Comparison of Package-on-Package Technologies Utilizing Flip Chip and Fan-Out Wafer Level Packaging

Amy P. Lujan
SavanSys Solutions LLC
Austin, Texas, USA
amyl@savansys.com

***Abstract*— Fan-out wafer level packaging (FOWLP) has been an area of focus in the electronics packaging industry for multiple years now. As the technology matures, the number of applications for which FOWLP is suitable is growing. Depending on the application, FOWLP may be competing against wire bond, flip chip, embedded, interposer-based, or 3D stacked technologies. In order to make an informed decision, designers must understand not only the value of FOWLP technology, but the value of the package it seeks to replace. Technical capabilities and final package cost are the two key factors to evaluate when making this decision.**

This analysis focuses on package-on-package (PoP) technologies. Amkor, TSMC, and STATS ChipPAC are all examples of OSATs and foundries providing FOWLP-PoP solutions. The incumbent technology against which FOWLP-PoP is compared is flip chip packaging with through mold vias, and both process flows will be discussed. A cost and yield analysis is carried out to determine the cost implications of different design attributes, and activity based cost modeling is used. With this type of cost modeling, a process flow is divided into a series of activities, and the total cost of each activity is accumulated.

The goal of this analysis is to understand the costs associated with variations of the FOWLP-PoP process, and to evaluate the design attributes that play a role in determining whether FOWLP-PoP or FC-PoP has the potential to be cost-effective.

Keywords-flip chip; FOWLP; PoP

I. INTRODUCTION

Fan-out wafer level packaging (FOWLP) is gaining the ability to compete with a variety of existing packaging solutions as the technology matures. Fan-out packages are being used in application processors, microcontrollers, data centers, and automotive applications—just to name a few areas—and the number of applications is growing [1, 2]. In cases where a FOWLP solution may be used, designers and suppliers must understand the design and cost implications of replacing an incumbent technology with a FOWLP solution.

One key area of focus within the FOWLP arena is FOWLP package-on-package (PoP). PoP technology has been around for over a decade [3], and the forecast for it continues to grow, as seen in Fig. 1.

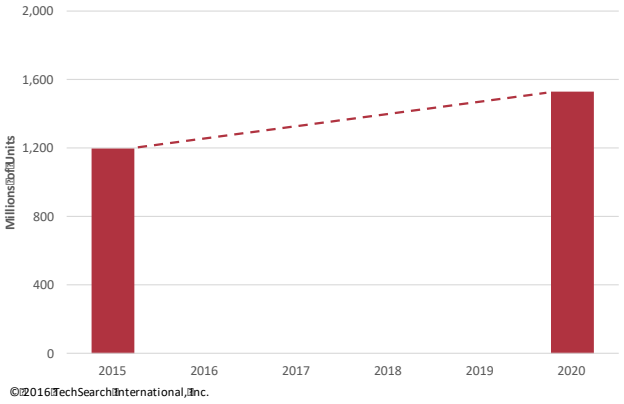

Figure 1. PoP Market Forecast from TechSearch

The first generation of PoP relied on wire bonded packages as both the top and bottom packages, with solder balls acting as the top-to-bottom connections [4]. Later generations began to use a flip chip package as the bottom package. They also began to utilize Through Mold Via® (TMV) technology, in which vias are created through the mold of the bottom package and filled with solder. Compared to placing two packages side-by-side, PoP technology allows for reduced signal noise, increased processing speed, and better use of board real estate [5].

Major industry players have been highlighting their FOWLP PoP solutions in recent years. TSMC is promoting InFO (integrated fan-out) PoP [6], AMKOR highlights the abilities of SLIM (silicon-less integrated module) and SWIFT (silicon wafer integrated fan-out technology) to use TMVs or copper pillars to become PoP solutions [7], and eWLB (embedded wafer level ball grid array) PoP is an option as well [8].

The primary goal of this analysis is to understand the cost and yield differences associated with flip chip and FOWLP PoP technologies. The scope of this analysis is the creation of the bottom package, and the creation of the top-to-bottom connection; this analysis does not include assembly of the top package.

II. PROCESS FLOWS

A. *Traditional PoP*

The traditional PoP solution against which FOWLP PoP is compared in this analysis is flip chip PoP that utilizes TMVs. The process flow is detailed in Fig. 2.

There are many differences between the flip chip process flow and the FOWLP PoP process flow that will be discussed next. Two key differences that are of particular note with regard to this cost and yield analysis are the number of scrap opportunities in the traditional PoP flow, and the fact that the die to be placed in the flip chip package must be bumped.

There are three scrap opportunities in the flip chip process flow. Bad die may be scrapped after the incoming wafer is bumped and diced; bad substrates may be scrapped before any die are assembled to them; bad packages may be scrapped after assembly. Having multiple scrap opportunities represents potential cost savings, because when a defect is found early in the process, the cost of additional processing does not have to occur on that bad package.

Additionally, the cost to bump a wafer is not insignificant. A die must have bumps or copper pillars to be placed in a flip chip package. Therefore, the cost of adding bumps to that incoming wafer is considered to be a part of the cost of a flip chip package. This is a cost that is not necessary in the FOWLP flow that will be evaluated next.

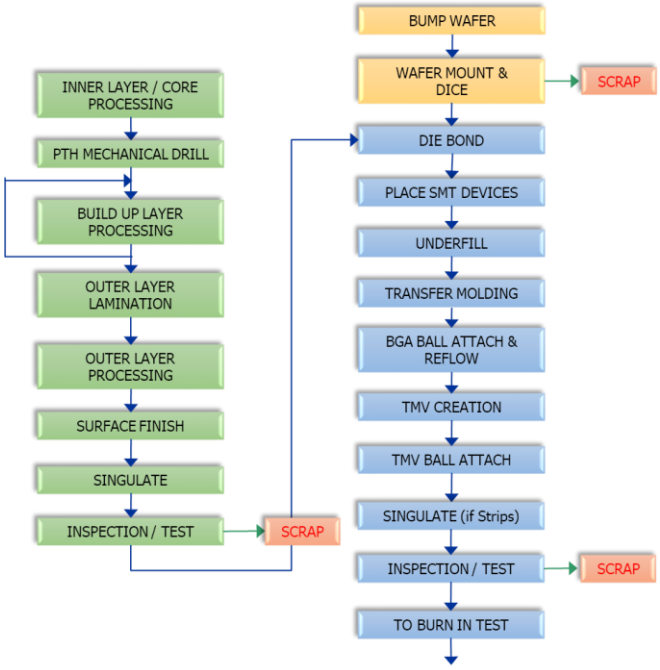

Figure 2. Flip chip PoP with TMVs Process Flow

B. *FOWLP PoP*

There are multiple types of fan-out technology. The detailed process flow for die-first, face down processing is shown in Fig 3.

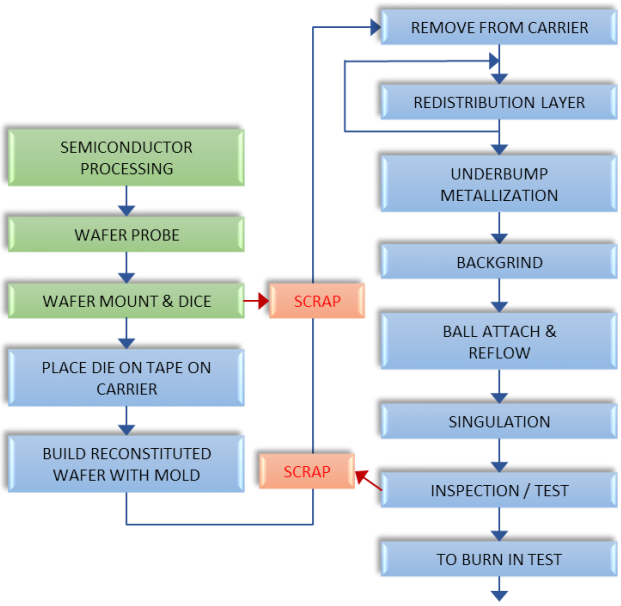

Figure 3. Die-first Face down FOWLP

In die-first face down FOWLP, the die does not have to be bumped. It is placed face down on a carrier, mold is applied to create a reconstituted wafer around the placed die, and then the carrier is removed and the newly-formed reconstituted wafer is flipped over, exposing the bottom face of the die. Connections are formed directly on the die, without the necessity of bumps. Not having to do the additional processing of adding bumps or copper pillars represents a cost savings over the flip chip process flow.

There are also only two scrap opportunities, versus the three scrap opportunities seen in the flip chip flow. The first scrap point occurs after testing the incoming die. The second and final scrap point does not occur until the end, after the package has been formed around the placed die and singulation has occurred. This increases potential scrap costs, because any defect created early in the process flow will not be discovered until the rest of the processing has been done.

One variation of fan-out technology is die-first processing in which the die is placed face up. This variation requires the die to come in with bumps or copper studs or pillars. There are other small adjustments in the face up process flow, such as the fact that the mold must be ground down to reveal the bumps or copper studs or pillars. This type of FOWLP also allows for the use of adaptive patterning, which requires an inspection step after mold application and copper stud reveal so that the redistribution layer (RDL) pattern can be adjusted for die shift.

Die-last FOWLP is another alternative. This process has similarities to the flip chip process. In this variation, the RDL is built first, utilizing a temporary carrier. Incoming die must have bumps or copper pillars, and if an inspection step is added after RDL creation, those die can be placed on known good RDL locations.

To translate any of the above fan-out technologies to a PoP technology, vias may be added in the mold to create

TMVs, or copper pillars may be plated up prior to mold application.

C. *Design Differences*

There are a variety of factors that differ between traditional and FOWLP PoP that cannot be captured by a cost comparison. Electrical performance, thermal output, and package height are all examples of considerations that will not be equivalent when comparing similar traditional and FOWLP PoP packages [9]. Furthermore, FOWLP tends to allow for finer features than traditional PoP technology [10]. Ultimately, a designer must know which technology will support the design requirements for a given product.

III. Activity Based Cost Modeling

Activity based cost modeling was used to construct generic die-first face-down FOWLP PoP and traditional PoP cost models. With activity based cost modeling, a process flow is divided into a series of activities, and the total cost of each activity is calculated. The following cost components are analyzed for each activity.

- Equipment cost and throughput – The throughput determines how long a wafer consumes the equipment. The total equipment cost determines how much depreciation should be allocated based on how long the equipment is consumed.
- Labor cost – The activity time and percentage of an operator required determine the labor component of the activity.
- Material cost – Both permanent material costs and consumable material costs are included. Permanent material includes mold, dielectric layers, copper, solder balls, etc. Consumables include chemicals for cleaning, flux, etch baths, etc.
- Yield loss – The cost of scrapped packages is allocated to the remaining good packages.
- Overhead and indirect costs – Typical overhead and indirect costs are accounted for.

Note that no overhead or indirect assumptions are made for any of the cost models used in this analysis. The comparisons focus on direct cost on

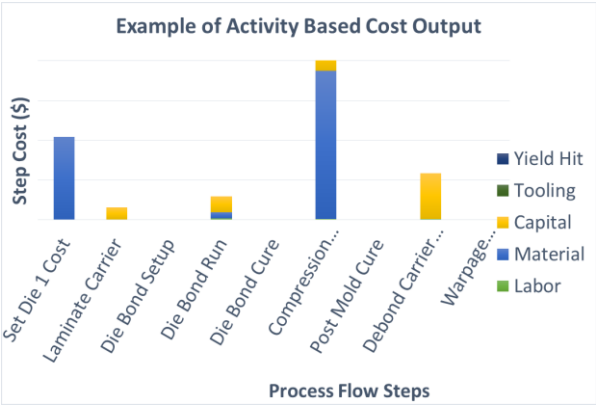

Figure 4. Sample of Activity Based Cost Modeling Results

Fig. 4 gives an example of the level of detailed output provided by an activity based cost model. The chart shows the first few steps of a die-first FOWLP process. The type of cost associated with each activity is broken down by color. In the steps selected, it can be seen that there is a high capital cost associated with the debond step, and a high material cost associated with the compression molding step.

IV. Cost Comparison

A. *Flip Chip PoP versus Die-last FOWLP PoP with TMVs*

As introduced in the previous section, there are a variety of FOWLP PoP options available. The most equivalent comparison that can be carried out for the two technologies being compared is flip chip PoP with TMVs versus die-last PoP with TMVs. In both cases, the die must come in bumped. Additionally, both processes require the use of underfill. Molded underfill or capillary underfill with later mold application may be used. The comparison in this analysis uses capillary underfill followed by mold application.

Three designs and five die price points were selected for analysis. The charts in Fig. 5 detail the results of this comparison. The results are presented as relative numbers and represent the cost of the die-last FOWLP PoP package as a percentage of the flip chip PoP cost. This means that any bars higher than 100% are cost-effective as a flip chip PoP; once the bar is below 100%, die-last FOWLP PoP is more cost effective.

These graphs allow for a few key observations. The first conclusion that can be drawn is that package size is a major cost driver for FOWLP PoP. The second is that die-last FOWLP PoP may have an advantage when handling more expensive die.

When examining the three charts in Fig 5., it can be seen that the largest package is always cost-effective as a flip chip PoP. This is because FOWLP PoP cost is almost entirely dependent on package size. In fan-out packaging, many of the activities are done on the entire wafer at once. Therefore, when that wafer is diced up into larger packages, the cost is spread out over fewer packages than if a smaller package were being fabricated.

Conversely, for the smallest package tested, which is 8x8mm, die-last FOWLP PoP is always cost-effective. For that smaller package size, the benefit of performing activities on the entire wafer at once is a clear benefit.

Like the largest package sizes, the medium-sized package is always cost-effective as a flip chip PoP. However, for the more expensive die, the bar starts to approach the 100% line. This reveals that die-last FOWLP PoP has a yield benefit compared to flip chip PoP—as the incoming die cost goes up, the yield benefit of die-last FOWLP PoP becomes more important.

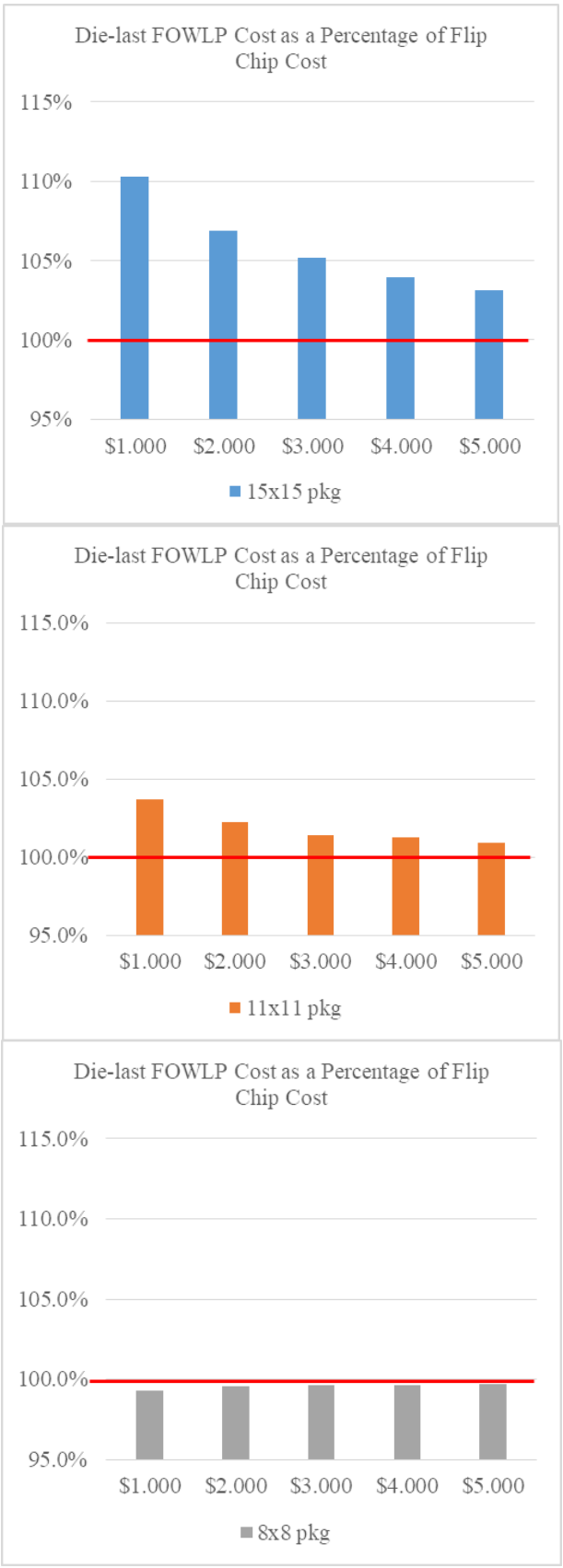

Figure 5. Die-last PoP cost as a percentage of Flip Chip PoP cost

It should be noted that one of the cost drivers in both PoP processes is the use of capillary underfill and a separate mold application step. One alternative is to use a molded underfill process. This reduces the number of steps, and tends to be a cost-effective solution [11]. However, the comparison above does not change even if molded underfill is used, because a switch to molded underfill benefits both package types being compared.

To take the comparison one step further, it is worth comparing the costs for process flow segments that appear in both of these technologies. Although flip chip and FOWLP are very different technologies, they share many of the same activities. There are also segments that are not the same in each technology, but can be considered equivalent (e.g. the substrate portion of a flip chip flow is similar to the UBM and RDL portions of a fan-out flow, though the activities are different). Table I shows a cost comparison for different segments.

TABLE I. COMPARISON OF PROCESS FLOW SEGMENT COSTS

Segment	***FOWLP***	***1-2-1 FC***	***Comments***
Support wafer	Only in FOWLP		
Substrate		Similar to UBM + RDL Cost	
UBM	Similar to substrate cost		
Bottom RDL			
Die prep	Same	Same	
Die bond	Same	Same	
Underfill	Similar	Similar	
Mold	Less expensive	More expensive	Transfer molding done in strips for flip chip is more expensive than compression molding the entire wafer at once
Debond	Only in FOWLP		
Misc.	More expensive	Less expensive	Low cost in both, comprised of miscellaneous cleans and inspections
Grind	Only in FOWLP		
TMV	Less expensive	More expensive	Costs are fairly close, but because this is done in strips for flip chip, the overall process takes longer and drives cost to be higher
Ball attach	Similar	Similar	
Dice and pack	Less expensive	More expensive	Low cost in both flows, but slightly higher in flip chip due to processing in strips
Yield loss on $1 die	Less expensive	More expensive	Both processes have high yields, but the die is placed later in the FOWLP process, which creates a yield benefit

B. *Flip Chip PoP versus Die-first face down FOWLP PoP with TMVs*

The comparisons in the previous section were repeated using the same flip chip PoP process flow as before, but this time comparing it to die-first face down FOWLP PoP (similar to an eWLB-style process). In all cases, the die-first face down FOWLP PoP solution was cost-effective.

Die-first face down fan-out tends to be more cost-effective than other styles of fan-out, which is why the results versus flip chip PoP depend on the style of fan-out.

One area in which die-first face down processing has a cost benefit is in wafer preparation. A die must be bumped to be placed in a die-last fan-out process flow; this is not the case for die-first face down process. The cost to bump an incoming wafer so that it can be placed in a die-last fan-out package is a cost that doesn't factor into a die-first face down fan-out package.

Another area in which die-first face down processing has a process and cost advantage is in temporary handling. The temporary handling process for a die-first face down process is straightforward. The die are placed on tape on a carrier, mold is applied, and then the reconstituted wafer is removed from the carrier tape. For a die-last process, a more robust handling system is necessary because the entire RDL is built up on a carrier, die are placed, and molding occurs. It is only near the end of the process, when the package is nearly complete, that debonding occurs. Carrying out the debond process without introducing any defects is crucial to keeping the cost of the package down.

On the other hand, die-last has a yield benefit, which translates to a cost benefit, when compared to die-first face down fan-out. Die placement occurs at the very beginning of the die-first face down process, which means any processing defects that occur along the way, such as during RDL creation, will result in the loss of already-placed die. In die-last processing, the RDL is built up first and die can be placed on known good RDL locations. This means that die-last processing tends to be more cost-effective than die-first face down processing when handling more expensive die—the yield benefit of die-last outweighs the increased processing costs [12].

Table II summarizes some of the key differences between these two technologies.

TABLE II. KEY COST/YIELD DIFFERENCES BETWEEN DIE-FIRST AND DIE-LAST FOWLP

Segment	***Comments***
Carrier/debond	Die-first has the cost benefit – Die-last carrier handling is more complex
Wafer preparation	Die-first has the cost benefit – Die-last processing requires the expensive bumping process for the incoming wafer
Yield hit	Die-last has the cost benefit – Die-first processing has a greater potential for yield loss because die placement occurs early in the process

C. *Die-first face up FOWLP PoP*

Now that die-first face down and die-last FOWLP have both been discussed and compared to flip chip PoP with TMVs, the natural progression is to investigate how die-first face down FOWLP competes. The technology of most focus within the industry that matches this style is TSMC's InFO PoP process. The InFO process utilizes a TIV—through InFO via—in place of a TMV [13]. This TIV is a large copper pillar. It is plated up on the carrier, die with copper pillars are placed face up, mold is applied, and then the mold is ground down to reveal the copper pillars. The remainder of the process is similar to the die-first face down process shown in Fig. 1.

The addition of large copper pillars in place of TMVs means this technology will not be cost-effective compared to traditional flip chip PoP with TMVs. The comparison of fan-out PoP with flip chip PoP is already a comparison of two very different technologies. Using copper pillars pushes that comparison farther apart, but it also allows for different designs and technical capabilities. Furthermore, the addition of copper pillars allows for a topside RDL, which is not supported by TMVs. Adding a topside RDL will be an added cost, but it will also allow for more designs and technical specifications.

Therefore, instead of comparing the cost of die-first face up fan-out PoP to flip chip PoP with TMVs, the cost of changing from a fan-out PoP with TMV technology to a fan-out PoP with copper pillar was analyzed.

The first table shows the cost increase when changing from a die-first face down fan-out flow using TMVs and molded underfill to a die-first face up fan-out flow utilizing large copper pillars. The second table is a similar comparison, except it includes the cost to add a topside RDL to the die-first face up fan-out design.

TABLE III. COST INCREASE FROM DIE-LAST TO DIE-FIRST

$1 die being packaged	**Cost increase from: die-last, face down w/MUF + TMVs to die-first, face up w/large Cu pillars**
Large design	6.9%
Medium design	4.3%
Small design	1.8%

TABLE IV. COST INCREASE FROM DIE-LAST TO DIE-FIRST WITH ADDITIONAL RDL

$1 die being packaged	**Cost increase from: die-last, face down w/MUF + TMVs to die-first, face up w/large Cu pillars + topside RDL**
Large design	18.9%
Medium design	12.7%
Small design	7.4%

V. SUMMARY

A variety of technologies were compared in this analysis. Fig. 6 summarizes the main results of the comparisons.

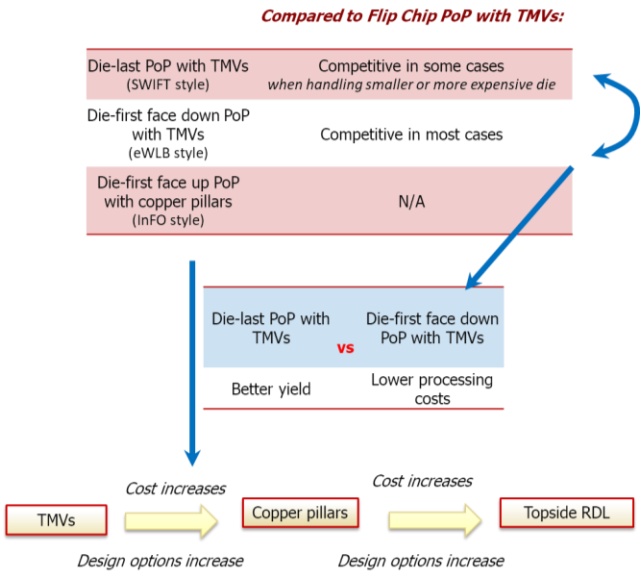

Figure 6. Summary of Technologies Compared

The process flows for FOWLP and traditional PoP technology utilizing a bottom flip chip package and TMVs were introduced. Three main FOWLP variations—die-first face up, die-first face down, and die-last—were also described.

The most equivalent comparison between FOWLP PoP and flip chip PoP was determined to be die-last FOWLP PoP versus flip chip PoP, with both using TMVs and capillary underfill. For the largest package tested, flip chip PoP was always cost-effective; for the smallest package test, FOWLP PoP was always cost-effective. This highlighted the fact that package size is a major cost driver for FOWLP. The analysis also showed that the more expensive the die being packaged, the more likely FOWLP PoP would be cost-effective over flip chip PoP. This is because FOWLP has a yield benefit, so it is more likely to be cost-effective when handling more expensive die.

Die-first face down FOWLP PoP was also compared to flip chip PoP with TMVs, and the FOWLP option was shown to always be the cost-effective choice. Die-first face down FOWLP has some cost advantages over die-last FOWLP, such as not requiring the expensive application of bumps or copper pillars to the incoming die and boasting a simpler temporary handling process.

Die-last FOWLP PoP was included in the analysis. The die-last FOWLP PoP process analyzed utilized large copper pillars as vias instead of TMVs. Because large copper vias will always be more expensive than vias drilled through mold and filled with solder, die-last FOWLP PoP was not compared directly to traditional flip chip PoP with TMVs. Instead, the increase in cost incurred by changing from die-first face-up PoP with TMVs to die-last PoP with copper pillars and a topside RDL was evaluated. The use of copper pillars and a topside RDL allow for designs not supported by other PoP options, so the cost increase may be worth it.

REFERENCES

[1] D. Butler, “Fan-Out Wafer Level Packaging: Breakthrough advantages and surmountable challenges,” Solid State Technology, August 2016.

[2] E. J. Vardaman, “A new era in packaging: FO-WLP emerges,” TechSearch International, Inc, ICSJ 2017.

[3] F. Carson, M. Bunyan, “Package-on-package trends and technology,” Solid State Technology, July 2006.

[4] L. Smith, “Package on package (PoP) applications, requirements, infrastructure and technologies,” Amkor Technology, Inc, SMTA Oregon chapter meeting, March 2011.

[5] J. Sprovieri, “’POP’ goes the future,” Assembly Magazine, September 2008,

[6] D. Lammers, “Fan-out is a game changer,” Nanochip Fab Solutions, Decemer 2015.

[7] Amkor website, “Silicon Wafer Integrated Fan-out Technology.”

[8] Y. Lin, C. Kang, L. Chua, W. K. Choi, S. W. Yoon, “Advanced 3D eWLB-PoP technology,” ECTC 2016.

[9] “Advanced Packaging “Update,” TechSearch International, Inc, September 2016.

[10] M. Lu, “Wafer level fan-out and embedded technology for portable/wearable/IoT devices,’ SPIL, Semicon Taiwan 2015.

[11] “Molded underfill (MUF) technology for flip chip pckages in mobile applications,” STATS ChipPAC Inc, ECTC 2010.

[12] A. Lujan, “Yield comparison of die-first face-down and die-last fan-out wafer level packaging,” ECTC 2017.

[13] L. Wood, “Analysis of Apple's A10 iPhone 7 Plus Application Processor: The TSMC Integrated Fan-Out (inFO) Package - Research and Markets,” Business Wire, November 2016.