Chip Scale Review®

ChipScaleReview.com

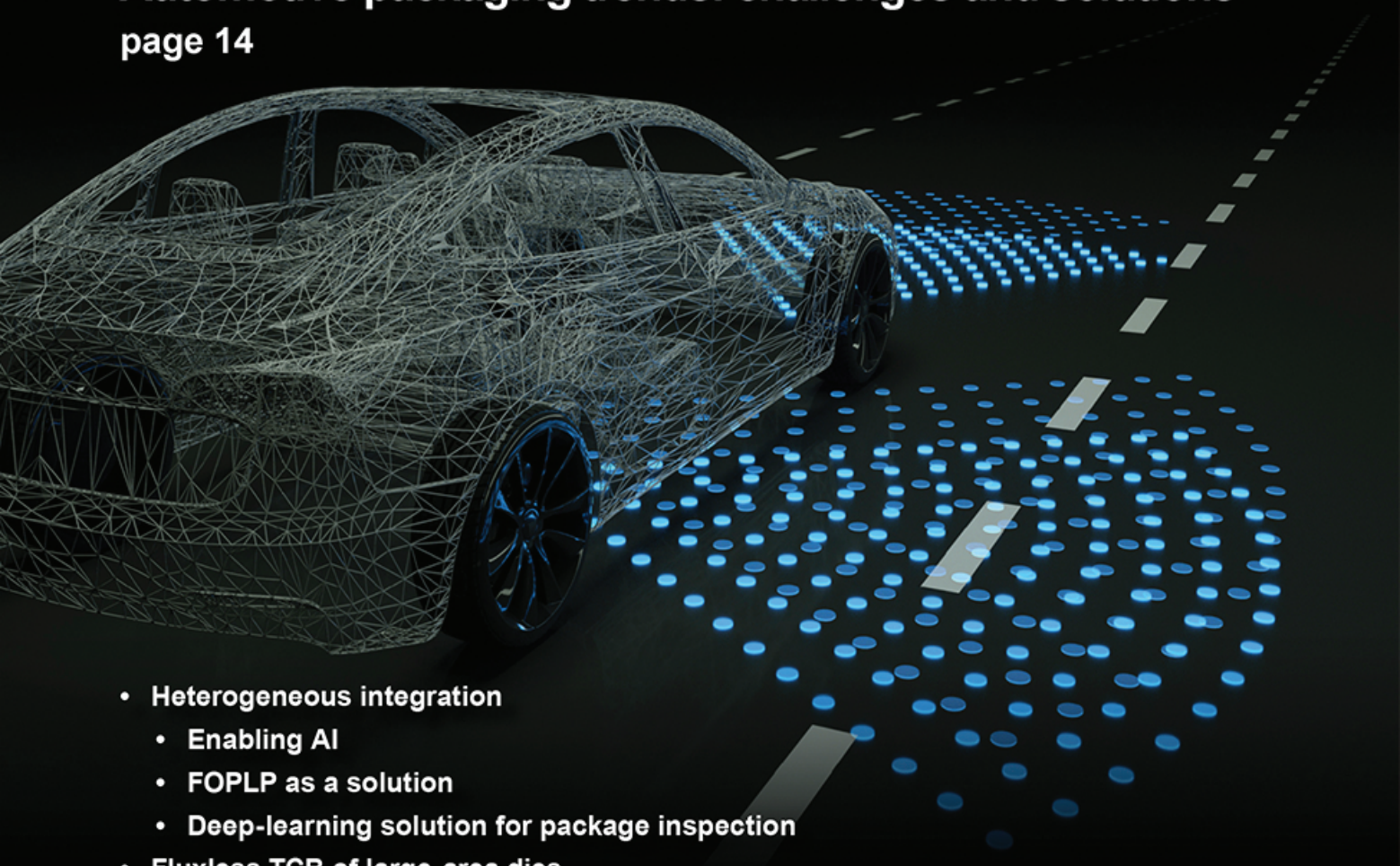
The Future of Semiconductor Packaging

Volume 24, Number 5

September • October 2020

Automotive packaging trends: challenges and solutions

page 14



- Heterogeneous integration
  - Enabling AI
  - FOPLP as a solution
  - Deep-learning solution for package inspection
- Fluxless TCB of large-area dies
- Moving automotive chip quality to zero defects
- A paradigm shift in simulation techniques of semiconductor test sockets

KATALYST™

DELIVERS INDUSTRY LEADING PERFORMANCE

Kulicke & Soffa's **Katalyst**™ delivers the industry leading performance in accuracy and speed for flip-chip placement. Its hardware and technology enable < 3 µm accuracy on substrate or wafer, offering lower Cost-of-Ownership in the industry.

YOUR ALLY IN TECHNOLOGICAL DISCOVERY, ALWAYS.

Kulicke & Soffa®

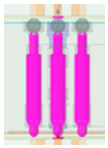
CONTENTS

September • October 2020
Volume 24, Number 5

The market growth in advanced driver assistance systems, electric vehicles, and autonomous vehicle technologies is driving the need for a zero defects strategy in the manufacture of ICs. Implementing such a strategy will require a holistic approach to automation that also meets the investment constraints facing legacy manufacturing facilities.

Photo courtesy of iStock/Just_Super

DEPARTMENTS



FEATURE ARTICLES

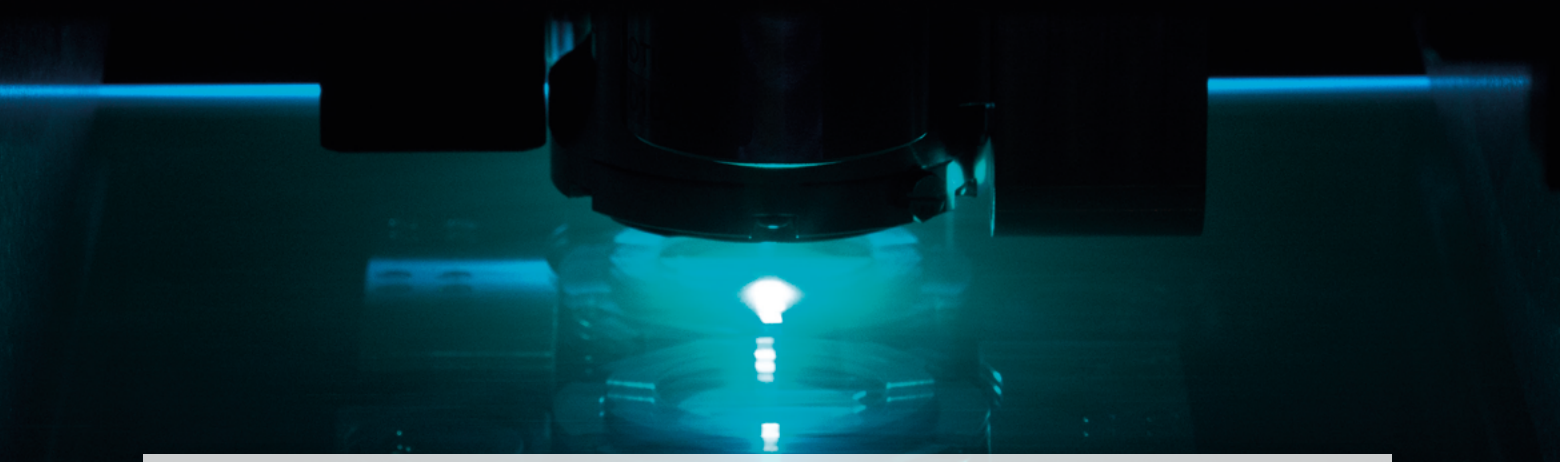
EVG®

invent
innovate
implement

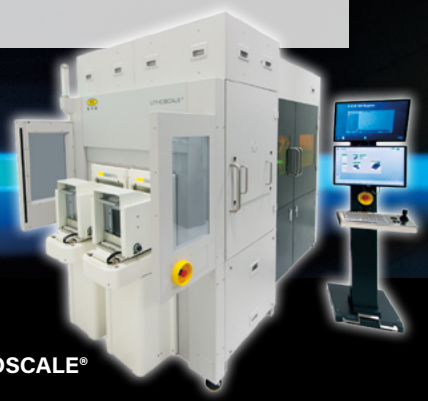
www.EVGroup.com

LITHOSCALE®

MLE™ Maskless Exposure Lithography System



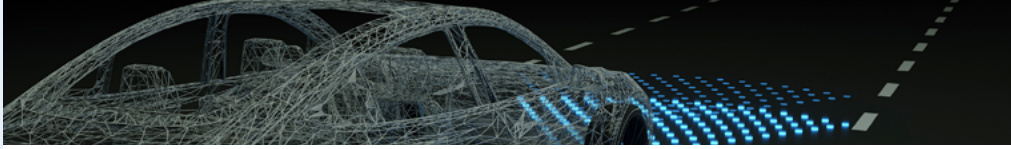
- Featuring next-generation digital lithography technology for high-volume manufacturing
- Combining high resolution with powerful digital processing to enable next-generation heterogeneous integration applications
- Scalable solution accommodating R&D and HVM needs in one system without adding on footprint
- Addressing lithography needs for Advanced Packaging, MEMS, Biomedical and High-density PCB applications



GET IN TOUCH to discuss your manufacturing needs
www.EVGroup.com

EVG® LITHOSCALE®

Chip Scale Review®
ChipScaleReview.com
The Future of Semiconductor Packaging



Volume 24, Number 5
September • October 2020

Subscriptions in the U.S. are available without charge to qualified individuals in the electronics industry. In the U.S. subscriptions by first class mail are $125 per year. Subscriptions outside of the United States are $225 per year to other countries.

Chip Scale Review, (ISSN 1526-1344), is published six times a year with issues in January-February, March-April, May-June, July-August, September-October and November-December. Periodical postage paid at Los Angeles, Calif., and additional offices.

POSTMASTER: Send address changes to Chip Scale Review magazine P.O. Box 2165, Morgan Hill, CA 95038

Printed in the United States

FEATURE ARTICLES

AD20T/S
Semi-automatic
Dicing Machine

SS10
Semi-automatic
Dicing Machine

A-CS-300
Automated
Wafer cleaning
machine

ML300EX FHWH
Laser Dicing Machine

To a new world

ACCRETECH DICING MACHINES SERIES

AD3000T-PLUS
Full-automatic
Dicing Machine

AD2000T/S
Full-automatic
Dicing Machine

PS 300
Package
Singulation Machine

ACCRETECH

Accretech America Inc.
2280 Campbell Creek Blvd., Suite 300, Richardson, TX 75082, U.S.A.
Tel: + 1-214-459-1688 Fax: + 1-214-459-1696
https://www.accretech.jp/english/index.html

A paradigm shift in simulation techniques of semiconductor test sockets

By Kevin DeFord, Khaled Elmadbouly, Jiachun (Frank) Zhou, Robert Friedt [Smiths Interconnect]

With the rise of the Internet of Everything (IoT), 5G, artificial intelligence (AI) and augmented reality (AR), high-performance test socket technology must keep pace. Performance specifications for test sockets published by suppliers should only be used as a general guideline in selecting the product family for a test application. Once a socket technology is chosen based on the general specification, suppliers should provide a more representative simulation of the socket based on the customer's package layout with a focus on the high-speed areas of the device. Typically, simulations focus on the socket in a vacuum and do not consider other features such as the printed circuit board (PCB) pad, vias or the ball grid array (BGA) ball. As data rates and bandwidth continue to increase, a paradigm shift is emerging in the industry that is driving socket suppliers to provide more detailed simulations, which include the device package and PCB interfaces in the analysis because of their impact on final socket performance in the system.

Most socket suppliers publish socket characterization data based on an optimum signal/return pin layout (e.g., signal with surrounding returns) as shown in **Figure 1**. Socket test and characterization are usually done in the design validation stage of development with test coupons that represent the pin and socket structure, but do not include any of the parasitic effects caused by the PCB and package alignment features of the socket. The measurement produces S parameters of the entire test setup (Fixture A – device under test [DUT] – Fixture B). The test socket is then de-embedded and gated in order to provide basic performance data, such as the -1dB insertion loss (IL), -10dB return loss (RL), loop inductance (L), and mutual capacitance (Cm) IL, without the effects of test fixtures and/or PCBs. Signal integrity performance data shown in the product collateral regarding the test sockets and contacts usually comes from measurements using this methodology. This information can then be used by the customer to understand the basic performance of the socket and pin in an ideal condition in order to help identify which socket family to use for a given application.

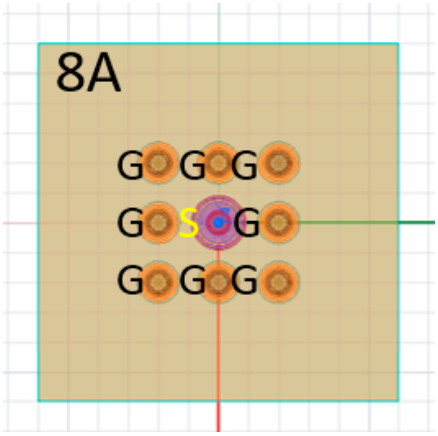


Figure 1: Example of an optimum signal/return pin layout.

Once the buyer chooses the socket technology based on initial performance, they can move into the next phase, which requires a detailed pin-out map of their device for simulation and mechanical socket design. The pin-out map provides a detailed layout of the signal, return and power pins as shown in **Figure 2**, which will later determine the worst-case radio frequency (RF) performance of the socket that can be expected. As the signal/return pattern changes in the socket, the standard specification data that pertains to the RF performance is no longer relevant as the bandwidth changes, depending on the return layout of the customer's device.

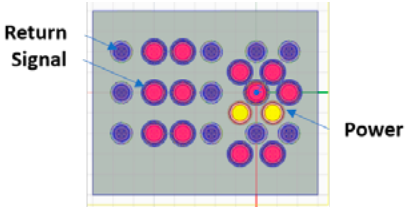


Figure 2: Example of a pin-out map.

It is good practice to design symmetrical return paths around the high-speed lanes of the customer device so the PCB and test socket can be designed to provide impedance-controlled differential pairs and guard neighboring lanes against cross talk, but real estate constraints don't always allow chip designers to follow these design rules. Socket designers turn to simulation tools such as HFSS to determine the performance of these complex layouts because it is hardly feasible to measure every instance of a customer's layout due to package to package variability. It is equally important that these simulation tools are calibrated through correlation to measured data to ensure there is confidence in the result that drives the final decision. Suppliers traditionally ignored BGA and PCB pad during simulation because they do not have control over the design of those features. Once the layout is established, and the simulation completed, the IL, RL and any cross talk effects are plotted and used to determine if the design is suitable for the application. In some cases, the S-parameter files are provided to the customer for Spice simulation at the system level, but even at this level of analysis, they do not capture the complete system performance due to the parasitic effects caused by the BGA and PCB pad and via.

As we shift our focus from simulating test socket performance in a vacuum to include the parasitic effects that PCB pads and the device present to the socket, customers will need to

provide these details, along with the pin-out map, in partnering with socket suppliers in the development of the socket technology in order to achieve the best possible system performance. When the socket designer develops a customer solution, a design standard is created to define the mechanical attributes of the socket family that are critical for the customer's package. Customers are beginning to provide simulation standards that dictate PCB copper thickness, pad diameter, via diameter and length, dielectric constants and loss tangents, and target impedance for the design. It is critical that socket impedance is matched as closely as possible to the PCB impedance to ensure good signal transfer. The target impedance from customer to customer does not always follow the traditional 50Ω or 100Ω standard that is assumed in the absence of such information. The socket designer can then focus on optimizing the interfaces and tuning the socket design to account for these effects. This also assures the customer that simulations are done to their standard and there is consistency from one SI engineer to another.

We developed a physical test sample to include a PCB with a short via to interface with the bottom of the socket and BGA spheres, and a PCB with a short via transition on the top and simulated the same structure to validate the model. An example of the new simulation set-up used to validate the measurement is shown in **Figure 3**. The simulation and measured results match very well as shown in the single-ended TDR plot in **Figure 4**. The original TDR measurement without the PCB and BGA used in our standard characterization method to produce the specification data is shown in **Figure 5**. When comparing the two TDRs, it is obvious that the PCB and BGA create many impedance discontinuities on the signal transmission, which impacts the overall performance. When the socket designer sees the whole stack, the physical design can be optimized to account for these transition points.

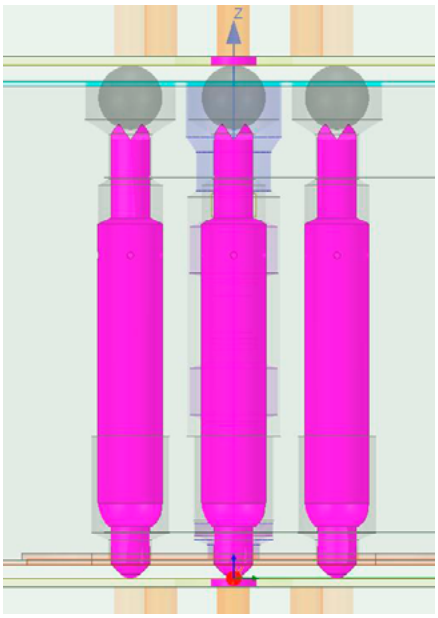


Figure 3: Example of a new simulation set-up used to validate measurements.

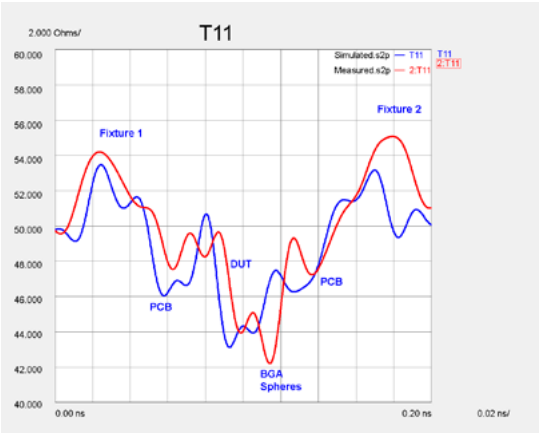


Figure 4: A single-ended TDR plot.

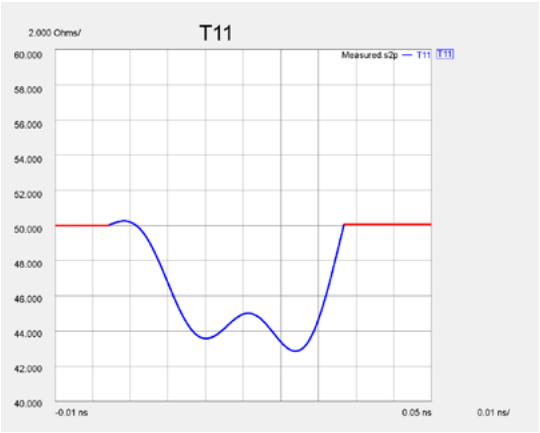


Figure 5: The original TDR measurement without the PCB and BGA used in the standard characterization method to produce the specification data.

In conclusion, socket suppliers provide specification sheets for socket and pin families to provide customers a way of comparing one design to another. These specifications should only be considered as a reference point in making an initial decision on the preferred socket technology for an application. The customer should then provide a detailed pin-out map of their device showing the location of high-speed data lines that provides the SI engineer information necessary for further device specific simulation. Simulation tools should always be validated by correlating the results with actual measurements to ensure confidence in the results. Traditionally, simulations have only focused on the socket without including parasitic effects caused by PCB pads and the BGA. This could set the expectation for socket performance which, when included in the customer's Spice simulation, may fall short of reality. IC manufacturers are beginning to set standards for simulations and provide the information needed to create more accurate models, which gives the socket designer the opportunity to optimize interfaces, as well as the entire system's performance. When PCB pads, via transitions and BGAs are included, the customer is given a more representative model that can be used to determine real-world effects and total system performance. As data rates continue to increase, chip and socket manufacturers will need to continue to partner closely to resolve the ever-evolving complex issues facing the industry.

Biography

Kevin DeFord serves as Director of Engineering Test for Smiths Interconnect. He has more than 25 years' experience in interconnect qualification and 15 years in Semiconductor Test. Email Kevin.DeFord@smithsinterconnect.com

PROCESS SOLUTIONS, EQUIPMENT AND SERVICES

Keeping It *Cool* for "Beyond Moore" Advanced Packaging

Lower Temp Cure with Vacuum

VertaCure XP
High-volume automated system for polyimide vacuum cure

PB Series
Manual system for polyimide vacuum cure

Your roadmap looks cooler than ever.

YES's industry-leading vacuum cure lets you take the temperature down, to make those advances real. Reduce stress/warpage in 3D stacks. Leverage the all-copper interconnect trend. Wafer or panel, R&D lab or high-volume fab – we've got you covered. YES supports your move to low-temp technology with unrivalled film quality, excellent yield and superior throughput.

Yield Engineering Systems, Inc.

Call: **1-510-954-6889** (worldwide) or **1-888-YES-3637** (US toll free)

www.yieldengineering.com

Automotive – Driving Zero Defects

Chip Scale Review asked David F. Hanny, Director of Marketing at Applied Materials, Automation Products Group, to provide insight into how market growth in advanced driver assistance systems (ADAS), electric vehicles (EV), and autonomous vehicle (AV) technologies is driving the need for a zero defects strategy in the manufacture of integrated circuits (ICs).

CSR: Because ADAS/EV/AV market growth is raising the complexity of ICs—as well as how they are used in systems that must make almost instantaneous decisions in traffic situations—what are the most significant limiting factors with respect to achieving a zero defects strategy in their manufacture? How can you overcome those limiting factors?

DH: We see three primary limiting factors on quality as we move towards zero defects in manufacturing. First is the slow development of new technologies and materials for new product introduction. In automotive chip manufacturing, new product yield begins as low as 40% for a period before it moves up to typical yields in the 88-92% range. Next is the introduction of new raw materials, along with a third factor being errors in human decisions. Each are inhibitors of quality in the fab. These challenges can be addressed with increased requirements, measures, and validation over supplier materials and quicker learning cycles of anomalies. Moving decisions from offline human decisions to real-time decisions based on data patterns enables the factory to increase product quality (**Figure 1**).

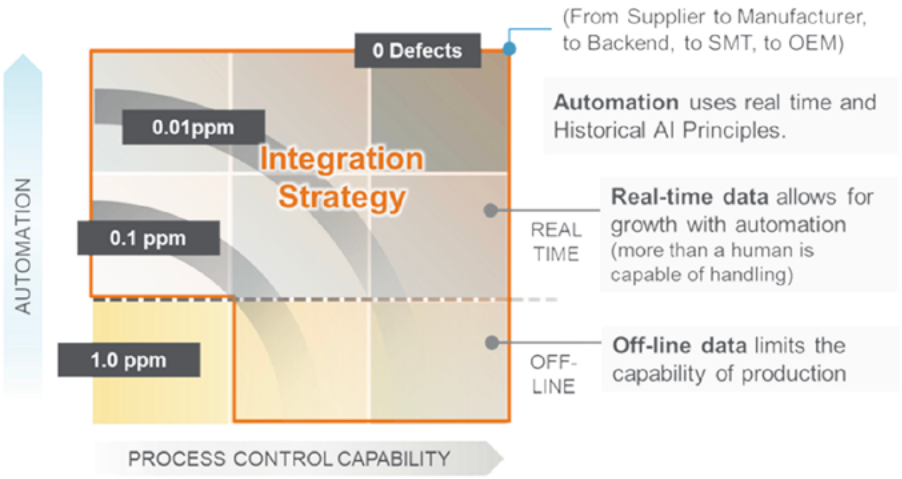


Figure 1: An illustration of end-to-end quality. Moving decisions from offline human decisions to real-time decisions based on data patterns enables factories to increase product quality.

CSR: How can the industry improve the way field failure data is married to quality issues with respect to semiconductor processes in the fab or at the outsourced semiconductor assembly and test (OSAT)/packaging supplier?

DH: The real challenge with performing failure analysis is that it relies heavily on the genealogical granularity of the data throughout the supply chain. Not every factory has the same level of ability to diagnose, and the process can be very manual. 300mm factories have developed more tools and have greater access to this data. Often the node in the supply chain that can't afford to answer the question gets stuck with the bill. To combat this challenge many packaging factories and surface mount technology (SMT) lines are beginning to increase their automation capabilities. At least one major original equipment manufacturer (OEM) has increased the automation requirements on their packaging suppliers, requiring them to add more sensor monitoring (like fault detection) to maintain their status as a valued supplier. A common trend for packaging and SMT lines is seeking for more advanced quality capabilities.

CSR: What role is artificial intelligence (AI) playing in the end-to-end quality chain? Can you describe in more detail how AI techniques are being developed and improved upon from earlier approaches?

DH: Factories operate in varying degrees of automation from operator-driven to the early stages of full automation (see phase 3 in **Figure 2**). Many companies are challenged to have the kind of end-to-end quality to leverage AI. The primary reasons are due to the economics and infrastructure of today's factories. Chips that are highly specialized for automotive applications such as ADAS and light detection and ranging (LIDAR) are typically manufactured in 300mm fabs that have infrastructure and systems running at

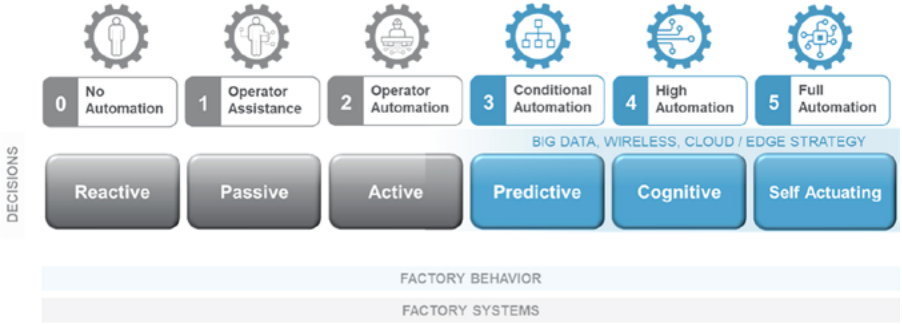


Figure 2: The roadmap to full automation: Factories operate in varying degrees of automation from operator-driven to the early stages of full automation.

varying degrees of conditional automation and higher (Figure 2). These factories have become equipped to take the next step towards AI. Most chips for automotive applications are being manufactured in less sophisticated factories. These factories lack data systems and enough of the right resources (money and people) to master AI. For the industry to move forward, increased levels of commitment are required. Furthermore, an understanding of the applications that leverage Industry 4.0 concepts and technologies are required to move towards AI. Companies we have seen succeed typically begin when they have a "technology visionary" followed by executive sponsorship and a willingness to change to something better than the status quo.

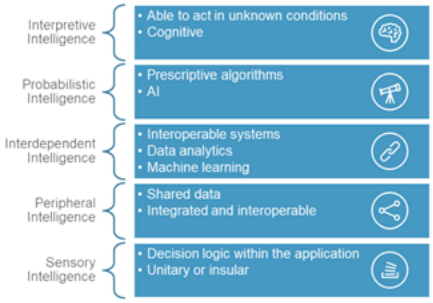


Figure 3: Quality intelligence levels: A holistic quality approach can be addressed in terms of levels of intelligence, starting with sensory intelligence, in which quality systems can make decisions within the scope of their data.

CSR: How would you characterize the way advanced excursion control, along with metrology capacity and dynamic metrology, achieve a holistic approach to a customer's quality strategy?

DH: We define a holistic quality approach as one that can learn and utilize three fundamental types of definitions of intelligence. We talk about these in terms of levels of intelligence (Figure 3). At the first level – sensory intelligence – quality systems can make decisions within the scope of their data. This is a passive automation system (Figure 2). When two or more systems share data to increase the effectiveness of a decision, we call this peripheral intelligence. Next, interdependent intelligence occurs when systems work together to achieve common objectives. At this level, machine learning begins and is supported by data analytics. In our view, this level provides a holistic approach to customer quality and maps to the predictive automation state. For factories, these are stepping stones of continuous improvement to increase quality.

CSR: Are there any other significant aspects to achieving a zero defects quality control strategy for automotive applications?

DH: There are no silver bullets, nor does anyone start at the top of the mountain. Our experience is that achieving zero defects really does start with the technology visionary who understands the manufacturing business impact that technology can deliver. Selecting the right partners and suppliers is critical. This process often results in replacing legacy systems (homegrown or under-capable) that will leverage industry learning, as opposed to single fab learning. This effort often requires a shift in culture and change in processes and skill profiles. Finally, it is important to not underestimate the challenge of change. But for those willing, the reward of increased quality can be realized.

Biography

David F. Hanny is Director of Marketing at Applied Materials, Automation Products Group, Salt Lake City, UT. He has been working in the nano-manufacturing industries for the past 25 years as a supplier of automation software, bringing multiple innovative software products to market that are being used in semiconductor, display, and pharma industries. He has done extensive work in the areas of material handling controls, RFID, and advanced short-interval scheduling. He has a BS degree in Computer Science from Weber State U. Email David_Hanny@amat.com

Moving automotive quality to zero defects

By Selim Nahas [Applied Materials, Automation Products Group] Manan Dedhia [Analog Devices]

The last 10 years have experienced an explosive demand for automotive electronic parts (Figure 1). In 2016, for example, a top-of-the-line Bentley required 110 pounds of wiring with 90 computers to connect. In 2020, similar wiring and connectivity requirements are prevalent in most cars that people buy. The automotive industry is among the least dependent on leading-edge supply chain technologies, instead choosing to use parts made on legacy nodes with proven reliability. Most of these legacy node fabs are semi-automated to manual and have been in production for 15 to 30 years. The economics that drive the decision to use 150mm and 200mm facilities is also changing—although not for the foreseeable next five years. Within the Automotive Electronics Council (AEC) and International Standards Organization (ISO), recent announcements on increased safety, design for test (DFT), and design for manufacturability (DFM) reflect the inevitable rise of Level 4 and Level 5 autonomous vehicles and automotive original equipment manufacturers (OEMs) and

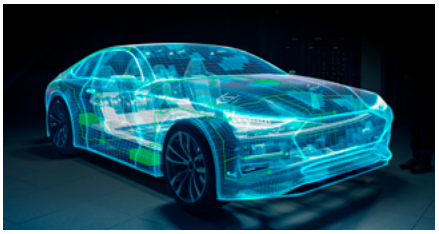


Figure 1: The automotive electronic revolution.

Tier 1s. These announcements have rightfully doubled down on a zero-defect mindset to reduce the cost of non-quality.

LB Semicon

World No.1 Specialized IC Packaging Company

Look Beyond
Your Current Needs to Make your Future Business Successful

▶ Webpage QR code
www.lbsemicon.com

www.facebook.com/lbsemicon

138, Cheongbuksandan-ro, Cheongbuk-eup, Pyeongtaek-si, Gyeonggi-do, Korea
TEL +82-31-680-1600 FAX +82-31-680-1798

Anatomy of field returns

The International Automotive Task Force (IATF) standards essentially require us to strive for zero defects. If we look at where the industry performs today, data suggests that the automotive supply chain resides at approximately one defective parts per million (PPM) or above. When reviewing field returns that constitute either warranty returns or zero-kilometer failures (Figure 2), the field

Figure 2: Fab and packaging contributions to field returns.

estimates suggest that roughly 25% of the failures come from the front end fab. Within this data set, 50% of failures that leave the facility essentially have a parametric test, but somehow elude our ability to detect the problem. Another 30% have no test coverage and therefore, no detection. Another 15% are undefined, meaning we can't assign the failure to any specific cause. This third category is essentially the unknown. In these cases, no real corrective action is deployed and the gap in detection persists. And finally, approximately 5% of the failures reflect disagreements within the supply chain as to the origin of the failure.

What eludes the supply chain

The constituents of the current supply chain – fab, packaging, and electrical test – can be evaluated as detection gates, with an increasing granularity of detection, but also in the order of increasing cost.

From a wafer fab perspective, automotive components can be broken out into the following buckets: safety/advanced driver assistance systems (ADAS), propulsion, and infotainment. First, the safety/ADAS bucket includes sensors (microelectromechanical systems [MEMS], optical, temperature) that are made on larger nodes, or radio-frequency (RF) components in specialty fabs (GaAs, GaN, SiGe, and so forth) with a lower than average level of automation and detectability. Second, propulsion is an increasingly large group with the advent of electric vehicles (EV) and can be anywhere from 180nm to 32nm. This bucket includes power components and engine control units (ECUs). These items would be considered mission-critical components in a way, requiring a higher level of reliability. Finally, infotainment does not warrant the same level of reliability and has some flexibility in terms of sampling the latest and greatest fab nodes.

Given the wide range of fab nodes that are sampled in the automotive supply chain, we are subject to a variety of available quality levels. Reducing cost per component to maximize profit also means reducing the acceptable quality level to an acceptable bare minimum. Every detection step is a non-value-added step, and hence the cost is passed off further downstream. This means that even with the best fabs in the business, PPM-level detection is neither offered, nor discussed for parts that are in volume production. This situation deteriorates further as older nodes or specialty technologies are used.

The best chance for detecting defects on a part is with electrical testing, which entails wafer probe or final automated test equipment (ATE) in packaged form. Conceptually, if the part has been characterized thoroughly and built on a known technology with a low defect rate, then any remaining PPM-level failures can be captured at the electrical test step. Characterization depends on: 1) the design failure mode and effects analysis (DFMEA) being able to simulate all failure modes; 2) product engineers being able to test the parts to cover all customer mission profiles; and 3) with increasing software components in the parts, ensuring that data fidelity is maintained throughout.

Most parts do not have wafer-level traceability, which further degrades the ability to tie back failures to fab processing as warranted. In the face of meeting ultra-aggressive customer timelines and cost pressures, we again see the acceptable quality level of this stage reduced to the bare minimum, which ensures that the outgoing product does not receive the full benefit of detection at electrical test. This translates to most customer failures being test-related issues and increased costs of non-quality and further results in the motivation to move towards a holistic approach as a matter of progress and safety.

Framing the issue

If we are to consider a strategy to move the needle closer to a zero-defect concept, then several things must change from the way facilities operate today. Most of these legacy facilities are semi-automated or manual. This means that they fundamentally have a multitude of point solutions and disciplines to govern their quality standards. Simply put, they neither capture all the data needed to govern their processes effectively, nor analyze this data in a manner that allows for speedy and effective feedback. Based on field returns and internal failures, we have seen how far this strategy can take us. It is unlikely that we will be able to breach the 1PPM barrier consistently without rethinking the process entirely.

To paraphrase the problem, we have an inability to scale our detection, coupled with the inability to assure our test coverage compounded by an inability to assign root cause with total certainty in too many of our cases. All these problems result from a patchwork approach to quality and a siloed perspective on quality data management. A holistic approach would streamline information sharing and facilitate first-time-right decision-making (Figure 3). So why haven't systems evolved to be holistic?

Adopting the needed changes

The fundamental flaw with point solutions is that they don't address the challenging issues observed at the fab level, with packaging and surface mount lines. The solution to this problem originates from a holistic approach. Holistic in this case refers to the ability to look at the entire manufacturing line. Beyond the single facility, holistic means the entire supply chain as a single entity. Moreover, we must adopt new principles of signal detection to successfully manage chart scaling. Most systems today are predicated on the idea that if a measurement chart is set up properly for a known parameter, it will detect anomalies accordingly. While in principle this is true, it becomes a daunting problem to manage 100,000 charts in a single facility with a skeleton crew. Each measurement value irrespective of control and specification limit

is given a meaning that relates to the overall quality of the device. To put it another way, it's the ability to review the effect the sum of variation has on the performance and reliability of a part. This approach is not currently practiced in the industry.

Cost of implementation and training

The rate that semiconductor fabs accommodate new automation solutions poses a barrier to being able to quickly change the quality capabilities of the fab. Most automation systems today are still fundamentally built around the same paradigm seen over the last 20 years—charts driven by Western Electric rules and out-of-control action plans (OCAPs) that are based around the errant chart. By the time the industry accessed streamlined automation solutions, legacy manufacturing facilities had long since amortized their building costs and depreciation regimes, running the business with just the operational essentials. This situation left minimal resources for new development because the revenue capacity of these facilities does not easily support the required investment to resolve any singular problem.

Holistic automation solutions address systemic problems rather than single gaps. Half a million dollars can easily represent half a percent in profits depending on the facility, which is a significant erosion to margins. Furthermore, the return on investment is either too small to justify the risk or takes too long to achieve. Point solutions can easily range from $150K to $750K, which is a difficult barrier to breach. Understanding the requirements and the intricacies of the domain and the information technology (IT) infrastructure required to support it takes a substantial investment. There is a significant cost to define and test disruptive systems that can replace the multitude of current automation practices. Unless a solution can provide a systemic resolution to quality – meaning resolve several high-value targets – these facilities will not be able to invest. These paradigms need to straddle the entire supply chain.

Most personnel in these facilities are not focused on developing new automation solutions, but rather on manufacturing reliable parts cost effectively. So, accessing the technology to build a holistic and streamlined quality system is not a realistic expectation for these facilities. It would require a financial and manufacturing mindset change. A strong holistic quality program requires both automation and learning systems for those responsible for operating it. While numerous opportunities exist to automate tasks and decisions that historically were manual, the expectation will remain high on understanding the meaning of signals and quality metrics.

Moving to zero defects

There is certainly a hierarchy of automation need in legacy facilities. Automated data acquisition is a good starting point, followed by a centralized alarm management system to connect signals and prevent moving materials into process tools that should not run product. A significant need also exists to reduce human error as part of the processing. Recipe management systems play a key role as does centralized configuration management. After the automation layer is in place, the ability to expand the capabilities increases. For instance, once the statistical process control (SPC) practices of the fab access the raw data of a wafer measurement, the engineering community can make use of it to isolate within-wafer variation problems. Qualifying new products is time consuming and preparing for production can be slow and prone to missing opportunities to define needed tests. The more understood the variation sources of a facility, the more opportunity to quickly qualify new products. This knowledge will directly impact the gap of test coverage that allows 30% of field failures to persist.

Moving these traditional good practices to a holistic approach requires some integrated infrastructure. The first step is to define what holistic means in this context. Holistic is the ability to understand the interdependent behavior of the process steps. The control plan illustrates the known measurements associated with any product. The automation system will need to provide users with the ability to define interdependency relations of different process steps, as outlined in the control plan and process failure modes and effects analysis (FMEA). At runtime, the system will consume the data coming from multiple process steps, as defined in the control plan, and outline which data does not fit the population that we expect for the specific processes. This will need to be done using real-time data tools because a single decision will require users to assess 15 parameters with 12 to 20 sites for each. Each site will need to have a statistic and be broken down to divulge within-wafer variance profiles, wafer-to-wafer variation within the same process step, and finally, variation inherited from upstream steps. This approach will reduce the variation arriving at final test and therefore, reduce the chance of failing parts leaving the facility. More important is the change that this approach will impose on final test validity. A more stringent variation verification will become more effective to capture a greater degree of nonconformance.

Another key difference is that holistic systems are not only error driven, but also sensitive to variation within the specification limits. Data show that this is happening in facilities that have yields ranging from 88% to 92%. The front-end fabs combined with the packaging operations account for an approximately 0.76PPM failure rate. Devices within the specification limits will ultimately be functional, but not necessarily the same from a performance or longevity perspective. Process problems such as whiskers and bridging will elude many of these tests, causing shorts in the field. So, the automation system provides users the ability to define qualitative stack-up of step attributes, which is a measure of acceptable variation for a given problem.

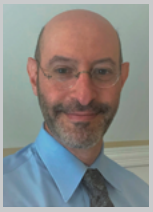
It is true that other factors play a role in the overall failure rate including electrostatic damage and other forms of mishandling that can occur in several places throughout the supply chain. This highlights the need to have rapid genealogy of quality attributes that include design. Design is the concept of parts that have been made for a long time but are now used in a new way that is not suitable for their reliability. This constitutes an approximately 0.15PPM contribution to failures in the field. The cases that are "unknown" represent an approximately 0.21PPM contribution to the failures and are a major liability to the vendor. In the case where no cause can be assigned, the small vendors will be held accountable for the failure and will have the cost deducted from their agreement with the big car manufacturers. If this ability could be automated to straddle the front end and back end and ultimately include the surface mount technology (SMT) line, then rapid identification would become possible for any given device and cause. The speed of resolution in this case will impact cost of liability and more importantly, safety. Unless the design of new automation systems adopts these guiding principles, it will be difficult to expect any supply chain to converge effectively towards zero defects (Figure 4).

Figure 3: Moving to holistic and intelligent systems.

Figure 4: Striving for zero defects.

Reference

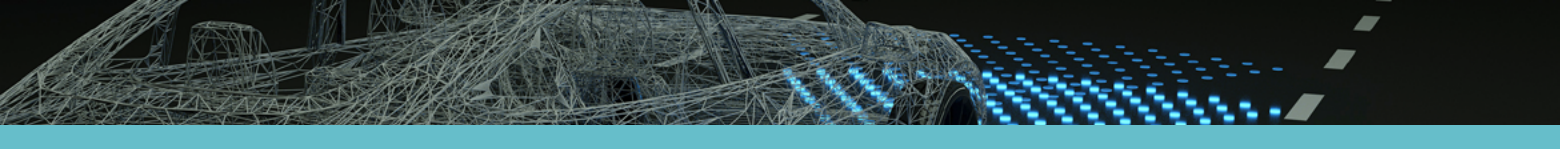
Original Equipment Suppliers Association Page 54 - https://www.oesa.org/sites/default/files/automotive-defect-recall-report-2019.pdf



Biographies

Selim Nahas is the head of the Process Quality Group, which is part of the Applied Materials, Automation Products Group. He has 25 years of experience in Semiconductor factory automation systems including Material Execution Systems, Automated Material Handling Systems and Advanced Process Control. Selim holds degrees in Industrial Engineering and Arts and Sciences from Northeastern U., Boston, MA. Email Selim_Nahas@amat.com

Mana Dedhia is a Quality Engineer at Analog Devices, Inc., Wilmington, MA. He was born in India and completed his Bachelors in Electronics Engineering from the U. of Pune and his MS in Electrical Engineering from the U. of Minnesota, Twin Cities. He started his semiconductor journey at Globalfoundries in Fab 8, Malta, NY in 2012.



Automotive packaging trends: challenges and solutions

By Thorsten Meyer, Ulrich Abelein, EungSan Cho, Bernhard Knott, Stefan Macheiner [Infineon Technologies AG]

Electronic components entered the automotive area in the 1950s and 1960s with the introduction of semiconductor transistors in car radios and power diodes in alternators. Since then, electronics have spread into all relevant areas of automotive transportation. Today, up to 80% of all innovations in a modern car are supported by electronics that address applications in all areas of motor and chassis functions, comfort, security, and safety. The majority of these innovations in automotive applications support three megatrends: 1) autonomous driving, 2) electro-mobility, and 3) connectivity.

Autonomous driving. The implementation of autonomous driving is divided into five different levels according to the Society of Automotive Engineers (SAE) [1]. While in level 1, the driver is only supported by some assistance systems (e.g., anti-lock braking system [ABS], etc.) during the car's operation. In the highest level 5, however, the driver will become a passenger in a fully-automated vehicle. The higher the automation level, the more support is needed from advanced driving assistance systems (ADAS). Therefore, ADAS is generating a strong demand for high-performance computing power, as well as various sensor technologies, preferably in complex system in package (SiP) solutions with multiple integrated components.

On the one hand, driverless operation requires highly-reliable, high-performance packaging solutions to cope with the expected use time extension. On the other hand, the rising gap between ambient and junction temperature requirements (because of enhanced self-heating), means that the exposed mounting locations (e.g., for sensors) with direct contact to corrosives demand highly-reliable, top-quality components.

Electro-mobility (eMobility). Electro-mobility will provide a big step towards the vision of zero emissions. In addition to driving and parking, which are status quo for traditional combustion engines, eMobility requires additional operating states like on-grid parking, vehicle-preconditioning (for the battery, as well as for driver comfort, e.g., cabin heating) and charging. The inevitably increasing operating times and the need for highly-efficient power electronics (e.g., SiC) with rising operational temperatures up to 200°C will drive innovation, especially in the materials area.

By adding sensors and microelectronic components to eMobility systems, the heterogeneity and complexity is increasing without losing sight of the need for fast time to market and low cost. These requirements will require highly innovative solutions in chip design, technology, and especially in packaging. Novel operating states, e.g., for vehicle charging, are coming with the applications developed within this framework. The results are significantly extended lifetime requirements. The AEC-Q100/101 stress test conditions are no longer suitable to qualify a package according to these mission profiles.

Connectivity. Connectivity will develop from connected infotainment to "car-to-x" communication. There will be a strong link to autonomous driving as well, e.g., the use of swarm intelligence. Software updates have to be possible "over the air," therefore, the vehicle has to be permanently accessible for any requests from the backbone/customer. Connectivity will be one of the main drivers that increases the operational time of automotive packages (and ECUs), mainly SiP at small nodes in this area.

All three megatrends in the automotive arena require increased integration of components in order to fulfill the performance and dimension requirements. Packages have to fulfill increased reliability requirements caused by the extension of operational times and thermal/electrical requirements. As complexity increases, there will be no single package solution that fulfills all needs. System integration with technology and packaging features from consumer electronics that have been adapted and qualified for the harsh conditions of automotive applications will therefore be one approach for future uses. Each of these megatrends is discussed in the sections below.

Autonomous driving

Autonomous driving means fully-automated driving—the people in the car become passengers. How will semiconductor packaging be affected by this development? The answer is that, because all tasks related to driving a car, such as accelerating, braking, or steering will be taken over by the car itself, the amount of sensors, actuators and controllers will see a dramatic increase. In addition, the car needs to recognize situations and act accordingly. Control units with increased computing power are needed to deal with this "big data" requirement. Furthermore, safety critical applications like steering or braking need to be redundant to ensure the highest safety level.

As a consequence of the above considerations, autonomous driving will not only increase the amount of semiconductor components, and therefore, the number and different kinds of semiconductor packages, but will also demand challenging requirements with respect to power density, heat dissipation, and current capability—all combined in miniaturized packages that meet the highest automotive reliability standards. Examples of automotive package developments that support upcoming requirements for autonomous driving are as follows:

Miniaturization. Miniaturization involves both flip-chip attach on lead

DIE BONDING AND DISPENSING SOLUTIONS

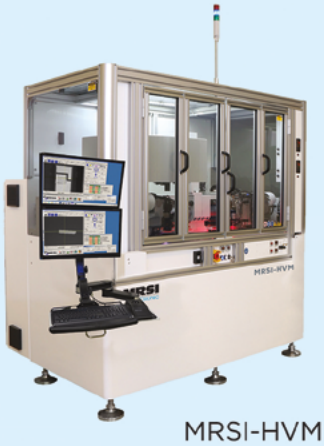
SCALE WITH US

0.5 MICRONS WAFER LEVEL PACKAGING

HIGH VOLUME

NEW

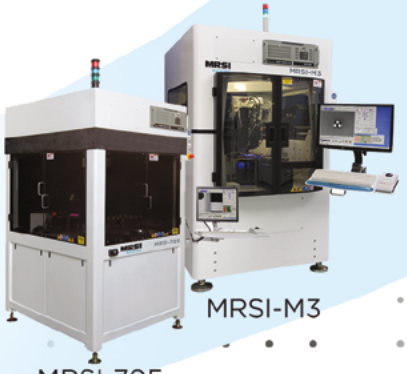
MRSI-S-HVM



MRSI-HVM

R&D/ MEDIUM VOLUME

MRSI-H



MRSI-M3

MRSI-705

Sensors
Photonics
Optoelectronics
Microelectronics
Microwave and RF

MRSI MYCRONIC

+1 978 667 9449 | sales.mrsi@mycronic.com | mrsisystems.com
YOUR STANDARD IN PRECISION AUTOMATION®

XPERI®

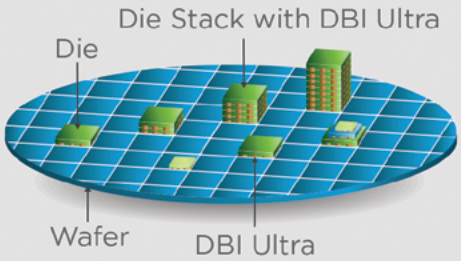
Semiconductor Technologies | invensas™

DBI® Ultra Die to Wafer Hybrid Bonding

The Ultimate 2.5D & 3D Integration Technology for High Performance Computing

High Bandwidth, High Capacity, Thin Profile, Low Power, Low Cost

Data Centers | AI, Machine Learning & Deep Learning Hardware | Automotive | Gaming | Consumer Electronics | Industrial & Scientific

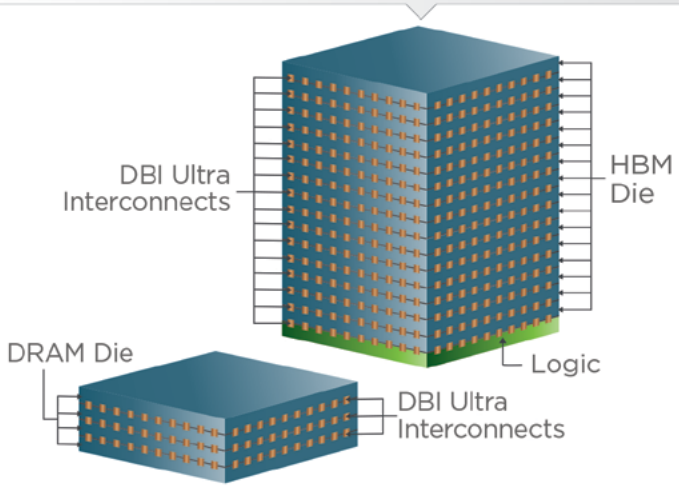


Ultimate Integration Flexibility

Accommodates various die sizes, wafer sizes, process technology nodes, etc.

Enabling 3D Stacked Memory Solutions

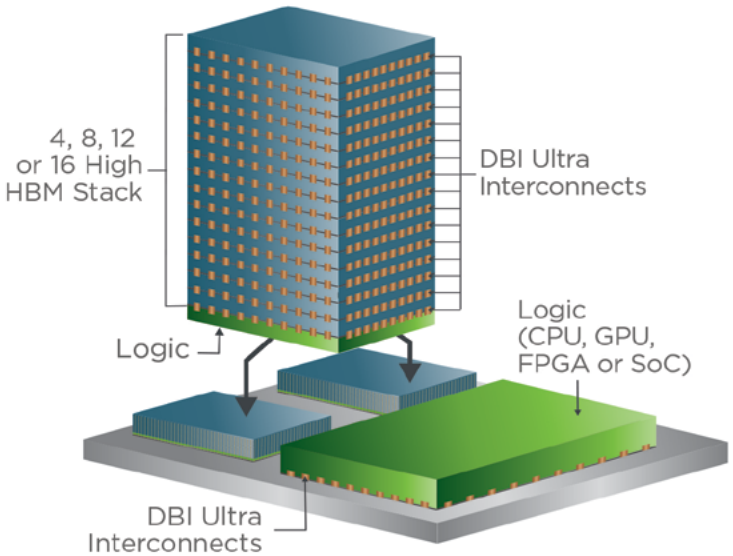
4, 8, 12, 16 or more high HBM2, HBM3 & beyond



4 High 3DS DRAM

Enabling Next Generation High Permanence Computing

2.5D Integration with DBI Ultra



invensas@xperi.com | www.invensas.com

3025 Orchard Parkway San Jose, CA 95134 +1 408.321.6000

frame-based packaging, and the use of an integrated half-bridge. These topics are discussed below.

Face-down chip assembly enables two major improvements compared to wire bonded solutions: 1) optimized/smallest package footprint, and 2) the shortest interconnect technology (Figure 1). The biggest challenge of flip-chip attach packages is the limited reliability on the printed circuit board (PCB). Because

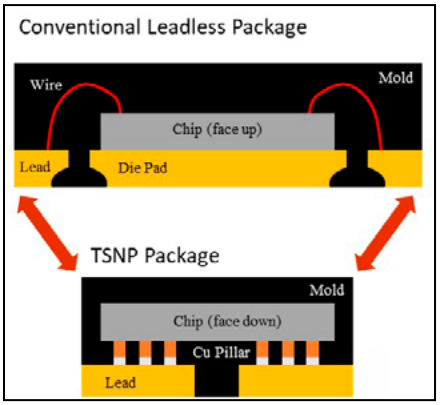


Figure 1: Package shrink-wire bond versus flip chip.

the chip is soldered face-down on the substrate, copper pillars need to buffer the stress from the coefficient of thermal extension (CTE) mismatch between the Si/chip (~3ppm/K) and the laminate/PCB (~16ppm/K). An intensive design study was performed and improvements with respect to bump design, bump layout on chip, and process optimizations were implemented. The results of the study showed that the temperature cycle on board (TCoB) performance allows usage in automotive applications, even in a high-temperature environment (e.g., an ambient temperature of 150°C).

The second topic with respect to miniaturization is use of an integrated half-bridge. Electrical power steering (EPS) functionality as a safety-critical application must be guaranteed under all circumstances. For autonomous driving, this would mean that such an application needs to be redundant to ensure safe operation, even if one component is failing.

Having system redundancy simply means to double up the system. Figure 2 shows an EPS system in which the bridge consists of 6 MOSFETs. The redundant system would also be built with 6 MOSFETs, so, 12 MOSFETs in total. With an integrated half-bridge solution, not only

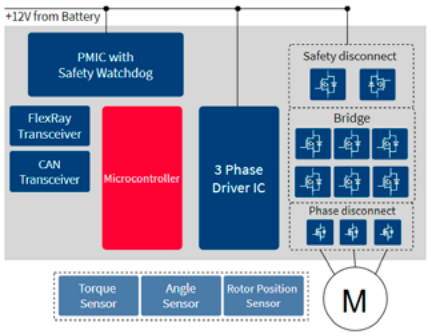


Figure 2: An EPS system diagram [2].

can the board space be optimized, but stray inductance and minimized switching losses with advanced electromagnetic interference (EMI) performance could also be realized (Figure 3).

Increased heat dissipation. To manage increasing power densities, and thereby the resulting heat dissipation, the heat input into the PCB must be reconsidered to avoid a thermal

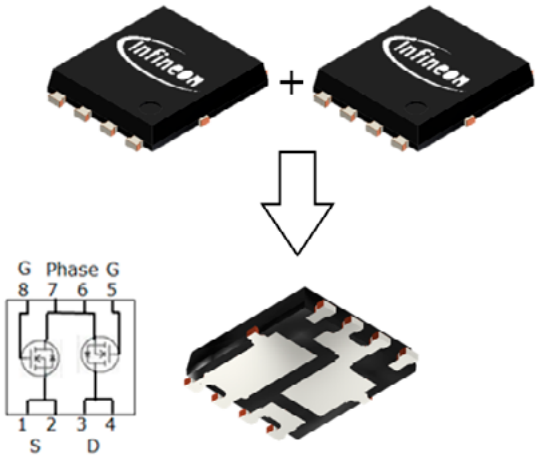


Figure 3: nfineon's integrated half-bridge in a TDSON8 package (IP generated).

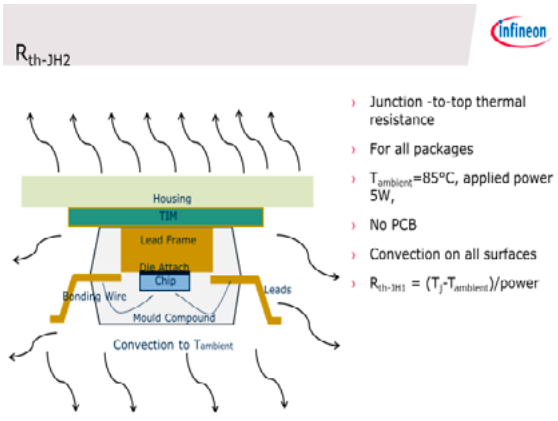


Figure 4: Top side cooling (TSC).

overload. Instead of dissipating the heat to the bottom and into the PCB, it could also be done the other way round, i.e., dissipate the heat to the top (Figure 4). With leaded packages, this would mean a simple reverse bending of the leads.

Leaded packages such as, system on integrated circuit (SoIC™), or quad flat package (QFP), typically come with a standoff (i.e., the distance between the bottom surface of the package and the bottom surface of the leads) of up to 200µm.The tolerance is related to the bending process of the leads. The tolerance of the overall package height, including the package body and the standoff, can add up to about 300µm. If one also considers tolerances from the board mounting process (solder thickness, PCB flatness, housing, etc.) this can easily add up to 500µm and more. In a topside cooling application, the thermal interface material (TIM) must compensate for this tolerance build-up to ensure a proper contact between the topside of the package to the cooling area (Figure 5). As a consequence, using a thicker TIM increases the thermal resistance of the thermal path between the topside of the package and the cooling area, thereby reducing the thermal performance of the application.

One way to minimize the package height tolerances is by way of a so-called negative standoff (Figure 6). In this case, no lead will extend the bottom surface of the package. The whole package height tolerance is determined by the package body height. With this topside cooling TOLT package, a 20% reduced Rth compared to bottomside cooling can be realized (Figure 7).

eMobility

The three megatrends we have been discussing make different contributions with respect to reliability requirements, yet they all result in a movement in the same direction. There is an overall clear visible trend

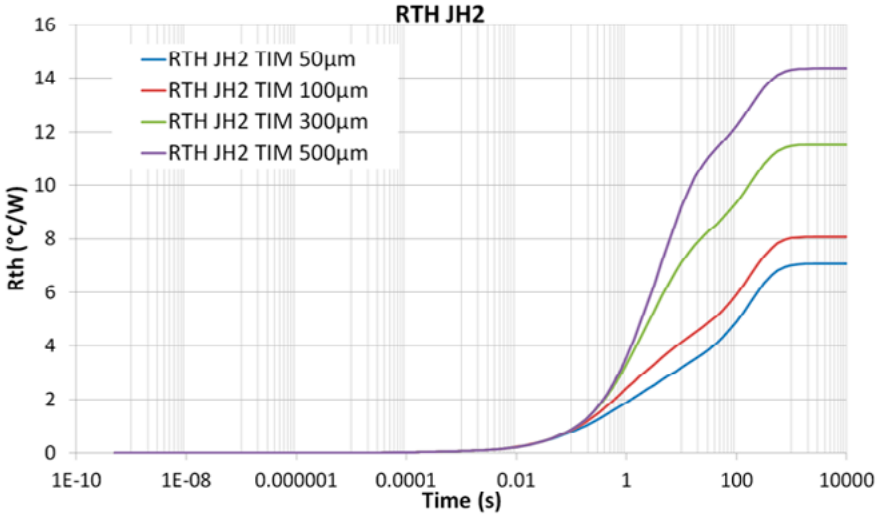


Figure 5: Impact of TIM thickness on Rth (T_{amb} 85°C; P_{diss} 5W).

Figure 6: Example of a negative stand-off.

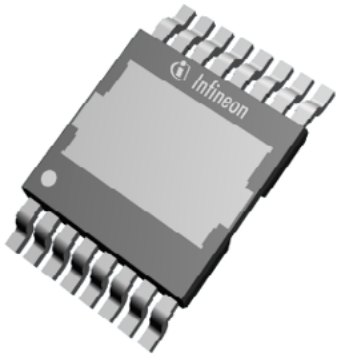


Figure 7: TOLT top side cooling with a negative stand-off (IP generated).

for automotive electronics towards: 1) longer operating times and 2) higher performance, with 3) no compromises in quality and reliability.

The following example demonstrates how changes in how a device is used influences qualification and development targets for automotive packages. **Table 1** shows a possible temperature mission profile of a microcontroller in an onboard charging system for an electric vehicle. The system is active during driving, as well as charging the battery with leads, which results in an increase of operating time to 40,000 hours compared to a typical value of 10,000 hours for the engine control of a combustion vehicle. The first important question is the degree of coverage of this temperature mission profile by a standard qualification test according to AEC-Q100. If we use Arrhenius' Law to determine the necessary equivalent test times at a defined stress test temperature, the acceleration factor A_T, depending on the activation energy E_a, can by calculated by:

$$A_T = e^{\left[-\left(\frac{E_a}{k_B T}\right)\left(\frac{1}{T_{test}} - \frac{1}{T_{operating}}\right)\right]}$$

$T_{junction}$ / °C	Time / hours
Operating	
145	400
140	3,200
76	26,000
43	8,000
-20	2,400
Non Operating	
85	914
80	7,312
60	59,410
23	18,280
-40	5,484

Table 1: Example of a mission profile.

Failure mechanism	Ea / eV	EST at 150 °C / hours
Cu-Al IMC growth	1.1	2,176
<standard>	0.7	4,437
Mold compound oxidation	0.45	10,059

Table 2: Equivalent stress time depending on the failure mechanism.

For E_a=0.7eV, this leads to an equivalent test time for high-temperature storage testing of 1,521h at 175°C. This value is well above the current AEC-Q100 requirements.

A second critical aspect resulting from increased operating times for the qualification is the increasing influence of high and low accelerated failure mechanisms on package qualification.

To understand the criticality of this phenomenon, we compared the equivalent stress times (EST) for two degradation mechanisms with high and low activation energies with the standard value of 0.7eV. **Table 2** summarizes the results.

The spread in the values of the equivalent test times depending on the failure mechanism is of course, not new as the physics did not change. However, it became much more relevant as the necessary test times to prove the required reliability became so long for the low accelerated failure mechanisms, that the high accelerated ones reach end of life well within this timeframe. Therefore, the fulfillment of a qualification test at a specified stress test condition and time might no longer be a meaningful design target. Instead, the mission profile becomes the central element of the design and validation process. Therefore, the number of customer-specific qualifications that exceed the standard qualification is growing, while the relevance of the standard is continuously decreasing. Depending upon when these customer-specific requirements are put in place, their acceptance can lead to the following: 1) prolonged time to market; 2) additional qualification efforts; and 3) the need for product changes. For the customer, this situation can lead to: 1) reduced product availability, 2) increased costs, and 3) risk for their development timeline.

A possible way out of the project management vs. reliability engineering dilemma described above could be the standardization of reference mission profiles. An extension of the standard with

reference mission profiles to cover most of these extended lifetime requirements is beneficial for both supplier and user because this approach brings certain advantages compared to standardized extended stress test conditions:

- Standardized mission profiles are usable in established processes (today used for customer specific mission profiles);
- Failure mode specific test evaluation becomes possible (so field application-relevant qualification failures can be discarded on a solid basis);
- It allows the use of a knowledge-based qualification using a structured generic data approach (i.e., an adapted robustness validation approach). This may improve time to market without any compromises on reliability.

In this section, the topic was analyzed only based on an exemplary temperature mission profile. For future automotive package designs, it will be essential to have tools and methods in place to deal with thermal cycling profiles, temperature-humidity profiles, and profiles based on other stressors, as well as those that depend on the application.

Connectivity

Despite all the changes brought about by the use of semiconductors, a car remains a car and comfort and safety remain the key expectations of the users. High-end cars will continue to be the early adopters of high-end comfort and safety features, but those features will trickle down to mid- and low-end cars over time.

Connectivity will enable a vehicle to access the internet and to communicate with smart devices, as well as other cars, and road-based infrastructures will provide "swarm intelligence" thereby enabling the collection of real-time data from multiple sources (**Figure 8**). Such connectivity will develop from bare connected infotainment, to a "car-to-anything" communication. There will be a strong link to autonomous driving

Figure 8: Requirements for connected cars [3].

that will require permanent accessibility of the car over the air for requests from the customer or the backbone.

Data security will play a major role in this area. Basic security considerations have to be implemented. A secure on-board communication, undisrupted car-2-cloud, car-2-infrastructure, and car-2-car communication has to be provided. A basic protection of the single ECUs is important, as well as a firewall and gateway, and a separate infotainment protection capability.

From a technology point of view, advanced packaging, such as SiP along with small node size chip technology, will be used. Components that have originally been designed for consumer electronics, will make their way into the automotive sector. One example is the embedded wafer-level ball grid array (eWLB) technology, which had originally been developed for wireless applications. It has found its way into the automotive sector for radar applications—with modifications to serve the increased reliability requirements.

With the requirement of permanent accessibility, connectivity will be one of the main drivers that increases the operational time of packages (and ECUs) for automotive applications. For example, software updates can be done during parking, using a wireless local area network.

Summary

The three megatrends: 1) autonomous driving, 2) eMobility, and 3) connectivity, are coming with specific requirements that will need adjustments for developing and enabling new, effective packaging solutions. Advanced packages for miniaturization, integration for fulfillment of safety requirements, and redundancy and advanced heat dissipation options are key for future automotive solutions. For these future automotive package designs, it will be essential to have tools and methods in place to deal with thermal cycling profiles, temperature-humidity profiles, and profiles based on other stressors, as well depending on the applications. Security and increased reliability will also play a major role in connectivity solutions for automotive applications. Packaging technology will finally make the difference.

Acknowledgement

Originally published at the International Device Packaging Conference, Fountain Hills, Arizona, March 2020.

References

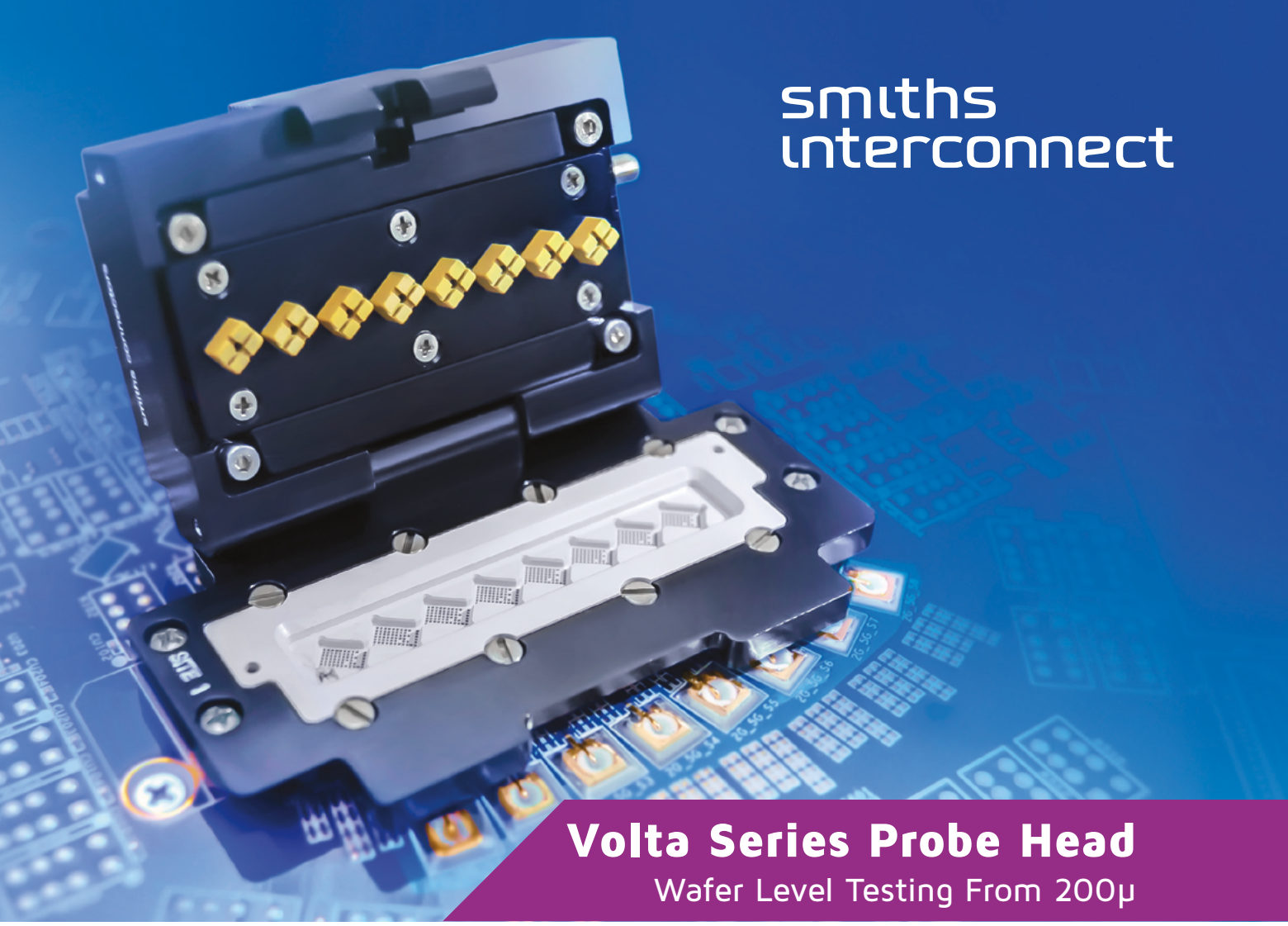
1. https://de.wikipedia.org/wiki/SAE_J3016
2. https://www.infineon.com/cms/en/applications/automotive/chassis-safety-and-adas/electric-power-steering-eps/
3. K. Konrad, "The impact of automated, driverless cars on semiconductors – an overview," IS-Test Workshop, May 2019.

Biographies

Thorsten Meyer is Lead Principal Engineer at Infineon Technologies AG, Regensburg, Germany. He holds a Diploma in Production Engineering. His focus lies in the development of new package concepts and he is the author of multiple publications and holds more than 150 patents and patent applications in the area of advanced packaging. Email Thorsten.Meyer2@infineon.com

Ulrich Abelein is Principal Engineer at Infineon Technologies in Neubiberg, Germany. He received degrees in Electrical Engineering from the Technical U. of Munich, 2003 and in economics from the U. of Hagen.

smiths interconnect



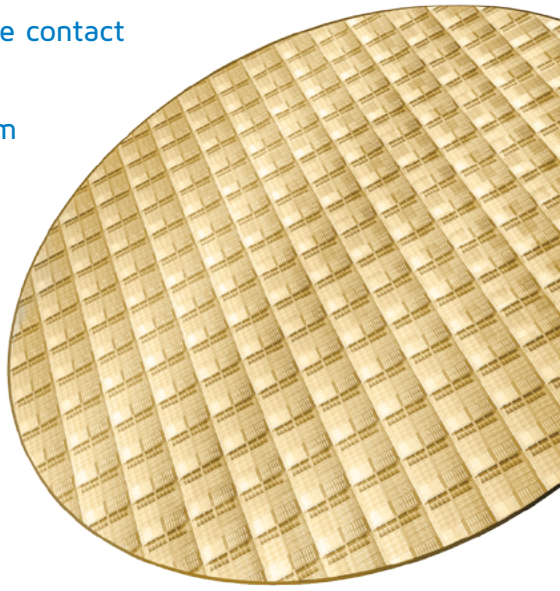
Volta Series Probe Head

Wafer Level Testing From 200µ

Increased Throughput, Reduced Test Time

Smiths Interconnect's Volta Series Probe Head provides improved efficiency in high reliability WLP, WLCSP and KGD testing.

- Extremely short signal path (≤3.80mm) enables low and stable contact resistance, high current carrying capacity and longer life cycle
- Enhanced planarity allows increased site to site test parallelism
- Reduced test set-up time, simple maintenance and field replacement ensure lower cost of ownership
- Allows sorted die test of all sites simultaneously resulting in increased test yields in a shorter time



Visit us at IWLPC Virtual Event October 13th-14th 2020

smithsinterconnect.com

Enabling artificial intelligence with heterogeneous integration

By Nelson Fan [ASM Pacific Technology Ltd.]

We are at the dawn of the artificial intelligence (AI) era! It will not be long before highly-intelligent cars enabled by AI will be cruising on our highways. Imagine that our future AI-enabled homes will automatically cool down to the optimal temperature just before we arrive home after a hard day's work, and dinner is prepared and ready to serve. At a large scale, cities built with embedded AI will be operating efficiently with tremendous amounts of devices and robots connected through 5G infrastructure with enough bandwidth for data management and transfer. High-performance computing (HPC) devices – one of the essential elements required for both end-terminals and edge computing – are needed for analyzing large amounts of data coming from massive numbers of (50B units by 2025) Internet of Everything (IoT) devices. It is critical that our industry address the design and assembly of HPC chips to enable the bright AI-enabled future that we desire.

Bottleneck in Moore's Law scaling

For many decades, Moore's Law has guided the semiconductor industry. It has been the norm to expect that the semiconductor node performance will double every 18 to 24 months—until now. Moore's Law is approaching its limit where the node size gets smaller than 14nm/10nm. Currently, advanced node development is still continuing. For example, TSMC has started the sampling of flagship application processor (AP) system on chip (SoC) devices with 5nm technology. Furthermore, 3nm node development has also been announced with risk builds to start by 2021. This continued node development is in question, however, as the commercial returns are not commensurate with the extremely high capital investment required. Today, the TSMC investment in the 3nm fab has exceeded US$20B, and by the time the fab is completed, the total investment will have reached a staggering US$50B. What is the way forward?

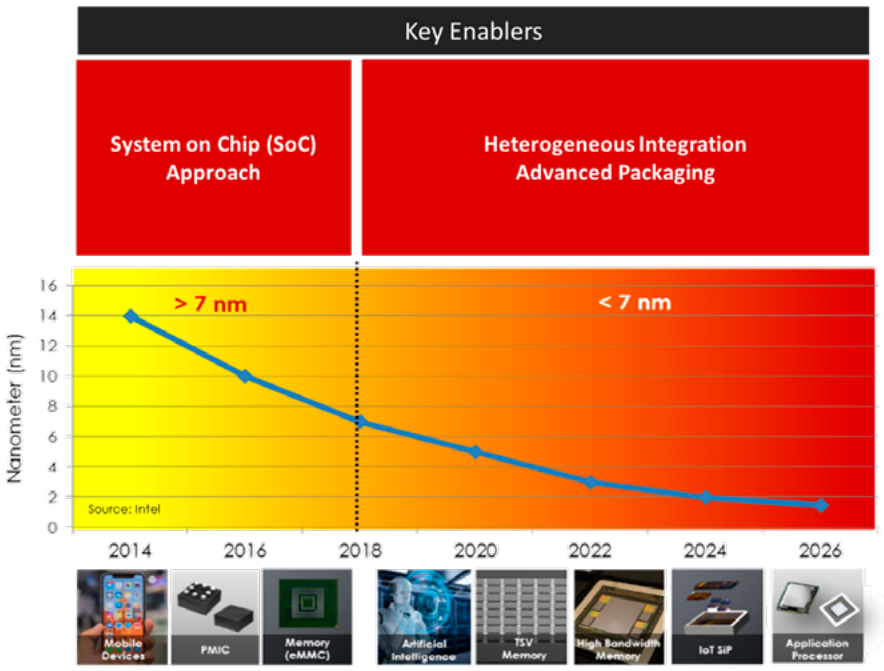


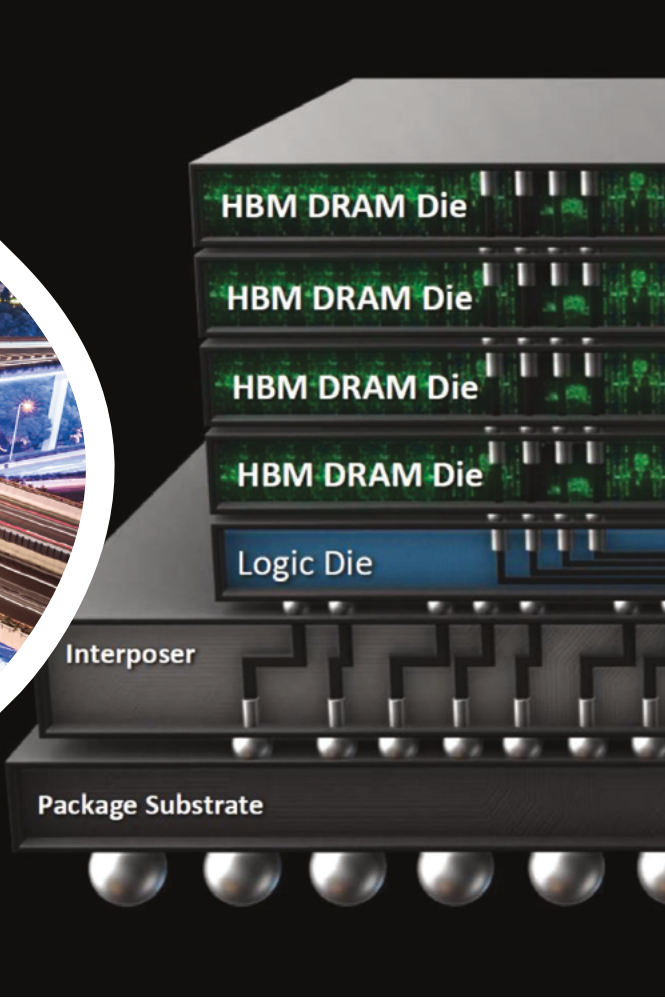
Figure 1: Packaging evolution vs. fabrication technology. SOURCE: Intel

Heterogeneous integration: a back-end scaling approach

To reduce the financial commitment for future node development, an effective approach is switching from "only front-end node scaling" to "combination with back-end scaling." Heterogeneous integration (HI) is a way forward. HI is a back-end approach by means of advanced packaging technologies that enable the integration of multiple chiplets with different functionalities and each fabricated with the best-fit node in terms of technology and economics, to reassemble an SoC-like function (**Figure 1**). Different approaches to achieving HI are described in the sections below (**Figure 2**). Some of these approaches are already in volume production.

Interposer. The use of an interposer is a 2.5D-IC package concept. An interposer is used to interconnect the chiplet on its top side, and the bottom side of the interposer is connected to the high-density build-up substrate (**Figure 3**). There are different types of interposer technologies. The one being implemented for volume production today is the through-silicon via (TSV) interposer by TSMC. It is a passive interposer that has been developed for close to ten years. TSMC has named this structure chip-on-wafer-on-substrate (CoWoS®).

The passive TSV interposer was first developed for field-programmable gate array (FPGA) devices to address the wafer yield issue. FPGA die were very large monolithic die. The TSV interposer was developed and designed with redistribution layer (RDL) fine linewidth and space routing, which provided a way to integrate smaller homogeneous partitioned FPGA chiplets. The resultant



When Moore's Law no longer gets your ICs where they need to go.

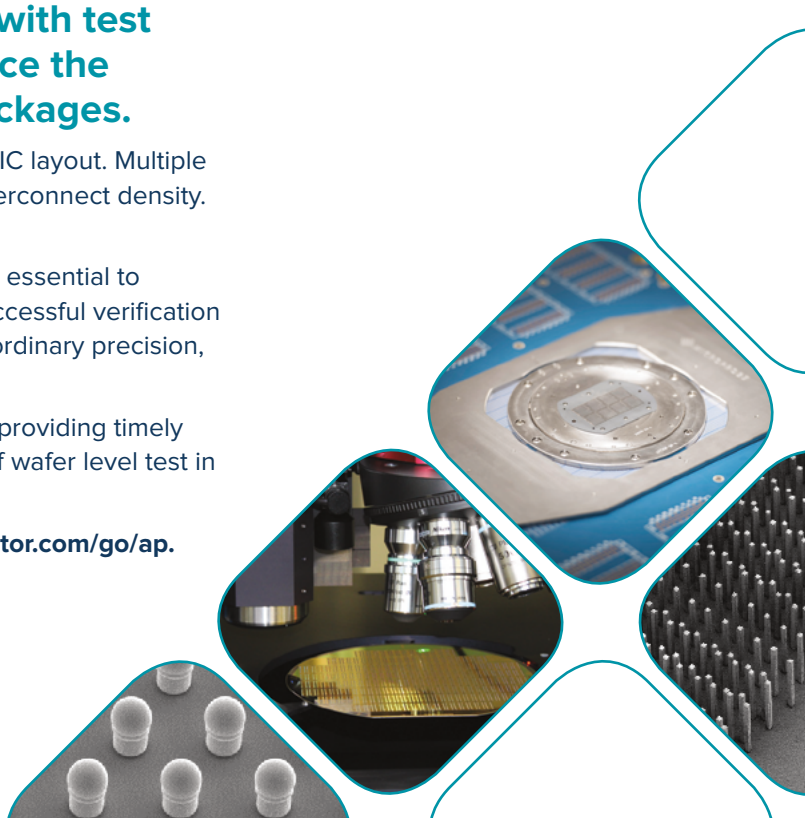
FormFactor takes on the challenge with test and measurement solutions to reduce the manufacturing cost of advanced packages.

Advanced packaging adds a new vertical dimension to IC layout. Multiple dies merge into single systems with unprecedented interconnect density. Performance goes up. Power consumption goes down.

And wafer-level test and measurement becomes nearly essential to guarantee cost-effective fabrication and packaging. Successful verification at this level requires probing and measuring with extraordinary precision, optically, electrically and mechanically.

FormFactor not only understands the problem, it's also providing timely solutions to advance yield knowledge at every phase of wafer level test in advanced IC packaging.

Let us help you rise to the occasion. **Visit www.formfactor.com/go/ap.**



FormFactor™
Test insight from lab to fab.

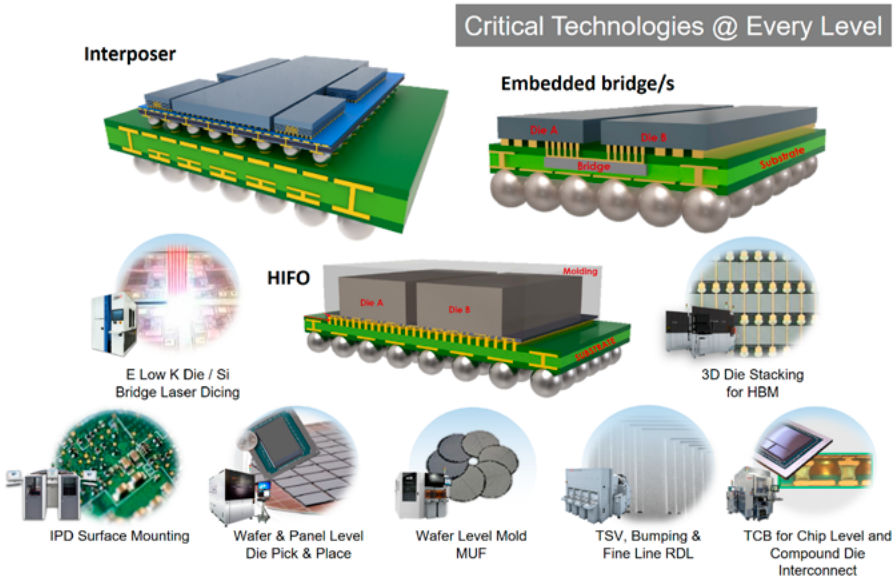


Figure 2: Packaging technologies in different HI levels.

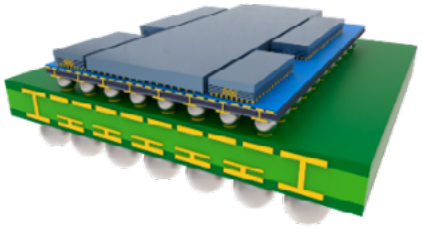


Figure 3: Typical structure of an interposer.

wafer yield with smaller die size had shown significant improvement, while the FPGA performance was unaffected with the CoWoS® structure.

One of the important design aspects of the CoWoS® TSV interposer is to maintain the die-to-die electrical communication performance. Both signal and power integrity are very important design aspects. Another important advantage of using a TSV interposer is the matching of its coefficient of thermal expansion (CTE) with the chiplet because they are both silicon. For over a decade, besides usage in FPGAs, both high-end graphics processing units (GPUs) and network processors made use of CoWoS® structures to be integrated with high-bandwidth memory (HBM) components. There are other types of interposer materials, such as glass and organic interposers, that are being developed. The former is enabled by through-glass via (TGV) technology, and the latter make use of thin-film technology. One of the common objectives of these new types of interposers is to reduce the manufacturing cost. Because the components to be integrated are increasing in quantity, the resulting interposer size is also getting larger. The challenge of using such large interposer integrated die is severe warpage that occurs when it is being flip-chip attached onto the substrate. Thermal compression bonding has been a proven way to overcome the integrated die warpage issue. The industry's interposer roadmap indicates that they are expected to increase in size to 100mm x 100mm, however not many large interposers can fit onto a 12" Si interposer wafer. Therefore, the scalability of the TSV interposer is limited.

Embedded bridge approach. The embedded bridge approach is considered a good approach to mitigate the TSV interposer scalability limitation (**Figure 4**). In 2014, Intel introduced its embedded multi-die interconnect bridge (EMIB) invention through a processor module named Lakefield, in which an EMIB bridge connects a Radeon Vega CPU with HBM. The size of the EMIB bridge is only around 6mm x 6mm! There are two important assembly processes involved in making an EMIB. First, in the panel format, single or multiple bridges are embedded into a semi-finished high-density build-up organic substrate with a very high-precision die attach tool. It is then followed by a high-density and fine line and space electrochemical metal deposition (ECD) process for the RDL formation. Moving forward it is foreseeable that when the geometry of chip I/O bump pitch features shrink further, the requirement on bridge placement accuracy will become more stringent.

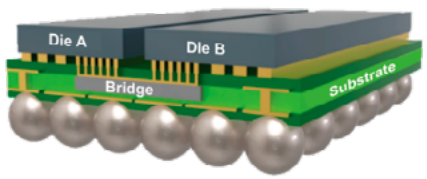


Figure 4: Typical structure of the embedded bridge approach.

Although EMIB is ideal for integrating multiple components without the scaling limitations found with CoWoS®, there is also concern in deploying this technology. The chip layout needs to align with the package architecture. This means that the I/Os of those connected chips need to couple well with the multiple interconnecting bridges. As a result, there are concerns with respect to how far the embedded bridge solution can be utilized in the open market.

Heterogeneous integrated fan-out (HIFO) approach. The HIFO approach is based on high-density wafer-level fan-out (HD WLFO) technology with both die-first and RDL-first processes, together with the structure to integrate multiple components with RDL fine line and space geometry (**Figure 5**). This approach has been widely deployed by outsourced semiconductor assembly and test suppliers (OSATS) like ASE, which has developed fan-out chip-on-substrate FOCoS technology. A core WLFO package size is typically around 5mm x 5mm, however a HIFO package size can achieve much larger body sizes. ASE – using its FOCoS structure – has demonstrated package sizes of more than 25mm x 25mm. The HIFO integrated die is not directly mounted on a board, but is mounted on

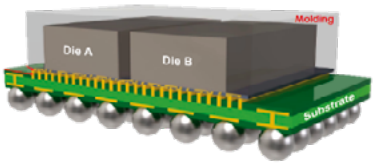


Figure 5: Typical structure of heterogeneous integrated fan-out (HIFO) technology.

top of an organic substrate. Therefore, the board-level stress can be mitigated by the substrate, which can be considered a stress buffer.

Emerging approach: 3D-IC integration. Besides package-level integration, the 3D-IC integration approach is being advocated in order to achieve the next level of integration with direct chip-to-chip interconnection without any external interposer or RDL routing (**Figure 6**). TSMC has announced its system on integrated chip (TSMC-SoIC™), and Intel has introduced its Foveros packaging technology. Both of these structures have the chiplet directly interconnected to a bottom active chip: TSMC's platform is through a copper-to-copper joint, while Intel's uses micro pillar solder joint interconnects.

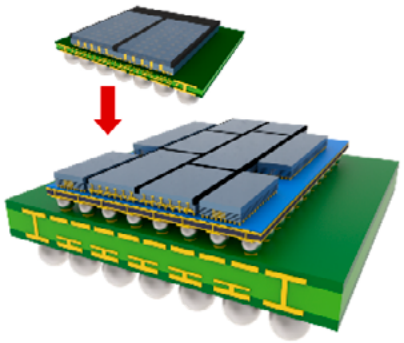


Figure 6: Typical structure of an emerging approach.

In high-end processor units such as GPUs and APs, almost 50% of the SoC area is occupied by static random access memory (SRAM). SRAM, however, might not benefit from advanced node technology. One way to improve its economics is to first extract the SRAM out from the SoC, and then have the SRAM portion fabricated with the best-fit node technology to achieve the optimal cost of ownership, and finally integrate it directly back on top of the logic portion, which is designed and fabricated with the more advanced node, to achieve the best performance. The resultant 3D-IC structure will continue to follow Moore's Law. Besides SRAM, there are many other chiplets with different functions at various nodes that can be mixed and matched by interconnecting them together like the SRAM example given above. Another way to further improve the economics is to reuse the intellectual property (IP) from not only in-house, but also from multiple third-party chiplets. However, adequate amounts of standards are required to define the logical and physical interfaces in order to enable this capability.

3D-ICs with copper-to-copper joint formation are based on a hybrid bonding process that pushes the envelope for the development of backend equipment capability and processing technology. It requires not only extremely high-precision placement accuracy from the micrometer to nanometer scale, but it also demands a very high level of cleanliness performance—basically equivalent to front-end equipment capability. Ultimately, the bond line thickness of the joint interface between the top and bottom dies is literally zero—it leaves no buffer for any minute particle. The 3D-IC integration can further evolve by combining the 3D-IC into a 2D or 2.5D packaging structure; hence, Moore's law can be further extended.

Impact of HI evolution on the supply chain

With the continued development of HI, we see significant impact on the supply chain for the industry. In the past, each player such as the wafer foundry, OSAT and substrate supplier, would remain in its own area of focus. Today, with HI, the boundaries among them have blurred. Foundries are stepping into the advanced package assembly business; substrate suppliers are also getting into the assembly process by offering passive and active embedding capability along with their substrate business. Although OSATS seem to be in a difficult position facing competition from both the foundries and substrate suppliers, they are also pushing the development of more cost-effective packaging technology to cope with the emerging but massive requirements of edge and IoT devices required by the AI world.

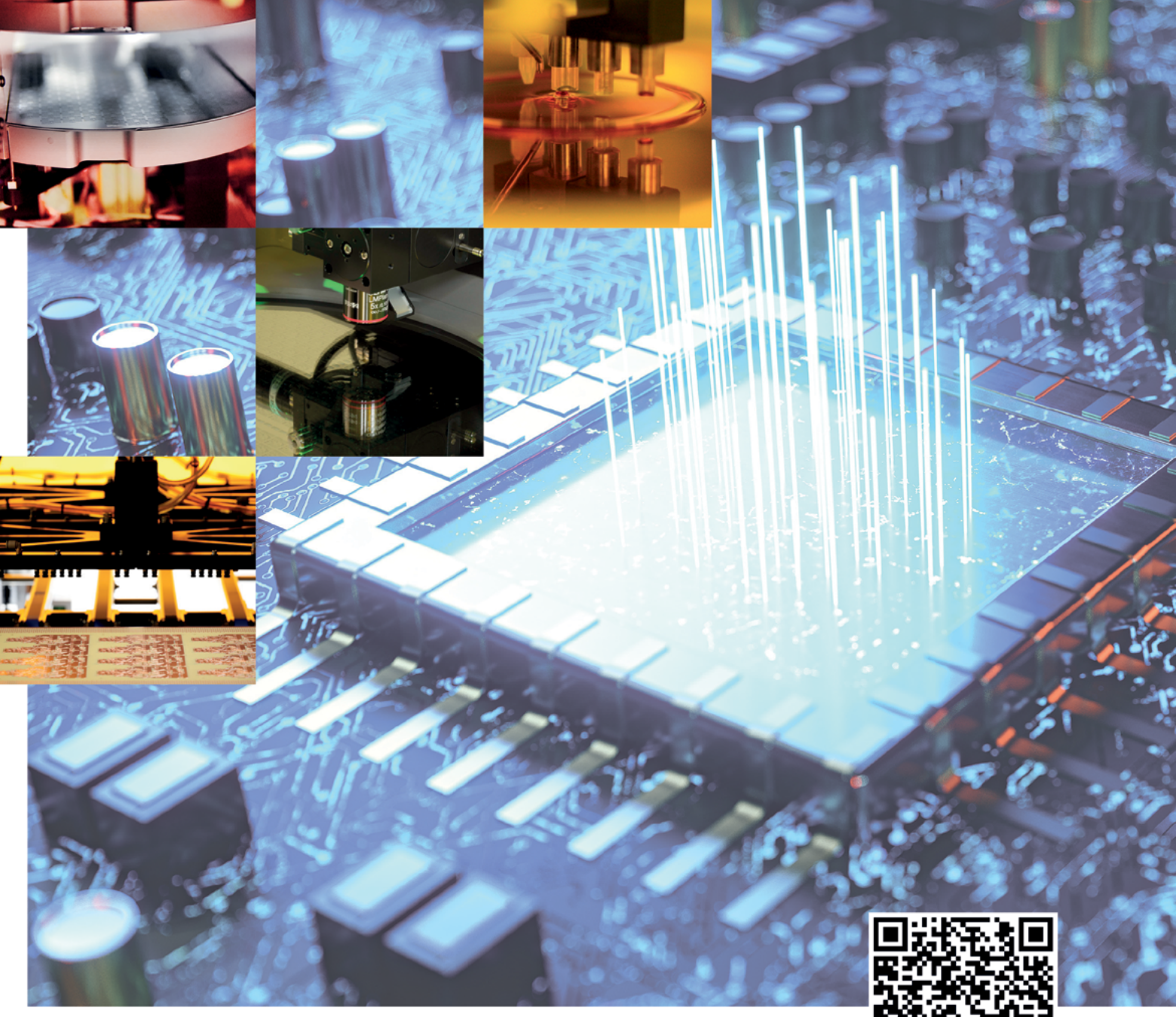
Summary

With the continuous evolution of heterogeneous integration, more business opportunities for each player are opened up because of the broadening of their serviceable available market. Without a doubt, there will be many new technical challenges with the associated processes, critical materials and enabling equipment. As a result, one will expect greater collaboration and co-development work among foundries, OSATS, and substrate suppliers, together with the semiconductor equipment manufacturers and materials suppliers.

Over the last decade, we have demonstrated our FIREBIRD TCB product and processing technology with a worldwide high-volume manufacturing (HVM) installation. We have also seen benefits from the ASM Fan-Out Technology Consortium—its panel-level fan-out bonder, NUCLEUS XL, is known to be the process of record (POR) for the high-precision panel-level packaging process. ASMPT offers a total interconnect solution experience as well as technical know-how established through the collaboration with many top-tier HI enablers to overcome their challenges, achieve lower cost of ownership and a shorter time to market capability.

Biography

Nelson Fan is VP of Business Development in Advanced Packaging Technology (APT), Semiconductor Solutions of ASM Pacific Technology Ltd., Hong Kong. With more than 30 years of experience in the semiconductor industry, he has held positions in different senior engineering roles in package and process development, as well as in manufacturing operations. He holds more than 40 US patents in semiconductor packaging technologies. Email: nelson.fan@asmpt.com



SUSS EQUIPMENT SOLUTIONS ENABLING LEADING-EDGE PACKAGING TECHNOLOGIES

SUSS MicroTec is a leading supplier of equipment and process solutions for the semiconductor industry enabling state-of-the-art packaging technologies from R&D to high volume production.

With next generation lithography solutions for FOWLP and leading edge wafer bonding technologies for 2.5D and 3D stacking, we contribute to the advancement of innovations in the key areas of packaging and system integration.

Always be one step ahead — with solutions from SUSS MicroTec.
Contact us for more information!

SUSS MicroTec
info@suss.com
www.suss.com

SÜSS MicroTec

WinWay Technology

High Heat Density Thermal Lid

- Air-Cooled & Water-Cooled solution
- High performance cooling module
- Customized lid design
- Device power up to 1200W

Active Thermal Controller

- 800W Heat dissipation solution
- Active thermal management for DUT
- Mass production for Handler
- Stand-alone operation

sales@winwayglobal.com

www.winwayglobal.com

FOPLP as a solution for heterogeneous integration

By Michael Hsu [Powertech Technology Inc.]

It's been several decades since the invention of integrated circuits (ICs)—but the proliferation of them has reshaped our way of life. Though Moore's Law, which observed this proliferation, is now facing difficulty moving forward, we did not come this far to see it end prematurely, so the industry has been working and collaborating for its continuation.

There are two trends in the industry that have been seeking to extend Moore's Law: 1) "More than Moore" led by foundries, and 2) "More than Moore" spearheaded by outsourced semiconductor assembly and test suppliers (OSATS). The former focuses on continuing the scaling, and the latter focuses on heterogeneous integration.

Heterogeneous integration involves a kind of "breaking the system on chip (SoC)" apart into discrete chips with diverse functions, and then reintegrating them back together with advanced packaging technologies. By doing so, we can maximize the economy aspect of chip fabrication, saving resources so that IC designers can divert more focus on the kinds of intellectual property (IP) that need advancement. To further this endeavor, there are many different technologies available—each with its own advantages.

In this paper we will present a thorough investigation on several advanced packaging technologies that are based on ball grid array (BGA) system in package (SiP), 2.5D/3D ICs, and fan-out packaging. The investigation will include the physical, electrical, and economic aspects of each technology.

Introduction

Quite sometime ago, our electronic systems consisted of IP blocks that were discretely packaged, and heterogeneously assembled on the board. The result of this was bulky electronic products that were not easily portable. Then, with the scaling of transistors, we were able to put multiple IP blocks into one chip, namely the SoC—and our electronics started to become portable, more powerful, and cheaper.

The scaling of transistors, however, is now facing physical challenges, and the resources and time needed to develop a newer iteration of SoC are becoming prohibitive [1,2] (Figure 1) because large demand quantities are needed to amortize the cost quickly. The

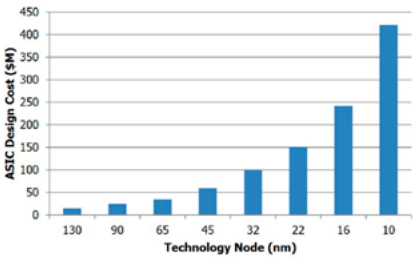


Figure 1: Design cost over technology node [1].

industry now focuses on splitting the SoC IP blocks apart, and reintegrating them back into a common packaging platform, where each IP can use the best choice of optimized wafer node (Figure 2). These now discrete

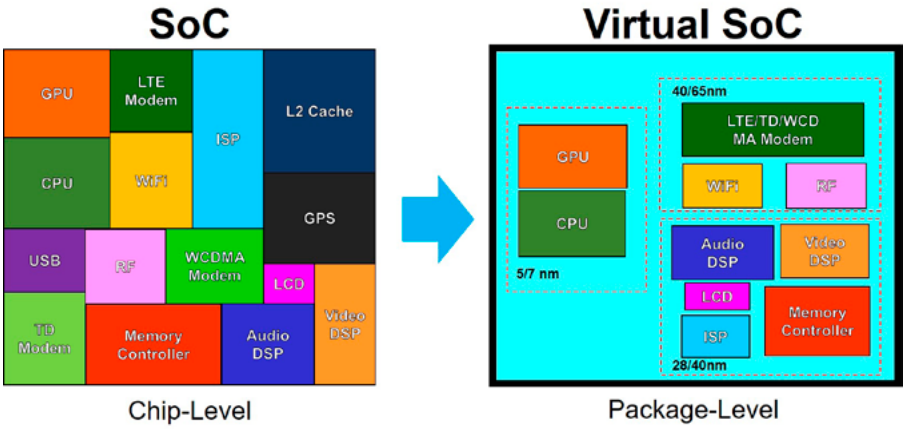


Figure 2: Splitting the SoC and reintegration.

IP blocks can be co-packaged together to achieve a similar performance as a SoC. The packaging solutions include (Figure 3) ball grid array (BGA) system in package (SiP), 2.5D through-silicon interposer (TSI), and 3D through-silicon via (TSV), fan-out wafer-level packaging (FOWLP), and fan-out panel-level packaging (FOPLP), as discussed in the sections below.

BGA SiP. Organic substrate stands as one of the most matured platforms for heterogeneous integration. The chips are either wire bonded or flip-chip bonded to the substrate. The wire bonded chips tend to suffer from parasitic capacitance or inductance at higher frequencies, and such

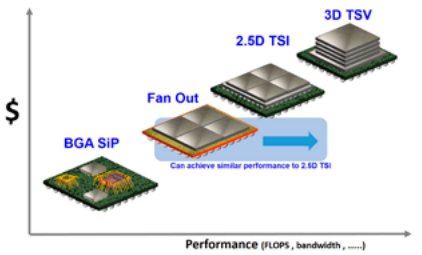


Figure 3: Potential packaging solutions.

chips also have rather limited bandwidth and input/output (I/O) density. Flip-chip technology can minimize parasitics while reaching higher bandwidth and I/O density, but the traces on the substrate would generate high transmission resistance, thereby resulting in higher power consumption. Heterogeneous integration would also require more routing layers for the substrate that will further increase the cost and decrease the substrate manufacture yield.

3D/2.5D IC. 3D/2.5D IC technologies use silicon as a platform, but the platform itself may also be an active chip. 3D IC provides the shortest electrical path for the adjacent dies using bumps and TSVs,

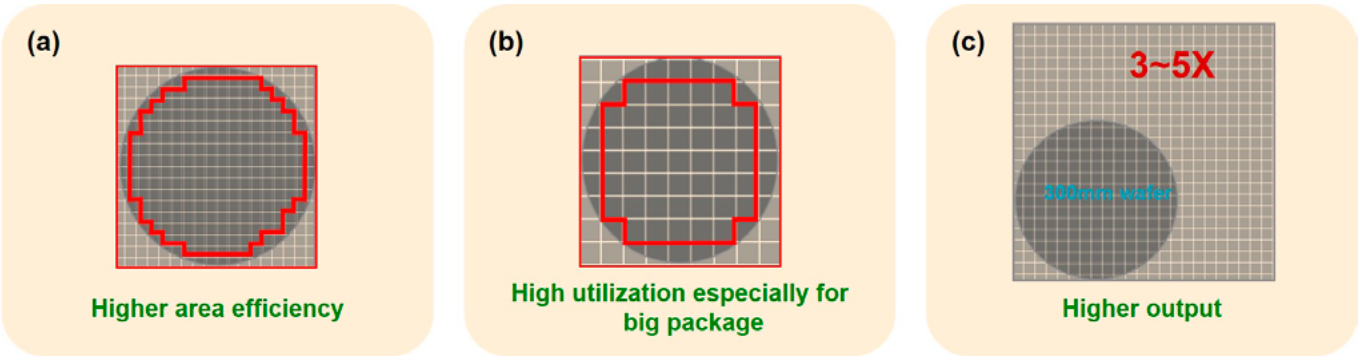


Figure 4: The geometrical benefit of panel-level packaging.

providing excellent electrical performance. The drawback is increased difficulties in chip design. It is especially difficult to integrate chips with different functions without adding cost on the wafer frontend process.

On the other hand, 2.5D IC uses a passive TSI with front end of line (FEOL) traces to interconnect chips—also providing excellent electrical performance. The interposer cost, however, limits its application to premium markets such as graphics processing unit (GPU) cards for high-end gaming and high-performance computing (HPC).

Fan-out wafer-level package (FOWLP). Fan-out packaging is characterized by utilizing the redistribution layers (RDLs) as a substrate, with I/Os distributing outward from the boundary of the chip area. Unlike organic substrates, RDLs can be much thinner while providing finer traces to interconnect the chips. This packaging solution naturally provides the smallest form factor while maintaining a decent electrical performance and power consumption. FOWLP isn't a new concept, but it gained fame rather recently with TSMC's debut of Integrated Fan Out (InFO) for iPhone 7's application processor engine (APE) [3]. Since then, the industry has started an "arms" race in an attempt to gain market share in this emerging field.

Fan-out panel-level package (FOPLP). FOPLP sought to overcome the limits of geometry from wafers, where the geometry caused wafers to fall short in terms of throughput, while the large rectangular panel enjoys the benefit of fully utilizing every area it carries. The panel can yield 3~5 times higher throughput, depending on the package size (Figure 4).

Another unique advantage of FOPLP is that the panel can be easily divided into smaller sub-panels, allowing it to be handled like a standard strip. This enables flexibility in the manufacturing process. Table 1 shows a brief summary comparing the potential packaging solutions for heterogeneous integration. It is believed that FOPLP provides the best balance among each of the metrics.

Challenges of FOPLP

The following sections discuss major challenges with the implementation of FOPLP technology.

Lack of Infrastructure. Unlike wafers, there are many different sizes of panels (Figure 5), so the equipment for FOPLP inevitably required a great deal of customization. Although SEMI Standards (e.g., SEMI 3D20, more details can be found at https://store-us.semi.org/products/3d02000-semi-3d20-en-specification-for-panel-characteristics-for-panel-level-packaging-plp-applications) have recently narrowed down the size to 510x515mm and 600x600mm, the industry still has to agree on one universal size. The processing capability needed for FOPLP can be covered by a wide range of equipment that were originally meant for other purposes, for example, wafer foundries, liquid crystal displays (LCDs), and the printed circuit board (PCB) substrate industries. Therefore, careful selection of the equipment must be exercised to build a production line that makes the most sense to outsourced semiconductor assembly and test suppliers.

There are some situations where the needed equipment has no adaptable platform existing on the open market, which will require the OSAT supplier to entirely design the new ones based on its own requirement. One notable example would be our own automated guided vehicle (AGV) and overhead transport (OHT) used for transporting panels from stage to stage. We also had our equipment vendor customized an equipment front end module (EFEM) for us.

	BGA SiP	3DIC	2.5DIC	Fan Out
I/O Density	Mid	Higher	Higher	High
Trace	Coarse	FEOL	FEOL	Fine Line
Chip Partition	Possible	Difficult	Possible	Possible
Bandwidth	Low	High	High	High
Power Consumption	High	Mid	Mid	Low
Package Thickness	Mid	Thick	Mid	Thin
Overall Cost	Mid	High	High	Justifiable

Table 1: Comparison of packaging solutions.

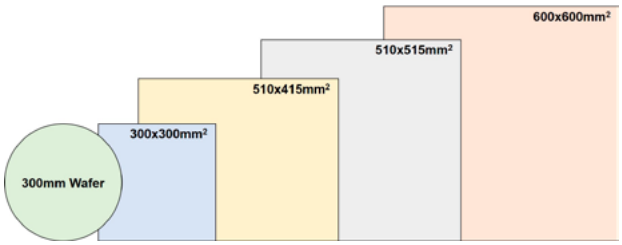


Figure 5: Different panel sizes.

Panel warpage control. Without proper mitigation measure, warpage can be a critical issue and it becomes more and more severe with the increasing number of RDLs and larger package sizes. Figure 6 shows a typical chip-first face-up fan-out process, where several thermally-induced areas of warpage will occur, such as post-mold curing, mold grinding, polyimide (PI) curing during patterning, and reflow during ball mounting. Existing methods to overcome the panel-level warpage includes physical suppression using a conveyer with a guide roller and vacuum chuck to keep the panel flat during the process steps (Figure 7).

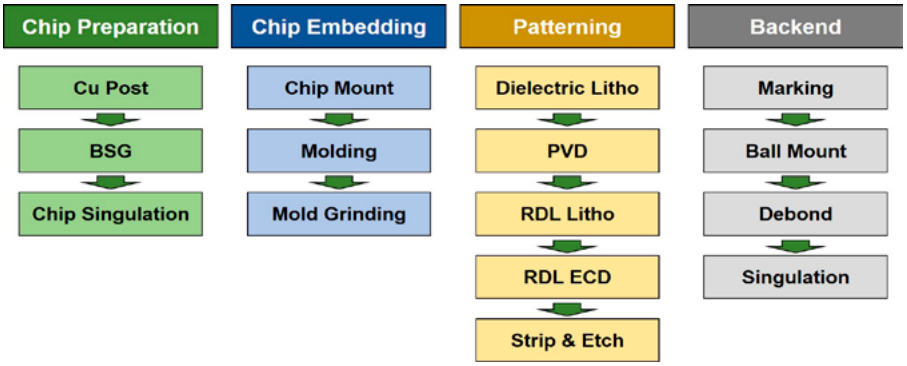


Figure 6: Chip-first face-up fan-out process flow.

Chip shift and alignment. The lack of precision during pick and place can cause some challenges for the subsequent lithography and alignment steps. The situation can be further worsened by thermally-induced chip shifting, whereby the chip shifts away from its original bonded position prior to patterning. There are several solutions to address this issue; for example, one can select an adhesive to prevent the shifting, but such adhesives may be hard to remove afterwards. Another method would be optimizing the process conditions and opting for better material selection to prevent the coefficient of thermal expansion (CTE) mismatch issue.

Some of the more indirect approaches include chip placement with offsets, and real-time adaptive mask alignment of the lithography tool. The former uses simulation to predict the potential shifting vector and feeds back the value to the pick and place machine. The latter approach simply enables alignment to the local fiducials of the die for adaptive patterning.

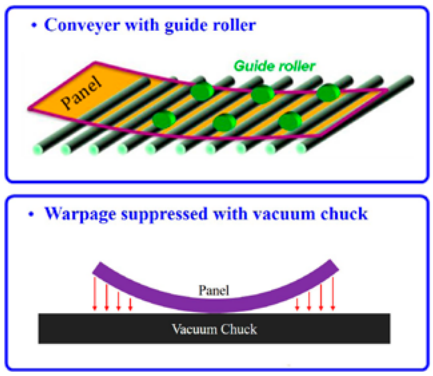


Figure 7: Physical warpage suppression methods.

Applications and architectures

Based on the type of architectures, FOPLP can be categorized into four major types: chip-first face-up, chip-first face-down, chip last, and chip middle. And there are many derivatives based on these different types.

Chip first. Most of the fan-out packaging we see today uses the chip-first architecture. Chip first represents the lowest cost solution in fan-out technology, where the chips are bonded to the carrier, either facing up or down, prior to the RDL formation. Chip-first face-down uses the least number of process steps, whereas chip-first face-up will require an additional post-mold grinding step to reveal the embedded chips. Generally, the chip-first face-down approach is used for smaller chips with RDL L/S ≥10/10µm, whereas chip-first face-up approaches are used for larger chips. However, it does have limitation in I/O density, chip size, and RDL L/S, and concerns in misalignment caused by the aforementioned chip shifting issue.

Applications like accelerated processing unit (APU), central processing unit (CPU), GPU, baseband processor, application specific integrated circuit (ASIC), power management IC (PMIC), radio frequency (RF), and analog, are ideal for the chip-first structure. It also has the largest share in the fan-out market because of the early development and matured process yield. PTI has two available technologies to address this market: CHIEFS® (Chip Integration Embedded Fan out Solution) and BF^{2}O® (Bump Free Fan Out). The former technology is a chip first face up with Cu post on chip, and the latter is chip first face down without chip bumping.

Chip last. As the name suggests, the chip-last approach is to build the RDL before embedding the chips—therefore the chips must have bumping. The RDLs can be tested before chip bonding, eliminating the loss of known good die (KGD). Additionally, the metallic bonding between the chip and RDL also prevents the dreaded chip shifting issue. Building the RDL on the flat carrier also allows finer RDL L/S and higher I/O density as low as the sub-micron ranges. The chip-last process also makes it much easier to attach passive devices of varying sizes. We have a standing solution called CLIP® (Chip Last Integration Package).

For heterogeneous integration, the chip-last approach is an ideal platform whereby many different chips and passives can be easily integrated in a common RDL substrate, and is one in which the fine-line RDL allows them to be put in close proximity. It is also a good alternative to the 2.5D IC, which requires a very expensive Si interposer with TSV. For example, for GPU and high-bandwidth memory (HBM) integration, the fine-line RDL allows high-density interconnection between the chips and can achieve decent performance at a much lowered cost. The chip-last process, however, is a more complex process, therefore it has a higher cost than the chip-first approach. The application of chip last, as a result, mostly targets high-end markets that emphasize performance, such as high-end processors, ASICs, and field-programmable gate arrays (FPGAs) for networking, artificial intelligence (AI), and HPC to replace 2.5D ICs.

Chip middle. The need for double-sided RDL fan-out solutions gave birth to the chip-middle process, where the chip is embedded and sandwiched by top and bottom RDL layers. The method of connecting the top and bottom RDLs varies, such as using a tall Cu pillar and via frames [4].

The chip-middle process allows heterogeneous integration in the vertical axis, where the chips can be embedded with passives or other chips bonded to the topside RDL. One of the most notable applications would be the APU for the mobile market, for which we offer the PiFO® (Pillars in Fan Out), where the tall Cu pillar can provide a much better electrical performance and reliability over the solder-based through-mold via (TMV).

Chip-embedded FO. A natural extension of chip-last and chip-middle technologies would be chip-embedded fan-out. A silicon chip embedded between the top and bottom RDLs can serve as a localized high-density interconnect for the chips it connects on top. This provides even better electrical performance than the chip-last solution, and does not require the expensive large Si interposer. The embedded chips can be much smaller than the interposer and be placed only where they are needed. The tall Cu pillars can also serve as a power conduit for the top chips—this is a significant advantage over the TSVs in the Si interposer, where the

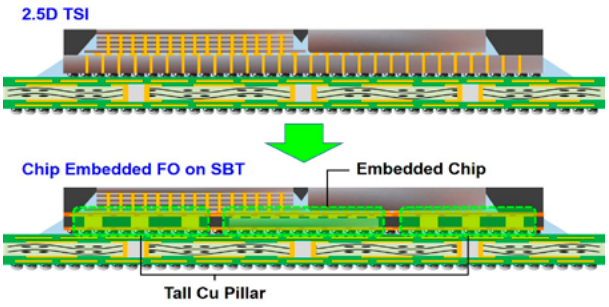


Figure 8: Chip embedded fan-out packaging.

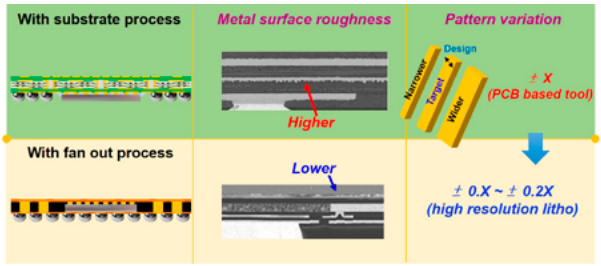


Figure 9: Advantage of fan-out AiP.

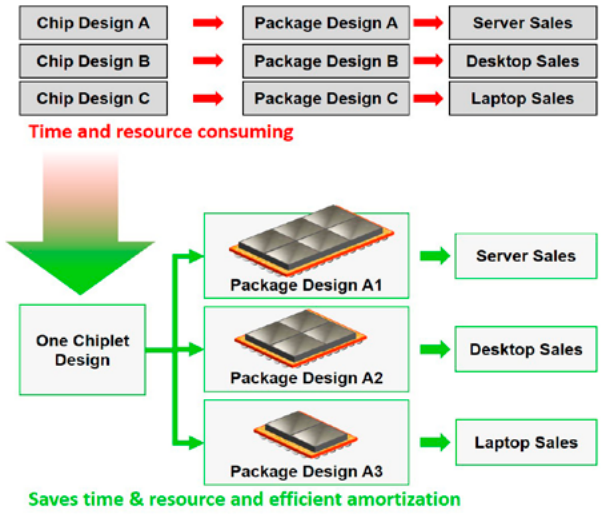


Figure 10: Chiplet integration scenario.

TSV with a small diameter will suffer from high resistance when pumping power from substrate to the chip (Figure 8).

FO AiP. Last year we witnessed the first commercialization of antenna in package (AiP) with Qualcomm's QTM052. Though technically more of a module than a package, this was still an important milestone for the adoption of mmWave technology. Currently, the industry has focused its efforts on the flip-chip chip-scale package (FCCSP) style of AiP, where patch antennas are on the top surface of the substrate with dipole antennas (Yagi-Uda antennas) around the peripherals. However, the lossy nature of the organic substrate and large form factor can be problematic for the 5G mobile application.

Fan-out packaging can build a patch antenna and Yagi-Uda antenna with RDL and the result is a structure with a very thin profile. The fine-line interconnection also allows a much lower transmission loss, thereby exhibiting excellent RF performance. The process control window for the fan-out process is also much tighter than that of the substrate, thereby allowing better design flexibility when considering impedance matching, insertion loss, and return loss (Figure 9).

Chiplet integration. For companies with large and diverse product lineups, chiplet integration makes perfect sense. AMD has demonstrated an example in which the CPU chiplet can easily scale from 2 chiplets to 8 chiplets, depending on the intended applications [5,6]. Not only the chips, but the package itself must be easily scalable. Fan-out technology allows highly flexible package design, as long as the resulting package size falls within the reticle size of the lithography equipment. Customers can simply pick up the chiplets they want for their system, and OSATS can use the fan-out technology to design the RDL to interconnect these chiplets, potentially saving new tooling costs, and several months' worth of lead time for a new substrate design (Figure 11).

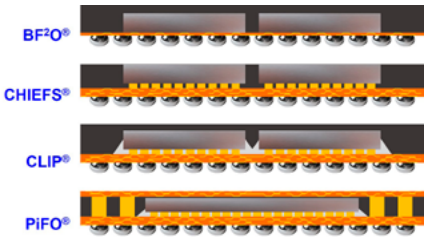


Figure 11: PTI's FOPLP lineup.

Summary

The advantages and the challenges of panel-level fan-out packaging have been thoroughly investigated in this paper. As future semiconductor growth is driven by many different technologies and applications, FOPLP stands as an ideal platform to meet the needs of heterogeneous integration and very large throughput at the same time (Figure 11). Combining chiplet integration with FOPLP, OSATS can actively help customers to save nonrecurring engineering (NRE) costs and expedite the new product design cycle time. The industry needs to work together to completely build up the ecosystem of both FOPLP and chiplets, so the supply chain and the end customers can enjoy the fruits brought by these technologies.

References

1. D. S. Green, "Common Heterogeneous Integration and Intellectual Property (IP) Reuse Strategies (CHIPS)," CHIPS Proposers Day, 2016.
2. S. Sutardja, "The Future of IC Design Innovation," ISSCC, 2015.
3. C. F. Tseng et al., "InFO (wafer-level integrated fan-out) technology," ECTC, 2016.
4. Y. Kim, et al., "Samsung fan-out panel-level package solutions," EDTM, 2019.
5. S. Naffziger, et al., "AMD chiplet architecture for high-performance server and desktop products," ISSCC, 2020.
6. D. Suggs, et al., "The AMD "Zen 2" Processor," IEEE Micro, 2020.

Biographies

Michael Hsu is the Senior Director of Advanced Package Product and Design at Powertech Technology Inc (PTI), Taiwan. He has over 25 years of experience in R&D and process in the semiconductor assembly industry. Before joining PTI , he worked for Amkor Technology and Everlight Electronics. He holds a Master's degree from National Tsing Hua U., Hsinchu Taiwan. Email michaelhsu@pti.com.tw

A deep-learning solution for heterogeneous package inspection

By Shahab Chitchian [INTEKPLUS Corporation Ltd.]

Heterogeneous integration through the use of system in package (SiP) technology has been effectively adopted by the mobile industry. Recently, chiplet packaging has become a key technology to continue Moore's law by improving yield and reducing total package/product cost.

In this article, the latest heterogeneous inspection results covering our deep-learning inspection process are reviewed. First, we introduce our deep-learning algorithms for two steps of segmentation and classification, followed by their training methods. In the second part of the article, the integration of deep learning and machine vision is presented. Furthermore, we show that by applying segmentation and classification in series, different defect modes can be distinguished and within each mode, different classes of true defects and overkills can be differentiated. Lastly, some case studies including micro-crack and bump damage detection capabilities are presented. By mitigation of under- and over-rejection for critical defect modes, our AI solution results in yield improvement for our semiconductor customers, which means significant cost reduction for such large form factor and expensive multi-chip packages.

The deep-learning process

To understand the AI inspection process and results in detail, our AI process is briefly summarized below in five sections. The sections are: 1) deep-learning algorithms; 2) model training and validation; 3) deep-learning combined algorithms and results; 4) second-inspection process and results; and 5) deep-learning deployed on edge computing vs. cloud computing.

Deep-learning algorithms. Our deep-learning approach consists of segmentation and classification. On the one hand, segmentation distinguishes target features (objects) from background. It is considered a pixel-level classification algorithm to calculate the probability value of a target feature for all pixels in the image, then to classify it as the target object if the probability is greater than the threshold value set by the user. A single model can be trained to segment several objects, but the performance is usually poor. So, our segmentation model classifies pixels into binary classes of defects and background. On the other hand, the classification determines different features (objects) in the image. It is used to classify features with the highest probability related to each class. Therefore, it has higher performance for multiple classes in one model compared to segmentation. Classification is applied when it is necessary to distinguish among different features, e.g., in the case of bump damage to determine a defective bump from other non-defective (overkill) bumps.

Figure 1a shows our segmentation steps and example defects. The

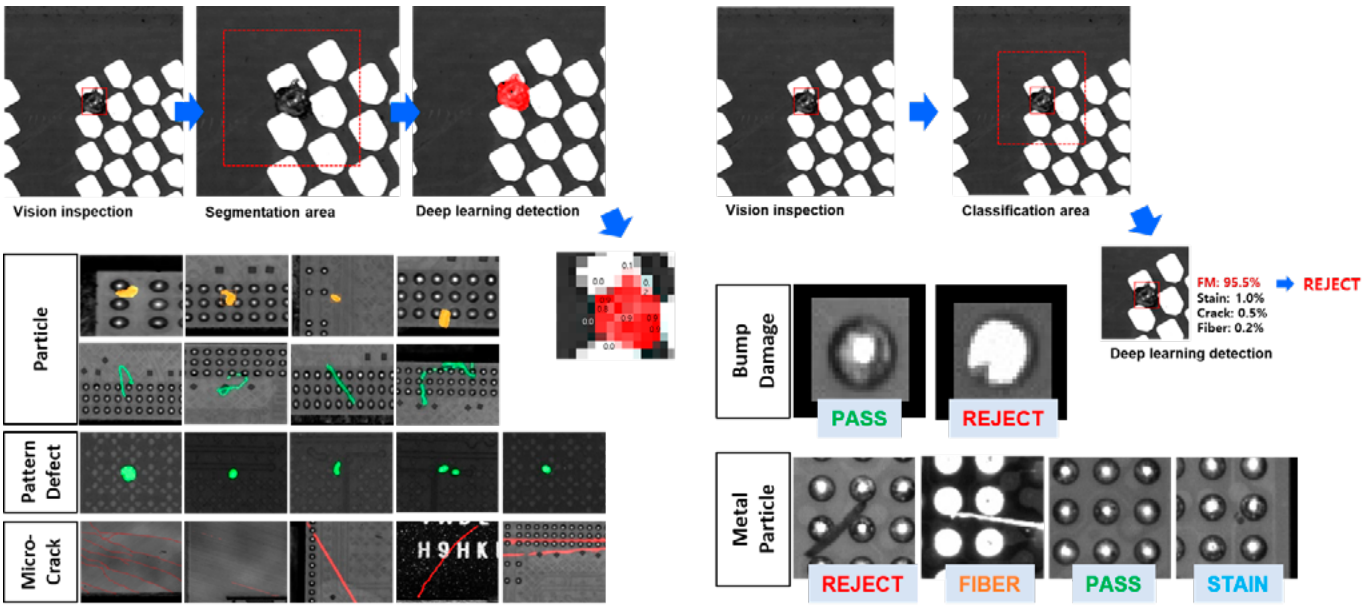


Figure 1: a) (left) The diagrams show segmentation inspection progress and results. Defect size can be measured accurately to determine a "Good" or "NG" defect based on the specification; b) (right) The diagrams show classification inspection progress and results. Different classes for each defect and multiple classes within each defect can be classified as applicable.

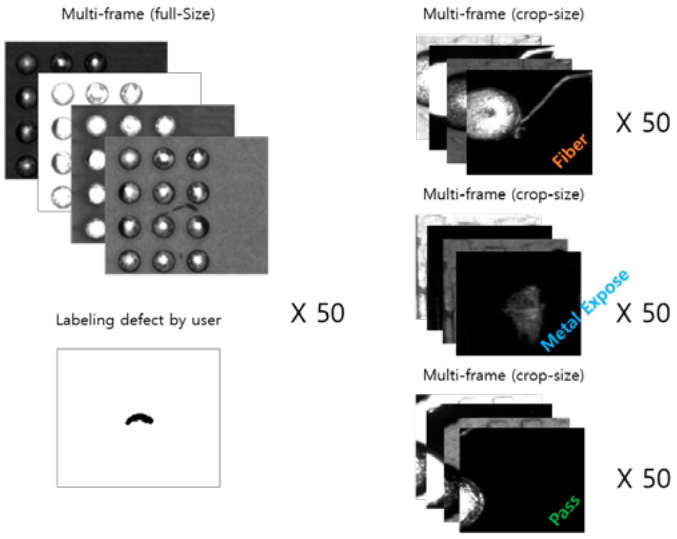


Figure 2: The deep-learning model training process: a) (left) segmentation; b) (right) classification.

algorithm checks every pixel in the image and calculates the defect probability value. If the probability value is higher than a threshold, it is marked as an NG (defect) pixel. In **Figure 1a** (top), pixels with probability values greater than 0.8 are segmented as being defective. Micro-cracks can be accurately detected both on the chip and on the mold surface. By using our AI solution, we can correctly differentiate micro-cracks from other overkill modes like grinding marks on the chip (see **Figure 1a**, bottom). In contrast, the classification algorithm classifies defects into corresponding defined classes. For each detected defect, the probability values of all classes are calculated. The defect is then assigned to the class with the highest probability value. **Figure 1b** shows classification steps and a bump area defect case study. In **Figure 1b** (top), the FM (particle) class has the highest probability value, so the defect is classified as FM mode. **Figure 1b** (bottom) depicts a bump damage reject and a metal particle defect, which are classified as true rejects compared to other acceptable modes, e.g., fiber, stain.

Model training and validation. Deep-learning model training includes four steps: loading, annotation, learning, and validation. For segmentation training, we load images to annotate defects on each image to achieve pixel-level ground truth. For classification training, pixel-level ground truth is not required—instead, we need to crop each defect and label it per the image. Model learning is performed in the third step, followed by testing in the last step. Segmentation testing is done by comparing annotated areas with segmentation results. Classification testing is carried out by comparing marked classes and test results classes. We enhance inspection performance of our AI solution by using multi-frame image capturing. Six frames with different lighting conditions are captured for each defect, then a minimum of three frames, including poor defect visibility, are selected for deep-learning model training. Inspection speed is reduced by increasing the number of frames. As shown in **Figure 2a**, for the segmentation model, a minimum of 50 multi-frame full-size images (different units of same product) and for the classification model, a minimum of 50 multi-frame crop-size images (per defect type) are required (see **Figure 2b**).

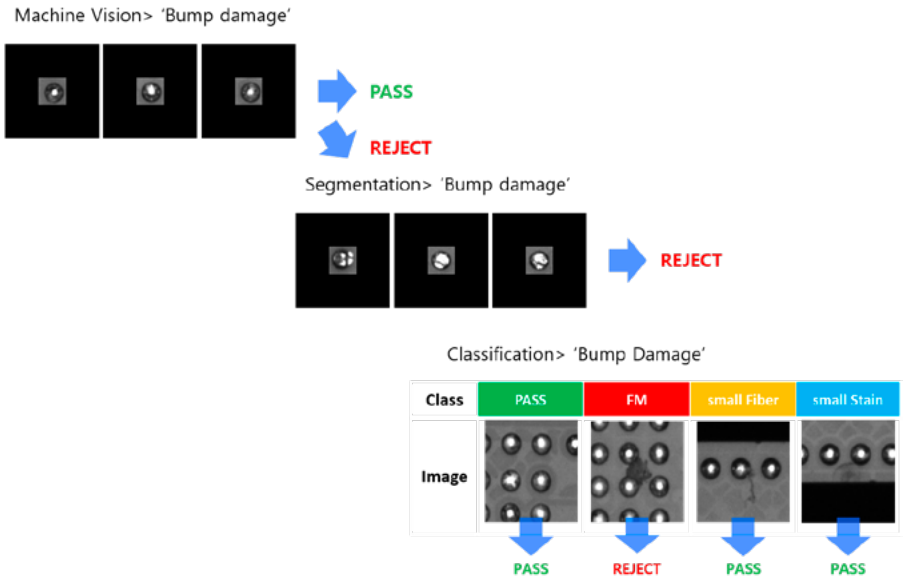


Figure 3: Deep-learning inspection progress using segmentation and classification combined algorithms.

Under-reject and over-reject validations are important steps before applying the AI model in a real inspection scenario like high-volume manufacturing inspection. After image capturing using vision software and collecting the images for model learning, the training process is completed. Next, we perform inspection by using the AI model. For under-reject validation, all logged images by the review software are reviewed for defects that were not detected. Model learning is reinforced after labeling the defective exported images. We then replace the existing model in the learning software. Finally, detection capability is again verified by re-inspecting the under-reject units. For over-reject validation, we review all images logged by the review software for overkill images. We then reinforce model learning after adding the exported images without labeling. The AI model is replaced and detection capability is verified again. In the case of over-reject validation, removing similarly-labeled overkill images before reinforcing the model learning effectively reduces overkills.

Combined algorithms and results

Our deep-learning approach includes combined segmentation and classification algorithms in series. Based on customer criteria and the trained model, segmentation determines the defect area. Based on different defect features (classes) and the trained model, the classification differentiates true defect modes and determines pass (overkill) or reject. For an example such as FM (particle), other defect modes of stain, fiber, and crack can easily cause overkill and underkill situations.

Combined segmentation and classification algorithms in series first checks all rejects (by machine vision) one more time based on customer criteria (segmentation trained model) and finally decides on whether the classification should be a true reject or pass (overkill) based on defect classes (classification trained model) and the customer specification. The AI inspection progress using combined algorithms is shown in **Figure 3**. As an example, for the bump damage case study, machine vision determines the pass units. Reject bumps are inspected by the segmentation algorithm, followed

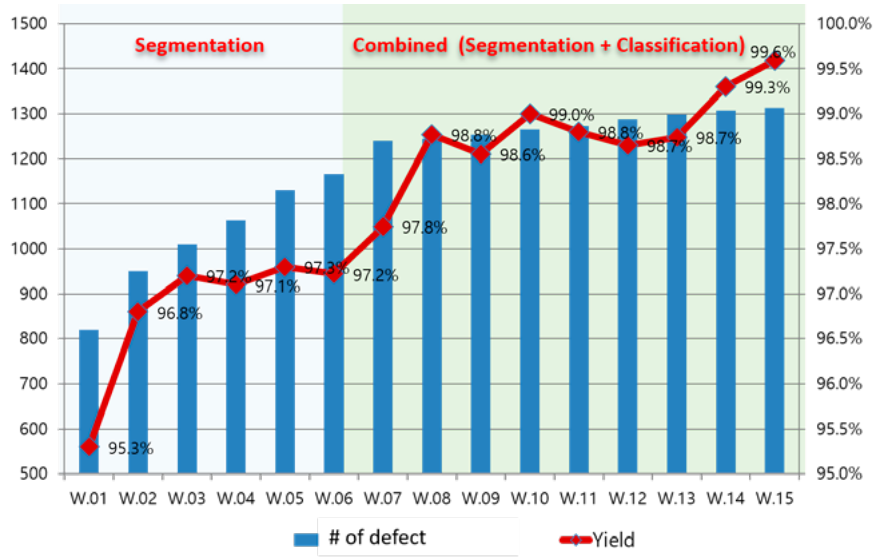


Figure 4: Deep-learning and combined algorithms progress over an evaluation period of fifteen weeks.

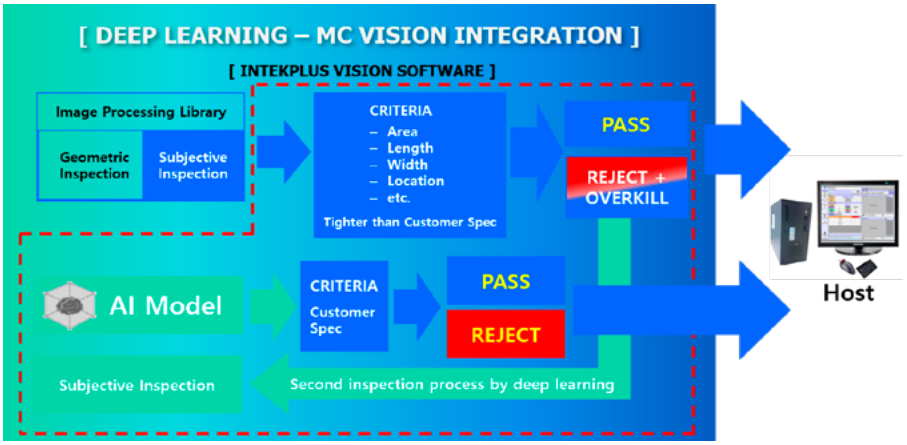


Figure 5: The deep-learning second-inspection process.

by bump damage classification modes to differentiate true reject and pass units (see **Figure 3**, right).

Figure 4 depicts the progress of deep learning and combined algorithms over fifteen weeks of evaluation time. The use of AI gradually improves the quality by learning over the first weeks of use, thereby increasing the number of defects found vs. time. Starting from week 6, when we apply the combined algorithms in series, product yield is enhanced a few percent by overkills mitigation. The reason this is so is because classification helps to identify the defects class (mode) and judge them correctly based on the customer specification, while segmentation just re-checks defects detected by machine vision, to judge the existence of the defect, but not the defect mode. Therefore, the integration of two algorithms into our AI system effectively enhances detection capability and improves product yield.

Second-inspection process and results

One of our key achievements with respect to AI development is to integrate our machine vision and deep learning in a unique way. **Figure 5** shows deep learning and machine vision integration as a second-inspection process. Our image processing library includes geometric and subjective inspection. Geometric inspection is clear judgment by machine vision. Examples for these inspection modes are package XY size, bump width, bump XY position offset, etc. In contrast, subjective inspection involves those defect modes for which adding AI to machine vision can significantly improve detection capability. Examples of subjective modes are FM (particle), crack, pattern/copper exposed, etc. At the first stage in **Figure 5**, subjective inspection is done with criteria tighter than the customer specification by machine vision. Reject and overkill results are sent to the AI engine that is already trained. The AI model re-inspects (second-inspection process) reject units based on the customer specification to differentiate final reject and overkill (pass) units. In this approach, deep learning is implemented as a closed-loop feedback to vision software. Therefore, the AI model distinguishes between true reject and overkill so it directly impacts package/product yield.

Deep-learning second-inspection results are shown in **Figure 6**. We have reviewed yield improvements of a certain product over a period of one month. During the first half of the month, a single-digit yield increase could be attained. Furthermore, the AI engine deepens its learning by inspecting more units and increasing the number of defects during this period. On the 17th day of the month, when process excursion or escapee (under-reject) has occurred, tightening the prime inspection criteria in our AI system (**Figure 5**) is an effective way to overcome excursion and under-reject (escapee) risks. By doing so, the yield loss percentage jumped drastically. Therefore, the second-inspection process by our AI system compensates for a two-digit yield loss and significant yield improvements are achieved every day in the second half of the month. In this particular case, the top yield gain

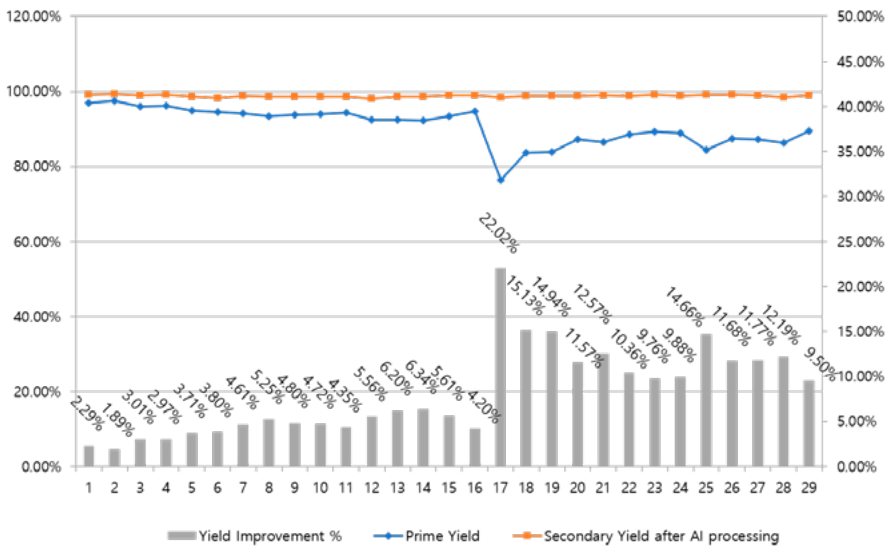


Figure 6: Deep-learning second-inspection process results.

defect modes are scratch, FM (particle), mold scratch, mold crack, incomplete mold, Cu exposed, and chip-out, respectively.

Deploying deep learning on edge vs. cloud computing

Last, but not the least feature of our AI system is edge computing, demonstrated in **Figure 7**. Some semiconductor manufacturers perform AI deployed on cloud computing by their yield management system. Our coherent integration of machine vision and deep-learning engine (**Figure 5**), is considered as edge computing (processing) prior to a customer's yield management process.

In a "regular" cloud-computing system, all data processing is done by an AI server. In this configuration, the units per hour (UPH) is impacted because of the data transferring time between the inspection machine and the AI server. If the server is down, all machines are down too. INTEKPLUS edge computing means the inspection engine has self-computing power for image capturing and data processing. UPH is not impacted because there is no need to send inspection images to a server. The cloud server performs AI learning and model training so there is no need for a high-performance server. In addition, if the server is down, all machines can continue running.

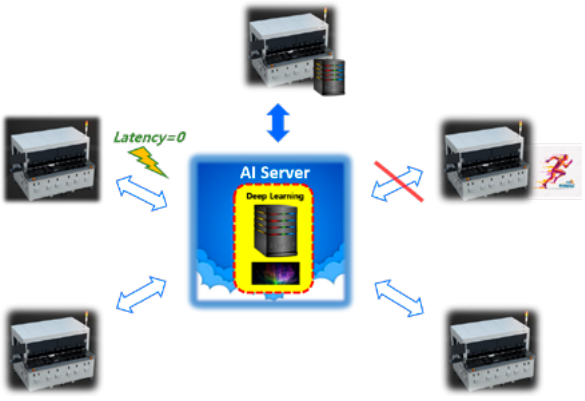


Figure 7: Deep learning deployed on edge computing vs. cloud computing.

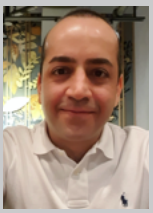
Summary

Heterogeneous packaging is going to be the main driver for today's and future semiconductor packaging in a variety of applications from system in package (SiP) to chiplet packaging. The main process challenge for these multi-chip expensive packages remains how to keep the package cost as low as possible so we can reference Moore's Law as being economically valid, even 50-plus years after the invention of integrated circuits. Final package inspection is the last process step to determine heterogeneous package/product yield. Therefore, having the most capable solution for final package 2D/3D inspection is inevitable in order to enable heterogenous packaging and drive the industry forward.

Our key vision capabilities such as large field of view (FOV), stitching, and high throughput, have been specialized by our heterogenous package inspection. In addition, integration of our deep-learning technology and machine vision into one process has resulted in yield improvement and significant cost savings for our customers: top heterogeneous chip makers.

References

1. P. Wesling, et al., "Heterogeneous integration roadmap," Ch. 8: Single-chip and multi-chip integration," IEEE 2019 Edition.
2. S. Chitchian, et al., "Combined image processing algorithms for improved optical coherence tomography of the prostate nerves," J. Biomedical Optics 15, 2010.
3. S. Chitchian, et al., "Segmentation of optical coherence tomography images for differentiation of the cavernous nerves from the prostate gland," J. Biomedical Optics 14, 2009.
4. S. Chitchian, et al., "Denoising during optical coherence tomography of the prostate nerves via wavelet shrinkage using dual-tree complex wavelet transform," J. Biomedical Optics 14, 2009.



Biography

Shahab Chitchian is Chief Strategy Officer and R&D Director at INTEKPLUS Corporation Ltd., Daejeon, South Korea. Previously, he was Senior Staff Process Engineer at Samsung Electronics, Semiconductor Test and System Package division where he worked on advanced FOWLP development and high-bandwidth memory packaging. He also was a Senior Process Engineer at Intel Co., Assembly and Test Technology Development (ATTD), worked on EMIB package development. He received his MSc and PhD in Electrical and Optical Engineering, from the U. of North Carolina at Charlotte. Email shahab@intekplus.com

Katy Crist
Director, Marketing & Communications
Tokyo Electron

YOU SEE WHAT'S NEXT.

WE ATTRACT WHO'S NEXT.

Attract, hire and retain your 21st century workforce with SEMI.

No one does MORE to fill the talent pipeline for the semiconductor industry. SEMI connects industry, talent and education for members worldwide. To access the next generation of leading minds, visit **semi.org/semiismore**.

SEMI IS MORE

semi™

Fluxless TCB of large-area dies with localized in situ oxide reduction

By Bob Chylak, Adeel Bajwa, Tom Colosimo, Tom Palumbo [Kulicke & Soffa Industries, Inc.]

To meet the demands for ever growing data, greater computing performance, lower power consumption, high bandwidth, and low latency in computationally challenging applications such as server, mobile, graphics, and artificial intelligence, etc., modern semiconductor chips often include large amounts of complex circuitry. This leads to large chip sizes, which for performance scaling reasons often come with reduced interconnect sizes and pitches that are usually assembled in a flip-chip fashion by either mass reflow, or more commonly through a thermal compression bonding (TCB) process. Both methods typically require the application of flux and post-bonding flux residual removal steps, which in general adds process complexity. Even more so, the shrinking contact size and pitch results in short die standoff heights, e.g., 30-50µm, and this makes it extremely difficult to clean the flux residues.

To put this into perspective, a full reticle size die (e.g., 32mm X 28mm) would contain approximately 442,000 contacts at 45µm pitch. We are proposing a fluxless TCB bonding process solution, which eliminates the need for flux application and therefore, post-bonding flux residual cleaning steps. These flux residuals pose severe package reliability concerns. Our proposed fluxless method relies on localized in situ reduction of oxides, using formic acid vapors, from the contact surfaces just prior to and during the bonding process. This method prevents flux process-related overheads and at the same time, achieves the same result.

Challenges: large area dies/high-density interconnects

The fundamental requirement for any flip-chip based TCB process, regardless of the die size and the interconnect pitch, is to get rid of oxides from the mating interfaces so that they can wet and bond reliably. Fluxing is the most common way to remove oxides and is accomplished via dipping the solder-capped pillars into a cavity filled with flux, or by spraying it on the surface of the substrate. A key requirement of a reliable assembly process is to completely clean the flux residues before the bonded die is underfilled. In general, it is very hard to completely remove these residues, especially when they are trapped underneath the die

Several factors, such as flux residue chemistry, die standoff height, die area, and interconnect density, etc., can affect the cleaning process. For large-area die (e.g., ≥1000mm^2) with high interconnect density and tight interconnect pitch, the entrapped flux residues are almost impossible to clean. Even more so, the trends in increasing interconnect densities and decreasing die standoff heights will make flux residue cleaning more and more challenging. Furthermore, the underfill material does not adhere to the surface, which is contaminated with flux residues. The inability to clean residues can result in formation of voids during the molding process. These voids pose serious reliability challenges because the solder will potentially leak into these voids during the reflow, and the solder will often short to the adjacent interconnect resulting in a failure. If the failure is not caught in testing, then it can potentially occur when the chip is soldered to the printed circuit board (PCB), again, resulting in a field failure.

In response to these issues, flux manufacturers have developed "no-clean" fluxes. While these fluxes do leave less residues, the residues that are left are even more difficult to clean. The no-clean fluxes, when used improperly, can lead to electrochemical migration and dendritic growth [1]. Currently, the semiconductor industry is looking at two potential approaches to address these issues. First, there has been R&D activity to eliminate solder-based materials by developing direct Cu-to-Cu bonding technology. This would eliminate the requirement for fluxing altogether and it will further facilitate the transition to finer pitches and smaller die standoff heights. The second approach is to develop methods that can reduce the oxides on interface materials, such as solder and copper, by means other than using flux. Inert environment chambers (e.g., nitrogen gas), plasma treatments (e.g., Ar, H_2), and noble metal finishes (e.g., Au), are some of the prominent approaches that are being investigated both in academia and industry.

Previous work on fluxless bonding

A number of research papers highlight the challenges with fluxless bonding. A few are discussed in the sections below.

Reducing and inert environment approaches. Although inert gases such as Ar and N_2 provide an excellent inert medium for soldering, they are still limited in terms of removing the existing oxides on the bonding surfaces. Some commercially-available tools are equipped with inert environment chambers, but they do not prevent the use of flux materials. Furthermore, they often require very high gas flows (i.e., ≥1000L/min) and the gas consumptions are not economically viable.

The flux application on a substrate is generally always required before it makes its way to the bonding chamber. The reducing gas (e.g., formic acid [FA]) vapor-based mass reflow technology has been around for quite some time, which eliminates the need for pre-reflow fluxing and post-reflow flux residue cleanup steps.

Noble metal finishes. Intel introduced the concept of solder preforms that are intended to be used for fluxless bonding applications. These preforms consist of a low melting point metal (e.g., tin, indium, etc.) and are further capped with a more noble metal finish (e.g., gold, palladium, etc.) to protect it against the oxidation. The preforms are sandwiched between the mating surfaces and the entire assembly is brought to the melting point of the preform, which upon melting, dissolves the noble

metal and simultaneously forms a bond. The complexities involved during the material preparation make it extremely difficult to implement in a production environment [2]. Another approach, also proposed by Intel, relies on breaking thin native oxides on In and Sn surfaces by applying sufficiently high pressure before reaching the melting point in an inert environment. For some processes, the temperature is raised close to, but lower than the melting point of the solder to promote solid-state diffusion. This method also requires a noble metal finish on one of the mating surfaces, and additionally requires an inert environment for a reliable bond [3].

Metal-based interconnects. Metal-metal TCB provides the unique advantage of eliminating the solder materials. Consequently, the reliance on flux materials to remove oxides can be completely eliminated. Other than using the noble metal finishes (e.g., Au, Pt, etc.), most metals (e.g., Cu, Sn, etc.) would require some method to remove the existing native oxides prior to and during the bonding process. In addition, metal-metal bonding has some very stringent requirements for reliable bonding, which include planarity, atomic-level flatness and removal of native oxide. The TCB process further facilitates the solid-state diffusion. Out of available material choices, Cu is of keen interest because it offers excellent material properties and is economically viable. Oxidation prevention on Cu-surfaces, however, is a major challenge. Bajwa, et al. [4], have shown that using a noble metal finish (e.g., gold) protects it from oxidation, but this approach adds severe processing complexities. Noble metal finish is currently not a standard practice in foundries and it will require noble metal finishing on the chips as well as the substrates to achieve its purpose.

More recently, plasma pre-treatments and other in situ oxide reduction techniques (e.g., forming gas reduction) have been used to reduce and prevent oxidation, but they are more applicable to wafer-to-wafer (W2W) processes [6-9]. More often, these methods are employed in controlled environments such as vacuum, N_2, etc. Another major hurdle during metal-metal bonding is the roughness of the joining surfaces and planarity of the chip to the target substrate. Though chemical mechanical planarization (CMP) can effectively resolve this issue, it is only applicable to semiconductor materials (e.g., Si) and cannot be extended to laminates or PCBs. Other techniques, such as fly-cutting of the copper pillars, can be potentially useful, but there is very limited data on its applicability.

Formic acid vapor delivery system

K&S has developed an FA vapor delivery system that can be integrated with both our chip-to-wafer (C2W) and chip-to-substrate (C2S) TCB machines. The delivery system allows injection of FA vapor directly onto the target surfaces immediately before the TCB process, thereby cleaning the metal as well as eliminating solder oxides, so that the use of flux is eliminated. Localized delivery of FA vapors and creating a localized reducing mini-environment is achieved through a custom-designed shroud that fits over a standard TCB bond head.

In the next sections we will discuss the chemistry of the FA-based tin oxide reduction process and the functionalities of the shroud and the gas delivery system. We will also provide a set of experimental data to clarify the performance of the fluxless TCB process specifically for large die (i.e., 900mm^2 bonding case). Additionally, future applicability of the process technology for Cu-to-Cu interconnect will be discussed as will the experimental data showing Cu-to-Cu bonds for different devices.

Formic acid reduction mechanism. The chemical reactions through which formic acid vapor reduces the Sn or Cu surface are given in (1), (2) and [11-13].

$$2HCOOH_{(g)} + SnO_{(s)} \rightarrow (HCOO)_2Sn_{(s)} + H_2O_{(g)} (1)$$

$$(HCOO)_2Sn_{(s)} \rightarrow Sn_{(s)} + 2CO_{2(g)} + H_{2(g)} (2)$$

FA in vapor form reacts with the tin oxide and leaves a thin layer of tin formate. This formate layer covers the bare solder surface and is subsequently removed by raising the surface temperature above 150°C. A very similar chemical reaction also happens for the copper oxide reduction process [5].

Formic acid vapor delivery system and shroud. The schematic of a FA delivery system is shown in Figure 1. The FA vapor is generated by passing an inert carrier (N_2: nitrogen) gas through a bubbler containing formic acid (HCOOH ≥95%) solution. Depending on the bubbler temperature, the N_2 gas coming out is completely saturated with formic acid, e.g., 3.5% FA, 96.5% N_2 at 22°C, which is ultimately transferred to a shroud mounted onto the bond head as shown in Figures 2 and 3.

The FA delivery system is designed to alter the percentage of FA vapor in the carrier gas by further diluting it with N_2 gas. The existing bond heads on current C2W and C2S machines have been modified to include a shroud. The shroud consists of three channels as shown in Figure 1. The innermost channel supplies FA vapor that reduces oxide layers on solder and metal surfaces prior to and/or during the TCB process. The middle channel serves as an exhaust collection port for residual FA vapor and other gaseous reaction byproducts. The outermost channel provides the shielding N_2 gas around the shroud. This helps to maximize the containment of residual FA vapor and other reaction by-products under the shroud area. The escaped FA vapors, inside the machine environment, are ultimately expelled through the facility-provided exhaust system, which, along with the FA concentration inside the machine, is continuously monitored for safety reasons.

The flow rates of the shielding gas, vacuum, and FA vapor are optimized to maximize the formic acid concentration over the desired region while simultaneously reducing the inclusion of oxygen. The shroud design is based on verification of the computational fluid dynamics (CFD)-based gas flow simulations and the experiments. Figure 4 shows a CFD simulation of the saturated FA vapor mass fraction over the die region under operating conditions.

Process flow for fluxless TCB. The bond head, carrying the die, is aligned with the target substrate, which is usually kept at 80-120°C. Then, it is brought to a specified separation distance between the chip and the substrate, which ensures the creation of an effective FA mini-environment.

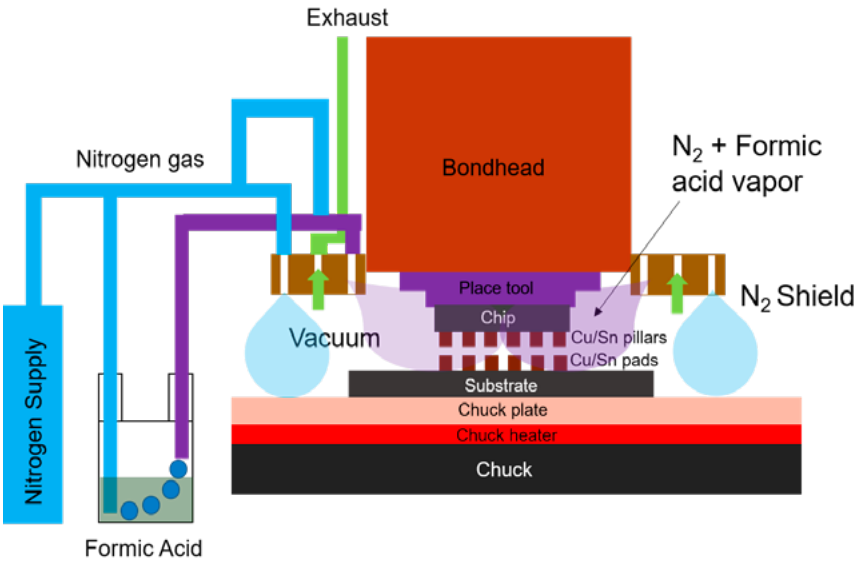


Figure 1: Schematics of a formic acid delivery system with a bond head shroud.

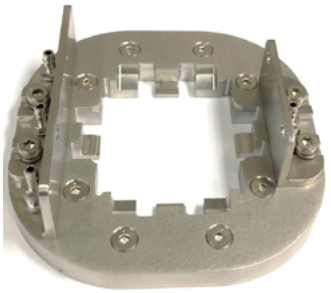


Figure 2: A bond head shroud.

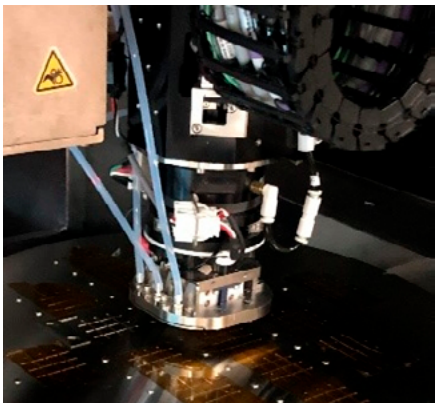


Figure 3: Shroud mounted on the bond head.

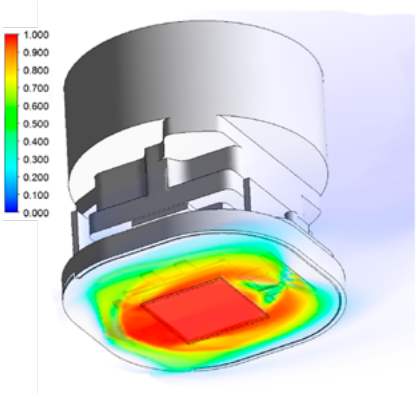


Figure 4: Mass fraction of FA vapor in N_2 over the chip area.

At this point, the gas delivery system triggers the supply of FA vapors. The bond head temperature is raised above 150°C to clean the oxides on the solder caps. If the substrate pad metallization is copper, FA vapor reacts to form a thin copper formate layer that remains on the surface and is cleaned during the TCB process when molten solder makes contact with the copper pad. During TCB, FA vapor supply continues, the bond head temperature is raised above the melting temperature of the solder, and a bond is formed between the chip

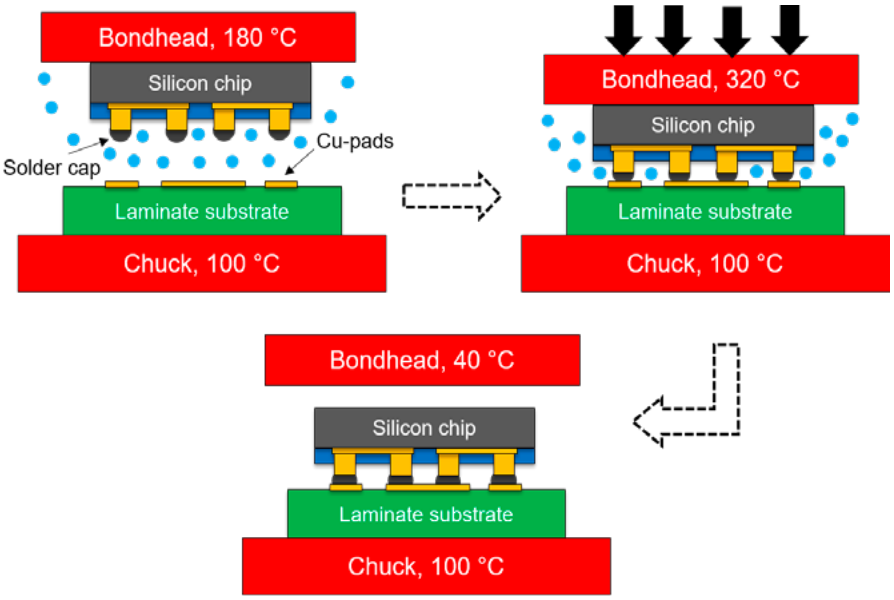


Figure 5: Fluxless TCB process.

and substrate in a position-limited force-controlled fashion. The FA vapor delivery is discontinued after the completion of the TCB step. Figure 5 shows the process flows.

Bonding of large-area die

To demonstrate the advantages of our fluxless bonding process for large area dies, we have used a Si-based test die with an area of 900mm^2. The die was terminated with solder (SnAg) capped copper pillars of 36µm diameter and contained two variations of pitches: 55µm and 80µm. A die-to-die bonded assembly was formed using FA vapor-based oxide reduction. Figure 6 shows a 30mm die bonded to another 30mm die using the fluxless TCB process.

The cross section of the bonded assembly revealed excellent joints as shown in Figure 7—superior to the joints made using the conventional flux-based TCB process as shown in Figure 8.

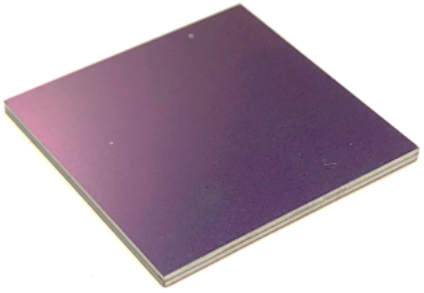


Figure 6: A 30mm die-to-die bonded assembly.

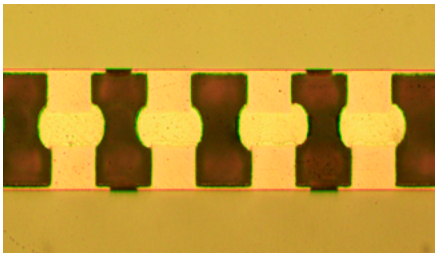


Figure 7: A fluxless TCB sample.

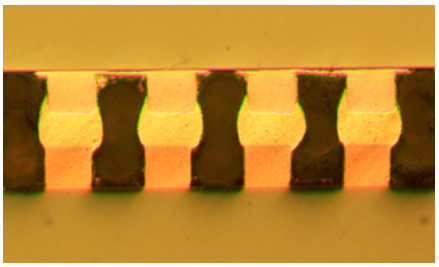


Figure 8: A flux-based TCB sample.

Cu-Cu direct bonding

The following sections discuss some of the challenges with respect to Cu-Cu direct bonding.

Main challenges for Cu-Cu bonding. Maintaining the copper surfaces to be oxide free is extremely challenging in a C2W or C2S machine because the surrounding environment is mostly air. An inert gas such as N_2 can be used to keep the environment oxide free, but very large flow rates (e.g., ≥1000L/min) are required. Moreover, using an inert gas does not reduce the existing natural oxide layers. Another issue for Cu-Cu bonding is to keep the surface atomically flat and planarized across the entire chip area. Today, it is done through a CMP process, which can only be utilized for the cases where both chip and substrate are made up of silicon. For laminates and PCBs, this still remains an issue. The CMP process adds more complexities and it is not a common practice of outsourced semiconductor assembly and test suppliers (OSATS), though

foundries use this process regularly for chip manufacturing. Currently, it makes more sense to mainly rely on foundries to provide the samples for direct Cu-Cu bonding.

Another less explored method is the fly-cutting to planarize the Cu-surfaces and it results in reasonably flat surfaces (e.g., 10-20nm) in comparison with CMP, which can result in ≤1nm surface roughness. This means that we would require very large pressures to flatten the asperities on a copper surface. Developing a cost-effective method to achieve a CMP-like surface finish still remains a challenge. The flatness requirement becomes even more stringent as the die areas, and consequently, the total I/O contact area sizes, increase. Rougher surfaces would require a very large bonding force – several tens of kN, for example. The chemistry of the plated-Cu contacts also plays an important role for defining the ductility of the copper. In general, more ductile contacts would require less pressure. Annealing the copper contact after plating also makes a huge difference because it makes the grain size larger and the copper more ductile.

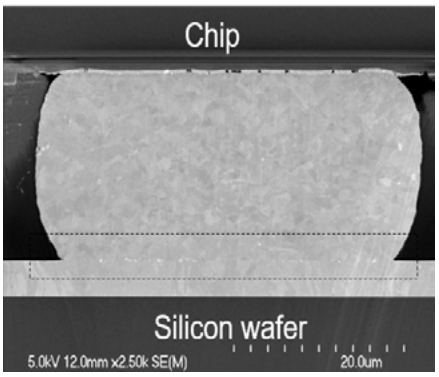


Figure 9: Chip-to-wafer as-plated copper pillar bond. The interface is highlighted with a dotted line.

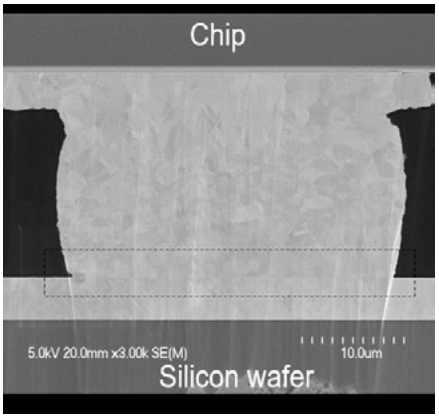


Figure 10: Chip-to-wafer fly-cut plated copper pillar bond. The interface is highlighted with a dotted line.

Cu-Cu bonding with in situ oxide reduction. We have extensively investigated the effectiveness of the FA vapor-based in situ copper oxide reduction process and it is found to be equally effective for the elimination of copper oxides. Our FA vapor delivery system along with the bond head mounted shroud successfully reduces the surface copper oxides by creating a reducing mini-environment. FA vapors react with Cu surface oxides to create a copper formate layer, which is eventually decomposed by elevating the temperature above 180°C. The Cu-surface roughness dictates a key TCB process parameter: bonding pressure. We have developed direct Cu-Cu TCB processes for a variety of surface finishes

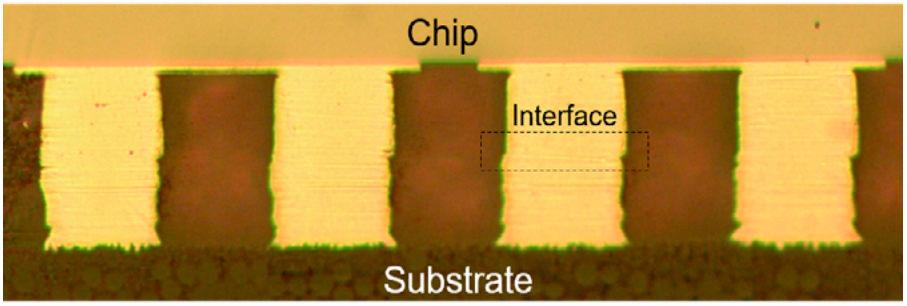


Figure 11: Chip-to-substrate as-plated copper pillar bond.

(e.g., as-plated copper, fly-cut copper, etc.) for both C2W and C2S applications. The rougher surfaces require relatively higher bonding pressure for a successful bond. For instance, as-plated Cu required at least 250MPa of pressure to achieve a reliable bond, while fly-cut Cu required only 80MPa to achieve the same results. In both cases, C2W and C2S, the bonding interface was nearly void free and exhibited very high shear strength (i.e., ≥150MPa). Figures 9 and 10 show examples of as-plated and fly-cut die pillars bonded to a blank Cu-plated Si wafer. Figure 11 shows an example of a fly-cut Cu pillar bonded to an as-plated Cu pillar on a laminar substrate.

Summary

We have successfully demonstrated a flux-less TCB process that utilizes in situ formic acid vapor application to reduce the oxide from solder as well as the copper surfaces. This method provides an opportunity to bond large area dies with high-density interconnects. Using flux-based TCB would require post-bond flux residuals cleanup, which is very challenging to achieve for these types of applications. Furthermore, FA-based in situ cleaning is equally effective for Cu-Cu bonding technology and we have shown promising results for both C2W and C2S applications.

References

1. M. Bixenman, et al., "Why clean a no-clean flux," Proc. of the International Conf. on Soldering and Reliability (ICSR) 2016, pub. by SMTA.
2. Lie, et al., "Approaches for fluxless soldering," US 6848610 B2, Feb. 1, 2012.
3. Jadhav, et al., "Methods including fluxless chip attach processes," US 7534715 B2, May 19, 2009.
4. A. A. Bajwa, S. Jangam, S. Pal, N. Marathe, M. Goorsky, T. Fukushima, et al., "Heterogeneous integration at fine pitch (≤10µm) using thermal compression bonding," 2017 IEEE 67th Electronic Comp. and Tech. Conf. (ECTC), Orlando, FL, 2017.
5. S. Jangam, A. A. Bajwa, U. Mogera, P. Ambhore, T. Colosimo, R. Chylak, "Fine-pitch (≤10µm) direct Cu-Cu interconnects using in situ formic acid vapor treatment," 2019 IEEE 69th Elec. Comp. and Tech. Conf. (ECTC), Las Vegas, NV, USA, 2019, pp. 620-627.
6. C-T. Ko, K-N. Chen, "Low-temperature bonding technology for 3D integration," Microelectronics Reliability 52, no. 2, 2012, pp. 302-311.
7. C. S. Tan, L. Dau Fatt, X. F. Ang, J. Wei, K. C. Leong, "Low-temperature CuCu thermo-compression bonding with temporary passivation of self-assembled monolayer and its bond strength enhancement," Microelectronics Rel. 52, no. 2, pp. 321-324, 2012.
8. K. Tanaka, W-S. Wang, M. Baum, J. Froemel, H. Hirano, S. Tanaka, et al., "Investigation of surface pre-treatment methods for wafer-level Cu-Cu thermocompression bonding," Micromachines 7, no. 12, pp. 234, 2016.
9. S. L. Chua, G. Y. Chong, Y. H. Lee, C. S. Tan, "Direct copper-copper wafer bonding with Ar/N_2 plasma activation," 2015 IEEE Inter. Conf. on Electron Devices and Solid-State Circuits (EDSSC), Singapore, pp. 134-137, 2015.
10. Y. Arai, M. Nimura, H. Tomokage, "Cu-Cu direct bonding technology using ultrasonic vibration for flip-chip interconnection," 2015 Inter. Conf. on Electronics Packaging and iMAPS All Asia Conf. (ICEP-IAAC), Kyoto, pp. 468-472, 2015.
11. W. Yang, M. Akaike, T. Suga, "Effect of formic acid vapor in situ treatment process on Cu low-temperature bonding," in IEEE Trans. on Components, Packaging and Manufacturing Tech., vol. 4, no. 6, pp. 951-956, June 2014.
12. M. Schmeißer, "Reduction of copper oxide by formic acid an ab-initio study," Fraunhofer, Chemnitz U. of Technology, 2012.
13. T. G. A. Youngs, S. Haq, M. Bowker, "Formic acid adsorption and oxidation on Cu (110)," Surface Science, no. 10, pp. 1775-1782, 2008.

Biographies

Bob Chylak is the CTO at Kulicke & Soffa Industries, Inc., Fort Washington, PA. He is responsible for the research and development of packaging solutions across all K&S products. Bob has more than 30 years of experience in the semiconductor industry and has BS and MS degrees in Electrical Engineering from Penn State U. He completed Management Studies at Stanford U. and MIT Sloan. He has published more than 50 papers and holds multiple patents. Email bchylak@kns.com

Adeel Bajwa is a Staff Engineer in the R&D Engineering Department at Kulicke & Soffa Industries, Inc., Fort Washington, PA. He is currently working on fluxless TCB and fine-pitch metal-metal interconnect technologies. He holds a BSc degree in Electrical Engineering and MSc and PhD degrees in Microsystems Engineering from the U. of Freiburg, Germany. He has published more than 50 papers in scientific journals and conference proceedings.

Reliability testing and data analysis of lead-free solder joints

By John H. Lau [Unimicron Technology Corporation]

Thirty years ago, the author asked key people in solder materials, solder mechanics, and solder manufacturing to write a book chapter in their areas of expertise. The result was a book called *Solder Joint Reliability: Theory and Applications* [1], published in 1991. Today, students, engineers, and researchers all over the world still use the book as their reference.

In the past 30 years, however, soldering technology has changed dramatically. In order to comply with the restriction of the use of certain hazardous substances (RoHS) regulations, the most challenging one is from tin-lead solders to lead-free solders. Therefore, the reliability of lead-free solder joints is under scrutiny.

Conducting reliability engineering of lead-free solder joints consists of three major tasks: design for reliability (DFR), reliability testing and data analysis, and failure analysis (see **Figure 1**). Usually, the procedure starts with a design of the interconnects of a particular semiconductor integrated circuit (IC) package with, e.g., the given chip size, the solder alloys, the molding compound, and the corresponding printed circuit board (PCB) and demonstrates that the design is electrically, thermally, mechanically, and chemically sound. The next step in the process is for a certain number of samples of the sound or reliable design to be built and tested under certain conditions for a certain period of time. The test data (failures) are then analyzed and fitted into a life-distribution designation for the interconnects. Next, failure analysis should be done on the failed samples to find out the root cause and understand the reason for their failure. In this study, the focus is on reliability testing and data analysis. Emphasis is placed on acceleration models and factors. Some recommendations in these areas will also be proposed.

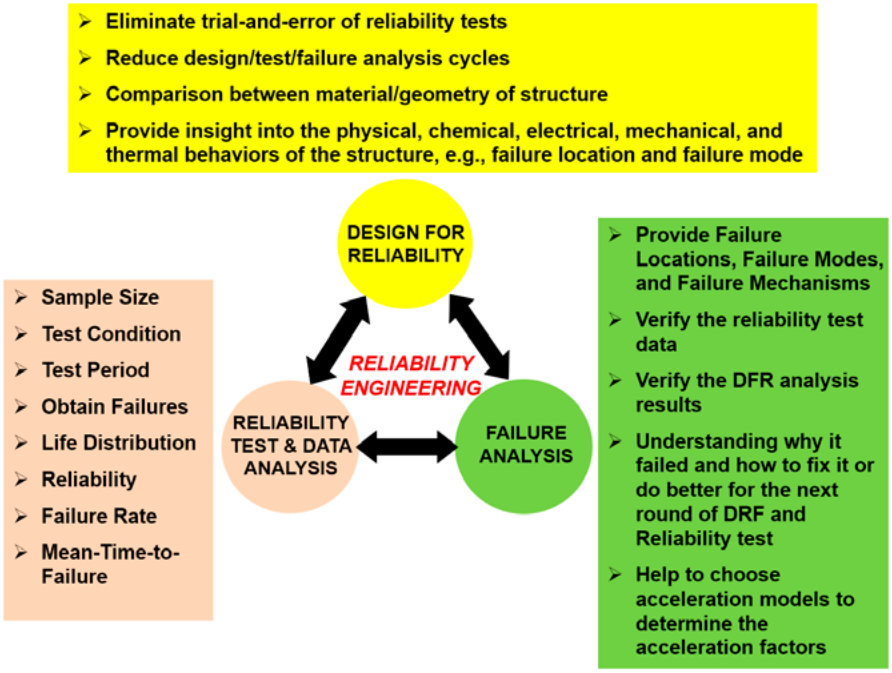


Figure 1: Reliability engineering: design for reliability, reliability testing and data analysis, and failure analysis.

Reliability testing and data analysis

This section discusses various aspects of reliability testing and data analysis. Examples of thermal cycling and drop tests of a fan-out lead-free package are also given.

Definition of reliability. In this study, reliability of an interconnect (e.g., solder bump, solder joint, or microbump) of a particular semiconductor package in an electronic product is defined as the probability that the interconnect will perform its intended function for a specified period under a given operating condition without failure [2-5]. Numerically speaking, reliability is the percent of survivors; that is, $R(x) = 1 - F(x)$, where $R(x)$ is the reliability (survival) function and $F(x)$ is the cumulative distribution function (CDF). Life distribution is a theoretical population model used to describe the lifetime of an interconnect, and is defined as the CDF, that is, $F(x)$ for the interconnect population. Therefore, the one and only way to determine the interconnect reliability is by reliability testing to determine the $F(x)$.

Objective of reliability testing. The objective of reliability tests is to obtain failures (the more, the better) and to best fit the failure data to determine the parameters of the CDF of a chosen probability distribution (e.g., Weibull). The number of items (i.e., sample size) to be tested should be such that the final data are statistically significant. The reliability test time is unknown, but usually takes a while (e.g., a few months for thermal cycling tests). It should be noted and emphasized that as soon as the life distribution $F(x)$ of the interconnects is estimated by reliability testing, the reliability $R(x)$, failure rate, cumulative failure rate, average failure rate, mean time to failure, etc., of the interconnects are readily determined [2-5].

Most reliability tests are accelerated tests, with increased intensity of exposure

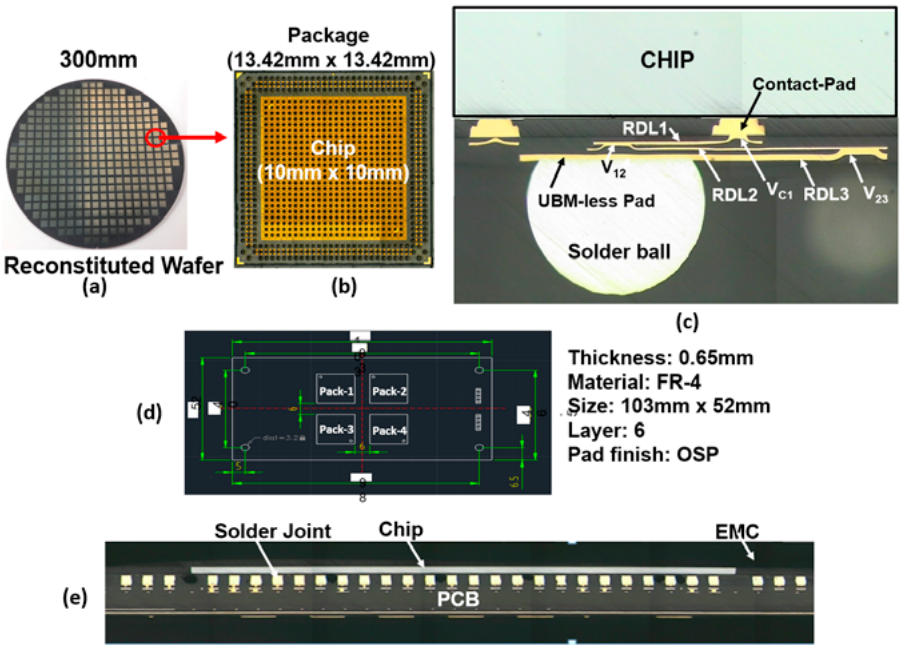


Figure 2: Fan-out package with lead-free solder joints on a PCB.

to aggressive environmental conditions and realistic sample sizes and test times. Acceleration models, therefore, are needed to map (transfer) the failure probability, reliability function, failure rate, and mean time to failure from a test condition to an operating condition. In establishing the acceleration models for lead-free interconnects, their surrounding materials (e.g., solder, molding plastic, ceramic, copper, fiber-reinforced glass epoxy, and silicon), loadings (e.g., stress, strain, temperature, humidity, current density, and voltage), and failure mechanisms and modes (e.g., overload, fatigue, corrosion, and electromigration) must be considered.

Objective of qualification testing. Unlike reliability tests, the objective of qualification tests is "PASS" or "NOT PASS" and the test time is well-defined ahead of time. As soon as there is a failure before the agreed test time, the test will usually stop and failure analysis is performed to find out why the failure occurred. After all the changes, e.g., redesign, a new qualification test will start again. The sample size of qualification tests is usually less than that of reliability tests. In short, the objective of qualification tests is not intended to obtain failures nor life distribution (or reliability).

Thermal cycling test example. Thermal cycling is the most common test in solder joint reliability. For example, the solder joint reliability of a fan-out wafer-level package (FOWLP) of a 10mm x 10mm chip on a PCB, as shown in **Figure 2**, subjected to thermal cycling has been reported [5,6]. The package dimensions are 13.42mm x 13.42mm (see **Figure 2b**), and there are three redistribution layers (RDLs) (see **Figure 2c**) and 908 solder balls at a diameter of 0.2mm on 0.4mm-pitch. The solder balls are made of Sn3Ag0.5Cu.

The dimensions of the PCB are 103mm x 52mm x 0.65mm (see **Figure 2d**). The PCB has 6 layers and is made of FR-4, and the pad finishing is an organic solderability preservative (OSP). It is NSMD (non-solder mask defined) and the solder mask opening diameter is 0.28mm. There are four packages on a PCB as shown in **Figure 2d**.

Figure 2e shows the cross section of one of the fan-out package PCB assemblies. It can be seen that the solder joints are properly made (no bridging and head-in-pillow) and the daisy chain on the PCB pads is clearly seen. The pads on both package and PCB are interconnected in an alternating pattern so as to provide a daisy chain connection (for continuous measurement

during testing) after they are assembled. There is no underfill between the package and the PCB.

Fifteen boards (each with four packages) are used for the temperature cycling tests. The sample size is 60 fan-out packages. Thermal cycling is performed in a Votsch 7027-15 environmental chamber as shown in **Figure 3a**. The temperature input to the chamber (measured in the air of the chamber) goes from room temperature to 85ºC and stays there for 15 minutes; the temperature then ramps down to -40ºC and stays there for 15 minutes and then ramps up again to 85ºC and stays there for 15 minutes, and so forth. The ramp up and ramp down times are 15 minutes each. The cycle time is one hour, i.e., one hour per cycle. The acquisition system is an Agilent 30970A data logger as shown in **Figure 3b**.

The Weibull cumulative distribution function *F(x)*, reliability function *R(x)*, failure rate *h(x)*, and mean-time-to-failure *(MTTF)* are given by, respectively [5]:

$$F(x) = 1 - exp\left[-\left(\frac{x-\gamma}{\theta}\right)^{\beta}\right]$$

$$R(x) = exp\left[-\left(\frac{x-\gamma}{\theta}\right)^{\beta}\right]$$

$$h(x) = \frac{\beta}{\theta}\left(\frac{x-\gamma}{\theta}\right)^{\beta-1}$$

and

$$MTTF = \gamma + \theta\Gamma\left(1+\frac{1}{\beta}\right)$$

where x is the random variable (e.g., life or cycles), γ is the expected minimum value of x (it is also referred to as the location parameter), θ is the characteristic value (at 63.2% failures) or the scale parameter, which could be used to represent the quality of a product, and β is the Weibull slope, or the shape parameter, which is a measure of the uniformity of a product (the larger the slope, the more uniform the product). It is frequently reasonable to assume that the expected minimum value of life (γ) of the population, i.e., the lower bound of life of the population, is equal to zero, then we have the two-parameter (θ, β) Weibull distribution. Γ is the gamma function.

The thermal cycling test of the lead-free fan-out package stops at 1,100 cycles and there are 14 failures (including one early failure at 58 cycles). The failure criterion is when the resistance of the daisy-chain of the PCB fan-out package assembly increases by 50%. The cycle at which the first solder joint of the package failed is considered as the cycle-to-failure

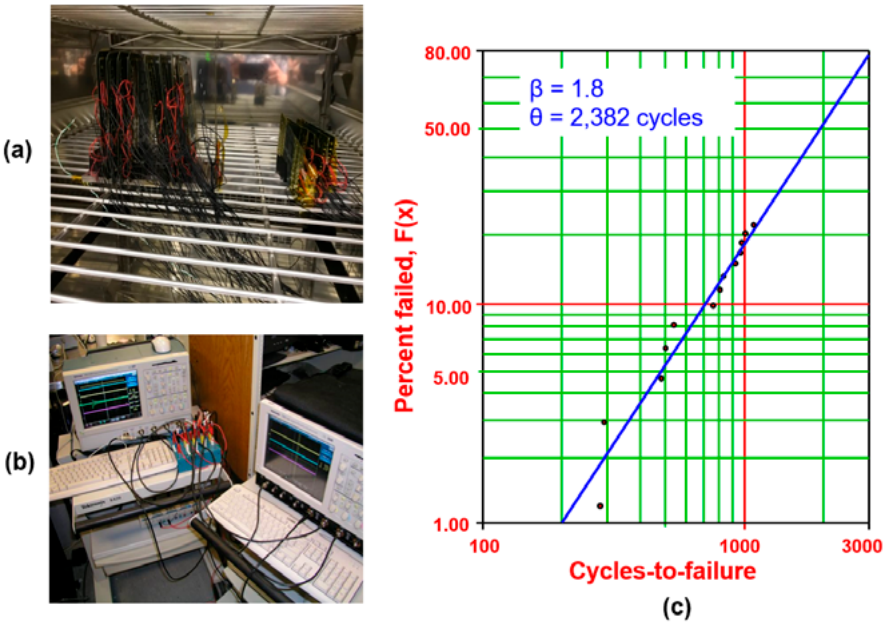


Figure 3: a) Thermal cycling chamber and samples; b) Data acquisition system; c) Weibull plot of the thermal cycling results for the fan-out package with lead-free solder joints on a PCB without underfill.

of the lead-free package. It can be seen from **Figure 3c** that the Weibull slope (β) and characteristic life (θ) of the lead-free fan-out package are, respectively, 1.8 and 2,382 cycles. Once the parameters of the life distribution for the fan-out lead-free package have been estimated by a reliability test, important reliability questions such as the following can be readily answered:

1. What is the probability that the lead-free fan-out package will fail by 400 cycles? $F(400) = 1 - \exp[-(400/2382)^{1.8}] = 0.039$, i.e., 3.9% of the lead-free fan-out package will fail at 400 cycles.
2. If we use 1000 units of them, how many do we expect to fail in the first 400 cycles? We will expect 1000 x 0.039 = 39 units will fail in the first 400 cycles.
3. What is the probability that the lead-free fan-out package will survive 300 cycles? $R(300) = \exp[-(300/2382)^{1.8}] = 0.976$, i.e., 97.6% of the lead-free fan-out package will survive at 300 cycles.
4. What is the failure rate of the lead-free package at 365x24 cycles? $h(8760) = (1.8/2382)(8760/2382)^{0.8} = 2141808 \times 10^{-9}$ per cycle = 2141808 FITs (ppm/k).
5. What is the *MTTF* of the lead-free fan-out package? $MTTF = 2382\Gamma(1 + 1/1.8) = 2382 \times 0.97084 = 2313$ cycles.

The failure locations occur at the package corner solder joints (see **Figure 4a**), and at the corner solder joints right underneath the chip corners. The failure mode is the cracking of the solder interface between the RDL3 of the fan-out package and the bulk solder as shown in **Figure 4b**.

Drop test example. For mobile products, completing a drop test is very important. The fan-out lead-free package and the PCB are the same as those [5,6] for the thermal cycling test. The drop test setup is according to JEDEC Standard JESD22 - B111 as shown in **Figure 5**. After more than 20 tries, the correct height of the drop table is obtained, which yields the drop spectrum with 1500G/ms as shown in **Figure 6**.

The drop condition for the test described above is 1,000 drops. There are 24 samples. The ones without underfill failed very early and the failure mode was

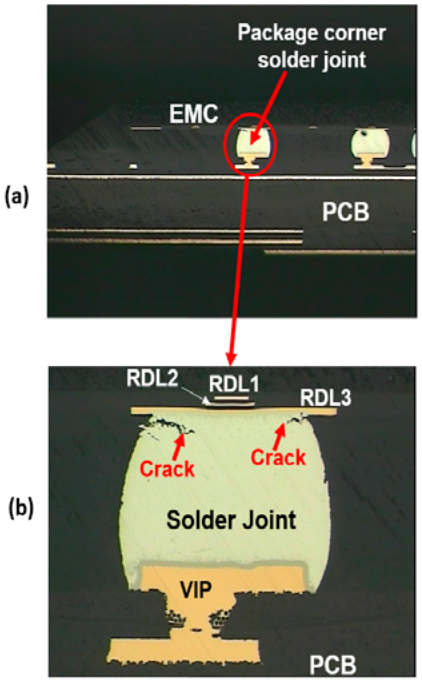


Figure 4: a) Failure location, and b) failure mode of the fan-out package with lead-free solder joints on a PCB subjected to thermal cycling.

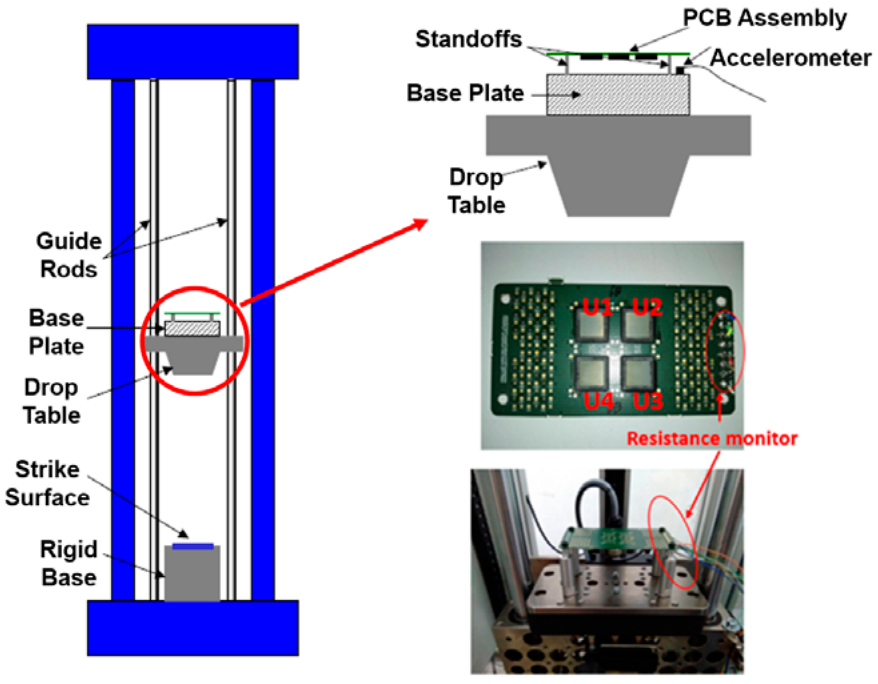


Figure 5: Drop test setup.

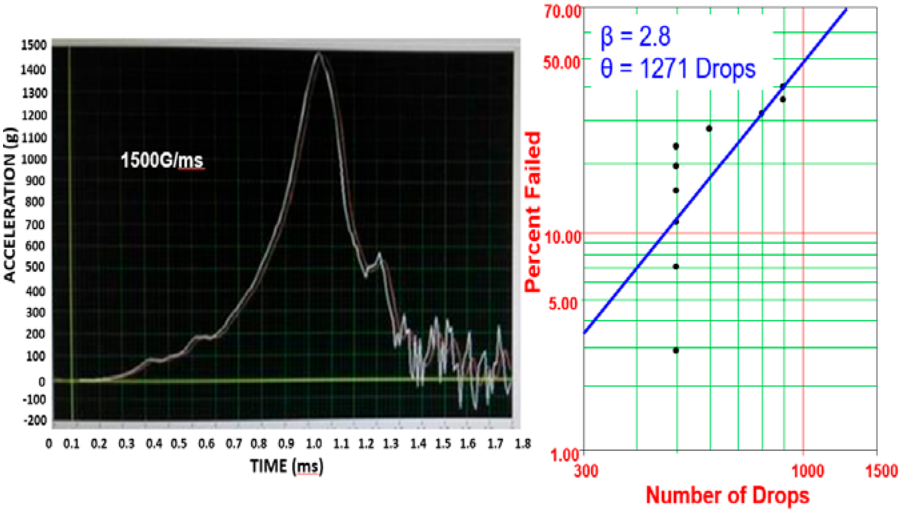


Figure 6: Drop spectrum and Weibull plot of the drop test results for the fan-out package with lead-free solder joints on a PCB with underfill.

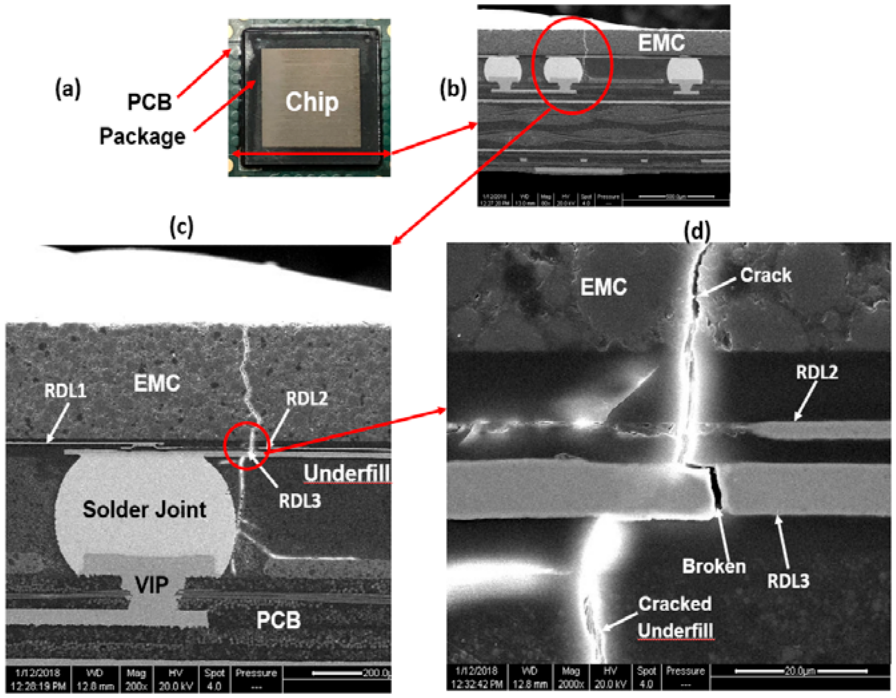


Figure 7: Failure modes of the fan-out lead-free package PCB assembly subjected to a drop test (>500 drops): a) The fan-out package PCB assembly; b) Cross section of the assembly; c) and d) Close-up of a cracked RDL; a cracked EMC; and a cracked underfill (but no crack in the solder joint).

the breaking of the Cu conductor wiring of the RDL in the fan-out package near the solder joint. Another 24 packages and their PCB assemblies (samples) were made and underfilled. The material properties of the underfill are: the filler content = 25%, the maximum filler size = 5µm, the average filler size = 1-2µm, the curing time and curing temperature = 8 minutes @ 135ºC, or 5 minutes @ 150°C. The Young's modulus, Poisson's ratio, and coefficient of thermal expansion (CTE) are respectively, 4-5GPa, 0.35, and 50-52x10^{-6}/°C. The drop condition is also 1,000 drops. The results (with a Weibull distribution) are shown in **Figure 6**. It can be seen that the Weibull slope is 2.8 and the characteristic life is 1,271 drops, and all 24 samples passed 480 drops without failure. The first failure occurs after 500 drops and the failure modes are shown in **Figure 7**. It can be seen that the RDLs of the fan-out package are broken. In this case, the EMC of the fan-out package has cracks and the underfill between the package and PCB also has cracks. However, the solder joint did not fail during the drop test.

Failure criteria

As mentioned earlier, the one and only way to determine the reliability, failure rate, characteristic life, mean life, etc., of lead-free solder joints is by conducting reliability tests. The most important factor in reliability tests is the failure criteria.

During reliability tests, we continuously perform resistance measurements. The failure criterion is defined as the resistance increases to certain (e.g., 1 to ∞) percentages of the original resistance. Usually daisy chains, which connect the solder interconnects of the chip/package and substrate/PCB, allow the measurements. Because most solder joints under fatigue loadings will go through crack initiation, crack propagation, and crack rupture sooner or later, the daisy chains' resistance will increase accordingly, that is, from small to large, and become infinite when the solder joint is totally cracked (opened). (Most people unintentionally pick the last one.) Because the exact "resistance vs. crack length" relations of various lead-free solder joints don't exist today, people just randomly pick a number, which is from 1 to an infinite percentage of the initial resistance. That's one of the reasons why there are so many different Weibull plots in the research literature, even with the same package, PCB, solder paste, sample size, test condition, and test period.

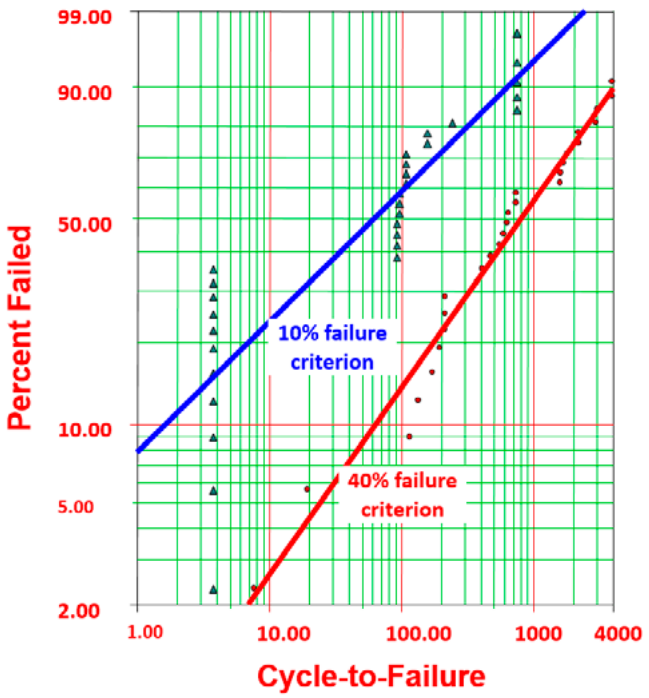


Figure 8: Weibull plots of a 208-pin PQFP with 10% and 40% resistant change failure criteria.

Figure 8 shows the Weibull plots of a 208-pin plastic quad flat pack (PQFP) with a lead-free solder paste subject to -40 ⇄ 125°C thermal cycling. A computer stored all the resistance measurements. Let's define one failure criterion as 10% resistance increase and the other as a 40% increase. It can be seen from the Weibull plots that: 1) the life distribution is many times different from the failure criterion based on the 10% resistance increase and the 40% resistance increase; 2) for the same percent failures, the life taken from the failure criterion based on higher resistance increases is longer; and 3) as expected, there are more failures with the failure criterion based on lower resistance increases.

Another important factor affecting the results of reliability testing is data extraction. During tests, how many measurements should we take at each cycle? It could be four measurements or eight measurements. In order to avoid false failures, it is recommended to compare the measurements of every channel at every two sequential cycles (no matter if it is four measurements or eight measurements). A three-cycle moving average method is recommended. Please read [5] for more details.

Why acceleration models?

In reliability tests, the most ideal situation is to have the test conditions very close to the use (operating) conditions. However, because of time-to-market and cost savings, this is almost impossible. The practical way is to run accelerated tests with increased intensity, and realistic sample sizes and test times. (Most reliability tests, therefore, are accelerated tests.) In this case, the price to pay is to construct the acceleration factors to map (transfer) the failure probability, reliability function, mean life, and failure rate from a test condition to the service operating (use) condition. Acceleration models are needed for determining the acceleration factors [4,5,7].

As mentioned earlier, the best-fit Weibull life distribution of a set of lead-free interconnects is

$$F_T(x_T) = 1 - exp\left[-\left(\frac{x_T}{\theta_T}\right)^{\beta_T}\right]$$

where F_T, x_T, θ_T, and β_T are the life distribution, time-to-failure, characteristic life, and shape parameter (or Weibull slope), respectively, under the test conditions. Let's consider the following simple acceleration model (transformation or mapping) [7]:

$$x_o = \alpha x_T^{\eta}$$

where x_o is the time-to-failure at operating condition, α is the acceleration factor, and η is a real number larger than zero, e.g., 8.888. Then, we have

$$F_o(x_o) = 1 - \exp\left[-(x_o/\theta_o)^{\beta_o}\right]$$

$$R_o(x_o) = \exp\left[-(x_o/\theta_o)^{\beta_o}\right]$$

$$h_o(x_o) = \frac{\beta_o}{\theta_o}\left(\frac{x_o}{\theta_o}\right)^{\beta_o - 1}$$

$$\beta_o = \frac{\beta_T}{\eta}$$

$$\theta_o = \alpha\theta_T^{\eta}$$

where F_o, R_o, h_o, x_o, θ_o, and β_o are the life distribution, reliability, failure rate, time-to-failure, characteristic life, and Weibull slope, respectively, at the operating conditions.

Linear acceleration. When $\eta = 1$, i.e., linear acceleration [7]

$$x_o = \alpha x_T$$

then we have

$$\beta_o = \beta_T$$
$$\theta_o = \alpha\theta_T$$

$$F_o(x_o) = 1 - \exp\left[-(x_o/\alpha\theta_T)^{\beta_T}\right]$$

$$h_o(x_o) = \frac{\beta_T}{\alpha\theta_T}\left(\frac{x_o}{\alpha\theta_T}\right)^{\beta_T - 1}$$

It can be seen that, for a linear acceleration ($\eta = 1$), the Weibull CDF plots of the operating condition and testing condition should have the same Weibull slope (i.e., the shape parameter remains the same), and the characteristic life is different by a factor of the acceleration factor. It should be pointed out and emphasized that the same Weibull slope is not an assumption but rather just came out naturally for linear acceleration [7]. So, if two or more different test cells (with different temperature conditions) of the same lead-free interconnects yield very different Weibull slopes, then either the linear acceleration model is not good (i.e., $\eta \neq 1$), or the Weibull distribution is not fitted with the test data, or both. Usually, the Weibull distribution adequately represents fatigue failures, therefore, nonlinear acceleration models may need to be considered. All the papers in the literature have been using the linear acceleration, i.e., $\eta = 1$, or $x_o = \alpha x_T$ or $N_o = \alpha N_T$, where $N_o = x_o$ and $N_T = x_T$.

In most solder joint reliability tests of integrated circuit (IC) packages on PCBs, the luxury of running multiple thermal cycling conditions is not available (due to the test time, chamber occupation, and manpower). Therefore, acceleration models are required to predict the life distribution and failure rate of the solder joints under anticipated service conditions. In the literature, for SnPb and

lead-free solders, the linear acceleration has been assumed.

Other acceleration models. It should be emphasized that the linear acceleration, $x_o = \alpha x_T$, is only one form of acceleration. There are many other accelerations,e.g., $\eta \neq 1$ and $x_0 = \alpha x_T^{\eta} + \zeta x_T^{\xi}$. More reliability tests should be performed for a variety of IC packages and temperature cycling conditions to establish the suitable acceleration model [5].

Acceleration factors

To determine α (the acceleration factor), you need an acceleration model, which has been discussed in the previous section. How do you choose an acceleration model? The answer is tests and failure mechanisms, which are discussed below.

Well-known linear acceleration factor for SnPb based on frequency and maximum temperature. For SnPb solder joints, the following Norris-Landzberg linear acceleration factor (Equation (2.68) of [4]), has been frequently used:

$$\alpha = \left(\frac{f_o}{f_T}\right)^q \left(\frac{\Delta T_T}{\Delta T_o}\right)^c \exp\left[1414\left(1/T_o - 1/T_T\right)\right]$$

In this equation, T_T, f_T, ΔT_T, and T_o, f_o, ΔT_o, are the maximum temperature during cycling (in degrees Kelvin), the temperature cycling frequency, and the temperature range (in degrees Celsius), respectively, at the testing conditions and at operating conditions. For SnPb solders, q = 1/3, and c = 1.9 ~ 2.0 have been used.

Linear acceleration factor for Sn3Ag0.5Cu in terms of dwell time and maximum temperature. For Sn3Ag0.5Cu, Pan et al., [8] proposed a modification of the classical Norris-Landzberg equation by replacing the cyclic frequency (f) with the dwell time at high temperature (t). A nonlinear curve fit of the temperature cycling data of the ceramic ball grid array (CBGA), chip scale package (CSP), and thin small outline package (TSOP) on a PCB at various temperature ranges such as 0 ⇄ 60ºC, 0 ⇄ 100ºC, and -25 ⇄ 125ºC, Pan et al., [8] obtained the following linear acceleration factor for Sn3Ag0.5Cu:

$$\alpha = \left(\frac{\Delta T_t}{\Delta T_o}\right)^{2.65} \left(\frac{t_t}{t_o}\right)^{0.136} exp\left[2185\left(\frac{1}{T_{max,o}} - \frac{1}{T_{max,t}}\right)\right]$$

where t_t and t_o are the dwell time at high temperature, ΔT_t and ΔT_o are the temperature range (in degrees Celsius) during thermal cycling, and $T_{max,t}$ and $T_{max,o}$ are the maximum (peak) temperatures (in degrees Kelvin) attained during thermal cycling, respectively under testing conditions and under operating conditions. It should be noted that the effect of dwell time on the acceleration factor is (t_t/t_o). Because t_o (operating dwell) is usually longer than t_t (testing dwell), that means that for a positive power, (t_t/t_o) is a deceleration factor. This is just like the effect of (f_o/f_t) in the classical Norris-Landzberg equation; usually f_t (testing frequency) is larger than f_o (operating frequency).

As an example based on the above discussion, consider the test conditions are: 0 ⇄ 100ºC with the dwell time at high temperature = 15 minutes and the operating conditions are: 20 ⇄ 70ºC with the dwell time at high temperature = 720 minutes. Then the linear acceleration factor is:

$$\alpha = \left(\frac{100}{50}\right)^{2.65} \left(\frac{15}{720}\right)^{0.136} exp\left[2185\left(\frac{1}{273+70} - \frac{1}{273+100}\right)\right] = 6.18$$

For other surface mount devices (SMDs) and test conditions, Miremadi, et al., [9] obtained the following linear acceleration factor with Sn3Ag0.5Cu solder alloy:

$$\alpha = \left(\frac{\Delta T_t}{\Delta T_o}\right)^{A} \left(\frac{t_t}{t_o}\right)^{B} exp\left[C\left(\frac{1}{T_{max,o}} - \frac{1}{T_{max,t}}\right)\right]$$

The constants A, B, and C are shown in Table 1 for three SMDs and test conditions.

Linear acceleration factor for Sn3Ag0.5Cu in terms of frequency and maximum temperature. For Sn3Ag0.5Cu, some of the researchers still use the classical Norris-Lanzberg acceleration model, i.e., in terms of the cyclic frequency and maximum temperature during thermal cycling. For PBGA, flip-chip plastic ball grid array (fc-PBGA), CBGA, CSP, quad flat pack (QFP), flip-chip, etc. with test conditions such as ΔT = 135ºC and ΔT = 180ºC, Lall et al., [10] obtained the following linear acceleration factor for Sn3Ag0.5Cu:

$$\alpha = \left(\frac{\Delta T_t}{\Delta T_o}\right)^{2.3} \left(\frac{f_o}{f_t}\right)^{0.3} exp\left[4562\left(\frac{1}{T_{max,o}} - \frac{1}{T_{max,t}}\right)\right]$$

For example, in the thermal cycling test of the fan-out lead-free (Sn3Ag0.5Cu) package of the 10mm x 10mm chip, the test conditions are -40 ⇄ 85ºC with 24 cycles per day, and we know that the product can survive 280 cycles. However, we also need to know if this value of 280 cycles is sufficient to meet the 5-year operating condition given by 20 ⇄ 60ºC with 1 cycle per day (see the calculations below).

$$\alpha = \left(\frac{125}{40}\right)^{2.3} \left(\frac{1}{24}\right)^{0.3} exp\left[4562\left(\frac{1}{273+60} - \frac{1}{273+85}\right)\right] = 13.79$$

The product with SAC305 will survive 13.79 x 280 cycles per day = 3861 ÷ 365 = 10.5 years > 5 years.

Linear acceleration factor for Sn3Ag0.5Cu and other lead-free

Test Components and Conditions	A	B	C
(i) For CSP component with a sample size of 10 and test condition = -25 ⇆ 100oC	2.86	0.077	4532
(ii) For LCCC and CBGA with a sample size of 67 and test condition = -25 ⇆ 100oC	1.07	0.18	4286
(iii) For TSOP and TQFP with a sample size of 8 and test condition = -25 ⇆ 100oC	2.14	0.21	274

Table 1: The constants A, B, and C for Sn3Ag0.5Cu tested in various SMDs and conditions in the linear acceleration factor based on dwell-time and maximum temperature.

Lead-Free Solders	a	b	c
SAC305	2.728	0.345	1602
SAC405	3.380	0.352	2234
SAC205	2.974	0.274	1888
SAC105	3.292	0.309	1883
SAC0307	3.094	0.230	1926
SN100C	3.126	0.215	1766
SN100C-SAC305	2.833	0.240	1379
SAC105-Ni	2.517	0.302	1287
SAC107	2.738	0.341	2078

$$\alpha = \left(\frac{\Delta T_t}{\Delta T_o}\right)^{a} \left(\frac{f_o}{f_t}\right)^{b} \exp\left[c\left(\frac{1}{T_{mean,o}} - \frac{1}{T_{mean,t}}\right)\right]$$

Table 2: The constants a, b, and c for various lead-free solders in the linear acceleration factor based on frequency and mean temperature.

solders in terms of frequency and mean temperature. Osterman, et al., [11] proposed another modification of the classical Norris-Landzberg equation by replacing the maximum temperature during thermal cycling with the mean temperature during thermal cycling. The revised equation takes the following form:

$$\alpha = \left(\frac{\Delta T_t}{\Delta T_o}\right)^a \left(\frac{f_o}{f_t}\right)^b exp\left[c\left(\frac{1}{T_{mean,o}} - \frac{1}{T_{mean,t}}\right)\right]$$

where a, b, and c are the constants for the temperature range (Δ_T), frequency (f), and mean temperature during cycling (T_{mean}), respectively. For lead-free solders such as SAC305, SAC405, SAC205, SAC105, SAC0307, SN100C, SN100C-SAC305, SAC105-Ni, and SAC107, a nonlinear curve fit of the thermal cycling data of the CABGA and CTBGA on PCB at various temperature ranges such as 0⇄100ºC, -40⇄100ºC, -40⇄125ºC, 25⇄125ºC, and -15⇄125ºC, the constants a, b, and c of the above equation have been obtained by Osterman, et al., [11] and are tabulated in **Table 2**.

For example, the test conditions are: 0⇄100ºC with 24 cycles per day and find the product with SAC305 solder that survives 990 cycles. Is it sufficient for a 10-year operating condition of 20⇄60ºC with 1 cycle per day?

$$\alpha = \left(\frac{100}{40}\right)^{2.728}\left(\frac{1}{24}\right)^{0.345} exp\left[1602\left(\frac{1}{273+40} - \frac{1}{273+50}\right)\right] = 4.73$$

The product with SAC305 will survive 4.73 x 990 cycles per day = 4682.7 ÷ 365 = 12.82 years > 10 years.

Summary and recommendations

Some important results and recommendations are summarized follows:

- Reliability engineering consists of three major tasks: design for reliability, reliability testing and data analysis, and failure analysis.
- The reliability of an interconnect (e.g., solder joint) of a particular package in an electronic product is defined as the probability that the interconnect will perform its intended function for a specified period of time under a given operating condition without failure.
- The one and only way to determine the interconnect (e.g., solder joint) reliability is by reliability testing to determine the parameters of a life distribution, *F(x)*. Once *F(x)* is estimated, the reliability, failure rate, cumulative failure rate, average failure rate, mean-time-to-failure, etc., of the interconnect are readily determined.
- The life distribution, *F(x)*, is package/component dependent. Actually, it is also affected by the chip size, solder alloy, type of pastes, PCB material, PCB thickness, number of copper layers in the PCB, reflow condition, solder joint volume, voids in the solder joint, test condition, continuity measurement, number of measurements during each cycle, the data acquisition system, failure criteria, data analysis method, etc.
- For a given confidence level, the method to determine the true Weibull slope, true characteristic life, and true mean life can be found in [5].
- The procedure for determining the confidence when comparing the mean life of two difference populations can also be found in [5].
- All the papers in the literature, whether intentional or not, are dealing with liner acceleration, i.e., $x_o = \alpha x_T$ or $N_o = \alpha N_T$, where α is the linear acceleration factor. Other accelerations and factors have to be experimentally investigated for different SMDs, lead-free solder alloys, and test conditions.
- Linear acceleration factors for various lead-free solder alloys based on: a) frequency and maximum temperature, b) dwell time and maximum temperature, and c) frequency and mean temperature have been systematically presented.

References

1. J. H. Lau, (Ed.), *Solder Joint Reliability: Theory and Applications*, Van Nostrand Reinhold, New York, April 1991.
2. J. H. Lau, "Design for reliability, reliability testing and data analysis, and failure analysis of solder joints," NEPCON West Workshops, Anaheim, CA, Feb. 1990.
3. J. H. Lau, E. Schneider, T. Baker, "Shock and vibration of solder bumped flip-chip on organic coated copper boards," ASME Trans., J. of Elec. Packaging, Vol. 118, June 1996, pp. 101-104.
4. J. H. Lau, Y. Pao, *Solder Joint Reliability of BGA, CSP, Flip Chip, and Fine Pitch SMT Assemblies*, McGraw-Hill, New York, 1997.
5. J. H. Lau, N. C. Lee, *Assembly and Reliability of Lead-Free Solder Joints*, Springer, New York, 2020.
6. J. H. Lau, M. Li, M. M. Li, I. Xu, T. Chen, Z. Li, et al., "Design, materials, process, fabrication, and reliability of fan-out wafer-level packaging," IEEE Trans. on CPMT, Vol. 8, No. 6, June 2018.
7. J. Lau, W. Dauksher, P. Vianco, "Acceleration models, constitutive equations and reliability of lead-free solders and joints," IEEE/ECTC Proc., May 2003, pp. 229-236.
8. N. Pan, G. Henshall, F. Billaut, S. Dai, M. Strum, R. Lewis, et al., "An acceleration model for Sn-Ag-Cu solder joint reliability under various thermal cycle conditions," SMTA Inter. Conf. Proc., Sept. 2005, pp. 876-883.
9. J. Miremadi, G. Henshall, A. Allen, E. Benedetto, M. Roesch, "Lead-free solder-joint-reliability model enhancement," IMAPS Proc., Oct. 2009, pp. 316-323.
10. P. Lall, A. Shirgaokar, D. Arunachalam, "Norris-Landzberg acceleration factor and Goldmann constants for SAC305 lead-free electronics," ASME Trans., J. of Elec. Packaging, Vol. 134, Sept. 2012, pp. 1-8.
11. M. Osterman, "Modeling temperature cycle fatigue life of select SAC solders," SMTA Inter. Conf., Sept. 2018.

Biography

John H. Lau is the CTO at Unimicron Technology Corporation, Taoyuan City, Taiwan (ROC). He has more than 40 years of R&D and manufacturing experience in semiconductor packaging, 500 peer-reviewed papers, 30 issued and pending US patents, and 20 textbooks. He is an ASME Fellow, IEEE Fellow, and IMAPS Fellow. He earned a PhD degree from the U. of Illinois at Urbana-Champaign. Email John_Lau@unimicron.com

Register now for the 17th annual IWLPC recognized as the premier semiconductor packaging conference and exhibition focused on advanced wafer-level packaging technology.

This year's conference theme, **"Bridging the Boundaries: Wafer, Panel and Beyond"** reflects the role of advanced wafer level packaging in the enablement of 5G communications, AI, and IoT, automotive and more.

The IWLPC has always provided a dynamic environment for learning, networking and technical exchange and this year will be no different. Well, maybe a little different. The 2020 conference will be the first IWLPC to be held virtually. The conference committee has arranged for a high-quality virtual conference experience to deliver the technical content and to facilitate communication and networking. As in prior years, the conference comprises three major parts: the technical program, the professional development courses, and the technology exhibition. The technical program has three parallel tracks with approximately 50 presentations on wafer-level packaging, 3D integration, and advanced manufacturing and test technologies. This year the technical presentations will be available on demand from October 13 through October 30. A chat feature will enable attendees to interact with the speakers, exhibitors and other attendees.

The 2020 IWLPC will kick off with a keynote talk titled "Trends, Challenges, Opportunities in Advanced Packaging for Smart Computing Era" given by Dan Oh, Ph.D., Engineering VP of Test & System Package (TSP), Samsung Electronics. Dr. Oh's speech will be broadcast live at Tuesday, October 13, 2020 at 9:00 am PDT.

In addition, Jan Vardaman of TechSearch International will moderate a LIVE panel discussion entitled "Meeting Future Advanced Packaging Challenges: What's Next?"

This event will be live on Wednesday, October 14 @ 9:00am PDT.

As the industry moves into the next silicon nodes and enters the era of heterogeneous integration, packaging plays an increasingly important role. Material selection, design, and fabrication of features, inspection, test, and reliability will be critical. The industry struggles with options to achieve high-density substrate to support high-bandwidth memory (HBM) plus logic. New versions of FO-WLP are being adopted. The panel members will discuss views on the challenges and possible solutions. Don't miss this once in a lifetime international panel discussion!

Panel members are:

Tim Olson Founder and CEO	DECA	Rahul Manepall, Ph.D Sr. Director of Engineering, SPTD & Sr. Principal Engineer	intel
Tanja Braun, Ph.D Group Manager,	Fraunhofer IZM	Shin-Puu Jeng, Ph.D Director	tsmc
Hong Xie VP Research Institute for Adv. Packaging	TF 通富微電	Max Min, Ph.D Director, Package and SI/PI Architect	SAMSUNG

For more about the panelists: www.smta.org/mpage/iwlpc-panel

Finally, the technology exhibition this year will feature many leading companies from across the advanced packaging supply chain. Attendees will be able to browse the virtual exhibition hall and engage in live chat sessions with the exhibitors. We look forward to seeing you this October in cyberspace.

For additional information and conference details: www.iwlpc.com

- Conference information: Jaclyn Sarandrea – jaclyn@smta.org
- Exposition information: Mckenna Hill – mckenna@smta.org



Participating Sponsors

Platinum Sponsor

SUSS MicroTec

Gold Sponsor

EVG PacTech YES

Silver Sponsor

APPLIED MATERIALS ERS micross XPERI

Supported By

IEEE ELECTRONICS PACKAGING SOCIETY

Supporting Media

3DInCites GLOBAL SMT&PACKAGING i-Micronews MEPTEC

The seventh IEEE International Workshop on testing three-dimensional, chiplet-based, and stacked ICs, 3D & chiplets test virtual workshop and continuation of the popular 3D-TEST Workshop will be held in conjunction with ITC / Test Week 2020 on November 6, 2020.

The 3DC-TEST Workshop focuses exclusively on test of and design-for-test for three-dimensional, chiplet-based, and stacked ICs (3D-SICs), including systems-in-package (SiP), package-on-package (PoP), and 3D-SICs based on through-silicon vias (TSVs), micro-bumps, and/or interposers. While these stacked ICs offer many attractive advantages with respect to heterogeneous integration, small form-factor, high bandwidth and performance, and low power dissipation, there are many open issues with respect to testing such products.

The 3DC-TEST Workshop offers a forum to present and discuss these challenges and (emerging) solutions among researchers and practitioners alike. 3DC-TEST will take place in conjunction with the IEEE International Test Conference (ITC) and is sponsored by the IEEE Philadelphia Section in concurrence with the Test Technology Technical Council (TTTC).

Key dates and deadlines:
- Submission deadline: September 28th
- Notification of acceptance: October 1
- Early registration deadline: October 5th
- Camera-ready material: October 23rd

For additional information:
Visit: www.3dtest.tttc-events.org
Contact: Erik Jan Marinissen, imec
General Co-Chair
Tel.: +1 32 16-288755
E-mail: erik.jan.marinissen@imec.be

ADVERTISER INDEX

November December 2020
Space Close November 6th
Materials Close November 13th

For Advertising Inquiries
ads@chipscalereview.com

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION
(All Periodicals Publications Except Requester Publications)

1	Publication Title: Chip Scale Review		
2	Publication Number: 1526-1344		
3	Filing Date: 10/1/2020		
4	Issue Frequency: Bi-Monthly		
5	Number of Issues Published Annually: 6		
6	Annual Subscription Price: $125/Year, $225/year international		
7	Mailing Address of Known Office of Publication: 12455 Watsonville Road, Gilroy, CA 95020 Contact Person: Kim Newman, 408.846.8580		
8	Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer): 12455 Watsonville Road, Gilroy, CA 95020		
9	Publisher: Kim Newman PO Box 2165 Morgan Hill, CA 95038		
	Editor: Debra Vogler 74-478 Hwy 111 #363 Palm Desert, CA 92260		
	Managing Editor: Lawrence Michaels PO Box 2165 Morgan Hill, CA 95038		
10	Owner: Haley Publishing Inc., PO Box 2165 Morgan Hill, CA 95038		
11	Known Bondholders, Mortgagees and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities: None		
12	Tax Status: Has not changed during preceding 12 months		
13	Publication Title: Chip Scale Review		
14	Issue Date for Circulation Data Below: Issue Date for Circulation Data Below: July - August 2020		
15	Extent and Nature of Publication	Average No. Copies Each Issue During Preceding 12 Months to Filing Date	No. Copies of Single Issue Published Nearest to Filing Date
a	Total Number of Copies (Net Press Run)	12,331	12,100
b	Paid and/or Requested Circulation	0	0
(1)	Paid Requested Outside-County Mail Subscriptions on Form 3541	0	0
(2)	Paid In-County Subscriptions (Form 3541)	0	0
(3)	Sales Through Dealers and Carriers, Street Vendors, Counter Sales and Other Non-USPS Paid Distribution	0	0
(4)	Other Classes Mailed Through the USPS	0	0
c	Total Paid and/or Requested Circulation	0	0
d	Free Distribution by Mail	11,790	11,497
(1)	Outside-County (on Form 3541)	0	0
(2)	In-County (on Form 3541)	11,790	11,497
(3)	Other Classes Mailed Through the USPS	0	0
(4)	Free or Nominal Rate Distribution Outside the Mail	585	603
e	Free Distribution Outside the Mail	0	0
f	Total Free Distribution (Sum of 15c and 15e)	11,790	11,800
g	Copies Not Distributed	346	303
h	Total (Sum of 15f and g)	12,136	12,103
i.	Percent Paid	0	0
j.	Percent Paid and/or Requested Circulation (15C divided by 15G times 100)	0.00	0.00
16	This Statement of Ownership will be printed in the September - October 2020 issue of this publication.		
a	Paid Electronic Copies	0	0
b	Total Paid Print Copies (Line 15c) + Paid Electronic Copies (Line 16a)	0	0
c	Total Print Distribution (Line 15f) + Paid Electronic Copies (Line 16a)	0	0
d	Percent Paid (Both Print & Electronic Copies) (16b divided by 16c Ì 100)	0	0
	☐ I certify that 50% of all my distributed copies (electronic and print) are paid above a nominal price.		
17	Publication of Statement of Ownership	☒ the publication is a general publication, publication of this statement is required. Will be printed in the September October 2020 issue of this publication.	
18	Signature and Title of Editor, Publisher, Business Manager, or Owner	Kim Newman, Publisher *Kim Newman*	

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

FOR TESTING ESD-SENSITIVE DEVICES

ES-P™ SERIES

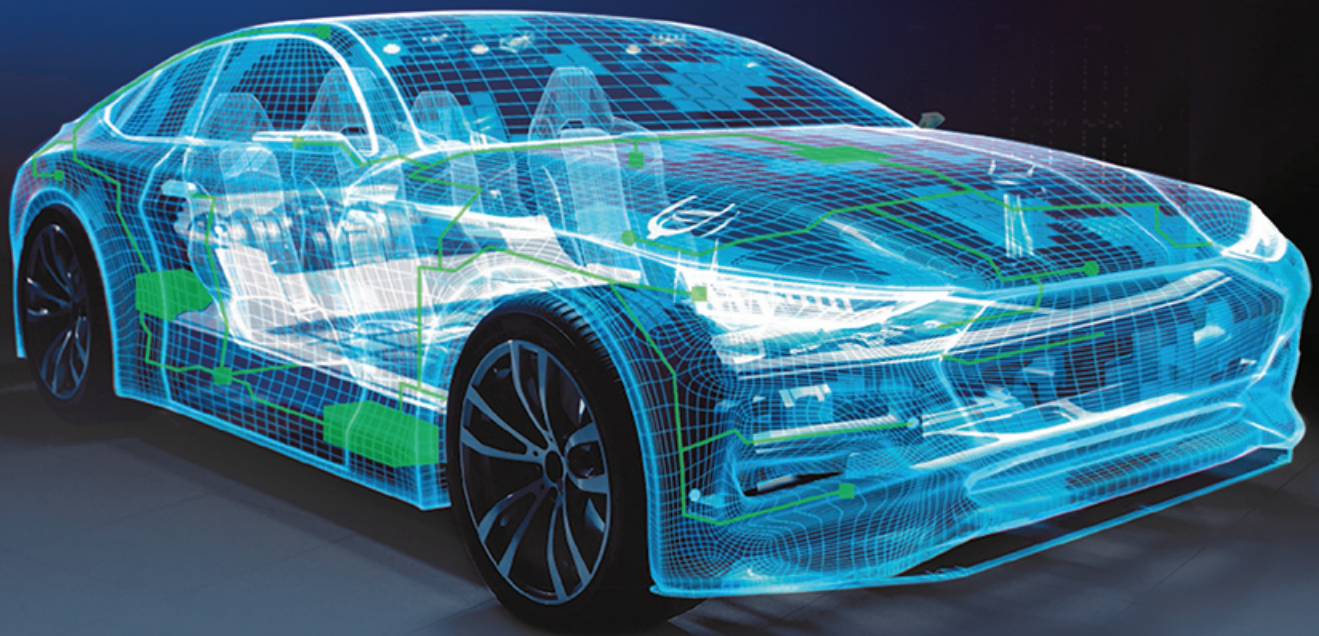
MEET THE FUTURE

Johnstech's patented ROL® technology built with ES-P™ material reduces the buildup of static electricity that may damage your device. This is ideal for devices that are susceptible to electrostatic discharge events. Take ESD protection to the next level and contact your local Johnstech sales representative for a customized test solution today.

CERTIFIED ISO 9001:2015 CERTIFIED

FOR MORE INFO, VISIT JOHNSTECH.COM

Steering the automotive electronic revolution



Applied SmartFactory® is the most comprehensive control and productivity suite in manufacturing. Our customers are increasing yield and product output, reducing cycle times and costs, and increasing profitability. Are you ready to maximize your automotive quality control strategy? Contact us at SmartFactory@amat.com

APPLIED MATERIALS®