High First-Pass Yields In A Lead-Free Environment

by Mario Scalzo, Indium

With the decrease in both the size and the pitch of area-array devices, new problems may arise in both printing and flux technology, such as head-in-pillow and graping defects. Print transfer efficiency of the solder paste must be maximised and flux activators have to be redesigned for high-tin alloys. We take a look at a statistical approach to formulating solder paste rheology and flux chemistry resulting in high print efficiencies and appropriate activity release to produce high first-pass yields.

While the electronics manufacturing industry has been occupied with the challenge of RoHS compliance and with it, Pb-free soldering, established trends of increasing functionality and miniaturisation have continued. The increasing use of ultra-fine pitch and area-array devices presents challenges in both printing and flux technology. With the decrease in both the size and the pitch of said components, new problems may arise, such as head-in-pillow and graping defects. Print transfer efficiency of the solder paste must be maximised to maintain consistent solder volumes and acceptable solder joints; process optimisation may also be a factor. Flux activators have to be redesigned to allow for the less-than-ideal soldering performance of high-tin alloys. These challenges have been met by a statistical approach to formulating solder paste rheology and flux chemistry resulting in high print efficiencies and appropriate activity release to produce high first-pass yields.

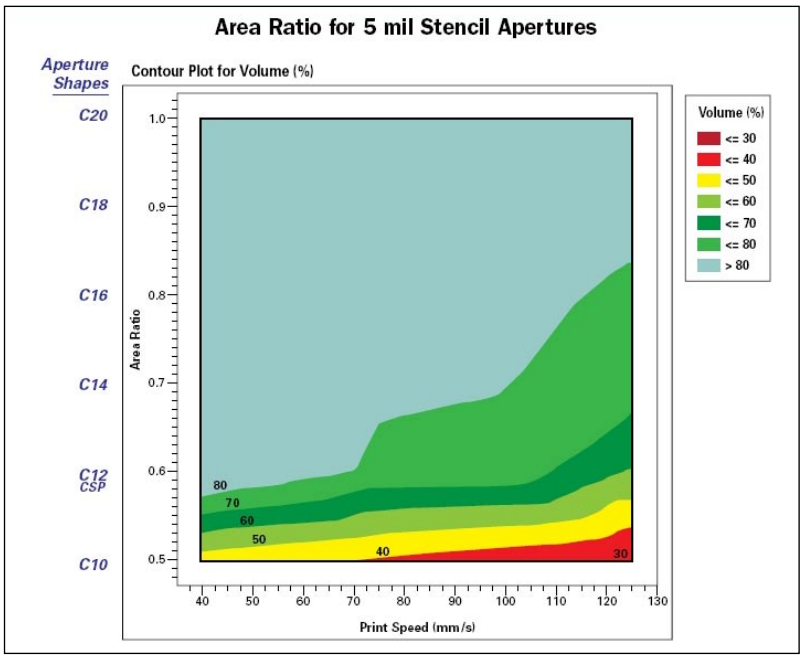
Transfer efficiency

Although stencils may offer a small increase in the amount of paste applied, it is the paste itself that usually makes the difference. There are three types of stencils that we normally come across when visiting customers: laser cut, laser cut with electro-polish, and electro-deposited (or e-fab). Stencil release, often-called transfer efficiency or TE, can be tracked through a paste measurement system, and the TE and stencil manufacturing costs commonly increase across the three types, respectively.

By feeding the stencil details into the paste measurement system at onset, the system can calculate the theoretical amount of paste that should be deposited, creating a percentage (efficiency) from measuring the amount of paste that was actually deposited. Transfer efficiency (Figure 1) is just now becoming something that we are tracking scientifically (read statistically). Some variables that can affect transfer efficiency are stencil type, atmospheric conditions, and the paste itself. Room temperature, and sometimes humidity, also affects transfer efficiency as the viscosity usually drops when solder paste is warmer. Solder paste also becomes less tacky at warmer temperatures. Humidity affects water-washable paste in the same ways, so much so that cold slump may be induced.

Measuring transfer efficiency can be an effective way of ensuring the maximum amount of return from a solder paste, but, transfer efficiency can only be measured once the variation has been taken into account. It is the variation in solder paste deposits that can be observed as an indicator of the defect level. The consequence of excessive variation is that ultra-fine pitch defect levels tend to be most susceptible. Because there is no single answer for all applications, we must look at the individual paste printing processes. Only then can the transfer

Figure 1 – Example transfer efficiency contour plot of print speed versus volume



Calculating the Upper and Lower Spec Limits (USL & LSL):
LSL = (Ratio) x (Nominal Height) = (50%) x (0.005") = 0.0025"
USL = (Ratio) x (Nominal Height) = (150%) x (0.005") = 0.0075"

If the LSL is 0.0025" and USL is .0075", then we can calculate the upper and lower Capability Indices, for 6 Standard Deviations (6 sigma):
Cpu = (Height % – 50%) / (3 x Std. Dev.) = (100-50) / (3 × 10) = 50 / 30 = 1.7
Cpl = (150% – Height %) / (3 x Std. Dev.) = (150-100) / (3 × 10) = 50 / 30 = 1.7
Cpk = (Cpu + Cpl) / 2 = 1.7

Figure 2 – Calculating upper & lower specification limits and capability indices

efficiency's upper and lower specification limits be defined. Each individual process has a different tolerable lower specification limit with regard to transfer efficiency and must be established through testing. To recognise the amount of variation in solder paste height, for example, eliminating the variation due to all other causes outside the actual printing process must occur.

Some of the more common causes for variation are related to board support and squeegees. As control of these basic features of the solder paste deposition process is improved, the variation in solder paste height will be minimised and defect levels will decrease for the components. An easy way of calculating the specification limits is to think of it like this: we have a target height of 0.005" (5 mils), which is the stencil thickness, therefore, for this application 0.005" solder paste height is 100%.

Knowing the theoretical volume, the capability indices for the various ultra-fine pitch components can be calculated using the standard deviation (σ) of the ratios. In Figure 2, we have reasonable print tolerances, which are plus or minus one-half of the 5 mil thickness, or 2.5-7.5 mils. Using a tolerance for 10% standard deviation could be used to characterise the solder paste prints (until further process measurement shows otherwise).

A Cpk (process capability index) of 1.7 is commonly acceptable, but the measured average height will indicate whether we need to make the tolerances higher or lower. A wider tolerance will give you a higher Cpk, but again, a Cpk just below 2.0 will be easier to monitor. It is common knowledge that no matter what you use for your Cpk (as long as it's tracked), it will go down when your process is losing control and up when the process improves (or your measurement system has been compromised). To maximise the Cpk; a good stencil printer cleaning, new squeegee blades, and better board support will make marked improvements. Starting with these features not only takes a lot of the variation out of the process but also will give you a better baseline.

Defect recognition

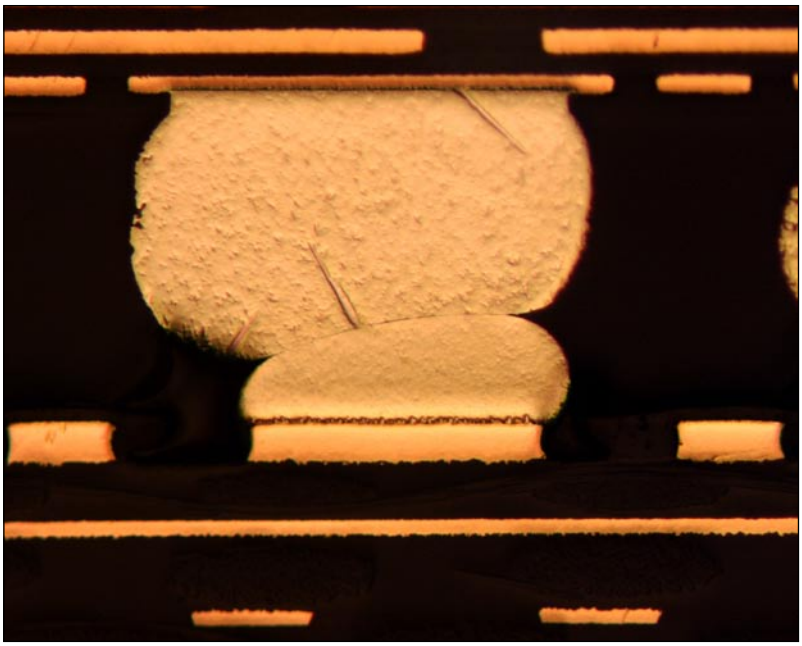
A head-in-pillow defect is the incomplete wetting of the entire solder joint or a Ball-Grid Array (BGA), Chip-Scale Package (CSP), or even a Package-on-Package (PoP). From cross-sections, it actually looks like a head has pressed into a soft pillow (Figure 3).

Two issues may cause head-in-pillow defects: poor wetting and warping of the component. They both look the same, but you can identify them because random head-in-pillow defects are from poor wetting and edge or centre defects are caused by warping. All the head-in-pillow defect types can be separated into three areas:

- ***Supply issues*** are everything before you put them on the line. This includes oxidation, or any other oxidation/hydroxide effects.
- ***Process issues*** are everything that is on-line. Including printing, placement and reflow.
- ***Material issues*** are everything that has to do with the soldering itself, such as wetting capacity or flux exhaustion.

Flux capability can be split into three areas; oxidation barrier, solvent boiling point and activator package. These are the critical factors determining the quality of solder joints produced during reflow. In addition to the flux chemistry, other factors influence fluxing performance, including solder particle size, deposit volume, and reflow profile. If the reflow process conditions are extreme enough to

Figure 3 - Example of head-in-pillow defect



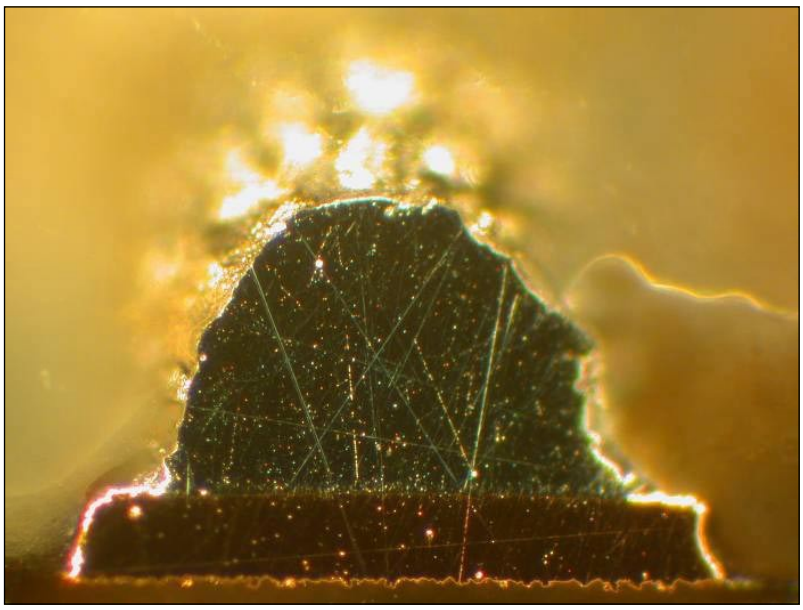


Figure 4 – Cross-section of an acceptable solder joint and surface graping

reduce the effectiveness of fluxing, the result may surface as poor wetting, solder balling, voiding, poor tensile strength and / or incomplete solder coalescence.

Among these, incomplete coalescence or "graping" (Figure 4), has arisen in increasing numbers, particularly with increased adoption of ultra-small component packages such as 0201 and 01005 passive components. This occurs when the outer layer of alloy powder spheres have lost their protective coating of flux and have not fully coalesced into the main bulk of the solder during reflow, producing an irregular surface finish similar to a cluster of grapes. Some CSPs have been affected by this same system, but it has proved difficult to diagnose and may be seen as a head-in-pillow defect.

In cross-sections of these joints where graping has appeared, it seems that the wetting, inter-metallic layer, and tensile strength have not been affected and electrical continuity is maintained. This suggests that the graping is merely superficial. Manufacturers should still try to prevent graping because quality assurance inspectors usually correlate an abnormal solder joint appearance with a "cold" or defective solder joint, and visual inspection equipment is usually trained to look for such characteristics in solder joints and report them as defective. These process interruptions cause significant line-down situations, as well as increased hand rework just to "shine" up a solder joint.

Reflow optimisation

Reflow is a balancing act. A good profile is a split between too little and too much.

Typically, we configure a reflow profile to work with the available solder and components to give the highest tensile strength possible. We know what the end goal is and we adjust what we have to achieve that goal. Besides tensile strength, some secondary goals are good wetting, solid intermetallic formation, homogeneous solder joint, and a small, tight crystalline structure (Figure 5). All of these are achieved through process management of the reflow process.

There are four parts of the reflow process that are adjusted to achieve the goals we have in mind, namely highest possible tensile strength. They are ramp rate, time above liquidus (TAL), peak temperature, and cool-down rate. Each one of these has its own effect on the final solder joint and each one is important.

Ramp rate is literally the first step in the four-part reflow process and plays an important role in the formation of the intermetallics. Ramp rate, from room temperature to peak, needs to be watched for a few reasons. Ramp rates determine the spread and volatisation of the flux, as well as having a sinister hand in voiding, graping, and oxidation build-up.

A slow ramp tends to allow more solvent volatisation, or "out gassing". This keeps the flux close to where it was applied, reducing spread and slump. It also gives enough time for the full volatisation of the solvents in the flux, usually reducing voiding, as well as keeping the T of the board well under 10°C. All this extra time may have a detrimental effect on some other points of interest, though, especially oxide build-up of component and substrate metallisations, as well as the solder alloy itself, which is the main factor in graping. To the opposite, a fast ramp reaches the softening temperature of the flux quicker, and therefore the flux (and paste) spread to cover a greater area, which increases the area of the joint. It may also allow for some of the activators to be saved for the actual liquidus of the alloy. Of course, there are downsides to this approach, namely the possibility of voiding and an increase in the T across the board.

TAL and peak temperatures have the same type of affect on the solder joint. Look at it as "total heat input", as you can have a longer TAL and lower peak, or a higher peak and shorter TAL. As it is, together they play arguably the most vital role of the reflow process. The name of the game is heat. Heat is responsible for solid intermetallic formation and a homogeneous solder joint, as well as proper flux deactivation.

A short TAL or low peak may result in insufficient intermetallic formation, which results in low tensile strength. It is the intermetallic that

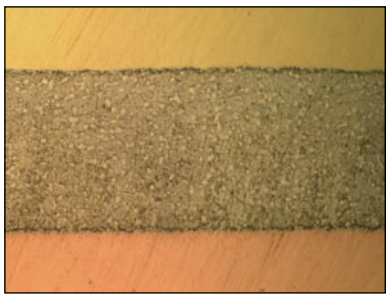


Figure 5 - Cross-section of a large-crystalline grain structured solder joint

gives the joint its strength, as you always want the joint to fail during testing at either the board-side of the pad or in the middle of the solder joint, not along the intermetallic. This is the same for the homogeneity of the joint, which is a metal solution. If the joint is not thoroughly mixed, then it is usually along the edges of the metal layer where it fails. Another issue with a short TAL or low peak is flux deactivation. Incomplete flux deactivation causes a multitude of sins, including poor surface insulation resistance (SIR) and continued etching of the metals.

A long TAL or high peak temperature may increase the dissolution of the base metallisations and possibly increase the melting point (MP) of the final joint. Too much dissolution of the base metals also forms a higher number of, and larger, intermetallics. Eventually, this may lead to the complete dissolving of the pad or component lead. Any time you increase the size of the intermetallic crystals, it is easier for them to fracture along the layer. A long TAL or high peak also increases joint stress, again giving another avenue for fracturing.

Cool down is the last line of defense against a poor solder joint. This is because the cool down ramp rate alone controls the formation of the crystalline structure of the metal lattice. The smaller, tighter, and denser we can make the crystal lattice, the higher the joint strength. Because it is along these facets of the crystal that the joint fractures, and the longer, larger, and sparser the crystal facets are, the easier they are to cleave. One way of visually investigating whether the solder joint is tight is to look at the post-reflow surface finish of the solder joint. A joint that seems to have good wetting and good flow yet is grainy and gray may have been exposed to a slow cool down. One method to test for this is to heat it up with a soldering iron. After it goes molten, remove the heat. If it becomes brighter and shinier, it probably needs a faster cool down. This may also happen if the joints around the perimeter of the board, or where the components are lightly populated, are bright and shiny and the densely populated areas have solders joints that are dull and grainy. This is because the more densely populated areas take longer to cool off and affect the cool down rate of the board. To avoid this, reposition the thermocouples used in profiling to the denser area and re-map the profile to meet their cooling needs.

Conclusion

Viewing a process through the statistical microscope really accomplishes two important goals. First, it forces an outside perspective view on each part of the process by focusing on the details and breaking each segment down into steps, increasing the understanding of the product. Secondly, by honing each step of these processes and paring away the waste, it streamlines the process flow and sharpens our cost-saving pencils. Polishing the printing process, where the majority of all solder issue can be traced, sets the foundation for success. Once ensuring that the prints are consistent, then other issues such as graping and head-in-pillow defects can be eliminated through optimisation of the reflow parameters or possibly evaluating a solder paste with an enhanced oxidation barrier, longer tack life or better wetting performance. Starting with a solid statistical foundation, using strong, stable building materials and sound manufacturing controls, a long-lasting and robust process can be built to weather the waves of variation.

Approaches for Managing Pb-Free Alloy Alternatives

The International Electronics Manufacturing Initiative (iNemi), an industry-led consortium focused on identifying and closing technology gaps, has outlined a set of recommendations to help the electronics industry manage Pb-free alloy alternatives. iNemi's recommendations support the guidelines developed by the EMS Forum to address these same issues. Although SAC 305/405 have been the most commonly used Pb-free alloys to date, they do not meet all of industry's needs for all applications, and new alloy solutions continue to be introduced. To help manufacturers manage the use of multiple solder alloys, iNemi recommends to: drive convergence of Pb-free alloys, develop an industry-standard assessment methodology, establish performance guidelines, update standards, identify and differentiate alloys.

Accordin to iNemi, the industry's experience with Pb-free processing is still in its infancy and there is much to be learned. It is inevitable that industry will continue to innovate and will develop new solder formulations and fluxes as its experience with Pb-free grows. However, the use of multiple alloys poses several challenges in the manufacturing process and industry needs some way to manage these challenges and limit their impact.

iNemi
Tel. +1 207 871 1260
cwilliams@inemi.org
www.inemi.org