Optosurf wafer roughness measurements

Carlos Reyes

2023-06-13

Table of contents

# 1. Wafer roughness measurement

In this document a new methodology to characterize wafer roughness is presented. This methodology uses a scattered light sensor that enables optical, contactless measurement of different types of surfaces (Optosurf)

# 2. Optosurf 32 pixel line detector

The Optosurf head is able to detect scattered light in order to measure roughness. The roughness of a sample can be determined from the shape of the optical field that goes into the detector, for rougher samples the optical field will be wider. The optosurf is also able to determine the lateral shift of the optical field, this is equivalent to the incoming angle of the sample’s reflected light.

In order to sample the optical field, the optosurf head has a 32-pixel linear detector within +-15 deg on-axis. The optical field is sampled by performing a window integration over each pixel, obtaining 32 sampling points. The measured parameter to characterize the sample roughness is called Aq and is calculated by reconstructing a histogram of the sampling points. The sampling process of the optical field and Aq value calculation is illustrated in [Figure 2.1](#fig-2-1).

|  |
| --- |
| Figure 2.1: a. Optical field definition. b. Window integration. c. Histogram |

## 2.1 Optical field and window integration simulation

The optosurf signal is reconstructed from the 32 sampling points. Such sampling points characterize the roughness of any given sample through the parameters and . The parameter relates to the incoming angle of the sample’s reflected light, this is equivalent to a lateral shift of the optical field. The parameter relates roughness of the sample and determines the width of the optical field.

This is shown in [Figure 2.2](#fig-2-2). When a wafer is rotated the optosurf signal is displaced along the linear detector, while increasing the sample roughness widens the optosurf signal through the parameter.

|  |
| --- |
| Figure 2.2: a. Effect of changing the angle of the incoming light. b. Effect of increasing sample roughness |

In order to simulate this, the equation of an optical field is defined as:

Then, different parameters for and are used to simulate different optical fields over a range of 30 degrees from -15 to 15, as shown in [Figure 2.3](#fig-2-3-histogram) (a). Notice when changing from -4.0 to 4.0 the optical field is displaced along the linear detector. When changing from 1.0 to 2.5 the optical field is widened.

In order to simulate the linear array detector, a window integration was performed over the optical field representing each pixel. This is represented by the green rectangles in [Figure 2.3](#fig-2-3-histogram). Notice how for optical field with narrower peaks there are fewer sampling points (green points), while for wider peaks there are more sampling points. This directly relates to the Aq parameter that is calculated from the histogram of the sampling points as shown in [Figure 2.3](#fig-2-3-histogram) b.

For the case of very smooth wafer the type of signal that is expected is a very narrow signal (small ), hence the Aq parameter will be limited to only a few sampling points along the main peak and across the tails of the signal.

|  |  |  |
| --- | --- | --- |
| |  | | --- | | (a) Optical field window sampling |  |  | | --- | | (b) Aq histogram reconstruction |   Figure 2.3: Window sampling and Aq histogram reconstruction for different optical fields. |

## 2.2 Aq parameter limitations

From the simulation of the linear array, it was learnt that the Aq parameter depends on the width of the optical field. In the case of a smooth wafer, the optical field is very narrow and the Aq parameter is limited to only a few sampling points. Additionaly, the Aq parameter is dependent on the incoming angle of the sample’s reflected light, hence being angle sensitive. This is illustrated in [Figure 2.4](#fig-2-4):

|  |
| --- |
| Figure 2.4: Aq values as function of mu and standard deviation |

From the previous figure the Aq has the following limitations:

* The variation of Aq parameter is small for wider optical fields (larger ).
* In contrast, the variation of Aq parameter increases for narrower optical fields (smaller ).
* The Aq parameter is angle sensitive, specially for optical fields with narrower peaks (smaller ).

Because of this limitations the Aq parameter would be able to distinguish between samples with very different roughness, but it would not be able to distinguish between samples with similar roughness, specially with a nm roughness variation.

# 3. Base function

The sampled optical field on the line detector was previously simulated, the observations were:

1. There are 32 sampling points as the line detector has the same number of pixels. This was simulated using a window integration over a defined optical field defined by: .
2. The parameters and determine the lateral shift and width of the gaussian function respectively. These parameters are related to the incoming light angle from the sample and the material roughness.
3. The parameter was swept to simulate different incoming light angles. With this data, a histogram was then reconstructed and its standard deviation value was calculated.
4. An oscillating pattern was observed in the histogram’s standard deviation (Aq parameter). This is not ideal as this indicated a roughness change, however only the angle of the incoming light was being swept.
5. Hence, a new parameter that is constant as a function of has to be found.

In order to obtain such new parameter a base function is proposed as shown in [Figure 3.1](#fig-3-1) :

|  |
| --- |
| Figure 3.1: a. Experimental data acquired at different incoming angles. b. Base function reconstruction |

The idea of the base function consists in reconstructing a reference function obtained from experimental data of a very smooth wafer (0.1 nm in roughness) at different angles in order to obtain a function that is constant as a function of . This function will then be used as a reference parameter to compare with the experimental data of rougher wafers (1 nm). The reconstruction of the base function is illustrated in the previous image.

## 3.1 Base function simulation

The mathematical procedure for the base function reconstruction consists in sampling the optical fields at different incoming angles and the reshifting such points to the 0 deg reference, obtaining a higher density of sampling points instead of the limited 32 sampling points. This is shown in [Figure 3.2](#fig-3-2) and [Figure 3.3](#fig-3-3) (higher density of sampling points).

|  |
| --- |
| Figure 3.2: Reconstruction of base function from shifted optical fields |

|  |
| --- |
| Figure 3.3: Base function with higher density of sampling points |

## 3.2 Experimental base function

The base function is meant to be a reference function in order to characterize wafer roughness. For such purpose, a very smooth wafer (0.1 nm in roughness) was measured at different angles. The base function was then reconstructed using the same procedure as the simulated base function. The experimental data is shown in [Figure 3.4](#fig-3-4). The collected angles were from -1.0 to 1.0 degrees.

|  |
| --- |
| Figure 3.4: Smooth wafer experimental data for base function reconstruction |

The procedure from the simulated base function was then applied to the experimental data. The results are shown in [Figure 3.5](#fig-3-5). Notice that there are some overlaps in the sampling points, this is due to innacurate angle measurements. However, this base function can be use as a reference for further data correction.

One of the main difference between using just the Aq parameter compared to the base function is the number of sampling for very smooth wafers. The base function has a higher density of points in the tails as well as in the main peak.

|  |
| --- |
| Figure 3.5: Reconstructed base function from experimental data |

## 3.3 On-axis and off-axis datasets

One of the limitations of the previous data, was that the exact incoming angle was not known, hence there was overlap in the sampling points. To compensate for this, a setup was built and shown in [Figure 3.6](#fig-3-6) :

|  |
| --- |
| Figure 3.6: OS500 head is attached to off-axis motor, and points at centre of rotation. Sample is attached to on-axis motor, and positioned at centre of rotation. Additionally there are two linear stages for positioning. All motors are PC controlled |

With this setup it is possible to acquire datasets along the optosurf axis (on-axis) at different vertical off-axis misalignment positions. Both axes are illustrated in [Figure 3.7](#fig-3-7) :

|  |
| --- |
| Figure 3.7: On-axis datasets at off-axis positions. Data is collected along the on-axis position with scanning angles of 20 deg. The same dataset is collected at different off-axis positions for example at +- 0.15 deg. Notice that the amplitude decreases for off-axis positions. |

One of the reasons for acquiring the dataset in both on-axis and off-axis positions is due to the very narrow peak obtained from the scattered light of the Si wafer. The idea is to acquire a dataset along the on-axis at 4000 angle positions (each position has a 32 points data trace) within +-20 deg, and at different off-axis positions for example between +- 0.15 deg. With these datasets it will be posible to reconstruct a base function at different off-axis positions and obtain the angles at which the new RMSE parameter will be valid for.

An example of a on-axis dataset is shown in [Figure 3.8](#fig-3-8). Different traces are also shown for different positions along the mechanical axis. It is with the on-axis dataset that the base function will be reconstructed.

|  |
| --- |
| Figure 3.8: On-axis dataset. The gray planes are cross-sections that represent that 32 sampling points optosurf traces |

The off-axis data is taken along the perpendicular axis of the linear array. When the head is aligned at the center the intensity of the optical field will be maximum. As the mechanical axis scans, the amplitude will start decreasing due to the ‘off-axis’ alignment. The dataset is shown in [Figure 3.9](#fig-3-9).

|  |
| --- |
| Figure 3.9: On-axis dataset. The gray planes are cross-sections that represent that 32 sampling points optosurf traces |

# 4. Minimization methods

The setup shown in [Figure 3.6](#fig-3-6) was used to obtain on-axis data to reconstruct base functions at different off-axis position. The methodology for this is described in [Figure 4.1](#fig-4-1) :

|  |
| --- |
| Figure 4.1: Methodology to acquire a base function |

1. **Starting base function:** The reconstructed base function from [Figure 3.5](#fig-3-5) was used as a starting function even if the sampling points were not perfectly aligned with their corresponding angles.
2. **Starting optosurf axis:** These are the 32 pixels that go from -15 to 15 deg with 1 deg steps.
3. **On-axis dataset:** An on-axis dataset of very Si smooth wafers (0.1 nm roughness). The mechanical axis was acquired within an angle range with a certain numbers of slices, for example +-20 deg (equivalent to 40 deg due to reflection angle) with 4096 slices. Each slice has a 32 sampling points optosurf trace.

**Minimization function:**

1. **Iteration over the slices of the on-axis dataset:** Each slice of the on-axis dataset is iterated and is used as the reference experimental data to be fit.
2. **Applying cost/minimization function:** This function takes the experimental data and makes a fit using three parameters:

* x0 - Equivalent to a lateral displacement of the base function.
* A0 - A factor multiplying the amplitude of the base function.
* pchip interpolator - A pchip (Piecewise Cubic Hermite Interpolating Polynomial) interpolator is an algorithm used to smooth the base function to a 32 points xaxis (starting optosurf axis + x0).

1. **Output variables:** The output variables of the minimization function are:

* The mechanical angle (corresponding to the mechanical scanning)
* x0 - Equivalent to a lateral displacement of the base function. Notice that x0 = 0 is the reference to know where the light is normally incident to the optosurf head.
* A0 - A factor multiplying the amplitude of the base function.
* RMSE - The optimization error measurement

**Displacement correction**

1. **Find x0 = 0 (normal incident plane)**
2. **Find the mechanical angle corresponding to x0 = 0:** This defines a displacement correction angle for the mechanical axis such that zero degrees correspond to the x0 of the optosurf axis.

## 4.1 Datasets

The acquired on-axis and off-axis datasets are shown in [Figure 4.2](#fig-4-2) :

|  |
| --- |
| Figure 4.2: On-axis datasets. Notice that as the roughness increases the amplitude decreases as well as becoming broader |

These datasets were acquired along the on-axis within an angle of 20 deg (40 deg due to double reflection angle) and at 6 different off-axis positions from -0.150 to 0.150 deg in 0.05 deg steps. Notice how for much smoother wafers (0.1 nm) the signal amplitude is much higher (around 35,000 in amplitude). As the sample becomes rougher the amplitude decreases and the signal becomes wider as well. This is illustrated in the Ann 1 min - Ann 5 min. The time represents the polish time, as the sample becomes smoother with higher polishing time the amplitude increases.

The parameters for each dataset are shown in [Table 4.1](#tbl-1). The datasets with a roughness of 0.1 nm are used as the reference to reconstruct the base function. The RMSE is then calculated with respect to these base functions for rougher wafers.

Table 4.1: On-axis and off-axis datasets parameters

|  | roughness (nm) | points | angle |
| --- | --- | --- | --- |
| wafer |  |  |  |
| W1\_0.11nm | 0.11 | 4096 | 20 |
| W2\_0.13nm | 0.13 | 4096 | 20 |
| W3\_0.1 nm | 0.10 | 4096 | 20 |
| W4\_1nm | 1.00 | 4096 | 20 |
| W5\_0.97nm | 0.97 | 4096 | 20 |
| W6 | NaN | 4096 | 20 |
| W7\_315nm | 315.00 | 4096 | 20 |
| W8\_5min | NaN | 4096 | 20 |
| W8\_4min | NaN | 4096 | 20 |
| W8\_3min | NaN | 4096 | 20 |
| W8\_2min | NaN | 4096 | 20 |
| W8\_1min | NaN | 4096 | 20 |

## 4.2 Minimization 1st iteration

Once the datasets have been acquired the 1st step is to apply the minimization function (steps 4 and 5) of [Figure 4.1](#fig-4-1) for each of the 4096 slices of the dataset. The output of the minimization function is the mechanical angle, x0, A0 and RMSE. The normal incidence is defined at x0 = 0, with this a displacement correction factor is calculated for the mechanical axis (step 7 and 8 of [Figure 4.1](#fig-4-1)).

The output of the minimization function for the wafers used a base function (0.1 nm roughness) is shown in [Figure 4.3](#fig-4-3) :

|  |
| --- |
| Figure 4.3: 1st optimization iteration for 0.1 nm Si wafers |

The meaning of the plots is:

1. x0 vs mechanical angle. The mechanical angle is the known scanning angle position from the scanning along the on-axis. x0 is the fit parameter of the minimization function. Notice that x0 = 0 is the reference to know where the light is normally incident to the optosurf head. If they both cross at 0,0 then the mechanical axis is aligned with the optosurf axis. If they don’t cross at 0,0 then there is a displacement between the mechanical axis and the optosurf axis. This displacement is corrected by applying a correction factor to the mechanical axis.
2. difference vs mechanical angle (mechanical angle-x0). This is the plot of difference between the mechanical angle and the x0. Ideally, after more optimization iterations the difference will converge to zero. With this difference a new optosurf axis is calculated.
3. RMSE vs mechanical angle. This plot is the RMSE per slice calculated from the minimization function. Ideally, after more optimization iterations the RMSE will converge to zero.
4. amplitude vs mechanical angle. This is the amplitude correction factor calculated from the minimization function.

The optosurf shifted axis for the 1st iteration is shown in [Figure 4.4](#fig-4-4) :

|  |
| --- |
| Figure 4.4: Optosurf shifted axis |

## 4.3 Base function 1st iteration

Once the minimization parameters have been calculated (x0, A0) a new shifted axis was calculated. The process to reconstruct the base function is to take the slices from the on-axis dataset that cover a one degree range. The data is then plot with respect to the new shifted axis and recentered with respect to the known mechanical angle.

With this a highest density of sampling point is obtained for the base function. An average is then calculated and finally an interpolation is done. Examples of base functions are shown in [Figure 4.5](#fig-4-5) :

|  |
| --- |
| Figure 4.5: Base functions for 0.1 nm Si wafers |

## 4.4 RMSE optimization

Once the 1st iteration has been completed there are several parameters that have been calculated:

1. x0 - equivalent to the correction factor for the mechanical axis.
2. A0 - equivalent to the amplitude correction factor.
3. RMSE - the error of the minimization function.
4. Base function - the reconstructed base function.

All of these variables are taken and the procedure shown in [Figure 4.1](#fig-4-1) is re-iterated until the x0 and the mechanical axis have a zero difference between each other. This will reduce the RMSE per iteration. The final RMSE which is to be used as the error base function for roughness characterization is shown in [Figure 4.6](#fig-4-6) :

|  |
| --- |
| Figure 4.6: Optimized RMSE for 0.1 nm Si wafers |

The final result for RMSE shown in plot b represents the reference RMSE calculated for very smooth wafers (0.1 nm). This error will be used to compare with rougher samples. Notice that the RMSE increases for angles greater than +-0.45 deg, so that RMSE parameter is valid only within this range.

# 5. RMSE-roughness characterization

The final step after having reconstructed base functions from smooth samples (0.1 nm) and calculating a reference RMSE is to compare rougher samples with respect to this. The roughness parameters are shown in [Table 5.1](#tbl-2). All the datasets were acquired along the on-axis at +-20 deg and at 7 off-axis positions from -0.15 to 0.15 deg with 0.05 deg steps.

Table 5.1: Roughness parameters

|  | roughness (nm) | points | angle |
| --- | --- | --- | --- |
| wafer |  |  |  |
| W1\_0.11nm | 0.11 | 4096 | 20 |
| W2\_0.13nm | 0.13 | 4096 | 20 |
| W3\_0.1 nm | 0.10 | 4096 | 20 |
| W4\_1nm | 1.00 | 4096 | 20 |
| W5\_0.97nm | 0.97 | 4096 | 20 |
| W6 | NaN | 4096 | 20 |
| W7\_315nm | 315.00 | 4096 | 20 |
| W8\_5min | NaN | 4096 | 20 |
| W8\_4min | NaN | 4096 | 20 |
| W8\_3min | NaN | 4096 | 20 |
| W8\_2min | NaN | 4096 | 20 |
| W8\_1min | NaN | 4096 | 20 |

The roughness was measured using an atomic force microscope (afm) as shown in [Figure 5.1](#fig-afm)

|  |
| --- |
| Figure 5.1: AFM measurements |

## 5.1 RMSE-roughness bands

[Figure 5.2](#fig-5-1) shows RMSE-roughness characterization using Rockley\_ref (0.1 nm) as a reference base function. Notice that there are several bands representing different roughness levels. [Figure 5.2](#fig-5-1) a shows three band levels for the different roughness values including 0.1 nm, 1 nm and . It is possible to differentiate between 0.1 nm and 1 nm levels which is a limitation when using only the Aq values. The 0.1 nm band is between 0 and 15 RMSE, while the 1 nm is 35 and 50 RMSE.

[Figure 5.3](#fig-5-2) b shows the RMSE normalized with respect to the amplitude correction factor and using a log scale. In this case the band spacing is wider due to the log scale.

|  |
| --- |
| Figure 5.2: Roughness RMSE bands |

A comparison of the same roughness values but using the Aq values is shown in [Figure 5.3](#fig-5-2). Notice the Aq values have a cosine shape and the 0.1 nm and 1 nm bands are not clearly separated. This is because for very smooth wafer the sampling points are limited over the main curve, hence the Aq values are not very sensitive to roughness changes in the nm scale. When focusing only in +-0.1 deg the Aq values seem to have a band behaviour, however because of the very small Aq values the resolution is too small to clearly distinguish between the different roughness levels.

|  |
| --- |
| Figure 5.3: Aq values |

## 5.2 RMSE-roughness wafer polishing

[Figure 5.4](#fig-5-3) shows the relationship between roughness and polishing time of a wafer. The more polishing time the smoother the wafer becomes as expected. A time longer than 5 min is required to obtain a roughness below 1 nm. With this propose methodology it is possible to characterize different wafers after polishing and compare them with respect to a reference wafer.

|  |
| --- |
| Figure 5.4: Roughness RMSE bands |

## 5.3 Repeatability and data conditions.

In order to know the requirements of the off-axis aligment different off-axis datasets as the one shown in [Figure 3.9](#fig-3-9) were acquired. The RMSE was then calculated using the same method. The results are:

|  |
| --- |
| Figure 5.5: Off-axis angle limitations |

In this case the results imply that the data needs to be collected at off-axis positions withing +-0.05 deg in order to obtain reproducible results. Notice outside this range the error increases significantly. Hence, in practice there has to be a precise control over the scanning angles using a vacuum in a clean room to make sure the data is collected within the required scanning on-axis and off-axis angles. This limitation is illustrated in [Figure 5.6](#fig-5-5) , in this case the Ann10 wafer at 0 deg was used as the reference base function. Then the same wafer was fit and the RMSE was calculated but at different off-axis positions. Notice how the RMSE results are not consistent with the reference base function.

|  |
| --- |
| Figure 5.6: Repeatibility for 0.1 nm roughness at different off-axis positions |

A similar situation is illustrated in [Figure 5.7](#fig-5-6) , in this case the Ann10a was used as a reference base function. Notice how the orange curve is outside the 0.1 nm band. This is due to scratches/dust in the wafer that can affect the roughness characterization. Hence, it is important to have a clean wafer under control enviroments that can be met in a controlled clean room.

|  |
| --- |
| Figure 5.7: Repeatibility of RMSE roughness in the same wafer |

Finally when using a 0.1 nm wafer as a refence function at 0 deg off-axis and fitting Ann 5 (1 nm) at different off-axis positions the result is not repeatable. For different off-axis positions the RMSE is not consistent and outside the expected 1 nm band. This is illustrated in [Figure 5.8](#fig-5-7).

|  |
| --- |
| Figure 5.8: Repeatibility for 0.1 nm roughness at different off-axis positions |

# 6. Conclusions

1. A new methodology has been presented in order to characterize wafer with roughness values between 0.1 to 1 nm.
2. This was accomplished by extending the capabilities through an optical setup and software in order to extend the measuring capabilities of the optosurf head.
3. In order to obtain reproducible result, the measurements need to be taken in a controlled way. Avoiding wafer scratches as well as being careful with vibrations.

# 7. Introduction

This is a book created from markdown and executable code.

See Knuth ([1984](#ref-knuth84)) for additional discussion of literate programming.

[1] 2

See author1, author2, and author3 ([2016](#ref-test))

s

# References

author1, author2, and author3. 2016. *Deep Learning*. MIT Press.

Knuth, Donald E. 1984. “Literate Programming.” *Comput. J.* 27 (2): 97–111. <https://doi.org/10.1093/comjnl/27.2.97>.