Assembly Equipment Requirements for Next Generation Advanced Packaging

Bob Chylak, Horst Clauberg and Tom Strothmann
Kulicke and Soffa Industries, Inc.
1005 Virginia Dr, Fort Washington, PA, 19034, USA
Phone: +1-215-784-6446
Email: bchylak@kns.com

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2016/1/000321/2254250/isom-2016-wp35.pdf by guest on 22 November 2022

Abstract

Device packaging is undergoing a proliferation of assembly options within the ever-expanding category of Advanced Packaging. Fan Out-Wafer Level Packages are achieving wide adoption based on improved performance and reduced package size and new System in Package products are coming to market in FOWLP, 2.5D and 3D package formats with the full capability to leverage heterogeneous integration in small package profiles. While the wide-spread adoption of thermocompression bonding and 2.5D packages predicted several years ago has not materialized to the extent predicted, advanced memory modules assembled by TCB are in high volume manufacturing, as are some high-end GPUs with integrated memory on Si interposer. High accuracy flip chip has been pushed to fine pitches that were difficult to imagine only three years ago and innovation in substrates and bonder technology is pushing the throughput and pitch capability even further. The packaging landscape, once dominated by a few large assembly providers, now includes turn-key packaging initiatives from the foundries with an expanding set of fan-out packing options. The fan-out processes include face-up and face-down methods, die first and die last methods and 2.5D or 3D package options.

Selection of the most appropriate packaging technology from the combined aspects of electrical performance, form-factor, yield and cost presents a complex problem with considerable uncertainty and high risk for capital investment. To address this problem, the industry demands flexible manufacturing solutions that can be modified and upgraded to accommodate a changing assembly environment.

This presentation will present the assembly process flows for various packaging options and discuss the key aspects of the process that influence throughput, accuracy and other key quality metrics, such as package warpage. These process flows in turn impose design constraints on submodules of the bonder. It will be shown that thoughtfully designed machine architecture allows for interchangeable and upgradeable submodules that can support nearly the entire range of assembly options. As an example, a nimble, low weight, medium force, constant heat bondhead for high throughput FOWLP can be interchanged with a high force, pulse heater bondhead to support low stress/low warpage thermocompression bonding. The various configuration options for a flexible advanced packaging bonder will be reviewed along with the impact of configuration changes on throughput and accuracy.

Key words

Advanced Packaging, Fan-out, Thermocompression Bonding

I. Introduction

Recent years have seen a dramatic increase in innovation and development of packaging technologies, collectively referred to as Advanced Packaging. Fan-out packaging is but one of the larger categories and itself spans a huge range of packaging options for many devices types, including single chips and systems in package. This paper will briefly review the process flows and in more detail discuss the implications for the optimal design of the die placement equipment. Accuracy, flexibility and cost-of-ownership (COO) are the main considerations that factor into both machine architecture and sub-system engineering. These three factors largely compete with each other, making it imperative that machine users and machine providers work together closely to fully understand the needs and that strong efforts are made to create accurate roadmaps for

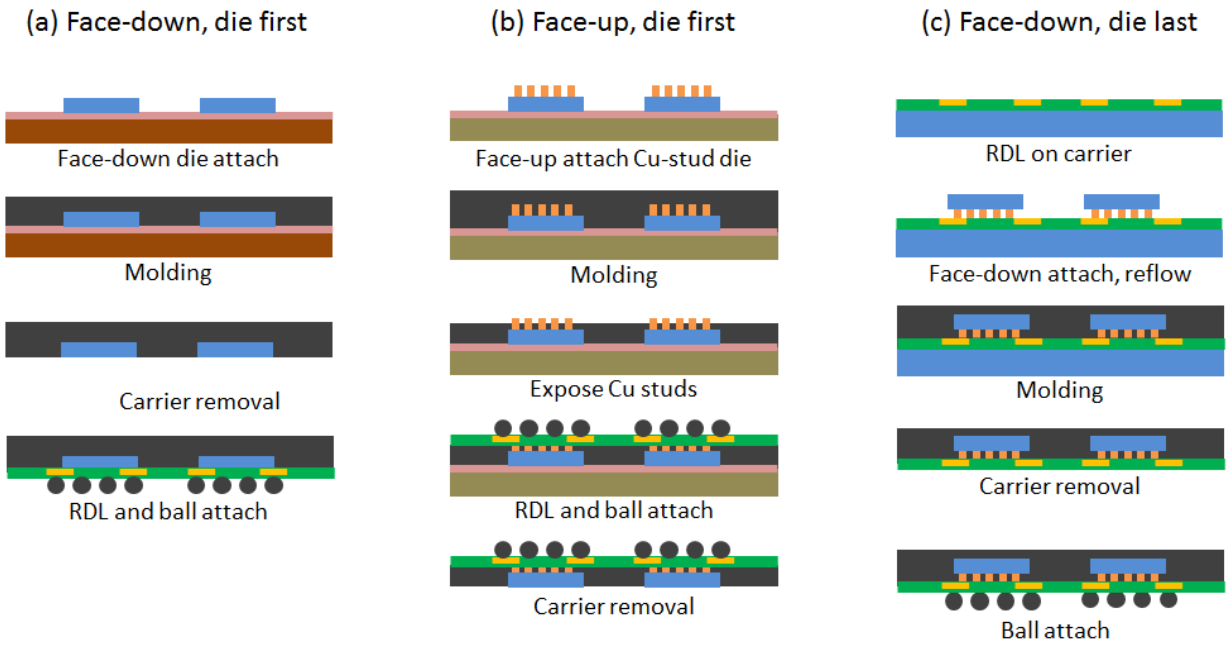


Fig. 1: Fan-out packaging assembly flows. (a) Face-down, die first, (b) face-up, die first and (c) face-down, die last

expected process requirements.

II. Fan-out Process Flows

A. Overview of fan-out package types

Fig. 1 provides a simplified overview of three broad categories of fan-out process flows. Traditional wafer-level fan-out, as exemplified by the eWLB process first introduced by Infineon, places a die onto a temporary adhesive on a carrier with the active side against the adhesive [1]. The carrier is then over-molded, the carrier and adhesive film are removed and then redistribution layers (RDL) are built directly onto the mold compound. This process is characterized as die-first, face-down.

An alternative process places the die face-up onto a carrier, over-molds and then grinds back to short interconnect studs on the die. RDL is then again applied directly onto the mold compound. TSMC's InFO process is the most familiar example of this process flow.

Finally, there are die-last processes that first create the RDL on a carrier, then bond die by either mass reflow or possibly thermocompression bonding onto the RDL. Only then is mold compound applied, followed by carrier removal.

Within these process flows, there are multiple variants, especially when the final product is a PoP (package-on-package) device. For example, the process flow in Fig 1(b) can be modified to where the die are placed onto an existing RDL with tall Cu pillars on the RDL, to allow placement of a second device on top and solder balls on the bottom.

B. Advantages and disadvantages of the process flows

All of the process flows just mentioned have been introduced in the market place and continue to be the subject of active development projects. The implications of the process flow on the requirements for die placement will be discussed below. Aside from these considerations, the overall process flows have advantages and disadvantages, the detailed discussion of which is beyond the scope of this paper. The die first processes are likely to be lower cost, since they can achieve very high speed die placement through the use of global alignment and since the overall assembly process is simpler. But die first risks good die to the RDL formation process, whereas die last allows for placing good die only on known-good sites. RDL yields are very high for one or two layers, but more layers can be a challenge to a die first process. Hence one of the main considerations between die first and die last is the cost of die relative to complexity and yield of the RDL process. High I/O, high cost die would favor die last processes, while low I/O die would favor die first processes.

II. Assembly equipment implications for die placement

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2016/1/000321/2254250/isom-2016-wp35.pdf by guest on 22 November 2022

Table 1: Die bonder requirements by advanced packaging application

	Face-down Die First	Face-up Die First	Face-down Die Last Mass reflow	Thermocompression Bonding (Face-down)
Flip/Unflip	Flip	Unflip	Flip	Flip
Substrate heat	Typically No	Typically Yes	No	Yes
Pulse heater	No	No	No	Yes
Fluxing	No	No	Yes	Yes – TC-CUF No – TC-NCP/NCF
Force	Low	Med	Low	Low TC-CUF High TC-NCP/NCF
Accuracy (µm)	5 – 10	5 – 10	3-10	2-3
Alignment	Global	Global	Local	Local
Substrate form factor	330mm carrier Future – panel	300mm carrier	330 mm carrier Future - panel	Substrate 300mm wafer
UPH	Current: 4000 Future: 6000	Current: 3000 Future: 6000	Current: 4000 Future: 6000	Current: 1000 Future: 2000

A. Global vs local alignment

Close examination of the various process flows reveals that each of the processes require die bonders with substantially different configurations and features. Table 1 gives an overview of the requirements by process flow. The die-first process flows, both face-up and face-down, generally require a global alignment to the carrier, meaning that only a few fiducials are provided on the carrier and all die are placed relative to these features. Accuracy requirements vary from better than 10µm to better than 5µm, depending on target application and assembly house. Deca Technologies' process flow is a notable exception with respect to accuracy, in that dynamic patterning allows for a substantial relaxation of the accuracy requirement [3].

The die last process can work with local alignment, since RDL is already in place and the die can be aligned to these features.

Achieving high accuracy is much more straight-forward for a local alignment process. In global alignment, the alignment features are 300mm or more apart, so that accuracy of only a few microns must be maintained as mechanisms move across such large distances. With local alignment, split vision systems can simultaneously look down at the alignment features on the carrier and up at the alignment features on the die when the placed head is already holding the die above the place position. Errors in the motion of the split vision system no longer factor into accuracy and bondhead accuracy is only important over the very slight correction moves associated with bringing the Z axis down from where the die was aligned and where it is placed.

Local alignment is also much more forgiving with respect to drift over time. In global alignment, many die area placed for every alignment relative to the global alignment features, whereas with local alignment, the alignment to the local features necessarily occurs for every die. While there is a throughput benefit associated with aligning only occasionally, the machine is now also required to maintain accuracy over the time it takes to place many die. The throughput advantage of the global alignment is not as much as might be thought, since each visit to the global alignment fiducial is relatively costly in time because the features are >300mm apart. In local alignment, the motions of the vision system are much shorter and therefore faster. Local alignment is usually only used when there is existing RDL, such as in a die last process. Local alignment with die first would have a significant disadvantage to global alignment schemes if there is a high product mix since local alignment would require a custom carrier for each product type to be run.

B. Face-down vs face-up

Die are typically picked from wafer film frame in the face-up orientation. Handing them over from the picker to the placer then puts them in the face-down orientation – the "flip chip" process. To achieve face-up placement therefore requires the addition of an unflipper mechanism. While the

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2016/1/000321/2254250/isom-2016-wp35.pdf by guest on 22 November 2022

unflip mechanism is relatively simple, the unflip process does introduce additional complexity. First, it introduces an additional hand-over of the die. More importantly, once the die is face-up against the place tool, the alignment features on the die are hidden to the vision system. Initial alignment therefore has to occur while the die is still on the unflipper. The hand-over to the place tool then has to occur with minimal loss of accuracy or, more desirably, a second vision system checks for alignment shifts before and after the handover. But this system adds both cost to the machine and time to the process.

Finally, bonders are required to handle a variety of die sizes. This requires that the unflipper tool either have a complex design that allows for robust handling of many die sizes or an additional automatic tool changer mechanism.

C. Hot vs room temperature placement

Perhaps the most significant difference in the die attach processes is whether the die are placed at room temperature or require heat to be adhered to the carrier. Generally, but not universally, face-down processes only require placement with both the carrier and the place tool at room temperature. The temporary adhesives used in face-up processes generally require that the carrier is at an elevated temperature, often between 100ºC and 150ºC. A high temperature carrier, especially in conjunction with global alignment, creates a substantial challenge to robust high accuracy alignment. The problem arises from thermal expansion. The most obvious is that the carrier temperature must be both very well controlled and highly uniform. Although a variety of carrier materials are used, several have coefficients of thermal expansion near 9ppm/ºC. For such a material, a 300mm disc would expand by 8µm for just a 3ºC rise in temperature. This expansion needs to be rationalized with a typical placement accuracy requirement of +/- 5µm. At room temperature, it's relatively straight-forward to maintain temperature control and uniformity of a few degrees Celsius. Maintaining high uniformity for a chuck at temperatures above 100ºC requires very careful machine design. The problem is exacerbated by air flows in the machine that provide internal cooling, ionized air for electrostatic charge control and removal of outgassing vapor from adhesives. The air flow must be designed and controlled extremely well. It should be noted that the thermal expansion properties of the carrier are somewhat fixed by down-stream requirements, so that choosing a material with lower CTE may not be an option.

Machine engineering considerations related to high carrier temperature go well beyond just carrier thermal uniformity. The work holder and carrier become heat sources that can raise the temperature of many other structures within the machine. These include the mounting system for the work holder, the vision system and the place head. All of these components are now subject to thermal expansion. The mounting mechanism must now be protected from heat and made of material with lower CTE that are more expensive than those needed for a bonder designed for room temperature bonding. Moreover, the vision system and place head do not remain in fixed position relative to the work holder. Especially the vision system, because of its close proximity to the carrier, must be both shielded from the heat of the carrier and maintained at a constant temperature. Both systems may need to incorporate lower thermal expansion materials.

All combined, the additional cost associated with the more expensive low CTE materials, shielding, heating systems, etc easily reach 10% of the total cost of the bonder. Thus the need for hot carrier should be very carefully considered. An alternative is to use a heated place tool instead of a heated chuck to reduce the thermal management constraints on the equipment.

D. Accuracy and throughput

The accuracy specification should be considered very carefully since it has enormous implications for cost-of-ownership (COO). Both throughput and cost of the bonder are affected, a double hit to the COO calculation. Especially for accuracy below about 5µm, the cost associated with building a machine increase exponentially as accuracy is improved. Similarly, alignment processes, re-calibrations, motion speeds and machine vibration control add substantial cycle time. Moreover, vibration and thermal management considerations add weight to moving parts of the machine, reducing motion speeds.

There are clear trade-offs between accuracy and throughput, but it should be noted that many of these trade-offs occur during the machine design phase. The extent to which throughput can be increased by dialing down accuracy on a given machine is limited. For example, while some high-end SMT-type assembly machines can achieve accuracy well below 10µm at very high throughputs, the basic architecture of these machines would make it extremely challenging to achieve better than 5µm accuracy. Similarly, flip-chip and thermocompression type bonders used for high accuracy fan-out applications can achieve well below 5µm accuracy, but their speeds are limited by their architecture to about 1/3 of that of a SMT-type machine running at about 7µm accuracy. In face down applications such as eWLB with moderate accuracy requirements of 7-10µm, high-end SMT equipment is the clear choice with throughput capability as high as 27k UPH. However, these machines do not have the required carrier heat or force required for die

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2016/1/000321/2254250/isom-2016-wp35.pdf by guest on 22 November 2022

placement in some demanding face up applications. For high accuracy applications demanding less than 5µm accuracy, as is common in both face up and RDL first process flows, the placement speed can range from 4k to 6k UPH. These high accuracy machines are able to also provide heat and force for those applications that require it.

III. Flexibility vs Dedicated Design

The previous discussion highlights the diverse design considerations required for the variety of packaging process flows under the broad category of fan-out packaging. The great variety of process flows, and market uncertainty regarding which one of these process flows will represent the highest volumes in production, force the equipment manufactures to produce machines that are configurable, flexible and even field retrofitable. Simply adding the ability to re-configure a machine to accommodate a new fan-out process adds cost and time-to-market that all machines need to carry, irrespective of whether they will eventually run an alternate process or not. Many structural and thermal management systems cannot be modified in the field, so again, cost is added, independent of whether the end-use requires all of these features. While some of these features can be specified as a process specific configuration at the time of purchase, providing many alternate configurations has a substantial adverse impact on the economies of scale for each specific machine configuration. The most cost effective production machine will be one that is optimized for a specific process flow; however this is difficult to achieve until dominant process flows emerge from the current uncertainty in the market. At this time flexible machine platforms provide the lowest risk for capital investment by the FOLWP technology providers.

III. Conclusion

The proliferation of advanced packaging flows has created a need for highly versatile and re-configurable die placement machines. Depending on the specific application, high end SMT-type machines can provide the required accuracy at throughputs that are far higher than those of high-end flip chip and thermocompression flip chip bonders. Accuracy in the 2 to 5µm range do require the latter machines that have architectures suited to dealing with the complexities of achieving high accuracies. Highly specialized equipment specifically designed for a specific process flow would provide a substantially cost-reduced and throughput-optimized solution. More clarity in roadmaps and wider adoption of the most favorable process flows would enable development of such machines.

References

[1] S. Kroehnert, “Fan-out WLP — The enabler for system-in-package on Wafer Level (WLSIP),” Elect. Sys-Int. Tech. Conf. (ESTC, 2012).

[2] C-F. Tseng, C-S. Liu, C-H. Wu and D. Yu, “InFO (Wafer Level Integrated Fan-Out) Technology,” 66th Elect. Comp. and Tech. Conf. (ECTC. 2016).

[3] C. Bishop, B. Rogers, C. Scanlan and T. Olson, “Adaptive Patterning Design Methodologies,” 66th Elect. Comp. and Tech. Conf. (ECTC, 2016).

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2016/1/000321/2254250/isom-2016-wp35.pdf by guest on 22 November 2022