Reliability challenges in 3D IC packaging technology

K.N. Tu

Department of Materials Science and Engineering, University of California, Los Angeles, CA 90095-1595, USA

1. Introduction

Microelectronic industry has two major technologies; one is chip technology and the other is packaging technology. The progress in chip technology has been guided by Moore's law, which states that the density of 2-dimensional integrated circuits (2D IC) on Si chips will double every 18–24 months. It has been true in the past 30–40 years, as shown in Fig. 1, which is a plot of circuit density of memory and logic devices on Si chips against year. For memory devices, we see that it started at 1 k (thousand) per chip from 1975 and reached 1 m (million) per chip at 1990. The increase of three orders of magnitude took 15 years. Since we have 2^10 = 1024, it should take 10 × 18 months = 15 years. Fig. 1 also shows that from 1 million to 1 billion, it took 20 years, from 1990 to 2010. Beyond 1 billion, it is hard to project the future since we are advancing into nanotechnology; the critical feature size is down to 22 nm, approaching the physical limit of miniaturization. Furthermore, economical considerations also become very unfavorable.

Today, the microelectronic industry cannot but look for a new direction to grow and the new direction is 3-dimensional integrated circuits (3D IC) [1–12]. In 3D IC, the chip technology and packaging technology are merged closer and closer together. Thus, it is of interest to ask whether or not there is a Moore's law for the packaging technology. If so, what is it and what is its prediction of the future?

It seems that we might consider the size of a solder joint or the number of solder joints per unit area in electronic packaging technology [13,14]. Fig. 2 shows the SEM image of an array of solder bumps on a chip surface. If we start from ball-grid-array (BGA) solder joints, its diameter is about 760 µm. The smallest pitch between the solder bumps can be assumed to be the same as the diameter, so we have about 10–100 joints/cm². Next is the C-4 joints used in flip chip technology, and the diameter and the pitch of C-4 solder balls are about 100 µm, so we have 10^5–10^6 joints/cm². Fig. 3 shows SEM cross-sectional images of the 2-level packaging technology, in which C-4 solder joints are used in the 1st level chip-to-module packaging and BGA joints are used in the 2nd level module-to-board packaging. The BGA joints are shown at the lower right corner, and the C-4 joints are shown at the upper left corner of the figure. The upper right corner shows a cross-sectional image of multi-layered Cu interconnects, and the Si chip is on the top side.

Today, electronic industry is developing micro-bumps for TSV (through-Si-via) technology for 3D IC and the diameter of a micro-bump is about 20–10 µm, so the density is about 10^5–10^6 joints/cm² [1–12]. It is likely that the diameter of micro solder bumps can be reduced to 1 µm and the density will be 10^7–10^8 joints/cm². Thus, we will have two orders of magnitude reduction in the diameter of solder bumps and four orders of magnitude increase in density. Assuming that the Moore's law can be applied to the rate of change of solder joints, it might take 15–20 years to do so, hence this is the new direction of growth or the new paradigm of the future of microelectronic industry. Actually, major electronic companies have already worked on 3D IC.

There are plenty of opportunities of research and development in 3D IC. In terms of applications, it is abundant in the area of consumer electronic products, for example, medical electronics. From the point of view of processing 3D IC or the manufacturing of TSV and micro-bumps, reproducibility and high yield are very
challenging at the moment. Not only the etching of the very large aspect ratio of via holes in Si wafers and the plating of Cu or W or doped poly-Si into the via holes, but also the alignment of the vias among the stacking of several chips is non-trivial due to chip warpage. From the point of view of device reliability, many challenging issues are expected due to the very large gradient forces exist in the 3D structure. In a dense 3D integration, joule heating will be the most serious problem and the heat must be conducted away. We must have temperature gradient to do so. If we have just 1 °C across a micro-bump of 10 μm in diameter, the temperature gradient will be 1000 °C/cm, which is huge in term of thermomigration. Besides, we expect very large voltage gradient (electric field), chemical potential gradient, and stress potential gradient too. Since it is a 3D structure, the reliability problems are new and very little experimental data are available.

2. Vertical interconnects by TSV and micro-bumps

In 3D IC, the processing challenge is in vertical interconnection, which is made by using micro-bumps to join the TSV (through-Si-vias) chips in stacking. We might recall that in Al and Cu intercon-

![Fig. 4. Schematic diagram of the cross-section of vertical interconnects between the stacking of three wafers.](image1)

![Fig. 5. (a) A synchrotron radiation X-ray tomography of an array of micro-bumps and (b) an enlarged X-ray tomography of the cross-sectional view of a row of micro-bumps. (Courtesy of Tian Tian, UCLA and Alastair Macdowell, ALS/LBNL.)](image2)

![Fig. 3. SEM cross-sectional images of two levels of packaging technology where solder joints are used. (Courtesy of Yishao Lai, ASE, Taiwan.)](image3)
connect technology, there are vertical W-vias in multi-level Al interconnects and vertical Cu-vias in dual damascene Cu interconnects. Yet the height of these vias is about 1 μm. In TSV, the height of the vias is about 200 μm or it is the same as the thickness of the Si chip. The diameter of vias in TSV is now 20 μm and may approach 5 μm soon.

Fig. 4 depicts the cross-section of vertical interconnects between the stacking of three chips. The Cu columns in TSV are joined by micro-bumps. Fig. 5a shows a synchrotron radiation X-ray tomography of an array of micro-bumps, and Fig. 5b is an enlarged X-ray tomography of the cross-sectional view of a row of micro-bumps. Fig. 6 shows a focused-ion-beam (FIB) polished cross-section of the metallization in a micro-bump. On both sides of the solder joint, the layer of Ni on the Cu via is shown. In between the Ni layers is Ni–Sn IMC and the unreacted Pb-free solder. If a post-soldering aging at 150 °C for 1000 h is performed, it will transform the entire joint into IMC completely. In Fig. 6, we can observe a line of crack near the middle of the solder joint. No doubt, the crack is undesirable from the point of view of mechanical properties of the joint. The reason of formation of the crack is unclear yet. One possible reason is Kirkendall (or Frenkel) voids formation during the joining to form the micro-bump. In processing the micro-bumps, Pb-free solder was deposited on the Ni which is on the Cu via. Then, two chips with the solder on the vias were positioned to face each other and joined by thermal compression and followed by the post-soldering aging. There are three major reliability causes given below on the basis of the above 3D structure.

2.1. Joule heating

Due to the stacking of chips and dense packing of circuits on each chip, joule heating is extremely serious and the heat must be removed effectively. Joule heating will have at least three issues. (a) It increases the device temperature and affects atomic diffusion which in turn will affect interfacial reactions and electromigration, thermomigration, and stress-migration. (b) It will generate thermal stress due to different thermal expansion, which in turn will lead to interfacial fracture and chip warpage. (c) Device engineers need to conduct heat away from the chip side by cooling, so a large temperature gradient will be created and thermomigration can be serious in the micro-bumps [15,16].

2.2. Chip warpage

When through-Si-vias (TSV) are used for vertical interconnects, it will lead to chip warpage since the distribution of the vias is non-uniform in the wafer as shown in Fig. 5. Typically, the TSV is arranged on the peripheral of a chip or in the center of a chip. Owing to different thermal expansion between the Si matrix and the Cu column in the vias, it will lead to compressive stress in the chip so the chip will buckle. The warpage increases with the number of the stacking chips. Consequently, some of the micro-bumps will be under tension and will crack easily. Furthermore, warpage will make the alignment of stacked chips difficult, and misalignment will generate shear stress.

2.3. Intermetallic compound (IMC) formation in micro-bumps

The diameter of via in TSV is 20 μm and is expected to be reduced to 10 μm soon. So the diameter of solder micro-bumps to join the Cu in the vias will be 20–10 μm. On the other hand, the temperature and time of post-soldering aging of the micro-bumps remain the same as those used for flip chip solder joints. Under the same processing condition, the micro-bumps will be transformed completely from Pb-free solder to Sn-based IMC such as Cu₆Sn₅, Cu₃Sn or Ni₃Sn₄, or the ternary IMC of Cu–Ni–Sn. IMC joints are brittle. Besides, we do not have enough data on the reliability of IMC.

The above causes will lead to serious reliability concerns in 3D IC technology. We cannot say that these are old reliability problems and we already know how to fit them by taking into account the reliability concerns in the device design stage, so to have the concerns under control. This is because the packaging structure in 3D IC is different from that in 2D IC, especially the vertical interconnects. Without a systematic and thorough study, we will not have the confidence needed in a large scale applications and mass production of 3D IC in consumer electronic products.

Indeed, while microelectronic industry has had the paradigm change from 2D IC to 3D IC, the practice will be slow and cautious. At the moment, the most challenging packaging reliability issue comes from chip-packaging interaction due to thermal stress induced cracking of ultra-low k dielectric layer in the multi-layered Cu/ultra-low k interconnect structure, and TSV can serve as an interposer to reduce the problem.

Previously, when thin film of SiO₂ was used as the dielectric, its mechanical property is stronger than that of solder, so in the flip chip technology the thermal stress tends to deform the solder joints but not the dielectric. Yet, as the industry advances by adopting ultra-low k dielectric materials, it has first to overcome the integration issue between Cu and ultra-low k, which is not easy at all. However, when the integration is successful and when the chip having the Cu/ultra-low k interconnect technology is joined by flip chip solder joints to a substrate, the thermal stress between the chip and the substrate tends to crack the ultra-low k dielectric layer, instead of the solder joints. This is because the ultra-low k materials are mechanically weaker and can crack easily. If we introduce a TSV chip as an interposer between the Si device chip and the substrate, the chip-packaging interaction issue will be reduced. This is because the thermal stress between the Si device chip and the TSV chip is much less. No doubt, we still have thermal stress between the TSV chip and the packaging substrate, nevertheless, it will be of less concern as long as there is no active device on the TSV chip as the interposer.

3. Joule heating

On the basis of irreversible processes, joule heating is entropy production in electric conduction [17]. Usually, the power of joule heating is written as

\[
P = I^2 R = \dot{f} \rho V
\]
where \( I \) is applied current, \( j \) is the current density \( (j = I/A) \), \( A \) is the cross-sectional area of the conductor, \( R \) is resistance of the sample \( (R = \rho l/A) \), \( \rho \) is resistivity, and \( l \) is the length of the sample, so the volume of the conductor \( V = Al \). Thus, \( I^2R \) is joule heating per unit time of the entire sample, and \( j^2 \rho \) is joule heating per unit volume per unit time of the sample and its unit is \( J/cm^3s \). The unit of power is \( \text{watt} = \text{joule/s} \).

On a Si chip, the elements which generate heat are the transistor, the contact metallization, the multi-layered Cu and Al interconnect, and the solder joints. In a mainframe computer or a server, the heat will be conducted away with an elaborated cooling system using water and liquid He or liquid N in an air-conditioning room. But for consumer electronic products such as hand-held wireless devices, typically a thermal interface material (TIM) will be inter-posed between the back-side of the Si chip and a piece of metallic plate or the metallic case of the hand-held device to conduct the heat away.

On 3D IC structure, it is not heat generation, rather it is heat reduction is challenging. Heat generation can be bigger by a factor of the number of chips in stacking, yet heat dissipation becomes much more difficult. If we consider TIM, it helps the top piece of the number of chips in stacking, yet heat dissipation becomes reduction is challenging. Heat generation can be bigger by a factor of metallic plate or the metallic case of the hand-held device to conduct the heat away.

In Fig. 2, if we assume the diameter of a solder bump is 100 \( \mu \)m and the spacing is also 100 \( \mu \)m, we can have 2500 bumps/cm\(^2\). Yet only power bumps which will carry current by design, and most signal bumps will carry very little current. If we assume there are 100 power bumps, each has a current density \( j = 1 \times 10^5 A/cm^2 \) and \( \rho = 2 \times 10^{-5} \Omega cm \) for SnAgCu solder, we have \( \rho j^2 = 10^3 J/cm^3s \). The total volume of the 100 power solder bumps is \( V = 100 \times (10^{-2} cm)^3 = 10^{-4} cm^3 \). Thus \( P = \rho j^2 V = 0.2 J/s = 0.2 W \). To include uncertainty in the calculation, we may say that the joule heating from solder joints is in the order of 1 W per chip. While the heat is much lower than that of the transistor and interconnect metallization, it affects locally the underfill surrounding the solder bumps. The polymer-based underfill has a low glass transition temperature, so the joule heating may change the viscosity of the underfill. The flow of the underfill will reduce its role in the protection of the bump as well as the chip.

We can make a similar calculation of the contact metallization and the Cu vias in TSV, and they will not be large. Therefore, most of the joule heating in a Si VLSI device is from the transistors. Joule heating is a kind of waste heat, it cannot be used to do work, however, it generates heat and will increase the temperature of the conductor. Knowing the heat capacity of the conductor, we can calculate the temperature increase, yet the real temperature change depends on heat dissipation. To know the real temperature change, we need temperature sensor and modeling of heat dissipation. Since the resistance of the conductor increases with temperature, it in turn will cause more joule heating. An increase of temperature of the conductor will also lead to thermal expansion of the conductor, and this is the basic reason of thermal stress in devices when various materials of different thermal expansion coefficients are used. Stress and stress potential gradient may lead to fracture and creep, respectively. When the thermal stress is cyclic because semiconductor devices are being turned on and off frequently, fa-
Elastic limit at 0.2%, Cu–Sn or Ni–Sn IMC is close to that of Al, we expect that grain growth occurs quickly on the cathode contact on the chip side. When it reaches $1 \times 10^4$ A/cm², electromigration failure will occur in a few hours. While thick Cu UBM and Cu column bumps have been used to reduce the effect of current crowding, electromigration can enhance the dissolution of Cu into solder and lead to failure.

When the diameter of a micro-bump is 10 μm, the remaining thickness of solder will be a few microns and it is within the critical thickness of electromigration, where the back-stress induced by electromigration may be large enough to counter-balance electromigration. This has been found to be true in Al interconnects, and we have

$$j \Delta x = \frac{Y \Delta e \Omega}{Z e \rho}$$

where $j$ is current density, $\Delta x$ is the critical length of the sample, $Y$ is Young’s modulus, $\Delta e$ is elastic strain and we can take it to be the elastic limit at 0.2%, $\Omega$ is atomic volume, $Z$ is the effective charge number of electromigration, $e$ is electron charge, and $\rho$ is resistivity. Since the parameters on the right-hand side of Eq. (2) are known for Al, Cu, and Sn (or Pb-free solder), we can calculate the critical product, $j \Delta x$, to be about 1000 A/cm for Al and Cu, and the critical product of Sn to be about two orders of magnitude smaller. This is why electromigration can occur in solder joints with a two orders of magnitude lower current density than that in Al and Cu interconnects, as discussed in the last paragraph, when we assume the critical length to be similar at 10 μm.

The back-stress occurs in Al interconnects due to its protective oxide. Since Sn also has a protective oxide, we expect back-stress can occur in Sn solder joints. Indeed Sn whiskers have been squeezed out at the anode end in flip chip solder joints of 100 μm in diameter [18]. However, in the vertical interconnect of TSV and micro-bumps, there are reactive interfaces on both sides of the Sn. If the interfaces are effective vacancy sources and sinks, the effect of stress induced by electromigration will be greatly reduced. We recall that there is no reactive interfaces on Al interconnects. Hence, whether or not a critical length of electromigration can be found in micro-bumps requires a careful study.

The rest of the solder joints in between TSV wafers have no current crowding because the current distribution is more uniform, so they should fail more or less like the solder joints in Cu column bumps. However, if temperature increase due to joule heating cannot be reduced, the effect of high temperature will enhance the failure. Furthermore, if a higher current density is used in TSV samples, electromigration will be faster. When vacancies are accumulated at the cathode side of the joint, the mechanical properties of the cathode interface is weak.

When the micro-bump is transformed completely to IMC, we must consider electromigration in IMC. Since the melting point of Cu–Sn or Ni–Sn IMC is close to that of Al, we expect that grain boundary electromigration will occur in IMC provided that the current density is about $10^3$ A/cm². Whether Sn or Cu (or Ni) is the dominant diffusing species in electromigration is unclear. We might find phase transformation, for example, from Cu₃Sn to Cu₅Sn or vice versa, to accompany electromigration. In forming Cu₃Sn, Kirkendall (or Frenkel) void formation due to Cu diffusion may occur. Besides electromigration, we expect thermomigration to occur under a large temperature gradient. Again thermomigration in IMC is new.

On cyclic thermal stress induced fracture, its effect on the bottom row of solder joints, i.e., the row between the last chip and the substrate, could be more serious than the rest since these solder joints are between a chip and a substrate due to different thermal expansion, yet the solder joints are much large in diameter, so the thermal stress issue may not be more serious than those micro-bumps in between chips because of chip warpage.

## 5. Warpage of TSV chips

When a chip is under tension, no warpage will occur. Warpage occurs under compression. This is well known, for example, in etching an oxidized Si chip to have a SiO₂ window and the window buckles. When the chip is thin and the arrangement of arrays of TSV is asymmetrical, warpage occurs. How can we avoid warpage or the compressive stress in TSV chips is a challenging issue. Since the vias in a TSV chip are arranged non-uniformly and owing to the different thermal expansion between the Si matrix and the Cu column in the vias, it can lead to compressive stress in the chip so the chip will buckle. We need to study and to simulate the stress generation and relaxation in TSV in order to understand the stress–strain evolution in processing a stack of TSV chips.

We can use synchrotron radiation (SR) micro X-ray diffraction to measure the strain in a TSV chip before and after thermal cycling [19,20]. The curvature can be determined by the diffraction, and we can calculate the bending stress. The correlation to fracture in micro-bumps can be made as a function of cyclic thermal stress. The curvature can lead to tensile stress and fracture in some of the micro-bumps. We may also use impact and drop test to study the fracture toughness of the micro-bumps in TSV samples and follow it by SR X-ray tomography examination.

## 6. Thermal stress in cu, W, or doped poly-Si vias

The thermal expansion coefficients of Si, Cu and W are $2.6 \times 10^{-6}$/C, $16.6 \times 10^{-6}$/C, and $4.5 \times 10^{-6}$/C, respectively. From the point of view of difference in the coefficients, clearly if we fill the TSV with heavily-doped poly-Si, we have much less thermal stress, as compared to that of Cu and W. Yet the electrical resistivity of heavily doped poly-Si is three orders of magnitude higher than that of Cu, so joule heating will be an issue as we have discussed in Section 3. If we cannot have the heat removed effectively, Cu vias are better. Furthermore, to join poly-Si vias together, we will need to prepare silicide contacts first before we can use micro-bumps. About W vias, it is more rigid. From the point of view of electromigration, the wear-out mechanism in Al interconnects with W vias will come back on the top piece of the chip stack.

Mechanical strength of TSV chip is an issue when the thickness of the chip becomes too thin. Its strength depends on the design of the array of TSV and tends to decrease with increasing density or number of vias. When it is weak, it fractures easily.

## 7. IMC formation in micro-bumps

The Cu columns in TSV are joined by micro-bumps between chips. The diameter of the micro-bumps is about 10 μm now.
Fig. 6 shows a focused-ion-beam polished cross-section of the metallization in a micro-bump. If a post-soldering aging at 150 °C for 1000 h is performed, it will transform the entire joint into IMC [13,14]. Reliability if IMC joints is mostly unknown. At the least, they are brittle.

8. 3-D imaging of failure in TSV technology

We can examine the 3-D structure by cross-sectioning using SEM and FIB. We can also use synchrotron radiation X-ray tomography to examine the structure and its reliability issues before and after failure. Fig. 8 shows an SR X-ray tomography image of an array of flip chip solder joint. The diameter of the solder bumps is 200 μm, and the second bump on the upper left corner has failed by electromigration. Fig. 9a and b shows SR X-ray tomography images of the top view and side view, respectively, of a pancake-type of void growth in a flip chip solder joint under electromigration. We will be able to measure the total volume of the void and in turn its rate of growth or propagation under electromigration. Knowing the total volume, V, we have

\[
V = \Omega j A t = \Omega \left( C \frac{D}{kT} Z e j \right) A t
\]

where \(\Omega\) is volume of a vacancy, \(j\) is vacancy flux and it has the unit of \(\#\) of vacancy/cm² s, and \(A\) is cross-sectional area of the void where the void growth occurs, and \(t\) is the time of growth. In other words, if we divide the void volume by vacancy volume, we obtain the number of vacancies needed to grow the void per unit area per unit time. Knowing the flux of vacancies, we will be able to calculate the driving force of electromigration, where \(J = C(D/kT)F\), \(C\) is concentration of vacancy, \(D\) is vacancy diffusivity, \(kT\) is thermal energy, \(F\) is the driving force of electromigration \((F = Z e j)\), \(Z\) is the effective charge number in electromigration, \(\rho\) is resistivity, and \(j\) is current density.

9. Summary

Vertical interconnect by using TSV and micro-bump is the new technology in 3D IC. The most serious reliability concern in 3D IC is joule heating. How to remove the heat is the most challenging problem. In order to remove heat, a temperature gradient must exist in the packaging and a large temperature gradient can lead to thermomigration. Equally challenging reliability issues are electromigration, stress-migration, and chip warpage under compressive stresses. Complete IMC formation in micro-bumps will affect the fracture toughness of the brittle joints.

Acknowledgment

The author would like to thank Dr. Yi-Shao Lai at ASE, Taiwan for helpful comments.

References

[14] Lai Yi-Shao, Tong Ho-Ming, Tu KN. Edited a special section on recent research advances in Pb-free solders. Microelectron Reliab 2009;49:221–322.