Chapter 4: Transparent Film Profilometry

4.1 Research background and goal

When a sample with a transparent film on the surface is measured, two interference signals due to the film surface and the film back-surface are superposed, and accurate surface profile measurement becomes difficult. For example, Fig. 4-1 shows the results of white-light interferometry on the surface of a semiconductor IC. Since the metal pattern wiring having high reflectance exists under the transparent film, the interference signal of the film back-surface is strong and the surface at that portion is depressed.

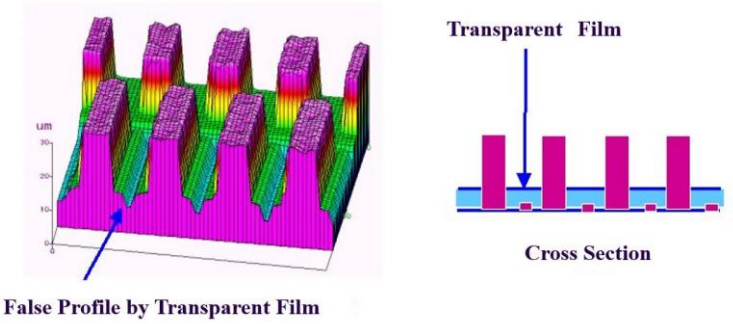


Fig. 4-1 Effect of transparent film

Figure 4-2 shows a schematic diagram of the film thickness and interferogram when a transparent film is present on the substrate as in this example. The same is true for an independent film without a substrate. As shown in this figure, if the film thickness is thicker than the coherence length of white light (about 1 μm), two peaks, surface reflection and back-surface reflection, appear in the interferogram. In this paper, this is called a thick film, and a film with a film thickness less than this value is called a thin film.

In the case of thick film, it has long been known that film thickness and/or film shape can be measured from two peak positions [Flournoy 1972, Tsuruta 1975, Lee 1990, Nakamura 1992, Itoh 1995]. However, what has been reported so far is a method of visually determining the peak position [Flournoy et al. 1972, Tsuruta et al. 1975] or a method of using a complicated calculation algorithm [Nakamura et al. 1992, Itoh et al. 1995], and there is no practical automatic measuring device on the market. The authors investigated a fully automatic peak detection algorithm when two peaks exist.

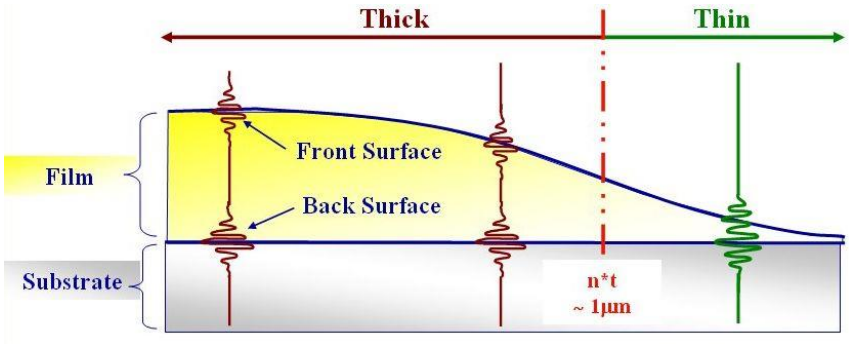


Fig. 4-2 Interferogram of film-covered surface

4.2 Algorithm for thick film（KF method）

We propose an algorithm for automatically detecting multiple peak positions in an interference waveform (named KF method from thicK Film) [Kitagawa 2003, Kitagawa 2004]. Figure 4-3 shows an interferogram for the test, which was obtained by measuring an oxide film with a thickness of about 1 μm formed on a Si wafer with a white light source. The Z-axis scanning speed is 2.4 μm/s (sampling point interval = 80 nm).

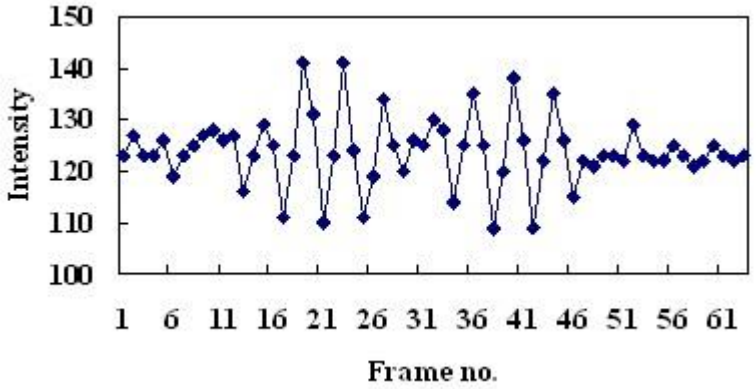


Fig. 4-3　Interferogram used for the KF algorithm test. An oxide film with a thickness of approximately 1 μm on a Si wafer was measured. The sampling interval was 80 nm.

4.2.1 Procedure of KF method

This algorithm uses the following procedure.

(1) Amplitude calculation

Obtaining the AC component of the intensity data and squaring it gives the amplitude graph in Fig. 4-4.

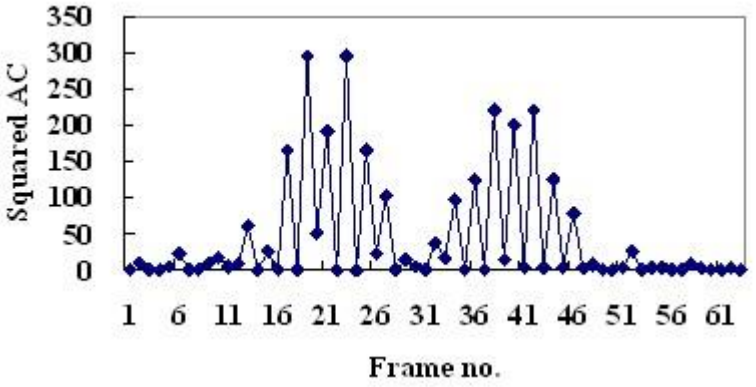


Fig. 4-4　Squared AC component

(2) Peak separation

Taking the above amplitude data as a histogram in which the horizontal axis is the observed value and the vertical axis is the frequency value, determine the optimal threshold for peak separation, using the method of discriminant analysis (Otsu's formula) [Otsu 1980] in the field of image processing.

More particularly, when the number of observation value levels is L and the frequency value

$n(i)$ ($i = 1, L$) of the observation value level i are given, we take the intraclass variance, expressed by the following equation as an objective function, and choose the threshold that minimizes the function:

$$f = \omega_1\sigma_1^2 + \omega_2\sigma_2^2, \qquad (4\text{–}1)$$

where ω_1, ω_2 is the frequency of each class, and σ_1^2, σ_2^2 is the variance of each class.

The values of the objective function are shown by the solid-line curve in Fig. 4-5, with the data number 30 being the threshold.

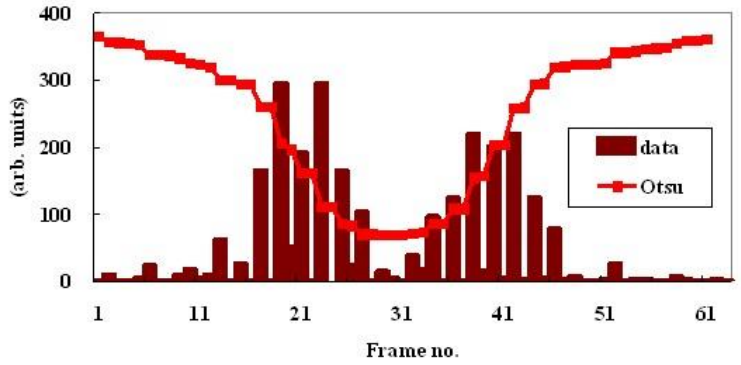


Fig. 4-5 Amplitude data and Otsu's objective function for automatic threshold selection

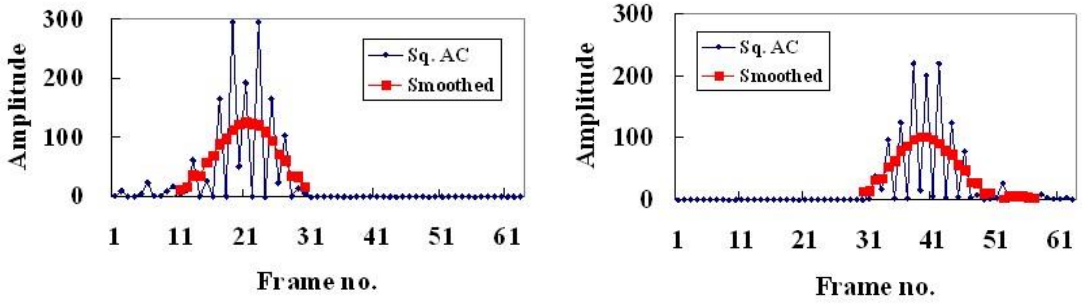
(3) Peak position detection

The threshold value (data number 30) obtained above is used as a boundary to divide into left and right peak regions, and the peak position detection similar to the conventional method is performed for each of the separated peak regions. Here, the waveform is smoothed by using the low-pass filter method, which is the most general peak position detection method. The calculation formula used is a 19-point quadratic polynomial smoothing method and is represented by the following formula [Savitzky et al. 1964; South 1997].

$$y(n) = (1/W)\sum_{i=-m}^{m} w_i x(n+i), \qquad (4\text{-}2)$$

where $m = 9$; W is a normalization coefficient and equal to the sum of w_i; w_i is the weighting coefficient, and in the case of $m = 9$, the weighting coefficients are 269, 264, 249, 224, 189, 144, 89, 24, −51, −136 in order from w_0.

As a result, a smooth curve was obtained as shown in Fig. 4-6, and two peak positions were obtained as data numbers 21 and 40. Furthermore, a quadratic curve was applied to three data points in the vicinity of peaks in order to determine interpolated peak positions, with the resolution being smaller than the sampling interval, obtaining values of 21.17 and 39.61.



(a) Left peak (b) Right peak

Fig. 4-6 Smoothed signals of separated peaks

(4) Front and back-surface heights and thickness

From the heights z_{p1} and z_{p2} converted from the peak position and the refractive index n of the film, the following three values can be obtained.

(a) Transparent film surface height = z_{p1}

(b) Film thickness $t = (z_{p1} - z_{p2})/n$

(c) Back-surface height of transparent film = $z_{p1} - t$

In this test, from the peak positions of 21.17 and 39.61, the sampling point interval of 0.08 μm, and the refractive index of the film n = 1.5, the three values become:

Surface height = 5−21.17×0.08 = 3.31 μm,

Film thickness = (39.61-21.17) × 0.08/1.5 = 0.98 μm,

Back-surface height = 2.33 μm

4.2.2 Comparison with other peak separation algorithms

Besides the above-mentioned Otsu's method, we have a method of Kittler et al. [1986] as another algorithm for automatic threshold selection based on the gray-level histogram [Takagi et al. 1991]. The Kittler method takes the following equation as the objective function in stead of Eq. (4-1):

$$f = \omega_1 \log(\sigma_1 / \omega_1) + \omega_2 \log(\sigma_2 / \omega_2), \tag{4-3}$$

where ω_1 and ω_2 are the frequencies of each class, and σ_1^2 and σ_2^2 are the variances of each class.

Comparison was made between these two methods for various cases. For the data of this experiment, both methods make no difference as shown in Fig. 4-7. Figures 4-8 to 4-10 show other cases. The Kittler method is more effective when the variances of two peaks are much different from each other as shown in Fig. 4-10. For interference waveform, however, little difference is found normally in the variances. Therefore, we adopted a discriminant analysis method that can obtain good results even in the cases shown in Figs. 4-8 and 4-9.

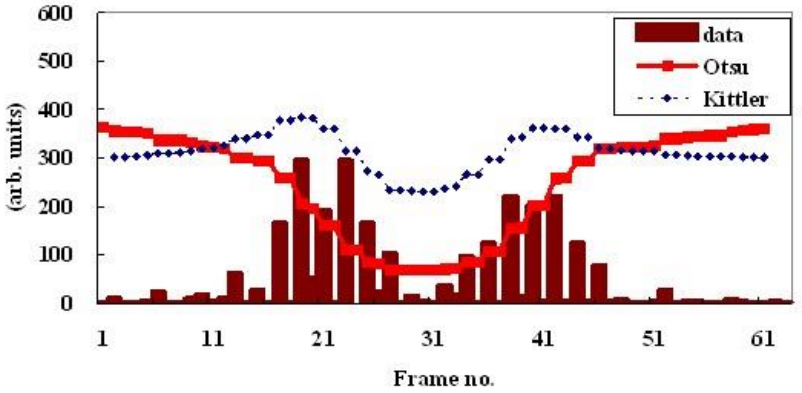


Fig. 4-7 Comparison of two algorithms (case 1): in the case of balanced peak heights

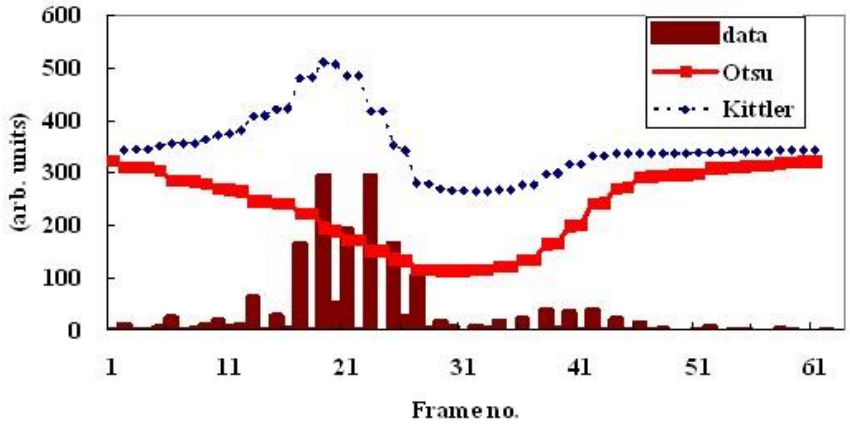


Fig. 4-8 Comparison of two algorithms (case 2): in the case of unbalanced peak heights

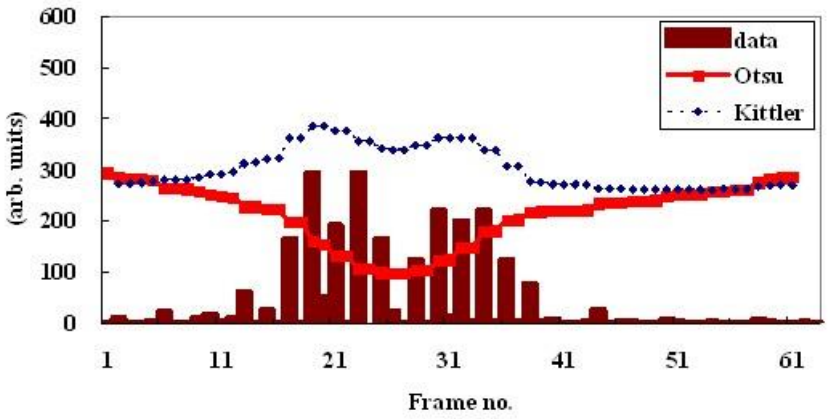


Fig. 4-9 Comparison of two algorithms (case 3): in the case of thin thickness

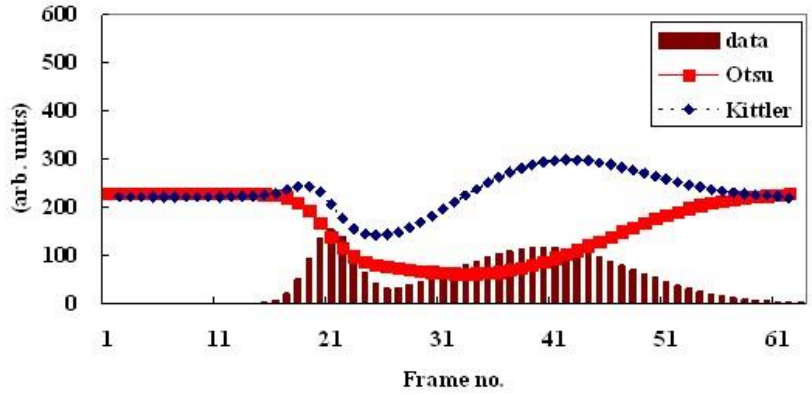


Fig. 4-10 Comparison of two algorithms (case 4): in the case of unbalanced peak width

4.3 Other calculation algorithms -Model function fitting method-

In the KF method in the previous section, if the two peaks are unevenly distributed on one side of the scanning range, the correct separation threshold cannot be obtained. Therefore, it is necessary to set the search area corresponding to the peak position. In addition, when the film thickness is reduced to about 1 μm, the skirts of the two peaks overlap, and the peak separation method has an error in the results. As an algorithm without such a problem, we examined the model function fitting method described below. This method finds the peak position by fitting the model function to the peak of the interference amplitude data by least squares.

4.3.1 Model Function

The squared data of the AC component is regarded as envelope data, and the Gaussian function is adopted as its model function. The Gaussian function has three parameters: surface height (peak position) x_0, peak width w, and peak amplitude a, and is represented by:

$$q(x) = a\exp\left[-\pi\{(x-x_0)/w\}^2\right]. \tag{4-4}$$

4.3.2 Two peaks fitting

If there are two peaks in the interference waveform, the model function is the sum of two Gaussian functions,

$$q(x) = a_1\exp\left[-\pi\{(x-x_{01})/w_1\}^2\right] + a_2\exp\left[-\pi\{(x-x_{02})/w_2\}^2\right], \tag{4-5}$$

where, the first term on the right side is for the first peak, and the second term is for the second peak.

The sum of squared errors is expressed by the following equation:

$$f = \sum_{i=1}^{n}\left[q_i - q(x_i)\right]^2, \tag{4-6}$$

where q_i is the i-th observed data, and x_i is the i-th position.

To find the six unknown parameters that minimizes this equation is a nonlinear programming

problem [1], and there are many calculation algorithms and commercial software packages. In this experiment, the optimization software Solver attached to Excel was used.

It is necessary to properly select the initial values of unknown parameters. The initial values of the peak position and peak amplitude can be obtained by the method shown in Section 4.3.4. Since the peak width is a device parameter that is almost determined by the optical system such as the illumination light source as the coherence length, an approximate value is experimentally obtained in advance and used as the initial value.

4.3.3 Fitting results

The results for this test data are shown in Fig. 4-11 and Table 4-1. As a result, peak positions of 21.14 and 39.84 were obtained. This value is very close to the result of the KF algorithm.

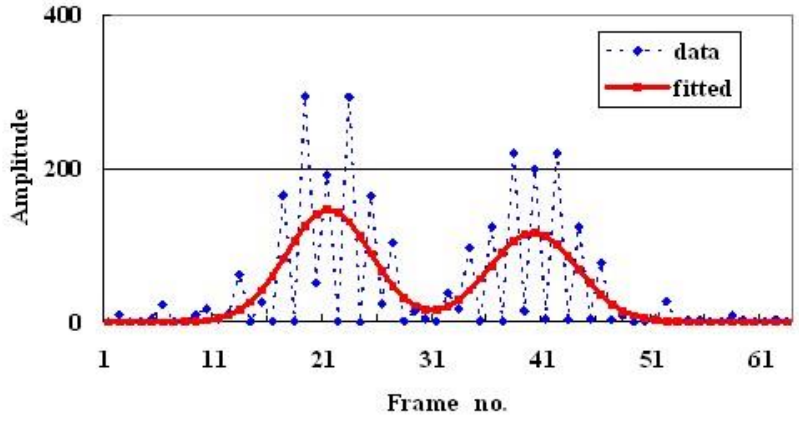


Fig. 4-11 Result of model function fitting

Table 4-1 Result of model function fitting

	Initial values		Estimated values	
	Peak 1	Peak 2	Peak 1	Peak 2
Position *x*	21	39	21.14	39.84
Amplitude *a*	167	117	147.35	116.98
Width *w*	13	13	9.69	10.09

4.3.4 Initial value calculation

In order to properly set the initial values of peak position and peak amplitude, we examined a method for obtaining a sufficiently smooth envelope. If this is obtained, the peak position and peak amplitude can be easily obtained by using the differential method. In this study, the smoothing method consisting of the following three steps was used.

(1) Data-shift and add method

As in this experiment, when the sampling point interval is approximately 1/4 of the fundamental period of the interference waveform, shifting one sample value horizontally

[1] There is a method of linearizing the least squares fitting problem of the Gaussian function by logarithmic transformation, and the nonlinearization problem of the evaluation function by logarithmic transformation can be solved by adopting the weighted least squares method [Jurs 1970]. However, it cannot be applied to the matching problem of the sum of two Gaussian functions.

corresponds to shifting the phase of the waveform by 90°. The sum of squares of the original waveform and the phase-shifted waveform is the squared amplitude of the approximate sine waveform as shown in the following equation, and the approximate envelope can be obtained by a simple calculation:

$$\begin{aligned} q(i)+q(i+1) &= f(x)^2 + f(x+\pi/4)^2 \\ &= [a\sin x]^2 + [a\sin(x+\pi/4)]^2 \\ &= [a\sin x]^2 + [a\cos(x)]^2 \\ &= a^2 \end{aligned} \tag{4–7}$$

This method is effective when the sampling point interval is 1/4, 1/8, 1/16, ..., and can be said to be a simplified version of the peak detection method using the Hilbert transform [Chim et al. 1992].

The calculation results of this test data are shown in Fig. 4-12. A simple calculation yields a very good approximation envelope.

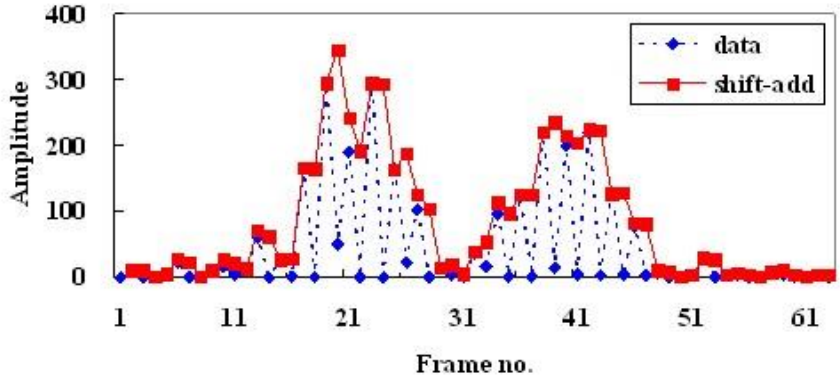


Fig. 4-12　First estimate of envelope obtained by data-shift and add method

(2)　Maximum filter

Remove the small concave points in the calculated approximate envelope by the following three-point neighborhood maximum value calculation:

$$q(i) = \max[q(i), \{(q(i-1)+q(i+1)/2\}]. \tag{4-8}$$

Fig. 4-13 shows the result of extending this idea to the five points and applying it to the approximate envelope of Fig. 4-12. The recesses have disappeared and the envelope is fairly smooth.

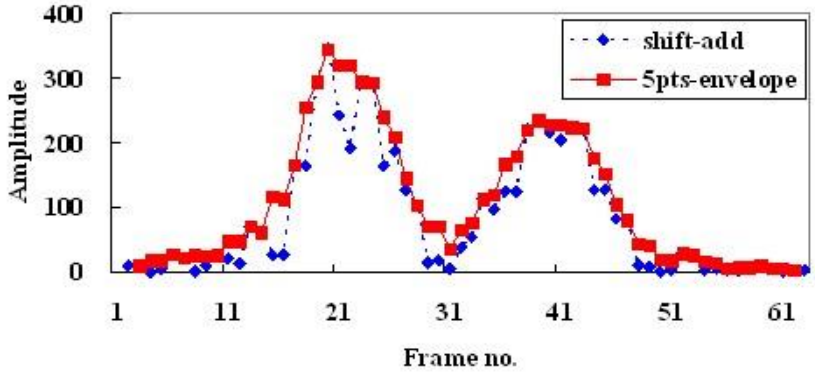


Fig. 4-13　Second estimate of envelope obtained by maximum filter

(3)　Polynomial smoothing

The waveform is further smoothed by the second-order polynomial smoothing method described in Section 4.2.1 (3). The result of convolving the coefficients of the 5-point quadratic approximation is shown in Fig. 4-14. As a result, a sufficiently smooth envelope waveform can be obtained, and the first peak (21, 333) and the second peak (39, 234) are obtained as the position and height of the peak by a simple differentiation method. From this, the initial value of the previous section was obtained by calculating the peak amplitude as 1/2 of the envelope height.

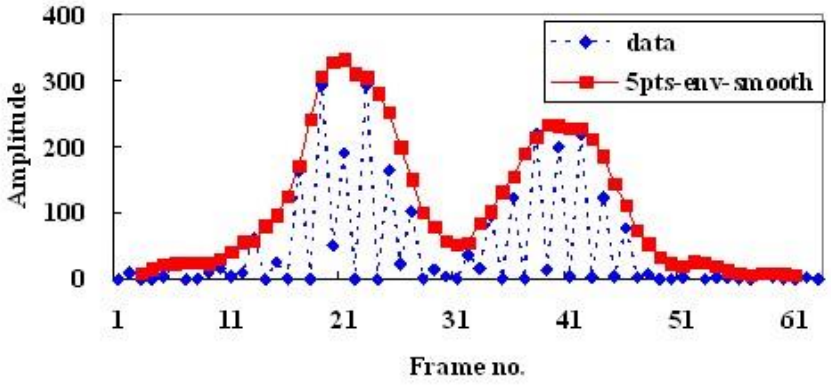


Fig. 4-14　Final estimate of envelope obtained by smoothing filter

4.4 Comparison of algorithms

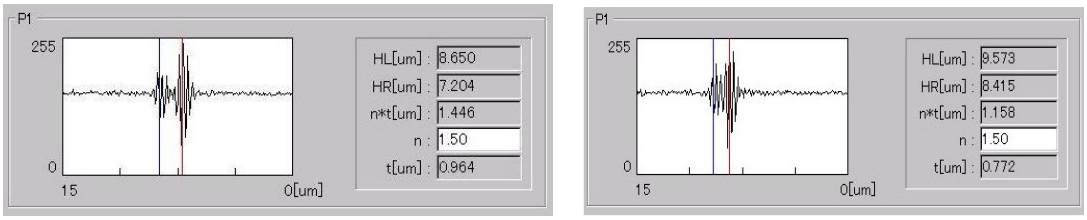
A comparison of the above two algorithms is shown in Table 4-2, and the peak separation method is practical in terms of stability and computational cost. Therefore, the peak separation method (KF method) was installed as the standard software in our profiler. The function fitting method is superior when the film thickness is so thin that the two peaks overlap.

Table 4-2　Comparison of two methods

Method	Robustness	Comp. cost	Multi-layers	Versatility
Peak-separation	○	○	○	×
Function fitting	×	×	△	○

4.5 Measurable film thickness range

Since this measurement method is based on the premise that the peaks of the interferogram can be separated, there is a lower limit of measurable film thickness. The interferogram and measurement results of the oxide film on the Si wafer are shown in Fig. 4-15(a)(b). In Fig. 4-15(b), the optical film thickness is measured as 1.16 μm (physical film thickness 0.77 μm), which is almost the lower limit.



(a) Optical thickness=1.45 μm (b) Optical thickness=1.16 μm

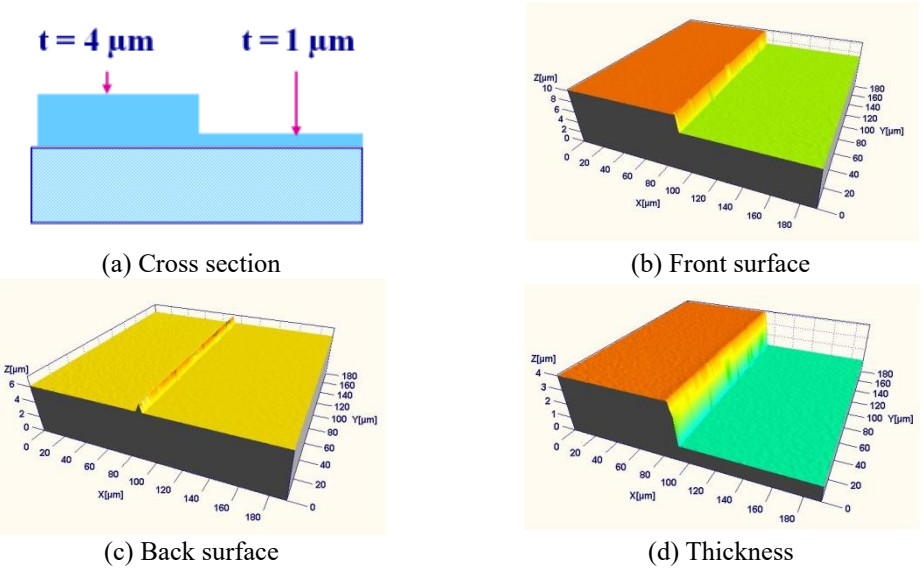
Fig. 4-15 Interferograms of S_iO_2 films

4.6 Experimental results

Various thick film samples were measured by incorporating the above algorithm into surface profiler SP-500. Unless otherwise specified, the vertical scanning speed is 2.4 μm/s.

4.6.1 Film thickness step samples

Figure 4-16 shows the results of measuring the step thickness of the oxide film formed on the Si wafer. The left film thickness is 4 μm, the right film thickness is 1 μm, the surface step is 3 μm, and the back-surface is measured as a flat surface correctly.



(a) Cross section (b) Front surface

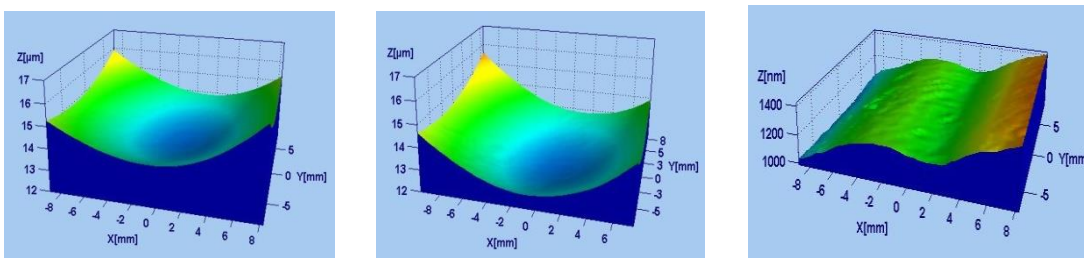
(c) Back surface (d) Thickness

Fig. 4-16 Measurement results of a SiO_2 film step on a Si wafer

4.6.2 Tapered film thickness sample

Figure 4-17 shows the measurement results of the tapered oxide film on the Si wafer using a wide-view type interference optical system with a field of view size of 200 mm (see Fig. 5-2).

A large concavely curved surface, a back-surface shape, and a tapered film thickness distribution are measured.

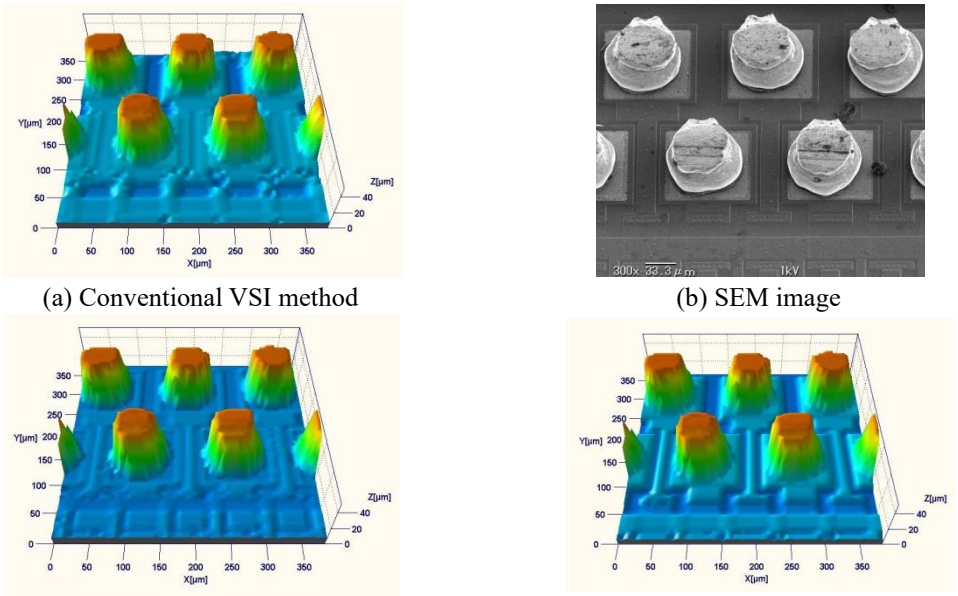


(a) Front surface (b) Back surface (c) Thickness

Fig. 4-17 Measurement results of a tapered oxide film on a Si wafer

4.6.3 Semiconductor IC sample

The semiconductor IC shown in Section 4.1 was measured and the results were compared with the results of the conventional method and SEM images (Fig. 4-18). Even in the region where the transparent film exists, the surface profile was measured without being affected by the sub-film pattern, which is consistent with the observation result of the SEM image.



(a) Conventional VSI method (b) SEM image

(c) Surface by KF method (d) Back surface by KF method

Fig. 4-18 Measurement results of IC surface partially covered by transparent film

4.6.4 Refractive index estimation method

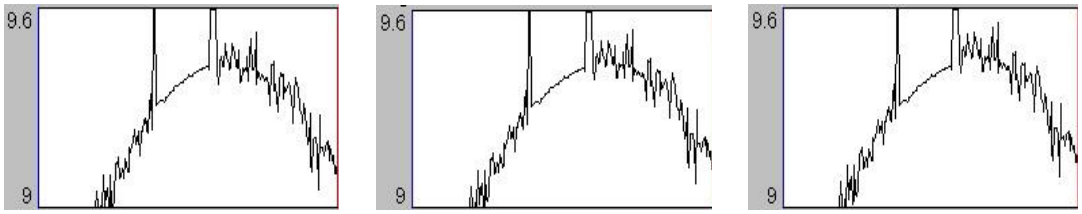
An attempt was made to measure the film shape of the resist film applied on the glass substrate. Since the refractive index of the resist film was unknown, part of the film was peeled off and the film shape was measured. The step (= physical film thickness t) of the exfoliated part was about 1.6 μm, and the optical film thickness nt was about 3.0 μm. From this, the refractive index n is estimated to be about 1.9.

In addition, noting that the substrate surface of the exfoliation part and the film back-surface are in the same plane with the correct refractive index, the back-surface profiles were calculated

with different refractive indexes as shown in Fig. 4-19. By observation, the refractive index n was estimated to be 1.90±0.02. Figure 4-20 shows the results of surface/back-surface shape and film thickness distribution measured using the obtained refractive index.

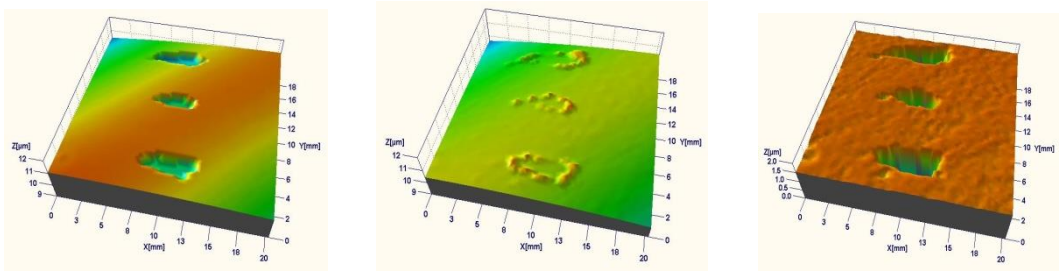
This refractive index estimation method can be automatically calculated if the back-surface is on the same plane and there are two regions with different film thicknesses. That is, when the surface step is Δs and the optical film thickness step is Δt, Δs is the step of the physical film thickness, so the refractive index n is obtained by:

$$n = \Delta t / \Delta s . \tag{4-9}$$



(a) n=1.88 (b) n=1.90 (c) n=1.92

Fig. 4-19 Back surface profiles calculated with different refractive indexes (n). The peeled area is located at center left.



(a) Front surface (b) Back surface (c) Thickness

Fig. 4-20 Measurement of resist film with the estimated refractive index

4.6.5 Measurement of water droplet surface

Figure 4-21 shows the measurement results of the water droplet on the contact lens. The diameter is about 1.5 mm and the height is about 8 μm. This is a good example showing this method's feature of high-speed 3D measurement, compared with the conventional film thickness meter that performs point measurement.

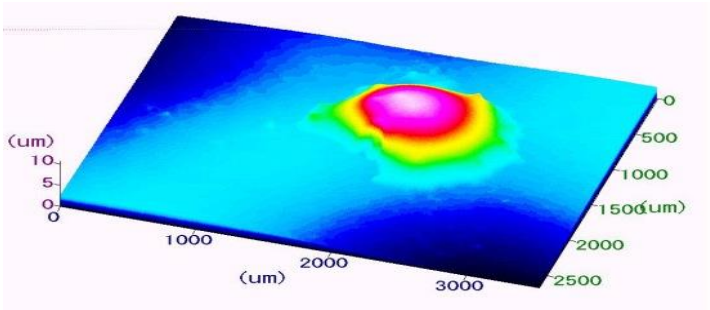


Fig. 4-21 Measurement of a water droplet

4.6.6 Repeatability

An oxide film sample with a nominal film thickness of 4 μm formed on a Si wafer was repeatedly measured. The repeatability of the average film thickness of 100 × 100 pixels is shown in Fig. 4-22. The average value = 4.056 μm, the standard deviation σ = 3 nm, and the CV value = 0.1% were obtained, confirming the very good repeatability.

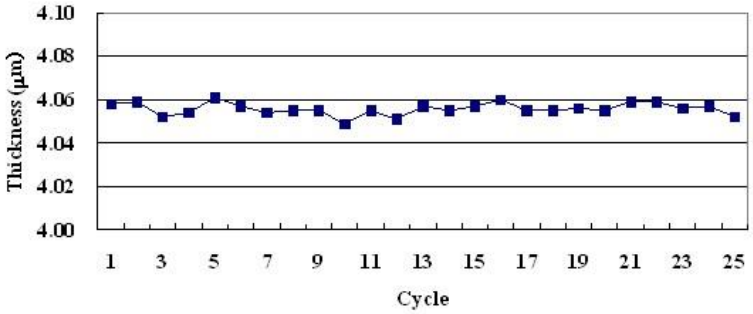


Fig. 4-22 Repeatability of film thickness measurement

4.7 Thin film algorithm（NF method）

When the film thickness of the transparent film is thinner than the coherence length, two peaks overlap as shown in Fig. 4-2, and it is difficult to separate the peaks by the KF method. The following techniques have been reported regarding the measurement method that targets such thin film [2)].

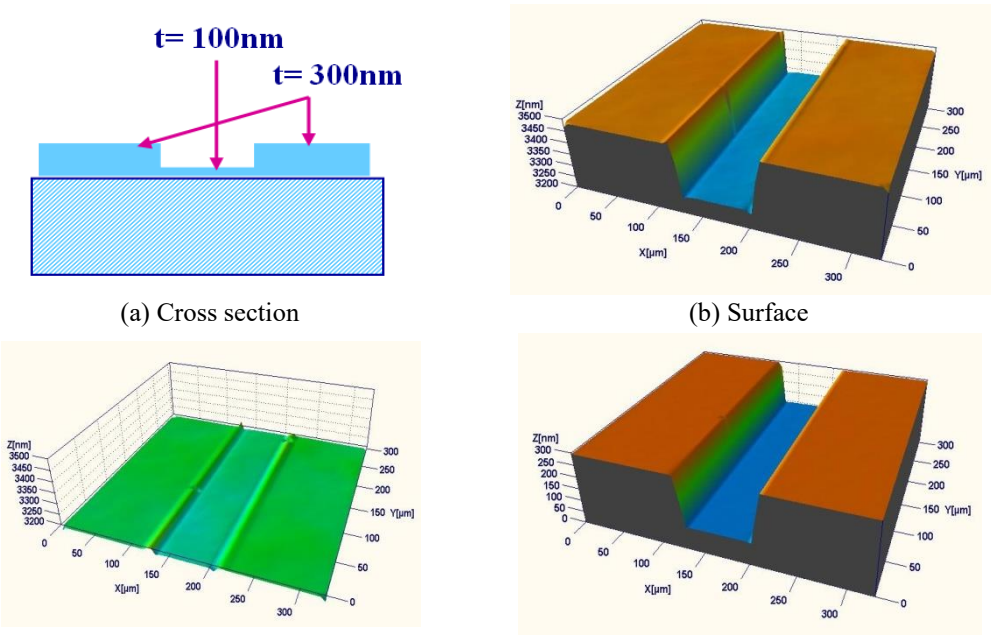
Kim et al. (1999) calculated the height and film thickness by Fourier transforming the interferogram of white light interference and non-linear least squares fitting the physical model to the obtained phase data. In the measurement experiment, a film thickness of 300 to 1000 nm is measured. Kim et al. (2002) used wavelength scanning interferometry that uses an AOTF (Acousto-Optic Tunable Filter) to sample two spectra with the reference optical path cut off and on. The film thickness is obtained from the spectrum at the path on and the surface profile is obtained from the spectrum at the path off. However, the film thickness of the experiment is 0.5 to 2 μm, and the result of thin film has not been reported. Akiyama et al. (2005) calculated the height and film thickness by Fourier transforming the interferogram of the wavelength scanning interferometer using AOTF and fitting the model function to the obtained amplitude-phase spectrum by nonlinear least squares method. The film thickness in the measurement experiment is 1 to 4 μm [3)].

In this way, there are problems such as computational cost and the need for special hardware, and no practical automatic measuring device can be found in the market.

[2)] Later, Ghim et al. (2006) proposed a method to obtain the height and film thickness simultaneously by Fourier transforming the spectrum obtained by the polarization spectroscopic interferometer with two built-in spectrometers, and least squares fitting the obtained phase with the theoretical model. The film thickness of the measurement experiment is 500 to 4000 nm, and the minimum measurable film thickness is about 500 nm. Furthermore, Mansfield (2006) proposed a thin film measurement method using the 'helical conjugate field' (HCF) function.

[3)] Later, Ueno et al. (2008) expanded the scanning wavelength band and measured a thin film of about 470 nm.

We have developed a thin film measurement algorithm based on a physical model (named NF algorithm from thinN Film) and released it in 2006 as the industry's first thin film surface profiler [Kitagawa 2006a]. The lower limit of measurable film thickness is about 50 nm for Si oxide film. Figure 4-23 shows the measurement results of the Si oxide film step of about 300 nm/100 nm. A surface step of about 200 nm and a flat back-surface shape are measured almost correctly.



(a) Cross section (b) Surface

(c) Back surface (d) Thickness

Fig. 4-23 Measurement results of a 300/100nm thickness step of SiO_2 film on a Si wafer

4.1 Summary

We have developed a signal processing algorithm (KF method) that achieves simultaneous measurement of surface height, back-surface height, and film thickness distribution of the surface covered with transparent film by positively utilizing the reflection signal from the film back-surface, which was conventionally treated as "disturbance". It can be applied to thick films with a film thickness of about 1 μm or more, and was put into practical use in 2002 as the world's first film profiler. We also proposed a refractive index estimation method that can be used when the back-surface is flat. Furthermore, we developed the NF algorithm applicable to thin film with a film thickness of about 1 μm or less, and put it into practical use in 2006 as the world's first practical thin film profiling algorithm.

This is an English translation of the article:
Katsuichi Kitagawa, "Chapter 4: Transparent Film Profilometry ",
in "Study on industrial surface profile measurement method based on optical interferometry", pp. 36-49, Doctoral Dissertation, University of Tokyo (2011)
(original in Japanese; translated by the author in Aug. 2020)