Wafer Level Assembly Technique Development for Fine Pitch Flip Chip 3D Die-to-Wafer Integration

Zhaozhi Li+, Brian J. Lewis, Paul N. Houston, Daniel F. Baldwin,
Eugene A. Stout *, Theodore G. Tessier * and John L. Evans+.
ENGENT, Inc.
3140 Northwoods Parkway, Suite 300A
Norcross GA 30071
www.engentaat.com

*FLIP CHIP INTERNATIONAL LLC
3701 E. University Dr.
Phoenix, AZ 85034
www.flipchip.com

+Industrial and Systems Engineering
Auburn University
Auburn University, AL 36849

ABSTRACT
Three Dimensional (3D) Packaging has become an industry obsession as the market demand continues to grow toward higher packaging densities and smaller form factor. In the meanwhile, the 3D die-to-wafer (D2W) packaging structure is gaining popularity due to its high manufacturing throughput and low cost per package. In this paper, the development of the assembly process for a 3D die-to-wafer packaging technology, that leverages the wafer level assembly technique and flip chip process, is introduced.

Research efforts were focused on the high-density flip chip wafer level assembly techniques, as well as the challenges, innovations and solutions associated with this type of 3D packaging technology. Processing challenges and innovations addressed include flip chip fluxing methods for very fine-pitch and small bump sizes; wafer level flip chip assembly program creation and yield improvements; and set up of the Pb-free reflow profile for the assembled wafer.

100% yield was achieved on the test vehicle wafer that has totally 1,876 flip chip dies assembled on it. This work has demonstrated that the flip chip 3D die-to-wafer packaging architecture can be processed with robust yield and high manufacturing throughput, and thus to be a cost effective, rapid time to market alternative to emerging 3D wafer level integration methodologies.

Key words: 3D Packaging, Die-to-Wafer, Silicon on Silicon, Wafer Level Assembly, Flip Chip on Wafer

INTRODUCTION

3D packaging has become a leading packaging solution as the market demand for high packaging density, low cost, small form factor continues to grow[1,2]. The 3D die-to-wafer integration, which enables high manufacturing throughput by attaching individual silicon die to the silicon substrate at the wafer scale, has become one of the popular 3D packaging techniques.

A 3D die-to-wafer packaging technology with the interconnections realized by the Through Silicon Vias (TSV) is discussed in [3]. The dies are placed on to the landing wafer and the bonding is performed with adhesive, heat and pressure. However, due to the fine pitch characteristics and non-uniform distribution of the dies on the base wafer, this type of die-to-wafer integration requires very high alignment accuracy and a specially engineered, highly-leveled wafer level bonding tool, as the bonding force needs to be homogeneous, controlled, and perpendicular. All these factors have made this type of 3D TSV die-to-wafer packaging technology challenging and costly.

A novel 3D wafer level chip scale packaging (3D-WLCSP) technology that realizes the 3D die-to-wafer integration by flip chip process is introduced in this paper. Compared with the 3D die-to-wafer packaging discussed previously that utilizes the TSVs as interconnect method, the 3D-WLCSP technology realizes the die-to-wafer integration by leveraging both low cost flip chip process and the existing infrastructures of high throughput wafer-level packaging techniques. The structural view of the 3D-WLCSP is shown in figure 1. The active side of a flip chip die is face-to-face mounted on the active side of the WLCSP on the wafer scale that carries hundreds of identical WLCSP substrates. Due to the face-to-face bonding structure, both the signal transmission path and the electrical transmission path in the

package can be greatly shortened so that the overall electrical performance of the package is substantially improved. The die-to-wafer interconnects are realized by the fine pitch flip chip solder joints. High levels of packaging density, small package footprints, and thin package profile can be achieved through this packaging technology.

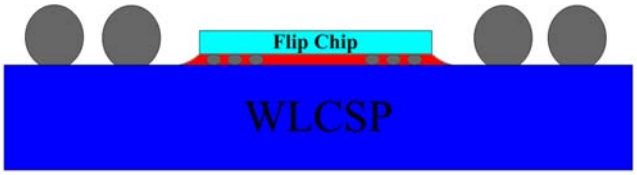

Figure 1. The Schematic View of 3D-WLCSP Technology: Flip Chip Silicon to Silicon Wafer Level Integration

The assembly process of this 3D die-to-wafer flip chip packaging process is shown in Figure 2. The 3D-WLCSP technology discussed in this paper can dramatically increase the manufacturing throughput due to the fact that all the flip chip interconnections across the whole wafer can be formed in one single reflow process.

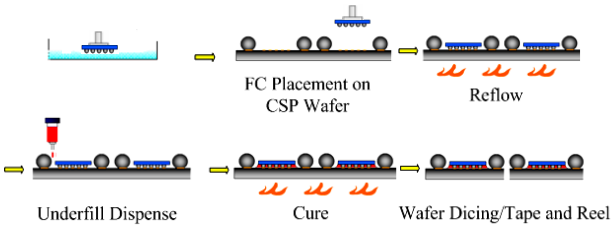

Figure 2. Capillary Underfill Process for Flip Chip 3D Die-to-Wafer Packaging (3D-WLCSP)

Previous publications have discussed the material selection (such as flux and underfill) and the reliability evaluation of this novel 3D packaging architecture [4-6]. The previous development efforts were performed on 2x2 and 3x3 WLCSP substrate tiles, in order to facilitate the process development and save the test vehicle wafers. Based on the assembly process parameters and experiences gained on the 2x2 and 3x3 tiles assembly, in this paper, the development of wafer level assembly process for the 3D-WLCSP technology is carried out. The focus is put on the assembly program creation, fluxing activities on the wafer, and the wafer level assembly yield improvement.

TEST VEHICLES

Figure 3 shows the test vehicle wafer which has a total of 1,876 WLCSP substrates fabricated on it. The test vehicle wafer is divided into four different quadrants with different layouts fabricated on each quadrant. The four test vehicles were simultaneously manufactured on the same wafer by dividing the WLCSP carrier substrate and complimentary flip chip test wafer into quadrants (listed as Quadrant 1 to Quadrant 4 in Figures 4 and 5). The four test vehicles have different pitches, bump sizes and layouts, which are designed to allow researchers to compare and study the effects of these design features on different field applications.

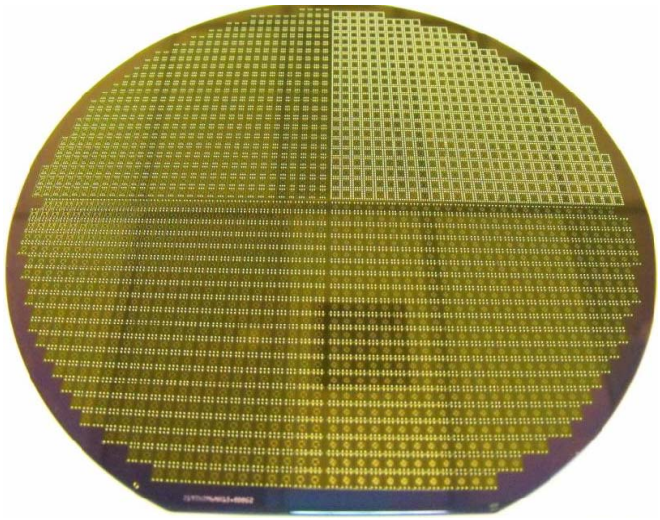

Figure 3. Test Vehicle WLCSP Substrate Wafer

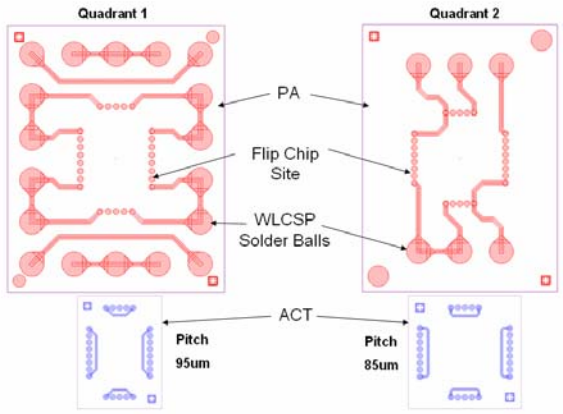

Figure 4. Quadrant 1 and Quadrant 2 Test Vehicle

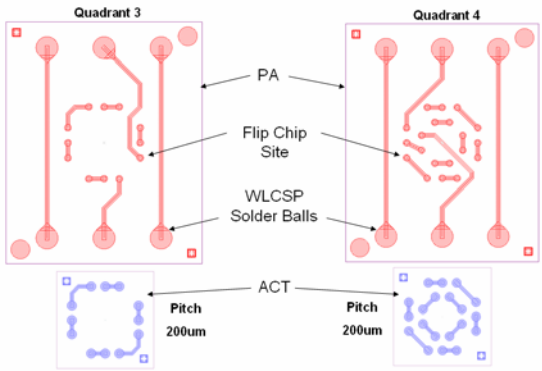

Figure 5. Quadrant 3 and Quadrant 4 Test Vehicle

The Passive Component (PA) device forms the base substrate for the 3D flip chip to WLCSP package and is a solder balled wafer level CSP in wafer form, as shown in Figure 3. The active component (ACT) dies are face-to-face

mounted to the PA wafer by the flip chip process. All of the devices used for this work contain Pb-free SAC 305 (96.5% tin, 3% silver, 0.5% copper) solder bumps. The details of the ACT and PA test vehicles are outlined in Table 1 and Table 2.

Table 1 Flip Chip Die Test Vehicle Information

Quadrant	Pitch/Bump Diameter (um)	Flip Chip I/O count	Flip Chip I/O array	UBM Diameter (um)	Die Size (um)
1	95/70	26	Peripheral	55	1100 x 1500
2	85/63	26	Peripheral	53	1300 x 1300
3	200/105	12	Peripheral	100	1300 x 1300
4	200/95	20	Full	90	1300 x 1300

Table 2 WLCSP Substrate Test Vehicle Information

Quadrant	CSP Pitch (um)	I/O Count	CSP I/O Array	UBM Diameter (um)	Die Size (um)
1	500	18	Peripheral	300	2500 x 3100
2	500	6	Peripheral	300	2270 x 3150
3	750	6	Peripheral	300	2500 x 3100
4	750	6	Peripheral	300	2500 x 3100

EXPERIMENTAL APPROACH AND RESULTS

1. Wafer Level Assembly Program Creation

ESEC Micron 2 die attach machine was used in this research as the means of pick and place activities. The operation of the ESEC Micron 2 machine requires that the placement location of each individual chips be known to the machine. The most straightforward way is to calculate each individual placement location on the wafer with reference of the wafer fiducials and input the X-Y coordinates of each location into the program. Two disadvantages are associated with this method. The first drawback is the time consumed on verifying and inputting in the X-Y coordinates for over 1,800 locations. If the location data were input wrongly, the placement accuracy would be compromised resulting yield loss. The second drawback found with the direct input method is that during the X-Y location data entry, the machine gets slower in response to the data keyed in as insufficient memory could be allocated. In this case, the assembly program creation is quite inefficient. Both drawbacks indicate that the direct key-in method is not applicable with the ESEC Micron 2 chip placement machine.

A new way to program the placement activities on the whole wafer was developed. As four different test vehicles were fabricated on the same WLCSP wafer with each test vehicle taking up one quadrant of the wafer, the wafer programming was split into four different programming activities, with each wafer quadrant programmed separately. For example, the first quadrant of the wafer has a total of 506 Q1 WLCSP substrates. The second quadrant of the wafer has a total of 470 Q2 WLCSP substrates. Both the third and fourth wafer quadrants have 450 WLCSP substrates fabricated. Thus the program was created based on the quadrant rather than the whole wafer, as each quadrant wafer has different die and substrate size as well as the layout. The image of quadrant 1 wafer is shown in Figure 6.

In order to program Quadrant 1 efficiently and save machine response time during program execution, the program for Quadrant 1 was divided into 2 sub-programs. Group “A”, as shown in Figure 6, was laid out in a rectangular shape, which has a 17x18 array (306 pieces) of CSP substrates. The 17x18 locations in section “A” were programmed by a “step and repeat” programming function provided by the ESEC Micron 2 machine. The first column is programmed (only 18 locations in the first column keyed into the program) and the remaining 17 columns in area “A” were “imaged” from the first column with only the interval (distance between the first column to the imaged column) input to the program. Therefore, instead of inputting 306 (=18x17) locations to the program, only 35 (=17+18) locations were input to sub-program 1. This method greatly increased the program efficiency. For areas B and C on Quadrant 1, the location of each individual substrate was directly input to sub-program 2, as neither area is in a rectangular shape.

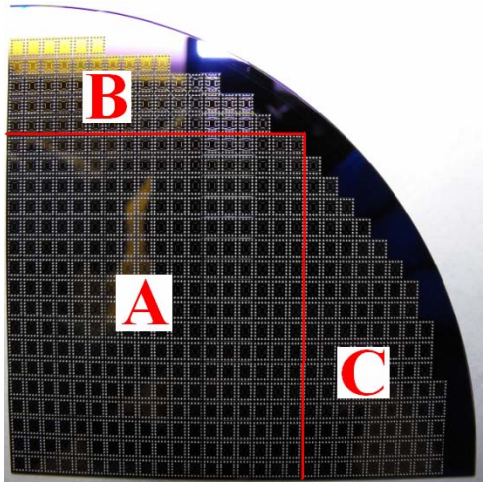

Figure 6. WLCSP Wafer Quadrant 1 Test Vehicle

The same programming method was used on Quadrant 2, 3 and 4 partial wafers. In total, there were 8 sub-programs for the whole wafer-scale programming. By dividing the whole wafer into sections and programming according to the layout of each section, the wafer assembly program was accomplished with high efficiency and reduced machine response time.

2. Flux Dwell Time Study

The result of the dry run on the initial wafer assembly program indicated that the estimated time for the ESEC Micron 2 to finish 1,876 times of pick, die vision, dip fluxing and placement is approximately four hours with full machine placement speed. This means the chip placed on the first location shall wait for around four hours before it goes to the reflow oven. Thus it is important to learn if the flux material is still functional after four hours of dwell time in the air. The functionality of the flux after dwell time is critical for the wafer level assembly yield. If the flux material becomes dried out and loses its fluxing capability before the reflow of assembled wafer, wafer level assembly

yield loss could occur. Therefore, an experiment was conducted to learn how long the flux dwell time can be in order to maintain 100% yield on the 3D-WLCSP first-level assembly.

Quadrant 3 test vehicles were utilized in this study. A total of 45 pieces of Quadrant 3 first-level packages were assembled. The flux dip height was 45 microns. After the end of every hour of dwell time, 9 pieces of first-level packages were sent to the reflow oven with nitrogen environment, following the reflow profile condition shown in Table 3. The reflow parameter settings are based on the studies reported in previous publications [4-6]. The outline of the experiment and results are shown in Table 4.

Table 3. Reflow Profile Parameter Settings for 3D Die-to-Wafer Flip Chip Assembly

Reflow Parameter	Ramp Rate	Soak Temp	Soak Time	Time Above Liquidus	Peak Temp
	1.2°C/s	150 ~ 180 °C	50 sec	65 sec	240°C

Table 4 WLCSP Substrate Test Vehicle Information

Dwell Time (Hours)	Assembly Yield
1	9/9
2	9/9
3	9/9
4	9/9
5	9/9

The results, as shown in Table 4, indicate that the tacky flux material selected for the 3D-WLCSP wafer level assembly has a long dwell time that will suffice the dwell time needed for the wafer level assembly process.

3. Initial Wafer Level Assembly Evaluation

The initial wafer level assembly was conducted after the creation of the program and the study of the flux dwell time. The reflow profile parameters are shown in Table 3. As discussed previously, the program was created based on each wafer quadrant. Thus the initial assembly was also conducted on the quadrant wafers. The full machine speed was utilized to evaluate the machine capability for a high-speed, high throughput wafer level assembly requirement. The images of each quadrant wafer after assembly are shown in Figure 7.

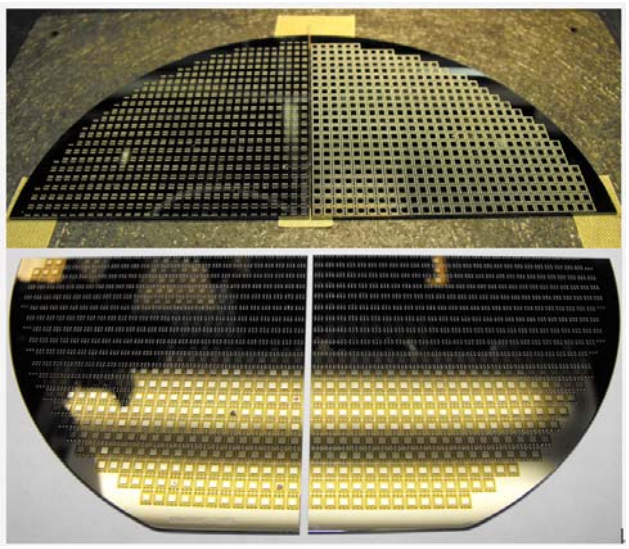

Figure 7 Initial Assembly on Quadrant Wafers

The initial yield on each quadrant wafer was calculated based on the probing of resistance across the daisy chain on each individual flip chip to WLCSP package. The initial yield is shown in Table 5.

Table 5 Initial Assembly Yield on Quadrant Wafer

Quadrant	Packages Assembled	Packages Good	Yield
1	470	442	94.04%
2	506	501	99.01%
3	450	445	98.89%
4	450	446	99.11%

The initial yield was satisfactory for Quadrants 2, 3, and 4. However, Quadrant 1 had the greatest yield loss. A destructive shear test was performed on the failed Quadrant 1 packages. After removal of the flip chip die, it was found that the die was placed inaccurately as the marks (flux residue) that the solder bumps left on the CSP substrate were “shifted” from the pad location, as shown in Figure 8. This indicates that either the placement program or the machine vision process had a problem.

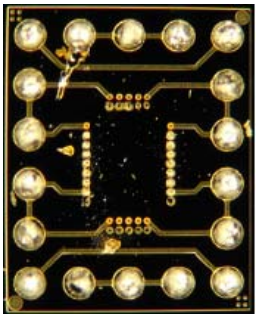

Figure 8 Destructive Failure Analysis of Quadrant 1 Failed Assembly

No similar failure was observed on the failed packages of Quadrants 2, 3, or 4. However, it was found from the flip

chip die inspection that some bumps were either missing or damaged, as shown in Figure 9. Figure 10 shows a cracked Quadrant 2 die captured by the die inspection. The daisy chain was opened because of the die crack. The defects on the flip chip die accounted for the failures on Quadrant 2, 3 4 initial assemblies.

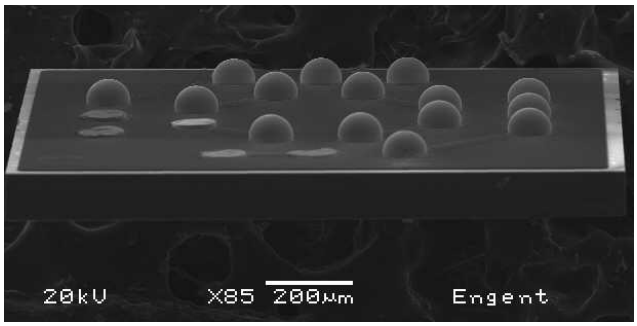

Figure 9 Missing Bumps on Quadrant 4 Flip Chip Die

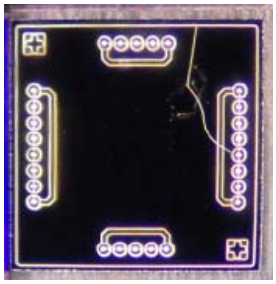

Figure 10 Die Crack on Quadrant 1 Flip Chip Die

4. Wafer Level Assembly Yield Improvement

Based on the failure analysis performed previously, the placement inaccuracy was the main reason accounting for the yield loss on Quadrant 1 partial wafer. The cause of the placement inaccuracy is the machine vision activity of local fiducials on the first column during the "step and repeat". As only the first column was visioned, if the wafer X-Y coordinates are not perfectly aligned with the machine X-Y coordinates, the "shift" will happen during the "repeat" action. To solve this problem, three global fiducials on the Quadrant 1 wafer were used in the modified program instead of local fiducials on the first column.

Careful die inspection was performed on Quadrants 1, 2, 3, and 4 flip chip dies. The dies that had the bump non-uniformity issue or other defects such as die crack were picked out. The assembly yields on different wafer quadrants after process improvement are shown in Table 6. 100% yield was achieved on all the packages across the whole wafer. The image of the assembled whole wafer that had 100% assembly is shown in Figure 11. Totally 1,876 silicon-on-silicon packages, which account for 39,776 flip chip solder joints, were formed during the wafer reflow process.

Table 6 Assembly Yield on Quadrant Wafers after Process Improvement

Quadrant	Packages Assembled	Packages Good	Yield
1	470	470	100.00%
2	506	506	100.00%
3	450	450	100.00%
4	450	450	100.00%

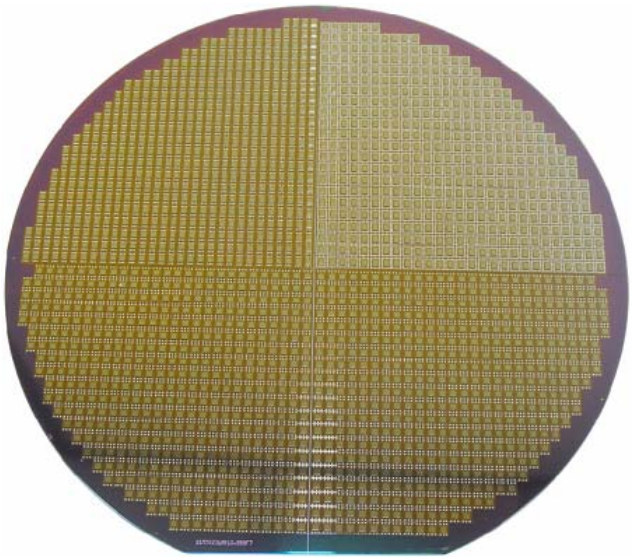

Figure 11 Assembled 3D-WLCSP Wafer Showing 100% Yield

The assembly of the whole wafer (1,876 placement sites) took three hours and thirty-seven minutes. Considering the reflow process of 6 minutes, on average the process time for each individual 3D-WLCSP first level assembly can be calculated as follows:

$$(3 \times 60 + 37 + 6) \times 60 / 1876 = 7.13 \text{ seconds/package}$$

For the standard flip chip on board assembly process, the chip placement time plus the reflow time would be around 6-7 minutes. Thus, it can be seen that the 3D die-to-wafer assembly process with 3D-WLCSP technology can dramatically improve the throughput to around 50 times or higher.

DISCUSSION AND CONCLUSIONS

The wafer level assembly process development for a 3D die-to-wafer integration technology (3D-WLCSP) is discussed in this paper. Development effort was focused on the pick and placement program creation for the wafer level assembly, flux dwell time study and the wafer level assembly yield analysis and improvement.

A "step and repeat" program technique was utilized for the programming on quadrant wafers. This technique greatly reduced the data entry and increased the machine operation efficiency.

The flux dwell time study showed that the test vehicle tiles can maintain 100% yield with a dwell time of as long as 5 hours before entering the reflow oven. This study confirmed that the flux used for the wafer level assembly can maintain its fluxing capability throughout the wafer level assembly process.

The initial assembly yield loss on Quadrant 1 was induced by the vision alignment of the local fiducials on the first column during the "step and repeat" process. A program improvement was made on the Quadrant 1 pick-and-placement program, with the use of global fiducials instead of local fiducials. Flip chip die defects identified in the Quadrant 2, 3 and 4 flip chip test vehicles accounted for the yield loss during the initial wafer level assembly. Careful die inspection was performed to eliminate the die with defects. After these efforts, 100% yield was achieved on the test vehicle wafer that carries 1,876 silicon-on-silicon packages. The full speed of the ESEC Micron 2 machine was utilized in the wafer assembly which ensures the high throughput of the 3D-WLCSP wafer level assembly.

This part of the work demonstrated that the 3D die-to-wafer assembly through the 3D-WLCSP technology can be processed with high yield and thus is a proven good solution for the 3D die-to-wafer integration.

References

[1] Yoon, S. W., Ganesh, V. P. et al, "Packaging and Assembly of 3-D Silicon Stacked Module for Image Sensor Application," *IEEE Transactions on Advanced Packaging*, Vol. 31, No. 3 (2008), pp. 519-526.

[2] Karnezos, M., "3-D Packaging: Where All Technologies Come Together," 29th *Electronics Manufacturing Technology Symposium*, San Jose, CA, July, 2004.

[3] http://www.evgroup.com/en/solutions/logic_memory/c2w_bonding/

[4] Li, Z., Lee, S., Lewis, B. J., Houston, P. N., Baldwin, D. F., Stout, G., Tessier, T., Evans, J.L. "Impact of Pb Free Reflow Profile Parameters Toward Flip Chip on Silicon Assembly Yield, Reliability and Intermetallic Compound Characteristics," *Proceedings of 60th Electronic Components and Technology Conference*, Las Vegas, NV, June, 2010.

[5] Li, Z., Houston, P. N., Baldwin, D. F., Stout, G., Tessier, T., Evans, J.L. "Design, processing and reliability characterizations of a 3D-WLCSP packaged component," *Proceedings of 59th Electronic Components and Technology Conference*, San Diego, CA, May, 2009.

[6] Houston, P. N., Li, Z., Baldwin, D. F., Stout, G., Tessier, T., "A 3D-WLCSP Package Technology: Processing and Reliability Characterization," *Proceedings of 58th Electronic Components and Technology Conference*, Lake Buena Vista, FL, May, 2008.