective lens excitation current in the unit of percentage from the full excitation of the lens on instrument C with the biprism voltage of 120 V.

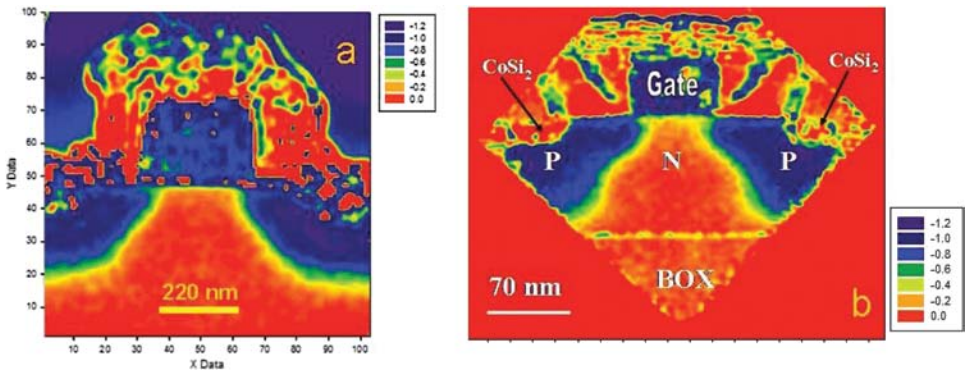

Figure 6: Mean inner potential maps of pFET devices. (a) With 220 nm gate width and (b) with 70 nm gate width.

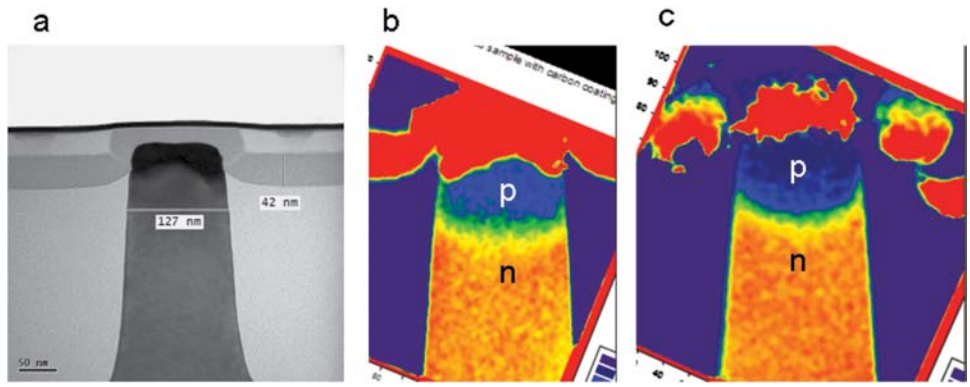

Figure 7: P-diffusion to N-well profile for the Si diffusion region. At the edge of the Si near SiO_2 wall, the junction is slightly tilted toward $CoSi_2$. (a) TEM image, (b) process condition 1, (c) process condition 2. Where high leakage to the substrate was observed in process condition 2, the junction is more curved upward near the edge of the diffusion Si region.

the $CoSi_2$ than the one shown in Figure 7b. In this case, we found that leakage from $CoSi_2$ toward the substrate is three orders of magnitude higher for the case of Figure 7c than the case of Figure 7b.

Example 3: Shallow extension junction profile. Recently, shallow junction devices have become a tool to control the source to drain off current leakage when the gate becomes narrow. However, that kind of device, without a raised source/drain, cannot be properly manufactured as shown in the following example. Figure 8a is a TEM image of a shallow junction device. Normally, $CoSi_2$ is processed in the source/drain region, where high implant dose is used to reduce the contact resistance. The shallow extension junction of low dose is implanted before the source/drain region. Figure 8b shows that junction position is between blue and yellow. Red indicates the presence of $CoSi_2$. From this image, $CoSi_2$ is very close to the junction position near the intersection of the extension and source/drain regions. Figure 8c is a drawing of the junction position relative to the $CoSi_2$ for this kind of device. This kind of device is not a stable device because a slight movement of $CoSi_2$ on either side of the device could cause device asymmetry leakage. This would act as a diode device instead of a transistor if one side of $CoSi_2$ breaches the junction position. Therefore, to make the shallow junction device work properly, a raised source/drain is needed to avoid the silicide breaching the shallow junctions. Recently, the semiconductor industry has adopted raised source/drain technology to help shrink the device further.

Strain Mapping by Dual-Lens Dark-Field Electron Holography

Strain mapping along the <220> direction in Si provides channel strain information important to the semiconductor industry. For pFET, compressive strain improves device mobility; whereas for nFET, tensile strain improves device performance. In the following, we provide two examples to show how to use dual-lens electron holography to measure strain along the <220> direction at high spatial resolution.

Example 4: Compressive strain measurements. One of the methods to provide compressive strain for pFET is to introduce epitaxial growth of SiGe in the source and drain region because the Ge atom size is larger than the Si. Because of this, the Si-Si atomic bonding in the channel under the gate is compressed when Ge is introduced to the lattice. One of the questions related to strain engineering at the micro scale is how the strain changes as a function of the SiGe profile. Early SiGe profiles had a rectangular shape, with a vertical wall along the edge of the source or drain region. Second generation SiGe profiles have a facet shape, which brings SiGe closer to the channel in hopes that it increases the strain in the channel region. That process is known as sigma-shaped

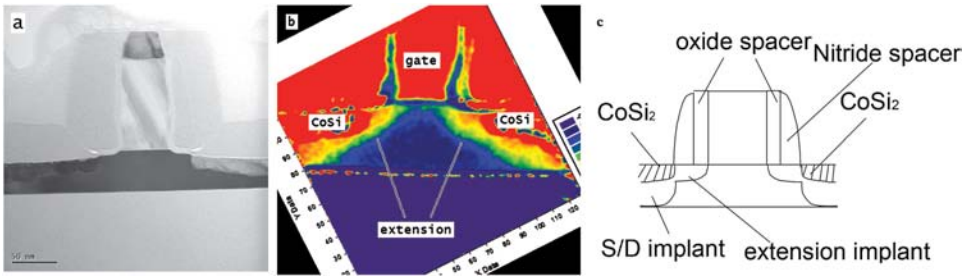

Figure 8: nFET shallow extension junction profile. (a) TEM image of the shallow junction, (b) junction map that shows $CoSi_2$ is too close to the junction at the intersection of the extension implant and the source/drain implant. (c) Drawing of shallow junction of Figure (8b).

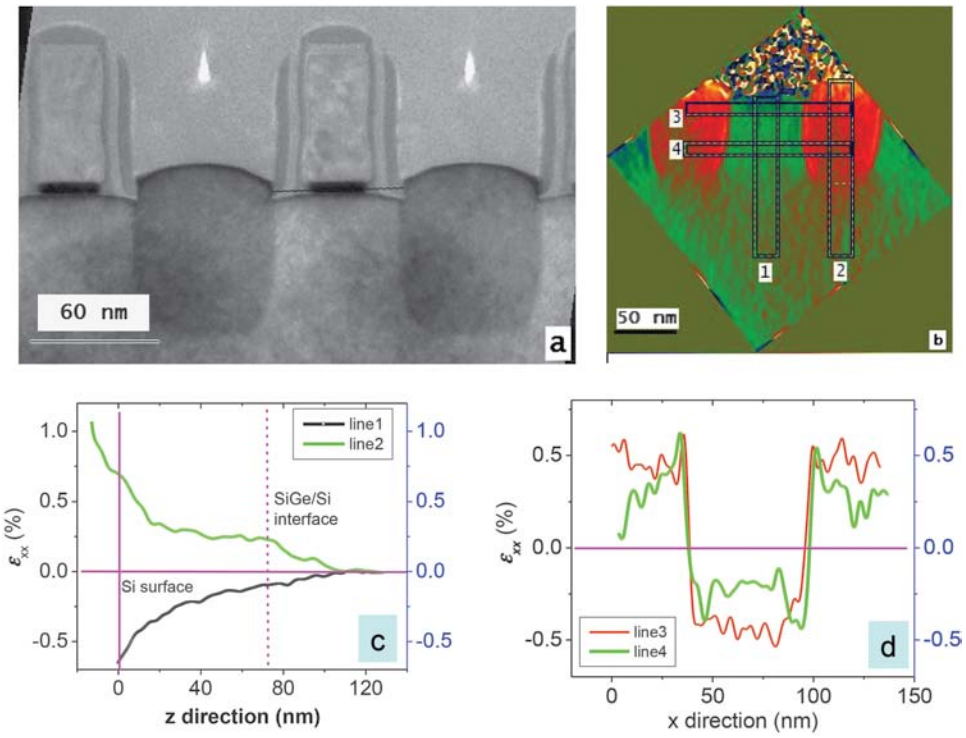

Figure 9: Strain maps of box SiGe after SiGe epitaxial growth. (a) TEM image, (b) <220> strain maps where blue color indicates lattice compression, and red color indicates lattice expansion. (c) Vertical profiles of <220> strain at positions 1 and 2 in the channel and in the source/drain region. (d) Line scan 3 and line scan 4 shown in Figure 9b.

SiGe. The question that needs to be answered is how strain is distributed in these two kinds of SiGe processes.

Figure 9 shows <220> strain measured right after SiGe growth with a box shape. Figure 9a is the TEM image, 9b is the strain map, and 9c is the vertical profile of the strain distribution in the channel and in the source drain region. The compressive strain increases to a maximum at the Si surface in the channel (to -0.65%). Figure 10d shows the line scan for the strain parallel to the channel direction. At the SiGe and Si boundary, the lattice constant along the <220> direction is discontinuous in Figure 9d, whereas at the bottom of the SiGe box, the lattice constant along the 220 direction is continuous between Si and SiGe, but the derivation of the lattice constant is not as shown by the line scan 2 in Figure 9c.

Figure 10a is the TEM image of the sigma profile SiGe device. Figure 10b shows the the strain map along the <220> direction, and Figure 11c shows the vertical profile of the strain map. The compressive strain for this device at the Si surface is about -1.7%. Compared with the box-shaped SiGe shown in Figure 9, the absolute strain is higher (more negative), and the compressive region is more concentrated in the channel region with the sigma SiGe process than with the box SiGe process. This is consistent with the electrical data, which shows that the device with sigma profile SiGe has better performance characteristics than the device with the rectangular-shaped SiGe process. The dual-lens electron holography allows measurement of the strain distribution in a very small region with detailed strain distribution.

Example 5: Strain after stress memorization. Tensile strain under the channel helps to improve device performance for nFET devices. One way to achieve tensile strain for nFET devices is to use stress memorization techniques to generate tensile strain in the channel of Si device [18–22]. This technique involves heavy implantation to amorphize Si in the source/drain region, as well as poly-Si in the gate. The Si of both the source/drain region and the gate is then recrystallized through a high-temperature anneal with stressed SiN deposited on top of the source/drain region and the poly-Si gate. After annealing, the SiN film is removed, and tensile strain in the channel remains. The mechanism of how strain is memorized is under debate. Early papers suggested that the strain was memorized in the poly gate [18, 19]. However, recent papers proposed that the strain is memorized in the source/drain region through dislocations generated by the stress memorization process [20, 21]. The high spatial resolution strain mapping helps to provide deeper understanding of the mechanism of the strain memorization techniques [22].

Figure 11a is the <220> strain map of a device processed through strain memorization techniques. From the map, we noticed the highest strain at the Si surface is about 1.1%. In addition to that, the strain near the dislocation region is about 2%. The strain radiates toward the channel region from both sides of the gate. Figure 11b is the profile of strain near the dislocation core, which shows high compressive strain under the dislocation core to about -15%- to -20% and tensile strain above dislocation core of 15% to 20%. Figure 11c is the vertical strain profile in the middle of the device shown in Figure 12a. Based on this picture, we concluded that the strain is memorized in the dislocation generated during the strain

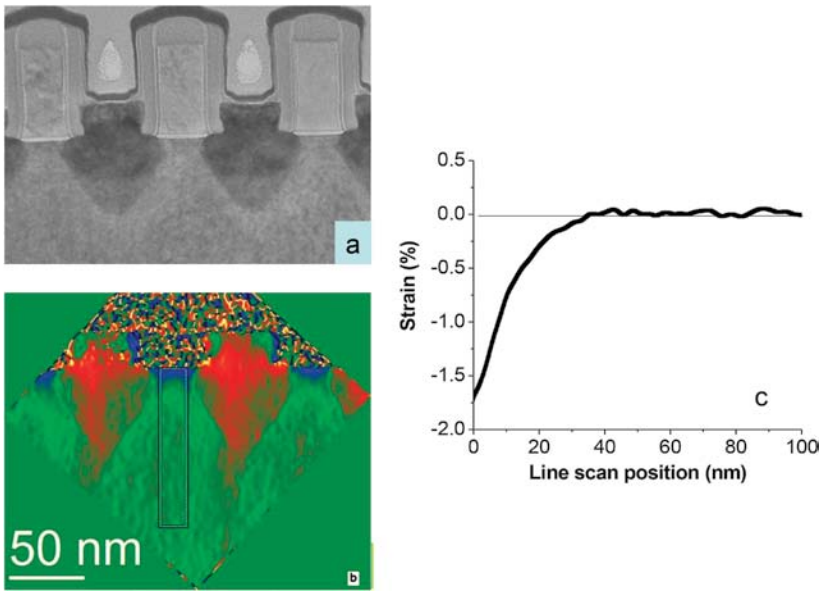

Figure 10: Strain measurement on sigma-shaped SiGe structure. (a) TEM image, (b) <220> strain maps where blue color indicates lattice compression, and red color is lattice expansion. (c) Vertical profile of strain distribution.

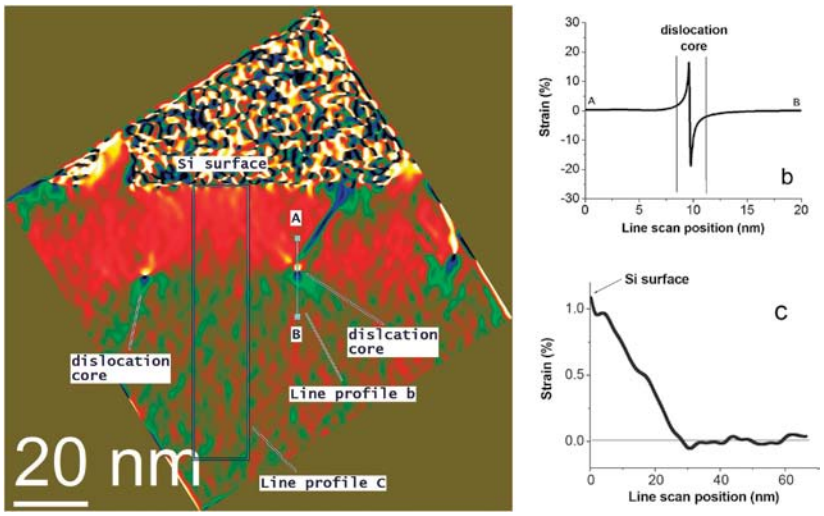

Figure 11: Strain map on a device processed by the stress memorization process. Red/yellow color shows lattice expansion, and blue indicates lattice compression. High strain near a dislocation is observed. Right top profile is the strain near the dislocation core; right bottom profile is the vertical strain distribution of the <220> direction. Vertical scale of the right figures is strain.

memorization process step. In addition to that, we noticed that additional tensile strain of 0.2% is generated after the poly-Si gate is removed, as required by the replacement gate process [22], consistent with the dislocation model for stress memorization technique [20].

Conclusion

Dual-lens electron holography with a wide FOV and good spatial resolution is critical for semiconductor device junction profiling and strain mapping. This method has the advantage of allowing the fringe width and fringe spacing in electron holograms to be adjusted independently from the fringe contrast. This results in better spatial resolution and signal-to-noise for the characterization of semiconductor devices of 2-D potential profiling and strain mapping. The high spatial resolution mapping by dual-lens electron holography reveals significant new insight about semiconductor device characteristics.

Acknowledgments

We would like to acknowledge IBM managers A. Katnani, M. Lagus, A.C. Lamberti, and K. O'Buckley for their support on the dual-lens electron holography development at IBM microelectronics division. We would also like acknowledge F. Baumann of IBM and D. Cooper of LETI in France for discussions on strain measurement.

References

[1] M Hytch et al., *Nature* 453 (2008) 1086–89.
[2] D Cooper et al., *Appl Phys Lett* 95 (2009) 053501–053501-3.
[3] D Cooper et al., *Appl Phys Lett* 96 (2010) 113508–113508-3.
[4] YY Wang et al., *Ultramicroscopy* 101 (2004) 63–72.
[5] YY Wang et al., US patent: US 7,015,469 B2 (2006).
[6] YY Wang et al., *Ultramicroscopy* 124 (2013) 117–29.
[7] GF Missiroli, G Pozzi and U Valdre, *J Phys E Sci Instrum* 14 (1981) 649–71.
[8] H Lichte, *Ultramicroscopy* 64 (1996) 79–86.
[9] WD Rau and H Lichte. in *Introduction to Electron Holography*, eds. E Volk, LF Allard, and DC Joy, Kluwer Academic/Plenum Publishers, New York, 1999, p. 208.
[10] H Lichte, *Advances in Optical and Electron Microscopy*, Vol. 12, Academic Press, New York, 1991, p. 25.
[11] E Voelkl and D Tang, *Ultramicroscopy* 110 (2010) 447–59.
[12] YY Wang et al., *JEOL News* 47 (2012) 9.
[13] MR Mccartney, *Ultramicroscopy* 53 (1994) 283–89.
[14] WD Rau et al., *Phys Rev Lett* 82 (1999) 2614–17.
[15] A Beche et al., *Ultramicroscopy* 111 (2011) 227–38.
[16] MJ Hytch, E Snoeck, and R Kilaas, *Ultramicroscopy* 74 (1998) 131–46.
[17] J Rouviere and E Sariggiannidou, *Ultramicroscopy* 106 (2005) 1–17.
[18] K Ota et al.. International Electron Devices Meeting (IEDM), 2002, 27–30.
[19] C Chen et al.. Symposium on VLSI Technology, 2004, 56–57.
[20] CE Weber et al., International Electron Devices Meeting (IEDM), 2011, 801.
[21] S Flachowsky et al., Ultimate Integration on Silicon (ULIS) conference, University of Glasgow, (2010) 149.
[22] YY Wang et al., *Appl Phys Lett* 103 (2013) 052104–052104-5.

MT

https://doi.org/10.1017/S1551929514000352 Published online by Cambridge University Press