April 30, 2010
RT5313
Engineering Technology pages

Research Report

Three-Dimensional Die-to-Wafer Integration Technology for High Throughput and Yield

Katsuyuki Sakuma

IBM Research - Tokyo
IBM Japan, Ltd.
1623-14 Shimotsuruma, Yamato
Kanagawa 242-8502, Japan

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Three-Dimensional Die-to-Wafer Integration Technology for High Throughput and Yield

Katsuyuki Sakuma

IBM Research – Tokyo, 1623-14, Shimotsuruma, Yamato, Kanagawa, 2428502, Japan

E-mail: sakuma2@jp.ibm.com

1. Background

Semiconductor devises have continued to rapidly get smaller with higher integration following Moore's Law. As a result, performance including speed, functionality, and power consumption has improved while costs have also been reduced. However, as the size of these devises has decreased (such as 90nm and 65nm), which causes management issues such as increased investment and increased research and development costs, and has also caused technical issues such as increased transistor leakage current. Therefore, it has become more difficult to minimize the size at the same pace as the speed. It is believed that the development of the three dimensional System-in-Package (SiP), which combines multiple LSIs three dimensionally, is an effective solution to these issues. The reason for this is that a 3-D SiP is a high-density package and can be used to achieve better functions than a System-on-Chip (SoC) through optimal chip design. There are three types of this 3D integration process; the Chip-to-Chip process, Chip-to-Wafer process, and Wafer-to-Wafer process. In Japan, Chip-to-Chip processing is further advanced than the other two processes due to the package orientation. However, technology verification and new technology development in the Chip-to-Wafer and Wafer-to-Wafer processes have not advanced much due to certain issues (yield and costs) despite studies being conducted.

2. Summary of Invention

In this invention, multiple laminated chips are inserted inside of a cavity when the chips are stacked three dimensionally, and isotropic pressure is applied from the top of the chip through an elastic material using a gas or liquid as the pressure medium in order to perform pressure bonding and stack multiple layers of chips. It also includes chip stacking technology that equalizes the pressure during Chip-to-Wafer bonding in order to increase the yield of bump bonding.

1. Templates with cavities are aligned with the wafers and are then temporarily fixed.
2. The LSI layer is created beforehand for stack, and chips are inserted and stacked inside of the cavity. The cavity is designed to hold the chips so that they do not slide laterally when the chips are laminated and bonded. The cavities are also the standard used for lateral alignment.
3. Isotropic pressure and heat are applied at the same time from the top of the highest chip toward the bottom through the elastic material using a gas or liquid covered with a flexible sheet as the pressure medium. Heat is also applied from the bottom of the silicon wafer where the chips are stacked by a pulse heater which heats the stage.

(The conventional flip chip bonding method was used for pressure and heat application for stacked chips. In the flip chip bonding method, the protruding terminals referred to as bumps are aligned with the pads on the substrate when each chip and substrate is electrically connected. Chips are then adsorbed to the flat plate that is parallel with the stage, are moved parallel in the z axis, and the bumps and pads come into contact for receiving heat application and pressure bonding. In this invention, a cavity holds the laminated chips so that they do not slide in the lateral direction, and pressure and heat are applied from the top to the all stacked chips. With the conventional method, bonding failures often occurred with the chip surface because the flat stage and plate are not parallel or dust exists on the flat plate or stage. It was confirmed that this invention is effective at high yield bump bonding with no bonding failures.)

4. After heat and pressure are applied, the flexible sheet that covers the stacked chips are removed, and the template is removed from the wafer.

5. Chip-to-Wafer lamination bonding with high throughput and high yield is completed.

3. Description

There are three types of this 3D integration processes; Chip-to-Chip process, Chip-to-Wafer process, and Wafer-to-Wafer process. Figure 1 shows a schematic diagram of the Wafer-to-Wafer process and Chip-to-Wafer process. In the Wafer-to-Wafer process, LSI through silicon vias and bumps are created at the wafer level, the process for making the layers thinner is executed, and alignment and bonding are executed between the wafers. The laminated wafers are then diced and three dimensional LSI chips are created. In this method, processes are executed at the wafer level before dicing, so it has advantages such as high throughput. The yield ratio for the laminated wafers is a multiple of each wafer's failure rate, which causes the overall yield to be low. The Chip-to-Wafer process is used to increase the yield of whole stacked chips. In the Chip-to-Wafer process, LSI silicon through vias and bumps are created at the wafer level, the process for making the layers thinner is executed, and dicing is executed to create chips. After separating the good products from defects, only good chips are stacked to the wafer at the bottom layer in order to create three dimensional LSI chips. Yield after stack is high because stack is executed after good chips are selected. However, when the good chips are stacked, the heating/pressurizing process and the cooling process need to be executed repeatedly according to the number of stacked layers in order to align the chips with the wafer on the bottom layer, and in order to bond with the bumps. In addition, if the stacked chips are installed in an array on the wafer, the same process needs to be executed repeatedly according to the number of stacked chips that is installed in the array. Therefore, when the flip chip packaging method is used for the Chip-to-Wafer process, it takes a long time, which results in low manufacturing throughput. Figure 2 shows the configuration of the flip chip bonding method for the Chip-to-Wafer process. Table 1 shows the advantages and disadvantages of the Wafer-to-Wafer process and Chip-to-Wafer process. In the flip chip bonding method, the protruding terminals referred to as bumps are aligned with the pads on the substrate when each chip and substrate is electrically bonded. Chips are then adsorbed to the flat plate that is parallel with the stage, are moved parallel in the z axis, and the bumps and pads come into contact for receiving heat application and pressure bonding. Figure 2 also shows the configuration of the flip chip bonding method used for the Chip-to-Wafer process.

Figure 3 shows the processes for Chip-to-Wafer chip integration uses the cavity related to this invention. The

cavity is a hole where the chips are inserted and can be used as a standard for alignment when the chips are stacked. It also holds the chips so that they do not slide laterally during the heating/pressurizing process for bonding. When the cavity is used for 3D chip stack, chips that are inserted into the cavity according to the number for stack layers, heat and pressure are applied, and the 3D stacked chips are created. By using the cavity as the standard surface for alignment, it is not necessary to execute a process for aligning the cavity and each chip that is above or below the cavity. Therefore, it possible to stack multiple chips at the same time. When the stacked chips are created in an array on the wafer, all processes can be executed together by using cavities laid out in the shape of the array. Therefore, manufacturing throughput is greatly improved, and devices with alignment functions are no longer needed. As a result, three dimensional laminated chips can be manufactured at a lower cost. Figure 4 shows a cavity (Hole size: 7.15mm x 6.15mm) that was made using molybdenum. In addition to using processed silicon that is the same silicon used for the stacked chips, metals that are similar to silicon CTE such as molybdenum, glass, and ceramics can all be used as the cavity material.

Figure 5 (a) and (b) show the configuration of the heating/pressure bonging equipment, which is for when the cavity is used for Chip-to-Wafer chip lamination. Usually, the flat plate, which is parallel to the stage where the chips are inserted to the cavity, is used as the jig for heating/pressure bonding of the stacked chips. However, if the flat plate is not parallel with the stage such as the alignment for parallelization degree is not sufficient or some dust exists in the resin on the flat plate or stage, or if the chip is inclined, a contact failure will occur such as the bump on the chip will not contact or bond with the opposite electrode pad. Such problems occur significantly when the bump height is low. For example, if the bump height is 100um, a problem will not occur because the parallel declination can be adsorbed by the height of the bump. However, if the height of the bump is 10um or less, a problem will likely occur, which can also negatively affect the bonded chip area. When three-dimensional lamination is used, the thickness of the whole laminated chip needs to be thin and the height of the bump also needs to be smaller for releasing heat. Figure 6 shows the height of the bumps and the materials used for the chip stack test. Sn-Cu was used as the material for the bumps, and the height was less than 7um. Figure 7 shows a photo of the sample bonding when the flip chip bonder was used for bonding. Chips and substrates were bonded using the flip chip bonding method, and then a shear test was conducted. The photo shows where the substrate contacted the bumps. The silver circles clearly seen in the top left are the imprints where the bumps bonded. The silver outlines seen at the bottom right indicate that the bumps did not have full contact. This indicates that bump bonding on the chip surface was not even and all bumps were able to bond equally.

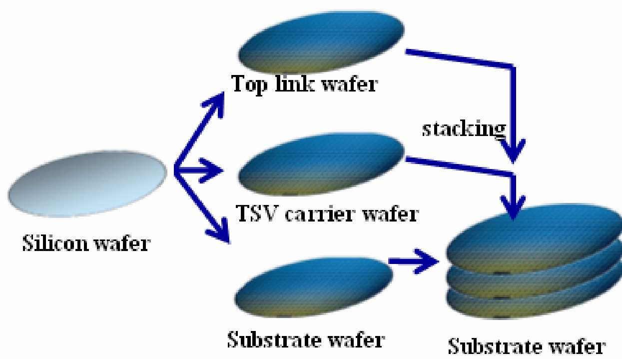
This invention proposes lamination technology for equalizing the weight during Chip-to-Wafer bonding, which will increase the yield for bump bonding. Figure 8 shows the Chip-to-Wafer chip lamination technique where the chip is laminated with equal pressure using the cavity. In this method, isotropic pressure is applied from the top of the chip using a gas or liquid as the pressure medium. Lateral slippage is avoided by inserting multiple laminated chips into the cavity. As a result, multiple layer chips can receive pressure bonding and lamination at the same time. Using a gas or liquid as the pressure medium is not by itself the chip lamination method. We need both cavity technology and a device which provide isotropic pressure for stacked chips in order to get high throughput and yield. This is because the pads are placed in the narrow pitch between the stacked chips where the thin/thick chips are layered, and the chips can easily slip when a small external force is applied in one

direction when we use conventional bonding tool to stack chips at once. And chips can easily slip due to static from the tweezers. For example, pads with a 200um pitch where the chips are layered moved easily when a force was applied. Our goal was to establish a Chip-to-Wafer 3D integration technique that can achieve high throughput for bonding/laminating processes for multi-layered chips, and achieve high yield from bump bonding. It is believed that this method for alignment, bonding, and lamination using a cavity (which has already been submitted for patenting) along with using a gas or liquid as the pressure medium is a new technique for achieve this. This invention can improve both throughput and bump bonding yield between chips, which are issues with the convention Chip-to-Wafer process, when three-dimensional chips are being manufactured.

Figure 9 shows the temperature and load for the practice of the invention. In this test, heat and pressure were applied through an elastic material using gas as the pressure medium. The laminated chips were inserted into the cavity, and the elastic material for applying the pressure was placed and fixed in the fixture. Then, thick metallic plates were used for pressing above and below the fixture. Here, the heating time was long because the heat capacity of the thick metallic plates was large. However, it is possible to reduce the heating time by using thinner metallic plates with pulse heat. Figure 10 shows a photo of the fixture that was used for this test. As Figure 10 shows, the template with the cavity was held by the fixture, and the valve for sending the compressed gas that was installed to the fixture can also be seen. N2 gas was used in this test as the pressure medium, but it is also possible to use other gasses such as carbon dioxide. Based on the conditions shown in Figure 9 and the fixture shown in Figure 10, the chips were inserted into the cavity, isotropic pressure was applied from the top of the chip through the elastic material using gas as the pressure medium, and the multi-layered chips in the cavity were given heat and pressure bonding at the same time. Figure 11 shows a photo of sample bonding when this invention was used. The imprints from where the bumps bonded can be clearly seen on the substrate for the whole chip. This indicates that the bump bonding was equal over the whole chip surface and that all bumps bonded evenly.

■ Effect of the Invention

It has become possible to apply pressure evenly during Chip-to-Wafer bonding and to increase bump bonding yield. By using the cavity method, it is possible to laminate multi-layer chips at the same time, and to execute the Chip-to-Wafer process with high throughput and high yield.



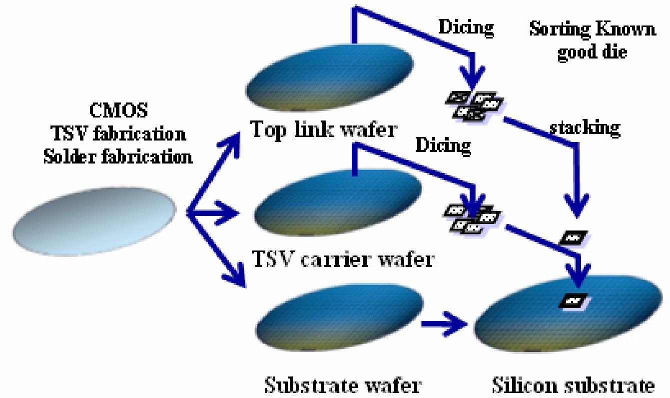


Figure 1. Schematic illustration of (a) wafer-to-wafer 3D integration and (b) die-to-wafer 3D integration processes.

Table 1. Features of 3D integration technologies

	Wafer-to-Wafer	*Die-to-Wafer*
Alignment required	Wafer scale	Die size
Yields	Low	High
Throughput on stacking	High	Low
Suitable for	Common size	Both common and dissimilar size

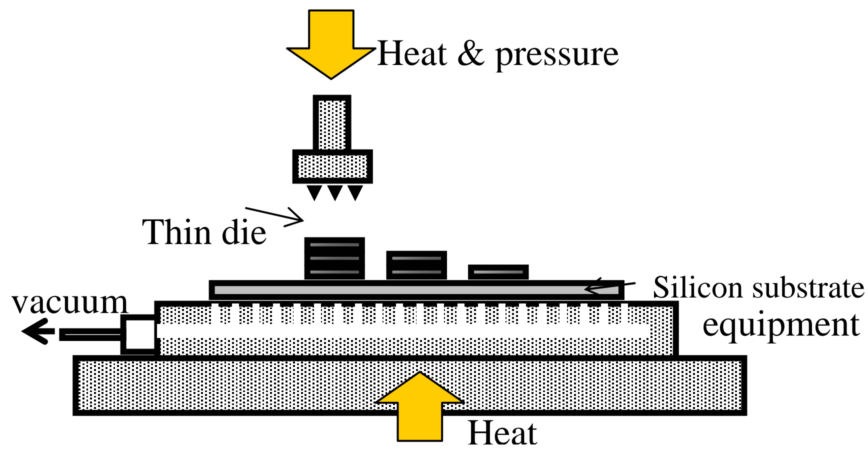


Figure 2. Configuration of conventional die-to-wafer lamination.

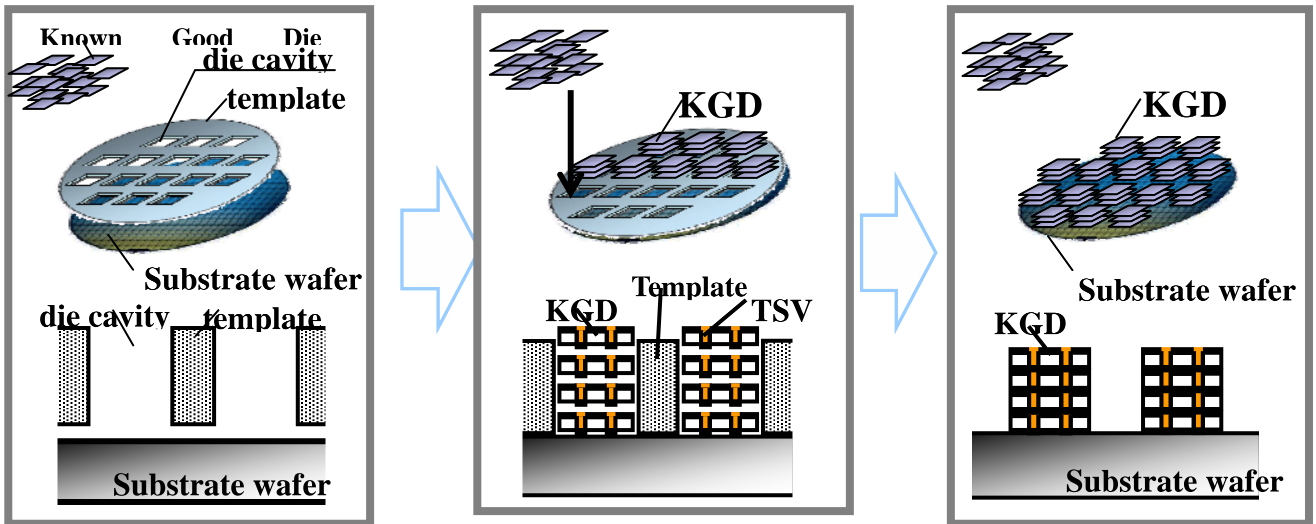


Figure 3. Schematic illustration of die-to-wafer 3D integration process with the die cavity technology.

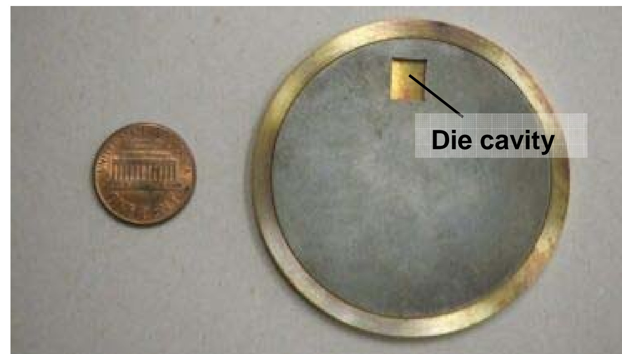


Figure 4. Photo of the molybdenum cavity prototype.

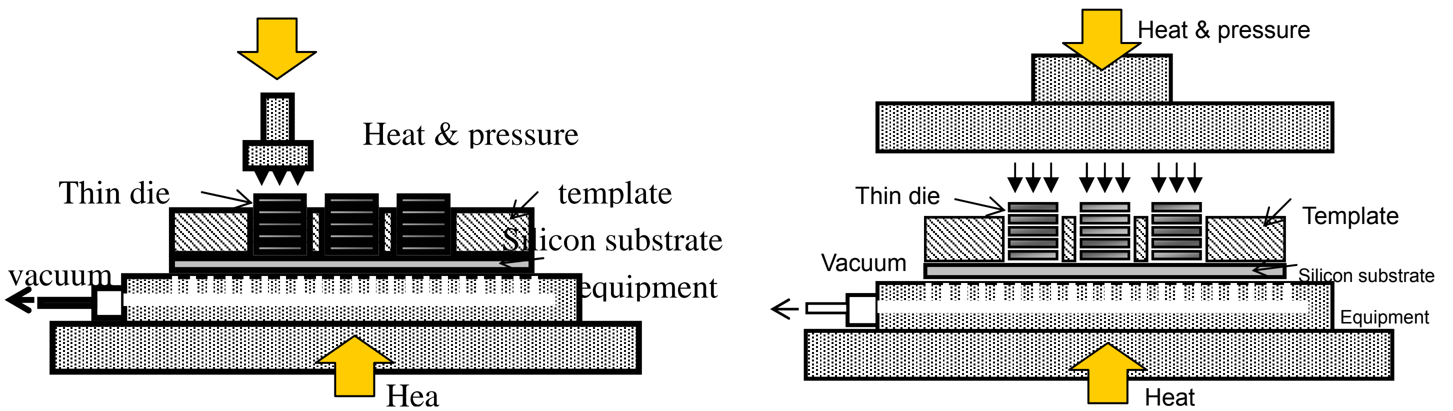
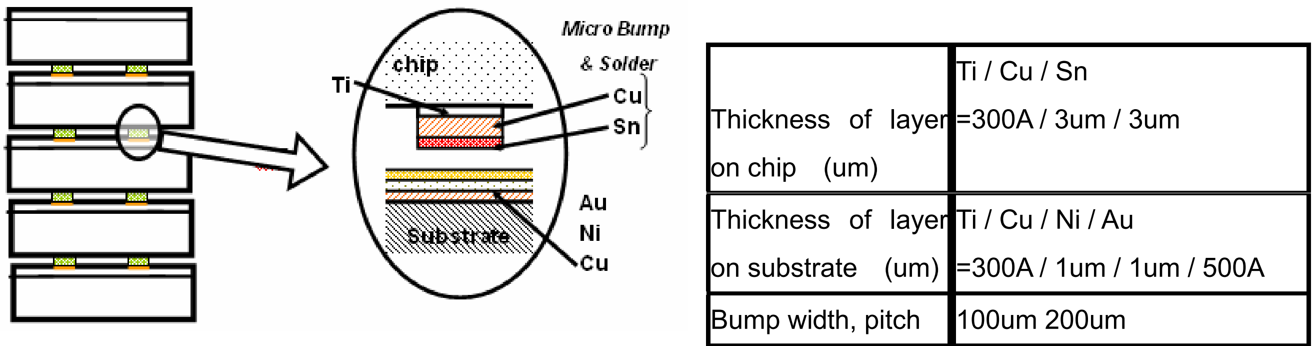


Figure 5(a)(b). Configuration of the die-to-wafer lamination using the die cavity technology.



Thickness of layer on chip (um)	Ti / Cu / Sn =300A / 3um / 3um
Thickness of layer on substrate (um)	Ti / Cu / Ni / Au =300A / 1um / 1um / 500A
Bump width, pitch	100um 200um

Figure 6. Film thickness and bump materials used for the bonding test.

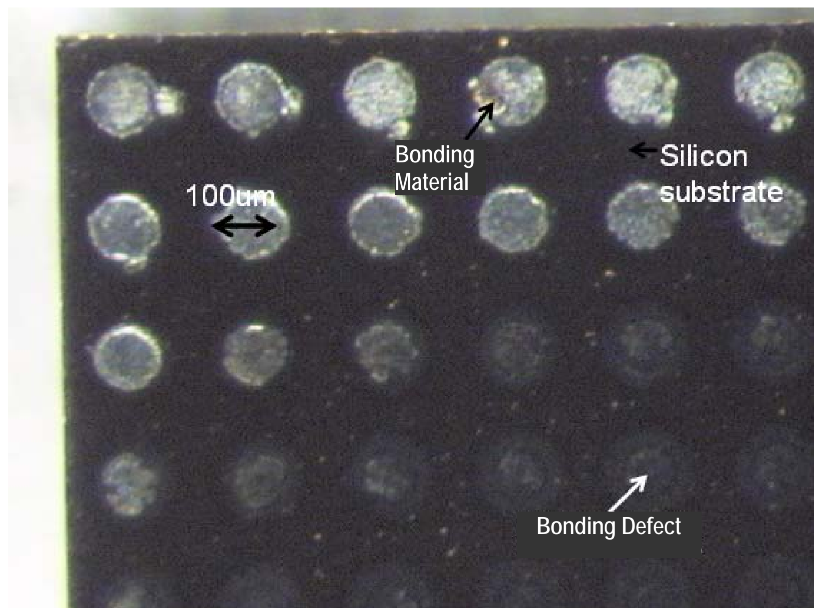


Figure 7. Photo of sample bonding using the flip chip packaging method (sample surface after the shear test).

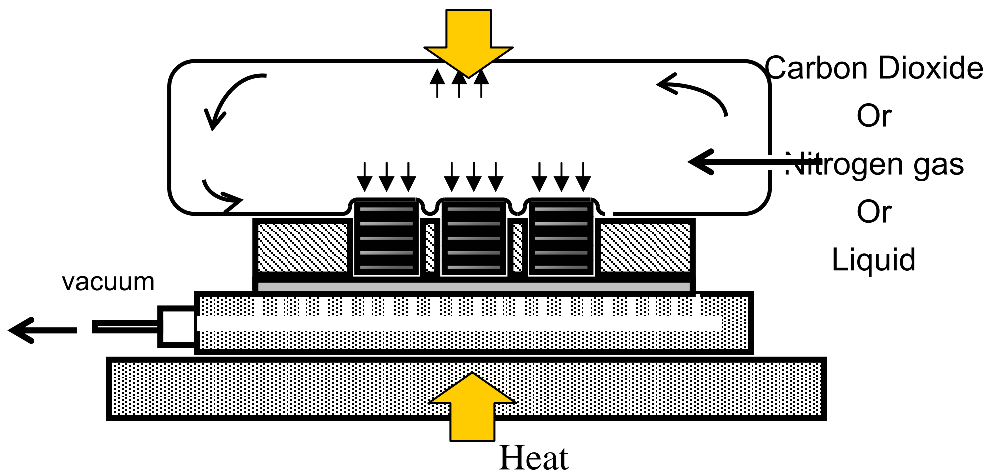
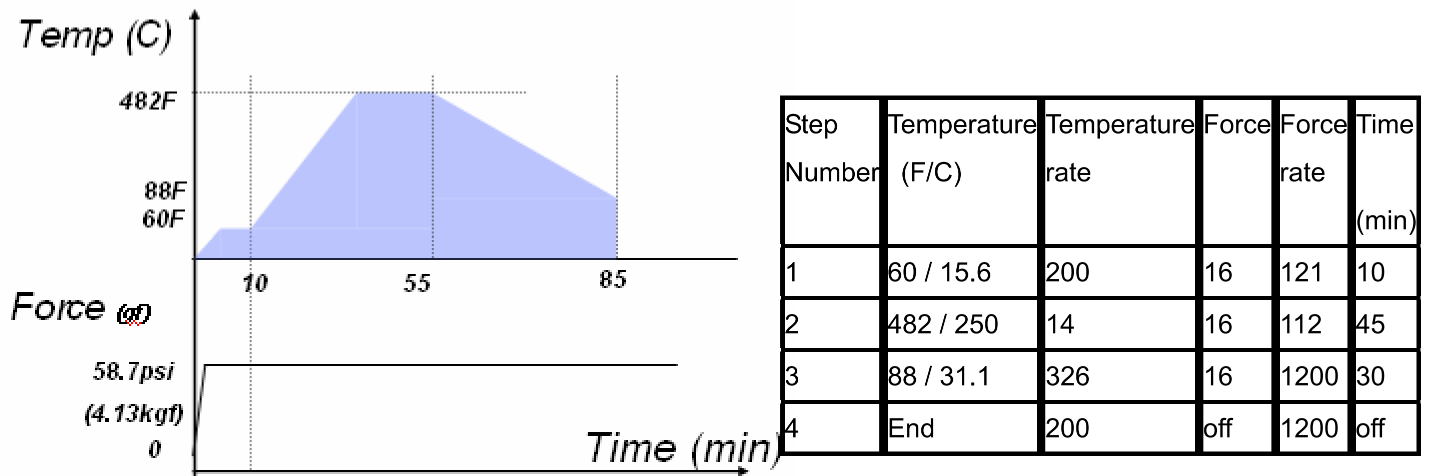


Figure 8. Configuration of die-to-wafer lamination using the die cavity technology with autoclave lamination.



Step Number	Temperature (F/C)	Temperature rate	Force	Force rate	Time (min)
1	60 / 15.6	200	16	121	10
2	482 / 250	14	16	112	45
3	88 / 31.1	326	16	1200	30
4	End	200	off	1200	off

Figure 9. Bonding load and conditions.

Figure 10. Photograph of an early fixture for lamination and bonding.

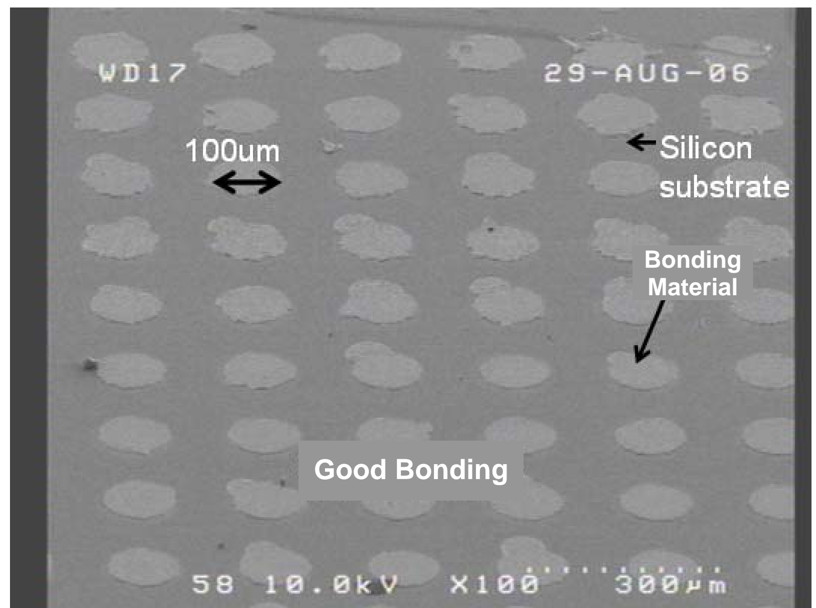


Figure 11. Photo of sample bonding using the invention (sample surface after the shear test).