A Study of SiCN Wafer-to-Wafer Bonding and Impact of Wafer Warpage

Abstract—Wafer to wafer bonding studies were carried out on blanket and patterned highly warped wafers using SiCN as bonding dielectric material to gain a deeper understanding of SiCN-to-SiCN direct bonding and to assess the impact of wafer shape on bonding overlay results. Regarding the fundamental understanding of SiCN bonding, characterization data show that the SiCN-to-SiCN interface is oxidized, suggesting strongly bonded due to the presence of covalent bonds, already at room temperature, in contrast to what has been hypothesized for SiO2 for which this reaction would start only by subjecting the bonded pair to an anneal of 150 °C. As concerns the impact of shapes on final bonding results, bonding experiments are carried out by combining different wafer shapes. Tools such as patterned wafer geometry (PWG) and a lithography scanner were used to measure the distortion signature of the wafers before and after bonding. Bonding recipe parameters were optimized to minimize overlay errors, on two different bonding tool configurations, for nominally flat wafers. The same parameters were used to bond the warped wafers to investigate the impact of wafer warpage. By using one of the two tool’s configurations, overlay results can be significantly reduced for flat wafers. When wafers with different shapes are bonded, recipes must be optimized to obtain tighter overlay specifications. Such optimization is needed to counteract the effect of the incoming distortion caused by the different incoming shapes and measured by the scanner. However, it should be noted that the incoming distortion is not solely determined by the shapes, but also by the kind of dielectric stack used, meaning that the same shape with a different stack will result in a different incoming distortion, as seen in the two different test vehicles used in this study.

Tendency of the bond wave to propagate according to the Si-orientation is visible in the distortion signatures after bonding.

Keywords—direct bonding, SiCN, saddle shape, high bow, 3D integration, wafer distortion, 3D NAND.

I. INTRODUCTION

Fusion wafer bonding is drawing a lot of interest as a key fabrication technique to realize 3D integration circuits and since its introduction, it is a technology booster in the manufacturing processes used to enable different applications as image sensors [1], microLED displays [2] or high-performance computing [3].

Historically, SiO2 is being used as bonding dielectric material but there is interest in using other dielectrics tailored around the different applications. For instance, Imec introduced PECVD SiCN for the implementation of 3D integration schemes and in this case SiCN would be more appropriate for its intrinsic barrier property against Cu diffusion [4]. The layer has been proved to be an excellent bonding dielectric demonstrating low surface roughness post CMP enabling high chemical bond strength [5]. A fundamental understanding of the reasons behind these performances is needed to eventually tune the properties of other dielectrics which could be used for different applications.
For a new fast-growing application like high-performance computing wherein 3D NAND technology [6] is used, bonding will have to be realized in non-standard conditions due to the incoming mechanical stress build up on the wafers during fabrication. Such stress will confer to the wafers a typical saddle shape in which the convex and concave shapes are simultaneously present on two different planes. Different shape mitigation techniques have been proposed but often those do not result in a perfectly flat shape.

In the first part of this paper, we will show some characterization results obtained by inspecting the interface of the SiCN/SiCN by means of different techniques (ERD, TEM, XPS, EELS and HAXPES) at different post bond annealing temperatures. In particular, we observe that an oxidation of the XPS, EELS and HAXPES) at different post bond annealing the SiCN/SiCN by means of different techniques (ERD, TEM, characterization results obtained by inspecting the interface of the bonding mechanism of SiO2, such oxidation would start only at 150 °C [7]. Moreover, we observe that by increasing the post bond annealing temperature the amount of oxygen at the interface is increasing, probably due to the reaction of the water created as a byproduct of the bonding process. Such results provide a better understanding of the mechanism behind SiCN-to-SiCN bonding.

In the second part of the paper, we show an initial study performed in an attempt to understand the effect of the wafer shape together with the bonding configuration on the final bonding overlay results. It needs to be highlighted that in this study optimization of the bonding recipe was performed only for reference wafers with low warpage and non-optimized results for wafers with high warpage will be shown as well.

The study was possible thanks to the creation of short loop test material which mimics the effect of mechanical stress observed during manufacturing of memory or logic wafers and at the same time optimized for the bonding characterization. In some cases, the creation of the test material was supported by a simulation study which led to the identification of the correct pattern to induce the complex wafer shapes typically observed in production. Tools such as patterned wafer geometry (PWG) and a lithography scanner were used to measure the distortion signature of the wafers before and after bonding.

Interesting considerations on the distortion induced on these wafers by changing the shapes could be drawn. Such considerations should be taken into account in the development of overlay prediction models based on shape [8].

From the bonding results we observe that the incoming shape will have a big impact on the final distortion results and a fine tuning of the recipes will be needed in order to counteract such effect. Interaction of dynamic events due to the interaction of the shape with the bond wave propagation are visible in the bonding results. However, a more accurate characterization will have to be done in the future to understand in detail the impact of the wafer shapes.

II. DESCRIPTION OF THE TEST MATERIAL

A. Sample preparation TEM, ERD and XPS

To assess dielectric bonding on 300mm Si wafers, a densified 120 nm SiCN layer was deposited on top of a 300nm thick SiO2 layer. Densification was done in N2 environment for 2 hours at 250 °C. After deposition and densification wafers went through a chemical mechanical polishing step during which ~20 nm of material was removed and a roughness of ~0.12 nm was achieved.

The bonding was performed in an EVG GEMINI® FB XT tool, and a process of reference activation sequence, featuring a N2 plasma step and a DI water rinse step, was used prior to bonding. The entire activation sequence was performed inside the bonding cluster.

After bonding, four different splits were created depending on the post bond annealing conditions: one wafer pair was not subjected at all to an annealing step, one was annealed at 110 °C, one at 250 °C and the last one at 350 °C. All the wafer pairs went through the same thinning step. Initially a grinding step was used to reduce the top wafer thickness to 50 μm Si. By using a dry etching process the Si was completely removed. Thanks to a HF based wet etch process the SiO2 layer was completely removed exposing the SiCN. Top SiCN was further reduced to 50 nm remaining SiCN thickness. Afterwards coupons were created by using a dicing step and submitted to different characterization techniques such as transmission electron microscopy (TEM) coupled with electron dispersive x-ray spectroscopy (EDS), electron recoil detection (ERD) and X-ray photoelectron spectroscopy (XPS).

For XPS analysis in order to further reduce the thickness of the top SiCN layer (~20 nm), a final thinning step was performed by using a SF6/O2 etch.

B. Sample preparation bowed and warped wafers

Shortloop test material has been created to characterize the impact of the wafer shape on the bonding signature. The basic test material has a single metal layer, used to define alignment and overlay markers, embedded in a dielectric stack, planarized and finished with a polished process of reference (POR) SiCN dielectric[9][10].

Two different test vehicles have been used for the experiments, A and B. The main difference between the two is the mask used and the type and thickness of the isolation layer used between the Si and the metal module (see Fig. 1). The two masks are characterized by a different die dimension and metal density. The A mask is characterized by a die dimension of 27 mm * 14 mm, while the B test vehicle is characterized by an extremely dense metal grid full of overlay markers, with a square unit cell dimension of 7.2 mm. Such a mask was introduced in order to have many more overlay markers distributed over the wafer.
To create symmetric shapes an uniform SiN layer, characterized either by a tensile stress or a compressive stress, and different thicknesses, has been deposited on the back side of the wafers to generate a range of stresses and therefore different wafer shapes. Different levels of wafer bow ranging from -120 μm to 240 μm were fabricated by using the A mask while only two symmetric shapes -220 μm and 300 μm were fabricated by using the B test vehicle.

The creation of the asymmetric shape has been supported by theoretical simulations [6] which led to the selection of the correct materials, stack and pattern to use to obtain the desired shape (see Fig. 2a). The pattern selected was a regular stripe pattern in which compressive SiO2 (-120 MPa) and SiN layer (1.1 GPa) were alternated. According to simulations a critical factor to achieve saddle shapes is to limit the maximum stripe width to dimension of 1μm or below. To serve this purpose, a predefined mask, characterized by long vertical stripes running through the entire die with a lateral dimension of 0.8 μm at a distance of 1.2 μm, was selected. Such a stress pattern was placed under the metal module as shown in Fig. 2b. First, 280 nm of SiN layer was uniformly deposited on a SiO2 layer. Afterwards the layer was partly etched away according to the mask layout, then filled and planarized with SiO2. Considering that the lateral dimension of one of the two stripes was a bit larger than what proposed, in order to counteract such non-ideality in the design, it was decided to increase the SiO2 thickness in the final stack balancing in this way the slight excess of tensile stress present in the stack. A perfect saddle shape was obtained in the pre-bonding state as it is shown in Fig. 2c.

C. Bonding experiments warped and bowed wafers

In this study we evaluated the bonding results obtained by using two different chuck systems labeled here as configuration A and configuration B, both installed on an EVG GEMINI® FB bonding cluster featuring a SmartView ® NT2 aligner.

Twelve wafer-pairs by using A mask, characterized by different bow combinations, have been bonded by using a configuration A chuck system, while six wafer pairs, manufactured by using B mask, have been bonded by using a configuration B chuck system.

In configuration A, the wafers are held flat before the bonding and the top wafer is slowly released during the bonding process.

In configuration B the chuck system is engineered in such a way that the scaling component, described in the next section, can be controlled, and minimized down to almost 0 ppm.

III. RESULTS AND DISCUSSION

A. TEM, EDS, ERD and XPS results

TEM/EDS inspection results are presented in Fig. 3. A ~7 nm oxidized bonding interface can be recognized from the EDS analysis. The thickness and the chemical composition appear stable over the three post bond annealing temperatures.

The same samples used for TEM/EDS analysis have been characterized by ERD and also in this case the three interfaces appeared to be the same from a chemical viewpoint. In Fig. 4 the result of one of the three samples is presented.

From the ERD analysis the thickness of the oxidized layer is estimated to be ~ 4 nm. Such calculation is done considering the mass density of SiO2. The fact that with ERD we calculate 4 nm SiO2, whereas TEM estimates 6 nm interface, could hypothetically mean that SiO2 agglomerates are incorporated in a matrix/environment of pristine SiCN material. This would be in accordance with previous EELS analysis of the interface in
Fig. 4. ERD concentration depth profile for the sample which was not subjected to any post bond anneal temperature.

which an evolution of chemical bond properties across the bond interface [11][12] could be detected. From these measurements it was observed that the oxidized interface does not have a stable nature across the entire thickness. Only an inner one could be potentially ascribed to pure SiO$_2$.

For ERD analysis it needs to be considered that the maximum accuracy in terms of detection of atomic density is decreasing in depth due to energy straggling, a fundamental physical phenomenon in the analysis. It has to do with the statistical nature of an energetic ion losing energy through gentle collisions with the lattice. Besides, because of experimental limitations, the depth resolution will be limited to around 10 nm.

In order to improve accuracy, a combined XPS and HAXPES (Hard-XPS) analysis was used. For these experiments top SiCN was thinned down to 20 nm. Initially a depth profile was performed using AlK$_\alpha$ radiation.

The following method was applied for the correction:

1) For the reference SiCN bulk, the O/Si, C/Si and N/Si ratio were calculated.

2) For the measurements close to the interface, the quantification was first performed using the O, C and N intensities and decomposing the Si intensities in a “Si-O” and a “SiCN” component.

3) Using the intensity of the SiCN Si1s component and the ratio calculated in (1), the expected intensities of the C, N and O component were calculated.

4) From the differences between the expected intensities (3) and the measured one (2), the excess concentration can be calculated.

Doing so, no excess concentration of C and N were detected, only O was observed to be in excess. Looking at the “Excess O concentration” ratio to the “Si-O” concentration measured on the Si1s peak, we observe a small increase for increasing annealing temperature (Table 1).

![Fig. 5. Concentration-depth profile showing the appearance of the oxidized interface as a function of sputter time.](image)

The XPS depth profiles from the three measured samples appeared to be the same, no significant differences are observed. An example of one of the three measurements is shown in the Fig. 5.

After this step, on a second position, partial sputtering was performed in order to reduce further the top SiCN to get closer to the oxidized interface and measurements were performed with high energy XPS (Cr K$_\alpha$).

Figure 6 presents the Si1s spectrum from the bulk of the SiCN layer (black) together with the spectra recorded close to the oxidized interface for the three samples. The additional peak at higher BE energy corresponds to the SiO$_2$ layer. For the quantification, the Si, C, N and O spectra were recorded and the SiCN layer was taken as reference to correct the intensities from the three spectra close to the interface.

![Fig. 6. Si1s spectrum from the bulk of the SiCN layer (black) together with the spectra recorded close to the oxidized interface for the three samples.](image)

<p>| TABLE I. Excess O concentration ratio to the Si-O concentration measured on the Si1s peak |
|---------------------------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>O excess in Si</th>
<th>110 °C</th>
<th>250 °C</th>
<th>350 °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>2.3</td>
<td>2.6</td>
<td></td>
</tr>
</tbody>
</table>
B. Incoming wafer distortion for bowed and warped wafers

Incoming grid distortion was characterized for each wafer right before bonding by reading out the alignment markers on the scanner. The backside (BS) SiN films or the stressing pattern introduced in the stack will not only induce a different shape but also a substantial in-plane-distortion (IPD) when these warped wafers are chucked flat on the scanner. Such distortion is partially corrected by the scanner itself using a linear alignment model when one would normally expose the wafers. In our case we modeled the uncorrected results by using a 4-parameter model featuring translation in X (T_x) and Y (T_y), rotation (R) and scaling (S).

An example of a distortion map before the correction for a saddle shaped wafer can be seen in Fig. 7a in which the corresponding vector plot represents the deviation of the marks measured with respect to their expected positions. In Fig. 7b the same map after correction, by using only the S parameter (-2.83 ppm) is presented. As it appears from the example most of the errors can be attributed to the scaling component. Similar analysis was done for the remaining wafers used for these experiments.

Fig. 7. Incoming distortion before (a) and after (b) the correction of the scaling parameter which in this case was calculated to be -2.83 ppm.

It is found that, for a uniform SiN layer deposition, as in the case of A, an incoming scaling of the grid is induced, and a strong linear correlation can be observed with the bow of the wafers (see Fig. 8). In the case of the B test vehicle, we had a limited number of wafers with different shapes produced. However, for the symmetrically bowed wafers it is observed that the bow-incoming scaling correlation would not be the same as the one observed for the A test vehicle. Such difference is probably due to the different stack and dielectrics used in the two test vehicles conferring different mechanical properties.

This observation should be taken into account when wafer grid deformation prediction models are developed[8] based on the shape of the wafer. The “conversion factor” will depend on the state of the wafer and the processes used to create it and will need to be determined experimentally.

Interestingly, it is possible to observe that even if the nominal bow of the saddle shape wafer is low (~5 μm), the incoming scaling for this wafer is comparable to a wafer with a uniform bow of ~150 μm.

C. Bonding results according to chuck configurations

W2W overlay is measured after bonding by using an infrared tool included in the bonder cluster. The data consists of a set of measuring coordinates (x1, y1; x2, y2; etc.) and corresponding misalignment values (Δx1, Δy1; Δx2, Δy2; etc.). Often to represent the results the vector length V, defined as the square root of the sum of the squared Δ, is used. Also, in this case the data were modelled with 4 parameter model to compare with the results before bonding. The difference between the overlay measured and the model is indicated as non-linear components.

In order to purely take into account the contribution of bow, the bonding recipe has been optimized for a reference wafer pair per each test vehicle and configuration. These reference wafers were not subjected to any stress. The as-created bonding recipes were kept constant towards the bonding of the remaining wafers characterized by different shapes.

Bonding experiments, by combining different shapes, were defined, as mentioned, by using the A test vehicle and bonded in configuration A tool. It can be noticed that the overlay gets worse when different shapes are bonded to each other while it stays approximately in the same range when the same shapes are bonded. Also, in this case most of the errors can be attributed to the scaling. It has been observed that if we define the difference between the bow of the top wafer and the bow of the bottom wafer as ΔBow, then the scaling after bonding will be fairly correlated to this parameter. By increasing ΔBow the scaling S will increase as \[ S = -0.8 + 0.03 \Delta \text{Bow} \], as can be observed in Fig. 9, meaning that an incoming bow change of 60 μm will add ~1 ppm scaling error in the final scaling after bonding. 1 ppm would correspond to 150 nm scaling error at the edge of a 300 mm wafer.

Fig. 8. Correlation between wafer bow, measured by PWG system, and incoming scaling of the wafer grid, measured by the scanner for the different test vehicles.
Considering that in configuration A wafers are kept flat we can easily understand that during the bonding we cancel out most of the incoming scaling error when wafers with the same shape are combined. On the other hand, when we combine wafers with different shapes, we can either amplify or reduce the impact of the incoming scaling errors into the final scaling obtained after bonding.

Nevertheless, in this chuck configuration we keep the scaling error due to the bonding process itself which is due to a relative difference in bonding wave speed between the top and the bottom wafer. In particular, because of gravity the top wafer bond wafer will have a tendency to propagate faster. Such error can be minimized to almost 0 ppm by using a different chuck system introduced by EVG to easily correct for the scaling error in the bonding recipe[13]. This is possible by changing the relative propagation speed between the two wafers realized by adjusting the geometry of the bottom chuck (configuration B).

A comparison between what can be achieved by using configuration A versus configuration B is presented in Fig. 10. Each box plot represents the total overlay measured and the part of the overlay attributed to the non-linear components, expressed in vector length values, for the reference wafer pairs bonded in configuration A and configuration B, respectively.

It can be noticed that the vector length magnitude is significantly reduced in configuration B and most of the errors can be attributed to the non-linear components which represents 70 % of the total vector length (if we consider the average vector length). As previously mentioned, this has been possible thanks to a better control of the scaling parameter, as can be seen in the inset graph where Δx and Δy are represented: The distribution of the measured points for configuration B is more clustered compared to the one observed for configuration A.

Due to a limited number of test material the bow-scaling characteristic for configuration B could not be extracted.

D. Shape evolution through thinning steps after bonding

A graphical representation of the bonding experiments performed in configuration B tool are shown in Fig. 11 where PWG images before bonding, after bonding, grinding and extreme thinning are presented. It can be noticed that the bow after bonding for all the bonding combinations gets reduced to almost 0. Nevertheless, a certain and non-uniform shape is observed after bonding even when two symmetric shapes are combined.

Fig. 11. Representation of the bonding experiments performed in configuration B tool by using the PWG images. In the figure images are representative of the status of the wafer in the bonding chamber before bonding. Thinning impact on the shape is presented too.

According to theoretical assumptions, considering only the static contribution of the bonding, mechanics and forces involved, the bow of the wafers after bonding should be 0. The fact that we still have a certain shape is ascribed to the dynamic contribution of the bonding process itself, implying that these shapes could give important information of the bonding process.

By thinning down the top wafers, the shape of the bonded stacks will always resemble more that of the bottom wafers, due to the gradual release of the mechanical strength related to the top Si substrate. However, the shape will not become completely the same as the shape of the bottom wafer. This is mostly because after extreme thinning, we will still have some layers belonging to the top wafers in the final stack.

E. Scanner data after bonding

In Fig. 12 the overlay residual maps are shown for the different wafer pairs bonded in configuration B (except for the saddle to saddle combination due to process related errors). The residual error measured on pairs made by combining different shapes is four times larger than the residual error measured on the reference pair made of nominally flat wafers. These error
maps exhibit a 8-fold rotational symmetry (45 degrees) which develops around a center characterized by a smaller error.

Such an error pattern is mostly related to the dependency of the bond wave propagation to the different mechanical properties that vary with a 45 degree periodicity on a (100) Si crystal plane.

For the two configuration where a saddle shape top wafer is used, even if the shape of the bottom wafers is different, the maps look almost the same. When a different top wafer is used, one can observe that the pattern is slightly changed. However, a clear understanding on how the impact of the shape will translate into the final overlay results, was not established. In order to elucidate the effect of wafer shape on overlay further, new bonding experiments will have to be implemented with a focus on bonding a specific wafer shape to a flat wafer.

IV. CONCLUSIONS

We carried out an investigation on the nature of the bonding interface in the pre-annealed state by using different analysis techniques. Sample preparation was possible thanks to the high bond strength that is typical for the SiCN-to-SiCN interface. TEM, EDS and ERD characterization techniques did not reveal significant differences between the interface which was not subjected to any post bond anneal temperature and the ones which were subjected to anneal of 250 °C and 350 °C. Only a combined XPS and HAXPES analysis was able to detect a slightly different oxidation state of these three interfaces. In particular, the amount of oxygen at the interface is increasing as a function of post bond anneal temperature.

We further initialized a study on the impact of different shapes on the bonding overlay supported by simulation studies and involving PWG, IR overlay measurements and scanner experimental characterizations.

It has been observed that incoming shapes are drastically changing the incoming distortion of the wafers. This incoming distortion will not be solely determined by the shapes but also by the dielectric stack used for the experiments, meaning that the same shape with a different stack will result in a different incoming distortion.

Information on the incoming distortion will be needed to tune the bonding recipes. Depending on the chuck configuration, such incoming distortion will play a different role in the final overlay results.

The second chuck configuration (configuration B) analyzed in this study turned out to be really efficient in the correction of the scaling error which represent the highest contribution in the incoming wafer distortion. Non-linear components represent the majority of the error in the final overlay result. The tendency of the bondwave to propagate according to the Si-orientation will have to be counteracted in the bonding chamber.

Moreover, the study provided a better understanding of the complexity involved in achieving the original goal of analyzing how the wafer shape and bonding configuration impact the final bonding overlay results.

ACKNOWLEDGMENT

This work is carried within the frame of the imec 3D System Integration Industrial Affiliation Program and within a Joint Development Project between Imec and EVG.

The author would like to thank Sander Van Gompel and Juan Fernando Gomez Granados for the help in depositing the SiN layers, Johan De Backer for the support received in the litho steps, Patrick Verdonck and Nicolò Angelo Pinna for the help in preparing the samples for XPS analysis.

REFERENCES

10.1063/1.4794319.


