

Laser Based Assembly of Ultra Fine Pitch Bumped ICs For Chip-on-Chip Proximity Coupled Applications

Andrew Strandjord, Thorsten Teutsch, and Ghassem Azdasht
Pac Tech – Packaging Technologies, Inc.
328 Martin Avenue
Santa Clara, CA USA 95050

Matt Giere
Sun Microsystems
9515 Towne Centre Drive
San Diego, CA 92121

Abstract:

This paper describes a process flow for creating ultra-fine-pitch solder interconnects between stacked die. The application for this die stack is Proximity Communication (PxC). Tight geometric control of both the bumping and assembly processes are required in all three dimensions, x-y-z, to enable efficient capacitance communication between the die. The test vehicle for this program consists of a bridged-stack design, where one chip is used as communication link (bridge) between two other die (island). The solder bumps perform several critical functions in this application: 1) provide for electrical transmission of power, ground, and low frequency signals between chips, 2) create precise x-y alignment between the stacked die which allows for good overlap of the non-bumped pads which are used for proximity coupled data transmission between die, and 3) provide for a narrow and tightly controlled standoff distance between stacked die to enable efficient capacitive coupling. The bumping process consisted of depositing an e-Ni/Au layer as the UBM (3 μm) on both the bridge and island dies, creating micro bumps of Sn (3 μm) on the bridge die, and creating a second set of flip chip sized bumps (100 μm SnAgCu) on the island die for assembly to a ceramic substrate. The bumped wafers were thinned to 150 μm , singulated, laser assembled together using the micro bumps to create a three chip unit, and then laser assembled to a ceramic substrate using the flip chip bumps. Assembly is accomplished by aligning the bumped bridge die to the corresponding unbumped island die and laser heating the micro bump area to reflow the Sn solder to form the fine pitch interconnect. The second assembly to the ceramic substrate is also accomplished using the laser heating process. This laser technology allows for highly localized heating, which allows for independent reflow and assembly of the two different solder bump regions, i.e. Sn micro-bumps and SnAgCu flip chip bumps.

Keywords: chip-on-chip, solder bumping, capacitive coupling, laser assembly

Introduction:

The transmission of signals between two or more semiconductor chips is traditionally accomplished through the use of either conventional wirebonds or solder interconnects, to connect the individual chips to a secondary substrate (or interposer), where another set of interconnects (metal traces) carry the electrical signals between the chips. The size and density of these chip interconnects (backend) are continually being reduced in an attempt to match the size reduction and increase functionality of new integrated circuit

designs (frontend). With many experts now anticipating an absolute limit on further silicon scaling, (i.e. the perceived end to Moore's Law), the industry has quickly moved toward building circuitry in the Z direction. This enables added functionality, while still maintaining the same footprint in the X-Y dimensions. This is often referred to as chip-on-chip or 3D packaging. The integration of chips in the Z direction has resulted in some fairly exotic chip-to-chip interconnect schemes. In some examples, a seemingly

chaotic maze of wirebonds is used to carry the signals between chips. One of the main issues with this method is that the relatively tall standoff of the wirebonds still makes the die stack tall in the Z direction. On the other hand, the use of solder interconnects can significantly reduce this standoff, especially when combined with thru-silicon-vias (TSV) and/or incorporating alternative signal transmitting technologies, like capacitive and inductive coupling [1-4].

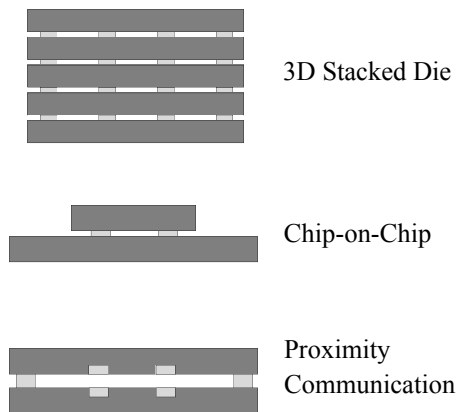


Figure 1. Schematic drawings of applications for ultra fine pitch bumping (micro bumps).

For most traditional solder bumping applications on semiconductor wafers, the bumps range in height from around 50 μm up to 300 μm . The solder bumps at the lower end of this range are often referred to as flip chip bumps (FC) and those at the higher end of this range are called wafer level chip scale packaging bumps (WLCSP). For 3D applications, the use of these ‘large’ solder bumps is not ideal because it still increases the overall size of the system in the Z dimension, limits interconnect density, and also contributes to parasitics which affect the electrical performance. By making the interconnect structure between the die very thin, one can create a device with improved performance in all of these areas.

There are several technologies which can be used to make this thin interconnect structure, including: copper to copper thermo-compression bonding and metal-solder micro-bumps. The selection between these options is often dictated by where one is in the supply and manufacturing chain. For example: if one is a vertically integrated company manufacturing memory die and has the capabilities to make thru-silicon-vias (TSV) at the fab (commonly referred to as front-end-of-line vias, FEOL), one may choose using copper-to-copper interconnects. Whereas, if one is using a foundry for their wafers and is creating vias in completed wafers (back-end-of-line vias, BEOL) one

may choose to use the solder micro-bump technology. Many other scenarios exist which will also affect the choice of interconnect technology.

The use of micro-bumps to create this thin interconnect is one method that allows for some versatility. By using wafer-level-packaging techniques (WLP) to create these micro-bumps, one can take advantage of the infrastructure that already exists within packaging subcontractors. These wafer level processes include: redistribution, UBM deposition, solder bumping, thin film deposition, and wafer thinning.

There are a number of different WLP technologies used to create solder bumps on integrated circuits. The most common include: printing solder pastes, dropping preformed spheres, and electroplating metal alloys. In general, the paste printing and sphere dropping technologies result in solder bumps that span the entire range listed above, i.e. 50 to 300 μm tall bumps. The electroplating technique is fairly versatile in that micro solder bumps can be formed which are very thin, i.e. 1 to 50 microns. This is evident in the fact that there has been a lot of activity in plating copper pillars followed by plating a thin layer of tin on top of the copper. The limitations of this electroplating technique are related to plating uniformity, high capital expenditure, and alloy composition.

In this paper, we describe the use of e-Ni/Au and a sputtered liftoff process to create very thin bumps at the wafer level. Electroless nickel/gold is first deposited onto the bond pads of the die using a series of sequential wet chemical techniques [5-6]. A liftoff resist is then applied to the wafer and photo lithographically patterned. This is then followed by the sputter deposition of thin layers of metallic solder and stripping of the liftoff resist.

Assembly of stacked devices with micro-solder interconnects poses its own set of challenges. As one stacks die, one upon another, multiple reflows are often necessary to build the entire stack. This can lead to the buildup of solder intermetallics, increase stress levels, and reducing reliability. In this paper we describe a laser assembly technology which significantly reduces all these effects [7].

Experimental:

The test vehicle for this program is a three-chip package as shown below in Figure 2. Two high-powered ‘island’ chips are solder connected to an alumina ceramic substrate using 2,512 flip chip solder bumps. One ‘bridge’ chip is used to communicate data

between each island chip, via the capacitively couple interconnects created using micro-solder bumps. The bridge chip and island chips were thinned to 150 μm . The substrate is an alumina ceramic substrate with tungsten alloy traces. A cavity 400-600 μm deep allows for the mechanical clearance for the bridge after reflow.

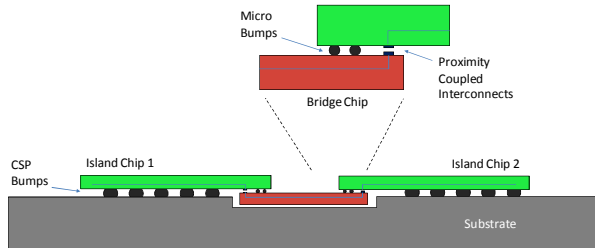


Figure 2. Schematic drawing of the proximity coupled test vehicle.

To produce efficient coupling of proximity interconnects, the pads on both the bridge and island chips need to be accurately aligned with respect to each other. The metal pads on adjacent chips act as parallel plates of a simple capacitor. The transmitter circuit on one chip drives its pad, and the corresponding receiver circuit on the adjacent chip detects the coupled voltage on its pads. This chip-to-chip communication requires neither solder, nor wires, but rather uses free space capacitive coupling. Because the capacitance between plates depends on the overlap and proximity of the chips, their alignment must be carefully controlled in all three translational and all three rotational axes. Experimental measurements indicate that for any particular choice of pad size, an X-, Y-, and Z-alignment of less than 1/3 of a pad pitch, or 8 μm for 24 μm pads, must be maintained [1]. Micro-solder interconnects are an ideal technology for aligning the stacked chips to each other and easily meeting these tolerances.

The island chip and bridge chips were fabricated on 8" silicon wafers. The area designed for micro-bumps consist of a field of 45 μm square aluminum pads on a 50 μm pitch. These are passivated with silicon nitride, with a 35 μm pad opening.

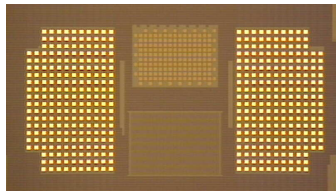


Figure 3. Optical image of micro-bump pad area on the bridge die.

Electroless Nickel/Gold Processing:

Most semiconductor devices that are to be bumped with solder also require a protective metal layer between the solder and the pad metallurgy (under-bump-metallurgy or UBM). Although there are many different materials and techniques for applying this layer, the most common are sputtered Al/NiV/Cu, electroplated Cu, and electroless nickel/gold [1-3]. In this study, we used the electroless nickel/gold process in all of the studies. This allows one to easily control and vary the height of the solder interconnect, and thus control and vary the standoff between the island and bridge die. The micro-bump and flip chip areas were both plated with the same thickness of nickel and gold.

For aluminum based integrated circuits, the chemical sequence for depositing the electroless nickel layer is to first clean off the pads of any organic or silicon based contaminants which may occur due to wafer handling, storage, or variations in the manufacturing process. This is typically performed using acid based chemicals.

The second step is to remove any native oxide that may have built up on the aluminum pad surface. This is typically performed using caustic based etching chemicals. The next step is to activate the surface of the aluminum. The most commonly used wet chemical system for this is "zincation", where a zinc oxide solution is used to replace some of the pad aluminum with zinc in an electrochemical reaction. Empirical research has shown that by stripping this zinc layer off and then reforming it in a second zincation step, one creates a higher quality layer of zinc. This is often referred to as "double zincation". This zinc layer changes the electric potential of the aluminum pad, and when immersed in a nickel sulfate solution, nickel replaces this zinc and an autocatalytic nickel reaction continues.

By adjusting the time, temperature, pH, and chemical concentrations of the nickel plating bath, one can create nickel structures between 1 and 25 microns tall. A nickel height of 3 microns was used in this study (2 μm above the 1 μm thick silicon nitride passivation).

For most applications, the deposition of solder does not immediately follow the nickel deposition process. The nickel surface will oxidize fairly quickly, and therefore a thin layer of a noble metal is typically deposited on top of the nickel to protect the surface from oxidation. There are two common metals (Pd and Au) which are compatible with the electroless nickel process and can be deposited sequentially within the same plating line

using either immersion or electroless based processes. A 600Å thick layer of gold was used in this study.

In this study both the island die and the bridge die were plated with 3µm on Ni/Au.

Resist Processing:

There are several techniques which one can use to deposit and pattern a thin layer of solder. The subtractive etching of a sputtered or evaporated metal is the most common technology used to pattern metals like copper and aluminum. A lift-off process is not as common for use in patterning interlayer metallurgies in semiconductors, but is fairly common for backend applications like redistribution and pad resurfacing. The general principle associated with the lift-off technique enables the use of nearly any metal, metal alloy, or metal stack.

In this study, a two layer lift-off process was used. A positive resist is first coated onto the wafers and flood exposed. This is followed by a second resist layer which is pattern exposed and then developed to produce a two layer system where the second resist layer extends beyond the first in the area that is opened.

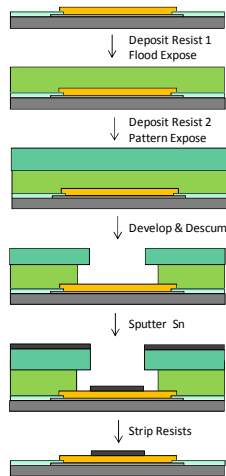


Figure 4. Schematic of process flow used for patterning the sputtered solder bump.

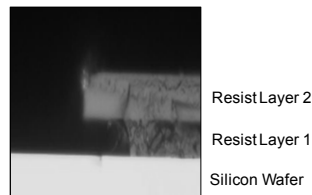


Figure 5. Optical image of the resist profile used for patterning the sputtered metal solder layer.

Sn/Au Sputtering and Resist Liftoff:

Three microns of tin was sputtered deposited onto the wafers using a standard magnetron sputtering system. One observes a fairly coarse grain structure in the sputtered tin film with good adhesion to the nickel/gold UBM layer. The lift off resist was removed by soaking the wafers in an acetone bath at room temperature. The use of ultrasonics greatly increased the rate of resist dissolution. NMP based strippers were also evaluated with similar results. Only the island die was bumped with Sn. The total height of the bump above the 1 µm passivation was 5 microns: 3 µm of NiAu + 2 µm of Sn.

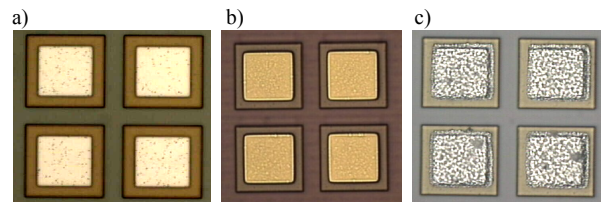


Figure 6. Optical images of: a) As received aluminum bond pads on bridge chip, b) Ni/Au plated bond pads, c) Sputtered Sn on the bond pads.

Flip Chip Solder Bumping:

Solder bumping of the flip chip sized pads on the island die was performed at the wafer level using a solder sphere placement systems (Ultra-SB², trademark of PacTech). This technology uses a patterned vacuum tooling fixture to simultaneously pick up preformed spheres and transfer them all over to the wafer at once. Figure 7 outlines the basic steps in the transfer process. First, a tooling plate (vacuum stencil) is lowered into the sphere reservoir. Vacuum is applied to the tooling plate to selectively pick up the spheres (a). This fixture is then inspected for missing or extra spheres. The tooling is then aligned to the wafer (b) and lowered to bring the solder spheres into contact with the wafer pads (c). Finally, the vacuum is turned off; the tooling plate is raised (d); and the solder spheres are reflowed (e).

The tooling plate is patterned with openings that correspond directly with the location of the I/O pads on the wafer. This tooling is created using similar methods to that of making a nickel plated surface-mount stencil. The size of openings in the tooling is designed to be slightly smaller than the size of spheres that will be placed onto the wafer. Because this tooling plate can be made with such accuracy and because the Ultra-SB² equipment has a placement accuracy of better than

$\pm 15\mu\text{m}$, the process can be used for both WLCSP and flip chip applications.

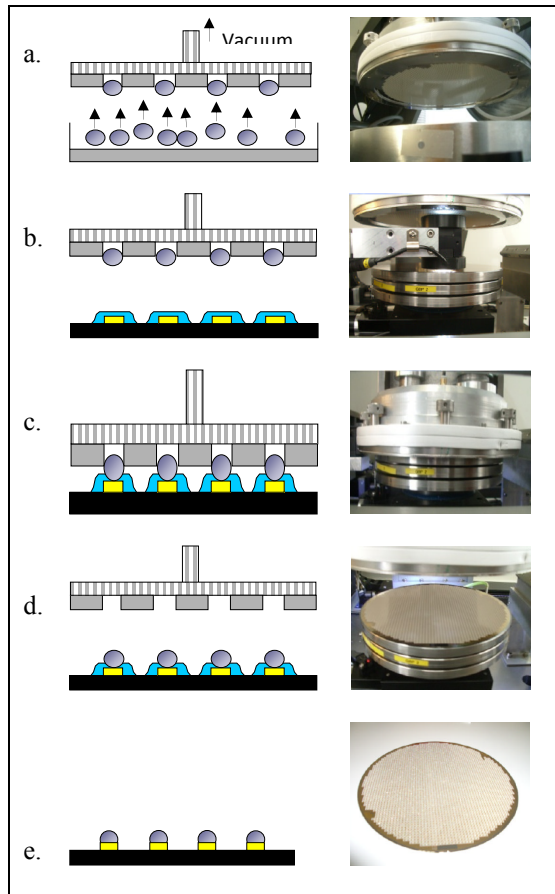


Figure 7. Process flow for flip chip bumping using Ultra-SB² technology.

Wafer Thinning and Dicing:

The wafers were thinned to 150 μm total silicon thickness using a Disco 5600 tool, followed by fine grit polishing for stress relief.

Wafer dicing was performed using a Disco double spindle system. The die were left on the sawing tape.

Assembly:

Assembly of the die was carried out using a LaPlace™ (trademark of PacTech) laser bonding tool [8-10]. Figure 8 shows the bond head region of the laser assembly tool. The principle behind laser bonding is to pick up and physically align the solder bumped die to the mating die (or substrate). This used nearly identical technology that is associated with most high speed pick and place tools. The difference comes in the way that

the solder bumps are reflowed. The LaPlace tool uses a laser pulse to selectively heat up the die and solder. By controlling the laser power, focus, and pulse width, one can further localize the heating to a fairly narrow region in both the x-y and z dimensions. This allows for independent assembly of different solder bumped areas on the same die or package.

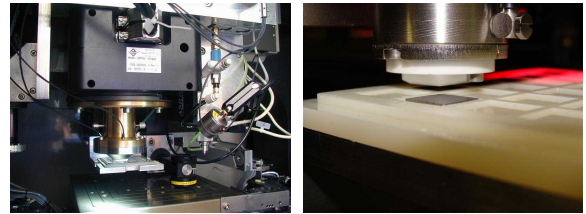


Figure 8. Images of the Laplace laser assembly bond head and tooling.

The LaPlace tool is equipped with a Nd:YAG laser system operating in the infrared, which allows the laser to be directed with little absorption through the backside of the die onto the metals and solders which do absorb this radiation and heat up. The tool uses feedback technology to tightly control the temperature of the die and adjusts to maintain the desired temperature profile. Typical reflow times for most applications are in the millisecond time frame. Figure 9 shows the energy response curves collected by the LaPlace tool that is used to maintain a constant temperature at the die.

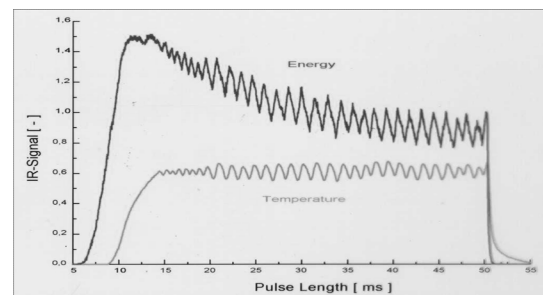


Figure 9: Energy response curve of the LaPlace laser assembly tool.

Two sets of ceramic fixturing were made to pickup and hold one bridge die and one to hold a bridge-island unit. The process flow for assembling the 3 chip unit was as follows:

1. Picking up bridge chip
2. Aligning bridge chip to first island chip
3. Bonding bridge chip to first island chip by laser reflow
4. Picking up bonded bridge-island sub-assembly

5. Aligning bridge-island sub-assembly to second island chip
6. Bonding bridge-island sub-assembly to the second island chip by laser reflow

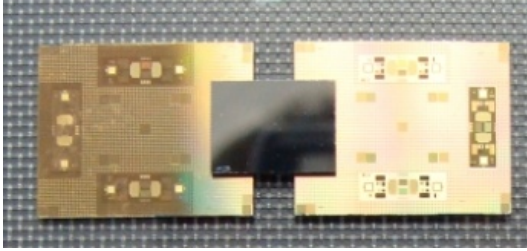


Figure 10. Image of assembled three chip unit: two island die and one bridge die.

Initial work was performed using bridge die which were fabricated using only 3 μ m of Sn as the solder for the micro bumps. Laser assembly of these devices proved very difficult due to poor reflow and wetting characteristics. Studies were undertaken to evaluate the wetting properties of these pads using thermal and laser energy, liquid fluxes, and various reflow atmospheres. All attempts to create good bonds between the bumped bridge die and the island die failed.

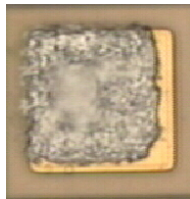


Figure 11. Optical image of Sn solder on Ni/Au UBM showing poor wetting after thermal reflow.

Addition of a thin layer of Au on top of the Sn layer proved essential in creating a thin Sn bump that had good wetting characteristics. This layer was < 500 Angstroms in total thickness. Use of the liftoff technique allowed for easy implementation of this added layer.

Fluxed samples with this thin gold layer showed good pullback onto the pad when reflowed at 245 °C in a nitrogen atmosphere, see Figure 12a. Analysis of this reflowed bump revealed a layer which was nearly 100% composed of a tin-nickel intermetallic.

Unfluxed samples which were reflowed using the LaPlace laser system also showed good pull back onto the pad and good wettability. In this example, only a thin layer of tin-nickel intermetallic was formed, see Figure 12b. This sample exhibited a classic

hemispherical shape, whereas the sample reflowed in the convection oven were essentially flat. The laser reflow in this example is approximately 20 microseconds in duration, which does not create the excessive amounts of intermetallics like the thermal reflow.

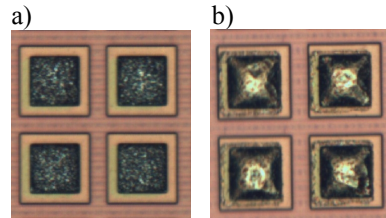


Figure 12. a) SnAu sample reflowed in convection oven, b) SnAu sample reflowed using the LaPlace laser system.

Figure 13 shows an example of an island die (only NiAu UBM) and a bridge die (3 μ m NiAu, 3 μ m Sn, and 500 Å Au) which were laser assembled and then disassembled. Note that there is solder on both die, indicating a good solder joint was formed during laser assembly.

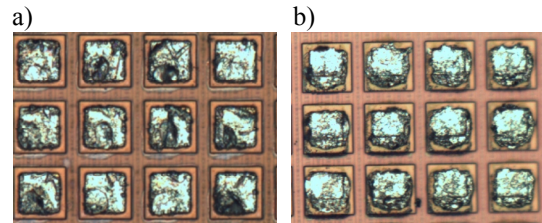


Figure 13. Images of a bridge chip and an island chip that were assembled using laser reflow and then disassembled to show solder wetting properties on all the pads.

Assembly of the 3 die stacked unit onto the ceramic substrate was accomplished using a third fixture to pick up the sub-assembly and align it to the ceramic substrate. Laser reflow was used to selectively heat up and reflow the flip chip size bumps on the island die. Figure 14 and 15 show a bare substrate and an assembled substrate.

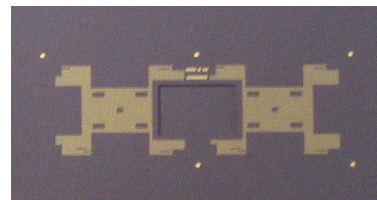


Figure 14. Image of ceramic substrate.

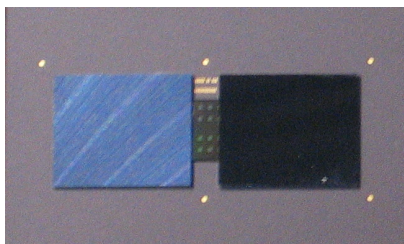


Figure 15. Image of a fully assembled device.

No shorts or opens were observed for the assembled device.

Conclusions:

Micro bumps were created using electroless nickel/gold and sputtered SnAu. The Au layer was essential in order to obtain good wetting properties. Assembly of the bumped die was accomplished using a laser based assembly system (LaPlace™). The advantage of using a laser to reflow the solder instead of direct thermal heating is based on 1) the extremely high selectivity between heating the various die within the 3D stack, 2) the tight control in both the x-y plane and z dimensions, and the tight control on intermetallic formation. Further work is ongoing to characterize the assembled devices and also to optimize the process for volume throughput.

References:

1. T. Sze, M. Gier, B. Guenin, N. Nettleton, D. Popovic, J. Shi, S. Bezuk, R. Ho, R. Drost, and D. Douglas, "Proximity Communication Flip-Chip Package with Micron Chip-to-chip Alignment Tolerances", Proceeding of ECTC 2009, San Diego CA, May 25-29th, 2009, pp. 966-971.
2. Stephen Mick, Dr. Paul Franzon, Alan Huffman, "Packaging Technology for AC Coupled Interconnection", IEEE Flip Chip Conference, July, 2002.
3. Jian Xu, John Wilson, Stephen Mick, Lei Luo and Paul Franzon, "2.8Gb/s Inductively Coupled Interconnect for 3-D ICs", VLSI Symposium, Japan, June 2005.
4. Stephen Mick, Lei Luo, John Wilson, and Paul Franzon, "Buried Bump and AC Coupled Interconnection Technology", Volume 27, IEEE Transactions on Advanced Packaging, February, 2004.
5. A.J.G. Strandjord, M. Johnson, H. Lu, D. Lawhead, R. Hanson, and R. Yassie, "Electroless Nickel-Gold Reliability UBM, Flipchip, and WLCSP, (Part I of III)" ", Proceedings of IMAPS 2000 San Diego CA, October, 2006.
6. T. Oppert, E. Zakel, and T. Teutsch, "A Roadmap to Low Cost Flip Chip and CSP using Electroless Ni/Au", Proceedings of the International Electronics Manufacturing Technology Symposium (IEMT) Symposium, Omiya, Japan, April 15-17, 1998.
7. T. Teutsch, R. G. Blankenhorn, G. Azdasht, P. Penke, E. Zakel, J.-D. Kim, Y.-N. Kim, J.-W. Lee, J.-H. Park, H.-G. Kim and J.-O. Kim, "LAPLACE – A New Assembly Method using Laser Heating for Ultra Fine Pitch Devices" Proceedings of the International Symposium on Microelectronics (36th IMAPS), Boston MA, November 16-20, 2003
8. E. Zakel, T. Teutsch, G. Azdasht, P. Penke, "A New Low Stress and Flexible Assembly Method using Laser Heating for NCP, ACF and Solder Interconnection", SEMICON Europe, Munich, Germany, March 31 - April 5, 2003.
9. Thorsten Teutsch, Elke Zakel, Ghassem Azdasht "Pulsed-laser Heating for Flip Chip Assembly", Advanced Packaging, pp. 40-43, Pennwell Corp., May/June 2006.
10. T. Teutsch, E. Zakel, G. Azdasht, P. Penke, "Laser Assisted Assembly for NCP, ACF and Solder Interconnection", Proceedings SMTA 2003, Oahu, Hawaii, 2004.