1. Introduction

The latest ultra-mobile electronics systems including smartphones and mobile internet devices contain a variety of thin, small and high density packages including the ball grid array (BGA), package-on-package (PoP), system-in-package (SiP), stacked die chip scale package (CSP) and wafer level packages.[1, 2] Designers have selected the optimal packaging technologies by performing the appropriate tradeoffs with the system performance requirements, dimensional requirements, manufacturability and system cost. What will be the package of choice in the future to meet the continuing demand for higher density and lowest cost?

Flip chip on board, or direct chip attach (DCA) on a printed wiring board (PWB) is becoming increasingly interesting as a possible system assembly solution. DCA contributes to the system miniaturization by having a higher number of input/output (I/O) pin counts in a chip size package with the benefit of lower parasitic interconnects. By comparing with CSP or BGA, DCA does not require a packaging substrate, so it enables a thinner system and lowers the package cost. If the system designer can optimize the DCA bump pitch to the motherboard design rules, the motherboard size can be decreased which can lower the overall system cost. For smartphones, consumer electronics and embedded devices that value an ultra-small form factor, DCA will bring new business opportunities for SOC applications.

Currently, the area array DCA flip chip is not the package of choice for logic devices to replace CSP or BGA packages. This is because current flip chip assembly technology can be difficult to master by custom assemblers or Original Design Manufacturers (ODM). Flip chip requires a special set of processes, materials, tools and knowledge of silicon that most customers are missing. For example, the interconnect bump pitch for logic flip chip device is less than 0.25 mm pitch, while an advanced BGA for the smartphone has only a 0.4 mm pitch. The existing PCB assembly capabilities for motherboards will not support such fine pitch.

Furthermore, advanced silicon technology requires a great deal of assembly integration for the process, material and design, particularly for the co-design of silicon and packaging. For example, today’s advanced silicon has low-k interlayer dielectric (ILD), strained silicon transistors, high-k metal gate, thick metal global wiring on the silicon surface, and lead-free interconnect technologies.[3] Significant effort is required for each product and silicon generation to enable the advanced silicon process and assembly technologies to co-exist.[4, 5]
This paper will discuss a proposed WLP structure and process, including front side wafer molding with exposed copper bumps, die back side film (DBF) and micro balls for the interconnect. A photograph of a prototype WLP structure is shown in Figure 1 and a cross-sectional view and layer structure of the proposed WLP is shown in Figure 2. WLP technology can provide the benefit to our customers of better mechanical handling, lower assembly stress on the silicon interconnect, improved die level warpage and the potential opportunity to simplify the board assembly process. WLP technology will enable a DCA flip chip process for advanced logic silicon devices using a more customer friendly process.

2. WLP Processing

2.1 Process target

The WLP process consists of six modules (i.e., front side mold, copper bump exposed process, micro solder ball attach, back grinding, die back side film lamination and saw singulation). Figure 3 shows the proposed process flow and process options. An experimental strategy was defined to narrow down process options, followed by the development of the full loop process for the WLP process demonstration. The test vehicle for this process proof of concept was a daisy chain metal only 45 nm node silicon chip in a 300 mm wafer format. Die size was 9.5 × 9.5 mm², die bump construction was copper columns with 47 μm height and the minimum die bump pitch was 227 μm.

The thickness targets for the WLP process development were 200 μm, 100 μm, and 75 μm to understand the process capability. The motivation for thinning the WLP package was to lower the board assembly profile. The 200 μm silicon thickness meets the 300 μm height of passive devices specified by the Electronics Industry Association (EIA) 0201 form factor (0.6 × 0.3 × 0.3 mm). A 100 μm die thickness would target the 200 μm assembly height of the EIA 01005 spec (0.4 × 0.2 × 0.2 mm).

2.2 Mold technology

We selected compression mold technology[6] for wafer front side mold to enable thin and uniform mold formation with good adhesion and flexible material selection. Figure 4 indicates the outline of compression mold tool and process flow.

The mold material was delivered onto the bottom of the molded die while in 300 mm wafer format and the wafer was immersed in this epoxy pool in a vacuum condition. Mold pressure was then applied at an elevated temperature. By design, there was no mold flow during the compression process. This helps achieving a uniform mold thickness, and excellent filler material distribution. The equipment can place a release film between the mold chase and the epoxy resin, allowing use of mold material which is free of releasing agent, thus enabling mold material with good adhesion properties. The mold condition used in this experiment was 7 kgf/cm² at 175 deg.C.
The mold material shown in Table 1 has a relatively lower modulus than other materials evaluated for wafer mold or encapsulation to reduce the wafer warpage.

An exposed bump mold process was studied to simplify the processing. The process used a release film between the wafer and the mold chase. After clamping the mold chase, the wafer was dipped in the epoxy and then top of copper bump was pressed against the release film. The mold material was extruded by the mold pressure and then cured, so the copper bump was exposed when the mold epoxy release film was removed. Figure 5 shows the bump exposed mold process.

Thin wafer mold capability was required in the WLP process steps such as the wafer back side encapsulation, but the handling of a thin wafer during compression mold was not well understood. For this gross reality check, we used blank wafers with thickness of 760 μm, 400 μm, 200 μm, and 100 μm to evaluate the wafer warpage after mold.

2.3 Front side mold grinding

The alternative to the bump exposed mold process was grinding back the mold surface after the front side wafer mold. The initial mold thickness was designed to be 100 μm thick and was then ground down to 40 μm to expose the top of the copper bumps. The final mold thickness and copper bump height was defined by the front side grinding tool capability. The experiment was done using an in-feed surface grinder used for Si back grinding process.

2.4 Micro solder ball mount

Micro ball mount was done with a metal stencil ball alignment and reflow soldering process. The first step was the flux dot printing on the copper pads by aligning the metal stencil. The second step was the ball mount on flux by realigning the ball mount stencil. The micro balls were set on the stencil and then dropped into the stencil cavity by squeegee designed for this application. The micro balls were held by the flux tackiness and the stencil was released. The last step was the wafer reflow in nitrogen atmosphere followed by deflux cleaning.

The micro ball mount flux was designed to have enough tackiness to hold the micro balls on the wafer. This was achieved through characterization of the viscosity and thixotropy. To have good solder wettability on an oxidized copper surface after the front side grinding, the flux activity had to be improved. Figure 6 shows the scanning electron microscope (SEM) image of micro ball wettability test with X-ray analysis. We observed no significant issue with respect to voids. The micro ball we used in this study was 100 μm diameter composed of SAC alloy. The test vehicle for the WLP experiments had 1,050 bumps per die, which resulted in a total of more than 700,000 micro balls per wafer.

2.5 Back grinding and DBF

The back grind tape used was selected to enable the planarization of the 100 μm micro balls on the wafer during the back grinding process. This experiment was also done using the in-feed surface grinder. After the wafer back grinding, DBF was laminated on the wafer back side to protect the die from chipping, cracking and any damage during the handling. Wafer saw was then used for die singulation.

3. Experiment Results

3.1 Compression mold results

Two approaches were evaluated for exposed bump molding using compression molding. Ultimately, exposed bump molding was not chosen. However, the evaluation of work in this area is summarized in the following sections.
3.2 Bump exposed mold

The feasibility of the bump exposed mold process has been demonstrated using a compression mold and release film on 300 mm wafer. Key process issues were identified with the mold thickness uniformity and the mold flash on the copper bumps. Figure 7 shows the mold thickness uniformity data within a wafer. The mold thickness variation within the wafer ranged from 24 μm to 53 μm in thickness with a 40 μm average. The experiment was conducted with a general purpose 300 mm wafer mold chase. The challenge for the bump exposed mold is designing the mold chase to control a very small volume of mold material such as 4.0 grams per 300 mm wafer for a 40 μm target thickness. Figure 8 shows the mold residue on top of the bump after mold. Mold de-flash cleaning using a wet blast technology was shown to be effective.

3.3 Thin wafer mold capability

The wafer thickness significantly impacts the wafer curvature after mold and we found 200 μm thickness of wafer is the process cliff. Figure 9 indicates the wafer curvature at different wafer thickness. Although a full thickness wafer is very flat after mold, a 200 μm wafer changed to 18.0 mm curvature, and less than a 200 μm wafer had very significant curvature or even wafer cracking after mold and after post mold cure.

However, wafer molding at full thickness with subsequent wafer thinning provided a very flat and thin wafer. Wafer thinning down to 100 μm showed only 2.0 mm curvature. It is very important for WLP architecture to optimize mold and DBF material and thickness in determining the proper trade-off between the on-die stress and the die level warpage performance.

3.4 Front side mold and mold grinding

Front side mold and mold grinding to expose the copper bump was demonstrated to have good thickness uniformity and only a minor copper residue issue. Table 2 shows the mold thickness uniformity before and after the mold grinding for both a blank wafer and bumped wafer. The compression mold thickness uniformity within wafer averaged 111.0 μm with a standard deviation of 8.03 μm. The analysis of variation within a wafer showed a consistent one directional mold thickness gradient in a wafer, potentially due to the mold chase design and setting issue. The mold chase improvements will be the next step to further improve the mold thickness uniformity.

After the copper bump exposed mold grinding process, the remaining mold thickness averaged 26.5 μm with a 0.66 μm standard deviation. This demonstrated that the
mold thickness non-uniformity could be successfully planarized through front side grinding process.

The grinding process required grinding of both copper bumps and epoxy mold compound simultaneously. Characterization of grinding wheel was required to avoid copper shear droop and was optimized through proper loading on the wheel face while grinding copper. Two step grinding was characterized to demonstrate the concept. However, we observed minor copper residue on the mold surface in the range of less than 10 μm size. Figure 10 displays the surface scratches with copper residues.

Further characterization of the grinding process to minimize the impact of the board level reliability such as moisture related failures is required.

3.5 Wafer thinning

The wafer thinning process with bumped wafer was successfully demonstrated on the 200 μm through 75 μm wafer thickness legs without any showstopper type of issue. The 200 μm thickness target leg showed an average thickness of 178 μm with a standard deviation of 3.19 μm. To better target the average wafer, we will need to measure the front side mold thickness accurately before silicon thinning. SEM photographs for both the 200 μm thickness leg and the 75 μm thickness leg are shown in Fig. 11.

3.6 Board level assembly

A reflow chip attach process was demonstrated on the 200 μm silicon thickness WLP with 200 μm core matrix mold array package. The board level assembly was done using water soluble flux with substrate pre-solder. Underfill was applied after chip-attach reflow and deflux cleaning. Figure 12 shows the photograph of a WLP mounted on a test board using a standard reflow process, and a cross sectional view of the board assembly for both the 200 μm silicon thickness leg and 75 μm thickness leg. No significant solder voids were observed.

4. Discussion

4.1 Die protection for DCA mounted on PCB

The potential DCA failure modes after SMT include silicon edge failures and interconnect failures between die and motherboard. In most DCA applications, underfill is typically optional and not the preferred solution at SMT. Without underfill, the edge of silicon chip, particularly at the saw side wall of the metal layer and interlayer dielectric, are exposed to the environment with the risk of thermo-mechanical and moisture related failures. The WLP architecture required a solution for the silicon edge failures.
Figure 13 is an alternative process to improve the die robustness using a fully molded WLP structure. The process steps would be as follows:

a) Start with full thickness wafer,
b) half depth cut the saw street
c) front side mold and grind,
d) thin the wafer to expose the half cut street
e) back side film lamination,
f) wafer bumping
g) saw singulation.

Since all of the die edges are completely molded after this process, it will improve the susceptibility to die edge cracking or chipping and offer enhanced environmental protection.

4.2 Next step

As described, we have demonstrated a feasible WLP process using a 300 mm wafer. Data will continue to be collected on warpage behavior, mechanical properties such as die breakage strength, the board level assembly process using a standard surface mount infrastructure, and on the WLP on PCB system.

5. Conclusions

To enable future low cost packages for DCA on board, we have demonstrated a viable 300 mm WLP process. The WLP is thin, small, and low cost by eliminating the package substrate. The WLP’s reliability was enhanced due to the protection offered by the epoxy compound on both the front side and back side of the die. The WLP is easy to assemble as it has solder micro balls on a very flat surface.

Key challenges going forward include the front side and back side grinding thickness variance control. Optimization of mold and DBF material and thickness will be required to manage on-die stress and the die level warpage performance. In addition, alternative bumping technologies[8–10] scalable to less than 0.3 mm pitch will be investigated as a lower cost option for high pin counts. Finally, we will need to define the minimum WLP thickness for handling and assembly using current surface mount infrastructure.

Acknowledgement

The authors would like to thank the people who have contributed to the project. All this work would not have been possible without several tool and material suppliers, who supported the processing of our wafers. We thank Charles Gealer and Nicholas Holmberg, Assembly and Test Technology Development (ATTD), for the great support to demonstrate the concept. And thank Weston Roth and Damion Searls, ATTD, for support to SMT. Finally special thanks go to Tsuyoshi Abe, Technology and Manufacturing Group Japan, for supporting flawless virtual lab processing at suppliers.

Reference

2002.
