Available online at www.sciencedirect.com
SCIENCE DIRECT®
Mechatronics 15 (2005) 271–290
MECHATRONICS

Broadband optical end-point detection for linear chemical–mechanical planarization (CMP) processes using an image matching technique

Jingang Yi *, C. Shan Xu

Lam Research Corporation, 4540 Cushing Parkway, Fremont, CA 94538, USA

Accepted 1 September 2004

Abstract

In this paper we discuss an end-point detection (EPD) method for the dielectric linear chemical–mechanical planarization (CMP) processes. The proposed EPD algorithms utilize the interferometry optical signals to determine the post-CMP film thickness. A set of collected broadband spectral signals are formed as a spectral image. An image-matching technique is then used to match the processed signal image to the reference image template obtained at the target film thickness. Several matching criteria are discussed and compared. We find that the image correlation coefficient is a good indicator to determine the process end-point. We also consider the impact of the material removal rate variations on the interferometry spectral signals. An analytical calculation is carried out to find an extraction and compression searching range of the spectral image to compensate for the removal rate uncertainties. The correctness and effectiveness of the proposed algorithms have been demonstrated through applications to an inter-metal dielectric (IMD) device CMP process. Compared with other optical EPD methods, the proposed image-matching method is robust to the CMP process variations.

* Corresponding author.
E-mail addresses: jingang.yi@cal.berkeley.edu (J. Yi), shan.xu@lamrc.com (C.S. Xu).

0957-4158/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mechatronics.2004.09.006

Keywords: Semiconductor manufacturing; Chemical–mechanical planarization (CMP); End-point detection (EPD); Interferometry; Image matching

1. Introduction

Chemical–mechanical planarization (CMP) is one of the important semiconductor manufacturing processes. During a CMP process, the wafer surface is planarized by pushing the wafer down against a moving polishing pad. Fluid chemical slurry is dropped on the polishing pad to assist material removal on the wafer surface. Due to lack of in situ sensors and varying environmental conditions, it is difficult to monitor the process in real time.

One of the most important CMP process requirements is to determine when to stop the polishing process, namely, end-point detection (EPD). Over-polishing a wafer will result in deviation from the target film thickness thus degrading device performance or yield. On the other hand, under-polished wafers lead to re-work and an increase in the IC fabrication cost. Therefore, EPD plays a critical role for a CMP process. Various methods, such as optical end-point detection, friction sensing, electrical, and acoustic methods, have been proposed and partially applied in production. For a review of EPD methods for CMP processes, readers can refer to Hetherington and Stein [1] and references therein. In this paper, we focus on the optical EPD methods.

In Bibby and Holland [2], several interferometry methods were described for various CMP applications. For example, for a single layer transparent film such as blanket oxide wafers, a mathematical interferometry model can be built and model-based estimation can be matched with the optical signal to determine the film thickness. For multiple layer devices that mix transparent layers with metal layers such as inter-level dielectric (ILD) or inter-metal dielectric (IMD), it is difficult to obtain such a mathematical model. An empirical approach was utilized by checking the characteristic of the ratio of interferometry signals at two different wavelengths [2]. The accuracy to determine the end-point is heavily dependent on a stable material removal rate and benign process running conditions. Unfortunately, for most CMP processes, it is difficult to maintain such conditions.

In this paper, we enhance the optical EPD methods proposed by Bibby and Holland [2]. First, a set of time-consecutively collected broadband spectral signals are formed as a spectral image. We believe that the broadband spectral signals contain much richer film thickness information than only one wavelength that has been used in Bibby and Holland [2]. An image-matching technique is then used to match the processed signal image to the reference image template obtained at the target film thickness. Several matching criteria are proposed and compared. We find that the image correlation coefficient is a good candidate to determine the process end-point. We also consider the impact of the material removal rate variations on the interferometry spectral signals. The change of the material removal rate could distort the image frame. An analytical calculation is carried out to find an extraction and

compression range of the spectral image to compensate for the removal rate uncertainties in processes and to therefore find the exact process end-point.

The paper is organized as follows. In Section 2, we discuss the linear planarization technology (LPT) and the EPD system setup. In Section 3 we first present the processing of interferometry signals and formation of the signal image. Matching the image frame with the pre-defined template at the target film thickness is then discussed. Section 4 investigates how the material removal rate and in-coming film thickness variations affect the formed image frames. Using a scaling factor to compensate for the variations, an image matching scheme is also discussed. Some testing results are presented in Section 5. Finally we conclude this study in Section 6.

2. Optical end-point detection (EPD) systems

2.1. Linear chemical–mechanical planarization

The linear planarization technology provided by Lam Research Corporation is different from other CMP technologies. Fig. 1 shows the schematic of the Lam LPT setup. The polishing pad is moving linearly against the wafer while the wafer carrier is rotating. Both the polishing pad and the wafer carrier speeds are constant in a process. An air-bearing supports the polishing pad from underneath the air platen (see Fig. 1(b)). The wafer is vacuumed face-down by the wafer carrier against the polishing pad. In widely-used rotary CMP polishers, the polishing pad is on a rotating table with a relatively large radius. The wafer carrier is rotating against the polishing pad table. The polishing pad table is normally rigid, and the wafer head carrier sometimes can be tilted or controlled by the inside air zones. Compared with regular rotary CMP tools, the LPT design can provide a wide range of polish pad speed and polish pressure, and therefore increases the throughput as well as the planarization performance [3].

2.2. Optical EPD system

Fig. 2 shows the schematic of the optical EPD systems. A broadband spectral lamp is used to create the visible light (with a wavelength range 300–800 nm). Two

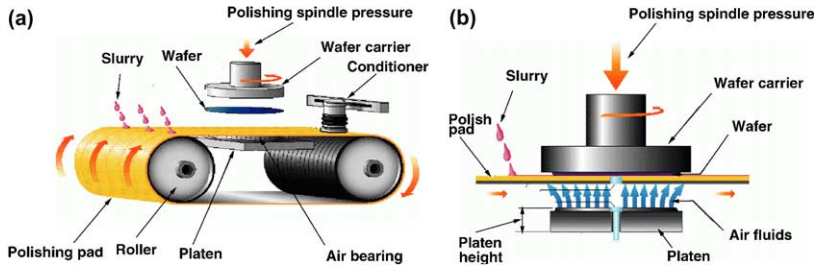

Fig. 1. (a) A schematic of Lam linear chemical–mechanical planarization systems, (b) polishing platen assembly.

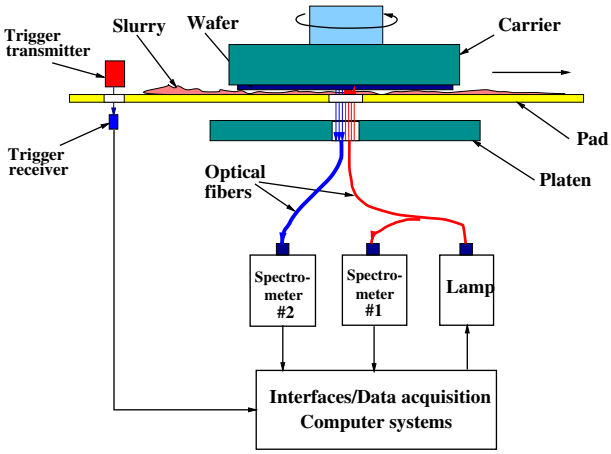

Fig. 2. A schematic of optical end-point detection systems.

spectrometers are used in the systems: spectrometer #1 measures the lamp source light intensity and spectrometer #2 is used to measure the reflective light intensity from the polishing wafer. A data acquisition and computer system samples the signal and a trigger system is used to control flashing the lamp. One or more EPD windows are opened at the center line of the polishing pad in order to let the light go through the platen and the pad. The optical fibers transit the incident light to the platen window and spectrometer #1 and also transmit back the reflective light to the spectrometer #2. Besides the wafer surface films, the incident and reflective lights go through several media including air, platen window, pad window, and slurry films. These media could distort the reflective light signals and make the optical EPD system less reliable and robust. As an example, Fig. 3 shows a typical broadband incident and reflective light intensities for an IMD device CMP process.

During a CMP process, the polishing pad is moving and per each revolution the EPD trigger system can collect one or more shots depending on how many EPD windows on the polishing pad. At each shot, both broadband incident and reflective signal intensities are sampled and stored in the computer system. We use these signals to develop a robust and reliable EPD algorithm.

3. Optical EPD method with image matching

3.1. Optical signal processing

At each EPD flash shot, we can measure the reflection light spectral information. We denote the reflection light intensity as $I_{\text{wafer}}(t;\lambda)$ and lamp intensity as $I_{\text{lamp}}(t;\lambda)$, respectively. We also define the silicon reference intensity, denoted as $I_{\text{Si}}(t;\lambda)$, by pushing a bare silicon wafer on the wafer carrier against the polishing pad and shoot-

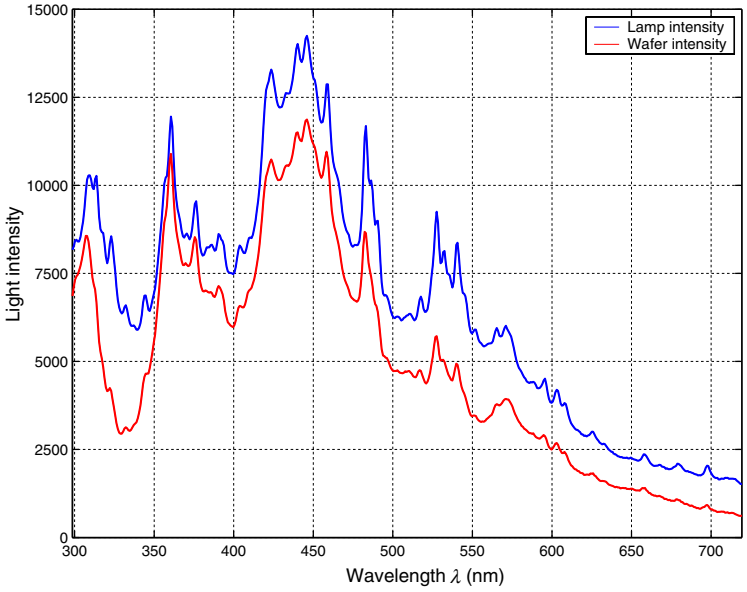

Fig. 3. An example of broadband incident and reflective intensities for an IMD CMP process.

ing the same lamp incident light $I_{\text{lamp}}(t;\lambda)$. We can then calculate the normalized reflection spectra as $I_{\text{norm}}(t;\lambda)$.

$$I_{\text{norm}}(t;\lambda) = \frac{\frac{I_{\text{wafer}}(t;\lambda)}{I_{\text{lamp}}(t;\lambda)}}{I_{\text{Si}}(t;\lambda)}. \tag{1}$$

It is observed that the use of normalized reflection spectra $I_{\text{norm}}(t;\lambda)$ can reduce the effects of various media in the optical path.

If we collect N shot interferometry signals in a CMP process and assemble the normalized reflection spectra in time sequence $I_{\text{norm}}(t_i;\lambda)$, $0 = t_0 \leqslant t_1 < t_2 < \cdots < t_N$, we can form a two dimensional (in time and wavelength domains) image. Normally, $\Delta T_s = t_i - t_{i-1} = \frac{L_B}{m v_B}$, $i = 1, 2, \ldots, N$ and L_B is the polishing pad length, v_B is the polishing pad linear speed, and m is the number of the EPD windows on the pad. Two low-pass filters [1] are utilized to smooth the normalized reflection spectra in both time and wavelength domains. Since the normalized EPD spectra image is correlated to the film thickness, we can then use these images to detect the process end-point by comparing the normalized spectral images with the template image at the target film thickness. This is the basic idea for the image matching based EPD method. For example, Fig. 4 shows three 2D normalized spectral images for three different IMD wafers at post-CMP thicknesses, 10,228 Å, 10,921 Å and 11,230 Å, respectively.

[1] In this application, a 5th order low-pass Butterworth digital filter is used to get rid of the system running noise.

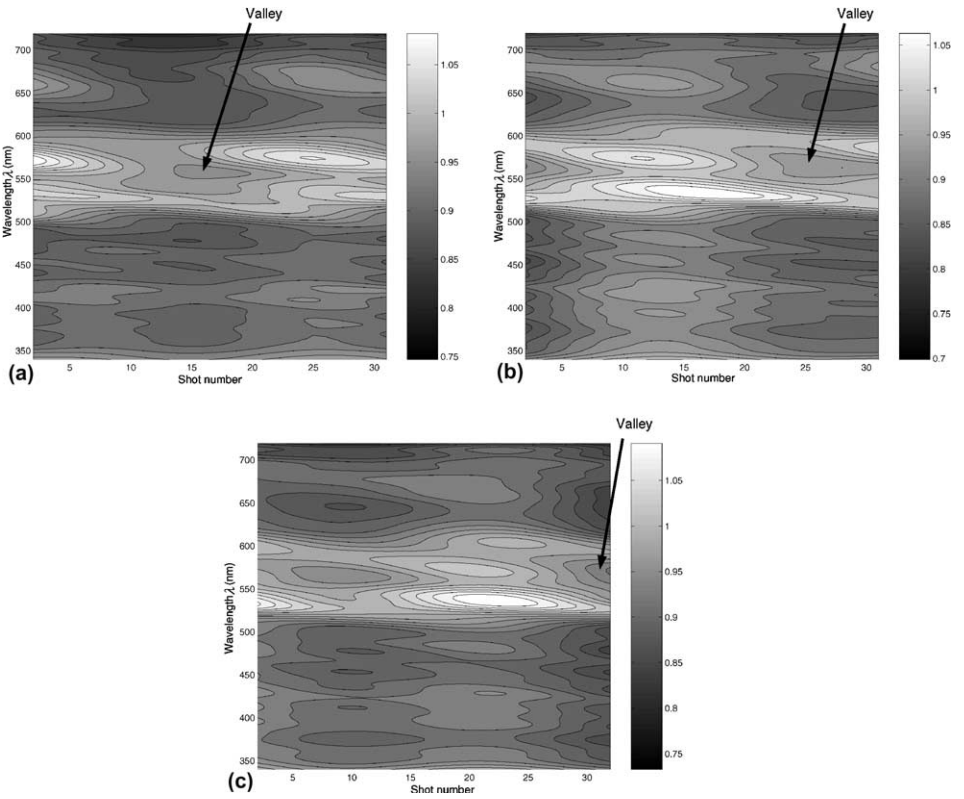

Fig. 4. 2D images of EPD data: (a) wafer #1, (b) wafer #2, (c) wafer #3.

If we take wafer #3 as the reference template, then we can find that at the time when polishing stops, the EPD signal reaches a "valley" around wavelength 570 nm. If we consider this "valley" on the other two wafer images (Figs. 4(a) and 4(b)), we can approximately see that the location of the "valley" is related to the post-thickness of the processes. If we want to determine the end-point of different wafers with larger thickness variations, we have to look through multiple wavelength information other than one fixed wavelength.

In order to enhance the signal-to-noise ratios, we can further improve the normalized reflection spectra $I_{\text{norm}}(t;\lambda)$ by dividing the broadband reflective light density by the intensity at a particular wavelength λ_0. Denote the relative normalized reflection spectra as $R_{\text{norm}}(t;\lambda)$ with

$$R_{\text{norm}}(t;\lambda) = \frac{I_{\text{norm}}(t;\lambda)}{I_{\text{norm}}(t;\lambda_0)}. \tag{2}$$

We can consider $R_{\text{norm}}(t;\lambda)$ as a gray-scale image since for each pixel there is only one numeric value that corresponds to the magnitude of the normalized reflection light intensity.

3.2. Comparisons among different image matching criteria

To formalize the development, we denote the template image as a matrix $\mathscr{T} \in \mathbb{R}^{M\times N}$ and a set of compared images as matrices $\mathscr{R} \in \mathbb{R}^{M\times N}$, where M and N are the numbers of spectrometer pixels and EPD shots used in the template and comparing images, respectively. Without loss of generality, we consider the template and the images to have the same size [2]. Choosing the size of wavelength range M and EPD shot number N depends on the real process setup and the EPD signals from CMP applications.

Suppose we have two images $\mathscr{R}$ and $\mathscr{T}$ with the same size. We need to compare and determine how similar the image $\mathscr{R}$ is to the template image $\mathscr{T}$. Both image values may have different magnitudes. Therefore, we need to normalize their values into the same scale. Moreover, due to the fact that there exist shot-to-shot variations, we have to smooth the EPD signals, or images, to eliminate these variations. We normalize the gray-scale image value between 0 and 1 by dividing the maximum element value in each of the image matrices.

Given a gray-scale image template $\mathscr{T}$ and an image $\mathscr{R}$, there are various algorithms to match $\mathscr{R}$ to $\mathscr{T}$. For example, the maximum distance d_{MD} (MD), defined as the maximum element distance between the image $\mathscr{R}$ to template image $\mathscr{T}$, is one indicator.

$$d_{\mathrm{MD}} = \max_{\substack{1\leqslant i\leqslant M\\ 1\leqslant j\leqslant N}} |\mathscr{R}_{ij} - \mathscr{T}_{ij}|. \tag{3}$$

If image $\mathscr{R}$ matches template $\mathscr{T}$ then the maximum distance should reach a minimum, or ideally zero. The sum of squares difference (SSD) is also a good indicator of the image matching.

$$\mathrm{SSD} = \sum_{1\leqslant i\leqslant M}\ \sum_{1\leqslant j\leqslant N} (\mathscr{R}_{ij} - \mathscr{T}_{ij})^2. \tag{4}$$

When $\mathscr{R}$ matches template $\mathscr{T}$, the SSD reaches its minimum value. In [4], a Hausdorff distance between two images d_{HD} is first calculated

$$d_{\mathrm{HD}} = \max_{\substack{1\leqslant i\leqslant M\\ 1\leqslant j\leqslant N}}\ \min_{\substack{1\leqslant k\leqslant M\\ 1\leqslant l\leqslant N}} |\mathscr{R}_{ij} - \mathscr{T}_{kl}|.$$

Then a Hausdorff fraction F_δ is defined as the portion of the image $\mathscr{R}$ is within a distance δ under the Hausdorff measure.

$$F_\delta = \frac{K_\delta}{MN}, \tag{5}$$

[2] Notice that in some cases we might compare a different size of EPD shots to the template due to the polishing removal rate drift; however, we can re-scale the image sizes to dimension $\mathbb{R}^{M\times N}$. We will discuss scaling images in Section 4.

where K_δ is the number of elements in $\mathscr{R}$ such that $d_{\mathrm{HD}} < \delta$. When F_δ is closest to 1, the image $\mathscr{R}$ is matching the template $\mathscr{T}$. We can also directly calculate the 2D correlation coefficient C between images $\mathscr{T}$ and $\mathscr{R}$

$$C = \frac{\sum_{1\leqslant i\leqslant M}\sum_{1\leqslant j\leqslant N}\mathscr{R}_{ij}\mathscr{T}_{ij}}{\left(\sum_i\sum_j\mathscr{R}_{ij}^2\sum_i\sum_j\mathscr{T}_{ij}^2\right)^{1/2}}. \tag{6}$$

Obviously, when $C = 1$, image $\mathscr{R}$ is the closest to $\mathscr{T}$. Maximum likelihood distance (ML) can be also used to measure the similarity of the image $\mathscr{R}$ and the template $\mathscr{T}$ [5]

$$\mathrm{ML} = \sum_{1\leqslant i\leqslant M}\sum_{1\leqslant j\leqslant N}\ln p(D_{ij}), \tag{7}$$

where $p(d) = \frac{1}{2\pi\sigma^2}\mathrm{e}^{-\frac{d^2}{2\sigma^2}}$ and distance D_{ij} is calculated as

$$D_{ij} = \min_{\substack{1\leqslant k\leqslant M\\ 1\leqslant l\leqslant N}}\left(|k-i| + |l-j| + \xi|\mathscr{T}_{ij} - \mathscr{R}_{kl}|\right).$$

In the above equation, ξ is the weight constant to ensure the same standard deviation for all directions of x, y, and z when calculating distance D_{ij}. When image $\mathscr{R}$ matches the template $\mathscr{T}$, ML reaches its maximum value.

Depending on the application, one of the above image matching criteria could be better than others. It is easy to see that the most computationally efficient algorithms are the maximum distance (MD), sum squares difference (SSD) and correlation coefficient C, as they only require $O(mn)$ computations. For a Hausdorff fraction calculation, it requires more comparisons and calculations than $O(mn)$ but less than $O(m^2n^2)$. The most expensive algorithm is maximum likelihood distance (ML) which requires $O(m^2n^2)$ computations. For effectiveness of the algorithms, it depends on the application. Generally speaking, maximum distance (MD) is the worst since it is very sensitive to the randomly large maximum difference between image $\mathscr{R}$ and template $\mathscr{T}$; and correlation coefficient C is the best for such random noises because it calculates a relative ratio, not a direct difference between $\mathscr{R}$ and $\mathscr{T}$.

In order to choose which method is suitable for the optical EPD application, we test various methods using experimental data. We polish six IMD wafers by a fixed time (around 30 EPD shots) and collect the EPD signals to compare various matching methods. The post-CMP thicknesses of these six wafers are listed in Table 1. We choose wafer #6 as the reference wafer for the image template and the template $\mathscr{T}$ is

Table 1
The most matching EPD shot numbers among wafers #1–#6 ($M = 136$ and $N = 10$) using correlation coefficient C

Wafer #	1	2	3	4	5	6
Post-thickness (Å)	10,228	10,489	10,407	10,639	10,921	11,230
Estimated EPD shot #	13	19	18	20	23	30
Estimated thickness at EPD (Å)	11,503	11,314	11,307	11,389	11,446	11,230
$C_{\max}$	0.967	0.997	0.997	0.998	0.997	1

calculated based on the reflective EPD signals between wavelength 464–611 nm and shots #21–30. Therefore, for template $\mathscr{T}$, $M = 136$ and $N = 10$. The reason we choose wafer #6 and shot #30 as the template $\mathscr{T}$ is that this wafer has the thickest post-polished film (11,230 Å) at shot #30 and thus we can always find a matched

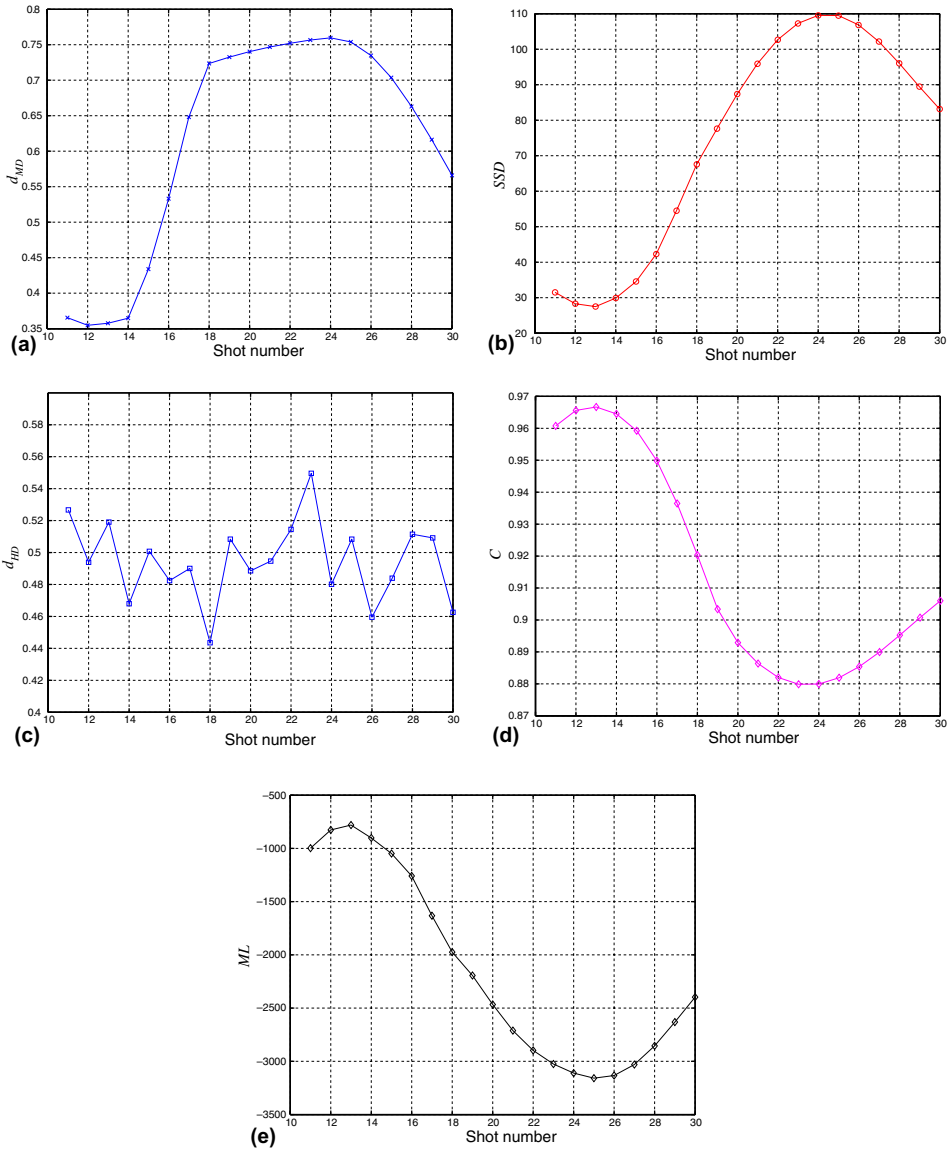

Fig. 5. Comparisons of different matching criteria for an example wafer #1 with template of wafer #6: (a) MD, (b) SSD, (c) HF, (d) *C*, (e) ML.

EPD shot for each of other five wafers. We then calculate the comparison results using different image matching algorithms discussed above.

Table 1 lists the best matching EPD shot for each wafer using the correlation coefficient C. In Table 1 we also estimate the film thickness at the matched EPD shot for each wafer. We find that using the maximum correlation coefficient $C_{\max}$ as the indicator to match image $\mathscr{R}$ with template $\mathscr{T}$ the estimated post-polished film thicknesses are very close to the target thickness (11,230 Å). The maximum error is less than 173 Å (for wafer #1). Fig. 5 shows one example of comparisons among various image matching algorithms for wafer #1. We compare the wafer #1 EPD image with the same size template. The best matching image for wafer #1 with the template from wafer #6 is at shot #13. From Fig. 5, we find that the minimum of MD is at shot #12, minimum of SSD is at shot #13, maximum of HF is at shot #23, maximum of C is at shot #13 and maximum of ML is also at shot #13. Using MD and Hausdorff fraction (HF) does not predict at the same EPD shot as other methods and, therefore, MD and HF are not suitable to determine the end-point. From the definition of MD and HF, we can find that these two methods use the maximum of comparison between $\mathscr{R}$ and $\mathscr{T}$, and are sensitive to the random noises in $\mathscr{R}$ and $\mathscr{T}$. Using SSD, correlation coefficient C and ML, we obtain the same EPD shot #13.

In this study we choose 2D correlation coefficient C to be the best image matching indicator for the EPD application for the following reasons. First, correlation coefficient C gives a normalized number between 0 and 1 to define the similarity between $\mathscr{R}$ and $\mathscr{T}$ and it is easy to use for comparison. Both SSD and ML define an absolute number (sum of difference) to compare $\mathscr{R}$ with $\mathscr{T}$. Therefore, using correlation coefficient C is more robust to any noises in $\mathscr{R}$ and $\mathscr{T}$. Moreover, using correction coefficient C has less computational complexity than the ML method.

4. Image matching algorithm with image scaling

4.1. Optical images scaling

During a CMP process, the material removal rate (RR) is changing with time due to uncertainties in the environmental conditions. Moreover, the removal rate is also drifting with consumables (such as polishing pad and conditioning disk) life-time. In the previous section, it has been assumed that the material removal rate for each process run follows the template trace exactly and, therefore, at a particular time the image can match with the template at the target film thickness. In order to compensate for removal rate variations and some other uncertainties in a process, we can re-scale the image within a range and match the re-scaled image with the template. In this section, we give an analytical discussion on how to find a scaling range and then outline the EPD algorithm.

The relative normalized reflection spectra $R_{\text{norm}}(t;\lambda)$ given by Eq. (2) is calculated by dividing the normalized intensity $I_{\text{norm}}(t;\lambda)$ by the normalized intensity at a specified wavelength λ_0. For a given wavelength λ, we can consider the bi-wavelength ratio $R_{\text{bw}}(t)$ as

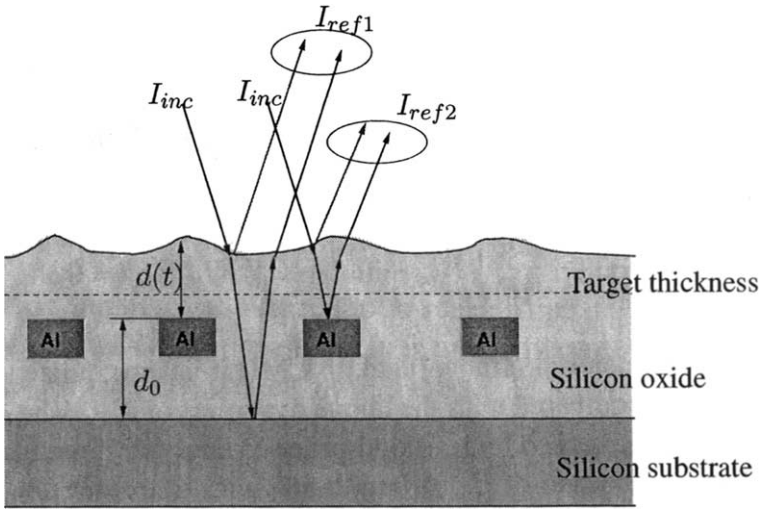

Fig. 6. Schematic of EPD optical interferometry signal path for one layer IMD film structure.

$$R_{\mathrm{bw}}(t) = R_{\mathrm{norm}}(t;\lambda).$$

In order to analyze the bi-wavelength signal $R_{\mathrm{bw}}(t)$, we need to find a mathematical model of $R_{\mathrm{bw}}(t)$ and then calculate the dynamics of the signal. As an example, Fig. 6 shows a simplified EPD interferometry optical path for an IMD film structure. The metal (aluminum) structure is embedded in the silicon oxide (SiO_2) film. The CMP process removes the oxide film topography to a target thickness while achieving the surface planarization.

Accurately capturing the model for the optical signals for IMD film is difficult due to the variations of micro-structures and CMP process conditions. Considering a simplified optical signal when the EPD incident light hits on the IMD wafer as shown in Fig. 6, the interferometry signals can be represented as a trigonometric function of film thickness. We can obtain the interferometry signals $R_{\mathrm{ref}}(t)$ at two different wavelengths λ and λ_0, respectively, as [3]

$$R_{\mathrm{ref}}(\lambda) = f(d(t)) = \alpha\cos(\omega d(t)) + \beta\cos(\omega(d(t)+d_0)), \tag{8}$$

$$R_{\mathrm{ref}}(\lambda_0) = g(d(t)) = \alpha\cos(\omega_0 d(t)) + \beta\cos(\omega_0(d(t)+d_0)), \tag{9}$$

where $\omega = \frac{4\pi n}{\lambda}$ and $\omega_0 = \frac{4\pi n}{\lambda_0}$ with $n = 1.52$ being the silicon oxide film reflective index. $d(t)$ is the oxide layer thickness from the metal bulk to the wafer surface and d_0 is the fixed oxide thickness under the metal bulk (see Fig. 6). $f(d(t))$ and $g(d(t))$ are

[3] For simplicity, here we directly write the interferometry signal $R_{\mathrm{ref}}(t) = I_{\mathrm{ref1}} + I_{\mathrm{ref2}}$ into a trigonometric function. I_{ref1} is the interferometry light intensity between reflective light from wafer surface and silicon substrate, and I_{ref2} is the interferometry light intensity between reflective light from wafer surface and metal bulk (Fig. 6).

functions of the film thickness $d(t)$ and light wavelength λ. The normalized bi-wavelength signal $R_{\mathrm{bw}}(t)$ can be represented as

$$R_{\mathrm{bw}}(t) = \frac{R_{\mathrm{ref}}(\lambda)}{R_{\mathrm{ref}}(\lambda_0)} = \frac{f(d(t))}{g(d(t))}. \tag{10}$$

Taking time derivatives twice for R_{bw} and from Eq. (10), we obtain [4]

$$\ddot{R}_{\mathrm{bw}} g + 2\dot{R}_{\mathrm{bw}} g' \dot{d} + \dot{d}^2 (R_{\mathrm{bw}} g'' - f'') = \ddot{d}(f' - R_{\mathrm{bw}} g'). \tag{11}$$

Notice that from Eqs. (8) and (9) we have

$$f'' + \omega^2 f = 0, \quad g'' + \omega_0^2 g = 0. \tag{12}$$

Plugging the above equations into Eq. (11), we can obtain the following dynamics for the bi-wavelength ratio signal R_{bw}

$$\ddot{R}_{\mathrm{bw}} + \left(2\frac{g'(d)}{g(d)} - \frac{\dot{RR}}{RR}\right)\dot{R}_{\mathrm{bw}} + (\omega^2 - \omega_0^2) RR^2 R_{\mathrm{bw}} = 0, \tag{13}$$

where $RR = \dot{d}(t)$ is the material removal rate. It is appropriate to assume that RR changes slowly compared with the above dynamics of R_{bw}. The dynamics of R_{bw} is approximately governed by a second-order ordinary differential equation and we can find that the solution of R_{bw} is oscillating with frequency ω_R

$$\omega_R(t) = RR\sqrt{(\omega^2 - \omega_0^2) - \left(\frac{g'(d)}{g(d)}\right)^2}. \tag{14}$$

The oscillating period of R_{bw} is

$$T(t) = \frac{2\pi}{\omega_R(t)}.$$

Normally we can pick λ and λ_0 such that $\frac{g'(d)}{g(d)}$ is small, and then we can estimate the oscillation period of R_{bw} as

$$T(t) \approx \frac{2\pi}{RR\sqrt{\omega^2 - \omega_0^2}} = \frac{\lambda\lambda_0}{2nRR\sqrt{\lambda^2 - \lambda_0^2}}. \tag{15}$$

With the theoretical estimate of the oscillation period of bi-wavelength signal R_{bw}, we can easily obtain the capability of the proposed image-matching EPD algorithm under RR drift. From Eq. (15), we know that the time period of R_{bw} is proportional to the reciprocal of material removal rate, $1/RR$. Denoting RR_{eq} as the nominal removal rate, we determine the oscillation period of R_{bw} as T_{eq} by Eq. (15). $RR_{\mathscr{R}} = (1-\epsilon)RR_{\mathrm{eq}}$, $|\epsilon| < 1$; we can then estimate the oscillation period $T_{\mathscr{R}}$ of R_{bw} correspondingly as

[4] Here the notation $f' = \frac{d(f(d(t)))}{d(d(t))}$ for the derivative of function $f(d(t))$ with respect to its argument $d(t)$ and $\dot{R}_{\mathrm{bw}} = \frac{dR_{\mathrm{bw}}(t)}{dt}$ for the time derivative of function $R_{\mathrm{bw}}(t)$. The same notation is used for higher derivatives and for other functions.

$$T_{\mathscr{R}} = \frac{\lambda\lambda_0}{2nRR_{\mathscr{R}}\sqrt{\lambda^2 - \lambda_0^2}} = \frac{T_{\text{eq}}}{1-\epsilon} = S_{\mathscr{R}} T_{\text{eq}},$$

where $S_{\mathscr{R}} = \frac{1}{1-\epsilon}$ is the scaling factor. Suppose that the image template length is $T_{\mathscr{T}}^{\text{win}} = N\Delta T_s$; from the above equation, we can estimate the process image size $T_{\mathscr{R}}^{\text{win}}$ as follows

$$T_{\mathscr{R}}^{\text{win}} = S_{\mathscr{R}} T_{\mathscr{T}}^{\text{win}}. \tag{16}$$

For a normal CMP process, the combined run-to-run and within-run removal rate variation should be less than 10%, namely, $-0.1 \leqslant \epsilon \leqslant 0.1$. We can then obtain the range of the $T_{\mathscr{R}}^{\text{win}}$ as

$$0.91T_{\mathscr{T}}^{\text{win}} \leqslant T_{\mathscr{R}}^{\text{win}} \leqslant 1.11T_{\mathscr{T}}^{\text{win}}.$$

The above equation implies that we need to re-scale the process image size about $\pm 10\%$ of its nominal value. For example, if $N = 10$, we need to search for the matching image with the size between 9 and 11 shots.

The incoming wafer film thickness is also varying due to the film deposition processes. From Eq. (15), it is easy to see that different film thicknesses change the phase angle of the signal $R_{\text{bw}}(t)$ and, therefore, shift the match image $\mathscr{R}$ with respect to the template image $\mathscr{T}$. However, the maximum film thickness shift is given by the delay of one EPD shot and can be estimated as $\Delta d_{\max} = RR \times \Delta T_s$. $\Delta d_{\max}$ is relatively small because the inter-shot period ΔT_s is small. For example, for a process with belt speed $v_B = 300\,\text{ft/min}$ (with pad length $L_B = 2.36\,\text{m}$ and one EPD window $m = 1$) and a nominal removal rate $RR = 3000\,\text{Å/min}$, $\Delta d_{\max} = RR \times \frac{L_B}{mv_B} = 78\,\text{Å}$.

4.2. Image match EPD algorithms

Considering the image matching EPD method discussed in Section 3 and the image scaling presented in the previous section, we propose EPD algorithms that calculate and determine the process end-point as follows. During a CMP process, we first collect the EPD signals. Once comparable shots (with pre-defined image template $\mathscr{T}$ at the target film thickness) have been collected to form an image $\mathscr{R}$, we then pre-process and re-scale the image data, and search the maximum correlation coefficient $C_{\max}$ with the template image within a given range. If $C_{\max}$ exceeds some threshold values, an end-point is called. Otherwise, we continue the process and update the image $\mathscr{R}$ with the newly collected EPD shots. If no active EPD is called, the process will be terminated by a fixed time [5]. Fig. 7 shows a schematic flow chart of the proposed image matching EPD algorithm.

[5] In this case, the process could be terminated by the previous wafer EPD process time or a fixed pre-defined processing time.

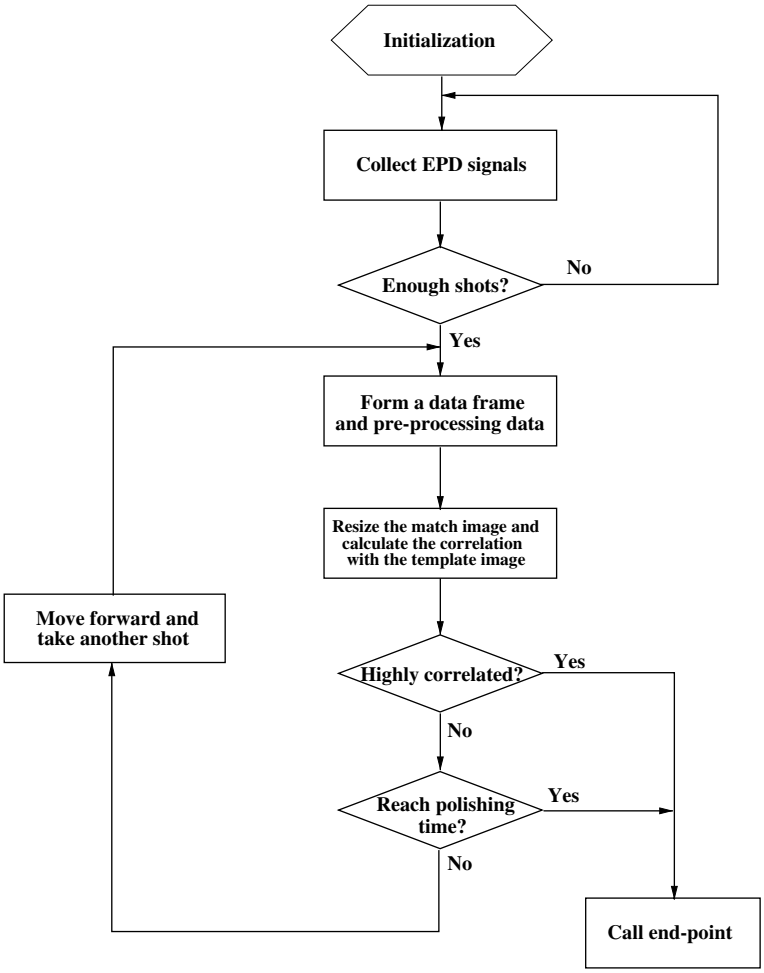

Fig. 7. A flow chart of the image matching EPD algorithm.

5. Experimental results

In this section, we illustrate a numerical example of the application of the proposed image matching based EPD method to an IMD device CMP process. We apply the proposed EPD algorithm for one lot of 25 IMD wafers from one production fab.

First we measured the pre-thicknesses d_i^{pre} of the ith IMD wafer and polished ith wafer for a certain time T_i. There is one EPD window on the processing pad ($m = 1$) and the belt speed $v_B = 300\,\text{ft/min}$. We measured the post-thickness d_i^{post} and estimated the average material removal rate $\overline{RR}_i$ for the ith wafer as

$$\overline{RR}_i = \gamma \frac{d_i^{\text{pre}} - d_i^{\text{post}}}{T_i}, \qquad i = 1, 2, \ldots, 25,$$

Table 2
Thickness estimate of the most matching image with wafer #24 with estimate thickness 11,442 Å ($M = 131$ and $N = 10$)

Wafer #	1	2	3	4	5	6	7	8	9	10	11	12
Est. d_i (Å)	11,642	11,821	11,697	11,751	11,742	11,680	11,649	11,657	11,665	11,874	11,617	11,649
Err. e_i (Å)	200	379	255	309	300	238	207	215	223	432	175	207
$C_{\max}$	0.938	0.999	0.991	0.993	0.989	0.999	0.998	0.995	0.973	0.996	0.999	0.982
$\overline{RR}$ (Å/min)	3250	3063	3141	3068	3092	3106	3082	2997	3149	3013	3066	3021
$S^*_{\mathscr{R}}$	1.08	0.96	1.01	0.91	0.91	1.01	1.02	1.05	0.91	0.91	1.02	0.91
	13	14	15	16	17	18	19	20	21	22	23	25
Est. d_i (Å)	11,726	11,706	11,506	11,789	11,679	11,892	11,571	11,848	11,887	11,814	11,457	11,675
Err. e_i (Å)	284	264	64	347	237	450	129	406	445	372	15	233
$C_{\max}$	0.985	0.995	0.985	0.999	0.949	0.991	0.998	0.995	0.998	0.989	1	0.977
$\overline{RR}$ (Å/min)	3071	3008	3062	2922	3051	2999	2994	2957	2975	2953	3013	3109
$S^*_{\mathscr{R}}$	0.91	1.06	1.10	1.04	1.02	1.07	1.07	1.03	0.91	0.96	1.08	1.02

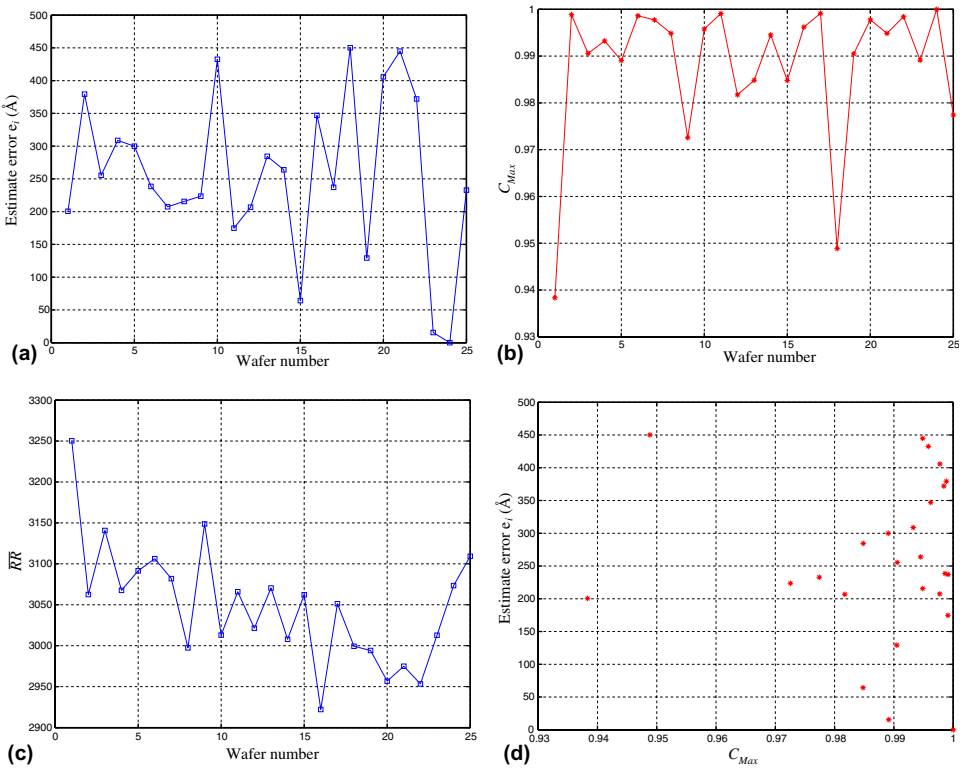

Fig. 8. Image matching results with wafer #24 (with 10-shot template and 0.9–1.1 re-scaling): (a) thickness estimate errors, (b) maximum correlation coefficients $C_{\max}$, (c) estimate removal rate $\overline{RR}$, (d) relationship between estimate thickness error e_i and maximum correlation coefficients $C_{\max}$.

where $\gamma = 0.45$ is a factor for averaging the wafer film topography effect on the removal rate [6]. During the polishing process, we collected optical EPD signals for each wafer. We picked the thickest post-polishing wafer #24 as the template image $\mathscr{T}$. Since a total of 30 EPD shots have been collected during polishing wafer #24, we choose shots #21–30 ($N = 10$) to form the template $\mathscr{T}$. In the test, we choose the wavelength λ between 505 nm and 612 nm for the template $\mathscr{T}$ and $\lambda_0 = 505$ nm. There are a total of $M = 131$ spectrometer pixels in this wavelength range.

Then we use the proposed EPD algorithm to capture the end-point shots for each wafer in the same lot. Denote the estimated end-point shot for wafer i as s_i; then we can estimate the film thickness d_i at shot s_i as

[6] For an IMD pattern wafer, the material removal rate is decaying during a CMP process due to the wafer surface topography [6]. After the wafer surface is planarized, the removal rate is close to the blanket oxide wafer removal rate. Therefore, we use the factor γ to compensate for the changing of wafer surface topography. The value of topography factor γ depends on the pattern characteristic and is varying among different device wafers.

$$d_i = d_i^{\text{post}} + \overline{RR}_i \times (T_i - t_{s_i}), \quad i = 1, 2, \ldots, 23, 25,$$

where t_{s_i} is the polishing time at shot s_i. Table 2 shows the thickness estimate error

$$e_i = d_i - d_{24}^{\text{post}}, \quad i = 1, 2, \ldots, 23, 25$$

for the other 24 wafers based on the "best" image matching of the template. In Table 2 we also list the scaling factor $S^*_{\mathscr{R}}$ at which the correlation coefficient reaches the maximum value $C_{\max}$.

In Table 2, we use an image $\mathscr{R}$ of a size 131×10, and a range of 0.9–1.1 re-scaling (with bicubic interpolation between data points) has been used in the process. Table 4 illustrates the sum of square errors (SSE) of the thickness estimate with the template wafer #24 for the incoming and post-thicknesses. SSE is calculated as

$$\text{SSE} = \sqrt{\frac{\sum_{i=1}^{25} e_i^2}{25}}.$$

Fig. 8 shows the image matching results with wafer #24 as the template. The maximum estimate error is around 450 Å. If we take a look at the correlation coefficients, we find that most wafers have high correlation coefficients. There is no significant relationship between maximum correlation coefficients and the estimate thickness errors (see Fig. 8(d)). From Fig. 8(a) and (d), we can see that the proposed image matching algorithms work well with the estimate of the film thickness.

Fig. 9 shows a normalized bi-wavelength optical signal for three wafers: #6, 9 and 24. The calculation of such a signal is given by Eq. (2) with $\lambda = 570\,\text{nm}$ and

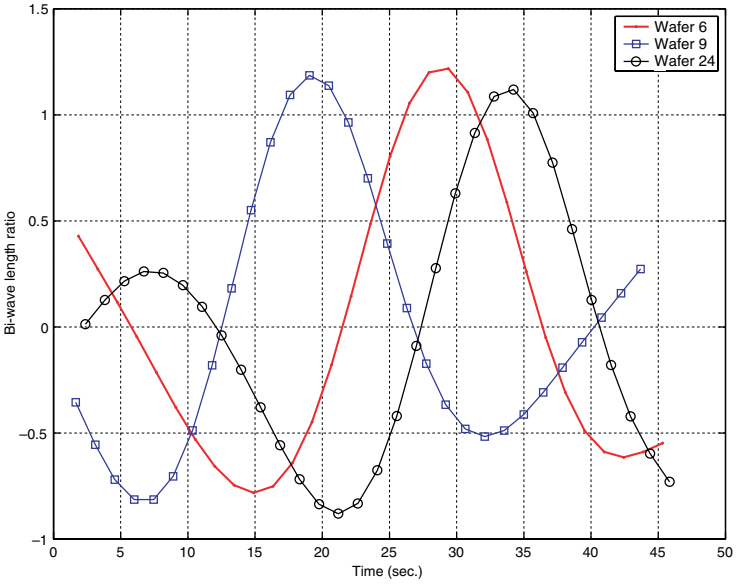

Fig. 9. Bi-wavelength ratio signal for three wafers #6, 9 and 24.

Table 3
Thickness estimate of the most matching bi-wavelength with wafer #24 with a post-thickness 11,442 Å

Wafer #	1	2	3	4	5	6	7	8	9	10	11	12
Est. d_i (Å)	11,907	11,941	11,885	11,971	12,082	12,127	11,945	11,951	12,043	12,161	11,866	12,056
Err. e_i (Å)	464	499	443	529	640	685	503	509	601	719	424	614
	13	14	15	16	17	18	19	20	21	22	23	25
Est. d_i (Å)	12,010	12,087	11,979	12,153	12,102	12,355	11,953	12,186	12,293	12,130	11,909	11,760
Err. e_i (Å)	568	645	537	711	660	913	511	744	851	688	467	318

Table 4
SSE calculations of estimated post-thicknesses with the template wafer #24

	Image matching method	Single-wavelength matching method
SSE (Å)	280	597

$\lambda_0 = 520\,\text{nm}$ [7]. We also subtract the mean value of the signal. The post-thicknesses for these three wafers are 11,230, 10,905 and 11,219 Å, respectively. From Fig. 9 we find that the normalized signal R_{norm} is like a periodic sinusoidal curve and that we have to specify the particular time period in order to determine the end-point. Otherwise, a bi-wavelength algorithm cannot give out a correct EPD location. Moreover, since the bi-wavelength signal is periodic, it is difficult to use the bi-wavelength signals to distinguish two wafers that have significantly different in-coming thicknesses. We can pre-define an algorithm activation time $T_a = 25\,\text{s}$ to rule out false EPD calls due to the periodic signals. In order to compare with the estimation results by the image matching method, we then compute the bi-wavelength signal for the 25-wafer set and match with the wafer #24. Table 3 illustrates the estimated post-thicknesses d_i and estimation errors e_i at the end-point predicted by the bi-wavelength matching method. We also calculate SSE for the bi-wavelength matching method as shown in Table 4. The comparison results in Table 4 clearly demonstrate that the CMP processes with the proposed EPD method have more uniform post-CMP film thicknesses than those of using bi-wavelength matching method.

6. Conclusion

In this paper, we proposed and developed a method to detect the end-point of dielectric CMP processes using an image matching technique. Optical interferometry signals have been collected as the raw data for the proposed method. Some data processing techniques are utilized to eliminate the signal disturbance from the slurry fluid films and other components in the optical path. A two-dimensional image (in time and wavelength) is formed based on the normalized EPD signals. Among different image matching algorithms, we demonstrated that the correlation coefficient is the best indicator for the relationship between the real-time EPD signals and the pre-defined image template collected at the target film thickness. We also considered the effects of the material removal rate variations on the EPD image signals. An analytical calculation was utilized to find an extraction and compression range of the spectral image to compensate for the removal rate uncertainties. The proposed algorithms have been tested through a numerical example using the CMP data of an inter-metal dielectric (IMD) device from fab production wafers. Compared with other optical EPD methods, the proposed image-matching method is robust to the CMP process variations.

[7] The reason that we choose these two particular wavelengths is that under such a choice the normalized bi-wavelength signal R_{norm} has a clear variation.

References

[1] Hetherington DL, Stein DJ. Recent advances in endpoint and in-line monitoring techniques for chemical–mechanical polishing processes. In: Proceedings of CMP-MIC Conference, Santa Clara, CA, 2001. p. 315–23.
[2] Bibby T, Holland K. Endpoint detection for CMP. J Electron Mater 1998;27(10):1073–81.
[3] Jensen A, Renteln P, Jew S, Raeder C, Cheung P. Empirical-based modeling for control of CMP removal uniformity. Solid State Technol 2001;44(6):101–6.
[4] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. IEEE Trans Pattern Anal Machine Intell 1993;15(9):850–63.
[5] Olson CF. Maximum-likelihood image matching. IEEE Trans Pattern Anal Machine Intell 2002;24(6):853–7.
[6] Kim SY, Seo YJ. Correlation analysis between pattern and non-pattern wafer for characterization of shallow trench isolation-chemical–mechanical polishing (STI–CMP) process. Microelectron Eng 2002;60:357–64.