Soldering Challenges Caused by Warpage and Deformation of Large-Size Server Integrated Circuits

Qu (Wisdom) Yanhong and Chris Nash
Indium Corporation
Suzhou, China
wqu@indium.com

Ronald C. Lasky, Ph.D., P.E.
Indium Corporation and Dartmouth College
NH, USA
Ronald.C.Lasky@dartmouth.edu

ABSTRACT
With the rapid deployment of 5G cellular networks, artificial intelligence, and increased internet traffic, the highly integrated design of server central processing unit (CPU) and graphics processing unit (GPU) chips that support these technologies have posed a series of challenges to the electronic assembly industry. Among them, problems such as the increase in integrated circuit (IC) package size, printed circuit board (PCB) high density interface (HDI) design, and thermal expansion coefficient mismatch between electronic components and PCBs, have become increasingly severe. These challenges can lead to excessive IC component warpage during the manufacturing process, which can lead to excessive head-in-pillow (HiP) or bridging defects during soldering. This article will explore the challenges posed by warpage of large-size IC packages, typically in the ball grid array (BGA) format, in servers and propose corresponding solutions.

Key words: server, GPU, warpage, head-in-pillow (HiP) defect, bridging, short circuit

INTRODUCTION
As integrated circuit component size, most notably BGA size, continues to increase in order to reduce thermal deformation of these BGAs and their mating PCBs, the industry has begun to adopt low-temperature soldering processes.

One example is bismuth-tin-based low-temperature solder as it has a lower melting point. This lower melting temperature can reduce thermal stress during the soldering process, thereby reducing thermal deformation of BGAs and PCBs. However, it is not known if bismuth-tin soldering processes can solve the problem of HiP defects and/or bridging caused by warpage. This situation raises the question: are there any other options for low-temperature soldering besides tin-bismuth based alloys? This article will seek to address this question by analyzing the process difficulties of large-size BGAs in terms of BGA deformation patterns, soldering challenges, and low-temperature solder characteristics, and proposing optimization solutions.

DEVELOPMENT TRENDS OF BGA SIZES
As IC integration continues to increase, more and more functions and components are integrated into a single IC that is typically packaged in a BGA. The application of multi-core processors and the demand for higher data throughput and memory capacity has driven server application scenarios to require higher data computing power and data processing efficiency as well as larger memory capacity. To meet these demands, ICs in BGA component packages inevitably develop larger sizes, as shown in Figures 1 and 2.

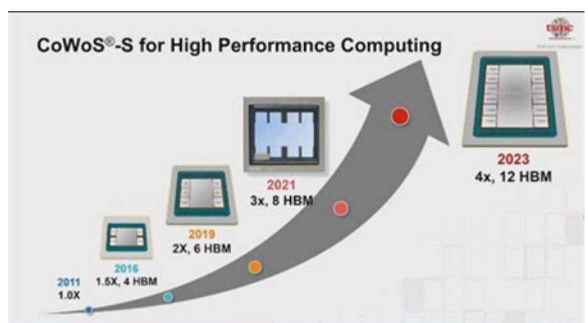


Figure 1. The development trends of high-performance BGAs

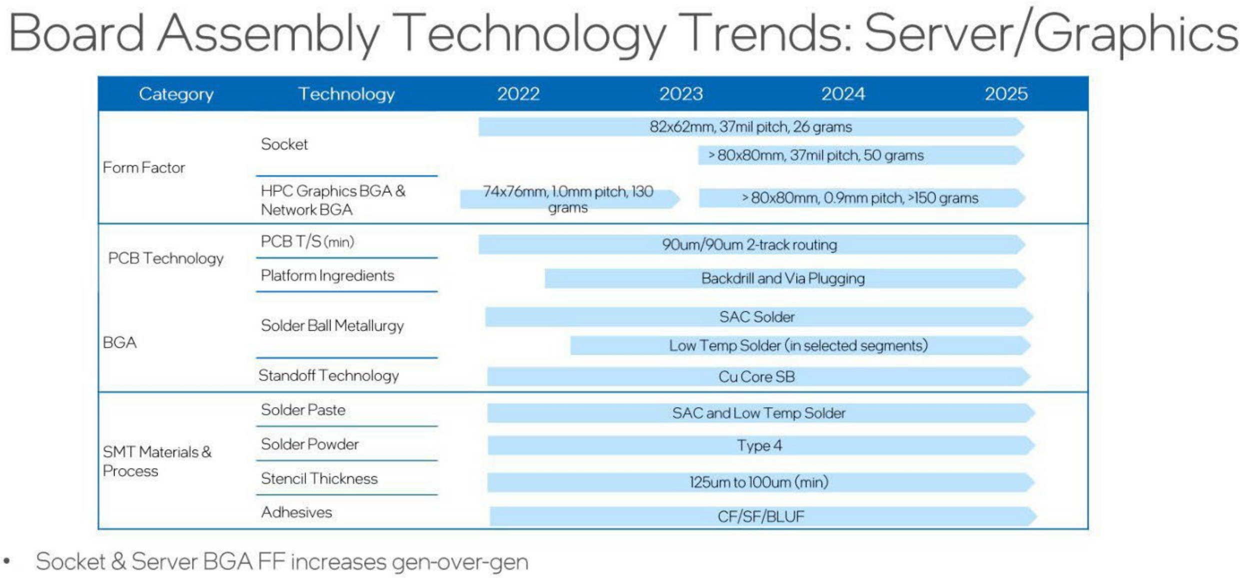


Figure 2. The technology development of Intel server chips.

BGA DEFORMATION MODES DURING ASSEMBLY

Due to the different thermal expansion coefficients of BGA substrates, IC silicon, and plastic packaging materials, the typical challenge brought by the increase in IC component package size is warping deformation during reflow soldering. The deformation of smaller BGAs is generally a "frown" or a "smile", while the deformation mode of larger BGAs becomes more complex. In order to test the warpage deformation of the BGA, shadow Moiré techniques are often used as per JESD22-B112A. To obtain the deformation of the BGA, as shown in Figure 3, the shadow Moiré uses the geometric interference generated by the reference grid and the projection of the reference grid on the sample to be tested. The shadow Moiré then uses a charged coupled device (CCD) to obtain the digital graphic image. The images are processed using a shifting, unwrapping algorithm and shape information extraction algorithm to obtain the BGA package warpage and deformation under different temperature changes that the BGAs are exposed to.

In the paper "Achieving Warp-Free Packaging: A Cover Mold Flip-Chip Packaging Design" published by the Latin American University and NVIDIA [2], one can see the process of BGA type deformation during the assembly processes, as shown in Figure 4.

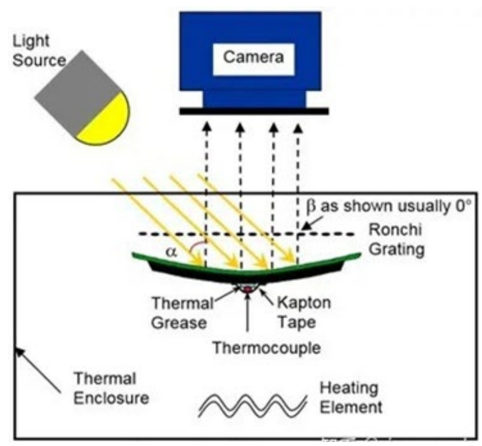


Figure 3. The principles of the shadow Moiré technique.

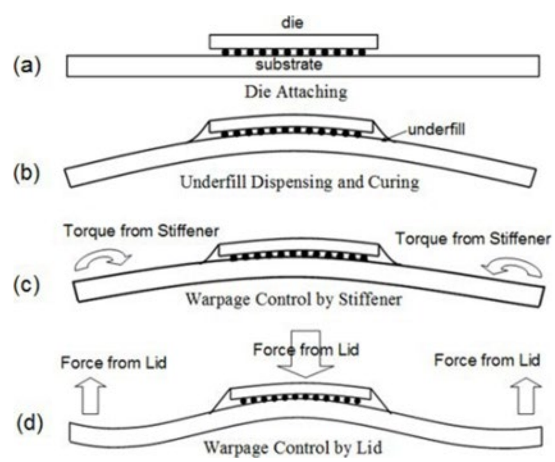


Figure 4. BGA type and PCB warpage during assembly processes.

Experiments Performed

An 80 mm x 80 mm BGA sample was used in this first experiment, and the warpage deformation of the BGA at different temperatures was observed through the shadow Moiré technique. As can be seen in Figure 5, in the initial stage at room temperature, the BGA had the largest warpage, showing a typical frown shape, which is about 500 µm of deformation. As the temperature slowly increased, the BGA exhibited the lowest warpage, hence the highest flatness between 180-217°C. When the BGA temperature was raised to 255°C, the deformation of the BGA showed a larger wave shape. When the temperature was reduced to room temperature, the BGA returned the frown shape and once again exhibited maximum warpage. Due to the large size of the BGA, the warpage pattern was dynamic, making the surface mounting process more complex.

The initial frown or smile deformation can usually be compensated for by increasing the stencil aperture openings to print more solder paste to achieve better soldering results.

However, the deformation of large-sized BGAs is considerable, and increasing the stencil aperture openings may not be enough to solve the problem of soldering defects caused by excessive deformation. Due to the large deformation of larger BGAs, the probability of occurrence of HiPs during the soldering process, and the risk of short circuits (bridging) also increases. To address these potential defects, we need to understand and tackle the BGA/ printed wiring board (PWB) warping process more thoroughly.

One approach to minimize these potential defects is to lower the assembly process temperature. Therefore, tin-bismuth low-temperature soldering processes have been considered. Most tin-bismuth low-temperature alloys melt at close to the eutectic temperature, and the melting point range is usually between 138°C and 146°C. The peak temperature of low-temperature reflow soldering is generally below 180°C. This peak temperature is almost 50°C lower than SAC solders. This lower peak temperature helps to reduce thermal deformation of the BGA and the PCB, thereby potentially avoiding open and short circuits caused by warpage deformation.

In a second experiment, we used a BGA with dimensions of 80 mm x 80 mm x 4.8 mm, the PWB pad size was 0.45 mm, the pitch was 1.0 mm, and the PWB board size was 525 mm x 400 mm x 3.2 mm.

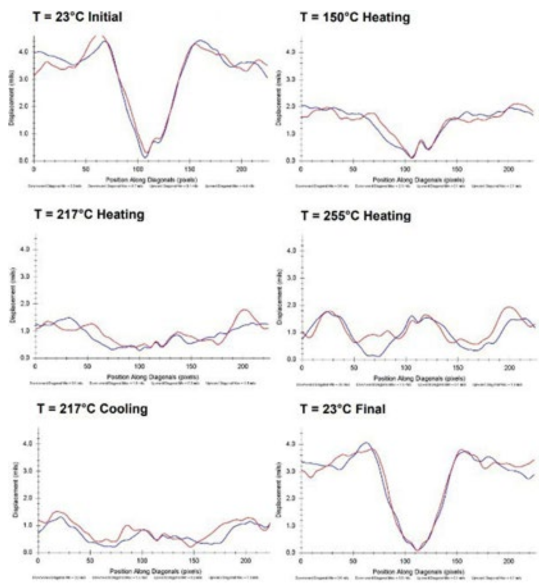


Figure 5. BGA warpage at different temperatures.

During the experiment, we selected three different types of low-temperature solders and SAC305 as samples. Among them, solder pastes A and C were tin-bismuth eutectic alloys containing 56% to 57% bismuth, with the balance of tin, and around 1% of silver and some proprietary dopant elements.

Solder paste B was a non-eutectic alloy containing 50% bismuth, with the balance of tin, and 1% of silver and proprietary dopant elements. All three low-temperature solders use peak temperatures below 200°C for reflow soldering. The height and shape of the solder joints measured by cross sectioning is shown in Table I and Figure 6.

Table I. Table displaying soldering heights of different low-temperature solders.

Paste Type	Melt Point (C°)	Peak Temp (C°)	Standoff (µm)		
			L	M	R
A	138-144	190	553	618	557
B	138-170	190	588	645	584
C	138-148	190	562	620	560
SAC305	217-220	245	506	475	490

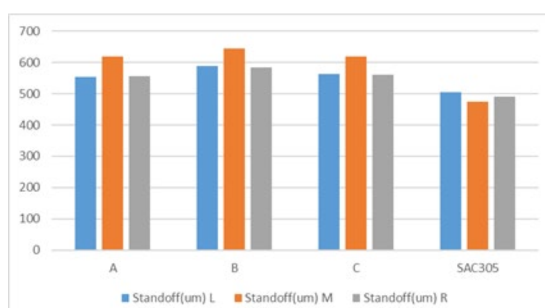
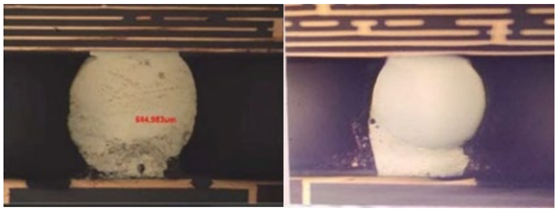


Figure 6. Graph displaying soldering heights of different low-temperature solders.

In the experiment, the three low-temperature solders all exhibited frown-shaped warpage on the BGAs; that is, the warpage is highest in the middle and lower around the edges. SAC305 samples exhibited the BGA warped in the shape of a smile, that is, the warpage is smallest in the middle and largest around the edges. This is similar to what was previously observed using shadow Moiré analysis.

At the same time, during the cross-sectioning process, we found that the solder joints had HiP and thermal tears when using low-temperature solders, as shown in Figure 7.



a) b)

Figure 7. a) A micrograph showing HiP defect and b) A micrograph showing tears, both in low-temperature solder assembly.

As for the HiP defect with low-temperature solders, the main reason it exists is as shown in Figure 8. At room temperature, the BGA already has a certain level of warpage. The tin-bismuth-silver low-temperature solder paste alloy cannot fully contact the BGA solder ball, as the initial frown BGA warpage is not eliminated at the low temperatures in reflow. Even if the solder has reached the liquidus point, it cannot form a solder joint with the BGA bump, which will be the main reason for HiP defects.

Figure 8. BGA warpage simulation diagram.

Figure 7 shows a HiP solder joint and fracture of the IMC layer after low-temperature soldering. It has been pointed out that the temperature difference between the solidus line and the liquidus line of non-eutectic tin-bismuth silver low-temperature solder is large. When the BGA or PCB is cooled, they tend to warp, and the solder joint is pulled apart. This situation is typically referred to as hot tearing. However, in this experiment, solder AB is a eutectic alloy, so why does hot tearing also occur? This hot tearing is mainly due to the fact that BGA solder balls are usually a tin-silver-copper (SAC) alloy. When the diffusion reaction occurs with the SAC alloy, more tin from the BGA balls enters the tin-bismuth-silver solder, causing the melting point of the alloy to change. The eutectic phase becomes a non-eutectic phase. The non-eutectic phase is more susceptible to hot tearing due to it being liquid over a large temperature range.

The BGAs, when using the SAC305 solder paste, were found to have a short circuit after soldering, and a HiP defect was observed through cross-sectioning. The short circuit mainly occurs in the middle of the BGA. From the relationship between warpage and temperature, Figure 9 shows that the BGA produces a smile type warpage at 245°C [2]. The BGA ball places excessive force on the solder paste in the middle of the BGA and creates bridging which will lead to a short circuit defect. On the edges of the component, the solder balls warp away from the board and solder paste, and an open circuit is the result in the form of HiP defects.

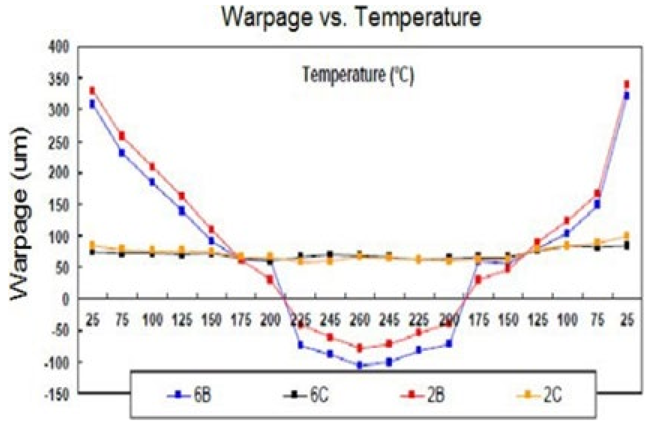


Figure 9. The relationship between warpage and temperature

MEDIUM TEMPERATURE SOLDER AND LARGE SIZE BGA SOLDERING

The warpage of BGA components becomes greater as the size increases. Often, the warpage is controlled by adding fixtures on the BGA. This not only increases the manufacturing cost of the BGA, but also has an effect on the wave-shaped warpage. Since the deformation of BGAs is smoothest at 200-220°C, this situation raises the question: is there a solder that can achieve soldering within a 220°C peak temperature?

To answer this question, another experiment was performed with a BGA of 80 mm x 80 mm x 4.8 mm, while the pad size was 0.45 mm, the pitch 1.0 mm, and the PCB board length was 525 mm x 400 mm x 3.2 mm.

We then employed a mixed alloy powder solder paste, which is a mid-temperature alloy system. This solder paste technology incorporates a low-melting point indium-based alloy and a higher-melting point tin-silver-copper alloy, which are mixed together with a flux vehicle. The indium-based alloy provides a lower melting point, and the tin-silver-copper alloy provides increased solder joint strength. With its melting range between 187°C and 206°C, the mixed alloy powder solder paste can achieve a minimum peak temperature reflow of 200°C, if this low a temperature is needed. After reflow, the solder joints have a high resistance to drop impact and thermal fatigue as shown in Figures 10 and 11.

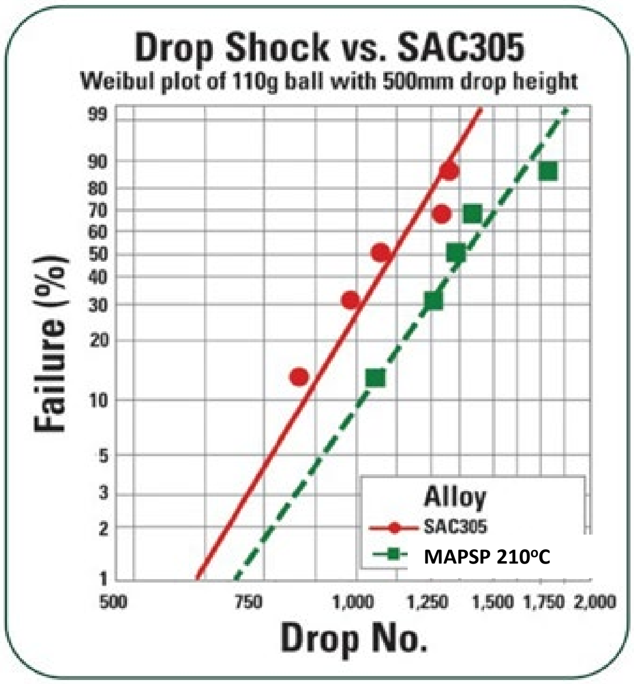


Figure 10. Drop shock performance of the mixed alloy powder solder paste.

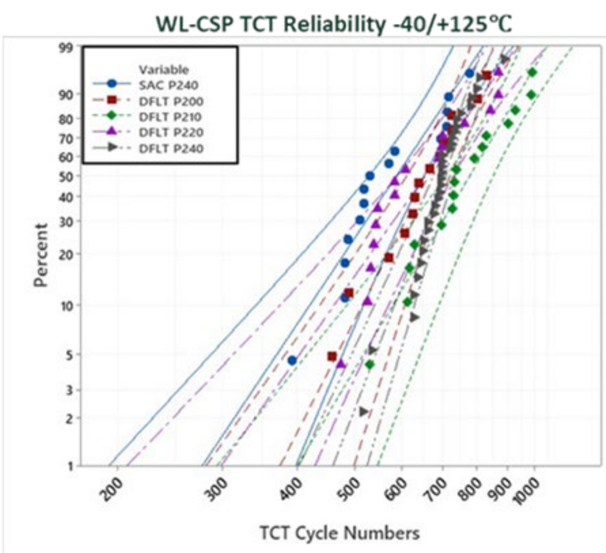


Figure 11. Thermal cycle performance of mixed alloy powder solder paste.

The reflow oven profile temperature curve is shown in Figure 12. Considering that the BGA shows minimal warpage at a temperature of about 220°C, this peak reflow soldering temperature can effectively reduce non-coplanarity issues during soldering.

In addition, the melting point of BGA balls is 217-220°C. So, we set the peak temperature during reflow to 220°C to ensure that the solder balls and solder melt and completely homogenize.

Zone	1	2	3	4	5	6	7	8	9	10	11	12
Top	210	210	250	250	300	280	250	240	230	230	230	230
Bottom	210	210	250	250	300	280	250	240	230	230	230	230
Conveyor Speed (cm/min): 80.0												

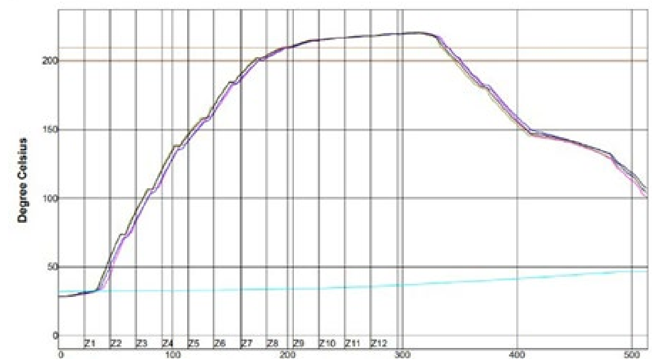


Figure 12. The reflow oven temperature profile for the experiments with mixed alloy powder solder paste.

After the soldering was completed, we observe the shape of the solder joints through cross-sectioning, as shown in Figure 13. The height and shape of the solder balls was relatively consistent and regular.

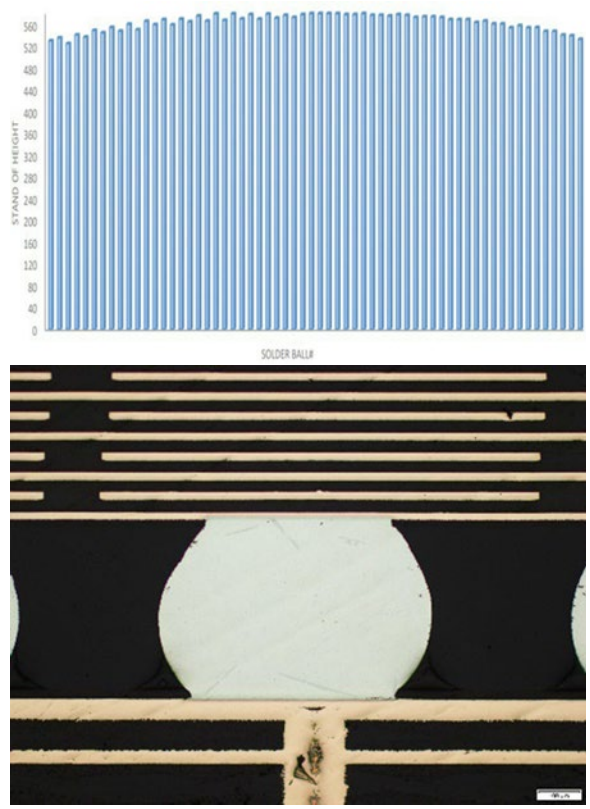


Figure 13. Solder ball height and shape.

CONCLUSION

In order to meet the high-performance, high-density integration, and heat dissipation requirements of servers, BGA component size is getting larger and larger. To minimize warping of the BGA and mating receiving PWB, low-temperature tin-bismuth solders have been considered.

In our experiments, however, we found that BGAs exhibit minimum warpage at a reflow temperature of 200-220°C, not the lower temperature of about 180°C that tin-bismuth solders require. Therefore, the best process production window can be obtained by using a solder alloy with a melting point of 200°C and producing it under a peak reflow soldering curve of 220°C. The mixed alloy powder solder paste is such a solder. In addition, the mixed alloy powder solder paste has outstanding drop shock and thermal cycle performance, making it an ideal solder for large BGA components.

REFERENCES

[1] S. Perng, X. Weidong, "Warpage Studies on Large BGA Site and Corner Bridging Mitigation," SMTAI 2014, Rosemont, IL.
[2] Y. Shen, L. Zhang, and X. Fan, "Achieving Warpage-Free Packaging: A Capped-Die Flip Chip Package Design," IEEE Electronic Components and Technology Conference, 2015.
[3] https://www.youtube.com/watch?v=BN3
DGC gDls4
[4] Hongwen Zhang, etal, patents US 11712762B2
And US11267080B2

APPENDIX

Low-Melt, High-Re-melt Solder Alloy: The Mixed Alloy Powder Solder Paste

Even among some knowledgeable people, it is not widely known that some metals can "melt" or "dissolve" other metals at room temperature. A classic example is mercury dissolving gold (see Figure 14). This phenomenon is especially impressive considering that gold melts at 1,064°C. This phenomenon is used in artisanal gold mining. The miner will mix mercury will soil that contains gold dust. The mercury will dissolve the gold, and since mercury is heavy, the mercury with dissolved gold will sink to the bottom of the container. The mercury is then separated by boiling it away to release the gold. Unfortunately, the mercury vapors released are often breathed by the miners, causing lasting health issues.

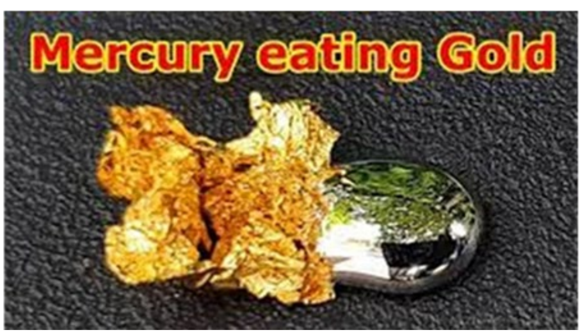


Figure 14. Mercury dissolving gold. This phenomenon is used in artisanal gold mining [3].

Hongwen Zhang and his collaborators exploited this concept to develop a new solder alloy that melts at a low temperature, in the 200-210°C range. See Figure 15 and Figure 16. This mixed alloy powder solder paste system uses a patented combination of powdered SAC alloys and a powdered indium alloy [4].

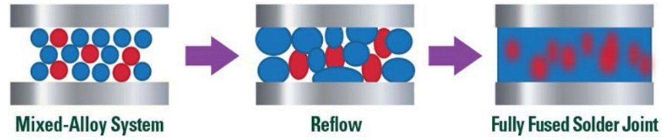


Figure 15. The indium alloy (blue) melts at a low temperature during reflow.

In its melted stage, it partially dissolves the SAC alloy (red). Upon cooling, a fully-fused solder joint is formed. The kinetics of this solder paste system is shown in Figure 16. After the first reflow, this resulting alloy has a high remelt temperature and has superior mechanical and reliability properties to SAC305 (see Figures 10 and 11).

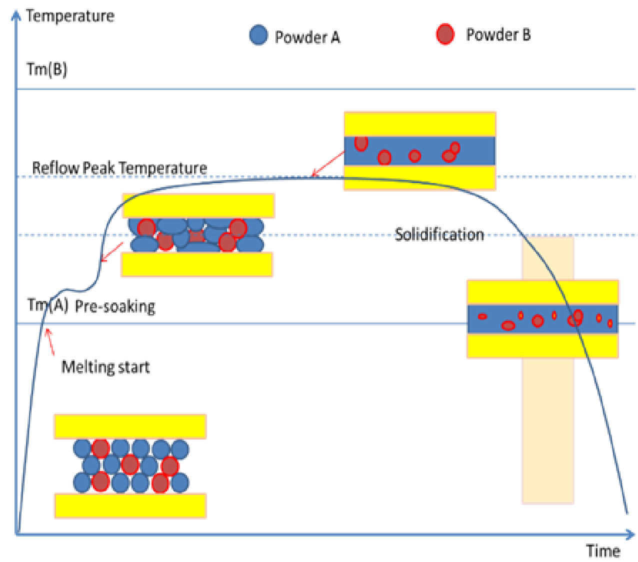


Figure 16. The kinetics of the alloys in the mixed alloy system during reflow.