Design and Reliability in Wafer Level Packaging

Xuejun Fan1,2, Qiang Han1
1,2Department of Mechanics, South China University of Technology, Guangzhou, P.R. China
1Department of Mechanical Engineering, Lamar University, PO Box 10028, Beaumont, Texas 77710, USA
xuejun.fan@lamar.edu

Abstract

Wafer Level Packaging (WLP) has the highest potential for future single chip packages because the WLP is intrinsically a chip size package. The package is completed directly on the wafer then singulated by dicing for the assembly. All packaging and testing operations of the dice are replaced by whole wafer fabrication and wafer level testing. Therefore, it becomes more cost-effective with decreasing die size or increasing wafer size. However, due to the intrinsic mismatch of the coefficient of thermal expansion (CTE) between silicon chip and plastic PCB material, solder ball reliability subject to temperature cycling becomes the weakest point of the technology. In this paper some fundamental principles in designing WLP structure to achieve the robust reliability are demonstrated through a comprehensive study of a variety of WLP technologies. The first principle is the ‘structural flexibility’ principle. The more flexible a WLP structure is, the less the stresses that are applied on the solder balls will be. Ball on polymer WLP, Cu post WLP, polymer core solder balls are such examples to achieve better flexibility of overall WLP structure. The second principle is the ‘local enhancement’ at the interface region of solder balls where fatigue failures occur. Polymer collar WLP, and increasing solder opening size are examples to reduce the local stress level. In this paper, the reliability improvements are discussed through various existing and tested WLP technologies at silicon level and ball level, respectively. The fan-out wafer level packaging is introduced, which is expected to extend the standard WLP to the next stage with unlimited potential applications in future.

1. Introduction

Wafer level packaging (WLP) is one of the fastest growing segments in semiconductor packaging industry due to the rapid advances in integrated circuit (IC) fabrication and the demands of a growing market for faster, lighter, smaller, yet less expensive electronic products with high performance and low-cost packaging [1]. Its primary advantage today is its small form factor because WLP is intrinsically a chip size package. Another advantage is its low packaging cost, compared to die-level packaging (e.g. wird-bond type packaging). With die-level packaging, the packaging cost has always become a greater percentage of IC cost as the die size decreases or the wafer size increases. The packaging cost per wafer in WLP remains a relatively constant percentage of the total IC cost, which means it can become more cost-effective with decreasing die size or increasing wafer size. Because of these advantages, many IC makers are incorporating WLP into their designs. As such, a paradigm shift to wafer level packaging is apparent [2].

Typical semiconductor devices utilizing WLPs today are limited to a lower number of I/O per die at 0.5mm pitch or less. There are two reasons for this. The smaller a given component is, the higher the associated die count per wafer. As mentioned above, WLP technologies by their very nature result in a lower cost per die (vs. traditional wirebond) when the die count per wafer is high. The second reason is that as the number of I/O per die increases (and thus the Distance to Neutral Point increases), the WLP may not achieve prescribed solder joint reliability requirements.

Solder joint thermal cycling reliability of standard WLPs is the weakest point of the technology. The ability of the solder joint to survive the required thermal cycle testing has limited WLP to the products having relatively small die sizes and a small number of I/O. However, there has been a demand for a WLP package with more integrated functionality, which will result in relatively larger number of I/Os and a larger size of die. In order to improve solder joint reliability (without underfill), new WLP technologies are needed to be developed.

Fig. 1 is a typical standard WLP structure, which is similar to a flip chip package on substrate without the presence of underfill. WLP has the simplest structure over all other types of IC packages since a WLP (on board) composes of only three major elements: silicon die, solder ball and the PCB. The difference in the coefficient of thermal expansion (CTE) between silicon (~2.6ppm/°C) and PCB (~17ppm/°C) determines that the solder ball thermal cycling fatigue performance will be limited. In order to resolve this issue, a variety of WLP technologies have been devised, among which two categories can be identified, fan-in WLP and fan-out WLP technologies, respectively. For the category of fan-in WLP technologies, there are several types in terms of processes and designs, such as ball on I/O, ball on polymer, and encapsulated Cu post WLP, etc. The fan-out WLP technologies have extended the standard WLP to the next stage. Fig. 2 is schematic of the horizontal view of the comparison between a standard fan-in WLP and a fan-out WLP [3,4].
In this paper, an overview of a variety of WLP technologies will be presented, with the emphasis on the fundamental principles to improve the solder ball reliability by introducing various designs. The paper will start from various designs at silicon level, in which several existing technologies are described to illustrate the improvements on the solder ball reliability. Then the considerations at ball level to achieve the better reliability will be discussed, which include ball shape optimization, polymer-core solder ball technology, and double-bump WLP technology.

2. Silicon Level Design Considerations

2.1 Ball on I/O - Standard Flip Chip

Ball on I/O WLP is a standard wafer level packaging technology and the process is very similar to a typical flip chip technology. The standard flip chip (SFC) process utilized by FlipChip™ International [5], formerly known as the Flex-on-Cap (or FoC) process, was created in the mid-1960's by Delco for use in the automotive industry. In this process, the ball is attached to the aluminum pad directly through a standard under bump metal (UBM) structure, such as shown in Fig. 3.

On some die, the I/Os are not located where one needs to have the balls. This is especially true when you take an existing die that is wire bonded and would like to convert it to WLP. The Redistribution Line (or RDL) process adds "redistribution metallization" (often called "runners' or "traces") that let you re-route the signal path from the die peripheral I/O to the new desired bump locations. This process is usually seen as a transitional solution between a die that is designed for wire bonding and a die that is designed for flip chip. One of those structures is the bump-on-nitride (BON) structure, consisting of solder bump and under bump metal (UBM) seated on the thin inorganic passivation as shown in Fig. 4 [2].

Although two structures between ball on I/O and ball on nitride with RDL are different, the added redistribution layer and the passivation layer in Fig. 4 do not provide stress buffer effect to the solder ball. The solder balls in these two structures are connected to the silicon base through ‘hard’ materials such as aluminum pad or nitride passivation. When thermal cycling takes place, the solder balls in both cases will support all shear stresses caused by the thermal mismatch between silicon and PCB material. As a result, reliability performance would not have significant difference between these two cases. For a standard WLP design, many studies have shown that the package is limited to 5x5 or less area arrays at 0.5mm pitch [6]. Beyond this limit, the thermal cycling performance requirement will not be met.

2.2 Encapsulated Cu Post WLP

In order to further improve the WLP solder ball reliability and go beyond the limit of the standard WLP design, one of tested technology is so-called encapsulated Cu post WLP. Fujitsu has a trademark name SuperCSP [7]. Casio Computer Co. Ltd. has licensed its Cu post embedded wafer-level package (EWLP) technology to Renesas Technology. The manufacturing process of a Cu Post WLP involves the formation of the redistribution layer and a polyimide film and electrolytic plated metal trace. After redistribution, the resist is patterned and Cu post is formed by electrolytic plating. Then the whole wafer is encapsulated with an epoxy molding compound (EMC). Fig. 5 shows the process steps based on the Fujitsu SuperCSP technology [7]. Fig. 6 is a picture of the redistributed wafer with Cu post before molding based on CASIO WLP technology [8].
When compared it to the Fig. 3 for a standard ball on I/O WLP structure, there are two major differences:

1. The ‘molded wafer’ increases the overall coefficient of thermal expansion of the silicon, and therefore, the thermal mismatch with respect to the PCB is reduced. In other words, the thermal stresses can be released with Cu post structure, which acts like a stress buffer layer between silicon base and solder ball. When the component is subjected to thermal cycling, some energy will be absorbed by the composite structure of Cu post and epoxy. Such a composite layer increases the flexibility of the whole WLP when temperature cycling is applied. This reveals a fundamental principle that in order to reduce the stresses over solder joints: a more compliant or more flexible WLP structure will make the solder balls more robust against fatigue failure. When the whole WLP structure becomes flexible, the less stress will be applied on the solder balls.

2. The Cu post or epoxy height is about 70um to 100um. This can be considered to virtually increase the height of solder balls. Increasing solder ball height will make whole WLP structure more flexible, therefore the thermal cycling performance can be improved.

Finite element analysis (FEA) has been performed to analyze the solder joint stress and creep deformation. The energy dissipation on the critical layer close to the package side is used as a parameter to evaluate the damage of solder balls under temperature cycling [9]. Fig. 8 plots the normalized energy dissipation of the critical solder ball under thermal cycling from -40°C to 125°C for both standard ball on I/O and Cu post WLPs, respectively. The ball geometry, solder opening size, and other package dimensions are the same for both cases. It can be seen that with Cu post structure, the energy dissipation on the ball is reduced by 50% compared to a ball on I/O structure. Experimental results also showed the excellent reliability results as shown in Table 1 [7]. Overall, with the Cu post structure, the array size can be extended to 12x12 with 0.5mm pitch from the original limit 5x5 area array size. Such an improvement clearly demonstrates that a ‘stress buffer layer’ at silicon level can greatly improve the package reliability due to the improvement of overall flexibility of the WLP structure.

2.3 Ball on Polymer

FlipChip™ International has developed a so-called Spheron WLCSP technology. Spheron WLP incorporates a ball on polymer (BOP) structure, as shown in Fig. 9 [2,5]. In this structure the solder ball is placed over a layer of polymer dielectric material. A new dielectric material has been utilized, which offers improved capacitance, reliability, and better compatibility to polyimide passivation verses BCB. In addition to the advanced polymer, Spheron WLP™ incorporates a new metal structure, which offers improved strength and electrical performance.

Fig. 8 Energy dissipation comparison for Cu post WLP vs. standard ball on I/O

<table>
<thead>
<tr>
<th>Test</th>
<th>Test Condition</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature Cycle</td>
<td>-55°C to 125°C</td>
<td>N=20</td>
</tr>
<tr>
<td>Pressure Cooker</td>
<td>+121°C/85% RH</td>
<td>N=10</td>
</tr>
<tr>
<td>Bending</td>
<td>15mm</td>
<td>OVER</td>
</tr>
<tr>
<td>Cyclic Bending</td>
<td>50micron</td>
<td>OVER</td>
</tr>
<tr>
<td>Free fall</td>
<td></td>
<td>OVER</td>
</tr>
</tbody>
</table>

Table 1 Reliability performance of SuperCSP [7]

There is a fundamental difference between the ball on polymer of Fig. 9 and the ball on I/O with RDL in Fig. 4. In Fig. 9, there is no ‘hard’ connection between solder ball and silicon base. Instead, the solder balls sit on a ‘soft’ dielectric film layer, which will act as stress buffer. When thermal cycling is applied, the BOP WLP is more compliant since the soft dielectric film layer will absorb some energy. As a result, the stresses applied to the solder balls will be reduced greatly. Fig. 10 showed the Weibull plot of both Ball on I/O and Ball
on polymer WLPs [2]. It can be seen that the reliability can be improved significantly.

Through the acquisition of Unitive®, Amkor’s wafer level CSPs incorporate thin film redistribution films to route I/O pads to JEDEC / EIAJ standard pitches, and thereby, avoiding the need to redesign legacy parts for CSP applications. It is available in three options: Direct Bump on Pad, Bump on Repassivation/Redistribution and Bump on Thick Repassivation/Redistribution [10]. The latter two options have a similar concept to the ball on polymer to improve the reliability. It has been reported that with the ball on polymer structure [10], the array size can be extended to 12x12 with 0.5mm pitch. This illustrates that a compliant film layer over solder balls to avoid balls connected to silicon directly could dramatically influence the WLP reliability. It again reveals that making the WLP structure more flexible and compliant will improve the solder ball reliability.

There are several different bump-on-polymer technologies and terminologies, as shown in Figs. 11-13, such as bump on thick repassivation/redistribution (Fig. 11), bump on RDL (Fig. 12), and bump on silicone bump (Fig. 13). Finite element analysis can be used to analyze the fatigue performance for each case.

2.4 Polymer Collar WLP

An alternative way to improve the solder ball reliability is to reduce the local stress concentration of the solder ball without changing the overall compliance of the structure. Polymer Collar™ WLCSP is one of these examples [6], which is an enhancement to die bumped with the standard ball on I/O process. In this process, a ring of polymer reinforcement material is added to the base of the bumps during processing. Since the interface region between bump and UBM receives the most stress during the device lifetime, the addition of the Polymer Collar reduce the local stress concentration and will improve solder joint reliability. Fig. 14 is the process flow for polymer collar WLP, and Fig. 15 shows the pictures of solder balls with polymer collar [6].
As shown in Figure 16, the Polymer Collar WLP had a 89% increase in Weibull life vs. the standard WLP. This improvement is higher than the 50 and 64% achieved during the previous two tests and it is reasonable to conclude that the Polymer Collar is very repeatable in producing more than a 50% increase in solder joint life in the three consecutive TC tests. This example demonstrated the second principle in improving the WLP package reliability, i.e., the local enhancement at failure location of solder balls will reduce the solder ball stress concentration, and the overall compliance of the package will not be affected.

Similar to the Infineon’s fan-out WLP, Freescale named as redistributed chip packages (RCP) [11]. Fig. 19 demonstrates the RCP process flow. Singulated die are placed with active side face-down on a substrate. The die are then encapsulated with a silica-filled epoxy molding compound. After cure, the panel of the die is then ground to the desired panel thickness and then released from the substrate. The die panelization process is done with standard assembly tools such as a pick and place tool. The epoxy panel with die then undergo a redistribution process to route out the signals, power, and ground. The redistribution process is done with standard silicon manufacturing equipment. These processing steps consist of the deposition of copper metallization layers by electroplating techniques. The metal layers are separated by insulating layers comprised of a spin-coated, photoimageable dielectric, and patterned using batch process lithography. The metal layers connect the pads on the die surface to the pads placed on the surface of the package, providing the same function as the metal layers in the substrate of a ball grid array (BGA), but with a much finer resolution.

The fan-out WLPS are expected to have superior reliability subjected to thermal cycling since ‘molded die’ and redistribution layer will make the WLP more compliant and flexible. More detailed analysis on the reliability of fan-out WLPS will be reported in a separate paper.
3. Ball Level Design Considerations

3.1 Ball Shape Optimization

In order to make the WLP structure as flexible as possible, something can be done at ball level by selecting different solder materials or by optimizing ball shapes. It has been well-known that increasing ball height will enhance the solder ball thermal cycling reliability greatly [e.g. 9]. This is because that the higher the ball is, the more compliant the WLP is. When the solder ball opening size is fixed, the optimization of solder ball shape and the volume will have a great impact on solder joint reliability. Fig. 21 shows three scenarios of the solder ball shapes. The larger solder ball diameter means more solder volume, which may provide stronger support for solder ball against fatigue failures. However, one should be aware that, with the fixed solder ball height and solder ball opening size, the greater the ball diameter is, the less flexible the whole structure. Finite element analysis has been performed on those three scenarios and the results of energy dissipation were tabulated in Table 2. It clearly showed that the convex shape solder ball (more solder volume with a larger diameter) will result in the worst solder joint reliability performance. This is again confirms that less flexible structural design will negate the overall solder joint fatigue performance.

<table>
<thead>
<tr>
<th>Legs</th>
<th>Solder ball diameter (um)</th>
<th>Normalized energy dissipation</th>
<th>Fatigue life</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>260</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>310</td>
<td>1.00</td>
<td>↓</td>
</tr>
<tr>
<td>3</td>
<td>360</td>
<td>1.19</td>
<td>↓</td>
</tr>
</tbody>
</table>

On the other hand, however, similar to polymer collar effect, if local stress concentration or stress singularity can be minimized, the solder ball fatigue performance can be improved too. When the solder ball diameter and ball height are fixed, increasing the solder ball opening size will provide stronger support and reduce the overall local stress concentration at the interface region of solder ball and silicon chip.

3.2 Plastic Cored Solder Balls in WLPs

A plastic core solder ball consists of a large polymer core coated by a Cu layer and covered with eutectic and/or lead-free solder. The main advantages of such a system are higher reliability due to the relaxing of stress by the polymer core and a defined ball height after reflow [12-15]. These balls could improve the solder ball reliability significantly due to the compliant feature of balls. Fig. 22 is a schematic of horizontal view of polymer core ball. Fig. 23 shows the photos of the plastic core solder balls. Table 3 shows the comparison of Young's modulus of polymer core material and SnPb. It can be seen that polymer core material has only one-tenth of the modulus of SnPb, which will make the structure more flexible when subjected to temperature cycling.

![Fig. 22. Cross-section structure of polymer core ball](https://example.com/fig22.jpg)

Fig. 22. Cross-section structure of polymer core ball

![Table 2 Normalized energy dissipation for three scenarios in Fig. 21](https://example.com/table2.png)

![Fig. 21. Three scenarios of solder ball shapes with the fixed solder ball opening size and ball height](https://example.com/fig21.jpg)

Fig. 21. Three scenarios of solder ball shapes with the fixed solder ball opening size and ball height

2008 10th Electronics Packaging Technology Conference
Fig. 23. Real images of plastic core solder balls on WLP [15]

Table 3 Material properties for polymer core and SnPb solder

<table>
<thead>
<tr>
<th>Material</th>
<th>Young’s modulus (GPa)</th>
<th>Poisson’s Ratio</th>
<th>CTE (ppm/°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plastic Core</td>
<td>4.7</td>
<td>0.38</td>
<td>40.2</td>
</tr>
<tr>
<td>Solder (SnPb)</td>
<td>40.2</td>
<td>0.4</td>
<td>24.7</td>
</tr>
<tr>
<td>Substrate Board</td>
<td>24.5</td>
<td>10.8</td>
<td>14</td>
</tr>
</tbody>
</table>

Fig. 24 shows some new failure modes using plastic cored solder balls under temperature cycling. The failure modes vary somewhat according to the geometry of the solder surrounding the polymer-core ball. Fig. 25 is the Weibull plot of temperature cycling results for the comparison of polymer cored ball and conventional solder balls. It clearly indicates that the plastic core solder ball structures improve the fatigue life significantly. The Young’s modulus of plastic core is much smaller than that of metal, so it is superior in relaxing the thermal stress and enables much greater resistance to micro-cracks in the solder layer than that of conventional solder balls. In addition, it is superior in holding the gap flatly between upper and lower substrates, because the particle diameter of polymer core is very uniform.

3.3 Double-Bump WLP Technology

There are several versions of double-bump WLP technologies [16-18]. Such a WLP has been designed and fabricated by using an encapsulation material, which is applied directly to a bumped wafer, thereby eliminating the underfill process, and protecting all the bumps on the wafer at once in a batch process [16]. This material was designed to have the necessary elastic modulus and coefficient of thermal expansion required by this application. After application of the encapsulation, the wafer is then bumped again, creating a double bump structure that increases the overall bump height to improve the reliability further. Redistribution of bond pads from the die periphery to an area array using dielectric film and redistribution metal aids in eliminating the need for an interposer. Fig. 26 is a schematic of double-bump structure, and Fig. 27 is the real cross-section image from a double-bump WLP [18].

This wafer level package technology has been evaluated using a test vehicle, which has a 0.5 mm pitch of an 8×8 array.
of bumps on a 5×5 mm² die. Micro Moire Interferometry has shown that the encapsulation layer facilitates the distribution of stress throughout the wafer level bumps. The bump structure and package geometry have been optimized using simulation and validated by experimentation to insure contact between the encapsulation and first level bump, which is key to reducing stress and improving reliability. Fig. 28 showed the reliability results under thermal cycling. It indicated that the double-bump WLP had much better reliability than standard WLP packages. Double-bump structure is similar to Cu post structure, which makes the overall silicon CTE increased and overall solder ball height increased. Therefore, the reliability performance is improved.

Fig. 28. Reliability performance of double-bump WLP vs. standard WLP

4. Conclusions
Wafer level packaging structural design has great impact on the package reliability. Solder ball temperature cycling reliability of WLPs is the weakest point of the technology due to the intrinsic feature of the CTE mismatch between silicon die and plastic PCB. In order to improve the solder ball reliability, the WLP structure should be designed as flexible as possible so that the stresses on the solder ball can be minimized. This is the first principle that should be followed. The second principle in designing a WLP structure is to reduce the local stress concentration without changing the overall structural compliance. Both principles can be realized through silicon back-end designs at silicon level or through ball designs at ball level. Same principles can be applied at PCB level. For example, lowering PCB CTE or adding stress buffer layer in PCB will also enhance the solder ball reliability.

References
8. Casio Embedded Wafer Level Packaging (EWLP), files.ipc.org/epug/2006_02/Casio%20EWLP%2002%2006.ppt
10. Amkor’s Wafer Level Packaging, http://www.amkor.com/Products/all_products/cspnl.cfm

2008 10th Electronics Packaging Technology Conference