Advanced Second Level Assembly Analysis Techniques - Troubleshooting Head-In-Pillow, Opens, and Shorts with Dual Full-Field 3D Surface Warpage Data Sets

Ken Chiavone
Akrometrix
Atlanta, Georgia

Abstract
SMT assembly planning and failure analysis of surface mount assembly defects often include component warpage evaluation. Coplanarity values of Integrated Circuit packages have traditionally been used to establish pass/fail limits. As surface mount components become smaller, with denser interconnect arrays, and processes such package-on-package assembly become prevalent, advanced methods using dual surface full-field data become critical for effective Assembly Planning, Quality Assurance, and Failure Analysis. A more complete approach than just measuring the coplanarity of the package is needed. Analyzing the gap between two surfaces that are constantly changing during the reflow thermal cycle is required, to effectively address the challenges of modern SMT assembly.

To fully understand and characterize variation in the gap that is the main cause of solder joint defects such as head-in-pillow, shorts, and opens, analysis of the interface between the mating surfaces needs to be:

• **Full-field:** A high density of data represents surface warpage much more finely than an approximation such as coplanarity, allowing area-specific review and analysis

• **Dual-surface**: Making assumptions about how the "bottom" surface is shaped when mounting a component neglects the complex behavior the land area might exhibit during the thermal cycle, and is a deficiency when attempting thorough assembly planning or failure analysis

• **Full-profile:** Making assumptions about when the single temperature when warpage of surfaces is important during the thermal cycle can lead to missing critical parts of the reflow process, overlooking when defects such as head-in-pillow can be caused

• **Statistical:** A large enough quantity of samples should be measured to allow confident calculation of expected average, maximum, minimum, and extreme (3σ) gaps between surfaces. After measuring warpage of multiple surface mount components and land areas separately, the collected data can be combined into statistical summaries for each temperature point.

Reviewing the combined data at the start of assembly planning can provide an overview of dual-surface warpage at each temperature, and for the entire thermal profile. The measurement and analysis results that follow include a package-to-board assembly interface case study, and calculations, graphs, and methodology highlighting the use of gap limits and pass/fail maps to visualize areas with potential assembly issues.

This analysis method provides new capabilities when planning and monitoring the assembly interface across a full reflow cycle that can help predict and compensate for defects such as head-in-pillow.

Introduction
The APEX 2009 publication "Telecommunications Case Studies Address Head-in-Pillow (HnP) Defects and Mitigation through Assembly Process Modifications and Control" is one of the most thorough published reviews of the head-in-pillow defect problem and how those responsible for troubleshooting poor yields analyze the problem and mitigate its severity. [1] In the past few years since it was published, the problem of head-in-pillow, or head-on-pillow as it is also called, has become more widely documented and discussed. It continues to grow as a concern and a detriment to both initial yield and end-user field reliability, as trends in electronics design push the limits of component size, thinness, and materials, with ever-decreasing interconnect pitches and solder ball sizes. All these forces make the reliable assembly of parts during the reflow process more challenging. The following is a review of a relatively new dual-surface analysis method that electronics manufacturers are now using to more fully understand component shapes throughout the reflow process, during the pre-production design, assembly planning and troubleshooting, and quality assurance phases of product development and manufacturing.

Warpage Standards
One of the major causes of head-in-pillow defects, opens, and shorts/bridging, is component warpage.[2] Since packages began to be manufactured without ceramic substrates, warpage has become an important factor when assembling BGA devices, both die-to-substrate, and second level package-to-PCB. The effects of warpage have continued to increase. One of the biggest product design, production yield, and reliability issues is warpage, since excess warpage and even small variations in the amount of warpage can cause an unacceptable defect rate.[3]

For many years, coplanarity has been the measure used to indicate warpage, or deviation from flatness, in the electronics industry. Simply stated, coplanarity is the distance between the highest point and the lowest point measured. Optionally, a '-' or '+' sign may also be assigned, indicating a 'smiling' or 'crying' shape, whether the shape is concave or convex when looked at from the side.[4]

All the related industry standards in wide use today deal specifically with package warpage. How to measure it, and in some cases the allowable limits a package can demonstrate and still assemble reliably.[5] Individual companies also have their own internal, proprietary limits for components, in many cases. These limits on coplanarity are derived through calculations, experimentation, or both, and given certain process controls they serve to 'ensure' an acceptable production yield.

But coplanarity limits have their own limitations. There is frustration in the industry because, while a standard that indicates a method for measuring warpage throughout the reflow cycle may be universally applicable, the standards that set specific numerical acceptance limits (such as 'package coplanarity must be less than 100 microns') can be useless or even harmful to decision-making in some cases.[6] It is very likely that an assembly of components demonstrating a coplanarity within 'standard limits' could fail in certain cases, and also likely that packages with more than the established coplanarity can assemble with zero defects. Those responsible for assembly, with millions of production units to assemble and high-pressure deadlines, cannot rely on the current published or proprietary standards to make the best decisions related to component shape compatibility and expected reliability.

Why traditional coplanarity limits are too limited
There are several reasons that setting coplanarity limits doesn't work as well as those setting the limits intend. Some of them are:

- The limits are generated from a relatively small set of components, under controlled conditions. All the variables in the process are not recorded, such as exact flux types and equipment settings, but the limits are derived from work done within those controlled conditions.
- The limits assume a low, symmetrical warpage from the land area (such as a PCB surface). Standards may indicate a warpage 'budget' for the land area of a small percentage of the total warpage, such as 20%.[7] It is common for a BGA land area to demonstrate much higher warpage than this limit, even with conventional PCB materials.[8] Newer 'cutting-edge' materials tend to demonstrate even higher warpage than standard materials.
- The limits assume peak warpage/coplanarity is the important measure, and that it occurs at a particular temperature point, such as peak temperature. Defects, especially head-in-pillow and non-wet-opens, can develop at critical points during the reflow cycle other than at peak temperature. The solder ball oxidation that results in head-in-pillow defects, for example, happens earlier in the cycle.[9]
- The limits address the warpage of a single surface, not the shape compatibility of the two surfaces that are actually attaching during assembly. With a single surface, knowing the gap that will exist at an interconnect location at a particular temperature point, which is critical to predicting reliability, is impossible.

'Gap per Interconnect' analysis prerequisites
Head-in-pillow and most other assembly defects traditionally associated with warpage are caused by the gap between attaching surfaces at each interconnect, at different points during the reflow assembly process. For reasons including those listed above, a thorough analysis of SMT component assembly must include identifying these gaps. To be effective, the method needs to be:

- **Dual-surface** – for anyone who wants the true gap at each interconnect, making assumptions about half of the attachment area is not sufficient. Both surfaces must be measured, then the two surfaces must be analyzed together in a shared (x,y,z) coordinate system.
- **Full-field** – modern methods of measuring warpage typically include optical technologies with a fairly high data density, usually greater than 1 point per square mm. This is worth listing as a requirement though, since surfaces can have complex shapes, and checking only a few points, such as at the corners and center, cannot register shapes sufficiently to reliably establish the gap at each interconnect. Data density must be sufficient to represent the full-field shape of the attach area.

- **Full-profile** – understanding the wide range of defects that can happen due to gaps during the reflow process requires analyzing gaps at various temperature points throughout that process. In most cases, evaluating gaps at multiple temperature points is necessary for thorough understanding of how the surfaces interact.
- **Statistical** – enough samples must be measured, of both attachment surfaces, to be confident that they represent a statistically relevant sampling of the component population. When checking gaps at the interconnect locations, it is useful to have surfaces that represent the maximum gap expected, minimum gap, average gap, and even statistically calculated gaps such as average gap + 3*(standard deviation of the gap) for each interconnect. Depending on the intent of the analysis, or the nature of the troubleshooting issue, different statistical gaps could be of more use than others.

Methodology and Data

For best results, dual-surface gap analysis requires sets of two surfaces, both Top Surfaces and Bottom Surfaces, which are the measured shapes of the physical sides of the components that attach during SMT reflow assembly. Ideally these will include many representative samples measured at all relevant temperature points in the reflow cycle, for both Top and Bottom surfaces. What follows is an example of how surface data can be analyzed when looked at together, in a shared coordinate system. The data shown comes from real, production components measured during separate reflow cycles. The Top Surface is from a conventional BGA with solder balls removed to provide a smooth, continuous surface. The Bottom Surface is a PCB land area from a production computer. (Note that this is an example illustrating dual-surface analysis methodology, but the surfaces do not represent an actual production assembly. These particular packages and PCBs are not necessarily intended to be assembled together.)

When performing dual-surface analysis with the intent of planning or troubleshooting assemblies, reviewing actual measured surface pairings (the measured shapes of a single top component and a single land area) is of limited use. Because defects tend to happen at the extremes of mechanical behavior, evaluating combinations of statistically-calculated shapes is a more streamlined, efficient method of reviewing shape compatibility throughout reflow assembly.

Measurement Technique

Shadow moiré measurement equipment was used to acquire the full-field surface data shown below. To remove the offset angle from perfectly level from each component, due to its position during measurement, a least squares fit (LSF) rotation was applied, so each surface data set is rotated to a common coordinate system. This process ensures the actual shape of the surface, not including its angle sitting in the oven, is being analyzed. Figure 1 illustrates the least squares fit rotation process.
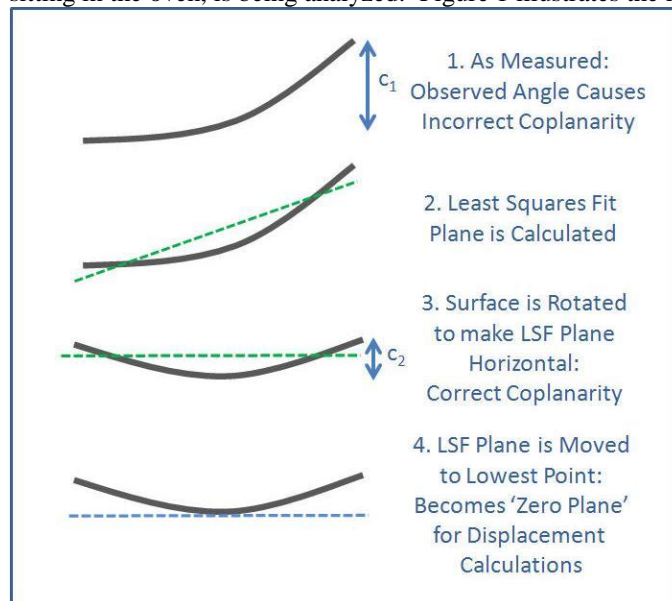

1. As Measured: Observed Angle Causes Incorrect Coplanarity c_1

2. Least Squares Fit Plane is Calculated

3. Surface is Rotated to make LSF Plane Horizontal: Correct Coplanarity c_2

4. LSF Plane is Moved to Lowest Point: Becomes 'Zero Plane' for Displacement Calculations

Figure 1: Least Squares Fit Rotation Process

Sample sets

For this example, 5 temperature points, each consisting of 10 Top Surfaces and 10 Bottom Surfaces, are shown. In separate thermal runs, the Top and Bottom samples were measured at the same temperature points during the reflow profile.

Mathematical Analysis

To compare the surfaces, both within their surface sets, and between the Top and Bottom surfaces, a new measurement convention was needed. Each LSF plane used to rotate a surface to a shared 'level' to establish coplanarity was then moved mathematically to the lowest point on the surface (for the Top Surfaces) or to the highest point of the measured surface (for the Bottom Surfaces). This becomes the 'Zero Plane' for each surface, and allows identification of which interconnect has 'more gap' or 'less gap' based on the vertical distance from the zero plane to that interconnect's location on the surface. Displacement is the term used to describe the 'distance from the zero plane' measured vertically for any (x,y) point analyzed.

The figures below show what look like normal surfaces, but they don't necessarily represent any particular surface measured. Instead, they are statistical representations of what was measured for each XY location in the measured area. For example, for point (4, 22) there will be a maximum displacement from the zero plane, for all the surfaces in the Top data set. If that maximum displacement is 85 microns, the Maximum statistical surface for the Top Surfaces will have an (x,y,z) point at (4, 22, 85 microns). The point beside it, (4, 23) may derive its z-coordinate from the same surface as (4,22) did, or from any other surface in the Top data set, whichever has the maximum z-displacement from the zero plane for that (x,y) coordinate.

Statistical surfaces calculated
Minimum surface – composed of (x,y,z) points where z is the minimum displacement from the zero plane for each (x,y) point across all surfaces in the data set compared.

Maximum Surface – composed of (x,y,z) points where z is the maximum displacement from the zero plane for each (x,y) point across all surfaces in the data set compared. (For the Bottom surfaces, Maximum means 'more negative', since displacement distance is always positive.)

Average Surface – composed of (x,y,z) points where z is the average displacement from the zero plane for each (x,y) point in the data set compared.

'Three Sigma' Surface – composed of (x,y,z) points where z is the average displacement from the zero plane plus 3X the standard deviation of the displacement, for each (x,y) point in the data set compared.

Following the above method, a total of 8 statistical surfaces were generated, 4 for the Top Surface and 4 for the Bottom Surface, for each temperature point.

(Note that for this example, solder balls have been removed from the Top Surface components. Because solder ball placement consistency is well controlled, the solder ball-free surface can represent the surface that would be created from the extreme points of the solder balls, for most purposes. That is, the surface without solder balls is basically the same shape as the effective surface at the interconnect locations would be with solder balls in place, just offset by the diameter of the solder ball, which has no effect on gap calculations. The visual and mathematical analysis is greatly simplified and is more easily understood using two continuous surfaces rather than the complex surfaces that solder balls, pad, and paste would create, so the clearer approach is presented here.)

Moving the surfaces together
There are options for how to move the surfaces relative to each other, once they are displayed with a shared coordinate system, on the same graph. In this example, the surfaces are moved toward each other in the vertical, z-direction until they touch at one point. In some cases, such as when the top and bottom touch in the middle at the closest point, this results in a configuration that looks 'right' and not off-balance. In other cases, such as when the surfaces meet at just a corner, the surfaces look 'off-balance'. While there are other z-orientation methods that can be used for deeper analysis of more complex shape combinations, that discussion is outside the scope of this work. For most analyses, even when they look odd when viewed together, orienting surfaces with a single point touching can provide useful, actionable results.

Gap Calculation
Once all this mathematical setup is complete, gap calculation is simple. The gap at any (x,y) location is the vertical distance between surfaces at that point. Since interconnect failure can happen with a gap between paste and solder ball during reflow of just 25 microns, or even less in some cases, the difference between a good joint and a defect can be a very thin margin.[10] Checking the gaps for all points in the BGA attach area, for different scenarios, is possible by changing the Top and Bottom statistical surfaces that are brought together for gap calculation. For a combination of components that demonstrate inverse warpage, such as concave for the top surface and convex for the bottom surface, that touch in the middle, choosing to evaluate the Maximum Top Surface and Maximum Bottom Surface together would allow estimation of the maximum gap possible for the measured components, around the outside edges. Similarly, evaluating the Minimum Top and Minimum Bottom statistical surfaces together would illustrate a situation where the two component attach areas are as flat as they can be at each (x,y) point, showing the least gap that would be present in the measured sample sets.

Results

The example results that follow illustrate reviewing the gap values at various temperature points, and extracting the cross-sectional gap along various chords (linear paths of interest during analysis). This in-depth analysis provides information not available using traditional single-surface analysis methods. Dual-surface analysis is useful for aspects of design, assembly planning and troubleshooting, and quality assurance.

Figure 2 shows the basic dual surface graph layout. The top 3D set of points is a statistical surface representing the inside (attach) surface of the upper surface mount component, in this case a package with solder balls removed. The bottom 3D set is the corresponding BGA land area, also represented by a statistically-derived surface. To better see both surfaces, a not-to-scale visual gap is introduced that does not affect gap calculations.
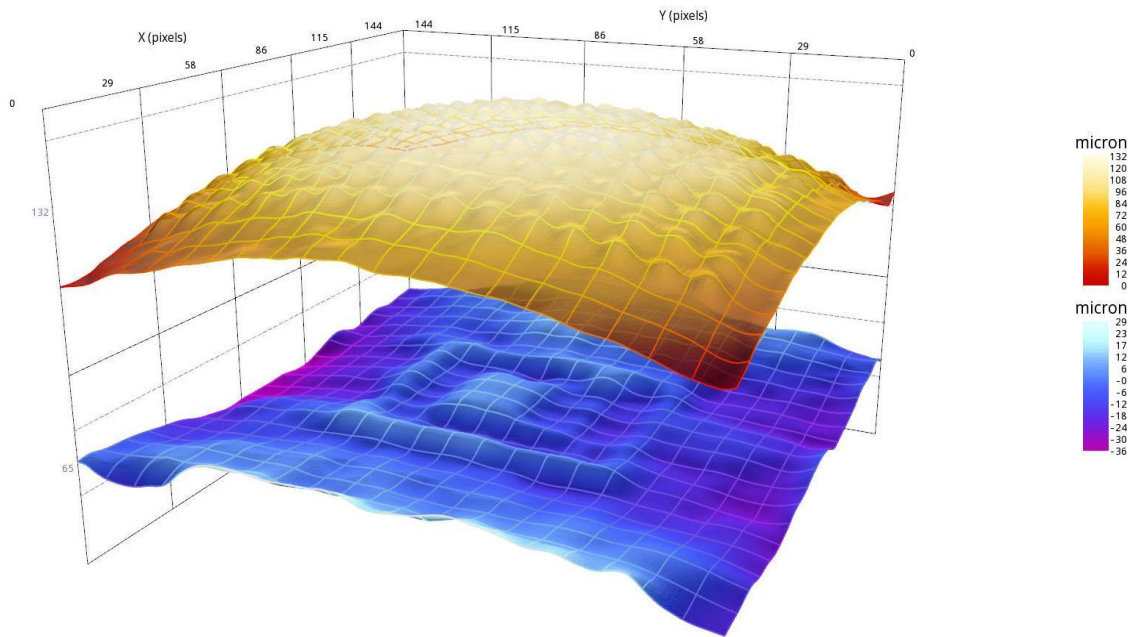
Figure 2: Dual Surface 3D Graph at Room Temperature

Figure 3 shows 'chords' on the surfaces. Lines of data that can be extracted from the 3D data set and plotted in 2D. (Chord plots are not affected by any visual gap shown in the 3D graphs. Actual gaps
are shown in the 2D plots.)
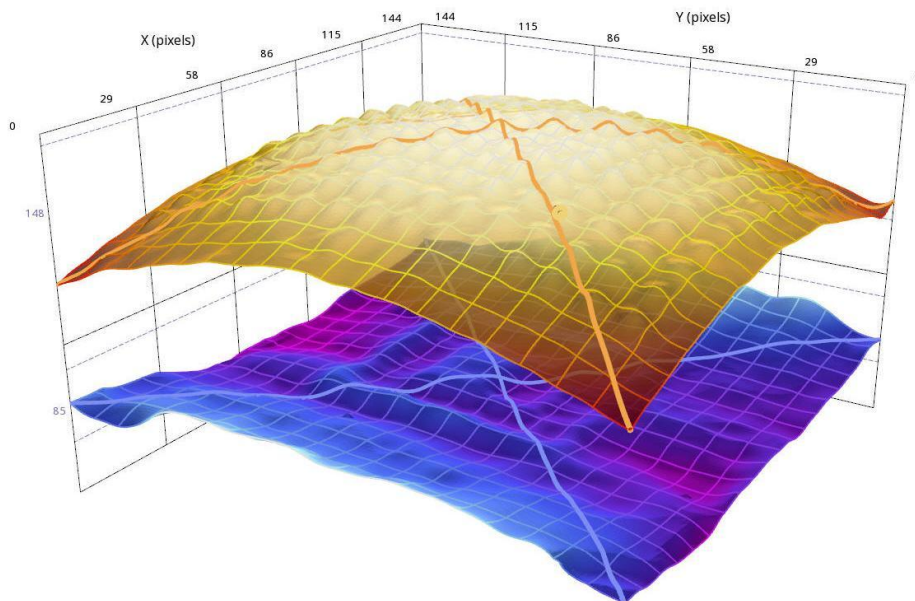
Figure 3: 3D Graph with Chord Lines Shown

Figure 4 shows 2D chord data extracted from the 3D data sets in Figure 3.
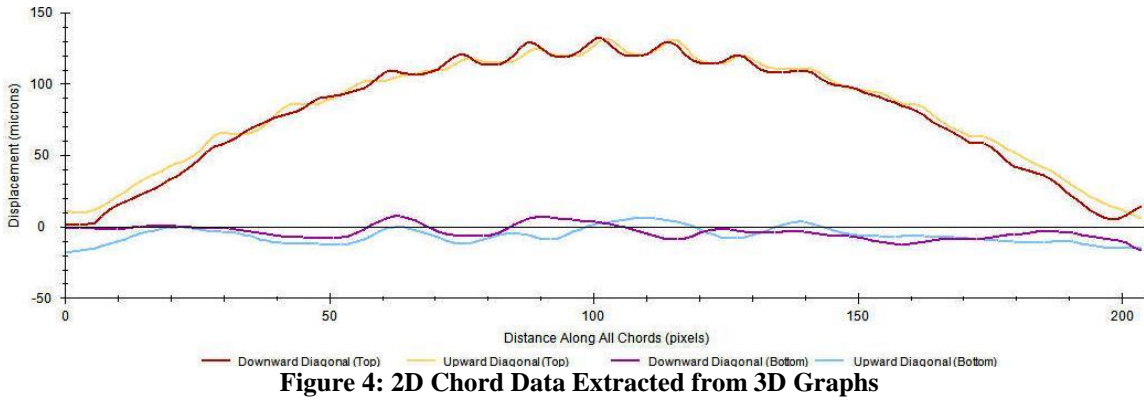

Figure 4: 2D Chord Data Extracted from 3D Graphs

Figure 5 shows the 3D and 2D chord data from the samples at RT, before the reflow profile starts. A maximum gap of approximately 125 microns occurs near the center of the measured area.
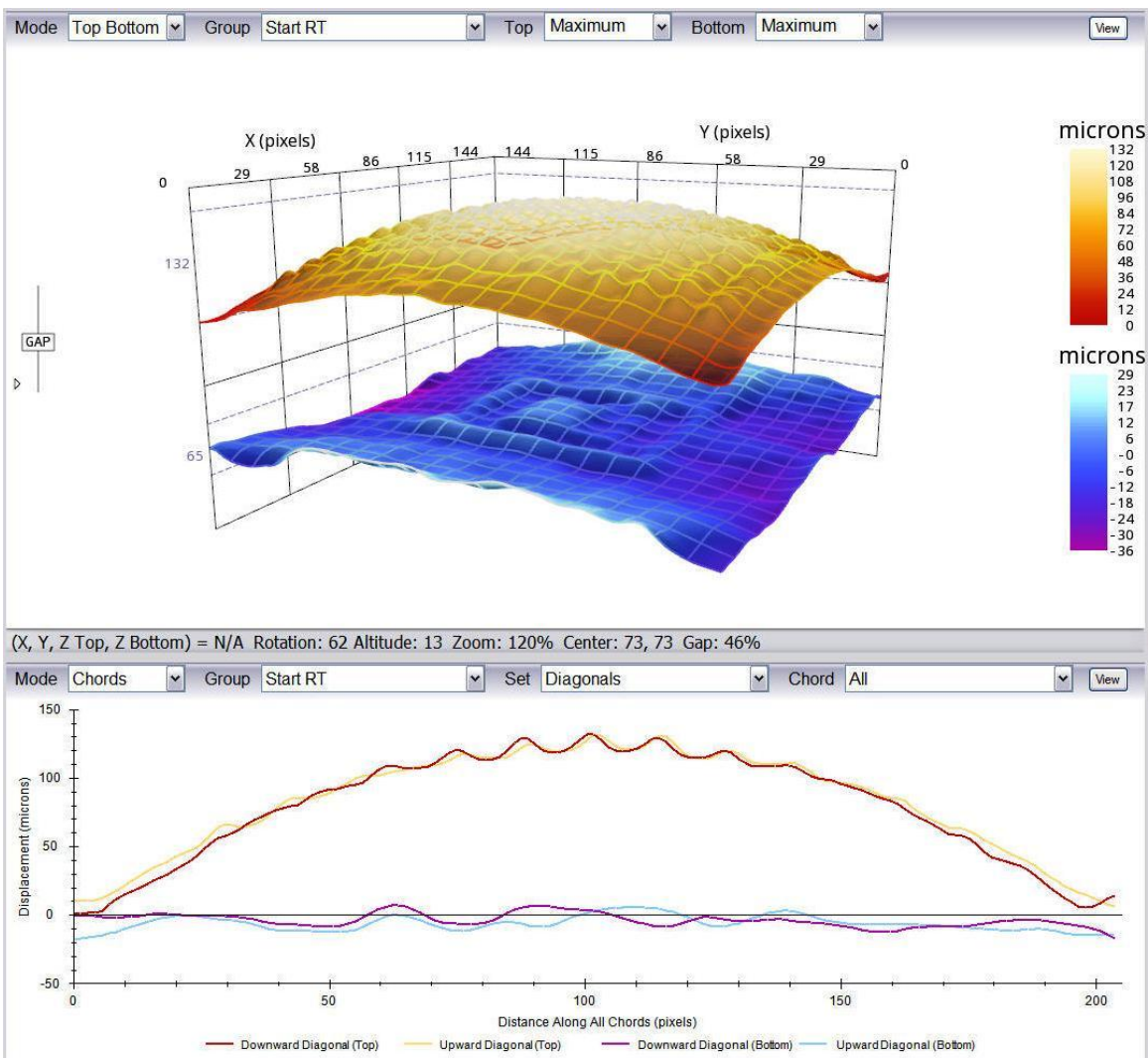

Figure 5: Room Temperature Start

Figure 6 shows the samples in the heating stage of the profile. At 160°C, the package flattens, reducing gaps in the BGA area to about 50-75 microns. During this heating and soak phase, head-in-pillow tends to begin to occur due to oxidation, if there is an excessive gap between solder balls and the paste on the land area, so warpage at these temperatures is important.
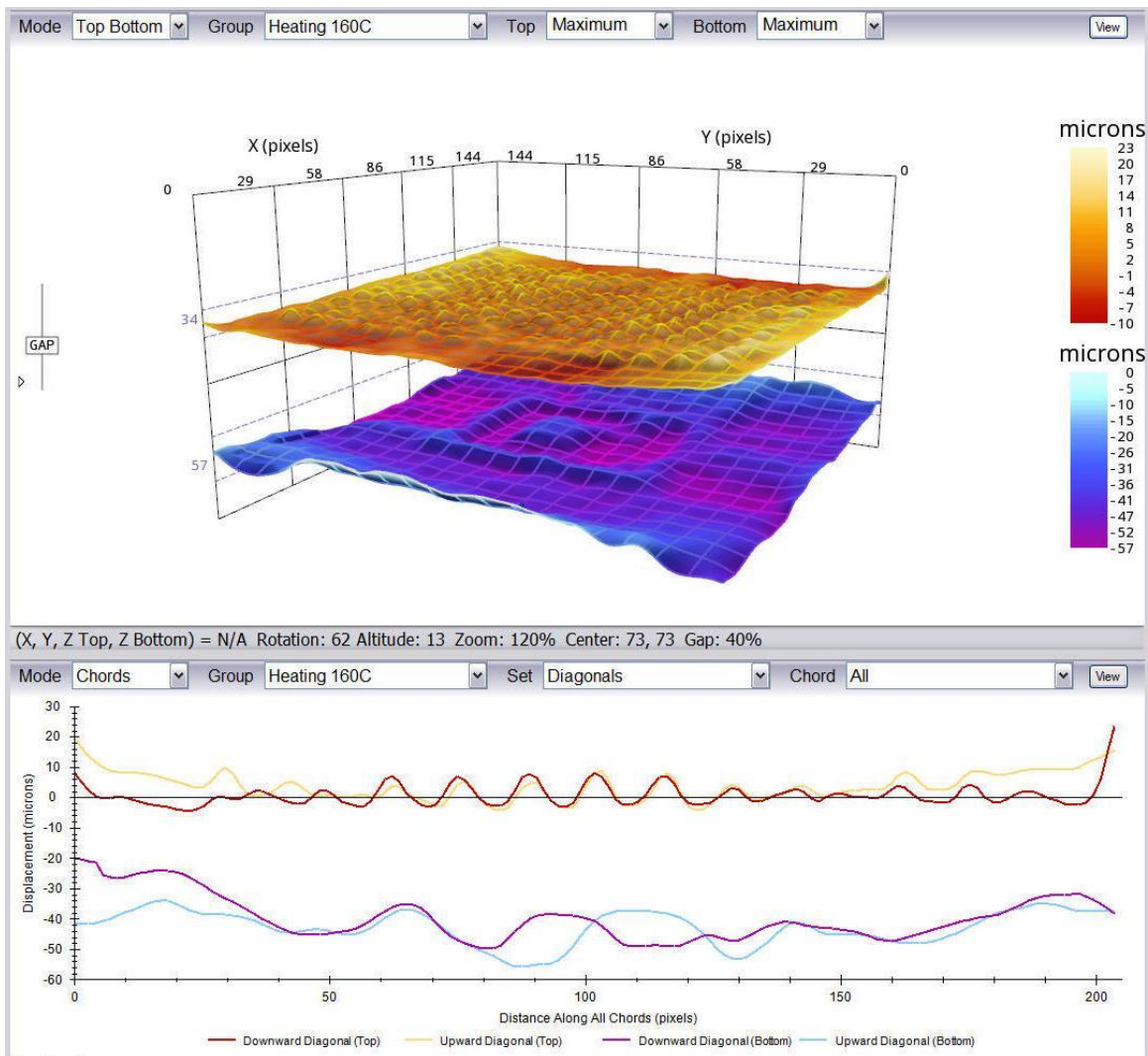
Figure 6: Ramp Up Heating Phase

Figure 7 shows the Maximum Top and Maximum Bottom statistical surfaces at 260°C peak temperature. 'Open' defects of various types tend to form around the peak temperature, so gaps at this phase are important and can inform decision-making about process control variables. In this example, the package and PCB area shape compatibility is very good, with an approximate gap of 50 microns overall, and an additional gap of 40 microns in places, caused by the slight asymmetrical 'saddle' shape of the PCB land area.
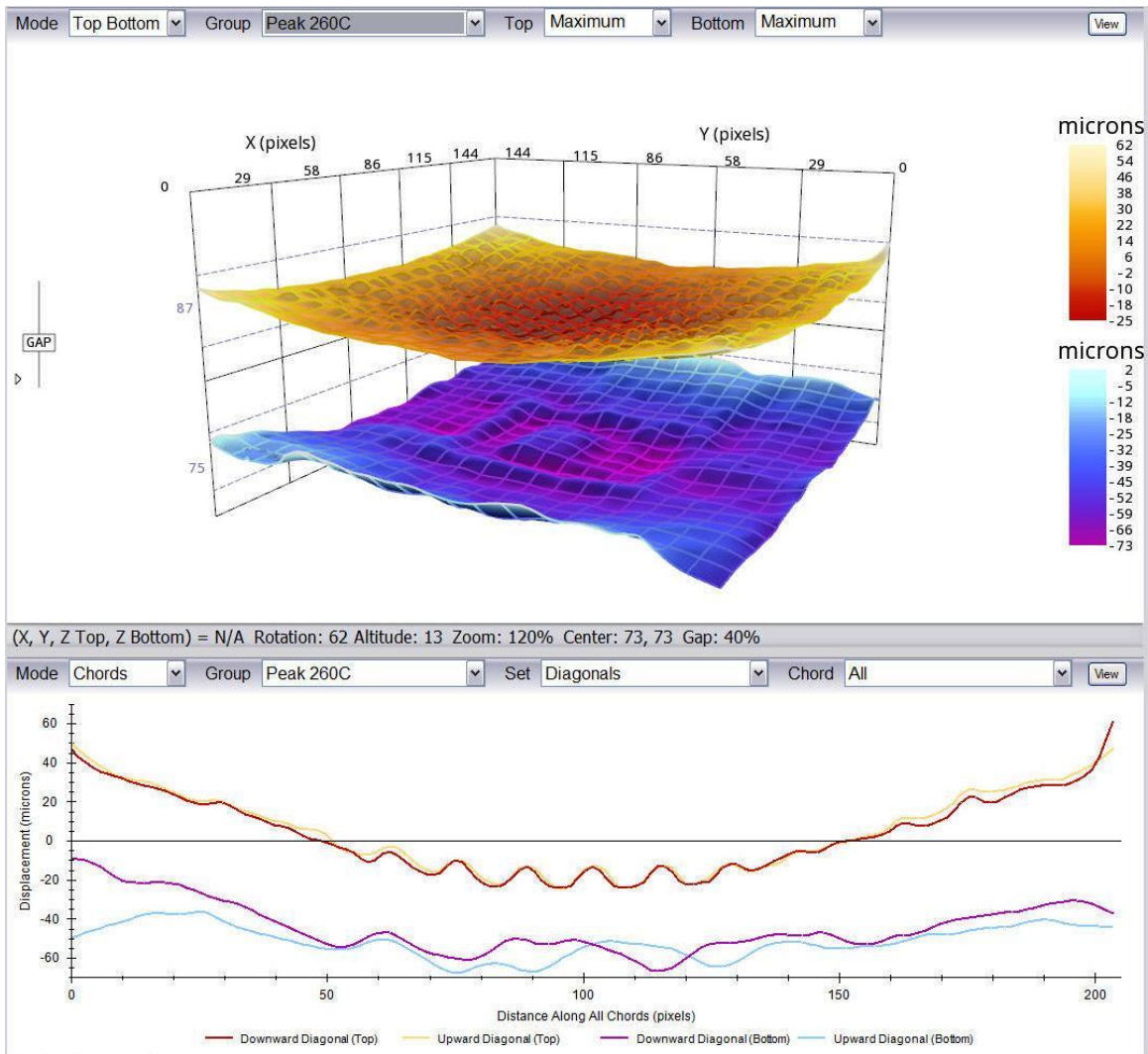
Figure 7: Gap Analysis at Peak Temperature

Figure 8 illustrates the effect of the PCB area asymmetry acting as a standoff to the package at two edges. If a feature like this caused an assembly problem, redesign or adjustment of process variables to increase the shape compatibility would be required.
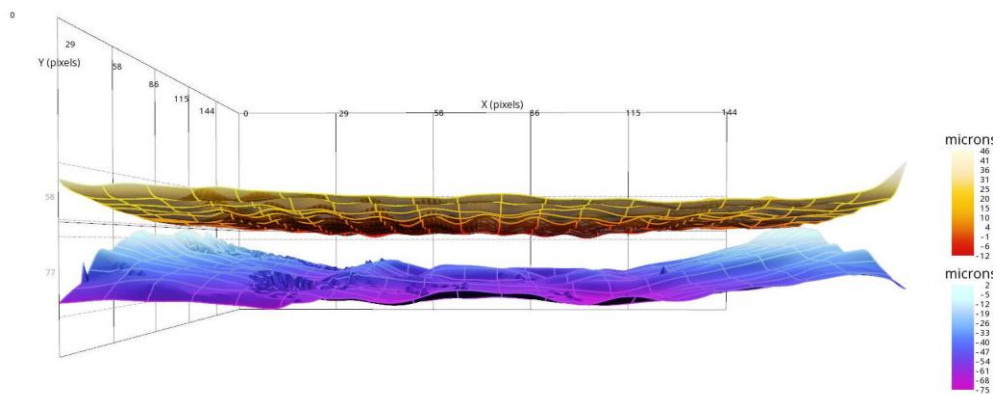

Figure 8: Side View of the Gap Caused by PCB Area's 'Saddle' Shape

Figure 9 shows the components cooling down, at 160°C. Gaps in the extreme areas are 55-100 microns. By this stage, past solder *liquidus*, most warpage-related defects will have already formed. Shorts/Bridging can happen near the end of *liquidus*, as the components cool and change their shapes and corresponding interconnect gaps before the solder has completely re-solidified. The amount of concern with post-peak temperature defect formation will determine how in-depth analysis needs to be to adequately understand the two surfaces' mechanical interaction before the solder solidifies.
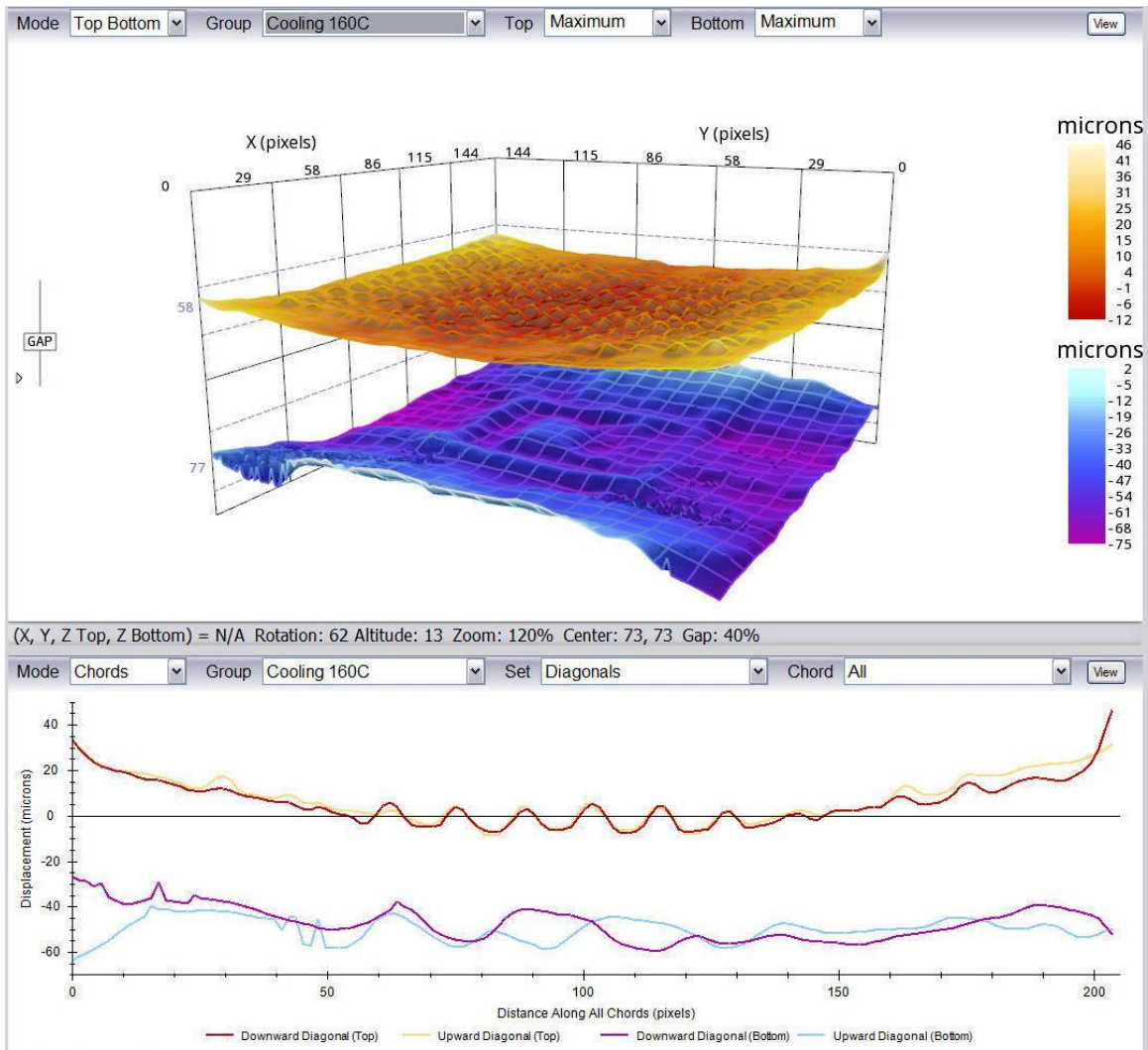

Figure 9: Cooling Phase at 160°C

 Figure 10 shows the final measurement, at room temperature (25°C). The components revert almost completely back to their original shapes. When actually attached in reflow, the two components will demonstrate a different warpage at this stage, since the solder will bond them together. But studying what the two components do when measured separately can be useful for estimating residual stresses that are carried by the solder joints, for purposes such as reliability engineering. In this case, it appears there could be a high level of residual stress in the solder, in the finished assembly, since the measured shapes are so different at final room temperature.
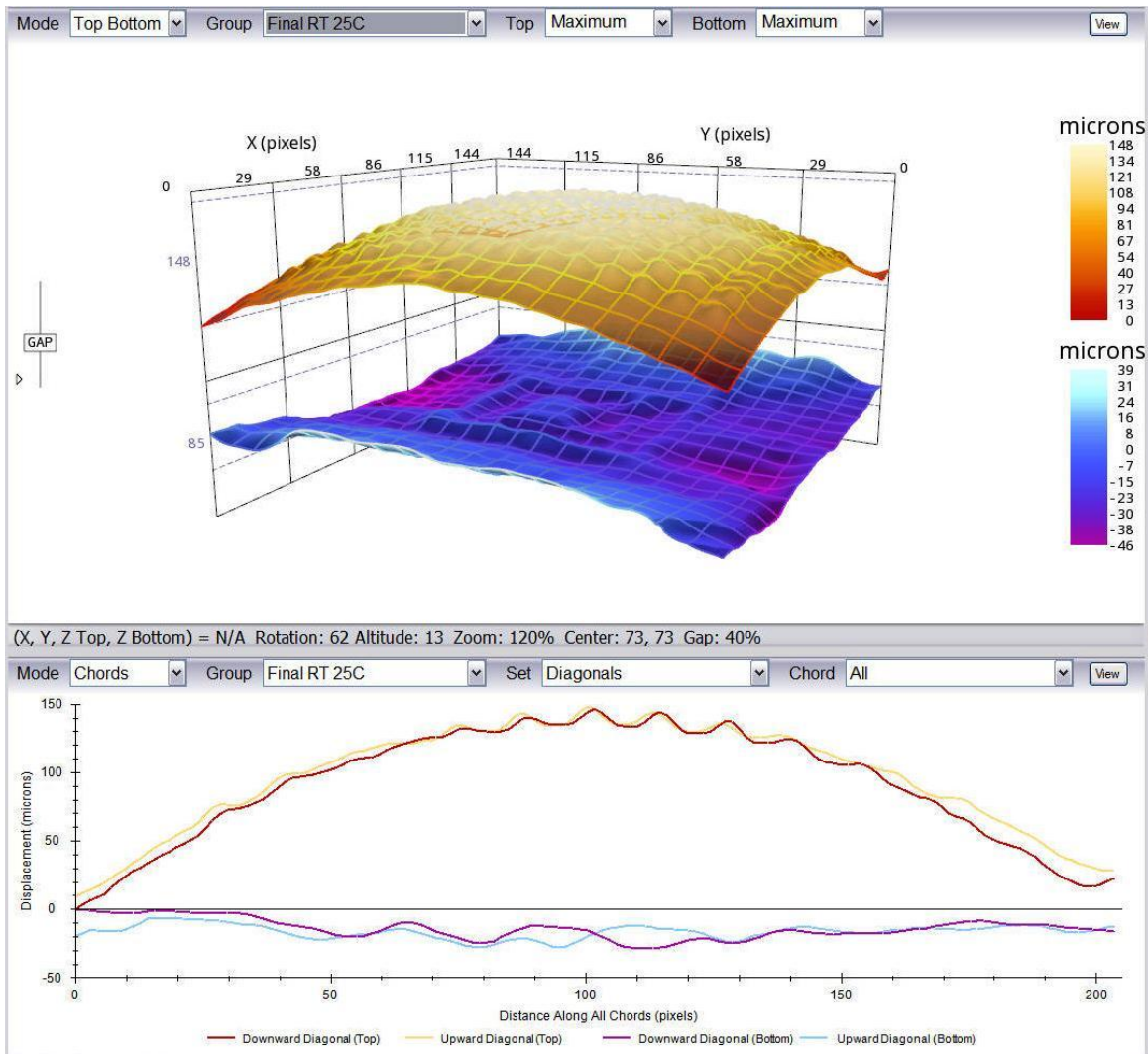
Figure 10: Room Temperature Final Graph

Gap Limits Example

In place of coplanarity limits, it can improve the process of evaluating full-field gaps to assign 'gap limits' to the graphs. Figures 11-13 show the passing (green), warning (yellow), and failing (red) gap areas using a colored plane between the Top and Bottom surfaces. Gap upper limits have been arbitrarily set at 50 microns for 'warning' and 75 microns for 'failing'.
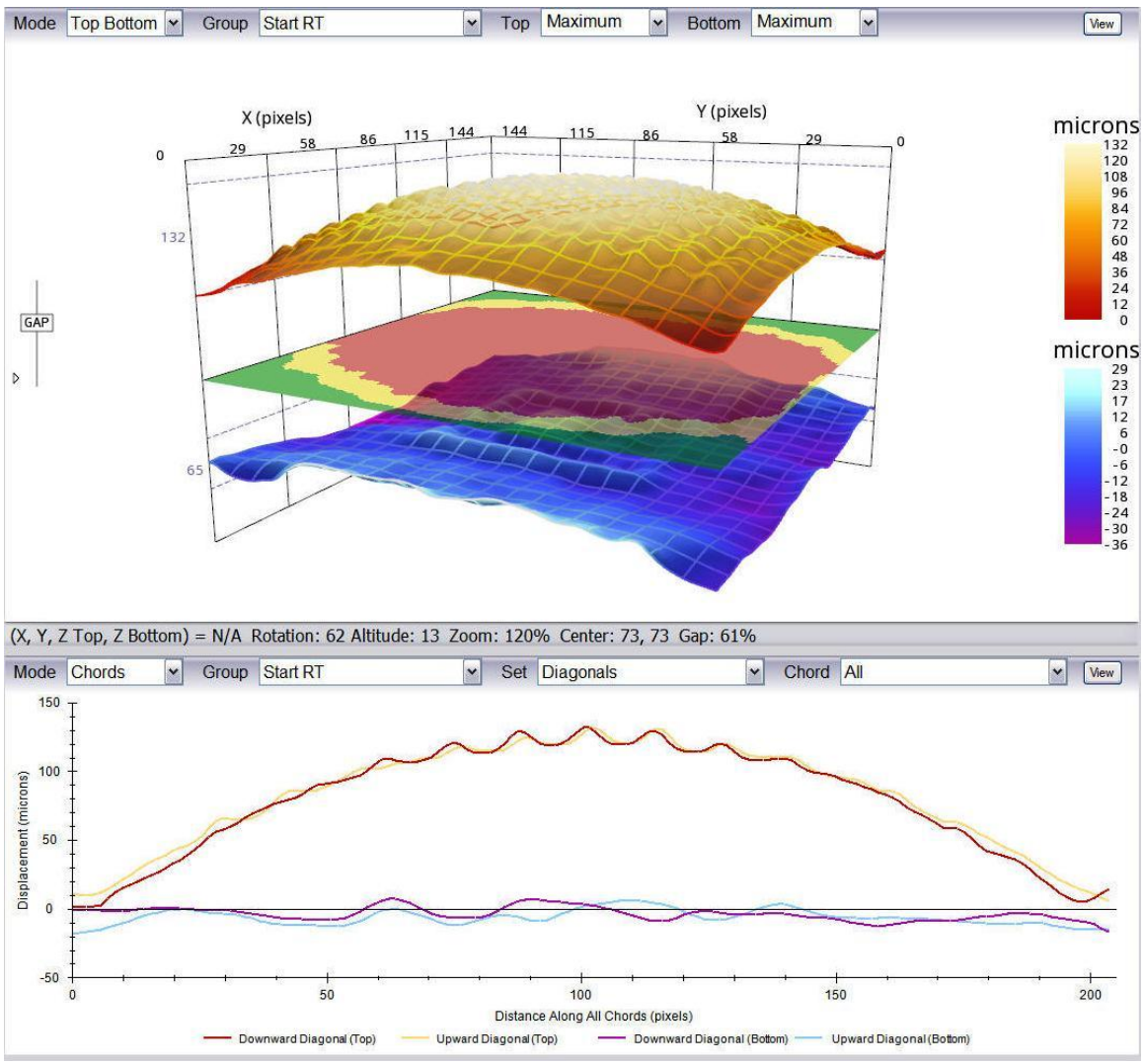
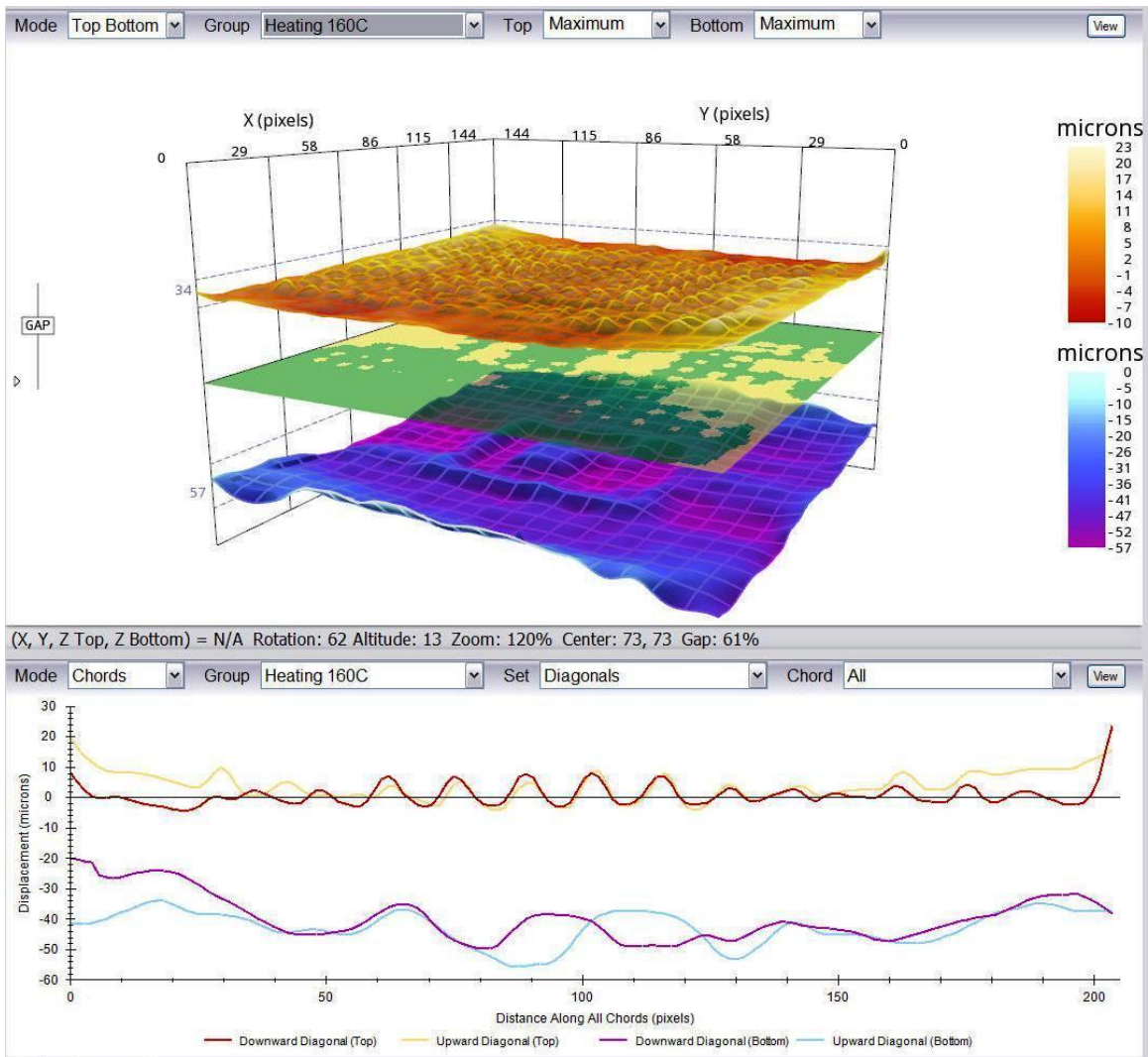**Figure 11: Starting Room Temperature Graph with Gap Limits Map**

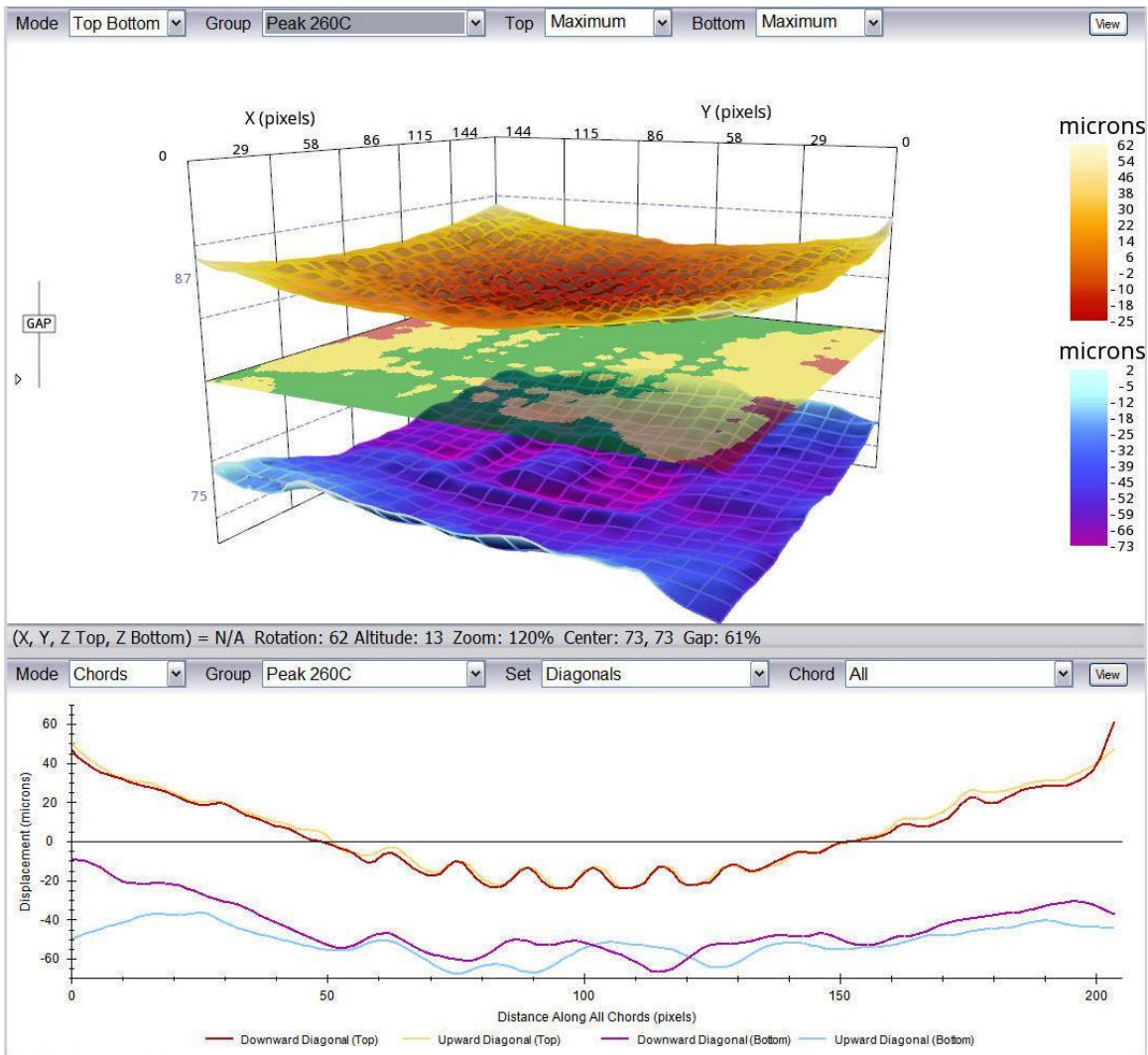**Figure 12: Heating at 160°C Graph with Gap Limits Map**

Figure 13: Peak Temperature Graph at 260°C with Gap Limits Map

Asymmetrical Warpage

Figure 14 shows an example of the Top Surface from the component used above, and a different PCB attach area, from a different design. This figure illustrates a case where, due to the complex shape of the land area involved, special processing decisions, such as adjusting paste volume, might be needed to optimize assembly yield.
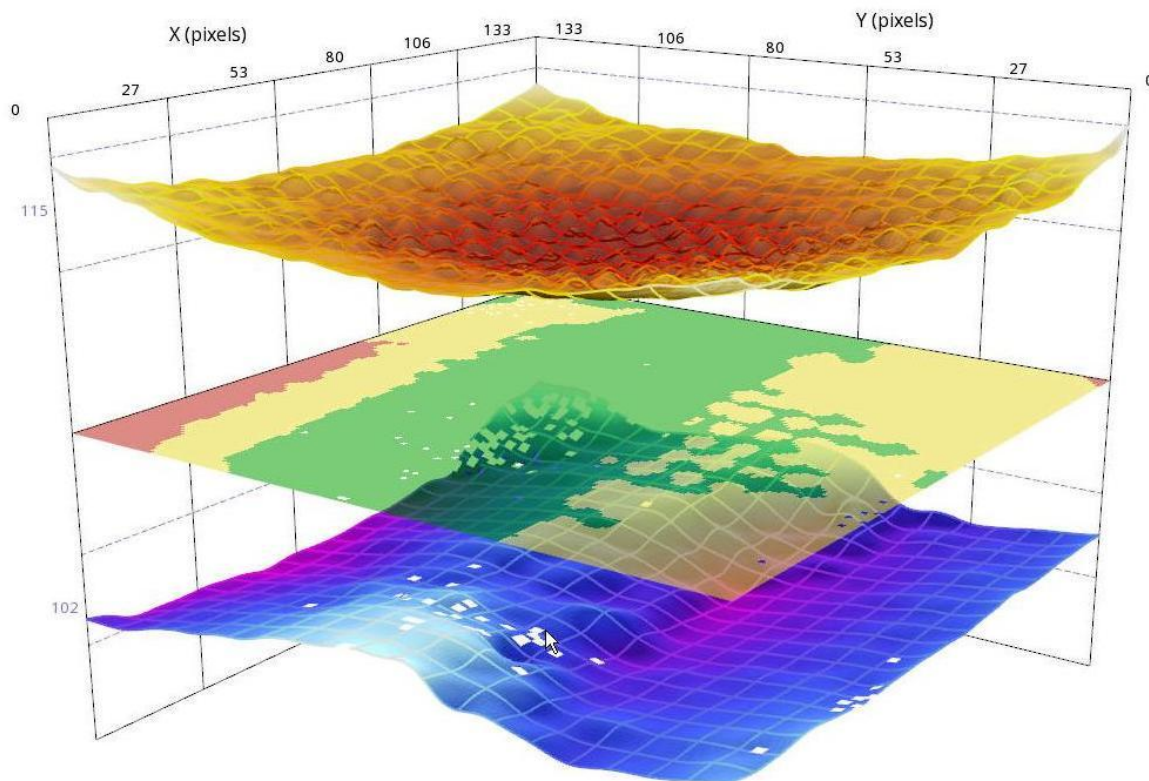
Figure 14: Example of Asymmetrical PCB Land Area Warpage with Gap Limits Map

(Note that warpage/shape data collected past the *liquidus* stage of the profile is not indicative of what will happen during actual assembly in the post-attach phases. Since the Top and Bottom surfaces are measured separately, the effect of the attach process and subsequent warpage combination of the attached components cannot be represented. Because head-in-pillow, shorts, and opens are formed prior to the end of *liquidus*, this is not a concern regarding the analysis process shown.)

Conclusion

The problem of head-in-pillow presented at APEX 2009, and all the associated process challenges with BGAs including fine pitch BGAs it described, have not gone away. Instead, the increasing use of new and thinner materials and finer ball pitches has made defects caused by warpage even more of a reflow assembly problem. The industry trend toward thinner, more powerful mobile devices, and the associated need for reliable package-on-package (PoP) assembly, is another reason head-in-pillow remains a critical defect type that must be avoided during design of assemblies, and through adjustment and control of process variables.

Many SMT electronics manufacturers have measured the full-field shape of surfaces throughout reflow for the past ten years or longer. Today, most major manufacturers are measuring the shape of their components' attach surfaces across reflow temperatures, as a daily routine. To best avoid and compensate for designs that have a tendency to develop shorts, opens, or head-in-pillow defects, companies responsible for development of components on either side of the assembly also need to plan for how the shape of that component will match with its mating part, to ensure that expected gaps at each critical temperature point are appropriate. Because of the small interconnect gaps the structure of surface mount products allows, 'designing for assembly' is critical. Those responsible for the reflow assembly of SMT components also need similar interconnect gap analysis at various temperatures, to adjust various process parameters to produce the most reliable solder joints possible.

The analysis method shown here is much more complex and involved than just measuring coplanarity and checking if it is less than a certain limit. Critical temperature points must be reviewed. Gaps per interconnect must be examined, possibly with a variety of statistical surface combinations to review the range of possibilities that will exist in production. Analyzing dual surfaces is certainly more challenging than traditional warpage analysis. But trends in electronics product design and manufacturing are creating more difficult SMT assembly and reliability scenarios.

These more complex problems create the need for more complex analysis methods, including the need to collect, combine, and review data from both components involved in each SMT assembly. In many cases, just analyzing package coplanarity at elevated temperature is no longer effective. By looking at both surfaces together, product designs can be validated, assembly process variables can be planned, and reliability and compatibility of components can be monitored, through use of full-field, dual-surface gap analysis across the critical temperatures of the reflow profile.

References

(1) Russell Nowland, Richard Coyle, Peter Read, George Wenger, "Telecommunications Case Studies Address Head-in-Pillow (HnP) Defects and Mitigation through Assembly Process Modifications and Control", APEX 2009

(2) Dudi Amir, Raiyo Aspandiar, Scott Buttars, Wei Wei Chin, Paramjeet Gill, "Head – And – Pillow SMT Failure Modes", P*roceedings of SMTA International*, 409-421, San Diego, CA, (2009)

(3) Dongji Xie, Dongkai Shangguan, David Geiger, "Head-in-Pillow (HIP) and Yield Study on SIP and PoP Assembly", ECTC 2009

(4) JESD22-B112A, "High Temperature Package Measurement Methodology", JEDEC Solid State Technology Association, 2009

(5) JEDEC Publication 95, SPP-024 Issue A, "Reflow Flatness Requirements for Ball Grid Array Packages", JEDEC Solid State Technology Association, 2009

(6) John Davignon, Ken Chiavone, Jiahui Pan, James Henzi, David Mendez, Ron Kulterman, "PCB Dynamic Coplanarity at Lead-Free SMT Temperatures", *Proceedings of SMTA International*, Ft. Worth, TX, (2011)

(7) JEITA ED-7306E, "Measurement Methods of Package Warpage at Elevated Temperature and the Maximum Permissible Warpage", Japan Electronics and Information Technology Association, 2007

(8) John Davignon, et al., "PCB Dynamic Coplanarity at Lead-Free SMT Temperatures", *Proceedings of SMTA International*, Ft. Worth, TX, (2011)

(9) Ranjit Pandher, Rahul Raut, Michael Liberatore, Navendra Jodhan, and Karen Tellefsen, "A Procedure To Determine Head-In-Pillow Defect And Analysis Of Contributing Factors", *Proceedings of SMTA International,* (2010)

(10) Dongji Xie, et al., "Head-in-Pillow (HIP) and Yield Study on SIP and PoP Assembly", ECTC 2009