Distribution of Local Thermal Residual Stress in Thin Chips Stacked by Flip Chip Structures

doi: 10.1109/IMPACT.2006.312210

Distribution of Local Thermal Residual Stress in Thin Chips Stacked by Flip Chip Structures

Hideo Miura, Nobuki Ueta, and Yuhki Sato
Fracture and Reliability Research Institute, Graduate School of Engineering, Tohoku University
6-6-11-712 Aoba, Aramaki,
Aoba-ku, Sendai, Miyagi 980-8579,
Japan

ABSTRACT

Mechanical reliability issues such as cracking of LSI chips and shift or deterioration of electronic performance of them caused by mechanical stress and strain in multi devices sub-assembly (MDS) structures were discussed analytically and experimentally. Local thermal deformation due to thinning of the LSI chips for mobile application causes large distribution of residual stress from –300 MPa to +150 MPa in the chips. The values of the maximum and the minimum stresses are strong functions of the thickness of the LSI chips and period of area-arrayed small bumps. In flip chip assembly structures, periodic stress or strain distribution appears in the thinned chips depending on the period of the area-arrayed bumps. The amplitude of the stress often exceeds 100 MPa, and it may cause the change of electronic performance and reliability of devices. In addition, both the amplitude and the average stress vary among the three-dimensionally stacked thin chips due to macroscopic bending of the assembled structure. Therefore, it is very important to optimize the MDS structures to minimize the stress and thus, to improve the reliability of products.

INTRODUCTION

Electronic products such as mobile phones and PCs have been miniaturized continuously and their functions have been improved drastically. Three dimensionally stacked structures such as multi-chip modules and multi-chip packages are indispensable for increasing the assembly density. There are various kinds of structures for the multi chip assembly. One of the most important application of the multi chip structures is a system in package (SiP) that is composed of the combination of CPU chips and memory chips. The conventional methods of the interconnection between an LSI chip and a substrate or another chip are wire bonding or small metallic bumps. A stacked structure is indispensable for maximizing the assembly density. However, the total thickness of the stacked structure is strictly limited especially for mobile application. To minimize the total thickness of the modules or packages, each chip has been thinned to less than 100 μm. However, since the mismatch in the thermal expansion coefficient between the chip and materials used in the modules such as molding compounds printed circuit boards, small bump and underfill for interconnection is large, large local deformation of the chip caused by thermal stress increases drastically because of the decrease of bending modulus of the thinned chips. Such a localized high stress or strain gives rise to fractures such as delamination or cracking of materials composing the modules. In addition, since high stress and strain deform the crystallographic structure of the semiconductor and dielectric materials, the electronic functions and reliability of the chips may deteriorate due to the change of band gap of the materials [1]-[3].

In this paper, the effect of thinning of the LSI chips on the residual stress in stacked thinned chips is analyzed using a finite element method. Local deformation of the LSI chips is the most important issue to be discussed for minimizing the residual and thermal stresses in the stacked structures. The estimated stress distribution is proved by stress measurement of Si chips mounted by flip chip technology by using stress sensor chips that are composed of 168 poly-crystalline-silicon-film gauges of 10-μm in length.

STRUCTURAL ANALYSIS

Since the number of I/O pins of LSI chips has been increasing drastically, a flip chip structure is the most important structure for the multi chip assembly application. This is because it is easy to miniaturize a diameter of metallic bumps such as Au bumps and Cu bumps to less than 100 μm as shown in Fig. 1. For mobile application, since most LSI chips are mounted on low-cost organic substrates, metallic bumps are pressed heavily to make stable contact with interconnections on the substrate during fabrication. The pressed bumps are surrounded by an underfill material such as epoxy resin, ACF (Anisotropic Conducting Film) and NCF (Non Conducting Film) to maintain compressive stress at the connected interfaces [4].

Conventionally, the bending modulus of an LSI chip is much higher than that of an organic substrate. Thus, the Cu interconnection on the organic substrate deforms seriously. The elastic deformation of the interconnection layer gives rise to compressive stress at the connected interface between the Cu bump and the substrate or a Si chip. This deformation also minimizes the change of residual stress in LSI chips and thus, no serious damage occurs in them after the assembly. However, since the thinning of an LSI chip decreases the bending modulus of the chip, the LSI chip starts to deform locally when the thickness of the chip is thinner than a certain critical value. Such a local deformation may cause sharp gradient of residual stress around the Cu bumps and thus, various failures. Therefore, the residual stress in an LSI chip after flip chip bonding was analyzed using a finite element method.

Figure 2 shows an example of a three-dimensional finite element model for the stress analysis [5]. Assuming the symmetry of the total structure, one fourth of the total structure was modeled for the stress analysis. Area arrayed bump layout was assumed. The diameter and the height of Cu bumps were fixed at 100 μm and the period of the bump was varied from 20 μm to 500 μm. The width of the chip was assumed to 5 mm. The chip was mounted on an organic printed

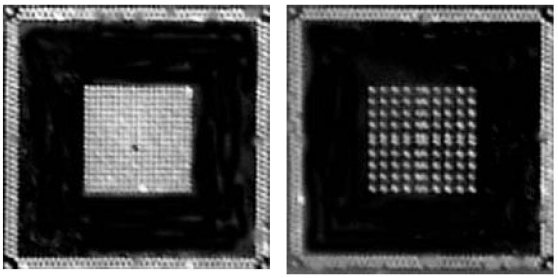

Fig. 1 Examples of the Cu bump alignment on substrates for evaluating the local displacement at a surface of a Si chip

1-4244-0735-4/06/$20.00 ©2006 IEEE

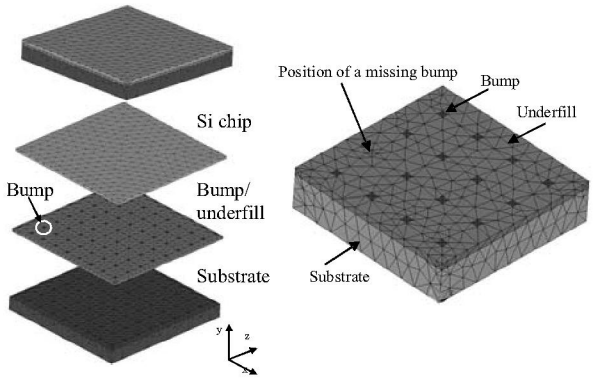

Fig. 2 Three-dimensional finite element model for analysis of residual stress in a Si chip mounted on a substrate by flip chip technology

circuit board at 150°C. Both a single-side and a double-sided mounting structure were assumed. In the latter case, the base line of the substrate was fixed. This boundary condition prohibits the substrate to bend. The thickness of the chip was varied from 280 μm to 50 μm. To analyze the local stress in a Si chip between two bumps precisely, a sub structure was also modeled as is also shown in Fig. 2. Only an area of sixteen (4 x 4) bumps was analyzed. The deformation of the side surface of this sub-model calculated using the whole structure is given as a boundary condition in this sub-model analysis. The total number of the nodes of the three dimensional model and elements were 27378 and 24548, respectively, and those of sub model were 31433and 28224. The materials constants used in the stress analysis is summarized in Table 1. Most materials were assumed to be elastic. Only Cu bumps were modeled as elastic-plastic material. Yield stress of the Cu bump was assumed to be 250 MPa.

Figure 3 indicates the definition of the relative displacement, in other words, the difference of the local deformation at a surface of a silicon chip. First, the local deformation at the surface of a Si chip in the regular (no defect) structure was calculated. In this figure, three bumps A, B, C are aligned at a constant interval. Then the center bump B is eliminated to model a missing bump. The open space is filled up by underfill. Then, the local deformation of this structure is analyzed. The relative displacement is defined as the difference of the magnitude of the surface deformation of the Si chip at the center position of a bump B as shown in the figure. When this relative displacement is larger than 10 nm, it can be measured by applying laser technique.

The effect of the thickness of a Si chip on the relative displacement at a chip surface is analyzed by varying the thickness only. This means that the bumps are still missing as a line defect. The calculated relative displacement is summarized in Fig. 4. When the thickness of a Si chip is thicker than 500 μm, the relative displacement at a surface of the chip is less than 50 nm. On the other hand, it increases drastically, when the thickness is decreased to less than 200 μm. The maximum relative displacement increases monotonically with decrease of the thickness of a Si chip, and it reaches about 600 nm when the thickness is decreased to 50 μm. In addition, the full width at a half maximum (FWHM) also decreases with decrease of the chip thickness from about 1 mm to about 200 μm when the chip thickness is less than 100 μm. This value is the same as the period of the area-arrayed bumps. Therefore, it is possible to detect the lack of bumps in a flip chip bonding structure by measuring this relative displacement, when the thickness of a Si chip becomes less than 200 μm.

The magnitude of this relative displacement does not change so much even when two chips are stacked as shown in Fig. 5. In this

Table 1 Materials constant

Material	Young's modulus (GPa)	Poisson's ratio	Thermal expansion coefficient (10^{-6}/K)
Si	167	0.07	3.0
Cu bump	117	0.3	17
Underfil	3	0.3	20
Substrate	15	0.3	15

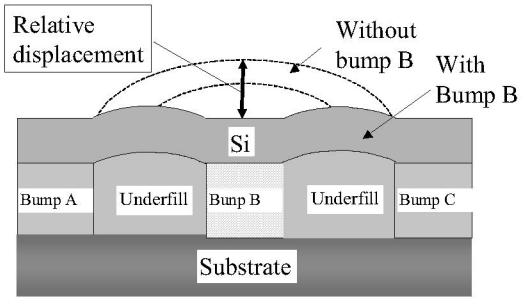

Fig. 3 Definition of relative displacement

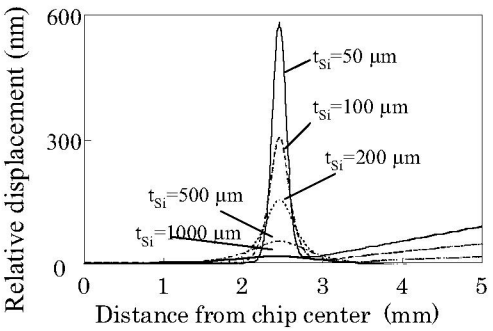

Fig. 4 Effect of thickness of a silicon chip on the relative displacement at a back surface of the chip

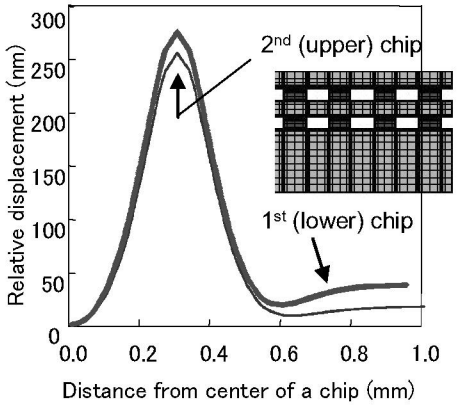

Fig. 5 Surface relative displacement of stacked chips

analysis, the thickness of both chips is fixed at 100 μm. When the period of the bumps are the same between the first and the second interconnection layers, the relative displacement of the second (upper) chip is almost the same as that of the first chip. This is because the bending modulus of the stacked chips is much smaller than that of the substrate. However, the amplitude of the relative displacement of the second chip changes drastically when the period

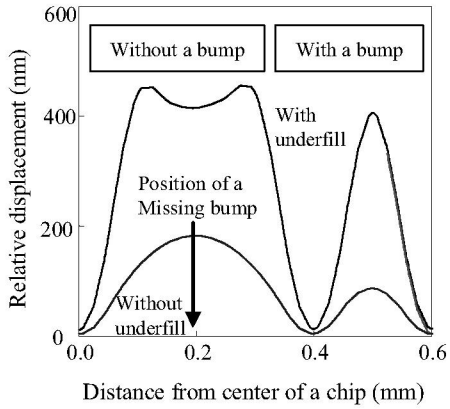

Fig. 6 Residual stress distribution in stacked chips

of the bumps in the second interconnection layer is different from that of the first layer.

An example of the effect of the layout design of bumps on the local relative displacement is shown in Fig. 6. In this analysis, bumps are aligned with a constant period of 200 μm. Only one bump at 0.2 mm from a center of a chip is eliminated. The surface of the chip is deformed locally by about 400 nm because of the mismatch of thermal expansion coefficient among interconnection materials. When a bump is eliminated, in other words, the period of the bumps is changed, the local deformation changes drastically. The amplitude of the local deformation at a position where the bump is eliminated becomes about twice of that at regular position even before filling up by underfill resin. After the filling of underfill resin, the deformation pattern changes depending on the period of bumps in the area of interest. Since the local strain in the silicon chip is determined by the local deformation of the chip, this fluctuation of the surface deformation causes a complicated strain (stress) distribution in the chip, and thus , fluctuation of the electronic performance of element devices in the chip.

Such a local deformation (relative displacement) of a silicon chip gives rise to a periodic stress distribution as shown in Fig. 7.The main reason for this periodic stress distribution is the mismatch in the thermal expansion coefficient between a metallic bump and underfill resin. The amplitude of the periodic distribution is a strong function of the thickness of the stacked chips. This amplitude also increases drastically when the thickness of a silicon chip becomes less than 100 μm. In this analysis, the thickness of both chips is 100 μm. A clear periodic stress distribution occurs in the upper chip, and the amplitude of the stress reaches about 100 MPa. This value agrees well with the result of a single chip assembly with the same bump alignment [6][7]. On the other hand, the amplitude in the lower chip is much smaller than that, about 30 MPa. This is because that the local deformation at the top surface of the lower chip is strictly limited by the upper chip. In addition, the average residual stress in the upper chip is higher than that in the lower chip by about 60 MPa. This shift is caused by the macroscopic deformation of the assembled structure. Since the thermal expansion coefficient of the substrate is larger than that of a silicon chip, the macroscopic shape of the assembled structure is convex upward when a chip is mounted on a single side of a substrate. And, besides, the distance from the neutral axis of the upper chip is larger than that of the lower chip. Thus, the average stress of the upper chip stays tensile stress side.

This difference of the stress distribution between the stacked chips is also a strong function of an alignment of bumps. Even when the period of the bump is the same between the upper (second) interconnection and the lower (first) interconnection, the position of the bumps affects the residual stress in the stacked chips. Figure 8

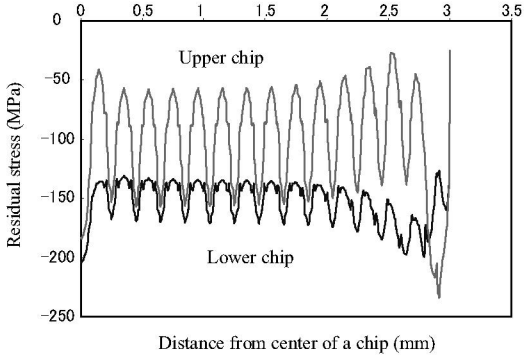

Fig.7 Distribution of residual normal stress along a stacked chip surfaces. (The phase of the bumps with a constant period of 200 μm is the same between the first and second interconnection layers.)

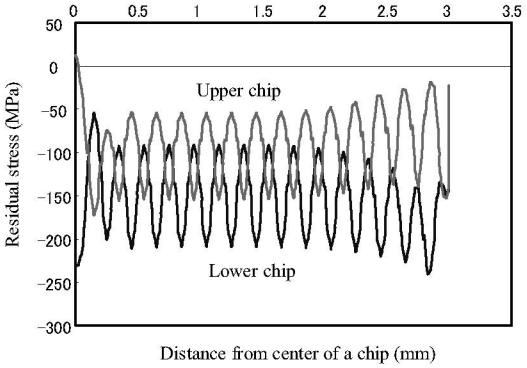

Fig.8 Distribution of residual normal stress along a stacked chip surfaces. (The phase of the bumps with a constant period of 200 μm is shifted by 180 degrees between the first and second interconnection layers.)

shows another stress distribution in the stacked chips. In this analysis, the phase of the bumps in the upper interconnection layer is shifted by 180 degrees. This means that underfill resin in the second layer is located above a bump in the first interconnection layer and vice versa. In this structure, the stress distribution in the upper chip is almost the same as that shown in Fig. 7. While, the periodic stress distribution becomes clearer in the lower chip comparing with the result shown in Fig. 7, though the average stress does not change so much. Therefore, the interconnection structure (the layout of bumps) is another important structural parameter that determines the residual stress in stacked chips.

Since the electronic functions that are determined by band gap structure in an active layer in a silicon chip are a function of stress (strain) in the layer, such a difference in the residual stress in stacked chips causes the layer-by-layer shift of the electronic functions of the chips. When the same memory chips are stacked, for example, the performance or reliability of these chips sometimes differs significantly with each other depending on the interconnection structure. It is very important, therefore, to optimize the structure of the stacked chips such as the thickness of each chip and the structure of each interconnection layer to assure the reliable performance of products.

EXPERIMENT

In order to prove the occurrence of a periodic distribution of the residual stress in a Si chip, stress sensor chips were developed using a polycrystalline silicon thin films by applying the piezoresistive effect of silicon [6]-[7]. The average grain size of the polycrystalline

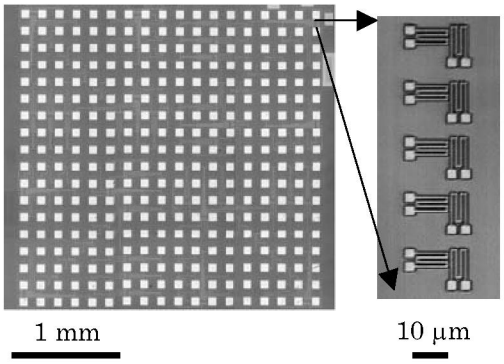

Fig. 9 Outlook of a test chip for measuring local stress distribution at a surface of the chip between two bumps. Five gauges are aligned between two bumps with a pith of 100 μm.

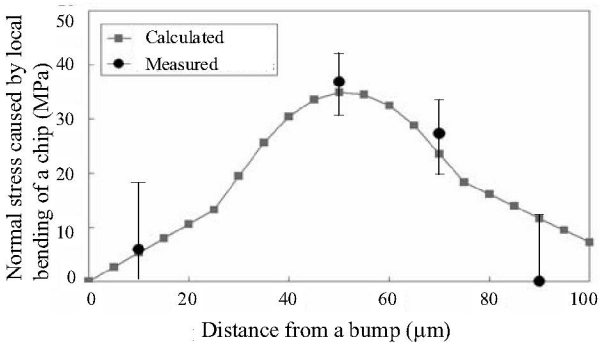

Fig. 10 Comparison of the measured stress with a calculated result

film was less than 0.1 μm. A basic structure of strain gauges is shown in Fig. 9. The length of each gauge is 10 μm to measure the distribution of residual stress between two bumps that are apart from 100 μm with each other. The width of the gauge is 0.5 μm. The total length of each gauge is about 40 μm. The sheet resistance of the film was controlled by ion implantation of As, and it was about 420 Ω/square. Thus, the average resistance of each gauge was about 40 kΩ. These gauges were made of an 1-μm thick polycrystalline film that was deposited on a thermally oxidized silicon wafer. 296 gauges were formed on a 4 mm x 4 mm Si chip as shown in Fig. 9. Thickness of the chip was 280 μm. 324 Cu bumps are area-arrayed on a chip by electroplating. Both the width and the thickness of the bump are 100 μm, and the interval of the bump is also 100 μm. 5 gauges are formed between the two bumps at most. 168 gauges can be used at a one chip. The gauges of interest are connected to the bumps by thin film interconnection.

Figure 10 shows the comparison of the measured stress with the calculated result. In this figure, only the difference of the local stress from the average stress is summarized. The horizontal axis of this figure is the distance from the edge of one bump, and the left side bump is close to the center of a sensor chip. The calculated stress distribution is not symmetric between the bumps because of the macroscopic bending of the chip. The measured results are also plotted in this figure with error bars. Each error bar indicates the maximum and the minimum measured value among 20 samples. The maximum fluctuation of the measured results is about 10 MPa. The measured results agree well with the calculated results. The maximum difference between the average measured results and the calculated one is about 10 MPa. Therefore, the authors concluded that the local stress distribution appears at a surface of a Si chip mounted on a substrate by an area-arrayed bump structure when the thickness of a Si chip is less than 200 μm.

COCCLUSIONS

Local residual stress at a surface of a silicon chip assembled to a substrate using area-arrayed metallic bumps was analyzed by a finite element method and measured using a stress sensor chip. Both the maximum and the minimum principal stresses that occur in an assembled chip increase monotonically with decrease of a thickness of the chip. In addition, a periodic stress distribution occurs at the chip thinner than 200 μm because of the mismatch of thermal expansion coefficient between the metallic bump and underfill material that surrounded the bump. The amplitude of the stress is a strong function of the thickness of a silicon chip and the intervals of the bumps. It exceeds 100 MPa when the thickness of a Si chip is less than 50 μm. In addition, both the local surface deformation and the distribution of the residual stress among the stacked plural chips varies drastically depending on the layout of bumps in each interconnection layer. The occurrence of the local stress distribution between two bumps was proved by applying a stress sensor chip that includes 168 strain gauges, 10 μm long and 0.5 μm wide. Therefore, it is very important to minimize both the amplitude of the residual stress in one chip and the fluctuation of the residual stress among the stacked chips to improve the reliability of products.

ACKNOWLEDGMENT

This research work is partly supported by .the Grants-in-Aid for Scientific Research and the 21COE program of Tohoku University. The authors thank Profs. M. Koyanagi, S. Kiyono, and W. Gao of Tohoku University for their helpful suggestions and discussions.

REFERENCES

[1] A. Hamada, T. Furusawa, N. Saito, and E. Takeda, "A new aspect of mechanical stress effects in scaled MOS devices," *IEEE Trans. on Electron Devices*, Vol. 38, No. 4 (1991), pp. 895-900.

[2] H. Miura and A. Nishimura, "Device Characteristic Changes Caused by Packaging Stress," *ASME, Mechanics and Materials for Electronic Packaging*, Vol. AMD-195, (1994), pp.101-109.

[3] Y. Kumagai, H. Ohta, H. Miura, and A. Shimizu, "Stress-Optimization of a Thin Film Semiconductor," *Proc. of The Int. Symp. On Micro-Mechanical Engineering*, Tsuchiura, (Dec., 2003), pp.427-432.

[4] N. Tanaka, K. Kawano, H. Miura, Y. Kada, and I. Yoshida, "A Highly Reliable Design for a Nonmetallurgical Contact-Joint Structure Consisting of an Adhesive Film," Trans. ASME, J. of Electronic Packaging, Vol. 126, (2004), pp. 82-86.

[5] Y. Sato and H. Miura, "Non-destructive Inspection Method for Detecting Open Failures in Flip Chip Structures," Proc. of ASME InterPACK '05, (2005), pp. 1-6 (CDROM).

[6] N. Ueta and H. Miura, "Measurement of Local Residual Stress of A Flip Chip Structure Using A Stress Sensing Chip," Proc. of ASME InterPACK '05, (2005), pp. 1-6 (CDROM).

[7]. H. Miura, N. Ueta, and Y. Sato, "Local Thermal Deformation and Residual Stress of a Thin Si Chip Mounted on a Substrate Using An Area-Arrayed Flip Chip Structure," Proc. of EMAP 2005, (2005), pp.220-225

[8] H. Miura, M. Kitano, A. Nishimura, and S. Kawai, "Thermal stress measurement in silicon chips encapsulated in IC plastic packages," *ASME, J. of Electronic Packaging*, Vol. 115, (1993), pp. 9-15.

[9] J. C. Suhling and R. C. Jaeger, "Silicon Piezoresistive Stress Sensors and Their Application in Electronic Packaging," *IEEE, Sensors Journal*, Vol. 1, No. 1 (2001), pp. 14-30.