A scanning wafer thickness and flatness interferometer

Maarten J. Jansen^a, Han Haitjema^a, Peter H.J. Schellekens^a

^aEindhoven University of Technology, Precision Engineering Section, Eindhoven, The Netherlands

ABSTRACT

Since the wafer industry is using an increasing amount of double side polished wafers, the future of wafer metrology is very likely to shift from capacitance gauging techniques to optical measurement techniques. Participating in an international project the Technical University of Eindhoven is developing a self calibrating, traceable double side wafer flatness and thickness measurement device. By using proper measurement principles and advanced software a robust and traceable wafer thickness and flatness measurement instrument is created which combines high lateral resolution, nanometer accuracy, high speed and low cost.

Keywords: wafer measurement, scanning interferometer, thickness, flatness, unwrapping, stitching

1. INTRODUCTION

For characterization of wafer thickness and flatness many challenges arise. Capacitance gauging techniques will give problems measuring wafers of strongly varying conductivity. Transmissive thickness measurements using an infrared interferometer will fail for non-transparent heavily doped wafers. A solution for measuring these wafers is found in simultaneously measuring the front side and back side flatness of a wafer using a Fizeau interferometer. With both flatness maps aligned accurately a thickness map can be derived. For chip production wafer thickness variation is the most critical parameter to be measured. Since the price of optical components tends to grow exponentially with size, the cost of ownership of a full size double side interferometer for measuring up to 300 mm silicon wafers is considerable. Also for a full size aperture interferometer large flatness deviations may result in a fringe density which cannot be measured in a single measurement. Due to the increased lateral resolution of a scanning interferometer it has the potential of higher tolerances on tilt alignment than a full size interferometer.

To overcome the difficulties, the Eindhoven University of Technology has developed a scanning double side surface interferometer for measuring the flatness and thickness variation of up to 300 mm double side polished wafers.

In this paper the working of the wafer geometry measuring device is explained. The flatness measurement and the stitching algorithm for combining local measurements into a global flatness map are described. Also the calculation of wafer thickness variation and its error compensation techniques are explained.

2. MEASUREMENT PRINCIPLE

In figure 1 the schematic setup is shown. The parallel measurement beam from a stabilized HeNe laser is split into a left part and a right part. The left part measures the distance between the rear reference flat and the back side of the wafer (A), while the right part measures the distance between the front reference flat and the front side of the wafer (B). Equation 1 describes how the wafer thickness (T) can be derived from the flatness measurements (A and B) and the calibrated distance between the reference flats (C) (See figure 1).

$$T = C - A - B \tag{1}$$

Further author information: (Send correspondence to Maarten J. Jansen.)

M.J.Jansen: E-mail: m.j.jansen@tue.nl, Phone: +31 (0) 40 247 4507, Address: Eindhoven University of Technology, Precision Engineering Section, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Han Haitjema: E-mail: h.haitjema@tue.nl, Phone: +31 (0)40 247 3715,

Peter H.J. Schellekens: E-mail: p.h.j.schellekens@tue.nl, Phone : +31 (0) 40 247 2620

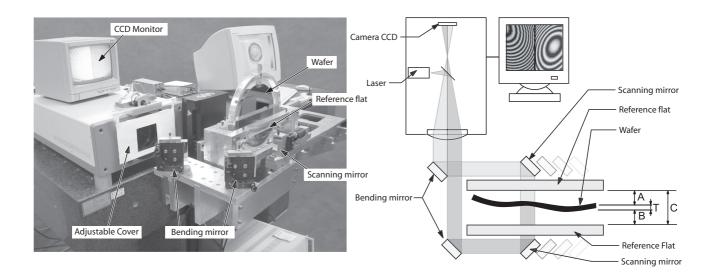


Figure 1. Left: Pre-Prototype measurement setup, Right: A schematic view S_0 S_{90} S_{180} S_{270} Phasemap

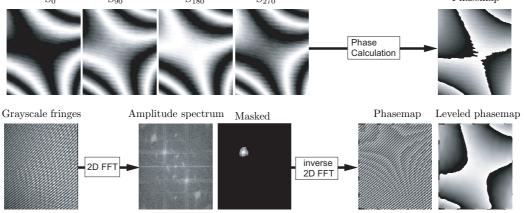


Figure 2. Phase acquisition, Top: Temporal phase shifting method, Bottom: Spatial carrier fringe method

Both sides of the wafer are measured simultaneously. The interference fringes from the front side and back side measurement are projected on a single CCD, eliminating the need for synchronizing electronics. The height difference between neighbouring fringes equals half the wavelength of the light. The flatness maps can be determined by means of a phase shifting method or by means of a spatial carrier fringe method^{1,2} For the phase shifting method multiple phase shifted interferograms are recorded. The phase shift is introduced by translating the reference flats relatively to the wafer by means of a piezo element. The spatial carrier fringe method can be used by tilting the wafer. The flatness of the wafer on both sides can now be derived from only a single interferogram. The pre-prototype version is a step and measure instrument, which measures a surface area of 25x45 mm at a time. Horizontal scanning is done by moving the scanning mirrors alongside of the rectangular reference flats. The vertical scanning is done by moving the wafer in between the reference flats. This way a full coverage of wafers up to 300 mm diameter is obtained. To reduce the fringe density while measuring heavily bowed wafers, the machine has a tilt adjusting mechanism. The wafer is locally aligned to be approximately parallel to the reference flats at all times. As a result the amount of fringes is minimized and the alignment specifications of the front side flatness map on the back side flatness map for obtaining thickness is reduced. This also enables the thickness and flatness measurement of wafers with rather bad flatness figures.

3. PHASE UNWRAPPING

To convert a phasemap into a flatness map phase jumps of magnitude 2π have to be removed from the wrapped phasemap. This conversion step is called unwrapping.

3.1. basic unwrapping

A basic least-means-squares estimate of an unwrapped phasemap is made by evaluating the x-gradients and y-gradients of a phasemap.³ For this the gradients are normalized between $-\pi$ and $+\pi$ by adding or subtracting a multiple of 2π for each element of the gradient of the wrapped phasemap ϕ^w .

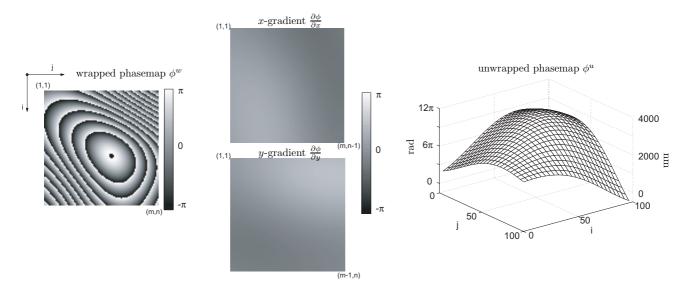


Figure 3. Left: wrapped phasemap ϕ^w , center: normalized phase gradients $\frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y}$, Right: unwrapped phasemap ϕ^u , obtained from the phasemap gradients

$$\left. \frac{\partial \phi}{\partial x} \right|_{i,j} = \left(\phi_{i,j+1}^u - \phi_{i,j}^u \right) = -\pi + \left(\pi + \phi_{i,j+1}^w - \phi_{i,j}^w \right) \text{ modulo } 2\pi$$
(2a)

$$\left. \frac{\partial \phi}{\partial y} \right|_{i,j} = \left(\phi_{i+1,j}^u - \phi_{i,j}^u \right) = -\pi + \left(\pi + \phi_{i+1,j}^w - \phi_{i,j}^w \right) \text{ modulo } 2\pi$$
(2b)

Setting the gradients of the unwrapped phasemap to be equal to the normalized gradients of the wrapped phasemap a set of linear equations 2a and 2b can be defined for all neighbouring phases. For solving these equations a set of linear equations can be written (eq. 3) in the form of $H\hat{x} = \hat{b}$.

$$\begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & 1 & \\ & & & \ddots & \ddots & \\ & & & -1 & 1 & \\ & -1 & \dots & 1 & & \\ & & -1 & \dots & 1 & \\ & & & \ddots & & \ddots & \\ & & & & -1 & \dots & 1 \\ \hline 1 & 0 & \dots & \dots & 0 \end{bmatrix} \cdot \begin{bmatrix} \phi_{1,1}^{u} \\ \phi_{1,2}^{u} \\ \phi_{1,3}^{u} \\ \vdots \\ \phi_{2,2}^{u} \\ \vdots \\ \phi_{m,n}^{u} \end{bmatrix} = \begin{bmatrix} \frac{\partial \phi}{\partial x} \Big|_{1,1} \\ \frac{\partial \phi}{\partial x} \Big|_{1,2} \\ \vdots \\ \frac{\partial \phi}{\partial x} \Big|_{m,n-1} \\ \frac{\partial \phi}{\partial y} \Big|_{m,1} \\ \vdots \\ \frac{\partial \phi}{\partial y} \Big|_{m-1,n} \end{bmatrix}$$
(3)

1

H is a connection matrix which connects all unknown neighbouring phases of \hat{x} according to eq. 2a and 2b. \hat{b} contains the normalized phase gradients of ϕ^w . The last row of matrix H creates a ground reference to make the set of equations solvable. The unwrapped phases are now calculated by $\hat{x} = (H^T H)^{-1} H^T \hat{b}$.

3.2. robust unwrapping

The robustness of the unwrapping algorithm is very important for working in an automated environment. Many 2D unwrapping methods have been studied and developed. To improve the robustness of a phase unwrapping algorithm first the unwrapping problem is simplified by splitting the wrapped phasemap into three parts:

- estimate of the unwrapped phasemap
- noise term
- wrapped residual phasemap

Most phasemaps can be unwrapped using equation 3. To further improve the robustness and to make the unwrapping algorithm insensitive to noise a smooth estimate of the unwrapped phasemap is constructed.

3.3. estimate of the unwrapped phasemap

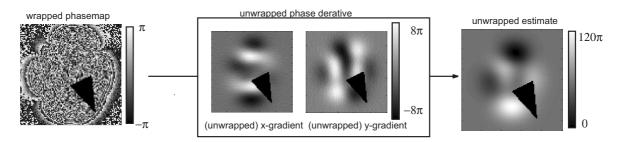


Figure 4. Creating an estimate (right) of a wrapped phasemap (left) by observing the unwrapped normalized phase gradients (center)

If phase jumps appear in the phase gradient maps, then the initial phasemap has one or more subsampled regions. In this subsampled example the gradients of the phasemap have to be unwrapped as well using equation 3 before continuing. The obtained phase gradients now exceed the $[-\pi, +\pi]$ range. From these gradients a unwrapped phase estimate is obtained using equation 3.

3.4. extracting noise from a phasemap

For high noise levels the least-means-squares unwrapping method of section 3.1 is limited, because the normalized phase gradient may not be an accurate enough estimator for the true gradient. However for oversampled phasemaps noise can easily be separated from the wrapped phasemap ϕ^w by using a averaging phase convolution filter. High noise levels in a subsampled phasemap generally require an iterative approach. A noise term is obtained by applying a convolution filter of the phasemap minus the estimated phasemap. For subsampled phasemaps a noise level of up to 40% is allowed by the unwrapping algorithm.

3.5. unwrapping of the wrapped residual

The wrapped residual phasemap generally is smooth and has none or only a few phase jumps. This phasemap can be unwrapped by a simple zigzag unwrapping algorithm. If the unwrapped residual phasemap still appears to have unwrapping errors, then the described unwrapping approach can be repeated until a satisfactory phase estimate is obtained. This is likely to occur for subsampled phasemaps with a high noise ratio.

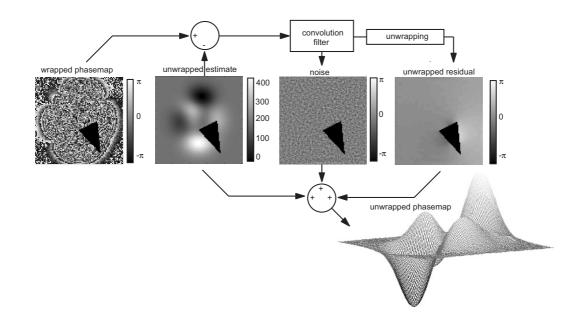


Figure 5. The unwrapped phasemap is obtained after recombining the phase estimate the noise term and the unwrapped residual phasemap

3.6. Unwrapping result

Now we can sum the estimated unwrapped phasemap, the noise term and the unwrapped residual to obtain the unwrapped phasemap (figure 5). For large phasemap it can be favourable to use a blockprocessing unwrapping algorithm. The phasemap is then cut into blocks that have overlapping regions with neighbouring blocks. All blocks are then unwrapped separately and combined again by a surface stitching algorithm (see section 4.1). The blockprocessing approach results in lower memory usage and higher speed for very large phasemaps.

4. FLATNESS MEASUREMENT

The interferometer measures frontside and backside flatness of a small area of the wafer surface. The field of view of the interferometer is limited by the aperture size of the interferometer and by the size of the scanning mirrors. Multiple measurements have to be conducted to cover a measuring area that is larger than the field of view.¹ For wafers with a small flatness deviation all flatness phasemaps can be recorded without realigning the wafer tilt position. The recorded phasemaps can simply be put together to obtain the phasemap of the wafer flatness. For measuring heavily bowed wafers a tilt alignment mechanism is required to prevent the interferometer to get out of range. Now the separate flatness maps must be connected using a surface stitching algorithm. Due to small measurement errors a propagation of errors will occur using the stitching procedure. The correction of lens aberration will prove to reduce the error propagation significantly.

4.1. Surface Stitching

A surface stitching software package has been developed for reconstructing the flatness map of an entire wafer from separate locally aligned flatness maps. The stitching software adjusts the tilt and height of partly overlapping measurements so that the overlapping areas have the same tilt and height position. The surface grids of all measurements are resampled to a uniform grid. This facilitates the stitching of surfaces of different rotational orientation or different gridsize.

Figure 7 show how the stitch points are chosen. A stitch point is a point where two surfaces are jointed together in z-direction. Each stitch point fixes one degree of freedom. A surface is assumed to have three degrees of freedom (that is x-tilt, y-tilt and z-position). One stitch point creates a joint-hinge with two degrees of

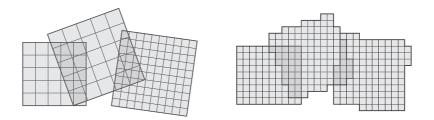


Figure 6. Left: Original surface grids, Right: Surface grids resampled to a uniform grid

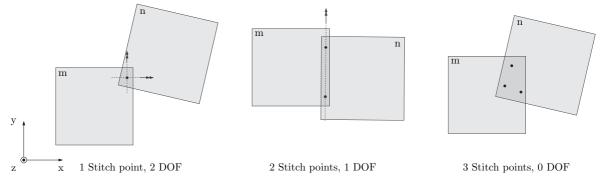


Figure 7. Left: Ball joint (2 DOF), Center: Line joint (1 DOF), Right: Fixed joint (0 DOF)

freedom. It allows free relative rotation around the x-axis and y-axis. Two stitch points create a joint-line with one degree of freedom. It allows free rotation around the intersecting line. Three or more stitch points completely fix one surface onto another. To automate the generation of stitch point positions the overlapping region is split

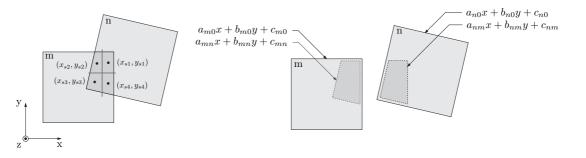


Figure 8. Left: Position of stitch points, Right: Regions through which planes are fitted (see also eq. 4)

in its center into four quadrants. One stitch point is chosen in the center of each quadrant (left of figure 8). Equation 4 can be written for a stitch point between a surface m and a surface n at the coordinate (x_{sj}, y_{sj}) :

$$-\begin{bmatrix} x_{sj} & y_{sj} & 1 \end{bmatrix} \begin{bmatrix} a_m \\ b_m \\ c_m \end{bmatrix} + \begin{bmatrix} x_{sj} & y_{sj} & 1 \end{bmatrix} \begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix} = b_{mnj}, \ b_{mnj} = \begin{bmatrix} x_{sj} & y_{sj} & 1 \end{bmatrix} \begin{bmatrix} a_{m0} - a_{mn} \\ b_{m0} - b_{mn} \\ c_{m0} - c_{mn} \end{bmatrix} - \begin{bmatrix} x_{sj} & y_{sj} & 1 \end{bmatrix} \begin{bmatrix} a_{n0} - a_{nm} \\ b_{n0} - b_{nm} \\ c_{n0} - c_{nm} \end{bmatrix}$$
(4)

$a_{n0}x + b_{n0}y + c_{n0}$:	plane fitted through surface n before stitching
$a_{m0}x + b_{m0}y + c_{m0}$:	plane fitted through surface m before stitching
$a_{nm}x + b_{nm}y + c_{nm}$:	plane fitted through the overlapping region of surface n with surface m before stitching
$a_{mn}x + b_{mn}y + c_{mn}$:	plane fitted through the overlapping region of surface m with surface n before stitching
$a_n x + b_n y + c_n$:	unknown best fit plane through surface n after stitching
$a_m x + b_m y + c_m$:	unknown best fit plane through surface m after stitching
(x_{sj}, y_{sj})	:	coordinates of a stitchpoint

When combining the equations of all stitch points between all surfaces a set of linear equations can be written in matrix notation of the form $H\hat{x} = \hat{b}$:

Γ	$-x_{s1}$	$-y_{s1}$	-1	x_{s1}	u_{s1}	1		0	0	0 -	1	$\begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix}$		[b ₁₂₁]	
		$-y_{s2}$	-1	x_{s2}				0	0	0		-		b_{122}	
	$-x_{s3}$	$-y_{s3}$	-1	x_{s3}	y_{s3}	1		0	0	0		a_2		b_{123}	
	$-x_{s4}$	$-y_{s4}$	-1	x_{s4}	y_{s4}	1	• • •	0	0	0		b_2		b_{124}	
	÷	÷	÷	÷	÷	÷	۰.	0	0	0	.	c_2	=	:	(5)
	$-x_{sj}$	$-y_{sj}$	-1	0	0	0		x_{sj}	y_{sj}	1				b_{1nj}	
	1	0	0	0						0		:		a_1	
	0	1	0	0						0				b_1	
	0	0	1	0						0		a_n		c_1	
												b_n			
												$\lfloor c_n \rfloor$			

In the first 4 rows of matrix H, 4 stitch points are generated between surface 1 and surface 2. The matrix H is filled similarly for all other stitch points between other surfaces. The last 3 rows of matrix H of eq. 5 create a boundary condition. The tilt parameters and z-position of the fitted plane through surface 1 are set to a_1, b_1 and c_1 as a ground reference. From the solution matrix H and the known parameters \hat{b} the stitched tilt and altitude coefficients \hat{x} can now be determined for all surfaces by a least-means- squares estimate: $\hat{x} = (H^T H)^{-1} H^T \hat{b}$. After changing the best fitted planes of the surfaces to the calculated parameters of \hat{x} the stitched surface is obtained (figure 9).

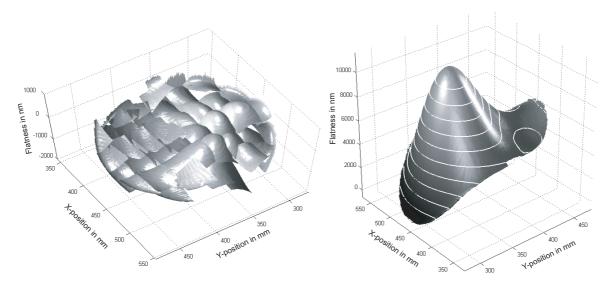


Figure 9. Left: Measured local flatness maps, Right: Stitched flatness map

4.2. Error Propagation

To study the effect of error propagation of the flatness stitching procedure due to measurement errors a set of measurements without error propagation and with error propagation are compared. For this a horizontal stroke of a relatively flat wafer is measured.

One set of measurements is measured without tilt alignment of the wafer (fig. 10a). Because the tilt alignment between neighbouring measurements is unchanged, the recorded phasemaps can be connected smoothly without using the stitching procedure. The flatness of figure 10b is obtained after unwrapping the phasemap of figure 10a. This measurement will be called the reference measurement.

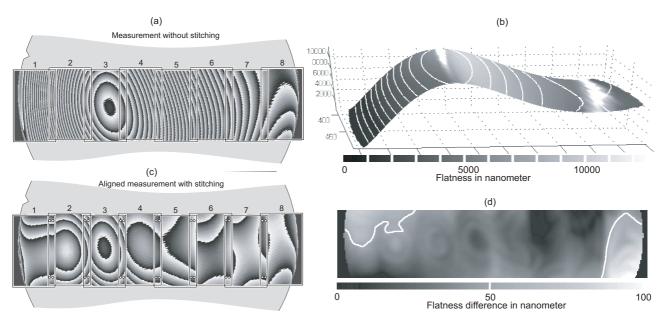


Figure 10. a: 8 phasemaps recorded without realignment, b: flatness obtained without stitching. c: Locally aligned measurements, d: Difference between flatnessmap b and stitched flatnessmap of measurement c.

For the other set of measurements the tilt of the wafer is adjusted for each measurement (fig. 10c). The wafer tilt is minimized by locally aligning the wafer for each measurement. Because the wafer alignment is different for neighbouring measurements these measurements have to be combined using the stitching algorithm as described in section 4.1. An overlap of 17% is used. When we compare both measurements the errors introduced by the stitching procedure appear (see figure 10d). A stitching error of approximately 100 nm appears for a wafer with a 10 μ m flatness deviation.

4.3. Lens calibration

For flatness measurement the accuracy of a surface interferometer is affected by aberrations of the optical system. To study the full aperture aberrations of the interferometer, measurements are performed outside of the scanning measurement setup. The effect of the aberrations can be made visible by comparing flatness maps that are recorded at a different tilt positions of a measurement object. When measuring the same surface with a positive and a negative horizontal tilt (figure 11a) the difference between both measurements after subtracting a least-means-squares surface should be zero. However, due to the field distortion of the optical system a residual appears. The same measurements can be preformed for a positive and negative vertical tilt (figure 11b).

To measure the field distortion accurately the measurements are performed at several different tilt positions. The displacement maps shown in figure 4.3 are obtained. The right part shows the barrel shape field distortion applied to a rectangular grid.

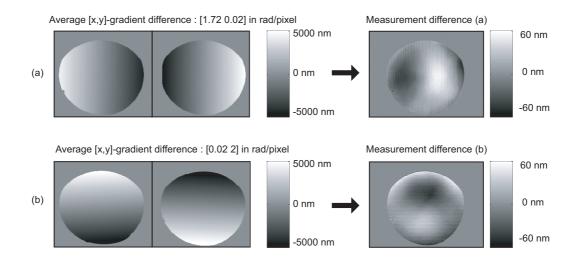


Figure 11. Measurement difference between horizontally tilted measurements (a) and vertically tilted measurements (b). Measured x-displacement map Measured y-displacement map Resulting field distortion (exaggerated)



Figure 12. Left: x-displacement map, Center: y-displacement map, Right: Field distortion applied to a grid.

For stitched measurements systematic measurement errors caused by spherical aberrations propagate quadratically. The influence of the field distortion on the measurement accuracy is higher for a tilted measurement than it is for an aligned measurement. The effectiveness of lens correction can therefore best be demonstrated for stitched measurements that are recorded using the spatial carrier fringe method.

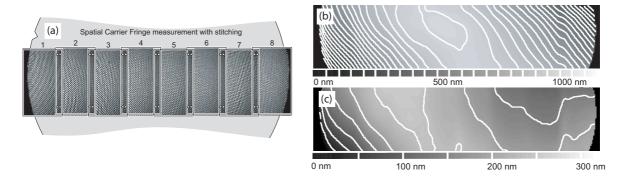


Figure 13. Left: eight carrier fringe measurements, Right: Difference with the reference measurement of figure 10b and stitched carrier fringe measurements without (b) and with (c) lens correction.

The difference between the flatness measurement of figure 13a and the reference measurement of figure 10b is a result of propagation of measurement errors that are introduced by the lens aberrations. By correcting for the field distortion the error propagation of the stitched spatial carrier fringe measurements is reduced from 1200 nm (figure 13b) to 300 nm (figure 13c).

5. THICKNESS MEASUREMENT

5.1. Reference flat calibration

To determine the thickness variation of a wafer the distance variation between the two reference flats is measured (term C of eq. 1). For this the wafer is positioned outside the viewing area of the interferometer. Depending on which part of the beam from the interferometer is covered interferograms will appear as shown in figure 14. When the left side of the interferometer is covered the distance between the reference flats is measured from the

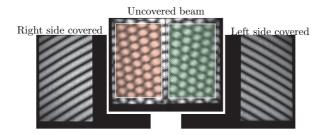


Figure 14. Reference flat measurement (left or right side covered), pre-alignment measurement (uncovered)

front side. When the right side of the interferometer is covered the same distance is measured from the back side. Therefore the left and the right images of figure 14 are each other's mirror. While making a horizontal scan multiple interferograms are recorded. A thickness reference map can be calculated from the recorded interferograms by using a spatial carrier fringe algorithm. (figure 15). This scanning calibration measurement of distance C can be done periodically to eliminate the effect of thermal or mechanical drift. When no wafer is present and the interferometer is not covered a mixture of interferograms appears (figure 14). The symmetry of these mixed interferograms is observed to make a pre-alignment of the front side flatness map on the back side flatness map.

5.2. Optical system calibration

1

To determine the effect of alignment errors, lens aberrations, scaling errors and rotation errors the wafer is measured at varying horizontal and vertical tilt positions (left of figure 16). Principally the derived wafer thickness should be independent of tilt or bow of the wafer surface. Due to the mentioned error sources however the true wafer thickness is not measured correctly. Figure 16 shows the relation between the measured thickness and the phase gradient of the local flatness map ϕ_f . It is assumed that the true thickness is measured when there is no tilt of the wafer $(\frac{\partial \phi_f}{\partial x} = 0 \text{ and } \frac{\partial \phi_f}{\partial y} = 0)$. Then the relationship is described well by the linear equation 6:

measured thickness = true thickness +
$$a\frac{\partial\phi_f}{\partial x} + b\frac{\partial\phi_f}{\partial y}$$
 (6)

The calibration parameters a and b of equation 6 can be estimated by fitting a plane through the data of figure 16. A curve is fitted for each pixel coordinate independently. When observing the calibration parameters for all CCD pixel coordinates the contribution of the alignment errors, lens aberrations, scaling errors and rotation errors can

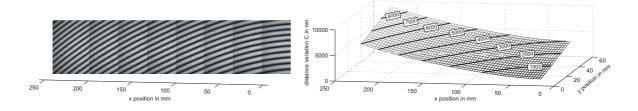


Figure 15. Left: Connected interferograms, Right: Calculated thickness reference values

Flatness phase maps ϕ_f for different wafer tilt

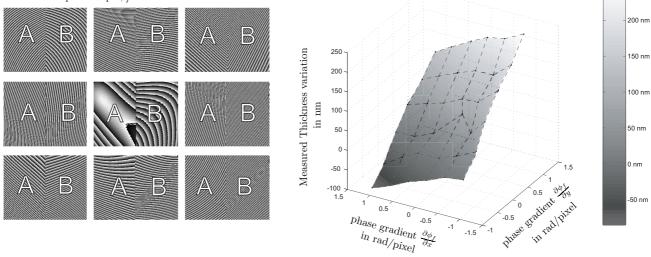


Figure 16. Left: Recorded phasemaps at different wafer tilt positions, Right: Measured wafer thickness for varying gradients of the front side flatness phasemap ϕ_f at a specific CCD pixel coordinate.

be separated. Using the calibration parameters and the measured local phase gradient it is now possible calculate the true thickness variation of a wafer independently of wafer tilt or bow. Figure 17 shows the construction of

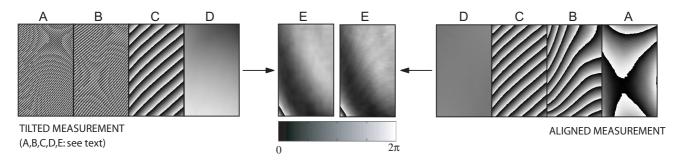


Figure 17. Left: Tilted measurement (carrier fringe method), Right: Aligned measurement (phase shifted method)

a thickness phasemap from an aligned measurement and from a strongly tilted measurement. The measured wafer thickness variation is obtained from both flatness phasemaps (A and B). After applying correction for the thickness reference (C) and a correction (D) for the mentioned error sources the true thickness (E) is obtained. Both calculated thickness phasemaps show a good resemblance. A global thickness phasemap of the entire wafer is constructed by combining all locally measured thickness phasemaps. The local thickness phasemaps connect smoothly without stitching. After unwrapping the global phasemap the thickness variation of the entire wafer is obtained (figure 18). Note that in contrast to the stitched flatness measurements fundamentally no error propagation occurs for the thickness measurement.

6. CONCLUSION

A new wafer geometry measuring concept is presented. By using proper measurement principles a robust, traceable and self-calibrating wafer flatness and thickness measurement instrument is created which combines

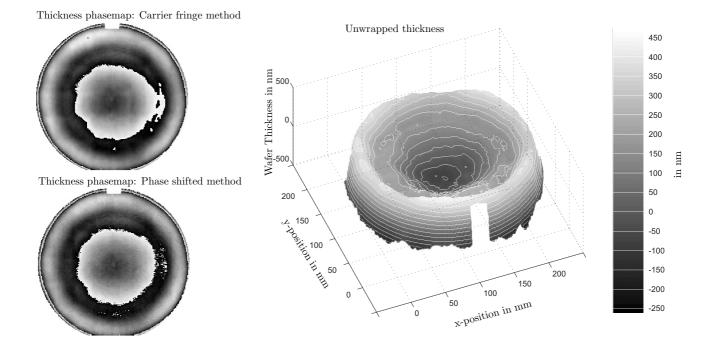


Figure 18. Left: wrapped wafer thickness phasemap acquired with the carrier fringe method and with the phase shifted method, Right: unwrapped wafer thickness

high lateral resolution, high measurement speed and low cost price. For wafer flatness measurements the main error source is introduced by the stitching procedure. For aligned measurements the propagation error of stitched measurements is around 100 nm. Further study of wafers of varying quality is required for proper quantization of measurement uncertainties of the stitching procedure. The removal of systematic lens errors has shown to reduce the error propagation. For future versions an attempt can be made to further reduce the effect of error propagation by integrating a reference object in the wafer holder around the wafer. The thickness measurement is fundamentally unaffected by error propagation. The measured thickness accuracy can therefore be assumed to be as good as that of a full size aperture interferometer. However, using a scanning setup principally a higher lateral resolution can be obtained, which can be in favour of the alignment accuracy. To get an idea of the accuracy of the current demonstration setup: a reproduceblity of approximately 40 nm has been a measured. This means a comparison between a thickness measurement before and after reassembly and recalibration of the optical setup. The wafer was measured at a different rotational orientation.

ACKNOWLEDGMENTS

This work has been financially support by EC Project Nr. GRD1-2000-25292, Acronym: "SuperWafer". Metrology workpackage in cooperation with IBS Precision Engineering.

REFERENCES

- 1. J.Kato and I.Yamaguchi, "Real-time fringe analysis based on electronic moire and its applications," Proc. of the 2nd International Workshop on Automatic Processing of Fringe Patterns 19, pp. 66–71, October 1993.
- 2. W. Osten, Digitale Verarbeitung und Auswertung von Interferenzbildern, Akademie Verlag GmbH, Berlin, 1991.
- 3. Z. D.Malacara, M.Servin, Interferogram analysis for optical testing, Marcel Dekker, Inc., New York, 1998.