Trends in Manufacturing that affect the selection of BGA/CSP rework equipment

Patrick McCall, Senior Product Manager
PACE, Incorporated
Laurel, MD 20723

Introduction

Grid Array packages (BGA and CSP/micro BGA) are gaining in popularity and acceptance in electronics manufacturing because they offer higher I/O counts per unit area and smaller footprints than conventional surface mount components. As grid array (GA) packages continue to gain acceptance, costs will come down, making them more accessible to all manufacturing.

Contract Manufacturers and OEMs are finding that GA devices prove to be more reliable and offer higher production yields than leaded devices. The most common cause of GAs having to be reworked during production is due to IC or other failure, not because of unsuccessful placement or reflow. In informal conversations, several contract manufacturers have reported that when using GAs, the need to rework packages due to bad placement or bad joints has dropped as much as 50 to 60%. In the future GAs will be preferred due to the higher production yields, higher performance and wider functionality. Today, GAs are seen in high end products. As manufacturing costs continue to come down, GAs will become the component of choice in all types of products.

On the surface, the cost of purchasing a rework station for GA components may not appear to be justified as the number of operations (and therefore problematic assemblies) is much lower than when using leaded devices. However, the higher dollar value associated with GAs and related assemblies is significant enough to justify the purchase of GA rework equipment. This, in turn, allows manufacturers to save valuable components and/or board assemblies and realize higher overall yields.

Because GAs are the package of choice for intelligent silicon, rework is often needed for other non-production-related reasons. In many cases, the performance or function of the final assembly can be greatly improved simply by modifying the internal software. Designers often allow for these types of upgrades in assemblies used in communications or monitoring applications. Therefore, the ability to rework a GA allows manufacturers the ability to upgrade or enhance an entire final assembly. When rework is performed for this reason, it is critical to ensure the process is successful and no damage or unnecessary thermal stress is applied to the assembly.

Factors that should be evaluated when selecting GA rework equipment include:

1. System Flexibility
2. Paste Stenciling and Flux
3. Flexibility in Component Placement
4. Full board Pre-Heating Capability
5. Flexible Nozzle Configurations

System Flexibility

Grid arrays, CSPs in particular, are most often thought to be used in size sensitive applications even though the benefits of using these packages go far beyond size considerations. As BGAs and CSPs become more accepted and cost effective, they will appear as a mix on all PCB types and are expected to displace the majority held by leaded devices in 3 to 5 years. Contract manufacturers and OEMs need to consider this carefully when selecting equipment for reworking these packages.

Flexibility in rework is a key consideration that can be and should be a major driver in the selection process. Typically, decisions are made regarding rework equipment selection that end up leaving 50 percent of rework opportunities out in the cold. The ability to successfully rework small components is often sacrificed by selecting a machine with large component capability. Another common example is that perceived portability is often chosen at the expense of the ability to work with large boards and provide full board preheating.

Machines that are able to provide full board preheating are tied to large power consumption requirements, i.e. 220 volt, 60 Hz, which is not usually readily available plant wide. This eliminates the portability of the machine and creates the need for centralized rework centers and logistics management.

Manufacturing facilities are often running a mix of boards and components through the same lines. In order for all production lines to have immediate access to large fixed BGA rework machines, one machine per two lines as a minimum is needed. Portability should be considered as rework machines that are portable and utilize standard and readily accessible power supplies can be moved easily between lines and/or departments. As a result, fewer machines may need to be purchased.

By selecting rework machines that are flexible, portable and truly meet the needs of the facility, manufacturers have to purchase fewer machines as they can be shared between production lines and departments. Fewer machines means that fewer highly trained employees are required to operate those machines and less labor is required to manage the logistics of a centralized rework center.

Paste & Flux

GAs bring with them special requirements when using flux and paste. It is critical that both be applied precisely and in controlled quantities. Traditionally, the application of paste is completed using spot stencils and smearing paste over the prepared lands. This is neither a simple or repeatable operation and relies heavily on the operator's skill and experience. All too often, the stencil is not square to the board and the paste is uneven from one edge of the stencil aperture to the other. It is also very common for excess paste to be forced under the edge of the stencil that the paste is drawn against. This is not usually a problem for standard surface mount devices as bridges can be easily corrected; however, this is a significant problem when using GA packages. Paste that is applied unevenly on a land can also affect alignment of the component to the board when using an optical overlay system.

Another problem with using templates to apply paste to the land patterns is that as boards become denser, there just isn't enough space to fit a spot stencil onto the board. In situations where a stencil cannot fit, an alternative commonly sought is to use liquid dispensing equipment that places a drop of paste on each land. When automated, this becomes very expensive and time consuming and when a manual process is used it becomes unreliable.

There are several possible solutions, the first of which is to not use paste. While this is an acceptable solution for most packages with eutectic solder balls, flux must still be used. Additionally, gold and 90/10 (high-temp) solder bumps on packages will become common place in the future and both require the use of paste.

As a favorable, cost effective alternative, paste can be applied directly to the solder ball on the underside of the package using a template. See Figure 1. This eliminates the space and spot stencil access concerns on dense boards and allows for a more precise measure of paste to be applied. Paste depth is not reliant on the stencils orientation to the board, but rather the thickness of the template which can be controlled exactly.

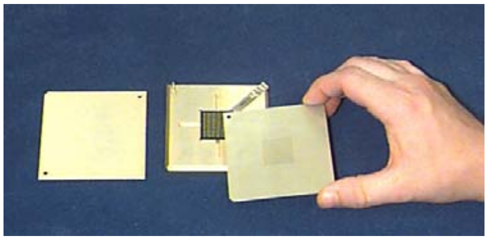

Figure 1.

This is possible because of the interface of the solder ball to the template. The solder ball actually penetrates the aperture of the stencil, forming a precise mechanical connection that minimizes and in most cases eliminates the likelihood that solder paste can be forced into areas that should be protected by the template

The precise and controlled application of flux is just as important to a successful process as the precise and controlled application of solder paste. Typically, the operator is directed to apply flux with a brush over the lands and rely on the process being used to ensure activation and drying of excess flux before reflow. If excess flux is not driven off before reflow, out gassing can occur, resulting in voids within the solder joint itself. With the introduction of gel flux, applying precise amounts of flux to balls can be accomplished and repeated consistently.

One method of ensuring even flux depth is to use an application device known generically as a “gel-flux applicator”. These devices are essentially a metal block that is milled to a precise depth, typically one-third the ball diameter. Gel flux is applied and a squeegee is used to distribute the flux evenly across the milled depression. The component is then picked up using a vacuum tool or pick and set down into the even layer of gel flux. When the component is lifted, flux has been applied to the solder balls to the appropriate depth and consistent amounts have been applied to each of the balls. The component is then ready for placement. This process closely mimics automated placement machines.

Flexibility in Component Placement

GAs offer many more challenges when placing them then their leaded cousins. The main difficulty is that the leads cannot be seen as they are on the underside of the component. Standard packages with pitches of 1.27 mm or greater with 0.7.mm or greater balls, can be placed successfully using templates that orient the solder ball array to the land array. This usually requires some level of technical competency but can be accomplished. Packages with ball diameters under .8 mm are very challenging to place using templates and packages under .6 mm are virtually impossible and success is hit-or-miss at best.

In order to guarantee successful placement, an optical alignment system must be employed. Split image systems are readily available with many different options. Typically, split image systems are comprised of a prism that us used to collect two images (one above and one below) and project them onto a series of mirrors where they are then projected into the lens of a camera. The images are then displayed on a monitor and appear as two separate images overlaid on one another. Either the component or the board is repositioned until the ball and land array patterns match exactly. See Figure 2.

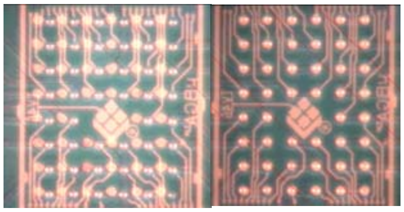

Figure 2.

Most optical systems used today also have a “split vision” capability, which allows the image to be split into smaller segments; e.g. only the corners of a component are viewed and magnified. This is valuable when placing large GAs or large QFPs as it allows two opposite corners to be viewed, ensuring proper alignment of components that are too large to be seen in their entirety. This is accomplished by positioning an additional set of mirrors (usually through manual means) that literally clip the image so only the corners are directed onto the mirrors that project the images into the camera.

Systems that use mirrors to direct images into the camera are delicate and require constant calibration to ensure the mirrors are in exact alignment. When the component is brought down to the board, the balls and lands will not match exactly if the mirrors, prism, or camera is out of alignment, even slightly.

Recalibrating an optical system is more often than not a lengthy and difficult process. Relocating a rework machine with an optical alignment system can easily misalign the mirrors taking the machine off line. In some cases, over time, the simple back and forth action of the optics housing as it is moved into position and out of the way again in the course of normal use is enough to cause the mirrors to become misaligned. This is but another reason why most rework stations must be kept in stable and stationary locations and are not suitable for today's production environment.

By incorporating an optical system that utilizes a PC, the "calibration calamity" can be avoided. Systems of this type direct the images from the prism directly into the camera and use software in the PC to manipulate the images. Software can even be used to produce a "split image" when placing large components, thus eliminating the mechanical mechanisms and an additional level of complexity from the system. PC based optics systems are easily calibrated as there are only two components, the camera and the prism, to align, thus reducing the compounding effect of the mirrors. This also reduces the cost of the equipment as well as the cost of ownership.

In order for a machine to be effective with a wide variety of component sizes, the optics system should have a wide range of zoom or magnification capability. For standard 1.27 mm pitch GAs with .8 mm solder balls magnification capability of 35x is required and a magnification of 80x or higher is required when placing CSPs.

An equally important piece in the placement puzzle is the table/board holder. The table/board holder should be stable, durable, adjustable, and able to hold a variety of board shapes either directly or by using optional carriers. Board holders should also be spring-loaded and be capable of holding a wide variety of board sizes, adding to the overall flexibility of the machine. Board holders should allow for the use of fully adjustable board supports

Pre-Heating

Proper pre-heating is an essential component of a successful process. Proper preheating:

- ensures homogenous temperatures across the board and components,
- it eliminates warpage, twisting, flexing, and bowing of PCBs during the process which is essential for maintaining planarity of the reflow site, and
- it allows for successful reflow at lower temperatures ensuring the safety of the PCB and component.

All too often, surface temperatures in one location of a board are measured to determine if the desired pre-heat temperature has been achieved. While measuring the temperature on the topside of the PCB indicates warming through the board, measuring in only one location does not ensure homogenous temperatures across the entire board. In many cases heat sinks such as ground planes or shielding can pull heat from surrounding areas creating cool spots which will result in warping or twisting.

When creating a profile, it is important to measure the topside temperatures at several locations including locations close to and away from known heat sinks. Only when the temperatures of all locations have a variation of less than 3 to 5 degrees centigrade and show signs of stability can we say with certainty that the entire board is warmed.

There are several options for applying heat to the underside of boards that include conductive, convective and radiant.

- Conductive methods are typically ineffective as they are difficult to control and require a very close proximity of the board to the hot surface.

- Convective methods are effective but are not necessarily efficient, as air is a poor medium to transfer heat. If full board preheating of boards over 10 x 10 is required and convective pre heating methods are used, large amounts of power are required. This typically means a dedicated 220 volt power supply. Additionally, the effectiveness of convective heat application can vary easily as ambient temperature change or if fans or other devices that can disturb air patterns (even people walking by) in the facility are present. This means that a profile developed in one location in a facility, such as a laboratory, will not have the same results when run in another location.

- Radiant heating sources are efficient for pre-heating and are effective when used on large boards. They are controllable, can be operated at 110 volts, 60 Hz or 230 volts, 50 Hz with average current draw and is an effective and stable medium for transferring heat as it is not disturbed by changes in environment, i.e. fluctuations in air patterns. As a result, it helps to ensure a repeatable process.

  When working with double-sided boards, the distance between the heat emitter and board is usually increased which can make radiant and convective methods ineffective. Radiant emitters cover a larger area than either convective or radiant methods and are more effective over greater distances.

  Additionally, radiant medias penetrate materials to a certain extent so in many cases pre-heat temperatures are reached faster and more evenly without subjecting the bottom surface of the board to high temperatures for longer periods of time as with convective methods.

Nozzle Configurations

Nozzles are used to direct hot gas from the heating source to the component/reflow site. Some of the more popular options include:

- using the nozzle to direct the hot gas straight down over the reflow site. The component is typically well below the nozzle, >3 millimeters. With this method, there are no means to direct the hot gas once it leaves the opening of the nozzle.
- using the nozzle as an oven in an attempt to force air under the component and re-create the same thermal environment as a reflow oven,
- the venting method where small slits or vents are present in the sidewalls of the nozzle to allow air to escape. The component is positioned either just inside or just below the nozzle opening, and
- Hybrid configurations that are a blend of the above three.

Each of these methods have benefits and detractors associated with them. There are specific scenarios where each may be the best choice.

When a nozzle is used to direct hot gas straight down over a component the main problem is reflowing adjacent components. While this is not desirable, it is often unavoidable to some extent when there is insufficient space to fit a nozzle around a component due to component density on the board. See Figure 3. Usually it is acceptable if adjacent components see temperatures above solder melt as long as all joints reach a liquid state. Should reflow occur in only a portion of the joints on a component, the planarity of the component may be effected. In the case of a GA, open joints can occur as a result. If solder joints are not heated to a state of complete liquidity, brittle joints will result and ultimately the assembly will fail.

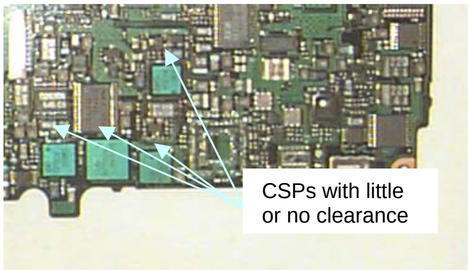

Figure 3.

Nozzles can only be used as an oven when sufficient space around the component exists. Oven style nozzles are oversized, requiring more clearance than a straight or vented style of nozzle. For the “oven concept” to work, the bottom edge of the nozzle must touch the board on all four sides and remain in contact through-out the entire reflow phase. In some types of oven nozzles, a rubber gasket is attached to the edge of the nozzle, adding to the space required around the component.

Temperature monitoring when developing a profile is extremely important to ensure overheating does not occur. Likewise, a system that delivers consistent and repeatable results is required to ensure a safe process. Should overheating occur, the component or solder mask on the board can be damaged. De-lamination of the board and component can also result from overheating. Component and board damage can readily occur when using oven style nozzles if operators are not trained or if profiles are not developed properly.

Component manufacturers should be queried to determine the maximum temperature a given component can withstand safely. The maximum temperature minus 20 degrees Celsius should be identified as a maximum temperature during reflow. Another parameter that must be monitored is the ramp rate. It is easy to exceed the recommended ramp when using the “oven concept”.

Typically, vented nozzles do not have the same spatial requirements as oven style nozzles. This is because the nozzle wall does not usually extend past the substrate of the component and the nozzle wall is significantly thinner than the rubber gasket along the bottom edge of an oven style nozzle. Adjacent component reflow is also controlled much better than straight nozzles as vented air is usually directed out the sides at an upwards angle. One of the drawbacks of using vented nozzles is that when components are fitted tightly inside the nozzle airflow can be affected. When using vented nozzles it is important that air be able to flow evenly across and around the component. Should airflow be blocked or restricted on one or more sides, reflow may not occur on that side of the component.

Hybrid nozzle styles can actually be used in the same ways as several of the three main classifications. Flexibility in nozzle design is preferred as the same nozzle can be used in many situations and board configurations.

Conclusion

As manufacturing processes and component packaging technology change, the need for rework equipment that is reliable, repeatable, and portable is clear. Rework equipment must also be flexible and adaptable to a meet the challenges of today’s rework and repair environment. While “product advancement” is visible, it has also come in the form of products that are over-engineered, overly complicated and not cost effective. Simplifying rework is an important step the technology life cycle. When grid array packages can be reworked easily and equipment becomes cost effective, acceptance of GA packages will increase rapidly. Today, service and repair centers are not able to replace faulty components or upgrade chip sets because the equipment needed to do so is well out of their price range. This deters manufacturers from readily adopting advanced packaging technology and limits the rate in which new packaging technology is developed. It is important that the manufacturers of GA rework equipment continue to upgrade their product offerings and stay current with today’s production needs and packaging innovations.

References

1. Kevin Towle, Line Configuration for Today’s Highly Complex Manufacturing, SMT Magazine, 11/98
2. Japan Printed Circuit Association, Report on the Technology Road Map for Advanced System Integration and Packaging, 1998
3. Mike Bixenman and Treliant Fang, Wafer Solder Bumping, Advanced Packaging, 6/99
4. Mike Sauer and Kristina Bergman, Rework & Repair, SMT Magazine, 7/99
5. Paul Wood, New Considerations in CSP Rework, Circuits Assembly, 7/99